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)

Nota: lo que entregamos a continuación es la memoria reescrita y no directamente la estructura “comentario – respuesta al comentario”, ya que consideramos que los comentarios abarcan varios (por no decir muchos) puntos de la memoria anterior, como puntos con respuestas erróneas, puntos con respuestas *vacías*, respuestas en puntos equivocados… Por lo que todo o casi todo es una respuesta nueva en sí.

# 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.

Sin embargo, como nueva empresa que somos en el mercado, comenzaremos fijándonos únicamente en los conciertos y festivales de España. En EventPeak, confiamos en construir un proyecto sólido y ambicioso que establezca unos precios de entradas justos y competitivos para todos los eventos y, por ello, incluiremos la implementación de un sistema de aprendizaje automático.

## Objetivo de negocio de la organización

El objetivo de EventPeak es ser el nuevo portal de venta de entradas de confianza para todo el mundo. Una de las metas propuestas por EventPeak es alcanzar la reputación y ampliar las fronteras de su principal competidor (Ticketmaster, con gran importancia sobre todo en Europa), para convertirse en la principal empresa mundial de venta de entradas para diferentes tipos de espectáculos como entradas de cine, teatro, conciertos y eventos deportivos, entre otros.

El objetivo final es optimizar la fijación de precios de las entradas, que beneficien a la empresa maximizando sus ingresos y, a su vez, que satisfaga al mayor número de clientes posibles aumentando su satisfacción y mejorando su percepción sobre los precios.

## Objetivo del sistema de aprendizaje automático

Si bien EventPeak busca una magnífica reputación debido a una tarificación muy precisa, será de gran importancia para ello la implementación de un sistema de aprendizaje automático que se encargue de dicha precisión.

Sin lugar a duda, la incorporación de un sistema computacional de estas cualidades será la clave para lograr los objetivos de la empresa, pues, con precios más precisos y justos, se llegará a una mayor satisfacción y confianza por parte del público.

## 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- que tratarán de llevar este proyecto a buen puerto.

La función principal del sistema de AA será generar estimaciones precisas de precios de entradas basadas en análisis de datos 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

Casi todos los días del año sale al menos un concierto o evento a la venta, pero no se suele superar la cifra de aproximadamente 10 conciertos nuevos publicados en un día, aun así, en ningún caso, un volumen complejo de gestionar.

Por otro lado, la mayoría de los conciertos se produce en la temporada de verano, por lo que estaríamos en un pico, y la minoría de ellos se produce durante el invierno, probablemente a finales de año y principios del que sigue, por lo que se trataría de un valle. Identificar correctamente estos momentos del año tan dispares es importante ya que pueden ocasionar cambios abruptos en el volumen.

Por último, el tiempo que se tarda en generar la predicción es un requisito que no hace mucha falta tener en cuenta ya que el público esperará igualmente al momento de la apertura de compra de entradas, independientemente de que se produzca en una fecha o en otra relativamente próxima. Por esto mismo, el sistema puede operar con cierta latencia sin afectar significativamente la experiencia del usuario.

## Requisitos legales

Uno de los requisitos legales más importantes que EventPeak debe tratar es la transparencia a la hora de fijar los precios. En innumerables ocasiones, el dominio de una oligarquía dentro de un mercado hace que se establezcan tarifas abusivas de cara al público, sin oportunidad alguna de acudir a un sitio mejor donde se ofrezcan mejores condiciones, específicamente, económicas. No queremos que esto pase en EventPeak: no queremos que haya un abuso de mercado y por ello apostamos por la transparencia y equidad en la fijación de precios.

EventPeak debe cumplir con las leyes y regulaciones generales y específicas de cada país, por ejemplo, el Decreto 17/2019 en España, por el cual EventPeak tendrá que hacerse cargo de promover la accesibilidad universal y de reunir las condiciones de seguridad, de salubridad e higiene en el lugar de celebración del espectáculo, entre muchos otros puntos igualmente importantes.

# 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.

La regresión, aunque es sensible ante valores atípicos y ruido en los datos (para evitarlo, se realizará el preprocesado de los datos), permite entender la relación entre las variables predictoras (como 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.

## 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 euros ya que comenzaremos el negocio en España.

Aquí surge una posible duda, dado que hay varios tipos de eventos: los eventos con precio único fijado y los eventos con varios precios disponibles en función de, por ejemplo, el tipo de entrada. En cualquiera de los casos, la intención es calcular el rango del precio, es decir, el precio mínimo y el precio máximo (aunque para ello se requerirá la implementación de dos modelos).

