Loan default prediction using Artificial Intelligence and Machine Learning

Kiran Bhausaheb Shinde

Date – 21/06/2023

"Some people call this artificial intelligence, but the reality is this technology will enhance us. So instead of artificial intelligence, I think we will augment our intelligence."

– Ginni Rometty

1.Abstract

With the increasing complexity and volume of financial data, the traditional methods used for loan default prediction have become less effective. In recent years, there has been a growing interest in **utilizing Artificial Intelligence (AI) techniques to enhance the accuracy and efficiency in financial systems** across the globe. AI offers the potential to extract valuable insights from large datasets, identify hidden patterns, and make more accurate predictions.

This report aims to provide an overview of the use of AI in loan default prediction and utilizing it to run as a successful business model. I explore various AI techniques, such as machine learning algorithms, deep learning models, and natural language processing, that should be applied for better prediction of potential loan defaults.

2.Problem Statement

The Reserve Bank of India (RBI) has reported a sharp decline in gross non-performing assets (NPAs), or bad loans, in the banking system in the last two years, **but wilful defaults have shot up** with more legacy loan accounts now getting **added to the wilful default category**. There has been a rise **of 38.50 per cent, or Rs 94,000 crore, in wilful defaults in the last two years,** reflecting the gaps in loan appraisals and risk management in the banking sector.

The condition across the world is the same. The reasons behind the loan defaults are complex, however because of increase in technology in financial sector furthermore made situation complex.

As mentioned above the upcoming technologies of Machine Learning, AI, Data Science and NLP would be essential to reduce the loan defaults. This technologies by using previous data of the costumer can predict the loan defaulters in advance.

3.Market/Customer/Business Need Assessment

Since the advent of **AI, its uses are multidimensional**. The AI can be utilized to predict everything under the sun, from predication of sun spots cycles to observing the patterns of microorganisms. There never has been a better decade than the current one to invest & leverage AI to its best & for as many needs as wanted, needed. No matter what field AI is used in, it never ceases to surprise us.

For instance, **utilizing AI in Loan screening process would help**-

- Better Application screening
- Better assessment of borrower's credit history and credit score to determine their past repayment behaviour and overall creditworthiness
- Better financial assessment where lenders can evaluate the borrower's financial position by analysing income, employment history, assets, and liabilities.
- Better assessment of Loan to value ratio of a collateral loans.
- Can be utilize in risk assessment, wherein lenders evaluate the overall risk associated with the loan application, considering factors such as creditworthiness, financial stability, collateral, and loan amount.
- Most importantly AI can be a better tool for decision makers in Banking system for enhancing their decision-making ability.

However, my model of Machine learning would help the lenders in banking system for better prediction of borrower's future behaviour. Whether he/she will default or not.

By using **following data sets of borrowers-**

- Credit history
- Type of employment
- Monthly income
- Annual income
- Other income

2

- Repayment of previous loan instalments
- Family credit history. Etc.

4.Target Specifications and Characterization.

The **traditional market and customer base for this business model would be**-

- **Commercial Banks**- Commercial banks are the most common type of lenders in the banking system. They accept deposits from customers and provide a wide range of loan products, including personal loans, mortgages, business loans, lines of credit, and more.
- **Credit Unions**- Credit unions are member-owned financial cooperatives that offer similar services to commercial banks. However, credit unions are typically more community-oriented and have membership requirements. They often provide loans at competitive rates and may have a focus on specific groups
- **Microfinance Institutions**- Microfinance institutions (MFIs) are specialized lenders that provide small loans, often referred to as microloans, to individuals and businesses in underserved communities or developing countries.
- **Non-Bank Financial Institutions (NBFI)-** Non-bank financial institutions (NBFI) are entities that offer financial services and products like traditional banks but do not hold a banking license. They include entities such as finance companies, leasing companies, and factoring companies. NBFI's focus may be on specific sectors or niche markets, and they often provide specialized lending solutions. etc.

However, as mentioned earlier with the advent of internet **the new players in banking system can also utilize this business model such as**-

- **Neo banks**- Neo banks, also known as digital banks or challenger banks, are a type of new banking entity that operates entirely online without physical branches. They leverage technology and innovative business models to provide a range of banking services and products to customers.
- **Online Lenders**- With the advancement of technology, online lenders have emerged as a popular alternative to traditional brick-and-mortar lenders. These lenders operate exclusively online, allowing borrowers to complete loan applications and receive funds digitally.
  Etc.

