



CREDIT RISK ANALYSIS OF LOANS (HOUSE, CAR & PERSONAL) USING MACHINE LEARNING

¹Kamatham Akhil Kumar Reddy, ²Anannya Sinha, ³Swaroop Reddy, ⁴Dr. Surbhi Ghai

¹Student, ²Student, ³Student, ⁴Assistant professor

¹Masters in Business Administration

¹Lovely professional university, Punjab, India

Abstract: This research investigates credit risk in housing, car, and personal loans at Canara Bank using machine learning methods. Credit risk remains a major concern for banks, and this study applies a Decision Tree Classifier to analyze key financial attributes influencing loan repayment patterns. The dataset, sourced from Kaggle, includes variables such as debt-to-income ratio (DTI), income, credit score, loan grade, and homeownership status. The model achieved an accuracy of 89.5%, with DTI emerging as the most influential factor, followed by loan grade and income. The findings underscore the potential of machine learning in improving credit risk evaluation, minimizing non-performing assets, and enhancing the speed and accuracy of loan approvals. This approach allows for early identification of high-risk borrowers and enables more effective, personalized lending strategies. Overall, the study offers valuable insights into modernizing credit assessment practices to strengthen financial stability and decision-making in the banking sector

Keywords : Credit Risk, Debt-to-Income Ratio (DTI), Decision Tree Classifier, Machine Learning

Introduction:

In today's dynamic financial ecosystem, effectively managing credit risk has become essential for maintaining the stability and growth of banking institutions. Among various credit products, housing loans hold significant importance due to their substantial contribution to a bank's loan portfolio and their impact on profitability. This research delves into the credit risk associated with housing loans offered by Canara Bank, utilizing the power of machine learning techniques to derive meaningful insights. The health of the banking industry depends on how well the banks manage their credit risks. The risk of credit is the possibility of loss when there

is a failure on the part of borrowers to repay the debts. Financial institutions have faced significant challenges in managing credit risk and minimizing losses. Over the years, Indian banks have experienced a decline in interest income, which has contributed to issues such as slow economic growth and an increase in non-performing assets.

The main goal of this study is to examine the financial factors that impact loan repayment behavior, helping to identify the underlying risks associated with borrowers. Machine learning methods provide a sophisticated and data-driven approach to analyze large datasets, revealing patterns, interpreting decision-making rules, and uncovering intricate relationships that might not be immediately apparent through traditional methods. By adopting these techniques, the research seeks to enhance the operational efficiency of the loan approval process, facilitating quicker and more accurate decisions. This can only be achieved when banks assess the sufficiency of capital and loan loss reserves at any given moment. Credit risk management has become a very important aspect of the banking sector as the banks face financial crisis globally. The regulators started demanding more transparency and they set stringent rules to know whether banks have a thorough knowledge of the customers and their associated risks. To survive in the banking industry, banks should follow various techniques for managing credit risk. The main function of credit risk management is to decide what amount of credit is to be given to the borrower with the agreement to repay the amount at a specific term. Banks can survive in stiff competition if they can formulate policies and procedures that reduce credit risk and curb the non-performing assets. When a company experiences significant defaults due to borrowers' failure to repay, it negatively impacts the country's economic development. Loans are typically granted to farmers, industrialists, and self-help groups after evaluating their financial situation and capacity to repay the borrowed funds.

Banks provide loans to individuals, industrialists and businessmen, after assessing their financial status and verifying and collecting various documents from the borrower. To know the position or risk level of the borrowers, the providers usually collect a huge amount of information from them. Despite following all procedures and disbursing loans, banks may still face risks due to borrowers' failure to repay. The risk primarily lies with the lender and includes loss of both interest and principal, disruption of cash flows, and increased collection costs. Given the uncertainty of the future, banks must confront significant credit risks and can only survive by managing them effectively. By reducing non-performing loans, banks can maintain stability and strength. Credit risks directly impact a bank's performance, as the default of even a few large customers can lead to significant challenges. Banks cannot predict the exact percentage of defaulted loans, and fluctuations in defaults will cause variations in their earnings, leading to unstable profits.

