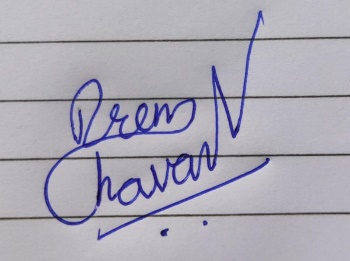
horizontal line

Project Summary

|  |  |
| --- | --- |
| Batch details | PGP-DSE-FT PUNE Nov’22 |
| Team members | Prem Chavan, Rupal Nikum, Chaitanya Thipse, Ashish Bikkad, Atharva Shastri, Prakhar Choudhary |
| Domain of Project | Finance |
| Proposed project title | INTEREST RATE PREDICTION |
| Group Number | Capstone\_Group\_4 |
| Team Leader | Prem Chavan |
| Mentor Name | Anjana Agrawal |

Date: 27/02/2023



Signature of the Mentor Signature of the Team Leader

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr. NO** | **Topic** | **Page No** |
| 1 | Overview | 3 |
| 2 | Business problem goals | 5 |
| 3 | Topic survey in depth | 7 |
| 4 | Critical assessment of topic survey | 9 |
| 5 | Methodology to be followed | 10 |
| 6 | References | 12 |

**Project Details**

# OVERVIEW:

**1.1 Objective:**

The objective of this project is to build a machine learning model that predicts the interest rate category (1/2/3) of a loan application based on various borrower features.

**1.2 Data:**

The dataset contains the following variables:

* Loan\_ID: A unique identifier for the loan.
* Loan\_Amount\_Requested: The listed amount of the loan applied for by the borrower.
* Length\_Employed: Employment length in years.
* Home\_Owner: The home ownership status provided by the borrower during registration. Values are: Rent, Own, Mortgage, Other.
* Annual\_Income: The annual income provided by the borrower during registration.
* Income\_Verified: Indicates if income was verified, not verified, or if the income source was verified.
* Purpose\_Of\_Loan: A category provided by the borrower for the loan request.
* Debt\_To\_Income: A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested loan, divided by the borrower’s self-reported monthly income.
* Inquiries\_Last\_6Mo: The number of inquiries by creditors during the past 6 months.
* Months\_Since\_Deliquency: The number of months since the borrower's last delinquency.
* Number\_Open\_Accounts: The number of open credit lines in the borrower's credit file.
* Total\_Accounts: The total number of credit lines currently in the borrower's credit file.
* Gender: Gender.
* Interest\_Rate: Target variable: Interest rate category (1/2/3) of the loan application.

**1.3 Methodology:**

Data Preprocessing: Clean, transform, and prepare the data for analysis.

Exploratory Data Analysis (EDA): Analyze and visualize the data to gain insights into the relationships between the variables.

Feature Engineering: Create new features or modify existing ones to improve the model's predictive power.

Model Building: Train and test different machine learning models to identify the best performing one based on metrics such as accuracy, precision, and recall.

Model Evaluation: Evaluate the model's performance on a separate validation dataset and fine-tune the model if necessary.

Model Deployment: Deploy the final model and integrate it into the existing system.

**1.4 Expected Outcome:**

The expected outcome of this project is a machine learning model that can accurately predict the interest rate category of a loan application based on various borrower features. The model can help the lending institution to make informed decisions about loan approvals and interest rates.

# Business problem statement (GOALS) :

The difficulty of increasing earnings and reducing risks for banks and lending organizations which are dealing with significant administrative costs and a fiercely competitive environment is our company's concern. Our objective is to find methods that can assist banks in utilising machine learning and AI to enhance business processes, boost productivity, and keep a profit.

**2.1 Business Problem Understanding:**

With its wide range of accounts, loans, and investment services, the banking industry is essential to the economy. Nonetheless, there are a number of difficulties that banks must overcome, such as high overhead expenses, fierce competition, and the necessity to preserve profitability. The old method of charging a difference in interest rates to generate income is no longer sufficient, and banks must look into innovative approaches to increase profits and reduce risks. Machine learning and AI are being used by numerous banks all around the world to streamline processes, cut costs, and improve customer satisfaction. Before integrating new technologies, banks must carefully weigh the risks and rewards because doing so necessitates a large investment in infrastructure and talent.

**2.2 Business Objective:**

Our goal is to assist banks and lending organisations in determining the best methods for utilising AI and machine learning to increase revenues and reduce risks. By thoroughly analysing the current market trends, weighing the advantages and disadvantages of various technologies, and making implementation suggestions, we hope to achieve this.

