INTEREST RATE FORECASTING

# Acknowledgement

The work “would not have been possible” without the contribution of **(project teacher name)** of the university **(insert college/university name)**. I am indebted to **(insert other teachers who have some contribution)** who have offered continuous support while preparing the project.

I am also grateful to all those "with whom" I had the opportunity to do the work and complete the project. 'Each member" of the "dissertation committee" have offered and provided' "professional guidance" and have given me great advice while completing the project.

On a personal note, I am also grateful to my family members who have offered me continuous support while I was completing the project. Without the help and support of them, this project would not have been completed.

# Abstract

Abstract interest forecasting is the main subject of the research that has been elaborated using python.  In the introduction chapter, the background of the study which provides a deep origin of the research has been mentioned. Along with the aim of the study - to develop and evaluate advanced ML models for accurate prediction of interest rates using the Lending Club dataset to enhance decision-making processes in financial markets and lending institutions. The literature review chapter unleashed the rich literature available related to the topic. Followed by the methodology chapter, that outlines the method and approaches implemented in the research in order to gain maximum benefits.  It describes different tools that have been used in analysing the data.  The implementation chapter describes the implementation of the tool done in order to outsource comprehensive results.  The result and evaluation chapter explains the results of various data that has been driven from the lending club dataset. Finally the conclusion chapter deals with the discussion of the  result and touches on   recommendation  and  future work that can be initiated in future ahead.

**Table of Contents**

[CHAPTER 1: INTRODUCTION 5](#_Toc179046058)

[1.1 Background of the Study 5](#_Toc179046059)

[1.2 Problem Statement 5](#_Toc179046060)

[1.3 Aim and Objectives 6](#_Toc179046061)

[1.4 Research Questions 6](#_Toc179046062)

[1.5 Scope of the Study 6](#_Toc179046063)

[1.6 Significance of the Study 7](#_Toc179046064)

[1.7 Structure of the Study 8](#_Toc179046065)

[CHAPTER 2: LITERATURE REVIEW 9](#_Toc179046066)

[2.1 Introduction 9](#_Toc179046067)

[2.2 Empirical Study 9](#_Toc179046068)

[2.3 Theories and Models 15](#_Toc179046069)

[2.4 Literature Gap 16](#_Toc179046070)

[2.5 Summary 16](#_Toc179046071)

[CHAPTER 3: RESEARCH METHODOLOGY 17](#_Toc179046072)

[3.1 Introduction 17](#_Toc179046073)

[3.2 Methodology 17](#_Toc179046074)

[3.2.1. Data Selection 17](#_Toc179046075)

[3.2.2. Modelling 18](#_Toc179046076)

[3.2.3. Evaluation 19](#_Toc179046077)

[3.3 Tools 19](#_Toc179046078)

[3.3.1 Python (platform) 19](#_Toc179046079)

[3.3.2 Scikit Learn 19](#_Toc179046080)

[3.3.3 Pandas 20](#_Toc179046081)

[3.3.4 Numpy 20](#_Toc179046082)

[3.3.5 Matplotlib 20](#_Toc179046083)

[3.3.6 Seaborne 20](#_Toc179046084)

[3.4 Summary 21](#_Toc179046085)

[Chapter 4: Implementation 22](#_Toc179046086)

[4.1 Introduction 22](#_Toc179046087)

[4.2 Description of Dataset 22](#_Toc179046088)

[4.3 Exploratory Data Analysis 23](#_Toc179046089)

[4.4 Data Partitioning 23](#_Toc179046090)

[4.5 Model Implementation 24](#_Toc179046091)

[4.6 Summary 25](#_Toc179046092)

[CHAPTER 5: RESULTS AND EVALUATION 27](#_Toc179046093)

[5.1 Introduction 27](#_Toc179046094)

[5.2 Model Output 27](#_Toc179046095)

[5.3 Summary 47](#_Toc179046096)

[CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS 48](#_Toc179046097)

[6.1 Introduction 48](#_Toc179046098)

[6.2 Discussion and Conclusion 48](#_Toc179046099)

[6.3 Recommendation 49](#_Toc179046100)

[6.4. Future Work 49](#_Toc179046101)

[References 51](#_Toc179046102)

**Table of Figures**

[Figure: Research Structure 8](#_Toc179045384)

[Figure : Data Implementations 27](#_Toc179045385)

[Figure : Data Imports and process 28](#_Toc179045386)

[Figure : Data fetchsing 29](#_Toc179045387)

[Figure : Implementing data shapes 30](#_Toc179045388)

[Figure : Data implementations results 31](#_Toc179045389)

[Figure : Data Dropdowns implementations 34](#_Toc179045390)

[Figure : Data informations and generations 36](#_Toc179045391)

[Figure : Annual income statements 37](#_Toc179045392)

[Figure: Distributions Rates 38](#_Toc179045393)

[Figure : Interest rate calculations 39](#_Toc179045394)

[Figure : Annual income and interest rate comparisons 40](#_Toc179045395)

[Figure : SVR vs MLP Performance 41](#_Toc179045396)

[Figure : calculate MAE 43](#_Toc179045397)

[Figure : Interest and loan infrastructure 44](#_Toc179045398)

[Figure : loans generations 45](#_Toc179045399)

[Figure : Averages interest rate over time 46](#_Toc179045400)

[Figure : Visualise key features influencing interest rates 47](#_Toc179045401)

# 

# CHAPTER 1: INTRODUCTION

Financial markets and lending institutes have actively taken part in lending, borrowing and investing money in recent times. The research deals with evaluating advanced ML models for accurate prediction of interest rates using the Lending Club dataset. The introduction chapter lays a strong foundation that will discuss the background of the topic, problem statement, aim, objective, research question and scope of the study.

## 1.1 Background of the Study

The shift from the traditional approach to the modern approach where the financial market is more focused on data-driven decision-making to enhance overall accountability is used by all major financial institutions like banks and investment companies. The growth in interest rates from 0% in 2022 to 4-5% in 2024 has been seen in the central banks of major countries like India, Brazil, Canada and China (Statista, 2024). The use of machine learning is done in the form of algorithm trading, fraud detection, portfolio management, underwriting loans, and other areas of finance. Machine learning has gained notable importance in analysing large data, discovering patterns, and limiting errors, which results in enhanced workability for these institutes. However, as new areas have been explored like digital loan processing to customers, it is important to have proper information about the customer and its financial history. This information not only helps in understanding the overall capability of the customer but also aids in facilitating a broader range of services. Predictive analysis is used by financial institutes to get a better visualisation of future outcomes. As finance-related tasks are sensitive in nature, the fusion of predictive analysis has accelerated the overall performance by simplifying and providing valuable insight into every transaction. The implementation of machine learning in the evaluation of large data sets and how it acts as a growth factor is important to investigate.

## 1.2 Problem Statement

The rise in interest rates and diversified range of loans can be a complex task to manage and predict its future outcomes. The accuracy of machine learning cannot be questioned as it uses data analytics to process data, however, accurate data interpretation can be a difficult task if the data is large in volume.

## 1.3 Aim and Objectives

**Aim**

The aim of the study is to develop and evaluate advanced ML models for accurate prediction of

interest rates using the Lending Club dataset to enhance decision-making processes in financial

markets and lending institutions.

**Objective**

* To conduct comprehensive exploratory data analysis (EDA) by using the Lending Club dataset to identify key features influencing interest rates
* To execute a Support Vector Regression (SVR) model for predicting interest rates, assessing the model’s performance and accuracy
* To implement and optimise a Multilayer Perceptron ANN model for interest rate prediction, comparing its efficacy with the SVR model
* To analyse and compare the predictive performance of the SVR and MLP models using appropriate metrics, providing insights into their suitability for financial forecasting

## 1.4 Research Questions

1. What is the process for conducting comprehensive exploratory data analysis (EDA) by

using the Lending Club dataset to identify key features influencing interest rates.

2. How to execute a Support Vector Regression (SVR) model for predicting interest rates,

and assessing the model’s performance and accuracy.

3. Why implement and optimise a Multilayer Perceptron ANN model for interest rate

prediction, comparing its efficacy with the SVR model?

4. How to analyse and compare the predictive performance of the SVR and MLP models

using appropriate metrics, providing insights into their suitability for financial

forecasting?

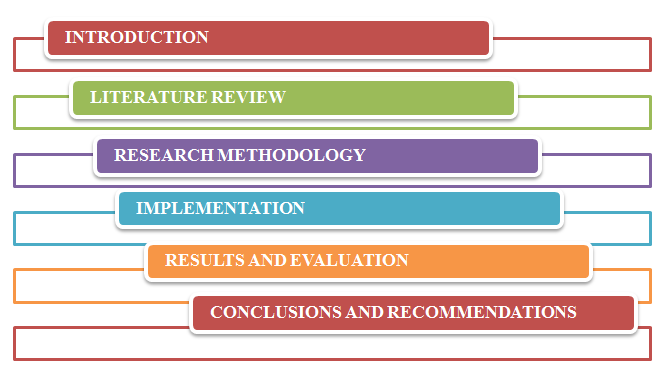
## 1.5 Scope of the Study

Interest forecasting is a new trend that is used by banks and investment companies. The scope of the study is to identify the role of ML in analysing large datasets. Along with this, the study will evaluate how financial institutions are making remarkable profits through interest forecasting and some of its drawbacks. The efficiency of the support vector regression (SVR) model for predicting interest rates and its comparison with the multilayer perceptron ANN model is important to discuss and can be used in algorithmic trading.

## 1.6 Significance of the Study

Interest rate forecasting is an integral part of transactions related to loans. Errors in interest forecasting can impact financial institutions negatively as wrong predictions can lead to inaccurate decision-making. The research significantly discusses the engagement of ML in performing accurate predictive analysis. Correct interest forecasting has the capability to design strategies that will always gain profits. Moreover, the forecasting feature can be used by both small-size and large-size finance companies so that every loan is processed without any doubt and concussion. As interest is being used explicitly in all types of transactions, be it home loans, car loans or EMIs, it is vital to examine the role of ML in the lending and borrowing process. An example of losses due to wrong decision-making is Toyota, it was reported that the company was holding debts for a longer period of time. In the year 2023, the Short-term and current portion of long-term debt was ¥12,305,639 and in the year 2024, it is ¥15,406,284 (Toyota, 2024). In this case, the intervention of machine learning could easily process data and aid in better decision-making.

## 1.7 Structure of the Study



#### Figure: Research Structure

(Source: Self- Created )

# 

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

The increasing competitiveness and the frequency of uncertain events have been crucial factors contributing to the financial distress of organisations. Hence predictive financial models are necessary for mitigating the risks associated with financial losses. Thus, this chapter will provide an empirical study of the existing literature such as journals and scholarly articles, on forecasting the interest rate. It will further highlight the significance of MLP and SVR models in offering accurate financial forecasting through a detailed analysis and comparison between the two. It will also identify the relevant theories and models on the significance of money and other assets.

## 2.2 Empirical Study

**Concept of Interest Rate and its Role in Global Economics**

According to Hofmann et al. (2021), the rate of interest can be referred to the cost of borrowing money or the return obtained from investments being done. It assumes an important part in the financial decisions made by both individuals as well as the global economy. The rates of interest help in the regulation of the economy and are determined by the central banks. The rate of interest can influence the way in which consumers and businesses can spend and save. The rates of interest that are higher can make it quite expensive to borrow and can discourage spending as well as investment (Hofmann *et al.,* 2021). The rates of interest that are lower can stimulate borrowing by causing a reduction in the cost of loans. The rates of Interest can also impact the value of currency with rates of interest that are attracting more foreign investment. This leads to an increase in the demand for the currency of a country and helps in strengthening the markets across the globe.

According to Lee and Werner, 2023, a depreciation in currency can be caused by a lower rate of interest. This can impact international trade with currencies that are stronger making exports much more expensive with imports becoming cheaper with a currency that is weaker having a completely opposite impact. Countries having higher rates of interest are more likely to attract foreign direct investment (FDI) with the countries having lower rates experiencing outflows of capital as investors invest in other countries for earning better returns (Lee and Werner, 2023). The rates of interest can help in regulation of economic growth with inflation being controlled and trade and investment being balanced.

