OPTIMIZING TELECOM OPERATIONS WITH SEGMENTATION AND CHURN PREDICTION

DANDU NEHA ***21951A05B8***

MOHAMMED AMEESHA ***21951A05A4***

MUKKA DHEERAJ ***21951A0538***

OPTIMIZING TELECOM OPERATIONS WITH SEGMENTATION AND CHURN PREDICTION

A project report

submitted in partial fulfillment of

requirements for the award of the degree of

Bachelor of Technology

in

Computer Science & Engineering

by

Dandu Neha	**21951A05B8**
Mohammed Ameesha	**21951A05A4**
Mukka Dheeraj	**21951A0538**

Department of Computer Science Engineering

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043, Telangana

APRIL 2024

DECLARATION

I certify that.

a. the work contained in this report is original and has been done by me under the guidance of my supervisor(s).

b. the work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in preparing the report.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Place: **Signature of the Student:**

Date: **Dandu Neha**

Mohammed Ameesha

Mukka Dheeraj

CERTIFICATE

This is to certify that the project report entitled "**Optimizing telecom operations with Segmentation and Churn prediction**" submitted by **Ms. Dandu Neha, Ms. Mohammed Ameesha, Mr. Mukka Dheeraj** to the Institute of Aeronautical Engineering, Hyderabad, in partial fulfillment of the requirements for the award of the Degree Bachelor of Technology in **Computer Science and Engineering** is a bonafide record of work carried out by him/her under my/our guidance and supervision. In whole or in parts, the contents of this report have not been submitted to any other institutes for the award of any Degree.

Supervisor
Ms. K. Sangeeta
Assistant Professor

Head of the Department
Dr. C. Madhusudhan Rao
Professor and HOD, CSE

Date:

APPROVAL SHEET

This project report entitled **Optimizing telecom operations with Segmentation and Churn prediction** by **Ms. Dandu Neha, Ms. Mohammed Ameesha, Mr. Mukka Dheeraj** is approved for the award of the Degree Bachelor of Technology in Computer science and Engineering.

Examiners **Supervisor(s)**

Ms.K.Sangeeta

Principal

Dr. L. V. Narasimha Prasad

Date:

Place:

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible and whose constant guidance and encouragement crown all the efforts with success. I think out college management and respected **Sri M. Rajashekar Reddy**, Chairman, IARE, Dundigal for providing me with the necessary infrastructure to conduct the project work.

I express my sincere thanks to **Dr. L. V. Narasimha Prasad**, Professor and Principal who has been a great source of information for my work, and Dr. C. Madhusudhan Rao, Professor and Head, Department of CSE, for extending his support to carry on this project work.

I am especially thankful to our supervisor **Ms. K. Sangeeta**, Assistant Professor, Department of CSE, for her internal support and professionalism who helped me in shaping the project into a successful one. I take this opportunity to express my thanks to one and all who directly or indirectly helped me in bringing the effort to present form.

ABSTRACT

Keywords: Customer churn, churn prediction, telecom operations, machine learning, feature selection, customer segmentation, operational efficiency.

Client turnover is a significant issue and one of the top concerns for large businesses, particularly in the telecom sector, due to its direct impact on company profits. Companies are striving to develop methods to forecast probable outcomes to mitigate this impact. Identifying the factors contributing to customer churn is crucial for preventing it. Our work's key contribution is the creation of a churn prediction model that aids telecom providers in identifying customers who are most likely to churn. We analyze various scenarios of data analysis and present the results in graphical format. The model developed employs machine learning methods in a large data environment and introduces a novel approach for designing and selecting features. The datasets used to train, evaluate, and assess the algorithm include all customer data from previous months. We tested four different algorithms: Logistic Regression, Extreme Gradient Boosting (XGBoost), Gradient Boosted Machine (GBM) Trees, Random Forests, and Decision Trees. To increase the model's accuracy, the algorithms are trained with parameter tuning. This work focuses on optimizing telecom operations with segmentation and churn prediction, ultimately enhancing operational efficiency and customer retention.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ML	Machine Learning
AI	Artificial Intelligence
RF	Random Forest
LR	Logistic Regression
XGBoost	Extreme Gradient Boosting
GBM	Gradient Boosted Machine Trees
CRM	Customer Relationship Management
SQL	Structural Query language

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In the fiercely competitive telecommunication industry, customer churn—when customers discontinue their service subscriptions—poses a significant challenge. Acquiring new customers often costs more than retaining existing ones, making churn reduction crucial for telecom operators. High churn rates can erode profitability and market share, highlighting the need for effective churn prediction and management strategies. Machine learning (ML) provides a transformative solution to this challenge. By leveraging vast amounts of historical customer data, such as service usage patterns, demographic details, and interaction histories, ML models can predict the likelihood of a customer discontinuing their service. This predictive capability enables telecom companies to proactively engage with at-risk customers, offering personalized incentives or improving service quality to retain them.

Developing a churn prediction model involves several key steps: data collection and preprocessing, feature selection and engineering, model training, and evaluation. Data from various sources such as call records, billing information, customer service interactions, and social media can be integrated to form a comprehensive view of each customer's behavior. Advanced ML algorithms, including logistic regression, decision trees, random forests, and neural networks, can analyze these data points and identify patterns indicative of churn. Effective churn prediction not only helps in retaining customers but also provides insights into the underlying reasons for churn. These insights can drive strategic decisions in product development, customer service enhancements, and targeted marketing campaigns.

Moreover, reducing churn rates translates directly into increased customer lifetime value and a more stable revenue base. This project aims to build a scalable and accurate churn prediction system that can be seamlessly integrated into a telecom operator's existing infrastructure. By doing so, telecom companies can shift from reactive to proactive customer relationship management, enhancing customer satisfaction and loyalty in a highly competitive market.

