**A PROJECT REPORT**

**on**

**“CHURN PREDICTION”**

**Submitted to**

**KIIT Deemed to be University**

**In Partial Fulfillment of the Requirement for the Award of**

**BACHELOR’S DEGREE IN**

**COMPUTER SCIENCE**

**BY**

**SHRIJAN POUDEL**

**PASHUPATI SAH**

**RIZAN KHANAL**

**UTKARSH SHRESTHA**

**SHUBHAM DAWADI**

**21053459**

**21053467**

**21053464**

**21053466**

**21053473**

**SOURAV KUMAR GIRI**



**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**

**April 2024**

A PROJECT REPORT

on

“CHURN PREDICTION”

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR’S DEGREE IN

COMPUTER SCIENCE

BY

SHRIJAN POUDEL

PASHUPATI SAH

RIZAN KHANAL

UTKARSH SHRESTHA

SHUBHAM DAWADI

21053459

21053467

21053464

21053466

21053473

SOURAV KUMAR GIRI



SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA -751024

April 2024

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024



CERTIFICATE

This is certify that the project entitled

“CHURN PREDICTION“

SUBMITTED BY

SHRIJAN POUDEL

PASHUPATI SAH

RIZAN KHANAL

UTKARSH SHRESTHA

SHUBHAM DAWADI

21053459

21053467

21053464

21053466

21053473

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during the year 2023-2024, under our guidance.

Date:11 / 4 / 2024

(SOURAV KUMAR GIRI)

Project Guide

**Acknowledgement**

We are profoundly grateful to **SOURAV KUMAR GIRI** of KIIT University for his expert guidance and continuous encouragement throughout to see that this project meets its target since its commencement to its completion. .....................

SHRIJAN POUDEL

PASHUPATI SAH

RIZAN KHANAL

UTKARSH SHRESTHA

SHUBHAM DAWADI

**ABSTRACT**

Churn prediction is a critical task for businesses looking to retain customers and maximize revenue. This paper introduces a machine learning-based approach to churn prediction, utilizing various predictive modeling techniques to accurately forecast customer churn. The proposed framework leverages historical customer data, including demographics, transaction history, and interactions, to train predictive models capable of identifying potential churners.

The machine learning process involves several iterative phases that work together to achieve effective results. First, data must be meticulously cleaned and prepared, including resolving inconsistencies and modifying features for optimal model performance. Multiple model types, such as logistic regression and decision trees, should be assessed, as each has unique strengths and weaknesses that suit various tasks. To avoid overfitting and identify the most effective configuration, cross-validation and precise tuning of model parameters are essential to ensure reliable outcomes.

Businesses can reduce churn by closely monitoring customer behavior for signs of disengagement. Advanced analytics pinpoint the specific actions, experiences, or characteristics that correlate with a higher risk of churn. Armed with this knowledge, companies can proactively reach out to these customers, offering personalized solutions or incentives designed to strengthen their loyalty.

This churn prediction approach proves highly successful in pinpointing customers likely to leave. With this information, businesses can proactively address dissatisfaction with personalized solutions and offers. This translates into stronger customer relationships, increased customer lifetime value, and ultimately, a healthier bottom line.

Contents

| 1 | Introduction | | | 1 |
| --- | --- | --- | --- | --- |
| 2 | Basic Concepts/ Literature Review | | | 3 |
|  | 2.1 | Types of Churn | | 4 |
|  | 2.2 | Unmitigated Churn | | 5 |
|  |  | Literature Review | | 6 |
| 3 | Problem Statement / Requirement Specifications | | | 9 |
|  | 3.1 | Project Planning........................... | | 9 |
|  | 3.2 | Project Analysis (SRS) | | 11 |
|  | 3.3 | System Design | | 12 |
|  |  | 3.3.1 | Design Constraints | 13 |
|  |  | 3.3.2 | System Architecture | 13 |
| 4 | Implementation | | | 14 |
|  | 4.1 | Methodology / Proposal ........................... | | 20 |
|  | 4.2 | Testing / Verification Plan ……………. | | 21 |
|  | 4.3 | Result Analysis / Screenshots …………. | | 21 |
| 5 | Standard Adopted | | | 22 |
|  | 5.1 | Design Standards . . . . . . . . . . . . . . . | | 22 |
|  | 5.2 | Coding Standards . . . . . . . . . . . . . . | | 23 |
| 6 | Conclusion and Future Scope | | | 24 |
|  | 6.1 | Conclusion ……………………….. | | 24 |
|  | 6.2 | Future Scope ………………………. | | 24 |
| References | | | | 25 |
| Individual Contribution | | | | 26 |
| Plagiarism Report | | | | 33 |

*CHURN PREDICTION*

Chapter 1

Introduction

Today, customer retention is the key to thriving in a world of endless choices. A strong customer base means consistent revenue, but more importantly, it's a sign your company is delivering real value that keeps people coming back. Investing in loyalty isn't just smart business, it's the only way to build a brand that stands the test of time.

Customer retention transcends mere financial metrics, embodying a company's commitment to excellence across its entire operation. From product innovation to customer service, every interaction shapes the customer experience and influences long-term loyalty.

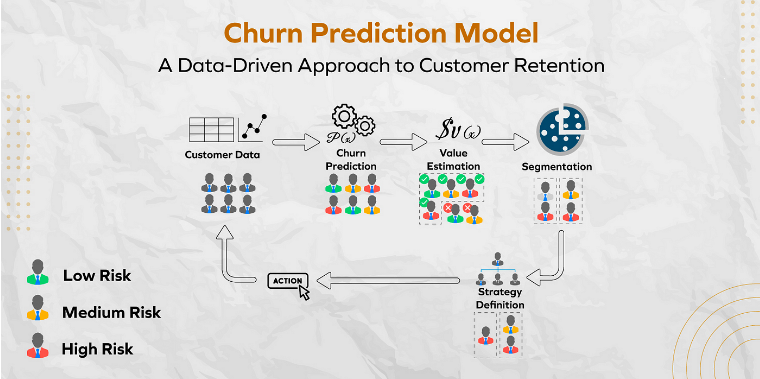


Figure 1.1: WORKING OF CHURN PREDICTION MODE

However, the looming threat of customer churn presents a formidable challenge for businesses of all sizes and industries. Churn, marked by the departure of customers, signifies more than just a loss of revenue—it signals a breakdown in the essential bond between a company and its clientele. Whether driven by dissatisfaction, competition, or shifting market dynamics, each instance of churn represents a missed opportunity and a potential harbinger of future obstacles.

To combat this challenge, forward-thinking organizations are increasingly turning to predictive analytics as a cornerstone of their customer retention strategies. By harnessing data-driven insights and advanced algorithms, businesses can proactively identify churn signals, address customer concerns, and strengthen relationships before they deteriorate.

The urgency of addressing churn cannot be overstated, especially in an era defined by rapid digitization and evolving consumer expectations. Failing to adopt proactive churn mitigatistrategies risks not only financial losses but also damage to a company's reputation and market relevance.

Against this backdrop, the imperative for predictive analytics to anticipate and mitigate churn has never been more apparent. Through a strategic blend of technological innovation, analytical rigor, and customer-eccentricity, businesses can navigate the challenges of customer retention with confidence and foresight

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

Chapter 2

Basic Concept:

A churn prediction model provides businesses with a glimpse into the future dynamics of their customer relationships. By analyzing historical customer data such as demographics, purchasing behaviors, and interaction patterns, the model identifies nuanced signals that could indicate a customer's potential departure. Through understanding these indicators, the model evaluates each customer and predicts their likelihood of staying loyal or leaving. This actionable insight empowers companies to proactively intervene, engaging at-risk customers with personalized strategies and offerings designed to enhance retention. By minimizing churn rates, businesses not only boost customer satisfaction but also secure consistent revenue streams and strengthen their overall stability.