Whenever banks provide loans to businesses and consumers, they will experience the risk of loan losses when borrowers fail to repay the loans provided to them. Furthermore, the study aims to tackle pressing issues such as reducing default rates and mitigating financial losses, which are critical to ensuring the bank's long-term sustainability. This involves identifying high-risk borrowers at an early stage, thereby allowing the bank to implement proactive measures such as loan restructuring or enhanced monitoring. Additionally, the research considers how these advanced technologies can streamline customer interaction processes, ensuring a seamless experience that builds trust and loyalty among borrowers

Ultimately, this research aspires to contribute to the field of credit risk management while simultaneously improving customer satisfaction by streamlining loan processes and making them more reliable and efficient. These findings are expected to offer actionable insights that not only support Canara Bank but also advance the broader understanding of credit risk in the banking sector, paving the way for more robust and future-ready risk assessment frameworks

CHAPTER 2

OBJECTIVES

- To explore financial factors that define /compose loan repayment behavior.
- To analyze credit risk through machine learning techniques.
- To suggest ways for enhancing operational efficiency and customer satisfaction by providing quicker loan decisions.

CHAPTER 3

REVIEW OF LITERATURE

Understanding and managing credit risk has become increasingly significant in the banking industry due to the growing complexity of loan products and stricter regulatory demands. Numerous studies have examined various aspects of credit risk, including the challenges posed by non-performing assets (NPAs), operational risks, and the integration of advanced machine learning technologies. Dr. Bhadrappa Haralayya explored trends in NPAs across various loan sectors, including housing loans, agricultural loans, and personal loans. His study highlighted a significant reduction in NPAs from 2018 to 2020, with personal loans showing the highest default rates, and noted improvements in vehicle loan recovery. He recommended adherence to Basel-II norms and a focus on sectors with better recovery status to mitigate credit risks (Haralayya, 2022). Addressing operational risks, Kumar emphasized the importance of regulatory compliance to mitigate fraud and emerging cybersecurity threats in Indian banks (Kumar, 2023), while Mulabagula et al. highlighted the inefficiencies of traditional loan management systems and suggested that artificial intelligence (AI) and machine learning (ML) are crucial for modern credit appraisal processes (Mulabagula *et al.*, 2024). (Joshi et al,2022). assessed the effectiveness of machine learning models, such as LogitBoost and Cart, in predicting operational efficiency, achieving high accuracy rates of 78.47% and 74.30%, respectively, in forecasting loan repayment behaviors (Joshi *et al.*, 2022). Hamid and Ahmed demonstrated the efficacy of decision tree algorithms, with J48 achieving a classification accuracy of 78.38% for loan risk classification, underscoring the potential of data mining techniques in credit risk management (Hamid & Ahmed, 2016).

Corporate social responsibility (CSR) initiatives have also been found to enhance banking credibility and customer retention. Belas et al. and Egri and Ralston noted that CSR activities, such as those implemented by Union Bank through its Financial Literacy and Credit Counseling (FLCC) centers and skill development programs, foster stronger community ties and indirectly influence credit risk management (Belas *et al.*, 2021; Egri & Ralston, 2021). Studies by Alzaidi and Malali and Gopalakrishnan underscored AI's potential in improving fraud detection, risk assessment, customer experience, and financial decision-making. While

Alzaidi observed that Middle Eastern banks are in the early stages of AI adoption (Alzaidi, 2018), Malali and Gopalakrishnan emphasized the transformative role of AI in wealth management and virtual assistant services (Malali & Gopalakrishnan, 2021). Alagic et al. incorporated mental health data into supervised machine learning algorithms for loan approval predictions, finding that Random Forest achieved the highest accuracy at 85%. Their feature importance analysis revealed that mental health factors and age were critical predictors in modern credit evaluations, highlighting the relevance of non-traditional variables (Alagic *et al.*, 2024).

Deep learning approaches have further advanced predictive accuracy in credit risk management. Dendramis et al. proposed artificial neural network (ANN) models employing ReLU activation functions and multiple hidden layers to predict defaults in small business loans. Their study found behavioral attitudes and moral hazards to be significant factors influencing defaults, with ANN models showing superior accuracy in differentiating default events (Dendramis *et al.*, 2021). Similarly, Hamzic et al. demonstrated that machine learning algorithms such as AdaBoost, Gradient Boosting, and XGBoost can enhance credit risk analysis, with XGBoost achieving an accuracy of 84%. They emphasized preprocessing techniques like label encoding and one-hot encoding to improve model accuracy and addressed the ethical considerations surrounding the use of sensitive data (Hamzic *et al.*, 2024). Lakhani et al. explored deep learning applications for lending bank loans, incorporating macroeconomic variables such as inflation and GDP growth. Their findings confirmed the efficacy of ANNs in mitigating financial losses caused by misclassification errors and reducing defaults (Lakhani *et al.*, 2021).

