

**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**  
JNANA SANGAMA, BELAGAVI – 590 018, KARNATAKA, INDIA



# **INTERNSHIP REPORT**

**30/08/2022 – 30/09/2022**

**Thaniya Technologies**

**Submitted in partial fulfillment of the requirements for the award of  
BACHELOR OF ENGINEERING  
in  
COMPUTER SCIENCE & ENGINEERING**

**Name**  
**DAYANANDA A**

**Submitted By**

**USN**  
**4VP20CS402**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**  
**VIVEKANANDA COLLEGE OF ENGINEERING & TECHNOLOGY**

[A Unit of Vivekananda Vidyavardhaka Sangha, Puttur (R)]  
Affiliated to Visvesvaraya Technological University and Approved by AICTE New Delhi & Govt. of Karnataka  
Nehru Nagara, Puttur – 574 203, DK, Karnataka, India

**APRIL 2023**

# VIVEKANANDA COLLEGE OF ENGINEERING & TECHNOLOGY

[A Unit of Vivekananda Vidyavardhaka Sangha, Puttur (R)]

Affiliated to Visvesvaraya Technological University and Approved by AICTE New Delhi & Govt. of Karnataka

Nehru Nagara, Puttur – 574 203, DK, Karnataka, India

## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING



### CERTIFICATE

Certified that the Internship/ Professional Practice, carried out by **Mr. Dayananda A**, bearing USN **4VP20CS402** a bonafide student of **Vivekananda College of Engineering & Technology, Puttur**, in partial fulfillment for the award of **Bachelor of Engineering in Computer Science & Engineering** of the **Visvesvaraya Technological University, Belagavi** during the year 2022-23. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library.

The Internship report has been approved as it satisfies the academic requirements in respect of Internship/ Professional Practice (18CSI85) prescribed for the said Degree.

\_\_\_\_\_  
Signature of the Internship  
Guide  
Prof. Shrinidhi A

\_\_\_\_\_  
Signature of the Internship  
Coordinator  
Prof. Thapaswini P S

\_\_\_\_\_  
Signature of the HOD  
Prof. Krishna Mohana A J

\_\_\_\_\_  
Signature of the Principal  
Dr. Mahesh Prasanna K

### EXTERNAL VIVA

Name of the Examiners

Signature with date

1.....

.....

2.....

.....

## INTERNSHIP COMPLETION CERTIFICATE

To whom it may concern

Thaniya Technologies certifies that **Mr. Dayananda A (4VP20CS402)** student of **VIVEKANANDA COLLEGE OF ENGINEERING AND TECHNOLOGY, PUTTUR** had undertaken and successfully completed the Internship on **Machine Learning** from **30 August 2022 till 30 September 2022**.

**Mr. Dayananda A** displayed professional traits during his internship period and managed to complete all assigned tasks as requested. He was hardworking, dedicated, and committed. It was a pleasure having him with us in this short period.

We wish him every success in life.



Sincerely,

**SRINIVAS**

**Project manager**



Shri Kalyani Opp Amara Jyothi Clinic SadanandaKamath  
Road Karkala, Karnataka, India



+91 7019582399



contact@thaniyatech.com  
info@thaniyatech.com



## EXECUTIVE SUMMARY & ACKNOWLEDGEMENT

For five weeks from 30/08/2022 to 30/09/2022, I did an internship at Thaniya Technologies. Thaniya Technologies is an online training and internship platform that offers a specialized training in various domains and provide internship in 4+ domains like Block chain, Machine Learning, Embedded System & IOT, Full Stack Development and more.

I am very appreciated to **Mr. Nishanth**, my mentors in Thaniya technologies for topics involving Statistic and Machine Learning. He guided me to deal with real time problem using machine learning models. He gave in-time feedback on my work and helped to present my ideas and achievements.

I would like to express my sincere gratitude to the internship guide **Prof. Shrinidhi A** Assistant Professor, Department of Computer Science & Engineering for his guidance and inspiration.

I would like to express my sincere gratitude towards our internship coordinator **Prof. Thapaswini P S**, Assistant Professor, Department of Computer Science & Engineering for providing excellent guidance, encouragement and inspiration throughout the Internship.

I would like to express my sincere gratitude to **Prof. Krishna Mohana A J**, Head, Department of Computer Science & Engineering for his guidance and inspiration.

