REVA UNIVERSITY
Bengaluru, India

A Project Report on

Investigating Super learner for Credit Risk Modeling in Mortgage Scenario

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration
In Business Analytics

Submitted By
Lalit Aggarwal
R19DM003

Under the Guidance of
Ravi Shukla
Consultant, Data Science at Dell Technologies Pvt. Ltd

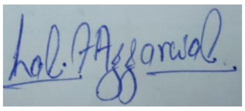
REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Lalit Aggarwal** hereby declare that I have completed the project work towards the second year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled **Investigating Super learner for Credit Risk modeling in Mortgage Scenario** under the supervision of **Ravi Shukla, Consultant, Data Science at Dell Technologies Pvt. Ltd.**This report embodies the original work done by me in partial fulfillment of the requirements for the award of a degree for the academic year **2022.**



Place: Bengaluru | Name of the Student: Lalit Aggarwal

Date: 27th Aug 2022 | Signature of Student

REVA UNIVERSITY
Bengaluru, India

Certificate

This is to Certify that the Project work entitled **Investigating Super learner for Credit Risk modeling in Mortgage Scenario** carried out by **Lalit Aggarwal** with **R19DM003,** is a bonafide student of REVA University, is submitting the second year project report in fulfillment for the award of **Master of Business Administration in Business Analytics** during the academic year **2022**. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said Degree.

Ravi Shukla

Signature of the Guide	Signature of the Director
Name of the Guide	Name of the Director
Ravi Shukla	**Dr. Shinu Abhi**

External Viva

Names of the Examiners

1. Dr. Santosh Nair, Founder, Analytic Edge
2. Rajkumar Dan, Data Scientist Consultant, Dell

Place: Bengaluru

Date: 27th Aug 2022

Acknowledgment

Among many people who helped my research journey on virtual try-on to be a success, without whom I would not have managed to complete this research nor would I have accomplished my MBA.

I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, **Dr. P Shayma Raju**, Pro-Vice Chancellor, **Dr. M. Dhanamjaya**, Registrar, **Dr. N. Ramesh.**

My special thanks go to **Dr. Shinu Abhi, Director of REVA Academy of Corporate Excellence** for believing in me and for providing me with all the guidance and support throughout my MBA program.

I am grateful to my mentor **Ravi Shukla, Consultant, Data Science at Dell Technologies Pvt. Ltd.** for all his help and guidance in the project and for showing me a path I could walk.

The team of mentors at REVA Academy of Corporate Excellence, including **Dr. J. B. Simha, Ratnakar Pandey, Pradeepta Mishra, MithunD.J.,** and others, taught us some awesome methodologies throughout the academic year. I would like to extend my gratitude to the educational department at Reva Academy of Corporate Excellence, REVA University campus, for providing ample opportunities and guidance.

Throughout the project work, I was grateful for the support, suggestions, and friendly advice offered by my classmates.

Lastly, I want to thank my family for supporting me through this research process and for believing in me.

Place: Bengaluru

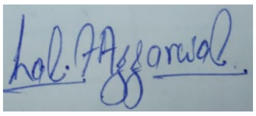
Date: 27th Aug 2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **Investigating Super learner for Credit Risk modeling in Mortgage Scenario** was scanned for similarity detection. Process and outcome are given below.

Software Used: **Turnitin**
Date of Report Generation:
Similarity Index in %: 11%
Total word count: **8,914**
Name of the Guide: **Ravi Shukla**



Place: Bengaluru
Date: 27th Aug 2022

Name of the Student: **Lalit Aggarwal**
Signature of Student

Verified by: **Dincy Dechamma**

Signature
Dr. Shinu Abhi,
Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	GBM	Gradient Boosting Machine
2	GLM	Generalized Linear Models
3	DRF	Distributed Random Forest
4	RF	Random Forest
5	XRT	Extremely Randomized Trees
6	DL	Deep Learning
7	SHAP	Shapley Additive Explanations
8	ICE	Individual Conditional Expectations
9	PDP	Partial Dependence Plot
10	AutoML	Automated machine learning

List of Figures

Abstract

In the present industry Credit risk analysis is very important for the organization's business as well as its reputation in the market. In general, credit risk modelling is a method that lenders employ to assess the degree of credit risk involved in making a loan to a borrower. When a business or individual borrower doesn't fulfil their loan obligations, credit risk develops. It is the likelihood that a lender won't get the principal and interest payments needed to pay off a loan that has been issued to a borrower. Credit risk will affect the lender's cash flow and raise collection expenses since it may be required to work with a debt collection agency to enforce the collection. The lender may suffer a loss of all or a portion of the loan that was given to the borrower, depending on the circumstances.

The turnaround time for delivering a statistical model in the general machine learning approach is somewhat long. Creating machine learning models that can reliably anticipate the outcomes is always difficult and time-consuming, even for expert data scientists. There are several stages to the intricate workflow that goes into machine learning models. While Automatic Machine Learning (AutoML) is more concerned with gathering data and making predictions. The AutoML platform will abstract every step that occurs between these two phases. AutoML models require extremely little time to develop models and tune hyperparameters. Prediction can also be extremely well conveyed with the use of various outcome interpreter approaches.

The objective of this project is to investigate Super learners for Credit Risk modeling in Mortgage scenarios, with the help of Automation Machine learning. Different super learners can be defined as the ensemble models of different base models and investigated against the credit risk dataset. For explaining the prediction of the super learner different result interpretation techniques i.e. SHAP, PDP and ICE have been used.

With the Use of H2O AutoML and the credit risk data, multiple machine learning models have been used, which comprise 25 models including Ensemble machine learning models “StackEnsemble_BestOfFamily” and “StackEnsemble_AllOfFamily” as well as base statistical machine learning models DeepLearning, DRF, XRT, GBM, and GLM, here define

AUC is defined as a stopping metric. Out of that "StackEnsemble_BestOfFamily" was giving an AUC of 71.08% and in base models, DRF was giving the highest Accuracy 88.02%, and an AUCof 70.5%. While in the interpretation SHAP, PDP, and ICE techniques gave a very good explanation for every individual result as well as a prediction of the whole data set.

Hence with the help of AutoML techniques, multiple machine learning models have been created in a very short time without wasting much time on data preparation, Data exploration, Feature engineering, model selection, model training, and hyperparameter tuning. With the help of SHAP, PDP, and ICE any individual result could be explained to the customer or the end-user.

Keywords:* *AutoML,* *metalearner,* *StackEnsemble_BestOfFamily, StackEnsemble_AllOfFamily, DeepLearning, DRF, XRT, GBM, GLM, AUC, SHAP, PDP, and ICE.

Contents

Chapter 1: Introduction

Credit risk is the likelihood that the borrower won't be able to pay back his debt on time or in whole. It refers to the possibility that the lender won't get the principal loaned or the interest due in a timely manner. This has the result of interfering with creditors' cash flow and raising the cost of collection. In dire circumstances, it can be necessary to write off all or a portion of the debt, incurring a loss for the lender. Knowing with absolute certainty how likely someone is to fail on a debt is very tough and complex. A competent evaluation of credit risk can also lessen the chance of losses due to default and late payments. The lender receives interest payments from the borrower as compensation for taking on the credit risk. Lenders or investors will either charge a higher interest rate or decline the loan opportunity entirely if the credit risk is higher. For instance, an applicant for the same loan with excellent credit and consistent income will pay a lower interest rate than one with a poor credit history (Team, 2022).

A person's credit risk is influenced by a wide range of variables. As a result, determining the borrower's credit risk is an extremely challenging undertaking. Credit risk modelling has become important since so much money depends on our ability to predict a borrower's credit risk accurately. Credit risk modelling is the practise of utilising data models to ascertain two crucial facts. The likelihood that the borrower will miss a loan payment is the first. The second is how this default will affect the lender's financial situation (Christian Bluhm, 2010).

To assess the credit risk of potential borrowers, financial institutions use credit risk models. Based on the validation of the credit risk model, they determine whether or not to authorise the loan and the loan's interest rate. With the development of technology, new approaches to modelling credit risk have appeared, such as credit risk modelling in R and Python. These include modelling credit risk utilising the newest analytics and big data tools. The way credit risk is modelled has also been impacted by other variables, such as the growth of economies and the ensuing rise of various types of credit risk (Tomasz R. Bielecki, 2013).

Financial organisations have created sophisticated methods for calculating and controlling credit risk across all product categories of companies. A good understanding of frequently employed methods would, in the eyes of a regulator, increase the oversight of financial institutions. The necessity to calculate the amount of capital required to support the bank's exposures initially sparked interest in credit risk models. The job of the credit risk model is to take general economic conditions and specific variables as inputs and produce a credit spread as an output. There are two primary categories of credit risk models in this regard: structural and scale models. Based on the value of a company's assets and liabilities, structural models are used to determine the likelihood that the company would fail. If a company's assets are worth less than the amount of debt that must be repaid, it goes into default. Reduced models consider a random, exogenous source of failure (Omomehin, 2021).

With the help of the credit risk modelling methodology, risk measurement and management may be done in a flexible and personalised way. Models respond to changes in business lines, credit quality, market factors, and the economic climate as a result of their design and are influenced by these changes (J.Teece, 2018).

Additionally, the models reflect concentration risk within the portfolio and give banks the ability to examine marginal and absolute risk contributions. These model characteristics can help the bank's overall credit culture. The degree to which credit management has embraced models. Between banks, there are significant differences in how economic capital is distributed. While some banks have put in place systems to capture the majority of exposures across the organisation, only others do so for a specific business line or legal entity. In addition, banks frequently create distinct models for corporate and retail exposures, and not all banks account for both kinds of exposures. Internal model applications range from straightforward to straightforward and sophisticated (BCBS, Calculation of RWA for credit risk, 2020).

Only a tiny portion of the banks the working group investigated already use the results of credit risk models in active portfolio management, but a sizable portion stated they intended to do so in the future. Applications now used include:

a)Calculating exposure and concentration limits;
b) Defining holding objectives for syndicated loans;
c) Pricing based on risk;
d) Enhancing the portfolio's risk and return profile;

e) Evaluation of business lines' or managers' risk-adjusted performance using risk-adjusted return on equity ("RAROC")

f) Capital allocation in the economy. Institutions also use model estimates for direct computations or for validation purposes to generate or confirm reserves for loan losses. (Chen, 2021).

The Working Group acknowledges that improved internal risk management in banking organisations can be a result of credit risk modelling. However, before employing the models in the process of establishing regulatory capital needs, the key challenges related to data limitations and model validation must be overcome. Description of the default process and other elements that affect loans The absence of previous credit and loan performance data as well as other modelled variables greatly restricts quality. Since credit risk is measured across lengthy time horizons, which means the data span many years, the specification issues become more severe. To accurately estimate important factors, several credit cycles can be necessary. As a result of the present constraints, model parameters frequently involve some degree of simplification of assumptions and information collection from various sources. Since doing sensitivity testing of the model's vulnerability to such assumptions is not currently conventional practise, the impact of these alternatives on model risk estimation is uncertain (Bessis, 2010).

Chapter 2: Literature Review

In a commercial bank, credit management is a challenging functional area. It requires skilled handling, accurate risk assessment at each level, and sufficient assurance of the security of monies exposed. Despite greatest attempts, it is unable to create solid security requirements, which causes credit to be unstable or unpaid on a frequent basis. Therefore, effective risk management, asset and liability management techniques, and continual search for safer criteria for risk elimination constitute credit management. Such risk management expertise has been developed and put into use to help mitigate risk rather than eradicate it (Circulars, 2008).

Credit risk is typically understood as the risk of default, or the risk of losing money if the borrower or counterparty fails to pay the bank the amount owed (principal or interest) according to a pre-arranged repayment schedule, on time. Value risk, or the danger of losing value as a result of the borrower's shift to a lower credit rating (opportunity costs connected with incorrectly pricing a new loan risk level), would be added to the definition to make it more complete.(Amendment to the capital accord to incorporate market risks, Basel Committee on Banking Supervision, 2005).

Banks have devised procedures that enable them to quantify these risks and so derive the necessary amount of capital to maintain their company, known as economic capital, in order to hedge against volatility in default/impairment levels (as well as other types of risk).The method for determining the required minimum regulatory capital is outlined in Pillar 1.Basel-I stated that this computation only took into account credit risk; however, in 1996, a calculation for market risk was introduced. Basel II increases the price to reflect operational risk.

A lot of effort needs to be done at the bank level as the entire financial sector worldwide works to implement the 2004 directive-based II Accord in some shape and intensity. Credit Risk Management gives users the tools they need to develop the necessary management structure, rules, processes, and practises for credit risk measurement(BIS, 2009).

One of Basel II's main advances is that it gives lenders an option between I and II, in contrast to Basel I, which only provided a single method for computing regulatory capital for credit risk. Using a standardised approach, Basel I is expanded upon by classifying exposures into several risk categories. Nevertheless, each risk category historically had a defined risk weight, e.g. The risk weights for the three Basel II categories (loans to governments, businesses, and banks) are established by the borrowers' given external credit ratings. Loans backed by residential real estate, among other categories that continue to be subject to fixed risk weights under Basel II, will bear a risk weight of 35%, up from 50% previously, if Loan-To-Value (LTV) climbs to 80%.This reduced weighting is in acknowledgement of historically low loss rates that frequently originated in residential mortgage loan portfolios across a wide range of economic situations in many nations (Reddy, 2006).

2. A Basic approach Internal Rating Based (IRB) - The IRB approach will enable lenders to create own models for calculating their regulatory capital need. Lenders assess the Probability of Default (PD) under the foundation IRB approach, and supervision sets the numbers for Loss Given Default (LGD), Exposure At Default (EAD), and Maturity Exposure (ME).For each exposure or kind of exposure, these numbers are fed into the lender's appropriate risk weight function to produce risk weights (Chatterjee, 2022).

3. Advanced IRB approach - For a more sophisticated IRB method, the lender will estimate PD, LGD, EAD, and M. Lenders with the most extensive risk and risk management modelling expertise will be able to use this technique. Only estimations of PD, LGD, and EADs are needed for retail portfolios; this method is referred to as retail IRB. It is not unexpected that the Basel II rule pushes lenders to switch to the IRB approach and ultimately advance or retail IRB access given that one of Basel II's primary goals is to strengthen the culture of risk management. In order to achieve this, banks should experience some regulatory capital relaxation while switching from a standardised approach to a basic IRB and then to an advanced or retail IRB approach. Three additional methods for calculating credit risk have been authorised.

Under this strategy, banks can utilise credit risk limiters (collateral, guarantees, and credit derivatives) to lower capital depending on the market risk of the collateral instrument that others hold and the external credit assessment of recognised guarantors. Residential mortgages, small and medium-sized enterprises (SME), and retail exposures all had their risk weights reduced. This method significantly differentiates exposures from transactions in order to increase the resulting risk-sensitive capital ratios.(Zhang, 2017).

Internal evaluations of the major risk concerns facing banks serve as the main inputs for calculating capital under the IRB approach. Combining quantitative information from banks and algorithms recommended by the committee, risk weights and the ensuing capital requirements are calculated.

The following four factors are used to calculate IRB risk-weighted assets for exposures to sovereigns, banks, or business entities:

1. Probability of default (PD), It calculates the probability that a borrower will default within a specific time frame.

2. Loss Given Default (LGD), the percentage of an exposure that will be lost in the event of a delay.

3. Exposure at default (EAD), determines how much of a credit commitment will likely be used as a resource in the case of a default.

