

Andrade, Andrés, Calvache, Reynaldo, Illera, José, Potosí, Ángel, Arias, Christian, Gil, Gustavo, (2021). Datamart para el análisis de los hábitos transaccionales de los clientes de una entidad bancaria con el uso del producto tarjeta débito. Cuaderno Activa, 13, 103-111.



Datamart para el análisis de los hábitos transaccionales de los clientes de una entidad bancaria con el uso del producto tarjeta débito

Datamart for the analysis of the transactional habits of Banking Entity customers with the use of the debit card product.

Andrés Felipe Andrade Mayor¹, Reynaldo Manzano Calvache², José Manuel Illera Recalde³, Ángel Antonio Potosí Fajardo⁴, Christian Gustavo Arias Iragorri⁵, Gustavo Eduardo Gil Prado⁶

Recibido: 11 mayo 2021. **Aprobado:** 23 mayo 2021.

Resumen: En este artículo se presentan los resultados del proyecto "Datamart para el análisis de los hábitos transaccionales de los clientes de una entidad bancaria con el uso del producto tarjeta débito", desarrollado como trabajo final en el programa de Especialización en Administración de la Información y Bases de Datos, de la Institución Universitaria Colegio Mayor del Cauca. El objetivo del proyecto fue construir un Datamart que le permitiera a una entidad bancaria generar reportes que apoyaran el proceso de medición y

monitoreo de las costumbres transaccionales de los clientes que hacen uso del producto tarjeta débito. Se empleó la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), una de las más utilizados para trabajar con minería de datos, que proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos. Esta metodología consta de seis fases: Definición de necesidades del cliente o comprensión de negocio, estudio y comprensión de los datos, análisis de los datos y

1 Ingeniero de Sistemas. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: aandrade@unimayor.edu.co

2 Matemático. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: reynaldomanzano@unimayor.edu.co

3 Ingeniero de Sistemas. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: joseillera@unimayor.edu.co

4 Ingeniero en Informática. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: aapotosi@unimayor.edu.co

5 Doctor en Ingeniería. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: charias98@unimayor.edu.co

6 Ingeniero de Sistemas. Institución Universitaria Colegio Mayor del Cauca. Popayán, Colombia.

Correo electrónico: gustavo.gilp@unimayor.edu.co

selección de características, modelado, evaluación y despliegue. Se implementó un modelo de topología en estrella en el Datamart utilizando como gestor de base de datos MYSQL. Asimismo, se construyó un sistema ETL (*Extract, transform and load*) que permitió la carga, limpieza y transformación de los datos para que quedaran almacenados en el Datamart. Finalmente, se evaluaron y se interpretaron los resultados con la herramienta Power BI. Este proyecto permitirá a una entidad bancaria analizar de manera detallada el flujo transaccional de los clientes que hacen uso del producto tarjeta débito.

Palabras clave: CRISP-DM, Datamart, ETL, Tarjeta débito, Topología en estrella.

Abstract. This article presents the results of the project "Datamart for the analysis of the transactional habits of Banking Entity customers with the use of the debit card product" developed as a final project in the program of the Especialización en Administración de la Información y Bases de Datos of the Institución Universitaria Colegio Mayor del Cauca that aimed to build a Datamart that allows a Banking Entity to generate reports that support the process of measuring and monitoring the transactional habits of customers who use the product. debit card. The CRISP-DM (Cross Industry Standard Process for Data Mining) methodology was used, one of the most used methods to work with data mining, it provides a standardized description of the life cycle of a standard data analysis project. This methodology consists of six phases: Defining customer needs or understanding the business, studying and understanding the data, analyzing the data and selecting features, modeling, evaluating and deploying. A star topology model was implemented in the Datamart using MYSQL as the database manager. Likewise, an ETL (Extract, transform and load) system was built that allowed the loading, cleaning and transformation of the data to be stored in the Datamart in a clean way. Finally, the results were evaluated and interpreted with the Power BI tool. The project carried out will allow a Banking Entity to analyze in detail the transactional flow of customers who make use of the debit card product.

Keywords: CRISP-DM, Datamart, Debit Card, ETL, Star Topology.

