



**UNIVERSIDAD AUTÓNOMA  
DE AGUASCALIENTES**

Centro de Ciencias Básicas

Departamento de Sistemas de Información

**TESIS**

**Aplicación de técnicas computacionales en la predicción de procesos  
industriales críticos para optimizar la calidad del producto, un paso hacia  
adelante en la industria 4.0**

PRESENTA

I.S.C. Karla Estrella Campos Campos.

Para obtener el grado de

**Maestro en Informática y Tecnologías Computacionales**

TUTOR

MIE. José de Jesús Palos García

COMITÉ TUTORAL

PhD. Luis Eduardo Bautista Villalpando

MC. Luis Enrique Arámbula Miranda

Aguascalientes, Ags., 09 de junio de 2022.

CARTA DE VOTO APROBATORIO  
INDIVIDUAL

M. en C. Jorge Martín Alférez Chávez  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *TUTOR* designado del estudiante *KARLA ESTRELLA CAMPOS CAMPOS* con ID 118839 quien realizó la tesis titulada: *APLICACIÓN DE TÉCNICAS COMPUTACIONALES EN LA PREDICCIÓN DE PROCESOS INDUSTRIALES CRÍTICOS PARA OPTIMIZAR LA CALIDAD DEL PRODUCTO, UN PASO HACIA ADELANTE EN LA INDUSTRIA 4.0*, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE  
"Se Lumen Proferre"

Aguascalientes, Ags., a día 10 de junio de 2022.



MIE. José de Jesús Palos García  
Tutor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado



UNIVERSIDAD AUTÓNOMA  
DE AGUASCALIENTES

CARTA DE VOTO APROBATORIO  
INDIVIDUAL

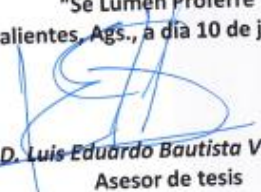
M. en C. Jorge Martín Alférez Chávez  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como **ASESOR** designado del estudiante **KARLA ESTRELLA CAMPOS CAMPOS** con ID 118839 quien realizó la tesis titulada: **APLICACIÓN DE TÉCNICAS COMPUTACIONALES EN LA PREDICCIÓN DE PROCESOS INDUSTRIALES CRÍTICOS PARA OPTIMIZAR LA CALIDAD DEL PRODUCTO, UN PASO HACIA ADELANTE EN LA INDUSTRIA 4.0**, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el **VOTO APROBATORIO**, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE  
"Se Lumen Proferre"  
Aguascalientes, Ags., a día 10 de junio de 2022.

  
PhD. Luis Eduardo Bautista Villalpando  
Asesor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

Elaborado por: Depto. Apoyo al Posgrado.  
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.  
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-FO-07  
Actualización: 01  
Emisión: 17/05/19

CARTA DE VOTO APROBATORIO  
INDIVIDUAL

M. en C. Jorge Martín Alférez Chávez  
DECANO DEL CENTRO DE CIENCIAS BÁSICAS

PRESENTE

Por medio del presente como *ASESOR* designado del estudiante *KARLA ESTRELLA CAMPOS CAMPOS* con ID *118839* quien realizó la tesis titulada: *APLICACIÓN DE TÉCNICAS COMPUTACIONALES EN LA PREDICCIÓN DE PROCESOS INDUSTRIALES CRÍTICOS PARA OPTIMIZAR LA CALIDAD DEL PRODUCTO, UN PASO HACIA ADELANTE EN LA INDUSTRIA 4.0*, un trabajo propio, innovador, relevante e inédito y con fundamento en el Artículo 175, Apartado II del Reglamento General de Docencia doy mi consentimiento de que la versión final del documento ha sido revisada y las correcciones se han incorporado apropiadamente, por lo que me permito emitir el VOTO APROBATORIO, para que ella pueda proceder a imprimirla así como continuar con el procedimiento administrativo para la obtención del grado.

Pongo lo anterior a su digna consideración y sin otro particular por el momento, me permito enviarle un cordial saludo.

ATENTAMENTE  
"Se Lumen Proferre"

Aguascalientes, Ags., a día 10 de junio de 2022.



M.C. Luis Enrique Arámbula Miranda  
Asesor de tesis

c.c.p.- Interesado  
c.c.p.- Secretaría Técnica del Programa de Posgrado

Elaborado por: Depto. Apoyo al Posgrado.  
Revisado por: Depto. Control Escolar/Depto. Gestión de Calidad.  
Aprobado por: Depto. Control Escolar/ Depto. Apoyo al Posgrado.

Código: DO-SEE-PO-07  
Actualización: 01  
Emisión: 17/05/19



DICTAMEN DE LIBERACION ACADÉMICA PARA INICIAR LOS TRAMITES DEL EXAMEN DE GRADO



Fecha de dictaminación dd/mm/aaaa: 10/06/22

**NOMBRE:** Karla Estrella Campos Campos **ID:** 118839

**PROGRAMA:** Maestría en Informática y Tecnologías Computacionales **LGAC (del posgrado):** Gestión de sistemas y tecnologías de información para mejorar competitividad, innovación y cambio organizacional.

**TIPO DE TRABAJO:** (  ) Tesis (  ) Trabajo Práctico

**TÍTULO:** Aplicación de técnicas computacionales en la predicción de procesos industriales críticos para optimizar la calidad del producto, un paso hacia adelante en la industria 4.0

**IMPACTO SOCIAL (señalar el impacto logrado):** Este trabajo proporciona técnicas de análisis de procesos industriales y reducción de variables para la optimización de los procesos productivos.

INDICAR	SI	NO	N.A. (NO APLICA)	SEGÚN CORRESPONDA:
<i>Elementos para la revisión académica del trabajo de tesis o trabajo práctico:</i>				
SI				El trabajo es congruente con los LGAC del programa de posgrado
SI				La problemática fue abordada desde un enfoque multidisciplinario
SI				Existe coherencia, continuidad y orden lógico del tema central con cada apartado
SI				Los resultados del trabajo dan respuesta a las preguntas de investigación o a la problemática que aborda
SI				Los resultados presentados en el trabajo son de gran relevancia científica, tecnológica o profesional según el área
SI				El trabajo muestra más de una aportación original al conocimiento de su área
SI				Las aportaciones responden a los problemas prioritarios del país
SI				Generó transferencia del conocimiento o tecnológica
SI				Cumple con la ética para la investigación (reporte de la herramienta antiplagio)
<i>El egresado cumple con lo siguiente:</i>				
SI				Cumple con lo señalado por el Reglamento General de Docencia
SI				Cumple con los requisitos señalados en el plan de estudios (créditos curriculares, optativos, actividades complementarias, estancia, predoctoral, etc)
SI				Cuenta con los votos aprobatorios del comité tutorial, en caso de los posgrados profesionales si tiene solo tutor podrá liberar solo el tutor
SI				Cuenta con la carta de satisfacción del Usuario
SI				Coincide con el título y objetivo registrado
SI				Tiene congruencia con cuerpos académicos
SI				Tiene el CVU del Conacyt actualizado
NO				Tiene el artículo aceptado o publicado y cumple con los requisitos institucionales (en caso que proceda)
<i>En caso de Tesis por artículos científicos publicados</i>				
N.A.				Aceptación o Publicación de los artículos según el nivel del programa
N.A.				El estudiante es el primer autor
N.A.				El autor de correspondencia es el Tutor del Núcleo Académico Básico
N.A.				En los artículos se ven reflejados los objetivos de la tesis, ya que son producto de este trabajo de investigación.
N.A.				Los artículos integran los capítulos de la tesis y se presentan en el idioma en que fueron publicados
N.A.				La aceptación o publicación de los artículos en revistas indexadas de alto impacto

Con base a estos criterios, se autoriza se continúen con los trámites de titulación y programación del examen de grado: SI  No

FIRMAS

Elaboró: **\* NOMBRE Y FIRMA DEL CONSEJERO SEGÚN LA LGAC DE ADESCRIPCIÓN:** Dr. Cesar Eduardo Velázquez Amador

**NOMBRE Y FIRMA DEL SECRETARIO TÉCNICO:** MITC. Jorge Eduardo Macías Luviano

**Revisó:** **NOMBRE Y FIRMA DEL SECRETARIO DE INVESTIGACIÓN Y POSGRADO:** Dra. Haydee Martínez Rivalcaba

**Autorizó:** **NOMBRE Y FIRMA DEL DECANO:** M. en C. Jorge Martín Alférez Chávez

**Nota: procede el trámite para el Depto. de Apoyo al Posgrado**  
 En cumplimiento con el Art. 205C del Reglamento General de Docencia que a la letra señala entre las funciones del Consejo Académico: ... Quitar la eficiencia terminal del programa de posgrado y el Art. 205F las funciones del Secretario Técnico, llevar el seguimiento de los alumnos.

## Agradecimientos

A Dios y a la Virgen, por estar presentes, a cada momento y en todo lugar.

A Jesús, por acompañarme siempre y desde el inicio en este sueño.

A mis compañeros y amigos Fernando Vázquez, Sergio Galvis, Angel Estrada, Nancy Correa y Christian Hernández por formar parte de este camino desde el principio.

A mi tutor y maestro MIE. Jesús Palos, por alentarme en este proceso y brindarme su apoyo en cada etapa.

Al Dr. Eduardo Bautista, por su dedicación en este proyecto y sus valiosas aportaciones en la codirección de esta investigación.

Al MC. Enrique Arámbula por su apoyo en la realización del presente trabajo.

A la empresa de manufactura por el conocimiento ofrecido y la oportunidad de abordar la investigación a través de sus procesos, personal de Ingeniería y de las diferentes áreas.

A la Universidad Autónoma de Aguascalientes por las oportunidades brindadas para continuar con mi desarrollo académico.

A mis padres Carlos y Ester  
y a mis hermanos Carlos y Aldo.

¡Gracias!

Dedicatorias



A mi Luz, mi guía...

## Índice General

Índice de tablas .....	3
Índice de Figuras .....	5
Acrónimos .....	10
Resumen .....	11
Abstract .....	12
Introducción .....	13
Objetivos de la Investigación .....	14
Objetivo General .....	14
Objetivos Específicos .....	14
Tipo de Investigación .....	15
Justificación .....	15
Capítulo 1. Antecedentes.....	18
1.1 Industria 4.0 .....	18
1.2 Procesos Industriales.....	23
1.3 Predicción de Procesos .....	24
1.4 Trabajos Relacionados.....	25
Capítulo 2. Fundamentos.....	26
2.1 Calidad .....	26
2.2 Aprendizaje automático .....	29
2.3 Minería de datos .....	31
2.4 Control de procesos .....	33
Capítulo 3. Identificación y selección de un proceso Industrial .....	34
3.1 Moldeo .....	34
3.2 Tecnología de Montaje Superficial.....	46
3.3 Ensamblaje de Tarjetas con componentes. ....	49



3.4	Selección del proceso: Tecnología de Montaje Superficial.....	54
3.4.1	Descripción del proceso: Impresión de Soldadura en Pasta – Inspección de Impresión.....	56
3.4.2	Selección de modelo.....	61
3.4.3	Selección de línea de producción .....	62
3.4.4	Descripción de defectos generados por el proceso.....	63
Capítulo 4.	“Voz del proceso”.....	66
4.1	Generación de datos.....	66
4.2	Recolección de datos.....	71
4.3	Descripción de datos.....	74
4.3.1	Entradas .....	74
4.3.2	Salidas .....	79
Capítulo 5.	Modelos estadísticos y de aprendizaje automático.....	86
5.1	Modelado de datos .....	86
5.1.1	Análisis de regresión.....	86
5.2	Uso de aprendizaje automático .....	95
5.2.1	Normalización de los datos.....	96
5.2.2	Técnica de selección de variables: Medias y Varianzas.....	96
Capítulo 6.	Aplicación de técnicas computacionales en el proceso de Tecnología de Montaje superficial.....	100
6.1	Análisis de datos mediante el uso de técnicas estadísticas .....	100
6.2	Aplicación de técnica de selección de variables: Medias y Varianzas .....	118
6.2.1	Implementación del Algoritmo.....	123
6.2.2	Escalabilidad.....	141
Discusión de Resultados	.....	143
Conclusiones	.....	146
Glosario	.....	148
Referencias Bibliográficas	.....	149

Índice de tablas

Tabla 1. Variables significativas para el proceso de soldadura de ola..... 52

Tabla 2. Parámetros de entrada. .... 74

Tabla 3. Estadísticos descriptivos para datos de entrada..... 75

Tabla 4. Datos de inspección de impresión. .... 79

Tabla 5. Estadísticos descriptivos para datos de salida. .... 82

Tabla 6. Conjunto de datos ejemplo..... 97

Tabla 7. Definición de subconjuntos de datos..... 97

Tabla 8. Fórmulas para calcular la media de los datos. .... 98

Tabla 9. Fórmulas para calcular la varianza. .... 98

Tabla 10. Desviación estándar por componente (valores más altos)..... 101

Tabla 11. Gráficas Individuales para parámetros de Impresión. .... 104

Tabla 12. Definición de clases para porcentaje de volumen. .... 118

Tabla 13. Definición del conjunto de datos para aplicación de técnica de selección de variables.  
..... 119

Tabla 14. Clasificación de datos..... 120

Tabla 15. Definición de subconjuntos de datos..... 121

Tabla 16. Planteamiento de ecuaciones para comparación de varianzas..... 122

Tabla 17. Planteamiento de ecuaciones para prueba de medias y varianzas..... 123

Tabla 18. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 50 tarjetas..... 127

Tabla 19. Parámetros candidatos a reducción para una muestra de 50 tarjetas..... 128

Tabla 20. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 100 tarjetas..... 131

Tabla 21. Parámetros candidatos a reducción para una muestra de 100 tarjetas..... 131

Tabla 22. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 200 tarjetas..... 133

Tabla 23. Parámetros candidatos a reducción para una muestra de 200 tarjetas..... 133

Tabla 24. Resultados obtenidos para varianzas y evaluación de medias y varianzas para cada muestra..... 135

Tabla 25. Variables candidatas a reducción..... 136

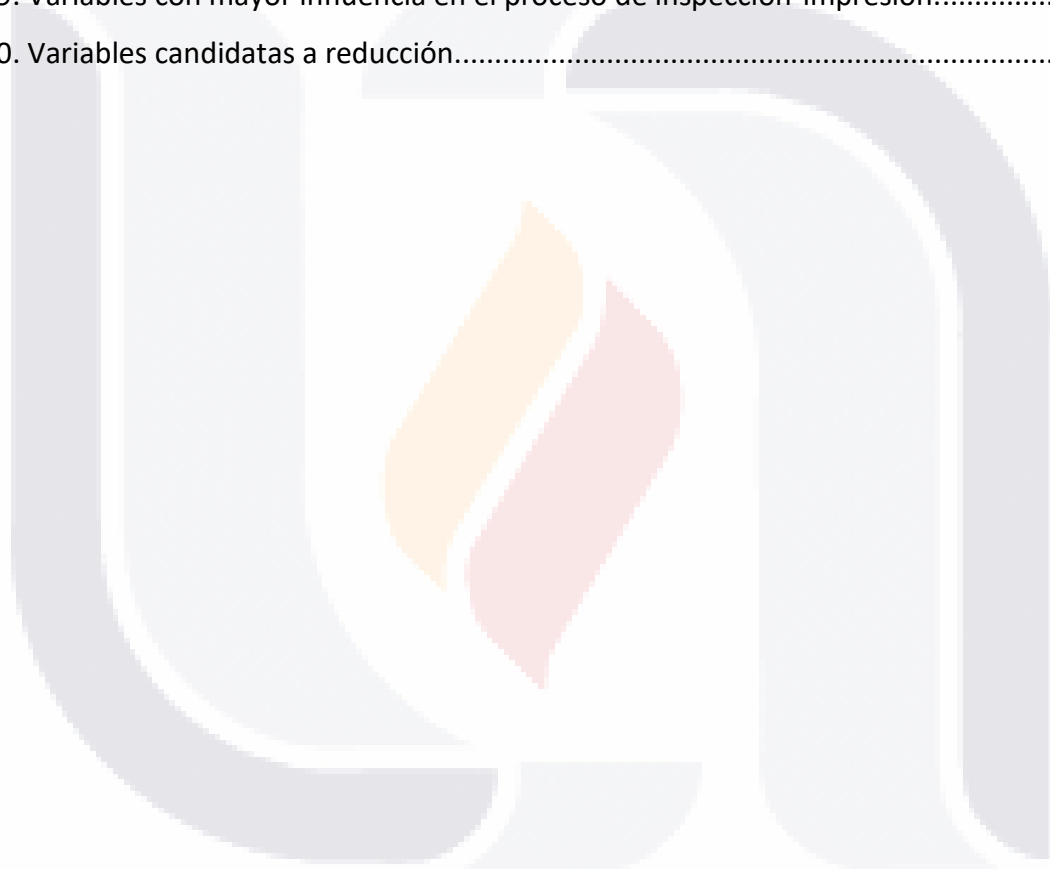
Tabla 26. Variables con mayor influencia en el proceso. .... 136

Tabla 27. Periodos utilizados para la generación de dataset de producción no consecutiva.... 137

Tabla 28. Resultados obtenidos para varianzas y evaluación de medias y varianzas para muestras de diferente periodo..... 139

Tabla 29. Variables con mayor influencia en el proceso de inspección-impresión..... 140

Tabla 30. Variables candidatas a reducción..... 140



## Índice de Figuras

Figura 1. Revoluciones Industriales (Ortiz Clavijo et al., 2018) .....	18
Figura 2. Características de la Industria 4.0 (Naya, 2018) .....	19
Figura 3. Representación Esquemática de un proceso (Pardo, 2012).....	23
Figura 4. Esfuerzo en proceso de minería de datos(Kantardzic, 2011). .....	32
Figura 5. Esquema general del proceso de moldeo. Elaboración Propia. ....	35
Figura 6. Dimensionado fuera de especificación.....	41
Figura 7. Registro de pieza fuera de especificación.....	41
Figura 8. Fisura.....	42
Figura 9. Grasa en la pieza plástica.....	42
Figura 10. Rebaba. ....	43
Figura 11. Hilillo. ....	43
Figura 12. Hilillo, reparable.....	43
Figura 13. Faltante de plástico.....	44
Figura 14. Pieza incompleta.....	44
Figura 15. Malformación en la pieza.....	44
Figura 16. Líneas de flujo. ....	45
Figura 17. Puntos de contaminación en película plástica.....	45
Figura 18. Proceso general de Tecnología de Montaje Superficial. Elaboración Propia. ....	47
Figura 19. Proceso general de Ensamblaje de Tarjetas con Componentes (PTH). Elaboración Propia.....	50
Figura 20. Resultado de prueba de penetración de flux, baja penetración. ....	52
Figura 21. Resultado de prueba de penetración de flux, exceso de flux.....	53
Figura 22. Residuos de flux. ....	53
Figura 23. Crecimiento dendrita.....	53
Figura 24. Corrosión.....	54
Figura 25. Tecnología de Montaje Superficial: Impresión de pasta e inspección. Elaboración Propia.....	55

Figura 26. Esquema general del proceso de impresión de pasta e Inspección. Elaboración Propia.  
 ..... 56

Figura 27. Impresión de Pasta utilizando plantillas. .... 61

Figura 28. Esquemmatización de una línea de producción. Elaboración Propia. .... 63

Figura 29. Insuficiencia. .... 63

Figura 30. Daño del lado A. .... 64

Figura 31. Exceso de soldadura..... 64

Figura 32. Desplazamiento. .... 65

Figura 33. Interfaz de configuración de salida de datos de Impresión de Pasta. .... 67

Figura 34. Generación de archivo con datos de impresión de pasta. .... 68

Figura 35. Log generado por la Impresora de pasta. .... 68

Figura 36. Interfaz de configuración de salida de datos de Inspección de Impresión de Pasta... 69

Figura 37. Log generado como resultado de la inspección de la impresión de pasta..... 70

Figura 38. Transferencia de datos desde Impresora de Pasta..... 73

Figura 39. Transferencia de datos desde equipo de Inspección de Impresión. .... 73

Figura 40. Esquemmatización de proceso de recolección de datos. Elaboración Propia. .... 73

Figura 41. Histograma de la temperatura..... 75

Figura 42. Histograma de la presión frontal. .... 76

Figura 43. Histograma de la presión trasera..... 76

Figura 44. Histograma de la humedad..... 76

Figura 45. Histograma de la velocidad de separación. .... 77

Figura 46. Histograma del tiempo de ciclo de impresión. .... 77

Figura 47. Histograma de la posición de la mesa. .... 78

Figura 48. Histograma de la velocidad de impresión frontal..... 78

Figura 49. Histograma de la velocidad de impresión trasera. .... 78

Figura 50. Componente u30 sin impresión de pasta. .... 80

Figura 51. Relación de pads con parámetros igual a cero. .... 80

Figura 52. Relación de datos sin referencia a un componente. .... 81

Figura 53. Log de inspección de impresión sin componentes referenciados..... 81

Figura 54. Histograma del tiempo de inspección de impresión. .... 82

Figura 55. Histograma de la altura..... 82

Figura 56. Histograma del Área. .... 83

Figura 57. Histograma del porcentaje de área. .... 83

Figura 58. Histograma del volumen..... 84

Figura 59. Histograma del porcentaje de volumen. .... 84

Figura 60. Histograma del desplazamiento en X. .... 85

Figura 61. Histograma del desplazamiento en Y. .... 85

Figura 62. Gráfica del modelo de regresión lineal simple. Ejemplo. .... 88

Figura 63. Gráficas de las funciones polinomiales..... 92

Figura 64. Promedio del porcentaje de volumen de componentes r178 por cada tarjeta..... 101

Figura 65. Promedio del porcentaje de volumen de componentes r188 por cada tarjeta..... 102

Figura 66. Promedio del porcentaje de volumen de componentes r2 por cada tarjeta..... 102

Figura 67. Promedio del porcentaje de volumen de componentes r201 por cada tarjeta..... 102

Figura 68. Promedio del porcentaje de volumen de componentes fb4 por cada tarjeta..... 102

Figura 69. Gráfica de temperatura. .... 103

Figura 70. Gráfica de temperatura para un periodo extendido. .... 103

Figura 71. Modelo de regresión lineal múltiple y detalle de coeficientes para porcentaje de volumen y parámetros de Impresión ..... 106

Figura 72. Resumen del análisis de regresión lineal múltiple para porcentaje de volumen y parámetros de impresión. .... 107

Figura 73. Análisis de regresión lineal múltiple para porcentaje de volumen y parámetros de Impresión: temperatura, humedad, posición de la mesa. .... 108

Figura 74. Análisis de regresión lineal simple para porcentaje de volumen y humedad..... 109

Figura 75. Gráfica de línea ajustada para el modelo de regresión lineal simple. Porcentaje de volumen y humedad. .... 110

Figura 76. Análisis de regresión cuadrática para porcentaje de volumen y humedad. .... 110

Figura 77. Gráfica de línea ajustada para el modelo de regresión cuadrática. Porcentaje de volumen y humedad. .... 111

Figura 78. Análisis de regresión polinomial de grado 3 para porcentaje de volumen y humedad.  
 ..... 111

Figura 79. Gráfica de línea ajustada para el modelo de regresión polinomial de grado 3. Porcentaje de volumen y humedad..... 112

Figura 80. Modelo de regresión lineal simple. Porcentaje de volumen y temperatura..... 113

Figura 81. Modelo de regresión polinomial de grado 2. Porcentaje de Volumen y Temperatura.  
 ..... 113

Figura 82. Modelo de regresión polinomial de grado 3. Porcentaje de volumen y temperatura.  
 ..... 113

Figura 83. Modelo de regresión lineal simple. Porcentaje de volumen y posición de la mesa. 114

Figura 84. Modelo de regresión polinomial de grado 2. Porcentaje de volumen y posición de la mesa..... 114

Figura 85. Modelo de regresión polinomial de grado 3. Porcentaje de volumen y posición de la mesa..... 115

Figura 86. Varianzas para 50 tarjetas..... 125

Figura 87. Medias y varianzas para 50 tarjetas. .... 126

Figura 88. Varianzas para 50 tarjetas con clasificación de 65 a 135% de volumen. .... 126

Figura 89. Medias y varianzas para 50 tarjetas con clasificación de 65 a 135% de volumen. ... 127

Figura 90. Varianzas para 100 tarjetas. .... 129

Figura 91. Medias y varianzas para 100 tarjetas. .... 129

Figura 92. Varianzas para 100 tarjetas con clasificación de 65 a 135% de volumen. .... 130

Figura 93. Medias y varianzas para 100 tarjetas con clasificación de 65 a 135% de volumen. . 130

Figura 94. Varianzas y medias y varianzas para 200 tarjetas. .... 132

Figura 95. Varianzas y medias y varianzas para 200 tarjetas con clasificación de 65 a 135% de volumen. .... 132

Figura 96. Varianzas y medias y varianzas para 1000 tarjetas. .... 134

Figura 97. Varianzas y medias y varianzas para 2000 tarjetas. .... 134

Figura 98. Gráficas de varianzas y medias y varianzas para 48 tarjetas electrónicas no consecutivas..... 137

Figura 99. Gráficas de varianzas y medias y varianzas para 96 tarjetas electrónicas no consecutivas..... 137

Figura 100. Gráficas de varianzas y medias y varianzas para 200 tarjetas electrónicas no consecutivas..... 138

Figura 101. Gráficas de varianzas y medias y varianzas para 1000 tarjetas electrónicas no consecutivas..... 138

Figura 102. Gráficas de varianzas y medias y varianzas para 2000 tarjetas electrónicas no consecutivas..... 138





## Acrónimos

CMM	<i>Coordinate Measuring Machine</i> (Máquina de medición por coordenadas).
FCT	<i>Functional Circuit Test</i> (Prueba de circuito Funcional).
IBM	<i>International Business Machines Corporation</i> (Empresa multinacional de tecnología, consultoría e innovación).
ICT	<i>In Circuit Test</i> (Prueba de Circuito)
IoT	<i>Internet of Things</i> (Internet de las cosas).
IP	<i>Internet Protocol</i> (Protocolo de Internet).
LAN	<i>Local Area Network</i> (Red de área local).
OCAP	<i>Out Of Control Action Plan</i> (Plan de Acción fuera de Control).
OGP	<i>Optical Gaging Products</i> (Sistema de Medición Dimensional)
PCB	<i>Printed Circuit Board</i> (Tarjeta/Placa de Circuito Impreso).
PCBA	<i>Printed Circuit Board Assembly</i> (Ensamblaje de Tarjeta/Placa de Circuito Impreso)
PTH	<i>Pass Through Hole</i> (Pasar a través del agujero, proceso en el cual se colocan componentes de lado a lado de la tarjeta sobre sus orificios).
SMT	<i>Surface Mount Technology</i> (Tecnología de Montaje Superficial).
TI	Tecnología de la Información.
UPH	<i>Unit per Hour</i> (Cantidad de unidades producidas en una hora).
VLAN	<i>Virtual LAN</i> (Red de área local virtual).

## Resumen

El término industria 4.0 es mencionado con frecuencia dentro de las grandes compañías dedicadas a la producción, sin embargo, existen procesos y equipos para los que no existe conectividad alguna, por lo que, no existe visibilidad de los datos si es que son generados.

Una de las tareas importantes del presente trabajo, es computarizar datos generados por los procesos industriales, establecer conectividad con otros sistemas de la empresa para la obtención automática de los datos, con la finalidad de reconocer el proceso a través de ellos, y a través de la visibilidad analizar el proceso de Tecnología de Montaje Superficial.

El reconocimiento de algunos procesos industriales permite identificar qué datos es necesario recolectar del proceso a través de la identificación de factores críticos y cuáles de ellos pueden o deben ser computarizados para con su análisis, simplificar las variables que intervienen en el con el objetivo de mejorar y optimizar los procesos.

Contrastar técnicas estadísticas con técnicas de minería de datos, permite ofrecer a los expertos de procesos, alternativas para encaminar los procesos a una mejora continua, convergiendo en una mejor calidad de los productos generados por los procesos y la optimización de recursos que intervienen en ellos.

La selección de variables significativas en los procesos es sin duda, una tarea necesaria para la aplicación de futuras técnicas importantes para el análisis de datos, ya que cualquier proceso, por sencillo que parezca es producto de la interacción de múltiples factores bajo diversas condiciones.

## Abstract

Nowadays Industry 4.0 is frequently mentioned in large companies focused on production of goods, however, it exists processes and machines which do not interact with other components, this is why there is not data generated when it is supposed to exist.

One of the main objectives of the current research work, is to compute data created by industrial processes establishing connectivity relation with other systems in the organization to gather data automatically. The final purpose to recognize the process through data and perform analysis of Superficial Mountain Technology process making valuable data available.

Know some industrial processes allow us to identify data that is needed to collect, based on the critical parameters detection and which of them should be computed to be analyzed to simplify variables that impact the process, following the objective to improve and optimize processes.

Use of statistical techniques and machine learning, enable us to provide to process experts, ways to drive process improvements, resulting in product quality and resources optimizations for the processes.

Features selection of critical parameters is an essential task for future techniques to be applied to perform data analysis, because all processes, even if they seem simple, is the result of many factors which interacting different conditions.

## Introducción

Hoy en día, existe un gran número de productos disponibles en los diferentes mercados, de diversas índoles, que cumplen diferentes propósitos, desde cubrir las necesidades básicas hasta mejorar la calidad de vida de los seres humanos, la industria alimentaria, farmacéutica, electrónica, médica, son ejemplos de diversos tipos de industrias cuyo propósito principal es la producción.

Día con día los procesos industriales requieren ser mejorados, la definición adecuada del proceso permite obtener mejores resultados en los productos, evitando grandes cantidades de defectos que impactan en procesos posteriores, tiempo de reprocesamiento y recursos.

La integración vertical es de suma importancia en grandes empresas que generan unidades básicas para la construcción de ensamblajes de mayor tamaño.

Asegurar la calidad de estos productos, facilita los procesos subsecuentes, y contribuye a los buenos resultados, que, se traducen en beneficios económicos para las empresas y confiabilidad.

La caracterización de un proceso industrial, a través de la obtención de datos generados por el proceso mismo, es sin duda, una aportación significativa a los diferentes procesos productivos, encontrando una forma sencilla de definir el proceso, mediante la identificación de factores críticos para este, continuando los avances en la implementación de la industria 4.0 en las medianas y grandes compañías.

Así pues, utilizar los recursos existentes para lograr el conocimiento del proceso además de buscar la estandarización, ha sido uno de los grandes retos de la presente investigación.

Dado este escenario, se ha iniciado con la identificación de varios procesos industriales, en los cuales, es posible aplicar diversas técnicas computacionales para describir los mismos, proponiendo técnicas para recolección, limpieza, transformación y análisis de datos.

Procesos como moldeo, impresión de pasta y ensamblaje de componentes, han sido reconocidos como indispensables en los procesos de la compañía que se tomó como base para el desarrollo del presente trabajo.

La selección de un proceso particular ha facilitado el desarrollo de una serie de instrucciones que definan el análisis de otros procesos de importancia para la empresa tomando el proceso seleccionado como base.

La aplicación de técnicas de análisis de datos como la de aprendizaje automático representa una gran ventaja respecto a las técnicas estadísticas tradicionales para el análisis de los procesos.

## Objetivos de la Investigación

### Objetivo General

- ✓ Analizar un proceso industrial con la finalidad de identificar factores significativos que contribuyan a mejorar la calidad del producto, implementando diversas técnicas computacionales, integrando elementos de la industria 4.0 en la empresa.

### Objetivos Específicos

- ✓ Describir un proceso industrial: identificar los recursos necesarios, variables de entrada y salidas del proceso, causas de variación y metodología para la producción de una o gran cantidad de piezas.
- ✓ Proporcionar técnicas computacionales de utilidad para el reconocimiento del proceso.
- ✓ Realizar el análisis del proceso, para reducir el número de variables, tomando en cuenta las más significativas para el proceso, descartando las que no presentan impacto en las salidas, haciendo uso de aprendizaje automático.

## Tipo de Investigación

El enfoque de la investigación es cuantitativo, ya que se busca aplicar métodos precisos encaminados a la predicción de los procesos y mediante la réplica de las condiciones favorables, obtener siempre los resultados que nos ayuden a optimizar la calidad de los productos.

Los datos a utilizar serán cuantitativos, ya que requerirán valores numéricos, se utilizarán Series de Tiempo, puesto que la observación en varios periodos, bajo diferentes condiciones permitirá identificar las diferencias entre los datos y con ello, las variables significativas para el proceso; correlacional, se identificará la relación entre variables de entrada y salida, enfocándose en aquellas que generen valor para optimizar los resultados. Los datos serán proporcionados por diferentes métodos de adquisición de datos, de forma experimental en tanto el proceso lo permita, obteniendo distintos valores de las variables de entrada, para determinar los efectos de las variables consideradas significativas para el proceso como salidas propias y entradas a otros procesos.

## Justificación

Mucho se ha hablado sobre la industria 4.0 dentro las empresas, sobre todo, en aquellas empresas de gama mundial.

“Numerosos sectores se han tenido que enfrentar ya a un cambio tecnológico acelerado que ha supuesto la ruptura de su cadena de valor tradicional”. (Palao Gil, 2016).

En tiempos actuales, el valor de una empresa radica en la facilidad de adaptarse a los cambios externos, y en el caso de manufactura, también en mejorar sus procesos de producción con la finalidad de obtener mejores ganancias empleando la menor cantidad de recursos. Mejorar la calidad de los productos, al obtener “0 defectos”, es un factor importante en la evolución del proceso productivo.

La empresa, está obligada a ser innovadora si quiere sobrevivir. (Castells & Pasola, 2004).

La industria 4.0, puede hacerse presente en diversidad de procesos en una empresa, pero el valor radica en permitir la escalabilidad de la solución a otros procesos y por qué no, en otras empresas.

Si bien, algunas aplicaciones incluyen uno o varias tecnologías y diferentes ramas de la industria como la alimentaria (Guerrero Cano et al., 2019) o se enfocan en un proceso como el de galvanizado (Martínez-de-Pisón et al., 2007), no se ha considerado la aplicación de aprendizaje automático para predecirlos.

El resultado esperado, es, al final de esta investigación, dar la pauta para la optimización del proceso de montaje superficial de componentes electrónicos mediante el uso de técnicas computacionales como minería de datos, aprendizaje automático, análisis predictivo, entre otras, como base para la predicción del proceso, que permita mejorar la calidad del producto como resultado del proceso y con ello, la calidad de los productos obtenidos mediante la integración en otros procesos.

Además, se pretende, sentar las bases para el análisis de otros procesos industriales, mediante la especificación del método a seguir, facilitando el desarrollo de los algoritmos para ser aplicado en otros procesos, contribuyendo a la definición de las bases para el complejo mundo de la predicción.

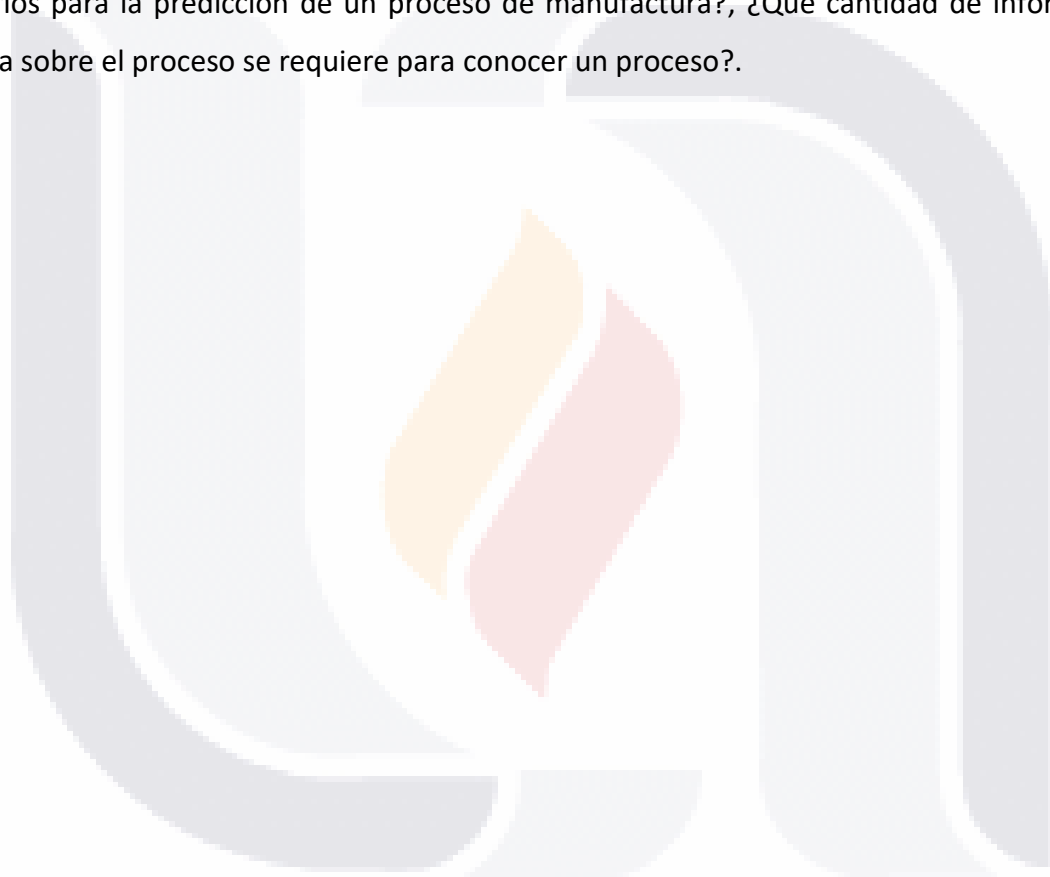
Contar con datos, para el análisis del proceso y la generación de los algoritmos, además de la selección de técnicas, es imprescindible en el desarrollo del presente trabajo, por lo que, se requiere el apoyo de ingeniería de procesos de la empresa y la intervención del personal técnico, operario y de calidad de la empresa.

La colaboración y retroalimentación es de vital importancia para llegar al objetivo general, y con ello, aportar un valor significativo a la compañía.

Aplicar aprendizaje automático en al menos un proceso crítico para la empresa, y además sentar las bases para escalarlo a otros, resulta un gran paso localmente, que puede ser compartido con otras compañías a lo largo de las diferentes plantas con que cuenta la empresa a nivel global, ya que, los mismos procesos son ejecutados en diferentes lugares, y por lo general, los defectos encontrados presentan similitudes. Además de contribuir con otras empresas del mismo giro.

La aportación de técnicas de recolección de datos, escalables a diversos procesos, permitirá además la integración con otros sistemas dentro de la empresa, facilitando el reconocimiento del proceso en tiempo real.

Algunas preguntas, que anteceden al presente trabajo, pretenden dirigir la investigación hacia la búsqueda de resultados: ¿Es posible aplicar aprendizaje automático para la optimización de un proceso de manufactura?, ¿Qué datos se requieren conocer para predecir un proceso de manufactura?, ¿Cómo se pueden identificar dichos datos?, ¿Cómo se pueden obtener los datos necesarios para la predicción de un proceso de manufactura?, ¿Qué cantidad de información histórica sobre el proceso se requiere para conocer un proceso?.





## Capítulo 1. Antecedentes

El avance tecnológico y computacional está presente en cualquier parte, aportando en la mayoría de los casos, alternativas de mejora en cualquier aplicación.

La industria, no está exenta de estas ventajas, pues este desarrollo ha tenido un avance importante en el área de manufactura, con la máquina de vapor y mecanización de los procesos, producción en masa, automatización y robótica y hoy en día la industria 4.0 (Industry et al., 2017).

### 1.1 Industria 4.0

La Industria 4.0 es la cuarta de una serie de revoluciones industriales (Figura 1); la primera introdujo los equipos mecánicos impulsados por agua y energía de vapor a la industria; la segunda, la producción en masa impulsada por el uso de energía eléctrica y el concepto de división del trabajo; y la tercera, por la producción automatizada gracias al uso de la electrónica y la informática. La cuarta revolución se caracteriza por las múltiples formas en que la tecnología conectada se integra con organizaciones, personas y activos para -luego de procesar información- orientar la acción inteligente. (Meneses Guzmán & Hernández Granados, 2020).

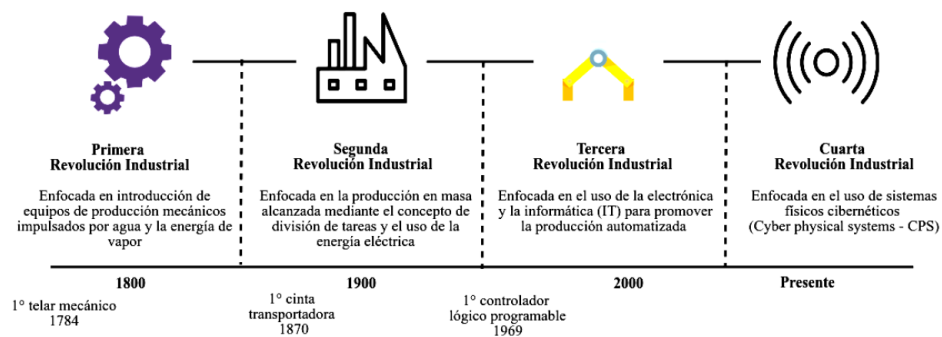


Figura 1. *Revoluciones Industriales* (Ortiz Clavijo et al., 2018)

Son nueve las tecnologías que están transformando la producción industrial, nombrándolas características, pilares o tecnologías (Figura 2).



