MINISTRY OF EDUCATION AND RESEARCH OF THE REPUBLIC OF MOLDOVA

Technical University of Moldova

Faculty of Computers, Informatics, and Microelectronics

Department of Software Engineering and Automation

Approved for defense
Department head:
Ion FIODOROV, phd, associate professor

„___” _____ 2025

PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS FOR AUTOMATED TICKET ROUTING SYSTEMS (DISTILBERT MULTILINGUAL CASED, INFOXLM, DISTILBERT BASE UNCASED)

Master's project

Student: _____ **Derevenco Serghei, IS-231M**

Coordinator: _____ **Besliu Corina, university lecturer**

Consultant: _____ **Cojocaru Svetlana, university assistant**

Chisinau, 2025

ABSTRACT

Automated ticket routing systems are vital for efficiently managing customer and IT support inquiries. This thesis evaluates the performance of three prominent Natural Language Processing (NLP) models: DistilBERT-base-uncased, DistilBERT-multilingual-cased and InfoXLM, tailored for automated ticket classification and routing. The study integrates a robust methodology comprising data cleaning, exploratory analysis and augmentation to prepare a real-world dataset. Advanced classification technique of fine-tuning was employed to optimize model performance across diverse ticket categories. The evaluation reveals that fine-tuning pre-trained models on domain-specific ticket data significantly enhances classification accuracy and routing precision. Additionally, performance optimization strategies, including stop-word removal, n-gram tokenization and batch size adjustments, were explored to further refine model accuracy and computational efficiency. By integrating domain-specific fine-tuning and scalable methodologies, this research underscores the transformative potential of machine learning in automated ticket routing systems. The findings provide actionable insights for implementing efficient, adaptable, and accurate ticket management solutions in dynamic support environments, ultimately improving service delivery and operational scalability.

REZUMAT

Sistemele automatizate de rutare a tichetelor sunt esențiale pentru gestionarea eficientă a cererilor de suport pentru clienți din diferite domenii, precum și cei din domeniul IT. Această teză evaluează performanța a trei modele importante de Procesare a Limbajului Natural (NLP): DistilBERT-base-uncased, DistilBERT-multilingual-cased și InfoXLM, adaptate pentru clasificarea și rutarea automată a tichetelor. Studiul integrează o metodologie eficientă, care include curățarea datelor, analiza exploratorie și augmentarea, pentru a pregăti un set de date eficient, extras dintr-o aplicație reală. Tehnică avansată de clasificare, fine-tuning, a fost aplicată pentru optimizarea performanței modelelor pe categorii diverse de tichete. Evaluarea expune faptul că ajustarea modelelor pre-antrenate pe date specifice domeniului îmbunătățește semnificativ acuratețea clasificării și precizia rutării. În plus, teza explorează strategii de optimizare a performanței, cum ar fi eliminarea stop-word-urilor, tokenizarea prin n-grame și ajustarea dimensiunii loturilor (batch size), îmbunătățind și mai mult acuratețea și eficiența modelelor. Prin integrarea metodei de fine-tuning specific domeniului și a celor mai noi metodologii, această cercetare subliniază impactul transformator al învățării automate asupra sistemelor de rutare a tichetelor. Rezultatele oferă perspective aplicabile pentru implementarea unor soluții eficiente, adaptabile și precise de gestionare a tichetelor, îmbunătățind în final livrarea serviciilor și scalabilitatea operațională în medii de suport dinamice.

CONTENTS

INTRODUCTION

In today's fast-paced corporate environment, customer service is critical to sustaining client happiness. As organizations expand and the volume of support tickets grows, efficient ticket management systems become increasingly important. Traditional manual ticket routing, which relies heavily on human judgment, can result in inefficiencies, longer response times, and probable mistakes in directing tickets to the appropriate support teams. To address these difficulties, this thesis investigates the use of Machine Learning (ML) models to improve automated ticket routing systems.

This research addresses the evaluation of several Natural Language Processing (NLP) models in automating ticket routing classification like: DistilBERT-base-uncased, DistilBERT-multilingual-cased, and InfoXLM. These models represent a spectrum of capabilities, from lightweight efficiency to multilingual and cross-lingual versatility. To assess their performance, the study utilizes a real-world dataset from a company. The dataset includes diverse ticket categories, metadata, and communication records, providing a rich foundation for exploring the complexities of ticket routing.

An essential part of the research involves data preparation, cleaning, and augmentation, which are critical for optimizing input data and ultimately, model performance. Data augmentation, in particular, is employed to overcome challenges such as class imbalance and limited data variety. Techniques of oversampling like random word insertion, augmentation with synonyms and antonyms, and undersampling like random articles deletion are used to diversify the dataset, helping to ensure that models are exposed to a wider range of linguistic patterns and ticket structures. This stage is crucial as it allows the models to better capture the nuances of ticket content, which can significantly influence the success of the classification and routing process.

The study then explores the application of ML methodologies, including zero-shot, few-shot learning, and fine-tuning, to gauge their effectiveness in the context of ticket routing. Zero-shot learning is evaluated for its ability to classify tickets without specific task-related training, a useful feature for rapidly adapting to new ticket categories without a labeled dataset. Few-shot learning is similarly tested for scenarios requiring domain-specific adjustments, leveraging a small amount of labeled data to refine performance. Fine-tuning, a more complicated strategy, is evaluated for its effectiveness in enhancing model accuracy by tailoring pre-trained models to the specific properties of the ticket data. By comparing various techniques, the study hopes to uncover their distinct strengths and potential trade-offs in terms of performance, resource consumption, and implementation complexity.

