



UNIVERSIDAD CATÓLICA
de Colombia

**PROPUESTA DE METODOLOGÍA PARA EL DESARROLLO DE PROYECTOS
DE ANALÍTICA PRESCRIPTIVA A PARTIR DE UN METAANÁLISIS**

JOHN NELSON CRISTANCHO LOPEZ

**PROGRAMA DE INGENIERÍA DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD CATÓLICA DE COLOMBIA
BOGOTÁ, MAYO
2021**

PROPUESTA DE METODOLOGÍA PARA EL DESARROLLO DE PROYECTOS
DE ANALÍTICA PRESCRIPTIVA A PARTIR DE UN METAANÁLISIS

JOHN NELSON CRISTANCHO LOPEZ

Trabajo de Grado para optar al título de
INGENIERO DE SISTEMAS

Asesor: LEONEL JOSE PAREDES MADRID Ph.D
ljparedes@ucatolica.edu.co

PROGRAMA DE INGENIERÍA DE SISTEMAS
FACULTAD DE INGENIERÍA
UNIVERSIDAD CATÓLICA DE COLOMBIA
BOGOTÁ, MAYO
2021



Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Advertencia](#).

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y construir a partir del material

La licenciente no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciente.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia](#).

<https://creativecommons.org/licenses/by-nc/4.0/deed.es>

NOTA DE ACEPTACIÓN

Jurado

Jurado

Leonel José Paredes Madrid, Ph.D
Asesor

Bogotá, Mayo 17, 2021

AGRADECIMIENTOS

Agradezco a Dios, a mi padre, a mi madre, a mi familia y amigos, que estuvieron siempre apoyándome en este largo camino. A mi tutor por su guía y al ingeniero Holman Bolívar por su inestimable ayuda en la gestación de este proyecto.

TABLA DE CONTENIDOS

GLOSARIO	11
RESUMEN	12
INTRODUCCIÓN	14
1. PLANTEAMIENTO DEL PROBLEMA	15
1.1. Descripción del problema.....	15
1.2. Formulación del problema	16
2. JUSTIFICACIÓN	17
3. OBJETIVOS	18
3.1. Objetivo general	18
3.2. Objetivos específicos.....	18
4. ALCANCES Y LIMITACIONES	19
4.1. Alcances.....	19
4.2. Limitaciones.....	19
5. MARCO CONCEPTUAL	20
5.1. Analítica Prescriptiva.....	20
5.2. Minería De Datos	21
5.3. Metodología KDD.....	22
5.4. Metodología CRISP-DM	24
5.5. Metodología SEMMA	26
6. MARCO TEORICO	27
6.1. Métodos estadísticos para combinación de estudios.....	27
6.1.1. Modelos de efectos fijos	27
6.1.2. Modelo de efectos aleatorios	27
6.1.3. Heterogeneidad.....	27
6.1.4. Meta-regresión.....	27
6.1.5. Análisis de subgrupos y sensibilidad	28
6.2. Calculo del tamaño del efecto	28
6.3. Evaluación de sesgo de publicación	28
7. ESTADO DEL ARTE	29
8. METODOLOGIA	31
8.1. Metaanálisis.....	31

9. DESARROLLO DEL PROYECTO	33
9.1. Formulación del problema	33
9.2. Búsqueda de información.....	33
9.3. Análisis estadístico	34
10. CONCLUSIONES	40
REFERENCIAS.....	42
ANEXOS.....	44

LISTA DE FIGURAS

<i>Figura 1. Etapas del modelo de proceso KDD. [10]</i>	22
<i>Figura 2. Etapas del modelo de proceso CRISP-DM. [11]</i>	24
<i>Figura 3. Etapas del modelo de proceso SEMMA. [12]</i>	26
<i>Figura 4. Aplicación de Criterios Inclusión / Exclusión. Años vs Numero de artículos. [El Autor]</i>	33
<i>Figura 5. Áreas de investigación presentes en los artículos. Área vs Numero de artículos. [El Autor]</i>	34
<i>Figura 6. Porcentajes de fases KDD utilizada en los artículos. [El Autor]</i>	35
<i>Figura 7. Porcentajes de fases CRISP-DM utilizada en los artículos. [El Autor]</i>	37
<i>Figura 8. Porcentajes de fases SEMMA utilizada en los artículos. [El Autor]</i>	38
<i>Figura 9. Uso de fases en los artículos KDD, CRISP_DM, SEMMA vs Numero de fases. [El Autor]</i>	39
<i>Figura 10. Metodología propuesta para proyectos de Analítica Prescriptiva. [El Autor]</i>	40

LISTA DE TABLAS

<i>Tabla 1. Etapas del modelo de proceso KDD presentes en los artículos. [El Autor].....</i>	<i>35</i>
<i>Tabla 2. Consolidado de fases KDD por artículo. [El Autor].....</i>	<i>35</i>
<i>Tabla 3. Etapas del modelo de proceso CRISP-DM presentes en los artículos. [El Autor]</i>	<i>36</i>
<i>Tabla 4. Consolidado de fases CRISP-DM por artículo. [El Autor].....</i>	<i>36</i>
<i>Tabla 5. Etapas del modelo de proceso SEMMA presentes en los artículos. [El Autor]</i>	<i>37</i>
<i>Tabla 6. Consolidado de fases SEMMA por artículo. [El Autor]</i>	<i>38</i>

LISTA DE ANEXOS

ANEXO 1: Artículos para el Metaanálisis	44
---	----

GLOSARIO

ANALÍTICA PRESCRIPTIVA: Es la parte de la analítica que se dedica a la automatización de la toma de decisiones, apoyándose principalmente en los sistemas de gestión de reglas de negocio y la optimización matemática.

CRISP-DM: Son las siglas de *Cross-Industry Standard Process for Data Mining*, es un método probado para orientar trabajos de minería de datos.

DOI: *Digital Object Identifier* es un enlace permanente en forma de código alfanumérico que identifica de forma única un contenido electrónico.

KDD: *Knowledge Discovery from Databases*, es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos.

METAANALISIS: Es una estrategia de revisión sistemática que pretende responder a diversas cuestiones y que es especialmente útil cuando los resultados de varios estudios son discordantes con respecto a la magnitud de la dirección de un efecto, cuando los tamaños muestrales de los estudios individuales es demasiado pequeño para detectar dicho efecto como significativo, o cuando un ensayo clínico de gran tamaño es demasiado costoso en términos económicos o de tiempo.

METODOLOGIA: Es un conjunto de procedimientos racionales utilizados para alcanzar el objetivo o la gama de objetivos que rige una investigación científica, una exposición doctrinal o tareas que requieran habilidades, conocimientos o cuidados específicos.

MINERIA DE DATOS: Es un proceso para detectar información de conjuntos grandes de datos, de la manera más automáticamente posible. Su objetivo es encontrar patrones, tendencias o reglas que expliquen el comportamiento de los datos en un contexto específico.

SEMMA: *Sample, Explore, Modify, Model, and Assess*, conjunto de herramientas funcionales de SAS Enterprise Miner para llevar a cabo las tareas centrales de la minería de datos.

RESUMEN

Actualmente el mundo se encuentra a puertas de la denominada cuarta revolución industrial o también llamada Industria 4.0, donde la comunicación y los grandes volúmenes de datos están en auge, datos que manejados correctamente pueden ayudar en el proceso de toma de decisiones, pero que para llegar hasta esa instancia deben utilizarse gran variedad de herramientas entre ella la Analítica Prescriptiva.

La Analítica Prescriptiva permite indicar diferentes posibles acciones según un determinado suceso, las cuales ayudan a encontrar una solución que sea acorde a las necesidades presentadas. Estas indicaciones o prescripciones ayudan al proceso de toma de decisiones teniendo en cuenta que además de predecir lo que va a suceder también indica porque va a suceder.

Con lo anterior en mente y al no haber una metodología estándar para el desarrollo de proyectos de Analítica Prescriptiva este trabajo propone una metodología a partir de un metaanálisis orientado a la Analítica Prescriptiva, en cual se reviso de manera sistemática el estado del arte, metodologías y usos en distintas áreas del conocimiento, encontrando patrones en sus procesos que son comunes a metodologías orientadas a Data Mining como lo son KDD, CRISP-DM y SEMMA.

