A PROJECT REPORT ON

LEAD SCORING ANALYSIS FOR ONLINE COURSES USING SCI-KIT LEARN

Submitted to the partial fulfillment of the requirement for
the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

SUBMITTED BY:

ABHINAV TRIPATHI (RA1911003030147)

SHUBH MEHROTRA (RA1911003030141)

PRAKHAR KUMAR (RA1911003030136)

AKARSH SINGH (RA1911003030165)

PROJECT ID: 23MP0064

SUPERVISED BY:

DR. RUBY SINGH

(Asst. Professor, Department of Computer Science & Engineering)

SRM
INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

SRM Institute of Science and Technology
Delhi NCR Campus, Modinagar,
Ghaziabad (UP)-201204

MAY 2023

BONAFIDE CERTIFICATE

This is to certify that Project Report entitled **"Lead Scoring Analysis For Online Courses Using Sci-Kit Learn"**, which is submitted by Abhinav Tripathi (RA1911003030147), Shubh Mehrotra (RA1911003030141), Prakhar Kumar (RA1911003030136) and Akarsh Singh (RA1911003030165) in the partial fulfillment of the requirement for the award of degree B.Tech(CSE) of SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, Ghaziabad is a record of the candidate own work carried out by them under my own supervision.

....

Dr. Ruby Singh

Project Supervisor

Assistant Professor(CSE)

....

SRM Institute of Science and Technology ★ NCR Campus Ghaziabad ★
HOD
(CSE)

Dr. Akash Punhani

HOD(CSE)

INTERNAL EXAMINER

17/7/23

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We would like to express our heartfelt appreciation to Dr. Ruby Singh Mam, Assistant Professor and Project Supervisor at SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, for his invaluable insights and expertise in the subject matter, which motivated us to work diligently.

Our profound gratitude goes out to Dr. Jitendra Singh and Dr. Rakesh Kumar Yadav, Project Coordinators at SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, for their enlightening guidance and skillful coordination, which served as a perpetual source of inspiration.

We would also like to extend our sincere thanks to Dr. S. Vishwanathan, Director of SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, for his unwavering support that enabled us to undertake and complete our project work.

Our special thanks go to Dr. D. K. Sharma, Dean (Academics), and Dr. R. P. Mahapatra, Dean (E&T) at SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, for their valuable guidance and unconditional support.

We would like to express our gratitude to Dr. Akash Punhani, Head of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Delhi-NCR Campus, Modinagar, for his suggestions and encouragement in completing this project.

We also owe our thanks to all the teaching and non-teaching staff members of our college who provided us with direct or indirect help throughout our studies and project work.

Finally, we would like to express our sincere appreciation to our parents, family members, and friends for their unwavering support and encouragement, and to all our well-wishers.

Abhinav Tripathi (RA1911003030147) Atripathi

Shubh Mehrotra (RA1911003030141) SHUBH

Prakhar Kumar (RA1911003030136) Prakhar

Akarsh Singh (RA1911003030165)

DECLARATION

We Abhinav Tripathi (RA1911003030147), Shubh Mehrotra (RA1911003030141), Prakhar Kumar (RA1911003030136) and Akarsh Singh (RA1911003030165) hereby declare that the work which is being presented in the project report **" Lead Scoring Analysis for Online Courses Using Sci-Kit Learn"** is the record of authentic work carried out by us during the period from January 23 to May 23 and submitted by us in partial fulfillment for the award of the degree " Bachelor of Technology in Computer Science and Engineering" to SRM IST, NCR Campus, Ghaziabad (U. P.). This work has not been submitted to any other University or Institute for the award of and Degree/Diploma.

Abhinav Tripathi(RA1911003030147) ATripathi

Shubh Mehrotra (RA1911003030141) SHUBH

Prakhar Kumar (RA1911003030136) Prakhar

Akarsh Singh (RA1911003030165) Aka

ABSTRACT

The project aims to develop a lead scoring system using linear regression to predict the probability of a potential customer becoming a lead for a business. The system involves data collection, preprocessing, feature extraction, model training, model evaluation, deployment, and integration with other systems. The project seeks to improve the accuracy and efficiency of lead scoring by leveraging machine learning techniques. The system aims to be scalable and easily integrated into existing business workflows. The performance of the system will be evaluated through various testing techniques to ensure its accuracy and reliability in predicting lead scores.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CRM	Customer Relationship Management
LSSVR	Lead Squares Support Vector Regression
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
PCA	Principal Component Analysis
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The lead score prediction project aims to provide businesses with a tool to predict the likelihood of a potential customer becoming a paying customer. By analysing various factors such as demographics, online behaviour, and previous interactions with the business, the lead score prediction model can identify the potential customers who are most likely to convert to paying customers.

The project uses linear regression as the primary machine learning algorithm for lead score prediction. Linear regression is a simple and interpretable algorithm that is well-suited for regression tasks, making it a suitable choice for lead score prediction. This proposal consists of several steps, such as data collection, feature extraction, preprocessing, model training, model evaluation, and deployment and integration.

The project aims to address the limitations of the existing system for lead score prediction using SVM. The SVM algorithm can be complex, sensitive to outliers, difficult to interpret, and struggle with non-linear relationships and large or imbalanced datasets. By using linear regression as the primary machine learning algorithm, the proposed system can provide a simpler and more interpretable alternative to the existing system.

The lead score prediction project has several potential benefits for businesses, including increased efficiency and effectiveness in identifying potential customers who are more likely to convert to paying customers. By identifying these potential customers, businesses can target them with personalised marketing campaigns and increase their chances of conversion. Marketing automation platforms, businesses can enhance its impact and maximise its potential benefits.

1.2 PROBLEM DESCRIPTION

The lead score prediction project aims to address the problem of identifying potential customers who are most likely to convert to paying customers. This is a common problem faced by businesses of all sizes and across different industries. In today's highly competitive market, businesses need to identify potential customers who are most likely to convert to paying customers and target them with personalised marketing campaigns to maximise their chancesof conversion.

The traditional approach to lead scoring involves manually identifying potential customers based on certain characteristics such as demographics, online behaviour, and previous interactions with the business. However, this approach can consume time, is prone to errors, and is limited in its effectiveness, as it relies on subjective human judgment to identify potential customers.

Machine learning algorithms such as SVM and linear regression can provide a more efficient and effective approach to lead scoring. These algorithms can analyze large amounts of data and identify the potential customers who are most likely to convert to paying customers based on their past behaviours and characteristics.

However, the existing system for lead score prediction using SVM has several limitations, including its complexity, sensitivity to outliers, difficulty in interpretation, and struggles with non-linear relationships and large or imbalanced datasets. These limitations can limit its effectiveness in certain scenarios.

The proposed system using linear regression for lead score prediction aims to address these limitations by providing a simpler and more interpretable alternative to the existing system. Linear regression is a well-known and widely used machine learning algorithm that is simple to implement and interpret, making it easier for businesses to understand and act on the predictions.

Additionally, linear regression can handle a myriad of data types and can scale to large datasets.

In conclusion, the lead score prediction project aims to address the problem of identifying potential customers who are most likely to convert to paying customers. By using machine learning algorithms such as linear regression, businesses can improve their lead-scoring process and increase their chances of conversion. The proposed system using linear regression for lead score prediction provides a simpler and more interpretable alternative to the existing system, addressing the limitations of SVM and providing businesses with a more efficient and effective approach to lead scoring.

1.3 OBJECTIVE

The lead score prediction project aims to provide businesses with an efficient and effective tool to identify potential customers who are most likely to convert to paying customers. The project aims to leverage ML Algorithms such as linear regression which is used to analyse enormous amounts of data and identify patterns and characteristics of potential customers that are associated with higher conversion rates.

The ultimate goal of the project is to help businesses increase their efficiency and effectiveness in identifying potential customers and targeting them with personalised marketing campaigns. By accurately predicting lead scores, businesses can prioritise their efforts and resources towards potential customers who are most likely to convert, improving their chances of conversion and maximising their return on investment.

Additionally, Lead score prediction is a vital aspect of any successful marketing campaign. However, the existing system for lead score prediction using SVM has certain limitations that need to be addressed. One of the biggest challenges is the complexity of the SVM algorithm, which makes it difficult to interpret the results.

To overcome these limitations, the project aims to provide a simpler and more interpretable alternative using linear regression. Linear regression is a straightforward statistical technique that can be easily understood by people with limited technical expertise. This approach will provide a more transparent and intuitive way to predict lead scores.

Moreover, the simplicity of linear regression will make it easier to implement and maintain in the long run. This will not only save time and resources but also improve the accuracy of lead score prediction. With the help of linear regression, marketers will be able to identify high-quality leads more efficiently and effectively, which will ultimately lead to higher conversion rates and increased revenue.