Introducción

Las entidades bancarias en Colombia atienden diferentes necesidades de los ciudadanos entre las que se encuentran el otorgamiento de créditos de una manera más fácil, rápida y oportuna, desde la introducción de las tarjetas débito y crédito, en la década de 1950, y en la década de 1970 para Colombia, lo cual produjo una revolución en la forma de pago de los consumidores [1], y permitió conocer las actividades que los clientes realizaban a través de estos medios.

Entre los nuevos retos que afronta una entidad bancaria en la actualidad está la creación de nuevas estrategias de negocio con su producto tarjeta débito, y esto tiene que ver con modelos que les permitan analizar la información transaccional de sus clientes. Por la falta de conocimiento acerca de estas estrategias se pierden nuevas oportunidades de negocio que se pueden producir con este tipo de análisis de información.

Con base en lo anterior, el presente proyecto pretende, por medio de un Datamart, apoyar el proceso de medición y monitoreo de los hábitos transaccionales de los clientes que hacen uso del producto *tarjeta débito*. El propósito es incentivar a las entidades bancarias en la creación de nuevas ofertas comerciales que despierten interés en los clientes a partir de conocer los beneficios que pueden obtener al usar los productos transaccionales que se ofrecen.

Metodología

El proyecto se realizó bajo la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*). El estándar incluye un modelo y una guía estructurados en seis fases, algunas de estas fases son bidireccionales, lo que significa que permiten revisar parcial o totalmente las fases anteriores. El proceso CRISP-DM proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los

modelos de ciclo de vida de desarrollo de software [2], [3], [4], [5]. En la Figura 1 se muestran las fases del proceso CRISP-DM.

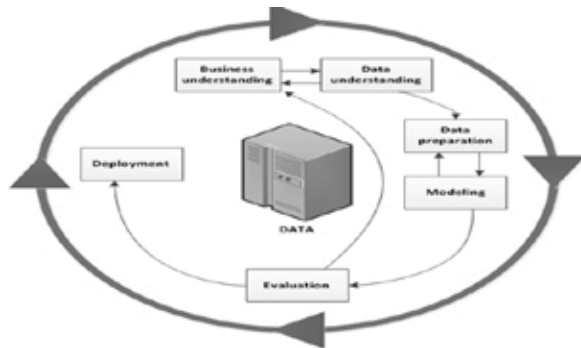


Figura 1. Fases de la metodología CRISP-DM.
 Fuente: tomado de [6].

Fase I. Comprensión del Negocio (*Business Understanding*): se realizan las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, ya que se convierten en objetivos técnicos y en un plan de proyecto [7].

Fase II. Comprensión de los Datos (*Data Understanding*): comprende la recolección inicial de datos con el objetivo de establecer un primer contacto con el problema [7].

Fase III. Preparación de los Datos (*Data Preparation*): en esta fase, una vez efectuada la recolección inicial de datos, se procede a su preparación para adaptarlos a las técnicas de *Data Mining* que se utilicen posteriormente [7].

Fase IV. Modelado (*Modelling*): en esta fase de CRISP-DM se seleccionan las técnicas de modelado más apropiadas para el proyecto de *Data Mining* específico [7].

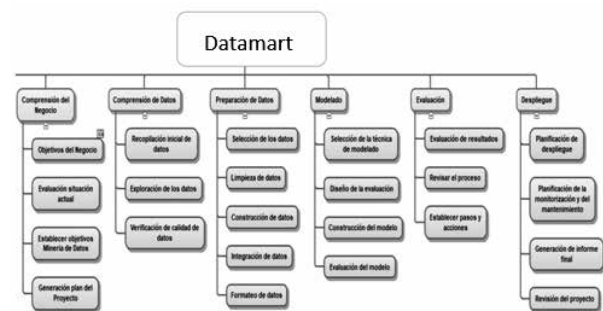
Fase V. Evaluación (*Evaluation*): en esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema [7].

Fase VI. Despliegue (*Deployment*): en esta fase, y una vez que el modelo ha sido construido y

validado, el conocimiento obtenido se transforma en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, o que el modelo se aplique a diferentes conjuntos de datos o como parte del proceso [7].

En la Figura 2 se despliega el alcance del proyecto y cada uno de sus entregables.

Figura 2. Alcance del proyecto.



