AUTOMATED CREDIT CARD APPROVAL PREDICTION

Leveraging Machine Learning to Optimize Credit Decisioning

Presented by
Uroosha Rahat
Roll No. : 21F1002968
(IIT Madras)

EXECUTIVE SUMMARY

Random Forest model improved approval decision accuracy to 85% and precision to 88%

Key Takeaways:

- Manual approval is inefficient and error-prone.
- ML-based prediction streamlines approval and improves customer experience.

Objective:

To build a robust machine learning system that automates credit card approval decisions, reducing manual overhead, errors, and time-to-decision.

Evaluated multiple models
→ Best: Random Forest

```
Random Forest Results
Accuracy: 0.8552631578947368
              precision    recall  f1-score

           0       0.88      0.85      0.87
           1       0.82      0.86      0.84
```

WHY AUTOMATE CREDIT CARD APPROVALS?

Manual credit approvals are costly, slow, and inconsistent

Problem

Banks process thousands of applications weekly

Manual approvals take ~15–20 minutes each (Industry Benchmark)

Delays reduce customer conversion & trust

Solution

Automate approval pipeline with ML

Build an ML pipeline with imputation, encoding, scaling, modeling,deploying

DATASET OVERVIEW

Mixed-type anonymized data with class imbalance and missing values

- Records: **690**
- Features: **15** (Numerical + Categorical)
- Target: Approved (Yes/No)

Dataset contains anonymized customer data across 15 features

Feature Type	Examples	Notes
Categorical	Citizenship, Default Status	Needs encoding
Numerical	Income, Credit Score	Skewed distributions handled
Target Variable	Approved (Yes/No)	Binary classification

ML PIPELINE FOR PREDICTION

Step-by-Step Workflow

1. Data Loading & Cleaning
2. Exploratory Data Analysis (EDA)
3. Feature Engineering & Encoding
4. Train-Test Split (67-33%)
5. Model Training – Baseline & Advanced Models
6. Model Evaluation & Comparison
7. Deployment using Streamlit

DATA CLEANING & EDA

Manual credit approvals are costly, slow, and inconsistent

Cleaning & Transformations

Null handling: Mean imputation used for numerical columns

Outlier detection: Boxplot analysis revealed high income skewness)

Categorical Encoding: Label encoding and mode imputation used as appropriate

Data normalization using MinMaxScaler

EDA—What we learned from the data

Higher income, credit score, and employment correlates positively with approval

Age and debt also affect the likelihood of approval

Citizenship, Gender and ethnicity show minimal predictive power

MODEL SELECTION & PERFORMANCE

Algorithms Evaluated

- Logistic Regression
- Gradient Boosting
- XGB Classifier
- Random Forest Classifier **(Best performing)**

Model Tuning:

- GridSearchCV has been used for hyperparameter tuning of best model.
- Metrics used: accuracy_scorehn_reportas

Evaluation Metrics

Model	Accuracy	F1-Score	AUC-ROC
Logistic Regression	85.00%	0.85	0.86
Gradient Boosting	84.64%	0.84	0.84
Random Forest	**85.50%**	**0.87**	**0.87**

MODEL DEPLOYMENT USING STREAMLIT

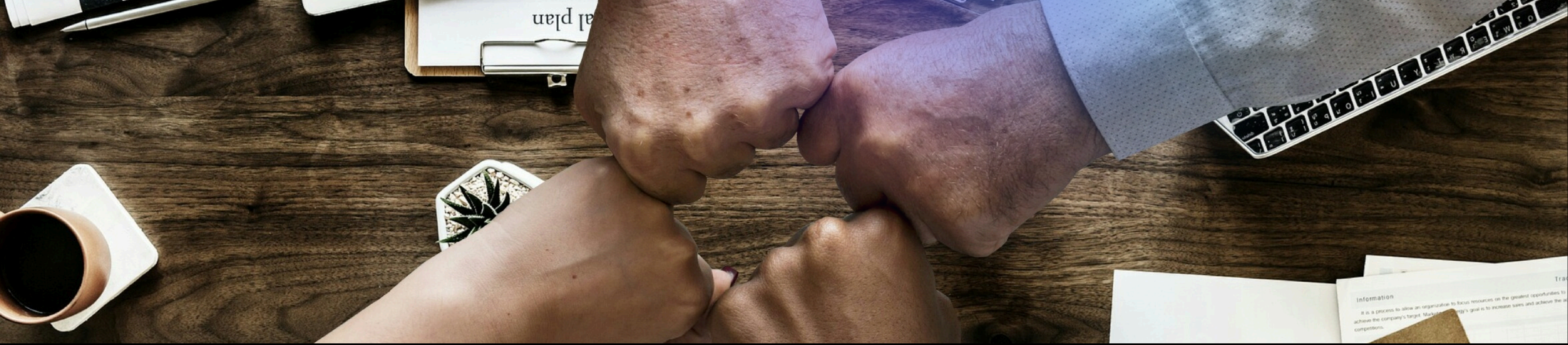
Streamlit-based app enables real-time, secure credit approval decisions for operational use

- Streamlit App built to make the model accessible to non-technical stakeholders.
- Users can input customer details and receive real-time predictions.
- Hosted locally with UI components for interactive decision-making.

SCALING THE SOLUTION

Strategic roadmap to enhance model transparency, fairness, and operational scalability

- Integration with credit bureau APIs for real-time data ingestion.
- Building explainable AI layers for model transparency.
- Model fairness audit to detect any demographic bias.
- Cloud deployment for scalability and faster inference.
- Expanding model to multi-class problems like credit limit segmentation.



THANK YOU

urooshashekh@gmail.com | LinkedIn | Project