**Lending Club and its Role in Global Economics**

According to Chang et al.2022*,* Lending Club is a platform that facilitates peer-to-peer (P2P) lending and has represented a significant transformation in the way in which individuals and businesses can access loans and assume a prominent part in global economics. “*One of the reasons why the P2P lending platform can break through the traditional lending market and attract many investors is its great ability to meet the capital demand*” (Chang *et al.,* 2022, pp. 304). The lending clubs can help small businesses to meet their capital demand by easily lending them the required funds. Lending Club can offer more benefits in comparison to traditional financial institutions as it is much more accessible and can help to obtain credit in a convenient manner. This is done by Lending Club by establishing a direct connection between the borrowers and investors. Benefits can be obtained from borrowers due to the rate of interest being lower. The investors are able to earn returns that are higher than conventional methods of saving and investment which results in democratisation of financial services.

As per Croux et al. (2020), platforms like Lending Club can help in strengthening finance in regions where limited access to credit through traditional banks is available. certain populations groups and small businesses that find it difficult to get access to credit facilities can easily get the funds that they need from lending clubs for the purpose of growth and investment. Economic activity can be conveniently stimulated when the access to credit can be improved which can lead to creation of jobs as entrepreneurial ventures are likely to increase with accessible credit facilities (Croux *et al.,* 2020). Lending Club can also help in better flow of capital by providing opportunities to international investors for participating in loans. This opportunity of making investments across different borders can help in allocation of capital in different regions having the potential of higher growth. As a result of this, these regions are likely to experience economic development.

**Importance of Interest Rate Forecasting**

According to Johannsen and Mertens, 2021, forecasting the rate of Interest can be considered important for not only individuals and businesses but also for investors. This forecast is likely to have a particularly relevant impact on the economic decisions as well as financial markets. The rate of interest can be predicted in an accurate manner by facilitating better planning on finances which can help in designing effective strategies on investment and risk management. Businesses will be able to make the right decision on having accurate knowledge on whether the rate of interest is likely to rise or fall. Right decisions can be made on borrowing, capital investments and managing debt with the businesses also getting the opportunity to make necessary adjustments to their strategies of pricing (Johannsen and Mertens, 2021). Companies having the obligation to pay back debt can take advantage of lower rates so that its interest expense can be reduced with an improvement in cash flow.

As per Medeiros et al. (2021), forecasting the rate of interest can be beneficial for investors as these rates are likely to influence the prices of assets like stocks, bonds, and real estate. The prices of bonds change inversely to the rate of interest which makes it important for investors to anticipate these changes so that they can manage risk. more stability in investment returns can be achieved by forecasting the fluctuations in returns. “*The responses of employment and inflation to interest rate decisions are arguably nonlinear on uncertainty*”(Medeiros *et al.,* 2021, pp.3). The monetary policy is shaped by the rate of interest which makes it highly essential to provide an accurate forecasting on the same so that economic stability can be managed. The central banks can figure out whether to raise or reduce the rate of interest for controlling inflation.

**Role of Exploratory data analysis (EDA) in Data Analysis**

According to Milo and Somech, 2020, it is important to understand the structure, various interconnections, and forms of patterns in a dataset. This is where Exploratory Data Analysis (EDA) becomes beneficial to understand the datasets in their entirety. “*Exploratory Data Analysis (EDA) is an essential process performed by data scientists in order to examine a new dataset up-close, better understand its nature and characteristics, and extract preliminary insights from it*”(Milo and Somech, 2020, pp. 1). Gaining fundamental information can be possible only if irregularities or missing values are identified. The techniques of summary statistics and visualisation of data is essential as well. These techniques can be obtained through EDA and can be applied for complicated models. Highlighting trends and providing information regarding the procedures of making certain decisions, is the primary focus of EDA. Such decisions are required to be taken for the analysis of subsequent data.

According to Patel et al. *(*2022), the primary objective of EDA is to provide a revelation of the distribution of data that can help with the identification of central tendencies like the mean and median. The relationships between variables that can be understood with the help of visualisations like histograms and scatter plots. This can also be beneficial for pointing out the presence of any pattern that is unusual (Patel *et al.,* 2022). Data can be cleaned and preprocessed with the help of EDA as it helps in identifying the values that are missing and the analysts can easily decide the way to manage them.

**Role of Support Vector Regression (SVR) model in Data Analysis**

According to Dash et val. (2023), Support vector machines (SVM) are commonly applied in machine learning classification issues, however, it can be also implemented with regression issues through support vector regression (SVR). The role of SVM in data analytics is related to predictive analysis where the data is continuous. Unlike the SVM where the primary motive is to classify data according to its characteristics, SVR focuses on numerical values that are not constant. . Kernels such as quadratic, radial basis function, and sigmoid are included in SVR models that can be easily used in data analytics. The efficiency of the SVR model in coordinating complex data that has non-linear relationships is relatively high as compared to other models pertaining to predictive performance. Time series forecasting is an integral part of data analysis that deals with financial institutions and sometimes m in weather prediction as well. Powerful tools such as SVR are proven to work effectively as they can handle data that are dimensional (Upreti *et al.*, 2022). The SVR regression model detects future data in variable value for example price of stock, interest increment and future rate of interest. As the usage of data analytics is widely observed, the use of SVR in data analysis is an important process that caters for accuracy and prevents error. Forecasting interest can be complex especially if the data pattern is large and diversified. Hence, The SVR model can also be applied to make the process simplified and provide data that are useful for financial institutes.

**Role of Multilayer Perceptron ANN model in Data Analysis**

According to Heidari et al. (2020), Multilayer perceptrons have notable capabilities as they consist of multiple layers for data analysis. The primary role of the data perceptron model is to classify data and it can also handle regression tasks like financial forecasting. Some of the advantages of ANN are its ability to approximate any continuous data, coordinate liners data and process data in volume. The model is extensively used by financial institutes in managing interest forecasting. However, forecasting is not limited to forecasting interest rates only, rather it is used in various parts of finance. Multilayer perceptrons are free neutral networking that helps in providing output from a set of inputs. The nodes are connected in a multilayer perceptron in layers and it uses backpropagation to enhance the result and networking system. This not only helps in predictive analytics but also increases the quality of the process involved in it. Perceptrons are commonly known for their algorithm but it should be noted that it originated to design image recognition. It is given the name aligning with its characteristic of perceiving, seeing, and identifying images. The model has mainly three kinds of layers hidden, output and input. The input layer processes the input that is given while the poutp[ut layer is responsible for performing predictive and classification data analysis. The application of multilayer perceptrons is one of the key factors as to why it is used in processing most data related to finance. In the finance sector, as the data are large and display dissimilar characteristics, the use of prediction through multilayer perceptrons is the best option.

**Analysis and comparison of the predictive performance of the SVR and MLP models using appropriate metrics focusing on their suitability for financial forecasting**

According to Palepu *et al.,* (2020), evaluation of past data can be crucial for predicting the financial performance of any organisation. Similarly, the SVR model transforms a given dataset into a hyperplane for an accurate representation of data to forecast the financial performance of the organisation precisely. A variety of financial metrics such as the flow of cash, revenue collected by the company, the sale of products, net profit margin and net working capital (NWC) have been used by the SVR model to project the financial performance of the organisation in the future (Adinyira *et al.,* 2021). These metrics are associated with the SVR model in determining a strategic financial plan to mitigate economic risks. Financial forecasting can be achieved through a detailed analysis of the former monetary information associated with the organisation. The cash flow of the organisation can be an essential determinant of its financial status in the future, through an estimation of the availability of cash. Likewise, the information on revenue before the current financial year can help the SVR model in estimating the growth in future revenue. An increase in revenue will automatically lead to the maximisation of profit, thereby positively impacting the overall financial development of the organisation. The prediction of sales can also be determined through the SVR model by estimating the amount of products sold in the previous financial year. NWC has been a key metric of the SVR model for calculating the financial capability of the organisation to carry out long-term expenses in the future.

According to Tashakkori *et al.,* (2024), the MLP model learns from algorithms to effectively regulate the biases resulting from errors associated with forecasted and actual outcomes. Likewise the SVR model, MLP also provides financial forecasting based on time series i.e., through the analysis of past algorithms. The evaluation of previously saved data can be useful for the organisation to make appropriate decisions to prevent any losses in the future. The stored data can help in understanding the patterns and values of certain products to promote effective loss mitigation strategies. Besides, the MLP model can incorporate strategies to mitigate financial distress during turbulent times. The proper use of the model can be vital to retrieve detailed financial information. Accordingly, the organisation can make effective investments for better returns. Debt to equity ratio, quick ratio, and net profit margin are some of the financial metrics used by the MLP model for predicting the finances of the organisation (Cece and Gençtürk, 2023). The debt-to-equity ratio is an essential metric used to measure the effectiveness of an investment through an assessment of previous inputs. It looks at the proportion of debt to the value of profit that the organisation can gain through certain investments. The quick ratio metric is a measure of financial resilience of an organisation during uncertain and chaotic times. This can help the MLP model in estimating the financial readiness of the organisation. The net profit margin is a key metric incorporated in the model for estimating the amount of profit that can be made by the organisation depending on the returns before the current financial year.

According to Dash *et al.,* (2023), both the SVR and MLP models are significant in providing accurate financial predictions for organisations, however, there are certain differences associated with its performance. The financial prediction offered by the MLP model has been more accurate in comparison to the SVR models. The former model can provide detailed forecasting despite the large dataset while the latter provides ideal results within a limited set of given data. The Root Mean Square Error (RMSE) i.e., the measure of differences between predicted and actual outcomes, is comparatively less in the MLP model than in the SVR model, thereby making it more accurate and efficient (Lee *et al.,* 2023). The MLP model is thus more difficult to understand and use due to its complex set-up, compared to the SVR model. However, such complexities can help in maximising its forecasting ability despite the large set of data.

## 2.3 Theories and Models

**Liquidity preference theory**

The ‘liquidity preference theory’ brings attention to the significance of money in the determination of the rate of interest. However, “*this theory has been developed in a framework of exogenous money*” (Missaglia *et al*., 2024, pp.1). As per the Keynesian liquidity preference theory, the rate of interest is the award that has been requested by the agents of the economy to give up their control over liquidity. However, there are distinct criteria demonstrated by Keynesians regarding holding money. Therefore, money holding has been distinguished by the three motives, which include ‘the transaction’, ‘speculation’, and the ‘precautionary motives’ (Saputra, 2020). The first and the last motives are based on a predictable function of income. The last motive, the speculative motive, is based on the perception of Keynes which indicates that it was ‘unpredictable’, along with putting a great emphasis on deciding market interest rates. It has been acknowledged that individuals have only two preferences for the medium where they can hold their funds, specifically money. It is regarded as liquid, however, earning no return (Marjohan *et al*., 2023). This theory also states that interest rates are regulated to balance the inclination to hold cash in contrast to the less liquid assets. The macroeconomic function of liquidity preference does not only look at the public’s inclination for cash, but it also looks at the liquidity preference of distinct financial organisations and banks. During unpredictable economic conditions, the need for liquidity has been generally raised. It may result in a lower rate of interest as the public and businesses prioritise holding cash instead of investing them.

**Liquidity premium theory**

The liquidity premium theory debates that an individual's preference is based on keeping assets in the form of a liquid. It includes cash instead of less liquid assets such as stocks, bonds, or real estate. “*Derived from the liquid premium theory, most of the investors choose short-term bonds rather than long-term bonds, and are willing to hold the latter at a premium, thus giving a positively-sloped yield curve*” (Kumar *et al*., 2021. pp. 3). The reason is long-term bonds are based on price risk along with these are relatively illiquid. However, the theory of market segmentation holds the concept that investors possess preferences concerning specific maturities more than others. Hence, investors who have chosen short-term bonds illustrated the formation of the interest rates concerning short-term bonds. However, those investors who have preferred long-term or intermediate bonds decide the yield curve concerning the long-term or intermediate bonds. The theory, while forecasting the rate of interest, contributes to affecting the yield in different parts of investing. In the forward rate of interest, this theory has been used to mark the differentiation between the additional holding return and the anticipated future spot rate. It has also been identified that the long-term bonds are more as compared to the expected and current future short-term rates. During periods of economic disturbances, investors have a tendency to grasp an elevated premium of liquidity as they intend to hold liquid assets or cash.

