HUMOUR DETECTION USING NLP

Amrita Veshin
22122104
3MScDS-B

Atharva Vetal
22122109
3MScDS-B

PROBLEM STATEMENT

The field of Natural Language Processing (NLP) has experienced a notable surge in interest in the domain of humour detection, aiming to computationally discern the nuanced aspects of humour in text. This research addresses the multifaceted challenge of understanding and identifying humour in written content, centring around jokes. The ultimate goal of this study is to advance our comprehension of humour detection using NLP techniques and provide practical solutions for applications that require the recognition of humour in text.

KEY OBJECTIVES

1. **Exploration of Humour in Text Data:** The primary objective of this research is to delve into the distinctive characteristics of humour, encompassing elements like wordplay, punctuation ratio, and tag ratio, by conducting an Exploratory Data Analysis (EDA) on a dataset of textual content classified as either humorous (jokes) or non-humorous (plain text).
2. **Feature Engineering for Humour Identification:** The study aims to develop effective feature engineering techniques, including Part-of-Speech (POS) tagging, tokenization, and punctuation ratio analysis, to gain insights into linguistic and structural attributes that distinguish humorous texts from non-humorous ones.
3. **Visualizing Elements that Construct Humour:** By constructing box plots and other visual representations, the research intends to depict the distribution of linguistic and structural features across humorous and non-humorous text categories, shedding light on the differences that contribute to humour identification.
4. **Supervised Learning Models:** Employing various supervised learning techniques, such as the Naive Bayes classifier, Random Forest, and advanced Pre-Trained Language Models like XLNet and BERT, the study seeks to classify text as humorous or non-humorous, providing a systematic approach for automating humour detection.
5. **Comparative Analysis of Models:** The research will undertake a comprehensive comparative analysis of the performance of different models in humour classification, aiming to determine which techniques and architectures are most effective for identifying humour in text.

SIGNIFICANCE

This research aims to bridge the gap between computational systems and human humour perception, contributing to the evolving field of humour detection using NLP. Humour Detection has numerous real-world applications, but is a challenging task in NLP. Understanding and automating the recognition of humour in text can enhance various applications such as sentiment analysis, chatbots, content filtering, and recommendation systems. Furthermore, the study's findings will provide valuable insights into the linguistic and semantic aspects of humour, advancing our comprehension of this complex facet of human communication.

ABSTRACT

The field of Natural Language Processing (NLP) has witnessed a surging interest in the elusive realm of humour detection. Humour, deeply rooted in the human experience, has posed a unique challenge for computational systems. Researchers have embarked on diverse endeavours to computationally recognize humour, ranging from the exploration of wordplay within jokes to the identification of humorous texts and satire in social media. These studies span a spectrum of methodologies and approaches, incorporating theories of humour, linguistic and semantic features and machine-learning techniques. From the computational recognition of wordplay through Raskin's theory of humour to the employment of word-association-based semantic features that surpass traditional methods, the research landscape is rich and diverse. Generative language models and pre-trained language models, such as BERT and Transformer architectures, play pivotal roles in this domain, offering novel avenues for humour detection. Furthermore, comparative studies explore the intersection of irony detection and established humour classification techniques, shedding light on their efficacy. This diverse array of methodologies and models collectively contributes to the evolving field of humour detection using NLP.

LITERATURE REVIEW

The realm of NLP has witnessed a growing fascination with the intricate and often elusive domain of humour detection. Researchers have embarked on diverse endeavours to computationally recognize humour, ranging from wordplay in jokes and satire in social media to the identification of humorous text in online reviews and Twitter posts. These studies delve into a variety of methods and approaches, encompassing theories of humour, linguistic and semantic features, machine learning techniques, and the integration of external text sources. As we delve into the literature on Humour Detection using NLP, we encounter a mosaic of innovative strategies and theoretical foundations that collectively advance our understanding of how machines can decipher the nuances of human laughter in the written word.

In their study titled "Computationally Recognizing Wordplay in Jokes," Taylor and Mazlack explore the computational recognition of humour, with a specific focus on wordplay within jokes. Their methodology concentrates on a select category of jokes, particularly "Knock Knock" jokes, and grounds itself in Raskin's Semantic Theory of Verbal Humour as the theoretical framework. The research employs an algorithm that learns statistical text patterns in N-grams and utilizes a wordplay generator to produce utterances resembling given words phonetically. Subsequently, a wordplay recognizer evaluates the validity of these utterances, and a joke recognizer assesses whether the wordplay transforms the text into a joke. The joke recognition process encompasses four steps, including joke format validation and wordplay sequence validation, concluding with last sentence validation. Furthermore, a punchline recognizer is designed to determine the meaningfulness of the last sentence with or without wordplay. Although the study exhibits reasonable success in recognizing wordplay, its proficiency in validating the utterances is comparatively less robust (Taylor & Mazlack, 2004)[1].

Word-association-based semantic features provide another novel approach to humour detection, surpassing traditional Word2Vec similarity methods. They excel in capturing intricate word relationships and contextual nuances, thus boosting the system's humour recognition capabilities. Cattle and Ma introduce this approach through their paper, "Recognizing Humour using Word Associations and Humour Anchor Extraction,". Interestingly, attempts to enhance these features with humour anchors were found to be counterproductive. Their word association-based humour classification system was evaluated on the Pun of the Day (PotD) and 16000 One-Liner (OL) datasets. Additionally, the paper establishes a baseline using document perplexity derived from a 3-gram language model trained on the WMT15 English news discussion corpus. The authors compare their approach to a baseline model inspired by Yang et al. (2015), encompassing various features, such as Word2Vec similarities, WordNet path similarities, sentiment lexicons, alliteration, rhyme chains, and word frequencies. These findings provide valuable insights into computational humour recognition, contributing to the development of more robust systems (Cattle & Ma, 2018)[2].

