

Ciencia de datos y geointeligencia en la mejora de la experiencia del cliente

Grisel A. Porras, Alberto L. Munoz

Instituto Tecnológico de Estudios Superiores de Monterrey,
Mexico

`griselporras89@gmail.com`, `amunoz@tec.mx`

Resumen Por medio de este artículo de investigación, se realizó una propuesta para la diversificación de servicios en gasolineras con la tecnología Swit en N.L, México, teniendo como eje central la aplicación de técnicas avanzadas de análisis de datos y geointeligencia. A nivel local, se identificaron patrones de consumo que proporcionaron información relevante para la toma de decisiones como la implementación de nuevos servicios dependiendo de las características demográficas y opiniones sobre el servicio.

Palabras clave: Geointeligencia, análisis geoespacial, ciencia de datos, geomarketing, gasolineras, diversificación, experiencia del cliente.

Leveraging Data Science and Geointelligence to Enhance Customer Experience

Resumen This research article presents a proposal for the diversification of services at gas stations using Swit technology in Nuevo León, Mexico, with a focus on the application of advanced data analysis and geointelligence techniques. At the local level, consumption patterns were identified, providing valuable insights for decision-making—such as the implementation of new services based on demographic characteristics and customer feedback.

Keywords: Geointelligence, geospatial analysis, data science, geomarketing, gas stations, diversification, customer experience.

1 Introducción

Hoy en día, la estrategia de diversificación juega un papel importante para el desarrollo financiero de una empresa. Con el avance de la tecnología y la necesidad de seguir a la vanguardia sobre la satisfacción del cliente, ha crecido la cantidad de empresas que sobrepasan los límites de las regiones, industrias y mercados, expandiendo así sus negocios.

Si bien es cierto que dicha estrategia juega un papel importante en el desarrollo prolífero de una empresa, es necesario mencionar que involucra incertidumbre, lo que genera que sea impredecible ante ciertas condiciones. Lo que se plantea es desarrollar conocimientos para contestar la pregunta inminente sobre la diversificación en gasolineras.

Esta estrategia puede realizarse cuando exista la naturaleza de implementar una mejor experiencia al mercado actual y que funcione como un diferenciador para los competidores. Entender estas fuentes de ventaja competitiva de las empresas ha derivado en la dirección del futuro económico, pues aprovechar estos activos permite responder a las oportunidades del entorno, evitando las debilidades y planteando la toma de decisiones consciente [7].

Las estaciones de servicio no han tenido grandes innovaciones en el mercado desde hace décadas, sin embargo, actualmente, el mercado se ha estado intentando abrir, de forma que ahora no solo pueden ofrecer un servicio, sino muchos. La diversificación de hoy en día puede incluir el mantenimiento del carro, servicios de lavado y tiendas de conveniencia.

Según información de PETROIntelligence, hasta junio 2023, se presentaron más de 13,400 estaciones de servicio en México, de las cuales, alrededor de 6,000 pertenecen a otra compañía diferente de Pemex, entre las más grandes se encuentran Oxxogas, Petro Seven y Corpogas. En comparación con años anteriores, la cantidad de estaciones incrementó aproximadamente 3 % cada una. En el caso de Nuevo León, se otorgaron concesiones por cada 88 km^2 de superficie. Además, en Nuevo León, se estima que la participación de marcas distintas a Pemex se ubica en un 72 % [10].

Conociendo un poco este contexto, abre la posibilidad de un nuevo mercado, en donde se aprovecha el área de oportunidad en la que los usuarios no reciben un servicio óptimo. De esta manera, se pueden capitalizar los tiempos muertos de espera durante el tiempo de abastecimiento.

Swit es un servicio que ha tomado forma desde hace unos años. Gracias a él, se han incorporado nuevas formas de adquirir y pagar servicios desde la gasolinera aprovechando los tiempos de espera, representando una estrategia de diversificación que mejora la experiencia del cliente.

Para mejorar estos esfuerzos, la ciencia de datos juega un papel fundamental para la toma de decisiones informadas. Por medio del análisis geoespacial, se busca implementar este servicio especializado, identificando áreas de enfoque prioritario para crear nuevas estrategias para la empresa. De este modo, se podrá implementar el servicio en una estación específica solo si se considera beneficioso para la zona en donde se evalúa.

Entran aquí factores importantes que se pueden o no relacionar entre sí y que subyacen dentro de la dinámica del sistema. Por ejemplo, la zona del país donde se encuentran, las restricciones de movilidad, la cercanía con otros puntos de interés así como el costo que cuesta el llenado. Estos son factores que, si son cuidadosamente considerados, pueden maximizar su éxito.

