ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS

INNOVACIÓN PARA LA EXCELENCIA

Análisis de la proyección de la demanda para la optimización de los procesos académicos de un centro de apoyo educativo

Alulema Chiluiza, Diana Verónica

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

Msc. Mazón Quinde, Inabel Karina

17 de noviembre del 2021

TESIS DIANA VERONICA ALULEMA CHILUIZA.docx

Scanned on: 22:3 February 21, 2022 UTC

Identical Words	181
Words with Minor Changes	0
Paraphrased Words	368
Omitted Words	1612

COPYLEAKS Website | Education | Businesses

Firma:

INABEL KARINA MAZON QUINDE

Firmado digitalmente por INABEL KARINA MAZON QUINDE
Fecha: 2022.02.21 20:14:56 -05'00'

Msc. Mazón Quinde Inabel Karina

DIRECTOR

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, **"Análisis de la proyección de la demanda para la optimización de los procesos académicos de un centro de apoyo educativo"** fue realizado por la ingeniera **Alulema Chiluiza, Diana Verónica** el mismo que ha sido revisado en su totalidad y analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 17 de Noviembre de 2021

Firma:

INABEL KARINA MAZON QUINDE

Firmado digitalmente por INABEL KARINA MAZON QUINDE
Fecha: 2022.03.14 16:01:17 -05'00'

Msc. Mazón Quinde Inabel Karina

DIRECTOR

C.C.: 0604597112

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo, **Alulema Chiluiza, Diana Verónica** con cédula de ciudadanía nº 1803440070, declaro que el contenido, ideas y criterios del trabajo de titulación "**Análisis de la proyección de la demanda para la optimización de los procesos académicos de un centro de apoyo educativo**" es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 17 de Noviembre de 2021

Firma:

Ing. Alulema Chiluiza Diana Verónica

C.C.: 180344007-0

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Alulema Chiluiza, Diana Verónica** con cédula de ciudadanía nº 1803440070, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación **"Análisis de la proyección de la demanda para la optimización de los procesos académicos de un centro de apoyo educativo"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 17 de Noviembre de 2021

Firma:

.....

Ing. Alulema Chiluiza Diana Verónica

C.C.: 180344007-0

Dedicatoria

El presente trabajo está dedicado al esfuerzo que permite alcanzar los objetivos.

Agradecimiento

A Dios y a todas las personas que me apoyaron en el camino para este logro académico.

Índice de Contenido

Índice de Tablas

Índice de Figuras

Resumen

El presente estudio se realiza en un centro de apoyo educativo dedicado al fortalecimiento de los conocimientos académicos en el área de ciencias exactas, en su funcionamiento diario cuando se solicita un servicio académico se decide si se acepta o se niega la inscripción del estudiante de acuerdo a la disponibilidad de aulas, maestros y horarios, sin embargo esta asignación de clases tiene un acierto promedio de solo el 33% ya que se realiza de manera manual e improvisada, basándose solamente en el criterio y conocimiento del personal; situación que ha generado problemas ya que en temporadas de alta demanda no se puede dar atención a todos los estudiantes, y en temporadas de baja demanda los recursos físicos y humanos se desperdician ya que no hay estudiantes.

El presente estudio plantea el diseño de un modelo de analítica de datos para la predicción de la demanda de estudiantes, en el cual mediante el análisis de los datos históricos se podrá conocer la fluctuación de la demanda, y obtener la proyección de la demanda para diferentes temporadas del año lectivo. De manera que la predicción que se obtiene con el modelo de analítica de datos permitirá realizar la planificación de la oferta del servicio académico, para optimizar los procesos académicos y mejorar la logística asociada a la disponibilidad de aulas, maestros y horarios, a fin de que el centro de apoyo educativo pueda brindar una atención eficiente para el beneficio tanto de la institución como de las personas que requieren las clases.

Palabras Clave:

- **ANALÍTICA DE DATOS**
- **MODELO DE PREDICCIÓN**
- **SERVICIO ACADÉMICO**

Abstract

This study has been performed on an educational support centre oriented to strength the academic knowledge related to exact sciences. The procedure begins when an academic service is required, next it is decided if the student is accepted or rejected according to room, teachers and schedules availability; being this a manual process and based on personal experience mainly, different issues have appeared a long the way, especially during high demand season when all the requirements cannot be covered, while during low demand the resources are wasted because of the lack of students. This work states a model to predict the demand of students based on data analytics, through an historical data analysis with the aim to understand the demand fluctuation; later, the demand projection is developed for the different seasons of the academic year. The model obtained will lead to optimize the academic processes inside the educational support centre, since a planning is possible in terms of offer and demand of the academic services subjected to data projections. On one side, this improves the logistics associated with room, teachers and schedules availability, while on the other side reduces customers attention times, which can be traduced as a win-win relationship between the centre and its clients.

Keywords:

- **DATA ANALYTICS**
- **PREDICTION MODEL**
- **ACADEMIC SERVICES**

Capítulo I

Introducción

Antecedentes

La necesidad de un aprendizaje eficiente ha dado lugar al surgimiento de varias organizaciones dedicadas a fortalecer los conocimientos académicos, siendo una de ellas el centro de apoyo educativo caso de estudio el cual fue creado en el año 2016 con la finalidad de brindar un servicio educativo de calidad, y ha encaminado sus esfuerzos a fortalecer el conocimiento, la seguridad y la motivación de los estudiantes, evidenciándose que los mismos logran vencer el temor a las materias, se sienten motivados con su aprendizaje, mejoran su rendimiento académico y mejoran su estado emocional; sin embargo debido a que no existe una planificación de la oferta del servicio académico, la asignación de clases se realiza de manera manual e improvisada de acuerdo al criterio y conocimiento del personal que atiende el requerimiento, lo cual ha ocasionado problemas asociados a la disponibilidad del servicio académico, ocasionando que en temporadas de alta demanda no se puede dar atención a todos los requerimientos de clases por falta de aulas o maestros, y en temporadas de baja demanda los recursos físicos y humanos se desperdicien porque no hay estudiantes.

La gestión de un negocio requiere de un conocimiento preciso sobre la demanda para mejorar la toma de decisiones y planificar las estrategias, en busca de mejorar la atención, optimizar los recursos y asegurar la supervivencia del negocio en mercados de alta competitividad; es así que se han venido desarrollado muchos estudios en torno al análisis de la información lo que ha dado lugar a una intensa evolución de la ciencia y la tecnología en el área de analítica de datos; la proyección de la demanda es una estrategia de la analítica de datos que ha tenido lugar en varios proyectos de diferentes sectores como el comercial, educativo, sanitario, entre otros (E3 Evolución Pymes,

2018). Por lo que resulta de gran interés para el centro de apoyo educativo la aplicación de la analítica de datos con la finalidad de obtener conocimiento sobre la fluctuación de la demanda del servicio académico.

En el año 2017 en Ecuador cerca de 3 millones de adultos iniciaron algún tipo de negocio o ya eran dueños de un negocio reciente, sin embargo el 90% no llegan a superar los tres años de funcionamiento, debido a que se presentan debilidades en la gestión de los nuevos negocios, lo que limita el crecimiento y la sostenibilidad de los emprendimientos (ESPAE Graduate School of Management, 2018).

En la Escuela Politécnica Nacional, se ha desarrollado un modelo de analítica de datos en el cual se considera la demanda de estudiantes por preferencia de carrera y el porcentaje de aprobación del curso de nivelación, para obtener una proyección del número de estudiantes que serán promovidos a primer semestre de cada carrera después de aprobar el curso de nivelación, para así mediante una regresión lineal determinar el número de aspirantes que deberían ingresar al curso de nivelación por preferencia de carrera (Sanchez, Sandoval, & Daza, 2017).

El Grupo Automotores y Anexos, Ayasa, reconoce a la tecnología como una pieza clave para el desarrollo de nuevos negocios, por lo que en los últimos tres años ha venido desarrollando dos proyectos tecnológicos de analítica de datos relacionados con las finanzas y con la cadena de abastecimiento de repuestos, los cuales se llevaron a cabo con éxito (iTahora, 2019).

En la Universidad Europea, se está desarrollando un ambicioso proyecto tecnológico que busca integrar los recursos de los diferentes departamentos a fin de analizar más de dos décadas de información y obtener conocimiento útil para mejorar la experiencia académica de los alumnos; el resultado del proceso analítico ayuda a mejorar la toma de decisiones y adelantarse a acontecimientos ya que permite proyectar la información a futuro y predecir comportamientos, además el proyecto

permite también la actualización tecnológica y evolución de la institución de manera que se garantice el presente y futuro de la misma. (Logicalis Spain, 2019).

ARC Airlines Reporting Corporation dedicada a la transacción de pasajes aéreos, aborda la nueva gestión de datos, ya que recopila las transacciones de aerolíneas como Delta, American Airlines, British Airways, Alaska Airlines y agencias de viajes como Expedia, con el fin de obtener información sobre el comportamiento de compra, y así conocer los lugares a donde se dirigen los viajeros, cuando viajan y cuánto pagan; ARC captura los datos, los analiza, los refina y finalmente crea informes personalizados para cada cliente (AWS, 2019).

Planteamiento del Problema

En el funcionamiento diario del centro de apoyo educativo la asignación de horarios, aulas, maestros e insumos en general se realiza de manera improvisada y subjetiva, basándose solamente en los requerimientos del día a día y en el criterio del personal, lo cual ha provocado una discordancia en la oferta del servicio académico, reflejándose en una efectividad tan solo del 33% en el desarrollo de la actividad académica. Por lo tanto se evidencia que se presentan problemas asociados al manejo de la fluctuación de la demanda de estudiantes, debido a que no existe un procedimiento objetivo para la planificación de la oferta del servicio académico, lo cual se debe tanto a la incertidumbre sobre la fluctuación de la demanda, como a la falta de herramientas tecnológicas que permitan automatizar los procedimientos académicos.

El hecho de no conocer cuándo va haber un alto requerimiento de clases o un bajo requerimiento de clases o cuándo no va haber requerimiento de clases, desencadena fallas en la logística de la institución, ya que no se puede realizar con certeza la asignación de los recursos físicos y humanos relacionados a los procedimientos académicos, de manera que no se puede anticipar el acondicionamiento

de aulas y la reserva del tiempo de maestros, lo cual conlleva a que la disponibilidad del servicio académico se vea afectada. Es así que cuando no hay estudiantes la ocupación en vano de las aulas y el desperdicio de tiempo de maestros se traduce en gastos innecesarios, y por el contrario cuando hay una alta demanda del servicio académico no se puede dar atención a todos los estudiantes debido a la falta de aulas y disponibilidad de tiempo de maestros, lo cual da como resultado pérdidas económicas y afectación a la imagen de la institución.

Justificación e Importancia

La falta de planificación de la oferta del servicio académico representa una debilidad en la gestión del centro de apoyo educativo, que afecta directamente a la logística relacionada con la asignación de recursos físicos y humanos, ya que no se puede realizar con anticipación la adecuación de aulas y reserva de maestros necesarios; lo que ha ocasionado fallas en la disponibilidad del servicio académico, reflejándose en una atención deficiente y desperdicio de recursos, ya que en temporadas de alta demanda se ha negado la atención a personas que solicitan las clases y por el contrario en temporadas de baja demanda se desperdicia el espacio físico así como el tiempo de maestros, lo cual implica pérdidas económicas y gastos innecesarios.

La principal causa que impide realizar una planificación adecuada de la oferta del servicio académico es la incertidumbre sobre la fluctuación de la demanda, razón por la cual en el presente trabajo mediante la analítica de datos se realizará el análisis de la información histórica del centro de apoyo educativo, a fin de conocer sobre la dinámica del negocio, y se obtendrá la predicción de la demanda del servicio académico para diferentes temporadas del año lectivo. Entonces la planificación de la oferta del servicio académico basada en la predicción de la demanda de estudiantes permitirá

mejorar la atención en el centro de apoyo educativo beneficiando el desarrollo de las actividades académicas, para que la institución pueda ser competitiva, mantenerse y desarrollarse, ya que si no se realizan reajustes para corregir las falencias se puede incrementar el riesgo del cierre de actividades de este emprendimiento.

La solución tecnológica de analítica de datos también puede contribuir con el desarrollo tecnológico de las pymes al optimizar y automatizar los procesos, para el beneficio tanto de la empresa ya que podrá evitar costos innecesarios y pérdidas económicas, así como también para el beneficio de los clientes ya que se podrá proporcionar una mejor atención (Iqbal, Kazmi, Manzoor, Soomrani, Butt, & Shaikh, 2018).

Además el presente proyecto de investigación será un aporte a los estudios de análisis para la proyección de la demanda enfocado al sector educativo y puede ser considerado como un prototipo modelo que toma en cuenta factores propios de nuestra realidad educativa, por lo que puede servir como base para el análisis de otras entidades del sector educativo y contribuir al marco de referencia de investigaciones de proyectos de minería de datos.

Objetivo General

Diseñar un modelo de proyección de la demanda para la planificación de la oferta del servicio académico de un centro de apoyo educativo, utilizando técnicas y algoritmos de minería de datos.

Objetivos Específicos

OE1: Obtener información sobre modelos, técnicas y algoritmos de analítica de datos para el análisis de la proyección de la demanda en empresas de servicios, mediante una revisión inicial de literatura.

OE2: Diseñar un modelo de analítica de datos para la proyección de la demanda mediante la recopilación y análisis de la información histórica del centro de apoyo educativo, de manera que se pueda tener una visión a corto plazo de la planificación académica.

OE3: Comprobar el modelo analítico diseñado para la proyección de la demanda mediante técnicas de validación implementadas en minería de datos.

Formulación del Problema

De acuerdo a los objetivos específicos del presente trabajo de investigación que trata sobre el análisis de la proyección de la demanda para la optimización de los procesos académicos de un centro de apoyo educativo, se respondieron las siguientes preguntas de investigación:

RQ1.1 ¿Cuáles son los estudios de analítica de datos existentes en la actualidad sobre modelos predictivos en empresas de servicios?

RQ1.2 ¿Cuáles son los algoritmos y herramientas de analítica de datos más utilizados en modelos predictivos?

RQ1.3 ¿Qué herramienta de análisis de datos se va a utilizar para el análisis de la proyección de la demanda?

