

**CUSTOMER CHURN PREDICTION WEBSITE**

***Mini Project - Report submitted by***

|  |  |  |
| --- | --- | --- |
| **AKSHITA** | **AVANI ACHARYA** | |
| (4NM21CS015) |  | (4NM21CS039) |
| **Dr. Sarika Hegde**  Professor | 6th Semester B.E. ***Under the Guidance of*** | **Ms. Jayapadmini Kanchan**  Assistant Professor Gd-I |

*In partial fulfilment of the requirements for the award of*

*the Degree of*

**Bachelor of Engineering in Computer Science Engineering**

*from*

***Visvesvaraya Technological University, Belagavi***

Department of Computer Science Engineering

NMAM Institute of Technology, Nitte - 574110

(An Autonomous Institution affiliated to VTU, Belagavi)

**MAY 2024**



(i)

**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

# CERTIFICATE

*Certified that the Mini project work entitled*

## “CUSTOMER CHURN PREDICTION WEBSITE”

*is a bonafide work carried out by*

**AKSHITA AVANI ACHARYA**

(4NM21CS015) (4NM21CS039)

*of 6th Semester B.E. in partial fulfilment of the requirements for the award of*

***Bachelor of Engineering Degree in Computer Science Engineering*** *prescribed by* ***Visvesvaraya Technological University, Belagavi*** *during the*



*year*

***2023-2024***

*.*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

***Signature of the Guide Signature of the HOD***

**Viva Voce Examination**

**Name of the Examiners Signature with Date**

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(ii)

### ABSTRACT

Customer churn, a critical metric in industries like telecommunications, signifies the rate at which customers stop using a company's services. Retaining existing customers is more cost-effective than acquiring new ones, especially considering annual churn rates of 15-25%. Despite the substantial financial losses associated with churn, predictive analytics offers a solution by identifying high-risk customers. This project aims to predict customer churn by leveraging predictive modelling as it would help businesses in strategically allocating resources to retain customers that are most likely to discontinue their services, optimizing their retention initiatives for maximum effectiveness.

(iii)

### ACKNOWLEDGEMENT

We wish to extend our heartfelt gratitude to those whose guidance and encouragement have propelled this project forward.

Foremost, our sincere appreciation goes to our project mentors, **Dr. Sarika Hegde**, Professor and **Ms. Jayapadmini Kanchan**, Assistant Professor, Department of Computer Science and Engineering, and Department of Computer Science and Engineering. Their unwavering support and insightful counsel have been invaluable throughout this endeavour.

We are also indebted to **Dr. Jyothi Shetty**, Head of the Department of Computer Science and Engineering at NMAM Institute of Technology, Nitte, for presenting us with this opportunity.

Furthermore, our heartfelt thanks are due to **Dr. Niranjan. N. Chiplunkar**, Principal of NMAM Institute of Technology, Nitte, whose continuous encouragement continues to inspire us in our pursuit of excellence.

We consider ourselves fortunate to have been surrounded by such dedicated and inspiring individuals who have made this project both educational and rewarding. Our gratitude knows no bounds, and we remain committed to continuing our efforts under their guidance.

(iv)

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| Cover page | i |
| Certificate page | ii |
| Abstract | iii |
| Acknowledgement | iv |
| Table of contents | v |
| Introduction | 1 |
| Literature survey | 2-3 |
| Dataset Description | 4 |
| Problem statement | 5 |
| Methodology | 6 |
| Implementation | 7-8 |
| Results | 9-13 |
| Conclusion | 14 |
| References | 15 |

(v)

### INTRODUCTION

**BRIEF OVERVIEW OF THE PROJECT:**

This project aims to establish “**ChurnSage**” a website that is designed to help businesses in the telecommunications, subscription services, or any industry where customer retention is critical. The project will start by collecting and cleaning historical customer data, including demographics, usage patterns, and interactions with the service. After preprocessing the data, various machine learning algorithms such as logistic regression, decision trees, random forests, KNN or XGBoosting will be employed to train predictive models. Once the models are trained and validated, they will be deployed into a production environment where they can be used to predict churn for new customers.

**OBJECTIVES OF THE PROJECT:**

* To determine the churn rates and identify the trends among customers who discontinue services versus those who remain active
* Investigate potential correlations between customer churn and other feature attributes of the dataset
* Analyze customer churn patterns and preferences based on the types of services offered.
* Evaluate the impact of various features and services on predicting customer churn
* Compare and select the most effective predicting model for accurately predicting churn behaviour.

**IMPORTANCE OF THE PROJECT:**

Predicting customer churn is essential for businesses to retain profitability and competitiveness in dynamic markets. By anticipating which customers are likely to leave, companies can implement proactive retention strategies to minimize revenue loss and maintain customer loyalty. Additionally, churn prediction enables organizations to allocate resources efficiently, focusing efforts on retaining high-value customers. By staying ahead of churn trends, businesses can adapt their strategies to evolving market conditions and sustain long-term growth. Ultimately, investing in churn prediction empowers companies to foster stronger customer relationships and drive sustainable business success.

1

### LITERATURE SURVEY

In our research on customer churn analysis, we referred to three key papers to gain insights into effective data analysis and model selection. These papers provided valuable knowledge on how to approach churn prediction, including the use of data mining techniques and machine learning models such as decision trees, logistic regression, XGBoost etc. By leveraging the findings from these papers, we were able to develop a comprehensive understanding of the methods and strategies essential for effective customer churn analysis in the telecom industry.

Paper 1:

Title: Customer Churn Analysis in Telecom Industry

Authors: Kiran Dahiya, Surbhi Bhatia

The paper examines the issue of customer churn in the telecom industry, emphasizing its significance for business sustainability and profitability. It delves into the role of Customer Relationship Management (CRM) in retaining customers and highlights the importance of data mining techniques for improving marketing strategies and network management. The authors propose a novel framework for churn prediction, outlining modules for data acquisition, preparation, preprocessing, extraction, and decision-making. They implement decision tree and logistic regression models on different datasets to predict churn, with decision tree outperforming logistic regression in accuracy. The study concludes that effective churn management is crucial for telecom companies, and decision tree models can be particularly useful in predicting and managing customer churn.

Paper 2:

Title: Customer churn prediction system: a machine learning approach

Authors: Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha, Pratyush Sethi

This paper discusses the challenges faced by telecommunication companies in retaining customers due to increasing competition and the importance of customer churn prediction in reducing potential churn. The authors emphasize the significance of machine learning techniques in predicting churn and propose a machine learning-based approach to tackle this problem. The paper explores various machine learning algorithms such as Logistic Regression, Naive Bayes, Support Vector Machine, Decision Trees, Random Forest, XGBoost, CatBoost, and AdaBoost for churn prediction.

2

Feature selection techniques, data pre-processing, and model evaluation using metrics like precision, recall, accuracy, and F-measure are also discussed. The results indicate that ensemble learning techniques like AdaBoost Classifier and XGBoost Classifier outperform other algorithms with an AUC score of 84%, showcasing their effectiveness in churn prediction. The paper concludes by highlighting the importance of machine learning in addressing churn prediction challenges and suggests future research directions in reinforcement learning and deep learning for further improving prediction accuracy.

Paper 3:

Title: A comparison of machine learning techniques for customer churn prediction.

Authors: T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, K. Ch. Chatzisavvas

This paper provides a detailed comparison of five machine learning classification methods for predicting customer churn in the telecommunication sector. The methods examined include Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees (DTs), Naïve Bayes, and Logistic Regression, along with boosting variations. The study finds that ANNs are effective for complex churn prediction tasks and outperform DTs and Logistic Regression. SVMs can outperform DTs and ANNs depending on data characteristics. Naïve Bayes shows good results, particularly compared to DT-C4.5, and Logistic Regression performs well with proper data transformation. Evaluation is based on precision, recall, accuracy, and F-measure from confusion matrices, and boosting algorithms significantly improve classifier performance. The study suggests further exploration of boosting algorithms, simulation schemes, and larger datasets for more robust results.

3

### PROBLEM STATEMENT

Customer churn is a significant concern for businesses, impacting revenue and growth. This project aims to develop a predictive model using machine learning to anticipate customer churn and integrate it into a user-friendly web application interface “ChurnSage”. The primary objective is to address the challenge of identifying customers at risk of leaving, enabling businesses to implement targeted retention strategies promptly.

Key challenges include:

* Early Prediction: Developing a robust predictive model capable of accurately identifying customers likely to churn, enabling businesses to intervene proactively and retain valuable customers.
* Accessibility: Creating an intuitive application interface that allows businesses to easily input customer data and receive actionable insights on churn risk, facilitating timely decision-making.
* Interpretability: Ensuring transparency and interpretability of the predictive model's results to build trust among users and enable informed decision-making.
* Personalization: Integrating features within the application to provide personalized recommendations and retention strategies tailored to individual customer profiles, enhancing effectiveness in mitigating churn.

By addressing these challenges, the project aims to empower businesses to reduce churn rates, improve customer retention, and ultimately, sustain long-term profitability and growth.

4

### DATASET DESCRIPTION

In the chosen dataset, there are 7043 customers and 21 features in the dataset.

Customer ID: A unique identifier for each customer.

**Demographic Information:**

* Gender: Indicates whether the customer is male or female.
* SeniorCitizen: Indicates whether the customer is a senior citizen (1 for yes, 0 for no).
* Partner: Indicates whether the customer has a partner (Yes or No).
* Dependents: Indicates whether the customer has dependents (Yes or No).

**Services Subscribed:**

* PhoneService: Indicates whether the customer has a phone service (Yes or No).
* MultipleLines: Indicates whether the customer has multiple phone lines (Yes, No, or No phone service).
* InternetService: Specifies the customer’s internet service provider (DSL, Fiber optic, or No).
* OnlineSecurity: Indicates whether the customer has online security (Yes, No, or No internet service).
* OnlineBackup: Indicates whether the customer has online backup (Yes, No, or No internet service).
* DeviceProtection: Indicates whether the customer has device protection (Yes, No, or No internet service).
* TechSupport: Indicates whether the customer has tech support (Yes, No, or No internet service).
* StreamingTV: Indicates whether the customer has streaming TV (Yes, No, or No internet service).
* StreamingMovies: Indicates whether the customer has streaming movies (Yes, No, or No internet service).

**Customer Account Information:**

* Tenure: The number of months the customer has been with the company.
* Contract: Specifies the contract term of the customer (Month-to-month, One year, Two years).
* PaperlessBilling: Indicates whether the customer has opted for paperless billing (Yes or No).
* PaymentMethod: Specifies the customer’s payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic)).
* MonthlyCharges: The amount charged to the customer monthly.
* TotalCharges: The total amount charged to the customer.

**Churn Prediction Target:**

* Churn: Indicates whether the customer has left within the last month (Yes or No)

5

#### METHODOLOGY

1. **Exploratory Learning and Understanding:**

To grasp the machine learning principles essential for the project, we engaged in learning YouTube videos, various tutorials, and research materials on customer churn prediction. This allowed us to understand the algorithms, predictive modelling mechanics, and relevant terminologies.

1. **Dataset Acquisition and Preprocessing:**

We obtained the Telco customer churn dataset from Kaggle, a platform hosting datasets. We conducted exploratory data analysis (EDA) on the dataset, including visualization techniques, to comprehend feature characteristics and distributions.

1. **Model Development and Evaluation:**

We experimented with various machine learning models to predict customer churn. The models included Logistic Regression, Gaussian Naive Bayes, Random Forest, Gradient Boosting, XGBoost, and Kernel SVM. We evaluated each model's performance using metrics such as precision, recall, F1 score, and ROC AUC score.

1. **Model Conversion and Integration:**

Upon achieving satisfactory performance, we converted the selected model (XGBoost) to the. pkl format for seamless integration into the deployment framework.

1. **Web Application Development:**

We developed a customer churn prediction web application using Streamlit and Flask. The application features user-friendly interfaces for user registration, login, and prediction. It employs the integrated XGBoost model to accurately predict customer churn based on input features.

The outlined methodology encapsulates the key steps involved in developing and deploying the customer churn prediction system using machine learning techniques. By leveraging exploratory learning, dataset preprocessing, model development, and web application integration, we successfully built a predictive tool to assist businesses in mitigating customer churn.

6

#### IMPLEMENTATION

1. **Dataset Acquisition and Preprocessing:**

* Obtain the Telco customer churn dataset from a reliable source like Kaggle.
* Perform exploratory data analysis (EDA) to understand the dataset's structure, feature distributions, and any missing values.
* Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features as required.

2. **Model Development and Evaluation:**

* Implement various machine learning models such as Logistic Regression, Random Forest, KNN, Decision Trees and XGBoost using libraries like scikit-learn.
* Split the dataset into training and testing sets and evaluate each model's performance using appropriate metrics like accuracy, precision, recall, and F1-score.
* Experiment with hyperparameter tuning techniques like grid search to optimize model performance.

3. **Model Conversion and Integration:**

* Select the best-performing model based on evaluation metrics and convert it to a serialized format (e.g., pickle file) for easy integration into the deployment framework.
* Ensure compatibility of the model with the deployment environment and any constraints related to memory usage or processing speed.

4. **Web Application Development:**

* Develop a web application ‘ChurnSage’ for customer churn prediction using a framework like Flask. - Design user-friendly interfaces for user registration, login, and inputting relevant features for churn prediction.
* Integrate the serialized machine learning model into the web application to enable real-time predictions.
* Deploy the web application on a web server (Streamlit) to make it accessible to users.

7

**Tech Stack Used:**

* Programming Languages: Python (for machine learning model development), HTML/CSS

(for web application development)

* Machine Learning Libraries: scikit-learn, pandas, NumPy, matplotlib and seaborn
* Web Framework: Flask
* Serialization: Pickle

5. **Continuous Monitoring and Improvement:**

* Implement mechanisms for monitoring model performance in production, such as tracking prediction accuracy and detecting concept drift.
* Collect user feedback and iterate on the application interface and features to improve user experience and prediction accuracy over time.

This plan outlines a structured approach to implementing customer churn analysis using machine learning techniques and deploying it in a web application for practical use, along with the technology stack involved.

8

#### RESULTS

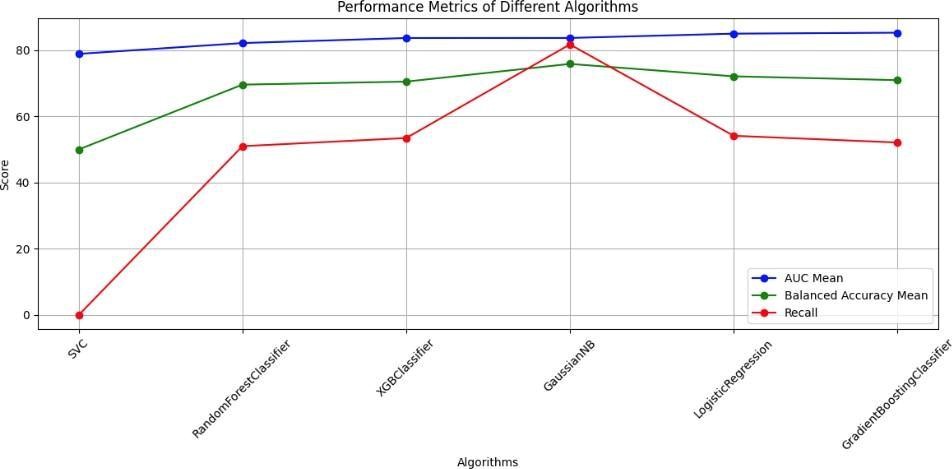


Fig 12.1: Performance metrics of different algorithms before the after SMOTEEN (over-sampling to deal with class imbalances)

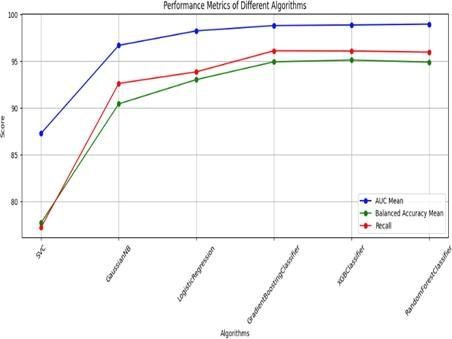


Fig 12.2: Performance metrics of different algorithms before the after SMOTEEN (over-sampling to deal with class imbalances)