Palabras Clave: Analítica Prescriptiva, CRISP-DM, KDD, Metaanálisis, Metodología, Minería de Datos, SEMMA.

ABSTRACT

Currently the world is at the doors of the so-called fourth industrial revolution or also called Industry 4.0, where communication and large volumes of data are on the rise, this data can be managed correctly can help in the decision-making process, but for this must be used a variety of tools including Prescriptive Analytics.

The Prescriptive Analytics allows to indicate different possible actions according to a certain event, which help to find a solution that is according to the presented needs. These indications or prescriptions help the decision-making process, keeping in mind that in addition to predicting what will happen, it also indicates why it will happen.

With the above in mind and as there is no standard methodology for the development of Prescriptive Analytics projects, this work proposes a methodology based on a meta-analysis oriented to Prescriptive Analytics, in which the state of the art, methodologies and uses in different areas of knowledge, finding patterns in their processes that are common to methodologies oriented to Data Mining such as KDD, CRISP-DM and SEMMA.

Keywords: Prescriptive Analytics, CRISP-DM, KDD, Meta-analysis, Methodology, Data Mining, SEMMA.

INTRODUCCIÓN

Actualmente el mundo se encuentra a puertas de la denominada cuarta revolución industrial o también llamada Industria 4.0, la cual “aparece como resultado del surgimiento, distribución, uso y apropiación de nuevas tecnologías (tecnologías digitales y tecnologías de Internet) que permiten procesos de producción completamente automatizados donde los objetos físicos (máquinas) interactúan sin que se de la participación humana” [1] Dentro de estos fenómenos tecnológicos se encuentra el Big Data, que se caracteriza por el manejo que se le da al volumen, variedad, velocidad, veracidad y valor presentes en los datos. Teniendo en cuenta las anteriores características es de vital importancia ahondar en las herramientas utilizadas en este tipo de tecnologías como la Analítica Prescriptiva.

La Analítica Prescriptiva permite indicar diferentes posibles acciones según un determinado suceso, las cuales ayudan a encontrar una solución que sea acorde a la necesidades presentadas. Estas indicaciones o prescripciones ayudan al proceso de toma de decisiones teniendo en cuenta que además de predecir lo que va a suceder también indica porque va a suceder.

La Analítica Prescriptiva aun esta en una fase de crecimiento, y se espera un futuro muy promisorio en su uso, debido a una característica fundamental que posee para la Industria 4.0 y es la automatización de la toma de decisiones, la cual se apoya en sistemas de gestión de las reglas de negocio y la optimización matemática, lo que le permite elegir cual es la decisión correcta de manera automática, decisión que también podría tomar un ser humano, pero debido a la cantidad y complejidad de información que actualmente se encuentra disponible hace que este proceso sea dispendioso, lo cual implicaría mayor tiempo tomando una decisión o incluso llegando a tomar una decisión posible mas no la ideal, circunstancias que no ocurrirían con la automatización de la toma de decisiones.

Las anteriores características permiten que la Analítica Prescriptiva sea utilizada con gran éxito en distintas áreas del conocimiento desde la inteligencia de negocios, pasando por cuidados para la salud, deportes, inteligencia artificial, incluso llegando a aplicaciones aeronáuticas entre muchos otros.

1. PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción del problema

En 1962 el estadístico estadounidense John W. Tukey (1915-2000) mientras enseñaba en la universidad de Princeton y a su vez trabajaba en los laboratorios Bell, publicó el artículo titulado “The future of Data Analysis”, donde en su introducción, definió el análisis de datos como “Procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planificar la recopilación de datos para hacer su análisis más fácil, más preciso o acertado, y toda la maquinaria y los resultados de las estadísticas (matemáticas) que se aplican al análisis de datos” [2].

Con el paso del tiempo, desde aquella definición, han venido apareciendo varios conceptos relacionados a los datos como los son: Big Data, Data Science, Analítica de Datos, Minería de Datos, entre otros y ha sido así debido a la creciente generación de datos en todos los ámbitos de la experiencia humana, “calculando alrededor de 1,7MB por persona cada segundo para el año 2020”, demostrando así el enorme potencial e importancia que tienen los datos en la actualidad.

Entre las herramientas utilizadas para limpiar, organizar y generar información a través de los datos, destaca la Analítica de Datos, que se define como “El proceso vital para comprender el conglomerado de datos que extraen y crean información valiosa para la toma de decisiones y la gestión eficaz en las organizaciones” [3]. Dentro de la Analítica de Datos existen tres tipos predominantes de analítica: la Analítica Descriptiva que se encarga de responder las preguntas ¿Que paso?, ¿Por qué pasó?, ¿Qué está pasando ahora?; la Analítica Predictiva que se encarga de responder las preguntas ¿Qué pasará?, ¿Por qué pasará?; y la Analítica Prescriptiva que se encarga de responder las preguntas ¿Qué debería hacer?, ¿Por qué debería hacerlo?

Actualmente la Analítica de Datos es ampliamente usada en distintos campos entre ellos la Analítica de Negocios donde “la gran mayoría de sus esfuerzos se dedican al análisis descriptivo y al análisis predictivo con metodologías que incluyen minería de datos, aprendizaje automático, inteligencia artificial y simulación” [4], demostrando el gran nivel de uso de las anteriores analíticas mencionadas, no siendo lo mismo para el caso de la Analítica Prescriptiva la cual esta mucho menos desarrollada [4].

“La Analítica Prescriptiva es capaz de sugerir (prescribir) las mejores opciones de decisión para aprovechar el futuro previsto e ilustra las implicaciones de cada opción de decisión” [4]. Al ser una herramienta tan poderosa y poco explorada y que de acuerdo a un estudio realizado por Market Research Future, empresa india de investigación de mercado global que combina información patentada con varias fuentes de datos para ofrecer resultados a sus clientes, el negocio global de la Analítica Prescriptiva alcanzara \$2,470 millones de dólares para el año 2022, lo cual remarca la crucial importancia de explorar, entender y profundizar en métodos de implementación de la analítica en cuestión.

Existen varios artículos donde ya han empezado a utilizar metodologías personalizadas para el proceso prescriptivo, debido a que no es común encontrar una metodología estándar para dicho proceso. Situación que ofrece la posibilidad de explorar en metodologías ya establecidas y probadas como es el caso de CRISP-DM, SEMMA, KDD, metodologías utilizadas para proyectos de minería de datos, concebidas entre finales de la década de los 90's y comienzo de la década del 2000. De acuerdo con Azevedo y Santos [5], las metodologías más utilizadas en proyectos de minería de datos son CRISP-DM y SEMMA.

Mariscal, Marbla y Fernandez [6], analizan que las metodologías utilizadas en proyectos de minería de datos se basan en la metodología KDD, por tal razón el presente trabajo busca identificar los elementos, características y fases de las tres metodologías antes mencionadas aplicables en proyectos de analítica prescriptiva a través de un metaanálisis de investigaciones citadas y publicadas en la base de referencia SCOPUS.

1.2. Formulación del problema

¿Como determinar los elementos, características y fases de las metodologías CRISP-DM, SEMMA Y KDD de Data Mining aplicables en el desarrollo de proyectos de Analítica Prescriptiva?

2. JUSTIFICACIÓN

En la actualidad no existen una metodología estándar para desarrollar proyectos con Analítica Prescriptiva, razón por la cual, esta investigación tiene como objetivo el realizar un metaanálisis orientado a la Analítica Prescriptiva, en cual se revisara de manera sistemática el estado del arte, metodologías y usos en distintas áreas del conocimiento, para de esta manera encontrar patrones en sus procesos que sean comunes a metodologías orientadas a Data Mining para posteriormente generar, de ser posible, una propuesta de metodología estándar que pueda ayudar a la implementación de procesos prescriptivos en distintas áreas del conocimiento.

El metaanálisis desarrollado en este trabajo ayudará a futuros estudios relacionados con Business Intelligence, Inteligencia Artificial entre otros que usen la Analítica Prescriptiva para la toma de decisiones, al mostrar una metodología estable que sea aplicable a este tipo de procesos.

3. OBJETIVOS

3.1. Objetivo general

Desarrollar una propuesta de metodología para el desarrollo de proyectos de Analítica Prescriptiva a partir de un metaanálisis.

