

LB/TH/43/2025
TH6010

**AUTOMATED USER REVIEW ANALYSIS TO
FACILITATE POTENTIAL MOBILE
APPLICATION EVOLUTION**

A.D.S.R.Gunathilaka

219337X

Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2024

**AUTOMATED USER REVIEW ANALYSIS TO
FACILITATE POTENTIAL MOBILE
APPLICATION EVOLUTION**

A.D.S.R.Gunathilaka

219337X

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

October 2024

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 02/27/2025

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Nisansa de Silva

Signature of the Supervisor:

Date: 27/02/2025

ACKNOWLEDGEMENT

I would like to express my profound gratitude to several individuals and institutions who have been instrumental in the successful completion of this thesis:

- First and foremost, I extend my heartfelt thanks to my supervisor, Dr. Nisansa de Silva, for his invaluable guidance, unwavering support, and insightful feedback throughout the course of this research. His expertise and encouragement have been pivotal in shaping this work.
- I am deeply grateful to the entire faculty and staff of the Department of Computer Science and Engineering for their dedication in delivering high-quality lectures, conducting rigorous examinations, and providing both major and minor support that has contributed significantly to the success of the MSc program.
- My sincere appreciation goes to all the staff members of the University of Moratuwa whose tireless efforts behind the scenes have been crucial in making this MSc program a resounding success.
- I would like to express my heartfelt thanks to my fellow MSc students for their camaraderie, intellectual discussions, and mutual support throughout this academic journey. Your friendship and collaboration have made this experience truly enriching.
- I extend my gratitude to the esteemed panel of examiners for their time, expertise, and constructive feedback in evaluating our research efforts. Your insights have undoubtedly enhanced the quality of this work.
- Last but not least, I wish to acknowledge the broader research and open-source communities worldwide. Your collective efforts in advancing scientific knowledge and providing accessible tools and resources have been instrumental in facilitating this research.

This journey would not have been possible without the support and contributions of all these individuals and groups. I am profoundly thankful for the opportunity to learn, grow, and contribute to the field through this research endeavor.

ABSTRACT

User reviews are crucial for mobile application evolution, but manually analyzing large volumes of feedback is time-consuming and inefficient. This thesis explores advanced techniques for automated user review analysis to facilitate potential mobile application evolution. We introduce a novel Convolutional Neural Network (CNN) based approach for Aspect-Based Sentiment Analysis (ABSA) on app reviews. Our model demonstrates significant improvements over existing baselines, with F1 scores of 0.62, 0.42, and 0.62 for aspect category classification in Productivity, Game, and Social Networking domains respectively. For aspect sentiment classification, we achieve accuracy scores of 0.80, 0.70, and 0.86 in the same domains. We provide empirical evidence on hyperparameter tuning, investigating the effects of batch size, number of epochs, and learning rate on model performance. To enhance our ABSA model, we investigate various word embeddings and data augmentation techniques. We find that Word2Vec embeddings and Round-trip Translation (RTT) augmentation yield the best results, offering insights for future research in this domain. Expanding our exploration of automated review analysis, we evaluate the potential of Large Language Models (LLMs). We provide a comprehensive comparison of state-of-the-art commercial and open-source LLMs in zero-shot and fine-tuned settings. Notably, we demonstrate the feasibility of using commercial LLMs as autonomous annotators, creating a high-quality dataset of 10,000 app reviews while achieving an accuracy of 81.89%. Our research also investigates the impact of various parameters on LLM performance for app review analysis, including training data size, number of epochs, Temperature, and Top_p. We find that fine-tuned open-source models can achieve performance comparable to commercial LLMs, with our best model reaching an F1 score of 0.83416. Our work contributes to the field of mobile application development by advancing automated user review analysis techniques to potentially improve the process of mobile application evolution. This work has the potential to lead to faster and more targeted mobile application improvements, enhancing developers' ability to respond to user needs effectively.

