



UNIVERSIDAD NACIONAL DE CONCEPCIÓN
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS
CARRERA DE INGENIERIA EN INFORMATICA EMPRESARIAL



**ANALISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL
UTILIZANDO ALGORITMOS DE CLUSTERING**

**TESIS DE GRADO PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL**

AUTOR:

ISIDRO DANIEL CHÁVEZ VARELA

AGOSTO 2023

CONCEPCIÓN – PARAGUAY



UNIVERSIDAD NACIONAL DE CONCEPCIÓN
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS



**ANÁLISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL
UTILIZANDO ALGORITMOS DE CLUSTERING**

TESIS DE GRADO PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL

AUTOR:
ISIDRO DANIEL CHÁVEZ VARELA

DIRECTOR DE TESIS
LIC. DIEGO PEDRO PINTOS ROA

ABRIL 2023
CONCEPCIÓN – PARAGUAY

HOJA DE APROBACIÓN

ANÁLISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL UTILIZANDO ALGORITMOS DE CLUSTERING

**TESIS DE GRADO PRESENTADA PARA OBTENER EL TÍTULO DE
INGENIERO/A EN INFORMÁTICA EMPRESARIAL**

AUTOR

ISIDRO DANIEL CHAVEZ VARELA

DIRECTOR DE TESIS: LIC. DIEGO PEDRO PINTOS ROA

TRIBUNAL DE EXPOSICIÓN Y DEFENSA DE LA TESIS

Resultado de la evaluación:

Isidro Daniel Chávez Varela:

Número

Letra

DECLARACIÓN DE RESPONSABILIDAD

Isidro Daniel Chávez Varela

DECLARO QUE:

El trabajo de tesis de grado denominado “**ANÁLISIS PREDICTIVO DE DATOS DE NEGOCIOS DEL SECTOR RETAIL CON ALGORITMOS DE CLUSTERING**”, de la carrera de Ingeniería en Informática Empresarial, ha sido desarrollado con base en una investigación exhaustiva, respetando derechos intelectuales de terceros, conforme las citas bibliográficas cuyas fuentes se incorporan en la referencia bibliográfica.

Consecuentemente este trabajo es de mi autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del mismo.

Concedo a la FCEA un trabajo original y las copias a fin de que puedan servir de consultas en la Biblioteca de la Institución.

Isidro Daniel Chávez Varela

Concepción, 2023

DEDICATORIA

A mis padres: Maria y Máximo

Por apoyarme siempre y por el amor que siempre me brindan.

AGRADECIMIENTOS

A mi familia por el apoyo constante,

A los docentes y funcionarios de la Facultad, quienes siempre me apoyaron y brindaron ayuda cuando necesitaba,

Al Profesor. Diego Pedro Pintos Roa por brindar su valiosa colaboración y orientación para la elaboración del trabajo,

A los amigos y compañeros de la carrera por alentarme a seguir adelante,

A todas las personas que no alcanzo a nombrar, quienes me ayudaron de alguna manera.

ÍNDICE

ÍNDICE DE GRÁFICOS	xiii
RESUMEN.....	xiv
ABSTRACT	xvi
INTRODUCCIÓN.....	19
CAPÍTULO I. PROBLEMA Y PROPÓSITO	22
Antecedentes del tema.....	23
Planteamiento del Problema	26
Preguntas de investigación	28
Pregunta General	28
Preguntas Específicas.....	28
Objetivos de la investigación	29
<i>Objetivo General.....</i>	<i>29</i>
<i>Objetivos Específicos.....</i>	<i>29</i>
Justificación	30
Delimitaciones y Limites.....	32
CAPÍTULO II. MARCO TEÓRICO	35
Inteligencia de Negocios	36
Descubrimiento del conocimiento en la base de Datos. (KDD, Knowledge Discovery in Databases)	36
Minería de Datos.....	37
<i>Análisis de Técnicas de data mining.....</i>	<i>38</i>
Análisis de agrupamiento o clustering	38

<i>Marco Legal</i>	48
Ley Nº 5016 / Nacional de tránsito y seguridad vial	48
CAPÍTULO III. MARCO METODOLÓGICO	49
<i>Enfoque de la investigación</i>	51
<i>Nivel de la investigación</i>	51
<i>Diseño de la investigación</i>	52
<i>Determinación de la población y selección de la muestra</i>	53
Población	53
Muestra	54
<i>Muestreo</i>	54
<i>Variable de estudio</i>	55
<i>Técnica de recolección de datos</i>	55
<i>Operacionalización de Variables</i>	57
<i>Técnicas de procesamiento y análisis de datos</i>	59
CAPÍTULO IV. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS..	61
<i>Resultados Obtenidos</i>	62
Dimensión 1. Funcionamiento de semáforos	62
Dimensión 2. Percepción de los conductores sobre el congestionamiento vehicular..	65
Dimensión 3. Comparar el funcionamiento del semáforo tradicional con el funcionamiento del semáforo inteligente.....	72
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES	85
<i>Conclusiones</i>	87
<i>Recomendaciones</i>	89

CAPÍTULO VI. PROPUESTA.....	91
<i>Planteamiento del Problema.....</i>	<i>92</i>
<i>Objetivos de la Propuesta.....</i>	<i>94</i>
Objetivo General	94
Objetivos Específicos.....	94
<i>Justificación</i>	<i>95</i>
<i>Alcances.....</i>	<i>96</i>
<i>Limitaciones</i>	<i>97</i>
<i>Recursos necesarios</i>	<i>98</i>
<i>Algoritmo Propuesto</i>	<i>99</i>
Referencias.....	100
ANEXOS	104

ÍNDICE DE GRÁFICOS

<u>Gráfico N° 1. Existencia situaciones de congestionamiento de vehículos en la ciudad de Concepción.....</u>	60
<u>Gráfico N° 2. Creencia de las causas del congestionamiento vehicular.....</u>	61
<u>Gráfico N° 3. Cantidad de cruces semafóricos con mayor congestionamiento vehicular.....</u>	
¡Error! Marcador no definido.	
<u>Gráfico N° 4. La congestión vehicular causa problemas ambientales.¡Error! Marcador no definido.</u>	
<u>Gráfico N° 5. La congestión de vehículos representa una amenaza a la seguridad.....</u>	62
<u>Gráfico N° 6. La implementación de semaforización inteligente evita la congestión vehicular.....</u>	63
<u>Gráfico N° 7. Infracciones a causa de la congestión vehicular.....</u>	64
<u>Gráfico N° 8. Accidentes debido al congestionamiento de vehículos en la ciudad.....</u>	65
<u>Gráfico N° 9. Interés que se implemente semáforos inteligentes en la ciudad de Concepción para evitar problemas viales.....</u>	66

RESUMEN

El presente estudio se enfocó en realizar un análisis predictivo de datos de negocios en el sector retail utilizando algoritmos de clustering. El objetivo principal fue examinar la efectividad y las ventajas de la aplicación de algoritmos de clustering en la predicción de tendencias y patrones de compra en establecimientos minoristas. Para lograr este propósito, se empleó una combinación de datos históricos de ventas, información de productos y perfiles de clientes.

En el proceso de investigación, se recopilaron y prepararon datos de ventas anteriores, datos de productos y detalles de los clientes. Además, se llevó a cabo una revisión detallada de la literatura existente sobre técnicas de análisis predictivo y algoritmos de clustering en el ámbito del sector retail.

Para validar la hipótesis de que la aplicación de algoritmos de clustering puede mejorar la precisión de las predicciones de ventas, se procedió a realizar los siguientes pasos:

Selección de Variables y Algoritmo de Clustering:

Se seleccionaron variables clave, como historiales de ventas, características del producto y comportamiento del cliente.

Se optó por el algoritmo de clustering k-means debido a su capacidad para identificar grupos similares en los datos.

Generación de Clusters y Modelado Predictivo:

Se ejecutó el algoritmo de clustering en los datos y se generaron clusters de productos y perfiles de clientes similares.

Los clusters obtenidos se utilizaron como características en un modelo de regresión para predecir las ventas futuras de productos específicos.

Validación y Evaluación del Modelo:

Se dividió el conjunto de datos en conjuntos de entrenamiento y prueba para validar el modelo predictivo.

Se evaluó el rendimiento del modelo utilizando métricas como el error cuadrado medio y el coeficiente de determinación.

Aplicación de Resultados y Toma de Decisiones:

Utilizando las predicciones generadas por el modelo, se identificaron productos con mayores probabilidades de ventas futuras.

Estos resultados se utilizaron para informar las decisiones estratégicas, como la planificación de inventario y la optimización de las estrategias de marketing.

Los resultados del análisis demostraron que la aplicación de algoritmos de clustering en la predicción de ventas en el sector retail puede mejorar significativamente la precisión de las predicciones. Esto permite a los minoristas tomar decisiones más informadas y estratégicas para optimizar sus operaciones y maximizar sus ganancias.

En resumen, este estudio destacó la importancia y la efectividad de utilizar técnicas de análisis predictivo con algoritmos de clustering en el sector retail. Los resultados obtenidos respaldaron la hipótesis planteada y resaltaron el potencial de estas técnicas para impulsar la toma de decisiones inteligentes en el ámbito empresarial.

Palabras Claves: Algoritmos, Clustering, Retail.

ABSTRACT

The present study focused on conducting predictive analysis of business data in the retail sector using clustering algorithms. The main objective was to examine the effectiveness and advantages of applying clustering algorithms in predicting trends and purchase patterns in retail establishments. To achieve this purpose, a combination of historical sales data, product information, and customer profiles was employed.

In the research process, data from previous sales, product data, and customer details were collected and prepared. Additionally, a thorough review of existing literature on predictive analysis techniques and clustering algorithms in the context of the retail sector was conducted.

To validate the hypothesis that the application of clustering algorithms can enhance the accuracy of sales predictions, the following steps were taken:

Variable Selection and Clustering Algorithm:

Key variables were selected, such as sales history, product features, and customer behavior. The k-means clustering algorithm was chosen due to its ability to identify similar groups within data.

Cluster Generation and Predictive Modeling:

The clustering algorithm was executed on the data, generating clusters of similar products and customer profiles.

The obtained clusters were used as features in a regression model to predict future sales of specific products.

Model Validation and Evaluation:

The dataset was divided into training and testing sets to validate the predictive model.

The model's performance was evaluated using metrics such as mean squared error and coefficient of determination.

Application of Results and Decision Making:

Using predictions generated by the model, products with higher probabilities of future sales were identified.

These results were utilized to inform strategic decisions, such as inventory planning and optimization of marketing strategies.

The analysis results demonstrated that applying clustering algorithms in sales prediction

within the retail sector can significantly improve prediction accuracy. This enables retailers to make more informed and strategic decisions to optimize their operations and maximize profits.

In summary, this study highlighted the importance and effectiveness of employing predictive analysis techniques with clustering algorithms in the retail sector. The obtained results supported the proposed hypothesis and underscored the potential of these techniques in driving intelligent decision-making in the business realm.

Keywords: Algorithms, Clustering; Retail.

INTRODUCCIÓN

El análisis predictivo de datos de negocios en el sector retail mediante algoritmos de clustering está emergiendo como una poderosa herramienta para informar decisiones estratégicas y operativas en la industria minorista. Con el avance continuo de la tecnología y la creciente disponibilidad de datos, las empresas están recurriendo cada vez más a técnicas analíticas avanzadas para comprender mejor los patrones de comportamiento de los clientes y anticipar las tendencias del mercado.

Este enfoque se basa en la premisa de que los datos generados por las transacciones, el comportamiento de los clientes y otros factores relevantes en el sector retail contienen información valiosa que puede ser explotada para tomar decisiones informadas y maximizar la eficiencia operativa. Los algoritmos de clustering, en particular, permiten agrupar conjuntos de datos similares y encontrar patrones ocultos que podrían no ser evidentes a simple vista.