In addition to the methodologies, the study offers a comparative analysis of the findings from the various approaches. This analysis helps to identify which methods are most effective for different types of ticket routing scenarios, allowing for insights into optimizing machine learning techniques based on specific operational requirements. The evaluation process involves testing on a range of performance metrics, such as classification accuracy, precision, recall, and overall routing efficiency. These metrics

offer a comprehensive view of the impact of each approach on the quality and reliability of the ticket routing system, which is essential for ensuring customer satisfaction in real-world applications.

Based on the comparative analysis, the best-performing model was selected for further optimization. To enhance its performance, separate approaches were explored, including adjusting the batch size to improve training efficiency and substituting the default classification algorithm with random forests to evaluate alternative decision-making strategies. Additionally, the number of training epochs was determined by analyzing the train and test accuracies to identify the point where the model achieves optimal generalization. These refinements not only improved the model's accuracy but also enhanced its robustness and adaptability to diverse ticket categories. This focused optimization highlights the iterative nature of machine learning workflows, emphasizing the importance of fine-tuning and exploring complementary strategies to achieve optimal results.

Through this comprehensive study, the thesis aims to contribute to the growing field of ML-driven ticket routing solutions by providing detailed insights into best practices for model selection and implementation. By demonstrating the transformative potential of ML in ticket routing, this research highlights the role of advanced algorithms in streamlining customer support processes, leading to significant improvements in service delivery, scalability, and operational accuracy. This exploration not only underscores the value of ML in automated ticket routing but also serves as a guide for future developments in this rapidly evolving domain.

BIBLIOGRAPHY

1. “AI-powered ticketing automation: A complete guide for 2024,” Zendesk. Accessed: Sep. 2, 2024. [Online]. Available: https://www.zendesk.com/blog/ai-powered-ticketing/

2. M. J. J. Bucher and M. Martini, “Fine-Tuned ‘Small’ LLMs (Still) Significantly Outperform Zero-Shot Generative AI Models in Text Classification,” Aug. 16, 2024, *arXiv*: arXiv:2406.08660. Accessed: Sep. 9, 2024. [Online]. Available: http://arxiv.org/abs/2406.08660

3. H. Naveed *et al.*, “A Comprehensive Overview of Large Language Models,” Apr. 09, 2024, *arXiv*: arXiv:2307.06435. Accessed: Sep. 9, 2024. [Online]. Available: http://arxiv.org/abs/2307.06435

4. J. Howard and S. Ruder, “Universal Language Model Fine-tuning for Text Classification,” May 23, 2018, *arXiv*: arXiv:1801.06146. doi: 10.48550/arXiv.1801.06146.

5. T. Schick and H. Schütze, “Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference,” Jan. 25, 2021, *arXiv*: arXiv:2001.07676. doi: 10.48550/arXiv.2001.07676.

6. S. Garg, S. Mitra, T. Yu, Y. Gadhia, and A. Kashettiwar, “Reinforced Approximate Exploratory Data Analysis,” Dec. 12, 2022, *arXiv*: arXiv:2212.06225. Accessed: Sep. 27, 2024. [Online]. Available: http://arxiv.org/abs/2212.06225

7. M. Komorowski, D. C. Marshall, J. D. Salciccioli, Y. Crutain. “Exploratory Data Analysis”, 2016.

8. H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

9. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Jun. 09, 2011, *arXiv*: arXiv:1106.1813. Accessed: Oct. 5, 2024. [Online]. Available: http://arxiv.org/abs/1106.1813

10. M. Werner and E. Laber, “Speeding up Word Mover’s Distance and its variants via properties of distances between embeddings,” May 08, 2020, *arXiv*: arXiv:1912.00509. Accessed: Oct. 12, 2024. [Online]. Available: http://arxiv.org/abs/1912.00509

11. J. Wei and K. Zou, “EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks,” Aug. 25, 2019, *arXiv*: arXiv:1901.11196. Accessed: Oct. 20, 2024. [Online]. Available: http://arxiv.org/abs/1901.11196

12. Van Dyk, D. A., & Meng, X. L. (2001). “The Art of Data Augmentation.” *Journal of Computational and Graphical Statistics*, vol. 10, pp. 1-50.

13. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” Mar. 01, 2020, *arXiv*: arXiv:1910.01108. Accessed: Oct. 22, 2024. [Online]. Available: http://arxiv.org/abs/1910.01108

14. T. Wolf *et al.*, “HuggingFace’s Transformers: State-of-the-art Natural Language Processing,” Jul. 14, 2020, *arXiv*: arXiv:1910.03771. Accessed: Oct. 23, 2024. [Online]. Available: http://arxiv.org/abs/1910.03771

15. Z. Chi *et al.*, “InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training,” Apr. 07, 2021, *arXiv*: arXiv:2007.07834. Accessed: Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/2007.07834

16. G. Lample and A. Conneau, “Cross-lingual Language Model Pretraining,” Jan. 22, 2019, *arXiv*: arXiv:1901.07291. Accessed: Oct. 27, 2024. [Online]. Available: http://arxiv.org/abs/1901.07291

17. S. Sarica and J. Luo, “Stopwords in Technical Language Processing,” *PLoS ONE*, vol. 16, no. 8, p. e0254937, Aug. 2021, doi: 10.1371/journal.pone.0254937.

18. S. Bhardwaj and J. Pant, “Sentiment Analysis Approach based N-gram and KNN Classifier,” *Int. J. Res. Electronics Comput. Eng.*, vol. 7, no. 2, pp. 2108, Apr. - Jun. 2019.

19. N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” Feb. 09, 2017, *arXiv*: arXiv:1609.04836. doi: 10.48550/arXiv.1609.04836.

20. C. Chen, A. Liaw, and L. Breiman, “Using Random Forest to Learn Imbalanced Data,” *Department of Statistics, UC Berkeley*, 2001. [Online]. Available: https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf

21. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929-1958, Jun. 2014. [Online]. Available: https://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf