

CIATEQ, A. C. Centro de Tecnología Avanzada
Dirección de Posgrado



*Análisis y optimización del proceso de ensamble de un
actuador magnético*

TESIS QUE PRESENTA

Ing. Armando Erik Rivera Paz
Asesor: Dr. Luis Alberto Cáceres Díaz

Para obtener el grado de

Maestro en
Manufactura Avanzada

San Luis Potosí, San Luis Potosí
junio, 2021

CARTA DE LIBERACIÓN DEL ASESOR



GOBIERNO DE
MÉXICO



CONACYT
Consejo Nacional de Ciencia y Tecnología



San Luis Potosí, SLP, 11 de mayo del 2021.

Mtro. Geovany González Carlos
Coordinador Académico de Posgrado
CIATEQ, A.C.

Los abajo firmantes, miembros del Comité Tutorial del Ing. Armando Erik Rivera Paz, una vez revisado su Proyecto Terminal de tesis, titulado “Análisis y Optimización del Proceso de ensamble de un actuador Magnético” autorizo que el citado trabajo sea presentado por el alumno para su revisión, con el fin de alcanzar el grado de **Maestría en Manufactura Avanzada**.

Sin otro particular por el momento, agradezco la atención prestada.

Firma

Dr. LUIS ALBERTO CÁCERES DÍAZ

Asesor Académico



CARTA DE LIBERACIÓN DEL REVISOR



San Luis Potosí, SLP, a 18 de junio del 2021.

Dra. María Guadalupe Navarro Rojero
Directora de Posgrado
CIATEQ, A.C.

Por medio de la presente me dirijo a usted en calidad de Revisor del proyecto terminal del alumno **Armando Erik Rivera Paz**, cuyo título es:

"Análisis y optimización del proceso de ensamble de un actuador magnético"

Después de haberlo leído, corregido e intercambiado información con el alumno, y realizado los cambios que le fueron sugeridos, puede ser autorizada su impresión, a fin de que se inicien los trámites correspondientes para su defensa.

Sin otro particular por el momento, y en espera de que mis sugerencias sean tomadas en cuenta en beneficio del estudiante y la Institución, agradezco la atención prestada.

Atentamente,

Firma *Raúl P B*

Dr. Raúl Pérez Bustamante

Eje 126 No. 225, Zona Industrial del Potosí, CP: 78395, San Luis Potosí, S.L.P., México.
Tel: +52 (444) 824 03 10 www.ciateq.mx



F31b Revisión: 01-Mar-2021

AGRADECIMIENTOS

Agradezco a CIATEQ, A.C., Centro de Tecnología Avanzada por su apoyo con la beca No. 832267, así también agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por su apoyo con la beca de titulación con número de folio 218MS04.

También agradezco de manera particular al programa CÁTEDRAS CONACyT y la asesoría otorgada por parte del Doctor Luis Alberto Cáceres Díaz, por los consejos y recomendaciones que me hizo, ya que me permitieron mejorar de manera significativa el contenido de la tesis.

Agradezco a mi familia por el apoyo recibido, por su comprensión y sus consejos para buscar y concluir este proyecto de forma exitosa.

A su vez agradezco a los ingenieros Juan Carlos Acosta Noyola y Michael Romankiewicz por su apoyo y enseñanzas técnicas durante gran parte del desarrollo de este proyecto.

Un agradecimiento al Doctor Juan Carlos Torres, que sumó gran importancia en el desarrollo profesional del proyecto, fungiendo como instructor, asesor y gran amigo.

Por último, agradezco especialmente a Gabriela Alavez Alemán, por impulsarme a concluir cada uno de los retos con su apoyo incondicional, comprensión y consejos.

RESUMEN

Los actuadores son dispositivos con la capacidad de recibir un tipo de energía (neumática, hidráulica, magnética, etc.) y transmitirla, realizando un efecto de fuerza con el objetivo de interactuar con un medio externo. Durante la fabricación de un actuador se genera mucha información que permite optimizar los procesos y calidad del dispositivo; sin embargo, el nivel de complejidad y de correlación de los datos dificulta la toma de decisiones. En este trabajo se desarrolló una metodología de optimización del proceso de ensamble de un actuador, el cual se llevó a través del departamento de ingeniería de proceso. Este método se acopla a la nueva tecnología industrial que involucra procesos automatizados y con uso de bases de datos. El uso de la ingeniería y tecnología existente se implementó de forma exitosa para crear la metodología de análisis estadístico, aplicando regresiones lineales simples y múltiples para evaluar el efecto de variables de ensamble sobre características críticas de funcionamiento de los actuadores magnéticos.

Usando las bases de datos existentes de los parámetros de ensamble del actuador, se obtuvo un modelo que cumplía con las condiciones de ensamble del actuador. A partir de este punto se propusieron las nuevas condiciones de sub-ensamble que llevaron a un producto con características mayormente estables y dentro las de las tolerancias especificadas para las características críticas del actuador.

Se tiene como resultado, la confirmación del uso del uso de técnicas de aprendizaje de máquina y bases de datos de fabricación disponibles en la línea de ensamble para la reducción del porcentaje de material de rechazo a un valor menor a 2%. A partir de las técnicas e información utilizados, se generó un modelo de regresión lineal múltiple con la capacidad de predecir en un 87.5 % la relación entre los valores medidos durante el proceso de ensamble y el resultado de una característica de funcionamiento del actuador.

Palabras clave: Ingeniería y tecnología, Tecnología industrial, Ingeniería de procesos, Actuador magnético, Regresión lineal múltiple, Predicción.

ABSTRACT

Actuators are devices with the ability to receive a type of energy (pneumatic, hydraulic, magnetic, etc.) and transmit it, performing a force effect in order to interact with an external environment. During the manufacture of an actuator a lot of information is generated that allows to optimize the processes and quality of the device; however, the level of complexity and correlation of the data makes it difficult to make decisions. In this work, an optimization methodology of the actuator assembly process was developed, which was carried out through the process engineering department. This method is coupled to the new industrial technology that involves automated processes and the use of databases. The use of existing engineering and technology was successfully implemented to create the statistical analysis methodology, applying simple and multiple linear regressions to evaluate the effect of assembly variables on critical operating characteristics of magnetic actuators using existing databases of actuator assembly parameters, a model was obtained that met the actuator assembly conditions. From this point on, the new sub-assembly conditions were proposed that led to a product with mostly stable characteristics and within the tolerances specified for the critical characteristics of the actuator.

The result is the confirmation of the use of machine learning techniques and manufacturing databases available on the assembly line to reduce the percentage of rejected material to a value less than 2%. From the techniques and information used, a multiple linear regression model was generated with the ability to predict by 87.5% the relationship between the values measured during the assembly process and the result of an actuator operating characteristic.

Keywords: Engineering and technology, Industrial technology, Process engineering, Magnetic actuator, Multiple linear regression, Prediction.

ÍNDICE DE CONTENIDO

RESUMEN	iv
ABSTRACT	v
ÍNDICE DE CONTENIDO	vi
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE TABLAS	x
GLOSARIO	xi
1. INTRODUCCIÓN	1
1.1 ANTECEDENTES	3
1.2. DEFINICIÓN DEL PROBLEMA	4
1.3. JUSTIFICACIÓN	7
1.4. OBJETIVOS.....	8
1.5. HIPÓTESIS.....	9
2. MARCO TEÓRICO	10
2.1. ESTADO DEL ARTE	10
2.1. ACTUADORES MAGNÉTICOS	11
2.1.1. Partes de un actuador magnético.....	11
2.1.2. Principio de funcionamiento	13
2.1.3. Características críticas de funcionamiento de un actuador magnético.....	14
2.1.4. Proceso de ensamble de un actuador magnético	16
2.2. CIENCIA DE DATOS	21
2.2.1. Minería de datos ("Data mining")	22
2.2.2. Proceso de minería de datos	23
2.2.3. Métodos de modelado y aprendizaje de máquina	24
2.2.4. Correlación lineal de variables	27
2.2.5. Modelo de regresión	28
2.2.6. Selección de predictores para modelo óptimo.....	29
2.2.7. Uso de software y lenguajes de programación	32
3.3. OPTIMIZACIÓN DE PROCESOS.....	33
3.3.1. Cantidad de material de rechazo (% de Scrap).....	34
3.3.2. Índices de capacidad (Uso de cpk y ppk)	35
3. PROCEDIMIENTO.....	37

3.1. ENTENDIMIENTO DEL NEGOCIO	38
3.3.1. Equipo de medición utilizado.....	42
3.2. ESTUDIO Y COMPRESIÓN DE LOS DATOS	46
3.3. PREPARACIÓN DE LOS DATOS.....	52
3.4. MODELADO	54
3.4.1. Análisis de correlación de los datos	54
3.4.2. Modelo de regresión lineal	55
3.4.3. Selección de predictores.....	55
3.4.4. Estimación de bondad de modelo	56
4. RESULTADOS	57
4.1. PRUEBA DE NORMALIDAD EN VARIABLES DE INTERÉS	57
4.2. RESULTADOS DEL ANÁLISIS DE CORRELACIÓN.....	61
4.3. RESULTADOS MODELOS DE REGRESIÓN.....	62
4.3.1. Selección de predictores.....	63
4.4. EVALUACIÓN DEL MODELO	65
4.5. DESPLIEGUE.....	71
CONCLUSIONES	73
RECOMENDACIONES	74
APORTACIÓN DE LA TESIS.....	75
APORTACIÓN SOCIAL DE LA TESIS.....	76
REFERENCIAS.....	77
ANEXO	83

ÍNDICE DE FIGURAS

Figura 1. Diagrama general de un actuador magnético lineal solenoide.....	2
Figura 2. Comparativa Producción Actuadores OK, No OK y Objetivo de No OK... 6	6
Figura 3. Elementos de un actuador magnético.	12
Figura 4 Representación líneas de campo magnético en una bobina.....	13
Figura 5 Evaluación del tiempo de reacción y conmutación.	16
Figura 6. Diagrama proceso de ensamble de actuador magnético.	17
Figura 7. Proceso de Embobinado de un actuador magnético.	18
Figura 8 Proceso soldadura por contacto.	19
Figura 9 Proceso de dispensado de silicón.....	20
Figura 10 Proceso de prensado.	21
Figura 11. Predicción de sistema de manufactura usando análisis predictivo y el método Watchdog Agent.	22
Figura 12 Esquema del Proceso KDD.....	23
Figura 13 Métodos y técnicas dentro de la minería de datos.	26
Figura 14 Selección de técnica estadística.....	27
Figura 15 Gráfica de dispersión con regresión lineal y no lineales.	31
Figura 16 Lenguajes de programación para Machine Learning.....	33
Figura 17 Diagrama causa efecto.....	35
Figura 18 Representación del índice del proceso.....	36
Figura 19 Implementación método CRISP-DM.....	38
Figura 20 Bases de datos disponibles y almacenamiento en línea de ensamble de actuadores magnéticos.	40
Figura 21 Envío de resultados entre celdas de ensamble.....	41
Figura 22 Sistema de visión.	43
Figura 23 Sensor de Luz.	43
Figura 24 Medición de tiempo de conmutación y reacción con sensores de luz..	44
Figura 25 Transductor de resorte.....	44
Figura 26 Medición de altura y profundidad con uso de transductor.	45
Figura 27 Sensor de fuerza.....	45
Figura 28 Representación medición de fuerza retracción pin.....	46

Figura 29 Estructura de base de datos de la línea de fabricación.	47
Figura 30 Ejemplo algoritmo para exploración de datos, Software Spyder.	48
Figura 31 Gráfica de distribución y distribución ajustada empírica con distribución normal.	50
Figura 32 Evaluación de ajuste adecuado y deficiente de distribución empírica.	51
Figura 33 Proceso de preparación de datos.....	53
Figura 35 Gráfica con valores atípicos para característica fuerza de retracción correspondientes al mes de marzo de 2019.....	54
Figura 35 Gráficas de distribución empírica por característica resistencia eléctrica.	57
Figura 36 Gráficas de distribución empírica por característica altura de pin retraído.....	58
Figura 37 Gráficas de distribución empírica por característica crítica fuerza de retracción.....	59
Figura 38 Matriz y mapa de calor de coeficientes de correlación de Pearson R... ..	61
Figura 39 Gráfica de distribución de residuos.....	66
Figura 40 Gráfica Q-Q de residuos en comparativa con cuantiles normales.....	67
Figura 41. Valores de predicción de datos en función del valor real.	69
Figura 42 Dispersión de los residuos del modelo.....	70

ÍNDICE DE TABLAS

Tabla 1 Comparativa desviación estándar, habilidad de proceso cpk y % material rechazado.	39
Tabla 2 Resumen datos faltantes en base de datos.....	48
Tabla 3 Clasificación de variables en la base de datos enero- junio 2020.....	49
Tabla 4 Resumen con clasificación de variables.	49
Tabla 5 Estadístico de características críticas.	50
Tabla 6 Reporte prueba Shapiro Whiks.	52
Tabla 7 Reporte Análisis de Curtosis y Asimetría.	52
Tabla 8 Resultados prueba Shapiro Wilk, correspondientes al mes de enero 2020.	60
Tabla 9 Resultados prueba análisis de curtosis y asimetría, correspondientes al mes de enero.	61
Tabla 10 Valores del primer modelo de regresión lineal.....	62
Tabla 11 Coeficientes parciales de la regresión múltiple.....	64
Tabla 12 Valores del modelo de regresión lineal optimizado.	65
Tabla 13 Resultado % de material rechazado.....	72

GLOSARIO

Actuador magnético: son dispositivos que tienen la capacidad de recibir y convertir energía, la cual realiza un efecto de interacción con un medio externo.

Aprendizaje de máquina: se refiere al uso de algoritmos y técnicas que nos ayudan a identificar patrones dentro de la base de datos y permiten la generación de modelos que describen dichos patrones.

Bondad de ajuste (R²): se refiere a la proporción en la cual se dispersa la variable respuesta explicada por el modelo y relativa a la varianza total.

Ciencia de datos: área de estudio que integra diferentes ramas de estudio tales como la programación, la estadística y los métodos exploratorios de datos.

Correlación lineal: método de análisis usado en modelos lineal para cuantificar el grado de relación que existe entre dos variables.

Covarianza: es un indicador que mide la variación que existe forma conjunta entre dos variables respecto al valor promedio de las mismas.

Desviación estándar: medida que se utiliza en la estimación de la variación o la dispersión de un conjunto de datos numéricos.

Dispensado: se produce al realizar el dispensado de un material de alta viscosidad sobre una superficie la cual se desea sellar.

Distribución normal: distribución de probabilidad de un estadístico particular el cual puede describirse por medio de la media μ y la varianza de los datos σ^2 .

Embobinado: se realiza cuando se enrolla un material conductor alrededor de una armadura, formando devanados con diferentes configuraciones.

Error (rmse): calcula la desviación media de cualquier punto por el modelo respecto de la recta de regresión.

Homocedasticidad: Se refiere a la estabilidad constante al realizar el cálculo del error de predicción de las observaciones.

Método CRISP-DM: se refiere a un modelo estandarizado, el cual tiene como objetivo detallar el proceso de exploración de datos. Es uno de los modelos analítico más usado en el área de minería de datos.

Minería de datos (data mining): es un conjunto de técnicas dentro de la ciencia de datos, la cual se aplica para el desarrollar el proceso de aprendizaje de la información a través de métodos que describen, correlacionan y descubren patrones o tendencias.

Parsimonia: se refiere a aquel modelo con la mejor capacidad de predicción basada en la optimización del uso de predictores para la formación del mismo modelo.

Prensado: El proceso de unión de dos materiales por medio del formado, se basa en la deformación mecánica de ambos materiales.

Regresión lineal: método de análisis que para modelar la relación lineal entre una variable dependiente con una ó varias variables independientes.

Soldadura por contacto: se lleva a cabo cuando se hace circular una cantidad de corriente a través de dos placas metálicas, las cuales se encuentran en contacto en sus extremos superiores con dos elementos denominados electrodos.

Variable de respuesta: variable que depende de k variables independientes.

Variable dependiente: se le denomina a aquella que describe una cantidad cuya magnitud se relaciona con otra magnitud.

1. INTRODUCCIÓN

Este estudio se enfoca en el uso de actuadores electromagnéticos lineales tipo solenoide, los cuales de acuerdo con la norma alemana VDE 0580, son aquellos que se mueven de un inicio hasta una posición final, como resultado de las fuerzas electromagnéticas, y su retorno se puede dar debido a la interacción del actuador con una fuerza mecánica de un dispositivo externo.

Los actuadores son dispositivos que tienen la capacidad de recibir y convertir energía, la cual realiza un efecto de interacción con un medio externo. Los actuadores electromagnéticos (a partir de aquí llamado solamente actuador magnético) son usados en componentes de seguridad, sistemas de inyección y compresión para vehículos automotores tipo SUV, camioneta y pick-up (Kocabicak & Demir, 2020; Wang et al., 2020). El principio de funcionamiento de un actuador electromagnético de solenoide se basa en la ecuación de Lorenz, la cual describe la relación de la magnitud de la fuerza generada en función de su campo magnético (Endesa Fundación, 2020; Gutiérrez & Iturralde, 2017). El ensamble de un actuador electromagnético se realiza en diferentes pasos, resumidos a continuación:

Subensamble A: enrollado del hilo conductor alrededor del núcleo magnético.

Subensamble B: ensamble de hilo-núcleo dentro de carcasa; ensamble magneto varilla dentro de carcasa; ensamble de actuador magnético completo.

Al final de cada proceso de ensamble se realiza una evaluación que determina si el ensamble cumple con las tolerancias requeridas. Los resultados de cada evaluación (aprobada y no aprobada) son registrados en una base de datos que se almacena en tiempo real. Los datos medidos de cada proceso de ensamble y de cada variable contiene información relevante para mejorar el proceso y para identificar las fallas que pueda tener el actuador magnético producido (Ismail et al., 2019).

En la Figura 1 se aprecia el diagrama general con las partes del actuador magnético.

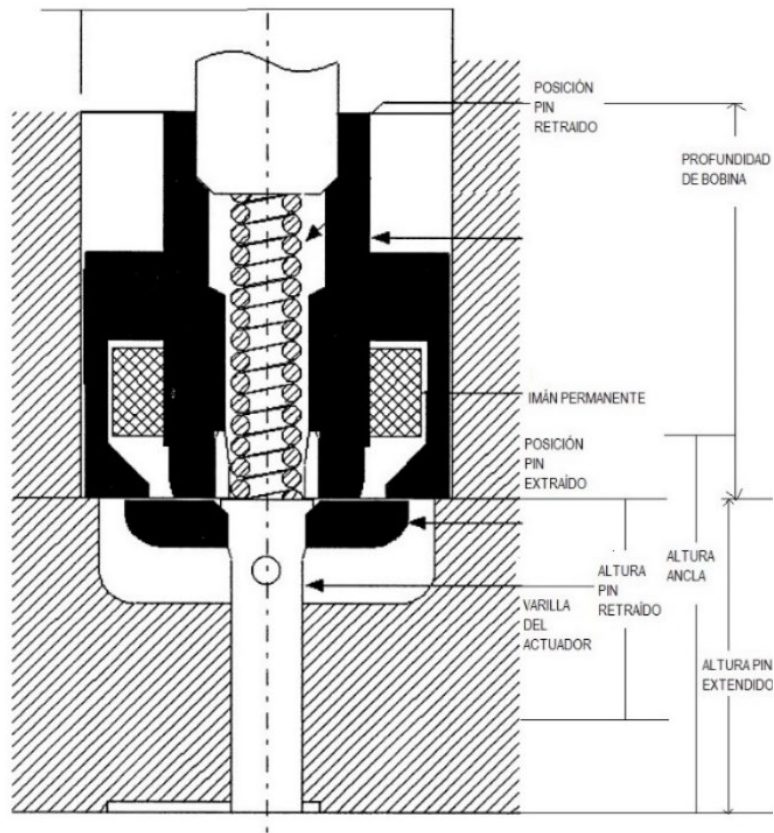


Figura 1. Diagrama general de un actuador magnético lineal solenoide.

Fuente: Imagen adaptada de (Howe, 2000).

El problema radica en que la cantidad de datos es extensa y la relación de codependencia que existe entre las variables aumenta la complejidad de los análisis. Para garantizar un proceso estable se requiere de una metodología que permita la toma efectiva de decisiones, la cual pueda estar basada en el análisis de la información obtenida durante la fabricación de las piezas.

Debido a lo anterior, y en conjunto con los constantes avances en la tecnología y las altas demandas de actuadores magnéticos en la industria automotriz, las operaciones de ensamble son realizadas en una línea de ensamble totalmente automatizada. Estas operaciones son comúnmente usadas para incrementar la competitividad y lograr un óptimo grado de flexibilidad, calidad y productividad (Costa et al., 2018).

El enfoque para la solución y el análisis de problemas en una línea de ensamble automatizada requiere de un método que permita la toma de decisiones rápida con base en un método definido. En consecuencia, se busca la posibilidad del uso de los datos generados dentro del mismo proceso, para realizar un análisis basado en ciencia de datos y que permita establecer la relación entre las características de funcionamiento del actuador magnético y sus parámetros de funcionamiento. Dicho análisis se realiza en conjunto con técnicas de aprendizaje de máquina, debido al gran volumen de datos que una línea automatizada puede generar.

Por su parte, está probado que el uso de técnicas de procesamiento de datos es efectivo para resolver problemas de aprendizaje de máquina (Jundong Li et al., 2018).

1.1 ANTECEDENTES

En México, la implementación de procesos automatizados en diferentes actividades se ha ido incrementando gradualmente. En 2018, se mostró en el trabajo de Frey y Osborne (Frey & Osborne, 2017) una estimación, la cual mostró que casi el 80% de las actividades realizadas en la industria manufacturera contaban con una alta probabilidad de ser reemplazadas por una automatización (Banco de México, 2018). Sin embargo, la inserción del uso de nuevas tecnologías como técnicas de análisis de datos basados en aprendizaje de máquina ("Machine learning") o inteligencia artificial, no se ha consolidado de manera eficiente.