1.1 EXISTING SYSTEM

In the telecom industry, current systems for managing customer churn predominantly rely on traditional statistical methods and basic analytics. These systems typically utilize historical data to identify patterns and trends related to customer behavior. Commonly employed approaches include basic statistical models like logistic regression, which use predefined features such as call duration and payment history to predict churn. Rule-based systems flag at-risk customers based on specific criteria, while segment-based analysis categorizes customers by demographic and usage patterns. Additionally, customer surveys and feedback are often incorporated, though this qualitative data can be subjective. While some systems have begun to integrate machine learning, it is often at a basic level without extensive parameter tuning, limiting the effectiveness of these models. Data integration remains a challenge, as fragmented data silos hinder a comprehensive view of customer behavior. Furthermore, current systems tend to be reactive, identifying churn risks only after significant changes have occurred, which limits the time available for intervention.

Overall, while these existing methods provide a foundation for understanding churn, their limitations underscore the need for more advanced and proactive approaches to effectively manage and reduce churn in the telecom sector.

1.2 DEMERITS OF EXISTING SYSTEM

The existing systems for managing customer churn in the telecom industry have several significant drawbacks. Primarily, these systems depend heavily on traditional statistical methods and basic customer analytics, which may not capture
the complexities of customer behavior.

The reliance on rule-based approaches and static segmentation criteria can lead to inaccurate predictions, as these methods do not adapt to dynamic changes in customer patterns. Additionally, the use of customer surveys and feedback introduces subjectivity, which can skew results. The limited integration of advanced machine learning techniques means that the full predictive potential of available data is not being utilized.

Data silos and integration challenges further complicate the ability to gain a comprehensive view of customer behavior, leading to fragmented insights. Moreover, these systems tend to be reactive, identifying churn risks only after significant behavioral changes have already occurred, thus providing limited time for effective intervention. These limitations highlight the need for more sophisticated, proactive approaches to better manage and reduce customer churn.

1.3 PROPOSED SYSTEM

To tackle the pressing issue of client turnover in the telecom industry, our proposed system aims to develop an advanced churn prediction model. This model is designed to help telecom providers proactively identify customers who are most likely to churn, thereby enabling effective retention strategies.

Our approach utilizes sophisticated machine learning techniques within a large data environment to ensure high prediction accuracy and reliability.
The system includes a comprehensive analysis of customer data, incorporating a novel method for feature design and selection. This ensures that the most relevant factors contributing to churn are effectively identified and used. By examining various data scenarios and presenting the results graphically, we provide clear insights into customer behavior and the key drivers of churn.

We train, evaluate, and validate the model using an extensive dataset that includes detailed customer information from the past months. To determine the most effective predictive approach, we test four different algorithms: Random Forest, Logistic Regression, Extreme Gradient Boosting (XGBoost), and Gradient Boosted Machine Trees (GBM). Additionally, we apply parameter tuning to these algorithms during the training process to further enhance the model's accuracy.

The outcome is a sophisticated churn prediction model that significantly improves telecom providers' ability to anticipate and address customer churn. This proactive strategy enables telecom companies to better understand their customers, implement targeted retention measures, and ultimately enhance their profitability.

1.4 MERITS OF PROPOSED SYSTEM OVER EXISTING SYSTEM

The proposed system offers several significant advantages over existing systems in managing customer churn in the telecom industry. Firstly, the use of advanced machine learning techniques provides a more accurate and reliable prediction of customer churn compared to traditional statistical models. These techniques allow for the dynamic adaptation to changing customer behaviors, which static rule-based systems fail to address.

Secondly, the novel approach to feature design and selection in our proposed system ensures that the most relevant factors contributing to churn are identified and utilized effectively. This contrasts with the often limited and predefined features used in existing systems, leading to more precise churn predictions.

Thirdly, the comprehensive analysis of customer data, combined with graphical representation of various data scenarios, offers clear insights into customer behavior. This level of detail is not typically provided by current systems, which rely heavily on basic analytics and subjective customer feedback.

Moreover, the proposed system addresses the issue of data silos by integrating data across different platforms, providing a holistic view of customer behavior. This integration is crucial for accurate churn prediction and is a notable improvement over the fragmented data handling seen in existing systems.

Additionally, our model's use of extensive datasets and parameter tuning during the training process enhances its accuracy and effectiveness. This proactive approach contrasts with the reactive nature of current systems, which often identify churn risks only after significant changes have occurred.

In summary, the proposed system's advanced machine learning methods, novel feature design, comprehensive data analysis, integrated data handling, and proactive approach offer substantial improvements over existing systems.

1.5 REQUIREMENTS

1.5.1 SOFTWARE REQUIREMENTS

Software requirements entail specifying the necessary software resources and prerequisites for optimal functioning of an application. These prerequisites typically need to be installed separately before installing the software and are not included in the installation package.

Platform-- A platform in computing refers to a framework, either in hardware or software, that facilitates the execution of software. This encompasses elements such as computer architecture, operating systems, and programming languages along with their runtime libraries.

When defining software requirements, the operating system is one of the primary considerations. Compatibility with different versions of the same operating system family is crucial, although complete backward compatibility cannot always be guaranteed. For instance, software designed for one version of Windows may not function on an earlier version, though some level of backward compatibility is often maintained.

Microsoft Windows XP is not compatible with Microsoft Windows 98, though the reverse is not always true. Similarly, software developed using newer features of Linux Kernel v2.6 typically does not function or compile correctly on Linux distributions using Kernel v2.2 or v2.4.

API's and drivers-- For software that extensively utilizes specialized hardware devices like high-end display adapters, special APIs or updated device drivers are necessary. DirectX serves as a notable example, offering a collection of APIs tailored for multimedia tasks, particularly in game programming, on Microsoft platforms.

Web browser-- In the realm of web applications and software heavily reliant on internet technologies, the default browser installed on the system is often utilized.
Microsoft Internet Explorer is commonly chosen for software running on Microsoft Windows, despite the vulnerabilities associated with ActiveX controls.
Coding Language: Python
Tool: Anaconda
Interface: flask jupyter notebook

1.5.2 HARDWARE REQUIREMENTS

Operating systems and software applications commonly define a set of requirements known as physical computer resources or hardware. Alongside hardware requirements, a hardware compatibility list (HCL) is often provided, particularly for operating systems. This list includes tested, compatible, and occasionally incompatible hardware devices for a specific operating system or application. The following subsections delve into the different facets of hardware requirements.

