



Tecnológico de Monterrey

Documento técnico

Datathon

The Outliers

Facundo Bautista Barbera | A01066843

Juan Marco Castro Trinidad | A01742821

Azul Fernanda Paniagua Bucio | A01424351

Patricio Mourra Cossío | A01424357

Fecha:

25 may 2025

1. Objetivo General

Diseñar y construir una solución integral basada en machine learning que permita predecir gastos del siguiente mes del cliente, con una interfaz amigable que facilite la visualización y el entendimiento de dichas predicciones, mejorando así la experiencia bancaria y fomentando la retención de usuarios.

2. Reconocimiento del Tipo de Problema

Este reto se aborda mediante:

- **Regresión:** para estimar el monto y la fecha aproximada.
- Predicción de valores continuos
- En algunos casos, se aplicaron **reglas heurísticas** para manejar patrones atípicos o fuera del entrenamiento supervisado.

3. Plan de Acción

A lo largo del desarrollo del proyecto, seguimos una serie de fases cuidadosamente estructuradas para garantizar una solución completa, efectiva y alineada con los objetivos del reto. Estas fases se describen a continuación:

Fase 1: Comprensión del problema

Objetivo: Alinear el equipo con la necesidad de negocio planteada por Hey Banco: anticipar gastos recurrentes como una estrategia para aumentar la retención de clientes.

Acciones realizadas:

- Revisión detallada del planteamiento del reto y criterios de evaluación.
- Discusión en equipo sobre la naturaleza del problema (clasificación + regresión).
- Identificación de oportunidades para generar valor accionable con predicciones:
 - Alertas automáticas.
 - Visualización personalizada.
 - Sugerencias de ahorro preventivo.

Resultado: Diseño de una solución basada en IA + interfaz accesible.

Fase 2: Análisis exploratorio de datos (EDA)

Objetivo: Detectar patrones de comportamiento en los datos de transacciones y perfiles de clientes.

Acciones realizadas:

- Limpieza de más de **346,000 transacciones**, identificando si se encontraban outliers y registros nulos.
- Generación de nuevas variables: frecuencia de gasto por comercio, última fecha de gasto, ciclo mensual típico, tipo de operación.
- Agrupamiento de datos por cliente y comercio para detectar secuencias repetitivas.
- Gráficas estáticas y dinámicas para visualizar mejor los datos

Herramientas usadas: Python, Pandas, Matplotlib, Seaborn.

Resultado: Identificación de comportamientos recurrentes y segmentos útiles para el modelo.

Fase 3: Desarrollo del modelo predictivo

Objetivo: Predecir, para cada cliente y comercio, si ocurrirá un gasto recurrente, cuándo y por cuánto.

Acciones realizadas:

- Definición de dos subproblemas:
 - **Regresión:** ¿Cuándo ocurrirá y de qué monto será?
 - **Clasificación binaria (Posible mejora):** ¿Ocurrirá nuevamente el gasto?
- Implementación de modelos con **LightGBM**:
 - Es un framework de aprendizaje automático basado en el algoritmo de gradient boosting, diseñado para ser eficiente y escalable, especialmente con grandes volúmenes de datos.
 - Combina múltiples árboles de decisión en secuencia.
- Divisiones estratificadas de entrenamiento/prueba por cliente.

- Evaluación mediante métricas:
 - **Regresión:** RMSE, MAE.
 - **Clasificación (Como posible mejora):** AUC, F1-score, log-loss.
- Ajuste de hiperparámetros por GridSearch (learning_rate, num_leaves, max_depth) mediante validación cruzada.

Resultado: Modelo eficiente y ligero con alta capacidad de generalización, listo para servir predicciones bajo demanda. Este modelo se puede ir mejorando con la recolección de más datos

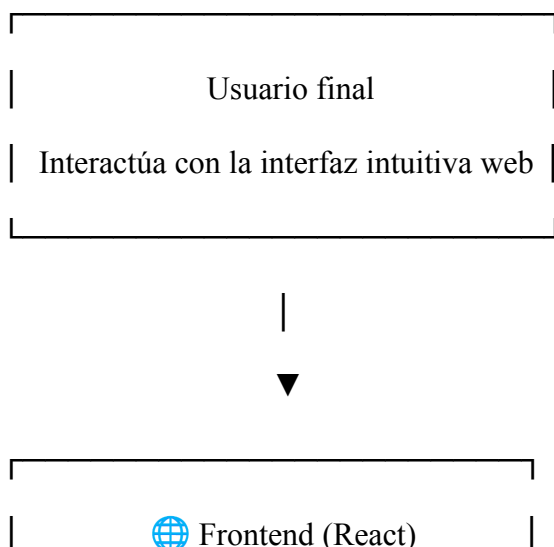
Fase 4: Desarrollo de la interfaz en Streamlit

Objetivo: Crear una experiencia visual simple, intuitiva y accesible desde dispositivos móviles y escritorio, conectada al modelo.

