##### Customer Churn Prediction Using Artificial Neural Network And Machine Learning

*A project report submitted in*

*The partial fulfilment of the requirements for the award of the degree of*

##### Bachelor of Technology in

**Computer Science & Systems Engineering**

###### Submitted by

R. SAI ROSHINI (19KD1A1552)

**P. HARSHITHA VARMA** **(19KD1A1549)**

G. AISHWARYA 19KD1A1521)

**M. RAM** **(19KD1A1541)**

###### Under the guidance of

Mrs. ANUPAMA

**Assistant Professor Department of CSIT**



##### Department of Computer Science & Engineering

LENDI INSTITUTE OF ENGINEERING & TECHNOLOGY

An Autonomous Institution, Accredited by NBA & NAAC with “A” Grade

Permanently Affiliated to JNTUK, Approved by A.I.C.T.E, Jonnada, Vizianagaram Dist.535005.

2019-2022



**LENDI INSTITUTE OF ENGINEERING &TECHNOLOGY**

**AN AUTONOMOUS INSTITUTION**

**Accredited by NAAC with “A” Grade, Accredited by NBA**

**Permanently Affiliated to JNTUK, Approved by A.I.C.T.E, Jonnada , Vizianagaram Dist - 535005.**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

## BONAFIDE CERTIFICATE

##### This is to certify that the project entitled “Customer Churn Prediction Using Artificial Neural Network and Machine Learning” is a Bonafide record of the work done by R. Sai Roshini (19KD1A1552),

##### P. Harshitha Varma (19KD1A1549), G. Aishwarya (19KD1A1521), M. Ram (19KD1A1541) under the supervision and guidance of Mrs.Anupama, Assistant Professor in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering from Lendi Institute of Engineering and Technology (Affiliated to JNTUK), Jonnada, Vizianagaram for the year 2023.

##### Internal Guide Head of the Department

Mrs. Anupama Dr. R. Rajendra

**Assistant Professor**

**Department of CSIT Department of CSSE**

**External Examiner**

## ACKNOWLEDGEMENT

With great solemnity and sincerity, we express our deepest sense of gratitude and pay our sincere thanks to our guide **Mrs. Anupama, Assistant Professor**, **Department of CSIT,** who evinced keen interest in our efforts and provided his valuable guidance throughout our project work.

We thank our **Prof. R. Rajendra**, Head of the Department of **Computer Science & Systems Engineering,** who helped us to complete our project work in a truthful method.

We thank our Mr.**S.Swaroop** , **Assistant Professor** of **Computer Science & Systems Engineering,** who helped us to complete our project work in a truthful method.

We thank our gratitude to our principal **Dr. V.V. Rama Reddy,** for his kind attention and valuable guidance to us throughout this course in carrying out the project.

We wish to express gratitude to our **Management Members** who supported us in providing good lab facilities.

We are thankful to **All Staff Members of the Department of Computer Science & Systems Engineering,** for helping us directly / indirectly to complete this project work by giving valuable suggestions.

All the above we greatly fully acknowledge and express our thanks to our parents who have been instrumental in the success of this project which plays a vital role.

R. SAI ROSHINI (19KD1A1552)

**P. HARSHITHA VARMA (19KD1A1549)**

G. AISHWARYA (19KD1A1521)

**M. RAM**  **(19KD1A1541)**

## DECLARATION

We hereby declare that the project work entitled “**Customer Churn Prediction Using Artificial Neural Network and Machine learning"** submitted to the JNTU Kakinada is a record of an original work done by **R. Sai Roshini (19KD1A1552), P. Harshitha Varma (19KD1A1549), G. Aishwarya (19KD1A1521), M. Ram (19KD1A1541)** under the esteemed guidance of **Mrs. Anupama, Assistant Professor**, Computer Science & Engineering, Lendi Institute of Engineering & Technology. This project work is submitted in the partial fulfilment of the requirements for the award of the **Bachelor of Technology in Computer Science & Systems Engineering.** This entire project is done to the best of our knowledge and is not submitted to any University for the award of degree.

R. SAI ROSHINI (19KD1A1552)

**P. HARSHITHA VARMA (19KD1A1549)**

G. AISHWARYA (19KD1A1521)

**M. RAM** **(19KD1A1541)**



**LENDI INSTITUTE OF ENGINEERING &TECHNOLOGY AN AUTONOMOUS INSTITUTION**

**Accredited by NAAC with “A” Grade, Accredited by NBA Permanently Affiliated to JNTUK, Approved by A.I.C.T.E, Jonnada, Vizianagaram Dist. - 535005.**

**DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING**

**VISION**

To be a frontier in computing technologies to produce globally competent computer science engineering graduates with moral values to build a vibrant society and nation.

MISSION

* Providing a strong theoretical and practical background in computer science engineering with an emphasis on software development.
* Inculcating professional behavior, strong ethical values, innovative research capabilities, and leadership abilities.
* Imparting the technical skills necessary for continued learning towards their professional growth and contribution to society and rural communities.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

**PEO-1:** Graduates will have strong knowledge and skills to comprehend the latest tools and techniques of Computer Engineering so that they can analyze, design and create computing products and solutions for real life problems.

**PEO-2:** Graduates shall have multidisciplinary approach, professional attitude and ethics, communication and teamwork skills, and an ability to relate and solve social issues through their Employment, Higher Studies and Research.

**PEO-3:** Graduates will engage in life-long learning and professional development to adapt to rapidly changing technology.

PROGRAM SPECIFIC OUTCOMES (PSOs)

**PSO-1:** Ability to grasp advanced programming techniques to solve contemporary issues.

**PSO-2:** Have knowledge and expertise to analyse data and networks using latest tools and technologies.

**PSO-3:** Qualify in national and international competitive examinations for successful higher studies and employment.

PROGRAM OUTCOMES (POS)

**PO-1** Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

**PO-2** Problem Analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**PO-3** Design/development of Solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations

**PO-4** Conduct Investigations of Complex Problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO-5** Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

**PO-6** The Engineer and Society: Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**PO-7** Environment and Sustainability: Understand the impact of professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**PO-8** Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.

**PO-9** Individual and Teamwork: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**PO-10** Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**PO-11** Project Management and Finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**PO-12** Life-Long Learning: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

**ABSTRACT**

Customer churn prediction has gathered greater interest in business, especially in the telecommunications industries. Many authors have presented different versions of the churn prediction models greatly based on the data mining concepts employing machine learning and meta-heuristic algorithms. The aim of this paper is to study some of the most important churn prediction techniques developed over the recent years. The primary objective is on the churn in telecom industries to accurately estimate the customer survival and customer hazard functions to gain the complete knowledge of churn over the customer tenure. Another objective is the identification of the customers who are at the blink of churn and approximating the time they will churn. This paper focuses on analyzing the churn prediction techniques to identify the churn behavior and validate the reasons for customer churn. This paper summarizes the churn prediction techniques to have a deeper understanding of the customer churn and it shows that most accurate churn prediction is given by the hybrid models rather than single algorithms so that telecom industries become aware of the needs of high-risk customers and enhance their services to overturn the churn decision.

Keywords- Telecome Retention, Churn Analysis , Exploratory Data Analysis, Telecustomer Dataset

Outcomes:

Our project titled ***“ customer churn prediction* ”**is mapped with the following outcomes:

**Program Outcomes :** PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8,

PO9, PO10, PO11, PO12

**Program Specific Outcomes :** PSO1, PSO2, PSO3.

**LIST OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No** | **Title** | **Page No** |
| **1** | **INTRODUCTION** | 13-14 |
| 1.1 Project Overview | 14 |
| 1.2 Project Deliverables | 14 |
| 1.3 Project Scope | 14 |
| **2** | **LITERATURE SURVEY** | 15-18 |
| **3** | **PROBLEM ANALYSIS** | 20-21 |
| 3.1 Existing System | 20 |
| 3.1.1 Challenges | 20 |
| 3.2 Proposed System | 20 |
| 3.2.1 Advantages | 21 |
| **4** | **SYSTEM ANALYSIS** | 22-27 |
| 4.1 Software Requirement Specification | 23 |
| 4.1.1 Functional Requirements | 24 |
| 4.1.2 Non - Functional Requirements | 24 |
| 4.2 Feasibility Study | 25 |
| 4.3 System Requirements | 26 |
| 4.3.1 Software Requirements | 26 |
| 4.3.2 Hardware Requirements | 27 |
| **5** | **SYSTEM DESIGN** | 28-41 |
| 5.1 Introduction | 29 |
| 5.1.1 System Architecture | 30 |
| 5.1.2 Use Case Diagram | 31-32 |
| 5.1.3 Activity Diagram | 33-35 |
| 5.1.4 Class Diagram | 36-38 |
| 5.1.5 Component Diagram | 39 |
| 5.1.6 Deployment Diagram | 39-41 |

|  |  |  |
| --- | --- | --- |
| **6** | **IMPLEMENTATION** | 43-57 |
| 6.1 Algorithm Description | 43 |
| 6.2 Technology Description | 47 |
| 6.2.1 Python | 48 |
| 6.2.2 Django Framework | 49 |
| 6.2.3 Html, Java script, CSS, j Query | 48 |
| 6.3 Sample Source Code | 54-57 |
| **7** | **TESTING** | 59-66 |
| 7.1 Introduction | 60 |
| 7.2 Test Cases | 55 |
|  | 7.3 Screenshot's | 58 |
| **8** | **OUTPUT SCREENS** | 68-71 |
| **9** | **CONCLUSION** | 72-73 |
| **10** | **REFERENCES** | 74-75 |
| **11** | **PUBLICATIONS (if ANY)** | - |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Fig No. | Name | Page no |
| 1 | System Architecture | 30 |
| 2 | Use case Diagram | 31-32 |
| 3 | Activity Diagram | 33-35 |
| 4 | Class Diagram | 36-38 |
| 5 | Component Diagram | 39 |
| 6 | Deployment Diagram | 39-41 |

**CHAPTER 1**

# INTRODUTION

# 

##### 1.INTRODUCTION

##### USECASE: TELECOME RETENTION

Churn prediction means detecting which customers are likely to leave a service or to cancel a subscription to a service. It is a critical prediction for many businesses because acquiring new clients often costs more than retaining existing ones. Once you can identify those customers that are at risk of cancelling, you should know exactly what marketing action to take for each individual customer to maximise the chances that the customer will remain.

Different customers exhibit different behaviours and preferences, so they cancel their subscriptions for various reasons. It is critical, therefore, to proactively communicate with each of them to retain them in your customer list. You need to know which marketing action will be the most effective for each customer, and when it will be most effective.

In the business world of telecommunication, Customer churn is a deliberate issue. Churn in the telecommunication industry refers to customers leaving their current service providers either due to better services or less costly services provided by the competing service providers. Customers have the right to choose from a vast number of service providers. So, the current telecommunication company faces a huge loss in revenue and profit. The companies employ data mining techniques to predict the customers who are most likely to change their service providers. Hence, by analysing these predictions, the companies can set up various techniques to retain their present subscribers intact. This is called Churn Prediction.

There are broadly two types of churners. These are voluntary churners and involuntary churners. Voluntary churners themselves initiate the service termination from their current service providers. These are further divided into deliberate and incidental churners. Incidental churners are those that change their service provider due to some incidents such as a change in location or change in financial position. Deliberate churners are the most impending churners which are of concern for companies. These churners deliberately move to other service providers due to changes in technology or price rate. Involuntary churners are those that are removed by the company from their subscribers' list due to some fraud.

**Project Overview:**  Our Customer churn prediction model overview is to train a machine learning model on the available data to train a machine learning model that will predict with a high accuracy which customers are about to churn, which in turn will help the business owner in making useful marketing decisions.

**Project Deliverables:** Customer churn analysis using machine learning Delivers companies with accurate forecasts of customer preferences: the key attributes that they are looking for in products / services, as well as the features that they are unsatisfied with. As a result, companies have important data that can be used for the optimization of an existing product or the creation of a new one. Using machine learning for customer churn prediction can also increase the company’s revenue. Improving the customer experience and better understanding their behavior and preferences automatically results in greater profits.

**Project Scope:** The Churn model can be developed across the spectrum of business. Other predictive models can be built to support both up streams and downstream analytics requirements.

**CHAPTER 2**

**LITERATURE SURVEY**

##### LITERATURE SURVEY

This section reviews papers based on customer churn prediction using machine learning techniques. Papers published in the last ten years are reviewed and analyzed based on the methodologies used. Kiran Dahiya and Surbhi Bhatia [1] go through churn analysis process. Data mining is a Knowledge Discovery Process where new insights are evolved from a large set of data. Churn stands for customers switching from one service provider to another. Service can be telecommunication, banking, etc. Churn analysis stands for analyzing the customers for their churn probability. The analysis consists of stages such as data acquisition, Data preparation, Data pre-processing, Data Extraction and Decision. The model is implemented using the decision tree and logistic regression. The insights found are passed to CRM to take necessary actions to reduce the churn rate. Out of the two implementations, the decision tree shows more accuracy and efficiency. Navid Forhad, Md. Shahriar Hussain, Rashedur M Rahman [2] discusses the process of predicting churners. This paper explains input data collection, methods of analysis, filtering and rule generation. The result is generated based on the frequency of bill payment. Later, possible obstacles while doing churn analysis are also discussed. The paper comes to an end with the view that a complete dataset is necessary for churn prediction with accuracy. A comparison of methods and mining tools are other areas to consider.

Nadeem Ahmad Naz, Umar Shoaib and M. Shahzad Sarfraz [3] focus on explaining the concept churn prediction. The main objective of services like telecommunication is to retain existing customers than attaining new ones due to saturation. Data mining techniques can be used for this customer churn analysis. The phases of data mining involve selection, pre-processing, data mining, transformation, and evaluation. Churners can be classified as voluntary and involuntary whereas voluntary can be either deliberate or incidental. Out of this, we apply data mining techniques mainly for the analysis of deliberate churners. There are various modeling techniques. We choose among them based on the objective of analysis. Decision Tree and Support Vector Machine is used to find true churn rate and false churn rate while logistic regression is used to find churn probability. Mitkees, Ibrahim & Badr, Sherif & Elseddawy, Ahmed [4] used data from the IBM Watson Analytics community and performed data mining phases. The data mining techniques include a classification for prediction and clustering and association for detection. Classification is done with the aid of Matlab software which uses the decision tree, logistic regression, and SVM. For clustering, the author uses k-means and DB-scan algorithms. Association rule is implemented using Apriori-type and FP-Growth algorithms. They are applied to Weka mining software. The efficiency of algorithms is represented using tables and graphs. David L.Garcia, Alfredo Vellido, and Angela Nebot [5] describe the design and development of the predictive model as 4 stages. First, one is identifying and obtaining the best data. As said, the data used determines the accuracy and efficiency of prediction. The second stage is the selection of attributes. It has two phases such as the search phase and evaluation phase. The data used for both phases should be different to avoid overfitting. The development of a predictive model is the third phase. For that, standard methods such as Decision tree and Regression analysis can be used or other soft computing methods. In the fourth stage, validation of the result is carried out by dividing the whole data into 70:30 ratios, where 70% of data is used for training the model and the rest 30% used for testing. N. L. R. Machado and D Ruiz [6] proposes a churn prediction method based on mobile application usage. It is based on the interest of companies to understand customers' usage of their application. To analyze customers’ behaviors, they are grouped based on their activity patterns. Prediction is carried out with algorithms such as K-Means, STREAM, and DBSCAN. Among these, K-Means presents the best performance. Further tests performed show an accuracy rate of 87% which is satisfactory. Abinash Mishra and U. Srinivasulu Reddy [7] compare ensemble-based classifiers with well-known classifiers. Also utilized those ensemble-based classifiers such as Bagging, Boosting and Random Forest for customer churn prediction in the telecom industry. The classifiers are compared based on error rate, specificity, sensitivity, and accuracy. Among all classifiers, Random Forest performs better. They also compared the performance of Random Forest with an existing churn prediction model called Classification and Regression Tree (CART) and its variants. Then Random Forest has a high-performance rate. G. Xia, H. Wang Jiang [8] set up the model based on the characteristics of amount and imbalance data and verify on the real data of telecom. By comparing with the Bayes, Decision Tree (DT), Artificial Neural Networks (ANN) and Support Vector Machine (SVM), the ensemble learning algorithms have the potential advantages. The effect of ensemble is obvious advantage especially the base classifiers are Support Vector Machines and has better hit rate, lift coefficient and accuracy rate. It can be used as an effective measure for customer churn prediction. The experiments are conducted in the environment based on MATLAB 2011a, DTree used classical C4.5, Bayes use Naive Bayesian, and ANN model used BP algorithm consists of a single hidden layer. P.K.D.N.M. Alwis, B.T.G.S. Kumara, H.A.C.S. Hapuarachchi [9] created a predictive churn model that obtains customer churn rate of five telecommunication companies. For model building, we classified the relevant variables with the use of the Pearson chi-square test, cluster analysis, and association rule mining. The C5.0 Decision tree algorithm tree, the Bayesian Network algorithm, the Logistic Regression algorithm, and the Neural Network algorithms were developed. The C5.0 algorithm of decision trees model proved optimal among the models with 85 percent accuracy. Kamya Eria and Booma Poolan Marikannan [10] identifies that Support Vector Machines, Naive Bayes, Decision Trees and Neural Networks are the most used CCP techniques. Feature selection is the most used data preparation method followed by Normalization and Noise removal. Support Vector Machines and Neural Networks as the most preferred prediction techniques. However, ensembles of these techniques improve the prediction accuracy of the models because of the combined advantages of the components.

**CHAPTER 3**

**PROBLEM ANALYSIS**

##### PROBLEM ANALYSIS

* 1. EXISTING SYSTEM:

In the Existing system, to improve the quality of a prediction system, a Hybrid approach and comparative results have been shown which depicts that the proposed approach shows an improvement in the quality and scalability of the prediction system.

3.1.1 CHALLENGES:

* + Inaccurate or messy customer data,
  + Weak attrition exploratory analysis,
  + Lack of information and domain knowledge,
  + Lack of a coherent selection of a suitable churn modeling approach,
  + Choice of metrics to validate churn model performance,
  + Line of Business (Lob) of services or products,
  + Churn event censorship,
  + Concept drift based on changes in customers behavior patterns driving churn,
  + Imbalance data (class imbalance issue).

PROPOSED SYSTEM:

* + 1. In this Proposing system, we are mainly focusing on a system that focus on the prediction problems, so independent variables are taken from the data of the current period The logistic regression algorithm produces a better prediction effect, based on which the level of importance of customer churn factors can be seen. In this part, the logistic regression model is used to predict the trend in customer churn, assist enterprises in finding out the early warning signals of customer churn, and determine the tendency of customer churn.
    2. ADVANTAGES:
       - In this Proposed System, Identify at-risk customers
       - Optimization products and services.
       - Increased revenue.
       - Determining the data mining methods frequently used in churn implementations.

### 

### 

### 

### 

### **CHAPTER 4**

#### SYSTEM ANALYSIS

##### 4.SYSTEM ANALYSIS

System analysis is a problem-solving technique that decomposes a system into its component pieces for the purpose of studying how well those component parts work and interact to accomplish their purpose. System analysis is the process of studying a procedure in order to identify its goals and purposes and create systems and procedures that will achieve them in an efficient way.

The development of a computer-based information system includes a systems analysis phase which produces or enhances the data-model which itself is a precursor to creating or enhancing a database. There are several different approaches to system analysis. When a computer-based information system is developed, systems analysis would constitute the following steps:

* The development of a feasibility study, involving determining whether a project is economically, socially, technologically and organizationally feasible.
* Conducting fact-finding measures, designed to ascertain the requirements of the system’s end-users. These typically span interviews, questionnaires, or visual observations of work on the existing system.
* Gauging how the end-users would operate the system (in terms of general experience in using computer hardware or software), what the system would be used for and so on.

##### Software Requirements Specifications

A software requirements specification (SRS) is a description of a software system to be developed. It lays out functional and non-functional requirements and may include a set of use cases that describe user interactions that the software must provide. Software requirements specification establishes the basis for an agreement between customers and contractors or suppliers on what the software product is to do as well as what it is not expected to do. Software requirements specifications permit a rigorous assessment of requirements before design can begin and reduces later redesign. It should also provide a realistic basis for estimating product

costs, risks, and schedules. Used appropriately, software requirements specifications can help prevent software project failure.

* + - CPU >= Pentium 1.5 Ghz
    - Operating System **>=** Windows 8/Linux

##### Functional Requirements

Functional requirements define what a system is supposed to do. Functional requirements are usually in the form of system shall do, an individual action of part of the system, perhaps explicitly in the sense of a mathematical function, a black box description input, output, process and control functional model.

* + - * The system needs a Data set of Telecommunication sector.
      * This model predicts the churners.
      * Will provide offers to the customers.

##### Non-Functional Requirements

Non-functional requirements define how a system is supposed to be. Non-functional requirements that specify criteria that can be used to judge the operation of a system, rather than specific behavior. Non-functional requirements are in the form of an overall property of the system as a whole or of a particular aspect and not a specific function.

**Performance:** The system efficiently processes the input and predicts the customers.

**Scalability:** The system must be able to provide instant recommendations to many users at any given time.

**Portability:** It is easily portable and compatible with all operating systems.

**Accuracy: It** is based on Quality.

**Reusability:** It uses some of the pre-existing videos and images from the internet.

##### Feasibility study

The feasibility study is an evaluation and analysis of the potential of a proposed project. It is based on extensive investigation and research to support the process of decision making. Feasibility studies aim to objectively and rationally uncover the strengths and weaknesses of an existing or proposed system, opportunities and threats present in the environment, the resources required to carry through, and ultimately the prospects for success. In its simplest terms, the two criteria to judge feasibility are cost required and value to be attained. A well- designed feasibility study should provide a historical background of a project, a description of a service and details of the operations. Generally, feasibility studies precede technical development and project implementation. A feasibility study evaluates the project’s potential for success. It must therefore be conducted with an objective, unbiased approach to provide information upon which decisions can be based.

Scalability:

Ability to process huge amounts of data and train the machine learning model.

Reliability:

It is an efficient way of processing data and classifying into a correct category.

Technical Feasibility:

Technical feasibility also involves evaluation of the hardware and the software requirements of the proposed system.

Economic Feasibility:

It serves as an independent project assessment and enhances project credibility. Economic feasibility is an assessment which typically involves cost/ benefits from the analysis of the project.

Operational Feasibility:

It measures how well the proposed system solves problems and takes advantage of the opportunities identified during scope definition. Operational feasibility studies analyze how the project plan satisfies the requirements identified in the requirements analysis phase of system development. To ensure success, desired operational outcomes must inform and guide design and development. These include design-dependent parameters such as reliability, maintainability, supportability, usability, disposability, sustainability, affordability and others.

##### Benefits of Conducting a Feasibility Study:

Conducting a feasibility study is always beneficial to the project as it gives you and other stakeholders a clear picture of your idea. Below are the key benefits of conducting a feasibility study:

* + - * Gives project teams more focus and provides an alternative outline.
      * Narrows the business alternatives.
      * Identifies a valid reason to undertake the project
      * Enhances the success rate by evaluating multiple parameters.
      * Aids decision-making on the project.

##### System Requirements

System requirements specification is a detailed statement of the effects that a system is required to achieve. A good specification gives a complete statement of what the system is to do, without making any commitment as to how the system is to do it.

A system requirements specification is normally produced in response to a user requirements specification or other expression of requirements and is then used as the basis for system design. The system requirements specification typically differs from expression of requirements in both scope and precision. The latter may cover both the envisaged system and the environment in which it will operate but may leave many broad concepts unrefined.

##### Software Requirements

CPU **>=** Pentium 1.5 Ghz GPU **>=** And R5/Nvidia 980M

Operating System **>=** Windows 8/Linux

##### Hardware Requirements

* + - * Memory : 4GB
      * Processor : Dual Core
      * Network : 1GB Ethernet

# 

# CHAPTER 5

#### SYSTEM DESIGN

* 1. **SYSTEM DESIGN**

##### Introduction

System design is the process of defining the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through that system. It is meant to satisfy specific needs and requirements of a business or organization through the engineering of a coherent and well-running system.

System design mainly concentrates on defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development.

System design implies a systematic approach to the design of a system. It may take a bottom-up or top-down approach, but either way the process is systematic wherein it considers all related variables of the system that needs to be created—from the architecture to the required hardware and software, right down to the data and how it travels and transforms throughout its travel through the system. System design then overlaps with systems analysis, systems engineering and systems architecture.

The system design approach first appeared right before World War II, when engineers were trying to solve complex control and communications problems. They needed to be able to standardize their work into a formal discipline with proper methods, especially for new fields like information theory, operations research and computer science in general.

**FIG: 5.1.3A SYSTEM ARCHITECTURE DIAGRAM**

###### **UML Diagrams**

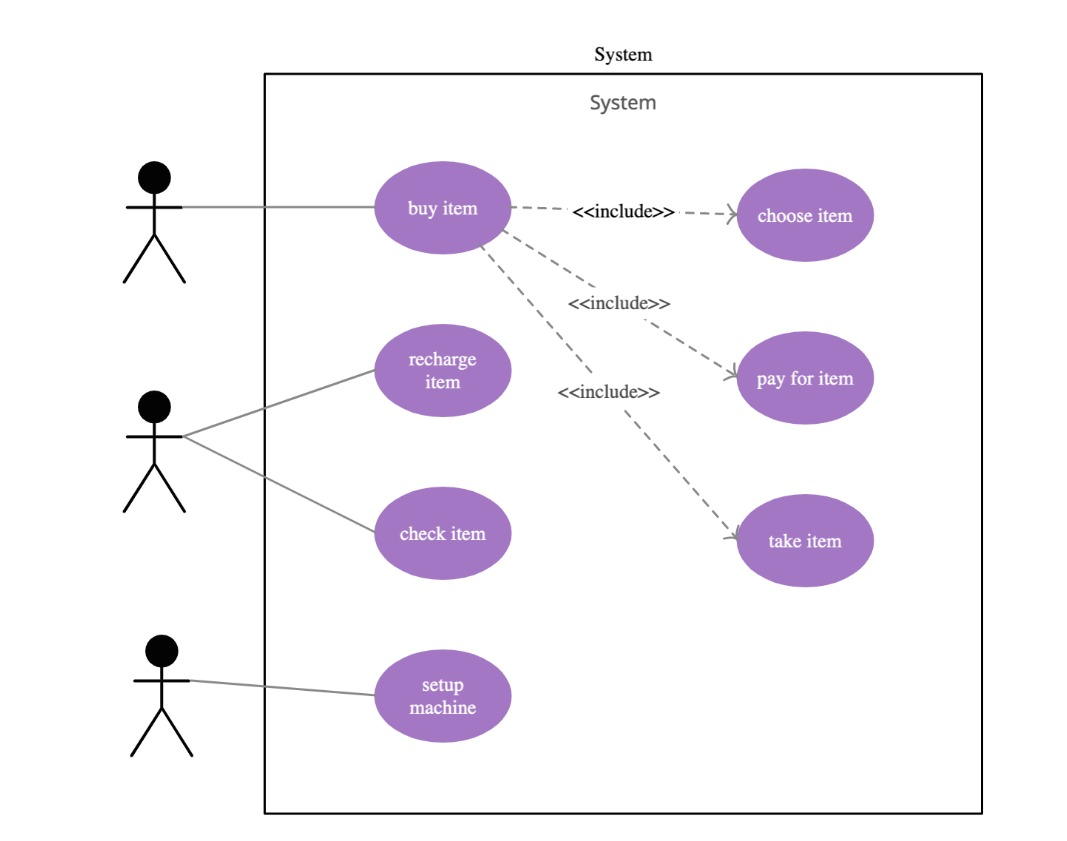
UML is a standard language for specifying, visualizing, constructing and documenting the Artifacts of software systems. The UML user mostly graphical notations to express the design of software project. It is a very important part of developing object-oriented software and the software development process. UML includes a set of graphic notation techniques to create visual models of object-oriented software-intensive systems.

