



**FACULTAD DE POSTGRADO
TRABAJO FINAL DE GRADUACIÓN**

**UN MODELO DE ANALÍTICA PARA PREDECIR EL
PRESUPUESTO EN LOS SECTORES DE SALUD Y
EDUCACIÓN EN HONDURAS A 2024**

SUSTENTADO POR:

**OSCAR EDUARDO ZÚNIGA GALLEGOS
SAMUEL HERIBERTO ZELAYA VALLE**

PREVIA INVESTIDURA AL TÍTULO DE

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

TEGUCIGALPA, FRANCISCO MORAZÁN, HONDURAS, C.A.

FEBRERO, 2024

UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

UNITEC

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

PRESIDENTE EJECUTIVO /

RECTORA

ROSALPINA RODRÍGUEZ

SECRETARIO GENERAL /

PRORRECTOR

ROGER MARTÍNEZ MIRALDA

VICERRECTOR ACADÉMICO NACIONAL

JAVIER ABRAHAM SALGADO LEZAMA

DIRECTORA NACIONAL DE POSTGRADO

ANA DEL CARMEN RETTALLY

**UN MODELO DE ANALÍTICA PARA PREDECIR EL
PRESUPUESTO EN LOS SECTORES DE SALUD Y
EDUCACIÓN EN HONDURAS A 2024**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE**

MÁSTER EN

ANALÍTICA DE NEGOCIOS

ASESORA METODOLÓGICA

ALBA GABRIELA GARAY ROMERO

MIEMBROS DE LA TERNA:

KEVIN EDUARDO FÚNEZ FÚNEZ

ERICK LEONARDO IZAGUIRRE ARDÓN

DANIEL ANTONIO LUNA RODRÍGUEZ

DERECHOS DE AUTOR

© Copyright 2024
Oscar Eduardo Zúniga Gallegos
Samuel Heriberto Zelaya Valle

Todos los derechos son reservados.



FACULTAD DE POSTGRADO

UN MODELO DE ANALÍTICA PARA PREDECIR EL PRESUPUESTO EN LOS SECTORES DE SALUD Y EDUCACIÓN EN HONDURAS A 2024

**Oscar Eduardo Zúniga Gallegos
Samuel Heriberto Zelaya Valle**

Resumen

La ejecución presupuestaria representa un pilar fundamental para los países en vías de desarrollo y es que a partir de esta se puede determinar el desempeño en sectores clave como lo son la educación y la salud. Asimismo, el auge que tiene actualmente la ciencia de datos es una oportunidad para que a partir de esta se brinden soluciones viables y que muestren resultados en pro de la ejecución presupuestaria a través de la creación y configuración de modelos de analítica como las series temporales y los bosques aleatorios, permitiendo efectuar predicciones de presupuestos y la identificación de objetos de gasto en los que más se invierten estos. Lo anterior significa que cualquier institución sea pública o privada que haga uso de esta tecnología cuenta con la alternativa de mejorar la ejecución presupuestaria y por ende ser más transparente en la rendición de cuentas.

Palabras claves: (Aprendizaje Automático, Bosques Aleatorios, Ejecución Presupuestaria, Modelos Predictivos, Series Temporales)



GRADUATE SCHOOL

AN ANALYTICAL MODEL TO PREDICT THE BUDGET IN THE HEALTH AND EDUCATION SECTORS IN HONDURAS FOR 2024

**Oscar Eduardo Zúniga Gallegos
Samuel Heriberto Zelaya Valle**

Abstract

Budget execution represents a fundamental pillar for developing countries, since it can be used to determine performance in key sectors such as education and health. Likewise, the current boom in data science is an opportunity to provide viable solutions that show results in favor of budget execution through the creation and configuration of analytical models such as time series and random forests, allowing budget predictions and the identification of spending objects in which they are most invested. This means that any institution, whether public or private, that makes use of this technology has the alternative of improving budget execution and therefore being more transparent in accountability.

Palabras claves: (Budget Execution, Machine Learning, Predictive Modeling, Random Forests, Time Series)

DEDICATORIA

A Dios padre todopoderoso, por darme la fortaleza y la guía necesaria para afrontar cada uno de los retos y vivencias afrontados durante la maestría.

A mis padres y el resto de mi amada familia, por brindarme siempre su amor y el apoyo incondicional en cada etapa de mi vida personal y profesional, por sentir cada uno de mis logros como si fuera suyo.

Oscar Eduardo Zúniga Gallegos

A Dios, primeramente, por haberme dado fuerzas para lograr todos y cada uno de los desafíos presentados durante todo el tiempo de la maestría.

A mis padres, esposa e hijo por ser ese motor que me impulsa cada día para seguir cosechando más logros y permanecer firme encaminado hacia el cumplimiento de mis objetivos profesionales y personales.

Samuel Heriberto Zelaya Valle

AGRADECIMIENTO

A cada uno de los profesionales que formaron parte del cuerpo docente de la Maestría en Analítica de Negocios, especialmente a Kevin Fúnez y Erick Izaguirre, ejemplos claros de éxito a nivel de docencia y profesionalmente, con amplia experiencia y con la habilidad de despertar el interés del maestrante con cátedras de gran nivel.

A mis compañeros de equipo a lo largo de la maestría: Samuel Zelaya, Martha Benítez, Ana Carrillo y Emili Flores, que aparte de brindarme su amistad, fueron pilares fundamentales para alcanzar el éxito en cada uno de los retos que enfrentamos en el transcurso de las asignaturas, con su dedicación, amplia experiencia y compromiso.

Oscar Eduardo Zúniga Gallegos

A todos los docentes que formaron parte en todo el proceso de la Maestría de Analítica de Negocios, especialmente a Kevin Fúnez, Julio Esteban Ramos, Erick Izaguirre su dedicación y conocimiento es valorado.

A mis compañeros Oscar Zúniga, Martha Benítez, Ana Carrillo y Emili Flores por haber sido ese complemento durante todas las jornadas de estudio, su apoyo y disposición fue pieza fundamental para el cumplimiento de este logro.

A mi amigo Daniel Gómez Santeli por brindarme siempre su apoyo y animarme a permanecer siempre luchando por mis objetivos.

Samuel Heriberto Zelaya Valle

ÍNDICE DE CONTENIDO

DEDICATORIA	1
AGRADECIMIENTO	2
ÍNDICE DE CONTENIDO	3
ÍNDICE DE TABLAS	7
ÍNDICE DE FIGURAS.....	8
CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN	10
1.1 INTRODUCCIÓN	10
1.2 ANTECEDENTES DEL PROBLEMA	10
1.3 DEFINICIÓN DEL PROBLEMA	11
1.3.1 ENUNCIADO DEL PROBLEMA.....	11
1.3.2 FORMULACIÓN DEL PROBLEMA	12
1.3.3 PREGUNTAS DE INVESTIGACIÓN.....	12
1.4 OBJETIVOS DEL PROYECTO.....	13
1.4.1 OBJETIVO GENERAL.....	13
1.4.2 OBJETIVOS ESPECÍFICOS.....	13
1. IDENTIFICAR LOS OBJETOS DE GASTO CON MAYOR AFECTACIÓN EN LA EJECUCIÓN DEL PRESUPUESTO DE LOS SECTORES DE SALUD Y EDUCACIÓN... 13	
2. DETERMINAR EL NIVEL DE EFICIENCIA EN LA EJECUCIÓN DEL PRESUPUESTO DE LOS SECTORES DE SALUD Y EDUCACIÓN.	13
1.5 JUSTIFICACIÓN.....	13
CAPÍTULO II. MARCO TEÓRICO	15
2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL	15
2.2 CONCEPTUALIZACIÓN	25
2.2.1 ANALÍTICA DE DATOS	25
2.2.2 APRENDIZAJE AUTOMÁTICO	25
2.2.3 ÁRBOL DE DECISIÓN	26
2.2.4 BOSQUES ALEATORIOS	26
2.2.5 CLUSTERIZACIÓN	26
2.2.6 CORRELACIÓN.....	26
2.2.7 EJECUCIÓN PRESUPUESTARIA	26

2.2.8	FUENTES DE DATOS	27
2.2.9	K-MEANS	27
2.2.10	MATRIZ DE CONFUSIÓN	27
2.2.11	MODELO NO SUPERVISADO.....	27
2.2.12	MODELO PREDICTIVO	27
2.2.13	MODELO SUPERVISADO	28
2.2.14	OBJETO DE GASTO.....	28
2.2.15	PRESUPUESTO.....	28
2.2.16	PRESUPUESTO VIGENTE	28
2.2.17	REGRESIÓN LINEAL	29
2.2.18	SUPPORT VECTOR MACHINE.....	29
2.2.19	VISUALIZACIÓN DE DATOS	30
2.3	TEORÍAS DE SUSTENTO	30
2.3.1	BASES TEÓRICAS	30
2.3.1.1	ECONOMÍA KEYNESIANA	30
2.3.1.2	TEORÍA DE LA PROBABILIDAD	31
2.3.1.3	TEORÍA DE LA ESTADÍSTICA	32
2.3.1.4	TEORÍA DE SISTEMAS	33
2.3.1.5	TEORÍA DE OPTIMIZACIÓN	33
2.3.1.6	TEORÍA DE APRENDIZAJE AUTOMÁTICO	34
2.3.1.7	TEORÍA DE MINERÍA DE DATOS	35
2.3.2	METODOLOGÍAS DESARROLLADAS	36
2.4	MARCO LEGAL	37
CAPÍTULO III. METODOLOGÍA		39
3.1	CONGRUENCIA METODOLÓGICA.....	39
3.1.1	MATRIZ METODOLÓGICA.....	39
3.1.2	OPERACIONALIZACIÓN DE LAS VARIABLES	41
3.2	ENFOQUE Y MÉTODOS.....	43
3.2.1	ENFOQUE	43
3.2.3	DISEÑO.....	44
3.2.4	INSTRUMENTO.....	44

3.2.5	TÉCNICA.....	44
3.3	DISEÑO DE LA INVESTIGACIÓN	45
3.3.1	POBLACIÓN	45
3.4	TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	45
3.5	FUENTES DE INFORMACIÓN.....	45
3.5.1	FUENTES PRIMARIAS	45
3.5.2	FUENTES SECUNDARIAS.....	46
CAPÍTULO IV. RESULTADOS Y ANÁLISIS		46
4.1	INFORME DE PROCESO DE RECOLECCIÓN DE DATOS	46
4.1.1	PREPARACIÓN Y LIMPIEZA DE CONJUNTO DE DATOS.....	48
4.2	RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS	49
4.2.1	RESULTADOS CUANTITATIVOS	49
4.2.1.1	SECTOR EDUCACIÓN	49
4.2.1.2	SECTOR SALUD.....	57
4.2.2	ANÁLISIS CUALITATIVO	64
4.3	ANÁLISIS INFERENCIAL Y MODELOS APLICADOS.....	64
4.3.1	SECTOR EDUCACIÓN	64
4.3.1.1	SERIES TEMPORALES	64
4.3.1.2	RANDOM FOREST.....	67
4.3.1.2.1	RANDOM FOREST (SIN DATOS DE AÑO 2023)	69
4.3.2	SECTOR SALUD.....	70
4.3.2.1	SERIES TEMPORALES	70
4.3.2.2	RANDOM FOREST.....	73
4.3.2.3	RANDOM FOREST (SIN DATOS DEL AÑO 2023)	75
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....		77
5.1	CONCLUSIONES	77
5.2	RECOMENDACIONES	78
CAPÍTULO VI. APLICABILIDAD.....		79
6.1	NOMBRE DE LA PROPUESTA	79
6.2	JUSTIFICACIÓN DE LA PROPUESTA	79
6.3	ALCANCE DE LA PROPUESTA	79

6.3.1	OBJETIVO GENERAL	80
6.3.2	OBJETIVOS ESPECÍFICOS	80
6.4	DESCRIPCIÓN Y DESARROLLO	80
6.4.1	DESCRIPCIÓN	80
6.4.2	DESARROLLO.....	80
6.4.2.1	<i>MODELO ARIMA</i>	83
6.4.2.2	<i>MODELO RANDOM FOREST</i>	91
6.4.2.3	<i>SELECCIÓN DE INFRAESTRUCTURA</i>	95
6.4.2.4	<i>ESPECIFICACIONES DE INFRAESTRUCTURA</i>	96
6.5	MEDIDAS DE CONTROL	96
6.6	CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO	99
6.6.1	TABLAS DE COSTOS	100
6.6.2	CRONOGRAMA DE ACTIVIDADES	101
6.6.3	ESTIMACIÓN TOTAL DE COSTOS DEL PROYECTO	102
6.7	PLAN DE COMUNICACIÓN.....	102
6.8	CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA	
	103	
	REFERENCIAS BIBLIOGRÁFICAS.....	105
	ANEXOS	109
	ANEXO 1 WORKFLOW DEL ANÁLISIS EXPLORATORIO DE LOS DATOS EN KNIME.	
	109

ÍNDICE DE TABLAS

Tabla 1. Matriz de Congruencia Metodológica	39
Tabla 2. Operacionalización de las variables.....	42
Tabla 3. Campos utilizados en conjunto de datos.....	47
Tabla 4. Resultados del Modelo.....	67
Tabla 5. Resultados del Modelo.....	68
Tabla 6. Resultados de Random Forest (Sin 2023).....	70
Tabla 7. Resultados del Modelo.....	72
Tabla 8. Resultados del Modelo.....	74
Tabla 9. Resultados del Modelo (Random Forest) - Sin 2023.....	76
Tabla 10. Especificaciones de Servidor On Cloud Azure.....	96
Tabla 11. Especificaciones de Sistema Gestor de Base de Datos.....	96
Tabla 12. Medición de riesgos detectados.	97
Tabla 13. Pago por hora recursos humanos.	100
Tabla 14. Costos en Tecnología (Mensuales) - OPEX.	100
Tabla 15. Costos en Tecnología - CAPEX	100
Tabla 16. Estimación Total.	102
Tabla 17. Concordancia de los segmentos de la investigación.....	104

ÍNDICE DE FIGURAS

Figura 1. Gasto Público en Salud 2022 de los países de norte américa.....	16
Figura 2. Gasto público de Educación 2022 en los países europeos.	16
Figura 3. TOP 5 Países con mejor educación	17
Figura 4. Comparación de población vs. gasto per cápita en la educación	18
Figura 5. Comparativa de Población vs Gasto Per cápita de salud.....	19
Figura 6. Ranking de países con mayor presupuesto en salud en América latina	20
Figura 7. Gasto público para educación y porcentaje sobre el PIB	21
Figura 8. Países con más inversión en educación en américa latina en relación con el PIB	22
Figura 9. Distribución de Centros Educativos por Departamento.	23
Figura 10. Relación de Presupuesto de Seguridad vs Salud en Honduras.....	24
Figura 11. Eficiencia de la Educación en Honduras	25
Figura 12. Definición del "Margen" entre clases: El criterio que los SVM intentan optimizar ...	30
Figura 13. Temas de Investigación Estadística.	32
Figura 14. Teoría de Sistemas.....	33
Figura 15. Proceso de Aprendizaje Automático.	35
Figura 16. Proceso de Minería de Datos.	36
Figura 17. Esquema de relación causal multivariada.....	40
Figura 18. Esquema de variable independiente	40
Figura 19. Esquema de variable dependiente	41
Figura 20. Enfoque y métodos	43
Figura 21. Ejecución presupuestaria del sector educación por año.	49
Figura 22. Serie de ejecución presupuestaria del sector educación por período.	50
Figura 23. Ejecución presupuestaria del sector educación por objeto de gasto.....	51
Figura 24. Ejecución presupuestaria del sector educación por grupo de gasto.	51
Figura 25. Ejecución presupuestaria del sector educación por institución.	52
Figura 26. Ejecución presupuestaria del sector educación por actividad/obra.	53
Figura 27. Ejecución presupuestaria del sector educación por función.....	53
Figura 28. Tendencia central de la ejecución presupuestaria en el sector educación.	54
Figura 29. Umbrales de Clasificación de la Ejecución Presupuestaria.....	54
Figura 30. % de ejecución presupuestaria por Admón. Central - Educación.	55
Figura 31. % de ejecución presupuestaria por Sector Público Descentralizado - Educación.....	56
Figura 32. Ejecución presupuestaria del sector salud por año.	57
Figura 33. Serie de ejecución presupuestaria del sector salud por período.	58
Figura 34. Ejecución presupuestaria del sector salud por objeto de gasto.....	59
Figura 35. Ejecución presupuestaria del sector salud por grupo de gasto.	59
Figura 36. Ejecución presupuestaria del sector salud por institución.	60
Figura 37. Ejecución presupuestaria del sector salud por actividad/obra.....	61
Figura 38. Ejecución presupuestaria del sector salud por función.....	61
Figura 39. % de ejecución presupuestaria por Admón. Central - Salud.	62
Figura 40. % de ejecución presupuestaria por Sector Público Descentralizado - Salud.....	63
Figura 41. Descomposición de la Serie Temporal - Sector Educación.....	65
Figura 42. Configuración del modelo ARIMA.....	66
Figura 43. Estimación de presupuesto vía S-ARIMA del sector educación.....	67
Figura 44. Configuración del Random Forest.....	68

Figura 45. Resultados de la predicción con Random Forest.....	69
Figura 46. Configuración de Random Forest.....	69
Figura 47. Gráfica de Serie Random Forest (Sin 2023).....	70
Figura 48. Resumen de la estacionalidad Sector Salud.	71
Figura 49. Configuración del modelo ARIMA.....	72
Figura 50. Estimación de presupuesto vía S-ARIMA del sector salud.....	73
Figura 51. Configuración del Random Forest.....	74
Figura 52. Resultados de la predicción con Random Forest.....	75
Figura 53. Configuración Random Forest (Sin 2023).	75
Figura 54. Resultados de Predicción con Random Forest (Sin 2023).	76
Figura 55. Bases de Presupuesto Ejecutado.....	81
Figura 56. Knime Workflow.....	82
Figura 57. Anaconda Navigator.....	83
Figura 58. Librerías Importadas.....	84
Figura 59. Función de Evaluación de Modelos.	85
Figura 60. Importar Dataset en un Dataframe.	85
Figura 61. Creación de dataframe filtrado por sector.	85
Figura 62. Transformación múltiple de datos de tiempo.	86
Figura 63. Agrupación de Datos.	86
Figura 64. Gráfica de Serie Temporal.....	87
Figura 65. Prueba de Dickey Fuller Aumentada.....	87
Figura 66. Prueba Dickey Fuller Aumentada (diff - 1).....	88
Figura 67. Gráfica de Autocorrelación (diff -1).....	88
Figura 68. Descomposición de la serie temporal.....	89
Figura 69. División de datos de prueba y entrenamiento.....	89
Figura 70. Generación de parámetros ARIMA.....	90
Figura 71. Implementación de Modelo.....	90
Figura 72. Predicción con Modelo.....	91
Figura 73. Métricas de evaluación del modelo.	91
Figura 74. Generación de variable con variación porcentual entre periodo	92
Figura 75. Gráfica de variación Porcentual de la serie temporal.	92
Figura 76. Características adicionales creadas.....	93
Figura 77. Datos de Entrenamiento y Prueba.	93
Figura 78. Implementación de Modelo.....	94
Figura 79. Gráfica de resultados del modelo.	94
Figura 80. Evaluación del Modelo.....	95
Figura 81. Parámetros para medir riesgos detectados.....	98
Figura 82. Diagrama EDT.....	99
Figura 83. Cronograma de actividades y distribución.	101
Figura 84. Preparación de datos EDA.....	109
Figura 85. EDA Sector Educación.....	110
Figura 86. EDA Sector Salud.....	111

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

La presente investigación está estructurada por varios capítulos con el fin de lograr los objetivos planteados en esta. En primer lugar, se realizó una revisión exhaustiva de la literatura existente, en donde se analizaron los estudios previos y las metodologías utilizadas para la predicción presupuestaria. También se llevó a cabo un análisis detallado de los datos disponibles, utilizando técnicas de análisis exploratorio y predictivo, los cuales proporcionaron las bases para desarrollar modelos predictivos y pronósticos que permitan estimar los presupuestos para los sectores de salud y educación para Honduras.

Una vez obtenidos los resultados de las predicciones, se procedió a la interpretación y análisis de estos, evaluando la calidad de los modelos desarrollados e identificando las variables que más influyeron en la investigación. Lo anterior permitió comprender de mejor manera el conjunto de datos histórico analizado en la investigación. Finalmente, se realizó una serie de conclusiones y recomendaciones acorde a los hallazgos obtenidos y su aplicabilidad en el contexto real.

1.2 ANTECEDENTES DEL PROBLEMA

En 2016 se realizó el estudio llamado “Análisis de la Implementación del Presupuesto por Resultados En Guatemala” cuyo objetivo fue analizar la ejecución del presupuesto en las instituciones públicas de Guatemala en los períodos comprendidos entre 2012 y 2015 concluyendo que la gestión de presupuesto por resultados es una opción viable debido a que a partir de esto es posible priorizar de forma adecuada los recursos financieros que poseen las administraciones públicas. (Sánchez W. , 2016)

Por su parte, en el estudio denominado “Análisis de la Ejecución Presupuestaria del Primer Nivel de Atención en el Ministerio de Salud de El Salvador, 2014 - 2017” tuvo como objetivo determinar el alcance y nivel de desarrollo de los procesos de gestión en la ejecución del presupuesto del primer nivel de atención del Ministerio de Salud de El Salvador en el período comprendido entre 2014 y 2017, en el cual se concluye que uno de los principales factores que intervienen directamente en la ejecución del presupuesto destinado a atención del primer nivel en el sector de salud es insuficiente de acuerdo a las necesidades que este implica. Además, recomiendan que las compras vayan acorde al período correspondiente y que estas no afecten a

períodos posteriores. (Alvarado, 2019)

Además, en el estudio llamado “Beneficios Ofrecidos por la Gestión del Big Data en las Instituciones Gubernamentales en la Era de la Digitalización” que tenía como objetivo conocer las ventajas que ofrece la tecnología de Big Data en la gestión de dichas instituciones, aplicando el análisis de contenido. Dicho estudio concluye que la aplicación de esta tecnología ofrece un apoyo significativo en la toma de decisiones, participación ciudadana, administración eficaz de los recursos y transparencia en todo el proceso de gestión gubernamental. (Borja y otros, 2019)

Por su parte, de acuerdo con el estudio que tiene como título “Ejecución Presupuestal de Gastos y su Incidencia en el Cumplimiento de Metas Institucionales de la Municipalidad Distrital de San Antonio” y cuyo objetivo fue determinar como la óptima ejecución presupuestaria de gastos contribuye al cumplimiento de metas institucionales de la Municipalidad Distrital de San Antonio en Callao, Perú. En dicho estudio se concluye que en la Municipalidad existe deficiencia en gastos para inversión pública, dejando sectores relevantes sin atender y devolver los fondos para ejecuciones de años posteriores. (Pacheco y otros, 2021)

Finalmente, Estefanía Montoya y Darwin Yáñez publicaron en 2022 el artículo que tiene como título “Analítica de datos: Una tendencia para la toma de decisiones empresariales en las organizaciones” en donde describen múltiples elementos y maneras de poner en práctica la analítica en entornos empresariales y la importancia de esta en las decisiones relevantes que determinan el rumbo de estas, concluyendo que una planeación financiera adecuada representa beneficios de gran valor tales como la evaluación y ejecución de presupuestos, obtención de rendimientos y rentabilidad y la identificación de recursos que son de poca visibilidad pero que podrían brindar una alternativa en casos fortuitos en donde sea necesario hacer uso de ellos. (Montoya & Yáñez, 2022)

1.3 DEFINICIÓN DEL PROBLEMA

En esta sección se expone el problema de investigación, que es el punto de partida de un estudio, definiéndose como una pregunta clara y concreta a la que se dará respuesta mediante la recopilación y el análisis de predicciones.

1.3.1 Enunciado del Problema

Se ha realizado un análisis inicial del conjunto de datos de la ejecución presupuestaria en Honduras obtenido del sitio web oficial de la Secretaría de Finanzas (SEFIN) y contando con información correspondiente a los períodos comprendidos entre 2010 y 2023 se logra observar que

los sectores de Salud y Educación no han sido de alta prioridad a lo largo de los años, lo que explica a priori el porqué de la desatención histórica que ha existido en estos.

En el sector salud, se dice que el estancamiento reciente del gasto público en este, según los expertos del Programa de las Naciones Unidas para el Desarrollo (PNUD) pasó de ser un 15% en 2009 y 2010 a 12.5% en el 2020, lo que se convierte en un déficit en el presupuesto. (Cuevas, 2022)

En cuanto a Educación, según El Heraldó, el 90% del presupuesto para dicho sector ha sido destinado para sueldos y salarios administrativos y del magisterio, lo que limita los fondos para las demás iniciativas del sector educativo. Las planificaciones interrumpidas son la matrícula gratis, infraestructura, dotación de insumos de bioseguridad, merienda escolar, material didáctico, etc. Lo anterior evidencia las limitaciones del presupuesto para la educación de Honduras. (El Heraldó, 2023)

1.3.2 Formulación del Problema

Una vez se conoce la situación que presenta la ejecución del presupuesto en el estado de Honduras, surge la siguiente interrogante ¿Es una buena alternativa predecir el presupuesto de los sectores de salud y educación en Honduras a 2024 haciendo uso de algoritmos de Aprendizaje Automático?

1.3.3 Preguntas de Investigación

1. ¿Es factible con una buena aproximación identificar a través de los objetos de gasto con mayor afectación en la ejecución del presupuesto de los sectores de salud y educación el presupuesto del año siguiente?
2. ¿Cuál es el nivel de eficiencia en la ejecución del presupuesto de los sectores de salud y educación?
3. ¿Es eficiente el uso de datos históricos de los objetos de gasto de mayor afectación para generar un modelo eficaz de predicción del presupuesto del año siguiente?

1.4 OBJETIVOS DEL PROYECTO

Los objetivos de la investigación son indispensables para una correcta ejecución de cada una de las etapas de esta, y es que a través de ellos es que se marca el horizonte al cual irá orientada, plasmando límites y el alcance.

1.4.1 Objetivo general

Predecir el presupuesto de los sectores de salud y educación en Honduras para 2024 haciendo uso de los algoritmos de Aprendizaje Automático.

1.4.2 Objetivos específicos

1. Identificar los objetos de gasto con mayor afectación en la ejecución del presupuesto de los sectores de salud y educación.
2. Determinar el nivel de eficiencia en la ejecución del presupuesto de los sectores de salud y educación.
3. Generar una predicción del presupuesto de Salud y Educación para el año 2024.

1.5 JUSTIFICACIÓN

Honduras es un país que se encuentra en vías de desarrollo, lo que significa que existen múltiples aspectos y desafíos a los que se enfrenta en cuanto a gestión pública se refiere. En busca de dicho desarrollo y brindar mejoras a la atención ciudadana, es necesaria la aplicación de estrategias y avances tecnológicos en medida de lo posible. La inminente predicción de un presupuesto óptimo para los sectores de salud y educación que históricamente han sido desatendidos en cierto modo es una solución potencialmente apta para mejorar la gestión de recursos acorde a las necesidades de la población.

Esta investigación busca proporcionar un algoritmo de predicción sólido que permita a las autoridades responsables de la aprobación y ejecución presupuestaria en Honduras planificar de forma más adecuada la asignación de recursos en los sectores de salud y educación al brindar la solución de presupuestos futuros. Dicho lo anterior, esto facilitará la asignación de presupuestos para dichos sectores de forma equitativa, permitiendo una distribución de los fondos del estado con mayor precisión y eficacia, dejando al estado fortalecer y mejorar los servicios e

infraestructura destinada a los rubros antes mencionados, teniendo un impacto positivo dentro del desarrollo socioeconómico de la población hondureña.

Otro punto es la adecuación de los enfoques en la asignación de objetos de gastos que serán identificados con la ejecución del estudio.

Además, en Honduras actualmente no existe registro de estudios de aplicación de técnicas de analítica de datos en este rubro, por lo que la indagación en un tema que es tendencia en la actualidad como lo son los modelos predictivos resulta innovador y útil para cualquier institución de esta índole, por tal razón la viabilidad se detalla a continuación:

- Viabilidad operativa: Esta investigación resulta viable a nivel operativo ya que la disponibilidad de la información será posible extraerla a través de los datos públicos extraídos de <https://www.sefin.gob.hn/ejecucion-presupuestaria/>.
- Viabilidad técnica: A nivel técnico es viable esta investigación debido a que los modelos predictivos son un tema que cada día alcanza mayor auge y, por ende, existen investigaciones, herramientas y avances tecnológicos a nivel mundial que invitan a adentrarse en ese espacio para extraer los mayores beneficios.
- Viabilidad financiera: La realización de la investigación será financiada por los investigadores, para obtener el título de Máster en Analítica de Negocios.

CAPÍTULO II. MARCO TEÓRICO

En el marco teórico de esta investigación se ilustra información importante sobre los factores internos y externos que afectan en la ejecución presupuestaria de los sectores de Salud y Educación en Honduras desde diferentes perspectivas. Posteriormente se ilustran las teorías que sustentan la investigación con respecto a los modelos de aprendizaje automático y aplicaciones de estos en la ejecución presupuestaria. Finalmente, se definirán todos aquellos conceptos de palabras utilizadas con el fin de brindar una mayor comprensión de los temas que aquí se presentan.

2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL

Partiendo desde un análisis descriptivo sobre la gestión del gasto público de los sectores de salud y educación en todos los países, hasta llegar a una perspectiva localmente que demuestre como se ejecuta el presupuesto en Honduras en estos dos sectores antes mencionados.

2.1.1 ANÁLISIS DE MACROENTORNO

A nivel mundial el a grandes rasgos el presupuesto público asignados a los sectores de salud y educación representa un desafío ya que existen países que según sus políticas existentes requieren más apoyo que otros. En principio según (Datos Macro, 2022) para el año 2022 Estados Unidos fue el país con mayor gasto público asignado al sector salud. Asimismo, se muestra que en el sector educación quien ha tomado las primeras posiciones en cuanto al gasto publico per cápita ha sido Noruega.

A continuación, se ilustra una gráfica donde se puede ver cómo están distribuidos los demás países con relación al gasto público en salud, y también como Estados Unidos se antepone frente a todos los demás países.

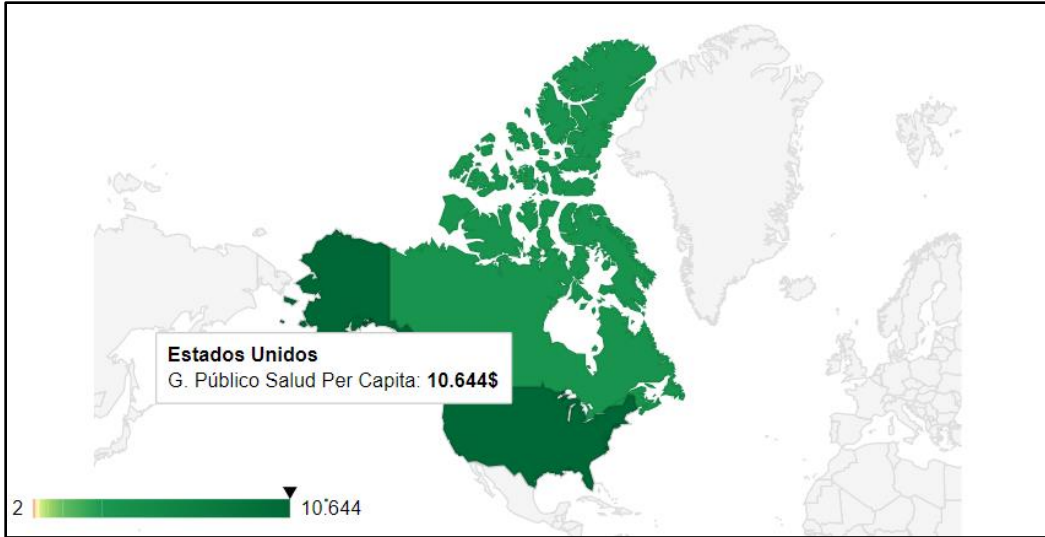


Figura 1. Gasto Público en Salud 2022 de los países de norte américa.
 Fuente: (Datos Macro, 2022)

En cuanto al gasto público en el sector educación a nivel de los países del mundo se identifica un comportamiento más bajo en relación con el sector salud el cual como se denota por el gasto per cápita de ambos países, en este caso del sector educación como se puede apreciar Noruega es el país con más gasto per cápita. Siendo este debido a que según (Datos Macro, 2022) el 65% de su población está conformada por edades de 15 a 64 años, lo cual genera una alta demanda por estos grupos en cuanto a la disposición de centros educativos.

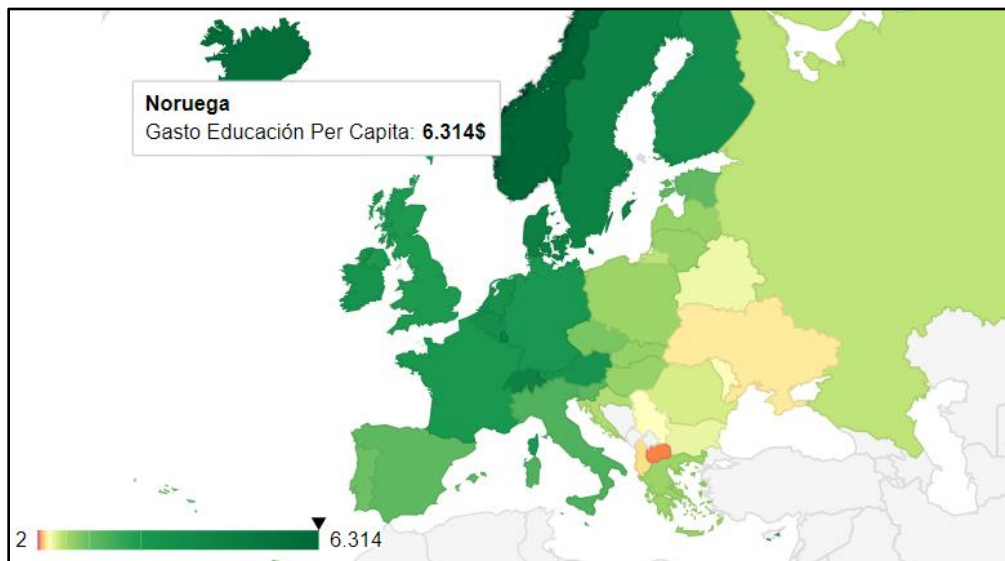


Figura 2. Gasto público de Educación 2022 en los países europeos.
 Fuente: (Datos Macro, 2022)

Ampliando el análisis en el sector educativo cabe mencionar que según (UNESCO, 2020) en el mundo el 86% de los adultos están alfabetizados como también un 92% de los jóvenes a nivel mundial. Sin embargo, existen muchos programas para calificar el nivel de educación de los países el cual es un referente mundial. Para el 2018 según (PISA (programme for international student assessment), 2018) los 5 mejores países catalogados con mejor nivel educativo representaban al 18% de la población mundial, como se puede ilustrar en la Figura 3. TOP 5 Países con mejor educación se muestran los países que conforman este bloque tan importante a nivel de la educación.

Puesto	Lectura	Matemáticas	Ciencias
1.	China	China	China
2.	Singapur	Singapur	Singapur
3.	Macao	Macao	Macao
4.	Hong Kong	Hong Kong	Estonia
5.	Estonia	Taiwán	Japón

Figura 3. TOP 5 Países con mejor educación

Fuente: (PISA (programme for international student assessment), 2018)

Realizando una comparación del gasto per cápita en Honduras con los 5 países que anteriormente se ilustraron, se puede ver que Hong Kong con 7 Millones de habitantes se convierte en el que más gasta en cuanto a la educación, siendo esto así, es notable la diferencia que tiene Honduras en relación con los otros países que lideran en educación. Y es aquí donde se demuestra que quizás el dimensionamiento al presupuesto en el sector de educación no es lo suficiente como para poder cubrir las necesidades dentro del país y lograr una mejora en dicho sector.