Other than that, it can be used by –

- **Government** for formulation of better reforms in banking system to reduce NPA's.
- **Credit rating agencies**- For better assessment of the creditworthiness and risk associated with issuing debt securities or borrowing entities, such as corporations, governments, or financial institutions.

Many other upcoming customer bases for the model can be utilize.

5-Bench marking alternate products- AI in banking system

- **SIA Chatbot** -From a customer chatbot perspective, SBI has launched SIA, an AI-powered chat assistant that addresses customer enquiries instantly and helps them with everyday banking tasks just like a bank representative
- **Robotic software**-ICICI Bank, India's second-largest private sector bank has deployed software robotics in over 200 business processes across various functions of the company
- **Thought Factory** of AXIS bank- to accelerate the development of innovative AI technology solutions for the banking sector.
- **Wei et al.** put forth a theoretical assessment of the impact that AI-based models (using big data generated through social media activity) might have on the quality of credit scores. They conclude that they might backfire, by inducing strategic changes in the patterns of activity on social media platforms by potential borrowers.
- **Óskarsdottir et al.** have modelled the odds of default on credit card debt, by using detailed mobile phone statements to reconstruct the social connections of cardholders.

6.Applicable Constraints

The use of financial data of individuals is subject to **various constraints** to ensure privacy, data protection, and prevent misuse. While the specific constraints may vary by country and region, here are some common constraints to consider-

- **Consent**- Generally, organizations must obtain the explicit consent of individuals before collecting, storing, and using their financial data. Consent should be informed, specific, and obtained through clear and transparent communication.
- **Purpose Limitation**- Organizations should collect and use financial data for legitimate and specified purposes only. They must not use the data beyond what was initially disclosed to the individual without obtaining additional consent or a legal basis.
- **Data Minimization**- Organizations should collect and process only the necessary financial data required to fulfil the intended purpose. Unnecessary or excessive data collection should be avoided.
- **Data Security**- Organizations must implement appropriate security measures to protect the financial data they collect. This includes measures to prevent unauthorized access, data breaches, and ensure the integrity and confidentiality of the data.
- **Data Retention**- Financial data should not be retained for longer than necessary to fulfil the purpose for which it was collected, unless required by law or for legitimate reasons. Organizations should establish data retention policies and securely dispose of data once it is no longer needed.
- **Data Sharing and Transfer**- When sharing or transferring financial data to third parties, organizations should ensure appropriate safeguards and contractual arrangements are in place to protect the data. This may include data protection agreements, confidentiality clauses, or ensuring the recipient is subject to similar data protection regulations.

- **Legal and Regulatory Compliance**- Organizations must comply with applicable laws, regulations, and industry standards related to the use of financial data. This includes data protection laws, financial sector regulations, and consumer protection regulations specific to financial services.
- **Individual Rights**- Individuals have rights over their financial data, including the right to access, rectify, restrict processing, and erase their data under certain circumstances. Organizations must provide mechanisms for individuals to exercise their rights and respond to their requests promptly.

7.Applicable Regulations

In India, financial data privacy is regulated primarily by the following **laws and regulations**-

- **The Reserve Bank of India (RBI) Guidelines**- The RBI issues guidelines and directives to regulate the collection, storage, and sharing of financial data by banks and other financial institutions. These guidelines outline the requirements for customer data privacy, data protection, and confidentiality.
- **The Information Technology Act, 2000**- The Information Technology Act, along with its amendments, provides a legal framework for electronic transactions and data protection in India. It includes provisions related to data privacy, security, and the punishment for unauthorized access and disclosure of financial data.
- **The Personal Data Protection Bill, 2019**- India has proposed comprehensive data protection legislation, known as the Personal Data Protection Bill, which is currently under review. Once enacted, it will establish rules and obligations for the collection, processing, and transfer of personal and financial data by entities in India.
- **The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016**- The Aadhaar Act governs the collection, storage, and usage of Aadhaar, a unique identification number issued to residents of India. It includes provisions related to privacy and security safeguards for Aadhaar data, including financial information linked to Aadhaar.

At the **global level, several regulations and frameworks** address financial data privacy. Here are some significant ones-

1. **Basel III**-Basel III is a global regulatory framework for banking, developed by the Basel Committee on Banking Supervision. While it primarily focuses on capital adequacy and risk management, it includes provisions related to the confidentiality and protection of customer financial data held by banks.
2. **Payment Card Industry Data Security Standard (PCI DSS)**- PCI DSS is a global security standard developed by major payment card companies to protect cardholder data. It applies to organizations that handle payment card information, including financial data, and outlines requirements for data security, encryption, and network protection.
3. **General Data Protection Regulation (GDPR)**- The GDPR is a comprehensive data protection regulation that applies to all European Union (EU) member states. It establishes rules for the processing and transfer of personal data, including financial data, by entities within and outside the EU.

