

SISTEMA INTELIGENTE PARA EL PRONÓSTICO DE ESCENARIOS DE
COSTOS Y FALLAS DE MANTENIMIENTO UTILIZANDO MACHINE LEARNING
COMO HERRAMIENTA GERENCIAL PARA UNA PLANTA DE BENEFICIO DEL
GREMIO PALMERO COLOMBIANO

CESAR AUGUSTO DÍAZ RANGEL

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS
ESCUELA DE INGENIERÍA MECÁNICA
MÁGISTER EN GERENCIA DE MANTENIMIENTO
BUCARAMANGA

2019

SISTEMA INTELIGENTE PARA EL PRONÓSTICO DE ESCENARIOS DE
COSTOS Y FALLAS DE MANTENIMIENTO UTILIZANDO MACHINE LEARNING
COMO HERRAMIENTA GERENCIAL PARA UNA PLANTA DE BENEFICIO DEL
GREMIO PALMERO COLOMBIANO

CESAR AUGUSTO DIAZ RANGEL

TRABAJO DE GRADO PARA OPTAR EL TÍTULO DE
MÁGISTER EN GERENCIA DE MANTENIMIENTO

DIRECTOR

DAVID ALFREDO FUENTES DÍAZ

PhD. Tecnología energética

CODIRECTOR

JESÚS ALBERTO GARCÍA NÚÑEZ

PhD. Ingeniería biológica y agrícola.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO MECÁNICAS
ESCUELA DE INGENIERÍA MECÁNICA
BUCARAMANGA

2019

DEDICATORIA

A mi familia por ser el motor de desarrollo y motivación diaria para los nuevos retos.

A mis padres por sus pilares morales.

A Cenipalma por su apoyo metodológico.

Al área de gestión humana de la federación por su voto de confianza.

AGRADECIMIENTOS

Especial reconocimiento a mi familia por su paciencia y vital apoyo durante las largas jornadas de trabajo.

A Cenipalma por el apoyo metodológico, científico y por el tiempo otorgado para el desarrollo de este proyecto.

A la planta de beneficio piloto del gremio palmero colombiano, por brindar la información de alto valor para el proyecto.

A la Universidad Industrial de Santander, especialmente al staff de apoyo del programa de maestría, por su paciencia y apoyo durante la especialización y maestría.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN	18
1 GENERALIDADES DEL PROYECTO	20
1.1 PLANTEAMIENTO DEL PROBLEMA.....	20
2 OBJETIVOS.....	22
2.1 OBJETIVO GENERAL	22
2.2 OBJETIVOS ESPECÍFICOS.....	22
3 JUSTIFICACION DEL PLAN PROPUESTO	23
4 ANÁLISIS DE LITERATURA RECOPIADA	25
4.1 MARCO TEÓRICO	25
4.1.1 Centro de investigación en palma de aceite, Cenipalma.	25
4.1.2 Contexto operacional.....	27
4.1.3 Estándar de procesos y productos obtenidos en planta de beneficio	30
4.1.4 Panorama de producción mundial de cosechas oleaginosas y costos de producción.	34
4.1.5 EGE: indicador clave de desempeño para el monitoreo de cambios en la disponibilidad de planta, rendimiento de equipos y calidad del proceso (Stamatis, 2010).	37
4.1.6 Diagnóstico y oportunidades de mejora detectadas en plantas de beneficio en el país.	43

4.1.7	Definición de estructura de costos de mantenimiento en una planta de beneficio.	50
4.1.8	Ciencia de los datos o Data Science aplicado a la industria.....	53
4.1.9	Definiciones de las principales áreas dentro de la Ciencia de los Datos: Machine learning, Deep learning, Data Mining y Big Data.	54
4.1.10	Bibliografía localizada acerca de machine learning y análisis de fenómenos tipo series de tiempo	67
4.1.11	Metodologías de pronóstico aplicadas a series de tiempo.	69
4.2	Marco conceptual	74
4.3	Marco legal y normativo.....	75
5	METODOLOGÍA.....	76
5.1	Plan de trabajo.....	76
5.2	Cronograma de actividades	79
6	PRESUPUESTO.....	80
7	RESULTADOS DEL PROYECTO	81
7.1	Especificación de datos de ingreso y de respuesta para los modelos de pronóstico	81
7.1.1	Análisis de normalidad para variables de respuesta del modelo	84
7.1.2	Análisis regresión de mejores subconjuntos.....	86
7.1.3	Análisis de correlación entre variables del modelo	88
7.2	Resultados de modelos estadísticos mediante regresión múltiple.....	90
7.2.1	Proceso de normalización de variables previo entrenamiento de red neuronal	95

7.3 Programación de aplicación inteligente para pronóstico de costos y tipos de fallas de mantenimiento a partir de información histórica.	96
7.3.1 Aplicación desarrollada en Matlab utilizando GUIDE y código M	96
7.3.2 Comparación modelos pronóstico regresión múltiple y red neuronal	111
7.4 Escenarios generados a partir del pronóstico de variables asociadas a costos de mantenimiento y eficiencia de procesos	112
7.4.1 Escenario 1: incremento de paradas y fallas, reducción en materia prima y de su calidad	112
7.4.2 Escenario 2: incremento de paradas y fallas, reducción en materia prima y de su calidad, descenso del precio internacional del aceite de palma (CIF Rotterdam)	114
8 CONCLUSIONES	117
BIBLIOGRAFÍA	119
ANEXOS	122

LISTA DE TABLAS

Tabla 1. Módulos de producción, servicios industriales, otras áreas y Biorrefinerías.	32
Tabla 2. Especificación de áreas de acuerdo con los módulos.	32
Tabla 3. Equipos y sistemas críticos por módulos.	34
Tabla 4. Referente de indicador EGE en otras industrias comparado con plantas de la agroindustria palmera.	41
Tabla 5. Tipos de funciones de activación neuronal.	55
Tabla 6. Descripción de configuraciones para redes neuronales.....	56
Tabla 7. Recursos bibliográficos disponibles.	67
Tabla 8. métodos por series de tiempo para pronósticos.	70
Tabla 9. Indicadores para medición de error estadístico.	71
Tabla 10. Presupuesto estimado para el proyecto.....	80
Tabla 11. Filtrado de variables con base en análisis de correlación y regresión múltiple.	81
Tabla 12. Resultados prueba de normalidad Anderson-Darling para variables de respuesta.	85
Tabla 13. Regresión de los mejores subconjuntos predictores.....	88
Tabla 14. Análisis de correlación total de variables.	90
Tabla 15. Resultados estadísticos regresión múltiple para variables de respuesta.	91
Tabla 16. Resultados estadísticos regresión múltiple para variables de respuesta.	111
Tabla 17. Estadísticos entre pronóstico regresión múltiple y red neuronal escenario 1.....	113
Tabla 18. Estadísticos entre pronóstico regresión múltiple y red neuronal escenario 2.....	115

LISTA DE FIGURAS

Figura 1. Estructura global de la federación	26
Figura 2. Estructura organizacional de Cenipalma.	27
Figura 3. Cadena productiva del aceite de palma.....	28
Figura 4. Panorama nacional de la palmicultura.	29
Figura 5. Productos y subproductos obtenidos a partir del fruto de la palma de aceite.	30
Figura 6. Esquema general de proceso en una planta de beneficio.	31
Figura 7. Producción mundial de cosechas oleaginosas.	35
Figura 8. Producción mundial de Aceite de Palma y participación por países.	36
Figura 9. Producción mundial de Aceite de Palma y participación por países.	36
Figura 10. Estructura de cálculo del indicador EGE y los indicadores base.	39
Figura 11. Producción mundial de Aceite de Palma y participación por países.	40
Figura 12. Conformación de EGE u OEE según Seis Grandes Pérdidas.	42
Figura 13. Muestra de plantas evaluadas a través del indicador EGE (OEE).	45
Figura 14. Indicadores específicos componentes del EGE.	45
Figura 15. Distribución de tiempo en paradas, fallas y tiempo de operación.	46
Figura 16. Análisis de causa raíz en plantas piloto: establecimiento de criticidad para paradas y fallas.....	47
Figura 17. Impactos sobre tiempo, velocidad y calidad del proceso.	48
Figura 18. Ejemplo de causa raíz identificada, objetos extraños y deficiencias en calidad de fruto en planta piloto.	49
Figura 19. Ambiente de intercambio entre metodologías y técnicas dentro de Data Science o Ciencia de los Datos.	53
Figura 20. Neurona biológica y el homólogo digital de red neuronal artificial.	54
Figura 21. Arquitecturas de configuración para redes neuronales.....	57
Figura 22. Proceso de desarrollo para Machine Learning.	58
Figura 23. Técnicas dentro de Machine Learning.	60

Figura 24. Aprendizaje supervisado.....	61
Figura 25. Aprendizaje no supervisado.....	62
Figura 26. Aprendizaje reforzado.....	63
Figura 27. Barcos cargueros tipo Panamax (Cap. Máx. 52 mil toneladas)	65
Figura 28. Diagrama metodología integrando ANN, FTA y FMEA.....	66
Figura 29. Pronósticos de temperatura de gases de exhosto.....	67
Figura 30. Tipos de métodos para pronósticos.	69
Figura 31. Configuración clásica de red neuronal NARX.....	73
Figura 32. Esquema planteado para desarrollo del proyecto.....	74
Figura 33. Esquema general del proyecto para el desarrollo del sistema de pronósticos.....	78
Figura 34. Cronograma de actividades.....	79
Figura 35. Histogramas para variables de respuesta.....	85
Figura 36. Respuesta modelo regresión múltiple Costos de Mantenimiento. En dichas gráficas, la variable base o de origen tiene línea continua color azul.....	92
Figura 37. Respuesta modelo regresión múltiple Eficiencia Global de Equipos (EGE)	92
Figura 38. Respuesta modelo regresión múltiple Tiempo paradas programadas mantenimiento.	93
Figura 39. Respuesta modelo regresión múltiple Tiempo paradas no programadas mantenimiento.	93
Figura 40. Respuesta modelo regresión múltiple Tiempo paradas programadas producción.	94
Figura 41. Respuesta modelo regresión múltiple Tiempo no programado producción.	94
Figura 42. Ventana principal sistema Matlab.....	97
Figura 43. Interfaz de desarrollo GUIDE en Matlab, entorno principal de aplicación.	97
Figura 44. Ventana principal de aplicación de pronóstico desarrollada.	98
Figura 45. Carga de archivo base de datos para entrenamiento de red neuronal.	99

Figura 46. Proceso de carga de base de datos completado.	99
Figura 47. Proceso de normalización de base de datos completado.	100
Figura 48. Menú de entrenamiento de red neuronal para elegir configuración predeterminada o específico por parámetro.	100
Figura 49. Ventana de configuración manual de red neuronal a entrenar.	101
Figura 50. Configuración de red neuronal para el proyecto.	103
Figura 51. Ventana de entrenamiento de red neuronal con indicadores de rendimiento, error, regresión, respuesta series de tiempo, entre otros.	104
Figura 52. Ventanas auxiliares durante entrenamiento de red neuronal.	105
Figura 53. Ventana de proceso de entrenamiento terminado con diagramas de redes en lazo abierto y lazo cerrado.	106
Figura 54. Configuración de red neuronal resultante con retroalimentación.	106
Figura 55. Ventana de proceso de resultados de entrenamiento por cada variable de respuesta.	107
Figura 56. Ventanas de gráficas para cada variable de respuesta, datos de entrenamiento contra respuesta modelo de pronóstico.	107
Figura 57. Comparación de pronóstico entrenamiento variable costos de mantenimiento.	108
Figura 58. Comparación de pronóstico entrenamiento variable Eficiencia Global de Equipos (EGE).	108
Figura 59. Comparación de pronóstico entrenamiento variable tiempo mantenimiento programado.	109
Figura 60. Comparación de pronóstico entrenamiento variable tiempo mantenimiento no programado.	109
Figura 61. Comparación de pronóstico entrenamiento variable tiempo programado de producción.	110
Figura 62. Comparación de pronóstico entrenamiento variable tiempo no programado producción.	110
Figura 63. Pronóstico de variables de respuesta ante nuevo escenario número 1. (Usando base de datos Anexo C)	112

Figura 64. Pronóstico de variables de respuesta ante nuevo escenario número 2.
(Usando base de datos Anexo D)114

LISTA DE ANEXOS

Anexo A: resultados regresión múltiple realizada con MINITAB	122
Anexo B: Base de datos para entrenamiento	133
Anexo C: Base de datos para validación escenario 1	133
Anexo D: Base de datos para validación escenario 2	133

RESUMEN

TÍTULO: SISTEMA INTELIGENTE PARA EL PRONÓSTICO DE ESCENARIOS DE COSTOS Y FALLAS DE MANTENIMIENTO UTILIZANDO MACHINE LEARNING COMO HERRAMIENTA GERENCIAL PARA UNA PLANTA DE BENEFICIO DEL GREMIO PALMERO COLOMBIANO¹

AUTOR: CESAR AUGUSTO DIAZ RANGEL²³

PALABRAS CLAVE: machine learning, redes neuronales y series de tiempo, costos de mantenimiento, eficiencia global de equipos (EGE), pronósticos.

DESCRIPCIÓN: La agroindustria del aceite de palma es el sector más representativo en el ámbito de cosechas oleaginosas a nivel internacional. La productividad industrial y estabilidad financiera han sido afectadas por diversos factores, entre los cuales se encuentra la falta de herramientas para el análisis de información de procesos para la toma de decisiones a nivel estratégico. Este proyecto tiene como propósito desarrollar una herramienta inteligente utilizando redes neuronales (machine learning) para el pronóstico de escenarios de costos y fallas de mantenimiento, la cual permitirá a la gerencia de planta extractora de aceite de palma y subproductos, estimar con mayor precisión el desempeño futuro a partir del análisis de parámetros operacionales y de costos para la gestión de producción y del mantenimiento, en conjunto con el comportamiento estacional de la oferta de fruto. Este proyecto tiene como fases: consolidación de bases de datos, exploración de patrones y extracción de características, selección de técnicas dentro de machine learning (redes neuronales y series de tiempo), entrenamiento de red neuronal y validación, generación de resultados a modo de posibles escenarios bajo condiciones reales de las plantas de beneficio.

¹ Trabajo de grado.

² Facultad de ingenierías fisicomecánicas. Escuela de ingeniería mecánica. Director: PhD. David Alfredo Fuentes Díaz.

³ Cenipalma. Programa de Procesamiento. Codirector: PhD. Jesús Alberto García Núñez.

ABSTRACT

TITLE: INTELLIGENT SYSTEM FOR ESCENARIOS FORECASTING OF COSTS AND MAINTENANCE FAILURES USING MACHINE LEARNING AS MANAGEMENT TOOL FOR A PALM OIL MILL OF THE COLOMBIAN OIL PALM FEDERATION.⁴

AUTHOR: CESAR AUGUSTO DIAZ RANGEL⁵⁶

KEYWORDS: machine learning, artificial neural network and time series, maintenance costs, overall equipment effectiveness (OEE), forecasting.

DESCRIPTION: The palm oil agroindustry is the most representative sector in the international field of oilseeds. The industrial productivity and financial stability have been affected by multiple factors, among which is the lack of tools for the analysis of process information for strategic decision making. The purpose of this project is develop an intelligent tool using neural networks (machine learning), to forecast of scenarios by costs and maintenance failures. This intelligent tool will provide to palm oil mill managers, more accurately to estimate future performance, based on the analysis of operational and cost parameters of production and maintenance process, considering the seasonal demand of oil palm fruit. This is the sections of this project: consolidation of databases, exploration of patterns and extraction of characteristics, selection of machine learning technique (neural networks and time series), neural network training and validation, generation of results as possible scenarios with real conditions of palm oil mills.

⁴ Master's Thesis.

⁵ Facultad de ingenierías fisicomecánicas. Escuela de ingeniería mecánica. Director: PhD. David Alfredo Fuentes Díaz.

⁶ Cenipalma. Programa de Procesamiento. Codirector: PhD. Jesús Alberto García Núñez.

INTRODUCCIÓN

Las plantas de beneficio demandan herramientas que les permitan tomar decisiones a nivel gerencial con el fin de incrementar la confiabilidad del negocio y por ende su rentabilidad, cuyo impacto deseado apunta hacia el incremento de la eficiencia de los procesos industriales ligados a la extracción de aceites y aprovechamiento de biomasa, entre otros frentes de negocios.

El siguiente proyecto propone el desarrollo de un sistema inteligente capaz de pronosticar escenarios de costos asociados a condiciones de producción y mantenimiento, permitiendo simular escenarios de desempeño para la planta piloto, bajo determinadas condiciones e indicadores previamente suministrados. La propuesta de valor está orientada hacia la construcción de un sistema para simulación de escenarios utilizando las herramientas y técnicas dentro de *machine learning*⁷, partiendo de datos existentes en planta de beneficio, transfiriendo los mismos a las etapas de consolidación de bases de datos, depuración y calidad de información, selección y aplicación de herramientas avanzadas, evaluación y validación, presentación de resultados en forma de escenarios de costos de mantenimiento y fallas en ventanas temporales asociadas a la fluctuación y disponibilidad del fruto de palma de aceite para la planta (Bell, 2015).

Como hipótesis a sustentar, se establece que la implementación de una herramienta para el pronóstico de escenarios de costos y fallas de mantenimiento, le permitirá a la gerencia estimar con mayor precisión el desempeño de la planta de beneficio, a partir del análisis del comportamiento estacional de la oferta de fruto en la zona junto con los parámetros operacionales de los procesos y gestión del mantenimiento.

⁷ Machine learning: aprendizaje automático o aprendizaje de máquinas. Es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. (Nilsson, 2016)

Este documento detalla el desarrollo del proyecto, iniciando con el marco teórico compuesto por el contexto del procesamiento y beneficio a partir del fruto de la palma de aceite, Cenipalma y Fedepalma como dos organizaciones con el mismo ADN institucional. Siguiendo a lo anterior se especifican los resultados del diagnóstico global de productividad industrial de las plantas de beneficio en Colombia, principales indicadores estratégicos para la toma de decisiones a nivel gerencial y directivo, oportunidades de mejora y aspectos de alto interés para el proyecto, hasta llegar al estándar y especificación de una planta de beneficio a nivel general. En este nivel se detalla la métrica de costos por cada uno de los procesos productivos y de soporte para el procesamiento en planta.

Posteriormente se abre el nuevo horizonte sobre Ciencia de los Datos (Data Science), en donde se mencionan las herramientas más promisoras de los últimos 10 años, iniciando con machine learning, Deep learning, data Mining y finalmente, Big Data. Las diversas configuraciones de redes neuronales, formas de medir la eficiencia de pronóstico y otros aspectos de gran importancia para el diseño de sistemas de aprendizaje supervisado, residen en este apartado.

Finalmente, se desarrolla la sección sobre diseño, programación, entrenamiento y validación de la red neuronal para el pronóstico de los escenarios de costos. Estos escenarios fueron diseñados con base en situaciones hipotéticas al día de hoy, pero con gran posibilidad de llegar a cumplirse dada la variabilidad económica y política del país, la cual finalmente termina impactando en la cadena de la palmicultura nacional.

1 GENERALIDADES DEL PROYECTO

1.1 PLANTEAMIENTO DEL PROBLEMA

Las plantas de beneficio demandan herramientas que les permitan administrar la información de los procesos y equipos para el posterior aprovechamiento de los indicadores generados. Uno de los principales componentes diferenciadores de la agroindustria palmera radica en la disponibilidad y fluctuación de la materia prima, es decir, el fruto de la palma de aceite. De acuerdo con lo indicado en el diagnóstico realizado por Cenipalma (Díaz R. & García N., 2016), el indicador componente del OEE⁸ con mayor impacto es la disponibilidad de planta, afectada principalmente por la fluctuación del fruto en las plantaciones, fenómenos a su vez asociados a ciclos de cosecha (incluye tipo de material genético sembrado y edad), condiciones climáticas, enfermedades y plagas, condiciones edafológicas, y condiciones socio-económicas de la región.

Es por lo anterior que en el mercado de los sistemas CMMS⁹ y ERP¹⁰ disponibles para la agroindustria, no se encuentran aplicaciones para pronóstico de escenarios a la medida del contexto palmero, debido a que las estructuras de los sistemas existentes han sido generadas a partir de otras industrias (ej. Oil&Gas) y sistemas de manufactura (ej. bebidas, alimentos, vehículos), sectores industriales en donde se cuenta con mayor control y confiabilidad en lo referente a materia prima para sus procesos.

Una solución flexible y con gran potencial de crecimiento iría más allá de gestionar indicadores, permitiendo a los gerentes y directores de planta constituir sistemas

⁸ OEE: acrónimo de Overall Equipment Effectiveness, en español, Eficiencia Global de Equipos (EGE).

⁹ CMMS: acrónimo de Computerized Maintenance Management System, en español, Sistema computarizado para gestión de mantenimiento.

¹⁰ ERP: acrónimo de Enterprise Resource Planning, en español, Sistema de planificación de recursos empresariales.

expertos para la asistencia en la toma de decisiones de forma oportuna y precisa, especialmente para planear actividades de mantenimiento para las épocas valle o de baja oferta de fruto en las zonas palmeras. Por otra parte, las plantas de beneficio cuentan con datos altamente variados y heterogéneos, partiendo desde la cantidad de fruto procesado al día y múltiples parámetros que se determinan a partir del mismo (ej. proveedor, lote, enfermedades, fecha de cosecha, calificación de calidad, potencial de aceite, material genético sembrado, entre otros), indicadores de programación y uso del tiempo (tiempo programado, efectivo, paradas programadas), indicadores de mantenimiento (MTBF¹¹, MTTR¹², MTTF¹³, OEE, intervenciones y rutinas, inventarios y repuestos, costos detallados por rubros), hasta llegar a indicadores de gestión ambiental. Uno de los problemas más críticos para la agroindustria palmera en el país radica en el desconocimiento de herramientas para la gestión y análisis inteligente de la información al igual que la ausencia de estándares nacionales en lo referente a su gestión, principalmente en lo relacionado con sistemas productivos y mantenimiento de plantas. Es por esto que la propuesta de un sistema inteligente basado en machine learning con técnicas de redes neuronales con series de tiempo junto con el análisis estadístico, se convertiría en una potente línea base para desarrollar múltiples simulaciones, partiendo de la generación de escenarios de costos vinculando rubros de producción, mantenimiento y calidad del proceso, contribuyendo a la línea de respuesta para la toma de decisiones empresariales. Además, el desarrollo de este proyecto podría sentar las bases para otros estudios posteriores relacionados con la generación de sistemas basados en otras técnicas, por ejemplo, utilizando redes bayesianas, árboles de decisión, lógica Fuzzy, regresión múltiple, entre otras.

¹¹ MTBF: acrónimo de Mean Time Between Fails, en español, Tiempo Promedio entre Fallas

¹² MTTR: acrónimo de Mean Time Between Repairs, en español, Tiempo Promedio entre Reparaciones (sistemas reparables)

¹³ MTTF: acrónimo de Mean Time To Fail, en español, Tiempo Promedio para Fallar (sistemas no reparables)

2 OBJETIVOS

2.1 OBJETIVO GENERAL

Desarrollar un sistema inteligente para el pronóstico de escenarios de costos y fallas de mantenimiento, aplicando machine learning con integración de redes neuronales y análisis estadístico, con el fin de presentar escenarios de costos de mantenimiento como punto de partida para la toma de decisiones a nivel gerencial.

2.2 OBJETIVOS ESPECÍFICOS

- Construir un inventario bibliográfico y tecnológico con las diversas alternativas para pronóstico de escenarios, al igual que la consolidación y depuración de la base de datos con KPIs a partir del estado del arte sobre los indicadores de disponibilidad y de fruto por proveedor, gestión de producción y mantenimiento en el contexto agroindustrial palmero.
- Desarrollar un análisis de sensibilidad estadística entre los indicadores de desempeño y oferta estacional de fruto, teniendo en cuenta el impacto sobre los costos de mantenimiento y frecuencia de fallas.
- Desarrollar la red neuronal con series de tiempo ajustada según el análisis de sensibilidad y relación con los vectores respuesta (costos y fallas), realizando posteriormente el entrenamiento, evaluación y validación con la base de datos de KPIs previamente normalizada.
- Consolidar los escenarios de costos de mantenimiento y los eventos asociados durante la ventana de análisis para cada escenario, teniendo en cuenta la época y disponibilidad de fruto para la planta piloto.

3 JUSTIFICACION DEL PLAN PROPUESTO

Este proyecto tiene como principal fin el desarrollar una solución enfocada al pronóstico de escenarios con costos y fallas de mantenimiento para una planta piloto de extracción de aceite de palma, a partir del uso de las herramientas y metodologías contenidas dentro de *machine learning*, específicamente relacionadas con análisis estadístico de variables y pronóstico a través de redes neuronales con series de tiempo(Bell, 2015). La estructura propuesta parte desde la consolidación de bases de datos, depuración y calidad de información, selección y aplicación de herramientas avanzadas, evaluación y validación, presentación de resultados relacionados con costos de mantenimiento y fallas asociadas a cada escenario generado.

Es posible hacer el desarrollo del proyecto a través de R-Studio¹⁴ utilizando librerías y módulos tales como Tensorflow¹⁵ para la integración de redes neuronales y Shiny¹⁶ para el desarrollo de GUI¹⁷ o interfaz visual para usuarios. También es posible realizar el proyecto utilizando la integración de R en Visual Studio¹⁸, migrando las bases de datos hacia el motor de SQLServer¹⁹. Es posible crear un proyecto utilizando lenguaje .NET bajo modelo MVC²⁰ con la integración de SQLServer. Sin embargo, por cuestión de tiempo para desarrollo del proyecto, se escoge el ecosistema para desarrollo de aplicaciones Matlab.

¹⁴ R-Studio: es un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos.

¹⁵ TensorFlow: es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos.

¹⁶ Shiny: es un paquete de R-Studio para el diseño de interfaz de usuario vía web.

¹⁷ GUI: acrónimo de Graphical user interface, en español, Interfaz gráfica para usuarios.

¹⁸ Visual Studio: es un entorno de desarrollo integrado para sistemas operativos Windows. Soporta múltiples lenguajes de programación, tales como C++, C#, Visual Basic, inclusive lenguajes no nativos como R.

¹⁹ SQL Server: es un sistema de manejo de bases de datos del modelo relacional, desarrollado por la empresa Microsoft.

²⁰ MVC: acrónimo de Modelo Vista Controlador.

En lo referente a la programación de la red neuronal y desarrollo de la aplicación con interfaz de usuario, se utilizará Matlab debido a que facilitará la integración tanto de bases de datos (migradas directamente hacia su Workspace), como la programación de las aplicaciones en su interior. Uno de los productos principales del proyecto es la generación de la aplicación con la cual se puedan pronosticar condiciones de mantenimiento y producción, a partir del ingreso de bases de datos previamente depuradas por el usuario. Se utilizó código tipo m.file para cristalizar el algoritmo de red neuronal por definir, posteriormente se realizará la interfaz de usuario usando las herramientas GUI dentro del ecosistema Matlab. Como producto entregable también se generó una aplicación ejecutable en plataforma Windows, a partir de la aplicación MCC (empacador de código en ejecutables) de Matlab, encargada de empacar y crear instaladores para aplicaciones desarrolladas dentro del mismo entorno.

Finalmente, la aplicación de pronóstico junto con los escenarios generados, servirán como línea base para la toma de decisiones a nivel de ingeniería directamente en los procesos de la planta, al igual que permitirá tener mayor certeza y fundamento técnico al momento de definir el presupuesto y prioridades de la planta para los próximos años.

4 ANÁLISIS DE LITERATURA RECOPIADA

4.1 MARCO TEÓRICO

4.1.1 Centro de investigación en palma de aceite, Cenipalma.

El centro de investigación Cenipalma ha cumplido 25 años desarrollando investigación, brindando servicios especializados e integrando al gremio palmero colombiano. Por su parte, la Federación nacional de palmicultores, Fedepalma, en conjunto con representantes importantes del gremio y del gobierno nacional, dieron fundación al centro de investigación. Posteriormente, Fedepalma se encargaría de los asuntos gremiales y políticos del gremio. Cenipalma inicialmente se conformó por la necesidad del gremio palmicultor por solucionar los problemas de sanidad ambiental, específicamente relacionada con la enfermedad de pudrición de cogollo la cual tuvo su impacto más alto por la afectación alrededor de 70 mil Ha entre Santander y Tumaco (Cenipalma, 2016). Posteriormente se conformaron programas de investigación especializados en cada etapa de la cadena productiva del aceite de palma, aceite de palmiste, almendra y subproductos. En la Figura 1 se puede apreciar la conformación global de la federación.

Figura 1. Estructura global de la federación



Fuente: documentación institucional Fedepalma.

El centro de investigación está organizado a modo de programas de investigación, extensión, campos experimentales, servicios técnicos especializados y servicios compartidos. En la Figura 2 se sustenta la organización de Cenipalma, en donde el Programa de Procesamiento es el encargado de investigar, desarrollar y generar soluciones a nivel de plantas de beneficio en general, relacionado principalmente con la productividad y competitividad del negocio, biorefinería (aprovechamiento de biomasa), calidad y usos del aceite y otros productos, nutrición y salud humana, y la generación de valor agregado a partir de los principales productos generados a partir del beneficio del fruto de la palma.

Figura 2. Estructura organizacional de Cenipalma.



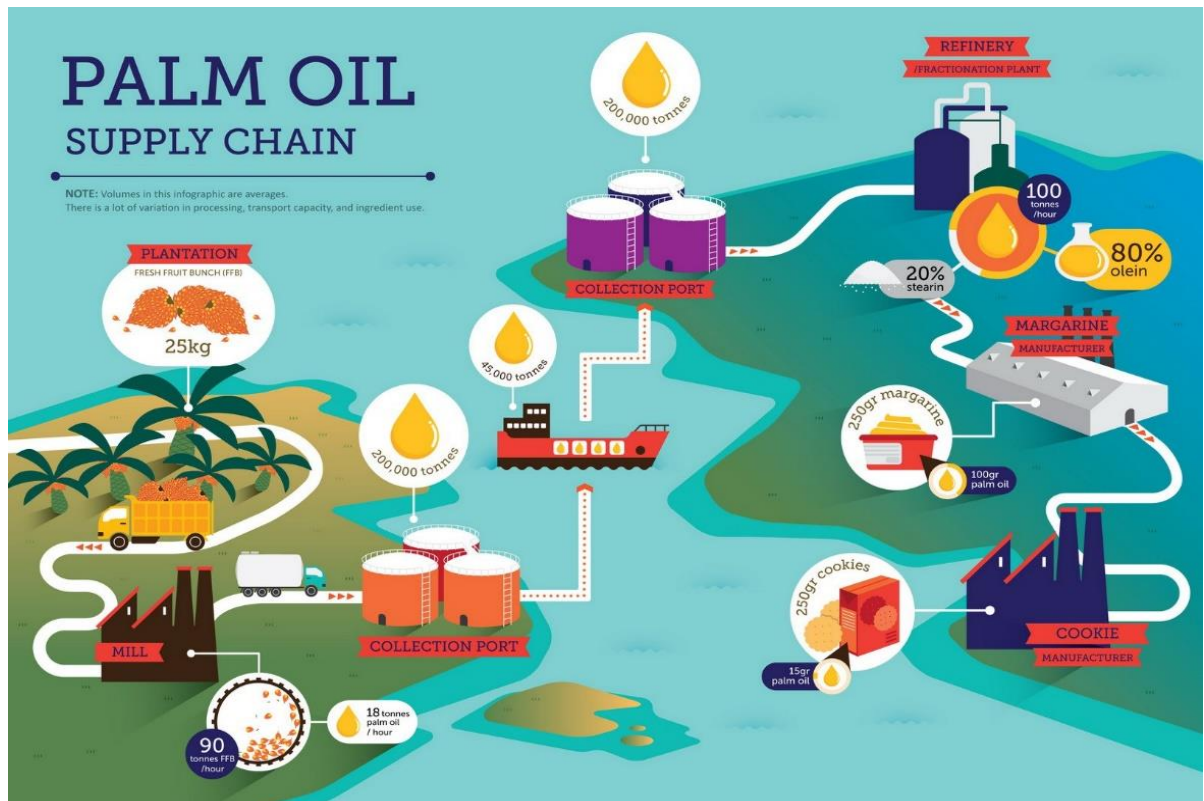
Fuente: documentación institucional Fedepalma.

4.1.2 Contexto operacional

El aceite de palma y aceite de palmiste son los principales productos generados por la palmicultura a nivel mundial. A nivel general, los productos oleaginosos y subproductos obtenidos del beneficio del fruto de la palma de aceite, parten desde el cultivo y cosecha en campo, etapa crucial para estos productos debido a que su cantidad de aceite por racimo y calidad final, dependen principalmente de las condiciones agroclimáticas, fertilización del suelo, prácticas agronómicas, entre otros factores que son motivo de estudio por parte de los programas de agronomía, biología y mejoramiento genético, entre otras áreas especializadas en Cenipalma,

Centro de investigación e innovación tecnológica en palma de aceite. En la Figura 3 se indican las etapas para obtención del aceite de palma.

Figura 3. Cadena productiva del aceite de palma



Fuente: RSPO Roundtable on Sustainable Palm Oil, 2017.

Una vez el fruto es cosechado, se transporta a las plantas extractoras, en donde se dará su beneficio para generar productos de alto valor tales como aceite de palma crudo, aceite de palmiste crudo, fibra y tusa. Una vez el aceite es extraído, se envía hacia plantas de refinación para obtención primaria de productos, conformando clientes nacionales e internacionales. A partir del aceite de palma y del aceite de palmiste se obtienen miles de productos, primordialmente a nivel alimenticio humano y animal, biocombustible (biodiesel) y subproductos, entre los cuales aparece la generación de energía eléctrica para su posterior interconexión con la red nacional de distribución, según el caso.

En la Figura 4 se detalla la distribución de plantaciones y extractoras en el país, separando las áreas a nivel de zonas palmeras las cuales integran departamentos, posteriormente a nivel de núcleos palmeros alrededor de las plantas de beneficio. En Colombia existen 4 zonas palmeras

Figura 4. Panorama nacional de la palmicultura.



Fuente: Cenipalma (Fedepalma, 2017).

4.1.3 Estándar de procesos y productos obtenidos en planta de beneficio

Una vez los racimos de fruto fresco (RFF) llegan a la planta de beneficio, se inician diversos procesos netamente mecánicos para la separación del fruto de los racimos vacíos (tusas), extracción del aceite crudo de palma, separación de nueces y fibra, recuperación de aceite, obtención de nuez y extracción de aceite de palmiste. En la Figura 5 se muestran los principales productos y subproductos generados en planta de beneficio.

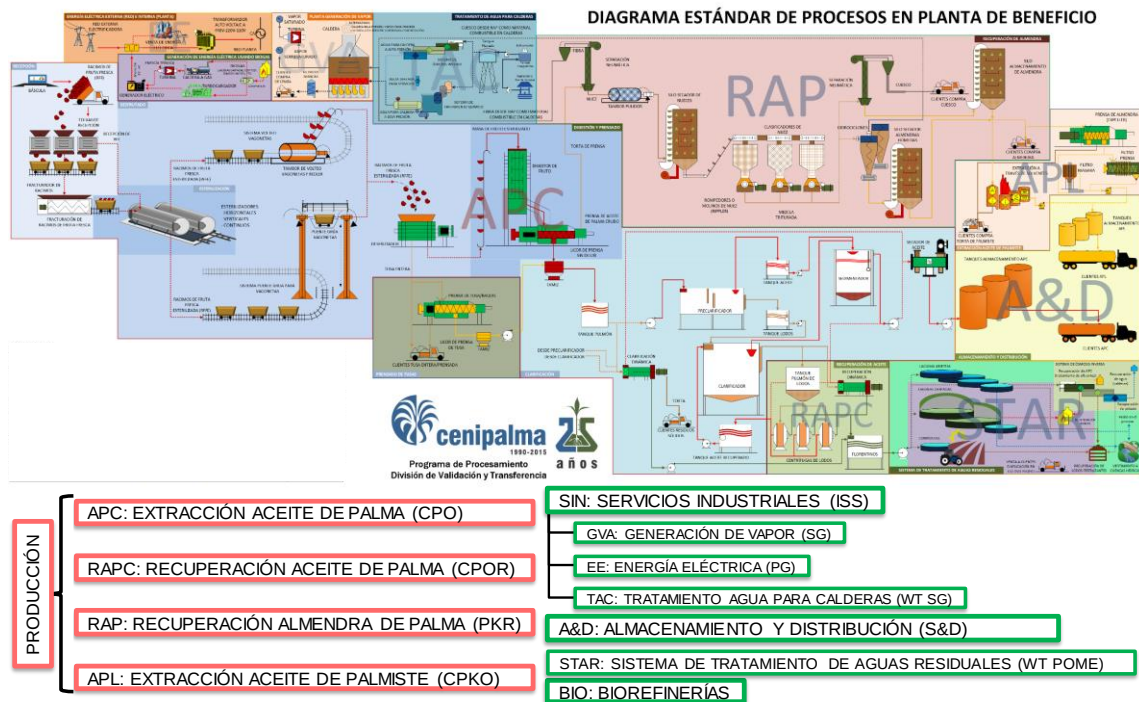
Figura 5. Productos y subproductos obtenidos a partir del fruto de la palma de aceite.



Fuente: Cenipama (Fedepalma, 2017).

Los procesos en planta de beneficio se pueden dividir a modo de módulos compuestos por áreas de trabajo (Figura 6), conformando unidades de proceso que intercambian materia prima y productos entre sí, áreas de servicios y soporte, sistemas relacionados con aprovechamiento mediante biorefinería (biomasa, biogás y subproductos).

Figura 6. Esquema general de proceso en una planta de beneficio.



Fuente: tomado de (Díaz R. & García N., 2016).

Las etapas que componen el procesamiento industrial de forma estándar en una planta de beneficio son:

- Recepción
- Esterilización
- Desfrutado
- Digestión
- Prensado
- Clarificación
- Recuperación de aceite
- Recuperación de nuez
- Extracción aceite de palmiste

En la Tabla 1 se especifican las definiciones de cada módulo, al igual que la delimitación y conformación interna a modo de áreas de proceso. Generalmente en

las plantas de beneficio, se registran indicadores de desempeño y eficiencia por procesos específicos, al igual que por producto y subproducto obtenido.

Tabla 1. Módulos de producción, servicios industriales, otras áreas y Biorrefinerías.

