

TELECOM CHURN PREDICTION USING DATA SCIENCE TECHNIQUES

ABSTRACT

Telecom churn prediction is the process of using data analytics and machine learning techniques to identify customers who are likely to cancel their service with a telecom company. The objective is to proactively target these at-risk customers with retention campaigns or other interventions before they leave. The input data for churn prediction can include demographic information, call history, billing information, customer usage patterns, and interactions with customer service. The output is a churn prediction score for each customer, indicating the likelihood that they will cancel their service. By reducing churn, telecom companies can improve customer loyalty, reduce customer acquisition costs, and increase revenue.

INTRODUCTION

The telecommunications industry is characterized by intense competition and a constant need to retain customers. Customer churn, or the rate at which customers switch to other service providers, is a significant challenge for telecom companies. Churn prediction models can help identify customers who are at risk of leaving, allowing companies to take proactive measures to retain them. Such models can also help telecom companies to allocate resources more efficiently and effectively, as well as improve customer satisfaction by addressing issues before they become critical.

Over the years, several studies have addressed the problem of churn prediction in the telecommunications industry, and different techniques have been proposed to address this issue. Traditional statistical methods such as logistic regression and decision trees have been used for churn prediction, and more recently, machine learning-based methods such as neural networks and ensemble models have also been explored. These techniques can leverage the power of data to identify patterns and relationships that are difficult to detect manually.

In this project, we propose a machine learning-based approach for telecom churn prediction and evaluate its effectiveness on a real-world dataset. We compare the performance of several algorithms, including logistic regression, decision trees, random forests, and neural networks. We evaluate the models using metrics such as accuracy, precision, recall, and F1-score and tune hyperparameters to optimize performance. Our results demonstrate the potential of machine learning techniques for this application and provide insights into the factors that drive customer churn in the telecommunications industry.

BUSINESS SCENARIO

In a business scenario for telecom churn prediction, a telecommunications company would use data analysis techniques to identify customers who are at high risk of leaving the company and taking their business to a competitor. By predicting customer churn, the company can take proactive measures to retain these customers and improve customer satisfaction.

Telecommunications companies face intense competition in the marketplace, and customer churn can have a significant impact on their revenue and profitability. In addition to losing the customer's business, the company may also incur costs associated with acquiring new customers to replace those who have left.

By implementing a telecom churn prediction model, the company can identify customers who are at high risk of churn and target them with personalized retention offers or customer service interventions. For example, the company may offer discounted rates, loyalty rewards, or improved customer service to customers who are at high risk of leaving.

Overall, a telecom churn prediction model can help a telecommunications company improve customer retention, reduce churn rates, and ultimately increase revenue and profitability.

RELATED WORK

Numerous studies have addressed the problem of churn prediction in the telecommunications industry. Some of the most widely used techniques include logistic regression, decision trees, and support vector machines. Logistic regression is a statistical method used to model the probability of a binary outcome, such as customer churn. Decision trees are a type of supervised learning algorithm that can be used for classification and regression tasks, and are particularly effective at handling non-linear relationships between features. Support vector machines are a type of machine learning algorithm that can be used for both classification and regression tasks, and are particularly effective at handling high-dimensional data.

More recently, deep learning-based methods such as neural networks have also shown promise in churn prediction. Neural networks are a type of machine learning algorithm that can learn complex relationships between features by using multiple layers of interconnected nodes. They have been successfully applied in various domains and are particularly effective at handling large, high-dimensional datasets.

While several studies have compared the performance of different churn prediction techniques, few have compared the performance of these techniques on the same dataset. Additionally, there is a need to explore more advanced feature engineering techniques and ensemble models to further improve model performance.

PRE-REQUISITE BACKGROUND

To understand the problem of telecom churn prediction and the proposed solution, some pre-requisite background in data analysis, machine learning, and telecommunications industry may be helpful.

Firstly, a basic understanding of data analysis is necessary to work with the datasets typically used for churn prediction. This includes familiarity with data cleaning and pre-processing techniques, as well as feature engineering, which involves selecting and transforming relevant features from the raw data. Exploratory data analysis (EDA) is also an important step in identifying patterns and relationships in the data.

Secondly, knowledge of machine learning algorithms and their application to classification tasks is important. This includes understanding of various models such as logistic regression, decision trees, random forests, and neural networks, as well as their strengths and weaknesses in different scenarios.

Lastly, some familiarity with the telecommunications industry is helpful to understand the unique challenges and opportunities in this domain. This includes knowledge of the different types of services provided by telecom companies, such as voice, data, and messaging services, as well as the various factors that can influence customer churn, such as pricing, network quality, customer service, and promotions.