Figura 2. Características de la Industria 4.0 (Naya, 2018)

- **Robots autónomos**

Todas las definiciones de robot industrial aceptadas actualmente consideran que es un brazo mecánico con capacidad de manipulación que incorpora un control más o menos complejo. Un sistema robotizado o robot, es un concepto más amplio pues engloba todos aquellos dispositivos construidos para realizar tareas de forma automática en sustitución de un ser humano y que son capaces de interactuar con otros robots o con humanos. (Garrell & Guilera, 2019).

La primera generación de robots son los brazos manipuladores, que repiten secuencialmente la tarea programada y no tienen en cuenta las posibles alteraciones de su entorno.

Los robots de segunda generación adquieren información limitada de su entorno y actúan en consecuencia.

La tercera generación, está formada por robots inteligentes, que se programan mediante el uso de un lenguaje natural y poseen la capacidad de auto programarse y adaptarse en tiempo real a entornos cambiantes para la planificación automática de sus tareas. Una variante de esta generación son los robots colaborativos, que controlan la seguridad del contacto físico con humanos y cualquier otro tipo de contactos.

Los robots serán cada vez más autónomos porque estarán dotados de inteligencia artificial y podrán hacer frente a tareas más complejas que impliquen eventuales tomas de

decisiones que actualmente están en manos exclusivas de humanos. Tendrán cada vez mayor interacción con otros robots y con humanos, estableciendo trabajo colaborativo y con capacidad de aprendizaje por las experiencias. Estarán interconectados para que puedan trabajar juntos y ajustar automáticamente sus acciones para adaptarse al siguiente producto inacabado en línea.

Las aplicaciones de robots humanoides están presentes en diversos sectores y empresas.

- **Simulación**

Permite ajustar y representar virtualmente el funcionamiento de un conjunto de máquinas, procesos y personas en tiempo real antes de ser puestos en marcha, lo que ayuda a prevenir posibles problemas, ahorrar tiempo y evaluar el resultado final en un entorno controlado. Es decir, permite reducir los costos asociados a procesos de aprendizaje de "prueba y error" mediante una representación virtual para el diseño de nuevos productos, o bien probar distintas configuraciones en las operaciones de la planta productiva. Como ejemplo, los operadores pueden probar distintas configuraciones hasta lograr una "configuración virtual óptima" que será luego plasmada en la línea física de producción. Además, las experiencias obtenidas en el mundo real servirán para mejorar el entorno virtual, generando una asociación de colaboración entre la planta física y su representación virtual. (Basco et al., 2018).

- **Sistemas de Integración Horizontales y Verticales**

Se refiere a la integración mediante redes universales de datos entre empresas, departamentos, funciones y tareas que evolucionan y permiten cadenas de valor verdaderamente automatizadas. (Meneses Guzmán & Hernández Granados, 2020).

- **Internet Of Things / Internet de las cosas**

El internet de las cosas (IoT) es definido como una red de objetos físicos conectados a través de internet, los cuales logran interactuar vía sistemas embebidos, redes de comunicación, mecanismos de computación de respaldo y aplicaciones típicamente en la nube. Permite a los objetos comunicarse entre sí, acceder a información de internet, capturar, almacenar y recuperar

datos e interactuar con usuarios humanos, así como con otros sistemas y aplicaciones, creando ambientes cada vez más conectados y más inteligentes (Muñoz, 2019).

- **Ciberseguridad**

Es fundamental para que todas las demás tecnologías logren una adecuada penetración en esta fase de digitalización. La evolución hacia una industria inteligente y la integración creciente de los actores de las cadenas de valor a través de internet, la computación en la nube y las plataformas digitales obliga a desarrollar mecanismos de la ciberseguridad en los entornos industriales. En la medida en que sean más los dispositivos, máquinas y personas conectadas, se valorará la oferta de herramientas preventivas que permitan detectar, anticipar y neutralizar amenazas sobre los sistemas de información de las empresas. (Basco et al., 2018).

- **La nube**

Ofrece almacenamiento, acceso y uso de servicios informáticos en línea. Puede expresarse en tres niveles diferentes, según el servicio provisto: infraestructura como servicio, plataforma como servicio y software como servicio. Esta tecnología permite a las empresas acceder a los recursos informáticos de una manera flexible con un bajo esfuerzo administrativo y desde distintos dispositivos, ofreciendo agilidad, interoperabilidad y escalabilidad. Esto es clave para aplicaciones industriales con elevados requerimientos informáticos. (Basco et al., 2018)

- **Impresión 3D**

La tecnología de fabricación aditiva presenta también un gran atractivo para la industria por sus características únicas, como la fabricación de piezas con geometría compleja, personalización en el diseño, y reducción de desperdicio de material. La Asociación Americana para Pruebas y Materiales (ASTM por sus siglas en inglés) indica que se puede usar impresión en 3D como sinónimo de fabricación aditiva en contextos no técnicos, ya que impresión en 3D se asocia a maquinaria de bajo precio y capacidad en general. Así también, la ASTM define a la fabricación aditiva como “un proceso de unión de materiales para hacer objetos a partir del modelado en 3D, generalmente capa por capa, en oposición a las metodologías de fabricación sustractivas y

fabricación formativa” (Rodríguez-Salvador & Mancilla-de-la-Cruz, 2018), mientras que impresión en 3D la define como “fabricación de objetos mediante la deposición de un material utilizando un cabezal de impresión, boquilla u otra tecnología de impresora” (ISO/ASTM 52900-2015: Additive manufacturing -- General principles -- Terminology, 2015).

- **Realidad Aumentada**

En muchas aplicaciones industriales y domésticas se dispone de una gran cantidad de información asociada a objetos del mundo real, la realidad aumentada se presenta como el medio que combina dicha información con los objetos del mundo real. Así, muchos de los diseños que realizan los arquitectos, ingenieros, diseñadores pueden ser visualizados en el mismo lugar físico del mundo real para los que han sido diseñados. La Realidad Aumentada está relacionada con la tecnología Realidad Virtual que está más extendida en la sociedad; presentando algunas características comunes como por ejemplo la inclusión de modelos virtuales gráficos 2D y 3D en el campo de visión del usuario. La principal diferencia es que la realidad aumentada no reemplaza el mundo real por uno virtual, sino por el contrario, mantiene el mundo real que ve el usuario complementándolo con información virtual superpuesta a la real. El usuario nunca pierde el contacto con el mundo real que tiene a la vista y al mismo tiempo puede interactuar con la información virtual superpuesta. (Basogain et al., 2000).

- **Big Data y Analítica**

Big Data, se define como los grandes volúmenes de datos, estructurados, no estructurados o semi estructurados y la necesidad de su captura, almacenamiento, procesamiento, análisis y visualización.

Al inicio se hablaba de un modelo de las 3Vs: Volumen, variedad y velocidad, luego IBM incluyó dos Vs más: Veracidad y Valor y, actualmente se han integrado dos más: Visualización y Viabilidad para llegar a un modelo de Big Data de las 7Vs. (Muñoz, 2019).

Aunque la mayor parte de la bibliografía propone nueve pilares tecnológicos, algunos otros consideran uno más, que es la Inteligencia Artificial. (Basco et al., 2018).

- **Inteligencia Artificial**

Se basa en el desarrollo de algoritmos que permiten a las computadoras procesar datos a una velocidad inusual, logrando además aprendizaje automático. Los algoritmos se nutren de datos y experiencias recientes y se van perfeccionando, habilitando a la máquina con capacidades cognitivas propias de los seres humanos como visión, lenguaje, comprensión, planificación y decisión en base a los nuevos datos.

La combinación de estas tecnologías, permiten obtener información en tiempo real, procesarla y analizarla, dejando atrás el modelo de prevención y permitiendo nuevas formas para predecir sucesos. Estos modelos predictivos se alimentan de información que puede provenir de fuentes propias de la empresa o de fuentes externas. (Basco et al., 2018).

El nuevo paradigma de Industria 4.0 se basa en la innovación, automatización y procesos más sofisticados que suponen un salto exponencial en la capacidad competitiva de la industria. Mantener un estado competitivo depende de la capacidad de adaptación al momento evolutivo en el que nos encontramos y, por lo tanto, a la capacidad de incorporar las nuevas tecnologías (Guerrero Cano et al., 2019).

### 1.2 Procesos Industriales

Los procesos constituyen los métodos de trabajo empleados por las organizaciones para aportar valor a sus clientes (externos e internos) (Figura 3). (Pardo, 2012).

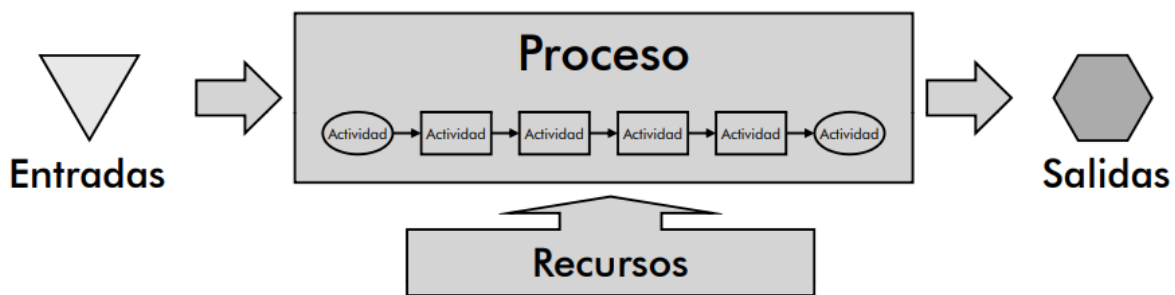


Figura 3. Representación Esquemática de un proceso (Pardo, 2012)

Dentro de los procesos industriales, existen actualmente varias herramientas que pueden ser utilizadas para posibles mejoras y diagnósticos, pero una de las principales es el uso de técnicas estadísticas que vienen a lo largo de los años, mejorando todo sistema operacional además de permitir tornar los productos fabricados más competitivos. (Hernández-Pedreira & Da Silva-Portofilipe, 2016).

La idea de usar técnicas de muestreo y análisis estadístico en un entorno de producción tuvo sus comienzos en la década de 1920. El objetivo de este concepto tan exitoso es reducir de manera sistemática la variabilidad y el aislamiento asociados con las fuentes de dificultades durante la producción. (Devore, 2008).

### 1.3 Predicción de Procesos

Un sistema predictivo, se trata de un método basado en la objetividad, con capacidad para efectuar análisis multivariable, de identificar factores críticos y de integrar en el conocimiento un aprendizaje continuo. La conjunción de estos elementos en una herramienta informática y el uso de algoritmos matemáticos avanzados le permiten actuar como herramienta de predicción (Zabala-Uriarte et al., 2009).

El control predictivo se puede considerar una técnica madura para sistemas lineales y no muy rápidos como los encontrados normalmente en la industria de procesos. Sistemas más complejos, tales como sistemas no lineales, híbridos, y o sistemas muy rápidos, eran considerados como fuera del alcance de los controladores predictivos. Durante los últimos años se han producido resultados espectaculares en estos campos. Recientemente han aparecido en la literatura aplicaciones de los controladores predictivos a sistemas no lineales o híbridos (Fernández Camacho & Bordóns Alba, 2004).

Como en otras muchas disciplinas científicas, en el aprendizaje automático el objetivo principal es modelar la relación entre un conjunto de variables medidas (entradas) y otro conjunto de variables relacionadas con estas (salidas). Una vez que se determinado dicho modelo matemático, es posible predecir el valor de las variables deseadas midiendo las observables. (Cano et al., 2019).

## 1.4 Trabajos Relacionados

Se han revisado algunos trabajos de investigación relacionados con el tema central, encontrando varios que aplican técnicas para mejorar algún proceso en específico, pero no necesariamente han aplicado alguna técnica computacional, predicción de procesos, o están enfocados a la industria 4.0.

Por ejemplo, se aplicó un análisis de regresión lineal múltiple para estimar la resistencia térmica de las incrustaciones en los enfriadores de ácido sulfhídrico (Sánchez-Escalona et al., 2018), a partir de los valores de las variables independientes. Se identificaron las variables que caracterizan al proceso y mediante la técnica experimental se registraron los valores de las variables.

Para el mismo análisis, se propuso otro modelo, que consiste en una red neuronal artificial.

Los coeficientes obtenidos en el análisis de regresión lineal permiten interpretar con facilidad las variables predictoras más influyentes, a diferencia de las redes neuronales artificiales que tienen un enfoque de “caja negra”.

En lo que respecta a la industria alimentaria (Guerrero Cano et al., 2019), se han aplicado algunas técnicas de predicción mediante la minería de datos, creando la oportunidad de reducir la dependencia de la supervisión humana al obtener información de fuentes de datos extensas, diversas y de transmisión directa.

Herramientas como la minería de datos y técnicas predictivas las organizaciones disponen de información que les ayuda a plantear estrategias empresariales más precisas, efectivas y aplicables en periodos de tiempos más cortos.

En un estudio para la reducción de problemas de adherencia en procesos de galvanizado, también se ha reportado el uso de minería de datos (Martínez-de-Pisón et al., 2007), en el que a diferencia de otras técnicas de minería de datos, los algoritmos generados de árboles de decisión pueden trabajar con un elevado número de atributos y muy pocos ejemplos. Esta característica los hace muy útiles en la búsqueda de conocimiento oculto dentro de históricos de procesos industriales donde se dispone de una enorme cantidad de variables y pocos casos erróneos.



## Capítulo 2. Fundamentos

De acuerdo con (Basco et al., 2018) la economía global transita una nueva fase que se caracteriza por la digitalización y la conectividad. Y, además, el análisis de datos y la toma de decisiones en tiempo real impactan positivamente en la eficiencia de toda cadena de valor.

Existen conceptos muy importantes a considerar para esta nueva era de la automatización, que no deben ser olvidados, para preservar la calidad de los productos, si bien, el proceso puede ser automatizado, optimizado y adaptado a las necesidades de la industria, el objetivo final siempre será asegurar la calidad del producto, evitando pérdidas de recursos, ya sea material, tiempo de producción, mano de obra o reprocesamiento.

### 2.1 Calidad

La ASQ, *American Society for Quality*, define calidad como “Un término subjetivo para el cual cada persona puede tener su propia definición. En un enfoque técnico, la calidad puede tener dos significados: 1. Las características de un producto o servicio que influyen en su capacidad para satisfacer necesidades declaradas o implícitas. 2. Un producto o servicio libre de deficiencias”. Cuando se trata de definir Calidad, a detalle, se puede encontrar que es un concepto intangible y complicado (Yang & Basem El-Haik, 2016).

Algunos autores, hacen referencia a las dimensiones de la calidad, listando en algunos casos, cinco o más dimensiones, por mencionar alguno, (Garvin, 1984) hace referencia a ocho dimensiones:

- Rendimiento

Se refiere a las características de operación principales de un producto, éstas corresponden a las características objetivo, la relación rendimiento y calidad, refleja las reacciones individuales de cada producto.

- Características

Son las particularidades secundarias que complementan las funciones básicas de un producto. En algunos casos las características principales o primarias como rendimiento son difíciles de distinguir de estas características secundarias, esta distinción normalmente se centra en la importancia para el usuario.

- Confiabilidad

Refleja la probabilidad de que falle un producto en un periodo de tiempo específico. Entre las mediciones de confiabilidad más comunes está el tiempo promedio para la primera falla (MTFF), tiempo promedio entre fallas (MTBF) y el porcentaje de fallas por unidad de tiempo.

- Conformidad

Se refiere al grado en que el diseño de un producto y las características de operación cumplen con sus estándares, incluye elementos tanto internos como externos. La conformidad es comúnmente medida por la incidencia de defectos: la proporción de las unidades que fallan, o no cumplen con las especificaciones y requieren ser retrabajadas o reparadas.

Confiabilidad y conformidad están altamente relacionadas. Mejoras en ambas mediciones se traducen en ganancia en calidad, y son mediciones objetivas, que reflejan en menor medida las preferencias individuales que las clasificaciones basadas en el rendimiento o características.

- Durabilidad

Es una medición de la vida del producto, tiene dimensiones técnicas y económicas. Técnicamente la durabilidad puede ser definida como la cantidad de uso que se puede obtener de un producto antes de que sea físicamente deteriorado. En ocasiones, la reparación es imposible, sin embargo, la durabilidad se vuelve más difícil de interpretar cuando a la reparación es posible, ya que la vida útil del producto puede variar con los

TESIS TESIS TESIS TESIS TESIS

cambios en las condiciones económicas. En este caso el ciclo de vida de un producto es determinado por los costos de reparación, evaluación personal de tiempo, inconvenientes, pérdidas debido al tiempo muerto, precios relativos y otras variables económicas, así como la calidad de los componentes o materiales.

Durabilidad y confiabilidad están fuertemente relacionados, es decir, un producto que presenta fallas frecuentemente es más probable que sea desechado más rápido que un producto más confiable.

- Servicio

Los consumidores no solo están preocupados acerca de las fallas o descomposturas de un producto, sino también, acerca del tiempo transcurrido antes de que los servicios sean restablecidos, las citas, llamadas, etc. Velocidad, cortesía y capacidad de reparación pueden ser términos para reconocer esta dimensión.

Algunas de estas variables pueden ser medidas objetivamente, otras reflejan diferentes estándares personales de lo que constituye un servicio aceptable. Un ejemplo objetivo de esta medición es MTTR (*Mean Time to Repair*).

- Estética

Es una medición subjetiva, que está estrechamente relacionada con la siguiente dimensión, percepción de la calidad, basada en la percepción del usuario.

Estética, se refiere a como “se ve” un producto, como se escucha o se siente, siendo claramente un juicio personal que refleja las preferencias individuales.

- Percepción de la calidad

Es también una medición subjetiva, ya que los consumidores no siempre poseen información completa acerca de los atributos de un producto, y confían frecuentemente en mediciones indirectas cuando comparan con otros productos. Por lo que no son evaluados objetivamente.

De acuerdo con el autor, cada dimensión es autónoma y distinta, y un producto puede ser clasificado como alto en alguna dimensión y bajo en otra.

## 2.2 Aprendizaje automático

El aprendizaje automático o *machine learning* es un área de la inteligencia artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través del entrenamiento con grandes volúmenes de datos. La construcción de modelos de aprendizaje automático requiere adaptaciones propias debido a la naturaleza de los datos o a la problemática que se aplica. (Russo et al., 2015).

“El trabajo de aprendizaje automático, converge de varias fuentes” (Nilsson, 2005), algunas de estas fuentes son: estadística, pues los métodos estadísticos pueden ser considerados instancias de machine learning, debido a que varios algoritmos están basados en reglas de estimación estadística. Modelos cerebrales, en los que importantes técnicas de *machine learning* están basadas como las redes de elementos no lineales, conocidas como redes neuronales, aproximando el aprendizaje al funcionamiento del cerebro humano. Los modelos psicológicos por su parte han sido tomados en algunos trabajos de aprendizaje por refuerzo y árboles de decisión, reconociendo el comportamiento en humanos y animales. Desde sus inicios, la inteligencia artificial también converge en el aprendizaje automático, con aportes significativos en varios algoritmos que son empleados en varios ámbitos, como juegos, toma de decisiones, sistemas expertos, lógica de programación inductiva, solución de problemas, entre otros; algunos autores, plantean *machine learning* como una disciplina de la inteligencia artificial. La teoría de control adaptativo es también otra de las fuentes presentadas por el autor, en la cual, se estudian problemas como control de procesos con parámetros desconocidos que deben ser estimados durante la ejecución, la teoría de control debe rastrear estos cambios. Los modelos evolutivos contribuyen a las técnicas como algoritmos genéticos o programación genética, en los que aspectos de la evolución biológica forman parte de las bases para este tipo de aprendizaje.

Existen tres grandes tipos de aprendizaje de algoritmos, supervisados, no supervisados y por refuerzo.

Los modelos de aprendizaje supervisado, se entrenan con datos en los que los resultados de salida son conocidos (Nuin et al., 2020), mientras que en los modelos de aprendizaje no supervisado, los algoritmos parten de datos no etiquetados para los que no existe información sobre clasificación o eventos. El trabajo principal del algoritmo es encontrar patrones de agrupación o *clusters* para elementos similares dentro del conjunto de datos. El aprendizaje supervisado es la técnica más común para entrenamiento de redes neuronales y árboles de decisión, las cuales son altamente dependientes de la información dada por clasificaciones predefinidas (Zhang, 2010).

De acuerdo con (Nuin et al., 2020), en el aprendizaje por refuerzo, el proceso de aprendizaje funciona mediante intervención humana, “premiando” o “penalizando” las decisiones parciales durante el aprendizaje, y las condiciones del algoritmo no son predecibles.

Por su parte, (Zhang, 2010) agrega otros tipos de aprendizaje como lo son aprendizaje semi supervisado, en el que se combinan ejemplos etiquetados y no etiquetados para generar una función o clasificación adecuada; algoritmos de transducción, similares al aprendizaje supervisado, con la diferencia que no se construye explícitamente una función, sino que trata de predecir nuevos resultados basados en entradas y salidas de entrenamiento y nuevas entradas; y por último, aprendiendo a aprender, en el que los algoritmos generan sus propios “prejuicios” o tendencias basados en experiencia previa.

Existe un sinnúmero de usos y aplicaciones de las diferentes técnicas de *machine learning*, algunos ejemplos de acuerdo con (Sandoval, 2018), son aplicaciones en cuanto a predicción de ventas de años siguientes analizando comportamientos actuales de clientes, prediagnósticos médicos que se basan en síntomas del paciente, preferencias de clientes de acuerdo con las operaciones realizadas en la red, cambios en comportamientos de aplicaciones móviles para adecuarse al usuario, predicción de tráfico urbano y búsqueda de alternativas, tendencias y motores de búsqueda, entre muchas otras.

Otra área que puede ser explotada usando *machine learning* es *big data* (Dhanda et al., 2019), término que se usa para describir datos con grandes volúmenes, si bien no todos los problemas de aprendizaje automático son adecuados para *big data* y no todos los conjuntos de datos son útiles para entrenar modelos, la conjunción de situaciones particulares puede conducir resultados extraordinarios, eliminando limitantes que normalmente afectan escenarios pequeños.

De acuerdo con (González-Marcos Fernando Alba-Elías, 2017), entre las técnicas de aprendizaje de máquina más utilizadas en el modelado de procesos industriales se encuentran las redes neuronales, las máquinas de vectores soporte, los árboles de decisión o los *random forest*. También señala, que ningún modelo o algoritmo puede o debe ser usado de modo exclusivo y que no existe el «mejor» modelo o algoritmo para un problema dado, ya que la propia naturaleza de los datos afectará a la elección de los modelos.

Mientras algunas técnicas propuestas hace más de veinte años continúan siendo refinadas, el incremento de tamaño, complejidad y variedad de los modelos actuales, demandan una reevaluación de las suposiciones y técnicas existentes (Sra et al., 2012).

### 2.3 Minería de datos

“Minería de datos es un proceso de descubrimiento de varios modelos, compendios y valores derivados de una colección de datos dada”. (Kantardzic, 2011).

La minería de datos descubre patrones de comportamiento mediante técnicas y algoritmos en los datos que normalmente son tratados como *big data* por su volumen, velocidad y validez, convirtiendo los datos en comprensibles y utilizables a menudo en el aprendizaje automático en la reproducción de patrones conocidos, como antecedente para algunos algoritmos, incluyendo los de predicción.

De acuerdo con (Kantardzic, 2011), avances recientes en computación, comunicaciones y tecnologías de almacenamiento digital y el desarrollo de tecnologías de adquisición de datos han hecho posible recopilar y almacenar datos en grandes volúmenes que crecen de forma exponencial, siendo demasiado grandes para tratarse de forma manual o de forma semiautomática, trabajar con grandes volúmenes de datos implica limitaciones prácticas para los algoritmos de minería de datos y aprendizaje de máquina, ya que aumentan el riesgo de soluciones de baja probabilidad que no cumpla con expectativas futuras además de la complejidad computacional para el tratamiento de dichos datos.

Existen varias formas y técnicas que pueden ser aplicadas a la reducción de grandes conjuntos de datos.

La fase inicial del proceso de minería de datos se enfoca en comprender los objetivos del proyecto, requerimientos del negocio, convirtiéndolos en una definición de minería de datos y plan preliminar enfocado en lograr los objetivos.

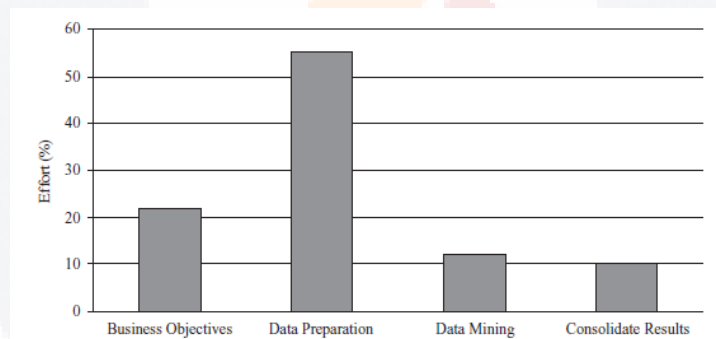


Figura 4. Esfuerzo en proceso de minería de datos(Kantardzic, 2011).

La Figura 4. Esfuerzo en proceso de minería de datos(Kantardzic, 2011)., hace referencia al compendio de varios autores que estiman que cerca del 20% del esfuerzo se enfoca en definir los objetivos del negocio, 60% en la preparación y conocimiento de los datos, lo cual, implica un importante consumo de tiempo en las tareas relacionadas con el análisis de datos, incluyendo fuentes, dueños, responsables de mantenimiento en la organización, costo, almacenamiento, tamaño, atributos, seguridad, restricción de uso, requerimientos de privacidad, problemas de calidad, tipos de datos, rango de valores, recolección de información, identificación de datos faltantes, datos inválidos, entre otras tareas, y otro 10% se enfoca en minería de datos y análisis de resultados de forma similar.

## 2.4 Control de procesos

“Las variaciones que se presentan en cualquier proceso pueden deberse a causas comunes o a causas especiales. La variación natural de materiales, maquinaria y personas da lugar a las causas comunes de variación. Las causas especiales, también conocidas como causas asignables, se deben en la industria a desgaste excesivo de las herramientas, a un nuevo operador, a cambios en los materiales, a nuevos proveedores, etc. Uno de los propósitos de las gráficas de control es localizar, y si es posible, eliminar las causas especiales de variación.” (Pérez et al., 2017)

De acuerdo con (Pérez et al., 2017), para llevar a cabo un análisis de la capacidad de un proceso, el proceso debe estar bajo control estadístico.

Se supone que las características del proceso que van a ser medidas tienen distribución normal. Esto se puede comprobar empleando pruebas de normalidad, como la prueba de Kolmogorov-Smirnov, la prueba de Ryan-Joiner o la prueba de Anderson-Darling. La capacidad del proceso es una comparación entre el desempeño del proceso y los requerimientos del mismo. Los requerimientos del proceso determinan los límites de especificación. El LSL y el USL (por sus siglas en inglés) son, respectivamente, el límite inferior de especificación y el límite superior de especificación. Los datos utilizados para determinar si un proceso está bajo control estadístico pueden emplearse para hacer el análisis de capacidad. A la distancia de 3 sigmas a ambos lados de la media se le conoce como dispersión del proceso. La media y la desviación estándar de las características del proceso suelen estimarse a partir de los datos obtenidos para el estudio del control estadístico del proceso.



## Capítulo 3. Identificación y selección de un proceso Industrial

Para comenzar con el análisis de un proceso industrial, primero, es necesario identificar el proceso a utilizar, con la diversificación de productos en las diferentes empresas, se pueden encontrar procesos de cualquier índole, que generan productos, que a su vez son parte de otros procesos.

Como parte de los trabajos iniciales, se ha reconocido físicamente diversos procesos, que pueden ser tan simples o complejos dependiendo que tan a fondo se estudien.

A simple vista, parece sencillo, si de acuerdo con la Figura 3. Representación Esquemática de un proceso (Pardo, 2012), vemos cualquier proceso como una caja negra, en la que intervienen diferentes recursos y suministrando entradas obtenemos ciertos valores de salida y producto del proceso. Sin embargo, al llevarlo a la práctica, después de un recorrido real en una empresa grande, es posible percibir que un simple cambio en los factores que intervienen en el proceso puede tener un gran impacto en sus salidas.

A continuación, se describirá brevemente las características sobresalientes de algunos de los procesos estudiados.

### 3.1 Moldeo

Como parte de la integración vertical en la empresa, el proceso de moldeo es uno de los procesos base, que genera diferentes piezas plásticas que, a su vez, sirven de suministro para otros procesos, integrándolos con otros materiales para formar ensambles finales de diferentes productos.

Existen varios factores, que al igual que en otros procesos, son de importancia para el proceso mismo.

Una prensa para moldeo es la máquina principal encargada de ejecutar el proceso de forma automática, una vez que los recursos son suministrados ya sea de forma manual o también automática, dependiendo de la pieza y de la máquina misma, al suministrar material plástico, que llamaremos resina, se funde para ser prensado posteriormente y al enfriarse, genera una pieza de plástico dependiendo del molde utilizado.

A pesar de que el proceso se realiza de forma automática y cada vez es más preciso, se pueden encontrar defectos de calidad, que, de no ser identificados, generan un defecto mayor durante los procesos subsecuentes en los que se utiliza la pieza moldeada.

Llevando el esquema inicial mencionado anteriormente al proceso de moldeo, se obtiene la Figura 5. En la que se pueden clasificar los diferentes factores.

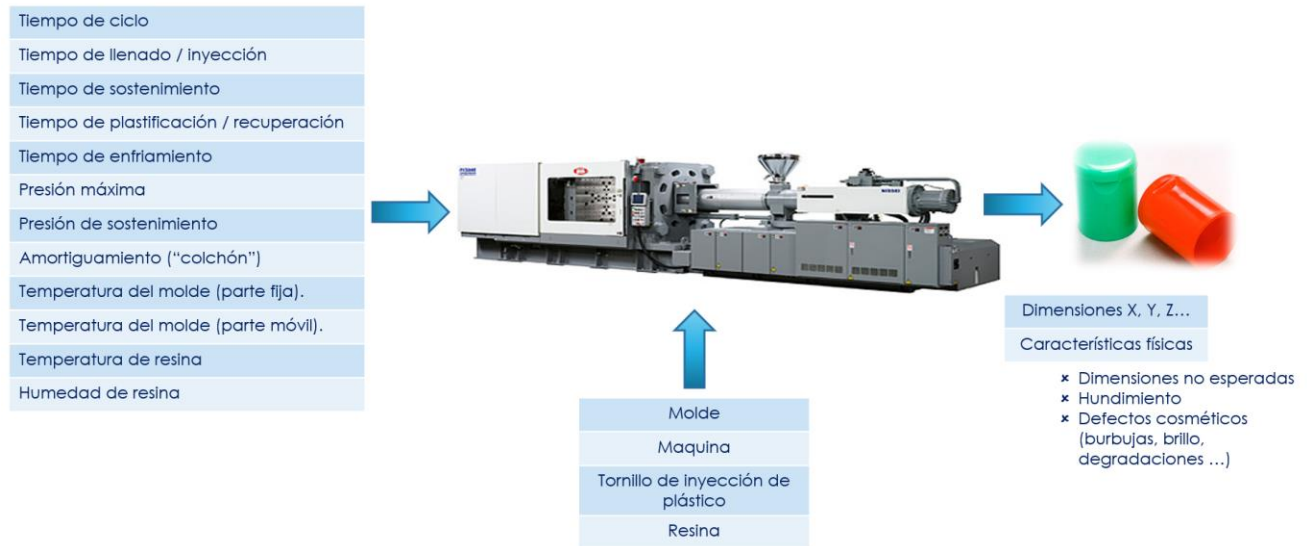


Figura 5. Esquema general del proceso de moldeo. Elaboración Propia.

**Recursos**

**A. Molde**

Es una pieza intercambiable en la prensa, que, da forma a la resina durante el proceso, dependiendo del producto esperado, es el molde que se debe emplear.

La empresa, cuenta con más de 200 piezas metálicas para generar productos, los cuales pueden generar de una hasta n número de piezas con el mismo molde, dependiendo de tamaño.

Presentar un daño en el molde puede ser catastrófico para la calidad de la pieza, ya que es el encargado de la modelación de la pieza, por lo que, se deben tener controles específicos como el mantenimiento y revisión constante de los mismos.

## B. Máquina

La máquina o prensa de moldeo, se encarga “convertir” la resina suministrada en la pieza moldeada, conjuntando los diferentes factores, de acuerdo a la configuración de esta.

Según el modelo de máquina, puede presentar diferentes funciones, que van desde las más austeras, hasta funciones de conexión por medio de red para generación de datos.

## C. Tornillo de inyección de plástico

Está conformado por barril y husillo, y es uno de los recursos esenciales para el proceso ya que es el encargado de inyectar la resina al molde y su desgaste, puede provocar severos defectos en el proceso, que, aunque los parámetros de entrada sean los adecuados, un problema en este elemento puede resultar catastrófico. Actualmente, el control del mantenimiento de esta pieza se realiza de forma manual, mediante mediciones dimensionales del barril y husillo, así como la separación entre ambos, al detectar alguna dimensión fuera de parámetro, se realiza el cambio de cualquiera de los dos elementos.

## D. Resina

Es el material plástico que normalmente se encuentra en forma de esferas que será fundido para formar la pieza plástica al final del proceso. Éste puede ser de diferentes colores y puede fundirse a diferentes temperaturas.

Durante la ejecución del proceso, se ha identificado, que, de acuerdo con la calidad del material y tipo, los demás factores que influyen en el proceso deben ser distintos, como lo es el caso de la temperatura a la que debe fundir el material o la presión de sostenimiento. La variación de estos parámetros, aun cuando son los mismos proveedores, se encuentra hasta por cada lote de este material, por lo que se considera de gran relevancia para el proceso este recurso.

## Entradas

### A. Tiempo de ciclo

Se conoce como tiempo de ciclo, el tiempo que tarda el proceso en generar un “disparo” del producto, dependiendo del tipo de producto y del molde, puede generar desde una hasta n piezas al mismo tiempo. El tiempo de ciclo se mide de disparo a disparo a partir de la carga de resina.

### B. Tiempo de llenado / inyección

Es el tiempo que transcurre durante la inyección de resina hasta antes de entrar a sostenimiento, puede ser detectado mediante sensores a través de las señales establecidas en cada punto de proceso.

### C. Tiempo de sostenimiento

Es el tiempo que transcurre a partir de que el molde queda “cerrado” para que la resina tome la forma del molde, este parámetro es de gran importancia para el proceso, ya que la resina debe solidificar, con un tiempo menor al adecuado, puede no generar la forma del producto, provocando problemas en la maquina con la resina líquida, o bien, con un tiempo mayor, puede provocar defectos cosméticos en la pieza y retraso en el tiempo de producción.

### D. Presión de sostenimiento

Se refiere a la presión en promedio que se ejerce en el molde después de ser inyectado para dar forma a la resina.

### E. Tiempo de plastificación / recuperación

Se conoce también como tiempo de dosificación o carga y es el tiempo que tarda el husillo en regresar a la posición de carga de resina.

F. Tiempo de enfriamiento

Este parámetro es definido de acuerdo con la pieza a ser moldeada y las características de la resina, y es el tiempo que debe estar la resina dentro del molde para solidificarse y dar forma a la pieza plástica.

G. Presión Máxima

Es el valor máximo de presión que se ejerce al iniciar la inyección después del llenado de resina.

H. Amortiguamiento / “Colchón”

Es el espacio que se debe asegurar entre el husillo y el barril en el cual debe existir resina durante el proceso, con la finalidad de evitar faltante de resina en la pieza al ser solidificada.

I. Temperatura del molde (parte fija).

La temperatura que debe presentar el molde en la parte fija, debe ser establecida de acuerdo con las necesidades del proceso, por medio de agua, se regula la temperatura entre cada uno de los disparos del proceso.

J. Temperatura del molde (parte móvil).

Es la temperatura a la que debe estar la parte móvil del molde para para lograr la generación de la pieza, esta es diferente dependiendo del producto y de la resina a utilizar, normalmente, no presenta variación en el proceso, al igual que la temperatura en la parte fija, y de los modelos estudiados, oscila entre 70 y 25 grados Fahrenheit.

K. Temperatura de Resina

Es la temperatura ideal a la que se debe encontrar la resina, antes de ser prensada por la máquina, esta debe ser fija dependiendo de la resina a utilizar y de la pieza a ser generada, sin embargo, puede presentar variación durante el proceso mismo.

La temperatura es establecida en la máquina de moldeo y en un caso ideal, debe mantenerse a lo largo de la producción de todo el lote del material.

L. Humedad de resina

Para que el proceso funcione con los parámetros configurados, la resina no debe presentar humedad, es decir debe estar seca. Esto se puede detectar al realizar una prueba de laboratorio a la resina antes de comenzar todo el proceso.

**Salidas**

La salida esperada del proceso de moldeo es por supuesto una pieza plástica, generada a partir del molde seleccionado, puede presentar distintos colores y texturas dependiendo de la resina a utilizar y del producto a generar, como se ha mencionado, es a su vez una entrada para los siguientes procesos dentro de la misma compañía por lo que, se espera no encontrar defecto alguno. Para asegurarnos, actualmente, se realizan algunas pruebas como resultante del proceso mismo, por lo que consideraremos como salidas las dimensiones y características física, también conocidas como atributos.

A. Dimensiones X, Y ...

Dependiendo del tipo de producto, dimensiones críticas son definidas, las cuales, deben estar dentro de un rango esperado y pueden ser medidas mediante diferentes herramientas como lo son *caliper*, *pin gage*, o maquinaria especializada y precisa como CMM, OGP y sistemas de visión.

Existen piezas muy sencillas, las cuales cuentan con una dimensión crítica como mínimo, sin embargo, otras piezas pueden tener hasta 50 dimensiones, las cuales deben ser aseguradas antes de iniciar un nuevo lote de producción.

El muestreo depende del producto a generar y de los procedimientos de calidad, que pueden ser tan sencillos o complejos dependiendo de la planta de la compañía y el sector en el que se encuentre, así como la criticidad para el siguiente proceso, en algunos casos,

se mide al iniciar el lote, dos veces por turno, otros solo al iniciar el turno o solo al iniciar el lote.

Normalmente uno o dos disparos son suficientes para identificar algún defecto inicial, sin embargo, el proceso se complica al realizar las mediciones por cada una de las cavidades del molde (número de piezas generadas), que pueden ser desde una hasta  $n$ , de los modelos estudiados, el máximo número de cavidades son 48, además de que, al ser una empresa que demanda productos moldeados para varios procesos, no se cuenta con solo una máquina, sino al menos 50 que generan diferentes productos, con diferentes lotes y demandas para cada una.

#### B. Características físicas

Además de la salida dimensional del producto, es necesario evaluar las características físicas o atributos con los que cuenta la pieza, lo cual, normalmente se identifica con una inspección visual, en esta parte específica del proceso, es complejo acertar en el resultado de la inspección ya que la mayoría de las veces recae el criterio del propio inspector, y frecuentemente, se rechazan lotes completos en el siguiente proceso debido a que la producción se generó de la misma forma.

#### **Defectos generados**

Algunos de los defectos más comunes encontrados en los productos derivados del proceso de moldeo, se pueden agrupar en dos grandes categorías, que son dimensionado fuera de especificación o defectos cosméticos que no permiten el uso de la unidad en los siguientes procesos, estos, se pueden identificar con ayuda de las siguientes imágenes.

#### **Dimensiones fuera de especificación**

Cuando las dimensiones de la pieza plástica están fuera de los límites de especificación considerando las tolerancias no pueden ser utilizados posteriormente, ya que su uso provoca fallas funcionales después de ensamblarlo en los siguientes procesos.

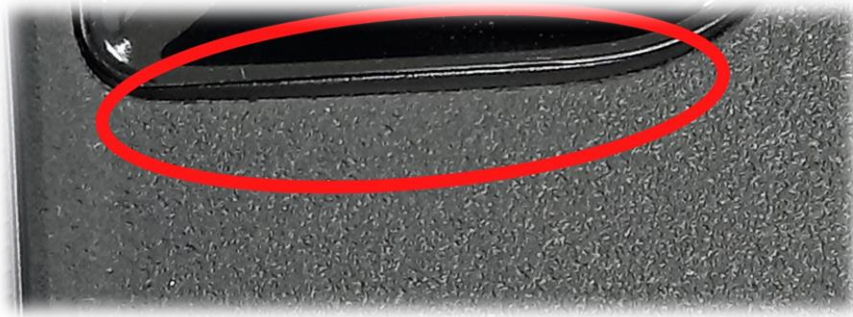


Figura 6. Dimensionado fuera de especificación.

Para evitar este tipo de defectos, la operación actualmente realiza una medición a algunas de las primeras 10 piezas producidas, el tamaño de la muestra varía en función del producto, con la finalidad de detectar problemas en el dimensionado, sin embargo, estos problemas pueden ocurrir a lo largo de la fabricación de todo el lote.

También resulta ser frecuente que existan errores al tomar la medición, ya que se utilizan diferentes herramientas, que no son digitales, y depende mucho de la habilidad del operador.

La siguiente figura, es un ejemplo del Registro de pieza fuera de especificación. Este es un formato convencional que se llena de forma manual basándose en el diagrama de la unidad.

No.	SYMBOL	DRAWING ZONE	NOMINAL	TOLERANCE (SPECIFICATIONS)				ACTUAL RESULTS				
				+ TOL.	- TOL.	UPPER	LOWER	SAMPLE 1	SAMPLE 2	SAMPLE 3	SAMPLE 4	SAMPLE 5
1		0-22	0.000	0.250	0.750	0.250	-0.750	0.59	0.52	0.57	0.51	0.50

Figura 7. Registro de pieza fuera de especificación.