El uso de la ciencia de datos en San Luis Potosí se ha utilizado ampliamente en sectores financieros, médicos, turísticos entre otras. Esto se ha realizado mediante el uso de plataformas o Software especializado, con el objetivo de optimizar sus operaciones y por ende rentabilidad. En 2018, se analizó el uso y ventajas de estas plataformas digitales en economías de colaboración en San Luis Potosí y con lo cual se concluyó en un beneficio con una aceleración en el aprendizaje continuo del negocio y reducción de costos (Alavez, 2018).

En el caso de la industria manufacturera su proceso ha sido más lento, sin embargo, este proceso se ha visto desarrollado a medida que las nuevas tecnologías se han

implementado en los nuevos procesos. Tal es el caso de empresas como la planta de BMW, con la integración de casi el 90% de sus procesos automatizados (Cluster Industrial, 2020).

Existen diferentes métodos y técnicas que se han implementado para la mejora de los procesos, la optimización de la productividad y control de calidad. Dentro de las mismas, destacan el uso de técnicas de predicción, uso de plataformas para compartir conocimiento y el uso de aplicaciones con datos en tiempo real (Wadhwa, 2019). El uso de técnicas de predicción contribuye a disminuir la cantidad de defectos de material y a la optimización de actividades de proceso, por medio de la generación de señales (llamadas en inglés "insights"), las cuales pueden predecir el desempeño del proceso o la prevención de fallas en el equipo (J. Lee et al., 2013).

1.2. DEFINICIÓN DEL PROBLEMA

El proceso de ensamble del actuador magnético descrito en este estudio se realiza en una línea totalmente automatizada, lo anterior debido a los requerimientos en las características críticas de funcionamiento y las tolerancias que influyen en el desempeño del mismo. Este proceso se lleva a cabo en seis estaciones o también llamadas celdas de ensamble, de las cuales 5 celdas son destinadas para el ensamble del actuador y la sexta y última celda se destina para realizar la evaluación respecto a sus características críticas de funcionamiento.

Es importante que cada uno de los actuadores cumpla con las tolerancias especificadas en cada una de las características críticas de funcionamiento, ya que estas garantizan el correcto desempeño del actuador al interactuar en los diferentes sistemas.

Al ser una línea de ensamble completamente automatizada, esta cuenta con la capacidad de medir y registrar los resultados de los diferentes procesos de fabricación de cada uno de los actuadores, lo anterior con el objetivo de garantizar la trazabilidad de las características críticas de cada uno de los actuadores fabricados.

Al finalizar la fabricación de cada uno de los actuadores, dichos registros son almacenados en diferentes bases de datos, las cuales pueden ser consultadas en caso de ser requeridas. Sin embargo, durante el año 2019 se detectó que la toma de decisiones ante problemas de funcionamiento en el equipo o bien desviaciones en la calidad del producto, no estaban basadas en ningún método implementado, derivando en largos tiempos para la solución de fallas y/o largo tiempo en aprobar cambios para la optimización del proceso.

En un inicio la información que se generaba en cada uno de los procesos no se utilizaba para generar algún indicador que llevara a la toma de decisiones en el control del proceso, análisis de fallas, etc.

Aunado al problema anterior, existió la falta de conocimiento en los métodos y estrategias, para el análisis de los datos y su traducción en un control de proceso basado en indicadores o señales. Al no realizar un ajuste apropiado en los procesos, el ensamble del actuador estuvo sujeto a diferentes cambios los cuales se vieron traducidos con un impacto significativo en el funcionamiento del actuador, dichos cambios al ser detectados en el proceso de inspección se tradujeron en características críticas fuera de tolerancia y por lo tanto material rechazado por el mismo motivo.

En consecuencia, el no realizar un ajuste apropiado, la cantidad de material rechazado debido a características críticas fuera de especificación se incrementó de por arriba de los objetivos (<2%). Lo anterior, tuvo varias implicaciones, como mayores costos de operación, reajustes en la planeación de entrega de materiales o incluso posibles retrasos en la entrega de producto terminado al cliente.

En la Figura 2 se muestran los resultados de producción en el año 2019, en la gráfica se representa en el eje X los meses del año, en el eje Y de lado izquierdo se muestran la cantidad de actuadores producidos y por último en el eje Y de lado derecho se observa el porcentaje total de producción.

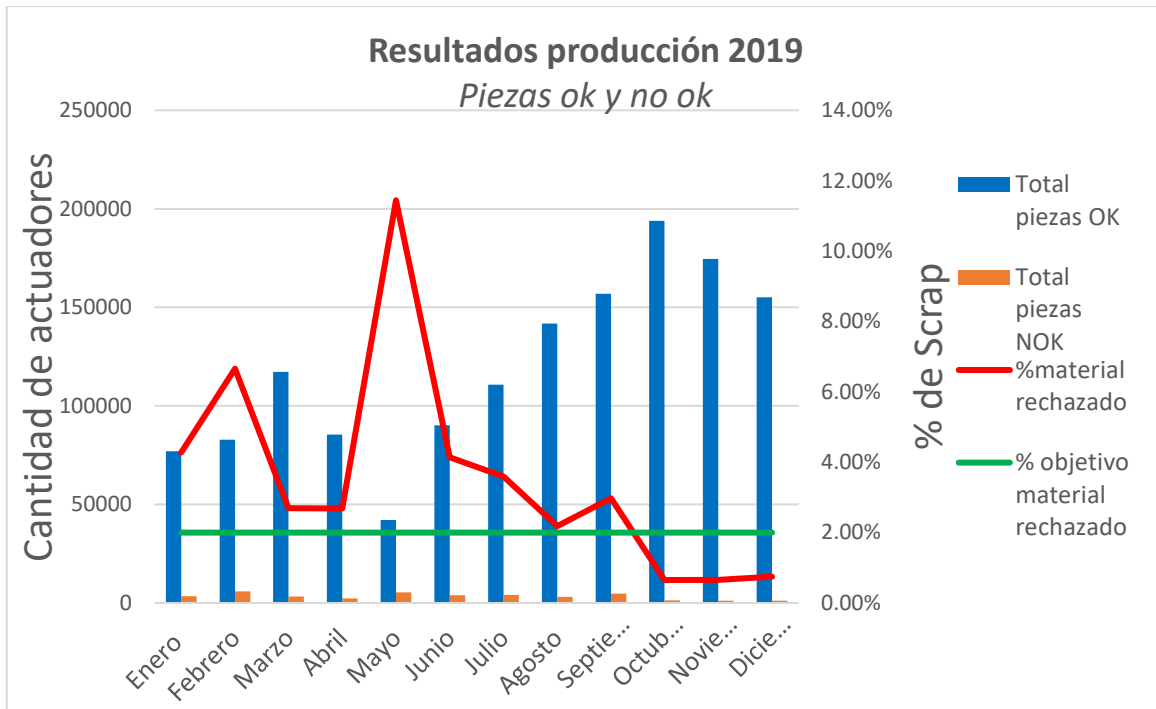


Figura 2. Comparativa Producción Actuadores OK, No OK y Objetivo de No OK.

Fuente: elaboración propia.

Como se puede observar en la gráfica, la cantidad de actuadores aprobados fabricados alrededor del año (barras azules) no presenta un comportamiento constante, mostrando variaciones significativas en la producción. Es el caso del mes de mayo con una producción de casi 50 mil actuadores, el cual se contrasta con la cantidad de actuadores fabricados durante el mes de octubre con una producción de casi 200 mil actuadores.

Lo anterior también se ve reflejado en el porcentaje de material rechazado (línea roja), en la cual se observan los mismos meses con casi un porcentaje de material de rechazo en mayo del 12% y el mes de octubre con un porcentaje debajo del 2% que se tiene como objetivo.

1.3. JUSTIFICACIÓN

Como se mencionó, la falta de conocimiento para convertir los datos de fabricación en indicadores de ajuste de proceso evita que se detecten variaciones y/o tendencias significativas, que puedan ser corregidas mediante acciones preventivas. Lo anterior incentiva a realizar un estudio con enfoque en la reducción del material rechazado y a describir las condiciones que lleven a un proceso estable. El no reducir la cantidad de material de rechazo no solo es poco rentable para la empresa, ya que al no ser ensamblado el material no cumple su ciclo de vida y debe ser desechado, lo cual también implica un impacto ecológico. El analizar los datos, puede traer beneficios como una curva de aprendizaje reducida en las operaciones de ensamble, lo cual deriva en control de proceso en menor tiempo en caso de tener nuevo personal. El tener una curva de aprendizaje corta, repercute directamente en la reducción de material de rechazo.

Por otro lado, al realizar un estudio de datos, da la posibilidad a describir los rangos en los cuales las condiciones de proceso resultan en un comportamiento estable y óptimo de las características críticas. Una vez obtenido los rangos óptimos de trabajo se pueden realizar y analizar propuestas que optimicen el desempeño del actuador, esto implica obtener una ventaja competitiva para crear procesos con mayor rentabilidad.

Debido a que la línea de ensamble es totalmente automática y los datos ya son generados, los costos que implican el desarrollo del o los indicadores son relativamente bajos en comparación con los beneficios previamente descritos. Por último, el tener un mayor conocimiento con el proceso y garantizando su estabilidad en producción, reduce el riesgo de no cumplir con las entregas con el cliente y generando también confianza para integrar nuevos proyectos.

1.4. OBJETIVOS

Se propone optimizar el proceso de ensamble de los actuadores con el uso de la información del proceso generada y almacenada en las bases de datos durante el periodo de enero a junio de 2020, empleando ciencia de datos. Con esta información se propone obtener un indicador que describa las condiciones óptimas de ensamble, para reducir la cantidad de actuadores completamente ensamblado rechazados debido a características críticas fuera de especificación.

- Objetivo general

Reducir la cantidad de material de rechazo debajo del 2%, generando un modelo matemático que permita predecir las condiciones comportamiento óptimo de cada Característica Crítica (C.C.) en relación a sus parámetros de ensamble.

- Objetivos específicos

Para llevar a cabo el desarrollo de un modelo que prediga el comportamiento, es necesario establecer una serie de objetivos clave para su desarrollo:

- a. Revisar la literatura disponible para establecer el estado del arte en cuanto a ciencia de datos y optimización de procesos.
- b. Establecer el proceso de extracción, tratamiento y análisis de datos, identificando aquellas variables que puedan tener influencia en las C. C. del actuador magnético.
- c. Evaluar que los predictores y C.C. cumplan con las condiciones necesarias para generar el modelo propuesto.
- d. Evaluar y establecer el grado de correlación entre las C.C. y los predictores escogidos.
- e. Proponer el modelo inicial que prediga el comportamiento de las C.C. y calcular el grado de error entre las predicciones y los valores reales.
- f. Evaluar y comparar el porcentaje de material rechazado, una vez optimizadas las condiciones de ensamble.

1.5. HIPÓTESIS

La toma de decisiones en manejo de alta densidad de información se puede optimizar considerando procedimientos de minería de datos y aplicando técnicas de aprendizajes de máquina. Con lo cual será posible predecir el comportamiento de las características críticas para el funcionamiento de un actuador magnético, considerando los parámetros de ensamble y sus posibles predictores.

2. MARCO TEÓRICO

El término de ciencia de datos ("data science") se aplica al campo de estudio que toma como materia prima "los datos", los cuales normalmente no cuentan con una estructura homogénea (Dhar, 2013). El proceso de análisis de datos comprende un proceso de interpretación de los datos, para determinar y estructurar la información que aporte relevancia para la toma de decisiones. De esta forma se crea un conjunto de datos, el cual pasa por un proceso de transformación o también llamado procesamiento. A partir del procesamiento de los datos se crea una nueva base de datos, la cual se analiza mediante el uso de algoritmos y técnicas estadísticas para encontrar interpretaciones basadas en patrones y de esta forma lograr una consolidación hacia un nuevo conocimiento (García et al., 2018). El uso de los datos para generar indicadores o también llamados Insights permite la toma de decisiones de forma efectiva. Por su parte, está probado que el uso de técnicas de procesamiento de datos es efectivo para resolver problemas de grandes volúmenes de datos a través de técnicas de aprendizaje de máquina (Jie Li et al., 2016).

Al realizar el estudio de los datos, estos permiten mostrar información para la toma de decisiones. El personal dedicado analizar los datos se le conoce como analista de datos (data scientists en inglés).

2.1. ESTADO DEL ARTE

La ciencia de datos es relativamente nueva en los análisis de problemas diarios, en ella se pueden resolver problemas por métodos como modelos predictivos, agrupamiento de los datos y el uso de reducción de dimensional.

El uso de la ciencia de datos en los procesos de manufactura puede optimizar considerablemente la rentabilidad de departamentos como calidad, logística y producción (Wadhwa, 2019). Hasta 2015 en México, la búsqueda de analistas de datos en los campos laborales era escasa, arrojando solo 7 resultados dentro de la oferta laboral. Sin embargo, la tendencia es que cada vez más empresas líderes tienden a invertir y desarrollar en este campo (Blancas, 2015).

El uso de la ciencia de datos en México es relativamente nuevo, sin embargo, su inserción en los diferentes campos ha tenido un gran auge en sectores como salud, finanzas, seguridad, entre otras. Debido a lo anterior, hoy en día se han creado ofertas educativas enfocadas a cubrir dichas necesidades tecnológicas, y en muchas ocasiones el uso de plataformas digitales facilita los avances y cambios. Esto ha generado gran demanda de expertos con alto conocimiento en lenguajes de programación, manejo de base de datos, estadística y matemáticas.

En San Luis Potosí, se han insertado una gran cantidad de empresas con tecnología cada vez más avanzada, con gran enfoque en el manejo de las nuevas tecnologías y la manipulación de los datos para generar un indicador. Sin embargo, esto no ha sido ampliamente diseminado en la mayoría de las en el manejo entidades el país.

El proceso de ciencia de datos se puede realizar por medio de diferentes herramientas estadísticas u otros métodos, entre los cuales se puede mencionar el análisis exploratorio, automatización por medio de aprendizaje de máquina o aprendizaje profundo o el uso de minería de datos para solución de problemas. La minería de datos puedes ser tomada como el resultado obtenido después de realizar a través de diferentes técnicas y con el uso de diferentes herramientas un análisis o investigación sobre una base de datos.

2.1. ACTUADORES MAGNÉTICOS

El principio de funcionamiento de un actuador electromagnético solenoide (a partir de aquí llamado solamente actuador magnético se basa en la ecuación de Lorenz, la cual describe la relación de la magnitud de la fuerza generada en función de su campo magnético (Gutiérrez & Iturralde, 2017).

2.1.1. Partes de un actuador magnético

De acuerdo con Li (H. R. Lee et al., 2016) un actuador magnético se compone esencialmente de un embobinado, un núcleo magnético, un magneto permanente o también llamado magneto, un pin o vástago, una guía y por último un sistema o dispositivo que permite el retorno de la unión pin - magneto a la posición original (posición de pin retraído).

El embobinado está formado por un material conductor, el cual se enrolla alrededor de un contenedor plástico llamado "bobina", la cual almacena en su interior un núcleo magnético, este embobinado en el núcleo genera un campo magnético inicial. El dispositivo de Unión entre magneto-pin, normalmente presenta un magneto con forma de arandela unido por la parte de en medio con el pin. El sistema o dispositivo que permite el retorno de la unión magneto –pin a la posición de inicio, puede estar formado por un sistema de resorte o puede estar sujeto a la aplicación de una fuerza por un dispositivo externo. Por último, para garantizar la dirección y el correcto desplazamiento del dispositivo se tiene una guía a través de la cual se desplaza la unión entre el pin y el magneto, al mismo tiempo esta guía define la posición final del recorrido del magneto.

En la Figura 3 se observa el conjunto de partes para un actuador magnético utilizado en la reducción de consumo de combustible.

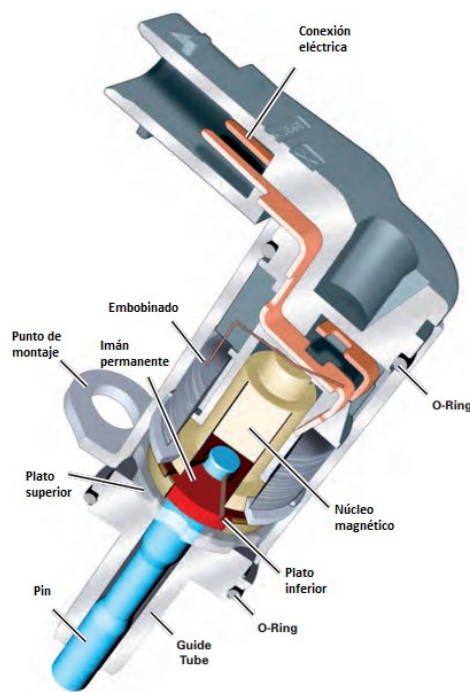


Figura 3. Elementos de un actuador magnético.

Fuente: Imagen adaptada de (Audi of America LLC, 2009)

2.1.2. Principio de funcionamiento

El proceso comienza en el estado inicial del actuador, en el cual existe una interacción entre el campo magnético del magneto permanente y el campo magnético generado en la bobina. En la Figura 4 se muestra la representación de las líneas de flujo magnético del embobinado, las cuales se presentan con una polaridad opuesta a las líneas de flujo del magneto, con lo cual el magneto se mantiene con una fuerza de atracción hacia el núcleo magnético.

Al aplicar un voltaje sobre la bobina, se induce un cambio en la polaridad del campo magnético de la bobina, con lo cual ambos campos magnéticos son similares. Como resultado el magneto es forzado a desplazarse en conjunto con la varilla en sentido opuesto a la dirección de la bobina y a través de la guía llega a la posición de pin extraído.

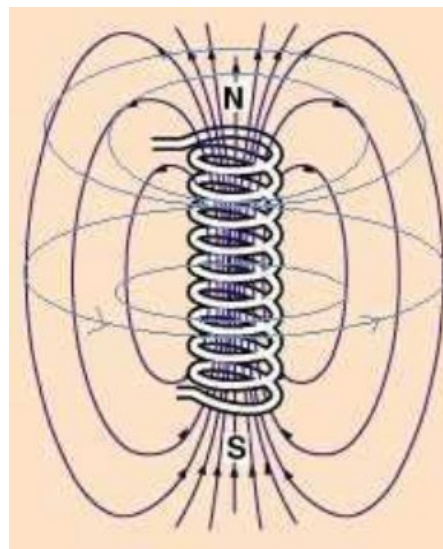


Figura 4 Representación líneas de campo magnético en una bobina.

Fuente: Imagen adaptada de (Hu, 2015).

Para poder regresar el magneto a su posición inicial existen dos posibilidades, ya sea invertir nuevamente el campo y que la atracción entre los polos opuestos regrese al magneto a su posición inicial, o a través de un medio mecánico forzar a que el magneto vuelva a su posición inicial. Por último, ambos elementos se encuentran dentro de un émbolo o también llamado camisa, el cual ayuda al igual que la guía a direccionar al magneto al realizar su recorrido.

Para garantizar el correcto funcionamiento del actuador magnético al interactuar con un sistema externo, se establecen características físicas las cuales son denominadas características críticas. Una característica crítica (a partir de ahora, C.C.), es aquella que juega un papel clave en el funcionamiento y desempeño del actuador magnético. Debido a que la aplicación puede variar, solo se mencionan en forma general las más importantes y su papel en el funcionamiento.

2.1.3. Características críticas de funcionamiento de un actuador magnético

Entre las características más importantes para el funcionamiento de un actuador magnético se encuentran:

- Resistencia eléctrica
- Altura de pin retraído y extendido
- Promedio de tiempo de reacción
- Tiempo de conmutación
- Fuerza de retracción

De acuerdo con la ley de Ohm, al aplicar una diferencia de potencial o bien llamado voltaje en un material conductor, la cantidad de corriente obtenida es proporcional a dicha diferencia de voltaje (Resnick et al., 1998).

$$I = \frac{V}{R}$$

Por ende, la resistencia del material conductor es un factor clave en la cantidad de corriente obtenida a partir del voltaje aplicado. La relación que existe entre la corriente de un material conductor y el campo magnético generado en el mismo se explican mediante la ley de Ampere y de Biot-Savart (Resnick et al., 1998).

En esta ley se establece que el campo magnético obtenido es igualmente proporcional la cantidad de corriente eléctrica en el material conductor.

$$\sum B\Delta l = \mu_0 I$$

La altura de pin en posición retraído y extendido se refieren a la distancia necesaria para lograr que el actuador cumpla su función de diseño y se permita interactuar con un medio externo. Cuando el actuador se encuentra en posición retraído se

garantiza que él actuador no está en contacto con ningún dispositivo externo y por ende no existe ninguna fuerza sobre el actuador en este estado. En el caso de la posición de pin extendido, se garantiza que él actuador puede interactuar con un medio externo y por ende lograr la función de diseño.

La posición del magneto dentro del actuador es importante, debido a que su interacción con la bobina depende también de la distancia a la que se encuentren inicialmente. De igual forma en que la posición inicial afectará en la influencia magnética, la posición final también se ve afectada para lograr que el magneto pueda regresar a su posición inicial. También esta característica afecta en caso de que el magneto interactúe con un sistema externo a través del Pin, a que este puede estar fuera de la tolerancia de la altura.

El tiempo promedio se refiere a la medición del tiempo a partir del momento en que fue inducido el voltaje en la bobina hasta el momento en que se detecta el primer movimiento del Magneto derivado de la influencia de los campos magnéticos. Para realizar la evaluación de dicha característica, se gráfica el tiempo tomado para realizar el cambio de posición del actuador.

Para garantizar el tiempo promedio de reacción sea un valor confiable, esta prueba se realiza un total de 5 veces y por lo tanto el promedio del total de mediciones es el valor obtenido al final de la prueba.

El obtener un tiempo de reacción confiable garantiza que él actuador puede recorrer la distancia en un tiempo preestablecido y por ende que él actuador podrá estar en contacto con un dispositivo externo en el momento que se requiera. Obtener un tiempo de reacción muy largo afecta en no garantizar que él actuador estará en contacto con el dispositivo externo en el momento que se requiera, por otro lado, si el actuador tiene un tiempo de reacción muy corto es posible que entra en contacto con el dispositivo externo antes de tiempo, lo cual también afecta su funcionamiento.

