High-Level Design (HLD)
Churn Analytics

Revision Number – 1.0

Last Date of Revision: 12 – 05 -2023

ANKIT KUMAR

Document Version Control

Date	Version	Description	Author
23– 02 - 2023	1.0	First Version of Complete HLD	Ankit Kumar

Contents

Document Version Control…….. 2

Abstract…… 4

1. Introduction ……. 5

1.1 Why this HLD Document ? …….. 5

1.2 Scope …… 5

1.3 Definitions …… 6

2. General Description …… 6

2.1 Problem Perspective ……. 7

2.2 Proposed Solution …….. 7

2.3 Technical Requirements ……. 7

2.4 Data Requirements …… 7

2.5 Tools Used …….. 8

3. Design Flow …… 8

3.1 Model Creation and Evaluation ……. 8

4. Performance Evaluation …… 8

4.1 Reusability …… 9

4.2 Application Compatibility ……... 9

4.3 Resource Utilization …… 9

4.4 KPIs …….. 9

5. Conclusion ……. 9

Abstract

In this study, we propose a machine-learning model for predicting customer churn in the Telecom industry. Customer churn, referring to the phenomenon where customers discontinue their subscription or leave a service provider, is a significant concern for businesses as it directly impacts revenue and market competitiveness. Leveraging a comprehensive dataset from the Telecom industry, we employ advanced machine learning techniques to develop an accurate and robust churn prediction model.

The proposed model utilizes various features related to customer behavior, service usage patterns, demographic information, and customer interaction history. By leveraging historical data, our model aims to identify key factors contributing to customer churn and provide insights for proactive customer retention strategy.

To construct the model, we adopt a multi-step approach involving data preprocessing, feature engineering, and model selection. We apply appropriate preprocessing techniques to handle missing values, outliers, and categorical variables. Feature engineering is employed to extract meaningful information from the dataset and enhance the predictive power of the model. Various machine learning algorithms, such as logistic regression, decision trees, random forests, and gradient boosting, are evaluated and compared to identify the most effective algorithm for churn prediction.

To evaluate the model's performance, we employ appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. Additionally, we conduct cross-validation and assess the model's generalization ability. We also investigate feature importance to gain insights into the factors driving customer churn in the Telecom industry.

The results demonstrate the effectiveness of our proposed model in accurately predicting customer churn. The insights gained from this study can assist Telecom industry stakeholders in developing targeted retention strategies, thereby minimizing customer attrition and improving overall business performance. Additionally, the methodology employed in this research can be adapted and applied to other industries facing similar churn prediction challenges, facilitating proactive customer management and enhanced business decision-making.

1. Introduction

1.1 Why this HLD Document ?

The main purpose of this HLD document is to feature the required details of the project and supply the outline of the Model Creation, Evaluation, and Deployment. This additionally provides a careful description of how the complete project has been designed end-to-end. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will :

- Present all of the design aspects and define them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and architectural design of the project.
- List and describe the non - functional attributes like :
    1. Security
    2. .Reliability
    3. Maintainability
    4. Portability
    5. Resource
    6. Utilization

1.2 Scope

The HLD documentation presents the structure of the system, such as database design, architectural design, application flow and technology architecture. The HLD uses non-technical terms to technical terms that can be understandable to the administrator of the system.

1.3 Definitions

Term	Description
Churn	The churn rate is a measure of the number of customers or employees who leave a company during a given period.
Database	Collection of all the information used by the System.
Jupyter - Notebook	The Jupyter Notebook is an open source web application that we use to create and share documents that contain live code, equations, visualizations, and text.

2. General Description

2.1 Problem Perspective

Churn rate analytics is a valuable tool for businesses to understand and manage customer attrition, also known as customer churn. Customer churn refers to the rate at which customers stop doing business with a company over a specific period of time. Analyzing churn rates helps businesses gain insights into customer behavior, identify potential issues, and develop strategies to improve customer retention. Here are some specific uses of churn rate analytics for a business:

1. Retention Strategy: Churn rate analytics helps identify the reasons behind customer churn. By analyzing patterns and customer behavior, businesses can uncover the factors leading to attrition. This information allows them to develop effective retention strategies to minimize churn. For example, if customers are leaving due to poor customer service, the company can focus on improving its support processes.
2. Customer Segmentation: Analyzing churn rates across different customer segments can provide valuable insights. By understanding which segments have higher churn rates, businesses can tailor their marketing and retention efforts accordingly. This may involve offering personalized incentives, targeted promotions, or improving product features to address the specific needs and concerns of different customer groups.
3. Product Improvement: Churn rate analytics can help identify product or service deficiencies that contribute to customer attrition. By examining feedback and behavior patterns, businesses can uncover areas where their offerings fall short or fail to meet customer expectations. This information can guide product

improvements, updates, or even the development of new offerings to enhance customer satisfaction and reduce churn.