RQ2.1 ¿Qué herramienta de analítica de datos es la más adecuada para el centro de apoyo educativo?

RQ2.2 ¿Cuál es el algoritmo de analítica de datos que mejor se ajusta a los datos del centro de apoyo educativo?

RQ3.1 ¿Es posible validar el modelo propuesto mediante técnicas de validación de minería de datos?

RQ3.2 ¿Cuál es el nivel de confianza con el que se va a trabajar para el análisis de los resultados?

Hipótesis de Investigación

El proceso analítico de proyección de la demanda mejorará la predicción de requerimientos de clases para realizar la planificación de la oferta del servicio académico en el centro de apoyo educativo.

Alcance de la Investigación

El alcance del presente trabajo de investigación es realizar un estudio comparativo de las técnicas, herramientas y algoritmos de analítica de datos implementados a nivel nacional e internacional, con el fin de proponer una solución fácil y entendible para la proyección de la demanda del servicio académico del centro de apoyo educativo, la misma que contribuirá con la planificación académica.

Variables de Investigación

- Variable Independiente: Algoritmos y herramientas analíticas.
- Variable Dependiente: Modelo analítico para la planificación del servicio académico.

Capítulo II

Marco Teórico

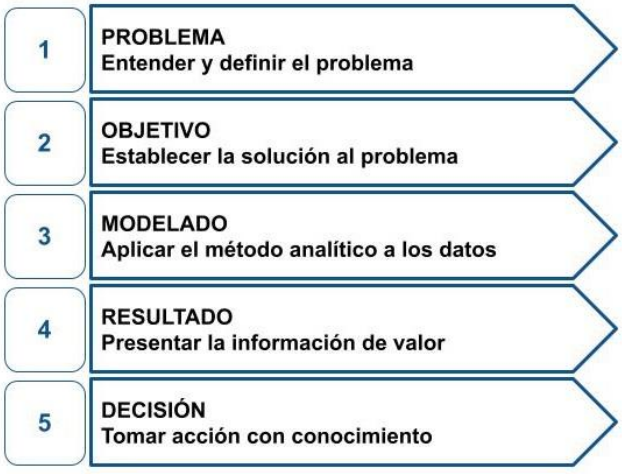
En este capítulo se presentan los contenidos teóricos y el estudio del estado del arte relacionado a las variables de investigación del presente trabajo, por lo tanto trata sobre la analítica de datos con especial énfasis en modelos, técnicas, algoritmos y herramientas que se enfocan en el proceso analítico para la proyección de la demanda.

Analítica de Datos

En la era tecnológica se experimenta el crecimiento exponencial de la cantidad de datos, ya que se generan de manera automática a través de sensores, máquinas, sistemas, redes interconectadas, clientes, proveedores, usuarios, etc., estos datos suponen un activo creciente para todas las empresas e instituciones, de ahí viene la importancia de la analítica de datos. En la Figura 1 se puede observar cuales son los pasos para los procesos analíticos.

Figura 1

Proceso de la analítica de datos



Definición del Problema

La mayoría de los problemas analíticos comienzan con cierta incertidumbre sobre la naturaleza del problema, por lo que el primer paso es eliminar la indeterminación del problema agregando criterios específicos a la comprensión del mismo, para evitar los callejones sin salida, los costos innecesarios del análisis de datos y más bien comenzar con un propósito claro y bien definido (Simon, 2017).

Definición de la Solución

Es una actividad de análisis que se utiliza para definir las necesidades del negocio, para comprender un problema mal definido o mal estructurado, con el fin de aumentar la comprensión e identificar preguntas que deben responderse para resolver la problemática. Raramente se encuentra una comprensión más profunda del problema en los datos, por lo que no es recomendable comenzar el análisis con los datos, más bien el análisis del problema está relacionado con la situación que causa el problema y los efectos negativos que ocasiona; una vez que está bien definido el problema se puede iniciar con el análisis de los datos y la búsqueda de soluciones (Simon, 2017).

Modelado de la Solución

Es la actividad en la cual efectivamente se realiza el análisis de los datos, en esta etapa se aplica matemáticas, estadísticas y algoritmos a los datos, y se comunica los resultados como visualizaciones de datos (Vercellis, Business intelligence, 2009).

Modelado Basado en Fórmulas. Los modelos basados en fórmulas se pueden construir utilizando herramientas básicas como Excel, estos modelos generalmente definen variables de entrada y medidas de salida, necesarias para realizar una actividad analítica como la simulación, así una secuencia de fórmulas construye la ruta desde la entrada hacia la salida (Vercellis, Business intelligence, 2009).

Modelado Basado en Algoritmos. Los modelos basados en algoritmos se crean utilizando herramientas más sofisticadas, estos modelos aplican algoritmos para encontrar patrones en actividades en las cuales se generan datos, muchos algoritmos se encuentran disponibles para realizar tareas de minería de datos como clasificación, agrupación, asociación, secuenciación y pronóstico (Vercellis, Business intelligence, 2009).

Presentación de Resultados

La visualización de datos está en el corazón de la analítica empresarial y consiste en la comunicación de la información de valor que se obtiene luego de haber convertido grandes cantidades de datos complejos en datos comprensibles para la empresa, la visualización es un lenguaje de imágenes que está a la par del lenguaje escrito y hablado, y que presenta mediante graficas los resultados que se obtiene de la analítica de datos (Power & Sharda, 2015).

Comprensión y Acción

La analítica de datos por sí sola no ofrece ningún impacto comercial, pero en conjunto con las personas que toman acción basándose en las ideas que se obtienen a través del análisis de los datos, se puede generar un impacto de negocios; así el análisis de los datos se torna en un agente de valor cuando impulsa la conversación que da forma a la toma colectiva de decisiones y medidas empresariales (Mishra, Hazra, Tarannum, & Kumar, 2016).

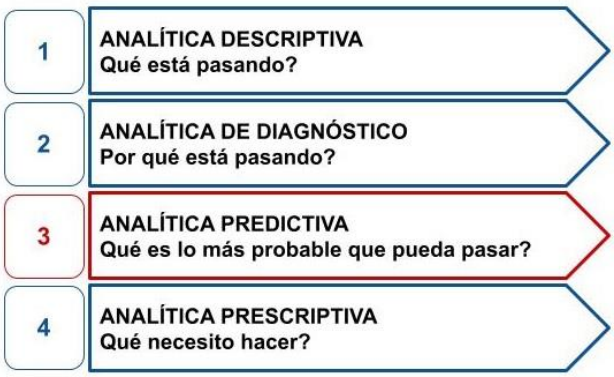
Tipos de Analítica de Datos

La analítica de datos se puede clasificar en cuatro categorías como se muestra en la Figura 2, resulta importante entender los diferentes tipos de analítica de datos, ya

que una de las características más relevantes de un analista de datos, es la habilidad de traducir datos en conocimiento útil para las personas que toma las decisiones.

Figura 2

Tipos de analítica de datos



Analítica Descriptiva

Esta analítica describe lo que sucedió en un periodo de tiempo del pasado, por lo que proporciona una visión y un contexto para poder entender los datos; así la analítica descriptiva permite visualizar las principales métricas de fracaso o éxito en un negocio como por ejemplo conocer información sobre ventas realizadas, ganancias y pérdidas (Balusamy, Nandhini Abirami, Kadry, & Gandomi, 2021).

Analítica de Diagnóstico

Esta analítica ayuda a comprender que está ocurriendo y porque está ocurriendo un suceso o fenómeno para poder tomar acciones correctivas, se ubica en un siguiente nivel de complejidad y valor añadido; en la analítica de diagnóstico el analista debe profundizar en los datos para encontrar la causa que origina un problema (Balusamy, Nandhini Abirami, Kadry, & Gandomi, 2021).

Analítica Predictiva

Esta analítica permite predecir en función de los datos lo que podría ocurrir en el futuro, como la previsión de una cantidad o la estimación de un punto en el tiempo en el que un evento podría ocurrir; esta analítica asciende un nivel más en términos de complejidad y valor añadido ya que utiliza técnicas estadísticas como el aprendizaje automático, modelado, inteligencia artificial y minería de datos (Balusamy, Nandhini Abirami, Kadry, & Gandomi, 2021).

Analítica Prescriptiva

Esta analítica va más allá del análisis de los datos y la predicción de resultados futuros, presenta una variedad de análisis de lo que podría suceder, lo cual constituye una guía sobre cómo las diferentes acciones podrían afectar al negocio, y a la mima vez sugiere la opción más óptima (Balusamy, Nandhini Abirami, Kadry, & Gandomi, 2021).
Conclusión. El presente trabajo de investigación se centra en el análisis de la proyección de la demanda, por lo que de acuerdo a los tipos de analítica de datos descritos, se elige la analítica predictiva para ser utilizada en el presente proyecto de investigación.

Modelado en Analítica de Datos

Para elaborar un modelo de analítica de datos se debe tomar en cuenta tres fases específicas: propósito, técnica y estructura; en la Figura 3, se muestra en detalle las opciones que se puede tener en cada fase.

Figura 3

Fases para el modelado en analítica de datos

Propósito

Generalmente un modelo de analítica de datos tiene uno de los tres propósitos que se describen a continuación.

- Encontrar Relaciones de Datos: Patrones ocultos en los datos que son útiles para obtener nuevo conocimiento y comprensión de situaciones.
- Encontrar Relaciones Causales: Aquellas relaciones entre los datos que son indicativas de causa y efecto.
- Encontrar Predictores: Son las variables en las relaciones causales que son útiles para predecir resultados futuros (Vercellis, Data mining, 2009).

Técnica

En un modelo de analítica de datos se puede utilizar las siguientes técnicas para el manejo de datos.

- Clasificación: Los datos se agrupan de acuerdo a un resultado que tiene en común.
- Asociación: Los datos se agrupan de acuerdo a las características que tiene en común.

- Pronóstico: En base a los datos históricos se realizan predicciones de tendencias y probabilidades futuras (Vercellis, Data mining, 2009).

Estructura

Un modelo de analítica de datos es una combinación de datos y software, que se utiliza para extraer inferencias de las relaciones de datos y para generar predicciones sobre el comportamiento de los sujetos de los datos; el software aplica la técnica de analítica de datos mediante la ejecución de algoritmos estadísticos o matemáticos (Vercellis, Mathematical models for decision making, 2009).

- Modelo: Un modelo se construye usando uno o más algoritmos.
- Algoritmo: Un algoritmo aplica una o más funciones estadísticas.
- Función Estadística: Una función estadística aplica una o más funciones matemáticas.
- Función Matemática: Una función matemática es una ecuación para la cual cualquier valor de entrada siempre produce un solo valor de salida correcto (Vercellis, Mathematical models for decision making, 2009).

Conclusión. Debido a que el objetivo del presente trabajo de investigación es el análisis de datos históricos para obtener tendencias y probabilidades futuras, se realiza la investigación de algoritmos analíticos relacionados a la técnica de pronóstico o también llamado forecast.

Aplicación del Estado del Arte

Debido a la gran diversidad de algoritmos estadísticos en el presente trabajo se realizó una revisión sistemática de la literatura relacionada a la *Técnica de Forecast de la Analítica de Datos,* como un medio para identificar y evaluar toda la investigación disponible y relevante sobre la técnica y algoritmos para analítica de datos predictiva.

Los criterios de inclusión definidos para la presente revisión sistemática de literatura son los siguientes:

- Se incluyen estudios y trabajos a partir del año 2009.
- El artículo hace referencia a proyectos realizados de toma de decisiones.
- El artículo describe información sobre modelos analíticos enfocados a predicciones.
- Se incluyen libros, artículos y conferencias, en el caso de conferencias deben tener una estructura de artículo científico y estar disponibles en la web.

Los criterios de exclusión definidos para la presente revisión sistemática de literatura son los siguientes:

- Artículos que no estén relacionados a analítica predictiva.
- Artículos que hagan referencia específicamente al estudio de un algoritmo analítico.
- Se excluye contenido que no esté en idioma inglés o español.

Al final de la revisión de literatura se ha seleccionado la siguiente cadena de búsqueda.

(“Prediction” or "Business Analytics" or "Business Intelligence Decision Support Systems") AND ("Companies" or "Industries" or "Manufacturing" or "Education") OR ("Real-time Systems ")

Para realizar la búsqueda de la cadena sugerida se utilizó las bases digitales SCOPUS y SPRINGER, debido a que cubren una amplia gama de publicaciones en el campo de la ciencia de la ingeniería y mantienen una base de datos completa y consistente. La aplicación de la cadena arroja alrededor de 213 documentos científicos en SCOPUS y 300 documentos en SPRINGER, filtrando información desde el 2009 y cuyos estudios son de revistas y congresos; sin embargo, se han seleccionado los estudios mediante 4 categorías: Comparación de algoritmos de minería de datos, Estudios de algoritmos de minería de datos en el área de medicina, Estudios de algoritmos de minería de datos en el área de educación y Estudios de algoritmos de

minería de datos en la industrial en general, escogiendo los más relevantes, los mismos que se presentan en la Tabla 1.

Tabla 1

Revisión de literatura en aplicación de algoritmos de minería de datos

Categoría	Autor, Año	Metodología	Objetivo del estudio
Comparación de algoritmos de minería de datos	(Wu, Yan, & Fan, 2012)	Series de Tiempo, Regresión Lineal	Desarrollo y comparación de diferentes modelos para pronosticar datos de ventas de nuevos productos con una tendencia de ventas creciente y múltiples entradas de predicción.
	(Tanwar & Kakkar, 2017)	Series de Tiempo	Comparación de rendimiento y estimación futura de datos de series de tiempo utilizando técnicas de minería de datos predictivos.
Estudios de algoritmos de minería de datos en el área de medicina	(Juang, Huang, Huang, Cheng, & Wann, 2017)	Series de Tiempo	Análisis predictivo de series de tiempo para pronosticar las visitas al servicio de emergencia de un hospital.
Estudios de algoritmos de minería de datos en el área de educación	(Kaur, Singh, & Josan, 2015)	Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree	Este estudio se enfoca en identificar a los estudiantes lentos mediante un modelo predictivo de minería de datos utilizando algoritmos basados en clasificación.
	(Saranya, Ayyappan, Narayanan, & Student, 2014)	Naive Bayes Algorithm	Análisis del pronóstico de crecimiento institucional representado gráficamente.
	(Alvarado & Jiménez, 2000)	Regresión lineal	Evaluación del rendimiento académico mediante la aplicación de algoritmos predictivos de minería de datos.
	(Sanchez, Sandoval, & Daza, 2017)	Regresión lineal	Análisis predictivo para la proyección del número de estudiantes que serán promovidos a primer semestre después de aprobar el curso de nivelación.
Estudios de algoritmos de minería de datos en la industria	(Francis & Kusiak, 2017)	Series de tiempo-ARIMA	Se desarrollan modelos que predicen el volumen de la demanda a partir de datos históricos.