**WHAT IS CHURN?**

Think of customer churn like a leaky bucket: every customer that leaves is lost revenue dripping away. It happens for many reasons – maybe the product falls short, a competitor swoops in with a flashier offer, or your customer service left them feeling frustrated. Sometimes, life just gets in the way, and a once-happy customer's needs simply change. Whatever the cause, the result hurts your bottom line. That's why companies fight so hard against churn. It's not just about keeping a list of names long; it's about proving the value you deliver. Loyalty programs, personalized outreach, and constantly improving the customer journey can stem the tide. But the smartest moves are proactive: using data-driven tools to spot those subtle hints someone is wavering, and offering solutions before they ever think of walking away.

*School of Computer Engineering, KIIT, BBSR*

**2.1 Types of Churn (Business perspective)**

Customer Churn:

Customer churn, also known as customer attrition, refers to the loss of customers or clients over a specific period. This type of churn is typically measured by the percentage of customers who cease to engage with a company's products or services, unsubscribe from services, or terminate their accounts.

Revenue Churn:

Revenue churn focuses on the loss of revenue attributed to departing customers. It measures the decrease in recurring revenue from existing customers due to cancellations, downgrades in service plans, or reduced usage of premium features.

Contractual Churn:

Contractual churn occurs when customers terminate contractual agreements prematurely before the agreed-upon term expires. This type of churn is common in subscription-based models, where customers commit to a fixed term but opt out before fulfilling the contract.

Product Churn:

Product churn occurs when customers discontinue the use of specific products or services within a company's portfolio. This can result from changes in customer preferences, dissatisfaction with product performance, or the availability of superior alternatives in the market.

Segment Churn:

Segment churn involves the loss of customers within specific demographic, geographic, or behavioral segments. Identifying segment-specific churn patterns enables businesses to tailor retention strategies to address the unique needs and preferences of different customer segments.

*School of Computer Engineering, KIIT, BBSR*

**2.2 Unmitigated Churn: Implications for Business Sustainability**

1. Revenue Loss:

Explanation: Unchecked customer churn directly translates to a decline in revenue streams for businesses. As customers leave, they take with them their purchasing power, leading to a reduction in sales and overall revenue. Without evaluating customer churn, businesses risk experiencing a continuous drain on their financial resources, hindering growth opportunities and profitability.

2. Diminished Brand Loyalty:

Explanation: Ignoring customer churn undermines efforts to build and maintain brand loyalty. Customers who churn are likely to share their negative experiences with others, tarnishing the reputation of the business and eroding trust among existing and potential customers. Without evaluating churn, businesses fail to address underlying issues driving customer dissatisfaction, further exacerbating the erosion of brand loyalty.

3. Decreased Market Competitiveness:

Explanation: In today's competitive marketplace, customer retention is a key differentiation for businesses seeking to gain a competitive edge. Failing to evaluate customer churn puts businesses at a disadvantage, as competitors capitalize on their inability to retain customers. Over time, this can lead to a loss of market share, reduced relevance, and diminished competitiveness compared to industry peers.

*School of Computer Engineering, KIIT, BBSR*

**Literature Review**

Churn prediction is a crucial task in various industries, particularly in telecommunications, banking, and e-commerce, where retaining customers is essential for business success. In recent years, machine learning (ML) techniques have gained significant attention for their effectiveness in predicting churn and identifying factors influencing customer attrition

Chih-Fong Tsai (2015) introduced novel hybrid neural network techniques aimed at predicting customer churn in a CRM dataset provided by American telecom companies. The study involved constructing two hybrid models by combining different neural network methods, namely back-propagation artificial neural networks (ANN) and self-organizing maps (SOM), for churn prediction. The hybrid models were created by combining ANN with ANN (ANN + ANN) and SOM with ANN (SOM + ANN). Specifically, the first hybrid model involved data reduction by filtering out unrepresentative training data, followed by using the outputs to build a prediction model based on the second technique. To evaluate model performance, three types of testing sets were employed: a general testing set and two fuzzy testing sets based on data filtered out by the first technique of the two hybrid models (ANN and SOM, respectively). The findings revealed that the hybrid models surpassed the single neural network baseline model in terms of prediction accuracy. Notably, the ANN + ANN hybrid model demonstrated superior performance compared to the SOM + ANN hybrid model.

A.H. Farquad introduced a hybrid methodology aimed at addressing the limitations of the conventional Support Vector Machine (SVM) model, which often produces opaque models that lack interpretability. The hybrid approach comprises three key phases:

*Firstly*, in the initial phase, SVM-Recursive Feature Elimination (SVM-RFE) is utilized to diminish the feature set. This process aims to select the most pertinent feature for further analysis.

*Subsequently*, in the second phase, the dataset with the reduced feature set is utilized to construct an SVM model, from which support vectors are extracted. SVM is recognized for its effectiveness in handling classification tasks, including churn prediction.

*School of Computer Engineering, KIIT, BBSR*

*Finally*, in the last phase, rules are generated utilizing a Naive Bayes Tree (NBTree), which amalgamates decision tree techniques with a Naive Bayesian classifier. This step is intended to enhance the interpretability of the model by transforming the SVM-derived model into human-comprehensible rules.

The dataset employed in this study originates from a bank credit card customer dataset obtained from the Business Intelligence Cup 2004. Notably, this dataset exhibits a significant imbalance, with 93.24% of customers labeled as loyal and only 6.76% labeled as churned. Experimental findings indicated that while the hybrid model exhibited promising performance in churn prediction, it encountered scalability challenges when applied to larger datasets. This limitation underscores the necessity for further optimization and refinement of the proposed methodology to ensure its effectiveness on larger datasets.

WouterVerbeke (2016) explored two algorithms, Ant-Miner+ and ALBA, for building accurate and understandable churn prediction models. Ant-Miner+ leverages ant colony behavior for high accuracy and allows incorporating expert knowledge, resulting in clear and interpretable models. ALBA combines the accuracy of Support Vector Machines with interpretable rule sets. When compared to other methods, ALBA with RIPPER achieved the highest overall accuracy, while C4.5 and RIPPER (oversampled) maximized true churner identification. Ant-Miner+ offered the benefit of incorporating domain knowledge and generating smaller, interpretable rules, while RIPPER produced concise but potentially counterintuitive rules. Overall, both Ant-Miner+ and ALBA provide valuable options for balancing accuracy and interpretability in churn prediction models

In a study by Lu (2017), a technique called boosting was explored to improve customer churn prediction models. Boosting assigns weights to data points, focusing the model on those most likely to be misclassified. Here, Lu separated customers into two clusters based on these weights. This approach effectively identified a group of high-risk churners.

Logistic regression, a common statistical method, served as the foundation for the model within each cluster. The experiment demonstrated that the boosting approach led to a clearer separation of churn data compared to a single logistic regression model.

*School of Computer Engineering, KIIT, BBSR*

Lee et al. (2010) investigated a method for churn prediction using Partial Least Squares (PLS) regression. PLS is a technique suited for handling datasets with highly correlated variables. Their goal was to develop a  concise and accurate predictive model for customer churn.  The study went beyond just the model, proposing a  straightforward and actionable churn marketing program. This program allows marketing managers to  effectively minimize churn through targeted marketing interventions

De Bock (20??) proposed GAMensPlus, a novel ensemble classification method for churn prediction. It combines the strengths of both performance and interpretability. GAMensPlus builds upon existing techniques like bagging, random subspace methods, and generalized additive models (GAMs). To enhance interpretability, De Bock introduced two tools: generalized feature importance scores and bootstrap confidence bands for smoothing splines.

