Grupo 4

CARLOTA SALAZAR, YUSHAN YANG, JULIA YI

PD1 - GIDIA

Grupo 4

EVENTPEAK



Contenido

[1. Contorno del proyecto 2](#_Toc158114894)

[Organización destinataria 2](#_Toc158114895)

[Objetivo de negocio de la organización 2](#_Toc158114896)

[Objetivo del sistema de aprendizaje automático 2](#_Toc158114897)

[Uso del sistema de aprendizaje automático 2](#_Toc158114898)

[Requisitos de rendimiento 2](#_Toc158114899)

[Requisitos legales 3](#_Toc158114900)

[2. Tarea de aprendizaje automático 3](#_Toc158114901)

[Definición de la tarea de aprendizaje automático 3](#_Toc158114902)

[Definir la variable objetivo 4](#_Toc158114903)

[Identificar los datos de entrada 4](#_Toc158114904)

[Fuente de datos principal y secundarias 4](#_Toc158114905)

[Factores que afecten al rendimiento 5](#_Toc158114906)

[3. Evaluación del Sistema 5](#_Toc158114907)

[Evaluación de éxito del sistema 5](#_Toc158114908)

[Evaluación de éxito de la tarea de aprendizaje automático 5](#_Toc158114909)

# Contorno del proyecto

## Organización destinataria

EventPeak es una empresa nueva que surge en el sector del entretenimiento y los eventos en vivo. Nos especializamos en la venta de entradas para una amplia variedad de eventos, incluyendo conciertos, festivales, obras de teatro, eventos deportivos y mucho más. Nuestra plataforma en línea proporciona a los usuarios acceso fácil y conveniente a una amplia gama de experiencias culturales y de entretenimiento en todo el mundo.

En EventPeak, trabajamos en estrecha colaboración con organizadores de eventos, promotores y sedes para ofrecer una selección diversa y emocionante de eventos para nuestros clientes.

Como nueva empresa en el mercado que somos, comenzaremos fijándonos únicamente en los conciertos y festivales de Estados Unidos. En EventPeak estamos constantemente buscando nuevas formas de innovar y mejorar nuestra plataforma y, por ello, incluiremos la implementación de un sistema de aprendizaje automático para estimar precios de entradas de manera más precisa, basándonos en multitud de datos diversos.

## Objetivo de negocio de la organización

El principal objetivo es optimizar la fijación de precios de las entradas para los eventos que se ofrecen en la plataforma. Mediante el análisis de múltiples variables, como la demanda histórica, la popularidad del artista o grupo, la ubicación, la fecha y hora del evento, entre otros, lograremos generar estimaciones de precios más precisas y ajustadas a las condiciones del mercado en tiempo real. Esto permitirá a EventPeak maximizar los ingresos por venta de entradas al mismo tiempo que garantiza una demanda óptima y una participación masiva en los eventos.

Otro objetivo igual de importante es mejorar la experiencia del cliente al ofrecer precios de entradas precisos y competitivos. Esto aumenta la satisfacción y la percepción de valor para los clientes, generando confianza, lealtad y relaciones comerciales a largo plazo en la industria del entretenimiento en vivo.

## Objetivo del sistema de aprendizaje automático

El objetivo del sistema de aprendizaje automático en EventPeak es ser capaz de analizar múltiples variables y patrones de datos históricos para generar estimaciones precisas de precios de entradas. Esto implica considerar factores como la demanda previa para eventos similares, la disponibilidad de entradas, la popularidad del artista o evento, la ubicación y la fecha del evento, entre otros.

## Uso del sistema de aprendizaje automático

El sistema de aprendizaje automático será utilizado por el equipo de precios y estrategia de EventPeak, compuesto por tres expertas en ingeniería de datos e inteligencia artificial: Carlota Salazar, Yushan Yang y Julia Yi. Su función principal será generar estimaciones precisas de precios de entradas basadas en análisis de datos históricos y variables relevantes. El equipo supervisará el rendimiento del sistema y realizará ajustes según sea necesario para garantizar estimaciones óptimas y competitivas en el mercado.

## Requisitos de rendimiento

Los requisitos de rendimiento para el sistema de EventPeak incluyen:

**Capacidad**: El sistema debe manejar un alto volumen de solicitudes de predicciones de precios de entradas, especialmente durante eventos populares y períodos de alta demanda. Se necesita capacidad para procesar una cantidad sustancial de predicciones por unidad de tiempo, especialmente durante los fines de semana y temporadas de vacaciones.