9

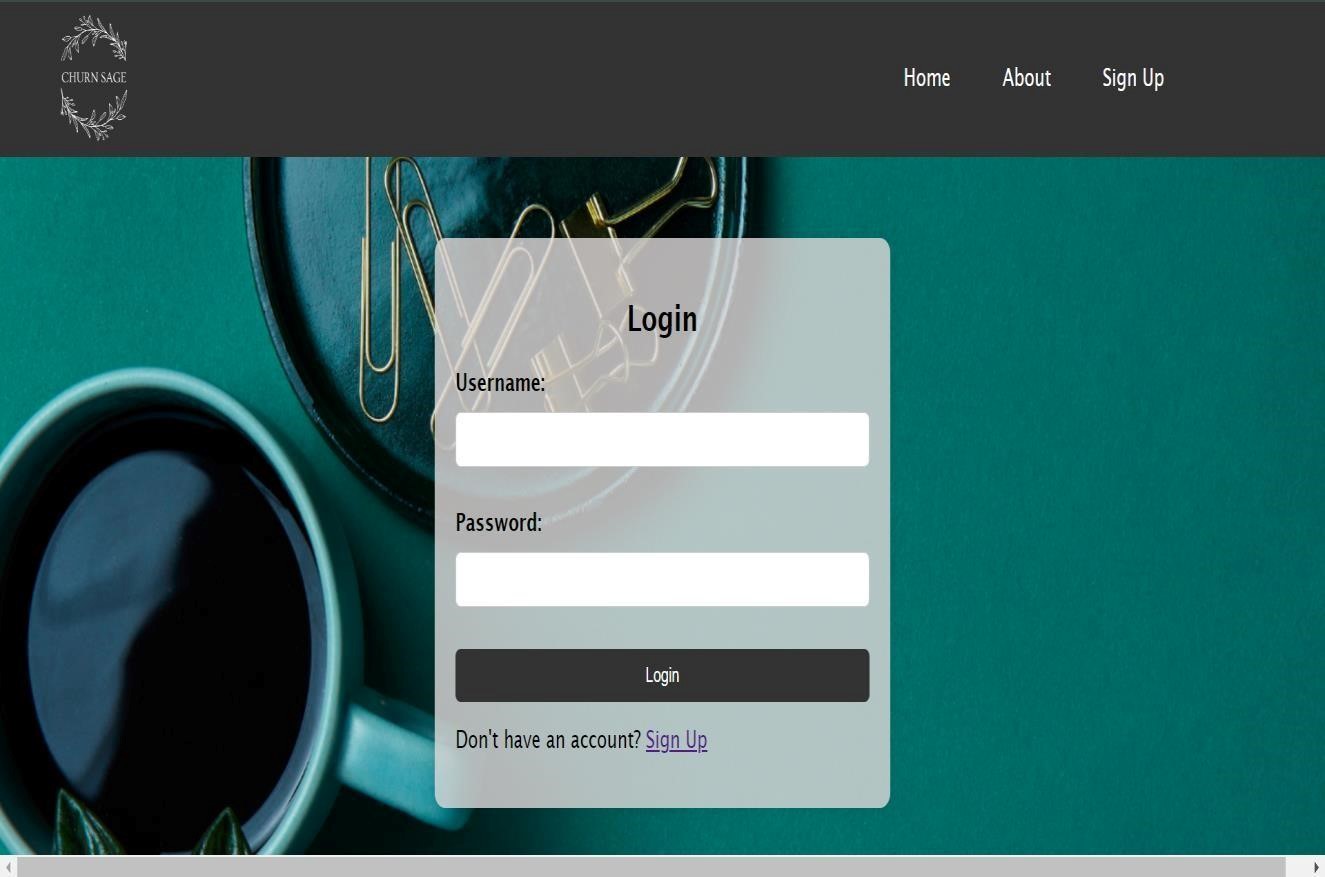


Fig 12.3: Login page of the ChurnSage website