* 1. **Approach:**

We'll take the next step in order to accomplish our goal:

Perform a thorough examination of the banking industry to pinpoint the major difficulties and opportunities.

Analyze the possible advantages and disadvantages of applying AI and machine learning to the banking industry, taking into account the effects on customer experience, operational effectiveness, and profitability.

Get case studies of successful machine learning and AI use in banking operations, as well as best practises.

Provide advice to banks and lending companies on how to use AI and machine learning to increase revenues and reduce risks while taking into consideration their unique business requirements and resources.

**2.4 Conclusions:**

By leveraging machine learning and AI, banks and lending firms can improve their operations, reduce costs, and enhance customer experience, thereby increasing profitability and competitiveness. However, these technologies also come with risks, including data privacy concerns, cyber security threats, and regulatory compliance issues. It is essential for banks to carefully evaluate the risks and benefits before adopting these technologies and to ensure that they are used ethically and transparently to maintain customer trust and confidence.

# TOPIC SURVEY IN BRIEF:

**3.1 Problem understanding:**

Banks and lending firms offer various kinds of accounts and provide loans based on the requirements. Apart from it, there are other various activities like investments in market and different funds. Overall, the banking sector has a wide impact on the economy directly and indirectly. We know that banks have massive overheads, with thousands of employees to pay and hundreds of branches to maintain. To maintain profitability, banks must take large margins on the money that passes through them. Earning out of the difference in interests (what it pays to depositors and what it charges from borrowers) is the main source of revenue for any bank and has been the key element in the functioning of all traditional financial institutions. We can use strategies which will permit these clubs maximise profits and minimize the risks involved in the sector.

**3.2 Current solution to the problem:**

Financial firms like banks and lending groups are generally free to determine the interest rate they will pay for deposits and charge for loans, but they must take the competition into account, as well as the market levels for numerous interest rates and policies.

There are many types of interest rates and loan products. When it comes to setting rates, certain loans, such as residential home mortgage loans, may not be based on the prime rate fixed by the firm but rather according to the policies set by the government.

These firms use an array of factors to set interest rates mainly to maximize profits for their shareholders. On the flip side, consumers and businesses seek the lowest rate possible. They start from client inputs, such as credit score, collateral provided, down payment and duration for the loan, employment status, assets owned and so on to calculate the optimum interest rates.

**3.3 Proposed solution to the problem:**

So our problem is dealing with the identifying the customer according to their loan dispensing category in other words we can say that we have a to assign a label to the customer given various attributes. To overcome this problem we will be applying Supervise learning classification ML models on the given dataset. We'll be going through various steps for that as in the data set there our missing values so we'll be imputing them , we have to develop new features as well , compare the accuracy among different models and should come to conclusive model which will be predicting the best and giving a good accuracy.

**3.4 Reference to the problem:**

The data for the problem is an example of Peer to peer lending (or P2P lending) Club which is one of the most innovative financial products of recent times. It enables creditworthy borrowers lower their cost of loans and individual lenders/investors to lend directly to their peers and community thereby earning higher returns. Lending clubs provide a virtual market place where borrowers and lenders can interact directly, without having to go through the traditional financial intermediaries like banks, who have become such behemoths in today’s time that they dictate all terms and conditions for both borrowers and lenders. The project will use machine learning algorithms that leverage different determining factors of a loan applicant. Selection of significant factors will help develop a prediction algorithm which can estimate loan interest rates based on clients’ information. On one hand, knowing the factors will help consumers and borrowers to increase their credit worthiness and place themselves in a better position to negotiate for getting a lower interest rate. On the other hand, this will help lending companies to get an immediate fixed interest rate estimation based on client’s information. Here, our goal is to use a training dataset to predict the loan rate category (1 / 2 / 3) that will be assigned to each loan in our test set. We will use combination of the features in the dataset to make our loan rate category predictions.

# 

# CRITICAL ASSESSMENT OF TOPIC SURVEY:

**4.1 Find the key area, gaps identified in the topic survey where the project can add value to the customers and business:**

One of the main gaps in interest rate prediction is the lack of accuracy. Customers and businesses would benefit greatly from a model that is more accurate in predicting interest rates. By improving the accuracy of interest rate prediction, businesses can make better decisions regarding investments, lending, and other financial activities. Along with there are some more key gaps in interest rate prediction are timing of the predictions. By providing timely predictions, businesses can make better decisions based on current market conditions, another gap is customers and businesses may have unique needs and preferences when it comes to interest rate prediction.

