Dynamic Customer Churn Prediction Using Machine Learning Algorithm

[1]Khan Asif, [2]Khan Anas, [3]Prof. Sneha Sankhe

[1] [2] UG student, [3]Assistant Professor

[1] [2] [3] Department of Information Technology,

[1] [2] [3] Theem College of Engineering, Boisar, India

Abstract: Customers are the most important assets in any industry since they are considered as the main profit source. Companies are working hard to survive in today's competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues: (1) acquiring new customers, (2) upselling the existing customers, and (3) increasing the retention period of customers. However, comparison of these strategies has shown that retaining an existing customer costs much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer's churn. Customer churn is a term that refers to the loss of a client or customer—that is, when a customer ceases to interact with a company or business. Similarly, the churn rate is the rate at which customers or clients are leaving a company within a specific period of time. Customer churn is one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customers that could churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn.

Keywords—Regression, Decision Trees, K-Nearest Neighbors and Support Vector Machine algorithms.

I. INTRODUCTION

A churn prediction model is developed in this project which can assist companies to predict customers who are most likely to churn. It uses machine learning techniques such as Logistic Regression, Decision Trees, K-Nearest Neighbors and Support Vector Machine algorithms to identify the primary determinants of customer churn along with the algorithm fit for such predictions. The dataset contains demographic details of customers, their total charges and they type of service they receive from the company. It comprises of churn data of over 7000 customers spread over 21 attributes obtained from Kaggle. Further on this investigation, the usage of the above mentioned algorithms is described for predicting customer churn. To conclude, the results of the algorithms for predicting customer churn are outlined in the form of accuracy, recall score, precision, f1scoreand kappa metrics using interactive graphs.

II. LITERATURE SURVEY

A literature survey on bank customer churn prediction is an essential step in understanding the latest developments and trends in this field. Churn prediction in the banking industry is crucial for retaining customers, improving customer satisfaction, and increasing profitability. Here is an overview of some key research papers and themes in this area:

1) **Churn Prediction Models:**

a. "Customer Churn Prediction in Retail Banking: A Case Study in the German Market" by Hadden and Tiwari, 2019.

b. "Predicting Customer Churn in Retail Banking: A Model Selection Approach" by Yeh and Hsu, 2019.

2) **Data and Feature Engineering:**

a. "Feature Engineering for Predictive Modeling of Customer Churn in Retail Banking" by Zhang and Chen, 2020.

b. "Customer Churn Prediction Using Ensemble Methods with Feature Engineering in Retail Banking" by Duan et al., 2018.

3) **Machine Learning Techniques:**

a. "Customer Churn Prediction in Retail Banking: A Comparative Analysis of Machine Learning Algorithms" by Khashei et al., 2016.

b. "A Deep Learning Approach for Customer Churn Prediction in the Banking Industry" by Li et al., 2019.

4) **Time Series Analysis:**

a. "Predicting Bank Customer Churn in Advance Using Time Series Data" by Tsai et al., 2017.

b. "Churn Prediction in Retail Banking Using Time Series Analysis" by Rastogi and Agrawal, 2019.

5) **Customer Segmentation:**

a. "Customer Segmentation and Churn Prediction in Retail Banking: A Comparative Study" by Cheng et al., 2019.

b. "A Customer Segmentation-Based Approach for Churn Prediction in the Banking Sector" by Akın et al., 2019.

6) **Customer Sentiment and Social Media:**

a. "Bank Customer Churn Prediction Based on Social Media Sentiment Analysis" by Coussement and Van den Poel, 2018.

b. "Customer Churn Prediction in the Banking Industry: A Social Media Approach" by Xu et al., 2017.

7) **Ethical and Regulatory Aspects:**

a. "Fairness and Bias in Customer Churn Prediction" by Verma and Rubin, 2018.

b. "Challenges and Opportunities for Customer Churn Prediction in the Context of GDPR" by Garcia et al., 2019.

8) **Real-World Case Studies:**

a. "Churn Prediction and Prevention in a Retail Bank: A Case Study" by Chen and Cheng, 2016.

   b. "Reducing Customer Churn in the Banking Sector: A Case Study of a Large Bank" by Li et al., 2017.