El tiempo de conmutación se refiere al tiempo total medido desde que se induce un voltaje en la bobina hasta el momento en que el pin alcanza la posición final (posición pin extendido). Este tiempo garantiza no sólo que él actuador

reaccionará en el tiempo preestablecido, sino que el recorrido que realiza el magneto a largo de la guía siempre se realiza en el mismo tiempo.

Para garantizar que el tiempo medido sea confiable esta prueba también realiza un total de 5 veces, cuál el promedio de los valores obtenidos es el valor como resultado de esta prueba. Las consecuencias de no cumplir esta característica son similares a las del tiempo promedio de reacción.

En la Figura 5 se observa las gráficas de medición de cambio de posición al aplicar un voltaje sobre el actuador magnético, la cual se utiliza para obtener el tiempo de reacción y tiempo de conmutación.

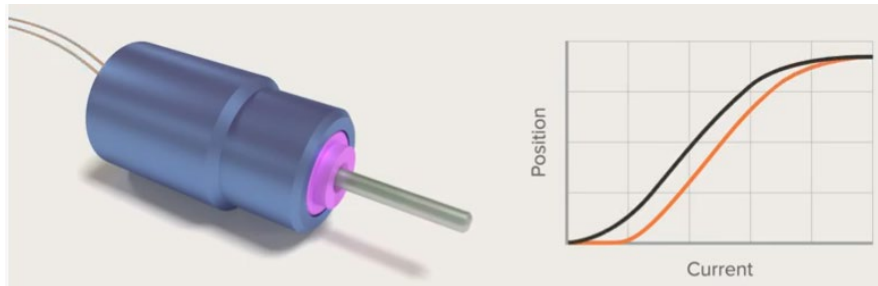


Figura 5 Evaluación del tiempo de reacción y conmutación.

Fuente: Imagen adaptada de (TLXTechnologies, 2020).

La fuerza de retracción se refiere a la cantidad física necesaria para desplazar La Unión de pin magneto de la posición ping extendido hasta la posición de pin retraído. Esta característica se relaciona a la fuerza de atracción entre la guía y el magneto, la cual es necesaria vencer para desplazar nuevamente el magneto a su posición original.

El no cumplir con dicha característica implica que durante la interacción del pin con un dispositivo externo el actuador no pueda repetir su función y por ende inhabilitando la función dispositivo externo.

2.1.4. Proceso de ensamble de un actuador magnético

El ensamble del actuador magnético se realiza en una línea de producción totalmente automatizada, lo cual implica que los procesos de ensamble son realizados por elementos autónomos "robots", sin la interacción de un elemento

humano. Cada uno de los procesos es realizado en una estructura cerrada, la cual se denomina celda de ensamble o simplemente "celda".

En general, se reconocen en el ensamble de un actuador magnético los siguientes procesos:

- Embobinado
- Soldadura por contacto
- Sellado
- Ensamble por prensa

En la Figura 6 se muestra el total los procesos de ensamble de un actuador magnético, en el cual se realizan procesos de ensamble en las primeras 5 celdas y al final se realiza el proceso de evaluación de las características críticas.

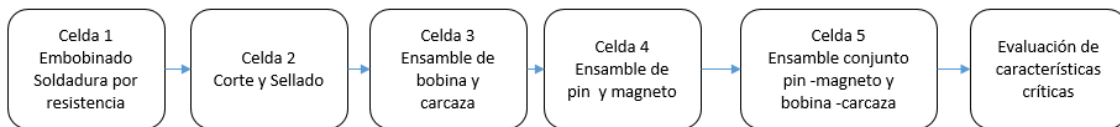


Figura 6. Diagrama proceso de ensamble de actuador magnético.

Fuente: elaboración propia.

De acuerdo con Jürgen (Hagedorn et al., 2018) el proceso de embobinado se realiza cuando se enrolla un material conductor alrededor de una armadura, formando devanados con diferentes configuraciones. Existen diferentes usos en embobinado, ya sea para motores eléctrico, solenoides, antenas, etc. Como se observa en la Figura 7 el alambre del material conductor es dirigido por medio de una aguja, el proceso de embobinado comienza cuando el conductor se fija en la armadura y esta gira de acuerdo con un patrón de enrollado.

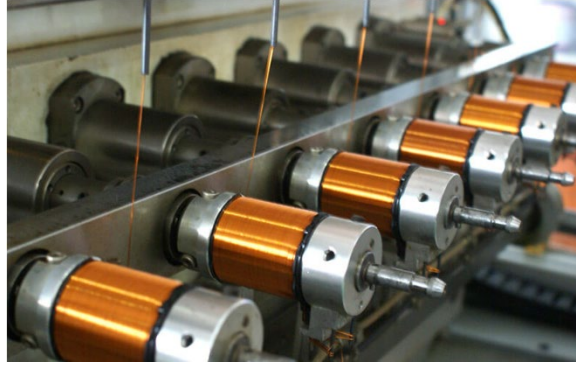


Figura 7. Proceso de Embobinado de un actuador magnético.

Fuente: Imagen adaptada de (KAIDI ELECTROMAGNETIC, 2019).

El proceso de embobinado influye directamente en la magnitud de la influencia electromagnética de los actuadores magnético, por lo cual es de gran importancia que las propiedades del material sean garantizadas, así como garantizar la geometría final del embobinado. El proceso de embobinado es principalmente afectado por parámetros de ajuste como cantidad de bobinas, velocidad de giro en el embobinado y tipo de embobinado. El proceso de embobinado puede afectar a características de funcionamiento de un actuador magnético como resistencia eléctrica, tiempo de conmutación y tiempo de reacción.

El proceso de soldadura por resistencia se lleva a cabo cuando se hace circular una cantidad de corriente a través de dos placas metálicas, las cuales se encuentran en contacto en sus extremos superiores con dos elementos denominados electrodos. Los electrodos se componen de un material con muy buena capacidad de conducción eléctrica. Como se observa en la Figura 8 el proceso de circular corriente a través de los materiales se ve afectado por la resistencia eléctrica de los materiales, por lo cual se genera una zona de calor entre la unión de las placas metálicas y así logrando un cambio de estado en la zona de contacto para posteriormente formar la unión de ambos materiales.

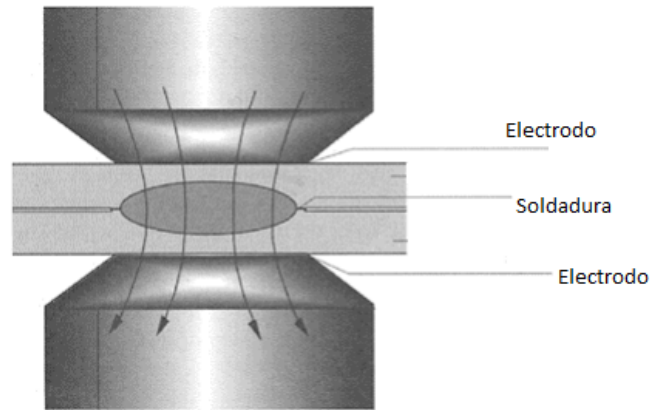


Figura 8 Proceso soldadura por contacto.

Fuente: Imagen adaptada de (Zhang & Senkara, 2011).

El proceso de soldadura por contacto se ve principalmente afectado por la corriente de soldadura, el tiempo de contacto durante la soldadura y la resistencia de los materiales que se desean soldar. El proceso de soldadura por contacto puede afectar a características de funcionamiento como la resistencia eléctrica.

El proceso de dispensado se produce al realizar el dispensado de un material de alta viscosidad sobre una superficie la cual se desea sellar (Scheugenpflug, 2007).

Al realizar el proceso de dispensado sobre la superficie, se espera que el adhesivo en estado líquido sea sometido a un proceso de solidificación en un ambiente controlado, a dicho proceso se le conoce también como curado.

Como se aprecia en la Figura 9 el proceso de dispensado se realiza de forma automatizada, por lo cual se debe de tomar en cuenta que dicho proceso se ve principalmente afectado por la velocidad con la cual se ve dispensado el material, las condiciones ambientales durante el dispensado, el tipo flujo con el cual se dispensa el material, las propiedades del material a dispensar y por último las propiedades del curado del material.



Figura 9 Proceso de dispensado de silicón.

Fuente: Imagen adaptada de (DEMAK, 2021).

Este proceso puede afectar directamente a características de funcionamiento del actuador magnético como fuga interna y en algunos casos tiempo de reacción.

El proceso de unión de dos materiales por medio del formado se basa en la deformación mecánica de ambos materiales. Como se observa en la Figura 10, este proceso se lleva cabo cuando un material externo (material rosa) se desea alojar internamente en un segundo material (material verde) por medio de una fuerza ejercida por una prensa, el material interno realiza un ensanchamiento logrando una zona de interferencia entre ambos materiales. Este proceso se ve afectado por parámetros como presión de ensamble, carrera de la prensa (longitud entre punto de inicio y punto final).

El proceso de ensamble por prensa afecta directamente a variables de funcionamiento como altura de pin extendido, altura de pin retraído, profundidad de bobina y altura de ancla.

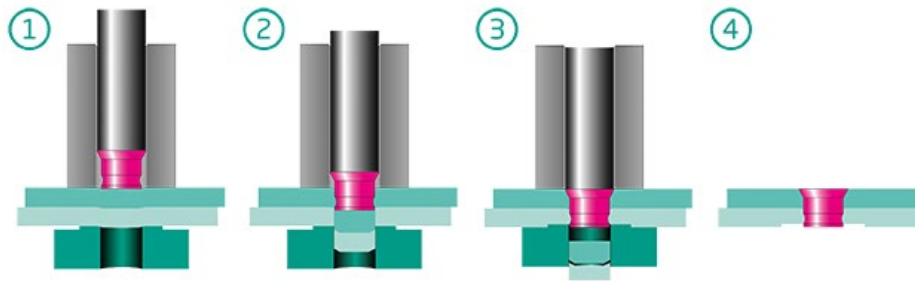


Figura 10 Proceso de prensado.

Fuente: Imagen adaptada de (TOX PRESSOTECHNIK, 2020).

2.2. CIENCIA DE DATOS

La ciencia de datos se define como un área que integra diferentes ramas de estudio tales como la programación, la estadística y los métodos exploratorios de datos (ORACLE, 2020). Esta se puede clasificar como un subconjunto de métodos en el área de Inteligencia artificial debido a que engloba diferentes técnicas con el propósito de encontrar el significado y obtener conocimiento a partir de los datos. La importancia de los datos radica en que estos almacenan información y conocimiento, los cuales mediante técnicas y métodos pueden ser descritos de forma cualitativa o cuantitativa, por ende, existe la posibilidad de obtener señales que lleven a la toma de decisiones o a la distribución de información dentro de los grupos de interés (García et al., 2018).

El uso de estas técnicas ha llevado a métodos para la predicción de eventos basado en modelos entrenados con información previamente obtenida, siendo el objetivo reducir el error de las predicciones máximo posible. Existen sistemas creados para el análisis y predicción de posibles fallos en sistemas de manufactura, los cuales tienen como objetivo asegurar funcionamiento adecuado del proceso de fabricación, en la Figura 11 se observa el sistema de Watchdog para aseguramiento de sistemas de manufactura.

En la Figura 12 se muestran el listado de los métodos, enfoques y pasos para el desarrollo de análisis por minería de datos.

DM Methodology	Pre-Processing	Main-Processing	Post-Processing
CRISP-DM <i>Cross-Industry Standard Process for Data Mining</i>	Business Understanding Data Understanding Data Preparation	Model	Evaluation, Deployment
KDD <i>Knowledge Discovery in Database</i>	Selection, Pre-Processing Transformation	Data Mining	Interpretation and Evaluation
SEMMA <i>Sampling, Exploration, Modification, Model, Verification</i>	Sample, Explore, Modify	Model	Assess
VC-DM <i>Virtuous Cycle of Data Mining</i>	Identify	Transform (Pre-Processing and Main-Processing)	Act, Measure

Figura 12 Esquema del Proceso KDD.

Fuente: Imagen adaptada de (Azevedo & Santos, 2008).

2.2.2. Proceso de minería de datos

Los métodos para realizar minería de datos están basados en la extracción de forma sistemática de la información, de la cual se carecer de conocimiento previo, sin embargo, con un potencial uso. A continuación, se listan los pasos a seguir para realizar el proceso de minería de datos:

- Entendimiento del negocio
- Entendimiento de los datos
- Modelado

De acuerdo con Espinosa (Espinosa Zúñiga, 2020) la fase de entendimiento del negocio es una de las etapas clave para el correcto inicio del análisis, durante esta etapa se define y describe el problema existente en el negocio, con el objetivo de entender el tipo de análisis que se requerirá (el tipo de predicciones que se pretende realizar). Complementando la información anterior, Amat (Amat, 2016) menciona que es necesario definir los datos disponibles y que involucran al proceso que se intenta optimizar y realizar una estrategia para evaluar el modelo que se desea optimizar. De igual forma se explican cuáles son las variables que existen alrededor del indicador que se quiere optimizar.

De acuerdo con Amat (Amat, 2016) antes de realizar el proceso de modelado es necesario considerar los siguientes puntos durante el análisis exploratorio:

Almacenaje de valores incorrectos en la base de datos, ya sea texto almacenado en una posición numérica o viceversa (Amat, 2016). Estos valores pueden ser generados debido a diferentes causas, durante el proceso de lectura y/o escritura de información (interrupción de proceso de envío de datos, dispositivos de lectura dañados, etc.).

Valores no almacenados o valores perdidos, dichos valores pueden ser generados y/o almacenados por las mismas causas que el caso anterior, adicional también se pueden originar debido a un problema con el sistema de medición (calibración de sistema, influencias de condiciones climáticas, etc.).

El tipo de variables con los cuales se cuenta en la base de datos ya sea una variable continua o numérica la cual se expresa como una cantidad numérica o una variable categórica, la cual está sujeta a una clasificación.

Por último, definir dentro de la base de datos las variables que son independientes que formarán parte de las variables predictoras para el modelo de regresión y las variables dependientes, las cuales serán la respuesta, y por lo tanto la variable a evaluar dentro del modelo de regresión.

Por último, se realiza un análisis estadístico de la base de datos para confirmar que no existan valores atípicos. Se pueden utilizar indicadores estadísticos como el cálculo de la media, la desviación estándar, conteo de número de observaciones, encontrar los valores máximo y mínimo de la base de datos.

2.2.3. Métodos de modelado y aprendizaje de máquina

El aprendizaje de máquina se refiere a uso de algoritmos y técnicas que nos ayudan a identificar patrones dentro de la base de datos y permiten la generación de modelos que describen dichos patrones. Al emplear sistemas entrenados con una base de datos pasada, se pretende realizar la predicción de nuevos eventos basados el reconocimiento de escenarios previamente descritos (Amat, 2016).

Dentro de la clasificación básica de los métodos de aprendizaje de máquina se pueden mencionar los siguientes tipos: métodos de aprendizaje supervisado, métodos de aprendizaje no supervisado y por último aprendizaje reforzado.

Los métodos de aprendizaje supervisado se refieren al conjunto de técnicas, con la cualidad de generar modelos que predicen el resultado de salida en base a ejemplos previos de esa misma salida.

Por otro lado, los modelos no supervisados se refieren a los algoritmos que pueden clasificar los datos de acuerdo a un conjunto de características. Los algoritmos no supervisados se caracterizan por tener la capacidad de detectar posibles anomalías dentro de las tareas.

Entre las ventajas del uso de métodos supervisados, se menciona la facilidad con la cual pueden ayudar a definir una ruta en la transformación de los datos en información que se procese.

Dentro de los métodos y técnicas de análisis utilizados en la minería de datos, se encuentra la descripción, clasificación, regresión, agrupamiento (clustering), y asociación. Estas técnicas son compartidas dentro de las técnicas utilizadas en el aprendizaje de máquina.

A continuación, se muestra la Figura 13 con el diagrama de los métodos utilizados para la minería de datos, los cuales se clasifican a través de métodos descriptivos, clasificación, regresión, clustering o agrupación y por último, métodos de asociación. Cada uno de los métodos involucra diferentes técnicas para su desarrollo y aunque sus requerimientos en muchos casos son similares, se debe tener en claro cuáles son los tipos de datos que se pueden obtener del proceso para determinar el método que se adecúe mejor.

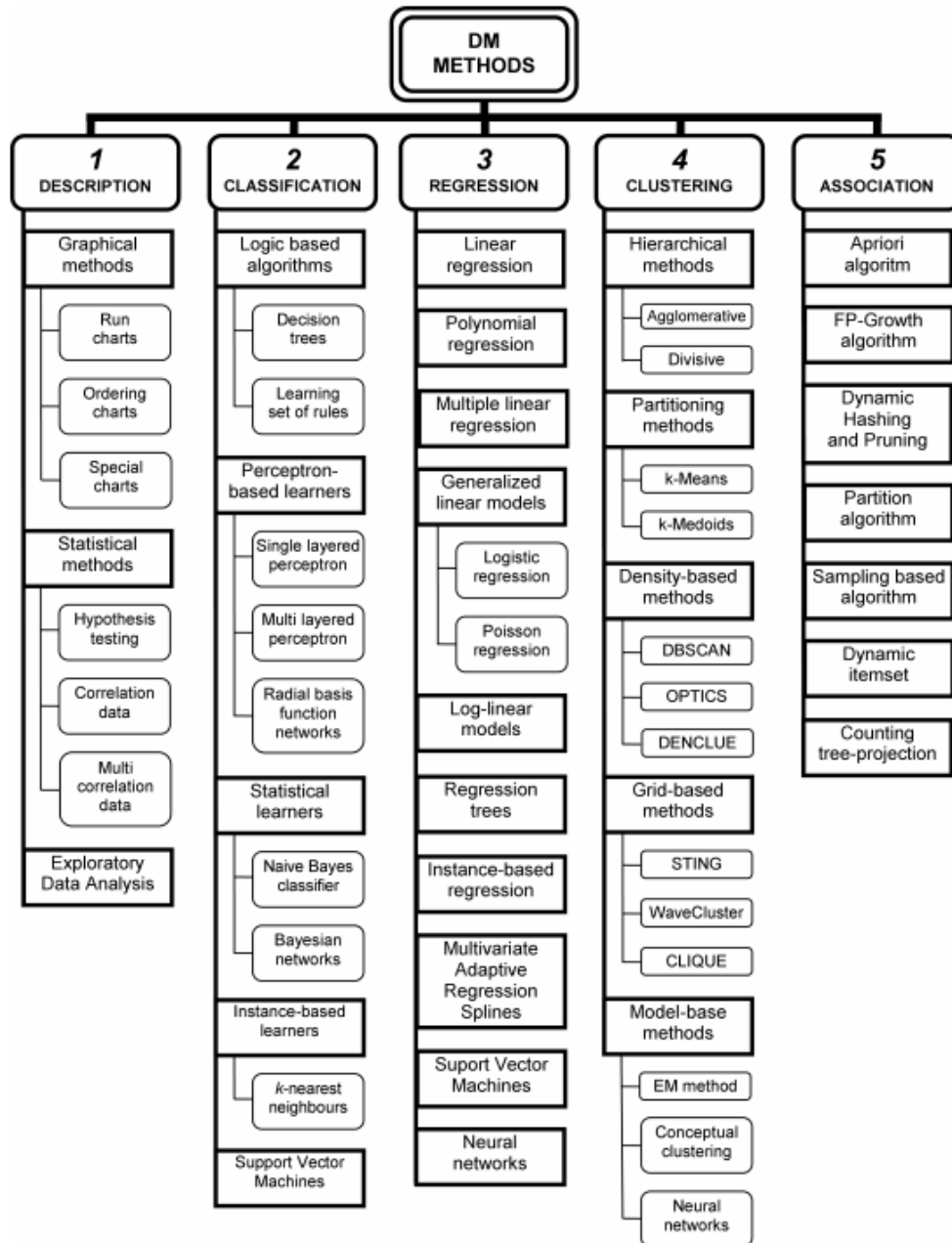


Figura 13 Métodos y técnicas dentro de la minería de datos.

Fuente: Imagen adaptada de (Rogalewicz & Sika, 2016).

Para realizar la selección del método estadístico, se puede utilizar la guía propuesta por Byrne (Byrne & Wilkins, 1998), para selección de método óptimo de análisis de datos. En Figura 14, se muestra la Tabla a partir de la cual se selecciona el método de análisis partiendo de las características de la variable que se intenta predecir.

Flow-chart for popularly used statistical tests

Q1, Univariate /Multivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)	Q 6 , No. of groups	Q 7 ,sample size	Valid Tests	
Univariate	Difference	Independent (un-paired)	Continuous (Normal)	2		Student's t-test	
				> 2		One-way ANOVA	
			Continuous (Non-normal) / Ordered categorical	2		Mann-Whitney U test	
				> 2		Kruskal-Wallis H test	
			Nominal	2	<20	Fisher's exact test	
			≥2	≥20	Chi-square test		
			Time to Event			Log-Rank test (Kaplan-Meier plot)	
	Dependent (paired)	Continuous (Normal)	2			Paied-t test	
			> 2			Repeated measured ANOVA Mixed effect Regression	
		Continuous (Non-normal) / Ordered categorical	2			Wilcoxon signed-rank test	
				> 2		Friedman test	
				2		McNemar's test	
Correlation	Continuous (Normal)				Pearson's correlation (r)		
	Continuous (Non-normal) /ordered				Spearman's correlation (rs)		
	Nominal (2 levels)		2		Spearman/Kappa (Agreement)		
Multivariable	Independent (un-paired)	Continuous (Normal residuals)				Linear Regression	
		Continuous (Non-normal residuals)				Linear Regression *	
		Ordered categorical				Ordered Logistic Regression	
		Nominal	(2 levels)				Binary Logistic Regression
			(> 2)				Multinomial Logistic Regression
		Time to Event				Cox Proportional Hazard Regression	
	Dependent (paired)	Continuous (Normal residuals)				Linear Mixed Effect Regression	
		Continuous (Non-normal residuals)				Linear Mixed Effect Regression *	
		Ordered categorical				Generalized Estimation Equation (GEE)	
		Nominal (2 levels)				Generalized Estimation Equation (GEE)	

Figura 14 Selección de técnica estadística.