### Defectos cosméticos

Los defectos cosméticos, son más frecuentes durante el proceso de moldeo, y solo son percibidos mediante inspección visual de las diferentes piezas, esta es una tarea rápida en la que los operadores deben descartar todas las piezas con defectos, que no puedan ser reparadas, desechándolas sin opción a recuperación o reparación. Son pocos los defectos que pueden ser reparados, pero si suele haber.



**Fisura**

Un defecto común es la fisura, la cual es generada por un corte incorrecto de la pieza cuando se inyectan varias al mismo tiempo, ya sea manual o automático, evitando que la pieza pueda ser utilizada en los siguientes procesos.



*Figura 8. Fisura.*

**Grasa**

Si bien, este defecto no afecta en algún ensamble, por parte de calidad no es permisible, ya que las piezas no están total mente limpias, generando una apariencia no deseable.

Esto se puede prevenir con la limpieza adecuada a los moldes y máquina, ya que un exceso de grasa en los mismos mancha la pieza como se muestra en la Figura 9.



*Figura 9. Grasa en la pieza plástica.*

**Rebaba**

Son defectos ocasionados por el tiempo de llenado, o el desgaste del molde, son excesos pequeños de plástico en una o más partes de la pieza, en ocasiones pueden ser limados para corregirlos, sin embargo, durante este proceso la mayor parte de las piezas son dañadas, por lo que, dependiendo la expertiz de los operadores, se intentan o no reparar.



Figura 10. Rebaba.

### Hilillos

Es exceso de plástico en forma de hilo debido a problemas en el punto de inyección, generando un exceso en la misma parte, el problema se torna mayor si la pieza plástica es pintada aún con este defecto, ya que genera un hilo más grueso, evitando la recuperación de la pieza.

En ocasiones, si los hilillos son identificados antes de la pintura, la pieza puede ser reparada, mediante procesos manuales como limado.



Figura 11. Hilillo.



Figura 12. Hilillo, reparable.

### Faltante de plástico

Por el contrario de los hilillos y rebabas, la falta del tiempo de llenado o presión no adecuada, pueden ocasionar faltantes de plástico que no forman completamente la pieza, este tipo de defecto, no permite una reparación, ya que la pieza moldeada debe ser generada en una sola parte.

La resina y tiempo de inyección son recursos que ya no son recuperables al encontrar este tipo de defecto.



Figura 13. Faltante de plástico.

En el caso de la Figura 14, al pasar por alto este primer defecto, el ensamble de la pieza no puede darse de manera uniforme, provocando también el desperdicio de esta y la pieza plástica ensamblada.



Figura 14. Pieza incompleta.

**Malformaciones**

Es común en el proceso, también encontrar malformaciones en las piezas, es decir, formas no esperadas en alguna parte de la pieza, provocadas por temperaturas no adecuadas o algún tiempo de proceso mayor o menor al necesario para la producción. Este no es un defecto reparable.



Figura 15. Malformación en la pieza.

### Líneas de flujo

Se pueden encontrar las conocidas como líneas de flujo o rayones en cualquier tipo de pieza de plástico, las cuales, pueden ser generadas por velocidades inadecuadas en el proceso, marcando visiblemente la pieza. Tampoco este defecto puede ser reparado.



Figura 16. Líneas de flujo.

### Puntos de contaminación

Existen partículas que se quedan atrapadas en las piezas plásticas o en el resultado de algún ensamble, que comúnmente no afectan el funcionamiento de la pieza, sin embargo, cosméticamente no es aceptable.



Figura 17. Puntos de contaminación en película plástica.

### Hundimiento

Es un defecto generado en las piezas, que no es generalmente identificado en el dimensionado, y genera una especie de cuna en la pieza, que no permite el uso en procesos subsiguientes.

Una vez generadas las piezas plásticas, son transportadas a los siguientes procesos dentro de la empresa de manufactura o almacenadas dependiendo del uso y la demanda de los ensambles finales. Enviar piezas de buena calidad, facilitan los procesos siguientes y disminuyen la generación de defectos de las unidades en las que se utilizan.

### 3.2 Tecnología de Montaje Superficial

La Tecnología de Montaje Superficial (SMT) es el método de fabricación de dispositivos electrónicos más utilizado en la actualidad, fue desarrollado en la década de 1960, pero no fue hasta finales de 1980 que tuvo amplia aplicación en la industria, ya que su uso se limitaba a circuitos híbridos de bajo volumen de producción, debido a la limitante de equipo automatizado que pudiera realizar este tipo de montaje. A fines de la década de 1990, se volvió el método dominante para la producción de tarjetas electrónicas, permaneciendo hasta la actualidad con la existencia gran variedad de equipos que permiten su implementación en mayor número de industrias para grandes volúmenes de producción, IBM ha sido uno de los pioneros de este tipo de fabricación.

Algunas de las ventajas de este tipo de fabricación, se centran en el tamaño reducido de los componentes y la ausencia de "hilos", mejora de inductancia y aumento de resistencia, el método de fabricación, además, permite el uso de ácidos, disolventes y limpiadores que facilitan la limpieza de residuos resultantes del proceso de soldado.

Dentro del proceso de SMT, podemos encontrar diversos procesos asociados, que dependen del tipo de producto a fabricar, entre los cuales destacan: Aplicación de soldadura en pasta, colocación de componentes, soldadura de componentes, limpieza e inspección y reparación.

Como parte de los trabajos de investigación, se ha recorrido físicamente el proceso de Tecnología de Montaje superficial para diferentes productos de la compañía, en diferentes líneas de producción. Se describirá brevemente cada uno de los subprocesos, en los cuales, se pudo identificar claramente la teoría de SMT.



Figura 18. Proceso general de Tecnología de Montaje Superficial. Elaboración Propia.

### I. Impresión de Soldadura en pasta

Es la primera etapa del proceso general, la cual puede considerarse como un proceso completo, que requiere conexión con otros procesos para la producción del producto final.

Se realiza de forma automática, a partir de que un operador suministra una tarjeta electrónica (PCB) a la máquina encargada de realizar la impresión de la pasta, tomando como guía una plantilla o esténcil, que presenta diferentes aberturas y formas dependiendo del circuito a imprimir.

Las tarjetas electrónicas pueden ser encontradas de forma individual o agrupadas en paneles, dependiendo de su tamaño, permitiendo la impresión de más de una durante el mismo ciclo.

Una vez colocado el esténcil en la posición correcta al igual que las tarjetas, la impresora, suministra pasta y aplica una presión para cubrir las aberturas del esténcil con la pasta necesaria.

### II. Inspección de Impresión

La inspección de la impresión de pasta es realizada para asegurar el suministro de pasta necesario para el producto, una vez impresa, la tarjeta pasa a través de una banda transportadora (*conveyor*) a la máquina encargada de realizar la impresión, en la cual,

mediante un equipo automático, se identifica la cantidad suministrada en cada uno de los pines de pasta, con la finalidad de evitar defectos en los siguientes procesos.

### III. Colocación de Componentes

Si la impresión de la tarjeta o panel de tarjetas ha sido clasificada como buena, se continua en la colocación de componentes de forma automática.

Componentes como transistores, diodos, resistencias, capacitores son “cargados” en la maquina encargada de colocarlos, en forma de rollo y durante este proceso, son colocados en la PCB de acuerdo con la posición indicada para cada uno de productos a generar.

### IV. Proceso de Reflujo

Una vez colocados los componentes, son soldados a la placa de circuito impreso, aplicando altas temperaturas con el objetivo de fundir la soldadura y adherir los diferentes componentes a la placa.

### V. Inspección Óptica

Cuando los componentes fueron soldados a la tarjeta, se procede a realizar una inspección óptica de la PCBA, es decir una revisión visual realizada de forma automática, con la finalidad de detectar defectos, entre los cuales se puede destacar, componente equivocado, desfase de componente, componente mal colocado o falta de componente que afectan directamente el funcionamiento esperado de la PCBA.

### VI. Inspección de Rayos X

Algunos productos, deben ser inspeccionados por una máquina de rayos X, después de la inspección óptica convencional, con la finalidad de garantizar un alto nivel de confiabilidad del producto ya sea considerado terminado a nivel tarjeta o como subensamble de uno mayor. Mediante la emisión de radiación electromagnética de alta frecuencia, los elementos densos en la tarjeta absorben más fotones y se muestran como

imagen en color negro, mientras que los menos densos son atravesados por los fotones, permitiendo una inspección multicapa que permite identificar defectos a más profundo nivel.

## VII. Reparación

Como parte del proceso de SMT, la reparación puede ser efectuada en ciertos casos del proceso, ya que no todas las piezas son reparables, esto depende del daño y del defecto presentado.

La primera intervención, se realiza si se encuentra algún defecto durante la inspección de impresión, en que, se debe realizar una limpieza de pasta en toda la tarjeta de forma manual, sin embargo, no todas las veces resulta ser efectiva, ya que la tarjeta puede sufrir daños irreversibles durante este proceso.

Una segunda reparación es efectuada cuando se detecta algún defecto durante la inspección óptica o más tarde durante el proceso de rayos X, en ambos casos, es realizada por un técnico de forma manual, en la que, dependiendo del defecto, puede ser regresada el flujo del proceso o debe ser desechada sin recuperación alguna de componentes.

Una vez terminado el proceso a nivel SMT, dependiendo del producto final esperado, las tarjetas pueden continuar su proceso de fabricación con procesos de pruebas, como ICT o FCT, PTH, cortado, lavado, procesos de ensamble hasta completar el producto, empackando las unidades, ya sea para enviarlas a los clientes o a los almacenes de la compañía.

### 3.3 Ensamblaje de Tarjetas con componentes

Como continuación al proceso de SMT, puede encontrarse para algunos modelos, el proceso de PTH en el cual, se realiza una inserción de componentes en los orificios de la tarjeta diseñados para este fin, esto puede ser realizado por panel de varias unidades o bien, después de realizar la depanelización de cada una en función del tamaño y diseño de cada tarjeta para facilitar su manejo.



La tarjeta es colocada en una base conocida como pallet para luego colocar los componentes, esperando ser sometida a un proceso de soldadura automática mediante una máquina soldadora por ola, al finalizar el soldado la tarjeta es enviada a una estación de inspección retoque en la que se realiza manualmente el soldado de componentes que después de ser inspeccionados no cubren los requerimientos de calidad, en esta misma estación se realiza la limpieza de residuos que pudieran haberse generado durante el proceso de soldadura cuando para el modelo de tarjeta se permite realizar un retoque manual.

Al finalizar este proceso, las tarjetas que cumplen con la calidad debida son enviadas a una estación de Inspección Óptica automática y posteriormente, cuando el modelo así lo requiere se envían a proceso de inspección por rayos X, en algunas situaciones solo se envía una cierta cantidad de tarjetas como muestra para disminuir el tiempo y recursos en este proceso.

Las tarjetas que presentan alguna falla ya sea mediante inspección manual o automática, son enviadas a una estación de reparación, y es ahí donde se define si la falla puede ser reparada para ser enviada nuevamente a inspección y retoque, en caso contrario, la tarjeta es desechada mediante el procedimiento establecido. Es necesario señalar que, dependiendo del modelo, es permitido realizar una reparación, ya que, en segmentos y piezas críticas, la reparación manual no es permitida en el proceso de fabricación.

Durante este proceso, son varios los factores que intervienen en cada uno de los subprocesos o etapas particulares. Para simplificar la comprensión del proceso, se muestra la Figura 19.

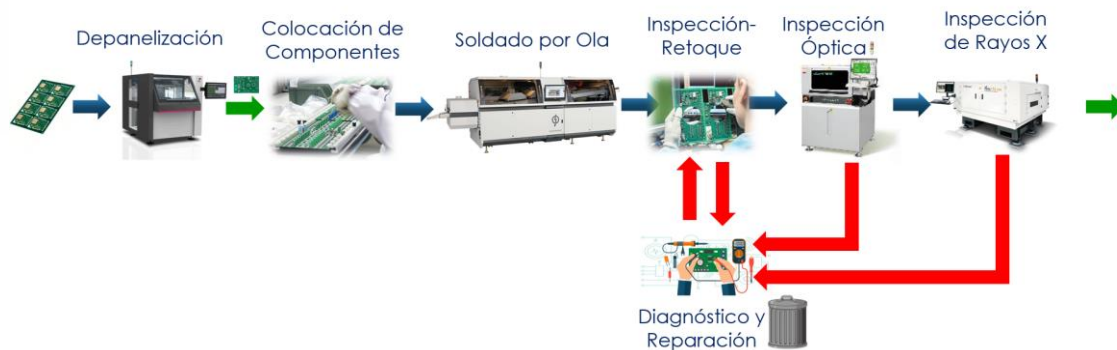


Figura 19. Proceso general de Ensamblaje de Tarjetas con Componentes (PTH). Elaboración Propia.

Dentro del proceso de PTH como se ha mencionado existen algunos subprocesos que se realizan de forma manual y además existen varias inspecciones que son parte del proceso general, por lo que, de forma automática, la parte crítica del proceso se centra en la soldadura por ola, en este proceso se realiza la limpieza de la tarjeta y componentes, removiendo partículas contaminantes para ser soldados posteriormente en la tarjeta, de modo que las tres tareas principales son:

I. Aplicación de flux

Como primer paso para realizar el soldado de componentes, se aplica flux de forma automática a cada una de las tarjetas que serán ingresadas a la soldadora, con el propósito de remover la oxidación de la PCB y componentes, reducir la tensión de la superficie, ayudar en soldado, evitando cortos e insuficiencias, ayudar a refluir a la soldadura después de realizar la unión y finalmente, proteger los elementos de la tarjeta para evitar re-oxidación después de ser soldada.

II. Pre calentamiento de la tarjeta electrónica

En esta etapa, se realiza un calentamiento de las partes a ser soldadas para activar el flux, normalmente el calentamiento en esta etapa es convección por lado inferior de la tarjeta y radiación por el lado superior.

El pre calentamiento es necesario principalmente para evitar choque térmico en los componentes, secar y activar el flux para la operación de soldado, evaporar partículas de flux de otros solventes como el agua, facilitar el relleno de los orificios de soldadura y establecer la temperatura adecuada en la PCBA.

III. Soldadura

Después del pre calentamiento de la tarjeta, la soldadura fundida es liberada a la parte inferior de la placa, permitiendo el soldado de los cables y el patrón de tierra a la tarjeta. Existen dos tipos de soldadura por onda, Chip y Lamda, que es aplicada de acuerdo con la complejidad de la placa, tipo de fundente y acabado.

Las variables identificadas como significativas para la soldadura de ola se muestran en la

Tabla 1.

Tabla 1. Variables significativas para el proceso de soldadura de ola.

Variable
Tasa de flujo de flux
Penetración de flux
Cobertura de flux
Depósito de flux
Tipo de flux (a base de agua, a base de alcohol)
Temperatura de rampa
Temperatura del lado superior de la PCBA
Temperatura del lado inferior de la PCBA
Flux activo
Potencia ultrasónica
Aire del tanque (define la tasa de flujo)
Presión de aire en la boquilla
Ancho de PCB
Cobertura de pallet
Desplazamiento del conveyor
Temperatura del baño de soldadura
Distancia de la boquilla
Spray ON
Spray OFF
Tiempo de soldado
Tiempo de contacto de soldadura
Tiempo de enfriamiento
Tiempo de permanencia
Tiempo de humidificación

Los defectos principales derivados del proceso de soldadura son:

- Baja penetración de flux

Un primer defecto generado durante este proceso es una baja penetración de flux, que desencadena defectos de soldadura, provocando la generación de piezas defectuosas al final del proceso. Este defecto puede ser identificado mediante una prueba de penetración.



Figura 20. Resultado de prueba de penetración de flux, baja penetración.

- Exceso de flux

Contrario a la baja penetración de flux, la prueba de penetración indica un defecto derivado de la aplicación de flux, que es exceso, provocando defectos en las siguientes tareas del proceso.

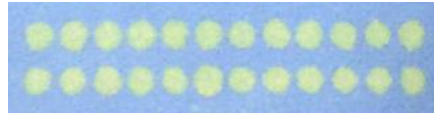


Figura 21. Resultado de prueba de penetración de flux, exceso de flux.

- Residuos

Un proceso deficiente de soldadura de ola deja residuos en la tarjeta electrónica, que evitan el buen funcionamiento, en la mayoría de los casos, al encontrar este tipo de defecto, la pieza no es recuperable mediante ningún proceso de reparación.

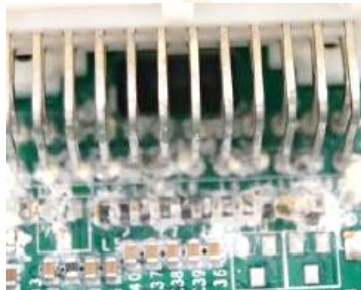


Figura 22. Residuos de flux.

- Crecimiento de dendrita

Es generado a causa de la reacción de residuos con la humedad y voltaje mediante la activación de flux en componentes PTH.

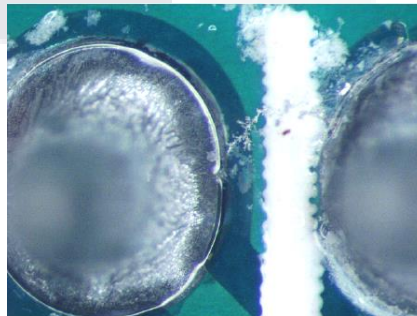
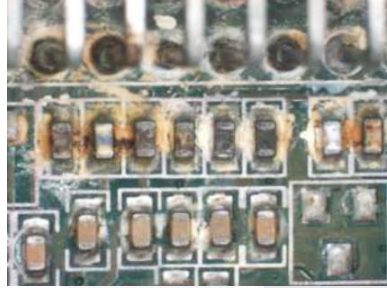


Figura 23. Crecimiento dendrita.

- Corrosión

Es causada por residuos en la PCBA que reaccionan con la humedad y el voltaje, dejando la tarjeta totalmente fuera de las expectativas de calidad.



*Figura 24. Corrosión.*

El proceso de soldadura es la parte esencial en el proceso de PTH para obtener resultados de calidad, conservando la funcionalidad y resistencia de la tarjeta, integrado con los demás subprocesos, encaminan el producto generado a través de ellos a los ensambles finales para los que la PCBA ha sido diseñada, no sin antes cumplir con las pruebas de circuito y funcionalidad.

### 3.4 Selección del proceso: Tecnología de Montaje Superficial.

Después de conocer varios procesos de la compañía, se ha optado por seleccionar un subproceso de SMT.

Aunque los procesos descritos anteriormente, son de gran importancia para la industria, la velocidad del proceso de SMT, así como la repetibilidad del proceso en varias plantas de la misma empresa, han sido factores importantes en la selección de nuestro proceso base.

En el caso del proceso de moldeo, la factibilidad de la obtención de datos representa un punto negativo para nuestro estudio, ya que, la evaluación de este proceso se basa en muestras y las piezas a fabricar son muy variables y los planes de producción no son del todo confiables debido a que la demanda y la producción de los ensambles tiene el común denominador del constante cambio.

Respecto al proceso de dispensado, debido a la localización geográfica de la planta, se vuelve compleja la interacción con el proceso y por lo tanto, los datos necesarios para el análisis no son alcanzables, a pesar de que el proceso es cien por ciento automatizado y de gran

importancia para la empresa, el impacto de la mejora alcanzaría solamente dos de las plantas con que cuenta la compañía.

Siendo el proceso de SMT el adecuado para fines de la investigación, la importancia de este proceso, a nivel local, ha alentado a obtener resultados tangibles para el presente trabajo, permitiendo escalabilidad localmente en 16 líneas de producción y más de mil productos diferentes, sin adentrar en la importancia global para la empresa.

Dado por sentado el proceso, se ha elegido el subproceso de Impresión de Soldadura en pasta, desde la introducción de la PCB hasta su inspección, como se muestra en la Figura 25.



Figura 25. Tecnología de Montaje Superficial: Impresión de pasta e inspección. Elaboración Propia.

El común denominador en las diferentes líneas de producción para distintos productos es encontrar defectos a nivel de Inspección Óptica, generados por los procesos que anteceden esta tarea.

Mediante un recorrido en las diferentes líneas de producción, se identificó que el proceso de Inspección de Impresión de pasta no cuenta con los controles necesarios para determinar si la impresión cumple con los parámetros establecidos para continuar con el proceso, esto debido a que la cantidad de unidades a producir no permite dar la atención debida a cada parte del proceso. Se ha encontrado además que, a pesar de contar con equipos automáticos para realizar esta tarea, que, si bien no son de lo más avanzado, deberían facilitar el proceso, se deja a criterio

de los operadores la distinción de tarjetas buenas y malas, ya que pueden o no representar fallas en los siguientes procesos.

Así mismo, no se cuenta con un criterio de aceptación estándar para los diferentes productos y líneas, debido a la gran cantidad de productos, cambios en la demanda y factores que intervienen en el proceso.

La detección temprana de problemas durante el proceso permite la mejora del mismo desde sus inicios, evitando perdidas de recursos que se van acumulando a lo largo del proceso, incrementando la cantidad de desperdicio de material y tiempo de operación, alentando a la disminución de retrabajo por defectos en este y procesos subsecuentes, además de que todos los productos a ser fabricados mediante esta tecnología tienen su inicio en este proceso.

Es así como se iniciará entonces, un análisis detallado del proceso Impresión de Soldadura en Pasta.

### 3.4.1 Descripción del proceso: Impresión de Soldadura en Pasta – Inspección de Impresión

Volviendo al diagrama inicial de un proceso, podemos definir inicialmente cada elemento en el siguiente Esquema general del proceso de impresión de pasta e Inspección. Elaboración Propia.



Figura 26. Esquema general del proceso de impresión de pasta e Inspección. Elaboración Propia.

TESIS TESIS TESIS TESIS TESIS

A pesar del tiempo que tiene en desarrollo la Tecnología de Montaje Superficial, siempre se encuentran variantes en el proceso, que no permiten definir todos los casos totalmente, por lo que no se ha logrado una implementación del proceso con cero defectos.

Los factores que influyen en este proceso se listan a continuación.

## Recursos

### A. Plantilla o esténcil

Es uno de los recursos principales para efectuar el proceso, ya que permite la impresión de pasta en cada uno de los orificios que resultarán útiles para la colocación de componentes en el siguiente proceso.

La apertura de cada uno de los elementos de esténcil, así como el grosor, material, desgaste y limpieza, son factores claves en el proceso de Impresión.

A lo largo del reconocimiento del proceso, se pudo identificar, que no hay un control para el manejo del esténcil, limpieza o mantenimiento, ni la identificación de piezas que se generan con cada uno.

### B. Pasta

La calidad de la pasta y las características propias, influyen directamente en el proceso, existen algunas variantes de la pasta según los proveedores y el modelo a producir. Para que un proceso se mantenga dentro de control, es necesario que la pasta no tenga una fecha de vida expirada.

### C. Impresora de Soldadura en Pasta

La máquina que realiza el proceso de impresión de forma automática se considera en sí, un recurso para el proceso y tiene impacto en el mismo, este es un recurso indispensable, que debe estar en perfectas condiciones y contar con un mantenimiento preventivo y correctivo adecuado.



#### D. PCB

La tarjeta electrónica, al igual que la pasta, es uno de los insumos principales del proceso, que, en conjunto, generará los productos y las condiciones de salida en el proceso.

El tamaño de la tarjeta, la cantidad de componentes, el grosor entre otros factores, influyen directamente en el proceso de impresión.

#### E. Equipo de inspección de Impresión de Pasta

Considerando la impresión e inspección como un proceso, el equipo de inspección de impresión de pasta es otro de los recursos necesarios para ejecutar el proceso, el cual, debe estar en buenas condiciones para realizar una correcta inspección, existen controles para llevar a cabo el mantenimiento preventivo, sin embargo, en las líneas recorridas, no existe un control del mantenimiento que se genera de forma correctiva.

Existen diferentes equipos en la empresa visitada, de diferentes fabricantes y diferentes versiones, a lo largo de 16 líneas de producción dentro de la misma empresa, se observaron equipos Parmi y Koh Young, los cuales pueden tener diferentes limitantes y factores que pueden precisar la evaluación de las tarjetas durante la inspección.

### **Entradas**

#### I. Temperatura

Se refiere a la temperatura a la que se encuentra el interior de la impresora, esta puede ser ajustada durante el proceso en función del tipo de pasta a utilizar, tipo de estencil y tarjeta, normalmente se presenta en grados centígrados o grados Fahrenheit.

#### II. Presión frontal

Es la presión que se ejerce sobre el estencil para realizar la expansión de pasta en la parte frontal.

III. Presión Trasera

Al igual que la presión frontal, se registra la presión trasera al momento de realizar la impresión, éstas pueden ser diferentes en cada una de las unidades generadas.

IV. Humedad

La humedad es uno de los factores que influyen en el proceso de impresión, ya que la pasta, el estencil y la misma tarjeta son sensibles a este factor. Normalmente, este es un parámetro no controlable en este proceso debido a los costos de adquisición y mantenimiento de humidificadores.

V. Velocidad de separación

Es la velocidad con que se separa la PCB del estencil después de ser impresa.

VI. Tiempo de Ciclo de Impresión

Los segundos que tarda una tarjeta en ser impresa desde su inserción en la impresora hasta su salida, influyen también en la calidad del proceso.

VII. Posición de la mesa

Se refiere a la posición de la mesa como soporte de impresión.

VIII. Velocidad de impresión frontal

Es un parámetro que indica la velocidad en la que se realiza la impresión frontal, en términos coloquiales podría decirse que indica la velocidad de “ida” en el suministro de pasta en la tarjeta.

IX. Velocidad de impresión trasera

Se refiere ahora a la velocidad de “vuelta” en la impresión de pasta en cada una de las tarjetas electrónicas.

## Salidas

### A. Tiempo de Ciclo de inspección

Este parámetro, indica el tiempo en que tarda la tarjeta en ser inspeccionada por el equipo de inspección, puede ser distinto en función del tipo de tarjeta a inspeccionar y el programa a utilizar.

### B. Área

Se refiere al área que ocupa la pasta en la tarjeta y es diferente según el componente a ser soldado. Debido a la cantidad de pines de una tarjeta, es necesario considerar el área con un porcentaje en lugar de una medida absoluta, para identificar la cobertura de la pasta en el pad.

### C. Altura

Indica la altura que tiene la pasta en la tarjeta, alturas elevadas o falta de altura refieren una mala impresión.

### D. Volumen

El volumen de impresión, se considera una variable de salida, ya que indica el volumen que ocupa la pasta en los componentes de la tarjeta, al igual que el área, se busca encontrar el porcentaje de volumen de pasta suministrado para cada uno de los pines de la tarjeta, ya que la evaluación precisa de cada componente aumenta la complejidad de la inspección.

### E. Desplazamiento X

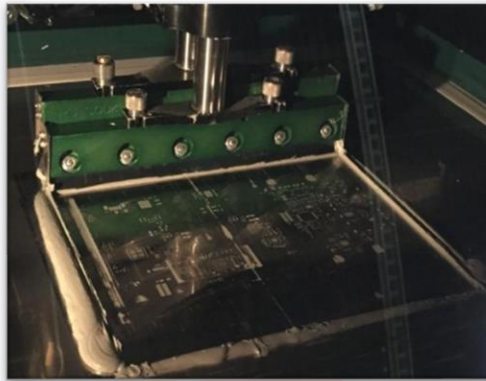
Indica si el componente fue colocado correctamente en la ubicación definida o si existe algún desplazamiento en la coordenada X.

## F. Desplazamiento Y

Indica si el componente fue colocado correctamente en la ubicación definida o si existe algún desplazamiento en la coordenada Y.

Conjuntando los valores de entrada y salida del proceso, así como los recursos, se ha logrado tener una definición general, aplicable a varios productos, factores adicionales pueden influir dependiendo del producto, obteniendo variaciones durante el proceso general.

En la Figura 27 se puede observar la impresión de pasta, a través de la Tecnología de montaje superficial mediante el uso de un estencil.



*Figura 27. Impresión de Pasta utilizando plantillas.*

### 3.4.2 Selección de modelo

Con la finalidad de orientar la investigación a un producto en específico, se ha seleccionado el modelo al que nos referiremos de aquí en adelante como ABC-123-456-00-TOP. El cual es uno de los modelos que presenta mayor cantidad de problema durante el proceso y es, además, en el que, por experiencia del área de ingeniería, no se ha logrado controlar como algunos otros.

La producción a nivel SMT del modelo seleccionado, requiere procesamiento por ambos lados de la tarjeta, bottom y top, o lado A y B, o lado 1 y 2.

La cantidad de componentes es mayor del lado B o top de la tarjeta, por lo que, en el proceso, se imprime primero el lado A, posteriormente lado B.

Por la naturaleza del proceso, el lado B, es el que presenta mayor cantidad de problemas de calidad, y al ser en orden cronológico el segundo en ser impreso, produce desperdicio de material y recursos tanto para el lado A como para lado B.

Se ha identificado un esquema general de los elementos del modelo ABC-123-456-00-TOP como complemento de la identificación de la tarjeta física, en los que se puede apreciar que la tarjeta cuenta con 2567 pines, de los cuales se indica colocar 2511 componentes en los procesos siguientes, de los componentes indicados, de acuerdo al programa de impresión, no se colocará pasta para 16 de los 2511 componentes, ya que, para esta tarjeta en particular de la serie, la función no requiere esos 16 restantes.

#### 3.4.3 Selección de línea de producción

Después de seleccionar el modelo, se ha identificado que al igual que la mayoría, puede ser producido en cualquier línea de producción de las más de 15 líneas habilitadas, sin embargo, los planes de producción sugieren la utilización de la línea C como línea primaria.

La línea C cuenta con un equipo de impresión DEK del fabricante ASM y un equipo de inspección Parmi.

La línea está proyectada para generar 105 unidades por hora para el lado B de la tarjeta seleccionada, cifra que normalmente se alcanza si no existe algún problema en alguno de los equipos que se requieren durante el proceso. Sin embargo, la eficiencia del proceso llega a máximo 90%, mostrando cantidades de hasta 60% después de la inspección, lo cual, tiene un impacto negativo al final del proceso productivo.

Por la cantidad de unidades generadas, y el costo al nivel de tarjeta impresa, no se ha tratado el problema de calidad de la forma en que se requiere, ya que el análisis se presenta hasta este punto del proceso, sin embargo, cuando una tarjeta con este defecto continua el proceso de fabricación, el costo se incrementa de los 10 a los 64 dólares, provocando pérdidas monetarias y de recursos en la línea para satisfacer las demandas en la producción aunado de los desechos generados, ya que el material ya no puede ser recuperado o utilizado en algún otro proceso.

La Figura 28 muestra el esquema general de la línea d producción 6 para el proceso de SMT.

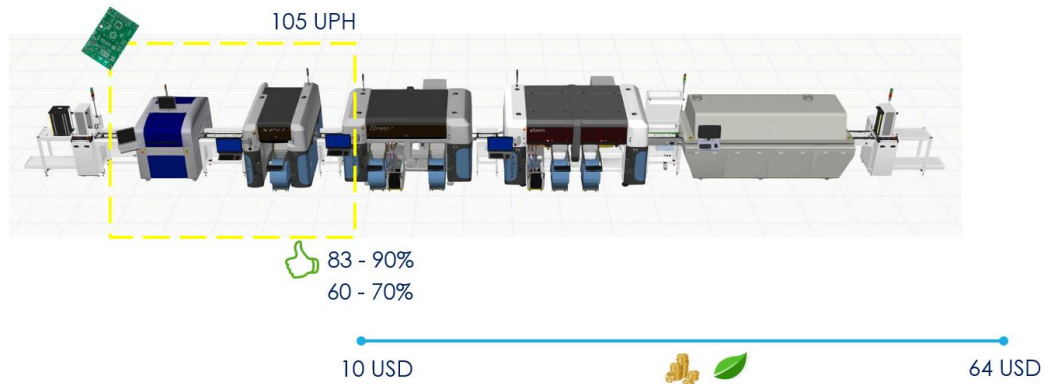


Figura 28. Esquematación de una línea de producción. Elaboración Propia.

### 3.4.4 Descripción de defectos generados por el proceso

Al revisar algunas tarjetas generadas en la línea C para el modelo definido anteriormente, se han encontrado diferentes defectos a lo largo de varios lotes de producción, en diferentes periodos. Se realizará una breve descripción sobre los más frecuentes.

#### Insuficiencia de soldadura

Se genera cuando no se logra el suministro suficiente de pasta para uno o varios pines de la tarjeta. La Figura 29 muestra este defecto en más de un pin.

El contar con al menos uno dentro de la tarjeta, supone un mal funcionamiento de la misma y fallas durante las pruebas. Sin embargo, es a consideración del operador, si el pin puede ser “retocado” de forma manual, es decir si puede colocarse pasta para evitar el deshecho de la tarjeta. Es necesario mencionar, que, por el tamaño del componente, no siempre se puede realizar esta operación, y que, en ocasiones, realizarlo de forma manual genera otro tipo de defectos como cortos por el exceso.

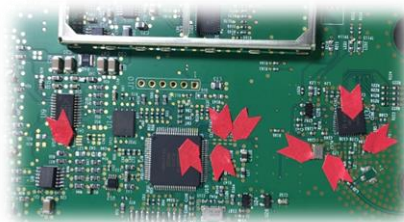


Figura 29. Insuficiencia.

**Daño del lado A**

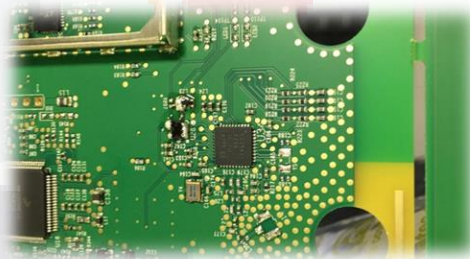
Como se ha mencionado, la tarjeta cuenta con dos lados, y al realizar la impresión del segundo lado o alguna limpieza o reparación manual, el primero lado sufre un daño, el cual es irreparable.



*Figura 30. Daño del lado A.*

**Exceso de soldadura**

Por el contrario de la insuficiencia, en este defecto se puede percibir que se coloca soldadura demás en alguno de los pines, provocando cortos al unirse con uno o varios pines, generalmente, este defecto no puede ser reparado de forma individual, sin embargo, se puede realizar una limpieza general de la tarjeta, que en la mayoría de los casos resulta ser contraproducente al dañar el lado A de la tarjeta.

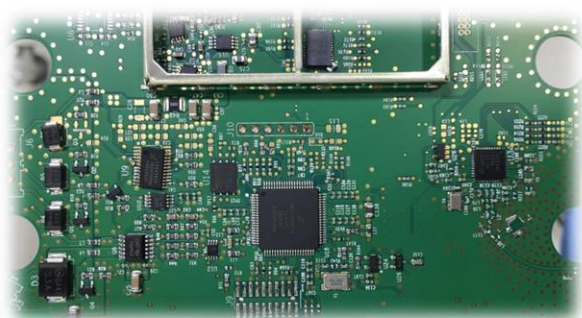


*Figura 31. Exceso de soldadura.*

**Desplazamiento**

Otro defecto común es el desplazamiento del componente, que si bien, no es propio de la impresión, es provocado por la misma ya que en su mayoría, esto es el resultado del desplazamiento en la pasta, que se identifica en la posición X y Y, provocando que el componente sea mal colocado y se identifique como desplazado.

La detección temprana de este defecto y mejor aún evitarlo a través de la mejora del proceso, permitiría la pérdida de recursos adicionales posteriores a la impresión.



*Figura 32. Desplazamiento.*





## Capítulo 4. “Voz del proceso”.

Un proceso no puede ser mejorado si no es medible, es decir si no se tiene reconocimiento del proceso a través de los datos.

Hasta el momento se ha recorrido el proceso de forma física, pero, no es posible realizar un análisis si no es de forma empírica en la línea de producción, donde solamente las salidas del proceso nos ayudan a identificar los problemas que pudieron haber ocurrido, los cuales generan defectos que ya no pueden ser reparados.

Es entonces donde se encuentra la necesidad de registrar los datos que se generan a lo largo del proceso para realizar el análisis y “escuchar” lo que puede decir el proceso a través de estos datos que, al ser interpretados, generen información de utilidad, aportando a la mejora del mismo y con ello de la calidad del producto.

Es aquí donde las diferentes técnicas computacionales comienzan a mostrar importancia para la empresa, el almacenamiento, procesamiento, transferencia y datos deben ser integradas a los procesos existentes de la empresa, sin afectarlos, buscando realizar la menor cantidad de cambios posibles. Un cambio, por menor que sea en el ámbito industrial, supone inversión de recursos, desde humanos, tiempo y hasta monetaria para ser aplicado, los cuales no siempre son factibles por la velocidad de los procesos mismos.

### 4.1 Generación de datos.

La siguiente tarea, es generar la mayor cantidad de información posible para el proceso de impresión – inspección de la línea C para el producto ABC-123-456-00-TOP.

La industria 4.0 parte de la visibilidad de los datos, pasando por la compresión, optimización y la optimización automática, conceptualmente, parece ser sencillo, sin embargo, en el ambiente real en las grandes industrias, se puede observar que no siempre los datos están computarizados y mucho menos conectados, como lo es el caso del proceso a analizar.

Se pretende hacer uso del equipo existente para generar datos que sean de utilidad para el análisis, con la menor afectación posible de los procesos, optimizando la utilización de los equipos involucrados en el proceso.

**Impresora de Pasta**

El modelo de impresora que se utiliza en línea C, es capaz de generar un archivo con datos sobre la impresión de cada una de las tarjetas, al configurar el equipo, seleccionado los parámetros para los que se requiere obtener datos, es posible guardarlo dentro del equipo.

Es posible leer el formato generado con extensión .DEK de forma lineal, para cada parámetro, sin embargo, estos datos no representan valor para el proceso si se encuentran aislados en la impresora, además de que almacenarlos en el equipo, puede llegar a afectar el funcionamiento de la impresora si no se tiene cuidado en el manejo de estos archivos, ya que los equipos cuentan con una capacidad limitada de almacenamiento.

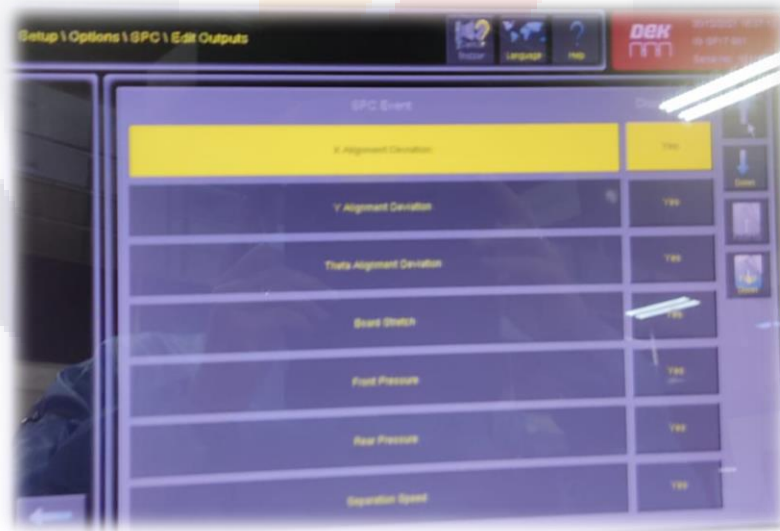


Figura 33. Interfaz de configuración de salida de datos de Impresión de Pasta.

Para esta etapa, como prueba, se genera una carpeta de forma temporal para guardar los archivos.

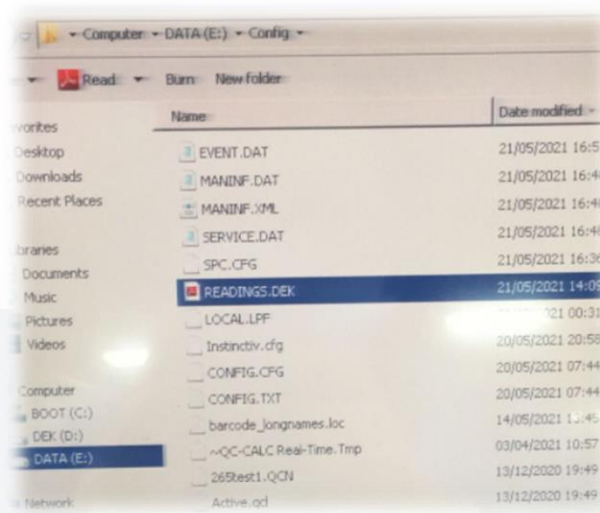


Figura 34. Generación de archivo con datos de impresión de pasta.

Al paso de algunas tarjetas se observa que en la carpeta se encuentra solo un archivo, agregando la información para cada una de las tarjetas, conservando siempre uno solo.

```

READINGS.DEK - Notepad
File Edit Format View Help
START, 44560.2135764

USER,FACTOR1,Squeegee,1
USER,FACTOR2,Batch Count,9770
USER,FACTOR3,2D Inspect,0000
USER,FACTOR4,Align Inspect,0000
USER,FACTORS,Offset,None

USER,FACTOR,S,6,Remote Barcode,NO_CODE
USER,FACTOR,S,7,Machine ID,051159
USER,FACTOR,S,8,Machine Location,
USER,FACTOR,S,9,Stencil Barcode,NO_CODE
USER,FACTOR,S,10,Software Version,09 SP17
USER,FACTOR,S,11,Fwd Offset,0.010:0.000:0.000
USER,FACTOR,S,12,Rear Offset,0.015:0.010:0.000

END,44560.4927546

START, 44560.2135764

DATA,X Error (mm) ,0.0000,0.0000,0.0250,-0.0250,0,4
DATA,Y Error (mm) ,0.0000,0.0000,0.0250,-0.0250,0,4
DATA,Theta Error (arcsec.) ,0.0000,0.0000,31.0000,-31.0000,0,4
DATA,Board Stretch (mm) ,0.0000,0.0000,0.1000,-0.1000,0,4
DATA,Front Pressure (kg) ,11.99086,12.00000,0.20000,-0.20000,0,5
DATA3,Front Pressure (kg) Range ,11.99086,12.00000,0.20000,-0.20000,12.10183,11.65796,(null),(null),0.16000,-0.16000,0,5
DATA,Rear Pressure (kg) ,12.78814,12.00000,0.20000,-0.20000,0,5
DATA3,Rear Pressure (kg) Range ,12.78814,12.00000,0.20000,-0.20000,12.93734,12.68930,(null),(null),0.16000,-0.16000,0,5
DATA,Sep. Speed (mm/s) ,3.0,3.0,0.5,-0.5,0,1
    
```

Figura 35. Log generado por la Impresora de pasta.