III. SYSTEM ARCHITECTURE

The system architecture for bank customer churn prediction typically involves a combination of data collection, pre-processing, model development, and deployment. Here's a high-level overview of the system architecture for predicting customer churn in a bank:

1. Data Collection

2. Data Pre-processing

3. Data Storage

4. Machine Learning Model Development

5. Model Evaluation.

6. Deployment

7. Customer Interaction

8. Reporting and Visualization

9. Feedback Loop

10. Compliance and Security

.
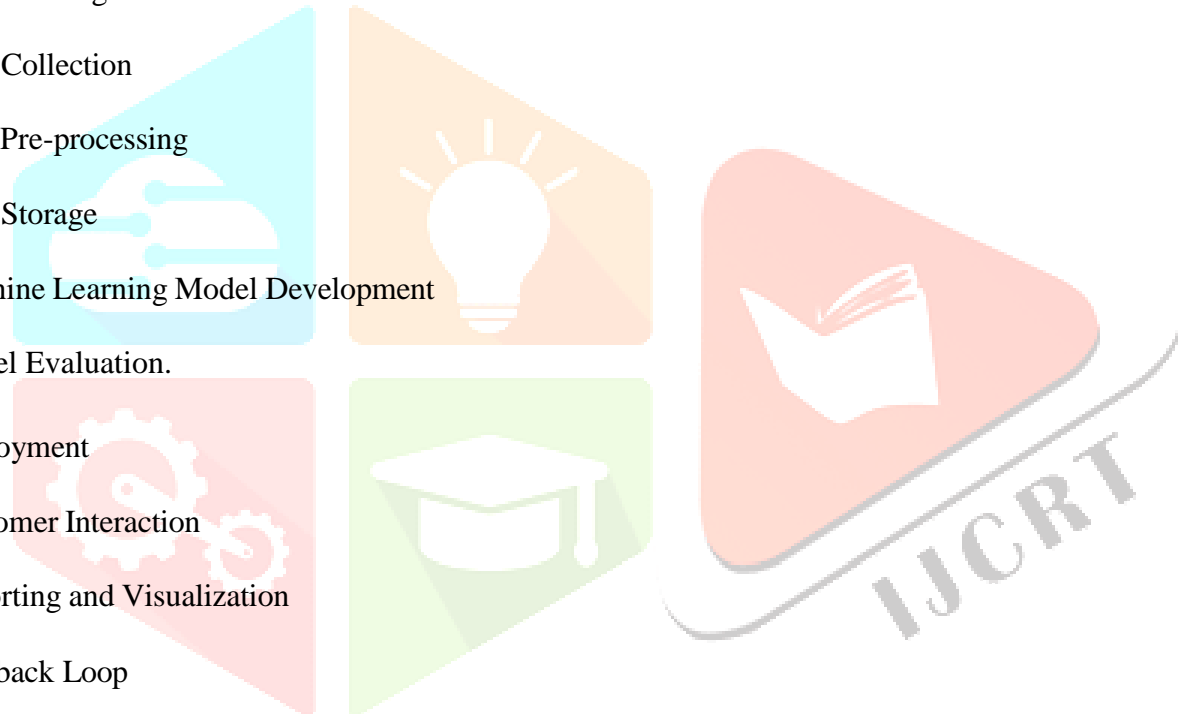
A. Design

To build the architecture required by a project, we use incremental process model in which we test each prototype and then clubbed with the actual model on observing a correct output. Each Prototype is built along each model and then clubbed together with an actual model. In this way project is built using incremental model.

B. Requirement Analysis

In the software development lifecycle, demand analysis is one of the most important phases. It's used to identify and define the software. For any software design, there are different kinds of conditions to be fulfilled to insure the smooth handling of the processes. easily defined conditions are important labels on the road to a successful design.

Hardware Requirements

- Windows 10 or latest version
- 8 GB RAM
- Intel i5 Processor
- 100GB free Hard Disk
- Wi-Fi Router

Software Requirements
- Anaconda
- IDE
- Operating System: Windows 10/8/7 (incl. 64-bit), Mac OS, Linux
- MySql

Functional Requirements

- Python
- Heroku

Non-functional Requirements
- Security
- Privacy

C. Proposed System

Predicting bank customer churn is a crucial task for financial institutions. Churn prediction helps banks retain existing customers and reduce customer acquisition costs. Below is a proposed system for bank customer churn prediction:

Gather historical customer data, including transaction history, demographic information, customer interactions, and any other relevant information.

Include information on customers who have churned (closed accounts) and those who have stayed. Clean and preprocess the data by handling missing values, outliers, and data quality issues.