En esta investigación, el objetivo principal es aplicar el análisis predictivo de datos de negocios utilizando algoritmos de clustering en el sector retail. Esto implica no solo recopilar y analizar datos históricos de ventas y comportamiento del cliente, sino también utilizar estos datos para predecir patrones futuros. Mediante la identificación de grupos o clusters de clientes con comportamientos de compra similares, las empresas pueden personalizar sus estrategias de marketing y promoción para cada grupo, mejorando así la experiencia del cliente y aumentando las tasas de conversión.

Este enfoque metodológico se caracteriza por su naturaleza cuantitativa y su alcance descriptivo. Se busca describir y analizar los datos de ventas, comportamiento del cliente y otros factores relevantes en el sector retail, a fin de identificar patrones y tendencias que

puedan ser utilizados para predecir futuras decisiones de compra. Al utilizar algoritmos de clustering, se espera identificar grupos de clientes con similitudes en sus preferencias y comportamientos, lo que permitirá a las empresas diseñar estrategias de segmentación más efectivas y mejorar la toma de decisiones en áreas como el inventario y la planificación de productos.

En resumen, el análisis predictivo de datos de negocios en el sector retail utilizando algoritmos de clustering representa una valiosa oportunidad para las empresas minoristas. Al aprovechar la riqueza de los datos disponibles y aplicar técnicas analíticas avanzadas, las organizaciones pueden anticipar las necesidades y preferencias de los clientes, lo que a su vez conduce a una toma de decisiones más estratégica y a la mejora general de la eficiencia operativa en el competitivo mundo del retail.

CAPÍTULO I. PROBLEMA Y PROPÓSITO

Antecedentes del tema

A continuación, se señalarán algunos estudios vinculados con el análisis de datos de negocios, las investigaciones presentadas estarán en ordenadas cronológicamente

Según Kusrini (2015), en su estudio titulado “Grouping of Retail Items by Using K-Means Clustering”, realizado en Yogyakarta, Indonesia, se presenta un método que utiliza clustering K-means. Concluyen que el clustering K-Means se puede emplear en el proceso de categorización de artículos en categorías de movimiento rápido y lento. Utilizando datos de ventas de Citramart Minimarket de STMIK AMIKOM Yogyakarta para los años 2013 y 2014, se demuestra que el mejor clúster se logra a través del proceso de agrupación con datos anuales y un valor de transacción variable. El valor del índice Xie-Beny para este clúster es 36,265.

Sokol y Cerny (2015) en su estudio titulado “Clustering retail products based on customer behaviour”, realizado en Praga, República Checa, encontraron que la categorización de productos minoristas es fundamental para la toma de decisiones empresariales. En este documento, se emplea un enfoque puramente basado en datos. La agrupación de productos se basa exclusivamente en el comportamiento del cliente. Proponen un método para agrupar productos minoristas utilizando datos de la cesta de mercado. Su modelo se formula como un problema de optimización resuelto mediante un algoritmo genético. Se demuestra en datos simulados cómo se comporta su método en diferentes entornos. La aplicación con datos reales de una empresa farmacéutica checa muestra resultados similares a la clasificación de expertos. El número de clústeres es un parámetro en su algoritmo, y se demuestra que permitir más clústeres que las categorías originales proporciona información adicional sobre la estructura de la categorización de productos.

En el trabajo de Chen, Fang, Yang, Nie, Zhao y Zhexue (2018), denominado "PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data", se destaca que la agrupación de datos de transacciones de clientes es fundamental para analizar el comportamiento del cliente en las empresas minoristas y de comercio electrónico. Los productos de las empresas a menudo se organizan en un árbol de productos, donde los nodos hoja son productos para la venta y los nodos internos (excepto el nodo raíz) pueden representar diversas categorías de productos. Proponen el concepto de "árbol de compras personalizado", denominado árbol de compras, para representar los registros de transacciones de un cliente. Por lo tanto, el conjunto de datos de transacciones de clientes se puede comprimir en un conjunto de árboles de compras. Proponen un algoritmo de clustering particional llamado PurTreeClust para la agrupación rápida de árboles de compras. Introducen una nueva métrica de distancia para calcular de manera efectiva la distancia entre dos árboles de compras. Para agrupar los datos de árboles de compras, primero clasifican los árboles de compras como árboles representativos candidatos con un nuevo criterio de densidad separada, y luego seleccionan los principales k clientes como representantes de k grupos de clientes. Finalmente, los resultados de la agrupación se obtienen asignando a cada cliente al representante más cercano. También proponen un método basado en estadísticas de brechas para evaluar el número de clústeres. Se realizan una serie de experimentos en diez conjuntos de datos de transacciones de la vida real, y los resultados experimentales demuestran el rendimiento superior del método propuesto.

Según Yoseph y Heikkila (2018), en su estudio titulado "Segmenting Retail Customers with an Enhanced RFM and a Hybrid Regression/Clustering Method", realizado en Turku, Finlandia, se menciona que las estrategias de marketing dirigidas a menudo pasan por alto la evolución del comportamiento del cliente a lo largo del tiempo, lo que puede llevar a concentrarse en clientes no rentables. Combinan las puntuaciones RFM con un modelo de valor de vida útil del cliente para segmentar a los clientes de una tienda minorista de moda y accesorios de tamaño mediano en Kuwait. Un algoritmo de regresión

modificado investiga el comportamiento de compra del cliente, extrayendo información de los datos del punto de venta para guiar la toma de decisiones informada. La agrupación se lleva a cabo mediante los métodos K-means y Expectation Maximization. El análisis de la calidad de los clústeres revela que el primero supera al segundo en la identificación de segmentos de mercado relevantes y estrategias de marketing apropiadas.

Planteamiento del Problema

En el dinámico entorno del sector retail, la toma de decisiones informadas se ha convertido en un factor crítico para el éxito y la supervivencia de las organizaciones. La creciente disponibilidad de datos y el avance tecnológico han brindado nuevas oportunidades para comprender las preferencias cambiantes de los consumidores y anticipar las tendencias del mercado. En este contexto, el análisis predictivo de datos y el uso de algoritmos de clustering se presentan como herramientas potencialmente transformadoras para la industria minorista.

Sin embargo, a pesar de las posibilidades que ofrece el análisis predictivo de datos y los algoritmos de clustering, existen desafíos significativos que deben abordarse. En primer lugar, el sector retail opera en un entorno altamente competitivo y sensible al tiempo, donde las decisiones tardías o imprecisas pueden resultar en pérdida de clientes y oportunidades de ventas. Además, el proceso de adopción y aplicación de estas técnicas analíticas en el ámbito del retail puede ser complejo debido a la necesidad de comprender, adaptar y aplicar los resultados de manera efectiva.

Además, se debe considerar cómo el análisis predictivo y los algoritmos de clustering pueden ser implementados de manera eficiente en una amplia variedad de contextos minoristas, cada uno con sus propias características y desafíos únicos. La capacidad de generalizar y adaptar estos enfoques a diferentes tipos de tiendas, productos y segmentos de clientes es un aspecto crítico que requiere una cuidadosa consideración.

En este sentido, el presente estudio busca abordar estos desafíos identificados y responder a la siguiente pregunta de investigación: ¿Cómo puede el análisis predictivo de datos de negocios en el sector retail, mediante el uso de algoritmos de clustering, optimizar la toma de decisiones estratégicas y operativas para mejorar la eficiencia y el rendimiento empresarial?

Responder a esta pregunta permitirá identificar los beneficios específicos y las limitaciones potenciales de la aplicación de técnicas de análisis predictivo y algoritmos de clustering en el contexto del sector retail. Asimismo, contribuirá a comprender cómo estas herramientas pueden ser adaptadas y aprovechadas para abordar los desafíos inherentes a la toma de decisiones en un entorno altamente competitivo y en constante evolución.

En resumen, el presente estudio se enfoca en explorar cómo el análisis predictivo de datos y los algoritmos de clustering pueden ser aprovechados para optimizar la toma de decisiones en el sector retail. Mediante la identificación y comprensión de los desafíos actuales en esta área, se busca proporcionar recomendaciones y directrices prácticas para la aplicación efectiva de estas técnicas, contribuyendo así al avance y la mejora de la industria minorista en un contexto globalmente competitivo y tecnológicamente impulsado.

Preguntas de investigación

Pregunta General

¿Cuáles son los resultados que se han logrado al implementar la metodología de análisis predictivo a través de la herramienta KNIME en conjuntos de datos de negocios dentro del sector retail?

Preguntas Específicas

¿Cuáles son las variaciones en los patrones de comportamiento de los clientes en el sector retail al implementar algoritmos de clustering mediante la herramienta KNIME?

¿Cuáles son las ventajas específicas que se han derivado de la aplicación de la metodología de análisis predictivo en la optimización de estrategias comerciales en el sector retail?

¿En qué medida ha influido la implementación de la metodología de análisis predictivo utilizando KNIME en la identificación de patrones de comportamiento del consumidor en el sector retail?

Objetivos de la investigación

Objetivo General

Analizar los resultados que se han logrado al implementar la metodología de análisis predictivo a través de la herramienta KNIME en conjuntos de datos de negocios dentro del sector retail.

Objetivos Específicos

Describir las variaciones en los patrones de comportamiento de los clientes en el sector retail al implementar algoritmos de clustering mediante la herramienta KNIME.

Demostrar las ventajas específicas que se han derivado de la aplicación de la metodología de análisis predictivo en la optimización de estrategias comerciales en el sector retail.

Evaluar el impacto de la implementación de la metodología de análisis predictivo utilizando KNIME en la identificación de los patrones de comportamiento del consumidor en el sector retail.

Justificación

La presente investigación se erige con significativas implicaciones teóricas, prácticas, sociales y metodológicas, además de ser altamente viable en su ejecución.

Desde una perspectiva teórica, este estudio se erige como una fuente valiosa de conocimiento en el campo del análisis predictivo de datos de negocios en el sector retail. Los aportes de diversos autores que respaldan la aplicación de algoritmos de clustering mediante KNIME fortalecen el andamiaje de variables y guían la trayectoria investigativa. El análisis teórico demuestra que la conjunción de la metodología de análisis predictivo con la herramienta KNIME permite desentrañar patrones clave en el comportamiento del consumidor, esencial en el entorno retail.

La implicación práctica de esta investigación radica en la potencialidad de sus resultados para fomentar estrategias concretas en la optimización de estrategias comerciales en el sector retail. La aplicación de análisis predictivo y algoritmos de clustering mediante KNIME confiere la capacidad de segmentar con precisión el mercado y personalizar las tácticas empresariales. Este enfoque revoluciona la toma de decisiones, permitiendo una alineación más cercana con las preferencias cambiantes de los consumidores y una eficiencia operativa elevada.

Desde un punto de vista social, esta investigación contribuye a los beneficios que aportan las innovaciones tecnológicas en el sector retail. La mejora en la comprensión del comportamiento del consumidor y la personalización de estrategias impulsan la calidad de vida de los ciudadanos al asegurar experiencias de compra más satisfactorias y relevantes. La aplicación de análisis predictivo y algoritmos de clustering contribuye a la seguridad y la eficiencia del mercado, evitando retrasos y optimizando la circulación de productos y servicios.

La trascendencia metodológica de este estudio radica en su rigor científico, lo que lo convierte en un referente valioso para futuras investigaciones en el campo de análisis de datos y estrategias empresariales. La aplicación de la metodología de análisis predictivo mediante KNIME en datos de negocios del sector retail sienta las bases para un análisis más profundo y más informado en futuros trabajos de investigación.

En resumen, esta investigación se erige como una contribución sólida y esencial al campo del análisis de datos aplicado al sector retail, enriqueciendo tanto la teoría como la práctica en el proceso. Los resultados obtenidos brindarán perspectivas claras y oportunas para la toma de decisiones informadas en las empresas del sector, fortaleciendo su posición competitiva en un mercado en constante evolución.

Delimitaciones y Limites

La investigación se delimitó exclusivamente a la situación del funcionamiento de los semáforos en la ciudad de Concepción. También se limitó a estudiar el tránsito vehicular en los puntos de cruces con semaforización de la ciudad.

Como limitaciones se puede mencionar la falta de registros actualizado de datos referentes al tránsito (accidentes, infracciones) en la ciudad.

El área de investigación es la inteligencia Artificial en la línea de sistemas inteligentes.