4. Maturity Exposure (ME), determines the exposure's remaining economic maturity (BCBS, 2003).

Chapter 3: Problem Statement

Existing Credit Risk prediction models are mainly based on a few basic ML and /or Neural networks because of the long time-consuming process for each model preparation and training that they lack in the interpretability of their prediction. It's hard to find which input features are playing important roles in predicting.

Obstacles to effective credit risk management

1. Irregular data handling - Problematic delay results from being unable to access the appropriate data when it is required.

2. There is no framework for modelling group-wide risk. Without it, banks are unable to develop extensive and relevant risk measurements and obtain a full understanding of group risk.

3. Ongoing revisions - The inability of analysts to quickly alter model parameters causes unnecessary effort duplication that is detrimental to the bank's efficiency ratio.

4. Inadequate risk management tools - Without good risk management, banks are unable to recognise portfolio concentrations or conduct portfolio reassessments regularly enough to manage risk.

5. Serious reporting - Analysts and IT are burdened by manual reporting processes based on spreadsheets.

   A review of the practises used in credit risk management and how they affect the performance of the Indian banking sector after Basel II's implementation. It is unclear how modelling presumptions affect estimations of the distribution's extreme ends. It is unclear if high target credit loss quantiles, which are used to measure credit risk, can be estimated with a sufficient level of accuracy to produce estimates of economic capital. Back testing offers a mechanism to regularly assess the model's performance for market risk. There isn't already a framework for regular credit verification risk models that is widely acknowledged. The development of both qualitative and quantitative standards may be necessary to ensure that modelling processes are adequate and the quality of outputs is comparable across banking institutions if the models were to be used for regulatory purposes. Supervisors would have to rely on the models for internal and external validation procedures if they were to be used for regulatory purposes.

Chapter 4: Objectives of the Study

Goals can be broad or detailed. What the researcher normally hopes to accomplish is stated in the research study's overall objective. Specific goals break down the overall goal into more manageable, logically connected components that methodically address various facets of the issue.

Most of the institutions the working group analysed did not distribute all of their capital across all of the different types of products or business lines because the majority of the credit risk models under investigation are still in their infant stages. Banks may review their judgments about the allocation of economic capital every month or every year. The majority do not allocate capital and profit/loss at the bank level, but rather on a somewhat micro basis, such as at the sub-portfolio, business, or product line level.

1. To evaluate the current state of the processes for credit risk management and the framework for bank lending.

2. To look into how much the methods and models for recognising, estimating, and tracking the credit risk of Indian banks' performance in the Indian banking sector have changed.

3. To determine the effect of Basel II's credit risk management framework on the banks' financial performance.

Banks occasionally rely on confidential data that consultants have collected. Additionally, the fact that validation analyses are typically incomplete raises questions about the effectiveness and thoroughness of the supervisory process.

The objective of this project is to investigate Super learners for Credit Risk modeling, with the help of Automation Machine learning. Different super learners can be defined as the ensemble models of different base models and investigated against the credit risk dataset. For explaining the prediction of the super learner different result interpretation techniques i.e. SHAP, PDP and ICE have been used.

Chapter 5: Project Methodology

A methodology is a set of methods, practices, processes, techniques, procedures, and rules. In project management, methodologies are specific and rigorous, and usually contain a series of steps and activities for each phase of the project life cycle. Projects can be brought to a successful conclusion in different ways. But the best and most popular project management methodologies, methods, and frameworks are constantly changing. New concepts are constantly emerging. Behind all successful projects lies a whole range of methods, tools, and techniques. A project management professional will probably use more than one of them in their lifetime.

In the traditional projects we follow following methodology:

1. Desk research:
Desk research is a research method that involves the use of already existing data. Existing data are summarized and compared to increase the overall effectiveness of the research.

2. Questionnaire construction:
Questionnaire construction is related to the design of a questionnaire to collect statistically useful information about a given topic. If questionnaires are properly constructed and responsibly administered, they can provide valuable data about any given subject.

3. Pilot study and testing of the questionnaire
Pilot testing allows the researcher to assess their entire questionnaire under survey conditions. The primary benefit of pilot testing is to identify problems before implementing a full survey. Pilot testing seeks to verify the validity of each question.

4. Finalization of the questionnaire
Finalization is a very important part of developing any question because it is a phase where researchers select the best out of testing questionnaire, which shall further allow researchers to get better information from their survey audience.

5. Data collection

Data collection is the process of collecting, measuring, and analyzing accurate knowledge for research using standard proven techniques. The researcher can evaluate his hypothesis based on the collected data.

6. Data Preparation

Data preparation involves searching for correct data to make sure that analytics applications provide meaningful information and actionable insights for business decision-making. Data is often enriched and optimized to make it more informative and useful—for example, by blending internal and external datasets, creating new data fields, eliminating outliers, and dealing with unbalanced datasets that could bias analysis results.

7. Data analysis and interpretation

Data interpretation is the process of assigning meaning to data. It includes an explanation of discovered patterns and trends in the data. Data analysis is the process of discovering patterns and trends in data.

8. Compilation of results and Report writing

After the survey, a compilation of results and forming a report out of the results derived through the analysis of the survey is a very important part of the project methodology.

9. Submission of progress reports, various statements, final reports, etc.

In case of Machine learning methodology, Data related steps same as in traditional approach but in place of manually understanding the relationship between data trends and their outcomes different Machine learning algorithms has been used which gives good in predictions based on different inputs. But the Machine learning algorithms are tedious and time consuming and it difficult for non- technical person to develop a statistical model which can give good accuracy.

In this project in place of traditional methodology, AutoML has been used. With the use of AutoML most of the Data related steps are same as traditional approach but for the selecting different machine learning algorithms, training and hyperparameter tuning could be automated.

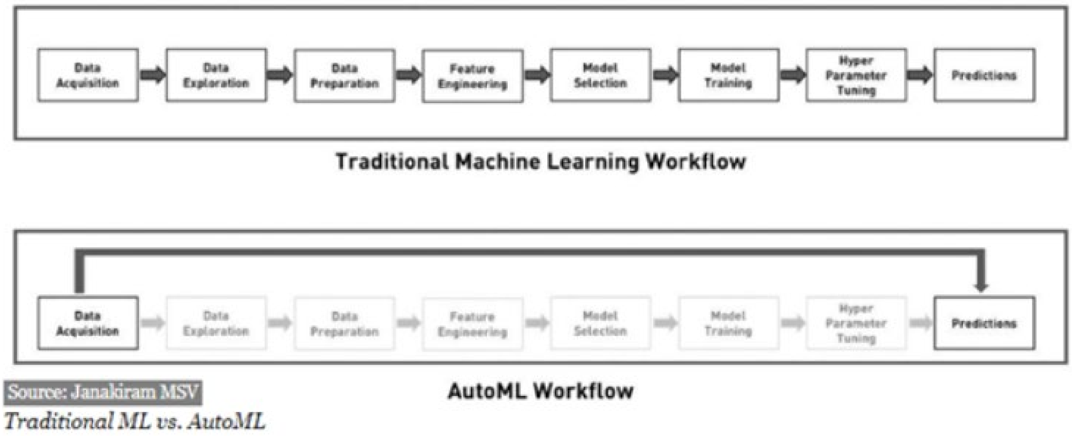


Figure 5.1 AutoML vs Traditional Machine learning workflow

In Figure 5.1 it has been shown that in the case of AutoML basic time-consuming steps of data modeling like Data explorations, Data Preparation, Feature Engineering, Model selection, Model training, and Hyperparameter tuning can be bypassed. So in this way, a lot of time could be saved that researcher generally waste during these processes and concentrated on Data Collection and deployments of the best model (Fernandez-Reyes, 2018).

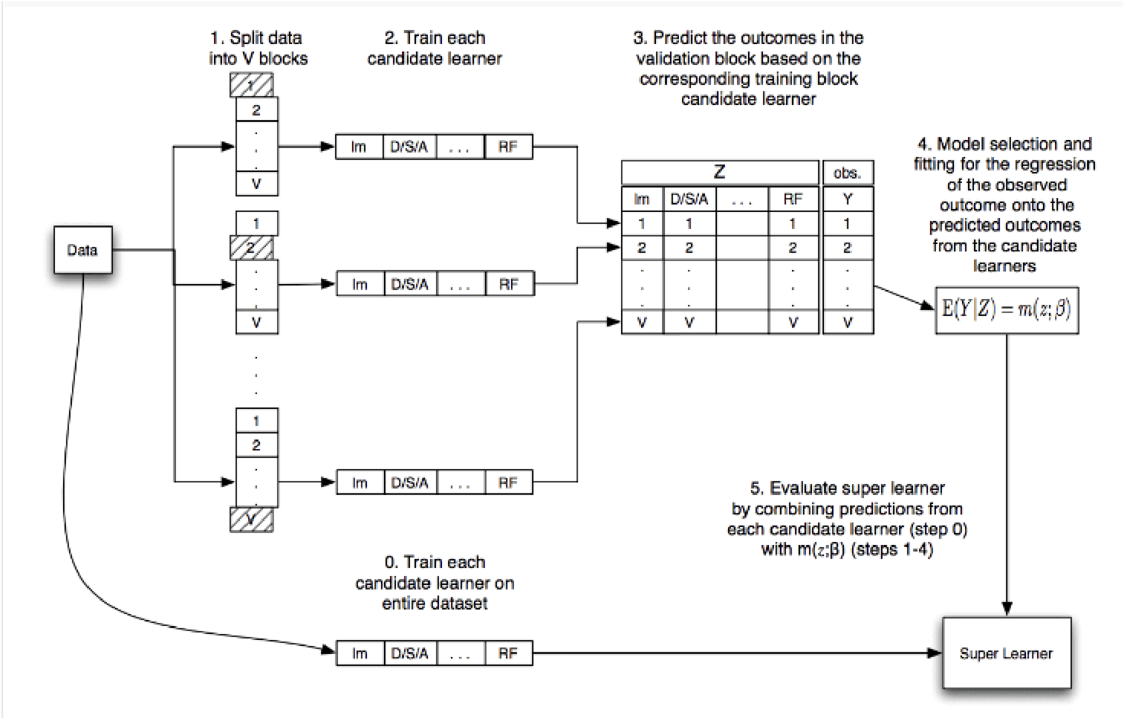


Figure 5.2: Process flow in AutoML

In Figure 5.2 Super learners can be described as the stack of different ensemble models, In this process, certain steps have been followed:

1. The cross-validation has been selected for the training dataset.
2. The maximum number of base models has also been specified.
3. For every base model these set has been followed:
    3.1.Base models have been evaluated by using cross-validation
    3.2.Store their predictions
    3.3.Then trained the base models on the training dataset and stored predictions.
4. Metalearners on the stored predictions
5. The prediction has been done on the Test Data set(Polley, 2011).

Figure 5.3: AutoML Workflow

In Figure 5.3 steps followed by AutoML have been shown, they are as follows:

1. Load and train and Test data.
2. Specify the response feature.
3. Running the AutoML by specifying the stopping criteria e.g. Maximum number of models, the maximum time to train is stopping metric or stopping rounds.
4. Analyze different models on the leader board and explore them based on our requirements usually ensemble stack model best of the family or All models give better accuracy.
5. The model could be saved and deployed in production for further uses (Pandey, 2019).

Chapter 6: Business Understanding

Assessing credit risk requires access to precise business data. Customers must provide their business details to suppliers as part of a credit application. Suppliers frequently check trade credit bureaus in addition to self-reported data to familiarise themselves with a prospective client. Here are several signs that can be used to gauge the degree of risk associated with extending company credit:

1. Business examples: Vendors, banks, landlords, and other partners can inform company credit bureaus about their payment experiences. Trade references are important sources of information for suppliers to review before issuing trade credit since they might reveal late payments or defaults on prior trade obligations. Although self-reported business references are typically required in the loan application, business references are a component of a credit report from a business credit agency. Although businesses are aware that their clients will only provide favourable references for a loan application, it is crucial to have these favourable references to balance out any unfavourable references that may have been submitted to the appropriate authorities (Commercial Reports – Experian, 2022).

2. Financial and Banking Information: To verify the relationship between the bank and the applicant, all complete loan applications will need banking information. The majority of credit applications also demand financial statements (although some do not, for smaller lines of credit).Credit managers are aware that small businesses are more likely to submit financial statements generated by management than those audited by CPAs, according to the Credit Research Foundation (Wagner, 2021).

3. Internet Research: A loan officer will probably look up a new loan application on the internet in general. The applicant will benefit if the search pulls up a local publication that reports that they have received an award or that they have grown their firm. Other headlines, such frequently changing management or recently closed agreements, could be a hint of problems with the applicants' financial stability and ability to pay.

4. Business credit agencies take into account a variety of elements when determining a company's credit score and rating. Statistical models evaluate a company's performance in

comparison to that of other companies operating in the same sector or locale. Predictive analytics can show whether a company is underperforming or offers a greater danger to its finances. A corporation can use these scores and ratings to reduce the need for independent research and more quickly comprehend the risks involved in working with another company (Johnson, 2022).

Of course, a company's business references, publicly available data, or corporate credit scores and ratings can also provide reassuring information. In either case, credit managers should decide on credit limits depending on how confident they are in the client's financial stability. Of course, a company's business references, financial records, and credit scores and ratings can also provide reassuring information. In any event, credit limitations should be set by business owners based on their belief in the client's capacity to pay.

Since insufficient credit risk policies continue to be the primary cause of major issues within the EU banking system, effective credit risk management has received more attention in recent years. By keeping credit exposure within reasonable level limitations, an effective credit risk management programme must have as its primary goal maximising the bank's risk-adjusted rate of return. Banks must also control both the risk associated with individual loans and transactions as well as the credit risk across the entire portfolio. The goal of this overview is to determine how far banks have come in creating and implementing effective credit risk management strategies. The recent adoption of Basel's new minimal regulatory capital requirements. A considerable dispute between academics, decision-makers, and business professionals was sparked by the Committee on Banking Supervision (also known as Basel II).The significance of these regulations for global banking systems and the fact that the new regulations mark a significant divergence from the current framework both contribute to this attention (Basel I).In the research that focuses on the theoretical merits of the framework, on particular sections of the Agreement, on the potential influence on banking systems, and on in-practice implementation issues, a number of diverse features have lately emerged. This research study is driven by the need to evaluate the new loan capital requirements, debate their adoption status, and examine some practical implementation challenges in the Indian banking sector. It focuses solely on credit risk management methods under Basel Convention II.

The broad objectives of this research study are three in particular:

a. Summarize some theoretical and empirical developments in the area of credit risk management strategies.

b. Briefly describe Basel II's handling of credit risk.

c. Describe the implementation challenges, issues, future prospects, and some policy implications for the performance of Indian banks.

Loss resulting from a loan default or other type of borrower credit is referred to as credit risk. Lenders to consumers, Lenders to enterprises, Businesses, and even Individuals face credit risks. However, credit risks are what lenders generally deal with, particularly in financial sector organisations like banks. Therefore, credit risk management is both a necessity and a solution in the banking industry. (Rehman, 2019).

The largest and most evident sources of credit risk, according to the Basel Committee on Banking Supervision, are loans, while other sources can be found in a variety of activities that the bank itself engages in. Every bank in the globe must be aware of the necessity to assess, monitor, and control credit risk while figuring out how credit risks may be decreased. As a result, the bank must protect itself appropriately against these risks and be fairly compensated for taking such risks. This is outlined in Basel II, which establishes the minimum amount of capital that banks must set aside to safeguard themselves against various forms of operational and financial risks. Advanced Credit Risk Management Techniques in India Under Basel II

The framework can be expected to provide the impetus for the adoption of more of the coming day's sophisticated credit risk management techniques in banks. However, these guidelines and recommendations only provide broad parameters and leave much to be desired room for individual banks to exercise their discretion and judgment. Any bank can develop its systems compatible with its risk architecture and expertise.