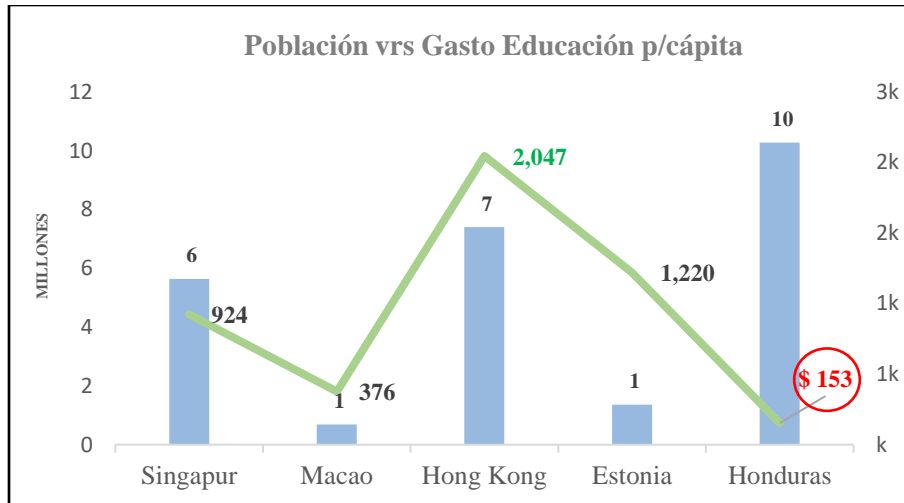


Figura 4. Comparación de población vs. gasto per cápita en la educación
Fuente: Elaboración Propia

En el sector salud según (Saint Leo University, 2023) de acuerdo con el índice de atención médica y eficiencia en los sistemas de salud, se generó un ranking de los países/ciudades con mejores sistemas de salud a nivel mundial, para los cuales se detalla a continuación:

1. Corea del Sur
2. Taiwán
3. Dinamarca
4. Austria
5. Japón

Adicional a ello se muestra la comparación del gasto público en salud de Honduras respecto a los Top países que antes se mencionan donde se puede notar que Honduras solo representa un 2% del gasto más alto per cápita que lo tiene Dinamarca con 5k USD por habitante.

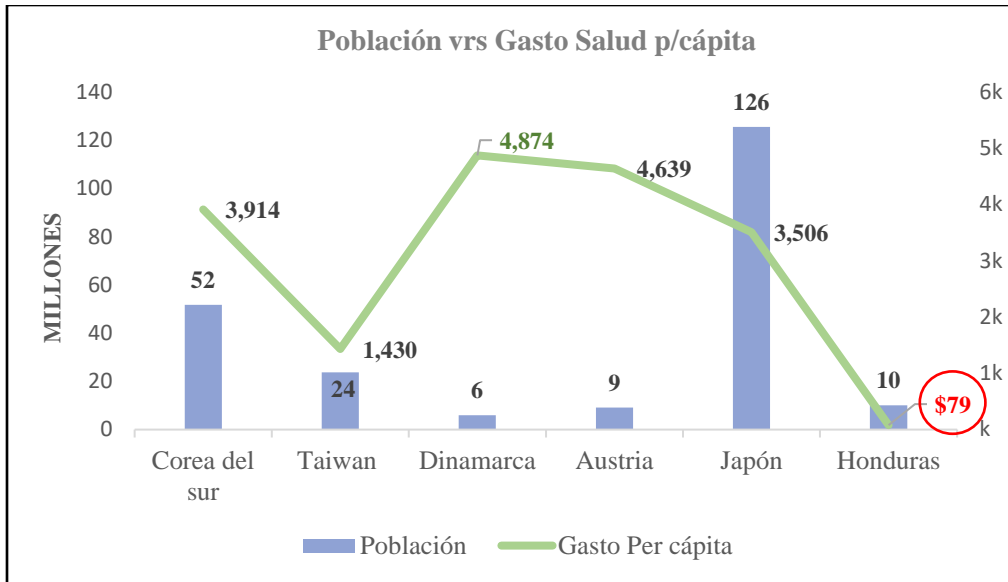


Figura 5. Comparativa de Población vs Gasto Per cápita de salud

Fuente: Elaboración propia

En esencia para que Honduras esté dentro del promedio de gasto per cápita debe tener un incremento per cápita de un 98% en su gasto y así de esta manera logrará equipararse con países como Austria o Dinamarca. No obstante, esto debe venir complementado con iniciativas bien definidas y estrategias que estén enfocadas a lograr una mejor salud en el país.

2.1.2 ANÁLISIS DE MICROENTORNO

El gasto público en salud a nivel latinoamericano es muy variable, sin embargo, en unos es más aprovechado en lugar de otros. Como ser el caso de Honduras donde se estima que solamente existe 0.31 médicos por cada 1,000 habitantes, muy diferente a con 8.19 médicos por cada 1000 habitantes.

Estas cifras son alarmantes con el simple hecho de pensar que el país está siendo afectado en uno de los pilares más importantes para la sociedad. Y por ello se vuelve de gran interés el estudio y análisis del comportamiento en la ejecución del presupuesto, esto debido a que en los medios de comunicación se afirma que la mayor parte del presupuesto de salud es destinado a pagos de salarios y transferencias a dependencias que no prestan servicios de salud. (Expediente Publico, 2020)

A continuación, se muestra una gráfica donde se ilustran los países de América latina que más invierten en salud.

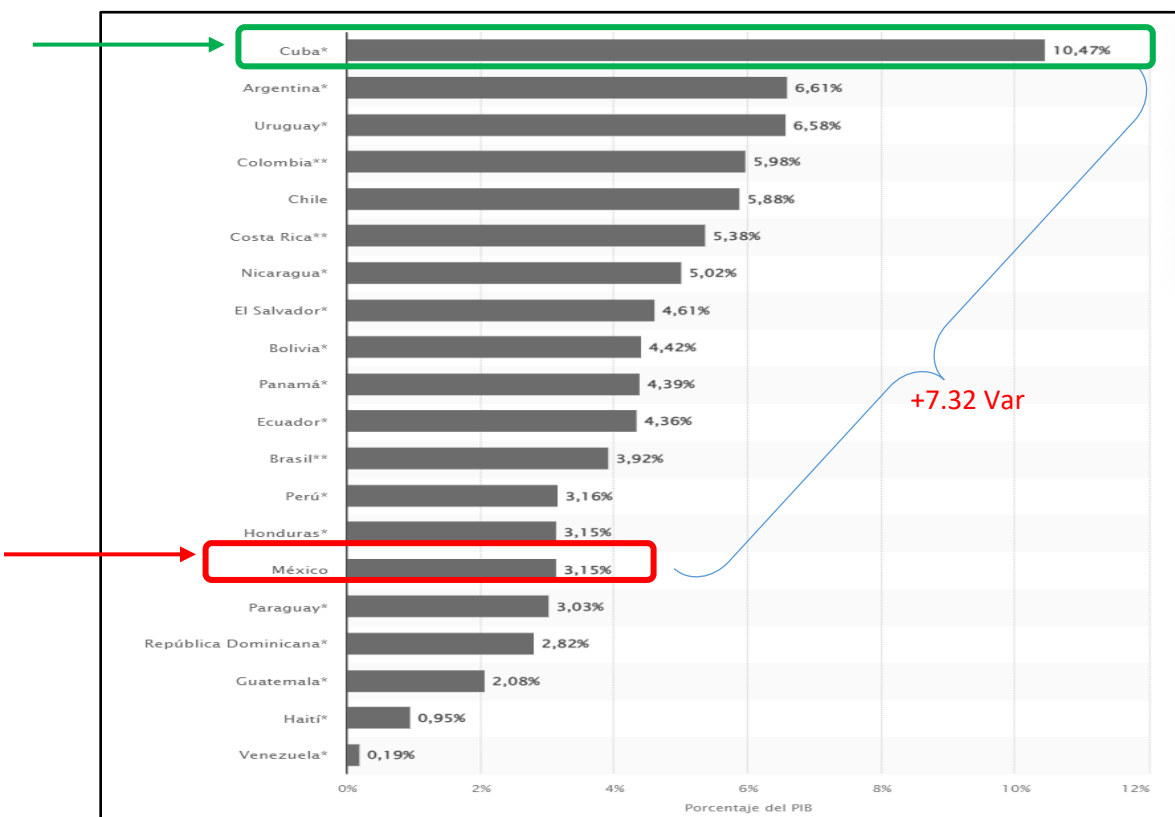


Figura 6. Ranking de países con mayor presupuesto en salud en América latina

Fuente: (statista, 2020)

En Cuanto a la educación es interesante ver como en América latina se desarrolla ya que existe un gran desafío para todos los países en relación con la calidad, equidad y la gestión eficiente de los recursos asignados. Es evidente que cada país enfrente sus propios retos, sin embargo, la educación debería ser una prioridad fundamental ya que a través del fortalecimiento de esta se logra el desarrollo sostenible y genera competitividad a nivel general.

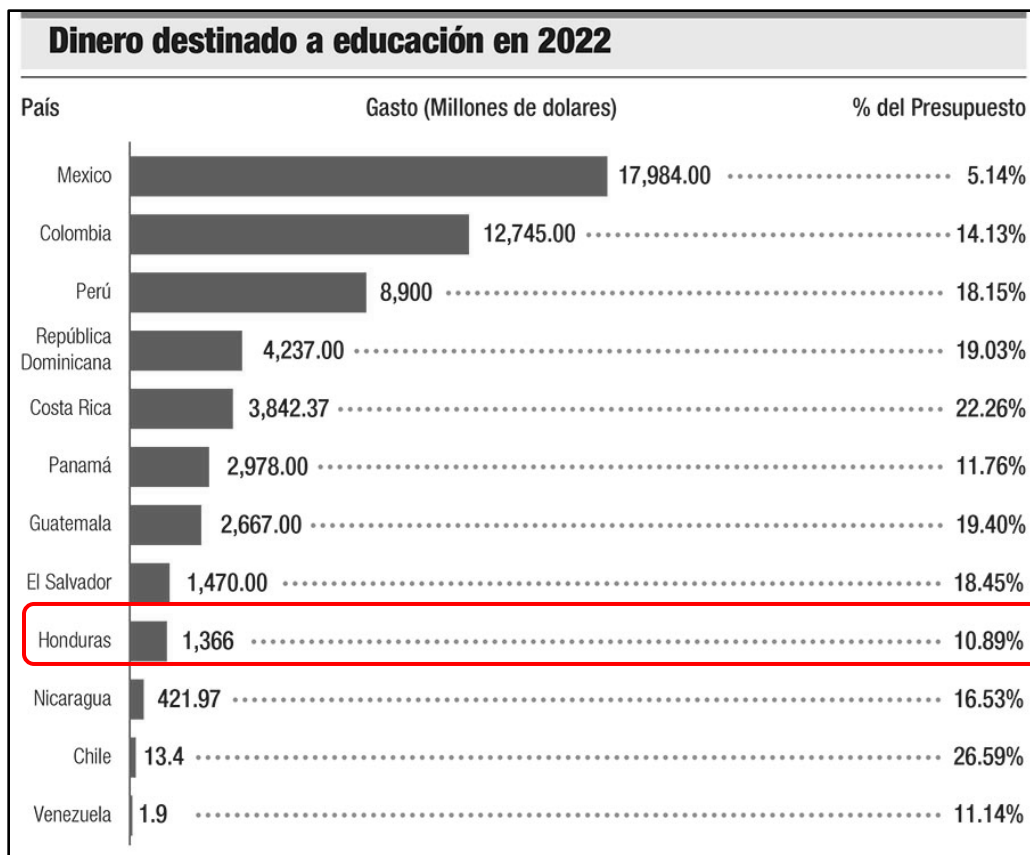


Figura 7. Gasto público para educación y porcentaje sobre el PIB
 Fuente: (Cognición, 2022).

Según se muestra en la gráfica Chile es el país que más porcentaje del presupuesto general destina a educación con un 26.59% seguido de Costa Rica con un 22.26%. Esto claramente denota los intereses en los que se focalizan las autoridades de cada uno de estos países. Costa Rica ha sido un referente en temas de tecnología e innovación en los últimos años a nivel centroamericano, lo que convierte en opciones para sus habitantes y genera oportunidades a nivel internacional, dando como resultado una mejor formación profesional en las personas que deciden ser parte de estas estrategias gubernamentales.

Ahora bien, el gasto público en educación también se relaciona directamente con el PIB que tiene cada país. Bajo este esquema los números cambian siendo Costa Rica el #1 que más inversión tiene en la educación con relación a su PIB. Honduras, queda en la posición #3 con un 5% de inversión en relación con el PIB.

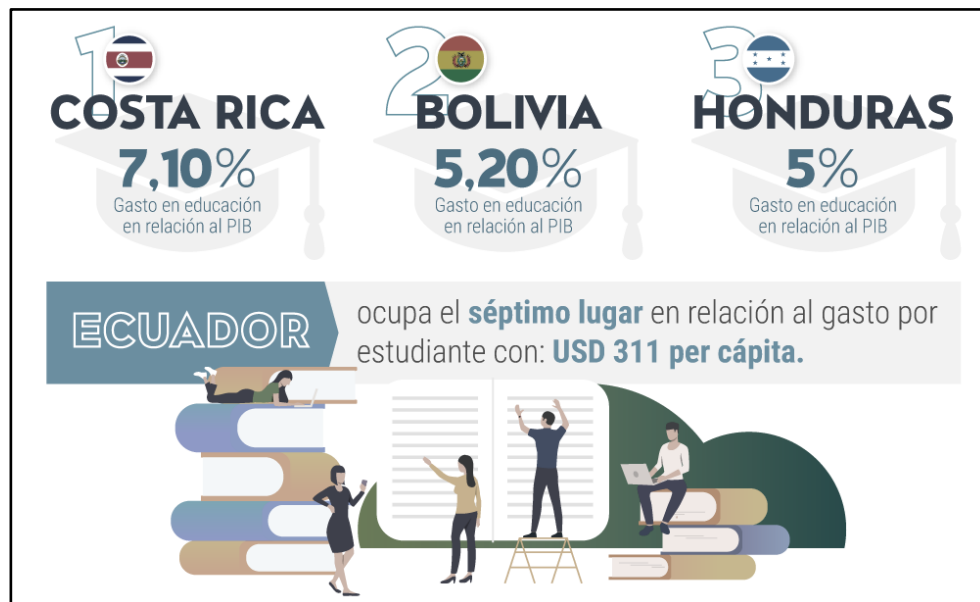


Figura 8. Países con más inversión en educación en América Latina en relación con el PIB
Fuente: (EKOS, 2019).

2.1.3 ANÁLISIS INTERNO

Honduras cuenta con un total de 49,000 centros educativos distribuidos a lo largo de todos los departamentos en los cuales Francisco Morazán, Cortés, Olancho, El Paraíso, Yoro, Santa Bárbara y Comayagua conforman casi el 60% de todo el universo de centros en el país (Figura 9. Distribución de Centros Educativos por Departamento).

Cantidad de Centros Educativos por Departamento

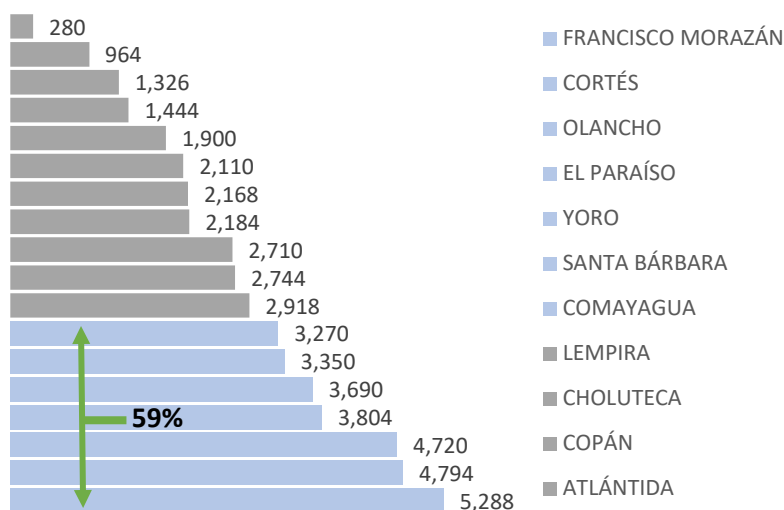


Figura 9. Distribución de Centros Educativos por Departamento.

Fuente: Elaboración propia.

Es interesante ver como la oferta de centros educativos que existe en Honduras, y aun así permanecen problemas de analfabetismo y índices altos de deserción escolar, causando cada día impacto en el desarrollo de la sociedad actual y las nuevas generaciones. Y es que existe un fenómeno que actualmente tiene preocupado a todos los economistas según (El Criterio HN, 2022) en el presupuesto general de la república para el 2023 se ha presupuestado un incremento de un 7% para el sector salud y un 8% para la educación, contrario a los sectores de seguridad y defensa con un 17% y 14% de incremento respectivamente. Es claro el panorama con estos números, el gobierno de manera directa no se está preocupando por los sectores importantes para el desarrollo de una sociedad desarrollada. Está enfocando sus esfuerzos a lograr e imponer sus ideales políticos entre las diferentes instituciones al priorizar sectores como defensa y seguridad, lo cual ya se sabe que el fin es blindarse a todo nivel, con tal de lograr su objetivo.

A continuación, se muestra una gráfica donde se puede identificar que los intereses por parte de las autoridades han estado destinados a temas de seguridad más que a temas de salud, lo cual es un causante también del declive que ha traído este sector por varios años. Llegando a situaciones tan complejas que recientemente se ha anunciado por los medios la liquidación de uno de los centros hospitalarios más importantes en Tegucigalpa (HEU), con el objetivo de poder

integrarlo a programas de “Equidad” misteriosos que solamente el gobierno actual conoce.



Figura 10. Relación de Presupuesto de Seguridad vs Salud en Honduras

Fuente: (Expediente Publico, 2020)

Otra variable importante dentro de la gestión de los presupuestos tiene que ver con la porción destinada a infraestructura de educación y salud, la cual forma parte de ese pequeño fragmento del presupuesto asignado a estos sectores. Lo que al final se reduce en aproximadamente un 2% del presupuesto asignada cada sector. Es preocupante el estado de los centros de salud y educación en Honduras ya que, aunque el presupuesto es reducido no se logra ni siquiera completar todas aquellas obras que se prometen mejorar en los gobiernos, creando falsas expectativas en la población.

Según (ASJ (Asociación para una Sociedad Justa), 2022) el ritmo que Honduras trae en el progreso de la educación estaría alcanzando un 100% de cobertura dentro de los próximos 50 años. Ver esta cifra es un dato preocupante ya que esto indica que los intereses de los gobiernos en general no se han centrado o enfocado en mejorar este indicador de cobertura, siendo este un pilar importante para el desarrollo del país a nivel competitivo.



Figura 11. Eficiencia de la Educación en Honduras

Fuente: (ASJ (Asociación para una Sociedad Justa), 2022)

2.2 CONCEPTUALIZACIÓN

En el siguiente apartado se muestran los conceptos básicos y que tienen relación directamente con la analítica de negocios los cuales se detallan a continuación:

2.2.1 Analítica de Datos

La analítica de datos es un término que se utiliza para referirse a un conjunto de aplicaciones que van desde la inteligencia empresarial básica, el procesamiento analítico en línea, la creación de informes para toma de decisiones y la analítica de alto nivel. Tiene un enfoque parecido al de la analítica de negocio, pero su alcance es más amplio, llegando a una gran variedad de campos y disciplinas. (Stedman, 2021)

2.2.2 Aprendizaje Automático

Derivado de la Inteligencia Artificial, brinda a las computadoras la capacidad de encontrar patrones en grandes conjuntos de datos y a partir de ello elaborar predicciones, permitiéndoles efectuar actividades de forma no supervisada. El término no es nuevo, ya que en 1959 se utilizó por primera vez. Sin embargo, en los últimos años ha alcanzado su pico debido a la importancia de los datos en el mundo actual. (Iberdrola, 2023)

2.2.3 Árbol de Decisión

Es un algoritmo de predicción que se representa como un árbol a nivel gráfico y en el que cada uno de los patrones que se detectan corresponde a una rama, siendo las predicciones que arroja el modelo las que se representan en las hojas. (González, 2021)

2.2.4 Bosques Aleatorios

El algoritmo de machine learning denominado Random Forest fue desarrollado por Leo Breiman y Adele Cutler, y se utiliza comúnmente en problemas de clasificación y regresión. Una de las características principales de este algoritmo es que está compuesto por un conjunto de árboles de decisión, logrando crear un bosque no correlacionado de estos. Una de las principales diferencias entre Random Forest y Árboles de Decisión es que este selecciona un subconjunto de características en la clasificación, mientras que los árboles las consideran en su totalidad. (IBM, 2023)

2.2.5 Clusterización

El agrupamiento, también conocido como Clustering, es una técnica importante en la analítica de datos que consiste en dividir por grupos de acuerdo con similitudes y estableciendo diferencias significativas con los miembros de los otros grupos. El objetivo de esta técnica es el realizar nuevas clasificaciones a partir de los grupos formados. (Mamani y otros, 2017)

2.2.6 Correlación

La correlación es una medida que indica el nivel de relación lineal entre dos variables. Esta medida estadística se utiliza a menudo para describir relaciones simples entre dos variables, sin implicar causalidad. El coeficiente de correlación, r , se utiliza para cuantificar la intensidad de esta relación. Es importante destacar que las correlaciones no permiten determinar la presencia o el efecto de otras variables diferentes a las dos que se están analizando. (JPM Statistical Discovery, 2023)

2.2.7 Ejecución Presupuestaria

Corresponde a la ejecución de los ingresos y gastos presupuestados por parte de la entidad ejecutora que tiene la obligación de efectuarlo. (INEC, 2023)

En cuanto a la aplicación de este “campo” en el modelo de datos resulta sumamente importante, ya que es a partir de este que se establece la variable objetivo en la predicción basada

en los modelos y es la ejecución presupuestaria el principal insumo para obtener resultados coherentes, ya que es la que dicta el patrón de cómo el presupuesto que se aprueba a inicio de cada año se va ejecutando con el pasar de los meses.

2.2.8 Fuentes de Datos

Representan el insumo principal para cualquier algoritmo de predicción. Está conformado por los datos históricos que se utilizarán para el entrenamiento de los modelos y así garantizar que la detección de patrones sea efectiva. (González, 2021)

2.2.9 K-Means

Es un método de Clustering que se utiliza para agrupar un conjunto de datos según patrones de afinidad. El número de grupos que se obtienen puede ser configurado según las necesidades de análisis de la información. (Mamani y otros, 2017)

2.2.10 Matriz de Confusión

La matriz de confusión es un elemento útil para valorar qué tan bueno resulta ser un modelo de clasificación de aprendizaje automático. Es representada en una tabla de dos dimensiones que muestra los resultados de la clasificación en forma de verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN), permitiendo identificar distintos tipos de error. (Recuero de los Santos, 2021)

2.2.11 Modelo No Supervisado

El modelo no supervisado se caracteriza por utilizar conjuntos de datos que no están etiquetados, explorándolos con la finalidad de encontrar estructuras y formas lógicas de ordenarlos, a través de características similares. Son usados en proyectos de segmentaciones muy bien definidas, como campañas de mercadeo. (González, 2021)

2.2.12 Modelo Predictivo

Luego del entrenamiento del algoritmo, es necesario se cree un modelo que se encargará de efectuar las predicciones. Este recibe una serie de datos nuevos para realizar las clasificaciones de acuerdo con los patrones detectados en la etapa de entrenamiento. Por ejemplo, si se usa un modelo entrenado con datos históricos de clientes para evaluar el riesgo de cancelación de una tarjeta de crédito, el modelo categorizará a los nuevos clientes en función de su comportamiento y preverá el riesgo de cancelación. (González, 2021)

2.2.13 Modelo Supervisado

Es ampliamente utilizado para hacer predicciones futuras basadas en comportamientos o características presentes en los datos históricos. Este método permite buscar patrones en los datos relacionando todas las variables con un campo especial llamado "campo objetivo". Por ejemplo, en la clasificación de correos electrónicos como "spam" o "legítimo", se analizan las características y patrones presentes en los correos ya etiquetados con ambas etiquetas. Se pueden determinar patrones, como que los correos de spam provienen de ciertas direcciones IP, tienen ciertas relaciones texto/imágenes y contienen ciertas palabras, entre otras características. Una vez que se han identificado todos los patrones relevantes durante la fase de aprendizaje, los correos electrónicos nuevos se comparan con los patrones y se clasifican en función de sus características como "spam" o "legítimos". (González, 2021)

2.2.14 Objeto de Gasto

Corresponde a una clasificación sistemática y homogénea de los diferentes bienes, características o transferencias que desarrolla durante su gestión la administración pública. (SEFIN, 2013)

En el conjunto de datos obtenido del sitio oficial de SEFIN existe una serie de objetos de gasto correspondientes a los sectores de salud y educación, destacando considerablemente el concepto de sueldos y salarios y los complementos como décimo tercer y cuarto mes.

2.2.15 Presupuesto

Este término hace referencia al monto de dinero que es necesario para afrontar una serie de gastos originados de uno o varios proyectos. En otras palabras, corresponde a montos anticipados resultantes de la estimación para la realización de los objetivos propuestos generalmente en el inicio de períodos. (Sánchez, 2020)

En SEFIN, la ejecución presupuestaria se realiza tomando en cuenta diferentes campos de la siguiente forma: el "monto aprobado" corresponde a los valores generalmente obtenidos de períodos anteriores y el cual es presentado ante las autoridades competentes para su aprobación o modificación respectivas, guardando el resultado del proceso en "aprobado ministro".

2.2.16 Presupuesto vigente

En el sitio oficial de la Secretaría de Finanzas (SEFIN) se define el presupuesto vigente de la siguiente forma:

Comprende las modificaciones presupuestarias (ampliación, disminución y traslados) de la estimación de los recursos y de los créditos presupuestarios autorizados por el Poder Legislativo en el Presupuesto General de Ingresos y Egresos de la República, una vez iniciada la ejecución del mismo para un período fiscal. (SEFIN, 2023)

A nivel del conjunto de datos analizado, el campo “vigente” sufre actualizaciones generalmente durante todos los meses de cada período, debido que cada cambio o acción que se lleve a cabo alterará el valor de este campo.

2.2.17 Regresión Lineal

La regresión lineal es un método analítico que permite predecir el valor de una variable desconocida a partir de otra variable relacionada que se conoce. Este proceso se lleva a cabo mediante la construcción de una ecuación lineal que modele la variable dependiente en función de la variable independiente. Por ejemplo, si se dispone de datos sobre los gastos e ingresos de un individuo en el último año, la técnica de regresión lineal puede analizarlos para determinar que los gastos son la mitad de los ingresos. A partir de esta relación, se puede utilizar un ingreso futuro conocido para estimar un gasto futuro desconocido dividiéndolo por dos. (Amazon, 2023)

2.2.18 Support Vector Machine

El SVM es un algoritmo de aprendizaje supervisado ampliamente utilizado en problemas de clasificación y regresión. Su objetivo principal es encontrar un hiperplano que pueda separar de la mejor manera posible dos clases diferentes de puntos de datos. Para lograr esto, el SVM busca el hiperplano con el margen más amplio entre las dos clases, que se representa mediante los signos positivos y negativos de la Figura 5. Este margen se define como la mayor distancia entre el hiperplano y la región paralela que no tiene puntos de datos. Sin embargo, este algoritmo sólo funciona en problemas que permiten la separación lineal, por lo que, en la mayoría de los casos prácticos, se maximiza el margen flexible permitiendo un número limitado de clasificaciones erróneas. (MathWorks, 2023)

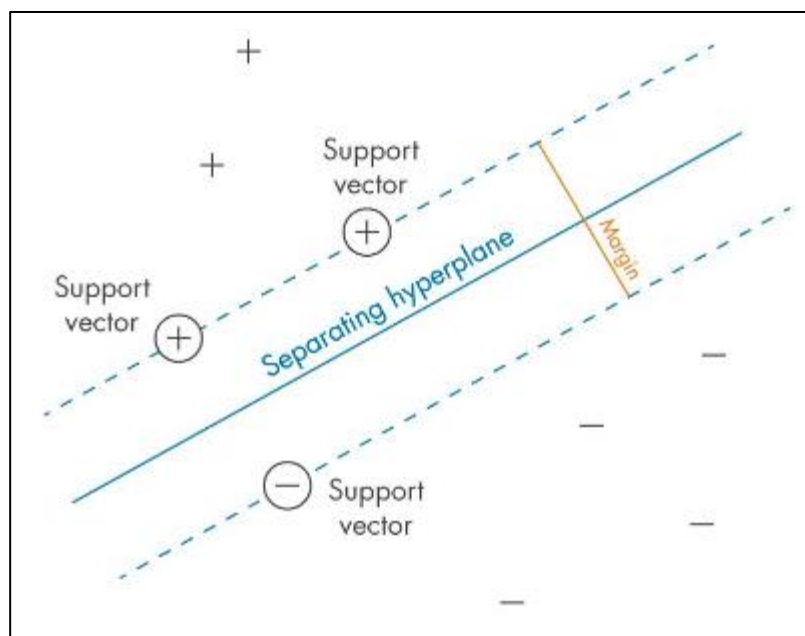


Figura 12. Definición del "Margen" entre clases: El criterio que los SVM intentan optimizar

Fuente: recuperado de <https://es.mathworks.com/discovery/support-vector-machine.html>

2.2.19 Visualización de Datos

La visualización de datos es un procedimiento mediante el cual se utilizan imágenes como gráficos o mapas para mostrar datos. Este método permite simplificar los datos complejos, de gran volumen o numéricos en una forma visual más fácil de comprender. Los programas y herramientas de visualización automatizan y mejoran el proceso de comunicación visual para lograr una mayor precisión y detalle. La visualización de datos se utiliza para extraer información útil de los datos sin procesar. (Amazon, 2023)

2.3 TEORÍAS DE SUSTENTO

En el siguiente apartado se describen las teorías de sustento relacionadas a la investigación, ya que estas ofrecen argumentos y fundamentos para demostrar por qué el reto de la investigación es importante y cómo se relaciona con los conocimientos existentes en la materia.

2.3.1 BASES TEÓRICAS

2.3.1.1 Economía Keynesiana

Esta teoría nace con el objetivo de dar respuesta a la Gran Depresión de los años 30, en la que se vivió un derrumbe significativo a nivel mundial. Uno de los puntos más relevantes de esta

teoría es que es que el acumulado de los gastos del gobierno, las empresas y los hogares representan el motor en cualquier economía. (Jahan y otros, 2014)

Además, dicha teoría se basó en el término de política fiscal que consistía en controlar la economía de naciones en tiempo de recesiones, haciendo uso de la ejecución del gasto presupuestario. Los keynesianos se enfocaban en realizar la ejecución presupuestaria en grandes proyectos de infraestructura, con el fin de generar desarrollo a través de empleos y estabilización de salarios. (Jahan y otros, 2014)

2.3.1.2 Teoría de la Probabilidad

La visión de Keynes con relación a la probabilidad se basa en que esta es completamente distinta a la frecuencia y es que, resulta como un complemento de la lógica deductiva. Define a la probabilidad como una fuerte relación entre premisas y conclusiones determinadas de forma racional y cuyos límites se dictaminan a partir de la imposibilidad o certeza. (Keynes, 1921)

Un punto interesante es que Keynes no le daba a la probabilidad un enfoque matemático, sino que definía los grados de creencia enfocándose en proposiciones y no en eventos, a través de afirmaciones que pueden ser verdaderas o falsas y que añade directamente el concepto de creencia racional que consiste básicamente en definir a la probabilidad como algo objetivo. (Jacobo, 2022)

Y en cuanto a conocer si una proposición es probable o no, Keynes (1921) menciona lo siguiente:

Una proposición no es probable porque pensemos que lo es. Una vez están dados los hechos que determinan nuestro conocimiento, lo que es probable o improbable en estas circunstancias ha sido fijado objetivamente, y es independiente de nuestra opinión. La teoría de la probabilidad es entonces lógica, porque se refiere al grado de creencia que es *racional* mantener en unas condiciones dadas, y no meramente a las creencias actuales de individuos particulares, que pueden ser o no racionales. (p. 4)

Una vez conocidos los conceptos de la teoría se puede ejemplificar suponiendo que “p” es una proposición y “h” es una hipótesis. También debe existir una proposición denominada secundaria “q” que surge a través de la relación que existe entre “p” y “h”. Para que exista racionalidad debe haber grados de certeza entre “h” y “p” para que estas sean racionales (Jacobo, 2022).

2.3.1.3 Teoría de la Estadística

Adán Barreto (2012) define a la estadística como “la ciencia cuyo objetivo es reunir información cuantitativa concerniente a individuos, grupos, series de hechos, etc., para deducir de ello, gracias al análisis de estos datos, significados precisos o previsiones para el futuro.” (Barreto, 2012)

La teoría estadística se divide en descriptiva e inferencial, siendo la primera la recopilación de un conjunto de datos para obtener una descripción específica de cada una de sus características, mientras que la inferencial corresponde a la toma de decisiones a través de los resultados aplicados en una población o muestra. (Barreto, 2012)

Para Anders Hald (1998) la definición del objeto de los métodos estadísticos se representa de la siguiente forma:

La reducción de los datos, lo cual se logra al considerar los datos disponibles como una muestra aleatoria de una población hipotética infinita, cuya distribución con respecto a las características bajo discusión es especificada por relativamente pocos parámetros. Establece tres tipos de problemas: a) especificación sobre la escogencia de la distribución de probabilidad para la población; b) estimación que describe como calcular los “estadísticos” de la muestra aleatoria para representar los parámetros de la población teórica y c) distribución sobre distribuciones muestrales de los “estadísticos”. (p. 713)

En la Figura 1 se observa el futuro de la estadística de acuerdo con el artículo publicado por Bradley Efron en 1998 y en el cual se observan los temas actuales relacionados con las investigaciones estadísticas.

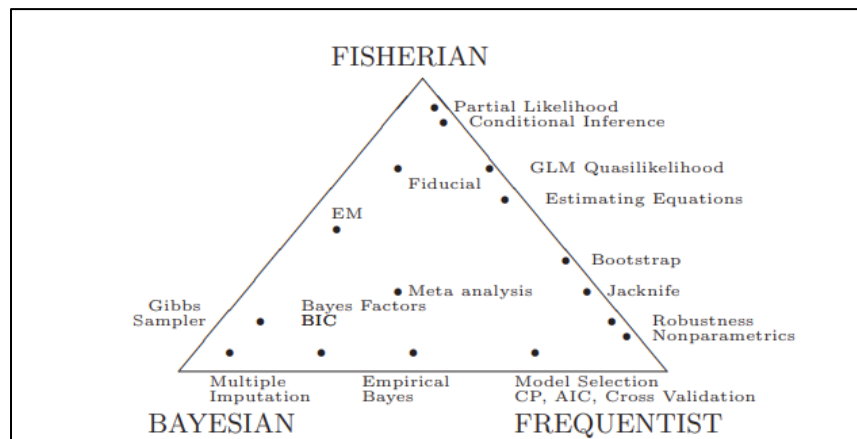


Figura 13. Temas de Investigación Estadística.

Fuente: recuperado de (Efron, 1998).

2.3.1.4 Teoría de Sistemas

Esta teoría trata acerca de que todo fenómeno o situación puede ser analizado como un sistema relacionado con la realidad a la que este pertenece. Se conoce que dentro de un sistema hay una serie de elementos interrelacionados y que cada elemento tiene un papel fundamental en el proceso, siendo indispensable para la correcta ejecución. Las interacciones que se dan entre las partes para formar un todo son la base fundamental de esta teoría, concluyendo que no es posible que exista un sistema si hay vacíos o errores en los componentes. Los elementos son clasificados como entradas, salidas y retroalimentación. (Luhmann, 1997)

Otro elemento importante en la teoría es el factor tiempo, y es que, un sistema no puede ser estático pues se caracterizan principalmente por ser dinámicos, ofreciendo una salida. Dicho lo anterior, los sistemas deben tener la capacidad de adaptabilidad de acuerdo con las circunstancias en las que se encuentre, apegándose a normativas vigentes. Sin embargo, se debe tomar en cuenta que, aunque los sistemas están sujetos a cambios, deben apegarse a su comportamiento original. (Luhmann, 1997)

En la Figura 2 se muestra a la teoría de sistemas de forma gráfica y en la cual se puede observar que el flujo de un sistema inicia con una entrada, se cumple un proceso y finaliza con una salida.

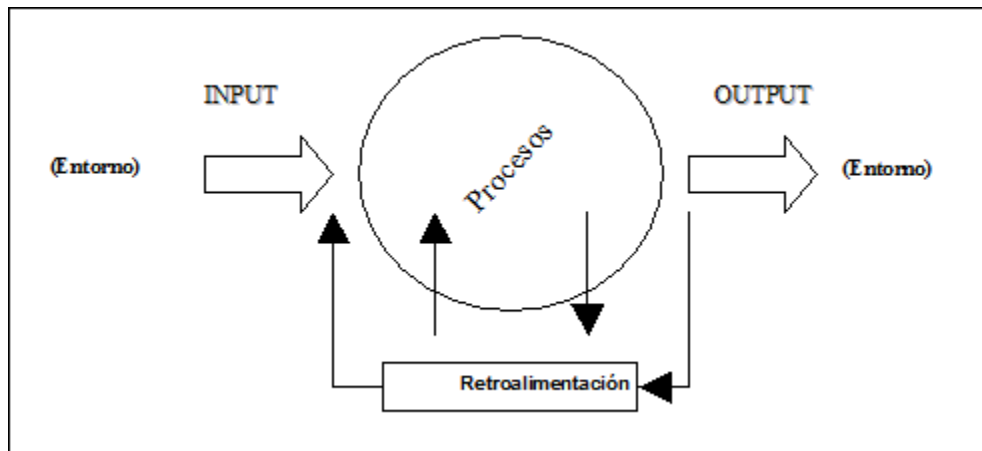


Figura 14. Teoría de Sistemas.

