**Churn Prediction in Telecommunications Sector using Machine Learning Techniques**

**SD21037**

**WONG ZI MING**

**BSD2712 RESEARCH METHODOLOGY**

**ASSIGNMENT 1**

**CHAPTER 1**

**INTRODUCTION**

* 1. **Research Background**

In today's highly competitive business environment, customer loyalty is crucial for maximizing revenues. This makes customer retention a top priority for businesses seeking to maintain a competitive edge. Acquiring new customers is expensive while retaining existing ones is cost-effective and easier.

The telecommunications industry has become increasingly important in developed countries due to rapid technological advancements and the growing number of operators. As competition intensifies, businesses in this industry are constantly looking for strategies to retain their customer base and minimize customer churn, which refers to customers switching to another provider. Three main strategies have been proposed to generate more revenues: acquiring new customers, upselling to existing customers, and increasing customer retention period. However, research has shown that retaining existing customers is the most profitable strategy, as it costs less than acquiring new customers and is generally considered easier than upselling.

To implement the strategy of customer retention, businesses need to accurately predict customer churn, i.e., identify customers who are likely to stop using their services soon. Churn prediction is a crucial task for many businesses, especially in the telecommunications sector, as it allows them to proactively identify customers at risk of churn and take targeted actions to retain them. Customer churn can have a significant impact on a company's bottom line, leading to decreased revenue and increased costs associated with acquiring new customers.

Machine learning, a subfield of artificial intelligence, offers powerful tools for predicting customer churn in the telecommunications sector. By analysing large and complex datasets, machine learning algorithms can uncover hidden patterns and relationships that may not be apparent to human analysts. Common algorithms used for churn prediction include Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and gradient boosting, each with its characteristics in terms of assumptions, complexity, interpretability, and accuracy. These algorithms can be applied to historical customer data on behaviours, preferences, demographics, and interactions to build predictive models that estimate the probability of churn for each customer. These models can then be used to identify customers at risk of churn and design targeted interventions to retain them. The research background in this area encompasses data collection and pre-processing, feature engineering and selection, model development and evaluation, and model deployment and maintenance.

However, there are challenges and opportunities in this research area. Dealing with imbalanced and noisy data, incorporating customer feedback and sentiment analysis, exploring deep learning and ensemble methods, and developing explainable and robust models are some of the challenges that researchers and practitioners need to address. Despite these challenges, churn prediction in the telecommunications sector using machine learning holds promise in improving customer satisfaction, loyalty, and profitability for telecom companies. By accurately identifying customers at risk of churn and taking proactive actions to retain them, businesses in the telecommunications industry can enhance their competitiveness and drive sustainable growth.

* 1. **Problem Statement**

Customer retention is a significant challenge for telecommunications companies, as customers have multiple options for better and cheaper services and churn, or discontinuation of services, can lead to revenue loss. The primary causes of churn are often non-satisfaction with service or better offerings from competitors, resulting in customers feeling undervalued by their service provider. This not only impacts the revenue of telecommunications companies but also affects the market competition, leading to higher prices, reduced service quality, and negative economic impact. To address this problem, machine learning techniques can be utilized to predict customer churn and proactively take measures to retain customers. Various machine learning algorithms, such as Logistic Regression, Random Forest, Decision Tree, Neural Networks, K-means clustering, and Support Vector Machine, can be employed for churn prediction in the telecommunications sector. However, there is no one-size-fits-all solution, and it is crucial to explore different options and evaluate their performance using appropriate metrics and validation methods. This research aims to develop an effective churn prediction model using machine learning techniques, identify the key factors that contribute to customer churn, and develop strategies for reducing churn and improving customer retention.

加多一句benefits to who?

* 1. **Research Questions**

The research questions in this study are:

1. How to identify the churn key factors that contribute to customer churn in the telecommunications sector?
2. How to apply different machine learning techniques to predict customer churn in telecommunications sector?
3. What are the key factors that contribute to customer churn in the telecommunications sector? 换去How can the final output be visualize?……
   1. **Research Objectives**

The research objectives in this study are:

1. To identify the churn key factors that contribute to customer churn in the telecommunications sector through a comprehensive review of existing literature and analysis of customer data.
2. To develop and evaluate different machine learning techniques for predicting customer churn in the telecommunications sector.
3. To explore how demographic and geographic factors influence customer churn in the telecommunications sector and how machine learning can be used to identify these factors. 换去To visualize the final output using Tableau…..
   1. **Scope of Research**

The duration of the research will vary depending on factors such as data size, the complexity of models, resource availability, and stakeholder objectives, with an estimated timeframe of six months to one year based on previous studies. Theories that may be employed include logistic regression, decision trees, and neural networks, each with its strengths and limitations. The population for this research will be customers of telecommunications companies, with a large and diverse sample expected to be included from multiple regions and demographic groups. The research location will be global, considering the global nature of the telecommunications industry, and may focus on specific regions or countries based on data availability and research needs. Limitations of the study may include data availability and quality, the complexity of the telecommunications sector, ethical concerns related to customer data usage and privacy, and the generalizability of research findings to different companies and contexts. 放在其中里面The software that will be used in this study is Tableau and Python3 through Jupyter Notebook.

* 1. **Significance of Research**

The significance of this research lies in its potential to develop an effective churn prediction model using machine learning techniques, which can provide valuable insights for telecom companies to improve customer retention, optimize resources, and increase revenues. The research can contribute to the advancement of knowledge and practice in churn prediction and machine learning by exploring different models, data sources, and techniques in a real-world setting. The findings may also lead to the development of new machine-learning models that can be applied in other industries and contexts. Furthermore, the research can benefit customers by improving their overall experience with telecom companies through personalized offers and recommendations, leading to increased satisfaction and loyalty. Additionally, the research can contribute to the academic community by providing new insights into customer behaviour analysis, marketing, and machine learning, and inspiring future research in related areas.