Having a solid understanding of these concepts is important for effectively implementing a churn prediction model and interpreting the results.

PROPOSED SOLUTION AND EXPERIMENT

The proposed solution for telecom churn prediction involves the application of machine learning algorithms to predict which customers are most likely to churn. The first step is to collect and pre-process a dataset that contains customer information, such as demographic data, service usage, billing history, and customer complaints. Feature engineering techniques are applied to extract relevant features from this data, such as customer tenure, usage patterns, and customer service metrics.

Several machine learning algorithms are then trained on the pre-processed dataset, including logistic regression, decision trees, random forests, and neural networks. Hyper-parameters are tuned to optimize performance, and the models are evaluated using metrics such as accuracy, precision, recall, and F1-score. To address issues such as class imbalance, techniques such as oversampling and under-sampling are also explored.

In our project, we use a real-world dataset from a telecommunications company to evaluate the performance of the proposed solution. The dataset contains information on over 7043 customers, including demographic data, service usage, billing history, and customer complaints. The target variable is a binary indicator of churn, indicating whether the customer has cancelled their service within the last month.

We pre-process the dataset by encoding categorical variables, scaling continuous variables, and handling missing values. We then apply feature engineering techniques to extract relevant features, such as customer tenure, usage patterns, and customer service metrics.

Our results show that the machine learning algorithms perform significantly better than a baseline model that predicts no churn, with the best model achieving an F1-score of 0.569. The log regression algorithm outperforms the other algorithms in terms of both accuracy and F1-score, and oversampling improves model performance for some algorithms. The results provide insights into the factors that drive customer churn in the telecommunications industry and demonstrate the potential of machine learning techniques for this application.

DISCUSSION AND OUTLOOK

The proposed solution and experiment for telecom churn prediction demonstrate the potential of machine learning algorithms for predicting customer churn in the telecommunications industry. However, there are several areas for discussion and future research.

One limitation of the proposed solution is the reliance on structured data, which may not capture all the relevant factors that drive customer churn. Unstructured data such as customer feedback and social media sentiment may provide additional insights into customer behaviour and preferences. Future research could explore the use of natural language processing and sentiment analysis techniques to incorporate unstructured data into the churn prediction model.

Another area for future research is the development of more interpretable models that can provide insights into the factors that drive customer churn. Many machine learning algorithms, such as neural networks, are inherently complex and difficult to interpret. Techniques such as decision tree visualization and feature importance analysis can provide some level of interpretability, but more research is needed in this area.

Finally, the proposed solution could be extended to incorporate real-time data and provide proactive interventions to prevent customer churn. For example, the model could be used to identify customers who are at high risk of churn and target them with personalized retention offers or customer service interventions. This could potentially improve customer satisfaction and reduce churn rates in the long term.

In summary, the proposed solution and experiment provide a promising approach to predicting customer churn in the telecommunications industry, but there are several areas for future research and development to improve the accuracy and interpretability of the model, as well as its potential impact on reducing customer churn.

TECHNIQUE	ACCURACY	AUC(ROC)	AUC(PR)	F1-SCORE
KNN	75.1%	0.764	0.533	0.521
LOG REGRESSION	79.1%	0.831	0.630	0.569
RANDOM FOREST	78.3%	0.808	0.602	0.539
SVM	78.6%	0.776	0.588	0.554

Table : Performance of Machine Learning Techniques on Test Set

CONCLUSION

In this project, we proposed a solution for telecom churn prediction using machine learning algorithms. We collected and pre-processed a real-world dataset of customer information, applied feature engineering techniques, and trained several machine learning algorithms to predict customer churn.

Our experimental results demonstrated that the proposed solution significantly outperformed a baseline model that predicts no churn, with the best model achieving an F1-score of 0.569. The log regression algorithm showed the best performance among the tested algorithms, and oversampling improved the performance of some algorithms. These results indicate the potential of machine learning techniques for predicting customer churn in the telecommunications industry.

Overall, this study contributes to the growing body of research on customer churn prediction, and provides insights into the factors that drive customer churn in the telecommunications industry. The proposed solution and experimental results offer practical applications for telecommunications companies seeking to improve customer retention and reduce churn rates.

