A dissertation submitted to the **University of Greenwich**
in partial fulfilment of the requirements for the Degree of

Master of Science
in
Data Science

Leveraging NLP for Enhanced Company Performance: Transformer-Based Sentiment Analysis and Comparative Review

Name: Kelaniyage Perera
Student ID: 001273076

Supervisor: Dr. Konstantin Kapinchev
Submission Date: December 2023
Word count: 12107

LEVERAGING NLP FOR ENHANCED COMPANY PERFORMANCE: TRANSFORMER-BASED SENTIMENT ANALYSIS AND COMPARATIVE REVIEW

Computing & Mathematical Sciences, University of Greenwich, 30 Park Row, Greenwich, UK.

(Submitted 22 December 2023)

ABSTRACT. The project aims to harness sophisticated machine learning and natural language processing approaches to enhance company performance through sentiment analysis. Leveraging the Glassdoor company reviews dataset and employing five transformer-based models (DistilBERT, BERT, RoBERTa, DeBERTa, and XLNet), alongside topic modelling and aspect-based analysis, this study seeks to extract insights from employee sentiments and perceptions across 500 UK-based companies.

Key objectives involve the training and evaluation of a transformer-based sentiment analysis model, employing topic modelling techniques to identify prevalent themes in reviews, and implementing aspect-based sentiment analysis techniques to delve into personal and organisational aspects. Notably, XLNet demonstrated superior performance with a 76% accuracy rate among the transformer-based models used.

This study addresses the gap in employing advanced NLP models for comprehensive analysis of reviews, aiming to provide actionable insights and recommendations for companies. The project's potential extends to benefiting multiple stakeholders: companies refining products and strategies, customers receiving improved services, investors making informed decisions, and enhancing online review platforms like Glassdoor.

Aligned with the MSc Data Science program, this project demonstrates practical application of theoretical knowledge in data science, AI, and machine learning. Moreover, it adheres to Masters' Project guidelines by showcasing advanced data science methodologies, contributing to the field's knowledge, and demonstrating the application of transformative data science techniques to real-world scenarios.

Keywords: Sentiment analysis, Company reviews, NLP, Transformer-based models, Topic modelling, Aspect-based sentiment analysis, Product, and service quality improvement.

Preface

This project is an investigation into the field of natural language processing, which is a core requirement of the Master of Science (MSc) program and complies with the National Qualifications Framework requirements. The primary focus and main objective of this study revolve around sentiment analysis, exploring topic modelling and aspect-based sentiment analysis. This approach utilizes sophisticated NLP techniques aimed at identifying sentiment patterns associated with specific entities in textual data. The fundamental premise of this research lies in the imperative skill of understanding and extracting subtle emotions from unstructured text, which holds immense significance in today's data-driven world. The essence of this report is to encapsulate the journey undertaken during this investigation, providing comprehensive insights into the intricate nuances of sentiment analysis within the context of entity recognition. Moreover, it ensures adherence to the criteria outlined by the National Qualifications Framework while fulfilling the academic requisites of the MSc program.

Acknowledgements

I would especially express my gratitude to Dr. Konstantin Kapinchev for consenting to supervise me and for his continuous support, advice, and direction during the course of this MSc data science project.

Table of Contents

List of Tables

List of Figures

Glossary

ABSA - Aspect-Based Sentiment Analysis

AWS - Amazon Web Services

BERT - Bidirectional Encoder Representations from Transformers

BoW - Bag of Words

CNN – Convolution Neural Networks

DeBERTa - Decoding-enhanced BERT

GPU – Graphical Processing Unit

KNN - K-Nearest Neighbours

LDA - Latent Dirichlet Allocation

ML - Machine Learning

MTA - Metropolitan Transportation Authority

NLP - Natural Language Processing

NMF - Non-Negative Matrix Factorization

PoS - Part of Speech

RNN – Recurrent Neural Networks

RoBERTa - Robustly Optimised BERT

VADER - Valence Aware Dictionary and Sentiment Reasoner

CHAPTER 1 – INTRODUCTION

1.1 Study overview

Sentiment analysis has gained prominence within the realm of computational linguistics due to the vast amount of sentiment-related data originating from various social media platforms, including Twitter, Facebook, online forums, and blogs, as highlighted in paper (Wangs, 2013). This analysis holds substantial significance for customers, allowing them to access feedback regarding products or services before a purchase. Similarly, companies leverage sentiment analysis to gauge employer opinions about their offerings, enabling them to assess employer satisfaction and make necessary improvements to enhance their working environment (Sahayak *et al*,2015).

In this study, the Glassdoor dataset was utilised to analyse company reviews from 500 companies. Five transformer-based approaches, namely DistilBERT, BERT, RoBERTa, DeBERTa, and XLNet, were employed, along with topic modelling and aspect-based analysis. The novelty of this study is the utilisation of transformer-based approaches within the domain of company reviews. The primary goal was to analyse Glassdoor reviews of UK companies across industries to extract insights regarding employee sentiments and company perceptions.

The beneficial stakeholders would extend to other companies. If commercially deployed, this approach could be managed by a central entity. This entity would analyse data from various companies and provide the output via a web application or interactive dashboard, facilitating the dissemination of analysed insights to the respective companies.

The significant contribution of this study to the body of literature by employing transformer-based methodologies. Specifically, employing aspect-based sentiment analysis coupled with NER and LDA facilitates the identification of negative and positive tags associated with specific named entities, such as individuals or organisations. This enables a more nuanced understanding of sentiment related to these entities within the reviews.

Additionally, the integration of topic modelling techniques enables the identification and extraction of the most prevalent phrases or topics discussed within the entirety of the review dataset. This aids in discerning the key themes or subjects that garner the most attention or discussion among reviewers.

The project on leveraging NLP techniques for enhanced company performance holds promise for a multitude of stakeholders. Companies and businesses stand to gain substantial insights into customer sentiments, preferences, and opinions, allowing them to refine products, fine-tune marketing strategies, and ultimately bolster overall customer satisfaction. Concurrently, customers themselves are poised to benefit from improved products and services, tailored more accurately to their needs through companies' proactive responses to their feedback. Moreover, this endeavour contributes significantly to the field of NLP and sentiment analysis, serving as a practical application of transformer-based models, thereby advancing the realm of research and development. Such comprehensive insights derived from this analysis also hold significance for investors and decision-makers, empowering them with informed data for strategic decision-making, partnerships, and investments. Additionally, this initiative could potentially elevate the quality and utility of online review platforms, benefitting platforms like Glassdoor, by enhancing sentiment analysis tools. Ultimately, the project's far-reaching application aims to optimise business operations, elevate customer satisfaction, and augment the efficiency of decision-making processes across various sectors.

1.2 Research Background and Motivation

Background:

The contemporary landscape of job searching, and company evaluation has been significantly transformed by online platforms like Glassdoor. These platforms have become instrumental by providing invaluable insights derived from employee reviews, thereby profoundly influencing the decision-making process for both job seekers and most importantly employers.

Motivation:

Despite the wealth of data available on such platforms, there exists a noticeable absence in employing the most recent advancements in NLP models, particularly the transformer-based models, for comprehensive analysis. This gap highlights the untapped potential benefits that could be derived from leveraging these advanced analytical techniques on these reviews. Implementing such models could offer richer insights for companies, and the platform itself, consequently aiding in reducing employee turnover.

1.3 Research Question

How can the analysis of Glassdoor reviews be utilised to extract insights beneficial for companies?

1.4 Research Objectives and Aims

1.4.1 Objective: To train a Transformer-Based Sentiment Analysis Model

- Research and explore existing Transformer-based models for sentiment analysis in NLP.
- Preprocess and tokenize the provided dataset of company reviews.
- Fine-tuning transformer-based model using the training and validation dataset.
- Analyse performance metrics such as F1 score, accuracy, precision, and recall.
- Deliverables: Trained sentiment analysis model, evaluation report documenting model performance.

1.4.2 Objective: To conduct Topic Modelling on Company Reviews

- Apply topic modelling techniques (e.g., LDA or NMF) to extract prevalent topics from the company reviews.
- Analyse and interpret the identified topics to gain insights into the common themes discussed across different companies.
- Visualise the topics using appropriate visualisation techniques (e.g., word clouds or heatmaps).
- Deliverables: Topic modelling results, visualisation of topics, insights, and interpretation of the identified themes.

1.4.3 Objective: To conduct Aspect-Based Sentiment Analysis

- Implement methods such as NER or POS tagging to identify relevant aspects (e.g., person, organisation, department, time) in the reviews.
- Determine the sentiment of each aspect using the pre-trained sentiment analysis model.

- Analyse the sentiment of specific aspects to gain insights into areas of excellence and areas that need improvement for each company.
- Deliverables: Aspect identification results, sentiment analysis of aspects, insights on company-specific areas of improvement.

1.4.4 Objective: To Develop a React-based Prototype for a Web Application with Visualizations

- Implement a prototype using React to create a feature-rich web application for visualizing sentiment analysis, topic modelling, and company comparison results.
- Utilize static visualizations such as charts, graphs, and word clouds to present findings in a comprehensive manner.
- Ensure the web application provides detailed explanations alongside visualizations for a comprehensive understanding of the insights derived.
- Enable a user-friendly interface allowing seamless navigation and exploration of the project's results.
- Deliverables: Prototype of a web application showcasing comprehensive project findings, utilizing React-based interactive visualizations and detailed explanations for enhanced user engagement and understanding.