Fuente: Imagen adaptada de (Byrne & Wilkins, 1998)

2.2.4. Correlación lineal de variables

La correlación lineal es un método estadístico a partir del cual se puede cuantificar el grado de relación lineal entre dos variables (Amat, 2016). Un coeficiente de correlación establece que el grado de asociación, sin embargo, este no se puede interpretar como la dependencia causal entre las variables de interés (Restrepo & González, 2007).

Existen diferentes indicadores que permiten explicar la relación entre las variables, entre los cuales se puede mencionar la Correlación de Pearson, Spearman y de Jackson(Hernández, 2018). De acuerdo con Amat (Amat, 2016) las principales características entre los tres indicadores son:

Correlación de Pearson se aplica para el estudio de variables continuas, las cuales se explica por medio de una distribución paramétrica de tipo normal y es susceptible al efecto en valores cercanos a los extremos.

Correlación de Spearman se utiliza para variables agrupadas para dentro de un rango o cuando la distribución de los valores no se puede parametrizar por medio de una distribución normal.

Correlación de Kendall es también utilizada cuando la distribución de los datos no se parametriza por medio de un comportamiento normal. Cumple también con la característica de poder explicar la relación entre variables cuando la cantidad de observaciones no es limitada.

Desde el punto de vista de Amat (Amat, 2016) el uso del coeficiente de correlación de Pearson debe cumplir con las condiciones de para dicho análisis se deberán cumplir con las condiciones de linealidad entre las variables a describir, distribución normal en los valores observados, valores se explican adelante, ya que se comparten con los requerimientos para generar un modelo de regresión.

El coeficiente de correlación de Pearson fluctúa entre $-1 \leq r \leq 1$, de aquí se deduce que los valores cercanos a los extremos explican una alta asociación, siendo una asociación inversamente proporcional en el caso de valores cercanos a -1 y directamente proporcional en caso de valores cercanos a 1.

Por último, de acuerdo con Restrepo y Gonzáles (Restrepo & González, 2007) se recomienda que dicho valor no sea extrapolado fuera de los rangos de los valores observados, cuando se efectúa la inferencia.

2.2.5. Modelo de regresión

De acuerdo con Montgomery (Montgomery, 2004), los modelos de regresión se enfocan en el ajuste de los modelos empíricos. El modelo de regresión lineal, describe la relación entre una variable de respuesta o variable dependiente (Y) y una o varias variables independientes o variables predictoras ($X_1, X_2, X_3, \dots, X_n$), a través de los parámetros estimados para cada una de las variables predictoras los cuales son llamados coeficientes de regresión parciales ($B_0, B_1, B_2, \dots, B_n$). El parámetro de B_0 se conoce como el punto de intersección entre los planos de los ejes ordenados. Existen diferentes métodos de regresión para calcular los parámetros del modelo, entre los cuales se puede mencionar mínimos cuadrados, regresión de Ridge, mínimos cuadrados parciales, entre otros. Este estudio se

enfoca en el análisis a través del método de mínimos cuadrados para regresiones lineales.

Dentro de las condiciones necesarias para realizar un modelo de regresión se mencionan:

- No colinealidad entre variables independientes
- Relación lineal entre variable predictora y variable de interés
- Distribución normal de la variable
- Variabilidad constante de los residuos (homocedasticidad)
- No auto correlación
- Valores atípicos, con alto leverage o influyentes
- Parsimonia

2.2.6. Selección de predictores para modelo óptimo

Para garantizar que el modelo de regresión cuenta con la proporción correcta de datos para el entrenamiento y validación, existen métodos de validación de la muestra, entre los cuales se pueden mencionar validación simple, validación Leave One Out Cross Validation, K-Fold Cross Validation, entre otros. El método de Validación Simple consiste en dividir la base de datos en Entrenamiento – Test (Ej 80%/20%). Por su parte el método de validación Leave One Out Cross Validation (LOOCV) utiliza $n-1$ datos para entrenar el modelo y el último valor lo utiliza para confirmar el modelo (Browne, 2000).

El proceso se repite tantas veces como observaciones disponibles. Este método tiene la desventaja de requerir un alto costo computacional debido que realiza un proceso de forma iterativa, sin embargo, tiene como ventaja el reducir la variabilidad entre las validaciones. El método de K-Fold Cross – Validation, consiste en dividir los datos de forma aleatoria en “k” grupos de aproximadamente el mismo tamaño k- grupo se emplean para entrenar el modelo. El proceso genera k estimaciones del error cuyo promedio se emplea como estimación final (Amat, 2016), con esto se consigue un balance entre bias y varianza dependiendo del número de subgrupos generados (se recomienda entre 5-10 grupos).

De acuerdo Montgomery (Montgomery, 2004), un modelo de regresión debe de cumplir con los siguientes requerimientos:

- No colinealidad entre variables
- Relación lineal entre variables predictoras y variable de interés
- Variabilidad constante de los residuos (homocedasticidad)
- No auto correlación
- Distribución normal en la variable de respuesta
- Parsimonia en el modelo
- No existan valores atípicos

Al realizar un modelo de regresión por método de mínimos cuadrados para explicar la relación de varias variables, se debe confirmar que los predictores no estén relacionados entre ellos, dicho comportamiento es conocido como la no colinealidad entre variables. La no colinealidad en los variables predictoras se puede interpretar cuando dos variables presentan alto grado de correlación entre ellas, lo cual establece y confirma que la relación que se establece con la variable de interés no pueda ser directamente una contribución individual por parte de la variable predictora, sino que exista una contribución asociada a ambas variables.

El efecto de la colinealidad existe cuando los predictores se relacionan de forma lineal, este representa un problema, debido a que complica el proceso para cuantificar la aportación individual que realiza cada variable predictora a la variable de interés. Para determinar si existe colinealidad entre dichas variables, se sugiere realizar el proceso de cuantificación de correlación entre las variables, con lo cual si se observa un valor de correlación cercano a uno se puede confirmar que existe colinealidad perfecta o casi perfecta (Amat, 2016). Como resultado, se propone que si existe un coeficiente de determinación R^2 alto, esto señala que existe una colinealidad.

Al visualizar la interacción entre las variables predictoras y la variable de respuesta, se debe observar que ambos estén linealmente relacionados. Sin embargo, no se debe ver que el comportamiento de los demás predictores tenga variación.

Para confirmar la relación entre la variable predictora y su variable de interés, se puede realizar la visualización de dichas variables en una gráfica lineal o una gráfica con la dispersión de los datos, con lo cual se puede confirmar el uso de un método de regresión lineal.

Se propone para tal efecto visualizar los residuos en cada uno de los predictores. Por tal motivo, una relación lineal debe de distribuir los valores residuales de forma aleatoria respecto a 0. A continuación, se presenta la figura 15 con la representación de las gráficas de una relación lineal y no lineales (cuadrático y cúbico).

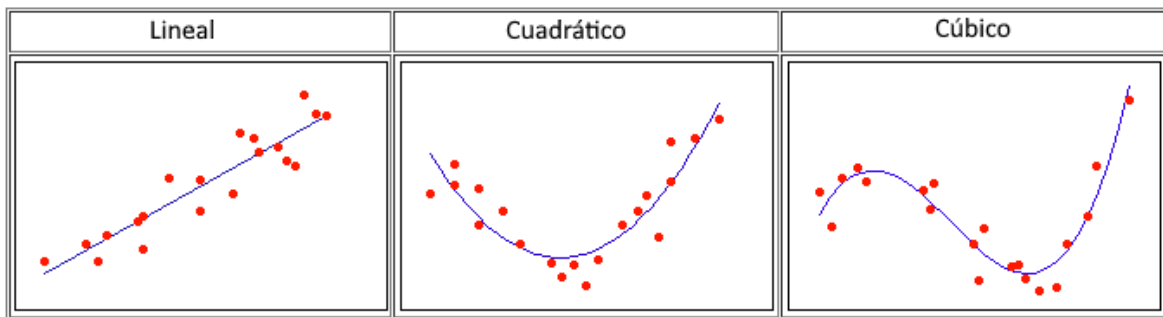


Figura 15 Gráfica de dispersión con regresión lineal y no lineales.

Fuente: imagen adaptada de (Addlink Software Científico, 2020).

Para establecer la referencia y confirmar que los datos de las observaciones obtenidas de la variable de respuesta mantienen una distribución normal se puede recurrir a un método gráfico (histograma, diagrama Q-Q) a realizar una prueba de hipótesis (ejemplo método de Shapiro Whiks).

Al realizar la distribución de los residuos, estos deben de cumplir con una distribución normal, donde su valor de media debe de ser igual a 0. Existen varias formas de comprobar que los valores de residuos cumplan con esta condición, ya sea con histogramas, Pruebas de hipótesis o gráficos de cuantiles.

Debido a que las condiciones del modelo se comprueban con los valores residuales, se propone en la literatura realizar un proceso iterativo, el cual realiza un ajuste de los predictores, tomando en cuenta los valores de residuos que se obtienen en la prueba. Al obtener los valores de residuos con mejor resultado, se

garantiza que se obtuvieron los valores de los coeficientes óptimos del modelo se pueden encontrar.

El valor de la varianza a lo largo de todos los valores residuales observados debe ser constante, esto se propone visualizar con una gráfica de residuos. La visualización de los datos debe de mantenerse de forma aleatoria y con un grado de variación. Otro método descrito es el de realizar la prueba de hipótesis Breusch-Pagan.

La auto correlación indica que los datos medidos en el proceso de fabricación fueron realizados de forma individual e independiente, con lo cual los valores no se ven relacionados.

La presencia de valores atípicos en las bases de datos afecta e influyen el resultado del proceso de modelado, por este motivo dichos valores deben ser identificados y tratados durante el proceso de modelado y de esta forma garantizar que el modelo pueda explicar el comportamiento del proceso en forma estable.

La Parsimonia hace referencia al que el modelo cuenta con las variables necesarias para lograr el óptimo desempeño durante el proceso de predicción.

Esta característica se puede explicar mediante el cálculo de los indicadores como coeficiente de correlación o también llamado R^2 , puesto que dicho valor indica con que eficiencia el modelo capaz de explicar las variables de respuesta.

2.2.7. Uso de software y lenguajes de programación

El uso de software para el desarrollo de técnicas de aprendizaje de máquina está ligado a los lenguajes de programación, como se muestra en la Figura 16 entre los lenguajes de programación más populares se encuentran Python, R, C++, Matlab, entre otros. En el desarrollo de este estudio se optó por el uso del lenguaje de programación Python, a través del Software Anaconda y Google Colab.

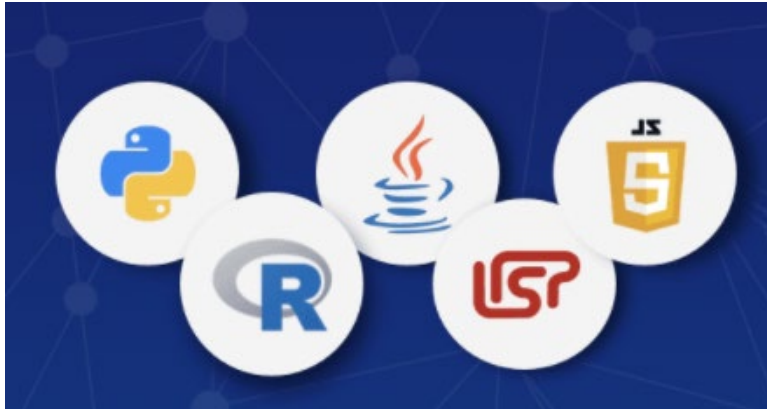


Figura 16 Lenguajes de programación para Machine Learning.

Fuente: Imagen adaptada de (Duomly, 2019).

De acuerdo con Matloff (Iacobucci & Robert, 2012), las ventajas del uso de la programación (como es el caso del lenguaje R) en el análisis de datos ofrece ventajas como:

- Rápida y fácil implementación del lenguaje entre estadistas profesionales
- Disponible en sistemas operativos más comunes (Windows, Linux, Mac)
- En su mayoría lenguaje de apertura libre, lo cual facilita su acceso a nuevos métodos y funciones de análisis
- Automatización de análisis y creación de nuevas funciones basado en características existentes

3.3. OPTIMIZACIÓN DE PROCESOS

Los procesos de fabricación forman parte de un proyecto con un ciclo de vida definido, este ciclo de vida está conformado por diferentes etapas antes y el inicio de producción (Kane, 1986).

Este proceso comienza durante un periodo de evaluaciones para conocer las limitaciones naturales del proceso y con el objetivo de definir los límites y especificaciones para la fabricación del producto.

El análisis del desempeño de proceso es de vital importancia, de aquí se confirma que el proceso garantiza las condiciones de las C.C. y que cada uno de los procesos de ensamble fueron realizados dentro de las especificaciones definidas.

Por otro, se reconoce que todos los procesos cuentan con la generación de material con C.C. fuera de la especificación, el cual no puede ser utilizado y por ende es desechado, a este material se le conoce como "Scrap".

3.3.1. Cantidad de material de rechazo (% de Scrap)

El % de material rechazado es uno de los elementos del costo derivado de la falta de calidad en los procesos, este puede llegar a tener un alto impacto en los costos de proceso (Green et al., 2019).

La cantidad de material rechazado que se genera afecta la rentabilidad del proceso, ya que este influye directamente en los costos de operación, ya sea en la demanda de materia prima, logística en movimientos de material e incluso implicaciones ecológicas, entre otras. Es por eso que el cálculo de este indicador es clave en su evaluación.

El cálculo de este indicador puede ser calculado de diferentes formas, ya sea basado en el costo de la materia prima o en los costos de los procesos agregados al material. En nuestro caso el cálculo del % de material rechazado será realizado a través de la relación entre el Número de Piezas de Rechazo y la suma del total de piezas fabricadas.

Una de las formas para analizar las posibles causas de material rechazado es mediante la elaboración de un diagrama causa – efecto o también llamado diagrama de Ishikawa.

En la Figura 17 se muestra un ejemplo con el diagrama de Ishikawa, donde se establecen las posibles causas que estén relacionadas al efecto directo de la causa de material rechazado. En la práctica para guiar la elaboración del mismo se inicia clasificando las posibles causas como problemas ligados a la medición, material, personal, medio ambiente, método de trabajo y por último a la máquina.

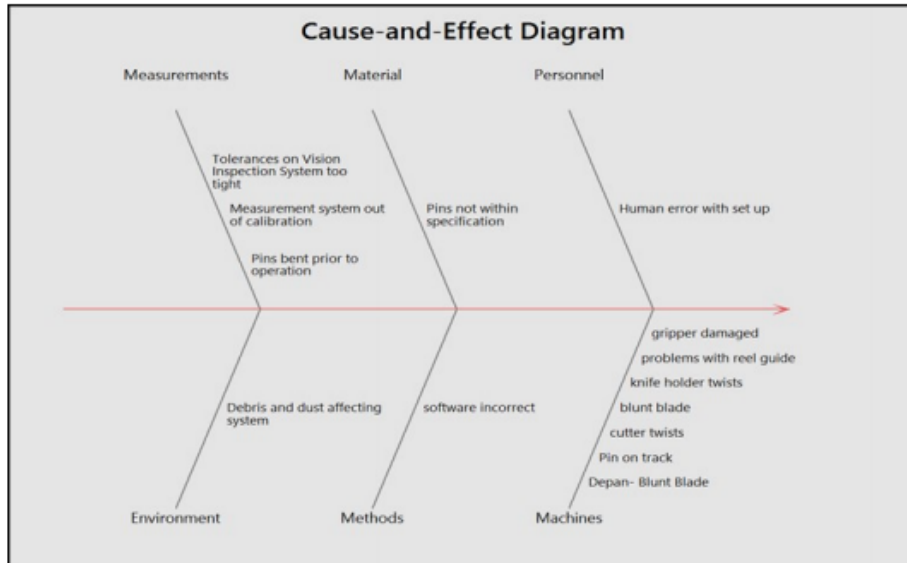


Figura 17 Diagrama causa efecto.

Fuente: Imagen adaptada de (Green et al., 2019).

3.3.2. Índices de capacidad (Uso de cpk y ppk)

Los índices de capacidad están ligados a los valores límite especificados para la característica continua, medida en la fabricación de un producto. La cual se define como la relación entre la distancia desde el punto donde el proceso se encuentra centrado hasta los límites de especificación dividido entre la medida de dispersión entre las observaciones realizadas.

Esto se puede representar de forma gráfica con el uso de un histograma, como se aprecia en la figura 18 De acuerdo a la distribución de los valores del proceso. En esta Figura se puede observar como los valores medidos para la una variable de interés son distribuidos en el Eje X del histograma y la ocurrencia de cada valor es representada en el eje Y. El proceso se encuentra bajo la condición de un comportamiento normal, lo que significa que el valor promedio de las observaciones se encuentra centrado y por lo tanto se espera que la distribución de los valores tenga una desviación de $\pm 3\sigma$ de acuerdo a la referencia del promedio.

Por último, para localizar el comportamiento de la variable, los valores se comparan respecto a la tolerancia mínima y máxima de la variable de interés, con lo cual se

puede saber si el proceso cumple la desviación requerida y si esta se encuentra dentro de los límites especificados.

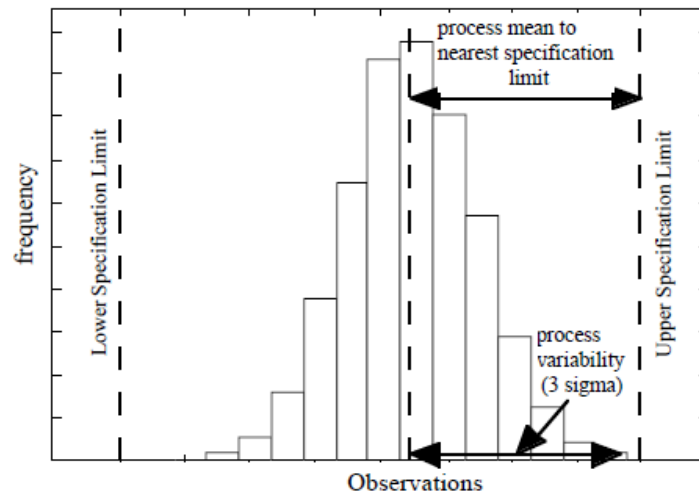


Figura 18 Representación del índice del proceso.

Fuente: Imagen aptada de(Steiner & Mackay, 1998)

3. PROCEDIMIENTO

El desarrollo del trabajo se estableció con base al método de minería de datos CRISP-DM el cual consta de 5 fases:

La primera fase realizó el proceso de entendimiento del negocio, utilizando la base de datos de 2019, se relacionaron los indicadores clave como habilidad de proceso y porcentaje de material rechazado con las características críticas de funcionamiento de un actuador magnético. Por otra parte, se realizó la descripción del proceso de generación de datos en el ensamble de un actuador magnético, así como las bases de datos disponibles dentro del proceso mismo. Se definió la cantidad de bases de datos disponibles, así como el método de recolección, la frecuencia de recolección y el almacenamiento de los datos. Por último, el objetivo y la expectativa de resultados del modelo de regresión.

En la segunda fase se realizó el análisis para lograr un entendimiento de los datos disponibles, en este proceso se clasificaron las variables como continuas y discretas, esto con el objetivo de utilizar sólo las variables continuas, las cuales podían aportar información valiosa para el modelo de regresión. Por último, se clasificaron las variables continuas como variables predictoras o variables de interés.

La tercera fase comprendió el proceso de preparación de los datos para generar el modelo de regresión con las variables necesarias. Para realizar el proceso de preparación de datos, se utilizó un algoritmo basado en el lenguaje Python y con el uso paqueterías ya existentes para el manejo de la información. Este algoritmo permitió el analizar y buscar datos inconsistentes o faltantes, eliminar variables discretas y por último realizar el cálculo de las estadísticas básicas una vez que la información fue filtrada.

La cuarta fase, comprendió la generación de los modelos de correlación entre las variables predictoras y las variables de interés. Utilizando el método de Pearson se generaron mapas de calor para visualizar la influencia entre ambas variables, de esta forma se realizó el primer modelo de regresión con las variables de mayor influencia sobre la variable de interés. este proceso fue repetido añadiendo nuevas

variables que pudieran aportar un mejor valor de predicción al modelo de regresión.

La Quinta y última fase realizó la evaluación y despliegue del modelo de regresión, con el objetivo de confirmar si dicho modelo cumplía las condiciones iniciales del modelo de regresión. Lo anterior se realizó a través del análisis de residuos entre los valores reales y de predicción. Una vez concluida la evaluación, se realizó el despliegue del modelo para optimizar el proceso de ensamble, lo anterior por medio de propuestas de ajuste en el proceso y con el objetivo de alcanzar los valores en las variables de proceso (variables predictoras) que permitan disminuir la variación en las características críticas de funcionamiento.

A continuación, se muestra en la figura 19 las distintas fases del proceso de minería de datos con el método CRISP-DM

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background</i> <i>Business Objectives</i> <i>Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i> Describe Data <i>Data Description Report</i>	<i>Data Set</i> <i>Data Set Description</i> Select Data <i>Rationale for Inclusion/Exclusion</i>	Select Modeling Technique <i>Modeling Technique</i> <i>Modeling Assumptions</i> Generate Test Design <i>Test Design</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria</i> <i>Approved Models</i>	Plan Deployment <i>Deployment Plan</i> Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Assess Situation <i>Inventory of Resources</i> <i>Requirements, Assumptions, and Constraints</i> <i>Risks and Contingencies</i> <i>Terminology</i> <i>Costs and Benefits</i>	Explore Data <i>Data Exploration Report</i> Verify Data Quality <i>Data Quality Report</i>	Clean Data <i>Data Cleaning Report</i> Construct Data <i>Derived Attributes</i> <i>Generated Records</i>	Build Model <i>Parameter Settings</i> <i>Models</i> <i>Model Description</i> Assess Model <i>Model Assessment</i> <i>Revised Parameters</i>	Review Process <i>Review of Process</i> Determine Next Steps <i>List of Possible Actions</i> <i>Decision</i>	Produce Final Report <i>Final Report</i> <i>Final Presentation</i> Review Project <i>Experience</i> <i>Documentation</i>

Figura 19 Implementación método CRISP-DM.