	(Alon, Qi, & Sadowski, 2001)	Series de Tiempo (ARIMA) de Box-Jenkins y la regresión multivariante.	Proyección de ventas minoristas agregadas, mediante una comparación de redes neuronales artificiales y métodos tradicionales
	(Ngai, Hu, Wong, Chen, & Sun, 2011)	Series de tiempo, regresión lineal y árboles de decisión	Aplicación de técnicas predictivas de minería de datos en la detección de fraudes financieros, un marco de clasificación y una revisión académica de la literatura.
	(Choudhury & Jones, 2014)	Series de tiempo, Regresión Lineal	Predicción del rendimiento de cultivos.
	(Mazón, Guun, Arroyo, & Raura, 2020)	Series de tiempo-ARIMA	Arquitectura para la predicción de la demanda en la optimización de la producción dentro de una empresa de manufactura.

De acuerdo al análisis realizado se puede evidenciar que los estudios e investigaciones en torno a minería de datos se están expandiendo en varios sectores tanto en medicina, educación y la industria en general; además se revela que para análisis predictivos los modelos más utilizados a nivel general son los que utilizan métodos de Series de tiempo y Regresiones, por tal motivo se profundizará en estos dos métodos, y serán objeto de pruebas en el presente trabajo de investigación.

Modelos de Series de Tiempo

Los métodos de series de tiempo se utilizan en conjuntos de datos donde la variable que se quiere pronosticar depende del tiempo, el objetivo de los modelos de series de tiempo es identificar patrones en los datos históricos, con el propósito de hacer predicciones para períodos en el futuro; entre las aplicaciones de predicción utilizando series de tiempo se tiene ventas futuras de productos y servicios, tendencias en indicadores económicos y financieros, o secuencias de medidas características en ecosistemas (Vercellis, Time series, 2009).

Modelos muy Simples. El promedio móvil consiste en la media aritmética de las observaciones sucesivas de la serie de tiempo y permite identificar el componente de tendencia del conjunto de datos, además se puede desarrollar en una hoja de cálculo sin necesidad de conocimientos estadísticos avanzados; sin embargo el promedio móvil como modelo predictivo no es practico ya que se obtiene un nivel de precisión más bajo que el que se obtiene con otros métodos simples (Vercellis, Time series, 2009).

Modelos de Suavizado Exponencial (ETS). Los modelos de suavizado exponencial están entre los métodos de predicción más versátiles y precisos, a pesar de que en un inicio se formularon sobre una base empírica e intuitiva, actualmente con el avance de la investigación se han cimentado sobre bases teóricas más sólidas; los modelos de suavizado exponencial presentan diferentes niveles de complejidad dirigidos a identificar los componente de tendencia y estacionalidad de una serie de tiempo, además incluyen factores como la autocorrección y la adaptabilidad (Vercellis, Time series, 2009).

Modelos de Box Jenkins (ARIMA). Estos modelos consisten en la combinación de regresión dinámica y promedio móvil, al igual que los modelos de suavizado exponencial pueden encontrar patrones de tendencia y estacionalidad, usan auto correlación con el fin de encontrar el mejor ajuste de una serie de tiempo, mediante coeficientes de regresión que representan la relación entrada-salida y coeficientes de auto correlación que representan el patrón en el tiempo; así el pronóstico del modelo es un promedio ponderado de los valores históricos de la serie de tiempo, en general presentan un buen desempeño en los conjuntos de datos históricos grandes y estables, y al contrario no presentan tan buen desempeño en conjuntos de datos ruidosos e inestables (Pankratz, A Primer on ARIMA Models, 1991).

Modelos de Demanda Intermitente de Croston. Estos modelos son utilizados para predecir la demanda intermitente, trabajan con series de tiempo donde algunos valores

son cero y la ocurrencia de la demanda no es identificable; estos modelos implican dos pronósticos separados de suavizado exponencial simple, uno para predecir el tamaño de la demanda diferente de cero y otro para predecir el tiempo entre cada demanda diferente de cero; este modelo no pronostica cuándo ocurrirá la próxima orden pero en ocasiones puede presentar mejores resultados que otros modelos de series de tiempo (Shenstone & Hyndman, 2005).

Modelos de Regresión

Los modelos de regresión son el origen de la analítica predictiva y sirven para resolver problemas de pronóstico que presenten una relación causa-efecto, donde la variable dependiente se relaciona directamente con una o más variables independientes; este método consiste en el establecimiento de una ecuación matemática que representa la interacción entre las diferentes variables, de manera que los valores de salida se corresponden con una combinación de los valores de entrada (Moreira, Carvalho, & Horváth, 2018).

Modelos de Regresión Lineal. La regresión lineal analiza la relación existente entre las variables dependiente e independiente, cuando se cuenta con una sola variable independiente se tiene regresión lineal simple y cuando se cuenta con un conjunto de variables independientes se tiene una regresión lineal múltiple; el objetivo de este método es encontrar la línea recta que mejor se ajusta a los datos, es decir la recta que más se aproxime a la relación entre las variables de manera que la suma del cuadrado de los errores en relación a la línea de referencia estimada sea el menor posible, esto se conoce como estimación de mínimos cuadrados ordinarios (Moreira, Carvalho, & Horváth, 2018).

Modelos de Regresión Dinámica. Estos modelos combinan la regresión lineal simple con la capacidad para incorporar factores causales como precios, promociones e

indicadores económicos en el pronóstico, esta habilidad de usar términos dinámicos permite realizar un pronóstico con mayor precisión cuando los factores causales impulsan la demanda; un modelo de regresión dinámica proporciona una comprensión más profunda de las relaciones entre variables que afectan la demanda, ya que permite modelar escenarios que incluyen variables explicativas, con lo cual se puede obtener pronósticos para diferentes escenarios de las variables independientes seleccionadas como variables explicativas (Pankratz, Estimation and Forecasting, 1991).

Modelos de Regresión Vectorial. Estos modelos utilizan un poderoso algoritmo que combina la regresión lineal y un vector de soporte, cuyo objetivo es minimizar los coeficientes del vector para encontrar una línea apropiada que se ajuste a los datos, así la función Support Vector Regression (SVR) se concentra en encontrar los coeficientes y mas no en minimizar el error de la suma de los cuadrados, por lo que permite la flexibilidad de definir cuanto error es aceptable en el modelo, tanto a través de un margen de error como mediante el ajuste de la tolerancia a caer fuera de esa tasa de error (Smola & Schölkopf, 2004).

Conclusión. En el apartado de Series de Tiempo, el modelo de Promedio Móvil es un cálculo manual que no tiene sustento estadístico sino más bien está centralizado en la experiencia del analista, que es lo que actualmente realiza el centro de apoyo educativo y no tiene una buena precisión; el modelo de Suavizado Exponencial tiene buen desempeño y es fácil de aplicar, se basa en los componentes de nivel, tendencia y estacionalidad, y le da mayor importancia a los datos históricos más recientes que a los datos históricos más lejanos; el modelo de Box-Jenkins-ARIMA a más de basarse en los componentes de nivel, tendencia y estacionalidad, le da mayor importancia a la auto correlación que existe entre los datos, además tiende a tener un mejor desempeño en conjuntos de datos históricos largos y estables; y finalmente el modelo de Demanda Intermitente de Croston se utiliza para series de datos donde la demanda es

comúnmente cero y la ocurrencia de la demanda no es identificable. En el apartado de Regresión, los modelos proporcionan una comprensión más profunda de las relaciones entre variables que afectan la demanda, por tal motivo también serán sujetos de análisis en el presente proyecto de investigación.

Herramientas para Analítica de Datos

Debido a que existe una gran variedad y cantidad de herramientas tecnológicas en el mercado, se considera conveniente tomar como referencia los estudios realizados por empresas expertas en el análisis e investigación de los productos tecnológicos de diferentes fabricantes, a fin de filtrar las mejores soluciones y herramientas para el análisis de datos.

La corporación Gartner es una empresa de consultoría y de investigación de las tecnologías de la información, que se encarga de analizar las fortalezas y debilidades de soluciones y herramientas tecnológicas de diferentes fabricantes así como también de investigar las tendencias de mercado (Gartner, 2020).

El cuadrante mágico de Gartner permite a las empresas que contratan servicios y soluciones de tecnologías de la información conocer sobre los mejores productos o servicios en una determinada área tecnológica para poder tomar las mejores decisiones en sus procesos de transformación tecnológica digital; este cuadrante proporciona información sobre las características de los proveedores en cuanto a visión de mercado y poder de implementación (Gartner, 2020). La Figura 4 presenta el cuadrante mágico de Gartner que muestra la clasificación de los proveedores con las mejores soluciones y productos en el área de ciencia de datos según lo menciona Gartner en su informe de Enero de 2021.

Figura 4

Plataformas de ciencia de datos y aprendizaje de máquina

Nota. Tomado de (Maloney, 2021).

A continuación, se detalla las características de cada cuadrante y una breve descripción de los proveedores más representativos que se encuentran en cada uno de estos.

Los Líderes

Tienen un negocio establecido con una amplia base de clientes y un fuerte posicionamiento en el mercado, debido a que han demostrado tener una visión adecuada para mantener su posición de acuerdo a la evolución de los requisitos del mercado (Gartner, 2019).

Los Aspirantes

Tienen una gran capacidad de ejecución pero carecen de una visión sólida, aun no demuestran un entendimiento real de hacia dónde va el mercado, por lo que pueden convertirse en lideres si sus planes a futuro mejoran (Gartner, 2019).

Los Visionarios

Son negocios pequeños que tiene habilidad para visualizar como evolucionara un mercado por lo que pueden introducir nuevas tecnologías, servicios y modelos comerciales, sin embargo carecen de la capacidad para ejecutar sus ideas por completo o con éxito (Gartner, 2019).

Los Jugadores de Nicho

Usualmente son nuevos negocios que se enfocan y tiene éxito en un segmento de mercado específico por lo que no adquieren una visión global del mercado, además no se caracterizan por innovar o por superar a otros competidores y tienen una capacidad limitada de implementación y soporte (Gartner, 2019).

Alteryx. Presenta una sólida visión de la empresa y el producto, especialmente en la automatización de procesos y en el área conjunta de ciencia de datos y machine learning, también continua teniendo un nivel de ingresos más alto en comparación con los otros proveedores de este cuadrante, además de realizar una expansión significativa en el año 2019 con la adquisición de ClearStory Data and FeatureLabs (Piatetsky, 2020).

Databricks. Se basa en Apache Spark y ofrece una Plataforma Unificada de Análisis de Datos que comprende ciencia de datos, machine learning e ingeniería de datos; se movió al cuadrante de líderes debido a su sólida ejecución, crecimiento y a una

estructura de socios de más de 500 empresas, además es líder en habilitar la escalabilidad del machine learning (Piatetsky, 2020).

Dataiku. Se basa en la facilidad de uso, visión y gobernabilidad de los datos, además tiene la capacidad para proporcionar un ambiente para el trabajo colaborativo de múltiples tipos de usuarios desde ingenieros de datos y científicos de datos hasta usuarios comerciales, su producto principal es Data Science Studio (Piatetsky, 2020).

MathWorks. Es una plataforma de MATLAB totalmente integrada que permite la preparación de datos, la construcción de modelos, la simulación e implementación, y se basa en su adaptabilidad para el uso de las últimas tecnologías como el deep learning y el reinforcement learning, además de una gran capacidad de ejecución (Piatetsky, 2020).

SAS. SAS Visual Data Mining y Machine Learning son productos para el área de ciencia de datos y machine learning; los productos de SAS tienen un alto grado de preparación empresarial y brindan un alto valor comercial a los clientes, y a pesar de la competencia de las alternativas de código abierto SAS sigue siendo un competidor fuerte (Piatetsky, 2020).

TIBCO. TIBCO Data Science es una poderosa plataforma que se ha formado mediante la adquisición de empresas de bussiness intelligence como Jaspersoft y Spotfire, y proveedores de analítica y ciencia de datos como Insightful, Statistica y Alpine Data (Piatetsky, 2020).

IBM. Watson Studio es una plataforma que integra varios productos como Watson Machine Learning, SPSS Modeler, SPSS Statistics, IBM Decision Optimization e IBM Streams (Piatetsky, 2020).

DataRobot. Esta plataforma proporciona automatización para todo el proceso de analítica de datos, lo cual permite que los científicos de datos así como también los usuarios comerciales realicen el análisis de datos que necesitan (Piatetsky, 2020).

Domino. Es una plataforma con fortaleza en el área industrial y ofrece características mejoradas para el trabajo punto a punto ya sea en la nube o en sitio; puede servir como plataforma central en empresas con grupos de ciencias de datos grandes e independientes (Piatetsky, 2020).

Google. Google Cloud AI es una plataforma que incluye Cloud AutoML, BigQuery ML, y TensorFlow; además tiene su propio hardware TPU y una infraestructura de nube masiva (Piatetsky, 2020).

H2O.ai. Esta plataforma ofrece H2O Driverless AI que es un producto comercial y tambiénSparklingWater que es un producto de código abierto; además proporciona soporte al cliente de manera global y también cuenta con un canal para la comunicación de la comunidad en slack (Piatetsky, 2020).

KNIME. Ofrece una plataforma analítica de código abierto y un servidor como una extensión comercial, con funciones avanzadas que incluyen colaboración, automatización e implementación (Piatetsky, 2020).

Microsoft. Su producto principal es Azure Machine Learning que incluye productos de soporte como Azure ML Studio, Azure Data Factory, Azure HDInsight, Azure Databricks, Power BI; ha tenido avances tanto en visión como en implementación (Piatetsky, 2020).