## 2.4 Literature Gap

The above literature review has addressed the concept of interest rate and their contribution to the global economy. Apart from the wider research it has been achieved that there is a gap in incorporating ML models for appropriate predictions of interest rates. Therefore, according to the research done by Missaglia and Botta (2024), gaps including integrating ML models to predict interest rates have put constraints on gaining insight into the changing nature of the economy and its relationships, restricting successful investment and policy-making strategies in the fast-changing landscape of finance. On the other hand, Lee *et al*., (2023) stated the relationship between finance and growth. However, this literature has not demonstrated the consensus on the financial growth nexus which is relevant to improve decision-making processes in the financial market. As per Milo, (2020), the advancement in ML has developed profound opportunities concerning EDA as well as completely automating processes. However, these articles have not mentioned the open questions to deal with the reduction of the manual effort needed for EDA which is significant to forecast the interest rate.

## 2.5 Summary

The enhanced market competition has shown financial distress in baking organisations. This study has used vital predictive models to culminate these uncertain risks within organisations. The utilisation of several predictive models has been done with the help of existing literature and journals. It has addressed the role of interest rates in the worldwide economy. It helps in lowering the rate of interest as well as stimulates more individuals to gain a mortgage for essential assets. Overall literature and different theories have been used to predict the decision-making process of the interest rates.

# CHAPTER 3: RESEARCH METHODOLOGY

## 3.1 Introduction

The methodology chapter outlines how the research has been conducted and tools were implemented to get the best results. Data collection, modelling, and evaluation will be discussed in the chapter to provide a better understanding. The tools that have been used in evaluating the dataset of lending clubs and how they have been used will be explained respectively.

## 3.2 Methodology

### 3.2.1. Data Selection

Data selection refers to the process of selecting suitable data and sources to collect data (Reel *et al.*, 2021). It also focuses on essential instruments that will work accurately with data to outsource results. Data selection is an integral part of analysing datasets as it helps in arranging and preparing the data in such a manner that will work effectively. In this research, data has been collected from secondary sources such as Kaggle. The dataset of the lending club has been taken into consideration for analysis. The selected data is related to Interest rates. Secondary data has been used in the research to process Python work. Secondary data has a number of advantages that are useful in research to meet its objective steadily. Firstly the scope of secondary data is wide as it is not possible to conduct primary data and get all the valuable information which is available easily. The use of secondary data can build a connection as it requires continuous interaction with studies and previous research. This continuous interaction helps in understanding the subject thoroughly. Hence, the effect of it is observed in the current research by providing better results with accuracy. Finally, the authenticity of secondary data cannot be questioned as it is already checked and interrogated. This helps in minimising the chance of any kind of error that may affect the present research. Since the data set of the lending club has been taken from Kaggle. It is important to mention some usefulness of Kaggle in the field of data analysis. kaggle datasets are easily available in various types of file formats and models that are provided to online researchers. Kaggle can cater data that are suitable for your data collection (Bojer and Meldgaard, 2021). For example, the image selection, data types and other filtration are present that provide accurate data according to the demand. kaggle is considered to be the one-stop solution for learning skills related to data analysis. Unlike other websites where Data are available at a certain cost, Kaggle provides quality data in large amounts for free. Likewise, the dataset of lending clubs has been taken from Kaggle which has been used in the research to predict interest rates. The data taken for the research purpose consists of interest forecasting of different types like interest, and credit cards.

### 3.2.2. Modelling

The SVR model has been used in data analysis. A support vector model is a kind of support vector machine that is used in the process of regression (Zheng *et al.*, 2021). The model discovers the function that can correctly predict the data included which is numerical or continuous. Essentially, it can use non liners and liners kernels which makes it a useful tool for the data analysis method. At the machine learning price, SVR models are extensively used as they have large memory and are easily workable in analysing unbalanced data. As the dataset of the lending club was unbalanced and had large memories, SVR models have been implemented to navigate the issue. Moreover, there was a significant variation in the data obtained from the lending club, hence SVR model helps in analysing the data smoothly without giving wrong results. Variation in data obtained from the high dimensional space in the data taken for research was effectively managed through the SVR model and it also provided a clear margin of separation.

Along with the SVR, the MLP (multilayer perceptron) model has been used in analysing the data. MLP can easily manage missing data and it is effective in analysing key trends in datasets that are particularly useful for professionals (Nosratabadi *et al.,* 2021). Different input means variation in numerical and categories that are difficult to process at the same time, however, MLP modelling that has been used in the dataset of the lending club has proven beneficial in analysing it. MLPs are extensively used in analysing complex data that is related to finance and interest rate forecasting. The dataset of lending clubs has “diversified income value”, “debt to income” and “ent length” which is difficult to analyse because of its complex nature. Thus, the implementation of the MLP model has significantly helped in getting the current interest value and making the research more effective.

### 3.2.3. Evaluation

The evaluation of both the model - SVR and MLP has been through mean square error (MSE) and mean absolute error(MAE). Analysis that is related to regression and machine learning uses MSE as a metric for analysing the predictive performance of the model. MSE is usually used in measuring the average difference in the square between the actual target and predictive model within a dataset (Hodson, 2022). The quality of the model is accessed through MSE and it helps understand how the data is aligned with the ground truth.

Likewise, MAE has been used for evaluation purposes of both models. MAE is a convenient statistic for regression models as it is simple to interpret and acts as a beneficial tool in predictive analysis (Qi *et al.,* 2020). The only difference between MSE and MAE is, that MAE is not influenced by extreme results. Hence it is suitable for datasets that have characteristics such as outliers or extreme values.

## 3.3 Tools

### 3.3.1 Python (platform)

Python is a programming language that is simple to understand and facilitates rapid results. It is used in web development, software development and data analysis. The data analytic part of the dataset of the lending club has been analysed through Python. One of the commonly used languages in data analysis, python has large libraries and is highly scalable which helps in analysing large and diversified data without any problems (McKinney, 2022). In the research, the structure of data has been created through Python. Moreover , python is considered to be one of the best tools when it comes to mathematical operations in data analysis.

### 3.3.2 Scikit Learn

In order to analyse large data it is important to have robust data libraries for machine learning. Thus, Scikit Learn has been applied in Python to analyse the data used in the research. The tool is considered to be effective and correctly result-oriented as it includes statistical modelling like classification, regression, clustering and dimensionality (Pölsterl, 2020). The libraries are built upon Numpy and Matplotlib. Scikit Learn works efficiently with other Python libraries to provide a seamless experience regarding data analysis. Moreover, it also helps in comparing data with one another to select the data that are closely related to the analysis, sometimes it also offers process data that help in the overall result of data that are selected. In this case, the dataset of the lending club has been selected and large libraries of scikit have been implemented in it.

### 3.3.3 Pandas

Pandas are an important part of Python that deal with data visualisation, manipulation and analysis. It is an open library that is used in many fields of data analysis like machine learning, programming and data science. Pandas has a significant feature of concluding the analysis based on statistical theories (Stepanek, 2020). As the dataset of the lending club was complex and messy, the implementation of pandas as a library has helped in cleaning it so that it is easy to read. Moreover, the relevance of data is incurred through pandas libraries which is important in data analysis.

### 3.3.4 Numpy

Proper training of machine learning models is essential in data analysis and it is done through Numpy. Numby is crucial as it is commonly used in forming matrices, performing mathematical operations and in the preprocessing of data (Harris *et al.,* 2020). In the research, the requirement for preprocessing of data was fulfilled by applying Numpy as a strong tool to inculcate the mathematical part of the lending data set.

### 3.3.5 Matplotlib

Matplotlib is a Python library that is used in creating data visualisation. Data visualisation is of various kinds such as plots, histograms, and bar charts (Lemenkova, 2020). Matplotlib can easily combine with other libraries of Python like pandas and numpy to facilitate better results. Likewise, it is considered to be a flexible option that can be customised in data analysis.

### 3.3.6 Seaborne

Seaborne libraries are tools that help in creating graphical representations and for analysing data that are exploratory (Sial *et al.,* 2020). In this case, the dataset of the lending club was exploratory, hence the tool was implemented. It should be noted that seaborne is a part of Python that does not require manipulation when compared to matplotlibData analysis often involves exploratory analysis of data in order to get better outcomes.

## 3.4 Summary

In this chapter, the methodology used in the research has been explained. The different tools used like Python and libraries like pandas, number, and seaborne have been examined to facilitate a clear concept of these instruments in data analysis. The modelling and its evaluation have been mentioned to understand the role of MSE and MAE in data analytics.

# Chapter 4: Implementation

## 4.1 Introduction

This chapter suitably outlines the techniques employed to execute advanced models of “Machine Learning” (ML) for the prediction of interest rates by using the Lending Club dataset. This chapter significantly includes the description of the dataset and it also details the procedures of exploratory data analysis in an integrated manner. This chapter also consists of the data preprocessing, transformation, partitioning and implementation of models for the evaluation of performance thereby highlighting the practical implementations of “machine learning” in financial prediction.

## 4.2 Description of Dataset

The dataset which has been used for the forecasting of interest rates the number of rows is 10001. The number of columns in this dataset is 60. This dataset suitably highlights diverse aspects which mainly capture a range of economic conditions which specifically makes it valuable for proactively understanding the dynamics of the lending markets. This dataset also includes a diverse range of categorical variables which includes “home ownership”, “annual income”, “verified income” and “debt to income” These are the critical factors in the dataset which notably impact the forecasting of interest rates. Homeownership suitably indicates financial responsibility and stability which frequently correlates with low factors of risk. This potentially leads to lower rates of interest for the borrowers (Roll *et al*., 2021). The lenders may oblige the homeowners because of their accepted creditworthiness. The annual income also notably serves as a primary predictor of the ability of a borrower to repay the loans. The high incomes commonly suggest low risk thereby allowing for more favourable rates of interest in a clear, concise and credible manner. The verified income plays a vital role in increasing the reliability of the income data. This significantly provides the lenders with dependence on the financial status of the borrower. This process of verification significantly helps in assessing the risk thereby setting applicable rates of interest. “Debt-to-income” ratios highly reflect the financial health of a borrower thereby contrasting obligations of monthly debt to gross monthly income.

## 4.3 Exploratory Data Analysis

Exploratory data analysis is an effective method of obtaining knowledge by exploring various datasets. The method systematically analyses data by utilising a “*series of analysis operations*” (Milo and Somech, 2020). The implementation of useful data analysis techniques makes EDA appropriate for the complete understanding of datasets. The method has enabled a satisfactory evaluation of the dataset that has been used to forecast interest rates in this research. The independent variables considered in the study are four important “features”, namely “annual income”, “debt-to-income”, “loan amount”, and “total credit lines”. On the other hand, the dependent variable is the “interest rate”, which has been the “target” of the study. This means that the interest rate of an individual is dependent on their annual income, debt-to-income ratio, loan amount, and credit lines.

Exploratory data analysis has several advantages. Firstly, this method uses a systematic approach to analyse datasets. This facilitates the evaluation and interpretation of large datasets. EDA involves the comparison of steps to ensure that the right steps are utilised for the data analysis (Abukmeil, *et al.* 2021). This allows the timely detection of flaws and improves the efficiency of the overall analysis.

The “annual income” serves as an important variable in the process of analysing the data using the EDA method. The maximum value of annual income analysed during the study is Rs. 4,18,000, whereas its minimum value is Rs. 3300. The median annual income is observed to be Rs. 60,000. The mean value of annual income is Rs. 68,034. On the other hand, the mean “debt-to-income” value during the analysis is found to be 30.879. The maximum “debt-to-income” is 489.09, whereas its minimum score is 1.94. The third quartile value of “debt-to-income” is 36.98. The minimum and maximum “earliest credit line” values are found to be 1978 and 2014 respectively.

