CUSTOMER CHURN PREDICTION

MANISH KUMAR

0. ABSTRACT

This project aimed to develop a model for classifying customer churn using four different machine learning algorithms: random forest, SVM, logistic regression, and KNN. However, the initial accuracy of these models was not satisfactory due to the imbalanced dataset. Therefore, a resampling technique called SMOTE-ENN was used to address this issue. The results showed that the model's accuracy improved significantly after applying this technique. The SMOTE-ENN resampling technique effectively balanced the dataset and reduced the impact of the class imbalance, resulting in better performance. The proposed model can be used by companies to predict customer churn and take proactive measures to retain their customers. This project highlights the importance of using appropriate techniques to address imbalanced data and their impact on model performance.

1. INTRODUCTION

The customer who ceases a product or service for a given period is referred to as a churner. In a telecommunication company, the individual who has opted for service from a firm is referred to as Churn. The individual who probably intends to depart from the firm in near future was predicted by the churn model. Many industries build a model like churn as a common application for data mining techniques. Mobile telephone organizations present across the globe are almost on the verge of building their own churn model. Furthermore, to retain customers, churn results can be efficiently utilized for various other goals. The Churn Management approach is actually the first step in building a model. In general, the project needs a churn model in the best way instead of taking a single method with the best lift. So here we have built a model for the long run. In this digital era, the client of one company may also be a consumer of one or more telecommunication firms. Some of us may use different carriers based on the distance and some others may use different carriers based on the different plans they offer. While performing the analysis using machine learning customer experience tends to provide valuable insights. Some people will change their service providers from time to time. An increase or decrease in the calling rate will also depend on different job responsibilities. Based on the availability of the data various situations may reflect.

2. PROBLEM STATEMENT

The telecom service churn problem refers to predicting which customers are likely to switch to a competitor's service or terminate their subscription with a telecom service provider. Churn is a common problem for telecom service providers, as losing customers can have a significant impact on their revenue and profitability. Telecom service providers can use historical customer data such as demographics, usage patterns, call and text logs, and customer service interactions to address this problem to build predictive models. These models can then be used to identify customers who are at a high risk of churning and take proactive measures to retain them, such as offering personalized incentives or improving customer service. The telecom service churn problem is a classification problem, where the goal is to classify customers as either churned or retained. The predictive model's performance can be evaluated using metrics such as accuracy, precision, recall, and F1 score. Feature engineering, hyperparameter tuning, and ensemble learning are some techniques that can be used to improve the performance of the model. Solving the telecom service churn problem can help telecom service providers improve customer retention, reduce customer acquisition costs, and ultimately increase revenue and profitability.

3. PRODUCT IDEA

Churn Predict is a cloud-based software product that leverages machine learning algorithms to help businesses predict customer churn and take proactive measures to retain valuable customers. It provides actionable insights and recommendations to reduce churn rates and increase customer retention.

Key Features:

1. Data Integration: Churn Predict seamlessly integrates with the business's existing customer data sources, such as CRM systems, transaction databases, customer support logs, and usage metrics, to gather the necessary information for churn prediction.

2. Machine Learning Models: The product employs advanced machine learning techniques to build predictive models that analyze historical customer data, including demographics, transaction history, customer

interactions, and behavior patterns, to identify churn patterns and predictors.

3. Churn Risk Assessment: Churn Predict assigns a churn risk score to each customer, indicating the likelihood of churn within a specified time frame. This score is calculated based on the customer's individual attributes and historical patterns observed in similar customers.

4. Real-time Monitoring: The software continuously monitors customer activities and behavior to update churn risk scores in real-time. This enables businesses to identify immediate churn risks and take timely actions to prevent customer attrition.

5. Actionable Insights: Churn Predict provides businesses with actionable insights and recommendations tailored to each customer. These insights may include personalized retention strategies, such as offering loyalty incentives, tailored communication campaigns, or targeted promotions to retain at-risk customers.

6. Dashboard and Reporting: The product offers an intuitive dashboard that visualizes churn risk trends, customer segments, and the effectiveness of implemented retention strategies. Detailed reports provide in-depth analysis and performance metrics to help businesses evaluate the success of their churn reduction efforts.