En relación con lo anterior, la presente investigación plantea analizar las características del cliente para obtener información más detallada de la industria

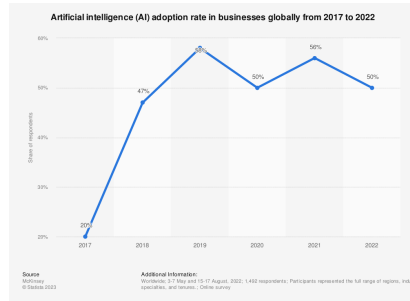


Fig. 1. Tasa de adopción de Inteligencia Artificial en empresas del 2017 al 2022.

y así tomar decisiones. Se propone realizar el desarrollo del análisis a partir del posicionamiento de diferentes variables que determinan en gran medida la rentabilidad del servicio que ofrece Swit.

2 Hipótesis

El conjunto de series de tiempo, clasificación, NLP y clustering de datos, sustentado por la información derivada de fuentes públicas para el análisis geoespacial, permitirá identificar tendencias de consumo en la industria energética, lo que se traducirá en una predicción más acertada de las necesidades del usuario y, por ende, en un incremento significativo en la experiencia del cliente dentro de las estaciones de servicio en evolución.

3 Estado del arte

Históricamente, la ciencia de datos en negocios siempre ha sido utilizada de forma empírica. Actualmente, según información de McKinsey & Company (Figura 1) comentan que la Inteligencia Artificial ha sido bien recibida como una herramienta que funciona como valor agregado en las empresas. Como bien se puede ver en la gráfica, desde el 2017 el valor de la tasa de adopción en los negocios se ha duplicado hasta el 2022. [5] Igualmente, aplicar estas técnicas permite generar estrategias que traen consigo mayor eficiencia y experiencias de cliente más positivas [1]. Por ejemplo, como se puede ver en la siguiente gráfica, las ganancias relacionadas al uso de inteligencia artificial aplicadas en negocios de todo el mundo ha traído consigo resultados satisfactorios para las mismas (véase Figura 2).

Esta aplicación de nuevas estrategias con Inteligencia Artificial son beneficiosas en más de un departamento de una empresa. Como se observa en la siguiente gráfica (Figura 3) que se deriva del estudio realizado por McKinsey, se puede ver que el máximo incremento de las ganancias con el uso de programas con Inteligencia Artificial fue en el departamento de Marketing y ventas, así como en el Desarrollo de productos o servicios. [5]

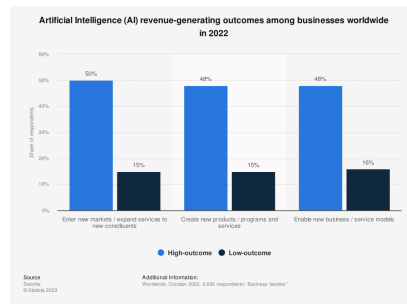


Fig. 2. Resultados de aplicar Inteligencia Artificial en empresas al rededor del mundo durante el 2022.

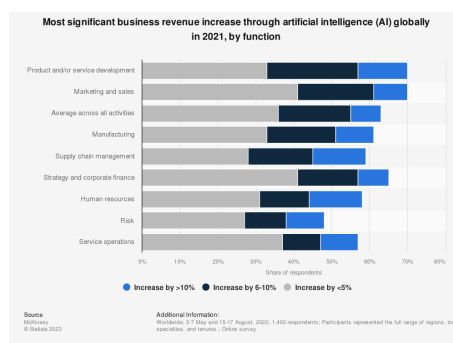


Fig. 3. Incremento de ganancias por el uso de Inteligencia Artificial en Negocios durante el 2021.

Aunque se espera que el mercado de Inteligencia Artificial tenga un crecimiento en los próximos años, aún queda un largo camino por recorrer, ya que muchas empresas todavía no ponen en práctica estos conocimientos. En la encuesta realizada por Forbes sobre el *'Deployment status of data science and machine learning in organizations worldwide as of 2019, by function'* se comenta que solo el 26.5% dentro de las empresas encuestadas usan técnicas de ciencia de datos y machine learning. Por otro lado, solo el 32.5% de las compañías encuestadas las utiliza en el departamento de Business intelligence competency center. [11]

En cuanto a las gasolineras, existe poca documentación relacionada con la diversificación de servicios, ya que se trata de un caso muy específico. No obstante, el artículo *Constructing marketing decision support systems using data diffusion technology: A case study of gas station diversification* [8] expone los esfuerzos realizados en torno a esta transformación.