## 4.4 Data Partitioning

Data partitioning is a critical step in data analysis specifically in “machine learning” as it highly involves splitting the dataset into testing and training subsets (Birba, 2020). In this scenario, the test and train data have been divided. For the forecasting of interest rates, the size of the train data is 80%. Further, the size of the test data is 20%. The role of data partitioning significantly lies in its increasing ability towards effectively improve scalability and optimise performance in an integrated manner. This also guarantees that the models are trained on one fragment of data and evaluated on another unconventional portion. This separation largely helps towards alleviating the overfitting in which a model performs notably well on training data. Partitioning also provides a mechanism towards dividing the data by the pattern of usage. The advantages of data partitioning lie in developing and improving performance by diminishing contention of server load and localisation of data. Scalability is yet another aspect in which partitioning plays a prominent role in terms of allowing the additional servers towards combining seamlessly. The test train divide plays a vital role in terms of improved “decision-making” about the necessary adjustments which are required for improvement.

## 4.5 Model Implementation

The Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models have been utilised for analysing data for the research. The SVR model has remarkable efficiency in different kinds of predictable analysis that involve complex data. The SVR is a “*machine learning method for classification and regression*” (Rahbar, *et al.* 2022, pp. 8). The primary aim of using the SVR model is to discover a target value with the most notable variation from “*the actual targets for all training samples*” (Huang, *et al.* 2022, pp. 8). SVR is the most effective model for analysis procedures involving non-linear data. The model is widely used for forecasting of target variables due to its efficiency and systematic approach (Xu *et al.* 2020). SVR has significantly facilitated this study by offering suitable analysis techniques. It uses various crucial metrics to develop a successful plan for an organisation’s financial growth. Thus, this model greatly helps in improving the financial performance of organisations. This explains the need for implementing SVR for accurate forecasting analysis and enhanced financial plans of organisations. Due to its efficiency and accuracy, the SVR model is applied in several vital forecasting departments, such as the meteorological department and in the health sector (Lo, *et al.* 2020). Another important benefit of implementing SVR is its ability to deal with complex datasets conveniently to predict results accurately. After several years of development, the SVR model has attained a level of remarkable perfection in performing complex forecasting tasks (Wei and He, 2023). This study has chosen this model to enhance the accuracy of forecasting “interest rates”. It has been highly beneficial to the purpose of the research.

On the other hand, the Multilayer Perceptron (MLP) model has been implemented due to MLP’s remarkable forecasting techniques. An efficient implementation of this model provides useful financial information (Orrù, *et al.* 2020). This model was introduced around 1950 but did not gain much appreciation back then. However, with continuous development and technological advancements, MLP has evolved into a highly successful forecasting tool (Safar *et al. 2023*). It has been noted that the financial prediction offered by the MLP model is more accurate in comparison with the SVR model. An important metric used in the MLP technique is the “debt-to-equity” ratio which analyses the suitability of a financial investment. The model is highly effective in dealing with complex patterns of data. MLP’s versatility is another reason for the extensive usage of the model. It can perform a wide range of functions with considerable accuracy (Hajirahimi and Khashei, 2020). The primary architecture of MLP comprises three layers, namely the inner layer, the hidden layer, and the external layer. In each of the three layers, the “*unique transfer functions*” are arranged systematically. On the other hand, SVR follows a basic linear format. A considerable number of studies have regarded SVR as an “*alternative model to MLP*” ((Sananmuang, *et al.* 2024).

Both SVR and MLP have proved beneficial in the forecasting of interest rates in this study. Both models utilise effective techniques to predict the outcome of the data analysis. Significant differences have been identified between the functioning of SVR and MLP. While SVR performs accurately even with large datasets, MLP functions efficiently within a specific limit of data. The difference between the predicted and actual outcomes is found to be lesser in the case of MLP, thus making it more accurate for the forecasting process. However, it has been detected that MLP is more complex and harder to grasp than SVR. Hence, it can be stated that SVR and MLP have their own strengths in forecasting target variables. An integration of the two models has enhanced the process of forecasting the “interest rates” in this research.

## 4.6 Summary

This chapter has provided a detailed description of the dataset that was analysed for the research. The primary economic conditions included in the dataset have been highlighted. The importance of “annual income” and “debt-to-income” ratio has been discussed. The chapter has also provided useful insights into the “exploratory data analysis” (EDA) method. The overall importance of EDA in this study for forecasting “interest rates” has been discussed. The significance of “data partitioning” has been analysed in this research. Finally, the chapter has explored the two models that have been implemented in this study, namely SVR and MLP.

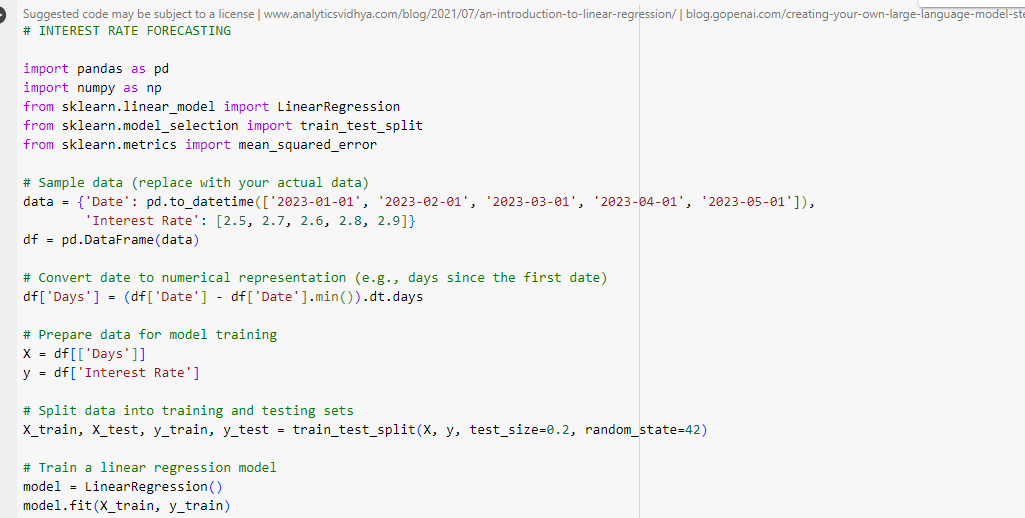
# 

# CHAPTER 5: RESULTS AND EVALUATION

## 5.1 Introduction

In this chapter, the result that has been generated by the analysis of the data of the lending club will be discussed and interpreted. The process of data implementation, data fetching and importing shall be analysed. An explanation of the result will provide a comprehensive understanding of the data that has been processed in Python.

## 5.2 Model Output

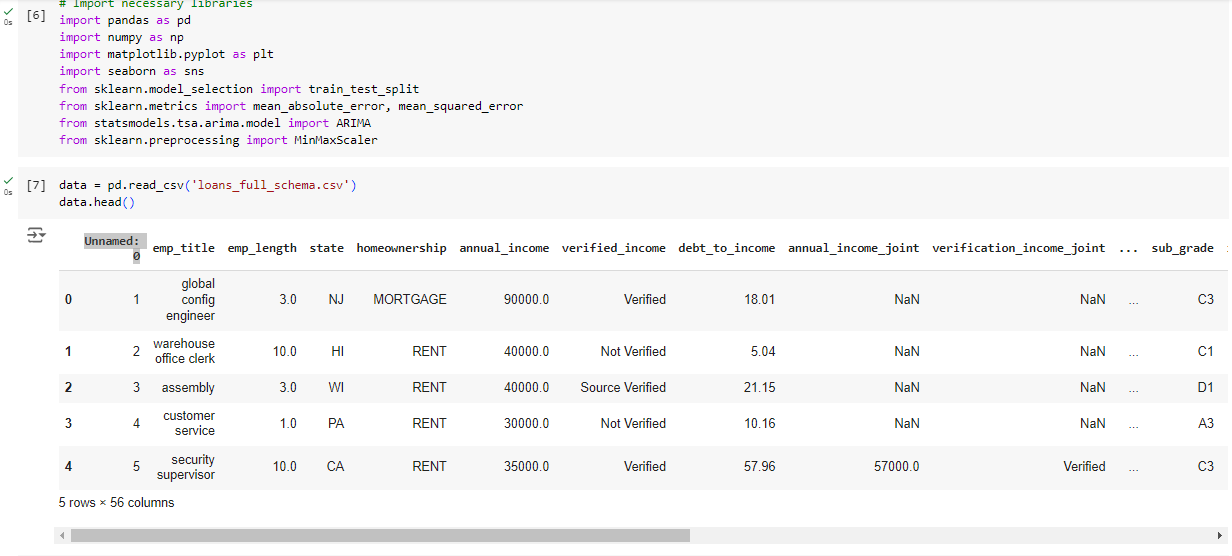




#### Figure : Data Implementations

(Source: Self Created)

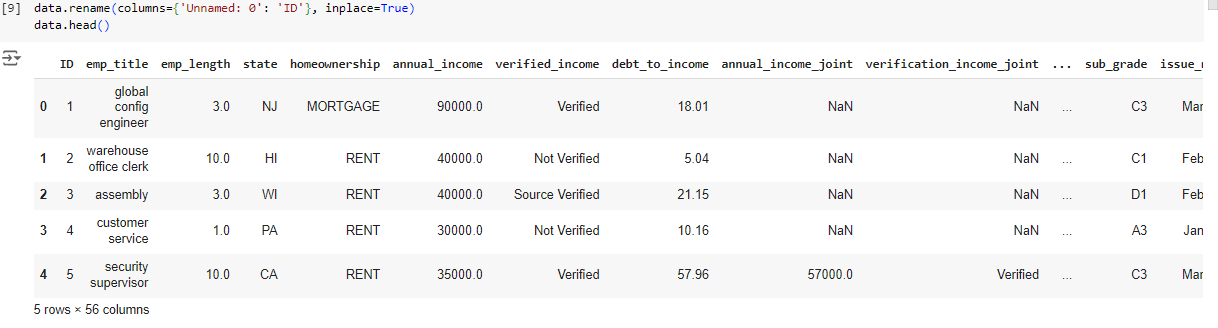
In the above figure, it can be observed that data implementation is being done in python. The interest forecasting pf lending club data set is being loaded through panda libraries. The large and complex data can easily be processed through pandas libraries. pandas in getting imported as pd from the “sk learn” as linear regression. Then the sample data is replaced by actuarial data in the second process (Kaluarachchi, 2022). The replacement of the data includes two important information such as data frame and data time. In the next process as shown in the figure, the data that is collected is converted to numerical values. The model training is then implemented followed by forecasting that can be seen in the figure. forecasting of the interest is bound to be accurate as it uses pandas library for analysis. Pandas libraries can easily work in proccering complex data that have numerica, and mathematical values in it.



#### Figure : Data Imports and process

(Source: Self Created)

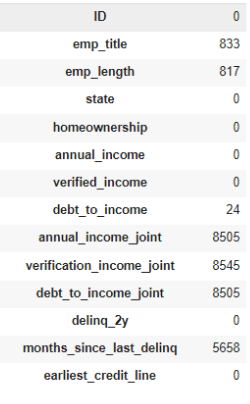
In the above figure data import and process has been shown that was done in python. pandas has been imported as pd and Mean squared error has been processed regarding different rent. The forecasting related to rent and mortgage has been done through python for getting an accurate value. There are five titles that are taken into consideration for analysis . The five titles are global configuring engineering, warehouse of office work, assembly, customer service and security supervision. The “empt\_ length” is 3.0 for global engineering and 10.0, 3.0, 1.0, and 10.0 respectively for the other 4 titles. The state is important to mention as it defines the data and its type. (Fathi *et al.,* 2022)The state are present in abbreviations like NU, HI, WI, PA and CA.

