

Documento de Requerimiento: Modelo de Machine Learning para Satisfacción de Clientes de Aerolínea

Versión: 1.0

Fecha: 08/05/2025

Cliente: AeroLíneas Horizonte S.A.

Proveedor: Consultora Analytics Avanzados

1. Introducción

1.1 Propósito

Este documento define los requerimientos para el desarrollo de un modelo de machine learning que permitirá a AeroLíneas Horizonte S.A. predecir y analizar la satisfacción de sus clientes basado en múltiples variables recopiladas durante la experiencia de vuelo.

1.2 Alcance

El sistema desarrollado analizará datos históricos de clientes para identificar patrones y factores que influyen en la satisfacción, permitiendo a la aerolínea tomar decisiones estratégicas para mejorar la experiencia del cliente y su lealtad a la marca.

1.3 Definiciones y Acrónimos

- **ML:** Machine Learning
- **IA:** Inteligencia Artificial
- **API:** Application Programming Interface
- **NPS:** Net Promoter Score
- **ETL:** Extract, Transform, Load

2. Descripción General

2.1 Perspectiva del Producto

El modelo de ML será integrado con los sistemas existentes de la aerolínea para recibir datos de clientes en tiempo real y proporcionar análisis predictivos sobre satisfacción, permitiendo intervenciones proactivas y mejoras en el servicio.

2.2 Funciones del Producto

- Predecir la satisfacción del cliente basado en variables demográficas y de servicio

- Identificar los factores más influyentes en la satisfacción del cliente
- Segmentar clientes según patrones de satisfacción
- Proporcionar recomendaciones para mejorar servicios específicos
- Generar informes automáticos de tendencias de satisfacción

2.3 Características de los Usuarios

El sistema será utilizado por:

- Equipo de experiencia del cliente
- Gerentes de operaciones
- Equipo de marketing y fidelización
- Ejecutivos de alto nivel para toma de decisiones estratégicas

3. Dataset y Variables

3.1 Origen de los Datos

Los datos provienen del sistema CRM de la aerolínea, encuestas post-vuelo, y sistemas de reserva, compilados en un dataset unificado que contiene 103,904 registros de clientes con las siguientes variables:

3.2 Variables Disponibles

#	Variable	Registros	Tipo	Descripción
1	id	103,904	int64	Identificador único del cliente
2	Gender	103,904	object	Género del cliente
3	Customer Type	103,904	object	Tipo de cliente (frecuente, ocasional, etc.)
4	Age	103,904	int64	Edad del cliente
5	Type of Travel	103,904	object	Propósito del viaje (negocios, personal)
6	Class	103,904	object	Clase de vuelo (económica, ejecutiva, primera)
7	Flight Distance	103,904	int64	Distancia del vuelo en kilómetros
8	Inflight wifi service	103,904	int64	Puntuación del servicio de wifi a bordo (1-5)
9	Departure/Arrival time convenient	103,904	int64	Puntuación de conveniencia de horarios (1-5)
10	Ease of Online booking	103,904	int64	Puntuación de facilidad de reserva online (1-5)
11	Gate location	103,904	int64	Puntuación de ubicación de puerta de embarque (1-5)
12	Food and drink	103,904	int64	Puntuación de comida y bebida (1-5)
13	Online boarding	103,904	int64	Puntuación de proceso de embarque online (1-5)
14	Seat comfort	103,904	int64	Puntuación de comodidad del asiento (1-5)
15	Inflight entertainment	103,904	int64	Puntuación de entretenimiento a bordo (1-5)
16	On-board service	103,904	int64	Puntuación de servicio a bordo (1-5)
17	Leg room service	103,904	int64	Puntuación de espacio para piernas (1-5)
18	Satisfaction	103,904	object	Variable objetivo (Satisfecho/Insatisfecho)

4. Requerimientos Específicos

4.1 Requerimientos Funcionales

4.1.1 Desarrollo del Modelo

- **RF-01:** El modelo debe predecir la satisfacción del cliente con una precisión mínima del 85%.
- **RF-02:** El modelo debe identificar los 5 factores más influyentes en la satisfacción del cliente.
- **RF-03:** Se deben probar al menos 3 algoritmos diferentes (ej. Random Forest, XGBoost, Redes Neuronales) para determinar el más eficaz.
- **RF-04:** El modelo debe ser capaz de procesar nuevos datos para predicciones en tiempo real.

