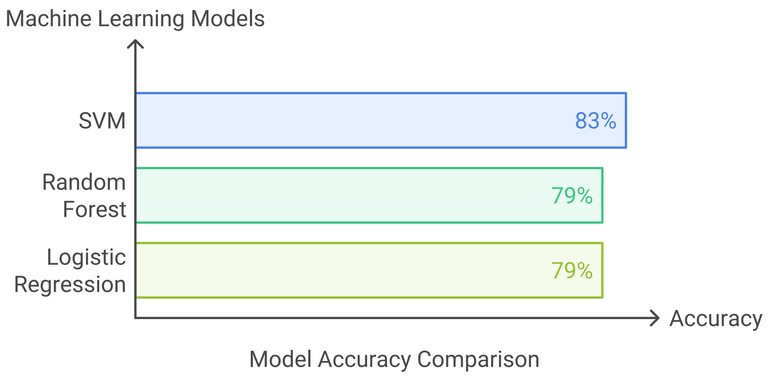
Process Overview for Each Model

1. **Logistic Regression:**
   - **Data Preprocessing**: Handled missing values in 'Age' and 'Embarked', encoded categorical variables ['Sex' and 'Embarked'].
   - **Feature Selection**: Selected features included 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', and 'Embarked_Q', 'Embarked_S' (dummy variables for 'Embarked').
   - **Model Training**: The Logistic Regression model was trained on the scaled features using **StandardScaler** for feature scaling.
   - **Cross-Validation**: Cross-validation was performed to evaluate model performance, resulting in an average accuracy of **79%**.
   - **Prediction**: After training, predictions were made on the validation set.
2. **Random Forest:**
   - **Data Preprocessing**: The same preprocessing steps as in Logistic Regression were followed for handling missing values and encoding categorical features.
   - **Hyperparameter Tuning**: A **Grid Search** with cross-validation was performed to find the best hyperparameters (such as **n_estimators**, **max_depth**, and **min_samples_split**) for the Random Forest model.
   - **Model Training**: The Random Forest model was trained using the best parameters found by GridSearchCV.
   - **Cross-Validation**: Cross-validation was applied to assess the model's performance, achieving an accuracy of **79%**.
   - **Feature Importance**: The model's feature importance was plotted to understand the contribution of each feature.
   - **Prediction**: Predictions were made separately on the test set.
3. **Support Vector Machine (SVM):**
   - **Data Preprocessing**: Similar preprocessing steps were applied: missing values were handled, categorical variables were encoded, and features were scaled using **StandardScaler**.
   - **Model Training**: An SVM model with a default RBF kernel was trained on the scaled features.
   - **Cross-Validation**: Cross-validation was applied to the model to evaluate its performance, and the model achieved an accuracy of **83%**, which was the highest among the three models.
   - **Prediction**: Predictions were made on the validation set after training the SVM model.

Model Performance:

- **SVM (Support Vector Machine)**:
  - **Best Model**: The SVM model achieved an accuracy of **83%**, outperforming the other models. It was able to effectively separate the data with its non-linear decision boundary, providing the best results.
- **Random Forest**:
  - **Accuracy**: The Random Forest model achieved an accuracy of **79%**. Despite performing well, it could not surpass the SVM in terms of accuracy.
  - **Feature Importance**: The model's performance was largely influenced by the importance of features like 'Fare' and 'Age', which are key predictors in the dataset.
- **Logistic Regression**:
  - **Accuracy**: Logistic Regression achieved an accuracy of **79%**, similar to Random Forest. The linear nature of Logistic Regression made it less powerful in capturing complex patterns in the data, resulting in a performance comparable to Random Forest.



Conclusion:

- The **Support Vector Machine (SVM)** model proved to be the best model for this dataset, with an impressive accuracy of **83%**, demonstrating its effectiveness in handling non-linear relationships within the data.
- Both **Random Forest** and **Logistic Regression** performed similarly, with an accuracy of **79%**, showing that both models were able to capture the main trends in the data, but without the same level of performance as SVM.

Thus, **SVM** is the recommended model for this particular classification task, providing the best overall accuracy.