Fuente: Imagen adaptada de (Wirth & Hipp, 2000).

3.1. ENTENDIMIENTO DEL NEGOCIO

Para comenzar la fase de entendimiento del negocio, se partió del planteamiento del problema, debido a que se pretende conocer la causa por lo cual los valores entre observaciones de cada característica crítica se encontraron dispersos y con valores alejados del valor promedio.

A continuación, se presenta la Tabla 1, donde se monitoreo el valor de la desviación estándar poblacional para determinar la influencia sobre la habilidad de proceso y el porcentaje de material rechazado.

Resistencia_Electrica												
Indicador/ mes	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
% de material rechazado	4.28%	6.66%	2.69%	2.69%	11.44%	4.14%	3.59%	2.19%	2.98%	0.65%	0.66%	0.74%
cpk	0.74981522	0.7022	1.0141	1.9378	1.0382	2.018	1.4467	1.4418	1.5134	1.2872	2.6692	4.6043
Desviación Estándar	0.04001558	0.0682	0.0702	0.0416	0.071	0.0579	0.0683	0.0907	0.0637	0.0812	0.0415	0.021

Tabla 1 Comparativa desviación estándar, habilidad de proceso cpk y % material rechazado.

Fuente: elaboración propia.

En esta Tabla se identificó que la variación en la desviación estándar durante el año 2019 generó un aumento en la producción de material de rechazo. Por tanto, el porcentaje total producido anual fue de 3,56% contra el objetivo del 2%, lo cual confirmó que existió una necesidad de disminuir la cantidad de material de rechazo en la línea de producción.

De aquí se planteó la propuesta de predecir las condiciones de fabricación que permitieran reducir la desviación estándar en las características críticas y por ende una optimización en los indicadores clave. De acuerdo a lo anterior, se establece que el procedimiento debe enfocarse en encontrar las causas que generan la dispersión en los valores de las características críticas.

Una vez planteado el objetivo de disminuir la desviación estándar con un valor menor a 0.04, se revisaron las bases de datos disponibles, de las cuales se obtendrían los datos para generar el modelo de regresión.

La línea de ensamble de actuadores magnéticos cuenta con 5 bases de datos provenientes de cada una de las celdas de ensamble, dichas bases de datos almacenan los valores obtenidos de las mediciones del proceso de forma local como se muestra en la Figura 20.

Por último, se cuenta con una base de datos en la celda de evaluación final, dicha base de datos almacena los valores de medición de las características críticas de funcionamiento y los valores correspondientes a la medición de algunos procesos que son clasificados como críticos.

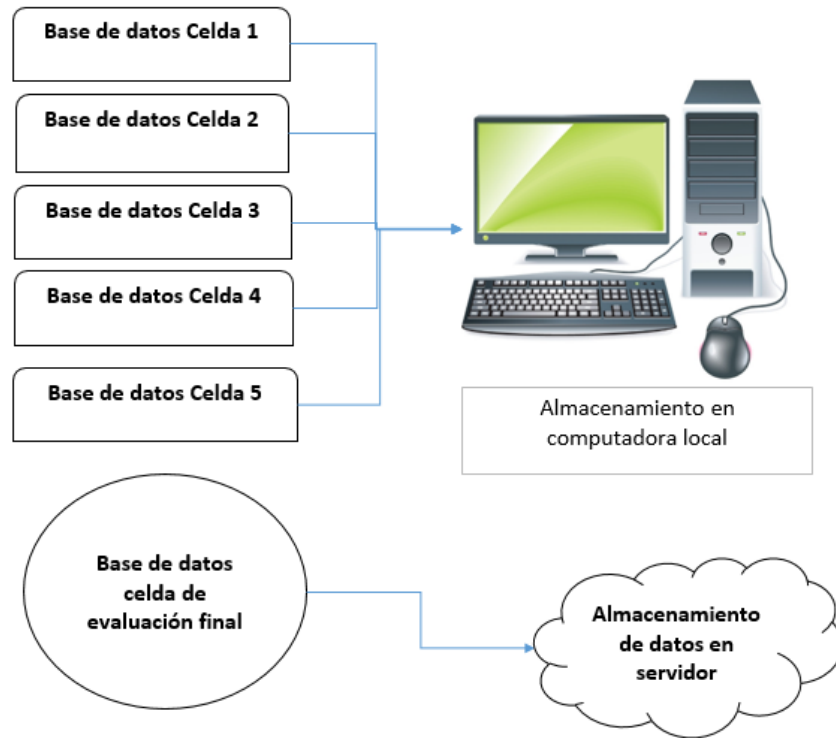


Figura 20 Bases de datos disponibles y almacenamiento en línea de ensamble de actuadores magnéticos.

Fuente: elaboración propia.

Debido a la anterior inicialmente se contaron con 6 bases de datos para analizar, con el objetivo disminuir la cantidad de datos a analizar y por ende la cantidad de recursos necesarios para generar el modelo de regresión, se planteó el objetivo de solo usar los datos necesarios.

De acuerdo a lo anterior, se observó que los resultados obtenidos en las celdas de ensamble se clasificaban como pieza aprobada o no aprobada dentro de la base de datos de la evaluación final. Como se muestra en la Figura 21, la base de datos almacenada en el servidor contenía el resultado de las pruebas de las celdas de ensamble, el cual se clasificaba como Ok ó No Ok.

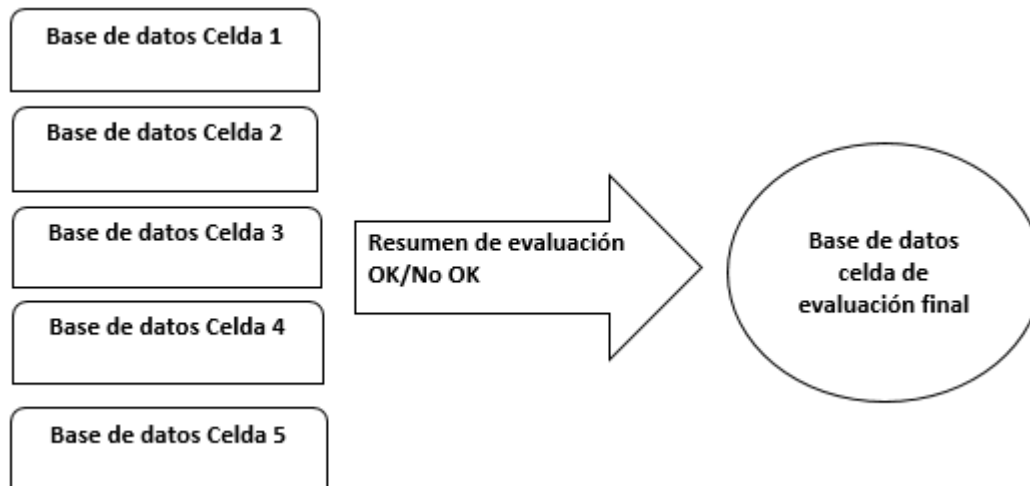


Figura 21 Envío de resultados entre celdas de ensamble.

Fuente: elaboración propia.

Por este motivo, se tomó la decisión de usar sólo la base de datos de la celda de evaluación final y solo usar las características críticas con los parámetros críticos.

- Cálculo de índices de capacidad y % de material rechazado

Para realizar el cálculo dl índice de capacidad se tomó el total de la base de datos (material aprobado + material rechazado), y De acuerdo a la siguiente formula:

$$\text{Process Capability} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right),$$

Donde:

USL: se refiere al límite de especificación superior de la característica medida.

LSL: se refiere al límite de especificación inferior de la característica medida

σ :se refiere a la desviación estándar* evaluada en la base de datos.

Para realizar el cálculo del porcentaje de material rechazado se utilizó el total de base de datos obtenida (material aprobado + material rechazado), De acuerdo a la siguiente formula se realizó el cálculo:

$$\% \text{ de material de rechazo} = \frac{\#Piezas \text{ Rechazadas}}{\#Piezas \text{ buenas producidas} + \#Piezas \text{ Rechazadas}}$$

3.3.1. Equipo de medición utilizado

Para realizar las evaluaciones en cada una de las observaciones, se utilizaron diferentes equipos de medición, los cuales realiza el proceso de medición y comunicación.

El primer proceso se realiza de acuerdo con la característica que se desea inspeccionar, el cual se realiza de acuerdo a un rango previamente establecido y con una resolución de acuerdo a la capacidad del equipo. El segundo proceso realiza la comunicación entre el sistema de medición y el control lógico del equipo (llamado PLC), el cual permite recibir y comparar el valor obtenido del sistema de medición, a partir de esto la pieza puede ser aprobada o rechazada de acuerdo a una tolerancia previamente establecida. Por último, el control envía el valor obtenido a una base de datos ya sea local o externa.

Entre los múltiples equipos utilizados para realizar las mediciones se pueden mencionar sistemas de visión, transductores, sensores basados en fibra de luz y de sensores.

Como se observa en la Figura 22, el sistema de visión compara y mide característica como el diámetro de embobinado, de acuerdo a una imagen de referencia, este sistema depende principalmente del uso de luz, contraste y resolución.

La respuesta obtenida después de la medición de cada observación se clasifica como OK para mediciones aprobadas y No OK para piezas mediciones reprobadas.



Figura 22 Sistema de visión.

Fuente: imagen adaptada de (Cognex, 2021).

El sistema de visión se utilizó en la evaluación de los procesos como embobinado y dispensado.

El uso de sensores fotoeléctricos fue utilizado para medir el tiempo promedio de reacción y el tiempo de conmutación. En la Figura 23 se observa un sensor de luz el cual requiere de una calibración previa donde se ajusta la cantidad de luz inicial y final entre los sensores, De acuerdo al tipo de prueba que se realizará.



Figura 23 Sensor de Luz.

Fuente: Imagen tomada de (Panasonic, 2020)

Este sensor realiza su medición cuando se detecta un cambio en la cantidad de luz entre emisor y receptor. Dicho sensor genera una señal que cuenta el tiempo

trascendido, desde que se realiza el movimiento desde la posición del pin retraído hasta la posición del pin extendido.

A continuación, se muestra en la Figura 24, donde se muestra el proceso de medición con sensores de luz.

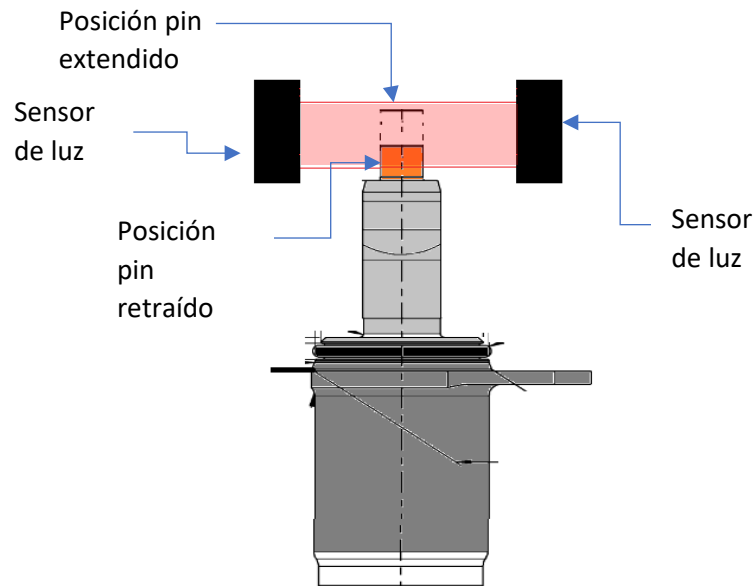


Figura 24 Medición de tiempo de conmutación y reacción con sensores de luz.

Fuente: Imagen elaboración propia.

El resultado de las mediciones con el sensor de luz en cada actuador se expresa en milisegundos.

En la Figura 25 se muestra un transductor de carrera ajustable, el cual se utiliza para realizar la medición de longitud, este transductor cuenta con una resolución de +/- 0.001 mm.



Figura 25 Transductor de resorte.

Fuente: Imagen Adaptada de (Peter Hirt, 2020)

Este transductor fue utilizado durante la medición de altura de pin extendido, altura de pin retraído, profundidad de bobina y altura de ancla, como se muestra en las siguientes imágenes.

En la Figura 26 se muestra el proceso de medición para las características de profundidad de bobina y altura de ancla, el resultado obtenido para cada una de las mediciones con el transductor se expresa en mm.

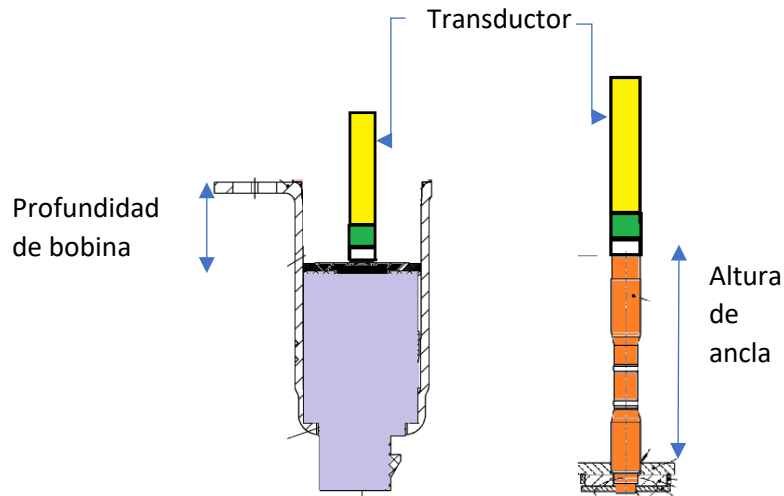


Figura 26 Medición de altura y profundidad con uso de transductor.

Fuente: Imagen elaboración propia.

En la Figura 27 se muestra un sensor para la medición de la fuerza de retracción de pin, el cual se colocó dentro de un dispositivo mecánico para realizar la prueba de fuerza. El resultado de la fuerza ejercida para cada una de las mediciones se expresa en Newton.



Figura 27 Sensor de fuerza.

Fuente: Imagen adaptada de (TOX PRESSOTECHNIK, 2020)

A continuación, se muestra en la Figura 28 el proceso de medición de fuerza de retracción de pin. Para realizar esta medición el sensor de fuerza se encuentra instalado dentro de un dispositivo por medio de actuadores que ejercen una fuerza sobre el pin hasta llevarlo de posición de pin extendido a retraído.

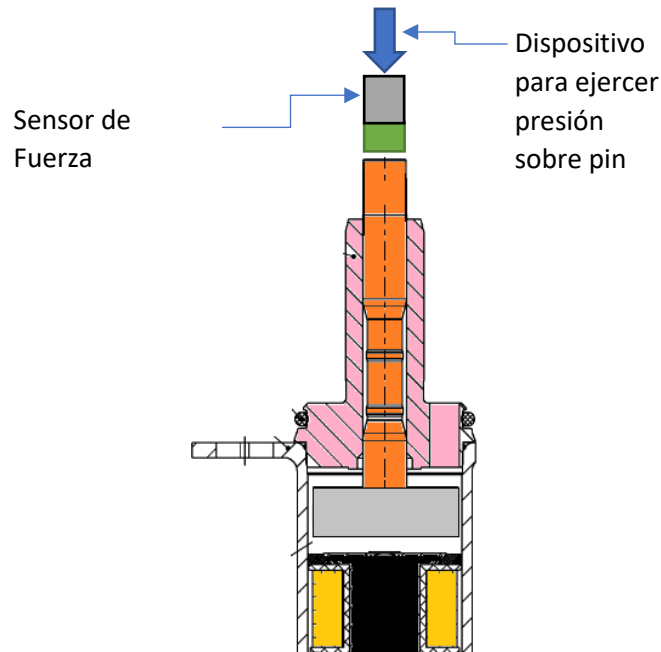


Figura 28 Representación medición de fuerza retracción pin.

Fuente: elaboración propia.

3.2. ESTUDIO Y COMPRESIÓN DE LOS DATOS

Previamente fue necesario realizar un estudio de las bases de datos obtenidos en el proceso de fabricación de un actuador magnético.

De acuerdo a lo anterior, se logró realizar un análisis descriptivo, el cual permitió entender la información contenida dentro de la base datos. El análisis exploratorio es importante ya que a partir del mismo se puede obtener la información de cada variable de observación, así como la temprana detección de valores atípicos. (Amat, 2016).

Para comenzar, se revisó la estructura o esqueleto de cualquier base de datos obtenida en la línea de producción. Se encontró que todas las bases de datos contenían una estructura similar, en la cual cada una de las columnas representaba la información de cada variable de interés y por otro lado cada renglón estaba representado por cada una de las observaciones realizadas durante el proceso de fabricación.

En la Figura 29 se muestra una de las bases de datos obtenidas en la línea de ensamble de los actuadores magnéticos, en la cual se puede observar cómo cada columna corresponde a una variable diferente, y cada renglón corresponde a cada observación registrada en el proceso de ensamble.

La base de datos utilizada en la celda de evaluación contaba con una dimensión de 142 columnas y alrededor de 80 mil datos.

Nº de Referencia	Status	Fecha	Hora	Nº de Batch	Diámetro embobinado	Resistencia
20011421059B	IO	15.01.2020	00:00:06	1002554	OK	7,75
20011421060B	IO	15.01.2020	00:00:40	10001531	OK	7,78
20011421061B	IO	15.01.2020	00:00:48	1002554	OK	7,76
20011421062B	IO	15.01.2020	00:00:55	10001531	OK	7,76
20011421065B	IO	15.01.2020	00:01:03	10000688	OK	7,73
20011421066B	IO	15.01.2020	00:01:10	1002554	OK	7,76
20011421067B	IO	15.01.2020	00:01:18	1002554	OK	7,75

Figura 29 Estructura de base de datos de la línea de fabricación.

Fuente: elaboración propia.

Sin embargo, se observó que 133 columnas mostraban información no relevante para el modelo (Variables categóricas) y solo 9 aportaban información para generar el modelo. Como resultado, se creó una base de datos que incluyera solo las variables de interés. Se realizó un pre-procesamiento de los datos, para descartar valores atípicos o datos faltantes de la base de datos.

Para realizar dicha exploración de datos, se utilizó un algoritmo basado en lenguaje de Python y utilizando el IDE Spyder y Google Colab como se muestra en la Figura 30, dicho al algoritmo contenía la librería de Pandas, con la cual se obtuvo un resumen de los valores faltantes:

```
df=pd.read_csv(link,delimiter=';',engine='python')
print(df.describe())
print(df.dtypes)
print('Valores Faltantes: ')
print(len(df.isnull()))
df2=df.describe()
df3=df3.append(df2)
```

Figura 30 Ejemplo algoritmo para exploración de datos, Software Spyder.

Fuente: Imagen adaptada de (AppDividend, 2021)

A continuación, se muestra la Tabla 2 un ejemplo del resumen del análisis de datos para cada una de las variables, en la cual la primera columna se listan las variables de tipo continúa encontradas en la base de datos y en la segunda columna la cantidad de datos faltantes para cada una de las variables:

Valores Faltantes:	
Resistencia_Electrica	0
Altura_Pin_Retraido	0
Altura_Pin_Extendido	0
Promedio_Tiempo_Reaccion	0
Tiempo_Conmutacion	0
Fuerza_Retraccion	0
Tipo de Falla	0
Altura_Ancla	0
Fuga	0
Profundidad_Bobina	0
dtype:	int64

Tabla 2 Resumen datos faltantes en base de datos.

Fuente: Imagen propia

Realizando la exploración de los datos, se confirmó que la base de datos no contenía datos faltantes.

Cómo siguiente paso se realizó el análisis y descripción de las variables contenidas en la base de datos, con el objetivo aquellas de tipo continuo y aquellas de tipo discreto. Esto se realizó con el mismo algoritmo previamente utilizado. A continuación, se muestra una Tabla con el resumen de la clasificación de variable obtenido a partir del análisis de la base de datos. Como se muestra en la Tabla 3, en la primera columna se muestran las variables encontradas en la base de dato y en la segunda columna se muestra la clasificación con el tipo de variable. Las variables que se describen como object son todas aquellas de naturaleza discretas

y aquellas variables que aparecen como float64 o int64 son clasificadas como variables de tipo continuas.

```
[8 rows x 10 columns]
Fecha                object
Hora                 object
Resistencia_Electrica  float64
Altura_Pin_Retraido   float64
Altura_Pin_Extendido  float64
Promedio_Tiempo_Reaccion float64
Tiempo_Conmutacion    float64
Fuerza_Retraccion     float64
Tipo de Falla        int64
Altura_Ancla         float64
Fuga                 float64
Profundidad_Bobina   float64
dtype: object
```

Tabla 3 Clasificación de variables en la base de datos enero- junio 2020.

Fuente: elaboración propia.

A continuación, se muestran en la Tabla 4 un ejemplo del resumen con las dimensiones de las bases de datos analizadas. En esta Tabla 4 se observa que la primera cifra como el total de observaciones encontradas y la segunda cifra como el total de variables analizadas en cada una de las observaciones. De acuerdo a lo anterior, con esto se confirmó que la base de datos contenía la información del total de actuadores fabricados.

El tamaño de la base de datos es (Filas, Columnas)	
138974	12

Tabla 4 Resumen con clasificación de variables.

Fuente: elaboración propia.