NOMBRE DEL MÓDULO		SIGLA EN INGLÉS	SIGLA EN ESPAÑOL	DEFINICIÓN DEL MÓDULO
1.	MODULO EXTRACCIÓN ACEITE DE PALMA CRUDO	CPO (Crude Palm Oil Extraction)	EAPC (Extracción de Aceite de Palma Crudo)	Módulo de producción encargado de la extracción de aceite de palma crudo a través de las prensas de RFF (Racimos de Fruto Fresco) digerido. En este módulo se incluyen los sistemas de clarificación dinámica (extracción de aceite)
2.	MÓDULO DE RECUPERACION DE ACEITE DE PALMA CRUDO	CPOR (Crude Palm Oil Recovery)	RAPC (Recuperación de Aceite de Palma Crudo)	Módulo de producción encargado de la recuperación de aceite de palma crudo a partir de los lodos usando centrifugas, equipos de recuperación dinámica y prensado de raquis/tusas, entre otros equipos para este fin. En este módulo se incluyen los sistemas de separación dinámica (recuperación de aceite y separación de lodos)
3.	MÓDULO RECUPERACION DE ALMENDRA DE PALMA	PKR (Palm Kernel Recovery)	RAP (Recuperación de Almendra de Palma)	Módulo de producción encargado de la recuperación de almendra a partir de la nuez que ha quedado en la torta separada previamente de la fibra a través de la columna neumática.
4.	MÓDULO EXTRACCIÓN ACEITE DE PALMISTE	CPKO (Crude Palm Kernel Oil Extraction)	EAPL (Extracción de Aceite de Palmiste)	Módulo de producción encargado de la extracción de aceite de palmiste a partir de la almendra previamente seca y limpia en los silos de almacenamiento.
5.	MÓDULO SERVICIOS INDUSTRIALES	ISS (Industrial Services and Support)	SIN (Servicios Industriales)	Módulo que comprende las áreas de Generación de Vapor, Energía Eléctrica, Tratamiento de Agua para Calderas, Tratamiento de Efluentes y Recuperación de Agua.
6.	MÓDULO DE BIOREFINERIAS	BIO (Biorefinery)	BIO (Biorefinería)	Módulo encargado de desarrollar escenarios de biorefinería: - Biogás - Biogás + cogeneración - Biogás + compostaje - Biogás + pellets - Biogás + pirólisis lenta (biocarbón) - Biogás + pirólisis rápida (bioaceites)
7.	MODULO ALMACENAMIENTO, DISTRIBUCIÓN Y COMERCIALIZACIÓN DE ACEITE DE PALMA CRUDO, ACEITE DE PALMISTE Y BIOMASA	S&D (Storage and Distribution)	A&D (Almacenamiento y Distribución)	Módulo encargado del almacenamiento, distribución y comercialización de APC, APL y biomasa (tusa, cuesco, fibra y torta para alimentación animal).

En la Tabla 2 a continuación se especifican las áreas que componen los módulos en planta de beneficio. Es importante tener en cuenta este estándar debido a que el proyecto especifica información e indicadores por cada una de estas áreas.

Tabla 2. Especificación de áreas de acuerdo con los módulos.

MÓDULOS		ÁREAS	SIGLA EN INGLÉS	SIGLA EN ESPAÑOL	DESCRIPCIÓN DE ÁREA
1.	APC	RECEPCIÓN	FFB RECEPTION (Fresh Fruit Bunches)	REC (Recepción)	Área que comprende la recepción de RFF previamente pesada en báscula.
2.		ESTERILIZACIÓN	STERILIZATION	EST (Esterilización)	Área encargada de la esterilización de RFF, ayudando a la separación de la nuez y almendra, facilitando la separación de los frutos de las tusas, y se reduce el efecto de la enzima lipasa encargada de promover la acidez y degradación del aceite a extraer.

3.		DESFUTADO	FRUIT SEPARATION	DES (Desfrutado)	Área encargada de desprender los frutos de las tusas o raquis.
4.		DIGESTIÓN Y PRENSADO	DIGESTION AND PRESSING	DPR (Digestión y Prensado)	Área encargada de digerir y prensar el fruto, obteniendo el licor de prensado (mezcla de aceite, agua, sólidos livianos y sólidos pesados).
5.		CLARIFICACIÓN	CLARIFICATION	CLA (Clarificación)	Área encargada de clarificar el licor de prensa a través de la clarificación estática o clarificación dinámica.
6.	RAPC	RECUPERACIÓN DE ACEITE DE PALMA CRUDO	CPOR (Crude Palm Oil Recovery)	RAPC (Recuperación de Aceite de Palma Crudo)	Área encargada de recuperar el aceite de palma crudo a partir de los lodos extraídos anteriormente.
7.		PRENSADO DE TUSA/RAQUIS	EFB PRESSING	PRT (Prensado de Tusas)	Área encargada de recuperar el aceite presente en las tusas/raquis.
8.	RAP	RECUPERACIÓN DE ALMENDRA DE PALMA	PKR (Palm Kernel Recovery)	RAP (Recuperación de Almendra de Palma)	Área encargada de recuperar la almendra a través del rompimiento de la nuez. Este proceso inicia partir de la separación neumática de la torta de prensado en nueces y fibra.
9.	APL	EXTRACCIÓN ACEITE DE PALMISTE	PKO (Palm Kernel Oil Extraction)	EAPL (Extracción de Aceite de Palmiste)	Área encargada de la extracción del aceite de palmiste.
10.		GENERACION DE VAPOR	SG (Steam Generation)	GVA (Generación de Vapor)	Área encargada de la generación de vapor para el proceso. Para el caso de las calderas a baja presión se usa el vapor para el proceso, en el caso de las calderas a alta presión el vapor además del proceso, se usa para cogeneración de energía eléctrica a través de energía térmica.
11.		ENERGÍA ELÉCTRICA INTERNA	PG-INSIDE (Power Generation at Palm Oil Mill)	EE INT (Energía Eléctrica Interna)	Área encargada del suministro de energía eléctrica para uso en la planta partiendo desde el transformador eléctrico.
12.		ENERGÍA ELÉCTRICA EXTERNA	PG-OUTSIDE (Power Generation from Electrical Company)	EE EXT (Energía Eléctrica Externa)	Servicio proveniente desde la empresa que suministra energía eléctrica en la región.
13.		TRATAMIENTO DE AGUA PARA CALDERAS	WT-SG (Water Treatment for Steam Generation)	TAC (Tratamiento de Agua Calderas)	Área encargada del tratamiento de agua requerida en condiciones físico-químicas específicas para su uso en las calderas.
14.	BIO	TRATAMIENTO DE EFLUENTES Y RECUPERACIÓN DE AGUA PARA VERTIMIENTOS	WT-POME (Water Treatment and Palm Oil Mill Effluents)	STAR (Sistema de Tratamiento de Aguas Residuales)	Módulo encargado del tratamiento de efluentes del proceso y recuperación de agua para vertimientos o reuso en el proceso.
15.		BIOREFINERÍAS	BIO (Biorefinery)	BIO (Biorefinería)	Área encargada de la obtención de beneficios a partir de los escenarios de biorefinería.
16.	MODULO ALMACENAMIENTO, DISTRIBUCIÓN Y COMERCIALIZACIÓN DE ACEITE DE PALMA CRUDO, ACEITE DE PALMISTE Y BIOMASA	ALMACENAMIENTO Y DISTRIBUCIÓN	S&D (Storage and Distribution)	A&D (Almacenamiento y Distribución)	Módulo encargado del almacenamiento, distribución y comercialización de APC, APL y biomasa (tusa, cuesco, fibra y torta para alimentación animal).

Fuente: documentación interna proyecto Plan Excelencia Industrial, Cenipalma (Díaz, Montero, Guevara, Cepeda, & Barrera, 2013).

El anterior estándar de procesos y productos por módulos y áreas de procesos, contienen equipos críticos designados de acuerdo al impacto generado a causa de paradas y fallas (Díaz et al., 2013). Para los siguientes equipos indicados en la Tabla 3, se generan indicadores de seguimiento dado el alto impacto a la continuidad del proceso, confiabilidad operacional, calidad del producto dado que el tiempo perdido

contribuye a la degradación química del aceite, impactando finalmente en la rentabilidad del negocio.

Tabla 3. Equipos y sistemas críticos por módulos.

No.	Equipo	Módulo	Área	Velocidad de Procesamiento	Materia Prima Procesada	Productos Obtenidos
1.	Prensa de APC	EAPC	Digestión y Prensado	Toneladas de RFF/h	Toneladas de RFF	Principal: toneladas de APC Secundario: tusas, torta de prensado
2.	Prensa de tusas/raquis	RAPC	Prensado de Tusas	Toneladas de Tusa/h	Toneladas de Tusa/Raquis	Principal: toneladas de APC recuperado. Secundario: tusas prensadas
3.	Centrífuga	RAPC	Recuperación de APC	m ³ de lodos/h	m ³ de lodos	Principal: toneladas de APC recuperado. Secundario: lodos y agua
4.	Separación dinámica	RAPC	Recuperación de APC	m ³ de lodos/h	m ³ de lodos	Principal: toneladas de APC recuperado. Secundario: lodos y agua
5.	Molinos rompedores de nuez/rippers	RAP	Recuperación de Almendra de Palma	Toneladas de Nuez/h	Toneladas de Nuez	Principal: toneladas de Almendra recuperada. Secundario: cuesco
6.	Prensa de APL/expellers	EAPL	Extracción de Aceite de Palmiste	Toneladas de Almendra/h	Toneladas de Almendra	Principal: toneladas de APL. Secundario: toneladas de torta de palmiste
7.	Sistema de extracción química de APL	EAPL	Extracción de Aceite de Palmiste	Toneladas de Almendra/h	Toneladas de Almendra	Principal: toneladas de APL. Secundario: residuos para tratamiento específico.

Fuente: documentación interna proyecto Plan Excelencia Industrial, Cenipalma (Díaz et al., 2013).

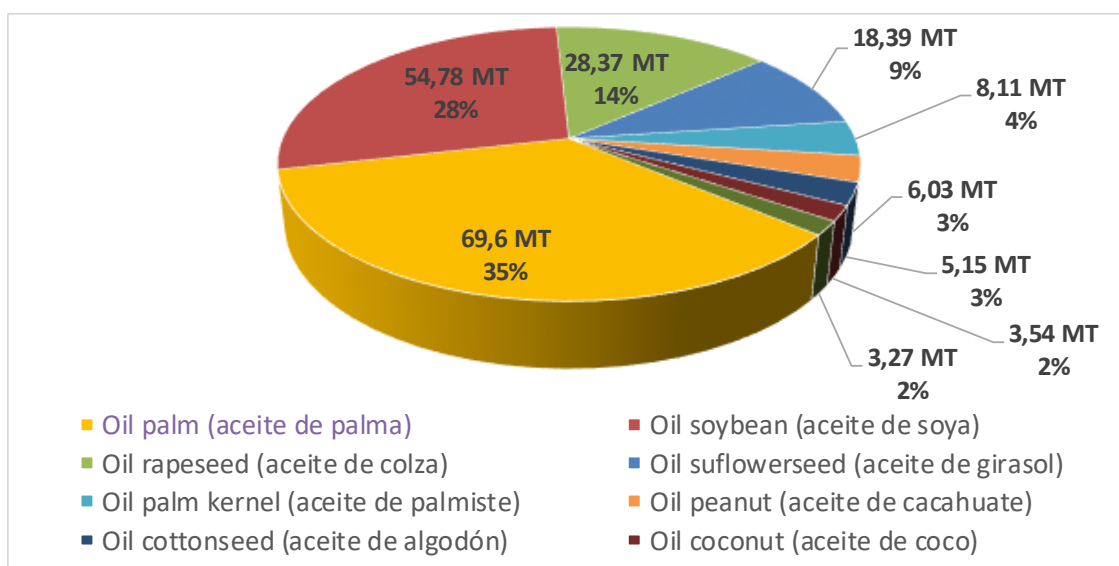
4.1.4 Panorama de producción mundial de cosechas oleaginosas y costos de producción.

La agroindustria oleaginosa a nivel mundial²¹, produce alrededor de 573 millones de toneladas anuales acumuladas a final del 2017 (USDA, 2018), posicionándose como uno de los productos con mayor participación en el grupo de cultivos. Dentro del grupo de cosechas oleaginosas y su posterior beneficio, el sector agroindustrial

²¹ En orden descendente en volumen de producción, los cultivos oleaginosos están compuestos por palma de aceite, soya, colza, girasol, palmiste (almendra de palma), cacahuete, algodón, coco y oliva.

de la palma de aceite y sus derivados, tienen al Aceite de Palma Crudo (APC)²² como el producto más comercializado con producción de 69,72 millones de toneladas anuales acumuladas a final del 2017, equivalentes al 35% de la producción mundial de productos oleaginosos (USDA, 2018). Por su parte el Aceite Crudo de Palmiste (APL), segundo producto de la cadena oleaginosa de la palma, ha registrado 8,11 millones de toneladas correspondientes al 4% como cuota en este grupo (USDA, 2018).

Figura 7. Producción mundial de cosechas oleaginosas.

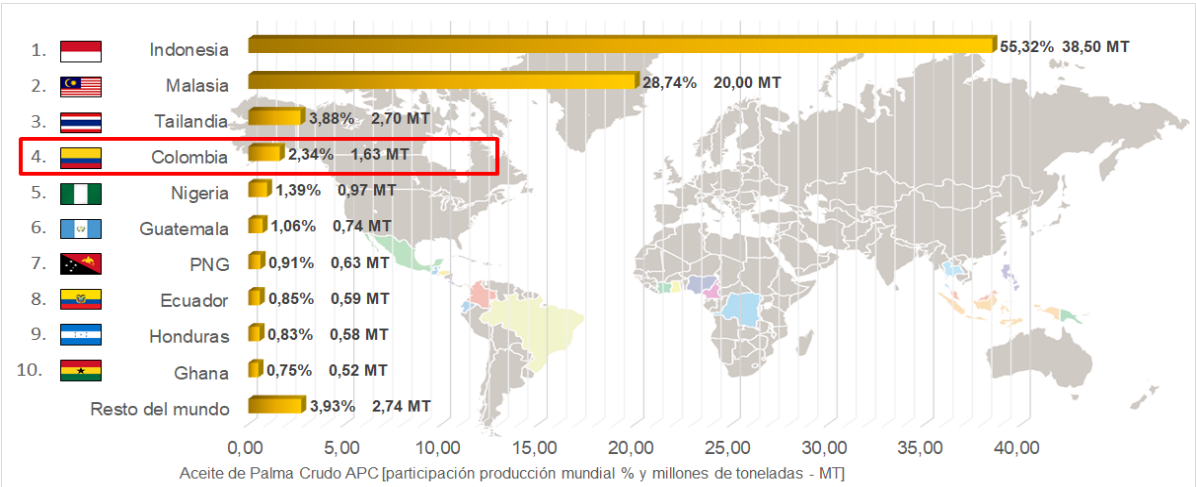


Fuente: indicadores anuales, USDA. (USDA, 2018)

En las plantas de beneficio en Colombia, al finalizar el año 2017, se han extraído alrededor de 1,6 millones de toneladas de APC (t APC) (Fedepalma, 2017), ubicando al país en la cuarta posición en el listado mundial de mayores productores de APC, cuyos referentes principales en igual escala de tiempo son Indonesia con 38,5 millones de toneladas y Malasia con 20 millones de toneladas de este producto (USDA, 2018) (Figura 8).

²² APC: abreviatura en español de Aceite de Palma Crudo, En inglés se conoce como CPO, Crude Palm Oil.

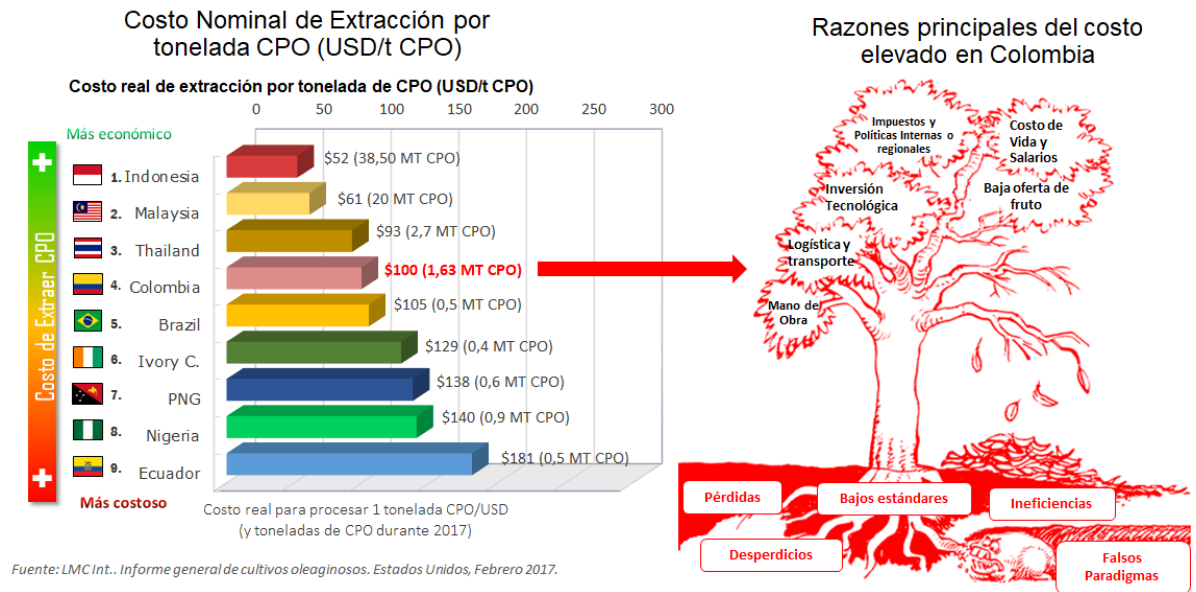
Figura 8. Producción mundial de Aceite de Palma y participación por países.



Fuente: tomado de (LMC International, 2017; USDA, 2018).

Uno de los indicadores más relevantes en la agroindustria y usado en el ejercicio de benchmarking es el costo nominal para procesar en planta de beneficio 1 t de APC, el cual para Colombia se encuentra alrededor de 100 USD (Mosquera, M., et al. 2016) (equivalente al 25% de los costos a través de la cadena productiva, cuyo valor total es de 398 USD/t APC), en contraste con los valores alcanzados por Indonesia y Malasia de 52 y 61 USD/t APC respectivamente (LMC International, 2017).

Figura 9. Producción mundial de Aceite de Palma y participación por países.



Fuente: LMC Int... Informe general de cultivos oleaginosos. Estados Unidos, Febrero 2017.

Fuente: ajustado de (Díaz R. & García N., 2016) e informe general cultivo oleaginosos LMC (LMC International, 2017).

La brecha de costos entre Colombia y los referentes asiáticos ha sido objeto de análisis por parte de Cenipalma, debido a que parte de las oportunidades de mejora diagnosticadas a nivel de producción y mantenimiento se ven reflejadas en este indicador. Partiendo del diagnóstico realizado por Cenipalma utilizando el indicador de Eficiencia Global de Equipos (EGE)²³(Díaz R. & García N., 2016), los resultados al 2016 con plantas piloto en las diversas zonas palmeras indicaron valores de OEE alrededor del 63%, valor influenciado por la estacionalidad del fruto en las zonas. El posterior análisis de causa raíz mostró múltiples oportunidades en cuanto a planeación y programación de actividades de producción y de mantenimiento (cumplimiento de planes por debajo del 60%), fallas y paradas no programadas (superior al 30% del tiempo de producción programado), variación de capacidad de procesamiento (inferior al 87% de capacidad nominal) y factores operacionales diversos (ausencia de estándares y prácticas). Actualmente algunas plantas se encuentran en proceso de mejoramiento continuo gracias al diagnóstico anterior, sin embargo, se ha generado la necesidad de contar con herramientas inteligentes para la toma de decisiones a nivel gerencial y directivo, basadas en datos históricos e integrando recursividad de sistemas expertos para la optimización de procesos de depuración, selección, análisis y generación de respuestas.

4.1.5 EGE: indicador clave de desempeño para el monitoreo de cambios en la disponibilidad de planta, rendimiento de equipos y calidad del proceso (Stamatis, 2010).

- ***Generalidades sobre el indicador***

El indicador EGE u OEE es una jerarquía de métricas que se centran en la eficacia con que se utiliza una operación de fabricación o sección del sitio. Los resultados

²³ OEE: acrónimo de Overall Equipment Effectiveness, en español, Eficiencia Global de Equipos.

se expresan en una forma genérica que permite la comparación entre unidades de fabricación en diferentes departamentos, organizaciones, máquinas e industrias. En esencia:

- EGE es una medida que identifica el potencial del equipo.
- EGE identifica y rastrea las pérdidas.
- EGE identifica ventanas de oportunidad.

Y su objetivo principal es aumentar la productividad:

- Disminuir el costo
- Aumentar la conciencia de la necesidad de la productividad de la máquina
- Aumentar la vida útil del equipo

Los resultados de estos objetivos son:

- Aumentar ganancias
- Obtener (o mantener) una ventaja competitiva
- Identificar la propiedad del equipo
- Reducir los gastos

Como medida, el OEE mide cómo se usa el equipo de capital efectivo mediante la identificación de restricciones y cómo las restricciones afectan al EGE. La efectividad se mide al multiplicar la disponibilidad y la eficiencia del rendimiento por la tasa de producción de calidad del producto. Los cálculos reales son:

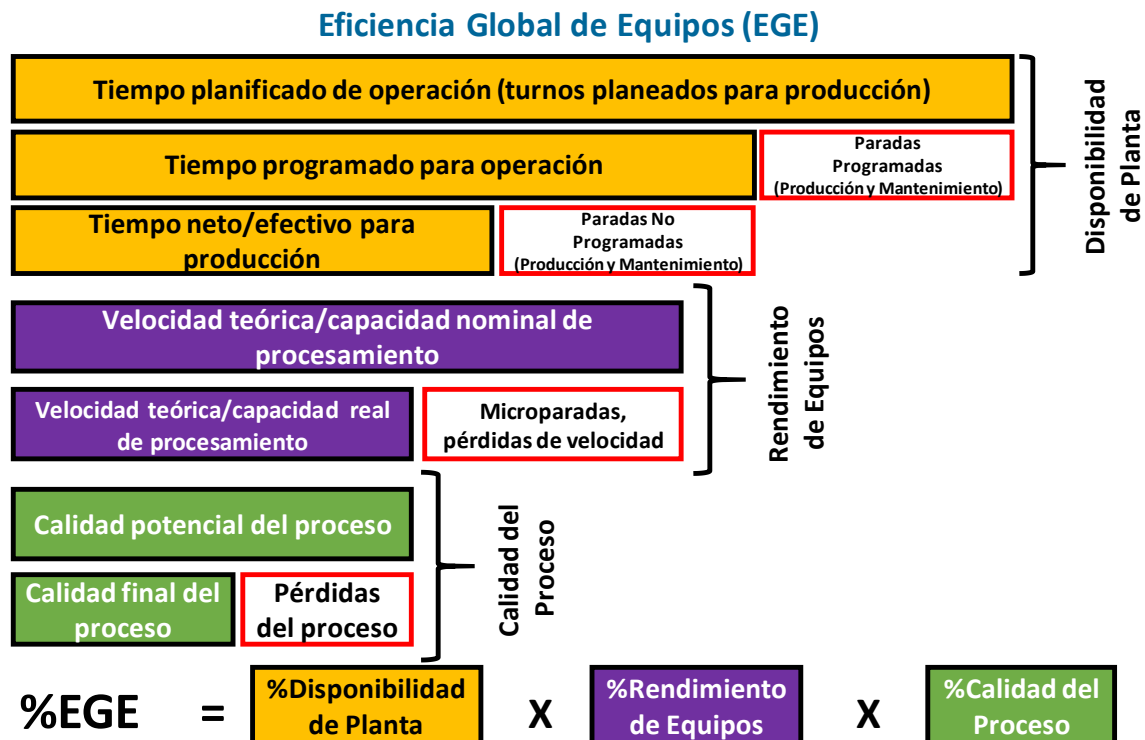
$$\begin{aligned} \text{Disponibilidad de Planta} &= \frac{\text{Tiempo Neto Producción}}{\text{Tiempo Planificado Operación}} \\ &= \frac{(\text{Tiempo disponible} - \text{Tiempo de inactividad})}{\text{Tiempo Planificado Operación}} \end{aligned} \quad (4-1)$$

$$\text{Rendimiento de Equipos} = \frac{\text{Velocidad o capacidad de equipos real}}{\text{Velocidad o capacidad de equipos nominal}} \quad (4-2)$$

$$\text{Calidad de Proceso} = \frac{\text{Unidades defectuosas o materia prima por fuera de estándar}}{\text{Unidades totales fabricadas o materia prima potencial posible}} \quad (4-3)$$

En la Figura 10 se muestra la estructura de cálculo para el indicador.

Figura 10. Estructura de cálculo del indicador EGE y los indicadores base.



Fuente: tomado de (Díaz R. & García N., 2016)

▪ **EGE comprende las Seis Grandes Pérdidas**

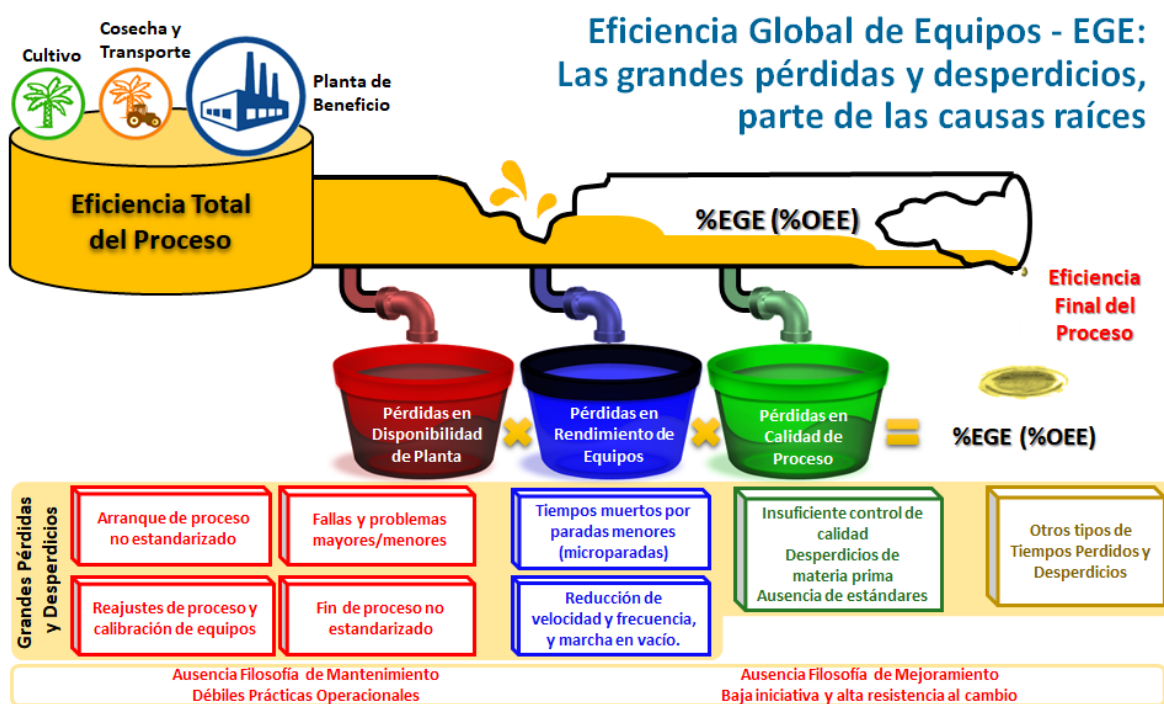
Las seis grandes pérdidas en equipos o máquinas en general son principalmente:

1. Fallas del equipo que producen pérdidas inesperadas de tiempo.
2. Ajuste de equipos, o tiempos muertos, que producen pérdidas de tiempo al iniciar una nueva operación. Por ejemplo, el cambio de formato para procesar otro tipo de materia prima.
3. Marchas en vacío, esperas y detenciones menores (fallas menores) ocurridas en el transcurso de la operación. Se producen pérdidas de tiempo, ya sea por problemas en la instrumentación, pequeñas obstrucciones, etc.
4. Velocidad de operación reducida, cuando el equipo no funciona a su capacidad máxima. Esto produce pérdidas productivas al no obtenerse la velocidad diseñada en el plan de producción.

5. Defectos en el proceso que producen pérdidas de producción al tener que rehacer partes de él, o al reprocesar productos defectuosos o completar actividades no terminadas.
6. Pérdidas de tiempo relacionadas con la puesta en marcha de un proceso nuevo, marcha en vacío, período de prueba, etc.

A continuación en la Figura 11 se muestra el esquema de grandes pérdidas dentro del concepto del indicador de Eficiencia Global de Equipos u OEE.

Figura 11. Producción mundial de Aceite de Palma y participación por países.



*EGE y OEE: se toma como producción máxima teórica en el tiempo planificado de producción (según sumatoria de turnos programados por la planta).
*TEEP: se toma como producción máxima teórica en el tiempo total calendario (24 horas * 365 días), usado en sistemas de manufactura.

Fuente: tomado de (Díaz R. & García N., 2016)

Se considera que un valor EGE de clase mundial es del 85% o superior. En la Tabla 4 se muestran los referentes de las industrias en general, hasta llegar a los resultados del diagnóstico realizado por Cenipalma, incluyendo los resultados consolidados por un autor en Malasia (Baluch, Abdullah, & Mohtar, 2012).

Tabla 4. Referente de indicador EGE en otras industrias comparado con plantas de la agroindustria palmera.

%EGE (%OEE)	Calificación	Sector Industrial	%OEE	Nivel de Competitividad
↑ 95%	Excelencia Clase Mundial	Industria Aeronáutica (fabricantes y ensambladores).	99,999999%	Excelente (Clase Mundial)
↕ 95%	Buena	Generadoras y distribuidoras de energía eléctrica.	99,999%	
↕ 85%		Industria Gas&Oil (cadena productiva completa).	98,987%	
↕ 85%	Aceptable	Industria productora de papel.	95,94%	Buena (Camino a Clase Mundial)
↕ 75%		Procesadoras de alimentos en general para consumo humano.	93,93%	
↕ 75%	Regular	Plantas de beneficio en Colombia, %EGE (APC*) promedio de 4 plantas con información en CeniSiC Cenipalma a 2016 (4 de 70 plantas existentes en el país).	49,50%	Inaceptable (Muy baja competitividad) 45 mil t APC ≈ \$36 millones USD
↕ 65%			(Meta inicial 75%) Se debe establecer según la planta Ver detalle EGE por planta a continuación	
↓ 65%	Inaceptable	Plantas de beneficio en Malasia, %EGE (APC) promedio de 41 plantas con información a 2011 según estudio de Nazim Baluch (373 plantas activas, del total 434 existentes).	62,00% (Meta 85%) Según literatura encontrada	Inaceptable (Muy baja competitividad) Pérdidas por ineficiencias alrededor de 7,05 billones** USD

*APC: Módulo Aceite de Palma Crudo o CPO, Crude Palm Oil.
 ** 1 billón USD = \$1,000,000,000 COP.
 *** N. BALUCH, Maintenance Management Performance of Malaysian Palm Oil Mills, University Utara Malaysia, 2012.

Fuente: tomado de (Díaz R. & García N., 2016).

La medición EGE también se usa comúnmente como un indicador de rendimiento clave (KPI) junto con los esfuerzos de fabricación ajustada para proporcionar un indicador de éxito. Esto se puede ilustrar mejor con una breve discusión de las seis métricas que componen el sistema. La jerarquía consta de dos medidas de nivel superior y cuatro medidas subyacentes. EGE y el rendimiento total del equipo efectivo (TEEP) son dos mediciones estrechamente relacionadas que informan la utilización general de las instalaciones, el tiempo y el material para las operaciones de fabricación. Estas métricas de vista superior indican directamente la brecha entre el rendimiento real y el ideal.

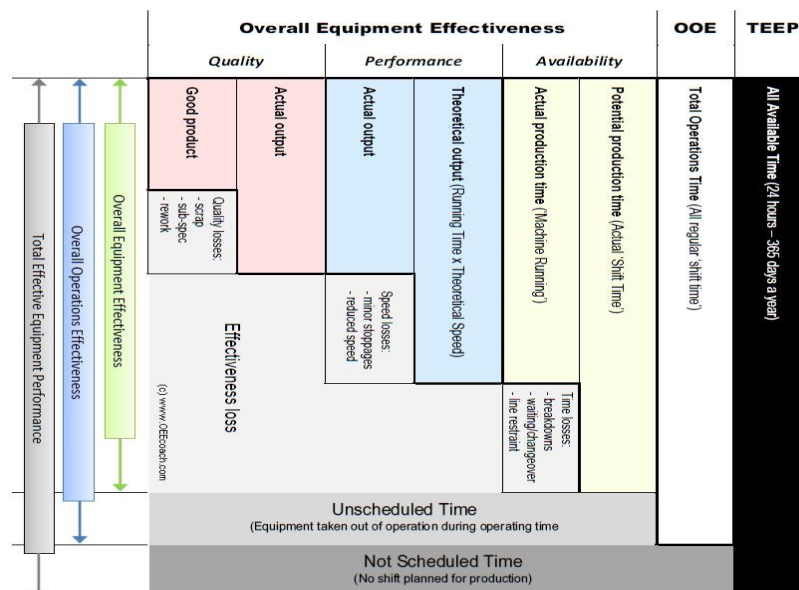
El EGE cuantifica el rendimiento de una unidad de fabricación en relación con su capacidad diseñada, durante los períodos en los que está programado su funcionamiento. Por otro lado, TEEP mide la efectividad de EGE en horas calendario, es decir, 24 horas por día, 365 días por año.

Además de las medidas anteriores, hay cuatro métricas subyacentes que proporcionan una comprensión de por qué y dónde existen las brechas de rendimiento EGE y TEEP. Estas medidas son:

- Tiempo: la parte de la métrica TEEP que representa el porcentaje del tiempo total del calendario que está realmente programado para la operación.
- Disponibilidad: la parte de la métrica EGE representa el porcentaje de tiempo programado para que la operación esté disponible para operar. A menudo se denomina tiempo de actividad.
- Rendimiento: la parte de la métrica EGE representa la velocidad a la que el área de trabajo se ejecuta como un porcentaje de su velocidad diseñada.
- Calidad: la parte de la métrica EGE representa las Unidades buenas producidas como un porcentaje de las Unidades totales iniciadas. Comúnmente conocido como Rendimiento de primer pase o Rendimiento de primer tiempo (FTT).

En la Figura 12 se muestra la composición para el indicador EGE y TEEP

Figura 12. Conformación de EGE u OEE según Seis Grandes Pérdidas.



Fuente: tomado de (Arno, 2015).

- **Cálculo del indicador EGE:**

El EGE divide el rendimiento de una unidad de fabricación en tres componentes separados pero mensurables: disponibilidad, rendimiento y calidad. Cada componente apunta a un aspecto del proceso que puede ser objeto de mejora. El EGE se puede aplicar a cualquier área de trabajo individual, o enrollarla a niveles de Departamento o Planta. Esta herramienta también permite profundizar para análisis muy específicos, como un número de pieza en particular, turno de trabajo o cualquiera de varios otros parámetros. Es muy poco probable que cualquier proceso de fabricación pueda ejecutarse al 100% del EGE. Muchos fabricantes comparan su industria para establecer un objetivo desafiante; 85% no es raro. La fórmula de cálculo para el EGE es:

$$OEE = Disponibilidad\ de\ Planta \times Rendimiento\ de\ Equipos \times Calidad\ del\ Proceso\ (4-4)$$

4.1.6 Diagnóstico y oportunidades de mejora detectadas en plantas de beneficio en el país.

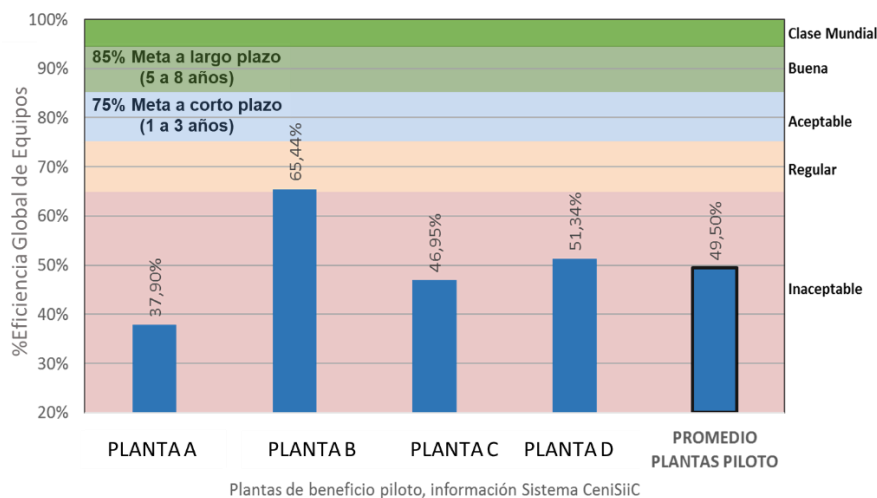
A partir del estudio sobre las condiciones durante el procesamiento del fruto de la palma de aceite llevado a cabo por Cenipalma desde el 2012, se utiliza la información cuyo diagnóstico parte desde el indicador de costo real por tonelada procesada de APC. Este indicador condensa efectos exógenos tales como normatividad y contexto social según el país, como aquellos que se han logrado identificar según las Seis Grandes Pérdidas del TPM para la etapa industrial. Esto ha sido uno de los objetivos del proyecto denominado Plan de Excelencia Industrial (PEI) desarrollado por Cenipalma, quienes, en conjunto con las plantas de beneficio, han llevado a cabo diversas investigaciones dirigidas hacia el diagnóstico de las causas raíces y la generación de soluciones a los problemas de alto impacto. Dentro de los estudios realizados, a partir del 2012 se han caracterizado alrededor de 14 plantas piloto alrededor del país (a 2017 existen 71 plantas de beneficio en Colombia), con relación al desempeño y eficiencia global de los procesos,

correlacionados a su vez con otros parámetros de referenciación internacional (Díaz R. & García N., 2016).

Dentro de los diversos KPIs estudiados, uno de los indicadores más representativos es el OEE, cuya estructura se conforma a través de la multiplicación porcentual de indicadores secundarios: Disponibilidad de planta (programación y aprovechamiento del tiempo para producción), Rendimiento de equipos (velocidad y capacidad de procesamiento), y Calidad del proceso (eficiencia y pérdidas de calidad durante el proceso). Dada la importancia para promover el mejoramiento continuo, este indicador ha sido usado y mejorado por organizaciones e industrias de Clase Mundial, cuyos estándares de eficiencia se encuentran por encima del 95%, estableciendo diversos rangos de calificación transversales a cualquier proceso productivo o de desarrollo.

De acuerdo con los resultados consolidados al corte del 2016, para las plantas piloto en Colombia el indicador OEE fue de 49.50%, cuyo decrecimiento respecto al 59.33% alcanzado en el 2015, fue parcialmente causada por la baja oferta de racimos de fruto fresco (RFF) provenientes desde las plantaciones, al igual que por la alta variabilidad climática y problemas relacionados con el manejo agronómico del cultivo (sanidad ambiental y aspectos de su biología). Por parte de la fase industrial y de procesamiento, el PEI se ha centrado en el diagnóstico e identificación de las causas raíces que son de alcance para las plantas, al igual que la planeación y programación de acciones encaminadas hacia la solución de los problemas encontrados, entre los cuales figuran la persistencia de falsos paradigmas.

Figura 13. Muestra de plantas evaluadas a través del indicador EGE (OEE).