The broader evolution of AI technologies and their investment trends were analyzed by Kamal Singh, who highlighted the transformative impact of machine learning, natural language processing (NLP), and predictive analytics in the banking sector. Singh emphasized AI's role in enhancing data analysis, optimizing decision-making, and improving customer experiences (Singh, 2020). In the context of housing finance, Joseph Mini P. G. Head used multivariate discriminant analysis and the CAMEL ranking system to identify factors determining the financial performance of housing finance companies (HFCs), emphasizing ICT integration, economic indicators, and retail credit risk diversification. His study also highlighted the linkage between housing finance and employment generation (Joseph Mini P. G. Head, 2020).

Community-based housing finance models have also been explored. Yap and de Wandeler investigated grassroots financial systems such as chit funds, underscoring their sociocultural and financial benefits in promoting timely payments and advocating for alternative, community-driven approaches to support vulnerable populations (Yap & de Wandeler, 1994). Addressing housing finance gaps, Mahadeva identified disparities favoring urban populations and white-collar workers while underscoring the need for inclusive financial instruments like subsidies and rural mortgages to improve housing stock in underserved areas (Mahadeva, 2024). Chandrasekar, Kumaraswamy, and Nayan analyzed trends in housing loan disbursement between Scheduled Commercial Banks (SCBs) and HFCs, noting a decrease in SCBs from 2004 to 2015 but an increase in HFC growth. Their study used ANOVA and CAGR methods to demonstrate the evolving dynamics in housing finance institutions (Chandrasekar, Kumaraswamy, & Nayan, 2023).

CHAPTER 4

Conceptual model:

Numerous financial and personal factors affect a borrower's capacity to repay a loan. The main factors influencing loan repayment behaviour are described in this conceptual framework along with how they interact. Three important independent variables (factors) that affect the dependent variable make up the framework:

Independent variables:

Income: Having more money makes it easier to repay debt.

Experience: Responsible repayment behaviour is improved by financial experience.

Credit Score: A high credit score reduces the chance of default and is indicative of sound financial management.

Dependent Variable:

Loan Repayment Behaviour: Timely payment, delayed payment, or default. Framework:

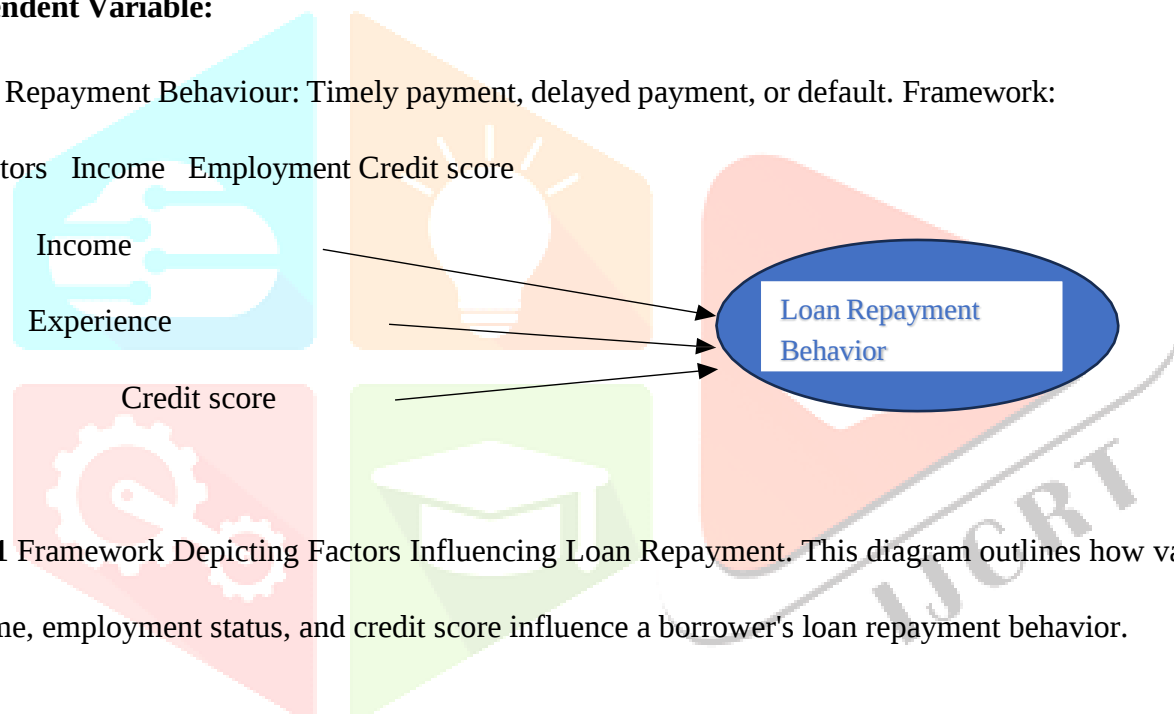


Fig. 1 Framework Depicting Factors Influencing Loan Repayment. This diagram outlines how variables like income, employment status, and credit score influence a borrower's loan repayment behavior.

CHAPTER 5

Research methodology:

Objectives:

This research employs a decision tree model to assess credit risk by analysing the impact of the debt-to-income ratio and various financial factors. The study aims to determine key predictors of loan default, enhancing risk evaluation in financial lending.

Data Collection:

This study utilizes secondary data obtained from Kaggle, a widely recognized platform for data science and machine learning datasets. The dataset titled [Credit risk] was accessed from Kaggle, ensuring a reliable and comprehensive data source for this research. making it suitable for the objectives of this study. The data was pre-processed and cleaned to ensure accuracy before analysis.

The dataset comprises financial records, including:

- Income
- Loan amount
- Credit history
- Homeownership status
- Debt-to-income ratio

Missing values were managed using appropriate imputation techniques to maintain data consistency and accuracy.

Source of dataset:

Model Development:

A Decision Tree Classifier was implemented using financial attributes to predict loan default probability. The following steps were undertaken:

- Data Preprocessing: Addressing missing values, encoding categorical data, and normalizing numerical features.
- Feature Selection: Identifying significant predictors influencing loan default.
- Model Training: Splitting data into training and test sets an 80:20 ratio.
- Evaluation Metrics: Measuring performance using accuracy, precision, recall, and feature importance analysis.
- Model mainly uses debt to income formula where it talks about ratio of your monthly income goes towards debt payments.

Formula:

$$\text{DTI Ratio} = (\text{Total Monthly Debt Payments} / \text{Gross Monthly Income}) * 100$$

Total monthly debt refers to all debt-related payments, including those for mortgages, rent, car loans, and other financial obligations.

Gross monthly income: This is your total monthly income.

Python Code Implementation

The following Python code was used to develop the decision tree model and visualize its structure:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score

# Load dataset
df = pd.read_csv('/content/credit risk data set (1).csv')

# Handling missing values
df_clean = df.dropna(subset=['Experience', 'loan_int_rate'])

# Compute Debt-to-Income Ratio (DTI)
df_clean['debt_to_income_ratio'] = df_clean['loan_amnt'] / df_clean['person_income']

# Encode categorical variables
label_encoders = {}
categorical_cols = ['person_home_ownership', 'loan_grade', 'loan_intent']
for col in categorical_cols:
    le = LabelEncoder()
    df_clean[col] = le.fit_transform(df_clean[col])
    label_encoders[col] = le

# Feature selection
features = ['person_income', 'Experience', 'cb_person_cred_hist_length',
            'cb_person_default_on_file', 'loan_percent_income', 'debt_to_income_ratio']
features.extend(categorical_cols)
target = 'loan_status'

# Split dataset into training and test sets
X_train, X_test, y_train, y_test = train_test_split(df_clean[features], df_clean[target], test_size=0.2, random_state=42)

X_train, X_test, y_train, y_test = train_test_split(df_clean[features], df_clean[target], test_size=0.2, random_state=42)

# Train Decision Tree Classifier
clf = DecisionTreeClassifier(max_depth=5, random_state=42, class_weight='balanced')
clf.fit(X_train, y_train)

# Model evaluation
y_pred = clf.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f'Model Accuracy: {accuracy * 100:.2f}%')

# Decision Tree Visualization
plt.figure(figsize=(20,10))
plot_tree(clf, feature_names=features, class_names=['No Default', 'Default'], filled=True, rounded=True, fontsize=8)
plt.savefig('decision_tree.png') # Save the visualization
plt.show()
```

Fig. 2 Visual Representation of the Decision Tree Model. The decision tree highlights key decision points such as debt-to-income ratio, loan grade, and income in predicting loan defaults.

Model Performance:

The decision tree classifier achieved an accuracy of 89.5%, demonstrating its strong ability to differentiate between defaulters and non-defaulters.