8. Business Model (Monetization Idea)
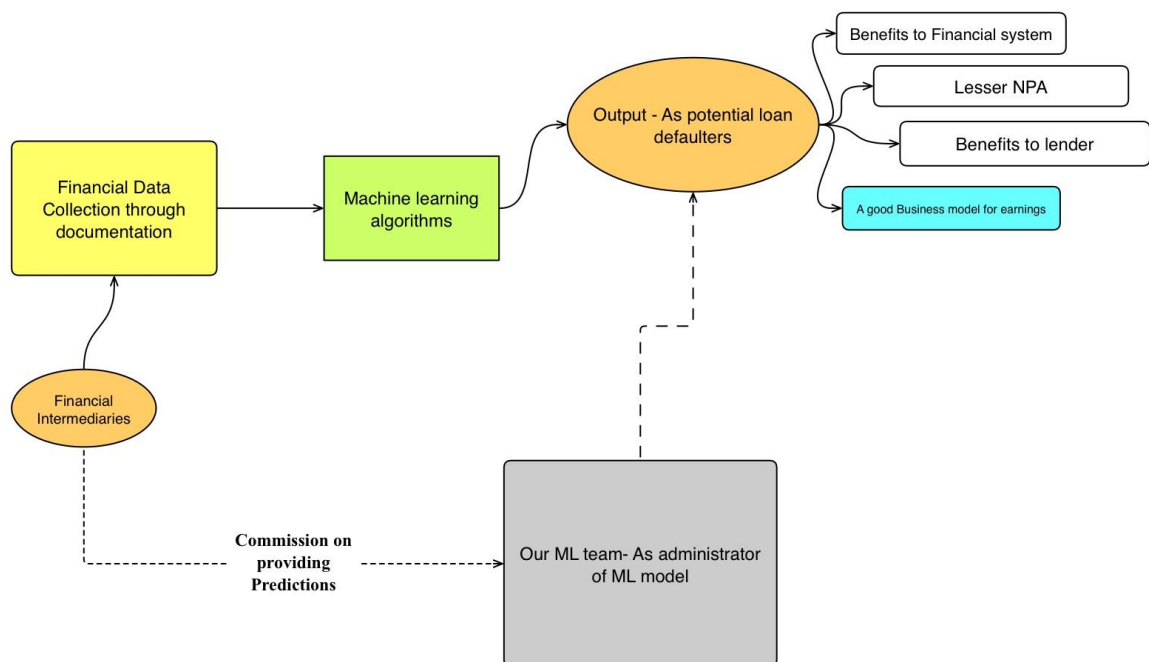
9. Final Product Prototype

To begin with the ML based predictions of potential, **we must follow the following steps**-

- First, the **primary collection of data** of borrowers is done by financial intermediaries (FI's). The ways to collect data is through documentation, credit rating agencies etc. The following data can be collected-
  - Credit history
  - Type of employment
  - Monthly income
  - Annual income
  - Other income
  - Repayment of previous loan instalments
  - Family credit history
      Etc.
- Followed by **preliminary analysis** of given data of everyone.
- Then, the data is forwarded by FI's to our ML team.
- ML team will prepare data also it will **clean data** to perform other functions.
- Then **data is processes using python libraries** such as Pandas, Seaborn & NumPy.
- Next step for the prototype would be **data exploration.**
- Followed by **feature engineering**. In this step we can use various financial ratios to further explore the data such as Loan to Value ratio, Debt coverage ratio etc.
- The final step of the model is **to select the appropriate machine learning model**-
  - Logistic regression

- o Random forest
- o Decision tree
- o Gradient boost classifier

Above mentioned models would be a good option for the prediction of the potential loan defaulters.

Based on the accuracy, we can select the appropriate model for the prediction.