RapidMiner. RapidMiner Studio cuenta con una edición gratuita y una edición comercial, también ofrece un servidor como una extensión empresarial para implementar y mantener modelos en un trabajo colaborativo; esta plataforma incluye RapidMiner Real-Time Scoring, RapidMinerRadoop, Rapidminer Auto Model y tiene una comunidad de usuarios grande y activa (Piatetsky, 2020).

Anaconda. Anaconda Enterprise es un entorno de desarrollo de ciencia de datos basado en el concepto de cuaderno interactivo, es adecuado para científicos de datos expertos que usan Python o R ya que se puede aprovechar el avance continuo y nuevas capacidades del código abierto (Piatetsky, 2020).

Altair. Su producto principal es Altair Knowledge Studio y está dirigido a grupos de analítica que desean aumentar las soluciones analíticas existentes con potentes capacidades de creación de perfiles y segmentación de datos o que deseen utilizar los arboles de decisiones patentados de Altair (Piatetsky, 2020).

El informe de Gartner sobre el Cuadrante Mágico para Plataformas de Ciencia de Datos y Aprendizaje de Maquina de Noviembre de 2019 incluyó solo a proveedores con productos comerciales y no consideró plataformas de código abierto como Python y R, a pesar de que son muy populares entre los científicos de datos y los profesionales del Machine Learning; sin embargo la mayoría de las herramientas expuestas tienen la opción de programación e inclusión de estas plataformas de código abierto (Piatetsky, 2020).

Capítulo III

Diseño del Modelo Analítico

En este capítulo, se detalla el análisis comparativo de las herramientas y algoritmos de analítica avanzada descritos en el capítulo dos, a fin de seleccionar lo más adecuado para el diseño de un modelo analítico predictivo que permita obtener la proyección de la demanda de estudiantes, con el objeto de mejorar la planificación de los procesos académicos del centro de apoyo educativo.

Situación Actual de la Planificación Académica

Actualmente el centro de apoyo educativo ingresa la información del negocio mediante macros de Excel, los cuales directamente guardan la información en este mismo software; y el proceso de planificación académica consiste en realizar macros y gráficos manuales en Excel para analizar los datos y tomar decisiones relacionadas con la oferta del servicio académico, sin embargo este proceso de planificación toma mucho tiempo de trabajo y no es efectivo.

Análisis de Herramientas de Ciencias de Datos

De acuerdo al primer filtro realizado mediante la presencia de herramientas de analítica de datos en el mercado, se puede mencionar que las herramientas de ciencia de datos y aprendizaje automático con un fuerte posicionamiento en el mercado y gran capacidad de ejecución son: *Alteryx, Dataiku, IBM, Databricks, MathWorks, SAS y TIBCO,* de acuerdo a la clasificación realizada por la empresa consultora Gartner reconocida a nivel mundial, en su último estudio publicado en Enero de 2021 (King, 2021).

Para la evaluación de las herramientas de analítica de datos se conformó un focus group con 4 profesionales de Business Analytics quienes trabajan como

consultores independientes o forman parte de una empresa de servicios en esta rama, la selección de los participantes se realizó mediante la técnica de muestra de conveniencia. Los participantes del focus group han trabajado en varios proyectos de inteligencia de negocios y analítica avanzada, y no tienen ninguna inclinación o preferencia por las herramientas analizadas, en la Tabla 2 se presenta el perfil profesional de los mismos.

Tabla 2

Participantes del focus group

Integrantes	**Rol**	**Perfil Profesional**
Diana Alulema	Moderador	Ingeniera en Electrónica y Redes de Información e investigador en el presente trabajo de investigación.
Karina Mazón	Participante	Máster en Sistemas de Información e Inteligencia de Negocios, con más de 7 años de experiencia, actualmente es Gerente de producción en una empresa de servicios en donde ha implementado alrededor de 15 proyectos de inteligencia de negocios y analítica avanzada.
Felipe Lemarie	Participante	Economista, con más de 10 años de experiencia en el análisis de datos y la realización de modelos cuantitativos y analíticos utilizando herramientas estadísticas y econométricas avanzadas.
Yoann Leny	Participante	Ingeniero en Sistemas, con más de 10 años de experiencia en proyectos de Big Data; ha realizado implementaciones de inteligencia de negocios y analítica avanzada en países como Australia, Francia y Estados Unidos.
Hugo Vera Flores	Participante	Master en Gerencia en Sistemas y Arquitectura Empresarial, actualmente se desempeña como consultor senior en Data Warehouse y Business Intelligence, además es fundador de la empresa Business Information Solutions S.A, la cual se dedica a proveer soluciones de Business Intelligence, Data Warehouse, Enterprise Performance Management y Big Data.

Para la evaluación se ha seleccionado en conjunto con el focus group, las características más importantes que deben estar presentes en la selección de una herramienta analítica; el detalle de la valoración de cada una de las características

evaluadas de las herramientas analíticas: *Alteryx, Dataiku, IBM, Databricks, MathWorks, SAS y TIBCO,* se muestra en la Tabla 3.

Tabla 3

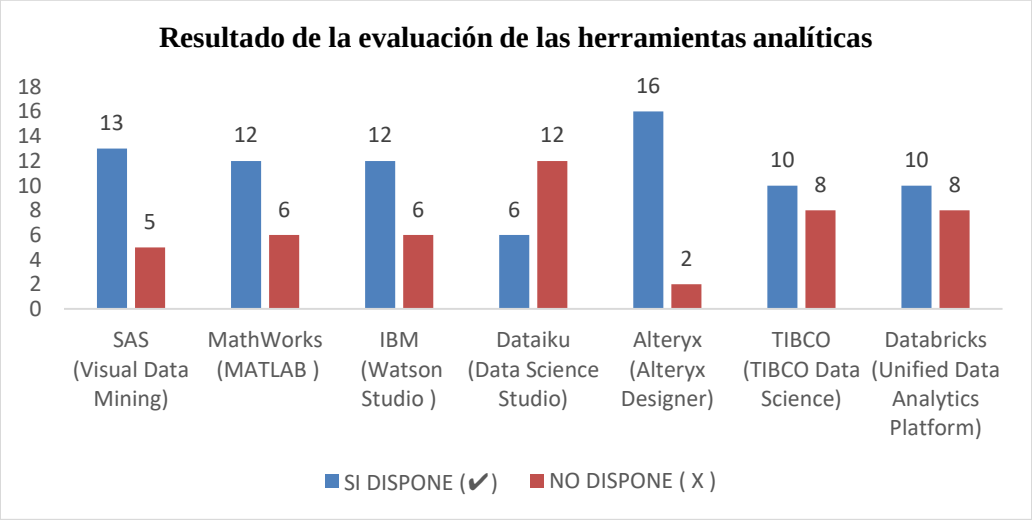
Detalle de evaluación de las herramientas analíticas

CARACTERÍSTICAS	**HERRAMIENTAS**						
	SAS (Visual Data Mining)	**MathWorks** (MATLAB)	**IBM** (Watson Studio)	**Dataiku** (Data Science Studio)	**Alteryx** (Alteryx Designer)	**TIBCO** (TIBCO Data Science)	**Databricks** (Unified Data Analytics Platform)
Licencia libre	X	X	X	X	X	X	X
Licencia para pruebas	✓	✓	✓	✓	✓	✓	✓
Multiplataforma	✓	✓	X	✓	X	✓	X
Combinación de modelos	✓	✓	X	X	✓	✓	X
Módulo incorporado de Técnicas Descriptivas	✓	✓	✓	✓	✓	✓	✓
Módulo incorporado de Técnicas Predictivas	✓	✓	✓	X	✓	✓	✓
Interfaz amigable	✓	✓	✓	X	✓	X	X
Visualización de datos	✓	✓	✓	✓	✓	✓	✓
Fácil de Configurar	X	✓	✓	X	✓	X	X
Fácil de Instalar	✓	✓	✓	✓	✓	✓	✓
Conversión fácil de datos	✓	✓	✓	X	✓	X	X
Integración con R	X	✓	X	X	✓	✓	✓
Personalización de módulos de predicción incorporados	X	X	X	X	✓	X	X
Creación de macros, módulos y la posibilidad de la reutilización de los mismos.	X	X	X	✓	✓	X	✓
Procesamiento de grandes volúmenes de información	✓	X	✓	X	✓	✓	✓
Soporte en Ecuador, Partners según lo enuncian sus páginas oficiales	✓	X	✓	X	✓	X	X
Casos de éxito expuestos en sus páginas oficiales	✓	X	✓	X	✓	X	✓
Validación del modelo	✓	✓	✓	X	✓	✓	✓

El resultado de la evaluación realizada por los participantes del focus group se visualiza en la Figura 5.

Figura 5

Resultado de la evaluación de las herramientas analíticas



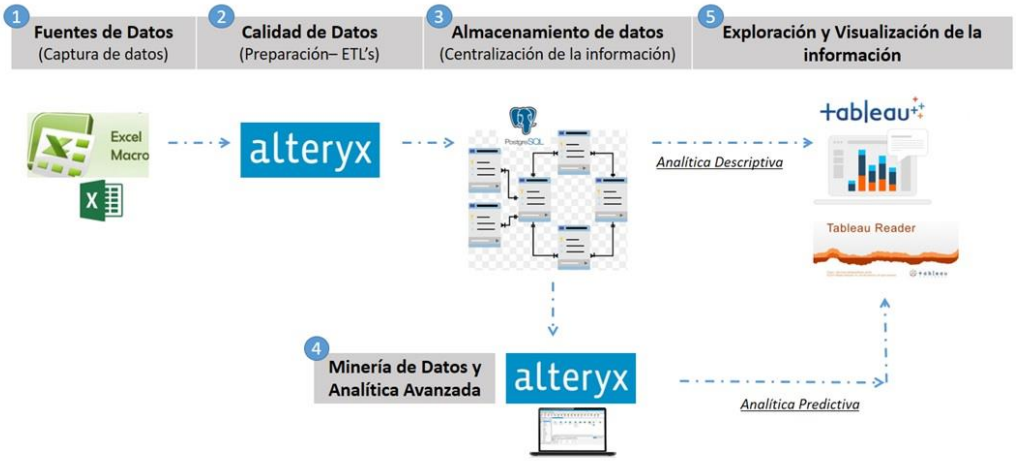
Mediante la evaluación de las herramientas analíticas realizada por el focus group, se pudo definir que la herramienta analítica Alteryx Designer obtuvo la mejor calificación tal como se indica en la Figura 5, por lo tanto se selecciona la herramienta analítica Alteryx Designer para ser utilizada en el diseño del modelo analítico del presente trabajo de investigación, debido a que presenta las características más importantes para un proyecto de analítica avanzada exitoso.

Arquitectura Propuesta de Toma de Decisiones

A continuación se plantea una arquitectura para la toma de decisiones dirigido a pequeñas y medianas empresas, que puede contribuir con la gestión del negocio y ayudar a realizar toma de decisiones mucho más rápidas y efectivas, esta arquitectura se visualiza en la Figura 6.

Figura 6

Arquitectura propuesta para el manejo de información



Primera Capa. Corresponde a las *Fuentes de Datos*, las cuales constituyen el núcleo de una solución de minería de datos y analítica avanzada, estas fuentes de datos principalmente suelen ser bases de datos transaccionales y archivos Excel, aunque lo más común en las pymes son los archivos Excel.

Segunda Capa. Corresponde a la *Calidad de Datos*, para la cual se plantea Alteryx Software herramienta de data quality y analítica avanzada, esta herramienta puede ser utilizada para realizar la fase de analítica avanzada y la misma puede ser aplicada para la preparación de datos; además está siendo utilizada por pequeñas y medianas empresas debido a que puede conectarse y limpiar datos de almacenes de datos, aplicaciones en la nube, hojas de cálculo y otras fuentes, unir fácilmente estos datos y luego realizar análisis (predictivos, estadísticos y espaciales) utilizando la misma interfaz de usuario intuitiva sin escribir código.

Tercera Capa. Corresponde al *Almacén de Datos*, en el cual se centraliza la información del negocio, para este almacenamiento se plantea PostgreSQL, puesto que es una base de datos libre y fácil de utilizar.

Cuarta Capa. Corresponde a los procesos de *Minería de Datos y Analítica Avanzada*, para la cual se plantea la herramienta Alteryx Software debido a que es una herramienta potente para los procesos de analítica avanzada y la mejor evaluada de acuerdo al análisis realizado en el presente trabajo de investigación.

Quinta Capa. Corresponde a la *Exploración y Visualización de la Información*, para lo cual se plantea Tableau Software herramienta de visualización de datos que ayuda a mejorar la comprensión de la información y a agilizar la toma de decisiones; esta herramienta lidera el mercado de inteligencia de negocios por 8 años consecutivos según lo menciona Gartner en su informe de Febrero de 2020.

Desarrollo del Modelo de Análisis Predictivo

En el presente trabajo de investigación para el desarrollo del modelo predictivo se aplicó la metodología CRISP-DM (Cross Industry Process - Data Mining) ya que es una de las metodologías más utilizadas para proyectos de minería de datos; en la Figura 7 se puede observar que la metodología CRISP-DM consiste en una serie de seis fases, el orden de ejecución de las fases no es estricto, de hecho en los proyectos a más de utilizarse iterativamente se puede avanzar y retroceder entre las fases si es necesario.

Figura 7

Fases de la metodología CRISP-DM

Comprensión del Negocio. Consiste en comprender el objetivo del proyecto en base a los requerimientos del negocio, para entonces definir un problema de minería de datos y diseñar un plan inicial que permita alcanzar el objetivo del proyecto (Villena-Román, 2016).

Comprensión de Datos. Consiste en la recolección inicial de datos y las actividades para familiarizarse con los mismos a fin de que el análisis de datos sea más preciso; el estudio de los datos permite identificar problemas en la calidad de los datos, encontrar conocimientos preliminares en los datos o encontrar subconjuntos interesantes de hipótesis de información oculta (Villena-Román, 2016).