Several studies in the literature have consistently compared the performance of irony detection models with established humour classification techniques to evaluate their effectiveness. In one such study, "Automatic Detection of Irony and Humour in Twitter," authors Barbieri and Saggion explore the automatic detection of irony and humour on the social platform Twitter. They employ a classification-based approach and introduce an extensive array of features for text interpretation and representation, successfully improving state-of-the-art performance in cross-domain classification experiments. Their dataset comprises 40,000 tweets spanning four topics: Irony, Education, Humour, and Politics, utilizing features such as frequency, written-spoken style, intensity, structure, sentiments, synonyms, and ambiguity. Additionally, the paper delves into figurative language filtering and the information gain associated with each feature. This research provides valuable insights into the realm of humour and irony detection in Twitter, contributing to the field of natural language processing (Barbieri & Saggion, 2014)[3].

Authors Barbieri, Ronzano, and Saggion provide a very interesting and unique perspective on humour detection, by highlighting humour and irony as weapons deployed for criticism across domains. Through their research paper, "Do we criticise (and laugh) in the same way? Automatic Detection of Multi-lingual Satirical News in Twitter,", they undertake a comprehensive exploration of automatic satirical news detection across multiple languages, including English, Spanish, and Italian, using language-independent (word usage frequency, etc.) and language dependent features (lemmas, bigram, skip-gram). Their approach hinges on a sophisticated modelling of tweets, incorporating lexical, semantic, and usage-related features, enabling the system to effectively classify tweets into satirical and non-satirical categories. Impressively, this system outperforms a baseline reliant on mere word-based analysis, attaining robust accuracy in discerning satirical content on Twitter. The study's notable contribution lies in its cross-lingual analysis, revealing the variances and commonalities in the utilization of satire within distinct linguistic contexts. Moreover, the authors introduce a novel and adaptable framework for satire detection on the Twitter platform and offer a valuable dataset for satirical and non-satirical news tweets in English, Spanish, and Italian (Barbieri, Ronzano, & Saggion, 2015)[4].

Transformer-based models have played a crucial role in understanding humour as a language element throughout the literature. Weller and Seppi through their paper "Humour Detection: A Transformer Gets the Last Laugh," introduce a Transformer architecture-based model that learns to identify humorous jokes using ratings from Reddit pages and demonstrates effectiveness comparable to human performance. It surpasses previous work in humour identification, achieving F-measures of 93.1% for the Puns dataset and a remarkable 98.6% for the Short Jokes dataset. The paper acknowledges diverse prior methods in humour identification, ranging from statistical and N-gram analysis to convolutional neural networks. It acknowledges the inherent challenge of defining the universal humour value of a joke but highlights the model's ability to identify jokes that resonate humorously with specific subsets of the population. Additionally, the paper underscores the significance of Reddit's rJokes thread as a valuable resource for gauging responses to jokes in a large group setting. Leveraging advancements in natural language processing and neural network architecture, this study explores humour in areas such as classification, generation, and social media (Weller & Seppi, 2019)[5].

Generative language models in NLP play a vital role in humour detection by facilitating the recognition of nuanced linguistic constructs, wordplay, and incongruities, enabling the detection of humour in text. Researchers Morales and Zhai employ generative language models in their paper "Identifying Humour in Reviews using Background Text Sources,". They utilize background text sources like Wikipedia descriptions to create effective features for humour identification. These features outperform existing literature, achieving an impressive 86% accuracy in classifying reviews as humorous or not. Additionally, the study highlights the broader applicability of humour predictions in identifying helpful reviews. The paper leverages recent advances in Wikification to access Wikipedia pages of entities mentioned in the reviews, computes language models for each aspect using Wikipedia descriptions, and demonstrates the superiority of unigram features over content-based features. By focusing on external text sources, this research advances our understanding of

humour detection and offers valuable insights into the selection of features employed in previous work (Morales & Zhai, 2017)[6].

Pre-trained language models such as BERT are quite popular in the field of humour detection. In the paper titled "ColBERT: Using BERT Sentence Embedding for Humour Detection," author Annamoradnejad introduces an innovative method for humour detection in short texts, employing BERT sentence embeddings and a neural network with parallel hidden layers. This approach achieves outstanding results, with a 98.2% F1 score in humour detection. The authors construct a substantial dataset of 200,000 formal short texts for balanced evaluation. Their 8-layer model with 110 million parameters surpasses baseline models, emphasizing the importance of incorporating text structure in machine learning. Preprocessing steps, including contraction expansion and punctuation cleaning, are detailed. The dataset is rigorously curated by removing duplicates and filtering based on character and word length, with news headlines converted to sentence case formatting. Notably, the dataset's size and diversity contribute to its effectiveness, as character count and sentiment features do not significantly correlate with the target value (Annamoradnejad, 2020)[7].

The worthiness of Deep Neural Networks (DNN) in this domain has been adeptly demonstrated by Authors Miraj and Aono through their paper titled "Integrating Extracted Information from BERT and Multiple Embedding Methods with the Deep Neural Network for Humour Detection,". They introduce a novel framework called IBEN, designed for humour detection in short texts sourced from news headlines. This framework ingeniously combines information extracted from various layers of BERT, employing a Bi-GRU neural network along with external embedding models and multi-kernel convolution techniques to generate higher-level sentence representations. The primary goal of IBEN is to assess the level of funniness in written sentences, making it a valuable foundation for generating humorous text. By fusing deep learning techniques, including multi-kernel convolution, Bidirectional GRU, and BERT, the framework adeptly captures contextual information, significantly enhancing humour detection performance. The paper provides comparative results for both original and edited news headlines, highlighting their most effective approach. It also delves into the utilization of word embedding techniques like Word2Vec, Glove, and FastText, which play a pivotal role in capturing word context and semantic similarity within the proposed framework, showcasing its comprehensive approach to humour detection (Miraj & Aono, 2021)[8].

