**Prediction of Survival**

**Prepared by**

SHAH YASHASVI

(16IT130)

**Under the supervision of**

Prof. Ravi patel

A Report Submitted to

Charotar University of Science and Technology

for Partial Fulfillment of the Requirements for the

Degree of Bachelor of Technology

in Information Technology

IT345 Software Group Project-II (5th sem)

**Submitted at**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**Chandubhai S. Patel Institute of Technology**

**At: Changa, Dist: Anand – 388421**

**October 2018**



**CERTIFICATE**

This is to certify that the report entitled “**Prediction of Survival**” is a bonafied work carried out by **Ms. Shah Yashasvi (16IT130)**under the guidance and supervision of **Prof. Ravi patel** for the subject **Software Group Project-II(IT345)** of **5th** Semester of Bachelor of Technology in **Information Technology** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate **herself**, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

|  |
| --- |
| Under supervision of,  Prof. Ravi patel  Assistant Professor  Dept. of Information Technology  CSPIT, Changa, Gujarat. |
| Prof. Parth Shah  Head & Associate Professor  Department of Information Technology  CSPIT, Changa, Gujarat. |

**Chandubhai S Patel Institute of Technology**

At: Changa, Ta. Petlad, Dist. Anand, PIN: 388 421. Gujarat

**Abstract**

The report highlights the works carried out during the entire duration of this semester. It is mainly focuses on Analysis of different machine learning Algorithms and select best algorithm for my dataset so that successful outcomes I can have. Prediction about survival of a person on titanic dataset gives a categorical value so that I need to apply classification algorithms which comes in supervised learning. To make data visualization I have used graph plotting and outcome also I have shown using graph so that we have better idea about model and it’s outcomes analysis using different algorithm is also plotted.

**Acknowledgement**

Knowledge in itself is a continuous process. At this moment of our substantial enhancement, We rarely find words to express our gratitude towards those who were constantly involved with us.

The completion of any inter disciplinary project depends upon coordination, cooperation and combined efforts of several resources of knowledge, creativity, skill, energy and time. The work being accomplished now, We feel our most sincere urge to recall and knowledge through these lines, trying our best to give full credit wherever it deserves.

We would like to thank our project guide **Mr. Ravi Patel** who advised and gave us moral support through the duration of our project. Without their constant encouragement we could not have been able to achieve what we have.

It’s our good fortune that we had support and well wishes of many. We are thankful to all and those names which have been forgotten to acknowledge here but contributions have not gone unnoticed.

With Sincere Regards,

16IT130

Shah Yashasvi

TABLE OF CONTENTS

* Abstract iii
* Acknowledgement iv
* Chapter 1 Introduction 1

1.1 Project Overview 2

1.2 Scope (what it can do and can’t do) 4

1.3 Objective 5

* Chapter 2 System Analysis 8

2.1 User Characteristics (Roles of users who is dealing with the system) 9

2.2 Tools & Technology (Hardware/ Software Requirements- Minimum

requirements to run the system) 12

* Chapter 3 System Design 17

3.1 Flow of System (Graphical representation of project) 17

3.2 Data Dictionary (Table & Relationship)/Diagrams 19

* Chapter 4 Implementation 21

4.1 Implementation Environment 21

4.2 Coding Standards (Sample code of module) 26

4.3 Snapshots of project 26

* Chapter 5 Constraints and Future Enhancement 28
* Chapter 6 Conclusion (Learning Outcome-In your own words) 29
* References 30

**LIST OF TABLES**

* Table 3.1 10
* Table 3.2 11

**LIST OF FIGURES**

* Fig 3.1 11
* Fig .1 14

**Chapter:1 Introduction**

* 1. **PROJECT OVERVIEW**

**Project Name:** Prediction of Survival

**Project Description:** Project is focuses on Analysis of different machine learning Algorithms and select best algorithm for my dataset so that successful outcomes I can have. Prediction about survival of a person on titanic dataset gives a categorical value so that I need to apply classification algorithms which comes in supervised learning. To make data visualization I have used graph plotting and outcome also I have shown using graph so that we have better idea about model and it’s outcomes analysis using different algorithm is also plotted. To identify accuracy and errors RMSE is used and from that accuracy is found. In this project I learn about numerical data manipulation with numpy library of python, for plotting data Matplotlib library and for dataset handling pandas library working of python.

* 1. **SCOPE**

This project can predict a person can survive or not according it’s features like Age, Sex, pclass, parch, sibsp for that first applying suitable algorithm for our dataset and output variable and train our machine so that it learns about given data and gives output for new data. It is better if we train machine using enough amount of data so that it can be predict for any type of value of our data.

