Final Report

Space Y

SpaceX Launch Analysis

1. Project Introduction:

SpaceX offers a launch service using the Falcon 9 rocket at a cost of $62 million, which is much lower than its competitors, whose launch costs reach $165 million. The main reason for this savings is the reuse of the first stage of the rocket.

In this project, our goal is to analyze the public data available on Falcon 9 launches to predict the success of the first stage landing, and to verify the accuracy of the company's official statements.

2. Project Objective:

- Predict the success of the launch of SpaceX rockets based on available historical data, which helps in improving future launches.

- Improve risk management and better resource allocation.

3. Project Methodology:

We used data science methodologies such as:

- Data collection via techniques such as API web scraping .

- Exploratory Data Analysis (EDA).

- Predictive modeling using machine learning algorithms.

- Deploying models using techniques such as MLOps , Gradio and MLflow.

4. Tools and techniques:

- Python for data analysis and model development.

- Scikit-learn for model selection and prediction.

- GridSearchCV for model optimization.

- MLflow for experiment tracking.

- Gradio for creating an interactive interface for users.

5. Project steps:

5.1. Data collection and exploration (week 1):

- **Data Collection:**
  Data was obtained from SpaceX API and other sources such as Wikipedia via web scraping.
- **Data Exploration:**
  - **Load the Data**
  - **Understand the Structure**
    Check the shape , View the first few rows & Check the data types
- **Deliverables**:
  Data included important variables such as Launch date, booster type, launch location, payload mass, and Mission outcome (success/failure).
  Target variable was the success or failure of the launch (1 = success, 0 = failure).

5.2. Data Analysis and Visualization (week 2):

- **Data Cleaning:**
  - **Handling missing values** using techniques such as Replacing missing values with mean for some columns such as Payload Mass.
  - **Filter out rows** with Falcon 9
  - Final storage of the cleaned data in CSV files.
- **Data Analysis:**
  - Count the number of launches by site and orbit, and gather landing outcomes.
  - Identify bad landing outcomes.
  - Create a binary classification based on landing outcomes.
  - Calculate overall success rate.

- **Data Visualization**:
  - Analyzing the relationship between Payload Mass and Flight Number using scatter plots.
  - Analyzing different Launch Sites and the success of launches for each site.
  - Evaluating the relationship between orbit type and mission success.
- **Tools**: Python (Matplotlib, Seaborn), Power PI.
- **Deliverables**:
  - A cleaned dataset and a saved as CSV file.
  - Notebook with Interactive visualizations and dashboards displaying key insights.

5.3. Machine Learning Model Development (week 3):

- **Model Selection:** We tested several models including:
  - **Decision Tree**
  - **K-Nearest Neighbors (KNN)**
  - **Logistic Regression**
  - **Support Vector Machine (SVM)**
- **Model Training:**
  - The models were evaluated using Confusion Matrix.
  - The accuracy of each model was compared, and the optimal parameters for each model were saved in JSON files.
- **Model Optimization:**
  - We also used **GridSearchCV** with **StratifiedKFold** to perform hyperparameter tuning, ensuring the models are tested with various configurations for optimal performance.
- **Deliverables**:
  - Notebook on model training and evaluation with details on model performance,.

5.4. MLOps, Deployment, and Final Presentation (week 4):

- **MLOps Implementation:**
  - For tracking our experiments, we utilized **MLflow**. This tool helps us log crucial information, such as:-
    hyperparameters, model accuracy, and the **best-performing models**.
  - It enables us to compare model performance and experiment results effectively.
- **Deployment**:
  - We used **Gradio** to build an interface where users can input mission details to receive predictions in real-time.
- **Deliverables**:
  - **MLflow** runs folder
  - **Gradio Interface**:Simple web interface where users provide:
    - Mass Payload (kg)
    - Orbit (LEO, GTO, ISS, etc.)
    - Launch Site (KSC LC 39A, etc.)
    - Predicts "Success" or "Failure" based on the input

6. Results and Recommendations:

- We were able to build an accurate model to predict launch success.
- We were able to improve the model using techniques such as GridSearchCV, and achieved good predictive results.
- Through the interactive application created, users can easily predict the success of any launch based on the entered data.

7. Future Improvements:

- Using more advanced models to improve accuracy.
- Collecting more recent data to improve the effectiveness of predictions, especially with the continuous changes in technology at SpaceX.

8. Team Roles:

- Ahmed Nashat: Collecting and cleaning data using techniques such as web scraping.
- Ahmed Nagi: Exploratory and statistical analysis of data.
- Awab Khalil: Designing graphs and visual analysis of results.
- Suhaila Arabi: Developing predictive models using machine learning algorithms.
- Aya Reda: Deploying models and managing the process using MLOps and Gradio techniques.

9. Conclusion:

In this project, we were able to build an effective model to analyze SpaceX launch data and predict landing success. This model can be a useful tool to improve future launches, with the possibility of further improvement and development in the future.