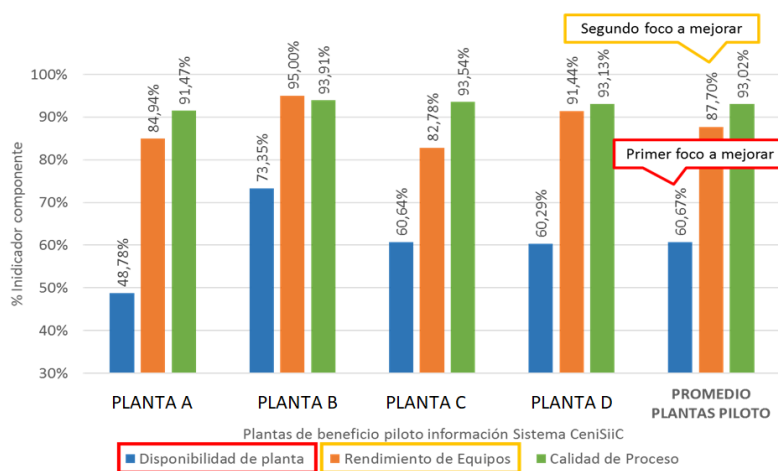


Fuente: tomado de (Díaz R. & García N., 2016)

Específicamente el valor de 49.50% del indicador OEE, para 4 plantas piloto evaluadas, estuvo conformado por la Disponibilidad de Planta con 60,67%, parámetro afectado por 8.185 horas de paradas (producción y mantenimiento), equivalentes al 39% de las 20.811 horas totales programadas para producción y desarrollo de sus operaciones. En la

Figura 14 se indican los resultados por cada una de las plantas piloto.

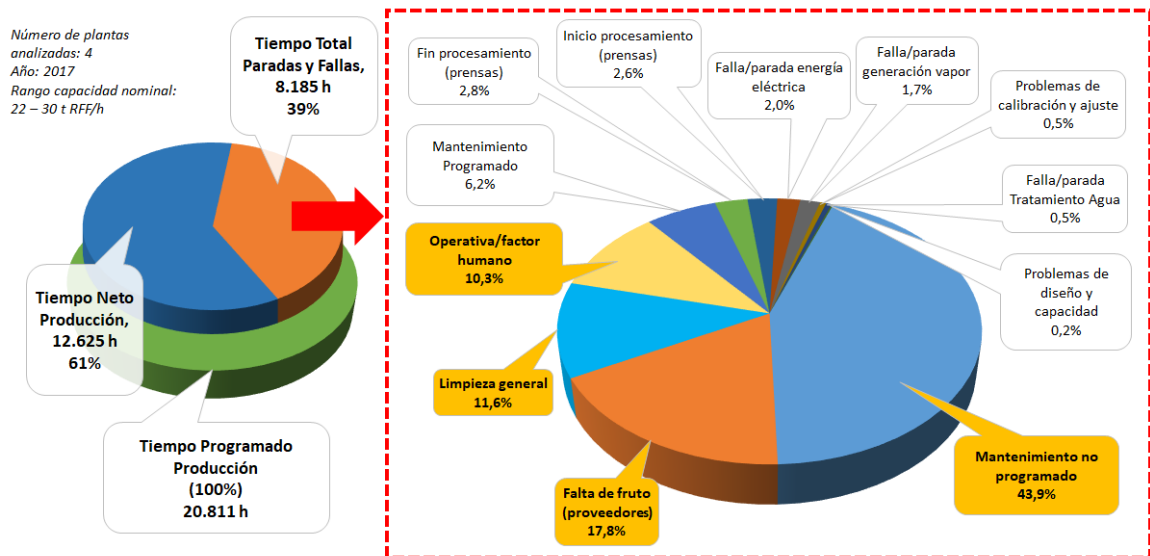
Figura 14. Indicadores específicos componentes del EGE.



Fuente: tomado de (Díaz R. & García N., 2016)

En la Figura 15 se especifica la segmentación de paradas y fallas, entre otras variables de interés, partiendo del tiempo programado para procesar, tiempo de paradas de producción y fallas de mantenimiento.

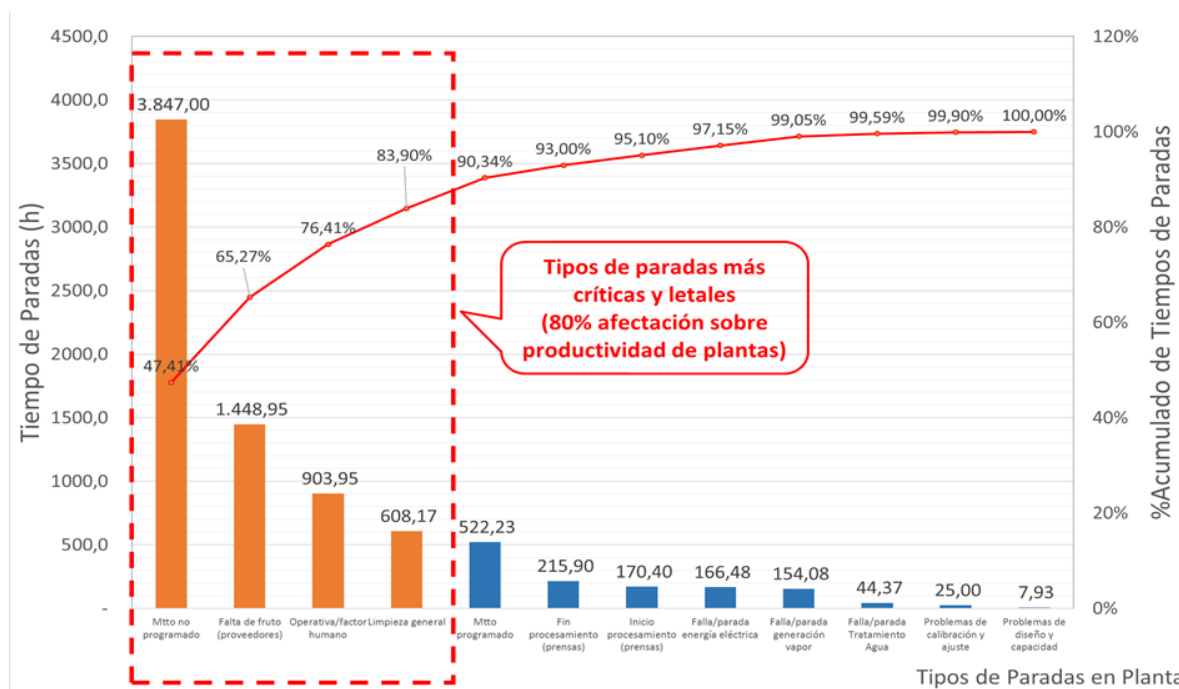
Figura 15. Distribución de tiempo en paradas, fallas y tiempo de operación.



Fuente: tomado de (Díaz R. & García N., 2016)

El análisis de causa raíz (ACR), ha identificado 4 tipos de paradas relacionadas con producción y mantenimiento, donde las paradas de producción fueron 3,743 horas (46% del tiempo total de paradas). En este grupo, los principales problemas se relacionaron con la insuficiencia de materia prima o falta de RFFs (18% del tiempo total de paradas con 1.459 horas), problemas operacionales y factor humano (10% del tiempo paradas con 840 horas), y limpieza y preparación de arranque del proceso (12% del tiempo total de paradas con 946 horas). Por otra parte, el tiempo perdido por fallas de mantenimiento fue de 4.442 horas (54% del tiempo total de paradas), siendo la parada con mayor frecuencia y criticidad identificada, el mantenimiento no programado con 3.594 horas (44% del tiempo total de paradas).

Figura 16. Análisis de causa raíz en plantas piloto: establecimiento de criticidad para paradas y fallas



Fuente: tomado de (Díaz R. & García N., 2016)

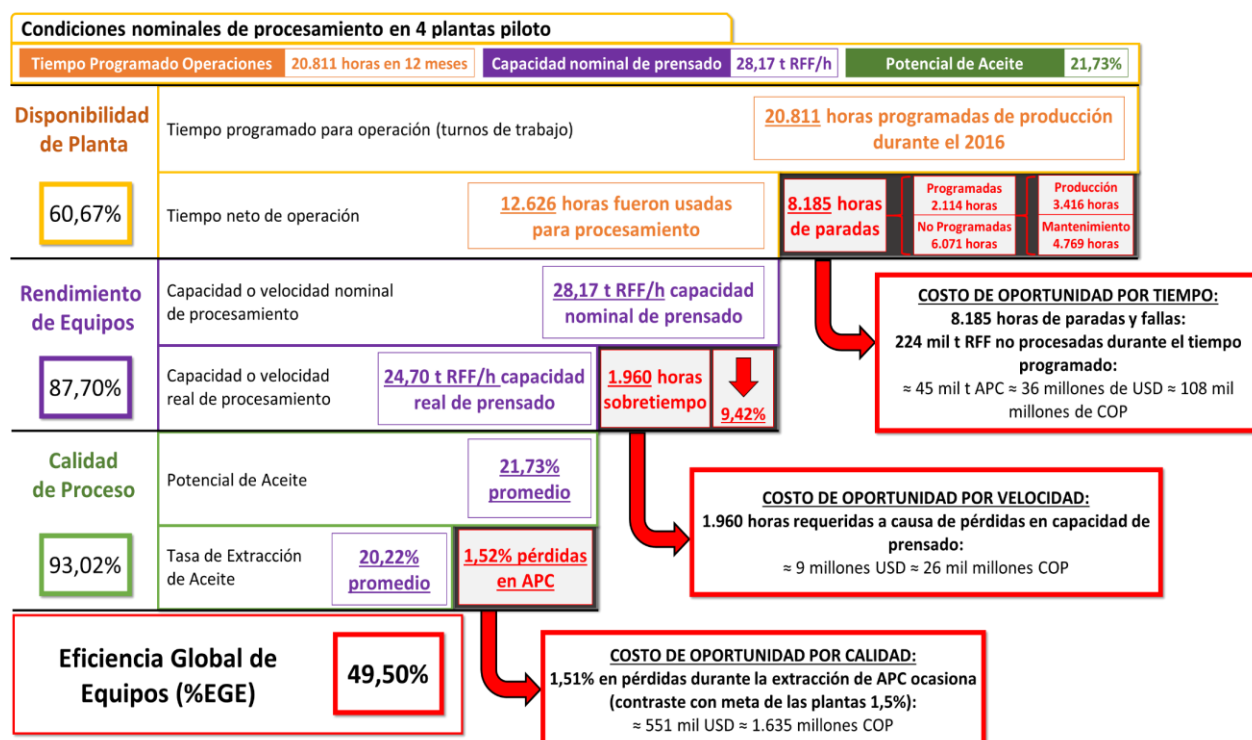
A nivel general, la acumulación de impactos por las paradas de producción y por las fallas de mantenimiento, ha generado el costo de oportunidad de 224.000 toneladas de RFF (equivalentes a 45.000 toneladas de APC no extraídas), materia prima que no fue procesada durante el tiempo programado de operación, cuya potencial comercialización equivaldría a 40 millones USD, ingreso para las 4 plantas piloto de la muestra.

La velocidad de procesamiento o el rendimiento de equipos del 87,70%, se ha visto afectado principalmente por las micro-paradas, cambios de frecuencia (según los valores nominales / de diseño), trabajo en vacío e inestabilidad en la capacidad de procesamiento. Estos problemas han causado como impacto negativo 1.604 horas adicionales (jornadas de trabajo extendidas y horas extras). Alrededor de 350.000 toneladas de RFF fueron procesadas por las plantas piloto de muestra, sin embargo, podrían haber procesado 586.000 toneladas de RFF, donde la producción potencial

de 237.000 toneladas de RFF podría haber sido aprovechada, y por consiguiente se hubiesen extraído alrededor de 50.000 toneladas de APC (7.5 millones de USD como potencial por comercialización).

De manera similar, el indicador de calidad del proceso fue superior al 90% de acuerdo con la información registrada por el sistema CeniSiiC®²⁴, en las 4 plantas del análisis, por lo que la prioridad del estudio se ha centrado en proporcionar soluciones a los problemas relacionados con la disponibilidad de planta y el rendimiento de equipos.

Figura 17. Impactos sobre tiempo, velocidad y calidad del proceso.



Fuente: tomado de (Díaz R. & García N., 2016)

²⁴ CeniSiiC: sistema integrado de información para las plantas de beneficio en Colombia, desarrollado por Cenipalma.

En la Figura 18 se muestra un ejemplo de las causas raíces identificadas en una de las plantas piloto analizadas. En este caso se puede observar la presencia de objetos extraños dentro de los cargamentos de RFF al llegar a la extractora, lo cual evidencia la ausencia de buenas prácticas agronómicas al momento de la cosecha y relacionado con la logística del fruto. El impacto generado sobre la confiabilidad de equipos, disponibilidad de planta y rentabilidad del negocio está alrededor de 500 mil USD durante el 2015.

Figura 18. Ejemplo de causa raíz identificada, objetos extraños y deficiencias en calidad de fruto en planta piloto.



Fuente: tomado de (Díaz R. & García N., 2016)

4.1.7 Definición de estructura de costos de mantenimiento en una planta de beneficio.

A continuación, se especifican los indicadores y rubros relacionados con costos por tonelada de RFF y por tonelada de Aceite, en el contexto de las plantas de beneficio en Colombia:

$$\text{Indicador 1} = \frac{\text{costo (COP)Proceso}}{\text{Ton RFF}} \quad (4-5) \quad \text{Indicador 2} = \frac{\text{costo (COP)Proceso}}{\text{Ton Aceite}} \quad (4-6)$$

Estos indicadores comprenden únicamente el proceso de extracción de APC el cual consta de:

1. Recepción.
2. Esterilización.
3. Desfrutación y prensado de tusas.
4. Digestión y prensado.
5. Clarificación.
6. Recuperación de almendra.
7. Servicios Industriales.

Los Indicadores están definidos de la siguiente manera:

$$\text{Indicador 1} = \frac{\text{Costo (COP)Proceso}}{\text{Ton RFF}} = \frac{C_1 + C_2 + C_3 + C_4 + C_5}{\text{Ton RFF}} \quad (4-7)$$

$$\text{Indicador 2} = \frac{\text{Costo (COP)Proceso}}{\text{Ton CPO}} = \frac{C_1 + C_2 + C_3 + C_4 + C_5}{\text{Ton CPO}} \quad (4-8)$$

Donde:

$C_1 = \text{Costo (COP) Mano de Obra}$

$C_2 = \text{Costo (COP) Energía}$

$C_3 = \text{Costo (COP) Mantenimiento}$

$C_4 = \text{Costo (COP) Otros}$

$C_5 = \text{Costo (COP) Transportes}$

Cada uno de los costos (C_1 a C_5), se deberán especificar de la siguiente manera:

▪ **COSTOS MANO DE OBRA (C_1):**

Comprende operarios desde la báscula hasta la recuperación de almendra, operarios de calidad, operarios de mantenimiento, jefes de turno en producción y en mantenimiento. Criterio general: Todo el personal de las secciones descritas que devenguen horas extras. El costo incluye prestaciones de ley.

▪ **COSTOS ENERGÍA (C_2):**

Definido como el costo de energía CPO (hasta recuperación de almendra) por tonelada de RFF y tonelada de aceite.

$$\frac{C_2}{\text{Ton RFF}} = \frac{C_G + C_R + C_{PE}}{\text{Ton RFF}} \quad (4-9) \qquad \frac{C_2}{\text{Ton CPO}} = \frac{C_G + C_R + C_{PE}}{\text{Ton CPO}} \quad (4-10)$$

Donde:

C_G = Costo (COP) Energía generada

C_R = Costo (COP) Energía Red Pública

C_{PE} = Costo (COP) Energía Planta de Emergencia

- Los costos para cada uno de estos ítems se calculan como:

$$\text{Energía (kW - hr) consumida} \times \frac{\text{costo (COP)}}{\text{kW - hr}} \quad (4-11)$$

- Se requiere especificar para cada planta el costo del kW/hr
- En cuanto al costo de energía generada por turbina, cada planta la calcula individualmente teniendo en cuenta que hay condiciones similares. Si se encuentran diferencias muy grandes se someterá a discusiones en ejercicios posteriores.

▪ **COSTOS DE MANTENIMIENTO (C_3)**

Los elementos que componen el costo por mantenimiento serán:

- Costos consumibles: Incluyendo lubricantes, soldaduras etc.

- Costos repuestos: Incluye materiales y herramientas.
- Incluye servicios externos y outsourcing

$$\frac{C_3}{Ton RFF} = \frac{C_c + C_{Rep} + C_s}{Ton RFF} \quad (4-12)$$

$$\frac{C_3}{Ton CPO} = \frac{C_c + C_{Rep} + C_s}{Ton CPO} \quad (4-13)$$

Donde:

$C_c = \text{Costo (COP) Consumible.}$

$C_s = \text{Costo (COP) Servicios.}$

$C_{Rep} = \text{Costo (COP) Repuestos.}$

▪ **COSTOS OTROS (C_4)**

En este ítem, se incluirían otros elementos o insumos necesarios para el buen desarrollo de la operación. A saber:

$$\frac{C_4}{Ton RFF} = \frac{C_{comb} + C_{IP} + C_{ILab} + C_{EL} + C_{EPP}}{Ton RFF} \quad (4-14) \quad \frac{C_4}{Ton CPO} = \frac{C_{comb} + C_{IP} + C_{ILab} + C_{EL} + C_{EPP}}{Ton CPO} \quad (4-15)$$

Donde:

$C_{comb} = \text{Costo (COP) Combustible}$

$C_{ILab} = \text{Insumos de Laboratorio}$

$C_{IP} = \text{Costo (COP) Insumos de Proceso}$

$C_{EL} = \text{Elementos de Limpieza}$

$C_{EPP} = \text{Elementos protec. personal}$

▪ **COSTOS TRANSPORTE (C_5)**

Contemplan todos los costos asociados al transporte de personal y subproductos.

$$\frac{C_5}{Ton RFF} = \frac{C_{Per} + C_{Sub}}{Ton RFF} \quad (4-16)$$

$$\frac{C_5}{Ton CPO} = \frac{C_{Per} + C_{Sub}}{Ton CPO} \quad (4-17)$$

Donde:

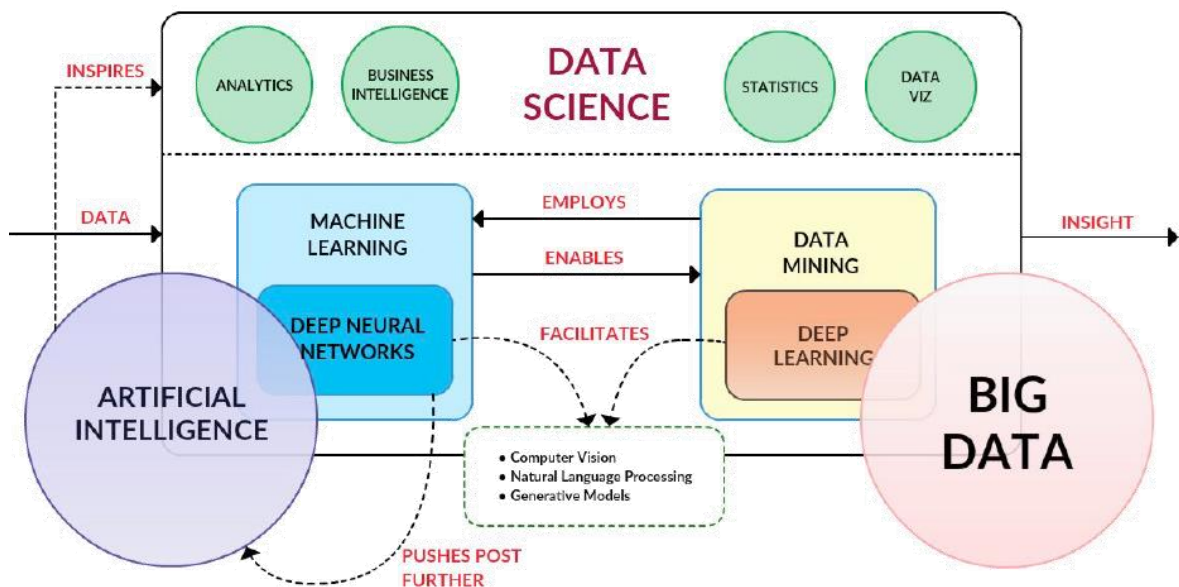
$C_{Per} = \text{Costo (COP) Personal.}$

$C_{Sub} = \text{Costo (COP) Subproductos.}$

4.1.8 Ciencia de los datos o Data Science aplicado a la industria

La ciencia de los datos o Data Science es un campo en crecimiento constante, el cual involucra métodos científicos, procesos y sistemas para extraer información a partir de datos de entrada, estructurados o no estructurados (Liu, 2015). La ciencia de los datos se concibe como la unificación de estadísticas, análisis de datos, aprendizaje automatizado y cualquier método relacionado para comprender y abstraer información de los fenómenos reales, a través de técnicas, metodologías y teorías adaptadas desde otros campos con mayores antecedentes como las matemáticas, estadística, ciencia de la información y de la computación. (Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., 1998). En la Figura 19 se aprecia el ambiente dentro del cual interactúan las diversas áreas que componen la ciencia de los datos.

Figura 19. Ambiente de intercambio entre metodologías y técnicas dentro de Data Science o Ciencia de los Datos.



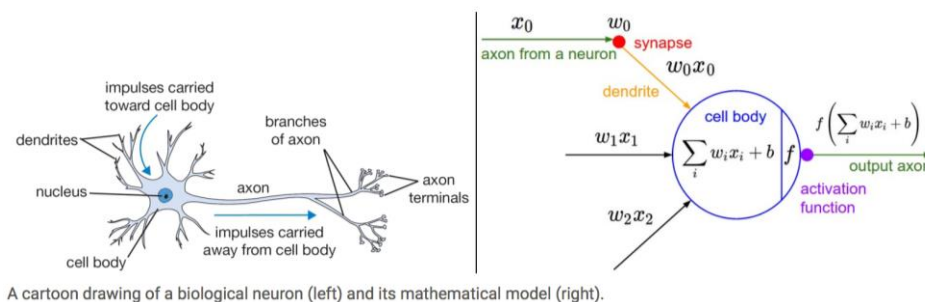
Fuente: Conceptual map of data science. Data science puzzle explained/2. Disponible en <http://www.kdnuggets.com/2016/03/data-science-puzzle-explained.html/2>. Noviembre 2018.

4.1.9 Definiciones de las principales áreas dentro de la Ciencia de los Datos: Machine learning, Deep learning, Data Mining y Big Data.

- **Redes neuronales**

Las redes neuronales artificiales son un modelo computacional inspirado en el comportamiento observado en la unidad biológica indispensable para el procesamiento de información: neurona (Figura 20). Estas redes están conformadas por diversas neuronas conectadas entre sí, con el fin de transmitir señales que posteriormente integrarán un dato específico, el cual previamente fue normalizado o estructurado a un nivel entre 1 y 0, con el fin de reducir la redundancia y pérdida de información por cifras decimales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones para tratamiento de señales) produciendo valores de salida, los cuales generalmente se reajustan a las condiciones reales o estudiadas a través de estas redes (Van Gerven & Bohte, 2017). Cada neurona está conectada con otras a través de enlaces, en donde el valor de salida de la neurona anterior es multiplicado un valor llamado peso. Estos pesos en los enlaces pueden incrementar o inhibir el estado de activación de las neuronas adyacentes. Del mismo modo, a la salida de la neurona, puede existir una función limitadora o umbral, que modifica el valor resultado o impone un límite que se debe sobrepasar antes de propagarse a otra neurona. Esta función se conoce como función de activación.

Figura 20. Neurona biológica y el homólogo digital de red neuronal artificial.

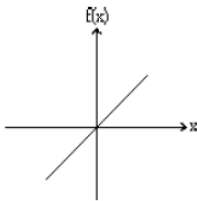
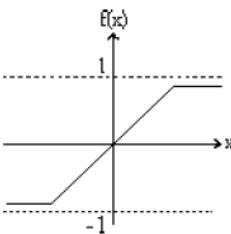
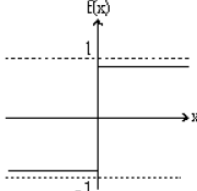
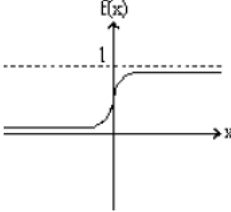
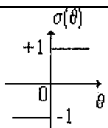
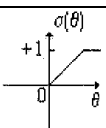
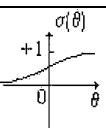
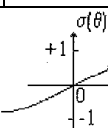


Fuente: tomado de Activation Functions in Neural Networks, disponible en https://isaacchanghau.github.io/post/activation_functions/. Febrero, 2019.

▪ **Funciones de activación**

Las funciones de activación se utilizan para mapear, escalar, normalizar o liminar el dominio de cada neurona, respecto a la información que procesará, verificando los efectos a la salida de la neurona.

Tabla 5. Tipos de funciones de activación neuronal.

Función	Representación matemática	Gráfica
Lineal: Generalmente se utiliza cuando se desea incrementar o decrecer rápidamente la proporción de los datos a analizar, dada la ganancia α .	$f(x) = \alpha * x$	
Rampa: Cuando una función lineal es limitada al rango $[-g, +g]$, se convierte en la función rampa, donde $g(-g)$ es el valor máximo y mínimo respectivamente que puede tomar la neurona. Esos valores son los que se refieren al nivel de saturación de la señal.	$f(x) = \begin{cases} +\gamma & x \geq \gamma \\ x & x < \gamma \\ -\gamma & x \leq -\gamma \end{cases}$	
Escalón: Esta función solamente responde a la señal de entrada, entregando $+g$ si la entrada es positiva y entregando $-g$ si es negativa, donde d y g son escalares positivos.	$f(x) = \begin{cases} +\gamma & x \geq 0 \\ -\delta & x \leq \text{cualquier otro} \end{cases}$	
Sigmoide: Esta es la función de activación más comúnmente utilizada, con forma de S, es acotada y no decreciente, provee una respuesta no lineal.	$f(x) = (1 + e^{-x})^{-1}$	
Combinación de funciones a partir de Sigmoide	 (a)  (b)  (c)  (d) a. Limitador duro c. Función logística b. Umbral lógico d. Tangente hiperbólica	

Fuente: tomado de (Van Gerven & Bohte, 2017)

- **Arquitecturas de configuración para redes neuronales**

En la Tabla 6 se indican los diversos tipos de configuración de redes neuronales.

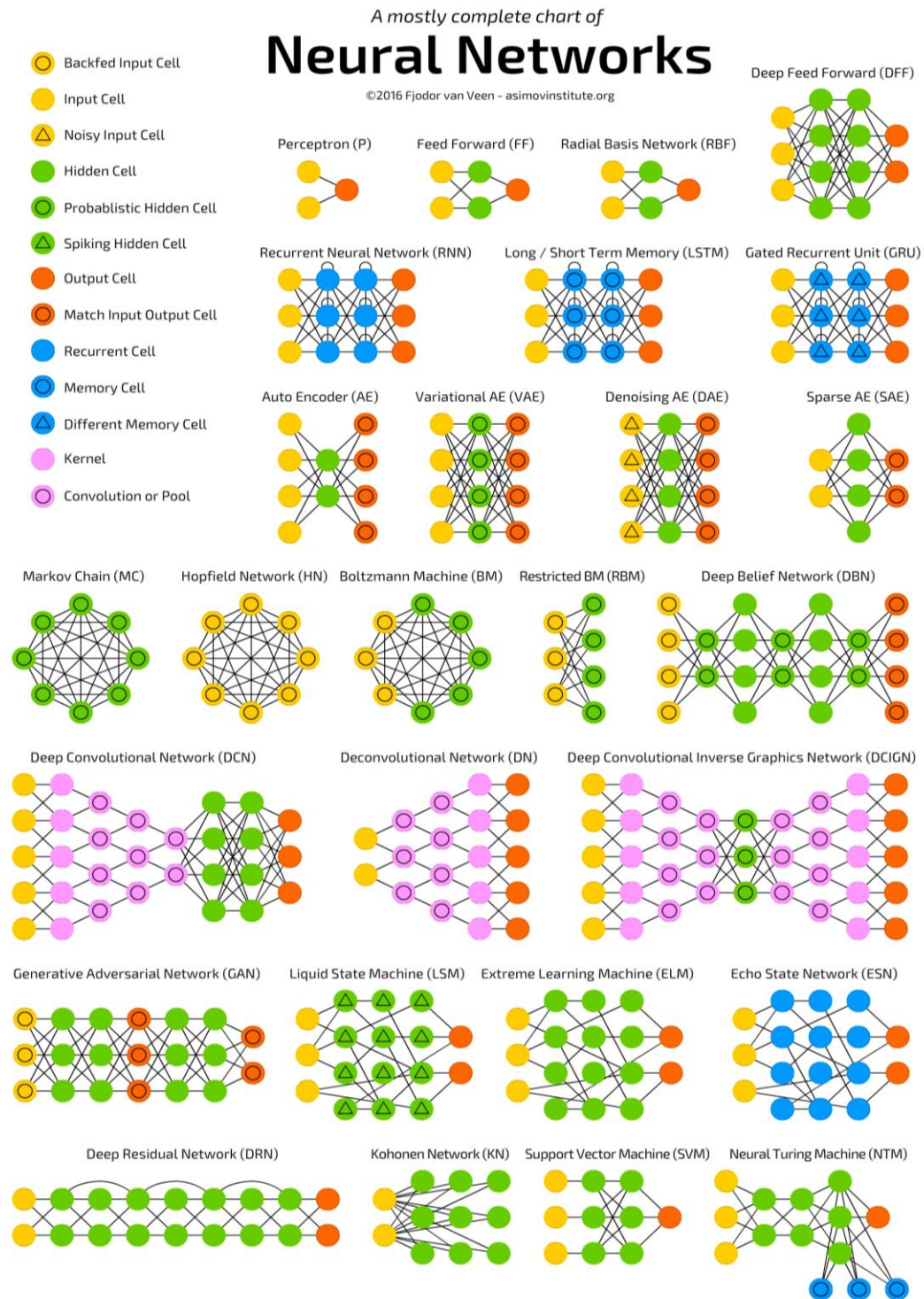
Tabla 6. Descripción de configuraciones para redes neuronales.

Tipo	Descripción
Alimentación hacia adelante y perceptrones	Es la red más simple, utilizada para modelar compuertas lógicas. Se utilizan en conjunto con otras redes más complejas.
Función de base radial	Se utilizan en conjunto con otras redes más complejas como unidad elemental.
Hopfield	Es una de las configuraciones más complejas por sus múltiples conexiones. Funcionan a modo de memoria asociativa similar a la humana. Se utilizan para reconocimiento de texto.
Cadenas de Markov en tiempo discreto	No tienen memoria a diferencia de otras redes (propiedad de Markov). Funciona similar a las cadenas de Markov.
Máquinas de Boltzmann	Funciona con asignación de pesos aleatorios y su aprendizaje se da a través de la propagación de errores hacia atrás. Los patrones de activación son binarios.
Auto codificada dispersa	A diferencia de la auto codificada, se expande el análisis en las capas ocultas. Estos tipos de redes se pueden usar para extraer muchas características de un conjunto de datos.
Las redes neuronales convolucionales profundas	Ampliamente utilizadas para el procesamiento de imágenes, pero también se pueden usar para otros tipos de entrada.
Redes de gráficos inversos convolucionales profundos	Estas redes intentan modelar características en la codificación como probabilidades, de modo que pueda aprender a producir una imagen, por ejemplo, un gato y un perro juntos, habiendo visto solo una de las dos en imágenes separadas
Máquinas neurales de Turing	En lugar de codificar una célula de memoria directamente en una neurona, la memoria se separa. Es un intento de combinar la eficiencia y la permanencia del almacenamiento digital regular y la eficiencia y el poder expresivo de las redes neuronales.
Las máquinas de estado líquido	Son un tipo de redes neuronales de punta, las activaciones sigmoideas se reemplazan con funciones de umbral y cada neurona es también una célula de memoria que se acumula.
Máquinas de vectores de soporte	Óptimas para problemas de clasificación.
Redes de Kohonen	Utilizan el aprendizaje competitivo para clasificar los datos sin supervisión.

Fuente: tomado de Azimov Institute. The Neural network zoo. Disponible en <http://www.asimovinstitute.org/neural-network-zoo/>. Febrero, 2019.

En la Figura 21 se muestran las diferentes configuraciones para las redes neuronales.

Figura 21. Arquitecturas de configuración para redes neuronales.

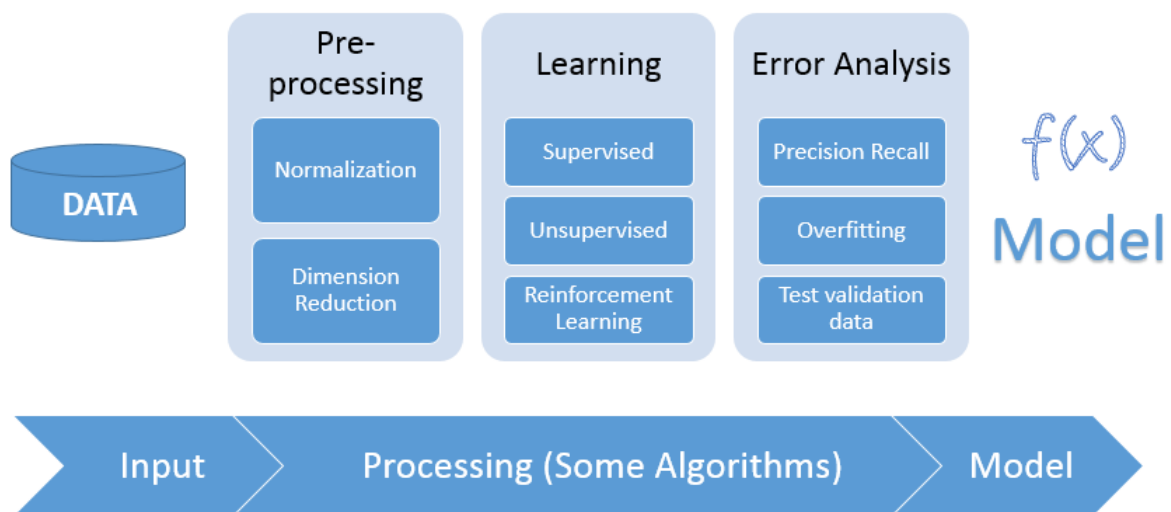


Fuente: tomado de Azimov Institute. The Neural network zoo. Disponible en <http://www.asimovinstitute.org/neural-network-zoo/>. Febrero, 2019.

- **Machine Learning**

Machine Learning o Aprendizaje Automático es una rama de la Inteligencia Artificial dedicada al aprendizaje de forma recurrente o automatizada. Normalmente es la rama central en la mayor parte de las aplicaciones de Inteligencia Artificial, pero no la única. Se suele además combinar con otras tecnologías. Por ejemplo, para entender el lenguaje escrito, lo que una persona quiere decir, es necesario aplicar conocimientos de morfología, de sintaxis, de sinonimia, de estadística, etc. Eso sí, una vez modelado, se puede utilizar para ayudar a reconocer y clasificar nuevos textos con base en el aprendizaje previo. En la Figura 22 se observan las etapas generales que se abordan en un proyecto de Machine Learning.

Figura 22. Proceso de desarrollo para Machine Learning.



Fuente: tomado de (Bell, 2015).

A continuación, se indican los pasos mínimos para un proyecto con *Machine Learning*:

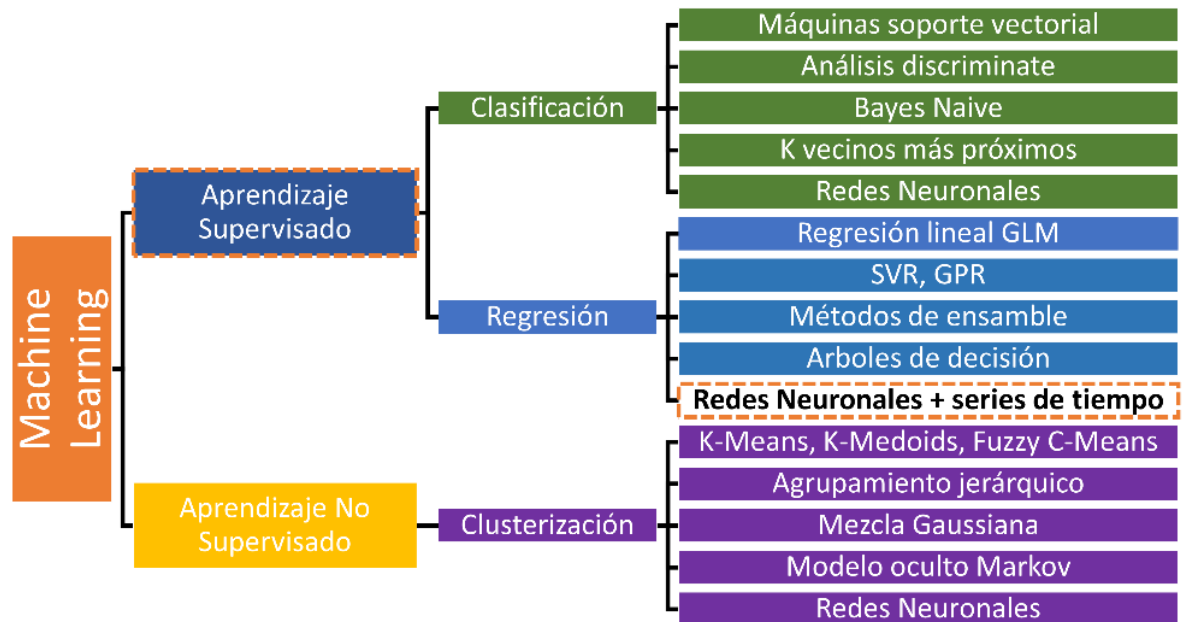
- **Recopilación de datos:** el conjunto de datos que tiene variedad, densidad y volumen de datos relevantes ayudará a un mejor aprendizaje.
- **Preparación de los datos:** esto implica solucionar problemas con el conjunto de datos recopilados, p. manejo de valores atípicos y gestión de puntos de datos

faltantes. Divida el conjunto de datos limpio en dos partes, una para capacitación y otra para evaluar el programa.

- **Entrenando un modelo:** elija un algoritmo apropiado y una representación de datos en forma del modelo adecuado para su problema. Use el conjunto de datos de entrenamiento para entrenar el modelo.
- **Evaluación del modelo:** para probar la precisión y exactitud del modelo, use el conjunto de datos de prueba que se guardan en el paso 2.
- **Mejora del rendimiento:** puede implicar elegir diferentes modelos y algoritmos, o introducir más variables y / o datos para entrenar el modelo.
 - ***Métodos populares de aprendizaje automático:***

Las tareas de aprendizaje automático generalmente se clasifican en tres grandes categorías, dependiendo de la naturaleza de la "señal" de aprendizaje o "retroalimentación" disponible para un sistema de aprendizaje. Los modelos basados en datos se basan en técnicas estadísticas y de aprendizaje. Con el creciente número de sensores en un sistema del mundo real, aumenta la posibilidad de detectar su comportamiento y el estado actual. Por lo tanto, la mayoría de los enfoques de la literatura reciente sobre la detección del tipo de falla y el mantenimiento predictivo tienen en cuenta los modelos basados en datos. Estos modelos son más genéricos que los modelos físicos y de base de conocimiento. Los algoritmos detrás de los modelos basados en datos se determinan a partir de los datos proporcionados. En la Figura 23 se especifican las técnicas disponibles dentro de Machine Learning.

Figura 23. Técnicas dentro de Machine Learning.



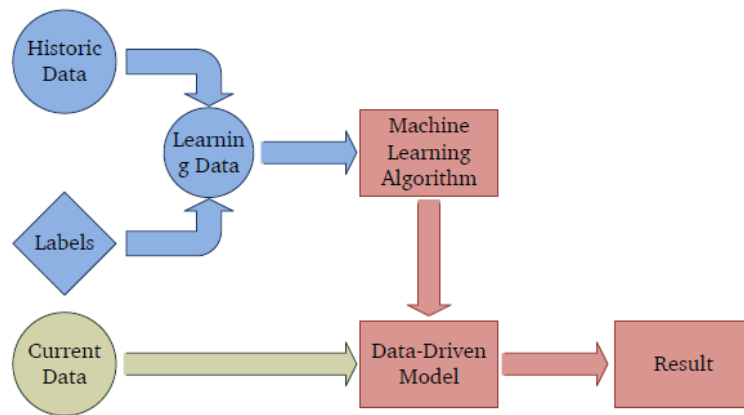
Fuente: Ajustado de *What Is Machine Learning?. Machine Learning, Matlab*. Disponible en: <https://www.mathworks.com/discovery/machine-learning.html>. Febrero, 2019.