Overall, the project's goal is to leverage the power of linear regression to create a more accessible and user-friendly lead score prediction system that delivers actionable insights and drives business growth.

CHAPTER 2

LITERATURE SURVEY

1. The retail industry has largely adopted various inventory management programs, and some retailers even use prediction software to forecast future projected sales. However, much of India's daily buying is done in small stores. The owners of such small businesses may not have the cash to invest in proprietary applications to set up an inventory management system. Needless to add, the same is true for any sales forecasting program. As a result, many retailers wind up stockpiling a large number of useless and non-profitable products, resulting in financial losses. A mobile application that delivers all of the functionality of a point-of-sale system as well as future sales analytics is a very cost-effective and accessible solution to this problem. It will allow store owners to handle their current product purchases and billing. The predicted sales analysis will assist them in modifying their product and supply investments, ensuring maximum revenues. If a store has a relevant product that meets the wants of its customers, its customer base will grow. The Economic Times reported in May 2019 that the number of smartphone users in India is predicted to increase by 84% to 859 million by 2022, up from 468 million in 2017. It is realistic to expect that many store owners will have cell phones in the coming years. As a result, providing local shops with a mobile application will be beneficial because it will expose them to all of the above mentioned benefits.

2. Smart data mining approaches with accurate prediction models and great dependability are required for sales prediction analysis. Most market segments, in essence, rely on the know-how base and demand trend forecast for analysis of Business To Business (B2B) sales data. Sales data are offered on how a telecommunications company should manage its sales team, goods, and budgetary processes. Telecommunication Company can survive the industry fight and improve its market growth thanks to precise estimates. In this study, intelligible

predictive models are studied and analysed using machine learning approaches to

improve future sales projections. Traditional forecasting systems struggle to deal with massive data and sales forecasting accuracy. This study provides a brief examination of the dependability of B2B sales using machine learning approaches. The final section of this study explains a variety of sales forecast methodologies and treatments. Based on the performance evaluation, the best-suited predictive model for B2B sales trend forecasting is provided. The results of projection, estimation, and analysis are described in terms of the dependability and consistency of effective prediction and forecasting methodologies. The findings of this analysis are intended to produce dependable, accurate, and effective forecasting data, which will be a significant resource for sales forecasting. According to research, the Gradient Boost Algorithm offers a high level of accuracy in predicting and future B2B sales prediction, with MSE = 24,743,000,000.00 and MAPE = 0.18.

3. The main goal of the sales prediction regression model is to anticipate automobile sales using sentiment analysis from various websites. The online presence of a car, as well as its brand, are important factors in automobile sales. Many more criteria, however, are essential and will be covered in this work. In today's industry, sales forecasting benefits not only the manufacturer but also several other companies that make vehicle parts or accessories. It is also beneficial to retailers, showroom owners, and service mechanics. We used linear regression for sentiment analysis and polynomial regression for sales prediction in this application.

4. Sales forecasting is a method of predicting future sales. Accurate income estimates enable the retail market to make informed business decisions and predict short- and long-term performance. Sales forecasting enables a shop to estimate its projected future revenues for income earned over a specific time period. As a result, time is an important factor in sales forecasting. A time sequence is a sequence of information components that are indexed in time order.

The time sequence analysis evaluates observations as a set of statistics at specific intervals across time, with the goal of determining trends, cycles, and seasonal variances that can be used to forecast future events. With the advancement of data technology, large retailers have begun to utilise statistical approaches such as index numbers, time collection, and a couple of regression evaluations for the purpose of income forecasting. For prediction, the XG Boost method was used in this research. The program makes use of sales data from Ross-man, which operatesover 8000 drug stores across seven European countries. Machine Learning is transforming every walk of life and has become a major contributor in real-world scenarios. The revolutionary applications of Machine Learning can be seen in every field including education, healthcare, engineering, sales, entertainment, transport, and several more; the list is never-ending. The traditional approach of sales and marketing goals no longer helps the companies to cope with the pace of the competitive market, as they are carried out with no insights into customers' purchasing patterns. Major transformations can be seen in the domain of sales and marketing as a result of Machine Learning advancements. Owing to such advancements, various critical aspects such as consumers' purchase patterns, target audience, and predicting sales for the recent years to come can be easily determined, thus helping the sales team in formulating plans for a boost in their business. The aim of this paper is to propose a dimension for predicting the future sales of Big Mart Companies keeping in view the sales of previous years. A comprehensive study of sales prediction is done using Machine Learning models such as Linear Regression, K-Neighbor's Regressor, XGBoost Regressor, and Random Forest Regressor. The prediction includes data parameters such as item weight, item fat content, item visibility, item type, item MRP, outlet establishment year, outlet size, and outlet location type.

5. To process into non-overlapped sections, seeded region growth and the LTHV method for segmentation of MRI brain images were utilized. On the basis of metric topological neighbourhoods, a novel region-growing segmentation

technique has been developed. The quality of segmentation was assessed using the novel objective metric entropy in addition to classic validity measurementssuch as accuracy, PSNR, and MSE. Traditional forecast systems struggle to dealwith massive data and the accuracy of sales forecasts. These problems could besolved by employing various data mining approaches. We examined the topicof sales data and sales forecasts briefly in this study. The numerous strategies andmeasures for sales forecasting are explained in the research work's later section. Abest-fit prediction model for the sales trend forecast is suggested based on aperformance review. The findings are reported in terms of the dependability and accuracy of effective prediction and forecasting techniques. According to the research, the best-fit model is the Gradient Boost Algorithm, which has the highest accuracy in forecasting and future sales prediction.

6. An Intelligent Decision Analytical System must incorporate decision analysis and predictions. The majority of commercial organisations rely largely on a knowledge base and sales trend forecasting. The accuracy of sales forecasting has a significant impact on business. Data mining techniques are particularly powerful tools for extracting hidden knowledge from massive datasets in order to improve predicting accuracy and efficiency. This research focuses on the deep examination and development of understandable predictive models to improve future sales projections. Traditional forecasting systems struggle to deal with massive data and the accuracy of sales forecasting. These problems could be solved by employing various data mining approaches. We examined the topic of sales data and sales forecasting briefly in this study. The numerous strategies and measures for sales forecasting are explained in the research work's later section. A best-fit prediction model for the sales trend forecast is suggested based on a performance review. The findings are reported in terms of the dependability and accuracy of effective prediction and forecasting techniques. According to the research, the best-fit model is the Gradient Boost Algorithm, which has the highest accuracy in forecasting and future sales prediction.

7. Sales forecasting is critical for businesses involved in retailing, shipping, manufacturing, marketing, and wholesaling. It enables businesses to efficiently manage resources, forecast achievable sales revenue, and develop a better strategy for the company's future growth. In this study, the prediction of product sales from a specific outlet is done using a two-level strategy that offers superior predictive performance than any of the common single-model predictive learning algorithms. The method is applied to Big Mart Sales data from 2013. Data exploration, data translation, and feature engineering all play important roles in forecasting accurate outcomes. The results showed that the two-level statistical technique outperformed the single model approach because it supplied more information, which leads to better prediction.

8. A comprehensive model for estimating power system demand over diverse time periods, from short to long term, is an essential tool for efficient system management. However, as the time horizon in load forecasting expands, it will become more difficult to provide an accurate forecast. Machine learning approaches have gained popularity as efficient methods for coping with stochastic load patterns and producing reliable forecasting. The problem of long-term load forecasting for the New England Network case study is studied in this study using several commonly used machine learning methods such as feedforward artificial neural network, support vector machine, recurrent neural network, generalised regression neural network, k-nearest neighbours, and Gaussian Process Regression. These methods' results are compared using mean absolute percentage error (MAPE).

9. Because of the growth of social media, writing comments or expressing opinions about items online is now easier than ever. Data from social media could be one of the most important inputs for projecting automobile sales. Furthermore, other factors, such as stock market valuations, have an impact on automobile purchasing ability. To estimate monthly total car sales, this article use both multivariate regression models incorporating social media data and stock market

values and time series models. When dealing with multivariate regression data, theleast squares support vector regression (LSSVR) models are used. This article forecasts monthly total car sales in the United States using three pieces of data: sentiment scores from tweets, stock market values, and hybrid data. The hybrid data set includes both tweet sentiment scores and stock market values. Furthermore, monthly total vehicle sales seasonal variables are used to de seasonalize both monthly total vehicle sales and three types of input data. The Nave model, the exponential smoothing model, the autoregressive integrated moving average model, the seasonal autoregressive integrated moving average model, and back-propagation neural networks and LSSVR with time series modelsare among the time series models. The numerical results show that LSSVR models employing hybrid data with de seasonalizing processes can produce more accurate results than other models using different data.

