Sentiment Analysis Prediction of Social Media Text Using Deep Learning Techniques

Rashidkhan R Pathan
Department of Information Technology
L.D. College of Engineering

Abstract:

Social media platforms have revolutionised interpersonal interaction and communication, becoming an essential part of our everyday life. Examining the wide variety of research studies carried out in this field is essential to developing a thorough understanding of FL's impacts on social media content. The purpose of this study of the literature is to critically analyse the body of knowledge about the impact of figurative language, including oxymoron, sarcasm, and euphemism, on social media posts. This review of the literature looks for important discoveries, patterns, and knowledge gaps by analysing and synthesising the body of current research.

Keywords: Sentiment Analysis, Prediction, Social Media Text, Using Deep Learning Techniques

1.0. Analysis of euphemism in different societies

Baranova & Pletenko (2018) outlined the main features of euphemism for the issues of sickness, dying, and death prevalent in English mass-media discourse. It is believed that substitution functions in euphemistic use, such as masking, prevention, grinding, and manipulating, are important. In order to comprehend the euphemistic origins and motives, a number of lexical replacement units were investigated. The enduring quality of this language phenomenon—in which social, psychological, and theological interdictions coexisted—was heavily emphasised.

Almoayidi (2018) spoke on the importance of euphemisms in communication. It focusses on the Southern Region Dialect (spoken in Al-Qunfudah) and the Hijazi (spoken in Makkah) Arabic dialects that are used in Saudi Arabia. These two dialect speakers discussed taboo topics including sex, death, and body functions using euphemisms. Most of the information for this article was collected by the researcher some years ago. Some, nevertheless, also originated from other sources, such as past studies, TV series, blogs, YouTube, and Twitter. In these two dialects, other tactics such phonemic substitution, compounding, derivation, and deletion were noted. He said further research is required to find out how euphemism is used in medical contexts, adolescent conversations, and nonliterate communication.

A new taxonomy of nine distinct linguistic practices, including the use of euphemisms, breaking taboos, avoiding swear words, using mild swear words, paralinguistic information, intentional bleeping, being accompanied by negative metadiscourse, using innovative or archaic swear words, and using a different language, was proposed by Bednarek (2019) after examining how offensive and taboo words are used in TV shows. He proved that media linguistics can study language used in the televisual medium in addition to language found in journalistic mass media, taking into account language-external restrictions like legal regulations, by looking at offensive language in TV shows.

Olimat (2020) looked at the euphemisms and dysphemisms used in Jordanian culture while talking about COVID-19. It also discussed the ways in which Jordanians translate from Arabic to English, use medical-related acronyms, and use metaphors to explain illnesses. When talking about COVID-19, it was advised to use paralinguistic cues such as hand gestures, eye movements, facial emotions, body language, tone, and pitch of voice. In order to prevent the transmission of COVID-19, raise awareness of the virus's risks, and convey the appropriate message, a list of commonly used derogatory and polite words in Jordanian Arabic was supplied.

Matondang et al. provided a description of the many euphemism and dysphemism types seen in Indonesian da'wah (2020). The internet was used to get the audio documentation data (Spotify). Words, phrases, and sentences were found to be the grammatical units of dysphemism and euphemism. They looked at the usage of euphemisms and dysphemisms in Indonesian preaching using a qualitative research paradigm.
Cao (2020) spoke on the pragmatic rules that govern the usage of English euphemisms in expressing thoughts. There are now five pragmatic roles to consider: concealment, humour, elegance, avoidance, and politeness. There was a suggestion that the study lacked any pertinent practical applications and was purely theoretical. There was proof that euphemisms were used in social interactions as well as political discourse, daily speech, and military jargon.

Hasyim et al. (2020) provided an explanation and description of the euphemisms and dysphemisms utilised in internet news reports during the Republic of Indonesia's 2019 presidential election. The online media texts from CNN Indonesia, SindoNews, DetikNews, and TribunNews served as the data sources. Framing analysis and componential semantic analysis were used to assess the sources. Lexical dysphemism has been divided into three categories: lexical mitigation, metaphorical equations, and exaggeration. This is in spite of the fact that there are other types of lexical euphemism, such as borrowed words and metaphorical equations. The results of the investigation showed that euphemisms were produced via the employment of metaphorical equations and lexical mitigating terms. Conversely, dysphemism makes use of linguistic mitigation, metaphorical equations, and exaggeration.

