Customer Churn Prediction

Introduction

Artificial Intelligence has spurred many sectors affecting human endeavours into profound capabilities and advancement, helping individuals and organisations to reevaluate data in all its cycles, from generation to use, in order to get insights that drive decision making (West and Allen). However, AI is not without its own ethical challenges arising from bias, privacy violations, safety issues, transparency loss, to inequality and security problems *(Fedorko et al.[4])*. These risks can be managed through governmental and business policies.

Machine Learning (ML), a subdivision of AI, helps software applications to improve their prediction capacity without a prior programming intention to do so. Anticipation of future and possible outputs using historical data is the basis of ML *(J. Faritha Banu et al.[3])*. One major area, AI and ML is making a profound impact is in commerce. Shaping customer attitudes in choice of certain brands and products is the focal point of this impact.

Retail businesses have enjoyed this grace in recent times *(Fedorko et al.[4])*. Hence, customers play a vital role in any business growth and profit, their belief and trust is pivotal for business continuity in any organisation. Customer satisfaction must be the interest and goal of every establishment *(Gilaninia et al.[1])*. To increase business performance, there is a need to focus more on customer satisfaction and other factors that may affect such satisfaction in relation to customers' retainability.

Customer churn rate is one of such factors that can tell us more about customer satisfaction, and brands and products perception. Predicting this rate will be of immense importance to many businesses, especially to their revenue and continuity. AI and ML tools will help us to achieve this. In the following sections, we shall give a detailed description of the problem as well as the method we are going to use while highlighting all the respective tasks involved from data sourcing, through model building to its deployment for use.

Problem Description

Customer churn prediction presents several key challenges. Firstly, data quality and availability are paramount. Without clean, accurate, and up-to-date data, the predictions are likely to be unreliable. Imbalanced data is another issue, as churned customers often represent a small fraction of the total customer base. This can lead to biased models, so it's essential to employ techniques like oversampling or undersampling to balance the dataset and select appropriate evaluation metrics. Feature selection and data preprocessing are critical to identify the most relevant features and prepare the data for modeling. Ensuring that the chosen machine learning algorithm is suitable for the specific problem, tuning its hyperparameters, and addressing issues like overfitting are crucial for model performance.

In addition to technical challenges, addressing the business side of customer churn prediction is equally vital. It's important to collaborate with domain experts to understand the unique aspects of the industry and market. Moreover, identifying the cost of false positives and false negatives is essential for setting the right prediction threshold, as it impacts the business's retention strategies. Furthermore, customer segmentation should be considered, as customers may have diverse reasons for churn. Implementing a system for continuous monitoring and model deployment ensures long-term effectiveness and the ability to adapt to evolving customer behaviour. Finally, adhering to data privacy regulations and ethical considerations is a must when handling customer data.

To sum up, successfully conducting customer churn prediction involves overcoming data-related challenges, optimising modeling techniques, and bridging the gap between technical and business aspects. It requires a comprehensive understanding of data quality, feature engineering, model selection, and dealing with imbalanced data. Additionally, a strong partnership between data scientists and domain experts is essential to align predictive models with the unique needs of the business, while addressing ethical and regulatory concerns to ensure customer data is handled responsibly.

Proposed Method

This project follows the machine learning lifecycle from data sourcing to model deployment. Each stage of the machine learning process is described below.

- **DATA SOURCING**: This is synonymous to data gathering. Here, datasets relating to customer churning are sought. These datasets could be gotten from *kaggle* or by *scraping* the web of prominent businesses and selectively seeking historical data of customer behaviour. Sometimes, the data sourced may not be sufficient and the model overfits. More data may have to be sourced to prevent this unwanted behaviour of the model. Thus, this process is iterative.
- **DATA CLEANING**: Also known as data wrangling. The sourced data is prepared in such a way that the machine learning algorithm can process it easily. Categorical features would be one hot encoded and noisy/unwanted data are discarded. The required tools for this phase are *numpy* and *pandas*. In cases where the dataset contains some null columns, dummy data would be inserted where necessary or such feature is discarded completely if it has no utility to the target column.
- **DATA VISUALIZATION**: Here, we will pull out actionable insights from our dataset. We can make informed decisions from the customer's behaviour. Pooling the data into visual charts and illustrations will help us see patterns in customer behaviour and we can draw conclusions. Customers suspected of abandoning our business would be given greater attention to prevent them from leaving. There is a multiplicity of tools that we can use in doing this but we would use *seaborn* and *matplotlib*, this is because it allows users to create complex and customised and interactive plots quickly. The observations from the dataset alongside the domain expertise will guide our model choices.
- **MODEL BUILDING**: We build machine learning models here through an iterative training process of the training data. The model learns customer behaviour through the input features and predicts customers who would consider leaving the business given their past feedback, complaints or actions. We save our model after testing it on a validation and cross validation data sets. We shall try more than one model and compare results before concluding on which one we shall consider for deployment, or perhaps build a Hybrid model for deployment.
- **MODEL DEPLOYMENT**: Here, the saved machine learning model is deployed on a server. We might spin up a Flask backend application and deploy the model on the server or use streamlit. The deployed model would be exposed through a set of API endpoints.