3.2. Objetivos específicos

- Elegir las metodologías de Minería de Datos aptas para ser usadas en procesos prescriptivos.
- Analizar artículos científicos relacionados con el uso de la Analítica Prescriptiva como insumo para el metaanálisis.
- Realizar un Metaanálisis de la Analítica Prescriptiva orientado al uso de metodologías de Minería de Datos en su proceso de aplicación.

4. ALCANCES Y LIMITACIONES

4.1. Alcances

- El presente trabajo de investigación solo analizara metodologías usadas en la Minería de datos.
- Los artículos científicos consultados serán publicaciones presentes en la base de datos SCOPUS hechas en los últimos 8 años.

4.2. Limitaciones

- La realización del metaanálisis se realizara durante el primer semestre de 2021.
- El tiempo de realización esta limitado por el periodo académico.
- Dada la naturaleza tan amplia de los proyectos que utilizan la analítica prescriptiva, la propuesta de metodología a desarrollar procurará, en la medida de lo posible, presentar un marco general de desarrollo, sin embargo, es posible que algunos proyectos no puedan ser incluidos en la metodología a desarrollar.

5. MARCO CONCEPTUAL

5.1. Analítica Prescriptiva

El término Analítica Prescriptiva fue acuñado por IBM y se define como un conjunto de técnicas matemáticas que determinan computacionalmente un conjunto de acciones o decisiones alternativas de alto valor dado un conjunto complejo de objetivos, requisitos y restricciones, con el objetivo de mejorar el rendimiento empresarial.

Según Atanu Basu CEO y presidente de AYATA, compañía americana de software orientado a la Analítica Prescriptiva, existen 5 pilares que se deben tener en cuenta para que los procesos de Analítica Prescriptiva tengan éxito:

- **Manejo de Datos Híbridos:** Son datos que están compuestos por datos estructurados como números y categorías y datos no estructurados como textos, imágenes, videos y audios. El procesamiento de datos híbridos permite a las empresas utilizar todos los datos disponibles para tomar las mejores decisiones posibles, al utilizar para dicho procesamiento tecnologías como: Machine Learning, Investigación de Operaciones, Visión por Computador Procesamiento de Lenguaje Natural, Estadística Aplicada, Procesamiento de Señales, Procesamiento de Imágenes, Metaheurísticas, entre otras.
- **Predicciones y Prescripciones integradas:** Son funciones que deben trabajar de manera conjunta para que el análisis sea coherente y su resultado sea valioso para la toma de decisiones.
- **Prescripciones y efectos secundarios:** Para generar las prescripciones, que se definen como acciones recomendadas dependientes del tiempo para mejorar el futuro, se utilizan varios métodos entre ellos la Investigación de Operaciones, la cual tiene en cuenta objetivos, restricciones y variables de decisión y que a través de la optimización y la simulación produce el mejor camino posible, una prescripción, evitando de esta manera efectos secundarios indeseados.
- **Algoritmos adaptativos:** Debido a la gran cantidad de datos que diariamente se generan, se hace necesario que los algoritmos utilizados en la Analítica Prescriptiva sean recalibrados automáticamente de manera adaptativa ya sea de forma dinámica y/o continua para ayudar al éxito del proceso del negocio.
- **Mecanismo de Retroalimentación:** Automatización de las decisiones basado en las prescripciones.

5.2. Minería De Datos

La Minería de Datos es un conjunto de metodologías y herramientas que mediante el análisis de grandes cantidades de datos ayuda a obtener patrones de comportamiento o tendencias ocultas que pueden ser muy útiles en la toma de decisiones [7]. Zaki y Meira en su libro DATA MINING AND ANALYSIS Fundamental Concepts and Algorithms, la catalogan como parte de un proceso de descubrimiento de conocimiento amplio, que incluye tareas de preprocesamiento como extracción de datos, limpieza de datos, fusión de datos, reducción de datos y construcción de características, así como pasos de posprocesamiento como interpretación de patrones y modelos, confirmación y generación de hipótesis, etc. [8].

Las técnicas en la Minería de Datos se podrían clasificar de tres tipos:

- **Técnicas de Clusterizado:** A partir de una base de datos de observaciones, se busca agruparlas dentro de un número de clases preestablecidas o no, generalmente mediante criterios de distancia o similitud, de forma que las observaciones que pertenezcan a una de las clases sean similares entre sí y distintas con las de otras clases. Algunas técnicas muy utilizadas son: K- Means, las Redes SOM (Self Organization Maps), Sistema de clasificación automática Bayesiana (Autoclass), Teoría de Resonancia Adaptativa (ART y ART2), otro tipo de redes neuronales, etc.
- **Clasificadores:** Tratan de buscar unas leyes o patrones que expliquen, dentro de una cierta incertidumbre y partiendo de una base de datos de observaciones, el comportamiento de unas variables frente a otras. Se dividen fundamentalmente en: generadores de árboles de decisión, generadores de reglas, modelos de regresión, redes neuronales, redes funcionales, clasificadores bayesianos, algoritmos genéticos, métodos visuales, clasificadores fuzzy, etc.
- **Técnicas de Reducción de Dimensión y Visualización de la Información:** Tienen como objetivo reducir el número de variables y visualizar la nube de puntos N-Dimensionales para poder detectar estructuras o características de forma visual. Entre las técnicas más utilizadas son: Análisis de Componentes Principales (PCA), proyector Sammon, PCA no lineal, proyector Andrews, Proyección Pursuit, RADVIZ, Gráficos de Coordenadas Paralelas, etc.

5.3. Metodología KDD

El término *Knowledge Discovery on Data* surgió a finales de los años 80s en el primer workshop de KDD [9], para enfatizar que el conocimiento es el producto final de un descubrimiento impulsado por datos y es un punto conjunto de diferentes áreas de investigación enfocadas en el análisis de datos y la extracción de conocimiento de diferentes puntos de vista, como bases de datos, estadística, matemáticas, lógica o inteligencia artificial [6]. A mediados de los 90s Fayyad y Piatetsky-Shapiro, entre otros autores, definieron a KDD como el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y finalmente comprensibles en los datos [10].

KDD es un modelo de proceso interactivo e iterativo que está compuesto por nueve etapas:

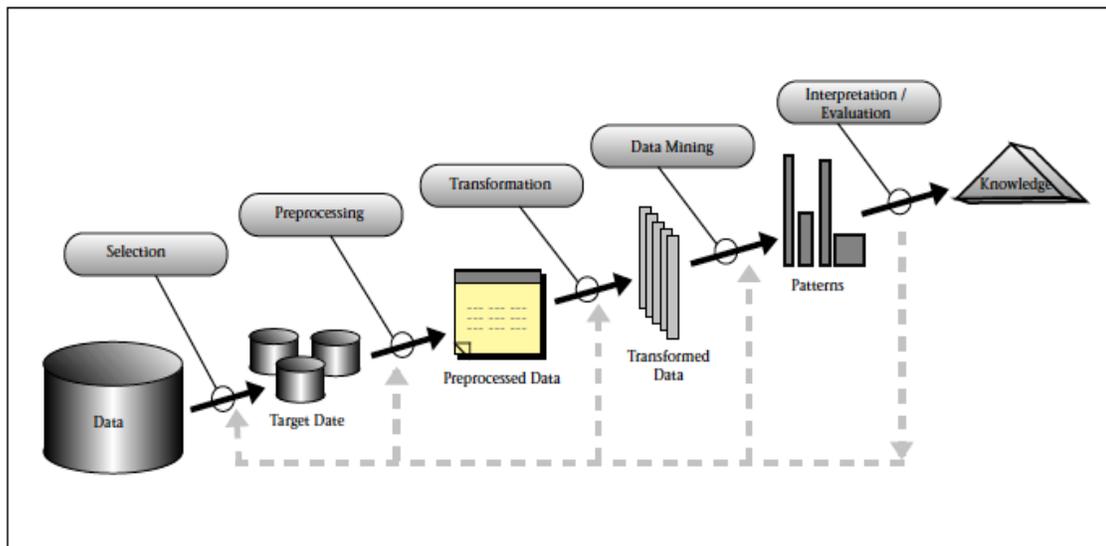


Figura 1. Etapas del modelo de proceso KDD. [10]

- **Desarrollo y comprensión del dominio de la aplicación**

Los objetivos se definen desde el punto de vista del cliente y se utilizan para desarrollar y comprender el dominio de la aplicación y su conocimiento previo.