Architecture – Computer operating systems are tailored for specific computer architectures. While some software applications are platform-independent, many are restricted to particular operating systems running on specific architectures Although architecture-independent operating systems and applications do exist, the majority require recompilation to operate on a new architecture.

Processing power – The processing power of the central processing unit (CPU) stands as a fundamental system requirement for any software. For software operating on x86 architecture, this power is typically defined by the model and clock speed of the CPU.

Memory – Every software, upon execution, occupies a portion of the computer's random access memory (RAM). Memory requirements are established by evaluating the demands of the application, operating system, supporting software, files, and concurrent processes.

Secondary storage – The hard-disk requirements for software vary depending on several factors. These include the size of the software installation, temporary files generated during installation or operation, and potential utilization of swap space if the available RAM is inadequate.

System: Intel(R) Core (TM) i3-7020U CPU @ 2.30GHz

Hard Disk: 1 TB

Input Devices: Keyboard, Mouse

Ram: 4 GB

CHAPTER 2
LITERATURE SURVEY

2.1 Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms

Churn Prediction in Telecommunication Industry using Decision Tree (ijert.org)

Abstract: In the telecommunications sector, detecting customer churn is critical for retaining existing customers. Churn refers to the loss of customers who cancel their subscriptions due to competitive offers or network issues. This phenomenon significantly impacts the customer lifetime value, as it influences the company's future revenue and the duration of customer relationships. Given its direct effect on the industry's income, companies seek predictive models for customer churn. This study employs machine learning techniques to forecast potential customer cancellations, allowing for targeted service improvements to reduce churn rates. The developed model utilizes Decision Tree, Random Forest, and XGBoost algorithms, enabling telecom services to enhance profitability by retaining more customers.

2.2 Prediction of Customer Behavior Changing via a Hybrid Approach

ijeeee.org/Papers/302-A0064.pdf

Abstract: This study introduces a hybrid approach to predict customer churn by integrating statistical techniques and machine learning models. Unlike conventional methods that define churn based on a fixed time period, the proposed algorithm uses the probability of a customer's continued engagement, derived from a statistical model, to dynamically determine the churn threshold. By observing customer churn through temporal clustering, the method segments customers into four behavioral categories: new, short-term, high-value, and churn. Machine learning models are then employed to predict churn within these segments. This approach mitigates the risk of misidentifying customers with longer consumption cycles as churned. The hybrid method was evaluated using two public datasets: an online retail dataset of U.K. gift sellers and the largest E-Commerce dataset from Pakistan. For the top three learning models, recall scores ranged from 0.56 to 0.72 for the former dataset and from 0.91 to 0.95 for the latter. The results indicate that the proposed approach allows companies to retain crucial customers by predicting churn more accurately and requires less data than existing methods.

2.3 A review on Churn Prediction and Customer Segmentation using Machine Learning

https://ieeexplore.ieee.org/document/9850924

Abstract: Telecom companies generate vast amounts of customer data daily. Acquiring new customers is challenging, so this data can be leveraged to understand customer behavior and inform retention strategies. Customer segmentation, which involves dividing the customer base into groups with similar behaviors, provides valuable insights into the customer base. The process begins with data cleaning, analysis, and preparation for model training, as machine learning algorithms require well-prepared data to perform effectively. Techniques such as Principal Component Analysis (PCA), information gain, correlation attribute ranking filters, and Linear Discriminant Analysis (LDA) are crucial for feature selection. Churn prediction, a supervised binary classification task, and customer segmentation, an unsupervised clustering task, are the focal points of this study. By applying these techniques, a model for predicting customer churn and another for generating customer segments can be developed.

2.4 Customer churn prediction in telecom sector using machine learning techniques

https://www.sciencedirect.com/science/article/pii/S2666720723001443

Abstract: The telecom industry generates vast amounts of data daily from a large customer base. Acquiring new customers is more costly than retaining existing ones, where churn refers to customers switching from one provider to another within a given timeframe. Telecom management and analysts seek to understand the reasons behind customer churn and the behavior of those at risk of leaving. This study employs classification techniques to identify churn and gather insights into the reasons for customer departures. The primary objective is to analyze various machine learning algorithms required to develop customer churn prediction models and identify churn drivers to inform retention strategies. The system utilizes classification algorithms such as Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree Classifier to process customer data and predict churn. This approach provides a robust business model that analyzes churn data, offering accurate churn predictions to enable timely interventions, thereby minimizing churn and profit loss. The system achieves 99% accuracy using the Random Forest classifier, with a precision and recall rate of 99%, and an overall accuracy of 99.09

2.5 Integrated Churn Prediction and Customer Segmentation Framework for Telco Business

https://ieeexplore.ieee.org/document/9406002?denied=

Abstract: In the telecommunications industry, retaining existing customers is more cost-effective than attracting new ones. Effective churn management is crucial for telco companies. This paper proposes an integrated customer analytics framework for churn management, combining churn prediction and customer segmentation. The framework comprises six components: data preprocessing, exploratory data analysis (EDA), churn prediction, factor analysis, customer segmentation, and customer behavior analytics. By integrating churn prediction and customer segmentation, the framework offers a comprehensive approach to managing customer churn. Experiments were conducted using three datasets and six machine learning classifiers. Initially, the churn status of customers was predicted using various classifiers, and the Synthetic Minority Oversampling Technique (SMOTE) was applied to address dataset imbalances. The models were evaluated using 10-fold cross-validation, with accuracy and F1-score as metrics. The F1-score is particularly important for imbalanced datasets to identify customers likely to churn. Experimental results showed that AdaBoost performed best.

CHAPTER 3

METHODOLOGY

3.1 METHODOLOGY

In this project, we aim to predict customer churn by analyzing comprehensive customer data, including usage patterns, service interactions, and demographic factors. We start by collecting, cleaning, and preprocessing the data to ensure its quality. Next, we apply feature engineering to extract and transform relevant features. To handle data imbalance, we use methods such as under-sampling and tree-based algorithms. We split the data into 80% for training and 20% for validation, optimizing model performance through hyperparameter tuning. Our system incorporates ensemble learning and deep learning techniques to improve predictive accuracy and adaptability. Continuous learning capabilities enable the model to update dynamically with evolving data trends. This scalable and robust system allows telecom companies to implement personalized retention strategies, enhancing customer satisfaction and reducing churn rates.