```
<ipython-input-2-5c3b30374690>:16: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
df_clean['debt_to_income_ratio'] = df_clean['loan_amnt'] / df_clean['person_income']
<ipython-input-2-5c3b30374690>:23: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
df_clean[col] = le.fit_transform(df_clean[col])
<ipython-input-2-5c3b30374690>:23: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
df_clean[col] = le.fit_transform(df_clean[col])
<ipython-input-2-5c3b30374690>:23: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
df_clean[col] = le.fit_transform(df_clean[col])
Model Accuracy: 89.51%
```

Fig.3. Python Code for Building the Credit Risk Model. Displayed code snippet shows the implementation of a decision tree classifier for assessing loan default risk using financial data.

Highly Influential Factors

Loan Grade (41%)

Lower grades indicate higher default risk. Debt-to-Income Ratio (33.9%)

Higher DTI ratios are strongly associated with loan default. Person's Income (12%)

Higher income correlates with reduced default probability. Home Ownership (7.5%)

Individuals owning homes exhibit lower default risk. Loan Intent (4.1%)

The purpose of the loan influences repayment behavior.

Less Influential Factors:

Credit History Length (<1%) Minimal effect on loan default risk. Past Default History (0%)

No notable influence on default likelihood. Loan Percent Income (0%)

Found to be an insignificant predictor in this dataset.

Decision Tree Visualization output:

The resulting decision tree structure reinforces the importance of debt-to-income ratio and loan characteristics in predicting loan default. The model's output serves as a tool for enhancing financial risk assessments.

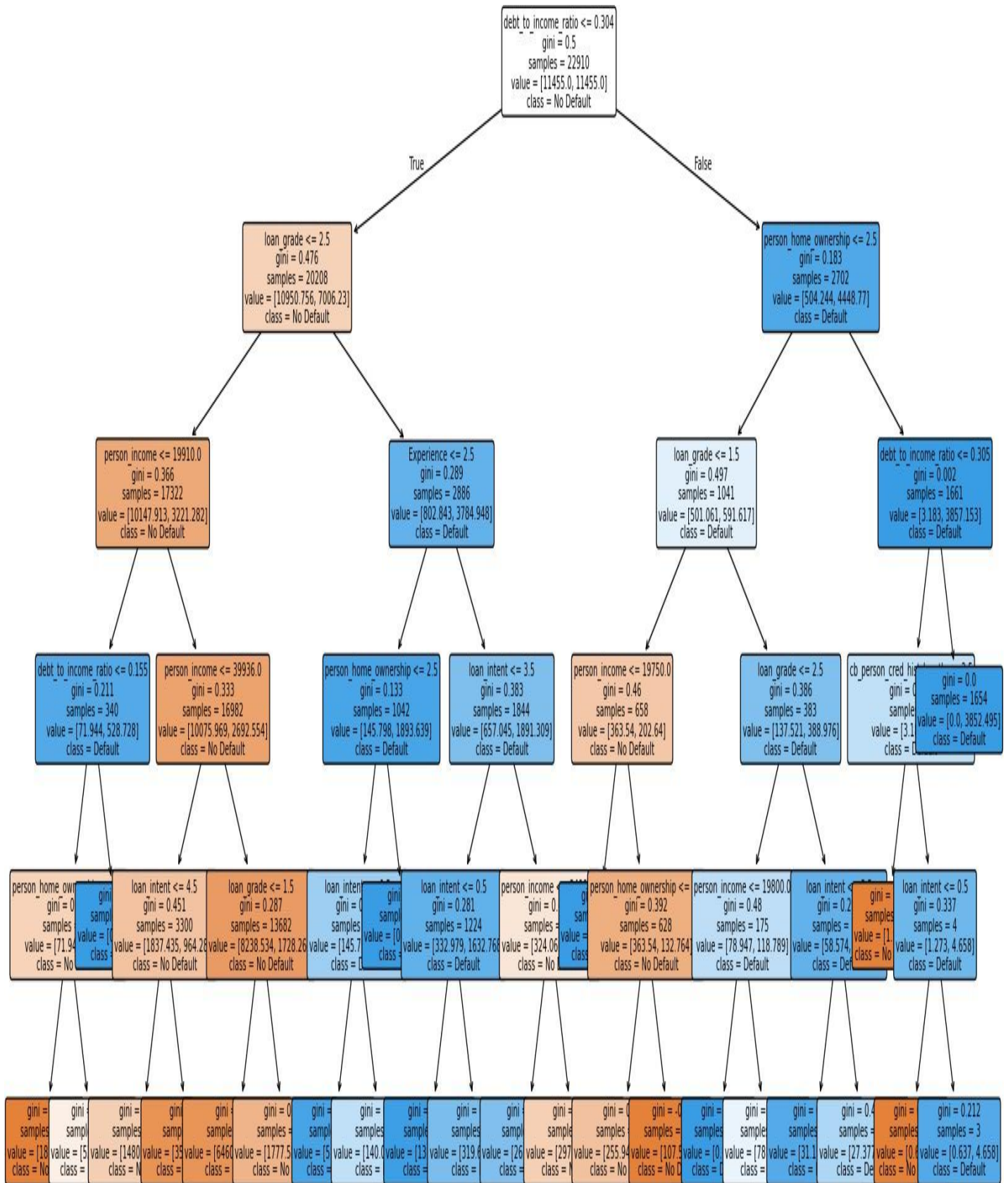


Fig.4: Gini Impurity Reduction Across Decision Tree Nodes, This chart demonstrates how Gini impurity decreases through the tree branches, improving the model's classification accuracy.

CHAPTER 6

Results & Discussion:

Debt-to-Income (DTI) Ratio:

At the root of the decision tree, the debt-to-income ratio emerges as the most influential predictor of loan repayment outcomes. Borrowers with lower DTI ratios (≤ 0.394) exhibit a stronger likelihood of timely repayments, whereas those with higher DTI ratios face an increased probability of default.

- Root Node: Debt-to-Income ratio (DTI)
- Branch Nodes: Loan Grade, Income
- Leaf Nodes: Default vs. No Default classification

Loan Grade:

Loan Grade plays a pivotal role in assessing credit risk. Individuals with higher loan grades (≤ 2.5) demonstrate more responsible repayment behavior, while those with lower grades are more susceptible to default.

Income Levels:

A borrower's income significantly affects loan repayment capacity. Higher income brackets (e.g., ≤ 19910 or ≤ 39936) correlate with a reduced risk of default, reinforcing the notion that financial stability enhances repayment reliability.

Homeownership:

Homeownership is another crucial factor, with homeowners generally showing lower default rates. This suggests that asset ownership reflects greater financial security.

Decision Pathways:

- Borrowers with **low DTI ratios** combined with **high loan grades** are strongly inclined to repay their loans punctually.
- Conversely, individuals with **high DTI ratios, poor loan grades, and lower incomes** are at a heightened risk of default.
- Notably, **experience** appears to have a secondary influence on loan repayment behavior, indicating that while financial literacy is relevant, concrete financial metrics like DTI, income, and loan grade carry greater weight in predicting defaults.

Gini Impurity Analysis:

- The Gini impurity decreases progressively as the tree branches out, signifying a rise in classification accuracy.
- Leaf nodes distinctly categorize borrowers into **default** (blue) and **no default** (orange) groups, highlighting the model's effectiveness in segmenting risk profiles.

Implications:

Lenders: This model underscores the importance of scrutinizing DTI ratios, loan grades, and income levels during the loan approval process to mitigate default risks.

Borrowers: The findings suggest that improving one's DTI ratio, securing a stable income, and maintaining a strong credit score are key strategies for boosting loan approval chances.

Risk Assessment Models: The relatively lower weight given to borrower experience implies that financial institutions should prioritize quantitative financial data over qualitative factors when evaluating loan applications.

This analysis not only enhances the understanding of loan default dynamics but also offers practical insights for both lenders and borrowers to effectively manage the complexities of financial risk.

Debt to income ratio calculator:

- Calculator helps individuals assess their financial health by determining The portion of their income allocated to debt repayment.
- Easy to assess their financial performance and exposure to loan burden.
 - Lenders use this calculator to assess whether debtor qualifies for loan of credit. Python code for creating a calculator:

```
def calculate_dti():  
    print("Debt-to-Income (DTI) Ratio Calculator")  
    print("-----")  
  
    try:  
        # Get user input for monthly income  
        income = float(input("Enter your gross monthly income: ₹"))  
  
        # Validate income  
        if income <= 0:  
            print("Error: Income must be greater than 0")  
            return  
  
        # Get user input for monthly debt payments  
        print("\nEnter your monthly debt payments:")  
        mortgage = float(input("Mortgage/Rent payment: ₹"))  
        car_loan = float(input("Car loan payment: ₹"))  
        credit_card = float(input("Credit card payment: ₹"))  
        student_loan = float(input("Student loan payment: ₹"))  
        other_debt = float(input("Other debt payments: ₹"))  
  
        # Calculate total monthly debt  
        total_debt = mortgage + car_loan + credit_card + student_loan + other_debt  
  
        # Validate debt  
        if total_debt < 0:
```

```

        print("Error: Debt payments cannot be negative")
        return

    # Calculate DTI ratio
    dti_ratio = (total_debt / income) * 100

    # Display results
    print("\nResults:")
    print(f"Total Monthly Debt: ₹{total_debt:.2f}")
    print(f"Gross Monthly Income: ₹{income:.2f}")
    print(f"Your Debt-to-Income Ratio: {dti_ratio:.2f}%")

    # Provide interpretation
    print("\nDTI Ratio Interpretation:")
    if dti_ratio <= 36:
        print("Good - Your DTI is at a manageable level")
    elif 36 < dti_ratio <= 43:
        print("Moderate - You may face some lending restrictions")
    elif 43 < dti_ratio <= 50:
        print("High - You might have difficulty getting approved for loans")
    else:
        print("Very High - Consider reducing debt before applying for credit")

except ValueError:
    print("Error: Please enter valid numerical values")
except Exception as e:
    print(f"An error occurred: {e}")

# Run the calculator
if __name__ == "__main__":
    calculate_dti()

```

Fig.5: User Interface of the Debt-to-Income (DTI) Calculator. The calculator interface helps evaluate the portion of income allocated to debt, supporting financial health assessment.

Example output:

Calculators give output based on these factors: Monthly debt payments:

- Rent/mortgage payments
- Car loan
- Student loan payments
- Other debt payments

```

Debt-to-Income (DTI) Ratio Calculator
-----
Enter your gross monthly income: ₹50000

Enter your monthly debt payments:
Mortgage/Rent payment: ₹12500
Car loan payment: ₹8500
Credit card payment: ₹4000
Student loan payment: ₹0
Other debt payments: ₹14000

Results:
Total Monthly Debt: ₹39000.00
Gross Monthly Income: ₹50000.00
Your Debt-to-Income Ratio: 78.00%