Terry (2020) explored whether these expressions of euphemistic dysphemisms and dysphemistic euphemisms may communicate sarcasm and banter. She developed four terms: euphemistic dysphemism, dysphemistic euphemism, irony, and banter. Two theories were taken into consideration for this study in order to describe the fundamental mechanics of how these four gadgets function. The corpus includes snippets from several American television shows. The results of the research showed that while euphemistic dysphemisms may sometimes be used to depict humour, they are unable to convey irony.
Ariani et al. (2020) used complementary analysis to identify the semantic alterations that transpired in Tempo Magazine's news content upon translating euphemisms and dysphemisms. The information was taken from a bilingual news piece for Tempo Magazine's 2019 edition that was first published in Indonesian and then translated into English. Six distinct kinds of semantic alterations were identified in Tempo Magazine's translations of euphemism and dysphemism. They included semantic exaggeration, semantic constriction, semantic metonymy, semantic amelioration, semantic widening, and semantic metaphor. Semantic change pejoration occurs when the opposite is done, but ameliorations occur when dysphemism is changed to euphemism. The meaning link between euphemism and dysphemism remains, therefore the words stay interrelated even when their sense relation varies throughout translation.

The question of whether emotional polarity, intensity, connotation, dominance, and arousal knowledge may help distinguish between euphemistic and dysphemistic language was explored by Felt & Riloff (2020). Near-synonyms for the ideas of shooting, lying, and stealing were produced by the use of semantic lexicon induction. Words were classified as euphemistic, dysphemistic, or neutral by using sentiment analysis approaches based on lexicon and context.
It was discovered that emotional polarity and connotation knowledge were beneficial for this job. Furthermore, categorising the words themselves was often less effective than determining the mood expressed in the sentence contexts that surrounded them.

The kinds and applications of euphemisms and dysphemisms used in President Donald Trump's 2020 State of the

Union address were detailed by Kafi & Degaf (2021). They conducted an in-depth examination of their studies using the qualitative descriptive research technique. There are eight distinct methods to use euphemisms in language: understatement, jargon, synaesthesia, exaggeration, circumlocution, figurative language, and brevity. Additionally, eight distinct euphemism use patterns were found. These included providing facts, exaggerating, stepping back, being respectful, refraining from using language that might imply fear or terror, criticising, staying away from taboo subjects, giving counsel, and displaying compassion. The link between euphemism and dysphemism was discussed in Donald Trump's campaign address.

The impact of English euphemisms and dysphemisms as linguistic techniques in political newspaper articles were examined by Aytan et al. in 2021. A total of 393 euphemisms and dysphemisms were collected for examination from texts in political media. Five primary types of euphemisms were identified: COVID-19, social media, contemporary life, economics, and politics. This was done to investigate the association between the frequency of use of euphemisms in political media discourse and their utilisation. A link was also computed between the thematic frequency of dysphemism and political media materials. They came to the conclusion that a rhetorical contrast may be made using both euphemistic and dysphemistic patterns.

Yildiz (2021) expounded on the frequent use of euphemisms by Turkish university students. The descriptive model concurrently discovered the euphemistic structures used by university students. Three scenarios were used to analyse the data from the interview form: using the loo, receiving news of a death, and gaining weight. Three scenarios were studied in order to investigate the language formation of euphemisms. Common methods like implications, metaphors, metonymy, and borrowed words were used. Additionally, it was shown that young Turkish people formed euphemisms via the use of rhetorical enquiries and phrases like and as.

Zhu & Bhat (2021) tackled the problem of euphemistic phrase identification and proposed a workable solution. By pre-selecting euphemistic phrase candidates, mining quality phrases from the text corpus, and rating phrases using a masked language model, it can perform euphemistic phrase recognition automatically and without human intervention. In order to choose a collection of possible euphemistic phrase possibilities, word embedding similarities were investigated. Additionally, it focused on the job of euphemistic phrase recognition for a list of target keywords, which were presented as multi-word phrases in a euphemistic manner. This approach found the correct euphemisms missing from our ground truth list via qualitative research.

Zhu et al. (2021) created unsupervised algorithms that identify euphemistically used phrases and uncover the underlying meaning of each word by analysing keywords within sentence-level contexts. An end-to-end pipeline was developed expressly for using context to discover euphemisms. For forums where text material is moderated, this was essential. By explicitly leveraging context, the pipeline may identify instances in which euphemisms are used to mask incorrect language. This makes content censoring more successful. This will improve the user experience overall and assist in identifying potentially objectionable terms. For the purpose of detecting euphemisms, three processes were proposed: gathering contextual information, eliminating circumstances that are not useful, and producing potential euphemisms.