Encode categorical variables (e.g., one-hot encoding) and normalize numerical features. Create relevant features that can help in churn prediction. Some common features include:

- Customer transaction frequency and volume

- Average account balance

- Customer tenure

- Customer demographics

- Interaction history (customer service calls, visits to the bank, etc.)

Perform EDA to gain insights into the data, such as identifying patterns, correlations, and potential factors influencing churn.
Split the dataset into a training set and a testing set for model evaluation.

Choose appropriate machine learning or deep learning models for churn prediction. Common models include:
- Logistic Regression

- Random Forest

- Gradient Boosting (e.g., XGBoost)

- Neural Networks

Train the selected models on the training dataset.

Evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and ROC AUC.
Use techniques like cross-validation to ensure robustness of the model. Optimize model hyperparameters to improve performance.
Deploy the trained model in a production environment to make real-time churn predictions.

Continuously monitor the model's performance and retrain it periodically to adapt to changing customer behavior.
Ensure that the model's predictions can be explained to bank stakeholders and regulators, especially if the model is used to make important decisions.
Segment customers based on their predicted likelihood of churn. This allows the bank to tailor retention strategies to specific customer groups.
Develop customer retention strategies, such as targeted marketing, personalized offers, or customer support interventions, to reduce churn.
Establish a feedback loop to gather data on the effectiveness of retention strategies and use this information to refine the churn prediction model and strategies over time.

Ensure that the data and model comply with data privacy regulations (e.g., GDPR) and internal bank policies.
Create dashboards and reports for bank stakeholders to monitor churn prediction results and the impact of retention strategies.
Document the entire process, including data sources, preprocessing steps, model details, and results, for future reference and regulatory compliance.

Remember that the success of a churn prediction system depends on the quality of data, the choice of modeling techniques, and the effectiveness of retention strategies implemented by the bank. Continuous improvement and adaptability are key factors in achieving long-term success in customer churn prediction.
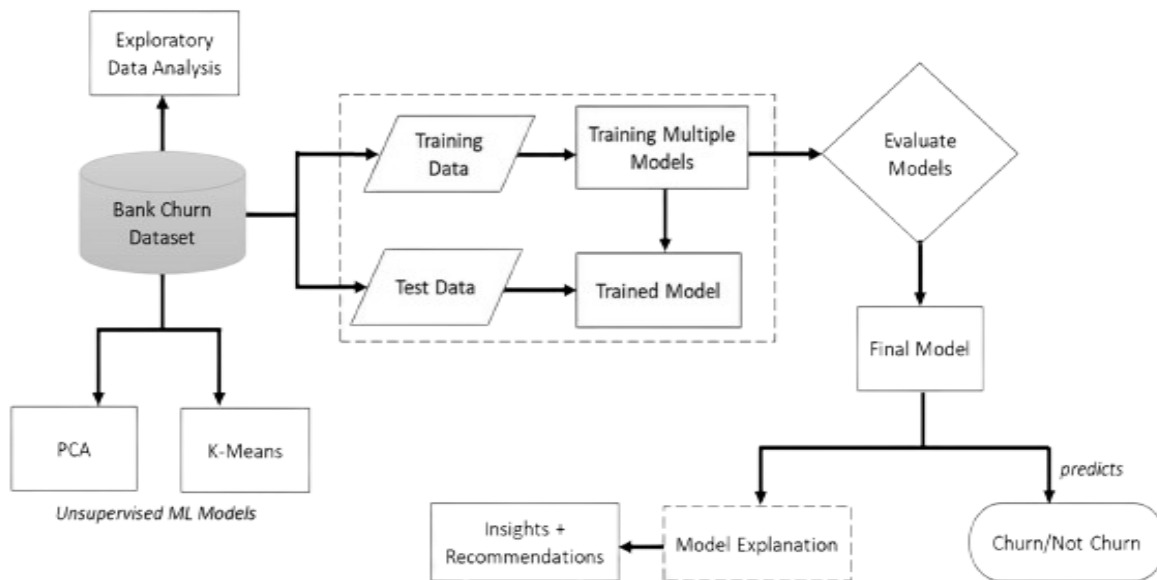
Fig 1. Proposed system of Dynamic Bank Customer Churn prediction.