**Tiempo**: Es crucial que el sistema genere predicciones de precios lo más rápido posible para garantizar respuestas rápidas a las consultas de los usuarios. Se deben establecer métricas de tiempo objetivo para la generación de predicciones y optimizar continuamente el rendimiento del sistema para cumplir con estos requisitos.

**Alta demanda**: El sistema debe poder manejar picos repentinos de solicitudes. Se requieren medidas de escalabilidad para gestionar cargas de trabajo inesperadas sin comprometer la calidad del servicio. Durante períodos de baja demanda, el sistema debe poder escalar hacia abajo para optimizar recursos y reducir costos operativos.

**Disponibilidad y fiabilidad**: La disponibilidad del sistema es crucial para garantizar una experiencia sin interrupciones para los usuarios. Se deben implementar medidas de redundancia y recuperación ante desastres para mitigar interrupciones en el servicio. Es necesario establecer acuerdos de nivel de servicio para garantizar altos niveles de disponibilidad y fiabilidad del sistema.

## Requisitos legales

Al desarrollar un sistema de predicción de precios de entradas, EventPeak debe considerar requisitos legales y éticos, como:

**Privacidad de los datos del usuario**: Cumplir con regulaciones de privacidad, obtener consentimiento explícito para recopilar y procesar datos personales, implementar medidas de seguridad sólidas.

**Transparencia y equidad en la fijación de precios**: El sistema debe ser transparente sobre los factores que influyen en los precios, evitar sesgos injustos en la fijación de precios (no abusar en ello) y promover la equidad y accesibilidad.

**Cumplimiento normativo**: Cumplir con leyes y regulaciones relevantes, incluidas las relacionadas con la venta de entradas y la comercialización de eventos.

**Mitigación de sesgos sociales**: Implementar medidas para identificar y mitigar posibles sesgos en el sistema de predicción de precios.

**Cumplimiento de contratos y acuerdos**: Respetar contratos y acuerdos con proveedores, socios y clientes, asegurando que el sistema no viole acuerdos existentes.

# Tarea de aprendizaje automático

## Definición de la tarea de aprendizaje automático

La tarea de aprendizaje automático que parece más adecuada para el objetivo del sistema y los objetivos de la compañía de EventPeak es la regresión. La regresión es útil cuando se quiere predecir un valor numérico, como el precio de las entradas, que es el caso de EventPeak.

Permite entender la relación entre las variables predictoras (como la demanda histórica, la ubicación del evento, la popularidad del artista, etc.) y la variable objetivo (precio de la entrada). Proporciona una estimación cuantitativa de cómo los cambios en las variables predictoras afectan al precio de las entradas. Sin embargo, la regresión es sensible ante valores atípicos y ruido en los datos.

## Definir la variable objetivo

La variable objetivo para el sistema de aprendizaje automático de EventPeak es el "precio de las entradas". Esta variable representa el valor monetario que se espera que los clientes paguen por las entradas a los eventos que se ofrecen en la plataforma de EventPeak, y se medirá en dólares ya que la mayoría de los conciertos se sitúan en Estados Unidos. Los valores de esta variable pueden variar dependiendo del evento, la ubicación, la demanda, la disponibilidad de entradas y otros factores. Es probable que existan valores extremos para eventos de alta demanda o eventos exclusivos, así como eventos con precios más moderados. La distribución podría mostrar clases dominantes de eventos con precios promedio y clases infrarrepresentadas de eventos muy caros o baratos.

Los datos sobre el precio de las entradas se obtendrán de fuentes internas de EventPeak, incluidos registros de transacciones de venta de entradas, información proporcionada por organizadores de eventos y promotores, así como datos recopilados de eventos anteriores. Además, se pueden utilizar datos de mercado y de la competencia para contextualizar y validar los precios de las entradas.

## Identificar los datos de entrada

Los datos se obtendrán a partir de una multitud de características reunidas sobre las que se aplicará el sistema de aprendizaje automático para la predicción del precio de entrada. Algunas de estas variables podrían ser: el artista principal, su popularidad o reputación o tendencia actual (tipo de noticias dedicadas, número de oyentes en plataformas de música, número de premios musicales obtenidos…), la demanda histórica, la fecha, hora y lugar del evento, la capacidad y calidad del recinto, la disponibilidad de entradas, etc.