- **Creación de un conjunto de datos objetivo**

Selecciona un conjunto de datos o se centra en un subconjunto de muestras de datos o variables.

- **Limpieza y preprocesamiento de datos**

Incluye operaciones básicas, como eliminar ruido o valores atípicos si corresponde, recopilar información necesaria para modelar o contabilizar el ruido, decidir estrategias para manejar los campos de datos faltantes y contabilizar la información de secuencia de tiempo y los cambios conocidos, así como decidir los problemas del sistema de gestión de la base de datos, como los tipos de datos, el esquema y el mapeo de datos perdidos y valores desconocidos.

- **Transformación de datos**

Se centra en la transformación de datos de una forma a otra para que los algoritmos de Minería de Datos se puedan implementar fácilmente, usando métodos de reducción y transformación sobre los datos objetivo.

- **Elección de la técnica de Minería de Datos adecuada**

Incluye elegir una técnica de Minería de Datos como: clasificación, agrupamiento, regresión, resumen, etc., teniendo en cuenta los objetivos definidos en la primera etapa.

- **Elección del algoritmo de Minería de Datos adecuado**

Se seleccionan uno o más algoritmos de Minería de Datos que se usaran para buscar patrones en los datos.

- **Implementación de la Minería de Datos**

Se implementan las técnicas y algoritmos seleccionados.

- **Interpretación de los patrones minados**

Se interpretan y evalúan los patrones descubiertos. Además, en esta etapa posiblemente se tenga que regresar a alguna etapa anterior. También es posible eliminar patrones redundantes o irrelevantes.

- **Uso del conocimiento descubierto**

Incluye incorporar este conocimiento en el sistema de desempeño, tomar acciones basadas en el conocimiento o simplemente documentarlo y reportarlo a las partes interesadas, así como verificar y resolver posibles conflictos con conocimiento previamente creído o extraído.

5.4. Metodología CRISP-DM

A mediados de los años 90s un grupo de empresas relacionadas con la Minería de Datos como Teradata, SPSS -ISL-, Daimler-Chrysler y OHRA, propusieron una guía de referencia para desarrollar proyectos de Minería de Datos llamada CRISP-DM (*Cross Industry Standard Process for Data Mining*) [11]. En 1999 la versión 1.0 fue publicada y totalmente documentada. CRISP-DM es considerada por muchos autores como el estándar *de facto* para desarrollar proyectos de Minería de Datos y de descubrimiento de conocimiento (Knowledge Discovery).

Esta metodología consta de seis etapas que son:

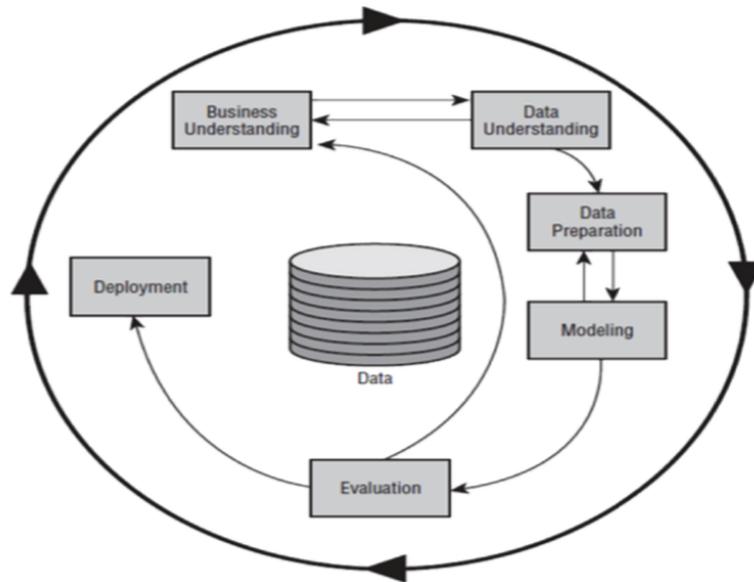


Figura 2. Etapas del modelo de proceso CRISP-DM. [11]

- **Comprensión del negocio**

Se enfoca en comprender los objetivos y requisitos del proyecto desde una perspectiva del negocio, luego en convertir este conocimiento en una definición de problema de minería de datos y un plan preliminar diseñado para lograr los objetivos.

- **Comprensión de los datos**

Esta fase comienza con la recopilación inicial de datos y continúa con actividades que le permiten familiarizarse con los datos, identificar problemas de calidad de los datos, descubrir los primeros conocimientos

sobre los datos y / o detectar subconjuntos interesantes para formar hipótesis sobre información oculta.

- **Preparación de los datos**

Cubre todas las actividades necesarias para construir el conjunto de datos final, datos que se introducirán en las herramientas de modelado, a partir de los datos en bruto iniciales. Es probable que las tareas de preparación de datos se realicen varias veces y no en ningún orden prescrito. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado.

- **Modelado**

Se seleccionan y aplican diversas técnicas de modelado, y sus parámetros se calibran a valores óptimos. Normalmente, existen varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requisitos específicos sobre la forma de los datos. Por lo tanto, a menudo es necesario volver a la fase de preparación de datos.

- **Evaluación**

Se han creado modelos que parecen tener alta calidad desde la perspectiva del análisis de datos. Antes de proceder a la implementación final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, para asegurarse de que el modelo logre adecuadamente los objetivos del negocio. Un objetivo clave es determinar si existe algún tema del negocio importante que no se haya considerado suficientemente. Al final de esta fase, se debe tomar una decisión sobre el uso de los resultados de la minería de datos.

- **Despliegue**

La creación del modelo generalmente no es el final del proyecto. Incluso si el propósito del modelo es aumentar el conocimiento de los datos, el conocimiento adquirido deberá organizarse y presentarse de manera que el cliente pueda utilizarlo. Según los requisitos, la fase de implementación puede ser tan simple como generar un informe o tan compleja como implementar un proceso de minería de datos repetible en toda la empresa.

5.5. Metodología SEMMA

SEMMA (*Sample, Explore, Modify, Model, and Assess*) creada por el Instituto SAS, que la define como una organización lógica del conjunto de herramientas funcionales de SAS Enterprise Miner para llevar a cabo las tareas centrales de la minería de datos [12].

SEMMA ofrece y permite la comprensión, organización, desarrollo y mantenimiento de proyectos de Minería de Datos y consta de cinco etapas:



Figura 3. Etapas del modelo de proceso SEMMA. [12]

- **Muestra**

Es una etapa opcional. Una parte de un gran conjunto de datos se toma lo suficientemente grande como para extraer información significativa y lo suficientemente pequeña como para manipularla rápidamente.

- **Explorar**

Ayuda a obtener la comprensión y las ideas, así como refinar el proceso de descubrimiento mediante la búsqueda de tendencias y anomalías.

- **Modificar**

Modifica los datos mediante la creación, selección y transformación de variables para enfocar el proceso de selección del modelo. Aquí se buscan valores atípicos y se reduce el número de variables.

- **Modelo**

En esta etapa el software busca automáticamente la combinación de datos.

- **Evaluar**

Evalúa la confiabilidad y la utilidad de los hallazgos y estima el desempeño.

6. MARCO TEORICO

6.1. Métodos estadísticos para combinación de estudios

6.1.1. Modelos de efectos fijos

En este modelo los estudios se combinan considerando que no existe heterogeneidad entre ellos, y que por lo tanto todos ellos constituyen estimaciones de un efecto real, cuya magnitud se desea conocer. El método general para este tipo de modelo es el método ponderado por la inversa de la varianza. En el cada estudio interviene en la estimación global de forma ponderada, con un peso proporcional a su precisión, que en este caso se mide con la inversa de la varianza, ya que la varianza es una medida de dispersión y por lo tanto su inversa una medida de precisión [13] .

6.1.2. Modelo de efectos aleatorios

En este modelo la inferencia se basa en suponer que los estudios incluidos en el análisis constituyen una muestra aleatoria del universo de estudios posibles, y sus resultados son mas conservadores al tener en cuenta una fuente extra de variación, ya que ahora se incluyen dos posibles fuentes de variación: La existente dentro de los estudios y la variación entre los estudios [13] .