Preparación de Datos. Consiste en todas las actividades para construir desde los datos iniciales el conjunto de datos final que será utilizado para el modelado, consta de la selección de tablas, registros y atributos, y de la transformación y limpieza de datos; las actividades para la preparación de los datos pueden ser realizadas muchas veces y

sin un orden específico a fin de adaptar los datos a las técnicas analíticas que se van a utilizar (Villena-Román, 2016).

Modelado. Consiste en seleccionar y aplicar la técnica de modelado que sea apropiada para el problema definido, con el correspondiente ajuste de sus parámetros a los valores óptimos; sin embargo pueden existir varias técnicas analíticas para un mismo problema de minería de datos, por lo que a menudo suele ser necesario volver a la fase de preparación de datos, ya que cada técnica puede tener requerimientos específicos sobre los datos (Villena-Román, 2016).

Evaluación. Consiste en evaluar el modelo desde la perspectiva de la analítica de datos y revisar los pasos ejecutados para la creación del mismo; además consiste en evaluar el modelo de acuerdo a los objetivos del negocio, ya que al concluir esta fase se debería poder tomar una decisión respecto al uso de los resultados de la minería de datos (Villena-Román, 2016).

Despliegue. Consiste en la finalización del proyecto y puede ser tan simple como la generación de un reporte que presente la información de valor, o tan complejo como la ejecución de un proceso automatizado y repetible de analítica de datos (Villena-Román, 2016).

Fase I: Comprensión del Negocio

En esta fase se realiza el análisis de la situación actual del negocio a fin de obtener conocimiento sobre los problemas y requerimientos del mismo, para así poder definir adecuadamente el objetivo del proyecto de minería de datos. Por lo cual se realizó una reunión de inicio del proyecto de analítica de datos con la directiva del centro de apoyo educativo, a fin de determinar las falencias en el proceso de planificación académica y definir el alcance del proyecto de analítica de datos; en esta reunión se enfatizó que el problema principal es la falta de conocimiento del número de

estudiantes a tener en los próximos meses, y en el alcance del proyecto de analítica de datos se definió que se diseñará un modelo analítico para la proyección de la demanda de estudiantes mediante la utilización de las herramientas tecnológicas Alteryx Designer para la proyección de la demanda, PostgreSQL como almacén de datos y Tableau Desktop para visualización de los resultados, cuyo objetivo es tener una predicción del número de estudiantes dentro de un periodo de 30 días.

Fase II: Comprensión de Datos

En esta fase se realiza el descubrimiento de las fuentes de datos relacionadas con el proyecto de analítica de datos, seguidamente se realiza el análisis de los datos y los metadatos a fin de lograr una buena comprensión de datos. Para el presente proyecto de investigación el centro de apoyo educativo proporcionó información relacionada al proceso académico con un historial de 4 años, que consta de tres archivos Excel correspondientes a la inscripción del estudiante, la contratación del servicio académico y la asistencia del estudiante a clases; cabe indicar que para cumplir con la Norma de Protección de Información los archivos Excel proporcionados por el centro de apoyo educativo no contienen información confidencial, en la Tabla 4 se detallan estas fuentes de información.

Tabla 4

Fuentes de datos para el proyecto de analítica de datos

Tema de Información	Canal	Tabla/Archivo
Inscripción	Excel	Información de inscripción del estudiante
Servicio	Excel	Detalle del servicio adquirido por el estudiante
Asistencia	Excel	Asistencia del estudiante al servicio adquirido

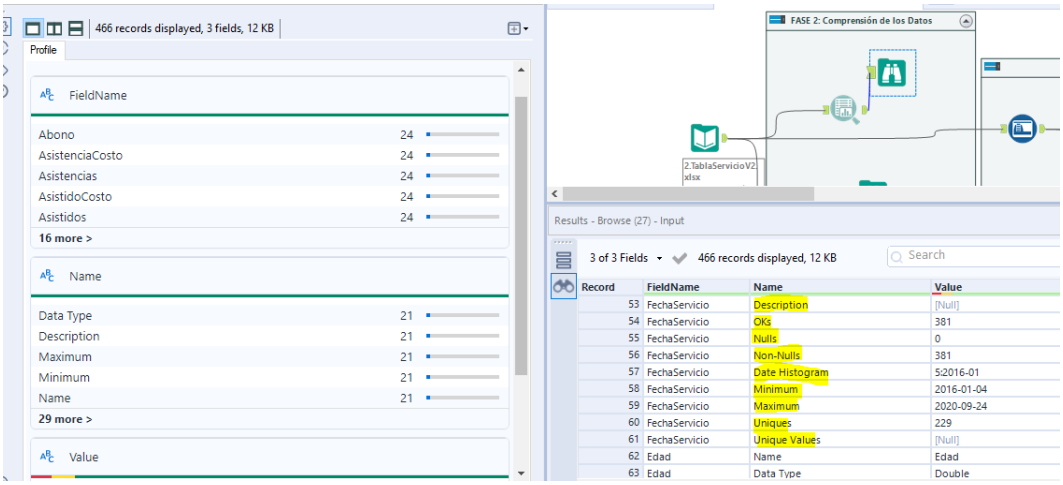
Recopilación de Datos. Las fuentes de datos elegidas para el proyecto de analítica de datos, son los archivos Excel proporcionados por el centro de apoyo educativo correspondientes a la inscripción del estudiante y a la contratación del servicio académico.

Descripción y Exploración de Datos. Luego de visualizar las fuentes de datos elegidas para el proyecto de analítica de datos, se procede con el análisis de cada una de ellas, lo cual implica realizar un perfilamiento de los datos y para este proceso se utilizó la herramienta analítica Alteryx que incorpora el componente *Data Investigation* para la exploración de la información.

El subcomponente *Basic Data Profile* analiza los datos y proporciona metadatos para cada columna de datos, en la Figura 8 se puede visualizar la presentación de los resultados de esta herramienta, donde para cada columna se tiene información como el tipo de dato, el tamaño del dato, el valor máximo y el valor mínimo, la cantidad de datos nulos y no nulos, la cantidad de datos válidos y en blanco, etc; este proceso permite analizar cómo está la calidad de datos de la información.

Figura 8

Perfilamiento de datos con Basic Data Profile



El subcomponente *Field Summary* analiza los datos y crea un resumen que contiene estadísticas descriptivas, los resultados se presentan en la Figura 9, donde se puede visualizar gráficamente las columnas de la información, con el fin de poder analizar anomalías o datos atípicos.

Figura 9

Análisis de datos con Field Summary

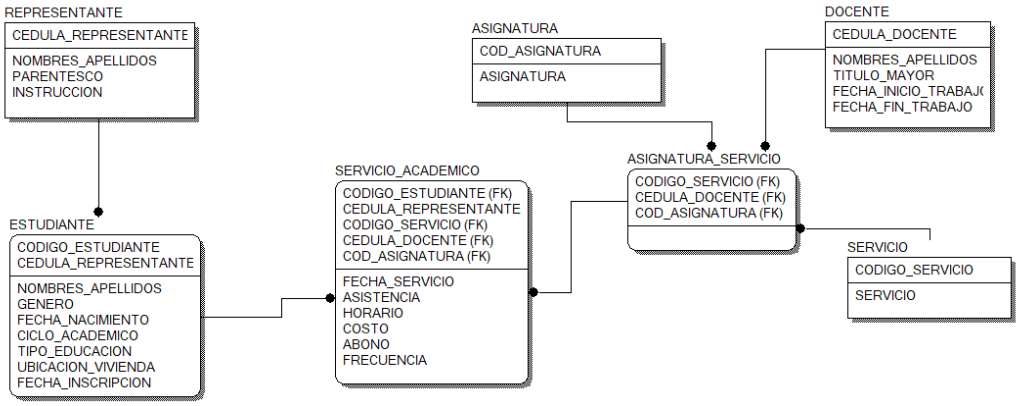
Fase III: Preparación de Datos

En esta fase se procede a la preparación de los datos para adaptarlos a las técnicas analíticas que se van a utilizar en la fase del modelado analítico, que en este caso particular son técnicas de predicción. Antes de proceder con la preparación de los datos, se realizó la creación de una base de datos para el proyecto de analítica de datos de manera que sea el nuevo input para el modelado analítico.

Base de Datos. Se diseñó un modelo entidad-relación en la herramienta CASE Erwin Data Modeler como se muestra en la Figura 10, este diseño servirá para la creación de la base de datos para el presente proyecto de analítica de datos, y también como una guía para que el centro de apoyo educativo implemente una base de datos transaccional en la que pueda almacenar de una manera más óptima su historia.

Figura 10

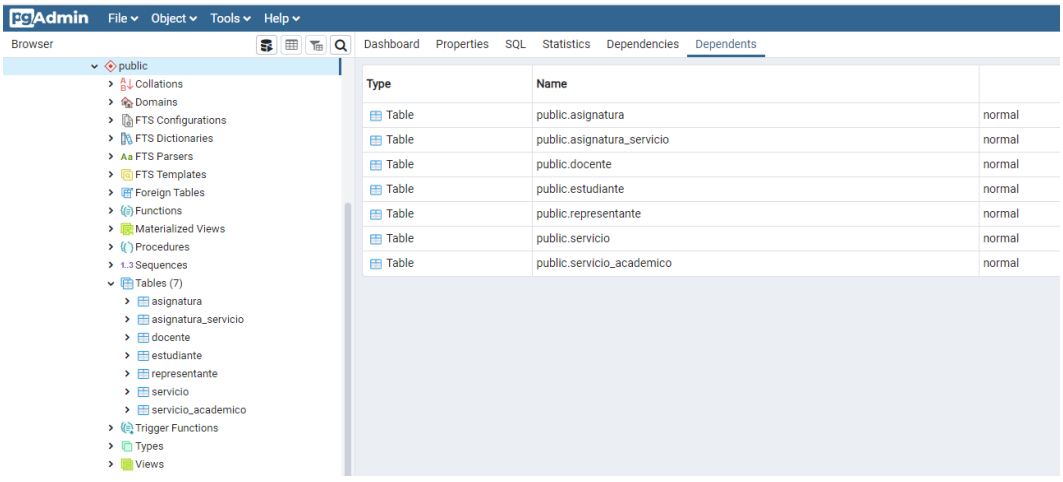
Modelo entidad-relación del centro de apoyo académico



Para la creación de la base de datos del proyecto de analítica de datos se utilizó PostgreSQL, una vez creada la base de datos se procedió con la generación de la estructura de los datos y las cargas de información proveniente de los archivos Excel, como se muestra en la Figura 11.

Figura 11

Estructura de datos en PostgreSQL



Posterior a la generación de la estructura de los datos y a la carga de información respectiva, se utilizó esta base de datos creada como fuente de datos para

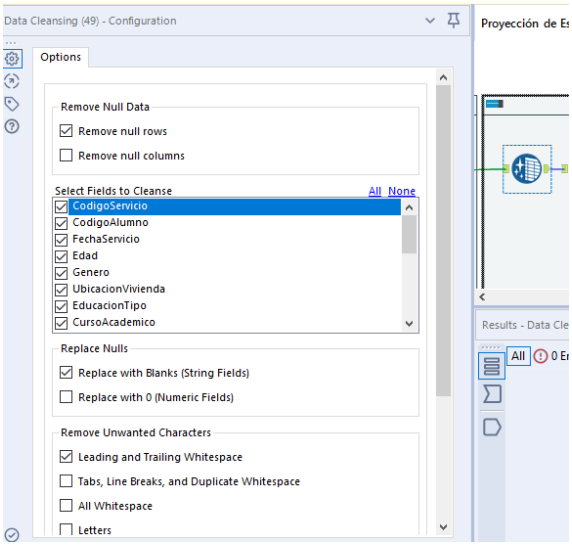
el proyecto de analítica de datos, y es desde aquí desde donde se realizaron los procesos de preparación de datos; cabe mencionar que por el tamaño de información que tiene actualmente el centro de apoyo educativo, no es necesario generar una base de datos analítica debido a que la transaccionalidad no es alta.

La preparación de datos depende de la técnica de modelado especifica que se va a aplicar, y consiste de tareas generales como la selección de datos, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato; para lo cual se utilizó la herramienta analítica Alteryx que incorpora los componentes *Preparation* y *Transform* específicos para estas tareas.

Limpieza de Datos. Para la limpieza de datos se aplicó el subcomponente *Data Cleansing* que soluciona problemas comunes de calidad de datos, con el fin de poder eliminar en las columnas de la data valores nulos, espacios en blanco, caracteres especiales; como se visualiza en la Figura 12, se procedió a eliminar registros nulos, remplazar datos nulos con espacios en blanco en los campos de cadena de caracteres y eliminar espacios en blanco del inicio o final del dato.

Figura 12

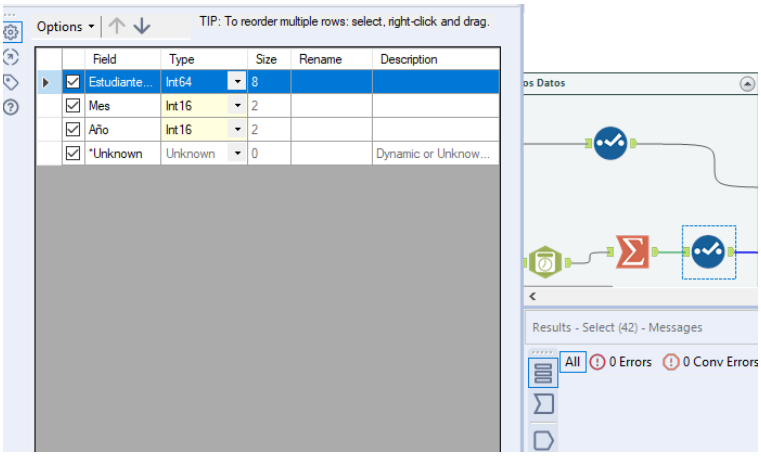
Calidad de datos con Data Cleansing



Formateo de Datos. Para las conversiones de tipos de datos se utilizó el subcomponente *Select* que permite modificar el tipo y tamaño de datos en una columna; como se visualiza en la Figura 13, los tipos de formateo de datos realizados fueron: la fecha y el código identificador de estudiante que se encontraban en la fuente de datos como cadena de caracteres.

