FINANCIAL ANALYTICS REPORT ID: CR-2026-01

Credit Risk Analysis

Loan Data 2007–2014

Models: Logistic Regression (SAGA) & Random Forest

PRIMARY OBJECTIVE

Identify key drivers of loan default and provide actionable risk policy recommendations through predictive modeling.

● N = 466,285 Records ● Data Period: 7 Years

METHODOLOGY OVERVIEW

Dataset & Technical Approach

Loan Data (2007-2014)

FEATURE SELECTION

FINANCIAL METRICS

- loan_amnt
- int_rate
- installment
- annual_inc
- dti (Debt-to-Income)

BORROWER PROFILE

- grade / sub_grade
- emp_length
- home_ownership
- verification_status
- purpose

FEATURE ENGINEERING

Ratio Calculation

```
loan_to_income = loan_amnt / annual_inc
installment_to_income = installment / annual_inc
```

Grade Grouping

Prime (A,B) | Near-Prime (C,D) | Subprime (E-G)

Modeling Pipeline

1 **DATA CLEANING & PREP**

Imputed missing values (median for revol_util), parsed dates, and standardized formats.

Remaining Data: 234,946 records

2 **TRAIN / TEST SPLIT**

Time-based splitting strategy to prevent data leakage.

Train < Jan 2013 | Test ≥ Jan 2013

3 **MODEL SPECIFICATIONS**

Logistic Regression

```
solver='saga'
class_weight='balanced'
```

Random Forest

```
n_estimators=150
max_depth=10
```

4 **EVALUATION METRICS**

ROC-AUC | Recall | Confusion Matrix

EXPLORATORY DATA ANALYSIS

Key Risk Drivers & Insights

● Good Loan ● Bad Loan (Risk)