La Figura 35 muestra parcialmente el contenido de uno de los archivos generados por la impresora, el cual, muestra los parámetros identificados como de entrada para nuestro proceso

entre algunos otros propios de la máquina, que, de acuerdo con el análisis del proceso al no presentar gran variación, serán descartados para este análisis. Además, incluye del nombre de la familia de modelos que se imprime. La hora se encuentra en un formato no legible comúnmente, pero ha sido convertido a un formato que pueda ser interpretado.

### Inspección de Impresión de Pasta

De forma similar, se busca generar datos sobre la inspección de pasta.

Los resultados de la investigación sobre este tipo de equipos han dado un gran número de opciones y alternativas de equipos avanzados que pueden proporcionar este tipo de datos y que pueden ser conectados con otros equipos en el proceso, sin embargo, el costo es elevado, además de que la compra supone también un esfuerzo en la inversión para cada una de las líneas de producción y recursos para su instalación y operación, modificación de la documentación existente para los procesos productivos actuales.

Entonces, intentando optimizar los recursos existentes, se exploran las opciones que parecen limitadas, pero no nulas.

La Inspección genera un archivo por cada una de las tarjetas que pasan por el mismo, con la referencia del pin de impresión, así como valores de volumen, desplazamiento, área y altura, el formato obtenido es tipo csv.

Se realiza la configuración correspondiente en la máquina para la generación del archivo.

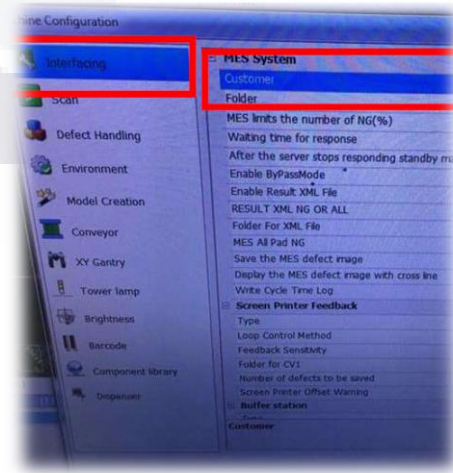


Figura 36. Interfaz de configuración de salida de datos de Inspección de Impresión de Pasta.

JOB PATH	BARCODE	INDEX	DATE	S.TIME	E.TIME	CYCLE	JOB	RESULT	USER	LOTINFO	MACHINE		
Barcode	BoardID	Component	PadID	Pin Number	Height	Area	Area(%)	Volume	Operator	Volume(%)	OffsetX	MACHINE001	Result
	0		0		113.3591	789.2778	106.4		80.3838	106.7	-8.011	17.7624	GOOD
	0		1		128.8057	255.6956	110.1		29.1787	123.7	8.5404	3.7156	GOOD
	0		2		109.1053	556.7728	103		54.806	99.8	-5.2623	27.846	GOOD
	0		3		111.4219	546.5769	101.1		55.1511	100.4	10.2428	37.2867	GOOD
	0		4		110.6557	564.1697	104.3		56.5274	102.9	-0.7989	22.054	GOOD
	0		5		109.2127	571.5667	105.7		56.3497	102.6	-8.8254	31.8025	GOOD
	0 r138		6	1	126.6264	256.6952	110.5		28.714	121.7	7.5013	10.3266	GOOD
	0 r138		7	2	126.6397	244.5002	105.3		27.316	115.8	-7.4076	10.8083	GOOD
	0 c95		8	1	120.9845	821.0649	110.7		87.5043	116.1	4.4603	-0.8146	GOOD
	0 c95		9	2	114.6261	821.0649	110.7		84.0974	111.6	1.8967	7.2418	GOOD
	0		10		104.494	585.9608	108.4		55.6077	101.2	-16.1607	30.2318	GOOD
	0 c94		11	1	107.4485	836.8585	112.9		82.0531	108.9	-3.2261	7.792	GOOD
	0 c94		12	2	105.0065	860.8486	116.1		82.2844	109.2	-3.7351	1.6902	GOOD
	0 r139		13	1	123.4402	255.8956	110.2		28.4622	120.6	12.5847	-0.7218	GOOD
	0 r139		14	2	120.7681	263.6924	113.5		28.5332	120.9	16.3207	19.1659	GOOD
	0		15		106.2788	568.368	105.1		54.5	99.2	-21.9992	13.1039	GOOD
	0 c93		16	1	111.5623	837.4582	112.9		83.7339	111.1	-5.6296	12.6936	GOOD
	0 c93		17	2	106.2196	845.0551	114		80.4187	106.7	-9.7901	13.3624	GOOD
	0		18		111.3226	547.7764	101.3		55.1483	100.4	-12.5605	26.4573	GOOD
	0 c92		19	1	110.5272	818.4659	110.4		81.2052	107.8	-10.8763	16.0888	GOOD
	0 c92		20	2	107.3958	812.8682	109.6		78.1458	103.7	-8.646	15.7874	GOOD
	0		21		103.7512	454.7736	107.6		52.713	98.1	-17.3688	26.9734	GOOD

Figura 37. Log generado como resultado de la inspección de la impresión de pasta.

Inicialmente, los archivos generados son guardados de forma temporal en la misma máquina, recordando que las capacidades de almacenamiento son limitadas, se genera un método de recolección de datos para transferir los datos a formatos más sencillos.

**Otros datos**

Se ha mencionado que existen otros datos como parte de los recursos del proceso que son de utilidad para el análisis del proceso, como lo es la relación del estencil con las unidades producidas, la limpieza y el desgaste.

En la línea se pueden utilizar 3 estenciles para la fabricación de la tarjeta, pero no se lleva un registro de los cambios en cada uno.

Se ha propuesto un mecanismo para el registro del mismo integrándolo al sistema de control de producción, mediante la identificación única de cada estencil, registrando el cambio en la línea, así como el lavado. Haciendo uso del sistema existente, se agrega una tarea que permita el reconcomiendo del mismo. Conceptualmente parece una tarea sencilla y se ha logrado implementar en el turno 1.

Con los cambios tan frecuentes de personal operario y mantenimiento, la implementación de esta tarea, no ha resultado ser de lo más eficiente, provocando que se realice el registro una y otra vez para desbloquear la línea sin resultados satisfactorios.

Como se ha comentado, cambios sencillos en el área operativa, requieren un esfuerzo integral de personal operario, ingeniería y calidad para su aplicación. Por lo cual, implementarlo

para un solo modelo, en una sola línea, genera confusión y afecta la estandarización de los procesos, por lo que se ha generado un plan general para llevar a cabo la implementación de estos controles, deshabilitándolo temporalmente en la línea piloto.

Por esta limitante, el estencil no será considerado en la generación de datos a ser analizados.

Al igual que el estencil, la pasta, no lleva un registro en la línea, solo se incluye la validación en el sistema de control de producción que la vida útil no tenga una fecha de expiración pasada, por lo que se asume, que es la adecuada para el proceso.

Se considerará también que los equipos de impresión e inspección se encuentran en condiciones óptimas para realizar el proceso, ya que, de no serlo, no existe un procedimiento para reemplazar los equipos de inmediato o tomar acciones que no estén contempladas en una reparación.

Partiendo entonces de estos datos, se continúa con la transferencia a formatos que puedan ser sencillos de analizar.

#### 4.2 Recolección de datos.

Antes de realizar la recolección de datos, se realiza una revisión de la infraestructura de la empresa, con la finalidad de aprovechar los recursos existentes y adaptarse a los recursos de Sistemas de Información con que se cuenta.

Las bases de datos estándar son bases de datos estructuradas, debido a que no se ha encontrado la necesidad hasta el momento de la utilización de las no estructuradas. SQL server es el estándar para manejo de bases de datos, y el licenciamiento está cubierto para diferentes servidores dentro de la empresa.

Los sistemas operativos son Windows y más del 90% de los servidores son también variantes de versiones de Windows para servidores de acuerdo con los diferentes propósitos.

Debido a las regulaciones de licenciamiento, no es posible agregar una nueva tecnología, sin antes realizar el procedimiento necesario y las compras de los mismos, ya que el uso de

tecnología o software no regulado debidamente puede implicar multas y suspensión de operaciones.

Contando con estos recursos, entonces, se procede con la solicitud de la conexión de la impresora de pasta y de la máquina de inspección para realizar la transferencia de datos. Se realiza la instalación de una nueva canaleta para bajar la conexión de red hasta la línea C, no sin antes explorar la factibilidad de la conexión por puerto Ethernet de ambos equipos.

Conectado el equipo, se ha solicitado un servidor en la misma empresa, ya que, los datos no pueden ser transferidos a sistemas externos por seguridad.

El servidor, en su etapa inicial de prueba, cubre con los requerimientos de sistema operativo y licenciamiento de la empresa.

Por el sistema operativo con que cuentan las máquinas y la criticidad de los equipos, se ha optado por generar una VLAN interna para los procesos contenidos en SMT, agregando inicialmente la impresora, la máquina de inspección y el servidor obtenido.

La conexión se realiza mediante IP fija, a cada uno de los equipos para establecer conexiones entre sí.

Con los recursos actuales se ha podido realizar la transferencia de archivos desde las máquinas al servidor mediante un servicio de Windows, instalado en el servidor, centralizándolos localmente para ser procesados.

Se ha solicitado un segundo servidor con SQL server, ajustándonos nuevamente a los recursos de la empresa, que si bien, no es requisito que utilice este tipo de base de datos, ha sido por estandarización, la selección de este tipo.

Generando una base de datos para la obtención de los parámetros, de forma separada por cada una de las máquinas, se ha complementado el servicio de Windows encargado de transferir los archivos con la transferencia de datos de cada archivo mediante un parser a la base de datos. Por la diferencia de formatos entre máquinas, se han creado servicios independientes para cada máquina.

La Figura 38 muestra un esquema general de la transferencia de archivos .DEK, generados por la impresora hacia al servidor central y posteriormente la base de datos.

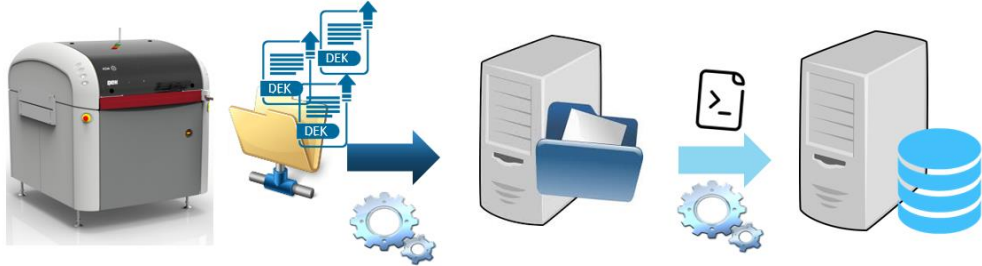


Figura 38. Transferencia de datos desde Impresora de Pasta.

De igual forma, como se ha descrito anteriormente, la siguiente figura muestra la transferencia de archivos de datos .csv a la base de datos, pasando por el servidor central.

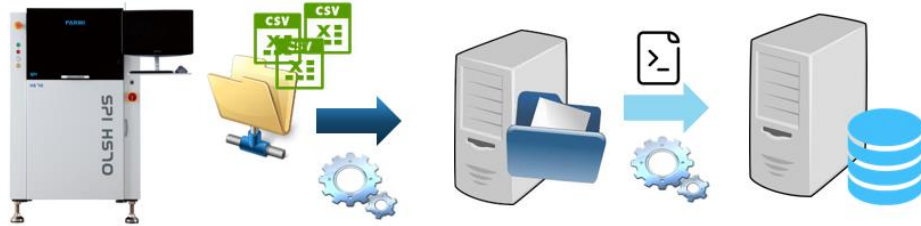


Figura 39. Transferencia de datos desde equipo de Inspección de Impresión.

Conjuntando los datos de los dos equipos, la adquisición de datos puede resumirse en con la siguiente Esquematzación de proceso de recolección de datos. Elaboración Propia.

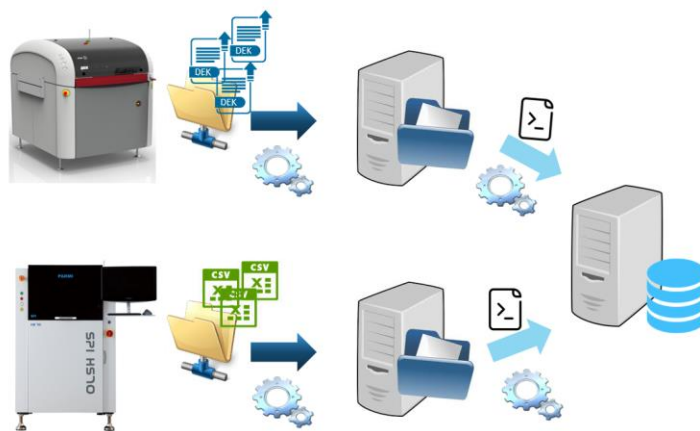


Figura 40. Esquematzación de proceso de recolección de datos. Elaboración Propia.



### 4.3 Descripción de datos.

Los datos obtenidos durante el procedimiento anterior se listan en las siguientes secciones.

#### 4.3.1 Entradas

De los datos generados por la impresora de pasta, se han tomado los que resultan ser más significativos y que en su mayoría son controlables.

Tabla 2. Parámetros de entrada.

No.	Parámetro	Tipo	Tipo de dato
	Familia de tarjeta	Entrada	Cuantitativo, fecha
<b>F.0</b>	Temperatura	Entrada	Cuantitativo, real
<b>F.1</b>	Presión Frontal	Entrada	Cuantitativo, real
<b>F.2</b>	Presión Trasera	Entrada	Cuantitativo, real
<b>F.3</b>	Humedad	Entrada	Cuantitativo, real
<b>F.4</b>	Velocidad de separación	Entrada	Cuantitativo, real
<b>F.5</b>	Tiempo de ciclo de impresión	Entrada	Cuantitativo, real
<b>F.6</b>	Posición de la mesa	Entrada	Cuantitativo, real
<b>F.7</b>	Velocidad de impresión frontal	Entrada	Cuantitativo, real
<b>F.8</b>	Velocidad de impresión trasera	Entrada	Cuantitativo, real
	Fecha y hora de salida de impresión		Atributo, fecha

Se ha agregado además la hora de inserción en la base de datos, con la finalidad de conocer el tiempo que tarda el dato en ser transferido desde la máquina, un identificador del nombre de la máquina y modelo, para permitir la escalación de la investigación a más máquinas de diferentes líneas y números de parte.

De los datos obtenidos, se ha realizado una purga, dejando solo los datos más recientes, como referencia, se han obtenido los estadísticos descriptivos para los datos de entrada tomados a partir del día 22 de Enero hasta el 04 de marzo, como una muestra significativa.

Tabla 3. Estadísticos descriptivos para datos de entrada

Variable	N	N*	Mean	SE Mean	St Dev	Variance	Minimum	Q1	Median	Q3	Maximum	Range
Temperature	26007	0	24.428	0.0178	2.873	8.254	21.000	21.800	22.500	27.100	30.500	9.500
Front Pressure	26007	0	10.995	0.000586	0.0945	0.00892	8.778	10.969	11.000	11.030	12.040	3.262
Rear Pressure	26007	0	10.994	0.000884	0.143	0.0203	0.000000	10.964	10.999	11.034	12.081	12.081
Humidity	26007	0	35.474	0.0246	3.961	15.689	27.000	32.800	34.900	37.500	56.100	29.100
Separation Speed	26007	0	2.9920	0.000676	0.1090	0.0119	1.5000	3.0000	3.0000	3.0000	3.0000	1.5000
Cycle Time	26007	0	18.192	0.0374	6.034	36.404	0.0000	17.800	18.000	18.100	479.400	479.400
Table Position	26007	0	0.012972	0.000021	0.003334	0.000011	-0.010000	0.010000	0.013000	0.015000	0.025000	0.035000
Front Print Speed	26007	0	109.74	0.0213	3.44	11.82	0.0000	110.00	110.00	110.00	110.00	110.00
Rear Print Speed	26007	0	109.76	0.0203	3.27	10.69	65.00	110.00	110.00	110.00	110.00	45.00

En Tabla 3, se muestran los Estadísticos descriptivos para datos de entrada, en los cuales se puede observar que, para 26007 datos representativos, se presenta una temperatura promedio de 24.428°, y los valores oscilan en un rango de 9.5°, lo cual, significa que si existe una variación significativa en este parámetro para el rango de datos seleccionado.

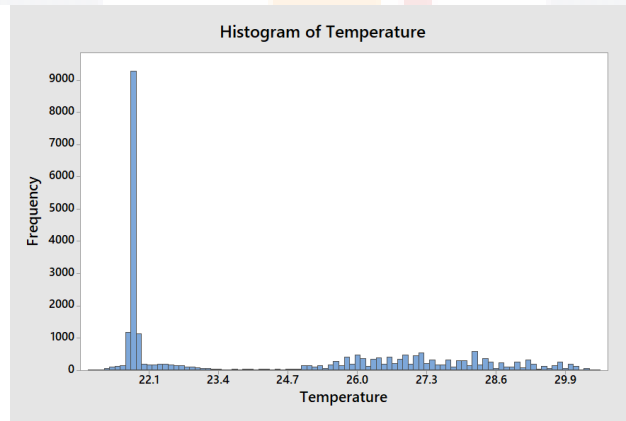


Figura 41. Histograma de la temperatura.

La presión frontal y trasera tienen una media similar, de 10.995 y 10.994 kg, la presión trasera por su parte muestra un valor mínimo de 0, lo cual puede representar un valor por descartar, ya que, de acuerdo con el proceso, siempre se debe ejercer una presión para realizar la impresión. Esto será tomado en cuenta en los rangos utilizados para posteriores conjuntos de datos.

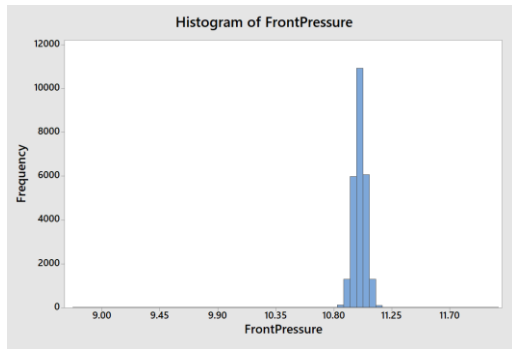


Figura 42. Histograma de la presión frontal.

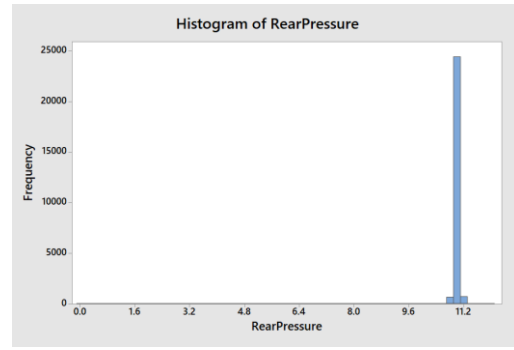


Figura 43. Histograma de la presión trasera.

Para el parámetro Humedad, se puede observar un rango de 29.100, presentando una varianza de 15.689, la cual es la segunda más alta del conjunto de parámetros, este caso en particular, refleja totalmente la realidad del proceso, ya que este es un parámetro que no está siendo controlado actualmente, es decir, el proceso no se encuentra algún cuarto o ubicación en la que la humedad pueda ser controlada, por lo cual, numéricamente, se ve reflejado el cambio abrupto entre los valores máximo y mínimo.

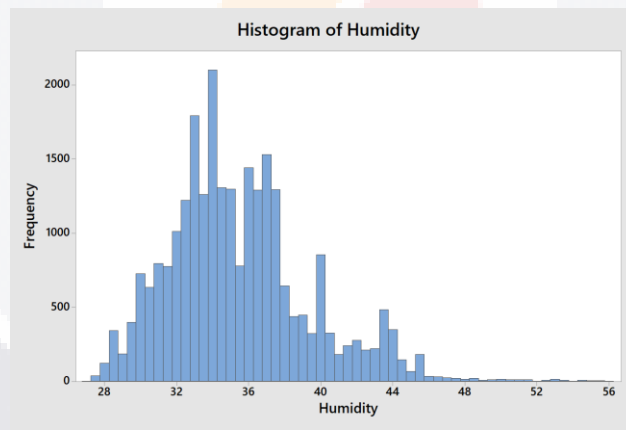


Figura 44. Histograma de la humedad.

La velocidad de separación presenta la segunda desviación estándar más baja del conjunto de datos, 0.1090 generando una varianza de 0.0119 en este conjunto de datos, indicando poca variación para este parámetro en el proceso.

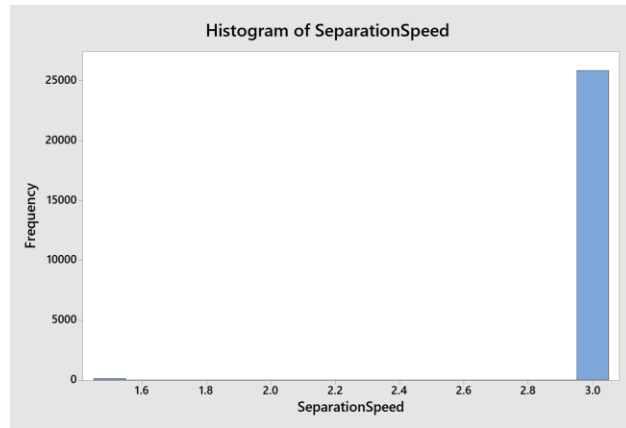


Figura 45. Histograma de la velocidad de separación.

El tiempo de ciclo promedio es de 18.192 segundos y la mediana de los datos es de 18 segundos, lo cual, describe, que la mayoría de las tarjetas tardan alrededor de 18 segundos en ser impresas, sin embargo, se encuentran valores extremos de 0 y 479.4, los cuales, se consideran anomalías durante el proceso o la obtención de datos, ya que una tarjeta electrónica no puede ser impresa en 0 segundos y los 479 segundos tomados para la impresión, representan problemas en la impresora, en los que la tarjeta sigue dentro de la maquina sin embargo no se realiza la tarea de forma usual.

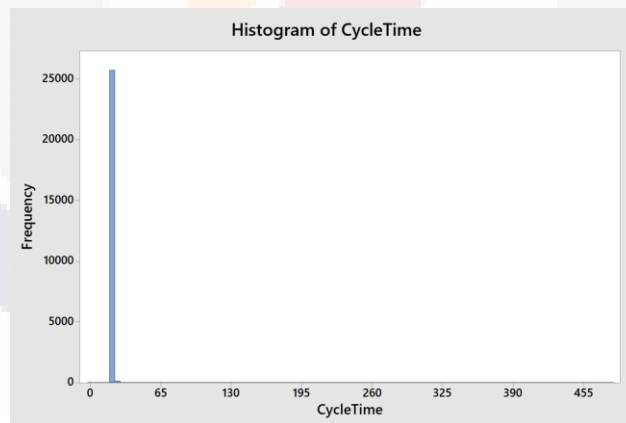


Figura 46. Histograma del tiempo de ciclo de impresión.

La posición de la mesa es el valor que muestra desviación estándar y varianza menor del conjunto de datos, acercándose a 0, lo cual, indica que, durante el proceso, la posición se mantiene estable o con variación prácticamente nula.

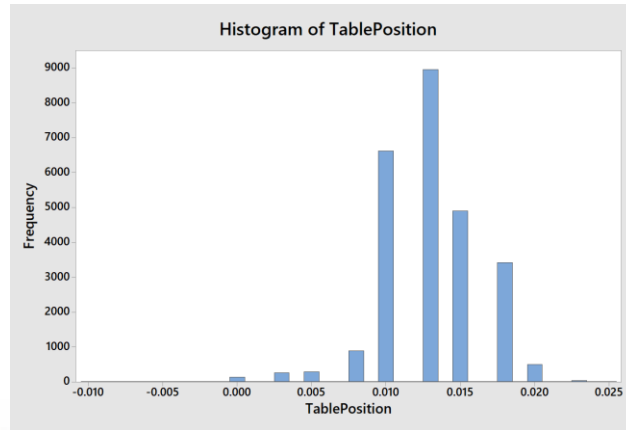


Figura 47. Histograma de la posición de la mesa.

Los datos para las velocidades de impresión frontal y trasera muestran en promedio 109.7 con una diferencia de 0.02 entre sí, sin embargo, para la presión frontal se observa un valor mínimo de 0.0, mientras que el la trasera el valor mínimo es de 65.0, lo cual puede ir de la mano con los valores mínimos obtenidos para la presión frontal y trasera descritos anteriormente, la velocidad para las presiones máxima definida en la máquina impresora es 110.0, manteniéndose como valor máximo para ambos parámetros.

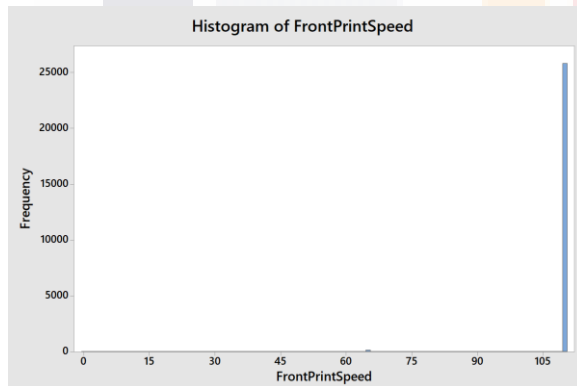


Figura 48. Histograma de la velocidad de impresión frontal.

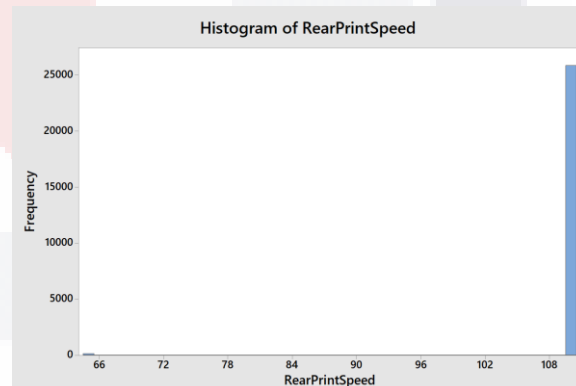


Figura 49. Histograma de la velocidad de impresión trasera.

#### 4.3.2 Salidas

Derivados del proceso de inspección, resumiendo los datos obtenidos por cada uno de los archivos generados, se presentan los datos listados en la Tabla 4.

Tabla 4. Datos de inspección de impresión.

No.	Parámetro	Tipo	Tipo de dato
	Ruta de programa de inspección		Atributo, texto
	Índice		Atributo, entero
	Hora de entrada a inspección		Atributo, fecha
	Hora de salida de inspección		Atributo, fecha
	Número de parte		Atributo, texto
	Componente		Atributo, texto
	Pad		Atributo, entero
	Número de Pin		Atributo, entero
<b>F.9</b>	Tiempo de ciclo de inspección	Salida	Cuantitativo, real
<b>F.10</b>	Altura	Salida	Cuantitativo, real
<b>F.11</b>	Área	Salida	Cuantitativo, real
<b>F.12</b>	Porcentaje de área	Salida	Cuantitativo, real
<b>F.13</b>	Volumen	Salida	Cuantitativo, real
<b>F.14</b>	Porcentaje de Volumen	Salida	Cuantitativo, real
<b>F.15</b>	Desplazamiento X	Salida	Cuantitativo, real
<b>F.16</b>	Desplazamiento Y	Salida	Cuantitativo, real

Se conoce además el número de máquina como dato tipo texto y el usuario que opera la máquina, también en formato de texto, sin embargo, se identifican como datos no relevantes, ya que siempre contienen el mismo valor.

De igual forma que se realizó en los datos de entrada, se agregó en la base de datos relacional, la fecha y hora de inserción además de los datos listados, así como el nombre de la máquina impresora de pasta y el usuario que procesa dichos datos.

Al realizar la validación de los datos obtenidos para las salidas del proceso en conjunto, se observa que existen valores en cero para diversas salidas, esto debido a que se utiliza la misma placa o tarjeta base para la impresión de modelos similares o de la misma serie.

Para nuestro modelo o número de parte bajo estudio, se ignorarán aquellos pads que no cuenten con pasta, en relación al componente u30, ya que no serán relevantes para el modelo a

obtener, los 16 pads son 2019, 2020, 2021, 2022, 2038, 2039, 2040, 2041, 2052, 2053, 2054, 2055, 2071, 2072, 2073 y 2074 correspondientes a los pines L12, L13, L14, M6, M7, M8, M9, N6, N7, N8, N9, P11, P12, P13 y P14 respectivamente, la Figura 51 muestra la relación de estos pads en un archivo.

JobPath	Barcode	Index	entertime	ExitTime	CycleTime	ModelName	UserFromLog	LotInfo	MachineFromLog	Height	Barcode	BoardID	Component	PadID	PinNumber
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2020	L12		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2022	L14		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2039	M7		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2041	M9		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2053	N7		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2055	N9		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2019	L11		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2021	L13		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2038	M6		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2040	M8		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2052	N6		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2054	N8		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2072	P12		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2074	P14		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2071	P11		
			2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator	MACHINE001	0.00000000000000000000	0	u30	2073	P13		

Figura 50. Componente u30 sin impresión de pasta.

JOB PATH	BARCODE	INDEX	DATE	S.TIME	E.TIME	CYCLE	JOB	RESULT	USER	LOTINFO	MACHINE	
			01/14/2022	23:49:49	23:49:58	9.77	GOOD		Operator		MACHINE001	
Barcode	BoardID	Component	PadID	Pin Number	Height	Area	Area(%)	Volume	Volume(%)	OffsetX	OffsetY	Result
	0	u30		2019 L11	0	0	0	0	0	0	0	0 GOOD
	0	u30		2020 L12	0	0	0	0	0	0	0	0 GOOD
	0	u30		2021 L13	0	0	0	0	0	0	0	0 GOOD
	0	u30		2022 L14	0	0	0	0	0	0	0	0 GOOD
	0	u30		2038 M6	0	0	0	0	0	0	0	0 GOOD
	0	u30		2039 M7	0	0	0	0	0	0	0	0 GOOD
	0	u30		2040 M8	0	0	0	0	0	0	0	0 GOOD
	0	u30		2041 M9	0	0	0	0	0	0	0	0 GOOD
	0	u30		2052 N6	0	0	0	0	0	0	0	0 GOOD
	0	u30		2053 N7	0	0	0	0	0	0	0	0 GOOD
	0	u30		2054 N8	0	0	0	0	0	0	0	0 GOOD
	0	u30		2055 N9	0	0	0	0	0	0	0	0 GOOD
	0	u30		2071 P11	0	0	0	0	0	0	0	0 GOOD
	0	u30		2072 P12	0	0	0	0	0	0	0	0 GOOD
	0	u30		2073 P13	0	0	0	0	0	0	0	0 GOOD
	0	u30		2074 P14	0	0	0	0	0	0	0	0 GOOD

Figura 51. Relación de pads con parámetros igual a cero.

La segunda consideración para los datos de salida es que no existe una referencia para la totalidad de componentes, como se observa en la Figura 52, el campo del componente, no contiene la referencia, sin embargo, si contiene datos sobre la impresión de pasta, que de acuerdo con el proceso, son ubicaciones en las que no se coloca componente en los procesos siguientes, sin embargo, pueden causar también algunos defectos como corto circuito o falso rechazo en las pruebas eléctricas, por lo que estos datos, si serán considerados en la generación del modelo, estos datos pueden ser tratados como *NoRef* para indicar que la ubicación No es referenciada. Mediante la revisión de algunos archivos, los componentes no referenciados son siempre los mismos, lo cual, confirma la consistencia de los datos a través de las diferentes tarjetas.

JobPath	Barcode	Index	entertime	ExitTime	CycleTime	ModelName	UserFromLog	LotInfo	MachineFromLog	Height	Barcode	BoardID	Component	PadiD	PinNumber	Area
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	115.89380000000000000000000000000000		0			147	2155.9199
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	112.44820000000000000000000000000000		0			150	2193.5046
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	117.66560000000000000000000000000000		0			153	2143.1252
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	111.30280000000000000000000000000000		0			154	2211.0974
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	116.77320000000000000000000000000000		0			329	695.9160
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	92.42560000000000000000000000000000		0			155	5002.7578
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	112.63400000000000000000000000000000		0			156	2142.3257
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	111.31310000000000000000000000000000		0			157	2169.5144
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	112.95110000000000000000000000000000		0			158	2202.3010
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	113.07890000000000000000000000000000		0			159	2155.9199
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	122.10610000000000000000000000000000		0			160	215.5120
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	102.62130000000000000000000000000000		0			161	2169.5144
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	102.47610000000000000000000000000000		0			162	2202.3010
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	107.28180000000000000000000000000000		0			163	2321.4524
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	106.13880000000000000000000000000000		0			164	2179.1106
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	103.43160000000000000000000000000000		0			165	2197.5032
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	120.26290000000000000000000000000000		0			166	211.3138
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	113.45130000000000000000000000000000		0			169	213.7128
	60964		2022-01-22 22:31:47.000	2022-01-22 22:31:56.000	9.72		Operator		MACHINE001	106.82630000000000000000000000000000		0			172	2376.6299

Figura 52. Relación de datos sin referencia a un componente.

JOB PATH	BARCODE	INDEX	DATE	S.TIME	E.TIME	CYCLE	JOB	RESULT	USER	LOTINFO	MACHINE					
			01/14/2022	23:49:49	23:49:58	9.77	GOOD		Operator		MACHINE001					
Barcode	BoardID	Component	PadiD	Pin Number	Height	Area	Area(%)	Volume	Volume(%)	OffsetX	OffsetY	Result				
	0				110.1874	754.0923	101.7	76.6332	101.7	-2.4153	35.3043	GOOD				
	0			1	120.2804	229.7062	98.9	24.8967	105.5	15.3091	26.94	GOOD				
	0			2	97.7316	579.1636	107.1	51.2718	93.3	27.9309	45.4976	GOOD				
	0			3	96.3317	581.1628	107.5	51.1195	93.1	24.8285	53.0142	GOOD				
	0			4	96.73	585.7609	108.3	51.7145	94.1	28.7225	49.7512	GOOD				
	0			5	101.6819	576.3647	106.6	53.4656	97.3	21.67	27.82	GOOD				
	0 r138			6	121.1336	243.7005	104.9	26.0877	110.6	11.4948	22.1716	GOOD				
	0 r138			7	117.7927	239.9021	103.3	25.1312	106.5	25.3586	16.9513	GOOD				
	0 c95			8	111.8195	776.0833	104.7	77.6902	103.1	28.3891	21.1481	GOOD				
	0 c95			9	109.4515	779.082	105.1	75.8946	100.7	29.6936	24.1295	GOOD				
	0			10	95.1074	567.5684	105	49.5229	90.2	13.9584	54.0955	GOOD				
	0 c94			11	104.1246	799.6736	107.8	75.7111	100.5	-6.7091	36.6512	GOOD				
	0 c94			12	101.9305	779.8817	105.2	73.5285	97.6	-12.2059	27.3095	GOOD				
	0 r139			13	112.9753	235.9037	101.6	24.2026	102.6	24.5096	23.4721	GOOD				
	0 r139			14	116.1562	224.9082	96.8	23.7832	100.8	28.6988	25.0589	GOOD				
	0			15	94.1465	577.1644	106.8	49.9461	90.9	7.547	46.3411	GOOD				
	0 c93			16	108.5393	796.6749	107.4	78.5405	104.2	-8.4456	35.9779	GOOD				
	0 c93			17	105.5875	796.8748	107.5	77.8348	103.3	-15.5263	30.8433	GOOD				
	0			18	96.7318	570.5671	105.5	50.6011	92.1	9.8432	45.314	GOOD				
	0 c92			19	105.6294	787.2786	106.2	75.7111	100.5	-23.9191	36.6304	GOOD				
	0 c92			20	104.0707	768.0865	103.6	74.085	98.3	-21.0277	29.0072	GOOD				
	0			21	97.6705	578.9637	107.1	51.8136	94.3	8.7475	54.8172	GOOD				
	0 r134			22	111.2621	218.9106	94.3	22.219	94.2	-14.0805	8.6836	GOOD				
	0 r134			23	111.4752	225.1081	96.9	23.0043	97.5	-18.77	11.2739	GOOD				
	0 c91			24	105.7536	779.8817	105.2	74.8186	99.3	-9.4856	33.7143	GOOD				
	0 c91			25	102.8977	781.2811	105.4	74.1538	98.4	-4.3089	29.9297	GOOD				
	0 r135			26	109.9886	228.5067	98.4	23.0637	97.7	3.0407	5.9758	GOOD				
	0 r135			27	109.0032	228.1069	98.2	22.7447	96.4	-8.8517	15.4018	GOOD				
	0 q10			28	106.4723	797.4745	101.7	76.9172	96.6	-6.7987	36.2939	GOOD				
	0 q10			29	103.7985	820.0652	104.6	78.0972	98.1	-8.4257	33.7997	GOOD				
	0 q10			30	110.7907	783.0804	99.9	78.2086	98.2	-16.1041	38.4888	GOOD				
	0 c90			31	105.3536	235.7038	101.5	22.1053	93.7	26.2014	24.0649	GOOD				
	0 c90			32	114.9853	243.3007	104.8	24.8313	105.2	24.0269	27.2227	GOOD				
	0 r137			33	112.3669	242.501	104.4	24.7779	105	21.8054	24.338	GOOD				

Figura 53. Log de inspección de impresión sin componentes referenciados.

Para realizar la descripción de los datos, nos enfrentamos al primer problema para un tratamiento tradicional, ya que la cantidad de datos, para los parámetros de salida, es 2548 veces mayor a los datos de entrada, ya que a pesar de la simplificación de pads, dado que todos los pines se imprimen bajo las mismas condiciones de entrada teóricas y la inspección se realiza de forma individual, la cantidad de datos es muy grande.

Como una referencia, se ha tomado poco más de dos horas de producción del día 15 de febrero a partir de las 00:00 horas, lo cual arroja 542724 registros, correspondientes a 213 tarjetas de 2458 pads considerados para cada una.

La siguiente tabla, muestra los Estadísticos descriptivos para datos de salida. más significativos para el proceso.



Tabla 5. Estadísticos descriptivos para datos de salida.

Variable	N	N*	Mean	SE Mean	St Dev	Variance	Minimum	Q1	Median	Q3	Maximum	Range
Output Cycle Time	542724	0	9.7404	0.000057	0.0421	0.00178	9.6400	9.7000	9.7300	9.7700	9.8300	0.1900
Height	542724	0	110.12	0.0151	11.10	123.19	0.0000	102.98	109.85	116.81	198.29	198.29
Area	542724	0	541.8	1.22	897.5	805574.4	0.00	81.2	231.5	736.3	12479.7	12479.7
Area Percentage	542724	0	96.034	0.0169	12.468	155.453	0.0000	88.400	98.800	104.500	152.300	152.300
Volume	542724	0	54.12	0.119	87.47	7651.84	0.000	7.53	23.84	73.35	1399.11	1399.11
Volume Percentage	542724	0	93.536	0.0221	16.310	266.007	0.0000	81.100	96.900	104.400	175.200	175.200
Offset X	542724	0	-10.641	0.0369	27.166	738.014	-139.459	-28.303	-12.184	7.914	101.388	240.847
Offset Y	542724	0	-42.607	0.0453	33.349	1112.132	-282.996	-60.800	-45.278	-27.728	175.653	458.649

El primer parámetro, tiempo de ciclo de salida o tiempo de ciclo de inspección de impresión, indica que cada tarjeta tarda en promedio 9.74 segundos en ser inspeccionada, con una varianza de 0.00178, resalta que, en dos horas consecutivas de producción, el tiempo de ciclo se mantiene en un rango pequeño de 9.64 y 9.83 segundos, es decir hay poca variación.

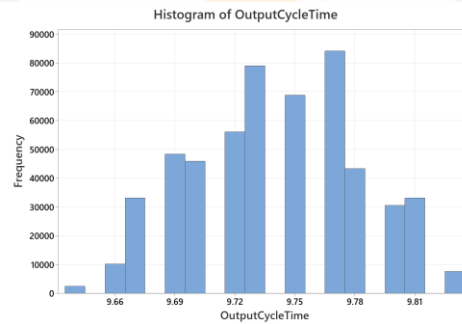


Figura 54. Histograma del tiempo de inspección de impresión.

El promedio de la altura de los componentes es de 110.12, y se pueden encontrar valores desde 0 hasta 198.29, lo cual muestra un rango bastante amplio; a pesar de existir variación denotada con varianza igual 123.19, no es tan alta como en los siguientes parámetros.