Fuente: recuperado de <https://sites.google.com/site/tecnologiатеатral/que-puedes-encontrar/teoria-de-sistemas-y-tecnologia-educativa-1>

2.3.1.5 Teoría de Optimización

La teoría de la optimización matemática está compuesta por un conjunto de métodos numéricos cuyo fin es encontrar el candidato ideal entre un conjunto de alternativas, de manera que no sea necesario evaluar cada una de ellas. Esta teoría resulta básica en las áreas de ingeniería debido a que ahí se diseñan sistemas y procesos buscando siempre la eficiencia. (Espinosa & Vásquez, 2016)

En esta teoría, aparece el término Problema de Optimización, el cual es definido por Gilberto Espinoza y Alejandro Vásquez (2016) de la siguiente forma:

Un problema de optimización es, en general, un problema de decisión. A partir del valor de una función, que se llama la función objetivo y que se diseña para cuantificar el rendimiento y medir la calidad de la decisión, se obtendrán valores para cierto número de variables, relacionadas entre sí mediante expresiones matemáticas, de manera que minimicen o maximicen esa función objetivo y, por lo general, teniendo en cuenta una serie de restricciones que limitan la elección de esos valores. (p. 3)

2.3.1.6 Teoría de Aprendizaje Automático

Ramos (2019) define al aprendizaje automático en su tesis que tiene como título “Aprendizaje automático para flujos de datos” de la siguiente forma:

El aprendizaje automático se centra en aplicar sistemáticamente algoritmos para sintetizar de forma automática las relaciones subyacentes en un conjunto de datos proporcionados en forma de ejemplos a través de una fase de entrenamiento, de tal forma que en el futuro se utilice esta información para la ejecución de predicciones de eventos desconocidos y una mejor toma de decisiones. Los campos de aplicación que existen de esta rama de la inteligencia artificial son muy variados. Algunos de ellos son la predicción bursátil, predicción meteorológica, detección de correos spam, construcción de sistemas de recomendación y detección de fraude en el uso de tarjetas de crédito. (p. 7)

Otro punto interesante del aprendizaje automático es que este se divide en dos tipos: supervisado y no supervisado. El supervisado se compone por variables predictoras que tienen como objetivo otorgar una variable de salida que sea de utilidad para la toma de decisiones, a través de una fase de entrenamiento que permite predecir a partir de dichas variables. Por su parte, el aprendizaje automático no supervisado es capaz de utilizar algoritmos predictivos sin etiquetar datos de entrenamiento, llegando a predecir de forma automática solo con valores en las variables predictoras. (Ramos, 2019)

En la Figura 3 se muestra un diagrama del proceso de aprendizaje automático el cual da inicio con el pre procesamiento en donde se obtienen los datos para entrenamiento y prueba, luego la etapa de aprendizaje del algoritmo para que sea evaluado y finalmente obtener las predicciones esperadas.

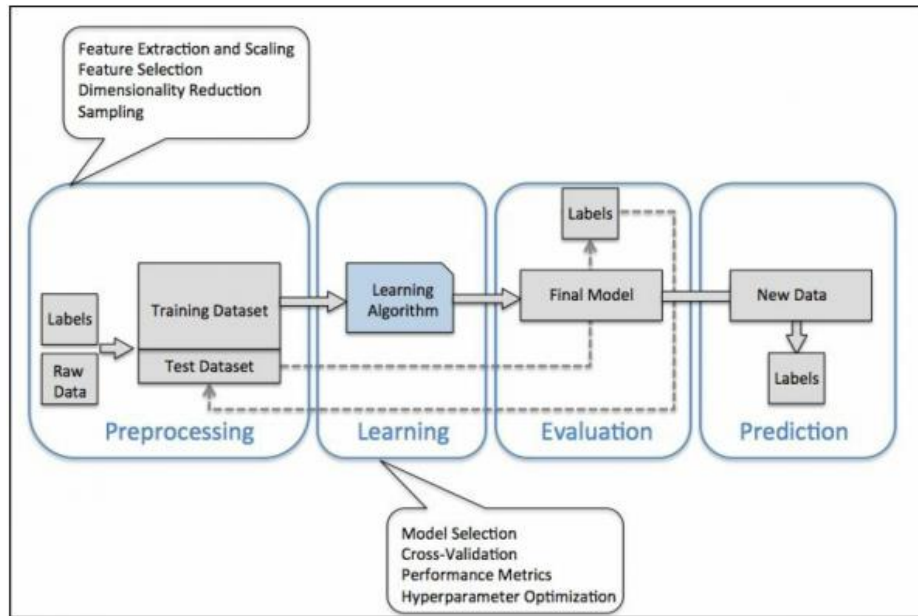


Figura 15. Proceso de Aprendizaje Automático.

Fuente: recuperado de <https://deoxyt2.livejournal.com/147612.html>

2.3.1.7 Teoría de Minería de Datos

Para definir a la minería de datos, se consulta a autores como Piatetsky y Frawley (1991) que indican lo siguiente:

La Minería de Datos se define como el proceso completo de extracción de información, que se encarga además de la preparación de los datos y de la interpretación de los resultados obtenidos, a través de grandes cantidades de datos, posibilitando de esta manera el encuentro de relaciones o patrones entre los datos procesados. (Piatetsky & Frawley, 1991)

Uno de los indicadores más importantes en la minería de datos es la bondad del resultado, y es que para aplicarlo se utilizan medidas como la precisión, la ratio de error, la varianza y matriz de confusión. Otro indicador es la relevancia del resultado, que incluye como métrica al coeficiente de cobertura, el de apoyo y el de significación. Estos coeficientes son de gran importancia en los resultados de la aplicación de minería ya que son capaces de medir la situación actual y la frecuencia de los resultados. (Padmanabhan, 1999)

Por su parte, los indicadores de novedad son utilizados cuando se tiene información en grandes cantidades y se complica la creación de reglas para la minería de datos con el objetivo de evitar redundancias. Finalmente, los indicadores de aplicabilidad de resultados son aquellos que miden la rapidez en la disponibilidad de la información que arrojan los modelos de minería de datos (Marcano & Talavera, 2007).

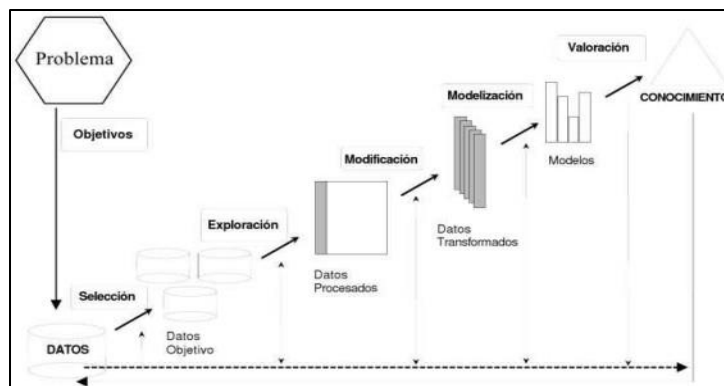


Figura 16. Proceso de Minería de Datos.

Fuente: recuperado <https://www.gestiopolis.com/que-es-data-mining/>

2.3.2 METODOLOGÍAS DESARROLLADAS

Durante el proceso de análisis de la información se emplearon técnicas relacionadas a la Analítica de datos, que apoyaron en la identificación de comportamientos y patrones en función de las variables que son parte de la investigación. Estas técnicas son denominadas descriptivas y predictivas que en su esencia lo que buscan es identificar correlaciones, tendencias, comportamientos y calcular probabilidades utilizando métodos de aprendizaje automático, como ser: el análisis y descomposición de series temporales a través de métodos ARIMA o VAR. Logrando con estos, identificar patrones estacionales, ciclos y tendencias que ayuden a obtener resultados lo más cercanos a la realidad.

Adicional, se empleó un modelo para analizar los componentes principales (PCA) con el objetivo de poder identificar las variables principales que influyen en el gasto público y de esta manera reducir la dimensionalidad de los datos en función del objetivo principal. Para cada técnica aplicada se examina el error obtenido al comprobar los resultados observados con los que el modelo predice, Esto con el fin de realizar ajustes necesarios a los parámetros de cada modelo y lograr mejor precisión en los resultados. Asimismo, se realiza una separación del conjunto de datos tomando un 70% para entrenamiento del modelo y un 30% para prueba.

En la evaluación del modelo se consideró la matriz de confusión y validación cruzada en el caso de los algoritmos de clasificación y para series temporales se utilizó Error Cuadrático Medio (MSE), Promedios Móviles (MA), Factores de Estacionalidad y grafica de residuos.

Se ha utilizado KNIME y Python como herramientas para la construcción y evaluación de los modelos predictivos.

2.4 MARCO LEGAL

En el **artículo 3** correspondiente al **Decreto No. 83-2004** publicado por la Gaceta se enuncian los objetivos de la Ley Orgánica de presupuesto:

- 1) Garantizar el cumplimiento del Principio de Legalidad, así como los criterios de eficiencia y eficacia en la obtención y aplicación financiera y económica de los recursos públicos.
- 2) Sistematizar las operaciones de programación, administración, evaluación y control de los recursos financieros del Estado, sin perjuicio de las acciones que correspondan a sus Entes Fiscalizadores.
- 3) Desarrollar sistemas que proporcionen información oportuna y confiable sobre el comportamiento financiero del Sector Público, que sea útil para la conducción de los órganos u Organismos correspondientes y para evaluar la gestión de los funcionarios responsables.
- 4) Establecer como responsabilidad propia del Poder Ejecutivo, por medio de los órganos competentes, según lo previsto en esta Ley, el mantenimiento y operación de:
 - a) Un sistema contable adecuado a las necesidades del registro e información de todos los actos con incidencia patrimonial, financiera o económica, acorde con la naturaleza jurídica y características operativas,
 - b) Un eficiente y eficaz control interno, previo, simultáneo o posterior, según el caso, de las operaciones de los distintos subsistemas que conforman la Administración Financiera del Sector Público;
 - c) Procedimientos adecuados que aseguren la conducción eficiente y eficaz de las actividades de los distintos Órganos u Organismos que intervienen en la administración financiera y la evaluación de los resultados de los correspondientes programas, proyectos o actividades; y,
 - d) Normas operativas que deberán cumplir obligatoriamente los Órganos u Organismos del Sector Público en cuanto ejecuten actividades propias de la administración financiera. (La Gaceta, 2004)

En el **Decreto No. 107-2021** publicado en el Diario oficial la Gaceta se enuncia lo siguiente:

Que el **Artículo 205** de la Constitución de la República Atribución 32) establece que es potestad del Congreso Nacional de Honduras la aprobación anual del Presupuesto General de Ingresos y Egresos de la República, tomando como base el proyecto que emita el Poder Ejecutivo, debidamente desglosado y resolver sobre su modificación. (La Gaceta, 2021)

Por su parte, en el artículo 9 de este decreto se indica la instrucción que dicta la forma en la que se debe reportar la ejecución presupuestaria:

Todas las instituciones del Gobierno: Central, Desconcentrado y Descentralizado, deberán remitir oficialmente dentro de los primeros quince (15) días del mes siguiente de finalizado el trimestre a la Secretaría de Estado en el Despacho de Finanzas, el informe trimestral en formato digital mediante la entrega de 3 CD, conteniendo su respectiva ejecución física y financiera del Plan Operativo Anual y del Presupuesto, así como cualquier otra información de relevancia que facilite el análisis de dicho informe, esta misma información debe estar disponible en los sitios WEB de

cada institución. Estos informes deberán ser elaborados trimestralmente y en forma acumulada por cada institución cumpliendo con el principio de Autoevaluación. El seguimiento y evaluación del Plan Operativo Anual y Presupuesto lo efectuarán las instituciones del Gobierno: Central, Desconcentrado y Descentralizado, sin excepción alguna, de acuerdo a las Normas Técnicas y Manuales de Procedimiento de los Subsistemas de Presupuesto, Tesorería, Contabilidad, Crédito Público e Inversión Pública aprobadas por la Secretaría de Estado en el Despacho de Finanzas, con base a la información registrada obligatoriamente en el Sistema de Administración Financiera Integrada. (La Gaceta, 2021)

Además, en el Decreto Ejecutivo Número PCM 19-2023 publicado en la Gaceta se enuncia lo siguiente:

Que en fecha 4 de mayo de 2016 fue publicado en el Diario Oficial “La Gaceta” el Decreto Legislativo No.25-2016 aprobado el 7 de abril del 2016, contentivo de la Ley de Responsabilidad Fiscal (LRF), el cual tiene como objetivo establecer los lineamientos para una mejor gestión de las finanzas públicas, asegurando la consistencia en el tiempo de la política presupuestaria y garantizar la consolidación fiscal, sostenibilidad de la deuda y reducción de la pobreza con responsabilidad, prudencia y transparencia fiscal. (La Gaceta, 2023)

CAPÍTULO III. METODOLOGÍA

3.1 CONGRUENCIA METODOLÓGICA

En el siguiente apartado, se presentará un detalle de la congruencia entre las variables y su operacionalización dentro del presente estudio. También se indicarán los instrumentos empleados en la investigación, así como las hipótesis planteadas. Además, se muestra la población utilizada objetivo y a su vez se explorarán las fuentes de información primaria y secundaria utilizadas. En este análisis podrá identificar los elementos esenciales en el proceso de la investigación, con el fin de lograr una comprensión completa de los enfoques metodológicos y las estrategias de recolección de datos aplicadas en el estudio.

3.1.1 MATRIZ METODOLÓGICA

En esta sección se expondrá la congruencia metodológica de la investigación. Esta es fundamental, ya que garantiza la coherencia y la validez de los resultados obtenidos. En este sentido, se mostrará cómo los métodos utilizados en la investigación son consistentes en relación con los objetivos y las preguntas planteadas.

Tabla 1. Matriz de Congruencia Metodológica

Matriz Metodológica					
Titulo	"Un Modelo de Analítica para Predecir el Presupuesto en los Sectores de Salud y Educación en Honduras a 2024"				
Problema	Objetivo General	Objetivos Específicos	Preguntas de Investigación	Variable Independiente	Variable Dependientes
¿Es una buena alternativa predecir el presupuesto de los sectores de salud y educación en Honduras a 2024 haciendo uso de algoritmos de Aprendizaje Automático?	Predecir el presupuesto de los sectores de salud y educación en Honduras para 2024 haciendo uso de los algoritmos de Aprendizaje Automático.	Identificar los objetos de gasto con mayor afectación en la ejecución del presupuesto de los sectores de salud y educación.	¿Es factible con una buena aproximación identificar a través de los objetos de gasto con mayor afectación en la ejecución del presupuesto de los sectores de salud y educación el presupuesto del año siguiente?	Modelos predictivos de analítica	Presupuesto de salud y educación
		Determinar el nivel de eficiencia en la ejecución del presupuesto de los	¿Cuál es el nivel de eficiencia en la ejecución del presupuesto de los sectores de salud y educación?		
		Generar una predicción del presupuesto de Salud y Educación para el año 2024.	¿Es eficiente el uso de datos históricos de los objetos de gasto de mayor afectación para generar un modelo eficaz de predicción del presupuesto del año siguiente?		

Fuente: Elaboración propia.

ESQUEMA DE VARIABLES DE ESTUDIO

A continuación, se presenta el esquema multivariado (Figura 17) los esquemas de las variables Independiente (Figura 18) y Dependiente (Figura 19) de estudio de la investigación.

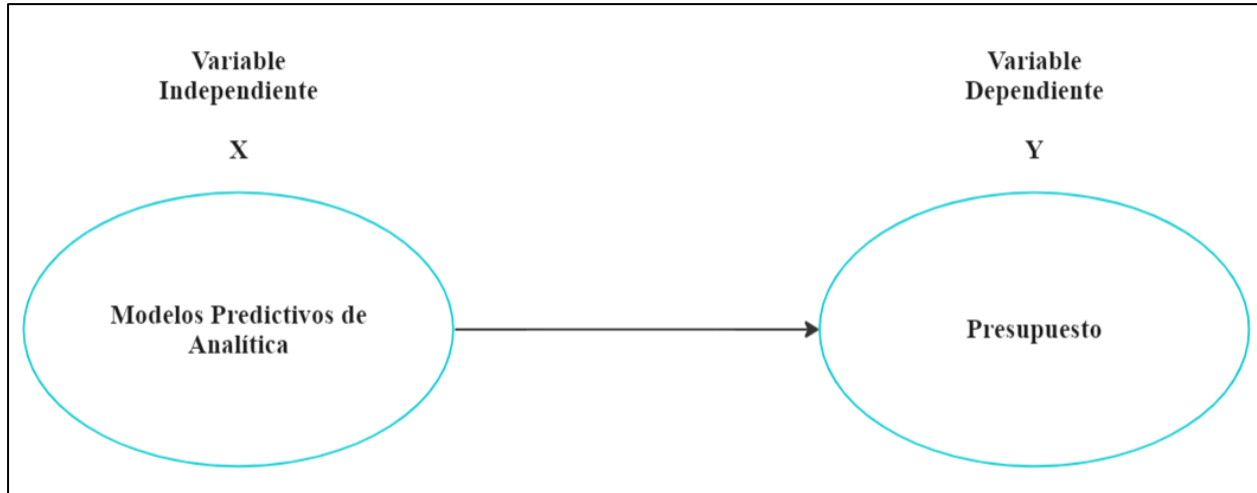


Figura 17. Esquema de relación causal multivariada

Fuente: Elaboración propia.

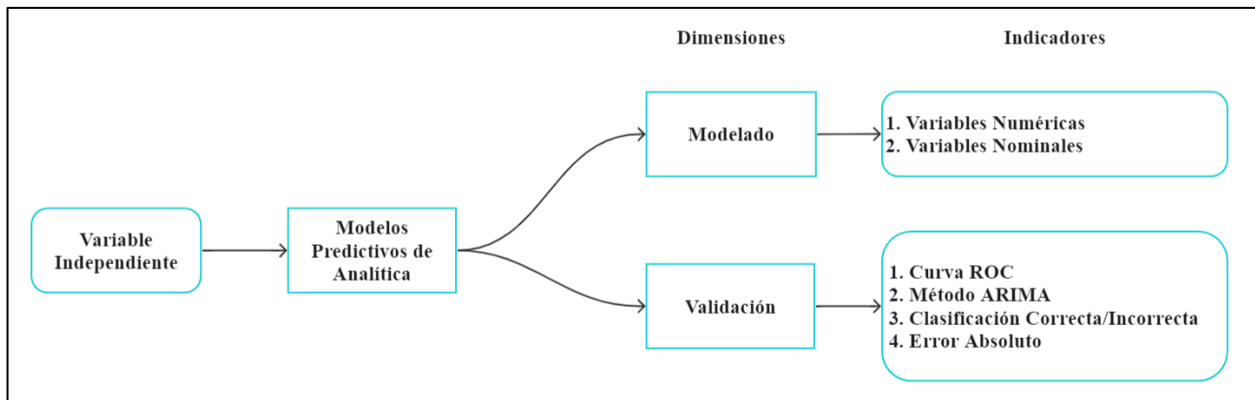


Figura 18. Esquema de variable independiente

Fuente: Elaboración propia.

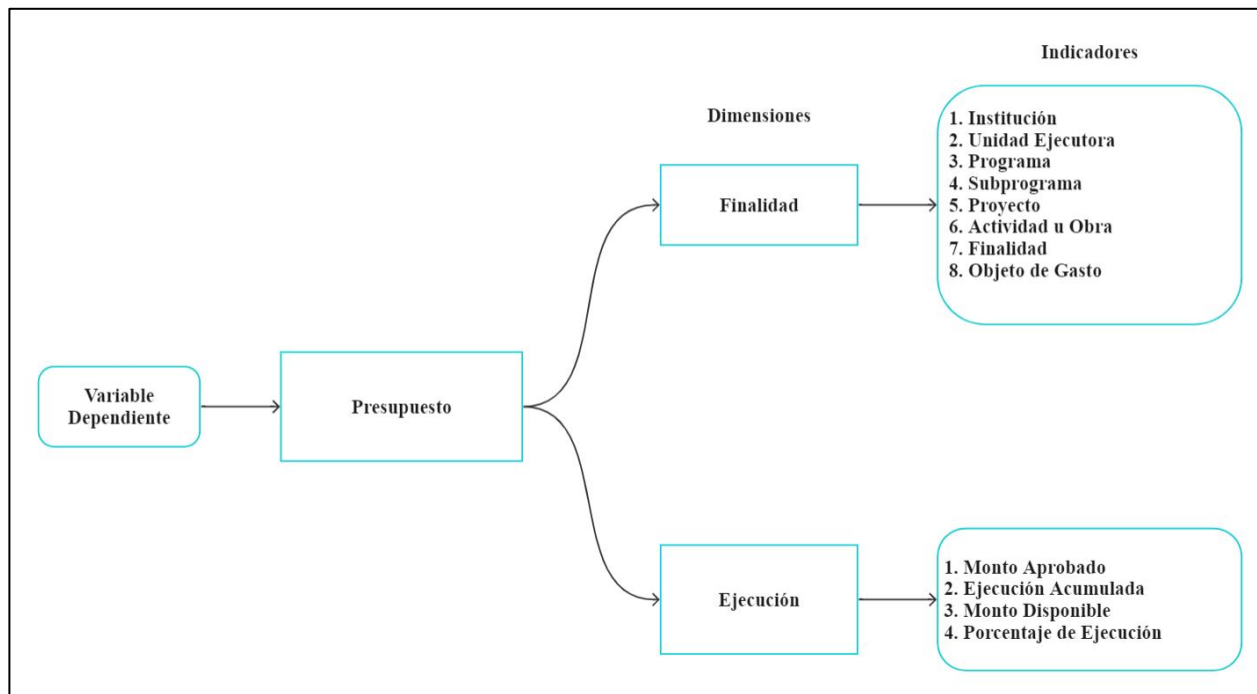


Figura 19. Esquema de variable dependiente

Fuente: Elaboración propia.

3.1.2 OPERACIONALIZACIÓN DE LAS VARIABLES

En la Tabla 2 se presenta la operacionalización de las variables, que consiste en definir de manera explícita la manera en la que se estudiará cada una de ellas en la investigación. Está compuesta por la definición conceptual y operativa de la variable, las dimensiones, indicadores y las preguntas/respuestas que serán de utilizas para otorgar valor.

Tabla 2. Operacionalización de las variables

Variable	Definición Conceptual	Definición Operacional	Dimensión	Indicadores	Pregunta	Respuesta	Instrumento
Independiente Modelos Predictivos de Analítica	Un modelo predictivo es el que se encargará de efectuar las predicciones recibiendo una serie de datos nuevos para realizar las clasificaciones de acuerdo con los patrones detectados en la etapa de entrenamiento. Por ejemplo, si se usa un modelo entrenado con datos históricos de clientes para evaluar el riesgo de cancelación de una tarjeta de crédito, el modelo categorizará a los nuevos clientes en función de su comportamiento y preverá el riesgo de cancelación. (González, 2021)	Es un modelo matemático que permitirá predecir el presupuesto para los sectores de educación y salud en Honduras, haciendo uso de algoritmos de aprendizaje automático.	Modelado	1. Variables Numéricas 2. Variables Nominales	1. ¿Cuántas Variables Numéricas hay? 2. ¿Cuántas Variables Nominales hay?	Conteo	Observación en Conjunto de Datos
			Validación	1. Curva ROC 2. Método ARIMA 3. Clasificación Correcta/Incorrecta 4. Error Absoluto	1. ¿Está el valor de la curva ROC en un punto óptimo? 2. ¿Es estadísticamente significativo cada término en el modelo? 3. ¿Cuál es el valor obtenido de precisión y sensibilidad en el modelo? 4. ¿Cuál es la media de los errores individuales obtenidos?	Valor	
Dependiente Presupuesto	Este término hace referencia al monto de dinero que es necesario para afrontar una serie de gastos originados de uno o varios proyectos. En otras palabras, corresponde a montos anticipados resultantes de la estimación para la realización de los objetivos propuestos generalmente en el inicio de períodos. (Sánchez, 2020)	Es una estimación del monto destinado a los sectores de salud y educación en Honduras.	Finalidad	1. Institución 2. Unidad Ejecutora 3. Programa 4. Subprograma 5. Proyecto 6. Actividad u Obra 7. Finalidad 8. Objeto de Gasto	1. ¿Cuántas instituciones forman parte del presupuesto de los sectores de salud y educación en Honduras? 2. ¿Cuántas unidades ejecutoras forman parte del presupuesto de los sectores de salud y educación en Honduras? 3. ¿Cuántos programas forman parte del presupuesto de los sectores de salud y educación en Honduras? 4. ¿Cuántos subprogramas forman parte del presupuesto de los sectores de salud y educación en Honduras? 5. ¿Cuántos proyectos forman parte del presupuesto de los sectores de salud y educación en Honduras? 6. ¿Cuántas actividades u obras forman parte del presupuesto de los sectores de salud y educación en Honduras? 7. ¿Cuántas finalidades forman parte del presupuesto de los sectores de salud y educación en Honduras? 8. ¿Cuántos objetos de gasto forman parte del presupuesto de los sectores de salud y educación en Honduras?	Conteo	Observación en Conjunto de Datos
			Ejecución	1. Monto Aprobado 2. Ejecución Acumulada 3. Monto Disponible 4. Porcentaje de Ejecución	1. ¿Cuál ha sido el monto aprobado de presupuesto de los sectores de salud y educación en Honduras en el período comprendido entre 2010 y 2023? 2. ¿Cuál ha sido la ejecución acumulada de presupuesto de los sectores de salud y educación en Honduras en el período comprendido entre 2010 y 2023? 3. ¿Cuál ha sido el monto disponible de presupuesto de los sectores de salud y educación en Honduras en el período comprendido entre 2010 y 2023? 4. ¿Cuál ha sido el porcentaje de ejecución de presupuesto de los sectores de salud y educación en Honduras en el período comprendido entre 2010 y 2023?	Valor	

Fuente: Elaboración propia.

Nota: En la operacionalización la respuesta para los indicadores se planteó de dos formas: “Valor” para cada una de las variables que estén relacionadas a la cantidad correspondiente a valores de validación y ejecución y “Conteo” hace referencia a las veces en las cuales se repite una circunstancia en los datos.

3.2 ENFOQUE Y MÉTODOS

En el siguiente apartado se describirá lo relacionado al enfoque y métodos de la investigación (Figura 20), incluyendo el alcance, el diseño, los instrumentos y las técnicas.

Enfoque	Cuantitativo
Alcance	Descriptivo
Diseño	No Experimental
Instrumento	Dataset obtenido de los datos abiertos de SEFIN
Técnica	Analítica de Datos

Figura 20. Enfoque y métodos

Fuente: Elaboración propia.

3.2.1 Enfoque

La investigación tiene un enfoque cuantitativo, ya que tiene por finalidad evaluar las variaciones en el presupuesto de los sectores de salud y educación a lo largo de los últimos años, es decir, que se basa en procesos tanto estadísticos como inferenciales de analítica de datos para cuantificar las relaciones de dependencia existentes entre los elementos que son relacionados directamente con el problema de investigación.

3.2.2 Alcance

El alcance de la presente investigación es de tipo descriptivo porque busca especificar propiedades, características y perfiles de objetos, perfiles de personas, etc. Es decir, que pretenden obtener información ya sea individual o colectiva de las variables en cuestión. (Mata, 2019)

3.2.3 Diseño

La investigación es de tipo no experimental, debido a que no existe manipulación deliberada de las variables, estas solo son observadas en su contexto. Además, la investigación reviste un corte de carácter transversal dado que la explicación del problema solo requiere de una única medición de datos en el tiempo. La valoración de las variaciones en la predicción del presupuesto puede estimarse con una única recolección de datos, en este caso, bastará con el conjunto de datos obtenido de los datos abiertos de SEFIN.

3.2.4 Instrumento

Como instrumento será utilizado un conjunto de datos obtenido de los datos abiertos de SEFIN para los períodos comprendidos entre 2010 y 2023, el cual contiene información fundamental y la que permitirá que la investigación se lleve a cabo, yendo desde la Institución, Objeto del Gasto, Finalidad, Montos de Presupuesto Aprobados y las ejecuciones correspondientes.

3.2.5 Técnica

La técnica aplicada para el procesamiento de la base de datos será la Analítica de Datos, cuyo término se utiliza para referirse a un conjunto de aplicaciones, desde la inteligencia empresarial básica, el procesamiento analítico en línea y la elaboración de informes, hasta diversas formas de analítica avanzada (modelos predictivos). En este sentido, su enfoque es similar al de la analítica de negocio, que abarca diversos enfoques del análisis de datos, pero su alcance es más amplio. Mientras que la analítica de negocio se centra en aplicaciones empresariales específicas, la analítica de datos se utiliza en una variedad de campos y disciplinas. (Stedman, 2021)

3.3 DISEÑO DE LA INVESTIGACIÓN

3.3.1 POBLACIÓN

En términos generales, si tiene la capacidad de reconocer y definir adecuadamente a un individuo, o incluso de manera más amplia, a un elemento, y luego observamos y medimos una característica o atributo específico de este, consideramos que cualquier grupo de estos individuos constituye una población. (Guerra Bustillo, 2003)

Partiendo de este concepto en la presente investigación se determina una población finita a partir una observación de datos, que los investigadores obtuvieron de los datos abiertos de la ejecución presupuestaria de SEFIN, conteniendo la información de aproximadamente 7,821,800 registros correspondientes al periodo entre 2010 y 2023.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

En la actualidad muchas organizaciones están apuntando por hacer uso de la analítica de datos y la inteligencia de negocios como parte del plan interno con el fin de tener una ventaja competitiva en el mercado. (Jaklič, 2018)

Se ha empleado una técnica basada en la analítica de los datos y en la observación del conjunto de datos abiertos publicado en el sitio oficial de SEFIN con la información correspondiente a la ejecución presupuestaria de 2010 a 2023 en Honduras, ya que se hace uso de modelos predictivos que buscan anticiparse a un resultado, que se obtiene del análisis del comportamiento de ciertas variables que son las determinantes en los resultados del modelo.

Se empleó como herramienta de validación un cuestionario de observación con el propósito de ir validando los resultados de cada uno de los indicadores con los cuales se van operacionalizando cada una de las variables establecidas para el presente estudio.

3.5 FUENTES DE INFORMACIÓN

A continuación, se detallan las fuentes primarias y secundarias en las que se centró la presente investigación:

3.5.1 FUENTES PRIMARIAS

Las fuentes primarias son aquellas que se conforman por información o datos que han sido publicados por primera vez y que no han sido evaluadas y/o modificados por nadie. Generalmente pueden encontrarse en formato impreso o digital. (Universidad de Guadalajara, 2023)

En este estudio al tratarse de un análisis haciendo uso de datos abiertos, no se cuenta con una fuente primaria.

3.5.2 FUENTES SECUNDARIAS

Están conformadas a partir de revistas, artículos, tesis, información en sitios web y libros, es decir la colección de referencia de una biblioteca. El fin de estas fuentes es que el acceso a las fuentes primarias sea más factible. Generalmente las fuentes secundarias son utilizadas cuando por alguna razón no se tiene acceso directo a las fuentes primarias, se tienen recursos limitados o si el investigador duda acerca de la confiabilidad de la fuente primaria. Desde otra perspectiva, suelen complementar la información obtenida de la fuente primaria. (Universidad de Guadalajara, 2023)

Para el presente estudio el principal insumo identificado como fuente secundaria es el conjunto de datos abiertos publicado en el sitio oficial de SEFIN con la información correspondiente a la ejecución presupuestaria de 2010 a 2023 en Honduras.

CAPÍTULO IV. RESULTADOS Y ANÁLISIS

4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS

Para este proceso se utilizaron datos abiertos de SEFIN, en el segmento de ejecución presupuestaria y de los que se obtuvieron las fuentes del sitio oficial de SEFIN sobre el presupuesto aprobado y ejecutado comprendido desde 2010 a 2023. En este conjunto de datos se encuentran variables numéricas, cadenas de texto y variables nominales/catóricas con un total de 56 campos por conjunto de datos anual.

Tabla 3. Campos utilizados en conjunto de datos

ID	NOMBRE	TIPO	DESCRIPCIÓN	ID	NOMBRE	TIPO	DESCRIPCIÓN
1	Gestión	Numérico	Año del presupuesto	29	Dia_4	Numérico	Cantidad de presupuesto ejecutado en día 4
2	Mes	Numérico	Mes del presupuesto	30	Dia_5	Numérico	Cantidad de presupuesto ejecutado en día 5
3	Administración	Texto	Nivel de Administración pudiendo ser central o instituciones descentralizadas.	31	Dia_6	Numérico	Cantidad de presupuesto ejecutado en día 6
4	Institución	Texto	Nivel que contiene los niveles contables o unidades ejecutoras.	32	Dia_7	Numérico	Cantidad de presupuesto ejecutado en día 7
5	Ger_Administrativa	Texto	Gerencia Administrativa	33	Dia_8	Numérico	Cantidad de presupuesto ejecutado en día 8
6	U_Ejecutora	Texto	Unidad responsable de ejecutar el presupuesto asignado.	34	Dia_9	Numérico	Cantidad de presupuesto ejecutado en día 9
7	Programa	Texto	Contiene los subprogramas, las actividades, los proyectos y las obras.	35	Dia_10	Numérico	Cantidad de presupuesto ejecutado en día 10
8	S_Programa	Texto	Se realizan productos finales parciales.	36	Dia_11	Numérico	Cantidad de presupuesto ejecutado en día 11
9	Proyecto	Texto	Generan una inversión o un bien y satisfacen una necesidad de la población al finalizar su elaboración	37	Dia_12	Numérico	Cantidad de presupuesto ejecutado en día 12
10	Actividad_u_Obra	Texto	Es donde se realizan las labores, acciones con las cuales se generaran los productos finales.	38	Dia_13	Numérico	Cantidad de presupuesto ejecutado en día 13
11	Organismo	Texto	Organismos fuentes del presupuesto.	39	Dia_14	Numérico	Cantidad de presupuesto ejecutado en día 14
12	F_Financiamiento	Texto	Permite al gobierno identificar la fuente de origen de los ingresos, así como la orientación de los mismos hacia la atención de las necesidades públicas.	40	Dia_15	Numérico	Cantidad de presupuesto ejecutado en día 15
13	BIP	Texto	Nombre de proyectos asignados en la ejecución del presupuesto.	41	Dia_16	Numérico	Cantidad de presupuesto ejecutado en día 16
14	Finalidad	Texto	Rubro de ejecución del presupuesto.	42	Dia_17	Numérico	Cantidad de presupuesto ejecutado en día 17
15	Función	Texto	Función específica de la finalidad de ejecución del presupuesto.	43	Dia_18	Numérico	Cantidad de presupuesto ejecutado en día 18
16	Grupo_de_Objeto	Texto	Agrupador de Objetos de Gasto.	44	Dia_19	Numérico	Cantidad de presupuesto ejecutado en día 19
17	Objeto_del_Gasto	Texto	Comprende a la ordenación sistemática y homogénea de los bienes, servicios y las transferencias que el Sector Público aplica en el desarrollo de su gestión.	45	Dia_20	Numérico	Cantidad de presupuesto ejecutado en día 20
18	Ben_Tranf	Texto	Es la clasificación que identifica a los beneficiarios de la operación.	46	Dia_21	Numérico	Cantidad de presupuesto ejecutado en día 21
19	Aprobado_Ministro	Numérico	Presupuesto Aprobado por los miembros de la cámara legislativa.	47	Dia_22	Numérico	Cantidad de presupuesto ejecutado en día 22
20	Aprobado	Numérico	Es el presupuesto aprobado por el Congreso Nacional anualmente.	48	Dia_23	Numérico	Cantidad de presupuesto ejecutado en día 23
21	Vigente	Numérico	Comprende las modificaciones presupuestarias (ampliación, disminución y traslados) de la estimación de los recursos y de los créditos presupuestarios autorizados por el Poder Legislativo en el Presupuesto General de Ingresos y Egresos de la República.	49	Dia_24	Numérico	Cantidad de presupuesto ejecutado en día 24
22	Eje_Enero	Numérico	Refiere al Devengado/Ejecutado Mensual.	50	Dia_25	Numérico	Cantidad de presupuesto ejecutado en día 25
23	Eje_Acum	Numérico	Es la base contable por la cual las transacciones y otros hechos son reconocidos cuando ocurren y no cuando se efectúa su cobro o su pago en efectivo o su equivalente.	51	Dia_26	Numérico	Cantidad de presupuesto ejecutado en día 26
24	Disponible	Numérico	Es el resultado de lo Presupuestado en la columna Vigente menos la columna Ejecutado_Acumulado.	52	Dia_27	Numérico	Cantidad de presupuesto ejecutado en día 27
25	Porc_Ejecucion	Numérico	Es el Porcentaje Ejecutado Acumulado	53	Dia_28	Numérico	Cantidad de presupuesto ejecutado en día 28
26	Dia_1	Numérico	Cantidad de presupuesto ejecutado en día 1	54	Dia_29	Numérico	Cantidad de presupuesto ejecutado en día 29
27	Dia_2	Numérico	Cantidad de presupuesto ejecutado en día 2	55	Dia_30	Numérico	Cantidad de presupuesto ejecutado en día 30
28	Dia_3	Numérico	Cantidad de presupuesto ejecutado en día 3	56	Dia_31	Numérico	Cantidad de presupuesto ejecutado en día 31

Fuente: Elaboración propia.