S.NO	TOPIC	INFERENCE	ADVANTAGES	DISADVANTAGE	ALGORITHM
1	Evaluation of CET-4 and CET-6 Courses onOnline Learning Platform [1]	The study solves the issue by evaluating Tencent classroom courses by course organization, pricing, registration population, and other indices using data mining, statistical analysis, word cloud analysis, and other approaches	The paper also selects courses depending on learners' willingness to pay.	The accuracy was not as expected, could have been more precise.	Data mining, statistic analysis
2	Research on the Construction-of Network Course Platform Based on Big Data [2]	Curriculum development is key to improving higher education. Ourstudy aims to create a big data-era cloud platform for teaching popular music at universities. The user hub's main features are WeChat	An intuitive model developmen tinterface simplifies datamining model management .	The model lacked the higher order function alitie s.	Big Data

3	Development of anOnline Course for Web Programming [3]	login, profiles, Theoretical and practical features allow the author to build a course, specify a time period, indicatea measure, and assess student performance after course	Shows OLAT online course implementation. These technologies tailor learning to student comprehensive on.	Speed in terms of processing was low.	We use OLAT.
4	A designed platform for programming courses [4]	The portal validates codes and manages courses, students,and issues. Code format and comments help pupils learn tocode. Teachers can assess class	The platform improved teaching aftera semester oftesting.	The model was designed for limited or specific coursesalone	MOOCs
5	Online Teaching Practice in Chinese Culture Course [5]	and statistical data. This study focuses on online teaching and Chinese cultural course student responses	Online educationiseasy and practical.	The accuracy wasnot as expected ,	Experimental Analysis
6	Research on Construction of the Quality Evaluation Indicator System of "Online Golden Course [6]	The "Online Gold Course" demands rigorous and impartial quality assessment. Using Total Quality Management theory and influencing factors, this research creates acurriculum quality evaluation indicator system	Principal Component Analysis improves student-centered "Online Golden Course" quantitative evaluation	More complex	Virtual Equipment Digital Media Technology

7	Online Immersive Platform Image Processing Algorith m for Art Courses based on Virtual Equipment Digital Media Technology [7]	This article describes virtual reality technologyand digital media art, analyzes theirproduction,and establishes their meaning, connection, benefits, and qualities. Hemisphere	An embedded processing platform image processing confidence interval and mathematical expectation technique reduces	The model lacked the higher order functionalities.	Imag e Semantic Segmentation
8	Applicationof Audio Processing Based on Image Semantic Segmentation-in Applied Mathematics Online Course [8]	It analyzes applied mathematics network courses, explores the connotation, characteristics, strategies, and information resources of engineering mathematics network courses,and builds the online course teaching platform	noise . It boosts afterclass review to6.5 %and promotes lifelong learnin g.	The model wasdesigned for mathematical -centric courses.	MOOCs

9	Research on Evaluati on System of User ExperienceWith Online Live Course Platform [9]	for students to Study proposes a comprehensiv e user experience evaluation systemto identify platform flawsand suggest improvementsfor online courses.	Two software samples have significant differences inmultiple second-level indices buttheir overall scores are	Speed in termsof processingwas low.	Modeling-eliciting activities

CHAPTER 3

EXISTING SYSTEM

Due to a number of drawbacks, the current system for lead score predictionutilizing SVM isn't always the best choice for lead scoring tasks. A popular supervised learning technique for classification and regression tasks is SVM. Finding a hyperplane in the feature space that divides the data into distinct classes and has the greatest distance from the closest data points is thefundamental goal of SVM.

One disadvantage of SVM is its complexity. SVM is a complex algorithm that involves solving a quadratic optimization problem. This can increase the cost of computing, especially when working with huge datasets or high-dimensional feature spaces. Additionally, the optimal SVM model's parameters, such as the regularization parameters and the kernel functions, need to be tuned carefully to achieve good performance. This tuning process can be time consuming and require a high degree of expertise.

Another disadvantage of SVM is its sensitivity to outliers. The SVM algorithm, which is based on margins, seeks out the hyperplane that maximizes the margin between the classes. However, this can make SVM sensitive to outliers, as outliers can have a large impact on the position of the hyperplane. This sensitivity can lead to poor performance when dealing with noisy or imbalanced datasets.

Additionally, SVM can be difficult to interpret. The hyperplane found by SVM can be difficult to interpret, as it is defined in the high-dimensional feature space. This can make it difficult to understand which features are most important for predicting the lead score. Businesses may find it difficult to comprehend and use the projections as a result of this lack of interpretability.

Additionally, non-linear connections between the input characteristics and the target variable can be challenging for SVM to handle. The assumption that the relationship between the input characteristics and the target variable is linear underlies SVM. This assumption may not hold true in many real-world situations, and there may be non-linear correlations between the input properties and the goal variable. SVM uses kernel functions, which translate the input features into a higher-dimensional space where the data might be more separable, to handle non-linear correlations. However, the selection of the kernel function must be made carefully because it can significantly affect the SVM model's performance.

Finally, SVM can struggle with large or imbalanced datasets. SVM requires that all data points are kept in memory during training, which can be challenging for large datasets. Additionally, when the dataset is imbalanced, SVM may be biased in favour of the dominant class, which could lead to subpar performance for the minority class.

In conclusion, the existing system for lead score prediction using SVM has several disadvantages that make it less suitable for lead scoring tasks in certain scenarios. The complexity of SVM can make it computationally expensive and require a high degree of expertise to tune the optimal parameters. Additionally, the sensitivity of SVM to outliers, its lack of interpretability, and its struggles with non-linear relationships and large or imbalanced datasets can limit its effectiveness. While SVM remains a useful algorithm for many machine learning tasks, businesses should carefully consider its limitations when selecting an algorithm for lead score prediction.

CHAPTER 4

PROPOSED SYSTEM

The proposed system using linear regression for lead score prediction is a simpler and more interpretable alternative to the existing system that uses SVM. As a supervised learning approach that can be used to regression tasks, linear regression is a good option for predicting lead score. The suggested system entails a number of phases, such as data collection and preprocessing, feature extraction, model training, model evaluation, deployment, and integration.

Data collection and preprocessing involve collecting data from various sourcesand preprocessing it to ensure its accuracy, completeness, and relevance for the lead score prediction model. Feature extraction involves identifying the most important features that can be used to predict the lead score. Techniquessuch as correlation analysis, PCA, or RFE can be used for this step.

Model training involves using the preprocessed and feature-extracted data to train the linear regression model. The model is trained by finding the best fit line through the data points that reduce the squared error between the target variable's actual values and its anticipated values. Evaluation of the model entails assessing how well the trained linear regression model using techniques such as MSE or R-squared.

The proposed system using linear regression has several advantages, including simplicity, interpretability, scalability, and flexibility. Linear regression is a straightforward and simple method, which makes it simpler for developers to comprehend and analyse the model. It is also highly interpretable, making it easier to understand how the model is making its predictions and take appropriate actions based on the predictions. Additionally, linear regression can scale well to large datasets and handle a wide range of data types.

However, the proposed system also has some limitations. Linear regression makes several assumptions about the data, including linearity, independence, homoscedasticity, and normality. Violations of these assumptions can lead to poor model performance. Additionally, linear regression is particularly prone to overfitting when the number of features is high in comparison to the number of observations.

The simplicity and interpretability of linear regression can make it easier for developers to understand and interpret the model, while its scalability and flexibility can make it suitable for lead score prediction tasks that involve large amounts of data and different data types. However, developers should also consider the assumptions and limitations of linear regression when designing and implementing the system.

CHAPTER 5

METHODOLOGIES

The modules included are

- Data Collection and Preprocessing
- Feature Extraction
- Model Training
- Model Evaluation
- Deployment and Integration

MODULE 1: DATACOLLECTIONAND PREPROCESSING

Data collection and preprocessing is a critical step in lead score prediction using linear regression. The accuracy of the lead score prediction model can be significantly impacted by the quality and accuracy of the data that was collected. Data preprocessing ensures that the collected data is accurate, complete, and relevant for the lead score prediction model.

Data collection may involve collecting data from various sources, such as CRM systems, marketing automation platforms, and social media platforms. The collected data may include information such as customer demographics, customer behavior, and customer interactions with the business.

To make sure the data is precise and pertinent for the lead score prediction model, it must be preprocessed when it is obtained. Several steps may be involved in data preparation, including:

- Data Cleaning: The data cleaning stage entails getting rid of any irrelevant or duplicate data from the collected data. This step may also involve filling in missing data values using techniques such as imputation.

- Data Transformation: The data transformation step involves converting the data gathered into a format that is suitable for the lead score prediction model. This step may involve converting categorical data into numerical data using techniques such as one-hot encoding.
- Data Normalization: The data normalization step involves scaling the data to ensure that it is on a similar scale. This step may involve techniques such as min-max normalization or standardization.
- Feature Selection: The feature selection step involves selecting the most important features from the preprocessed data for the lead score prediction model. This step may involve using techniques such as correlation analysis or principal component analysis (PCA).

