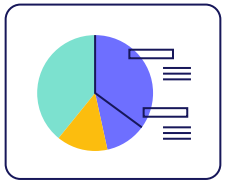


Loan Approval Prediction: A Data-Driven Approach to Credit Risk

Crouse : CDS-3513 Data Mining Technique
Instructor: Dr. Asif Malik

Group Members:
Hessa Khalfan – H00535166
Maryam Ali Abdulla – H00535836
Nourah Abdulla Alghfeli – H00535834

The Core Challenge: Balancing Profit and Risk

In the competitive financial services sector, banks and lending institutions face a fundamental challenge: managing credit risk while maximizing profitability.

The Business Reality:

- Banks generate revenue by lending money to borrowers
- Every loan carries inherent risk of default (borrower fails to repay)
- Default leads to significant financial losses

The Critical Dilemma:

Lend too aggressively:

- High default rates → Major financial losses

Lend too cautiously:

- Reject creditworthy applicants → Lost revenue and customers

Our Solution: A data-driven predictive model to help Credit Risk Managers make faster, more accurate, and more consistent lending decisions

Our Three-Fold Objectives

Business Objectives:

1. **Predict and Minimize Loan Defaults**
   - Identify high-risk applicants accurately
   - Reduce financial losses from defaults

2. **Maximize Approval of Creditworthy Applicants**
   - Avoid rejecting good customers
   - Increase profitable lending opportunities

3. **Enhance Decision-Making Efficiency**
   - Automate initial screening process
   - Enable faster, more consistent decisions

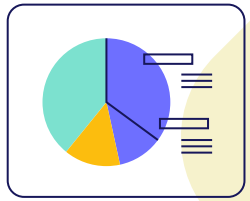
Data Mining Objectives:

1. **Develop a Classification Model**
   - Predict "Approved" or "Rejected" status
   - Based on applicant financial profile

2. **Compare Different Algorithms**
   - Build Decision Tree model
   - Build Logistic Regression model
   - Determine best performing algorithm

3. **Identify Key Influential Factors**
   - Analyze CIBIL score, income, assets
   - Provide actionable business insights

A Structured Approach: The CRISP-DM Framework



1

Business Understanding

Define objectives and requirements from business perspective

2

Data Understanding

Collect initial data and identify quality issues

3

Data Preparation

Clean, transform, and construct final dataset

4

Modeling

Build and train Decision Tree and Logistic Regression models

A Structured Approach: The CRISP-DM Framework

5

Evaluation

Assess model performance and validate business objectives

6

Deployment

Present findings through report, dashboard, and presentation

Why CRISP-DM?

- Industry-standard methodology for data mining projects
- Iterative and flexible approach allowing refinement
- Business-focused ensuring practical value

Preparing High-Quality Data for Modeling

Source: Kaggle Loan Approval Prediction Dataset by Archit Sharma

Size: 4,269 loan applications with 13 variables

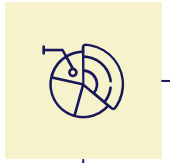
Quality: High-quality dataset with no missing values

Data Cleaning

Removed loan_id column (no predictive value)

Feature Engineering

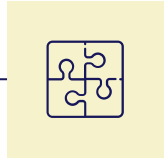
Created "Total Asset Value" variable
Sum of: residential, commercial, luxury, bank assets



Data Transformation

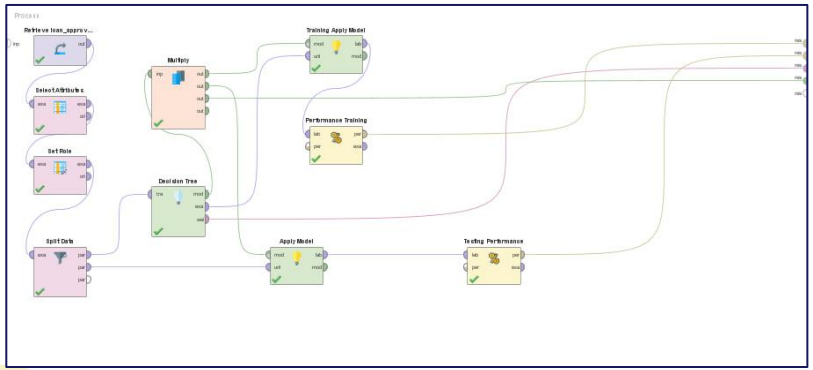
Converted categorical features to numerical format
Education (Graduate/Not Graduate) → Binary encoding
Self-employed status → Binary encoding