4.1.1 PREPARACIÓN Y LIMPIEZA DE CONJUNTO DE DATOS

En primera instancia, el mayor reto encontrado en el conjunto de datos fue la unificación en un solo archivo base, y es que en el sitio oficial de SEFIN se encuentran distribuidos por año, por lo que fue necesario utilizar la herramienta Knime y trabajar con más de 7 millones de registros.

Una vez unificado del conjunto de datos, fue necesario revisar la calidad de los registros existentes, detectando comportamientos tales como: los que tenían palabras que incluyen vocales tildadas se mostraban con caracteres extraños, asimismo las palabras que contienen la letra ñ mostraban caracteres que distorsionaban la información de cara al análisis de datos. Para solventar esta situación, se utiliza el nodo denominado “String Manipulation (Multi Column)” de Knime para que explore todos los caracteres resultantes de los escenarios planteados y dejarlos en un estado uniforme.

Otro punto importante fue identificar que campos clave como la “Finalidad” mostraba valores duplicados que se originaban por una coma extra, una letra tildada, etc. por lo que fue necesario aplicar el nodo “String Manipulation” de Knime para dejar cada finalidad como valor único en el conjunto de datos.

Además, para manejar de forma más adecuada los períodos de tiempo se transformaron los campos “ANIO” y “MES” a tipo cadena para concatenarlos posteriormente en un campo denominado “PERIODO”.

Finalmente, se detecta que la manera en la cual se efectúan los registros de la ejecución presupuestaria tiende a ser ambigua en algunos casos debido a que se engloban bajo la misma clasificación varios ámbitos. Por ejemplo, en el sector educación se observa que dentro de las funciones destinadas a la ejecución presupuestaria se engloba “90 – Educación, Investigación, Cultura y Actividades Recreativas” por lo que al momento de rendición de cuentas y transparencia podría provocar inconvenientes.

4.2 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

En este capítulo se presentan los principales hallazgos obtenidos del proceso de análisis de los datos de estudio con el fin de que sea posible elaborar conclusiones y recomendaciones fundamentadas y útiles para brindar respuesta al problema de investigación.

4.2.1 RESULTADOS CUANTITATIVOS

En este apartado se presentan los hallazgos, relaciones y los patrones más relevantes que se detectaron en el conjunto de datos, haciendo uso de la técnica de Análisis Exploratorio de Datos (EDA) cuyo propósito principal es explicar de forma eficaz los conjuntos de datos analizados.

4.2.1.1 Sector educación

En cuanto a ejecución presupuestaria a lo largo de los últimos años, en la Figura 21 se logra observar que la ejecución ha ido variando, mostrando una tendencia al aumento con el pasar de los años. Además, se observa que en los últimos seis años el presupuesto ha aumentado sustancialmente. Es importante aclarar que para 2023 se cuentan datos hasta septiembre, es por ello por lo que se ve una baja para dicho año.

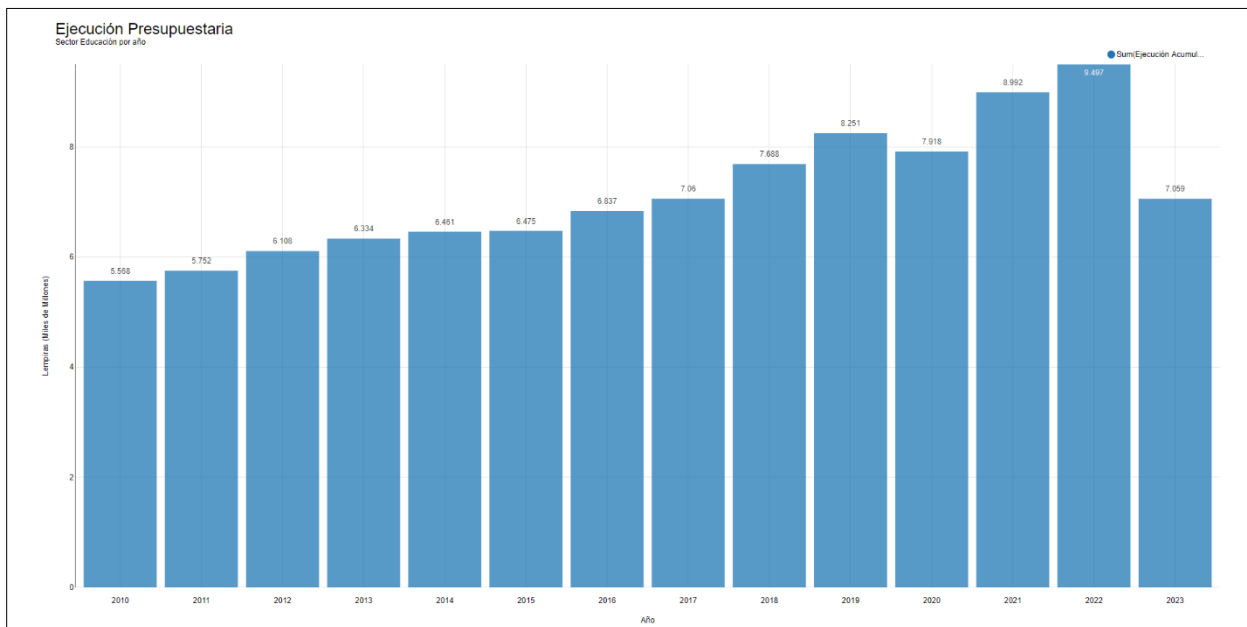


Figura 21. Ejecución presupuestaria del sector educación por año.

Fuente: Elaboración propia.

En la Figura 22 se observa una serie obtenida de la ejecución presupuestaria a lo largo de los períodos 2010 a 2023, mostrándose ciclos bien definidos de acuerdo con el presupuesto correspondiente a cada año.

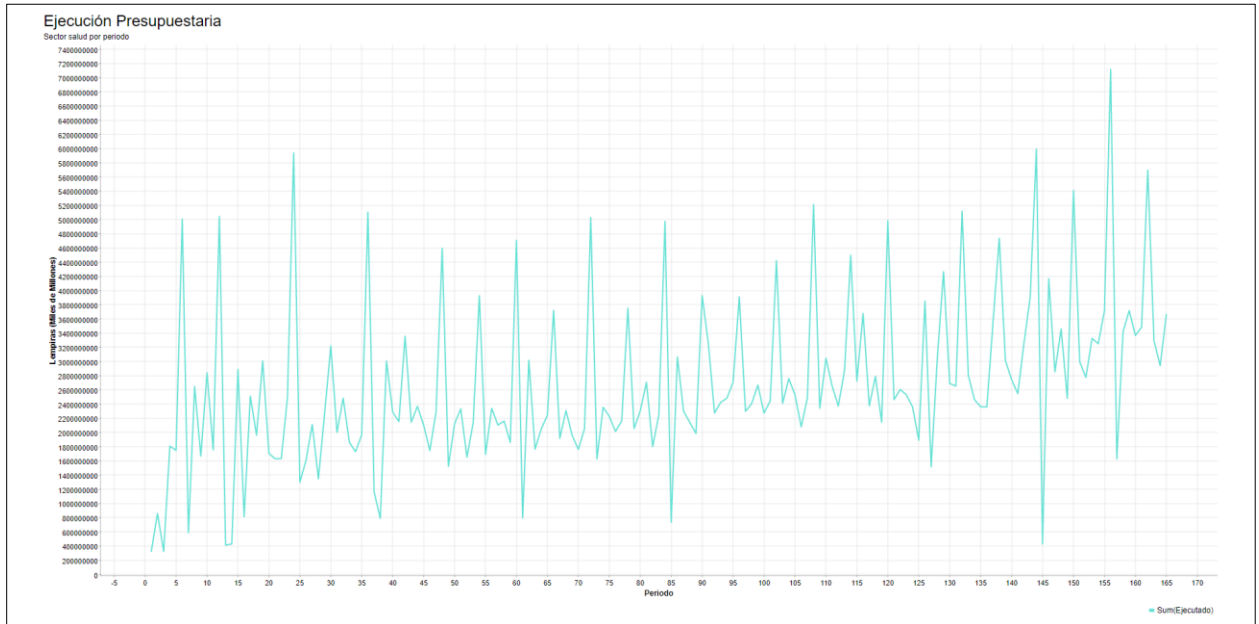


Figura 22. Serie de ejecución presupuestaria del sector educación por período.

Fuente: Elaboración propia.

El objeto de gasto representa a los bienes, servicios y transferencias que el estado ejecuta en la gestión presupuestaria, por lo que en la Figura 23 se logra observar que los objetos de gasto en los que más se ha orientado el presupuesto son los sueldos básicos de docentes (25.75%), adicionales (17.12%), sueldos básicos (10.90%) y transferencias corrientes a universidades (7.95%).

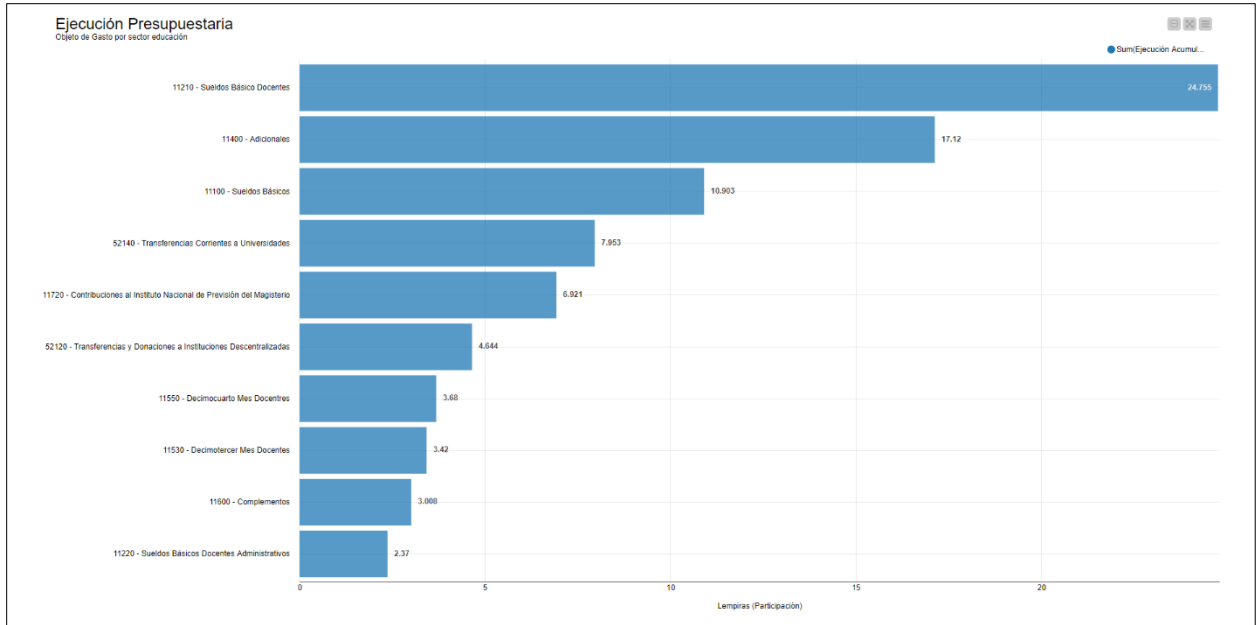


Figura 23. Ejecución presupuestaria del sector educación por objeto de gasto.
Fuente: Elaboración propia.

Por su parte, el grupo de gasto asocia la serie de objetos de gasto que el estado ejecuta en la gestión presupuestaria. En la Figura 24 se observa que los servicios personales (relacionados con los sueldos y salarios) acaparan el 78.84 % y las transferencias y donaciones con 16.79% son los grupos en los que se concentra principalmente el presupuesto.

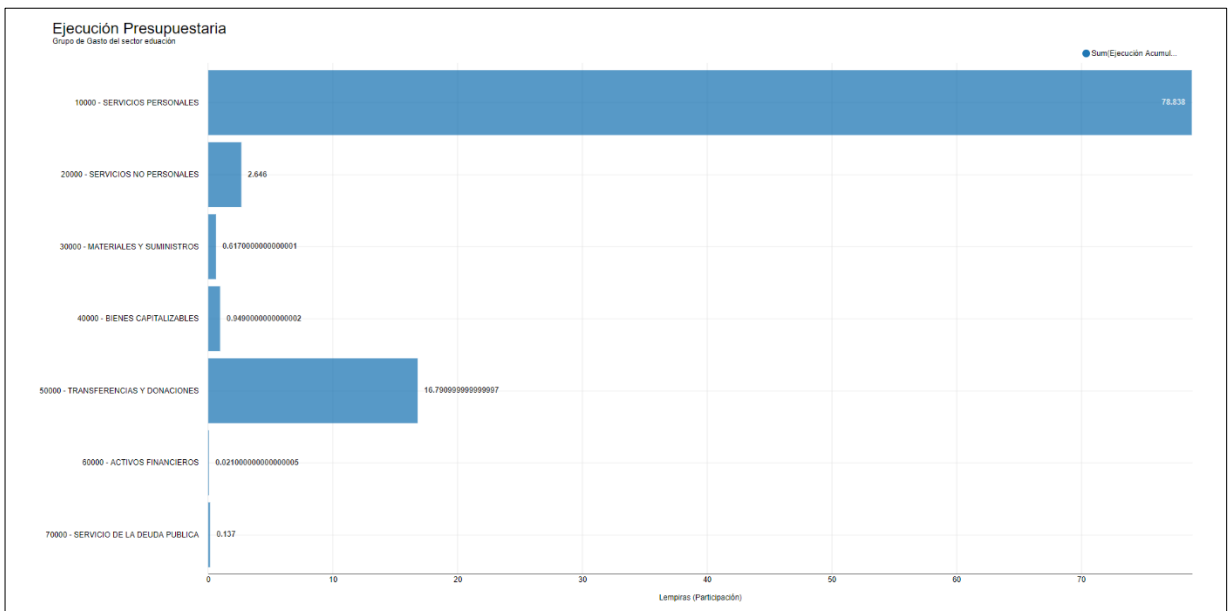


Figura 24. Ejecución presupuestaria del sector educación por grupo de gasto.
Fuente: Elaboración propia.

Además, en la Figura 25 se realiza el cruce de la ejecución presupuestaria por institución, resaltando que la Secretaría de Educación se lleva la gran parte (84.67%) debido a que esta es la encargada de la educación media, básica y prebásica, seguido de la UNAH (8.95%), INFOP (2.55%) y la UPNFM (1.76%).

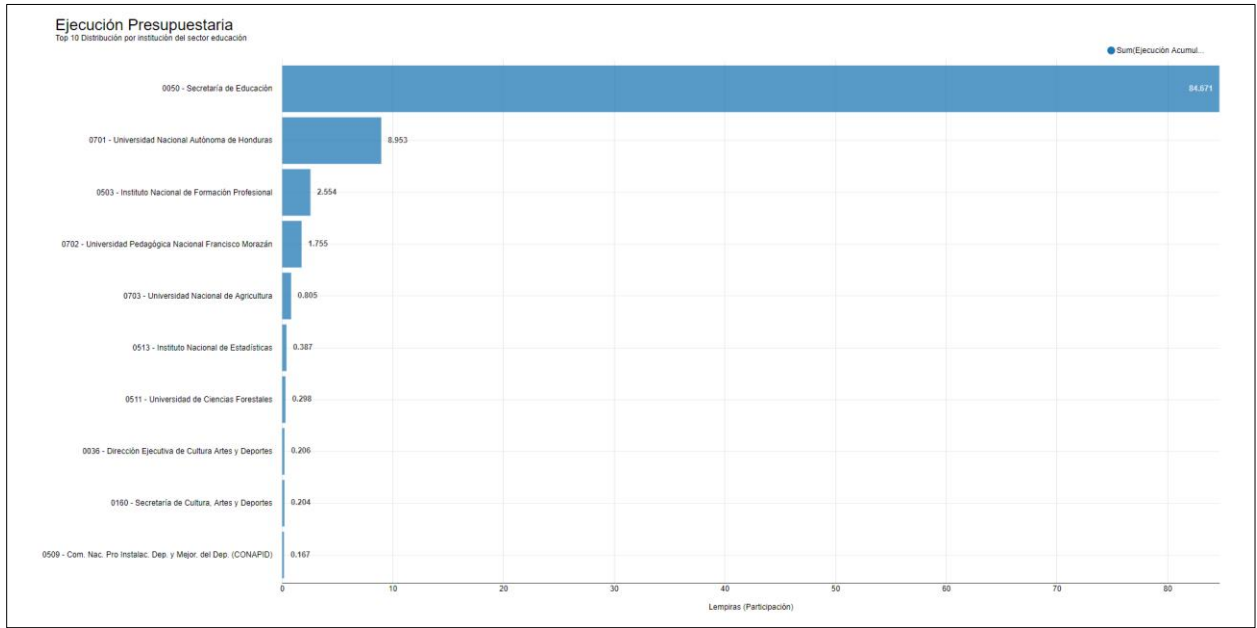


Figura 25. Ejecución presupuestaria del sector educación por institución.
 Fuente: Elaboración propia.

La actividad u obra representa los lugares/razones en donde se realizan los productos finales del sector. En la Figura 26 se observa que los servicios de educación básica (39.10%) son los que más presupuesto utilizan, seguido de los servicios de educación media (21.38%) y las transferencias a universidades (15.81%).

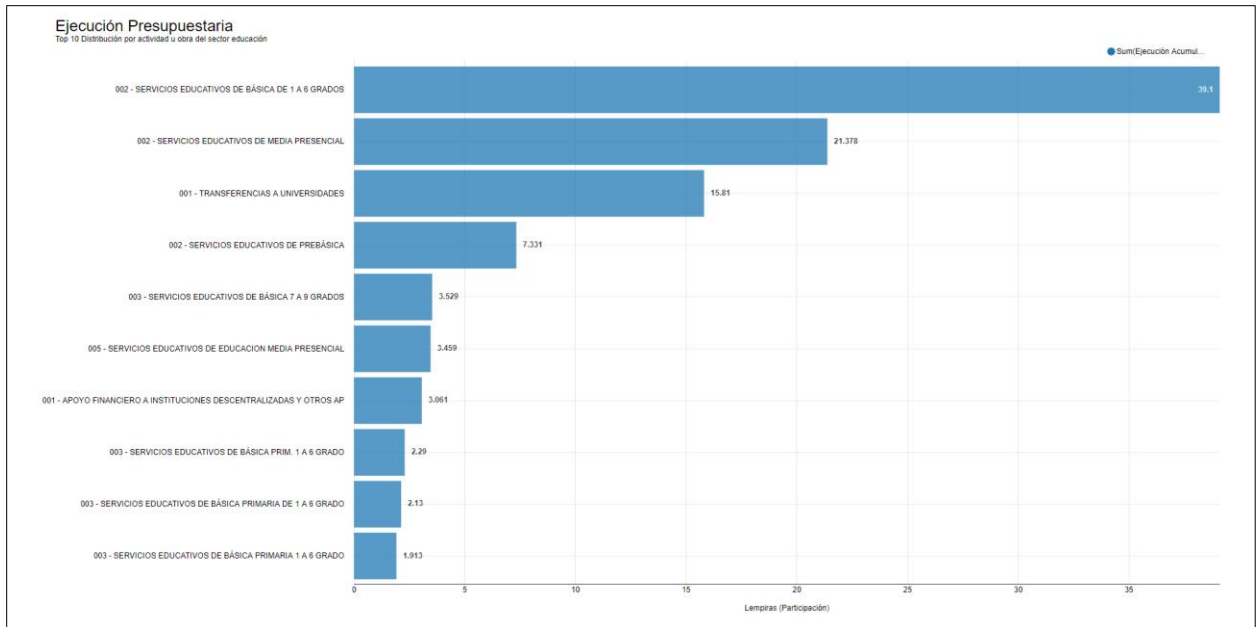


Figura 26. Ejecución presupuestaria del sector educación por actividad/obra.

Fuente: Elaboración propia.

Y para la función de la ejecución presupuestaria del sector educación se observa que gran parte del presupuesto va destinado a enseñanza preescolar y primaria (37.47%), seguido de educación, investigación, cultura y actividades recreativas (31.51%), también aparece la enseñanza secundaria (15.84%) y la enseñanza superior (10.91%). Lo anterior se observa en la Figura 27.

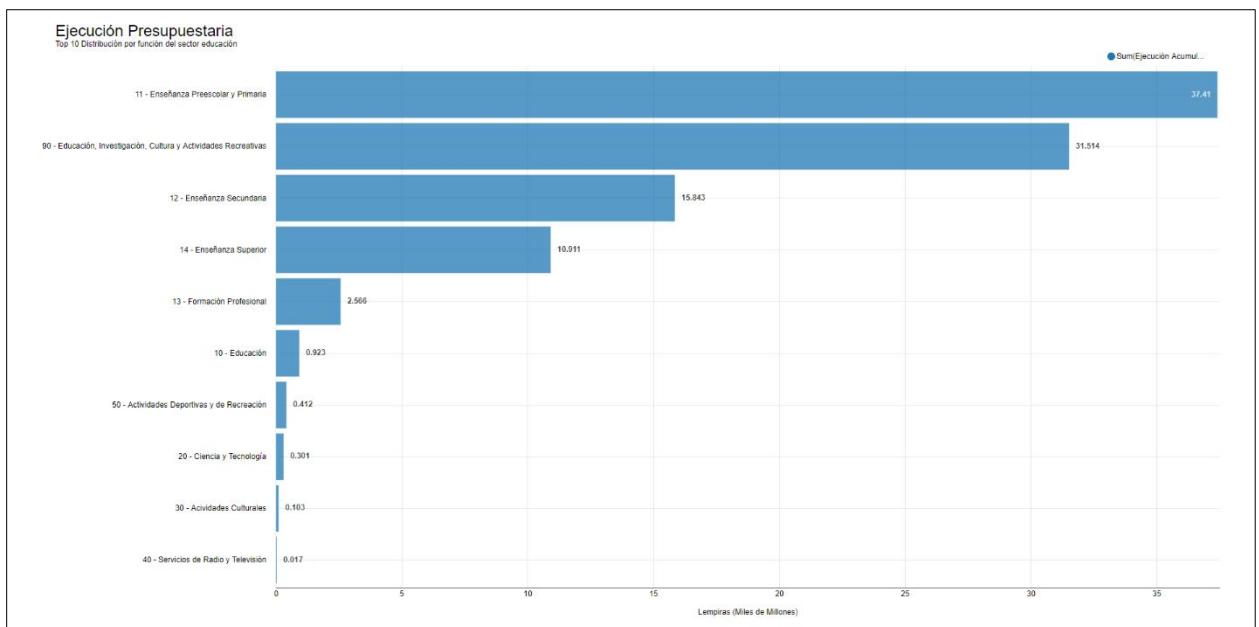


Figura 27. Ejecución presupuestaria del sector educación por función.

Fuente: Elaboración propia.

En la Figura 28 se observa la tendencia central de la ejecución presupuestaria la cual muestra una tendencia que va hacia el aumento año con año, con excepción de 2020 (Pandemia de COVID-19) y en 2023 (Se cuenta con datos hasta septiembre de dicho año). Lo anterior significa que ha existido un crecimiento sostenido de un año a otro a nivel general, lo que se traduce como una ejecución presupuestaria consistente y con tendencia en sentido positivo.

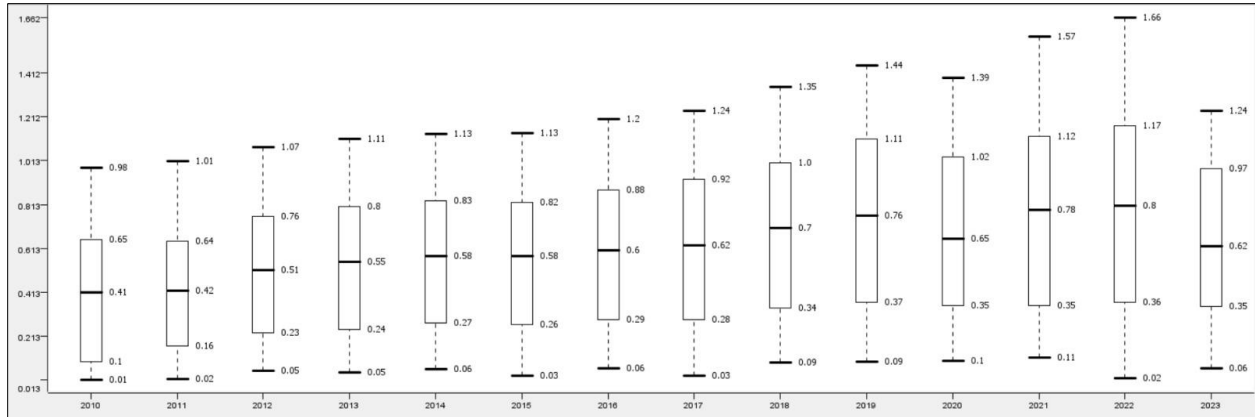


Figura 28. Tendencia central de la ejecución presupuestaria en el sector educación.

Fuente: Elaboración propia.

Es importante mencionar el porcentaje de ejecución presupuestaria a lo largo de los años, ya que a partir de este se realiza una calificación de la eficiencia en la ejecución presupuestaria que se muestra en la Figura 29, la cual describe los umbrales del indicador “Ejecución Presupuestaria” elaborado por las Empresas Públicas de Cundinamarca (EPC) de Colombia y en el que se realiza la clasificación entre Satisfactorio, Adecuado o Deficiente.

SEMÁFORO	RANGO	OPCIONES DE MANEJO
Satisfactorio	50%-100%	Cumple satisfactoriamente con la meta.
Adecuado	39%-49%	Es el rango en el cual se deberá realizar un plan de mejoramiento, para evitar el incumplimiento.
Deficiente	0%-38%	Es el rango en el cual se deberá realizar una acción correctiva para eliminar las causas de dicho incumplimiento.

Figura 29. Umbrales de Clasificación de la Ejecución Presupuestaria.

Fuente: recuperado de (EPC, 2021).

En la Figura 30 se visualiza la ejecución por Administración Central en la cual se observa que, en los años 2011, 2012 y 2017 el porcentaje de ejecución ha superado el 85%. No obstante, también denota que en 2013 y 2020 es en donde hubo menor ejecución, con valores menores al 55%. Dicho lo anterior y tomando la Figura 29 como referencia se puede clasificar como satisfactoria la ejecución presupuestaria en esta categoría.

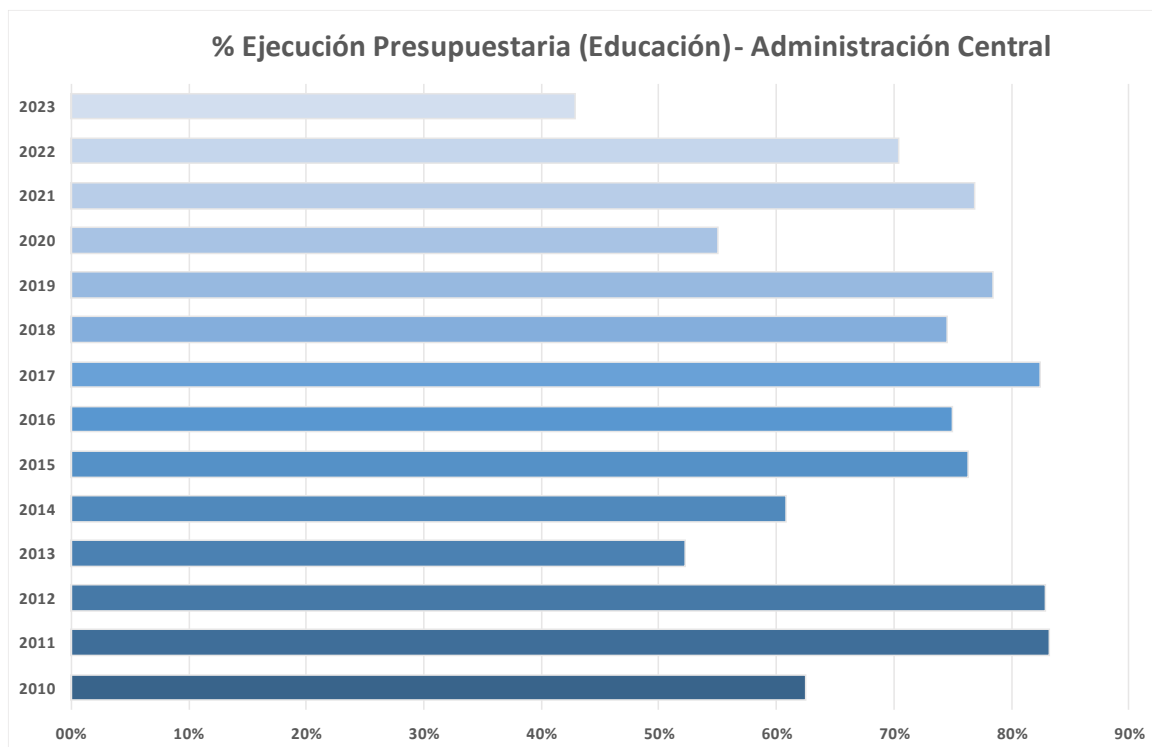


Figura 30. % de ejecución presupuestaria por Admón. Central - Educación.

Fuente: Elaboración propia.

En la Figura 31 se observa la ejecución por Sector Público Descentralizado en la cual se evidencia que, en los años 2011 y 2012 el porcentaje de ejecución ha superado el 60% siendo satisfactorios. No obstante, también denota que en 2013 y 2020 es en donde hubo menor ejecución, con valores menores al 30%, clasificándose como una ejecución deficiente. Para visualizar la clasificación ver Figura 29.

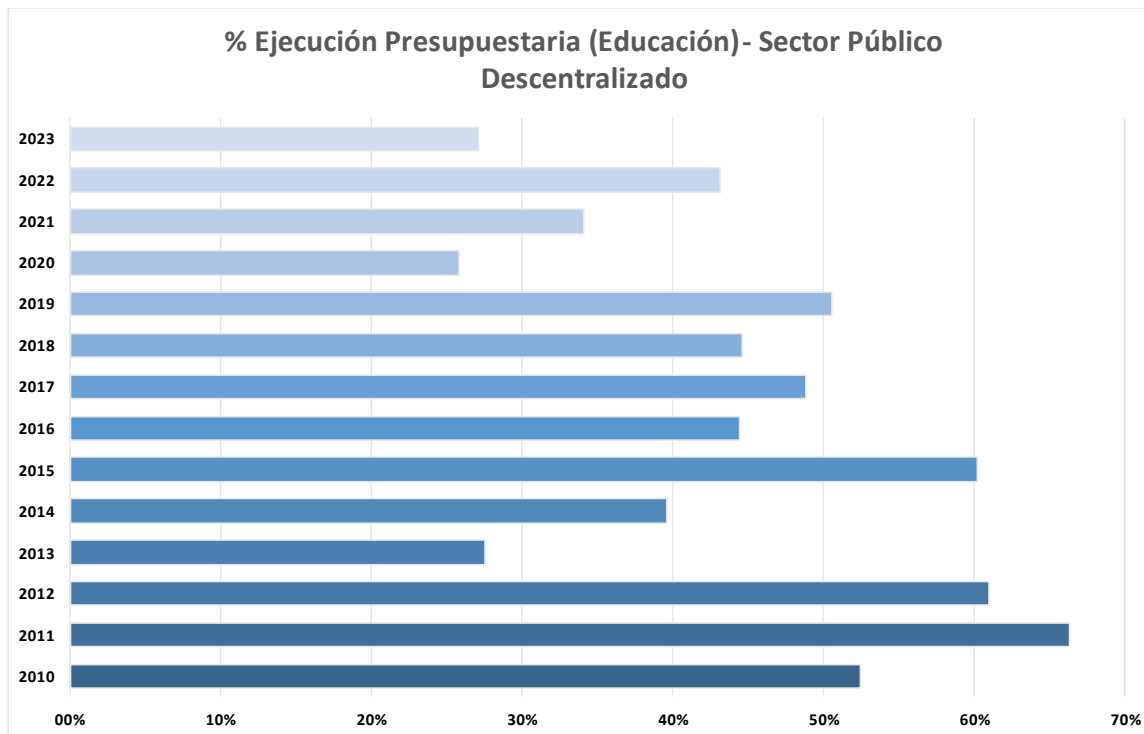


Figura 31. % de ejecución presupuestaria por Sector Público Descentralizado - Educación.
 Fuente: Elaboración propia.

4.2.1.2 Sector salud

En cuanto a ejecución presupuestaria a lo largo de los últimos años, en la Figura 32 se logra observar que la ejecución ha ido variando, mostrando una disminución en 2018 y 2019 en comparación a los demás años. Además, se observa que en los últimos cinco años el presupuesto ha aumentado sustancialmente. Es importante aclarar que para 2023 se cuentan datos hasta septiembre, es por ello por lo que se ve una baja para dicho año.

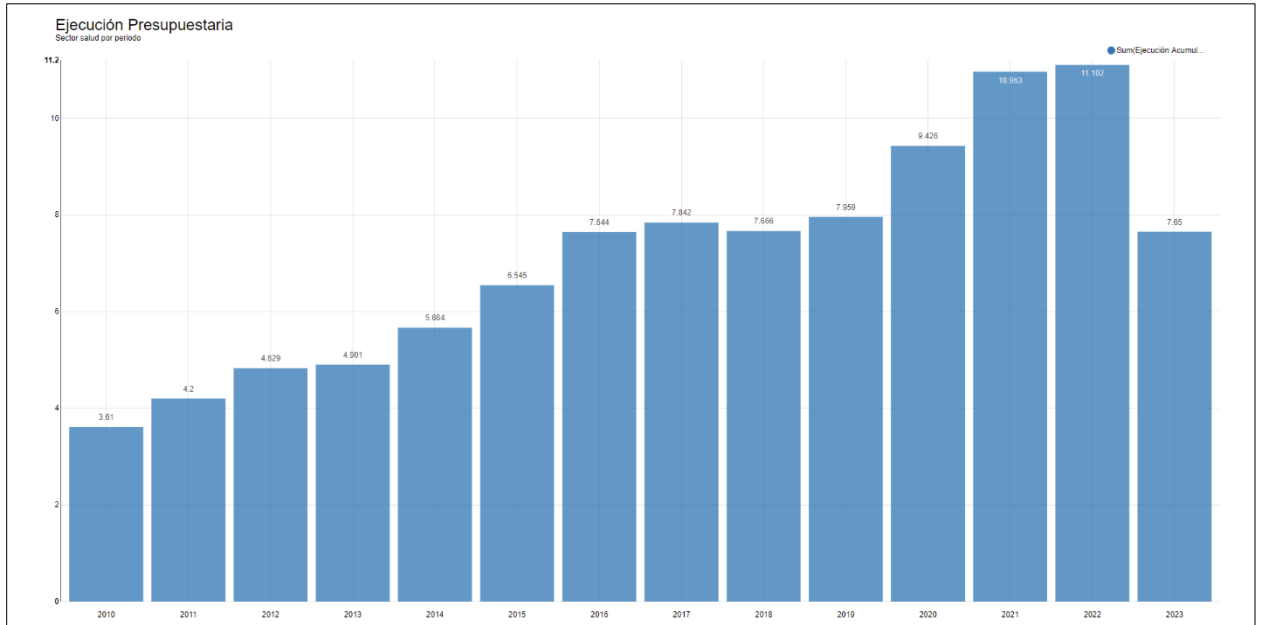


Figura 32. Ejecución presupuestaria del sector salud por año.

