

2025

Proyecto Final del Curso

**Análisis de Satisfacción de Pasajeros de Aerolínea**

Javier Cueto González

Contenido

[Introducción: 2](#_Toc199489106)

[El porqué de esta elección: 2](#_Toc199489107)

[Propuesta de proyecto: 2](#_Toc199489108)

[Obtención y tratamiento de datos: 3](#_Toc199489109)

[Evaluación de su calidad y adecuación. 3](#_Toc199489110)

[Procesado, limpieza, transformación y estructuración de los datos. 4](#_Toc199489111)

[Elección y justificación de modelos o técnicas de IA/ML. 4](#_Toc199489112)

[Resultados de Random Forest 5](#_Toc199489113)

## Introducción:

En este proyecto se tratará la creación y el estudio de un programa de predicción del nivel de satisfacción de los pasajeros de una aerolínea, utilizando para ellos algoritmos de aprendizaje automático sobre datos reales recogidos de las experiencias de usuarios. El objetivo es comprender qué factores influyen más en la percepción del cliente, y construir el modelo predictivo que permita anticipar si un pasajero estará satisfecho o no.

## El porqué de esta elección:

Sencillamente, de las pocas veces que he tenido la oportunidad de viajar en avión, he disfrutado de la experiencia de no solo volar en el avión sino también tanto lo que es el proceso de realizar el viaje (ya sea el propio embarque, la facturación, etc). Esto me llevó a, en el momento de tener que buscar un tema sobre el tratar este trabajo, a inclinarme a realizar este tema, que es la valoración y la satisfacción (o no) de aquellas personas que han ido en avión.

Más allá de cualquier motivación personal, es sabido que cualquier empresa busca, al menos en teoría, recibir feedback de sus clientes que les proporcione sobre la valoración de servicios dados y en que puntos estos se podrían mejorar para hacer mejor la experiencia de los clientes. En el sector aéreo, mejorarla puede suponer una ventaja competitiva significativa. En este contexto, el uso de IA y modelos predictivos ofrece a las empresas una herramienta poderosa para actuar de manera proactiva: mejorar la calidad del servicio, reducir costes y fidelizar clientes.

## Propuesta de proyecto:

Lo que se propone con este proyecto es utilizar las valoraciones que una aerolínea ha estado haciendo a sus usuarios, desde la mera experiencia de comprar los billetes online hasta la wifi del avión, terminando por preguntar a cada usuario simplemente ha estado satisfecho con la experiencia, si no lo ha estado o no sabe. Con todos estos datos reunidos en un conjunto de datos se busca crear un modelo de datos que, a partir de toda la información reunida, pueda predecir con exactitud, sin necesidad que el cliente mismo lo diga, si la experiencia ha sido satisfactoria o no.

Una vez realizado esto, se buscará, en base a los resultados del modelo, que factores son los que más influencian la percepción de los usuarios y en cuales la aerolínea podría actuar para mejorar la experiencia a sus clientes.

## Obtención y tratamiento de datos:

## Evaluación de su calidad y adecuación.

Este dataset contiene un total de 129.880 filas de datos y un total de 24 columnas. Las columnas incluidas en el conjunto de datos son:

* Gender: Género del pasajero (Female, Male)
* Customer Type: Tipo de cliente (Loyal customer, Disloyal customer)
* Age: Edad
* Type of Travel: Tipo de viaje (Personal, Business)
* Class: Clase del vuelo (Business, Eco, Eco Plus)
* Flight Distance: Distancia del vuelo
* Valoraciones del servicio del 0 al 5, donde 0 indica "No aplicable" (el cliente no sabe o no responde a la pregunta o sencillamente no da una valoración que se pueda usae.):
  + Inflight wifi service
  + Departure/Arrival time convenient
  + Ease of Online booking
  + Gate location
  + Food and drink
  + Online boarding
  + Seat comfort
  + Inflight entertainment
  + On-board service
  + Leg room service
  + Baggage handling
  + Check-in service
  + Inflight service
  + Cleanliness

Departure Delay in Minutes

Arrival Delay in Minutes

Satisfaction: Nivel de satisfacción general (Satisfied, Neutral or Dissatisfied)

Con esto se puede ver que se trata de un conjunto de datos bastante extenso. Del análisis que se hizo del dataset se pueden resaltar los siguientes puntos:

* Se observó que en las columnas de valoraciones del servicio requerían una interpretación específica, sobre los 0 en los niveles de satisfacción (por ejemplo, en Inflight wifi service). Según la documentación del dataset, el 0 no indica una mala experiencia, sino que el servicio no fue utilizado o no aplicaba al pasajero. Se decidió mantener estos valores, ya que aportan información útil sobre la disponibilidad o uso del servicio, en lugar de ser considerados como valores erróneos.
* El campo ‘Satisfaction’, el que será nuestro campo ‘objetivo’, es un campo binario que tiene dos valores: ‘Neutral or Dissatisfied’ y ‘Satisfied’. Del primer valor tenemos 73.452 líneas de clientes, del segundo 56.428. Se observa una distribución no muy desequilibrada entre ambas, lo que permite entrenar modelos sin necesidad de aplicar técnicas de balanceo como sobremuestreo o submuestreo.
* Respecto a valores nulos, solo se encontraron 393 en la columna ‘Arrival Delay in Minutes’.

## Procesado, limpieza, transformación y estructuración de los datos.

* Los valores nulos de la columna ‘Arrival Delay in Minutes’ son posiblemente vuelos sin retrasos registrado, por lo que se decidió rellenar los valores nulos con el cero. En general, el porcentaje de datos ausentes era bajo, lo que permitió trabajar con el conjunto original con mínimas modificaciones.
* La única columna del dataset que se elimino fue la columna ‘id’.
* Todas las variables categóricas, como ‘Type of Travel’ (‘Personal Travel’, ‘Business Travel’), fueron transformadas a formato numérico utilizando codificación, con el fin de que los modelos de aprendizaje automático pudieran procesarlas correctamente.

## Elección y justificación de modelos o técnicas de IA/ML.

Se eligió el modelo de aprendizaje supervisado, ya que el objetivo del proyecto es predecir una categoría específica – la satisfacción del cliente–, dada un conjunto de características observadas. Dado que la variable objetivo en los datos está etiquetada, ‘satisfaction’, el aprendizaje supervisado es la opción obvia.

Además, elegimos el algoritmo Random Forest porque tiene una alta precisión, ya que maneja bien tanto las variables categóricas como las numéricas, y ofrece una medida de la importancia de cada variable. Esto es importante para nuestro objetivo, ya que podemos identificar qué factor es el más influyente en la satisfacción del pasajero.

Accuracy Score:

|  |  |  |  |
| --- | --- | --- | --- |
| **Datos de entrenamiento** | **Datos normales** | **Datos escalados** | **Análisis de Componentes Principales**  **(95 % varianza)** |
| Random Forest Classifier | 95,58 | 95,54 | 91,63 |
| Decision Tree Classifier | 94,52 | 94,56 | 88,77 |
| KNeighbors Classifier | 64,2  (k = 20) | 93,0  (K = 9) | 82.0  (k = 20) |
| SVM Linear | - | 87,59 | 95,36 |

Viendo los resultados de los diferentes modelos se eligió Random Forest Classifier entrenado con datos no escalados.

## Resultados de Random Forest

El modelo obtuvo una precisión aproximada del 96%, lo cual indica una alta capacidad de acierto en la predicción del nivel de satisfacción de los pasajeros.

Otras métricas como Error Cuadrático Medio, Error Absoluto Medio o el Coeficiente de Determinación sugieren que el modelo también captura con eficacia la relación entre las variables de entrada y el objetivo, aunque por tratarse de un algoritmo de clasificación estas no son estrictamente necesarias, son más típicos de algoritmos de regresión:

Error Cuadrático Medio (mse): 0.04

Error Absoluto Medio (mse): 0.04

Coeficiente de Determinación (R2): 0.82

Matriz de Confusión:

Gráfico, Gráfico de rectángulos

El contenido generado por IA puede ser incorrecto.

Lo que más nos interesa del modelo resultante es como este realizo la predicción y que grado de importancia tuvo cada uno de los valores. Mediante el atributo ‘feature\_importances\_’ del modelo Random Forest podemos obtener la importancia relativa de cada variable (feature) en las predicciones del modelo, como se puede ver:

|  |  |
| --- | --- |
| **Característica** | **Importancia (%)** |
| Class | 16.25 |
| Online boarding | 14.55 |
| Inflight wifi service | 13.32 |
| Type of Travel | 7.81 |
| Inflight entertainment | 5.99 |
| Seat comfort | 4.12 |
| Ease of Online booking | 4.01 |
| Age | 3.49 |
| Customer Type | 3.44 |
| Flight Distance | 3.37 |
| Inflight service | 3.15 |
| Checkin service | 3.14 |
| Baggage handling | 2.78 |
| Cleanliness | 2.70 |
| Leg room service | 2.25 |
| On-board service | 2.23 |
| Gate location | 1.80 |
| Departure/Arrival time convenient | 1.47 |
| Arrival Delay in Minutes | 1.30 |
| Food and drink | 1.26 |
| Departure Delay in Minutes | 1.13 |
| Gender | 0.46 |