6.1.3. Heterogeneidad

Si la variabilidad entre los estudios es superior a la que se espera por puro azar, se ha de contrastar la hipótesis de homogeneidad a través de una prueba que fue desarrollada por Cochran en 1954, en la que se calcula una suma ponderada de las desviaciones del efecto determinado en cada estudio con respecto a la media global. También es posible verificar la homogeneidad de los resultados a través de otros medios como el análisis por subgrupos, el cual ofrecerá un resultado mas confiable [13] .

6.1.4. Meta-regresión

Describe el análisis en el que las características de los estudios o de los sujetos incluidos en cada estudio se usan como variables explicativas en un

modelo de regresión multivariante, en el que la variable dependiente es el efecto medido o alguna medida de la desviación del efecto de cada estudio respecto del efecto global [13].

6.1.5. Análisis de subgrupos y sensibilidad

Consiste en efectuar el metaanálisis en diferentes grupos de estudios, reunidos según características de los mismos (estudios publicados frente a no publicados, estudios de gran tamaño frente a de pequeño tamaño, estudios según nivel de calidad, según la fecha de publicación, etc.) [13].

6.2. Calculo del tamaño del efecto

El tamaño de efecto se puede definir como cualquier medida estadística que evidencia el grado con el que un evento dado esta presente en una muestra. El tipo de medida se llama efecto, y su magnitud es el tamaño del efecto. Los tamaños de efecto se calculan a partir de diferentes tipos de datos originales que pueden ser [13]:

- Medias, tamaños de muestras y desviaciones estándar de los grupos de control y experimental.
- Tablas de contingencia 2x2 que representan las cuatro posibles soluciones categóricas experimentales.
- Coeficientes de correlación.

6.3. Evaluación de sesgo de publicación

El sesgo de publicación es publicación selectiva de artículos mostrando cierto tipo de resultados sobre otros, situación que se presenta por la tendencia de las revistas de rechazar resultados no significativos y por la tendencia de los propios autores de no enviar resultados no significativos. Este comportamiento conlleva un déficit de resultados nulos, y una sobreestimación de resultados significativos. La presencia de sesgo de publicación se evalúa mediante métodos gráficos como histogramas ponderados, diagramas de embudo, gráficos de cuartiles normales, o métodos estadísticos como la prueba de correlación de rangos, Fail-safe number, Weigthed fail-safe number [13].

7. ESTADO DEL ARTE

Como se ha mencionado, la Analítica Prescriptiva en los últimos años ha comenzado a tener un gran auge, debido al poder que aporta principalmente en la toma de decisiones. Lo anterior se ve reflejado en la variedad de estudios que se han venido produciendo en distintas áreas del conocimiento, desde la inteligencia de negocios, pasando por cuidados para la salud, deportes, inteligencia artificial, incluso llegando a aplicaciones aeronáuticas entre muchos otros. A continuación, se mencionarán algunas de las investigaciones que se han realizado en el campo de la analítica prescriptiva.

Prescriptive analytics for recommendation-based business process optimization, Gröger, C., Schwarz, H., Mitschang, B., 2014, citas 29. Este artículo propone el concepto de Optimización de Procesos de Negocios basada en Recomendaciones (rBPO) el cual aprovecha la analítica prescriptiva y genera de forma proactiva recomendaciones de acción durante la ejecución del proceso para evitar una desviación de la métrica prevista. Se basa en un almacén de procesos holístico y emplea técnicas de clasificación para la generación de predicciones y recomendaciones en tiempo real. El rBPO se centra en la optimización de procesos basada en datos en tiempo de ejecución, no en la mejora del modelo de proceso clásico durante el tiempo de diseño o el análisis expost. [14].

Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework, Srinivas, S., Ravindran, A. R., 2018, citas 26. Este artículo propone un marco analítico prescriptivo, en el que se aprovecha el riesgo de no presentación específico de un paciente que se predice mediante el uso de múltiples fuentes de datos para desarrollar reglas dinámicas para la programación del paciente. Estas reglas de programación incorporan múltiples decisiones de diseño del sistema de citas, como el ajuste por no presentación y la secuenciación de pacientes, simultáneamente. Las reglas propuestas se evalúan considerando la suma ponderada del tiempo de desbordamiento de recursos (horas extra), el tiempo de inactividad de los recursos, el tiempo de espera del paciente y el número de citas denegadas. Si son efectivas, estas reglas serán de gran valor para los profesionales de la salud, ya que podrían ayudar a aliviar el problema de la sobrecarga de recursos, la congestión de la sala de espera, las infecciones adquiridas en el hospital y las demoras más prolongadas en las citas [15].

Prescriptive analytics for FIFA World Cup lodging capacity planning, Ghoniem, A., Ali, A., Al-Salem, M., Khallouli, W., 2016, citas 4. Este artículo desarrolla un marco de análisis de optimización que emplea secuencialmente dos modelos de programación de enteros para el análisis de espectadores extranjeros y los consiguientes requisitos de alojamiento. El marco se aplica para evaluar la preparación de la infraestructura de alojamiento en Qatar para FIFA 2022 [16].

Prescriptive analytics system for long-range aircraft conflict detection and resolution, Ayhans, S., Costas, P., Hanan, S., 2018, citas 7. Este artículo propone un sistema de análisis prescriptivo para abordar un problema de detección y resolución de conflictos de aeronaves de largo alcance (CDR). Dado un conjunto de trayectorias previstas, el sistema declara un conflicto cuando una zona protegida de una aeronave en su trayectoria es violada por otra aeronave. El sistema resuelve el conflicto prescribiendo una solución alternativa que se optimiza perturbando al menos una de las trayectorias involucradas en el conflicto. Para lograr esto, el sistema aprende de patrones descriptivos de trayectorias históricas y observaciones meteorológicas pertinentes y construye un Modelo Oculto de Markov (HMM). Mediante una variante del algoritmo de Viterbi, el sistema evita el volumen del espacio aéreo en el que se detecta el conflicto y genera una nueva trayectoria óptima libre de conflictos. El concepto clave sobre el que se basa el sistema es la suposición de que el espacio aéreo no es más que un conjunto de cubos de datos espacio-temporales concatenados horizontal y verticalmente donde cada cubo se considera una unidad atómica [17].

8. METODOLOGIA

8.1. Metaanálisis

En 1976 el estadístico americano G.V. Glass acuñó el término Metaanálisis como el análisis estadístico del conjunto de resultados obtenidos en diferentes ensayos clínicos sobre una misma cuestión, con la finalidad de evaluarlos de manera conjunta. Para poder realizar este análisis son necesarias las siguientes etapas:

- **Formulación del problema:** La pregunta que se pretende responder debe formularse de manera clara, además se deben definir los conceptos implicados en la misma. A partir de la pregunta se generan los objetivos a cumplir y la hipótesis que se quiere contrastar [18].
- **Búsqueda de la literatura:** La validez de un metaanálisis depende mucho de lo exhaustivo con que se identifique y localice los trabajos originales, para realizar dicha búsqueda se debe recurrir a diferentes fuentes de información como lo son de tipo informales, primarias y secundarias.
 - Fuentes informales: Se refieren a archivos personales, libros, contacto con expertos etc.
 - Fuentes primarias: Se refieren a revistas conocidas relacionadas con el tema y la realización de revisiones ascendentes.
 - Fuentes secundarias: Se refieren a bases de datos automatizadas que incluyen solamente revistas y no se incluyen tesis doctorales, trabajos no publicados elementos denominados literatura gris.

En esta fase se definen también los criterios de inclusión y exclusión, los cuales dependen del objetivo del metaanálisis, a partir de dichos criterios se comienza con la búsqueda bibliográfica [18].

- **Codificación de los estudios:** Después de seleccionar los estudios que cumplen con los criterios establecidos, se elabora un Manual de Codificación en el que se hagan explícitos los criterios mediante los cuales se van a codificar las características de los estudios, para de esta

manera comprobar que características de los estudios pueden estar moderando o afectando los resultados [18].

- **Cálculo del tamaño del efecto:** En esta etapa se calcula un índice estadístico el cual tiene que ser tal que pueda calcularse de forma homogénea en todos los estudios, de forma que sea capaz de poner en la misma métrica los resultados de los estudios [18]. Este índice se denomina índice del tamaño de efecto.