Proposed Split

1. **Data sourcing:** Assuming a real world scenario, we will gather and collect relevant customers data for churn prediction, hence the various tasks are as follows:
    1.1. Identifying customers data, such as CRM databases and billing records
    1.2. Customers information such as demographics, usage history and contact details
    1.3. Ensure compliance with data privacy regulation

    However, for the purpose of this project, we shall use publicly available datasets from Kaggle for example, to demonstrate our study, and the model build can always be adapted to a real world context. This stage will involve making the choice of at least two datasets from the available dataset in different domains, say finance and telecommunications for example.
2. **Data cleaning:** Clean and preprocess the collected data to prepare it for analysis, hence we have the following tasks:
    2.1. Handling missing values in the customers dataset
    2.2. Encode categorical variables (e.g., contract type) for machine learning.
    2.3. Standardise data formats and units (e.g., currency conversion).
    2.4. Engineer relevant features like customer tenure, usage patterns, and billing history.
3. **Model:** Building an AI/ML model that predicts customer churn. We shall achieve this thanks to the following tasks:
    3.1. Selecting appropriate algorithms for churn prediction (eg. logistic regression, random forests, or neural networks), inspired from the literature.
    3.2. Splitting the Dataset into training and testing sets
    3.3. Training and the model using appropriate techniques to improve model training like cross-validation for example.
    3.4. Evaluating model performance using metrics such as accuracy, precision, recall, and ROC AUC.
4. **Model Deployment (Frontend and Backend Skills Required):** This stage involves deploying the churn prediction model for use in customer management. After identification of the appropriate available tools to do this, either using Flask, Django or Streamlit, we shall have the following tasks:
    4.1. Frontend:
        4.1.1. Create a web-based dashboard for customer service representatives to input customer information and receive churn predictions.

4.1.2. Design visualisations and reports to present model results, after we have identified the relevant information to bring forward to be useful to domain experts in order to make informed business decisions.

4.2. Backend:

4.2.1. Develop an API using a framework like Flask or Django or Streamlit to serve the churn prediction model.

4.3. Integrate the model into the customer relationship management (CRM) system.

4.4. Monitor model performance in real-time and provide alerts for high-risk customers.

5. **Business Strategy:** Develop strategies and action plans based on churn predictions. Hence this will involve:
    5.1. Analyse the model's predictions to identify factors contributing to churn.
    5.2. Define retention strategies, such as targeted marketing campaigns or personalised offers.
    5.3. Collaborate with marketing and sales teams to implement customer retention initiatives.

Conclusion

Customer churn poses significant challenges for businesses across various sectors. To sustain growth, companies not only need to acquire new clients but also retain existing ones. Each departing client represents a substantial loss in terms of investment, necessitating additional resources for replacement. Predicting when a client might leave and providing incentives for them to stay can result in significant savings. This potential impact inspired our study, anticipating that our findings could benefit a broad range of industries.

Our primary objective is to build a comprehensive end-to-end machine learning project, a goal aligned with this training session. We believe that selecting a focus area based on the team's expertise will enhance our model analysis through domain-specific insights. Therefore, we've chosen to work with two datasets: one from the financial sector [10] and the other from e-commerce [6,8] . Both datasets are publicly available on *Kaggle*.

In summary, predicting customer churn is important. Effective action can be taken to retain the customer before it is too late. Thus, the ability to predict that a customer is at risk of churn while there is still time to do something about it represents a huge additional potential revenue stream for any online business.

References

1. Gilaninia, Shahram, et al. "Marketing Mix and Consumer Behavior." Kuwait Chapter of Arabian Journal of Business and Management Review, vol. 2, no. 12, Aug. 2013, pp. 53–58.
2. West, Darrell M, and John R Allen. "How Artificial Intelligence Is Transforming the World." Brookings, 24 Apr. 2018,*Accessed 12 Oct. 2023.*
3. J. Faritha Banu, et al. ``Artificial Intelligence Based Customer Churn Prediction Model for Business Markets." Computational Intelligence and Neuroscience, vol. 2022, 29 Sept. 2022, pp. 1–14. *Accessed 12 Oct. 2023.*
4. Fedorko, Richard, et al. "Artificial Intelligence in E-Commerce: A Literature Review." ResearchGate, unknown, July 2022. *Accessed 12 Oct. 2023.*
5. Kaggle Customer Churn Prediction Datasets.
6. Fridrich, M. (2023). Machine Learning in Customer Churn Prediction [Dissertation Thesis]. Brno University of Technology, Faculty of Business and Management. Supervised by Petr Dostál.
7. How to leverage churn prediction data to build a resilient business.
8. Ecommerce Customer Churn Analysis and Prediction
9. Zindi: Expresso Churn Prediction Challenge dataset also available on Kaggle.
10. Kaggle: Credit Card Churn Prediction

Mentor

- Foutse Yuehgoh

Members

- Jack Oraro
- Lucky Nkwocha
- Ayooluwa Jesuniyi
- Oyelayo Seye
- Adenike Ayodeji
- Elijah Mesagan