The research study makes the assumption that Basel II just aims to codify the current best practises in monitoring banking risks, even if the revised credit capital adequacy rules represent a dramatic change compared to the Basel I framework. However, fundamental flaws in the financial infrastructure that must be fixed as a matter of urgency prevent its effective adoption in many developing nations. The research project places special attention on providing a succinct overview of the Basel I and II frameworks and outlining the fundamental components of credit risk management. It provides an overview of the new Basel II loan capital regulations, addresses implementation challenges, evaluates the performance of a few Indian banks, and makes an effort to derive pertinent policy implications.

Chapter 7: Data Understanding

Businesses require solutions to extract, align, and distil what's necessary as well as swiftly arrive at analytical interpretations while managing data to derive value from data. Scoring models allow every financial institution for which they are built to optimise its operations.

CRIF (*CRIF Highmark is one of the four Credit Information Companies (CIC) in India.*)gives business analysts, from rookie to senior modellers, a full range of modelling tools and knowledge to develop, build, test, implement, and manage predictive models. The following list of score sheet types is available:

Application Scorecard: These are tools that enable businesses to forecast the possibility that a candidate would act in a particular way, assisting them in making wise automated decisions. A credit scorecard app evaluates the risk of default, i.e., it forecasts whether the consumer will pay or not. A numerical score is often supplied for each applicant in a credit risk application scorecard, with higher scores correlating to lower levels of anticipated risk. This enables lenders to decide whether to approve, evaluate, or reject applicants in an accurate and consistent manner. Application scorecards can also be used to forecast a wide range of additional variables, including applicant accessibility, potential future profitability, risk of loss, etc.

Behavioral Scorecard: Does the user know which clients are the most lucrative? Are the clients behind on their or other lenders' payments? The proper clients for our company may be found, kept, and even grown with the aid of behavioural insights. These quantify consumer behaviour to enhance customer and credit portfolio management. Lenders may increase risk exposure control, build a more effective pricing programme, target current and new customers for cross-selling initiatives, and make more customer-centric decisions using CRIF's behavioural scorecards.

Scorecard Collection: Decisions about debt management are simplified with these Scorecards. By considering prior behaviour, it can identify customers who pose a risk. Depending on their riskiness, appropriate treatments can then be started for clients as soon as feasible to protect corporate assets using the most cost-effective method available. By reducing credit losses, the collection card significantly contributes to the business's profitability. As a result, the company can reduce the amount of credit used for reserves. The allocation of capital that would otherwise be used to finance corporate expansion is directly impacted by provisions. By identifying and focusing on customers that have a higher tendency to pay using creative and specialised collection techniques, CRIF Collection Insights may assist organisations in prioritising collections, minimising defaults, maximising returns, and decreasing overhead expenses.

In order to facilitate securitization by group audit services and IFRS 9 assessments, fraud scoring models optimise fraud risk control and concentrate verification on a smaller number of cases;

tracking, monitoring, fine-tuning, recalibrating, and/or re-engineering of scorecards are all part of model management, which also includes trend, stability, and migration analysis; Analyzing the value of different data and evaluating one's own appraisal of the score are examples of feasibility studies.

	AGE	JOB	MARITAL	EDUCATION	DEFAULT	HOUSING	LOAN	LOAN_STATUS	Income
0	58	management	married	tertiary	no	yes	no	no	2194
1	44	technician	single	secondary	no	yes	no	no	6423
2	33	entrepreneur	married	secondary	no	yes	yes	no	728
3	47	blue-collar	married	unknown	no	yes	no	no	2036
4	33	unknown	single	unknown	no	no	no	no	669

Figure 7.1: Sample of Credit Risk Dataset

In Figure 7.1,a sample of the dataset has been shown, where 8 independent features and 1 dependent feature (Loan Status) have been used.

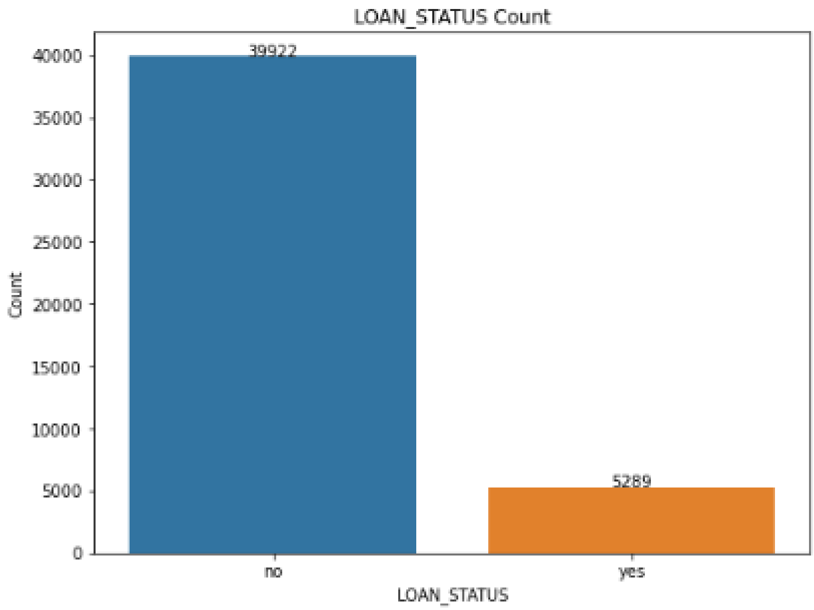


Figure 7.2: Loan Status value counts (No/Yes)

In Figure 7.2, the dependent feature count has been shown, there were 5289 whose loan has been passed based on their past financial data, and 39922 customers' loan was rejected due to some or other reasons.

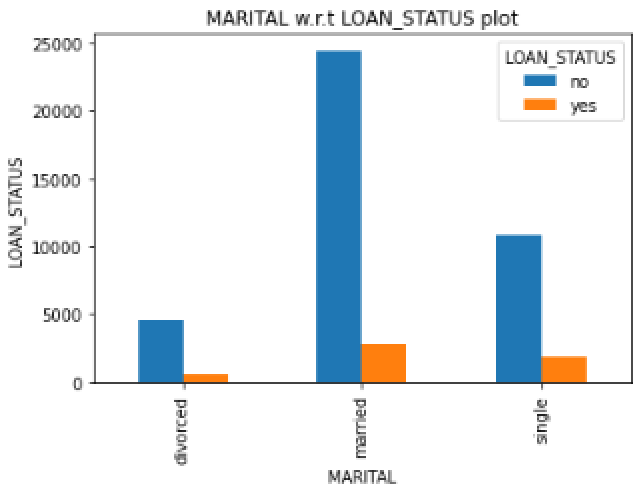


Figure 7.3: Marital Status vs Loan Status

In Figure 7.3, it was a bi-variant analysis between customer's Marital status vs dependent feature loan status, in this, it was noticed if the customer is married or single their loan approval chance was high as compared to the divorced customers.

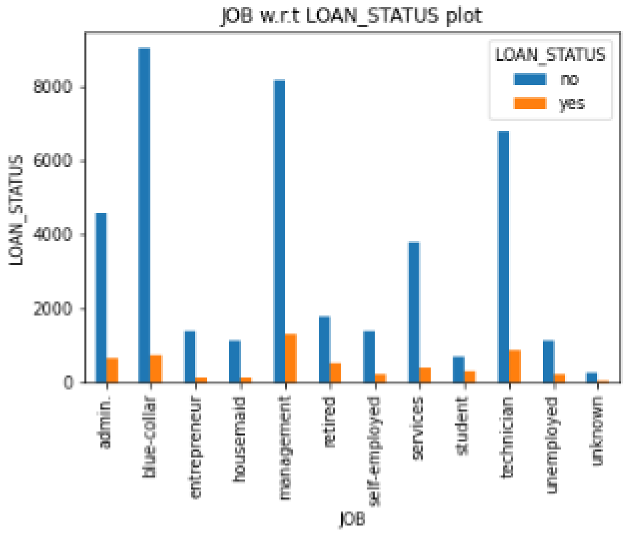


Figure 7.4: Job Status vs Loan Status

In Figure 7.4, it has been noticed that customers who were in the job of Admin, blue-collar, Management, or technician had a high probability of loan approval as compared to other customers.

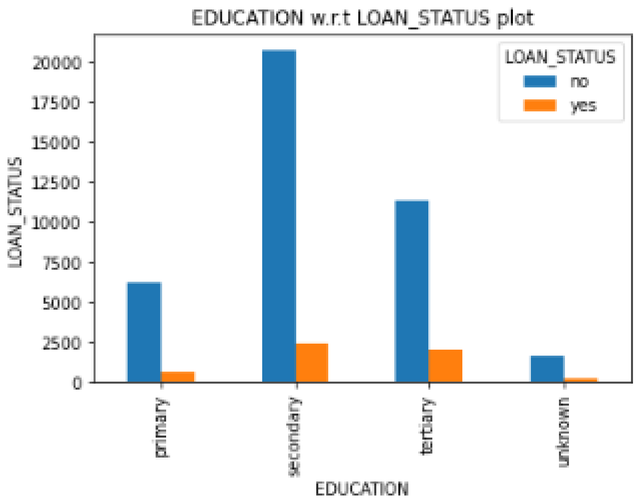


Figure 7.5: Education Status vs Loan Status

In Figure 7.5, it has been noticed that if the customer's education was secondary or tertiary their chance getting of loan approval was high as compared to primary educated or unknown.

Figure 7.6 Age Status vs Loan Status

In Figure 7.6, it has been noticed that in the data set most of the customers were from age 20 to 60 and the highest from age of around 30. The chance of loan approval was also high when the customer's age was between 30 to 40 years.

Chapter 8: Data Preparation

Data collection, combination, structure, and organisation are all steps in the process of preparing data for use in Business Intelligence (BI), analytics, and data visualisation applications. Data pre-processing, profiling, cleaning, validation, and transformation are some of the components of data preparation. It frequently includes entails acquiring data from multiple internal and external sources.

When integrating datasets for loading into a data warehouse, NoSQL database, or data lake storage, as well as when developing new analytics applications using those datasets, information technology (IT), business intelligence (BI), and data management teams execute data preparation. Additionally, self-service data preparation technologies are being used more frequently by business users, data scientists, data engineers, and other data analysts to gather and prepare their own data.

Process steps for preparing data

There are several steps involved in data preparation:-

Although the data preparation procedures described by various data experts and software providers vary somewhat, they often comprise the following activities:

Data Gathering:

Operating systems, data warehouses, data lakes, and other data sources are used to gather pertinent data. Data scientists, BI team members, other data experts, and the end users who collected the data should confirm that it is suitable for the purposes of the anticipated analytics applications during this step.

Data exploration and profiling:

The collected data must next be examined to determine what it includes and what needs to be done to make it suitable for its intended use. In order to resolve inconsistencies, anomalies, missing values, and other problems, data profiling finds patterns, relationships, and other properties in the data, as well as patterns, relationships, and other attributes in the data.

Cleaning data:

Additionally, detected data issues and errors are fixed to produce comprehensive and reliable datasets.For instance, incorrect data is removed or repaired, missing values are added, and conflicting records are harmonised as part of data set cleaning.

Data organisation:
The data now has to be organised and modelled to satisfy the analytical requirements. For instance, in order for BI and analytical tools to access data saved in comma-separated values (CSV) files or other file formats, tables must be created.

Transformation and enrichment of data:
Data typically needs to be formatted in addition to being changed into a standardised format. For instance, data transformation can entail developing fresh fields or columns that combine values from older ones. With the help of techniques like augmenting and appending data, data enrichment further enhances and optimises data sets as necessary.

Data validation and publication:
In this last step, automated processes are used to the data to ensure its accuracy, consistency, and completeness. The created information is subsequently kept in a data warehouse, data lake, or other repository and either used by the person who prepared it directly, or made accessible to other users.

Data management, which develops and controls data sets suitable to be used for BI and analytics, might incorporate or feed data preparation operations. In order to make data accessible to consumers, data management duties include indexing, classifying, and preserving data sets and the metadata that goes with them. A data curator collaborates with data scientists, business analysts, other users, and the IT and data management teams in some organisations. In some cases, business users themselves may handle data, as well as data managers, engineers, database administrators, or data scientists.

For the study, descriptive research will be used. The goal of descriptive research design is to identify and describe the traits of a specific person or group. It would consist with (studies in development, case studies, documentary analysis, and correlation studies).

Actions required:

1. Conducting a pilot study, creating cover letters, submitting the questionnaire, following up, reviewing the data, and creating a report are just a few of the first steps. temperament 18

2. Determining objectives: preparing the list of objectives and anticipating the kind of response to be received to value the entire study

3. Delimiting the sample: distribute questionnaires to the select representative sample i.e., 30% of the population.

4. Developing the closed- and open-ended questionnaire 6.3 Data Collection SourcesBoth primary and secondary data sources will be used in the investigation. Researchers will be able to acquire information directly from primary sources. The tools proposed to be used for Primary Data collection include the following Sources of data and tools: For the study, information will be gathered from both primary and secondary sources.

   a) Primary data:- The main information needed for the study will be gathered through
      i). A survey that incorporates an in-depth interview
      ii). Observation

   b) Secondary data: - Will be gathered from a variety of sources, including books, magazines, newspapers, journals, the RBI, the Bank for International Settlements (BIS), bank annual reports, and other official websites, such as the Indian Banking Association (IBA), etc. 6.4 Sampling Plan: The sampling plan for the research is as follows

5. Population Definition: Public Sector Banks (27 in number including 19 Nationalized Banks and 7 State Bank of India and its subsidiaries) in India.

6. Sampling Frame: Reserve Bank of India Documents and Database, Banks Data, Bank for International Settlement (BIS)data Base

Sampling Unit:

Nationalized Banks – 19 in number of SBI group along with its 07 in number

Sample size: 30% of the Population

Sample design: Non-probability Judgement Sampling 19

7. The Respondents are in the positions of General Manager, Deputy General Manager, Assistant General Manager, and Executive Director, and their major duty is to design the framework for credit risk management in their banks.

8. The responding Indian banks would be classified based onthe size as measured by the value of total advances Ownership – public-private Geographical spread as measured by no. of branches made by these branches till the year 2011-12 (Respondents to be chosen for the study subject to their exposure to strategic and operational issues related to Credit risk management system and practices in Commercial Banks)(Noman, 2014)

But in the case of AutoML, most of the data preparation steps were not involved. Data has been collected from different sources, cleaned, and did the exploration. Data has been analyzed to understand the current patterns.

.

Chapter 9: Data Modeling

Data modelling in software engineering is the practise of utilising formal approaches to simplify a diagram or data model of a software system. It involves communicating information and data using text and symbols. A data model offers a design framework for creating a new database or redesigning old applications. The process of developing data models in which data associations and limitations are specified and ultimately written for reuse is known as data modelling. Conceptually illustrating data's relationships through diagrams, symbols, or text. Thus, data modelling aids in improving naming, rules, semantics, and security consistency. In turn, this enhances data analysis. No matter how it is used, there is a focus on the necessity of data availability and management.