Here are some more examples of how 80% accuracy in a logistic regression model for telecom churn prediction can help telecom companies improve customer loyalty, reduce customer acquisition costs, and increase revenue:

1. Improving customer loyalty: With an accurate churn prediction model, telecom companies can proactively identify customers who are at risk of churning and offer them targeted retention campaigns. For example, if a customer has a high probability of churning due to poor network coverage in their area, the company can offer them a discount on their plan or upgrade their network coverage in that area to retain the customer.
2. Reducing customer acquisition costs: Acquiring new customers can be expensive, so it's important for telecom companies to focus on retaining their existing customers. By accurately predicting which customers are at risk of churning, companies can allocate their resources more efficiently to focus on retaining those customers rather than spending money on acquiring new ones.
3. Increasing revenue: Churn prediction models can also help telecom companies increase revenue by identifying customers who are likely to be receptive to targeted upsell or cross-sell offers. For example, if a customer has a high probability of churning because they are reaching the end of their contract, the company could offer them an upgrade to a higher-tier plan or a new device at a discounted rate to encourage them to stay.
4. Improving customer satisfaction: By proactively identifying and addressing customer issues, such as poor network coverage or billing discrepancies, telecom companies can improve customer satisfaction and loyalty. For example, if a customer has a history of billing issues and is at risk of churning, the company could offer them a credit or discount to resolve the issue and retain their business.

Overall, an accurate churn prediction model can help telecom companies improve customer loyalty, reduce customer acquisition costs, and increase revenue by enabling them to identify and address customer issues proactively.

In conclusion, the proposed solution and experiment demonstrate the potential of machine learning algorithms for telecom churn prediction and highlight the importance of data-driven approaches in the telecommunications industry.

REFERNECES

[1]<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

[2]https://en.wikipedia.org/wiki/Customer_attrition






KEY CITATIONS






[1]Hung, Shin-Yuan, David C. Yen, and Hsiu-Yu Wang. "Applying data mining to telecom churn management." *Expert Systems with Applications* 31.3 (2006): 515-524.

Dahiya, Kiran, and Surbhi Bhatia. "Customer churn analysis in telecom industry." *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*. IEEE, 2015.

[2] Ahmad, Abdelrahim Kasem, Assef Jafar, and Kadan Aljoumaa. "Customer churn prediction in telecom using machine learning in big data platform." *Journal of Big Data* 6.1 (2019): 1-24.

THE MACHINE LEARNING CANVAS

<p>PREDICTION TASK </p> <p>Type of task: The telecom churn prediction task is a binary classification task, where the goal is to predict whether a customer will churn or not churn.</p> <p>Entity on which predictions are made: The entity on which predictions are made is the individual customer of the telecom company.</p> <p>Possible outcomes: The possible outcomes are either churn (1) or not churn (0), which represents a binary classification problem.</p>	<p>DECISIONS </p> <p>To turn predictions into proposed value for the end-user in the context of telecom churn prediction, there are several parameters that can be considered:</p> <ol style="list-style-type: none"> 1.Customer Segmentation 2.Personalized Offers 3.Customer Service 4.Customer Acquisition 	<p>VALUE PROPOSITION </p> <p>The end-users of a telecom churn prediction system are typically the customer retention or marketing teams within the telecom company. The objective of these teams is to reduce customer churn and improve customer retention rates by identifying and addressing churn risks among their customer base.</p>	<p>DATA COLLECTION </p> <p>For telecom churn prediction, a strategy for creating an initial train set and continuous updating of the predictive model could be as follows:</p> <ol style="list-style-type: none"> 1. Initial train set. 2. Continuous update. 	<p>DATA SOURCES </p> <p>The raw information for telecom churn prediction can be obtained from various sources such as customer billing and payment history, customer call logs, network usage logs, customer demographics, and customer complaints, among others.</p>
--	---	--	---	---

<p>IMPACT SIMULATION </p> <p>Yes, machine learning models can be deployed for telecom churn prediction. After training and validating the model, it can be deployed in a production environment to make predictions on new data in real-time.</p>	<p>MAKING PREDICTIONS </p> <p>The timing for making real-time or batch predictions for telecom churn prediction depends on the specific use case and the available data.</p>		<p>BUILDING MODELS </p> <p>The number of production models needed for telecom churn prediction may vary depending on the specific business scenario and the complexity of the machine learning solution.</p>	<p>FEATURES </p> <p>The input representations available at prediction time for telecom churn prediction may include various features related to the telecom customer's usage and behaviour.</p>
	<p>MONITORING </p> <p>There are several metrics that can be used to quantify the value creation and measure the ML system's impact in production for telecom churn prediction such as Accuracy , Precision , Recall , F1-score , ROC AUC.</p>			