- Aprendizaje Supervisado (Modelos Predictivos):

En el aprendizaje supervisado, se utilizan los datos recopilados en el pasado de un sistema del mundo real. Este conjunto de datos se denomina datos históricos. Como no todos los registros de datos recopilados se pueden usar como ejemplos de aprendizaje, los datos deben filtrarse. Cada registro de datos de entrenamiento está etiquetado de acuerdo con su resultado esperado. El conjunto de datos resultante se denomina datos de entrenamiento. Se procesa mediante una técnica de aprendizaje automático. Esta técnica examina la relación entre un registro de datos y el resultado etiquetado, y luego crea un modelo basado en datos. Para cualquier dato nuevo, el modelo basado en datos intenta dar el mejor resultado en función de los datos aprendidos. La Figura 24 visualiza el flujo de trabajo descrito de un modelo basado en datos de aprendizaje supervisado.

En el área de la detección del tipo de falla y el mantenimiento predictivo, la técnica de aprendizaje supervisado es la técnica de aprendizaje más comúnmente utilizada, porque el sistema del mundo real es monitoreado y los datos históricos están disponibles. Además, los datos históricos pueden relacionarse con el estado del sistema del mundo real. Esto mejora la precisión antes de que el sistema se conecte.

Figura 24. Aprendizaje supervisado.



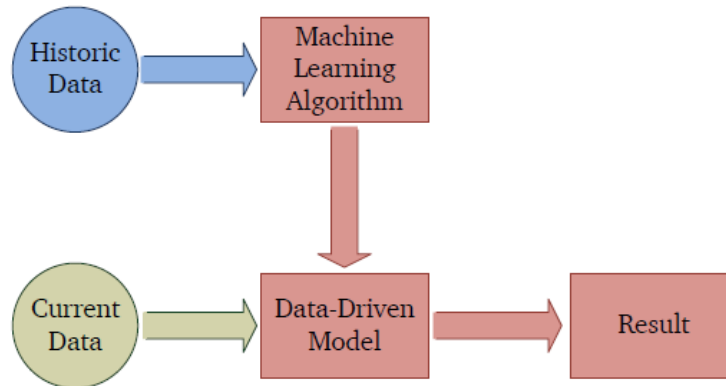
Fuente: tomado de (Patrick Jahnke, 2015).

- **Aprendizaje sin supervisión:**

En otros problemas de modelos basados en datos, los datos históricos no tienen valores objetivo correspondientes. El alcance de estos problemas de aprendizaje sin supervisión es encontrar o descubrir grupos de ejemplos similares dentro de los datos históricos. Este proceso se llama agrupamiento. Otro objetivo del enfoque de aprendizaje no supervisado puede ser determinar la distribución de datos dentro de un espacio de entrada, en un proceso denominado estimación de densidad. La Figura 25 muestra el flujo de trabajo de un modelo de aprendizaje basado en datos de aprendizaje no supervisado. Para la detección del tipo de falla y los problemas de mantenimiento predictivo, la técnica de aprendizaje no supervisado es una técnica de aprendizaje inusual, porque la estimación de clustering y densidad de los datos históricos no es eficiente para la detección precisa del tipo de falla y el mantenimiento predictivo. En la literatura reciente sobre la detección del tipo de falla

y el mantenimiento predictivo, no existe un enfoque que utilice la técnica de aprendizaje no supervisado.

Figura 25. Aprendizaje no supervisado.



Fuente: tomado de (Patrick Jahnke, 2015).

- ***Aprendizaje reforzado:***

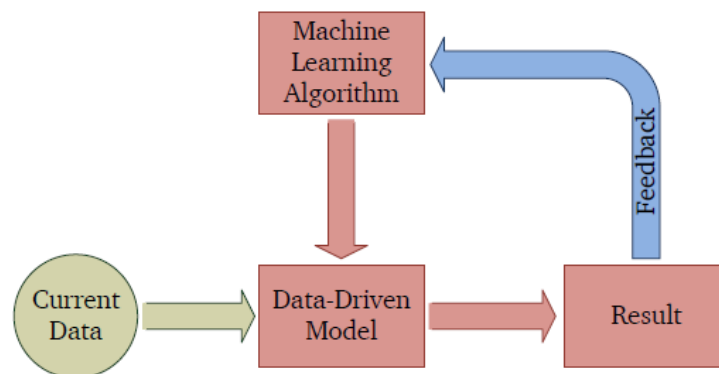
La técnica de aprendizaje de refuerzo ha intentado encontrar un resultado adecuado para los datos de entrada dados. A diferencia de otras técnicas de aprendizaje, no tiene una fase de aprendizaje de datos históricos. Por lo tanto, el algoritmo de aprendizaje automático se basa en un esquema de prueba y error. Al comienzo de la fase de aprendizaje, un algoritmo de aprendizaje automático, en relación con algunos aspectos, define un modelo basado en datos. En consecuencia, también se deben definir los aspectos del algoritmo de aprendizaje automático que definen el modelo basado en datos. Este modelo crea un resultado, dependiendo de los datos actuales en el sistema del mundo real. Ese resultado se evaluará si coincide con el resultado esperado. El algoritmo de aprendizaje automático recibió esta evaluación como retroalimentación sobre la calidad del resultado, a partir de la cual el algoritmo de aprendizaje automático propaga estrategias más exitosas y rechaza las no exitosas. La Figura 26 visualiza la técnica de aprendizaje de refuerzo.

Sin embargo, la técnica de aprendizaje supervisado o no supervisado se puede utilizar en combinación, por lo que el enfoque de aprendizaje automático primero

aprende de los datos históricos (etiquetados o no). El aprendizaje de refuerzo es ampliamente utilizado en modelos basados en datos, donde el resultado es una acción.

En la literatura reciente sobre la detección del tipo de falla y el mantenimiento predictivo, no existe un enfoque en el que se use la técnica de aprendizaje de refuerzo, porque el proceso de aprendizaje ocurre en condiciones de tiempo de ejecución. Esto implica que una estimación incorrecta con respecto al estado de un sistema del mundo real podría tener consecuencias desastrosas. Pero como se indicó anteriormente, la técnica de aprendizaje supervisado en combinación puede refinar un modelo basado en datos en el modo de tiempo de ejecución, lo cual es imposible usando únicamente el aprendizaje supervisado.

Figura 26. Aprendizaje reforzado.



Fuente: tomado de (Patrick Jahnke, 2015).

- **Deep Learning**

Deep Learning, o Aprendizaje Profundo, es una rama del Machine Learning. Es una de las más novedosas, aunque sus orígenes son de la década de los noventa con la aparición de redes neuronales. La novedad ahora es que gracias a nuevos avances en computación se pueden relacionar más campos de observación que antes debían ser tratados de forma independiente.

- ***Big Data***

Big Data o Datos Masivos se refiere a nuevas tecnologías capaces de procesar millones de datos. Es un concepto que ha cobrado fuerza desde el 2010 y coincide con una explosión de datos digitales. El análisis de esos datos puede ser meramente estadístico y no conlleva necesariamente el uso de aplicaciones de inteligencia artificial. Sin embargo, es común que para alguna de sus tareas, por ejemplo la detección de patrones a priori ocultos, sí se utilicen herramientas de machine learning.

4.1.9.1 Aplicación de machine learning en la industria y sectores productivos en general

En la industria y en otros sectores productivos, se encuentran múltiples ejemplos de soluciones desarrolladas utilizando machine learning, en conjunto con otras herramientas de acuerdo con la necesidad del sector.

A nivel comercial, es posible encontrar los siguientes desarrollos a modo de software para pronóstico especializado, bajo la nueva tendencia de Industria 4.0²⁵:

- SAP: Sistema ERP Leonardo (aprendizaje no supervisado, indicadores industriales y globales)
- Siemens: MindSphere (IIoT²⁶, deep learning)
- General Electric: Predix (IIoT, deep learning)
- Fanuc: robot manipulador y sistema de autoaprendizaje no supervisado.
- Kuka: robot de alta precisión fabricado hasta la actualidad (IIoT, aprendizaje no supervisado)

A continuación se muestra un ejemplo de desarrollo académico y de investigación aplicada, el cual tuvo aplicación y solución a un problema real.

²⁵ Industria 4.0: sinónimo Cuarta Revolución Industrial, en donde se conjuga la industria con la ciencia de los datos y automatización.

²⁶ IIoT: acrónimo de Industrial Internet of the Things, en español, Internet Industrial de las Cosas

- ***Predicción de la condición del sistema máquinas de un barco a través de herramientas de confiabilidad analítica y redes neuronales artificiales (Lazakis, Raptodimos, & Varelas, 2018)***

El objetivo de este trabajo es proporcionar un enfoque sistemático utilizando Machine Learning para identificar sistemas/componentes de maquinaria de barco críticos y analizar sus parámetros físicos. Además, para los sistemas / componentes principales del motor crítico y sus parámetros relevantes a monitorear, se identificarán mediante la combinación de Análisis mediante Árboles de Falla (FTA) y Análisis de Modos de Falla y Efectos (FMEA).

En la Figura 27 se muestra la embarcación tipo carguera Panamax, cuyo nombre se deriva del Canal de Panamá, dado que este tipo de barcos están diseñados para pasar por el canal sin problema de distancias entre puertos.

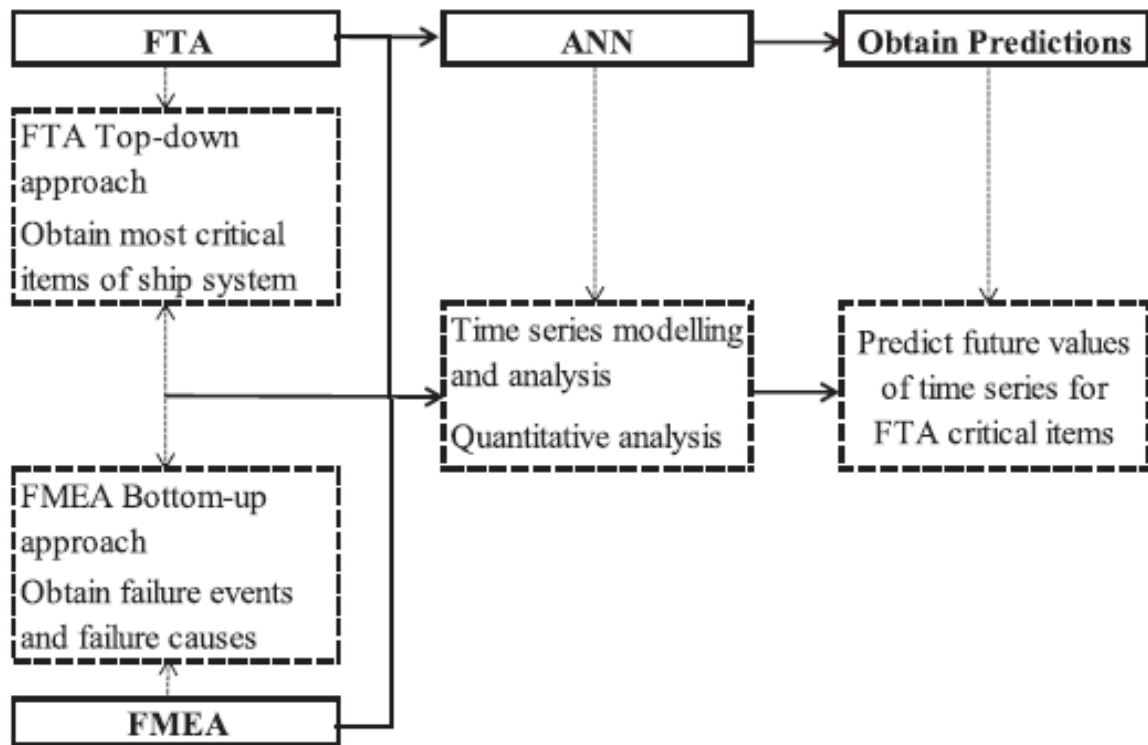
Figura 27. Barcos cargueros tipo Panamax (Cap. Máx. 52 mil toneladas)



Fuente: tomado de (Lazakis et al., 2018)

En la Figura 28 se muestra del diagrama que condensa la metodología que integra machine learning (ANN) con las técnicas y metodologías de mantenimiento FTA y FMEA. Precisamente ambas técnicas de mantenimiento generan las entradas a la red neuronal, obteniendo la estadística de fallas, modos de fallas, tiempo de funcionamiento, entre otros.

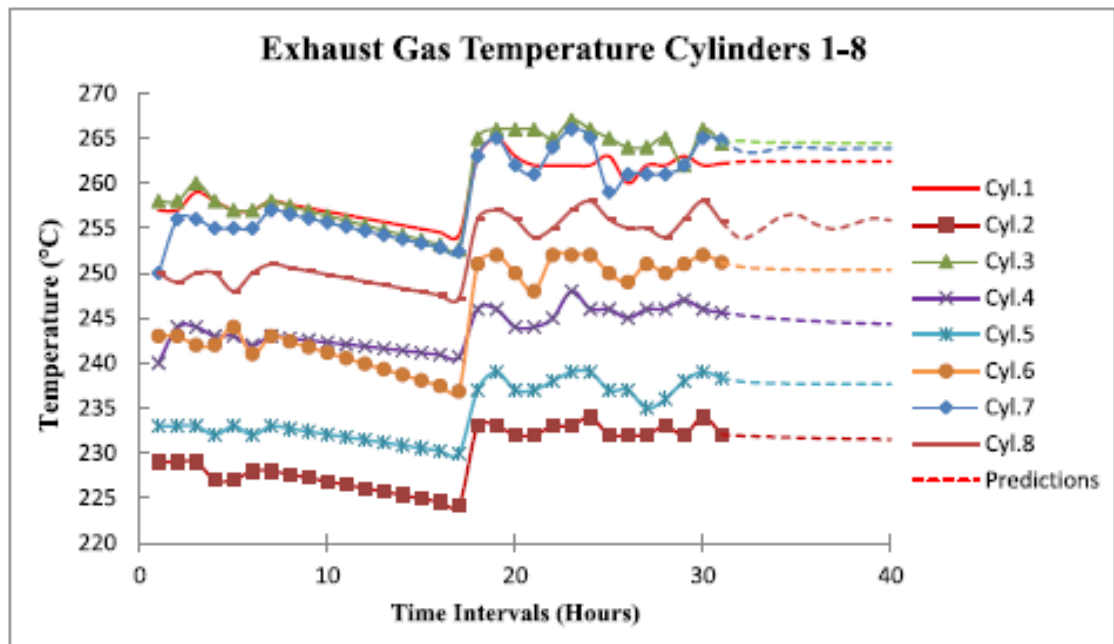
Figura 28. Diagrama metodología integrando ANN, FTA y FMEA.



Fuente: tomado de (Lazakis et al., 2018)

Finalmente en la Figura 29 se muestran los resultados de la red neuronal configurada, en donde se muestran los pronósticos con línea punteada para representar la temperatura de los gases de exhosto. Con el pronóstico de temperaturas se podrán generar por adelantado los escenarios posibles de impactos y consecuencias, generando información que podrá ser retroalimentada a la red neuronal para reentrenar y obtener resultados más cercanos a las condiciones reales.

Figura 29. Pronósticos de temperatura de gases de exhosto.



Fuente: tomado de (Lazakis et al., 2018)

4.1.10 Bibliografía localizada acerca de machine learning y análisis de fenómenos tipo series de tiempo

Dentro de la bibliografía localizada y disponible para el desarrollo de la técnica más adecuada para la construcción del sistema de pronóstico, se encuentran los recursos mencionados en la Tabla 7:

Tabla 7. Recursos bibliográficos disponibles.

Tipo de recurso	Fuente
Libro	Van Gerven, M., & Bohte, S. (2017). Artificial Neural Networks as Models of Neural Information Processing. <i>Frontiers in Computational Neuroscience</i> (Vol. 11). https://doi.org/10.3389/fncom.2017.00114
Artículo	Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., B. Y. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example (pp. 40–51). Springer Tokio. https://doi.org/10.1007/978-4-431-65950-1_3

Artículo	Kamble, L. V., Pangavhane, D. R., & Singh, T. P. (2015). Neural network optimization by comparing the performances of the training functions -Prediction of heat transfer from horizontal tube immersed in gas-solid fluidized bed. <i>International Journal of Heat and Mass Transfer</i> , 83, 337–344. https://doi.org/10.1016/j.ijheatmasstransfer.2014.11.085
Artículo	Abdulkadir, S. J., & Yong, S. P. (2015). Scaled UKF–NARX hybrid model for multi-step-ahead forecasting of chaotic time series data. <i>Soft Computing</i> , 19(12), 3479–3496. https://doi.org/10.1007/s00500-015-1833-z
Libro	Conway, D., & White, J. M. (2012). <i>Machine Learning for Hackers: Case Studies and Algorithms to Get You Started</i> . O' Reilly. https://doi.org/10.1038/nchem.1001
Libro	Attaway, S. (2017). <i>A Practical Introduction to Programming and Problem Solving</i> . En Elsevier (Ed.), <i>Matlab</i> (p. 598). Elsevier. https://doi.org/10.1016/B978-0-12-405876-7.00001-8
Libro	J. Bell, <i>Machine Learning: Hands-On for Developers and Technical Professionals</i> , vol. 1. 10475 Crosspoint Boulevard Indianapolis, IN 46256: John Wiley & Sons, Inc, 2015.
Artículo	Patrick Jahnke, "Machine Learning Approaches for Failure Type Detection and Predictive Maintenance", p. 83, 2015.
Libro	N. J. Nilsson, <i>Inteligencia Artificial una nueva síntesis</i> . 2016.
Artículo	I. H. Witten, E. Frank, y M. a Hall, <i>Data Mining: Practical Machine Learning Tools and Techniques</i> . 2011.
Libro	O'Reilly Media, <i>Big Data Now: 2012 Edition</i> . 1005 Gravenstein Highway North, Sebastopol, CA 95472.: O'Reilly Media, Inc., 2012.
Artículo	Z. Wang <i>et al.</i> , "Failure prediction using machine learning and time series in optical network", <i>Opt. Express</i> , vol. 25, núm. 16, p. 18553, 2017.
Artículo	E. D. Michie, D. J. Spiegelhalter, y C. C. Taylor, "Machine Learning , Neural and Statistical Classification", <i>Technometrics</i> , vol. 37, núm. 4, p. 459, 1994.
Compilación digital de archivos	Wikipedia, "Machine Learning Guide", 2017.
Artículo	C. M. Frau y R. M. Guajardo, "Comparación de redes neuronales y regresión lineal para estimar productividad de sitio en plantaciones forestales, utilizando geomática", <i>Bosque</i> , vol. 27, núm. 1, pp. 35–43, 2006.
Tesis de grado Maestría	D. A. G. Pasuy, "Diseño de una metodología basada en una técnica inteligente para análisis de los tiempos muertos de una línea de producción. Aplicación en una empresa del sector alimenticio de la zona centro de Colombia", Universidad Nacional de Colombia Facultad de Ingeniería y Arquitectura, 2017.
Libro	P. Kim, <i>MATLAB deep learning: with machine learning, neural networks and artificial intelligence</i> .
Libro	G. Peck, <i>Machine Learning Techniques Using Matlab</i> . .

4.1.11 Metodologías de pronóstico aplicadas a series de tiempo.

En este apartado se mostrarán en conjunto las alternativas desde los puntos de vista de la estadística y la ciencia de los datos (machine learning), debido a que los resultados más promisorios han sido obtenidos gracias a la aplicación conjunta de herramientas en ambas áreas.

▪ **Series de tiempo**

Una serie de tiempo es un conjunto de valores, correspondientes a una variable, ordenados según el instante en que fueron capturados, y normalmente tomados a intervalos regulares de tiempo, sin embargo, en la mayoría de casos d. Los datos se pueden comportar de diferentes formas a través del tiempo, puede que se presente una tendencia, un ciclo; no tener una forma aleatoria, variaciones estacionales (anual, semestral, etc).

4.1.11.1 **Métodos de pronóstico**

En la Figura 30 se muestran los tipos de métodos para realización de pronósticos.

Figura 30. Tipos de métodos para pronósticos.



Fuente: adaptado de Ingeniería industrial Online. Disponible en : <https://www.ingenieriaindustrialonline.com/herramientas-para-el-ingeniero-industrial/pron%C3%B3stico-de-ventas/>. Febrero, 2019.

Para el proyecto se eligieron los métodos cuantitativos, causales:

- Regresión múltiple
- Redes neuronales con series de tiempo
- **Métodos por series de tiempo**

Entre los métodos para pronosticar condiciones bajo series de tiempo, se detallan a continuación en la Tabla 1Tabla 8:

Tabla 8. métodos por series de tiempo para pronósticos.

Métodos por series de tiempo	Descripción del método	Usos	Exactitud en un plazo			Identificación del punto de retorno	Costo relativo
			Corto	Mediano	Largo		
Promedio móvil	El pronóstico de basa en un promedio aritmético ponderado de un número de puntos de datos del pasado	Planeación de corto a mediano plazo para inventarios, niveles de producción y programación. Es bueno cuando hay muchos productos	Mala a muy buena	Mala	Muy mala	Mala	Bajo
Suavización exponencial	Similar al promedio móvil y da un mayor peso exponencial a los datos más recientes. Bien adaptado para usarse con computadoras y cuando es necesario pronosticar un gran número de artículos	Igual que el promedio Móvil	Regular a muy buena	Mala a buena	Muy mala	Mala	Medio
Métodos matemáticos	Un modelo lineal o no lineal ajustados con los datos de series de tiempo, normalmente mediante regresión. Incluye las líneas de tendencia, polinomios, logaritmos lineales, series de Fourier, etc.	Igual que el promedio móvil pero con limitaciones debido al costo y su uso con pocos productos	Muy buena	Regular a buena	Muy mala	Mala	Bajo a medio
Box-Jenkins	Métodos de autocorrelación que se usan para identificar las series de tiempo subyacentes para ajustar el "mejor" modelo. Se necesitan aproximadamente 60 puntos de datos del pasado	Limitado debido al costo de los productos que requieren de pronósticos muy exactos a corto plazo	Muy buena a excelente	Regular a buena	Muy mala	Mala	Medio alto

Fuente: tomado de (SCHROEDER, 1996)

4.1.11.2 Análisis de regresión lineal múltiple

Una extensión natural del modelo de regresión lineal simple consiste en considerar más de una variable explicativa. Los modelos de regresión múltiple estudian la relación entre:

- Una variable de interés Y (variable respuesta o dependiente).
- Un conjunto de variables explicativas o regresoras $X_1, X_2, \dots, X_p$

En el modelo de regresión lineal múltiple se supone que la función de regresión que relaciona la variable dependiente con las variables independientes es lineal, es decir:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (4-18)$$

- **Evaluación del desempeño de pronóstico: Medición del error**

Para la evaluación del desempeño predictivo se emplean diferentes indicadores que la aproximación de la variable pronosticada respecto a serie de datos respectiva.

Tabla 9. Indicadores para medición de error estadístico.

Nombre de indicador	Ecuación	Utilidad
Suma Acumulada Errores (CFE)	$CFE = \sum_{t=1}^T e_t$	Indica si los pronósticos están centrados o descentrados respecto a los datos originales.
Desviación Media Absoluta (MAD)	$MAD = \frac{\sum_{t=1}^T e_t }{T}$ $e_t = \text{desviación acumulada}$ $T = \text{número de períodos analizados}$	Se utiliza para calcular cuánto varían de su media los valores de un conjunto de datos.
Promedio del Error Porcentual Absoluto (MAPE)	$MAPE = \frac{1}{T} \left(\sum_{t=1}^T APE_t \right) = \frac{1}{T} \left(\sum_{t=1}^T \frac{ Y_t^s - Y_t^a }{Y_t^a} \right) \times 100$ APE :error porcentual absoluto. Y_t^a :valor pronosticado de Y_t . Y_t^s :valor real de Y_t . T : número de períodos.	Mide el tamaño del error (absoluto) en términos porcentuales, muy útil especialmente para variables con valores a gran escala (millones).

Raíz Cuadrática Media de Error (RMSE)	$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t^s - Y_t^a)^2}$ Y_t^a :valor pronosticado de Y_t. Y_t^s :valor real de Y_t. T: número de periodos.	Es un estimador que mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima
Desviación Estándar del Error	$\sigma_e \sim 1.25MAD$	Es una estimación de la desviación estándar, la cual sirve para identificar si los datos están normalmente distribuidos.

Fuente: ajustado de (SCHROEDER, 1996)

4.1.11.3 **Redes neuronales con series de tiempo, técnicas dentro de Machine Learning.**

Debido a que el ambiente de desarrollo de la aplicación es Matlab, a continuación se mencionan las redes tipo NARX para análisis de datos tipo series de tiempo.

- **Redes neuronales no lineales autorregresivas con entrada externa**

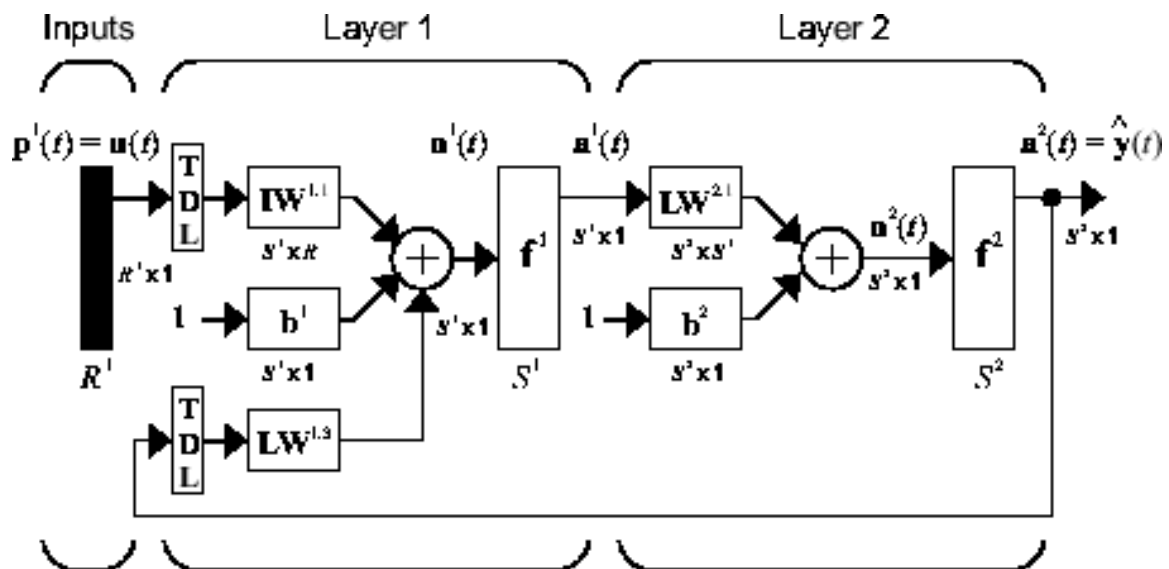
El modelo autorregresivo no lineal con entradas exógenas (NARX, Nonlinear autoregressive neural network with external input) es una red dinámica recurrente, con las conexiones de retroalimentación que incluyen varias capas de la red. El modelo NARX se basa en el modelo lineal de ARX, que comúnmente se utiliza en el modelado de series de tiempo. Estas redes pueden aprender a predecir una serie de tiempo dados valores pasados de la misma serie de tiempo, la entrada de retroalimentación y otra serie de tiempo, llamada serie de tiempo externa o exógena (Paluszek & Thomas, 2017). La ecuación para el modelo NARX es:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-n_y), u(t-1), u(t-2), \dots, u(t-n_u)) \quad (4-19)$$

El siguiente valor de la señal de salida dependiente $y(t)$ se regresa a los valores previos de la señal de salida y los valores previos de una señal de entrada independiente (exógena). Se puede implementar el modelo NARX utilizando una

red neuronal de retroalimentación hacia adelante, para aproximar la función $f()$. A continuación, se muestra un diagrama de la red resultante, donde se utiliza una red de avance de dos capas para la aproximación. Esta implementación también permite un modelo ARX vectorial, donde la entrada y la salida pueden ser multidimensionales. En la Figura 31 se muestra la configuración convencional para este tipo de redes.

Figura 31. Configuración clásica de red neuronal NARX.



Fuente: tomado de (Paluszek & Thomas, 2017)

Estas redes tienen diversos tipos de aplicaciones. Se puede usar como predictor para determinar el siguiente valor de la señal de entrada. También se puede utilizar para el filtrado no lineal, en el que la salida de destino es una versión sin ruido de la señal de entrada. Las redes NARX se utilizan de forma importante en el modelado de sistemas dinámicos no lineales.

4.2 Marco conceptual

A continuación, se presenta el esquema del marco conceptual para el proyecto, cuyo final se alcanza cuando se presentan los resultados indicados anteriormente. El proceso se podría reiniciar con el fin de continuar hacia el proceso de mejora continua, partiendo de los escenarios pronosticados como información de reentrenamiento del sistema, con el fin de incrementar la precisión del modelo de predicción previamente diseñado en la etapa de machine learning.

Figura 32. Esquema planteado para desarrollo del proyecto.



Fuente: proyecto.

4.3 Marco legal y normativo

En lo referente al marco legal, dentro del proyecto se tuvieron en cuenta las siguientes normas y estándares:

- ISO55001:2014, Asset management -- Management systems, Edition 1. 2014.
- UNE-EN 16646:2015, Maintenance - Maintenance within physical asset management. Versión 2015.
- ISO 14224 y The British Standards Institution, "ISO 14224:2016", Bsi, vol. 1999, p. 71, 2016.

5 METODOLOGÍA

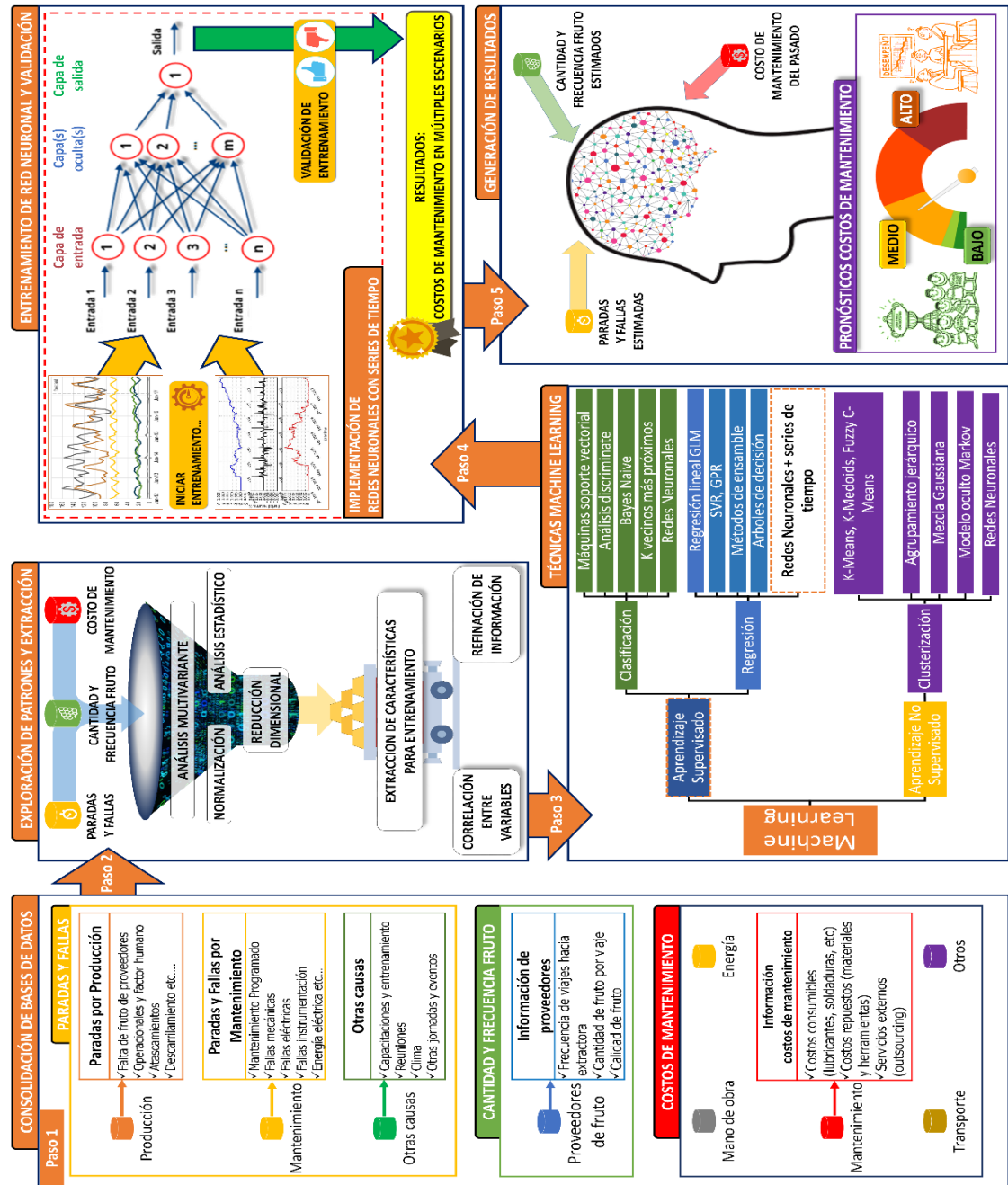
5.1 Plan de trabajo

La metodología del proyecto será la siguiente, la cual se ilustra en la Figura 33 con los detalles por cada paso:

- ***Consolidación de bases de datos (paso 1):***
 - De acuerdo con los archivos suministrados por la planta piloto, se procederá a consolidar los datos inicialmente en formato MS Excel para su posterior migración hacia Matlab.
 - Se organizarán por años, meses y tipo de dato por procedencia (oferta de fruto, mantenimiento, producción, costos, entre otros).
- ***Exploración de patrones y extracción de características (paso 2):***
 - Se aplicarán análisis de sensibilidad y otras técnicas estadísticas, con el fin de determinar las principales variables con mayor aporte estadístico sobre los valores finales de costos de mantenimiento.
 - Aplicación de normalización de datos para su procesamiento por la red neuronal.
- ***Implementación de herramienta Machine Learning (paso 3):***
 - Diseño de red neuronal con series de tiempo (aprendizaje supervisado), utilizando código M en Matlab, se incluirá este desarrollo dentro de una GUI como interfaz de usuario.
- ***Entrenamiento de red neuronal y validación (paso 4):***
 - Se analizará el desempeño de la red neuronal a partir de estadísticos tales como error medio (ME), desviación media absoluta (MAD), error medio cuadrado (MSE), error porcentual (PE) y MPE, entre otros parámetros para verificar desviación de respuestas según los datos respuesta existentes.

- Validación con datos fuera del rango de entrenamiento inicial. Verificar si se requiere reentrenamiento.
- ***Generación de resultados (paso 5):***
- Escenarios de costos de mantenimiento pronosticados: se presentarán de acuerdo con la fluctuación y disponibilidad de fruto, debido a que esta condición es preponderante para el desempeño y disponibilidad de la planta.

Figura 33. Esquema general del proyecto para el desarrollo del sistema de pronósticos.

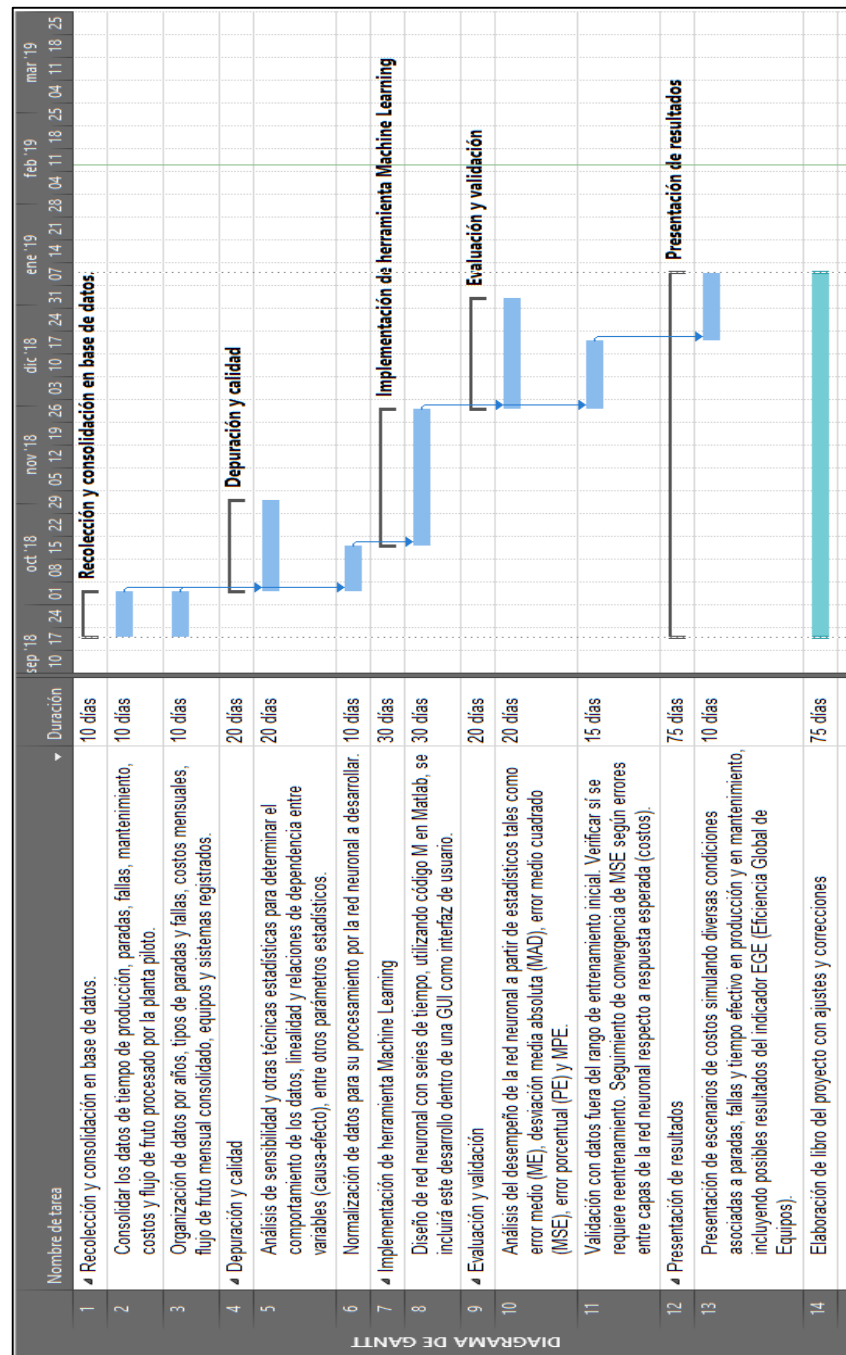


Fuente: tomado de póster expuesto durante XIX Conferencia Internacional sobre Palma de Aceite. Sitio disponible en <http://web.fedepalma.org/xviii-conferencia/es/la-conferencia>

5.2 Cronograma de actividades

En la Figura 34 se indica la asignación de tiempo por cada una de las actividades anteriormente especificadas.