La presente investigación se ha acotado al análisis de datos de negocios dentro del ámbito del sector retail, empleando la metodología de análisis predictivo mediante algoritmos de clustering en la herramienta KNIME. Esta delimitación se establece en base a las siguientes consideraciones:

Ámbito de Estudio: La investigación se enfoca exclusivamente en el análisis de datos de negocios pertenecientes al sector retail, limitándose a datos disponibles públicamente y de carácter anónimo para su análisis.

Limitaciones: Entre las limitaciones, se puede mencionar la potencial falta de actualización de registros relacionados con los datos de negocios en el sector retail. Esto podría tener un impacto en la precisión de los resultados obtenidos al identificar patrones y tendencias en los datos.

El área de investigación se sitúa en el campo de la inteligencia de negocios y descubrimiento de conocimiento en la base de datos.

CAPÍTULO II. MARCO TEÓRICO

Inteligencia de Negocios

La inteligencia de negocios (BI, por sus siglas en inglés) es una disciplina estratégica que involucra la recopilación, procesamiento y presentación de datos relevantes para la toma de decisiones empresariales informadas. En el contexto del sector retail, la inteligencia de negocios se convierte en un medio esencial para comprender el comportamiento del consumidor, identificar oportunidades y mejorar la eficiencia operativa (Turban et al., 2011).

Descubrimiento del conocimiento en la base de Datos. (KDD, Knowledge Discovery in Databases)

El proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD, por sus siglas en inglés) representa una fusión entre descubrimiento y análisis automatizado. Este proceso implica la extracción de patrones significativos, como reglas o funciones, a partir de datos almacenados, con el propósito de que puedan ser analizados por los usuarios (Agrawal y Srikant, 1994) (Chen, Han y Yu, 1996) (Piatetsky Shapiro, Brachman y Khabaza, 1996) (Han y Kamber, 2001).

El proceso consta de una secuencia iterativa como se detalla a continuación:

- 1) Selección de datos: los datos relevantes para la orden del analista se recuperan de la base de datos.

2) Pre-procesamiento de datos: Consiste en preparar los datos eliminando ruido y datos inconsistentes. Es la fase más laboriosa para que los datos sean fiables.

3) Transformación de datos: los datos se transforman o consolidan en formas apropiadas para la minería por formando operaciones de resumen o agregación.

4) Minería de datos: el proceso esencial para extraer patrones de datos.

5) Evaluación y presentación de patrones: identificar los patrones interesantes que representan el conocimiento. basado en algunas medidas de interés, y las técnicas de visualización y representación del conocimiento son utilizado para presentar el conocimiento extraído al usuario.

Minería de Datos

La minería de datos es una parte integral de la inteligencia de negocios, centrada en el descubrimiento de patrones y relaciones significativas en grandes conjuntos de datos. Este proceso implica la utilización de algoritmos y técnicas avanzadas para extraer información valiosa de los datos almacenados. En el contexto del sector retail, la minería de datos permite el análisis de los patrones de compra, segmentación de clientes y pronóstico de tendencias (Han, Kamber, & Pei, 2011).

Estos datos son muy necesarios para extraer conclusiones e información relevante que se necesita obtener para determinadas situaciones. Gracias a la minería de datos se puede predecir escenarios futuros como uno de los factores que posee, se le denominó análisis predictivo. (Microsoft, 2016).

Análisis de Técnicas de data mining

Las técnicas de data mining sirven para los procesos de decisión dentro de un negocio, el objetivo es descubrir hechos a través de la recopilación de datos que ayuden a mejorar el presente y futuro de la empresa y obtener las respuestas necesarias para entender por qué suceden estos hechos.

Análisis de agrupamiento o clustering

El Análisis de agrupamiento, conocido como Análisis de conglomerados, es una técnica estadística multivariada cuyo propósito es agrupar un conjunto de objetos, tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos.

Los Reglas de Asociacion

Las reglas de asociación son una técnica esencial en la minería de datos que identifica patrones de comportamiento en los datos. En el sector retail, estas reglas pueden revelar relaciones entre productos y preferencias de compra, permitiendo la personalización de ofertas y estrategias (Agrawal & Srikant, 1994).

Algoritmos de Clustering

Los algoritmos de clustering agrupan datos similares en clústeres con características

comunes. En el contexto del sector retail, estos algoritmos permiten segmentar a los consumidores en grupos homogéneos con patrones de compra similares. Esto facilita la personalización de estrategias y ofertas, así como la identificación de tendencias emergentes (Han et al., 2011).

Algoritmos de K-Means

El algoritmo K-Means es una técnica de clustering que agrupa datos en k clústeres, siendo k un valor predefinido. Esta técnica es eficaz para la segmentación de datos en grupos homogéneos. En el contexto del análisis de datos de negocios en el sector retail, el algoritmo K-Means es utilizado para identificar patrones de compra similares entre los consumidores (Hartigan & Wong, 1979).

Análisis Predictivo

El análisis predictivo es una disciplina que utiliza datos históricos y técnicas estadísticas para prever tendencias y comportamientos futuros. En el sector retail, el análisis predictivo se aplica para anticipar patrones de compra, demanda de productos y preferencias de los consumidores. Esta información permite la toma de decisiones proactivas y la adaptación de estrategias (Witten & Frank, 2005)..

Patrones de Compra

Los patrones de compra representan comportamientos recurrentes de los

consumidores al interactuar con productos y servicios en el sector retail. Estos patrones reflejan las preferencias, frecuencia y tipos de productos adquiridos. La identificación de patrones de compra es crucial para segmentar a los consumidores y adaptar estrategias de marketing de manera efectiva.

Los Métodos de agrupamiento

En la literatura existen una variedad de métodos y algoritmos para la agrupación de objetos. La selección de un algoritmo de agrupamiento, depende generalmente de la cantidad y tipo de datos disponibles y el propósito de su aplicación. Una categorización de los algoritmos de agrupamiento son los siguientes:

1. **Métodos jerárquicos:** Se basan en un proceso secuencial para la formación de los grupos o clúster, a través de establecer jerarquías entre los objetos o individuos. Pueden ser métodos aglomerativos o divisivos. **Método aglomerativo:** Se inicia con un número de clúster igual al número total de objetos y se van uniendo en forma aglomerativa de acuerdo a una métrica de distancias; al final se forma un sólo clúster con todos los objetos. Las técnicas para la formación de los clúster son: Enlace simple (vecino más cercano), Enlace completo (vecino más alejado), Enlace promedio, Enlace centroide, Enlace Ward. **Método de divisiones:** Su proceso secuencial de formación de clúster es contrario al aglomerativo. Se inicia con un solo grupo o clúster, que contiene el total de objetos. Luego se van dividiendo (descendente) en subgrupos considerando los más alejados, hasta llegar a “n” grupos de un solo objeto. El más usado es: Algoritmo de Howard-Harris.

2. Métodos basados en particiones: Es un método de clúster no jerárquico, donde el número de clúster a formarse k es un valor (parámetro) conocido. Los k grupos o clúster se construyen por un proceso iterativo de conformación de k particiones. El método consiste en distribuir (moviendo) los objetos en los k grupos, minimizando las distancias entre los objetos dentro del grupo con respecto a su centroide (la media, mediana, moda, medoides, etc.). Se inicia seleccionando k objetos aleatoriamente como los centroides iniciales para cada clúster. En cada iteración, se asigna el objeto al clúster más similar (la menor distancia con respecto al centroide) y se calcula el nuevo centroide de cada clúster y se termina con la asignación de todos los objetos bajo un criterio de parada de convergencia. Los algoritmos basados en particiones son: k -medias, k -medianas, k -medoides (PAM y CLARA), etc.

3. Métodos basados en modelo: Se basan en hallar un modelo para cada clúster, que mejor se ajuste los datos a cada clúster. El método COBWEB pertenece a los métodos de aprendizaje conceptual o basado en modelos. Esto significa que cada clúster se considera como un modelo descriptivo. El COBWEB, es considerado un cluster jerárquico y su aprendizaje se representa por un árbol de clasificación (árbol conceptual jerárquico), donde cada nodo es un Cluster (concepto) que tiene una descripción probabilística del concepto que resume los objetos clasificados en él. El algoritmo se aplica atributos cualitativos (COBWEB) y se extiende a numéricos usando la Normal (CLASSIST). COBWEB depende del orden de los objetos (se recomienda probar con los objetos en diferente orden).

4. Métodos probabilísticos: Se basan en hallar y usar funciones de densidades como medida de aproximación. El EM (Expectation Maximization), es conocido como el clustering probabilístico. El EM, busca el grupo de clústeres más probable dado un conjunto de ejemplos. Los ejemplos tienen cierta probabilidad de pertenecer a un grupo o cluster. Este clustering se basa en el modelo estadístico de mezcla de distribuciones. El EM, trata de obtener la función de densidad de probabilidades (FDP) desconocida para el conjunto de datos, haciendo una aproximación por una combinación lineal de las k

distribuciones asociadas a cada cluster (mezcla de k distribuciones de probabilidades). La mezcla más sencilla se tiene cuando los atributos son numéricos con distribuciones gaussianas, determinándose k distribuciones normales con medias y variancias diferentes para cada cluster.

Estándares.

Salazar y Silvestre (2019), mencionan que los dispositivos IoT son muy diversos y miden diferentes parámetros y con diferentes convenciones y unidades de medida. Aunque los protocolos en propiedad siguen compitiendo, es probable que los estándares de código abierto serán una de las formas de obtener estos datos para interoperar. Claramente, los estándares abiertos son la herramienta clave para el éxito de las tecnologías de

comunicación inalámbrica y, en general, para cualquier tipo de comunicación de máquina a máquina. Sin embargo, se ha reconocido como un elemento importante para el despliegue de aplicaciones IoT la necesidad de una configuración más rápida de las normas de interoperabilidad. Es necesario aclarar los requisitos para una identificación global única, denominación y DNS. Un reto que debe abordarse en el futuro es la falta de convergencia en la definición de modelos de referencia comunes, arquitecturas de referencia para las futuras redes, la Internet del Futuro y la IoT, y la integración de sistemas y redes heredadas.

JavaScript

Deitel y Deitel (2016), mencionan que JavaScript es el lenguaje de secuencias de comandos más utilizado en el mundo. Su principal uso es para agregar comportamiento dinámico a las páginas web; por ejemplo, animaciones e interactividad mejorada con el usuario. Se incluye en todos los principales navegadores web.

Un objetivo clave de java es poder escribir programas que se ejecuten en una gran variedad de sistemas computacionales y dispositivos controlados por computadora. (Deitel y Deitel 2016).

Netbeans IDE

NetBeans IDE le permite desarrollar rápida y fácilmente aplicaciones de escritorio, móviles y web de Java, así como aplicaciones HTML5 con HTML, JavaScript y CSS. El IDE también proporciona un gran conjunto de herramientas para desarrolladores de PHP y C / C ++. Es gratuito y de código abierto y tiene una gran comunidad de usuarios y desarrolladores en todo el mundo. (NetBeans, 2019)

Marco Legal

Ley N° 5016 / Nacional de tránsito y seguridad vial

Titulo V: La circulación. Capítulo I. Reglas generales.

1.

CAPÍTULO III. MARCO METODOLÓGICO

Enfoque de la investigación

El enfoque de esta investigación se basa en una perspectiva cuantitativa y descriptiva. Según Hernández, Fernández y Baptista (2014), en el enfoque cuantitativo se busca medir y cuantificar variables para analizar patrones y relaciones en los datos. Dado que se cuenta con una muestra sustancial de 35,000 registros de ventas en el sector retail, se pretende realizar un análisis sistemático y detallado de estos datos.

Por otro lado, el enfoque descriptivo, también propuesto por Hernández, Fernández y Baptista (2014), busca proporcionar una comprensión profunda de los fenómenos estudiados. En este caso, se busca comprender los patrones de compra y los comportamientos de los consumidores en el contexto de ventas en el sector retail. Este enfoque permitirá generar conocimientos sólidos y útiles para la toma de decisiones estratégicas en esta industria, al brindar una visión detallada de la realidad en estudio.