7. Integration with Business Systems: Churn Predict integrates seamlessly with existing business systems, allowing automated triggers and workflows based on churn risk scores. For example, businesses can automatically trigger retention campaigns or notify account managers to reach out to high-risk customers.

8. Customizability and Scalability: Churn Predict is designed to accommodate different business needs and can be customized to specific industries or customer segments. It is scalable to handle large volumes of customer data and can adapt to evolving business requirements.

9. Data Security and Compliance: Churn Predict prioritizes data security and complies with industry-standard security protocols to protect customer data. It ensures compliance with relevant data protection regulations

and provides options for customers to manage their data privacy preferences.

10. Ongoing Support and Updates: The product includes ongoing support, maintenance, and regular updates to improve the accuracy and performance of the churn prediction models. It also incorporates feedback from businesses to enhance its features and functionalities.

By leveraging Churn Predict, businesses can proactively address customer churn, optimize customer retention efforts, and foster long-term customer relationships, ultimately leading to increased profitability and business growth.

4. BENEFITS AND IMPACT

Customer churn prediction has several benefits and can have a significant impact on businesses. Here are some of the key benefits and impacts of customer churn prediction:

1.Retaining Customers: Customer churn prediction helps businesses identify customers who are likely to churn or discontinue using their products or services. By identifying these customers in advance, businesses can take proactive measures to retain them. This may involve implementing targeted retention strategies, such as personalized offers, loyalty programs, or improved customer support, to reduce churn rates.

2.Cost Savings: Acquiring new customers can be more expensive than retaining existing ones. By accurately predicting customer churn, businesses can focus their resources on retaining valuable customers rather than allocating them to acquiring new customers. This can lead to significant cost savings in terms of marketing, advertising, and customer acquisition expenses.

3.Improved Customer Experience: Churn prediction models allow businesses to gain insights into customer behavior and preferences. By analyzing customer data, businesses can understand the factors that contribute to churn and take proactive steps to address those issues. This can result in enhanced customer experience, satisfaction, and loyalty.

4.Revenue Optimization: Churn prediction helps businesses identify segments or groups of customers with a high churn probability. By understanding the characteristics and behaviors of these customers, businesses can develop

targeted marketing campaigns or product offerings to mitigate churn and increase customer lifetime value. This can ultimately lead to revenue optimization and improved financial performance.

5.Competitive Advantage: Businesses that effectively utilize churn prediction gain a competitive advantage in the market. By successfully retaining customers who would otherwise churn, they can establish a loyal customer base and differentiate themselves from competitors. This can lead to increased market share and sustainable business growth.

6.Enhanced Decision-Making: Churn prediction models generate valuable insights that can inform strategic decision-making processes. Businesses can use these insights to identify patterns and trends, make data-driven decisions, and develop effective retention strategies. This can lead to improved business operations, targeted marketing efforts, and overall better decision-making across various departments.

7.Customer-Centric Approach: By leveraging churn prediction, businesses can adopt a customer-centric approach. They can proactively address customer needs and concerns, personalize communication and offerings, and provide better customer service. This customer-centric focus can result in increased customer satisfaction, loyalty, and advocacy.

In summary, customer churn prediction offers numerous benefits to businesses, including customer retention, cost savings, improved customer experience, revenue optimization, competitive advantage, enhanced decision-making, and a customer-centric approach. By effectively predicting and managing churn, businesses can foster long-term relationships with their customers and drive sustainable growth.

5. TARGET SPECIFICATION AND SPECIALIZATION

Target specification and specialization in customer churn prediction involve defining the specific objectives, target audience, and focus areas for the churn prediction model. Here are some aspects to consider:

1. Objectives:
   - Clearly define the objectives of the churn prediction model. Is the goal to identify customers at risk of churn, understand the factors contributing to churn, or develop personalized retention strategies?

- Determine the specific metrics or key performance indicators (KPIs) that will be used to evaluate the success of the churn prediction model, such as churn rate reduction, customer retention rate improvement, or revenue impact.

2. Target Audience:
- Identify the target audience or customer segment for the churn prediction model. This could be specific to industry, business type, or customer characteristics.
- Consider whether the churn prediction model will focus on individual customers or aggregate insights at a higher level, such as customer segments or product categories.

