**Loan Approval Prediction System (Report)**

* Introduction :

Loan status prediction is a crucial aspect of financial risk management. Predicting whether a loan applicant will default or not helps financial institutions in making informed lending decisions. In this project, we aim to build predictive models using Support Vector Machines (SVM) and Logistic Regression to determine the likelihood of a loan being approved or rejected.

* Objective :

The primary objective of this report is to compare the performance of SVM and Logistic Regression models on the given dataset and identify the key features that influence loan status.

* Dataset Description :

The dataset used for this project is sourced from Kaggle, named “Loan\_data.csv” . Key features include:

- Applicant Information: Gender, education, marital status, dependents, etc.

- Loan Information: Loan amount, loan term, credit history, etc.

- Property Information: Property area, etc.

The target variable is `Loan\_Status`, indicating whether the loan was approved (`1`) or not (`0`).

* Data Preprocessing :
* **Missing Values:** Handling missing data by imputation or removal.
* **Categorical Variables:** Converting categorical columns to numerical values.
* Exploratory Data Analysis (EDA) :

Analyzed the distribution of approved and rejected loans.

* Methodology :

1. Logistic Regression - Logistic Regression is a linear model used for binary classification tasks. It estimates the probability of a binary outcome based on the input features.
2. Support Vector Machines (SVM) -SVM is a powerful classification algorithm that works by finding the hyperplane that best separates the data points of different classes.

* Model Training and Evaluation –
* **Data Splitting:** Splitting the dataset into training and testing sets.
* **Model Training:** Training Logistic Regression and SVM models on the training data.
* **Evaluation Metrics:** Using metrics like accuracy, precision to evaluate model performance.
* Results :

Finally, I got the accuracy score 81% for logistic regression for the training dataset, and 83% in case of SVM model. Also, found the accuracy score 83% (approx.) for testing dataset, in both the models.

Both models demonstrated similar performance on the testing dataset, with an accuracy score of approximately 83%. However, there are some nuances to consider when comparing these models:

* **Consistency in Performance:** The SVM model showed a slight improvement in accuracy on the training set compared to Logistic Regression, which suggests that SVM might have captured the decision boundary more effectively during training. However, the performance on the testing dataset was nearly identical for both models, indicating that both models generalize well to unseen data.
* **Model Complexity and Interpretability:**
  + **Logistic Regression** is a simpler model that provides interpretable results, as the coefficients can be directly linked to the log odds of the outcome. This interpretability is useful for understanding the impact of individual features on loan status.
  + **SVM**, while potentially offering better performance in some cases, is generally more complex and less interpretable. The decision boundary is not as straightforward to interpret as the coefficients in Logistic Regression.

Given the similar accuracy scores on the testing dataset, the choice between Logistic Regression and SVM may depend on factors beyond accuracy alone:

* **If interpretability is crucial**, especially in financial applications where understanding the influence of features is important, **Logistic Regression** may be the preferred choice.
* **If model performance is the primary concern** and slightly higher accuracy on the training dataset is considered a significant factor, **SVM** could be chosen, especially when the data is well-separated and the complexity of the model can be justified.

Overall, **both models perform similarly well in this case.**