1.5 Structure of the Report

The report consists of several chapters where each dedicated to specific aspects of the objectives:

Introduction: Provides an overview of the project's context, objectives, and fundamental research questions.
Literature Review: Analyses and synthesises existing literature on sentiment analysis, Transformer-based NLP models, methodologies in topic modelling, and aspect-based sentiment analysis, building upon the theoretical framework established in the previous section.
Theoretical Framework
Transformer-Based NLP Models: Discusses theoretical concepts and functionalities of various transformer-based models, including BERT, DistilBERT, RoBERTa, DEBERTa, and XLNet, highlighting their architectures, pre-training strategies, and applications in sentiment analysis.

Topic Modelling Techniques: Explores the theory of the techniques like Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), outlining their algorithms, assumptions, and applications in extracting prevalent themes from textual data.

Aspect-Based Sentiment Analysis: Explores the theoretical foundations of Aspect-Based Sentiment Analysis (ABSA) methodologies, particularly focusing on techniques like Named Entity Recognition (NER) and their role in identifying specific aspects within text data, and how sentiment analysis is conducted on these aspects.

Methodology: Elaborates on the methodologies, techniques, and tools employed in the study, detailing the practical application of Transformer-based models (BERT, DistilBERT, RoBERTa, DEBERTa, XLNet) for sentiment analysis, topic modelling techniques (LDA, NMF), and aspect-based sentiment analysis (NER) on the Glassdoor UK dataset.

Results and Discussion: Presents the outcomes derived from sentiment analysis, topic modelling, company comparisons, aspect-based sentiment analysis, and offers detailed interpretations and insights based on the practical application of the methodologies discussed.

Engages in an extensive discussion of the findings, their implications, limitations encountered during the study, and suggests potential future research directions, tying back the practical outcomes to the theoretical framework established.

Future work: Potential directions for advancing sentiment analysis methodologies were explored. Discussions revolved around fine-tuning domain-specific models for corporate sentiment analysis, integrating explainable AI techniques, expanding sentiment analysis to multilingual contexts, addressing ethical biases, enhancing user-centric interfaces, and optimizing models for deployment and scalability.
These proposed avenues aim to propel the field of sentiment analysis towards greater accuracy, inclusivity, user engagement, and ethical considerations in future research endeavours.

Conclusion: Summarises the key findings, contributions made by the study, reiterates their significance within the broader theoretical context, and proposes potential avenues for further exploration within the conceptual framework.

CHAPTER 2 - LITERATURE REVIEW

2.1 Work Satisfaction Impact on Company Success

Numerous scholars emphasise the shift towards a human-centred view of firms in modern economies, highlighting the significance of employee satisfaction (Filbeck and Preece, 2003; Edmans, 2011; Symitsi et al., 2018). Studies by Schneider et al. (2003), Edmans (2011), Guiso et al. (2014), Rubera and Kirca (2012), O'Reilly et al. (2014), and Popadak (2013) suggest that higher employee satisfaction levels are associated with better financial outcomes, innovation, revenue growth, and long-term firm value.

However, the conventional methods of measuring employee satisfaction via surveys and publicly available datasets like GPTWI and "Best Companies" rank are criticised for their limitations (Luo, Zhou, & Shon, 2016). These approaches are prone to selection bias and lack granularity in understanding the intricacies of employee satisfaction.

To address these limitations, leveraging social media data, such as mining employee reviews on platforms like Glassdoor, is proposed as a potential direction. This method could provide a more comprehensive and less biassed view of employee satisfaction within organisations (Luo, Zhou, & Shon, 2016).

All these studies emphasize how crucial employee satisfaction is to boosting business success and show how alternative data sources, like online employee evaluations, can provide rich details that standard surveys may fail to capture.

2.2 Online Platforms and Glassdoor Insights

The influence of social media spans diverse domains, impacting opinions, consumer sentiments, and even stock price predictions (Asur and Huberman, 2010; Bollen et al., 2011). Platforms like Twitter have become instrumental in gauging public opinions, predicting election outcomes, and delineating product advantages and drawbacks (Bermingham and Smeaton, 2011; Kim and Hovy, 2006).

Glassdoor, established in 2007, serves as a unique platform where employees can anonymously review companies and management (Glassdoor: About Us, 2016). Despite its potential, limited studies have harnessed Glassdoor's wealth of data. For instance, Huang et al. (2015) analysed Glassdoor employee ratings, establishing a link between employee satisfaction and companies' market value using specific rating dimensions (Huang et al., 2015). However, these studies often overlooked the rich textual content within employee reviews, limiting their scope and depth (Moniz, 2015).

In a pioneering move, Moniz (2015) conducted textual analysis on Glassdoor reviews, employing topic modelling to extract keywords related to "goal-setting." This study revealed a positive association between goals and firm value (Tobin’s Q), highlighting the untapped potential of text reviews in comprehending various dimensions of corporate value through employee satisfaction (Moniz, 2015).

2.3 Approaches for Sentiment Analysis

The study by Mekala et al. (2021) emphasises the significance of analysing user feedback from app repositories and social media platforms, particularly in extracting insights on product and service opinions. Traditional NLP and ML models often face challenges with accuracy due to the need for extensive labelled datasets. Addressing this limitation, the authors introduced a deep-learning-based approach utilising a BERT-based sequence classifier, achieving an impressive 87% classification accuracy for user feedback analysis. Their research showcases the efficacy of deep learning in distinguishing relevant from irrelevant content in user feedback, a task where traditional ML methods had previously been employed. Additionally, the study identifies potential improvements to enhance dataset quality and extend the classification accuracy to 95-100% for various classification tasks, aiming to surpass conventional methods and delve into more granular assessments of user feedback's relevance to software requirements. This work offers insights into leveraging deep learning models for efficient and accurate analysis of online user feedback, offering promising directions for future research in requirements engineering (Mekala et al., 2021).

Sahayak, Shete, and Pathan (2015) delved into sentiment analysis on Twitter, employing machine learning algorithms, including Weka-based models like unigram, tree kernel, and feature-based approaches. Their study aimed to categorise tweets into positive, negative, or neutral sentiments, proving valuable for companies seeking product feedback and customers evaluating opinions pre-purchase. However, challenges persist due to Twitter's brief, context-rich messages. The research highlighted the complexities of deciphering sentiments within limited text lengths and underscored the domain-centric nature of sentiment analysis, limiting its cross-platform applicability.

Ghazzawi and Alharbi (2019) explored customer complaints in the context of the MTA public service provider, employing Naïve Bayes, KNN, ID3, and Random Trees classification models. The study aimed to link complaint details to specific agencies. Results revealed that the ID3 classifier demonstrated the highest predictive capability, showcasing superior performance in accuracy, recall, and precision. Conversely, KNN exhibited the least favourable outcomes, indicating its unsuitability for the complaint dataset. Overall, the research successfully classified complaint records based on agency affiliations, highlighting a strong correlation between issue details and agency names within the dataset.

2.4 Structural Topic Modelling Approach

Ding et al. (2020) employed a unique approach, utilising structural topic modelling, to discern service quality attributes within Airbnb accommodations based on 242,020 reviews in Malaysia. Their innovative methodology extracted 22 service-related topics, revealing four previously unexplored facets. The study cross-validated a modified SERVQUAL questionnaire with topic modelling outcomes, affirming its ability to encompass general Airbnb service quality attributes. Additionally, the research delved into the preferences of Malaysian versus international Airbnb users. It found that Malaysian users emphasised property appearance and location, while international users prioritised accommodations for larger groups. Furthermore, the study highlighted the growing significance of communication with hosts in shaping Airbnb users' lodging experiences. The exploration of distinct service quality dimensions using this novel technique presents valuable insights into the nuances of Airbnb accommodation experiences.

2.5 Deep Learning Techniques for Named Entity Recognition

The research done by Li et al. (2020) focuses on application of the deep learning techniques NER. It explores several models and methodologies employed in NER tasks, including Multi-Task Learning, Transfer Learning, Active Learning, Reinforcement Learning, Adversarial

Learning, and Neural Attention mechanisms. Notably, their investigation highlights how Transfer Learning facilitates effective model adaptation from one domain to another with limited labelled data, while Active Learning significantly reduces annotation efforts while maintaining model accuracy. The research showcases the utility of these deep learning models in enhancing NER methodologies, demonstrating their versatility across diverse domains and languages.

The literature review explores various research domains, including the impact of employee satisfaction on company success, sentiment analysis in social media, service quality attributes through structural topic modelling, and deep learning methods in NER. Highlighting the significance of alternative data sources such as social media and Glassdoor reviews, this review serves as a foundation for future research. This study aims to further investigate advanced techniques like transformer-based approaches, seeking to advance the understanding and analysis of these domains.

