Computer Science and Information Systems 21(4):1411–1433 https://doi.org/10.2298/CSIS240328040K

Sentiment Polarity Analysis of Love Letters: Evaluation of TextBlob, Vader, Flair, and Hugging Face Transformer

Gaganpreet Kaur[1,⋆], Amandeep Kaur[1], Meenu Khurana[2], and Robertas Damaševičius[3]

[1] Chitkara University Institute of Engineering and Technology
Chitkara University, Rajpura, Punjab
gaganpreet.kaur@chitkara.edu.in
amandeep@chitkara.edu.in
[2] Chitkara University School of Engineering and Technology,
Chitkara University,Baddi, Himachal Pradesh
meenu.khurana@chitkara.edu.in
[3] Vytautas Magnus University
robertas.damasevicius@vdu.lt

Abstract. Sentiment analysis is the task of computationally identifying and quantifying the emotions and opinions expressed in text. However, existing sentiment analysis tools, while increasingly sophisticated, face challenges when applied to complex and personal domains such as love letters. This study investigates the performance and accuracy of four popular Python libraries for sentiment analysis (TextBlob, Vader, Flair, and Hugging Face Transformer) in determining the polarity and intensity of sentiments in love letters. A corpus of 300 love letters was collected and randomly sampled to provide 500 sentences for analysis. Due to the lack of labelled data, human experts participated in evaluating the quality and accuracy of the sentiment annotations.Inter-rater agreements were computed among four judges across randomly sampled sentence lots in two distinct blind rounds. The results reveal varying degrees of effectiveness and agreement among sentiment analysis tools (TextBlob, Vader, Flair, and Hugging Face) and human judges, with Cohen's Kappa values showing low to moderate agreement (ranging from 0.09 to 0.77), and each tool demonstrating unique strengths—Vader excelling in sentiment intensity and Flair with Hugging Face better at contextual nuances—in handling the emotional complexity of the texts. The study also highlights limitations and proposes some custom metrics for evaluating sentiment analysis tools in the context of love letters, such as tenderness index, passion quotient, nostalgia score, and others. The findings contribute to the emerging field of sentiment analysis and provide insights for developing natural language models better suited for personal and emotionally charged domains.
Keywords: Sentiment Analysis, Polarity Analysis, Natural Language Processing.

1. Introduction

Opinion mining, or sentiment analysis, is the process of computationally analysing text to identify the dominant feelings that are conveyed. Sentiment analysis in writing is ex-

⋆ Corresponding Author

tremely important for a variety of applications, from social media interactions and consumer reviews to literary analysis and interpersonal communication [9, 16, 22, 23, 29, 30]. In 2002, the first publication on sentiment analysis using computer-based methods was published. It was based on product reviews, which it used an unsupervised Pointwise Mutual Information - Information Retrieval algorithm to categorize as thumb up or down for positive and negative sentiment. [26].In the process of automating sentiment analysis, researchers and practitioners have created and utilized a number of algorithms, each with specific advantages and disadvantages [18] [17]. Social networking sites have become essential resources for exchanging feelings with people all over the world as a result of the Internet's explosive growth. Many people use text, photos, music, and video to express their ideas or perspectives. However, text communication through network media on the web might be a little overwhelming. Every second, social networking websites produce a significant amount of unstructured data on the Internet. To comprehend human behavior, data analysis must begin as soon as it is generated. Sentiment analysis can assist with this by identifying polarity in messages [20] [10].

The study begins with a thorough overview of related literature, outlining the present state of sentiment analysis algorithms as we proceed on our comparison journey. The technique is then covered in detail, including the dataset that was used, the evaluation measures that were used, and the particular setups for each algorithm. The next sections provide a thorough examination of each algorithm, contrasting its advantages and disadvantages in relation to sentiment analysis. Lastly, we present conclusions based on our research, providing information about the algorithmic decisions that best satisfy the requirements of sentiment analysis in highly emotive material. By conducting this research, we hope to add significant knowledge to the emerging field of sentiment analysis and assist practitioners in choosing algorithms that are appropriate for their particular use cases [28] [19].

Existing sentiment analysis tools, while increasingly sophisticated, face challenges when applied to complex and personal domains such as love letters. This study investigates the responses and accuracy of four popular tools (TextBlob, Vader, Flair, and Hugging Face Transformer) in determining the sentiment of love letters, specifically focusing on sentences.

- Accuracy of Sentiment Categorisation: Can these tools effectively categorise love letter sentences into neutral, positive, negative, and compound sentiment categories with high accuracy?
- Human-based Evaluation: In the absence of readily available labeled data, how does the sentiment analysis of these tools compare to the judgments of human experts on love letter language?
- Tool Strengths and Limitations: What are the specific strengths and limitations of each tool when analyzing the emotional complexity of love letters?
- Applicability and Model Development: Can the insights gained from this study be used to guide the development of natural language models better suited for sentiment analysis in personal texts like love letters?

By addressing these questions, this study aims to establish the suitability of existing sentiment analysis tools for personal and emotionally charged domains, paving the way for more accurate natural language models in the future [31, 32].

The study makes contributions to the field of sentiment analysis by rigorously evaluating the performance of four prominent sentiment analysis tools—TextBlob, Vader, Flair, and Hugging Face Transformer—specifically within the context of love letters, a complex and emotionally charged type of text. It introduces novel metrics such as the tenderness index, passion quotient, and nostalgia score, which are designed to measure subtle emotional nuances more effectively. The findings highlight each tool's strengths and limitations, revealing a varying degree of accuracy and the ability to handle the intricate emotional content of personal correspondence. These insights are crucial for advancing the application of sentiment analysis technologies in personal and nuanced domains, providing a foundation for future enhancements and tool customizations tailored to specific textual analyses. '

The novelty of this study lies in its focused examination of sentiment analysis tools specifically within the domain of love letters—a unique and emotionally nuanced form of personal communication. This is the first study, to the best of the authors' knowledge, to introduce and apply custom metrics such as the tenderness index, passion quotient, and nostalgia score, which are specifically designed to assess and quantify the subtle emotional nuances found in such texts. These innovations address the gap in existing sentiment analysis research by tailoring evaluation techniques to the specific challenges of analyzing deeply personal and emotional content, thereby expanding the applicability and precision of sentiment analysis tools in handling complex, non-standard text datasets.

The paper has VI Sections. Sections I and II covers the Introduction and Literature review. Section III and IV shows the Methodology and Results Discussions on this research. Section V and VI describe the limitations and detailed conclusion of the analysis.

2. Literature Review

The paper compares sentiment analysis technologies using rule-based and machine learning methodologies, revealing that tool performance and strategy don't always impact accuracy, but low accuracy in small datasets necessitates further analysis. [2] [13]. In reviewing recent advancements in sentiment analysis across different domains and modalities, we see a variety of innovative approaches that leverage modern computational techniques to enhance performance and adaptability. Firstly, Li et al. [22] and Karayiğit et al. [16] both implement BERT-based models, but in distinctly different contexts—Li et al. focus on extracting complex aspect-category-opinion-sentiment relationships for implicit sentiment analysis, demonstrating a marked improvement in F1 and recall scores, whereas Karayiğit et al. apply BERT to analyze COVID-19 related sentiments on social media, achieving superior classification accuracy over traditional and other deep learning models. Yang et al. [33] introduce a tri-modal model that incorporates contrastive learning and a transformer architecture to optimize multimodal sentiment analysis. This approach addresses alignment and semantic discrepancies across text, audio, and visual data, showing notable improvements on benchmark datasets.