D. System Design

System design is the process of planning system elements similar as armature, modules, and factors, the colourful interfaces of these factors, and the data passing through the system. The thing of the system design process is to give sufficient detailed information and knowledge. information about the system and its system elements so that the perpetration is compatible with the architectural units defined in the models and views of the system architecture.
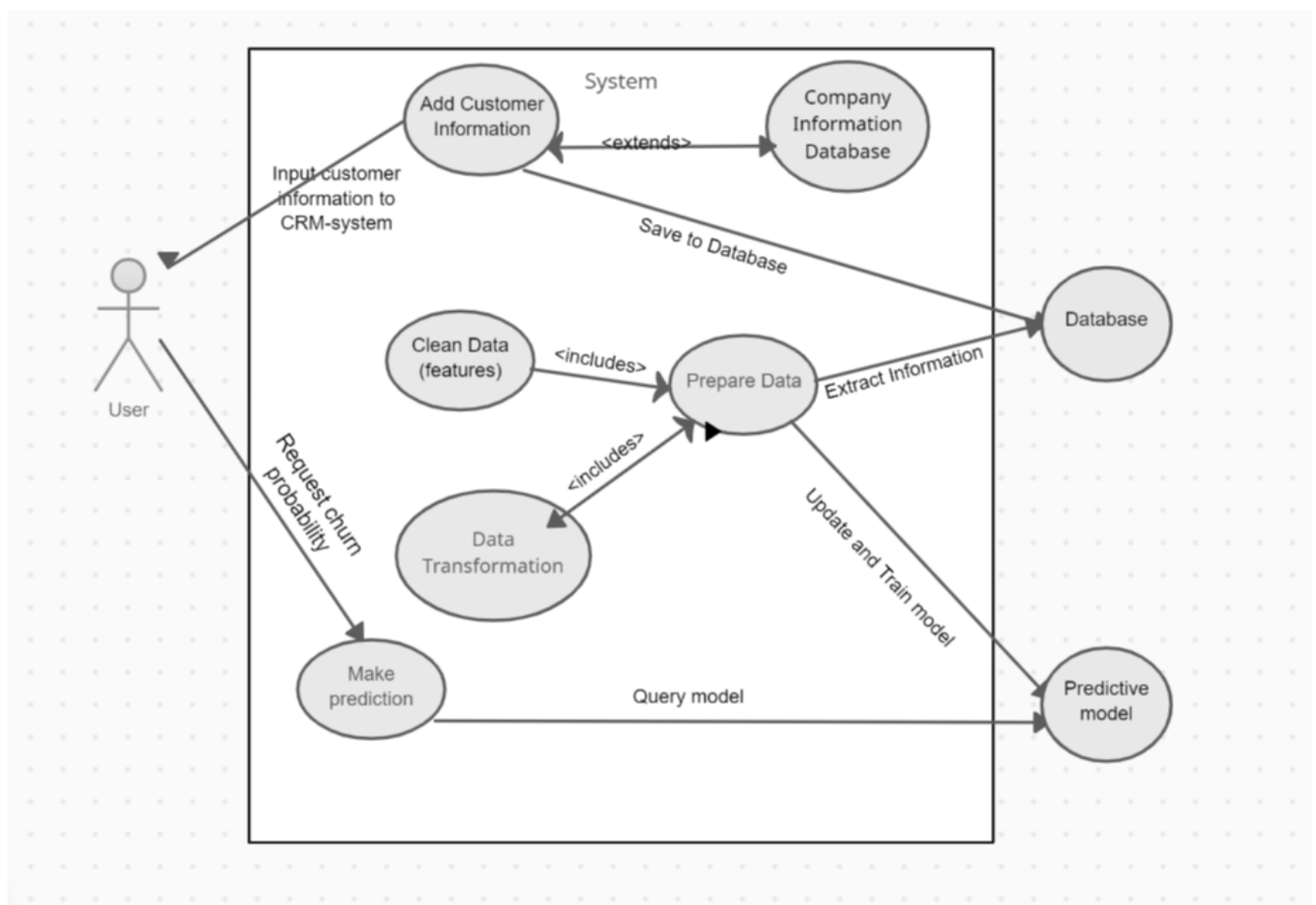
Fig 2. Use case of Bank Customer Churn Prediction

