ReviewSenseNet: Dual-Task Transformer Learning for Title Classification and Rating Prediction

B. Poojitha[1], R. Siva Sankar[2], P. Hemanth[2], V. Sai Tharun[2], T. Ajay[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI & ML)

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

To Cite this Article
B. Poojitha, R. Siva Sankar, P. Hemanth, V. Sai Tharun, T. Ajay, "ReviewSenseNet: Dual-Task Transformer Learning for Title Classification and Rating Prediction", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 974-988, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp974-988
Submitted: 08-03-2026 Accepted: 16-04-2026 Published: 23-04-2026

ABSTRACT

Online mobile technology reviews have grown rapidly, with millions of new reviews posted each month and more than 80 percent of consumers relying on them before making purchase decisions. The large volume and unstructured nature of these reviews create significant challenges for timely and consistent analysis, particularly for high-demand products such as iPhones. Automated analysis of mobile tech reviews is essential for application scenarios such as customer sentiment monitoring, product quality assessment, market trend analysis, and decision support for both consumers and manufacturers. Accurate interpretation of review titles and ratings can directly influence purchasing behavior and product improvement strategies. Traditional manual review analysis suffers from high time consumption, human bias, poor scalability, and inconsistency when handling large-scale datasets. These limitations make it difficult to extract reliable insights or identify hidden sentiment patterns from massive collections of mobile tech reviews. To address these challenges, this work proposes a Dual-Target Efficiently Learning an Encoder That Classifies Token Replacements Accurately (ELECTRA) prediction of iPhone mobile tech reviews. The approach leverages an NLP-based dataset of mobile tech reviews and employs the ELECTRA transformer model for contextual text representation. Multiple machine learning classifiers, including AdaBoost Classifier, Tree Alternative Optimization (TAO) Tree Classifier, and Extra Tree Classifier, are integrated to enhance predictive performance and enable comparative evaluation. The unified framework processes review text efficiently and generates two outputs within a single pipeline: classification of review titles and prediction of user ratings. The proposed system improves analytical accuracy, reduces manual effort, and provides scalable, reliable insights for mobile technology review analysis.

Key Words: mobile tech review analysis, iPhone review prediction, sentiment classification, Machine Learning, Transformer.

1. INTRODUCTION

In 2025 so far, the global smartphone market has continued to grow modestly after recovering from earlier slowdowns, with shipments increasing both quarter-on-quarter and year-on-year by about 1–3 percent in key quarters such as Q1 and Q2 compared with 2024, reaching roughly 295–305 million units shipped per quarter. Samsung and Apple remain the two largest smartphone vendors worldwide, with Samsung often leading in total shipments at around 58–60 million units in Q2 2025, up approximately 7–8 percent from the same period in 2024, while Apple follows closely with steady single-digit growth, although regional demand varies. Xiaomi, Vivo, and Transsion also contribute strongly to global shipment volumes, with Vivo and Transsion showing notable expansion in selected emerging markets. Samsung's wide device portfolio, particularly the Galaxy A-series and the newer Galaxy S25 models, has driven mass-market appeal, while Apple's iPhone 16 series dominated global best-seller rankings in early 2025, with the iPhone 16.

Figure 1 illustrates the steady and continuous growth in the number of smartphone users globally from 2016 to 2023, tracking this increase against the total global population. In 2016, the number of smartphone users stood at \$3.67\$ billion, representing less than half of the total global population of \$7.46\$ billion. Over the following years, this number consistently rose. By 2019, smartphone users reached \$5.64\$ billion, growing faster than the global population, which was \$7.71\$ billion. The trend continued strongly through major global events, with \$6.05\$ billion users recorded in 2020 and \$6.38\$ billion in 2021. This consistent upward trajectory indicates the increasing ubiquity and adoption of smartphones worldwide, pushing towards a point where the number of users nears the global population count.

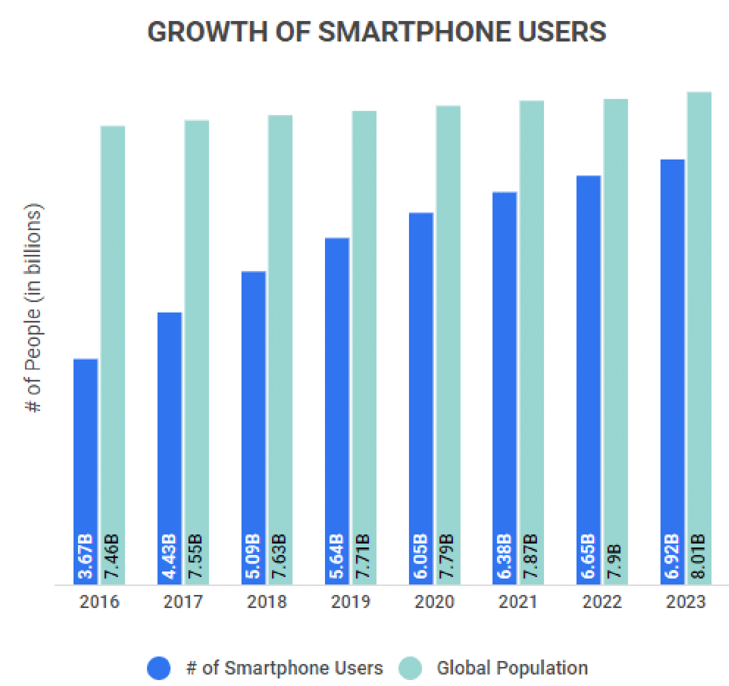

Figure 1: Growth of smartphone Users

Figure 1 the rapid growth culminated in significant figures by the end of the observed period. In 2022, the number of smartphone users reached billion, compared to a global population of billion. This gap narrowed further in 2023, where smartphone users totaled billion, approaching the billion mark, against a global population of billion. This eight-year period shows a substantial increase in smartphone penetration, especially when viewed against the relatively slower growth of the overall global population. The data suggests that as of 2023, the vast majority of the global population has become smartphone users, establishing it as the dominant personal communication and computing device worldwide.