Figura 34. Cronograma de actividades.



6 PRESUPUESTO

En la Tabla 10 se podrá encontrar el detalle del presupuesto estimado para la realización del proyecto, detallando la contribución para cada parte involucrada.

Tabla 10. Presupuesto estimado para el proyecto.

Recurso	Rol	Costo Bruto (COP)	Costo Neto	Dedicación en tiempo		
				%	COP	Costo HH (\$)*
Cesar Augusto Díaz Rangel	Estudiante maestría (investigador Programa de Procesamiento, Cenipalma)	4.500.000	6.840.000	80%	5.472.000	22.800
Jesús Alberto García Núñez	Codirector (Coordinador Programa de Procesamiento, Cenipalma)	12.000.000	18.240.000	5%	912.000	3.800
David Alfredo Fuentes Díaz	Director de proyecto (UIS)	12.000.000	18.240.000	15%	2.736.000	11.400
Costo mensual dedicación personal e intelectual (\$)					9.120.000	
Costo específico dedicación durante 75 días del proyecto (5 meses aprox.) (\$)			Cenipalma	70%	31.920.000	
			UIS	30%	13.680.000	
Costo total dedicación personal e intelectual (\$)					45.600.000	

*Hora Hombre:

Calculada a partir de 30 días X 8 horas

CONCEPTO	PRESUPUESTO (COP)
GASTOS GENERALES	
Papelería y útiles escritorio	850.000
GASTO DE OFICINA	
Utilización PC personal	620.000
Suscripción o compra de libros y estándares	1.400.000
Gastos de viajes y viáticos en general	2.100.000
Impresión y presentación de documentos	250.000
GASTOS DE PERSONAL Y ASESORÍA	45.600.000
Imprevistos	2.000.000
PRESUPUESTO TOTAL (\$ COP)	\$52.820.000

7 RESULTADOS DEL PROYECTO

7.1 Especificación de datos de ingreso y de respuesta para los modelos de pronóstico

En la Tabla 11 se muestra la selección de variables de entrada y variables de respuesta, considerando el objetivo principal del proyecto enfocado a pronosticar costos de mantenimiento, el cual será más preciso al agregar otros parámetros tales como tiempo total programado y no programado tanto para mantenimiento como para producción, eficiencia global de equipos (EGE) dada la importancia de conjugar eficiencia en utilización del tiempo, velocidad de equipos según los valores nominales y la calidad del proceso. Este proceso de selección fue soportado en el análisis de regresión de mejores subconjuntos y en las correlaciones cruzadas entre variables.

Como resumen de la Tabla 11 se configuran 18 variables de entrada (color verde) y 6 variables de respuesta (color amarillo).

Tabla 11. Filtrado de variables con base en análisis de correlación y regresión múltiple.

VARIABLES INICIALES	AGRUPACIÓN	VARIABLES SELECCIONADAS [Unidad medida]
TIEMPO PLANEADO OPERACIÓN PLANTA	INDICADORES GLOBALES DE DESEMPEÑO	TIEMPO PLANEADO OPERACIÓN PLANTA [h]
FRUTO PROCESADO (t RFF)		FRUTO PROCESADO [toneladas de RFF]
FRUTO A PROCESAR TEÓRICO (t RFF)		
TIEMPO EFECTIVO DE PROCESO (h)		TIEMPO EFECTIVO DE PROCESO [h]
TIEMPO PROGRAMADO MANTENIMIENTO		TIEMPO PROGRAMADO MANTENIMIENTO [h]
TIEMPO NO PROGRAMADO MANTENIMIENTO		TIEMPO NO PROGRAMADO MANTENIMIENTO [h]
TIEMPO PROGRAMADO PRODUCCION		TIEMPO PROGRAMADO PRODUCCION [h]
TIEMPO NO PROGRAMADO PRODUCCION		TIEMPO NO PROGRAMADO PRODUCCION [h]
CAPACIDAD INSTALADA DE PRENSADO (t RFF/h)		CAPACIDAD INSTALADA DE PRENSADO [h]
CAPACIDAD EFECTIVA DE PRENSADO (t RFF/h)		CAPACIDAD EFECTIVA DE PRENSADO [h]
% TEA		% TEA
% PÉRDIDAS AC/RFF		

% DISPONIBILIDAD DE PLANTA		
% RENDIMIENTO DE EQUIPOS		
% PÉRDIDAS DE PROCESO		
% OEE		% OEE
TRM (USD)	INDICADORES ECONÓMICOS GENERALES	
CPO CIF ROTTERDAM (USD/TON CPO)		CPO CIF ROTTERDAM (USD/TON CPO) [USD]
CPKO CIF ROTTERDAM (USD/TON CPKO)		
INFLACION ANUAL		
Inicio y fin proceso	TOTAL TIEMPO PROGRAMADO PRODUCCIÓN (h)	TIEMPO TOTAL PARADAS PROGRAMADAS PRODUCCION
Aseo		
Capacitación		
Equipo no utilizado		
Falta de fruto fresco	TOTAL TIEMPO NO PROGRAMADO PRODUCCIÓN, OPERACIONES Y OTROS (h)	TIEMPO TOTAL PARADAS PRODUCCIÓN, OPERACIONES Y OTROS
Falta de fruto esterilizado		
Arranque y parada de planta		
Atascamiento de equipo por cuerpos extraños		
Disparo por sobrecarga de equipos		
Operación inadecuada		
Vagoneta descarrilada		
Atascamiento en ciclón y columna de fibras		
Alto nivel de Lodos		
Ausencia de personal		
Mantenimiento programado	TOTAL TIEMPO PROGRAMADO MTTO Y SERVICIOS (h)	TIEMPO TOTAL PARADAS MANTENIMIENTO PROGRAMADO
Mesa de transferencia	RECEPCIÓN	RECEPCIÓN [h]
Tolvas fruto, compuertas, unidad hidráulica		
Esterilizador N° 1	ESTERILIZACIÓN	ESTERILIZACIÓN [h]
Esterilizador N° 2		
Esterilizador N° 3		
Puente grua	DESFRUTACIÓN	DESFRUTACIÓN [h]
Tambor desfrutador		
Sinfín bajo desfrutador		
Sinfín inclinado N° 1		
Transportador tusa N° 1 y No 2		
Elevador de fruto	EXTRACCIÓN	EXTRACCIÓN [h]
Sinfín distribuidor fruto		
Digestor N° 1		
Digestor N° 2		

Digestor N° 3		
Prensa N° 1		
Prensa N° 2		
Prensa N° 3		
Transportador de torta	DESFRIBRADO	DESFRIBRADO [h]
Ciclon de fibras y ducto		
Esclusa de fibras		
Sinfines fibras N° 1, N 2, N 3, N4		
Tamiz circular crudos	CLARIFICACIÓN	CLARIFICACIÓN [h]
Bomba crudos		
Preclarificador		
Bombas preclarificador		
Clarificador N° 1		
Clarificador N° 2		
Bomba aceite humedo		
Bomba aceite terminado		
Tamiz circular de lodos		
Bomba desarenadora		
Ciclón desarenador		
Tanque pulmon de lodos		
Filtrocepillo N° 1		
Filtrocepillo N° 2		
Deslodadora N° 1		
Deslodadora N° 2		
Deslodadora N° 3		
Bomba y tanque aceite recuperado N° 1		
Bomba y tanque aceite recuperado N° 2		
Bomba y tanque aceite recuperado N° 3		
Elevador y Sinfín nuez humeda	PALMISTERÍA	PALMISTERÍA [h]
Tambor pulidor		
Caldera N° 1	GENERACIÓN DE VAPOR	GENERACIÓN DE VAPOR [h]
Caldera N° 2		
Caldera N° 3		
Caldera N° 4		
Suministro de agua para calderas	SUMINISTRO DE AGUA	SUMINISTRO DE AGUA [h]
Suministro de agua para proceso		
Transformador y transferencia electrica	TRANSMISIÓN Y GENERACIÓN ALTERNA EE	TRANSMISIÓN Y GENERACIÓN ALTERNA EE [h]
Plantas Emergencia N° 1 y N a 2		
Daños eléctricos en motores	OTRAS FALLAS DE MANTENIMIENTO	OTRAS FALLAS DE MANTENIMIENTO [h]

CONSUMIBLES A	TOTAL COSTOS CONSUMIBLES	TOTAL COSTOS DE MANTENIMIENTO [millones COP]
MATERIALES DE ASEO PARA MTTO		
CONSUMIBLES B		
PINTURAS		
MATERIALES PARA CONSTRUCCION		
SOLDADURAS		
TUBERIAS Y ACCESORIOS		
CORREAS		
ACEITES Y LUBRICANTES		
LAMINAS		
TORNILLERIA		
MATERIALES VARIOS		
ELEMENTOS DE LABORATORIO		
MAQUINARIA Y EQUIPOS PLANTA	TOTAL COSTOS REPUESTOS	
CHUMACERAS Y EMPAQUES		
BOMBAS Y ACCESORIOS		
RODAMIENTOS		
REPUESTOS MECANICOS		
ELEMENTOS DE INSTRUMENTACION		
REDUCTORES Y ACCESORIOS		
REPUESTOS GENERALES		
REPUESTOS ELECTRICOS		
MOTORES Y ACCESORIOS PARA MOTORES		
SERVICIOS MECANICOS	TOTAL COSTOS SERVICIOS	
TALLER		
SERVICIOS GENERALES		
SERVICIOS ELECTRICOS		

7.1.1 Análisis de normalidad para variables de respuesta del modelo

Antes de aplicar la técnica de regresión múltiple, se revisa la normalidad de las variables de respuesta, con base en la prueba de Anderson-Darling la cual mide qué tan bien siguen los datos una distribución específica. Para un conjunto de datos y distribución en particular, mientras mejor se ajuste la distribución a los datos, menor será este estadístico (Minitab-Inc, 2016). A continuación en la Tabla 12 se muestran los resultados de dicho análisis.

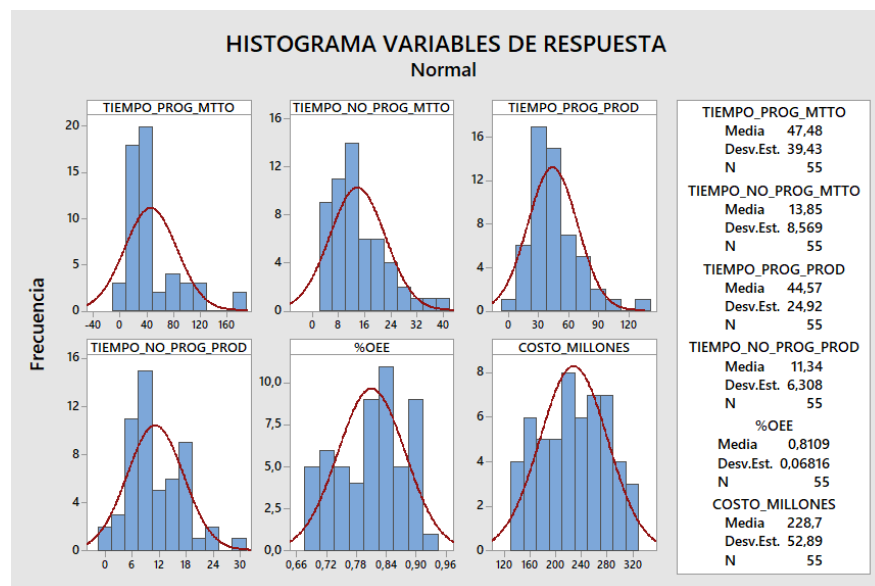
Tabla 12. Resultados prueba de normalidad Anderson-Darling para variables de respuesta.

Grupo	N	Media	IC 95%	Desv.Est.	IC 95%	Mín.	Mediana	Máx.	Prueba de normalidad P	Decisión
PROG MTTO	55	47,481	(36,822. 58,140)	39,429	(33,194. 48,571)	6	36	189,03	<0,005	No pasa
NO PROG MT	55	13,854	(11,537. 16,171)	8,5689	(7,2139. 10,556)	3,06	11,11	39,41	<0,005	No pasa
PROG PROD	55	44,571	(37,835. 51,308)	24,920	(20,979. 30,697)	4,1	40,43	129,58	0,008	No pasa
NO PROG PR	55	11,335	(9,6298. 13,040)	6,3080	(5,3105. 7,7704)	0,25	10	31,03	0,106	Pasa
%EE	55	0,81091	(0,7925. 0,8293)	0,068158	(0,0574. 0,0840)	0,68	0,82	0,92	0,047	No pasa
COSTOS MTT	55	228,69	(214,39. 242,99)	52,892	(44,528. 65,155)	133,57	229,38	328,84	0,302	Pasa

Fuente: desarrollado en Minitab.

De acuerdo con lo anterior, es preciso considerar que la medida de tendencia central válida para las variables de Tiempo no programado de producción y Costos de Mantenimiento, será la media. Sin embargo, para las demás variables de respuesta, se deberá usar la mediana, dado que los conjuntos de datos no tienen comportamiento normal al menos por esta prueba (el valor p es menor al nivel de significancia de 0.005). Esta consideración se establece con el fin de evitar errores al momento de requerir un valor estadístico representativo por variable. En la Figura 35 se muestran las distribuciones para las variables de respuesta, confirmando lo indicado anteriormente.

Figura 35. Histogramas para variables de respuesta.



Fuente: desarrollado en Minitab.

7.1.2 Análisis regresión de mejores subconjuntos

Uno de los criterios para sustentar la elección de las variables fue el resultado de la matriz de análisis de regresión para mejores subconjuntos. Este método identifica los modelos de subconjuntos que generan los valores de R^2 más altos a partir de un conjunto completo de las variables predictoras que se especifiquen. La regresión de los mejores subconjuntos es una manera eficiente de identificar modelos con los que pueda alcanzar sus metas utilizando el menor número posible de predictores. Los modelos de subconjuntos pueden estimar los coeficientes de regresión y pronosticar las respuestas futuras con varianzas más pequeñas que el modelo completo utilizando todos los predictores (D.C. & R.E., 2008).

Respecto a los indicadores estadísticos, se consideró para el análisis (Minitab-Inc, 2016):

- **R-cuad.:** mientras mayor sea el valor de R^2 , mejor se ajustará el modelo a los datos. El R^2 siempre se incrementa cuando se agregan predictores adicionales a un modelo.
- **R-cuad.(ajust):** se usa el R^2 ajustado cuando se requiere comparar modelos que tengan diferentes números de predictores. El R^2 siempre aumenta cuando se agrega un predictor al modelo, incluso cuando no haya una mejora real en el modelo. El valor de R^2 ajustado incorpora el número de predictores del modelo para ayudar a elegir el modelo correcto.
- **R-cuad.(pred):** se utiliza el R^2 pronosticado para determinar qué tan bien el modelo predice la respuesta para nuevas observaciones. Los modelos que tienen valores más grandes de R^2 pronosticado tienen mejor capacidad de predicción. Un R^2 pronosticado que sea sustancialmente menor que el R^2 puede indicar que el modelo está sobreajustado. Un modelo sobreajustado se produce cuando se agregan términos para efectos que no son importantes en la población.
- **Cp de Mallows:** el Cp de Mallows compara la precisión y el sesgo del modelo completo con los de los modelos que tienen los mejores subconjuntos de

predictores. Un valor del Cp de Mallows que está cerca del número de predictores más la constante indica que el modelo produce estimaciones relativamente precisas y no sesgadas.

- **S varianza:** se utiliza S para evaluar qué tan bien el modelo describe la respuesta. Utilice S en lugar de los estadísticos R^2 para comparar el ajuste de los modelos que no tienen una constante. S se mide en las unidades de la variable de respuesta y representa la distancia que separa a los valores de los datos de los valores ajustados. Mientras más bajo sea el valor de S, mejor será descrita la respuesta por el modelo.

En última fila de la Tabla 13 se muestran los valores máximos para R^2 con 65.4%, R^2 ajustado con 53.1%, R^2 pronóstico con 44.1%. Al elegir el valor de R^2 más alto se indica que deberían elegir 18 predictores (variables de entrada) para alcanzar dicho valor como límite mínimo, a pesar de que el valor de varianza (S) no sea el más bajo. Respectivamente, el valor Cp Mallows esta cerca de la cantidad de predictores sugerida, dado que el valor resultante es 19 (generalmente es Cp + constante).

Tabla 13. Regresión de los mejores subconjuntos predictores.

Vars	R-cuad.	R-cuad. (ajustado)	R-cuad. (predicción)	Cp de Mallows	S (varianza)	RECEPCION	ESTERILIZACION	DESFRUTACION	EXTRACCION	DESFIBRADO	CLARIFICACION	PALMISTERIA	GVA	PTA	GEE	OTRAS_MITO	TIEMPO_EFECTIVO	TIEMPO_PLANEADO	CAP_PRENSADO_NOMINAL	CAP_PRENSADO_EFECTIVA	RFF_PROCESADO	TEA	CPO_CIF_ROTTERDAM
1	18,60	17,00	12,40	33,70	4,82E+07	X																	
1	6,30	4,50	0,00	46,60	5,17E+07	X																	
2	25,70	22,90	17,30	28,30	4,64E+07	X	X																
2	25,20	22,30	14,50	28,90	4,66E+07	X	X																
3	38,70	35,10	30,40	16,80	4,26E+07	X	X	X															
3	33,70	29,80	22,20	22,00	4,43E+07	X	X	X															
4	45,80	41,50	35,50	11,40	4,04E+07	X	X	X	X														
4	42,90	38,40	32,10	14,40	4,15E+07	X	X	X	X														
5	48,20	42,90	35,80	10,90	4,00E+07	X	X	X	X	X													
5	48,20	42,90	36,50	11,00	4,00E+07	X	X	X	X	X													
6	53,30	47,40	42,50	7,60	3,84E+07	X	X	X	X	X	X												
6	51,90	45,90	35,60	9,10	3,89E+07	X	X	X	X	X	X												
7	56,90	50,50	42,10	5,80	3,72E+07	X	X	X	X	X	X	X											
7	54,80	48,10	41,90	8,00	3,81E+07	X	X	X	X	X	X	X											
8	58,90	51,80	44,10	5,70	3,67E+07	X	X	X	X	X	X	X	X										
8	58,90	51,70	41,70	5,80	3,67E+07	X	X	X	X	X	X	X	X										
9	60,80	53,00	41,70	5,80	3,63E+07	X	X	X	X	X	X	X	X	X									
9	60,30	52,40	42,40	6,30	3,65E+07	X	X	X	X	X	X	X	X	X									
10	61,80	53,10	41,00	6,80	3,62E+07	X	X	X	X	X	X	X	X	X	X								
10	61,60	52,90	41,70	7,00	3,63E+07	X	X	X	X	X	X	X	X	X	X								
11	62,50	52,90	37,20	8,10	3,63E+07	X	X	X	X	X	X	X	X	X	X	X							
11	62,40	52,80	39,40	8,10	3,63E+07	X	X	X	X	X	X	X	X	X	X	X							
12	63,20	52,70	36,20	9,30	3,64E+07	X	X	X	X	X	X	X	X	X	X	X	X						
12	63,00	52,40	37,00	9,50	3,65E+07	X	X	X	X	X	X	X	X	X	X	X	X						
13	63,70	52,20	34,80	10,80	3,66E+07	X	X	X	X	X	X	X	X	X	X	X	X	X					
13	63,70	52,20	34,80	10,80	3,66E+07	X	X	X	X	X	X	X	X	X	X	X	X	X					
14	64,30	51,80	33,90	12,20	3,67E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
14	64,10	51,50	33,80	12,40	3,68E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
15	64,70	51,20	32,70	13,70	3,70E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
15	64,60	50,90	33,20	13,90	3,70E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
16	65,00	50,20	32,10	15,50	3,73E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
16	65,00	50,20	27,90	15,50	3,73E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
17	65,40	49,50	29,20	17,00	3,76E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
17	65,10	49,00	27,70	17,40	3,78E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
18	65,40	48,10	24,10	19,00	3,81E+07	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	65,4	53,1	44,1		3,62E+07	35	33	31	29	27	25	23	21	19	17	15	13	11	9	7	5	3	1
	MÁXIMO	MÁXIMO	MÁXIMO		MÍNIMO	CONTEO DE PREDICTORES																	

Fuente: análisis mediante Minitab

7.1.3 Análisis de correlación entre variables del modelo

El análisis mediante coeficiente de correlación o indicador de Pearson permite identificar el comportamiento entre variables sin importar el tamaño o unidad de medida, dado que el indicador de Pearson se normaliza entre 0 y 1 (D.C. & R.E., 2008). Este valor de correlación (r) se interpreta de varias formas de acuerdo con el signo:

- Si $r = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva.
- Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.
- Si $r = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante.

De acuerdo con la Tabla 14, la franja color amarillo agrupa a las 6 variables de respuesta, identificando que las relaciones con las demás variables de entrada, incluso entre el mismo grupo de variables de salida, contiene relaciones variadas de incremento y decremento. De forma general se aprecia que la relación entre los costos totales de mantenimiento y las variables de tiempo de paradas por cada área es negativa, lo cual es lógico dado que, al reducir las fallas y tiempo perdido en estas áreas, el costo de mantenimiento también se reduce dado que requerirá menos insumos, repuestos y mano de obra respecto al escenario en donde las fallas llegan al punto más alto de ocurrencia. Con este análisis se comprueba la selección de variables para su posterior integración al modelo de pronóstico.

Tabla 14. Análisis de correlación total de variables.

	RECEPCION	ESTERILIZACION	DESFUTACION	EXTRACCION	DESFIBRADO	CLARIFICACION	PALMISTERIA	GVA	PTA	GEE	OTRAS_MITTO	TIEMPO_EFECTIVO	TIEMPO_PLANEADO	CAP_PRESADO_NOMINAL	CAP_PRESADO_EFECTIVA	RFF_PROCESADO	TEA	CPO_CIF_ROTTERDAM	%OEE	TIEMPO_PROG_MITTO	TIEMPO_NO_PROG_MITTO	TIEMPO_PROG_PROD	TIEMPO_NO_PROG_PROD	COSTO_MITTO	VALOR MINIMO	VALOR MAXIMO	
RECEPCION	1,000																										
ESTERILIZACION	0.213	1.000																							0.213	0.213	
DESFUTACION	0.238	0.261	1.000																						0.238	0.261	
EXTRACCION	0.377	0.098	0.435	1.000																					0.098	0.435	
DESFIBRADO	0.005	-0.038	0.079	0.000	1.000																				-0.038	0.079	
CLARIFICACION	0.175	0.069	0.141	-0.005	0.023	1.000																			-0.005	0.175	
PALMISTERIA	-0.090	0.296	0.271	0.284	0.436	0.172	1.000																		-0.090	0.436	
GVA	-0.022	-0.066	-0.034	-0.014	0.034	0.256	-0.098	1.000																	-0.098	0.256	
PTA	-0.079	-0.039	0.080	-0.052	-0.011	0.011	0.314	-0.080	1.000																-0.080	0.314	
GEE	0.058	-0.065	-0.036	-0.076	0.002	0.030	-0.053	-0.028	0.142	1.000															-0.076	0.142	
OTRAS_MITTO	-0.118	0.079	0.270	-0.139	0.112	0.016	0.255	-0.025	0.595	0.299	1.000														-0.139	0.595	
TIEMPO_EFECTIVO	0.077	0.128	0.288	0.309	0.368	-0.052	0.488	0.170	-0.022	-0.017	-0.130	1.000													-0.130	0.488	
TIEMPO_PLANEADO	0.095	0.159	0.240	0.283	0.176	-0.044	0.339	0.196	-0.126	-0.019	-0.270	0.863	1.000												-0.270	0.863	
CAP_PRESADO_NOMINAL	-0.112	0.090	0.142	-0.115	-0.002	0.050	0.028	0.201	0.142	0.133	0.386	-0.284	-0.260	1.000											-0.284	0.386	
CAP_PRESADO_EFECTIVA	-0.058	0.087	0.349	0.033	0.191	0.192	0.134	0.253	0.019	-0.031	0.230	-0.114	-0.217	0.702	1.000										-0.217	0.702	
RFF_PROCESADO	0.054	0.184	0.456	0.310	0.435	0.068	0.520	0.286	-0.025	-0.042	-0.016	0.862	0.689	0.091	0.400	1.000									-0.042	0.862	
TEA	0.004	-0.015	-0.132	-0.031	-0.349	0.175	-0.258	0.031	-0.211	-0.109	-0.368	-0.206	-0.170	0.047	0.133	-0.124	1.000								-0.368	0.175	
CPO_CIF_ROTTERDAM	-0.188	0.037	0.034	-0.159	0.373	0.098	0.146	-0.079	-0.230	-0.230	-0.100	-0.059	-0.019	-0.145	0.182	0.053	-0.118	1.000							-0.230	0.373	
%OEE	0.119	-0.017	0.207	0.162	0.210	0.090	0.088	0.095	-0.162	-0.202	-0.297	0.352	0.222	-0.455	0.262	0.459	0.147	0.335	1.000						-0.455	0.459	
TIEMPO_PROG_MITTO	0.044	-0.003	-0.203	-0.087	-0.392	-0.081	-0.336	-0.076	-0.165	0.044	-0.242	-0.276	0.192	-0.201	-0.505	-0.498	-0.065	-0.008	-0.309	1.000					-0.505	0.192	
TIEMPO_NO_PROG_MITTO	0.261	0.283	0.733	0.542	0.519	0.375	0.641	0.144	0.108	-0.017	0.228	0.554	0.401	0.067	0.337	0.694	-0.209	0.127	0.282	-0.417	1.000				-0.417	0.733	
TIEMPO_PROG_PROD	-0.146	-0.101	-0.225	-0.239	-0.348	0.070	-0.355	0.107	-0.157	-0.081	-0.191	-0.541	-0.168	0.413	0.264	-0.366	0.323	0.136	-0.185	0.424	-0.409	1.000			-0.541	0.424	
TIEMPO_NO_PROG_PROD	0.077	0.271	0.011	-0.128	-0.265	-0.079	-0.319	-0.036	0.034	0.066	0.126	-0.485	-0.468	0.340	0.349	-0.277	0.267	0.009	-0.171	-0.101	-0.244	0.250	1.000		-0.485	0.349	
COSTO_MITTO	-0.204	-0.250	-0.100	-0.036	0.033	-0.082	0.111	-0.018	-0.114	-0.013	-0.035	-0.008	0.108	0.247	0.016	-0.008	-0.431	-0.026	-0.327	0.195	-0.106	0.208	-0.183	1.000		-0.431	0.247

Fuente: análisis mediante Minitab

7.2 Resultados de modelos estadísticos mediante regresión múltiple

A través de la herramienta Minitab, se determinaron los coeficientes para las ecuaciones de regresión múltiple de las 6 variables de respuesta, según el Anexo A, siendo las ecuaciones:

MODELO TIEMPO PROGRAMADO MANTENIMIENTO	
$\text{TIEMPO_PROG_MITTO} = 145,2 - 0,6398 * \text{TIEMPO_EFECTIVO} + 0,6556 * \text{TIEMPO_PLANEADO} - 4,194 * \text{CAP_PRESADO_EFECTIVA} - 375 * \text{TEA}$	7-1
MODELO TIEMPO NO PROGRAMADO MANTENIMIENTO	
$\text{TIEMPO_NO_PROG_MITTO} = -2,56 + 1,264 * \text{ESTERILIZACION} + 0,9460 * \text{DESFUTACION} + 0,994 * \text{EXTRACCION} + 0,9037 * \text{DESFIBRADO} + 1,106 * \text{CLARIFICACION} + 0,596 * \text{PALMISTERIA} + 0,357 * \text{GVA} + 0,0965 * \text{OTRAS_MITTO} + 0,01272 * \text{TIEMPO_EFECTIVO}$	7-2
MODELO TIEMPO PROGRAMADO PRODUCCION	
$\text{TIEMPO_PROG_PROD} = -144,5 - 4,60 * \text{ESTERILIZACION} - 1,390 * \text{DESFUTACION} - 0,3866 * \text{TIEMPO_EFECTIVO} + 0,3758 * \text{TIEMPO_PLANEADO} + 3,389 * \text{CAP_PRESADO_EFECTIVA} + 342 * \text{TEA}$	7-3
MODELO TIEMPO NO PROGRAMADO PRODUCCION	
$\text{TIEMPO_NO_PROG_PROD} = -67,0 + 5,21 * \text{ESTERILIZACION} - 1,018 * \text{PALMISTERIA} + 0,1407 * \text{TIEMPO_EFECTIVO} + 3,36 * \text{CAP_PRESADO_EFECTIVA} - 0,00620 * \text{RFF_PROCESADO}$	7-4

MODELO OEE	
%OEE=0,6876-0,00425*PALMISTERIA-0,02011*CAP_PRENSADO_NOMINAL+0,02080*CAP_PRENSADO_EFECTIVA+0,000006*RFF_PROCESADO+0,438*TEA	7-5
MODELO COSTO MTTO	
COSTO_MTTO=763312878-39772253*ESTERILIZACION-4468131*DESFIBRADO+10394940*PALMISTERIA-31121542*PTA-2108764*OTRAS_MTTO+8130122*CAP_PRENSADO_NOMINAL-3485353*CAP_PRENSADO_EFECTIVA-3105277934*TEA0,8421-0,512*ESTERILIZACION-0,395*DESFIBRADO+0,501*PALMISTERIA-0,333*PTA-0,427*OTRAS_MTTO+0,3336*CAP_PRENSADO_NOMINAL-0,241*CAP_PRENSADO_EFECTIVA-0,748*TEA	7-6

En la Tabla 15 se muestran los indicadores estadísticos para los datos pronosticados a través de las regresiones múltiples. Cabe recordar que estos modelos se utilizarán como referente para comparar los resultados generados posteriormente por la red neuronal luego de su entrenamiento, en conjunto con la referencia de los datos originales. Las celdas resaltadas en color rojo para el Tiempo no programado producción y para los Costos de Mantenimiento, indican valores de R2 bajos, lo cual indica que el modelo de regresión para estas variables requerirá de otras variables distintas para dar mayor soporte a los pronósticos por regresión múltiple. El valor MAPE alto señalado confirma el requerimiento de agregar reformular la cantidad de variables de entrada para el modelo asociado a dicha variable.

Tabla 15. Resultados estadísticos regresión múltiple para variables de respuesta.

INDICADOR ESTADÍSTICO	VARIABLES DE RESPUESTA					
	EGE	TIEMPO MTTO PROGRAMADO	TIEMPO MTTO NO PROGRAMADO	TIEMPO PROGRAMADO PRODUCCION	TIEMPO NO PROGRAMADO PRODUCCION	COSTOS MTTO
S (VARIANZA)	0,091	0,057	0,044	0,085	0,149	0,188
R-cuad.	0,908	0,935	0,972	0,838	0,520	0,588
R-cuad. (ajustado)	0,899	0,929	0,965	0,818	0,471	0,516
R-cuad. (pred)	0,885	0,915	0,942	0,778	0,384	0,438
MSE	4,20E-04	1,02E+02	2,05E+00	9,44E+01	1,88E+01	1,13E+03
RMSE	2,05E-02	1,01E+01	1,43E+00	9,72E+00	4,34E+00	3,36E+01
MAD	0,016	7,371	1,101	6,690	3,416	26,113
CFE	-2,69E-04	1,39E-01	-1,66E-02	9,62E-02	-7,83E-02	4,58E-03
MAPE	1,98%	25,60%	10,52%	21,51%	83,67%	12,64%
DESV. EST ERROR	0,020	9,214	1,377	8,362	4,270	32,642

A partir de la Figura 36 hasta la Figura 41, se muestran las respuestas de los modelos de regresión múltiple.

Estas respuestas ofrecen un pronóstico aceptable, dado que los modelos en cada caso continúan la tendencia de los datos originales, a pesar de no llegar a los valores exactos sobretodo en máximos y mínimos de los datos base.

Figura 36. Respuesta modelo regresión múltiple Costos de Mantenimiento. En dichas gráficas, la variable base o de origen tiene línea continua color azul.

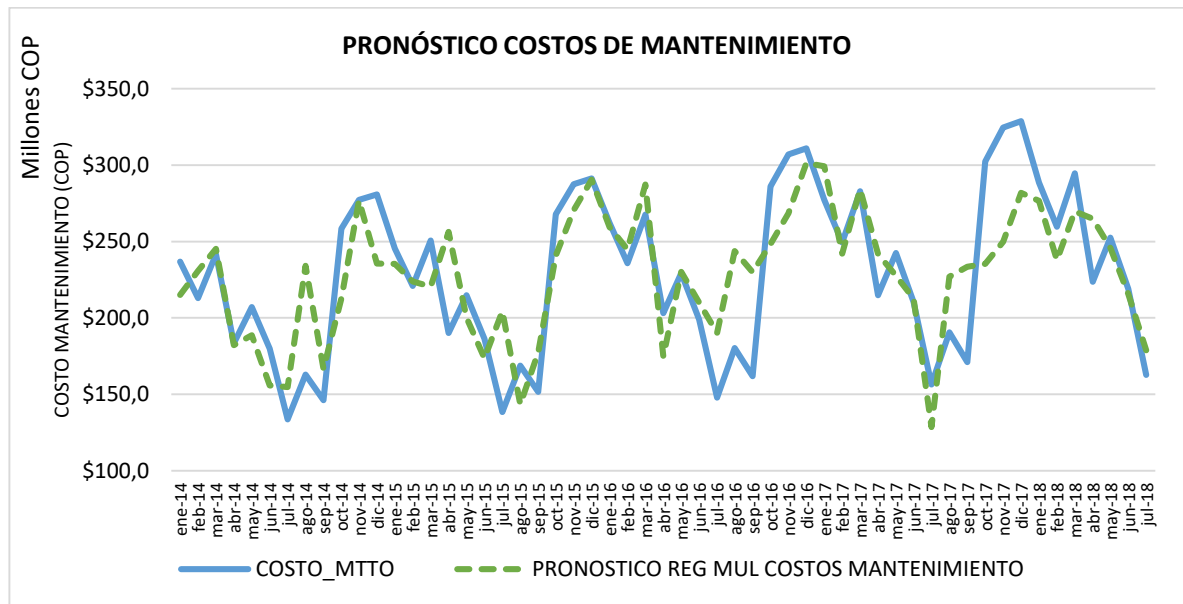


Figura 37. Respuesta modelo regresión múltiple Eficiencia Global de Equipos (EGE)

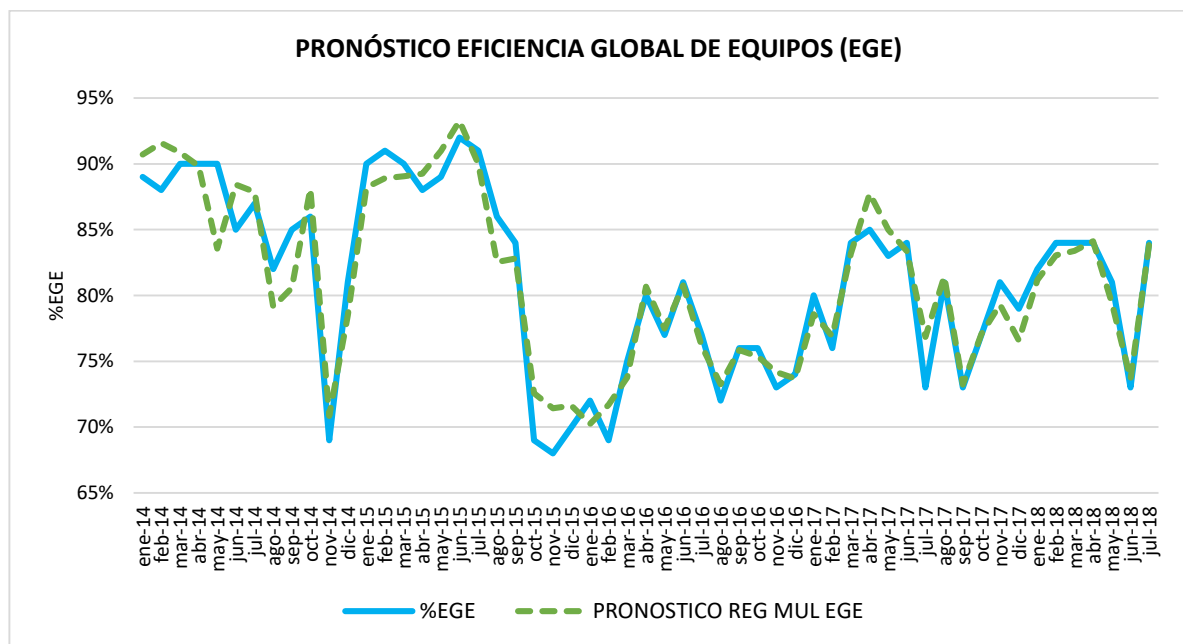


Figura 38. Respuesta modelo regresión múltiple Tiempo paradas programadas mantenimiento.

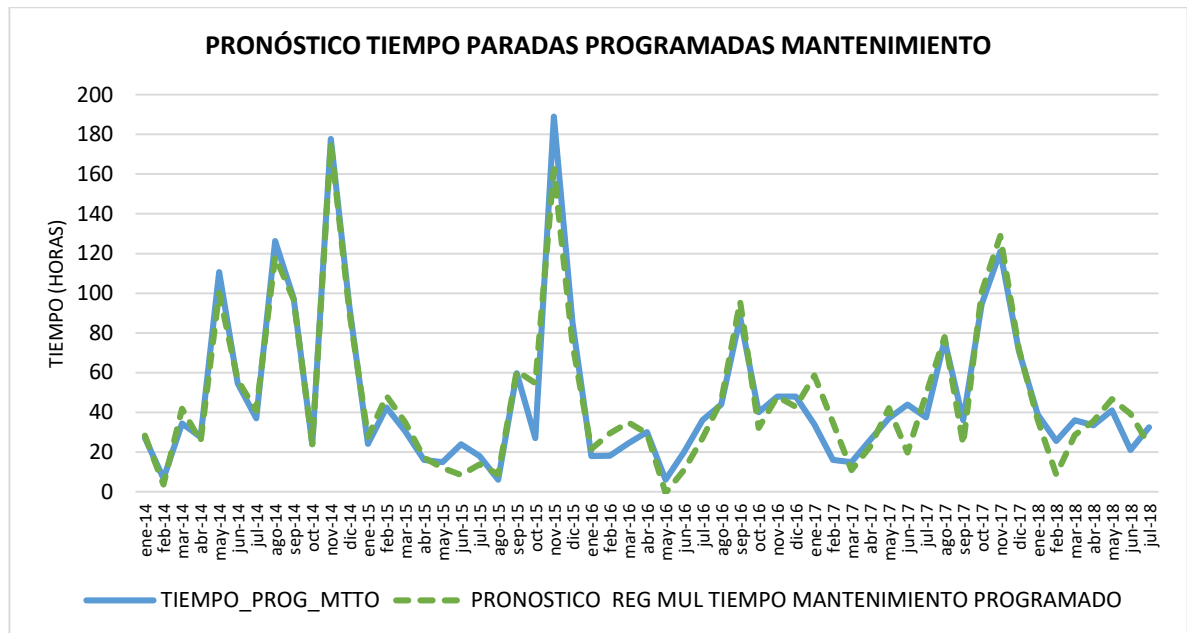


Figura 39. Respuesta modelo regresión múltiple Tiempo paradas no programadas mantenimiento.