Por último, se observa la Tabla con el análisis estadístico de la base de datos, esta Tabla solo aplica a las variables enteras y flotantes. Dicha Tabla contiene en cada columna el nombre de cada variable analizada y en cada renglón se refiere a la medición realizada de acuerdo al estadístico. Se realiza el análisis estadístico de las variables, donde se realiza el conteo total de observaciones ("count"), se calcula

el promedio ("mean") y la desviación estándar ("std"), los valores de los cuartiles primero, segundo y tercero (25%, 50%, 75%). Adicional se encuentra el valor mínimo y máximo ("min" y "max") respectivamente de la base de datos.

A continuación, se muestra un ejemplo de la Tabla obtenida para la característica resistencia eléctrica:

Resistencia_Electrica	
count	69163.000000
mean	7.754053
std	0.054790
min	7.520000
25%	7.730000
50%	7.760000
75%	7.780000
max	8.040000

Tabla 5 Estadístico de características críticas.

Fuente: elaboración propia.

En la Figura 31a se muestra la gráfica paramétrica con la distribución de los datos de acuerdo a la distribución normal. Los rangos de los valores obtenidos durante la evaluación son representados en el eje X y la frecuencia de observación obtenida para cada uno de los valores es representada en el eje Y. En la Figura 31b se muestra la gráfica de ajuste empírico para la misma distribución, con esta gráfica se observa como los valores de la variable analizada se ajustan y acercan a la distribución normal.

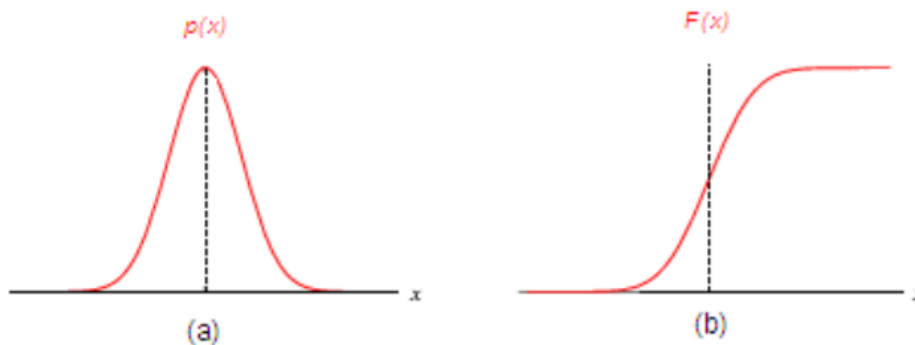


Figura 31 Gráfica de distribución y distribución ajustada empírica con distribución normal.

Fuente: Imagen adaptada de (Pei & Zaiane, 2006).

En la Figura 32a se puede observar a través de la gráfica con color rojo la distribución empírica con un comportamiento normal. Esta gráfica se muestra y utiliza como referencia para comparar con la gráfica de los datos de una variable de interés que se representa con color azul, la cual fue ajustada empíricamente. Con esto se observa como ambas gráficas tienden a permanecer unidas en su recorrido y con lo cual se confirma que el comportamiento de los datos ajustados empíricamente pertenece a una distribución normal.

Por otro lado, se observa en la gráfica 32b, en la cual se observa como las gráficas de las variables ajustadas empíricamente no se mantienen unidas a lo largo de la distribución de acuerdo a la referencia de una curva normal.

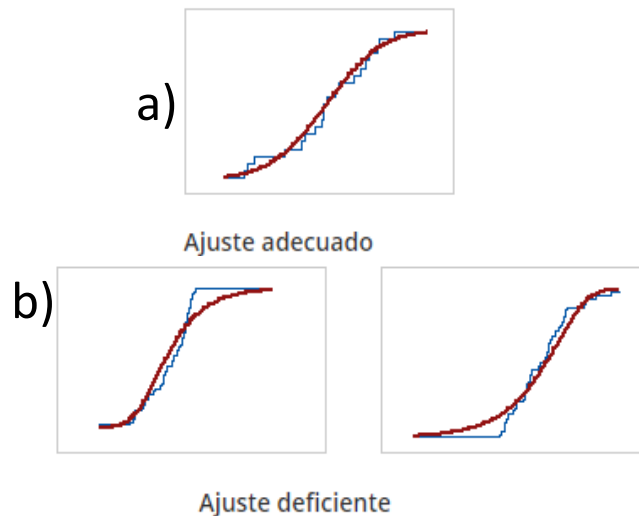


Figura 32 Evaluación de ajuste adecuado y deficiente de distribución empírica.

Fuente: Imagen adaptada de (Minitab, 2020)

Ambas gráficas se utilizaron como referencia para confirmar normalidad en las variables de las características críticas. Para garantizar que las variables analizadas cumplieran con la condición de normalidad, se realizaron los tests de Shapiro Whiks por medio de una prueba hipótesis.

- Ho: Los datos analizados de la variable analizada son datos con distribución normal
- Ha= Los datos analizados de la variable analizada no son datos con distribución normal

La prueba se realizó a través del algoritmo generado en Google Colab y con las paqueterías cargadas para el análisis de datos. El resultado de la prueba de

hipótesis se resumió como se muestra en la Tabla 6 en donde se aprecia en la primera columna el nombre de la variable analizada, en la segunda columna el valor de p-value para el intervalo de confianza en el que se realizó la prueba. Dicho valor puede aprobar la hipótesis en caso de que el valor sea mayor a >0.01 o en rechazar en caso contrario, con lo cual se puede confirmar o no si los datos cumplen con una distribución normal.

Resultado prueba Shapiro Wilk (statistic , p -value)	
Variable	p-value

Tabla 6 Reporte prueba Shapiro Whiks.

Fuente: elaboración propia.

Por último, se realizó el análisis de curtosis y asimetría de las variables, con lo cual se pretendió encontrar posibles desviaciones en la normalidad de los datos. A continuación, se presenta en la Tabla 7 en el cual se lista en la primera columna las variables analizadas y en la segunda columna se agregan los resultados analizados.

	Variable 1
Curtosis	-0.903666114
Asimetría	-0.266422435

Tabla 7 Reporte Análisis de Curtosis y Asimetría.

Fuente: elaboración propia.

3.3. PREPARACIÓN DE LOS DATOS

Durante la fase de exporación de datos se reconoce que existieron valores atípicos, los cuales se generaron durante condiciones especiles de fabricación (ejemplo calibración y/o mantenimiento de equipos de medición), sin embargo mes a mes la base de datos se filtro con el uso de un algoritmo, excluyendo dichos valores de la base de datos usada para crear el modelo de regresión.

En la Figura 33, se muestra como se realizó el proceso de reducción de la base de datos original excluyendo las variables que no aportaban información al modelo de regresión.



Figura 33 Proceso de preparación de datos.

Fuente: elaboración propia.

Para confirmar que la nueva base de datos contenía la información correcta, nuevamente fue analizada dicha base de datos con el algoritmo previamente utilizado. En caso de obtener un valor faltante ó atípico, el proceso de filtro se repetía para excluir dicho valor y nuevamente se ejecutaba el proceso de descripción de la base de datos hasta confirmar que los valores de la base de datos se encontraba dentro de los valores esperados. A continuación, se muestran la Figura 35 la gráfica lineal de la característica de Fuerza de Retracción del mes de marzo del año 2019, en esta gráfica se puede apreciar la presencia de valores atípicos (círculos en azul), los cuales se localizan en puntos extremos y en algunos casos fuera del rango de especificación. Estos valores deben de ser excluidos del proceso de análisis de regresión, debido a que los datos son generados cuando el proceso se encuentra en condiciones especiales (calibración, mantenimiento, etc.) y por ende no son representativos del proceso normal.

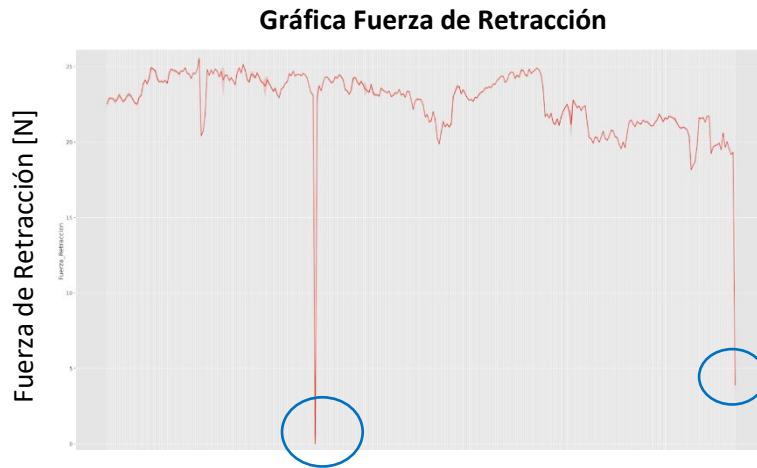


Figura 34 Gráfica con valores atípicos para característica fuerza de retracción correspondientes al mes de marzo de 2019.

Fuente: elaboración propia.

3.4. MODELADO

3.4.1. Análisis de correlación de los datos

Se empleó el método de correlación de Pearson para poder explicar la relación lineal que presentan los datos. Previamente se analizó el indicador más adecuado a utilizar para explicar la relación entre las variables predictoras y de respuesta, ya que se contaron con diferentes opciones para el análisis del mismo (correlación de Pearson, Spearman, Kendall, entre otras) (Hernández, 2018). Se realizó la selección del método de Pearson, debido a que las características de las variables de respuesta debían de responder a un comportamiento con una distribución de datos normales, lo cual garantiza la estabilidad del proceso y por ende la posibilidad del cálculo de indicador de habilidad de proceso.

Según Amat (Amat, 2016) el cálculo del coeficiente de Pearson se explica como la covarianza estandarizada, la cual se define mediante la fórmula (Jundong Li et al., 2017):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - x_{prom}) * (y_i - y_{prom})}{\sqrt{\sum_{i=1}^n (x_i - x_{prom})^2 * \sum_{i=1}^n (y_i - y_{prom})^2}}$$

3.4.2. Modelo de regresión lineal

Para poder plantear este modelo de regresión se analizaron los valores obtenidos en la correlación de Pearson. La evaluación del modelo de regresión lineal se realizó a través de los residuos, los cuales son el resultado de la diferencia entre el valor real de la observación y el valor de predicción del modelo. El análisis de los residuos no solo permite visualizar la variación entre predicción y valor real, sino que también aporta información para confirmar que se cumplen las condiciones necesarias para generar un modelo de regresión lineal (Amat, 2016).

Modelo de regresión múltiple. En la literatura se define como el modelo que ajusta modelos lineales entre variables dependientes y variables independientes.

Este se representa mediante la siguiente fórmula (García et al., 2018):

$$y_i = \beta_0 + \beta_1 * x_1 + \beta_2 x_2 + \dots \beta_n x_n + \epsilon_i$$

Donde y_i es la variable dependiente (en nuestro caso la variable de interés altura de pin), β_0 el valor constante ordenado al origen y $\beta_1, 2, \dots, i$ representan los valores de los coeficientes de regresión parcial, que se determinan a partir de las variables predictoras ($x_{1,2,i}$) (Montero R, 2016).

3.4.3. Selección de predictores

Para realizar el proceso de selección de predictores óptimos en el modelo de regresión, se utilizó el método iterativo Step Forward (SF), con el cual analizó la contribución parcial de los cada uno de los predictores para explicar el modelo.

El método SF en la búsqueda del mejor rendimiento del sistema y el cual se explica mediante la siguiente fórmula:

$$J_{max} = \max_{X \in (\mathcal{X} - Z_{l-1})} J_{nrep}(Z_{l-1} + \{X\}).$$

Fuente : Adaptada de (Ververidis & Kotropoulos, 2005).

3.4.4. Estimación de bondad de modelo

Para realizar la evaluación del modelo de regresión, se utilizó el método de Selección hacia adelante, con el cual se agregaron las variables De acuerdo a la importancia dentro del modelo de Regresión.

$$R^2 = 1 - \frac{\sum(y - y_i)^2}{\sum(y_i - y_{promedio})^2}$$

Para evaluar la importancia de cada variable, se evaluó el valor de Ajuste de Bondad para modelo simple De acuerdo a la siguiente fórmula:

$$R^2_{ajustado} = R^2 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

4. RESULTADOS

4.1. PRUEBA DE NORMALIDAD EN VARIABLES DE INTERÉS

En la gráfica se observan como los valores de la característica se encuentran distribuidos en el eje X y la frecuencia de observación para dichos valores se representa mediante el eje Y.

A continuación, se presentan las gráficas de los valores observados correspondientes al mes de enero de 2020 para las características de resistencia eléctrica, altura de pin retraído y fuerza de retracción.

En la Figura 35a se observa como los valores se ven distribuidos en el rango de 7.60 hasta 8.1 ohm, sin embargo, la frecuencia con que los datos se presentan a lo largo de la distribución no es de acuerdo con una curva normal. Se analiza el valor de resistencia igual a 7.85 ohm, el cual se encuentra con una frecuencia de casi 20000 observaciones lo cual representa un valor muy por encima de las frecuencias de observación (<10000). En la Figura 35b se observa la gráfica con el ajuste empírico de la variable, en esta gráfica se observa cómo el 90% de los valores se encuentran en el rango de 7.7 a 7.9, esto vuelve a confirmar que la distribución obtenida no presenta un comportamiento normal.

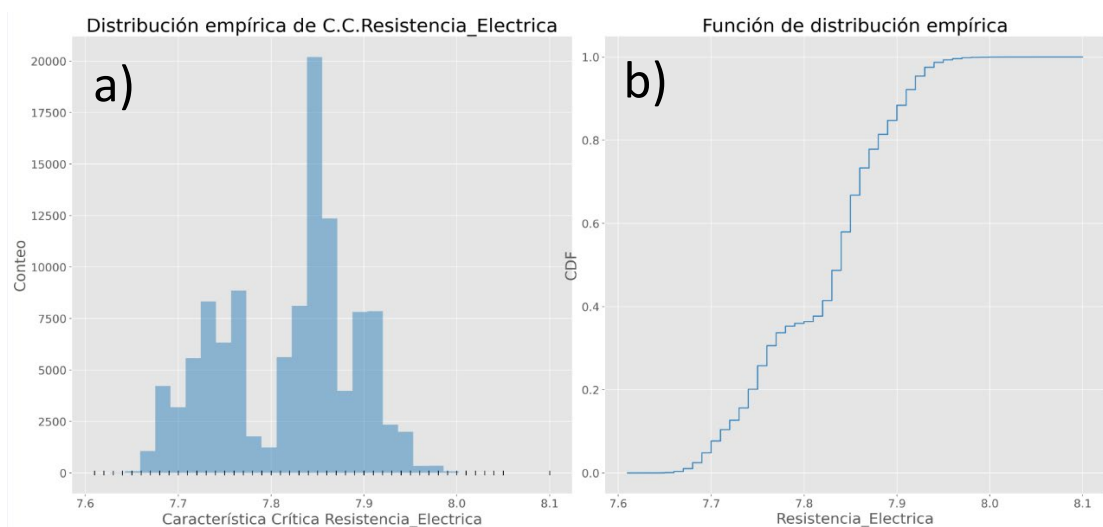


Figura 35 Gráficas de distribución empírica por característica resistencia eléctrica.

Fuente: elaboración propia.

Debido a que el proceso de ensamble tiene una baja influencia sobre el resultado de la resistencia eléctrica y se asume que en su mayoría se relaciona directamente a las características del material, se puede concluir que la mayoría de las observaciones no presentan normalidad para esta característica, lo cual concluye que esta variable no puede ser predicha a través de las características de ensamble.

En la Figura 36a se puede apreciar la gráfica de distribución de los para la característica de altura de pin extendido, en esta gráfica se observa que el rango de los valores se extiende desde 33.15 hasta 33.45 mm. La frecuencia de la distribución de los datos muestra una tendencia a formar una curva con mayor similitud a la curvatura normal. Sin embargo, la presencia de valores menores a 33.25, ubicados en el extremo izquierdo del rango, se intuye que dicho proceso no es completamente normal. Al analizar la gráfica 36b con los valores ajustados empíricamente, se observa como los valores ubicados en el extremo izquierdo no representan ni siquiera el 5% del total de observaciones. Con esto se intuye que estas observaciones pueden no ser significativas para analizar el resto del total de observaciones analizadas para esta característica. Al observar la distribución de los datos mayores a 33.25 mm, se observa que el 90% de los datos se encuentran entre el rango de 33.25 hasta 33.40 mm, con lo cual se opta por continuar con el análisis de normalidad por medio de prueba de hipótesis de Shapiro Whiks y D'Agostino.

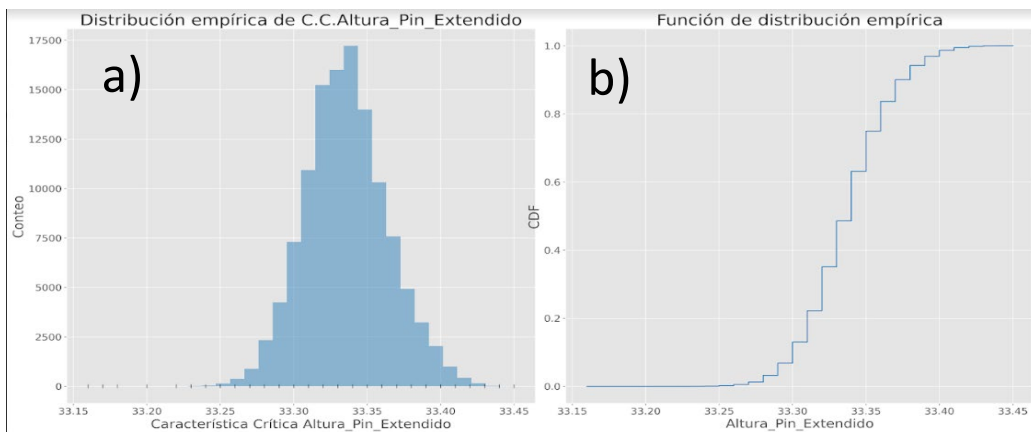


Figura 36 Gráficas de distribución empírica por característica altura de pin retraído.

Fuente: Elaboración propia.

Por último, se muestra el resultado de la gráfica de la característica de fuerza de retracción. En la Figura 37a se observa la distribución de los datos en el rango desde 7.5 hasta 22.5 N, sin embargo, nuevamente la frecuencia de las observaciones en la distribución se encuentra mayormente en el rango de 15 hasta 22.5 N con una distribución con curvatura similar a la curva normal.

Al observar la gráfica con la distribución de los datos ajustados empíricamente, se observa como los datos nuevamente en el rango de valores entre 7.5 y 15 N no acumulan ni siquiera el 5% de los valores observados. Por otro lado, el rango de valores entre 16 y 21 N acumula casi el 90% del rango de los valores observados. Por este motivo se concluye en realizar un análisis de los datos de acuerdo a la prueba de hipótesis para confirmar normalidad en los datos.

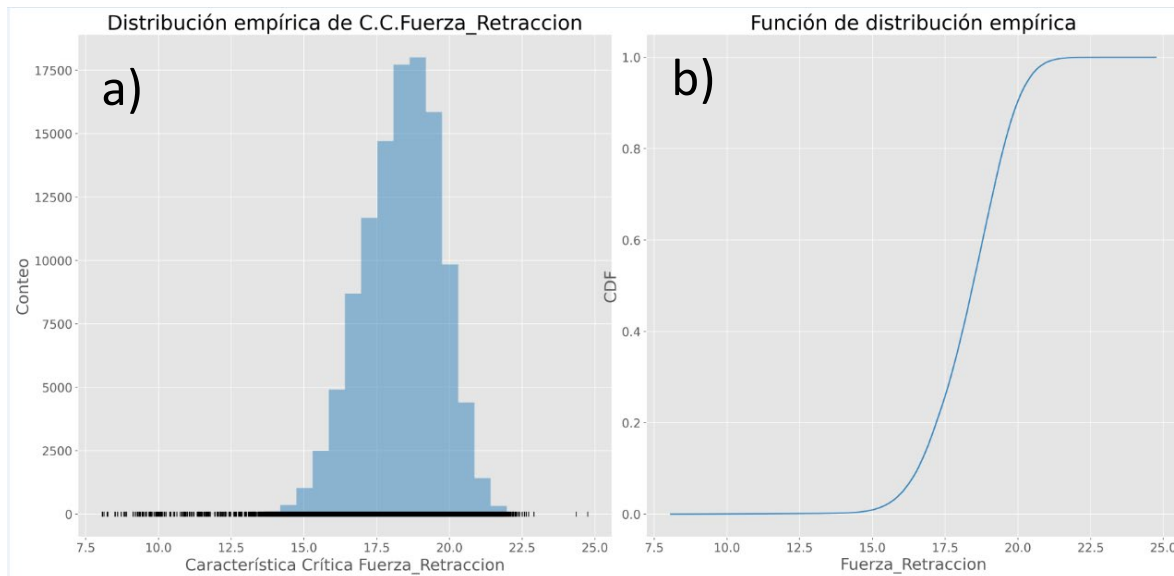


Figura 37 Gráficas de distribución empírica por característica crítica fuerza de retracción.

Fuente: Elaboración propia.

A continuación, se muestran los resultados de la prueba de normalidad de acuerdo con la prueba de hipótesis de Shapiro Whiks y D'Agostino's, para las características de resistencia eléctrica, valor de pin extendido y fuerza de retracción de acuerdo al mes de enero de 2020. Los resultados de la prueba de Shapiro Whiks se resumen en una Tabla donde se indica la característica crítica, el valor estadístico alcanzado y el por último el p-value con el cual se realizó la prueba de hipótesis.

En la Tabla 8 se observan los resultados de la prueba de normalidad para las tres variables analizadas, con los valores p-value obtenidos de la prueba de hipótesis. Después de haber realizado un proceso de normalizado de los datos, se observa que el valor p-value para la variable de Resistencia eléctrica no es mayor a 0.01 por lo cual se puede concluir y confirmar que los datos no cuentan con una distribución normal.

En el caso de las variables valor de pin extendido y fuerza de retracción, la prueba de Shapiro confirma que los valores de p-value alcanzados confirman la hipótesis de que los datos cumplen con una distribución normal.