CHAPTER 3 – THEORETICAL FRAMEWORK

3.1 Overview of Natural Language Processing

Natural Language Processing (NLP) exists at the convergence point of computer science and artificial intelligence, focusing on the comprehension and manipulation of human language. As delineated by Lane et al. (2019), natural languages diverge markedly from computer programming languages, serving as mediums for human communication rather than being directly translatable into finite mathematical operations like their programming counterparts. While programming languages instruct machines with precision, natural languages are employed by humans for everyday communication, sharing information, expressing sentiments, and conveying complex ideas without adhering to a rigid syntax governed by compilers or interpreters.

NLP is a research area within Computer Science and Artificial Intelligence as defined by Lane et al. (2019) for processing natural languages. This process mainly encompasses converting natural languages into computer understandable data, allowing machines to comprehend and learn about the world through the representation of linguistic information in numerical formats. The understanding garnered from this data transformation is utilized to generate natural language text reflective of the underlying comprehension.

Moreover, a NLP system is often visualized as a pipeline, comprising multiple stages of processing. Lane et al. (2019) likens this system to a pipeline due to its sequential stages where raw natural language inputs enter one end, undergo various computational processes, and yield processed outputs at the other end. This metaphorical pipeline encapsulates the flow of language comprehension, from initial input to final processed output.

In essence, NLP systems decode, understand, and generate human languages, empowering machines to comprehend, interpret, and manipulate linguistic information—a critical domain involving computer science, artificial intelligence, and linguistic analysis.

3.2 Fundamentals of Transformer-based Models

In the domain of natural language processing (NLP), Transformer-based models have emerged as a groundbreaking innovation for processing sequential data, specifically excelling in tasks such as translation, text summarization, and language generation (Vaswani et al., 2017). Prior to the introduction of Transformers, RNNs, including LSTM networks, were extensively utilized to capture word order in language-related tasks. However, the limitations inherent in RNNs became apparent when faced with processing large text sequences, as their sequential nature led to computational inefficiency and an inability to retain long-range dependencies (Hochreiter & Schmidhuber, 1997).

The paradigm shift arrived with Transformers, as they presented a novel architecture capable of parallel computation, enabling the training of significantly larger models using GPUs (Devlin et al., 2018). This departure from the constraints of RNNs, particularly in efficiently processing extensive text sequences without losing context, marked a significant milestone in natural language understanding and analysis. The introduction of Transformers, spearheaded by Vaswani et al. (2017), revolutionized the field by leveraging attention mechanisms for enhanced contextual comprehension, making them a linchpin in various applications, notably within sentiment analysis (Chen et al., 2019).

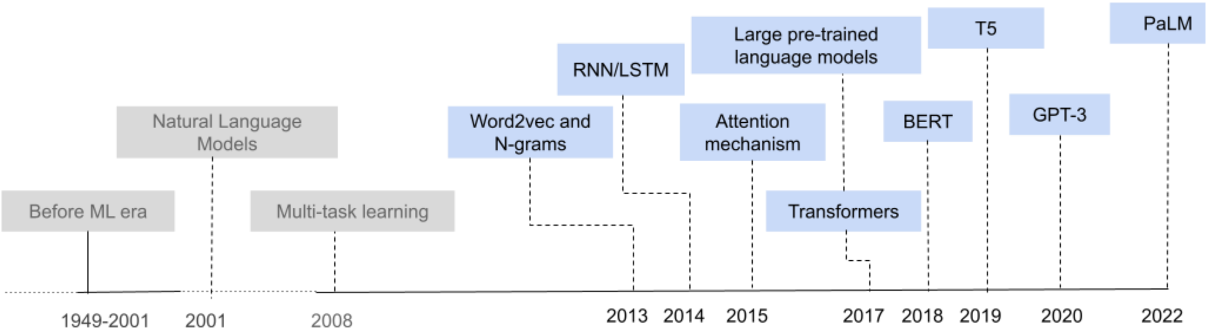


Figure 1: Language Modelling History

3.2.1 Conceptual Basis of Transformers

The conceptual basis of Transformers lies in several key components that revolutionized sequence transduction models. These components include positional encoding, attention mechanism, and self-attention.

Positional Encoding

Positional encoding addresses the absence of inherent sequence order information in the Transformer model due to its lack of recurrent or convolutional structures. Vaswani et al. introduce positional encodings to imbue the model with information about the relative or absolute positions of tokens within a sequence. They utilize sine and cosine functions to generate positional embeddings, allowing the model to consider sequence order. This enables the Transformer to discern positional relationships, despite lacking inherent sequential computations. (Vaswani et al., 2017)

Attention

Attention serves as a pivotal mechanism in the Transformer architecture, enabling the model to focus on relevant parts of the input sequence while performing computations. This attention mechanism allows the model to weigh different input elements based on their relevance to a specific output, offering a way to model dependencies without being constrained by their distances in the sequence. (Vaswani et al., 2017)

Self-Attention

Self-attention, or intra-attention, stands as a central feature in the Transformer, facilitating the computation of sequence representations. By relating different positions within the same sequence, self-attention aids in generating comprehensive sequence representations. It has been effectively applied in a multitude of tasks, encompassing reading comprehension, summarising information, and acquiring task-agnostic sentence representations (Vaswani et al., 2017).

The fundamental component of the transformer architecture is illustrated in Figure 2. Within the encoder and decoder, there exists a series of stacked layers. Each of these layers comprises a multi-head self-attention mechanism, succeeded by a position-wise feed-forward network. These individual layers are encompassed by residual connections and layer normalization. Through the self-attention mechanism, the model gains the capability to concurrently attend to diverse segments within the input sequence. This stands in contrast to the approach of traditional recurrent neural networks, which handle the input sequence token by token. By employing the self-attention mechanism, the model can effectively apprehend extensive dependencies present in the input sequence. This becomes indispensable for numerous natural language processing tasks.

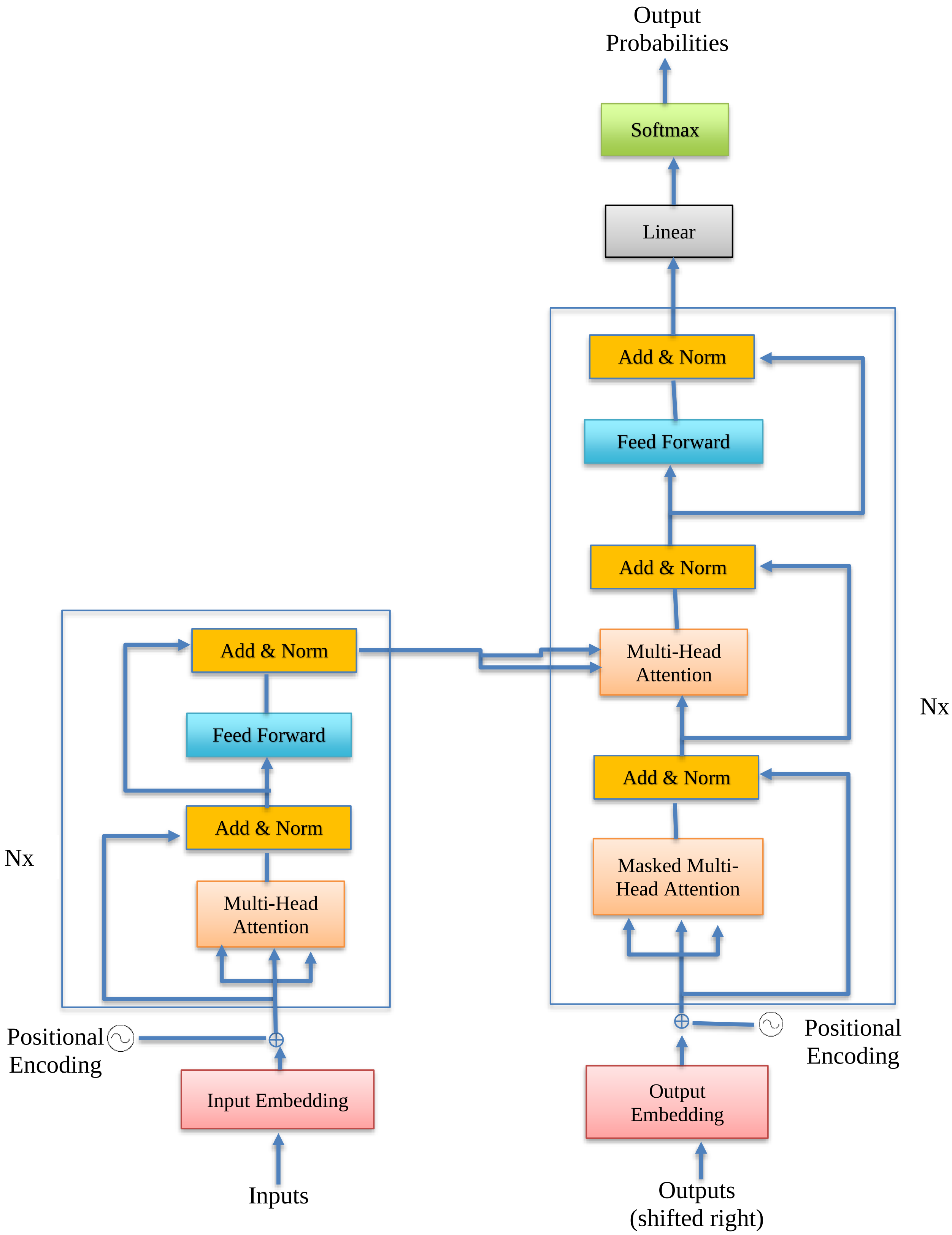


Figure 2:Transformer Model Architecture (Vaswani et al., 2017)

The position -wise feed forward neural network constitutes a two-layered structure pivotal for learning non-linear relationships among input tokens. This feature holds significant importance, particularly in tasks like machine translation, where the model's capability to generate new words absent in the original input sequence is essential. Introducing residual connections and layer normalization significantly contributes to enhancing the stability of model training. Residual connections play a critical role in mitigating gradient vanishing or explosion issues during the training process, while layer normalization ensures the stabilization of neuron activations within the network.