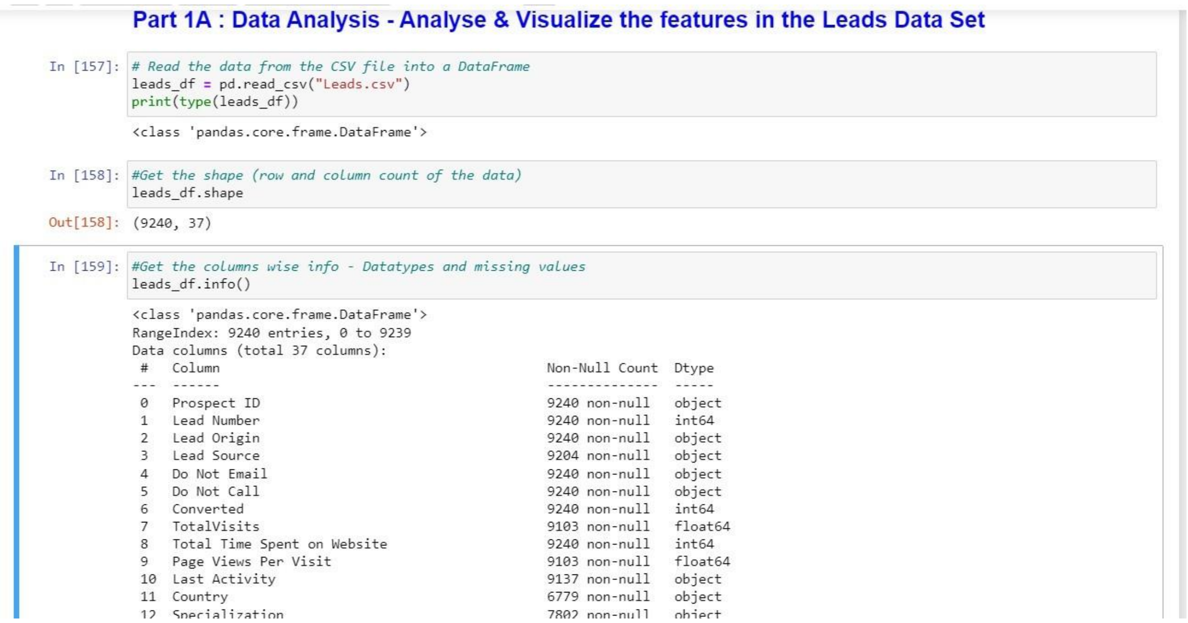
Another important consideration in data collection and preprocessing for lead score prediction using linear regression is data bias. Data bias refers to the presence of any systematic error or prejudice in the collected data, which can result in inaccurate or biased predictions. Data bias can be caused by a number of things, including selection bias, measurement bias, and sampling bias.

To mitigate the risk of data bias, developers should carefully evaluate the sources of data and the methods used to collect the data. They should also consider the potential impact of data bias on the lead score prediction model and take appropriate steps to reduce or eliminate any bias that may be present.

In addition to data bias, developers should also consider the scalability and maintainability of the data collection and preprocessing pipeline. As the amount of data grows, it may become necessary to implement more scalable data collection and preprocessing techniques, such as distributed computing or parallel processing. Additionally, developers should ensure that the data collection and preprocessing pipeline is maintainable and can be easily updated or modified as needed.

Once the data has been collected and preprocessed, it is important to evaluate the quality of the data and ensure that it is suitable for use in the lead score prediction model. This may involve conducting exploratory data analysis to identify any patterns or trends in the data, as well as conducting hypothesis testing to validate the assumptions underlying the linear regression model.

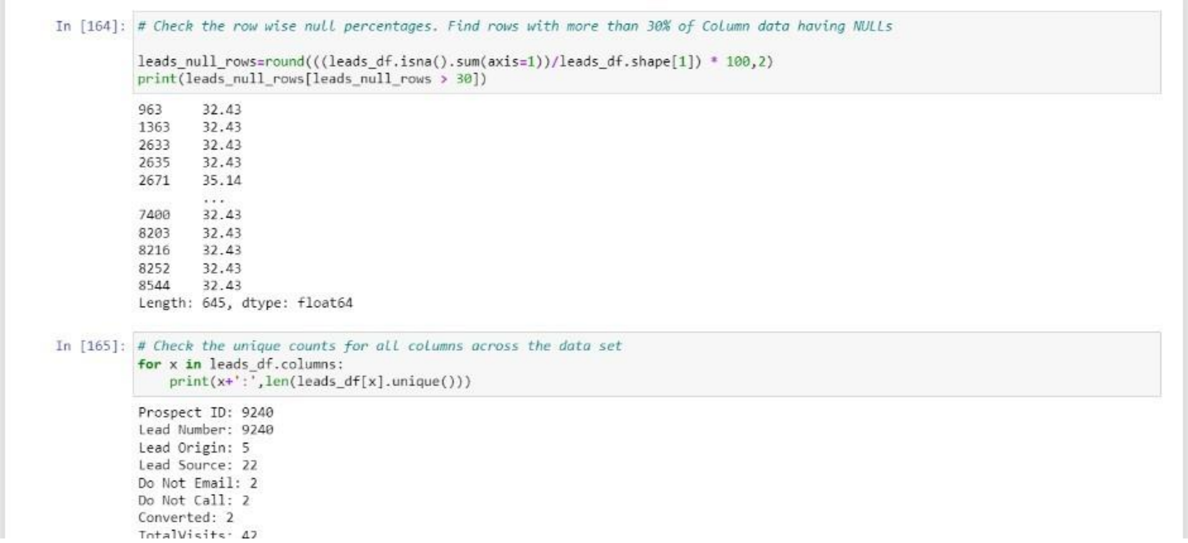
Overall, data collection and preprocessing are critical steps in lead score prediction using linear regression. By carefully collecting and preprocessing data, developers can ensure that the lead score prediction model is accurate, reliable, and effective. Additionally, by considering factors such as data bias, scalability, and maintainability, developers can ensure that the data collection and preprocessing pipeline is robust and can support the needs of the business over the long term.



Part 1A : Data Analysis - Analyse & Visualize the features in the Leads Data Set

```
In [157]: # Read the data from the CSV file into a DataFrame
          leads_df = pd.read_csv("Leads.csv")
          print(type(leads_df))

          <class 'pandas.core.frame.DataFrame'>

In [158]: #Get the shape (row and column count of the data)
          leads_df.shape

Out[158]: (9240, 37)

In [159]: #Get the columns wise info - Datatypes and missing values
          leads_df.info()

          <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 9240 entries, 0 to 9239
          Data columns (total 37 columns):
           #   Column                                        Non-Null Count  Dtype
          ---  ------                                        --------------  -----
           0   Prospect ID                                   9240 non-null   object
           1   Lead Number                                   9240 non-null   int64
           2   Lead Origin                                   9240 non-null   object
           3   Lead Source                                   9204 non-null   object
           4   Do Not Email                                  9240 non-null   object
           5   Do Not Call                                   9240 non-null   object
           6   Converted                                     9240 non-null   int64
           7   TotalVisits                                   9103 non-null   float64
           8   Total Time Spent on Website                   9240 non-null   int64
           9   Page Views Per Visit                          9103 non-null   float64
           10  Last Activity                                 9137 non-null   object
           11  Country                                       6779 non-null   object
           12  Specialization                                7802 non-null   object
```

Figure 5.1: Data Analysis



```
In [164]: # Check the row wise null percentages. Find rows with more than 30% of Column data having NULLs

          leads_null_rows=round(((leads_df.isna().sum(axis=1))/leads_df.shape[1]) * 100,2)
          print(leads_null_rows[leads_null_rows > 30])

          963     32.43
          1363    32.43
          2633    32.43
          2635    32.43
          2671    35.14
                  ...
          7400    32.43
          8203    32.43
          8216    32.43
          8252    32.43
          8544    32.43
          Length: 645, dtype: float64

In [165]: # Check the unique counts for all columns across the data set
          for x in leads_df.columns:
              print(x+':',len(leads_df[x].unique()))

          Prospect ID: 9240
          Lead Number: 9240
          Lead Origin: 5
          Lead Source: 22
          Do Not Email: 2
          Do Not Call: 2
          Converted: 2
          TotalVisits: 42
```

Figure 5.2 : Checks the row wise null percentages.

MODULE 2: FEATURE EXTRACTION

Feature extraction is a crucial step in lead score prediction using linear regression. Finding the most crucial features from the preprocessed data that can be used to forecast the lead score is known as feature extraction. Reduce the dimensionality of the data while maintaining as much pertinent information as you can is the aim of feature extraction.

