

Informe Final: Predicción del Aterrizaje de la Primera Etapa del Falcon 9 de SpaceX

Autora: Marcia Katherine Leal Normand

Afiliación: Universidad Andrés Bello

Contacto: marcia.leal@email.com

Fecha: 12 de mayo de 2025

Tabla de Contenido

1. Resumen Ejecutivo
2. Introducción
3. Metodología
4. Análisis Exploratorio de Datos (EDA)
5. Modelado Predictivo
6. Resultados y Discusión
7. Conclusión
8. Referencias
9. Apéndices

1. Resumen Ejecutivo

Este informe presenta un análisis completo del historial de lanzamientos del cohete Falcon 9 de SpaceX, enfocado en predecir si la primera etapa logra aterrizar exitosamente. A través de técnicas de ciencia de datos que incluyen análisis exploratorio, visualización interactiva (Folium y Dash), y clasificación mediante modelos de Machine Learning, se construyó un modelo predictivo con alta precisión. Los datos fueron procesados y estandarizados para mejorar la interpretabilidad y rendimiento del modelo. Se aplicó validación cruzada y selección de hiperparámetros. Finalmente, se presentan recomendaciones y oportunidades futuras de mejora.

2. Introducción

SpaceX ha revolucionado la industria aeroespacial mediante el uso de cohetes reutilizables. La capacidad de predecir si un aterrizaje será exitoso tiene valor estratégico para reducir costos y planificar operaciones. Este informe responde a la pregunta: ¿Es posible predecir con precisión si la primera etapa del Falcon 9 aterrizará exitosamente usando datos históricos?

3. Metodología

- Fuente de datos: Dataset público proporcionado por IBM Skills Network (CSV)
- Procesamiento: Limpieza de datos nulos, normalización (StandardScaler), creación de variables dummy
- División de datos: 80% entrenamiento, 20% prueba
- Modelos usados: Regresión logística, SVM, árbol de decisión
- Visualizaciones: Seaborn, Plotly Dash, mapas Folium
- Evaluación: Accuracy, matriz de confusión, GridSearchCV

4. Análisis Exploratorio de Datos (EDA)

- Visualizaciones clave:
- FlightNumber vs Class por sitio de lanzamiento
- PayloadMass vs Class por tipo de órbita
- Tasa de éxito por tipo de órbita
- Relación entre variables categóricas (LandingPad, Serial, Orbit)
- Mapas interactivos con Folium:
- Ubicación de todos los sitios de lanzamiento
- Distancia a ciudad, autopista, costa, ferrocarril

5. Modelado Predictivo

- Regresión logística: mejor precisión con $C=1$, solver='lbfgs', penalty='l2'
- SVM: mejor kernel: linear
- Árbol de decisión: Precisión en test: 83.33%
- Evaluación final: Regresión logística obtuvo mejor rendimiento general en validación cruzada

6. Resultados y Discusión

- El modelo predictivo permite anticipar con buena precisión si una misión logrará aterrizar la primera etapa
- Los mapas Folium revelan que todos los sitios de lanzamiento están cerca de la costa y alejados de ciudades
- Las variables más importantes en la predicción fueron: tipo de órbita, sitio de lanzamiento, número de vuelos y masa de carga
- Las visualizaciones con Dash permiten interactuar por rango de carga y sitio seleccionado

7. Conclusión

Este proyecto demostró que es viable predecir el éxito del aterrizaje de la primera etapa de un cohete Falcon 9 con datos históricos. Los modelos desarrollados ofrecen precisiones aceptables y las visualizaciones interactivas agregan valor interpretativo. A futuro se podría integrar información meteorológica o de sensores técnicos para mejorar el rendimiento.

8. EDA con Resultados SQL

A continuación se presentan ejemplos de consultas SQL aplicadas sobre el dataset cargado en SQLite, junto con Cambria (Cuerpo) y sus hallazgos clave:- Cantidad total de lanzamientos a órbita geosíncrona (GTO): `SELECT COUNT(*) FROM SPACEXTBL WHERE Orbit = 'GTO';` Resultado: 10 lanzamientos a GTO- Payload máximo registrado: `SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEXTBL;` Resultado: 15600 kg- Promedio de payload para cada versión del booster: `SELECT Booster_Version, AVG(PAYLOAD_MASS_KG_) AS Promedio_Masa FROM SPACEXTBL GROUP BY Booster_Version;` Resultado: permite identificar qué versiones transportan mayor carga- Recuento de resultados de misión por sitio de lanzamiento: `SELECT Launch_Site, COUNT(Mission_Outcome) FROM SPACEXTBL GROUP BY Launch_Site;` Resultado: se observa mayor frecuencia de lanzamientos desde CCAFS SLC-40- Órbitas con mayor tasa de éxito: `SELECT Orbit, AVG(CAST(Class AS FLOAT)) AS Tasa_Exito FROM SPACEXTBL GROUP BY Orbit` Informe Final - SpaceX Falcon 9 `ORDER BY Tasa_Exito DESC;` Resultado: las órbitas ISS y LEO presentan las tasas más altas Estos resultados fueron utilizados para confirmar tendencias identificadas en el EDA gráfico y orientar la construcción del modelo predictivo.

9. Referencias

- IBM Skills Network. (2025). SpaceX Falcon 9 Dataset. [CSV]
- McKinsey & Company. (2024). The business case for reusability in space
- Scikit-learn Documentation. <https://scikit-learn.org/>
- Plotly Dash Docs. <https://dash.plotly.com/>

10. Apéndices

- Figura 1: Mapa interactivo con sitios de lanzamiento (Folium)
- Figura 2: Dashboard con selección de carga y sitio (Dash)
- Código fuente en GitHub: https://github.com/tu_usuario/spacex-landing-prediction.