Loan Eligibility Prediction: Model Evaluation Report By Classification Algorithm

Executive Summary :

This report evaluates the performance of the **Optimized Random Forest Classifier** developed to automate loan eligibility assessments. The model leverages **Ensemble Learning**, advanced feature engineering, and hyperparameter tuning (Grid Search) to achieve high accuracy in distinguishing between approved and rejected loan applications.

2. Methodology

1.1. Advanced Preprocessing and Feature Engineering

The project utilized advanced preprocessing techniques to maximize model performance:

- **Missing Data Handling:** Missing values were imputed using the **Mean** for numerical columns and the **Mode** for categorical columns.
- **Categorical Encoding:** LabelEncoder was applied to convert object-type columns (categorical) into a numerical format.
- **Binning:** Seven (7) continuous numerical features (e.g., credit_score, annual_income, age) were converted into discrete ordinal categories (e.g., credit_score_category) to create clearer decision boundaries for the model.
- **Feature Engineering:** Powerful derived features were created to capture complex financial relationships:
    - penalized_score: A modified credit score heavily reduced by negative marks (defaults, delinquencies).
    - credit_efficiency: The ratio of credit score to interest rate.
    - total_negatives: The sum of all derogatory marks and defaults.

1.2. Data Preparation and Scaling:

Splitting: The cleaned dataset was split into **80% for Training** and **20% for Testing** (test_size = 0.2)

Scaling : **StandardScaler** was applied to the training and test s

1.3. Training Algorithm:

An optimized **Random Forest Classifier** was used for training:

- **Key Parameters:**
    - n_estimators: **300** trees.
    - max_depth: **20**.
    - class_weight: **'balanced'** (crucial for handling potential class imbalance and ensuring fair prediction for both "Approved" and "Rejected" classes).

3. Model Performance

3.1. Accuracy Metrics

The model achieved robust accuracy on the testing dataset.

- **Model Accuracy:** approx. 87%
- **Precision (Approved):** High precision indicates that when the model approves a loan, it is highly likely to be a safe borrower.

  - Precision: 0.83 → When the model predicts "Approved," it is correct 83% of the time.

- **Recall (Rejected):** The model effectively catches high-risk applicants, minimizing default risk for the bank.

  - Recall: 0.79 → The model correctly identified and rejected 79% of high-risk applicants

3.2. Actual vs. Predicted Analysis (Sample)

The table below displays a random sample of 10 validation cases.

