SUBMISSION OF PROJECT SYNOPSIS AND GUIDE ACCEPTANCE FORM

(To be submitted to the Project Steering Committee)

PART A: Synopsis Registration

I. Student Details:

1. Name of the Program: Master of Business administration (MBA)

2. Name of the Student: Rahul Singh

3. Roll Number: 2314107786

4. Session & Year: Sep 2024 – Mar 2025

5. Elective: Data Science & Analytics

II. Project Details

6. Title of the Project: Customer Churn Prediction and Analysis for a Subscription-Based Business

7. Introduction and Review of Literature:

Introduction: In today's competitive business environment, customer retention is crucial, particularly for subscription-based companies where recurring revenue is essential. Customer churn—the process by which users discontinue their subscriptions—poses a significant challenge as it directly impacts revenue and growth. Studies show that retaining existing customers is often more cost-effective than acquiring new ones, making churn prediction a valuable tool for business sustainability.

This project aims to predict churn in a subscription-based business using data science and machine learning techniques. By analyzing customer data such as usage patterns, demographic information, and engagement history, this project seeks to identify the key indicators of churn. Predictive analytics provides a proactive solution, allowing businesses to identify customers at risk of churning and implement targeted retention strategies, such as personalized offers and enhanced customer support, to increase loyalty and reduce churn rates.

Review of Literature

Research on customer churn prediction is extensive, especially within subscription-based and SaaS (Software as a Service) industries. Three major themes have been identified in the literature: churn prediction techniques, factors influencing churn, and case studies on successful churn reduction.

1. **Churn Prediction Techniques**

Various machine learning algorithms have been used in churn prediction, including logistic regression, decision trees, support vector machines (SVM), and ensemble methods like random forests and gradient boosting. Hadden et al. (2007) suggest that logistic regression is popular for its interpretability, though machine learning models such as decision trees and neural networks generally provide better accuracy in churn prediction. In recent years, neural networks have gained attention for their ability to detect complex patterns in data (Ghaziri & Laroche, 2018). Additionally, hybrid approaches like stacking or boosting combine multiple algorithms to improve prediction accuracy (Huang et al., 2018).

2. **Factors Influencing Customer Churn**

The literature identifies common factors associated with customer churn, such as engagement levels, usage frequency, payment history, and customer satisfaction scores. Socio-demographic variables like age, location, and income level have also shown predictive value (Kim et al., 2016). For instance, low engagement or delayed payments may signal potential churn. Subscription tenure—the duration of a customer's relationship with the business—is another key variable, with research by Gupta et al. (2019) suggesting that new subscribers are at a higher risk of churn compared to long-term users.

3. **Case Studies and Applications**

Numerous case studies illustrate successful applications of churn prediction models across industries like telecommunications, financial services, and SaaS. Verbeke et al. (2012) demonstrated that implementing churn prediction in the telecom industry, combined with retention strategies, reduced churn rates by up to 15%. In financial services, Zhang et al. (2020) applied churn prediction in credit card subscriptions, where identifying high-risk customers and providing tailored retention offers significantly boosted customer retention.

Conclusion of Review of Literature

The literature emphasizes that effective churn prediction combines the right algorithms with an understanding of domain-specific factors contributing to customer disengagement. This project builds on existing insights to develop a tailored churn prediction model for a subscription-based business. By leveraging data analytics and advanced machine learning techniques, the model will seek to identify high-risk customers and support strategic interventions to improve retention and foster long-term customer loyalty.

8. Objectives of the Study: The primary objective of this study is to develop a predictive model for identifying customers at risk of churning in a subscription-based business. By leveraging data science and machine learning techniques, this project seeks to analyze various customer attributes, including behavioral patterns, demographic data, and subscription history, to uncover insights into the factors contributing to churn.

Key objectives include:

1. **Identify Key Indicators of Churn**: Analyze customer data to determine the most significant factors associated with churn, such as engagement frequency, payment history, and customer satisfaction levels.

2. **Build a Predictive Model**: Utilize machine learning algorithms to create a model capable of accurately predicting high-risk customers, enabling early intervention to improve customer retention.

3. **Enable Targeted Retention Strategies**: Provide actionable insights that allow the business to design personalized retention strategies, such as customized offers or proactive customer support, aimed at high-risk segments.

4. **Enhance Business Profitability**: By reducing churn rates, the study aims to help increase customer lifetime value, contributing to long-term profitability and growth.

This study will contribute to a deeper understanding of churn patterns and help in creating data-driven retention strategies tailored to the subscription-based business model.

9. Research Methodology and References: Research Methodology

The research methodology for this study on customer churn prediction in a subscription-based business involves a multi-step approach that combines data collection, data processing, model development, and evaluation. The aim is to build a predictive model that accurately identifies high-risk customers, allowing the business to implement proactive retention strategies.