Fuente: Elaboración propia.

En la Figura 33 se observa una serie obtenida de la ejecución presupuestaria a lo largo de los períodos 2010 a 2023, mostrándose ciclos que han ido incrementando de acuerdo con el presupuesto correspondiente a cada año.

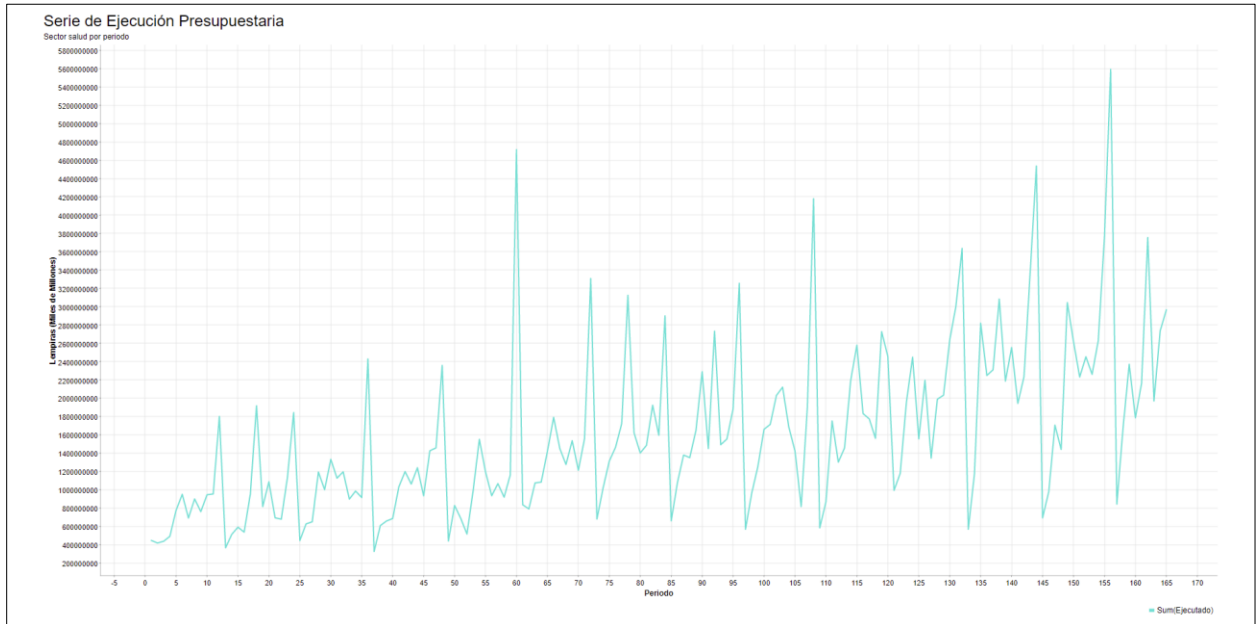


Figura 33. Serie de ejecución presupuestaria del sector salud por período.

Fuente: Elaboración propia.

El objeto de gasto representa a los bienes, servicios y transferencias que el estado ejecuta en la gestión presupuestaria, por lo que en la Figura 34 se logra observar que los objetos de gasto en los que se ha orientado el presupuesto en gran medida son sueldos básicos (30.41%), productos farmacéuticos (7.03%), servicios médicos (3.46%) y contribuciones al INJUPEMP (3.34%).

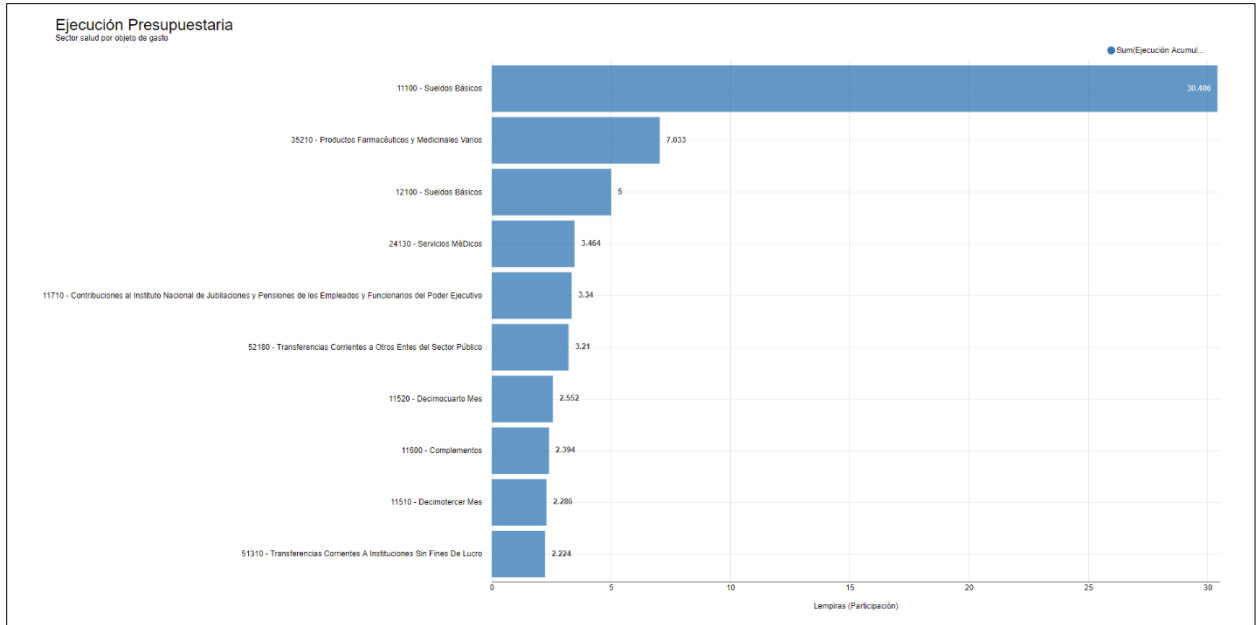


Figura 34. Ejecución presupuestaria del sector salud por objeto de gasto.

Fuente: Elaboración propia.

Por su parte, el grupo de gasto asocia la serie de objetos de gasto que el estado ejecuta en la gestión presupuestaria. En la Figura 35 se observa que los servicios personales (relacionados con los sueldos y salarios) abarcan el 51.94%, mientras que los materiales y suministros (15.67%) y las transferencias y donaciones (16.79%) son los grupos en los que se concentra principalmente el presupuesto.

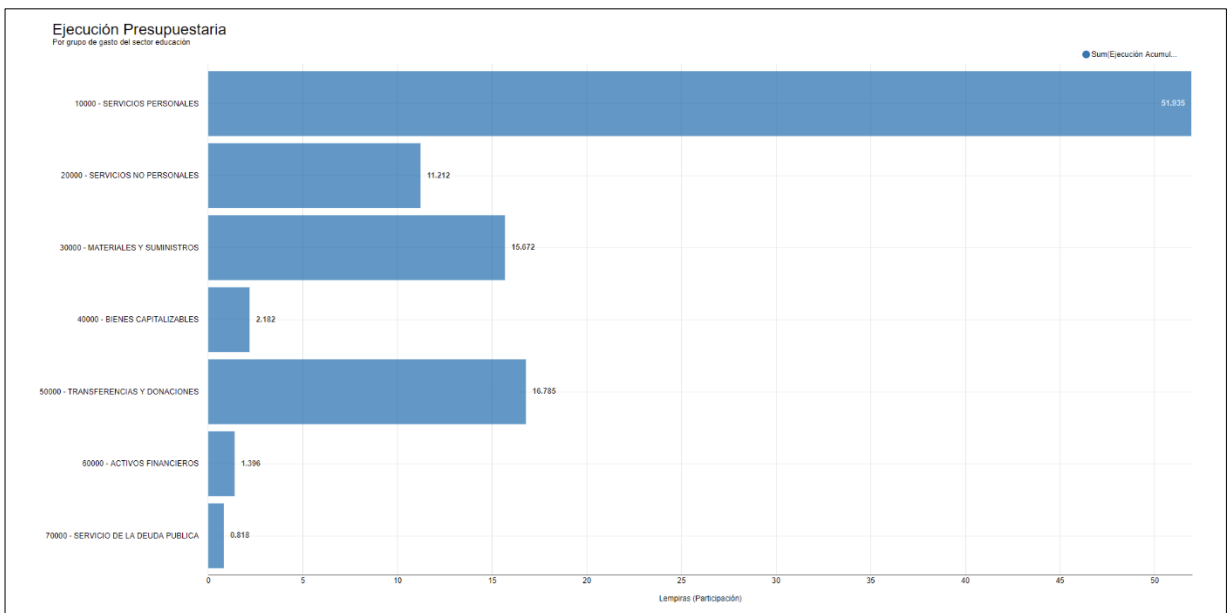


Figura 35. Ejecución presupuestaria del sector salud por grupo de gasto.

Fuente: Elaboración propia.

Además, en la Figura 36 se realiza el cruce de la ejecución presupuestaria por institución, resaltando que la Secretaría de Salud se lleva la gran parte debido a que esta es la encargada de centros de salud y demás (71.59%), seguido del Instituto Hondureño de Seguridad Social (IHSS) que tiene el 20.73% y el Hospital Escuela Universitario con un 3.58%.

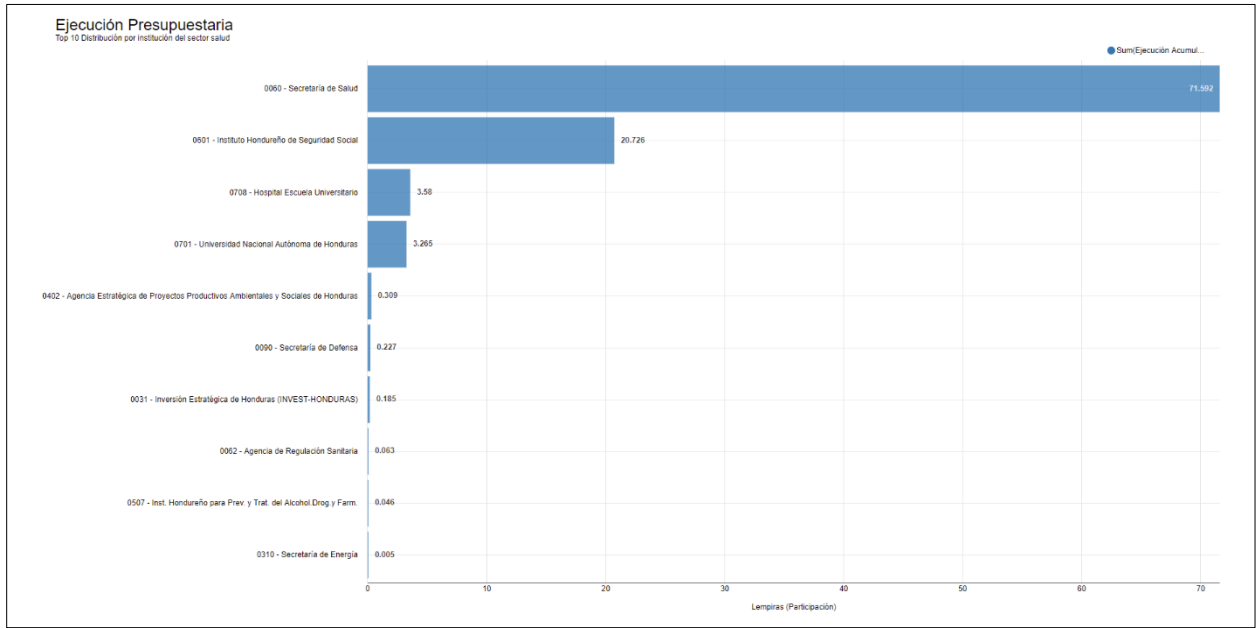


Figura 36. Ejecución presupuestaria del sector salud por institución.
Fuente: Elaboración propia.

La actividad u obra representan los lugares/razones en donde se realizan los productos finales del sector. En la Figura 37 se observa que dirección y coordinación son los que más presupuesto utilizan (31.39%), seguido de los servicios administrativos y contables (8.49%), consultas médicas (7.63%) y la atención hospitalaria (6.21%).

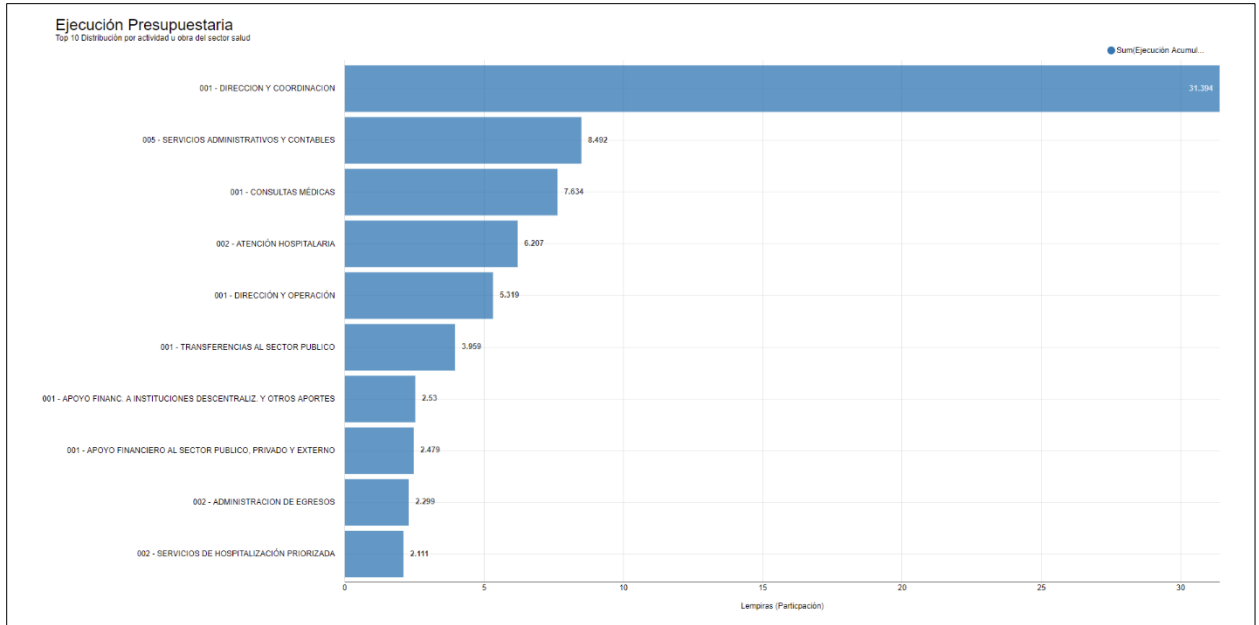


Figura 37. Ejecución presupuestaria del sector salud por actividad/obra.
Fuente: Elaboración propia.

Y para la función de la ejecución presupuestaria del sector salud se observa que gran parte del presupuesto va destinado a servicios de atención médica hospitalaria (50.29%), seguido de servicios de salud (31.48%) y la promoción y prevención de la salud con un 18.23%. Lo anterior se observa en la Figura 38.

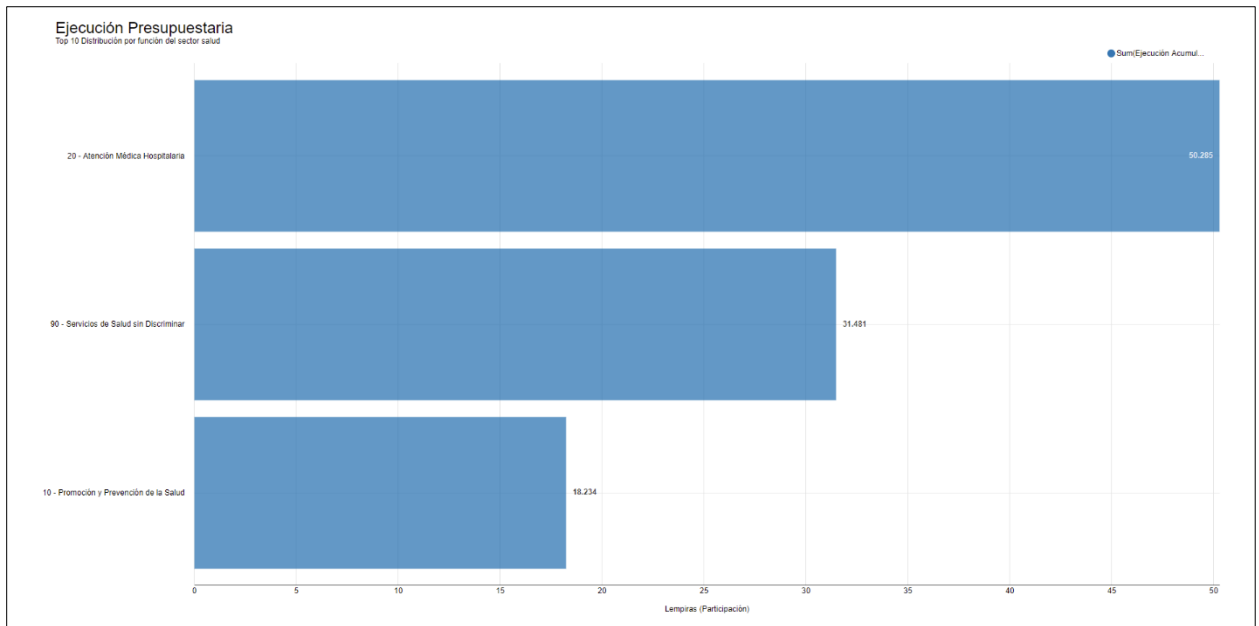


Figura 38. Ejecución presupuestaria del sector salud por función.
Fuente: Elaboración propia.

Es importante mencionar el porcentaje de ejecución presupuestaria a lo largo de los años, ya que a partir de este se logra observar la eficacia. En la Figura 39 se visualiza la ejecución por Administración Central en la cual se observa que, en los años 2011, 2015, 2017, 2020 y 2021 el porcentaje de ejecución ha superado el 70% siendo clasificado como satisfactorio. No obstante, también denota que en 2013 es en donde hubo menor ejecución, con valores menores al 30% clasificándose como una ejecución deficiente. Para visualizar la clasificación ver Figura 29.

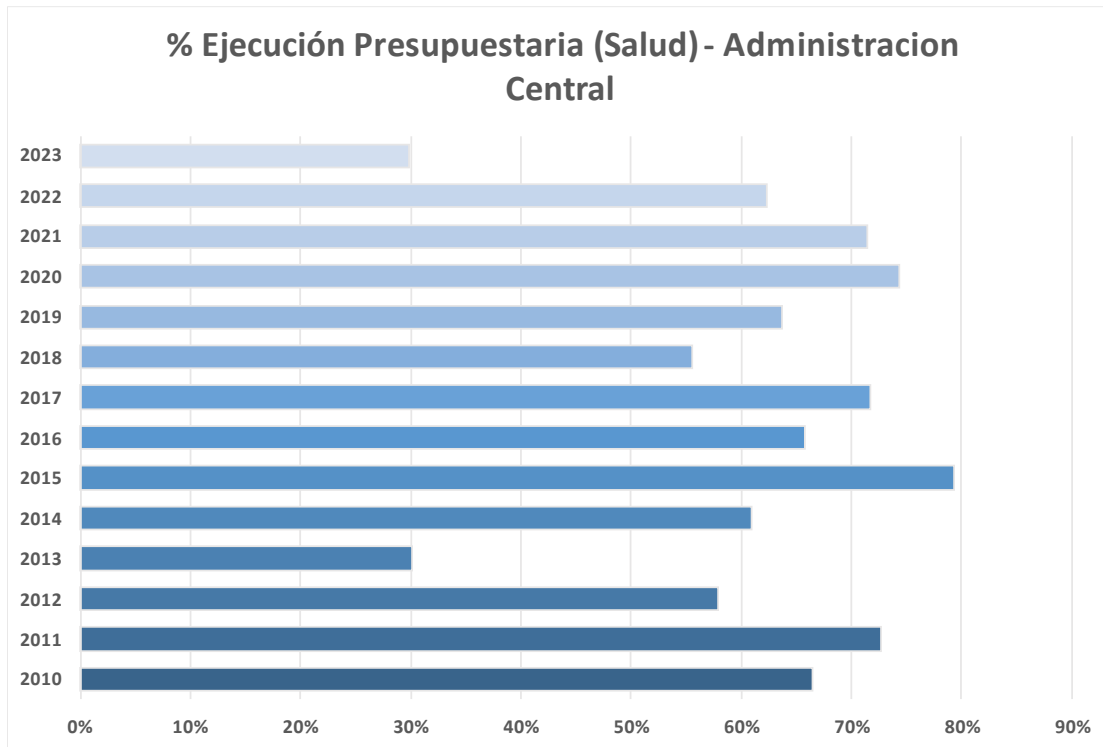


Figura 39. % de ejecución presupuestaria por Admón. Central - Salud.
Fuente: Elaboración propia.

En la Figura 40 se observa la ejecución por Sector Público Descentralizado en la cual se evidencia que, en los años 2017, 2018, 2021 y 2022 el porcentaje de ejecución ha superado el 80% clasificándose como ejecución satisfactoria. No obstante, también denota que, en 2010 y 2011 es en donde hubo menor ejecución, con valores menores al 39% clasificándose como una ejecución deficiente. Para visualizar la clasificación ver Figura 29.

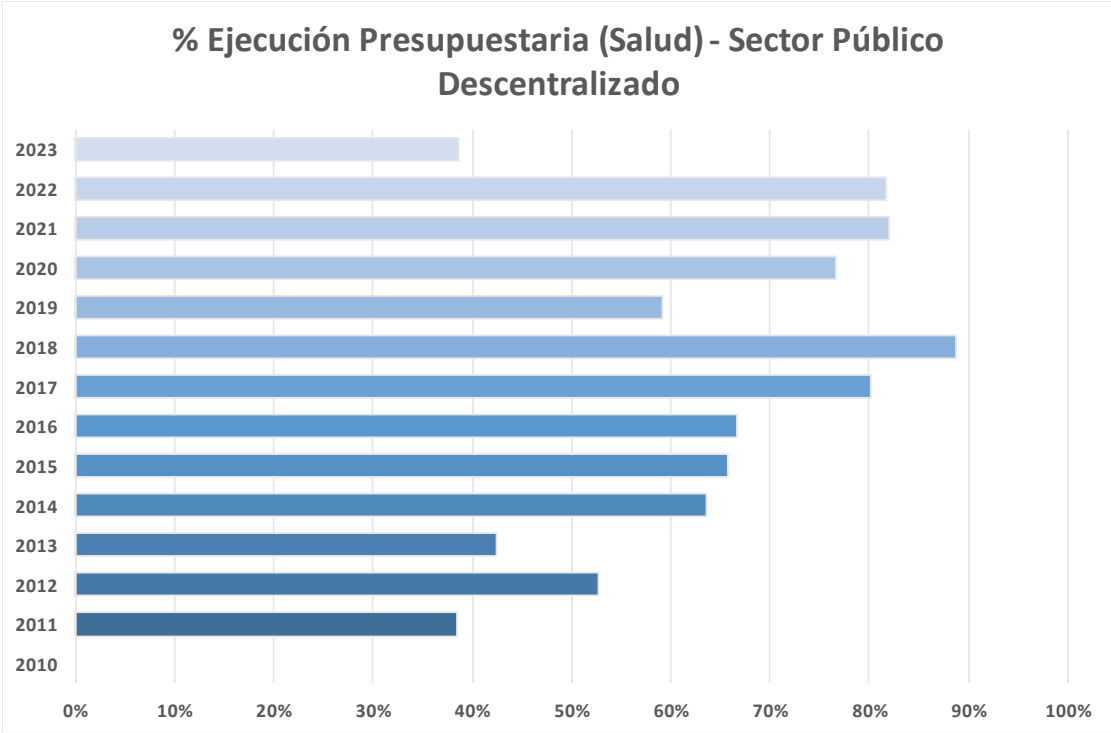


Figura 40. % de ejecución presupuestaria por Sector Público Descentralizado - Salud.
 Fuente: Elaboración propia.

4.2.2 ANÁLISIS CUALITATIVO

En análisis cualitativo comprende el proceso en el que los investigadores organizan e interpretan la información recolectada con el fin de dar respuesta a la pregunta de investigación, esto con el fin de establecer las relaciones existentes, lograr establecer hallazgos y conclusiones a partir de la interpretación. El dato es clasificado como “Cualitativo” cuando está compuesto por archivos de audio, video o transcripciones. (Rodríguez y otros, 2005)

En este estudio al tratarse de un análisis haciendo uso de datos abiertos, no se cuenta con este segmento.

4.3 ANÁLISIS INFERENCIAL Y MODELOS APLICADOS

Debido a que la estadística y los análisis de los datos van más allá de la descripción de las distribuciones de las variables, la generalización de resultados y la comprobación de los supuestos requieren de estimaciones a partir de los estadígrafos, evidenciando con ello, las relaciones y probabilidades de ocurrencias de los eventos en distribuciones normales y no, para a partir de métodos paramétricos y no paramétricos, aceptar o rechazar las hipótesis de la investigación. (Sampieri y otros, Metodología de la Investigación, 2010)

4.3.1 SECTOR EDUCACIÓN

4.3.1.1 *Series temporales*

En primera instancia para la predicción del presupuesto se utilizó el algoritmo ARIMA correspondiente al modelo de analítica denominado Descomposición de series temporales.

La descomposición de series temporales es un método que consiste en tomar una serie de tiempo definida a través de varios periodos y descomponerla ya sea en tendencia, estacionalidad, ciclos o residuos. Una vez se realiza la acción descrita anteriormente, se procede a determinar si es aditiva o multiplicativa, resultando un modelo ideal en la realización de predicciones. Dentro de esta técnica resalta la tendencia debido a que es a través de esta que se logra visualizar un comportamiento a largo plazo del conjunto de datos analizados. (Pepió, 2009)

En la Figura 41 se observa un resumen de la estacionalidad. En la sección “Original” se ve el comportamiento regular del conjunto de datos. En “Trend” se evidencia la tendencia que en este caso es multiplicativa. Por su lado en el bloque “Seasonal” es posible identificar los factores estacionales, siendo en este ejercicio un factor estacional anual muy marcado. Finalmente se

observa el “Error residual” que se caracteriza por alejarse del 0 significativamente en las estacionalidades iniciales de la serie.



Figura 41. Descomposición de la Serie Temporal - Sector Educación.

Fuente: Elaboración propia.

En la Figura 42 se muestra la configuración y ejecución del mejor modelo de predicción ARIMA, indicando la data de entrenamiento y en donde:

- p es el orden del componente autorregresivo.
- d es el orden de diferenciación.
- q es el orden del componente de media móvil.

```

M modelo_auto=auto_arima(train_data,start_p=0,d=1,start_q=0,
    max_p=4,max_d=2,max_q=4, start_P=0,
    D=1, start_Q=0, max_P=2,max_D=1,
    max_Q=2, m=12, seasonal=True,
    error_action='warn',trace=True,
    suppress_warnings=True,stepwise=True,
    random_state=20,n_fits=100)
print(modelo_auto)

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[12] : AIC=470.689, Time=0.46 sec
ARIMA(1,1,0)(1,1,0)[12] : AIC=349.338, Time=0.13 sec
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Time=0.58 sec
ARIMA(1,1,0)(0,1,0)[12] : AIC=390.989, Time=0.04 sec
ARIMA(1,1,0)(2,1,0)[12] : AIC=346.758, Time=0.25 sec
ARIMA(1,1,0)(2,1,1)[12] : AIC=337.427, Time=0.60 sec
ARIMA(1,1,0)(1,1,1)[12] : AIC=336.288, Time=0.24 sec
ARIMA(1,1,0)(0,1,1)[12] : AIC=336.164, Time=0.19 sec
ARIMA(1,1,0)(0,1,2)[12] : AIC=336.673, Time=0.34 sec
ARIMA(1,1,0)(1,1,2)[12] : AIC=336.827, Time=0.65 sec
ARIMA(0,1,0)(0,1,1)[12] : AIC=inf, Time=0.23 sec
ARIMA(2,1,0)(0,1,1)[12] : AIC=329.939, Time=0.39 sec
ARIMA(2,1,0)(0,1,0)[12] : AIC=384.912, Time=0.10 sec
ARIMA(2,1,0)(1,1,1)[12] : AIC=330.373, Time=0.46 sec
ARIMA(2,1,0)(0,1,2)[12] : AIC=330.673, Time=0.49 sec
ARIMA(2,1,0)(1,1,0)[12] : AIC=344.316, Time=0.12 sec
ARIMA(2,1,0)(1,1,2)[12] : AIC=331.240, Time=0.79 sec
ARIMA(3,1,0)(0,1,1)[12] : AIC=302.064, Time=0.32 sec
ARIMA(3,1,0)(0,1,0)[12] : AIC=353.912, Time=0.08 sec
ARIMA(3,1,0)(1,1,1)[12] : AIC=302.661, Time=0.44 sec
ARIMA(3,1,0)(0,1,2)[12] : AIC=302.952, Time=0.61 sec
ARIMA(3,1,0)(1,1,0)[12] : AIC=311.235, Time=0.19 sec
ARIMA(3,1,0)(1,1,2)[12] : AIC=303.856, Time=0.87 sec
ARIMA(4,1,0)(0,1,1)[12] : AIC=290.358, Time=0.32 sec
ARIMA(4,1,0)(0,1,0)[12] : AIC=344.287, Time=0.08 sec
ARIMA(4,1,0)(1,1,1)[12] : AIC=291.351, Time=0.52 sec
ARIMA(4,1,0)(0,1,2)[12] : AIC=291.530, Time=0.83 sec
ARIMA(4,1,0)(1,1,0)[12] : AIC=301.925, Time=0.24 sec
ARIMA(4,1,0)(1,1,2)[12] : AIC=292.613, Time=1.26 sec
ARIMA(4,1,1)(0,1,1)[12] : AIC=275.843, Time=0.67 sec
ARIMA(4,1,1)(0,1,0)[12] : AIC=inf, Time=0.44 sec
ARIMA(4,1,1)(1,1,1)[12] : AIC=277.027, Time=0.70 sec
ARIMA(4,1,1)(0,1,2)[12] : AIC=277.182, Time=1.36 sec
ARIMA(4,1,1)(1,1,0)[12] : AIC=285.082, Time=0.57 sec
ARIMA(4,1,1)(1,1,2)[12] : AIC=278.419, Time=1.81 sec
ARIMA(3,1,1)(0,1,1)[12] : AIC=274.511, Time=0.50 sec
ARIMA(3,1,1)(0,1,0)[12] : AIC=inf, Time=0.27 sec
ARIMA(3,1,1)(1,1,1)[12] : AIC=275.750, Time=0.51 sec
ARIMA(3,1,1)(0,1,2)[12] : AIC=275.888, Time=1.28 sec
ARIMA(3,1,1)(1,1,0)[12] : AIC=285.257, Time=0.53 sec
ARIMA(3,1,1)(1,1,2)[12] : AIC=277.149, Time=1.43 sec
ARIMA(2,1,1)(0,1,1)[12] : AIC=282.385, Time=0.69 sec
ARIMA(3,1,2)(0,1,1)[12] : AIC=276.143, Time=0.69 sec
ARIMA(2,1,2)(0,1,1)[12] : AIC=282.080, Time=0.61 sec
ARIMA(4,1,2)(0,1,1)[12] : AIC=inf, Time=1.26 sec
ARIMA(3,1,1)(0,1,1)[12] intercept : AIC=inf, Time=1.02 sec

Best model: ARIMA(3,1,1)(0,1,1)[12]
Total fit time: 26.104 seconds
ARIMA(3,1,1)(0,1,1)[12]

```

Figura 42. Configuración del modelo ARIMA.

Fuente: Elaboración propia.

En cuanto a los resultados, luego de la ejecución de los modelos se obtuvo un R^2 de 0.795 lo que según López (2019) “Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar”, por lo que se intuye que la predicción está siendo efectiva (Ver Tabla 4).

Tabla 4. Resultados del Modelo

Indicador	Valor
R ²	0.795

Fuente: Elaboración propia.

La Figura 43 representa la predicción resultante del modelo SARIMA, en la cual es posible observar que el presupuesto para el sector educación para el período 2024 tiene una tendencia similar a los períodos anteriores.

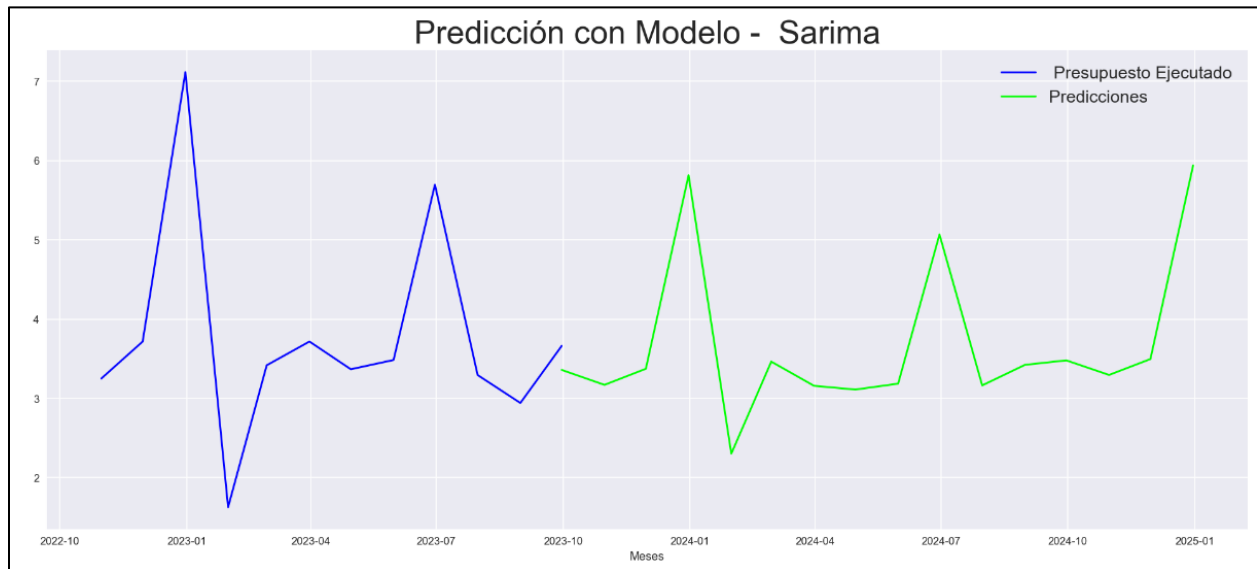


Figura 43. Estimación de presupuesto vía S-ARIMA del sector educación.

Fuente: Elaboración propia.

4.3.1.2 Random Forest

En segunda instancia para la predicción del presupuesto se utilizó el modelo de analítica denominado Random Forest.

El algoritmo de aprendizaje automático denominado Random Forest se utiliza comúnmente en problemas de clasificación y regresión. Una de las características principales de este algoritmo es que está compuesto por un conjunto de árboles de decisión, logrando crear un bosque no correlacionado de estos. Una de las principales diferencias entre Random Forest y Árboles de Decisión es que este selecciona un subconjunto de características en la clasificación, mientras que los árboles las consideran en su totalidad. (IBM, 2023)

En la Figura 44 se visualiza la configuración del Random Forest para efectuar la predicción, en la cual es necesario incluir la cantidad de árboles de decisión a incluir para conformar el bosque,

en este caso se utilizan 400. Además, es necesario indicar la data de entrenamiento y la data de pruebas.

```
# Random Forest

# Creamos el modelo con 400 árboles
rfr = RandomForestRegressor(n_estimators=400)

# Entrenamos el modelo
rfr.fit(train_feature, train_target)

# Hacemos las predicciones
fcst = rfr.predict(test_feature)
```

Figura 44. Configuración del Random Forest.

Fuente: Elaboración propia.

En cuanto a los resultados, luego de la ejecución del modelo se obtuvo un R^2 de 0.877 lo que según López (2019) “Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar”, por lo que se intuye que la predicción está siendo efectiva (Ver Tabla 5).

Tabla 5. Resultados del Modelo

Indicador	Valor
R^2	0.877

Fuente: Elaboración propia.

Dicho lo anterior y luego de efectuar la predicción del presupuesto con Random Forest, en la Figura 45 es posible visualizar que los resultados son muy cercanos a los datos actuales, lo que significa que la predicción del presupuesto para el período 2024 en el sector educación está por encima del rango planteado en la hipótesis.