Overall, the transformer architecture stands as a versatile and flexible neural network design extensively proven to be effective across a diverse spectrum of natural language processing tasks. Its pivotal self-attention mechanism serves as the cornerstone innovation, enabling the model to capture extensive dependencies within input sequences. Moreover, the position-wise feed-forward network substantially enables the model to discern and understand intricate non-linear relationships among the input tokens. Complemented by residual connections and layer normalization, the model's training stability is significantly bolstered.

3.2.2 Overview of Transformer Variants

This section delves into the theoretical underpinnings of prominent transformer-based models employed in sentiment analysis, including DistilBERT (Wolf et al., 2019), XLNet (Yang et al., 2019), BERT (Devlin et al., 2019), DeBERTa (He et al., 2020), and RoBERTa (Liu et al., 2019). Their distinct architectures, pre-training strategies, and advantages concerning sentiment analysis tasks were considered.

Bidirectional Encoder Representations from Transformers (BERT):

The cornerstone of this analysis is BERT, a bidirectional encoder architecture that leverages transformers to understand word context within a sentence. BERT excels in capturing long-range dependencies and generates comprehensive word representations through its multi-layered structure (Devlin et al., 2019). Its masked language modelling and next sentence prediction pre-training tasks empower it to analyse both intra-sentence and inter-sentence relationships (Liu et al., 2019). Notably, BERT demonstrates exceptional performance in sentiment analysis tasks due to its ability to account for sentiment nuances and contextual shifts within textual data (Joshi et al., 2020).

DistilBERT: A Lighter BERT:

Building upon BERT's success, DistilBERT offers a lighter, faster alternative, making it an appealing choice for resource-constrained environments. By reducing the number of parameters and attention heads, DistilBERT maintains comparable accuracy to BERT while achieving significantly faster inference speeds (Wolf et al., 2019). This reduced complexity makes DistilBERT suitable for real-time sentiment analysis applications, particularly in social media analytics or customer service chatbots (Khan et al., 2022).

RoBERTa: A Robustly Optimized BERT Pretraining Approach:

RoBERTa seeks to address potential biases inherent in the original BERT training data By eliminating the predictive objective focused on consecutive sentence prediction and instead training on extended document segments, RoBERTa significantly enhances its performance on subsequent tasks, such as sentiment analysis, surpassing its predecessor (Liu et al., 2019). This enhanced robustness makes RoBERTa a valuable option for situations where data quality and bias minimization are critical considerations (Sun et al., 2021).

DeBERTa: Decoupling Encoder-Decoder Representation from Bi-directional Encoder Representations:

DeBERTa introduces a disentangled encoder architecture that addresses the limitations of BERT's positional encoding strategy. By decoupling the absolute and relative positional information, DeBERTa facilitates improved long-range dependency modelling and achieves outstanding results in sentiment analysis tasks, particularly for complex or sentiment-ambiguous texts (He et al., 2020). This advancement allows DeBERTa to tackle challenging sentiment analysis scenarios where subtle nuances or sarcasm may be present.

XLNet: Generalized Autoregressive Pretraining for Language Understanding:

XLNet stands out amongst these models by employing a permutation language modelling objective during pre-training. This approach enables the model to explore every conceivable arrangement of the input sequence, resulting in improved understanding of complex syntactic structures and long-range dependencies. XLNet demonstrates competitive performance in sentiment analysis, particularly for tasks involving dialogue or conversational data (Yang et al., 2019).

While these transformer models share a common foundation in transformer architecture, their distinct pre-training objectives, parameter sizes, and architectural adaptations lead to nuanced differences in performance and suitability for specific sentiment analysis tasks. BERT reigns supreme in its balanced blend of accuracy and interpretability, while DistilBERT offers speed and efficiency for resource-constrained environments. RoBERTa prioritizes robustness to data bias, while DeBERTa excels in handling complex textual nuances. XLNet, with its permutation language modelling, shines in analysing intricate syntactic structures and conversational data.

3.4 Topic Modelling Techniques

Topic modelling is a statistical technique employed to recognize abstract subjects or themes present in a corpus of documents or texts. It aims to uncover hidden thematic structures and prevalent themes without prior labelling. LDA and NMF stand as popular techniques for topic modelling, where LDA assumes a probabilistic approach based on the distribution of words in topics, while NMF focuses on matrix factorization to uncover underlying topics within the data (Blei et al., 2003; Lee & Seung, 2000).

3.4.1 Introduction to Latent Dirichlet Allocation (LDA):

LDA is a probabilistic model that operates on collections of discrete data, particularly text corpora. It models every document within a corpus as a stochastic combination across latent topics. Each of these topics characterizes a probability distribution encompassing words within the corpus. This hierarchical Bayesian model encompasses three levels, describing the generation of words within a document based on underlying topic mixtures (Blei et al., 2003).

3.4.2 Basic generative process of LDA

LDA proposes a generative process for constructing documents allowing documents to be associated with multiple topics, thereby capturing the complex relationships and inherent structures within the corpus:

1. Document-Level Representation: Every document is initially represented as a mix of latent topics randomly, characterized by a Dirichlet distribution.
   $\theta \sim \mathrm{Dir}(\alpha)$

2. Topic Assignment: Every word within a document, a topic is chosen from the document's topic mixture.
   zn~Multinomial(θ)

3. Word Generation: Words are generated based on a multinomial distribution conditioned on the selected topic.
   wn~p(wn|zn,β)

3.4.3 Probabilistic Modelling and Inference in LDA

The joint distribution of the topic mixture, topics, and words is formulated, considering the Dirichlet priors for topics and multinomial distributions for word generation within documents. Estimating the model parameters, such as α (Dirichlet parameter) and β (word probabilities parameter), involves inference and parameter estimation, which can be challenging due to the intractability of exact inference (Blei et al., 2003).

$$\mathrm{p}(\theta, \mathrm{z}, \mathrm{w}|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n, \beta)$$

3.4.4 Graphical Representation and Levels of LDA

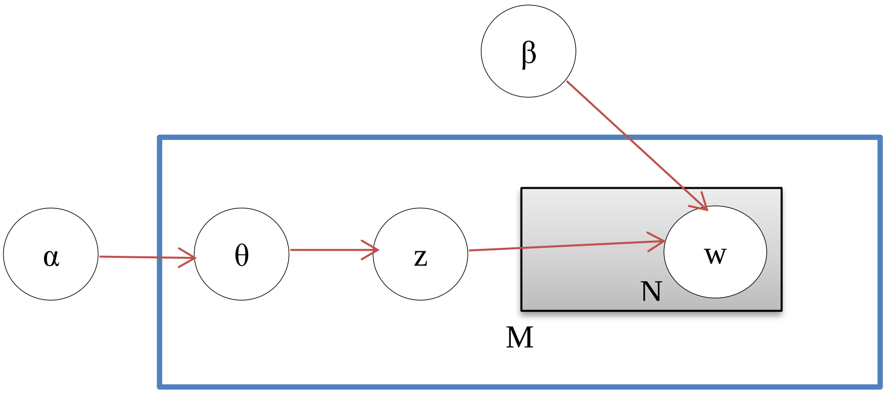


Figure 3: Graphical representation of LDA

A graphical model representation illustrates the three levels of LDA: corpus-level parameters (α and β), document-level variables (topic mixtures), and word-level variables (topics and words). This hierarchical structure distinguishes LDA from traditional clustering models by allowing multiple topic assignments for words within a document (Blei et al., 2003).

LDA, serving as a probabilistic generative model, encapsulates the essence of capturing complex structures within textual data. Its hierarchical nature and flexibility provide a foundation for exploring and extending the model's capabilities to address diverse challenges in text modelling and data representation.

3.4.5 Introduction to Non-negative Matrix Factorization (NMF)

NMF is a potent mathematical technique widely applied in data analysis and pattern recognition. This approach aims to factorize a given non-negative matrix V into two non-negative matrices W and H of lower dimensionality, as represented by the equation (Lee & Seung, 2000):

$$V \approx WH \rightarrow (1)$$

In the context of statistical analysis of multivariate data, NMF provides a way to approximate a given matrix V, which consists of n-dimensional data vectors arranged as columns, into a product of two matrices: W, an n×r matrix, and H, an r×m matrix. The objective is to identify a compressed representation of the original data, where r should be smaller than n or m to produce a more concise matrix.

Equation (1) represents an essential aspect of NMF: the linear approximation of each data vector V by a combination of columns in W weighted by components in H. Consequently, W holds a set of basis vectors optimized for approximating the data in V. A good approximation through W requires the discovery of latent structures inherent in the data.