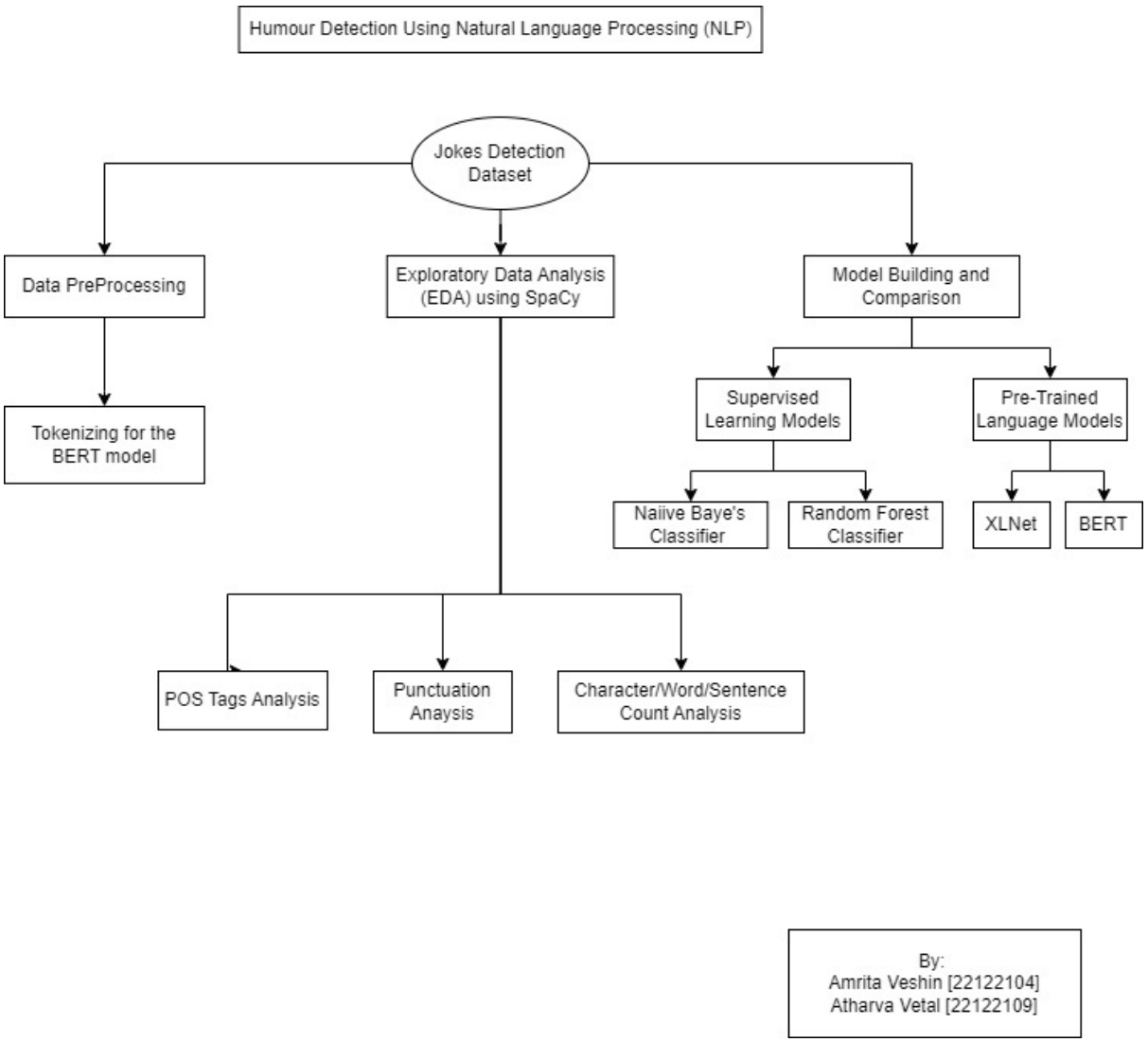
This research endeavours to comprehensively explore the intricate world of humour detection in textual data. It encompasses multiple key objectives, including the analysis of various features such as wordplay, punctuation and tag ratios, the application of feature engineering techniques such as POS tagging and tokenization, visualization of humour elements through visual representations like box plots, utilization of supervised learning models ranging from traditional classifiers to advanced Pre-Trained Language Models (PTLMs) like XLNet and BERT for automated humour detection, and conducting an in-depth comparative analysis to evaluate the effectiveness of different models. These objectives collectively aim to advance the understanding of humour detection in natural language text, contributing to the broader field of Natural Language Processing and its real-world applications.

ABOUT THE DATASET

The dataset that has been used in this research has been taken from Kaggle, titled ‘Jokes Detection’. It is a very popular dataset and is widely used by researchers in the NLP domain. It has 2,00,000 unique text records, which are classified as humorous or non-humorous (joke or plain text) via the ‘humour’ variable. The dataset is balanced, containing 50% True and 50% False values for the humour variable. Following is the snapshot of the data as a pandas data frame:

	text	humor
0	Joe biden rules out 2020 bid: 'guys, i'm not r...	False
1	Watch: darvish gave hitter whiplash with slow ...	False
2	What do you call a turtle without its shell? d...	True
3	5 reasons the 2016 election feels so personal	False
4	Pasco police shot mexican migrant from behind,...	False

