

PROPUESTA DE MONOGRAFÍA

Título del proyecto	Modelo predictivo de etiquetas para los clientes de Evertec PlacetoPay	
Estudiante 1		
Nombres completos	Laura Cristina Torres López	e-mail: laura.torresl@udea.edu.co
		GitHub: Laurtorr
Estudiante 2		
Nombres completos	Esteban Mateo Jiménez Marín	e-mail: esteban.jimenezm@udea.edu.co
		GitHub: Estebanjm77

1. Descripción del problema

Placetopay by Evertec es una plataforma transaccional de comercio electrónico y pagos no presenciales presente en Colombia y otros países de Latinoamérica. Esta solución permite que comercios grandes y pequeños habiliten canales de venta y cobro en línea desde cualquier dispositivo y cualquier método de pago. La plataforma ofrece servicios de seguridad transaccional, métodos de pago fácil y ajustados a las necesidades de los comercios.

Considerando lo anterior, es de afirmar que la plataforma tiene a su disposición un alto volumen de información, tanto de los comercios como de los usuarios que transan día a día, minuto a minuto. Con esta información, la organización brinda también servicios de análisis de datos para sus clientes con la intención de facilitar la toma de decisiones estratégicas por medio del entendimiento de la dinámica del negocio.

Como una apuesta adicional, la empresa se ha empeñado por incorporar modelos de Machine Learning en sus procesos internos y servicios externos. Sin embargo, en la actualidad aún existe mucho por aportar en este ámbito, debido a que no se cuenta con modelos robustos.

El Core del negocio está en sus servicios y soluciones transaccionales, y, son los comercios los que alimentan estos servicios. Por ende, la atracción y cierre de contratos con antiguos y nuevos clientes es fundamental para la subsistencia y crecimiento de la organización. En el ámbito de la negociación se vuelve fundamental contar con modelos de análisis de datos que orienten la toma de decisiones particulares. Especialmente, modelo de predicción o pronóstico, por ejemplo, modelos que permitan predecir cuántas transacciones procesará determinado comercio en el futuro. De contar con este insumo, se facilitaría la toma de decisiones estratégicas en las negociaciones y cierres de contrato con antiguos y nuevos clientes, y, como ya se enunció, se incidiría directamente en el crecimiento organizacional.

Sin embargo, para llegar a modelos refinados de predicción, se requiere de una base bien constituida. Actualmente, la organización no posee una madurez en la infraestructura ni un proceso establecido para proyectos de machine learning. Así mismo, nos encontramos con que los comercios no se encuentran clasificados ni segmentados. En línea con esto, nos propusimos crear un modelo de predicción para la segmentación de comercios, con esto, esperamos proveer una base sólida para el refinamiento de futuros modelos en la empresa.

2. Descripción del dataset

En el marco de esta monografía se utilizará la data que tiene la organización sobre los comercios y sus transacciones. Estos datos estarán anonimizados a través de un hash que transforma los datos de entrada en un código. Entre las variables contenidas en la base de datos encontramos: nombre (código) del comercio, número de usuarios, número de transacciones, monto, sector comercial, fecha o período, tasa de aprobación, número de transacciones fraudulentas, medio de pago, frecuencia de compra, grupo tarifa, entre otros.

Si los datos se presentan de forma anonimizada y detallada, el dataset puede alcanzar una volumetría de centenas de millones de datos. Si, son entregados de forma agrupada, el dataset puede alcanzar una volumetría de 100.000 registros. Ambas opciones serán evaluadas para la construcción del modelo una vez se tenga una guía u orientación del costo computacional y la metodología que se desee implementar.

3. Métricas de desempeño

Métrica de desempeño del modelo: Teniendo en cuenta que el objetivo de este desarrollo es lograr obtener un predictor de etiquetas de los clientes de Placetopay (vector de creación de etiquetas lo suficientemente valido que se ajuste a la distribución de los datos), se plantea inicialmente construir un modelo que utilice algoritmos de clustering o agrupamiento para el cual sería de gran importancia los criterios de separabilidad y compactación. Así pues, la métrica elegida para evaluar el desempeño del modelo de machine learning será el coeficiente de Silhouette ya que mide tanto la similitud entre los puntos de un mismo grupo como la diferencia entre los puntos de diferentes grupos. Para la clasificación de clientes, es importante que los clientes dentro de un mismo grupo sean similares entre sí en términos de sus características y comportamientos, y que los clientes de diferentes grupos sean diferentes en estos mismos aspectos. Con esta métrica (que toma valores entre -1 y 1) se define que valores cercanos a 1 indican una mejor agrupación.

Métrica de negocio: Dado que la empresa manifiesta no tener madurez en la ejecución de proyectos relacionados al desarrollo de modelos de machine learning y en consecuencia con las necesidades manifestadas por la misma, a nivel de negocio se entiende que el valor agregado que este trabajo de monografía aportará se encuentra en la aplicabilidad del modelo predictivo cómo base de los desarrollos subsecuentes que se pretenden realizar en Placetopay (cómo predecir cuándo un comercio tiene intención de transar o desarrollo de modelos de recomendaciones para los clientes) con el fin de tener insumos para la toma de decisiones a nivel estratégico.

4. Criterio de desempeño

Para el despliegue del modelo desarrollado, se busca inicialmente que la métrica del coeficiente de Silhouette sea lo más cercano posible a 1 ya que esto reflejaría una mejor calidad en la agrupación o etiquetado de los clientes. No obstante, aún se debe definir este criterio con la empresa.

5. Referencias

- Leongómez, E. (s.f.). Validación de Clustering [Presentación de diapositivas]. Departamento de Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia. Recuperado de https://disi.unal.edu.co/~eleonguz/cursos/mda/presentaciones/validacion_Clustering.pdf

- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.