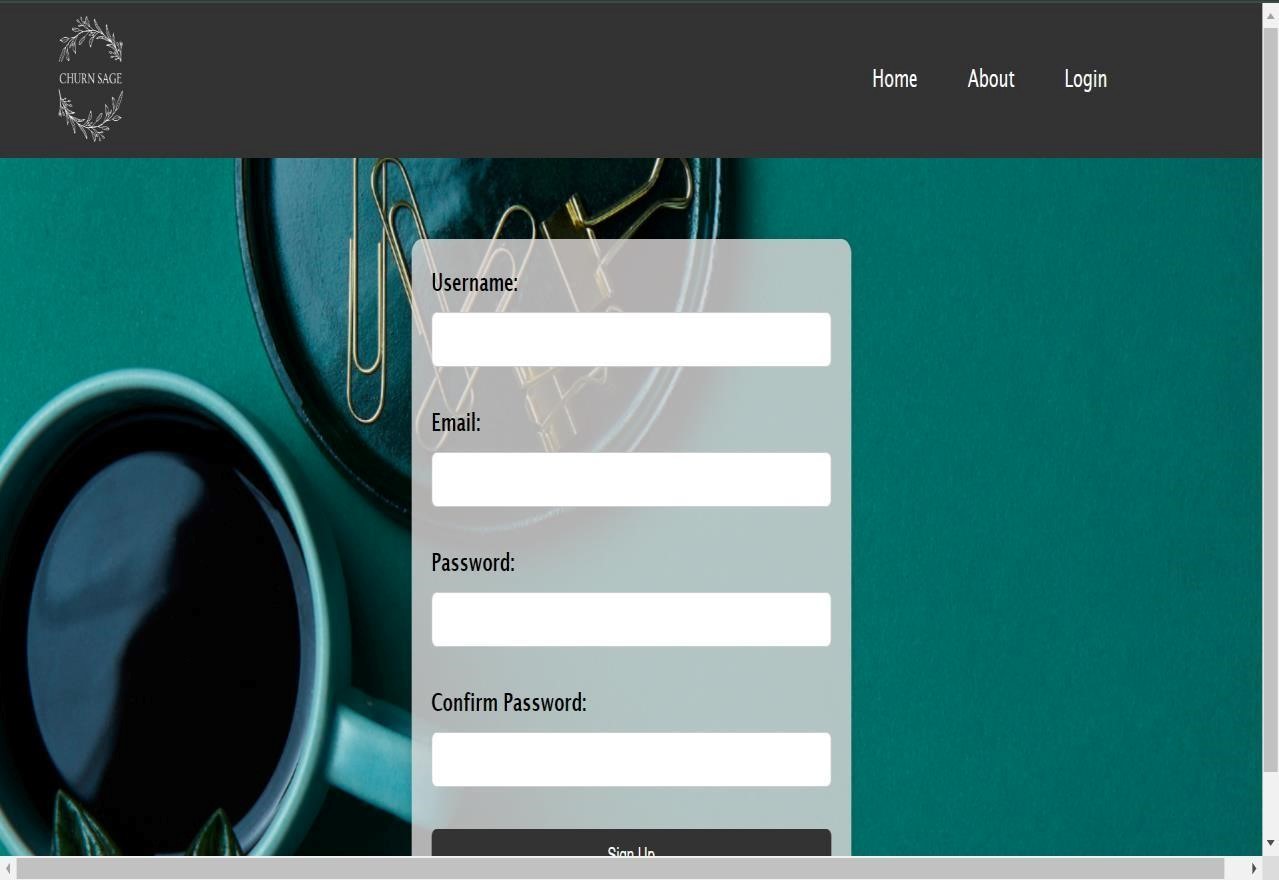


Fig 12.4: Signup page of the ChurnSage website

10

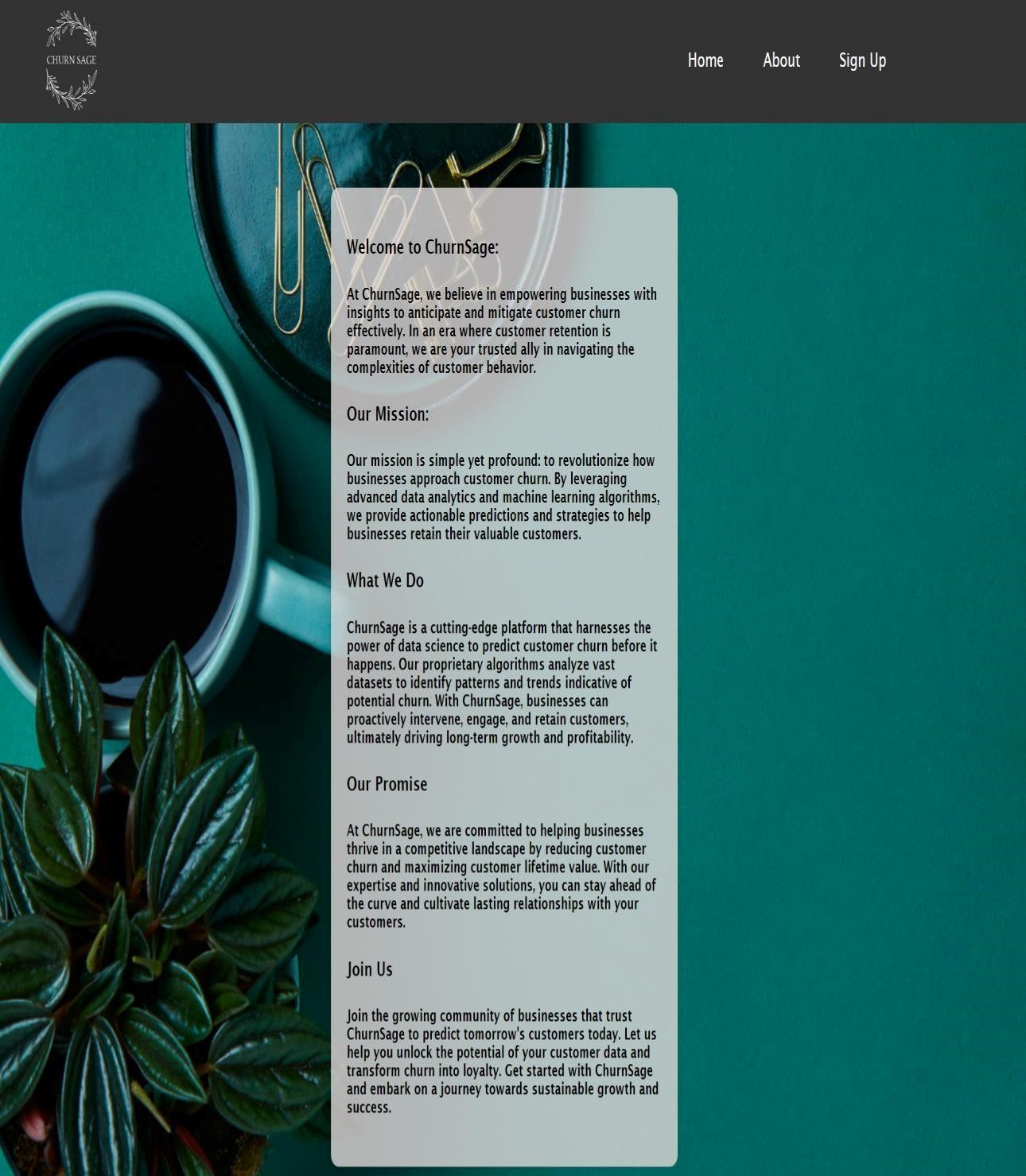