****

#### Figure : Data fetchsing

(Source: Self Created)

Data fetching is a process that deals with obtaining data from a centralised source in order to analyse it (Pierre *et al.*, 2021). It is selected in accordance with the suitability, for example data of interest are taken into consideration for analysing interest forecasting by the finance depart,net. There are 5 ids that have been taken into consideration. The forest ID is a global config engineer that has “empt\_ length ” of 3.0. The state is New Jersey and the home ownership is portage. The annual income of the id is 90000 and the income is verified as shown in the figure. the debt to income is 18.0 and the sub-grade is C. Like the other 4 id has been mentioned in the figure that show diversified characteristics.

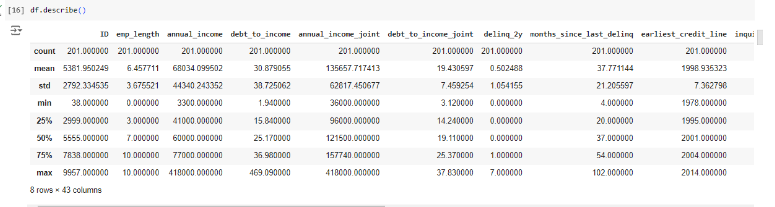




#### Figure : Implementing data shapes

(Source: Self Created)

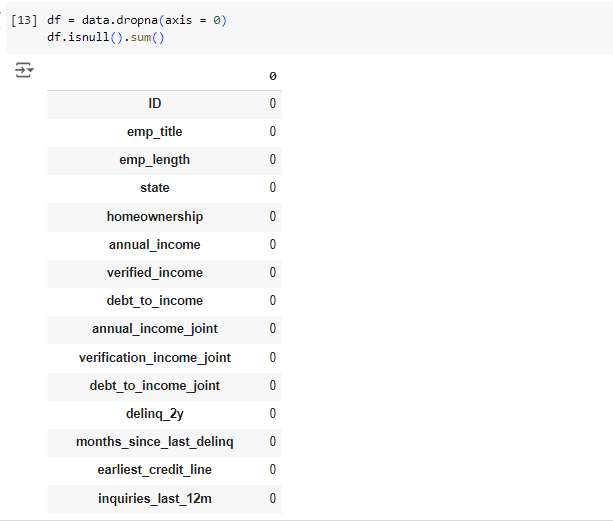
The above figure represents an interpretation of data shapes. The “emp title” has a score of 833, whereas the “emp length” value is 817. The “annual income joint” amounts to 8505. The “annual income”, “verified income”, “total credit limit” show a null value that has been dropped later during the research. The “debt-to-income” score is 24. The dataset shows 1271 “months since last credit inquiry”. The values of these variables have been implemented to forecast the “interest rates”. These are the “data shapes” on which the study has relied to develop suitable interpretations. The EDA method and the two effective models have facilitated the utilisation of the values in this dataset for forecasting the“interest rates”. The EDA method is highly effective in evaluating complex non-linear datasets (Bondu, *et al.* 2020), Hence, it has facilitated this data analysis process.



#### Figure : Data implementations results

(Source: Self Created)

The above figure displays the “data implementation results” of the study. The minimum value of “annual income” is 3300, whereas its maximum value is 418000. The mean “annual income” amounts to 68034. It is observed that the mean “date-to-income” score is 30.87. The first-quartile and third-quartile values of the “date-to-income” are 15.84 and 36.98 respectively. The std or “standard deviation” of “annual income” is 44,340. On the other hand, the std score of “debt-to-income” is 38.725. The minimum “debt-to-income joint” is 3.12, while its maximum value is 37.83. It is observed that the median “debt-to-income joint” is 19.11. The mean “annual income joint” is 135657.717. The maximum and minimum values of “emp length” are 10 and 0 respectively. The mean of “earliest credit line” is 1998.935. The values assigned to the variables in this dataset represent the outcomes of the data implementation process. These results have been crucial in guiding the research to forecast the “interest rates”. Prediction of interest rates is considerably significant for the economic stability of any organisation (Alaya, *et al.* 2021).



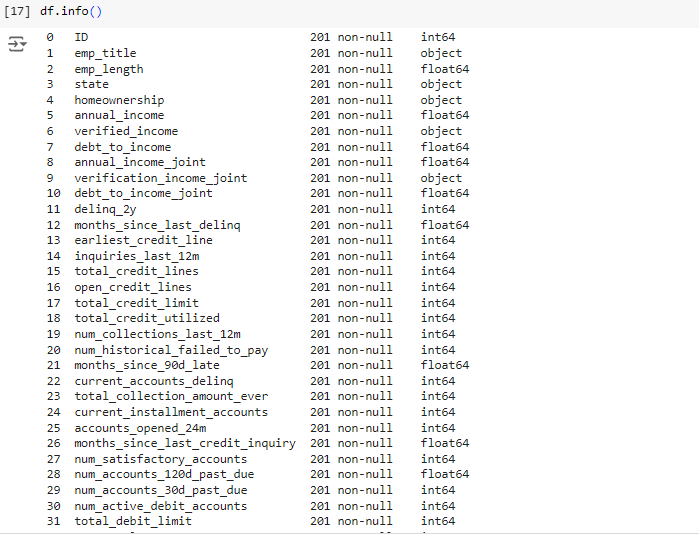


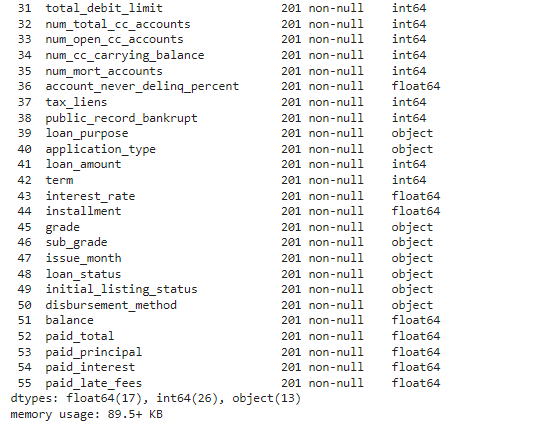


#### Figure : Data Dropdowns implementations

(Source: Self Created)

The above figure represents the data dropdown implementation where values of all the IDs are presented. Data dropdowns can show past, present, and future data sets that can be implemented in the present data to gain valuable information (Kim and Henke, 2021). it can be observed that the “emp \_ title ”, “emp\_ length” and state have given zero value. Zero value can also be stated as the null value in the Python programming system. The ownership income, annual income and verified income are in the null value. The data taken from the lending club data set has been provided with the above value. The loan purpose, application type and loan amount have been represented in the figure that was taken from data analysis done through Python. Moreover, the initial total, paid total and paid principle have been informed through the figure to get the best-regarding interest forecasting.

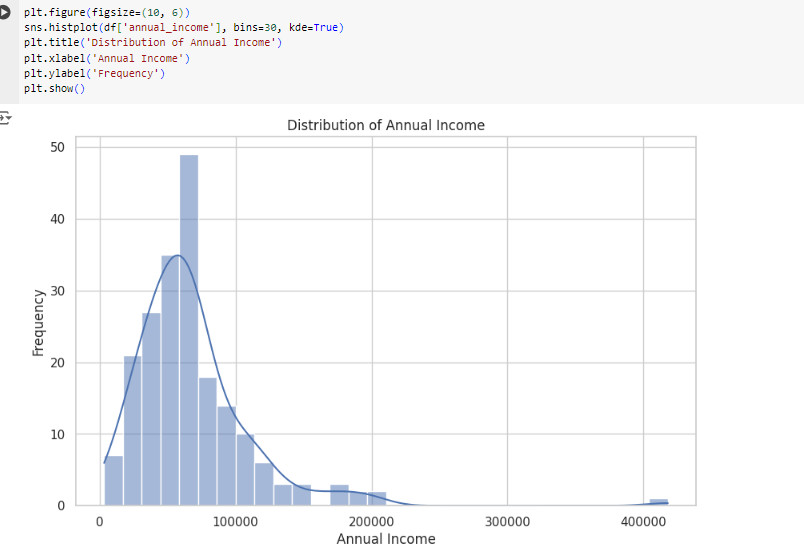




#### Figure : Data informations and generations

(Source: Self Created)

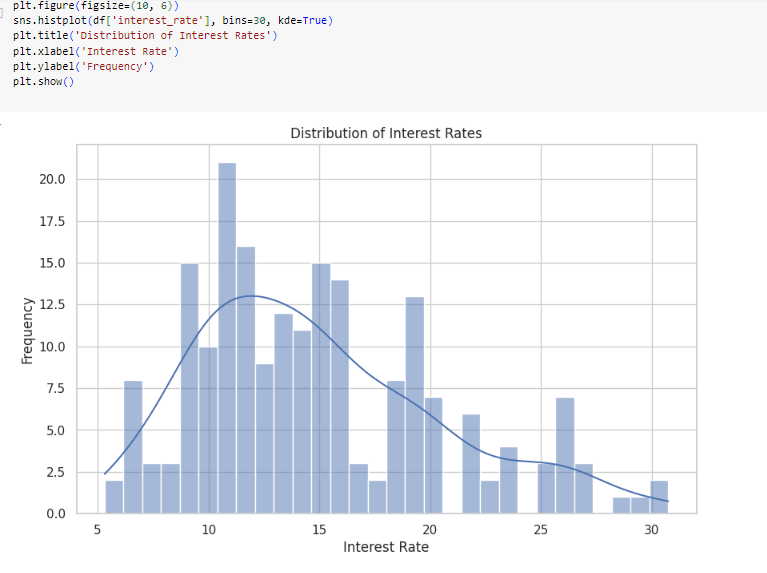
The above figure represents the data information and generation that has been generated from the dataset of the lending club. Data generation is the process of processing data with the help of software that uses algorithms. Data information is the process of grouping random data that do not have any specific information like name, origin and numerical values (Arendt *et al.*, 2020). Through Python, all the generation and information has been analysed with its outcomes. The emp title, emp length and state are 201 non-null, 201 non- and 201 non-null. Along with it, the loan purpose, Loan amount and paid late fees are identical to 201, 201 and 201 non-null. The total credit line, open credit line and distribution method is 201. In the figure there is a clear representation of num total cc accounts, num cc carrying balance and instalments as 201, 201, and 201 respectively.



#### Figure : Annual income statements

(Source: Self Created)

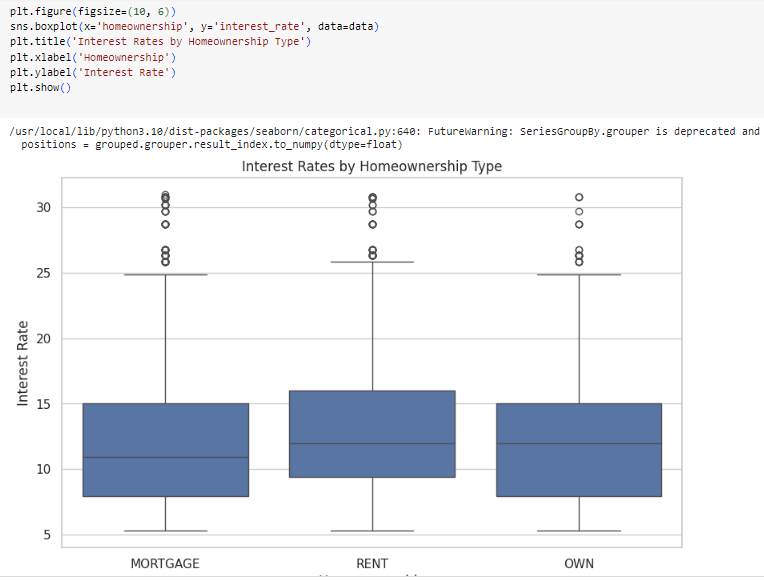
The above figure represents the income statement that has been generated from the dataset of the lending club. The income statement refers to the profitability and other financial numerical that the company or a group has encountered over some time (Smith *et al.,* 2021). It can be about a mid-sized company, small company or large company, however, it should be noted that the income statement is not only about the company. Sometimes it can refer to data that is collected by analysing a group of people or a large section of society. In the X axis, the annual income is represented as 0, 100000, 200000, 300000 and 400000. In the Y axis, the frequency has been presented such as 0, 10, 20, 30, 40, and 50. It can be observed that the highest annual income is of frequency 35 and the amount for it is 100000.