El estudio se centra en una empresa de Taiwán dedicada a la distribución de gasolina, la cual, debido a cambios en el mercado energético, la competencia creciente y la presión por la sostenibilidad, se ha visto obligada a diversificar sus

operaciones. Ante la disminución de márgenes en la venta de combustibles, la empresa ha tenido que explorar nuevas líneas de negocio dentro de sus estaciones de servicio, como tiendas de conveniencia, servicios de mantenimiento vehicular, lavado de autos, e incluso la instalación de cargadores eléctricos. Dado que se dispone de información limitada sobre el tema, el objetivo es evaluar la viabilidad de incorporar un nuevo servicio en alguna de sus estaciones. Para ello, se consideran distintas variables como:

- El precio es competitivo
- La cantidad de equipo es adecuado
- Los servicios son personalizados para la satisfacción del cliente
- El staff sirve a los consumidores de forma rápida
- El staff sirve a los consumidores con entusiasmo

Además se miden otras variables como la lejanía a una carretera, al pueblo más cercano y a la ciudad más cercana. Apuntando a cada gasolinera con estas características para clasificar en clases A, B y C las gasolineras. Como los datos son pocos, la propuesta es utilizar una técnica de mega-trend-diffusion (MTD) para construir la toma de decisiones a partir de una red bayesiana. Según los resultados la precisión del modelo aumentó al utilizar este enfoque.

De igual forma, existen otros enfoques relacionados con geointeligencia. Dentro de ellos se encuentran el artículo *'Site selection for small gas stations using GIS'* [2] donde se realizó un modelo para posicionar de forma estratégica la ubicación óptima de una estación de combustible con GIS en Irán. Se compararon diferentes áreas, evaluando factores en cada una de ellas como:

- Distancias a la avenida principal
- Distancias a carreteras
- Distancias a estacionamientos
- Distancias a la ciudad más cercana
- Distancias a otras gasolineras
- Distancias a escuelas
- Distancia a hospitales

Cada factor, contiene un peso asociado en el que se busca atribuir a cada lugar posible una calificación. En este enfoque, se utilizaron diferentes técnicas y se probó cada una. Se encontró que los métodos como el Fuzzy analytical hierarchy process (FAHP) y analytical hierarchy process (ANP) eran los más flexibles.

De igual forma, también se ha realizado una investigación sobre estrategias de geomarketing aplicadas a un tipo de zona específica. Los artículos *'Creating a geodemographic classification model within geo-marketing: the case of Eskişehir province'* [4] y *'Building a Geo-Demographic Segmentation Model: the Case of Hanoi City, Vietnam'* [6] tienen un enfoque parecido. En ambos casos, se estudia una subdivisión de dos países, en el primer caso, en la ciudad de Eskişehir, Turkia y Hanoi, Vietnam.

En relación con lo anterior, se utilizaron técnicas como K-means Clustering, Principal Component Analysis (PCA), Hierarchical Clustering Analysis que dio como resultado la segmentación geodemográfica de los barrios de estas dos ciudades, para encontrar características de personas compartidas. Características como la edad, nivel de educación redujeron el costo y el tiempo en el proceso de ingresar o expandir los negocios en una nueva área.

Por último, se ha realizado una investigación sobre el uso de la información recabada por dispositivos IOT en el artículo '*Exploring the application of IoT in the service station business*' [9], donde se hizo uso de series de tiempo para entender el perfil del cliente. Por ejemplo, ver la relación que existía con el precio de la gasolina en la llegada de clientes, las horas pico y el tráfico en general. En este estudio se buscaba segmentar al mercado para poder ofrecer servicios más especializados que mejoraran la satisfacción hacia la estación, como la optimización del horario de empleados.

4 Descripción de las metodologías

La metodología propuesta para esta investigación combina diferentes direcciones cualitativas y cuantitativas. Gracias a ellas, será posible procesar los datos para analizar y predecir las tendencias de la industria energética en estaciones de servicio.

Tomando esto en consideración, se utilizará la metodología Cross Industry Standard Process for Data Mining, por sus siglas en inglés CRISP-DM, la cual, combina todas las etapas aplicables en un proyecto. El modelo contiene seis etapas dependientes entre sí, donde las secuencias no es completamente estricta [12].

1. Comprensión del negocio: Se establece el objetivo del proyecto para identificar las tendencias de consumo en la industria energética para mejorar la satisfacción del cliente. Asimismo, se realiza una evaluación para comprender el contexto de las necesidades actuales del cliente, así como los desafíos que implican proponer diferentes soluciones para las oportunidades de mejora.
2. Comprensión de los Datos: Primero, se realizará una recopilación de los datos por medio de fuentes públicas. Para fines de este proyecto se estará utilizando información del Instituto Nacional Estadística y Geografía (INEGI) para recopilar datos sobre gasolineras e información geoespacial que tengan que ver con el comportamiento del usuario. Posteriormente, se estarán evaluando los datos crudos para comprender un poco mejor las tendencias.
3. Preparación de los datos: El siguiente paso es la limpieza de los datos, incluyendo los faltantes, outliers y otros errores que pudieran existir. En adición a esto, se realizarán las transformaciones necesarias.
4. Modelos: Se estarán utilizando diferentes técnicas de Ciencia de Datos mencionadas anteriormente.