##### Use Case Diagrams

Use case diagrams are usually referred to as behavior diagrams used to describe a set of actions that some systems should or can perform in collaboration with one or more external users of the system (actors). Each use case should provide some observable and valuable result to the actors or other stakeholders of the system.

Table: 5.1.1A Graphical Representations of Use case Diagrams

|  |  |  |
| --- | --- | --- |
| Actor | An actor in the Unified Modelling Language specifies a role played by a user or any other system that interacts with the subject. |  |
| Use case | A Use Case is the functionality provided by the system. Use Cases are depicted with an ellipse. The name of the Use Case is written in ellipse. |  |
| Association | Association is a relationship between classifiers which is used to show that instances of classifiers could be either linked to each other or combined logically or physically into some aggregation. |  |



**Fig 5.1.1B Graphical Notation of Use Case Diagram**

##### Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams are intended to model both computational and organizational processes (i.e., Workflows). Activity diagrams show the overall flow of control.

Activity diagrams are constructed from a limited number of shapes, connected with arrows. The most important shape types:

* + - * Rounded rectangles represent actions
      * Diamonds represent decisions
      * A black circle represents the start (initial state) of the workflow an encircled black circle represents the end (final state).

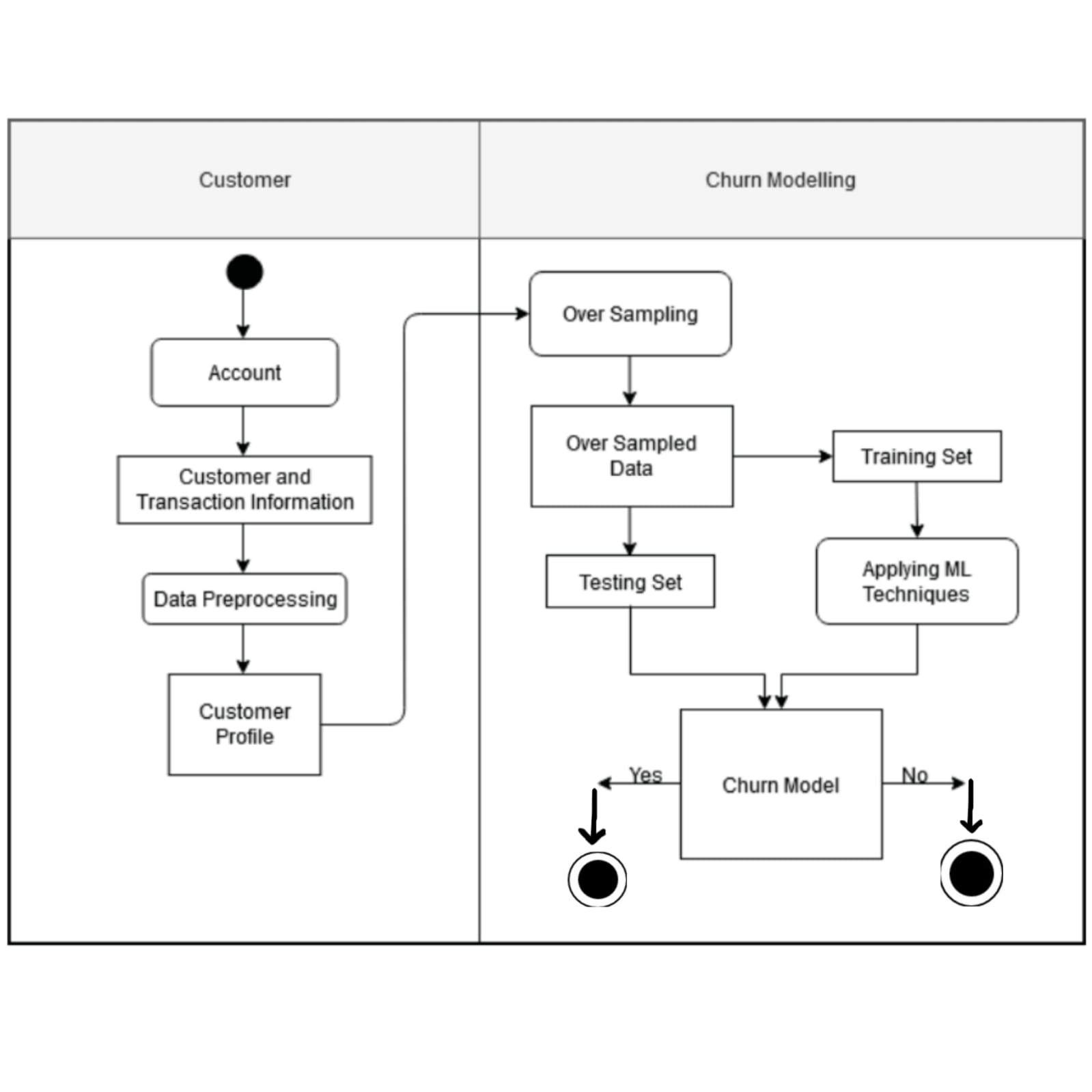
Activity diagrams may be regarded as a form of flowchart. Typical flowchart techniques lack constructs for expressing concurrency. However, the join and split symbols in activity diagrams only resolve this for simple cases; the meaning of the model is not clear when they are arbitrarily combined with decisions or loops.

An activity diagram is graphical representation of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams are intended to model both computational and organizational process (i.e., workflows)

An activity diagram visually presents a series of actions or flow of control in a system like a flowchart or a data flow diagram**.** Activity diagrams are often used in business process modelling. They can also describe the steps in a use case diagram. Activities modelled on can be sequential and concurrent. In both cases an activity diagram.

##### Table: 5.1.2A Graphical Representations of Activity Diagrams

|  |  |  |
| --- | --- | --- |
| Action State | It represents the execution of an atomic action. An action state is a simple state with an entry action whose only exit transition is triggered by the implicit event of completing the execution of the entry action. |  |
| Initial state | An initial node is a control node at which flow starts when the activity is invoked. An activity may have more than one initial state. |  |
| Transition | Transition is used to show the flow of action from one state to another. |  |
| Final state | A final node is a control node at which flow starts when the activity is invoked. |  |



**FIG NO 5.1.2B: Graphical Notation for Activity Diagram**

##### CLASS DIAGRAM:

The Class diagram is a static diagram. It represents the static view

of an application. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application. It describes the attributes and operations of a class and the constraints imposed on the system. The class diagrams are widely used in the modelling of object-oriented systems because they are the only UML diagrams which can be mapped directly with object-oriented languages. The class diagram shows a collection of classes, interfaces, associations, collaborations and constraints. It is also known as a structural diagram.

Purpose

The purpose of the class diagram is to model the static view of an application. The class diagrams are only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction.

UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application, but class diagram is a bit different. So, it is the most popular UML diagram in the coder community. So, the purpose of the class diagram can be summarized as:

* Analysis and design of the static view of an application.
* Describe the responsibilities of a system.
* Based on component and deployment diagrams.
* Forward and reverse engineering.

Active Class

Active classes initiate and control the flow of activity, while passive classes store data and serve other classes. Illustrate active classes with a thicker border.

Visibility

Use visibility markers to signify who can access the information which is in a class. Private visibility hides information from anything outside the class partition. Public visibility allows all other classes to view the marked information. Protected visibility allows child classes to access information which is inherited from a parent class.

Associations

Associations represent static relationship between the classes. Place the association names above, on or below the association line. Use a filled arrow to indicate the direction of the relationship. Place roles at the end of an association. Roles represent how the two classes see each other.

Multiplicity

Place multiplicity notations at the ends of an association. These symbols indicate the number of instances of one class linked to the instance of another class.

Constraint

Constraints are placed inside the curly braces {}.

Composition and Aggregation

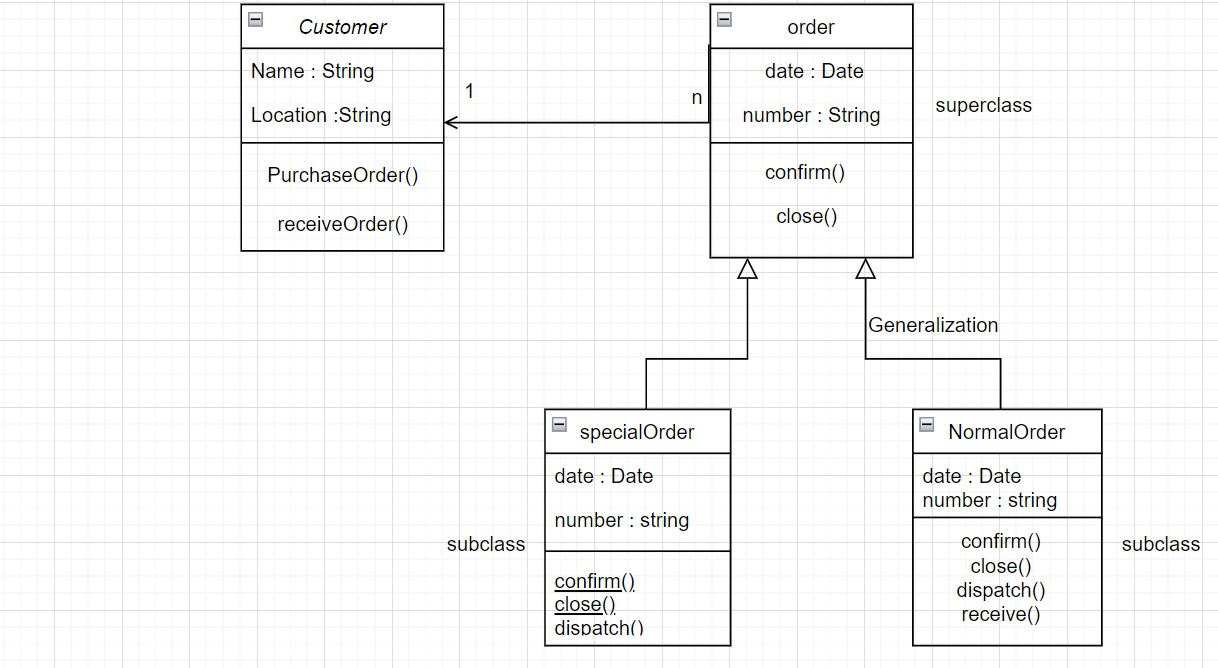
Composition and aggregation link a semantic association between two classes in UML diagram. They are used in class diagrams. They both differ in their symbols.

Generalization

It is a specification relationship in which objects of the specialized element (the child) are substitutable for objects of the generalization element (the parent). It is used in class diagrams.

In our class diagram classes are symptoms reader, symptoms Analyzer and Calculate vales. The responsibility of Symptoms reader is to read the symptoms from the patients. Symptoms Reader have the methods such as get Symptoms and add symptoms which are used to get the symptoms from the user and

add to the training data set. Symptoms Analyzer can analyze the symptoms entered by the user and calculate vale class can calculate the accuracy values to predict the disease and suggest Specialist.



**FIG: 5.1.3A CLASS DIAGRAM**

**5.1.5. COMPONENT DIAGRAM:**

A component diagram breaks down the actual system under development into various high levels of functionality. Each component is responsible for one clear aim within the entire system and only interacts with other essential elements on a need-to-know basis.

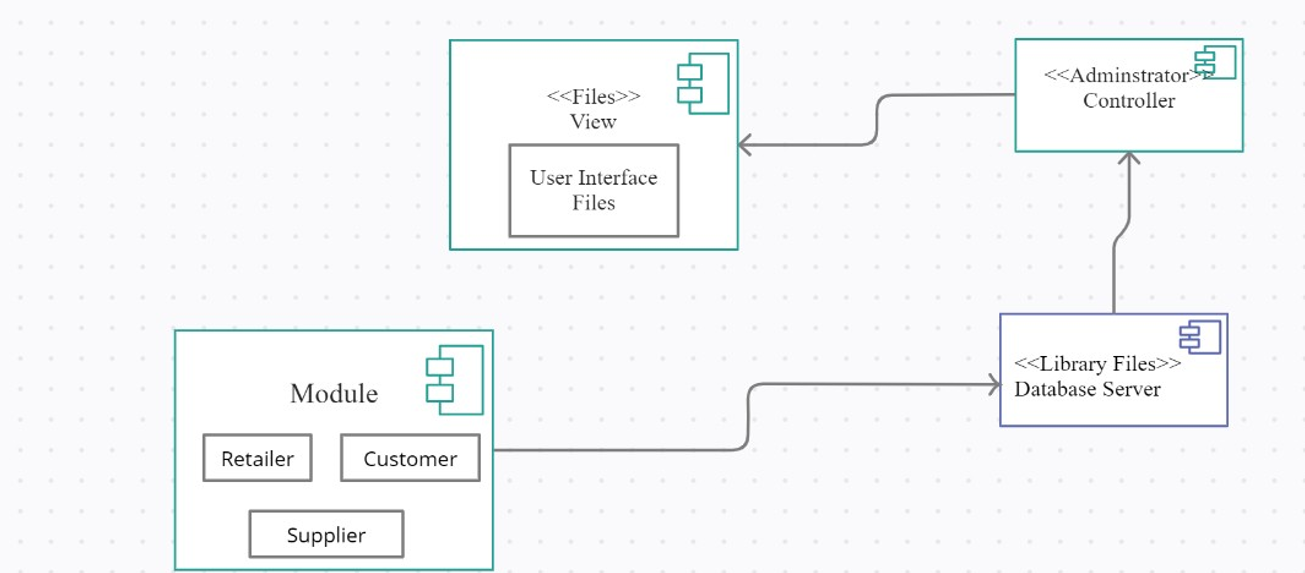
## Basic Concepts of Component Diagram:

A component represents a modular part of a system that encapsulates its contents and whose manifestation is replaceable within its environment. In UML 2, a component is drawn as a rectangle with optional compartments stacked vertically. A high-level, abstracted view of a component in UML 2 can be modeled as:

1. A rectangle with the component's name
2. A rectangle with the component icon
3. A rectangle with the stereotype text and/or icon

Looks of a Component

**FIG: 5.1.3A COMPONENT DIAGRAM**



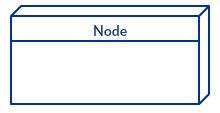
**5.1.6. DEPLOYMENT DIAGRAM:**

A deployment diagram is a UML diagram type that shows the execution architecture of a system, including nodes such as hardware or software execution environments, and the middleware connecting them. Deployment diagrams are typically used to visualize the physical hardware and software of a system. Using it you can understand how the system will be physically deployed on the hardware.