Data Splitting



Data Splitting
70% training set (2,988 records)
30% testing set (1,281 records)
Stratified sampling to maintain class distribution

Building Two Competing Predictive Models



Decision Tree Model:

How it works: Creates a set of hierarchical "if-then" rules

- Splits data into branches based on most important features
- Each path leads to a final decision (Approved/Rejected)

Advantage: Highly interpretable and easy to explain

- Non-technical users can understand the logic

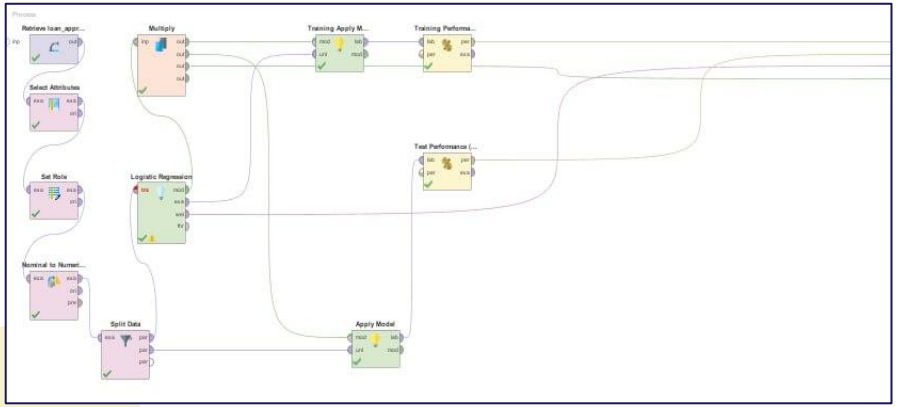
Example Rule: "IF CIBIL score < 550, THEN Reject“

- Clear, actionable decision criteria

Business Value: Transparent decision-making

- Can explain rejection reasons to applicants
- Complies with regulatory requirements for explainability

Building Two Competing Predictive Models



Logistic Regression Model:

How it works: Creates a mathematical to calculate approval probability

- Assigns weights to each feature (CIBIL, income, assets)
- Outputs probability score between 0 and 1

Advantage: Powerful statistical benchmark with proven track record

- Widely used in financial industry

Business Value: Industry-standard approach

- Provides baseline to measure Decision Tree performance
- Validates our findings against established methods

The Decision Tree Was the Clear Winner

What This Means:

Decision Tree: Only **40 errors** out of 1,281 test cases
Logistic Regression: **115 errors** out of 1,281 test cases
Decision Tree is superior at both minimizing risk and capturing opportunities
The **96.7%** accuracy makes it highly reliable for real-world deployment

	Decision Tree	Logistic Regression's
Accuracy	96.9%	91.0%
Precision	97.4%	94.3%
Recall	97.4%	91.1%

Accuracy: Overall correctness of predictions
Precision: Reliability when approving loans (minimizes bad approvals)
Recall: Ability to identify all good applicants (minimizes missed opportunities)

Decision Tree

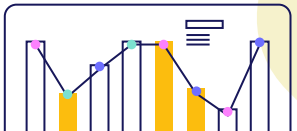
accuracy: 96.88%

	true Approved	true Rejected	class precision
pred. Approved	778	21	97.37%
pred. Rejected	19	463	96.06%
class recall	97.62%	95.66%	