Fuente: elaboración propia.

En el proyecto se establecieron los siguientes requerimientos para el proceso:

- R1. Cantidad de transacciones ATM realizadas por los clientes, por fecha, tipo de transacción y código de respuesta.
- R2. Cantidad de transacciones POS realizadas por los clientes, por fecha, tipo de transacción y código de respuesta.
- R3. Cantidad de transacciones ATM nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, nombre departamento y fecha.
- R4. Cantidad de transacciones POS nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, nombre municipio y fecha.
- R5. Cantidad de transacciones ATM relacionadas con retiros nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, fecha y municipio, comparado con el porcentaje de hurto de cada municipio y el porcentaje de hurtos en las entidades financieras.

R6. Cantidad de transacciones POS relacionadas con compras nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, fecha y municipio, comparado con el porcentaje de hurto de cada municipio y el porcentaje de hurtos en las entidades financieras.

R7. Cantidad de transacciones POS, por departamento, municipio, almacén, fecha, código de transacción y código de respuesta.

A continuación, en la Tabla 1 se muestra el desarrollo de la matriz de dimensiones, como solución a los requerimientos definidos en el proyecto.

Tabla 1. Matriz de dimensiones.

	TIEMPO	TIPO_TRNX_ATM	TIPO_TRNX_POS	MUNICIPIO	HURTOS COLOMBIA	HURTOS ENTIDADES FINANCIERAS	COD_RESP_POS	COD_RESP_ATM
R1	X	X						X
R2	X		X				X	
R3	X	X		X				X
R4	X		X	X			X	
R5	X	X		X	X	X		X
R6	X		X	X	X	X	X	
R7	X		X	X			X	

Fuente: elaboración propia.

En la Tabla 2 se presenta la descripción de las dimensiones de la Bodega de datos.

Tabla 2. Descripción de dimensiones.

NOMBRE	DESCRIPCIÓN
transacciones_atm	Tabla donde se encuentra el historial de transacción ATM
transacciones_pos	Tabla donde se encuentra el historial de transacciones POS
dim_ciudad	Esta dimensión contiene código departamento, nombre departamento, código municipio, nombre municipio, zona postal, código postal y tipo (rural o urbano)
dim_codigos_resp_atm	Códigos de respuesta diferentes para cada caso, que se manejan para saber si la transacción es aprobada o rechazada
dim_tipo_tmx_atm	Códigos de los diferentes tipos de transacciones ATM
dim_codigos_resp_pos	Códigos de respuesta diferentes para cada caso, que se manejan para saber si la transacción es aprobada o rechazada
dim_tipo_tmx_pos	Códigos de los diferentes tipos de transacciones POS
din_hurtos_colombia	Se extrae un histórico de hurtos del año 2019 de la pagina de la Policía de los diferentes municipios por entidades financieras y hurtos por delincuencia común, para obtener un porcentaje de criminalidad por cada municipio (número de
dim_hurtos_ef	
dim_tiempo	Esta dimensión contiene las fecha en diferentes formato en un intervalo de tiempo determinado

Fuente: elaboración propia.

En la Tabla 3 se muestra la definición de cubos.

Tabla 3. Definición de cubos.

Nº	NOMBRE	DESCRIPCIÓN	AGREGACIÓN
1	fac_req1_atm	Cantidad de transacciones ATM realizadas por los clientes, por fecha, tipo de transacciones y código de respuesta.	count (cod_tipo_trans)
2	fac_req2_pos	Cantidad de transacciones POS realizadas por los clientes, por fecha, tipo transacción y código de respuesta.	count (cod_trans)
3	fac_req3_atm	Cantidad de transacciones ATM nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, nombre, municipio y fecha.	sum (monto) count (cod_tipo_trans)
4	fac_req4_pos	Cantidad de transacciones POS nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, nombre de municipio y fecha.	sum (monto) count (cod_trans)
5	fac_req5_atm	Cantidad de transacciones ATM relacionadas con retiros nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, fecha y municipio comparado con el porcentaje de hurto de cada municipio y el porcentaje de hurtos en la entidades financieras.	sum (monto) count (cod_tipo_tans)
6	fac_req6_pos	Cantidad de transacciones ATM relacionas con compras nacionales realizadas por los clientes, por tipo de transacción, código de respuesta, monto, fecha, y municipio comparado con el porcentaje de hurto de cada municipio y el porcentaje de hurtos en las entidades financieras.	sum (monto) count (cod_tans)
7	fac_req7_atm	Cantidad de transacciones POS, por departamento, municipio, almacén, fecha, código de transacción y código de respuesta.	sum (monto) count (cod_tans)

Fuente: elaboración propia.