Evaluations on real-world data from six churn prediction projects demonstrated that GAMensPlus delivers strong classification performance. At least on par with the individual models it combines (logistic regression and GAM), GAMensPlus offers a valuable option for churn prediction tasks.

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

Chapter 3

Problem Statement

The SaaS companies are facing challenges in retaining customers, leading to a decline in revenue and market share. The companies wants to implement a proactive approach to reduce customer churn by identifying customers who are likely to churn in the near future. To address this problem, the company seeks to develop a machine learning-based predictive model that can accurately forecast customer churn.

3.1 Project Planning

To successfully execute the project development for churn prediction in a SaaS company, it is essential to follow a structured approach. The following steps outline the key tasks involved in planning and implementing the project:

Requirement Identification:

Engage with stakeholders to gather requirements and gain a thorough understanding of the business objectives.

Clearly define the key features and functionalities expected from the predictive model.

Data Collection and Preparation:

Retrieve historical customer data from the company's databases, ensuring data integrity and completeness.

Preprocess the collected data by addressing missing values, encoding categorical variables, and standardizing numerical features as needed.

Exploratory Data Analysis (EDA):

Conduct comprehensive EDA to explore data distributions, identify potential outliers, and uncover insights into feature relationships.

Visualize key metrics and trends to better understand the underlying patterns in the data.

*School of Computer Engineering, KIIT, BBSR*

Feature Selection:

Pinpoint the key factors behind churn: Use statistical analysis or feature importance tools (these can be named if relevant) to reveal which pieces of customer data actually signal 'leaving soon' risk.

Streamline your model: Focus on the features that pack the biggest predictive punch. Ditch any that are repetitive or don't add much value.

Model Selection:

Assess different machine learning algorithms that can be used for classification tasks, such as logistic regression, decision trees, random forests, and gradient boosting methods.

Select the best algorithm by considering performance indicators and computational resource requirements.

.

Hyperparameter Tuning:.

Fine-tune the hyperparameter of the selected model using techniques like grid search or randomized search to optimize performance and generalization.

Experiment with different parameter combinations to identify the optimal configuration.

.

Model Evaluation:

Put your model through its paces: Using cross-validation means testing your model on different slices of data, not just what it was trained on. This gives you a realistic picture of accuracy, precision, and other key metrics.

Check if it works in the wild: Real-world data is messy. Testing your model on a totally new dataset reveals if it's truly reliable, or if it only performs well under ideal conditions.

.

Deployment and Integration:

Install the finalized predictive model in the production environment, ensuring smooth integration with current systems and workflows.

Set up monitoring tools to observe model performance and identify any changes or irregularities as time progresses.

.

Testing and Quality Assurance:

Conduct rigorous testing to validate the model's functionality, accuracy, and stability under different conditions and use cases.

Perform quality assurance checks to ensure compliance with business requirements and regulatory standards.

.

Documentation and Reporting:

Document all phases of the project development process, including data preprocessing steps, model training methodologies, and evaluation results.

*School of Computer Engineering, KIIT, BBSR*

Prepare detailed reports and documentation to provide insights into the project's progress, outcomes, and recommendations.

Training and Knowledge Transfer:

.

Provide training sessions and workshops to stakeholders, including data analysts, data scientists, and business users, to familiarize them with the model's capabilities and usage.

Transfer knowledge about model maintenance, troubleshooting, and updates to ensure ongoing support and optimization.

3.2 Project Analysis

Requirement Clarity:

Review the gathered requirements or problem statement meticulously to ensure they are clear, comprehensive, and devoid of ambiguities.

Address any vague or unclear requirements that require further elucidation.

Feasibility Assessment:

Assess the feasibility of implementing the proposed solution within the stipulated constraints, encompassing factors such as time, budget, and resource availability.

Evaluate the realism and attainability of the project goals considering the current capabilities and limitations.

Risk Identification:

Identify potential risks and uncertainties that could impede the project's progress or affect its outcome.

Evaluate the likelihood and potential impact of each risk, categorizing them based on severity and priority.

Scope Definition:

Clearly define the project scope, delineating its boundaries and deliverables.

Ensure that the scope is well-defined and manageable within the constraints of the project.

Resource Allocation:

Evaluate the availability and allocation of resources, including human resources, tools, and infrastructure.

Determine whether additional resources or expertise are required to execute the project successfully.

*School of Computer Engineering, KIIT, BBSR*

Technology and Tools Assessment:

Assess the suitability of the chosen technology stack and tools for implementing the proposed solution.

Consider factors such as scalability, compatibility, and ease of integration with existing systems.

Requirement Validation:

Validate the requirements against the expectations and needs of stakeholders.

Seek feedback and input from key stakeholders to ensure consensus and alignment on the project objectives.

Gap Analysis:

Perform a comprehensive gap analysis to pinpoint differences between the existing state and the intended future state.

Develop strategies to bridge these gaps and achieve the project goals effectively.

Documentation Review:

Review all project documentation, including the requirements document, project plan, and other relevant artifacts.

Ensure that the documentation accurately reflects the project scope, objectives, and constraints.

3.3 System Design

Following hardware specifications are used to simulate the python code for all the models.

|  |  |
| --- | --- |
| **Component** | Specifications |
| **Processor** | 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz 2.80 GHz |
| **RAM** | 16.0 GB (15.6 GB usable) |
| **System type** | 64-bit operating system, x64-based processor |
| **Edition** | Windows 11 Home Single Language |
| **Version** | 23H2 |

**Table : Hardware Specification**

*School of Computer Engineering, KIIT, BBSR*

Following software specifications are used to simulate the python code for all the models.

|  |  |
| --- | --- |
| **Software used** | Python Notebook using google colab |
| **Packages used** | |
| **Package name** | **Description** |
| pandas | Deals with data frame |
| numpy | Deals with numeric operations |
| matplotlib.pyplot | Deals with data visualization, typically in the form of plots, graphs and charts |
| seaborn | It offers an accessible way to create visually appealing and insightful statistical graphics. |
|  |  |
| missingno | Offers various visualizations to examine the occurrence and distribution of missing data within a pandas DataFrame. |
| scatter\_matrix | Compactly plots all the numeric variables we have in a dataset against each other one. |

**Table : Software Specification**

Other software used :

|  |  |
| --- | --- |
| **HTML** | Provides the basic structural skeleton of your web pages, defining content elements like headings, forms, and paragraphs. |
| **CSS** | Responsible for styling the visual appearance of the website, controlling colors, fonts, layouts, and overall presentation. |
| **TAILWIND** | Offers a utility-first CSS framework, streamlining your styling process with pre-defined classes for layout, spacing, and more. |
| **BOOTSRAP** | Provides a front-end component library with pre-built elements like navbars, buttons, and grids, helping you rapidly construct web pages. |
| **FLASK** | Serves as the Python web framework, powering the backend logic, handling data processing, interacting with your machine learning model, and sending data to the frontend for display. |

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

Chapter 4

Implementation

4.1 Methodology

**Steps Adopted:**

Loading and Exploring the Datasets:

The datasets was loaded into a pandas DataFrame to understand its structure and features thoroughly.