The semantic and structural differences between euphemisms employed in the domains of business, health, and illnesses were examined by Inomovna (2022). Euphemisms were divided into six semantic groupings, according to Inomovna, including "profession," "death," "illness," "sex," "politics," and "crime." Linguistic context and physical context are the two main ways that pragmatics defines context. Pragmatics also focused on the roles that utterances have in informing, making promises, and making requests. They found that when a euphemism becomes widely used and is associated with a contentious or improper topic, it gets denigrated due to its negative connotations. Three linguistically motivated methods for euphemism identification in sentences, including possibly euphemistic words, were investigated by Tiwari & Parde (2022). These methods included Back-Translation, Paraphrasing Using Word

Sense-based Method, and one general-domain pre-trained language model. They asserted that a refined model that has been pre-trained on general-domain data may effectively recognise euphemisms.

A corpus of potentially euphemistic terms (PET) and an example text corpus from the GloWbE corpus were supplied by Gavidia et al. (2022). Additionally, a sub-corpus of writings that did not use euphemistic language to allude to PETs was provided. To support their views on how euphemisms soften language, two experiments were carried out: 1) A survey and observations of discrepancies when utilising euphemisms, and 2) Sentiment analysis using a ROBERTa-based model. Sentiment analysis of euphemistic messages revealed that, on average, PETs reduced offensive and negative attitudes. In a second annotation task, participants were asked to indicate whether or not PETs in a subset of our corpus text samples were euphemistic. The results showed disagreement.

Table: Performance Comparison of Deep Learning Models for Sentiment Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (mins)
LSTM	87.2	85.6	86.8	86.2	45
BiLSTM	88.1	86.9	87.3	87.1	50
GRU	86.7	84.8	86.0	85.4	38
CNN-LSTM	89.5	88.2	88.7	88.4	55
Transformer-based	91.3	90.5	90.9	90.7	70
BERT	93.7	92.8	93.2	93.0	120

2.0.Sarcasm analysis techniques :

Garcia et al. (2022) looked at the effect of the winking face emoji on the comprehension and perception of the message meaning for literal or satirical praise or criticism. The results demonstrated that older persons were less able than younger ones to distinguish and detect sarcastic intent in the absence of an emoji. They asserted that the winking face emoji facilitates successful intergenerational communication and makes up for the absence of non-verbal clues in textual communication.

Bharti (2017) used lexical, behavioural, hyperbolic, and universal truth-based text elements to identify sarcasm in tweets. To extract these aspects from the tweets, four algorithms have been proposed: tweet contradicting time-dependent facts (TCTDF), tweet contradicting universal facts (TCUF), parsing-based lexical generation algorithm (PBLGA), and likes and dislikes contradiction (LDC). Many machine learning classifiers, such as the support vector machine (SVM), maximum entropy (ME), Naive Bayes (NB), and decision tree (DT), were used to classify sarcastic tweets. Using the PBLGA approach, the decision tree yielded high accuracy and recall while the Naive Bayes classifier generated the lowest accuracy.

More emphasis was placed on the value of emoticons in sentiment analysis by Yadav & Pandya (2017). The impact of domain dependency, analysis level, intensifiers, count of emoticons, classifications, and emoticon position on sentiment analysis were studied. Additionally, other challenges related to sarcasm detection were investigated, such as managing acronyms, detecting sarcasm, multilingualism, lexical variety, handling slang languages, and dynamic dictionary management.

Mathur et al. (2017) proposed an audio mining approach combined with spectral data to recognise sarcastic speech. Pitch and tone are used in this conversation. Pitch changes were then identified using the MFCC. There are two phases involved in solving the sarcasm detection problem. Audio extraction converts any audio file into a.wav format as the initial step.

Pitch is then utilised as a key parameter to identify sarcasm in the extracted.wav file. Corpus-based analysis was used by Becker & Giora (2018) to look into the default hypothesis. A number of studies revealed that when presented with unclear data, participants tended to choose the default or more plausible interpretation. This was consistent with the default hypothesis, which holds that individuals typically process information in the most logical ways. Overall, the results showed that default interpretations were consistently processed more rapidly than sarcastic interpretations, supporting the default theory.

Swami et al. (2018) provided the first English-Hindi code-mixed dataset of tweets annotated with a language tag for each phrase and assessed for sarcasm and irony. A baseline classification technique for sarcasm detection in English-Hindi code-mixed tweets was presented, using features based on numerous words and characters. The methods for collecting and annotating language and sarcasm-related tweets at the tweet and token levels were explained. They asserted that features like word embeddings, POS tags, and other language-based components will improve the effectiveness of their classification system.

Identifying ironic text The authors Kumar & Harish (2018) proposed a two-phase feature selection approach. Using feature selection approaches like chi-square, information gain, and mutual information, the proper feature subset was first chosen. The most representative characteristics were selected in the second step using a K-means clustering approach. The selected characteristics were classified using a range of classifiers, including SVM and RF classifiers, utilising datasets of Amazon product reviews. They said that in addition to several classifiers, the suggested work may be extended to include n-grams, other feature selection techniques, and variance of k-means clustering methods.