The study aims at using Descriptive statistical tools using SPSS software. This will enable,

1. Presentations of data in graphics where graphs are used to compare data or summarise it.

2. Data is summarised in a tabular description using tables of numbers.

3. Statistical summaries (single numbers) that condense the data Various statistical methods, including the Chi-square test, t-test, ANOVA, correlation, multiple regression analysis, and discriminant analysis, as well as graphical approaches like bar diagrams, pie charts, and so on, will be used to examine the data gathered from the various sources. All Statistical analyses of the data will be done using SPSS version 18.0 and Microsoft Excel(Guangxu Li, 2018).

It facilitates simple data storage in the database and has a beneficial impact on data analysis. For data management, data governance, and data intelligence, it is crucial. Additionally, it allows for faster performance, less mistake, higher quality, and a clearer scope of data utilisation. It also improves the documentation of data sources. Data modelling guarantees that a business complies with relevant governmental legislation and industry rules from the perspective of compliance. Additionally, it empowers workers to make data-driven choices and plans, builds on business intelligence, and makes it possible to find new opportunities by enhancing data capabilities.

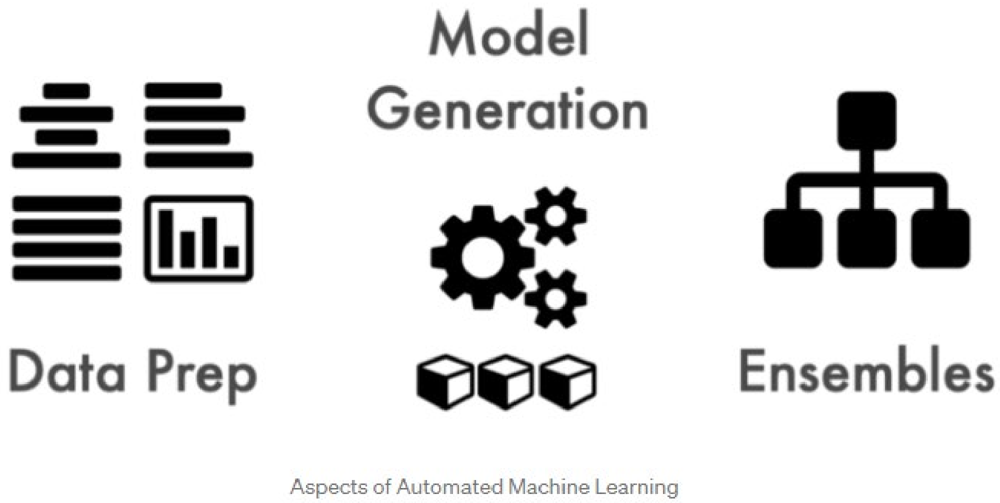


Figure 9.1: Aspects of Automated Machine Learning

In Figure 9.1, basic aspects of Automated machine learning have been explained, where data has to be collected from the source and prepared for the machine learning models. In Automated machine learning, different base and ensembled models will be generated(Pandey, A Deep dive into H2O's AutoML, 2019) .

model_id	auc	logloss	aucpr	mean_per_class_error	rmse	mse
StackedEnsemble_BestOfFamily_3_AutoML_8_20220716_204454	0.710815	0.326445	0.286215	0.351219	0.306801	0.0941268
StackedEnsemble_AllModels_2_AutoML_8_20220716_204454	0.710589	0.326455	0.286061	0.344115	0.306765	0.0941046
StackedEnsemble_AllModels_1_AutoML_8_20220716_204454	0.710347	0.326806	0.284202	0.351165	0.306946	0.0942157
StackedEnsemble_BestOfFamily_2_AutoML_8_20220716_204454	0.710149	0.326896	0.284905	0.346986	0.30696	0.0942247
DRF_1_AutoML_8_20220716_204454	0.704951	0.33437	0.286943	0.34251	0.309266	0.0956456
GBM_2_AutoML_8_20220716_204454	0.701234	0.331476	0.263136	0.352149	0.309638	0.0958756
XRT_1_AutoML_8_20220716_204454	0.700893	0.337806	0.271597	0.354919	0.312968	0.0979493
GBM_4_AutoML_8_20220716_204454	0.700447	0.334691	0.270076	0.351465	0.309524	0.095805
GBM_3_AutoML_8_20220716_204454	0.700335	0.332374	0.260778	0.353529	0.309812	0.0959832
StackedEnsemble_BestOfFamily_1_AutoML_8_20220716_204454	0.699799	0.33095	0.261124	0.358147	0.30917	0.0955859
GBM_1_AutoML_8_20220716_204454	0.69844	0.332399	0.261681	0.357177	0.309774	0.0959597
GBM_5_AutoML_8_20220716_204454	0.696926	0.332226	0.259248	0.364794	0.309984	0.09609
GLM_1_AutoML_8_20220716_204454	0.668986	0.340973	0.218211	0.369995	0.313912	0.0985409
DeepLearning_1_AutoML_8_20220716_204454	0.612563	0.397489	0.195069	0.415548	0.32298	0.104316

Figure 9.2: Leader Board of AutoML

In Figure 9.2,the Leader Board of the AutoML has been shown, where 14 machine learning models with different metrics have been shown, out of that Best of Family ensemble machine learning models outperformed other ensembles as well as base models.

Chapter 10: Data Evaluation

Plans for evaluation should outline the methods and sources used to gather data. Both quantitative and qualitative data has to be gathered within a structure that is in keeping with programme goals, stakeholder expectations, and project schedules. The Basel Committee on Banking Supervision produced the capital adequacy rules, which were approved by the Narasimhan Committee (BCBS).In 1988, BCBS introduced Basel I requirements, which are regarded as the first step toward risk-weighted capital adequacy standards. Basel I regulations were modified by BCBS in 1996, and Basel II, a comprehensive modification of Basel I's framework, was launched in 1999.In accordance with the recommendations of the Narasimham Committee, India accepted the Basel I standards for commercial banks in 1992, the Basel I market risk amendment in 1996, and has committed to implementing the Basel II standards beginning in March 2008.On April 27, 2007, the final RBI instructions for implementing Basel II were made public (Narasimham).

These recommendations state that banks in India should initially use SA for credit risk, and the RBI has supplied the 12 details of these techniques in its recommendations. Some banks may be permitted to transition to more advanced strategies like IRB once sufficient skills are gained by both banks and RBI. Indian banks will be expected to continuously maintain a minimum CRAR of 9% under the updated Basel II rule. Banks are additionally urged to reach a tier I CRAR of at least 6% by March 2010.The RBI has recommended the banks to operate the updated standards concurrently with the standards that are now in effect in order to facilitate a smooth transition to Basel II.

```
metalearner.varimp()
```

```
[('DRF_1_AutoML_8_20220716_204454',
  0.45089760422706604,
  1.0,
  0.5377771520927915),
 ('XRT_1_AutoML_8_20220716_204454',
  0.21615146100521088,
  0.4793803714609224,
  0.25779981093343934),
 ('GBM_2_AutoML_8_20220716_204454',
  0.10483089834451675,
  0.23249380205561107,
  0.12502985474869172),
 ('DeepLearning_1_AutoML_8_20220716_204454',
  0.03858282417058945,
  0.08556892697784141,
  0.046017013857779586),
 ('GLM_1_AutoML_8_20220716_204454',
  0.027984146028757095,
  0.062063195205323496,
  0.03337616836729786)]
```

Figure 10.1: Importance of different Base models in Best of Family Stacked Model.

In Figure10.1, it was shown how the different base models contribute to the best family stacked ensemble model. Here it has been noticed the DRF was the playing most important role in making the decision.

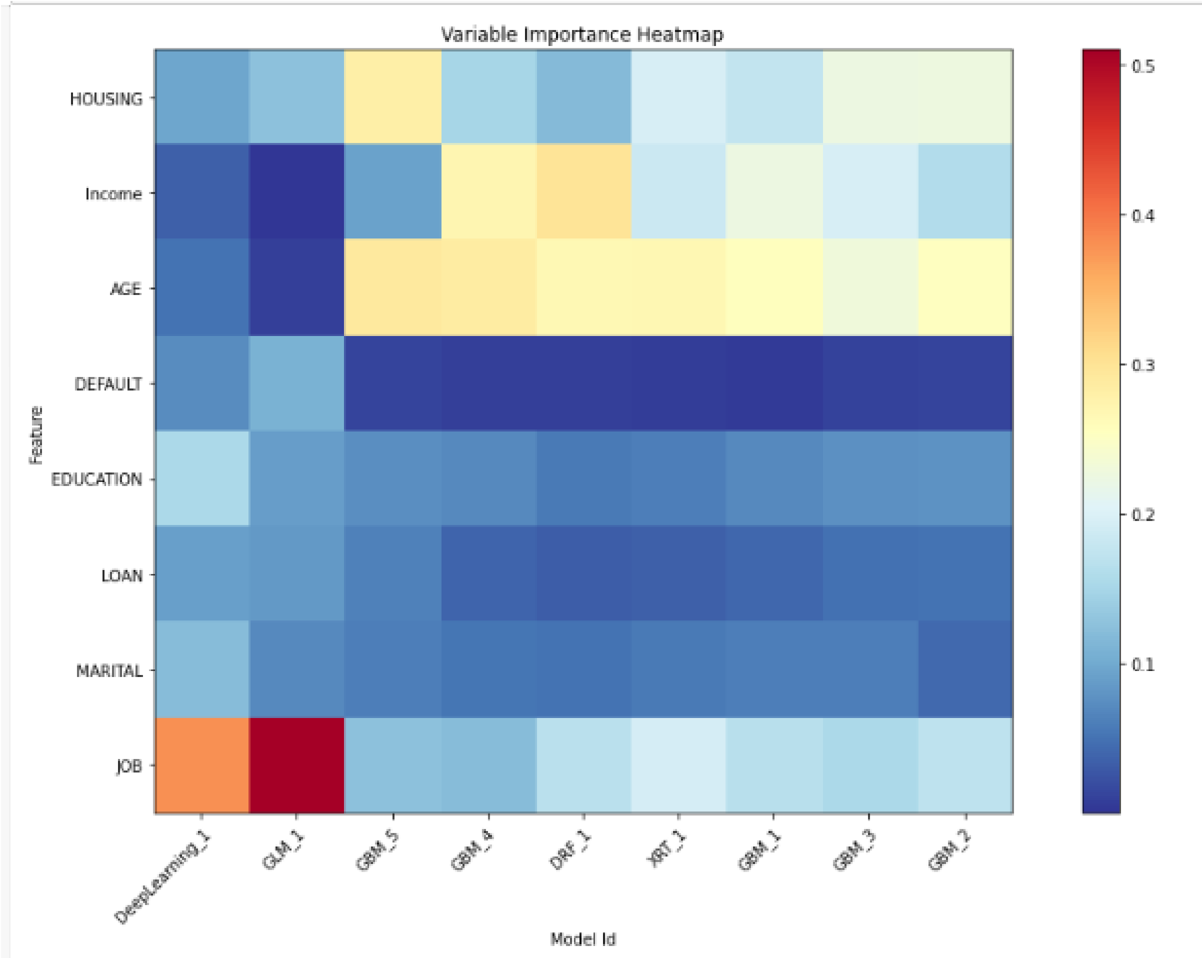


Figure 10.2: Variable importance Heatmap for Different base models in AutoML

In Figure 10.2 with the help of heatmap, it has been shown which base machine models give more importance to which feature while predicting the results. As DeepLearning and GLM_1 gives more importance to the Job feature, while GLM_5 gave more importance to Age and Housing.

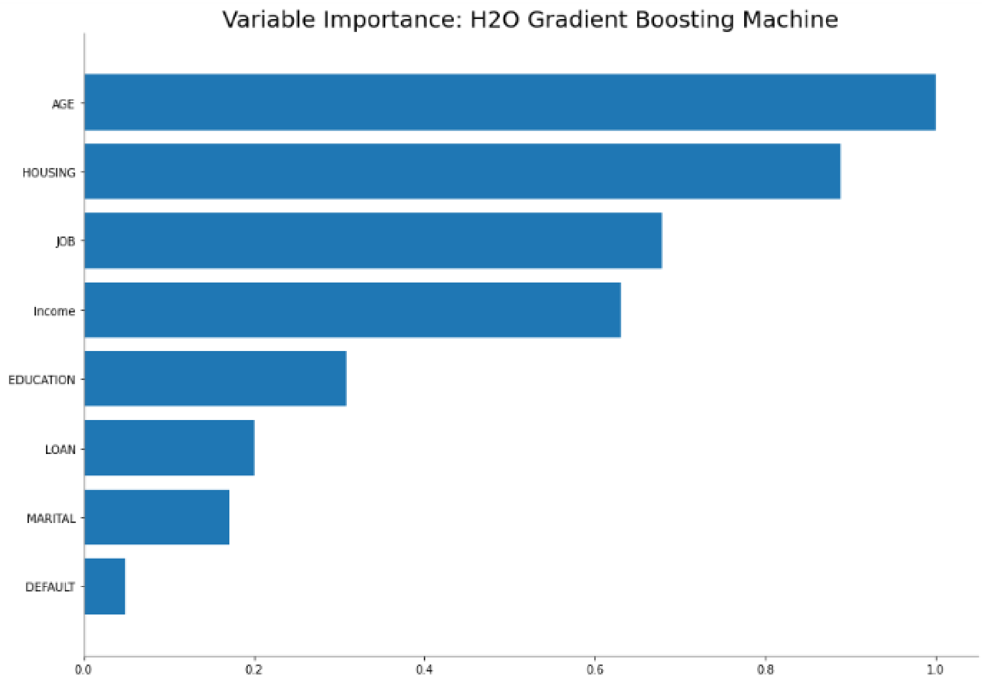


Figure 10.3: Variable Importance for the Gradient boosting model

In Figure 10.3, it is shown how the Gradient boosting base model treats to different independent features for making any decision. As Gradient boosting gave more importance to Age, housing and Job as compare to Marital, Loan and Education.

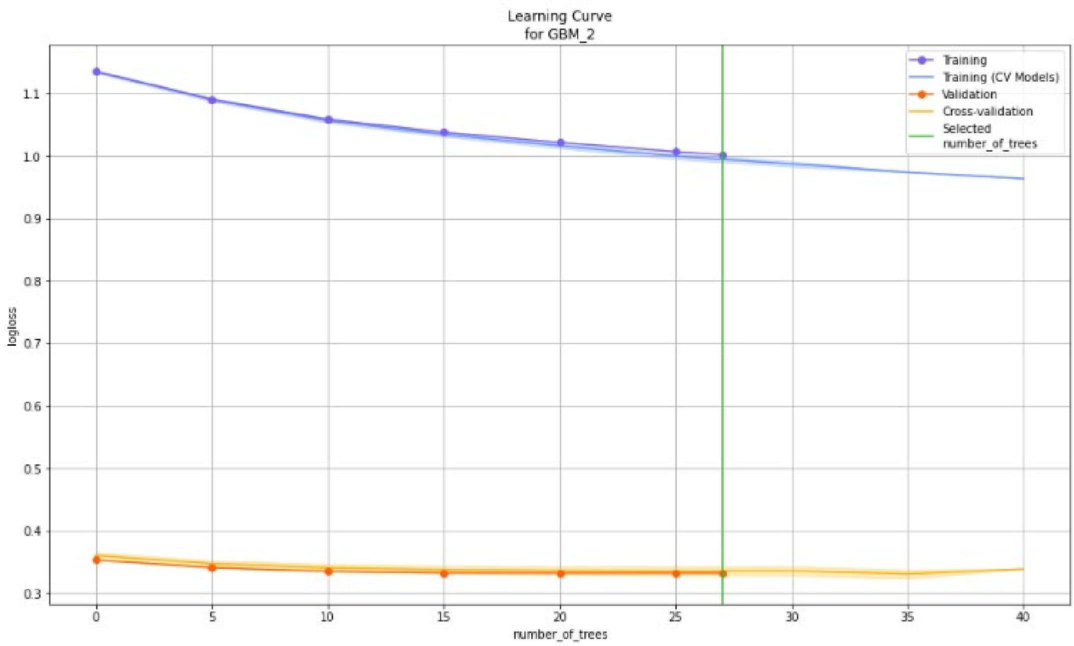


Figure 10.4: The learning curve of the Gradient Boosting model

In Figure 10.4, it is shown how the learning curve of the Gradient boosting model, and how many trees it takes in learning during training, validation, and cross-validation.