BAD VS GOOD LOANS

Interest Rate +20.6% Higher
13.26% → 15.99%

Debt-to-Income (DTI) +15.2% Higher
15.94 → 18.36

Loan-to-Income +21.9% Higher
0.201 → 0.245

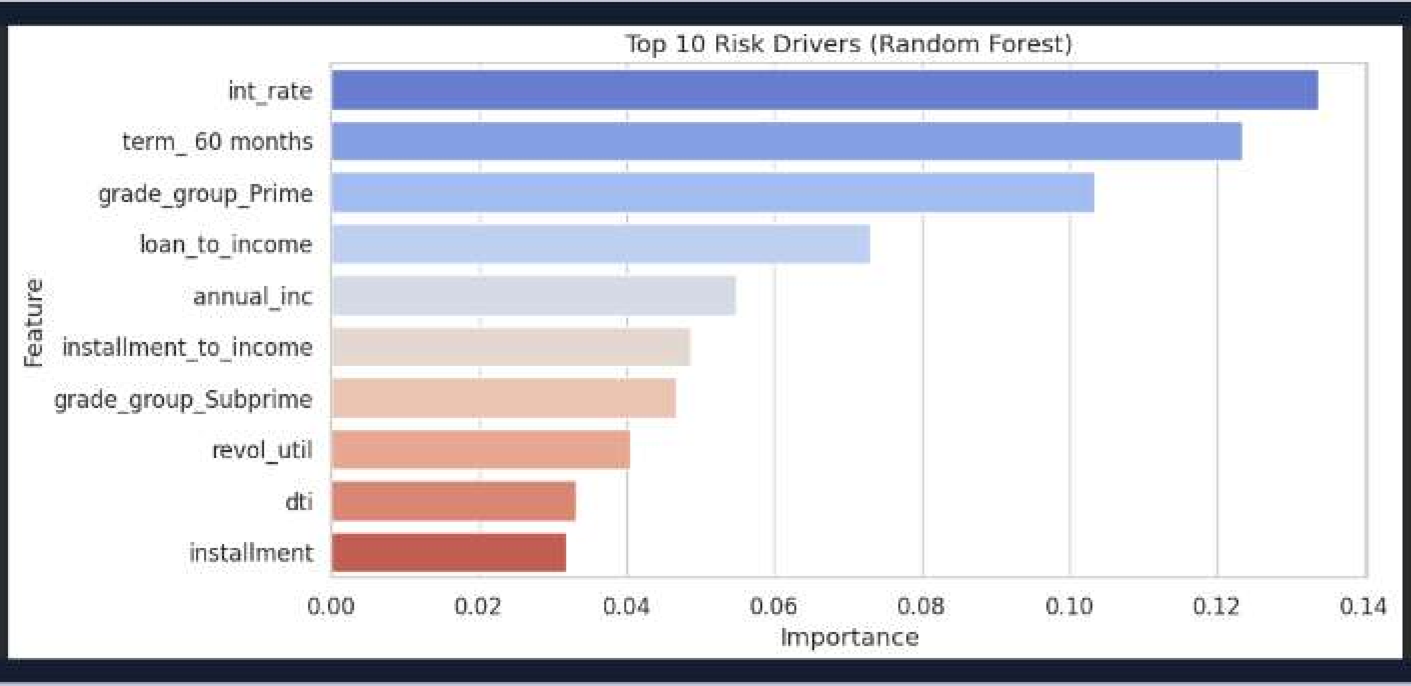
HIGHEST RISK PURPOSES

PURPOSE	DEFAULT RATE
1 Small Business	31.3%
2 Moving	23.9%
3 Other	23.8%
4 Medical	22.4%
5 Debt Consol.	22.4%

Feature Importance (Random Forest)

PRIMARY DRIVER

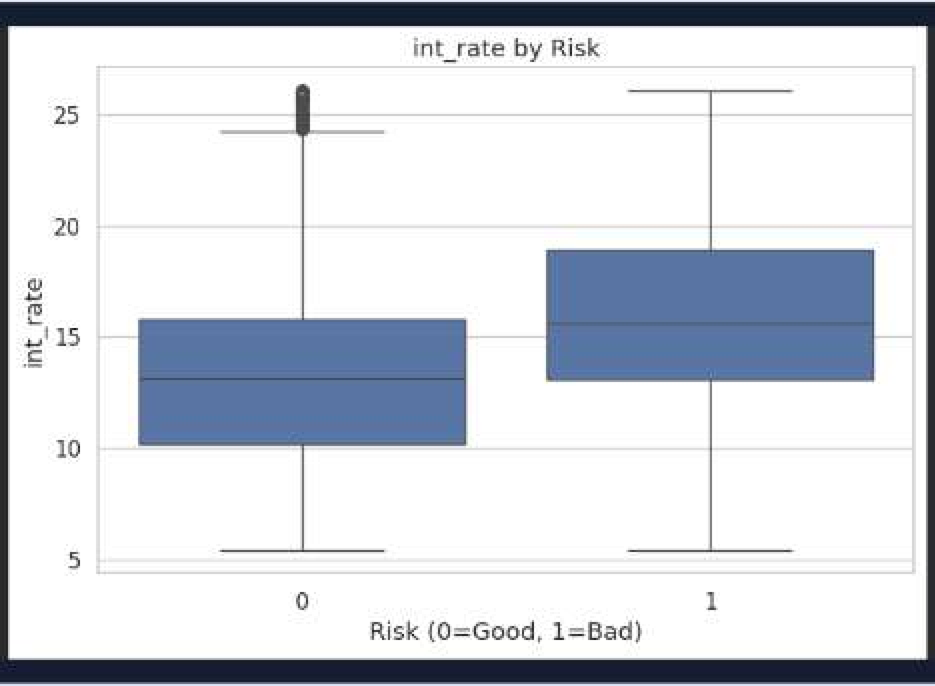
Top predictors contributing to model accuracy