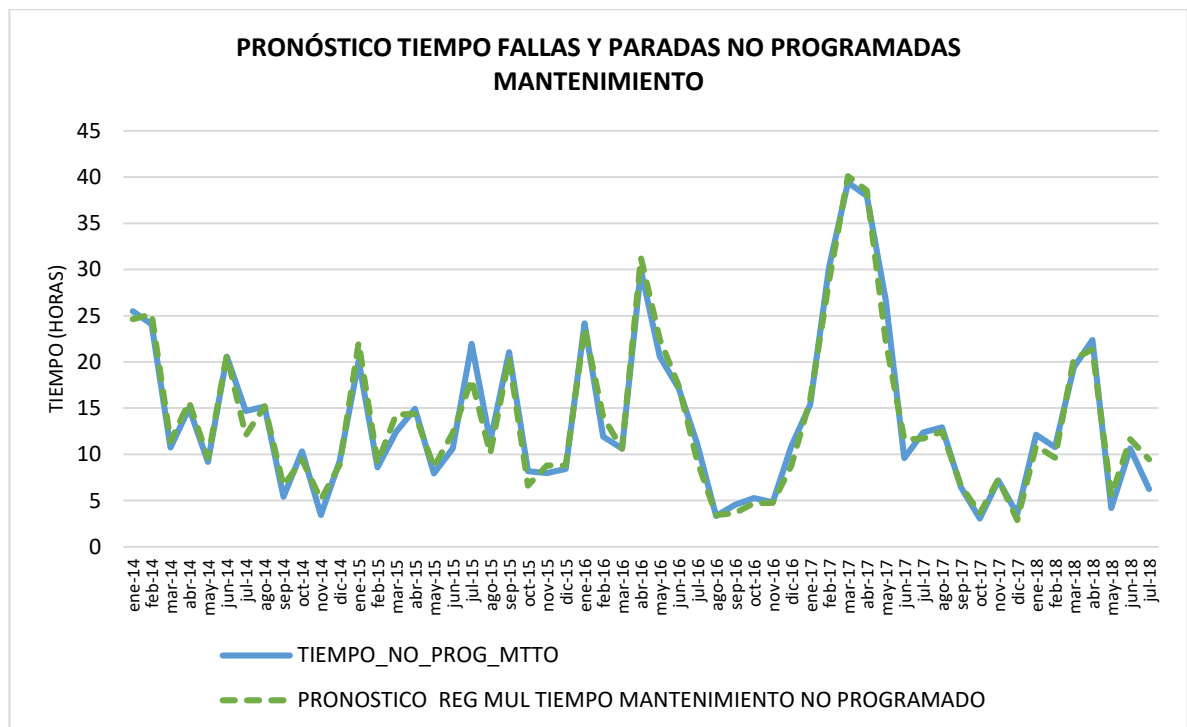


Figura 40. Respuesta modelo regresión múltiple Tiempo paradas programadas producción.

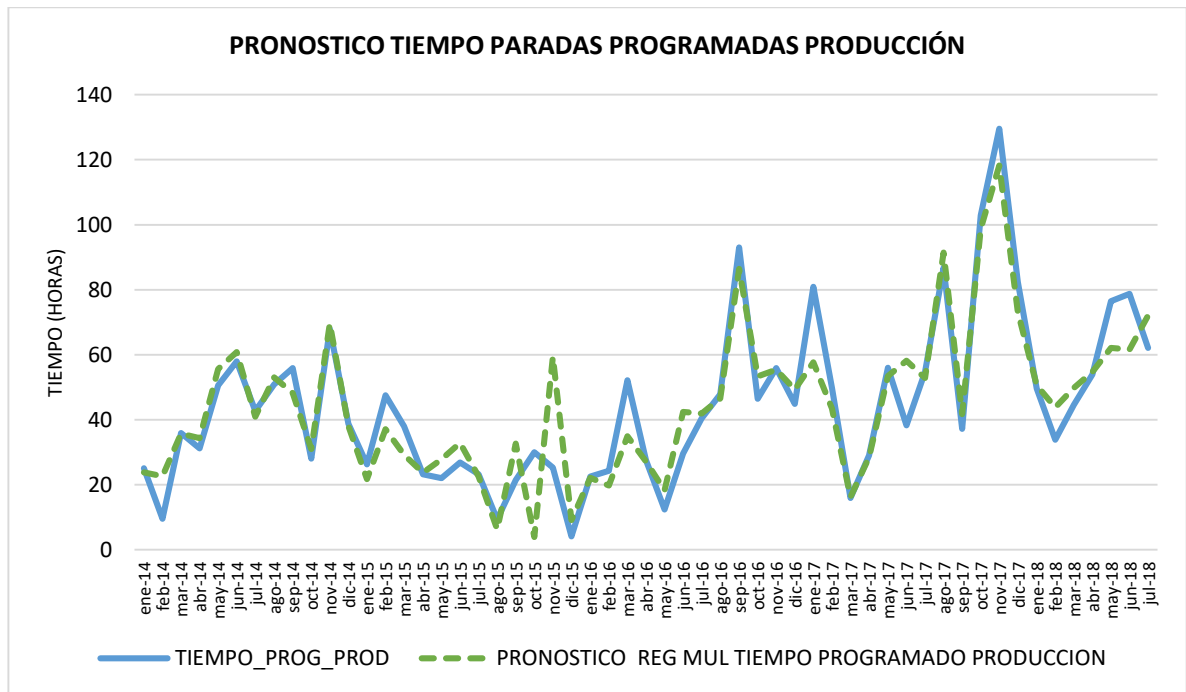
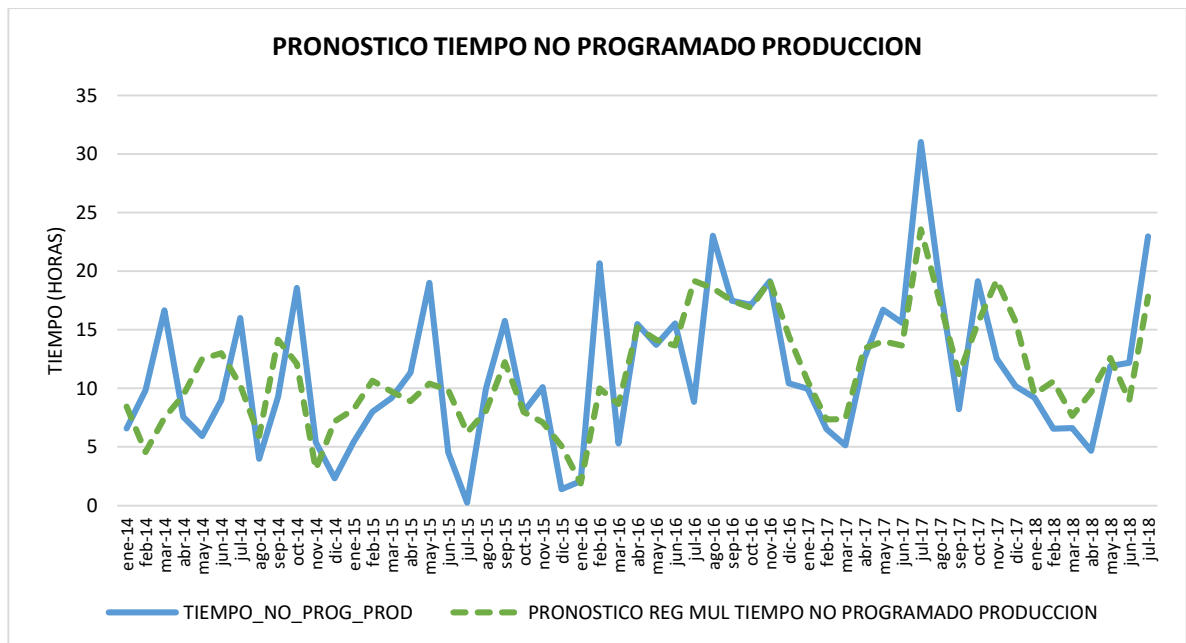


Figura 41. Respuesta modelo regresión múltiple Tiempo no programado producción.



7.2.1 Proceso de normalización de variables previo entrenamiento de red neuronal

Una etapa crucial para el pronóstico de datos a través de redes neuronales es la normalización de variables. Este proceso optimiza el procesamiento de datos por parte de los algoritmos de machine learning, dado que las operaciones matemáticas tendrán respuestas entre rangos preestablecidos e iguales para todas las variables, independiente de su unidad de medida. Problemas de re. Sin embargo, una mala aplicación de la normalización, o una elección descuidada del método de normalización puede modificar erróneamente los datos, y por ende el análisis. El método escogido en este proyecto fue el de Escalado de variables (Feature Scaling o MinMax Scaler) (Paluszek & Thomas, 2017).

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (7-7)$$

Tanto para las 18 variables de entrada como para las 6 variables de respuesta, se aplicó el mismo tipo de normalización, teniendo en cuenta que el comportamiento gráfico no fuese notablemente distinto entre datos originales y datos normalizados, es decir, no se observaron cambios abruptos al realizar la normalización.

7.3 Programación de aplicación inteligente para pronóstico de costos y tipos de fallas de mantenimiento a partir de información histórica.

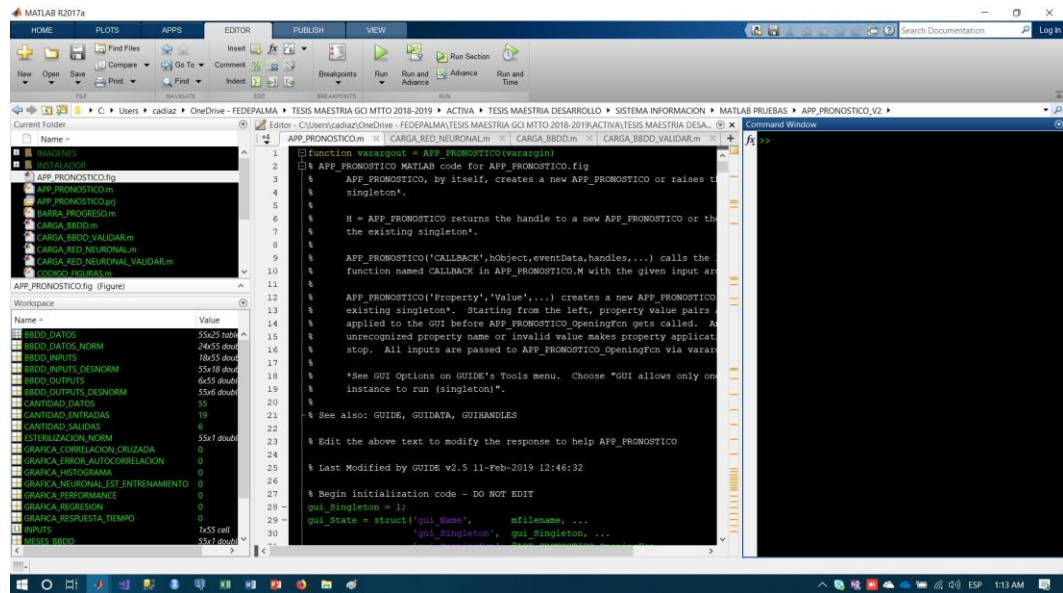
Según se mencionó en la justificación del proyecto, el desarrollo de la aplicación de pronóstico se realizó bajo ambiente del sistema computacional de matemática Matlab²⁷, específicamente utilizando la herramienta de interfaz visual GUIDE, junto con código M y librerías especializadas tales como Redes neuronales para series de tiempo (Neural network time series tool - ntstool). Una vez desarrollada la aplicación, se realizaron pruebas de simulación, depuración y posterior empaquetamiento en forma de aplicación utilizando la herramienta de Compilador de código Matlab para aplicaciones (MCC Builder). Una vez generado el paquete instalador de la aplicación de pronóstico, se integra con el paquete de pre requisitos y compilador sin requerir instalación de Matlab en el equipo anfitrión, a través del paquete Compilador y tiempo de ejecución (MCR Runtime). De esta forma la aplicación desarrollada en este proyecto podrá ser ejecutada en equipos con sistema operativo Windows 10 (preferiblemente), arquitectura 64 bits, procesador Core i5, i7 o i9, memoria RAM al menos de 8 GB, espacio en almacenamiento de al menos 500 MB.

7.3.1 Aplicación desarrollada en Matlab utilizando GUIDE y código M

Utilizando código M para el funcionamiento matemático general de la aplicación y la herramienta GUIDE para el diseño interactivo de la interfaz visual, se generó la siguiente aplicación la cual inicia con la ventana principal para configuración de Matlab (Figura 42). En este ambiente principal es posible controlar las variables que se van generando en tiempo real durante el cálculo, archivos externos utilizados, depuración de código M, ejecución de comandos generales en consola, entre otras funcionalidades que facilitan el desarrollo de aplicaciones.

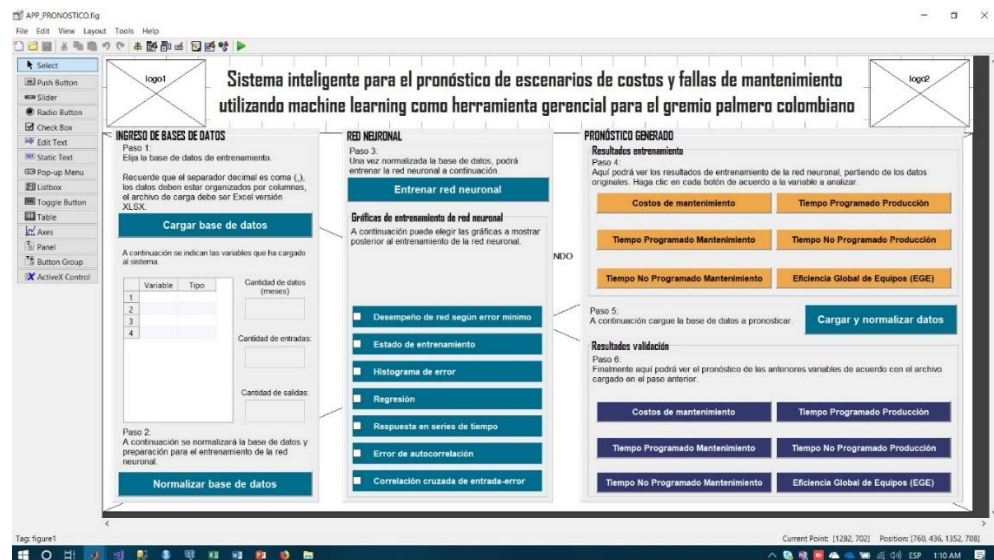
²⁷ Matlab: abreviatura en inglés MATrix LABoratory, en español significa laboratorio de matrices.

Figura 42. Ventana principal sistema Matlab.



GUIDE es otra herramienta de Matlab (Figura 43), especializada en el diseño de interfaz visual e interactiva para usuarios. La forma de programación visualmente es arrastrar y soltar elementos tales como botones, cuadros de texto, barras de desplazamiento, gráficas, entre otros. Una vez puesto dentro de una ventana principal, se genera el respectivo código M, en donde a través del editor de texto se configuran las acciones y operaciones deseadas por cada elemento.

Figura 43. Interfaz de desarrollo GUIDE en Matlab, entorno principal de aplicación.



Una vez se ejecuta la aplicación GUIDE junto con el código M, se genera una nueva ventana en donde se visualiza el resultado de la programación (Figura 44). En esta ventana se organizan por pasos cada una de las actividades a realizar. De acuerdo con el capítulo 5 sobre metodología del proyecto (página 76), el paso 1 consta de la carga de las bases de datos para entrenamiento, posteriormente en el paso 2 se normaliza la información de acuerdo con el apartado 7.3.1. sobre normalización de variables (página 95). En el paso 3 se configura e inicia el entrenamiento de la red neuronal con la opción de verificar el entrenamiento a través de varios indicadores estadísticos. En el paso 4 se verifican los resultados de entrenamiento a través del contraste de las variables base con las pronosticadas por la red. En el paso 5 se cargan nuevas bases de datos para realizar validaciones o nuevos pronósticos, los cuales se pueden visualizar en el paso 6.

Figura 44. Ventana principal de aplicación de pronóstico desarrollada.

Sistema de pronóstico escenarios de costos de mantenimiento para plantas de beneficio - Cenipalma. Proyecto de grado maestría Gerencia de Mantenimiento - Cesar Augusto Díaz Rangel.



Sistema inteligente para el pronóstico de escenarios de costos y fallas de mantenimiento
utilizando machine learning como herramienta gerencial para el gremio palmero colombiano



INGRESO DE BASES DE DATOS

Paso 1:
Elija la base de datos de entrenamiento.

Recuerde que el separador decimal es coma (,), los datos deben estar organizados por columnas, el archivo de carga debe ser Excel versión XL.SX.

Cargar base de datos

A continuación se indican las variables que ha cargado al sistema.

	Variable	Tipo
1		
2		
3		
4		

Cantidad de datos (meses):

Cantidad de entradas:

Cantidad de salidas:

Paso 2:
A continuación se normalizará la base de datos y preparación para el entrenamiento de la red neuronal.

Normalizar base de datos

RED NEURONAL

Paso 3:
Una vez normalizada la base de datos, podrá entrenar la red neuronal a continuación.

Entrenar red neuronal

Gráficas de entrenamiento de red neuronal

A continuación puede elegir las gráficas a mostrar posterior al entrenamiento de la red neuronal.

- ☐ Desempeño de red según error mínimo
- ☐ Estado de entrenamiento
- ☐ Histograma de error
- ☐ Regresión
- ☐ Respuesta en series de tiempo
- ☐ Error de autocorrelación
- ☐ Correlación cruzada de entrada-error

PRONÓSTICO GENERADO

Resultados entrenamiento

Paso 4:
Aquí podrá ver los resultados de entrenamiento de la red neuronal, partiendo de los datos originales. Haga clic en cada botón de acuerdo a la variable a analizar.

Costos de mantenimiento

Tiempo Programado Producción

Tiempo Programado Mantenimiento

Tiempo No Programado Producción

Tiempo No Programado Mantenimiento

Eficiencia Global de Equipos (EGE)

Paso 5:
A continuación cargue la base de datos a pronosticar.

Cargar y normalizar datos

Resultados validación

Paso 6:
Finalmente aquí podrá ver el pronóstico de las anteriores variables de acuerdo con el archivo cargado en el paso anterior.

Costos de mantenimiento

Tiempo Programado Producción

Tiempo Programado Mantenimiento

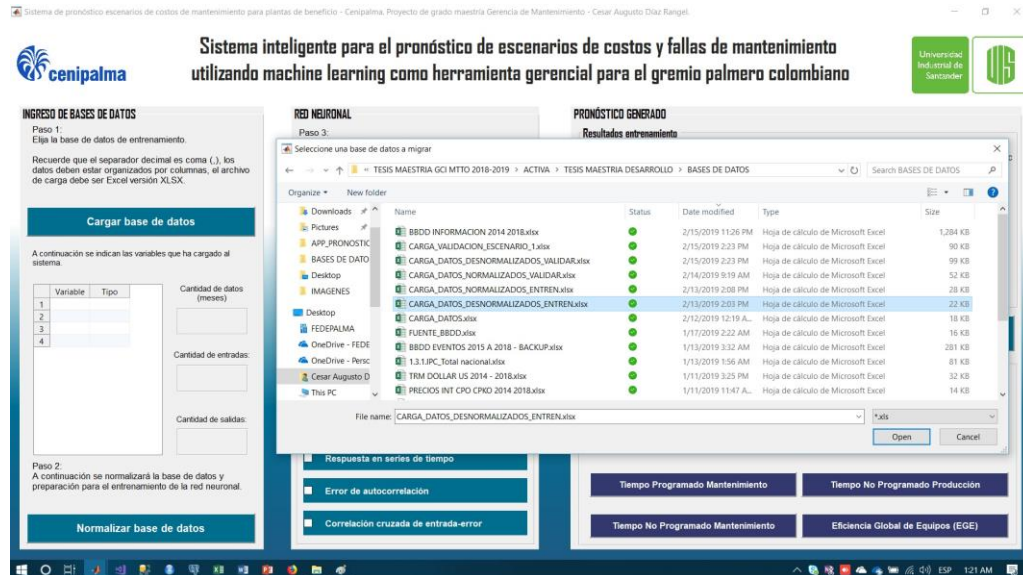
Tiempo No Programado Producción

Tiempo No Programado Mantenimiento

Eficiencia Global de Equipos (EGE)

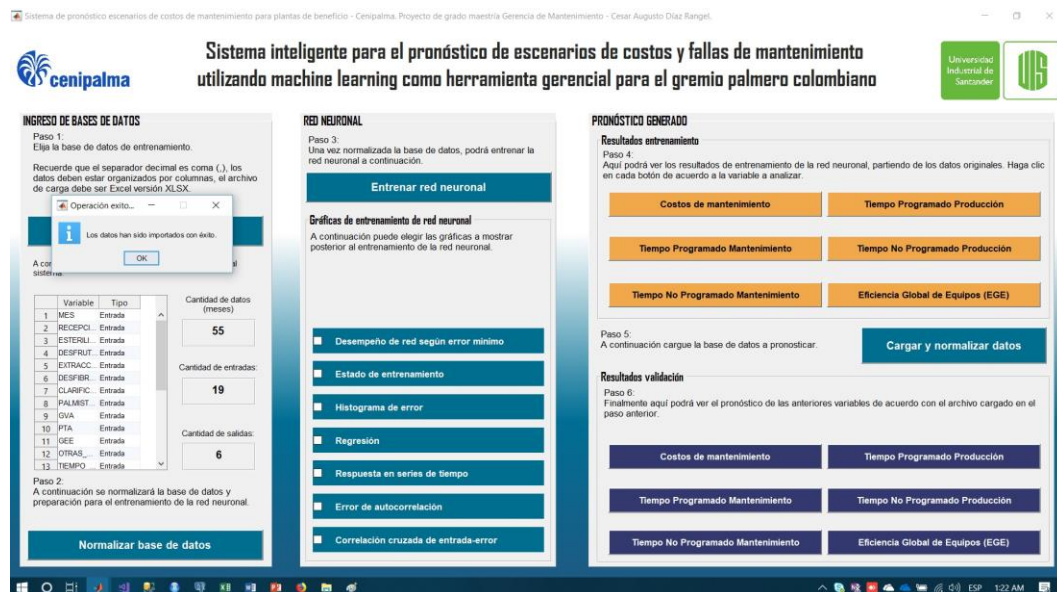
La Figura 45 corresponde al paso 1, en donde se muestra una ventana emergente para indicar el archivo a cargar. Se debe elegir un archivo en formato Excel (XLSX), en donde las variables se organizan por columnas y los datos en filas de forma cronológica (ver Anexo B).

Figura 45. Carga de archivo base de datos para entrenamiento de red neuronal.



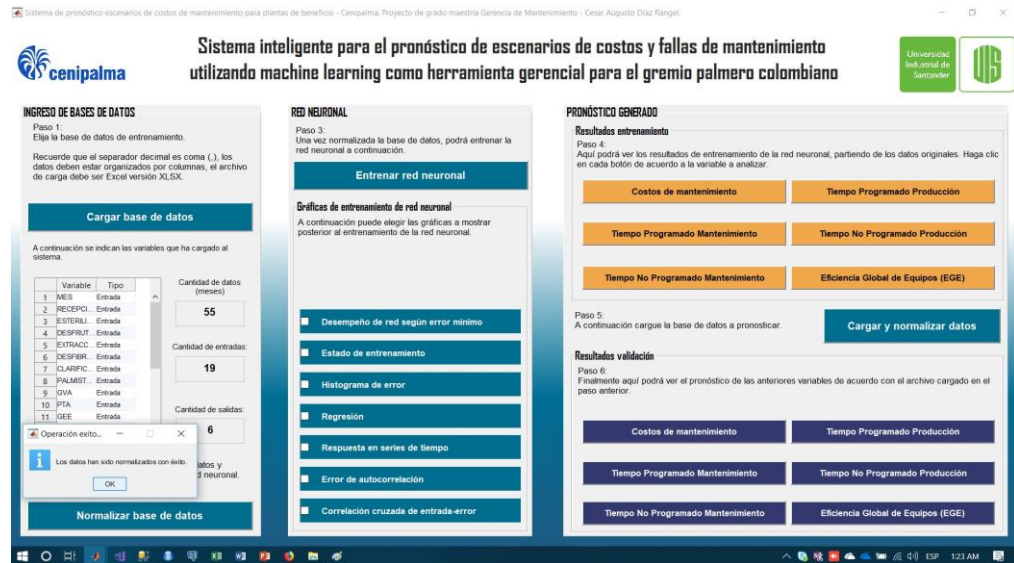
Proceso de carga de base de datos completado (Figura 46).

Figura 46. Proceso de carga de base de datos completado.



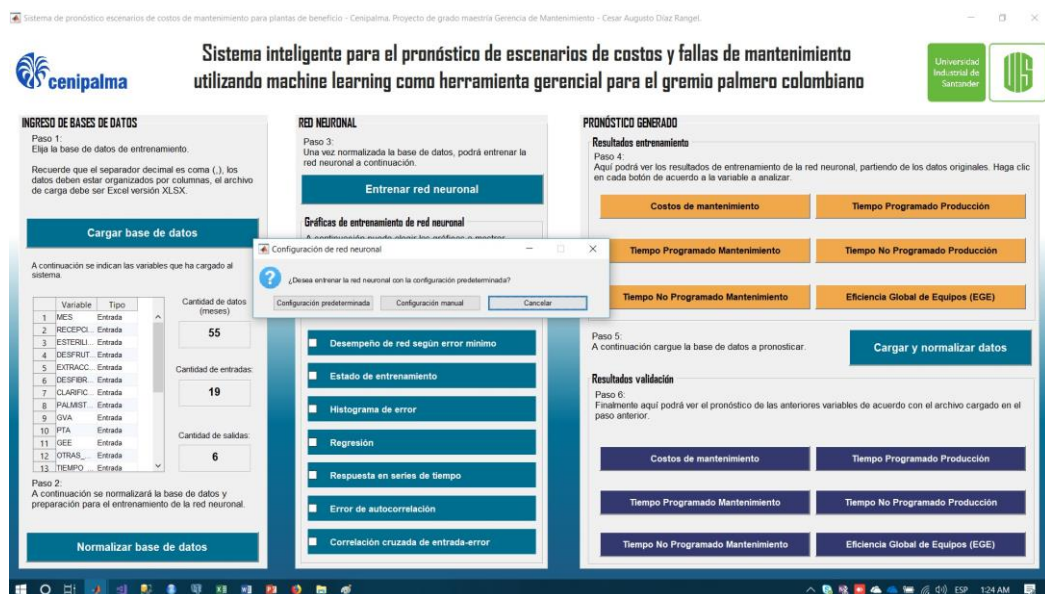
En la Figura 47 se muestra el mensaje de normalización de base de datos completado, paso obligatorio antes del entrenamiento de la red neuronal.

Figura 47. Proceso de normalización de base de datos completado.



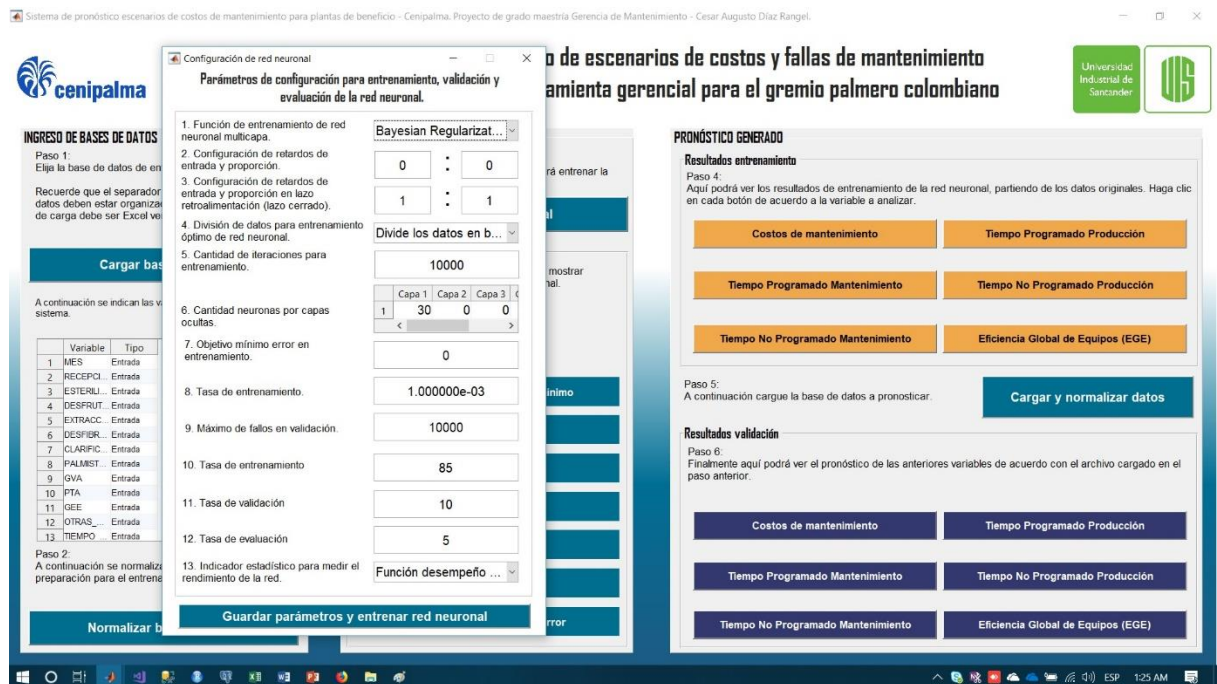
En la Figura 48 se muestra el menú de entrenamiento de la red neuronal, en donde es posible elegir entre el entrenamiento con parámetros predeterminados, o la configuración de parámetros de forma manual.

Figura 48. Menú de entrenamiento de red neuronal para elegir configuración predeterminada o específico por parámetro.



En la Figura 49 se muestra la ventana emergente para la opción de configuración manual.

Figura 49. Ventana de configuración manual de red neuronal a entrenar.



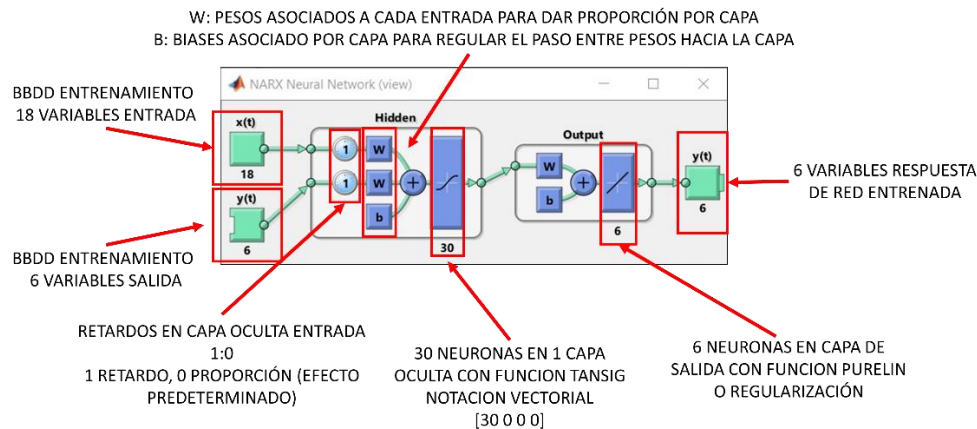
Entre las opciones configurables se encuentra:

1. Función de entrenamiento:
 - a. Bayesian Regularization ('trainbr'): (predeterminado) se recomienda este tipo de técnica de entrenamiento de acuerdo con referencias sobre redes neuronales para series de tiempo, incluso aplicables para series del tipo caótico o con estadístico media diferente en cada período de análisis. Los estadísticos R2 y RMSE muestran mejores comportamientos respecto al uso de otras funciones, incluso el tiempo de convergencia al error mínimo es más bajo (Kamble, Pangavhane, & Singh, 2015).
 - b. Levenberg-Marquardt ('trainlm')
 - c. BFGS Quasi-Newton ('trainbfg')
 - d. Resilient Backpropagation ('trainrp')
 - e. Scaled Conjugate Gradient ('trainscg')

- f. Conjugate Gradient with Powell/Beale Restarts ('traincgb')
 - g. Fletcher-Powell Conjugate Gradient ('traincgf')
 - h. Polak-Ribière Conjugate Gradient ('traincgp')
 - i. One Step Secant ('trainoss')
 - j. Variable Learning Rate Gradient Descent ('traingdx')
 - k. Gradient Descent with Momentum ('traingdm')
 - l. Gradient Descent ('traingd')
2. Configuración de retardos:
 - a. Entrada: cantidad de retardos y proporción de los mismos.
 - b. Lazo cerrado retroalimentado: cantidad de retardos y proporción de los mismos.
 3. División de datos para optimización de entrenamiento:
 - a. División proporcional en bloques contiguos (divideblock): (predeterminado) se usa a modo predeterminado dado que se ajusta a series de tiempo en donde el espaciamiento es conocido. En este caso, los datos están organizados por meses.
 - b. División aleatoria (dividerand)
 - c. División por intercalamiento de datos (divideint)
 - d. División de datos según índice (divideind)
 4. Cantidad de iteraciones durante entrenamiento: predeterminado 10.000 veces o ciclos, según la complejidad en convergencia al error indicado.
 5. Cantidad de neuronas por capas: es posible configurar hasta 5 capas ocultas, en donde cada celda indica la cantidad de neuronas por capa. De forma predeterminada se ha establecido solo 1 capa oculta con 30 neuronas. No se recomienda incrementar la cantidad de capas y neuronas sin proporción, dado que podría ocasionar entrenamientos más lentos, baja convergencia, valores de RMSE altos, distribución no normal de error, entre otros problemas. En la Figura 50 se indica la configuración de la red neuronal por defecto para el proyecto, sugerida dado que se alcanzó el entrenamiento más balanceado respecto a tiempo de

entrenamiento, error y desempeño final. De acuerdo con el apartado 4.1.9 (página 53), el tipo de arquitectura es una red perceptrón multicapa, con retroalimentación de error de adelante hacia atrás (MLP Neural Network with Backpropagation) (Paluszek & Thomas, 2017).

Figura 50. Configuración de red neuronal para el proyecto.



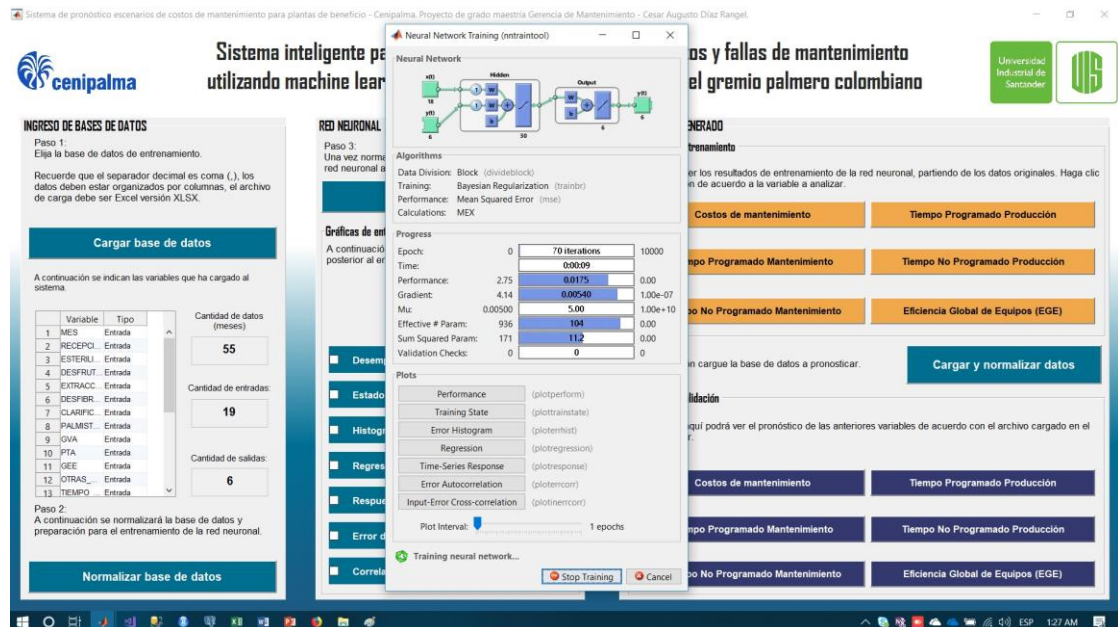
6. Objetivo error mínimo entrenamiento: predeterminado se ha establecido 0 como valor mínimo a buscar, es decir, la red deberá buscar que el indicador de error estadístico, generalmente RMSE, llegue lo más cercano posible a ese valor.
7. Tasa de entrenamiento: velocidad de iteración entre paso y paso de la red. Como valor predeterminado se estableció 0.001, dado que los indicadores de error convergieron entre 5 y 10 minutos de entrenamiento.
8. Mínimo de fallos en validación: por defecto se estableció 10.000, dado que es la cantidad de validaciones durante cada paso de entrenamiento.
9. Tasa de entrenamiento, validación, evaluación: (85, 10 y 5 por defecto)
 - a. El 85% de la base de datos se utilizará para entrenamiento.
 - b. El 10% se utilizará para validar que la red se está generalizando y para detener el entrenamiento antes del sobreajuste.
 - c. El último 5% se usará como una prueba completamente independiente de la generalización de la red.

10. Indicador estadístico para medir el error de la red durante el entrenamiento y validación:

- MSE: (predeterminado) error medio cuadrático. Se elige este indicador dado que sobre los demás, debido a que incorpora tanto la varianza del pronóstico como su sesgo.
- SSE: suma del error cuadrático.
- SAE: suma del error absoluto.
- MAE: media del error absoluto.

En la Figura 51 se muestra el proceso de entrenamiento de la red, mostrando la ventana emergente en donde se indica el avance del proceso en tiempo real.

Figura 51. Ventana de entrenamiento de red neuronal con indicadores de rendimiento, error, regresión, respuesta series de tiempo, entre otros.

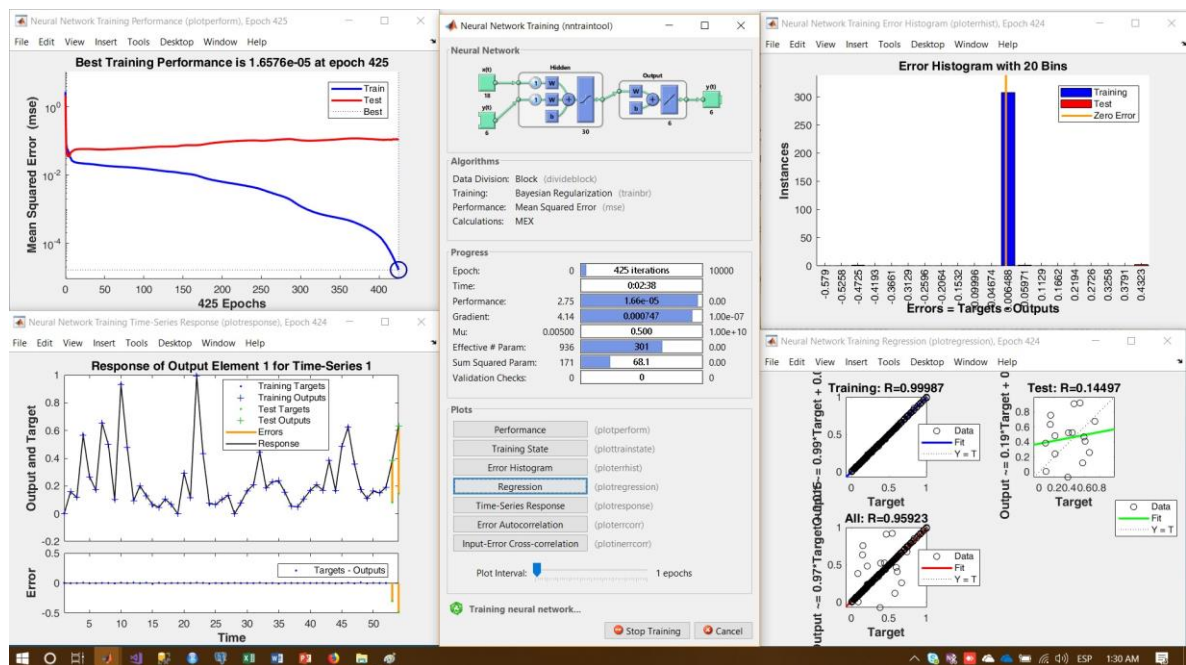


Entre las ventanas de gráficas a visualizar se aprecian (Figura 52):

- Desempeño de entrenamiento de acuerdo a indicador estadístico (MSE por defecto).
- Estados de entrenamiento: gráficas de optimización de entrenamiento y validación.