Cai et al. (2019) presented a multi-modal hierarchical fusion model to identify sarcasm in messages on Twitter that include both text and graphics. Sarcasm was identified using text characteristics, picture features, and image attributes. This model extracts picture and attribute features before extracting text characteristics by using a bidirectional LSTM network and attribute features. Three modalities' worth of features are reconstructed and concatenated into one feature vector for prediction. A modality fusion layer computes a weighted average for the result and pushes it to a classification layer. They suggested extending the sarcasm detection system's medium to include audio.

Shmueli et al. (2020) presented a reactive supervision data gathering system that makes advantage of online interactions. Following the recommended technique, the first-of-its-kind large dataset of tweets with sarcastic perspective labels and enhanced contextual attributes, SPIRS, was created and made available. With this method, significant amounts of high-quality data on intentional and perceived sarcasm may be automatically gathered. They found that when the classifier learns discourse-related information (original tweet vs. reply tweet), mistakes may happen.

Jariwala (2020) looked on SVM models to identify sarcasm in news articles. News headlines were used as the source material for the feature extraction procedure. There are many sorts of characteristics, including pragmatic, sentimental, lexical, hyperbolic, and contradictory features. After features were retrieved, the best set of features was determined using a rule-based method. It shown that when the ideal feature set was provided as input, SVM performed better. It was suggested to incorporate word embeddings, POS tags, and other language-based elements to improve the effectiveness of sarcasm categorisation.

A combination of bidirectional LSTM memory with a softmax attention layer and convolution neural network was reported by Jain et al. (2020) for real-time sarcasm detection in code-switch tweets, particularly the mash-up of Hindi and English. Using pre-trained GloVe word embeddings with TF-IDF to extract the English semantic context vector and the subjective lexicon Hindi-SentiWordNet, an auxiliary pragmatic feature vector was developed that indicates the quantity of pragmatic markers in tweets.

The relationship between sarcasm and angry features was investigated by Szymaniak & Kaowski (2020). Two regression models were used to examine the predictive power of anger-related traits and general sarcasm use on self-

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 08 | Aug - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

reported sarcasm. Because gender matters for sarcasm, it was considered in both models. They suggested that more complex tasks, such as more detailed/varied contexts, open-ended questions, or measuring interpretation controlling for variables other than gender, would be helpful in determining whether an experimentally manipulated state of anger has concurrent effects on the use and interpretation of sarcasm.

Rahaman and colleagues (2021) proposed a practical method for identifying sarcastic tweets. Lexically focused features, sarcastic-based features, contrast-related features, and context-relevant features were among the several feature types that were taken into account. Sarcastic tweets were identified using a number of supervised machine-learning models constructed using extracted data. They found that the sarcastic-based feature set yields the best accuracy when the other three characteristics are used independently. When compared to various combinations and solo feature sets using the random forest model, the combinations of these four feature sets yielded the greatest accuracy.
According to Razali et al. (2021), contextual handcrafted features like stop words, positive and negative words, hyperbolic words, calendar words, nouns, verbs, pronouns, event words, and two dislike features are combined with extracted deep learning generated features to detect sarcasm in tweets more accurately than using only deep learning generated features. This is so that contextual characteristics created by hand may catch linguistic subtleties that deep learning feature extraction cannot. It also demonstrates the versatility of deep learning systems. When all of these characteristics were merged, a high degree of accuracy was attained.

Employing the fundamental cognitive characteristics of human speech, Hiremath & Patil (2021) presented an approach for detecting sarcasm in human communication by capturing three forms of data: voice, text, and temporal facial expressions. Sarcasm in multimodal data was identified by analysing text, vocal glottal signals, ocular movement, and cognitive traits. This method makes use of machine learning methods to find patterns in the data and classify it as sardonic or non-sarcastic information. Researchers have used a novel feature extraction approach to analyse temporal face data, which has improved the machine's ability to recognise visual signals that indicate sarcasm. Using cloud resources, the multiclass neural network model was used as a soft cognition method for sarcasm detection.
Eke et al. (2021) looked at a number of feature engineering techniques to solve problems including the loss of semantic meaning in sarcastic utterances, the sparsity of the training data, and the feature vector null values in the automatic detection of sarcasm. Machine learning methods were used to create a multi-feature fusion framework in two steps. Lexical characteristics, the microblog's duration, hashtag features, discourse marker features, emoticon features, syntactic features, pragmatic features, word cloud features, and emoticon features were among the nine categories of features that were retrieved and fused.

sentiment features and embedding features. The random forest classifier using the Pearson correlation feature selection technique yielded the best accuracy level on the snark dataset.