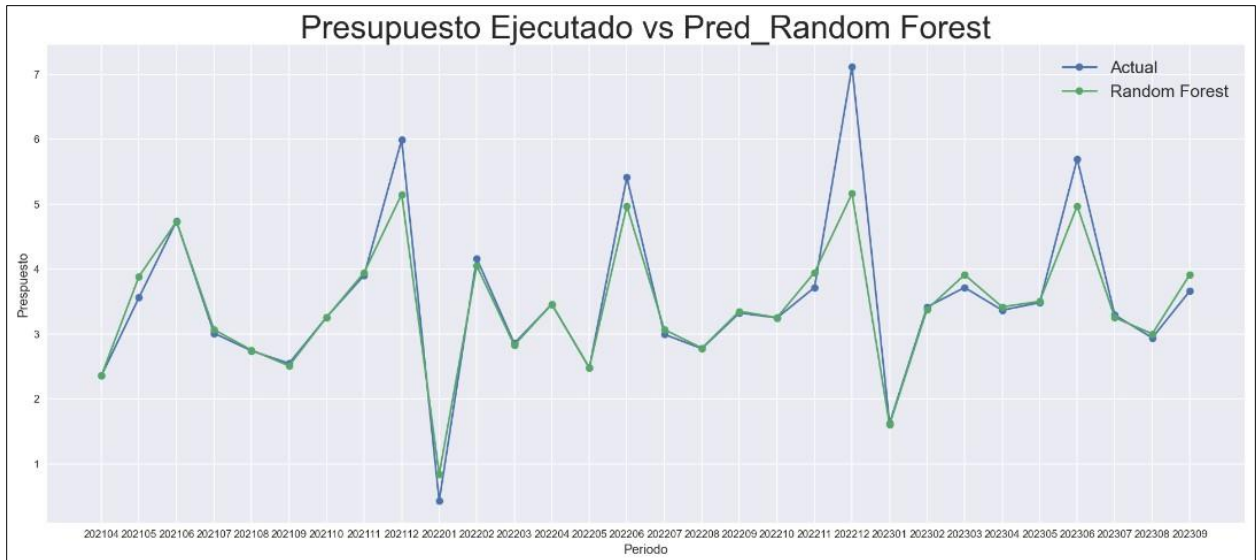


Figura 45. Resultados de la predicción con Random Forest.

4.3.1.2.1 Random Forest (Sin datos de año 2023)

Como segunda prueba para confirmar el comportamiento de los datos y lograr una mayor continuidad de los datos se procedió a aplicar el modelo excluyendo del conjunto de datos la información correspondiente al año 2023.

```
# Random Forest

# Creamos el modelo con 400 árboles
rfr = RandomForestRegressor(n_estimators=400)

# Entrenamos el modelo
rfr.fit(train_feature, train_target)

# Hacemos las predicciones
fcst = rfr.predict(test_feature)
```

Figura 46. Configuración de Random Forest

Fuente: Elaboración Propia.

Una vez separado los datos de entrenamiento y prueba del modelo se proceden a aplicar la técnica de creación de características a partir de los rezagos. Con el cual se logró un comportamiento muy similar al patrón real de la serie.

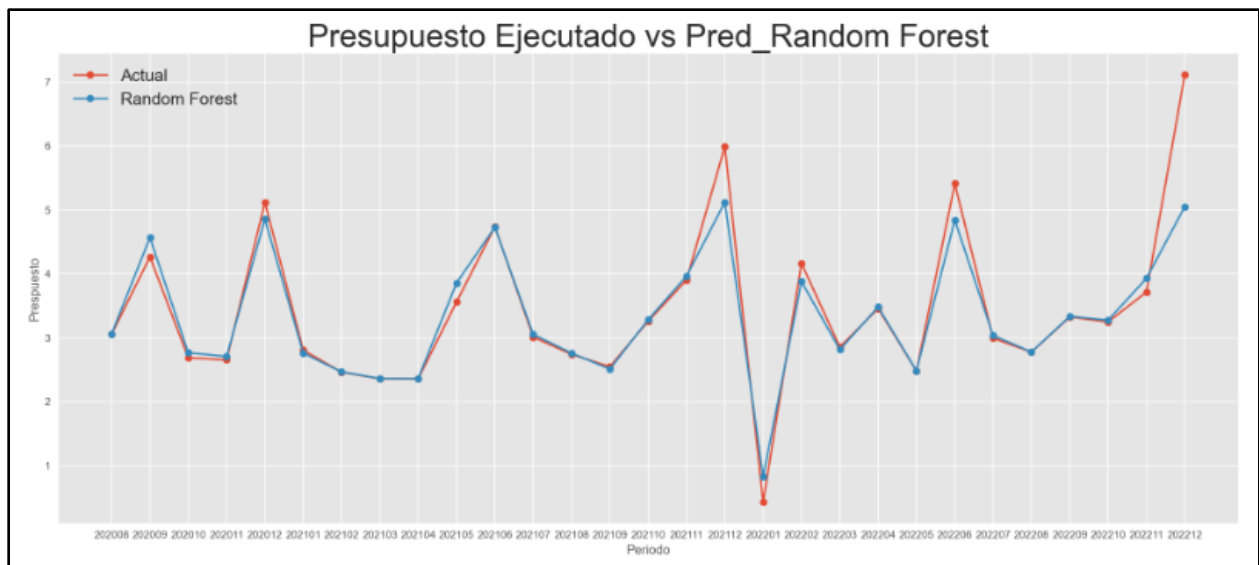


Figura 47. Gráfica de Serie Random Forest (Sin 2023)

Fuente: Elaboración Propia.

Una vez analizado el comportamiento entre cada punto de los periodos evaluados, se procede a generar la validación del modelo por medio de los indicadores, en el cual como se muestra en la Tabla 6 se obtuvo un valor de R2 muy similar y cercano al que se obtuvo tomando en cuenta la información del 2023. Lo cual demuestra que la afectación es mínima al excluir el año 2023 que se encuentra incompleto.

Tabla 6. Resultados de Random Forest (Sin 2023)

Indicador	Valor
R ²	0.875

Fuente: Elaboración Propia

4.3.2 SECTOR SALUD

4.3.2.1 Series temporales

En primera instancia para la predicción del presupuesto se utilizó el algoritmo ARIMA correspondiente al modelo de analítica denominado Descomposición de series temporales.

La descomposición de series temporales es un método que consiste en tomar una serie de tiempo definida a través de varios periodos y descomponerla ya sea en tendencia, estacionalidad, ciclos o residuos. Una vez se realiza la acción descrita anteriormente, se procede a determinar si es aditiva o multiplicativa, resultando un modelo ideal en la realización de predicciones. Dentro de esta técnica resalta la tendencia debido a que es a través de esta que se logra visualizar un

comportamiento a largo plazo del conjunto de datos analizados. (Pepió, 2009)

En la Figura 48 se observa un resumen de la estacionalidad. En la sección “Original” se ve el comportamiento regular del conjunto de datos. En “Trend” se evidencia la tendencia que en este caso es multiplicativa. Por su lado en el bloque “Seasonal” es posible identificar los factores estacionales, siendo en este ejercicio un factor estacional anual muy marcado. Finalmente se observa el “Error residual” que se caracteriza por alejarse del 0 en las estacionalidades.



Figura 48. Resumen de la estacionalidad Sector Salud.

Fuente: Elaboración propia.

En la Figura 49 se muestra la configuración y ejecución del modelo de predicción ARIMA, indicando la data de entrenamiento y en donde:

- p es el orden del componente autorregresivo.
- d es el orden de diferenciación.
- q es el orden del componente de media móvil.

```

ARIMA(1,1,4)(0,1,0)[12] : AIC=inf, Time=0.49 sec
ARIMA(2,1,0)(0,1,0)[12] : AIC=266.785, Time=0.04 sec
ARIMA(2,1,0)(0,1,1)[12] : AIC=222.756, Time=0.18 sec
ARIMA(2,1,0)(0,1,2)[12] : AIC=223.773, Time=0.67 sec
ARIMA(2,1,0)(1,1,0)[12] : AIC=249.143, Time=0.15 sec
ARIMA(2,1,0)(1,1,1)[12] : AIC=224.209, Time=0.31 sec
ARIMA(2,1,0)(1,1,2)[12] : AIC=224.689, Time=0.81 sec
ARIMA(2,1,0)(2,1,0)[12] : AIC=227.370, Time=0.24 sec
ARIMA(2,1,0)(2,1,1)[12] : AIC=221.159, Time=0.72 sec
ARIMA(2,1,1)(0,1,0)[12] : AIC=inf, Time=0.49 sec
ARIMA(2,1,1)(0,1,1)[12] : AIC=inf, Time=0.79 sec
ARIMA(2,1,1)(0,1,2)[12] : AIC=inf, Time=1.93 sec
ARIMA(2,1,1)(1,1,0)[12] : AIC=inf, Time=0.60 sec
ARIMA(2,1,1)(1,1,1)[12] : AIC=inf, Time=0.69 sec
ARIMA(2,1,1)(2,1,0)[12] : AIC=inf, Time=1.35 sec
ARIMA(2,1,2)(0,1,0)[12] : AIC=inf, Time=0.47 sec
ARIMA(2,1,2)(0,1,1)[12] : AIC=inf, Time=0.70 sec
ARIMA(2,1,2)(1,1,0)[12] : AIC=inf, Time=1.21 sec
ARIMA(2,1,3)(0,1,0)[12] : AIC=inf, Time=0.64 sec
ARIMA(3,1,0)(0,1,0)[12] : AIC=262.121, Time=0.05 sec
ARIMA(3,1,0)(0,1,1)[12] : AIC=216.250, Time=0.21 sec
ARIMA(3,1,0)(0,1,2)[12] : AIC=217.563, Time=0.96 sec
ARIMA(3,1,0)(1,1,0)[12] : AIC=241.955, Time=0.14 sec
ARIMA(3,1,0)(1,1,1)[12] : AIC=217.795, Time=0.44 sec
ARIMA(3,1,0)(2,1,0)[12] : AIC=225.430, Time=0.27 sec
ARIMA(3,1,1)(0,1,0)[12] : AIC=inf, Time=0.55 sec
ARIMA(3,1,1)(0,1,1)[12] : AIC=inf, Time=1.33 sec
ARIMA(3,1,1)(1,1,0)[12] : AIC=inf, Time=1.03 sec
ARIMA(3,1,2)(0,1,0)[12] : AIC=inf, Time=0.47 sec
ARIMA(4,1,0)(0,1,0)[12] : AIC=252.592, Time=0.07 sec
ARIMA(4,1,0)(0,1,1)[12] : AIC=207.288, Time=0.40 sec
ARIMA(4,1,0)(1,1,0)[12] : AIC=233.697, Time=0.25 sec
ARIMA(4,1,1)(0,1,0)[12] : AIC=inf, Time=0.39 sec
ARIMA(5,1,0)(0,1,0)[12] : AIC=249.911, Time=0.08 sec

Best model: ARIMA(4,1,0)(0,1,1)[12]
Total fit time: 68.268 seconds
ARIMA(4,1,0)(0,1,1)[12]

```

Figura 49. Configuración del modelo ARIMA.

Fuente: Elaboración propia.

En cuanto a los resultados, luego de la ejecución de los modelos se obtuvo un R^2 de 0.723 lo que según López (2019) “Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar”, por lo que se intuye que la predicción está siendo efectiva (Ver Tabla 6).

Tabla 7. Resultados del Modelo

Indicador	Valor
R^2	0.723

Fuente: Elaboración propia.

La Figura 50 representa la predicción resultante del modelo ARIMA, en la cual es posible observar que el presupuesto para el sector educación para el período 2024 tiene una tendencia similar a los períodos anteriores.

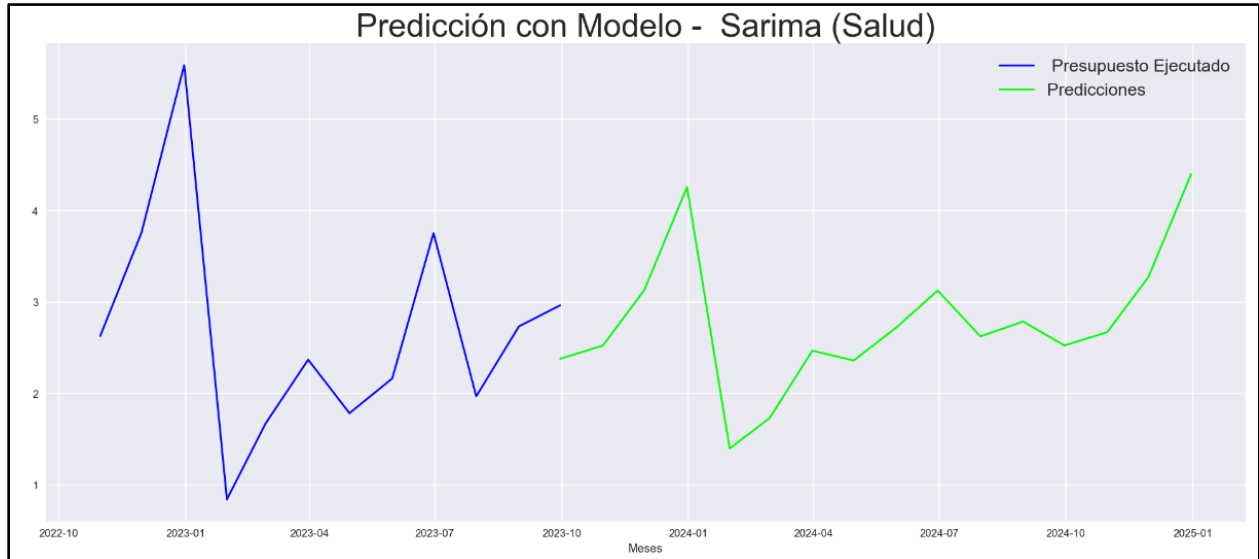


Figura 50. Estimación de presupuesto vía S-ARIMA del sector salud.

Fuente: Elaboración propia.

4.3.2.2 Random Forest

En segunda instancia para la predicción del presupuesto se utilizó el modelo de analítica denominado Random Forest.

El algoritmo de aprendizaje automático denominado Random Forest se utiliza comúnmente en problemas de clasificación y regresión. Una de las características principales de este algoritmo es que está compuesto por un conjunto de árboles de decisión, logrando crear un bosque no correlacionado de estos. Una de las principales diferencias entre Random Forest y Árboles de Decisión es que este selecciona un subconjunto de características en la clasificación, mientras que los árboles las consideran en su totalidad. (IBM, 2023)

En la Figura 51 se visualiza la configuración del Random Forest para efectuar la predicción, en la cual es necesario incluir la cantidad de árboles de decisión a incluir para conformar el bosque, en este caso se utilizan 400. Además, es necesario indicar la data de entrenamiento y la data de pruebas.

```

# Random Forest

# Creamos el modelo con 400 árboles
rfr = RandomForestRegressor(n_estimators=400)

# Entrenamos el modelo
rfr.fit(train_feature, train_target)

# Hacemos las predicciones
fcst = rfr.predict(test_feature)

```

Figura 51. Configuración del Random Forest.

Fuente: Elaboración propia.

En cuanto a los resultados, luego de la ejecución del modelo se obtuvo un R^2 de 0.871 lo que según López (2019) “Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar”, por lo que se intuye que la predicción está siendo efectiva (Ver Tabla 7).

Tabla 8. Resultados del Modelo

Indicador	Valor
R^2	0.871

Fuente: Elaboración propia.

Dicho lo anterior y luego de efectuar la predicción del presupuesto con Random Forest, en la Figura 52 es posible visualizar que los resultados son muy cercanos a los datos actuales, lo que significa que la predicción del presupuesto para el período 2024 en el sector salud está por encima del rango planteado en la hipótesis.

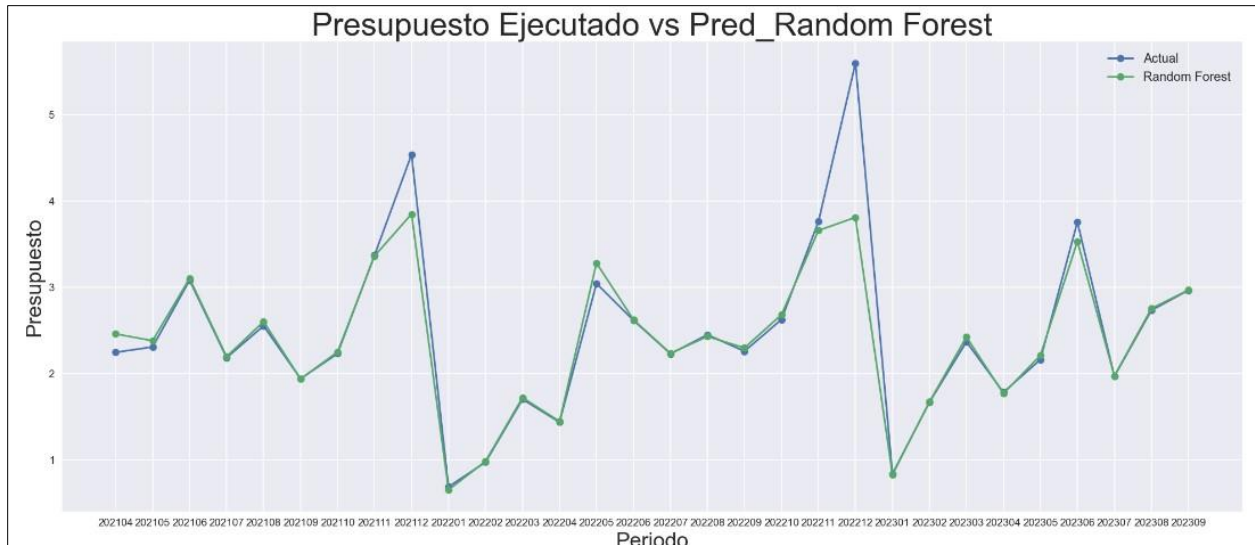


Figura 52. Resultados de la predicción con Random Forest.

Fuente: Elaboración Propia.

4.3.2.3 Random Forest (Sin Datos del año 2023)

En esta sección se realizó un ejercicio aplicando el algoritmo de Random Forest excluyendo de todo el set de datos el año 2023 con el objetivo de validar la afectación o no afectación al resultado debido a la incompletitud de este año.

```
# Random Forest

# Creamos el modelo con 400 árboles
rfr = RandomForestRegressor(n_estimators=400)

# Entrenamos el modelo
rfr.fit(train_feature, train_target)

# Hacemos las predicciones
fcst = rfr.predict(test_feature)
```

Figura 53. Configuración Random Forest (Sin 2023).

Fuente: Elaboración Propia.

Una vez separado los datos de entrenamiento y prueba del modelo se proceden a aplicar la técnica de creación de características a partir de los rezagos. Con el cual se logró un comportamiento muy similar al patrón real de la serie.

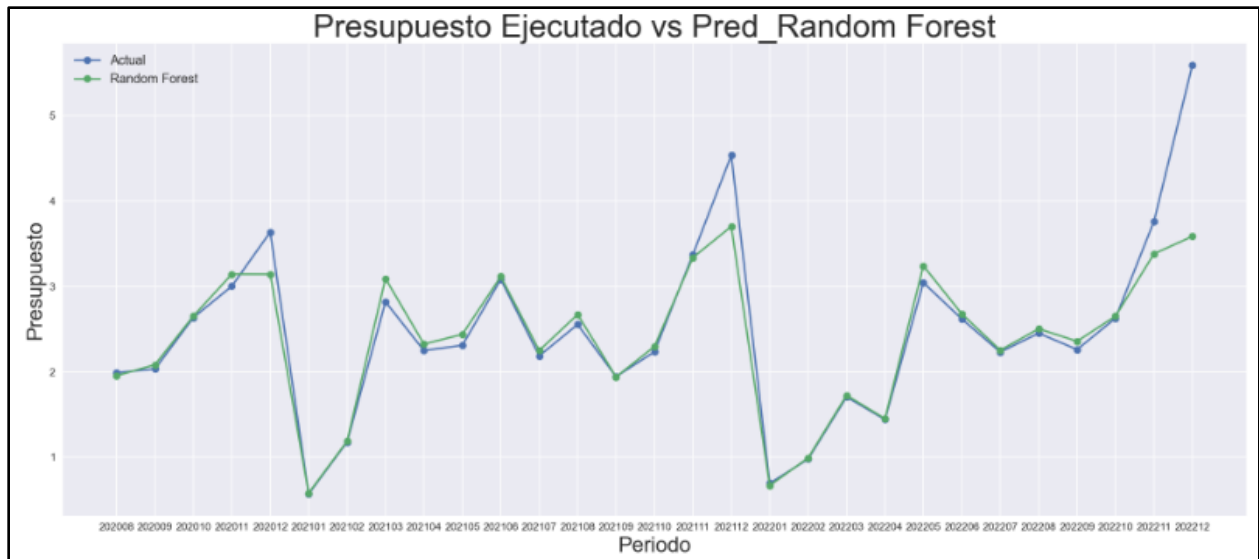


Figura 54. Resultados de Predicción con Random Forest (Sin 2023).

Fuente: Elaboración Propia.

Una vez analizado el comportamiento entre cada punto de los periodos evaluados, se procede a generar la validación del modelo por medio de los indicadores, en el cual como se muestra en la Tabla 9 se obtuvo un valor de R2 muy similar y cercano al que se obtuvo tomando en cuenta la información del 2023. Lo cual demuestra que la afectación es mínima al excluir el año 2023 que se encuentra incompleto.

Tabla 9. Resultados del Modelo (Random Forest) - Sin 2023

Indicador	Valor
R ²	0.836

Fuente: Elaboración Propia.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- Los resultados de esta investigación indican que sí es posible realizar predicciones de presupuesto utilizando datos abiertos como insumo, empleando en primer lugar el modelo SARIMA de series temporales, obteniendo coeficientes de determinación de 0.79 y 0.72 para los sectores de educación y salud respectivamente. Sin embargo, al emplear el algoritmo de Random Forest, se observa una mejora significativa en las predicciones, con coeficientes de determinación de 0.88 y 0.87 para los mismos sectores. Esta mejora se atribuye a la capacidad del algoritmo de adaptarse a la complejidad inherente de los datos, como el ruido y la no estacionariedad, factores comunes en las series temporales.
- El análisis revela que existen categorías específicas que tienen una influencia significativa en la ejecución presupuestaria de cada sector. En el sector educación, los objetos de gasto más relevantes incluyen los sueldos básicos de docentes (25.75%), los adicionales (17.12%), los sueldos básicos (10.90%) y las transferencias corrientes a universidades (7.95%). Por otro lado, en el sector salud, se destacan los sueldos básicos (30.41%), los productos farmacéuticos (7.03%), los servicios médicos (3.46%), y las contribuciones al INJUPEMP (3.34%). Estos hallazgos resaltan la importancia de gestionar cuidadosamente estos objetos de gasto para mejorar la eficiencia en la ejecución presupuestaria de ambos sectores.
- En función de determinar el nivel de eficiencia en la ejecución del presupuesto de los sectores de salud y educación, el análisis efectuado muestra resultados variables en diferentes niveles de administración. En el sector educación, a nivel de administración central, se observa un nivel satisfactorio de eficiencia, con un desempeño superior al 85%. Sin embargo, en el sector público descentralizado, los resultados son más variables, con niveles que oscilan entre adecuados y satisfactorios, superiores al 40%, aunque se destacan los años 2013 y 2020 con porcentajes deficientes, menores al 39%. En el sector salud, a nivel de administración central, se evidencia un nivel satisfactorio de eficiencia, con un desempeño superior al 70%. De forma similar, en el sector público descentralizado, se observan resultados satisfactorios, superiores al 80%. No obstante, se exceptúan de lo

anterior los años 2010, 2011 y 2013 con porcentajes deficientes.

- El presupuesto generado a través de la predicción oscila a L 34,523,479,118.00 para el sector educación y L 25,039,700,912.00 para el sector salud, por lo que al realizar la comparación con la publicación de SEFIN correspondiente a 2024 se tiene un porcentaje de efectividad del 64% y 61% respectivamente.

5.2 RECOMENDACIONES

- Para SEFIN, se recomienda realizar una reestructuración en el registro de datos correspondientes a la ejecución presupuestaria, ya que, actualmente se encontró que hay descripciones de campos que se duplican, llevan el mismo nombre, que agregan preposiciones de más, tienen leyendas poco descriptivas o ambiguas (por ejemplo, en el sector educación existe la función “Educación, investigación, cultura y actividades recreativas”), etc. Lo anterior se sugiere para que exista mayor claridad en los datos abiertos que sirven de insumo para investigaciones y rendición de cuentas, logrando mayor transparencia y calidad en la ejecución.
- En la aplicación del modelo SARIMA, se sugiere que se apliquen más iteraciones que las desarrolladas en este estudio, esto con el fin de obtener mayor precisión en un futuro, ya que así se otorgaría más información y por ende mejores resultados.
- Se sugiere que para la implementación de este modelo y al tratarse de un tema sumamente sensible como los presupuestos, se diseñe posteriormente a este estudio un modelo de ajuste considerando variables adicionales que los expertos en el tema consideren necesarias para alcanzar mayor precisión en las métricas resultantes.
- Se recomienda a los maestrantes e investigadores que continúen realizando análisis de los datos abiertos de SEFIN con el fin de seguir contribuyendo a encontrar mejoras y así brindar productos que sean de utilidad a la sociedad.

CAPÍTULO VI. APLICABILIDAD

6.1 NOMBRE DE LA PROPUESTA

“Aplicación de un Modelo de Analítica para Predecir el Presupuesto en los Sectores de Salud y Educación en Honduras a 2024”.

6.2 JUSTIFICACIÓN DE LA PROPUESTA

La aplicación de un modelo de analítica en un ámbito tan importante como lo es la ejecución presupuestaria en sectores clave para el desarrollo de cualquier país como lo son la educación y la salud resulta una opción viable haciendo uso del auge que está teniendo la ciencia de datos hoy en día. Estos modelos brindan la capacidad de realizar una predicción de presupuestos con un rango aceptable de métricas, posibilitando una mejor gestión de los objetos de gasto en los cuales se subdividen los montos, lo que a su vez también permite identificar con gran precisión esos objetos de gasto que representan los factores críticos al momento de las gestiones.

Otro punto importante es que la aplicación de un modelo de analítica refleja en cierto punto la evaluación de la eficiencia a lo largo de los años, marcando un patrón e indicando que la eficiencia de ejecución de los años posteriores al estudio no debe ser inferior en ninguna medida a dicho patrón. Además, al realizar una comparación con valores correspondientes a 2024 es posible obtener una validación que brinde fiabilidad en cuanto a los resultados de los modelos predictivos. Dicho lo anterior, con la aplicación del modelo predictivo es posible una optimización de recursos presupuestarios, generando enfoques claros y así evitar la sub o sobreutilización de fondos del estado de Honduras.

6.3 ALCANCE DE LA PROPUESTA

El alcance de esta propuesta consiste en proporcionar una guía general en la que se indique cada uno de los pasos correspondientes para que SEFIN o cualquier otra institución que cuente con fuentes de datos relacionados con ejecución presupuestaria puedan aplicar un modelo de analítica (Series Temporales o Random Forest) y que a partir de estos sea posible realizar predicciones precisas con el fin de mejorar el proceso actual haciendo uso de los avances tecnológicos relacionados con el análisis de datos y técnicas de aprendizaje automático. Es importante mencionar que esta sección corresponde únicamente a una propuesta en la que se brindan los insumos necesarios para que investigadores o instituciones efectúen modelos con el

fin de realizar predicciones posteriormente, por lo que queda fuera del alcance cualquier tipo de implementación actual por parte de los investigadores implicados en este estudio.

6.3.1 OBJETIVO GENERAL

Elaborar una guía detallada para que individuos o instituciones puedan crear modelos de aprendizaje automático relacionados con la ejecución presupuestaria.

6.3.2 OBJETIVOS ESPECÍFICOS

- Describir cada uno de los pasos necesarios para la construcción/configuración de los algoritmos de series temporales y random forest en la predicción de presupuestos de los sectores salud y educación.
- Estructurar un cronograma de actividades que evidencie el tiempo y los costos que implicaría la aplicación de un modelo predictivo de presupuestos en ambientes de producción.

6.4 DESCRIPCIÓN Y DESARROLLO

6.4.1 DESCRIPCIÓN

Las series temporales se conforman de una secuencia de N observaciones ordenadas de manera equidistante y cronológicamente, dando lugar a observaciones en intervalos regulares en una escala de tiempo.

Estas son de gran utilidad para diversos campos, desde las finanzas, las operaciones hasta la medicina. Donde se aplican para analizar patrones, tendencias a medida transcurre el tiempo, y lo esencial de esta herramienta es aprender a anticipar cambios en múltiples escenarios, impulsando la capacidad predictiva en un mundo donde los datos son el principal insumo en el proceso de la toma de decisiones.

6.4.2 DESARROLLO

Se realizó un estudio investigativo tomando como referencia el presupuesto ejecutado en el sector salud y educación desde 2010 a Sept de 2023, Con el objetivo de identificar patrones de comportamiento en relación con el tiempo que ha transcurrido e identificar estacionalidades que estén marcadas al pasar de los años.

A continuación, se detallan los pasos a seguir para el seguimiento del presente estudio:

1. Extracción de Conjuntos de Datos

Inicialmente se localizaron las bases del presupuesto ejecutado y aprobado histórico, dentro de la página de SEFIN (<https://www.sefin.gob.hn/ejecucion-presupuestaria/>) y se realizó la extracción de estos utilizando solamente los archivos con el indicados de “Mes” con el valor “0”. Esto debido a que este archivo consolida toda la información de los todos los meses del año indicado.

AÑO	MES	DESCRIPCIÓN	ARCHIVO XLS	ARCHIVO CSV	ARCHIVO JSON	Fecha de Actualización	
2023		11	Ejecución Mensual	Ver	Ver	Ver	04/12/2023
2023		10	Ejecución Mensual	Ver	Ver	Ver	09/10/2023
2023		9	Ejecución Mensual	Ver	Ver	Ver	06/09/2023
2023		8	Ejecución Mensual	Ver	Ver	Ver	06/09/2023
2023		7	Ejecución Mensual	Ver	Ver	Ver	02/08/2023
2023		6	Ejecución Mensual	Ver	Ver	Ver	02/07/2023
2023		5	Ejecución Mensual	Ver	Ver	Ver	16/06/2023
2023		4	Ejecución Mensual	Ver	Ver	Ver	16/05/2023
2023		3	Ejecución Mensual	Ver	Ver	Ver	16/05/2023
2023		2	Ejecución Mensual	Ver	Ver	Ver	16/05/2023

Figura 55. Bases de Presupuesto Ejecutado.

Fuente: Sitio Oficial SEFIN.

2. Consolidación de Conjuntos de Datos

Posteriormente al tener los archivos descargados se consolidan en un solo archivo conteniendo la información de todo el histórico de 2010 a Sept 2023. Se utiliza la herramienta llamada KNIME para realizar esta consolidación.

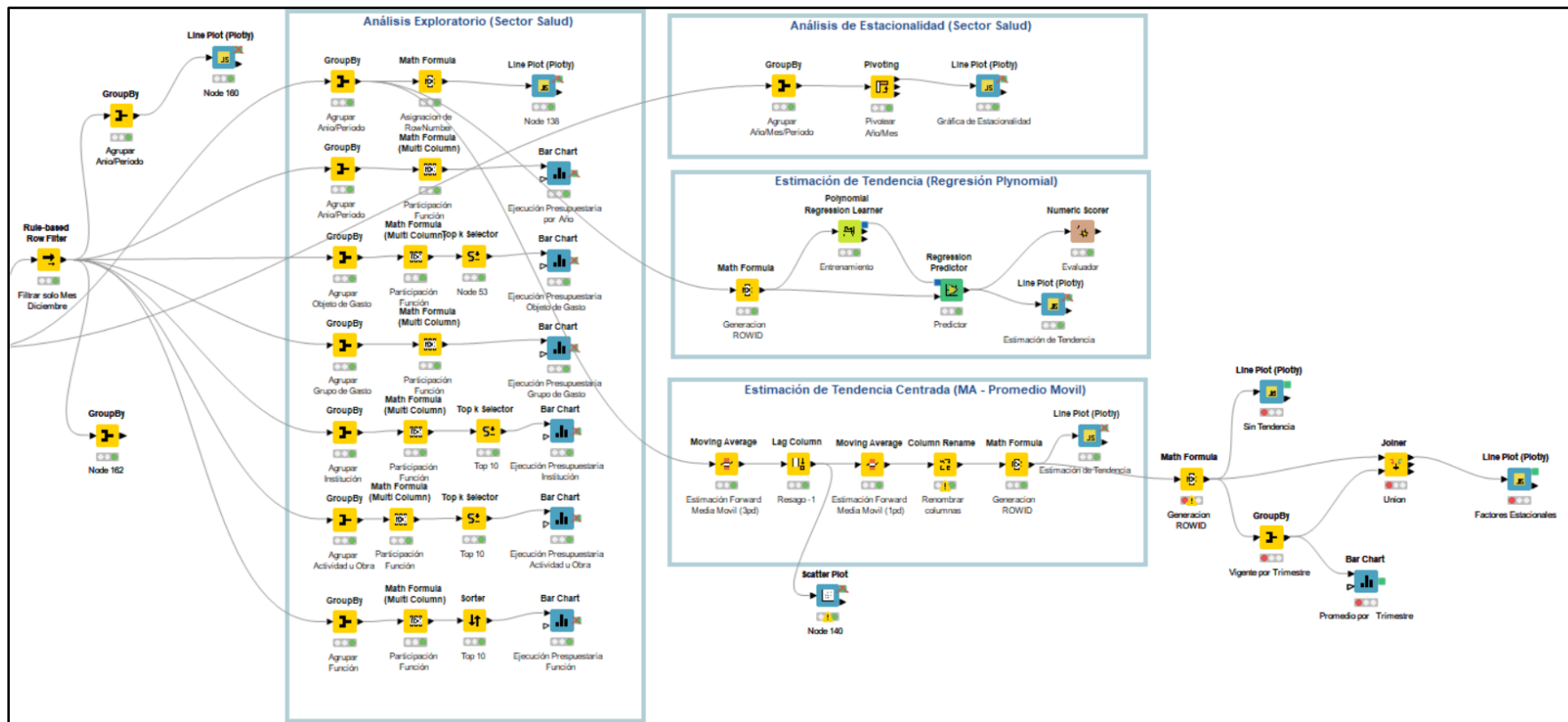


Figura 56. Knime Workflow.

Fuente: Elaboración Propia.

3. Ejecución de “Jupyter Notebook”

Teniendo la información consolidada con todos los años de historia, se procede a construir el notebook (Python) que se utilizará para la elaboración del modelo.

Se utiliza el software llamado “Anaconda” y el modelo de “Jupyter Notebook” como entorno de desarrollo para el lenguaje Python, y se crea un nuevo Notebook.

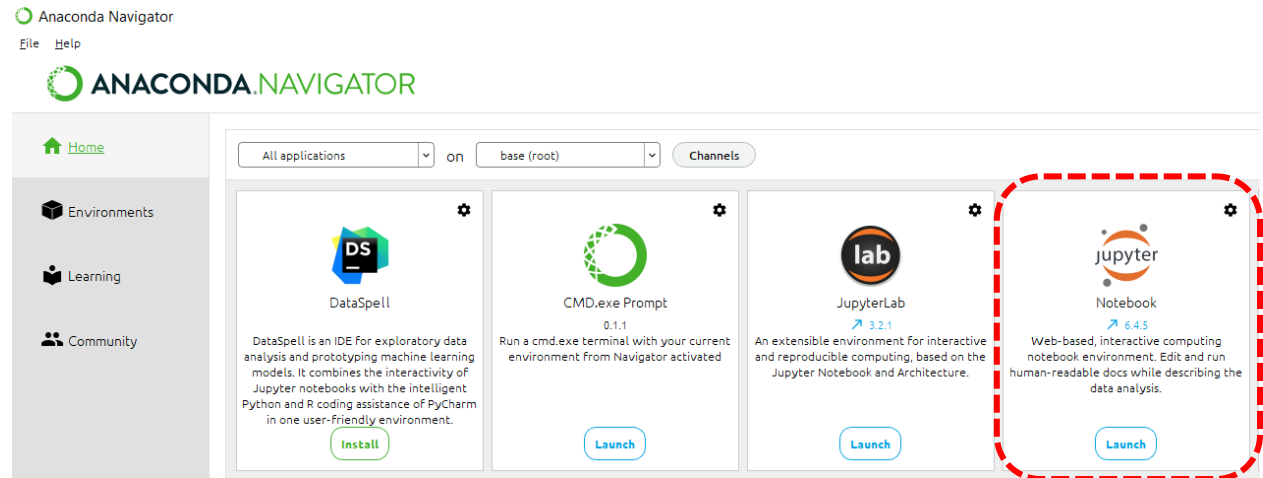


Figura 57. Anaconda Navigator.

Fuente: Elaboración Propia.

6.4.2.1 Modelo ARIMA

1. Importar Librerías

Como paso inicial dentro del Notebook ahora se procede a importar todas las librerías que serán utilizadas en tanto para crear el modelo como para manipulación de los datos.

Librerías utilizadas:

Manipulación y Tratamiento de Datos

- Numpy, Pandas

Visualización de Datos

- plotly.express, matplotlib.pyplot, scipy.stats

Modelado

- statsmodels, pmdarima, sklearn.

Evaluación

- sklearn.metrics, statsmodels.tools.eval_measures, sklearn.metrics.

```

# Manipulación y tratamiento de Datos
import numpy as np
import pandas as pd

# Visualización de datos
#!pip install plotly
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')
from scipy.stats import boxcox

# Modelación Arima
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller

# Métrica de Evaluación
from sklearn.metrics import mean_squared_error
from statsmodels.tools.eval_measures import rmse
from sklearn import metrics

# No presentar advertencias
import warnings
warnings.filterwarnings("ignore")

# Modelo Auto-Arima
!pip3 install pmdarima
#python -m pip install --trusted-host files.pythonhosted.org --
#https://stackoverflow.com/questions/49324802/pip-always-fails-
import pmdarima as pm
from pmdarima import auto_arima

# Target Train-Test split
from pandas import read_csv

# Creando Train test split
from sklearn.model_selection import train_test_split

# Random Forest
from sklearn.ensemble import RandomForestRegressor
from sklearn import metrics

import statsmodels.api as sm
from statsmodels.tsa.stattools import acf
from statsmodels.tsa.arima_model import ARIMA
import statsmodels.tsa.api as smt

```

Figura 58. Librerías Importadas.

Fuente: Elaboración Propia.

2. Creación de función para evaluar modelos

Se crea función con las métricas de evaluación para los modelos que se implementaron en la investigación.

Métricas:

- MSE: Error Cuadrático Medio
- MAE: Error Absoluto Medio
- RMSE: Raíz de Error Cuadrático Medio
- MAPE: Porcentaje de Error Absoluto Medio
- R^2 : Coeficiente de Determinación

```
def evaluacion_metrica(y_true, y_pred):  
  
    def mean_absolute_percentage_error(y_true, y_pred):  
        y_true, y_pred = np.array(y_true), np.array(y_pred)  
        return np.mean(np.abs((y_true - y_pred) / y_true)) * 100  
    print('Evaluation metric results:-')  
    print(f'MSE is : {metrics.mean_squared_error(y_true, y_pred)}')  
    print(f'MAE is : {metrics.mean_absolute_error(y_true, y_pred)}')  
    print(f'RMSE is : {np.sqrt(metrics.mean_squared_error(y_true, y_pred))}')  
    print(f'MAPE is : {mean_absolute_percentage_error(y_true, y_pred)}')  
    print(f'R2 is : {metrics.r2_score(y_true, y_pred)}',end='\n\n')
```

Figura 59. Función de Evaluación de Modelos.

Fuente: Elaboración Propia.

3. Importar Conjunto de datos consolidado en un dataframe

Se Carga el DATASET consolidado en un dataframe conteniendo la información de la ejecución presupuestaria del sector salud y educación desde el 2010 a septiembre de 2023.

```
df = pd.read_csv("DatasetTFGPresupuestoSEFINV3.csv", delimiter = ',',encoding='latin-1')
```

Figura 60. Importar Dataset en un Dataframe.

Fuente: Elaboración Propia.

4. Crear Dataframe por “Finalidad” según el sector a evaluar

Según sea el sector que se está estudiando se debe filtrar el dataframe y crear uno nuevo solamente con la información filtrada que se requiere, en este caso del ejemplo es por “sector salud”.

```
valor_a_buscar = '06 - SERVICIOS DE SALUD'|  
df = df[df['Finalidad'] == valor_a_buscar]  
df.head()
```

Figura 61. Creación de dataframe filtrado por sector.

Fuente: Elaboración Propia.

5. Transformaciones múltiples en variables de tiempo

Para analizar una serie temporal es importante desagregar la variable de tiempo. Es decir, convertir a año, Periodo, Mes, Trimestre, etc. Ya que de esta manera se puede realizar múltiples evaluaciones tomando como referencia los diferentes espacios de tiempo.

```
#Renombrar columna de Gestión ==> Año
df.rename(columns={'Gestión': 'Anio'}, inplace=True)

#Convertir a formato de dos digitos el Mes
df['Mes'] = df['Mes'].astype(str).str.zfill(2)

#Convertir a formato de dos digitos el Mes
df['Anio'] = df['Anio'].astype(str).str.zfill(4)

#concatener el año con el mes para generar el dato de periodo
df['Periodo'] = df['Anio'].astype(str) + df['Mes']
```

Figura 62. Transformación múltiple de datos de tiempo.

Fuente: Elaboración Propia.

6. Agrupación de Conjunto de Datos a nivel de variable “Periodo”.

Es necesario agrupar por la variable que contiene la serie (Periodo) y por la variable que contiene los datos de la ejecución presupuestaria que ha sucedido al pasar del tiempo. Asimismo, se define el “Periodo” como el índice del conjunto de datos dado que es un dato único.

```
#Agrupar el conjunto de datos por Año, Mes y Periodo con La variables "Ejecucion de enero" sumariada
df_1=df.groupby(['Anio','Mes','Periodo'])['Ejecución de enero'].sum().reset_index()

#Columna Periodo se convierte en el indice de La tabla
df_1 = df_1.set_index('Periodo')
```

	Anio	Mes	Ejecución de enero
Periodo			
201001	2010	01	0.445384
201002	2010	02	0.417288
201003	2010	03	0.436342
201004	2010	04	0.490573
201005	2010	05	0.773669

Figura 63. Agrupación de Datos.

Fuente: Elaboración Propia.

7. Graficar Serie Temporal.

Se genera un gráfico de línea con el objetivo de visualizar la serie temporal para obtener las primeras suposiciones sobre los ciclos, la estacionalidad y tendencia.

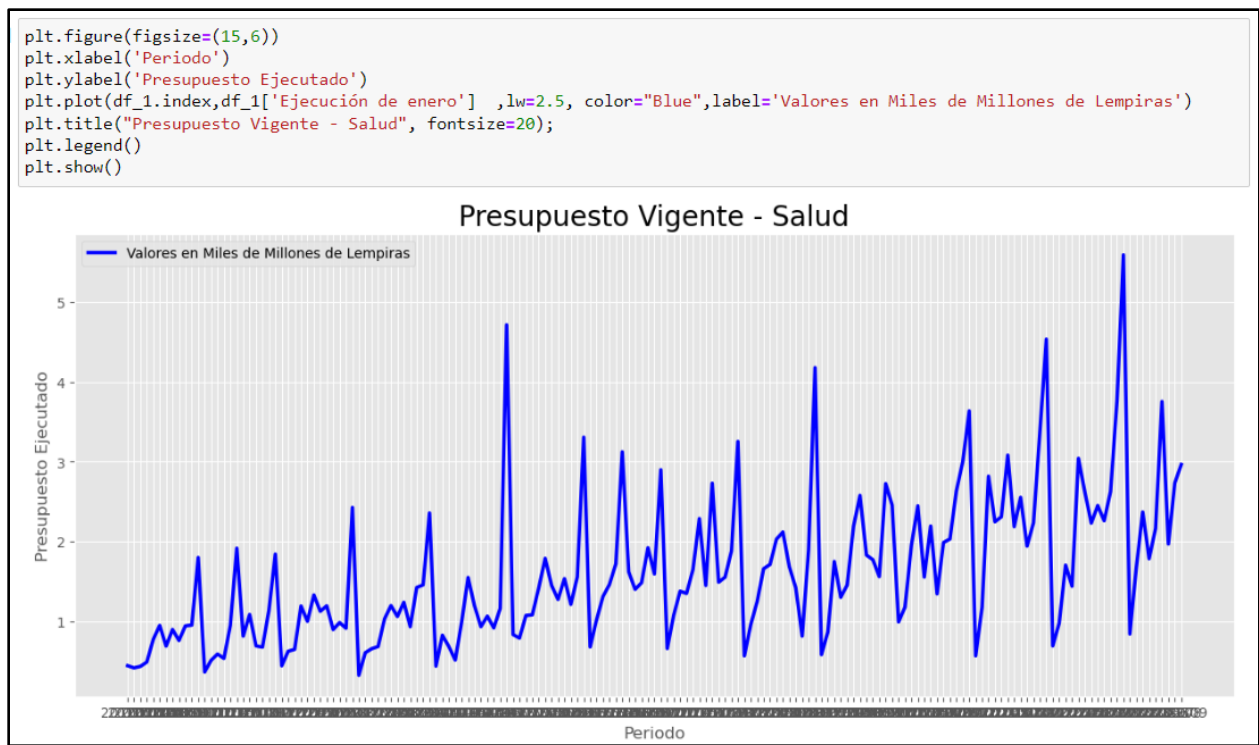


Figura 64. Gráfica de Serie Temporal.
 Fuente: Elaboración Propia.

8. Prueba de ADF “Dickey Fuller Aumentada”.

En los análisis de series temporales es importante identificar ciertos aspectos que ayudarán a elegir el mejor modelo para el pronóstico. En este caso se realizó la prueba ADF con el fin de identificar estacionariedad en la serie, y como se ve a continuación la misma no es estacionaria.

```
Resultados de la prueba de Dickey-Fuller para columna: Ejecución de enero
Test Statistic          0.283527
p-value                 0.976576
No Lags Used            11.000000
Número de observaciones utilizadas  153.000000
Critical Value (1%)    -3.473830
Critical Value (5%)    -2.880623
Critical Value (10%)   -2.576945
dtype: float64
Conclusion:====>
No se puede rechazar la hipótesis nula
Los datos no son estacionarios
```

Figura 65. Prueba de Dickey Fuller Aumentada.
 Fuente: Elaboración Propia.

9. Prueba ADF “Dickey Fuller Aumentada” con diferenciación de primer orden.

En el análisis de series temporales cuando se tiene una NO estacionariedad existe un método llamado “diferenciación” el cual consiste en realizar una serie con sus variaciones o diferencias, esto con el objetivo de lograr una estacionariedad en la misma y poder mejorar la precisión del modelo.

```
Resultados de la prueba de Dickey-Fuller para columna: Ejecución de enero
Test Statistic          -1.399911e+01
p-value                 3.890976e-26
No Lags Used            1.000000e+01
Número de observaciones utilizadas  1.530000e+02
Critical Value (1%)     -3.473830e+00
Critical Value (5%)     -2.880623e+00
Critical Value (10%)    -2.576945e+00
dtype: float64
Conclusion:====>
Rechazar la hipótesis nula
Los datos son estacionarios
```

Figura 66. Prueba Dickey Fuller Aumentada (diff - 1).

Fuente: Elaboración Propia.

10. Autocorrelación con diferenciación de primer orden.

Una vez que se ha analizado la estacionariedad de la serie se realiza un análisis de autocorrelación con el objetivo de identificar la correlación entre en la misma serie con sus rezagos, y como podemos ver en el rezago 0 se logra la autocorrelación de la serie diferenciada en primer orden.

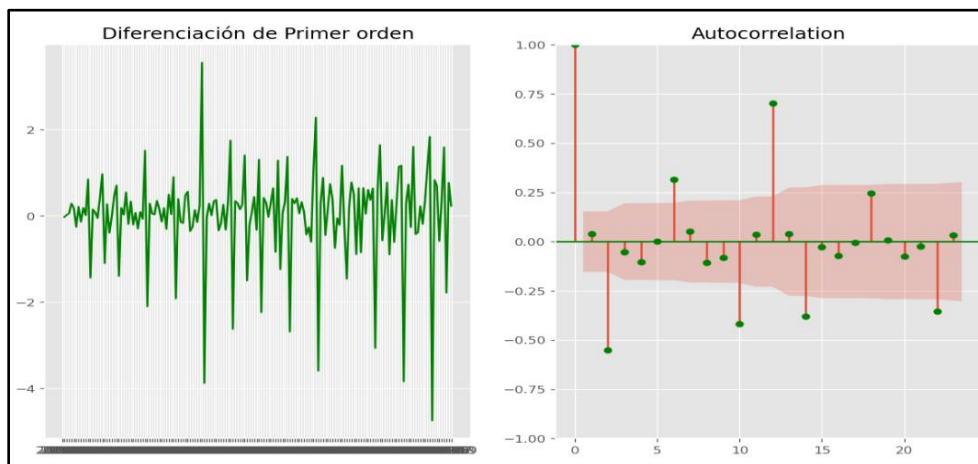


Figura 67. Gráfica de Autocorrelación (diff -1).

Fuente: Elaboración Propia.

11. Descomposición de la Serie Temporal.

Como parte exploratoria de la serie temporal se realiza una descomposición de la serie

temporal donde como se muestra en la Figura 64 se estima la tendencia, se identifican los factores estacionales y se visualizan los errores residuales y su comportamiento.

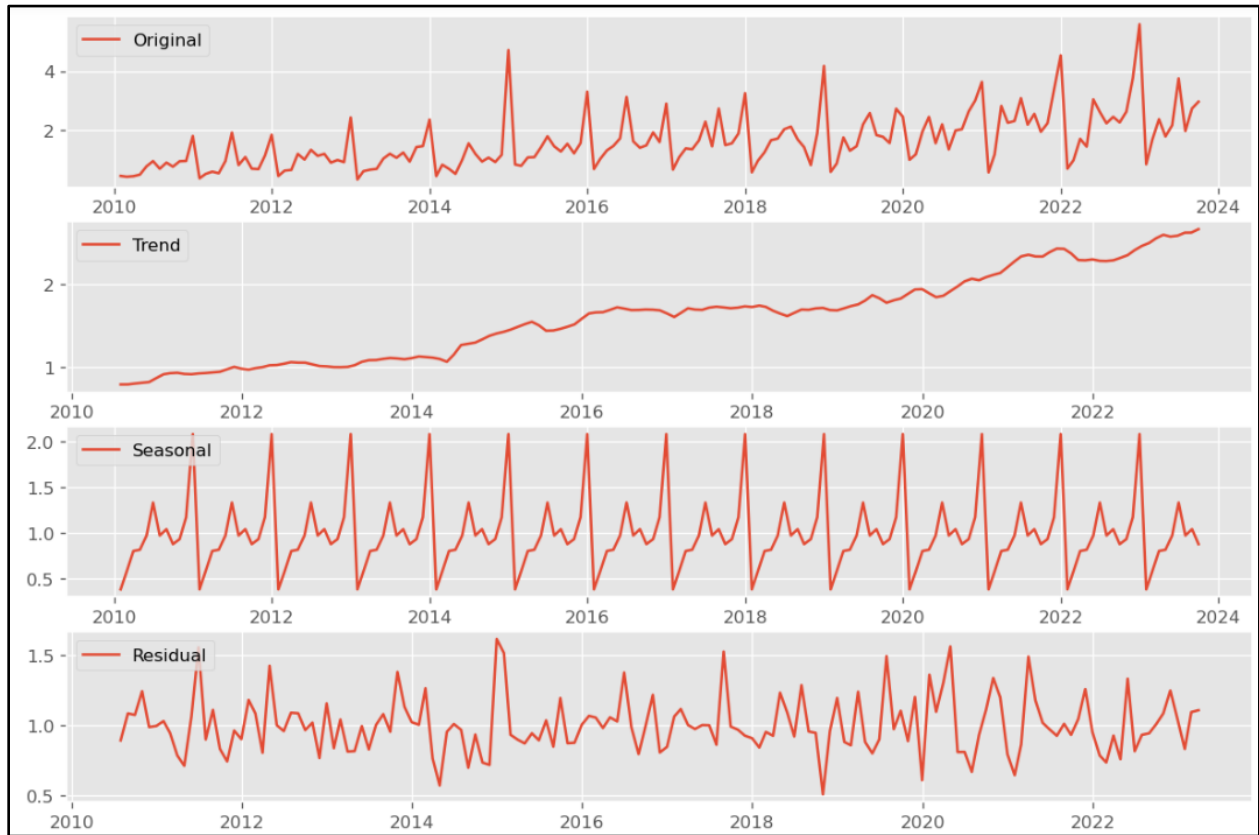


Figura 68. Descomposición de la serie temporal.

Fuente: Elaboración Propia.

12. Modelado: División de datos de Prueba y Entrenamiento.

Para el modelo es necesario dividir el conjunto de datos en 2 particiones, una para entrenamiento y otra para prueba.

```
train_data.shape, test_data.shape
((153, 1), (12, 1))
```

Figura 69. División de datos de prueba y entrenamiento.

Fuente: Elaboración Propia.

13. Generación de parámetros S-ARIMA.

En los pronósticos con modelos S-ARIMA se buscan en las diferentes iteraciones los mejores parámetros de p, d, y q con el fin de identificar el mejor ajuste del modelo para una buena precisión. Se utilizó la función AUTOARIMA la cual consiste en generar múltiples iteraciones de

con diferentes combinaciones de estos parámetros hasta lograr el menor valor de AIC.

```

ARIMA(3,1,1)(0,1,0)[12] : AIC=inf, Time=0.55 sec
ARIMA(3,1,1)(0,1,1)[12] : AIC=inf, Time=1.33 sec
ARIMA(3,1,1)(1,1,0)[12] : AIC=inf, Time=1.03 sec
ARIMA(3,1,2)(0,1,0)[12] : AIC=inf, Time=0.47 sec
ARIMA(4,1,0)(0,1,0)[12] : AIC=252.592, Time=0.07 sec
ARIMA(4,1,0)(0,1,1)[12] : AIC=207.288, Time=0.40 sec
ARIMA(4,1,0)(1,1,0)[12] : AIC=233.697, Time=0.25 sec
ARIMA(4,1,1)(0,1,0)[12] : AIC=inf, Time=0.39 sec
ARIMA(5,1,0)(0,1,0)[12] : AIC=249.911, Time=0.08 sec
Best model: ARIMA(4,1,0)(0,1,1)[12]
Total fit time: 68.268 seconds
ARIMA(4,1,0)(0,1,1)[12]

```

Figura 70. Generación de parámetros ARIMA

Fuente: Elaboración Propia.

14. Implementación del Modelo S-ARIMA.

Una vez se ha identificado el modelo con sus parámetros óptimos, se procede a implementar para generar la función que sirve para proporcionar los valores a predecir del modelo. A continuación, se ilustra el resultado de los valores predichos por el modelo.

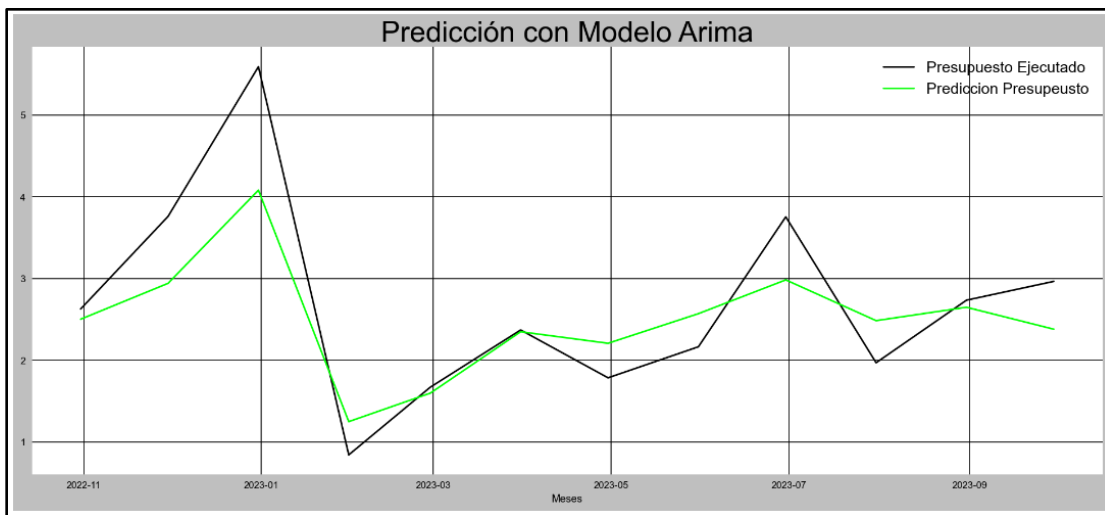


Figura 71. Implementación de Modelo.

Fuente: Elaboración Propia.

15. Predicción con Modelo.

Para confirmar los valores de predicción se realizó una predicción a futuro de 12 meses hacia delante con el fin de identificar el comportamiento de la ejecución presupuestaria al final del 2024, como se muestra en la figura 68 el patrón se mantiene con relación al año 2023.

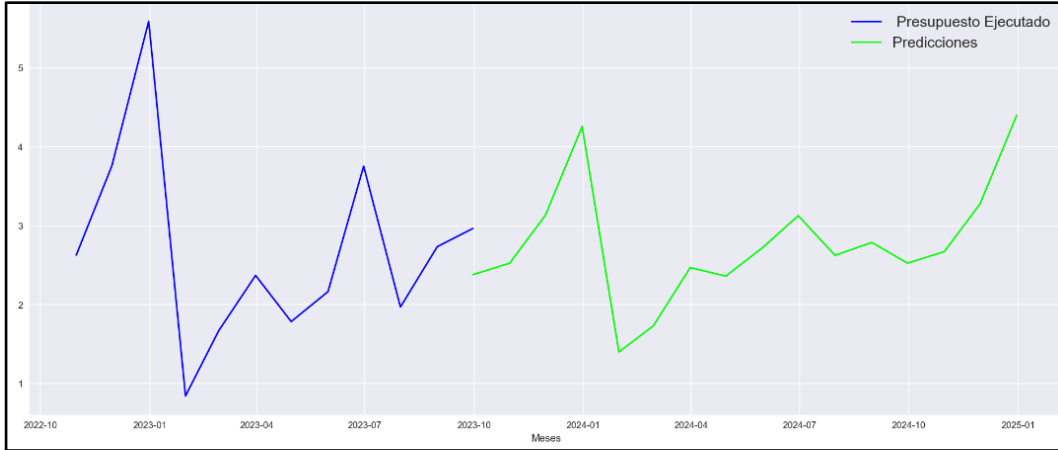


Figura 72. Predicción con Modelo

Fuente: Elaboración Propia.

16. Métricas de evaluación de modelo.

Como último paso, se realizó una evaluación de ciertos indicadores o marcadores del modelo que marcarán la pauta de la eficiencia y precisión del modelo que se ha generado a partir de las iteraciones antes mencionadas. En la figura 69 como se puede ver se obtuvo un R2 aceptable en función de la hipótesis de la investigación. Adicional, se logra una precisión del modelo de un 53% dado que el error porcentual que arrojó fue de 47% lo cual indica que no es el mejor en cuanto al manejo de los errores.

```

Evaluation metric results:-
MSE is : 0.3913281774280657
MAE is : 0.47863057057444197
RMSE is : 0.6255622890073104
MAPE is : 47.218749035060625
R2 is : 0.7231214461221657

```

Figura 73. Métricas de evaluación del modelo.

Fuente: Elaboración Propia.

6.4.2.2 Modelo Random Forest

El modelo de Random Forest es un algoritmo de clasificación, sin embargo, también es utilizado en la regresión ya que este permite tener mayor flexibilidad y capacidad para manejar las relaciones no lineales con patrones complejos y cambiantes a lo largo del tiempo. Es por ello por lo que se ajusta muy bien en aplicaciones con series temporales.

1. Extracción, Transformación y Carga de los datos

Como punto de partida se repite los pasos desde el 1 al 6 del modelo ARIMA donde está comprendida la extracción, transformación y carga del conjunto de datos.

2. Creación de variable con variación porcentual entre periodos

En este paso se genera una variable adicional a la serie conteniendo la variación porcentual entre cada periodo, con el objetivo de poder normalizar los datos. Adicional, como se eliminan filas con valores NAN.

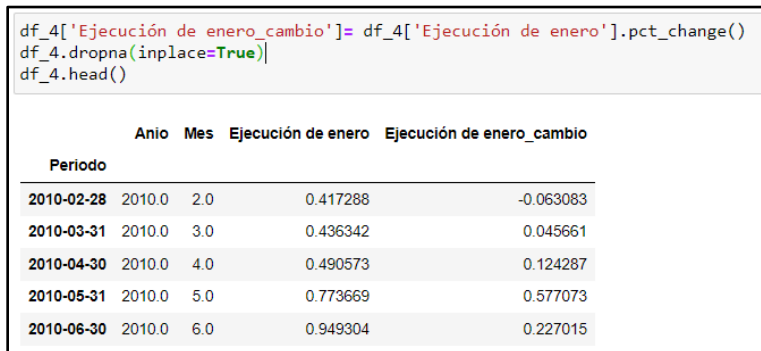


Figura 74. Generación de variable con variación porcentual entre periodo

Fuente: Elaboración Propia.

3. Gráfica de variación porcentual de la serie

Se genera una gráfica con los datos de la variación porcentual de la serie para visualizar comportamiento de estos a medida pasa el tiempo.

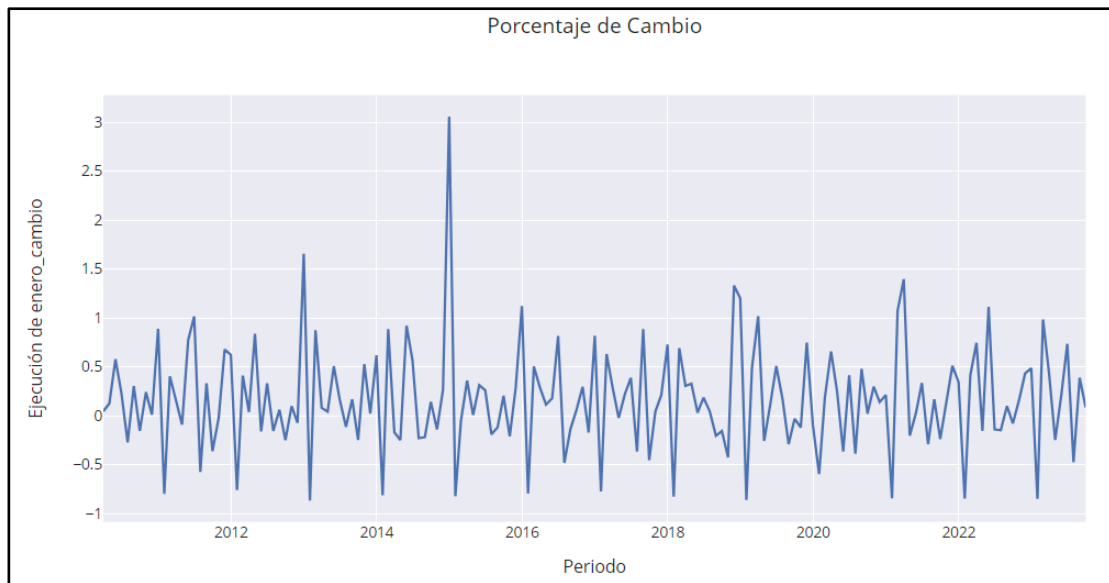


Figura 75. Gráfica de variación Porcentual de la serie temporal.

Fuente: Elaboración Propia.

4. Técnica de ingeniería de características.

En el siguiente paso se realiza una ingeniería de características a partir de los datos de la variación porcentual, con el objetivo de encontrar un detalle más a fondo en la variación para que el modelo aprenda de una mejor manera, capturando información estacional, datos categóricos y mejorar también la interpretabilidad. Se generaron variables adicionales que están conformadas por los valores de los 12 meses según cada rezago.

Anio	Mes	Ejecución de enero	Ejecución de enero_cambio	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
2012	12	2.427936	1.656454	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	01	0.324511	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	02	0.607887	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	03	0.657707	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	04	0.685083	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	05	1.031777	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN	NaN
2013	06	1.198086	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN	NaN
2013	07	1.061741	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN	NaN
2013	08	1.239877	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN	NaN
2013	09	0.933064	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN	NaN
2013	10	1.424616	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454	NaN
2013	11	1.457254	0.022910	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454
2013	12	2.358146	0.618212	0.022910	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343

Figura 76. Características adicionales creadas.

Fuente: Elaboración Propia.

5. Modelado (Entrenamiento y Prueba).

Se crean 2 conjuntos de datos, uno solamente conteniendo los datos originales de la serie (Target) y el otro con los datos de las características rezagadas que se crearon previamente (Variables Predictoras). Con este proceso se está indicando al modelo que la variable a predecir es la del primer conjunto a partir de las variables predictoras que son las variables rezagadas.

```
X2.head(3)
```

	Ejecución de enero	Ejecución de enero_cambio	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	
Periodo														
201312	2.358146	0.618212	0.022910	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343	1.656454
201401	0.438814	-0.813916	0.618212	0.022910	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239	-0.866343
201402	0.827201	0.885082	-0.813916	0.618212	0.022910	0.526816	-0.247454	0.167777	-0.113802	0.161186	0.506061	0.041624	0.081957	0.873239

```
y2.head()
```

Periodo	Ejecución de enero
201312	2.358146
201401	0.438814
201402	0.827201
201403	0.685990
201404	0.515738

Name: Ejecución de enero, dtype: float64

Figura 77. Datos de Entrenamiento y Prueba.

Fuente: Elaboración Propia.

6. Implementación de Modelo

Una vez se ha separado ambos conjuntos se procede a implementar el modelo, en la presente investigación se utilizaron 400 árboles para poder predecir el valor de la ejecución presupuestaria.

```
# Random Forest  
  
# Creamos el modelo con 400 árboles  
rfr = RandomForestRegressor(n_estimators=400)  
  
# Entrenamos el modelo  
rfr.fit(train_feature, train_target)  
  
# Hacemos las predicciones  
fcst = rfr.predict(test_feature)
```

Figura 78. Implementación de Modelo.

Fuente: Elaboración Propia.

7. Gráfica de Resultados con el modelo.

Como se puede ver en la Figura 75 el modelo logra predecir de una buena manera los valores de la ejecución presupuestaria a medida va transcurriendo cada periodo. Esto indica que las características que se definieron logran un buen ajuste con el modelo logran de esta manera acercarse a la realidad.

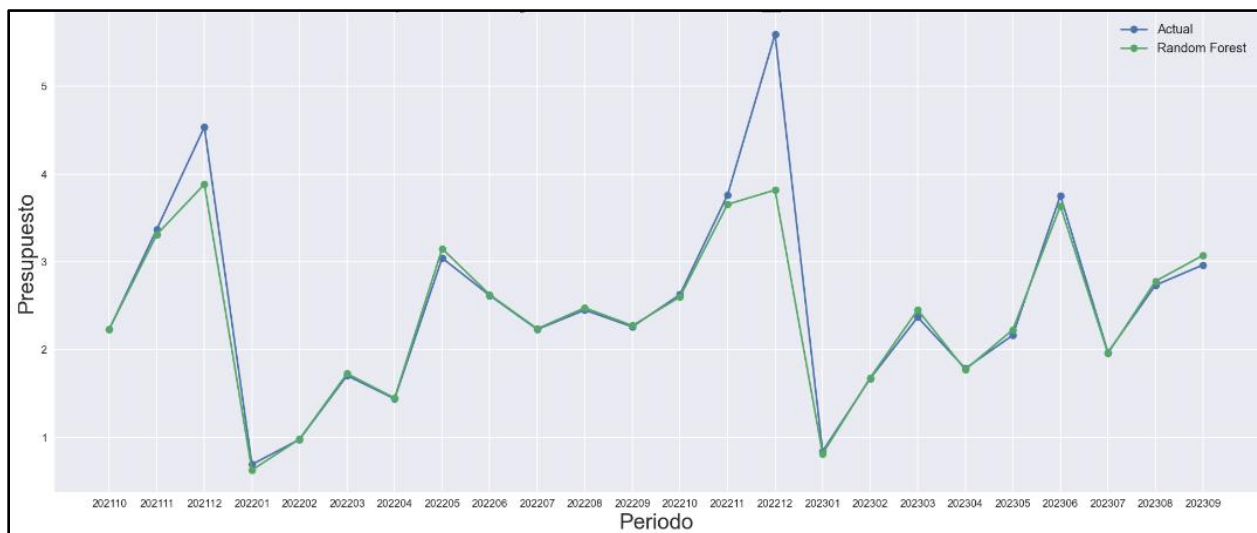


Figura 79. Gráfica de resultados del modelo.

Fuente: Elaboración Propia.

8. Evaluación del Modelo.

Al evaluar las métricas y la precisión del modelo, se puede ver que se apega bastante a lo que sería un modelo óptimo para implementación. El porcentaje de errores absolutos es relativamente bajo lo que indica una buena precisión, sumado al valor de R2 por encima del 80%.

```
Evaluation metric results:-  
MSE is : 0.1518069524915092  
MAE is : 0.14005124426312007  
RMSE is : 0.3896241169274679  
MAPE is : 3.718352225908176  
R2 is : 0.8791285362018881
```

Figura 80. Evaluación del Modelo.

Fuente: Elaboración Propia.

Nota: cada uno de los pasos descritos anteriormente es aplicable para la ejecución presupuestaria de acuerdo con los conjuntos de datos que proporciona SEFIN, haciendo el filtro por el sector de interés y por ende cada resultado tendrá variantes de acuerdo con los datos que se utilicen.

6.4.2.3 Selección de Infraestructura

Se establecerá un contrato con Azure, cuyos proveedores o intermediarios en Honduras es el equipo de Software One, con quienes se realizó la cotización de un plan mensual que ofrece los siguientes requerimientos nivel de infraestructura:

- Modelo: Servidor con núcleo virtual
- Nivel de procesamiento: Aprovechado
- Tipo de Hardware: Serie Standard Gen 5
- Instancia: 8 vCore
- Redundancia: 1 instancia local
- Almacenamiento: 32 GB
- Almacenamiento por Respaldo: 10 GB
- Tamaño promedio de la base de datos durante el período de retención: 10 GB

6.4.2.4 Especificaciones de Infraestructura

En el presente proyecto se ha realizado un análisis con el fin de dimensionar los requerimientos mínimos a nivel tecnológico, para ello se presenta a continuación un desglose de todos estos aspectos que deben tomarse en cuenta como requerimiento previo para la puesta en marcha del chatbot, a nivel de servidor y sistema gestor de base de datos:

Tabla 10. Especificaciones de Servidor On Cloud Azure.

Almacenamiento	2 TB
Procesador	E5 Xeon v3 Ls-series 8 vCore
Memoria RAM	16 GB
Sistema Operativo	Windows Server 2019
Tipo	Virtual

Fuente: Elaboración propia.

Tabla 11. Especificaciones de Sistema Gestor de Base de Datos.

Versión	Microsoft SQL Server 2019 Estándar Edition
---------	--

Fuente: Elaboración propia.

6.5 MEDIDAS DE CONTROL

Para una correcta puesta en marcha del proyecto es necesario que se definan una serie de medidas de control durante la ejecución de este:

1. Seguimiento continuo de los avances: tomando como referencia los hitos incluidos en el EDT y en el cronograma de actividades. Se sugiere utilizar listas de control y reuniones periódicas como técnicas para el proceso.
2. Control de presupuesto: verificar que los costos y gastos se apeguen a lo establecido en la estimación presupuestaria.
3. Calidad en los conjuntos de datos utilizados: aplicar la mayor cantidad de técnicas de preparación y limpieza de datos con el fin de obtener resultados precisos.
4. Validación de los modelos: haciendo uso de data de prueba externa en relación con los conjuntos de datos utilizados en la etapa de entrenamiento.
5. Capacitaciones: al recurso humano que se encargará de utilizar el modelo en ambientes

productivos, con el fin de que la institución obtenga el mayor provecho posible.

6. Gestión de riesgo: se busca detectar riesgos potenciales en la aplicación de los modelos para evaluar su criticidad de la siguiente forma:

Tabla 12. Medición de riesgos detectados.

Riesgo	Aparición Probabilidad	Gravedad (Impacto)	Valor del riesgo	Nivel de riesgo
Caída en el sistema.	2	5	12	Importante
Pago impuntual/inexistente de la licencia mensual de base de datos lo cual impedirá acceder a la información almacenada.	2	4	8	Apreciable
Que la infraestructura tecnológica de los modelos sufra daños.	2	5	12	Importante
Enfermedad de algún miembro del equipo de expertos contratado para el proyecto.	3	2	6	Apreciable

Fuente: Elaboración Propia.

Para la medición anterior se utilizó la clasificación detallada en la Figura 81, la cual dicta los parámetros de acuerdo con las variables Aparición y Gravedad, indicando el nivel de Riesgo que puede ir desde Marginal (Muy leve) hasta Muy grave.

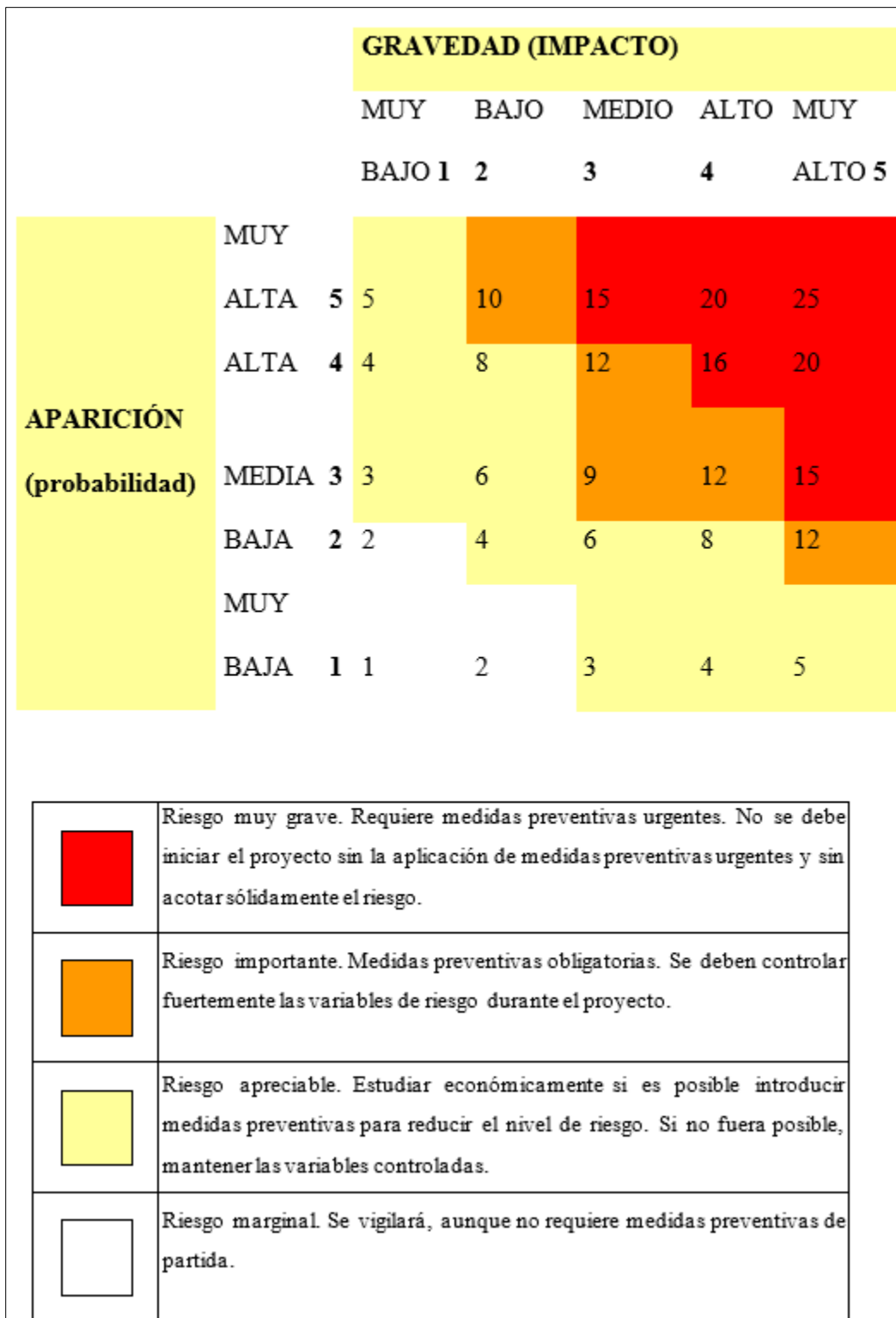


Figura 81. Parámetros para medir riesgos detectados.

Fuente: Elaboración Propia.

6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

Para este proyecto se utiliza en primera instancia una estructura del desglose del trabajo (EDT) el cual permite una visualización más clara de los entregables.



Figura 82. Diagrama EDT.

Fuente: Elaboración Propia

6.6.1 TABLAS DE COSTOS

A continuación, se muestra de forma detallada cada uno de los costos de recursos humanos y tecnología.

Tabla 13. Pago por hora recursos humanos.

Puesto/Cargo	Pago por Hora (Dólares Americanos)
Encargado de proyectos	\$ 20.00
Analista Jr. de Proyectos	\$ 11.00
Jefe de infraestructura	\$ 15.00
Técnico 1 de Arquitectura CLOUD	\$ 11.00
DBA 1 (Administrador de Base de Datos)	\$ 13.00
Técnico 1 en Ciencia de Datos	\$ 13.00
Técnico 2 en Ciencia de Datos	\$ 13.00

Fuente: Elaboración propia.

Para definir los costos tecnológicos es importante identificar la diferencia entre gastos operativos OPEX y las inversiones de capital CAPEX. Los OPEX básicamente son los servicios en la nube adquiridos por la empresa mientras que los CAPEX son los modelos tradicionales (locales) de TI que se adquieren y se incluyen en la planificación financiera de las empresas a largo plazo. (García R. , 2023)

Tabla 14. Costos en Tecnología (Mensuales) - OPEX.

Infraestructura	Costo Mensual Estimado
Azure Data Services	\$ 1,066.73
Licencia de Microsoft SQL Server	\$ 583.80
Restauración de Base de Datos	\$ 7.68
Almacenamiento en la Base de Datos	\$ 4.42
Total a pagar (Mensual)	\$ 1,662.63

Fuente: Elaboración propia.

Tabla 15. Costos en Tecnología - CAPEX

Infraestructura	Costo Mensual Estimado
Licencia Knime	\$ 0.00
Licencia Python	\$ 0.00
Total a pagar	\$ 0.00

Fuente: Elaboración propia.

6.6.2 CRONOGRAMA DE ACTIVIDADES

Aplicación de Modelos de Analítica								
Actividades	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Responsable	Dias trabajados	Monto a pagar
1. Inicio del Proyecto								
1.1 Definir Objetivos						Encargado de proyectos	15	\$ 2,400.00
1.2 Establecer el alcance						Encargado de proyectos	10	\$ 1,600.00
2. Planificación del proyecto								
2.1 Definir las actividades						Analista Jr. De Proyectos	5	\$ 440.00
2.2 Estimar los costos						Encargado de proyectos	5	\$ 800.00
2.3 Seleccionar plataforma tecnológica						DBA 1	5	\$ 520.00
2.4 Estimar recursos humanos						Analista Jr. De Proyectos	5	\$ 440.00
2.5 seleccionar la infraestructura						T1, T2	5	
3. Modelado del Proyecto								
3.1 Preparación de los Datos						DBA 1	5	\$ 520.00
3.2 Análisis Exploratorio de los Datos						T1C	5	\$ 520.00
3.3 Construcción del Modelo de Series Temporales						T2	5	\$ 520.00
3.4 Construcción del Modelo de Random Forest						T1	5	\$ 520.00
3.5 Definir los elementos de arquitectura								
3.6 Evaluar los riesgos								
3.7 Monitorear funcionamiento						T1, T2	5	\$ 520.00
4. Ejecución del Modelado								
4.1 Configuración de Datos de Entrenamiento						T1	2	\$ 208.00
4.2 Análisis de los Resultados						T2	2	\$ 208.00
5. Plan de comunicación								
5.1 Diseñar el Plan						Encargado de Proyectos	3	\$ 480.00
5.2 Ejecutar el Plan						Analista Jr. De Proyectos	2	\$ 176.00
						Total	84	\$ 9,872.00

Figura 83. Cronograma de actividades y distribución.

Fuente: Elaboración propia.

6.6.3 ESTIMACIÓN TOTAL DE COSTOS DEL PROYECTO

Tabla 16. Estimación Total.

Resumen de Costos	Cantidades
Costos Tecnológicos	\$ 8,313.15
Costos de Recursos Humanos	\$ 9,872.00
Inversión Total Proyecto	\$ 18,815.15

Fuente: Elaboración propia.

Nota: Es importante mencionar que la estimación total de costos descrita en la Tabla 16 está orientada a la capacidad instalada de un departamento de analítica de datos nuevo en la institución, por lo que, si está ya cuenta con recurso humano con la experiencia profesional y académica adecuados e incluso cuenta con equipo tecnológico local o servicios en la nube vigentes con características óptimas, la estimación se reduciría considerablemente.

6.7 PLAN DE COMUNICACIÓN

Este consiste en dar a conocer la nueva metodología y forma de gestión de los modelos analítica de datos en la institución. El plan de comunicación interno constará de:

- Una reunión de presentación.
- Un memorándum.

Se muestra cada uno de ellos:

MEMORÁNDUM

PARA: MIEMBROS DE LA INSTITUCIÓN

DE: DIRECTOR GENERAL

ASUNTO: PUESTA EN MARCHA DE UN MODELO PREDICTIVO

FECHA: 31/03/2024

Enfocados al crecimiento y mejora continua que tiene la institución como visión para este 2024 me complace informales por este medio que a partir del 01 de abril del presente año empezaremos a utilizar una nueva herramienta tecnológica como apoyo al proceso de ejecución

presupuestaria, un *modelo de analítica de datos*.

El modelo fue creado en la plataforma open source *Python* y en *MS SQL Server 2019* con infraestructura de almacenamiento en la nube en Azure Data Services y cuya estructura tiene como fin elaborar predicciones para que la elaboración de presupuestos sea óptima en la mayor medida posible.

Este proyecto tiene como propósito mejorar y transformar el proceso actual de la elaboración y ejecución de presupuestos, mediante el diseño de modelos de analítica que provean una alternativa en pro de conseguir el bienestar de la ciudadanía.

En los próximos días les convocaremos a una reunión para capacitarles sobre todos los aspectos relacionados, acerca de cómo funcionará el nuevo proceso.

Esperamos contar con su apoyo y trabajar juntos como equipo hacia el crecimiento y logro de metas de la institución. Al ser una tecnología nueva para ustedes, se brindará todo el acompañamiento necesario recibiendo constantes capacitaciones y talleres para que puedan sacarle el mayor provecho y adaptarse eficientemente a la misma.

Atentamente,

Director General

6.8 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

En el siguiente apartado se muestra de forma resumida cada uno de los elementos de la presente investigación con el fin de que se evidencie la concordancia existente en cada uno de los capítulos para determinar la fiabilidad de dicha investigación, otorgando un resultado final que logre ser beneficioso para futuras investigaciones e incluso ser utilizado por instituciones públicas y privadas.

Tabla 17. Concordancia de los segmentos de la investigación.

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título	Objetivo General	Objetivos Específicos	Teorías de Sustento	Variables	Población	Técnicas	Conclusiones	Nombre de la Propuesta	Objetivos
"Un Modelo de Analítica para Predecir el Presupuesto en los Sectores de Salud y Educación en Honduras a 2024".	Predecir el presupuesto de los sectores de salud y educación en Honduras a 2024 haciendo uso de los algoritmos de Aprendizaje Automático.	Identificar los objetos de gasto con mayor afectación en la ejecución del presupuesto de los sectores de salud y educación.	Enonomía Keynesiana. Teoría de la Probabilidad. Teoría Estadística. Teoría de Sistemas. Teoría de Optimización. Teoría de Aprendizaje Automático. Teoría de Minería de Datos.	Independiente: Modelos predictivos de analítica. Dependiente: Presupuesto de salud y educación.	Datos abiertos de la ejecución presupuestaria de SEFIN, conteniendo la información de aproximadamente 7,821,800 registros correspondientes al periodo entre 2010 y 2023	Modelos de analítica: Series de tiempo y Random Forest.	Si es posible realizar predicciones con modelos de analítica. Con series temporales se obtuvo 0.79 en educación y 0.72 para salud. Por su parte, con Random Forest se obtuvo 0.88 y 0.87 respectivamente.	"Aplicación de un Modelo de Analítica para Predecir el Presupuesto en los Sectores de Salud y Educación en Honduras a 2024".	Elaborar una guía detallada para que individuos o instituciones puedan crear modelos de aprendizaje automático relacionados con la ejecución presupuestaria.
		Determinar el nivel de eficiencia en la ejecución del presupuesto de los sectores de salud y educación.					Lo objetos de gasto que demuestran mayor repercusión para el sector educación son los sueldos básicos de docentes (25.75%), adicionales (17.12%), sueldos básicos (10.90%) y transferencias corrientes a universidades (7.95%). Por su parte, en el sector salud destacan los sueldos básicos (30.41%), productos farmacéuticos (7.03%), servicios médicos (3.46%) y contribuciones al INJUPEMP (3.34%).		
		Generar una predicción del presupuesto de Salud y Educación para el año 2024.					La eficiencia de ejecución presupuestaria en el sector educación a nivel de administración central muestra un nivel satisfactorio, con un desempeño superior al 85%. Por su parte, el sector público descentralizado es variable mostrando resultados entre adecuados y satisfactorios, superiores al 40%. De lo anterior se exceptúan los años 2013 y 2020 que muestran porcentajes deficientes (menores a 39%). En cuanto al sector salud a nivel de administración central se obtuvo un nivel satisfactorio, con un desempeño superior al 70%. Por su parte, el sector público descentralizado también evidencia resultados satisfactorios, superiores al 80%. De lo anterior se exceptúan los años 2010, 2011 y 2013 que muestran porcentajes deficientes (menores a 39%).		
							El presupuesto generado a través de la predicción oscila a L 34,523,479,118.00 para el sector educación y L 25,039,700,912.00 para el sector salud, por lo que al realizar la comparación con la publicación de SEFIN correspondiente a 2024 se tiene un porcentaje de efectividad del 64% y 61% respectivamente.		Describir cada uno de los pasos necesarios para la construcción/configuración de los algoritmos de series temporales y random forest en la predicción de presupuestos de los sectores salud y educación. Estructurar un cronograma de actividades que evidencie el tiempo y los costos que implicaría la aplicación de un modelo predictivo de presupuestos en ambientes de producción.

Fuente: Elaboración propia.

REFERENCIAS BIBLIOGRÁFICAS

- Alvarado, J. (2019). Análisis de la Ejecución Presupuestaria del Primer Nivel de Atención en el Ministerio de Salud de El Salvador, 2014 - 2017. CIES - UNAN Managua.
- Amazon. (09 de Mayo de 2023). *amazon web services*. ¿Qué es la regresión lineal?: <https://aws.amazon.com/es/what-is/linear-regression/>
- Amazon. (09 de Mayo de 2023). *amazon web services*. Qué es la visualización de datos: <https://aws.amazon.com/es/what-is/data-visualization/>
- AndBank. (28 de Junio de 2012). *ANDBANK Private Bankers*. <https://www.andbank.es/observatoriodelinversor/que-son-las-divisas/>
- ASJ (Asociación para una Sociedad Justa). (13 de 09 de 2022). La Realidad del Sector Educativo en Honduras: <https://confidencialhn.com/wp-content/uploads/2023/02/PPT-Estado-de-Pais-Educacion-2023.pdf>
- Barreto, A. (2012). *El progreso de la Estadística y su utilidad en la evaluación del desarrollo*. Toluca: Universidad Autónoma del Estado de México.
- Bohórquez, M., Torys, J., & Paredes, M. (2020). Modelos de Predicción de Deserción de Clientes para una Administradora de Fondos Ecuatoriana. *Revista Compendium: Cuadernos de Economía y Administración*, 1-11.
- Borja, M., Mercado, M., & Rodríguez, R. (2019). Beneficios Ofrecidos por la Gestión del Big Data en las Instituciones Gubernamentales en la Era de la Digitalización. *Propiedad Inmaterial*(28), 273-293.
- Castro, O., & Guerrero, D. (2021). *Modelo Predictivo de Propensión de Ahorro e Inversión en Productos de Banca Patrimonial*. Bogotá: Universidad de los Andes.
- CMF. (09 de Mayo de 2023). *EDUCA: Portal de Educación Financiera*. <https://www.cmfchile.cl/educa/621/w3-article-27152.html#:~:text=El%20cr%C3%A9dito%20es%20un%20pr%C3%A9stamo%20de%20dinero%20que%20una%20parte,que%20no%20tuvo%20ese%20dinero.>
- Cognición. (15 de 08 de 2022). ¿Cuánto invertirán los países de América Latina en educación en 2022? <https://sarrauteducacion.com/2022/08/15/cuanto-invertiran-los-paises-de-america-latina-en-educacion-en-2022/>
- Cuevas, J. (19 de Julio de 2022). *El sistema de salud en Honduras está en crisis, pero hay cifras que también muestran mejoras*. Voz de América: <https://www.vozdeamerica.com/a/sistema-salud-honduras-tesis-tambien-mejoras-6661257.html>
- Datos Macro. (31 de 12 de 2022). *Expansion*. Gasto Público en Salud: <https://datosmacro.expansion.com/estado/gasto/salud>
- Efron, B. (1998). *Fisher in the 21 st Century*. *Statistical Science*.
- EKOS. (04 de 08 de 2019). *Países con más inversión en educación en América Latina*. <https://ekosnegocios.com/articulo/paises-con-mas-inversion-en-educacion-en-america-latina>
- El Criterio HN. (22 de 11 de 2022). *Salud y Educación, entre los presupuestos que menos crecerán de cara a 2023*. <https://criterio.hn/salud-y-educacion-entre-los-presupuestos-que-menos-creceran-de-cara-a-2023/>
- El Heraldo. (04 de Enero de 2023). *L 1,360 millones destinarán para atender la crisis educativa en el país*. El Heraldo: <https://www.elheraldo.hn/honduras/presupuesto-tesis-educacion-infraestructura-honduras-DG11607224>

- EPC. (2021). *epc*. Empresas Públicas de Cundinamarca S.A. E.S.P.: <https://www.epc.com.co/intranet2012/sig/Financiera/Documentos/Indicador%20Ejecucion%20Presupuestal.pdf>
- Espinosa, G., & Vásquez, A. (2016). *Aplicaciones de Programación no Lineal*. México: Omnia Science.
- Expediente Publico. (20 de 07 de 2020). *La salud pública en Honduras, desdibujada y desmantelada en el gobierno de “la vida mejor”*. <https://www.expedientepublico.org/la-salud-publica-en-honduras-desdibujada-y-desmantelada-en-el-gobierno-de-la-vida-mejor/>
- García, L. (1 de Septiembre de 2021). *Universidad Central*. ¿Qué son las redes neuronales y cómo funcionan?: <https://www.ucentral.edu.co/noticentral/redes-neuronales>
- García, R. (19 de Octubre de 2023). *Telefónica Tech*. Los 6 errores más comunes a la hora de presupuestar un proyecto Cloud: <https://telefonicatech.com/blog/6-errores-mas-comunes-presupuestar-un-proyecto-cloud#:~:text=Los%20servicios%20Cloud%20son%20gastos,planificaci%C3%B3n%20financiera%20a%20largo%20plazo.>
- González, A. (2021). *cleverdata*. Conceptos básicos de Machine Learning: <https://cleverdata.io/conceptos-basicos-machine-learning/>
- Guerra Bustillo, C. W. (2003). *Estadística*. La Habana, Cuba: Editorial Félix Avila .
- Iberdrola. (09 de Mayo de 2023). *iberdrola.com*. Descubre los principales beneficios del 'Machine Learning': <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- IBM. (09 de Mayo de 2023). *ibm.com*. ¿Qué es el random forest?: <https://www.ibm.com/mx-es/topics/random-forest>
- INEC. (2023). *Conceptos y definiciones*. Instituto Nacional de Estadística y Censo - Panamá.
- Jacobo, J. (2022). *Probabilidad e incertidumbre en Keynes. Una revisión de tipo bayesiano*. Bogotá: Revista de Economía Institucional.
- Jahan, S., Mahmud, A., & Papageorgiou, C. (2014). *¿Qué es la economía keynesiana?* International Monetary Fund.
- Jaklič, J. G. (2018). The role of compatibility in predicting business intelligence and analytics use intentions. *International Journal of Information Management*, 305–318.
- JPM Statistical Discovery. (09 de Mayo de 2023). *Portal de formación estadística*. https://www.jmp.com/es_co/statistics-knowledge-portal/what-is-correlation.html#:~:text=La%20correlaci%C3%B3n%20es%20una%20medida,afirmaciones%20sobre%20causa%20y%20efecto.
- Keynes, J. M. (1921). *A treatise on probability*. Cambridge: Cambridge University Press for the Royal Society.
- La Gaceta. (28 de Junio de 2004). Decreto No. 83-2004. *La Gaceta*.
- La Gaceta. (17 de Diciembre de 2021). Decreto No. 107-2021. *La Gaceta*.
- La Gaceta. (25 de Abril de 2023). Decreto Ejecutivo No. PCM 19-2023. *La Gaceta*.
- López, J. (2019). *Coeficiente de determinación (R cuadrado)*. *economipedia*: <https://economipedia.com/definiciones/r-cuadrado-coeficiente-determinacion.html>
- Luhmann, N. (1997). *Sociedad y Sistema: la Ambición de la Teoría*. Barcelona: Paidós.
- Mamani, Z., Del Pino, L., & Cortez, A. (2017). *Minería de datos distribuida usando clustering k-means en la predictibilidad del proceso petitorio en una organización pública*. Perú: Facultad de Ingeniería Industrial - UNMSM.
- Marcano, Y., & Talavera, R. (2007). *Minería de Datos como soporte a la toma de decisiones*

- empresariales*. Zulia: Universidad de Zulia.
- Mata, L. (02 de Julio de 2019). *Investigalia*. Profundidad o alcance de los estudios cuantitativos: <https://investigaliacr.com/investigacion/profundidad-o-alcance-de-los-estudios-cuantitativos/>
- MathWorks. (09 de Mayo de 2023). *Hiperplanos óptimos como límites de decisión*. es.mathworks.com: <https://es.mathworks.com/discovery/support-vector-machine.html>
- Miranda, J., Rey, P., & Weber, R. (2005). Predicción de Fugas de Clientes para una Institución Financiera mediante Support Vector Machines. *Revista Ingeniería de Sistemas*, 49-68.
- Montoya, E., & Yáñez, D. (2022). *Analítica de datos: Una tendencia para la toma de decisiones empresariales en las organizaciones*. Unibersidad Libre (Seccional de Cúcuta).
- Muñoz, D. (19 de Septiembre de 2018). *Universidad Pontificia Bolivariana*. <https://www.upb.edu.co/es/central-blogs/ortografia/como-se-dice-migrante-emigrante-inmigrante>
- Oberst, T. (2014). *La importancia del ahorro : teoría, historia y*. Biblioteca Digital de la Universidad Católica de Argentina.
- Ordóñez, R., & Pastor, M. (2016). *Sistema de predicción de clientes desertores de tarjetas de crédito para la banca peruana usando Support Vector Machine*. Lima: Universidad Nacional Mayor de San Marcos.
- Ortega, C. (02 de Noviembre de 2023). *¿Qué es el muestreo por conveniencia?* QuestionPro: <https://www.questionpro.com/blog/es/muestreo-por-conveniencia/>
- Pacheco, M., Quispe, T., & Soto, R. (2021). *Ejecución Presupuestal de Gastos y su Incidencia en el Cumplimiento de Metas Institucionales de la Municipalidad Distrital de San Antonio*. Universidad Nacional del Callao.
- Padmanabhan, T. (1999). *Unexpectedness as a measure of Interestingness in Knowledge Discovery System*.
- Pariona, J. (2017). *Clasificación de Fuga de Clientes en una Entidad Financiera Utilizando el Algoritmo Smote para Datos Desbalanceados en una Regresión Logística*. Lima: Universidad Nacional Agraria La Molina.
- Pepió, M. (2009). *Series Temporales*. Universidad Politècnica de Catalunya. <https://doi.org/9788483016367>
- Piatesky, G., & Frawley, W. (1991). *Knowledge Discovery in Databases*. Cambridge.
- PISA (programme for international student assessment). (2018). *OECD*. Resultados de PISA 2018 (Volumen I): Lo que los estudiantes saben y pueden hacer: <https://www.oecd.org/pisa/pisa-es/>
- Ramos, J. (2019). *Aprendizaje automático para flujo de datos*. Universidad Politécnica de Madrid: Madrid.
- Recuero de los Santos, P. (13 de Diciembre de 2021). *Telefónica Tech*. Cómo interpretar la matriz de confusión: ejemplo práctico: <https://empresas.blogthinkbig.com/como-interpretar-la-matriz-de-confusion-ejemplo-practico/>
- Rodríguez, C., Lorenzo, O., & Herrera, L. (2005). *Teoría y práctica del análisis de datos cualitativos*. Universidad Autónoma de Tamaulipas.
- Román, V. (25 de Abril de 2019). *medium*. Algoritmos Naive Bayes: Fundamentos e Implementación: <https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fundamentos-e-implementaci%C3%B3n-4bcb24b307f>
- Saint Leo University. (05 de 01 de 2023). *Los 10 países con el mejor sistema de salud en el mundo*. <https://worldcampus.saintleo.edu/noticias/cuales-son-los-10-paises-con-mejor->

- sistema-de-salud-en-el-mundo
- Sampieri, R., Fernández, C., & Baptista, P. (2002). *Metodología de la Investigación*. Distrito Federal: McGraw-Hill.
- Sampieri, R., Fernández, C., & Baptista, P. (2010). *Metodología de la Investigación*. Mc Graw Hill.
- Sánchez, J. (01 de Abril de 2020). economipedia:
<https://economipedia.com/definiciones/presupuesto.html>
- Sánchez, J. (01 de Mayo de 2020). *economipedia*. Banca:
<https://economipedia.com/definiciones/banca.html>
- Sánchez, W. (2016). *Análisis de la Implementación del Presupuesto por Resultados En Guatemala*. Universidad de Chile.
- SEFIN. (2013). *Clasificador por Objeto de Gasto para Gobiernos Locales*. Secretaría de Finanzas.
- SEFIN. (13 de Noviembre de 2023). *Datos Abiertos Ejecución Presupuestaria*.
<https://www.sefin.gob.hn/ejecucion-presupuestaria/>
- Sevilla, A. (01 de Septiembre de 2021). *economipedia*. Producto interior bruto (PIB):
<https://economipedia.com/definiciones/producto-interior-bruto-pib.html>
- statista. (2020). *Ranking de los países de América Latina y el Caribe con mayor gasto público en salud como porcentaje del PIB en 2020*.
<https://es.statista.com/estadisticas/1270377/paises-con-mayor-gasto-sanitario-como-porcentaje-del-pib-en-latinoamerica/>
- Stedman, C. (Mayo de 2021). *Análisis o analítica de datos*. ComputerWeekly.es:
<https://www.computerweekly.com/es/definicion/Analisis-o-analitica-de-datos#:~:text=En%20este%20sentido%2C%20su%20naturaleza,tiene%20un%20enfoque%20m%C3%A1s%20amplio.>
- UNESCO. (2020). *Informe de seguimiento de la educación en el mundo*. GEM Report:
<https://gem-report-2020.unesco.org/es/seguimiento/#:~:text=A%20nivel%20mundial%2C%20el%2086,en%20la%20generaci%C3%B3n%20m%C3%A1s%20joven.>
- Universidad de Guadalajara. (02 de Noviembre de 2023). *Clasificación general de las fuentes de información*. Biblioteca Virtual del Sistema de Universidad Virtual:
<http://biblioteca.udgvirtual.udg.mx/portal/clasificacion-general-de-las-fuentes-de-informacion>

ANEXOS

Anexo 1 Workflow del Análisis Exploratorio de los Datos en Knime.

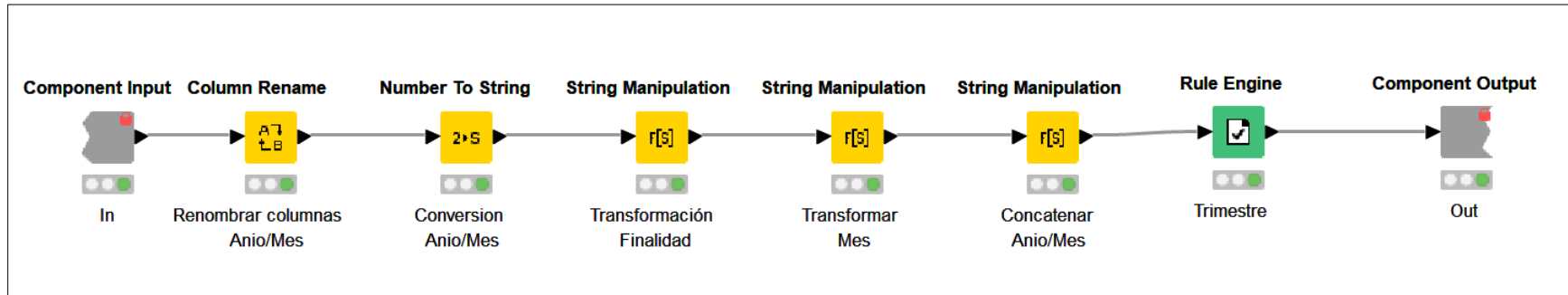


Figura 84. Preparación de datos EDA.

Fuente: Elaboración propia.

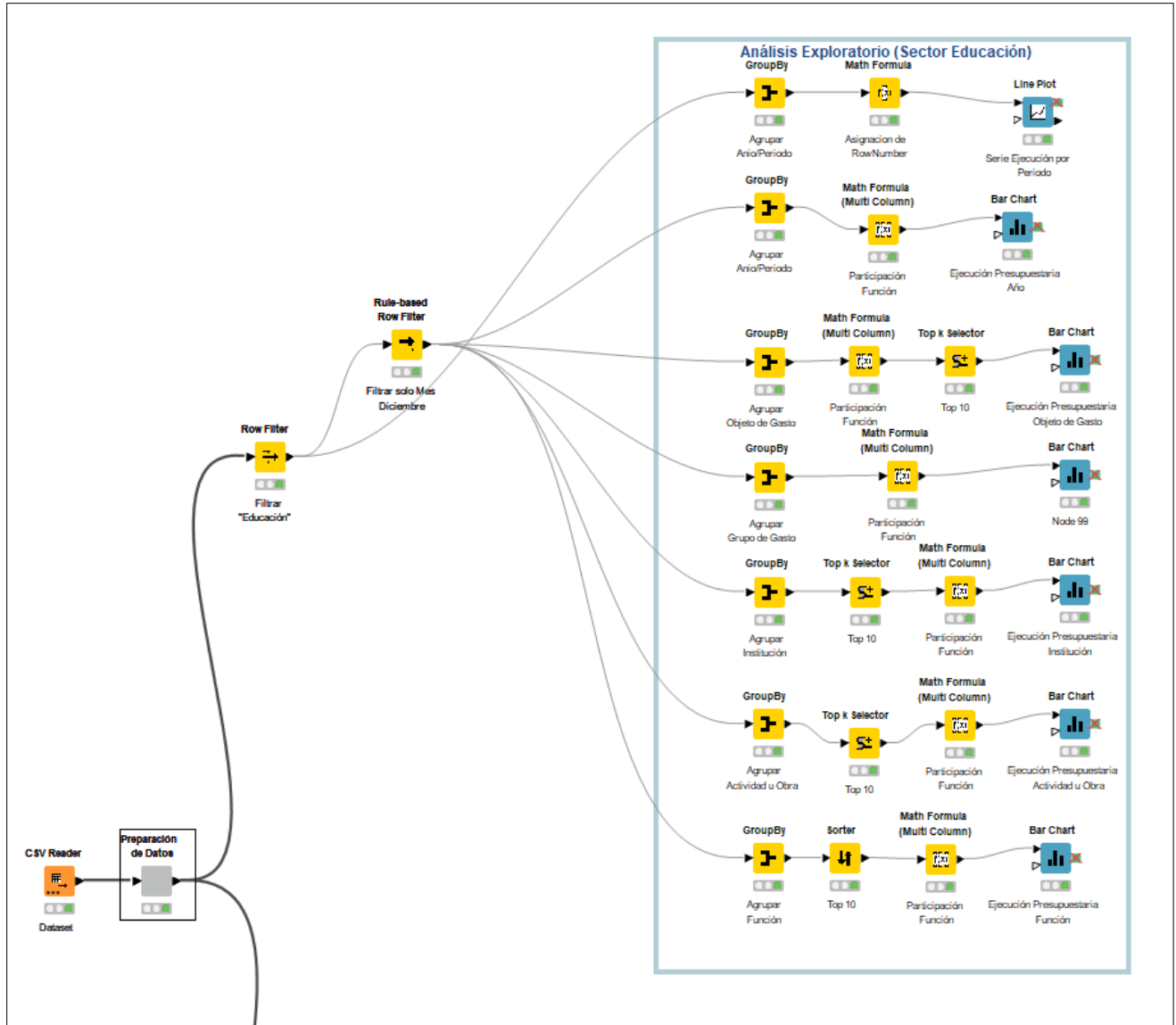


Figura 85. EDA Sector Educación.

Fuente: Elaboración propia.

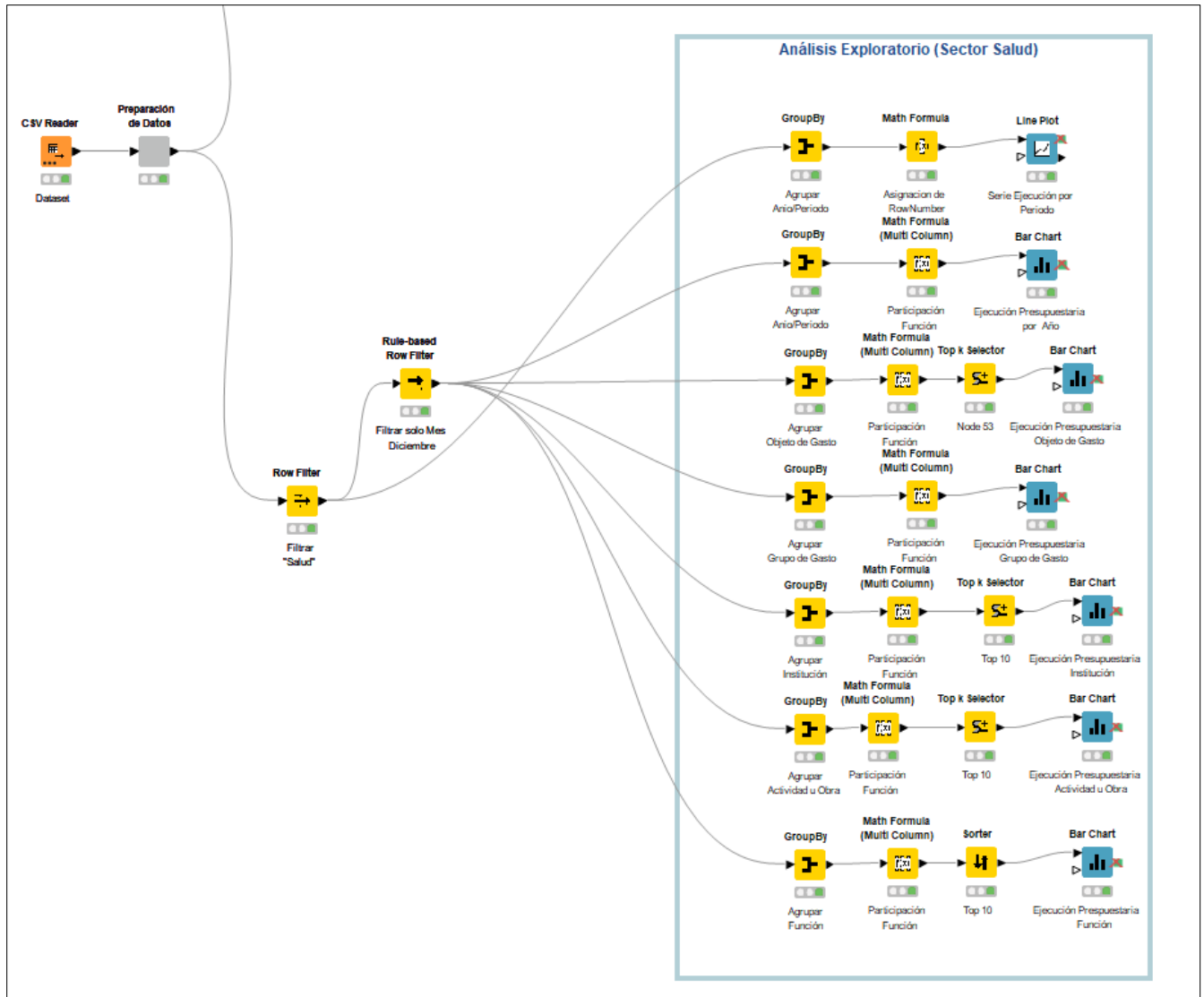


Figura 86. EDA Sector Salud.
Fuente: Elaboración propia.