#### Figure: Distributions Rates

(Source: Self Created)

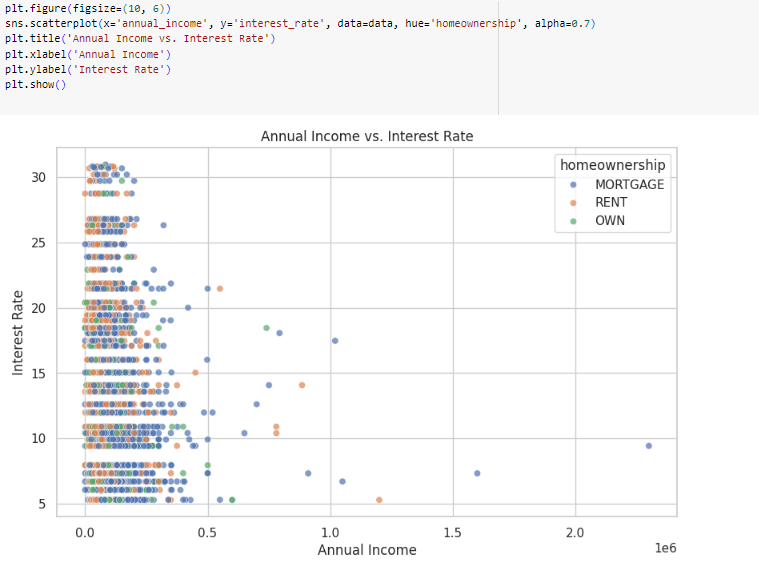
The above figure represents the distribution rate of the interest in data that has been analysed through Python. The distribution of interest rate refers to the amount of interest that has been provided to one member or group of members (Benidis *et al.,* 2020). It also refers to the return on investment that can be generated by investing in a particular thing. In the X axis, the interest has been represented as 5, 10, 15, 20, 25 and 30. Similarly in the Y-Axis, the frequency has been mentioned such as 0.0, 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, 17.5 and 20.0 the bar chart represents that the distribution rate was minimum in the initial stage while a rise is seen of 10 and then the graph has gone downward from 15 to 30



#### Figure : Interest rate calculations

(Source: Self Created)

The above figure represents the interest rate calculation that has been applied in the data set of lending clubs. Interest rate calculation in interest forecasting is an integral part as it provides information that can aid accuracy in predictive performance. The interest rate helps in understanding the position and power of the borrower and lender in the marketplace (Mehrotra and Sergeyev, 2021). In the figure the interest rate of 3 subjects has been taken into consideration- mortgage, rent and own. The x-axis represents the subject while the y-axis represents the frequency of interest rates. The rate of the mortgage as mentioned in the dataset is 15 and for rent, the rate is 15.1. Finally, the rate for ownership is relatively low at 14.8. The interest rate has been calculated by taking home ownership as the main aspect.

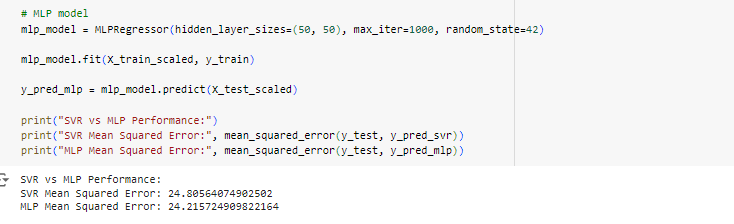


#### Figure : Annual income and interest rate comparisons

(Source: Self Created)

In the above figure, the comparison of interest rate and annual income has been represented by python the annual;l income is the total amount generated in one year from various sources such as salaries, business, incentives, etc (Bauer and Rudebusch, 2020). on the other hand is considered to be the cost of borrowing money or the reward for saving money over time. In the above figure, the three main subjects of ownership have been taken into account. The subjects are mortgage, rent and ownership. The x-axis represents the annual income while comparing it with the interest rate shown in the y-axis. The interest rate is taken as 5, 10, 15, 20, 25 and 30 while the annual income is taken as 0.0, 0.5, 1.0, 1.5, 2.0. the annual interest rate and annual income of the data set obtained from the lending club are more or less similar, there is not a huge difference between them.



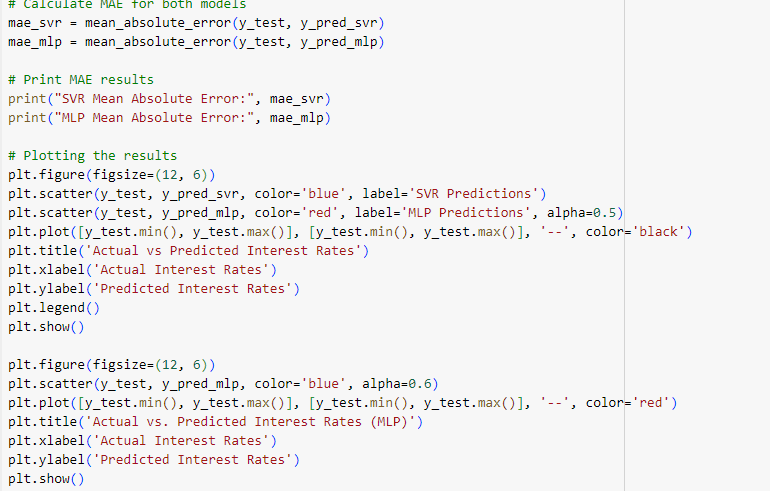


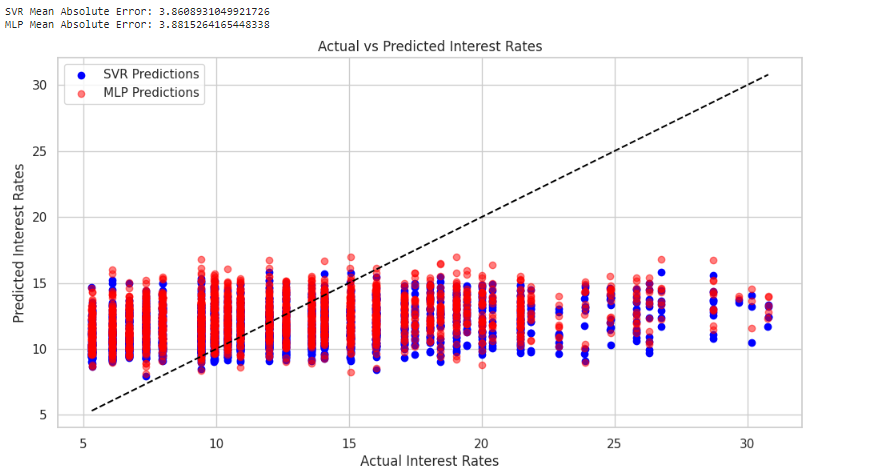
#### Figure : SVR vs MLP Performance

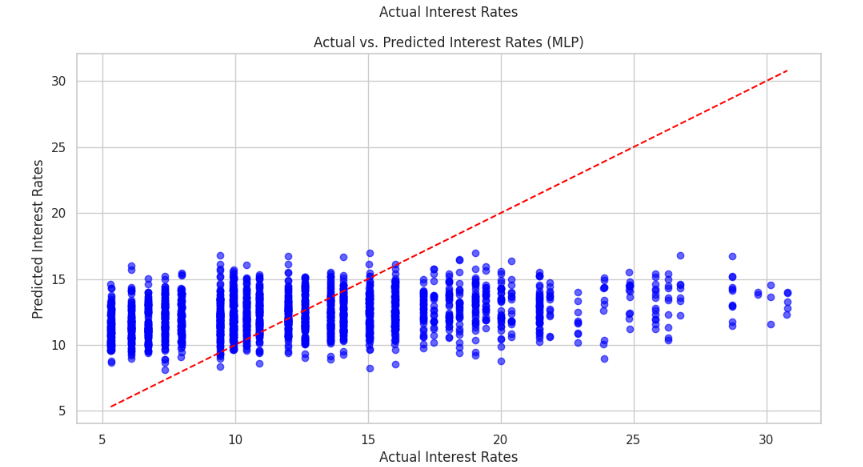
(Source: Self Created)

The SVR and ML modelling has been implemented in analysing the data set of lending clubs.

Selection of the figure and targeting of the data split has been performed in the process as shown in the figure. Svr is a modelling type of machine learning algorithm that is used for regression analysis whereas MLP modelling is supervised algorithm that is used in dimensional data to analyse and facilitate accurate data about the information inputted (Oukhouya and El Himdi, 2023).



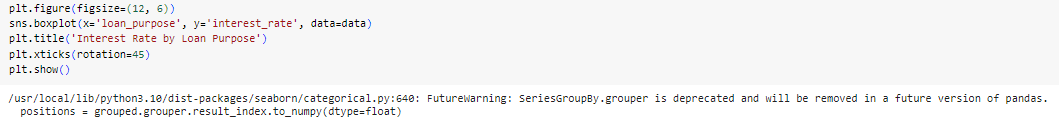


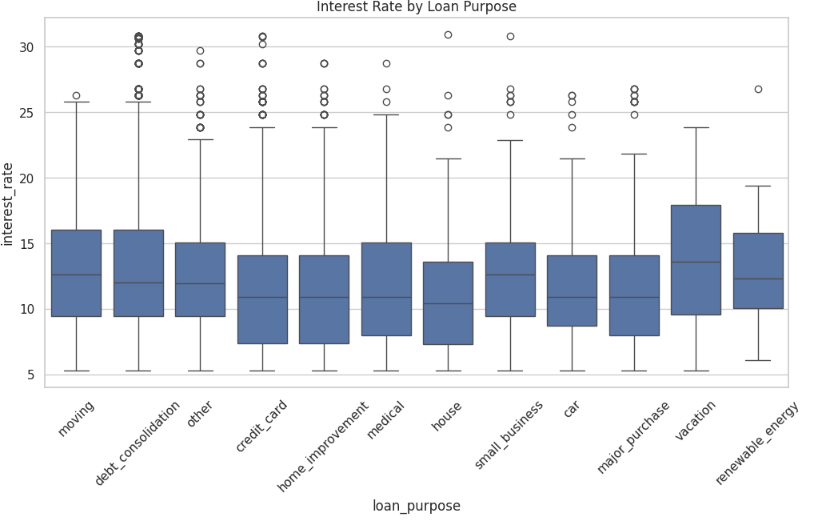


#### Figure : calculate MAE

(Source: Self Created)

The above represents the calculation of the MAE of the data set related to the lending club that has been taken from Kaggle. This means the absolute error is referred to as a metric that is used in calculating the average difference between the predicted and actual values of a set of observations (Baade *et al.*, 2020). The SVR means absolute error is 3.860 and MLP means absolute error is 3.881. Two predictions have been performed in the observation such as SVR prediction and MLP prediction that provide a comprehensive data integration regarding the data that has been selected for analysis. The use of mae in the SVR mol helps in evaluating the subject more explicitly. However, there is a slight difference between the evolution of the data with MLP and SVR. The data that are evaluated with SVR are more accurate as SVR modelling can easily simplify complex data. However, in the case of MLP modelling, dimensional data can easily be analysed in it.