- **Análisis estadístico e interpretación:** En esta etapa primero se debe describir las características de los estudios que se han codificado, con objeto de obtener una imagen prototípica de los estudios. Teniendo en cuenta los resultados de etapas anteriores a continuación se crea una base de datos en la que las filas son los estudios y las columnas son las variables potencialmente moderadoras de los resultados de los estudios, así como el tamaño del efecto obtenido en cada estudio. Posteriormente el análisis estadístico pasa por tres fases [18]:
 - 1) Cálculo del tamaño del efecto medio con su intervalo de confianza y valoración de su significación estadística.
 - 2) Análisis de la heterogeneidad de los tamaños del efecto.
 - 3) Si los tamaños del efecto son heterogéneos, se hace una búsqueda de variables moderadoras de tal variabilidad.

- **Publicación:** La última etapa en la realización de un metaanálisis, como la de cualquier otra investigación, es su publicación. Al tratarse de una investigación empírica, las secciones que debe incluir el informe escrito del metaanálisis son las típicas de un estudio empírico: introducción, método, resultados y discusión y conclusiones [18].

9. DESARROLLO DEL PROYECTO

9.1. Formulación del problema

Que metodología CRISP-DM, SEMMA, KDD o sus fases son las mas apropiadas para desarrollo de proyectos de Analítica Prescriptiva?

9.2. Búsqueda de información

La búsqueda de artículos se realizo en la base de datos Scopus debido a que cubre un mayor número de revistas que otras bases de datos. Además el idioma de consulta fue el inglés debido a que es el utilizado en dicha base de datos.

9.2.1. Criterios de inclusión / exclusión

Al consultar en la base de datos usando los criterios de búsqueda Article title, Abstract y Keywords de manera simultanea arrojó mas de 400 resultados, razón por la cual se determinó utilizar los siguientes criterios para tener una mejor fiabilidad y precisión en la búsqueda (Ver Anexo A).

- **Primer Criterio:** Utilizar solamente el criterio de búsqueda Article Title en el cual el termino *Prescriptive Analytics* fue utilizado como termino clave para la búsqueda.
- **Segundo Criterio:** Que la fecha de publicación del articulo estuviera desde el año 2013 hasta el año 2020.
- **Tercer Criterio:** Seleccionar artículos que no se incluyan en libros y/o conferencias debido a la dificultad para acceder a dichos recursos.
- **Cuarto Criterio:** Seleccionar artículos que tengan un proceso de desarrollo de una solución utilizando analítica prescriptiva, evitando de esta manera revisiones, estados del arte entre otras.

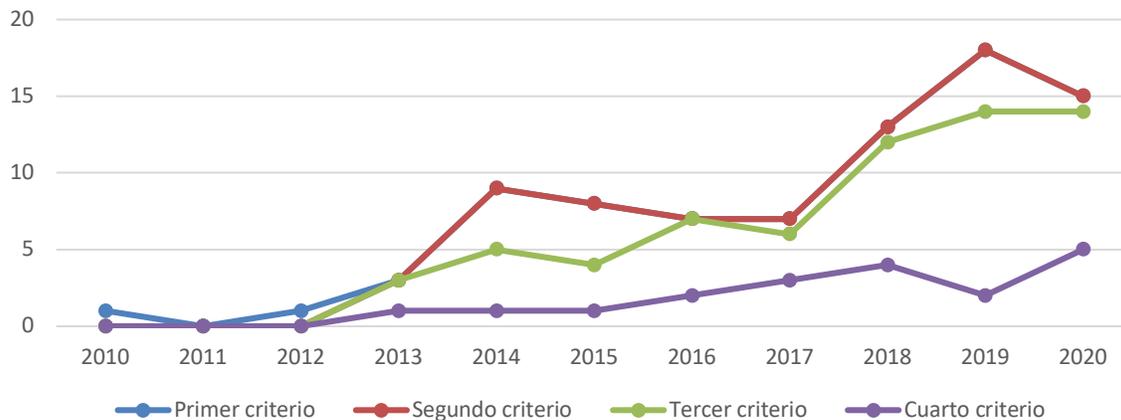


Figura 4. Aplicación de Criterios Inclusión / Exclusión. Años vs Numero de artículos. [El Autor]

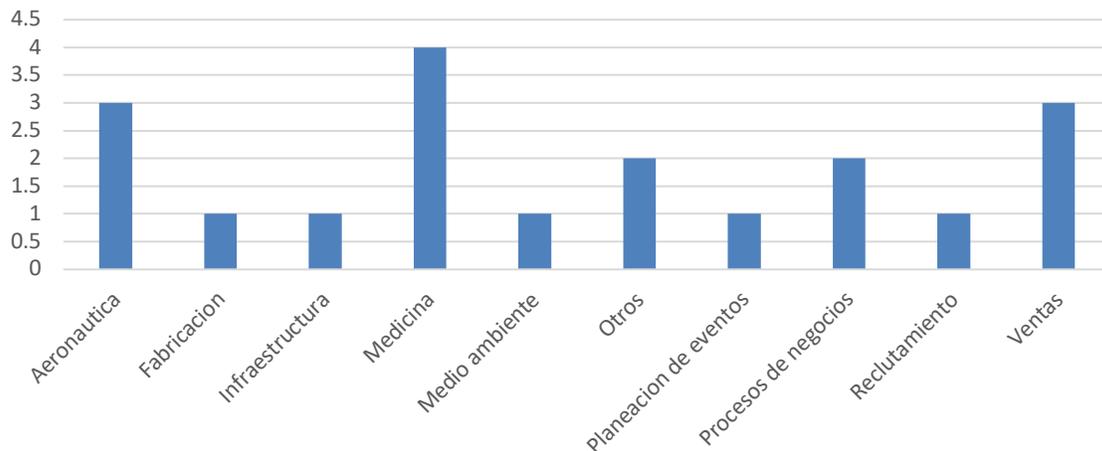


Figura 5. Áreas de investigación presentes en los artículos. Área vs Numero de artículos. [El Autor]

9.3. Análisis estadístico

Para determinar las similitudes que hay entre los procesos de desarrollo presentes en los artículos de proyectos de analítica prescriptiva y las fases presentes en las metodologías de minería de datos KDD, CRISP-DM y SEMMA, se procedió a leer dichos artículos y realizar una matriz para cada metodología donde se muestran las fases de cada una y se relaciona con las fases halladas en los artículos a través de dos valores, 1 si la fase esta presente y 0 si la fase no esta presente en el flujo del desarrollo expuesto en los artículos, para de esta manera poderlos evaluar de una manera estadística.

9.3.1. Análisis de etapas de las metodologías presentes en los artículos

9.3.1.1. KDD

ID ARTICULO	Selección	Preprocesamiento / Limpieza	Transformación / Reducción	Minería de datos	Interpretación / Evaluación
GrSM2014	1	1	0	1	1
BLPW2020	1	1	0	1	1
MPGThWGKPSHV2020	1	1	1	1	1
PSAchShB2020	1	1	0	1	1
VAOPN2020	1	0	1	1	1
SrR2018	1	1	0	0	1
KrWL2016	1	1	0	1	1
KSSV2013	1	0	0	0	1
AyCS2018	1	0	0	0	1
JQH2016	1	0	1	1	1
VBPRFr2015	1	1	0	0	1
CSACrDOPS2019	1	1	0	0	1

GhAA-SKh2017	1	0	0	0	1
WY2017	1	0	1	1	1
AS2018	1	1	0	1	1
dAGrc2017	1	0	0	1	1
GLSchT2020	1	1	0	1	1
WP2019	1	0	0	1	1
HRGNgKV2018	1	1	0	1	1

Tabla 1. Etapas del modelo de proceso KDD presentes en los artículos. [El Autor]

Fase	Valor
Selección	19
Preprocesamiento / Limpieza	11
Transformación / Reducción	4
Minería de Datos	13
Interpretación / Evaluación	19

Tabla 2. Consolidado de fases KDD por artículo. [El Autor]

De acuerdo a los datos de la tabla de consolidados de fases KDD por artículo, las fases más comunes utilizadas en los artículos son las fases de Selección e Interpretación / Evaluación, situación que es representada en la siguiente gráfica.