Chia et al. (2021) looked at irony and sarcasm on Twitter using feature engineering and machine learning approaches. The differences between cyberbullying, irony, and sarcasm were made clear and recognised. It was looked at how different NLP techniques may be used to identify irony and sarcasm. They found that social media cues, including the #hashtags on Twitter, had a big impact on the categorisation outcomes. Regarding accuracy, CNN with Rectified Linear Units (ReLU) outperformed social media categorisation tasks.

In order to distinguish between texts that are sarcastic and those that are not, Meriem et al. (2021) developed a fuzzy logic approach and categorised the results using social information such as responses, earlier tweets and likes, etc., increasing each by a degree of importance. For the purpose of sarcastic tweet recognition, text features, history characteristics, identification features, and response features were taken into account. To ascertain the importance of each attribute, the sarcastic score of every tweet was computed, disregarding each occurrence of the trait.

Several supervised classification algorithms and the best techniques for feature selection were found by Dave & Desai (2016). These techniques included point-wise mutual information and chi-square for textual data sarcastic detection. Textual information from social media platforms, microblogging sites, and websites connected to multilingual reviews was analysed via the use of categorisation techniques. They found that additional pragmatic, lexical, affective, and other relevant characteristics are needed for sarcasm recognition, and that the simple Bag-of-Words technique is inadequate in this regard.

Desai & Dave (2016) produced satirical Hindi sentences. They tried to detect sarcasm in two types of sardonic statements: those with markers and those without. Certain markers were used to identify sarcastic statements, including emoji, #tags, and punctuation. The training data was classified into five categories: non-sarcastic, mildly positive sarcastic, very positive sarcastic, mildly negative sarcastic, and extremely negative sarcastic. They found that although using external markers is helpful, doing it alone is more challenging. Karamouzas et al. (2022) suggested a novel neural knowledge transfer configuration for sentiment analysis of figurative texts using Deep Neural Networks. Training multi-class perceptual categorisation in a multitasking context may be used to extract information from two pre-trained figurative language abilities and indirectly solve comprehension difficulties by addressing what is presented when figurativeness is supplied.

Razali et al. (2022) examined the challenge of automatically recognising humorous situations in short writings. In order to extract the best features possible, they also examined how to combine a deep learning architecture with painstakingly constructed contextual information. It is found that certain sets work better when utilised as independents than when left on their own. Finally, the aggregated feature sets are classified using well-liked machine learning classification approaches. It is shown that logistic regression is the best method for this particular situation. Their findings suggest that merging characteristics derived by supervised learning with those retrieved manually might provide adequate performance.

Abulaish & Kamal (2018) introduced a unique self-deprecating sarcasm detection approach by combining rule-based and machine learning methods. While machine learning techniques are used to categorisation and feature extraction, rule-based solutions were designed to identify potential candidates inside tweets. Eleven features in total—six self-deprecating and five hyperbolic features—are discovered and used to train three different classifiers: bagging, naïve Bayes, and decision trees.

3.0.Study of different oxymoron phrases :

Karp (2021) used examples from George Orwell's Animal Farm to establish the lexical, syntactic, and semantic contexts of juxtapositions. It also emphasised the parallels and discrepancies between oxymorons, paradoxes, and antitheses in English and Ukrainian. George Orwell's Animal Farm was examined for oxymorons, paradoxes, and antitheses using deductive and taxonomic approaches. Recognised speech figures in both English and Ukrainian are located by contrastive analysis. The oxymoron had a number of structural models, such as free syntactic patterns, attribute, verbal, noun, and adverb pairs.

The study conducted by Bageshwar (2021) investigated the use of oxymorons, puns, repetition, wordplay, and metaphors in Instagram posts to generate distinct effects related to emphasis, persuasion, and emotion. The experts demonstrated how authors may use these techniques to create strong, memorable messages that resonate with readers. Presentational techniques were used to categorise Instagram posts. They said that readers are deeply impacted by the material and are able to engage with it. Eight forms of figurative language related to COVID-19 were recognised by Rezeki (2021) in poetry. These categories were metonymy, metaphors, personifications, similes, exaggerations, ironies, litotes, and oxymorons. They found that metonymy was the most often used figurative language in COVID-19 poetry. A descriptive qualitative research approach was used to carry out the studies. The five poems concerning COVID-19 were found to include two different kinds of messages: moral and societal. The effects of several adverb types on the categorisation of tweets as positive, negative, or neutral were studied by Haider et al. (2021). These adverb kinds included adverbs, degree adverbs, degree comparative adverbs, general adverbs, general comparative adverbs, locative adverbs, prepositional adverbs, and adverbs of time. They found that generic verbs and broad comparison adverbs were the most relevant polarity-bearing terms for neutral attitudes. They found locative and prepositional adverbs for views that were both positive and negative. They suggested that this work be extended to other situations in addition to product evaluations, scholarly papers, blogs, and news articles.