I would like to thank our Principal **Dr. Mahesh Prasanna K** for providing all the facilities and a proper environment to work in the college campus.

I am thankful to all the teaching and non-teaching staff members of Computer Science & Engineering Department for their help and needed support rendered throughout the internship.

# TABLE OF CONTENTS

<b>Title</b>	<b>Page No.</b>
<b>List of Figures</b>	<b>I</b>
<b>CHAPTER 1 INTRODUCTION</b>	<b>01</b>
1.1 Introduction to Machine Learning	01
1.1.1 Input dataset	01
1.2 About the Company	03
1.3 Company Profile	03
<b>CHAPTER 2 SYSTEM STUDY AND ANALYSIS</b>	<b>04</b>
2.1 Introduction to Project	04
2.2 System Requirements	05
2.3.1 Functional Requirements	05
2.3.2 Non-Functional Requirements	06
2.3 Software Requirements	06
2.4 Hardware Requirements	07
2.5 Design	07
2.5.1 Flow Chart Diagram	07
2.5.2 System Architecture	08
<b>CHAPTER 3 PROJECT IMPLEMENTATION</b>	<b>10</b>
3.1 Introduction	09
3.2 Algorithm Used	09
3.2.1 Workflow	11
3.3 Pseudo Codes	11
3.3.1 Logistic regression	11
3.3.2 Decision Tree Regressor	12
3.3.3 Random Forest Regressor	12
3.4 Results/Snapshots	13
<b>CHAPTER 4 ASSESSMENT &amp; CONCLUSION</b>	<b>15</b>
4.1 Assessment of Internship	15
4.2 Conclusion	16
<b>REFERENCES</b>	<b>16</b>

# LIST OF FIGURES

Figure No.	Title	Page No.
2.1	Flowchart of the System	07
2.2	System Architecture of the System	08
3.1	Reading Data from Dataset	13
3.2	Understanding Data Using Bar Graph	13
3.3	Train-Test-Split Data	14
3.4	Implementing Logistic Regression	14
3.5	Implementing Decision Tree Regressor and Random Forest Regressor	14

# INTRODUCTION

## CHAPTER 1

### INTRODUCTION

#### 1.1 Introduction to Machine Learning

Machine Learning is a subfield of Artificial Intelligence (AI) that involves training computers or machines to learn from data and make predictions or decisions without being explicitly programmed. Machine Learning algorithms are designed to automatically learn from data and improve their performance over time by adjusting their internal parameters based on the data they are exposed to.

There are three main types of Machine Learning:

1. **Supervised Learning:** This involves training the algorithm on a labeled dataset, where the input data and the desired output are provided. The algorithm learns to predict the output based on the input data.
2. **Unsupervised Learning:** This involves training the algorithm on an unlabeled dataset, where the input data is provided without any desired output. The algorithm learns to find patterns and structure in the data without any explicit guidance.
3. **Reinforcement Learning:** This involves training the algorithm to make decisions in an environment, where the algorithm receives feedback in the form of rewards or penalties based on its actions.

Machine Learning has a wide range of applications in various fields, such as finance, healthcare, transportation, and entertainment. Some of the common applications include image recognition, natural language processing, predictive analytics, fraud detection, and recommendation systems.

##### 1.1.1 Input dataset

Dream Housing Finance company deals in all home loans. They have presence across all urban, semi urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan. Company wants to automate the loan eligibility process based on customer detail provided while filling online application form. These



details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. The dataset we will be working on has 615 rows & 13 columns. The Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. It's a classification problem , given information about the application we have to predict whether the they'll be to pay the loan or not. approved (Y/N)

Variable	Description
Loan_ID	Unique loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Dependents	Number of dependents
Education	Applicant Education (Graduate/ Under Graduate)
Self_Employed	Self employed (Y/N)
Applicant Income	Applicant income
Coapplicant Income	Co applicant income
Loan Amount	Loan amount in thousands
Loan_Amount_Term	Term of loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi Urban/ Rural
Loan_Status	Loan approved (Y/N)

## 1.2 About the Company

Thaniya Technologies is an IT solutions based company which specializes in performing custom design and development of projects which perfectly matches client business requirements. Their subsidiaries include Rooloo designs, Rooloo Fixme and Drony Aeriels. Their team uses cutting edge technologies to customize everything in a product development process whether it's about the look & feel feature of front-end to skilled back- end programming. Their customized methods and technologies result in full-functional, highly dynamic and interactive solution. They offer variety of services such as web design, web development, app development, machine learning, Internet of things, game development and event management. With a team of qualified professionals that are dedicated to delivering quality services for the success of client's projects [4].