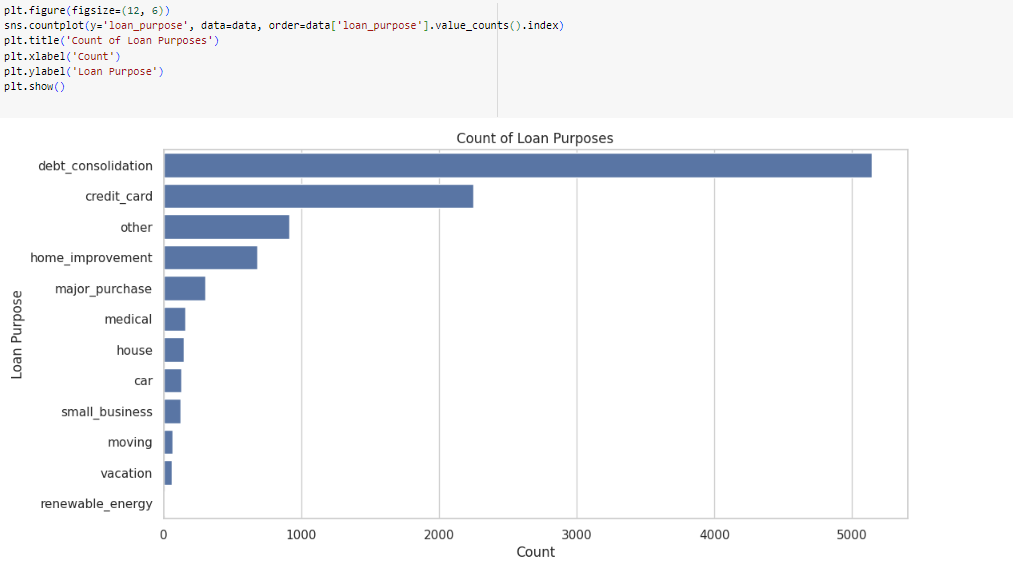




#### Figure : Interest and loan infrastructure

(Source: Self Created)

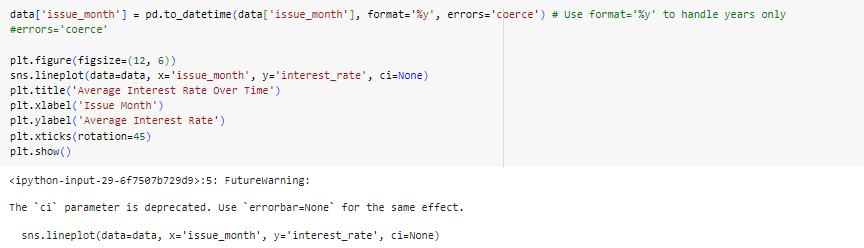
The above figure represents the purpose of the loan that has been taken and the interest rate that is charged. In the x-axis, some subjects are considered to be the loan purpose such as moving, debt consolidation and credit card (Anguelov, 2021). Likewise, in the y-axis, the interest rate has been mentioned that is charged against different services. The interest rate for renewable energy is 17, for cars the rate is 14 while for houses the rate is 13. On the other hand, the interest for home improvement is 14, for credit cards the rate is 14.5 and for medical it is 15.

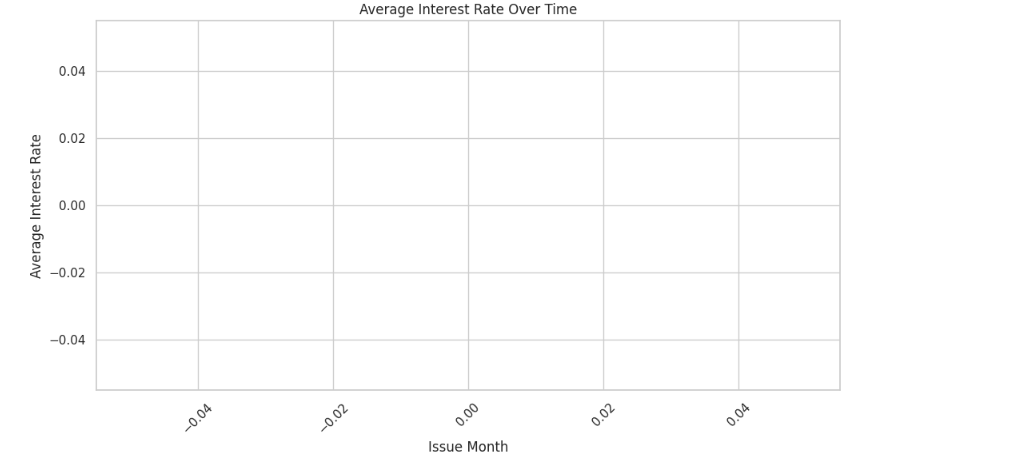


#### Figure : loans generations

(Source: Self Created)

The above figure demonstrates the loan generations that have been conducted through SVR modelling in Python. The number of loans, their value and their purpose are the first things that were used as input in Python for analysing data. The dataset of lending clubs has been used in outsourcing the loan generation of different variables. The x-axis in the figure represents the loan count while in the y-axis, the purpose of the loan is mentioned. It has been observed that the highest loan count is for debt consolidation. The count for loans for debt consolidation is 5000 and the count for credit cards is approximately 2500. Likewise, the count of the loans for home improvement and major purposes is approximately 1000. The loan count for medical, house and car is the lowest at 0 to 500 approximately.

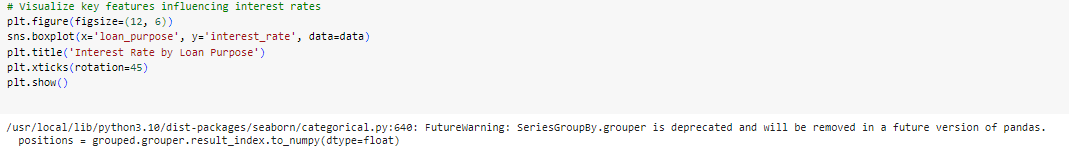


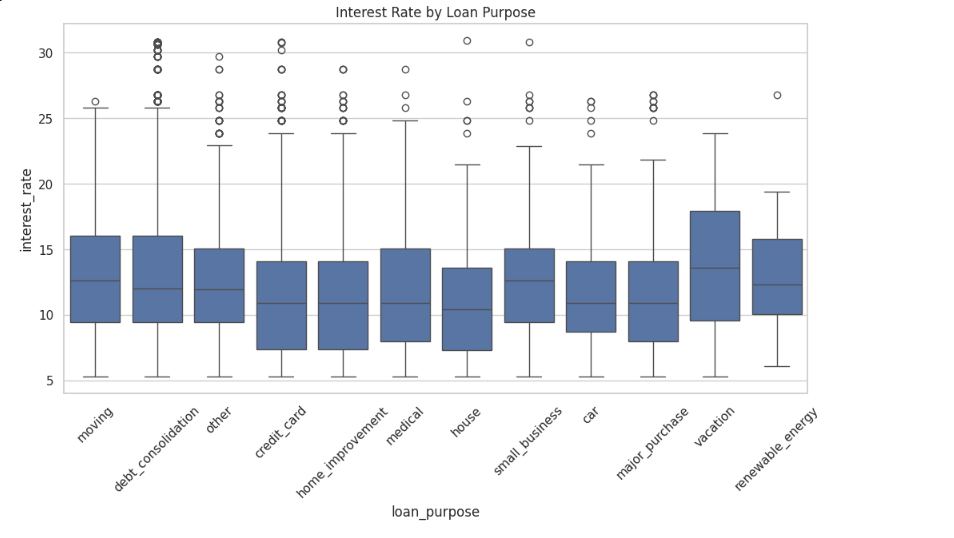


#### Figure : Averages interest rate over time

(Source: Self Created)

The above figure represents the average interest rate over time that has been observed in the dataset of lending clubs. In the field of interest forecasting the average interest rate is of great importance as it helps in identifying the current pattern and trend (Vayanos and Vila, 2021). The x-axis represents the issue month while the y-axis represents the average interest rate. The issue month is as follows - o.o4, o.o3, 0.00, 0.02 and 0.04 while the average interest rates are 0.04, 0.02, 0.00, 0.02, and 0.04 respectively. Generically the average interest rate related to interest forecasting is between 5 to 10 %. However, different data can possess varied information as is seen in the current dataset that has been taken from analysis.



****

#### Figure : Visualise key features influencing interest rates

(Source: Self Created)

In the above figure, the key features that are influencing interest rates are represented. The interest is influenced by many factors including demand, inflation, purpose and the current protocol of the market. However, in the dataset of ending clubs that have been analysed in this research, the main influencing factors are credit cards, small business, vacation, home improvement and car loans.

## 5.3 Summary

The methodology chapter outlines the method that has been used in performing the Python work. The Python work has been performed in the dataset of the lending club that has been obtained from Kaggle. The SVR modelling and MLP modelling have been implemented in the data set to outsource valuable information like annual income, interest rate and comparison. The chapter deals with all the figures that have been included in the research like loan infrastructure, calculation of MAE, SVR and MLP modelling performance and other data analysis for the research. Finally, the overall approach that has been taken in the entire research is a secondary analysis of the data.

# CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

## 6.1 Introduction

In the conclusion chapter the overall outcomes of the research and what has been performed in the research will be discussed. Discussion of the research topic and how it has met the objectives while fulfilling every guideline will be explained in the chapter. Along with its contribution, this research has been undertaken and future work will be examined.

## 6.2 Discussion and Conclusion

Interest forecasting is an important exercise done by all financial institutes and money lending companies. It helps in understanding the data related to interest and contributes to the overall performance of the business. In this research, the role of interest forecasting has been displayed by taking the dataset of the lending club and analysing it through Python. In the Python data analysis process, several models have been used like SVR and MLP. The SVR model helps in regression analysis while the MLP model processes data that are nonlinear and dimensional . The research explained a rich literature that is present on the subject and showcased the view of different authors on interest forecasting through machine learning. The dataset that has been obtained from Kaggle also denotes the importance of Kaggle as an independent source for open source data that can be analysed by researchers and students. All the aims and objectives of the research have been met by examining various aspects of machine learning and its roles in the finance department. It should be mentioned that machine learning has not only changed the circumstances on a small level but rather it has influenced the global market as well. In recent times where there are tons of transactions taking place, it's become hard to predict and track every data manually . Hence, machine learning is implemented in the form of Python or any other language to bridge the gap of untraceability and misunderstanding. The research has extensively outlined the importance of interest forecasting and has an implementation of Python in it. The result chapter explicitly explained different kinds of data that are outsourced through Python. Data implementation, data fetching and data import have been undertaken in order to properly analyse different perspectives such as average loan, loan type and subject on which loans are taken. The subject can be of various types like car loans, home improvement loans, and debt consolidation. It has been found as demonstrated in the result chapter that the most amount of loans are taken for debt consolidation. Debt consolidation is the main factor that is playing a significant role in changing the dynamic of interest rates in the global market. Laon generation and visualisation of average loans taken from a period has been done which signifies that the loans are generated immensely from the financial agents. The dataset has been taken into consideration for creating a critical study about interest forecasting and how it is done in the marketplace. Finally it has been observed that the interest rate in recent years has increased as previously it was between 0 to 5 % and at present it is 5 to 15 %. In a more specific interest, cars are now at 8 to 9 % whereas previously it was 2 to 6%. Likewise, interest on home is 15% while before it was 4 to 8%. The interest on debt consolidation is 14% while before it was 8% respectively.

## 6.3 Recommendation

Based on the study it is found that the general trends of the interest rate is increasing throughout the years. This also shows that the operational risks of the financial institutions may tend to increase considerably. Analysing and predicting the rate of interest of lending money is important for the management to make useful decisions. The lending organisations can easily implement different models of python to enhance the predictive analysis to a great event. In the end, the inclusion of machine learning training as a profession for handling interest forecasting can make a huge difference in the coming year pertaining to financial work.

## 6.4. Future Work

Due to a lack of primary data, the research was not as accurate as seen in the case of primary research. It is important to mention that primary data is processed personally, hence, credibility is more important. Moreover, in the research , two models have been used which are the SVR and MLP. However, the inclusion of models like decision trees and logistic regression could have impacted positively in understating the subject clearly. The decision tree and logistics regression would provide better results in analysing the data of the lending club.

# 

# 

# 

# References

Abukmeil, M., Ferrari, S., Genovese, A., Piuri, V. and Scotti, F., 2021. A survey of unsupervised generative models for exploratory data analysis and representation learning. *Acm computing surveys (csur)*, *54*(5), pp.1-40.

