**CNN-Based Depression Detection System**

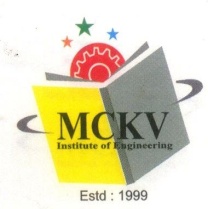
Submitted as a partial fulfillment of Bachelor of Technology in Computer Science & Engineering (Data Science)

of

MCKV Institute of Engineering

(An Autonomous Institute under UGC Act, 1956; Approved by AICTE,

Affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal)



**Project Report** **On**

**ML based Churn Analysis for Customer Retention**

***Submitted by***

***Group No 13 of CSE-Data Science(2020-2024)***

|  |  |  |
| --- | --- | --- |
| **Name of Students** |  | **Examination Roll No.** |
| ABHINABA SARKAR |  | 11600320001 |
| ARNAB PAL |  | 11600320011 |
| DIPANJAN NATH |  | 11600320013 |
| HIMANSHU SHEKHAR METE |  | 11600320015 |
| UTTAM SOREN |  | 11600320051 |

Under the supervision of

Mr. Nilay Kr. Nag

B.Sc(Phy), AMIE(ECE), ADST(CEDT,Govt. of India), MS(Software Systems)

21 years of experience in teaching (6 years in Industry)

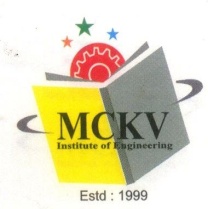
MCKV Institute of Engineering

**Department of Computer Science & Engineering,**

**MCKV Institute of Engineering**

**243, G.T. Road(N)**

**Liluah, Howrah - 711204**



**Department of Computer Science & Engineering**

**MCKV Institute of Engineering**

**243, G. T. Road (N),**

**Liluah, Howrah-711204**

**CERTIFICATE OF RECOMMENDATION**

I hereby recommend that the term paper prepared under my supervision by the students listed below

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Name** | **Signature** |
| 1. | Abhinaba Sarkar |  |
| 2. | Arnab Pal |  |
| 3. | Dipanjan Nath |  |
| 4. | Himanshu Shekhar Mete |  |
| 5. | Uttam Soren |  |

for the project entitled **ML based Churn Analysis for Customer Retention** be accepted in partial fulfilment of the requirements for the degree of Bachelor of Technology in Computer Science & Engineering (Data Science).

|  |  |  |
| --- | --- | --- |
| ----------------------------------------------------------- |  | -------------------------------------- |
| Mr. Avijit Bose |  | Project Guide |
| Assistant Professor & Head of the Department, |  | Name of Project Guide |
| Computer Science & Engineering |  | Designation |
| MCKV Institute of Engineering, Howrah |  | Computer Science & Engineering Dept. |



**MCKV Institute of Engineering**

**(An Autonomous Institute under UGC Act, 1956**

**Approved by AICTE**

**Affiliated to Maulana Abul Kalam Azad University of Technology,**

**West Bengal)**

**CERTIFICATE OF APPROVAL**

**(B.Tech Degree in Computer Science & Engineering (Data Science))**

This project term paper is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed, and conclusion drawn therein but approve the project report only for the purpose for which it has been submitted.

|  |  |  |
| --- | --- | --- |
| COMMITTEE ON FINAL EXAMINATION FOR EVALUATION OF  PROJECT TERM PAPER | 1. |  |
| 2. |  |
| 3. |  |
| 4. |  |

# ACKNOWLEDGEMENT

We would like to extend our heartfelt gratitude to the numerous individuals and organizations who have provided unwavering support throughout our Final Year Project.

First and foremost, we would like to express our deepest appreciation to our mentor, Mr. Nilay Kr. Nag, for his exceptional guidance, unwavering support, and constant encouragement throughout the duration of this project. His expertise and extensive knowledge in the field of machine learning have been invaluable to us, and we are truly grateful for his willingness to dedicate his time and share his expertise.

We would also like to express our gratitude to the members of our project team for their unwavering dedication and hard work. Together, we have achieved significant milestones, and their contributions have been instrumental in the success of this project.

Furthermore, we would like to extend our thanks to the following individuals and organizations for their invaluable support:

* MCKVIE, for providing us with the necessary resources and support to complete this project. Their commitment to academic excellence and research has been instrumental in our growth and development.
* Our esteemed Head of Department, Mr. Avijit Bose, for his continuous encouragement and unwavering support throughout the project. His guidance and belief in our abilities have been pivotal in our journey.

We acknowledge that without the support of these remarkable individuals and organizations, this project would not have been possible. Their guidance, expertise, and encouragement have played a significant role in shaping our project's outcomes, and we are deeply grateful for their contributions.

We would like to express our sincere appreciation to all those who have supported us throughout this project. Your assistance and guidance have been invaluable, and we are truly grateful for your presence in our journey.