Cost Functions in NMF:

To assess the accuracy of the approximation in NMF, two primary cost functions are employed:

1. Euclidean Distance:

$$\left||V - WH|\right|^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})\text{^}2 \rightarrow (2)$$

2. Kullback-Leibler Divergence:

$$D(V||WH)) = \sum_{i,j} \left(V_{ij} log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij} \right) \rightarrow (3)$$

These cost functions allow the measurement of the difference between V and the approximated WH. Minimizing these functions with respect to W and H under the non-negativity constraints W, H≥0 is fundamental in NMF.

Utilizing the principles of NMF elucidates an innovative avenue within the analysis of the Glassdoor dataset. By applying NMF techniques to this dataset, the intention is to delve into uncovering latent topics and prevalent themes. This methodology offers a promising approach to distil the intricate feedback and reviews present within the Glassdoor platform, ultimately aiming to extract and understand underlying topics contributing to a comprehensive analysis of user sentiments and opinions.

3.5 Aspect-based Sentiment Analysis Approach

ABSA is a specialized technique within NLP that aims to discern and evaluate sentiment polarity regarding specific aspects or entities within textual data (Liu et al., 2018). This methodology is instrumental in extracting nuanced sentiments associated with different facets or features mentioned in text, providing a more granular understanding of opinions expressed.

At its core, ABSA focuses on Named Entity Recognition (NER) as a pivotal technique for identifying and categorizing specific entities or aspects within a given text corpus (Pontiki et al., 2014). NER operates by discerning and classifying entities, such as products, services, or attributes, that are the subjects of opinions or sentiments expressed in the text. This process typically involves a predefined set of entity categories or patterns recognized through machine learning algorithms or rule-based systems.

Once the relevant entities or aspects are identified, sentiment analysis is conducted specifically on these entities to gauge the sentiment polarity associated with each (Wang et al., 2016). Sentiment analysis within ABSA aims to determine whether opinions expressed towards these specific aspects exhibit positivity, negativity, or neutrality. This analysis can involve various methods, such as lexicon-based approaches, machine learning models, or hybrid systems combining both approaches.

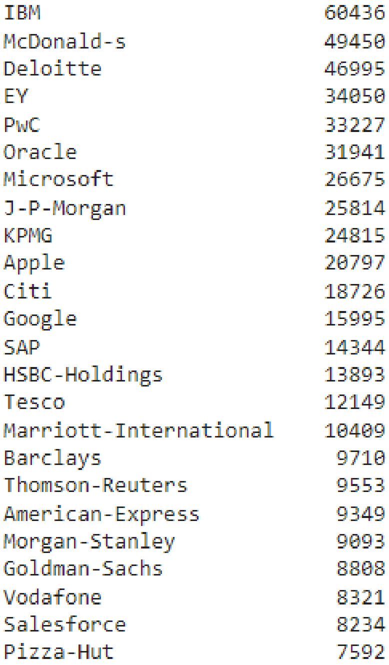
The sentiment analysis process may include the utilization of lexicons or dictionaries specialized in sentiments, containing scores associated with words or phrases (Liu et al., 2018). These scores help assess the polarity of sentiments expressed towards aspects. Additionally, machine learning models like SVM, RNN, or Transformer-based models like BERT, GPT, or their variants may be employed to infer sentiments related to specific aspects within the text.

Overall, ABSA amalgamates the methodologies of NER and sentiment analysis to enable a fine-grained analysis of sentiments, providing insights into the diverse opinions regarding various aspects or entities present in textual data.

CHAPTER 4 – METHODOLOGY

4.1 Data Sources and Dataset Selection

The Glassdoor dataset was chosen which presents a superior choice owing to its extensive features and comprehensive information, distinguishing it from simpler datasets like those found on Kaggle. Its authenticity, offering genuine reviews from employees about their workplace experiences, provides valuable insights into company cultures and overall employee satisfaction, a rarity in standard datasets. Housing 800,000 records across 18 columns, this dataset encapsulates employee reviews, company ratings, salary information, and more, affording a broader context for comprehensive analysis. Covering 428 unique UK companies, it allows for the examination of diverse industries, company sizes, and cultures, enabling nuanced insights that smaller datasets may lack. For job seekers, the dataset aids in informed decision-making by offering real employee experiences to assess crucial factors such as job satisfaction and company culture. Moreover, its depth supports businesses and researchers in analysing industry trends, understanding market sentiments, and making informed strategic decisions. The dataset's size and depth also open doors for intricate analyses, machine learning, and NLP tasks, providing extensive research opportunities beyond the scope of simpler datasets.

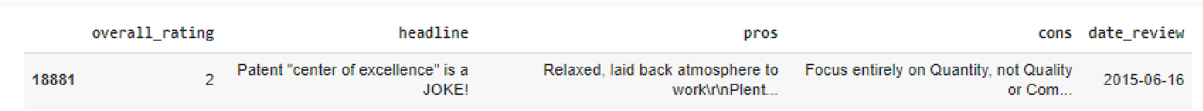


```
IBM                         60436
McDonald-s                  49450
Deloitte                    46995
EY                          34050
PwC                         33227
Oracle                      31941
Microsoft                   26675
J-P-Morgan                  25814
KPMG                        24815
Apple                       20797
Citi                        18726
Google                      15995
SAP                         14344
HSBC-Holdings               13893
Tesco                       12149
Marriott-International      10409
Barclays                     9710
Thomson-Reuters              9553
American-Express             9349
Morgan-Stanley               9093
Goldman-Sachs                8808
Vodafone                     8321
Salesforce                   8234
Pizza-Hut                    7592
```

Figure 4: Sample of the companies and count of reviews in Glassdoor

4.2 NLP Data Preprocessing, Tokenization, Stemming, and Feature Extraction

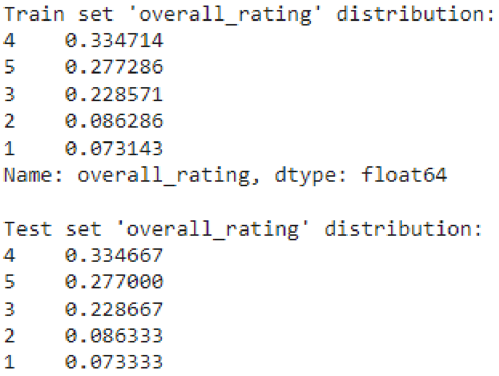
To this study, a fraction of the data was selected to ensure feasibility in training transformer models. Overall rating, date of the review (date_review) features with the concatenation of 'cons', 'pros', and 'headline' columns was performed to gather comprehensive review information for analysis.



	overall_rating	headline	pros	cons	date_review
18881	2	Patent "center of excellence" is a JOKE!	Relaxed, laid back atmosphere to work\r\nPlent...	Focus entirely on Quantity, not Quality or Com...	2015-06-16

Figure 5: Overview of the extracted features and target

Duplicates and null values were addressed to ensure data quality. Removing duplicates to ensure data integrity and eliminate redundancy, while removing null values helps maintain data consistency and prevents potential biases in the analysis. a sample size of 10,000 instances was selected from the dataset, adhering to a 70:30 train-test split ratio to facilitate model training and evaluation. The sample was derived while preserving the original distribution of the 'overall_rating' feature to ensure representativeness across both the train and test sets. By employing a stratified splitting approach, the 'overall_rating' distribution remained consistent in both partitions. This methodology was implemented to maintain the proportionality of different 'overall_rating' categories in the training and testing subsets. As depicted in the analysis, the 'overall_rating' distributions in the train and test sets reflect this preservation of proportions, ensuring that the model's learning and generalization are based on a balanced representation of the original dataset's target variable.



```
Train set 'overall_rating' distribution:
4    0.334714
5    0.277286
3    0.228571
2    0.086286
1    0.073143
Name: overall_rating, dtype: float64

Test set 'overall_rating' distribution:
4    0.334667
5    0.277000
3    0.228667
2    0.086333
1    0.073333
```

Figure 6: Record distribution per target categories

The 'overall_rating' column was remapped into a binary classification to serve as the target variable for sentiment analysis. Figure 2 depicts the original data distribution and how it looks after mapping.