### 1.2.1 Services

- **App Development:** Company develops Custom made Application for Business.
- **Graphics Design:** Logo, Creative Poster, Visiting Cards, brochure services are provided.
- **Web Development:** They build websites for online businesses. Design-Host-Maintain.
- **Aerial Videography:** Aerial Photography, Drone Photography/ Videography can be provided to required ones.
- **Digital Marketing:** It includes Search Engine Optimization, Translation and quality assurance of texts, Research work and keyword strategies.

## 1.3 Company Profile

**Company Name:** Thaniya Technologies, Karkala

**Office :** Shri Kalyani Sadananda kamath Road, opp. Amara Jyothi Clinic, Karkala,  
Karnataka 574104.

**Email:** info@thaniyatech.com

**Year stand up:** 2020.

**Company Category:** IT Software/Embedded

# **SYSTEM STUDY AND ANALYSIS**

## CHAPTER 2

### SYSTEM STUDY AND ANALYSIS

#### 2.1 Introduction to Project

Loan Distribution is the main business part of many banks. The main portion of banks income comes from the loan distributed to customers. These banks apply interest on loan which are distributed to customers. The main objective of banks is to invest their assets in safe customers. Up to now many banks are processing loans after regress process of verification and validation. But till now no bank can give surety that the customer who is chosen for loan application is safe or not. So to avoid this situation we introduced a system for the approval of bank loans known as Loan Prediction System Using Python. Loan Prediction System is a software which checks the eligibility of a particular customer who is capable of paying loan or not. This system checks various parameters such as customer's martial status, income, expenditure and various factors. This process is applied for many customers of trained data set. By considering these factors a required model is built. This model is applied on the test data set for getting required output. The output generated will be in the form of yes or no. Yes indicates that a particular customer is capable of paying loan and no indicates that the particular customer is not capable of paying loan. Based on these factors we can approve loans for customers. In India, peoples are highly applying for loans due to certain reasons. It is being hard for bank employees to check and predict whether the customer trustworthy to approve the loan with interest rate. The Data analysis helps to reduce the complexity of data and give appropriate results helpful for user. The data analysis technique used to analyze the behavior of data is quite exploratory. The main motive behind the paper is to identify the nature of loan applicant; Loan prediction system is very useful tool for the bank-employee and also to the customer. It reduces the risk factor. Loan prediction system provides results with the help of trained models for loan approval. The datasets used in purpose to train the models are the past records collected.

Loan approval prediction using machine learning has several potential benefits, including improving the efficiency of loan approval processes, reducing the risk of lending to uncreditworthy borrowers, and automating decision-making to reduce human bias.

## 2.2 System Requirements

System Requirements are the configurations that a system must have in order.

### 2.2.1 Functional Requirements

Functional requirements for loan approval prediction using machine learning typically include the following:

- 1 Data input: The model should be able to take in data related to the Bank details, such as loan ID, Gender ,Married ,Dependents ,Education ,Self-employed, Loan Amount
- 2 Data pre-processing: The model should pre-process the input data by cleaning it, converting it to numerical values, and scaling or normalizing the data.
- 3 Model Training: The system should be able to train machine learning algorithms on the collected and preprocessed data to create predictive models. This may involve selecting appropriate algorithms such as logistic regression, decision trees, random forests, support vector machines, or neural networks, and tuning their hyperparameters to optimize performance.
- 4 Prediction and Decision-Making: The system should be able to use the trained models to predict loan approval outcomes for new loan applications. This may involve inputting the applicant's information into the trained model and obtaining a predicted probability or binary classification (e.g., approved or denied) as the output.
- 5 Model Deployment: The system should be able to deploy the trained models into a production environment, where they can be used to predict loan approval outcomes for real-time loan applications.
- 6 Model Interpretability: The system should provide some level of model interpretability, allowing lenders to understand the factors that contribute to the loan approval decisions made by the machine learning models.
- 7 Monitoring and Maintenance: The system should include mechanisms for monitoring the performance and accuracy of the deployed models, and regular maintenance to ensure that the models continue to provide accurate predictions over time.

### **2.2.2 Non-Functional Requirements**

The non-functional requirements define how the system will perform certain task. They are also called as quality attributes.

- 1 Accuracy: The accuracy of the machine learning model is crucial for loan approval prediction as it directly affects the reliability of the decision-making process.
- 2 Performance: The performance of the machine learning model is important in terms of its speed and efficiency.
- 3 Scalability: The machine learning model should be scalable to accommodate varying levels of data and increasing workload. It should be able to handle large datasets, multiple concurrent requests, and increasing number of loan applications without a significant degradation in performance.
- 4 Robustness: The machine learning model should be robust and able to handle noisy or incomplete data, as loan application data may contain missing or inconsistent information.
- 5 Security: Loan application data can contain sensitive and confidential information.
- 6 Maintainability: The machine learning model should be designed in a way that makes it easy to maintain and update.

## **2.3 Software Requirements**

A Software requirements specification document describes the intended purpose, requirements, and nature of software to be developed.

1. Editor: Jupyter Notebook or Visual Studio Code.
2. Code Language: Python.

## **2.4 Hardware Requirements**

Hardware system requirements often specify the operating system version, processor type, memory size, available disk space and additional peripherals any, needed.

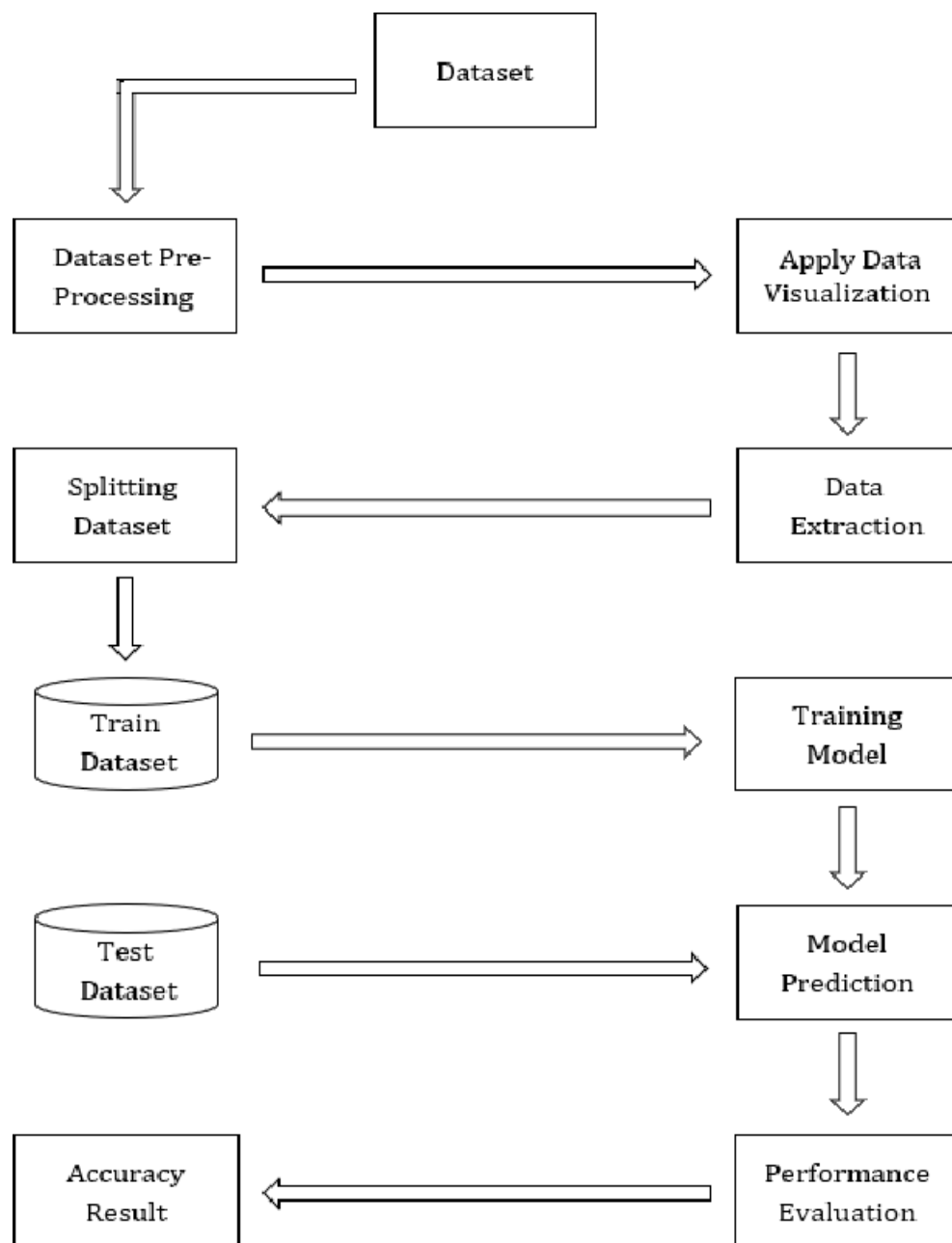
1. Processor: AMD Ryzen
2. Installed RAM: 8.00 GB (6.91 GB usable)
3. System type: 64-bit operating system, x64-based processor
4. Windows: Windows10

## 2.5 Design

Project design is an early phase of the project where a project's key features, structures, criteria for success. The major deliverables are all planned.

### 2.5.1 Flow Chart Diagram

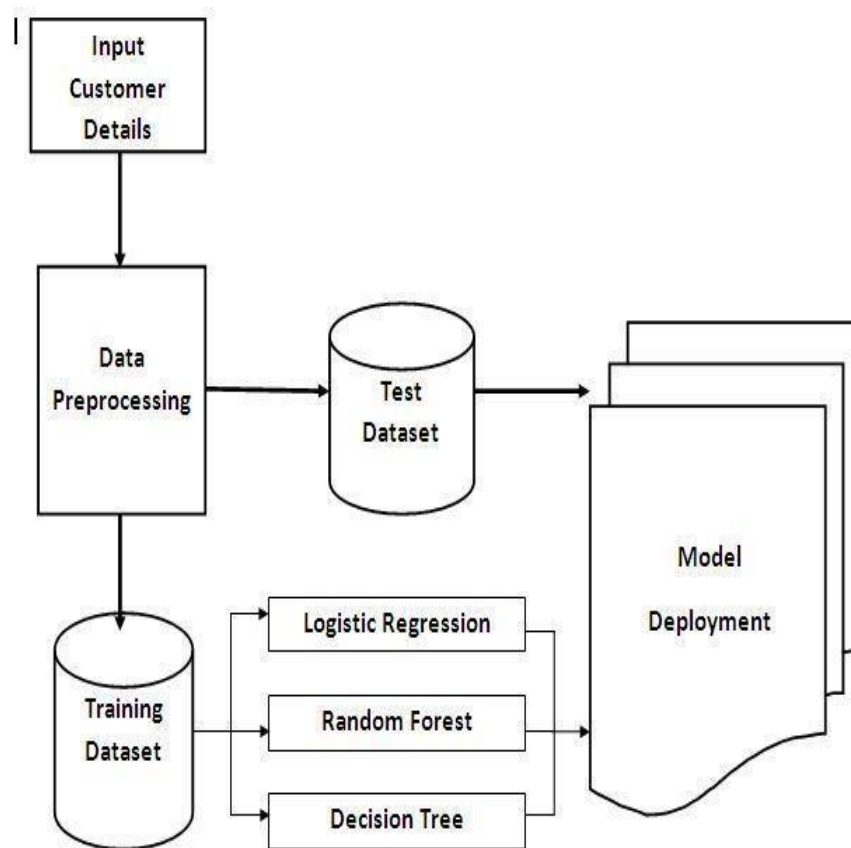
A flowchart is a diagram that depicts a process, system or computer algorithm. Figure 2.1 shows the flowchart of the system.



**Figure 2.1: Flowchart of the System**

## 2.5.2 System Architecture

System architecture refers to the overall structure and organization of a computer system or software application. It defines the various components of the system, how they interact with each other, and how data flows between them. The architecture also outlines the overall design philosophy and guiding principles of the system. Figure 2.2 describes overall system architecture of project.