Acciones realizadas:

- Diseño con enfoque **mobile-first**, usando la paleta de colores, tipografía y anuncios de la app oficial de Hey Banco.
- Implementación de vistas: resumen de predicciones, visualización por categoría, calendario de próximos gastos, y detalles por comercio.
- Creación de una API REST que permite consultar las predicciones por ID de cliente.
- Pruebas en distintos navegadores y dispositivos para asegurar compatibilidad.

Arquitectura:



- Visualiza histórico de gastos	
- Muestra comercios frecuentes	
- Grafica gastos por mes	
- Integra predicción del próximo mes	
- Solicita datos desde la API REST	

| HTTP requests (JSON)

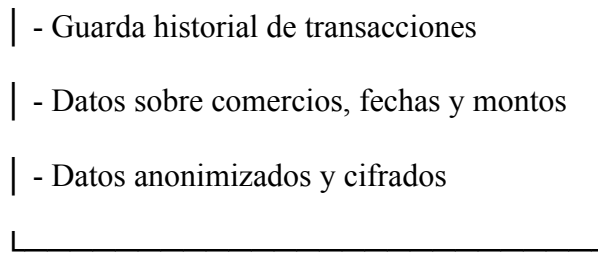


Backend (Flask API)	
- Maneja peticiones del frontend	
- Consulta base de datos PostgreSQL	
- Procesa y limpia datos	
- Aplica modelo de ML para predicción	
- Responde al frontend con datos	
Seguridad:	
- Cifrado TLS en toda la comunicación	
- Cifrado/descifrado de datos al consultar	
la base de datos	

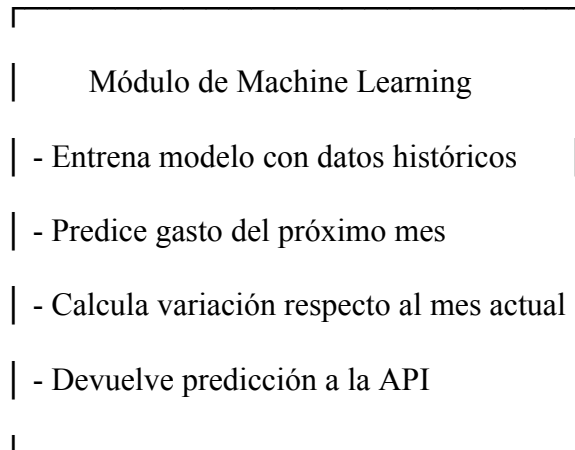
| Encrypted queries



Base de Datos (PostgreSQL)	



Flujo de Datos Seguro



Resultado: Interfaz funcional, atractiva y fácilmente escalable.

Fase 5: Integración, pruebas y ajuste final

Objetivo: Verificar la integración completa entre el modelo, la API y la interfaz.

Acciones realizadas:

- Verificación del flujo: input del usuario → llamado al modelo → respuesta con predicción → render en Streamlit.

- Visualizaciones de historial de compras en un período mensual. Esto incluye una tabla de gasto mensual en la que se compara el monto de los gastos de cada mes vs el mes anterior, categorías más consumidas por mes.
- Predicción del monto y su variación para el siguiente mes

Resultado: Solución completa, estable y con experiencia coherente al entorno digital de Hey Banco.

Fase 6: Elaboración de entregables: presentación, ficha técnica y documentación.

4. Resultados y Alcances

- Modelo predictivo funcional conectado a la interfaz.
- Capacidad de estimar **probabilidad, fecha y monto de futuros gastos**.
- Interfaz accesible desde celular o PC.
- Visualizaciones de datos históricos fáciles de entender.

5. Plan accionable

De acuerdo con las predicciones obtenidas, observamos que hay algunos usuarios que no tienen un gasto muy estable. Es decir, algunos meses se dispara y otros meses se mantiene bajo. Es por ello que al momento de hacer la predicción, se podría agregar una recomendación de algún producto ya existente de Hey, Banco para gestionar sus gastos, abrir una cuenta de inversión, etc.

Además,

6. Viabilidad y Comparativa

Comparativa con apps actuales (Ej: Fintonic, BBVA “Mi día a día”): pocas predicen con precisión futura ni ofrecen visualización sencilla de cargos estimados.