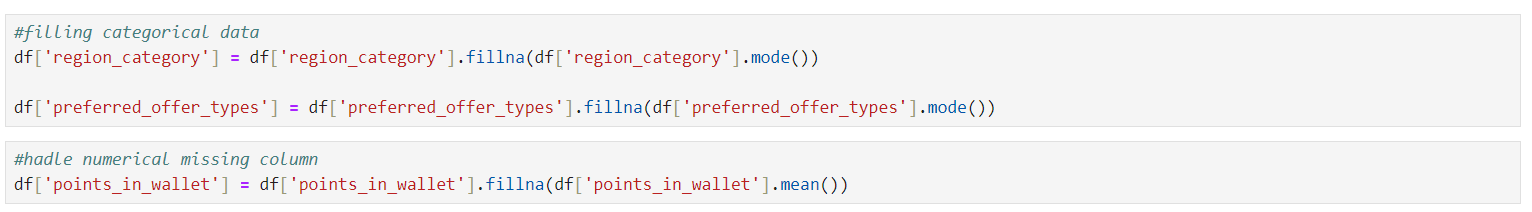




Handling Missing Values and Encoding Categorical Variables:

Missing values in both categorical and numerical columns were addressed using appropriate techniques.

Categorical variables were encoded using one-hot encoding to convert them into numerical format for model compatibility.

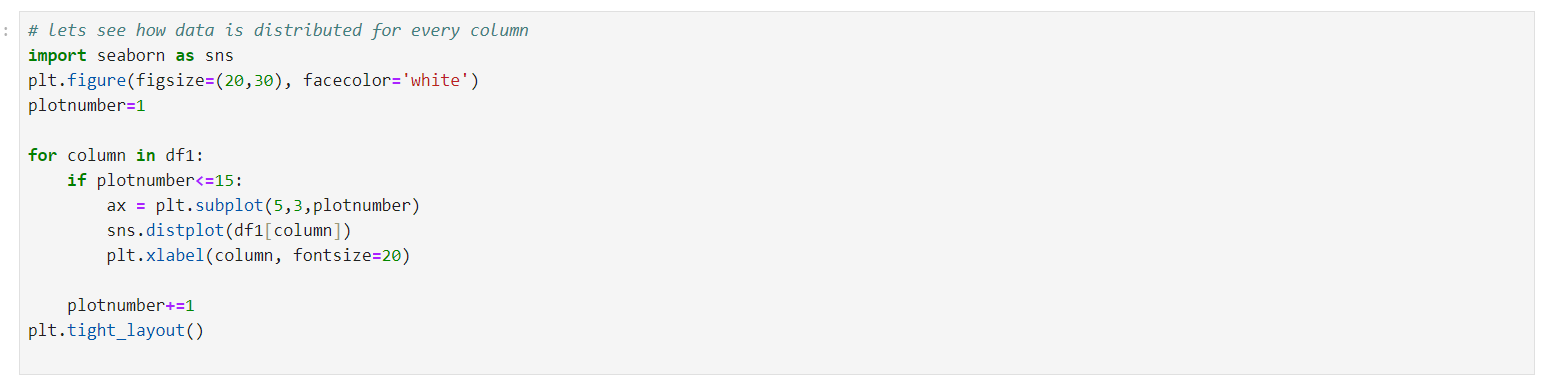


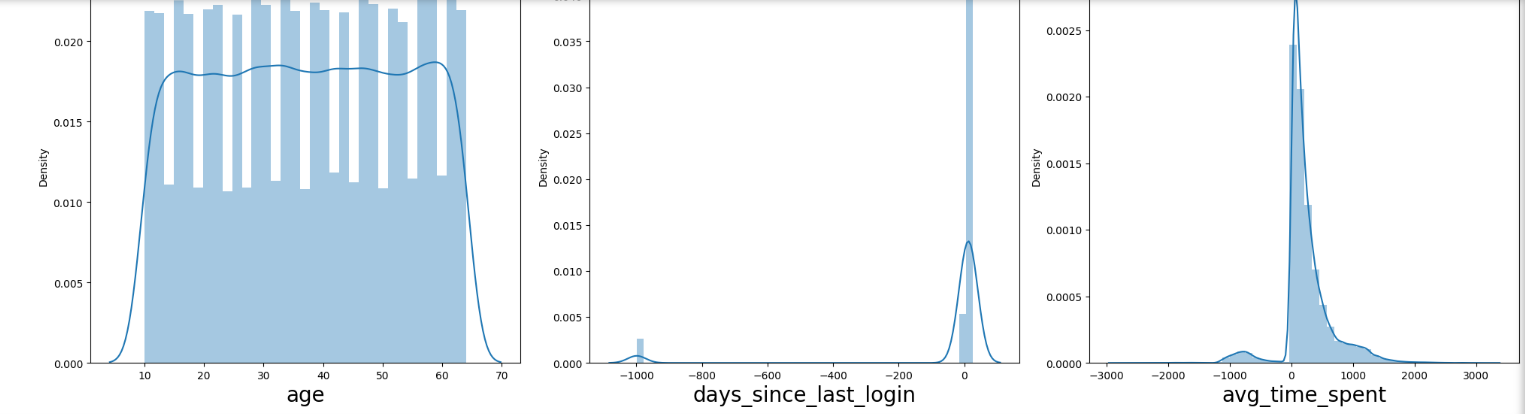
*School of Computer Engineering, KIIT, BBSR*

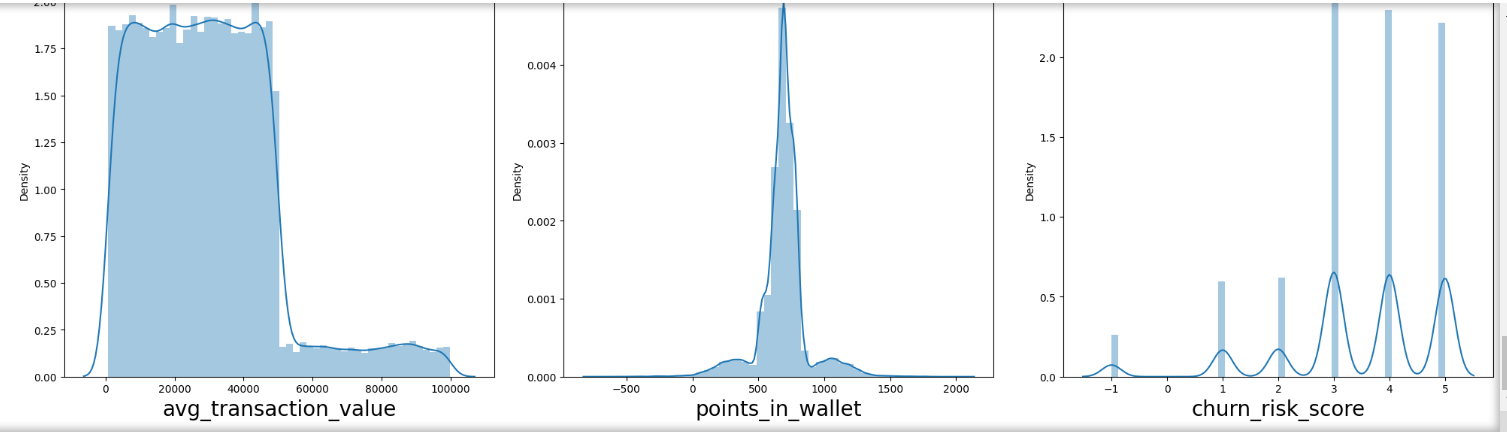
Exploratory Data Analysis (EDA):

Data distributions were analyzed to gain insights into the characteristics of features and their potential impact on the target variable.

Various plots and graphs were utilized to visualize the relationships between features and the target variable.