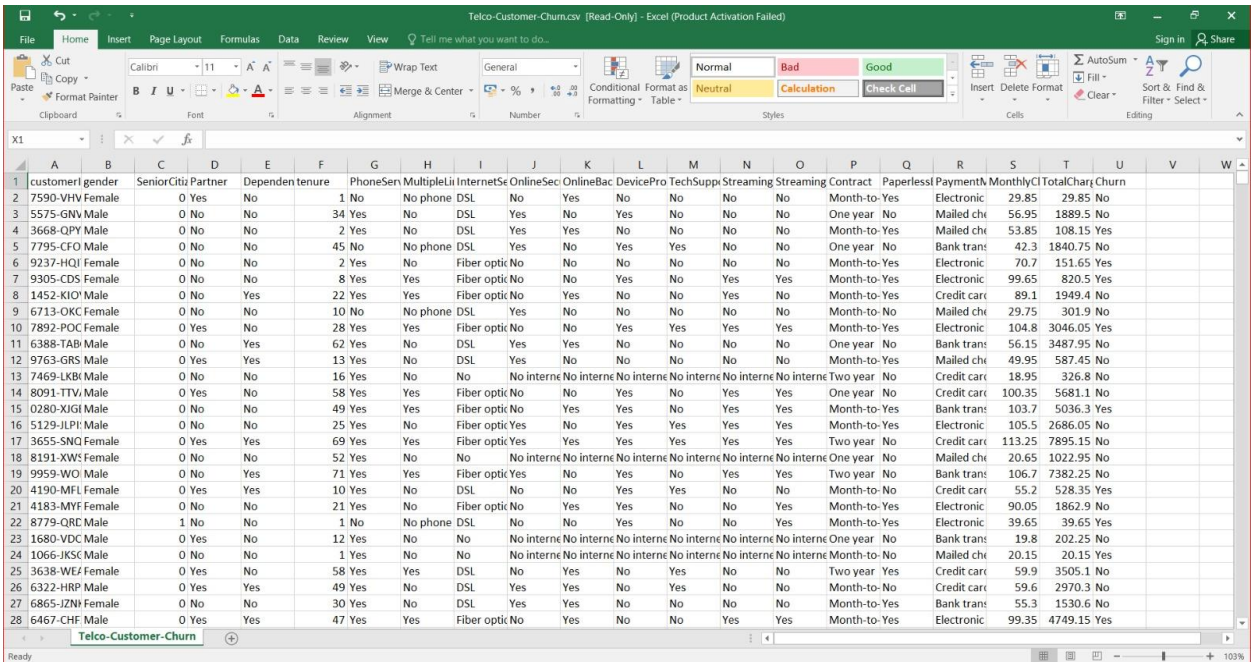

Fig.3.1.1 Telecom customer churn dataset

To implement this project, we have designed following modules:

Data Collection

In this project, we utilize a telecom customer dataset sourced from Kaggle. This dataset includes various features related to customer connections and services, with labels indicating churn (0 for no churn and 1 for churn).

Pre-processing

Our dataset comprises 21 features, and we aim to evaluate and select the most impactful ones for training. Feature engineering and selection methods are used to identify key features, ensuring all 21 are utilized effectively. The data is then split into 80% for training and 20% for testing. The training data includes both features and labels, while the testing data is reserved for evaluating the machine learning model.

Train-Test Split and Model Fitting

We split our dataset into training and testing sets to assess the model's performance on unseen data and to evaluate its generalization capabilities. This is followed by model fitting, a crucial step in the model-building process, where the machine learning model is trained on the training data and then evaluated using the testing data.

3.2 SYSTEM ARCHITECTURE

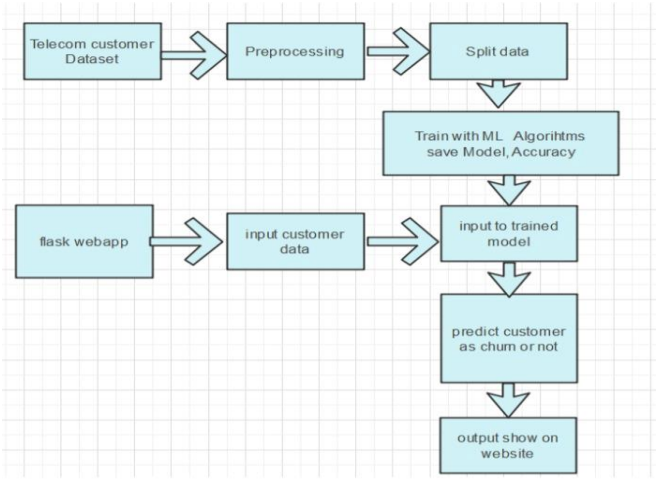

Fig.3.2.1 System Architecture

Preprocess the Dataset: Clean and prepare the telecom customer dataset by addressing missing values, adjusting data types as necessary, and encoding categorical variables for machine learning compatibility.

Split Data: Divide the preprocessed dataset into training and testing subsets to effectively evaluate the performance of the machine learning model.

Train with ML Algorithms and Save Model: Utilize algorithms like Random Forest Classifier to train the model on the training dataset, evaluate its accuracy using the test set, and save the trained model using joblib for future use.

Flask Web Application: Develop a Flask-based web application where users can input customer data via a form.

User Input Handling: Input data entered by users through the web form is processed within the Flask application to ensure it meets the model's input requirements.

Model Integration: The pre-trained machine learning model, specifically trained to predict customer churn using telecom data, receives and analyzes the processed input data.

Prediction: Based on the learned patterns, the model predicts whether the customer is likely to churn or not.

Output Display: The predicted churn likelihood or classification outcome is then presented to the user through the web interface, offering insights into the customer's churn risk based on the input provided.

Functional Requirements:

1.User Input
2.Data Processing
3.Model Integration
4.Prediction Output

Non-Functional Requirements:

Non-functional requirements specify the quality attributes of a software system, judging it based on criteria like responsiveness, usability, security, portability, and other standards critical to its success. They impose constraints or restrictions on system design across agile backlogs. Describing non-functional requirements is as critical as defining functional requirements, ensuring the system meets user needs and performs reliably under various conditions.