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

* Ya solo las tres primeras variables, ‘Class’, ‘Online boarding’ y ‘Type of Travel’, abarcan el 44 % del peso total del modelo. Sumándole las tres siguientes variables, ‘Type of Travel’ y ‘Inflight entertainment’, esta cifra sube al 58 %.
* A partir de la sexta variable, ‘Seat comfort’, el peso y la diferencia entre las variables va a disminuir, cada una rondando entre el dos y el cuatro por ciento del peso total.
* Los últimos valores, de ‘Departure/Arrival time convenient’ hacia abajo, no superan el 1,5%.

Con estos datos interpretar, a partir de los factores más importantes en el modelo, que elementos influyen en la más en la satisfacción de los pasajeros:

Tipo de cliente y tipo de billete:

La variable más importante fue la clase del billete (Class), que tuvo un peso del 16,25 %. Esto quiere decir que la clase en la que viaja la persona (Business o Economy, por ejemplo) influye mucho en si el pasajero está contento o no. Esto tiene sentido porque en clases más caras hay más comodidades, mejor trato, y todo en general es más cómodo, por lo que la gente suele estar más satisfecha.

También fue importante el tipo de viaje (Type of Travel), con un 7,81 %. Si una persona viaja por trabajo, normalmente quiere que todo funcione bien y rápido, sin retrasos, porque tiene cosas que hacer. Pero si viaja por vacaciones, puede estar más tranquilo. Por eso, el motivo del viaje cambia mucho cómo valora una persona la experiencia.

El tipo de cliente (Customer Type) también tuvo peso, con un 3,44 %. Esto diferencia entre clientes nuevos y los que vuelan a menudo. Los que ya conocen cómo funciona todo suelen estar más acostumbrados y, si las cosas salen bien, es más fácil que se queden fieles a la aerolínea. En cambio, los nuevos pueden sentirse más perdidos si no está todo claro.

Embarque y reservas online:

Una variable muy importante fue Online boarding, con un 14,55 %. Esto se refiere a cuando la persona consigue su tarjeta de embarque por internet, desde el móvil o el ordenador, sin pasar por el mostrador del aeropuerto. Esto es muy valorado porque la gente evita hacer colas, ahorra tiempo y se estresa menos. Hoy en día, mucha gente prefiere hacer las cosas por su cuenta desde su teléfono, y si el proceso es fácil y rápido, ya empieza bien el viaje.

Parecido a eso está Ease of Online booking, con un 4,01 %, que mide si es fácil reservar el billete por internet. Aunque no pesa tanto como lo del embarque, sigue siendo importante porque marca cómo empieza todo. Si la reserva online es complicada o da errores, la persona ya empieza con una mala impresión.

Conectividad durante el vuelo

Servicio wifi:

Otra variable que llamó la atención fue el wifi a bordo (Inflight wifi service), con un 13,32 %. Cada vez más personas quieren estar conectadas incluso cuando están volando. Algunos porque están trabajando, otros porque quieren hablar con familia o usar redes sociales. Si el wifi va mal o no funciona, eso puede afectar mucho la experiencia del vuelo, sobre todo para gente que lo necesita.

Conclusión

En general, los resultados muestran que lo más importante para los pasajeros no es solo lo que pasa dentro del avión, sino también cómo se sienten desde antes de volar. Cosas como la clase en la que viajan, si el viaje es por trabajo o por ocio, y si pueden hacer todo fácil desde el móvil (reserva y embarque) pesan mucho en cómo ven todo el servicio.

Las aerolíneas que inviertan en hacer más fácil todo el proceso digital, que ofrezcan wifi bueno y que cuiden más a los distintos tipos de clientes (frecuentes o nuevos), tendrán más posibilidades de dejar a la gente satisfecha y de que vuelvan a elegirlas. Porque hoy en día, el viaje no empieza cuando subes al avión, sino desde que compras el billete.