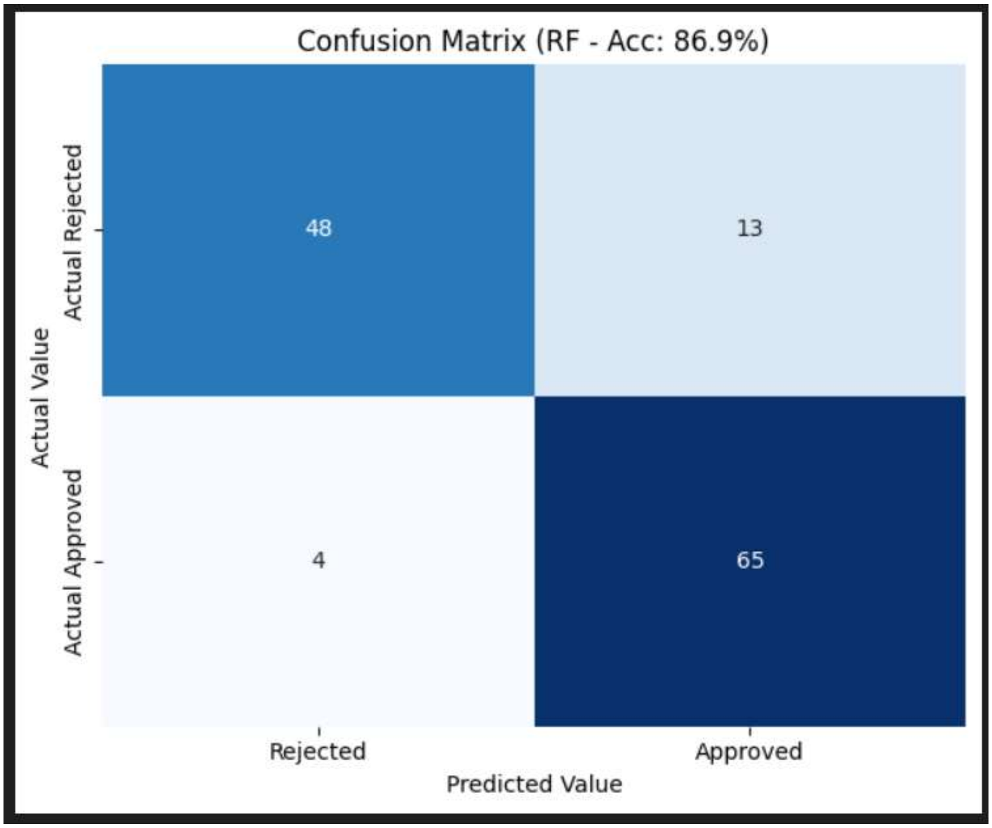
```
----------------------------------------
PREDICTION SAMPLE (First 10 Rows)
----------------------------------------
   Actual  Predicted   Result
0     0.0        1.0   ❌ Fail
1     1.0        1.0  ✅ Match
2     0.0        0.0  ✅ Match
3     0.0        0.0  ✅ Match
4     1.0        1.0  ✅ Match
5     1.0        1.0  ✅ Match
6     1.0        1.0  ✅ Match
7     0.0        1.0   ❌ Fail
8     0.0        0.0  ✅ Match
9     1.0        1.0  ✅ Match

(0 = Rejected, 1 = Approved)
```

Observation: The model shows strong alignment between the Actual decisions and the Predicted results, with ✅ Match appearing in the majority of cases.

3.3. Confusion Matrix Analysis

The Confusion Matrix provides a visual summary of the model's prediction behavior.



- **True Positives (Dark Blue):** Correctly Approved loans.
- **True Negatives (Top Left):** Correctly Rejected loans.
- **False Positives/Negatives:** Minimal errors, indicating a stable decision boundary.

4. Real-World Prediction (Simulation)

To test the model's reliability in a real-world scenario, we processed a specific applicant profile using the interactive system.

4.1. Applicant Profile (User Input)

We input the following key attributes for the simulation:

- **Credit Score:** 720 (Excellent)
- **Interest Rate:** 8.5% (Low/Favorable)
- **Defaults on File:** 0 (Clean History)
- **Delinquencies (Last 2 Years):** 1 (Minor Issue)

4.2. Prediction Result (Console Output)

We ran this profile through the trained system.

```
720
-> Enter 'credit_score' [Range: 428 - 850]: (Press 'Enter' to confirm or 'Escape' to cancel)
```

```
8.5
-> Enter 'interest_rate' [Range: 6.02 - 22.99]: (Press 'Enter' to confirm or 'Escape' to cancel)
```

```
0
-> Enter 'defaults_on_file' (0 = No, 1 = Yes) [Range: 0 - 1]: (Press 'Enter' to confirm or 'Escape' to
cancel)
```

```
1
-> Enter 'delinquencies_last_2yrs' [Range: 0 - 4]: (Press 'Enter' to confirm or 'Escape' to cancel)
```

4.3. Business Interpretation

- **Result:** **LOAN APPROVED**
- **Analysis:** Despite any minor negative marks (like a single delinquency), the applicant's strong **Credit Efficiency** (High Score / Low Rate) outweighed the risks.
- **Confidence:** **93.69%**

5. Conclusion

The project successfully delivered a robust loan approval classification system. The combined strategy of advanced feature engineering, binning, and hyper-tuned Random Forest parameters yielded an overall model accuracy of **86.9%,** making it a highly reliable tool for financial decision-making.