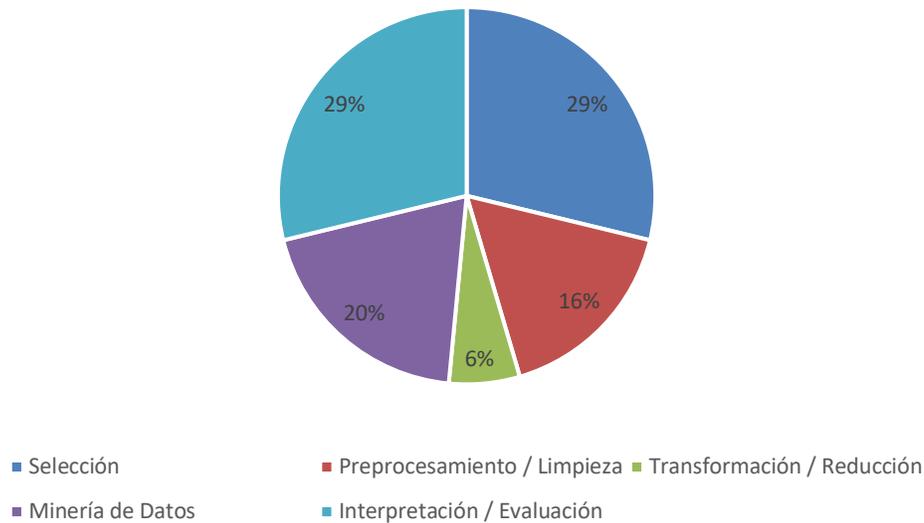


Figura 6. Porcentajes de fases KDD utilizada en los artículos. [El Autor]

9.3.1.2. CRISP-DM

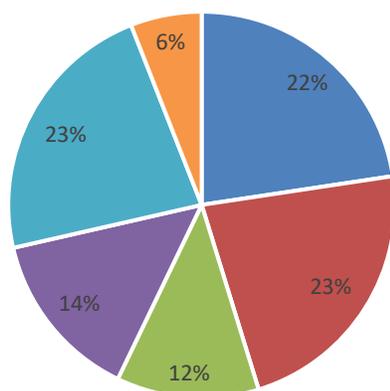
ID ARTICULO	Comprensión del negocio	Comprensión de los datos	Preparación de los datos	Fase de Modelado	Evaluación	Implantación
GrSM2014	0	1	0	1	0	1
BLPW2020	1	1	1	0	1	0
MPGThWGKPSHV2020	1	1	1	1	1	0
PSAChShB2020	1	1	1	1	1	1
VAOPN2020	1	1	0	1	1	1
SrR2018	1	1	0	0	1	1
KrWL2016	1	1	1	1	1	0
KSSV2013	1	1	0	0	1	0
AyCS2018	1	1	0	0	1	1
JQH2016	1	1	1	1	1	0
VBPRSF2015	1	1	0	0	1	0
CSACrDOPS2019	1	1	1	0	1	0
GhAA-SKh2017	1	1	0	0	1	0
WY2017	1	1	1	1	1	0
AS2018	1	1	1	1	1	0
dAGrc2017	1	1	0	1	1	0
GLSchT2020	1	1	1	1	1	0
WP2019	1	1	0	1	1	0
HRGNgKV2018	1	1	1	1	1	0

Tabla 3. Etapas del modelo de proceso CRISP-DM presentes en los artículos. [El Autor]

Fase	Valor
Comprensión del negocio	19
Comprensión de los datos	19
Preparación de los datos	10
Fase de Modelado	12
Evaluación	19
Implantación	5

Tabla 4. Consolidado de fases CRISP-DM por artículo. [El Autor]

De acuerdo a los datos de la tabla de consolidados de fases CRISP-DM por artículo, las fases más comunes utilizadas en los artículos son las fases de Comprensión del Negocio, Comprensión de los Datos y la Evaluación, situación que es representada en la siguiente gráfica.



- Comprensión del negocio ■ Comprensión de los datos ■ Preparación de los datos
- Fase de Modelado ■ Evaluación ■ Implantación

Figura 7. Porcentajes de fases CRISP-DM utilizada en los artículos. [El Autor]

9.3.1.3. SEMMA

ID ARTICULO	Muestra	Explorar	Modificar	Modelo	Evaluar
GrSM2014	0	1	0	1	1
BLPW2020	0	1	1	1	1
MPGThWGKPSHV2020	0	1	1	1	1
PSAChShB2020	0	1	1	1	1
VAOPN2020	0	1	1	1	1
SrR2018	1	1	0	0	1
KrWL2016	0	1	0	1	1
KSSV2013	0	1	0	0	1
AyCS2018	0	1	0	0	1
JQH2016	0	1	1	1	1
VBPRSFr2015	0	1	0	0	1
CSACrDOPS2019	0	1	1	0	1
GhAA-SKh2017	0	1	0	0	1
WY2017	0	1	1	1	1
AS2018	0	1	0	1	1
dAGrc2017	1	1	0	1	1
GLSchT2020	0	1	1	1	1
WP2019	0	1	0	1	1
HRGNgKV2018	0	1	1	1	1

Tabla 5. Etapas del modelo de proceso SEMMA presentes en los artículos. [El Autor]

Fase	Valor
Muestra	2
Explorar	19
Modificar	9
Modelo	13
Evaluar	19

Tabla 6. Consolidado de fases SEMMA por artículo. [El Autor]

De acuerdo a los datos de la tabla de consolidados de fases SEMMA por artículo, las fases más comunes utilizadas en los artículos son las fases de Explorar, y la Evaluar, situación que es representada en la siguiente gráfica.

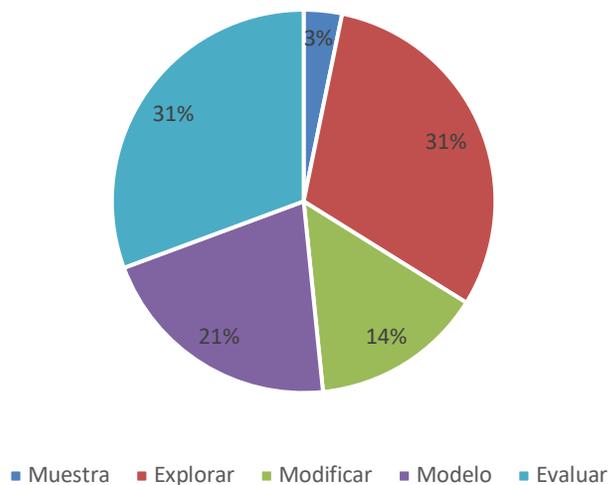


Figura 8. Porcentajes de fases SEMMA utilizada en los artículos. [El Autor]

9.3.2. Análisis de Metodología Vs Metodología

En la siguiente gráfica se puede apreciar cuántas fases son usadas por cada artículo de cada metodología, esto con el fin de visualizar que metodología se encuentra más completa en cada artículo, siendo para este caso CRISP-DM la metodología que presenta una mejor similitud con los procesos de desarrollo de cada artículo.

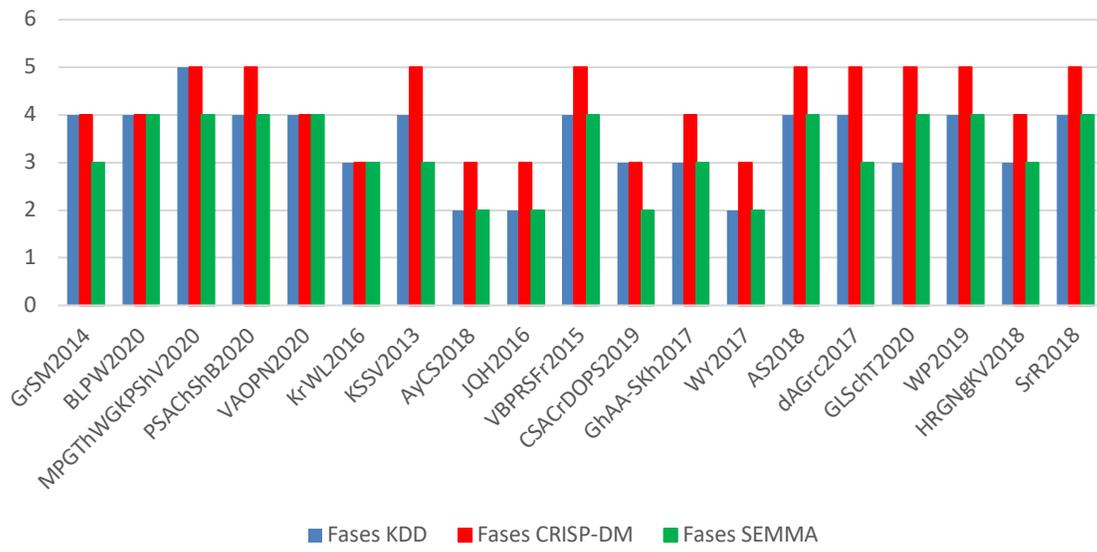


Figura 9. Uso de fases en los artículos KDD, CRISP_DM, SEMMA vs Numero de fases. [El Autor]