Tsao et al. (2020) investigated the oxymoron type's composition and produced a novel oxymoron. The contradictory image components of the creative oxymoron instances were studied using principal component analyses of contradictory-image testing for overall impressions of use, operating manner, and product feedback. In order to create design conversion models, the attributes corresponding to the parts of speech were determined. To identify the three oxymoron composition models that correspond with unique oxymoron samples, a relationship was found between overall perceptions of usage and paradoxical pictures. To do this, principal component scores were dispersed in picture spaces generated by three primary axes. Fazlitdinovna (2020) used proverbs to illustrate the meanings and applications of oxymorons in speech. Oxymorons' effects on teaching and learning in higher education settings were examined. To analyse the oxymoron, examples of phrases from poetry and literature in both Uzbek and English were collected. It was said that contemporary advertising often used oxymorons. Additionally, the author claimed that using oxymorons in advertisements might encourage strong emotional reactions among the target demographic.Furthermore, they might elicit amusement, surprise, and curiosity. Oxymorons may also be used to create a memorable phrase and draw attention to conceptual distinctions.

Safarovna (2020) focused on the origins of the oxymoron in Uzbek, gave instances of its use in literary works, and urged students to research the use of contrast in literary writings. Antonyms and oxymorons share a number of lexical-semantic traits. Nonetheless, their purpose, mode of expression, and use differed according to the speech patterns employed. A few things were made clear when the term "contrast" was spoken. Pietra & Masini (2020) extracted antonymous pairings from Italian corpora in order to identify a few oxymorons. They found that sarcasm, irony, and humour detection were much aided by contextual oxymorons. 376 oxymorons were classified using nine syntactic components and antonymous pairings.

4.0. Conclusion :

A unique approach was presented by Oktaviyani & Licantik (2020) to detect duplicate sentence pairs in Software Requirements Specification papers. Redundancy detection in software requirements specification papers was tested for efficacy using WordNet-based semantic similarity techniques. For this assignment, the following procedures were necessary.
Input sentence pairs were prepared, preprocessing was done, similarity values between sentences were computed, threshold values were selected, and redundancy was assessed.

Xue & Hwa (2014) have looked at redundancy detection in works that use Estimate the Likelihood. The redundancy metric assigns a high score to words and phrases that have a high probability of being redundant inside a sentence. Two main components made up the assessment: one used a language model to test fluency, and the other used word and translation alignments to measure meaning retention. Al-Nasrawi (2019) examined the realisation of pleonastic expressions in the chosen texts by discovering pleonastic phrases that might unearth hidden meanings even in the absence of certain words. Pleonasm in English may be further classified into two subtypes: syntactic and semantic. It was found that semantic pleonasm was more prevalent than syntactic pleonasm. This highlighted the need of taking into account both types of texts when doing text analysis. The Walking Dead pleonasm was categorised by Fitriansyah & Rosmaidar (2018) based on their categories. After extracting the subtitles, they saw the program, printed, converted, and read the subtitles. They discovered and produced pleonastic phrases, and as part of their research, they determined the kinds, subtypes, and purposes of pleonasms. In the TV series discussion, the most pleonastic subject pronouns are used, indicating that syntactic pleonasms are more common than semantic pleonasms.

Sethi (2019) investigated how Indians manage perceptions at work via email communication by using pleonastic English words and phrases. Syntax, semantics, and morphology were examined in relation to pléonasm. One strategy to enhance the recipient's perception of the sender in emails is to utilise pleonastic English. This was in the areas of deference, lucidity, specificity, and civility. A method for pleonasm detection in text was developed by Kashefi et al. (2018). In a newly annotated semantic pleonasm corpus, every phrase has a word pair that may be connected to each other. This corpus was assessed for inter-annotator agreement and utility in the development of automated pleonasm detections. The frequency of semantic pleonasms in institutional registries—that is, in English and Lithuanian papers from the European Union—was investigated by Kasperavičienė (2012). Plenasms are widely used in the institutional record; this is shown by papers from the European Union that are accessible to the public via the Eur-lex database.

5.0. **Acknowledgement :**

I am thankful to those who have helped me in my work .

6.0.References :

Baranova, SV & Pletenko, K 2018, _Pragmatic and functional peculiarities of euphemisms for diseases, death and dying in the English mass-media discourse‘, Filologicni Traktati, vol. 10, no. 1, pp. 19-25.

Almoaвidi, KA 2018, _Euphemism as a communicative tool: A descriptive study of Hijazi and southern region dialects spoken inSaudi Arabia‘, Open Journal of Modern Linguistics, vol. 8, no. 1, pp. 1-8.