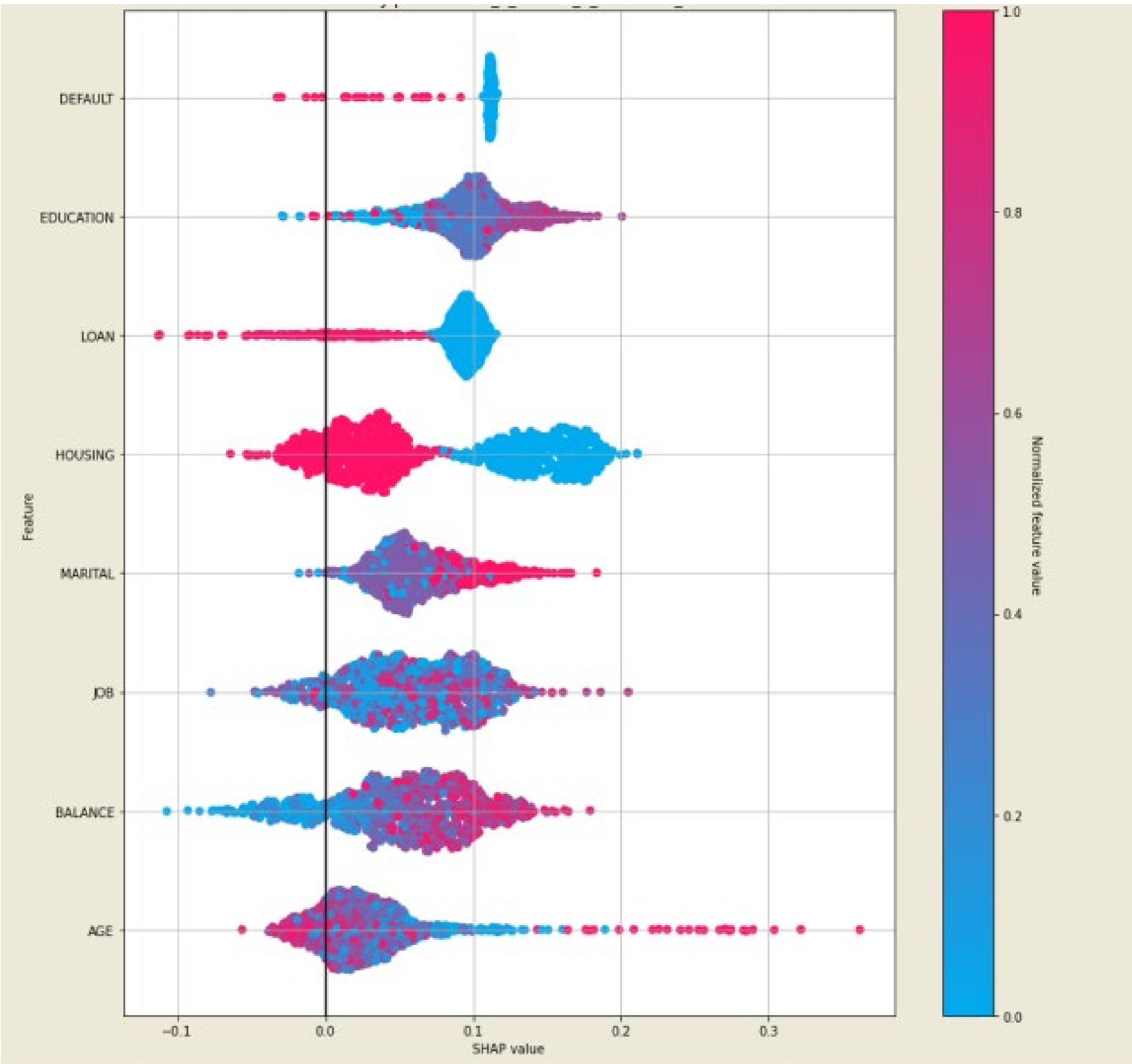


Figure 10.5: SHAP Explanation

In Figure 10.5,A positive SHAP number indicates a positive impact on prediction, which causes the model to predict 1, according to the SHAP explanation (e.g. Loan approved).The model predicts 0 since a negative SHAP score indicates a negative impact (e.g. Loan did not approve).

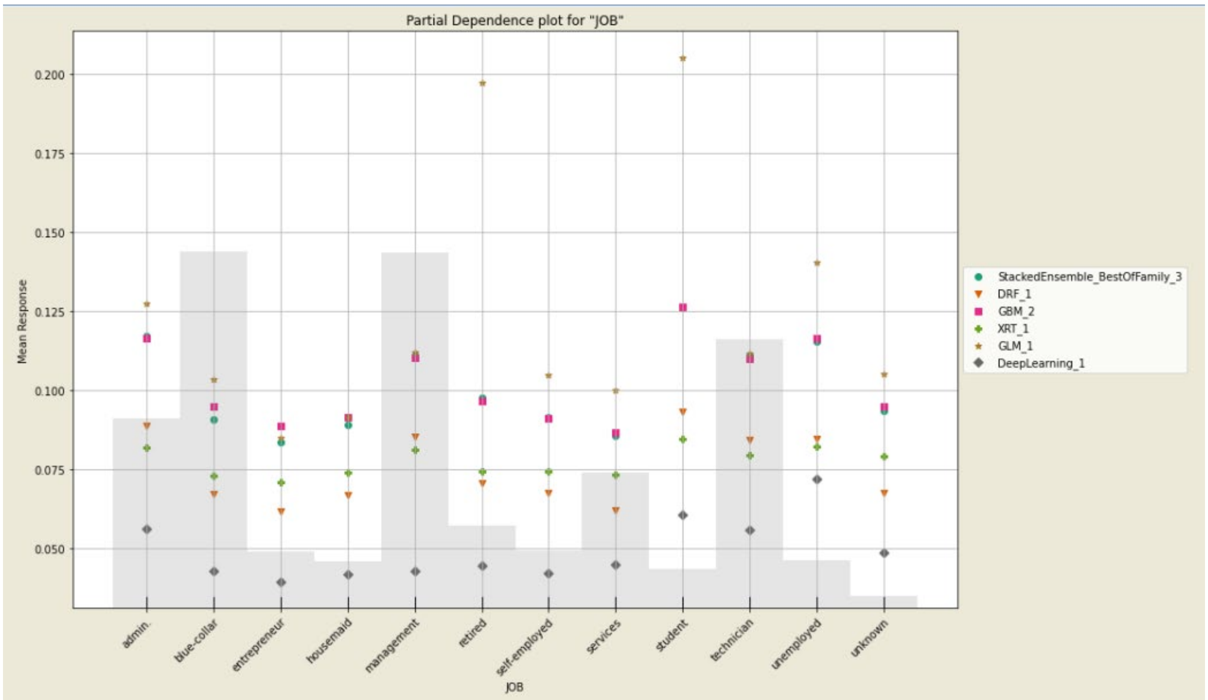


Figure 10.6: Partial Dependence Plot Effect of Job on different models

In Figure 10.6,it has been shown the effect of Job independent features on different ensembles as well as base models with the help of PDP. GLM_1 giving more importance Admin., self-Employed and unemployed person. Whereas GBM_2 gave the importance person who are working as Admin., management, student and unemployed.

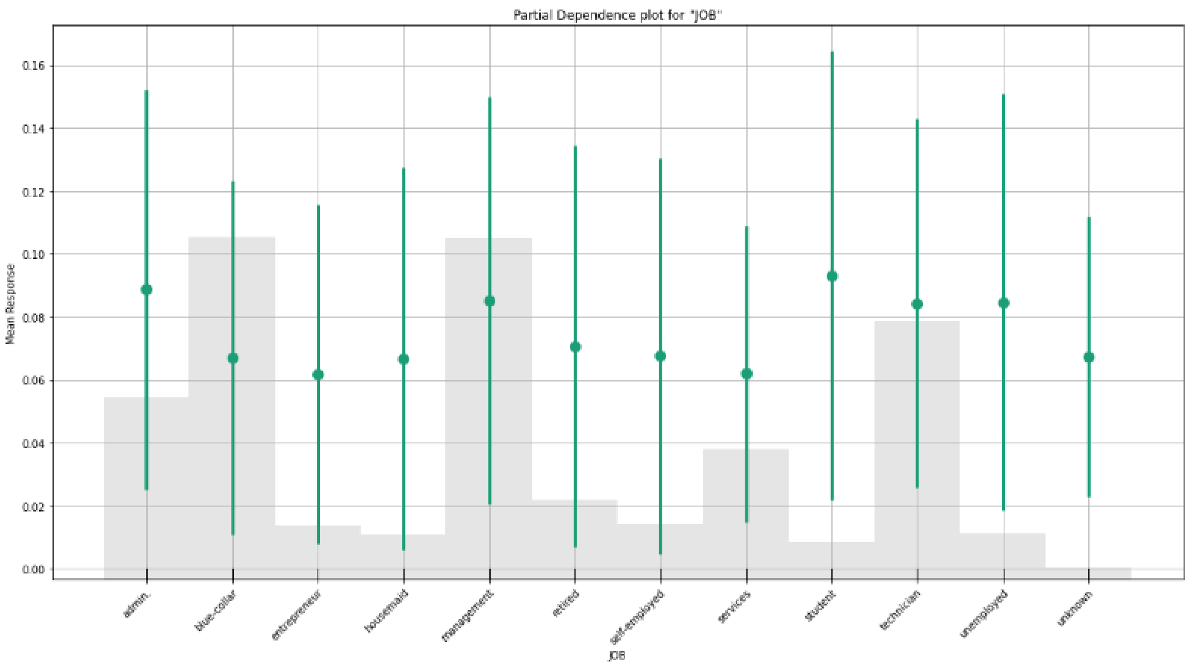


Figure 10.7: PDP plot for Job feature.

In Figure 10.7, it has been shown which jobs impact positively or negatively on loan approval for a customer. In this case person who are working as a admin, management and student have high chance of loan approval whereas who are working as blue-collar, entrepreneur, housemaid and services has less chance of getting loan approval.

Individual row prediction

AGE	JOB	MARITAL	EDUCATION	DEFAULT	HOUSING	LOAN	LOAN_STATUS	Income
42	admin.	married	secondary	no	yes	no	no	1173

predict	no	yes
no	0.934694	0.0653063

Figure 10.8: Single Row prediction by Best Base Model

In Figure 10.8,checking the individual customer result and their probability of approval and not approval with the best base model. In this case as per base model probability of rejecting the loan application is 93.47% and for getting approval is only 6.53%. Which is same as real Loan status.

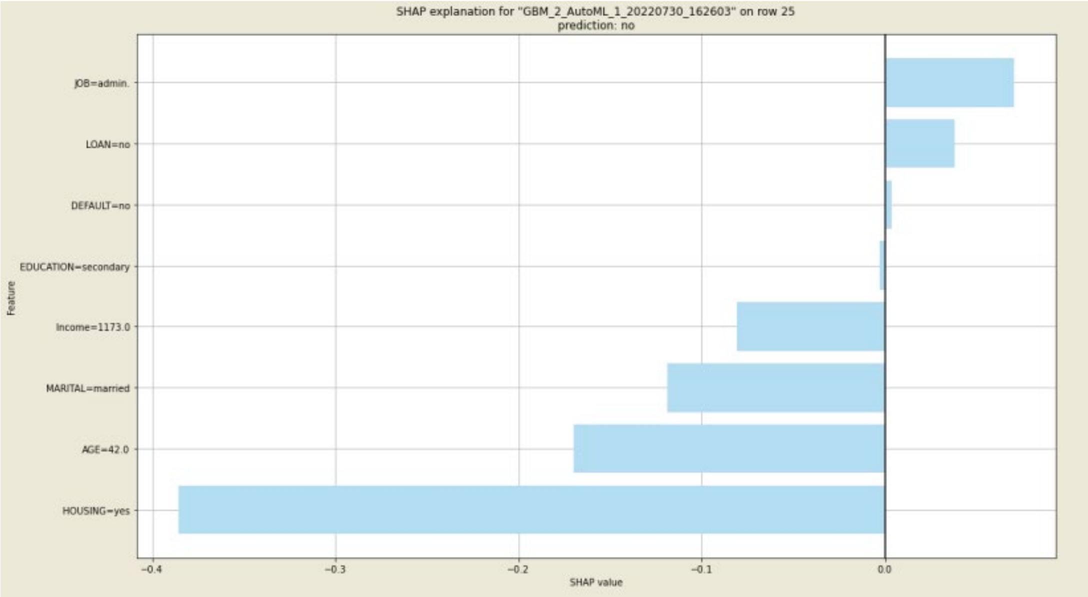


Figure 10.9: SHAP Explanation of the result

Figure 10.9, it has shown the explanation of individual customer results with the help of SHAP. Job and loan positively impact the chance of loan approval whereas Housing, age, marital status, and income effect are more in rejecting their loan request.

model_id	auc	logloss	aucpr	mean_per_class_error	training_time_ms	predict_time_per_row_ms	algo	predict	no	yes
StackedEnsemble_AllModels_2_AutoML_1_20220730_162603	0.68943	0.33347	0.25566	0.359944	13824	0.047491	StackedEnsemble	no	0.94048	0.059517
StackedEnsemble_AllModels_1_AutoML_1_20220730_162603	0.68932	0.33381	0.25357	0.361831	17929	0.030178	StackedEnsemble	no	0.9361	0.063897
StackedEnsemble_BestOfFamily_3_AutoML_1_20220730_162603	0.68886	0.3338	0.25414	0.364459	11105	0.02589	StackedEnsemble	no	0.93728	0.062718
StackedEnsemble_BestOfFamily_2_AutoML_1_20220730_162603	0.68819	0.33421	0.25247	0.36453	13113	0.02527	StackedEnsemble	no	0.93038	0.06962
StackedEnsemble_BestOfFamily_1_AutoML_1_20220730_162603	0.68546	0.33504	0.24878	0.371197	13703	0.018063	StackedEnsemble	no	0.94135	0.058653
GBM_2_AutoML_1_20220730_162603	0.68323	0.33611	0.24774	0.362919	1443	0.012393	GBM	no	0.93469	0.065306
GBM_5_AutoML_1_20220730_162603	0.68253	0.33645	0.24671	0.366798	1154	0.009996	GBM	no	0.92833	0.071669
GBM_1_AutoML_1_20220730_162603	0.68241	0.33653	0.24882	0.374916	2649	0.017497	GBM	no	0.95124	0.048763
GBM_3_AutoML_1_20220730_162603	0.68227	0.33688	0.24782	0.369229	1583	0.014941	GBM	no	0.93585	0.064146
XRT_1_AutoML_1_20220730_162603	0.67983	0.33902	0.24673	0.363123	1920	0.010926	DRF	no	0.94077	0.059228

Figure 10.10: Prediction by a different model of Leader board for the individual row (Including Stacked)

In Figure 10.10, it has been shown how different machine learning models predicting of a particular row. In all of the models giving the high probability of rejection of loan application.