Figura 13

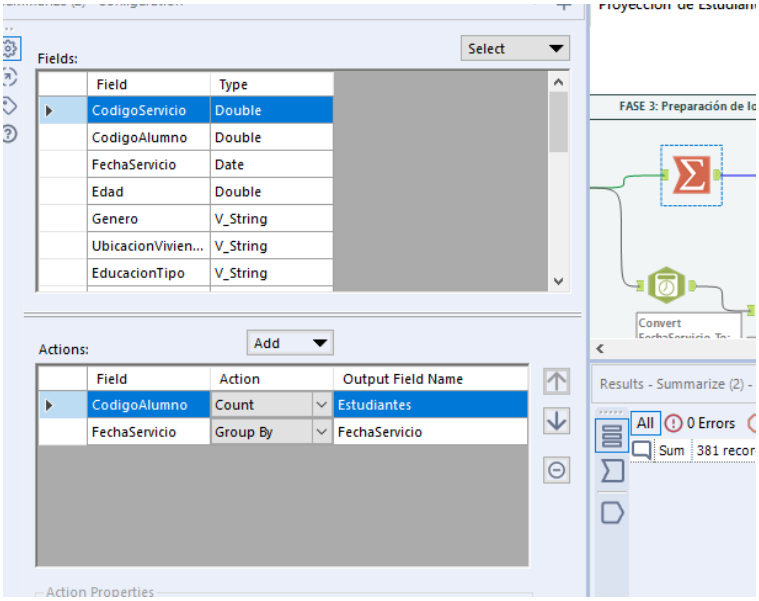
Conversión del tipo de dato con Select



Agrupación. Dentro de los procesos de preparación de datos también se realizaron agregaciones de la data con el subcomponente *Summarize* que permite agrupar una columna de datos por valores idénticos, de tal manera de tener un consolidado de la información; como se visualiza en la Figura 14, en los datos se realizó la agrupación por código del alumno y fecha del servicio, debido a que no era necesario toda la información de la base de datos.

Figura 14

Agrupación de campos con Summarize



Fase IV: Modelado

En esta fase se procede con la selección del algoritmo analítico y el diseño del modelo analítico predictivo, para posteriormente proceder con la aplicación del modelo analítico predictivo en la totalidad de la data y obtener la proyección de la demanda.

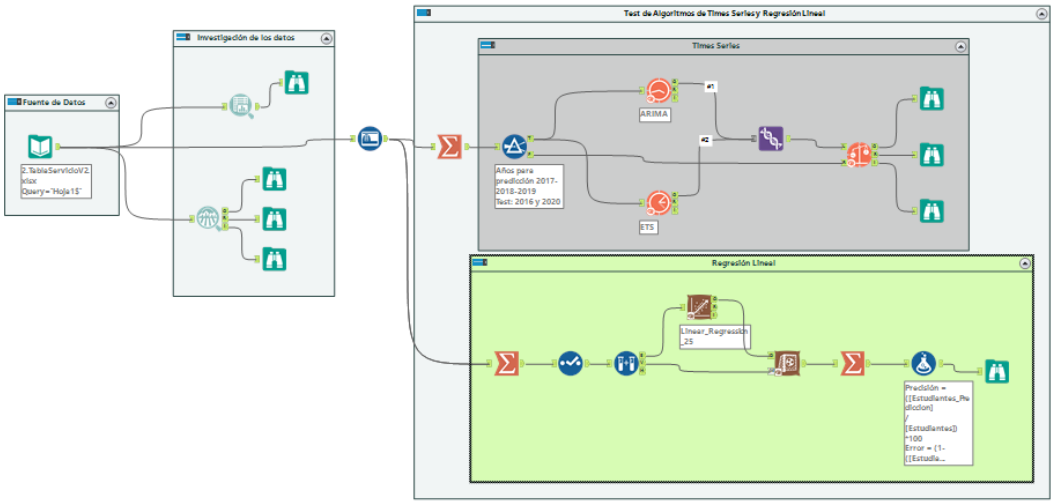
Análisis de Algoritmos Analíticos. Para la elección de los algoritmos analíticos a ser evaluados se consideró el estudio del estado del arte y el análisis teórico de los algoritmos analíticos, realizados en el capítulo dos sección 2.4. De acuerdo al estudio del estado del arte se pudo evidenciar que los algoritmos para pronóstico más utilizados por los científicos de datos son los algoritmos de *Series de Tiempo y Regresión Lineal*, según lo mencionan estudios e investigaciones realizadas referente a Business Analytics, Big Data, Data Mining, Data Science y Machine Learning a nivel mundial, lo cual constituye un primer filtro para la evaluación de los algoritmos de analítica de datos. Y de acuerdo al análisis teórico de los algoritmos analíticos de series de tiempo y regresión lineal, debido a sus características y ajuste con el negocio se optó por los

algoritmos analíticos de Regresión Dinámica ARIMA, Series de Tiempo Exponential Smoothing (ETS) y Regresión Lineal.

Para la evaluación de desempeño de los algoritmos analíticos se utilizó la herramienta de analítica avanzada Alteryx Designer y las fuentes de datos elegidas para el proyecto de analítica de datos. A fin de poder definir cuál es el mejor algoritmo analítico se realizó pruebas de precisión con los algoritmos mencionados: ARIMA, ETS y Regresión Lineal con un dataset de información del 80% de la data, como se visualiza en la Figura 15 se utilizó información para el entrenamiento 2017, 2018 y 2019, y para la evaluación 2016 y 2020.

Figura 15

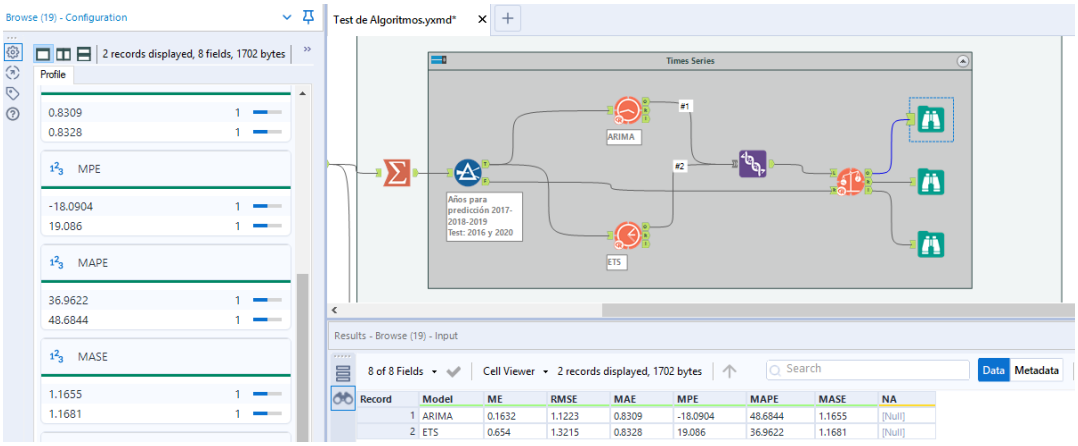
Prototipos de algoritmos de series de tiempo y regresión



En la Figura 16 se puede evidenciar la comparación realizada entre dos algoritmos de series de tiempo ARIMA y ETS, en estas evaluaciones se utilizó el error porcentual absoluto medio (MAPE) por ser el más utilizado para proyectos predictivos (Chase, 2016); como se puede observar el menor error lo presenta el algoritmo ETS con 36.96% frente al algoritmo ARIMA con 48.68%.

Figura 16

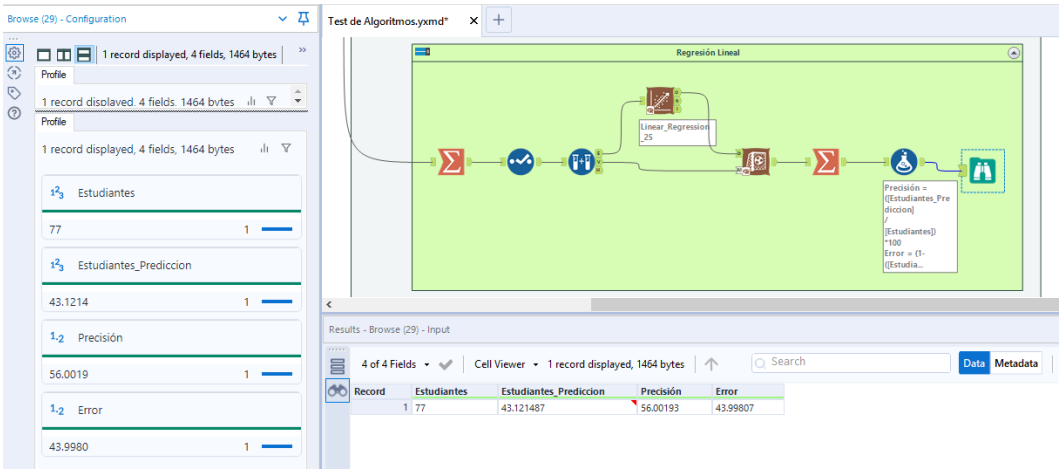
Resultados de los prototipos de algoritmos de series de tiempo



En la Figura 17 se puede evidenciar la aplicación del algoritmo de regresión lineal, el cual presenta un error del 43.99%, cabe mencionar que para realizar este prototipo se utilizó los mismos datos con los que se trabajó en los prototipos de series de tiempo, con el fin de que sean comparables.

Figura 17

Resultado del prototipo de algoritmo de regresión lineal



Los resultados de las pruebas de precisión realizadas para los algoritmos ARIMA, ETS y Regresión Lineal se describen en la Tabla 5.

Tabla 5

Resultado de la evaluación de los algoritmos analíticos

Modelo	**Algoritmo**	**Error Absoluto**	**% Precisión**
Series de Tiempo	ARIMA	48.68	51.32
	ETS	36.96	63.04
Regresión	Regresión Lineal	43.99	56.01

Mediante la evaluación de los algoritmos analíticos se pudo evidenciar que el algoritmo ETS o también llamado de Suavizado Exponencial, obtuvo la precisión más alta con respecto a ARIMA y a Regresión Lineal, tal como se indica en la Tabla 5, por lo tanto se selecciona el algoritmo *Time Series - ETS o Suavizado Exponencial* para ser utilizado en el diseño del modelo analítico por ser el algoritmo que más se ajusta al negocio.

Diseño del Modelo Analítico. El diseño del modelo analítico predictivo se realizó con los componentes predictivos de la herramienta de analítica avanzada Alteryx, donde se utilizó el algoritmo Time Series – ETS o Suavizado Exponencial ya que fue el algoritmo analítico mejor evaluado en el presente trabajo de investigación, debido a que presentó el menor error entre los algoritmos analíticos evaluados.

En las Figuras 18-19, se puede visualizar la aplicación del algoritmo ETS, en donde se configura los parámetros de variables como: nivel de frecuencia, tiempo de predicción, intervalos de confianza, etc, los mismos que permitieron realizar la predicción de la demanda. Cabe mencionar que para el aprendizaje del modelo solo se tomó información hasta el 2019-01-01, ya que el año 2020 no es adecuado para modelos predictivos debido a que por el tema de la pandemia existen muchos valores atípicos. Adicional la correlación entre las variables fue de 0.8, lo que muestra un patrón

lineal ascendente bastante fuerte, es decir a medida que el número de alumnos aumenta, también aumenta la demanda en la variable clases.

Figura 18

Aplicación del algoritmo ETS

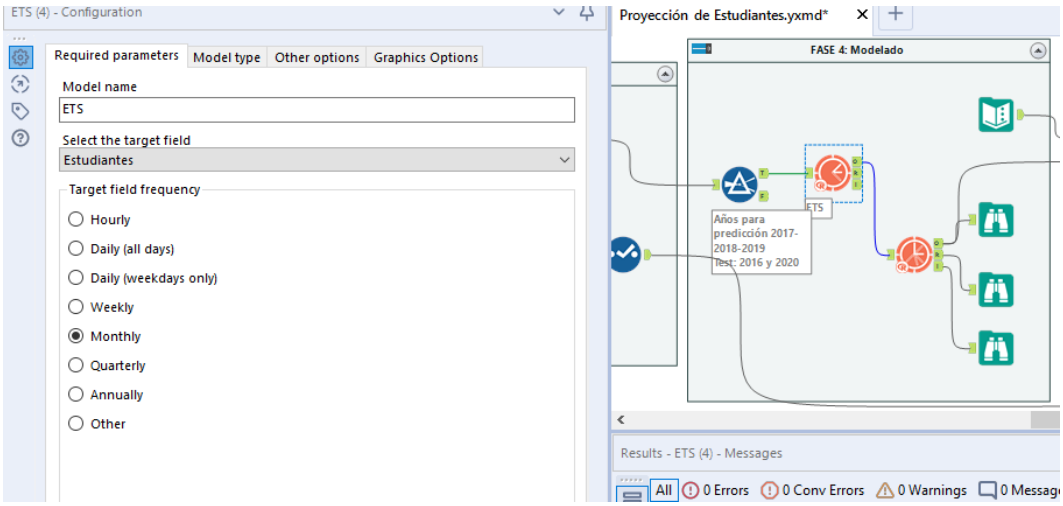
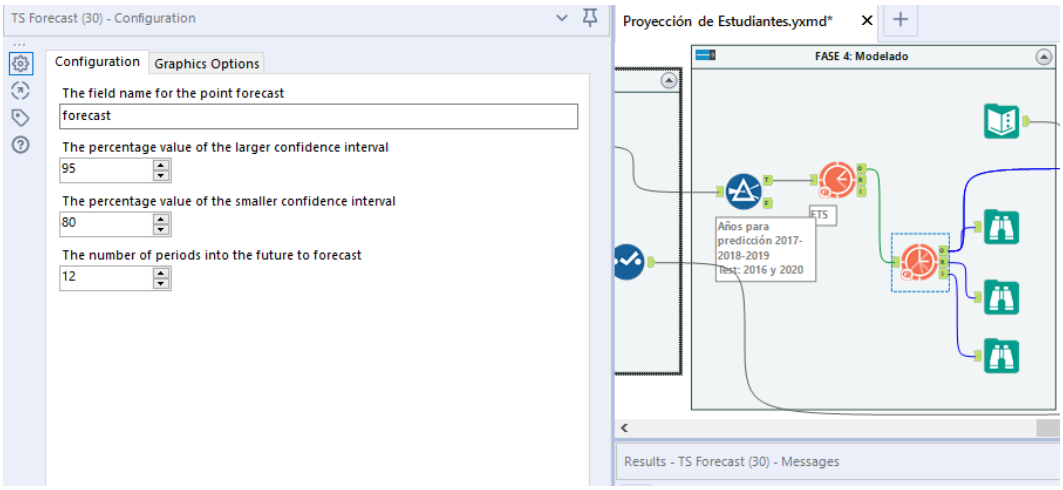


Figura 19

Aplicación del algoritmo ETS – Forecast



Como resultado se obtiene el modelo analítico predictivo para la proyección de la demanda de estudiantes, mediante la utilización de la herramienta de analítica

avanzada Alteryx y la aplicación del algoritmo Times Series - ETS con el uso de la técnica de pronóstico o también llamado forecast.