Resultado prueba Shapiro Wilk (statistic , p -value)	
Resultados estadísticos de Resistencia Eléctrica	0.009571606
Resultados estadísticos de Altura Pin Extendido	0.66914916
Resultados estadísticos de Fuerza Retracción	0.983470738

Tabla 8 Resultados prueba Shapiro Wilk, correspondientes al mes de enero 2020.

Fuente: elaboración propia.

Por último, se realizó el análisis de los indicadores de curtosis y asimetría, con los cuales se pretende detectar si los valores obtenidos pueden o no presentar alguna desviación significativa de la distribución de normalidad. Siendo la referencia que para indicadores se encuentren dentro del rango de -1 a 1 para ser considerados con ligera desviación de normalidad.

En la Tabla 9 se observa como los valores de resistencia eléctrica presentan un valor de curtosis muy cercano a -1, con lo cual se reafirma la conclusión obtenida en la prueba de hipótesis. En el caso de las variables de valor de pin extendido y fuerza de retracción se obtuvieron valores que presentan ligera desviación, sin embargo, no representan valores fuera del rango de referencia, por lo cual la se refuerza el resultado de la prueba de hipótesis.

	Resultados estadísticos de Resistencia_Electrica	Resultados estadísticos de Altura_Pin_Extendido	Resultados estadísticos de Fuerza_Retraccion
Curtosis	-0.903666114	-0.107927447	-0.065369069
Asimetría	-0.266422435	0.096049786	0.02169787

Tabla 9 Resultados prueba análisis de curtosis y asimetría, correspondientes al mes de enero.

Fuente: elaboración propia.

4.2. RESULTADOS DEL ANÁLISIS DE CORRELACIÓN

En la Figura 38, se presenta la matriz de coeficientes de correlación de Pearson (R) entre las variables de ensamble y las características críticas. El valor de correlación entre cada variable se presenta también como un mapa de calor que permite identificar las contribuciones más importantes entre las medidas (Usharani & Shanthini, 2020).



Figura 38 Matriz y mapa de calor de coeficientes de correlación de Pearson R.

Fuente: elaboración propia.

Se identifica en la barra de colores que desde $R=-1$, en color rojo, representa una correlación lineal negativa, lo cual significa que la relación que existe entre las variables es 1 a 1 o perfecta, pero en este caso es de forma inversa; es decir, cuando una variable aumenta la otra disminuye. Los valores de $R=0$ están representados en color blanco en el mapa de calor. Esto significa que no hay alguna dependencia entre las variables de interés o correlacionadas. Por su parte, los valores de $R=1$, en color azul, muestran una correlación lineal positiva perfecta, esto quiere decir que cuando una variable aumenta en una unidad de magnitud la de correlacionada aumenta en la misma cantidad.

4.3. RESULTADOS MODELOS DE REGRESIÓN

La Tabla 10 muestra el reporte del primer modelo de regresión lineal implementado para la variable de interés altura de pin (y) en función de la variable de ensamble altura de ancla. Este modelo es una regresión lineal implementada por el método de mínimos cuadrados ordinario (OLS) para una sola correlación (Df). El total de la población de datos empleado fue de 89431. El modelo de regresión se realizó con el objetivo de reducir el grado de error entre las predicciones del modelo y los valores reales. Esto se llevó a cabo mediante un análisis de residuos y seleccionando el método óptimo para dividir la base de datos para el entrenamiento del modelo y el test del mismo.

Reporte de primer modelo de regresión				
Variable dep:	y	R-cuadrada:	0.759	
Modelo:	OLS	Adj. R-cuadrada:	0.759	
Método:	min cuadrados	F-estadística:	281600	
N° Observ	89431	Prob (F-statistic):	0	
Df modelo:	1	rmse	0.02340	
Inter conf	[0.025 0.975]			
	coef	std err	t	P> t
Const	-5.1696	0.073	-71.254	0
Ancla	1.0602	0.002	530.614	0

Tabla 10 Valores del primer modelo de regresión lineal.

Fuente: elaboración propia.

Se trabajó con un intervalo de confianza estándar entre 2.5% y 97.5% de los datos. Se obtuvo una estimación de bondad del ajuste (R^2) de 0.759.

El valor del error de ajuste (rmse) fue de 0.0234, esto define cuanto se está desviando el valor del promedio de predicción respecto a los valores reales. El coeficiente parcial de regresión para el ancla fue de 1.0602. Este valor se emplea en la implementación del segundo modelo en el que incluye la contribución de todas las variables al comportamiento del pin del actuador magnético.

Después de identificar la correlación real entre la variable altura de pin con el ancla se implementó el método de regresión múltiple, en donde se agregó cada variable en función de las iteraciones. Esto se llevó a cabo considerando el método step-forward que implica evaluar cada iteración agregando los predictores de manera secuencial y revisando su efecto sobre la altura de pin, para saber si cada una de ellas tenía o no alguna contribución (Zhang Z, 2016).

4.3.1. Selección de predictores

En la Tabla 11 se presenta el resultado de los coeficientes parciales de la regresión lineal múltiple, obtenidos por el método de Step-Forward. Para este análisis se tuvo en cuenta la contribución de las variables de ensamble desde el coeficiente de correlación de Pearson mayor al menor.

Debido a la gran cantidad de datos generados en las líneas de producción y dado que la información contenía mucha dispersión, se tomó la decisión de establecer cuál era la correlación que existía entre las variables de evaluación para poder definir un primer método de regresión y esto se consideró teniendo en cuenta el valor de R.

Resultado				PB	ptR	tC	AA	VPD	RE	FR
Cambio	R2	Iteración	R	Coeficientes parciales de regresión lineal múltiple						
	0.76	1	0.87				1.0602			
0.010	0.77	2	0.67			0.0196	0.9477			
0.000	0.77	3	0.65		-0.0065	0.0254	0.9470			
0.079	0.85	4	0.36			0.0271	0.8684	0.5116		
0.003	0.85	5	0.28			0.0339	0.8485	0.5397	-0.0460	
0.023	0.87	6	-0.28	-0.3745		0.0320	0.8639	0.4044	-0.0449	
0.000	0.87	7	-0.04	-0.3739		0.0323	0.8611	0.4063	-0.0449	-0.0004

Tabla 11 Coeficientes parciales de la regresión múltiple.

Fuente elaboración propia.

En la Tabla 11 se presenta el valor de cambio, el coeficiente de correlación R2, la iteración correspondiente, el coeficiente de correlación de Pearson en orden descendente y los coeficientes parciales de correlación. Las interacciones se establecieron en orden, según el peso que tiene cada variable al modelo de regresión múltiple. Esa decisión se tomó con base en la matriz de correlaciones de Pearson (ver Figura 1). Se muestran los resultados de predicción de altura de pin empleando el modelo de regresión lineal múltiple, en el cual se tienen en cuenta las contribuciones no solo la altura de ancla sino también la profundidad de bobina, el promedio de tiempo de reacción, el tiempo de cambio, valor de pin dentro, resistencia eléctrica y fuerza de retracción.

La Tabla 12 muestra el reporte del modelo de regresión lineal optimizado a través del método de StepForward para la variable de interés de altura de pin (y) en función de las variables de ensamble altura de ancla, tiempo de conmutación, altura de pin retraído, resistencia eléctrica, profundidad de bobina y fuerza de retracción. El método de ajuste sigue siendo por mínimos cuadrados (OLS). Los datos analizados fueron 78252, en un intervalo de confianza entre 2.5% y 97.5%.

Se obtuvo un valor de bondad de ajuste (R2) de 0.874, el cual representa una optimización en comparativa con el modelo inicial, esto se explica debido a la contribución que se logra con la selección de los predictores. Por último, de acuerdo al modelo generado se obtuvo un valor de error de ajuste (rmse) de 0.0168, con lo cual se concluye que la desviación del promedio de predicción respecto a los valores reales ha disminuido.

Reporte de modelo de regresión optimizado				
Variable dep:	y	R-cuadrada:	0.874	
Modelo:	OLS	Adj. R-squared:	0.874	
Metodo:	min. Cuadrados	F-estadística:	9.06E+04	
No. observaciones:	78252	Prob (F-statistic):	0	
Df Model:	6	rmse	0.01684973	
Inter conf	[0.025 0.975]			
	coef	std err	t	P> t
const	-5.2567	0.115	-45.639	0
Altura_Ancla	0.8601	0.002	388.912	0
Tiempo_Conmutacion	0.0324	0	109.376	0
Altura_Pin_Retraido	0.4071	0.003	153.181	0
Resistencia_Electrica	-0.0451	0.001	-44.094	0
Profundidad_Bobina	-0.3728	0.003	-118.28	0
Fuerza_Retraccion	-0.0004	4.53E-05	-8.394	0

Tabla 12 Valores del modelo de regresión lineal optimizado.

Fuente: elaboración propia.

4.4. EVALUACIÓN DEL MODELO

El proceso de análisis de la información, se realizó para confirmar que las variables cumplieran con condiciones necesarias para establecer un modelo de regresión, así como poder calcular sus índices de habilidad de proceso y establecer correlación entre ellas.

Entre estos análisis se puede mencionar análisis de normalidad, colinealidad, homocedasticidad, etc.

La Figura 39 presenta la distribución de histogramas de residuos de los modelos empleados; regresión lineal simple en a) y múltiple en b).

La gráfica de distribución segmentada en la Figura 39a) muestra que el ajuste no es adecuado y que los datos no se agrupan de una forma normal. Hay distintas frecuencias alrededor de la media. Todos los valores predichos iguales al valor real indican que hay un valor 0 en los residuos y que el modelo está funcionando de forma correcta. La densidad de histogramas de residuos mostrados en la Figura 39b) corresponden a los valores obtenidos aplicando el modelo de regresión lineal múltiple e indican que la mayoría de los datos están dando alrededor de la media. En este caso al considerar todos los predictores contribuyendo a la altura de pin, el

modelo explica de mejor manera el comportamiento del pin y los residuos tienen una distribución más normal.

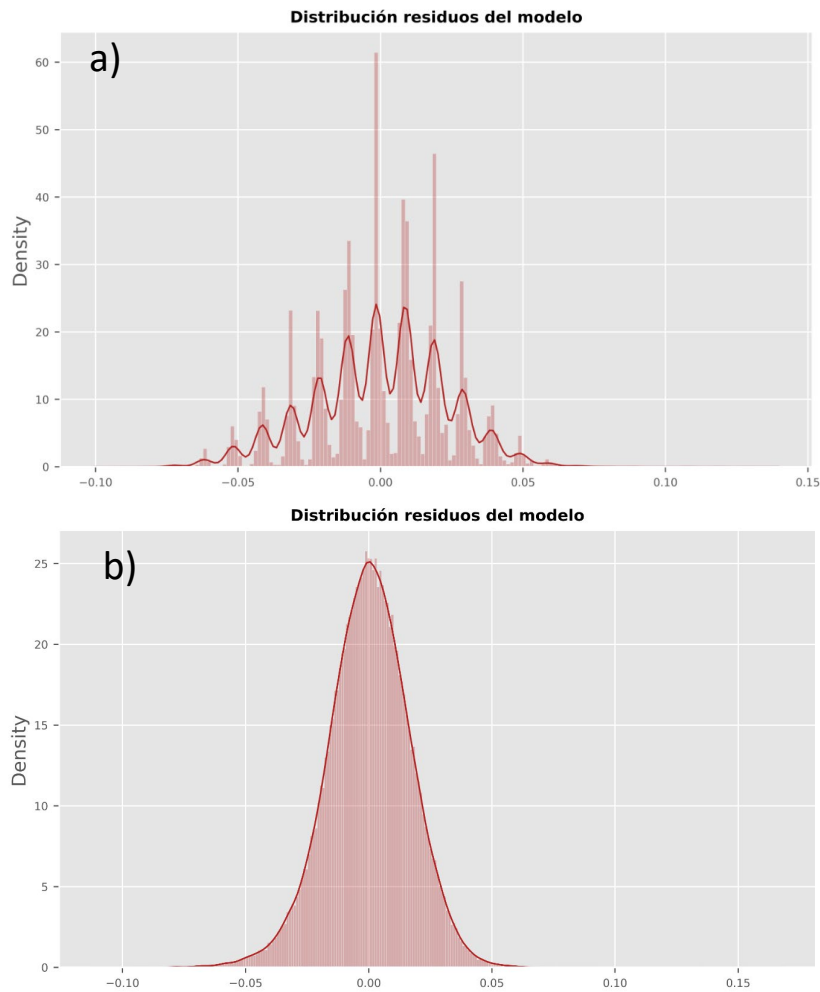


Figura 39 Gráfica de distribución de residuos.

Fuente: elaboración propia.

Complementando la gráfica anterior se realizó una gráfica Q-Q para comparar la distribución de los datos de acuerdo a los cuantiles (25%, 50%, 75%)

En la gráfica se observa la representación de los cuantiles en los ejes X y Y, los valores promedio centrados. A partir de aquí cada número en el eje representa un cuartil de la distribución. En la gráfica se observa la línea de referencia con color rojo que representa una distribución normal.

De esta forma los valores residuales se grafican para confirmar si estos se acercan y mantienen respecto a la línea de distribución normal.

En la gráfica 40a) se observa como los valores residuales se separan de la línea de referencia y por tanto no se ajustan a la línea de distribución normal de los datos. Tomando como referencia el primer cuartil ($x=1$) se observa como los valores residuales no se concentran en este punto, sino que se distribuyen en los extremos. Esto nos vuelve a confirmar que el modelo inicial carecía de una distribución normal en los valores residuales y por ende el modelo de regresión no cumplía del todo las condiciones iniciales.

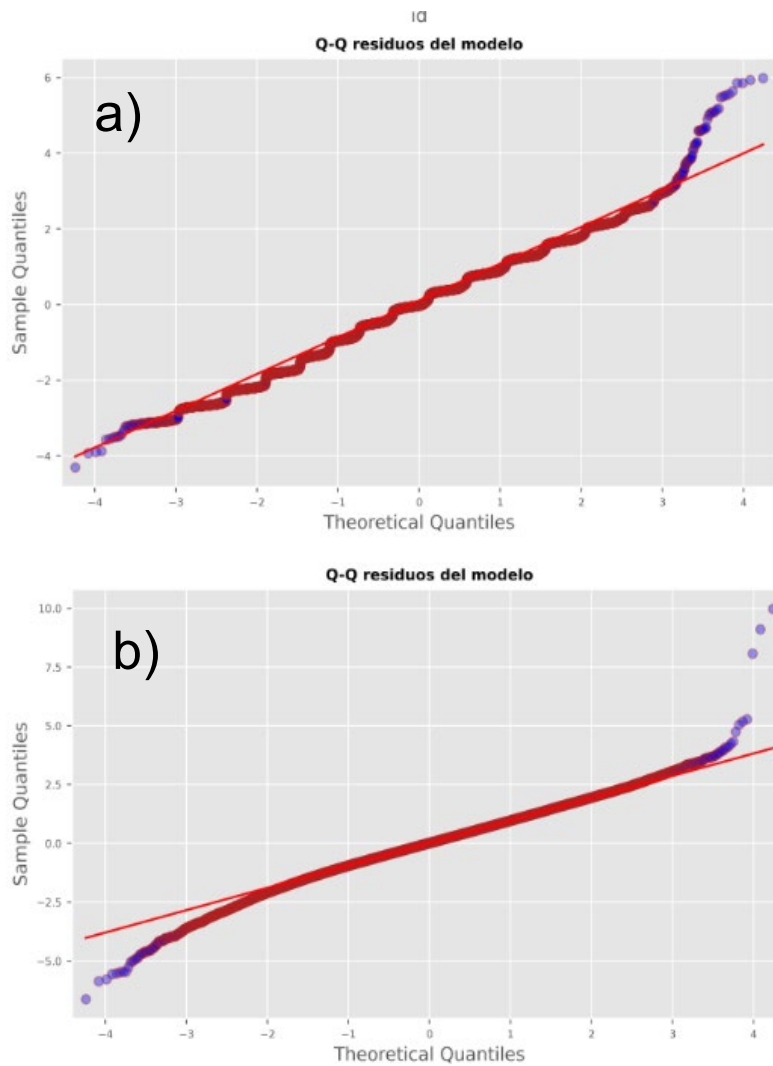


Figura 40 Gráfica Q-Q de residuos en comparativa con cuantiles normales.

Fuente: elaboración propia.

Sin embargo, al realizar el proceso de optimización con la selección de predictores se observa como el segundo modelo concentra los valores residuales en los primeros tres cuartiles de cada extremo, con lo cual se confirma que el modelo se encuentra más cercano a una distribución normal en la mayoría de los cuartiles.

En la Figura 41a) se muestra la predicción de los datos considerando los modelos de regresión lineal sencillo y múltiple. En la Figura 41a) se observó como el modelo de regresión simple ajustado a los valores medidos muestran una dispersión considerable respecto de la predicción y solamente tiene en cuenta el efecto del valor de ancla sobre la altura de pin extraído.

La confiabilidad de la predicción en este caso depende de la contribución del valor de la variable de ensamble altura de ancla. Si se tiene una concentración grande de puntos en la parte superior de la regresión lineal se obtendrá una gráfica de residuos más dispereja, lo cual no es recomendable para ajustes normales.

En la Figura 41b) se muestran los resultados de predicción de altura de pin empleando el modelo de regresión lineal múltiple, en el cual se tiene en cuenta el efecto de todas las variables.

Para este caso hubo una aproximación mucho más real y se tienen los valores más cercanos a la media. Con esto se genera un modelo que pueda predecir de forma más exacta los valores de la característica del pin.

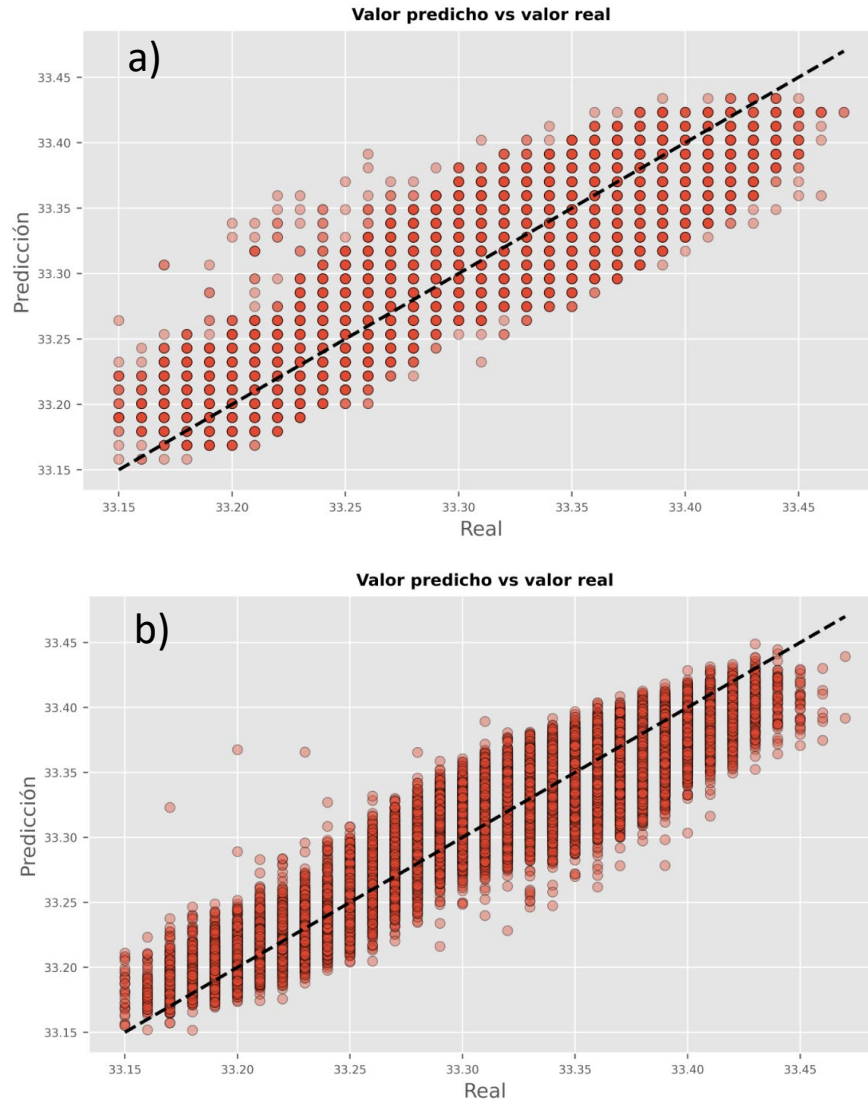


Figura 41. Valores de predicción de datos en función del valor real.

Fuente: elaboración propia.

En la Figura 42 se observan los residuos de los datos obtenidos del modelo en función de la predicción realizada para los dos casos de modelo de regresión empleados. En 42a) el modelo de regresión simple hace que los residuos se distribuyan entre -0.10 y 0.14, lo cual muestra que hay una varianza importante.

El valor predicho para la altura de pin considerando solamente el efecto del ancla muestra los datos de forma seccionada y esto hace que en la imagen se vean grupos verticales de datos, lo cual es un vicio del modelo y no representa la realidad de las medidas.

Las condiciones ideales para hacer regresiones indican que los valores de residuo siempre deben ser aleatorios. En nuestro análisis los valores de residuos presentan tendencias.

Al aplicar el modelo de regresión lineal múltiple, las contribuciones de todas las variables en la fabricación del actuador hacen que la predicción sea más confiable y con menor varianza. Aunque el modelo constituye un sistema de entrenamiento al ajustar las variables, es importante evitar procesos de sobre ajuste.