DTI Ratio Interpretation:
Very High - Consider reducing debt before applying for credit

```

Fig.6: Sample Output from DTI Calculator Analysis. The output example interprets the user's debt burden and suggests loan eligibility based on DTI values.

- It interprets according to your debt payments and income and suggests whether you can avail yourself of the loan or not and talks about your financial distress.

CHAPTER 7

Conclusion:

This study on evaluating credit risk for loans (housing, car, and personal) at Canara Bank through machine learning techniques offers significant insights into effective risk management within a constantly evolving financial environment. Utilizing a decision tree classifier, the research pinpointed critical financial elements affecting loan repayment patterns, with the debt-to-income (DTI) ratio standing out as the primary indicator, followed by loan grade and income levels. The model delivered a commendable accuracy of 89.5%, proving its effectiveness in separating borrowers likely to default from those who are not. Homeownership also emerged as a key factor, indicating financial solidity, whereas factors such as experience and credit history length had a lesser impact compared to measurable financial data.

The main aim of this research was to examine how financial factors influence credit risk, identify high-risk borrowers, and predict loan repayment behavior. While traditional credit assessment methods remain useful, they often struggle to account for the complexity of borrower behavior and financial stability. In contrast, machine learning techniques provide a more advanced, data-driven approach, capable of analyzing large datasets, uncovering complex patterns, and creating predictive models with greater accuracy. By integrating such methodologies, this study has demonstrated how banking institutions can improve their credit risk management processes, ultimately reducing default rates and mitigating financial losses.

The findings demonstrate how machine learning can boost the efficiency of loan approval workflows by supporting swift, evidence-based decisions and lowering default occurrences. For Canara Bank, this method provides a forward-thinking strategy to spot high-risk borrowers early, enabling customized solutions like loan adjustments or intensified oversight. In addition to reducing risks, the study enhances customer satisfaction by simplifying loan procedures and building confidence through consistent and clear decision-making processes. These outcomes not only fortify Canara Bank's approach to managing credit risk but also present a replicable framework that could benefit the wider banking industry in India, especially in tackling issues related to non-performing assets (NPAs) and economic volatility.

Scope for future study:

Although this study offers significant insights into credit risk management through machine learning, several aspects require further exploration. Future research can be investigated:

- **The Influence of Macroeconomic Variables** Examining the effects of external economic factors, including inflation rates, unemployment levels, and interest rate variations, on credit risk assessment and management
- **Integrated Credit Risk Models:** Designing and assessing hybrid models that merge conventional credit scoring methods with machine learning techniques to improve predictive performance and accuracy.
- **Behavioral Factors in Credit Risk:** Investigating the psychological and behavioral influences on borrower repayment patterns, incorporating behavioral finance concepts into machine learning-based risk assessment models.
- **Exploration of Sophisticated Algorithms:** While the decision tree classifier yielded solid results, testing more advanced techniques like Random Forest, XGBoost, or neural network-based deep learning models could elevate precision and manage larger, intricate datasets. Combining multiple models through ensemble approaches could also be examined for improved outcomes.
- **Enhancing Borrower Experience:** Exploring how these models can be tailored to offer individualized financial guidance such as tips to reduce DTI ratios could empower borrowers and improve satisfaction, aligning with the study's secondary goals.

References:

1. **Reducing NPAs Across Loan Categories**
 - Haralayya, Dr. Bhadrappa. (2022). *Trends in Non-Performing Assets (NPAs) across various loan sectors.*
2. **Operational Risks and AI in Credit Appraisal**
 - Kumar, Mani Bhushan. (2023). *Regulatory compliance and risk management in Indian banking.*
 - Mulabagula, M. Geeta et al. (2024). *The role of AI and ML in modern credit appraisal processes.*
3. **Machine Learning Models for Operational Efficiency**
 - Joshi, Ankur et al. (2022). *Predicting operational efficiency using LogitBoost and Cart algorithms.*
4. **Decision Trees in Loan Risk Classification**
 - Hamid & Ahmed. (2016). *J48 Decision Tree Algorithm for Loan Risk Classification.*
5. **Enhancing Banking Credibility Through CSR**
 - Belas et al. (2021). *CSR and its impact on banking credibility and customer retention.*
 - Egri & Ralston. (2021). *Financial Literacy and Credit Counseling (FLCC) initiatives.*

6. AI's Potential in Banking

- Alzaidi. (2018). *Adoption of AI in Middle Eastern Banks.*
- Malali & Gopalakrishnan. (2021). *AI transformations in wealth management and virtual assistants.*

7. Mental Health Factors in Credit Risk

- Alagic, Adnan et al. (2024). *The impact of mental health and age on loan approval predictions.*

8. Predicting Default Risks with Deep Learning

- Dendramis, Yiannis et al. (2021). *Artificial neural networks (ANN) in predicting small business loan defaults.*

9. Advanced Models for Loan Risk Analysis

- Hamzic, Dzenan et al. (2024). *Machine learning applications in credit risk assessment (XGBoost, AdaBoost, Gradient Boosting).*

10. Applications of ANN and Decision Trees in Risk Prediction

- Lakhani, Mohit et al. (2021). *Deep learning models in mitigating financial losses in lending institutions.*

11. Evolution and Investment Trends in Artificial Intelligence

- Singh, Kamal. (2020). *Market trends in AI-driven financial services.*

12. Financial Performance Determinants in Housing Finance Companies

- Mini P G, Joseph. (2020). *Multivariate discriminant analysis in housing finance companies.*

13. Community-Based Housing Finance Models

- Yap, K. S. & Wandeler, K. de. (1994). *Effectiveness of chit funds in grassroots financial systems.*

14. Gaps in Housing Finance for Rural Populations

- Mahadeva, M. (2024). *Disparities in housing finance access between urban and rural sectors.*

15. Trends in Housing Loan Disbursement

- Chandrasekar, Kumaraswamy, & Nayan. (2023). *Comparative analysis of Scheduled Commercial Banks (SCBs) and Housing Finance Companies (HFCs).*