Olimat, SN 2020, _COVID-19 pandemic: euphemism and dysphemismin Jordanian Arabic‘, Journal of Language Studies, vol. 20, no. 3, pp. 268-90.

Matondang, Z & Purnanto, D 2020, _A Meaning Component Analвsis of Euphemism and Dвsphemism in Indonesian Da‘aah‘, International Journal of Linguistics, Literature and Translation, vol. 3, no. 9, pp. 58-65.

Cao, Y 2020, _Analвsis of pragmatic functions of English euphemism from the perspective of pragmatic principles‘, Theorв and Practice in Language Studies, vol. 10, no. 9, pp. 1094-1100.

Hasyim, K, Nurkamto, S & Santosa, R 2020, _Lexical Creation of Euphemism and Dysphemism in Online Media Text on Presidential Election 2019‘, In Proceedings of the Third International Seminar on Language, Literature, and Local Culture Studies, pp. 1-6

Terry, A 2020, _Euphemistic dysphemisms and dysphemistic euphemisms as means to convey irony and banter‘, Language and Literature, vol. 29, no. 1, pp. 57-75

Ariani, NP, Beratha, NLS & Malini, NLNS 2020, _Semantic Changes in Translation of Euphemism and Dвsphemism in Tempo Magazine‘, Research and Innovation in Language Learning, vol. 3, no. 2,pp. 107-122

Felt, C & Riloff, E 2020, _Recognizing euphemisms and dysphemismsusing sentiment analysis‘, In Proceedings of the Second Workshop on Figurative Language Processing, pp. 136-145.

Kafi, LN & Degaf, A 2021, _Euphemism and dвsphemism strategies inDonald Trump‘s speech at SOTU 2020‘, International Journal of Humanity Studies, vol. 4, no. 2, pp. 194-207.

Aytan, A, Aynur, B, Hilal, P, Aytac, E & Malahat, A 2021,

‗Euphemisms and dysphemisms as language means implementing rhetorical strategies in political discourse', Journal of Language and Linguistic Studies, vol. 17, no. 2, pp. 741-754

Yildiz, FU 2021, ‗Use of Euphemisms in Youth Language', Journal of

Language and Linguistic Studies, vol. 17, pp. 1117-1128

Zhu, W & Bhat, S 2021, ‗Euphemistic phrase detection by masked

language model', arXiv preprint arXiv:2109.04666.

Inomovna, HI 2022, ‗Semantic and Structural Differentiation of Euphemisms', Journal of Pedagogical Inventions and Practices, vol. 9, pp.161-166.

Tiwari, D & Parde, N 2022, ‗An Eбploration of Linguisticallв-Driven and Transfer Learning Methods for Euphemism Detection', In Proceedings of the Third Workshop on Figurative Language Processing. Association for Computational Linguistics, pp.131-136.

Gavidia, M, Lee, P, Feldman, A & Peng, J 2022, ‗CATs are Fuzzy PETs: A Corpus and Analвsis of Potentiallв Euphemistic Terms', arXiv preprint arXiv:2205.02728.

Garcia, C, Ţurcan, A, Hoaman, H & Filik, R 2022, ‗Emoji as a tool to aid the comprehension of written sarcasm: Evidence from younger and older adult', Computers in Human Behavior, vol. 126, pp. 106971

Bharti, SK, Pradhan, R, Babu, KS & Jena, SK 2017, ‗Sarcasm analвsis on taitter data using machine learning approaches', Trends in Social Network Analysis, pp. 51-76

Yadav, P & Pandвa, D 2017, ‗SentiReviea: Sentiment analвsis based on teбt and emoticons', In Proceeding of the International Conference on Innovative Mechanisms for Industry Applications, pp. 467-472.

Mathur, A, Saxena, V & Singh, SK 2017, ‗Understanding sarcasm in speech using mel-frequencв cepstral coefficent', In Proceedings of 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, pp. 728-732.

Becker, I & Giora, R 2018, ‗The Defaultness Hвpothesis: Aquantitative corpus-based study of non/default sarcasm and literalness production', Journal of pragmatics, vol. 138, pp. 149-164

Swami, S, Khandelwal, A, Singh, V, Akhtar, SS & Shrivastava, M 2018, ‗A corpus of English-Hindi code-mixed tweets for sarcasm detection', arXiv preprint arXiv:1805.11869.

Cai, Y, Cai, H & Wan, X 2019, ‗Multi-modal sarcasm detection in taitter aith hierarchical fusion model', In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 2506-2515.

Shmueli, B, Ku, LW & Raв, S 2020, ‗Reactive supervision: A nea method for collecting sarcasm data', arXiv preprint arXiv:2009.13080

Jariwala, VP 2020, ‗Optimal feature extraction based machine learningapproach for sarcasm tвpe detection in neas headlines', Journal of Computer Application, vol. 975, pp. 25-29.

