



PREDICCIÓN DE ASISTENCIA A FESTIVALES EN ARGENTINA

PREDICCIÓN DE ASISTENCIA A FESTIVALES EN ARGENTINA

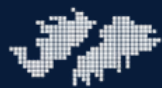
MATERIA: APRENDIZAJE AUTOMÁTICO
ESTUDIANTE: MARIANA ANAHÍ LÓPEZ
TECNICATURA SUPERIOR EN CIENCIA
DE DATOS E INTELIGENCIA ARTIFICIAL



CENTRO POLITÉCNICO SUPERIOR
MALVINAS ARGENTINAS

2025

Aplicación de técnicas de regresión supervisada sobre datos del Catálogo Nacional de Fiestas y Festivales (SInCA)



Predicción de la cantidad de asistentes a festivales y fiestas populares en Argentina

Contexto del Problema

En Argentina, los festivales y fiestas populares constituyen una parte fundamental de la vida cultural y social. Estos eventos, que se desarrollan en distintas regiones del país, atraen a miles de visitantes y contribuyen de manera significativa a la economía local a través del turismo, la gastronomía y el comercio.

Sin embargo, no siempre se dispone de información precisa sobre la cantidad de asistentes que participan en cada evento. La asistencia depende de múltiples factores, como la ubicación geográfica, la temática, la temporada del año o la organización del evento.

Predecir esta variable puede ser útil para la planificación de recursos, la promoción turística y la evaluación del impacto económico y cultural de las celebraciones.

En este contexto, el proyecto busca aplicar técnicas de Aprendizaje Automático para estimar la cantidad aproximada de asistentes a los festivales registrados en el Catálogo Nacional de Fiestas y Festivales (SInCA), utilizando las características descriptivas de cada evento.

Objetivo general

Desarrollar un modelo de regresión supervisada que permita predecir la cantidad aproximada de asistentes a un festival o fiesta popular en Argentina, a partir de variables culturales y organizativas contenidas en el dataset del SInCA.

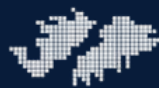
Objetivos específicos

- Analizar el dataset del Catálogo Nacional de Fiestas y Festivales, identificando las variables relevantes y los valores faltantes.
- Realizar un proceso de limpieza y codificación de las variables categóricas para su uso en modelos de regresión.
- Entrenar y comparar distintos modelos supervisados de regresión (lineal y no lineal).
- Evaluar el desempeño de los modelos utilizando métricas de error y determinar cuál ofrece mejores resultados.
- Aplicar el modelo seleccionado para estimar los valores faltantes de asistencia y analizar los resultados obtenidos.

Tipo de problema

El problema se enmarca en el Aprendizaje Supervisado, ya que la variable objetivo (cantidad aproximada de asistentes) se encuentra presente en el dataset para una parte de los registros.

Dado que se trata de una variable numérica continua, el proyecto corresponde a un problema de regresión.



Modelos propuestos:

Para abordar el problema de predicción de la cantidad de asistentes, se propone utilizar distintos modelos de regresión supervisada que permitan analizar las relaciones entre las variables del dataset y estimar un valor numérico continuo.

Los modelos considerados son:

- Regresión Lineal Múltiple: se utilizará como modelo base para evaluar la relación entre las variables independientes y la cantidad aproximada de asistentes.
- Random Forest Regressor: se incorporará como modelo alternativo para capturar relaciones no lineales y posibles interacciones entre las variables, ya que este tipo de modelo basado en árboles de decisión ofrece mayor flexibilidad ante datos categóricos codificados y relaciones complejas.

Ambos modelos serán comparados en función de su desempeño mediante métricas de error (MAE, RMSE y R^2), con el fin de determinar cuál se ajusta mejor a los datos y ofrece predicciones más precisas.

Dataset utilizado:

El dataset proviene del Catálogo Nacional de Fiestas y Festivales del Sistema de Información Cultural de la Argentina (SInCA). Obtenidos de la página: <https://datos.cultura.gob.ar/>

Contiene aproximadamente 1.000 registros de festivales en todo el país y las siguientes variables principales:

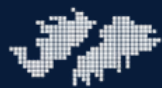
- Provincia
- Temática principal
- Mes de realización
- Duración (días)
- Tipo de entrada (paga o gratuita)
- Tipo de gestión (pública, privada o mixta)
- Periodicidad del evento
- Cantidad aproximada de asistentes (variable objetivo)

Alrededor del 50% de los registros no poseen valor en la variable objetivo, lo que representa un desafío adicional que será abordado mediante estrategias de imputación o modelado en etapas.

Relevancia del trabajo:

Este proyecto es relevante porque propone aplicar técnicas de Aprendizaje Automático a un ámbito poco explorado, como el cultural y turístico. Los festivales y fiestas populares tienen una gran importancia en Argentina: fortalecen la identidad de cada comunidad, impulsan la actividad turística y generan un impacto económico significativo, sobre todo en las provincias del interior.

Poder estimar la cantidad de asistentes a estos eventos a partir de sus



características resulta útil para la planificación y la gestión. Los gobiernos locales, las organizaciones culturales y los productores pueden anticipar la demanda, preparar mejor la infraestructura y diseñar estrategias de promoción más eficientes.

Además, el trabajo tiene valor formativo, ya que permite poner en práctica los conocimientos adquiridos en la materia a partir de un caso real, con desafíos propios del trabajo con datos abiertos: valores faltantes, variables categóricas y la necesidad de elegir el modelo más adecuado. En conjunto, busca mostrar cómo la ciencia de datos puede convertirse en una herramienta concreta para analizar fenómenos culturales y apoyar la toma de decisiones basadas en información