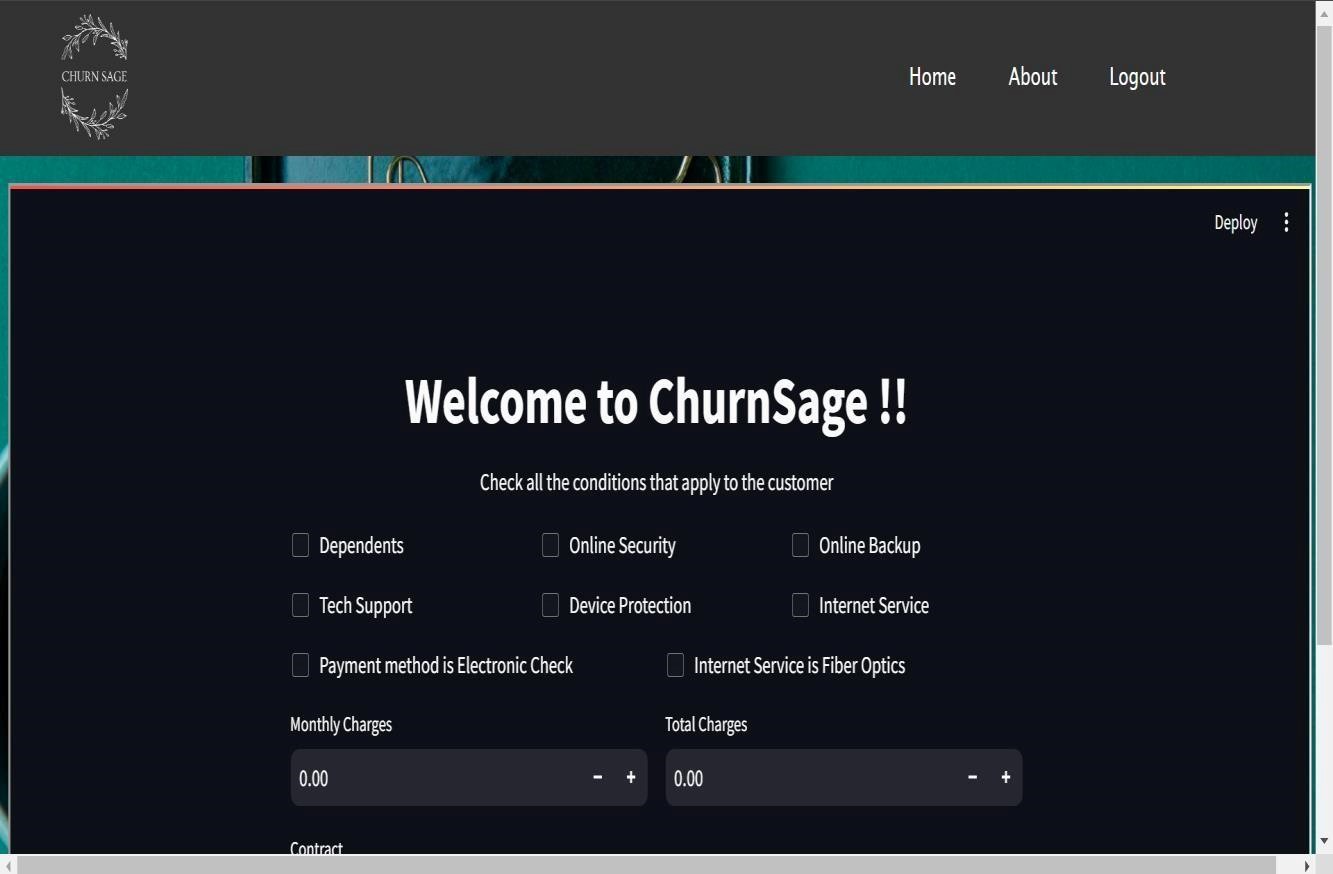
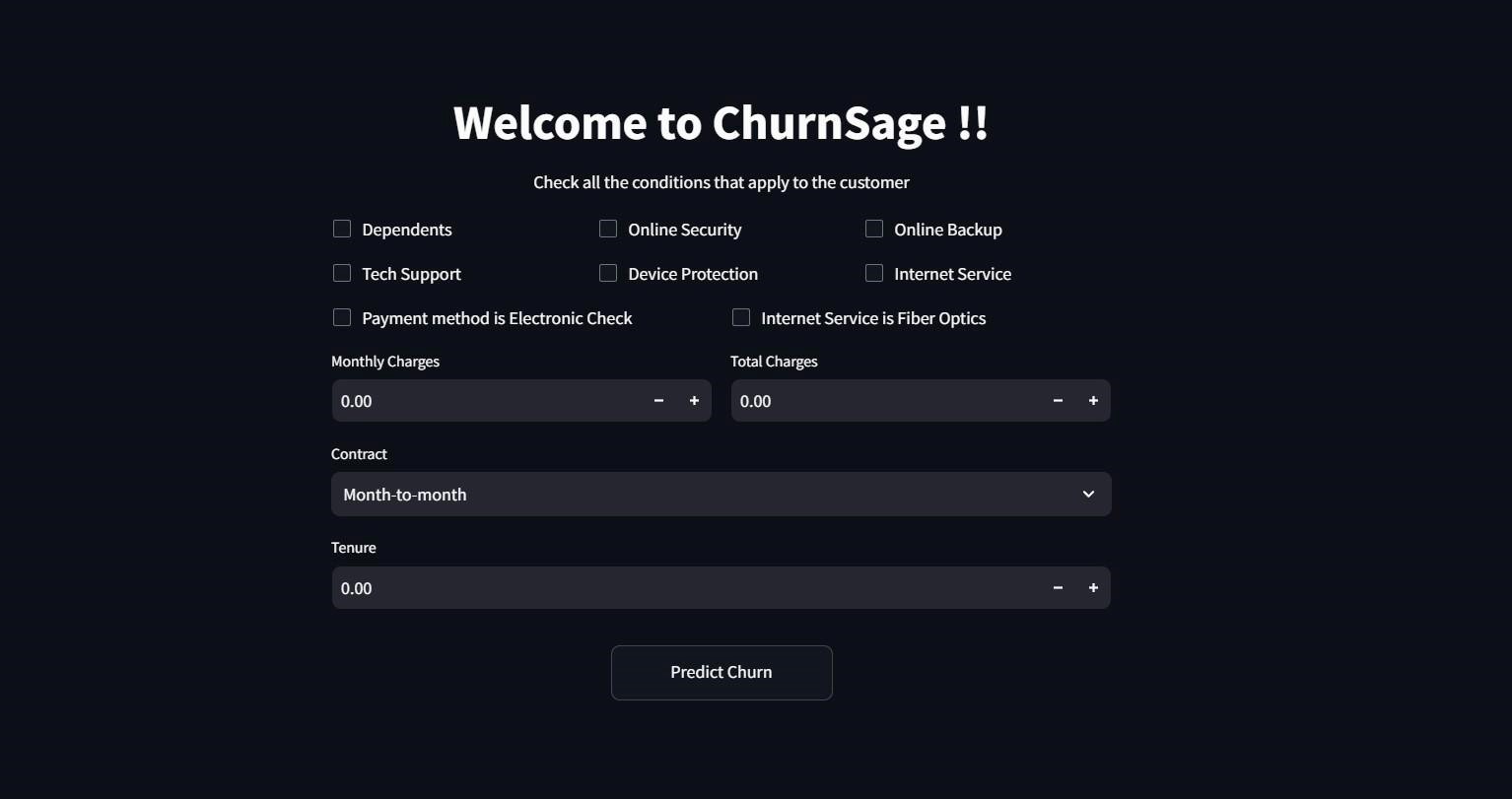