- A good accurate model of machine learning would predict the defaulters, **this information can be forwarded** to the lenders to further strengthen their strategies of loan recovery.
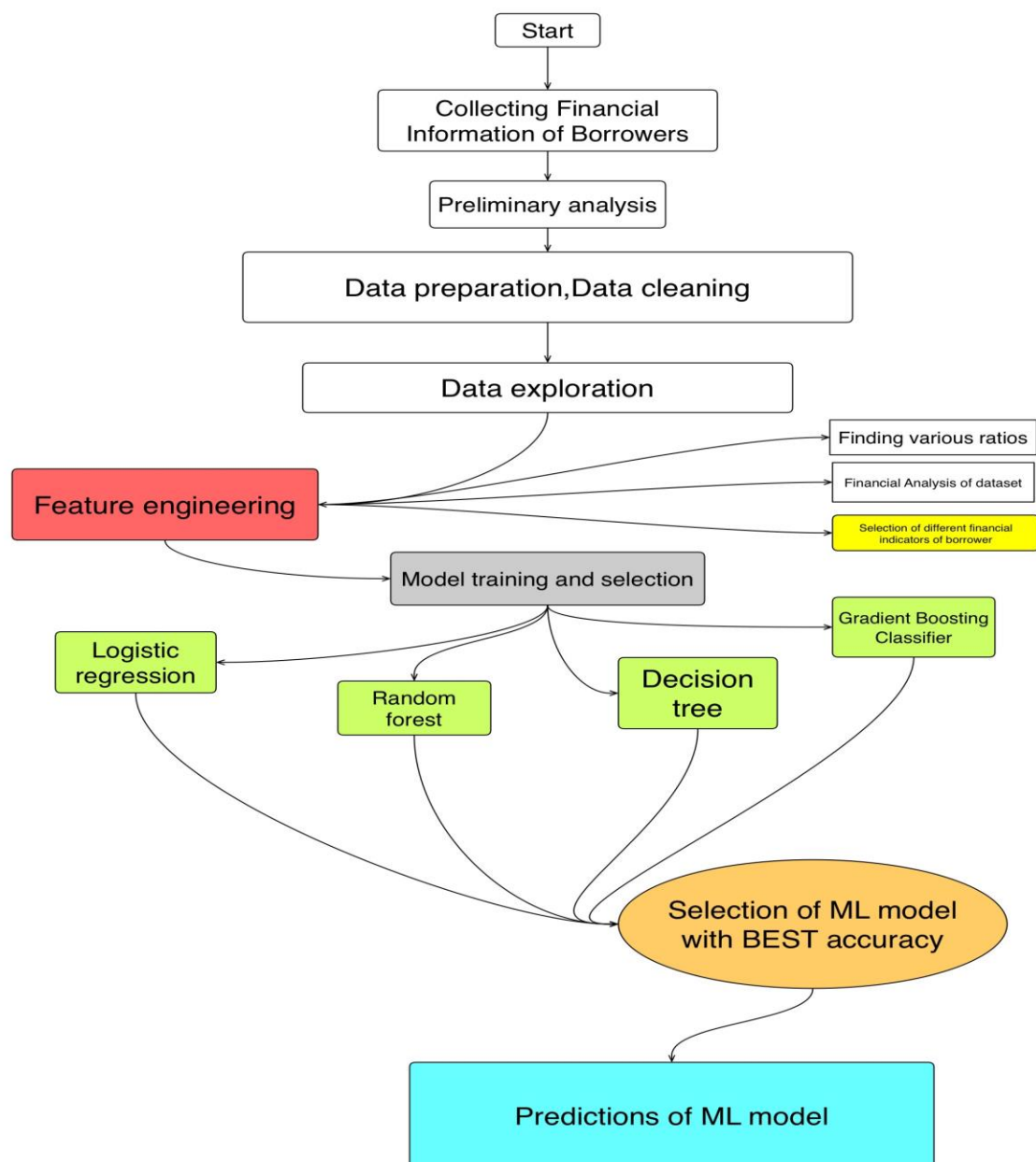
10.Conclusion

Machine learning (ML) models have revolutionized various industries, including finance, by providing advanced predictive capabilities and insights. ML models, such as those used in loan default prediction, leverage vast amounts of historical data and powerful algorithms to assess credit risk and make informed decisions. These models offer several advantages, including improved accuracy, automation of decision-making processes, and the ability to handle complex patterns and relationships in the data. By analysing a wide range of factors, such as financial data, credit history, and borrower characteristics, ML models can effectively assess the likelihood of loan default and assist financial institutions in managing risk.

However, it is important to note that ML models are not infallible and should be used as tools to support decision-making rather than as standalone solutions. The performance and reliability of ML models depend on the quality and relevance of the data used for training, as well as careful feature selection, model validation, and ongoing monitoring and maintenance.