RESEARCH DESIGN FLOWCHART



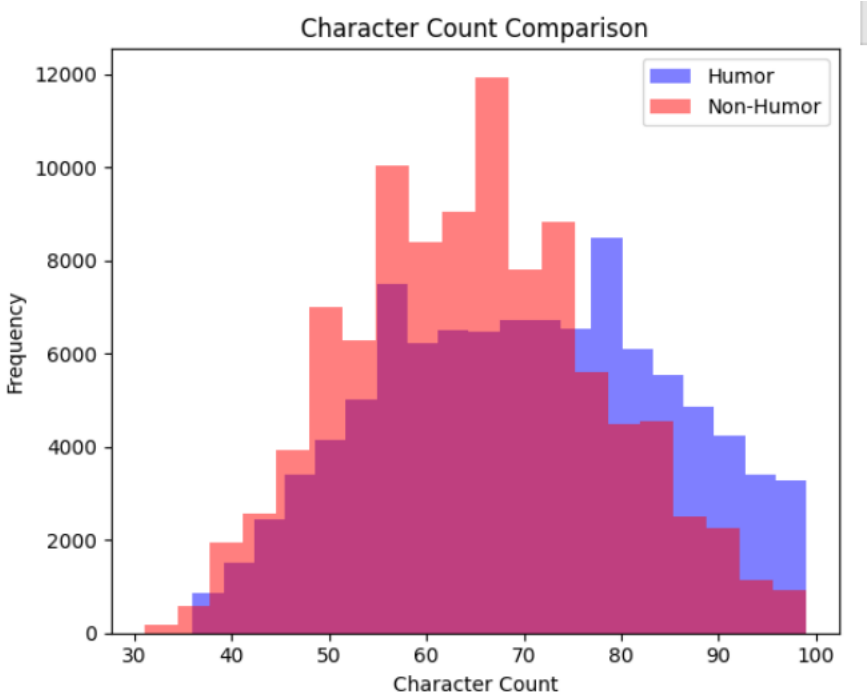
DATA PREPROCESSING

Tokenization: Tokenizing in NLP is the process of breaking down text into individual units, or tokens, like words or sub-words. Transformers, such as BERT, use sub-word tokenization to handle various languages and morphological variations efficiently. Tokenization in BERT is critical because it allows text to be transformed into numerical representations while preserving the context. These tokenized sequences are then used as input to deep neural networks, enabling BERT to capture rich contextual information and achieve state-of-the-art results in various NLP tasks. Tokenizing in BERT

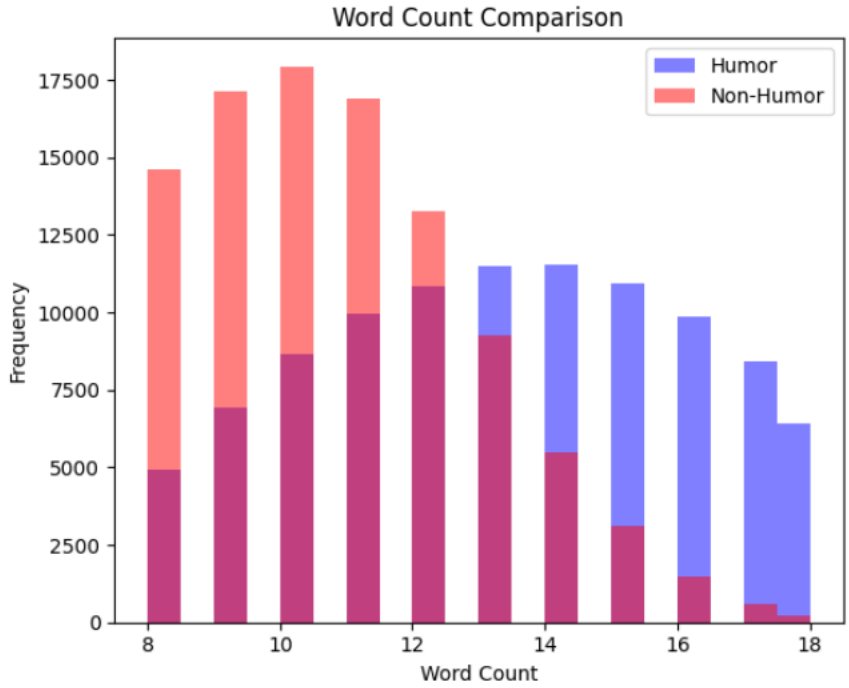
architecture serves as the initial step, allowing the model to comprehend language and extract intricate patterns within the text.

EXPLORATORY DATA ANALYSIS (EDA)

1. Wordplay Comparisons between Humorous and Non-Humorous Texts

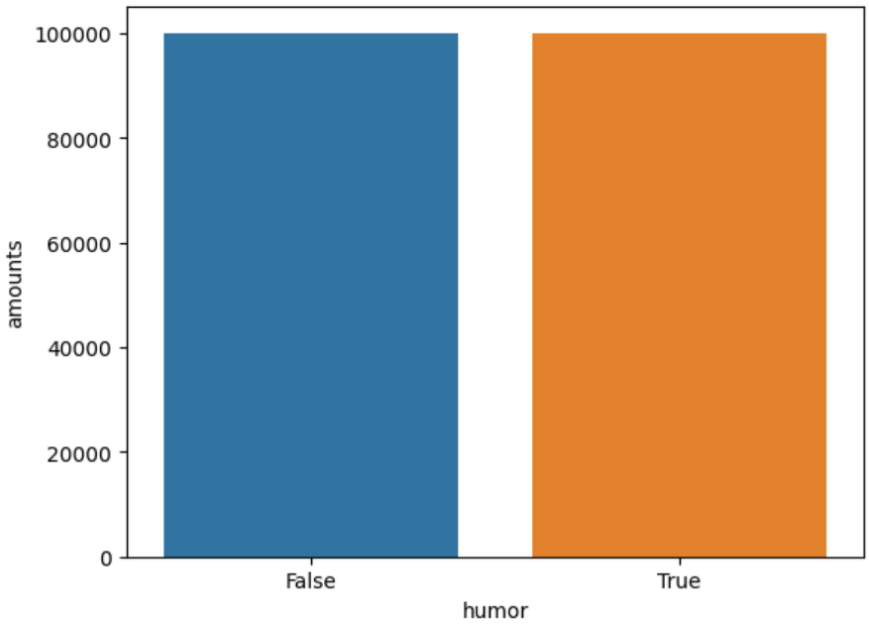


Taking a look at the character count comparison histogram, according to their frequencies, we observe that more no. of non-humorous texts have a lower character count as compared to the humorous texts, which tend to have a higher character count.