5. Evaluación: Evaluar el rendimiento de los modelos a partir de diferentes métricas. Ajustar hiperparámetros y comparar diferentes modelos para evitar underfitting y overfitting.
6. Despliegue: Implementar los modelos realizados, actualizarlos y monitorearlos para garantizar su buen funcionamiento.

5 Diseño de experimentos

Los experimentos se estructurarán en tres fases bien definidas durante el proyecto para lograr los objetivos planteados al inicio del documento. En la primera fase, los datos que se estarán analizando provienen de diferentes fuentes como el Directorio Estadístico Nacional de Unidades Económicas (DENUE) y el Censo de Población y Vivienda 2020 (SCINCE 2020) provenientes del INEGI con el objetivo de recopilar datos sociodemográficos para llevar a cabo una clasificación exhaustiva de las zonas geográficas. Esto será útil para segmentar al mercado de forma más precisa y adaptar las estrategias según las características de cada área. Las variables que se estarán utilizando se pueden encontrar en la siguiente tabla (*Tabla 1*)

La selección de indicadores sobre las características sociodemográficas de la población y las viviendas del país vienen desagregados hasta el nivel de manzanas. Los indicadores se dividen en diferentes categorías, sin embargo, solo se trabajaron con los temas relacionados a Educación, Características Económicas, Hogares Censales y Vivienda. Esto con el objetivo de entender a mayor profundidad sobre la situación actual de cada vivienda, así como el estilo de vida que lleva cada zona en específico. Para facilitar el manejo y análisis de los datos, se construyeron nuevas variables a partir de los indicadores originales relacionados con educación y edad. Estas transformaciones permitieron condensar la información y agrupar a la población en categorías significativas para el análisis sociodemográfico. Se generaron las siguientes variables: Población Infantil, correspondiente a personas de 0 a 17 años; Adultos Jóvenes, que agrupa a la población de 18 a 29 años; Adultos en Edad Productiva, que incluye a personas de 30 a 49 años; y Adultos Mayores Activos, que representa a la población entre 50 y 64 años. En cuanto al nivel educativo, se definió la variable Escolaridad Baja, que considera a personas de 15 años o más con educación básica incompleta o al menos un grado cursado en educación media superior; y Escolaridad Superior, que incluye a personas de 25 años o más con al menos un grado aprobado de nivel superior.

Posteriormente, para un análisis más profundo, se estará evaluando la información de las reseñas de Google de veinte gasolineras con el fin de extraer información relevante y patrones de opinión de los usuarios relacionados con el tiempo de espera y el servicio al cliente.

En la segunda fase del proyecto, se procederá al entrenamiento de modelos de inteligencia artificial. Estos modelos servirán de punto de partida para la automatización relacionada con la toma de decisiones. Finalmente, en la tercera fase, se realizarán pruebas en entornos controlados para evaluar la eficacia,

Tabla. 1. Variables utilizadas en la segmentación de las zonas geográficas.

Variable	Significado
$ECO1_R$	Población de 12 años y más económicamente activa.
$VIV29_R$	Porcentaje de viviendas particulares habitadas que disponen de automóvil o camioneta.
$VIV33_R$	Porcentaje de viviendas particulares habitadas que disponen de televisor.
$VIV34_R$	Porcentaje de viviendas particulares habitadas que disponen de computadora, laptop o tablet.
$VIV35_R$	Porcentaje de viviendas particulares habitadas que disponen de línea telefónica fija.
$VIV36_R$	Porcentaje de viviendas particulares habitadas que disponen de teléfono celular.
$VIV37_R$	Porcentaje de viviendas particulares habitadas que disponen de Internet.
$VIV82_R$	Porcentaje de viviendas particulares habitadas que disponen de servicio de televisión de paga.
$VIV83_R$	Porcentaje de viviendas particulares habitadas que disponen de servicio de películas, música o videos de paga por Internet.
$EDU34_R$	Población de 15 años y más con educación básica incompleta.
$EDU37_R$	Población de 15 años y más con educación básica completa.
$EDU43_R$	Población de 18 años y más con al menos un grado aprobado en educación media superior.
$EDU46_R$	Población de 25 años y más con al menos un grado aprobado en educación superior.
$POB1$	Población total.
$POB8$	Población de 0 a 14 años.
$POB9$	Población de 15 a 17 años.
$POB13$	Población de 18 a 24 años.
$POB14$	Población de 30 a 49 años.
$POB15$	Población de 50 a 59 años.
$POB16$	Población de 60 a 64 años.
$POB30$	Población de 25 a 29 años.
$POB31$	Población de 30 a 34 años.
$POB32$	Población de 35 a 39 años.
$POB33$	Población de 40 a 44 años.
$POB34$	Población de 45 a 49 años.