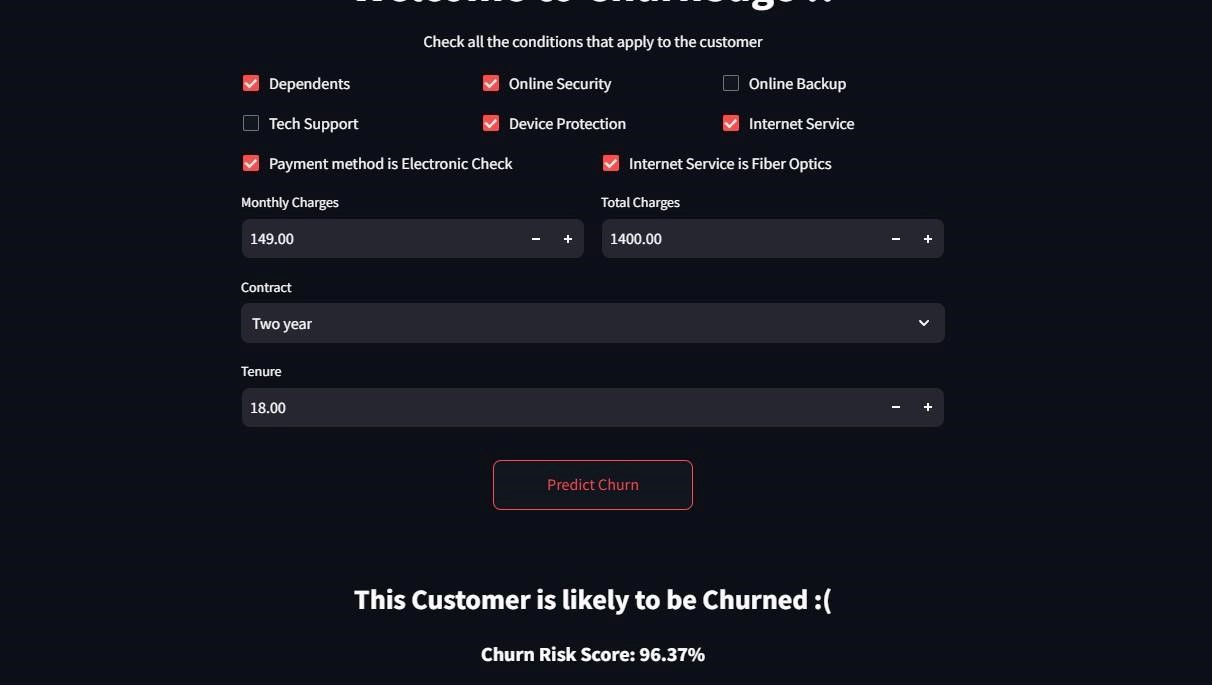
Fig 12.5: About page of the ChurnSage website