**Figure 2.2: System Architecture of the System**



# **PROJECT IMPLEMENTATION**

## CHAPTER 3

# PROJECT IMPLEMENTATION

### 3.1 Introduction

Loan Distribution is the main business part of many banks. The main portion of banks income comes from the loan distributed to customers. These banks apply interest on loan which are distributed to customers. The main objective of banks is to invest their assets in safe customers. Up to now many banks are processing loans after regress process of verification and validation. But till now no bank can give surety that the customer who is chosen for loan application is safe or not. So to avoid this situation we introduced a system for the approval of bank loans known as Loan Prediction System Using Python. To achieve this, the project will involve various tasks such as data collection, data cleaning and pre-processing, exploratory data analysis, feature selection, model selection, model training, and model evaluation. The final output of the project will be a prediction model that can be used by advertisers to estimate the number of views that their YouTube ads will receive. This will help them make informed decisions on their marketing strategies and optimize their ad campaigns to achieve better results.

### 3.2 Algorithm Used

“Logistic regression” is a mathematical approach that used in the describing the relationship between some independent variable to a numerous dependent in variable or a dichotomous dependent. The regression function is a employed because of the proposed covariates are combination of continuous and categorical random variable , whereas the dependent variable default is a dichotomous.

#### Approach of Logistic Regression:

1. Data preparation: Clean and preprocess the data.
2. Model specification: Specify the form of the logic regression model.

3. Parameter estimation: Estimate the model parameters using a method such as ordinary least squares.
4. Model evaluation: Evaluate the performance of the model by checking the assumptions, examining the residuals for patterns or heteroscedasticity.

Decision Tree Regressor is a popular supervised machine learning algorithm used for regression analysis. It works by recursively partitioning the data into subsets based on the values of the independent variables and creating a decision tree that predicts the value of the dependent variable for new data points. Each internal node of the tree represents a decision based on the value of an independent variable, and each leaf node represents a prediction of the value of the dependent variable.

**Approach Of Decision Tree Regressor:**

1. The algorithm recursively repeats the splitting process for each subset until a stopping criterion is met.
2. Once the tree is built, new data points are classified by traversing the tree from the root node to a leaf node based on the values of the independent variables.
3. The predicted value for a new data point is then calculated as the average of the values of the dependent variable of the training data points that belong to the same leaf node as the new data point.

“Random Forest Regressor [5]” builds multiple decision trees and merge their predictions together to get a more accurate and stable prediction rather than relying on individual decision trees. Each tree in a random forest learns from a random sample of the training observations. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

**Approach of Random Forest Regressor:**

1. Pick at random K data points from the training set.
2. Build the decision tree associated with those K data points.
3. Choose the number of decision trees you want to build and repeat step 1 & 2.

### 3.2.1 Workflow

1. Data collection: Collecting and gathering data related to the about bank holder details, phone.. etc.
2. Data preparation: Preprocessing the data to clean it up, removing any missing or irrelevant data, and transforming the data if required.
3. Data analysis: Exploratory data analysis to understand the correlation between the different features and identify any patterns.
4. Model selection: Selecting a suitable regression algorithm such as Logistic Regression, Decision Tree Regressor, or Support Vector Regressor that can best fit the data and make accurate predictions.
5. Model training: Training the model on the training data and tuning the hyper parameters for optimal performance.
6. Model evaluation: Evaluating the model's performance on the test data using metrics such as mean squared error, mean absolute error, and R-squared.
7. Model deployment: Deploying the model in a production environment, where it can make real-time predictions on new data points.

## 3.3 Pseudo Codes

Pseudo code is an informal high-level description of the operative principles of the algorithm and computer programs.

### 3.3.1 Logistic Regression

The pseudo code for the Logistic Regression [1] is shown below.

```
# Logistic Regression
from sklearn.linear_model import LogisticRegression

model = LogisticRegression()

classify(model, X, y)
```

### 3.3.1 Decision Tree Regressor

The pseudo code for the Decision Tree Regressor [4] is shown below.

```
# Decision Tree Regressor

from sklearn.tree import DecisionTreeRegressor

decision_tree = DecisionTreeRegressor()

decision_tree.fit(X_train, y_train)
```

### 3.3.1 Random Forest Regressor

The pseudo code for the Random Forest Regressor is shown below.

```
# Random Forest Regressor

from sklearn.ensemble import RandomForestRegressor

n_estimators = 200

max_depth = 25

min_samples_split=15

min_samples_leaf=2

random_forest = RandomForestRegressor(n_estimators = n_estimators, max_depth =
max_depth, min_samples_split=min_samples_split)

random_forest.fit(X_train,y_train)
```

### 3.1 Results/Snapshots

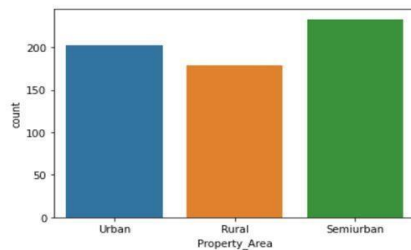
Figure 3.1 shows importing modules and reading dataset of loan approval prediction.