Deployment diagrams help model the hardware topology of a system compared to other UML diagram types which mostly outline the logical components of a system.

**Deployment Diagram Notations**

In order to draw a deployment diagram, you need to first become familiar with the following deployment diagram notations and deployment diagram elements.

**Nodes**

A node, represented as a cube, is a physical entity that executes one or more components, subsystems or executables. A node could be a hardware or software element.

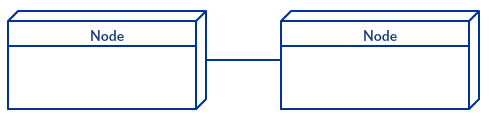
**Artifacts**

Artifacts are concrete elements that are caused by a development process. Examples of artifacts are libraries, archives, configuration files, executable files etc.

**Communication Association**

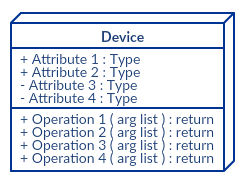
This is represented by a solid line between two nodes. It shows the path of communication

between nodes.

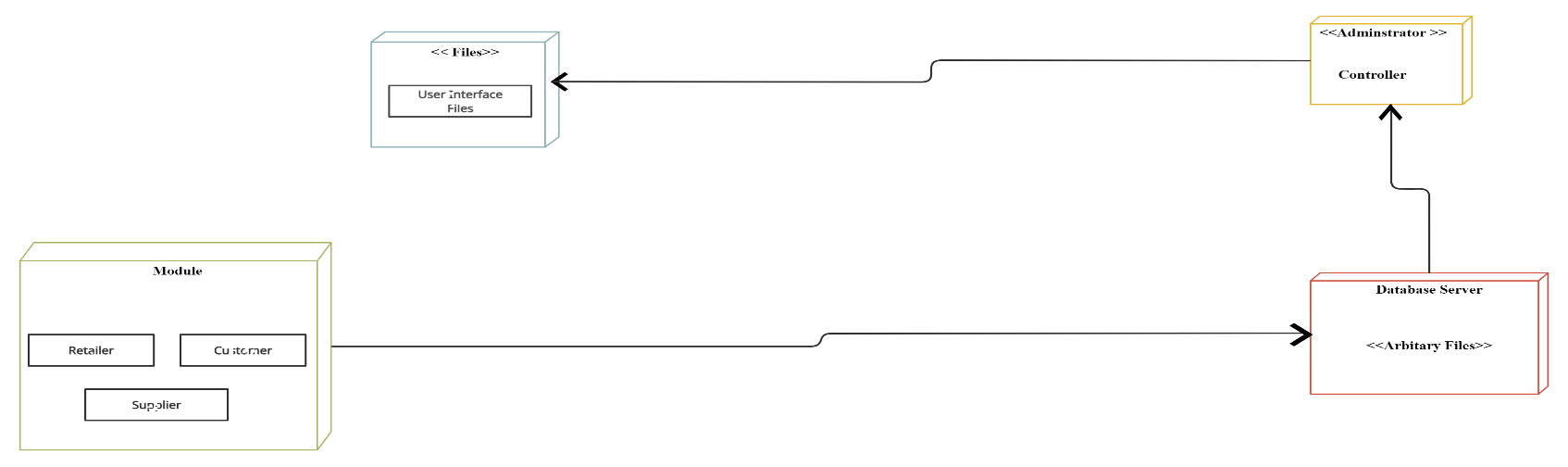


**Devices**

A device is a node that is used to represent a physical computational resource in a system. An example of a device is an application server.



**FIG: 5.1.6 DEPLOYMENT DIAGRAM**



##### 

**CHAPTER 6**

**IMPLEMENTATION**

**IMPLEMENTATION**

##### ALGORITHM DESCRIPTION

As customers use a subscription-based service they generate large amounts of structured data and metadata. From demographic data to usage and transaction data, it is relatively straightforward and efficient to collect and analyze even in a way which is privacy conscious. The idea underpinning this approach is that customers on the verge of dropping exhibit certain common **churn signals**. These can be identified in the structured data and metadata that they generate through the use of their account. [Streaming data analytics](https://www.bizdata.com.au/blogpost.php?p=real-time-vs-streaming-analytics) systems are typically best suited to handling this level of real-time data.

**DECISION TREE ALGORITHM :**

* Decision Tree is a **Supervised learning technique** that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where **internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**
* In a Decision tree, there are two nodes, which are the **Decision Node** and **Leaf Node.** Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
* The decisions or the test are performed on the basis of features of the given dataset.
* ***It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.***
* It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
* In order to build a tree, we use the **CART algorithm,** which stands for **Classification and Regression Tree algorithm.**
* A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree i nto subtrees.

**RANDOM FOREST CLASSIFIER :**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, ***"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."*** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

**Step-1:** Select random K data points from the training set.

**Step-2:** Build the decision trees associated with the selected data points (Subsets).

**Step-3:** Choose the number N for decision trees that you want to build.

**Step-4:** Repeat Step 1 & 2.

**Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

**Naïve Bayes**

A [comparison study](https://acadpubl.eu/jsi/2018-119-10/articles/10b/2.pdf) of the different machine learning approaches that are typically deployed for customer attrition analysis claims that the best method is a boosted **Support Vector Machines** model which had an accuracy score of 95% and an f-score of 84%. Conversely, the worst performing machine learning methods for the task were the **Naïve Bayes** and **Logical Regression** models

The research suggests that achieving the best results in machine learning churn prediction relies on using a large training dataset as well as implementing **boosting algorithms**, the use of which is associated in a substantial increase in f-score. The specific algorithm (or, more accurately a meta-algorithm) used in this case was **AdaBoost**, which refines the process over time. When implemented, it places more weight on those particular features of a dataset which have been shown to be more relevant. This reduces the dimensionality of the data and reduces the computing resources spent on processing features that are less relevant.

highlights two companies working with two different datasets which successfully predicted churn using **decision trees** learning model. The first applied customer service logs, work order details and contractual details to pull a list of clients that were predicted to churn with an accuracy level of 89%. The second parsed the activity logs of customers, using data such as number of searches, number of downloads and number of links following.

### **SEMANTIC ANALYSIS**

One [analysis](https://books.google.com.au/books?id=qDRyDwAAQBAJ&pg=PA156&lpg=PA156&dq=Customer+Churn+Prediction+Using+Sentiment+Analysis+and+Text+Classification+of+VOC&source=bl&ots=32nG6MHTAl&sig=ACfU3U18cWhJS2cOVl4whlfsLsrgCUYPxA&hl=en&sa=X&ved=2ahUKEwiRwsKHhZXjAhUEUI8KHacQC64Q6AEwA3oECAkQAQ#v=onepage&q&f=false) of these topics calls this space as the **Voice of the Customer (VOC)** and defines it as “call center calls, emails, questionnaires, web reviews and SMS”. It identifies that the existing research on this topic is still quite sparse and experimental, but highlights cases in which unstructured call center information applied to a churn prediction model led to an increase in its predictive power.

**KNN**

KNN is a simple algorithm. When a new customer is presented, the algorithm looks through the database for customers who are most similar to the target customer. It then predicts if the customer would churn based on whether those similar customers churned or not.

This contained 976 customers of a telecommunications company. 125 of them churned while 851 stayed.

To make the model, I split the data into two. 70 per cent was put in a training set and used to create the model. The model was then used to predict if the other 30% churned or not.

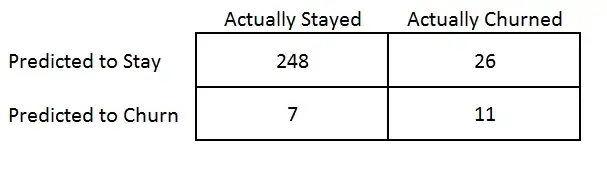
The variables used to predict whether the customer churned available were:

* State
* Account Length
* Area Code
* Phone
* Int’l Plan
* Vmail Plan
* Vmail Message
* Day Mins
* Day Calls
* Day Charge
* Eve Mins
* Eve Calls
* Eve Charge
* Night Mins
* Night Calls
* Night Charge
* Intl Mins
* Intl Calls
* Intl Charge
* Cust Serv Calls

KNN only works with numerical variables, so for this model I will remove all non-numerical variables. (There are techniques for using categorical variables like one-hot encoding, but we’ll ignore them here).

How did the model perform?

Here are the results:



**Accuracy = 89%**

## Optimising the model

If K is too small the model will be “overfit”. This means that the model will do well on the data you used to create it, but when it comes across new observations it will perform poorly. If K is too high, the model will also perform poorly. The best value of K was selected by picking one that is not to high or low.

## Variable Selection

Not all variable are useful in predicting if a customer will churn. For example, the customers phone number is completely useless in predicting because it is unique to each customer.

##### Technology Description

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage- collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

##### JUPYTER

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

While there are many ways to use Jupyter Notebook depending on your goals and intentions, the first step in using the notebook is getting to know the technology. One of the best parts of using Jupyter Notebook is its flexibility and versatility, as you can work with the technology by downloading [JupyterLab](https://blog.jupyter.org/jupyterlab-is-ready-for-users-5a6f039b8906), or simply opening your browser to use the notebook on the go through the original application. A tool that was made for collaboration and interactivity, Jupyter Notebook is also compatible with the most commonly used programming languages such as R, C++, Ruby, and Python.

Jupyter Notebook also gives the user access to a community of fellow users and open-source programming libraries. Once you begin using it, it is easy to find additional information and instructions on how to use the technology and integrate it into other components that may interest you. Divided into front end and back end interfaces, Jupyter Notebook not only gives users access to the outcome of their code but also assists in the process of tweaking and editing the code before it is executed. This list outlines the top five uses for the Jupyter Notebook both within and outside of the classroom.

##### HTML

HTML stands for Hyper Text Mark-up Language. It is used to design web pages using a mark- up language. HTML is the combination of Hypertext and Mark-up language. Hypertext defines the link between the web pages. A mark-up language is used to define the text document within tag which defines the structure of web pages.

HTML is a mark-up language that is used by the browser to manipulate text, images, and other content to display it in the required format.

**JAVA SCRIPT**

Java script is the most popular programming language in the world and that makes it a programmer’s great choice. Once you learnt Java script, it helps you developing great front- end as well as back-end software’s using different Java script-based frameworks like j Query, Node.JS etc.

**CSS**

Cascading Style Sheets, fondly referred to as CSS, is a simple design language intended to simplify the process of making web pages presentable.

CSS is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain.

**J QUERY**

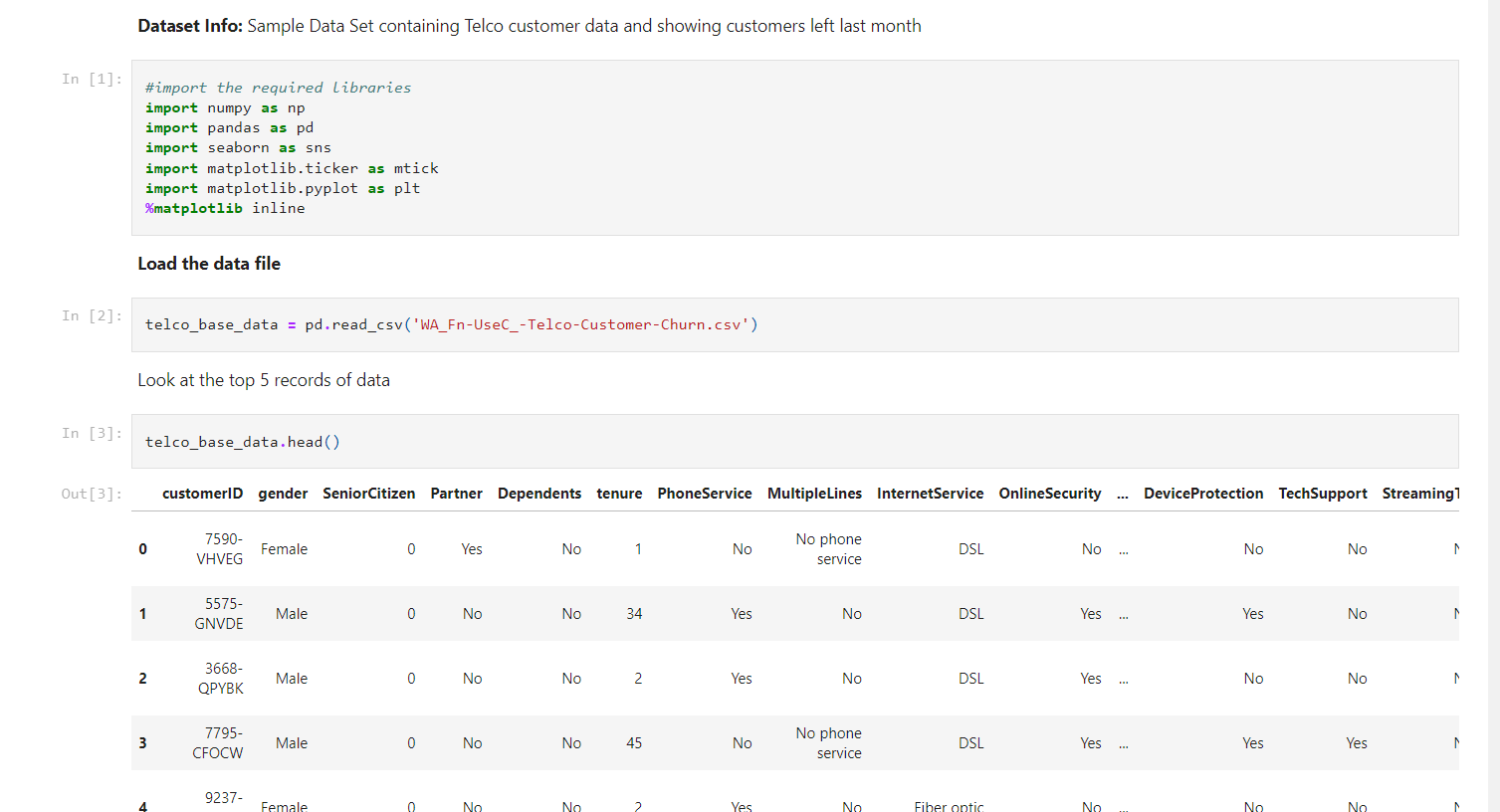
J Query is a fast and concise JavaScript library created by John Resign in 2006. jQuery simplifies HTML document traversing, event handling, animating, and Ajax interactions for Rapid Web Development.