In the realm of recommendation systems, Gu et al. [8] and Rosewelt et al. [29] blend sentiment analysis with graph neural networks and optimization algorithms, respectively, to refine e-commerce recommendations based on user sentiment extracted from reviews, which both show substantial enhancements in recommendation accuracy. Gunasekar and

Thilagamani [9] focus on cross-domain sentiment analysis by developing a network that integrates aspect and domain-invariant features, resulting in a model that outperforms traditional approaches in adapting to new domains without loss of sentiment analysis accuracy. Shi et al. [30] and Liang et al. [23] both extend the capabilities of sentiment analysis models through innovative learning strategies. Shi et al. introduce an adaptive prompt-based method that utilizes contrastive learning to effectively handle few-shot learning scenarios, while Liang et al. develop a multi-channel model that integrates pre-training mechanisms to better capture nuanced textual information from user-generated content.

The field of sentiment analysis has seen significant advancements through the development and application of rule-based, machine learning, and hybrid methodologies.

Rule-based sentiment analysis systems rely on manually crafted rules to identify and classify sentiments based on the presence of certain keywords, phrases, or patterns that have predefined sentimental values. Liu and Haig [24] proposed the use of fuzzy rule-based systems for interpretable sentiment analysis, highlighting the approach's alignment with the inherent uncertainty of language and its contribution to interpretability. Hutto and Gilbert [11] introduced VADER, a parsimonious rule-based model designed specifically for social media text, showcasing its effectiveness by outperforming individual human raters and various benchmarks in sentiment analysis tasks. Rule-based approaches offer interpretability and straightforward implementation but may lack flexibility and comprehensiveness.

The advent of machine learning, especially deep learning, has significantly advanced sentiment analysis by automating the extraction of complex features from text. Do et al. [6] provided a comprehensive review of deep learning techniques for aspect-based sentiment analysis, underscoring their ability to capture both syntactic and semantic features without the need for manual feature engineering. Similarly, Zhang et al. [34] surveyed deep learning applications in sentiment analysis, illustrating the technique's success across various sentiment analysis tasks due to its sophisticated data representation learning capabilities. Machine learning methodologies, particularly deep learning, provide powerful tools for capturing the complexities of language and sentiment but can suffer from a lack of interpretability and require substantial computational resources. Traditional and deep learning approaches are directly compared in the works of Kapočiūtė-Dzikienė et al. [15] and Pratibha et al. [28], with the former testing these methods on Lithuanian text sentiment analysis and finding traditional techniques slightly outperforming deep learning methods, while the latter curates a multimodal dataset for analyzing sentiments in tweets related to war and peace, demonstrating the growing importance of comprehensive datasets for deep pragmatic analysis.

Hybrid systems aim to combine the interpretability and simplicity of rule-based approaches with the powerful feature extraction and prediction capabilities of machine learning models. Chikersal et al. [4] developed SeNTU, a system that integrates a rule-based classifier with supervised learning (specifically, Support Vector Machines) for Twitter sentiment analysis, demonstrating how rules can refine machine learning predictions. Prabowo and Thelwall [27] combined rule-based classification, supervised learning, and machine learning into a novel method that improves classification effectiveness, showcasing the potential of hybrid approaches in enhancing sentiment analysis performance. Hybrid systems emerge as a promising middle ground, leveraging the strengths of both approaches to achieve both high performance and interpretability in sentiment analysis tasks.

The discussed articles collectively illustrate the diverse applications of sentiment analysis in improving systems' ability to interpret human emotions accurately across various digital platforms and modalities.

3. Methodology

Based on the identified problems and aims of this study, In this section we explain the steps taken for obtaining them (Figure 3). The section begins explanation of how the dataset for this research was created.The next step is to preprocess the data so that it can be subjected to sentiment analysis using these four tools.

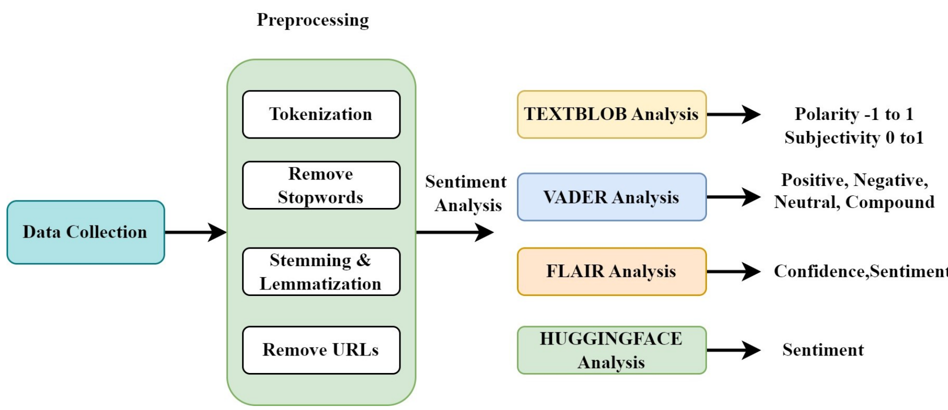

Fig. 1. Sentiment Analysis Flow

- **Dataset Collection**
  In this paper, we provide a carefully selected dataset of 300 love letters gathered from various sources, including well-known websites like Google and Quora. The wide and diverse spectrum of expressions of love captured in the dataset reflects the complex and multifaceted nature of this complex emotion. The dataset including these carefully prepared letters is available to the general public on Mendeley [DOI: 10.17632/rd5bjbnm35.3]. The dataset contains actual sentiments in each love letter, making it a vital resource for sentiment analysis research. We primarily assess how well four different Natural Language Processing (NLP) algorithms do in determining the sentiment conveyed in these love letters. This in-depth investigation seeks to further the field of sentiment analysis in the context of amorous emotions by shedding insight on how well these algorithms capture the complex intricacies of emotional language.
- **Preprocessing**
  For this research number of preprocessing steps were taken. It entails actions changing text to lowercase, tokenizing it, deleting punctuation, removing common words

(stop words), and stemming/lemmatization words to their simplest form. Managing numerals, special characters, contractions, HTML elements, and URLs are additional steps.The objective is to prepare text data for machine learning and sentiment analysis by cleaning and organizing it. [1].

– **Sentiment Analysis Tools** [5]
  Sentiment analysis tools known as software applications are used to examine text exchanges to determine the tone, intent, and emotion of each message. Businesses can gain from these technologies in a number of ways, such as more efficient feedback management, enhanced issue solving, higher-quality goods and services, well-informed business decisions, and real-time feedback analysis.

1. **TextBlob** TextBlob is a rule-based sentiment analysis tool that adheres to natural language processing principles. Unlike certain other sentiment analysis tools, TextBlob does not rely on pre-trained models but rather employs a rule-based methodology to determine sentiment in textual data. Even anyone without programming skills can use it because of its simple and intuitive design. Its primary benefit is its ease of integration with Python, despite its limited support for multiple languages, including English. TextBlob's sentiment analysis provides basic emotion identification and categorizes text as positive, negative, or neutral depending on its polarity. Its handling of denial and complex contextual nuances is fairly limited, and it might not be able to portray the subtle emotional nuances in more complex texts. [12].TextBlob is often used in academic contexts and small-scale projects where a quick and easy sentiment analysis is sufficient, even though it is lightweight and simple to use. Additionally, real-time applications are a good fit for it. The tool's robust documentation and active community further contribute to its success in certain applications.
2. **VADER** For use with social media text and short content segments, the rule-based VADER (Valence Aware Dic-tionary and Sentiment Reasoner) Sentiment Analyzer is designed. VADER is particularly good at processing sentiments expressed through colloquial language because of its rule-based architecture and pre-trained lexicon, which enable it to discern sentiment intensity with remarkable accuracy. Because it is specifically made to account for context and negations, this tool is perfect for sentiment analysis in short, passionate texts like tweets and online comments. VADER offers a compound score that accounts for the overall strength of the sentence in addition to polarity scores. [14].On a scale that goes from extremely negative to highly positive, emotions are grouped. VADER is good at capturing sentiment in real-time and doesn't require fine-tuning, but because it is rule-based, it might not be as suitable for longer or more formal writings. However, because of its ease of use and effectiveness in gathering sentiment subtleties, VADER is a popular choice for sentiment analysis tasks in social media and short text contexts.
3. **Flair** This machine learning-based sentiment analysis tool stands out for its advanced features and contextual embeddings. Rather than using rule-based techniques, Flair use pre-trained contextual embeddings to capture the semantic meaning of words and phrases in context. This helps to increase the understanding of sentiments by enabling Flair to recognize the intricate relationships that exist between words and their surroundings. Flair's multilingual features enable it to offer extensive language support for a large range of documents. Additionally, it provides the fine-tuning option, which lets users adjust the model to suit certain domains or activities.Because of its strength

in managing negation, context, and emotion recognition, it is helpful for sentiment analysis jobs in a range of scenarios, including those involving love letters with complex emotional expressions. However, it needs a lot of processing power, especially for real-time applications, and depending on the specific embeddings used, the model size may vary. Flair is well-liked in academia and business because to its sophisticated features and adaptability for sentiment analysis tasks. [25].