- Histograma de error: distribución estadística en tiempo real del error de entrenamiento.
- Regresión: se muestran las regresiones de entrenamiento, validación y evaluación, buscando los valores de R2 óptimos (lo más alto posible).
- Respuesta en series de tiempo: gráfica de respuesta ante entrenamiento, validación y evaluación, al igual que una gráfica sobre distribución del error.
- Error de autocorrelación: correlación cruzada de la señal consigo misma.
- Correlación cruzada de error de entrada: correlación cruzada de la señal consigo misma para señal de retroalimentación.

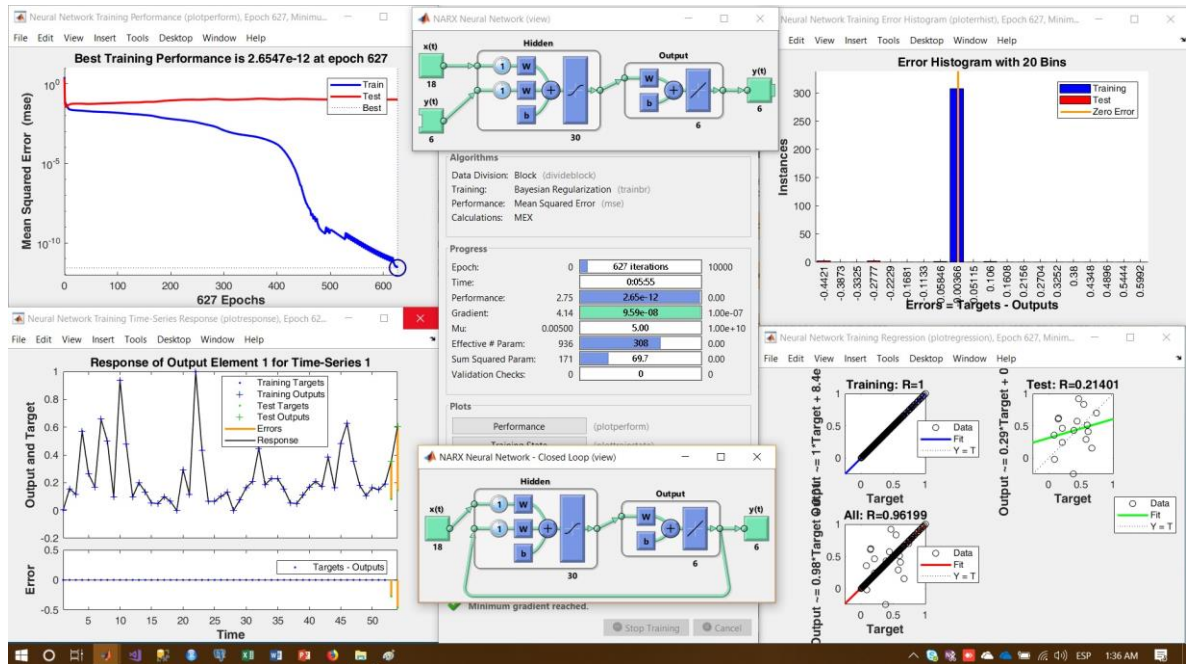
Figura 52. Ventanas auxiliares durante entrenamiento de red neuronal.



Estas ventanas se utilizaron para supervisar el entrenamiento, el cual duró entre 5 y 10 minutos. Se realizaron pruebas con más capas ocultas con más neuronas, sin embargo, la convergencia al error mínimo y los resultados luego de validación tuvieron un indicador de MSE por encima de 1, lo cual indica que la red tuvo problemas por saturación de neuronas. La configuración anterior predeterminada fue la mejor combinación entre MSE por debajo de 0.001 y tiempo de entrenamiento inferior a 10 minutos.

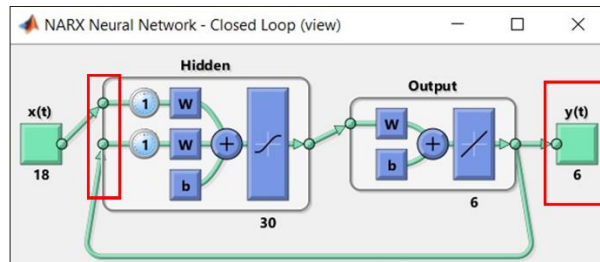
En la Figura 53 se muestra el resultado final luego de converger con $MSE = 2.65e-12$ al minuto 5:55.

Figura 53. Ventana de proceso de entrenamiento terminado con diagramas de redes en lazo abierto y lazo cerrado.



En la Figura 54 se muestra la red neuronal configurada inicialmente, con el lazo de retroalimentación, indicando que los errores generados al obtener las 6 variables de salida se tuvieron en cuenta para el reentrenamiento. Posteriormente para la validación de pronósticos se tuvo en cuenta la estructura de red con lazo cerrado.

Figura 54. Configuración de red neuronal resultante con retroalimentación.

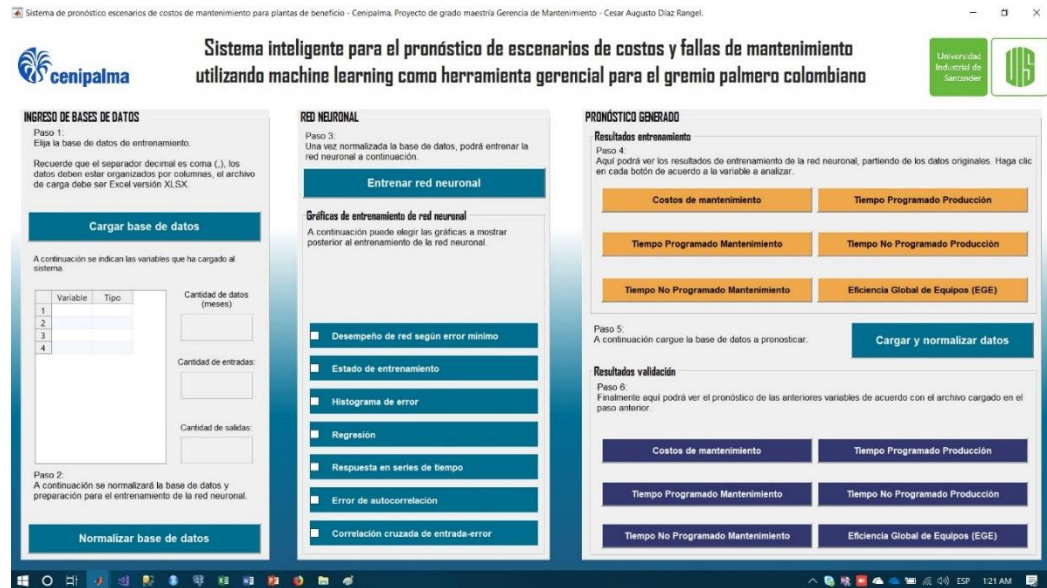


6 VARIABLES RESPUESTA
AJUSTADAS POR
REENTRENAMIENTO

LAZO CERRADO
SE RECIBEN NUEVAS VARIABLES DE
ENTRADA Y SE REAJUSTA RESPUESTA CON
ERROR DE SALIDA

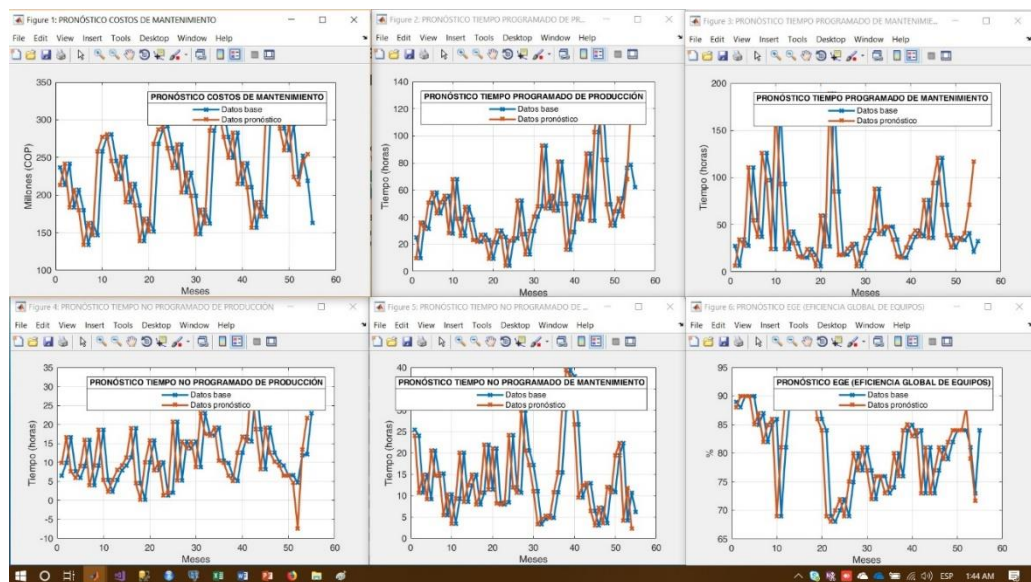
En el paso 4 se pueden graficar las variables de respuesta de forma individual, respecto a la base de datos de entrenamiento inicial.

Figura 55. Ventana de proceso de resultados de entrenamiento por cada variable de respuesta.



En la Figura 56 se muestran las 6 gráficas de respuesta pronóstico para entrenamiento, según la base de datos inicial.

Figura 56. Ventanas de gráficas para cada variable de respuesta, datos de entrenamiento contra respuesta modelo de pronóstico.



A continuación, desde la Figura 57 hasta la Figura 62 se muestran los comportamientos de pronóstico de forma detallada.

Figura 57. Comparación de pronóstico entrenamiento variable costos de mantenimiento.

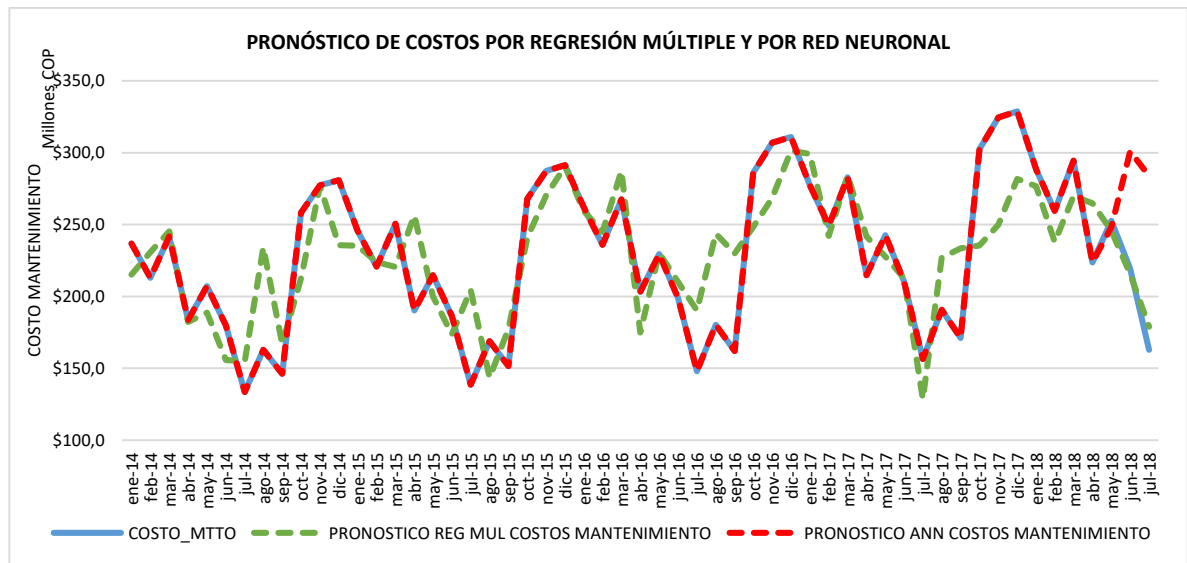


Figura 58. Comparación de pronóstico entrenamiento variable Eficiencia Global de Equipos (EGE).

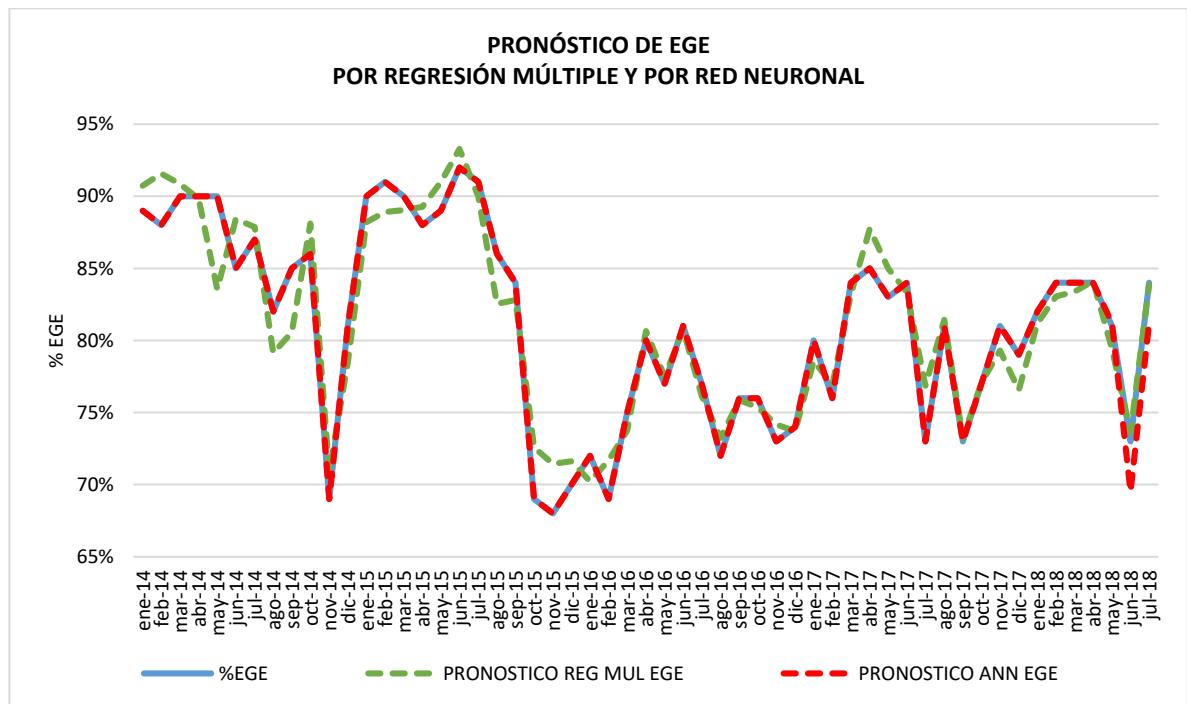


Figura 59. Comparación de pronóstico entrenamiento variable tiempo mantenimiento programado.

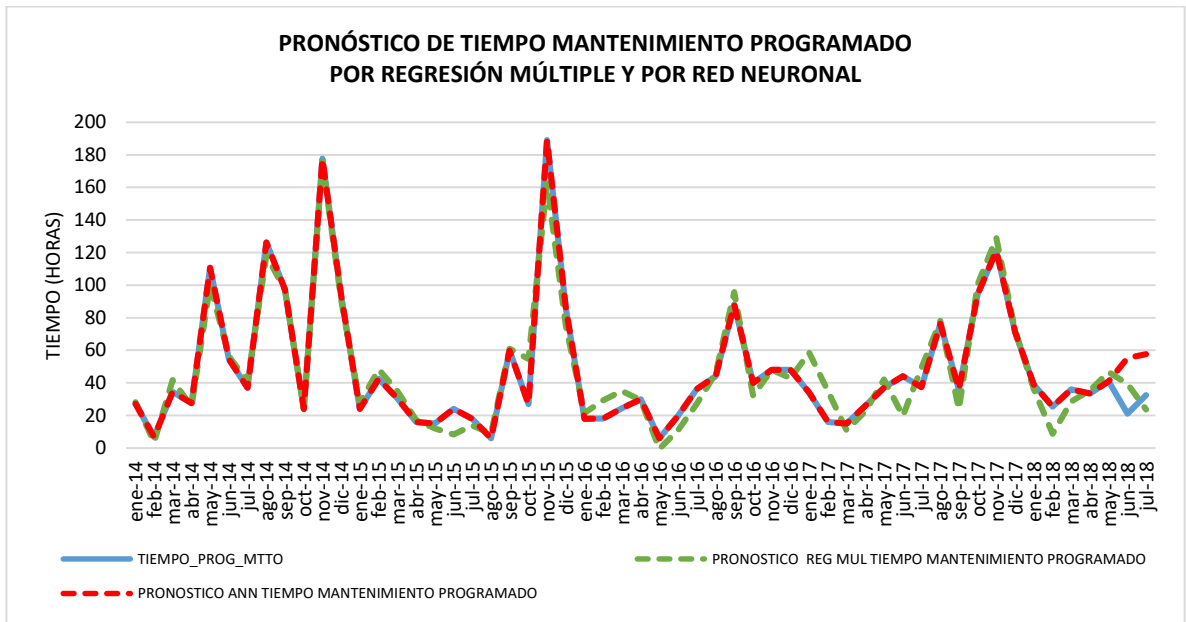


Figura 60. Comparación de pronóstico entrenamiento variable tiempo mantenimiento no programado.

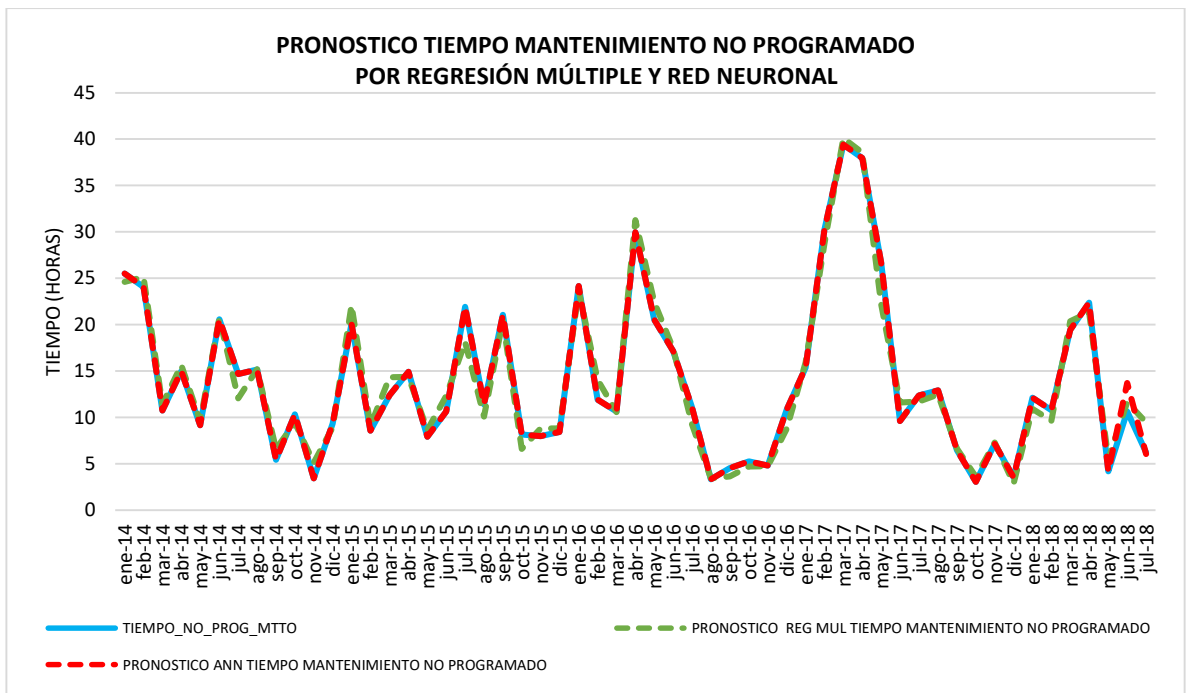


Figura 61. Comparación de pronóstico entrenamiento variable tiempo programado de producción.

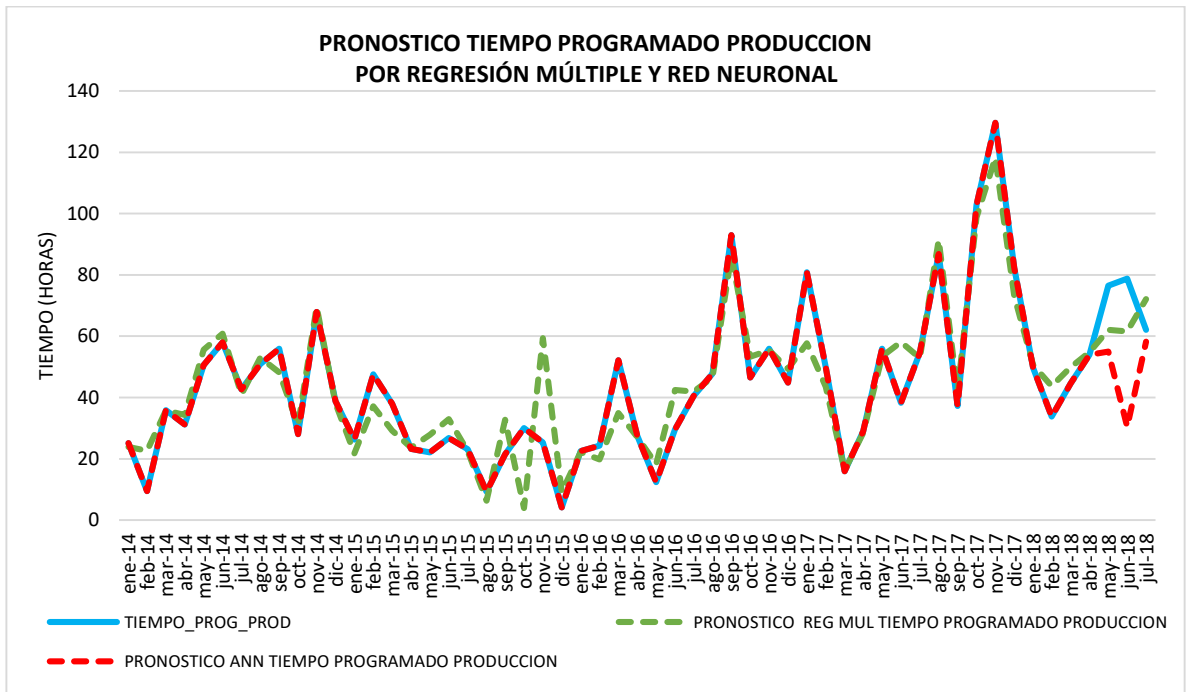
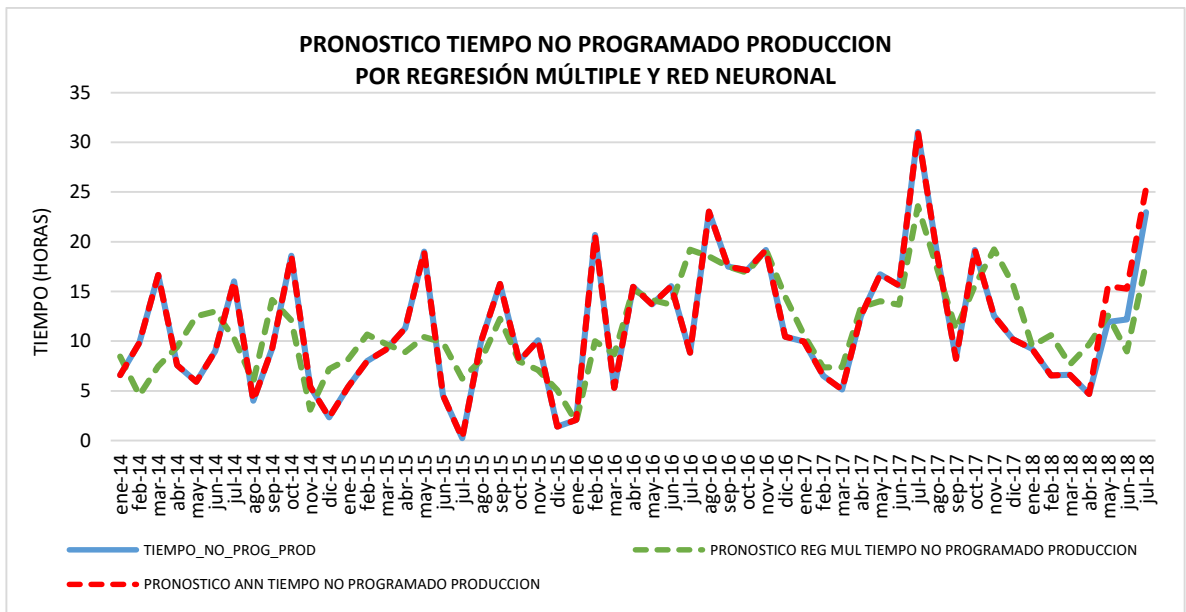


Figura 62. Comparación de pronóstico entrenamiento variable tiempo no programado producción.



7.3.2 Comparación modelos pronóstico regresión múltiple y red neuronal

En la Tabla 16 se muestran los indicadores estadísticos comparativos entre los métodos desarrollados, siendo regresión múltiple lineal (REG MUL) y red neuronal (RED NEUR).

Tabla 16. Resultados estadísticos regresión múltiple para variables de respuesta.

INDICADOR ESTADÍSTICO	VARIABLE DE RESPUESTA											
	EGE (REG MUL)	EGE (RED NEURONAL)	TIEMPO MANTENIMIENTO PROGRAMADO (REG MUL)	TIEMPO MANTENIMIENTO PROGRAMADO (RED NEUR)	TIEMPO MANTENIMIENTO NO PROGRAMADO (REG MUL)	TIEMPO MANTENIMIENTO NO PROGRAMADO (RED NEUR)	TIEMPO PROGRAMADO PRODUCCION (REG MUL)	TIEMPO PROGRAMADO PRODUCCION (RED NEUR)	TIEMPO NO PROGRAMADO PRODUCCION (REG MUL)	TIEMPO NO PROGRAMADO PRODUCCION (RED NEUR)	COSTOS DE MANTENIMIENTO (REG MUL)	COSTOS DE MANTENIMIENTO (RED NEUR)
S	0,091	0,009	0,057	0,005	0,044	0,004	0,085	0,008	0,149	0,015	0,188	0,018
R-cuad.	0,908	0,962	0,935	0,981	0,972	0,989	0,838	0,953	0,520	0,921	0,588	0,951
R-cuad. (ajustado)	0,899	0,911	0,929	0,961	0,965	0,977	0,818	0,931	0,471	0,907	0,516	0,937
R-cuad. (pred)	0,885	0,909	0,915	0,932	0,942	0,951	0,778	0,912	0,384	0,881	0,438	0,905
MSE	4,20E-04	3,66E-05	1,02E+02	3,29E+01	2,05E+00	1,73E-01	9,44E+01	5,11E+01	1,88E+01	5,16E-01	1,13E+03	3,88E+02
RMSE	2,05E-02	6,05E-03	1,01E+01	5,73E+00	1,43E+00	4,16E-01	9,72E+00	7,15E+00	4,34E+00	7,18E-01	3,36E+01	1,97E+01
MAD	0,016	0,001	7,371	1,080	1,101	0,059	6,690	1,340	3,416	0,165	26,113	3,764
CFE	-2,69E-04	1,11E-03	1,39E-01	-1,08E+00	-1,66E-02	-5,28E-02	9,62E-02	1,34E+00	-7,83E-02	-1,65E-01	4,58E-03	-3,61E+00
MAPE	1,98%	0,15%	25,60%	4,38%	10,52%	0,58%	21,51%	1,74%	83,67%	1,20%	12,64%	2,06%
DESV. EST ERROR	0,020	0,001	9,214	1,351	1,377	0,074	8,362	1,675	4,270	0,207	32,642	4,705

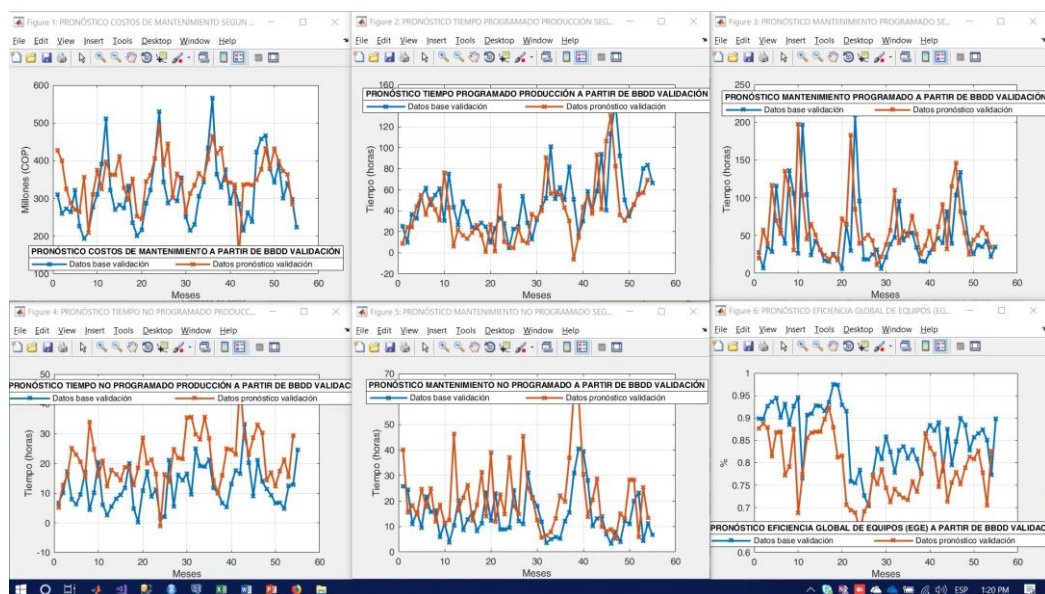
De acuerdo con lo anterior, se observa que el desempeño del modelo de pronóstico usando la red neuronal supera al modelo de regresión múltiple lineal, específicamente al comparar el indicador de RMSE, dado que en la red neuronal este valor es más bajo. Igualmente, al ver el indicador MAPE, la reducción porcentual del error es considerable, tal es el caso de la variable Costos de Mantenimiento, la cual pasó de 12.64% a 2.06% en reducción de error porcentual absoluto. Los coeficientes de determinación también tuvieron mejoría al desarrollar el modelo de red neuronal. Sin embargo, es necesario continuar reentrenando la red neuronal, incluso retroalimentar el sistema con errores de nuevos pronósticos, dado que el antecedente del tipo de redes para series de tiempo es su constante desactualización.

7.4 Escenarios generados a partir del pronóstico de variables asociadas a costos de mantenimiento y eficiencia de procesos

7.4.1 Escenario 1: incremento de paradas y fallas, reducción en materia prima y de su calidad

En la Figura 63 se muestran los resultados de pronóstico generados a través de la red neuronal, manteniendo la misma cantidad de períodos del entrenamiento (55 meses equivalente a 4 años y 7 meses). En el primer escenario de simulación se incrementó aleatoriamente el tiempo de falla entre el 15% y 25% de la base de datos original, para las áreas de esterilización, extracción, palmistería. Igualmente se redujo por igual la TEA en 1.5%, la cantidad de RFF procesado se redujo entre 10% y 20%. Las demás variables permanecieron similares a la base de datos inicial (misma usada en entrenamiento). En las gráficas para cada variable, la línea color azul corresponde al resultado estimado por regresión lineal múltiple, la línea color naranja corresponde al pronóstico por red neuronal. Se agregó el resultado por regresión múltiple con el fin de contar con una referencia adicional, dado que los datos pronosticados no existen en la base de datos original.

Figura 63. Pronóstico de variables de respuesta ante nuevo escenario número 1. (Usando base de datos Anexo C)



En la Tabla 17 se aprecian los estadísticos de error y desempeño, comparando los resultados por regresión lineal múltiple y resultados a través de red neuronal.

Tabla 17. Estadísticos entre pronóstico regresión múltiple y red neuronal escenario 1.

INDICADOR ESTADÍSTICO	VARIABLE DE RESPUESTA					
	EGE	TIEMPO MANTENIMIENTO PROGRAMADO	TIEMPO MANTENIMIENTO NO PROGRAMADO	TIEMPO PROGRAMADO PRODUCCION	TIEMPO NO PROGRAMADO PRODUCCION	COSTOS DE MANTENIMIENTO
MSE	5,91E-03	2,53E+02	8,18E+01	2,39E+02	1,25E+02	7,60E+03
RMSE	7,69E-02	1,59E+01	9,05E+00	1,55E+01	1,12E+01	8,72E+01
MAD	0,086	12,835	7,323	12,400	9,741	78,727
CFE	8,57E-02	-8,95E+00	-6,39E+00	9,55E+00	-8,83E+00	-2,67E+01
MAPE	9,89%	47,71%	56,17%	35,13%	208,92%	26,30%
DESV. EST ERROR	0,107	16,044	9,154	15,501	12,176	98,409

Dados los cambios mencionados anteriormente sobre la base de datos para simular el escenario 1, se observa que la variable de tiempo no programado de producción muestra la mayor diferencia entre ambos métodos, en donde el pronóstico por red neuronal resulta más alto que lo obtenido por regresión según el indicado MAPE.

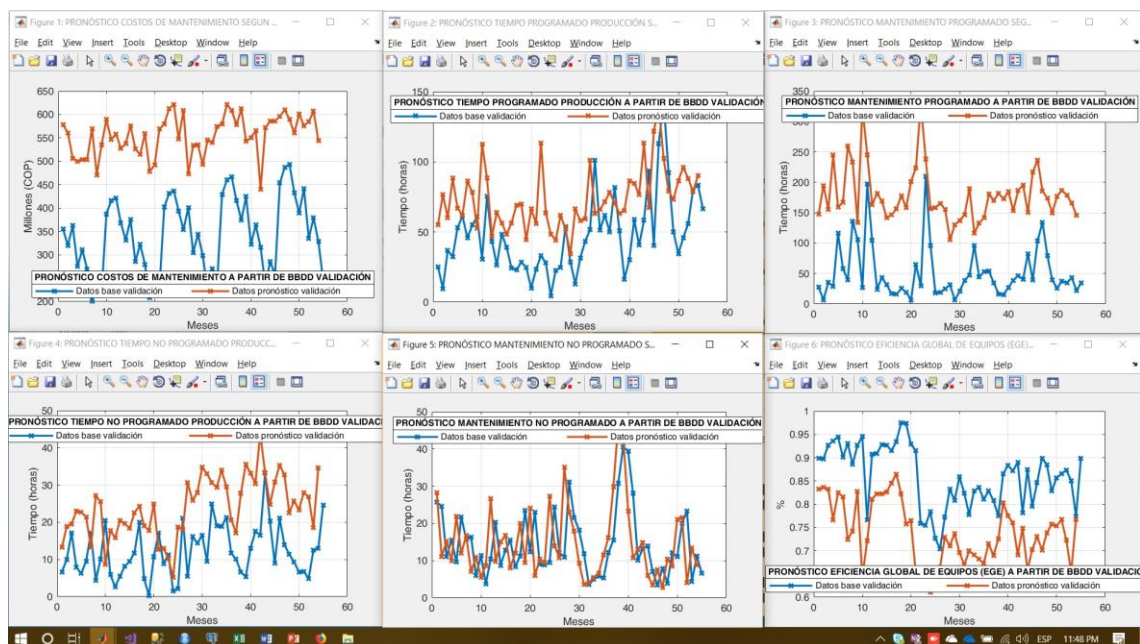
En las plantas de beneficio, uno de los factores que genera mayores problemas en la eficiencia general de la planta son las paradas no programadas y fallas de mantenimiento. Los eventos no contemplados tienden a solucionarse o mitigar sus efectos con servicios, repuestos y gastos administrativos de último momento, lo cual genera sobrecostos no contemplados en el presupuesto anual de la planta, sin olvidar los accidentes de trabajo y otros problemas colaterales.

Sumado a lo anterior, las condiciones de RFF procesado y TEA decrecieron de acuerdo con la base de datos inicial, afectando negativamente las finanzas de la planta. El factor TEA impacta directamente sobre la rentabilidad del negocio, dado que el precio de venta y negociaciones ante clientes, dependerá de los valores históricos de TEA. Este escenario es posible encontrarlo en al menos el 40% de las plantas de beneficio en Colombia (Díaz R. & García N., 2016).

7.4.2 Escenario 2: incremento de paradas y fallas, reducción en materia prima y de su calidad, descenso del precio internacional del aceite de palma (CIF Rotterdam)

En la Figura 64 se muestran los resultados de pronóstico generados a través de la red neuronal, manteniendo la misma cantidad de períodos del entrenamiento (55 meses equivalente a 4 años y 7 meses). En el segundo escenario de simulación se incrementó aleatoriamente el tiempo de falla entre el 25% y 35% de la base de datos original, para las áreas de esterilización, extracción, palmistería. Igualmente se redujo por igual la TEA en 3%, la cantidad de RFF procesado se redujo entre 40% y 50%. Las demás variables permanecieron similares a la base de datos inicial (misma usada en entrenamiento). En las gráficas para cada variable, la línea color azul corresponde al resultado estimado por regresión lineal múltiple, la línea color naranja corresponde al pronóstico por red neuronal. Se agregó el resultado por regresión múltiple con el fin de contar con una referencia adicional, dado que los datos pronosticados no existen en la base de datos original.

Figura 64. Pronóstico de variables de respuesta ante nuevo escenario número 2. (Usando base de datos Anexo D)



En la Tabla 18 se aprecian los estadísticos de error y desempeño, comparando los resultados por regresión lineal múltiple y resultados a través de red neuronal.

Tabla 18. Estadísticos entre pronóstico regresión múltiple y red neuronal escenario 2.

INDICADOR ESTADÍSTICO	VARIABLE DE RESPUESTA					
	EGE	MANTENIMIENTO PROGRAM	MTTO NO PROGRAMADO	TIEMPO PROGRAMADO PRODUCCION	TIEMPO NO PROGRAMADO PRODUCCION	COSTOS DE MANTENIMIENTO
MSE	1,53E-02	1,68E+04	4,88E+00	9,35E+02	1,77E+02	4,85E+04
RMSE	1,24E-01	1,30E+02	2,21E+00	3,06E+01	1,33E+01	2,20E+02
MAD	0,134	125,942	2,047	26,052	11,760	215,045
CFE	1,34E-01	-1,25E+02	1,82E-01	-2,47E+01	-1,14E+01	-2,02E+02
MAPE	15,55%	461,08%	13,75%	107,22%	267,21%	70,00%
DESV. EST ERROR	0,168	157,428	2,559	32,566	14,700	268,806

Dados los cambios mencionados anteriormente sobre la base de datos para simular el escenario 2, se observa que el cambio sobre la variable exógena del precio internacional del aceite generó afectación a nivel general sobre las 6 variables pronosticadas, especialmente sobre los costos de mantenimiento.