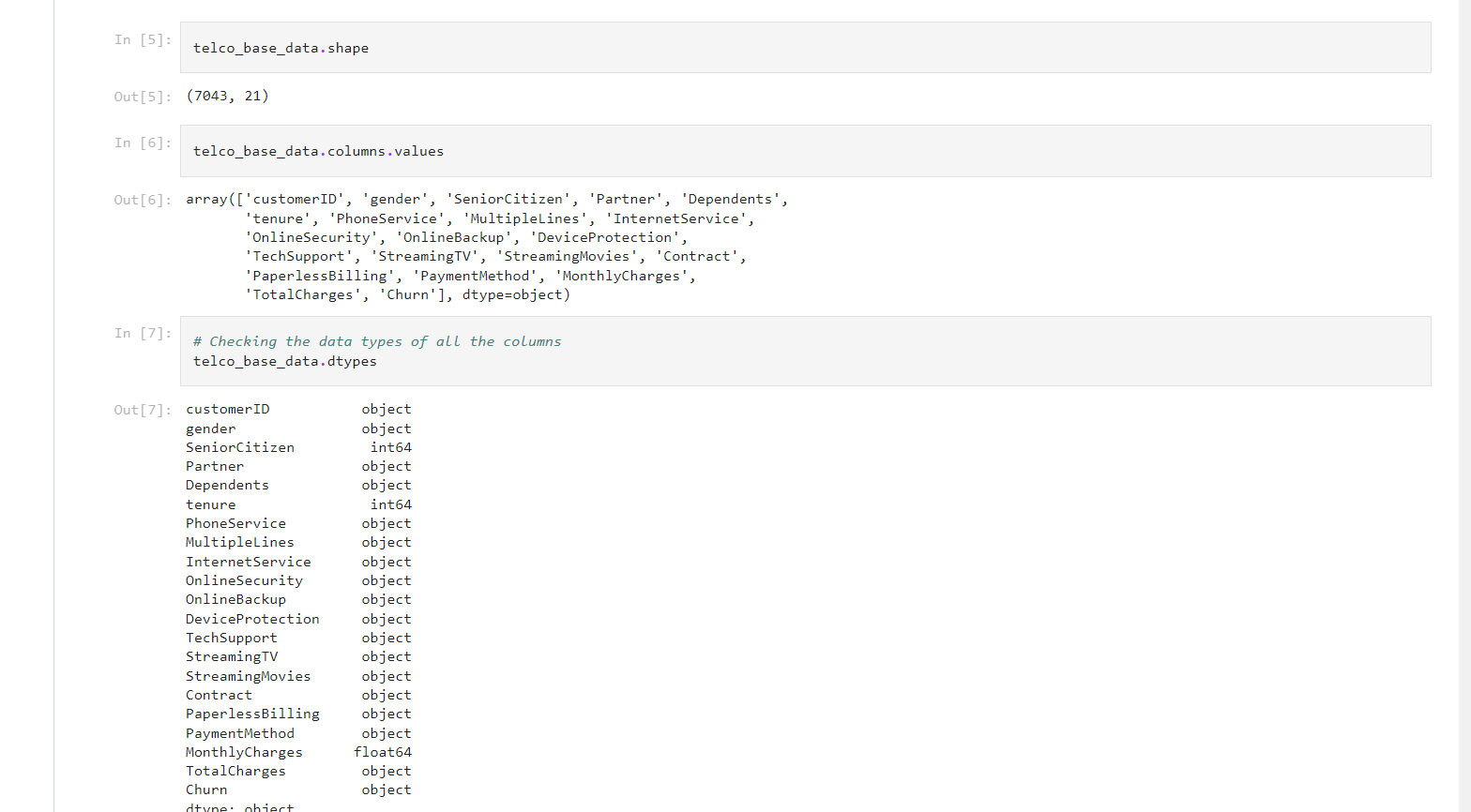
**FLASK**

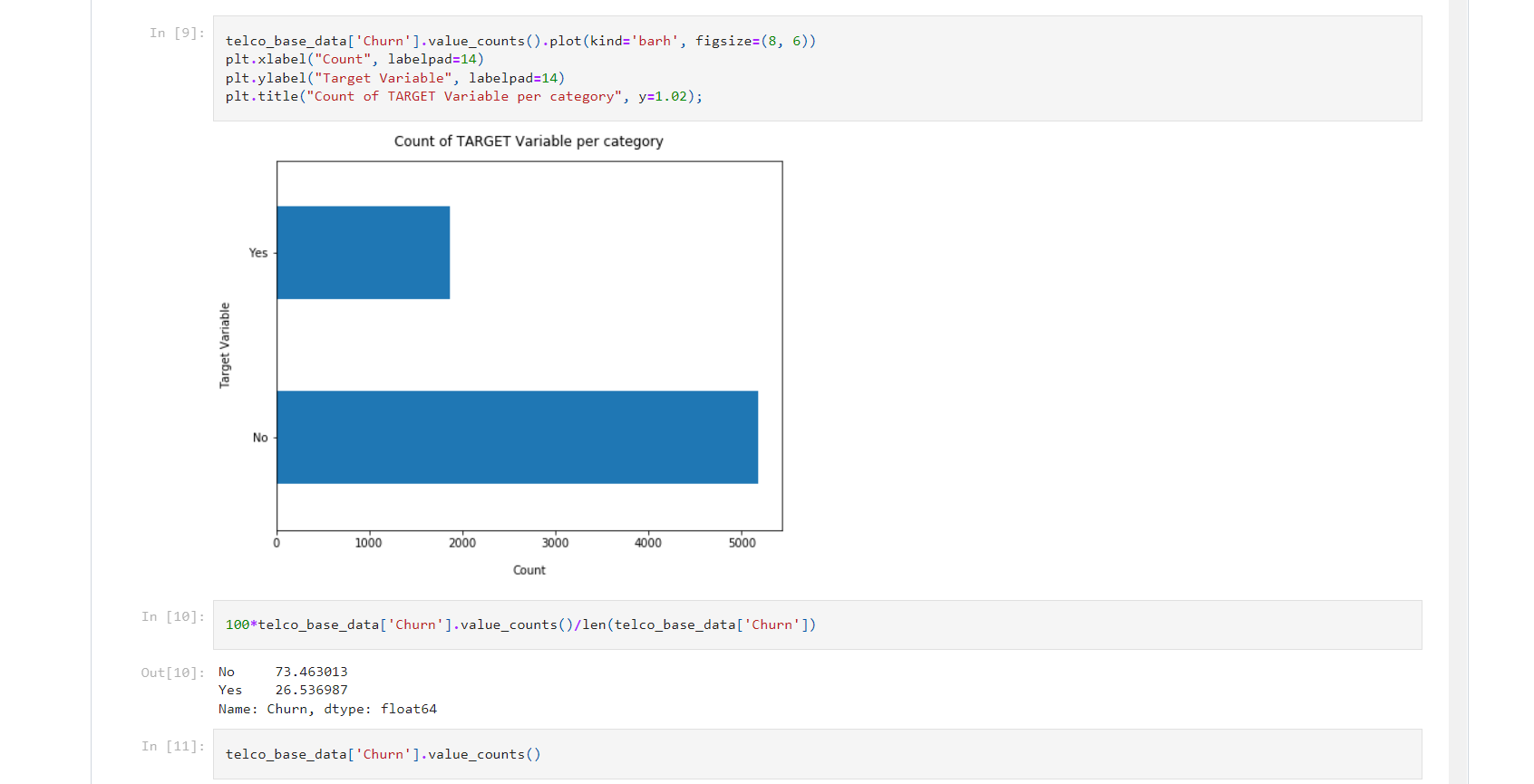
Flask is a web framework, it’s a Python module that lets you develop web applications easily. It has a small and easy-to-extend core: it’s a microframework that doesn’t include an ORM (Object Relational Manager) or such features.It does have many cool features like url routing, template engine. It is a WSGI web app framework.

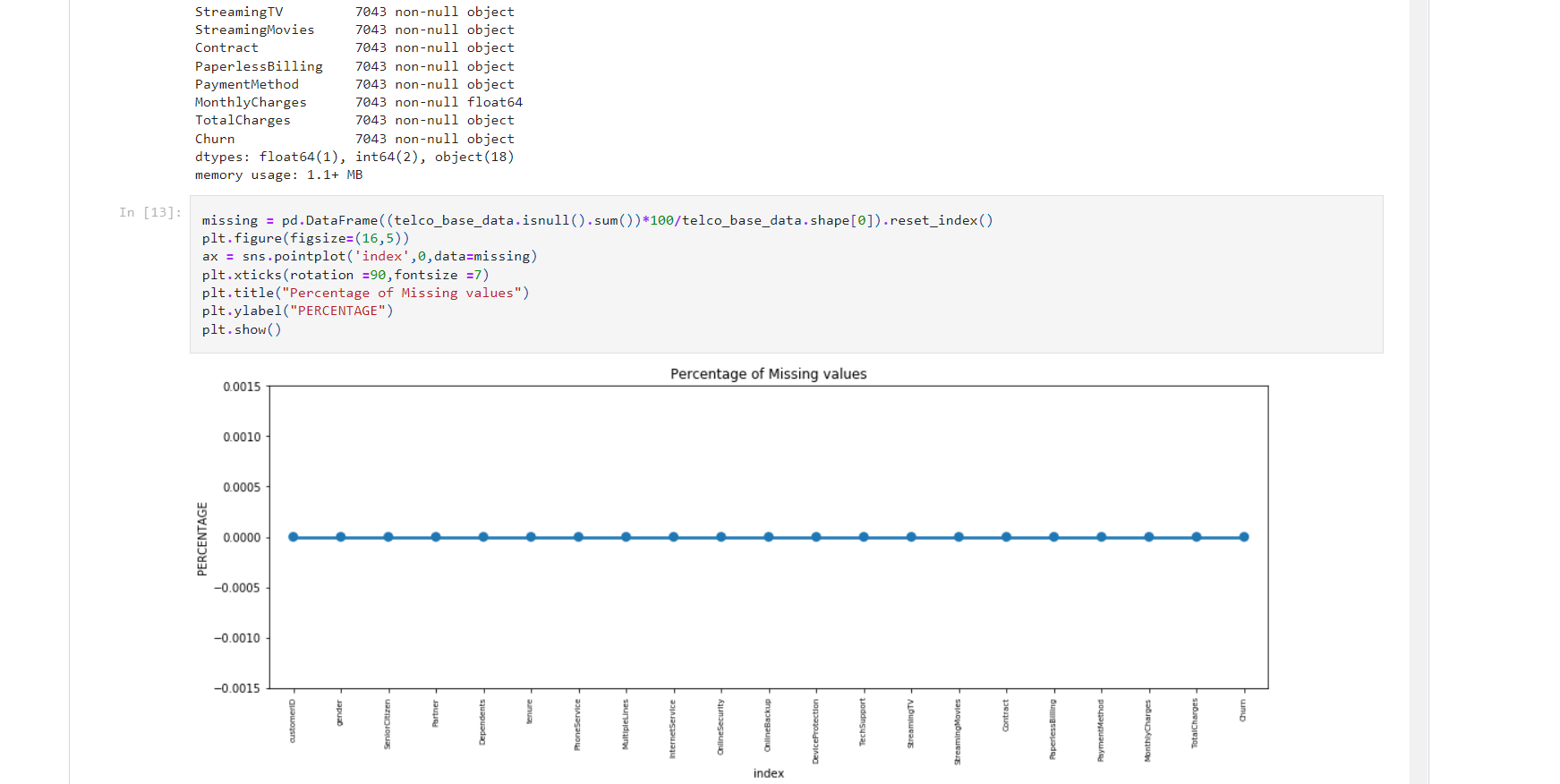
Flask is a web application framework written in Python. It was developed by Armin Ronacher, who led a team of international Python enthusiasts called Poocco. Flask is based on the Werkzeg WSGI toolkit and the Jinja2 template engine.Both are Pocco projects.