Resultados

Luego de llevar a cabo el proceso mencionado en la metodología, en esta sección se presentan los resultados obtenidos.

1. Como resultado principal se obtuvo un Datamart con un modelo dimensional en topología en estrella de la bodega de datos, que comprende las dimensiones con los movimientos transaccionales POS (*Point of sale*) y ATM (*Automated teller machine*) de las tarjetas débito de una Entidad Bancaria.

2. Un proceso ETL para la carga y transformación periódica de la información proveniente de archivos planos que contienen los movimientos transaccionales por tarjeta débito POS y ATM de una entidad bancaria.

3. La *Dashboard* que facilita la visualización, medición y monitoreo de los hábitos transaccionales de los clientes de una entidad bancaria.

En la Figura 2 se relaciona el modelo dimensional para la bodega de datos.

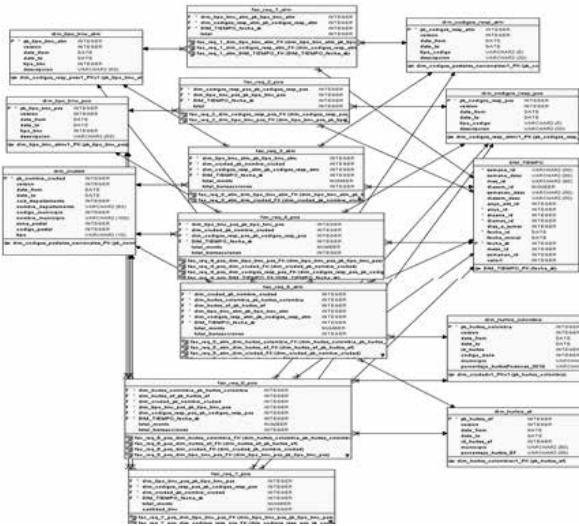


Figura 2. Modelo dimensional.

Fuente: elaboración propia.

En las Figuras 3, 4, 5 y 6 se relaciona el proceso ETL automatizado en la herramienta Pentaho Data Integration (PDI), con la cual es posible presentar los pasos ejecutados en el proceso de transformación que se utilizó en la elaboración de la Bodega de datos y, posteriormente, realizar el análisis de resultados.

En la Figura 3 se muestra el proceso ETL, y se especifica cada paso ejecutado para cargar el historial de transacciones ATM en la Bodega de Datos.

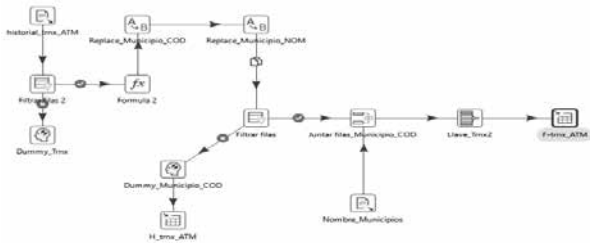


Figura 3. Proceso ETL de transacciones ATM.

Fuente: elaboración propia.

En la Figura 4, por su parte, se muestra el proceso ETL especificando cada paso ejecutado para cargar el historial de transacciones POS en la Bodega de datos.

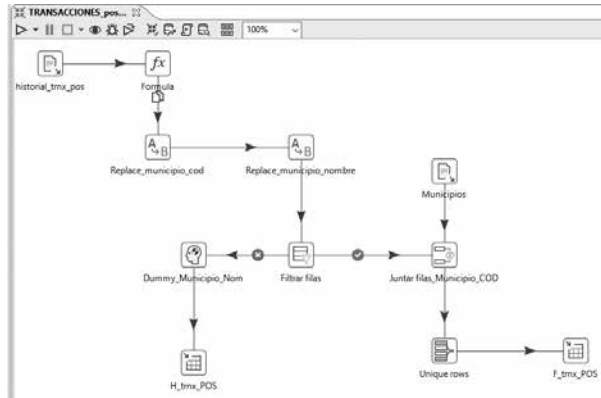


Figura 4. Proceso ETL de transacciones POS.