Interest Rate Distribution by Risk

KEY INDICATOR

Clear separation observed in interest rates between Good (0) and Bad (1) loans



RISK STRATIFICATION

Credit Grade Risk Patterns

Monotonic Trend Analysis

Monotonic Increase

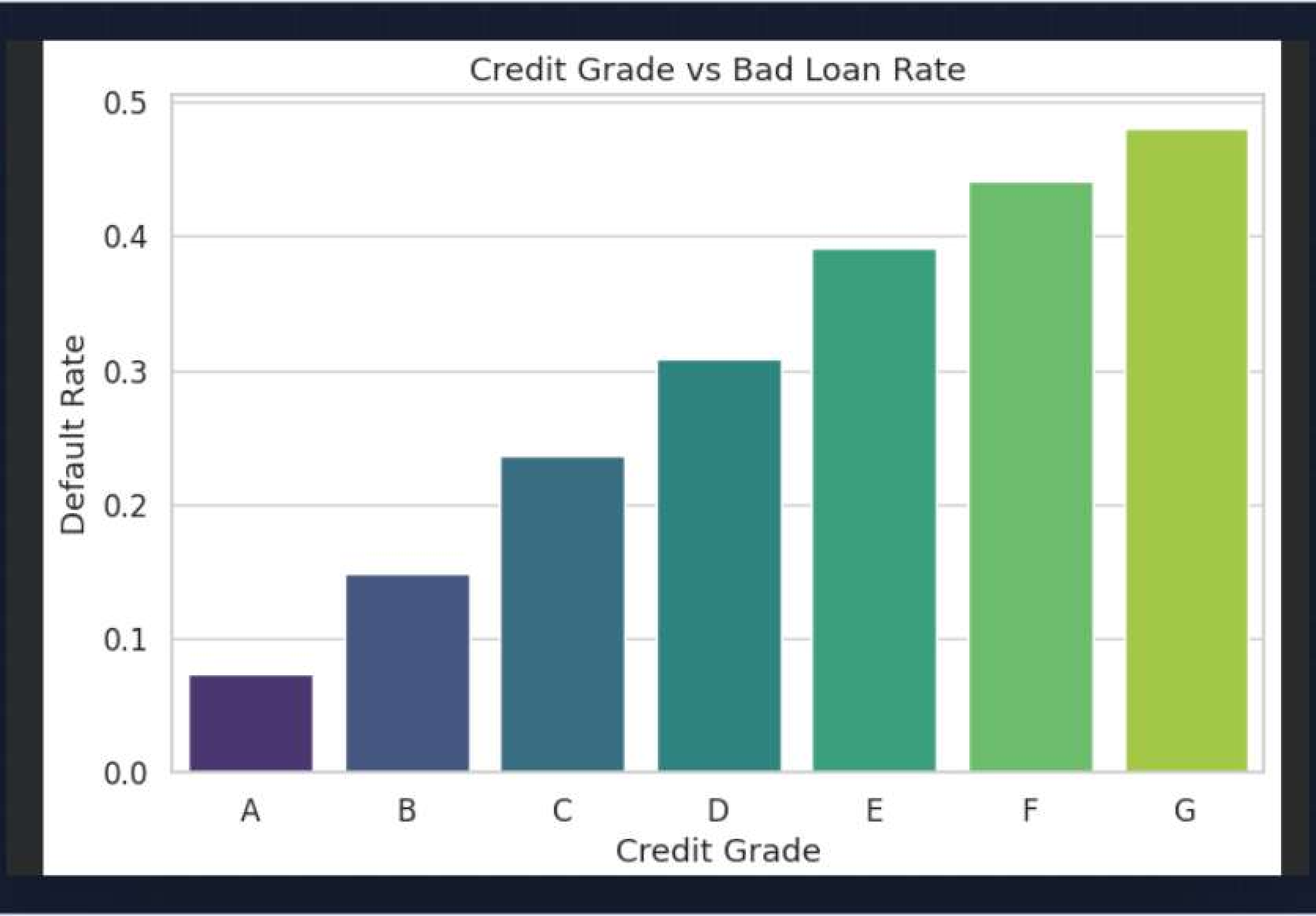
Default probability rises consistently as credit grade worsens from A to G, validating the grading system's effectiveness as a primary risk filter.

DEFAULT RATES BY GRADE

A	7.3%
B	14.9%
C	23.7%
D	30.9%
E	39.2%
F	44.1%

Visualization: Credit Grade vs Bad Loan Rate

n=234,946



Bars represent the proportion of defaulted loans (Risk=1) within each grade bucket.

