

Katydid Inspired BERT Optimization for Improved Sentence Semantics in Text Classification

¹Prativa Mishra, ²Gopalakrishnan T and ³Parameswaran T

^{1,2}Department of Information Technology, Manipal Institute of Technology Bengaluru,
Manipal Academy of Higher Education, Manipal, India.

³Department of Computer Science and Engineering, School of Engineering and Technology, CMR University,
Bengaluru, Karnataka, India.

¹prativa.mitblr2023@learner.manipal.edu, ²gopalakrishnan.t@manipal.edu, ³parameswaran.t@cmr.edu.in

Correspondence should be addressed to Gopalakrishnan T : gopalakrishnan.t@manipal.edu

Article Info

Journal of Machine and Computing (<https://anapub.co.ke/journals/jmc/jmc.html>)

Doi: <https://doi.org/10.53759/7669/jmc202606009>

Received 16 April 2025; Revised from 30 August 2025; Accepted 06 October 2025

Available online 16 October 2025.

©2026 The Authors. Published by AnaPub Publications.

This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Abstract – The proposed Katydid-Inspired BERT Optimization for Improved Sentence Semantics in Text Classification (KyBERT), aims to optimize the ability of the model to identify more complex linguistic patterns along with long-range dependencies. Bio-inspired optimization on KyBERT, where the inspiration is drawn from the katydid's sensory capabilities, enhances better handling with context-rich and resource-dense language. Generally these models enhance the values of key metrics like classification accuracy, F-measure, and Fowlkes-Mallows Index (FMI) since sentence semantics was poorly represented in simple models. This proposed model KyBERT has also confirmed that useful in creating a much-refined framework, that rarely gives false predictions and significantly improves text classification rate. The evaluation based on the Amazon Reviews dataset with 476,001 unique IDs established a high performance of text classification with 64.29%, F-measure of 64.67% and FMI of 64.68%.

Keywords – Sentence Semantics, KyBERT, Katydid Optimization, BERT, Text Classification, Sentiment Analysis, Bio Inspired Models.

I. INTRODUCTION

Online shopping has transformed the way trading occurs among humans, bridging geographical distances and offering unparalleled convenience. Unlike traditional shopping, which demands physical visits to stores, online platforms provide access to a wide variety of products from different parts of the world [1]. Traditional shopping often involves time-consuming activities, like waiting in lines or traveling to malls, while online shopping allows browsing and purchasing with a few clicks. This transformation that arises as evidence of human trade where marketplaces have transformed from a small, local environment to the huge space, which is popularly known as digital world. Online shopping offers competitive prices, customer reviews, which really helps in taking decisions with lot of options. The progression towards shopping from physical to digital is a significant step toward human trade and re-making consumerism in times ahead. [2].

Consumer reviews now become a part of modern trade since they summarize the experiences of previous consumers and guide potential consumers. In Traditional shopping, feedback was mainly passed through the word of mouth in their personal known circle [3]. Online platforms have enabled a global exchange of opinions, offering transparency in the purchasing process. Reviews don't just tell one about the quality of the products, but also generate trust between the seller and the buyer. Today's marketplace decision-making requires this virtual “word-of-mouth”. [4]. Human trade always relied on trust and experiences, similarly consumer reviews have become a solid basis of influence on purchasing behavior. The change from personal to global reviews is one fundamental change which represents the change in modern commerce[5].

A sentence is the foundation of communication, as it carries meaning and illustrates the thoughts clearly. Communicating complex ideas and emotions along with providing valuable information is bound within a well-formed sentence [6]. Sentence semantics refers to the deeper meaning behind a sentence, beyond just its grammatical structure. It tries to understand the interaction of words and how meaning is constructed in context. Therefore, semantics forms an integral part in the interpretation of every word uttered within a sentence [7]. Misunderstanding sentence semantics can lead to confusion or misinterpretation, which shows its importance in effective communication.

The study of sentence semantics enables a deeper understanding on how the meaning is constructed, this impacts in learning to develop a language and everyday conversation, this makes a way for clear and accurate communication in human interaction [8].

Sentiment analysis has emerged as a powerful tool in understanding emotions, opinions, and attitudes expressed in written text. Its importance lies in helping businesses and individuals gauge public opinion in regards to some product review or social media posting, or even in customer feedback [9]. By categorizing text as positive, negative, or neutral, sentiment analysis provides valuable insights into consumer behavior and market trends.

Text classification plays a broader role in organizing and labeling written content into predefined categories. This enables the filtering process on enormous amounts of data efficiently, which deals with the activities like sorting of emails to the categorization of articles [10]. Sentiment analysis and text classification transform raw text into actionable insights that can empower better decision-making across sectors. It has reshaped human interaction with digital content, refining the way in which information is understood and processed. The blend of emotional understanding with structured categorization enhances the clarity and relevance of modern communication [11].

Traditional language models depends on sequential processing of words and predicting the next word based on previous inputs [12]. These models often struggle with capturing long-range dependencies in language, affecting the understanding of complex sentences. Bidirectional Encoder Representations from Transformers (BERT) has transformed this approach to the next step by considering context from every direction, which increases the ability of understanding the accurate meaning of the context. BERT improves the ability to perform such tasks as question answering, text classification, and sentiment analysis much more accurately but also appears to better reflect even more nuance in understanding the language [13]. The change from traditional models to BERT marks a significant advancement in natural language processing, which improves contextual understanding in human communication [14].

Bio-inspired computing draws inspiration from natural systems to solve complex problems in computing[15]. Mimicking biological processes, such as neural networks, evolution, and swarm intelligence, this approach enhances optimization, learning, and problem-solving techniques[16]. It plays a vital role in various fields, including robotics, artificial intelligence, and data analysis, by providing adaptive and efficient solutions[17]. These algorithms mirror nature's ability to evolve, self-organize, and adapt to dynamic environments. Bio-inspired computing represents a fusion of biology and technology, shaping the future of computational models, while addressing challenges that traditional algorithms struggle to resolve, fostering innovation in human-designed systems[18].

Problem Statement

The real challenge in sentiment analysis is the interpretation of the semantics of sentences, which requires capturing nuances of context and tone and subtle linguistic patterns. More frequently, classical models put themselves in a situation like they cannot find a balance between true positive and false positives; hence they end up showing significant inconsistencies in classification accuracy in complex resource-rich language environments. This inconsistency impacts the ability to precisely detect sentiment, with gaps in real-time applications such as customer feedback analysis or social media monitoring. The need to optimize models for better sentence-level semantic comprehension has become critical, especially with growing data complexity. Improving models are capable of handling the long-range dependencies, minimizing false predictions, and adapting variations in sentence structures is a problem in advancing systems that provide sentiment analysis.

Motivation

This sections explains the motivation for the proposed work :

- The motivation for this work stems from the increasing need for more accurate sentiment analysis in real-time applications, such as customer feedback systems, social media monitoring, and market analysis.
- Current models often struggle with complex sentence structures and subtle shifts in tone, leading to unreliable sentiment classification. These limitations can impact decision-making processes in businesses and other sectors that rely on sentiment-driven data insights.
- The growing complexity of language, coupled with the rising volume of unstructured data, emphasizes the need for models that can better understand sentence semantics. This has driven the pursuit of developing optimized frameworks capable of capturing deeper linguistic nuances and delivering more precise sentiment analysis across diverse real-world scenarios.

Objective

This research has based on the following objectives:

- To develop a model that will help to understand the complex linguistic data and improving the accuracy of sentiment analysis by optimizing sentence-level semantics. This involves developing a model capable of capturing complex linguistic patterns, long-range dependencies, and subtle contextual cues within the text.
- To enhance the model's ability to reduce false predictions, especially in ambiguous or resource-dense language structures, thereby increasing overall classification accuracy. This extends to the enhancement on key metrics, such as F-measure and the Fowlkes-Mallows Index (FMI), into a more reliable model performance assessment. Generally

this should be in the direction of developing a strong framework that can satisfactorily comprehend nuanced sentiment in the heterogeneous yet real-world applications.

Paper Organization

Section 2 describes some of the literature review done in this area. Section 3 describes the importance of Optimized Katydid inspired BERT Model. Section 4 describes about dataset. Section 5 describes our implementation and results. Finally, we conclude in section 6.

II. LITERATURE REVIEW

“Emotion-Cognitive BERT” [19] integrates emotional and cognitive reasoning into the BERT model to enhance sentiment analysis of online public opinions during emergencies. The key contribution lies in combining emotion detection with cognitive understanding to interpret complex sentiments expressed in crisis situations. The working mechanism adds modules to the basic BERT architecture that capture emotional cues and patterns in cognitive processes from textual data. It is hence able to evaluate more than just the superficial sentiment but to evaluate deeper emotional states and thought processes reflected in language. “Imbalanced BERT” [20] discusses the deep-learning-approached sentiment analysis problem on imbalanced datasets by using techniques such as BERT and ensemble stacking. In particular, it focuses on developing a method to handle class imbalance, which is common in real-world data. The working mechanism thus involves leveraging BERT for powerful text representations and then applying ensemble stacking of multiple classifiers to improve predictive performance, especially on minority classes.

“Stream ETL” [21] provides a framework that uses big data technologies to perform real-time sentiment analysis of Twitter data. Its most significant contribution is a high-velocity optimized stream extract, transform, and load (ETL) system. Working mechanism of the system is continually extracts tweets, transforms the content using sentiment analysis algorithms, and then loads the results into a storage system for subsequent analysis. By using the big data tools and technology, the given framework can cope with the high volume and velocity of Twitter data. “Unsupervised Attention Aspect Analysis” [22] is an aspect-based customer review sentiment analysis related to an unsupervised attention neural framework. The key contribution to the work is providing an unsupervised model in which aspect and sentiment identification takes place with unlabeled data. The working mechanism is guided by an attention mechanism of the neural network to place proper emphasis on the relevant words or phrases that might be relevant to certain aspects with their corresponding sentiments. “GUDN” [23] a new guide network that introduced label reinforcement strategy for the purpose of extreme multi-label text classification. The core novelty was an innovation in Guide Unified Deep Network (GUDN) that handled very large texts attached with a high number of labels efficiently. It went through the operation of guiding the learning process using the strategy of label reinforcement, which captures and leverages label relations to further help the model understand and predict complex label dependency. “Patent SBERT” [24] proposes a deep NLP-based hybrid model for patent distance calculation and classification using an augmented SBERT. The key contribution is adapting SBERT for the specialized domain of patents, thus enhancing its ability to comprehend technical as well as legal languages. The working mechanism involves training SBERT on patent data to develop embeddings that adequately reflect semantic similarities between patents.

“Refined SBERT” [25] Manifold Representation explores representing Sentence-BERT embeddings in manifold space to improve semantic representation. The key contribution is with regards to refining sentence embeddings through a mapping of SBERT embeddings to a manifold that might capture the geometric nature of semantic relationships much better. In working mechanism, the manifold learning techniques were applied to the embeddings produced by SBERT to enhance its ability to capture complex nuances in semantic representations. “CrisisTransformers” [26] utilizes pre-trained language models and sentence encoders to process crisis-related social media texts. The model effectively handles large-scale, unstructured data during crisis events, focusing on filtering and analyzing critical information. The use of pre-trained models enhances the understanding of crisis-specific vocabulary, improving the analysis of social media posts in real time. “BlendCSE” [27] introduces a contrastive learning technique for the representation of sentence embeddings focus on rich semantic representation and high transferability across tasks. The model with contrasts between multiple learning methods is to obtain rich contextual understanding towards sentence embeddings with better transfer learning for different natural language processing tasks.

“Semantic Syntax” [28] integrates semantics, syntax, and external knowledge for aspect-based sentiment analysis. It combines deep semantic understanding with syntactical structures and leverages domain-specific knowledge, enabling more precise extraction of sentiment related to specific aspects within texts, enhancing the depth of sentiment analysis. “Graph Sentiment” [29] utilizes graph convolutional networks to learn from word semantics to sentence syntax for aspect-based sentiment analysis. The model builds a graph representation of text, allowing for the structured analysis of relationships between words and aspects. This framework enhances sentiment extraction by focusing on syntactical and semantic dependencies. “Lexicon Marathi” [30] introduces a lexicon-based approach for sentence annotation in aspect-oriented sentiment analysis, specifically for Marathi movie reviews. By employing a rich sentiment lexicon, the framework annotates sentences for sentiment polarity, enhancing the detection of aspect-related opinions in regional language datasets with limited resources.

“Parental Sentiment” [31] analyses in-depth media coverage at a national level on parental leave reform. It uses natural language processing to conduct sentiment, semantics, and contributor analysis of content and traces the dynamics of how contributors instill societal perceptions of the policy reforms through media texts. “Triple GAT” [32] introduces a triple-

channel graph attention network for enhancing aspect-level sentiment analysis. This model includes three distinct channels that capture abundant aspect information; it makes use of the attention mechanism in the process to be able to weigh relevant aspects. It enhances the extraction of sentiments tied to certain aspects in complex textual data. “Quantum Tensor” [33] proposes a quantum-like tensor compression model for sentence representation, by using functions of constraints to analyze the semantics. This proposed model compresses information at the level of sentence by tensor decomposition with the necessary semantic features preserved. “CSMF-SPC” [34] proposes a multimodal sentiment analysis model that incorporates the semantic context across various modalities along with sentiment polarity correction. To combine the diversified data sources, the framework leverages an effective combination of visual, textual, and contextual features. The sentiment polarity correction mechanism further refines the sentiment predictions for each mode towards the development of a strong solution based on sentiment analysis. “Twain-GCN” [35] uses twain-syntax graph convolutional networks for aspect-based sentiment analysis. The model constructs a dual graph structure that combines syntactical information, allowing detailed examination of relationships between words and aspects. This graph-based approach enhances understanding of sentiment polarity related to specific aspects, promoting more nuanced sentiment extraction.

“EBERT” [36] introduces an enriching ensemble model with features for Arabic word sense disambiguation; an extension of BERT through added linguistic features and statistical analysis. Contextual cues and linguistic patterns are infused into the model to enhance the disambiguation of words, especially those from the Arabic vocabulary that are rich in their morphological structures. Optimized data collection methods are used to gather training data for the above process to ensure robustness and minimize noise within the data set. Ensemble Approach This approach uses an ensemble of several classifiers. It tries to exploit the strengths of each and hence enhance word sense prediction with the collective power of all. “ACB” presents a structured approach to tackling linguistic complexities in aspect-based sentiment analysis. The framework systematically breaks down sentences, capturing semantic and syntactic components related to specific aspects. By organizing the analysis into distinct stages aspect identification, context extraction, and sentiment polarity determination the model offers a streamlined and coherent approach. Incorporating structured feature representation allows Aspect Struct to address diverse linguistic challenges, especially in complex or ambiguous reviews. This structured pipeline enhances the interpretability and accuracy of sentiment analysis, focusing specifically on the sentiments linked to product or service aspects.

III. OPTIMIZED KATYDID INSPIRED BERT MODEL (KY-BERT)

BERT is the latest advanced model in natural language processing, developed by Google. As opposed to most models, BERT takes the context of a word from both sides of the sentence instead of only one side of the sentence that is used to interpret text in traditional models. This two-way understanding enables BERT to gain more deeply insightful language meaning. It depends on the architecture of the transformer and uses attention mechanisms greatly. This mechanism pays much attention to the important words or parts of a sentence and their relationship to other parts of the sentence. As a result, BERT can handle complex language tasks, including text classification, sentiment analysis, and question answering. The functional phases of Traditional BERT are as follows.

- **Tokenization:** The first step in BERT is breaking the input text into smaller units called tokens. BERT uses WordPiece tokenization, which breaks words into subwords or individual characters to handle rare words or unfamiliar vocabulary.
- **Masking:** BERT has used the Masked Language Modeling (MLM) technique. In this process, some words in the text are randomly replaced by a [MASK] token. The model learns to predict masked words based upon their context, and that helps it understand which words relate with each other.
- **Next Sentence Prediction (NSP):** BERT is trained on predicting whether one sentence logically follows another. The model takes in two sentences and evaluates whether the second sentence really is the next sentence. Such a feature trains BERT on tasks dealing with pairs of sentences, including text classification and question answering.

Fig 1. Depicts the overall structure of traditional BERT.

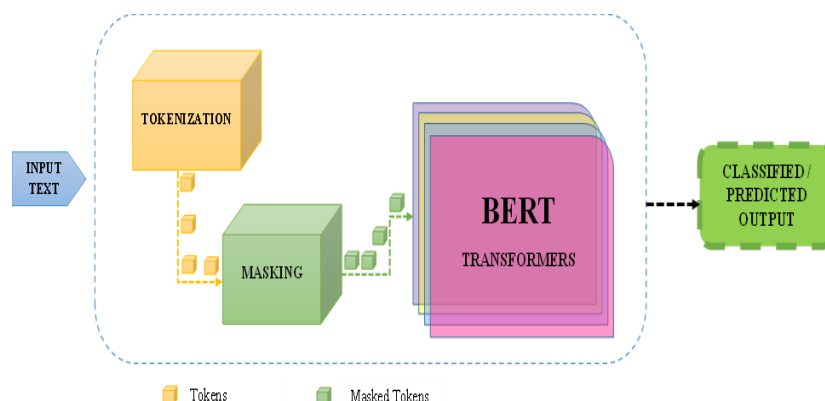


Fig 1. Traditional BERT Structure.

- *Training on Large Corpora:* BERT is pre-trained on huge datasets like Wikipedia in an unsupervised manner where patterns and structures of the language are learned.
- *Fine-Tuning:* Pre-trained BERT is highly adaptable to different natural language processing tasks without requiring massive changes through fine-tuning. BERT, after being pre-trained, can be fine-tuned for specific purposes. For instance, sentiment analysis can be performed by adding the classification layer on top of BERT and fine-tuning it with labeled data.

BERT's structure enables it to excel in understanding complex language tasks with more accurate predictions and contextual relevance.

Traditional BERT focuses on understanding language by considering context from both directions, improving tasks like text classification, question answering, and sentiment analysis. Inspired by the auditory system of katydids, Katydid-inspired BERT (KyBERT) to enhance the sensitivity of the network to even the finest linguistic patterns in the same way that katydids detect specific frequencies in their environment. KyBERT focuses more on enriching the attentional model from BERT through biologically-inspired mechanisms, where even further complex structures in language can be recognized even more precisely. This new approach toward natural language processing makes it model the sound-filtering and -processing mechanism of the katydid wing to capture even more nuanced issues in text-based interactions.

Contextual Camouflage

Contextual Camouflage within KyBERT takes inspiration from the katydid's natural ability to blend into its surroundings for survival, allowing the model to seamlessly integrate subtle and intricate meanings within varying contexts. This mechanism enhances KyBERT's capability to interpret complex, ambiguous sentences, ensuring more accurate text comprehension. In a sentence, context often determines the meaning of individual words. The same word can carry different meanings depending on its surroundings. This context sensitivity becomes critical in complex text structures, where multiple meanings coexist. Traditional models typically process each word individually, potentially leading to a loss of the overall semantic flow. The Contextual Camouflage layer in KyBERT avoids this issue by encoding contextual dependencies directly into the structure of the language model. Let a word embedding W represent the raw, unprocessed word. Traditional models would process this as $f(W) = W$, losing critical contextual cues. KyBERT, however, encodes the word W within its context, modeling the transformation as shown in Eq.(1).

$$C(W) = f(W, C_1, C_2, \dots, C_n) \quad (1)$$

where $C_1, C_2, \dots, C_n$ represent the surrounding words that contribute to the word's meaning. This equation highlights the transformation where the meaning of the word W is no longer isolated but fused with the surrounding words C_n .

The camouflage concept of katydids works by blending into their environment, not standing out, yet still retaining essential characteristics that allow them to survive and communicate. In KyBERT, contextual embedding takes a similar approach, embedding words in such a way that they maintain their relevance in the text while blending seamlessly into the surrounding information. KyBERT leverages this property to handle cases where multiple meanings might be possible. By embedding words contextually, the system learns to weigh possible meanings and select the most likely one. The contextual transformation relies heavily on encoding the relationship between each word in a sentence and its neighboring words. This relationship can be modeled as represented in Eq.(2).

$$T(W_i) = W_i + \sum_{j=1}^n \alpha_{ij} C_j \quad (2)$$

where W_i represents the word of interest, C_j the contextual neighbors, and α_{ij} the attention weight given to each contextual word. This enriches the representation of W_i to include relevant contextual information, thus allowing it to "camouflage" within its sentence as katydid blend into its surroundings without being easily noticed by predators. This kind of meaning will not just be derivable from intrinsic properties of the word but will be highly derived from the contextual environment. Through this process, the word W_i becomes a complex interaction of its immediate surroundings, much like how the katydid adapts its form to match the leaves around it.

The Contextual Camouflage in KyBERT also introduces the concept of dynamic weighting for each context by reflecting the proximity-relevance of nearby words. Each word C_j contributes differently in determining the meaning of W_i , depending on its position and usage in the sentence. This dynamic weighting can be represented as Eq.(3).

$$T(W_i) = W_i + \sum_{j=1}^n \beta_{ij} f(C_j) \quad (3)$$

where β_{ij} represents the dynamic weighting assigned to each word, which reflects its impact on the word W_i . This shows the ability that the system possesses to adapt according to changes in its surroundings just like the katydid adapts to its environment.

This process allows KyBERT to understand “polysemous words”, a word which have multiple meanings depending on their context. Contextual Camouflage is used by KyBERT for dynamically deciding which meaning would be most appropriate within a given sentence. This mechanism is very important for understanding complex text, where meaning is not always clear from the individual words alone but is used based on the overall sentence structure. The blending of word meanings into their context also allows for a more sophisticated understanding of ambiguous sentences, where traditional models might struggle. This process can be expressed mathematically through the vector transformation of each word into its contextual representation as shown in Eq.(4).

$$V(W_i) = \gamma \cdot W_i + (1 - \gamma) \cdot T(W_i) \quad (4)$$

where $V(W_i)$ represents the final vectorized form of the word after contextual embedding, and γ is a blending factor that controls the balance between the word’s original meaning and its contextual meaning. This blending ensures that each word maintains its identity while seamlessly merging into its surrounding context, optimizing for nuanced comprehension. Contextual Camouflage allows KyBERT to effectively process sentences where meaning shifts based on context.

Frequency-Driven Attention

This Frequency-Driven Attention step is inspired by the katydid’s ability to detect certain frequencies that help it in the recognition of important linguistic signals in the text. With this mechanism, it will be able to pay special attention to certain patterns or sequences of words similar to how species-specific sound frequencies are accorded higher priority by the katydids. This allows for KyBERT to identify and focus its attention on key structures of language in the process of improving the interpretation of heavy text. Frequency-driven attention in KyBERT can therefore be mathematically modeled towards capturing the key information in terms of frequency at which certain words or phrases appear in a particular context. This enables the model to give greater weight to some elements while making a certain hierarchy of importance with regards to the text. The system is optimized based on patterns in language use most relevant to the task by focusing on frequent linguistic patterns. The function for determining word frequency can be expressed mathematically in Eq.(5).

$$F(W) = \frac{\text{count}(w)}{\sum_{i=1}^n \text{count}(w_i)} \quad (5)$$

where $F(W)$ represents the frequency of the word w , which is calculated as the ratio of the word’s occurrence to the total occurrences of words in the input sequence. This formula creates a frequency distribution across the text, is very similar to how the katydid’s ear picks up sounds at different intensities, focusing on critical signals.

Once word frequency is established, KyBERT applies an attention mechanism to allocate importance based on this frequency. Frequency-driven attention assigns more weight to words or patterns that appear frequently and are likely to contribute meaningfully to the overall context. This weighting process is represented by the attention function as mathematically expressed in Eq.(6).

$$A(w_i) = \alpha \cdot F(w_i) + (1 - \alpha) \cdot R(w_i) \quad (6)$$

where $A(w_i)$ is the attention score for word w_i , α is a scaling factor controlling the balance between frequency $F(w_i)$ and relevance $R(w_i)$. The relevance $R(w_i)$ is derived from how closely the word matches the expected patterns within the context, allowing for finer adjustments to the weighting system.

Just like the katydid, which shifts its attention on the respective frequencies, KyBERT adjusts its attention weights with respect to both the frequency and contextual importance dynamically. Words having high frequency irrelevant in context obtain less attention, while words have lower frequency are highly relevant are emphasized. The balance of frequency and relevance ensures the model does not weight too much on those common and captures the critical content. With respect to the fine-tuning of frequency-driven attention, KyBERT adds in a temporal dimension to the process of weighting, which may realize that some words become more important as a sentence or paragraph unfolds. This temporal adjustment is very similar on how katydids respond to the changing patterns of sounds. The attention score evolves dynamically which is depend upon the context, which can be represented mathematically in Eq.(7).

$$A_t(w_i) = \beta_t \cdot A(w_i) + (1 - \beta_t) \cdot C_t \quad (7)$$

where $A_t(w_i)$ represents the time-adjusted attention score, β_t is a time-dependent scaling factor, and C_t accounts for the contextual shifts. This adjustment ensures that attention is recalibrated based on the evolving meaning of the text, reflecting the katydid’s ability to respond to environmental changes.

The system combines a frequency modulation factor to account for varying levels of importance across different linguistic structures. This factor, similar to the frequency shifts detected by katydids in their environment, which adjusts the weighting of different sentence components, allowing for nuanced attention distribution. This modulation process can be expressed mathematically in Eq.(8).

$$A_f(w_i) = A_t(w_i) \cdot (1 + \gamma \cdot \log(F(W_i))) \quad (8)$$

where $A_f(w_i)$ is the frequency-modulated attention score, and γ is the modulation factor that amplifies the attention given to particularly significant words based on their frequency. This amplification replicates the behaviour on how katydids focus on particular sound frequencies to enhance communication or detect threats, ensuring that key patterns within the text receive heightened attention. This step ensures that KyBERT captures essential linguistic signals in a manner similar to the katydid's auditory processing.

Ultrasonic Language Detection

This phase takes a direct inspiration from the katydid's ability to detect ultrasonic frequencies in its environment, which is essential for the survival and communication. This particular characteristic ensures that katydid's hears sounds beyond human senses of auditory range, at the same time it also permitting it to be aware on nearby predators or to communicate through delicate patterns of sound waves. This translates to a heightened sensitivity to pick up on rare and commonly overlooked linguistic signals, particularly ones that traditional models might not recognize as a result of their subtlety or infrequency.

In complex texts, especially domain-specific content or rare linguistic patterns, some words or structures carry important meaning but don't appear very often. Traditional models tend to ignore these signals, which are basically hidden in the noise, with common or high-frequency words captured. Ultrasonic Language Detection in KyBERT captures low-frequency but high-importance signals, ensuring that important albeit subtle information is processed equally prioritarily as a more frequent term. The basis of this step starts off by finding sparse linguistic cues, which can be quantitatively described using their inverse frequency as depicted in Eq.(9).

$$U(w) = \frac{1}{F(w)} \quad (9)$$

where $U(w)$ represents the ultrasonic importance of a word w , and $F(w)$ is its frequency in the text. Words with a lower frequency will have a higher $U(w)$, which clearly shows that these are the rare signals that need greater attention.

KyBERT then adjusts its attention to these ultrasonic signals by applying an attention weighting that emphasizes these low-frequency words. The attention mechanism for ultrasonic detection can be represented mathematically using Eq.(10).

$$A_u(w_i) = \alpha_u \cdot U(w_i) + (1 - \alpha_u) \cdot A(w_i) \quad (10)$$

where $A_u(w_i)$ represents the ultrasonic adjusted attention score for word w_i , then α_u being a scaling factor that controls the balance between the ultrasonic importance $U(w_i)$ and the regular attention score $A(w_i)$.

Contextual relevance in ultrasonic detection refines through evaluating such rare signals within the proper context. The relevance of the rare word within its sentence or paragraph can be modeled as represented in Eq.(11).

$$R_u(w_i) = U(w_i) \cdot C(w_i) \quad (11)$$

where $R_u(w_i)$ is the ultrasonic relevance, and $C(w_i)$ represents the contextual weight of the word w_i .

The next step contains dynamic weighting, where KyBERT adjusts its attention on ultrasonic signals based on the context and task requirements. This dynamic adaptation replicates that how the katydid adjusts its behavior based on ultrasonic cues, whether to avoid a predator or continue in communication. This dynamic adjustment is expressed mathematically in Eq.(12).

$$A_d(w_i) = \beta_u \cdot A_u(w_i) + (1 - \beta_u) \cdot A(w_i) \quad (12)$$

where $A_d(w_i)$ represents the dynamically adjusted attention score, β_u is the influence of ultrasonic detection, and $A(w_i)$ is the traditional attention. This dynamic balancing allows KyBERT to shift attention toward rare signals without overwhelming the broader comprehension of the text.

KyBERT integrates these ultrasonic signals into a broader representation of the text. This integration is essential to ensure that subtle linguistic patterns are seamlessly combined into the overall understanding. The integrated signal can be represented mathematically in Eq.(13).

$$I(w_i) = \gamma \cdot R_u(w_i) + (1 - \gamma) \cdot A_d(w_i) \quad (13)$$

where $I(w_i)$ represents the integrated attention score that combines ultrasonic relevance and dynamic attention, γ represents the blending factor. This integration ensures that rare but critical signals are combined into the overall comprehension process, enhancing KyBERT's ability to interpret complex, nuanced language.

Directional Context Recognition

This Directional Context Recognition is modeled after the highly advanced directional hearing of the katydid. Katydids have the special ability to identify the exact direction from which a sound emanates. This skill is used by katydids both for communication and predator evasion. In KyBERT, this ability means it can successfully detect the origin and flow of meaning within the text, where words and sentences will be understood based on the directional flow of context. Meaning

in complex sentences or documents often takes place based on the arrangement or sequence of the words. Words occurring earlier or later in a sentence can influence how a sentence would be interpreted as a whole. In Directional Context Recognition meaning is assigned based on how the context flows from one word to another, thereby making the meaning dynamically change as the sentence progresses, capturing nuance in the language. The first step in Directional Context Recognition involves calculating the directional weight of each word based on its position within a sentence. This weight can be modeled as shown in Eq.(14).

$$D(w_i) = \sum_{j=1}^i \alpha_j \cdot C_j \quad (14)$$

where $D(w_i)$ is the directional context of the word w_i , and C_j is the contextual influence of each preceding word, scaled by the weighting factor α_j . This captures the flow of meaning in the unfolding sentence, enabling it to give weight to the direction from which the context comes.

With direction context captured, KyBERT now evaluates how meaning shifts between words and ties attention to directional influence. This step can be mathematically expressed with the directional attention function as shown in Eq.(15).

$$A_d(w_i) = \beta_d \cdot D(w_i) + (1 - \beta_d) \cdot A(w_i) \quad (15)$$

where $A_d(w_i)$ is the attention score adjusted for directional context, β_d controls the balance between directional influence and standard attention $A(w_i)$.

KyBERT also incorporates forward and backward context recognition, so that it identifies meaning from words preceding and also from words that follow. This bidirectional approach replicates the katydid's unique behaviour which is to detect sounds from various directions. The forward context can be represented mathematically in Eq.(16).

$$F(w_i) = \sum_{k=i+1}^n \gamma_k \cdot C_k \quad (16)$$

where $F(w_i)$ represents the forward directional context for word w_i , with C_k is the contextual influence of each subsequent word and γ_k is the forward weighting factor. The combination of forward and backward context ensures that meaning is comprehended from both directions.

The final directional context for each word is derived from the combination of both backward and forward context, which can be represented mathematically in Eq.(17).

$$D_f(w_i) = \delta \cdot D(w_i) + (1 - \delta) \cdot F(w_i) \quad (17)$$

where $D_f(w_i)$ represents the final directional context score, and δ controls the balance between backward and forward contextual weighting. This combination ensures that KyBERT recognizes meaning holistically, with full consideration of how each word contributes to the overall message.

Context Embedding Strategy

Context Embedding Strategy parallels the katydid's precise egg-laying strategy, where eggs are carefully placed within plant tissue or soil, ensuring protection and successful growth. This embedding strategy takes form by embedding words in a structured manner within their context, ensuring that each word remains deeply connected to its surrounding information. This enables the model to extract and preserve nuanced meanings, otherwise might be ignored by traditional embedding techniques. To begin with, each word is initially represented as a basic embedding vector $E(w_i)$, which lacks contextual depth. The raw form can be expressed as Eq.(18).

$$E(w_i) = f(w_i) \quad (18)$$

where $f(w_i)$ denotes the non-contextual word embedding function for word w_i . This raw representation serves as the foundation for further transformations. KyBERT's approach embeds words into context by blending the word's intrinsic meaning with the information provided by surrounding words.

To capture the contextual embedding, KyBERT uses a weighted sum of surrounding words. The context embedding for a word w_i is represented mathematically in Eq.(19).

$$C(w_i) = \sum_{j=1}^n \alpha_{ij} \cdot E(w_j) \quad (19)$$

where $C(w_i)$ represents the context embedding of the word, α_{ij} is the attention weight between word w_i and surrounding word w_j , and n is the number of words in the surrounding context window.

This embedding strategy reflects the katydid behaviour to lay eggs into safe environments. KyBERT integrates each word deeply within its context, ensuring that its meaning is defined and preserved by the surrounding words. The model adjusts the embedding based on relevance, computed through attention scores dynamically for each surrounding word. The

next step is refinement of the embedding, capturing the directionality in the flow of context, similar to what katydid exploits with their directional hearing. Meaning is captured in the flow of a sentence. The forward context embedding can be modeled mathematically as shown in Eq.(20).

$$F(w_i) = \sum_{k=i+1}^n \beta_{ik} \cdot E(w_k) \quad (20)$$

where $F(w_i)$ represents the forward context embedding, β_{ik} is the forward attention weight. This forward embedding enables the model to capture how upcoming words influence the meaning of w_i , ensuring that future context contributes to the overall embedding. Simultaneously, the backward context embedding is calculated as shown in Eq.(21).

$$B(w_i) = \sum_{l=1}^{i-1} \gamma_{il} \cdot E(w_l) \quad (21)$$

where $B(w_i)$ represents the backward context embedding, γ_{il} is the backward attention weight. This backward embedding ensures that past words in the sentence are accounted for when determining the word's meaning.

KyBERT combines both forward and backward context embeddings to create the final context rich embedding for each word. This combined embedding can be represented mathematically in Eq.(22).

$$E_c(w_i) = \delta \cdot F(w_i) + (1 - \delta) \cdot B(w_i) \quad (22)$$

where $E_c(w_i)$ is the final context embedding, and δ is the weight controlling the balance between forward and backward contexts. This combination mirrors the katydid's strategy of embedding its eggs in environments that ensure long-term success, similarly embedding each word within the broader textual environment to preserve and enhance its meaning.

The context embedding strategy is further fine-tuned to specific task requirements such that the embeddings are optimized toward achieving the desired outcome. The task-specific adjustment can be represented mathematically in Eq.(23).

$$E_t(w_i) = \theta \cdot E_c(w_i) + (1 - \theta) \cdot T \quad (23)$$

where $E_t(w_i)$ is the task-specific embedding, θ is the task-weighting factor, and T represents the task-specific vector influencing the word's embedding. This step ensures that the embeddings are tailored to the specific task, aligning the context with the desired objective.

Nocturnal Language Processing

This process of KyBERT draws directly on the nocturnal behavior of a katydid. As it has enhanced sensory capabilities, the insect thrives on low-light settings. Hence, this step in KyBERT was intended for language comprehension in "low-information" scenarios, for which data might be less, ambiguous, or incomplete. Because of mimicking the katydid's adaptation to limited visibility, the process of KyBERT enhances its text processing ability with only partial contextual clues or fragmented data and situations wherein the meaning is not apparent.

While traditional language models rely on huge datasets or rich contextual information to be processed properly, Nocturnal Language Processing evolves a method of handling these low-light conditions in textual data so that KyBERT can draw meaning, even with ambiguity and lacking information. This is critical in real-world applications, such as legal documents, medical records, or historical texts, where complete clarity might not always be available.

The initial step in this process is assigning a confidence score to each word based on its surrounding context. In low-resource settings, KyBERT calculates this score by considering how available and relevant the context around the word is, as expressed in Eq.(24).

$$S(w_i) = \frac{\sum_{j=1}^n C_j}{n} \quad (24)$$

where $S(w_i)$ is the confidence score of the word w_i , C_j is the contextual relevance of the surrounding word w_j , and n is the number of words in the surrounding window. This score determines how well a word can be understood given the available context.

In low-information environments, where such confidence scores may be low, KyBERT would rely on a fallback mechanism to ensure proper interpretation. The fallback mechanism would pull in learned patterns from larger datasets and similar contexts to fill in the gaps of understanding. The fallback weight $F(w_i)$ is calculated as shown in Eq.(25).

$$F(w_i) = \lambda \cdot P(w_i) + (1 - \lambda) \cdot S(w_i) \quad (25)$$

where $F(w_i)$ represents the fallback weight, $P(w_i)$ is the word's learned pattern probability from the larger dataset, and λ is a scaling factor that balances learned patterns and contextual confidence. This mechanism allows KyBERT to utilize prior knowledge to compensate for missing data, much like how the katydid relies on heightened sensory abilities to navigate in the dark.

Nocturnal Language Processing Introduces Estimation Mechanism for Missing or Incomplete Data. In texts containing gaps or ambiguity, the KyBERT calculates the plausible meaning of a word or phrase using surrounding context as well as any known patterns. This estimation is mathematically expressed as Eq.(26).

$$E(w_i) = \frac{1}{n} \sum_{j=1}^n \beta_j \cdot C_j \quad (26)$$

where $E(w_i)$ is the estimated meaning of the word w_i , β_j is a weighting factor assigned to each contextual word C_j , and n is the number of words in the surrounding window.

The refinement layer also exists in KyBERT, and therefore, further adjustments to words or phrases are made within a low-information environment. Refinement is considered during the second pass of analysis in adjusting the estimated meaning to fine-tune interpretation through fallback patterns and other surrounding information. The refined meaning can be represented mathematically as Eq.(27).

$$R(w_i) = \theta \cdot E(w_i) + (1 - \theta) \cdot F(w_i) \quad (27)$$

where $R(w_i)$ is the refined meaning of the word w_i , and θ is a factor that balances the estimated meaning $E(w_i)$ and fallback weight $F(w_i)$. This refinement ensures that the model maintains accuracy even in ambiguous or incomplete contexts.

Another important feature of Nocturnal Language Processing involves uncertainty management in low information environments. Although scores do not rise above the expectations, KyBERT equips the system with an uncertainty management mechanism, where a weighted penalty on uncertain data points is assigned. This process is represented mathematically in Eq.(28).

$$U(w_i) = \mu \cdot (1 - S(w_i)) \quad (28)$$

where $U(w_i)$ is the uncertainty score for the word w_i , and μ is a penalty factor applied to words with low confidence scores. By calculating uncertainty, KyBERT reduces the impact of potentially unreliable interpretations, ensuring that the final meaning is less influenced by uncertain elements.

Long-Range Language Detection

The inspiration for Long-Range Language Detection comes from the katydid behaviour, which can sense vibrations and signals over long distances through its extended antennae. For a katydid to be able to survive and reproduce successfully, it would need to use its antennae to pick up a signal from predators or mates who might be miles away. In the same respect, long-range detection in KyBERT refers to the capturing of dependencies that are long-range in language, where words or phrases appearing far apart in a text can inform each other. So, it is extremely important for understanding complex long-range connections in documents or texts. This ability to understand long-range connections is crucial for grasping the full meaning of complex documents or texts. Traditional models tend to focus on nearby words, often struggling to maintain the semantic link between distant elements within a sentence or paragraph.

Long-Range Language Detection mechanisms track and interpret words and structures over long distances. First, it calculates dependencies between the word and all the other words in the text no matter how large the intervening distance. The dependency weight can be expressed mathematically in Eq.(29).

$$D(w_i) = \sum_{j=1}^n \alpha_{ij} \cdot E(w_j) \quad (29)$$

where $D(w_i)$ represents the long-range dependency of word w_i , α_{ij} is the weight of the dependency between words w_i and w_j , and $E(w_j)$ is the embedding of the word w_j .

To manage the complexity of processing these long-range dependencies, KyBERT scales mechanism with some relationships are more significant over the others, that helps to filter out weaker less relevant connections and focuses on the strongest long-range dependency. The scaling factor β_{ij} adjusts the importance of each dependency.

$$D_s(w_i) = \sum_{j=1}^n \beta_{ij} \cdot D(w_i) \quad (30)$$

where in Eq.(30), $D_s(w_i)$ is the scaled dependency for word w_i , β_{ij} determining the relevance of each relationship based on the overall context. This ensures that the model focuses on critical connections between words, similar to how a katydid's antennae selectively pick up important signals in its environment.

Long-range detection requires continuous tracking of contextual shifts throughout the text. KyBERT introduces a dynamic adjustment mechanism that modifies the dependency weights as new words are processed. This ensures that earlier words remain relevant even as the sentence evolves. The dynamic adjustment for long-range dependencies can be mathematically represented as Eq.(31).

$$D_d(w_i) = \gamma_t \cdot D_s(w_i) + (1 - \gamma_t) \cdot C_t \quad (31)$$

where $D_d(w_i)$ represents the dynamically adjusted dependency for word w_i , γ_t is the dynamic scaling factor based on time step t , and C_t is the contextual adjustment applied as the sentence progresses.

In addition to its long-range detection mechanism, KyBERT has a mechanism for reinforcement that builds up key ties between distant words. This can help the model reinforce all the relations amongst words or phrases that are importantly relevant to meaning in text. The reinforcement weight $R(w_i)$ is calculated as represented mathematically in Eq.(32).

$$R(w_i) = \lambda \cdot D_d(w_i) + (1 - \lambda) \cdot A(w_i) \quad (32)$$

where $R(w_i)$ represents the reinforced weight for word w_i , λ is the reinforcement scaling factor, and $A(w_i)$ is the regular attention score for the word.

In the case of KyBERT, a decay function is used to control distant dependencies, which are gradually lost over time. The decay function ensures that as distance increases between words of a dependency, the strength of connection is weakened unless this is reinforced by other contextual factors. The decay function is expressed mathematically in Eq.(33).

$$D_c(w_i) = e^{-\delta \cdot d_{ij}} \cdot R(w_i) \quad (33)$$

where $D_c(w_i)$ is the decayed dependency for word w_i , δ is the decay rate, and d_{ij} is the distance between words w_i and w_j . This function helps balance the importance of long-range dependencies, ensuring that only the most relevant connections persist over time.

Finally KyBERT captures the overall meaning of the sentence in terms of representing the sentence, by infusing the long-range dependency relations with the short-range and long-range relationships, towards coherent understanding. This integrated meaning is represented mathematically in Eq.(34).

$$M(w_i) = \theta \cdot D_c(w_i) + (1 - \theta) \cdot A_s(w_i) \quad (34)$$

where $M(w_i)$ represents the final meaning of word w_i , θ is a blending factor, and $A_s(w_i)$ is the short-range attention score for the word.

Resource-Rich Language Handling

KyBERT's Resource-Rich Language Handling emulates the katydid's complex mating ritual, where males present resource-rich nuptial gifts to females, improving chances of reproduction. In a similar manner, KyBERT processes language rich with resources dense, complex sentences filled with diverse elements—by leveraging this wealth of information to enhance understanding. The aim is to extract maximum meaning from intricate, multi-layered texts, akin to how the katydid uses the richness of its environment for survival and communication. In language, resource-rich text contains a multitude of syntactical structures, high-frequency terms, low-frequency specialized vocabulary, and complex interdependencies. KyBERT's resource-rich handling strategy ensures that it processes such dense information effectively, extracting relevant meaning without being overwhelmed by the abundance of data.

The process begins with assigning a richness score to each word or phrase, based on the complexity and amount of information it carries. This score, denoted as $R(w_i)$, is influenced by various factors, including the word's syntactic role, semantic weight, and frequency of occurrence. The richness score can be represented mathematically in Eq.(35).

$$R(w_i) = \alpha \cdot S(w_i) + \beta \cdot F(w_i) \quad (35)$$

where $S(w_i)$ represents the syntactic complexity of the word w_i , $F(w_i)$ represents its frequency within the text, and α , β are scaling factors that adjust the influence of these components. This formula quantifies how resource-rich each word or phrase is, allowing KyBERT to prioritize high-value elements for deeper analysis.

To ensure efficient processing of resource-rich language, KyBERT introduces a distribution mechanism that allocates attention according to each word's richness score. Words or phrases with higher scores receive more processing resources, enabling KyBERT to focus on key elements while avoiding unnecessary computation on trivial parts of the text. The distribution mechanism is defined mathematically in Eq.(36).

$$A_r(w_i) = \gamma \cdot R(w_i) + (1 - \gamma) \cdot A(w_i) \quad (36)$$

where $A_r(w_i)$ represents the attention score adjusted for resource richness, and γ is a scaling factor that balances between richness $R(w_i)$ and the normal attention score $A(w_i)$. This distribution ensures that the richest parts of the text receive the most attention during processing.

KyBERT also incorporates contextual weighting, ensuring that resource-rich words are evaluated not just based on their intrinsic complexity, but also in relation to their surrounding context. This mechanism allows the model to adapt to the dynamic flow of meaning in dense texts, highlighting parts that might gain or lose importance depending on their context. The contextual weighting function is expressed as Eq.(37).

$$C_r(w_i) = \delta \cdot R(w_i) + (1 - \delta) \cdot C(w_i) \quad (37)$$

where $C_r(w_i)$ represents the contextually adjusted richness score for word w_i , and $C(w_i)$ is the standard contextual embedding for the word. The contextual weighting ensures that KyBERT processes resource-rich words dynamically, adjusting their importance based on the evolving meaning of the text.

To further refine its handling of complex language, KyBERT introduces a hierarchical resource evaluation step. This step assesses how resources within a sentence or paragraph relate to each other, identifying key structures that contribute to the overall meaning. The hierarchical evaluation is expressed mathematically in Eq.(38).

$$H(w_i) = \sum_{j=1}^n \theta_{ij} \cdot R(w_j) \quad (38)$$

where $H(w_i)$ represents the hierarchical resource score for word w_i , and θ_{ij} is the weight assigned to the relationship between word w_i and its neighboring word w_j . This hierarchical approach allows KyBERT to understand how resources are distributed within a text, ensuring that key relationships between different elements are captured and analyzed.

In resource-rich language, it is common to encounter overlapping meanings, where multiple layers of meaning coexist within a single sentence or phrase. KyBERT introduces an overlap detection mechanism to resolve these conflicts and ensure that each layer of meaning is adequately addressed. The overlap detection score $O(w_i)$ is calculated as shown in Eq.(39).

$$O(w_i) = \frac{1}{n} \sum_{j=1}^n \zeta_{ij} \cdot R(w_j) \quad (39)$$

where $O(w_i)$ represents the overlap score for word w_i , and ζ_{ij} is the weight assigned to the overlapping meanings between word w_i and its neighbors. This score helps KyBERT manage overlapping resources, ensuring that each layer of meaning is processed without interference.

Finally KyBERT integrates all these factors into a final resource processing score that combines richness, contextual relevance, hierarchical evaluation, and overlap detection. This final score is used to guide the overall interpretation of the text. The final processing score $P(w_i)$ is represented mathematically in Eq.(40).

$$P(w_i) = \lambda \cdot C_r(w_i) + \mu \cdot H(w_i) + v \cdot O(w_i) \quad (40)$$

where $P(w_i)$ is the final resource processing score for word w_i , and λ , μ , and v are scaling factors that adjust the contribution of contextual relevance, hierarchical evaluation, and overlap detection, respectively.

Contrasting Traditional BERT with KyBERT

Traditional BERT is designed to understand the context of words by processing language in both directions, which allows it to capture intricate relationships within text.

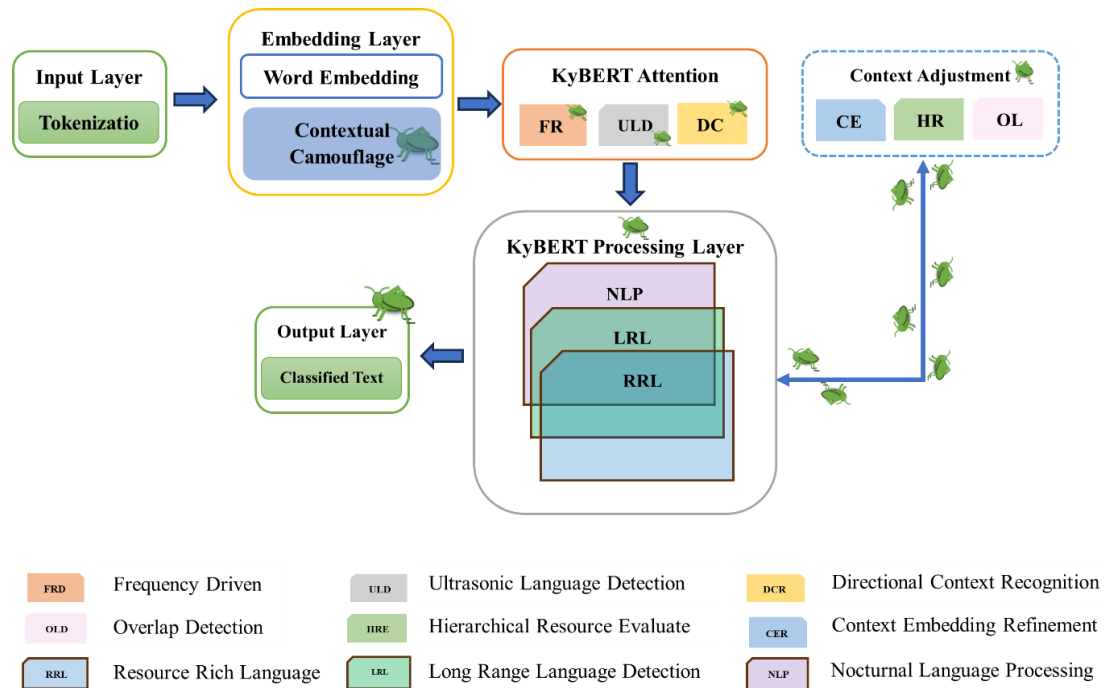


Fig 2. KyBERT Structure.

While BERT is highly effective in many natural language processing tasks, it tends to struggle with detecting long-range dependencies, handling sparse or low-information contexts, and effectively managing complex, resource-heavy language structures. These limitations reduce its efficiency in processing nuanced and multi-layered texts. KyBERT optimizes the traditional BERT framework by addressing these specific challenges. KyBERT introduces Contextual Camouflage, which refines the way words are embedded within their context, enhancing the model's ability to interpret subtle and ambiguous meanings. Frequency-Driven Attention adjusts attention based on linguistic patterns, ensuring that critical signals receive the necessary focus. The Ultrasonic Language Detection feature adds sensitivity to rare or specialized language, capturing meanings that traditional BERT might overlook. **Fig 2.** Shows the overall functional portion of KyBERT.

Further optimization is evident in Directional Context Recognition, where KyBERT enhances the ability to track meaning shifts throughout the text, ensuring greater accuracy in understanding sentence flow. Nocturnal Language Processing enables KyBERT to handle low-information environments effectively, filling gaps in meaning where data might be sparse. Long-Range Language Detection tackles BERT's challenge with distant dependencies, ensuring that even widely separated words influence the overall understanding. KyBERT also optimizes resource-heavy text structures through Resource-Rich Language Handling, ensuring complex, information-dense sentences are processed efficiently without loss of meaning. KyBERT builds upon traditional BERT by integrating bio-inspired optimizations such as context embedding, attention adjustments, and enhanced detection capabilities, making it more robust for nuanced, resource-dense, and long-range language processing tasks.

IV. DATASET DESCRIPTION

A dataset is a structured collection of data, primarily used to train models or analyze patterns. The Amazon Reviews dataset serves the purpose of examining customer opinions across a vast array of product categories. Containing 476,001 unique values, it is a highly comprehensive resource of reviews for sentiment analysis, customer behavior, and product trends analysis. An entry provides all the necessary information: the rating of the review, the textual feedback, and relevant product metadata. This dataset is highly useful for sentiment analysis, machine learning model training, and natural language processing applications. The wide variety of reviews allows comprehensive models to be developed that encompass consumer preferences on diverse products, which would greatly benefit businesses by the understanding of market trends and the needs of their customers. This could also be used in research to understand how different types of products affect customer satisfaction and behavior, thereby opening up insights that may be useful in improving customer service and product offerings.

V. RESULTS AND DISCUSSION

Results and discussion in sentiment analysis are typically carried out with the objective of evaluating the model's performance based on the key metrics such as Classification Accuracy (CA) and F-Measure (FM). Classification Accuracy refers to the portion of accurately classified instances among the total number of instances. F-Measure balances between false positives and false negatives by accounting for precision and recall. **Fig 3.** depicts CA and FM results obtained.

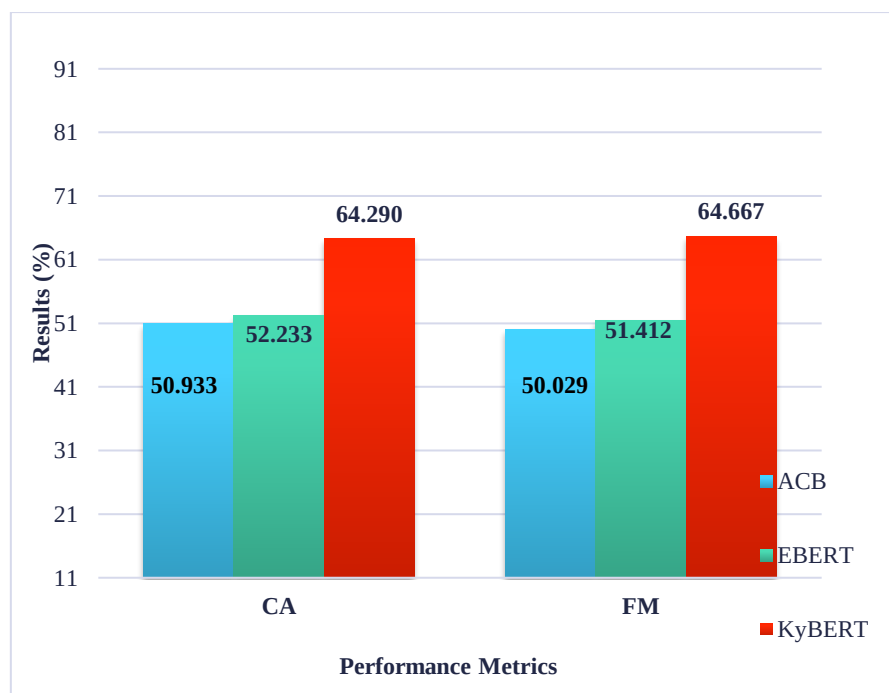


Fig 3. Classification Accuracy and F-Measure.

In this analysis, three classification algorithms ACB, EBERT, and KyBERT were compared using these metrics. The ACB algorithm achieved a CA of 50.933% and an FM of 50.029%, showing modest performance. EBERT slightly outperformed ACB, with a CA of 52.233% and an FM of 51.412%, indicating better overall classification accuracy and balance between precision and recall. **Table 1.** Presented the obtained results clearly in tabular format.

Table 1. Classification Accuracy and F-Measure

Classification Algorithms	CA	FM
ACB	50.933	50.029
EBERT	52.233	51.412
KyBERT	64.290	64.667

Best performance was shown by KyBERT, which achieved a CA of 64.290%, FM of 64.667%, and was far ahead of both ACB and EBERT. This result shows that KyBERT was better with understanding and predicting the sentiment, probably due to an improved contextual understanding and feature extraction procedure. This variation in performance between the algorithms develops the sense of carefully selecting models according to specific complexities of the respective tasks, as issues in sentiment analysis. Superior performance of KyBERT indicates its better ability to cope with sentiment classification in real applications and mostly those where nuanced sentiment is crucial. In sentiment analysis, Fowlkes-Mallows Index (FMI) and Matthews Correlation Coefficient are helpful measures of the quality of classifications obtained by classification models. FMI measures the degree of similarity between the predicted clusters and true clusters, and in addition to measurement of the type of precision for each class, it also evaluates a balanced performance. MCC evaluates the correlation between predicted and actual classifications, accounting for true and false positives and negatives, offering a comprehensive view of the model's performance.

This model compared three classification algorithms in which ACB, EBERT, and KyBERT were assessed in terms of FMI and MCC. The performance from the ACB was lowest at 50.052% FMI and 1.925 MCC, implying that this was not meaningful because it failed to meaningfully identify and classify the sentiment data. **Fig 4.** The obtained results for FMI and MCC in case of ACB, EBERT, and KyBERT. **Table 2.** The achieved results of FMI and MCC results.

Table 2. Classification Accuracy and F-Measure

Classification Algorithms	FMI	MCC
ACB	50.052	1.925
EBERT	51.433	4.520
KyBERT	64.678	28.618

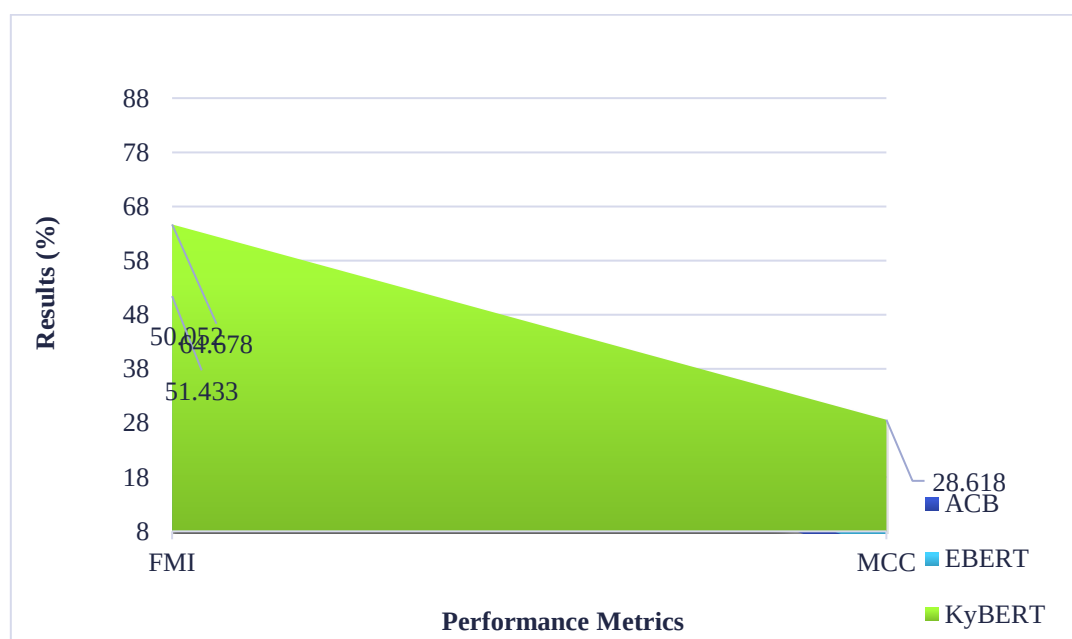


Fig 4. FMI and MCC.

EBERT outperformed ACB marginally with an FMI of 51.433% and an MCC of 4.520, implying that it brought about slight improvement in its classification performance with good balancing between precision and recall. KyBERT outperforms both the systems of ACB and EBERT, with an FMI of 64.678% and an MCC of 28.618, thus signifying a significant improvement in correlation and classification accuracy. In particular, its high MCC score depicts the ability to

handle imbalanced data and more accurate predictability. As per the results fetched KyBERT is the strong model for sentiment classification with difference based on FMI and MCC thus provides more reliability as well as precision in tasks pertaining to sentiment analysis.

There are four key metrics in evaluating performance of sentiment analysis: True Positive Rate, False Positive Rate, True Negative Rate, and False Negative Rate. These metrics provide insight into the model's ability to classify data accurately. TPR is also called sensitivity or recall it measures the proportion of actual positive instances that are classified correctly. A higher TPR means that the model is at a greater efficacy of correctly predicting positive sentiment in the dataset. TNR or specificity computes the proportion of the actually negative instances which are classified correctly. High TNR shows the strength of the model to recognize the negative sentiments. FPR represents the proportion of negative instances misclassified as positive, and FNR accounts for the proportion of the positively classified instances wrongly designated as negative.

In the analysis of three algorithms ACB, EBERT, and KyBERT these metrics illustrate varying levels of classification accuracy. ACB, with a TPR of 51.604%, an FPR of 49.677%, a TNR of 50.323%, and an FNR of 48.396%, displays modest performance. These figures suggest that ACB struggles with classifying both positive and negative sentiments, as evidenced by its relatively high FPR and FNR, indicating misclassification in nearly half of the cases. **Fig 5.** Depicts the obtained TPR, TNR, FPR, FNR results in pictorial format.

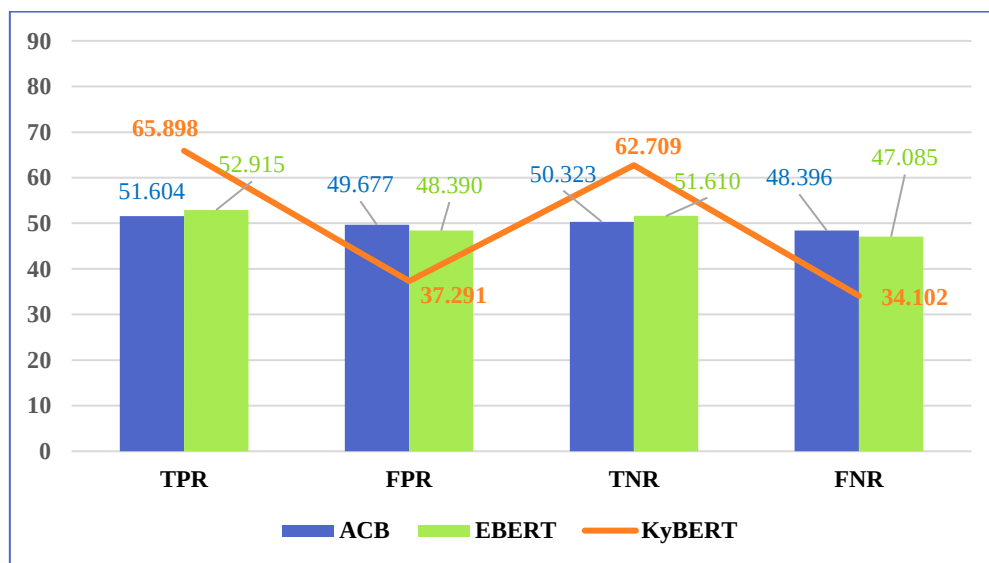


Fig 5. TPR, TNR, FPR and FNR.

EBERT performs slightly better, with a TPR of 52.915%, an FPR of 48.390%, a TNR of 51.610%, and an FNR of 47.085%. This indicates a slight improvement in recognizing both positive and negative sentiments, with marginally fewer misclassifications compared to ACB. **Table 3.** Contains the TPR, TNR, FPR and FNR results obtained by the models ACB, EBERT and KyBERT.

Table 3. TPR, TNR, FPR and FNR

Classification Algorithms	TPR	FPR	TNR	FNR
ACB	51.604	49.677	50.323	48.396
EBERT	52.915	48.390	51.610	47.085
KyBERT	65.898	37.291	62.709	34.102

KyBERT significantly outperforms the other two algorithms. With a TPR of 65.898%, an FPR of 37.291%, a TNR of 62.709%, and an FNR of 34.102%, KyBERT excels in identifying both positive and negative sentiments. Its lower FPR and FNR signify fewer misclassifications, reflecting the model's robust capability in distinguishing sentiment polarity. The superior performance of KyBERT in identification with maximum accuracy establishes the position that the model can particularly be of use where scenarios require the correct classification of both positive and negative sentiments. This shows that KyBERT is the most reliable model for real-world applications of sentiment analysis, providing increased precision and reliability.

VI. CONCLUSION

Sentiment analysis requires a deep understanding of contextual nuances and sentence semantics. In this work, models like ACB, EBERT, and KyBERT have been employed to address the complex relationships between words, capturing meaning effectively in sentiment-related tasks. Models have been tested using various performance metrics such as classification accuracy and F-measure, combined with other indicators that are important for a comprehensive analysis of their

capabilities in interpreting the sentiment. KyBERT, with its bio-inspired optimizations, aims to refine sentence-level semantics and capture subtle contextual variations in language, while ACB and EBERT provide solid benchmarks. Each model contributes uniquely to advancing the understanding of sentiment, bringing valuable insights into language complexities. The results indicate advancements in the ability of these models to recognize, classify, and evaluate sentiments effectively, thereby enhancing sentiment analysis approaches. The work overall highlights the progress made in optimizing natural language models to enrich sentiment comprehension, focusing on the depth of sentence semantics, ultimately benefiting real-world applications.

CRedit Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Prativa Mishra, Gopalakrishnan T and Parameswaran T; **Methodology:** Prativa Mishra; **Software:** Gopalakrishnan T and Parameswaran T; **Data Curation:** Prativa Mishra; **Writing-Original Draft Preparation:** Prativa Mishra, Gopalakrishnan T and Parameswaran T; **Visualization:** Gopalakrishnan T and Parameswaran T; **Investigation:** Prativa Mishra; **Supervision:** Gopalakrishnan T and Parameswaran T; **Validation:** Parameswaran T; **Writing- Reviewing and Editing:** Prativa Mishra, Gopalakrishnan T and Parameswaran T; All authors reviewed the results and approved the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data Availability Statement

The Datasets used and /or analysed during the current study available from the corresponding author on reasonable request.

Funding

No funding agency is associated with this research.

Consent to Publish

All authors gave permission to consent to publish.

References

- [1]. Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment analysis for customer review: Case study of Traveloka," *Procedia Computer Science*, vol. 216, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.
- [2]. D. Chen, H. Zhengwei, T. Yiting, M. Jintao, and R. Khanal, "Emotion and sentiment analysis for intelligent customer service conversation using a multi-task ensemble framework," *Cluster Computing*, vol. 27, no. 2, pp. 2099–2115, Jun. 2023, doi: 10.1007/s10586-023-04073-z.
- [3]. A. Patel, P. Oza, and S. Agrawal, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," *Procedia Computer Science*, vol. 218, pp. 2459–2467, 2023, doi: 10.1016/j.procs.2023.01.221.
- [4]. J. Bowden and R. Gemayel, "Sentiment and trading decisions in an ambiguous environment: A study on cryptocurrency traders," *Journal of International Financial Markets, Institutions and Money*, vol. 80, p. 101622, Sep. 2022, doi: 10.1016/j.intfin.2022.101622.
- [5]. A. J. Lak, R. Boostani, F. A. Alenizi, A. S. Mohammed, and S. M. Fakhrahmad, "RoBERTa, ResNeXt and BiLSTM with self-attention: The ultimate trio for customer sentiment analysis," *Applied Soft Computing*, vol. 164, p. 112018, Oct. 2024, doi: 10.1016/j.asoc.2024.112018.
- [6]. M. Wilksch and O. Abramova, "PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets," *International Journal of Information Management Data Insights*, vol. 3, no. 1, p. 100171, Apr. 2023, doi: 10.1016/j.jjimei.2023.100171.
- [7]. Y. Huang, X. Bai, Q. Liu, H. Peng, Q. Yang, and J. Wang, "Sentence-level sentiment classification based on multi-attention bidirectional gated spiking neural P systems," *Applied Soft Computing*, vol. 152, p. 111231, Feb. 2024, doi: 10.1016/j.asoc.2024.111231.
- [8]. M. Jang and P. Kang, "Sentence transition matrix: An efficient approach that preserves sentence semantics," *Computer Speech & Language*, vol. 71, p. 101266, Jan. 2022, doi: 10.1016/j.csl.2021.101266.
- [9]. J. Dai, Y. Zhao, and Z. Li, "Sentiment-topic dynamic collaborative analysis-based public opinion mapping in aviation disaster management: A case study of the MU5735 air crash," *International Journal of Disaster Risk Reduction*, vol. 102, p. 104268, Feb. 2024, doi: 10.1016/j.ijdr.2024.104268.
- [10]. P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," *IEEE Access*, vol. 11, pp. 108433–108447, 2023, doi: 10.1109/access.2023.3320738.
- [11]. A. Pani, S. N. Balla, and P. K. Sahu, "Decoding consumer-centric transition to electric mobility based on sentiment, semantic and statistical analysis," *Research in Transportation Business & Management*, vol. 51, p. 101069, Dec. 2023, doi: 10.1016/j.rtbm.2023.101069.
- [12]. T. Zhao, L. Meng, and D. Song, "Multimodal Aspect-Based Sentiment Analysis: A survey of tasks, methods, challenges and future directions," *Information Fusion*, vol. 112, p. 102552, Dec. 2024, doi: 10.1016/j.inffus.2024.102552.
- [13]. A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *Journal of Big Data*, vol. 10, no. 1, Jun. 2023, doi: 10.1186/s40537-023-00781-w.
- [14]. N. Habbat, H. Nouri, H. Anoun, and L. Hassouni, "Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106999, Nov. 2023, doi: 10.1016/j.engappai.2023.106999.
- [15]. S. P. Geetha, N. M. S. Sundari, J. Ramkumar, and R. Karthikeyan, "Energy Efficient Routing In Quantum Flying Ad Hoc Network (Q-Fanet) Using Mamdani Fuzzy Inference Enhanced Dijkstra ' S Algorithm (Mfi-Eda)," vol. 102, no. 9, pp. 3708–3724, 2024.
- [16]. J. Ramkumar, K. S. Jeen Marseline, and D. R. Medhunhashini, "Relentless Firefly Optimization-Based Routing Protocol (RFORP) for Securing Fintech Data in IoT-Based Ad-Hoc Networks," *International Journal of Computer Networks and Applications*, vol. 10, no. 4, p. 668, Aug. 2023, doi: 10.22247/ijcna/2023/223319.
- [17]. R. Karthikeyan and R. Vadivel, "Proficient Dazzling Crow Optimization Routing Protocol (PDCORP) for Effective Energy Administration in Wireless Sensor Networks," 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM), pp. 1–6, Aug. 2023, doi: 10.1109/elexcom58812.2023.10370559.

- [18]. P. Menakadevi and J. Ramkumar, “Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data,” 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1–5, Mar. 2022, doi: 10.1109/icacta54488.2022.9753203.
- [19]. B. Wan, P. Wu, C. K. Yeo, and G. Li, “Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies,” *Information Processing & Management*, vol. 61, no. 2, p. 103609, Mar. 2024, doi: 10.1016/j.ipm.2023.103609.
- [20]. A. Ismail, F. H. Sazali, S. N. Agos Jawaddi, and S. Mutalib, “Stream ETL framework for twitter-based sentiment analysis: Leveraging big data technologies,” *Expert Systems with Applications*, vol. 261, p. 125523, Feb. 2025, doi: 10.1016/j.eswa.2024.125523.
- [21]. A. Dey and M. Jenamani, “Aspect based sentiment analysis of consumer reviews using unsupervised attention neural framework,” *Applied Soft Computing*, vol. 167, p. 112259, Dec. 2024, doi: 10.1016/j.asoc.2024.112259.
- [22]. Q. Wang, J. Zhu, H. Shu, K. O. Asamoah, J. Shi, and C. Zhou, “GUDN: A novel guide network with label reinforcement strategy for extreme multi-label text classification,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 4, pp. 161–171, Apr. 2023, doi: 10.1016/j.jksuci.2023.03.009.
- [23]. H. Bekamiri, D. S. Hain, and R. Jurowetzki, “PatentSBERTa: A deep NLP based hybrid model for patent distance and classification using augmented SBERT,” *Technological Forecasting and Social Change*, vol. 206, p. 123536, Sep. 2024, doi: 10.1016/j.techfore.2024.123536.
- [24]. Y. Chu, H. Cao, Y. Diao, and H. Lin, “Refined SBERT: Representing sentence BERT in manifold space,” *Neurocomputing*, vol. 555, p. 126453, Oct. 2023, doi: 10.1016/j.neucom.2023.126453.
- [25]. R. Lamsal, M. R. Read, and S. Karunasekera, “CrisisTransformers: Pre-trained language models and sentence encoders for crisis-related social media texts,” *Knowledge-Based Systems*, vol. 296, p. 111916, Jul. 2024, doi: 10.1016/j.knosys.2024.111916.
- [26]. J. Xu, C. S. Zhanyi, L. Xu, and L. Chen, “BlendCSE: Blend contrastive learnings for sentence embeddings with rich semantics and transferability,” *Expert Systems with Applications*, vol. 238, p. 121909, Mar. 2024, doi: 10.1016/j.eswa.2023.121909.
- [27]. Z. Zhao, M. Tang, F. Zhao, Z. Zhang, and X. Chen, “Incorporating semantics, syntax and knowledge for aspect based sentiment analysis,” *Applied Intelligence*, vol. 53, no. 12, pp. 16138–16150, Dec. 2022, doi: 10.1007/s10489-022-04307-4.
- [28]. A. Dai, X. Hu, J. Nie, and J. Chen, “Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis,” *International Journal of Data Science and Analytics*, vol. 14, no. 1, pp. 17–26, Mar. 2022, doi: 10.1007/s41060-022-00315-2.
- [29]. N. T. Mhaske and A. S. Patil, “Sentence Annotation for Aspect-oriented Sentiment Analysis: A Lexicon based Approach with Marathi Movie Reviews,” *Journal of The Institution of Engineers (India): Series B*, vol. 105, no. 6, pp. 1669–1677, May 2024, doi: 10.1007/s40031-024-01072-5.
- [30]. A.-Q. Vermillet, S. E. Krejberg, J. S. Nielsen, and C. E. Parsons, “An analysis of national media coverage of a parental leave reform investigating sentiment, semantics and contributors,” *Scientific Reports*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-023-49356-y.
- [31]. C. Zhu, B. Yi, and L. Luo, “Triple-channel graph attention network for improving aspect-level sentiment analysis,” *The Journal of Supercomputing*, vol. 80, no. 6, pp. 7604–7623, Nov. 2023, doi: 10.1007/s11227-023-05745-8.
- [32]. Y. Yu, D. Qiu, and R. Yan, “A Quantum-Like Tensor Compression Sentence Representation Based on Constraint Functions for Semantics Analysis,” *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Jan. 2024, doi: 10.1007/s44196-023-00380-w.
- [33]. Y. Li, W. Weng, C. Liu, and L. Li, “CSMF-SPC: Multimodal Sentiment Analysis Model with Effective Context Semantic Modality Fusion and Sentiment Polarity Correction,” *Pattern Analysis and Applications*, vol. 27, no. 3, Aug. 2024, doi: 10.1007/s10044-024-01320-w.
- [34]. Y. Hou, F. Liu, X. Zhuang, and Y. Zhang, “Twain-GCN: twain-syntax graph convolutional networks for aspect-based sentiment analysis,” *Knowledge and Information Systems*, vol. 66, no. 9, pp. 5541–5560, May 2024, doi: 10.1007/s10115-024-02135-1.
- [35]. S. Kaddoura and R. Nassar, “EnhancedBERT: A feature-rich ensemble model for Arabic word sense disambiguation with statistical analysis and optimized data collection,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, p. 101911, Jan. 2024, doi: 10.1016/j.jksuci.2023.101911.
- [36]. K. Ahmed et al., “Breaking down linguistic complexities: A structured approach to aspect-based sentiment analysis,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101651, Sep. 2023, doi: 10.1016/j.jksuci.2023.101651.

Publisher’s note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.