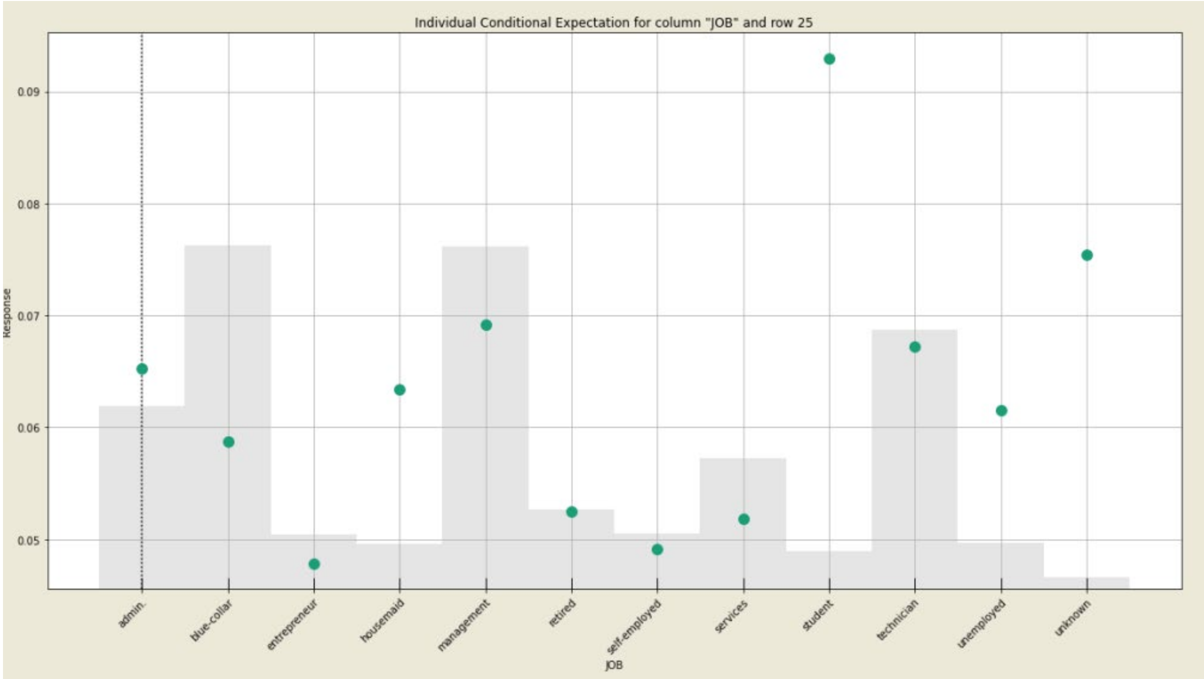


Figure 10.11: Model Individual Conditional Expectation for Job feature.

In Figure 10.11,the effect of different Jobs on loan approval has been shown with help of the ICE graph. In the explanation of different job on loan approval status by ICE is almost same as described by SHAP and PDP.

Chapter 11: Deployment

The concept of deployment is about applying a model to predict using data. The creation of the model generally does not end the project. Although the model's goal is to obtain a deeper understanding of the data, the information must still be organised and presented so that the client can make use of it. The deployment step can range in complexity from developing a repeatable data science process to something as simple as producing a report, depending on the requirements. The customer, not the data analyst, will frequently carry out the deployment stages. To swiftly identify transactions that are extremely likely to be fraudulent, a credit card firm, for instance, could want to use a trained model or group of models (such as neural networks or meta-learners). The customer must know in advance what steps must be completed in order to use the models developed, even if the analyst does not carry out the deployment. The proposed research study's advantages include:

1. The banking Industry as a whole - The capability to measure and predict the risks of any individual application. It allows banks to plan strategies to avoid a negative outcome. Using different credit scoring models, it is possible to find the best ones for the business and determine the level of credit risk.

2. The Commercial Banks plus R&D departments of these banks - It aids in calculating the risk component of each transaction and in making plans for how to handle a bad event.Also aids in creating credit models, which can be a useful tool for assessing the risk of lending.

3. Academicians, Academic Researchers - Will help them analyze which model of credit risk management is more effective further forming different models and educating people in the field.

4. Banking Training Institutions and Organizations - For routine banking procedures, risk is acceptable. Credit risk management functions as a preventive technique to lessen a risk's likelihood of happening or to mitigate a risk that is already likely to happen.

5. Educational institutions - Credit risk management courses can provide students with the skills and knowledge to enter the field. These courses may include credit risk analysis, financial statement analysis, and loan portfolio management.

Chapter 12: Analysis and Results

The research paper's discussion section should be guided by the results section, which should concentrate on summarising the findings without attempting to analyse or evaluate them. The findings of the analysis are reported in the results. The author explains how the data was used in the analysis section. Knowing what the analysis entailed is crucial for writing an analysis, but this does not imply that data is required. The analysis must be completed before writing the results section.

The contribution of the study lies in both the identification of benchmark practices and the credit risk management practices and the same being followed presently. The study would make a horizontal review of Credit risk practices across the banks in India. Such review will help banks management and bank supervisors both by revealing the range of credit risk management practices followed in the Indian Banking industry and shall provide useful information about the strength and weaknesses of alternative practices. The study would enable us to identify the gaps between benchmark practices and the present credit risk management practices of banks in emerging markets like India, that have undergone significant changes due to reforms and that have not been found in the existing literature. The research study would contribute to existing empirical analysis through

1. Examining the extensive array of credit risk management practices

2. An analysis of the level of credit risk management practises and how they affect Indian banks' financial performance.

3. And the financial performance of Indian banks would be through the time series dimension of data

In this research, it has been shown that with the use of Automate Machine learning technique in the combination of different explanations e.g. SHAP, PDP and ICE. Different complex, as well as base models, can be developed in a short time with minimal knowledge of programming and compared with different metrics. Researchers could save a lot of time in developing, training, or tuning the different machine learning models, they could spend that time on data collection and understanding it. On rejecting any loan application end user can explain the reason or features behind that so the customer can also be satisfied with the explanation.

Chapter 13: Conclusions and Recommendations for future work

The Basel-II Credit Risk Management practises research study would highlight the current progressive development and sophistication of the risk management configuration of the Indian banking sector. Better risk management capabilities would give banks a competitive edge in the market and better position them to take advantage of prospects for both organic and inorganic expansion .It is entirely up to the bank's adaptability, risk appetite, and competitiveness to grasp the nettle and upgrade the risk governance in their organisation in order to achieve a sharper risk-reward profile, even though the Basel II framework creates an enabling environment for enhancing the risk management capability in the banks by providing the right incentives. The study examines initiatives to recognise the new paradigm in risk management, corporate governance, and oversight of commercial banks in India.

So it has been shown Automated Machine learning can be a very useful technique in the field of financial and banking sectors. Researchers or machine learning engineers can use Automated Machine learning in a combination of explanation techniques in providing very efficient models to the industry and provide insight also how the complex model is working and explain every prediction done by it.

In the future, more realistic data can be used from the industry to prepare many more machine learning models with all the dependent features, in the different financial sectors they usually consider a lot of data or features to calculate the credit risk which cannot be available in the general to use and research.

Bibliography

Amendment to the capital accord to incorporate market risks, Basel Committee on Banking Supervision. (2005, 11 15). Retrieved from BIS: https://www.bis.org/publ/bcbs119.htm

BCBS. (2003). Basel II: The New Basel Capital Accord, Basel Committee on Banking Supervision. *https://www.bis.org/bcbs/bcbscp3.htm* , 226.

BCBS. (2020). Calculation of RWA for credit risk. *BIS* .

Bessis, J. (2010). *Risk Management in Banking.* West Sussex: John Wiley & Sons Ltd.

BIS. (2009). Basel II: Revised international capital framework,Basel Committee on Banking Supervision. *https://www.bis.org/publ/bcbsca.htm* .

Chatterjee, S. (2022). Modelling Credit Risk | Bank of England. *https://www.bankofengland.co.uk/ccbs/modelling-credit-risk* .

Chen, J. (2021, 02 25). *Risk-Adjusted Return On Capital (RAROC)*. Retrieved from investopedia: https://www.investopedia.com/terms/r/raroc.asp#:~:text=Risk%2Dadjusted%20return%20on%20capital%20(RAROC)%20is%20a%20risk,accompanied%20by%20higher%20expected%20returns

Christian Bluhm, L. O. (2010). *Introduction to Credit Risk Modeling.* New York: Taylor & Francis Group.

CIRCULARS, R. (2008). *RBI CIRCULARS*. Retrieved from Reserve Bank Of India: https://rbi.org.in/Scripts/BS_CircularIndexDisplay.aspx?Id=4682

Commercial Reports – Experian. (2022). *https://www.experian.in/commercial-reports* .

Fernandez-Reyes, F. C. (2018). Using an AutoML H2O Model to Predict Attrition and LIME to Explain the Predicted Class - DZone AI. *https://dzone.com/articles/using-an-automl-h2o-model-to-predict-attrition-and* .

Guangxu Li, G. K. (2018). A Group Decision Making Model for Integrating Heterogeneous Information. *https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7776941* .

J.Teece, D. (2018). Business models and dynamic capabilities. *ScienceDirect* , 40-49.

Johnson, H. D. (2022). What Is A Business Credit Score And How Does It Work? | Bankrate. *https://www.bankrate.com/finance/credit-cards/building-better-business-credit-score/* .

Narasimham, M. (n.d.). Narasimham Committee. *https://en.wikipedia.org/wiki/Narasimham_Committee* .

Noman, A. H. (2014). An Investigation of Credit Risk Management Strategies of Private Commercial Banks of Bangladesh. *GATR* .

Omomehin, V. (2021). *Credit Risk: Implementing Structural Models.* Cork: AIMS.

Pandey, P. (2019). A Deep dive into H2O's AutoML. *https://towardsdatascience.com/a-deep-dive-into-h2os-automl-4b1fe51d3f3e* .

Pandey, P. (2019). A Deep dive into H2O's AutoML. *https://towardsdatascience.com/a-deep-dive-into-h2os-automl-4b1fe51d3f3e* .

Polley, E. C. (2011). Super Learning. *https://link.springer.com/chapter/10.1007/978-1-4419-9782-1_3* , 43-66.

Reddy, D. Y. (2006). Challenges and Implications of Basel II for Asia (Central Bank Articles and Speeches). *BIS* .

Rehman, Z. U. (2019). Impact of Risk Management Strategies on the Credit Risk Faced by Commercial Banks of Balochistan." Financial Innovation 5(1). *Financial Innovation* .

TEAM, T. I. (2022). Credit Risk. *https://www.investopedia.com/terms/c/creditrisk.asp* .

Tomasz R. Bielecki, M. R. (2013). *Credit Risk: Modeling, Valuation and Hedging.* New York: Springer Finance.

Top 4 Major Challenges of Credit Risk Management in Banks: Why It's Important. (2021). *GDS Link* .

WAGNER, H. (2021). Analyzing a bank's financial statements. *https://www.investopedia.com/articles/stocks/07/bankfinancials.asp* .

Zhang, X. (2017). Essays in Credit Risk Management. *https://dokumen.tips/documents/zhang-xuan-2017-essays-in-credit-risk-management-abstract-credit-risk-management.html* .

Appendix

Plagiarism Report[1]

Investigating Super learner for Credit Risk Modeling in Mortgage Scenario

by Lalit Aggarwal

Submission date: 26-Aug-2022 10:37AM (UTC+0530)
Submission ID: 1887290814
File name: arner_for_Credit_Risk_Modeling_in_Mortgage_Scenario_-_Lalit.docx (845.42K)
Word count: 9033
Character count: 49485

[1] Turnitn report to be attached from the University.

Investigating Super learner for Credit Risk Modeling in Mortgage Scenario

ORIGINALITY REPORT

11%	7%	1%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ukcatalogue.oup.com Internet Source	2%
2	www.techtarget.com Internet Source	1%
3	www.bis.org Internet Source	1%
4	library1.nida.ac.th Internet Source	1%
5	Submitted to Heriot-Watt University Student Paper	1%
6	Submitted to Technological Institute of the Philippines Student Paper	1%
7	Submitted to University of Wollongong Student Paper	<1%
8	Submitted to Institute of Management Technology Student Paper	<1%

9	Sreejata Banerjee. "Basel I and Basel II compliance issues for banks in India", Macroeconomics and Finance in Emerging Market Economies, 2012 Publication	<1%
10	Submitted to University of Leeds Student Paper	<1%
11	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
12	www.indussoft.com Internet Source	<1%
13	Submitted to Segi University College Student Paper	<1%
14	rbidocs.rbi.org.in Internet Source	<1%
15	Submitted to Uttar Pradesh Technical University Student Paper	<1%
16	Submitted to Coventry University Student Paper	<1%
17	Submitted to London School of Commerce Student Paper	<1%
18	Submitted to University of Wales Swansea Student Paper	<1%

19	Submitted to University of Johannsburg Student Paper	<1 %
20	Submitted to RDI Distance Learning Student Paper	<1 %
21	www.unirepository.svkri.uniri.hr Internet Source	<1 %
22	Submitted to Arab Open University Student Paper	<1 %
23	Submitted to aun Student Paper	<1 %
24	www.cambridge.org Internet Source	<1 %
25	library.unisel.edu.my Internet Source	<1 %
26	Submitted to Universiti Teknologi MARA Student Paper	<1 %
27	www.mdpi.com Internet Source	<1 %
28	zimbabwe.wroclaw.pl Internet Source	<1 %
29	indianaccounting.org Internet Source	<1 %
30	research.gigaom.com Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

Publications in Conference

Paper submitted:

Lalit Aggarwal, Sanjeev Jha, J.B. Simha, “Investigating Super learner for credit Risk modeling in Mortgage Scenario”
9th International Conference on Business Analytics and Intelligence, IIMB.

Submission Date: 2nd October 2022.

Investigating Super learner for credit Risk modeling in Mortgage Scenario

Lalit Aggarwal
REVA Academy for Corporate Excellence – RACE, REVA University
Bangalore, Karnataka 560064
lalitaggarwal.ba05@reva.edu.in

Sanjeev Jha
REVA Academy for Corporate Excellence – RACE, REVA University
Bangalore, Karnataka 560064
sanjeev.ba05@reva.edu.in

J.B. Simha
REVA Academy for Corporate Excellence – RACE, REVA University
Bangalore, Karnataka 560064
jb.simha@reva.edu.in

***Abstract*—In the present industry Credit risk analysis is very important for the organization's business as well as its reputation in the market. In general, credit risk modelling is a method that lenders employ to assess the degree of credit risk involved in making a loan to a borrower.**

The objective of this paper is to investigate super learners for credit risk modeling in mortgage scenarios, with the help of AutoML. Different super learners can be defined as ensemble models of different base models and investigated against the credit risk dataset. For explaining the prediction of the super learner different result interpretation techniques i.e. SHapley Additive exPlanations (SHAP), Partial Dependence Plot (PDP) and Individual Conditional Expectation (ICE) have been used.