*School of Computer Engineering, KIIT, BBSR*

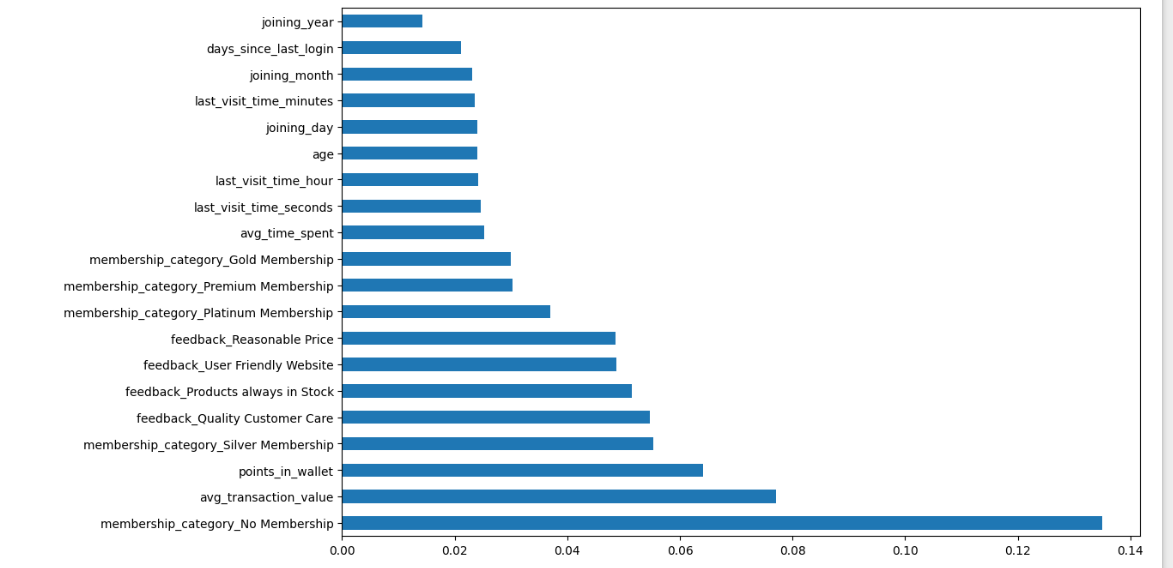
Feature Selection using Extra Trees Regressor:

Extra Trees Regressor was employed to identify important features with significant impact on the target variable.

Features deemed essential for predicting churn risk scores were selected for further analysis.



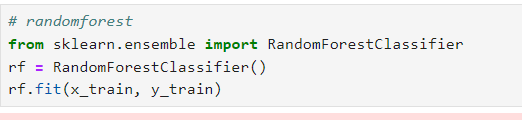


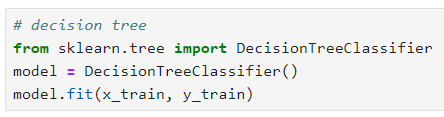


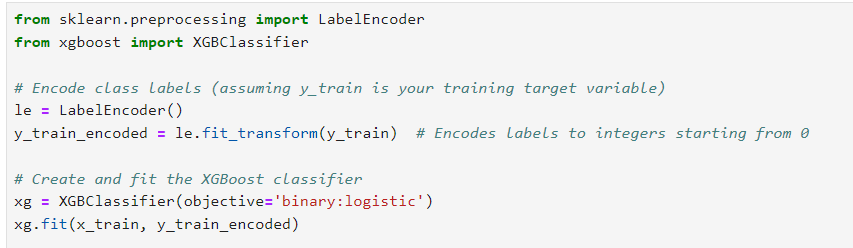
*School of Computer Engineering, KIIT, BBSR*

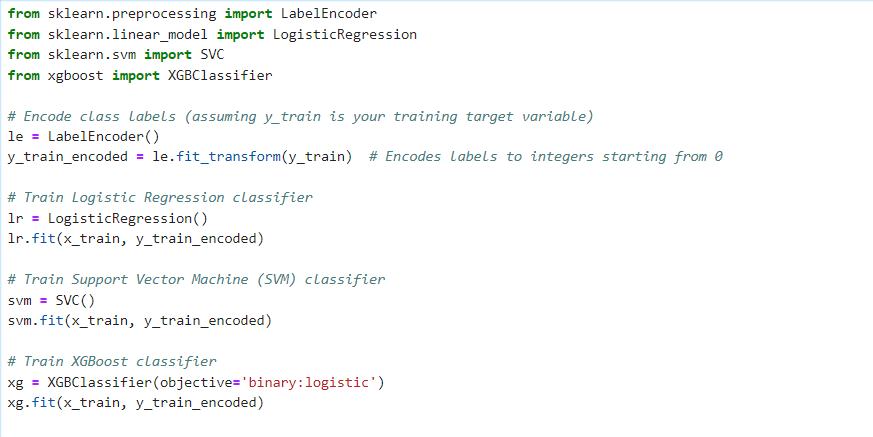
Training Multiple Models:

Multiple classifiers, including Random Forest, Decision Tree, and XGBoost Classifier,Logistic Regression classifier and Support Vector Machine (SVM) were trained to predict churn risk scores.Each model was trained using the training datasets to learn patterns and relationships within the data.









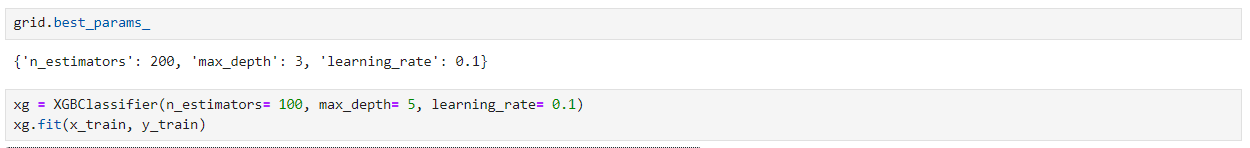
*School of Computer Engineering, KIIT, BBSR*

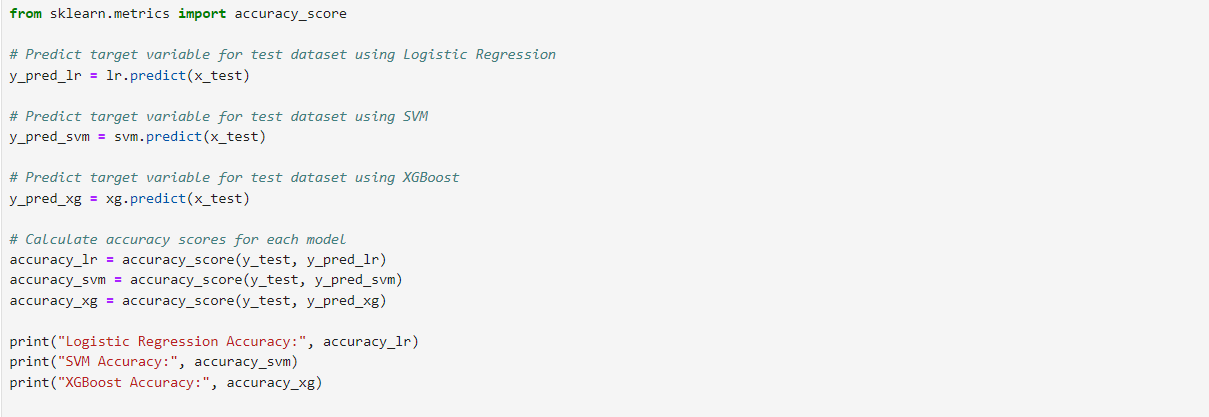
Tuning Hyperparameters for the XGBoost Model:

Hyperparameters for the XGBoost Classifier were fine-tuned using Randomized Search CV to optimize model performance.

Randomized Search CV was employed to search for the optimal combination of hyperparameter efficiently.