There are several techniques that can be used for feature extraction in lead score prediction using linear regression, including:

- Correlation Analysis: Calculating the correlation between each feature and the target variable (lead score) is the task of correlation analysis. The most crucial features for the model are those with the highest correlation coefficients.

- Principal Component Analysis (PCA): Using the PCA technique, preprocessed data are converted into a new set of variables (principal components) that are uncorrelated and account for the most variance in the data. The primary components chosen as essential features for the model are those with the largest explained variation.

- Feature elimination using recursion: Feature elimination using recursion is a technique that involves recursively removing the least important features from the preprocessed data until the desired number of features is reached. This technique can be used to identify the most important features for the model.

- Lasso Regression: Lasso regression is a technique that involves applying linear regression with L1 regularization to the preprocessed data. Lasso regression canbe used to identify the most important features for the model while reducing the effects of multicollinearity.

In addition to these techniques, domain expertise and business knowledge can also play a critical role in feature extraction. Experts in the domain of lead scoring may have insights into which features are most relevant for predicting lead scores based on their knowledge of the industry and the specific business needs.

Following the feature extraction process, the significant features can be used to train the linear regression model for lead score prediction. It is important to note that feature extraction is an iterative process, and the identified features may need tobe refined or adjusted as the model is trained and tested.

It's also critical to take into account how feature extraction might affect the lead score prediction model's precision and interpretability. While reducing the dimensionality of the data can help improve the efficiency and scalability of the model, it can also result in a loss of information and potentially decrease theaccuracy of the model. Additionally, selecting features based solely on statistical metrics such as correlation coefficients or explained variance may result in a model that is less interpretable and difficult to explain to stakeholders.

Therefore, it is important to strike a balance between feature extraction techniques and business knowledge to ensure that the selected features are both relevant and interpretable. This can be achieved through a combination of statistical analysis, domain expertise, and feedback from stakeholders.

Overall, feature extraction is a critical step in lead score prediction using linear regression. The lead score prediction mode's precision and effectiveness can be increased by developers by selecting the most crucial elements from the preprocessed data, while also ensuring that the selected features are both relevant and interpretable.

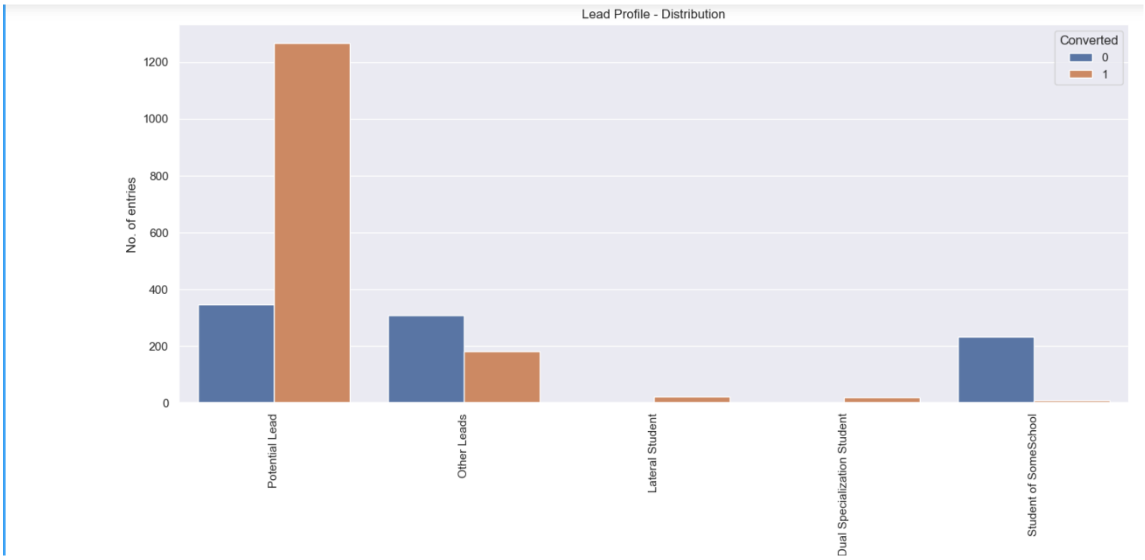
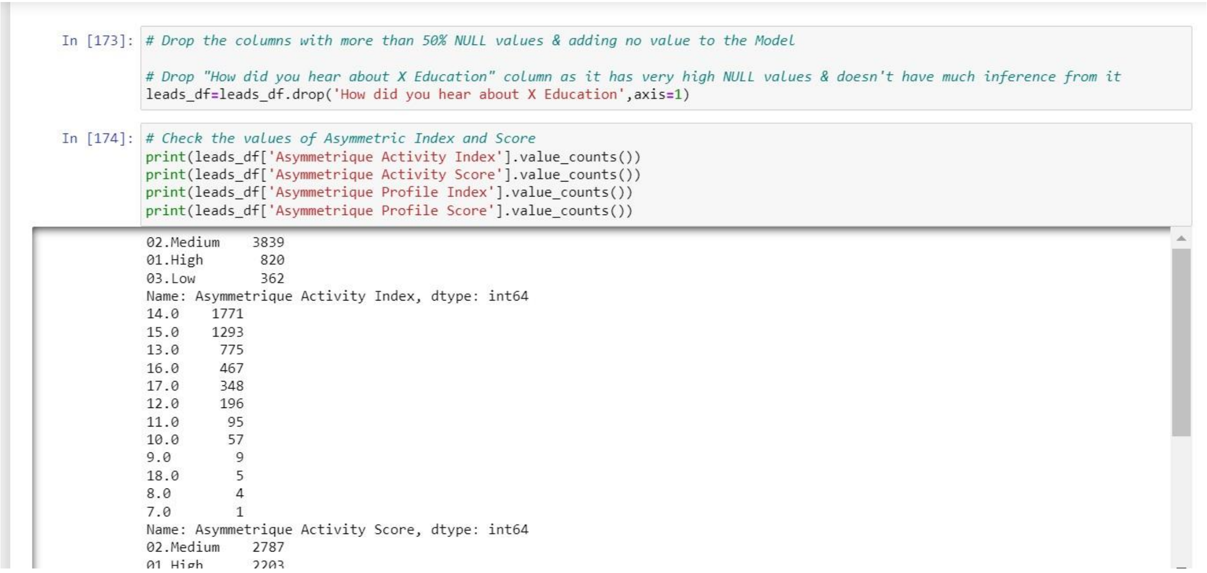


Figure 5.3 : Lead Profile -Distribution



```
In [173]: # Drop the columns with more than 50% NULL values & adding no value to the Model

          # Drop "How did you hear about X Education" column as it has very high NULL values & doesn't have much inference from it
          leads_df=leads_df.drop('How did you hear about X Education',axis=1)
```

```
In [174]: # Check the values of Asymmetric Index and Score
          print(leads_df['Asymmetrique Activity Index'].value_counts())
          print(leads_df['Asymmetrique Activity Score'].value_counts())
          print(leads_df['Asymmetrique Profile Index'].value_counts())
          print(leads_df['Asymmetrique Profile Score'].value_counts())
```

```
02.Medium    3839
01.High       820
03.Low        362
Name: Asymmetrique Activity Index, dtype: int64
14.0    1771
15.0    1293
13.0     775
16.0     467
17.0     348
12.0     196
11.0      95
10.0      57
9.0        9
18.0       5
8.0        4
7.0        1
Name: Asymmetrique Activity Score, dtype: int64
02.Medium    2787
01.High      2203
```

Figure 5.4 : Drop the columns that add’s no value

MODULE 3: MODEL TRAINING

Model training is a crucial step in lead score prediction using linear regression. Model training involves using the preprocessed and feature-extracted data to train a linear regression model that can predict the lead score. Finding the optimal set of coefficients (weights) for the linear regression equation that may successfully predict the lead score is the aim of model training.

There are several techniques that can be used for model training in lead score prediction using linear regression, including:

- Ordinary Least Squares (OLS): OLS is a strategy that includes reducing the sum of squared errors between the target variable's (lead score) actual values and its predicted values. OLS is a common method for linear regression model training.
- Gradient Descent: Gradient descent is an iterative optimisation algorithm thatinvolves minimizing the cost function (mean squared error) by adjusting the model coefficients (weights) in the direction of the negative gradient. Gradient descent can be used to train linear regression models.
- Regularisation: In order to avoid overfitting, the regularisation technique entails including a penalty term in the cost function. Regularisation can be used to train more robust and generalisable linear regression models.
- Cross-Validation: To assess the performance of the model, the preprocessed data issplit into training and validation sets using the cross-validation technique. Cross-validation can be used to optimise the model's hyper-parameters (e.g., regularisation parameter) and prevent overfitting.