precisión y adaptabilidad de los sistemas desarrollados. En primer lugar, el experimento se realizará en el estado de Nuevo León, se planea probarlo en dos entornos, Primero, en zonas conglomeradas con alto nivel poblacional (como el centro de Monterrey) donde se encuentran zonas específicas relacionadas con el nivel socioeconómico y otros aspectos relevantes como la edad.

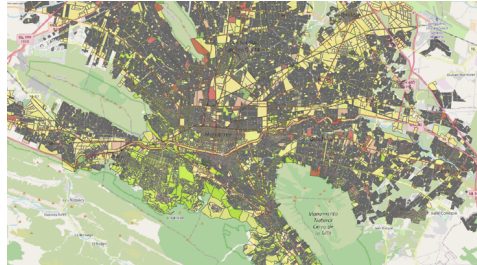


Fig. 4. Cluster de manzanas utilizado K-means, con $k = 5$.

Posteriormente se estará evaluando en zonas con menor accesibilidad a servicios como centros de conveniencia y centros de acceso para el pago de servicios, así como menor nivel poblacional, por ejemplo, cerca de Galeana, al sur de Nuevo León.

Después de realizar la limpieza y transformación de datos, se propuso revisar la implementación de diferentes modelos de Machine Learning para encontrar las manzanas donde sería viable aplicar el servicio de Swit. Como se mencionó anteriormente, las variables utilizadas recopilan información sociodemográfica como las características de vivienda, la edad y la educación, así como la densidad poblacional. Adicional a eso, se estudia la zona, considerando tiendas de conveniencia como Oxxo o 7-Eleven, escuelas y la cantidad de cajeros CFEmáticos que se encuentran a una distancia cercana. Esto permite identificar las zonas de interés y la influencia de los competidores cercanos.

Primeramente, se realizó la segmentación de las zonas de vivienda en Nuevo León. Se puede observar el resultado usando k-means como se ve a continuación en la siguiente imagen (*Figura 4*). En este caso la predicción de 0 arroja las manzanas que cuentan con todos los servicios, por ejemplo, San Pedro. Por el otro lado, mientras más se aleja de este número, se acerca a zonas con mayor escasez de servicios no básicos, por ejemplo, la colonia Independencia.

En el siguiente mapa se puede visualizar las zonas con amplio potencial para colocar el servicio, pues son zonas retiradas de la ciudad metropolitana. Como se puede observar (*Figura 5*), las zonas que presentan mayor área de oportunidad en la accesibilidad al pago de servicios es la zona sur.

6 Resultados del Modelo

Con el objetivo de identificar zonas con alto potencial para la implementación del servicio, se entrenaron y evaluaron dos modelos de clasificación supervisada: **Random Forest** y **Regresión Logística**. Estos modelos fueron seleccionados por su solidez teórica, interpretabilidad y buen desempeño en tareas con múltiples variables categóricas y numéricas. [3]

Para la evaluación de ambos modelos, se utilizó el **área bajo la curva ROC (ROC AUC)**, una métrica estándar que mide la capacidad del modelo para

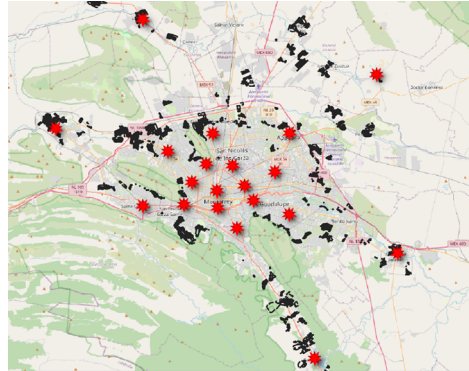


Fig. 5. Manzana con menos de 3 gasolineras y puntos referenciados de cajeros automáticos.

distinguir entre clases sin necesidad de fijar un umbral específico. Los resultados obtenidos fueron:

- **Random Forest:** ROC AUC = 0.922
- **Regresión Logística:** ROC AUC = 0.876

El modelo Random Forest demostró un desempeño superior, con una mayor capacidad discriminativa entre zonas óptimas y no óptimas. Para optimizar el rendimiento del modelo, se utilizó la técnica de Grid Search, la cual permite explorar de forma sistemática todas las combinaciones posibles de hiperparámetros definidos.