With the Use of H2O AutoML and the credit risk data, we are using multiple machine learning models, which comprise 25 models including Ensemble machine learning models "StackEnsemble_BestOfFamily" and "StackEnsemble_AllOfFamily" as well as base statistical machine learning models DeepLearning, Distributed Random Forest (DRF), Extremely Randomize Tree (XRT), Gradient Boosting Machine (GBM), and General Linear Models (GLM), here define AUC is defined as a stopping metric. Out of that "StackEnsemble_BestOfFamily" is giving an AUC of 71.08% and in base models, DRF is giving the highest Accuracy 88.02%, and an AUC of 70.5%. While in the interpretation SHAP, PDP, and ICE techniques are giving a very good explanation for every individual result as well as a prediction of the whole data set.

Hence with the help of AutoML techniques, multiple machine learning models are created in a short time without wasting a time on data preparation, data exploration, feature engineering, model selection, model training, and hyperparameter tuning. With the help of SHAP, PDP, and ICE any individual result can be explained to the customer or the end-user.

Keywords—AutoML, metalearner, StackEnsemble_BestOfFamily, StackEnsemble_AllOfFamily, DeepLearning, DRF, XRT, GBM, GLM, AUC, SHAP, PDP, and ICE.

I. INTRODUCTION

Credit risk is the likelihood that the borrower would not be able to pay back his debt on time or in whole. It refers to the possibility that the lender won't get the principal loaned or the interest due in a timely manner. This has the result of interfering with creditors' cash flow and raising the cost of collection. In dire circumstances, it can be necessary to write off all or a portion of the debt, incurring a loss for the lender. Knowing with absolute certainty how likely someone is to fail on a debt is very tough and complex. A competent evaluation of credit risk can also lessen the chance of losses due to default and late payments. The lender receives interest payments from the borrower as compensation for taking on the credit risk. Lenders or investors will either charge a higher interest rate or decline the loan opportunity entirely if the credit risk is higher. For instance, an applicant for the same loan with excellent credit and consistent income will pay a lower interest rate than one with a poor credit history [1].

A person's credit risk is influenced by a wide range of variables. As a result, determining the borrower's credit risk is an extremely challenging undertaking. Credit risk modelling has become important since so much money depends on our ability to predict a borrower's credit risk accurately. Credit risk modelling is the practise of utilising data models to ascertain two crucial facts. The likelihood that the borrower will miss a loan payment is the first. The second is how this default will affect the lender's financial situation [2].

To assess the credit risk of potential borrowers, financial institutions use credit risk models. Based on the validation of the credit risk model, they determine whether or not to authorise the loan and the loan's interest rate. With the development of technology, new approaches to modelling credit risk have appeared, such as credit risk modelling in R and Python. These include modelling credit risk utilising the newest analytics and big data tools. The way credit risk is modelled has also been impacted by other variables, such as the growth of economies and the ensuing rise of various types of credit risk [3].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Financial organisations have created sophisticated methods for calculating and controlling credit risk across all product categories of companies. A good understanding of frequently employed methods would, in the eyes of a regulator, increase the oversight of financial institutions. The necessity to calculate the amount of capital required to support the bank's exposures initially sparked interest in credit risk models. The job of the credit risk model is to take general economic conditions and specific variables as inputs and produce a credit spread as an output. There are two primary categories of credit risk models in this regard: structural and scale models. Based on the value of a company's assets and liabilities, structural models are used to determine the likelihood that the company would fail. If a company's assets are worth less than the amount of debt that must be repaid, it goes into default. Reduced models consider a random, exogenous source of failure [4].

With the help of the credit risk modelling methodology, risk measurement and management may be done in a flexible and personalised way. Models respond to changes in business lines, credit quality, market factors, and the economic climate as a result of their design and are influenced by these changes[5].

Additionally, the models reflect concentration risk within the portfolio and give banks the ability to examine marginal and absolute risk contributions. These model characteristics can help the bank's overall credit culture. The degree to which credit management has embraced models. Between banks, there are significant differences in how economic capital is distributed. While some banks have put in place systems to capture the majority of exposures across the organisation, only others do so for a specific business line or legal entity. In addition, banks frequently create distinct models for corporate and retail exposures, and not all banks account for both kinds of exposures. Internal model applications range from straightforward to straightforward and sophisticated [6].

Only a tiny portion of the banks the working group investigated already use the results of credit risk models in active portfolio management, but a sizable portion stated they intended to do so in the future. Applications now used include:

1. Calculating exposure and concentration limits;
2. Defining holding objectives for syndicated loans;
3. Pricing based on risk;
4. Enhancing the portfolio's risk and return profile;
5. Evaluation of business lines' or managers' risk-adjusted performance using Risk-Adjusted Return On Capital (RAROC)
6. Capital allocation in the economy. Institutions also use model estimates for direct computations or for validation purposes to generate or confirm reserves for loan losses [7].

The working group acknowledges that improved internal risk management in banking organisations can be a result of credit risk modelling. However, before employing the models in the process of establishing regulatory capital needs, the key challenges related to data limitations and model validation must be overcome. Description of the default process and other elements that affect loans The absence of previous credit and loan performance data as well as other modelled variables greatly restricts the quality. Since credit risk is measured across lengthy time horizons, which means the data span many years, the specification issues become more severe. To accurately estimate important factors, several credit cycles can be necessary. As a result of the present constraints, model parameters frequently involve some degree of simplification of assumptions and information collected from various sources. Since sensitivity testing of the model's vulnerability to such assumptions is not currently conventional practice, the impact of these alternatives on model risk estimation is uncertain [8].

II. LITERATURE REVIEW

In a commercial bank, credit management is a challenging functional area. It requires skilled handling, accurate risk assessment at each level, and sufficient assurance of the security of the monies exposed. Despite the greatest attempts, it is unable to create solid security requirements, which causes credit to be unstable or unpaid frequently. Therefore, effective risk management, asset and liability management techniques, and continual search for safer criteria for risk elimination constitute credit management. Such risk management expertise has been developed and put into use to help mitigate risk rather than eradicate it [9].

Credit risk is typically understood as the risk of default, or the risk of losing money if the borrower or counterparty fails to pay the bank the amount owed (principal or interest) according to a pre-arranged repayment schedule, on time. Value risk, or the danger of losing value as a result of the borrower's shift to a lower credit rating (opportunity costs connected with incorrectly pricing a new loan risk level), would be added to the definition to make it more complete [10].

Banks have devised procedures that enable them to quantify these risks and so derive the necessary amount of capital to maintain their company, known as economic capital, to hedge against volatility in default/impairment levels (as well as other types of risk). The method for determining the required minimum regulatory capital is outlined in Pillar.
1. Basel-I stated that this computation only took into account credit risk; however, in 1996, a calculation for market risk was introduced. Basel II increases the price to reflect operational risk.

A lot of effort needs to be done at the bank level as the entire financial sector worldwide works to implement the 2004 directive-based II Accord in some shape and intensity. Credit Risk Management gives users the tools they need to develop the necessary management structure, rules, processes, and practices for credit risk measurement [11].

One of Basel II's main advances is that it gives lenders an option between I and II, in contrast to Basel I, which only provided a single method for computing regulatory capital for credit risk. Using a standardized approach, Basel I is expanded upon by classifying exposures into several risk categories. Nevertheless, each risk category historically had a defined risk weight, e.g. The risk weights for the three Basel II categories (loans to governments, businesses, and banks) are established by the borrowers' given external credit ratings. Loans backed by residential real estate, among other categories that continue to be subject to fixed risk weights under Basel II, will bear a risk weight of 35%, up from 50% previously, if Loan-To-Value (LTV) climbs to 80%. This reduced weighting is in acknowledgment of historically low loss rates that frequently originated in residential mortgage loan portfolios across a wide range of economic situations in many nations [12].

2. A Basic approach Internal Rating Based (IRB) - The IRB approach will enable lenders to create their models for calculating their regulatory capital needs. Lenders assess the Probability of Default (PD) under the foundation IRB approach, and supervision sets the numbers for Loss Given Default (LGD), Exposure At Default (EAD), and Maturity Exposure (ME). For each exposure or kind of exposure, these numbers are fed into the lender's appropriate risk weight function to produce risk weights [13].

Under this strategy, banks can utilize credit risk limiters (collateral, guarantees, and credit derivatives) to lower capital depending on the market risk of the collateral instrument that others hold and the external credit assessment of recognised guarantors. Residential mortgages, Small and Medium-sized Enterprises (SME), and retail exposures all had their risk weights reduced. This method significantly differentiates exposures from transactions to increase the resulting risk-sensitive capital ratios [14].

III. METHODOLOGY

In the case of AutoML, data-related steps same as in the traditional approach. Case of manually understanding the relationship between data trends and their outcomes in different ML algorithms has been used. This gives good predictions based on different inputs. Machine learning algorithms are tedious and time-consuming and it is difficult for a non-technical person to develop a statistical model which can give good accuracy.

In this paper in place of traditional methodology, AutoML has been used. With the use of AutoML, most of the Data related steps are the same as the traditional approach but for selecting different machine learning algorithms, training and hyperparameter tuning could be automated.

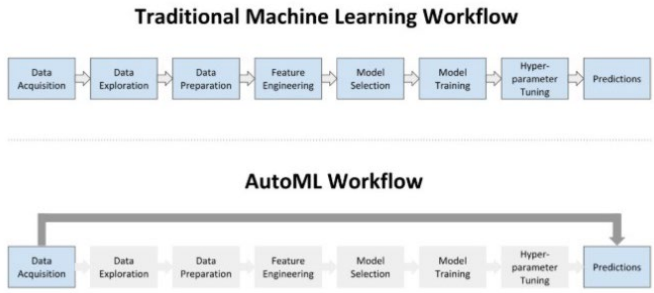


Figure 3.1 - AutoML vs Traditional Machine learning workflow

In Figure 3.1 it has been shown that in the case of AutoML basic time-consuming steps of data modeling like Data explorations, Data Preparation, Feature Engineering, Model selection, Model training, and Hyperparameter tuning can be bypassed. So in this way, a lot of time could be saved that researchers generally waste during these processes and concentrated on Data Collection and deployments of the best model [15].

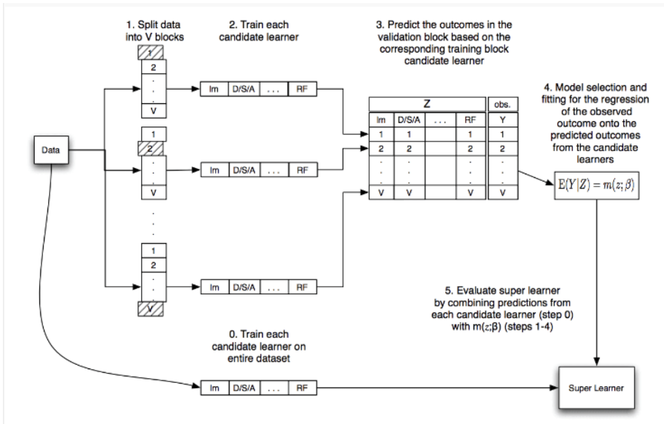


Figure 3.2: Process flow in AutoML

In Figure 3.2 Super learners can be described as the stack of different ensemble models, in this process, certain steps have been followed:

1. Cross-validation has been selected for the training dataset.
2. The maximum number of base models has also been specified.
3. For every base model these set has been followed:
   3.1. Base models have been evaluated by using cross-validation
   3.2. Store their predictions
4. Then trained the base models on the training dataset and stored predictions.
5. Metalearners on the stored predictions
6. The prediction has been done on the Test Data set [16].

Figure 3.3: AutoMLWorkFlow

In Figure 3.3 steps followed by AutoML have been shown, they are as follows:

1. Load and train and Test data.
2. Specify the response feature.
3. Running the AutoML by specifying the stopping criteria e.g. Maximum number of models, the maximum time to train is stopping metric or stopping rounds.
4. Analyze different models on the leader board and explore them based on our requirements usually ensemble stack model best of the family or All models give better accuracy.
5. The model could be saved and deployed in production for further use [17].

IV. Data Modeling

Data modelling in software engineering is the practice of utilising formal approaches to simplify a diagram or data model of a software system. It involves communicating information and data using text and symbols. A data model offers a design framework for creating a new database or redesigning old applications.

model_id	auc	logloss	aucpr	mean_per_class_error	rmse	mse
StackedEnsemble_BestOfFamily_3_AutoML_8_20220716_204454	0.710815	0.326445	0.286215	0.351219	0.306801	0.0941268
StackedEnsemble_AllModels_2_AutoML_8_20220716_204454	0.710589	0.326455	0.286061	0.344115	0.306765	0.0941046
StackedEnsemble_AllModels_1_AutoML_8_20220716_204454	0.710347	0.326806	0.284202	0.351165	0.306946	0.0942157
StackedEnsemble_BestOfFamily_2_AutoML_8_20220716_204454	0.710149	0.326896	0.284905	0.346986	0.30696	0.0942247
DRF_1_AutoML_8_20220716_204454	0.704951	0.33437	0.286943	0.34251	0.309266	0.0956456
GBM_2_AutoML_8_20220716_204454	0.701234	0.331476	0.263136	0.352149	0.309638	0.0958756
XRT_1_AutoML_8_20220716_204454	0.700893	0.337806	0.271597	0.354919	0.312968	0.0979493
GBM_4_AutoML_8_20220716_204454	0.700447	0.334691	0.270076	0.351465	0.309524	0.095805
GBM_3_AutoML_8_20220716_204454	0.700335	0.332374	0.260778	0.353529	0.309812	0.0959832
StackedEnsemble_BestOfFamily_1_AutoML_8_20220716_204454	0.699799	0.33095	0.261124	0.358147	0.30917	0.0955859
GBM_1_AutoML_8_20220716_204454	0.69844	0.332399	0.261681	0.357177	0.309774	0.0959597
GBM_5_AutoML_8_20220716_204454	0.696926	0.332226	0.259248	0.364794	0.309984	0.09609
GLM_1_AutoML_8_20220716_204454	0.668986	0.340973	0.218211	0.369995	0.313912	0.0985409
DeepLearning_1_AutoML_8_20220716_204454	0.612563	0.397489	0.195069	0.415548	0.32298	0.104316

Figure 4.1: Leader Board of AutoML

In Figure 4.1, the Leader Board of the AutoML has been shown, where 14 machine learning models with different metrics have been shown, out of that Best of Family ensemble machine learning models outperformed other ensembles as well as base models.

V. Data Evaluation

Plans for evaluation should outline the methods and sources used to gather data. Both quantitative and qualitative data have to be gathered within a structure that is in keeping with program goals, stakeholder expectations, and project schedules.

```
metalearner.varimp()
```

```
[('DRF_1_AutoML_8_20220716_204454',
  0.45089760422706604,
  1.0,
  0.5377771520927915),
 ('XRT_1_AutoML_8_20220716_204454',
  0.21615146100521088,
  0.4793803714609224,
  0.25779981093343934),
 ('GBM_2_AutoML_8_20220716_204454',
  0.10483089834451675,
  0.23249380205561107,
  0.12502985474869172),
 ('DeepLearning_1_AutoML_8_20220716_204454',
  0.03858282417058945,
  0.0855689269778414l,
  0.046017013857779586),
 ('GLM_1_AutoML_8_20220716_204454',
  0.027984146028757095,
  0.062063195205323496,
  0.033376168367297786)]
```

Figure 5.1: Importance of different Base models in Best of Family Stacked Model.

Figure 5.1, it is shown how the different base models contribute to the best family stacked ensemble model. Here it has been noticed the DRF was the playing most important role in making the decision.