Keywords: Mobile App Review Analysis, Aspect-Based Sentiment Analysis, Convolutional Neural Networks, Large Language Models, Word Embeddings, Data Augmentation, Hyperparameter Tuning, GPT-3.5, LLAMA 2, Mistral, Fine-tuning, Software Evolution

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Acknowledgement	ii
Abstract	iii
Table of Contents	iv
1 INTRODUCTION	1
1.1 Background	1
1.2 Research Problem	2
1.3 Research Objectives	3
1.4 Contributions	3
2 Literature Survey	4
2.1 Importance of user feedback for apps development	4
2.2 Traditional Machine Learning Approaches	5
2.2.1 Topic Modeling-Based Approaches	5
2.2.2 Supervised Learning Approaches	5
2.2.3 Specialized Analysis Frameworks	6
2.2.4 Active Learning and Hybrid Approaches	6
2.2.5 Comparative Analysis	7
2.3 Deep Learning Approaches	9
2.3.1 CNN-Based Architectures for App Review Analysis	9
2.3.2 Transformer-Based Approaches and Transfer Learning	10
2.3.3 Comparative Analysis of Deep Learning Approaches	11
2.3.4 Key Insights and Implications	12
2.4 Pre-processing user reviews	13
2.5 Summery	14
2.6 LLM-based Annotation	17
2.6.1 Evolution of LLM Annotation Strategies	17
2.6.2 Advanced Integration Frameworks	17

2.6.3	Human-LLM Collaborative Approaches	18
2.6.4	Domain-Specific Applications and Evaluation	18
2.6.5	Limitations and Challenges	18
2.6.6	Comprehensive Perspectives and Future Directions	19
2.6.7	Comparative Analysis	19
3	PUBLICATIONS OF THE RESEARCH	21
3.1	Aspect-based Sentiment Analysis on Mobile Application Reviews	21
3.2	Automatic Analysis of App Reviews Using LLMs	21
4	Aspect Based Sentiment Analysis on Mobile Application Reviews	22
4.1	Introduction	22
4.2	Methodology	22
4.2.1	OverSampling the Data	23
4.2.2	Pre-Processing and Embedding	23
4.2.3	Feature extraction and classification	24
4.3	Experiments	25
4.3.1	Dataset	25
4.3.2	Evaluation Criteria	25
4.3.3	Data Over Sampling	25
4.3.4	Embeddings	26
4.3.5	Feature extraction and classification	26
4.4	Results and Analysis	27
4.4.1	Pre-processing	27
4.4.2	Optimization	27
4.4.3	Final Results and Error Analysis	28
4.5	Conclusion	29
5	Automatic Analysis Of App Reviews Using LLMs	32
5.1	Introduction	32
5.2	Methodology	32
5.2.1	LLMs Selection	33
5.2.2	Benchmarking Dataset	34
5.2.3	Dataset for Fine-Tuning Custom LLMs	35

5.2.4	Selection of Prompts	36
5.2.5	Fine-Tuning Open Source Models	40
5.2.6	Evaluation Strategy	41
5.3	Experiments	42
5.3.1	Evaluating Commercial Model Performance	44
5.3.2	Effects of Temperature and Top_p Parameters on Classification Accuracy of Commercial Models	44
5.3.3	Evaluating Dataset Quality through Inter-Annotator Agreement Analysis	48
5.3.4	Evaluating Open Source Models	49
5.3.5	Effect of Training Dataset Size on Model Performance	49
5.3.6	Effect of Number of Epochs on Model Performance	52
5.3.7	Evaluation of Explanation Quality	54
5.3.8	Effects of Temperature and Top_p on Fine-tuned Open-source Models	55
5.3.9	Model Performance Visualization	60
5.4	Conclusion	62
5.5	Limitations	62
5.6	Ethical considerations	63
6	Findings	65
6.1	Aspect-Based Sentiment Analysis on App Reviews	65
6.1.1	CNN-based Approach for Aspect-Based Sentiment Analysis	65
6.1.2	Impact of Word Embeddings	65
6.1.3	Effectiveness of Data Oversampling Techniques	66
6.1.4	Hyperparameter Tuning	66
6.2	Automatic Analysis of App Reviews Using LLMs	66
6.2.1	Evaluation of Popular Commercial Models	67
6.2.2	LLMs as Autonomous Annotators	67
6.2.3	Fine-tuning and Evaluating Open-source LLMs	67
7	Discussion and Conclusion	69