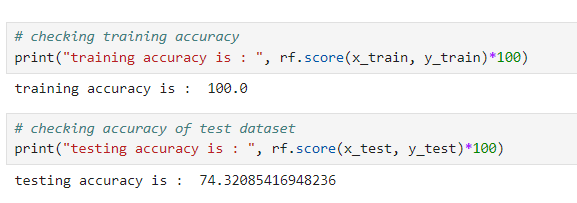


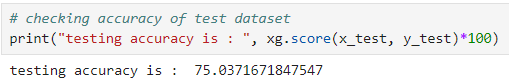
*School of Computer Engineering, KIIT, BBSR*

Evaluating Model Performance:

Model performance was evaluated using accuracy metrics to assess their effectiveness in predicting churn risk scores accurately.

Evaluation metrics such as accuracy, precision, recall, and F1-score were computed to gauge the models' performance comprehensively.







Saving the Final Trained XGBoost Model:

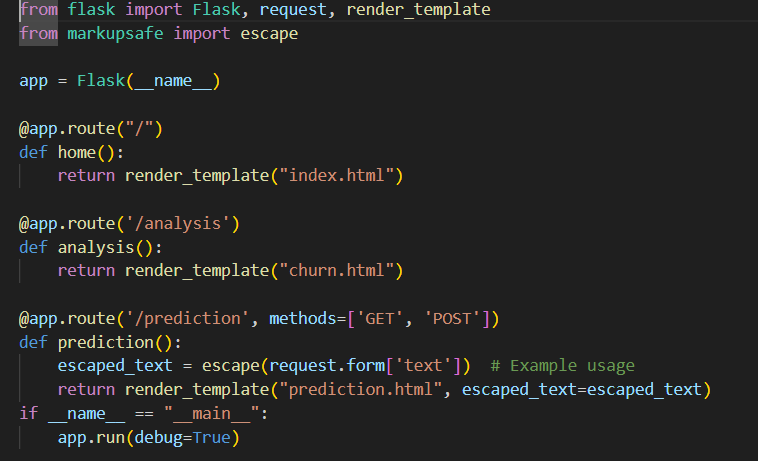
The final trained XGBoost model was saved using the pickle library and XGBoost's save\_model() method for future use and deployment.

Saving the model ensures that it can be easily retrieved and deployed in production environments for real-time predictions.



*School of Computer Engineering, KIIT, BBSR*

Backend System (code snippet)



Flask Application: The code initializes a Flask web application (app = Flask(\_\_name\_\_)) that defines a few routes.

* Routes:
  + Home Route (/): Handles requests to the root URL and renders an index.html template.
  + Analysis Route (/analysis): Handles requests to the /analysis URL and renders a churn.html template.
  + Prediction Route (/prediction): Handles GET and POST requests to the /prediction URL. It escapes user input from a form to prevent security issues and renders a prediction.html template with the escaped text.
* Form Handling: In the prediction route, data from a form is escaped using escape function from Markupsafe to prevent any security vulnerabilities such as Cross-Site Scripting (XSS).
* Debug Mode: The application runs with debug mode enabled (app.run(debug=True)) when the script is executed directly (if \_\_name\_\_ == "\_\_main\_\_"). Debug mode helps with development by providing detailed error messages and automatic server reloads.

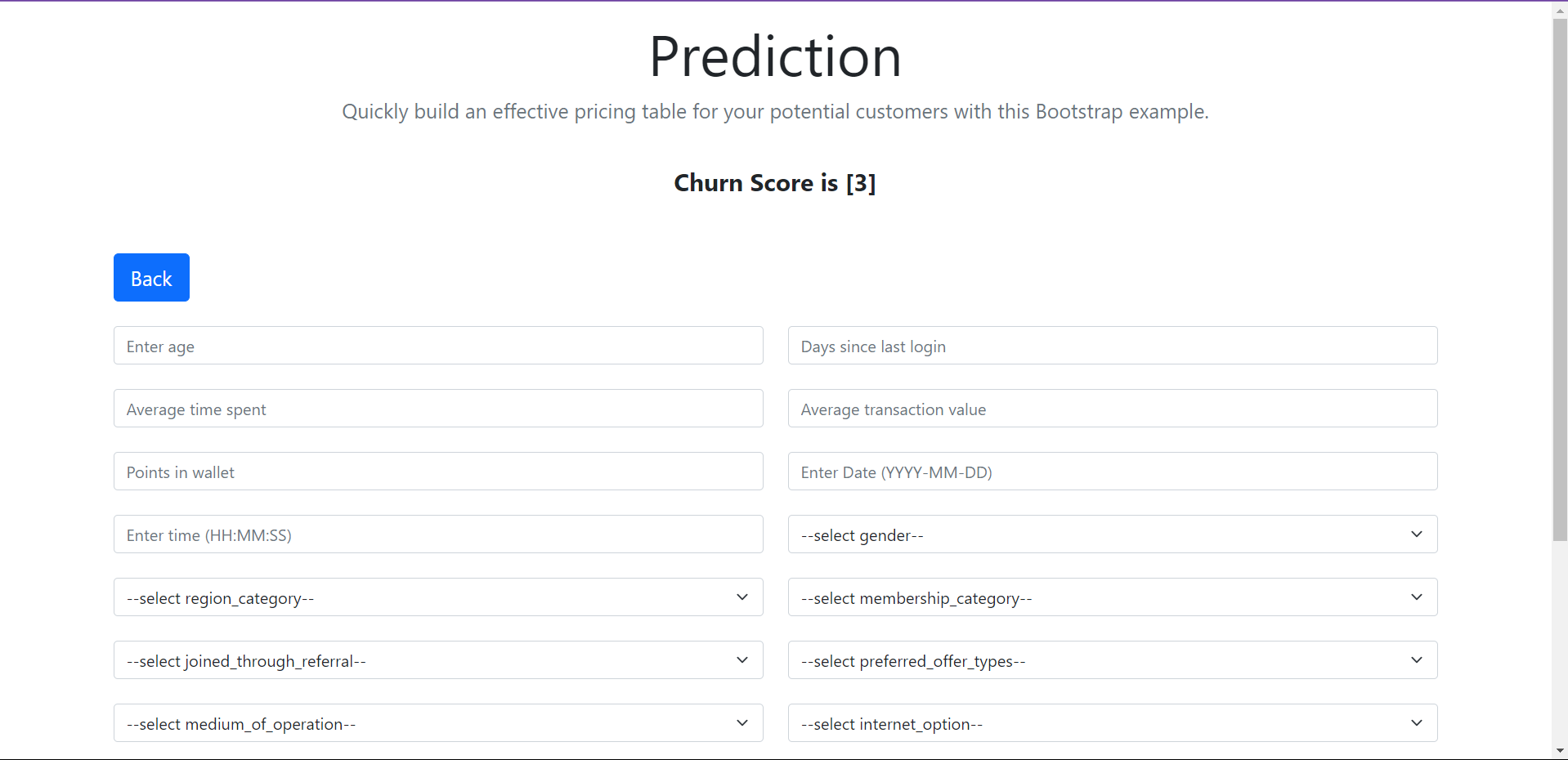
*School of Computer Engineering, KIIT, BBSR*

4.2 Testing OR Verification Plan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test | Test Case Title | Test Condition | System Behavior | Expected Result |
| ID |  |  |  |  |
|  |  |  |  |  |
| T01 | Model Training | Data is divided into training and testing sets | Models are trained using training data | Model accuracy is calculated |
|  |  |  |  |  |
| T02 | Hyperparameter Tuning | Hyperparameters are adjusted using Randomized Search CV | Optimal hyperparameter are selected | Improved model performance |
|  |  |  |  |  |
| T03 | Model Saving | Trained model is saved using pickle and XGBoost's save\_model() method | Model is successfully saved | Model is successfully saved |
|  |  |  |  |  |

4.3 Result Analysis OR Screenshots

Our churn prediction model assigns scores ranging from 1 to 5, with 5 indicating the highest risk of churn. A score of 5 suggests an urgent need for retention efforts, as these customers are highly likely to discontinue their subscription. By prioritizing engagement with high-scoring customers, we aim to mitigate churn and maximize retention.