4. **Hugging Face** It is a platform that houses a variety of cutting-edge natural language processing (NLP) models, including transformer-based models like BERT, GPT, and many others, even if it isn't a dedicated sentiment analysis tool.Because of their remarkable results on a range of NLP benchmarks, these models—which are available via the Hugging Face Transformers library—have emerged as the preferred options for sentiment analysis assignments. Hugging Face transformer-based models are characterized by their capacity to represent intricate contextual relationships in text, which makes them ideal for sentiment analysis of a wide range of complicated content, including love letters. By using deep learning techniques, these models frequently perform better than conventional rule-based or lexicon-based approaches [15]. Hugging Face platform offers a library of pre-trained models that span numerous languages and domains. Additionally, it facilitates fine-tuning, enabling customers to customize the models to meet their unique requirements. Because of the platform's comprehensive documentation, user-friendly interface, and a large community of contributors to its ecosystem, it is widely used in both academics and industry. Hugging Face is useful for practitioners looking for state-of-the-art solutions for sentiment analysis tasks, especially those involving love letters with rich emotional content because it uses transformer-based models to handle complex linguistic nuances. [3].
   Table[1] shows some important features of the above four tools.

Table 1. Features of Tools

Tools	Algo. Logic	Supportive Lang.	Pre trained	Fine tuning	Emo Recog.	Industry Acceptance
TextBlob	Rule-based	Limited	No	NA	Simple	Academic, Small-scale
Vader	Rule-based	English	Yes	NA	Sentiment intensity	Brief texts & social media
Flair	Machine Learning	multiple lang.	Yes	Yes	Advanced	Business and academia
Hugging face	Transformer-based	multiple lang.	Yes	Yes	Refined	Research and Business

- **Sentiment Polarity Analysis of Love Letters**

Python functions were developed to compute the polarity of the sentiments as shown in Table 2, which is the short view of the output. It contains the sentiment scores of each sentence (Segment_Content). The descriptions are as follows :

This dataset is used for the analyses of text sentiment through multiple tools. Each "Segmented_Content" segment receives a polarity score (-1 to +1) from TextBlob (measuring negativity/positivity) and a subjectivity score (0 to 1) from the same tool (measuring opinionatedness). VADER further categorises text as Positive, Negative, Neutral, or with a ”Compound” score (-1 to +1) combining these proportions. Flair and Hugging Face models also provide sentiment analysis and confidence levels [2].

Table 2. Tool Polarity and Judge Score(Partial view)

Segment_Content	TB_Pol	TB_Sub	VD_Pos	VD_Neu	VD_Neg	VD_Comp	F_Sent	F_Confd	HF	J1_T1	J1_T2	J1_T3	J1_T4
Dear My Girl, Yes, I do comprehend that you are experiencing quite a bit of discomfort in your first workplace, ...	0.17	0.38	0.10	0.76	0.12	-0.22	N	0.99	N	1	1	0	1
People will not respect your time and will try to take your attention, most of the time without any malice in their hearts,....	0.12	0.42	0.09	0.87	0.02	0.67	N	0.99	N	1	1	0	1
It’s possible that you’ll have feelings of guilt because you didn’t follow through with what you had planned to do.	0	1	0	0.9	0.1	-0.27	N	0.99	N	1	0	1	1
However, you are not to blame because you are a human being.Everyone gets their attention diverted.	0	0.1	0.12	0.88	0	0.25	P	0.82	N	1	0	0	0
In many respects, the fundamental split is between those individuals who are able to get back on track	0.33	0.39	0.09	0.85	0.05	0.20	P	0.90	N	1	1	0	1
You are going to get interrupted, but you have the option to keep what you say short.	0	0	0.16	0.78	0.05	0.61	P	0.99	P	0	1	1	1

In summary, the following methodology (Figure 2) has been implemented to obtain the goals of this study.

1. **Tool Selection and Parameter Configuration:** Selected tools for evaluation include: TextBlob, Vader, Flair, and Hugging Face Transformer and configuration of the relevant parameters for each tool was done to ensure consistency and fairness in the analysis.
2. **Dataset Preparation:**
   – Collect a corpus of 300 love letters to serve as the dataset for analysis.
   – Randomly select 500 sentences from the love letters to provide a representative sample.
3. **Sentiment Analysis of Love Letters:** The TextBlob, Vader, Flair, and Hugging Face Transformer were used to analyze and categorize selected sentences into neutral, positive, negative, and compound sentiments.
4. **Human-based Evaluation:**
   – Engagement of human experts for two blind rounds to independently evaluate the sentiment of the same set of sentences.
   – Collection of annotations from human judges regarding the sentiment categories for comparison with tool-generated results.
   – Calculate Inter-Rater Agreements between the sentiment analysis tools and human judges using metrics Cohen’s Kappa and analysis of the level of agreement/disagreement between tools and human experts.
5. **Strengths and Limitations Assessment:** Examination of the specific strengths and limitations of each sentiment analysis tool.

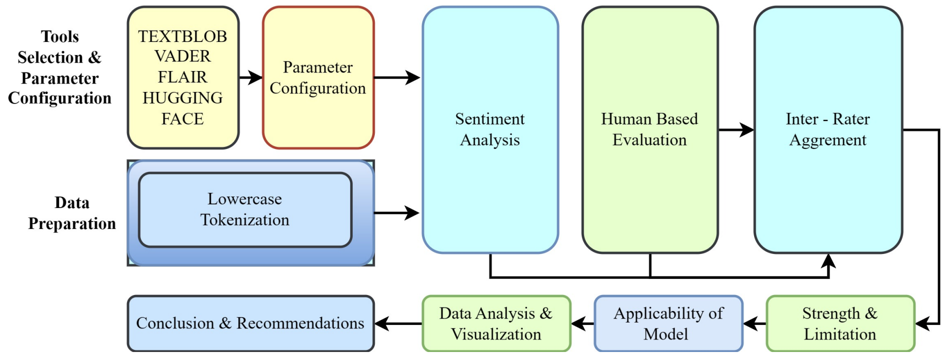

Fig. 2. Research Flow

6. **Applicability :** Based on the findings, derivation of the insights into the applicability of existing sentiment analysis tools for personal and emotionally charged domains like personal love letters, etc.
7. **Data Analysis and Visualization:**
   – Analyses of the collected data using statistical methods to derive quantitative insights.
   – Draw confusion matrices and histograms, to present the results in a comprehensible manner.
8. **Conclusion and Recommendations** for summarization of the study's findings, highlighting the suitability and limitations of sentiment analysis tools for love letters. In the next section, we discuss the outcome of the methodology applied.

4. Results & Discussion

Evaluating sentiment analysis tools in real-time NLP text data without a benchmark or a labeled dataset is hard due to the dependency of traditional supervised learning metrics on ground truth labels. However, other strategies can be used to approximate accuracy or evaluate performance, including Human expert evaluation. This involves using human annotators to assess the accuracy of sentiment predictions by comparing human annotations to the predictions made by each sentiment analysis tool. Hence, user feedback as a tool of evaluation was used in this research work. User feedback serves as an indirect measure of performance if sentiment analysis tools are integrated into an application or platform. Therefore, In this section, we discuss the outcome of the annotations that these tools assign to the segmented text of the love letters.