10. CONCLUSIONES

Al hacer la búsqueda para realizar el presente trabajo se pudo constatar el hecho del crecimiento que ha venido teniendo el interés por la analítica prescriptiva en el ámbito académico y comercial, hecho que queda representado en la Figura 4, donde fácilmente se puede apreciar como a partir del año 2013 ha tenido un crecimiento significativo. Adicional también queda de manifiesto como la analítica prescriptiva se esta usando en diferentes y variados campos tal como se muestra en la Figura 5, donde la temática de los artículos trabajados pertenecen a mas de 8 áreas del conocimiento.

Las tres metodologías aunque sus etapas se llaman de forma diferente si que son análogas entre si, y a nivel general se puede apreciar que la mayoría de las etapas pertenecen a tres grupos en particular: Datos, Modelado y Evaluación, situación que se hace visible en la graficas originadas a partir de las Tablas 2, 4 y 6, donde los mayores porcentajes se presentan en las etapas relacionadas con los grupos antes mencionados. De esta manera se hace notar que las metodologías de Minería de Datos son afines a los procesos de desarrollo de proyectos de analítica prescriptiva.

Teniendo en cuenta lo anterior se hace viable proponer una metodología de trabajo para proyectos de analítica prescriptiva a partir de las etapas presentes en las metodologías KDD, CRISP-DM y SEMMA.

La metodología que se propone surge a partir de las etapas mas usadas y robustas de las metodologías de Minería de Datos analizadas. Entre estas la metodología CRISP-DM mostro una mejor similitud con el flujo de desarrollo de los artículos consultados y que además de poseer una mejor documentación[11] y un uso generalizado en el mercado, la hacen la candidata mas apropiada para usar su etapas en la presente propuesta.



Figura 10. Metodología propuesta para proyectos de Analítica Prescriptiva. [El Autor]

Para generar el anterior flujo se tuvo en cuenta:

Comprensión de Datos: Debido a que mas allá de tener datos para usar se debe hacer un ejercicio juicioso de saber su origen y finalidad.

Preparación de los datos: Debido a que incluye la selección y limpieza de los datos.

Modelado: Debido a que en esta etapa se seleccionan y utilizan herramientas, métodos y técnicas pertinentes al problema en cuestión.

Evaluación: Debido a que en esta etapa se comprueba la efectividad del proceso realizado además de revisar todos los pasos ejecutados.

REFERENCIAS

- [1] Ministerio de tecnologías de la información y las comunicaciones. (2019). Aspectos Básicos de la Industria 4.0. Republica de Colombia.
- [2] Tukey, J. (1962). The future of the data analysis. Princeton University & Bell Telephone Laboratories.
- [3] Koochang, A. & Horn, J. (2020). Critical components of data analytics in organizations: A research model.
- [4] Lepenioti, K. Bousdekis, A. Apostolou, D. & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges.
- [5] Azevedo, A. & Santos, M. (2008). KDD, SEMMA AND CRISP-DM: A parallel overview.
- [6] Mariscal, G. Marban, O. & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies.
- [7] Martínez de Pisón, F. Ordieres, J. Castejon, M. De Cos, F. & Pernía, A. (2001) Gestión del Conocimiento y Minería de Datos. Murcia: Actas del XVII Congreso Nacional de Ingeniería de Proyectos.
- [8] Zaki, M. & Meira, W. (2014). Data Mining and Analysis Fundamental Concepts and Algorithms. Cambridge university Press.
- [9] Piatetsky-Shapiro, G. (1991). Report on the AAAI-91 Workshop on Knowledge Discovery in Databases. Technical report 6, IEEE Expert.
- [10] Fayyad, U. Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview.
- [11] CRISP-DM 1.0 - Step-by-step data mining guide (2000). SPSS.
- [12] SAS Institute 2005. Semma Data Mining Methodology. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>.
- [13] Metaanálisis. Sociedad Española de Medicina de Laboratorio.
- [14] Gröger, C. Schwarz, H. & Mitschang, B. (2014). Prescriptive Analytics for Recommendation-Based Business Process Optimization.

[15] Srinivas, S. & Ravindran, R. (2018). Optimizing Outpatient Appointment System using Machine Learning Algorithms and Scheduling Rules: A Prescriptive Analytics Framework.

[16] Ghoniem, A. Ali, A. Al-Salem, M. & Khallouli, W. (2016). Prescriptive analytics for FIFA World Cup lodging capacity planning.

[17] Ayhans, S. Costas, P. & Hanan, S. (2018). Prescriptive analytics system for long-range aircraft conflict detection and resolution.

[18] Sanchez-Meca, J. (2010). Como realizar una revisión sistemática y un meta-análisis.

ANEXOS

ANEXO 1: Artículos para el Metaanálisis

# ARTICULO	ID	TITULO DOCUMENTO	# REFERENCIADO	DOI	AÑO
1	GrSM2014	Prescriptive Analytics for Recommendation-Based Business Process Optimization	26	10.1007/978-3-319-06695-0_3	2014
2	BLPW2020	Prescriptive analytics for reducing 30-day hospital readmissions after general surgery	0	10.1371/journal.pone.0238118	2020
3	MPGThWGKPSHV2020	Price Investment using Prescriptive Analytics and Optimization in Retail	0	10.1145/3394486.3403365	2020
4	PSAChShB2020	Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming	1	10.1016/j.dss.2020.113290	2020
5	VAOPN2020	Prescriptive analytics for swapping aircraft assignments at all Nippon airways	0	10.1287/inte.2019.1016	2020
6	SrR2018	Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework	22	10.1016/j.eswa.2018.02.022	2018
7	KrWL2016	Prescriptive Control of Business Processes: New Potentials Through Predictive Analytics of Big Data in the Process Manufacturing Industry	21	10.1007/s12599-015-0412-2	2016
8	KSSV2013	Prescriptive analytics for allocating sales teams to opportunities	11	10.1109/ICDMW.2013.156	2013
9	AyCS2018	Prescriptive analytics system for long-range aircraft conflict detection and resolution	7	10.1145/3274895.3274947	2018
10	JQH2016	Prescriptive analytics for understanding of out-of-plane deformation in additive manufacturing	7	10.1109/COASE.2016.7743482	2016
11	VBPRSFr2015	An Information System for Sales Team Assignments Utilizing Predictive and Prescriptive Analytics	7	10.1109/CBI.2015.38	2015
12	CSACrDOPS2019	Prescriptive maintenance of railway infrastructure: From data analytics to decision support	4	10.1109/MTITS.2019.8883331	2019

13	GhAA-SKh2017	Prescriptive analytics for FIFA World Cup lodging capacity planning	3	10.1057/s41274-016-0143-x	2017
14	WY2017	The green fleet optimization model for a low-carbon economy: A prescriptive analytics	3	10.1109/ICASI.2017.7988358	2017
15	AS2018	Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights	2	10.1016/j.orp.2018.08.004	2018
16	dAGrc2017	Prescstream: A framework for streaming soft real-time predictive and prescriptive analytics	2	10.1007/978-3-319-62392-4_24	2017
17	GLSchT2020	Prescriptive analytics for inventory management in health	1	10.1080/01605682.2020.1776167	2020
18	WP2019	Prescriptive cluster-dependent support vector machines with an application to reducing hospital readmissions	1	10.23919/ECC.2019.8796082	2019
19	HRGNgKV2018	Prescriptive analytics through constrained bayesian optimization	1	10.1007/978-3-319-93034-3_27	2018