CREDIT RISK ANALYSIS

Ankit Kumar Verma
Bharath N
Karthik Chivukula

Slide 1: Title Slide

- **Title:** Credit Risk Analysis & Loan Approval Prediction
- **Presented by:** Ankit Kumar Verma, Bharath N, Karthik Chivukula
- **Date:** 04 FEB 2025

Slide 2: Project Overview

Objective: Build a classification model to predict loan approval based on applicant and credit bureau data.

Datasets Used:

- Applications Dataset (application_base.csv)
- Bureau Dataset (bureau.csv)

Slide 3: Business Problem

- Home Credit needs a reliable model to assess loan applicants.
- Goal: Reduce **loan defaults** while **maximizing approvals**.
- Use **historical credit behavior** to predict risk.

Slide 4: Data Understanding & Cleaning

- **Applications Data:** Demographic & financial details of applicants.
- **Bureau Data:** Past credit history at trade level.
- **Challenges:**
  - **Missing Values** → Handled using median imputation.
  - **Categorical Variables** → Encoded using Label Encoding.
  - **Class Imbalance** → Addressed using SMOTE.

Slide 5: Feature Engineering

- Bureau data is at **trade level**, needs aggregation.
- New features created:
  - **Total number of trades** per applicant.
  - **Number of active/closed trades**.
  - **Max overdue days**.
  - **Average days credit was held**.
- Merged aggregated **bureau features** with applications dataset.

Slide 6: Exploratory Data Analysis (EDA)

- **Income Distribution:** Applicants with higher income have lower default risk.
- **Credit Amount vs. Default Rate:** Higher loan amounts increase risk.
- **Feature Correlations:** Identified key influencing variables.

Slide 7: Handling Class Imbalance

Problem: Default cases (TARGET=1) are much fewer than TARGET=0

Solution:

- Applied SMOTE to generate synthetic examples.
- Improved recall for predicting high-risk applicants

Slide 8: Model Selection & Training

- **Models Used:**
  - **Random Forest** (Baseline Model)
  - **XGBoost** (Boosting Model)
  - **Logistic Regression** (Benchmark Model)
- **Data Split:** 80% Training | 20% Testing
- **Feature Scaling:** Standardized numerical features.

Slide 9: Model Evaluation Metrics

- **Classification Metrics Used:**
- **Precission** - 0.93
- **Recall** – 0.98
- **ROC-AUC** Score - 0.6240
- **Best Performing Model: XGBoost** (ROC-AUC = 0.6897)

Slide 10: Feature Importance (Random Forest)

Key Features Identified:

- AMT_CREDIT (Loan Amount Requested)
- DAYS_CREDIT(Past credit History Length)
- CREDIT_DAY_OVERDUE(Overdue Days in Past Loans)

Business Insights: Applicants with a history of long overdue payments are more likely to default.

Slide 11: Business Recommendations

1. **Auto-Approve:** Low-risk applicants (Credit Score > 0.7).
2. **Manual Review:** Medium-risk applicants (0.3 ≤ Score ≤ 0.7).
3. **Reject Loans:** High-risk applicants (Score < 0.3).
4. **Interest Rate Adjustment:** Increase rates for medium-risk applicants.
5. **Encourage Collateral:** For high loan amounts to reduce risk.

Slide 12: Conclusion & Next Steps

- Successfully built a **credit risk classification model**.
- **XGBoost** performed best with a **ROC-AUC of 0.6240**
- Business insights support **better loan approval strategies**.
- **Future Work:**
  - Deploy the model for **real-time loan approvals**.
  - Enhance feature selection with **deep learning techniques**.

Slide 13: Thank You!

- **Q&A**
  Contact:

  Ankit Kumar Verma [ankit.sain4@gmail.com]

  Bharath N [bharath.bglr.15@gmail.com]

  Karthik Chivukula [90karthik.c@gmail.com]

GitHub Repository:

https://github.com/bharathn15/Credit-Risk-Analysis