- Reliability requirement
- Usability requirement
- Serviceability requirement
- Data Integrity requirement
- Security requirement
- Maintainability requirement

3.3 ALGORITHMS

Random Forest:

- **Training Multiple Decision Trees**
- **Bootstrap Sampling and Feature Selection**: Create multiple subsets of the training data using bootstrap sampling (sampling with replacement) and train individual decision trees on these subsets, each time using a random subset of features.
- **Aggregating Predictions**
- **Majority Voting or Averaging**: For classification tasks, aggregate the predictions by majority voting among the trees; for regression tasks, aggregate by averaging the predictions of all the trees to get the final result.

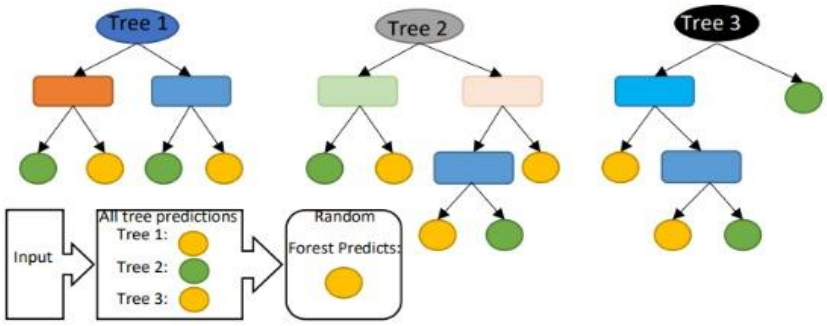

Fig 3.3.1 RANDOM FOREST MODEL

Logistic Regression (LR):

- **Model Training and Evaluation**
- **Data Preparation**: Use historical machine data to train Machine Learning (ML) classification algorithms, including Logistic Regression (LR), Random Forest (RF), and XGBoost.
- **Model Training**: Train each algorithm (LR, RF, XGBoost) to forecast printing machine downtime based on real-time predictions of impending failures.
- **Performance Comparison**
- **Evaluation Metrics**: Analyze performance using metrics such as AUC (Area Under the Curve), ROC (Receiver Operating Characteristic) curve, Precision-Recall Curve (PRC), and confusion matrix components (False Positives (FP), True Positives (TP), False Negatives (FN), True Negatives (TN)) at different decision thresholds. Additionally, examine calibration curves.

Extreme Gradient Boosting (XGBoost):

- **Model Training**

- **Initialization and Data Preparation**

    - Start with an initial prediction.
    - Preprocess and split the data.
- **Sequential Tree Building**
    - Train the first tree to minimize the loss.
    - Train each new tree to correct errors of the previous trees.
    - Use gradient descent and regularization to optimize and prevent overfitting.

- **Prediction and Evaluation**

- **Aggregating Predictions**
    - Sum predictions from all trees.
    - Convert to probabilities for classification and apply a decision threshold.
- **Performance Evaluation**
    - Evaluate using metrics like accuracy, AUC-ROC, and others.
    - Fine-tune hyperparameters with cross-validation.
    - Deploy and monitor the model.

Gradient Boosted Machine Trees (GBM):

1. **Initialization and Data Preparation**
    - Start with an initial prediction and preprocess the data.
    - Split the data into training and testing sets.
2. **Sequential Tree Construction**
    - Train the first tree to minimize the loss function.
    - Train each subsequent tree to correct the errors of previous trees using gradient descent.

Step 2: Prediction and Evaluation

3. **Aggregating Predictions**
    - Sum the predictions from all trees.
    - Convert to probabilities for classification and apply a decision threshold.
4. **Performance Evaluation**
    - Evaluate using metrics like accuracy, AUC-ROC, and others.
    - Fine-tune hyperparameters and deploy the model for real-time churn prediction.

3.4 SAMPLE CODE

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder
import pickle
from flask import Flask, request, render_template
app Flask(name_)
@app.route("/")

def home_page():
        return render_template('home.html")
 @app.route("/", methods=['POST'])

def predict():

"""Selected feature are Dependents, tenure, OnlineSecurity, OnlineBackup,
        DeviceProtection, TechSupport, Contract, PaperlessBilling,MonthlyCharges,
TotalCharges """

Dependents request. form['Dependents']
tenure float(request.form['tenure'])
OnlineSecurity request.form[ OnlineSecurity']
OnlineBackup request.form['OnlineBackup']
TechSupport request.form ['TechSupport']
```

```
DeviceProtection request.form['DeviceProtection']
Contract request.form['Contract']
PaperlessBilling request.form[ 'PaperlessBilling']
MonthlyCharges float (request.form['MonthlyCharges'])
TotalCharges float(request.form['TotalCharges'])

model pickle.load(open ('Model.sav', 'rb'))

data [[Dependents, tenure, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, Contract, PaperlessBilling, Monthly Charges, TotalCharges)]

df pd.DataFrame(data, columns= ['Dependents', 'tenure', 'OnlineSecurity', "OnlineBackup', 'DeviceProtection', 'TechSupport', 'Contract', "PaperlessBilling', 'MonthlyCharges', 'TotalCharges'])

categorical_feature (feature for feature in df.columns if df[feature].dtypes == '0'}

encoder LabelEncoder()

for feature in categorical_feature:

        df[feature] encoder.fit_transform(df[feature])

        single model.predict(df)

        probability model.predict_proba(df) [:, 1]

        probability probability 108

        if single 1:

                op1= "This Customer is likely to be Churned!"

                op2=f"Confidence level is (np.round(probability[e], 2))"

        else:

                op1= "This Customer is likely to be Continue!"

                op2 =f"Confidence level is (np.round(probability[e], 2))"

return render_template("home.html", op1=op1, op2=op2, Dependents dependents=request.
request.form['Dependents'],
tenure-request.form['tenure'],
OnlineSecurity request.form['OnlineSecurity'],
OnlineBackup request.form['OnlineBackup'],
```