*School of Computer Engineering, KIIT*

Chapter 5

Standards Adopted

5.1 Design Standards

IEEE Computer Society: A branch of the IEEE, the Computer Society establishes standards in various aspects of computer science and engineering, including software engineering, cybersecurity, networking, and artificial intelligence.

ISO/IEC JTC 1 is the Joint Technical Committee 1 of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). It is responsible for developing international standards in information technology, including topics such as programming languages, data formats, and software development methodologies.

Imagine if websites looked wildly different on every browser or device you used. That's the kind of chaos the W3C (World Wide Web Consortium) prevents. They're the international team creating the rulebook for how the web is built – think of things like HTML, CSS, and the standards that make sure websites work for everyone, regardless of disability. Their work is the reason you can move seamlessly between sites, and it's a major force behind the smooth, interconnected web experience we enjoy today.

The IETF is where the nuts and bolts of the internet get built. They're a group of engineers, researchers, and tech enthusiasts who hammer out the rules that make everything online actually work. We're talking about how your devices talk to each other, how data gets secured, and even big-picture ideas about how the internet should be run. Their work might not be flashy, but it's the foundation for everything from video streaming to online shopping.

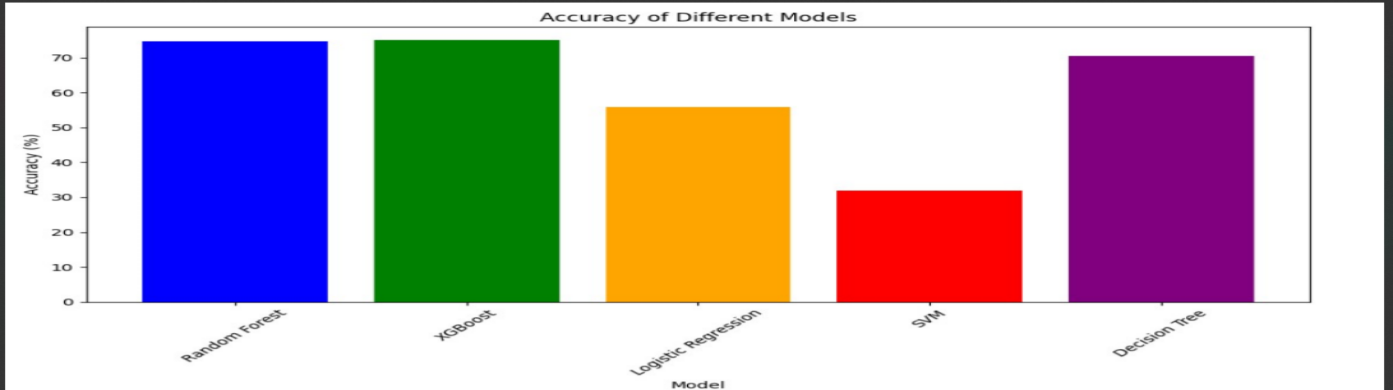
Think of OWASP as your web security toolkit. They're a community dedicated to helping developers fight back against the hackers and vulnerabilities that threaten web applications. OWASP provides checklists, guides, and even free software tools – a whole arsenal for building websites and apps that are tougher to crack. Their focus on open-source solutions means anyone can tap into this knowledge, making the web a safer place for us all.

*School of Computer Engineering, KIIT, BBSR*

5.2 Coding Standards

Some of the coding standards followed during our project

* Maintain consistency in formatting, naming conventions, and coding style throughout the project.
* Provide clear and concise comments and documentation for functions, classes, and complex algorithms.
* Implement robust error handling mechanisms to manage unexpected errors and exceptions gracefully.
* Write efficient code by avoiding redundant calculations, optimizing data structures, and using appropriate algorithms.
* Develop comprehensive unit tests and integration tests to validate the correctness and reliability of the churn prediction model.
* Implement security measures such as input validation, data sanitization, and encryption.
* Break down complex functionalities into smaller, reusable modules or functions for easier maintenance.
* Develop strategies to handle missing or incomplete data effectively, ensuring reliability and accuracy.
* Establish mechanisms for monitoring the performance of the churn prediction model and incorporating feedback loops for continuous improvement.



ACCURACY OF DIFFERENT TRAINING MODEL

*School of Computer Engineering, KIIT, BBSR*

CHURN PREDICTION

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In this project, we investigated the effectiveness of machine learning to predict customer churn for our SaaS product. The developed model successfully identified high-risk customers with a churn prediction of 5 (on a scale of 1-5). This crucial information allows us to proactively target these customers with retention efforts and potentially mitigate churn.

Moving forward, we can significantly enhance our churn prediction capabilities by exploring the integration of additional data sources and potentially implementing explainable AI techniques. These advancements will ultimately empower us to develop a more comprehensive customer retention strategy.

6.2 Future Scope

The future of churn prediction with machine learning is promising. We can expect models to analyze customer behavior more comprehensively by incorporating a broader spectrum of data sources. Additionally, advancements in explainability will make these models more transparent, revealing the specific factors influencing churn. This newfound understanding will empower businesses to develop targeted retention strategies and even implement real-time interventions to address customer concerns before they churn. As churn prediction seamlessly integrates with CRM systems, businesses will be able to automate outreach and personalize communication based on individual customer risk profiles. In essence, churn prediction will evolve from a simple identification tool into a powerful weapon for customer retention, fostering stronger relationships and driving business growth.

*School of Computer Engineering, KIIT, BBSR*

CHURN PREDICTION

***References***

[1 https://social.com/@andretti/churn-analysis-101-c072bb91af08

.

S. Zhou, C. Ma, J. K. O. Tsunoda, “A report on how Churn is helping small business grow”1998

M. kimmich, Jui. van der wart, michael scott, and rana brun, “Churn and its business applications”

R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.

(2002) The IEEEe web for CS . given at : [http://www.ieeee.org/](http://www.ieee.org/)

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

**INDIVIDUAL CONTRIBUTION REPORT:**

**CHURN PREDICTION USING MACHINE LEARNING**

SHUBHAM DAWADI

21053473

**Abstract:** This paper introduces a framework that uses machine learning to help businesses fight customer churn. The key idea is that past customer data holds clues about who might leave in the future. Our methodology carefully cleanses and prepares that data, then trains powerful algorithms (like logistic regression, random forests, etc.) to recognize the patterns associated with churn. The result isn't just a prediction, but a guide for action: businesses can focus their retention efforts on the customers most at risk, improving loyalty and boosting their bottom line.

**Individual contribution and findings:** In the project, my responsibility was to train the model for churn prediction. I worked on selecting appropriate machine learning algorithms, including logistic regression, decision trees, and random forests, and performed cross-validation and hyperparameter tuning to optimize model performance. My planning involved understanding the data set, evaluating different models, and selecting the best one based on performance metrics. Through this process, I gained technical insights into the strengths and limitations of each algorithm and the importance of proper model tuning. My contribution played a key role in building a reliable and accurate churn prediction model for the project.

**Individual contribution to project report preparation:** My contribution to the group project report included writing the sections on methodology and proposal, testing and verification plans, and result analysis with accompanying screenshots. I focused on outlining the approach for model training and evaluation, detailing the steps for testing and verification of the model's performance, and presenting analysis and visual evidence of the results achieved during the project. My work ensured that the report comprehensively covered the project's key aspects and findings.