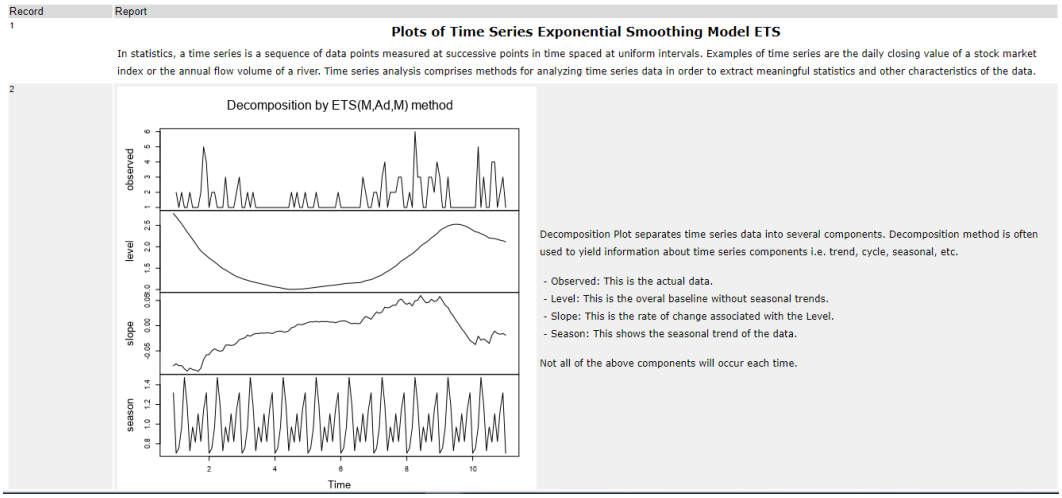
Fase V: Validación del Modelo Analítico

En esta fase se analiza y valida los resultados obtenidos mediante la aplicación del modelo analítico predictivo, así en la Figura 20 se puede observar las gráficas de los componentes de series de tiempo del algoritmo ETS con información sobre estadísticas significativas y otras características de los datos.

- Gráfico 1. Observado: son los datos reales.
- Gráfico 2. Nivel: es la línea de base general sin tendencias estacionales.
- Gráfico 3. Pendiente: es la tasa de cambio asociada con el nivel.
- Gráfico 4. Temporada: muestra la tendencia estacional de los datos, en la cual se puede ver que siguen un comportamiento similar atípico.

Figura 20

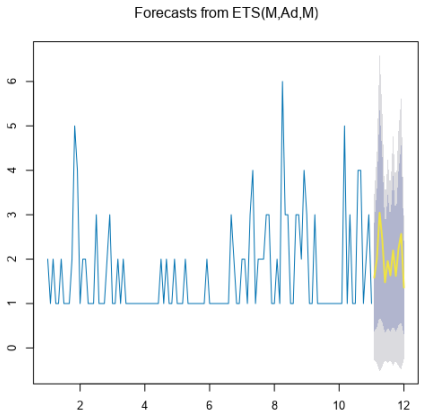
Gráfico de descomposición de ETS



En la Figura 21, se muestra el gráfico de los valores proyectados, es decir la predicción del número de estudiantes para un periodo de 12 meses, de manera que se pueda tener un vistazo de la proyección realizada.

Figura 21

Gráfico del forecast proyectado de ETS



The Forecast Plot shows the historic data in black and the expected value in blue. The orange in the plot shows the 90% confidence interval, and the yellow shows the 95% confidence interval.

En la Figura 22, se muestra una tabla con el resumen de errores del modelo ETS, en donde se puede visualizar que el error absoluto (MAPE) está en un 44.95%, lo que implicaría una precisión de 55,05%, la misma que es aceptable frente a la precisión de 33% que se maneja actualmente mediante cálculos en Excel.

Figura 22

Resumen de los resultados del modelo ETS

Summary of Time Series Exponential Smoothing Model ETS

Method:
ETS(M,N,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
0.0116521	0.9748936	0.68321	-21.2358269	44.9503147	0.8344922	0.0045628

Information criteria:

AIC	AICc	BIC
929.2891	932.1634	977.4314

Smoothing parameters:

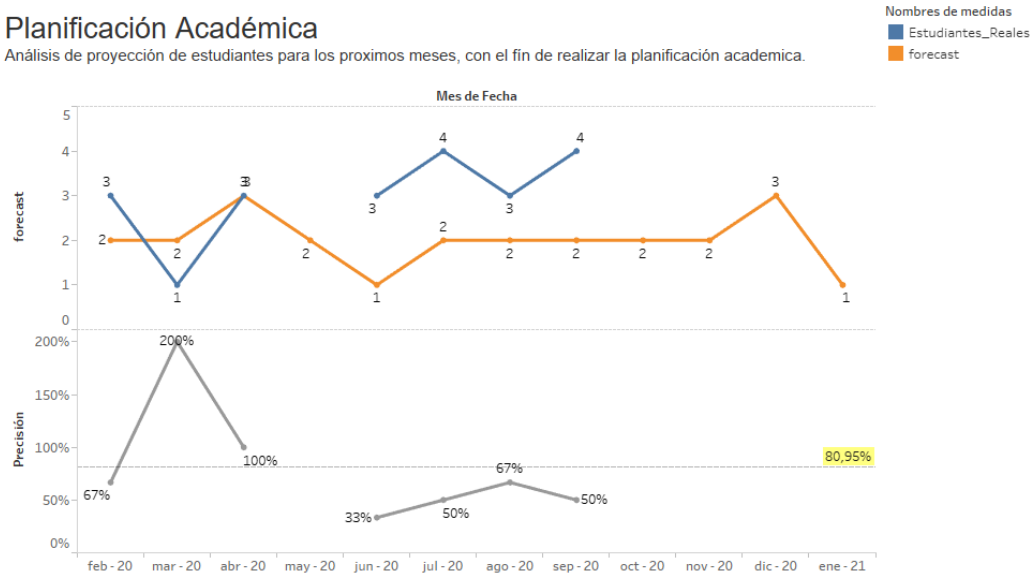
Parameter	Value
alpha	0.128252
gamma	1e-04

Fase VI: Despliegue de la Información

Finalmente, en esta fase se realiza el despliegue del descubrimiento de la información que se obtuvo con el modelo analítico predictivo, para lo cual se utilizó la herramienta de visualización Tableau, con el fin de lograr un análisis fácil y comprensible; así en la Figura 23, se puede visualizar en la primera gráfica un bechmark de inscripciones de estudiantes reales vs. forecast, y en la segunda gráfica la precisión del modelo.

Figura 23

Resultado del modelo predictivo con Tableau



Capítulo IV

Validación del Modelo Predictivo

En este capítulo se realiza el análisis y la validación de los resultados del modelo analítico predictivo mediante diferentes técnicas de minería de datos, para comprobar el modelo generado y determinar la validación exitosa del mismo.

Análisis de los Resultados del Modelo Predictivo

Validación del Modelo con Tableau

Para la evaluación del modelo analítico predictivo en la parte de la visualización se realizó un cálculo de precisión para conocer cuál es el porcentaje de aciertos de los procesos realizados, con lo cual se puede comprobar si el modelo analítico predictivo desarrollado es mejor que el proceso manual que se realiza actualmente, la fórmula para el cálculo de la precisión del pronóstico está determinada de la siguiente manera.

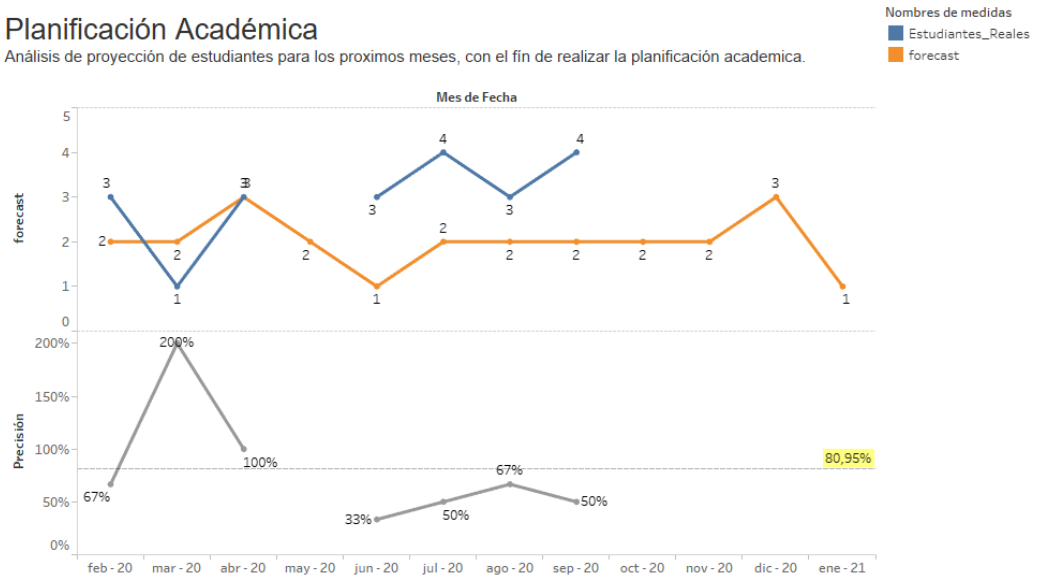
$$\boldsymbol{Precisión\ del\ Pronóstico}\ (\%) = (1 - \text{Error Absoluto}) * 100$$

$$\boldsymbol{Precisión\ del\ Pronóstico}\ (\%) = \left(1 - \frac{abs(estudiantes\ reales\ - forecast)}{estudiantes\ reales}\right) * 100$$

Como se mencionó anteriormente para el desarrollo del modelo analítico predictivo se utilizó un 80% de los datos para el aprendizaje del modelo y un 20% para la validación del mismo, dentro de este último segmento se predijo 12 meses para la evaluación del modelo, en donde se visualizó el comportamiento del modelo analítico predictivo generado con la herramienta analítica Alteryx; a continuación en la Figura 24, se muestra los resultados del modelo analítico mediante la herramienta de visualización Tableau, donde en la segunda gráfica se puede observar una precisión del pronóstico de 80.95%.

Figura 24

Precisión del pronóstico del modelo con Tableau



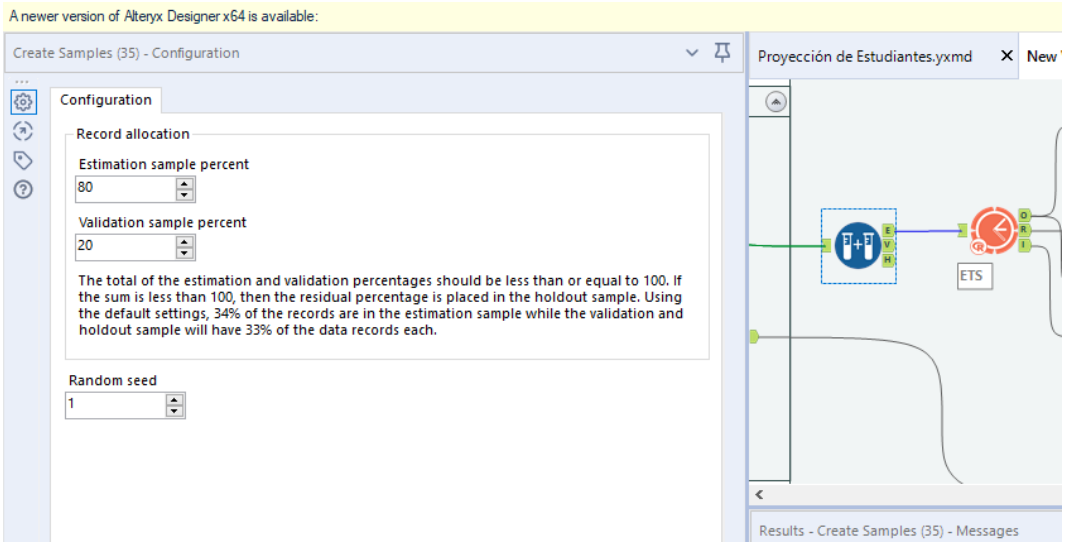
El promedio de precisión obtenido de la proyección de la demanda dentro de los 12 meses analizados fue de 80,95%, por tal motivo se puede concluir que el modelo generado con la herramienta analítica sobrepasa el valor de 33% de precisión que actualmente se tiene en la predicción de estudiantes de forma manual.

Validación del Modelo con Alteryx

Para la evaluación del modelo analítico predictivo se utilizó la herramienta analítica Alteryx que incorpora el subcomponente *Create Samples*, para dividir los datos de entrada en dos muestras aleatorias; donde se tomará para el aprendizaje del modelo el 80% de los registros de inscripciones de estudiantes y el otro 20% será para la validación, como se observa en la Figura 25.