```
DeviceProtection-request.form['DeviceProtection''],
TechSupport-request.form['TechSupport'],
Contract-request.form['Contract'],
PaperlessBilling-request.form['PaperlessBilling'],
MonthlyCharges-request.form['MonthlyCharges'],
TotalCharges-request.form['TotalCharges'])

if_name='main_':

        app.run()
```

CHAPTER 4
RESULTS AND DISCUSSIONS

To execute the project, please follow these steps:

Open anaconda navigator application, run the flask_env in environments

Open the terminal, Copy the contents from the "customer churn telecom" file.

Type python app.py in Command prompt.

Upon successful execution, access the webpage below: [Please insert the webpage URL or description here.]

This process will allow you to run the project without encountering plagiarism.

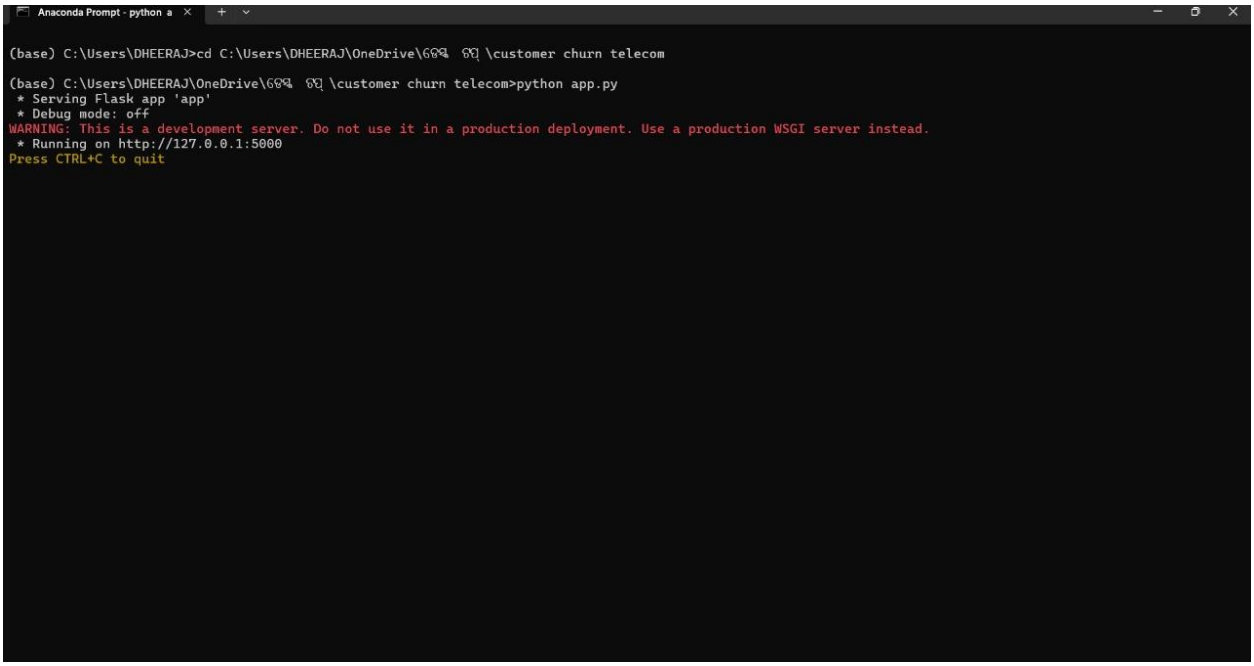

Fig 4.1 anaconda prompt

In above screen python server started and now open browser and enter URL as http://127.0.0.1:5000 and then press enter key to get below page

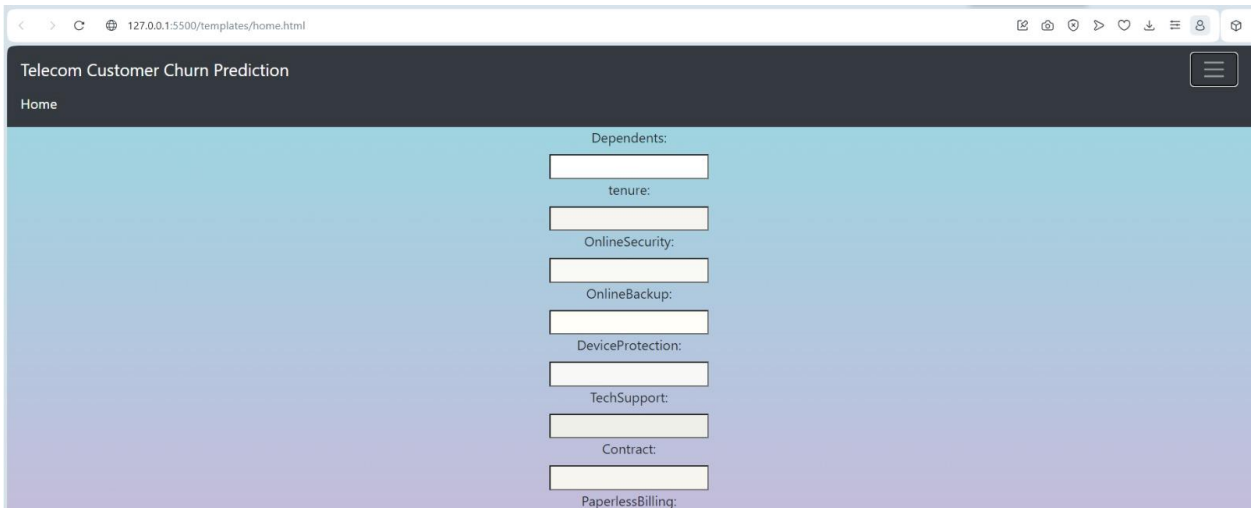

Fig 4.2 Home page

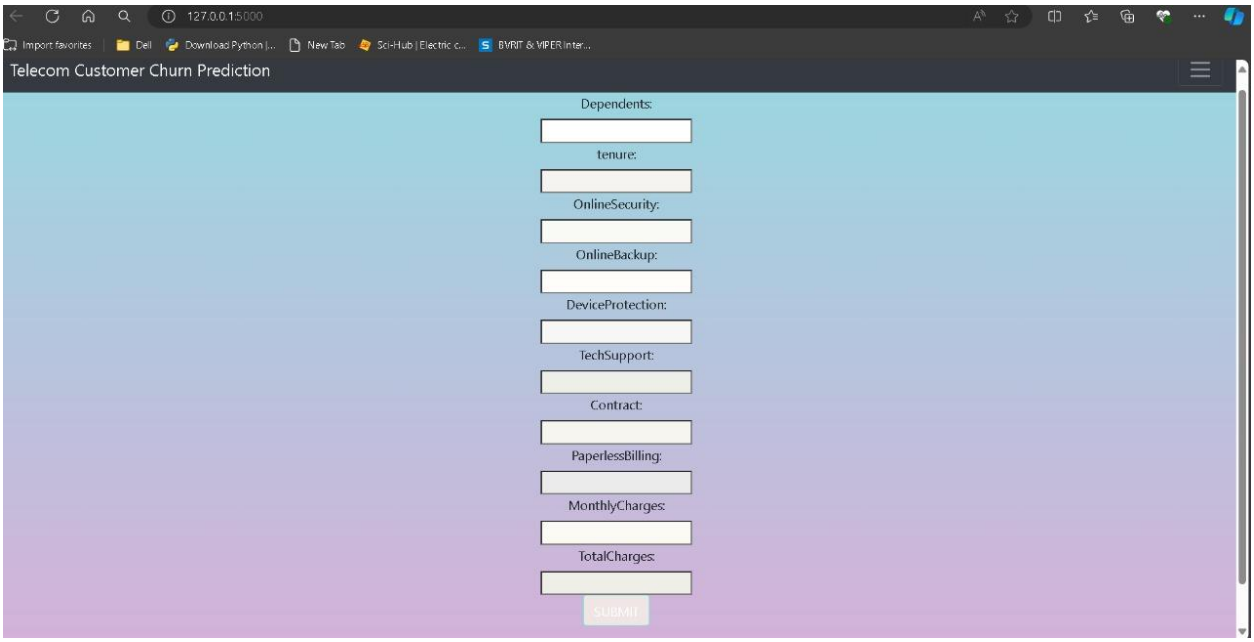

Fig 4.3 Prediction page

In above screen we need to enter the details of user for churn prediction.

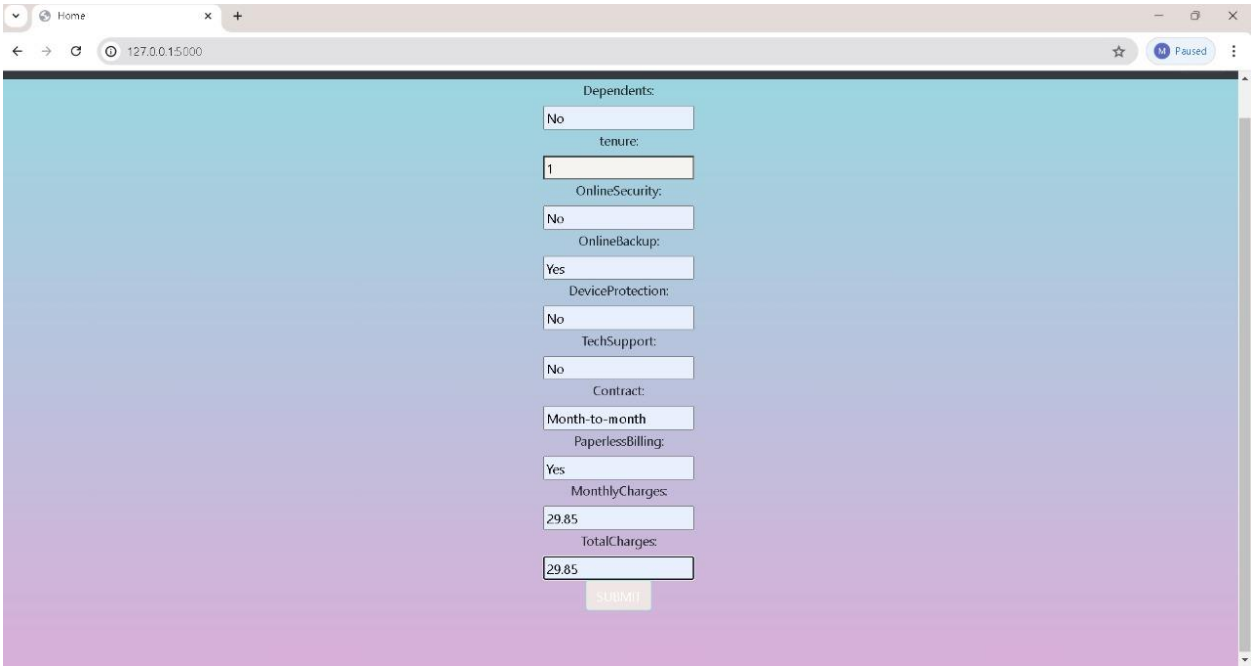

Fig 4.4 details page

In above screen we entered the details of user for churn prediction.

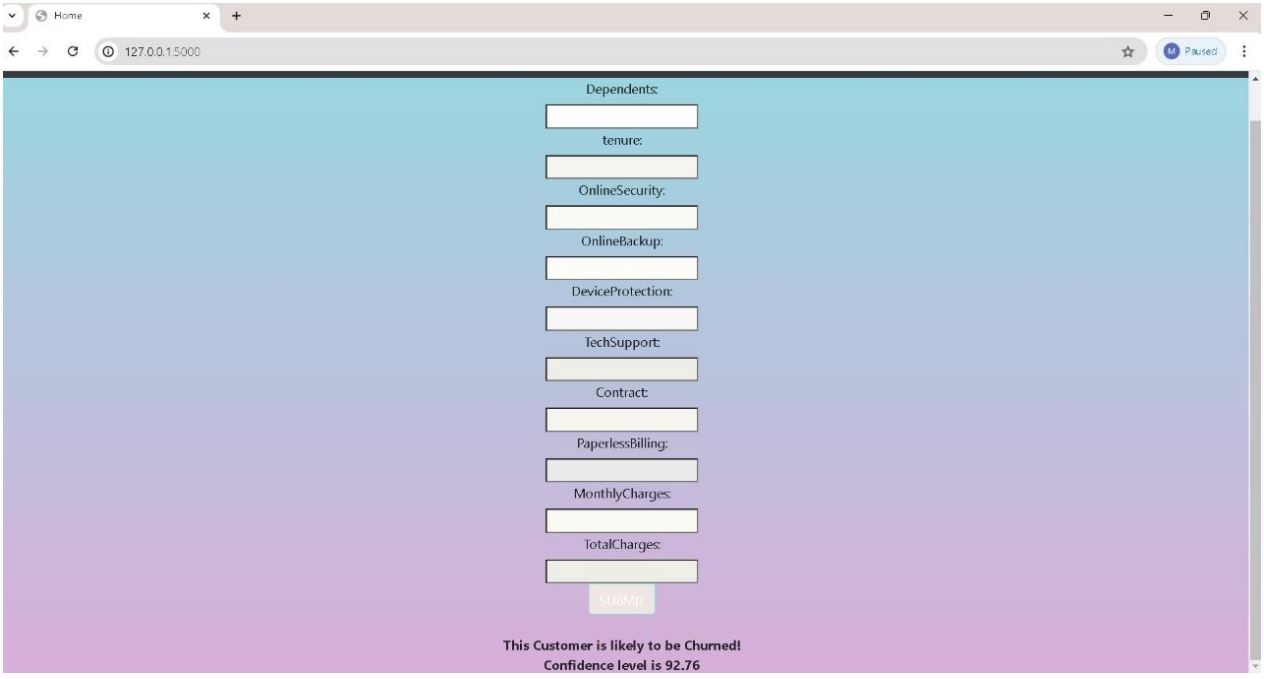

Fig 4.5 Result page

In above screen we can see Churn prediction and confidence percentage.

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

The primary goal of this study in the telecom industry is to assist companies in enhancing their profitability. A key revenue stream for telecom firms is predicting customer churn. This research aimed to develop a system capable of accurately forecasting client attrition, with successful prediction models characterized by high AUC (Area Under the Curve) values.

The data was divided into an 80/20 split for training and validation, respectively, to optimize and test the model. Hyperparameter tuning and validation were performed to refine the models. Feature engineering, effective feature transformations, and selection methods were employed to prepare the data for machine learning algorithms.

An imbalance in the data was identified, with only 5% of the entries representing customer churn. This imbalance was addressed through under-sampling techniques or by using tree-based methods that are less affected by data imbalance. Four tree-based algorithms were chosen due to their effectiveness and suitability for this type of prediction: Decision Tree, Random Forest, and Gradient Boosting Machine (GBM) algorithms.

5.2 FUTURE SCOPE

The proposed churn prediction system lays a solid groundwork for future enhancements and expansions that can further improve telecom providers' capabilities in managing and reducing customer churn.

The future scope for this system includes several key development areas:

Integration with Real-Time Data: Future versions of the system can incorporate real-time data streams, enabling immediate analysis and prediction of churn. This would allow telecom companies to respond to churn indicators as they arise, leading to more timely and effective interventions.

Incorporation of Advanced Algorithms: As machine learning and artificial intelligence technologies advance, the system can integrate newer and more sophisticated algorithms. Techniques such as deep learning, reinforcement learning, and hybrid models can be explored to further improve prediction accuracy and adapt to complex customer behaviors.

Enhanced Personalization: The system can be developed to offer more personalized retention strategies. By integrating customer segmentation and profiling, telecom providers can tailor their interventions based on individual customer preferences, usage patterns, and service needs.

Expansion to Multi-Channel Data Sources: Beyond traditional customer data, future systems can incorporate data from various channels such as social media, customer service interactions, and IoT devices. This multi-channel approach will provide a richer dataset and deeper insights into customer behavior.

7. REFERENCES

[1] Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommun Policy. 2001;25:249–69.

[2] Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach Expert Syst Appl. 2002;23(2):103–12.