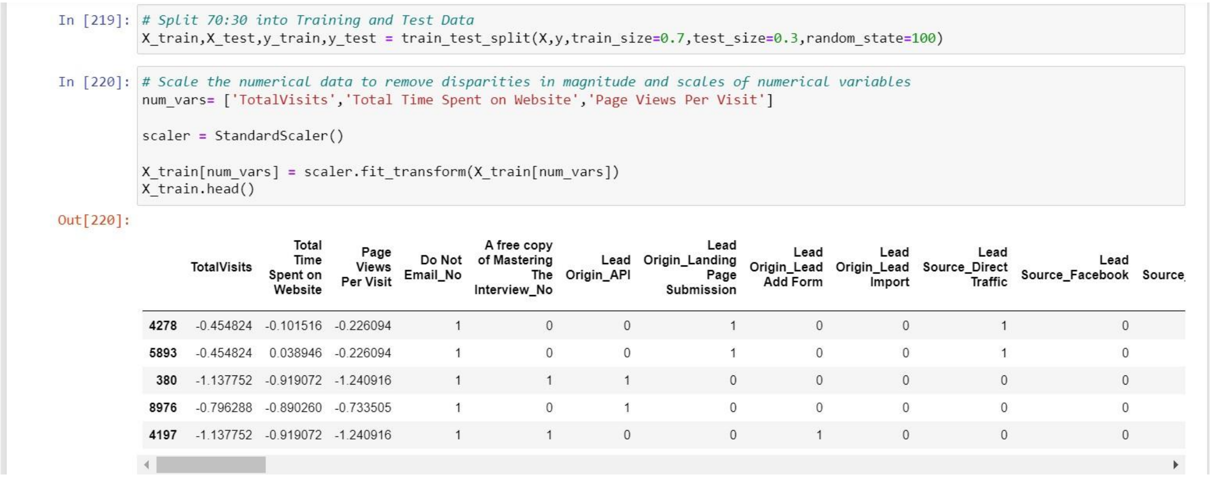
In order to make sure the trained model is accurate and dependable, it is crucial to assess its performance. This can be done by evaluating the model's performance ona test set of data that was not used during the training process. The accuracy and prediction capability of the model can be evaluated using performance metrics including mean squared error, R-squared, and root mean squared error.

It is also crucial to take into account the potential effects of model restrictions and assumptions on the lead score prediction model's accuracy. The assumptions of linear regression are that the characteristics and the target variable have a linear relationship and that the errors are normally distributed and homoscedastic. If these assumptions are violated, it can result in inaccurate or biased predictions.

To address these potential limitations, developers should carefully evaluate the model assumptions and use appropriate techniques such as residual analysis or transforming the data to ensure that the model is robust and accurate.

Another important consideration in model training is scalability and efficiency. As the amount of data grows, it may become necessary to implement more scalable and efficient techniques such as distributed computing or parallel processing. Additionally, developers should ensure that the trained model can be easily updated or modified as needed.

Overall, model training is a critical step in lead score prediction using linear regression. By using appropriate model training techniques and carefully evaluating the model's assumptions and limitations, developers can ensure that the lead score prediction model is accurate, reliable, and effective. By considering scalability and efficiency, developers can ensure that the model can handle large volumes of data and is scalable for future needs.

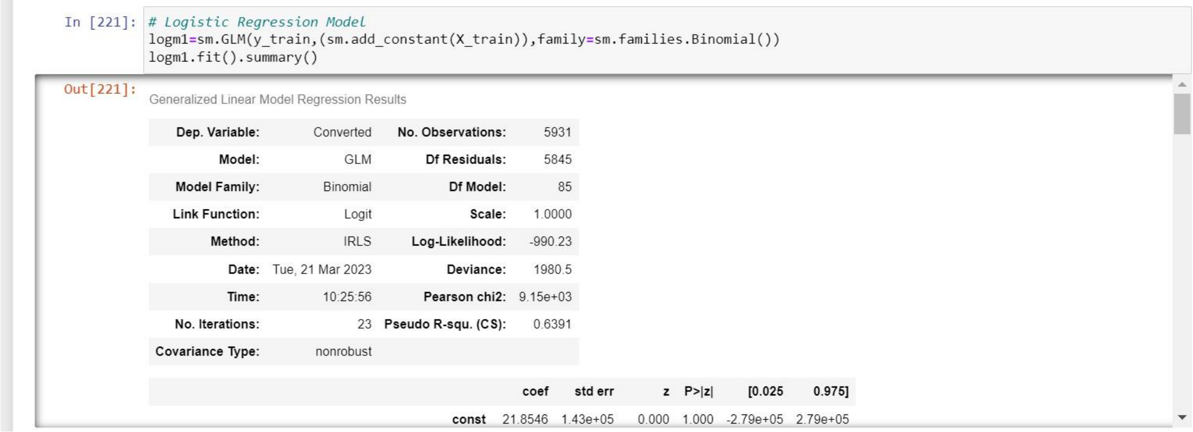


```
In [219]: # Split 70:30 into Training and Test Data
          X_train,X_test,y_train,y_test = train_test_split(X,y,train_size=0.7,test_size=0.3,random_state=100)

In [220]: # Scale the numerical data to remove disparities in magnitude and scales of numerical variables
          num_vars= ['TotalVisits','Total Time Spent on Website','Page Views Per Visit']

          scaler = StandardScaler()

          X_train[num_vars] = scaler.fit_transform(X_train[num_vars])
          X_train.head()
```

Out[220]:

	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Do Not Email_No	A free copy of Mastering The Interview_No	Lead Origin_API	Lead Origin_Landing Page Submission	Lead Origin_Lead Add Form	Lead Origin_Lead Import	Lead Source_Direct Traffic	Lead Source_Facebook	Source,
4278	-0.454824	-0.101516	-0.226094	1	0	0	1	0	0	1	0	
5893	-0.454824	0.038946	-0.226094	1	0	0	1	0	0	1	0	
380	-1.137752	-0.919072	-1.240916	1	1	1	0	0	0	0	0	
8976	-0.796288	-0.890260	-0.733505	1	0	1	0	0	0	0	0	
4197	-1.137752	-0.919072	-1.240916	1	1	0	0	1	0	0	0	

Figure 5.5 : Splitting the data into train and test



```
In [221]: # Logistic Regression Model
          logm1=sm.GLM(y_train,(sm.add_constant(X_train)),family=sm.families.Binomial())
          logm1.fit().summary()
```

Out[221]:

Generalized Linear Model Regression Results

Dep. Variable:	Converted	**No. Observations:**	5931
Model:	GLM	**Df Residuals:**	5845
Model Family:	Binomial	**Df Model:**	85
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-990.23
Date:	Tue, 21 Mar 2023	**Deviance:**	1980.5
Time:	10:25:56	**Pearson chi2:**	9.15e+03
No. Iterations:	23	**Pseudo R-squ. (CS):**	0.6391
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	21.8546	1.43e+05	0.000	1.000	-2.79e+05	2.79e+05

Figure 5.6 : Model Building

MODULE 4: MODEL EVALUATION

Model evaluation is a crucial step in lead score prediction using linear regression. Model evaluation involves measuring the performance of the trained linear regression model to determine how accurately it predicts the lead score. The goal of model evaluation is to assess the model's accuracy, reliability, and generalisation ability.

There are several techniques that can be used for model evaluation in lead score prediction using linear regression, including:

- Mean Squared Error (MSE): The average squared difference between the target variable's (lead score) projected and actual values is measured by the MSE metric. Alower MSE indicates a more accurate model.
- R-squared: R-squared is a metric that gauges how much of the target variable's (the lead score) variance the linear regression model is able to account for. A higher R-squared indicates a more accurate model.
- Residual Analysis: Residual analysis involves analyzing the residuals (the difference between the predicted values and the actual values of the target variable) to ensure that they are normally distributed and have constant variance. Non-normal residuals or heteroscedastic residuals may indicate model bias or poor model fit.
- Cross-Validation: Cross-validation can also be used for model evaluation by assessing the model's performance on unseen data. This technique can help identify any overfitting or underfitting issues and ensure that the model has good generalisation ability.

It is important to interpret the results of model evaluation in the context of the business problem being addressed. Developers should consider the potential impact of false positives and false negatives on the business and ensure that the model’s performance meets the required thresholds for accuracy and reliability.

Additionally, it may be necessary to optimize the model's hyperparameters (e.g., regularization parameter) to achieve the best possible performance. To do this, it may be necessary to use methods like grid search or randomized search to identify the hyperparameter values that will maximize the model's performance.

It is also important to compare the performance of the linear regression model with other machine learning models to ensure that it is the best approach for lead score prediction in the given context. For example, it may be necessary to compare the performance of the linear regression model with other approaches like support vector machines, random forests, or decision trees.

