

I. INTRODUCTION

The demand for telecommunications services in Indonesia has increased significantly in recent years. According to data from BPS (Statistics Indonesia) in 2022, 66.48% of Indonesia's population has accessed the internet, and 86.54% of households are connected to the internet. This statistic highlights the essential role of telecommunications in the daily lives of Indonesian citizens, as digital connectivity becomes a fundamental part of modern society [1]. This increase in internet penetration and reliance on digital services has led to an increasingly competitive telecommunications market, where providers must continuously adapt to meet the evolving expectations and needs of their customers.

In this competitive landscape, customer preferences for telecommunications services have diversified. Customers now have numerous options regarding service packages, pricing, and promotions offered by different providers [2]. As a result, customer behavior has become more dynamic, with increased churn rates as customers frequently switch providers in pursuit of better service or more attractive offers. Understanding and predicting customer churn has therefore become a critical area of focus for telecom providers, as retaining customers is often more cost-effective than acquiring new ones. Additionally, evaluating the financial implications of churn prediction through cost-benefit analysis can provide companies with actionable insights to optimize retention strategies and maximize profitability [3].

To address this challenge, telecom providers are increasingly adopting predictive models to identify customers at high risk of churn. Accurate churn prediction models enable providers to implement targeted retention strategies, potentially improving customer satisfaction and reducing churn rates. However, predicting churn becomes increasingly complex as data volume and feature dimensions grow, necessitating the selection of the most relevant features to improve model interpretability and performance. Recent studies have shown that attribute selection methods, such as Chi-squared and ANOVA tests, play a vital role in isolating significant features that drive churn, ultimately enhancing prediction accuracy when integrated into machine learning models like Support Vector Machine (SVM) [4]. This selective approach aids in reducing model complexity and maintaining high accuracy, even when working with large datasets.

Among various machine learning algorithms, the Random Forest (RF) algorithm has demonstrated superior performance for classification tasks, especially when compared to algorithms such as SVM, K-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA). Random Forest is particularly advantageous for handling datasets with a large number of variables and features, as it effectively eliminates irrelevant features to enhance model accuracy and performance [5]. Moreover, ensemble-based models like Random Forest have proven particularly suitable for the telecommunications sector due to their ability to handle high-dimensional and imbalanced datasets, offering robust solutions for telecom churn prediction

[6]. Metrics such as ROC-AUC are particularly important when evaluating these models, as they provide a comprehensive assessment of performance, even on imbalanced datasets [7].

The Random Forest (RF) algorithm, as an ensemble-based model, is designed to improve prediction accuracy and robustness by aggregating multiple decision trees. It leverages bootstrapping (bagging) techniques to construct a diverse set of trees and randomly selects features for each split, effectively minimizing overfitting and enhancing generalization. These properties make Random Forest particularly effective for complex and high-dimensional datasets, such as those frequently encountered in customer churn prediction, where model stability and precision are crucial for accurately identifying potential churners [8].

This study focuses on leveraging the Random Forest algorithm to predict customer churn using the Telco Customer Churn dataset. By identifying the most influential features that contribute to customer churn, the study aims to provide actionable insights for telecom providers to enhance their customer retention strategies. Through this research, we seek to bridge the gap in understanding churn behavior in the telecom industry and demonstrate the effectiveness of the Random Forest algorithm in this context.