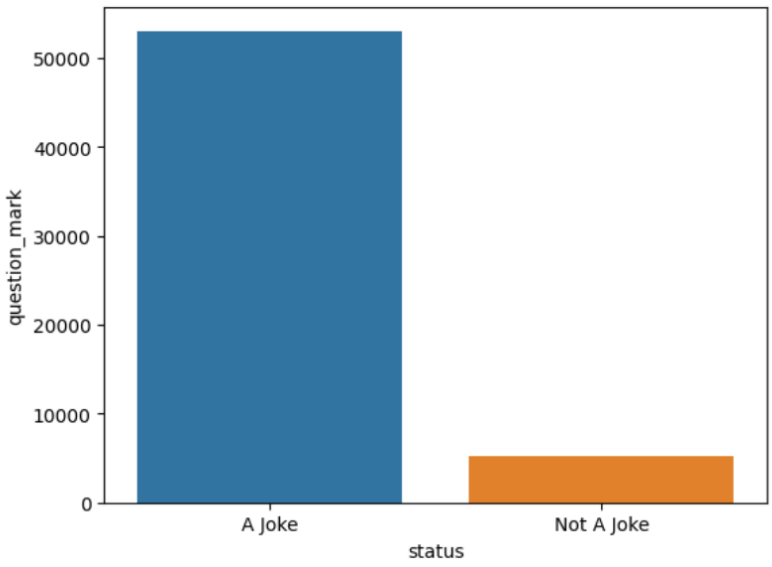
From the above Word Count Comparison Graph, we can observe that the frequency of low-word-counted (8 to 12) non-humorous texts is quite higher than the humorous texts, which tend to have a higher word count comparatively.

2. **How Proportionate is the Data**



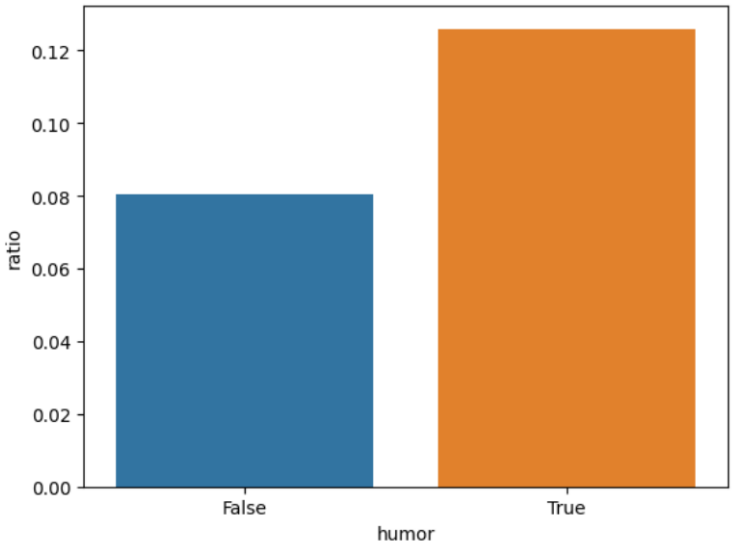
We can see from the above plot that the data is balanced between the two categories humorous and non-humorous texts. Having an equal balance between jokes and non-jokes is crucial. Understanding the data proportion is vital because if it's significantly imbalanced, the model may become biased towards the category with more data, impacting its performance. In such cases, the model is more likely to favour the label with a larger training dataset due to a better understanding of the same.

3. **How Frequently Does the Question Mark '?' Symbol Appear**

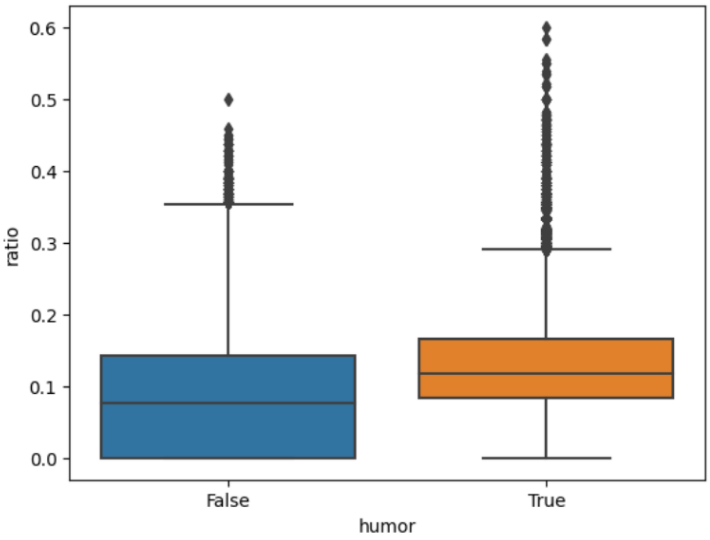


From the above barplot, it's evident that the question mark is more prevalent in data labelled as jokes. Out of 100,000 data samples, over 50,000 contain a question mark, while those labelled as "_not_ajoke" consist of only around 5,000. Consequently, we can conclude that text classified as jokes has a greater tendency to include question marks.