Adinyira, E., Adjei, E.A.G., Agyekum, K. and Fugar, F.D.K., 2021. Application of machine learning in predicting construction project profit in Ghana using Support Vector Regression Algorithm (SVRA). Engineering, Construction and Architectural Management, 28(5), pp.1491-1514.

Alaya, M.B., Kebaier, A. and Sarr, D., 2021. Deep calibration of interest rates model. *arXiv preprint arXiv:2110.15133*.

Anguelov, D., 2021. Banking ‘development’: the geopolitical–economy of infrastructure financing. Area Development and Policy, 6(3), pp.271-295.

Arendt, J.F.H., Hansen, A.T., Ladefoged, S.A., Sørensen, H.T., Pedersen, L. and Adelborg, K., 2020. Existing data sources in clinical epidemiology: laboratory information system databases in Denmark. Clinical epidemiology, pp.469-475.

Baade, A., Peng, P. and Harwath, D., 2022. Mae-ast: Masked autoencoding audio spectrogram transformer. arXiv preprint arXiv:2203.16691.

Bauer, M.D. and Rudebusch, G.D., 2020. Interest rates under falling stars. American Economic Review, 110(5), pp.1316-1354.

Benidis, K., Rangapuram, S.S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L. and Aubet, F.X., 2022. Deep learning for time series forecasting: Tutorial and literature survey. ACM Computing Surveys, 55(6), pp.1-36.

Birba, D.E., 2020. A Comparative study of data splitting algorithms for machine learning model selection.

Bojer, C.S. and Meldgaard, J.P., 2021. Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting, 37(2), pp.587-603.

Bondu, R., Cloutier, V., Rosa, E. and Roy, M., 2020. An exploratory data analysis approach for assessing the sources and distribution of naturally occurring contaminants (F, Ba, Mn, As) in groundwater from southern Quebec (Canada). *Applied Geochemistry*, *114*, p.104500.

Cece, O. and Gençtürk, M., 2023. Net Profit Margin Forecasting with Machine Learning Methods in Hospital Finance Management. Journal of Health Systems and Policies, 5(2), pp.103-119.

Chang, A.H., Yang, L.K., Tsaih, R.H. and Lin, S.K., 2022. Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. Quantitative Finance and Economics, 6(2), pp.303-325.

Croux, C., Jagtiani, J., Korivi, T. and Vulanovic, M., 2020. Important factors determining Fintech loan default: Evidence from a lendingclub consumer platform. Journal of Economic Behavior & Organization, 173, pp.270-296.

Dash, R.K., Nguyen, T.N., Cengiz, K. and Sharma, A., 2023. Fine-tuned support vector regression model for stock predictions. Neural Computing and Applications, 35(32), pp.23295-23309.

Dash, R.K., Nguyen, T.N., Cengiz, K. and Sharma, A., 2023. Fine-tuned support vector regression model for stock predictions. Neural Computing and Applications, 35(32), pp.23295-23309.

Fathi, M., Haghi Kashani, M., Jameii, S.M. and Mahdipour, E., 2022. Big data analytics in weather forecasting: A systematic review. Archives of Computational Methods in Engineering, 29(2), pp.1247-1275.

FINANCIAL SUMMARY FY2024. [Online]Available on:<<https://global.toyota/pages/global_toyota/ir/financial-results/2024_4q_summary_en.pdf>>

Hajirahimi, Z. and Khashei, M., 2020. Weighted MLP-ARIMA series hybrid model for time series forecasting. *Journal of Industrial Engineering and Management Studies*, *7*(2), pp.187-201.

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J. and Kern, R., 2020. Array programming with NumPy. Nature, 585(7825), pp.357-362.

Heidari, A.A., Faris, H., Mirjalili, S., Aljarah, I. and Mafarja, M., 2020. Ant lion optimizer: theory, literature review, and application in multi-layer perceptron neural networks. Nature-inspired optimizers: theories, literature reviews and applications, pp.23-46.

Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, 2022, pp.1-10.

Hofmann, B., Lombardi, M.J., Mojon, B. and Orphanides, A., 2021. Fiscal and monetary policy interactions in a low interest rate world.

Huang, J., Algahtani, M. and Kaewunruen, S., 2022. Energy forecasting in a public building: A benchmarking analysis on long short-term memory (LSTM), support vector regression (SVR), and extreme gradient boosting (XGBoost) networks. *Applied Sciences*, *12*(19), p.9788.

Johannsen, B.K. and Mertens, E., 2021. A Time‐Series Model of Interest Rates with the Effective Lower Bound. Journal of Money, Credit and Banking, 53(5), pp.1005-1046.

Kaluarachchi, Y., 2022. Implementing data-driven smart city applications for future cities. Smart Cities, 5(2), pp.455-474.

Kim, B. and Henke, G., 2021. Easy-to-use cloud computing for teaching data science. Journal of Statistics and Data Science Education, 29(sup1), pp.S103-S111.

Kumar, R.R., Stauvermann, P.J. and Vu, H.T.T., 2021. The relationship between yield curve and economic activity: An analysis of g7 countries. *Journal of Risk and Financial Management*, *14*(2), p.62.

Lee, H., Kim, D. and Gu, J.H., 2023. Prediction of food factory energy consumption using MLP and SVR algorithms. Energies, 16(3), p.1550.

Lee, K.S. and Werner, R.A., 2023. Are lower interest rates really associated with higher growth? New empirical evidence on the interest rate thesis from 19 countries. International Journal of Finance & Economics, 28(4), pp.3960-3975.

Lemenkova, P., 2020. Python libraries matplotlib, seaborn and pandas for visualization geo-spatial datasets generated by QGIS. Analele stiintifice ale Universitatii" Alexandru Ioan Cuza" din Iasi-seria Geografie, 64(1), pp.13-32.

Lo, W.L., Zhu, M. and Fu, H., 2020. Meteorology visibility estimation by using multi-support vector regression method. *Journal of Advances in Information Technology Vol*, *11*(2), pp.40-47.

Marjohan, M., Anggraini, A. and Dewi, S.K.S., 2023. Opportunity Set, Liquidity, Stock Return, Inflation As A Moderator Investment Risk, Investment. *Jurnal Manajemen*, *27*(2), pp.379-400.

McKinney, W., 2022. Python for data analysis. " O'Reilly Media, Inc.".

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á. and Zilberman, E., 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. Journal of Business & Economic Statistics, 39(1), pp.98-119.

Mehrotra, N.R. and Sergeyev, D., 2021. Debt sustainability in a low interest rate world. Journal of Monetary Economics, 124, pp.S1-S18.

Milo, T. and Somech, A., 2020, June. Automating exploratory data analysis via machine learning: An overview. In Proceedings of the 2020 ACM SIGMOD international conference on management of data (pp. 2617-2622).

Milo, T. and Somech, A., 2020, June. Automating exploratory data analysis via machine learning: An overview. In *Proceedings of the 2020 ACM SIGMOD international conference on management of data* (pp. 2617-2622).

Missaglia, M. and Botta, A., 2024. Households’ liquidity preference, banks’ capitalization and the macroeconomy: a theoretical investigation. *Review of Political Economy*, *36*(3), pp.1192-1215.

Naeem, S., Mashwani, W.K., Ali, A., Uddin, M.I., Mahmoud, M., Jamal, F. and Chesneau, C., 2021. Machine learning-based USD/PKR exchange rate forecasting using sentiment analysis of Twitter data. Computers, Materials & Continua, 67(3), pp.3451-3461.

Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C. and Mosavi, A., 2021. Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. Agriculture, 11(5), p.408.

Orrù, P.F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R. and Arena, S., 2020. Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability*, *12*(11), p.4776.

Oukhouya, H. and El Himdi, K., 2023, April. Comparing machine learning methods—svr, xgboost, lstm, and mlp—for forecasting the moroccan stock market. In Computer Sciences & Mathematics Forum (Vol. 7, No. 1, p. 39). MDPI.

Palepu, K.G., Healy, P.M., Wright, S., Bradbury, M. and Coulton, J., 2020. Business analysis and valuation: Using financial statements. Cengage AU.

Patel, H., Guttula, S., Mittal, R.S., Manwani, N., Berti-Equille, L. and Manatkar, A., 2022, August. Advances in exploratory data analysis, visualisation and quality for data centric AI systems. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4814-4815).

Pierre, J., Crooks, R., Currie, M., Paris, B. and Pasquetto, I., 2021, May. Getting Ourselves Together: Data-centered participatory design research & epistemic burden. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-11).

Pölsterl, S., 2020. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. Journal of Machine Learning Research, 21(212), pp.1-6.

Qi, J., Du, J., Siniscalchi, S.M., Ma, X. and Lee, C.H., 2020. On mean absolute error for deep neural network based vector-to-vector regression. IEEE Signal Processing Letters, 27, pp.1485-1489.

Rahbar, A., Mirarabi, A., Nakhaei, M., Talkhabi, M. and Jamali, M., 2022. A comparative analysis of data-driven models (SVR, ANFIS, and ANNs) for daily karst spring discharge prediction. *Water resources management*, *36*(2), pp.589-609.

Reel, P.S., Reel, S., Pearson, E., Trucco, E. and Jefferson, E., 2021. Using machine learning approaches for multi-omics data analysis: A review. Biotechnology advances, 49, p.107739.

Roll, S., Kondratjeva, O., Bufe, S., Grinstein-Weiss, M. and Skees, S., 2021. Assessing the short-term stability of financial well-being in low-and moderate-income households. Journal of Family and Economic Issues, pp.1-28.

Safar, A.A., Salih, D.M. and Murshid, A.M., 2023. Pattern recognition using the multi-layer perceptron (MLP) for medical disease: A survey. *International Journal of Nonlinear Analysis and Applications*, *14*(1), pp.1989-1998.

Sananmuang, T., Mankong, K. and Chokeshaiusaha, K., 2024. Multilayer perceptron and support vector regression models for feline parturition date prediction. *Heliyon*, *10*(6).

Saputra, A.J. and Harris, I., 2020. Analysis of The Development Cashless Transaction on The Need for Money Paper Based. In *Conference Series* (Vol. 3, No. 1, pp. 29-40).

Sial, A.H., Rashdi, S.Y.S. and Khan, A.H., 2021. Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. International Journal, 10(1), pp.277-281

Smith, S.A., Boyer, C.N. and Griffith, A.P., 2020. An Introduction to Basic Farm Financial Statements: Income Statement. The University of Tennessee.

Statista, 2024. [Online] Availbe On : <<https://www.statista.com/statistics/1317878/inflation-rate-interest-rate-by-country/>>

Stepanek, H., 2020. Thinking in Pandas. Berkeley, CA, USA: Apress.

Tashakkori, A., Talebzadeh, M., Salboukh, F. and Deshmukh, L., 2024. Forecasting Gold Prices with MLP Neural Networks: A Machine Learning Approach. International Journal of Science and Engineering Applications (IJSEA), 13, pp.13-20.

Upreti, K., Singh, U.K., Jain, R., Kaur, K. and Sharma, A.K., 2022. Fuzzy logic based support vector regression (SVR) model for software cost estimation using machine learning. In ICT Systems and Sustainability: Proceedings of ICT4SD 2021, Volume 1 (pp. 917-927). Springer Singapore.

Vayanos, D. and Vila, J.L., 2021. A preferred‐habitat model of the term structure of interest rates. Econometrica, 89(1), pp.77-112.

Wei, J. and He, X., 2023. Support vector regression model with variant tolerance. *Measurement and Control*, *56*(9-10), pp.1705-1719.

Xu, D., Zhang, Q., Ding, Y. and Huang, H., 2020. Application of a hybrid ARIMA–SVR model based on the SPI for the forecast of drought—a case study in Henan Province, China. *Journal of applied meteorology and climatology*, *59*(7), pp.1239-1259.

Zheng, J., Wang, Y., Li, S. and Chen, H., 2021. The stock index prediction based on SVR model with bat optimization algorithm. Algorithms, 14(10), p.299.