4.1. Objective Agreement Between the Tools :

This research work aims to evaluate sentiment analysis tools objectively and subjectively. We compare their agreement on polarity and relevance to human judgements in love letters. High agreement validates tool reliability, while disagreements expose limitations and biases, guiding model improvement. Sentiment can be complex, so disagreements may reflect natural language complexity and automated emotion capture challenges. By understanding these discrepancies, we can refine models, improve context understanding, and consider factors like irony or cultural references that influence sentiment. This section focuses on tool agreement levels on the same content.

Agreement of Random Samples (100) In this 100 sentences were randomly selected from a corpus of love letters to ensure an unbiased and representative analysis. This randomness is crucial as it prevents any sampling bias, guaranteeing that each sentence in the corpus has an equal chance of being chosen. Sentiment and polarity analyses were then conducted on these sentences using various Natural Language Processing (NLP) tools, including TextBlob, VADER, Flair, and Hugging Face. Each tool applies its methodology to evaluate the emotional tone of the text, yielding scores that indicate the sentiment's positivity or negativity. These scores were meticulously recorded in an Excel sheet for detailed analysis and comparison. The core of the study involved analysing the agreement and disagreement among these tools outcomes. This phase is essential to gauge the reliability and consistency of sentiment analysis methodologies. Statistical measures like Cohen's Kappa [21] have been used to quantify the agreement between the tool's sentiment assessments. The results of this analysis are critical; high agreement levels between tools could indicate their reliability, while significant disagreements might highlight the subjective nature of sentiment analysis or the diverse approaches of different tools. Ultimately, the study's structured approach, starting from the random selection of sentences to detailed sentiment analysis and inter-tool agreement assessment, offers a comprehensive understanding of the sentiments in the love letters.

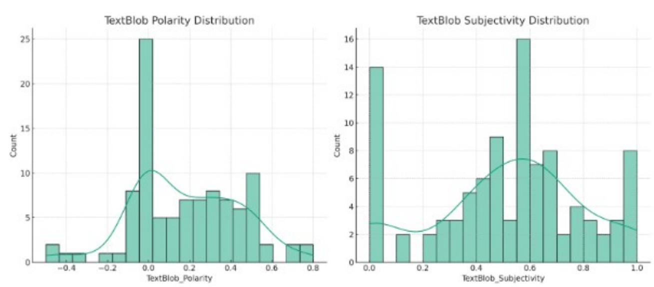

Fig. 3. Textblob Polarity and Subjectivity Distribution

- TextBlob Polarity: The polarity scores in Fig. 4 indicates that the majority of sentiments are positive, with a mean polarity of around 0.18. However, there is a wide distribution, with some sentiments being negative as well.
- TextBlob Subjectivity: The subjectivity scores in Fig 3. show a mean of about 0.51, suggesting a balance between objective and subjective sentiments in the dataset. The distribution is fairly uniform.
- VADER Scores: The VADER in Fig. 4 shows a higher distribution of neutral scores, with positive sentiments also being fairly common. Negative scores are less frequent,

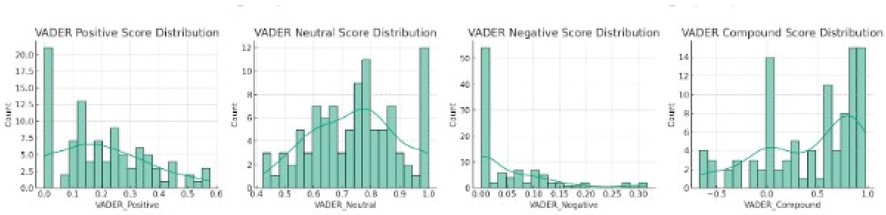

Fig. 4. Vader Score Distribution

and the compound score histogram reflects a skew towards positive compound sentiment.

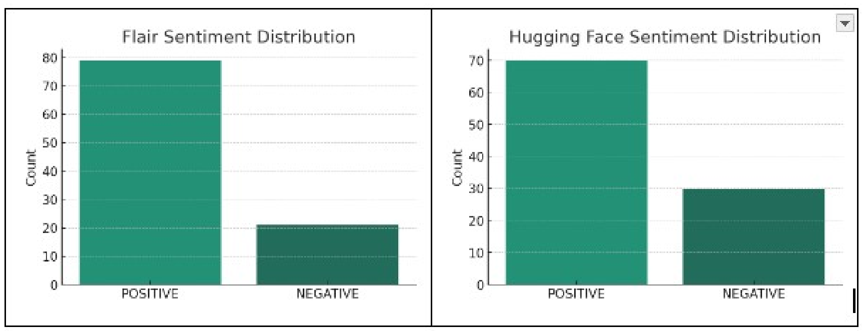

Fig. 5. Flair and Huggingface Distribution

- Flair Sentiment: The bar plot for Flair sentiment in Fig5. indicates that there is a predominance of one sentiment class over the other in the dataset.
- Hugging Face Sentiment: Similar to the Flair sentiment, the Hugging Face sentiment classifier in Fig. 4 shows a distribution skewed towards one sentiment class.

Using Cohen's Kappa score, the agreement with sentiment score from the four tools has been calculated.

The Cohen's Kappa Score Matrix (κ) is calculated to assess the agreement between two raters (or tools) who classify items into mutually exclusive categories. The formula for Cohen's Kappa is given by:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{1}$$

where: p_o is the relative observed agreement among raters, computed as the proportion of items upon which the raters agree. p_e is the hypothetical probability of chance agreement, calculated based on the probabilities of each rater randomly assigning items to categories. To calculate these probabilities, construct a $k \times k$ confusion matrix for k categories, where the element at the i^{th} row and j^{th} column (n_{ij}) represents the number of items Rater 1 classified into category i and Rater 2 into category j. Then, p_o is calculated as:

$$p_o = \frac{1}{N} \sum_{i=1}^{k} n_{ii} \tag{2}$$

where n_{ii} are the diagonal elements of the confusion matrix (where raters agree), and N is the total number of items. The probability of chance agreement, p_e, is computed as:

$$p_e = \sum_{i=1}^{k} \left(\frac{\sum_{j=1}^{k} n_{ij}}{N} \right) \left(\frac{\sum_{j=1}^{k} n_{ji}}{N} \right) \tag{3}$$

where the first term inside the summation is the proportion of items Rater 1 assigns to category i, and the second term is the proportion of items Rater 2 assigns to the same category. Cohen's Kappa score ranges from -1 (perfect disagreement) to +1 (perfect agreement), with 0 indicating that the observed agreement is exactly what would be expected by chance.

The outcomes are displayed in Table 4.1 and Fig. 6

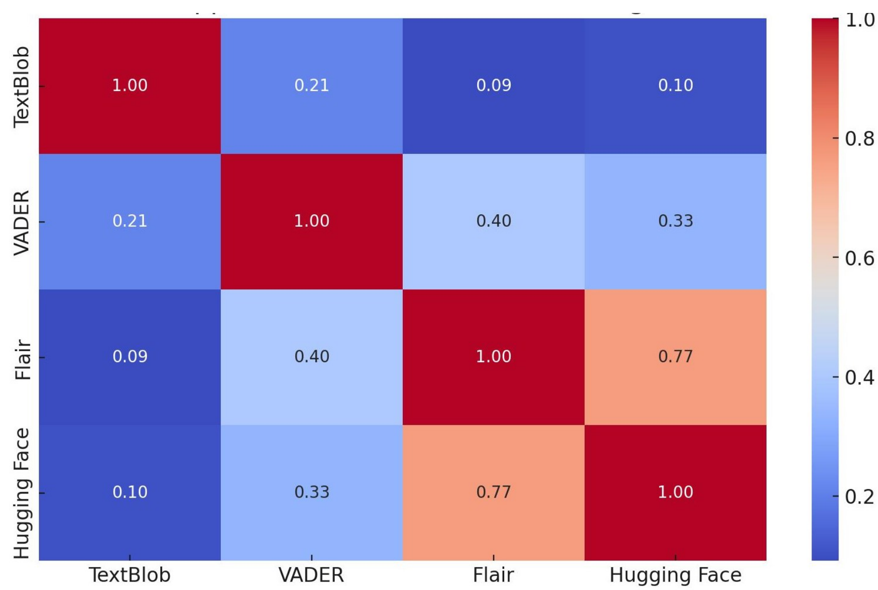

Fig. 6. Cohen's Kappa score Matrix for Sentiment Agreement

Agreement of Random Samples (200) As done before in the previous case of 100, 200 sentences were randomly selected from a corpus of love letters to ensure an unbiased and representative analysis for the second round of evaluation. In this case the following observations can be made.

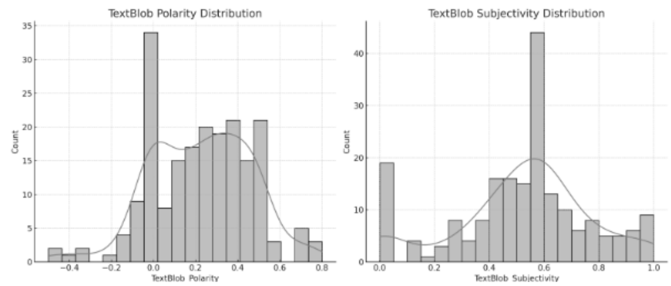

Fig. 7. Textblob Polarity and Subjectivity Distribution

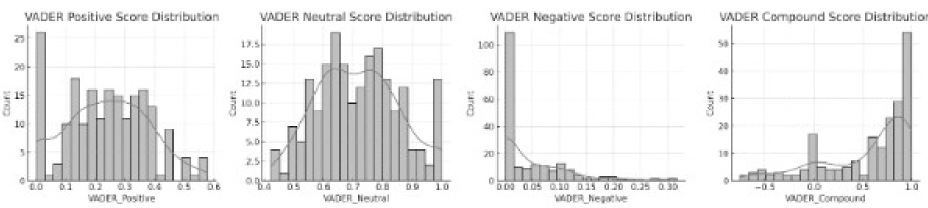

Fig. 8. Vader Score Distribution

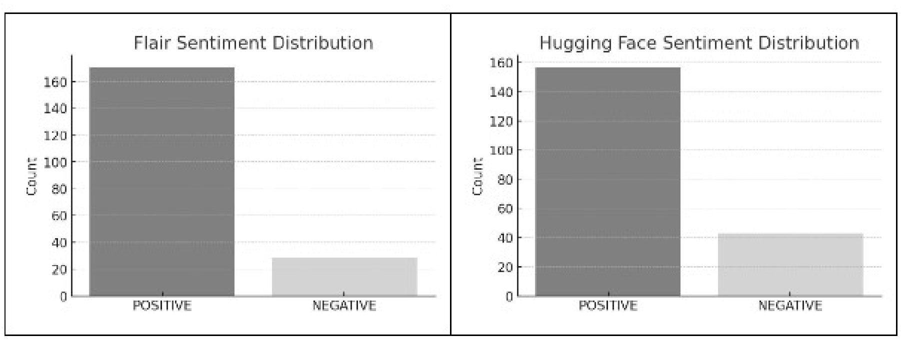

Fig. 9. Flair and HuggingFace sentiment Distribution

Table 3. Results of Cohen Kappa Score Matrix

TextBlob and VADER:	TextBlob and Flair	TextBlob and Hugging Face	VADER and Flair	VADER and Hugging Face	Flair and Hugging Face
K=0.21 K=0.21	K=0.09 K=0.09	K=0.10 K=0.10	K=0.40 K=0.40	K=0.33 K=0.33	K=0.77 K=0.77

- **TextBlob Polarity:** The polarity scores from TextBlob in Fig. 7 indicate a distribution with a slight skew towards positive sentiment.
- **TextBlob Subjectivity:** The subjectivity scores show a wide distribution, in Fig.7 indicating a mix of subjective and objective sentiments in the dataset.
- **VADER Scores:** The VADER sentiment analysis shows a variety of distributions for the positive, neutral, and negative scores. The compound score distribution indicates in Fig. 8 a skew towards positive sentiment overall.
- **Flair Sentiment:** The bar plot for Flair sentiment in Fig.9 shows the counts of each sentiment class.
- **Hugging Face Sentiment:** Similarly, the bar plot for Hugging Face sentiment in Fig. 9 shows the distribution of sentiment classes.
  The subsequent section of this analysis will focus on evaluating the extent of concordance between human expert judgments across different sentences from love letters.This examination is imperative to optimise and check the suitability of each tool's performance tailored to specific text genres such as love letter analysis.

4.2. Concordance Between Human Judges and Tools (Round 1 [100]) :

To analyse the concordance between human judges and tool scores, we will compute the correlation between the judges' ratings and the score of sentiments from the tools (TextBlob, Vader,and Flair). This will help us understand how well the sentiment analysis tools align with human judgement. For computations between the human judges themselves, we will calculate the correlation between the ratings of different judges to gauge the degree of agreement among them. This will give light on the consistency of human judgement across different individuals.

Concordance Between Human Judges and Tools Scores For this, the correlation between each judge's ratings and the sentiment score from TextBlob, VADER, and Flair is computed. Since the judge's ratings are binary, we will use methods suitable for binary data correlation with continuous data, such as point-biserial correlation (Figure 10).

- Correlations vary significantly across different judges and different tools in Fig.10 indicates that there is no consistent pattern of agreement or disagreement between human judges and automated sentiment analysis tools.
- TextBlob Polarity: Some judges show a weak to moderate positive correlation with TextBlob polarity, suggesting a slight agreement in assessing the overall sentiment (positivity/negativity) of the texts.

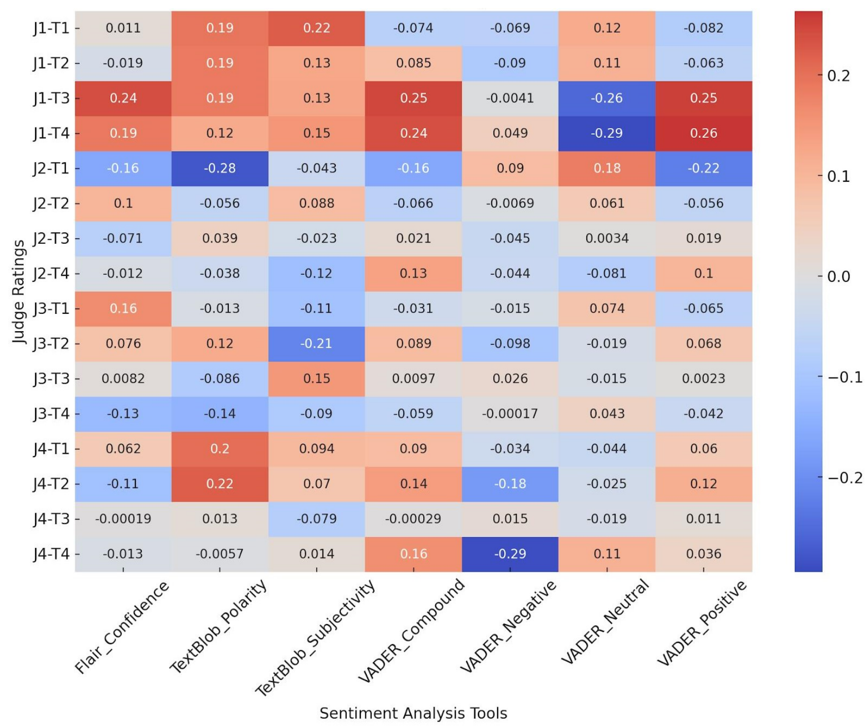

Fig. 10. Correlation between Human Judges and Tools

- VADER Scores: Correlations with VADER scores (positive, neutral, negative, compound) also vary. In some cases, there is a weak positive correlation, while in others, there is no significant correlation or even a weak negative correlation.
- Flair Confidence: The correlation with Flair's confidence in its sentiment predictions is generally low, indicating that the confidence level of the Flair model is not very aligned with human judgments.

Concordance Between Human Judges In this section, calculations regarding the correlation between the ratings of different judges are done (Figure 11) and for this correlating binary data, (Phi coefficient) is computed, [7] which is a measure of the association between two binary variables.

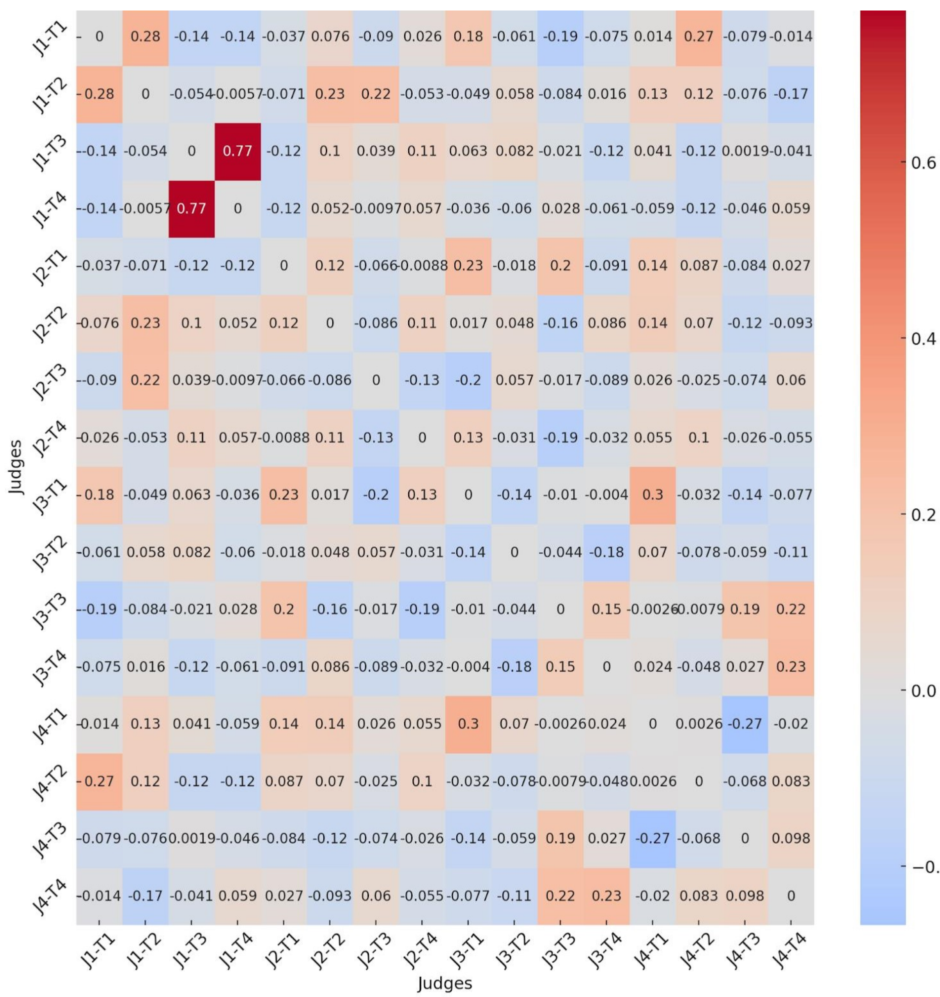

Fig. 11. Correlation between Human Judges

The Phi coefficient [7] values between pairs of judges are as follows:

- The correlations vary significantly, indicating differing levels of agreement between judges. However, in some cases, there is a moderate positive correlation, suggesting agreement between certain pairs of judges. At the same time, it is observed that many pairs of judges show low to negligible correlation, indicating a lack of consistent agreement in their ratings.
 item A few pairs show a high correlation in Fig.11 (e.g., 'J1-T3' and 'J1-T4' with a coefficient of 0.774), suggesting a strong agreement in their judgments and scoring.

4.3. Inferences from the Concordance Values

Concordance Between Judges and Tools

- There is no strong and consistent pattern of correlation between the judges' ratings and the sentiment analysis tool's scores. This suggests that there may be some level of agreement between human judges and automated tools, the concordance is not uniform across different tools or different judges.
- The variability in correlations suggests that sentiment analysis tools might interpret sentiment differently compared to human judgment, which can be due to the nuances and complexities of human language that automated tools may not fully capture.

Concordance Between Human Judges

- The varying levels of correlation between different pairs of judges highlight the subjective nature of sentiment analysis when performed by humans. Different judges may focus on different aspects of the text or interpret sentiments differently, leading to varying levels of agreement.
- The presence of both high and low correlations among judges pairs suggests that some judges may have similar criteria or thresholds for sentiment evaluation, others differ significantly in their judgement.

These findings demonstrate the challenges in sentiment analysis, both in terms of creating tools that can mimic human judgement and in achieving consistency among human evaluators. It highlights the need for clear guidelines and training for human judges and the importance of continuous improvement in sentiment analysis algorithms to better align with human interpretations.

4.4. Concordance Between Human Judges and Tools (Round II [200]):

In this round, 200 random sentences with their evaluations are analysed, similar to the previous round. It includes sentiment analysis scores from TextBlob, VADER, and Flair, along with evaluations from four judges.

Concordance Between Human Judges and Tools Scores in Round II The observations regarding the point-biserial correlation coefficients between the judge's ratings and the sentiment analysis tool scores in Round II are as follows:
It can be observed that generally, the Correlations are low, indicating a weak agreement between human judges and sentiment analysis tools. This is consistent across different judges and tools. TextBlob Polarity and Subjectivity: Some judges show a

slight positive correlation with TextBlob scores, but the correlation is generally weak. In VADER Scores, the Correlations with positive, neutral, negative, and compound scores vary and are generally low. In the case of Flair Confidence score analysis, the correlation with Flair's confidence level is also generally low, suggesting limited alignment with human judgments.

Concordance Between Human Judges in Round II The Phi coefficient values between pairs of judges in Round II are as follows:
The Variability in Correlations is similar to Round I, the correlations between judges in Round II vary, indicating differing levels of agreement. At the same time, it can be observed that some Moderate Correlations can be observed in case a few judge pairs show moderate positive correlations, suggesting agreement in their evaluations. Lastly, many Low Correlations can also be observed, clearly showing that many pairs of judges exhibit low to negligible correlation, indicating a lack of consistent agreement in their ratings. Based on the above observations, generally low correlations between human judges and analysis tools suggest that the tools' assessments of sentiment do not consistently align with human judgement. This might be due to the complexity of human language and sentiment, which automated tools might not fully capture.The variability in these correlations highlights the differences in how humans and machines interpret and evaluate sentiment. The variability in agreement among judges underscores the subjective nature of sentiment analysis. Different judges might prioritise different aspects of the text or interpret sentiments differently. The presence of both moderate and low correlations among judge pairs suggests varying criteria or thresholds for sentiment evaluation among different judges. The findings from Round II are inline with those from Round I, highlighting the challenges in achieving consistency in sentiment analysis, both among human evaluators and between human and automated tool assessments. This emphasises the guidelines for human sentiment analysis and continued improvement in automated sentiment analysis algorithms.
From the round I and round II of evaluation, it is abundantly clear that such tools are unsuitable for getting the sentimental value of the sentences embodied in the love letter. There is a necessity to customise and modify these tools to gain such availability.

5. Limitations and Future Applicability

Despite the significant advancements demonstrated in the reviewed sentiment analysis research, several limitations are evident, offering directions for future work. One recurrent challenge is the dependency on large, labeled datasets for training complex models. Although efforts like those described by Shi et al. [30] aim to mitigate this through adaptive prompt-based learning for few-shot scenarios, the generalization of such models across truly diverse or low-resource linguistic domains remains problematic. Moreover, while models such as those introduced by Yang et al. [33] improve multimodal sentiment analysis, the alignment of different modalities (text, audio, visual) and the effective fusion of such data without redundancy still pose considerable challenges.

The complexity of sentiment analysis models, especially those integrating deep learning and neural networks, often leads to issues with interpretability. Models like the RWESA-

GNNR proposed by Gu et al. [8] integrate sentiment analysis with graph neural networks to enhance recommendation systems, yet the interpretability of such models is not well addressed, which is crucial for trust and reliability in real-world applications. Furthermore, the performance of sentiment analysis tools across different cultural or demographic segments has not been thoroughly explored, which could affect the applicability and fairness of these tools in global or multicultural settings.

Future research could focus on developing more robust models that require fewer data yet are capable of high generalization across domains. This includes exploring unsupervised or semi-supervised methods that could leverage unlabeled data more effectively. There is also a pressing need for the development of models that provide better interpretability and explainability, which are essential for applications in fields such as healthcare, finance, and law where understanding model decisions is crucial. Additionally, enhancing the ability of sentiment analysis tools to handle multimodal data more efficiently, possibly through the development of novel neural network architectures or more sophisticated data fusion techniques, could significantly improve the accuracy and applicability of sentiment assessments. Lastly, the ethical implications and biases in sentiment analysis need rigorous examination to ensure these tools are fair and equitable across all user demographics.

Table 4. Sentiment Analysis Metrics for Love Letters

Metric	**Definition**	**Calculation**
Tenderness Index	A metric that measures the tool's accuracy in identifying and quantifying expressions of tenderness in love letters.	Evaluate the frequency and intensity of words associated with tenderness (e.g., gentle, affectionate, caring) in the sentiment predictions.
Passion Quotient	Assess the tool's effectiveness in capturing the passionate undertones within love letters.	Analyze the presence and intensity of words related to passion (e.g., fiery, intense, desire) in the sentiment predictions.
Nostalgia Score	A metric indicating how well the sentiment analysis tool captures nostalgic sentiments in love letters.	Consider the accuracy of identifying words and phrases associated with nostalgia (e.g., reminiscing, longing, memories) in the sentiment predictions.
Romantic Consistency	Measure the tool's ability to maintain sentiment consistency in segments of text expressing romantic feelings.	Evaluate the tool's performance in correctly predicting sentiment changes within the context of a romantic narrative.
Depth of Emotional Understanding	Assess the tool's capability to grasp the depth and complexity of emotions conveyed in the love letters.	Develop a metric that considers the tool's accuracy in capturing subtle shifts in sentiment intensity and emotion complexity.
Personalization Accuracy	Gauge how well the sentiment analysis tool adapts to and accurately reflects the unique personalization in individual love letters.	Evaluate the tool's precision in identifying sentiment nuances specific to the personal writing style and expressions of the letter.
Sentiment Diversity	Measure the variety and range of sentiments detected by each tool within the collection of love letters.	Assess the tool's performance in identifying a diverse array of sentiments, reflecting the multifaceted nature of love.

6. Conclusion

From the thorough analysis conducted during Round I and Round II evaluations, a clear pattern emerges, underscoring the inadequacy of existing sentiment analysis tools in capturing the nuanced sentimental value embedded within love letters. This revelation prompts a compelling argument for the imperative need to customise and modify these tools to enhance their suitability for discerning sentiments in the context of personal and emotionally charged texts. The key foundation of this argument lies in the examination of Phi coefficient values between pairs of judges in Round II. The Variability in Correlations, akin to Round I, persists in Round II, indicating divergent levels of agreement among the judges.

This variability is a crucial indicator of the challenges in achieving consensus, showcasing that different pairs of judges exhibit inconsistent levels of agreement in their evaluations.

A critical observation arises from the presence of Moderate Correlations in specific judge pairs, implying a degree of agreement in their assessments. However, the overarching trend reveals a multitude of Low Correlations, signifying that numerous judge pairs display low to negligible correlation. This discrepancy directly points to the lack of consistent alignment between human judges and sentiment analysis tools, implying that these tools' assessments do not reliably coincide with human judgments. The observed variability in correlations sheds insight on the complex structure of human language and sentiment, which automated tools may struggle to fully encapsulate. The disparities highlight the differing interpretations and evaluations of sentiment between humans and machines, emphasizing the nuanced and complex nature of emotional expression in love letters. Moreover, the consistency in findings between Round II and Round I accentuates the persistent challenges in achieving reliability and uniformity in sentiment analysis. This consistency underscores the ongoing struggle, both among human evaluators and between humans and automated tools, emphasizing the inherent difficulty in capturing the multifaceted emotional nuances present in love letters. Consequently, the argument strengthens the case for the necessity of clear guidelines for human sentiment analysis, facilitating a standardized approach and minimizing discrepancies among human judges. Simultaneously, it emphasizes the imperative for continuous enhancements in automated sentiment analysis algorithms. The presence of varying criteria and thresholds for sentiment evaluation among different judges, as suggested by the mixed correlations, reinforces the subjective nature of sentiment analysis and the need for further refinement in both manual and automated approaches. In essence, the argument emphasizes the pivotal role of customization and modification of sentiment analysis tools, drawing attention to the complex and subjective nature of sentiment analysis in the domain of personal and emotional expressions within love letters.

Table 5. Variable & Abbreviations

Variable	Abbreviations
Textblob_Polarity	TB_Pol
Textblob_Subjectivity	TB_Sub
Vader_Positive	VD_Pos
Vader_Neutral	VD_Neu
Vader_Negative	VD_Neg
Vader_Compound	VD_Comp
Flair_Sentiment	F_Sent
Flair_Confidence	F_Confd
Huggingface	HF
Judge	J
Tool	T

References

1. Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., Manicardi, S., et al.: A comparison between preprocessing techniques for sentiment analysis in twitter. (2016)
2. Borrelli, F.M., Challiol, C.: Comparing and evaluating tools for sentiment analysis (2023)
3. Chhabra, A., Chaudhary, K., Alam, M.: Exploring Hugging Face Transformer Library Impact on Sentiment Analysis: A Case Study. Auerbach Publications (2024)
4. Chikersal, P., Poria, S., Cambria, E.: Sentu: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. SemEval 2015 Task 10 pp. 647–651 (2015)
5. Dhiman, P., Kaur, A.: Text pre-processing techniques in addressing fake news. In: Computational Methods in Science and Technology: Proceedings of the 4th International Conference on Computational Methods in Science & Technology (ICCMST 2024), 2–3 May 2024, Mohali, India. vol. 1, pp. 142–148. CRC Press (2024)
6. Do, H.H., Prasad, P., Maag, A., Alsadoon, A.: Deep learning for aspect-based sentiment analysis: A comparative review. Expert Syst. Appl. 118, 272–299 (2019)
7. Fleiss, J.L.: Measuring agreement between two judges on the presence or absence of a trait. Biometrics pp. 651–659 (1975)
8. Gu, J., Xu, Y., Liu, W.: Rwesa-gnnr: Fusing random walk embedding and sentiment analysis for graph neural network recommendation. Information Technology and Control 53(1), 146 – 159 (2024)
9. Gunasekar, M., Thilagamani, S.: Improved feature representation using collaborative network for cross-domain sentiment analysis. Information Technology and Control 52(1), 100 – 110 (2023)
10. Gupta, D., Madhukar, M., et al.: Analysis of machine learning approaches for sentiment analysis of twitter data. Journal of Computational and Theoretical Nanoscience 17(9-10), 4535–4542 (2020)
11. Hutto, C.J., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media (2014)
12. Illia, F., Eugenia, M.P., Rutba, S.A.: Sentiment analysis on pedulilindungi application using textblob and vader library 2021(1), 278–288 (2021)
13. Ilyas, S.H.W., Soomro, Z.T., Anwar, A., Shahzad, H., Yaqub, U.: Analyzing brexit's impact using sentiment analysis and topic modeling on twitter discussion. In: The 21st Annual International Conference on Digital Government Research. pp. 1–6 (2020)
14. Joharee, I.N., Hashim, N.N.W.N., Shah, N.S.M.: Sentiment analysis and text classification for depression detection. Journal of Integrated and Advanced Engineering (JIAE) 3(1), 65–78 (2023)
15. Kapočiūtė-Dzikienė, J., Damaševičius, R., Woźniak, M.: Sentiment analysis of lithuanian texts using traditional and deep learning approaches. Computers 8(1) (2019)
16. Karayiğit, H., Akdagli, A., Acı, : Bert-based transfer learning model for covid-19 sentiment analysis on turkish instagram comments. Information Technology and Control 51(3), 409 – 428 (2022)
17. Kaur, G., Kaur, A., Khurana, M., et al.: A review of opinion mining techniques. ECS Transactions 107(1), 10125 (2022)
18. Kaur, G., Kaur, A., Khurana, M., et al.: A stem to stern sentiment analysis emotion detection. In: 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). pp. 1–5. IEEE (2022)
19. Kumar, A., Hooda, S., Gill, R., Ahlawat, D., Srivastva, D., Kumar, R.: Stock price prediction using machine learning. In: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). pp. 926–932. IEEE (2023)

20. Kumar, R., Sharma, C.M., Chariar, V.M., Hooda, S., Beri, R.: Emotion analysis of news and social media text for stock price prediction using svm-lstm-gru composite model. In: 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). pp. 329–333. IEEE (2022)
21. Kvålseth, T.O.: Note on cohen's kappa. Psychological reports 65(1), 223–226 (1989)
22. Li, J., Li, X., Du, Y., Fan, Y., Chen, X., Huang, D.: An aspect-category-opinion-sentiment quadruple extraction with distance information for implicit sentiment analysis. Information Technology and Control 52(2), 445 – 456 (2023)
23. Liang, S., Jin, J., Du, W., Qu, S.: A multi-channel text sentiment analysis model integrating pre-training mechanism. Information Technology and Control 52(2), 263 – 275 (2023)
24. Liu, H., Haig, E.: Fuzzy rule based systems for interpretable sentiment analysis. In: 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI). pp. 129–136 (2017)
25. Mahrukh, R., Shakil, S., Malik, A.S.: Sentiments analysis of fmri using automatically generated stimuli labels under naturalistic paradigm. Scientific Reports 13(1), 7267 (2023)
26. Marsanich, G.: Comparing Predictions of YouTube Video Like to Dislike Ratios Using Sentiment Analysis Tools. Ph.D. thesis, Tilburg University (2022)
27. Prabowo, R., Thelwall, M.: Sentiment analysis: A combined approach. J. Informetrics 3, 143–157 (2009)
28. Pratibha, Kaur, A., Khurana, M., Damaševičius, R.: Multimodal hinglish tweet dataset for deep pragmatic analysis. Data 9(2) (2024)
29. Rosewelt, L.A., Raju, D.N., Sujatha, E.: A new sentiment and fuzzy aware product recommendation system using weighted aquila optimization and grnn in e-commerce. Information Technology and Control 52(3), 617 – 637 (2023)
30. Shi, C., Zhai, R., Song, Y., Yu, J., Li, H., Wang, Y., Wang, L.: Few-shot sentiment analysis based on adaptive prompt learning and contrastive learning. Information Technology and Control 52(4), 1058 – 1072 (2023)
31. Tesfagergish, S.G., Damaševičius, R., Kapočiūtė-Dzikienė, J.: Deep learning-based sentiment classification in amharic using multi-lingual datasets. Computer Science and Information Systems 20(4), 1459 – 1481 (2023)
32. Tesfagergish, S.G., Kapočiūtė-Dzikienė, J., Damaševičius, R.: Zero-shot emotion detection for semi-supervised sentiment analysis using sentence transformers and ensemble learning. Applied Sciences (Switzerland) 12(17) (2022)
33. Yang, Z., Li, Z., Zhu, D., Zhou, Y.: Tri-clt: Learning tri-modal representations with contrastive learning and transformer for multimodal sentiment recognition. Information Technology and Control 53(1), 206 – 219 (2024)
34. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (2018)

Gaganpreet Kaur received her M.sc degree from Guru Nanak Dev University India,M.Tech degree from Guru Nanak Dev University India, and doing Ph.D. degree from Chitkara University, India. She has 12+ years of experience in academia. She is currently working as JRF in Chitkara University, India. She has published more than 10 research papers in reputed journals and conferences. She has 2 patents granted and published. Her primary research areas encompass machine learning, Deep Learning, Data Mining, NLP and artificial intelligence.

Amandeep Kaur currently holds the position of a Professor at the Chitkara University Institute of Engineering and Technology, Chitkara University Punjab. She earned her doctorate degree from I. K. Gujral Punjab Technical University, Jalandhar. Dr. Kaur's academic achievements include receiving both her M.Tech (Computer Science and Engineering) and B.Tech (Computer Science and Engineering) degrees with distinction. Additionally, she has successfully qualified for the UGC-NET in Computer Science.Dr. Amandeep Kaur boasts an extensive research portfolio, with approximately 100 publications in renowned international journals and fully refereed international conferences. She has accumulated 24 years of valuable experience in her field and has filed and published more than 107 patents.Moreover, Dr. Kaur has played a significant role in mentoring the academic growth of over 30 Ph.D. and PG students. Her primary research areas encompass medical informatics, machine learning, IoT (Internet of Things), artificial intelligence, and cloud computing.Notably, Dr. Amandeep Kaur has been recognized for her exceptional contributions, winning the Excellence Award in the "Filing Patent" category for three consecutive years (2021, 2022, and 2023) and the Best Ph.D. Supervision Award in 2023.Furthermore, she has achieved recognition on a global scale, as she is included in Stanford University's prestigious list of the top 2

Meenu Khurana received her B.E. (Honors) degree from Punjab Engineering College (PEC University) India, M.E. in Computer Science from Panjab University, Chandigarh India and Ph.D. degree from Chitkara University, India. She has 28+ years of experience in industry and academia. She is currently working as Professor Pro Vice Chancellor in Chitkara University, India. She is a Senior Member of IEEE, Professional Member of ACM, Life Member of ISTE.She has published more than 80 research papers in reputed journals and conferences. She has 12 patents granted and 3 published. She has won best paper awards in various conferences few of them are IMEIS 2022, ICAC3N 2021, NCASM 2020 and more.She has been keynote speaker at various conferences, the recent ones are ICAETC, 2023, India; ICADEIS 2023, Indonesia. She has been Technical Program Committee member and session chair at conferences of repute, few of them are SmarTechCon 2023, IEEE DELCON 2022, ICAN 2022. Dr. Meenu Khurana served as judge in Smart India Hackathon in 2022. She is the guest editor for special issue of SN Computer Science. Her research area includes vehicular adhoc networks, fog computing, MIMO systems, MU-MIMO, data deduplication, machine learning algorithms. She has delivered expert sessions and conducted workshops in the area of vehicular adhoc networks, MIMO systems, MU-MIMO, 5G Networks in VANETs.

Robertas Damaševičius received the Ph.D. degree in informatics engineering from the Kaunas University of Technology, Lithuania, in 2005. He is currently a Professor with the Department of Software Engineering, Kaunas University of Technology and the Department of Applied Informatics, Vytautas Magnus University, Lithuania, and an Adjunct Professor with the Faculty of Applied Mathematics, Silesian University of Technology, Poland. He also lectures software maintenance, human–computer interface, and robot programming courses. He is the author of more than 500 articles and a monograph published by Springer. His research interests include sustainable software engineering, human–computer interfaces, assisted living, and explainability. He is also the Editor-in-Chief of the Information Technology and Control journal.

Received: March 28, 2024; Accepted: July 29, 2024.