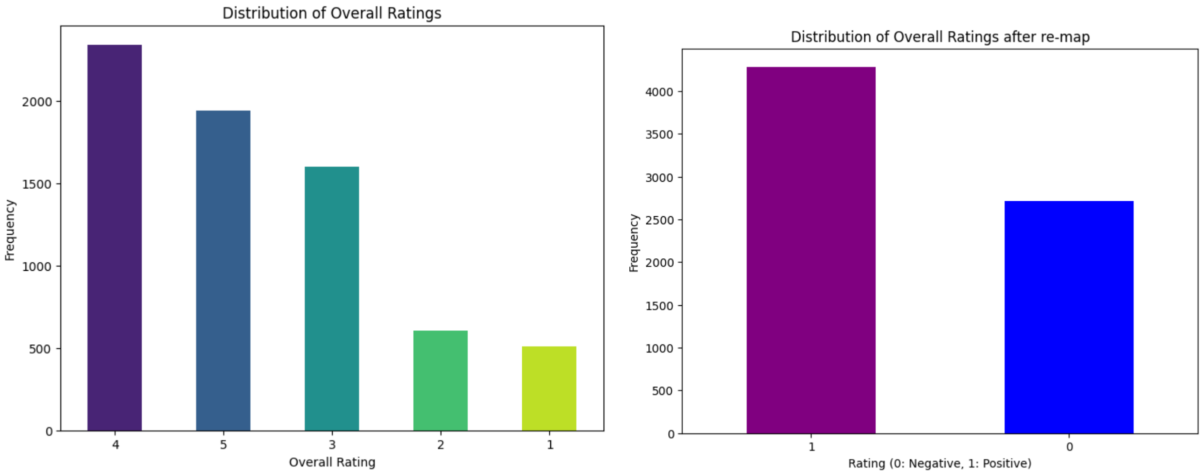


Figure 7: Distribution of overall ratings

The punctuation, special characters, and symbols were removed from the from the text and converted into lower case data. This step ensures uniformity by treating identical words in different cases (e.g., "Word" and "word") as the same, preventing potential duplicates during analysis. Stopwords are frequently encountered words (e.g., "are," "you," "it") that typically make a limited contribution to the overall context and significance of the text. The function utilizes the stopwords from English corpus from the NLTK library to filter out such stopwords. This process helps focus on more meaningful words and reduces noise in the text data.

Tokenization is the process of breaking down text into smaller units, typically words or phrases referred to as tokens. Tokenization is performed using using functions like word_tokenize available in libraries such as NLTK. Each sentence or comment is split into individual tokens, aiding in subsequent analyses by providing a structured representation of the text. Tokenization serves as a foundational step in NLP tasks, enabling text manipulation, analysis, and feature extraction. It forms the basis for various downstream processes, including but not limited to sentiment analysis, topic modelling, and POS tagging.

Lemmatization is the process of reducing words to their fundamental or base form, referred to as a lemma. For instance, verbs in different tenses or nouns in plural form are transformed into their base forms (e.g., "running" becomes "run," "mice" becomes "mouse"). The

WordNetLemmatizer from NLTK is utilized to perform lemmatization on tokenized words. Lemmatization aids in standardizing words to their canonical forms, reducing inflectional forms to a common base. This process helps in normalisation, reducing the vocabulary size, and improving the accuracy of analyses and models. It ensures that different inflected forms of a word are treated as the same, which is crucial for tasks like text classification, information retrieval, and machine learning models, where a smaller and more consistent vocabulary is beneficial.

The dataset underwent a partitioning process into training and validation sets, adhering to an 8:2 ratio, where 80% of the dataset was allocated for training the models, and the remaining 20% was set aside for evaluating the models' performance. The training subset was employed to optimize the model's parameters, while the testing subset was utilized to gauge the model's performance on unseen instances and assess its generalization capabilities.

4.3 Model Development, Training and Fine Tuning

The computations for this study were performed using Google Colab Pro, which provided access to advanced computational resources, including the Volta Tesla 100 GPU version. This enhancement in speed and scalability allowed for efficient model training and experimentation with complex deep learning architectures.

The development of transformer-based models for sentiment analysis involved several crucial steps and considerations:

Model Selection: Five transformer-based models, namely DistilBERT, BERT, RoBERTa, DeBERTa, and XLNet, were chosen as they are the most advanced transformer-based models with superior capabilities to understand the semantic parts of the text, making them suitable for task of sentiment analysis.

Model Training and Optimization: The training loop involved setting up the optimizer, initializing the model's parameters, and iterating through multiple epochs. During training, the models were fine-tuned on the training dataset to minimize the loss function using the AdamW optimizer. A linear scheduler was employed to adjust the learning rate, optimising the models' performance over the training iterations. The models were trained over 10 epochs to ensure comprehensive learning and optimization of parameters. The decision to use a batch size of 8 was made based on a balance between GPU memory constraints and computational efficiency.

Model Evaluation: The models' performance was evaluated on the testing dataset to assess their ability to generalize to new, unseen instances. Metrics such as accuracy, classification reports, and confusion matrices were generated to analyse the models' predictive capabilities and identify their strengths and weaknesses in sentiment classification.

Save the model for future use: The model was saved to preserve its architecture and trained weights for future use. Alongside the model, the tokenizer necessary for processing text inputs was also saved. Additionally, essential details about the model, such as its specifications and configurations, were stored in a separate file. Moreover, this metadata includes information such as the model's name, the count of labels, training epochs, batch size, and learning rate. These saved components ensure the model's reproducibility and facilitate its integration for subsequent analyses and applications. The stored information can be easily reloaded to utilize the model and its associated metadata efficiently.

Predicting the test data: The test data was loaded, and the same preprocessing steps were applied to ensure consistency in data preparation. Subsequently, the best model along with its tokenizer was loaded. Predictions were generated using this loaded model and tokenizer on the pre-processed test data to assess the model's performance on previously unseen data.

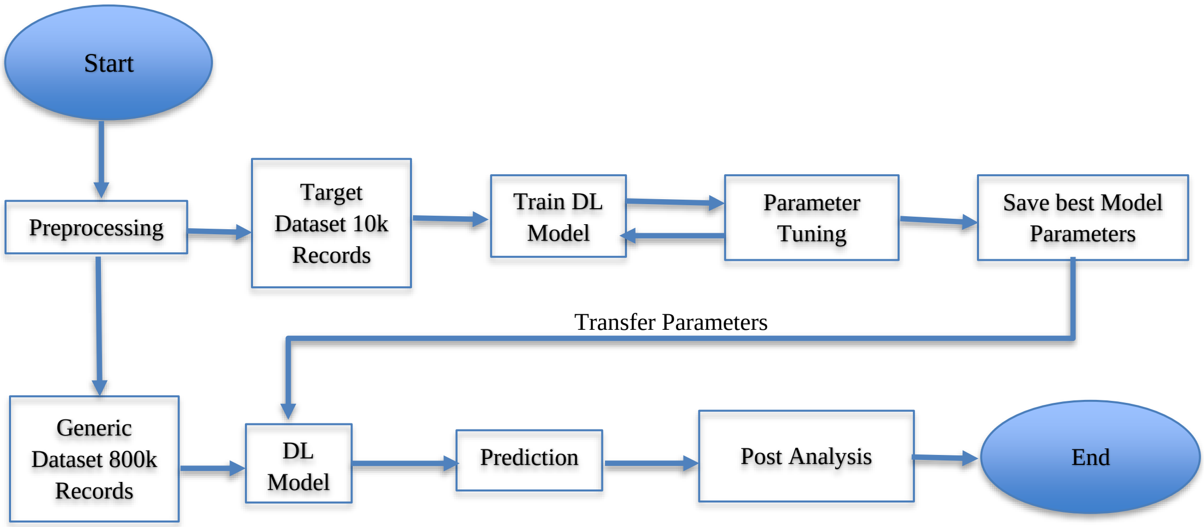


Figure 8: Overview of the development pipeline used.

Figure 8 depicts the ML model development and training pipeline was designed to effectively model the target dataset and generate meaningful predictions on unseen data. The initial stage involved preprocessing the data through tokenization and lemmatization, preparing it for model ingestion. Transformer-based models were trained on the 10,000-record target dataset, with the optimal model and its parameters being saved for subsequent use. To ensure consistency, unseen data underwent identical preprocessing steps before predictions were generated using the saved model. Post-analysis techniques, including word clouds, trend analyses, and other visualizations, were employed to explore and interpret the predictions. Further insights were extracted through LDA to discover latent topics within the review data, and aspect-based sentiment analysis to gauge sentiment towards specific features.

4.4 Identifying Prevalent Topics across Companies

Two distinct algorithms namely, LDA and NMF, were utilised to extract themes from the textual data. The process began with initializing CountVectorizer for LDA and TF-IDF Vectorizer for NMF, aiming to preprocess and transform the text data into numerical representations suitable for topic modelling. For LDA, a CountVectorizer was set up with parameters including the maximum document frequency, minimum document frequency, maximum number of features, and English stop words. Similarly, for NMF, a TF-IDF Vectorizer was set up with analogous parameters.

Subsequently, LDA and NMF models were fitted to their respective transformed data. LDA utilized the CountVectorizer-transformed data, while NMF leveraged TF-IDF transformed data. Each model extracted 15 topics for LDA and NMF. Top words associated with each topic were identified and stored, capturing the most representative words defining each topic.

Finally, to visually represent the topic prevalence, average topic probabilities across all documents were calculated for both LDA and NMF. A bar plot was generated, depicting the average prevalence of each topic identified by NMF. This method was preferred over LDA due to its robustness in handling matrix factorization, which often results in more interpretable topics, especially when dealing with text data.

4.5 Implementing Aspect based Sentiment Analysis

The aspect-based sentiment analysis was conducted by utilizing spaCy's named entity recognition (NER) capabilities and VADER sentiment analysis. The process involved loading a pre-trained spaCy model and extracting entities, such as persons, organizations, dates, etc. from the provided test dataset.

