Research Proposal Transcript

Introduction

My name is Munro Ross, I am currently studying Data Science at the University of Essex Online, and this is my Research Proposal Presentation for Research Methods and Professional Practice.

This proposal will outline a research plan on the impact of class imbalance techniques on a model's interpretability in churn prediction.

Background and Rationale

Class imbalance occurs in multiple real-world situations such as fraud detection, medical diagnosis and customer churn prediction. It refers to skewed distribution of classes in a dataset, with one class having a significantly higher number of occurrences than others. This normally causes the minority class, which is generally the class of interest, to be underrepresented in a machine learning models training and cause poor generalisation to this class (Awe, 2020).

Various methods can be utilised to combat the difficulties encountered from imbalanced datasets, however many high performing models that are less affected by data imbalance, such as Neural Networks and Ensemble models are not as transparent with decision making (Caigny, W. De Bock and Verboven, 2024).

Most existing literature focuses on how different class imbalance techniques affect model performance when applied to customer churn classification, there is not much research into how these techniques affect interpretability and explainability of models. The research in this proposal will therefore compare how different techniques used for class imbalance affect interpretability of models such as Logistic Regression and Decision Trees as well as compare the resulting performance of the models.

Aims and Research Questions

The aim of this research is to evaluate the impact of class imbalance techniques on the interpretability and explainability of models.

The research questions that will be explored are, how do class imbalance techniques affect the interpretability and explainability of churn prediction models, what are the trade-offs between performance and interpretability for each technique.

Objectives

The following are the objectives of the proposed research.

- Compare machine learning models trained using various imbalance-handling techniques. This will explore techniques such as resampling methods, synthetic data generation and cost-sensitive learning.
- Measure and analyse the interpretability and explainability of these models using explainability techniques such as SHAP and LIME for both white box and black box models as well as inherent interpretability techniques for the white box models.

- Compare model performance vs interpretability.
- Outline practical recommendations for real-world applications.

All this helping to understand the trade-off between model performance and trust for business stakeholders.

Review of Literature

Outlined in the next few slides are important points of key literature found during my recent literature review assignment for the Research Methods and Professional Practice module.

Customer churn refers to the number of customers lost over a period, acquiring new customers can cost a company up to 25 times more than customer retention (Gallo, 2014; O'Brien & Downie, 2024). This highlights the importance of customer retention and for companies to be able to predict churn.

Machine Learning is used for churn prediction due to its ability to learn patterns autonomously (Pecan, 2024). Models built using algorithms such as XGBoost, Random Forest and Decision Trees normally perform better when implemented in customer churn applications and where class imbalance is present (Pulkundwar et al., 2023).

Churn datasets are normally imbalanced, with the class of interest (churners) being the minority class. Imbalanced data leads to poor predictive performance for some models (Brownlee, 2020). This is normally due to bias towards the majority class in the dataset. Strategies to overcome this include oversampling, under sampling and cost-sensitive learning (Burkov, 2019).

Understanding why models make the predictions they do helps businesses make data driven decisions (Ribeiro et al., 2016). Especially in customer churn, the company want to identify customers that are likely to churn and the reasons, therefore it is good to understand why the model is making the decisions it is.

White box models are inherently interpretable with some examples being Linear models, Decision Trees and Generalized Additive models. Black Box models are more complicated and harder to interpret, these include Tree ensemble methods, Support Vector Machines and Neural Networks (Dwivedi et al., 2022). SHAP and LIME are popular Explainable Artificial Intelligence (XAI) tools to explain predictions (Özkurt, 2025). These techniques can be used on both white box and black box models.

To summarise the literature Boosting and Balancing techniques help with model performance when class imbalance is present (Ahmad et al., 2019). XAI can be used to explain decisions of both white box and black box models. But how balancing methods affect model interpretability or explainability is not well studied.

Methodology and Research design

This project does not involve primary data or collection methods, instead a publicly available secondary dataset will be used. The BlastChar (2017) Telco customer churn dataset from Kaggle will be used and a more recent version can be found via IBM

(2019). The dataset contains over 7000 observations and has 21 features, one being the target feature.

Using secondary data allows the focus to stay on the research aim of comparing class imbalance techniques on model interpretability. It eliminates ethical concerns that could arise from primary data collection of human participants and is well suited to my task.

The algorithms that will be used for this research are Logistic regression and Decision trees for white box models, XGBoost and Random Forest for black box models. While logistic regression performance may not be great for this use case, it should indicate whether a change in interpretability occurs when using class imbalance techniques. The inclusion of black box models will help compare SHAP and LIME techniques for explainability between models of different levels of interpretability.