4. Customer Lifetime Value (CLV): Churn rate analytics allows businesses to calculate customer lifetime value, which estimates the revenue a customer is likely to generate over their entire relationship with the company. By factoring in churn rate, businesses can assess the impact of customer attrition on their revenue and profitability. This information helps prioritize customer acquisition efforts and allocate resources to retain high-value customers.
5. Proactive Customer Retention: Churn rate analytics enables businesses to predict which customers are at a higher risk of churning in the future. By utilizing predictive models and algorithms, businesses can identify early warning signs and proactively engage with at-risk customers to prevent churn. This may involve personalized outreach, targeted offers, or special attention to ensure customer satisfaction and loyalty.

In summary, churn rate analytics provides valuable insights into customer behavior, helps businesses understand the reasons behind customer attrition, and enables them to develop targeted strategies for customer retention. By leveraging this information, businesses can enhance customer satisfaction, improve their products or services, and ultimately increase customer loyalty and long-term profitability.

2.2 Proposed Solution

To solve the problem I have used various BI tools and Machine Learning Techniques to derive insights from the data so that major steps could be taken by stakeholders to eliminate the problem of churn.

2.3 Technical Requirements

As for technical requirements, the system is fully checked to take the load of data and processing it through various machine learning algorithms.

2.4 Data Requirements

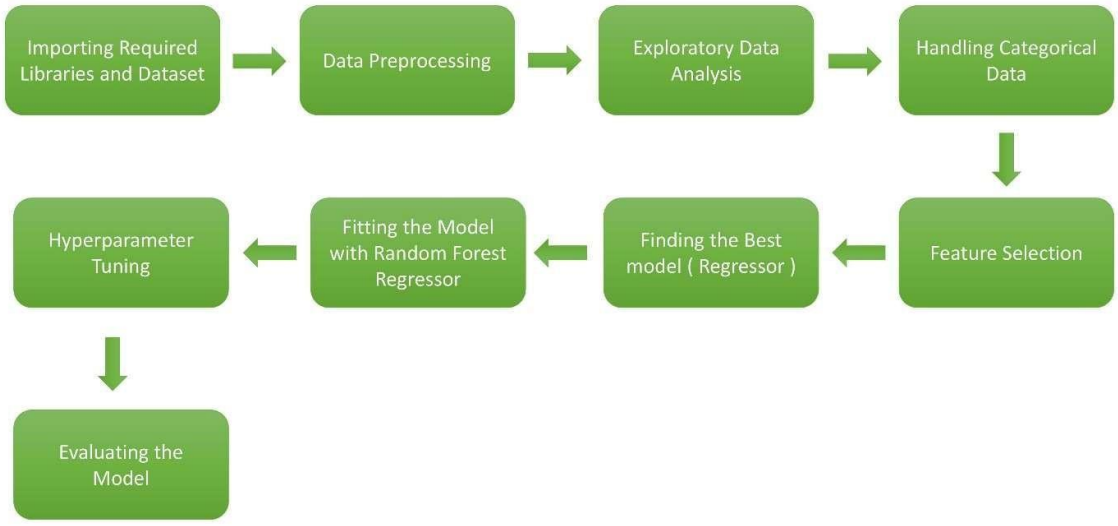
The Data requirements totally supported the matter statement and also the dataset is accessible on the Kaggle within the file format of (.xlsx). Because the main theme of the project is to induce the expertise of real time issues, we have a tendency to transform the information into the prophetess database and commerce it into csv format.

2.5 Tools Used

- Python 3.9 is employed because of the programming language and frameworks like NumPy, Pandas, Scikit - learn and alternative modules for building the model.
- Jupyter - Notebook is employed as an IDE.
- For Data visualizations, seaborn and components of matplotlib are getting used.
- For information assortment prophetess info is getting used.
- GitHub is employed for version management.

3. Design Flow

3.1 Modelling Process



.

4. Performance Evaluation

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different parts of the system are communicating or using Python as an interface between them. All the components have its own tasks to perform and it is the job of a Python to ensure proper transfer of data.

4.3 Resource Utilization

When a task is performed, it'll doubtless use all the process power offered till the process is finished.

4.4 KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Housing Price and its relationship with different

metrics

1. Impact of gender on Churn.

2. Impact of Senior Citizen on Churn.

3. Influence of Payment Methods on Churn.

5. Conclusion

The Flight Fare Prediction system will predict the churn for helping the company with the trained knowledge with set of rules. The company can use this information to manage their business and customer.