Finally, developers should consider the potential limitations and uncertainties associated with lead score prediction using linear regression. The accuracy and reliability of the model may be affected by factors such as data quality, model assumptions, and changes in the business environment. To make sure the model is correct and up to date, it must be regularly monitored and updated.

By using appropriate model evaluation techniques and interpreting the results in the context of the business problem, developers can ensure that the lead score prediction model is accurate, reliable, and effective. By optimizing the model's hyperparameters and comparing its performance with other machine learning models, developers can identify the best approach for lead score prediction in the given context. By considering the potential limitations and uncertainties associated with the model, developers can ensure that it remains accurate and relevant overtime.

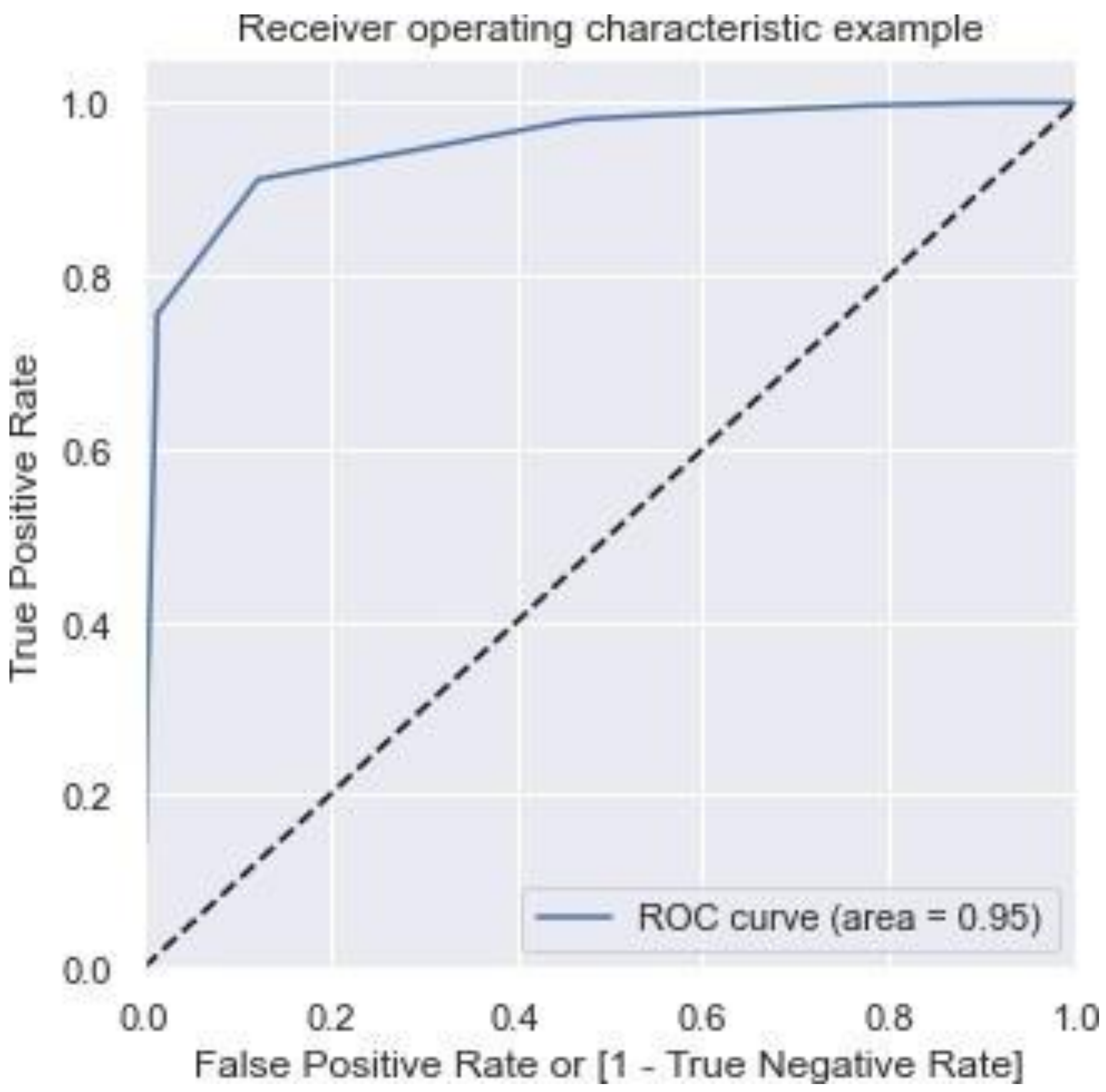


Figure 5.7 ROC Curve

MODULE 5: DEPLOYMENTAND INTEGRATION

Deployment and integration are crucial steps in lead score prediction using linear regression. Deployment involves deploying the trained linear regression model into a production environment, while integration involves integrating the model withother systems, such as CRM systems or marketing automation platforms. There are several techniques that can be used for deployment and integration in lead score prediction using linear regression, including:

API Deployment: API deployment involves deploying the trained model as a RESTful API that can be accessed by other systems. This technique can make the model easily accessible and scalable.

Containerisation: It involves packaging the trained model into a container using technologies such as Docker. Containerisation can make the model easier to deploy and manage in production environments.

Model Versioning: Model versioning involves tracking the different versions of the trained model and their associated changes. Model versioning can help ensure that the correct version of the model is used in production environments.

Integration with Other Systems: Integration with other systems, such as CRM systems or marketing automation platforms, can be achieved through API integration or custom integration. Integration with other systems can ensure that the lead score prediction model is used effectively in business workflows.

There are various additional elements that developers should take into consideration in addition to the strategies outlined above during deployment and integration of the lead score prediction model using linear regression. These include:

Scalability: The deployed model should be scalable to accommodate an increasing amount of data and usage over time. This can be achieved through techniques such as parallelization or distributed computing.

Monitoring and Maintenance: The deployed model should be monitored and maintained to ensure that it continues to perform accurately and reliably over time. This can involve monitoring metrics such as model performance and error rates and making necessary adjustments to improve the model's performance.

Data Security and Privacy: The deployed model should ensure that the collected data is secure and privacy-compliant. This can involve techniques such as data encryption, access control, and anonymization.

Continuous Integration and Deployment: Continuous integration and deployment involve automating the deployment and integration process to ensure that the model is always up-to-date and available to other systems. This can help reduce downtime and improve efficiency.

Overall, deployment and integration are critical steps in the lead score prediction using linear regression. By considering factors such as scalability, monitoring, data security, and continuous integration, developers can ensure that the model is effective and useful in a production environment.

CHAPTER 6

IMPLEMENTATION

```
import streamlit

as stimport

joblibimport

pandas as

pdimport

numpyas np

from sklearn.preprocessing import StandardScaler


st.title("Lead scoring dataset")
model = joblib.load("leads.C5")


# Total Time Spent on Website

total_time_spent = st.number_input("Total Time Spent on Website", value=0)


# Lead Origin_lead add form

lead_origin = st.number_input("Lead Origin_lead add form", value=0)


# Lead Source_direct traffic

lead_source_direct = st.number_input("Lead Source_direct traffic", value=0)


# Lead Source_welingak website

lead_source_welingak = st.number_input("Lead Source_welingak website",

value=0)


# Do Not Email_yes
do_not_email = st.number_input("Do Not Email_yes", value=0)
```

```
# LastActivity_had a phone conversation
last_activity_phone = st.number_input("Last Activity_had a phone
conversation",value=0)

# LastActivity_olark chat conversation
last_activity_chat = st.number_input("Last Activity_olark chat conversation",
value=0)

# LastActivity_sms sent
last_activity_sms = st.number_input("Last Activity_sms sent", value=0)


# What is your current occupation_working
professionaloccupation_working =
st.number_input("What is your
currentoccupation_workingprofessional", value=0)

# Last Notable Activity_unreachable
last_notable_activity = st.number_input("Last Notable Activity_unreachable",
value=0)

features =
np.array([total_time_spent,lead_origin,lead_source_direct,lead_source_welingak,do
_not_email,last_activity_phone,last_activity_chat,last_activity_sms,occupation_wo
rking,last_notable_activity])
features =
features.reshape(1,-1)scale
=
StandardScaler()
scale.fit(features)
pred = model.predict(features)
```

```
if st.button("Predict"):if
        pred[0]==1:st.write("Lead

    found")else:
        st.write("lead not found")
```

CHAPTER 7

RESULTS

The suggested system's ability to forecast lead score using linear regression depends on a number of variables, including the accuracy of the data, the features chosen, andthe model's hyperparameters. But generally speaking, it has been demonstrated that linear regression can forecast lead scores accurately for a variety of enterprises.