Es probable que existan valores extremos para eventos de alta demanda o eventos exclusivos, así como eventos más modestos con precios más moderados.

Sacar rango de todos los precios mínimos y ver media y mediana de ellos para medir simetría. Lo mismo para el máximo.

## Identificar los datos de entrada

Cabe destacar que los datos con los que se entrenará el modelo son datos de eventos actualmente en venta y, este modelo entrenado, se probará su fiabilidad con nuevos eventos que surjan en los próximos dos o tres meses.

Los datos de entrada que se usarán para crear el modelo que prediga precios de entrada podrían ser, en un principio:

* El artista principal, nombre y apellido (original o artístico según se le conozca) y la popularidad relacionada (con información como el número de búsquedas en internet, el número de oyentes en plataformas musicales, el número de premios obtenidos…)
* Demanda actual (otros eventos en curso)
* Fecha, hora, ciudad o estado y país de realización del evento.
* Nombre del lugar o recinto donde se celebra, que categorizaremos como una de las tres posibles etiquetas: “estadio”, “teatro”, “sala”, según la capacidad o tamaño grande, mediano o pequeño, respectivamente.

## Fuente de datos principal y secundarias

Para obtener datos reales para el proyecto de predicción de precios de entradas de conciertos, utilizaremos, en un principio, las siguientes fuentes:

* API de Ticketmaster (competencia de EventPeak), donde se obtendrá la información base o principal.
* Webs de los recintos o webs donde encontrar una descripción de los mismos, con el fin de obtener la capacidad y así poder clasificarlo como “estadio”, “teatro” o “sala”.
* Información relevante sobre el artista: número de seguidores en RRSS, número de oyentes mensuales, número de búsquedas en internet… En un principio, con número de oyentes y de búsquedas debería valer, pero no descartamos otros para completar.

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.

La demo del código para extraer los datos íntegros de Ticketmaster, el número de oyentes mensuales en Spotify de un artista y el porcentaje de búsquedas por comunidad autónoma del artista, se puede ver en el archivo <codigo_entrega_1.ipynb> o abriendo el enlace. Al ejecutar el código de Ticketmaster y descargar los datos, se obtiene una muestra de 767 filas que puede visualizarse en el archivo <datos_19_02.csv> o abriendo el enlace (aunque, mejor, abrir y visualizar directamente ejecutando el código de Ticketmaster). Cada día se suben entre 2 o 5 conciertos más, por lo que iremos recopilando nuevos datos para entrenar o probar el modelo.

Los datos extraídos de Ticketmaster incluyen los atributos: nombre del evento con nombre del artista, tipo del evento (nos quedaremos con los conciertos), localización, fechas y horas de realización, fechas y horas de la venta, el género o subgénero del artista, promotores, rango de precios (con o sin gastos de gestión, ambas opciones), información sobre si está a la venta o no el evento, etc.

Algunas de estas variables no nos serán de mucha utilidad, pero las que sí sean importantes, las recopilaremos junto a otras variables de otras extracciones de datos también importantes para desarrollar nuestro proyecto.

## 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

Para evaluar que el sistema tiene éxito o es fiable, compararemos los precios estimados por él con los precios publicados por Ticketmaster, para cada uno de los eventos.

Además, podríamos establecer una heurística que pueda realizar la tarea de forma aproximada y que también sirva para comparar con los precios estimados, por ejemplo, partiendo del precio medio que suele fijar una sala concreta para sus eventos o, también, el precio medio que suele fijar un artista para sus eventos.

Por otro lado, también podríamos evaluar el éxito del sistema indirectamente observando la tasa de ventas, la satisfacción del cliente o el incremento en los ingresos de EventPeak.

## Evaluación de éxito de la tarea de aprendizaje automático

Para evaluar el rendimiento de la tarea de aprendizaje automático, ya sea durante el entrenamiento o durante el uso del mismo, podremos utilizar métricas como la raíz del error cuadrático medio (RECM) o el coeficiente de determinación R2 ajustado.

Con estas métricas podremos medir el rendimiento, es decir, qué tan bien se ajusta el modelo a las observaciones reales que tenemos: un error relativamente pequeño significará un sistema muy preciso para fijar los precios, impactando así positivamente a la comunidad y aportando más confianza a los clientes con el servicio que ofrece EventPeak.

En resumen, el éxito del aprendizaje automático influye directamente en el logro de los objetivos comerciales y, por tanto, contribuye al éxito de la empresa.