Fuente: elaboración propia.

Mientras en la Figura 5 se muestra el ETL para cargar las dimensiones ciudad en la Bodega de datos.

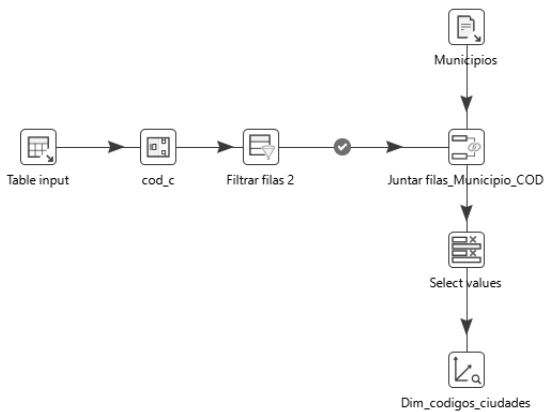


Figura 5. Proceso ETL para dimensión ciudad.

Fuente: elaboración propia.

En la Figura 6 se muestra el ETL para cargar la dimensión tiempo en la Bodega de datos.

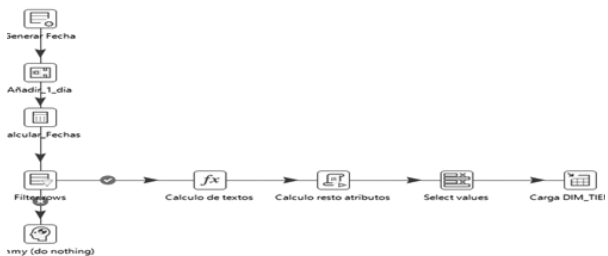


Figura 6. Proceso ETL para dimensión tiempo.

Fuente: elaboración propia.

En las siguientes Figuras (7, 8, 9, 10, 11, 12 y 13) se relacionan los resultados del *Dashboard* resultante de cada requerimiento, empleando la herramienta de Inteligencia de Negocios Power BI.

Informe Requerimiento 1: en el informe de la Figura 7 se identifica la cantidad de transacciones ATM realizadas por tipo de transacción, respuesta de la transacción y fecha.

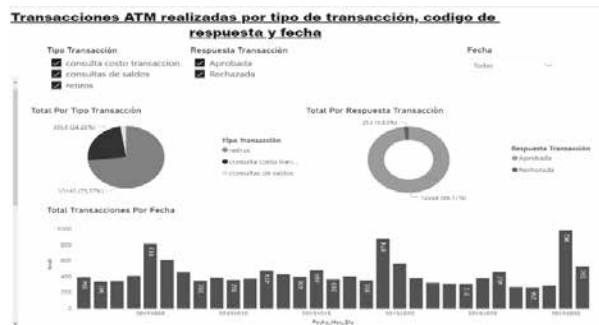


Figura 7. Total de transacciones ATM por tipo, código y fecha.
 Fuente: elaboración propia.

En este reporte, el mayor porcentaje de transacción ATM es de tipo retiro con un 73,37 % y en respuesta de transacciones aprobadas un 98,17 %, además se puede evidenciar el día con mayor número de transacciones ATM por fecha.

Informe Requerimiento 2: en el informe en la Figura 8 se identifica la cantidad de transacciones POS realizadas por tipo de transacción, respuesta de la transacción y fecha.

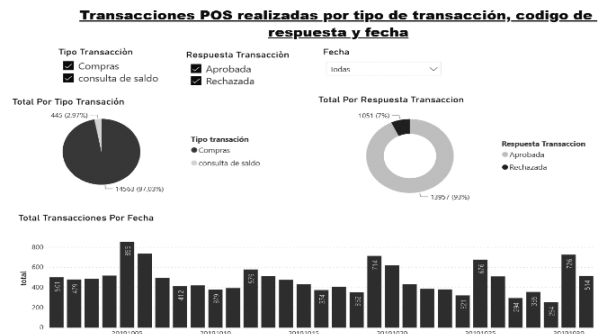


Figura 8. Total de transacciones POS por tipo, código y fecha.
 Fuente: elaboración propia.