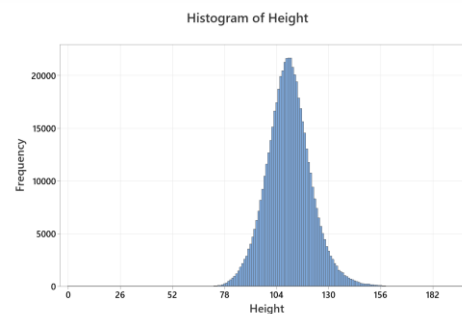


Figura 55. Histograma de la altura.

El área, por el contrario, muestra la mayor variación de todos los parámetros considerados como salidas, con una varianza de 805574.4, y su valor mínimo en 0, lo cual indica que, para algunos componentes, no ha sido impresa la más mínima cantidad de pasta.

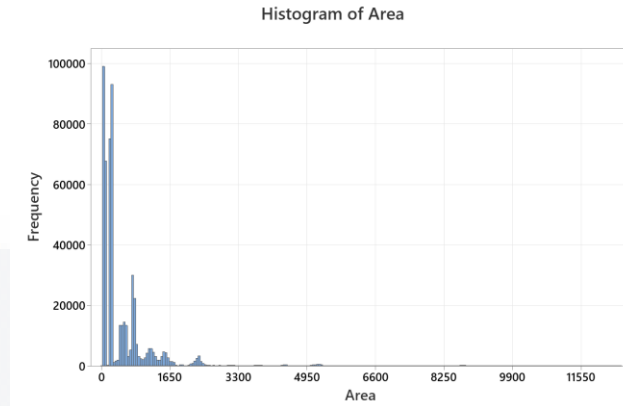


Figura 56. Histograma del Área.

En lo que respecta al porcentaje de área, es decir del porcentaje de pasta respecto al valor esperado para cada uno de los componentes, se puede apreciar que la media oscila alrededor de 96%, lo cual es un dato bastante cercano al 100% esperado.

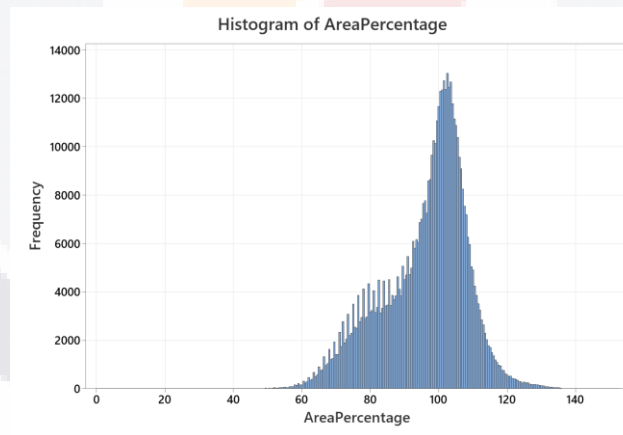


Figura 57. Histograma del porcentaje de área.

El parámetro de volumen, para las dos horas tomadas como muestra se encuentra entre 0 y 1399.11, esto se debe a que el volumen suministrado es en función del componente a colocar, los componentes más pequeños, deben contener menor volumen de pasta, lo cual, nos conduce a utilizar el porcentaje de volumen, en lugar del volumen absoluto de la tarjeta.

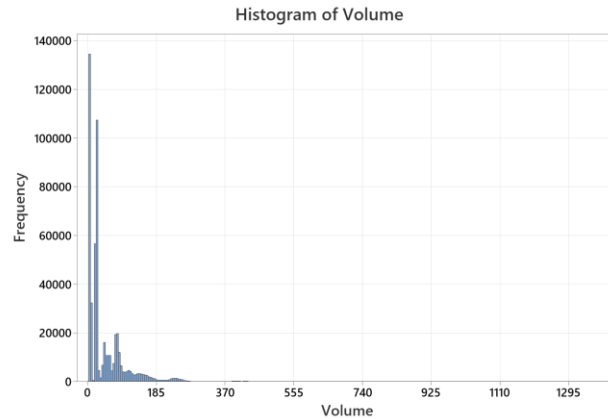


Figura 58. Histograma del volumen.

El porcentaje de volumen, que se encontró en las 213 tarjetas, no se encuentra en el caso ideal, ya que la media de los datos tiene un valor de 93.536 y existen valores desde 0 a 175.2% lo cual es demasiado variable, tomando en cuenta que las condiciones del proceso no varían en gran medida en dos horas y se está utilizando el mismo modelo como referencia.

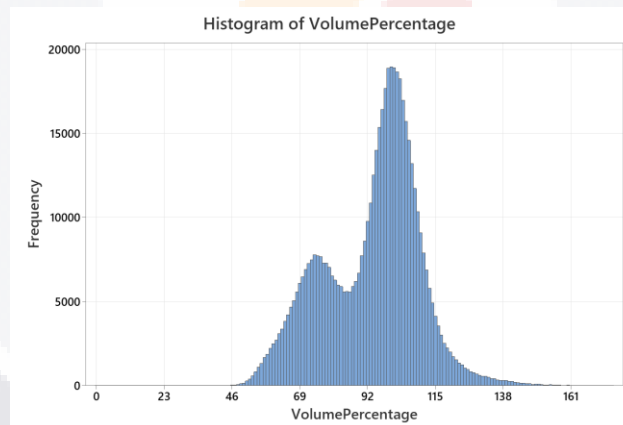


Figura 59. Histograma del porcentaje de volumen.

En el Histograma del porcentaje de volumen., se observa una distinción notable entre los valores, con varios cercanos a 70 y otros cercanos a 93, esto refleja lo que físicamente sucede en el proceso, ya que dependiendo del tipo o clasificación del componente se puede notar una falta de soldadura en mayor o menor grado, por la misma naturaleza, reflejando en el resultado de la inspección cambios significativos entre tipos de componentes.

En cuanto a los desplazamientos en las posiciones X y Y, se han encontrado valores muy distintos entre sus promedios, -10.641 y -42.607, datos que definen que existe mayor desplazamiento en Y que en X, comprobado con la varianza par ambos parámetros 738.014 y 1112.132 respectivamente.

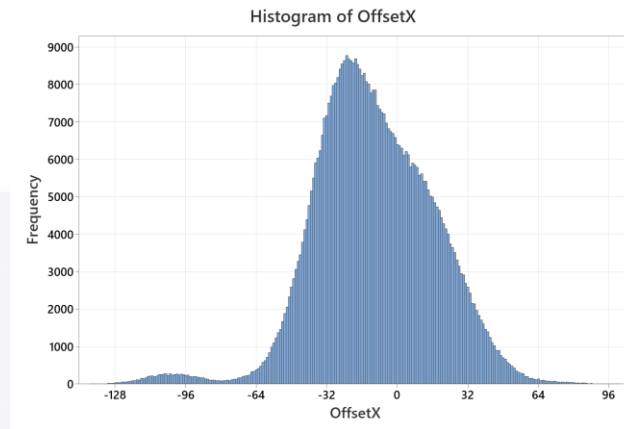


Figura 60. Histograma del desplazamiento en X.

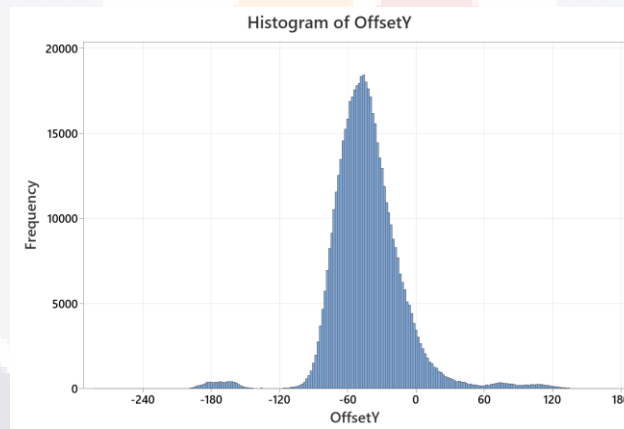


Figura 61. Histograma del desplazamiento en Y.

La tarea de generación, recolección y descripción de datos ha facilitado la identificación del proceso de tecnología de montaje superficial, permitiendo consolidar los datos en una sola fuente y realizar una limpieza previa para facilitar las siguientes etapas del análisis.

## Capítulo 5. Modelos estadísticos y de aprendizaje automático

Existen varias técnicas computacionales que nos permiten analizar datos provenientes de cualquier fuente, en las que, de acuerdo con el tipo de dato, volúmenes y objetivo del análisis facilitan la comprensión de la realidad, proporcionando información de gran importancia para alcanzar los objetivos finales de los datos.

### 5.1 Modelado de datos

De acuerdo con (Wisniewski & Rawlings, 1990), el modelado se refiere al desarrollo de expresiones matemáticas que describen el comportamiento de una variable aleatoria de interés, agregando que la complejidad matemática del modelo y el grado en que describe a la realidad, depende de qué tanto se conozca el proceso a ser estudiado y el propósito del modelado.

La variable de interés, generalmente en los modelos es denotada como  $Y$  o variable dependiente, y comúnmente el modelado tiene como objetivo describir como esta variable cambia a partir de las diversas condiciones que pueden presentar cambios.

Otras variables que probablemente proporcionan información sobre el comportamiento de la variable  $Y$ , se utilizan en el modelo como variables explicativas o predictores, y se denominan variables independientes, denotadas como  $X_1, X_2, X_3, \dots, X_n$  y se asume que son datos conocidos.

La mayoría de los modelos, involucran constantes desconocidas o parámetros que suelen controlar el comportamiento del modelo.

#### 5.1.1 Análisis de regresión

El análisis de regresión es una técnica estadística para investigar y modelar la relación entre variables y es de las técnicas más utilizadas para el análisis de datos multifactor, presentando diversas aplicaciones en varios campos, incluyendo ingeniería. (Douglas C. et al., 2012).

Los modelos de regresión permiten predecir el comportamiento de la variable dependiente a partir de otras variables.

Existen diversos tipos de regresión en las cuales se pueden destacar diferentes modelos para análisis distintos de acuerdo con el número de variables y su interacción, algunos ejemplos son modelo de regresión lineal simple, modelo de regresión lineal múltiple, modelo de regresión lineal multivariante, modelo de regresión lineal múltiple multivariante y modelo de regresión no lineal.

A continuación, se presenta de forma breve una descripción de los elementos de los modelos a utilizar.

- **Regresión lineal simple**

El modelo de regresión lineal simple se aplica cuando se busca explicar la relación existente entre la variable dependiente y una única variable independiente. Cuando existe relación entre las dos variables, es posible predecir un valor de respuesta a partir de un valor predictor con una exactitud mayor que la asociada a las probabilidades únicamente.

El modelo de regresión lineal es definido por la Ecuación 1:

$$Y = \beta_0 + \beta_1 X$$

*Ecuación 1. Ecuación de regresión lineal simple estimada.*

Dónde:

X = Predictor o variable independiente

Y = Respuesta o variable dependiente

$\beta_0$  = Intercepto

$\beta_1$  = Pendiente

El modelo de regresión proporciona la línea recta que mejor se ajusta al conjunto de datos.

El objetivo de la regresión lineal simple es minimizar la distancia vertical entre un conjunto de datos y la línea recta dada por la ecuación de regresión, es decir, la suma de las distancias al cuadrado desde cada punto a la línea debe ser lo más pequeña posible, indicado en el método de mínimos cuadrados.

Si a la ecuación estimada de regresión lineal se añade la suma de  $e$ , que es una variable aleatoria llamada error, que explica la variabilidad en Y que no puede explicar con la relación entre X y Y, se puede encontrar que el residual o error es denotado por la Ecuación 2.

$$e_i = Y_i - (\beta_0 + \beta_i X_i) \quad \text{Ecuación 2. Regresión lineal simple. Residual.}$$

La línea recta que mejor se ajusta a los datos, es la línea para la cual se minimiza la suma de los cuadrados de los residuos, es decir,  $\sum e_i^2 = 0$ .

Visualizar las gráficas residuales, antes de revisar las mediciones estadísticas de bondad de ajuste, pueden revelar patrones residuales no deseados que sugieren resultados sesgados más efectivos que los propios números, una vez aprobados los resultados en forma gráfica, se puede confiar en los resultados numéricos verificando los valores estadísticos de ajuste.

La Figura 62 muestra un ejemplo del modelo de regresión lineal simple en forma gráfica, que se conoce como línea de regresión estimada.

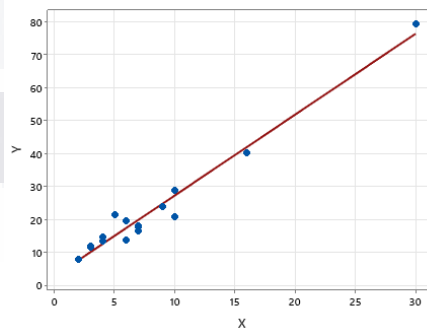


Figura 62. Gráfica del modelo de regresión lineal simple. Ejemplo.

El coeficiente de determinación o  $R^2$  es una medida de la bondad de ajuste de la recta estimada a los datos reales, que expresado en porcentaje se interpreta como el porcentaje de la

variación total de Y que se puede explicar mediante el modelo de regresión lineal, en forma simple,  $R^2$  se puede definir como se muestra en la Ecuación 3.

$$R^2 = \frac{\text{Variación explicada}}{\text{Variación total}} \quad \text{Ecuación 3. Coeficiente de determinación.}$$

Dónde:

$R^2 =$  Coeficiente de determinación

Ampliando la Ecuación 3, en términos de suma de cuadrados,  $R^2$  se puede definir como sigue.

$$SCE = \sum (Y_i - \hat{Y}_i)^2$$

$$SCT = \sum (Y_i - \bar{Y}_i)^2$$

$$SCR = \sum (\hat{Y}_i - \bar{Y}_i)^2$$

$$SCT = SCR + SCE$$

$$R^2 = \frac{SCR}{SCT}$$

$$R^2 = \frac{SCT - SCE}{SCT}$$

$$R^2 = 1 - \frac{SCE}{SCT}$$

Ecuación 4. Coeficiente de determinación en términos de suma de cuadrados.

Dónde:

SCE = Suma de cuadrados debida al error

SCT = Suma de cuadrados total

SCR = Suma de cuadrados debida a la regresión

$R^2$  siempre está situada en un rango de 0 a 100%, donde 0% indica que el modelo no explica la variabilidad de los datos de respuesta, mientras que 100% indica toda la variabilidad de los datos de respuesta, entre mayor sea  $R^2$ , mejor es el ajuste del modelo a sus datos.



Si bien,  $R^2$  es una medida práctica e intuitiva del ajuste del modelo al conjunto de datos, es necesario evaluarla junto con los gráficos del modelo y otros datos estadísticos en conjunto con conocimiento del área en la que el modelo está enfocado para proporcionar conclusiones precisas. Generalmente un  $R^2$  bajo implica un problema cuando se trata de generar predicciones en datos precisos, en los que el valor aceptable de  $R^2$  depende del ancho del intervalo de predicción y de la variabilidad de los datos, sin embargo, aún con un valor de  $R^2$  es bajo, si existen predictores significativos, es posible obtener conclusiones valiosas sobre la asociación de los cambios en los predictores con el valor de respuesta y otros coeficientes pueden ser significativos para el modelo.

El coeficiente de determinación no penaliza la inclusión de variables explicativas no significativas, presentando un aumento de  $R^2$  cuando se agregan variables explicativas que guardan poca relación con la variable independiente por lo que algunos autores, prefieren utilizar un coeficiente de determinación ajustado o  $R^2$  ajustado, que es la medida que define el porcentaje que explica la varianza de la regresión en relación con la varianza de la variable explicada, el cual contempla la penalización de variables.

El coeficiente de determinación ajustado está dado por la Ecuación 5:

$$R_a^2 = 1 - \left( \frac{n - 1}{n - p - 1} \right) * (1 - R^2)$$

*Ecuación 5. Coeficiente de determinación ajustado.*

Dónde:

$R_a^2$  = Coeficiente de determinación ajustado

$R^2$  = Coeficiente de determinación

n = Tamaño de muestra

p = Número de variables explicativas o predictores

De una forma más simple se puede ver como

$$R_a^2 = 1 - \frac{SCE (n - p - 1)}{SCT (n - 1)}$$

*Ecuación 6. Coeficiente de determinación ajustado en términos de suma de cuadrados.*

El número de variables explicativas no puede ser cero, ya que como mínimo se debe explicar una variable en función de otra, el valor mínimo de  $p$  es 1, por lo que generalmente el valor de  $R_a^2$  es menor que  $R^2$ .

El coeficiente de correlación lineal o  $R$ , es una medida descriptiva para la intensidad de asociación lineal entre la variable  $X$  y  $Y$ , estos valores se encuentran entre  $-1$  y  $1$ , si  $R$  es igual a  $-1$  significa una relación negativa perfecta, si  $R$  es igual a  $1$ , significa una relación lineal positiva perfecta, valores cercanos a  $0$ , indican que las variables  $X$  y  $Y$  no tienen relación lineal.

La Ecuación 7 muestra la relación entre el coeficiente de determinación y el coeficiente de correlación lineal:

$$R = (\text{signo de } \beta_1) \sqrt{R^2}$$

*Ecuación 7. Relación entre el coeficiente de determinación y el coeficiente de correlación lineal.*

En otros términos, el coeficiente de correlación se define como:

$$R = \frac{S_{XY}}{S_X S_Y}$$

*Ecuación 8. Coeficiente de correlación.*

Dónde:

$S_{XY}$  = Covarianza muestral

$S_X S_Y$  = Producto de las desviaciones típicas

El coeficiente de determinación es más general que el coeficiente de correlación lineal.

- **Regresión lineal múltiple**

Se utiliza cuando la relación lineal simple no es suficiente para comprender el comportamiento de la variable de interés, puesto que influye más de una variable o está correlacionada con el valor de alguna otra variable.

El modelo de regresión lineal múltiple con p variables predictoras se define por la Ecuación 9:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

*Ecuación 9. Ecuación de regresión lineal múltiple estimada.*

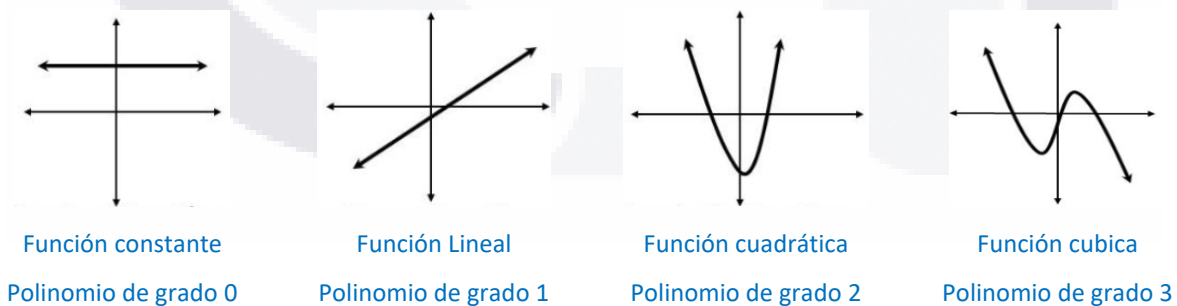
Dado que la regresión lineal múltiple es una extensión de la regresión lineal simple, el coeficiente de determinación y el coeficiente de correlación lineal siguen siendo de utilidad para describir el modelo.

Si existe una gran cantidad de predictores, antes de ajustar un modelo de regresión lineal, es necesario utilizar técnicas de selección de modelo para excluir los predictores que no estén asociados con las respuestas.

- **Regresión polinomial**

Es una forma de regresión lineal en la que la relación de las entre la variable Y las variables X, es modelada como un polinomio de grado n en X, un polinomio de grado n>1, convierte el modelo de regresión lineal en una curva, en la que los datos de X son elevados a un exponente, pero no los coeficientes, por lo que el modelo sigue siendo lineal.

La Figura 1 muestra algunos ejemplos de gráficas de funciones polinomiales.



*Figura 63. Gráficas de las funciones polinomiales.*

En el modelo cuadrático o polinómico de grado 2, la variable predictora X, es elevada al cuadrado para modelar la curvatura mediante la siguiente ecuación:

$$Y = \beta_0 + \beta_1X + \beta_2X^2$$

Ecuación 10. Ecuación de regresión polinomial de grado 2.

De igual forma, se puede modelar utilizando polinomios de grado mayor, como lo es el caso de la función cúbica, el cual tiene diversas aplicaciones en fenómenos reales.

$$Y = \beta_0 + \beta_1X + \beta_2X^2 + \beta_3X^3$$

Ecuación 11. Ecuación de regresión polinomial de grado 3.

Utilizar un modelo con polinomio de mayor grado de un predictor, mejorará el coeficiente de determinación frente a modelos de menor grado, sin embargo, esto se realiza con una pérdida de grados de libertad, por lo que para comparar si un modelo de orden superior ofrece un mejor ajuste, se pueden utilizar otros enfoques como referencia, como el coeficiente de determinación ajustado, error estándar de estimaciones o parsimonia.

### Pruebas de significancia para la regresión lineal

La ecuación de regresión lineal simple, indica que el valor esperado de Y es una función lineal de X, como se establece en la Ecuación 1. Ecuación de regresión lineal simple, por lo que, si  $\beta_1 = 0 \rightarrow Y = \beta_0$ , indicando que el valor medio de Y no depende del valor de X, demostrando que no existe una relación lineal. De forma alternativa, si  $\beta_1 \neq 0$ , por la ecuación de regresión lineal, existe una relación entre las variables.

El signo de cada coeficiente indica la dirección de la relación, los coeficientes de regresión indican los cambios medios en la variable de respuesta para una unidad de cambio en la variable X o predictora mientras se mantienen constantes los otros predictores en el modelo, para el modelo de regresión lineal simple, los coeficientes representan una pendiente.

Existen varias pruebas estadísticas para medir la significancia del modelo, entre ellas la prueba “t” de *student* o Test-T, que hace referencia a la influencia de cada una de las variables independientes en la variable dependiente.

La Ecuación 12 muestra el estadístico de la prueba t para cada uno de los coeficientes del modelo.

$$t_i = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$$

*Ecuación 12. Estadístico de la prueba t.*

Dónde:

t = Estadístico de la prueba T para el coeficiente i

$S_{\hat{\beta}_i}$  = Error estándar para el coeficiente estimado i

Debido a que la prueba T es una prueba marginal que solo determina la influencia de una variable independiente sobre la variable dependiente, aunque la prueba sea significativa, no se debe anticipar a la reducción de la variable sin revisar otros resultados.

La prueba “F” de Fisher, mide la significancia global del modelo de regresión mediante las fórmulas que convergen en la Ecuación 13.

$$CMR = \frac{SCR}{p}$$

$$CME = \frac{SCE}{n - p - 1}$$

$$F = \frac{CMR}{CME}$$

*Ecuación 13. Estadístico de la prueba F.*

Dónde:

F = Estadístico de la prueba F.

CMR = Cuadrado medio debido a la regresión.

CME = Cuadrado medio del error.

n = tamaño de muestra

p = Numero de factores

Usualmente los resultados se colocan en una tabla de análisis de varianza.

Tanto para la prueba F como para la prueba T, se busca un nivel de confianza de 95% y un valor p menor que 0.05 para que los modelos sean estadísticamente significativos.

El valor p (p-value) de cada termino comprueba la hipótesis nula de que el coeficiente  $\beta_i = 0$ , es decir, no tiene efecto en la regresión lineal, un valor p por debajo de 0.05, indica un rechazo de la hipótesis nula, mostrando una aportación significativa al modelo, debido a que los cambios del predictor están relacionados con cambios en la variable de respuesta. De la misma forma, un valor p-grande o insignificante, propone que los cambios en el predictor no están asociados con los cambios en la respuesta, es decir no existe una relación lineal.

En otros términos, el coeficiente  $\beta$  indica la intensidad y la dirección de la relación entre la variable independiente y la variable dependiente, entre más alejado se encuentre de 0, más fuerte es la relación existente.

Comúnmente, el valor p, sugiere los términos que deben mantenerse en el modelo de regresión, los valores menores a 0.05, deben permanecer en el modelo.

Para llevar a cabo el análisis estadístico de un volumen considerable de datos, existen diversas aplicaciones que facilitan esta tarea, algunos con mayor cantidad de funciones y ventajas que otros, permitiendo la optimización del tiempo y resultados al ejecutar estos y muchos otros tipos de análisis estadístico.

## 5.2 Uso de aprendizaje automático

Con el objetivo de identificar los patrones existentes ente los diferentes datos cuya característica es su volumen se ha optado por comenzar con la aplicación de técnicas de minería de datos en preparación a la implementación de aprendizaje automático.

Después de describir algunas técnicas estadísticas, dado el conjunto de datos, se busca realizar la reducción de variables, como fase de preprocesamiento al desarrollo de un modelo predictivo.

De acuerdo con (Kantardzic, 2011) existen varias técnicas de reducción de datos, que pueden ser aplicadas de acuerdo al conjunto de datos y los resultados esperados.

### 5.2.1 Normalización de los datos

Comúnmente, los datos a ser utilizados por las diversas técnicas se encuentran en formatos y valores distintos, que no siempre suponen una normalidad.

Para la aplicación de algunas técnicas, ya sean de minería de datos o de aprendizaje automático, se sugiere el uso de la Fórmula de normalización.

$$X_{norm_i} = \frac{X_i - \mu_i}{S_i}$$

*Ecuación 14. Fórmula de normalización.*

Dónde:

$X_i$  = Característica  $i$

$\mu_i$  = Media del conjunto de datos  $X$

$S_i$  = Rango de la característica  $i$  ( $\text{Max}X_i - \text{Min}X_i$ )

El resultado de la aplicación de la fórmula anterior se encuentra siempre en un rango de -1 a 1, proporcionando homogeneidad entre los valores de los conjuntos de datos.

### 5.2.2 Técnica de selección de variables: Medias y Varianzas

Después de evaluar algunas técnicas reducción, se ha seleccionado la técnica conocida como selección de características, selección de variables, reducción de características, selección de atributos o selección de subconjuntos variables, que incluye un conjunto de técnicas que seleccionan un subconjunto de características importantes para la reconstrucción de modelos de aprendizaje, mediante la eliminación de las características menos relevantes o redundantes del conjunto de datos. Esta técnica está basada en la comparación de medias y varianzas de los diferentes conjuntos de datos para cada parámetro.

De acuerdo con (Kantardzic, 2011), los objetivos de la técnica de selección de variables son: mejorar el desempeño del modelo de minería de datos; proporcionar un rápido y mejor proceso de aprendizaje; proporcionar una mejor comprensión del proceso que genera los datos.

La técnica de selección de variables para reducción de datos, parte de la clasificación de los datos para formar subconjuntos con características similares, de modo que los subconjuntos de datos estén integrados por los valores de acuerdo con la clasificación.

Tabla 6. Conjunto de datos ejemplo.

<b>X</b>	<b>Y</b>	<b>Clase</b>
<b>X<sub>1</sub></b>	<b>Y<sub>1</sub></b>	<b>A</b>
<b>X<sub>2</sub></b>	<b>Y<sub>2</sub></b>	<b>B</b>
<b>X<sub>3</sub></b>	<b>Y<sub>3</sub></b>	<b>A</b>
<b>X<sub>4</sub></b>	<b>Y<sub>4</sub></b>	<b>A</b>
<b>X<sub>5</sub></b>	<b>Y<sub>5</sub></b>	<b>B</b>
<b>X<sub>6</sub></b>	<b>Y<sub>6</sub></b>	<b>B</b>

Tomando como base el Conjunto de datos ejemplo. los subconjuntos se integran de la siguiente manera:

Tabla 7. Definición de subconjuntos de datos.

<b><math>X_A = \{X_1, X_3, X_4\}</math></b>
<b><math>X_B = \{X_2, X_5, X_6\}</math></b>
<b><math>Y_A = \{Y_1, Y_3, Y_4\}</math></b>
<b><math>Y_B = \{Y_2, Y_5, Y_6\}</math></b>

Una vez generados los subconjuntos, se realiza el cálculo de la media y la varianza para cada uno, asumiendo que son independientes para proceder a la comparación.

La Tabla 8, muestra las fórmulas tradicionales para calcular la media de una muestra y de una población, que pueden ser aplicadas de igual forma dependiendo de los datos a utilizar.



Tabla 8. Fórmulas para calcular la media de los datos.

Media de una muestra	Media de una población
$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{\mu} = \frac{\sum_{i=1}^n x_i}{n}$

Dónde:

$x_i$  = Valor de la característica i

$\bar{x}$  = Media de la muestra

$\bar{\mu}$  = Media de la población

$n$  = Tamaño de muestra

Para realizar el cálculo de la varianza, se sugiere la utilización de las fórmulas descritas en la Tabla 9, las cuales deben aplicarse con base en el conjunto de datos.

Tabla 9. Fórmulas para calcular la varianza.

Varianza de una muestra	Varianza de una población
$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$	$\sigma^2 = \frac{\sum(x_i - \bar{\mu})^2}{n}$

Dónde:

$s^2$  = Varianza de una muestra

$\sigma^2$  = Varianza de una población

$x_i$  = Valor de la característica i

$\bar{x}$  = Media de la muestra

$\bar{\mu}$  = Media de la población

$n$  = Tamaño de muestra

Cuando el valor de la media para cada una de las clases está muy separado, la técnica sugiere que la característica tiene potencial para distinguir entre dos clases, de lo contrario, la característica no es de interés para el modelo.

Definidas las medias y varianzas para cada subconjunto, se realiza la comparación para las dos clases de la misma variable, mediante el uso de la siguiente formula:

$$SE(A - B) = \sqrt{(var(A)/n_1 + var(B)/n_2)} \quad \text{Ecuación 15. Comparación de varianzas entre dos clases.}$$

Dónde A y B son los subconjuntos de los valores para cada una de las clases y  $n_1$  y  $n_2$  corresponden al número de muestras para cada una.

La ecuación debe ser aplicada a cada de las variables a ser estudiada.

Entre más grande sea el valor de la diferencia de varianzas para ambas clases de una variable, significa que tiene gran influencia dentro de la clasificación.

Calculada la diferencia de los valores obtenidos mediante la Ecuación 15 para cada subconjunto de cada característica, se culmina la prueba comparando las medias de ambos subconjuntos mediante la aplicación de la Prueba de medias y varianzas. En la cual, se define un valor umbral para realizar la evaluación referido también como *threshold value*.

$$TEST \frac{|mean(A) - mean(B)|}{SE(A - B)} < threshold \ value \quad \text{Ecuación 16. Prueba de medias y varianzas}$$

De acuerdo con esta técnica, una variable será candidata para reducción si la diferencia de medias en ambas clases es pequeña, las variables que presenten menor valor, pueden ser candidatas a reducción, sin embargo, una forma estricta de aplicar reducción o no, es realiza la prueba comparando el resultado con el valor umbral.

## Capítulo 6. Aplicación de técnicas computacionales en el proceso de Tecnología de Montaje superficial.

Descritos los datos, se pretende conocer el proceso a través de la aplicación de las técnicas computacionales puntualizadas en capítulo anterior, con la finalidad de obtener información sobre el comportamiento de los datos, para encaminar a la predicción.

### 6.1 Análisis de datos mediante el uso de técnicas estadísticas

Para realizar un análisis del proceso, conjuntando valores de entrada y salida obtenidos de forma automática, se utiliza *minitab*, para conocer el comportamiento del proceso mediante gráficas de control que nos ayuden a describirlo. Es a partir de este momento donde los datos, resultan ser importantes para modelar la realidad.

Contar con los datos en el mismo formato en la base de datos, facilita la transferencia a otros sistemas, por lo que se generan scripts para identificar primero la variación en la salida, seleccionando un periodo de tiempo en particular.

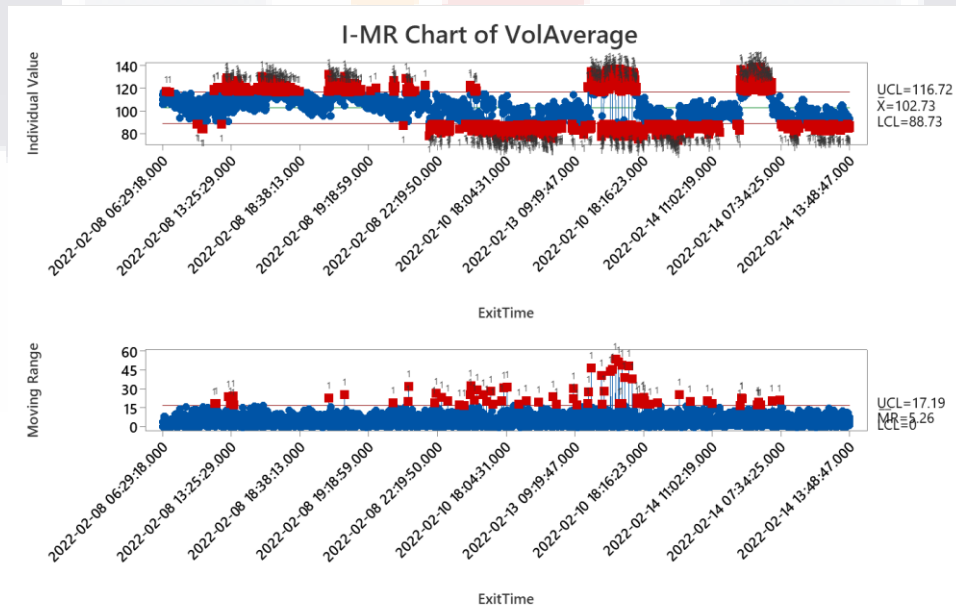
El primer criterio para analizar los datos ha sido identificar los componentes con mayor variación, debido a que existen varios componentes y como se ha descrito en el capítulo anterior, la diferencia de estos componentes si es significativa en el proceso y en los resultados, es importante tomarlos en cuenta de forma individual para observar una distribución normal, y además agruparlos por tarjeta, nos haría perder visibilidad de la particularidad de cada uno. Para conocer los componentes con más variación se calcula la desviación estándar.

La Tabla 10, muestra de mayor a menor, la desviación estándar de los componentes que presentan una mayor variación en los últimos 12 días, resaltando que los valores menores a 8.03, muestran una mínima variación entre sí.

Tabla 10. Desviación estándar por componente (valores más altos).

Referencia	Desviación Estándar
r178	14.2388
r188	13.1381
r2	11.5007
r201	11.0252
fb4	10.1470
fb1	9.8244
r7	9.5729
c2	9.5648
d1	9.1558
r197	8.2563
r195	8.2549
r199	8.0382
r196	7.9935
r179	7.9751
d8	7.9436
c67	7.8573

Para visualizar el comportamiento de las salidas, se realiza una gráfica para el promedio del volumen cada uno de los componentes o referencias con mayor variación.

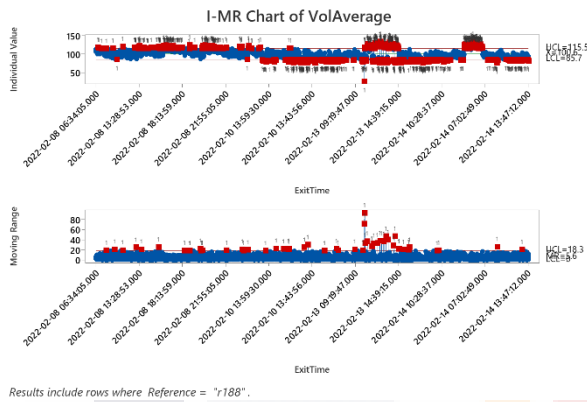


Results include rows where Reference = "r178".

Figura 64. Promedio del porcentaje de volumen de componentes r178 por cada tarjeta.

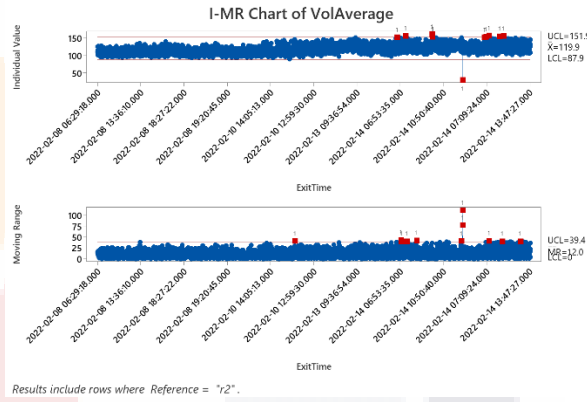
Para el componente r178 se observa una variación notable en el promedio del volumen, lo cual puede ser resultado de una variación en los parámetros de entrada, que serán analizados posteriormente, sin embargo, aunque existe variación, se observa que los valores si oscilan alrededor de 100% de volumen y son pocos los valores que salen del rango definido como bueno para el proceso 70 a 130%.

De igual forma, se realiza una gráfica de control para observar el comportamiento de los otros componentes, tomando como referencia los siguientes cuatro con mayor variación, se obtienen las siguientes gráficas.



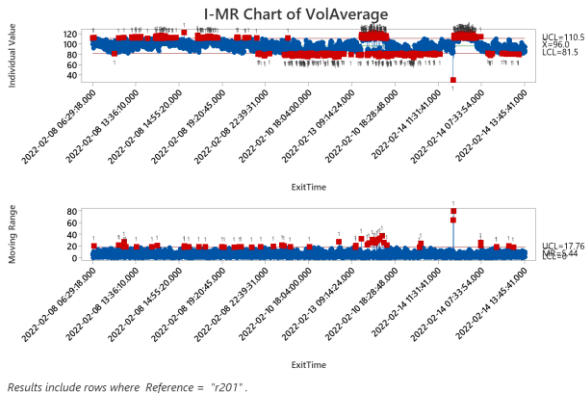
Results include rows where Reference = "r188".

Figura 65. Promedio del porcentaje de volumen de componentes r188 por cada tarjeta.



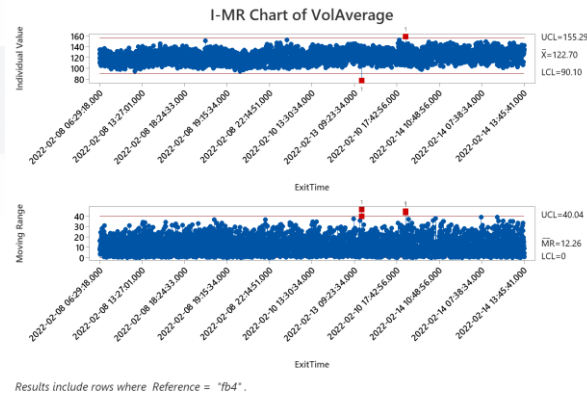
Results include rows where Reference = "r2".

Figura 66. Promedio del porcentaje de volumen de componentes r2 por cada tarjeta.



Results include rows where Reference = "r201".

Figura 67. Promedio del porcentaje de volumen de componentes r201 por cada tarjeta.



Results include rows where Reference = "fb4".

Figura 68. Promedio del porcentaje de volumen de componentes fb4 por cada tarjeta.

Cada uno de los componentes muestra diferentes valores y los rangos calculados mediante  $\pm 3\sigma$  se ve muy distinto para cada uno de los componentes, lo cual, demuestra gráficamente que hay una variación significativa en el proceso de impresión.

Mediante el conocimiento del comportamiento en la salida de forma gráfica, se pudo identificar que existe variación en el proceso, por lo que se procede a analizar las entradas, generadas por la impresora de pasta para el mismo periodo de tiempo.

Respecto a la temperatura suministrada durante la impresión, se puede observar que, si existe variación en el mismo periodo de tiempo, resaltando en la Gráfica de temperatura. un aumento en el último día de 22 a 26 grados, lo cual, sugiere una condición extraordinaria en el proceso respecto los días anteriores.

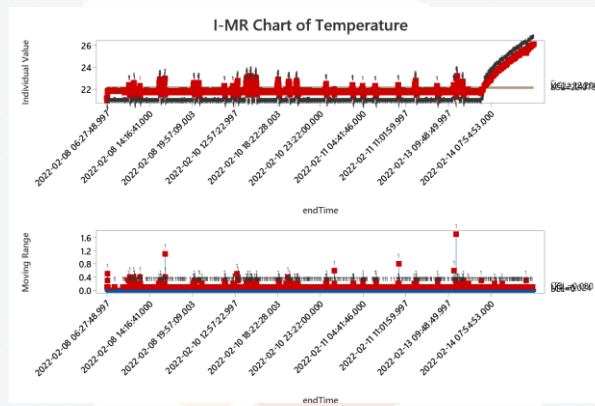


Figura 69. Gráfica de temperatura.