Balancing techniques that will be utilised and compared for this study are model baselines with no balancing, Random oversampling and Random under sampling, SMOTE (Synthetic Minority Oversampling Technique), ADASYN (Adaptive Synthetic Sampling) and cost-sensitive learning.

The evaluation metrics used to evaluate performance of each model before and after sampling techniques will be F1-score, Area under the Curve (AUC), precision and recall.

SHAP (Shapley Additive explanation) and LIME (Local Interpretable Model-agnostic Explanations) will then be used to evaluate explainability of each model trained using the different balancing techniques. SHAP being used to explain feature contribution to the predictions at a global level, while LIME will be used to provide local explanations of the individual predictions (Keita, 2023).

Stability and Fidelity will then be evaluated for both LIME and SHAP techniques used.

Ethical Considerations and Artefacts

While this project will not involve primary data collection, the topic revolves around ethical AI and focuses on techniques to improve model transparency striving towards more AI trustworthy systems. The dataset used will be public and anonymised meaning there ais no risk to individuals' personal data. All experiments conducted and the methodology will be documented to ensure reproducibility.

Artifacts created during this research will include Jupyter Notebooks documenting the full experimental procedure. Visualisation of performance and interpretability comparisons and a final report outline methodology and results achieved.

Timeline

The timeline for each stage of this research can be seen outlined on the slide and will span 8 weeks. The first week will be utilised to finalise the research plan, the second will be used to inspect and preprocess the dataset for ML applications, the baseline models will be trained in the third week. In the fourth week class balancing techniques will be applied and Interpretability tests done in the fifth week. The final three weeks will compare trade-off between performance and interpretability of each method applied and all results will be compiled.

Conclusion

To conclude, this project aims to explore the effect of class balancing techniques on model interpretability and explainability, a space with not so much research. By applying XAI techniques such as SHAP and LIME to real world churn data with imbalanced classes, the goal is to better understand the effect of preprocessing steps on model trustworthiness.

References used to compile the research proposal can be found on the last two slides, as well as in the transcript.

References

Ahmad, A.K., Jafar, A. and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, [online] 6(1). doi:https://doi.org/10.1186/s40537-019-0191-6.

De Caigny, A. W. De Bock, K. and Verboven, S. (2024). Hybrid black-box classification for customer churn prediction with segmented interpretability analysis. *Decision Support Systems*, 181, pp.114217–114217. doi:https://doi.org/10.1016/j.dss.2024.114217.

Awe, O. (2020). *Computational Strategies for Handling Imbalanced Data in Machine Learning VP-IASE VP of Global Engagement, -LISA 2020 Global Network, USA*. [online] Available at: https://isi-web.org/sites/default/files/2024-02/Handling-Data-Imbalance-in-Machine-Learning.pdf.

BlastChar (2017). *Telco Customer Churn*. [online] www.kaggle.com. Available at: https://www.kaggle.com/datasets/blastchar/telco-customer-churn.

Brownlee, J. (2019). *A Gentle Introduction to Imbalanced Classification*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/what-is-imbalanced-classification/.

Burkov, A. (2019). *THE HUNDRED-PAGE MACHINE LEARNING BOOK*. Andriy Burkov.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Rana, O., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G. and Ranjan, R. (2022). Explainable AI (XAI): Core Ideas, Techniques and Solutions. *ACM Computing Surveys*, 55(9), pp.1–33. doi:https://doi.org/10.1145/3561048.

Gallo, A. (2014). *The Value of Keeping the Right Customers*. [online] Harvard Business Review. Available at: https://hbr.org/2014/10/the-value-of-keeping-the-right-customers.

IBM (2019). *Telco customer churn (11.1.3+)*. [online] Ibm.com. Available at: https://community.ibm.com/community/user/blogs/steven-macko/2019/07/11/telco-customer-churn-1113.

Keita, Z. (2023). *Explainable AI - Understanding and Trusting Machine Learning Models*. [online] Datacamp.com. Available at: https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models.

Özkurt, C. (2025). Transparency in Decision-Making: The Role of Explainable AI (XAI) in Customer Churn Analysis. *Information Technology in Economics and Business*. doi:https://doi.org/10.69882/adba.iteb.2025011.

Pecan (2024). *We Used AI to Predict Customer Churn — Now What? Churn Reduction Initiatives*. [online] Pecan AI. Available at: https://www.pecan.ai/blog/churn-reduction-strategies-prediction-playbook/.

Pulkundwar, P., Rudani, K., Rane, O., Shah, C. and Virnodkar, S. (2023). A Comparison of Machine Learning Algorithms for Customer Churn Prediction. doi:https://doi.org/10.1109/icast59062.2023.10455051.

Ribeiro, M.T., Singh, S. and Guestrin, C. (2016). ‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16*, [online] pp.1135–1144. doi:https://doi.org/10.1145/2939672.2939778.