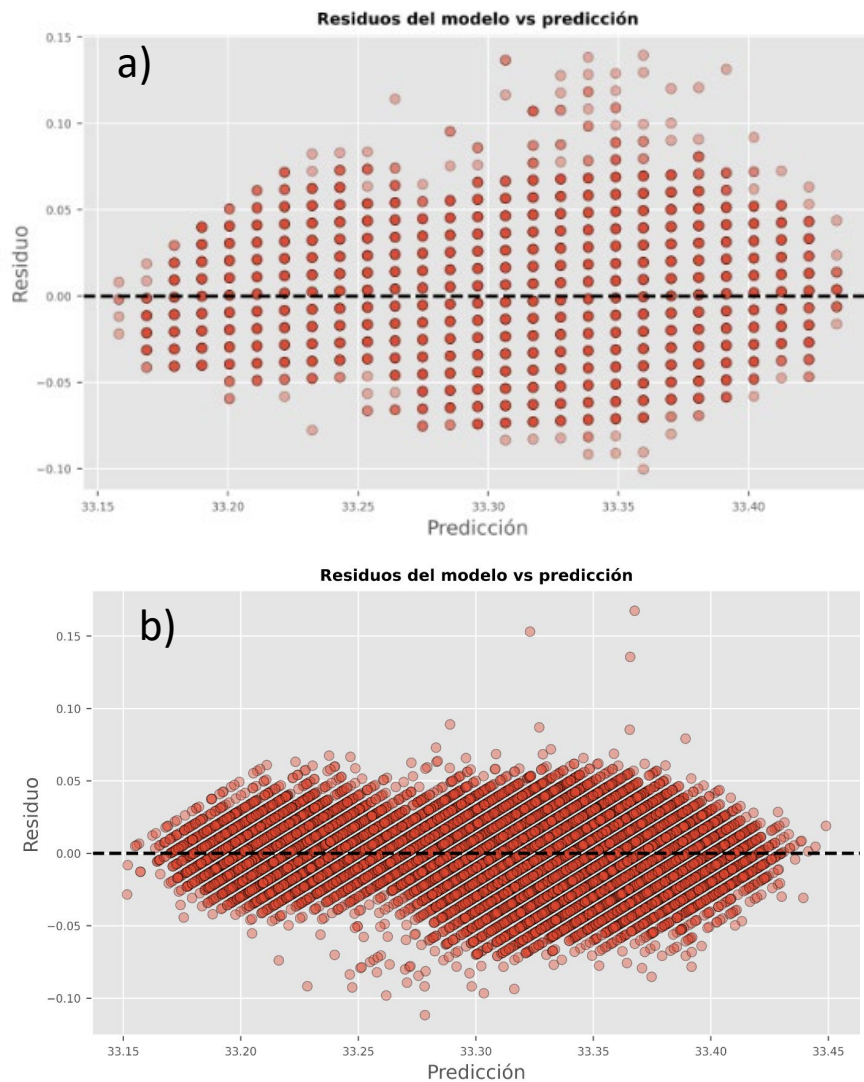


Figura 42 Dispersión de los residuos del modelo.

Fuente elaboración propia.

Los residuos no se deben ajustar forzando a disminuir la varianza porque el modelo quedaría sobre ajustado, lo cual ocurre cuando durante el entrenamiento de un sistema de aprendizaje automático se plantea únicamente la optimización de una variable (varianza, por ejemplo) y entonces el sistema queda con algoritmos anómalos y reproduce patrones que no son generales (Gonzalo, 2020). A pesar de que en la Figura 5b) se observa que los valores están dispersos alrededor del residuo central 0. Sin embargo, hay cierta cantidad de datos que se pueden descartar. No obstante, al tratarse de medidas reales tomadas en un proceso de ensamble, el modelo se debe entrenar y cuidar que no se encuentre sobre ajustado. Las 7 iteraciones realizadas integrando de manera sistemática las contribuciones de cada variable, permiten tener un modelo de regresión y predicción confiable y puede emplearse para análisis con otras variables que tengan codependencia de hasta 8 grados de libertad (Fernandez-Palacin et al., 2018).

4.5. DESPLIEGUE

Una vez obtenidos los modelos de regresión lineal que explicaban la contribución de cada una de las características de ensamble en relación a las características críticas de funcionamiento, se realizaron a partir del mes de enero de 2020 los primeros ajustes al proceso tomando como guía los siguientes criterios:

- Contribución de la característica de ensamble antes de ajustar el proceso
- Identificación de causas de valores atípicos antes del ajuste de proceso
- Documentación de acciones para controlar características de ensamble

Por último, se estableció el análisis de normalidad por medio de gráficas de distribución en las características de críticas de funcionamiento, como parte del método para detectar una posible desviación en el proceso.

A continuación, se presenta la Tabla 13 en donde se muestran los resultados obtenidos durante el desarrollo del estudio. En el primer renglón se observa la producción total de actuadores aprobados en la línea de ensamble. Con lo cual se puede observar que la producción fue incrementando en comparativa con 2019, cuando la producción cayó hasta 50 mil actuadores.

En el segundo renglón se observa la cantidad de material de rechazo producido por características críticas de funcionamiento a lo largo del periodo.

En el tercer renglón se observa el indicador de % de actuadores rechazados, el cual se calcula de acuerdo a la relación entre la cantidad de actuadores aprobados y rechazados. De acuerdo a lo anterior, se observa que el % de material rechazado se vio disminuido y con mayor estabilidad en comparativa con el año 2019, obteniendo valores dentro del objetivo establecido (<2%).

Estadística 2020

	Enero	Febrero	Marzo	Abril	Mayo	Junio
Total piezas OK	111789	118405	87781	127654	163518	166815
Total piezas NOK	716	583	735	391	666	267
% de material rechazado	0.64%	0.49%	0.83%	0.31%	0.41%	0.16%
% objetivo de material rechazado	2%	2%	2%	2%	2%	2%

Tabla 13 Resultado % de material rechazado.

Fuente: elaboración propia.

En conclusión, se observa una reducción significativa el porcentaje de material rechazado con respecto a 2019 y permaneciendo dicho indicador dentro del objetivo a lo largo del estudio.

CONCLUSIONES

En este trabajo se implementó de forma exitosa una metodología para el análisis de datos y predicción del comportamiento de un actuador magnético a partir de sus variables de ensamble y características críticas medidas en tiempo real.

Las principales conclusiones se presentan a continuación:

- Durante estudio de las variables Valor de Pin Extendido y Fuerza de Reacción, se encontraron valores atípicos, con lo cual se concluye que el pre procesamiento de estos datos es importante para generar el modelo de regresión
- Inicialmente se aplica el modelo de regresión lineal simple para estimar la bondad del ajuste y definir la pertinencia de aplicación del modelo de regresión lineal múltiple.
- Se estableció una matriz de correlación de dependencia a través de factores Pearson entre variables de ensamble y críticas que permiten definir los parámetros de entrada para el modelo step-forward.
- A partir del método step-Foward se encontraron variables con poca o nula aportación para optimizar el modelo de regresión, por lo cual dichos valores no son incluidos en el último modelo de regresión.
- El modelo de regresión lineal múltiple considera la contribución de las variables de ensamble y características críticas y permite predecir en un 87.5% y explicar la relación entre valores medidos durante el proceso de ensamble.
- El uso de métodos de regresión permite establecer señales que permiten reconocer previamente el efecto del ajuste de una posible variable predictora.
- El uso de modelos de regresión y análisis de variables permitió la reducción de material de rechazo en la línea de ensamble.

RECOMENDACIONES

El desarrollo de esta tesis fue realizado con un método de regresión lineal que relaciona varias variables. Durante la investigación del marco teórico se encontraron diferentes métodos de regresión, en las cuales la relación no solo no es lineal sino también se proponen el uso de variables categóricas en el proceso. Se recomienda seguir extendiendo el conocimiento y realizar nuevos métodos para generar más métodos de predicción en los procesos.

El estudio se realizó por medio de un método de regresión lineal, sin embargo, durante el desarrollo también se propuso el desarrollo del análisis por medio de programación lineal, específicamente métodos de optimización basado en condiciones (ej. Nelder Mead, método de la zona dorada).

También se recomienda realizar el estudio por medio de estos métodos, con lo cual no solo se confirman los resultados de optimización propuestas en esta tesis, sino se puede desarrollar un método híbrido que desarrolle la regresión lineal y una optimización con programación lineal.

APORTACIÓN DE LA TESIS

El desarrollo de este estudio contribuye con un método capaz de explicar la relación entre las variables de ensamble de un proceso y su posible efecto en el comportamiento del producto. El estudio presenta y propone una metodología de minería de datos para el análisis de procesos de manufactura, con lo cual se realiza un enfoque alternativo a las propuestas actuales de mejora de procesos. Esto conduce a la toma de decisiones basada en un previo análisis que considera factores de riesgo para la generación de material de rechazo y desempeño de indicadores de proceso. Por último, el uso de técnicas de aprendizaje de máquina para el análisis de procesos contribuye en la mejora continua de procesos automatizados y con una nueva metodología para el análisis y solución de problemas.

APORTACIÓN SOCIAL DE LA TESIS

El desarrollo de este estudio contribuye a la disminución de generación de material rechazado en el proceso de ensamble de actuadores magnéticos y con lo cual se reduce el impacto ambiental, logrando completar el ciclo de vida de los materiales utilizados para el ensamble del actuador. Por otro lado, también permite analizar el comportamiento del diseño del actuador actual durante su fabricación y su evaluación, permitiendo la propuesta de nuevos diseños optimizados y/o con nuevas propuestas en su proceso de ensamble. Lo anterior contribuye en la evolución a procesos optimizados con alternativas orientada a la reducción del costo y/o energías. Por último, este estudio contribuye con un método capaz de realizar el procesamiento de información en tiempo real y explicar desviaciones en la estabilidad del proceso mismo, lo cual una ventaja cuando existe una generación de grandes cantidades de información.

REFERENCIAS

- Addlink Software Científico. (2020). *Ajuste de curvas con regresión lineal y no lineal con Minitab*. <https://www.addlink.es/noticias/minitab/2981-ajuste-de-curvas-con-regresion-lineal-y-no-lineal-con-minitab>
- Alavez, G. (2018). *Economía colaborativa y microempresas de hospedaje temporal en San Luis Potosí*". www.ipn.mx
- Amat, R. (2016). *Machine learning con Python y Scikit-learn*. https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn.html
- AppDividend. (2021, May 24). *Pandas DataFrame describe() method In Python example*. <https://appdividend.com/2020/05/25/pandas-dataframe-describe-method-in-python/>
- Audi of America LLC. (2009). *The 2.0L 4V TFSI engine with AVS* (pp. 9–9). https://vwcampersite.files.wordpress.com/2015/01/ssp_922903_2-0l_4v_tfsi_engine.pdf
- Azevedo, A., & Santos, M. F. (2008). *KDD, SEMMA AND CRISP-DM: A parallel overview*. <https://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf>
- Banco de México. (2018). *La automatización en México desde una perspectiva regional*. <https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-las-economias-regionales/recuadros/%7BE3665296-DCDE-78FD-54CB-0420E1CD9A36%7D.pdf>
- Blancas, E. (2015, November). *Data science en México: retos y oportunidades [Blog]*. <https://medium.com/@edublancas/data-science-en-m%C3%A9xico-retos-y-oportunidades-4a68f683b777>
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology* 44(1). In *Journal of Mathematical Psychology* (Vol. 44). www.idealibrary.com
- Byrne, D., & Wilkins, W. (1998). *Publishing your medical research paper*. <https://www.sciencedirect.com/journal/archives-of-medical-research>
- Cluster Industrial. (2020, February). *Inteligencia artificial, más allá de la ficción: BMW SLP*. <https://www.clusterindustrial.com.mx/Noticia/2789/Inteligencia-Artificial-Mas-Alla-de-La-Ficcion-Bmw-Slp>

- Cognex. (2021). *Sistemas de visión in-sight 7000*. <https://www.cognex.com/es-mx/products/machine-vision/2d-machine-vision-systems/in-sight-7000-series>
- Costa, M. J. R., Gouveia, R. M., Silva, F. J. G., & Campilho, R. D. S. G. (2018). How to solve quality problems by advanced fully-automated manufacturing systems. *International Journal of Advanced Manufacturing Technology*, 94(9–12), 3041–3063. <https://doi.org/10.1007/s00170-017-0158-8>
- DEMAK. (2021). *Foam gaskets (FIPFG)*. <https://demakgroup.com/en/equipment/foam-gaskets-fipfg>
- Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64–73. <https://doi.org/10.1145/2500499>
- Duomly. (2019, March 3). *Which programming language is the best for machine learning and artificial intelligence*. [Blog]. <https://www.blog.duomly.com/which-programming-language-is-the-best-for-machine-learning-and-artificial-intelligence/>
- Endesa Fundación. (2020). *Electromagnetismo*. . <https://www.fundacionendesa.org/es/recursos/a201908-que-es-el-electromagnetismo>.
- Espinosa Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería Investigación y Tecnología*, 21(1), 1–13. <https://doi.org/10.22201/fi.25940732e.2020.21n1.008>
- Fernandez-Palacin, F., Lopez-Sanchez, M. A., & Muñoz-Márquez, M. (2018). Stepwise selection of variables in dea using contribution loads. *Pesquisa Operacional*, 38(1), 31–52. <https://doi.org/10.1590/0101-7438.2018.038.01.0031>
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, Á., & Padilla, W. (2018). *Ciencia de Datos. Técnicas analíticas y aprendizaje estadístico en un enfoque práctico*. <https://www.alfaomega.com.mx/default/editoriales/altaria/ciencia-de-datos-tecnicas-analiticas.html>
- Gonzalo, Á. (2020, June 20). *¿Qué es el sobreajuste u overfitting y por qué debemos evitarlo?* *Machine learning, data science y analítica avanzada*.

<https://machinelearningparatodos.com/que-es-el-sobreajuste-u-overfitting-y-por-que-debemos-evitarlo/>.

Green, S., Nabhani, F., & Bradley, G. (2019). *Reducing the scrap rate in an electronic manufacturing SME through Lean Six Sigma methodology* *Alireza Shokri, PhD Six Sigma in SCM, Senior lecturer in operations and supply chain management, Lean Six.
<https://core.ac.uk/download/pdf/46520216.pdf>

Gutiérrez, H., & Iturralde, K. (2017). *Fundamentos básicos de instrumentación y control: serie de textos académicos de la Facultad de Ciencias de la Ingeniería para estudiantes de ingeniería. Universidad Estatal Península de Santa Elena.*
<https://www.fnmt.es/documents/10179/10666378/Fundamentos+b%C3%A1sico+de+instrumentaci%C3%B3n+y+control.pdf/df746edc-8bd8-2191-2218-4acf36957671>

Hagedorn, J., Sell-Le Blanc, F., & Fleischer, J. (2018). *Handbook of coil winding technologies for efficient electrical wound products and their automated production.* <https://link.springer.com/book/10.1007/978-3-662-54402-0>

Hernández, D. (2018). *Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones.* *Revista Latinoamericana de Hipertensión* 37(5) (Vol. 37).
https://www.revistaavft.com/images/revistas/2018/avft_5_2018/25sobre_uso_adecuado_coeficiente.pdf

Howe, D. (2000). Magnetic actuators. *Sensors and Actuators A: Physical*, 81(1-3). In *Sensors and Actuators* (Vol. 81). www.elsevier.nl/locatersna

Hu, F. (2015). *How the magnetic field works atomic physics view project.*
<https://www.researchgate.net/publication/287205872>

Ismail, A., Truong, H. L., & Kastner, W. (2019). Manufacturing process data analysis pipelines: a requirements analysis and survey. *Journal of Big Data*, 6(1), 1–26. <https://doi.org/10.1186/s40537-018-0162-3>

KAIDI ELECTROMAGNETIC. (2019). *Do you really know solenoid?*
<http://www.solenoidsupplier.com/do-you-really-know-solenoid/>

Kane, V. E. (1986). *Process capability indices.* *Journal of Quality Technology*. 18(1), 41–52. <https://doi.org/10.1080/00224065.1986.11978984>

Kocabicak, Z. K., & Demir, U. (2020). Design and optimization of an electromechanical actuator for the latch of a foldable vehicle seat. *Materials Testing*, 62(7), 749–755. <https://doi.org/10.3139/120.111539>

- Iacobucci, A., & Robert, C. (2012). *The art of R programming: a tour of statistical software design by norman matloff* (l. no starch press, Ed.). <https://doi.org/10.1080/09332480.2012.685374>
- Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*. Wiley. <https://www.wiley.com/en-us/Discovering+Knowledge+in+Data%3A+An+Introduction+to+Data+Mining%2C+2nd+Edition-p-9780470908747>
- Lee, H. R., Ahn, J. H., & Kim, H. Y. (2016). Design of a solenoid actuator for a cylinder valve in a fuel cell vehicle. *Applied Sciences (Switzerland)*, 6(10). <https://doi.org/10.3390/app6100288>
- Lee, J., Lapira, E., Bagheri, B., & Kao, H. an. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 1(1), 38–41. <https://doi.org/10.1016/j.mfglet.2013.09.005>
- Li, Jie, Wang, Z. L., Zhao, H., Gravina, R., Fortino, G., Jiang, Y., & Tang, K. (2016). Networked human motion capture system based on quaternion navigation. *BodyNets International Conference on Body Area Networks*, 0(0). <https://doi.org/10.1145/0000000.0000000>
- Li, Jundong, Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: a data perspective. In *ACM Computing Surveys* (Vol. 50, Issue 6, pp. 1–45). Association for Computing Machinery. <https://doi.org/10.1145/3136625>
- Li, Jundong, Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature Selection. *ACM Computing Surveys*, 50(6). <https://doi.org/10.1145/3136625>
- Minitab. (2020). *Interpretar los resultados clave para gráfica de CDF empírica*. <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/graphs/empirical-cdf-plot/interpret-the-results/key-results/>
- Montgomery, D. C. (2004). *Diseño y análisis de experimentos* (L. Wiley, Ed.; Segunda Ed, Vol. 2). <https://libreria-limusa.com/producto/disen-y-analisis-de-experimentos-2a-ed/>
- ORACLE. (2020). *¿Qué es la ciencia de datos?* <https://www.oracle.com/mx/data-science/what-is-data-science/>
- Panasonic. (2020). *HG-T Sensor Láser de barrera*. <https://www.panasonic-electric-works.com/es/hg-t-sensor-laser-de-barrera.htm>
- Pei, Y., & Zaiane, O. (2006). *A synthetic data generator for clustering and outlier anayisis*. ERA. <https://doi.org/10.7939/R3B23S>
- Peter Hirt. (2020). *T521P*. <https://www.peterhirt.ch/en/products/t521p>

- Resnick, R., Halliday, D., & Walker, J. (1998). *Fundamentals of physics: Vol. Vol 1* (John Wiley, Ed.). <https://www.wiley.com/en-us/Fundamentals+of+Physics+Extended%2C+10th+Edition-p-9781118230725>
- Restrepo, L., & González, J. (2007). *De Pearson a Spearman*. . *Revista Colombiana de Ciencias Pecuarias* 20(2),. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-06902007000200010&lng=en&nrm=iso
- Rogalewicz, M., & Sika, R. (2016). Methodologies of knowledge discovery from data and data mining methods in mechanical engineering. *Management and Production Engineering Review*, 7(4), 97–108. <https://doi.org/10.1515/mper-2016-0040>
- Scheugenpflug. (2007). *Metering and dispensing technology*. *Direct Industry*. <https://pdf.directindustry.com/pdf/scheugenpflug-gmbh/metering-dispensing-systems/35264-611690.html>
- Steiner, S. H., & Mackay, R. J. (1998). *Understanding process capability indices*. <https://www.researchgate.net/publication/2497218>
- TLXTechnologies. (2020). *Solenoid fundamentals: how solenoids work*. <https://www.youtube.com/watch?v=Sq-CYfp9t4c>
- TOX PRESSOTECHNIK. (2020). *Accesorios para el TOX®-Powerpackage*. <https://es.tox-pressotechnik.com/productos/actuadores/actuadores-neumohidraulicos/accesorios/>
- Usharani, R., & Shanthini, A. (2020). *Machine learning approaches for predicting patient severity levels in T2dm complications neuropathy*. *European Journal of Molecular & Clinical Medicine*, 7(10), 2818-2833.
- Ververidis, D., & Kotropoulos, C. (2005). *Sequential forward feature selection with low computational cost*. *2005 13th European Signal Processing Conference*, 1-4. IEEE
- Wadhwa, D. (2019). *How to improve your productivity with AI and machine learning?*. *Big data made simple*. <https://bigdata-madesimple.com/how-to-improve-your-productivity-with-ai-and-machine-learning/>
- Wang, S., Weng, Z., & Jin, B. (2020). *A performance improvement strategy for solenoid electromagnetic actuator in servo proportional valve*. *Applied Sciences (Switzerland)*, 10(12). <https://doi.org/10.3390/app10124352>
- Wirth, R., & Hipp, J. (2000). *CRISP-DM: towards a standard process model for data mining*. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 29-39.

https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining

Zhang, H., & Senkara, J. (2011). *Resistance welding: fundamentals and applications*. Routledge & CRC Press.

<https://www.routledge.com/Resistance-Welding-Fundamentals-and-Applications-Second-Edition/Zhang-Senkara/p/book/9781138075245>

ANEXO

Ligas a código utilizado para el análisis.

A continuación se enlistan las ligas con las direcciones donde se encuentran los algoritmos utilizados durante el desarrollo del estudio:

- Limpieza de Datos:
 - <https://github.com/Erik1454/Tesis-/blob/Data-Analysis-Assembly-Process-Modeling-by-a-Multiple-Regression/LimpiezaDeDatos>
- Descripción de Base de Datos:
 - <https://github.com/Erik1454/Tesis-/blob/Data-Analysis-Assembly-Process-Modeling-by-a-Multiple-Regression/DescribirCleanData>
- Análisis de Correlación
 - <https://github.com/Erik1454/Tesis-/blob/Data-Analysis-Assembly-Process-Modeling-by-a-Multiple-Regression/PearsonCorrelationAnalysis>
- Test de Normalidad Shapiro Whiks
 - <https://github.com/Erik1454/Tesis-/blob/Data-Analysis-Assembly-Process-Modeling-by-a-Multiple-Regression/Test%20Shapiro-Whiks>
- Test de Colinealidad
 - <https://github.com/Erik1454/Tesis-/blob/Data-Analysis-Assembly-Process-Modeling-by-a-Multiple-Regression/Test-Colinealidad>