1. **Data Collection:** The study will use historical customer data, including demographic information (e.g., age, location), behavioral data (e.g., login frequency, usage patterns), and transactional data (e.g., payment history, subscription length). This data can either be obtained from internal business records or simulated datasets from public sources if access to real data is limited.

2. **Data Processing**: Raw data will be pre-processed to ensure accuracy, consistency, and completeness. This includes handling missing values, removing duplicates, and standardizing data formats. Exploratory Data Analysis (EDA) will then be conducted to identify trends and relationships within the data, including identifying potential churn indicators. Feature engineering will also be applied to create new variables, such as customer engagement scores or subscription tenure, that could improve the predictive model's performance.

3. **Model Selection and Development**: A range of machine learning algorithms will be considered, including logistic regression, decision trees, random forests, and gradient boosting models. Initial models will be trained using a subset of the data, with hyperparameter tuning to optimize model performance. Advanced algorithms, such as ensemble methods or deep learning models, may also be explored based on model accuracy and interpretability.

4. **Model Evaluation**: The performance of each model will be evaluated using standard metrics such as accuracy, precision, recall, and the F1-score. Cross-validation will be used to ensure the model's robustness across different data subsets, and the best-performing model will be selected for implementation.

5. **Interpretability and Insights**: The chosen model will be analyzed to identify key factors influencing customer churn. Interpretability

techniques, such as feature importance or SHAP (Shapley Additive Explanations) values, will help in understanding the impact of each variable on churn prediction. The insights will be used to recommend actionable retention strategies.

6. **Implementation and Recommendations**: Finally, a report will be generated with model insights and suggested retention strategies for high-risk customer segments. Business recommendations might include tailored offers, enhanced customer engagement initiatives, or targeted communication plans based on identified churn drivers.

References

- Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2007). "Computer assisted customer churn management: State-of-the-art and future trends." *Computers & Operations Research,* 34(10), 2902-2917.

- Ghaziri, H., & Laroche, P. (2018). "Application of Deep Learning Techniques in Customer Churn Analysis." *Journal of Business Research,* 87, 93-104.

- Gupta, S., Hanssens, D. M., & Hardie, B. G. (2019). "Modeling Customer Lifetimes and Customer Equity." *Journal of Marketing Research,* 56(4), 564-578.

- Huang, Y., Wang, H., & Lin, L. (2018). "Machine Learning Approaches for Customer Churn Prediction." *Knowledge-Based Systems,* 151, 10-22.

- Kim, Y., Park, H., & Choi, S. (2016). "Churn Prediction Model Using Support Vector Machine in Telecom Industry." *Expert Systems with Applications,* 38(6), 777-783.

- Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2012). "Building comprehensible customer churn prediction models with advanced rule induction techniques." *Expert Systems with Applications,* 38(3), 2354-2364.

- Zhang, Z., Lu, X., & Li, Y. (2020). "Using Machine Learning to Predict Customer Churn in Subscription-Based Business Models." *IEEE Transactions on Knowledge and Data Engineering,* 32(8), 1580-1586.

This methodology outlines a systematic approach to developing and implementing a churn prediction model, while the references provide foundational and recent studies that contribute to the understanding of customer churn analytics.

III. Guide Details:

10. Name of Proposed Guide: Saransh Kapoor

11. Guide Registration No. (If available): 722156

12. Designation: Senior Manager

13. Affiliation: RBL Bank Limited

14. Qualification: Master of Business administration (MBA)

15. Total Experience: 13+ Years

16. Communication Address: P3, 317, 4th Floor, Jalvayu Vihar, Sector 21, Noida - 201301

17. Contact No.: +91 78498 11211

18. E-mail ID: saranshkapoor87@gmail.com

PART – B: Guide Acceptance

I, Dr./Mr./Mrs. **Saransh Kapoor** with working as hereby confirm my willingness to guide Mr./ Ms. Reg No. **Rahul Singh (2314107786)** for the topic **"Customer Churn Prediction and Analysis for a Subscription-Based Business"** during the period 2024 – 2025.

Place: Noida

Date: 10-11-2024   (Signature of the Guide)

(Note: A Guide needs to get registered with the University if he/ she is guiding a MUJDOE project for the first time. Guide Registration form can be downloaded from the LMS portal)

DECLARATION

I hereby declare that this project synopsis is an original work carried by me and has not been/will not be submitted to any other University for fulfilment of any course of study.

Place: Noida

Date: 10-11-2024

(*Filled in application forms to be signed by both student and the Guide. Forms must be scanned in either .pdf/.doc format and submitted through the LMS student's Login. For uploading, please refer section 2.1.5 of this document)

Rahul Singh

(Signature of the Student)