Logistic Regression

accuracy: 91.02%

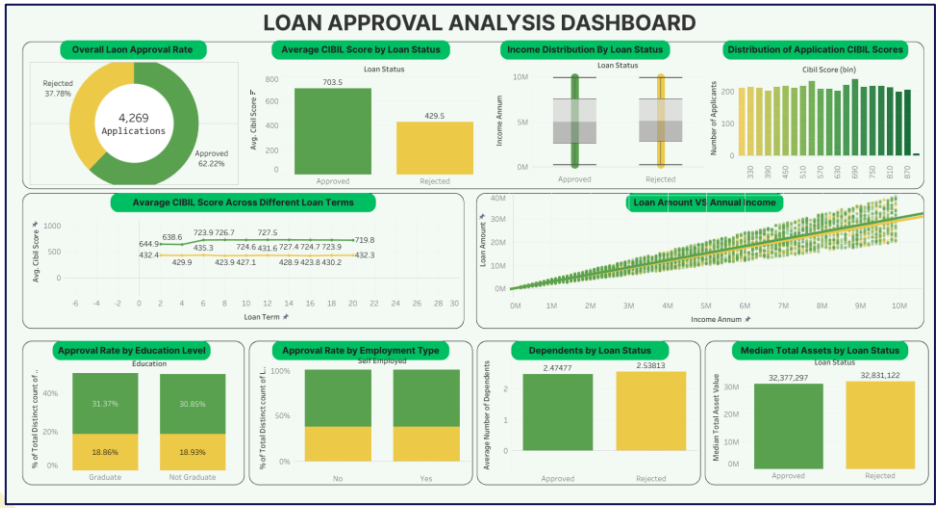
	true Approved	true Rejected	class precision
pred. Approved	726	44	94.29%
pred. Rejected	71	440	86.11%
class recall	91.09%	90.91%	



Interactive Dashboard: Turning Data Into Business Insights

Purpose:

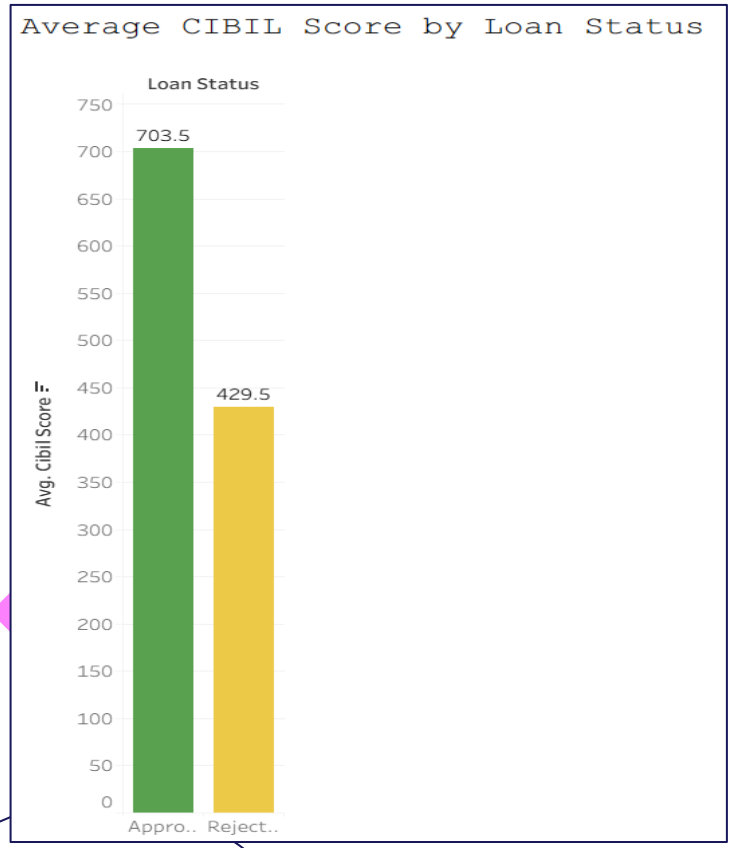
1. Make complex findings accessible to business users
2. Enable interactive exploration of lending patterns
3. Support data-driven decision making
4. Allow filtering and drill-down analysis by different segments



Key Performance Indicators Tracked:

- Overall approval rate
- CIBIL score distribution and impact
- Income and loan amount relationship
- Impact of education level on approval decisions
- Employment type analysis (salaried vs self-employed)
- Asset value influence

CIBIL Score: The Single Most Dominant Factor



The Evidence:

Approved applicants: Average CIBIL score of **704**

Rejected applicants: Average CIBIL score of **430**

Gap: **274** points a massive difference

What This Means:

The bank's lending decision **is overwhelmingly driven by credit history**

CIBIL score is the **primary decision criterion**

This factor alone explains most of the approval/rejection pattern

Business Implication:

Strong credit history → Almost always approved

Weak credit history → Almost always rejected

Key Insights from the Models

Income's Actual Role:

Income is **NOT** a primary factor for approval/rejection

Instead, income determines the **maximum loan amount** an applicant can receive

Higher income → Larger loan approved (if CIBIL score is good)

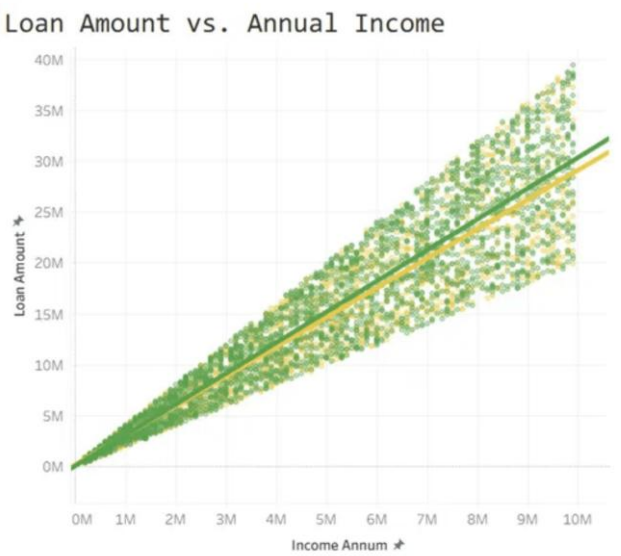
Demographics Have Minimal Impact:

Education Level: Almost no effect on approval

Employment Status: Minimal difference

Number of Dependents: No significant correlation

Conclusion: The lending model is purely financially driven



The **scatter plot** shows strong correlation between income and loan amount
Income acts as a limiting factor, not a deciding factor

Three Actionable Recommendations

1- Automate Pre-Screening:

Deploy the Decision Tree model for clear-cut applications
Fast-track high CIBIL scores, auto-decline low scores

Result: Significant time savings for analysts

2- Focus Human Expertise on ”Gray Area” Cases:

Have analysts focus on borderline applicants
Apply human judgment where it adds the most value

Result: Better decisions on complex cases

3- Targeted Marketing Strategy:

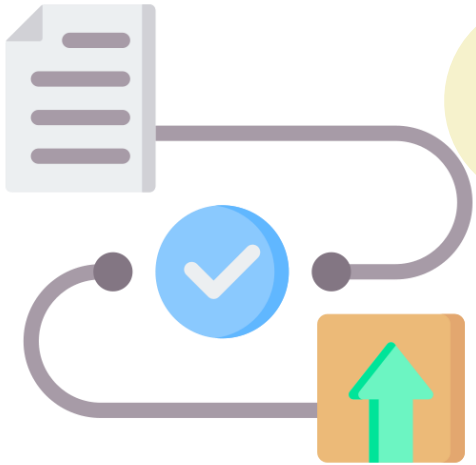
Focus campaigns on individuals with high CIBIL scores
Attract applicants most likely to be approved

Result: Higher approval rates and better customer satisfaction

Conclusion: A Data-Driven Path Forward

What We Achieved:

- ✓ Built a highly accurate (**96.7%**) and interpretable Decision Tree model
- ✓ Identified CIBIL score as the key driver of loan approvals
- ✓ Provided clear, actionable recommendations to improve efficiency and profitability



Key Takeaway:

Our analysis provides a data-driven foundation for optimizing the loan approval process, reducing risk, and increasing revenue through smarter decision-making.

Thank You