4. **Punctuations Analysis**

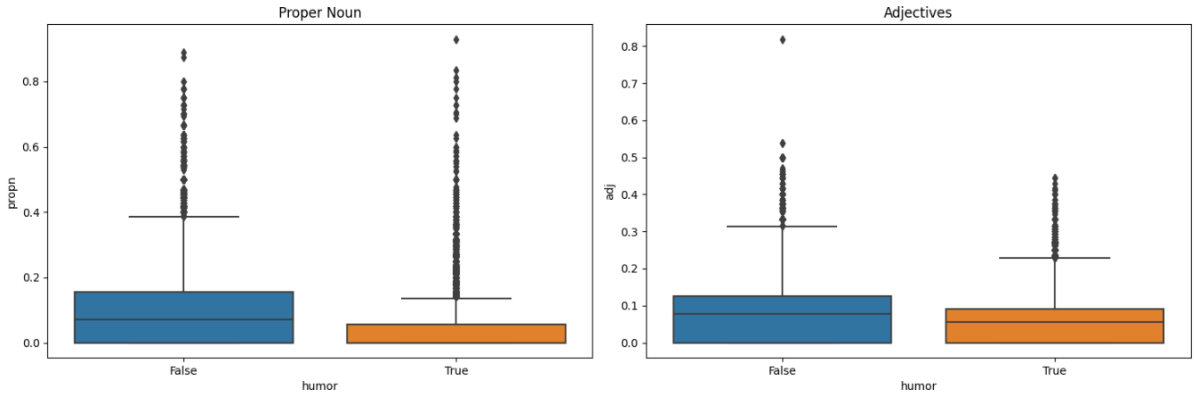


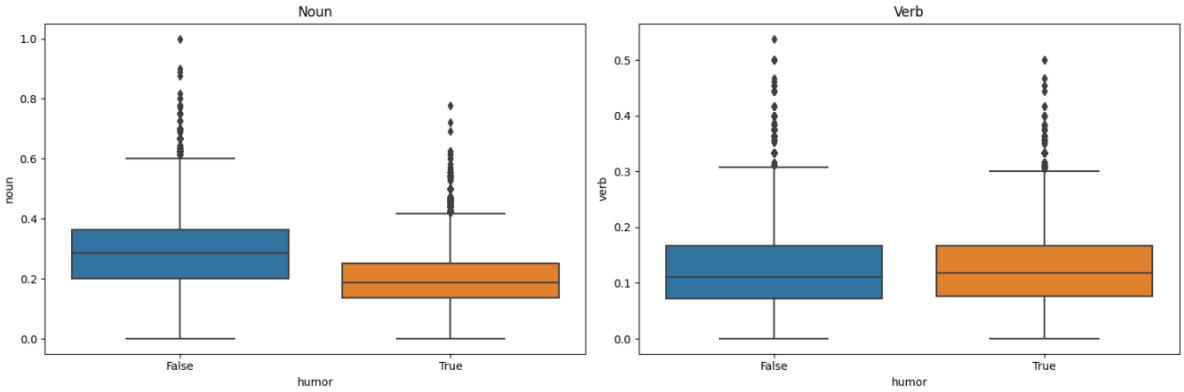
From the above plot, we can observe that the ratio of punctuations (0.12) is greater in humorous texts as compared the non-humorous texts (0.08), validating our previous observation as well.



The boxplot above reveals that, on average, humorous content contains more punctuation marks compared to non-humorous content. It's worth noting that jokes often have a specific structure. A well-crafted joke is expected to be somewhat unpredictable, but it should still adhere to a recognizable structure commonly found in modern comedy. This structure typically involves a series of setups leading to a punchline. To maintain this structural order, punctuation marks are frequently used. This explains why punctuations are more prevalent in jokes.

5. POS Tags Analysis





In the comparison above, we have examined the frequency of Nouns, Proper Nouns, Verbs, and Adjectives in texts with the help of boxplots. This analysis provides insights into the characteristics of jokes.

Based on the visualizations, it is evident that non-joke texts tend to contain a higher number of Proper Nouns, Adjectives, and Nouns. This suggests that non-joke texts may lean towards formality, resulting in an increased occurrence of these language elements. Consequently, we can infer that the formality of a text is associated with a greater presence of Proper Nouns, Adjectives, and Nouns.

CLASSIFYING TEXTS AS HUMOROUS AND NON-HUMOROUS

1. BERT Pre-Trained Language Model

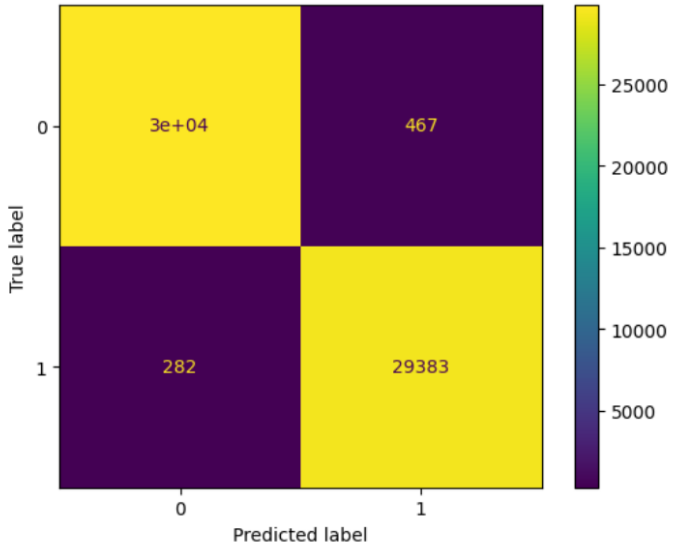
BERT, or Bidirectional Encoder Representations from Transformers, is a leading NLP model. It excels in humour detection by understanding the context and nuances of language.

In humour detection tasks, BERT is fine-tuned using labelled datasets that include humorous and non-humorous text examples. BERT's bidirectional nature enables it to recognize linguistic patterns and context that contribute to humour, making it effective at identifying humour in text, such as jokes in social media posts, funny content in scripts, or humour in chatbots.

i. Preprocessing: Cleaning and preparing the dataset (if required).
ii. Tokenization: Tokenize the text using a BERT tokenizer.
iii. Cross Validation: Splitting the dataset into training set and validation set
iv. Making Datasets and DataLoaders using PyTorch
v. Fine-tuning: Training the BERT-based model for the classification task.
vi. Evaluation: Assessing model performance using accuracy score and confusion matrix.
vii. Inference: Using the model for humour classification on new text data.

The accuracy score obtained via the BERT model is 98.7516%.

Following is the Confusion Matrix obtained for the BERT model:



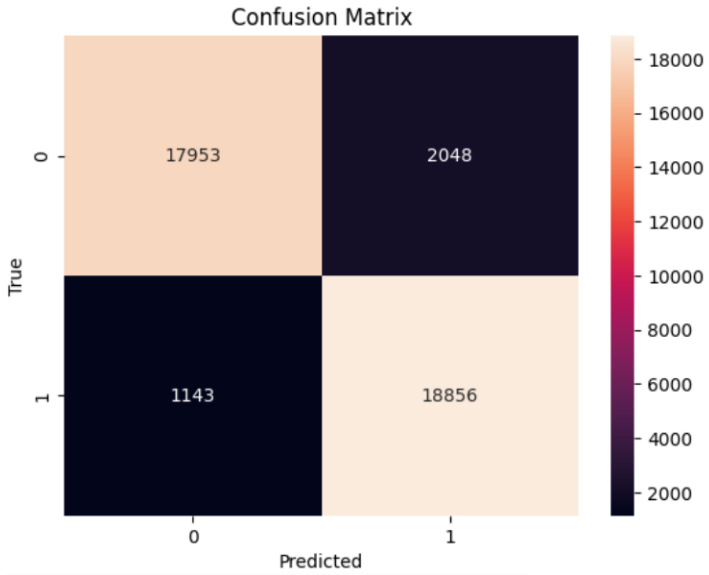
We observe that this model performs exceptionally well. However, when we analyse the confusion matrix, we notice that most of the incorrect predictions occur when the model mistakenly categorizes a non-joke as a joke. Considering that the number of such incorrect predictions is relatively low, it doesn't appear necessary to make any modifications to the model.

