**Capstone Project: Predicting Falcon 9 First Stage Landing Success**

This project aims to analyze SpaceX Falcon 9 landing outcomes and develop a machine learning model to predict whether the first stage will successfully land. The project involves data collection, preprocessing, exploratory data analysis, feature engineering, model selection, training, and evaluation.

**Project Workflow and Steps**

Here is a step-by-step breakdown of the entire process:

**Phase 1: Project Understanding and Planning**

1. **Define the Problem Statement**
   * Understand the significance of Falcon 9 reusability and cost savings.
   * Define the problem as a binary classification: **Will the first stage land successfully?** (Yes/No)
2. **Define Project Goals & Objectives**
   * Analyze the factors that contribute to a successful landing.
   * Develop a predictive model using machine learning.
   * Interpret model results in gaining insights into feature importance.
3. **Gather Requirements**
   * Programming languages: **Python**
   * Tools: **Jupyter Notebook, Pandas, Scikit-Learn, Matplotlib, Seaborn, SQL, TensorFlow/Keras (if needed)**
   * Data sources: **SpaceX API, Wikipedia, Historical flight data, NASA, Kaggle datasets**

**Phase 2: Data Collection and Preprocessing**

1. **Data Collection**
   * Use **SpaceX API** to collect Falcon 9 launch data.
   * Scrape additional data from **Wikipedia** and other sources.
   * Collect datasets on **weather conditions**, **landing site coordinates**, etc.
2. **Store and Organize Data**
   * Convert collected data into structured **CSV files** or store them in a database.
   * Load data using **Pandas**.
3. **Data Cleaning and Handling Missing Values**
   * Remove irrelevant columns.
   * Handle missing values (impute or drop).
   * Convert categorical values into numerical representations.
4. **Exploratory Data Analysis (EDA)**
   * Perform **summary statistics** and visualization using:
     + **Histograms**
     + **Box plots**
     + **Scatter plots**
     + **Correlation matrices**
   * Identify key factors affecting landing success:
     + **Launch site**
     + **Payload mass**
     + **Rocket version**
     + **Weather conditions**

**Phase 3: Feature Engineering and Data Transformation**

1. **Feature Selection**
   * Identify which features impact landing success.
   * Remove redundant or highly correlated features.
2. **Feature Engineering**
   * Create new features from existing ones.
   * Convert categorical features using **one-hot encoding**.
   * Normalize numerical values using **MinMaxScaler** or **StandardScaler**.
3. **Split the Dataset**

* Divide data into **training (80%)** and **testing (20%)** sets.

**Phase 4: Model Development**

1. **Choose Machine Learning Models**

* Try different models to find the best fit:
  + **Logistic Regression**
  + **Random Forest**
  + **Decision Trees**
  + **Support Vector Machine (SVM)**
  + **Neural Networks (optional)**

1. **Train the Model**

* Use training data to fit models.
* Tune hyperparameters to optimize performance.

1. **Evaluate Model Performance**

* Use metrics such as:
  + **Accuracy**
  + **Precision, Recall, F1-score**
  + **ROC-AUC Curve**
* Identify the best-performing model.

**Phase 5: Model Optimization and Testing**

1. **Hyperparameter Tuning**

* Use techniques like:
  + **Grid Search CV**
  + **Randomized Search CV**
  + **Bayesian Optimization**
* Optimize the model for better accuracy.

1. **Final Model Testing**

* Test the model on unseen data.
* Validate with cross-validation.

**Phase 6: Deployment and Documentation**

1. **Model Deployment (Optional)**

* Deploy the model using **Flask API, FastAPI, or Streamlit**.
* Create a simple **dashboard** to visualize predictions.

1. **Interpret Model Results**

* Explain **feature importance**.
* Provide insights on improving landing success.

1. **Document Findings**

* Write a **project report** or **presentation**.
* Include:
  + **Introduction**
  + **Data sources and processing**
  + **EDA findings**
  + **Modeling approach**
  + **Results and future work**

**Final Deliverables**

* Cleaned dataset
* Jupyter Notebook with analysis and model training
* Exploratory data analysis visualizations
* Final trained model
* Report summarizing findings and recommendations
* (Optional) Web-based deployment for predictions