E. Data Flow

A data flow diagram represents the data flow of a process or system usually an information system. A data flow diagram has no control flow - it has no decision rules and no loops. A data flow diagram (DFD) is a graphic or visual representation that uses a standardized set of symbols and notations to describe the operation of a business through the transmission of information. It gives a more clear idea of our project. It expands on each process to give detailed information about the process.
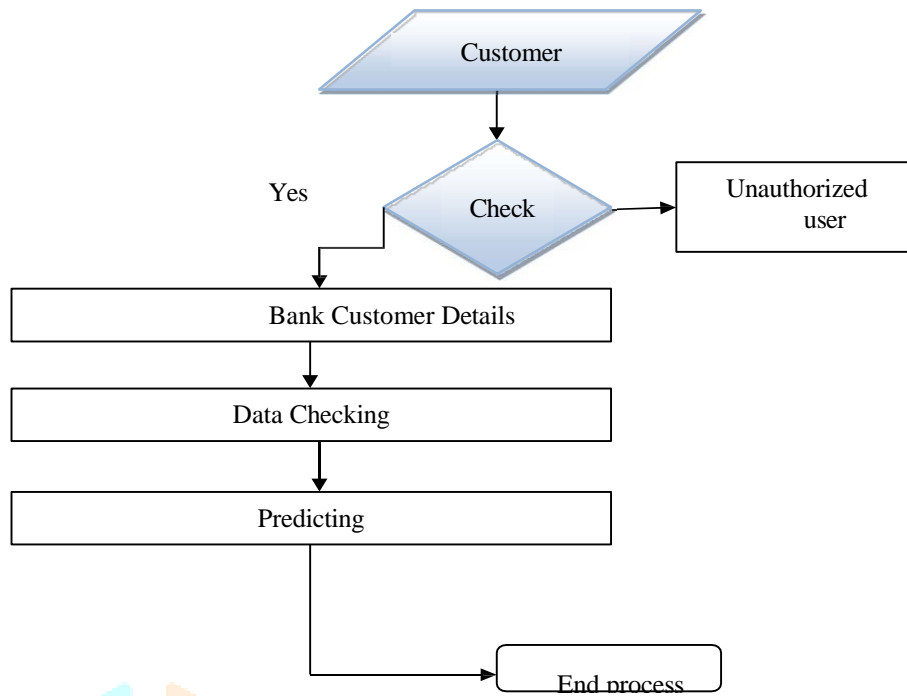
Fig 3. Data Flow of Bank Customer Churn Prediction

IV. RESULTS

In this project, It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant
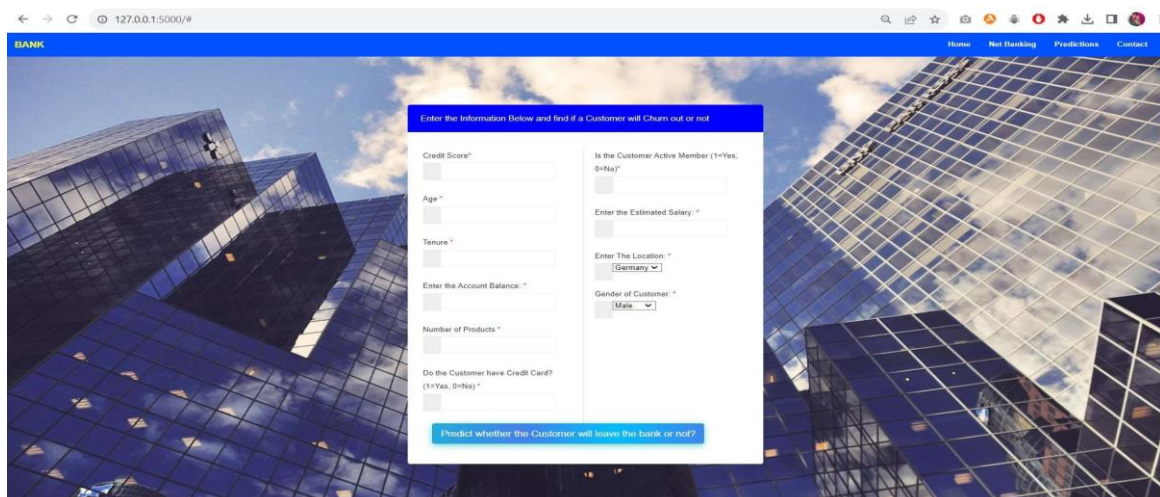