# Contents –

|  |  |
| --- | --- |
| 1. Title of the Project | 1 |
| 1. Certificate of recommendation | 2 |
| 1. Certificate of approval | 3 |
| 1. Acknowledgement | 4 |
| 1. Content page | 5 |
| 1. Abstract of the Project | 6 |
| 1. Introduction | 7 |
| 1. Scope of work | 8 |
| 1. Survey of the Previous Work | 9 |
| 1. Hardware and Software Details 2. Software required 3. Dataset to be used | 10  11 |
| 1. Roadmap of Future Work 2. Proposed Works for the Seventh Semester 3. Proposed Works for the Eight Semester | 11  11 |
| 1. References | 11 |

# Abstract –

This project focuses on churn analysis in the telecommunications industry, aiming to develop a predictive model that can identify customers at risk of churning and enable proactive measures for customer retention. The project utilizes Python programming language and various libraries such as pandas, matplotlib, and scikit-learn.

The project follows a systematic workflow, starting with data preprocessing to handle missing values and convert categorical variables into numerical representations. Exploratory data analysis techniques are employed to gain insights into the relationship between different features and churn. Visualizations are created to highlight patterns and trends in the data.

Feature selection techniques, specifically SelectKBest, are applied to identify the most important features for predicting churn. These features are used to train a neural network model with multiple layers. The model is optimized using appropriate algorithms and evaluated using various performance metrics such as accuracy, precision, recall, and F1-score.

The evaluation of the model's performance includes the analysis of a confusion matrix to provide insights into the model's predictions and its ability to accurately identify churn instances. The project serves as a comprehensive reference for anyone interested in implementing churn analysis in the telecommunications industry, showcasing the application of machine learning techniques and emphasizing the importance of proactive customer retention strategies.

# Introduction –

In the rapidly evolving telecommunications industry, customer churn poses a significant challenge for companies. Churn, which refers to customers discontinuing their services and switching to competitors, not only leads to revenue loss but also highlights the need for effective customer retention strategies. Recognizing the importance of addressing this issue, our organization has embarked on a churn analysis project to develop a predictive model that can identify customers at risk of churning. This report provides an overview of our ongoing project, including the previous works, the challenges faced, and our planned approach to overcome these challenges.

To lay the foundation for our churn analysis project, we conducted an extensive review of existing literature and industry practices. This allowed us to gain valuable insights into the factors influencing churn, the methodologies employed, and the potential limitations encountered. Additionally, we studied the specific challenges faced by our organization, including the complexity of the telecom landscape, the diversity of customer segments, and the availability and quality of data. By understanding these challenges, we aimed to develop a robust framework that addresses the unique aspects of our organization's churn analysis.

One of the primary challenges we encountered was the availability of comprehensive and reliable data. Telecom companies generate vast amounts of data, including customer information, service usage, billing details, and customer interactions. However, the data often exists in disparate systems and formats, making it difficult to extract meaningful insights. To overcome this challenge, we have planned to implement a data integration and preprocessing pipeline that consolidates and cleanses the data, ensuring its quality and compatibility for analysis. Additionally, we will employ advanced data mining techniques to uncover hidden patterns and relationships within the data, enabling us to identify the key drivers of churn.

**Here are some of the benefits of Churn Analysis :**

- **Improved Customer Retention:** Churn analysis helps identify customers at risk of discontinuing services, enabling proactive measures to retain them and enhance customer loyalty.

- **Cost Savings and Resource Optimization**: By leveraging churn analysis insights, companies can target their retention efforts effectively, optimizing resources and minimizing costs associated with acquiring new customers.

- **Competitive Advantage and Business Strategy Optimization:** Churn analysis provides a competitive edge by understanding customer behavior and preferences, allowing businesses to tailor offerings and experiences. This informs strategic decision-making, refining marketing, product development, and customer service strategies.

**However, there are also some challenges associated with Churn Analysis :**

- **Data Quality:** Churn analysis relies on accurate and reliable data. Challenges may arise due to incomplete or inconsistent data, requiring data cleansing and preprocessing techniques.

- **Feature Selection:** Choosing the most relevant features for predicting churn can be challenging. It requires domain knowledge and careful analysis to identify the key variables that contribute to customer attrition.

- **Class Imbalance:** Churn datasets often exhibit class imbalance, where the number of churn instances is significantly lower than non-churn instances. Handling this imbalance is crucial to avoid biased model performance and ensure accurate predictions.

# Scope of work –

**The scope of work for this project includes:**

- **Data Collection:** Gathering relevant customer information and churn status data from reliable sources in the telecommunications industry.

- **Data Pre-processing:** Cleaning, organizing, and transforming the collected data to make it suitable for analysis. This includes handling missing values, removing duplicates, and converting categorical variables into numerical representations.