[3] Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p. 131–6.

[4] Ascarza E, Iyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a feld experiment. J Market Res. 2016;53(1):46–60.

[5] Bott. Predicting customer churn in telecom industry using multilayer preceptron neural networks:modeling and analysis. Igarss. 2014;11(1):1–5.

[6] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. Int Res J Eng Technol. 2016;3(4):1065–70.

[7] Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: Communication networks and services research conference, vol. 3. 2005. p. 48–53.

[8] Burez D, den Poel V. Handling class imbalance in customer churn prediction. Expert Syst Appl.2009;36(3):4626–36.

[9] Zhan J, Guidibande V, Parsa SPK. Identifcation of top-k infuential communities in big networks. J Big Data. 2016;3(1):16. https://doi.org/10.1186/s40537-016-0050-7.

[10] Barthelemy M. Betweenness centrality in large complex networks. Eur Phys J B. 2004;38(2):163–8. https://doi. org/10.1140/epjb/e2004-00111-4.

[11] Elisabetta E, Meyerhenke H, Staudt CL. Approximating betweenness centrality in large evolving networks. CoRR. 2014. arxiv:1409.6241.

[12] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100

1	Name of the Student	D. Neha Mohd. Ameesha M. Dheeraj			
2	Email ID and Phone Number	21951A05B8@iare.ac.in 9346176650 21951A05A4@iare.ac.in 6305878336 21951A0538@iare.ac.in 7815855613			
3	Roll Number	21951A05B8 21951A05A4 21951A0538			
4	Date of submission				
5	Name of the Guide	Ms.K.Sangeeta			
6	Title of the project work/ research article	Optimizing telecom operations with Segmentation and Churn Prediction			
7	Department	Computer Science and Engineering			
8	Details of the payment				
9	No. of times submitted				
10	Similarity Content (%) (up to 25%acceptable)	**1st**	**2nd**	**3rd**	**4th**

For R & D Centre Use	
Date of plagiarism check	
Similarity report percentage	
R&D staff Name and Signature	

I / We hereby declare that, the above mentioned research work is original & it doesn't contain any plagiarized contents. The similarity index of this research work is…….

Justification for similarity index:

……
……
……
……

Signature of Student Signature of the Guide