Figura 25

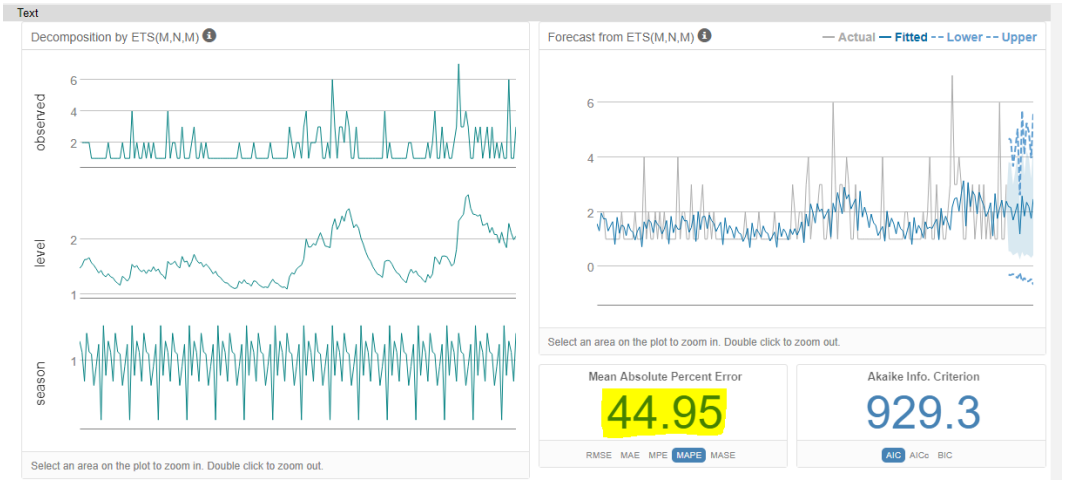
Porcentajes para aprendizaje y validación con Create Samples



En la Figura 26, se detalla los resultados del modelo analítico predictivo ETS, además se observa los errores y coeficientes generados.

Figura 26

Precisión del pronóstico del modelo con Alteryx



Con el error MAPE calculado anteriormente mediante un promedio de 12 meses se presenta una precisión del modelo de 80,95%, y con la validación de Alteryx

mediante una muestra del 80% de registros aleatorios se presenta un error de 44.95% lo que daría una precisión del modelo de 55.05%, este valor igualmente sigue siendo superior a la precisión de 33% que se tiene actualmente en la predicción de estudiantes de manera manual.

Conclusiones de la Validación del Modelo Predictivo

- Mediante una validación realizada aplicando cálculos con Tableau en una muestra de 12 meses analizados, utilizando datos desde el año 2016 hasta el año 2019 para el aprendizaje del modelo y el año 2020 para su validación, se obtuvo una precisión promedio de 80.95% utilizando el error MAPE.
- De la misma manera se realizó la validación del modelo utilizando la herramienta analítica Alteryx, mediante una muestra del 20% de datos para la evaluación y del 80% de datos para el aprendizaje, para lo cual se obtuvo una precisión de 55.05% con la observación del error MAPE.
- Finalmente se puede mencionar que mediante las diferentes validaciones realizadas, se obtuvo un promedio de precisión de 80.95% aplicando cálculos en Tableau, y una precisión de 55.05% mediante una validación de Alteryx; por lo tanto de acuerdo a los resultados de precisión obtenidos se puede decir que la hipótesis planteada es verdadera, ya que el modelo analítico predictivo diseñado mejora la precisión de 33% que se tiene con el uso de cálculos en Excel actualmente.

Conclusiones

- Alteryx es una herramienta para analítica de datos que se puede utilizar para ejecutar tanto procesos de calidad de datos como procesos de analítica avanzada, además presenta una gran capacidad de ejecución y un fuerte posicionamiento en el mercado de plataformas de ciencia de datos, según lo menciona Gartner empresa experta en el análisis e investigación de las fortalezas y debilidades de soluciones y herramientas tecnológicas de diferentes fabricantes.
- Para el diseño del modelo analítico predictivo se utilizó la herramienta analítica Alteryx y el algoritmo estadístico Times Series – ETS, ya que son los mejor valorados con respecto a las características de evaluación empleadas en el presente trabajo de investigación; además con la herramienta analítica Alteryx y el algoritmo estadístico Times Series – ETS se obtuvo el modelo predictivo de una manera rápida, fácil y con resultados que mejoran notablemente la predicción de estudiantes que se realiza actualmente.
- De acuerdo a las diferentes validaciones realizadas donde se obtuvo un promedio de precisión de 80.95% aplicando cálculos en Tableau y una precisión de 55.05% mediante una validación de Alteryx, esto debido que para los cálculos de Tableau no se tomó en cuenta para el aprendizaje del modelo el año 2020 por ser un año atípico por temas de la pandemia, mientras que con Alteryx se tomó una muestra aleatoria del 80%-20%; por tal motivo y de acuerdo a los resultados de precisión obtenidos se puede decir que la hipótesis planteada es verdadera, ya que el modelo analítico predictivo diseñado mejora la precisión de 33% que se tiene con el uso de cálculos en Excel actualmente.

- El modelo analítico diseñado en el presente trabajo de investigación puede servir como un prototipo modelo de análisis predictivo para otras entidades del sector educativo; ya que considera factores propios de la realidad educativa puede servir como base para análisis de datos que necesiten realizar proyección de la demanda estudiantil, proyección de la deserción estudiantil, etc.

Recomendaciones

- Por el momento el centro de apoyo educativo no necesita una base de datos analítica debido a que la transaccionalidad que maneja no es alta, sin embargo debería pensar la posibilidad de la implementación de un DWH (Data WareHouse), después de algunos años cuando los datos se incrementen y el tratamiento y análisis de datos sea inmanejable.
- Se debería utilizar e implementar en producción el modelo analítico predictivo diseñado en el presente trabajo de investigación, ya que de acuerdo a los resultados de precisión que se obtuvieron, podría mejorar la predicción de la demanda de estudiantes y por consiguiente mejorar la planificación académica del centro de apoyo educativo; de esta manera el modelo analítico predictivo podría contribuir en forma general con la automatización y optimización de los procesos académicos.
- Se recomienda implementar una base de datos transaccional y un sistema académico, para que el ingreso de las inscripciones de estudiantes y pagos del servicio académico se realice mediante un software especializado y no se siga realizando en Excel, a fin de lograr una mayor eficiencia en el negocio.
- Para obtener mejores resultados en la predicción de la demanda se recomienda definir cuidadosamente la frecuencia del pronóstico, ya que se debe considerar la propia naturaleza del fenómeno, y además se debe tener en cuenta que el error es mayor para pronósticos a largo plazo y también mientras mayor sea el nivel de agregación el error es más pequeño; adicional es sumamente importante no tomar en cuenta el año 2020 en los modelos predictivos debido a que es un año atípico por temas de pandemia, y por lo tanto la transaccionalidad de inscripciones de estudiantes no es comparable con otros años.

Referencias

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Retailing and Consumer Services, 8(3)*, 147-156, https://doi.org/10.1016/S0969-6989(00)00011-4.

Alvarado, J., & Jiménez, A. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema, 12(2)*, 248-525, https://www.redalyc.org/articulo.oa?id=72797059.

AWS. (2019). *Powering Business Transformation with Snowflake on AWS*. Obtenido de AWS: https://aws.amazon.com/es/partners/success/airlines-reporting-corporation-snowflake/

Balusamy, B., Nandhini Abirami, R., Kadry, S., & Gandomi, A. H. (2021). Big Data Analytics. En *Big Data: Concepts, Technology, and Architecture* (págs. 161-186). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119701859.

Chase, C. W. (2016). Performance Metrics. En *Next Generation Demand Management* (págs. 127-146). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119449591.

Choudhury, A., & Jones, J. (2014). Crop yield prediction using time series models. *Journal of Economic and Economic Education Research, 15(3)*, 53-68, https://www.abacademies.org/articles/jeeervol15no32014.pdf.

E3 Evolución Pymes. (2018). *Nuevas oportunidades de negocio para emprender basadas en big data.* Obtenido de ANDALUCÍA EMPRENDE: https://www.andaluciaemprende.es/wp-content/uploads/2018/07/ESTUDIO-NUEVAS-OPORTUNIDADES-DE-NEGOCIO-PARA-EMPRENDER-BASADAS-EN-BIG-DATA.pdf

ESPAE Graduate School of Management. (2018). *Global Entrepreneurship Monitor Ecuador 2017.* Obtenido de ESPAE Escuela de Negocios: https://www.espae.edu.ec/wp-content/uploads/2021/02/GemEcuador2017.pdf

Francis, H., & Kusiak, A. (2017). Prediction of engine demand with a data-driven approach. *Procedia Computer Science, 103*, págs. 28–35, https://doi.org/10.1016/j.procs.2017.01.005.

Gartner. (2019). *How markets and vendors are evaluated in Gartner Magic Quadrants*. Obtenido de Gartner: https://www.gartner.com/en/documents/3956304/how-markets-and-vendors-are-evaluated-in-gartner-magic-q

Gartner. (2020). *Positioning technology players within a specific market*. Recuperado el 20 de Noviembre de 2020, de Gartner: https://www.gartner.com/en/research/methodologies/magic-quadrants-research

Iqbal, M., Kazmi, S. H., Manzoor, A., Soomrani, A. R., Butt, S. H., & Shaikh, K. A. (2018). A study of big data for business growth in SMEs: Opportunities & challenges. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET).* https://doi.org/10.1109/ICOMET.2018.8346368.

iTahora. (2019). *La estrategia de Ayasa se sustenta en analítica avanzada*. Obtenido de iTahora: https://itahora.com/2019/04/23/la-estrategia-de-ayasa-se-sustenta-en-analitica-avanzada/

Juang, W.-C., Huang, S.-J., Huang, F.-D., Cheng, P.-W., & Wann, S.-R. (2017). Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *7(11)*, http://doi.org/10.1136/bmjopen-2017-018628.

Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science, 57*, 500–508, https://doi.org/10.1016/j.procs.2015.07.372.

King, T. (04 de Marzo de 2021). *What's Changed: 2021 Gartner Magic Quadrant for Data Science and Machine Learning Platforms*. Obtenido de Solutions Review: https://solutionsreview.com/business-intelligence/whats-changed-2021-gartner-magic-quadrant-for-data-science-and-machine-learning-platforms/

Logicalis Spain. (2019). *La UE, más cerca de convertirse en una universidad Data Driven gracias al proyecto de analytics implementado con Logicalis Spain*. Obtenido de LOGICALIS Architects of Change: https://recursos.es.logicalis.com/caso-de-exito-universidad-europea-0-0

Maloney, R. (11 de Abril de 2021). *H2O.ai logra gran posicionamiento en integridad de visión en el cuadrante Visionarios del Cuadrante Mágico de Gartner 2021 para Data Science y Machine Learning*. Obtenido de H2O.ai: https://www.h2o.ai/blog/h2o-ai-logra-gran-posicionamiento-en-integridad-de-vision-en-el-cuadrante-visionarios-del-cuadrante-magico-de-gartner-2021/

Mazón, I. K., Guun, S., Arroyo, R., & Raura, G. (2020). Architecture for demand prediction for production optimization: a case study. *Advances in Intelligent Systems and Computing, 1066*, págs. 1–11, https://doi.org/10.1007/978-3-030-32022-5_1.

Mishra, B. K., Hazra, D., Tarannum, K., & Kumar, M. (2016). Business Intelligence using Data Mining techniques and Business Analytics. *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, (págs. 84-89). https://doi.org/10.1109/SYSMART.2016.7894496.

Moreira, J. M., Carvalho, A., & Horváth, T. (2018). Regression. En *A General Introduction to Data Analytics* (págs. 161-185). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119296294.

Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decision Support Systems, 50(3)*, 559–569, https://doi.org/10.1016/j.dss.2010.08.006.

Pankratz, A. (1991). A Primer on ARIMA Models. En *Forecasting with Dynamic Regression Models* (págs. 24-81). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118150528.

Pankratz, A. (1991). Estimation and Forecasting. En *Forecasting with Dynamic Regression Models* (págs. 324-341). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118150528.

Piatetsky, G. (Febrero de 2020). *Leaders, changes, and trends in Gartner 2020 Magic Quadrant for data science and machine learning platforms*. Obtenido de KDnuggets: https://www.kdnuggets.com/2020/02/gartner-mq-2020-data-science-machine-learning.html

Power, D. J., & Sharda, R. (2015). Business Intelligence and Analytics. En *Wiley Encyclopedia of Management* (págs. 1-4). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118785317.weom070011.

Sanchez, T., Sandoval, I., & Daza, W. (2017). Modelo de proyección para la oferta de cupos en el primer año de las carreras de ingeniería de la Escuela Politécnica Nacional. *The 15th LACCEI International Multi-Conference for Engineering, Education, and Technology.* https://doi.org/10.18687/LACCEI2017.1.1.452.

Saranya, S., Ayyappan, R., Narayanan, K., & Student, M. (2014). Student progress analysis and educational institutional growth prognosis using data mining. *International Journal of Engineering Sciences & Research Technology, 3(4)*, https://www.academia.edu/6993773/Student_Progress_Analysis_and_Educational_Institutional_Growth_Prognosis_Using_Data_Mining.

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting, 24(6)*, 389–402, https://doi.org/10.1002/for.963.

Simon, P. (2017). A Framework for Agile Analytics. En *Analytics : the agile way* (págs. 113-130). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119424215.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14(3)*, 199–222, https://doi.org/10.1023/b:stco.0000035301.49549.88.

Tanwar, H., & Kakkar, M. (2017). Performance comparison and future estimation of time series data using predictive data mining techniques. *2017 International Conference on Data Management, Analytics and Innovation*, (págs. 9-12). https://doi.org/10.1109/ICDMAI.2017.8073477.

Vercellis, C. (2009). Business intelligence. En *Business Intelligence: Data Mining and Optimization for Decision Making* (págs. 3-19). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470753866.

Vercellis, C. (2009). Data mining. En *Business Intelligence: Data Mining and Optimization for Decision Making* (págs. 77-94). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470753866.

Vercellis, C. (2009). Mathematical models for decision making. En *Business Intelligence: Data Mining and Optimization for Decision Making* (págs. 65-75). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470753866.

Vercellis, C. (2009). Time series. En *Business Intelligence: Data Mining and Optimization for Decision Making* (págs. 187-219). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470753866.

Villena-Román, J. (2016). *CRISP-DM: La metodología para poner orden en los proyectos*. Obtenido de Sngular: https://www.sngular.com/es/data-science-crisp-dm-metodologia/

Wu, L., Yan, J., & Fan, Y. (2012). Data mining algorithms and statistical analysis for sales data forecast. *2012 Fifth International Joint Conference on Computational Sciences and Optimization*, (págs. 577-581). https://doi.org/10.1109/CSO.2012.132.