The generated bar plot in figure 9 illustrates the frequency distribution of named entity labels identified within the dataset using spaCy's natural language processing capabilities. Each bar represents a specific entity label, such as PERSON (referring to individuals), ORG (representing organizations), and DATE (indicating temporal expressions). The height of each bar denotes the occurrence frequency of the respective entity label, providing a visual depiction of the prevalence of different types of named entities within the analysed text data. This visualization offers insights into the diverse nature of entities present, highlighting the prominence of certain categories, thereby aiding in understanding the dataset's content and contextual nuances.

4.6 Utilised Python Libraries

The choice and integration of diverse libraries and packages in this study were instrumental in implementing a comprehensive NLP and ML framework. The TensorFlow, Transformers, Torch, and Scikit-learn libraries formed the backbone of the analysis, providing an array of tools for neural network implementation, model development, and efficient data manipulation. Utilizing the functionalities offered by these libraries, a range of transformer-based architectures including BERT, RoBERTa, DistilBERT, DeBERTa, and XLNet became readily available. This facilitated in-depth exploration of pretrained models and their customization for specific tasks. Additionally, NLTK, Gensim, and SpaCy facilitated essential text preprocessing tasks including tokenization, lemmatization, and stop word removal. Furthermore, Scikit-learn's machine learning utilities offered a comprehensive suite of tools for model evaluation, hyperparameter tuning, and data transformation, while the VaderSentiment package allowed for sentiment analysis. The utilization of these libraries collectively ensured a robust framework for NLP and ML tasks, leveraging their efficiency, scalability, and diverse functionalities to explore, preprocess, model, and evaluate textual data effectively. This amalgamation of libraries provided a rich toolbox, offering versatility and efficiency in handling various components of the research pipeline, ultimately contributing to a comprehensive and well-structured analysis.

CHAPTER 5 – RESULTS & DISCUSSION

Choosing the Glassdoor data: While datasets on platforms like Kaggle serve their specific purposes, the Glassdoor dataset stands out due to its authenticity, comprehensiveness, diversity, and depth. These attributes make it more suitable for conducting in-depth analysis, supporting decision-making, and fostering nuanced research in areas related to employee sentiments, company cultures, and market dynamics.

Importance of Binary Target: The decision to convert 'overall_rating' into a binary classification (positive/negative sentiment) rather than multiple targets is essential for sentiment analysis tasks. This binary classification simplifies the prediction task, enabling the models to distinguish between positive and negative sentiments effectively. It aligns with the goal of extracting insights regarding company perceptions based on employee sentiments, focusing on areas of improvement (negative sentiment) and positive aspects. The binary target enhances the interpretability of the model's output, facilitating a clearer understanding of sentiment distribution across reviews and aiding in actionable insights for companies.

Text Preprocessing:

Text preprocessing plays a pivotal role in NLP and text mining tasks. Its significance lies in enhancing the quality of textual data for subsequent analysis by machine learning models or other NLP techniques. The steps involved in preprocessing aim to clean, standardize, and prepare the text, enabling more effective feature extraction and pattern recognition. By removing irrelevant characters such as regular expressions, filtering stopwords and converting to lowercase, the preprocessing function streamlines the text data, making it more applicable for NLP tasks like sentiment analysis, topic modelling, classification, and information retrieval.

Data splitting:

This splitting strategy ensured that a considerable portion of the data was utilised for model training to learn patterns and relationships while being evaluated on a separate dataset to measure their performance and ability to generalize to new, unseen data. Additionally, a separate dataset, not used in the training or validating phases, was kept aside for further testing and as a completely unseen dataset to assess the models' performance in real-world scenarios.

Model outputs:

DistilBERT

```
Classification report:
              precision    recall  f1-score   support
           0       0.71      0.60      0.65       527
           1       0.78      0.85      0.81       873
    accuracy                           0.76      1400
   macro avg       0.74      0.73      0.73      1400
weighted avg       0.75      0.76      0.75      1400

Confusion Matrix:
               Predicted Negative  Predicted Positive
True Negative                 315                 212
True Positive                 128                 745
```

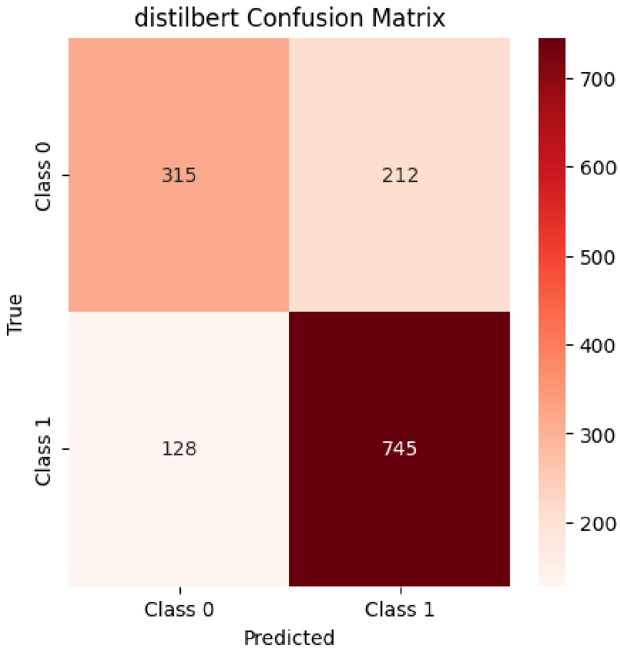


Figure 9: Distil BERT confusion metric.

The DistilBERT model obtained a 76% overall accuracy in its classification report, signifying its capability to distinguish sentiments as positive or negative within the reviews. Specifically, the precision and recall for negative sentiment (Class 0) were 71% and 60%, respectively, while for positive sentiment (Class 1), they were 78% and 85%. The model exhibited a better performance in correctly identifying positive sentiments, as evident from higher precision and recall values for Class 1. Despite a slightly lower precision and recall for negative sentiments, the model demonstrated a more robust ability to detect positive sentiments. The overall performance metrics, including the balanced accuracy, the average F1-score across different classes (macro-average), and the weighted F1-score, were recorded at 73%, 73%, and 75%, respectively, reflect a generally satisfactory performance in sentiment classification, with a slight inclination towards accurate prediction of positive sentiments.

BERT

```
Classification report:
              precision    recall  f1-score   support
           0       0.74      0.56      0.64       527
           1       0.77      0.88      0.82       873

    accuracy                           0.76      1400
   macro avg       0.75      0.72      0.73      1400
weighted avg       0.76      0.76      0.75      1400

Confusion Matrix:
               Predicted Negative  Predicted Positive
True Negative                 297                 230
True Positive                 106                 767
```

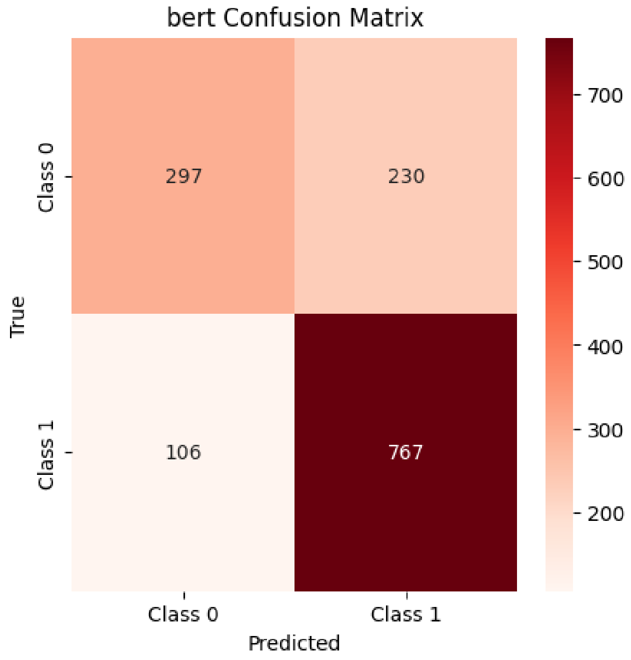


Figure 10:BERT confusion metric

The BERT model demonstrates favourable performance with an accuracy of 75%. It demonstrates higher precision, recall, and F1-score for predicting positive sentiments (class 1), achieving an 82% F1-score. Conversely, it presents relatively lower metrics for predicting negative sentiments (class 0), with a 64% F1-score. The macro-average F1-score, signifying balanced accuracy, stands at 73%. The weighted average F1-score is 75%, indicating a marginally stronger performance in identifying positive sentiments. Delving into the confusion matrix, the model accurately predicted 297 negative sentiments but misclassified 230 negatives as positive. Additionally, it correctly predicted 767 positive sentiments, yet misclassified 106 positives as negative. The model shows proficiency in identifying positive sentiments but reveals potential for enhancement in predicting negative sentiments. These results suggest the model's strength in discerning positive sentiments while indicating opportunities for refining its performance in capturing negative sentiments.