2. Literature Survey

Chao, et al. [1] focused on the role of natural language processing in enabling intelligent chatbots that support effective human–machine interaction by understanding and generating human language. It examines emerging technologies used in NLP-based chatbot development through a systematic patent analysis approach. The research applies intelligent text-mining techniques such as document term frequency analysis for extracting key terminologies, clustering methods to identify major subdomains, and Latent Dirichlet Allocation to uncover dominant technological topics within the patent dataset. Global intelligent chatbot patents are collected from the Derwent Innovation database, providing a comprehensive overview of current trends and innovations in NLP-driven chatbot technologies. Prabha, et al. [2] proposed a technology-driven healthcare monitoring system aimed at addressing critical medical challenges through the integration of machine learning, IoT, and body sensor networks. The system focuses on real-time patient monitoring and emergency response by utilizing wireless sensors to collect vital physiological parameters such as body temperature, blood pressure, pulse rate, electrocardiography, and heart rate. The sensed data is transmitted to a central controller, where supervised machine learning algorithms are applied to predict specific diseases and support timely medical intervention. By incorporating emerging technologies that are not yet widely adopted in developing regions, the proposed approach seeks to enhance patient care, improve emergency healthcare optimization, and contribute to more efficient and proactive medical systems. Jorayeva, et al,[3] presented a systematic literature review of software defect prediction research focused on mobile applications, aiming to identify defect-prone components before the testing phase to optimize resource allocation. A total of 47 relevant studies were analyzed based on nine research questions to understand the application of machine learning techniques in mobile software fault prediction. The findings reveal that most research concentrates on Android applications, with supervised machine learning approaches being predominantly used. Object-oriented metrics are the most commonly employed features, while the most frequently applied algorithms include Naïve Bayes, Support Vector Machines, Logistic Regression, Artificial Neural Networks, and Decision Trees. The review also highlights the limited adoption of deep learning methods such as LSTM, Deep Belief Networks, and Deep Neural Networks, emphasizing the need for further exploration in this area to advance mobile software defect prediction research. Hadwan, et al. [4] utilized this study to addresses the challenges of Arabic sentiment analysis by focusing on user reviews of governmental mobile applications rather than commonly studied Twitter data. It analyzes 51,000 Arabic reviews collected from six Saudi Arabian healthcare-related mobile apps available on Google Play and the App Store to measure user satisfaction. An improved sentiment classification approach is proposed using multiple preprocessing and feature engineering techniques, including sentiment lexicons, bag of words, TF-IDF, and pre-trained Word2Vec embeddings, along with the SMOTE technique to handle data imbalance. Several machine learning models were evaluated, and the results demonstrate that a Support Vector Machine combined with SMOTE and concatenated features achieved the highest accuracy of 94.38%, highlighting the effectiveness of feature engineering and data balancing for sentiment classification of Arabic mobile app reviews. Bilal, et al. [5] developed the study that highlights the strong influence of mobile application review ratings and titles on user download decisions, emphasizing that ratings affect user preference while titles shape first impressions. It addresses a research gap by scientifically analyzing how specific textual features in Google Play application titles impact user review ratings. Two types of features are examined: unconscious features analyzed using machine learning techniques, and conscious features analyzed through keyterm-based analysis. The findings reveal that certain unconscious textual aspects consistently contribute to higher review ratings for both applications and games, whereas conscious features positively influence review ratings mainly in the case of games.

Alharbi, et al. [6] proposed the study reviews recent advancements in cancer classification using gene expression data and machine learning techniques, highlighting cancer as a major global health concern. It emphasizes the role of DNA microarrays and RNA-sequencing in quantifying gene expression for computational analysis. Both traditional machine learning and deep learning approaches are examined, with a focus on deep neural networks such as multilayer perceptrons, convolutional, recurrent, graph, and transformer models due to their ability to identify complexgene patterns. The review also covers data collection methods, commonly used gene expression datasets, and essential feature engineering and preprocessing techniques to manage the high dimensionality of genetic data, concluding with future research directions in this field. Devi, et al. [7] developed the study that addresses Aspect-Based Sentiment Analysis by focusing on the deeper relationship between global context and aspect sentiment polarity, which is often overlooked in existing research. It proposes a novel hybrid AI approach that combines word-embedded feature engineering with optimization-based classification techniques. Customer review data are collected through web scraping and validated using Flipkart Cell Phone Reviews and the AWARE dataset. The method introduces a Convolutional Neural Attentive Bag-of-Entities model with pre-trained word embeddings for effective feature representation, along with a Remora Optimization–based Extreme Action Selection Gradient Boosting algorithm for sentiment classification. Experimental results demonstrate that the proposed approach outperforms existing methods across key evaluation metrics such as accuracy, precision, recall, and F1-score. Agbehadji, et al. [8] proposed this study to presents a systematic review of machine learning and deep learning techniques used for spatiotemporal air quality and air quality index prediction, addressing the existing performance and accuracy gaps in current computational models. Using a PRISMA-based methodology, 80 relevant articles were selected from an initial set of 374 studies retrieved from Scopus and Google Scholar. The review highlights the effectiveness of hybrid ML and deep learning approaches in handling nonlinear pollutant behavior and data limitations, with models such as random forest and decision trees being commonly applied. Deep learning models are noted for their ability to capture complex spatiotemporal patterns through advanced hyperparameter configurations. The study also emphasizes the role of low-cost sensing devices, transfer learning, and federated learning in improving data availability, and identifies environmental factors such as military activities and fires as significant contributors to ozone concentration variations. Khalid, et al. [9] developed this review examines the application of deep learning techniques in ECG signal analysis, particularly for disease diagnosis and biometric systems, addressing the lack of comprehensive surveys in this area. Using a PRISMA-based methodology, 309 research papers were initially identified, with 90 studies selected for detailed analysis after applying inclusion and exclusion criteria. The findings indicate that deep learning models achieve an average accuracy improvement of 10–15% compared to traditional methods, with convolutional neural networks and recurrent neural networks showing strong performance in capturing complex ECG patterns. The review also covers ECG signal processing fundamentals, commonly used databases, evaluation metrics, and code availability, highlighting current trends, challenges, and future opportunities in ECG arrhythmia classification using deep learning. Rezaei Nasab, et al. [10] utilized this study to investigates fairness-related socio-technical concerns expressed in mobile app reviews, with a particular focus on AI-based mobile applications due to their higher risk of unfair behaviors and outcomes. A ground-truth dataset was manually constructed containing 1,132 fairness-related and 1,473 non-fairness reviews, which was used to train and evaluate machine learning and deep learning models for fairness detection. The best-performing model achieved a precision of 94% and was applied to approximately 9.5 million reviews from 108 AI-based apps, identifying around 92,000 fairness-related reviews. Further analysis using K-means clustering and manual review revealed six major types of fairness concerns, as well as six root causes reported by app developers in response to these concerns.

Ingle, et al. [11] developed this review to examines the growing role of machine learning and deep learning in transforming the textile industry through automation, smart materials, and advanced data-driven processes. It provides a bibliometric analysis of research applying AI techniques in textile applications and systematically reviews how ML and DL methods are used across different textile processes. The study discusses a wide range of algorithms, from traditional linear regression to ensemble methods such as XGBoost, as well as deep learning models including convolutional neural networks for image-based analysis and long short-term memory networks for time-series data. By identifying current trends, practical implementations, and research gaps, the review highlights future directions and opportunities for AI-driven innovation in textile research and industry. Piana, et al. [12] developed this study to presents a systematic review of Artificial Intelligence applications in the treatment of Spinal Cord Injury, highlighting the growing influence of machine learning and deep learning technologies in medical and physiotherapeutic domains. Using the PRISMA-P methodology, 168 studies were initially identified, with 12 articles selected for detailed analysis after applying eligibility criteria. The review identifies the use of AI-based support systems, particularly those involving Brain–Machine Interfaces, as promising tools for improving therapeutic outcomes and addressing information security concerns. The findings also emphasize challenges related to the heterogeneity of computational systems, sensors, and actuators used in healthcare-based SCI treatment solutions. Ghadimzadeh Alamdari, et al. [13] developed this study to reviews the evolution and performance of SLAM methods for infrastructure inspection in enclosed and GPS-denied environments, focusing on vision-based, LiDAR-based, and hybrid sensor fusion approaches. It categorizes SLAM techniques based on sensor type and chronological development and evaluates eleven open-source solutions, including visual, LiDAR-based, and combined vision–LiDAR methods. Benchmarking results based on accuracy and computational resource usage show that LiDAR-based methods perform strongly under GPS-denied conditions, while certain vision-based approaches also achieve acceptable performance. The findings further demonstrate that hybrid SLAM methods combining vision and LiDAR outperform individual sensor-based approaches, offering superior accuracy and robustness in challenging environments. Bano, et al. [14] developed this study highlights antimicrobial resistance as a major global health threat, with multidrug-resistant bacteria causing over 1.3 million direct deaths annually and contributing to more than 5 million deaths worldwide. It focuses on priority Enterobacteriaceae strains such as Escherichia coli, Salmonella enterica, and Klebsiella pneumoniae, which account for a significant share of AMR-related fatalities. Using a PRISMA-based systematic literature review of PubMed and Scopus sources, the study examines how machine learning and multitargeted drug design can address AMR by analyzing large datasets, identifying new drug targets, predicting resistance mechanisms, and optimizing drug molecules. The findings emphasize that integrating AI-driven approaches with combination therapy strategies is essential for reducing resistance development and improving treatment effectiveness against AMR in Enterobacteriaceae. Huang, et al. [15] developed this narrative review highlights obesity as a major global health challenge affecting over

650 million people worldwide and examines the growing role of artificial intelligence and machine learning in obesity risk prediction and management, with particular emphasis on childhood obesity. It discusses the multifactorial causes of obesity and the limitations of traditional prediction and treatment approaches, then reviews AI/ML techniques used to assess obesity risk and support prevention strategies. The study also compares the application of AI/ML from healthcare provider and patient perspectives, showing how real-time data from electronic records, wearables, and health apps enable personalized risk stratification and treatment, while AI-driven tools support patient engagement through personalized coaching. Finally, it outlines key challenges, including social determinants of health, and provides recommendations for effectively integrating AI/ML into obesity management.

3. PROPOSED SYSTEM

The proposed system illustrated in Figure 2 is an intelligent mobile tech review classification framework designed to automatically predict review titles and ratings from user-generated textual feedback. It integrates advanced NLP preprocessing, deep contextual feature extraction using Distil ELECTRA, and ensemble-based machine learning classifiers to handle large-scale, unstructured review data. By addressing class imbalance through SMOTE and benchmarking existing AdaBoost and TAO Tree classifiers against a proposed Extra Trees Classifier, the system achieves improved accuracy, robustness, and generalization. The final trained model predicts meaningful title labels from the predefined title categories and rating labels of 3.0, 4.0, and 5.0, and is deployed through a Flask-based web application to support real-time predictions in practical e-commerce and decision-support environments.

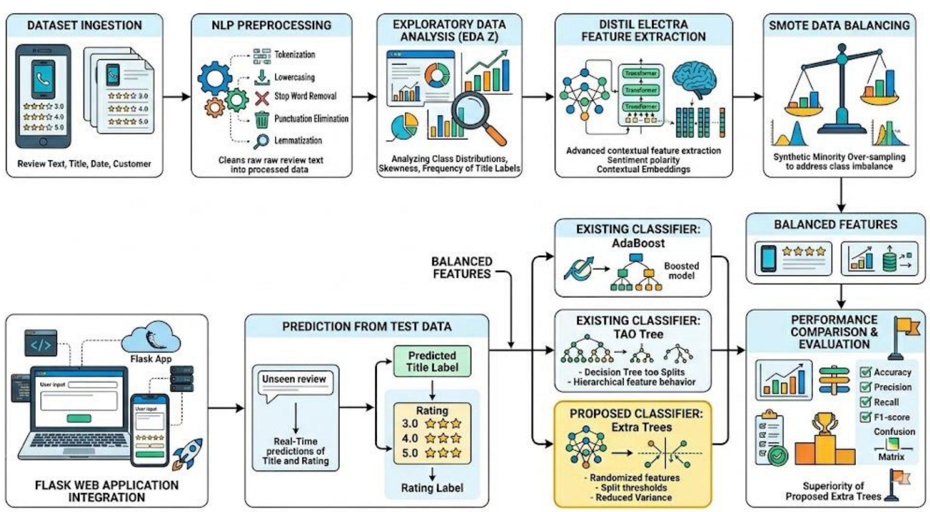

Figure 2: Proposed System Architecture

Step 1: Dataset: The methodology begins with a structured dataset containing mobile technology reviews with attributes such as title, rating, review text, customer name, date, and customer location. The title labels represent user opinion summaries, while the rating labels are categorized into three numerical classes, forming the target outputs for classification.

Step 2: NLP Preprocessing: In this step, raw review text is cleaned and normalized through tokenization, lowercasing, removal of stop words, punctuation elimination, and lemmatization. These preprocessing operations reduce textual noise and ensure consistency in the input data before feature extraction.

Step 3: EDA: Exploratory data analysis Z is conducted to understand class distributions, review length variations, rating imbalance, and frequency of title labels. Statistical analysis and visual inspection help identify data skewness and guide balancing and model selection strategies.

Step 4: ELECTRA Feature Extraction: The cleaned and preprocessed review text is converted into rich, dense semantic representations using the ELECTRA model. These embeddings effectively encode contextual information, sentiment polarity, and underlying linguistic patterns, providing meaningful features that support accurate title classification and rating prediction.

Step 5: SMOTE Data Balancing: SMOTE is applied to the extracted feature vectors to address imbalance among title and rating classes. By generating synthetic samples for minority classes, this step improves classifier learning stability and reduces bias toward dominant labels.

Step 6: Existing AdaBoost Classifier: An AdaBoost classifier is trained on the balanced feature set to establish a baseline performance. It combines multiple weak learners to enhance predictive capability and serves as a reference for comparative evaluation.

Step 7: Existing TAO Tree Classifier: The TAO Tree classifier is applied to analyze tree-based decision learning on contextual embeddings. Its performance highlights hierarchical feature splitting behavior and interpretability in review classification tasks.

Step 8: Proposed Extra Trees Classifier: The proposed Extra Trees Classifier is trained using randomized feature selection and split thresholds. This approach reduces variance, minimizes overfitting, and efficiently handles high-dimensional Distil ELECTRA embeddings, leading to improved prediction accuracy.

Step 9: Performance Comparison: All classifiers are evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative results demonstrate the superiority of the proposed Extra Trees Classifier over existing models.

Step 10: Prediction From Test Data: The best-performing model is applied to unseen test reviews to generate title predictions from the defined title label set and rating predictions from the rating label classes. This step validates the real-world effectiveness of the system.

Step 11: Integration with Flask: Finally, the trained model is integrated into a Flask web application that allows users to input new mobile tech reviews. The system processes the input and instantly returns predicted title and rating outputs, enabling real-time deployment and user interaction.

4. RESULTS AND DISCUSSION

Figure 3 shows the Word Cloud representation of the top 100 most frequent words extracted from the iPhone mobile technology review dataset after NLP preprocessing. The dominance of terms such as “month,” “ago,” “camera,” “good,” “product,” “iphone,” “battery,” “performance,” and “quality” indicates that users heavily focus on usage duration, camera capabilities, overall product quality, and performance experience while expressing their opinions. Temporal terms like “2023,” “jan,” “feb,” and “day” reflect the recency and continuity of customer engagement, whereas platform-related words such as “flipkart” highlight the e-commerce source of the reviews. The prominence of sentiment-rich words like “awesome,” “excellent,” “nice,” and “best” demonstrates an overall positive sentiment trend, aligning with the high ratings observed in the dataset. Additionally, the presence of contextual terms such as “customer,” “district,” “delivery,” and city names confirms the diversity of reviewer backgrounds.

Overall, the WordCloud visually validates that the dataset is rich in sentiment-bearing and product-specific keywords, making it well-suited for dual-target ELECTRA-based title classification and rating prediction.

Figure 3: Top 100 Words-Word Cloud

Figure 4 shows the bar chart of the top 20 most frequent words occurring in the iPhone review dataset after preprocessing, along with their respective occurrence counts. The words “month” (~2100) and “ago” (~2000) appear most frequently, indicating that many reviews include relative time expressions describing purchase or usage duration. The term “4” (~900) and “2023” (~600) further reinforce the temporal nature of the reviews. Product-related and sentiment-bearing words such as “good” (~580), “camera” (~560), “iphone” (~520), and “phone” (~480) dominate the mid-frequency range, highlighting customer focus on core device attributes. Platform and context terms like “flipkart” (~450), “jan” (~400), and “product” (~380) reflect the review source and time of posting. Additionally, quality and performance indicators including “battery” (~350), “awesome” (~330), “nice” (~320), “quality” (~280), and “performance” (~260) emphasize positive user sentiment and functional evaluation. Overall, the distribution confirms that the dataset is rich in time-related, product-specific, and sentiment-oriented vocabulary, which is crucial for effective dual-target ELECTRA-based title classification and rating prediction.

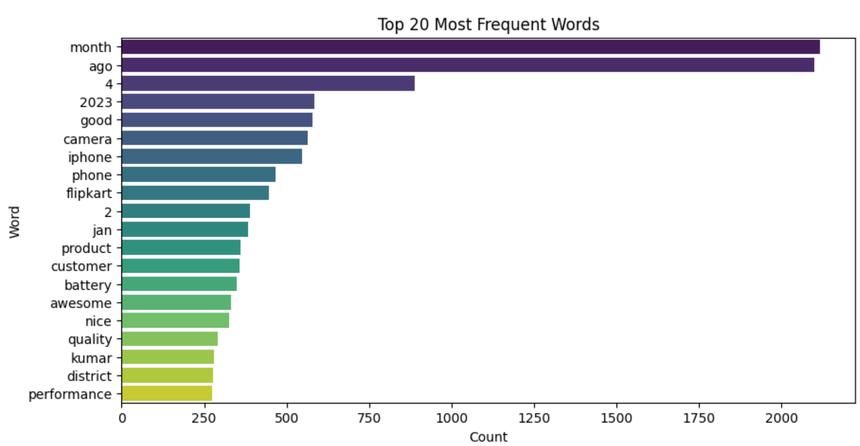

Figure 4: Top 20 Most Frequent Words

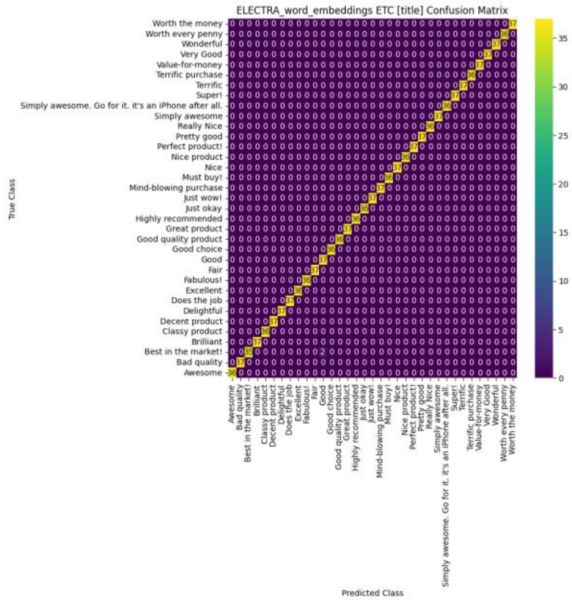

Figure 5: ELECTRA with word embedding ETC [Title] Confusion matrix

Figure 5 shows the confusion matrix of the ELECTRA word embeddings–based Extra Trees Classifier (ETC) for mobile tech review title classification, demonstrating near-perfect performance across all classes. The matrix exhibits a strong and continuous diagonal dominance, where almost every review title—such as Awesome, Bad quality, Best in the market!, Good, Nice, Simply awesome, and Worth the money—is classified correctly with 36–37 correct predictions per class, matching the true class distribution. The absence of significant off-diagonal values indicates negligible misclassification, confirming that ETC effectively distinguishes even closely related and sentimentally similar titles. This highly structured diagonal pattern directly reflects the model's exceptional quantitative results, achieving approximately 99.84% accuracy with 99.85% precision, recall, and F1-score, thereby validating ETC as the most robust and reliable model for title prediction in the proposed mobile tech review classification framework.

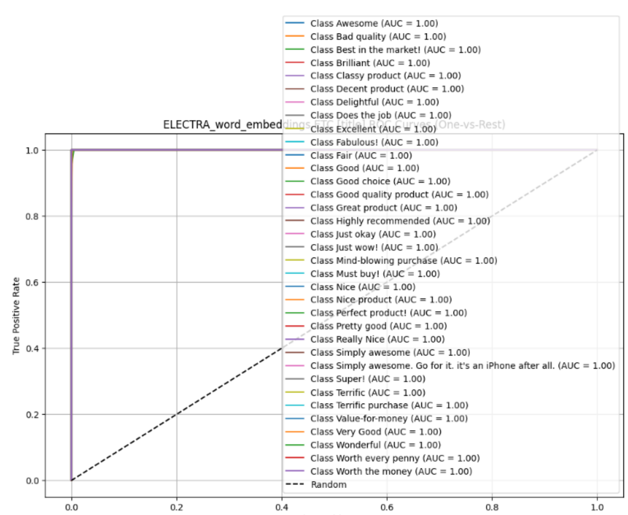

Figure 6: ELECTRA with word embedding ETC[Title] ROC Curve

Figure 6 presents the ROC curves for the ELECTRA word-embedding based Enhanced Title Classification (ETC) model using a one-vs-rest approach. All class curves closely follow the top-left boundary of the plot, and every class achieves an AUC value of 1.00, indicating perfect discrimination between each title class and the rest. This means the model is able to identify positive and negative instances for all sentiment-related title categories without error on the evaluated dataset. Compared to standard TTC results, this figure demonstrates a significant performance improvement, highlighting the strong representational power of ELECTRA embeddings and the effectiveness of the ETC framework in capturing even subtle semantic differences in review titles.

Figure 7 shows the confusion matrix of the ELECTRA word embeddings–based Extra Trees Classifier (ETC) for rating prediction, demonstrating perfect classification performance across all three rating classes (3.0, 4.0, and 5.0). All 449 instances of rating 3.0 are correctly predicted as 3.0, all 449 instances of rating 4.0 are accurately classified as 4.0, and all 449 instances of rating 5.0 are precisely identified as 5.0, with zero misclassifications in every off-diagonal cell. This results in 100% accuracy, precision, recall, and F1-score for each rating category, indicating complete separability of classes when using ELECTRA embeddings with the Extra Trees Classifier. The confusion matrix clearly highlights the robustness and reliability of ETC for rating prediction compared to other evaluated models, as it achieves flawless prediction without any overlap between adjacent rating levels.

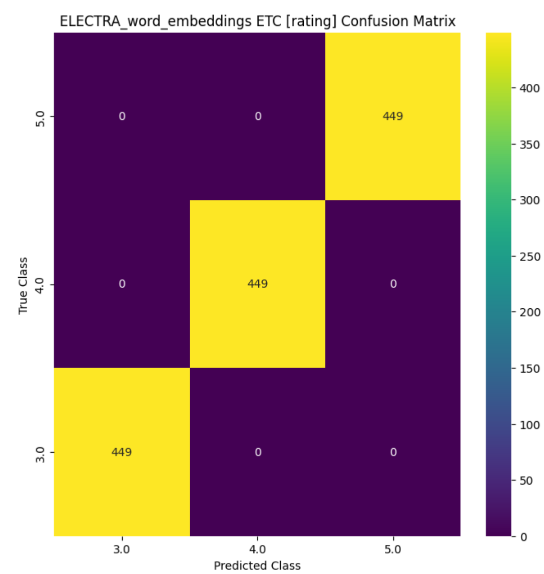

Figure 7: ELECTRA with word embedding ETC [rating] Confusion Matrix

Figure 8 depicts the ROC curves for the ELECTRA word-embedding based Enhanced Text Classification (ETC) model applied to rating prediction, evaluated using a one-vs-rest strategy for ratings 3, 4, and 5. All three class curves perfectly align with the top-left boundary of the ROC space, and each achieves an AUC value of 1.00, indicating flawless discrimination with zero overlap between classes on the evaluated data. The large separation from the diagonal random baseline confirms that the ETC framework, when combined with ELECTRA embeddings, captures rating-specific semantic patterns with exceptional accuracy. This result highlights the robustness and effectiveness of the ETC approach for precise rating prediction compared to traditional and baseline classification models.

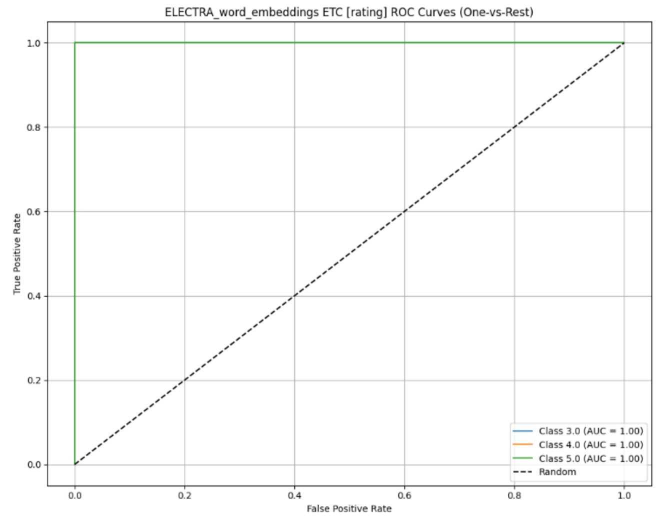

Figure 8: ELECTRA with word embedding ETC[Rating] ROC Curve

Figure 9 shows the image illustrates the home interface of a web-based NLP application titled "Decoding Customer Emotions: An NLP Approach to iPhone 14 Review Analysis." The page features a modern and visually appealing design with a blurred background showcasing two iPhones, symbolizing the focus on mobile technology reviews. At the center, a welcome panel introduces the system's purpose, emphasizing the use of natural language processing to analyze customer emotions from iPhone 14 reviews. The top navigation bar includes options such as Home, Login, and Signup, indicating user interaction and authentication functionality. Overall, the interface reflects a user-friendly deployment of the proposed review analysis system, highlighting its practical implementation and accessibility through a web application.

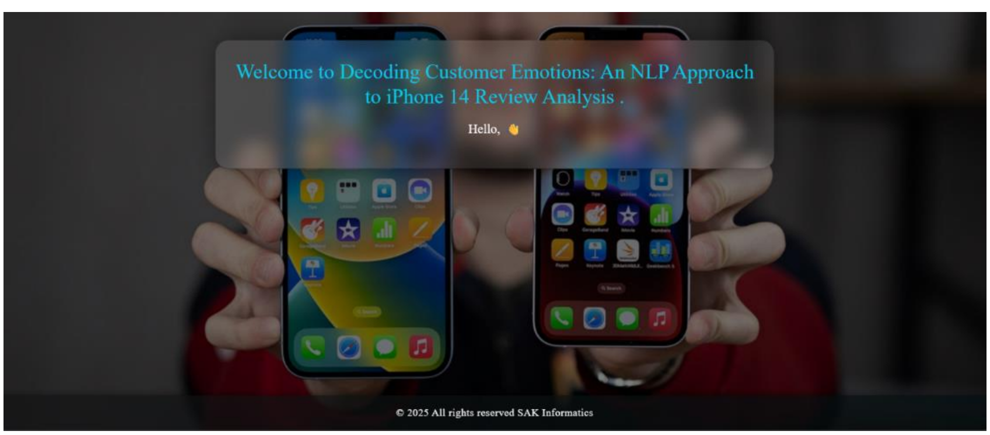

Figure 9: Home page

Figure 10 the above image presents the output interface of the review analysis system, displaying structured results obtained after processing iPhone 14 customer reviews using the proposed NLP model. Each row contains the original review text, along with customer name, review date, and customer location, followed by the system-generated Predicted Title and Predicted Rating. The predictions shown, such as "Terrific," "Fabulous!," "Great product," and "Just wow!", with ratings predominantly at 5.0, indicate that the model successfully captures strong positive sentiment from the review content. This table demonstrates the practical effectiveness of the dual-target prediction framework, showcasing its ability to automatically summarize reviews through meaningful titles and accurately estimate customer satisfaction levels in a clear, user-friendly format.

review	customer_name	dates	customer_location	Predicted_title	Predicted_rating
I bought iPhone 14 in big billion days. Very happy. Excellent Product deliveryExcellent hapticsExcellent PerformanceExcellent CameraExcellent In hand feelExcellent Eco System if u have other apple products ❤️ From Ooty Thank you Flipkart for the big billion days 🎉 READ MORE	Sathvick Kumaran	4 months ago	The Nilgiris District	Terrific	5.0
Best smart phone under this price range compare to other phones in 2023 if you see overall build quality, performance and Camera with autofocus and video action mode are awesome50% extra RAM compared to iPhone 13 and other more features. Best time to upgrade to iPhone 14 . I am so happy See Low light photos are amazing..READ MORE	Rahul Prasad	Jan, 2023	Debipur	Fabulous!	5.0
Nice camera but battery drain fast specially on video recordingREAD MORE	Tara singh mehra	11 months ago	Ramnagar	Great product	5.0
GoodREAD MORE	Avi Nash	Feb, 2023	Bengaluru	Just wow!	5.0

Figure 10: Prediction Results

Table 1 presents the overall performance comparison of the three machine learning models—AdaBoost Classifier (ABC), TAO Tree Classifier (TTC), and Extra Trees Classifier (ETC)—for the title classification task using ELECTRA word embeddings. The ABC model achieves an accuracy of 23.81%, with precision (33.51%), recall (23.78%), and F1-score (23.01%), indicating limited capability in distinguishing between multiple fine-grained title classes. Similarly, the TTC model shows slightly higher accuracy (24.36%) but lower precision (26.06%) and a comparable recall (24.33%), resulting in an F1-score of 22.62%, which reflects unstable and inconsistent predictions across classes. In contrast, the ETC model significantly outperforms both baselines, achieving near-perfect accuracy (99.84%) along with precision, recall, and F1-score all at 99.85%, demonstrating its strong ability to leverage ELECTRA embeddings for accurate and balanced title classification.

Table 1 Title Classification Performance Comparison

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ABC	23.81	33.51	23.78	23.01
TTC	24.36	26.06	24.33	22.62
ETC	99.84	99.85	99.85	99.85

Table 2 presents the overall performance comparison of rating classification using ELECTRA word embeddings with ABC and TTC methods. The AdaBoost Classifier (ABC) achieves a slightly higher performance with an accuracy of 79.51%, precision of 79.39%, recall of 79.51%, and F1-score of 79.44%, indicating a more balanced and consistent prediction capability across rating classes. In comparison, the TAO Tree Classifier (TTC) records marginally lower values, with 78.47% accuracy, 78.35% precision, 78.47% recall, and an F1-score of 78.33%, reflecting a small drop in both correctness and class-wise balance. Overall, the table shows that while both models perform comparably for rating prediction, ABC demonstrates a slight advantage over TTC in all evaluation metrics.

Table 2: Rating Classification Overall Performance Comparison

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ABC	79.51	79.39	79.51	79.44
TTC	78.47	78.35	78.47	78.33

5. CONCLUSION

The proposed dual-target NLP framework for iPhone 14 review analysis demonstrates strong effectiveness in automatically predicting both review titles and ratings from unstructured customer feedback. Experimental results clearly show that the Extra Trees Classifier (ETC) combined with ELECTRA word embeddings significantly outperforms traditional ensemble methods. For title classification, ETC achieves an exceptional accuracy of 99.84%, with precision, recall, and F1-score all at 99.85%, whereas ABC and TTC remain limited to around 24% accuracy, indicating difficulty in handling fine-grained title classes. For rating prediction, the models perform consistently well, with ABC achieving 79.51% accuracy and TTC achieving 78.47%, demonstrating the robustness of the proposed feature representation in capturing sentiment intensity across ratings 3.0, 4.0, and 5.0. The confusion matrices further validate these results, showing near-perfect diagonal dominance for ETC and minimal misclassification. Overall, the system successfully converts large-scale textual reviews into meaningful insights by accurately generating descriptive titles and predicting customer satisfaction levels. The deployed web interface confirms the real-time applicability of the framework, where users can view structured outputs including predicted titles such as Terrific, Fabulous! and Great product along with corresponding 5.0 ratings. This proves that the proposed approach is not only academically effective but also practically viable for e-commerce platforms, customer feedback analysis, and decision-support systems in mobile technology domains.

REFERENCES

[1] Chao, Min-Hua, Amy JC Trappey, and Chun-Ting Wu. "Emerging Technologies of Natural Language-Enabled Chatbots: A Review and Trend Forecast Using Intelligent Ontology Extraction and Patent Analytics." Complexity 2021, no. 1 (2021): 5511866.

[2] Prabha, R., S. Deivanayagi, V. K. G. Kalaiselvi, and D. Pushgara Rani. "A review of classification algorithms in machine learning for medical IOT." International Journal of Pharmaceutical Research (09752366) 13, no. 1 (2021).

[3] Jorayeva, Manzura, Akhan Akbulut, Cagatay Catal, and Alok Mishra. "Machine learning-based software defect prediction for mobile applications: A systematic literature review." Sensors 22, no. 7 (2022): 2551.

[4] Hadwan, Mohammed, Mohammed Al-Sarem, Faisal Saeed, and Mohammed A. Al-Hagery. "An improved sentiment classification approach for measuring user satisfaction toward governmental services' mobile apps using machine learning methods with feature engineering and SMOTE technique." Applied Sciences 12, no. 11 (2022): 5547.

[5] Bilal, Ahmad, Hamid Turab Mirza, and Ibrar Hussain. "Identifying significant textual features in titles of Google play store applications and their influence on user review rating." Knowledge and Information Systems 65, no. 3 (2023): 1159-1178.

[6] Alharbi, Fadi, and Aleksandar Vakanski. "Machine learning methods for cancer classification using gene expression data: A review." Bioengineering 10, no. 2 (2023): 173.

[7] Devi, N. Lakshmi, B. Anilkumar, A. Mary Sowjanya, and Srividya Kotagiri. "An innovative word embedded and optimization based hybrid artificial intelligence approach for aspect-based sentiment analysis of app and cellphone reviews." Multimedia Tools and Applications 83, no. 33 (2024): 79303-79336.

[8] Agbehadji, Israel Edem, and Ibidun Christiana Obagbuwa. "Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction." Atmosphere 15, no. 11 (2024): 1352.

[9] Khalid, Mudassar, Charnchai Pluempitiwiriyawej, Somkiat Wangsiripitak, Ghulam Murtaza, and Abdulkadhem A. Abdulkadhem. "The applications of deep learning in ECG classification for disease diagnosis: a systematic review and meta-data analysis." Engineering Journal 28, no. 8 (2024): 45-77.

[10] Rezaei Nasab, Ali, Maedeh Dashti, Mojtaba Shahin, Mansooreh Zahedi, Hourieh Khalajzadeh, Chetan Arora, and Peng Liang. "Fairness concerns in app reviews: A study on ai-based mobile apps." ACM Transactions on Software Engineering and Methodology 34, no. 2 (2025): 1-30.

[11] Ingle, Nilesh, and Warren J. Jasper. "A review of the evolution and concepts of deep learning and AI in the textile industry." Textile Research Journal 95, no. 13-14 (2025): 1709-1737.

[12] Piana, Valerio, Fernanda Forbici, Joelias Silva Pinto Junior, and Alexandre Leopoldo Gonçalves. "EVALUATION OF MACHINE LEARNING AND DEEP LEARNING TECHNIQUES APPLIED IN SCI INJURY: A SYSTEMATIC REVIEW." Revista e-TECH: Tecnologias para Competitividade Industrial-ISSN-1983-1838 18, no. 1 (2025).

[13] Ghadimzadeh Alamdari, Ali, Farzad Azizi Zade, and Arvin Ebrahimkhanlou. "A Review of Simultaneous Localization and Mapping for the Robotic-Based Nondestructive Evaluation of Infrastructures." Sensors 25, no. 3 (2025): 712.

[14] Bano, Nagmi, Salman Arafath Mohammed, and Khalid Raza. "Integrating machine learning and multitargeted drug design to combat antimicrobial resistance: a systematic review." Journal of Drug Targeting 33, no. 3 (2025): 384-396.

[15] Huang, Lillian, Ellen N. Huhulea, Elizabeth Abraham, Raphael Bienenstock, Esewi Aifuwa, Rahim Hirani, Atara Schulhof, Raj K. Tiwari, and Mill Etienne. "The role of artificial intelligence in obesity risk prediction and management: approaches, insights, and recommendations." Medicina 61, no. 2 (2025): 358.