Nuestra solución - Hey Banco

Aspecto	Descripción
Funcionalidad central	Predicción de gastos recurrentes personalizados (monto, fecha) y visualización en una interfaz interactiva.
Predicción futura	Sí, con IA (clasificación + regresión).
Visualización clara	Sí, con Streamlit en estilo mobile-first y adaptada a la imagen de Hey Banco.
Interactividad	Alta: el usuario puede consultar predicciones, ver detalles y filtrar.
Nivel de personalización	Alto: por cliente individual, en base a comportamiento histórico.
Tecnología usada	Modelo de IA + API + Streamlit.

BBVA - Herramienta “Mi Día a Día” (España y México)

Aspecto	Descripción
Funcionalidad central	Clasificación de transacciones por categoría (alimentación, transporte, etc.) y visualización de gastos pasados.
Predicción futura	No realiza predicción exacta de montos ni fechas.
Visualización clara	Sí, estilo dashboard móvil.
Interactividad	Limitada: muestra históricos, pero no permite planificar.
Nivel de personalización	Medio: categorías genéricas para todos los clientes.
Tecnología usada	Regla de categorización y dashboard en app.

Fintonic (España y Latinoamérica)

Aspecto	Descripción
---------	-------------

Funcionalidad central	Seguimiento de gastos, alertas de cobros inesperados, informes financieros.
Predicción futura	Limitada: notifica si detecta patrones, pero sin fecha/monto específico.
Visualización clara	Sí, tipo app de finanzas personales.
Interactividad	Media: reportes interactivos, pero no conecta con predicción detallada.
Nivel de personalización	Medio-alto: se adapta con uso, pero no basado en ML explícito.
Tecnología usada	Heurísticas + categorización financiera tradicional.

Revolut (Internacional)

Aspecto	Descripción
Funcionalidad central	Agrupación de gastos, presupuestos mensuales, alertas y estadísticas.
Predicción futura	No explícita. Puede estimar consumos promedio, pero no anticipa pagos.
Visualización clara	Sí, moderna y atractiva.
Interactividad	Alta en presupuestos, pero no en predicción.
Nivel de personalización	Medio: responde a hábitos, pero no se anticipa.
Tecnología usada	Análisis estadístico de hábitos y UX avanzada.

Conclusión de la comparativa:

Nuestra propuesta es única por su enfoque proactivo, no solo categoriza o analiza el pasado, sino que anticipa el futuro financiero del cliente, con base en inteligencia artificial, presenta resultados de forma clara, personalizada e interactiva, integrable a la experiencia Hey Banco y puede escalar para brindar más valor con funcionalidades como alertas automáticas, ajustes de presupuesto o sugerencias de ahorro.

Costos estimados:

Infraestructura

Recurso	Descripción	Costo estimado mensual
Servidor API/backend	Máquina virtual pequeña (por ej. AWS EC2 t3. small, 2 vCPU, 2 GB RAM) para servir el modelo	USD \$10 - \$15
Almacenamiento	Base de datos ligera (ej. PostgreSQL o SQLite gestionado), con datos de predicciones e historial	USD \$5 - \$10
Modelo en ejecución	Como el modelo puede predecir bajo demanda (No en tiempo real masivo), se puede usar sin GPU. Solo se requiere CPU.	USD \$0 - \$5
Front end	Puede montarse como app web ligera en el mismo servidor	USD \$5 - \$10
Total estimado mensual	-	USD \$20 - \$35

Costos Humanos (Equipo y mantenimiento)

Actividad	Recurso humano	Tiempo estimado	Costo estimado
Monitoreo del modelo	Data Scientist o MLOps(medio tiempo)	10h/mes	USD \$250 - \$400
Mejora de interfaz/UX	Frontend developer o persona técnica	6h/mes	USD \$150 - \$250
Soporte y ajustes	General (bugs)	4h/mes	USD \$250 - \$400
Total estimado mensual			USD \$500 - \$850

Costos totales

Categoría	Costo mensual estimado
Infraestructura	USD \$20 - \$35
Recursos humanos	USD \$500 - \$850
Total mensual	USD \$520 - \$885

7. Limitaciones y Mejoras Futuras

- La precisión aún puede mejorar en categorías de gasto atípicas.
- Integrar lógica adaptativa para usuarios nuevos sin historial.
- Escalar predicción con datos de tiempo real de pagos automáticos.
- Mejorar las métricas del modelo con más datos de entrenamiento.
- Agregar un modelo de clasificación en donde analizaríamos qué tan recurrente es que un cliente compre en cierta tienda

y/o con el mismo monto

8. Conclusión

Este proyecto demuestra cómo una solución de machine learning bien integrada en una interfaz clara puede generar valor inmediato tanto para el usuario como para el negocio, y está lista para escalar dentro del ecosistema digital de Hey Banco.