

PROYECTO FINAL

Daniel Felipe Cerquera Idrobo – Jose Andres Cardona Rojas

Conclusiones.

El trabajo desarrollado con el conjunto de datos del Titanic permitió recorrer de forma completa cada fase del proceso de análisis de datos, siguiendo los once pasos planteados en el proyecto. Para empezar, se realizó la carga de las librerías necesarias y del archivo utilizado, el cual proviene de una base pública muy conocida. Después se llevó a cabo una exploración inicial que permitió identificar el tamaño del dataset, los tipos de variables presentes y algunas características básicas de los registros. Más adelante se avanzó hacia un análisis exploratorio más profundo, que facilitó la detección de patrones, distribuciones destacadas y algunos valores atípicos que era necesario revisar. Esto condujo a la etapa de limpieza, en la que se trataron los datos faltantes y se corrigieron pequeñas inconsistencias. Luego se aplicaron las transformaciones necesarias, además de la codificación de variables y el escalamiento correspondiente para que el modelo pudiera procesarlas correctamente. Finalmente, se construyó un modelo predictivo, se evaluó su comportamiento y se interpretaron sus resultados para dar respuesta a la pregunta central del proyecto.

Durante el análisis exploratorio surgieron hallazgos clave que ayudaron a comprender mejor la dinámica de los datos. Una de las relaciones más evidentes fue el impacto del género en la probabilidad de supervivencia: las mujeres sobrevivieron en mayor proporción que los hombres, lo cual está alineado con los protocolos de evacuación que se aplicaron en ese momento histórico. Otro patrón importante fue la diferencia marcada entre las clases del barco. Los pasajeros de primera clase presentaron tasas de supervivencia superiores, lo que sugiere ventajas en el acceso a zonas de evacuación y botes salvavidas. La edad también mostró un efecto relevante: los niños pequeños, especialmente los menores de 10 años, tuvieron más probabilidades de sobrevivir. Además, variables como el valor del pasaje (Fare) y el número de familiares a bordo (SibSp y Parch) aportaron información importante al análisis y mostraron que la dinámica social del barco tuvo influencia directa en los resultados. En conjunto, los gráficos y estadísticas confirmaron tendencias muy claras que sirvieron como base para el modelado predictivo.

La pregunta central del proyecto buscaba identificar los factores que influyeron en la supervivencia de los pasajeros y determinar si un modelo de machine learning podía anticipar dicho comportamiento. Con base en los resultados obtenidos, la respuesta es positiva. Tras preparar adecuadamente el conjunto de datos, el modelo fue capaz de generar predicciones con un nivel de exactitud razonable y útil para el propósito del estudio. Las variables que más contribuyeron al proceso predictivo fueron el género, la clase, la edad y la tarifa del boleto, lo cual coincide con las observaciones del análisis exploratorio. Esto confirma que la supervivencia no dependió del azar, sino que estuvo condicionada por características sociales, demográficas y económicas. En términos prácticos, el modelo demostró que es posible estimar con buena aproximación si un pasajero hubiese sobrevivido o no, basándose únicamente en su perfil dentro del dataset.

Al interpretar estos resultados en el marco histórico del suceso, queda claro que los factores sociales y culturales de la época jugaron un papel esencial. La mayor supervivencia de mujeres y niños refleja la aplicación, al menos parcial, del conocido principio de evacuación que priorizaba a

estos grupos. Por otro lado, el hecho de que los pasajeros de mayor estatus económico tuvieran más oportunidades de sobrevivir evidencia desigualdades estructurales propias del contexto social del siglo XX. La edad, nuevamente, se convirtió en un factor determinante, pues los niños fueron considerados prioritarios al momento del rescate. Como consecuencia, los resultados del análisis no solo muestran patrones matemáticos, sino que también permiten entender comportamientos humanos, decisiones logísticas y condiciones sociales que terminaron influyendo en la tragedia del Titanic.

A pesar de todo lo logrado en el análisis, es importante reconocer las limitaciones del estudio. El dataset utilizado, aunque ampliamente difundido en entornos educativos, no incluye toda la información histórica y contextual relacionada con el caso. Por ejemplo, no se cuenta con datos sobre la ubicación exacta de cada pasajero en el barco, su condición física al momento del impacto o la rapidez con la que pudieron acceder a las zonas de evacuación. Además, fue necesario imputar valores faltantes, lo cual introduce cierto grado de incertidumbre. Desde el punto de vista técnico, el modelo empleado podría perfeccionarse mediante técnicas más avanzadas, un ajuste más fino de los hiperparámetros o la incorporación de algoritmos más sofisticados. La capacidad predictiva también está limitada por las variables disponibles, ya que el modelo no puede analizar información que no exista dentro del conjunto de datos original.

No obstante estas limitaciones, el proyecto deja planteadas varias recomendaciones útiles para análisis futuros. Sería conveniente trabajar con datasets más completos o complementar el actual con información adicional proveniente de investigaciones históricas. También sería interesante comparar distintos modelos de machine learning, como random forest, técnicas de boosting o incluso redes neuronales, para verificar si se puede aumentar el nivel de precisión obtenido. Otra recomendación sería utilizar métodos de imputación más complejos que reduzcan el impacto del manejo de datos faltantes. Asimismo, se podría profundizar en el análisis segmentando a los pasajeros por grupos específicos para identificar perfiles de supervivencia más detallados. Finalmente, incorporar validaciones cruzadas más robustas y análisis de sensibilidad ayudaría a fortalecer los resultados y hacerlos más confiables.

En conclusión el proyecto fue una oportunidad valiosa para comprender a fondo cada etapa del proceso de análisis de datos, desde la revisión inicial del dataset hasta la construcción e interpretación del modelo predictivo. Los hallazgos permitieron responder con claridad a la pregunta de investigación y aportaron una visión amplia sobre los factores que influyeron en la supervivencia de los pasajeros del Titanic. De igual forma, el estudio demostró la importancia de considerar tanto la técnica analítica como el contexto histórico para comprender mejor los resultados obtenidos. Aunque el dataset y el modelo presentan limitaciones, estas abren puertas a investigaciones futuras y enfatizan que el análisis de datos es un proceso que se perfecciona constantemente. En conjunto, las conclusiones reflejan un aprendizaje profundo y significativo sobre exploración, preparación, modelado y evaluación, al tiempo que permiten entender con mayor detalle uno de los episodios marítimos más recordados de la historia.