4.1.2 Procesamiento de Datos

- **RF-05:** Implementar un pipeline ETL automatizado para preprocesar los datos.
- **RF-06:** Desarrollar métodos para manejar valores atípicos y faltantes.
- **RF-07:** Crear variables derivadas que agreguen valor predictivo al modelo.

4.1.3 Visualización y Reportes

- **RF-08:** Desarrollar dashboard interactivo para visualizar predicciones y factores clave.
- **RF-09:** Generar informes automáticos mensuales sobre tendencias de satisfacción.
- **RF-10:** Implementar alertas para cuando métricas específicas caigan por debajo de umbrales predefinidos.

4.2 Requerimientos No Funcionales

4.2.1 Rendimiento

- **RNF-01:** El tiempo de respuesta para predicciones individuales debe ser menor a 1 segundo.
- **RNF-02:** El sistema debe soportar el procesamiento de hasta 10,000 predicciones por hora.

4.2.2 Seguridad

- **RNF-03:** Los datos de clientes deben estar cifrados en tránsito y en reposo.
- **RNF-04:** Implementar autenticación de dos factores para acceso al sistema.
- **RNF-05:** Cumplir con regulaciones GDPR y locales sobre protección de datos personales.

4.2.3 Usabilidad

- **RNF-06:** La interfaz debe ser intuitiva y requerir mínima capacitación para su uso.
- **RNF-07:** El sistema debe proporcionar explicaciones claras sobre cómo se generan las predicciones.

4.2.4 Escalabilidad

- **RNF-08:** El sistema debe ser escalable para manejar incrementos de hasta 50% en volumen de datos.
- **RNF-09:** La arquitectura debe permitir fácil integración de nuevas fuentes de datos.

5. Metodología y Entregables

5.1 Metodología de Desarrollo

Se utilizará una metodología ágil con sprints de 2 semanas y entregas incrementales para permitir revisión y feedback continuos.

5.2 Entregables

1. Análisis Exploratorio de Datos:

- Informe detallado de insights preliminares
- Identificación de patrones y correlaciones clave

2. Modelo de Machine Learning:

- Código fuente documentado
- Resultados de validación y pruebas
- Documentación técnica del modelo

3. Herramientas de Integración:

- API para integración con sistemas existentes
- Documentación de la API
- Scripts de ETL para procesamiento de datos

4. Dashboard y Reportes:

- Interfaz interactiva para visualización
- Plantillas de informes automáticos
- Guía de usuario

5.3 Calendario de Entregas

- **Semana 2:** Plan de proyecto detallado
- **Semana 4:** Análisis exploratorio de datos
- **Semana 8:** Primer prototipo del modelo ML
- **Semana 12:** Modelo refinado con validación completa
- **Semana 16:** Sistema integrado con dashboard y reportes
- **Semana 18:** Entrenamiento y documentación completa

6. Consideraciones Adicionales

6.1 Limitaciones

- El modelo se basa en datos históricos que pueden no reflejar cambios recientes en operaciones.
- Las predicciones tendrán un margen de error inherente que debe ser considerado en la toma de decisiones.

6.2 Suposiciones

- Se supone que los datos proporcionados son representativos de la base total de clientes.
- Se asume que los patrones de satisfacción identificados se mantendrán relativamente estables en el corto plazo.

6.3 Dependencias

- Acceso oportuno a todos los sistemas fuente de datos
- Disponibilidad de expertos en negocio para validar insights y recomendaciones
- Infraestructura técnica adecuada para implementación

7. Criterios de Aceptación

1. Precisión del modelo superior al 85% en datos de prueba.
2. Tiempo de respuesta del sistema conforme a los requerimientos de rendimiento.
3. Dashboard funcional con todas las visualizaciones requeridas.
4. Documentación completa y aprobada.
5. Capacitación exitosa a todos los usuarios designados.
6. Pruebas de integración completadas sin errores críticos.

8. Aprobaciones

Rol	Nombre	Firma	Fecha
Director de Proyecto	Ing. Alexandra Torres	Alext T.	11/05
Representante del Cliente	Ing. Johan Mendez	Mendez M	11/05
Líder Técnico	Ing. Jhonathan Rivera	JR	11/05
Oficial de Protección de Datos	Ing. Jorge Belltra	Beltran J.	11/05