Los hiperparámetros ajustados fueron: número de árboles, profundidad máxima de cada árbol, número mínimo de muestras para dividir un nodo, número mínimo de muestras requeridas en una hoja. Una vez entrenado el modelo con los hiperparámetros óptimos, se procedió al análisis del cutoff a fin de determinar el umbral de decisión que ofrezca el mejor equilibrio entre precisión, recall y utilidad práctica.

Aunque la curva ROC permite evaluar la calidad general del modelo, para utilizarlo en la práctica es necesario establecer un cutoff, es decir, un umbral de probabilidad a partir del cual se clasifica una observación como “zona óptima”.

Se probaron distintos valores de cutoff y se observaron sus efectos en métricas clave como precisión, recall y F1-score. El análisis reveló que un cutoff de aproximadamente 0.42 proporciona el mejor equilibrio entre exactitud y cobertura, permitiendo un balance efectivo entre falsos positivos y falsos negativos. Con este umbral, el modelo Random Forest obtuvo los siguientes resultados:

- **Precisión:** 0.83
- **Recall:** 0.88
- **F1-score:** 0.85

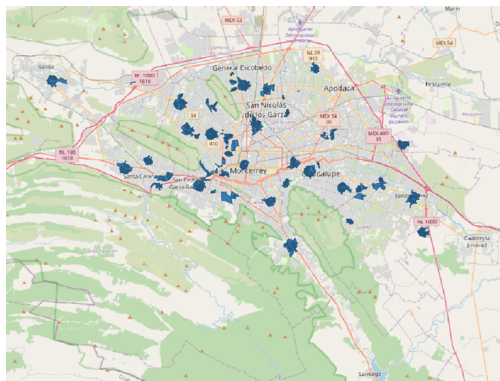


Fig. 6. Lugares potenciales para la aplicación del servicio.

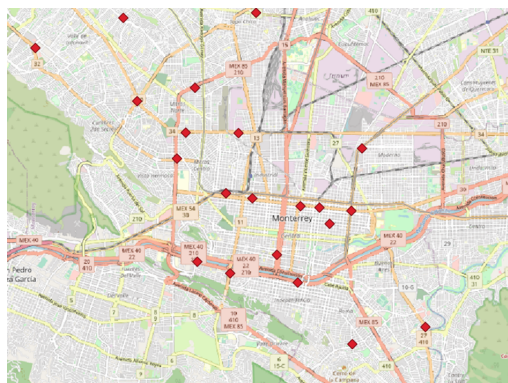


Fig. 7. Estudio de caso: 20 gasolineras.

– **Cost-matrix gain:** 0.48

Estos resultados validan el uso de Random Forest como herramienta central para apoyar decisiones estratégicas de localización, con una base sólida en datos geoespaciales y sociodemográficos.

Después de obtener la clasificación, en la Figura 6 se observa cuáles son las zonas donde se podría aplicar el servicio.

Se realiza entonces el siguiente procedimiento para determinar dónde colocar el servicio. Se estudia el caso de 20 gasolineras del centro de Monterrey elegidas de forma aleatoria. Estas se pueden observar a continuación en la Figura 7.

Para volver más completa la toma de decisiones, se obtuvieron alrededor de 400 comentarios sobre las reseñas de dichas gasolineras por medio de la API Outscraper, la cual utiliza diferentes técnicas de web scrapping. El producto resultante fue una base resultante de las reseñas, la cual contenía información de la calificación asignada por el usuario y el comentario correspondiente. Como primer acercamiento, se quitaron las palabras comunes como artículos

Símbolo	Valores	Leyenda
✓ 	2.371 - 5.638	2.4 - 5.6
✓ 	5.638 - 6.795	5.6 - 6.8
✓ 	6.795 - 8.302	6.8 - 8.3

Fig. 8. Simbología que contiene tres clases, basados en la calificación asignada a cada gasolinera.

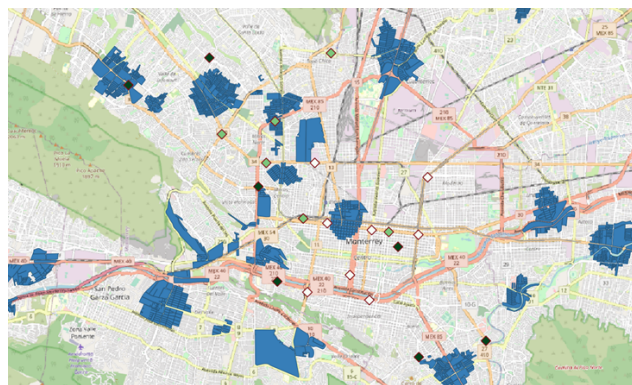


Fig. 9. Mapa que muestra las zonas potenciales junto con las gasolineras clasificadas.