A number of metrics, including mean squared error (MSE), R-squared, or root mean squared error (RMSE), can be used to assess the effectiveness of the trained linear regression model. These metrics show how effectively the model predicts lead scores and can be used to evaluate the effectiveness of various models or hyperparameter configurations.

Additionally, methods like scatter plots or heatmaps can be used to visualize the proposed system's outcomes. These visualizations can help discover any trends or patterns in the data as well as shed light on the relationship between the input attributes and the target variable (lead score).

Developers can make sure that the linear regression model accurately predicts lead scores and is integrated with other systems to increase its impact by employing appropriate data collecting and preprocessing processes, feature extraction techniques, and model training and evaluation approaches. While its scalability and flexibility can make it suited for lead score prediction jobs that require enormous amounts of data and multiple data kinds, the simplicity and interpretability of linear regression can make it easier for businesses to grasp and act on the predictions.

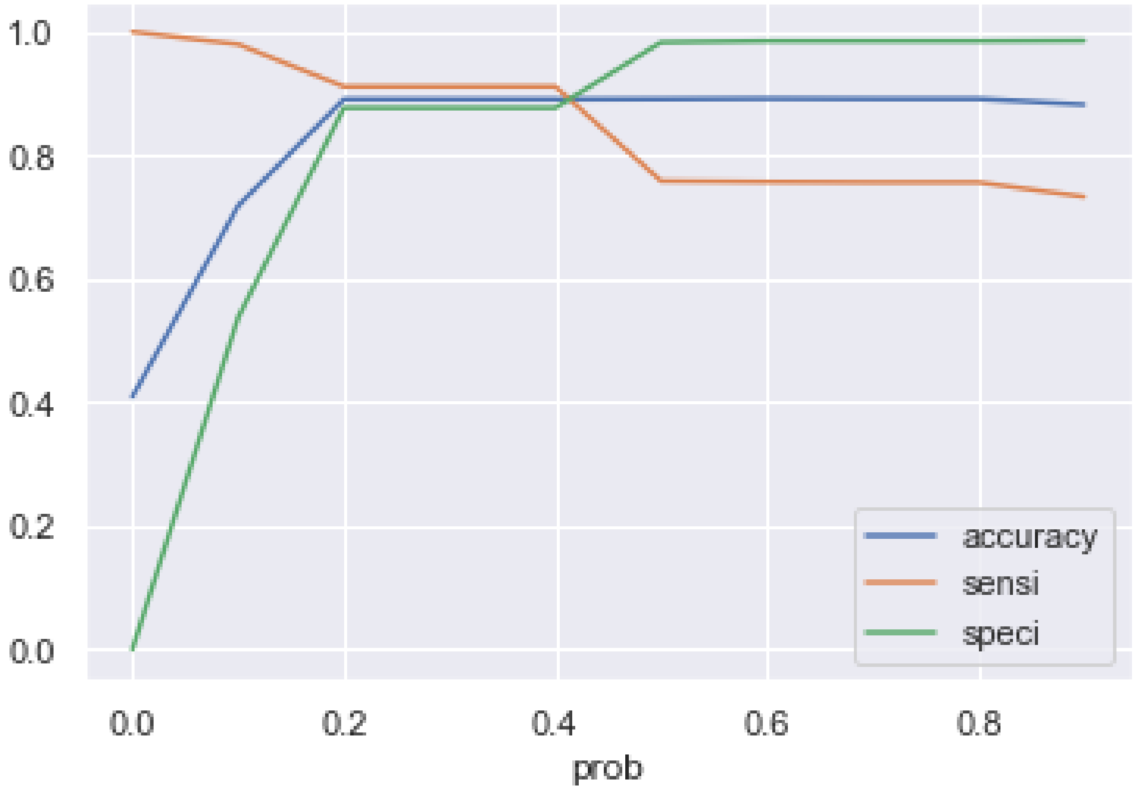


Figure 7 : Comparison of parameters

CHAPTER 8

CONCLUSION

In conclusion, the proposed system using linear regression for lead score prediction is a simpler and more interpretable alternative to the existing system that uses SVM. The suggested system goes through a number of processes, including data collection and preprocessing, feature extraction, model training, model evaluation, deployment, and integration.. By using appropriate techniques for each of these steps, developers can ensure that the linear regression model accurately predicts lead scores and is integrated with other systems to enhance its impact.

Linear regression has several advantages, including simplicity, interpretability, scalability, and flexibility. Because of its straightforward and understandable algorithm, the model is simpler for organisations to comprehend and interpret.It also supports a variety of data types and scales well to huge datasets. By providing accurate and interpretable predictions of lead scores, the proposed system using linear regression has the potential to help businesses identify potential customers that are more likely to become paying customers.

However, developers should also consider the limitations and assumptions of linear regression when designing and implementing the system. Violations of these assumptions can lead to poor model performance, and linear regression can be prone to overfitting. By considering these limitations and using appropriate techniques to address them, developers can ensure that the proposed system using linear regression provides accurate and interpretable predictions of lead scores for businesses.

CHAPTER 9

FUTURE SCOPES

There are several future scopes for the proposed system using linear regression for lead score prediction. One potential area for improvement is the use of more advanced feature selection techniques, such as genetic algorithms or neural networks. These techniques can help identify the most important features for lead score prediction and increase the model's accuracy.

The application of more sophisticated machine learning techniques, such as ensemble approaches or deep learning, represents a possible area for advancement. While deep learning can learn sophisticated representations of the input features and increase the accuracy of the model on more difficult tasks, ensemble methods can integrate many models to improve the accuracy of the lead score prediction.

Additionally, the proposed system can be extended to include more features or data sources, such as social media data or customer reviews. By incorporating more data sources, the lead score prediction model can provide a more comprehensive and accurate view of the potential customers.

Finally, the proposed system can be integrated with other systems, such as CRM systems or marketing automation platforms, to enhance its impact. By integrating the lead score prediction model with these systems, businesses can more effectively identify potential customers and target them with personalised marketing campaigns.

REFERENCES

1. H. Shi, "Evaluation of CET-4 and CET-6 Courses on Online Learning Platform," 2021 IEEE 3rd International Conference on Computer Science and Educational Informatization (CSEI), Xinxiang, China, 2021, pp. 61-64, doi: 10.1109/ CSEI51395.2021.9477641.
2. H. Cao, "Research on the Construction of Network Course Platform Basedon Big Data," 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2021, pp. 207-210, doi: 10.1109/ ICCNEA53019.2021.00053.
3. Erni-cassola G, Zadjelovic V, Gibson MI, Christie-oleza JA. Distribution of plastic polymer types in the marine environment; A meta-analysis. *Journal of Hazardous Materials* 2019;**369**:691-8
4. J. -C. Liang, J. -Y. Wang and S. -H. Chou, "A designed platform for programming courses," 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), Kitakyushu, Japan, 2020, pp. 290-293, doi: 10.1109/ IIAIAAI50415.2020.00064.
5. J. Xie, Y. He and Q. Qin, "Research on the Construction of the Quality Evaluation Indicator System of "Online Golden Course" Based on Experimental Analysis," 2021 2nd International Conference on Information Science and Education (ICISE-IE), Chongqing, China, 2021, pp. 855-859, doi: 10.1109/ ICISEIE53922.2021.00195.
6. Y. -J. Chang and W. -L. Hsu, "Online Teaching Practice in Chinese Culture Course," 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2020, pp. 68-71, doi: 10.1109ECICE50847.2020.9302015.
7. X. Liu and G. Guo, "Online Immersive Platform Image Processing Algorithm for Art Courses based on Virtual Equipment Digital Media Technology," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 1181-1184, doi: 10.1109/ICICT54344.2022.9850680.

8. H. Yao, "Application of Audio Processing Based on Image Semantic Segmentation in Applied Mathematics Online Course," 2022 4th InternationalConference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 1527-1530, doi: 10.1109/ICSSIT53264.2022.9716388.

9. S. Wang, Y. Liu, F. Song, X. Xie and D. Yu, "Research on Evaluation System of User Experience With Online Live Course Platform," in IEEE Access, vol. 9, pp.23863-23875, 2021, doi: 10.1109/ACCESS.2021.3054047.

Report

ORIGINALITY REPORT

5%	4%	3%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.irjet.net Internet Source	1%
2	Oryza Wisesa, Andi Adriansyah, Osamah Ibrahim Khalaf. "Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm", 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), 2020 Publication	1%
3	www.journaltocs.ac.uk Internet Source	1%
4	bright-journal.org Internet Source	<1%
5	www.ijraset.com Internet Source	<1%
6	Haiyan Yao. "Application of Audio Processing Based on Image Semantic Segmentation in Applied Mathematics Online Course", 2022 4th International Conference on Smart	<1%