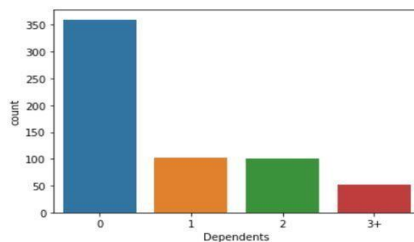
**Figure 3.1: Reading the Data from Dataset**

Figure 3.2 shows the Exploratory Data Analysis.

```
In [40]: sns.countplot(df['Property_Area'])
Out[40]: <AxesSubplot:xlabel='Property_Area', ylabel='count'>
```



```
In [38]: sns.countplot(df['Dependents'])
Out[38]: <AxesSubplot:xlabel='Dependents', ylabel='count'>
```



**Figure 3.2: Understanding Data Using Bar graph**

Figure 3.3 shows the train-test split

```
In [46]: ##Train-Test Split specify input and output attributes
X = df.drop(columns=['Loan_Status'], axis=1)
y = df['Loan_Status']

In [47]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

In [56]: # classify function
from sklearn.model_selection import cross_val_score
def classify(model, x, y):
    x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
    model.fit(x_train, y_train)
    print("Accuracy is", model.score(x_test, y_test)*100)
    score = cross_val_score(model, x, y, cv=5)
    print("Cross validation is", np.mean(score)*100)
```

**Figure 3.3: Train-Test-Split the data**

Figure 3.4 shows implementing Logistic Regression .

```
In [57]: from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
classify(model, X, y)

Accuracy is 77.27272727272727
Cross validation is 80.9462881514061
```

**Figure 3.4: Implementing Logistic regression**

Figure 3.5 shows implementing Decision Tree Regressor and Random Forest Regressor.

```
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
classify(model, X, y)

Accuracy is 72.72727272727273
Cross validation is 73.94908703185392

from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
classify(model, X, y)

Accuracy is 73.37662337662337
Cross validation is 76.55204584832734
```

**Figure 3.5: Implementing Decision Tree Regressor and Random Forest Regressor**

# **ASSESSMENT & CONCLUSION**



## CHAPTER 4

### ASSESSMENT & CONCLUSION

#### 4.1 Assessment of Internship

- Understood various information about python modules that are used in machine learning
- Gained practical knowledge on machine learning concepts.
- Importance of statistical method in machine learning.
- Working in team: company assigned a project which helped in gaining team work quality such as communication skills, decision making report building and problem solving.
- Work ethics: Constantly keeping in touch with team lead or mentor helped me to build cooperate work ethics, such as emphasis on quality of work, time management. As well as professionalism.

#### 4.2 Conclusion

The internship helped in understanding various machine learning methods that can be used in solving any real-world problem. It helped me in using various tools and techniques that are utilized during the process of machine learning project work. From the beginning of the internship mentors were friendly and taught every concept that is necessary while working in project. The project is a humble venture to satisfy the concept that are learnt in internship. Several user-friendly coding has also adopted. The objective of software planning is to provide a frame work that enables the manger to make reasonable estimates made within a limited time frame at the beginning of the software project and should be updated regularly as the project progresses.

## REFERENCES

- [1]. “An Approach for Prediction of Loan Approval about Machine Learning Algorithm,” Mohammad Ahmad Sheikh, Proceedings of the International Conference on Electronics and Sustainable Communication Systems (2020).
- [2]. PidikitiSupriya, “Loan Prediction using Machine Learning Models,” International Journal of Engg. and Techniques, Mar-Apr 2019.
- [3]. Ashlesha Vaidya, “Predictive and probabilistic approach by logistic regression : prediction of loan approval,” Computer Science Engineering SRM University, Chennai July 3-5, 2017, IIT Delhi.
- [4]. “Prediction of Loan Approval using Machine Learning” Rajiv Kumar, Vinod Jain, Premsagar Sharma, International Journal of Advanced Science and Technology 2019.
- [5]. “Bank loan analysis using customer usage data: A big data approach using Hadoop,” 2017 2nd International conference on Telecommunication & Networks.