* 1. **OBJECTIVE**
* To get information about survival of a person on the Titanic ship
* Learning of how machine learning algorithm works and implementation of that algorithms
* By applying various algorithm on Titanic dataset, finding of highest accuracy

**Chapter: 2 System Analysis**

**2.1 USER CHARACTERISTICS**

**Roles of user:** one who is running this project needs to know how to run this code and what will be the output is correct or not.

**2.2 TOOLS AND TECHNOLOGY**

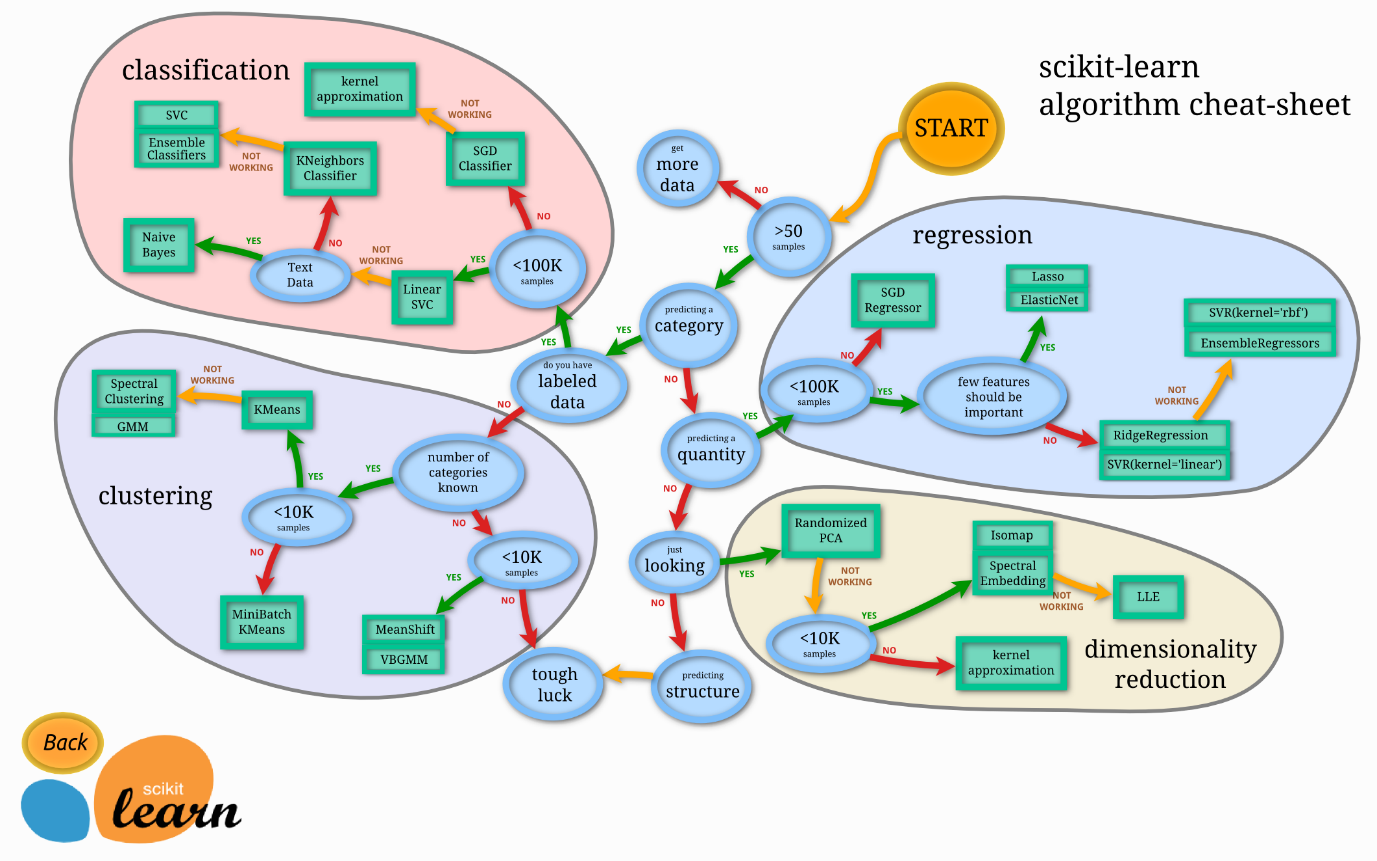
**Software Requirements:**

* Windows OS (for this piece of code)
* Python (version 3+)
* Titanic Dataset

**Chapter: 3 System Design**

**3.1 FLOW OF SYSTEM**

**Deciding Algorithm:**



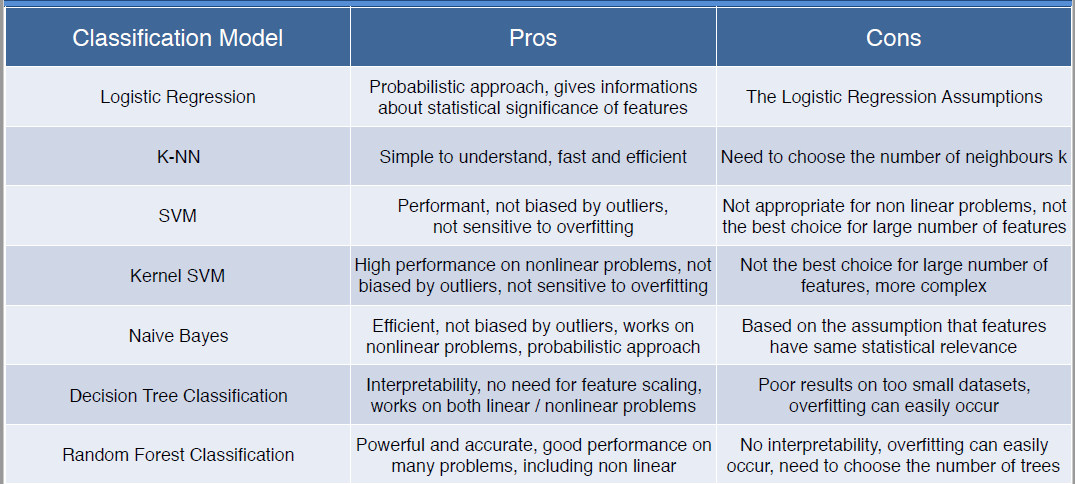


Table 3.1

Implementation is in the next chapter.

**3.2 DATA DICTIONARY(TABLE & RELATIONSHIP)/DIAGRAMS**

* **Accuracy and error of different classification Algorithms**

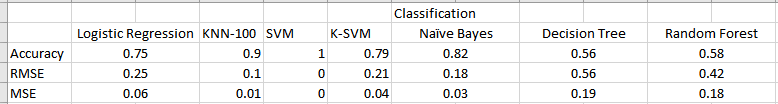


Table 3.2

* **Relationship Diagram of different Classification Algorithms**