Low Risk (A-C) 🟧 Med Risk (D) High Risk (E-G)

MODEL EVALUATION

Performance Comparison

🏆 Winner: Random Forest

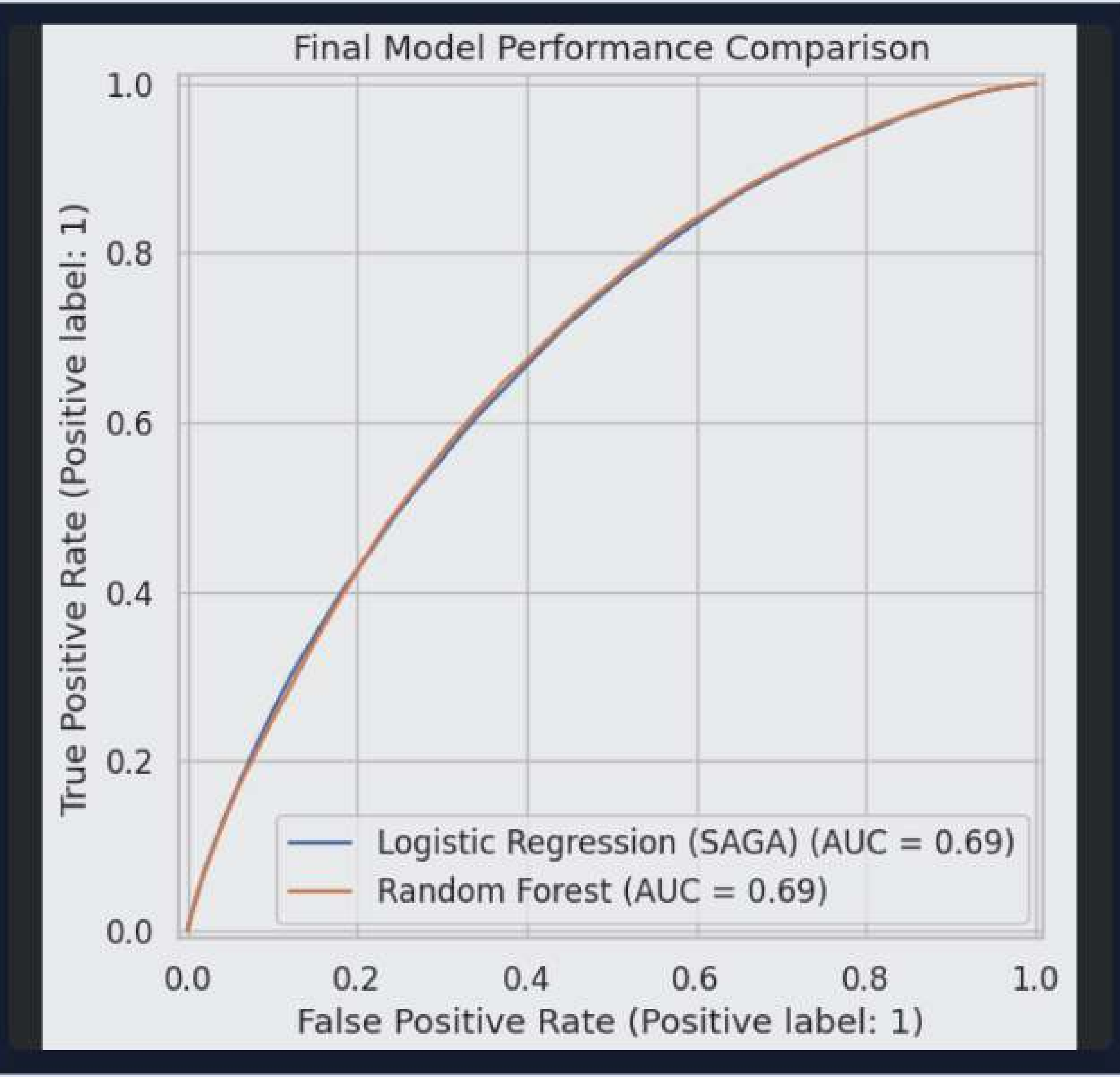
LOGISTIC REGRESSION Baseline

0.6854 AUC

RANDOM FOREST **Best**

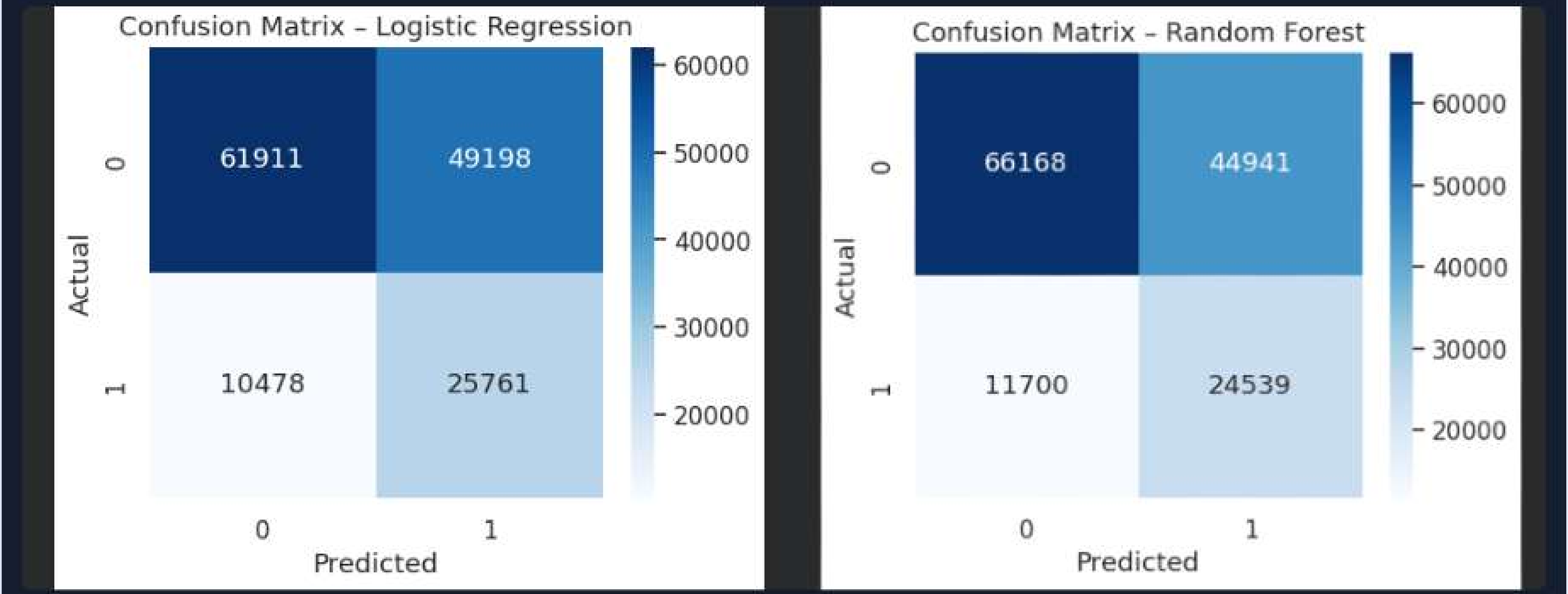
0.6870 AUC

ROC Curve Analysis



Confusion Matrices

■ True Neg ■ True Pos



PERFORMANCE EDGE

Random Forest achieves marginally higher AUC and demonstrates better stability across validation folds.

RECALL FOCUS

RF model correctly identifies a higher volume of bad loans, critical for minimizing default losses.

ACTION PLAN

Key Insights & Recommendations

Final Analysis

ANALYST FINDINGS

Risk Profile Indicators

Bad loans exhibit significantly higher interest rates (16% vs 13%), DTI ratios, and loan-to-income burdens.

Credit Grading Validity

Default risk is strictly monotonic across grades A through G, confirming the grading system's robustness.

Model Performance

Random Forest slightly outperforms Logistic Regression (AUC 0.687), offering better recall for bad loans.

Primary Risk Drivers

Key predictors include Interest Rate, Loan Term (60 months), Credit Grade, and Debt-to-Income leverage.

BUSINESS RECOMMENDATIONS

01

Optimize Pricing Strategy

Monitor interest rate tiers carefully; excessive rates correlate with default. Ensure risk-adjusted pricing accounts for this elasticity.

02

Stricter Affordability Checks

Introduce tighter caps on Debt-to-Income (DTI) and Loan-to-Income ratios for applicants in lower credit grades.

03

Enhanced Underwriting for Segments

Apply manual review or enhanced automated scrutiny for high-risk loan purposes like Small Business and Moving.

04

Leverage Credit Grades

Continue utilizing credit grades as the primary coarse filter, while using the Random Forest model for borderline cases.