## Fuente de datos principal y secundarias

Para obtener datos reales para el proyecto de predicción de precios de entradas de conciertos, utilizaremos varias fuentes de datos como: API de Ticketmaster (competencia de EventPeak) para contextualizar y validar los precios de las entradas obteniendo la información principal; webs públicas sobre las que realizar web scrapping como los sitios web de estadios o de artistas con el fin de obtener la capacidad y calidad del estadio, información relevante sobre el artista, etc.; datos de RRSS y plataformas de reseñas como Instagram, Spotify, Youtube, Yelp, Google Trends… para obtener la popularidad del artista, entre otros; bases de datos de eventos históricos como last.fm con información sobre conciertos pasados, etc.

Hay algunas restricciones que debemos tener en cuenta para obtener datos en Ticketmaster, nuestra fuente de datos principal, ya que la cuota predeterminada es de 5000 llamadas a la API por día, con un límite de 5 solicitudes por segundo. Además, hay una limitación en la paginación profunda que solo permite recuperar hasta el elemento 1000, (solo se puede obtener un máximo de 1000 elementos multiplicando el tamaño por el número de página).

La demo del código para extraer datos de Ticketmaster se puede ver en el archivo <demo_codigo_tm.ipynb> o abriendo el enlace. Al ejecutar ese código y descargar los datos, se obtiene una muestra de 100 filas que puede visualizarse en el archivo <muestra_datos_ticketmaster.csv> o abriendo el enlace.

El resto de las fuentes son secundarias y servirán para completar la información que aporte Ticketmaster. En el demo de código adjuntado anteriormente, puede verse una prueba para extraer datos de Spotify y de last.fm. Los datos obtenidos de ambos sitios junto con los datos obtenidos de otros (todavía pendiente de hacer), se recopilarán y analizarán para potenciar y mejorar la precisión de las predicciones de precios de entradas.

## Factores que afecten al rendimiento

Pueden existir factores o cambios relevantes que alteren el rendimiento y, por tanto, haya que reentrenar el sistema, como, por ejemplo: cambios en la preferencia del público, variaciones estacionales o climáticas, cambios en la disponibilidad de datos, nuevos competidores o cambios en la industria, actualizaciones en políticas o regulaciones…

# Evaluación del Sistema

## Evaluación de éxito del sistema

La medida cuantitativa principal que EventPeak puede utilizar para validar el éxito del sistema es la precisión de las predicciones de precios de entradas. Esto se puede medir utilizando métricas como el error medio absoluto (MAE), el error cuadrático medio (MSE) o el coeficiente de determinación R2. Además, la organización puede considerar métricas indirectas como la tasa de conversión de ventas, la satisfacción del cliente o el incremento en los ingresos por venta de entradas.

EventPeak podría comparar su sistema con una referencia basada en precios de entradas previos sin aprendizaje automático y evaluar mejoras en la precisión. También se considerarían enfoques heurísticos, como la fijación de precios según la demanda histórica o costos. La evaluación se realizaría dividiendo datos en conjuntos de entrenamiento y prueba, utilizando métricas como MAE o MSE para comparar las predicciones del modelo con los precios reales. Durante el entrenamiento, estas medidas ayudan a ajustar el modelo para minimizar errores. Durante el uso del modelo, estas métricas permiten evaluar su capacidad predictiva en datos no vistos. Este análisis permitiría determinar la eficacia del sistema en comparación con otras alternativas en términos de precisión y valor para EventPeak.

## Evaluación de éxito de la tarea de aprendizaje automático

Las medidas de éxito del aprendizaje automático están estrechamente relacionadas con las metas del sistema y los objetivos de negocio de EventPeak. Un modelo preciso y confiable puede mejorar la fijación de precios de las entradas, maximizando los ingresos y la satisfacción del cliente. Por otro lado, un rendimiento deficiente podría llevar a decisiones de fijación de precios inexactas, impactando negativamente en la participación del evento y la percepción de la marca. En resumen, el éxito del aprendizaje automático influye directamente en el logro de los objetivos comerciales al proporcionar estimaciones precisas y competitivas de precios de entradas.