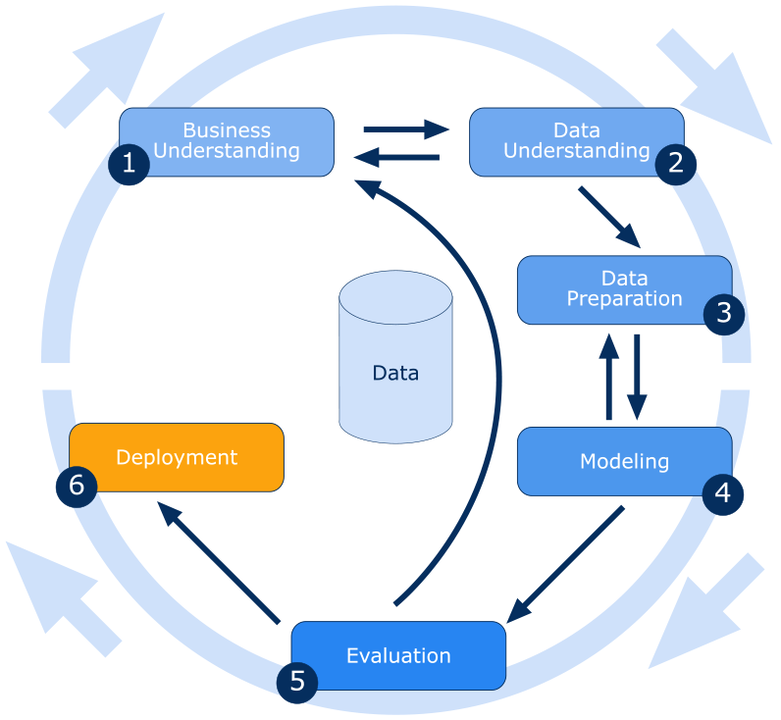
**4.2 What key gaps are you trying to solve?**

Lack of accuracy, Timeliness of predictions are some of the gaps which need to solve. By addressing these key gaps, the project can add significant value to customers and businesses, leading to improved financial outcomes and better decision-making, can improve the accuracy of their predictions, and make the predictions more accessible to a wider audience.

# 

# 

# METHODOLOGY TO BE FOLLOWED (Explain each steps from 1-5)



**5.1 Business Understanding:**

Understanding the business issue and determining the project's objective are the first steps. In this instance, the objective is to forecast, based on numerous borrower and loan parameters, the interest rate category (1/2/3) of a loan application. It's also critical to comprehend how the forecast will be applied in a professional setting, such as when approving loans or evaluating risks.

**5.2 Data Understanding:**

The following step is to compile and research the project's pertinent data. This include locating the data's sources, cleaning and preparing it for analysis, and then studying it to draw conclusions and spot potential problems.

The variables provided are: IN OVERVIEW SECTION above

**5.3 Data Preparation:**

Once the data has been gathered and explored, the next step is to prepare the data for modeling. This includes selecting the relevant features, handling missing values, encoding categorical variables, and scaling or normalizing the data.

In our Dataset there are 146201 Missing Values

In ‘Annual\_Income’ column is Right skew and have Outlier

Target Variable is not much balance Variable

**5.4 Modeling:**

The next step is to choose and train the suitable machine learning models using the prepared data. Decision trees, neural networks, and logistic regression are some examples of this. To evaluate the models' performance, they will be trained on a training dataset and tested on a different validation dataset.

**5.5 Evaluation:**

The performance of the models needs to be evaluated to determine which model is best suited for the problem at hand. This involves comparing the performance of different models on the validation dataset and selecting the one with the highest accuracy or other relevant metrics.

**5.6 Deployment:**

The model must then be put into production so that it can be utilized to generate predictions. This can entail including the model in a web application or other piece of software.

**Throughout the project, it is important to monitor and evaluate the performance of the model in the production environment and make adjustments as needed.**

# REFERENCES

* **Kaggle**:- The dataset is hosted on Kaggle, a platform for dataset and machine learning competitions. This platform provides a range of tools and resources for data scientist and machine learning practitioners, including datasets, notebooks, competitions etc.
* Link of Dataset:- <https://www.kaggle.com/code/kerneler/starter-interest-rate-prediction-72a18110-4/data?select=train.csv>

**Notes For Project Team**

*Sample Reference for Datasets (to be filled by team and mentor)*

|  |  |
| --- | --- |
| Original owner of data | Kaggle |
| Data set information | The dataset contains loan application information including loan amount, borrower employment and income, credit history, and interest rate. |
| Any past relevant articles using the dataset | Not Found |
| Reference | Kaggle |
| Link to web page | <https://www.kaggle.com/code/kerneler/starter-interest-rate-prediction-72a18110-4/data?select=train.csv> |

### 