Dado el comportamiento que se observa de una temperatura en aumento y debido a que las condiciones de impresión son las mismas para toda la tarjeta, se extiende el periodo para el conjunto de datos de entrada, quedando la temperatura como sigue:

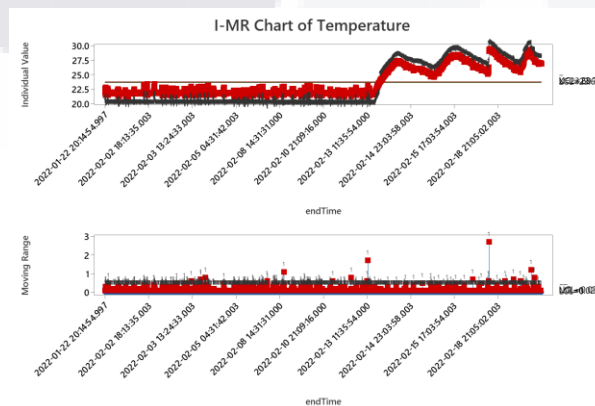
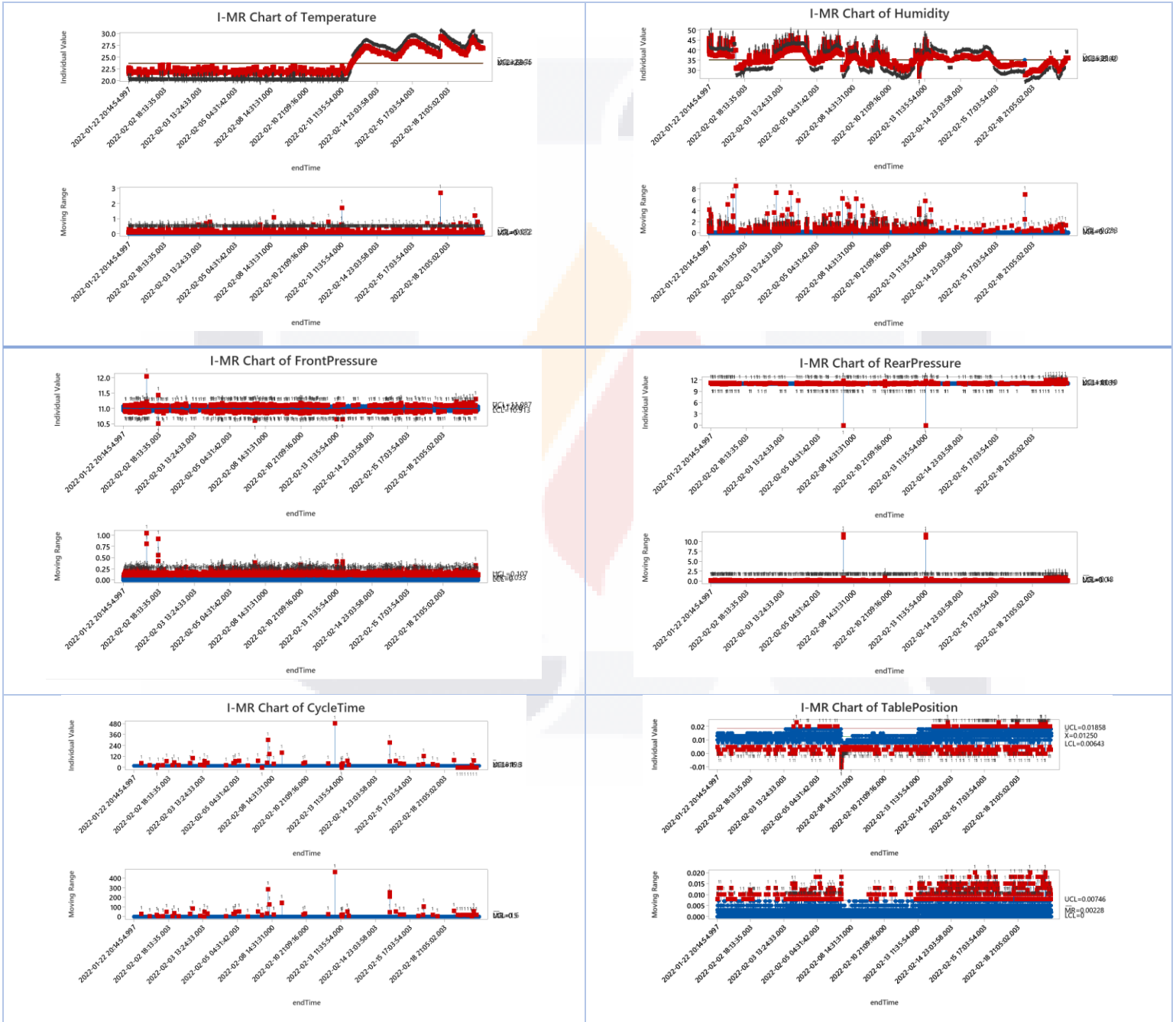


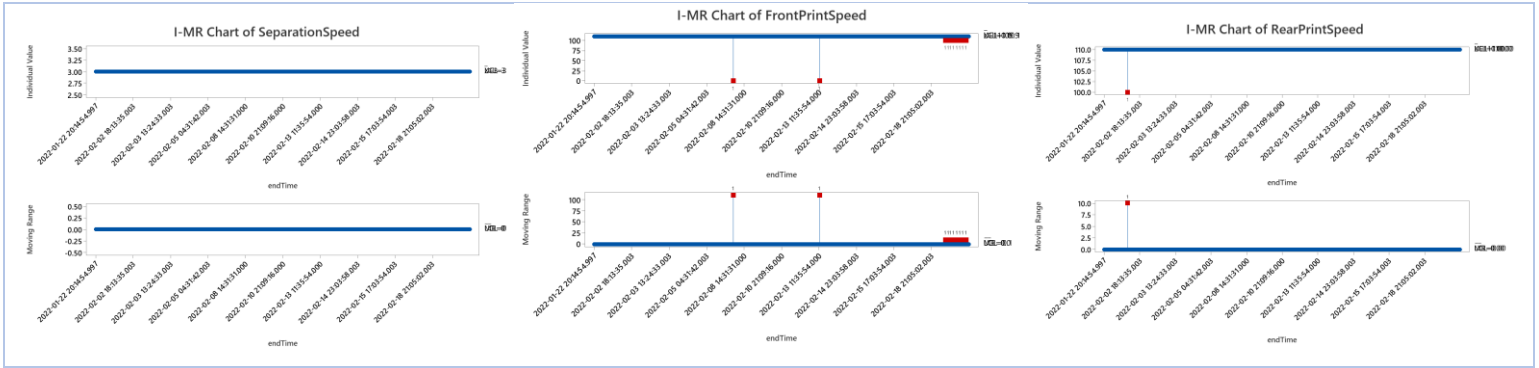
Figura 70. Gráfica de temperatura para un periodo extendido.

La Figura 70, muestra que después de que la temperatura tuvo un aumento significativo, estuvo oscilando entre valores que se pueden considerar no esperados, dado la tendencia que se tenía para el mismo proceso.

Se realiza un análisis gráfico individual para los parámetros de entrada que se puede simplificar en la Tabla 11. Gráficas Individuales para parámetros de Impresión.

Tabla 11. Gráficas Individuales para parámetros de Impresión.





En lo que respecta a la humedad, gráficamente, existe una variación notable por periodos de tiempo, el comportamiento refleja lo que físicamente sucede en el proceso, ya que, para este parámetro, no se cuenta con instalaciones que permitan controlar este factor dentro de la empresa, siendo un factor externo al proceso no controlable.

Referente a las presiones frontal y trasera, se observa una variación menor a la humedad y temperatura, y se encuentran algunos puntos dispersos que se disparan cuando existe alguna condición particular en la impresora de pasta, como alguna falla por detención o avance de tarjeta, mal funcionamiento o mala colocación del rollo, lo cual implica en la mayoría de las veces, la apertura de la máquina, eliminando la presión existente ya sea frontal o trasera. Si bien, la variación está marcada para estos parámetros, los rangos en los que se encuentran los valores no son muy amplios.

El tiempo de ciclo de impresión, también es diferente a lo largo del proceso, y presenta de la misma forma algunos valores disparados debido a que en ocasiones la tarjeta se detiene en la maquina por algún mal funcionamiento y el tiempo de ciclo sigue avanzando, a pesar de estos casos, los valores oscilan en los 18 segundos.

La posición de la mesa presenta variación frecuente durante el proceso, a pesar de ser un valor que debiera mantenerse fijo, sin embargo, el rango de sus valores de acuerdo con la gráfica de la tabla anterior es de -0.01 a 0.02.

Existen otros parámetros, que no presentan variación alguna a lo largo del proceso de impresión, como lo son las velocidades de separación y velocidades de impresión frontal y



trasera, que al igual que los parámetros anteriores muestran *outliers* conocidos también como anomalías, por las condiciones ya descritas.

Conociendo el proceso de forma separada utilizando los datos de forma automática, encontrando variación tanto en las entradas y salidas, se busca la relación entre los parámetros descritos, utilizando técnicas estadísticas.

Los parámetros que no presentaron variación alguna no se han considerado para este análisis, debido que no resultan ser relevantes para definir la correlación en el proceso por su condición constante.

Para encontrar una relación significativa entre los parámetros del proceso y la salida esperada de porcentaje de volumen se plantean las siguientes hipótesis:

$H_0$ = La relación existente entre las variables de entrada y la variable de salida es igual a 0, es decir no existe relación.

$H_a$ = La relación existente entre las variables de entrada y la variable de salida es diferente de 0, es decir, al menos una variable de entrada tiene un efecto significativo sobre la variable de salida.

Para un nivel de confianza de 95% se obtienen los siguientes resultados:

Regression Equation					
Vol Avg = 183.5 - 8.95 Temperature + 4.60 FrontPressure + 0.723 RearPressure + 1.925 Humidity - 0.0147 CycleTime - 373.1 TablePosition					
Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	183.5	43.5	4.22	0.000	
Temperature	-8.95	1.13	-7.95	0.000	1.35
FrontPressure	4.60	3.48	1.32	0.186	1.04
RearPressure	0.723	0.629	1.15	0.251	1.06
Humidity	1.925	0.109	17.60	0.000	1.71
CycleTime	-0.0147	0.0200	-0.73	0.464	1.01
TablePosition	-373.1	74.8	-4.99	0.000	1.34

Figura 71. Modelo de regresión lineal múltiple y detalle de coeficientes para porcentaje de volumen y parámetros de Impresión .

Model Summary			
S	R-sq	R-sq(adj)	R-sq(pred)
6.87547	19.21%	18.86%	17.73%

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	15563.1	2593.8	54.87	0.000
Temperature	1	2987.3	2987.3	63.19	0.000
FrontPressure	1	82.7	82.7	1.75	0.186
RearPressure	1	62.4	62.4	1.32	0.251
Humidity	1	14649.7	14649.7	309.90	0.000
CycleTime	1	25.3	25.3	0.54	0.464
TablePosition	1	1176.2	1176.2	24.88	0.000
Error	1385	65471.9	47.3		
Total	1391	81035.0			

Figura 72. Resumen del análisis de regresión lineal múltiple para porcentaje de volumen y parámetros de impresión.

De acuerdo con el valor probabilístico (p-value) para el modelo de regresión lineal múltiple, se rechaza la hipótesis nula, mostrando que si existe relación entre algunas variables de entrada y la salida. Las variables que muestran un valor menor a 0.05 son temperatura, humedad y posición de la mesa.

Sin embargo, de los resultados obtenidos se observa un coeficiente de autodeterminación ( $R^2$ ) de 19.21 para el modelo de regresión lineal, lo cual representa un porcentaje muy bajo en el ajuste de los datos en ese periodo, es decir, cerca del 19% del total de la variación en la variable de salida es explicada por el modelo de regresión:

$$Vol Avg = 183.5 - 8.95 Temperature + 4.60 FrontPressure + 0.723 RearPressure + 1.925 Humidity - 0.0147 CycleTime - 373.1 TablePosition$$

Ecuación 17. Modelo de regresión lineal múltiple para porcentaje de volumen y parámetros de impresión.

En la búsqueda de la simplificación del modelo y la mejora de  $R^2$ , se eliminan las variables no significativas del modelo anterior, buscando una correlación entre humedad, temperatura y posición de la mesa respecto al porcentaje de volumen.

Con una menor cantidad de variables, se repite el planteamiento anterior:

H<sub>0</sub>= La relación existente entre las variables de entrada temperatura, humedad y posición de la mesa y el porcentaje de volumen es igual a 0, es decir no existe relación.

H<sub>a</sub>= La relación existente entre las variables de entrada temperatura, humedad y posición de la mesa y el porcentaje de volumen es diferente de 0, es decir, al menos una variable de entrada tiene un efecto significativo sobre la variable de salida.

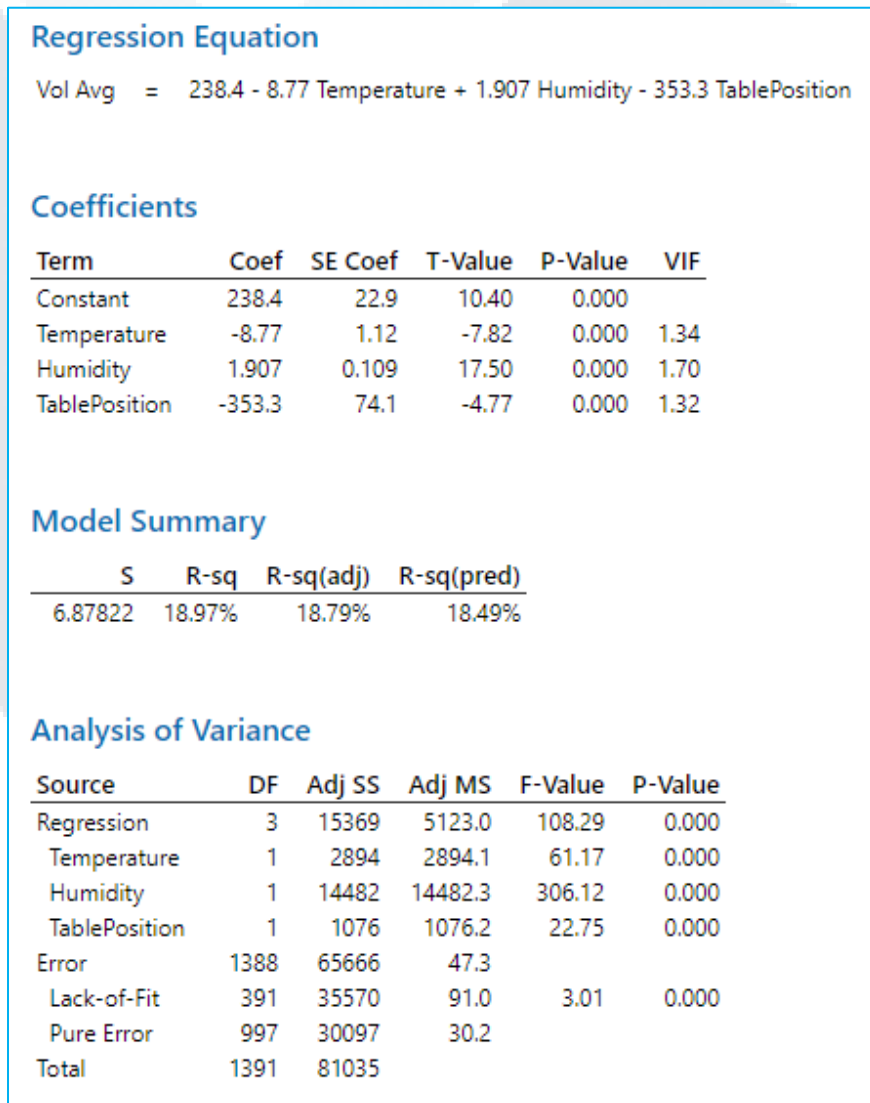


Figura 73. Análisis de regresión lineal múltiple para porcentaje de volumen y parámetros de Impresión: temperatura, humedad, posición de la mesa.

En el cual, dado que el valor probabilístico es menor a 0.05 se afirma que existe una relación entre las variables temperatura, humedad y posición de la mesa respecto al porcentaje de volumen, no obstante, esta relación, dada por el coeficiente de autodeterminación, solo explica el 18.97% de los datos, por lo que no se puede considerar un modelo valido para el conjunto de datos utilizado.

Para observar de mejor manera el comportamiento del proceso, dado que la temperatura, humedad y posición de la tabla son los valores significativos con un p-value menor a 0.05, se busca una relación de forma individual entrada-salida mediante modelos de regresión para cada uno de los tres parámetros en referencia al porcentaje de volumen.

Iniciando con la relación entre volumen y humedad, se realiza el modelo de regresión para los datos de humedad y porcentaje de volumen, obteniendo los siguientes resultados.

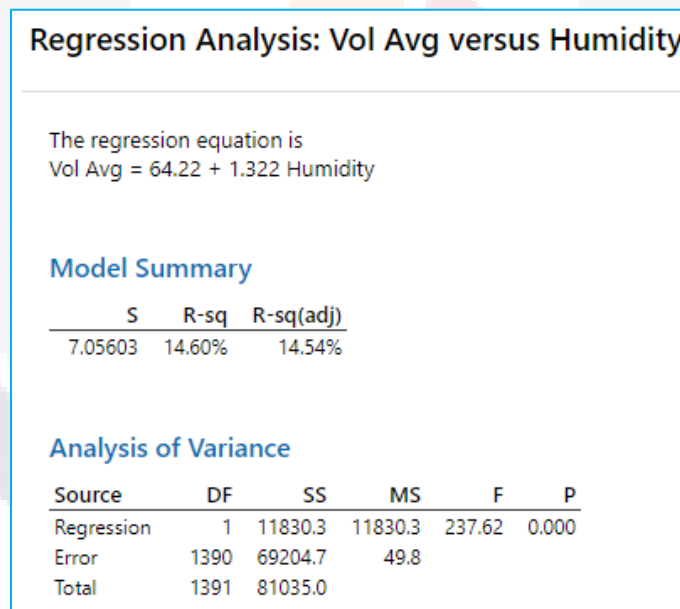


Figura 74. Análisis de regresión lineal simple para porcentaje de volumen y humedad.

A pesar de contar con un p-value menor que 0.05 en el que se sugiere una relación entre la variable dependiente e independiente, con el valor de  $R^2 = 14.60$ , el modelo representa en un mínimo porcentaje la realidad de los datos, que gráficamente se ve como sigue.

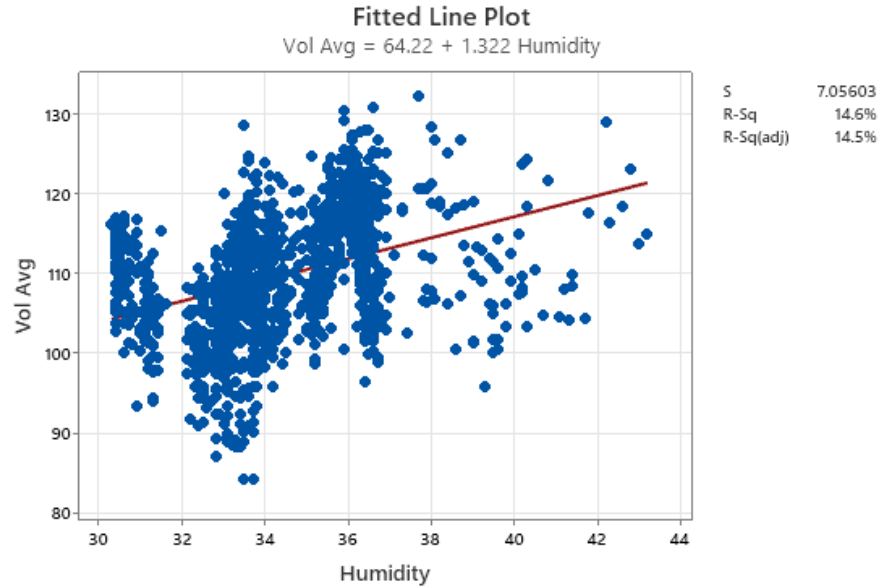


Figura 75. Gráfica de línea ajustada para el modelo de regresión lineal simple. Porcentaje de volumen y humedad.

Si bien la mayoría de los datos se concentran entre 32 y 36 para obtener un porcentaje de volumen entre 90 y 130 para el rango de datos, existen muchos otros datos que no se encuentran dentro de estos rangos y la ecuación de regresión lineal no logra ajustarse a la mayoría, por lo que se continúa en la búsqueda de una mejor definición del modelo, mediante el uso de regresión lineal cuadrática.

**Polynomial Regression Analysis: Vol Avg versus Humidity**

The regression equation is  
 $\text{Vol Avg} = 0.33 + 4.998 \text{ Humidity} - 0.05266 \text{ Humidity}^2$

**Model Summary**

S	R-sq	R-sq(adj)
7.04707	14.88%	14.75%

**Analysis of Variance**

Source	DF	SS	MS	F	P
Regression	2	12055.5	6027.76	121.38	0.000
Error	1389	68979.5	49.66		
Total	1391	81035.0			

**Sequential Analysis of Variance**

Source	DF	SS	F	P
Linear	1	11830.3	237.62	0.000
Quadratic	1	225.2	4.53	0.033

Figura 76. Análisis de regresión cuadrática para porcentaje de volumen y humedad.

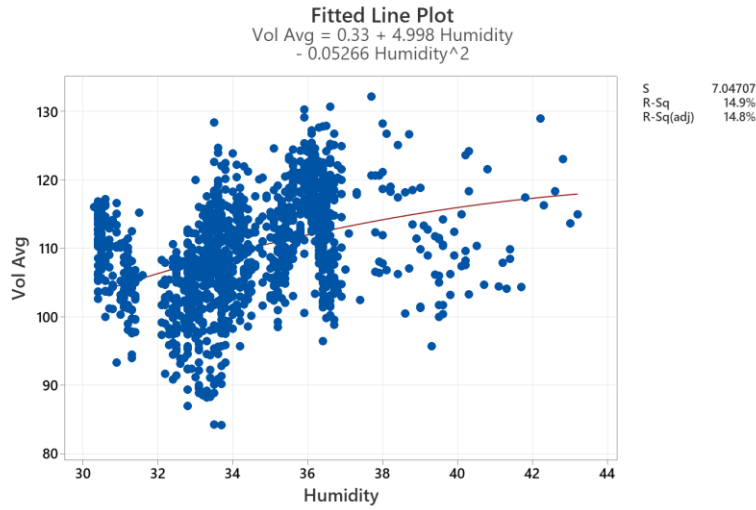


Figura 77. Gráfica de línea ajustada para el modelo de regresión cuadrática. Porcentaje de volumen y humedad.

De manera similar al modelo de regresión lineal anterior, el resultado de este modelo para la relación humedad y porcentaje de volumen, muestra un 14.88% en el valor de R<sup>2</sup>, lo cual, comparado con el modelo anterior, es ligeramente mejor, pero no suficiente para afirmar que la relación está dada por el modelo de regresión lineal cuadrático.

El siguiente modelo, muestra la regresión lineal cúbica, que describe en un 22.07% la relación de los datos y continua con un p-value menor a 0.05.

**Polynomial Regression Analysis: Vol Avg versus Humidity**

The regression equation is  
 Vol Avg = 3337 - 278.6 Humidity + 7.944 Humidity<sup>2</sup> - 0.07479 Humidity<sup>3</sup>

**Model Summary**

S	R-sq	R-sq(adj)
6.74536	22.07%	21.90%

**Analysis of Variance**

Source	DF	SS	MS	F	P
Regression	3	17881.2	5960.42	131.00	0.000
Error	1388	63153.8	45.50		
Total	1391	81035.0			

**Sequential Analysis of Variance**

Source	DF	SS	F	P
Linear	1	11830.3	237.62	0.000
Quadratic	1	225.2	4.53	0.033
Cubic	1	5825.7	128.04	0.000

Figura 78. Análisis de regresión polinomial de grado 3 para porcentaje de volumen y humedad.

En la gráfica se puede observar que al igual que los modelos anteriores, los datos se ajustan para un rango de 32 a 36 de humedad, sin embargo, los datos no ajustados, sugieren un cambio abrupto en los valores, lo cual, en la realidad muestra que, si bien la humedad afecta el proceso, no está directamente relacionado con el porcentaje de volumen obtenido después de realizar la inspección de la tarjeta.

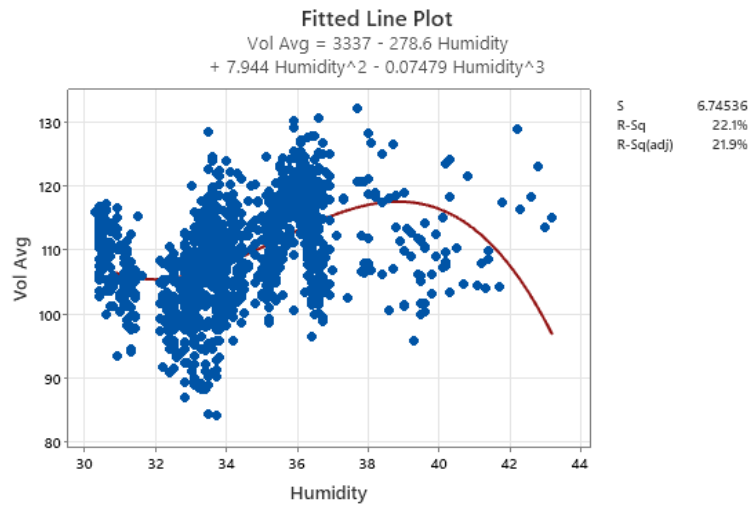


Figura 79. Gráfica de línea ajustada para el modelo de regresión polinomial de grado 3. Porcentaje de volumen y humedad.

De manera similar a los modelos de regresión para la humedad, se ha realizado el análisis para la temperatura y la posición de la mesa.

Para la temperatura, se pueden observar los diferentes resultados para modelos de regresión lineal simple (Figura 80), cuadrática (Figura 81) y cúbica (Figura 82), en los que de forma comparativa, el modelo de regresión cubica es el que mejor se ajusta a los datos presentados, con un valor  $p = 0.013$ , menor a 0.05, indicando que si hay una relación entre la temperatura y el porcentaje de volumen, seguido del modelo de regresión lineal cuadrático con un valor  $p=0.017$ , en el caso de la regresión lineal simple, para los datos dados, se concluye que no hay una relación.

La relación dada por los modelos cuadrático y cúbico para la temperatura y porcentaje de volumen solo explica el 0.58% y 0.77% de los datos, lo cual, no se considera un resultado satisfactorio para demostrar relación mediante un modelo de regresión.

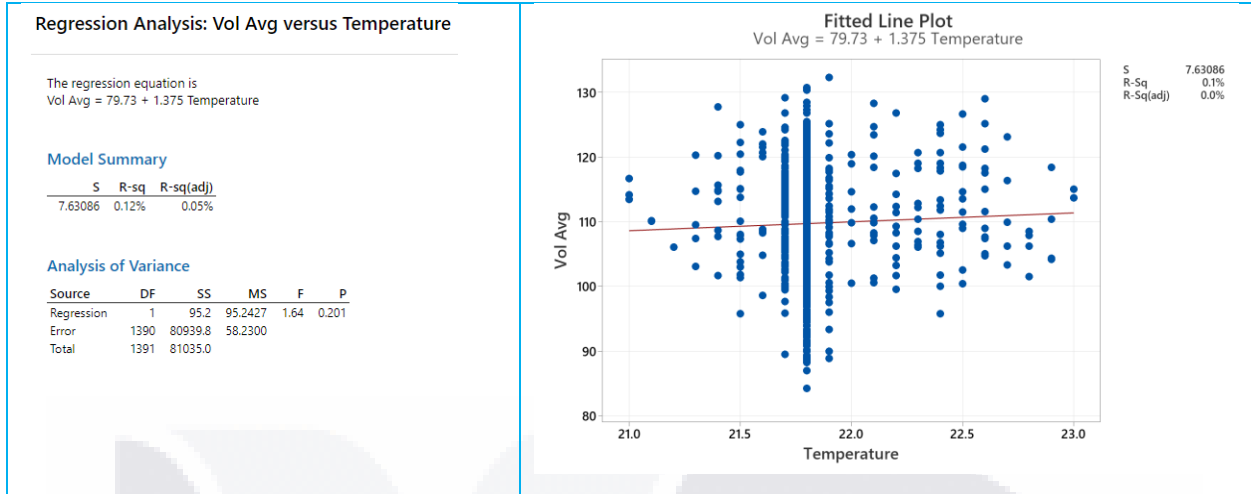


Figura 80. Modelo de regresión lineal simple. Porcentaje de volumen y temperatura.

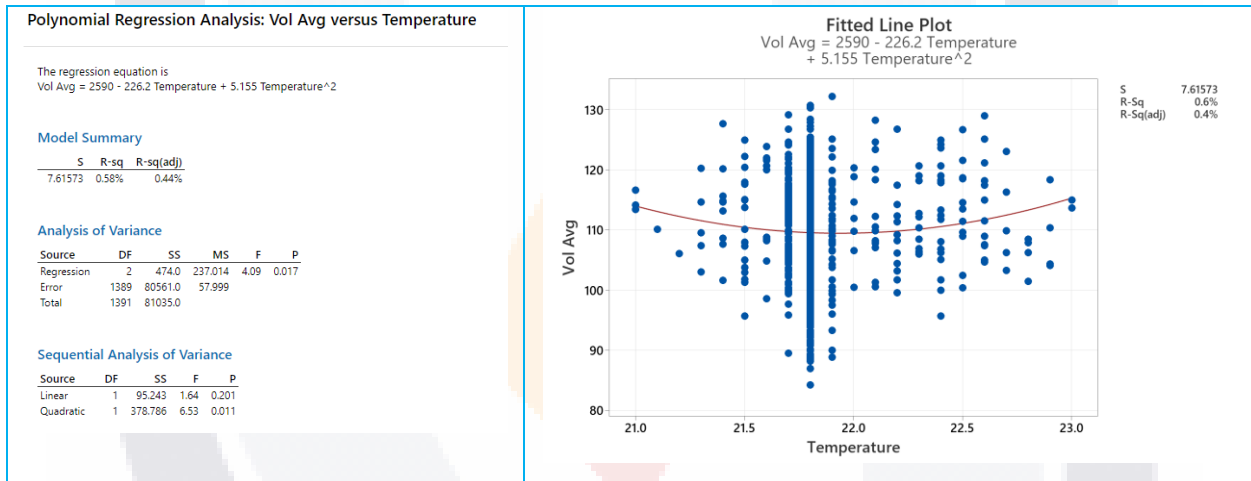


Figura 81. Modelo de regresión polinomial de grado 2. Porcentaje de Volumen y Temperatura.

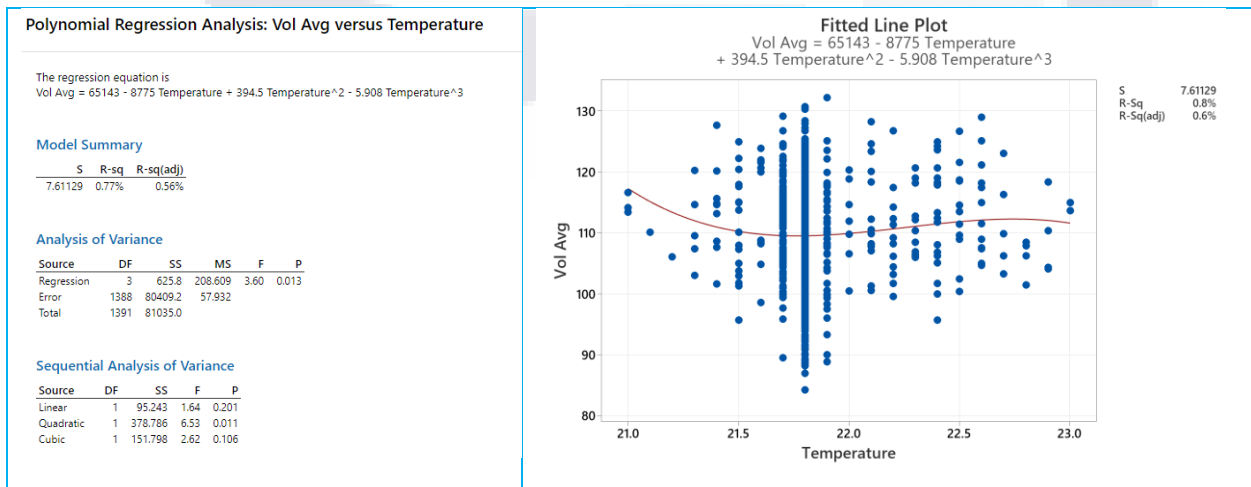


Figura 82. Modelo de regresión polinomial de grado 3. Porcentaje de volumen y temperatura.



Como tercer parámetro que presenta relación mediante el modelo inicial, se encuentra la posición de la mesa, para el cual, se realiza el mismo análisis de temperatura y humedad.

Los resultados del análisis para cada uno de los tipos de regresión arrojan coeficientes de correlación menores a 0.05, lo cual indica que, si existe una relación entre la variable dependiente e independiente, sin embargo, la  $R^2$  para cada uno de los modelos, 1.05, 4.25 y 4.28, sugieren esta correspondencia en un porcentaje mínimo para los datos, lo cual se puede confirmar en cada una de las gráficas correspondientes a los modelos.

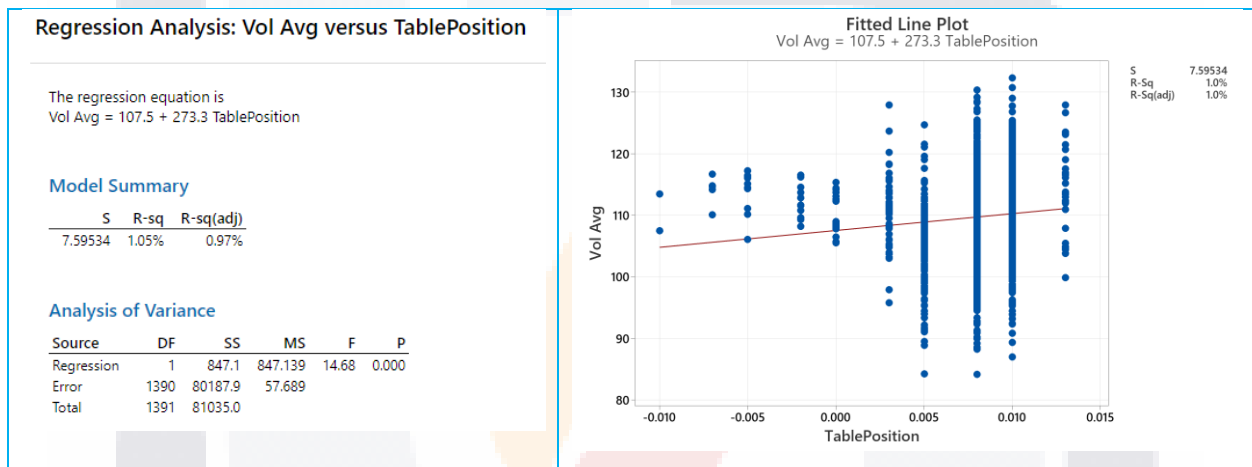


Figura 83. Modelo de regresión lineal simple. Porcentaje de volumen y posición de la mesa.

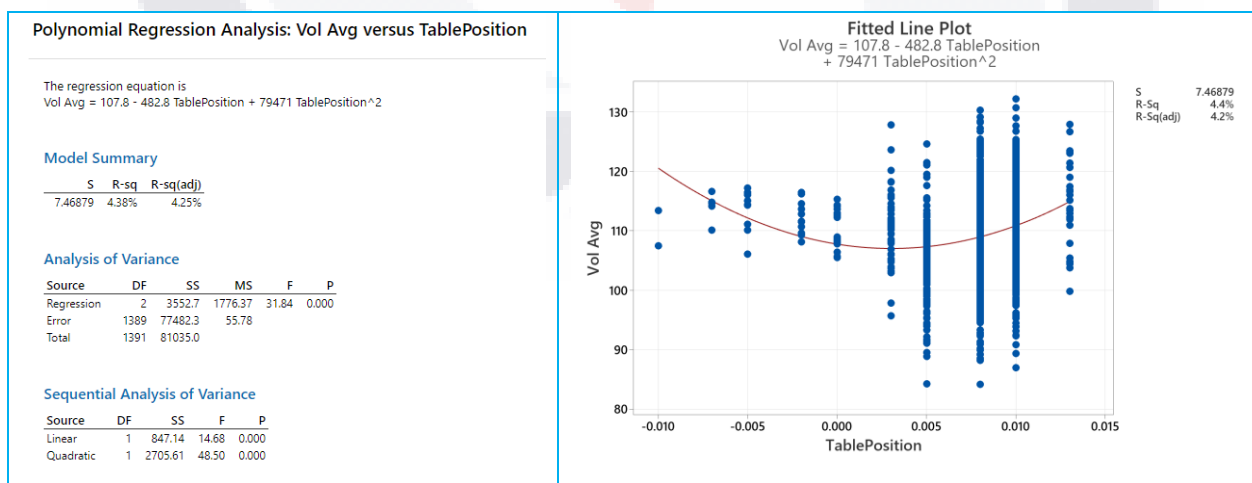


Figura 84. Modelo de regresión polinomial de grado 2. Porcentaje de volumen y posición de la mesa.

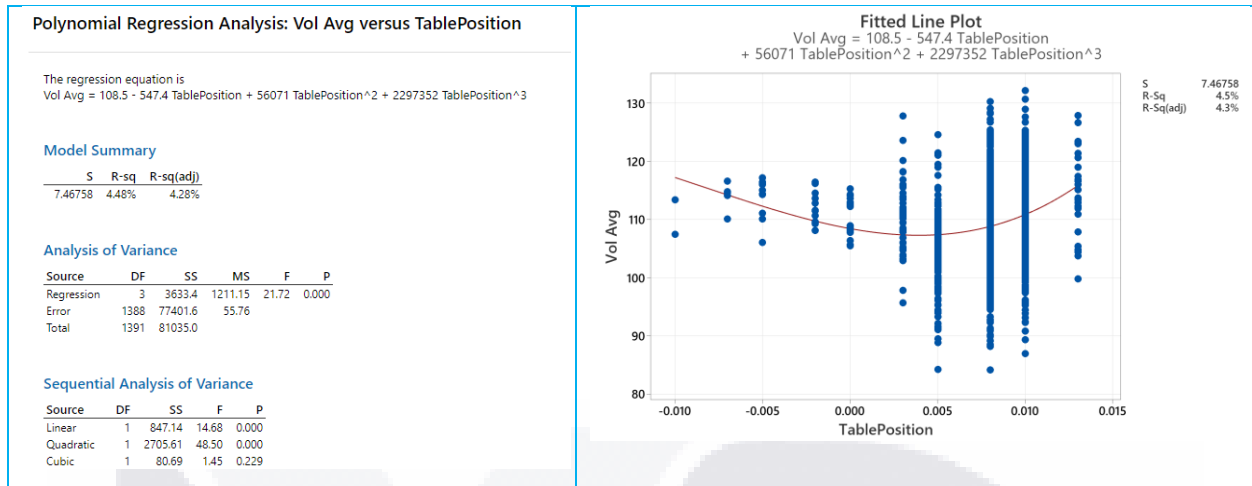


Figura 85. Modelo de regresión polinomial de grado 3. Porcentaje de volumen y posición de la mesa.

En el supuesto de que existe una relación, se ha identificado que tomar el promedio de los pines de una tarjeta del componente con mayor variación para realizar este estudio, no ha resultado del todo satisfactorio y no describe la mayor cantidad de datos, por lo que se busca otro criterio de parámetros.

Se ha repetido el mismo estudio para el pad que presenta mayor variación en lugar de componente, seguido del pad con mayor cantidad de volumen, así como el pad con menor cantidad, sin embargo, aunque por el valor probabilístico se demuestre que, si hay relación entre los datos, el porcentaje para  $R^2$ , no supera el obtenido para el análisis realizado por componente.

De igual forma, se ha buscado una dependencia entre los demás parámetros de salida, en el supuesto de que el porcentaje de volumen no fuera la mejor opción para determinar si una salida es o no correcta.

La variable dependiente ha sido cambiada entre el porcentaje de volumen, área, altura y offset en X y en Y para identificar la relación existente con las entradas sin obtener mejoras significativas en los modelos.

Los modelos de regresión lineal para una variable o varias dan como resultado un porcentaje de confiabilidad menor al 50%, bajo las distintas combinaciones, por lo que los datos utilizados para demostrar una relación no son de utilidad para la caracterización del mismo.

TESIS TESIS TESIS TESIS TESIS

El periodo seleccionado, puede ser un periodo muy corto dada la cantidad de los datos para el proceso, pero debido a las limitantes del software en cuanto a cantidad de datos se refiere, no es posible analizar periodos más extensos.

Por lo tanto, con las limitantes que conlleva, se puede concluir que, para la tarjeta seleccionada en el periodo de tiempo dado, las variables obtenidas de forma automática no son significativas con respecto a los resultados de inspección.

Continuando en la búsqueda de la caracterización de modelo, se pretende identificar si el uso y desgaste del esténcil influyen en las salidas, con la adversidad de que en proceso actual no se cuenta con un registro para ello, se propone agregar un control utilizando otra técnica computacional para adquisición de datos.

Mediante el análisis de los sistemas existentes, se encuentra factible agregar una validación en el sistema de control de producción para la línea C, en el que, se realice el registro del esténcil una vez que se completa el proceso de limpieza y se pone en producción, realizando también el registro del mismo cuando se ingresa a lavado para ser habilitado para su uso posteriormente mediante el mismo sistema, utilizando un número de identificación para cada uno de los tres diferentes esténciles con que se cuenta para la producción de la tarjeta seleccionada.

La pasta en sí es también uno de los factores que pueden afectar el proceso productivo, por lo que se busca la identificación del lote y proveedor del suministro a utilizar, sin embargo, actualmente no existe un registro del uso de ésta, por lo que al igual que en el esténcil, se debe implementar un mecanismo para la obtención de este dato y completar las variables que influyen en este proceso.

Se ha mencionado, que la organización y los recursos son una parte indispensable en la implementación de nuevas técnicas computacionales, lo cual se ha comprobado mediante la puesta en marcha de esta iniciativa, en la que, se complica el proceso productivo, con tan solo agregar esta validación.

Se ha explorado también la posibilidad de identificar otro modelo que sea importante para la compañía en el que sea más visible la correlación, para el cual, se requiere repetir la implementación de las técnicas computacionales utilizadas para esta línea y este modelo. Una siguiente alternativa es también utilizar otro de los procesos descritos para obtener resultados satisfactorios que puedan ser una aportación valiosa para la etapa de predicción, sin embargo, con la finalidad de contrastar los métodos, resultados y, sobre todo, el manejo de mayor cantidad de datos, se seguirá utilizando el proceso y variantes definidas en el uso de la siguiente técnica de *machine learning*, no sin antes implementar los avances logrados hasta el momento en procesos que puedan aportar mejoras significativas en las tareas diarias.