- **Model Training:** Developing a predictive model using machine learning techniques such as decision trees, logistic regression, or neural networks. The model is trained on a designated training set and optimized to accurately predict churn.

- **Model Evaluation:** Assessing the performance of the trained model using evaluation metrics like accuracy, precision, recall, and F1-score. Cross-validation techniques are employed to ensure the model's robustness and generalizability.

- **System Deployment:** Integrating the churn prediction model into a functional system or application that can be seamlessly incorporated into the telecommunications company's infrastructure. This involves creating user-friendly interfaces or APIs for easy access and utilization.

Challenges associated with this project include:

* Data Availability and Quality: Obtaining reliable and comprehensive data for churn analysis can be challenging due to limited sources and data inconsistencies.
* Feature Selection and Model Complexity: Selecting relevant features and handling complex models, such as deep learning architectures, require careful consideration and expertise.
* Class Imbalance: Imbalanced datasets, with fewer churn instances, pose a challenge for training accurate churn prediction models.
* Interpretability and Explainability: Interpreting and explaining model results, especially for complex models, can be challenging for stakeholders.
* Implementation and Deployment: Integrating the churn prediction system, ensuring scalability, and maintaining performance require collaboration between data scientists, IT teams, and business stakeholders.

Despite these challenges, CNN-based depression detection systems have the potential to revolutionize the way that depression is diagnosed and treated. By providing a reliable and affordable way to detect depression in real time, these systems can help to improve the lives of millions of people.

# Survey of previous works –

**PREDICTIVE ANALYSIS OF CUSTOMER CHURN IN TELECOM INDUSTRY USING SUPERVISED LEARNING**

**SHREYAS RAJESH LABHSETWAR**

**DEPARTMENT OF COMPUTER ENGINEERING**

**FR.C.RODRIGUES INSTITUTE OF TECHNOLOGY,INDIA**

In this paper, There is written about predictive analysis of customer churn in telecom industry. It’s main aim, retaining a loyal customer is far more important than acquiring the new customer. Finding the factor by which customer’s are loosed .

And it’s need a complex calculation to predict that whether a customer leave or stay. In this paper , there used is used 3 supervised machine learning algorithm.(1) EXTRA TREE CLASSIFIER,(2) XGBOOSTING,(3) SVM. AND the AUC score 0.843,0.787 and 0.735 respectively and low false negative. Here use logistic regression , Gaussian Naïve Bayes, Adaboost , XGB Classifier, SGD classifier, SVM .But we are using Artificial Neural Network to solve this complex problem.

**Customer churn prediction using improved balanced random forests**

**Yaya Xie , Xiu Li , E.W.T. Ngai , Weiyun Ying**

**a Department of Automation, Tsinghua University, Beijing, PR China**

**b Department of Management and Marketing, The Hong Kong Polytechnic University, Hong Kong, PR China**

**c School of Management, Xi’an Jiao tong University, Xi’an, PR China**

**Improved balanced random forest**

In this report they have used balanced random forest to predict the churn analysis. Balanced random forest is a combination of weighted random forest, where the minority class has a higher predication weight, and balanced random forest, where the minority class is oversampled. They got an accuracy rate 93.4%.

**Scope of improvement**

Here they used the IBRF model to predict the churn but they can improve the effectiveness and generalization ability more, which we can do more efficiently by using ANN.

**A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China**

**Research Hypothesis**

Based on the existing research, customer relationship management theory, and customer value,this paper puts forward the following research hypotheses from five dimensions:

**1)price factor:** Hypothesis 1. Price can significantly affect customers’ willingness negatively, and the increase of monthly consumption will increase the customer churn rate.

**2)product factor**: Hypothesis 2. Previous consumption has a positive impact on customers’ repeated purchase behavior. For customers with behavioral stickiness, the more they depend on the products, the lower the churn rate.

**3)customer factor:** Hypothesis 3. The statistical variable of customers will have an impact on customer loyalty. The higher the customer value, the lower the churn rate.

**4) business factor:** Hypothesis 4. Convergence business has an adverse impact on customer churn and the churn rate of customers signing the bundling contract is reduced.

**5)service factor:** Hypothesis 5. Relationship investment can significantly increase customer confidence, and there is a negative correlation between customer satisfaction and customer churn; i.e., the lower the customer satisfaction is, the higher the customer churn rate is.

**A Survey on Churn Analysis**

**arXiv:2010.13119v3 [cs.LG] 2 Dec 2020**

**A Preprint by Jaehuyn Ahn, Gyeonggi-do, Republic of Korea**

**Title: Churn Analysis in Various Business Fields**

Introduction:

Customer churn refers to customers who cease doing business with a company within a given time or contract. As competition increases and customer acquisition costs rise, preventing churn becomes crucial. Studies have shown that efforts to retain existing customers are more efficient than acquiring new ones. Churn prediction can help increase customer retention and the overall value of a company.

