

Classification Model

SpaceX Falcon 9 Landing Prediction







Our Process

From Problem Definition to Best Model Selection

ML Process Flowchart



Data Preparation (Tasks 1-3)

-  **Target (Y):** The `Class` variable was extracted. The goal is to predict if the landing will be a success (1) or a failure (0).
-  **Features (X):** 83 features were used, including `PayloadMass`, `Orbit`, `LaunchSite`, etc., all converted to numerical format.
-  **Standardization:** `StandardScaler` was applied to normalize the scale of all features, ensuring no single attribute dominates the model.
-  **Splitting:** Data was divided into 80% for training (72 samples) and 20% for testing (18 samples) using `train_test_split`.

Models Under Test (Tasks 4-10)



Logistic Regression

Robust and interpretable linear model, ideal for binary classification.



SVM

Support Vector Machine, effective in high-dimensional spaces.



Decision Tree

Non-linear, rule-based model, easy to visualize and interpret.



KNN

K-Nearest Neighbors, an instance-based model that classifies based on "neighborhood."

Optimization: GridSearchCV

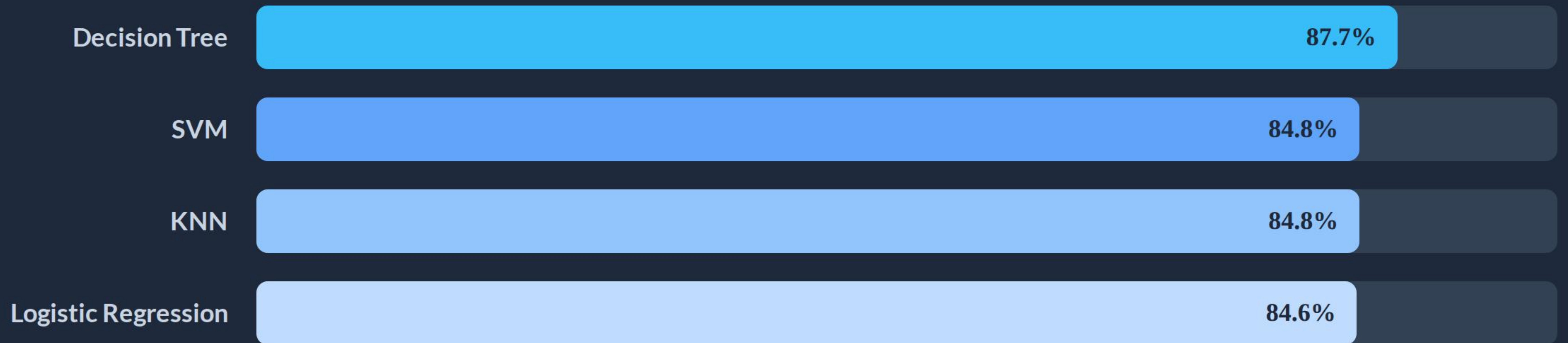
The Methodology

We used `GridSearchCV` to exhaustively test various hyperparameter combinations for each of the four models. This ensures we find the best possible configuration.

Cross-Validation (CV=10)

Cross-Validation (CV=10) was used to evaluate each combination. This divides the training data into 10 folds, training on 9 and testing on 1, ensuring stable performance and preventing overfitting.

Evaluation: Accuracy (Validation)



The Decision Tree had the best average performance in cross-validation, although all models remained close.

Evaluation: Accuracy (Test)

83.3%

Test Set Accuracy

A Surprising Result

Upon evaluating the optimized models on the 18 test data points (which the model never saw), all four algorithms achieved the ****exact same**** accuracy of 83.33%.

Analysis: Confusion Matrix

(All Models)	Predicted: Failure (0)	Predicted: Landing (1)
Actual: Failure (0)	3 (True Negative)	3 (False Positive)
Actual: Landing (1)	0 (False Negative)	12 (True Positive)

Key point: 0 False Negatives. The model never predicted a failure when the landing was a success. The challenge lies in the 3 False Positives.

Conclusion: The Best Model



Logistic Regression

Although all models had identical test performance (83.3%), the **Logistic Regression** model was selected.

It offers the best combination of simplicity, interpretability, and robust performance, in addition to having stable CV performance. (Params: `C=0.01, solver='lbfgs'`).

Summarized Process

Building > Evaluation > Optimization > Selection. An iterative process to find the best solution.

Questions?

Thank you.

Image Sources



https://images.stockcake.com/public/3/8/a/38a9735d-b15c-4812-8aa5-baffb5560f5e_large/victory-awaits-above-stockcake.jpg

Source: stockcake.com



https://elements-resized.envatousercontent.com/elements-video-cover-images/1673d26b-8432-4e55-89b4-031b8b90f7fc/video_preview/video_preview_0000.jpg?w=500&cf_fit=cover&q=85&format=auto&s=1113ea946990c0a7c537e591cc29cd7b5f612fba445cf9633de8419c1afc6ed6

Source: elements.envato.com