que no servían para el análisis y se obtuvo un texto condensado. Sobresalen valores importantes como “Servicio”, “Precio” y “Atención”. Estos comentarios representan un sentimiento relacionado que puede estudiarse como Positivo, Negativo o Neutro, según sea el caso. De este modo, se clasificó cada comentario con dicha clasificación. Posteriormente, se evaluaron las manzanas que tenían alrededor de 1500 unidades de distancia una gasolinera cercana para que de este modo se pudiera obtener el tipo de mercado al que podía estar dirigido y la distancia a puntos de interés. A continuación, se muestra el resultado de la calificación asignada a cada gasolinera (véase Figura 8). Son tres clases que fueron agrupadas según el valor de la calificación. El tono verde oscuro representa las gasolineras que deberían considerarse para aplicar el servicio de Swit según el modelo.

Si comparamos el mapa propuesto de zonas potenciales (Figura 9), los lugares que arroja el resultado pueden ser bastante útiles, pues se encuentran en las zonas arrojadas y cuentan con una experiencia satisfactoria del cliente, que representa una parte importante del éxito de un negocio.

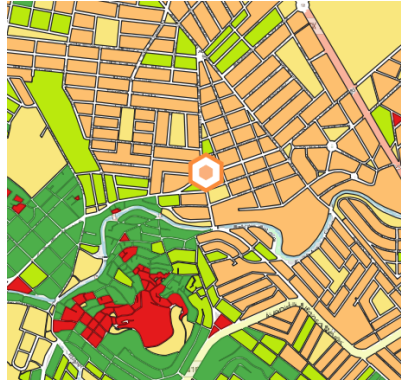


Fig. 10. Caso de estudio: Gasolinera Rio Aguanaval.

7 Descubrimientos

Actualmente cada día hay mayor atención a la toma de decisiones basadas en datos y surgen nombres como geomarketing, geointeligencia o ciencia de datos. Después de analizar el estado del arte, se propone el uso de variables similares para segmentar el mercado y proponer un servicio personalizado. La introducción de un servicio a un nuevo mercado fue el principal problema que fue identificado, ya que representa una gran incertidumbre, muchos de ellos proponían solo segmentar el mercado o solo utilizar puntos de interés sin considerar variables sociodemográficas. En este caso, se propuso un sistema de recomendación automatizado que utilizó diferentes técnicas de machine learning y geointeligencia que pudiera combinar las dos variables y sumarle la opinión del cliente, una variable muy importante a considerar. Este modelo implica un ahorro en el tiempo y una estrategia de ahorro de pérdida segura, pues disminuye el riesgo de que el servicio no funcione en el lugar escogido, no solo para las estaciones de servicio, si no para cualquier lugar en el que se pueda aplicar. Como se menciona anteriormente, se proporciona una metodología que tiene la capacidad de usarse en diferentes industrias y que representa un gran beneficio para sus usuarios.

Se estudia el resultado propuesto que obtuvo la mayor calificación (*Figura 10*). Representa un punto estratégico, pues se encuentra en el centro de varias manzanas con diferente clasificación socioeconómica; de un lado se encuentra la zona cercana a la universidad Tecnológico de Monterrey y la colonia Roma, y en el otro extremo se encuentra cercano el Cerro de la Campana. Se encuentra cerca de escuelas como el Tecnológico de Monterrey, Instituto Excelsior, además de varias primarias y secundarias públicas. Todos estos factores son importantes dentro del modelo. Del mismo modo también cuenta con diferentes rangos de edad pues representa una zona de mucho movimiento para jóvenes, así como personas adultas que reconocen estas manzanas como lugares accesibles y seguros donde se puede rentar. De igual forma, tomaron peso las reseñas encontradas en

Google, pues de las veinte gasolineras estudiadas fue la tercera con el rating más alto (3.9), lo que insta a considerarla como una gasolinera con un buen servicio.

Si bien cuenta con un Oxxo muy cercano es importante considerar que una gran cantidad de gasolineras cuentan con este tipo de servicios de conveniencia por lo cual no afectó tanto la calificación de la gasolinera. Otro punto a tomar en cuenta es que el análisis de las reseñas puede estar sesgado a la falta de comentarios para cada gasolinera. En posibles y futuros trabajos, se espera trabajar con un mayor volumen de reseñas, para adoptar estrategias más personalizadas con las bases de datos que se pudieron analizar sobre la venta de gasolina.

Este es uno de los más grandes desafíos que debe ser abordado en el futuro para tener resultados más confiables y precisos. En el futuro se puede probar el modelo para verificar su eficacia en más ciudades y en otros contextos, no solo en gasolineras; pues la metodología permite adaptarlo y/o extenderlo según sea necesario.