**Types of Churn:**

1. Contractual Churn: Customers who do not extend their contract or renew their subscription.

2. Non-Contractual Churn: Customers who leave a service or contract without time constraints.

**Churn Analysis in Business Fields:**

Churn analysis has been studied in various industries, including telecommunications, finance, insurance, gaming, music streaming, internet services, newspaper subscriptions, online dating, online commerce, Q&A services, and social networks.

**Benefits of Churn Prediction:**

Churn prediction helps improve business outcomes by identifying customers likely to churn and providing incentives to retain them. Different churn observation criteria, such as monthly, daily, and binary churn, are used to monitor customer status changes.

**Approaches to Churn Prediction:**

Churn prediction models have been developed using log data and machine learning techniques. Features such as customer behavior, retention time, customer lifetime value, and psychological factors are considered in churn prediction models.

**Conclusion:**

Churn analysis plays a crucial role in customer management and business success. Understanding the differences in churn definitions and techniques across various industries can aid in the development of effective churn prediction models. Further research and advancements in deep learning and parameter description are needed to improve churn prediction accuracy.

# Hardware & Software details –

Hardware:

* Computer: A high-performance computer system capable of handling data processing and model training tasks efficiently.
* Storage: Sufficient storage capacity to store the collected data, pre-processed data, and trained models.
* Memory (RAM): Adequate RAM capacity to accommodate the size of the dataset and support the computational requirements of the machine learning algorithms.

Software:

* Python 3.8.8: A programming language commonly used in machine learning and data analysis tasks.
* pandas 1.3.0: A powerful data manipulation and analysis library, particularly useful for handling and preprocessing tabular data.
* matplotlib 3.4.2: A data visualization library for creating plots and charts to analyze and display data.
* numpy 1.21.0: A fundamental package for scientific computing with Python, providing support for large, multi-dimensional arrays and mathematical functions.
* scikit-learn 0.24.2: A comprehensive machine learning library with tools for classification, regression, clustering, and model evaluation.
* tensorflow 2.5.0: An open-source deep learning framework for building and training neural network models.
* keras 2.4.3: A user-friendly deep learning library that can be used on top of TensorFlow or Theano.
* Django 3.2.4: A popular web framework that simplifies the development and deployment of web applications.
* MySQL: A popular relational database management system used for storing and retrieving data.
* Jupyter Notebook 6.4.0: An interactive development environment for writing and executing Python code in a web browser. Jupyter Notebook allows you to create and share documents that contain live code, equations, visualizations, and narrative text.

# Roadmap of Future work –

**7th Semester:** Implementing User Interface In the next semester, the focus will be on enhancing the project by implementing a user interface. The goal is to provide a more user-friendly and accessible way for users to interact with the churn prediction system. Several options can be explored for the user interface, including web-based interfaces using Django or mobile interfaces for phones.

* **Option 1: Web-based User Interface with Django**
* Explore using Django, a popular web framework in Python, to develop a web-based user interface.
* Utilize Django's built-in features for creating interactive web pages, handling user input, and displaying churn prediction results.
* Design an intuitive and visually appealing interface that allows users to input their information and receive churn predictions.
* **Option 2: Mobile User Interface**
* Consider developing a mobile application to provide users with a convenient way to access the churn prediction system.
* Explore frameworks such as React Native or Flutter to build cross-platform mobile apps that can be deployed on both iOS and Android devices.
* Design a user-friendly interface optimized for mobile screens, allowing users to input their data and receive churn predictions on their phones.

**8th Semester:** Implementing a Chatbot In the second semester, the focus will shift towards implementing a chatbot to enhance the user experience and provide an interactive and conversational interface for users.

* Incorporate Rasa, an open-source machine learning framework, to build the chatbot functionality.
* Train the chatbot to understand user queries related to churn prediction and provide accurate responses.
* Implement natural language processing (NLP) techniques to enable the chatbot to understand and interpret user input effectively.
* Design the chatbot to guide users through the process of inputting their data and provide churn predictions based on the inputs.

# References –

<https://ictactjournals.in/paper/IJSC_Vol_10_Iss_2_Paper_5_2054_2060.pdf>

<https://downloads.hindawi.com/journals/ddns/2021/7160527.pdf>

<https://adiwijaya.staff.telkomuniversity.ac.id/files/2014/02/Customer-Churn-Prediction-using-Improved-Balance-Random-Fores.pdf>

<https://www.researchgate.net/publication/295257947_Churn_analysis_Predicting_churners>

<https://www.researchgate.net/publication/343787983_Customer_Churn_PredictionA_Survey>