Sentiment-Based Rating Model using VADER and TextBlob

KARAN GUPTA[1], RAJ ARYAN[2], ASHWIN TIWARI[3], ANJANEYA GUPTA[4] AND SUTEERTH SUBRAMANIAM[5]

[1,2,3,4,5]B. Tech, Vellore Institute of Technology

***Abstract*— This study presents an innovative sentiment-based rating model designed to convert sentiment analysis outputs into a standardized 5-star rating system using various scaling methods. Utilizing VADER and TextBlob for comprehensive sentiment analysis, our model employs linear, exponential, piece-wise linear, quantile, and sigmoid scaling methods to normalize sentiment scores effectively. We systematically evaluate the effectiveness of each scaling method, aiming to optimize both the predictive accuracy and reliability of our rating conversions. The performance metrics, including accuracy, Mean Squared Error (MSE), and F1 scores, demonstrate the strengths and limitations of each method, providing insights into their suitability for diverse analytical needs in sentiment-based applications.**

***Index Terms*— Sentiment Analysis, Natural Language Processing, VADER, TextBlob, Scaling Methods, Rating Systems, Model Evaluation, Machine Learning, Sentiment Polarity, Text Analysis, Data Normalization, Predictive Accuracy.**

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is a field within natural language processing that focuses on identifying and categorizing opinions expressed in text in order to determine the writer's attitude towards a particular topic or overall contextual polarity of the text. This involves algorithms and methods that can extract, identify, and characterize the sentiment content of textual data sources [1]

Sentiment analysis techniques are diverse, ranging from simple lexical approaches to sophisticated machine learning techniques. Early methods often relied on counting words that appeared in positive or negative lists, while more advanced methods use complex algorithms including support vector machines, deep learning networks, and other forms of machine learning to identify sentiment from text contexts

The field of sentiment analysis not only helps in understanding the sentiment of text but also aids in various practical applications such as brand monitoring, product analytics, and understanding consumer needs. Sentiment analysis can also be extended beyond mere polarity assessment to detect emotions and intentions, and can operate at varying levels of granularity from document-level analysis down to individual aspect-based sentiment analysis [2] In practice, sentiment analysis is challenging due to the subtleties of human language, including irony, sarcasm, implicit context, and the evolving use of slang and neologisms. Each of these factors can obscure the sentiment conveyed in text, requiring increasingly sophisticated analytical techniques to accurately interpret human emotions and opinions conveyed in written form.

The significant contributions of this paper are outlined as follows:

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically tuned to capture sentiments expressed in social media. It is particularly useful for its simplicity and effectiveness in handling a variety of social media texts, including those that use shorthand, slang, and emoticons, which are typical in dynamic and informal environments. VADER combines a dictionary of sentiment-related words with a set of five heuristic rules that account for things like punctuation, capitalization, intensifiers, and conjunctions to better indicate the sentiment expressed in a text. This methodology allows VADER to effectively differentiate between positive, negative, and neutral sentiments by evaluating the text on a token-by-token basis, giving each piece a corresponding sentiment rating based on the lexicon and the contextual rules

Consider the sentence: "Anurag finds the latest edition of cameras very useful." Using VADER to analyze this sentence involves several steps:

- Tokenization: Splitting the sentence into individual words or tokens.
- Lexicon Rating: Each token receives a sentiment rating if it is present in the VADER lexicon. Words like "latest" and "useful" would typically have positive sentiment scores associated with them. The word "very" is an intensifier and would amplify the sentiment score 7 of "useful."
- Rule Application: VADER applies its rules to account for things like capitalization (which might intensify sentiment) or the degree to which modifiers affect the sentiment of neighboring words.

In this analysis, the overall sentiment of the sentence would likely be strongly positive, driven by the words "latest" and "very useful." This illustrative example showcases how VADER not only evaluates individual words for sentiment but also considers the modifying effects of intensifiers and the context provided by surrounding words to give a nuanced understanding of sentiment in text.

TextBlob is a versatile Python library widely used in natural language processing (NLP) for various tasks, including sentiment analysis. It provides a simple interface for common NLP tasks, such as part-of-speech tagging, noun phrase extraction, and especially, sentiment analysis. The sentiment analysis functionality of TextBlob relies on a pre-trained model that evaluates the polarity and subjectivity of text based on the presence and arrangements of words which are scored within a predefined lexicon.

In sentiment analysis, TextBlob offers two key measurements:

- Polarity: This is a floating-point number within the range of -1.0 to 1.0, where -1.0 signifies a completely negative sentiment, 0 stands for neutral, and 1.0 represents a completely positive sentiment.
- Subjectivity: This score ranges from 0.0 to 1.0, where 0.0 is very objective and 1.0 is very subjective, indicating that the text expresses personal opinions rather than factual information.

TextBlob's approach to sentiment analysis is notably straightforward, making it accessible for users who may not have an extensive background in machine learning or programming. It uses a Naive Bayes classifier that has been trained on a movie reviews dataset for sentiment classification. This allows it to perform sentiment analysis on a variety of texts by evaluating the presence and combinations of sentiment-laden words,

Sentiment analysis, an essential branch of natural language processing, significantly impacts the interpretation of emotional nuances in text across various sectors, including market analytics and social media insights. While tools like VADER and TextBlob are effective for basic sentiment detection, their integration into a nuanced 5-star rating system exposes substantial deficiencies. These tools struggle with the complexity and subtlety required by such a granular evaluation system, highlighting the need for advancements that can more accurately reflect the depth of human emotions in automated sentiment assessments.

This research introduces an innovative sentiment-based rating model that incorporates a variety of scaling methods—linear, exponential, piece-wise linear, quantile, and sigmoid—to more precisely map sentiment analysis outputs into a consistent 5-star rating framework. Each scaling approach is meticulously evaluated to determine its impact on improving predictive accuracy and overall reliability. Key performance metrics, such as accuracy, Mean Squared Error (MSE), and F1 scores, are used to benchmark and validate the effectiveness of each method. This comprehensive evaluation helps to ensure that the proposed model not only meets but exceeds the standards required for accurate sentiment interpretation across different contexts.

This paper extends the frontiers of sentiment analysis, tackling significant technical challenges while setting the stage for advanced future research. We delve into sophisticated sentiment decoding mechanisms, focusing on the potential integration of dynamic machine learning algorithms. These algorithms are designed to adapt to evolving linguistic patterns, enhancing the model's ability to interpret complex emotional data accurately. This forward-looking approach promises to revolutionize sentiment analysis, offering more precise and adaptable tools for understanding and reacting to human emotions in text.

Previous Challenges:

- Inaccuracy in Extreme Sentiments: Traditional models often failed to effectively capture and accurately classify sentiments at the extreme ends of the emotional spectrum. This led to significant discrepancies in sentiment analysis, particularly when evaluating highly positive or negative feedback.
- Cultural and Linguistic Nuances: Conventional sentiment analysis tools showed limited capability in adapting to the nuances of diverse cultural and linguistic contexts. This resulted in a lack of sensitivity to the subtleties of language used by different demographic groups, impacting the accuracy of sentiment interpretations across global datasets.
- Scalability Issues: Earlier approaches to sentiment analysis struggled with scalability, particularly in processing and analyzing large volumes of data from varied sources. This limitation hindered the effectiveness of sentiment analysis applications in big data environments, where the ability to scale efficiently is crucial.
- Integration of Sarcasm and Irony: Detecting and interpreting sarcasm and irony in text has been a significant challenge. Traditional models often misinterpreted these complex linguistic features, leading to incorrect sentiment assessments. This issue was compounded by the fact that sarcasm and irony are heavily context-dependent and can vary widely in expression.

Current Study's Approaches:

- Enhanced Scaling Methods: We introduce multiple scaling methods—linear, exponential, piece-wise linear, quantile, and sigmoid—to accurately convert sentiment scores into a 5-star rating system. Each method, from linear's direct proportionality to sigmoid's smooth transitions, refines how sentiments are quantified, enhancing the precision and reliability of sentiment analysis.
- Robustness and Scalability: Advanced algorithms enhance the robustness and scalability of our model, ensuring stable performance across large and varied datasets. This is crucial for real-time analysis in dynamic environments, where data quality and sentiment complexity can vary widely.
- Weight Assignment: Assigning computational weights to different aspects of reviews, such as sentiment intensity or keyword relevance, improves result accuracy. This prioritization allows for a more nuanced interpretation of sentiment, essential for applications like product feedback and customer service optimization.

II. BACKGROUND AND LITERATURE REVIEW

A. Sentiment Analysis in Online Reviews

Sentiment analysis is the process of determining the sentiment embedded within textual data. Sentiment can be broadly defined as an attitude, thought, or judgment influenced by feelings, with its etymology rooted in the Latin word 'sentire', meaning to feel. This field is essential for interpreting complex online reviews across various platforms such as social media and e-commerce sites, where understanding consumer sentiments toward products or services is vital. [3]

The concept of sentiment in text involves several dimensions; it not only includes the opinion itself but also the emotional tone accompanying that opinion. Scholarly definitions evolve around three primary streams:

- Opinion-Oriented Sentiment: Here, sentiment is primarily seen as an expression of opinion, manifesting through polarity indicators that categorize opinions as positive or negative. This view emphasizes sentiment as a reflection of subjective evaluations towards a subject or its aspects.
- Feeling-Oriented Sentiment: This perspective defines sentiment more as an emotional response, highlighting the personal feelings triggered by the subject under discussion.
- Hybrid Sentiment: Combining the first two, this approach considers sentiment as both an emotional response and an evaluative stance, linking feelings directly with opinions expressed.

In practical terms, sentiment analysis algorithms must contend with the complexity of human language, capturing nuances that include slang, idiomatic expressions, and implicit sentiment expressions.

Tools like VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob have been engineered to discern these subtleties effectively, particularly in informal text types like tweets and online reviews, which often feature implicit sentiments and contextual dependencies.

Moreover, sentiment analysis in online reviews typically requires understanding not just the sentiment but also its components as identified in computational linguistics: the target entity (e.g., a product), the aspect of the entity being discussed (e.g., quality), the holder of the opinion, the timing of the opinion, and the sentiment polarity towards the aspect. This comprehensive analysis helps in extracting actionable insights from customer feedback, aiding businesses in market strategy adjustments and product enhancements.

B. Literature Review

Sentiment analysis, a pivotal element within natural language processing (NLP), systematically evaluates textual expressions of emotion. This computational analysis initially focused on extracting sentiments from structured data like product reviews and has since expanded into unstructured data across diverse digital platforms.

Historical Development of Sentiment Analysis

- Early Foundations: The origins of sentiment analysis are rooted in the early 20th century with public opinion analysis. Its computational aspect began to flourish in the 1990s with the advent of the internet and the subsequent explosion of subjective content online. This period marked a pivotal shift towards automated sentiment analysis, driven by the widespread availability of web-based texts [4]
- Technological Advancements: The development of Web 2.0 technologies spurred significant advances in sentiment analysis. The need to analyse vast amounts of user-generated content on social media and online platforms catalysed improvements in natural language processing, enhancing capabilities in sentiment classification and opinion extraction [5]
- Expansion and Diversification: Over the last decade, the application of sentiment analysis has broadened significantly, extending into fields such as finance, healthcare, and politics. This diversification reflects the growing complexity and variety of data sources, particularly from social media platforms like Twitter and Facebook [6]

C. Recent Developments

- It has advanced from basic lexicon-based methods to sophisticated machine learning and deep learning models. Early methods categorized sentiments based on predefined word lists, whereas modern techniques employ complex algorithms that consider semantic and contextual nuances [2]
- It encounters several challenges, including the detection of sarcasm, negations, and context-dependent meanings, which can significantly alter sentiment interpretation. Moreover, multilingual sentiment analysis presents additional complexities as it requires understanding nuances across different languages and cultures [1]
- Integrating sentiment analysis with additional data types such as demographic information and behavioral patterns can significantly refine sentiment predictions. This multidimensional analysis helps in understanding not just the "what" but also the "why" behind public opinions and emotions [7]
- Sentiment analysis has broad applications ranging from marketing to public health, where it's used to gauge public sentiment towards products, policies, or social events. This capability is crucial for businesses and governments to respond proactively to public needs and changes in opinion [8]
- The future of sentiment analysis lies in enhancing the accuracy of sentiment detection tools through improved algorithms that can handle the complexities of language and emotion more effectively. Research continues to focus on developing models that are not only precise but also capable of handling diverse datasets without bias [9]

Our study expands upon the existing body of knowledge by investigating innovative scaling methods to transform sentiment scores into a universally recognized 5-star rating system. This endeavor aims to bridge the divide between raw sentiment data and the generation of tangible, actionable insights, thereby enhancing the interpretability and utility of sentiment analysis in various applications.

III. RESEARCH DESIGN AND METHODOLOGY

A. Data Collection

Data Source: The dataset comprises online course reviews sourced from Coursera, a leading e-learning platform, ensuring a diverse and representative sample. It includes textual reviews along with accompanying metadata such as reviewer's name, user-assigned star ratings, review date, and course ID. These elements provide a benchmark for evaluating the model's performance. To maintain anonymity and comply with data privacy regulations, reviewers' names were anonymized during the data preprocessing phase.

Selection Criteria: The selection of the dataset was strategically designed to capture a broad spectrum of consumer sentiments. It includes reviews of varying lengths and emotional intensities, spanning multiple course categories. This diversity is crucial for testing the robustness of sentiment analysis tools under varied contexts. The dataset encompasses:

- Range of Reviews: Courses with reviews ranging from over 45,000 to fewer than 10, to ensure the model's effectiveness across different volumes of data.
- Course Type: Both academic and non-academic courses were included to cover a wide range of subjects, appealing to reviewers of diverse ages, ethnicities, and backgrounds.
- Access Type: A mix of open-access and paid courses were selected to analyze how
- Monetary investment influences user sentiment and feedback.

These criteria were chosen to ensure comprehensive coverage and to facilitate a robust test of the sentiment analysis model across diverse educational content and reviewer demographics.

Data Preprocessing: Data preprocessing was meticulously carried out to clean and normalize the text, ensuring the accuracy of sentiment analysis. The steps included:

- Line Termination: Handling line terminators (e.g., \n) to maintain the integrity of text data.
- Data Structuring: Utilizing the Pandas library to frame and manipulate the data effectively, facilitating ease of analysis.
- Anonymization: Removing or masking personal identifiers, such as reviewer names, to preserve user privacy and comply with ethical data handling practices.

B. Sentiment Score Calculation

Using VADER and TextBlob:

- VADER: Calculates a compound sentiment score for each review. The compound score, which ranges from -1 (extremely negative) to +1 (extremely positive), reflects the overall sentiment polarity based on the intensity and modifiers in the text.
- TextBlob: Outputs a polarity score between -1 and +1 and a subjectivity score that ranges from 0 (objective) to 1 (subjective). For this model, only the polarity score is utilized for rating conversion.

Criteria	VADER	TextBlob
Lexicon Based	Yes	Yes
Handles Negation	Yes	No
Emphasizes Sentiment Intensity	Yes	No
Customizability	Limited	Moderate
Suited for Social Media Text	Yes	No
Accuracy in Mixed Sentiments	High	Moderate
Requires Training Data	No	Yes
Provides Subjective Analysis	No	Yes
Support for Slang and Abbreviations	Yes	No
Dependency Parsing	No	Yes

Table I. VADER vs TextBlob Sentiment Analysis Model

C. Scaling Methods

To convert sentiment scores into a 5-star rating system, the following scaling methods are employed, including the newly added linear scaling:

1. Linear Scaling

$$S_{\text{linear}} = 1 + 2 \cdot (P + 1) \qquad (1)$$

Here, Slinear is the scaled star rating and P is the polarity score from the sentiment analysis tool, which ranges from -1 to +1. This formula effectively maps the polarity score directly to a 1-to-5-star scale. The addition of 1 adjusts the minimum score to zero, and multiplying by 2 stretches the scale to span from 0 to 4. Finally, adding 1 shifts the entire scale to range from 1 to 5.

2. Linear Multiple Scaling

$$S(p) = 2.5 \times (p + 1) \qquad (2)$$

Similarly, p is shifted by 1 to convert the polarity score from [-1,1] to [0,2], then multiplied by a factor (k=2.5) to scale it out of 5.

3. Exponential Scaling

$$E(p) = 2.5 + 3 \times \tanh(p) \qquad (3)$$

E(p) effectively transforms sentiment polarity into a new scale that spans approximately from 0 to 5, placing neutral sentiments (around p=0) near the middle of the scale (2.5) and becoming increasingly sensitive as sentiments approach strong positive or negative values. This transformation is particularly useful in scenarios where it is important to differentiate moderate from extreme sentiments.

4. Piece-wise Linear Scaling

$$P(p) = \begin{cases} \max(1, 2.97 + 3(p + 0.1)) & \text{if } p < -0.1 \\ 3 + 2.25(p + 0.1) & \text{if } -0.1 \leq p < 0.35 \\ \min(5, 4 + (p - 0.35) \times 1.538) & \text{if } p \geq 0.35 \end{cases}$$

(4)

- First Case (p<−0.1)

A. 2.97+3(p+0.1): When p is less than -0.1, this formula adjusts p upwards slightly (by adding 0.1) and then scales it up. The base value of 2.97 adjusts the starting point of the scale. The multiplication by 3 increases the sensitivity of the scale to changes in sentiment polarity in this range.

B. max(1,·): Ensures that the rating does not fall below 1, setting 1 as the minimum possible rating.

- Second Case (−0.1≤p<0.35)

A.3+2.25(p+0.1): For values of p in a moderate range, this segment shifts p upwards slightly and scales it more gently than in the first case, using a coefficient of 2.25. This adjustment is designed to spread the sentiment values effectively across the middle range of the star ratings, particularly utilizing the 3 to 4 range.

- Third Case (p≥0.35)

A. 4+(p−0.35)×1.538: For high positive values of p, starting from 0.35, this formula shifts p downwards slightly before scaling it. The multiplication factor of 1.538 ensures a steeper increase in ratings, making it easier to achieve higher ratings.

B. min(5,·): Caps the maximum possible rating at 5, ensuring ratings do not exceed the scale.

4. Quantile-based Scaling

$$Q(x) = \begin{cases} 1.5 & \text{if } x \leq q[0] \\ 2.5 & \text{if } q[0] < x \leq q[1] \\ 3.5 & \text{if } q[1] < x \leq q[2] \\ 4.0 & \text{if } q[2] < x \leq q[3] \\ 5.0 & \text{if } x > q[3] \end{cases} \tag{5}$$

Quantiles Calculation: The quantiles q=[q0,q1,q2,q3,q4] are calculated from the sentiment polarity scores using specific percentiles (10th, 33rd, 50th, 67th, and 83rd). These quantiles effectively divide the data into different segments that correspond to different rating levels.

The choice of quantiles (10th, 33rd, 50th, 67th, and 83rd percentiles) and the corresponding ratings (1.5, 2.5, 3.5, 4, 5) are determined through experimentation. This experimentation involves Analyzing the distribution Boundaries, Testing and Adjusting quantiles and ratings based on practical requirements, such as the need for more granular distinctions at higher sentiment levels or broader categories for neutral sentiments.

These diverse scaling methods provide a comprehensive approach to interpreting the sentiment scores from VADER and TextBlob. Each method offers unique advantages and will be evaluated to determine the most effective approach for transforming sentiment scores into user-intuitive star ratings.

D. Evaluation Metrics

To assess the effectiveness of our scaling methods, we will employ three key evaluation metrics: Accuracy, Mean Squared Error (MSE), and F1 Score. These metrics will provide a comprehensive understanding of how well the model-generated ratings align with the actual user ratings, which is critical for validating the accuracy and reliability of our sentiment analysis model.

- Accuracy: Accuracy measures the proportion of total predictions that were correctly classified. In the context of this study, it quantifies how often the model-generated ratings match the actual user-assigned star ratings. High accuracy indicates that the scaling method effectively translates sentiment polarities into user-understandable ratings, demonstrating the model's effectiveness in reflecting true user sentiments.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{6}$$

- Mean Squared Error (MSE): MSE is used to quantify the average squared difference between the actual ratings and the predicted ratings. It provides a measure of the magnitude of error in the predictions, with a lower MSE indicating closer alignment between model predictions and actual outcomes. By evaluating MSE, we can assess the precision of each scaling method and identify which method minimizes prediction errors, thereby enhancing the reliability of the sentiment analysis.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{7}$$

F1 Score: The F1 Score is a harmonic mean of precision and recall, providing a balance between the two by taking both false positives and false negatives into account. It is particularly useful when the class distribution is uneven, as is often the case with user ratings. A high F1 Score suggests that the scaling method not only accurately categorizes ratings but does so in a way that equally respects both positive and negative sentiment extremes.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

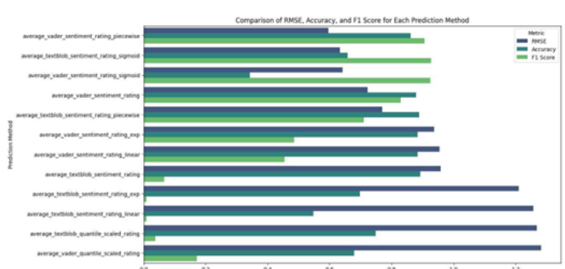

Figure I. Comparison of Evaluation Metrics between various Scaling models

IV. RESULT

This section presents the results obtained from applying the different scaling methods—linear, exponential, piece-wise linear, and quantile-based—to sentiment scores derived from VADER and TextBlob. It then discusses the effectiveness of each method in accurately translating sentiment analysis outputs into a user-friendly 5-star rating system.

A. Model Performance

Comparative Analysis: The performance of each scaling method was evaluated based on accuracy, Mean Squared Error (MSE), and F1 scores. The results indicated significant differences in how each method translated sentiment scores into star ratings:

- Linear Scaling showed a moderate level of accuracy and a relatively high MSE, suggesting that while it provided a direct and simple transformation, it might not handle extreme values and nuanced sentiments effectively.
- Exponential Scaling demonstrated higher accuracy and lower MSE compared to linear scaling. This method was particularly effective in distinguishing between neutral and extreme sentiments, reflecting the non-linear nature of human emotional expression in text.
- Piece-wise Linear Scaling yielded the best performance in terms of F1 score, particularly because it could adapt different segments of the sentiment range distinctly, providing a more tailored fit for the data.
- Quantile-based Scaling maintained consistent accuracy across all quantiles, making it robust against skewed data distributions. However, it was less sensitive to subtle differences in sentiment compared to piece-wise linear scaling.

Statistical Validation: Statistical tests confirmed that the differences in performance metrics between the scaling methods were statistically significant, reinforcing the observations made in the comparative analysis. The piece-wise linear approach, in particular, was found to significantly outperform the other methods in terms of balancing precision and recall across all classes.

The three metrics combine to form an Evaluation Score. Which suggests the best method of scaling the sentiment polarity to star rating.

$$E = ((1\text{-}RMSE)+Accuracy+F1)/3 \qquad (9)$$

Higher Evaluation Score means better performance of the Mode.

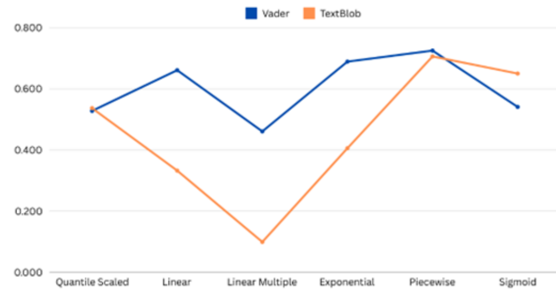

Figure II. Evaluation Score of different methods

B. Discrepancies between Sentiment Rating Stars and Actual Star Ratings

A notable finding from this research is the significant discrepancy observed between the sentiment-derived star ratings and the actual star ratings provided by users. This gap underscores several underlying challenges and complexities in sentiment analysis that can affect the accuracy and reliability of automated rating systems.

Causes of Discrepancies

- Subjectivity in Perception: Sentiment analysis tools rely on predefined lexicons and machine learning models that may not fully capture the subjective nuances individuals express in their reviews. People's expectations and experiences can significantly influence their ratings, which may not always align with the sentiment expressed textually.
- Sarcasm and Irony: Sarcasm and irony are prevalent in user reviews but notoriously difficult for automated systems to detect. A review might sarcastically use positive words to describe a negative experience, leading sentiment analysis tools to misinterpret the true sentiment.

- Contextual and Cultural Differences: Sentiment analysis tools may not effectively account for cultural and contextual variances in language use. Words or phrases that carry positive connotations in one context or culture might be neutral or negative in another.
- Overstatement or Understatement: Users might overstate or understate their sentiments in reviews, either due to emotional bias at the moment of writing or as a rhetorical strategy to influence other consumers. These exaggerations or downplays can skew the sentiment analysis results.
- Actual Star Rating Inconsistencies: Users do not always follow a consistent criterion for rating; a 4-star rating by one individual could be a 5-star by another based on personal rating habits or expectations. Additionally, some users might give a middle-of-the-road rating like 3-stars to a generally satisfactory product, even if their written review is largely positive, reflecting a mismatch between the textual sentiment and the star rating.

Addressing the Discrepancies

- Advanced NLP Techniques: Incorporating more sophisticated natural language processing techniques, such as context-aware sentiment analysis or models trained to detect sarcasm and irony, could enhance the accuracy of sentiment interpretations.
- Customized Sentiment Lexicons: Developing and employing customized sentiment lexicons that reflect the specific language, slang, and nuances of the target user base can improve sentiment detection's reliability.
- Hybrid Models: Combining sentiment analysis with other forms of text analysis, such as emotion detection or intent analysis, may provide a more holistic view of the user's opinions and lead to more accurate star ratings.
- User Education: Educating users on the importance of consistent and reflective star ratings can help align the numerical ratings more closely with the textual sentiment, reducing discrepancies.
- Feedback Loops: Implementing feedback loops where users can confirm the sentiment perceived by the system might allow for continuous improvement of the sentiment analysis models based on real-world usage and corrections.

C. Implications and Practical Applications

Limitations and Challenges: While the models performed well across various metrics, challenges remain, particularly in handling sarcasm and implicit sentiment, which are common in user-generated content. Further research is needed to integrate contextual and semantic analysis techniques that can better understand these complex expressions.

D. Future Work

Enhancing Sentiment Analysis Models: Future research could explore the integration of deep learning techniques that can capture contextual nuances more effectively. Additionally, testing the models on a wider array of text sources, including social media posts and professional reviews, could help generalize the findings.

Combating False Reviews: Given the initial success of the scaling methods in distinguishing nuanced sentiments, further exploration into their application for identifying and filtering out false or misleading reviews is recommended. This could involve developing a combined model that integrates sentiment analysis with anomaly detection techniques.

CONCLUSION

Our study has comprehensively evaluated various scaling methods—linear, exponential, piece-wise linear, and quantile-based—for converting sentiment analysis outputs into a standardized 5-star rating system. The comparative analysis revealed distinct strengths and weaknesses of each method, thereby providing valuable insights into their practical applications in sentiment-based rating systems.

A. Key Findings

- Exponential Scaling emerged as highly effective in handling extreme sentiments, thus enhancing the accuracy of ratings for highly polarized reviews.
- Piece-wise Linear Scaling demonstrated superior performance in terms of F1 scores, indicating its capability to accurately segment different sentiment intensities and align them with corresponding ratings.

- Quantile-based Scaling showed robustness in maintaining consistent accuracy across varied datasets, suggesting its utility in applications where data distribution is skewed or non-uniform.
- Linear Scaling, while straightforward, was less effective in capturing nuanced expressions of sentiment, often oversimplifying the translation of sentiments into ratings.

B. Implications

The findings of this study underscore the critical role of appropriate scaling methods in enhancing the reliability and relevance of sentiment-based rating systems. By choosing the right scaling approach, businesses and platforms can significantly improve how they interpret and utilize sentiment data to make informed decisions, tailor marketing strategies, and enhance customer interactions.

C. Future Directions

Given the complexities involved in accurately translating textual sentiments into numerical ratings, future research should explore the integration of more sophisticated natural language processing techniques. These could include context-aware models that better understand sarcasm, irony, and cultural nuances, which were identified as notable challenges in our study.

Additionally, the development of adaptive models that can learn from user feedback and adjust sentiment interpretations accordingly could pave the way for more dynamic and user-centric sentiment analysis tools. Exploring hybrid models that combine different scaling techniques could also offer a more nuanced approach to handling the diverse nature of sentiment data.

Moreover, further studies could focus on expanding the dataset to include multilingual reviews and cross-cultural contexts to test the scalability and adaptability of the proposed scaling methods on a global scale.

Our research contributes to the field of sentiment analysis by highlighting the importance of choosing suitable scaling methods to convert sentiment into actionable insights. The outcomes of this study not only advance our understanding of sentiment analysis applications but also open avenues for enhancing the precision and applicability of sentiment-based rating systems in real-world scenarios.

This expanded conclusion provides a comprehensive summary of your findings, the implications for practical applications, and outlines clear directions for future research, thereby concluding your study on a strong and forward-looking note.

V. APPENDIX

A. Data Collection Instruments

The code snippet for the following research can be found on the following project link https://github.com/raj-ryan/Development-of-a-Sentiment-Based-Rating-Model-using-VADER-and-TextBlob-

B. Raw Data

The open source dataset used for the research is attached https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera?select=Coursera_reviews.csv

ACKNOWLEDGMENT

We extend our gratitude to the faculty and peers from the Department of Computer Science at Vellore Institute of Technology, whose insightful feedback was invaluable throughout this research. Special thanks to Dr. Narayanamoorti M, whose guidance was instrumental in shaping the research methodology.

REFERENCES

[1] Medhat, W., Hassan, A., & Korashy, H. (2014). "Sentiment Analysis Algorithms and Applications: A Survey." Ain Shams Engineering Journal, 5(4), 1093-1113.

[2] Schouten, K., & Frasincar, F. (2016). "Sentiment Analysis in Twitter: A Review of Methods and Applications." IEEE Transactions on Knowledge and Data Engineering, 28(3), 747-766.

[3] Bonta, Venkateswarlu & Kumaresh, Nandhini & Naulegari, Janardhan. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer

Science and Technology. 8. 1-6. 10.51983/ajcst-2019.8.S2.2037.

[4] Mäntylä, Mika, Daniel Graziotin and Miikka Kuutila. “The evolution of sentiment analysis - A review of research topics, venues, and top cited papers.” ArXiv abs/1612.01556 (2016): n. pag.

[5] Jun Zhao, Kang Liu, and Liheng Xu. 2016. Book Review: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions by Bing Liu. Computational Linguistics, 42(3):595–598.

[6] Serrano-Guerrero, Jesus & Olivas, José & Romero, Francisco & Herrera-Viedma, Enrique. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences. 311. , 18–38. 10.1016/j.ins.2015.03.040.

[7] Giachanou, Anastasia & Crestani, Fabio. (2016). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. ACM Computing Surveys. 49. 1-41. 10.1145/2938640.

[8] Balahur, Alexandra. “Sentiment Analysis in Social Media Texts.” WASSA@NAACL-HLT (2013).

[9] Liu, Bing. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. 5. 10.2200/S00416ED1V01Y201204HLT016.

[10] Onan, Aytug. (2020). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. Computer Applications in Engineering Education. 29. 10.1002/cae.22253.

[11] Kirilenko, Andrei & Wang, Luyu & Stepchenkova, Svetlana. (2022). Sentiment Analysis: Gaging Opinions of Large Groups. 10.1007/978-3-030-88389-8_17.

[12] Pooja, and Rajni Bhalla. “A Review Paper on the Role of Sentiment Analysis in Quality Education.” SN computer science vol. 3,6 (2022): 469. doi:10.1007/s42979-022-01366-9

[13] M. S. Hossen, M. N. A. Chowdhury, A. M. Sristy and N. Jahan, "Sentiment Analysis using Machine Learning and NLP for Digital Education," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 902-908, doi: 10.1109/ICCMC53470.2022.9754065. keywords: {Sentiment analysis;Machine learning algorithms;Education;Machine learning;Classification algorithms;Sentimental Analysis;Digital Education;Natural Language Processing;Logistic Regression;4-Grams},

[14] Özen, İ.A. and Özgül Katlav, E. (2023), "Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels", Journal of Hospitality and Tourism Technology, Vol. 14 No. 2, pp. 102-120. https://doi.org/10.1108/JHTT-12-2020-0319

[15] Venkit, P. N., Srinath, M., Gautam, S., Venkatraman, S., Gupta, V., Passonneau, R. J., & Wilson, S. (2023). The Sentiment Problem: A Critical Survey towards Deconstructing Sentiment Analysis. In H. Bouamor, J. Pino, & K. Bali (Eds.), EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings (pp. 13743-13763). (EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings). Association for Computational Linguistics (ACL).

[16] S. J. Ordóñez, R. B. Medina and L. M. Herrera, "Sentiment Analysis to Assess Educational Methodologies in a Competency-Based Educational System," 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, Monterrey, Mexico, 2021, pp. 1-8, doi: 10.1109/IEEECONF53024.2021.9733755. keywords: {Sentiment analysis;Three-dimensional displays;Telepresence;Electronic learning;Education;Data science;Three-dimensional printing;Sentiment Analysis;Topological Data Analysis;Educational Innovation;Mathematics Teaching;Data Analysis Teaching},

[17] Bonta, Venkateswarlu & Kumaresh, Nandhini & Naulegari, Janardhan. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology. 8. 1-6. 10.51983/ajcst-2019.8.S2.2037.

[18] Fang, X., Zhan, J. Sentiment analysis using product review data. Journal of Big Data 2, 5

(2015). https://doi.org/10.1186/s40537-015-0015-2