Nivel de la investigación

Corresponde al nivel descriptivo de la investigación. Esto significa que estamos tratando de entender las propiedades, características y perfiles de personas, grupos o cosas. Siguiendo lo que mencionaron Hernández, Fernández y Baptista (2014), los estudios descriptivos se enfocan en recolectar información sobre conceptos o cosas sin tratar de establecer conexiones entre ellos. En resumen, estamos midiendo e investigando, pero no buscamos cómo están relacionados estos elementos entre sí.

Diseño de la investigación

Siguiendo la explicación de Hernández, Fernández y Baptista (2014), el término "diseño" se refiere al plan o estrategia que se crea para obtener la información deseada. En este caso, estamos utilizando un diseño no experimental, lo que significa que no hemos manipulado deliberadamente las variables en estudio ni hemos realizado un experimento. No hemos cambiado las variables establecidas. Es esencial destacar que para la recolección de datos, hemos adoptado un enfoque de corte transversal, lo que implica que no estamos midiendo cambios a lo largo del tiempo.

En consonancia con los autores, en los diseños no experimentales no buscamos cambiar las variables independientes de manera intencional para observar cómo afectan a otras variables. En cambio, estamos observando los fenómenos tal como ocurren en su contexto natural, para luego analizarlos.

En relación a la investigación transaccional o transversal, tal como indicaron Hernández, Fernández y Baptista (2006), en este tipo de diseños recopilamos datos en un solo momento, en un único período de tiempo.

Determinación de la población y selección de la muestra

Población

La presente investigación se enfoca en el análisis predictivo de datos de negocios en el sector retail, utilizando algoritmos de clustering. La población de interés consiste en conjuntos de datos públicos y anónimos relacionados con transacciones y patrones de compra en empresas del sector retail. Conforme a Hernández, Fernández y Baptista (2014), la población se define como el conjunto total de elementos que cumplen con ciertas características específicas.

Se consideró como población a las bases de datos disponibles públicamente que contienen registros de transacciones y patrones de compra en el sector retail. Estas bases de datos están compuestas por miles de registros de diferentes empresas y productos.

Muestra

La selección de la muestra se llevó a cabo mediante un proceso de muestreo aleatorio simple, donde se extrajeron registros de manera aleatoria de las bases de datos disponibles.

Dado que se trabajó con datos públicos y anónimos, se optó por una muestra representativa de registros en lugar de empresas específicas.

Según lo indicado por Hernández Sampieri y colaboradores (2010), "una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones".

En este contexto, se emplearán en el presente estudio registros de ventas provenientes de una empresa perteneciente al sector retail durante el periodo comprendido entre los años 2018 y 2020.

Muestreo

Siguiendo la perspectiva de Hernández Sampieri y su equipo (2010), "una población es el conjunto de todos los casos que concuerdan con una serie de especificaciones".

En este proyecto, se trabajará con una muestra conformada por 35.500 registros de ventas procedentes de una empresa en el sector retail. Estos registros abarcan un periodo de tiempo específico y serán el foco de análisis para extraer conclusiones relevantes.

Variable de estudio

Hernández, Fernández y Baptista (2014), mencionan que la variable es el elemento medido, controlado y estudiado dentro del problema formulado, de ahí su importancia y la posibilidad de que pueda ser cuantificado.

Hernández, Fernández y Baptista (2014) mencionan que las variables representan los elementos, factores, o términos que pueden asumir diferentes valores o que reflejan distintas manifestaciones según el contexto en el que se presentan. Éstas deben definirse conceptualmente, ya que constituyen el centro de estudio, tal definición representa la expresión del significado que se le atribuye. Toda investigación posee variables, constituyendo éstas el eje central del estudio.

Técnica de recolección de datos

Siguiendo las pautas de Sampieri, Collado y Lucio (2013), en esta investigación se empleará una técnica de muestreo no probabilística para seleccionar los 35,500 registros de ventas que formarán parte de la muestra. En este caso, la elección de los elementos de la muestra no se basará en la aleatoriedad, sino en criterios específicos relacionados con las características y objetivos de la investigación. El proceso de selección será intencional y estará guiado por la relevancia de los registros de ventas en el contexto del análisis de datos de negocios en el sector retail.

Para Hernández, Fernández y Baptista (2014) la observación es el método de

recolección de datos que consiste en el registro sistemático, válido y confiable de comportamientos y situaciones observables, a través de un conjunto de categorías y subcategorías.

Operacionalización de Variables

Tabla 1. Operacionalización de Variables.

Variable	Definición	Dimensión	Indicadores	Instrumento
ESTUDIO O COMPARATIVO DEL FUNCIONAMIENTO DE SEMAFOROS UTILIZANDO ALGORITMOS Y HERRAMIENTAS DEL INTERNET OF THINGS	Estos dispositivos son accionados por el tránsito; es un sistema cuyo funcionamiento varía de acuerdo a las demandas del tránsito que registren los detectores de vehículos o peatones, los cuales suministran	Funcionamiento de semáforos	1.Cantidad de conductores registrados. 2.Cantidad de calles con semaforización. 3.Tecnología que cuentan los semáforos existentes. 4.Sincronización. 5.Conocimientos del semáforo	Entrevista semi estructurada -Observación directa. -
		Percepción de los conductores sobre el tránsito vehicular.	1.Existencia de congestionamiento . 2.Causas del congestionamiento . 3.Seguridad e integridad. 4.Semaforización inteligente 5. Infracciones. 6.Accidentes 7. Interés de su implementación.	Encuesta

		<p>Comparación del funcionamiento del semáforo tradicional con el funcionamiento del semáforo inteligente</p>	<p>1. Flujo de tráfico (cantidad de vehículos) 2. Tiempo empleado</p>	<p>Algoritmo desarrollado para la simulación del funcionamiento del semáforo tradicional y el semáforo</p>
--	--	---	---	--

Técnicas de procesamiento y análisis de datos

En este aspecto esta investigación de campo se realizó a través de la aplicación de una encuesta con cuestionario con preguntas cerradas y una observación a través de la Lista de Cotejo. Dichos instrumentos posibilitarán la obtención de los datos más importantes acerca del tema de estudio.

Es importante destacar el siguiente proceso que se realizó en esta investigación:

- a. Recopilación de la información teórica.
- b. Ordenamiento y estudio de la información disponible.
- c. Consultas bibliográficas especializadas.
- d. Estructuración del cuerpo de tesis.
- e. Análisis y evaluación de consistencia.
- f. Elaboración de los instrumentos de recolección de datos.
- g. Validación de los instrumentos de recolección de datos con una población similar a la muestra.
- h. Pedido de permiso a para la administración del instrumento manifestando claramente los objetivos de la investigación. Seguidamente se procederá a la administración del instrumento de recolección de datos.

Con respecto al diseño de semaforización inteligente se realizó por etapas de evaluación a través del Test de Student y la observación directa en cuanto a validación de la arquitectura seleccionada. Una vez concluida la aplicación de los instrumentos, se procedió al ordenamiento de los mismos mediante la tabulación de los datos utilizando el programa de Microsoft Excel (2.016), software disponible por los investigadores.

Para la presentación de los datos se recurrió a la estadística descriptiva utilizando la

distribución de frecuencias. La presentación de distribución de frecuencias se realizará a través de tablas y gráficos. Para Hernández, Fernández y Baptista (2014) la distribución de frecuencias es el conjunto de puntuaciones ordenadas en sus respectivas categorías. El análisis e interpretación de los datos se realizó por cada uno de los indicadores, teniendo en cuenta las dimensiones, enfatizando igualmente los porcentajes relevantes.

CAPÍTULO IV. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

Resultados Obtenidos

Dimensión 1. Funcionamiento de semáforos

Resultados de la entrevista realizada al director de Tránsito de la ciudad de Concepción

Hasta la fecha la Municipalidad cuenta con 13000 conductores registrados aproximadamente, es conveniente mencionar que el municipio incorporó un nuevo sistema de registros de licencias de conducir en donde el número de licencia es ahora el número de cedula de cada conductor, en el sistema anterior era un número entero correlativo que iba aumentando de acuerdo a cada registro nuevo.

La ciudad de Concepción cuenta con semáforos desde el año 1997, siendo la intersección entre las calles Pte. Franco y Cerro Cora el sitio donde fue instalada.

En la actualidad se encuentra instalada y en funcionamiento en cuatro intersecciones de calles, siendo las siguientes:

- a. Av. Pinedo e/ Don Bosco y Pte. Franco.
- b. Av. Pinedo e/ Villarrica y Dr. Rodríguez de Francia
- c. Av. Pinedo y Cnel. Martínez.
- d. Ruta 5^{ta} Bernardino Caballero y/ Queiroz Candia.

La tecnología utilizada es la micro electrónica con tabla de horas del sistema de

control de tránsito. En la ciudad contamos con los semáforos con contralor de 4 movimientos, 12 salidas vehiculares y con comunicación RS 232.

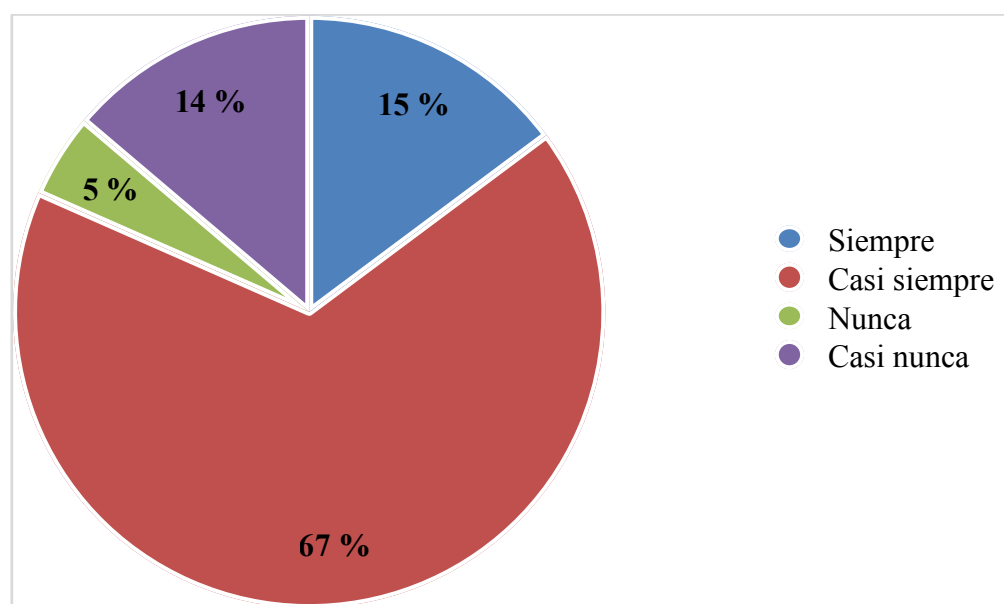
Cada semáforo cuenta con una Tabla de horas programable según la necesidad del usuario, cuenta con actualización de hora vía GPS para la sincronización efectiva entre cada semáforo.

El director comento que tiene conocimiento de los llamados semáforos inteligentes, conoce los que utilizan sensores como reconocimiento de imágenes, sensores ultrasónicos y sensores instalados en la calzada. Dijo que le parece una tecnología que traería muchos beneficios a la ciudad si se llega a implementar.

También dijo que cree que es una alternativa viable para su implementación en la ciudad, que ayudaría bastante en el control de tránsito, en un desarrollo ordenado del flujo vehicular que tenemos en la ciudad, y también beneficiaría directamente a los ciudadanos en el desarrollo de sus actividades diarias en la ciudad.

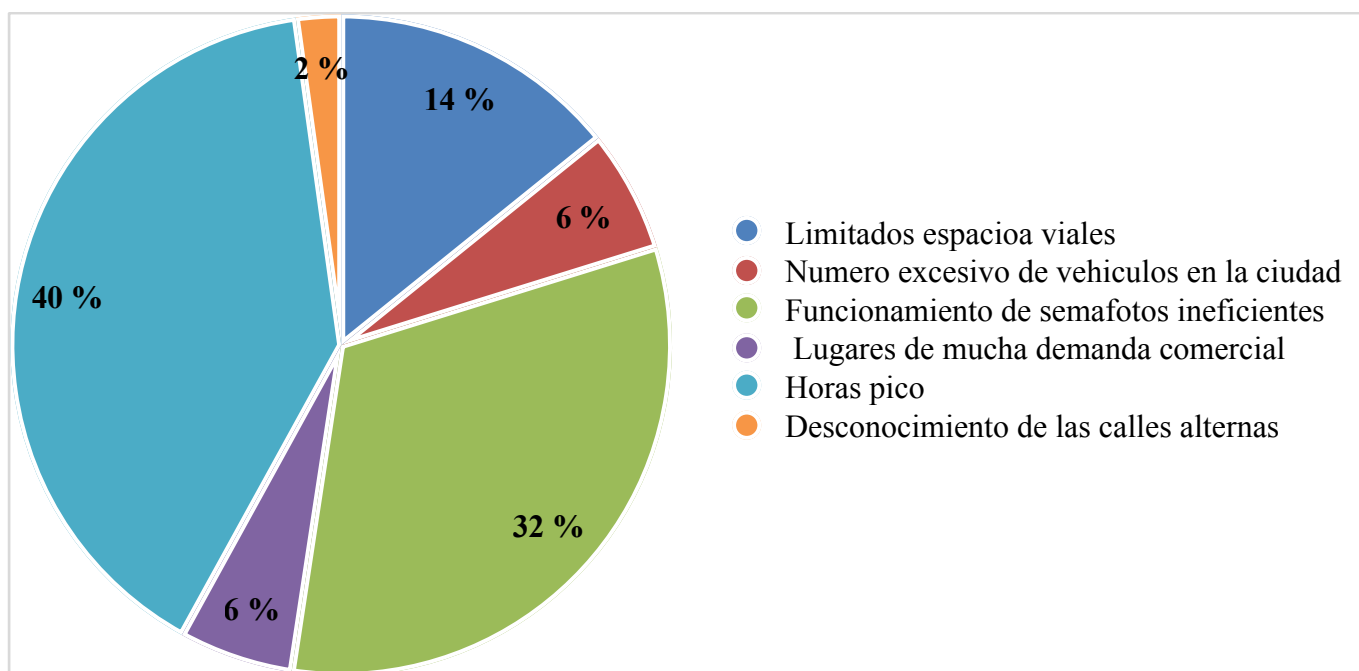
Dimensión 2. Percepción de los conductores sobre el congestionamiento vehicular.

Gráfico N° 1. Existencia situaciones de congestionamiento de vehículos en la ciudad de Concepción.



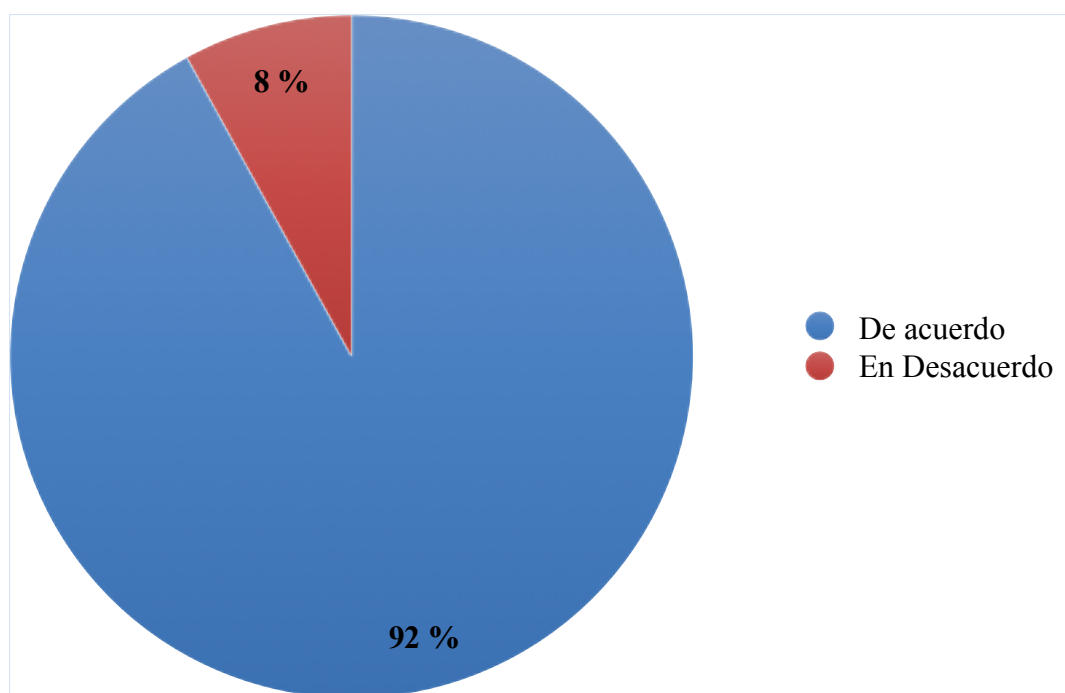
En el gráfico N° 1 se observa que el 67% de los conductores encuestados alegan que casi siempre existe congestionamiento de tránsito, mientras que el 15% dicen que siempre hay, en efecto el 14% sostiene que casi nunca; para el 4% restante de la población no existe nunca congestionamiento de tránsito.

Gráfico N° 2. Creencia de las causas del congestionamiento vehicular.



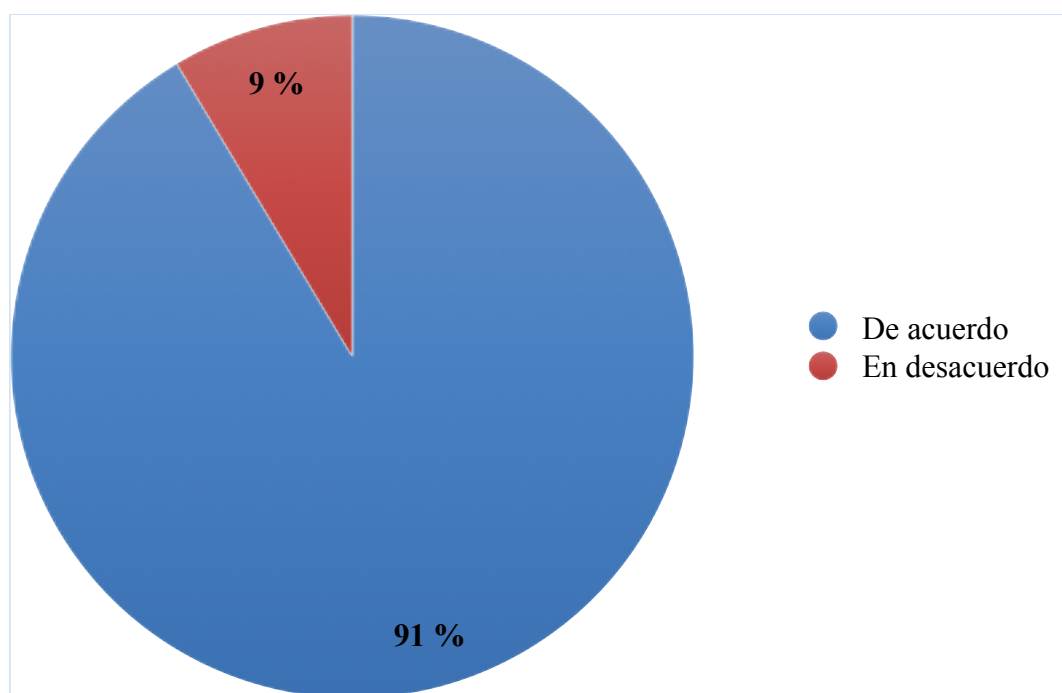
En el gráfico N° 2 se observa que el 40% de los conductores encuestados creen que la causa del congestionamiento es en hora pico, el 32% cree que es por el funcionamiento de semáforos ineficientes, mientras que el 14% cree que es por los limitados espacios viales, el 6% cree que la causa es por el número excesivo de vehículos en la ciudad, otro 6% cree que la causa sería por los lugares de mucha demanda comercial y finalmente el 2% restante cree que es por desconocimiento de las calles alternas.

Gráfico N° 3. La congestión de vehículos representa una amenaza a la seguridad e integridad.



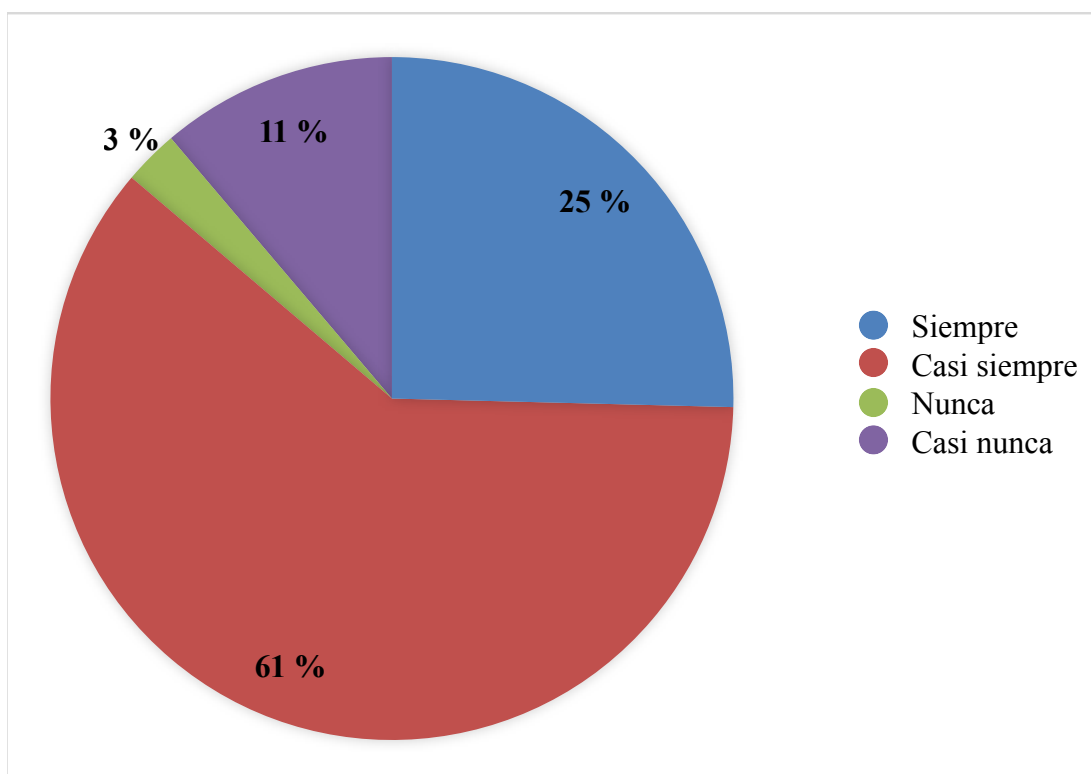
En el gráfico N° 3 se presenta los resultados obtenidos en el cual en un 92% de los encuestados afirman que los conductores están de acuerdo que la congestión vehicular representa una amenaza a la seguridad e integridad en tanto el 8% restante respondieron que se encuentran en desacuerdo.

Gráfico N° 4. La implementación de semaforización inteligente evita la congestión vehicular.



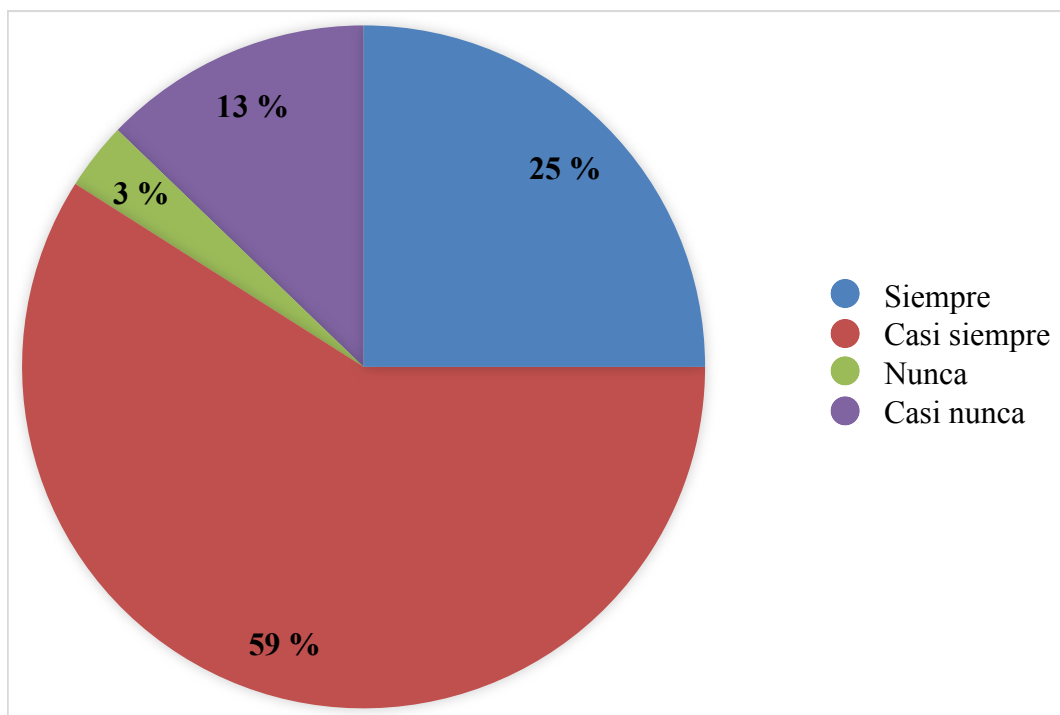
El gráfico N° 4 se presenta los resultados obtenidos de la pregunta: Estas de acuerdo con la afirmación de que la implementación de semaforización inteligente evita la congestión vehicular, a lo que una mayoría de 91% de los conductores encuestados respondieron que están de acuerdo, mientras que el 9% dicen estar en desacuerdo.

Gráfico N° 5. Infracciones a causa de la congestión vehicular.



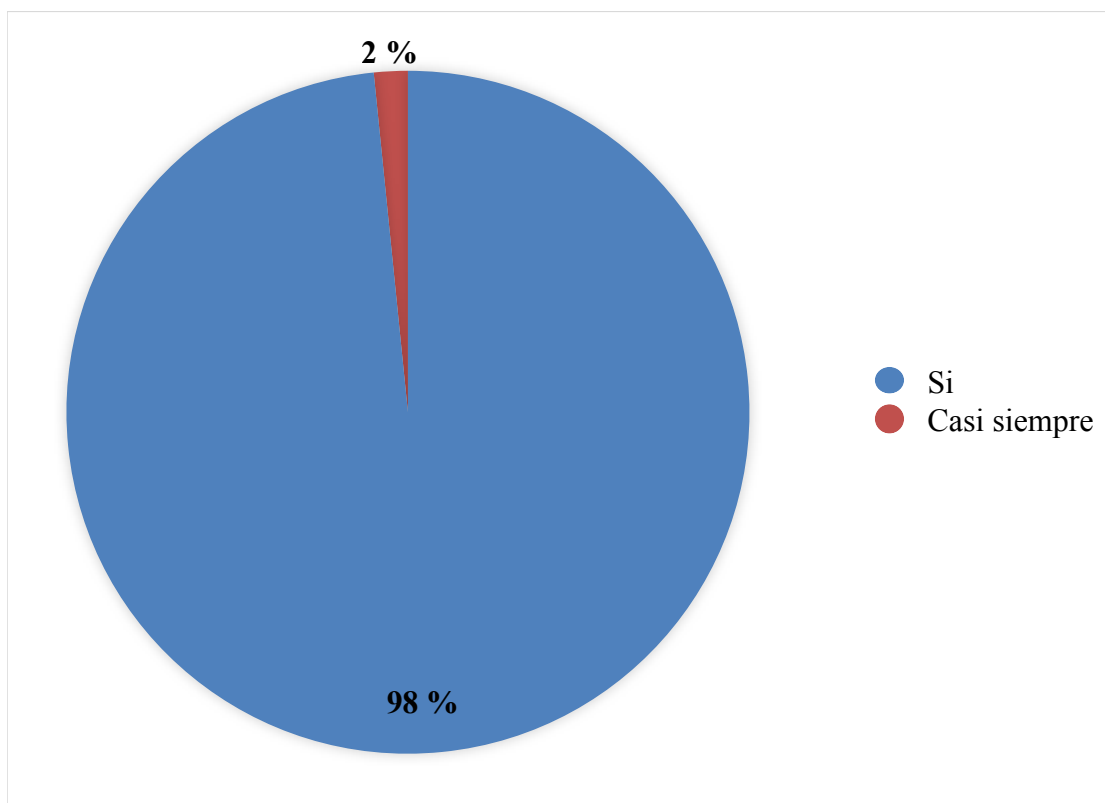
Según los datos que se obtuvieron en el gráfico N° 5 el 61% de los conductores encuestados dicen que suelen cometerse casi siempre infracciones a causa del congestionamiento, mientras que el 25% dicen que siempre suele cometerse infracciones, en efecto el 11% responde que casi nunca se comete infracciones y el 3% dice que nunca se comete infracciones a causa del congestionamiento vehicular.

Gráfico N° 6. Accidentes debido al congestionamiento de vehículos en la ciudad.



En el gráfico N° 6 se observa que el 59% de los encuestados responde que casi siempre ocurren accidentes debido al congestionamiento, el 25% dice que siempre ocurren accidentes en cambio el 13% dice que casi nunca ocurren accidentes y un 3% dice que nunca ocurren accidentes.

Gráfico N° 7. Interés que se implemente semáforos inteligentes en la ciudad de Concepción para evitar problemas viales.



El gráfico N° 7 se presenta el interés de los conductores encuestados de que se implemente semáforos inteligentes en la ciudad a lo que una mayoría de 98% de los conductores encuestados están de acuerdo con la implementación de semáforos inteligentes. Mientras que el 02% están en desacuerdo.

Dimensión 3. Comparar el funcionamiento del semáforo tradicional con el funcionamiento del semáforo inteligente

Para medir las ventajas del funcionamiento del semáforo inteligente contra el semáforo tradicional se desarrolló en el lenguaje de programación JAVA utilizando Netbeans IDE un simulador del funcionamiento de ambos sistemas de acuerdo a los datos y parámetros que se obtuvo de la observación del flujo de tráfico vehicular en una intersección semafórica de la ciudad. Y se procedió a realizar una simulación por software que se detalla en los siguientes pasos.

En la Tabla N° 1 del tráfico se observan las columnas “Av. Pinedo 1”, “Don Bosco” y “Av. Pinedo 2” que representan las arterias viales en el lugar donde se realizó la observación del flujo vehicular. Los datos numéricos en cada columna representan el número de vehículos en cada arteria vial que se generó aleatoriamente teniendo como límite mínimo de un vehículo y máximo de cuarenta vehículos. Se tuvo en cuenta para el mínimo y máximo de vehículos en espera que en 100 metros (una cuadra) entran un máximo de cuarenta vehículos (motos y autos conjuntamente). Los datos (flujo vehicular) recabados del semáforo tradicional son los mismos datos a utilizarse para la comparación con el semáforo inteligente.

A su vez se observan las siguientes columnas:

1. Ciclos: indica cuando todos los semáforos de la intersección dieron luz verde.
2. Fases: representa el ciclo de funcionamiento de un semáforo en luz verde. Según el resultado de la observación realizada cada semáforo de cada arteria tiene un tiempo de 30 seg. de encendido en luz verde.
3. “Av. Pinedo 1”, “Don Bosco” y “Av. Pinedo 2”: representa las arterias viales con semáforos. Los datos numéricos en cada columna representan el número de vehículos en cada arteria vial.
4. Cantidad de vehículos que cruzaron: esta columna representa el número de vehículos que cruzaron el semáforo una vez que este encendió la luz verde.
5. Arteria en verde: representa el semáforo de la arteria que fue accionado. Para nuestro análisis Av. Pinedo 1= 1, Don Bosco = 2 y Av. Pinedo 2= 3.
6. Tiempo de verde: representa los segundos que estuvo encendido el semáforo en luz verde. De acuerdo a la observación realizada en uno de los puntos con semáforos tradicional en la ciudad de Concepción constatamos que el tiempo de luz verde que habilita en verde es de 30 seg. para cada semáforo de las diferentes arterias.

Tabla N° 1. Simulación del funcionamiento del Semáforo Tradicional.

Simulación del funcionamiento del Semáforo Tradicional							
Ciclos	Fases	Flujo de Trafico por arterias			Cantidad de vehículos que cruzaron	Arteria en verde	Tiempo de Verde (segundos)
		Av. Pinedo 1 (Arteria 1)	Don Bosco (Arteria 2)	Av Pinedo 2 (Arteria 3)			
1	1	6	2	1	6	1	30
	2	9	2	12	2	2	30
	3	14	4	13	13	3	30
2	4	20	12	15	20	1	30
	5	5	15	21	15	2	30
	6	14	12	26	26	3	30

3	7	19	13	13	19	1	30
	8	0	27	21	27	2	30
	9	0	0	21	21	3	30
4	10	3	4	2	3	1	30
	11	0	4	4	4	2	30
	12	4	1	12	12	3	30
Simulación del funcionamiento del Semáforo Tradicional							
Ciclos	Fases	Flujo de Trafico por arterias			Cantidad de vehículos que cruzaron	Arteria en verde	Tiempo de Verde (segundos)
		Av. Pinedo 1 (Arteria 1)	Don Bosco (Arteria 2)	Av Pinedo 2 (Arteria 3)			
5	13	6	11	18	6	1	30
	14	5	17	25	17	2	30
	15	5	1	26	26	3	30
6	16	7	4	2	7	1	30
	17	2	4	6	4	2	30
	18	7	3	6	6	3	30
7	19	7	6	6	7	1	30
	20	9	10	6	10	2	30
	21	15	6	11	11	3	30
8	22	15	9	2	15	1	30
	23	4	9	14	9	2	30
	24	5	12	19	19	3	30
9	25	8	14	5	8	1	30
	26	3	16	13	16	2	30
	27	6	9	32	32	3	30
10	28	8	10	6	8	1	30
	29	7	14	6	14	2	30
	30	7	0	9	9	3	30
11	31	8	0	9	8	1	30

	32	0	1	20	1	2	30
	33	3	6	22	22	3	30
12	34	3	14	13	3	1	30
	35	2	14	18	14	2	30
	36	5	1	19	19	3	30
13	37	11	19	13	11	1	30
	38	0	19	13	19	2	30
	39	6	4	27	27	3	30
14	40	11	4	1	11	1	30
	41	0	4	3	4	2	30
	42	6	6	24	24	3	30
15	43	14	20	5	14	1	30
	44	0	20	7	20	2	30
	45	6	4	10	10	3	30
16	46	15	9	0	15	1	30
	47	2	20	0	20	2	30
	48	3	12	18	18	3	30
17	49	7	13	3	7	1	30
	50	4	13	10	13	2	30
	51	4	10	10	10	3	30
18	52	4	12	0	4	1	30
	53	4	17	3	17	2	30
	54	4	2	9	9	3	30
Simulación del funcionamiento del Semáforo Tradicional							
Ciclos	Fases	Flujo de Trafico por arterias			Cantidad de vehículos que cruzaron	Arteria en verde	Tiempo de Verde (segundos)
		Av. Pinedo 1 (Arteria 1)	Don Bosco (Arteria 2)	Av Pinedo 2 (Arteria 3)			
19	55	5	3	19	5	1	30
	56	0	8	20	8	2	30

	57	3	6	25	25	3	30
20	58	3	6	2	3	1	30
	59	3	10	18	10	2	30
	60	4	10	31	31	3	30
Total de Vehiculos					794	Total tiempo (seg)	1.800

Ahora procede a “correr” el simulador con los datos recabados:

1. En la primera fila se observa para “Av. Pinedo 1” 6 vehículos en cola, para “Don Bosco” 2 vehículos en cola y para “Av. Pinedo 2” 1 vehículo. Recordar, y para tener en cuenta el funcionamiento de un semáforo tradicional es de forma secuencial, es decir, ira dando luz verde en un orden preestablecido. Entonces en la columna “Cantidad de vehículos que cruzaron” indica que pasaron 6 vehículos, y en “Arteria en verde” se observa el 1 que representa a la arteria “Av. Pinedo 1”, y en la columna “Tiempo de verde” se observa 30 seg, que es el tiempo fijo de luz verde ya preestablecido.
2. En la segunda fila de la Tabla N° 1 se observa 9, 0 y 11 respectivamente, y en la segunda fila de la Tabla N° 3 se observa para “Av. Pinedo 1” 9 vehículos en cola, para “Don Bosco” 2 vehículos y para “Av. Pinedo 2” 12 vehículos en cola. Como explicamos anteriormente, el funcionamiento del semáforo tradicional es secuencial, y por ende dará luz verde al siguiente semáforo de acuerdo a la secuencia establecida. Por esta razón en la columna “Cantidad de vehículos que cruzaron” indica que pasaron 2 vehículos, y en “Arteria en verde” se observa el 2 que representa a la arteria “Don Bosco”, y en la columna “Tiempo de verde” se observa 30 seg,
3. En la tercera fila de la Tabla N° 1 se observa 5, 4 y 1 respectivamente, y en la tercera fila de la Tabla N° 3 muestra que para “Av. Pinedo 1” 14 vehículos en cola, para “Don Bosco” 4 vehículos y para “Av. Pinedo 2” 13 vehículos en cola. Ahora el semáforo da luz verde para el siguiente semáforo en la secuencia establecida, entonces se observa en la “Cantidad de vehículos que cruzaron”

indica que pasaron 13 vehículos, y en “Arteria en verde” tenemos el 3 que representa a la arteria “Av. Pinedo 2”, y en la columna “Tiempo de verde” se dió 30 segundos,

Esta secuencia de funcionamiento se ira repitiendo de la misma forma en las siguientes filas de la tabla hasta llegar al final de la misma.

En la Tabla N° 2 se observan las siguientes columnas:

1. Ciclos: indica cuando todos los semáforos de la intersección dieron luz verde.
2. Fases: representa el de funcionamiento de un semáforo cuando cambia de una luz a otra. Según el resultado de la observación realizada cada semáforo de cada arteria tiene un tiempo de 30 segundo de encendido en luz verde en el semáforo tradicional.
3. “Av. Pinedo 1”, “Don Bosco” y “Av. Pinedo 2”: representa las arterias viales con semáforos. Los datos numéricos en cada columna representan el número de vehículos en cada arteria vial.
4. Cantidad de vehículos que cruzaron: esta columna representa el número de vehículos que cruzaron el semáforo una vez que este encendió la luz verde.
5. Arteria en verde: representa el semáforo de la arteria que fue accionado, para nuestro análisis Av. Pinedo 1 = 1, Don Bosco = 2 y Av. Pinedo 2 = 3.
6. Tiempo de verde: representa los segundos que estuvo encendido el semáforo en luz verde, se consideró; de acuerdo a la observación realizada; los segundos de encendido en luz verde de 15 seg. para un número de vehículos menor o igual a 15, 20 seg. para un número de vehículos mayor que 15 y menor o igual a 25, y 30 seg. para un número de vehículos mayor que 25.

Para una mayor fluidez del tránsito vehicular y evitar en la mayor medida posible el congestionamiento, el semáforo inteligente debe analizar la cantidad de vehículos en cada arteria, y de acuerdo al resultado dar luz verde a la arteria con mayor cantidad de vehículos, y también dependiendo de la cantidad de vehículos en la arteria que se habilitó la luz verde, el semáforo inteligente dará un tiempo de luz en verde variable. También más arriba se indica que se generó aleatoriamente un número máximo de cuarenta vehículos para cada arteria, y ello no implica que el número máximo de vehículos en cola será cuarenta porque los vehículos pueden ir acumulándose mientras en otra arteria tiene el paso libre.

Sabiendo todo lo mencionado anteriormente procede a “correr” el simulador con los datos recabados.

Tabla N° 2. Simulación del funcionamiento del semáforo inteligente.

Simulación del funcionamiento del semáforo inteligente							
Ciclos	Fases	Flujo de Trafico por arterias			Cantidad de vehículos que cruzaron	Arteria en verde	Tiempo de Verde (segundos)
		Av. Pinedo 1 (Arteria 1)	Don Bosco (Arteria2)	Av Pinedo 2 (Arteria 3)			
1	1	6	2	1	6	1	15
	2	9	2	12	12	3	15
	3	14	6	1	6	2	15
2	4	20	8	16	20	1	20
	5	5	11	22	22	3	20
	6	14	23	5	23	2	20
3	7	19	1	18	19	1	20
	8	0	15	26	26	3	30
	9	0	15	0	15	2	15
4	10	3	4	2	4	2	15
	11	3	0	4	4	3	15
	12	7	1	8	7	1	15
5	13	2	11	26	26	3	30
	14	7	17	7	17	2	15
	15	7	1	8	7	1	15
6	16	2	4	10	10	3	15
	17	4	4	4	4	2	15
	18	9	3	4	9	1	15
7	19	0	6	10	10	3	15
	20	9	10	0	10	2	15
	21	15	6	5	15	1	15

8	22	0	9	7	9	2	15
	23	4	0	19	19	3	20
	24	5	12	5	5	1	15
9	25	3	14	10	14	2	15
	26	6	2	18	18	3	20
	27	9	11	19	9	1	20
10	28	2	12	25	25	3	20
	29	9	16	0	16	2	20
	30	9	0	3	9	1	15
Simulación del funcionamiento del semáforo inteligente							
Ciclos	Fases	Flujo de Trafico por arterias			Cantidad de vehículos que cruzaron	Arteria en verde	Tiempo de Verde (segundos)
		Av. Pinedo 1 (Arteria 1)	Don Bosco (Arteria 2)	Av Pinedo 2 (Arteria 3)			
11	31	1	0	12	12	3	15
	32	1	1	11	1	2	15
	33	4	6	13	4	1	15
12	34	0	14	26	26	3	30
	35	2	14	5	14	2	15
	36	5	1	6	5	1	15
13	37	6	19	19	19	2	20
	38	6	0	19	19	3	20
	39	12	4	14	12	1	15
14	40	5	4	15	15	3	15
	41	5	4	2	5	1	15
	42	6	10	23	10	2	20
15	43	14	14	28	28	3	30
	44	14	14	2	14	2	15
	45	20	4	5	20	1	20
	46	9	9	5	9	2	15

16	47	11	11	5	11	1	20
	48	1	23	23	23	3	20
17	49	5	24	3	24	2	20
	50	9	0	10	10	3	15
	51	9	10	0	9	1	15
18	52	0	12	0	12	2	15
	53	4	5	3	4	1	15
	54	0	7	9	9	3	15
19	55	1	8	19	19	3	20
	56	1	13	1	13	2	15
	57	4	6	6	4	1	15
20	58	0	6	8	8	3	15
	59	3	10	16	10	2	15
	60	4	10	29	4	1	15
Total de Vehículos					769	Total tiempo (seg)	1045

1. En la primera fila se observa para “Av. Pinedo 1” 6 vehículos en cola, para “Don Bosco” 2 vehículos en cola y para “Av. Pinedo 2” 1 vehículo, en la columna “Cantidad de vehículos que cruzaron” indica que la mayor cantidad de vehículos es 6 y por consecuente en “Arteria en verde” se observa el 2 que representa a la arteria “Av. Pinedo 1”, y en la columna “Tiempo de verde” se da 15 segundos, que dio ese tiempo porque la cantidad de vehículos mayor de las tres arterias fue menor a 20.
2. En la segunda fila de la Tabla N° 1 se observa 9, 0 y 11 respectivamente, y en la segunda fila de la Tabla N° 2 se observa para Av. Pinedo 1, 9 vehículos en cola, para “Don Bosco” 2 vehículos y para Av. Pinedo 2, 12 vehículos en cola. Para entender el funcionamiento y los números de la segunda fila de la Tabla N° 2 hay

que recordar que en el punto 1 el semáforo habilitó luz verde para la arteria Av. Pinedo 1 que tenía 6 vehículos en cola y estos cruzaron todos, pero en las otras arterias quedaron vehículos en espera, estos vehículos se suman al flujo inicial de la Tabla 1 obteniendo así como resultado los datos de la Tabla 2 en la arteria Don Bosco 2 vehículos (resultado de sumar 2 vehículos que quedo en espera más 0 vehículos del flujo inicial de la tabla N° 1) y en la arteria “Av. Pinedo 2” se puede observar 12 vehículos (resultado de sumar 1 vehículo que quedo en espera más 11 vehículos del flujo inicial de la tabla N° 1). Ahora se procede a procesar los datos, como en el primer punto la luz verde se habilitó para la arteria “Av. Pinedo 1” este ya no entra en comparación con las demás arterias, así se procede a comparar el número de vehículos que tiene la arteria Don Bosco y Av. Pinedo 2, y se observa como resultado que el mayor número de vehículos tiene la arteria Av. Pinedo 2 por consecuente la luz verde se habilita para esta arteria. Así la columna SI tiene 12 (de 12 vehículos que cruzaron), la columna SI Accionado tiene 3 (arteria Av. Pinedo 2 = 3) y la columna Tiempo de verde tiene 15 (se habilito 15 segundos porque el flujo de tráfico es menor a 20 vehículos).

3. En la tercera fila de la Tabla N° se observa 5, 4 y 1 respectivamente, y en la tercera fila de la Tabla N° 2 se observa para Av. Pinedo 1, 14 vehículos en cola, para “Don Bosco” 6 vehículos y para Av. Pinedo 1 vehículos en cola. Como en el anterior punto se volvieron a sumar los vehículos en cola que quedaron en espera, así para Av. Pinedo 1 quedo en 14 vehículos (de la suma de 9 vehículos que quedaron en espera más 5 vehículos del nuevo flujo), Don Bosco quedo en 6 vehículos (de la suma de 2 vehículos que quedaron en espera más 4 vehículos del nuevo flujo), y Av. Pinedo 2 quedo en 1 vehículo porque en esta arteria dio la luz verde y cruzaron todos los vehículos. En este punto el semáforo ya habilitó luz verde para las arterias Av. Pinedo 1 y Av. Pinedo 2, así que solo falta la arteria Don Bosco, por esta razón ya no se compara el flujo de tráfico y directamente se habilita la luz verde para esta arteria, para evitar una excesiva demora de los conductores que se encuentran aguardando en la arteria. Entonces se obtiene que para la columna SI tiene 6 (de 6 vehículos que cruzaron), la columna SI Accionado tiene 2 (arteria Don Bosco = 2) y la columna Tiempo de verde tiene

15 (se habilita 15 seg porque el flujo de tráfico es menor a 20 vehículos).

Para las siguientes filas se vuelve a repetir los mismos puntos anteriores teniendo en cuenta cada condición del funcionamiento del algoritmo del semáforo inteligente.

Así con los pasos anteriores y simulando una hora de funcionamiento del semáforo inteligente y el semáforo tradicional, se obtiene que la cantidad de vehículos que pasaron por el semáforo inteligente es un total de 2817 en un tiempo de 45min. 25seg. y por el semáforo tradicional pasaron un total de 2140 vehículos en 1 hora de tiempo. Aumentando el tiempo de funcionamiento de la simulación se puede observar los resultados resumida en la siguiente tabla:

Sistema Semáforo Inteligente		Sistema Semáforo Tradicional	
Flujo de trafico	Tiempo (hs)	Flujo de Trafico	Tiempo (hs)
9844	03:26:10	9843	06:00:00
20293	06:57:20	20287	12:00:00
40663	13:59:55	40656	24:00:00

Se puede observar que, para el mismo flujo de tráfico vehicular, hacer circular dichos vehículos, al semáforo inteligente le toma solo la mitad de tiempo comparado con el semáforo tradicional, lo que significa que con semáforo inteligente hay menos congestionamiento vehicular y por ende un ahorro de tiempo de espera para los conductores.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

Conclusiones

En este trabajo de investigación se obtuvo las siguientes conclusiones:

Sobre la situación del funcionamiento de semáforos en la ciudad, se encontró que en la ciudad se siguen utilizando semáforos de funcionamiento tradicional en sólo 4 intersecciones de la ciudad. La tecnología utilizada es la micro electrónica con tabla de horas del sistema de control de tránsito. En la ciudad se cuenta con los semáforos con contralor de 4 movimientos, 12 salidas vehiculares y con comunicación RS 232. Cada semáforo cuenta con una Tabla de horas programable según la necesidad del usuario, cuenta con actualización de hora vía GPS para la sincronización efectiva entre cada semáforo.