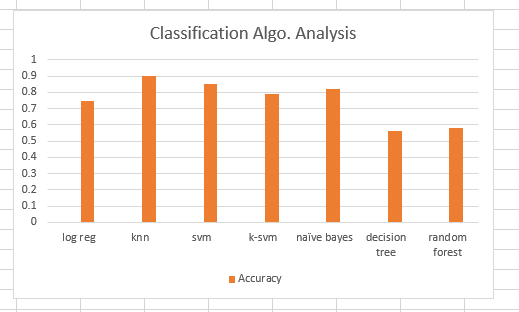


Figure 3.1

**Chapter: 4 Implementation**

**4.1 IMPLEMENTATION ENVIRONMENT**

* I have implemented my code in spyder IDE but there are so many options to implement python code as per user’s choice they can choose best for them like Anaconda, spyder, python IDLE, Jupyter Notebook, etc.
* **About Spyder IDE**

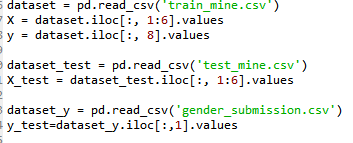
Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

Beyond its many built-in features, its abilities can be extended even further via its plugin system and API. Furthermore, Spyder can also be used as a PyQt5 extension library, allowing developers to build upon its functionality and embed its components, such as the interactive console, in their own PyQt software.

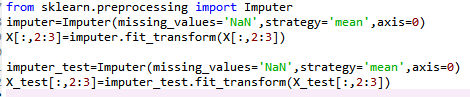
**4.2 CODING STANDARDS (SAMPLE CODE OF MODULE)**

**Code Snippets:**

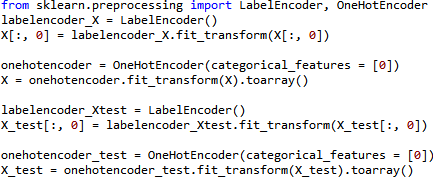
* Importing datasets:



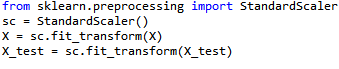
* Taking care of missing data:



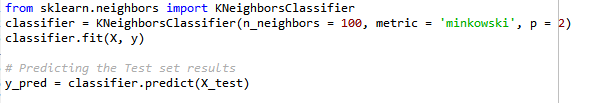
* Encoding categorical data:



* Feature Scaling of data:



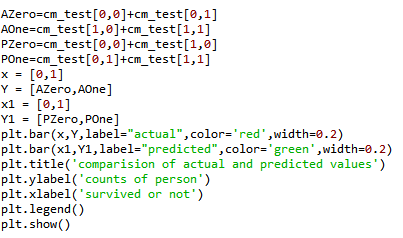
* Fitting Model and predicting test set results:



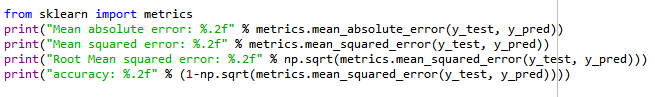
* Confusion Matrix:



* Plotting graph:

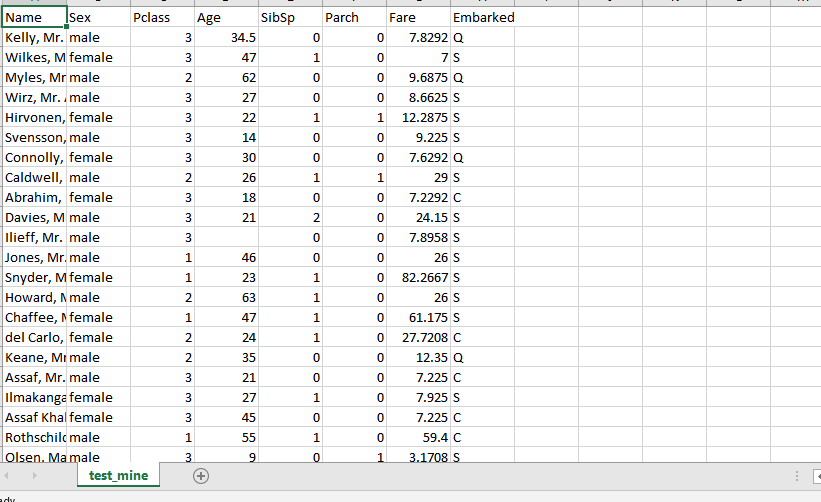


* Accuracy and Error:

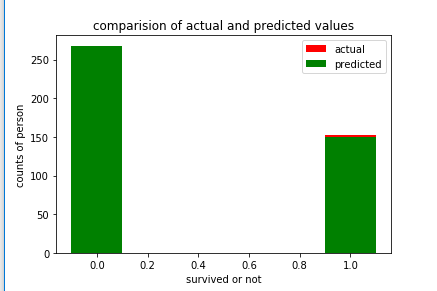


**4.3 SNAPSHOTS OF THE PROJECT**

* **Data in csv file:**

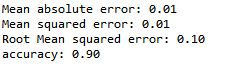


* **Graphical representation of actual and predicted values of survival**



**Figure 4.1**

* **Accuracy of model**



**Chapter : 5 Future Enhancement and Constraints**

**Future Enhancement:**

* Detect Outliers and analyze for various algorithms
* Improve Accuracy if Outliers exist
* Detection of Overfitting and Underfitting for various algorithms
* Improve Accuracy if model has overfitted or underfitted by it’s removal techniques
* Better Graphical Representation

**Constraints:**

* This is specifically limited prediction for Titanic dataset but we can also predict for different areas and requirements like this. As this is limited scope target it not contains more applications.

**Chapter : 6 Conclusion**

Using this small project one can have idea about how to analyze various algorithms and choose best from them suitable for their Objective and have idea about why k-nn algorithms is best for my dataset. Prediction about survival of a person can be done using this very efficiently using classification K-NN model and plotted output gives visualization for it.

**References**

* <https://campus.datacamp.com/courses/intermediate-python-for-data-science/matplotlib?ex=1>
* <https://docs.anaconda.com/anaconda/user-guide/getting-started>
* <https://www.superdatascience.com/machine-learning/>
* <https://www.youtube.com/watch?v=EvZvmFrjrus>
* <https://www.youtube.com/watch?v=73sHV_7sWmU>
* <https://www.datacamp.com/community/tutorials/kaggle-machine-learning-eda>
* [<https://www.kaggle.com/c/titanic/data>](https://www.kaggle.com/c/titanic/data)
* [<https://www.datacamp.com/courses/intro-to-python-for-data-science>](https://www.datacamp.com/courses/intro-to-python-for-data-science)
* <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>