En este escenario también se afectó negativa y notablemente la cantidad de RFF procesada y su calidad, al igual que se incrementó el tiempo no programado por paradas de producción y fallas de mantenimiento.

Es posible la ocurrencia de un escenario de este tipo, especialmente para las plantas de beneficio localizadas en el departamento de Santander y Cesar, igualmente a las plantas localizadas en Tumaco (Nariño). Dada la contingencia y alta incidencia de una de las enfermedades más graves que ha sufrido la palmiticultura colombiana, Pudrición de Cogollo (PC), se han sembrado nuevos materiales tipo híbrido o modificados con el fin de resistir este tipo de plagas (esta enfermedad arrasó con alrededor de 70 mil hectáreas entre Puerto Wilches y Tumaco (Mosquera et al., 2017)). Sin embargo, este nuevo cultivar trae consigo problemas al momento de su procesamiento en planta, debido a que los equipos

existentes en el país no están adaptados para procesar fruto distinto al convencional, pues las condiciones morfológicas y composición mesocarpio-nuez-almendra son distintas a los cultivares tradicionales. Lo anterior induce a un escenario en el cual las plantas trabajen por debajo de su punto de equilibrio financiero, los equipos no mantengan la eficiencia general por encima del límite nominal (EGE al menos en 60% (Díaz R. & García N., 2016)), la cantidad de fruto procesado sea inferior a lo proyectado un año atrás, la TEA se reduzca debido a la mezcla prolongada de fruto convencional (*Elais Guineensis*) y fruto híbrido, entre otras condiciones negativas para la estabilidad financiera del negocio.

8 CONCLUSIONES

El proyecto contemplado como el desarrollo de un sistema inteligente para el pronóstico de escenarios de costos y fallas de mantenimiento, logró mostrar más información de la establecida inicialmente. Este beneficio puede representar alto valor estratégico para la gerencia de planta de beneficio, dado que puede conjugar los efectos sobre las variables de tiempo programado y no programado de producción y mantenimiento, Eficiencia Global de Equipos (EGE) y costos de mantenimiento, con el fin de generar nuevos escenarios para estar adelantado a la situación actual de la palmicultura a nivel global.

Partiendo del inventario bibliográfico y tecnológico constituido por desarrollos académicos y aplicados a la industrial en general, se extrajeron las mejores características con el fin de establecer el plan de trabajo del proyecto. Esta metodología permitió establecer diversas fases, iniciando por la consolidación y normalización de bases de datos, las cuales previamente fueron analizadas según métodos e indicadores estadísticos específicos para series de tiempo.

Las bases de datos fueron cedidas por una planta de beneficio piloto, cuyo proceso de identificación inicial reveló la existencia de 50 variables, de las cuales se eligieron 24 haciendo uso de métodos estadísticos de correlaciones y mejores subconjuntos. Es importante considerar la aplicación de la estadística como herramienta de exploración del comportamiento de los datos, dado que el reconocer inicialmente su naturaleza, permitirá utilizar adecuadamente los indicadores y otros métodos matemáticos requeridos implícitamente por el entrenamiento de redes neuronales.

El método de regresión lineal múltiple proporcionó una referencia de alto valor para el proyecto, debido a que permitió identificar las relaciones entre variables, variables imprescindibles y aquellas con bajo impacto sobre el modelo de pronóstico.

El uso de Matlab como herramienta computacional y de desarrollo de software permitió revelar las altas prestaciones y el panorama prometedor por el uso de este sistema. Definitivamente la integración de matemática, programación de software, estadística disponible en cada etapa de la aplicación, proporcionó beneficios tales como optimización en tiempo de desarrollo, depuración y ajustes del software de pronóstico.

El uso más allá de los resultados de pronóstico por entrenamiento demostró el alto valor de este tipo de alternativas, las cuales sirven como un método de futurista para ayudar a diseñar los nuevos panoramas, considerando los cambios en variables dentro del resorte de las plantas, al igual que aquellas variables tales como el precio internacional del aceite de palma, valor tranzado por el referente más grande en el mundo, Malasia.

BIBLIOGRAFÍA

- Arno, K. (2015). OEE Coaching. Retrieved from <http://oeecoach.com/oeo-oeo-teep-explained/>
- Baluch, N., Abdullah, C. S., & Mohtar, S. (2012). Measuring OEE in Malaysian Palm Oil Mills. *Interdisciplinary Journal Of Contemporary Research In Business*, 4(2), 733–743.
- Bell, J. (2015). *Machine Learning: Hands-On for Developers and Technical Professionals*. *Machine Learning* (Vol. 1). 10475 Crosspoint Boulevard Indianapolis, IN 46256: John Wiley & Sons, Inc. Retrieved from <https://www.wiley.com/en-us/Machine+Learning%3A+Hands+On+for+Developers+and+Technical+Professionals-p-9781118889497>
- Cenipalma. (2016). *Mejores prácticas agroindustriales del cultivo de la palma de aceite en Colombia*.
- D.C., H., & R.E., W. (2008). *Hat Matrix. The Concise Encyclopedia of Statistics*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-32833-1_177
- Díaz, C., Montero, J., Guevara, F., Cepeda, A., & Barrera, J. (2013). *Modelo para medición de eficiencia real de producción y administración integrada de información en Planta de Beneficio Producción*. *Boletín técnico No. 33*.
- Díaz R., C. A., & García N., J. A. (2016). Plan de Excelencia Industrial y estudio de los principales indicadores de desempeño en plantas de beneficio piloto en Colombia. *Revista Palmas*, 37(2016), 57–68. <https://doi.org/10.13140/RG.2.1.3342.8566>
- Fedepalma. (2017). *Anuario Estadístico 2017 Fedepalma* (2017th ed.). (FEDERACIÓN NACIONAL DE CULTIVADORES DE PALMA DE ACEITE (FEDEPALMA), Ed.). Bogotá, Colombia: Fedepalma.

- Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., B. Y. (1998). What is Data Science? Fundamental Concepts and a Heuristic Example (pp. 40–51). Springer Tokio. https://doi.org/10.1007/978-4-431-65950-1_3
- Kamble, L. V., Pangavhane, D. R., & Singh, T. P. (2015). Neural network optimization by comparing the performances of the training functions -Prediction of heat transfer from horizontal tube immersed in gas-solid fluidized bed. *International Journal of Heat and Mass Transfer*, 83, 337–344. <https://doi.org/10.1016/j.ijheatmasstransfer.2014.11.085>
- Lazakis, I., Raptodimos, Y., & Varelas, T. (2018). Predicting ship machinery system condition through analytical reliability tools and artificial neural networks. *Ocean Engineering*, 152(November 2017), 404–415. <https://doi.org/10.1016/j.oceaneng.2017.11.017>
- Liu, A. (IBM). (2015). Data Science and Data Scientist. IBM DATA ANALYTICS. Retrieved from <http://www.researchmethods.org/DataScienceDataScientists.pdf>
- LMC International. (2017). *The LMC Oilseeds & Oils Report*.
- Minitab-Inc. (2016). Minitab 17: manual. Retrieved from https://www.minitab.com/uploadedFiles/Documents/getting-started/Minitab17_GettingStarted-es-es.pdf
- Mosquera, M., Valderrama, M., Fontanilla, C., Ruíz, E., Uñate, M., Rincón, F., & Arias, N. (2016). Actualización de costos de producción para el fruto de palma de aceite y el aceite de palma en 2015: estimación en un grupo de productores colombianos. *Revista Palmas*, (July).
- Mosquera, M., Valderrama, M., Ruíz, E., López, D., Castro, L., Fontanilla, C., & González, M. (2017). Costos de producción para el fruto de palma de aceite y el aceite de palma en 2015: estimación en un grupo de productores colombianos. *Palmas*, 38(2), 11–27. Retrieved from http://web.fedepalma.org/media/01-Palmas-38-2-2017_VF_sin_marcas.pdf

- Nilsson, N. J. (2016). *Inteligencia Artificial una nueva síntesis*. Retrieved from <http://lagranjill.blogspot.com/2016/04/inteligencia-artificial.html>
- Paluszek, M., & Thomas, S. (2017). *MATLAB Machine Learning*. Berkeley, CA: Apress. <https://doi.org/10.1007/978-1-4842-2250-8>
- Patrick Jahnke. (2015). Machine Learning Approaches for Failure Type Detection and Predictive Maintenance, 83.
- SCHROEDER, R. (1996). *Administración de operaciones, toma de decisiones en la función de operaciones*.
- Stamatis, D. H. (2010). *The OEE Primer. Understanding Overall Equipment Effectiveness, Reliability, and Maintainability*. (& T. & Francis, Eds.). Productivity Press.
- USDA. (2018). *Oilseeds markets and trades*. Retrieved from <https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf>
- Van Gerven, M., & Bohte, S. (2017). *Artificial Neural Networks as Models of Neural Information Processing. Frontiers in Computational Neuroscience* (Vol. 11). <https://doi.org/10.3389/fncom.2017.00114>

ANEXOS

ANEXO A:

Anexo A: resultados regresión múltiple realizada con MINITAB

Resultados para: REGRESION_MULTIPLE_G_COSTOS.MTW

Análisis de regresión: TIEMPO_PROG_ vs. RECEPCION. ESTERILIZACI. DESFRUTACION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	4	78362,8	19590,7	175,27	0,000
TIEMPO_EFECTIVO	1	56301,0	56301,0	503,69	0,000
TIEMPO_PLANEADO	1	47107,7	47107,7	421,44	0,000
CAP_PRENSADO_EFECTIVA	1	9448,5	9448,5	84,53	0,000
TEA	1	785,9	785,9	7,03	0,011
Error	50	5588,9	111,8		
Total	54	83951,7			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
10,5725	93,34%	92,81%	91,34%

Coefficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	145,2	32,8	4,43	0,000	
TIEMPO_EFECTIVO	-0,6398	0,0285	-22,44	0,000	4,08
TIEMPO_PLANEADO	0,6556	0,0319	20,53	0,000	4,16
CAP_PRENSADO_EFECTIVA	-4,194	0,456	-9,19	0,000	1,09
TEA	-375	142	-2,65	0,011	1,06

Ecuación de regresión

$$\text{TIEMPO_PROG_MTTO} = 145,2 - 0,6398 \text{ TIEMPO_EFECTIVO} + 0,6556 \text{ TIEMPO_PLANEADO} - 4,194 \text{ CAP_PRENSADO_EFECTIVA} - 375 \text{ TEA}$$

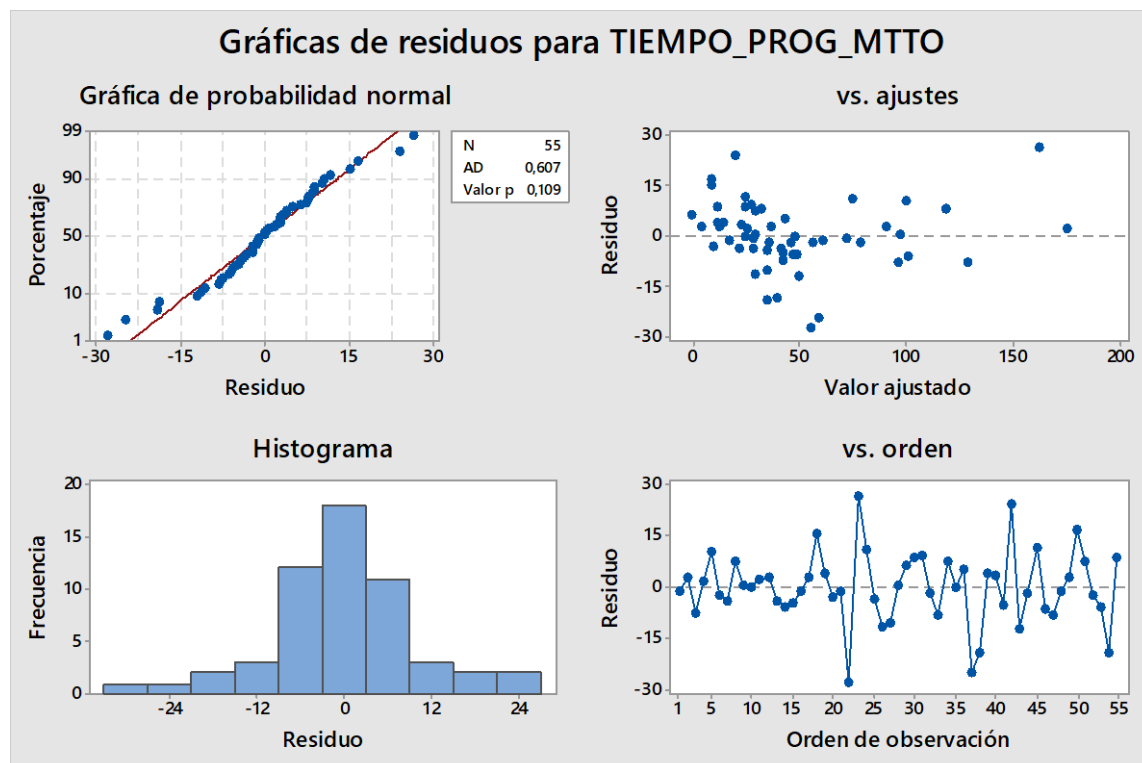
Ajustes y diagnósticos para observaciones poco comunes

Obs	TIEMPO_PROG_MTTO	Ajuste	Resid	Resid est.	
22	27,01	54,81	-27,80	-2,89	R
23	189,03	162,51	26,52	2,87	R
37	34,00	58,62	-24,62	-2,41	R
42	44,00	20,06	23,94	2,36	R

Residuo grande R

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 2,05872



Análisis de regresión: TIEMPO_NO_PR vs. RECEPCION. ESTERILIZACI. DESFRUTACION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	9	3847,44	427,493	163,57	0,000
ESTERILIZACION	1	20,70	20,705	7,92	0,007

DESFRUTACION	1	415,78	415,776	159,09	0,000
EXTRACCION	1	201,26	201,256	77,01	0,000
DESFIBRADO	1	360,79	360,790	138,05	0,000
CLARIFICACION	1	220,28	220,285	84,29	0,000
PALMISTERIA	1	30,66	30,656	11,73	0,001
GVA	1	18,63	18,632	7,13	0,011
OTRAS_MTTO	1	20,89	20,895	8,00	0,007
TIEMPO_EFECTIVO	1	46,51	46,513	17,80	0,000
Error	45	117,61	2,613		
Total	54	3965,04			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
1,61663	97,03%	96,44%	94,86%

Coefficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	-2,56	1,20	-2,13	0,038	
ESTERILIZACION	1,264	0,449	2,81	0,007	1,20
DESFRUTACION	0,9460	0,0750	12,61	0,000	1,71
EXTRACCION	0,994	0,113	8,78	0,000	1,54
DESFIBRADO	0,9037	0,0769	11,75	0,000	1,39
CLARIFICACION	1,106	0,120	9,18	0,000	1,29
PALMISTERIA	0,596	0,174	3,42	0,001	2,29
GVA	0,357	0,134	2,67	0,011	1,26
OTRAS_MTTO	0,0965	0,0341	2,83	0,007	1,53
TIEMPO_EFECTIVO	0,01272	0,00301	4,22	0,000	1,95

Ecuación de regresión

$$\text{TIEMPO_NO_PROG_MTTO} = -2,56 + 1,264 \text{ ESTERILIZACION} + 0,9460 \text{ DESFRUTACION}$$

$$+ 0,994 \text{ EXTRACCION} + 0,9037 \text{ DESFIBRADO} + 1,106 \text{ CLARIFICACION}$$

$$+ 0,596 \text{ PALMISTERIA} + 0,357 \text{ GVA} + 0,0965 \text{ OTRAS_MTTO} + 0,01272 \text{ TIEMPO_EFECTIVO}$$

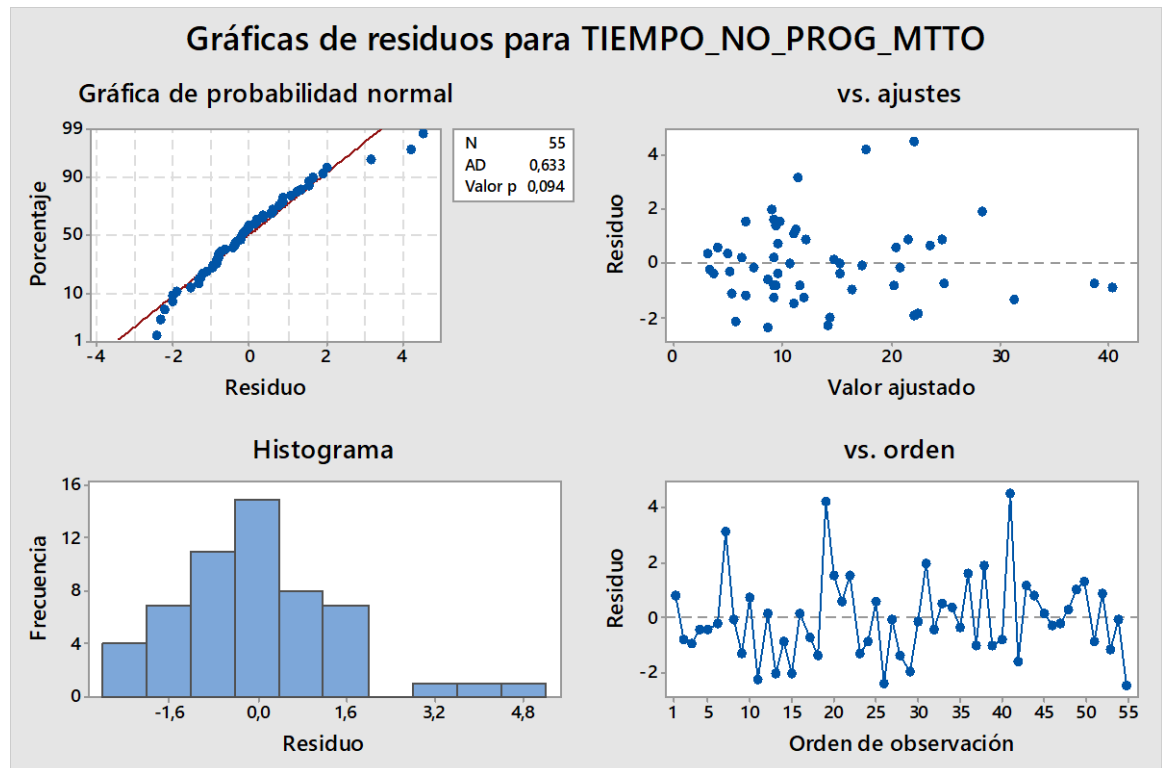
Ajustes y diagnósticos para observaciones poco comunes

Obs	TIEMPO_NO_PROG_MTTO	Ajuste	Resid	Resid est.	
6	20,580	20,770	-0,190	-0,25	X
7	14,670	11,501	3,169	2,04	R
19	21,940	17,707	4,233	2,75	R
41	26,710	22,177	4,533	3,05	R
52	22,370	21,478	0,892	1,38	X

Residuo grande R
 X poco común X

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 2,20611



Análisis de regresión: TIEMPO_PROG_ vs. RECEPCION. ESTERILIZACI. DESFRUTACION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	6	28345,9	4724,3	43,71	0,000
ESTERILIZACION	1	302,7	302,7	2,80	0,101
DESFUTACION	1	1101,8	1101,8	10,19	0,002
TIEMPO_EFECTIVO	1	20297,7	20297,7	187,81	0,000
TIEMPO_PLANEADO	1	15166,6	15166,6	140,33	0,000
CAP_PRENSADO_EFECTIVA	1	5065,0	5065,0	46,86	0,000
TEA	1	639,7	639,7	5,92	0,019
Error	48	5187,7	108,1		
Total	54	33533,7			

Resumen del modelo

	S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
10,3960		84,53%	82,60%	78,71%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	-144,5	32,4	-4,46	0,000	
ESTERILIZACION	-4,60	2,75	-1,67	0,101	1,09
DESFRUTACION	-1,390	0,435	-3,19	0,002	1,39
TIEMPO_EFECTIVO	-0,3866	0,0282	-13,70	0,000	4,13
TIEMPO_PLANEADO	0,3758	0,0317	11,85	0,000	4,24
CAP_PRENSADO_EFECTIVA	3,389	0,495	6,85	0,000	1,33
TEA	342	141	2,43	0,019	1,08

Ecuación de regresión

$$\text{TIEMPO_PROG_PROD} = -144,5 - 4,60 \text{ ESTERILIZACION} - 1,390 \text{ DESFRUTACION}$$

$$- 0,3866 \text{ TIEMPO_EFECTIVO} + 0,3758 \text{ TIEMPO_PLANEADO} + 3,389 \text{ CAP_PRENSADO_EFECTIVA}$$

$$+ 342 \text{ TEA}$$

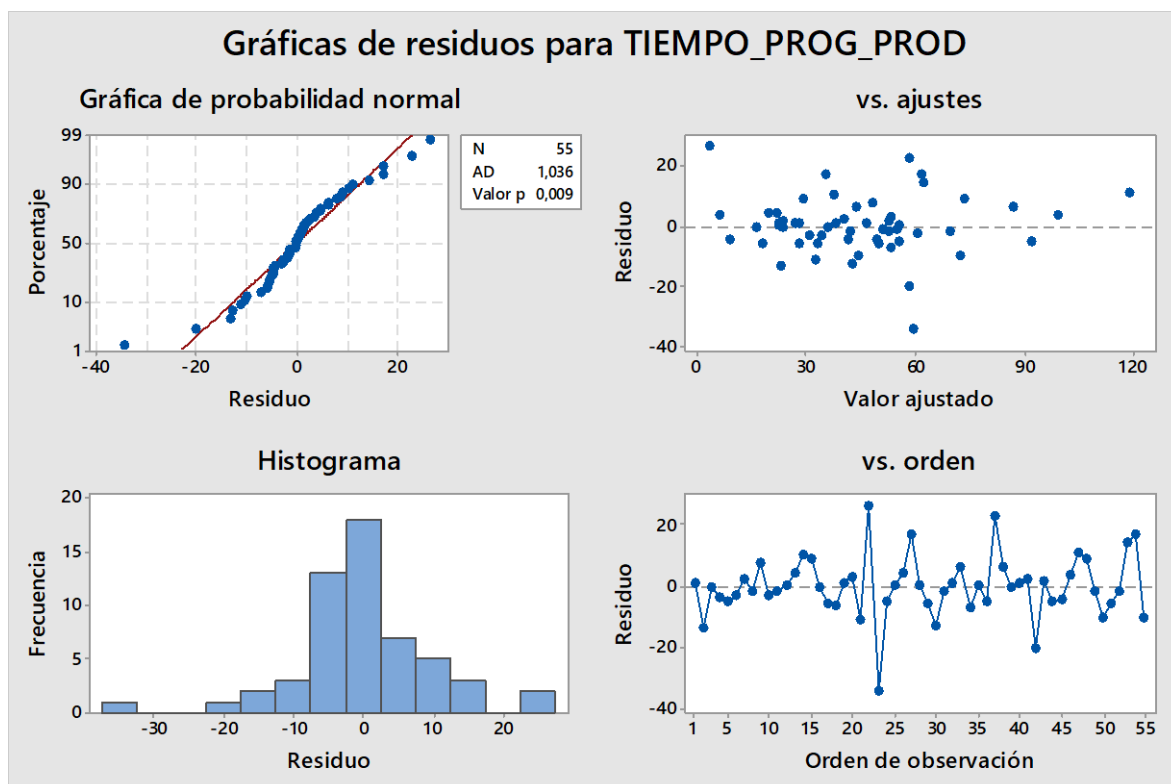
Ajustes y diagnósticos para observaciones poco comunes

Obs	TIEMPO_PROG_PROD	Ajuste	Resid	Resid est.	
22	30,04	3,62	26,42	2,79	R
23	25,33	59,50	-34,17	-3,76	R
37	80,91	58,14	22,77	2,29	R
40	28,98	27,94	1,04	0,13	X
42	38,27	58,26	-19,99	-2,02	R
43	54,67	52,93	1,74	0,22	X

Residuo grande R
 X poco común X

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 2,20431



Análisis de regresión: TIEMPO_NO_PR vs. RECEPCION. ESTERILIZACI. DESFRUTACION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	5	1115,43	223,09	10,58	0,000
ESTERILIZACION	1	350,90	350,90	16,64	0,000
PALMISTERIA	1	138,81	138,81	6,58	0,013
TIEMPO_EFECTIVO	1	63,80	63,80	3,03	0,088
CAP_PRENSADO_EFECTIVA	1	124,88	124,88	5,92	0,019
RFF_PROCESADO	1	85,63	85,63	4,06	0,049
Error	49	1033,26	21,09		
Total	54	2148,69			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
4,59205	51,91%	47,01%	38,17%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	-67,0	36,4	-1,84	0,072	
ESTERILIZACION	5,21	1,28	4,08	0,000	1,21
PALMISTERIA	-1,018	0,397	-2,57	0,013	1,47
TIEMPO_EFECTIVO	0,1407	0,0809	1,74	0,088	174,08
CAP_PRENSADO_EFECTIVA	3,36	1,38	2,43	0,019	52,98
RFF_PROCESADO	-0,00620	0,00308	-2,02	0,049	205,11

Ecuación de regresión

$$\text{TIEMPO_NO_PROG_PROD} = -67,0 + 5,21 \text{ ESTERILIZACION} - 1,018 \text{ PALMISTERIA} \\ + 0,1407 \text{ TIEMPO_EFECTIVO} + 3,36 \text{ CAP_PRENSADO_EFECTIVA} \\ - 0,00620 \text{ RFF_PROCESADO}$$

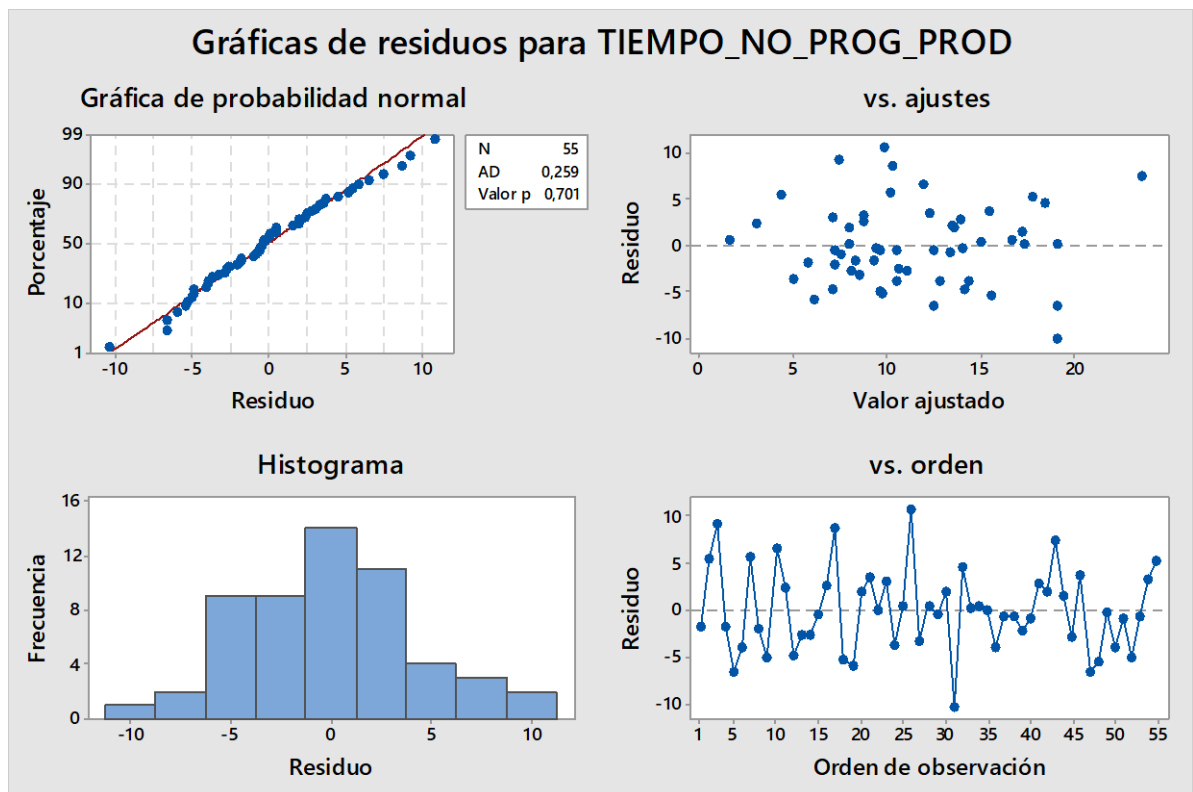
Ajustes y diagnósticos para observaciones poco comunes

Obs	TIEMPO_NO_PROG_PROD	Ajuste	Resid	Resid est.	
3	16,67	7,43	9,24	2,10	R
11	5,42	3,08	2,34	0,63	X
22	8,00	7,99	0,01	0,00	X
26	20,69	9,94	10,75	2,37	R
31	8,84	19,15	-10,31	-2,35	R
39	5,13	7,21	-2,08	-0,64	X
43	31,03	23,61	7,42	2,09	R X

Residuo grande R
 X poco común X

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 2,00751



Análisis de regresión: %OEE vs. RECEPCION. ESTERILIZACI. DESFRUTACION. EXTRACCION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	5	0,227786	0,045557	96,77	0,000
PALMISTERIA	1	0,002462	0,002462	5,23	0,027
CAP_PRENSADO_NOMINAL	1	0,159454	0,159454	338,70	0,000
CAP_PRENSADO_EFECTIVA	1	0,093723	0,093723	199,08	0,000
RFF_PROCESADO	1	0,007892	0,007892	16,76	0,000
TEA	1	0,001008	0,001008	2,14	0,150
Error	49	0,023068	0,000471		
Total	54	0,250855			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
0,0216975	90,80%	89,87%	88,51%

Coeficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	0,6876	0,0634	10,84	0,000	
PALMISTERIA	-0,00425	0,00186	-2,29	0,027	1,45
CAP_PRESADO_NOMINAL	-0,02011	0,00109	-18,40	0,000	2,20
CAP_PRESADO_EFECTIVA	0,02080	0,00147	14,11	0,000	2,70
RFF_PROCESADO	0,000006	0,000001	4,09	0,000	1,77
TEA	0,438	0,299	1,46	0,150	1,13

Ecuación de regresión

$$\%OEE = 0,6876 - 0,00425 \text{ PALMISTERIA} - 0,02011 \text{ CAP_PRESADO_NOMINAL} + 0,02080 \text{ CAP_PRESADO_EFECTIVA} + 0,000006 \text{ RFF_PROCESADO} + 0,438 \text{ TEA}$$

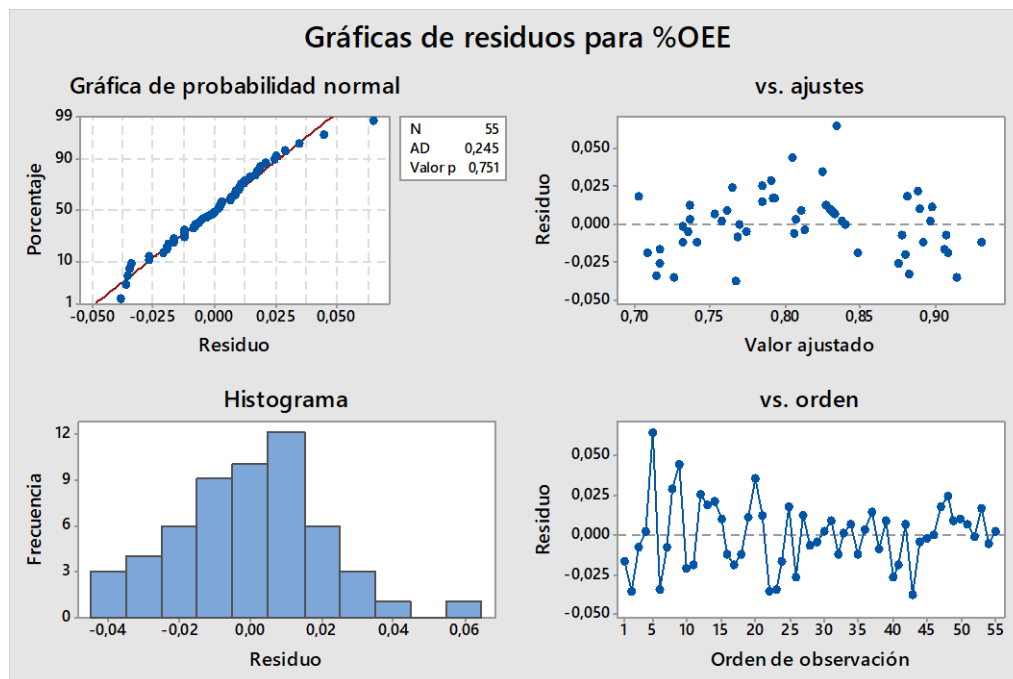
Ajustes y diagnósticos para observaciones poco comunes

Obs	%OEE	Ajuste	Resid	Resid est.	
5	0,90000	0,83536	0,06464	3,05	R
9	0,85000	0,80558	0,04442	2,11	R
39	0,84000	0,83115	0,00885	0,51	X

Residuo grande R
X poco común X

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 1,83772



Análisis de regresión: COSTO_MTTTO vs. RECEPCION. ESTERILIZACI. DESFRUTACION. EXTRACCION. ...

Selección de términos escalonada

α a entrar = 0,15. α a retirar = 0,15

Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Regresión	8	8,90238E+16	1,11280E+16	8,25	0,000
ESTERILIZACION	1	2,04182E+16	2,04182E+16	15,14	0,000
DESFIBRADO	1	7,70527E+15	7,70527E+15	5,71	0,021
PALMISTERIA	1	1,26485E+16	1,26485E+16	9,38	0,004
PTA	1	6,27463E+15	6,27463E+15	4,65	0,036
OTRAS_MTTTO	1	7,19055E+15	7,19055E+15	5,33	0,025
CAP_PRENSADO_NOMINAL	1	2,42498E+16	2,42498E+16	17,98	0,000
CAP_PRENSADO_EFECTIVA	1	3,05386E+15	3,05386E+15	2,26	0,139
TEA	1	3,91715E+16	3,91715E+16	29,04	0,000
Error	46	6,20446E+16	1,34879E+15		
Total	54	1,51068E+17			

Resumen del modelo

S	R-cuad.	R-cuad. (ajustado)	R-cuad. (pred)
36725940	58,93%	51,79%	44,15%

Coefficientes

Término	Coef	EE del coef.	Valor T	Valor p	VIF
Constante	763312878	115833242	6,59	0,000	
ESTERILIZACION	-39772253	10222207	-3,89	0,000	1,21
DESFIBRADO	-4468131	1869412	-2,39	0,021	1,59
PALMISTERIA	10394940	3394494	3,06	0,004	1,69
PTA	-31121542	14429128	-2,16	0,036	1,82
OTRAS_MTTTO	-2108764	913313	-2,31	0,025	2,12
CAP_PRENSADO_NOMINAL	8130122	1917415	4,24	0,000	2,36
CAP_PRENSADO_EFECTIVA	-3485353	2316301	-1,50	0,139	2,33
TEA	-3105277934	576220019	-5,39	0,000	1,46

Ecuación de regresión

COSTO_MTTTO = 763312878 - 39772253 ESTERILIZACION - 4468131 DESFIBRADO
+ 10394940 PALMISTERIA
- 31121542 PTA - 2108764 OTRAS_MTTTO + 8130122 CAP_PRENSADO_NOMINAL
- 3485353 CAP_PRENSADO_EFECTIVA - 3105277934 TEA

Ajustes y diagnósticos para observaciones poco comunes

Obs	COSTO_MTTTO	Ajuste	Resid	Resid est.	
8	162861108	233994899	-71133791	-2,02	R
20	168856267	144558613	24297654	0,94	X

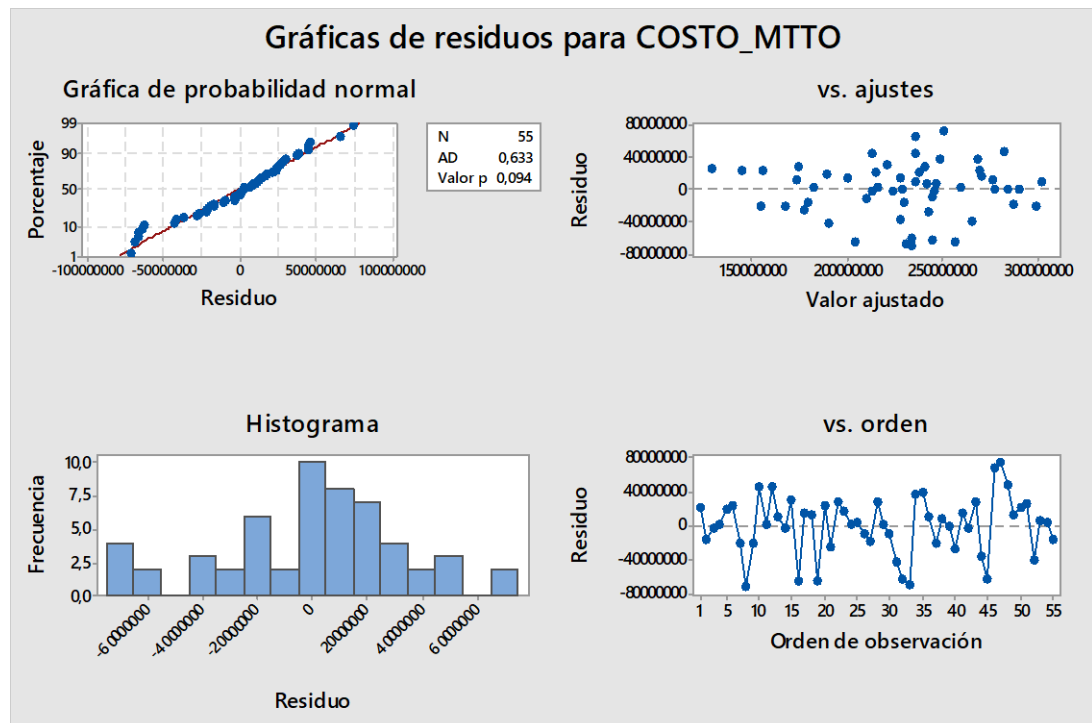
25	262147299	259098226	3049073	0,12	X
33	161881380	230637286	-68755906	-2,17	R
39	282943240	283841841	-898601	-0,04	X
47	324553068	250232529	74320539	2,11	R

Residuo grande R

X poco común X

Estadístico de Durbin-Watson

Estadístico de Durbin-Watson = 1,75897



Los siguientes anexos están adjuntos en el CD y puede visualizarlos en base de datos de la biblioteca UIS:

ANEXO B:

Anexo B: Base de datos para entrenamiento

Nombre de archivo digital contenido en disco del proyecto:

BBDD_ENTRENAMIENTO.xlsx.

ANEXO C:

Anexo C: Base de datos para validación escenario 1

Nombre de archivo digital contenido en disco del proyecto:

BBDD_VALIDACION_ESCENARIO_1.xlsx.

ANEXO D:

Anexo D: Base de datos para validación escenario 2

Nombre de archivo digital contenido en disco del proyecto:

BBDD_VALIDACION_ESCENARIO_2.xlsx.