La percepción de los conductores sobre el tránsito vehicular en la ciudad de Concepción, dio como resultado que el 67% de los conductores alegan que casi siempre existe congestionamiento de tránsito en la ciudad, 32% cree que es por el funcionamiento de semáforos ineficientes, y con relación a que la implementación de semáforos inteligente colabore para la disminución de la congestión en el tránsito, 91% de los conductores encuestados respondieron que están de acuerdo.

Las diferencias observadas entre el funcionamiento del semáforo tradicional y el funcionamiento del semáforo inteligente, utilizando algoritmos y herramientas del Iot.” a través de una simulación por software con la situación propia del tránsito de la ciudad de Concepción, arrojó como resultado que con el sistema tradicional, a un flujo de 9.843 vehículos les tomó 6 horas de tiempo cruzar por la intersección, mientras que al semáforo inteligente a un flujo de 9.844 vehículos solo invirtió 03 horas 26 minutos para realizar la tarea. Esto demuestra que se obtiene un ahorro significativo del tiempo con el uso del

semáforo inteligente respecto al tradicional, y también una mayor fluidez de desplazamiento de vehículos.

Por consecuente, mediante la investigación se ha logrado analizar los resultados obtenidos de la comparación del funcionamiento de los semáforos utilizando algoritmos y herramientas IoT, concluyendo así, que los semáforos utilizados en la ciudad de Concepción son prácticamente obsoletos para el desarrollo fluido del tránsito vehicular. Que los conductores de la ciudad perciben la existencia de congestionamiento en las calles y consideran que la utilización de semáforos inteligentes contribuiría a la disminución del congestionamiento en los cruces semaforicos de la ciudad.

Y que por medio de la simulación por software se obtuvo resultados muy favorables en ahorro de tiempo y una mayor fluidez de desplazamiento de vehículos.con el semáforo inteligente.

Recomendaciones

A la Municipalidad de Concepción:

La implementación de semáforos inteligentes para disminuir el congestionamiento y evitar así todos los problemas que ello acarrea consigo, y ser beneficiados con las bondades que nos ofrece un sistema innovador y moderno en la ciudad.

A los conductores de la ciudad de Concepción:

Respetar siempre las reglas de tránsito para una conducción tranquila por las calles de la ciudad, que nos permita a cada uno de nosotros cumplir con nuestras labores diarias, colaborar y apoyar el funcionamiento y la implantación de nuevas tecnologías para el ordenamiento del tránsito en la ciudad.

CAPÍTULO VI. PROPUESTA

Planteamiento del Problema

Hoy en día, todas las iniciativas para el desarrollo Smart de una ciudad se enfocan en las instituciones públicas por el impacto que tienen en la sociedad en general. Una ciudad inteligente es una iniciativa para administrar la ciudad, donde se realizan acciones y servicios, en función de la información tecnológica. La adaptación de las personas al Internet se ha vuelto indispensable para un desarrollo sostenible, en esta escena el (IoT) es un concepto que se refiere a la interconexión digital de objetos cotidianos con la Internet; en la actualidad las principales ciudades del mundo han identificado el impacto que genera la aplicación del Internet of Thing,

Según la Revista Cap (2022), el 18 de agosto, Interactive S.A, organiza un workshop, con el apoyo de Teisa, donde se lanzó oficialmente la red de conectividad para dispositivos IoT SIGFOX. En el evento se dio a conocer la nueva tecnología que potenciará los negocios de las empresas, permitiéndoles alcanzar metas más altas.

SIGFOX es una red mundial con base en Francia, que permite que cualquier dispositivo compatible con la tecnología de transmisión SIGFOX pueda conectarse y enviar información que el usuario del dispositivo pueda utilizar.

En el workshop se dio a conocer esta nueva tecnología, que está al alcance de todos los sectores de la Economía Paraguaya; primario (Agrícola, Ganadero), secundario (Industrias) y terciario (Servicios) tanto público como privado.

En el evento se conversó con Fernando Ramírez, quien encabeza el equipo de SIGFOX en Paraguay, y Jorge De León Llanes, coordinador comercial de SIGFOX en

Interactive S.A., empresa que se encargará de la distribución y venta de las soluciones IOT.

Jorge de Leon menciona que la tecnología IoT es de uso masivo en Europa, Asia y Norte América como base de gestión para la implementación de otras tecnologías como el Big Data, Machine Learning, Robótica. No obstante, en el Mercosur, así como en Paraguay, el proceso de expansión está en su fase inicial y en los próximos años ira acercándose a su nivel de madurez.

Una de las opciones que más contribuye para agilizar la movilidad en la ciudad de Concepción se plantea a partir de la implementación de un sistema semaforización inteligente, para esta implementación

Objetivos de la Propuesta

Objetivo General

Diseñar un prototipo de Semáforo inteligente utilizando algoritmos y Herramientas de Internet of Things.

Objetivos Específicos

Identificar los componentes necesarios para la implementación de un sistema de semaforización inteligente

Asegurar que la arquitectura de IoT cumpla con los aspectos de calidad necesarios para su funcionamiento.

Proponer el algoritmo para el prototipo que se ajuste a los requerimientos específicos identificados para el semáforo inteligente.

Justificación

Según la percepción de los conductores encuestados, el 67 % denota que hay situación de congestionamiento vehicular en las principales calles de la ciudad de Concepción, Y también de los resultados obtenidos en la comparativa del funcionamiento del semáforo tradicional y el funcionamiento del semáforo inteligente se puede ver que con un sistema de semaforización inteligente se reduce el tiempo de espera de los vehículos y por ende favorece al descongestionamiento de las calles

Es por ello que esta investigación se aboca a desarrollar un prototipo de semaforización inteligente con el fin de ofrecer una alternativa de solución a la problemática del tránsito vehicular en la ciudad de Concepción.

La investigación se basa en un estudio descriptivo atendiendo que se enfoca en resolver una de las problemáticas del tránsito vehicular, particularmente en las intercepciones semafóricas del centro de la ciudad, donde se va a describir los requerimientos para el desarrollo y el funcionamiento del prototipo de semáforos inteligentes.

Alcances

A fin de lograr los objetivos establecidos se presentará el diseño de un prototipo de semáforos inteligentes para la ciudad de Concepción. Todos los recursos y herramientas serán proporcionados para el desarrollo de este proyecto. El prototipo desarrollado está preparado para ser utilizado en todos los cruces semafóricos existentes en la ciudad de Concepción y también para futuras instalaciones.

Limitaciones

Esta propuesta presenta las siguientes limitaciones:

1. El prototipo propuesto es un sistema embebido.
2. Cada intersección semafórica funcionara como un nodo independiente, es decir, no tendrá conexión con el siguiente cruce semafórico por ende la toma de decisión será independiente entre un nodo y otro.
3. No se contempla una base de datos para el almacenamiento de datos del flujo vehicular.
4. Para realizar modificaciones de los parámetros debe realizarse localmente por cada nodo.
5. No posee un sistema de gestión (software).

Recursos necesarios

Para la propuesta se tuvo como referencia en el borrador de la norma IEEE P2413 - Estándar para un marco arquitectónico para la Internet de las cosas (IoT), donde se identifican las diferentes capas de aplicaciones verticales, sus relaciones. mediante las tecnologías informáticas de vanguardia, de igual manera con la investigación realizada en varias de las ciudades donde ya existe la implementación de un sistema inteligente de transporte se definen los requerimientos funcionales no funcionales que debe tener una arquitectura basada en IoT para la semaforización en Concepción.

- a. Sensor: El principal componente para el funcionamiento inteligente del semáforo será el Sensor. Más allá del tipo de sensor utilizado para detectar la presencia de los vehículos el prototipo desarrollado se encuentra adecuado para cualquier tipo de sensor a ser utilizado. Se profundizo más en la utilización del detector de espiras inductivas y el magnetómetro.

1.

Algoritmo Propuesto

```
//Declarar variables para entrada y salida de datos
```

```
//Arteria 1
```

```
int pinA1Rojo = 7;
```

```
//int pinA1Ambar = 1;
```

```
int pinA1Verde = 6;
```

```
int pinBtnA11 = 8;
```

```
//Arteria 2
```

```
int pinA2Rojo = 2;
```

```
//int pinA2Ambar = 1;
```

```
int pinA2Verde = 3;
```

```
int pinBtnA21 = 4;
```

```
int pinBtnA22 = 5;
```

```
//Arteria 3
```

```
int pinA3Rojo = 9;
```

```
//int pinA3Ambar = 1;
```

```
int pinA3Verde = 10;
```

Referencias

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, 487-499.
- Han, J., & Kamber, M. (2001). Data mining: concepts and techniques (2nd ed.). Morgan Kaufmann.
- Bhattacharya, S., & Bhattacharya, S. (2014). Customer segmentation and clustering techniques: A review. *International Journal of Business Information Systems*, 17(4), 412-438.
- Jain, S., & Luthra, S. (2016). A comprehensive review on customer purchase behavior analysis in retail sector using data mining techniques. *Procedia Computer Science*, 78, 507-512.
- Wu, D. D., & Olson, D. L. (2010). Business data mining—A machine learning perspective. *Business Intelligence Journal*, 15(1), 15-24.
- Kim, H. W., Lee, H. J., & Lee, I. (2017). Clustering-based customer segmentation approach for personalized recommendation systems. *Journal of Business Research*, 74, 1-10.
- Chen, K., Cheng, K. T., & Chen, T. C. (2013). Customer segmentation using modified RFM model: A case study on the wireless communication industry. *Expert Systems*

with Applications, 40(10), 3976-3985.

Xie, W., & Zhang, Y. (2016). Customer segmentation and target marketing using clustering methods. *Expert Systems with Applications*, 58, 265-273.

Quintero, J y Prieto, Lina (2015). “Sistemas *inteligentes de transporte y nuevas tecnologías en el control y administración de transporte*” Obtenido en diciembre del 2021 de: <http://puente.upbbga.edu.co/index.php/revistapuentes/article/download/220/16>

R. H. Weber, (2010). "Internet of Things - New Security and Privacy Challenges". Computer Law & Security Review 26: 23-30.

Salazar, J. y Silvestre, S. (2019). El mundo Internet of Things (IoT). Erasmus +. Primera edición. Obtenido de https://upcommons.upc.edu/bitstream/handle/2117/185120/LM01_R_ES.pdf;jse

Significados.com. (2023). Obtenido en <https://www.significados.com/analisis/>

Tarqui Vallejo, V. (2015). Modelo de Simulación de la dinámica del Tránsito Vehicular bajo el enfoque microscópico. La Paz, Bolivia

Thomson, Ian y Bull, A. (2001), La congestión de tránsito urbano: causas y consecuencias económicas y sociales, serie Recursos naturales e infraestructura, N° 25 (LC/L.1560-P), Santiago de Chile, Comisión Económica para América Latina y el Caribe (CEPAL). Publicación de las Naciones Unidas, N° de venta: S.01.II.G.105.

Ucha, F. (enero, 2015). Definición de Tránsito. Definición ABC. Obtenido de <https://www.definicionabc.com/social/transito.php>

Valle Pazmiño, S. (2019). Implementación de un prototipo IOT con semáforos inteligentes a través de una infraestructura de computación en la niebla. Obtenido de: <https://repositorioslatinoamericanos.uchile.cl/handle/2250/2794653>

ANEXOS