Como ayuda visual para el usuario de los valores significativos que se han obtenido de forma automática para el proceso, después de comprobación de la correlación de los valores obtenidos de forma automática, se pretende realizar la implementación de un sistema estadístico para el control de procesos, con la finalidad de mejorar la calidad del producto, mediante el reconocimiento temprano de las fallas obtenidas, ya que si bien, los datos no nos han sido de unidad para llegar a técnicas más avanzadas como una predicción, el usuario pueda contar con alertas visuales para mejorar el proceso, haciendo uso de las herramientas con las que cuenta la empresa.

La participación en la generación de un sistema propio de la empresa para el control estadístico de procesos ha permitido diseñar un sistema robusto en el que cada uno de los procesos pueda ser modelado en tiempo real, buscando siempre la mejora continua del proceso, contando con un sistema avanzado de almacenamiento de datos para facilitar el análisis del proceso a largo plazo, sin iniciar desde la obtención de datos o más atrás, desde la generación y transferencia de los mismos.

Los procesos seleccionados pueden ser definidos en la propia aplicación, aplicando una de las técnicas de adquisición de datos de acuerdo con el proceso, facilitando la consolidación de los datos y a la vez modelando el comportamiento actual del proceso, habilitando alertas a los usuarios para detección temprana de problemas en los procesos.

## 6.2 Aplicación de técnica de selección de variables: Medias y Varianzas

Por la naturaleza de los datos, que incluyen una gran cantidad de valores para cada una de las tarjetas electrónicas, así como variables para el proceso, se realizará la aplicación de medias y varianzas como técnica de selección de variables para reducción de datos.

De acuerdo esta técnica, se considerarán todas las variables como igual para el proceso, por lo que todos los valores del conjunto de datos que generen una medición serán considerados, agregando además un clasificador, que originalmente no está contenido en el conjunto de datos.

Para elaborar el conjunto de datos, como se ha descrito anteriormente, se ha conjuntado cada uno de los componentes de la tarjeta, reflejados en diferentes datos de salida, con el valor de entrada para la tarjeta, obteniendo para los parámetros de entrada iniciales, 2548 datos de salida distintos, tomando en cuenta la hora de impresión de la tarjeta y de inspección de impresión dado que no existe un número de serie o un identificador que proporcione con certeza la relación impresión-inspección, esta tarea, ha sido sumamente compleja, ya que en el proceso real el tiempo de ciclo de impresión no siempre es el mismo, ni el tiempo transcurrido entre la impresión y la inspección, sin embargo, todos los registros se han encontrado sincronizados.

De acuerdo con el proceso se realizará la clasificación como siguiente paso para la aplicación de la técnica, en la que se tomará en cuenta el volumen de impresión como referencia para definir las dos clases:

*Tabla 12. Definición de clases para porcentaje de volumen.*

Clase	Definición
<b>GOOD_VOLUME</b>	Elementos del conjunto de datos cuyo porcentaje de volumen es mayor o igual a 70% y menor o igual que 130%. Los elementos dentro de este rango presentan pocas fallas o nulas, en los siguientes procesos.

**UNEXPECTED\_VOLUME** Si el volumen es menor que 70% o mayor a 130%, se categorizará el elemento en esta clase, ya que, si bien presenta bajo volumen o alto en cierto pad de la tarjeta, este volumen no es esperado para cubrir el requerimiento de calidad de la tarjeta y se debe clasificar como una falla.

Suponiendo todos los parámetros como independientes y como un mismo conjunto de datos, agregando la clasificación, el conjunto de datos queda de la siguiente manera:

Tabla 13. Definición del conjunto de datos para aplicación de técnica de selección de variables.

No.	Parámetro	Tipo
F.0	Temperature	Característica
F.1	Front Pressure	Característica
F.2	Rear Pressure	Característica
F.3	Humidity	Característica
F.4	Separation Speed	Característica
F.5	Input Cycle Time	Característica
F.6	Table Position	Característica
F.7	Front Print Speed	Característica
F.8	Rear Print Speed	Característica
F.9	Output Cycle Time	Característica
F.10	Height	Característica
F.11	Area	Característica
F.12	Area Percentage	Característica
F.13	Volume	Característica
F.14	Volume Percentage	Característica
F.15	Offset X	Característica
F.16	Offset Y	Característica
	Category Vol Percentage	Clasificador

Ejemplificando la clasificación para un conjunto de datos pequeño, tubularmente se vería como sigue:

Tabla 14. Clasificación de datos.

F.0	F.1	F.2	F.3	F.4	F.5	F.6	F.7	F.8	F.9	F.10	F.11	F.12	F.13	F.14	F.15	F.16	Class
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	113.2016	494.3982	97.6	50.3244	97.8	20.5128	-70.9116	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	107.9718	522.1869	103.1	51.0964	99.3	24.1319	-46.0232	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	112.8774	499.9959	98.7	51.879	100.8	29.3841	-69.2045	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	130.626	1489.792	105.5	177.7864	124	-41.8368	-58.6553	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	116.9197	1507.385	106.8	164.3808	114.6	-26.4844	-54.7793	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	101.3226	1108.947	100.9	102.2118	91.5	-4.9279	-46.7386	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	105.0938	1112.746	99.8	103.5765	91.4	0.2187	-39.8829	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	103.2898	1115.145	100	104.6063	92.3	8.7338	-41.4106	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	103.1784	1106.348	100.7	104.9078	93.9	9.3176	-50.7181	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	123.6522	1467.401	104	167.7598	117	32.3454	-70.1217	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	106.671	786.479	106.1	76.7636	101.9	19.749	-53.8339	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	114.4192	1496.989	106.1	160.0366	111.6	-10.9195	-56.895	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	113.8808	1115.145	101.5	112.2415	100.5	-4.2893	-55.6974	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	141.1233	1235.096	110.8	153.212	135.2	19.6216	-79.7092	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	135.8958	1237.295	111	150.4467	132.8	1.9151	-73.7413	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	107.8511	1111.546	101.1	105.0769	94.1	-10.6735	-52.5421	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	118.3272	1196.312	108.8	129.532	116	10.8805	-71.0582	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	133.3553	1302.269	116.8	156.4716	138.1	9.0082	-60.0318	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	140.1997	1380.437	123.8	174.1326	153.7	-23.5016	-49.0168	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	121.8853	1157.328	105.3	128.433	115	1.2888	-54.6405	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	134.1488	1511.383	107.1	187.6435	130.8	-27.5794	-66.1059	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	110.2402	746.4953	100.7	75.2796	99.9	11.5181	-50.7533	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	121.4082	1519.38	107.6	171.3932	119.5	-6.7111	-58.4425	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	116.4397	1206.108	109.7	124.6173	111.6	19.7405	-64.9845	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	137.0243	1282.477	115	155.2044	137	-12.5139	-76.6981	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	132.6357	1417.621	127.1	163.2572	144.1	7.4631	-28.71	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	117.8815	1192.913	108.5	125.7459	112.6	-32.211	-50.1647	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	135.077	1508.984	106.9	189.7121	132.3	-27.2219	-73.6171	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	120.1945	1532.175	108.5	170.9951	119.2	-15.6001	-55.7555	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	106.4068	539.1799	119.3	51.0055	111.1	10.9567	-18.0962	GOOD_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	135.7374	1329.257	119.2	156.7608	138.4	10.919	-38.1109	UNEXPECTED_VOLUME
26.1	11.02631	10.96706	36.9	3	18.1	0.013	110	110	9.78	136.4772	1202.709	107.9	142.8536	126.1	18.6897	-28.6904	GOOD_VOLUME
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Debido a que los datos se encuentran en distintos rangos y unidades de medición, se aplicará la fórmula de normalización definida por la Ecuación 14, para estandarizar los datos, y obtener valores entre -1 y 1.

Una vez normalizados los datos, se realizan los subconjuntos para cada una de las clases en cada una de las variables por pares para cada una de las clases definidas, para cada una de las variables, como se muestra en la Tabla 15.

Tabla 15. Definición de subconjuntos de datos.

---

<b><math>F. 0_{GOOD\_VOLUME} = \{Value_1, Value_2, Value_3, Value_4, \dots Value_n\}</math></b>
<b><math>F. 0_{UNEXPECTED\_VOLUME} = \{Value_{14}, Value_{15}, Value_{18}, Value_{19}, \dots Value_m\}</math></b>
<b><math>F. 1_{GOOD\_VOLUME} = \{Value_1, Value_2, Value_3, Value_4, \dots Value_n\}</math></b>
<b><math>F. 1_{UNEXPECTED\_VOLUME} = \{Value_{14}, Value_{15}, Value_{18}, Value_{19}, \dots Value_m\}</math></b>
<b><math>F. 2_{GOOD\_VOLUME} = \{Value_1, Value_2, Value_3, Value_4, \dots Value_n\}</math></b>
<b><math>F. 2_{UNEXPECTED\_VOLUME} = \{Value_{14}, Value_{15}, Value_{18}, Value_{19}, \dots Value_m\}</math></b>
...
<b><math>F. 16_{GOOD\_VOLUME} = \{Value_1, Value_2, Value_3, Value_4, \dots Value_n\}</math></b>
<b><math>F. 16_{UNEXPECTED\_VOLUME} = \{Value_{14}, Value_{15}, Value_{18}, Value_{19}, \dots Value_m\}</math></b>

---

Identificados los subconjuntos de datos, se realiza el cálculo de la media y la varianza para cada uno haciendo uso de las fórmulas mostradas en la Tabla 8 y Tabla 9 respectivamente.

Es necesario destacar que, para esta definición, los valores obtenidos para la varianza en cada caso son distintos si se aplica la fórmula considerando el conjunto de datos como muestra o población, teniendo relación directa con la cantidad de datos a utilizar, cuando el conjunto de datos es pequeño, los valores entre varianzas de población y muestra, revelan diferencias significativas, sin embargo, en conjuntos de datos grandes, no se identifica gran diferencia por el uso de las fórmulas para calcular la varianza de los datos definidas en la Tabla 9.

Dado que no se considera el total de tarjetas producidas, siempre se estarán tomando muestras bajo distintas condiciones, se aplicará la fórmula de varianza para una muestra en cada conjunto de datos, así como la fórmula para la media de una muestra dada en la Tabla 8.



Posteriormente, se realiza la comparación para las dos clases de la misma variable, mediante el uso de la Ecuación 15. Comparación de varianzas entre dos clases., que aplicada al conjunto de datos, genera el conjunto de ecuaciones planteado en la Tabla 16:

Tabla 16. Planteamiento de ecuaciones para comparación de varianzas.

---


$$SE(F.0_{GOOD\_VOLUME} - F.0_{UNEXPECTED\_VOLUME})$$

$$= \sqrt{(var(F.0_{GOOD\_VOLUME})/n_1 + var(F.0_{UNEXPECTED\_VOLUME})/n_2)}$$
  

$$SE(F.1_{GOOD\_VOLUME} - F.1_{UNEXPECTED\_VOLUME})$$

$$= \sqrt{(var(F.1_{GOOD\_VOLUME})/n_1 + var(F.1_{UNEXPECTED\_VOLUME})/n_2)}$$
  

...

$$SE(F.16_{GOOD\_VOLUME} - F.16_{UNEXPECTED\_VOLUME})$$

$$= \sqrt{(var(F.16_{GOOD\_VOLUME})/n_1 + var(F.16_{UNEXPECTED\_VOLUME})/n_2)}$$


---

Con la diferencia de los valores obtenidos para cada subconjunto de cada característica, se culmina la prueba comparando las medias de ambos subconjuntos mediante la aplicación de la Ecuación 16. Prueba de medias y varianzas, para la cual, el valor umbral definido para llevar a cabo la evaluación en el presente análisis es 0.5.

Llevado al conjunto de datos inicial, se obtienen las ecuaciones que se muestran en la Tabla 17 en representación de cada una de las variables o características del conjunto de datos.

Tabla 17. Planteamiento de ecuaciones para prueba de medias y varianzas.

$$TEST \frac{|mean(F.0_{GOOD\_VOLUME}) - mean(F.0_{UNEXPECTED\_VOLUME})|}{SE(F.0_{GOOD\_VOLUME} - F.0_{UNEXPECTED\_VOLUME})} < 0.5$$

$$TEST \frac{|mean(F.1_{GOOD\_VOLUME}) - mean(F.1_{UNEXPECTED\_VOLUME})|}{SE(F.1_{GOOD\_VOLUME} - F.1_{UNEXPECTED\_VOLUME})} < 0.5$$

...

$$TEST \frac{|mean(F.16_{GOOD\_VOLUME}) - mean(F.16_{UNEXPECTED\_VOLUME})|}{SE(F.16_{GOOD\_VOLUME} - F.16_{UNEXPECTED\_VOLUME})} < 0.5$$

Dada la gran cantidad de datos y los recursos computacionales, representa una incomprensible labor, la ejecución de la técnica de forma manual para el conjunto de datos del proceso de SMT, por lo que se persigue la automatización de la técnica haciendo uso de las ecuaciones planteadas a través del desarrollo de un algoritmo.

### 6.2.1 Implementación del Algoritmo

Definida la técnica de minería de datos a implementar, se procede a la integración de algunas técnicas aplicadas hasta el momento, con la finalidad de generar un algoritmo que permita obtener los resultados derivados de la aplicación de medias y varianzas para cualquier conjunto de datos.

Mediante el funcionamiento del método de recolección de datos para impresión e inspección de impresión, se ha logrado obtener cada subconjunto de forma separada.

Para iniciar con la integración de los datos, se ha determinado iniciar con las primeras 50 tarjetas de un periodo de 2 horas, para revisar que todos los datos tengan relación para ambos procesos. La integración se realiza en SQL Server, para obtener el *dataset*, que dinámicamente pueda ser generado para diferente número de tarjetas, mediante diversos cálculos, aproximando

el tiempo que transcurre desde el final de la impresión hasta el inicio de la inspección, se ha logrado conjuntar los datos para distintos periodos y distinto número de tarjetas.

La construcción del algoritmo para la aplicación de la técnica de medias y varianzas se ha desarrollado mediante la utilización del lenguaje Python, conectando el programa con la base de datos de SQL server, definiendo el número de tarjetas a utilizar para el estudio.

En el mismo programa, se ha realizado la normalización de los datos, recorriendo el arreglo para todos los parámetros del conjunto de datos.

Con la construcción de un nuevo arreglo bidimensional, que contiene los datos normalizados, se ha efectuado la separación por clases de cada una de las características, obteniendo arreglos separados para cada clase que en el programa se denominarán XA y XB.

La parte medular del algoritmo realiza el cálculo de medias y varianzas de cada subconjunto por cada uno de los parámetros, indicando un *threshold* de 0.5 de acuerdo con las ecuaciones presentadas. Asimismo, se realiza la comparación de varianzas mediante la aplicación de la ecuación correspondiente y finalmente se realiza la prueba si el resultado obtenido por parámetro es menor que el *threshold*, evaluando si la característica es candidata o no a reducción.

Por el tamaño de cada conjunto de datos se identificado la necesidad de generar los resultados de forma gráfica, para resumir el resultado de la aplicación del algoritmo e identificar las características fácilmente.

Se ha realizado la programación modular, mediante el uso de funciones para separar la programación en cada bloque del algoritmo.

Para probar la programación descrita, se ha utilizado, un conjunto de datos pequeño, realizando manualmente los cálculos para comprobar los resultados matemáticos a nivel de programación.

Posteriormente, se ha realizado la prueba con diferentes conjuntos de datos que permitan modelar la realidad con la técnica seleccionada.

Utilizando un subconjunto de los datos correspondientes a la producción del 15 de febrero, se ha aplicado el algoritmo a diferente número de tarjetas para observar su variación.

Seleccionando 50 tarjetas con los 2548 componentes cada una, se ha obtenido una muestra de 127,400 registros para los parámetros definidos.

La Figura 86 muestra gráficamente el valor obtenido para la diferencia de varianzas para cada uno de los parámetros definidos para el proceso, en la cual se observan valores muy pequeños para cada variable, obteniendo como variable de mayor influencia para el proceso, el tiempo de ciclo de impresión, seguido de la temperatura y tiempo de ciclo de impresión.

Como ha sido esperado, en general, a excepción del tiempo de inspección de impresión, las variables que presentan mayor influencia en el proceso, son en su mayoría las variables de entrada del proceso, mostrando valores que significan mayor influencia excluyendo aquellos valores que no presentan variación, los cuales, al tener un valor constante, no representan cambio alguno en el porcentaje del volumen del proceso como lo es la velocidad de separación, y velocidades de impresión frontal y trasera.

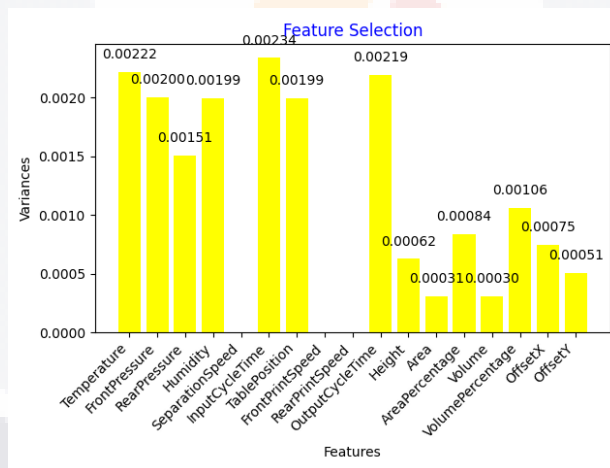


Figura 86. Varianzas para 50 tarjetas.

Los resultados para la diferencia de medias y varianzas para el mismo conjunto de datos, se puede visualizar en la Figura 87, la cual, no muestra ningún valor por debajo de umbral definido como 0.5, que, en un sentido estricto, no sugiere ninguna variable candidata para reducción. Si se realiza la selección basados en los valores menores, podría descartarse la presión frontal y trasera, así como el tiempo de impresión.

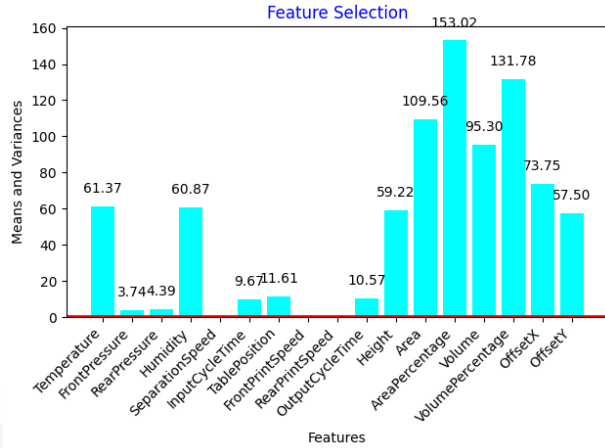


Figura 87. Medias y varianzas para 50 tarjetas.

A manera de experimento, se ha aplicado el mismo algoritmo tomando una clasificación distinta para las variables, realizando una evaluación con un rango de 65 a 135% de volumen, obteniendo los siguientes resultados.

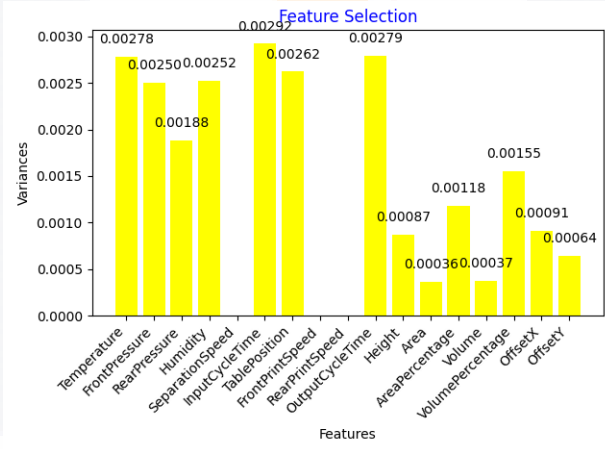


Figura 88. Varianzas para 50 tarjetas con clasificación de 65 a 135% de volumen.

La Figura 88, muestra las mismas características que la clasificación anterior con mayor influencia, que son tiempo de ciclo de impresión, temperatura y tiempo de ciclo de inspección de impresión, lo cual, para este conjunto de datos no representa un cambio significativo. Sin embargo, para el caso de la prueba, los valores muestran mayor diferencia respecto de la clasificación anterior manteniendo presiones frontal y trasera con valor menor, caso contrario del tiempo de ciclo de impresión ya que, su valor se ve por encima de 9.67 de la primera clasificación, permitiendo al tiempo de ciclo de inspección de impresión, formar parte de los tres

parámetros “posibles” candidatos a reducción mediante la obtención de los parámetros con menor valor.

Esto confirma, que la clasificación de las variables es de gran importancia en la aplicación del algoritmo para la técnica de medias y varianzas.

Y como resultado de la comparación de medias y varianzas para el mismo rango de 65 a 135% se obtiene la siguiente figura, corroborando que la clasificación tiene impacto directo en el algoritmo, en los valores obtenidos de la varianza y la comparación de medias y varianzas, por lo que es de gran importancia, definir inicialmente un criterio de clasificación adecuado para el conjunto de variables.

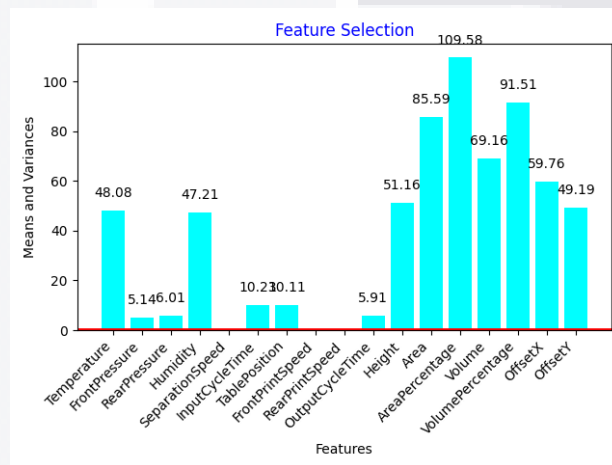


Figura 89. Medias y varianzas para 50 tarjetas con clasificación de 65 a 135% de volumen.

Las variables o parámetros con mayor influencia, para los distintos rangos se pueden observar en la Tabla 18, presentando presión frontal y posición de la mesa como distintas variables que influyen de acuerdo a los distintos criterios de clasificación.

Tabla 18. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 50 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Tiempo de ciclo de impresión	Tiempo de ciclo de inspección
Temperatura	Tiempo de ciclo de impresión
Tiempo de ciclo de inspección	Temperatura
Presión Frontal	Posición de la mesa
Humedad, Posición de la mesa	Humedad

Después de realizar la evaluación de medias y varianzas, se obtienen los cinco parámetros candidatos a reducción ordenados desde el menor valor hasta el mayor. Encontrando que coinciden los primeros cinco, sin embargo, si se tratara de reducir dos, tres o cuatro valores, las variables a reducir serían distintas en cada caso. Si se tratara de reducir solamente una variable, esta coincide para esta muestra como presión frontal, aunque es necesario destacar que el valor en ambos casos está por encima del *threshold* o valor umbral.

Tabla 19. Parámetros candidatos a reducción para una muestra de 50 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Presión frontal	Presión frontal
Presión trasera	Tiempo de ciclo de inspección
Tiempo de ciclo de impresión	Presión trasera
Tiempo de ciclo de inspección	Posición de la mesa
Posición de la mesa	Tiempo de ciclo de impresión

Dada la gran cantidad de datos, utilizar 50 tarjetas, no representa una muestra muy significativa, ya que, esto solo representa menos de media hora de producción de unidades en condiciones normales, por lo que se duplica la cantidad de datos para observar el comportamiento del algoritmo.

Se realiza el mismo experimento, duplicando la cantidad de datos para observar las diferencias, 100 tarjetas, es lo equivalente a poco menos de 1 hora de producción del modelo seleccionado, pero implica 2548 datos para cada variable por cada una de las tarjetas.

Para la primera clasificación, el volumen esperado está dentro del rango [70, 130], en el que se puede observar que el parámetro con mayor influencia es el tiempo de ciclo de inspección, seguido por la posición de la mesa y la temperatura.

Respecto a la muestra anterior, el tiempo de impresión, no parece ser tan significativo para el conjunto de datos como lo es el tiempo de inspección, sin embargo, los parámetros tiempo de impresión y de inspección siguen siendo de los primeros con mayor influencia, caso contrario de la temperatura, donde, para el conjunto de datos de 100, deja de ser de los primeros tres parámetros con mayor influencia, agregando la posición de la mesa en las primeras tres posiciones.

Extendiendo la selección a cinco parámetros, coinciden con los cinco de la muestra de 50 tarjetas, aunque en diferente posición, se sigue encontrando dentro de los primeros cinco, el tiempo de impresión, tiempo de inspección, posición de la mesa, temperatura y humedad.

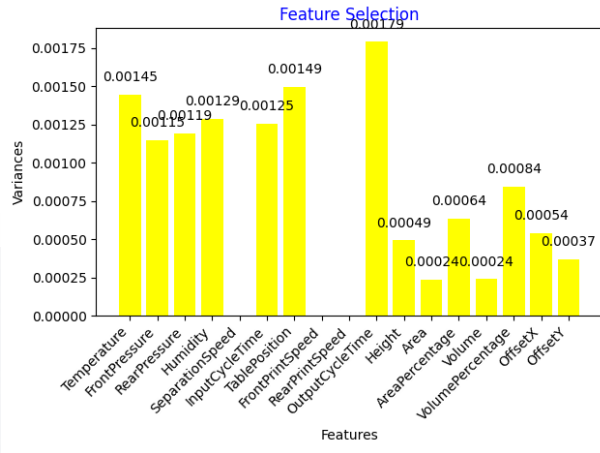


Figura 90. Varianzas para 100 tarjetas.

Lo mismo sucede con las medias y varianzas, como se puede observar en la Figura 91, tres parámetros candidatos a reducción son el tiempo de ciclo de inspección, presión frontal y tiempo de ciclo de impresión, extendiendo la selección a cinco parámetros, se observa que la posición de la mesa y la presión trasera son también candidatos a reducción. Comparando los cinco parámetros, se sigue manteniendo con el conjunto de datos anterior, aunque en diferentes posiciones, sin embargo, comparado con el *threshold* establecido, los valores siguen siendo superiores por lo que se continúa agregando tarjetas al conjunto en la búsqueda de algún valor que esté por debajo de 0.5.

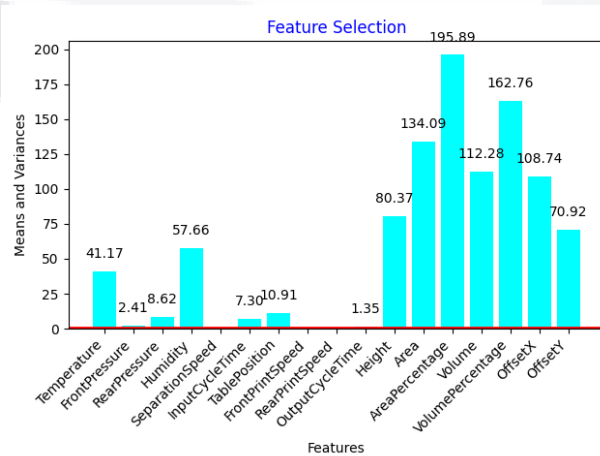


Figura 91. Medias y varianzas para 100 tarjetas.



Repitiendo la ejecución del algoritmo para el conjunto de datos con clasificación basada en un rango 5% más abierto, [65,135], se obtuvieron las siguientes gráficas.

En la Figura 92 se puede observar que los valores obtenidos para las varianzas son distintos para cada uno de los parámetros, al igual que sucede con las Medias y varianzas para 100 tarjetas con clasificación de 65 a 135% de volumen., por lo que se reafirma que la clasificación es importante para el algoritmo, sin embargo, a pesar de los valores distintos obtenidos, al seleccionar los cinco parámetros con mayor influencia y los cinco candidatos a reducción, se observa que son los mismos respecto a la clasificación anterior.

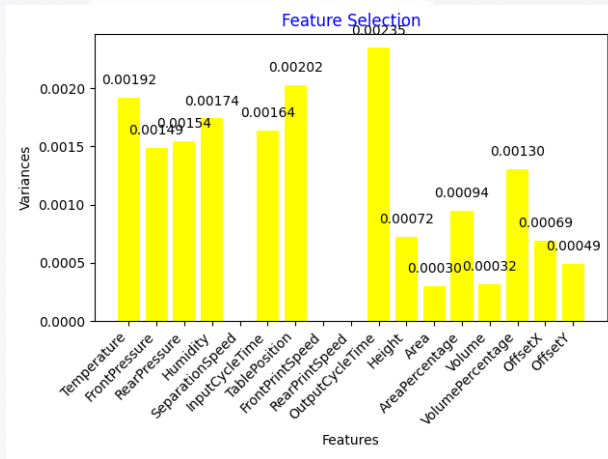


Figura 92. Varianzas para 100 tarjetas con clasificación de 65 a 135% de volumen.

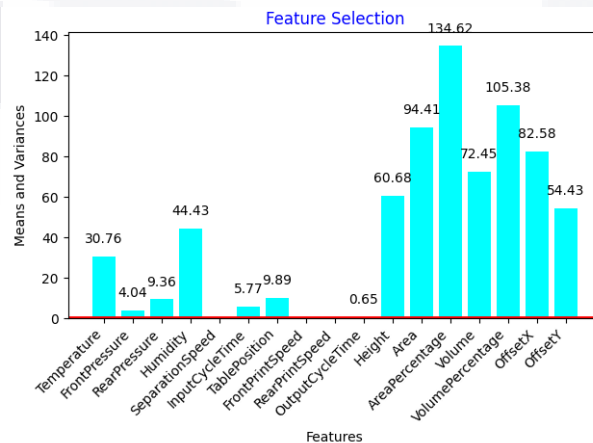


Figura 93. Medias y varianzas para 100 tarjetas con clasificación de 65 a 135% de volumen.

La siguiente tabla Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 100 tarjetas. muestra de forma sintetizada los parámetros con mayor influencia para cada clasificación.

Tabla 20. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 100 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Tiempo de ciclo de impresión	Tiempo de ciclo de inspección
Tiempo de ciclo de inspección	Posición de la mesa
Posición de la mesa	Temperatura
Temperatura	Humedad
Humedad	Tiempo de ciclo de impresión

La Tabla 21. Parámetros candidatos a reducción para una muestra de 100 tarjetas. muestra una coincidencia entre los parámetros candidatos a reducción en ambos rangos, que si bien, los valores obtenidos no son iguales, el orden de los parámetros a reducir indica que el tiempo de ciclo de inspección sería el primero, seguido de la presión frontal, tiempo de ciclo de impresión, presión trasera y posición de la mesa. Encontrando el valor más bajo en 1.35 y 0.65 para los rangos de +30 y +35% del valor esperado respectivamente, valores no tan distantes, pero mayores al valor umbral.

Tabla 21. Parámetros candidatos a reducción para una muestra de 100 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Tiempo de ciclo de inspección	Tiempo de ciclo de inspección
Presión frontal	Presión frontal
Tiempo de ciclo de impresión	Tiempo de ciclo de impresión
Presión trasera	Presión trasera
Posición de la mesa	Posición de la mesa

Duplicando la cantidad de muestras, para un primer rango, se ha obtenido las siguientes gráficas.

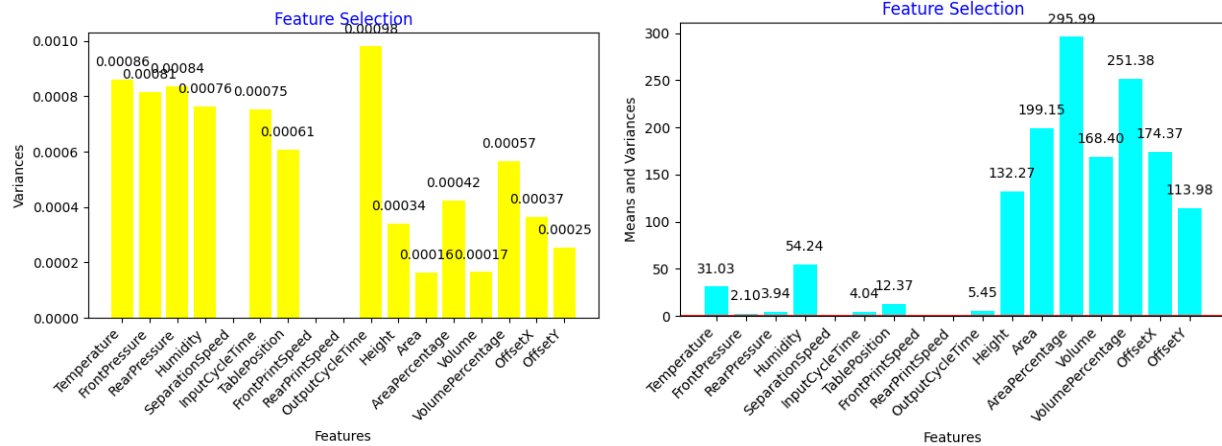


Figura 94. Varianzas y medias y varianzas para 200 tarjetas.

Utilizando la segunda clasificación con el rango [65,135], se presentan las siguientes gráficas como resultado de la ejecución del mismo algoritmo.

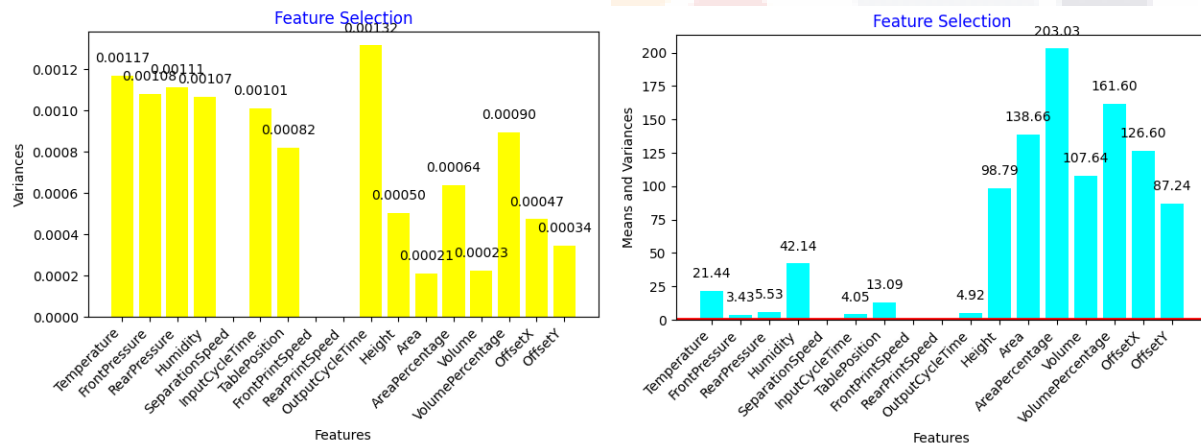


Figura 95. Varianzas y medias y varianzas para 200 tarjetas con clasificación de 65 a 135% de volumen.

En resumen, para una muestra de 200 tarjetas, se han obtenido los siguientes parámetros con mayor influencia para el proceso ordenados de mayor a menor, coincidiendo en los cinco registros, aunque no en mismos valores.

Tabla 22. Parámetros con mayor influencia de acuerdo con las varianzas obtenidas para una muestra de 200 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Tiempo de ciclo de inspección	Tiempo de ciclo de inspección
Temperatura	Temperatura
Presión trasera	Presión trasera
Presión frontal	Presión frontal
Humedad	Humedad

De igual forma que en las muestras de 50 y 100 tarjetas, la ejecución del algoritmo para 200, da como resultado la reducción de los mismos cinco parámetros, aunque permanece la constante que no el mismo orden, como se puede observar en la siguiente tabla.

Tabla 23. Parámetros candidatos a reducción para una muestra de 200 tarjetas.

Clasificación de acuerdo al rango [70,130]	Clasificación de acuerdo al rango [65,135]
Presión frontal	Presión frontal
Presión trasera	Tiempo de ciclo de impresión
Tiempo de ciclo de impresión	Tiempo de ciclo de inspección
Tiempo de ciclo de inspección	Presión trasera
Posición de la mesa	Posición de la mesa

Si bien, se ha demostrado que la clasificación es importante en el algoritmo, una variación de 10% en los rangos, no sugiere cambios significativos en el orden de las variables para definir si son o no de gran influencia para el proceso, así como si son candidatas a reducción, habiendo definido el rango inicial para el porcentaje de volumen de 70 a 130%, se continua la ejecución del algoritmo dentro de este mismo rango.

Ejecutando el algoritmo para mayor cantidad de datos, buscando un incremento a 1000 tarjetas, que representan 2,548,000 de registros, es necesario hacer una pausa para mencionar, que, a pesar de no ser una gran cantidad de datos, los recursos computacionales con los que se cuenta no lograron la ejecución exitosa, buscando un incremento en el intérprete para un procesador de 64 bits, esperando la ejecución completa de la técnica de machine learning seleccionada.

Volviendo a la ejecución del algoritmo para 1000 y 2000 tarjetas, se ha logrado culminar con el ejercicio de incrementar el número de muestras para la obtención de los parámetros.

La Figura 96 y Figura 97 presentan las gráficas que se han obtenido como resultado de la ejecución del algoritmo para los 2,548,000 y 5,096,000 registros respectivamente.

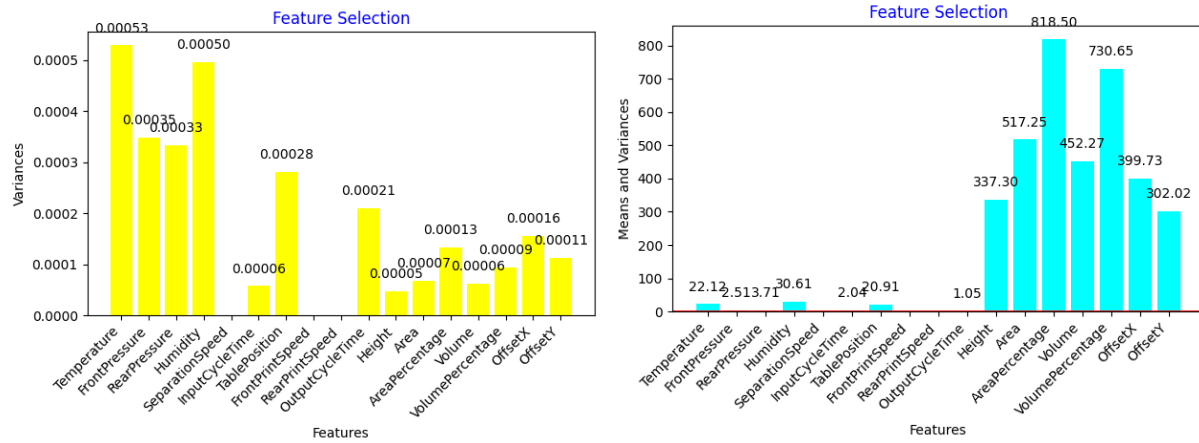


Figura 96. Varianzas y medias y varianzas para 1000 tarjetas.

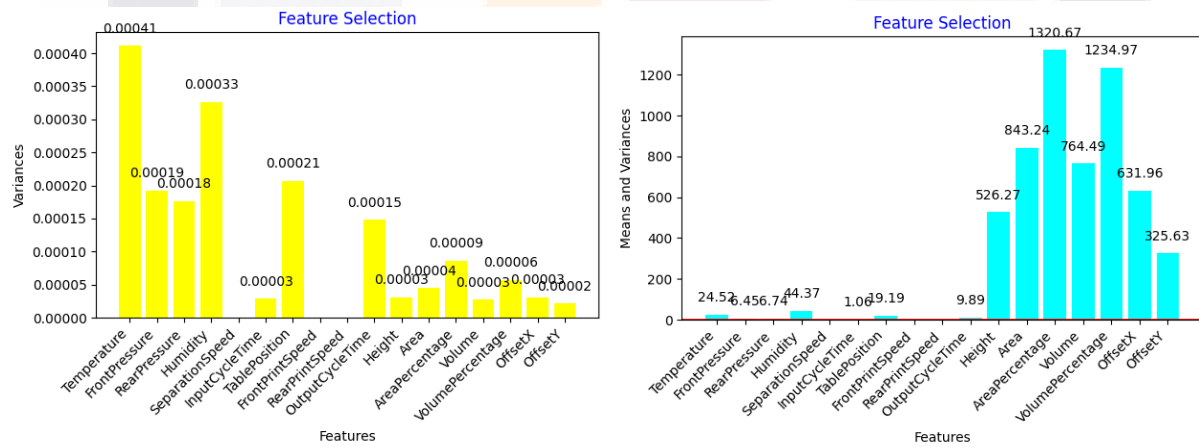


Figura 97. Varianzas y medias y varianzas para 2000 tarjetas.

Se puede continuar con el ejercicio para mayor cantidad de datos, cuando las capacidades del equipo de cómputo lo permitan, sin embargo, la repetición del ejercicio ha logrado coincidir en las variables candidatas a reducción, como se muestra en la Tabla 24. Resultados obtenidos

para varianzas y evaluación de medias y varianzas para cada muestra. Donde, aunque se pueden observar valores distintos, existe repetibilidad en las variables candidatas a reducción.