3. Data and Variables:
- Determine the data sources and variables that will be used for churn prediction. This may include customer demographic information, transactional data, customer behavior data, customer service interactions, or other relevant data sources.
- Define the key variables that are expected to have an impact on churn prediction, such as customer tenure, purchase frequency, customer complaints, or product engagement metrics.

4. Model Selection and Techniques:
- Select the appropriate machine learning techniques or models for churn prediction based on the available data and objectives.
- Common techniques include logistic regression, decision trees, random forests, support vector machines, or more advanced models like neural networks or gradient boosting algorithms.
- Consider whether the model will be built from scratch or if pre-built libraries or frameworks can be utilized.

5. Feature Engineering:
- Determine the specific features or variables that will be used as inputs to the churn prediction model. This may involve feature engineering techniques such as creating new variables, aggregating data, or transforming existing variables to improve predictive performance.

6. Retention Strategies:
- Based on the churn prediction model's insights, define the specific retention strategies or actions that will be implemented to mitigate churn. This could

involve personalized offers, targeted marketing campaigns, improved customer service, or product enhancements.

7. Performance Monitoring and Iteration:
   - Establish a process for monitoring and evaluating the performance of the churn prediction model over time.
   - Define how frequently the model will be updated or retrained to adapt to changing customer behavior and market dynamics.

Specializing in customer churn prediction involves understanding the unique challenges and dynamics of the specific industry or business domain. This includes considering industry-specific factors that impact churn, such as seasonality, product lifecycle, competitive landscape, or customer preferences. By tailoring the churn prediction model to the specific needs of the industry or business, it can provide more accurate and actionable insights.

In summary, target specification and specialization in customer churn prediction involve defining clear objectives, identifying the target audience, selecting appropriate models and techniques, engineering relevant features, and implementing targeted retention strategies. This specialization helps in creating a churn prediction model that is aligned with the specific requirements and challenges of the industry or business.

6.MARKET/CUSTOMER/BUSINESS NEED ASSESSMENT

Market/Customer/Business Needs Assessment for customer churn prediction involves several key points:

1. Market Analysis: Evaluate the competitive landscape, market trends, and demand for churn prediction solutions in your industry. Understand the potential market size and growth opportunities.

2. Customer Analysis: Analyze customer churn patterns, behaviors, preferences, and pain points. Assess the impact of churn on revenue and profitability. Identify the specific needs and expectations of your customers regarding retention and personalized experiences.

3. Business Goals and Objectives: Align churn prediction with business goals, such as revenue growth, customer satisfaction, or market positioning. Determine how reducing churn can contribute to these objectives.

4. Data Availability and Quality: Assess the availability and quality of customer data required for churn prediction. Consider the completeness, accuracy, and consistency of the data to ensure reliable analysis.

5. Implementation Feasibility: Evaluate the technical feasibility of implementing a churn prediction model. Assess existing infrastructure, technology stack, and resource capabilities. Consider scalability, maintainability, and potential partnerships or outsourcing needs.

By conducting a thorough assessment of the market, customer needs, and business goals, you can ensure that the churn prediction initiative is well-aligned, viable, and capable of delivering meaningful insights and impact for your business.

7.IMPLEMENTATION

The implementation of customer churn prediction involves several key steps:

1. Data Collection: Gather relevant customer data from various sources such as transactional records, customer interactions, demographics, and behavioral data. Ensure the data is comprehensive, accurate, and representative of your customer base.

2. Data Preprocessing: Cleanse and preprocess the data by removing duplicates, handling missing values, and performing necessary transformations. This step ensures the data is in a suitable format for analysis.

3. Feature Engineering: Extract and create meaningful features from the collected data that can help predict customer churn. This may involve calculating metrics such as customer lifetime value, recency, frequency, and monetary value of transactions.

4. Model Selection: Choose appropriate machine learning or statistical models for customer churn prediction. Popular models include logistic regression, decision trees, random forests, support vector machines (SVM), and gradient boosting algorithms like XGBoost or LightGBM.

