**PREDICTIVE ANALYSIS**

REPORT ON: LOAN APPROVAL

**Topic: Loan Approval Prediction**

**Abstraction**

This report outlines the development of a Loan Approval Prediction System using Support Vector Machines (SVM) for classification tasks. The dataset includes various features, such as applicant income, loan amount, and demographic information, that are used to predict the likelihood of loan approval. The system leverages preprocessing techniques, feature engineering, and hyperparameter optimization to achieve a robust predictive model.

The workflow includes cleaning and preparing the dataset, encoding categorical features, scaling numeric variables, and training the model. Visualization tools such as heatmaps and bar plots were employed for exploratory data analysis, and SVM hyperparameters were optimized using grid search to enhance accuracy.

**Introduction**

Predicting loan approval is a critical problem in financial and banking sectors. The goal of this project is to build a predictive model to classify loans as approved or not approved based on historical data. This involves analyzing patterns in the dataset and applying machine learning techniques for classification.

Support Vector Machines (SVM) were selected due to their effectiveness in handling high-dimensional spaces and clear separation of classes using kernel tricks.

**Literature Survey**

Loan approval prediction has been widely studied in financial domains. Traditional methods often rely on statistical approaches like logistic regression. However, machine learning models like SVM and Random Forest have gained popularity due to their accuracy and scalability. Feature engineering and preprocessing techniques such as scaling and encoding are crucial for improving model performance. Research suggests combining ensemble methods or fine-tuning hyperparameters can further enhance prediction accuracy.

**Proposed System**

The Loan Approval Prediction system is implemented as follows:

**Preprocessing**

1. Dropped irrelevant columns, such as Loan\_ID.
2. Handled missing values using mean and mode imputation.
3. Encoded categorical variables using LabelEncoder.
4. Scaled numerical variables using StandardScaler.

**Feature Engineering**

1. Introduced the Income\_to\_Loan\_Ratio feature to capture the relationship between income and loan amount.

**Model Training and Testing**

1. Splitted data into training and testing sets (60% training, 40% testing).
2. Trained an SVM classifier and optimized its hyperparameters using grid search.
3. Evaluated the model on both training and testing sets.

**Visualization**

1. Used heatmaps to identify correlations between features.
2. Bar plots for visualizing categorical feature distributions.

**Mathematical Foundations**

**Support Vector Machine (SVM)**

SVM is a supervised learning algorithm used for classification tasks. It identifies a hyperplane that best separates the classes in feature space. Key parameters include:

1. **Kernel**: Defines the function to project data into higher dimensions.
2. **C**: Regularization parameter to prevent overfitting.
3. **Gamma**: Influences the decision boundary curvature.

**Performance Metrics**

1. **Accuracy Score**: Evaluates the percentage of correctly classified samples.

IMPLEMENTATION:

