Dado que la velocidad de separación, velocidad de separación frontal y velocidad de separación trasera no presentan variación, los campos de la tabla aparecen repetidamente vacíos para diferentes tamaño de muestra, las casillas marcadas en amarillo, resaltan los parámetros con varianzas más altas para el tamaño de muestra indicado, mientras que los campos marcados en color azul, muestran los valores más bajos para el resultado de la prueba de medias y varianzas entre clases en comparación con los demás parámetros.

Tabla 24. Resultados obtenidos para varianzas y evaluación de medias y varianzas para cada muestra.

Tamaño de muestra	50		100		200		1000		2000	
parámetro	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas
Temperature	0.00222	61.37000	0.00145	41.17000	0.00086	31.03000	0.00053	22.12000	0.00041	24.52000
Front Pressure	0.00200	3.74000	0.00115	2.41000	0.00081	2.10000	0.00035	2.51000	0.00019	6.45000
Rear Pressure	0.00151	4.39000	0.00119	8.62000	0.00084	3.94000	0.00033	3.71000	0.00018	6.74000
Humidity	0.00199	60.87000	0.00123	57.66000	0.00076	54.24000	0.00050	30.61000	0.00033	44.37000
Separation Speed										
Input Cycle Time	0.00234	9.67000	1.00125	7.30000	0.00075	4.04000	0.00060	2.04000	0.00003	1.06000
Table Position	0.00199	11.61000	0.00149	10.91000	0.00061	12.37000	0.00028	20.91000	0.00021	19.19000
Front Print Speed										
Rear Print Speed										
Output Cycle Time	0.00219	10.57000	0.00179	1.35000	0.00098	5.45000	0.00021	1.05000	0.00015	9.89000
Height	0.00062	59.22000	0.00049	8037.00000	0.00034	132.27000	0.00005	337.30000	0.00003	526.27000
Area	0.00031	109.56000	0.00024	134.09000	0.00016	199.15000	0.00007	517.25000	0.00004	843.24000
Area Percentage	0.00084	153.05000	0.00064	195.89000	0.00042	295.99000	0.00013	818.50000	0.00009	1320.64000
Volume	0.00030	95.30000	0.00024	112.28000	0.00017	168.40000	0.00006	452.27000	0.00003	764.49000
Volume Percentage	0.00106	131.78000	0.00084	162.76000	0.00057	251.38000	0.00009	730.65000	0.00006	1234.97000
Offset X	0.00075	73.75000	0.00054	1008.74000	0.00037	174.37000	0.00016	399.73000	0.00003	631.96000
Offset Y	0.00051	57.50000	0.00037	70.92000	0.00025	113.98000	0.00011	302.02000	0.00002	325.63000

La Tabla 25, muestra las variables candidatas a reducción resultantes de la ejecución de la técnica de machine learning para diferente tamaño de muestra, coincidiendo en las cinco variables, aunque no siendo constante el orden, dada la diferencia de valores.

Para los diferentes conjuntos de datos, se sabe también que los parámetros referentes a las velocidades de separación y de impresión frontal y trasera, no existe variación alguna, por lo que, para la aplicación de futuras técnicas, no es necesario considerarlos, ya no representan influencia alguna en el resultado.

Tabla 25. Variables candidatas a reducción.

Variables candidatas a reducción
Tiempo de ciclo de inspección
Presión frontal
Presión trasera
Tiempo de ciclo de impresión
Posición de la mesa

Las variables que presentan mayor influencia, por su parte, son distintas para cada conjunto de datos, sin embargo, se observa una repetibilidad en la temperatura y la humedad, lo cual coincide con los primeros estudios estadísticos realizados, en los que ambas variables presentan una correlación con el porcentaje de volumen.

Tabla 26. Variables con mayor influencia en el proceso.

Variables con mayor influencia
Temperatura
Humedad

Los resultados obtenidos del *dataset* de producción consecutiva para diferente tamaño de muestra, aunque han mostrado resultados concretos, se observa que, en ningún caso, el valor umbral ha sido menor que el establecido, por lo que para la verificación de los resultados obtenidos y buscando un resultado que cumpla la prueba del valor umbral, se repite el ejercicio para los mismos tamaños de muestra, utilizando la misma clasificación en un rango [70,130]%.

El conjunto de datos se ha construido utilizando diferentes periodos en los que se ha producido la misma pieza, en los que se sabe existen diferentes condiciones para el proceso, tomando una muestra mínima de 500 tarjetas en cada periodo, para construir los *datasets* combinados cercano al tamaño de muestra propuestos inicialmente, para considerar el mismo número de tarjetas para cada subconjunto de datos, 52, 96, 200, 1000 y 2000 tarjetas.

Los periodos utilizados se muestran en la Tabla 27, generando un total de 2187 tarjetas impresas e inspeccionadas.

Tabla 27. Periodos utilizados para la generación de dataset de producción no consecutiva.

Hora de salida de impresión de primera tarjeta	Hora de término de inspección de última tarjeta	Total de datos	Número de tarjetas
2/14/2022 23:59:38.003	2/15/2022 2:01:15.000	1274000	500
2/2/2022 15:00:29.997	2/2/2022 21:01:00.000	1472744	578
1/22/2022 21:58:12.997	1/23/2022 3:01:30.000	1335152	524
3/3/2022 10:00:00.003	3/3/2022 17:01:16.000	1490580	585

Aplicando el algoritmo al número de tarjetas establecidas, se obtienen los siguientes resultados de forma gráfica.

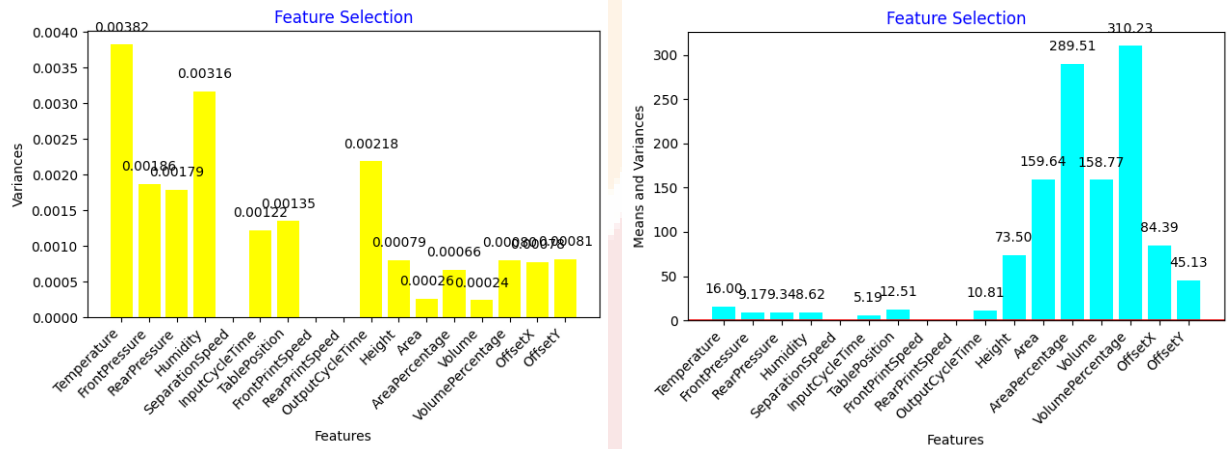


Figura 98. Gráficas de varianzas y medias y varianzas para 48 tarjetas electrónicas no consecutivas.

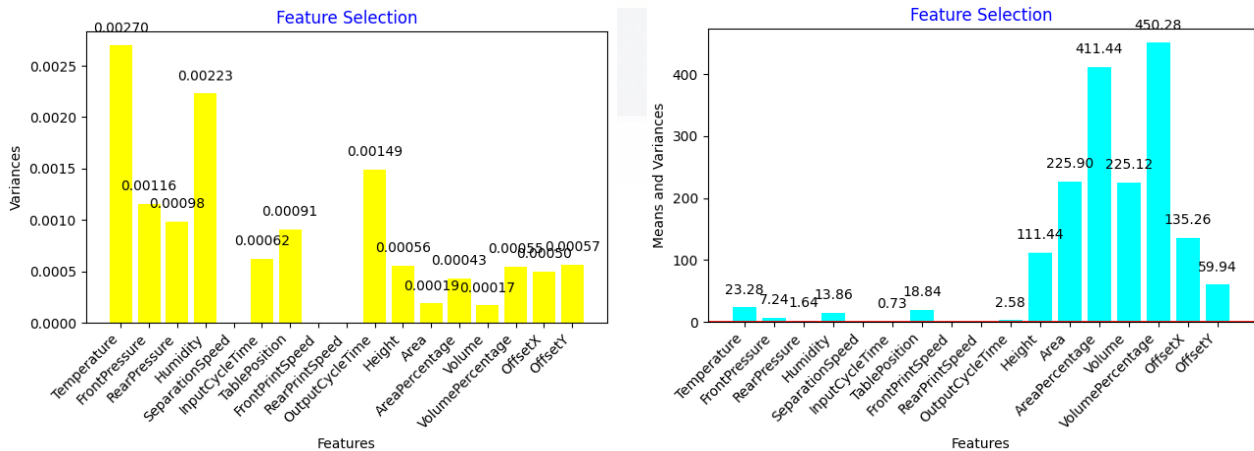


Figura 99. Gráficas de varianzas y medias y varianzas para 96 tarjetas electrónicas no consecutivas.



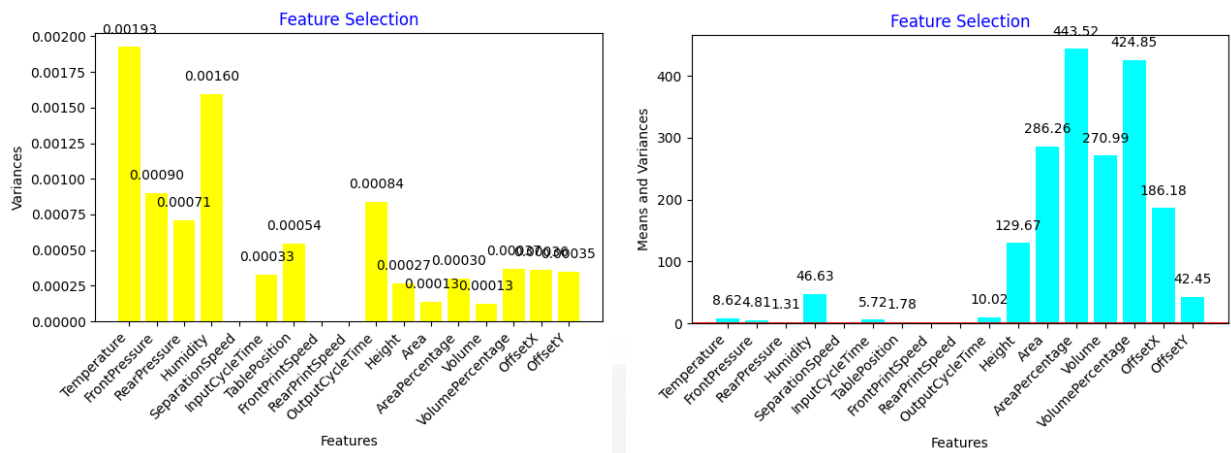


Figura 100. Gráficas de varianzas y medias y varianzas para 200 tarjetas electrónicas no consecutivas.

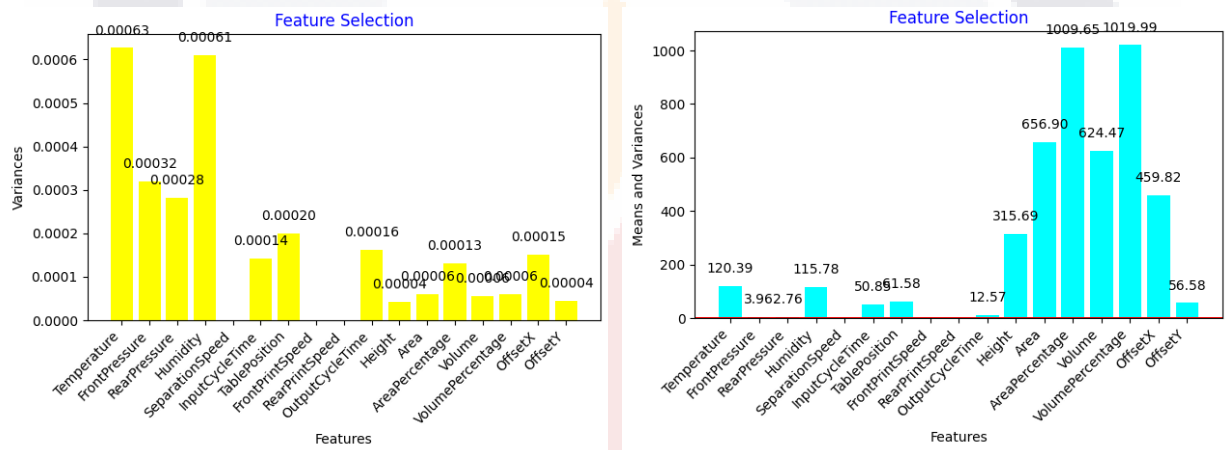


Figura 101. Gráficas de varianzas y medias y varianzas para 1000 tarjetas electrónicas no consecutivas.

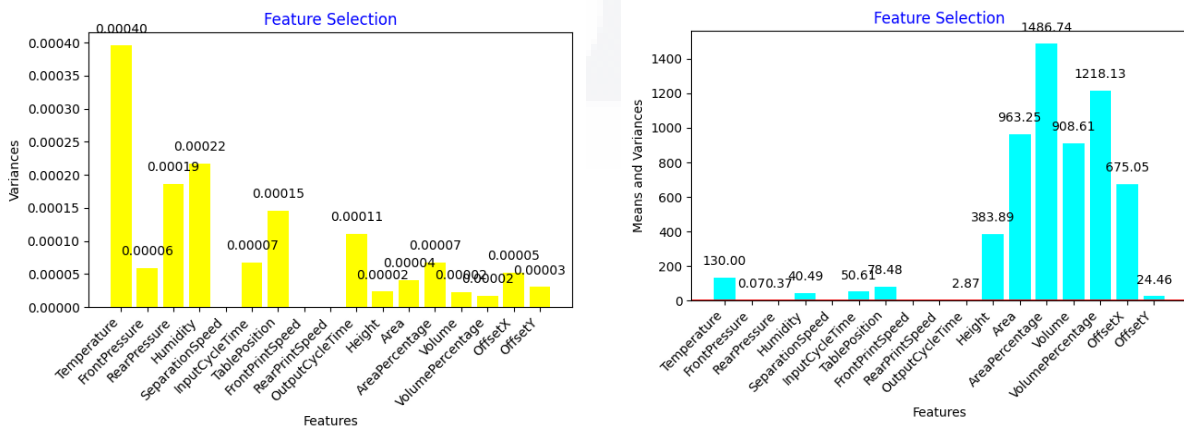


Figura 102. Gráficas de varianzas y medias y varianzas para 2000 tarjetas electrónicas no consecutivas.

Tomando como base la Tabla 24. Resultados obtenidos para varianzas y evaluación de medias y varianzas para cada muestra., se simplifican los resultados obtenidos para las muestras no consecutivas en la Tabla 28, utilizando los mismos criterios para apreciación de resultados.

Tabla 28. Resultados obtenidos para varianzas y evaluación de medias y varianzas para muestras de diferente periodo.

Tamaño de muestra	50		100		200		1000		2000	
parámetro	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas	Varianzas	Medias y Varianzas
Temperature	0.00382	16.00000	0.00270	23.28000	0.00193	8.62000	0.00063	120.39000	0.00040	130.00000
Front Pressure	0.00186	9.17000	0.00116	7.24000	0.00090	4.81000	0.00032	3.96000	0.00006	0.07000
Rear Pressure	0.00179	9.34000	0.00098	1.64000	0.00071	1.31000	0.00028	2.76000	0.00019	0.37000
Humidity	0.00316	8.62000	0.00223	13.86000	0.00160	46.63000	0.00061	115.78000	0.00022	40.49000
Separation Speed										
Input Cycle Time	0.00122	5.19000	0.00062	0.73000	0.00033	5.72000	0.00014	50.85000	0.00007	50.61000
Table Position	0.00135	12.51000	0.00091	18.84000	0.00054	1.78000	0.00020	61.58000	0.00015	78.48000
Front Print Speed										
Rear Print Speed										
Output Cycle Time	0.00218	10.81000	0.00149	2.58000	0.00084	10.02000	0.00016	12.57000	0.00011	2.87000
Height	0.00079	73.50000	0.00056	111.44000	0.00027	129.67000	0.00004	315.69000	0.00002	383.89000
Area	0.00026	159.64000	0.00019	225.90000	0.00013	286.26000	0.00006	656.90000	0.00004	963.25000
Area Percentage	0.00066	289.51000	0.00043	411.44000	0.00030	443.52000	0.00013	1009.65000	0.00007	1486.74000
Volume	0.00024	158.77000	0.00017	225.12000	0.00013	270.99000	0.00006	624.47000	0.00002	908.61000
Volume Percentage	0.00080	310.23000	0.00055	450.28000	0.00037	424.85000	0.00006	1019.99000	0.00002	1218.13000
Offset X	0.00078	84.39000	0.00050	135.26000	0.00036	186.18000	0.00015	459.82000	0.00005	675.05000
Offset Y	0.00081	45.13000	0.00057	59.94000	0.00035	42.45000	0.00004	56.58000	0.00003	24.46000

De acuerdo con los resultados de la Tabla 28, se puede observar que al igual que en las muestras consecutivas, las variables que más influencia tienen en el proceso son en su mayoría las variables de entrada, mostrando variación entre parámetros como presión frontal, posición de la mesa y tiempo de ciclo de salida pero permaneciendo constantes la temperatura y la humedad, dentro de las tres variables con mayor influencia, agregando un tercer parámetro que es la presión trasera aunque no dentro de las tres primeras posiciones, seguido de la presión frontal, en la que también se observa una repetibilidad en 4 de 5 ejecuciones.

En un sentido estricto y complementando los resultados obtenidos durante la implementación del algoritmo para muestras consecutivas, se definen dos variables con mayor influencia para el proceso de impresión-inspección de soldadura en pasta en la Tabla 29.

*Tabla 29. Variables con mayor influencia en el proceso de inspección-impresión.*

<b>Variables con mayor influencia</b>
Temperatura
Humedad

De los subconjuntos de datos que integran el conjunto de datos con distintos periodos, se puede resaltar que los parámetros definidos inicialmente como no variables, se mantienen en el mismo estado, comprobando que, a lo largo del proceso, a pesar de las diferentes condiciones, las velocidades de separación y de impresión frontal y trasera, se mantienen constantes, por lo cual, no son relevantes para futuras técnicas que puedan ser aplicadas que puedan estar basadas en la variación.

Los resultados obtenidos para las variables candidatas a reducción, muestran una diferencia respecto a los datos consecutivos, lo cual indica, que el proceso en realidad es un proceso que opera bajo distintas condiciones dependiendo el periodo de tiempo y sugiere que pueden existir otros factores no considerados en los conjuntos iniciales que puedan presentar constancia en variables a reducir, además de que las variables consideradas son importantes para el control del proceso mismo.

Cabe resaltar, que, aunque existe variación entre los parámetros a aplicar reducción, en la última ejecución del algoritmo, para 2000 piezas producidas en cuatro periodos distintos, se ha logrado un resultado favorable como resultado de la comparación de medias y varianzas con el valor umbral, siendo las siguientes variables candidatas a reducción, basados en esta ejecución.

*Tabla 30. Variables candidatas a reducción.*

<b>Variables candidatas a reducción</b>
Presión frontal
Presión trasera

Las cuales, se han mantenido en las primeras cinco variables candidatas a reducción en la ejecución de la técnica de reducción de variables para todas las muestras y además, para el último conjunto de datos, el resultado está por debajo de 0.5, valor umbral definido para todos los conjuntos de datos por igual.

El resultado obtenido para los parámetros presión frontal y presión trasera, de acuerdo con la técnica de reducción de variables, puede ser descartado al igual que los parámetros que no presentan variación, dado que, no son parámetros que tengan potencial para distinguir entre las dos clases de los diferentes conjuntos de datos, no son de utilidad en la aplicación de otras técnicas de machine learning.

Mediante la aplicación de la técnica de reducción de datos, es posible identificar que las variables con mayor influencia en el proceso de impresión-inspección son la temperatura y la humedad y las variables de las que se puede prescindir en la aplicación de otras técnicas son la presión frontal y presión trasera. En definitiva, las velocidades de separación e impresión frontal y trasera no presentan variación alguna durante el proceso.

Ha sido importante para esta investigación, utilizar distintos conjuntos de datos en periodos distintos, consecutivos y no consecutivos, para observar la influencia que pueden tener los distintos parámetros a lo largo del proceso.

### 6.2.2 Escalabilidad

Con recursos computacionales más elevados, es posible aplicar el mismo algoritmo para una mayor cantidad de datos, además con los métodos de recolección establecidos en tiempo real, es posible conocer los parámetros más significativos y la reducción de los mismos de forma inmediata.

Mediante la obtención de datos al igual que se ha realizado en la línea C, para cualquier otra línea de producción y cualquier otro modelo, es posible conocer si son las mismas variables que influyen de forma significativa en el proceso, y si es posible aplicar reducción de las mismas para otros modelos y otras máquinas tanto de impresión como de inspección.

La recolección de datos, por su parte, puede ser utilizada para cualquier proceso, adecuando el guardado de datos de cualquier formato a la base de datos propuesta inicialmente que ya está conectada con el programa del algoritmo.

La técnica de selección de variables puede ser aplicable también, a cualquier proceso, para el que se tengan identificadas las variables que posiblemente intervienen en el mismo, dando un resultado inmediato sobre las variables que son candidatas a reducción.

Mediante la obtención de datos del proceso en tiempo real para entradas y salidas de un proceso, la relación de estos y la aplicación del algoritmo para selección de variables, es posible obtener de forma inmediata resultados de cualquier proceso que se requiera analizar.

La flexibilidad del algoritmo permite el uso de cualquier conjunto de datos, sin limitar al uso de un número específico de variables o cantidad de datos, dependerá entonces del proceso el subconjunto de datos a utilizar y de las variables a explorar.

La clasificación por su parte para las variables del proceso es parte fundamental en la aplicación de la técnica de selección de variables, la cual también, dependerá del proceso y los valores de los parámetros que intervienen en él.

La reducción de datos es entonces parte fundamental en la fase de preprocesamiento de datos, en preparación a las siguientes etapas de *data mining*, simplificando los conjuntos de datos que permitirán procesar de forma más sencilla, los datos relevantes para las etapas siguientes.

## Discusión de Resultados

### **I. Transferencia automática de datos al sistema de control estadístico de procesos.**

Como primer resultado y derivado del método de recolección de datos, se ha logrado la transferencia de datos desde los archivos generados por la impresora de pasta y los datos generados por la máquina de inspección, mediante el uso de técnicas computacionales, al sistema de control estadístico de procesos, el cual, haciendo uso de los datos obtenidos, genera graficas de control para los valores críticos definidos.

Esto ha permitido un aporte significativo a para la empresa, ya que, los datos obtenidos, tardan alrededor de 2 minutos en ser reflejados en las gráficas de control, en el caso de las salidas que es donde se presenta mayor variación, permitiendo conocer en tiempo real el porcentaje de volumen, variable seleccionada como critica para este proceso.

Para los registros de entrada, tomando como ventaja la conexión de la impresora a red, y el método de recolección de datos, se ha decidido utilizar el mismo sistema de control de procesos, para monitorear el comportamiento de las variables de entradas definidas como significativas, mediante las técnicas descritas, las cuales son temperatura y humedad, de acuerdo con el análisis realizado y la variación presentada, se aplica solamente un muestreo para optimizar el uso de recursos tecnológicos de los que dispone la compañía.

Este nuevo método, ha disminuido considerablemente el tiempo de obtención de datos y la violación de políticas de TI como conexión de dispositivos extraíbles para transferencia de datos.

### **II. Registro electrónico de plan de acción integrado al sistema de control estadístico de procesos.**

Se ha implementado, el registro electrónico de acciones efectuadas en el proceso, cuando existe alguna anomalía, con la finalidad de identificar los ajustes que se realizan a lo largo de cada etapa del proceso, lo cual, al ser analizado, permitirá aplicar mejoras significativas en el mismo,

TESIS TESIS TESIS TESIS TESIS

basados en la frecuencia de ejecución de acciones inmediatas mediante el seguimiento del OCAP electrónico.

Este nuevo proceso, en conjunto con la integración con el sistema estadístico de control de procesos ha sido formalmente adoptado por la empresa, mostrando aceptación del área de Ingeniería y Calidad debido a que es posible tener conocimiento en corto tiempo del proceso a través de la obtención de datos de forma automática y registro de acciones para condiciones no favorables en el proceso mismo.

### **III. Notificación inmediata de anomalías a usuarios utilizando mensajería instantánea.**

Después de la puesta en marcha de la obtención de datos y aplicación en el sistema de control de procesos, ha sido necesario enviar una notificación inmediata cuando existe una anomalía en el proceso o la violación a alguna de las reglas para el control del proceso, por lo que se recurre nuevamente a una técnica computacional para el envío de un correo electrónico para notificar el problema a las áreas correspondientes haciendo uso de listas de distribución, aunado a esto, se ha desarrollado para complementar, un servicio mediante el uso de la API de *telegram*, para enviar la misma notificación a través del dispositivo móvil a diversos grupos.

### **IV. Reducción de tiempo de atención a incidentes detectados en el proceso derivado del desarrollo e implementación de un dashboard.**

Como parte de la solución integral y a manera de evaluación de la efectividad de la implementación anterior, se ha conectado a su vez el sistema de control de procesos a un dashboard desarrollado como parte de los esfuerzos y las contribuciones a la empresa, para reducir el tiempo en que se tarda en atender los incidentes, los procesos (que, aunque no son automáticos como estos primeros) que generen mayor tiempo de respuesta, y el cumplimiento de la resolución de estos problemas.

## **V. Extensión de implementación métodos de recolección de datos para otros procesos críticos.**

Derivado de la automatización en la transferencia de datos para la inspección e impresión, para complementar la contribución del presente trabajo a la empresa, se ha incluido la conexión de dos equipos más, fortaleciendo la adquisición de datos en SMT como parte del proceso inicial descrito anteriormente, los cuales son el porcentaje de unidades sin falla en la Inspección Óptica y las temperaturas de horno de reflujo en las diferentes zonas, extendiéndolo también al proceso de inspección de PTH, mediante la transferencia de datos derivados de este proceso.

## **VI. Introducción de técnicas de minería de datos en análisis de procesos industriales**

La implementación de la técnica de reducción de variables se ha adoptado como uno de los métodos de análisis para los nuevos procesos, aunado de los métodos tradicionales estadísticos, simplificando la tarea de selección de variables significativas para un proceso.

La escalabilidad de las soluciones presentadas abre paso al análisis de procesos de diferentes áreas, mediante el uso de gran cantidad de datos, dejando un antecedente para un proceso complejo, supone que los datos de cualquier proceso pueden ser utilizados para la aplicación de esta técnica, simplificando esta tarea.

Si bien, no existe una predicción concreta del proceso de SMT, la obtención de datos en tiempo real en conjunto con el análisis del proceso, marcan la pauta para la aplicación de técnicas subsecuentes de aprendizaje automático enfocadas a contribuir con una predicción acertada de cualquier proceso. La reducción de variables es un precedente importante que permite encaminar cualquier proceso a las etapas posteriores de un algoritmo de predicción.



## Conclusiones

El reconocimiento de los recursos computacionales actuales en la industria permite realizar un análisis adecuado y proponer las técnicas que mejor se adapten a los procesos existentes, sin que implique mayor afectación.

Conocer los sistemas con que cuenta la empresa y las necesidades, permite conjuntar las nuevas técnicas para aplicar mejoras disponiendo de los recursos existentes.

La aplicación de técnicas de aprendizaje automático, simplifican el tiempo de análisis de procesos y permiten utilizar mayor cantidad de datos, arrojando resultados más inmediatos que los métodos de análisis tradicional.

La aplicación de machine learning no está limitada a algún campo de estudio en particular, extendiéndose a prácticamente cualquier ámbito. Los procesos industriales, como se describió, son capaces de generar millones de datos de diversos tipos y volúmenes y no son la excepción.

Mediante la generación, consolidación y análisis de datos, es posible optimizar los procesos industriales a través de la aplicación de diversas técnicas de machine learning, desde la limpieza y reducción de datos hasta las técnicas avanzadas de adaptación independiente de datos. Dependiendo de la naturaleza de los datos arrojados por el proceso, se pueden proponer diversas técnicas que faciliten su análisis, ofreciendo una mejora significativa basada en los resultados obtenidos.

Para realizar una predicción de un proceso de manufactura, es necesario conocer todos los datos involucrados en el proceso, lo cual, desde un inicio no es una tarea fácil, debido a que no todas las variables se encuentran a la mano o son de fácil recolección, pues sugieren modificaciones a los procesos existentes, que no resultan ser de fácil adopción por el personal y la integración es compleja en un inicio. Se busca siempre realizar una obtención de datos mediante métodos automatizados que no impliquen cambios en los procesos actuales, explotando los beneficios que máquinas avanzadas para efectuar los procesos puedan proporcionar, sin embargo, de acuerdo con los procesos presentados, no siempre existen de forma inmediata, por lo que, se busca la generación de ellos buscando el menor impacto posible,

integrándolos a sistemas existentes que no generen problemas de adopción o costos de desarrollo elevados.

A lo largo del camino se ha experimentado, que los datos que en un principio parecen ser los más relevantes, resultan no serlo siempre al final, dejando opción abierta a que datos que parecen no cobrar relevancia en el proceso, sean, los que más influencia tengan sobre él, es por eso que las técnicas de selección de variables son de suma importancia como antecedente para una predicción.

La observación de los procesos de forma física, es sin duda, una de las oportunidades más valiosas para identificar los datos que influyen en el proceso, el análisis descriptivo de estos datos, contribuye también al reconocimiento del proceso a través de los valores que son generados por el mismo, recorrer los procesos dentro de la industria una y otra vez, permite reconocer los cambios que existen y los factores que intervienen en ellos, modelar los datos de forma gráfica y utilizar sistemas existentes para reconocer su variación, permiten también identificar si son o no de importancia para el proceso. La cantidad de datos por su parte depende de la frecuencia con que se realiza un proceso, siendo ésta reconocida también a través de la observación del proceso y de la cantidad de datos generados bajo diferentes condiciones en el proceso, no hay un número exacto para reconocer la cantidad de datos, sino que esté es dado por el mismo proceso y las condiciones que pueda presentar para generar variación.

El mundo de la industria es sin duda un perfecto escenario para la aplicación de varias técnicas computacionales, que son de utilidad cuando están correctamente direccionadas, persiguiendo un objetivo común que es la mejora constante aprovechando los recursos existentes e incluyendo cada vez más el uso de la tecnología y de los elementos más actuales en el ramo como lo es la industria 4.0.

Glosario

cluster	Conjunto o racimo, el término es empleado comúnmente en diversas áreas de informática y ciencias.
dendrita	Cristal metálico que se produce generalmente por solidificación, caracterizado por una estructura similar a la de un árbol con varias ramas.
depanelización	Proceso de separación de tarjetas que se encuentran unidas en una misma placa para simplificar los procesos de SMT y PTH.
fiducial	Marca en el diseño de una PCB, generalmente de cobre en forma redonda, utilizada como punto de referencia en los procesos de SMT, existen marcas fiduciales locales y globales.
flux	Químico fundente utilizado en proceso de soldadura para la fabricación de componentes electrónicos.
minitab	Software estadístico que permite realizar análisis básico y avanzado.
pad	Superficie de metal, normalmente cobre en un circuito impreso en el que se colocan las terminales a soldar de los componentes, existen pads de diferentes formas dependiendo del componente a ser colocado.
parser	Programa computacional para llevar a cabo el proceso de análisis sintáctico y transformación de una estructura específica en otra estructura diferente.
pin	Terminales en las que se colocarán los componentes de la tarjeta, normalmente cada componente requiere dos o más pines para ser colocados. Cada pin tiene diferente tamaño y forma específica de acuerdo con el pad requerido por el componente a ser colocado.
recurso	Elemento disponible para llevar a cabo una actividad. Medios de los que se dispone para la producción de bienes.
telegram	Plataforma de mensajería y voz por protocolo de Internet, enfocada en mensajería instantánea, envío de archivos y comunicación.

## Referencias Bibliográficas

- Basco, A. I., Beliz, G., Coatz, D., & Garnero, P. (2018). *Industria 4.0: Fabricando el Futuro - Ana Inés Basco, Gustavo Beliz, Diego Coatz, Paula Garnero - Google Libros*. Inter-American Development Bank. <https://books.google.com.co/books?hl=es&lr=&id=geiGDwAAQBAJ&oi=fnd&pg=PA8&dq=industria+4.0+pdf&ots=n-DS9mxbTG&sig=vbRVf5zZ1BEX6pcbBV5DEI4qwkk#v=onepage&q=industria 4.0 pdf&f=false>
- Basogain, X., Olabe, M., Espinosa, K., Rouèche, C., & Olabe, J. C. (2000). Realidad Aumentada en la Educación: una tecnología emergente Realidad Aumentada en la Educación. *Revista Mexicana de Tecnología*, 2(3), 14. <http://multimedia.ehu.es>
- Cano, G., Luque, S. ;, Ruiz, L., Ramón, J., & Roldán, C. (2019). Predictive Maintenance Using Machine Learning Techniques. In *Aeipro* (Vol. 23). [http://dspace.aeipro.com/xmlui/bitstream/handle/123456789/2293/AT03-020\\_2019.pdf?sequence=1&isAllowed=y](http://dspace.aeipro.com/xmlui/bitstream/handle/123456789/2293/AT03-020_2019.pdf?sequence=1&isAllowed=y)
- Castells, E., & Pasola, J. V. (2004). *Tecnología e innovación en la empresa*. Universitat Politècnica de Catalunya. Iniciativa Digital Politècnica. <https://books.google.com.mx/books?id=vFZsgeizTO8C>
- Devore, J. (2008). Probabilidad y estadística para ingeniería y ciencias. In *Cengage Learning: Vol. 7th Ed*. <https://doi.org/10.1017/CBO9781107415324.004>
- Dhanda, N., Datta, S. S., & Dhanda, M. (2019). *Machine Learning Algorithms*. Packt Publishing. <https://doi.org/10.4018/978-1-5225-7955-7.ch009>
- Douglas C., M., Elizabeth A., P., & G. Geoffrey, V. (2012). Introduction to Linear Regression Analysis, Fifth Edition. In *A JOHN WILEY & SONS, INC., PUBLICATION* (Vol. 81, Issue 2).
- Fernández Camacho, E., & Bordóns Alba, C. (2004). Control Predictivo: Pasado, presente y futuro. *Revista Iberoamericana de Automática e Informática Industrial ( RIAI )*, 1(3), 1–28.
- Garrell, A., & Guilera, L. (2019). *La Industria 4.0 en la sociedad digital*. ICG Marge, SL. <https://books.google.com.mx/books?id=YnSIDwAAQBAJ>
- Garvin, D. A. (1984). What Does Product Qual Mean.pdf. In *Sloan Management Review* (Vol. 26, Issue 1, pp. 25–43).
- González-Marcos Fernando Alba-Elías, A. (2017). Machine Learning En La Industria: El Caso De La Siderurgia. *Economía Industrial*, 55–63. <http://www.minetad.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/405/GONZALEZ MARCOS Y ALBA ELÍAS.pdf>
- Guerrero Cano, M., Luque Sendra, A., Lama-Ruiz, J. R., Beltrán, A. M. (Coordinador), Félix Ángel, M. (Coordinador), Beltrán, A. M., Félix Ángel, M., Diseño, U. de S. D. de I. del, & Innovación, U. de S. T. D. I. e I. del P. y la. (2019). *El nuevo paradigma de la industria 4.0 y su aplicación a la industria agroalimentaria*. <https://idus.us.es/xmlui/handle//11441/88922>
- Hernández-Pedreira, C., & Da Silva-Portofilipe, F. (2016). Aplicación del control estadístico de procesos (CEP) en el control de su calidad. *Application of Statistical Process Control (SPC) in It's Quality Control.*, 36(1), 130–145. <https://dibpxy.uaa.mx/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=112542396&lang=es&site=eds-live&scope=site>
- Industry, P., Aguilar-pereyra, F., & Larios-osorio, M. I. M. (2017). *El Entorno de la Industria 4 . 0 :*

- Implicaciones y Perspectivas Futuras*. 54, 33–45.
- ISO/ASTM 52900-2015: Additive manufacturing -- General principles -- Terminology, ISO/ASTM Standards (2015). <https://www.iso.org/obp/ui/#iso:std:iso-astm:52900:ed-1:v1:en>
- Kantardzic, M. (2011). *Data M*.
- Martínez-de-Pisón, F. J., Ordieres, J., Pernía, A., Alba, F., & Torre, V. (2007). Reduce of adherence problems in galvanised processes through data mining techniques. *Revista de Metalurgia*, 43(5), 325–336. <https://doi.org/10.3989/revmetalm.2007.v43.i5.77>
- Meneses Guzmán, M., & Hernández Granados, J. B. (2020). *Industria 4.0. Transformación digital, un cambio en el que participamos todos*. [https://revistas.tec.ac.cr/index.php/investiga\\_tec/article/view/5000](https://revistas.tec.ac.cr/index.php/investiga_tec/article/view/5000)
- Muñoz, O. Q. (2019). *Internet de las Cosas (IoT)*. Ibukku, LLC. <https://books.google.com.mx/books?id=vnnEDwAAQBAJ>
- Naya, S. (2018). Nuevo paradigma de Big Data en la era de la industria 4.0. *Revista TOG*, 15(May), 4–9. [www.revistatog.com](http://www.revistatog.com) <http://www.revistatog.com/num27/pdfs/editorial2.pdf>
- Nilsson, N. J. (2005). INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK Department of Computer Science. *Machine Learning*, 56(2), 387–399. <http://www.ncbi.nlm.nih.gov/pubmed/21172442>
- Nuin, J. J. B., Sanz, E. P., & Moreno, E. C. (2020). *Manual práctico de inteligencia artificial en entornos sanitarios*. Elsevier Health Sciences. <https://books.google.com.mx/books?id=88nSDwAAQBAJ>
- Ortiz Clavijo, L. F., Fernández Ledesma, J. D., Cadavid Nieto, S., & Gallego Duque, C. J. (2018). Computación en la Nube: Estudio de herramientas orientadas a la Industria 4.0. *Lámpsakos*, 20, 68–75. <https://doi.org/10.21501/21454086.2560>
- Palao Gil, F. J. (2016). La Cuarta Revolución Industrial y las Organizaciones Exponenciales. *Executive Excellence*, 132, 36–38. <http://www.businessinsider.com/the-extraordinary-size-of-amazon-in-one-chart-2017-1>
- Pardo, J. M. (2012). *Configuración y usos de un mapa de procesos*. [www.aenor.es](http://www.aenor.es)
- Pérez, A., Santamaria, E. K., Operario, D., Tarkang, E. E., Zotor, F. B., Cardoso, S. R. de S. N., Autor, S. E. U., De, I., Dos, A., Vendas, O. D. E., Empresas, D. A. S., Atividades, P. O., Artigo, N., Gest, G. N. R. M. D. E., Para, D. E. F., Miranda, S. F. da R., Ferreira, F. A. A., Oliver, J., Dario, M., ... Volk, J. E. (2017). Estadística de Schaum. In *BMC Public Health* (Vol. 5, Issue 1). <https://ejournal.poltektegal.ac.id/index.php/siklus/article/view/298> <http://repositorio.unan.edu.ni/2986/1/5624.pdf> <http://dx.doi.org/10.1016/j.jana.2015.10.005> <http://www.biomedcentral.com/1471-2458/12/58> <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&P>
- Rodríguez-Salvador, M., & Mancilla-de-la-Cruz, J. (2018). Presencia de la Industria 4.0 en la fabricación aditiva: análisis de tendencias tecnológicas. In *DYNA - Ingeniería e Industria* (Vol. 93, Issue 6, pp. 597–601). <http://search.ebscohost.com/login.aspx?direct=true&db=&AN=132948245&site=eds-live>
- Russo, C., Ramón, H., Alonso, N., Cicerchia, B., Esnaola, L., & Tessore, J. P. (2015). *Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning*. 131.
- Sánchez-Escalona, A. A., Góngora-Leyva, E., & Zalazar-Oliva, C. (2018). Predicción de la resistencia

térmica de las incrustaciones en los enfriadores de ácido sulfhídrico. *Prediction of the Fouling Thermal Resistance on the Sulphydic Acid Coolers.*, 34(3), 90–100. <https://dibpxy.uaa.mx/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=132932673&lang=es&site=eds-live&scope=site>

Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*, 11(0), 36–40.

Sra, S., Nowozin, S., & Wright, S. J. (2012). *Optimization for Machine Learning*. MIT Press. <https://books.google.com.mx/books?id=JPQx7s2L1A8C>

Wisniewski, M., & Rawlings, J. O. (1990). Applied Regression Analysis: A Research Tool. In *The Journal of the Operational Research Society* (Vol. 41, Issue 8). <https://doi.org/10.2307/2583482>

Yang, K., & Basem El-Haik. (2016). *Design for Six Sigma : Roadmap to product development, 2nd Edition* (Issue August).

Zabala-Uriarte, A., Suárez-Creo, R., & Izaga-Maguregi, J. (2009). La predicción de resultados en procesos industriales Una herramienta de control basada en el Conocimiento. *Dyna (Spain)*, 84(3), 209–218. <https://doi.org/10.15446/dyna.v84n200.56571>

Zhang, Y. (2010). *New Advances in Machine Learning*. IntechOpen. <https://books.google.com.mx/books?id=XAqhDwAAQBAJ>