**TELECOME CHURN ANALYSIS**

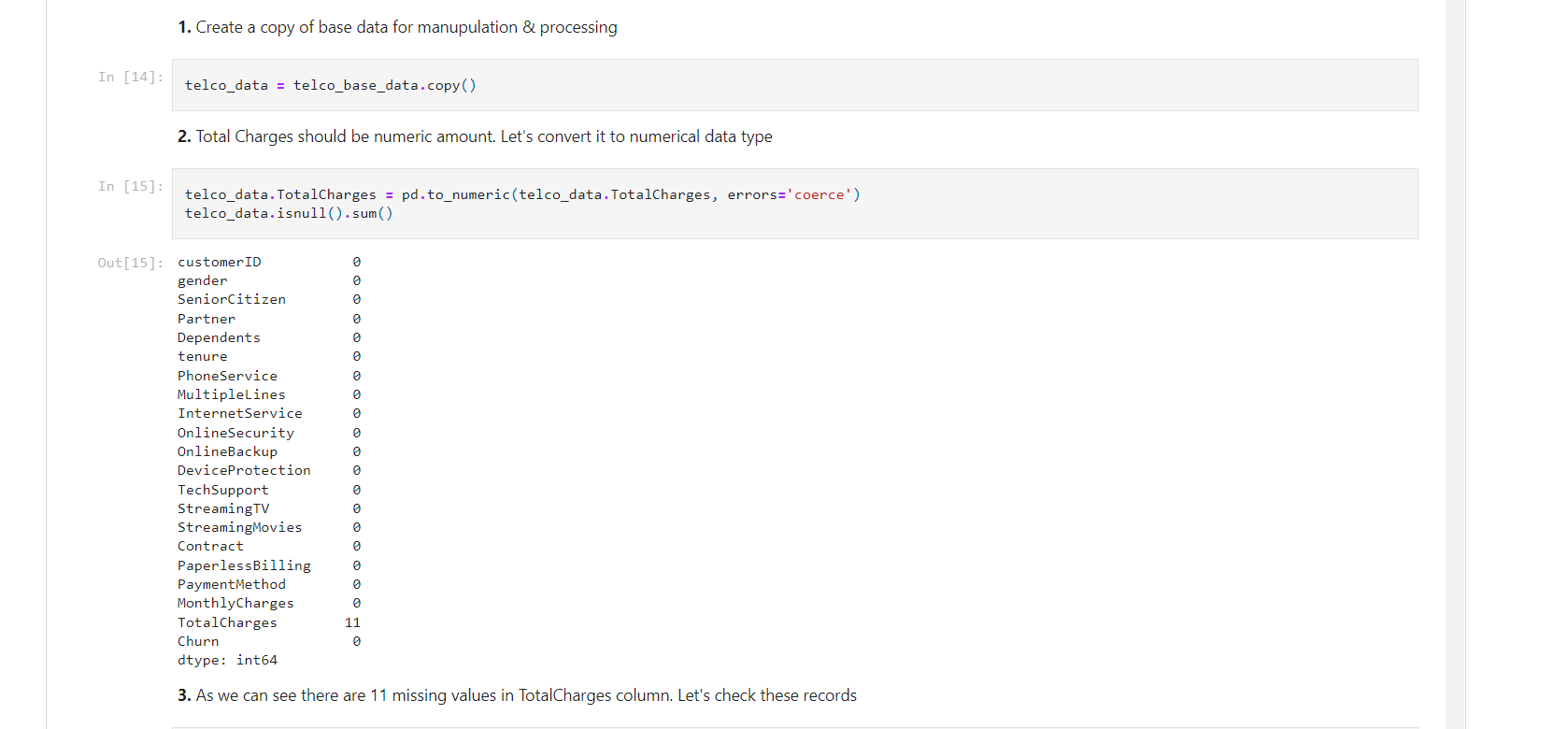




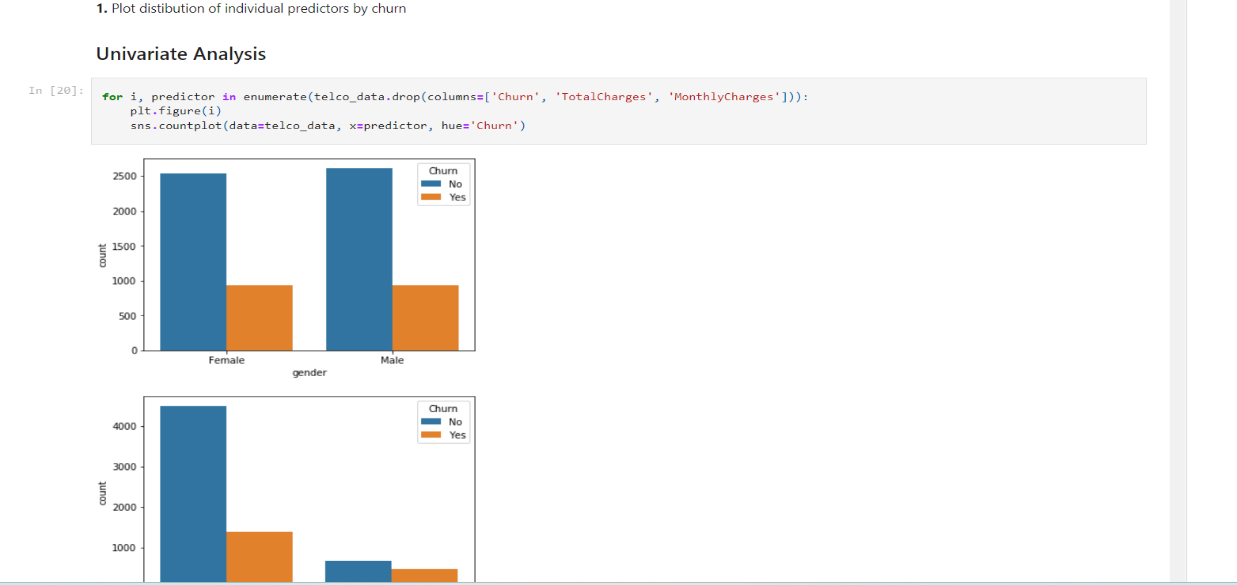


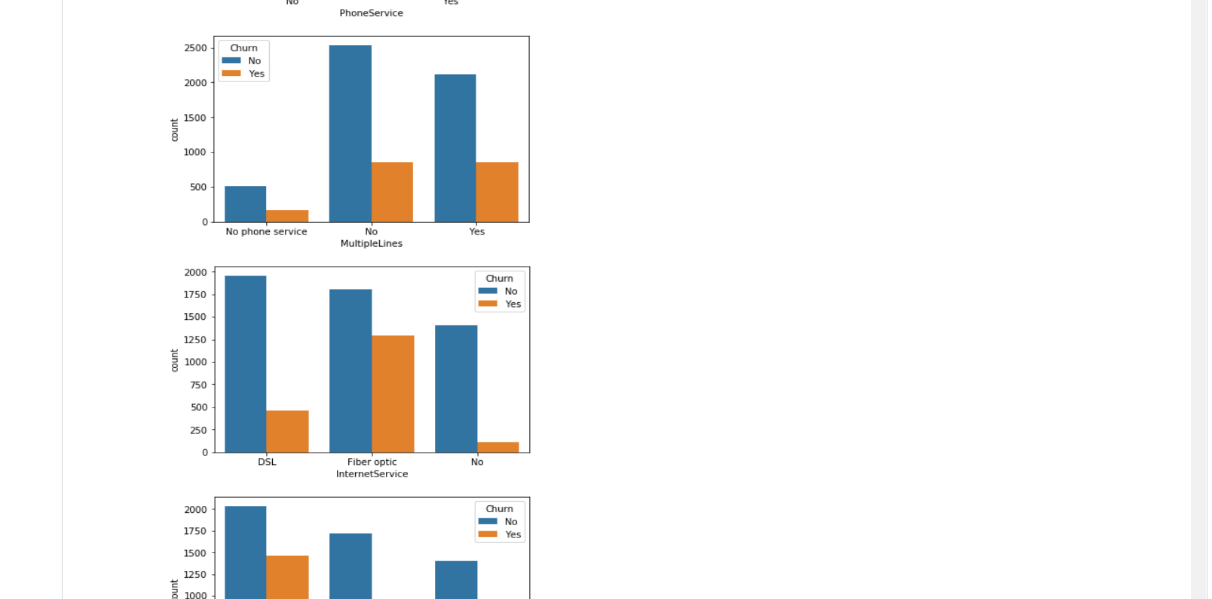


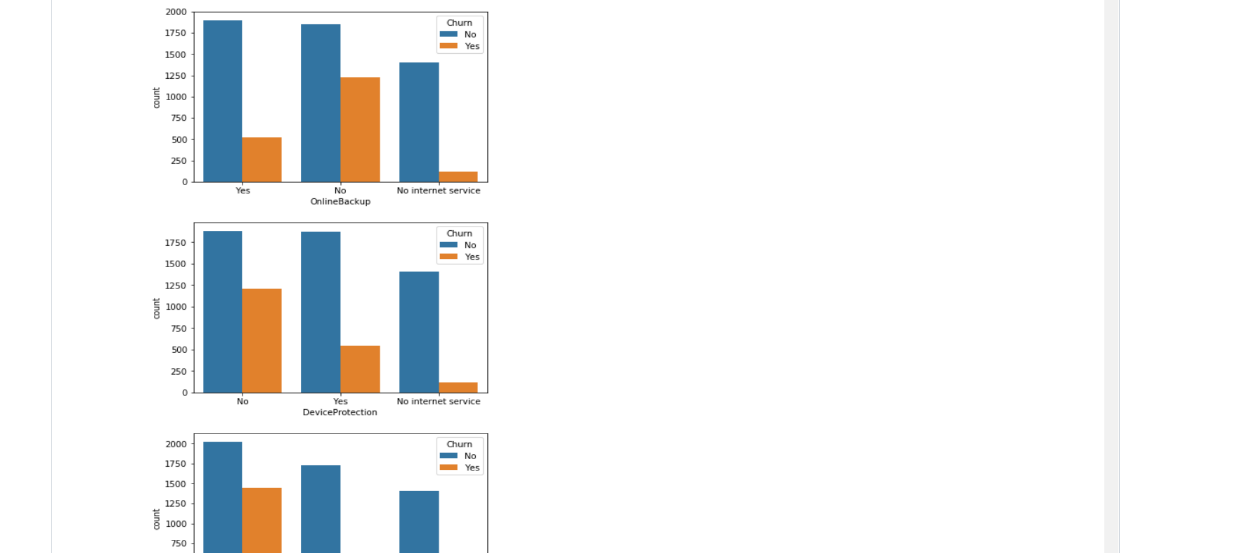
**DATA CLEANING**

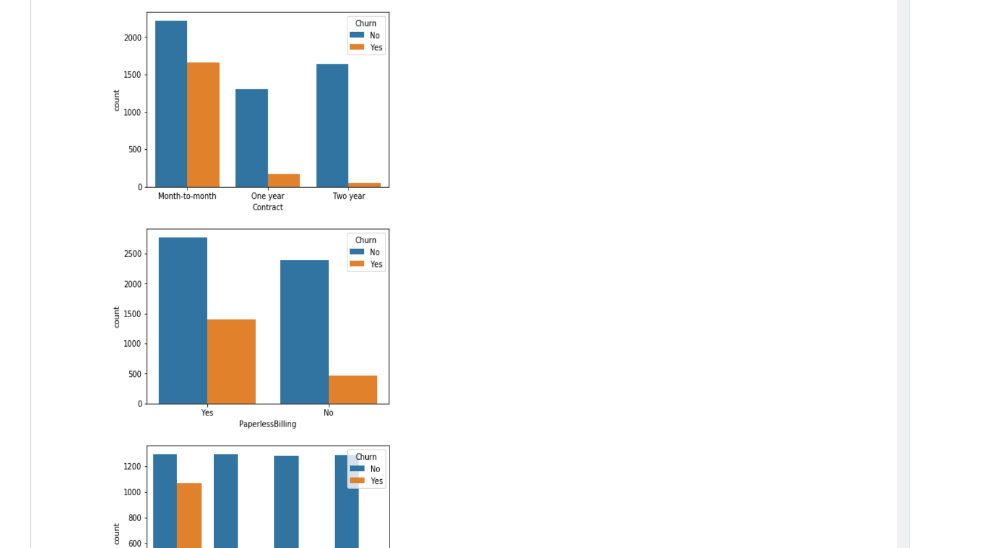


## Data Exploration

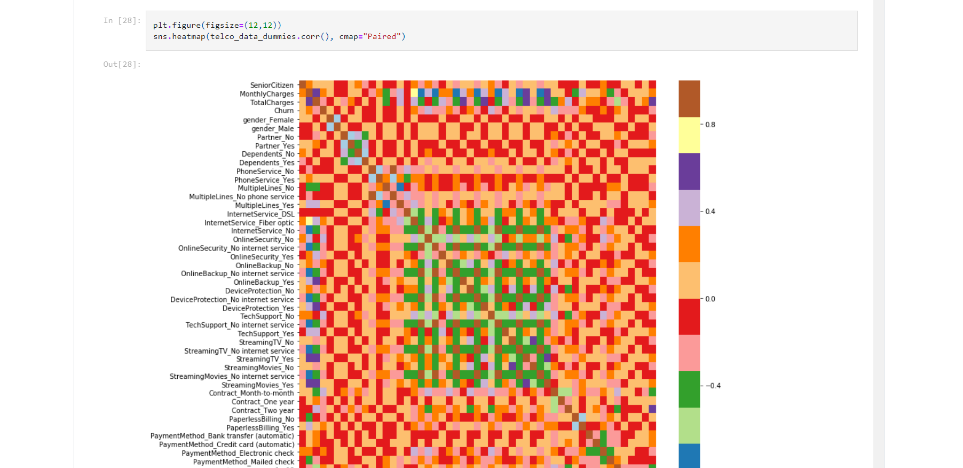


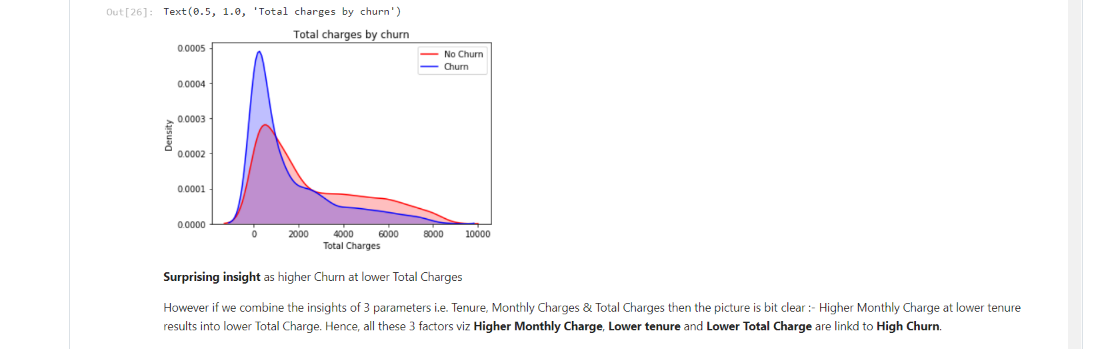


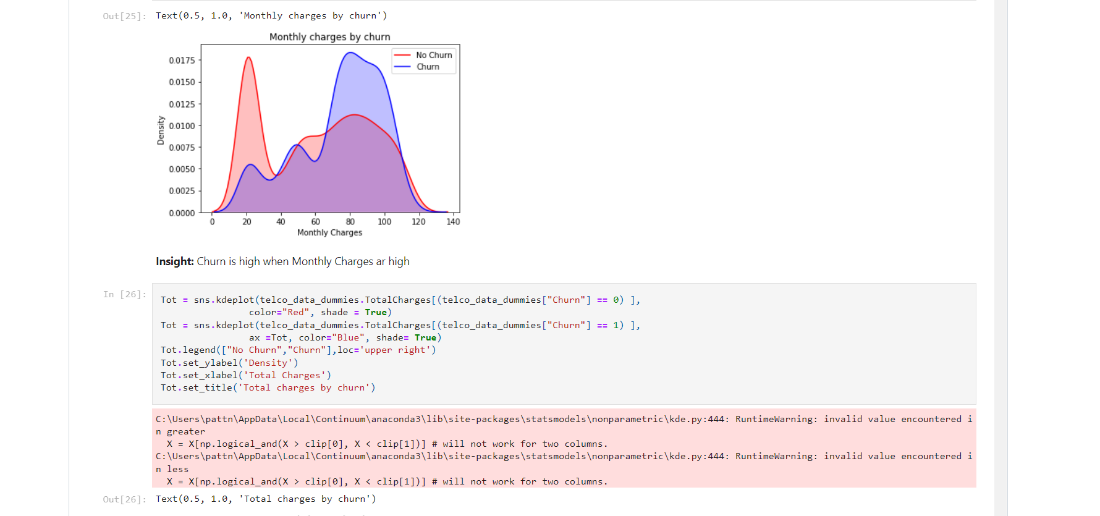




**INSIGHTS GATHERING DERIVED INSIGHTS**







##### Sample Source Code:

##### # coding: utf-8

##### 

##### import pandas as pd

##### from sklearn.model\_selection import train\_test\_split

##### from sklearn.ensemble import RandomForestClassifier

##### from sklearn import metrics

##### from flask import Flask, request, render\_template

##### import pickle

##### 

##### app = Flask("\_\_name\_\_")

##### 

##### df\_1=pd.read\_csv("first\_telc.csv")

##### 

##### q = ""

##### 

##### @app.route("/")

##### def loadPage():

##### return render\_template('home.html', query="")

##### 

##### 

##### @app.route("/", methods=['POST'])

##### def predict():

##### 

##### '''

##### SeniorCitizen

##### MonthlyCharges

##### TotalCharges

##### gender

##### Partner

##### Dependents

##### PhoneService

##### MultipleLines

##### InternetService

##### OnlineSecurity

##### OnlineBackup

##### DeviceProtection

##### TechSupport

##### StreamingTV

##### StreamingMovies

##### Contract

##### PaperlessBilling

##### PaymentMethod

##### tenure

##### '''

##### 

##### 

##### 

##### inputQuery1 = request.form['query1']

##### inputQuery2 = request.form['query2']

##### inputQuery3 = request.form['query3']

##### inputQuery4 = request.form['query4']

##### inputQuery5 = request.form['query5']

##### inputQuery6 = request.form['query6']

##### inputQuery7 = request.form['query7']

##### inputQuery8 = request.form['query8']

##### inputQuery9 = request.form['query9']

##### inputQuery10 = request.form['query10']

##### inputQuery11 = request.form['query11']

##### inputQuery12 = request.form['query12']

##### inputQuery13 = request.form['query13']

##### inputQuery14 = request.form['query14']

##### inputQuery15 = request.form['query15']

##### inputQuery16 = request.form['query16']

##### inputQuery17 = request.form['query17']

##### inputQuery18 = request.form['query18']

##### inputQuery19 = request.form['query19']

##### 

##### model = pickle.load(open("model.sav", "rb"))

##### 

##### data = [[inputQuery1, inputQuery2, inputQuery3, inputQuery4, inputQuery5, inputQuery6, inputQuery7,

##### inputQuery8, inputQuery9, inputQuery10, inputQuery11, inputQuery12, inputQuery13, inputQuery14,

##### inputQuery15, inputQuery16, inputQuery17, inputQuery18, inputQuery19]]

##### 

##### new\_df = pd.DataFrame(data, columns = ['SeniorCitizen', 'MonthlyCharges', 'TotalCharges', 'gender',

##### 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService',

##### 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',

##### 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',

##### 'PaymentMethod', 'tenure'])

##### 

##### df\_2 = pd.concat([df\_1, new\_df], ignore\_index = True)

##### # Group the tenure in bins of 12 months

##### labels = ["{0} - {1}".format(i, i + 11) for i in range(1, 72, 12)]

##### 

##### df\_2['tenure\_group'] = pd.cut(df\_2.tenure.astype(int), range(1, 80, 12), right=False, labels=labels)

##### #drop column customerID and tenure

##### df\_2.drop(columns= ['tenure'], axis=1, inplace=True)

##### 

##### new\_df\_\_dummies = pd.get\_dummies(df\_2[['gender', 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService',

##### 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup',

##### 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies',

##### 'Contract', 'PaperlessBilling', 'PaymentMethod','tenure\_group']])

##### 

##### 

##### #final\_df=pd.concat([new\_df\_\_dummies, new\_dummy], axis=1)

##### 

##### 

##### single = model.predict(new\_df\_\_dummies.tail(1))

##### probablity = model.predict\_proba(new\_df\_\_dummies.tail(1))[:,1]

##### 

##### if single==1:

##### o1 = "This customer is likely to be churned!!"

##### o2 = "Confidence: {}".format(probablity\*100)

##### else:

##### o1 = "This customer is likely to continue!!"

##### o2 = "Confidence: {}".format(probablity\*100)

##### 

##### return render\_template('home.html', output1=o1, output2=o2,

##### query1 = request.form['query1'],

##### query2 = request.form['query2'],

##### query3 = request.form['query3'],

##### query4 = request.form['query4'],

##### query5 = request.form['query5'],

##### query6 = request.form['query6'],

##### query7 = request.form['query7'],

##### query8 = request.form['query8'],

##### query9 = request.form['query9'],

##### query10 = request.form['query10'],

##### query11 = request.form['query11'],

##### query12 = request.form['query12'],

##### query13 = request.form['query13'],

##### query14 = request.form['query14'],

##### query15 = request.form['query15'],

##### query16 = request.form['query16'],

##### query17 = request.form['query17'],

##### query18 = request.form['query18'],

##### query19 = request.form['query19'])

##### 

##### app.run()

**CHAPTER 7**

**TESTING**

# TESTING

#### 7.1 Introduction

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs (errors or other defects).

Testing involves the execution of a software component or system component to evaluate one or more properties of interest. In general, these properties indicate the extent to which the component or system under test have the following

* + - Meets the requirements that guided its design and development
    - Responds correctly to all kinds of inputs
    - Performs its functions within an acceptable time
    - Is sufficiently usable
    - Can be installed and run in its intended environments and achieves the general result its stake Holders desire.

As the number of possible tests for even simple software components is practically infinite, all software testing uses some strategy to select tests that are feasible for the available time and resources. As a result, software testing typically (but not exclusively) attempts to execute a program or application with the intent of finding software bugs (errors or other defects).

Software testing can provide objective, independent information about the quality of software and risk of its failure to users and/or sponsors.

