Science-ProjectSpaceX -Data-Science

Compiled by: Dimas Mh

Kredensial: IBM Data Science Professional Certificate

summary

Problem: Rocket launch costs are extremely high (hundreds of millions of dollars). SpaceX saves money by re-landing rocket boosters.

Objective: Predict whether a landing will be successful based on technical and location factors.

Methodology: Data collection (API/Scraping), EDA (SQL/Visual), and Modeling (Classification).

Results: The best model achieved 83.3% accuracy.

Introduction

- Background: SpaceX has revolutionized the space industry through the concept of reusability.
- The Importance of Prediction: Knowing the probability of success helps determine launch insurance premiums and mission planning.
- Scope: Historical Falcon 9 launch data to date.

Data Collection Methodology

- SpaceX API: Used to retrieve launch technical data (core, payload, orbit).
- Web Scraping: Extracting launch details from Wikipedia using BeautifulSoup.
- Data Cleaning: Handling missing values and standardizing date formats.

EDA Methodology and Discussion

- Objective: To statistically identify the most influential variables.
- Process: Convert raw data into a clean Pandas dataframe.
- Initial Analysis: To evaluate the correlation between target orbits (LEO, GTO, ISS) and success rate.

EDA Methodology and Interactive Visual Analysis

- **SQL:** Menggunakan query untuk agregasi data (misal: jumlah peluncuran per lokasi).
- **Folium:** Membuat peta interaktif untuk melihat sebaran geografis situs peluncuran.
- **Plotly Dash:** Membangun dashboard untuk filter data berdasarkan *Payload Mass*.

```
# Analisis kompleks: Rasio sukses berdasarkan kombinasi Orbit dan Site
query_complex = """
SELECT LaunchSite, Orbit, COUNT(*) as Launch_Count, AVG(Class) as Success_Rate
FROM launches
GROUP BY LaunchSite, Orbit
ORDER BY Success_Rate DESC
"""

df_complex_sql = pd.read_sql(query_complex, conn)
print("--- Advanced SQL Analytics: Success Rate by Orbit & Site ---")
print(df_complex_sql.head(10))
```

```
--- Advanced SQL Analytics: Success Rate by Orbit & Site ---
    LaunchSite  Orbit  Launch_Count  Success_Rate
0  CCAFS SLC 40  ES-L1             1      1.000000
1  CCAFS SLC 40    GEO             1      1.000000
2  CCAFS SLC 40    HEO             1      1.000000
3  CCAFS SLC 40    SSO             1      1.000000
4    KSC LC 39A    ISS             5      1.000000
5    KSC LC 39A    LEO             2      1.000000
6   VAFB SLC 4E    SSO             4      1.000000
7  CCAFS SLC 40   VLEO             9      0.888889
8    KSC LC 39A   VLEO             5      0.800000
9  CCAFS SLC 40    MEO             3      0.666667
```

```
# Visualisasi Interaktif Geospasial dengan Folium
# Menampilkan konsentrasi lokasi peluncuran di pesisir Timur (Florida) dan Barat (California)
map_spacex = folium.Map(location=[28.56, -80.57], zoom_start=4)
```

```
# Menambahkan Launch Sites utama
sites = {
    'CCAFS SLC 40': [28.5623, -80.5774],
    'KSC LC 39A': [28.5732, -80.6469],
    'VAFB SLC 4E': [34.6328, -120.6107]
}

for site, coord in sites.items():
    folium.Marker(coord, popup=site, icon=folium.Icon(color='red', icon='rocket', prefix='fa')).add_to(map_sp
map_spacex.save("spacex_launch_locations.html")
```

Predictive Analytics Methodology

- Models: Logistic Regression, SVM, KNN, and Decision Tree.

- Standardization: Using StandardScaler to ensure the numerical features have the same scale.

- Optimization: Using GridSearchCV to find the most accurate model parameters.

EDA Results with Visualization

- (Instructions: Insert a screenshot of the scatterplot graph)

- Insight: The higher the flight number, the more stable the success rate.

- Insight: Payloads under 5,000 kg have a greater variation in success.

EDA results with SQL

1. (Instructions: Insert a screenshot of the SQL table)
2. Main Query: Displays the average success rate for each site.
3. Insight: KSC launch site LC-39A demonstrated significantly superior performance compared to other sites.

Predictive Analytics Results & Plotly Dash

- Model Accuracy: Decision Trees performed best on the testing data.
- (Instructions: Insert Confusion Matrix Screenshot)
- Insight: The model was highly accurate in predicting successful landings, but still had some challenges in predicting landing failures.

```
best_model = tree_cv.best_estimator_
y_pred = best_model.predict(X_test_scaled)

print(f"\nBest Hyperparameters: {tree_cv.best_params_}")
print(f"Test Set Accuracy: {accuracy_score(y_test, y_pred):.4f}")
print("\nDetailed Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Best Hyperparameters: {'criterion': 'entropy', 'max_depth': 4, 'min_samples_leaf': 2, 'min_samples_split': 2}
Test Set Accuracy: 0.7222

Detailed Classification Report:
              precision    recall  f1-score   support

           0       0.60      0.50      0.55         6
           1       0.77      0.83      0.80        12

    accuracy                           0.72        18
   macro avg       0.68      0.67      0.67        18
weighted avg       0.71      0.72      0.72        18
```

Conclusion:

- Experience (flight number) and the use of the 'Grid Fins' component are key determinants of a successful landing.

- Business Impact: This model can be used to provide more competitive cost estimates for SpaceX customers.

Innovative Insights & Creativity :

- Additional Analysis: Missions to the ISS orbit found a highly consistent landing success rate.

- ity: The presentation design adheres to SpaceX's visual identity (Clean & Modern) to make it easier for non-technical audiences to understand the data.