2. Naiive Baye's Classifier

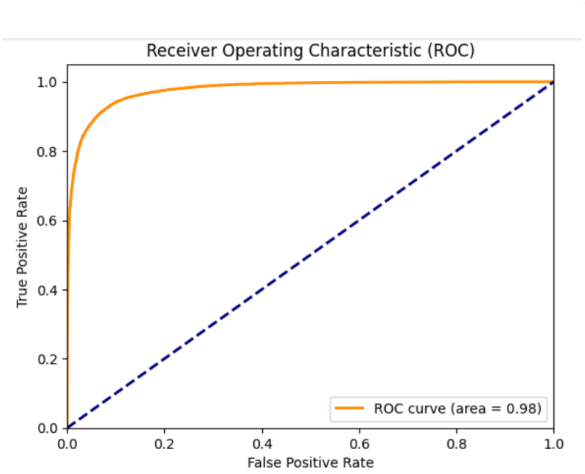
Accuracy: 0.920225

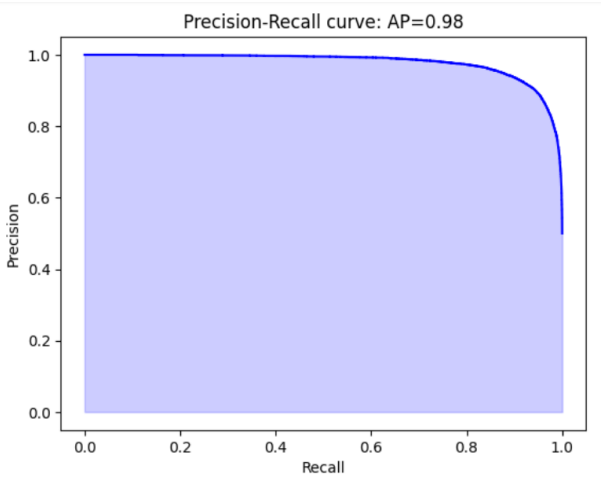
	precision	recall	f1-score	support
False	0.94	0.90	0.92	20001
True	0.90	0.94	0.92	19999
accuracy			0.92	40000
macro avg	0.92	0.92	0.92	40000
weighted avg	0.92	0.92	0.92	40000

From the above evaluation metrics table, we observe that the performance of the Naiive Bayes classifier is reasonably well (with 92% accuracy), although, not surpassing the BERT model.



Upon analyzing the confusion matrix, we notice that most of the incorrect predictions occur when the model mistakenly categorizes a non-joke as a joke. The number of such incorrect predictions is relatively more than the BERT model.



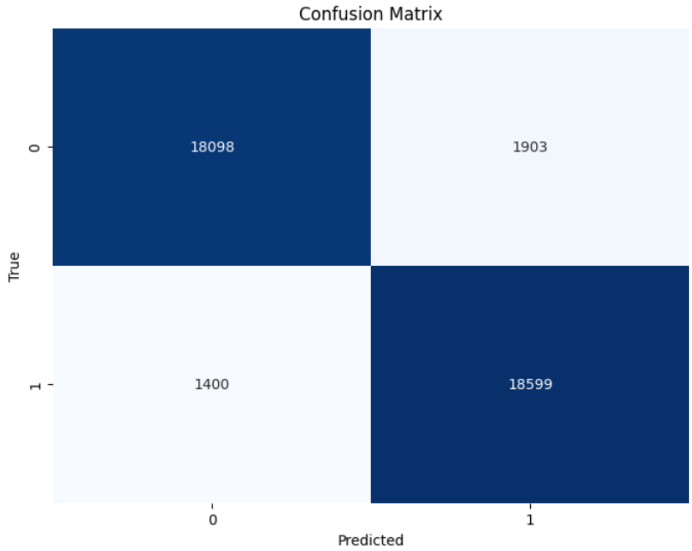


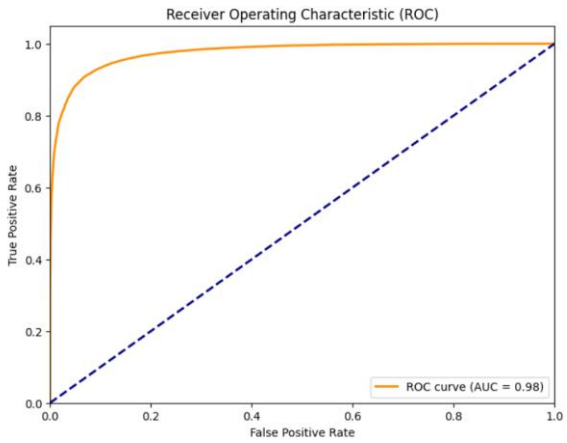
The ROC (Receiver Operating Characteristic) curve with an area under the curve (AUC) of 0.98 is an indication of a highly accurate binary classification model. The ROC curve is a graphical representation of the trade-off between a model's true positive rate (sensitivity) and its false positive rate (1-specificity) across different classification thresholds.