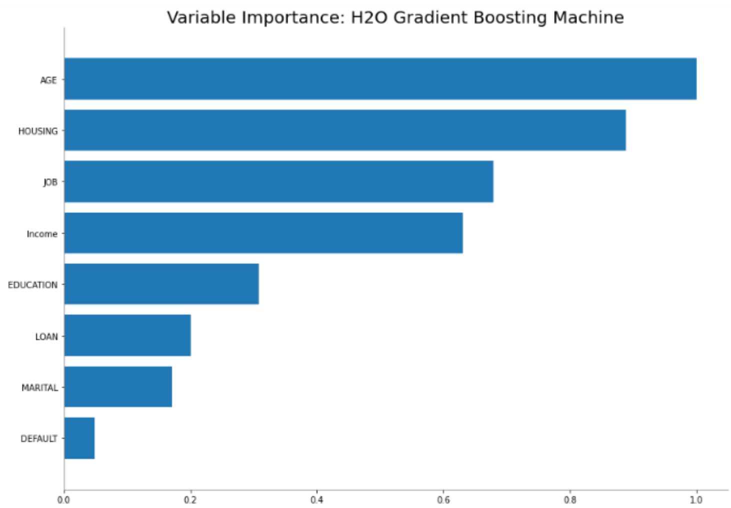


Figure 5.2: Variable Importance for the Gradient boosting model

In Figure 5.2, it is shown how the Gradient boosting base model treats different independent features for making any decision. As Gradient boosting gives more importance to Age, housing, and Job as compared to Marital, Loan, and Education.

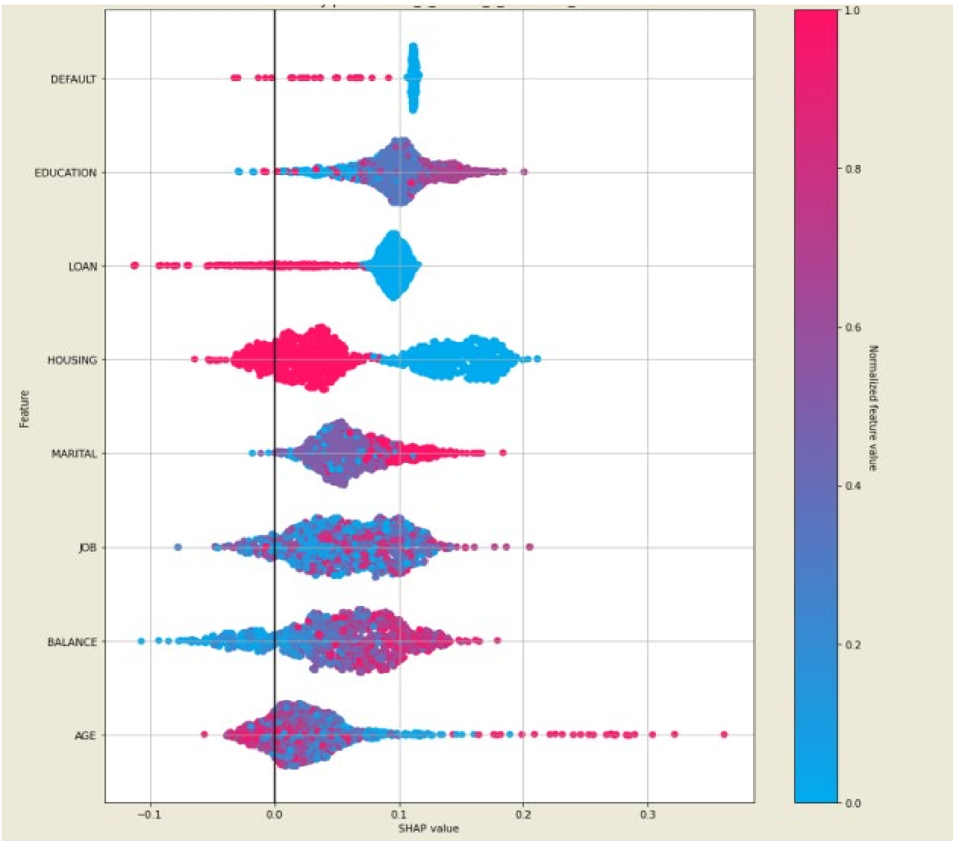


Figure 5.3: SHAP Explanation

In Figure 5.3, A positive SHAP number indicates a positive impact on prediction, which causes the model to predict 1, according to the SHAP explanation (e.g. Loan approved). The model predicts 0 since a negative SHAP score indicates a negative impact (e.g. Loan did not approve).

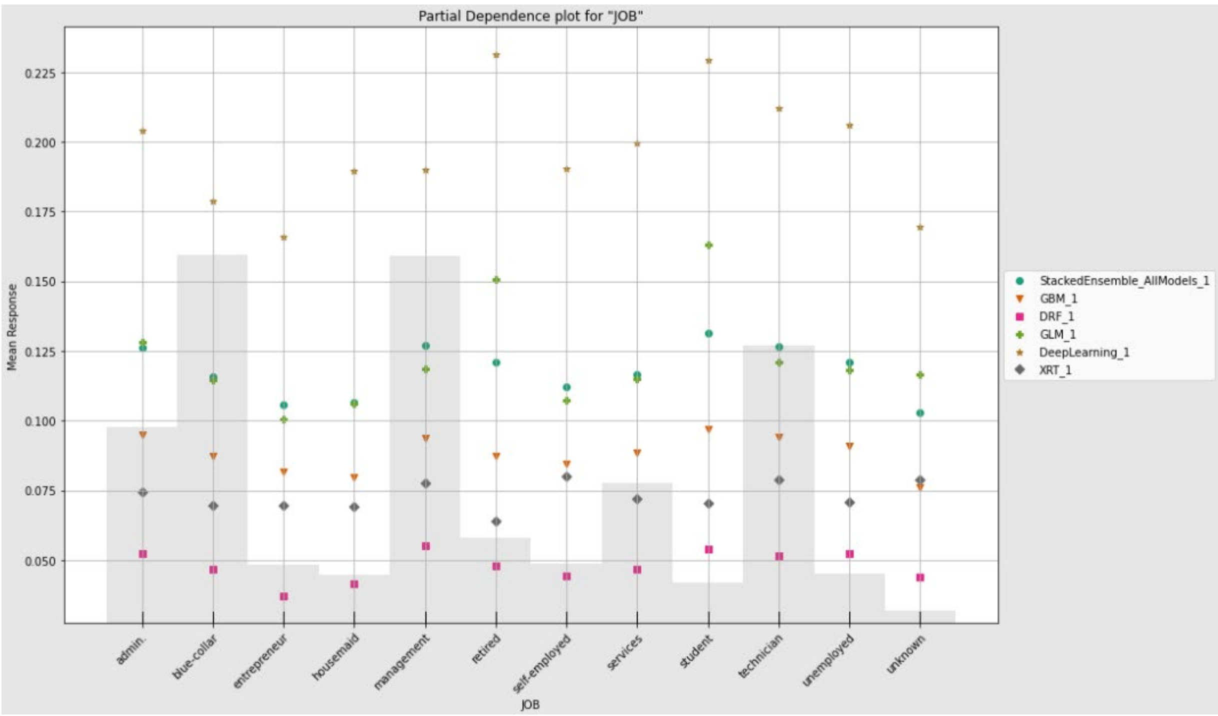


Figure 5.4: Partial Dependence Plot Effect of Job on different models

Figure 5.4, it has been shown the effect of Job independent features on different ensembles as well as base models with the help of PDP. GLM_1 gives more importance to Admin., self-Employed and unemployed persons. Whereas GBM_2 gave importance to the person who is working as Admin., management, student, and unemployed.

Individual row prediction

AGE	JOB	MARITAL	EDUCATION	DEFAULT	HOUSING	LOAN	LOAN_STATUS	Income
42	admin.	married	secondary	no	yes	no	no	1173

predict	no	yes
no	0.934694	0.0653063

Figure 5.5: Single Row prediction by Best Base Model

In Figure 5.5, checking the individual customer result and their probability of approval and not approval with the best base model. In this case, as per the base model probability of rejecting the loan application is 93.47% and for getting approval is only 6.53%. Which is the same as real Loan status.

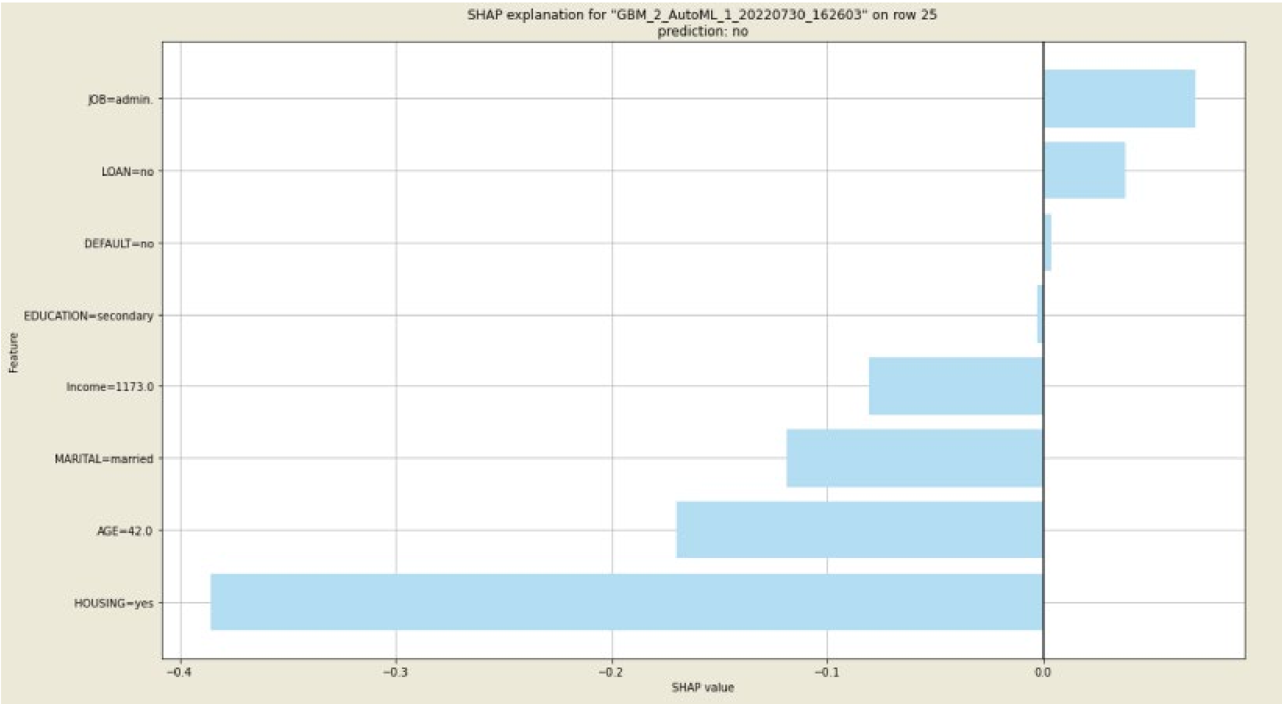


Figure 5.6: SHAP Explanation of the result

Figure 5.6, it has shown the explanation of individual customer results with the help of SHAP. Job and loan positively impact the chance of loan approval whereas Housing, age, marital status, and income affect more in rejecting their loan request.

VI. Analysis and Result

In this study, it has been shown that with the use of Automate Machine learning technique in the combination of different explanations e.g. SHAP, PDP and ICE. Different complex, as well as base models, can be developed in a short time with minimal knowledge of programming and compared with different metrics. Researchers could save a lot of time in developing, training, or tuning the different machine learning models, they could spend that time on data collection and understanding it. On rejecting any loan application end user can explain the reason or features behind that so the customer can also be satisfied with the explanation.

VII. Conclusions and Recommendations for Future Work

Hence it has been shown Automated Machine learning can be a very useful technique in the field of financial and banking sectors. Researchers or machine learning engineers can use Automated Machine learning in a combination of explanation techniques in providing very efficient models to the industry and provide insight also how the complex model is working and explain every prediction done by it.

VIII. References

[1] "Credit Risk: Definition, Role of Ratings, and Examples." https://www.investopedia.com/terms/c/creditrisk.asp (accessed Sep. 14, 2022).

[2] C. Bluhm, L. Overbeck, and C. Wagner, "Introduction to Credit Risk Modeling," *Introd. to*

Credit Risk Model., Apr. 2016, doi: 10.1201/9781584889939.

[3] "Credit Risk: Modeling, Valuation and Hedging - Tomasz R. Bielecki, Marek Rutkowski - Google Books." https://books.google.co.in/books?hl=en&lr=&id=GWzmCAAAQBAJ&oi=fnd&pg=PR5&dq=Credit+Risk:+Modeling,+Valuation+and+Hedging,+Tomasz+R.+Bielecki,+Marek+Rutkowski&ots=6Zobowx0pC&sig=StBEDSuUxXWSh9ibAuqeFcmUF6I#v=onepage&q=Credit Risk%3A Modeling%2C Valuation a (accessed Sep. 14, 2022).

[4] C. A. Fernandes and C. Dissertação, "Structural Models in Credit Risk," 2014.

[5] D. J. Teece, "Business models and dynamic capabilities," *Long Range Plann.*, vol. 51, no. 1, pp. 40–49, Feb. 2018, doi: 10.1016/J.LRP.2017.06.007.

[6] "RBC20 - Calculation of minimum risk-based capital requirements." https://www.bis.org/basel_framework/chapter/RBC/20.htm?inforce=20230101&published=20201126 (accessed Sep. 14, 2022).

[7] "Risk-Adjusted Return On Capital (RAROC)." https://www.investopedia.com/terms/r/raroc.asp (accessed Sep. 14, 2022).

[8] J. Bessis, "Risk Management in Banking," p. 840 p., 2009, Accessed: Sep. 14, 2022. [Online]. Available: https://hal-hec.archives-ouvertes.fr/hal-00494876

[9] "Reserve Bank of India - Index To RBI Circulars." https://www.rbi.org.in/scripts/BS_CircularIndexDisplay.aspx?Id=4964 (accessed Sep. 14, 2022).

[10] "Amendment to the capital accord to incorporate market risks," Nov. 2005, Accessed: Sep. 14, 2022. [Online]. Available: https://www.bis.org/publ/bcbs119.htm

[11] "Basel II: Revised international capital framework." https://www.bis.org/publ/bcbsca.htm (accessed Sep. 14, 2022).

[12] B. - Bank for International Settlements, "Y V Reddy: Challenges and implications of Basel II for Asia (Central Bank Articles and Speeches)," 2006.

[13] "Modelling credit risk | Bank of England." https://www.bankofengland.co.uk/ccbs/modelling-credit-risk (accessed Sep. 14, 2022).

[14] "Zhang, Xuan (2017) Essays in credit risk management. Abstract Credit risk management is becoming more and more important in recent years. Credit risk refers to the risk that an obligor." https://dokumen.tips/documents/zhang-xuan-2017-essays-in-credit-risk-management-abstract-credit-risk-management.html?page=1 (accessed Sep. 14, 2022).

[15] "Using an AutoML H2O Model to Predict Attrition and LIME to Explain the Predicted Class - DZone AI." https://dzone.com/articles/using-an-automl-h2o-model-to-predict-attrition-and (accessed Sep. 14, 2022).

[16] "Super Learning | SpringerLink." https://link.springer.com/chapter/10.1007/978-1-4419-9782-1_3 (accessed Sep. 14, 2022).

[17] "A Deep dive into H2O's AutoML. An overview of Automated Machine… | by Parul Pandey | Towards Data Science." https://towardsdatascience.com/a-deep-dive-into-h2os-automl-4b1fe51d3f3e (accessed Sep. 14, 2022).