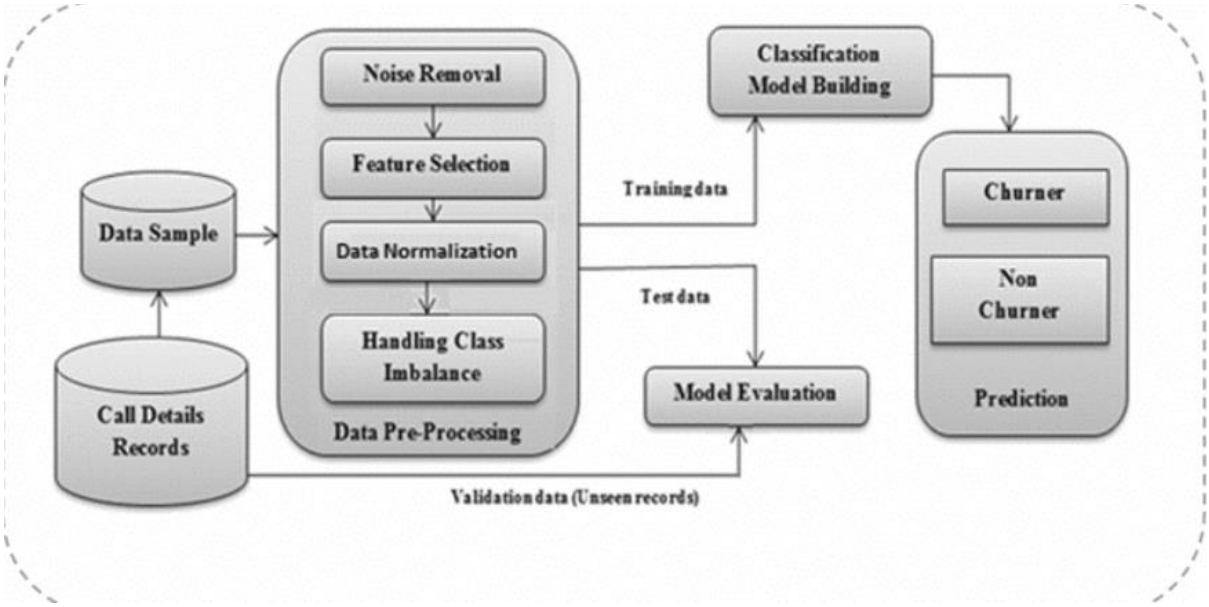
5. Model Training: Split the dataset into training and testing sets. Use the training set to train the selected model by fitting it to the data and adjusting its

parameters. Consider techniques such as cross-validation and hyperparameter tuning to optimize model performance.

6. Model Evaluation: Assess the performance of the trained model using evaluation metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC-ROC). Validate the model's effectiveness in predicting customer churn using the testing dataset.

7. Deployment and Monitoring: Integrate the churn prediction model into your operational systems or analytics platforms. Set up a process to regularly update the model with new data and monitor its performance over time. Continuously evaluate and refine the model as needed.

8.FINAL PRODUCT PROTOTYPE



9. CODE IMPLEMENTATION

The link to code Implementation
https://drive.google.com/file/d/1XMZqDN8KNoHf7TWDIWEav6TrRNeQdpnC/view?usp=sharing

10. CONCLUSION

In conclusion, this project aimed to classify customer churn using four different machine learning algorithms: random forest, SVM, logistic regression, and KNN. However, due to the imbalanced dataset, the accuracy of the models did not meet our expectations. To address this issue, we employed a resampling technique called SMOTE-ENN, which resulted in a significant improvement in the performance of our model. As a result of this technique, our model was able to provide accurate classifications on the test data. Thus, it can be concluded that SMOTE-ENN resampling is an effective approach to improve the accuracy of models when dealing with imbalanced datasets. This project highlights the importance of using appropriate techniques to address imbalanced data to obtain better results.

Model	Outputs	Precision	Recall	F1-Score
Logistic Regression	0	0.91	0.87	0.89
	1	0.90	0.93	0.91
KNN	0	0.98	0.98	0.98
	1	0.99	0.99	0.99
Random Forest	0	0.97	0.96	0.96
	1	0.97	0.98	0.97
SVM	0	0.95	0.94	0.94
	1	0.95	0.96	0.96

11. REFERENCES

1. Kim H S, Yoon C H. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. Telecommunications Policy,2004, 28(9): 751–765.
2. Rosset S, Neumann E. Integrating customer value considerations into predictive modelling. Third IEEE International Conference on Data Mining,2003: 1–8.
3. model for managers in mobile telecommunication. Networks and Services Research Conference, Proceedings of the 3rd Annual16-18 May 2005:48–53.