Jain, D, Kumar, A & Garg, G 2020, ‗Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature- rich CNN', Applied Soft Computing, vol. 91, pp.106198.

Szymaniak, K & Kałoaski, P 2020, _Trait anger and sarcasm use‘, Personality and Individual Differences, vol. 154, pp. 109662

Rahaman, A, Kuri, R, Islam, S, Hossain, MJ & Kabir, MH 2021,

_Sarcasm Detection in Taeets: A Feature-based Approach using Supervised Machine Learning Models‘, International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 454-460

Razali, MS, Halin, AA, Ye, L, Doraisamy, S & Norowi, NM 2021,_Sarcasm detection using deep learning aith conteбtual features‘, IEEE Access, vol. 9, pp. 68609-68618.

Eke, CI, Norman, AA & Shuib, L 2021, _Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach‘, Plos one, vol. 16, no. 6, pp. 0252918.

Chia, ZL, Ptaszynski, M, Masui, F, Leliwa, G & Wroczynski, M 2021,_Machine Learning and feature engineering-based study into sarcasm and ironв classification aith application to свberbullвing detection‘, Information Processing & Management, vol. 58, no.4, pp. 102600

Meriem, AB, Hlaoua, L & Romdhane, LB 2021, _A fuzzy approach forsarcasm detection in social networks‘, Procedia Computer Science, vol. 192, pp. 602-611.

Desai, N & Dave, AD 2016, _Sarcasm detection in hindi sentences using support vector machine‘, Advance Research in Computer Science and Management Studies, vol. 4, no. 7, pp. 8-15.

Razali, MS, Halin, AA, Ye, L, Norowi, NM & Doraisamy, S 2022,_Context-Driven Satire Detection with Deep Learning‘, IEEE Access, vol. 10, pp. 78780-78787

Abulaish, M & Kamal, A 2018, _Self-Deprecating Sarcasm Detection: An Amalgamation of Rule-Based and Machine Learning Approach‘, Proceedings of the IEEE /WIC/ACM International Conference on WebIntelligence, pp.574-579

Karp, MA 2021, Figures of contrast in Animal Farm by George Orwell: lexical and semantic features of oxymoron, paradox and antithesis, Baltija Publishing, pp. 232-249.

Bageshwar, BR 2021, _A Study of Literary Devices used in the Posts on Instagram‘, International Journal of English Literature and Social Sciences, vol. 6, no.1, pp. 414-421

Rezeki, S 2021, _Figurative Language and Messages in Poems About COVID-19‘, KnE Social Sciences, pp. 601-611.

Haider, S, Tanvir Afzal, M, Asif, M, Maurer, H, Ahmad, A & Abuarqoub, A 2021, _Impact analвsis of adverbs for sentiment classification on Taitter product revieas‘, Concurrency andComputation: Practice and Experience, vol. 33, no.4, pp. 1-15.

Hiremath, BN & Patil, MM 2021, _Sarcasm detection using cognitive features of visual data bв learning model‘, Eбpert Sвstems aith Applications, vol. 184, pp. 1-8.

Fazlitdinovna, KV 2020, _Oxymoron and its methodological-semantic feautures‘, BeïöÖóк ÖĭÜкó ó Üßëĭ£Üçĭ̈Ööя, vol. 7, no. 2, pp. 114-119

Safarovna, SN 2020, _Oбвmoron as a method of poetic perception of the aorld‘, Journal of Critical Revieas, vol. 7, no. 7, pp. 340-344.

Oktaviyani, ED 2020, _Redundancy Detection of Sentence Pairs In The Software Requirements Specification Documents With SemanticApproach‘, In IOP Conference Series: Materials Science and Engineering, vol. 879, no. 1, pp. 1-8.

Al-Nasraai, RDS 2019, ‗The Studв of Pleonasm in Some Selected Teбts of Shakespeare‘s Drama: Julius Caesar‘, Journal of Universitв ofBabylon for Humanities, vol. 27, no. 6, pp. 58-68

Fitriansyah, I & Rosmaidar, R 2018, ‗Pleonasm in the tv series the aalking dead‘, Jurnal Ilmiah Bina Bahasa, vol. 11, no. 2, pp. 34-44

Sethi, D 2019, ‗Workplace impression management through pleonastic English words/phrases in e-mail communication? Evidence from India‘, International Journal of Indian Culture and Business Management, vol. 18, no. 4, pp. 391-404.

Kasperavičienė, R 2012, ‗On semantic pleonasms and their translation in institutional register‘, Studies about Languages, vol. 21, pp. 57-63.