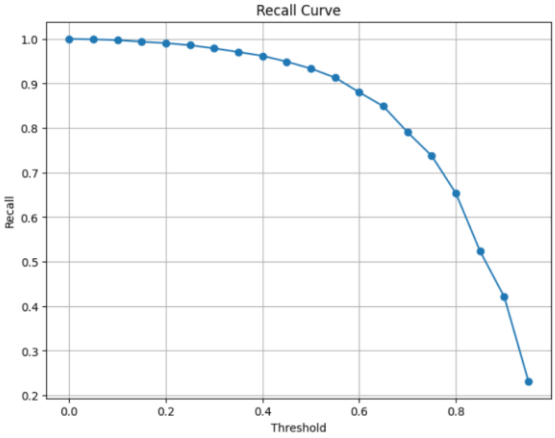
3. Random Forest Classifier

Accuracy: 0.917425

	precision	recall	f1-score	support
False	0.94	0.90	0.92	20001
True	0.90	0.94	0.92	19999
accuracy			0.92	40000
macro avg	0.92	0.92	0.92	40000
weighted avg	0.92	0.92	0.92	40000







On observing the Random Forest evaluation metrics, we can see and infer that although its accuracy is significant (91.7%), it is still lower than the Naiive Bayes Model. Therefore, we can conclude that the BERT pre-trained model is more effective than the Naiive Baye's and Random Forest classifier models.

RESULTS

1. The frequency of non-humorous texts having a lower character count is higher as compared to humorous texts, which tend to have a higher character count.
2. The frequency of low-word-counted (8 to 12 words) non-humorous texts is quite higher than the humorous texts (for the given dataset), which tend to have a higher word count comparatively.
3. Question marks were found to be more prevalent amongst texts classified as jokes. Consequently, we can conclude that humorous texts have a greater tendency to include question marks.
4. On average, humorous content contains more punctuation marks compared to non-humorous content. A good joke is both surprising and follows a familiar comedic structure, often consisting of setups that lead to a punchline. Punctuation plays a significant role in maintaining this structure, which is why jokes tend to have more of it.
5. Non-joke texts tend to contain a higher number of Proper Nouns, Adjectives, and Nouns. This suggests that non-joke texts may lean towards formality, resulting in an increased occurrence of these language elements.
6. The BERT pre-trained model is more effective than the Naiive Baye's and Random Forest classifier models. Amongst the latter two, Naiive Baye's Classifier proved to be slightly more effective than the Random Forest model.

CONCLUSION AND RECOMMENDATIONS

The field of natural language processing (NLP) is currently dominated by the Transformer architecture, which is evident by our research via implementing the BERT model and comparing it with the other Supervised ML techniques. This trend is expected to continue for several years. As demonstrated in this study, we've showcased the effectiveness of this approach. With a relatively short training time, we were able to achieve excellent model performance via the BERT pre-trained language model. In the realm of humour detection, we customized the BERT model and fine-tuned it for our specific task, which, in this instance, involves classifying text as either a joke or not.

Following are some recommendations based on our study:

1. **Leverage Transformer Architectures:** The study highlights the effectiveness of Transformer architectures, particularly the BERT model, for humour detection. It's recommended to explore and leverage these architectures for various NLP tasks, as they offer state-of-the-art performance.

2. **Fine-Tuning for Specific Tasks:** Fine-tuning pre-trained language models like BERT for specific tasks within humour detection can significantly boost model performance. Researchers and practitioners should consider fine-tuning models to align with their specific humour-related goals.
3. **Punctuation Analysis:** Given the prevalence of punctuation in humorous texts, researchers can further investigate the role of punctuation in humour detection. Developing models that explicitly capture and analyse punctuation can improve humour classification.
4. **Semantic Analysis:** Analysing the use of Proper Nouns, Adjectives, and Nouns can provide insights into the differences between humorous and non-humorous texts. Future studies may explore how semantic analysis can enhance humour detection accuracy.
5. **Model Selection:** While BERT outperformed Naive Bayes and Random Forest models in this study, it's essential to choose the right model based on the specific task and dataset. Researchers should experiment with various models to find the most suitable one for their context.
6. **Real-World Applications:** Humour detection has various real-world applications, such as sentiment analysis, content recommendation, and chatbots. Integrating humour detection into these applications can enhance user experience and engagement.
7. **Multilingual Humour Detection:** Extending humour detection models to multiple languages can have broader applications. Multilingual models should be explored to cater to diverse linguistic contexts.

REFERENCES

[1] J. M. Taylor and L. J. Mazlack, "Computationally recognizing wordplay in jokes," *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 26, no. 26, Jan. 2004, [Online]. Available: https://escholarship.org/content/qt0v54b9jk/qt0v54b9jk.pdf?t=op2j2a

[2] Cattle and X. Ma, "Recognizing Humour using Word Associations and Humour Anchor Extraction," *International Conference on Computational Linguistics*, pp. 1849–1858, Aug. 2018, [Online]. Available: https://www.aclweb.org/anthology/C18-1157.pdf

[3] F. Barbieri and H. Saggion, "Automatic detection of irony and humour in Twitter.," *ICCC*, pp. 155–162, Jan. 2014, [Online]. Available: http://computationalcreativity.net/iccc2014/wp-content/uploads/2014/06/9.2_Barbieri.pdf

[4] F. Barbieri, F. Ronzano, and H. Saggion, "Do we criticise (and laugh) in the same way? automatic detection of multi-lingual satirical news in twitter," *International Conference on Artificial Intelligence*, pp. 1215–1221, Jul. 2015, [Online]. Available: https://ijcai.org/Proceedings/15/Papers/175.pdf

[5] O. Weller and K. Seppi, "Humor Detection: A Transformer Gets the Last Laugh," arXiv (Cornell University), Aug. 2019.

[6] Morales and C. Zhai, "Identifying Humor in Reviews using Background Text Sources," *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 492–501, Jan. 2017, doi: 10.18653/v1/d17-1051.

[7] Annamoradnejad, "ColBERT: Using BERT sentence embedding for humor detection," *arXiv (Cornell University)*, Apr. 2020, [Online]. Available: https://arxiv.org/pdf/2004.12765.pdf

[8] R. Miraj and M. Aono, "Integrating Extracted Information from Bert and Multiple Embedding Methods with the Deep Neural Network for Humour Detection," *Social Science Research Network*, Apr. 2021, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3836515