Software testing can be conducted as soon as executable software (even if partially complete) exists. The overall approach to software development often determines when and how testing is conducted. For example, in a phased process, most testing occurs after system requirements have been defined and then implemented in testable programs. In contrast, under an Agile approach, requirements, programming, and testing are often done concurrently.

###### Testing Types

A software engineering product can be tested in one of two ways:

* + - * Black Box Testing
      * White Box Testing

###### Black box testing

Knowing the specified function that a product has been designed to perform, determine whether each function is fully operational.

###### White box testing

Knowing the internal workings of a software product determine whether the internal operation implementing the functions perform according to the specification, and all the internal components have been adequately exercised.

###### Testing Strategies

Four Testing Strategies that are often adopted by the software development team include: Testing Strategies Four Testing Strategies that are often adopted by the software development team include:

* + - Unit Testing
    - Integration Testing
    - Validation Testing
    - System Testing

**Unit Testing**

We adopt white box testing when using this testing technique. This testing was carried out on individual components of the software that were designed. Each individual module was tested using this technique during the coding phase. Every component was checked to make sure that they adhere strictly to the specifications spelt out in the data flow diagram and ensure that they perform the purpose intended for them.

All the names of the variables are scrutinized to make sure that they are truly reflected of the element they represent. All the looping mechanisms were verified to ensure that they were as decided. Beside these, we trace through the code manually to capture syntax errors and logical errors.

##### Integration Testing

After finishing the Unit Testing process, next is the integration testing process. In this testing process we put our focus on identifying the interfaces between components and their functionality as dictated by the DFD diagram. The Bottom-up incremental approach was adopted during this testing. Low level modules are integrated and combined as a cluster before testing.

The Black box testing technique was employed here. The interfaces between the components were tested first. This allowed identifying any wrong linkages or parameters passing early in the development process as it just can be passed in a set of data and checked if the result returned is an accepted one.

##### Validation Testing

The system has been tested and implemented successfully and thus ensured that all the requirements listed in the software requirements specification are completely fulfilled. In case of erroneous input corresponding error messages are displayed.

##### System Testing

System testing is a series of different tests whose primary purpose is to fully exercise the computer-based system. Although each test has a different purpose, all the work should verify that all system elements have been properly integrated and perform allocated functions. System testing also ensures that the project works well in the environment. It traps the errors and allows convenient processing of errors without coming out of the program abruptly.

Software testing is critical element of software quality assurance and represents ultimate review of specification, design and coding. Test case design focuses on a set of technique for the creation of test cases that meet overall testing objectives. Planning and testing of a programming system involve formulating a set of test cases, which are similar to the real data that the system is intended to manipulate. Test castes consist of input specifications, a description of the system functions exercised by the input and a statement of the extended output.

In principle, testing of a program must be extensive. Every statement in the program should be exercised and every possible path combination through the program should be executed at least once. Thus, it is necessary to select a subset of the possible test cases and conjecture that this subset will adequately test the program.

**TESTING CHURN RATE :**

To calculate churn rate, choose a time period, such as monthly or annual. You’ll need to know the number of customers you had at the beginning of the time period and the number you lost. Divide the following: Lost Customers ÷ Total Customers at the Start of Time Period. Then, multiply the number by 100.

Remember, the steps to calculate churn rate are:

* Determine a time period: monthly, annual, or quarterly.
* Determine the number of customers you had at the beginning of the time period.
* Determine the number of customers that churned by the end of the time period.
* Divide the number of lost customers by the number of customers you had prior to the churn.
* Multiply that number by 100.

**Customer Churn Rate = (Lost Customers ÷ Total Customers at the Start of Time Period) x 100**

*Customer Churn Rate = (50 ÷ 500) x 100*  
*Customer Churn Rate = (0.10) x 100*  
*Customer Churn Rate = 10%*

##### Guidelines for developing test cases

* + - Describe which feature or service your test attempts to cover
    - If the test case is based on a use case, it is a good idea to refer to the use case name.
    - Remember that the use cases are the source of test cases. In theory the software is supposed to match the use cases not the reverse. As soon as you have enough use cases

, go ahead and write the test plan for that piece

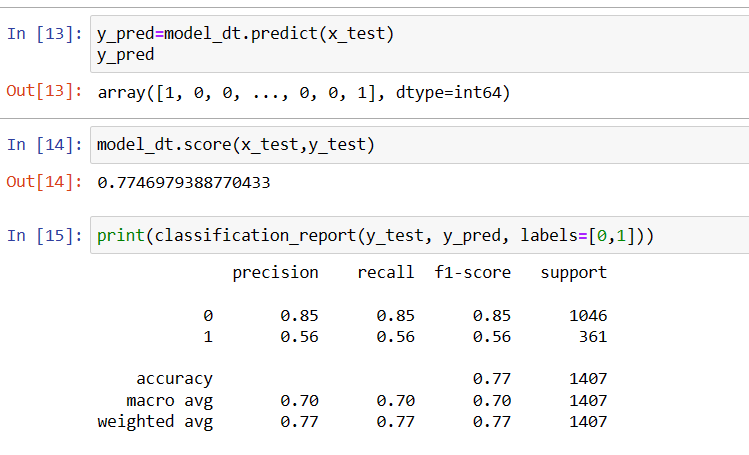
* + - Specify what you are testing and which particular feature. Then specify what you are going to do to test the feature and what you expect to happen.
    - Test the normal use of the object’s methods. Test the abnormal but reasonable use of the object’s methods.
    - Test the abnormal but unreasonable use of the object’s methods.
    - Test the boundary conditions. Also specify when you expect error dialog boxes, when you expect some default event, and when functionality till is being defined.
    - Test object’s interactions and the messages sent among them. If you have developed sequence diagrams, they can assist you in this process
    - when the revisions have been made, document the cases so they become the starting bases for the follow- up test

Attempting to reach agreement on answers generally will raise other what-if questions. Add these to the list and answer them, repeat the process until the list is stabilized, then you need not add any more questions.

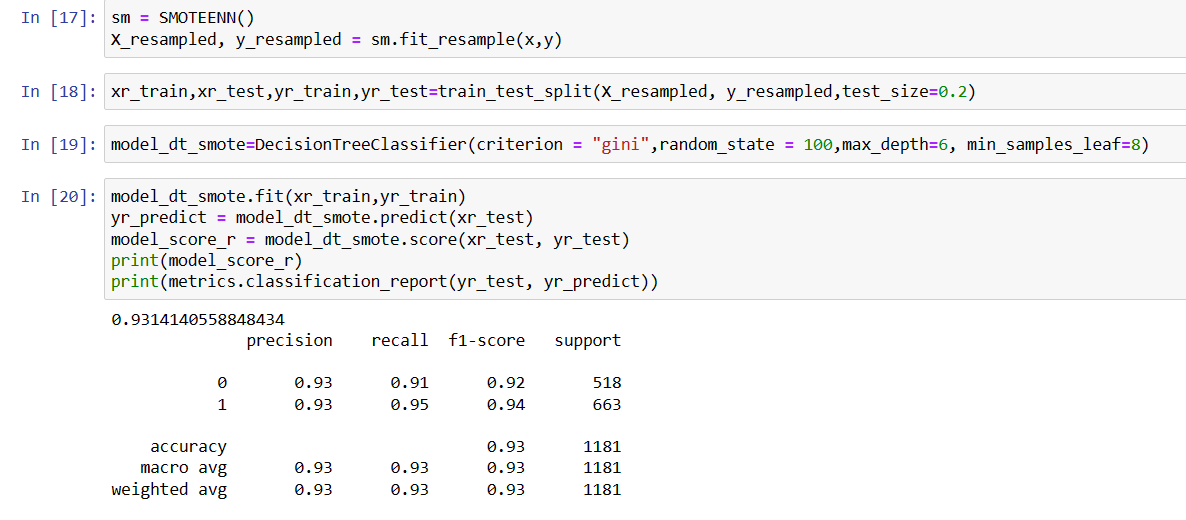
#### 7.2 Test Cases

Table: 7.2 Representation of Test cases and status

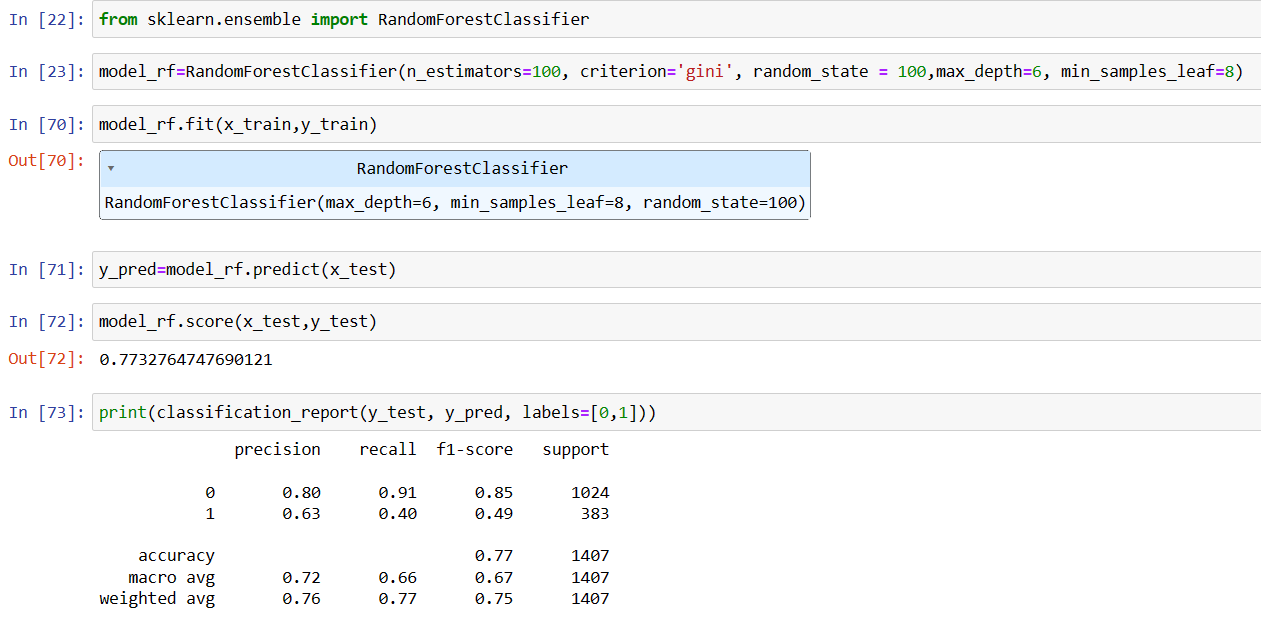
**TEST CASE1 :**



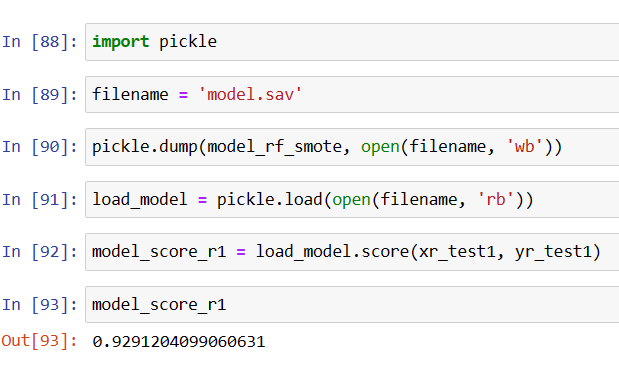
Test Case 2:



**Test Case 3:**



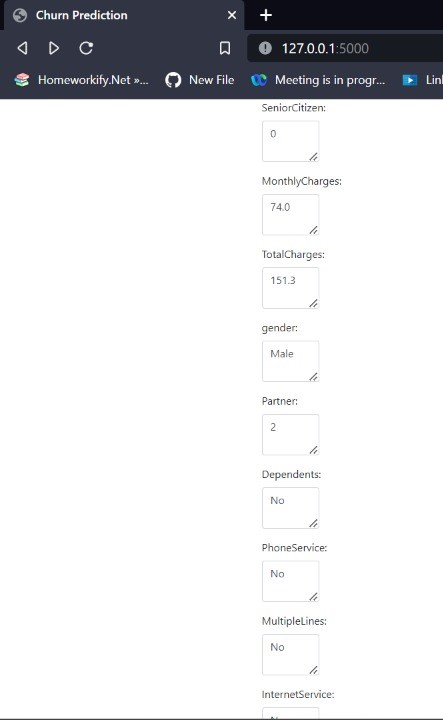
**Results:**

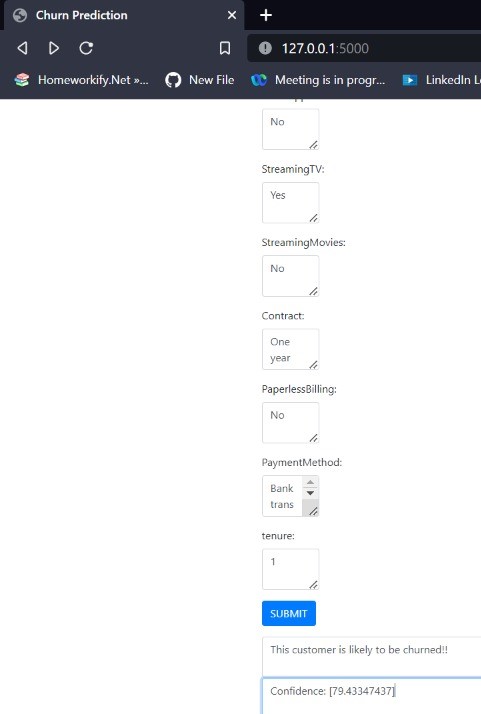


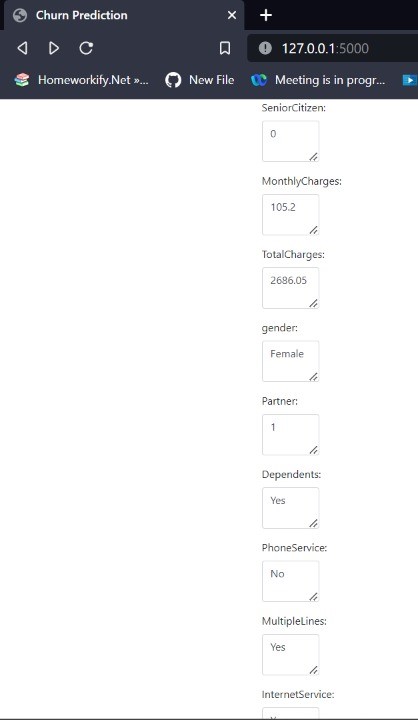
CHAPTER 8

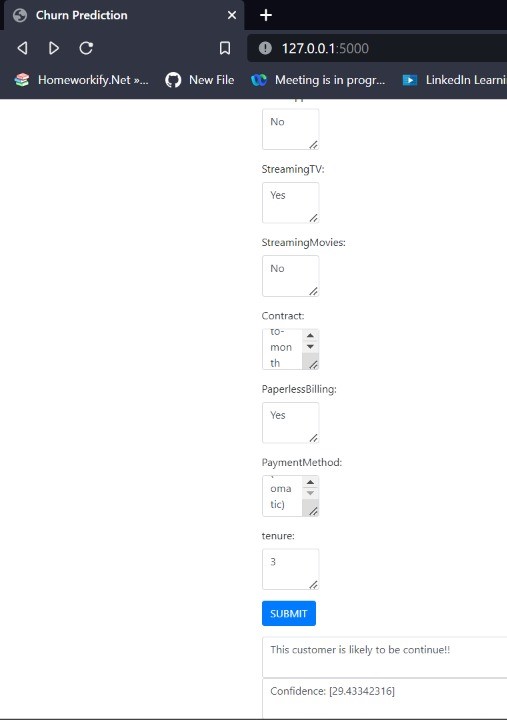
**RESULTS**

* 1. **RESULTS**









**CHAPTER 9**

**CONCLUSION**

##### CONCLUSION

These are some of the quick insights from this exercise:

1.Electronic check medium are the highest churners

2.Contract Type - Monthly customers are more likely to churn because of no contract terms, as they are free to go customers.

3.No Online security, No Tech Support category are high churners

4.Non senior Citizens are high churners

Note: There could be many more such insights, so take this as an assignment and try to get more insights :)

Customer churn analysis allows to minimize acquisition costs and increase marketing efficiency, preparing a solid base for future marketing analysis and campaigns.

In Telecom Company can have a clear view and can provide them some exiting offers to stay in that service.

The obtained results show that our proposed churn model produced better results and performed better by using machine learning techniques.In upcoming time it is necessary to reduce further more features in order to obtain better accuracy and introducing some more machine learning models for better performance Many churn prediction models and techniques have been presented to date.. In this paper, a simple model based on DM techniques was introduced to help a CRM department to keep track its customers and their behavior against churn. A data set of 5000 instances with 23 attributes isused to train and test the model. Using 3 different techniques which are DT, SVM, and NN for classification and simple K Means techniques for clustering results indicate that the best output for the data set in hand is SVM technique. The next stage of the authors’ research will involve performing a deeper analysis into the customer data to try to establish new churn prediction retention model that will use the predicted and clustered data to assign a suitable retention

strategies for each churner type.

Program Outcomes:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 1 | 2 |

**Program Specific Outcomes:**

|  |  |  |
| --- | --- | --- |
| **PSO1** | **PSO2** | **PSO3** |
| 3 | 2 | 1 |

**CHAPTER 10**

**REFERENCES**

#### REFERENCES

1. ***H. Abbasimehr, M. Setak, M. Tarokh. A Neuro-Fuzzy Classifier for Customer Churn Prediction***
2. ***V. Lazarov and M. Capota. Churn Prediction. Business Analytics Course. TUM Computer Science, December 2007***
3. ***.Carlo Vercellis, Business Intelligence: Data Mining and Optimization for Decision Making, John Wiley & Sons, Ltd. 2009 ISBN: 978-0-470-51138-1***
4. ***S. Gotovac. “Modeling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services,” vol. 51, no. 3, pp. 275-283, 2010***
5. ***H. Kim, and C. Yoon, “Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market.” Telecommunications Policy. Vol. 28 No.: PP. 751-765, 2004.***
6. ***M. Hassouna, Agent Based Modelling and Simulation: An Examination of Customer Retention in the UK Mobile Market. PhD thesis, Brunel University, UK, 2012.***
7. ***K. Muata, and O. Bryson, Evaluation of Decision Trees: AMulti Criteria Approach, Computers and Operational Research, 31, 1933-1945, 2004***
8. ***W. Au, C. Chan, and X. Yao, A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction, IEEE transactions on evolutionary computation, 7, 6, 532-545, 2003.***
9. ***R. Behara, W. Fisher, and J. Lemmink, Modelling and Evaluating Service Quality Measurement Using Neural Networks, International journal of operations and production management, 22, 10, 1162-1185, 2002.***