Fig 4. Home Page

First we will have a look at our projects User Interface. These are the screenshots of project outputs.
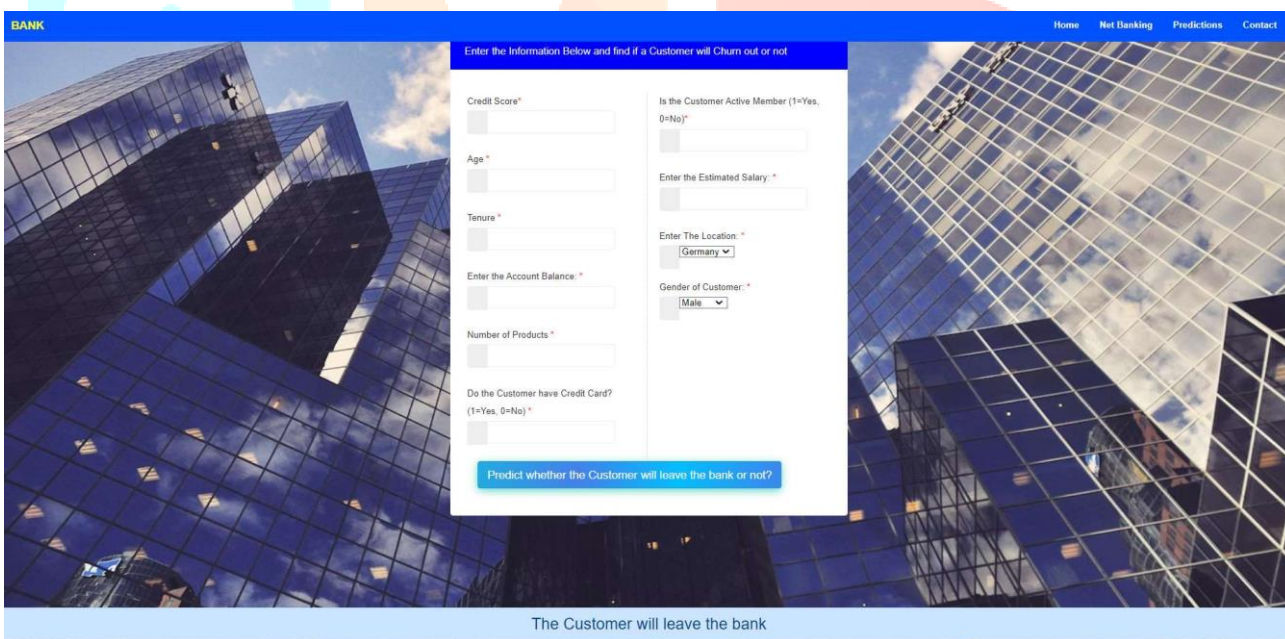

Fig 5. Form validation

Fig 6. Customer will leave the bank.

Fig 7. Customer will not leave the bank

V. CONCLUSION

We Customer churn prediction can help companies identify at-risk customers who are likely to cancel their subscriptions or close their accounts, which can help with preemptive retention strategies. This can help companies:

- Increase profits: As more customers stay longer, revenue should increase

- Improve the customer experience: Understanding why customers churn can help identify weaknesses and improve the overall customer experience

- Minimize acquisition costs: It costs more to acquire new customers than to retain existing ones

Our study looks into the potential applications of machine learning in the B2B domain to forecast customer attrition. As was previously indicated, one of the two components of customer churn management is churn prediction. In the context of business-to-business transactions, it would be intriguing to explore potential feature combinations and their effects on churn prediction in subsequent research. It would also be fascinating to look into what steps may be made in terms of retention strategies and how businesses should actively engage with consumers who are likely to leave, depending on the variables. What other techniques might be applied to feature selection and sampling, and how they might affect the outcome, is another topic that might be worthwhile to look into further. Furthermore, which 37 additional algorithms could

VI. ACKNOWLEDGMENT

REFERENCES

[1] J. Long, E. Shelhamer, and T. Darrell, "LANE DETECTION TECHNIQUES" – A Review." in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431– 3440.

[2] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, And P.H Torr, "A Layered Approach To Robust Lane Detection At Night." 2015, pp. 1529–1537.

[3] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixelwise labelling," arXiv preprint arXiv:1505.07293, 2015

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015

[5] Shopa, P., N. Sumetha and P.S.K Pathra. "Traffic sign detection and recognition using OpenCV", International Conference on Information Communication and Embedded Systems (ICICES2014), 2014.

.