En relación con lo anterior, se propone que el servicio ofrecido no tarde más que el tiempo promedio de espera de cada transacción para no afectar la opinión del usuario, así como ofrecer servicios según la zona. En esta gasolinera en específico se recomienda ofrecer las recargas, el pago de luz y la compra de boletos de cine para potenciar su ubicación estratégica. Del mismo modo, también se propone que se evite ofrecer el servicio en horas pico para no engrandecer el tiempo de espera en la fila y evitar pérdida de clientes.

Por último, tomando en cuenta las consideraciones éticas, es crucial considerar que la recopilación de datos sobre el comportamiento de los usuarios puede comprometer la privacidad de la empresa, pues son datos sensibles. La divulgación indebida de esta información podría tener consecuencias legales y dañar la confianza del cliente; es por esta razón que deben ser protegidos para prevenir el acceso no autorizado o el mal uso de estos.

8 Conclusiones

La diversificación representa una oportunidad a la que se enfrentan diferentes empresas. Gracias a la tecnología, se han desarrollado diferentes formas para facilitar la toma de decisiones conscientes que conlleven a la creación de estrategias flexibles y focalizadas a temas específicos para el desarrollo de una empresa. Por medio de diferentes herramientas de la ciencia de datos, fue posible crear un sistema de recomendación que, por medio del análisis de diferentes fuentes de información, lograra hacer un reconocimiento de las zonas potenciales donde aplicar un nuevo servicio. La ventaja de esta investigación es que tiene una metodología que puede aplicarse no solo a estaciones de servicio, sino a cualquier establecimiento que logre aprovechar los tiempos de espera largos, como bancos, hospitales, lava autos y oficinas de gobierno donde se realicen trámites o en conciertos.

Este reporte muestra la gran adaptabilidad que representa trabajar con Geointeligencia para las diferentes industrias, mostrando así una interactividad

con otros medios como la Ciencia de Datos. Esto abre nuevas posibilidades para la creación de propuestas que mejoren los resultados del modelo, haciéndolo más preciso y mucho más robusto, pues se pueden considerar más variables o ajustar otras dependiendo del caso.

Referencias

1. Ameen, N., Tarhini, A., Reppel, A., Anand, A.: Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, vol. 114, pp. 106548 (2021) doi: 10.1016/j.chb.2020.106548
2. Aslani, M., Alesheikh, A.: Site selection for small gas stations using gis. *Scientific Research and Essays*, vol. 6, pp. 1361–3171 (08 2011)
3. Couronné, R., Probst, P., Boulesteix, A.-L.: Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*, vol. 19, no. 1, pp. 270 (July 2018) doi: 10.1186/s12859-018-2264-5
4. Ergün, M., Uygucgil, H., atalık, O.: Creating a geodemographic classification model within geo-marketing: the case of eskişehir province. *Bulletin of Geography. Socio-economic Series*, pp. 45–61 (03 2020) doi: 10.2478/bog-2020-0003
5. Global, M.: The state of ai in 2022—and a half decade in review (2022), <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>
6. HANG, L., TUAN, B., MANH, V., CHI, N., BINH, B., DIEP, T.: Building a geo-demographic segmentation model: the case of hanoi city, vietnam (01 2021) doi: 10.2991/aebmr.k.211119.048
7. Le, H.: Literature review on diversification strategy, enterprise core competence and enterprise performance. *American Journal of Industrial and Business Management*, vol. 09, pp. 91–108 (01 2019) doi: 10.4236/ajibm.2019.91008
8. Li, D.-C., Lin, Y.-S., Huang, Y.-C.: Constructing marketing decision support systems using data diffusion technology: A case study of gas station diversification. *Expert Syst. Appl.*, vol. 36, pp. 2525–2533 (03 2009) doi: 10.1016/j.eswa.2008.01.065
9. Marques, R., de Paula Ferreira, W., Nassif, G., Armellini, F., Dungen, J., de Santa-Eulalia, L. A.: Exploring the application of iot in the service station business. *IFAC-PapersOnLine*, vol. 54, no. 1, pp. 402–407 (2021) doi: 10.1016/j.ifacol.2021.08.163
10. PetroIntelligence: Fotografía del sector gasolinero en mexico (2021), <https://petrointelligence.com/Contribuciones/fotografia-sector-feb2021.pdf>
11. Statista: Deployment status of data science and machine learning in organizations worldwide as of 2019, by function. Graph (September 9 2019), <https://www.statista.com/statistics/1053561/data-science-machine-learning-deployment-by-function/>, retrieved September 06, 2023
12. Wirth, R., Hipp, J.: Crisp-dm: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, (01 2000)