En este reporte, el mayor porcentaje de transacción POS es de tipo compras con un 97,03 %, y consulta de saldo, el 2,97 %, en cuanto a la respuesta de las transacciones: aprobadas el 93 % y rechazadas el 7 %, además se puede evidenciar el día con mayor número de transacciones POS por fecha.

Informe Requerimiento 3: en el informe de la Figura 9 se identifica la cantidad de transacciones ATM realizadas por monto, tipo de transacción, respuesta de la transacción, departamento y municipio.



Figura 9. Total de transacciones ATM por monto, departamento y municipio.
 Fuente: elaboración propia.

En este reporte de transacción ATM por monto y departamento, en este caso el Valle del Cauca, de un total de 1.246 transacciones, en el mayor monto por municipio se encuentra Pereira con un 7,2 % y le sigue la localidad de Usaquén con un 6,95 %.

Informe Requerimiento 4: en el informe de la Figura 10 se identifica la cantidad de transacciones POS realizadas por monto, tipo de transacción, respuesta de la transacción, departamento y municipio.



Figura 10. Total de transacciones POS por monto, departamento y municipio.
 Fuente: Elaboración Propia

En este caso, en este reporte de transacción POS se evidencia que el departamento con mayor monto de transacción es Casanare, y se puede conocer, además, el monto por almacén de un total de 676 transacciones que se realizaron en el mes de octubre de 2019.

Conclusiones

EL Datamart generado brinda información de calidad que ayudará a los directivos de la entidad bancaria a tomar decisiones acertadas y oportunas en lo que tiene que ver con la formulación de estrategias de negocio relacionadas con el producto de tarjetas débito.

El proyecto le permitió a la entidad bancaria analizar de forma detallada el flujo transaccional de los clientes que hacen uso del producto tarjeta débito, mediante la evaluación de la cantidad de transacciones ATM y POS, aprobadas o rechazadas, por departamento, municipio, monto y almacenes.

Se pudo, además, identificar una mayor percepción del riesgo transaccional por índice de hurtos a entidades financieras y personas, en ciudades con mayor flujo transaccional, como fue el caso de la localidad de Usaquén en Bogotá.

En la entidad bancaria se pudieron identificar los canales transaccionales por departamento, municipio y monto, donde se presentan la mayor cantidad de transacciones aprobadas y rechazadas, lo que permitió analizar las causas de por qué se presentan rechazos transaccionales.

Con los resultados obtenidos, se identificaron los almacenes con mayor flujo transaccional, por departamento, municipio, monto y fecha, que podrían generar una posibilidad de alianzas estratégicas comerciales con la entidad bancaria.

En la entidad bancaria se lograron evidenciar los departamentos y municipios con mayor flujo transaccional, lo que permite evaluar la gestión comercial de las agencias.

Referencias

- [1] J. E. Gómez-González, J. Jaramillo-Echeverri, y A. Meisel-Roca, "El uso de efectivo y tarjetas débito y crédito en Colombia," *Borradores Econ.*, n.º 950, 2016.
- [2] A. Azevedo and M. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview", in *Proceedings of IADIS European Conference on Data Mining*, Amsterdam, Netherlands, 2008, pp. 182-185.
- [3] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide", CRISP-DM consortium: NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA), and OHRA Verzekeringen en Bank Groep B. V., The Netherlands, 2000.
- [4] J. Hernández, M. Ramírez, y C. Ferri, *Introducción a la Minería de Datos*, Madrid: Pearson Educación, 2005. Disponible en: <http://dspace.ucbscz.edu.bo/dspace/handle/123456789/526>.
- [5] J. Villena, *CRISP-DM: La metodología para poner orden en los proyectos de Data Science*, 2016. Disponible en: <https://data.sngular.team/es/art/25/crisp-dm-la-metodologia-para-poner-orden-en-los-proyectos-de-data-science>
- [6] IBM, "IBM SPSS Modeler CRISP-DM," IBM Corp., 2016.
- [7] Smartbase Group, [En línea]. Disponible en: <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>