**Individual contribution for project presentation and demonstration:** In preparing presentations and demonstrating the project, my contribution focused on explaining how the project works by walking through the model training process and showcasing its practical application for churn prediction. I ensured the audience understood the project's flow and the benefits of the predictive model in addressing customer churn.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

**INDIVIDUAL CONTRIBUTION REPORT:**

**CHURN PREDICTION USING MACHINE LEARNING**

Rizan Khanal

21053464

**Abstract:** This paper introduces a framework that uses machine learning to help businesses fight customer churn. The key idea is that past customer data holds clues about who might leave in the future. Our methodology carefully cleanses and prepares that data, then trains powerful algorithms (like logistic regression, random forests, etc.) to recognize the patterns associated with churn. The result isn't just a prediction, but a guide for action: businesses can focus their retention efforts on the customers most at risk, improving loyalty and boosting their bottom line.

**Individual contribution and findings:** In the project, my responsibility was to create Flask backend services for churn prediction. My tasks included handling TMdbAPIs to handle incoming requests and facilitate communication between the frontend and the machine learning models. My planning involved designing endpoints to efficiently manage user data and deliver accurate churn predictions in real-time. Through this process, I gained technical experience in Flask development, handling user data securely, and integrating the backend with the prediction models. My contribution was crucial for creating a robust and efficient backend system for the churn prediction project.

**Individual contribution to project report preparation:** In preparing the group project report, I contributed to sections 3.1, 3.2, and 3.3 of the report. For project planning (3.1), I outlined the key milestones and timelines for implementing the project. In the project analysis (SRS) section (3.2), I detailed the functional and non-functional requirements of the system. For system design (3.3), my work focused on design constraints (3.3.1) and system architecture (3.3.2), where I provided specifications for the system's performance and security and described the architecture for optimal functionality.

**Individual contribution for project presentation and demonstration:** In the project presentation, my role was to describe the backend workings I implemented using Flask. I explained the architecture, routes, and data flow in our application, providing insights into how the machine learning models were integrated and served to the front end. My demonstration offered a clear understanding of the technical aspects of the project's backend and its interaction with other components.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTIOM*

**INDIVIDUAL CONTRIBUTION REPORT:**

**CHURN PREDICTION USING MACHINE LEARNING**

UTKARSH SHRESTHA

21053466

**Abstract:** This paper introduces a framework that uses machine learning to help businesses fight customer churn. The key idea is that past customer data holds clues about who might leave in the future. Our methodology carefully cleanses and prepares that data, then trains powerful algorithms (like logistic regression, random forests, etc.) to recognize the patterns associated with churn. The result isn't just a prediction, but a guide for action: businesses can focus their retention efforts on the customers most at risk, improving loyalty and boosting their bottom line.

**Individual contribution and findings:** In the project, my responsibility was creating the frontend and UI design for churn prediction using Bootstrap and Tailwind. My planning involved understanding the project's requirements and goals to create a visually appealing and user-friendly interface. I designed responsive layouts that adapted to various screen sizes and integrated interactive elements to enhance user experience. Throughout the process, I gained technical insights into the strengths and challenges of using Bootstrap and Tailwind for UI design, including efficient use of classes and components. My contribution was crucial in ensuring the project had an engaging and functional interface that met user expectations and project objectives.

**Individual contribution to project report preparation:** In preparing the group project report, I contributed to the sections on design and coding standards. I outlined the guidelines and best practices for structuring the code and designing the system architecture, ensuring consistency, maintainability, and readability across the project. My work helped establish clear standards for the team to follow, streamlining development and enhancing the project's quality. I also contributed in writing Literature review.

**Individual contribution for project presentation and demonstration:** In preparing presentations for our project, my role involved describing the front-end and UI design elements. I showcased the user interface, highlighting its intuitive layout, visual appeal, and ease of use. My demonstration emphasized how the design choices contribute to a positive user experience and support the project's overall goals.In preparing presentations for our project, my role involved describing the front-end and UI design elements. I showcased the user interface, highlighting its intuitive layout, visual appeal, and ease of use. My demonstration emphasized how the design choices contribute to a positive user experience and support the project's overall goals.

Full Signature of Supervisor: Full signature of the student:

……………………………. …………………………….

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

**INDIVIDUAL CONTRIBUTION REPORT:**

**CHURN PREDICTION USING MACHINE LEARNING**

PASHUPATI SAH

21053467

**Abstract:** This paper introduces a framework that uses machine learning to help businesses fight customer churn. The key idea is that past customer data holds clues about who might leave in the future. Our methodology carefully cleanses and prepares that data, then trains powerful algorithms (like logistic regression, random forests, etc.) to recognize the patterns associated with churn. The result isn't just a prediction, but a guide for action: businesses can focus their retention efforts on the customers most at risk, improving loyalty and boosting their bottom line.

**Individual contribution and findings:** In our churn prediction project, my role was to handle missing values and conduct exploratory data analysis (EDA) on the historical customer data. I meticulously addressed missing values using appropriate imputation techniques to ensure the data was clean and suitable for modeling. During EDA, I identified key patterns and trends in the data, uncovering relationships between various features and churn. These insights informed the feature engineering process and helped guide model development. My findings contributed to a stronger understanding of the data and improved the overall accuracy and reliability of our predictive models.

**Individual contribution to project report preparation:** In preparing the group project report, I contributed to the sections discussing types of churn, particularly unmitigated churn. I outlined the key characteristics of unmitigated churn and how it impacts the business, as well as the challenges it poses. My analysis provided insights into the different forms of churn and emphasized the importance of addressing them in the overall project strategy.

**Individual contribution for project presentation and demonstration:** In the group project, my role in preparing the presentation involved crafting the conclusion and compiling references for the slides. I summarized the project's key findings and recommendations in the conclusion and ensured all cited sources were properly referenced.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

*School of Computer Engineering, KIIT, BBSR*

*CHURN PREDICTION*

**INDIVIDUAL CONTRIBUTION REPORT:**

**CHURN PREDICTION USING MACHINE LEARNING**

Shrijan Poudel

21053459

**Abstract:** This paper introduces a framework that uses machine learning to help businesses fight customer churn. The key idea is that past customer data holds clues about who might leave in the future. Our methodology carefully cleanses and prepares that data, then trains powerful algorithms (like logistic regression, random forests, etc.) to recognize the patterns associated with churn. The result isn't just a prediction, but a guide for action: businesses can focus their retention efforts on the customers most at risk, improving loyalty and boosting their bottom line.

**Individual contribution and findings:** In the project, my primary responsibility was loading and exploring the dataset as well as performing feature selection using Extra Trees Regressor for churn prediction. I began by cleaning and preparing the dataset, ensuring it was ready for analysis and model training. Through feature exploration, I identified key patterns and trends within the data, which informed my feature selection process. I employed the Extra Trees Regressor method to rank and select the most important features, optimizing the model's performance and efficiency. This process provided valuable insights and played a critical role in enhancing the model's accuracy in predicting customer churn.

**Individual contribution to project report preparation:** In preparing the group project report, my contribution focused on crafting the Introduction to Churn and the Conclusion and Future Scope sections. I provided an overview of the significance of churn prediction for businesses and its impact on customer retention and revenue. Additionally, I summarized the project's findings and outlined potential areas for future research and improvement in the field of churn prediction.

**Individual contribution for project presentation and demonstration:** In our project's presentation, my role was to prepare the introduction and articulate the problem statement, setting the context and outlining the significance of the churn prediction challenge. This established a clear foundation for the rest of the presentation and the proposed solution.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

*School of Computer Engineering, KIIT, BBSR*