11

Fig 12.6: Home page of the ChurnSage website

Fig 12.7: Prediction Interface

12



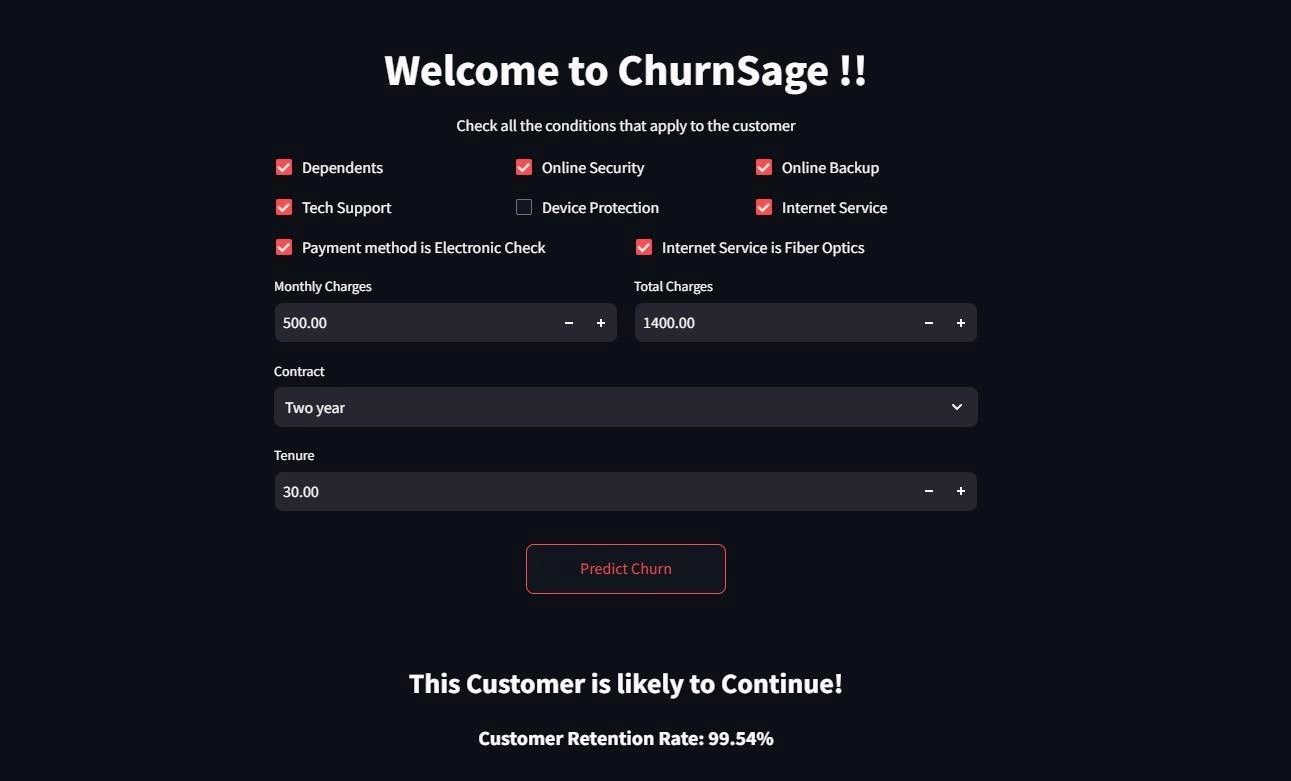


Fig 12.8 and 12.9: Churn prediction results

13

#### CONCLUSION

In conclusion, the customer churn prediction project marks a significant advancement in the domain of customer relationship management, offering a proactive solution for businesses to anticipate and mitigate customer attrition. By leveraging a diverse set of machine learning models and integrating them into a streamlined web application, the project addresses critical challenges in customer retention, including early identification of churn risks, personalized intervention strategies, and data-driven decision-making.

The development of robust predictive models, including Logistic Regression, Random Forest, Gradient Boosting, XGBoost, among others, empowers businesses to forecast customer churn with remarkable accuracy. By analyzing historical customer data and extracting meaningful insights, these models enable businesses to identify patterns and factors contributing to churn, thereby facilitating targeted retention efforts and resource allocation.

Moreover, the deployment of a user-friendly web application using Streamlit provides businesses with an intuitive platform to access churn prediction insights and take proactive measures to retain valuable customers. Through features like user registration, login, and prediction, the application streamlines the process of customer churn analysis and empowers businesses to make informed decisions in real-time.

By ensuring transparency and interpretability in the prediction results, the project builds trust and confidence in the predictive models, fostering a culture of data-driven decision-making within organizations. Furthermore, by facilitating the integration of predictive analytics into existing business processes, the project promotes a proactive approach to customer relationship management, ultimately leading to improved customer retention rates and enhanced business sustainability.

Overall, this project represents a significant step forward in the realm of customer churn prediction, offering businesses a powerful tool to anticipate and address customer attrition proactively. By harnessing the capabilities of machine learning and deploying them in a user-centric web application, the project enables businesses to stay ahead of the curve and build stronger, more enduring relationships with their customers.

14

**REFERENCES**

1. <https://ieeexplore.ieee.org/abstract/document/7359318>
2. <https://link.springer.com/article/10.1007/s00607-021-00908-y>
3. <https://www.sciencedirect.com/science/article/abs/pii/S1569190X15000386>
4. https://www.sciencedirect.com/science/article/abs/pii/S0308596106000760
5. <https://youtu.be/yBDHkveJUf4?si=VyZox0kqzpnM-2Mk>
6. <https://youtu.be/hQl2wyJvK5k?si=rq3BkEreOoE2EpQg>
7. <https://youtu.be/nZRygaTH2MA?si=Bj6Q0ANzfSNViae1>
8. <https://userpilot.com/blog/churn-prediction/>
9. <https://www.paddle.com/resources/churn-prediction>
10. <https://www.geeksforgeeks.org/top-10-algorithms-every-machine-learning-engineer-should-know/>

15