RoBERTa

```
Classification report:
             precision    recall  f1-score   support
           0       0.73      0.58      0.65       527
           1       0.78      0.87      0.82       873
    accuracy                           0.76      1400
   macro avg       0.75      0.73      0.74      1400
weighted avg       0.76      0.76      0.76      1400

Confusion Matrix:
              Predicted Negative  Predicted Positive
True Negative                308                 219
True Positive                113                 760
```

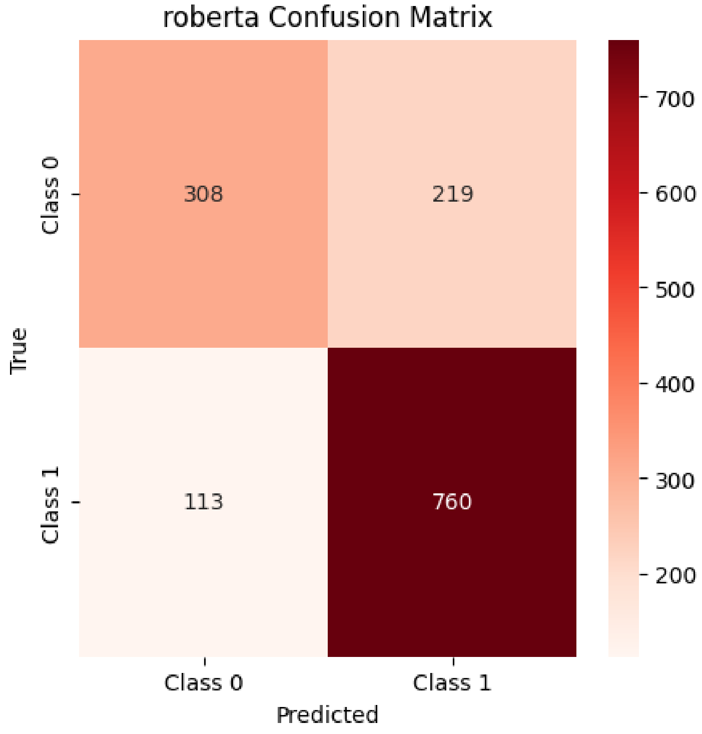


Figure 11: RoBERTa confusion metric

The Roberta model demonstrated an overall accuracy of 76.29% in sentiment analysis, indicating a moderate capacity for sentiment classification. However, its performance exhibited a notable disparity between classes. While precision for the negative class reached 73%, recall was notably lower at 58%, suggesting a greater tendency to overlook negative sentiments.

Conversely, precision for the positive class reached 78%, with a recall of 87%, signifying a stronger ability to identify positive sentiments accurately. This discrepancy could potentially stem from an imbalance in the training dataset, with more positive samples leading to a better grasp of that class. Alternatively, it might reflect challenges in capturing subtle linguistic cues or contextual nuances that signal negative sentiment. Further fine-tuning with balanced data and attention to negative language markers could potentially enhance model performance across both classes.

DeBERTa

```
Classification report:
              precision    recall  f1-score   support
           0       0.73      0.59      0.65       527
           1       0.78      0.87      0.82       873
    accuracy                           0.76      1400
   macro avg       0.75      0.73      0.74      1400
weighted avg       0.76      0.76      0.76      1400

Confusion Matrix:
               Predicted Negative  Predicted Positive
True Negative                 310                 217
True Positive                 115                 758
```

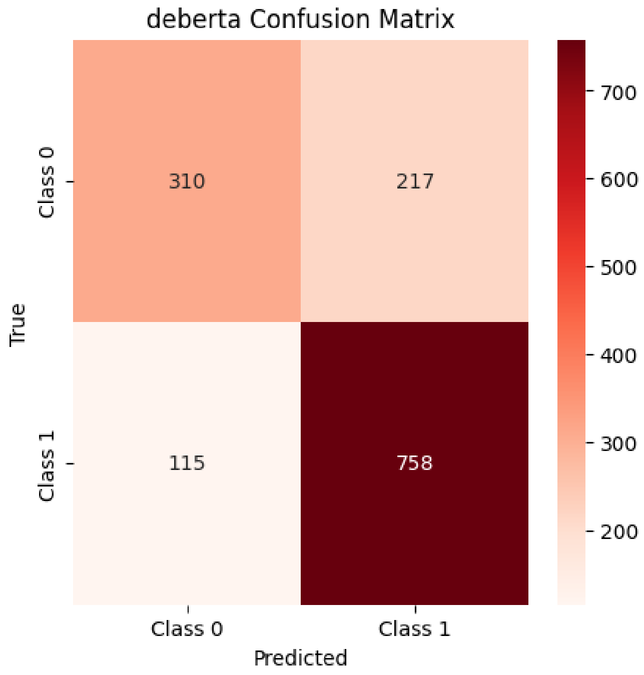


Figure 12: DeBERTa confusion metric

The DeBERTa model attained an overall accuracy of 76.29% in sentiment analysis. While precision for the negative class (0) was 73%, precision for the positive class (1) reached 78%.

Recall demonstrated a greater disparity, with 59% for the negative class and 87% for the positive class. The balanced accuracy, accounting for both precision and recall, was 72.83%, mirrored in the macro average F1-score of 74%. The weighted average F1-score of 76% also suggests the model's effectiveness in generalizing across classes. Notably, the confusion matrix reveals a tendency to predict positive sentiments more accurately than negative ones.

XLNet

```
Classification report:
              precision    recall  f1-score   support
           0       0.74      0.57      0.65       527
           1       0.77      0.91      0.83       873
    accuracy                           0.77      1400
   macro avg       0.76      0.73      0.74      1400
weighted avg       0.76      0.77      0.76      1400

Confusion Matrix:
               Predicted Negative  Predicted Positive
True Negative                 303                 224
True Positive                 105                 768
```

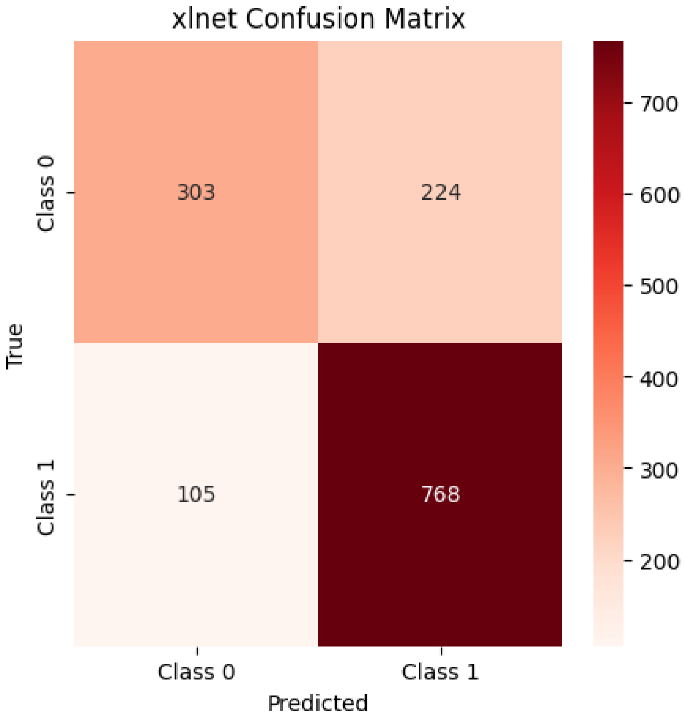


Figure 13: XLNet confusion metric

The XLNet model achieved an overall accuracy of 77% in sentiment analysis. However, its performance exhibited notable differences between classes. Precision for the negative class (0) was 74%, while precision for the positive class (1) reached 77%. Recall demonstrated a more pronounced disparity, with 57% for the negative category comparison to 91% for the positive category. This imbalance can be seen in the results, as shown by the balanced accuracy of 73% and the macro average F1-score of 74%. While the weighted average F1-score of 76% suggests the model's ability to generalize across classes, the confusion matrix clearly indicates a stronger tendency to accurately predict positive sentiments over negative ones.

Following equation denotes the equation for the balanced accuracy:
balanced_accuracy = (class_recall['0'] + class_recall['1']) / 2

Table 1: Model performance summary

Model	Training Time (seconds)	Accuracy	Balanced Accuracy	Precision	Recall	F1-score
DistilBert	1609.6	0.76	0.73	0.78	0.85	0.81
BERT	3071.2	0.76	0.72	0.77	0.88	0.82
RoBERTa	3062.2	0.76	0.73	0.78	0.87	0.82
DeBERTa	4952.5	0.76	0.73	0.78	0.87	0.82
XLNet	6715.4	0.77	0.73	0.77	0.91	**0.83**

Best model:

Among the evaluated models, XLNet appears to be the best-performing model for sentiment analysis based on several factors. While XLNet's F1-score of 0.83 is the highest among the models, indicating its robustness in precision and recall trade-off, other factors contribute to its recognition as the best model.

XLNet demonstrates a commendable balance between precision (77%) and recall (91%) for sentiment classification, particularly excelling in correctly identifying positive sentiments while maintaining reasonable accuracy in detecting negative sentiments.

Moreover, although XLNet requires a longer training time compared to some other models, such as DistilBERT or BERT, its overall performance by the metrices, including balanced accuracy, accuracy, precision, recall, and F1-score, collectively justify its slightly longer training duration.

Additionally, XLNet's ability to maintain a high F1-score while achieving a balanced accuracy of 73% suggests its effectiveness in handling imbalanced classes, ensuring reliable performance in both positive and negative sentiment identification.

Overall, XLNet stands out as the preferred model for sentiment analysis due to its superior F1-score, strong balance between precision and recall, and robust performance across multiple evaluation metrics, affirming its effectiveness in capturing nuanced sentiment expressions within the dataset.

Prediction Results:

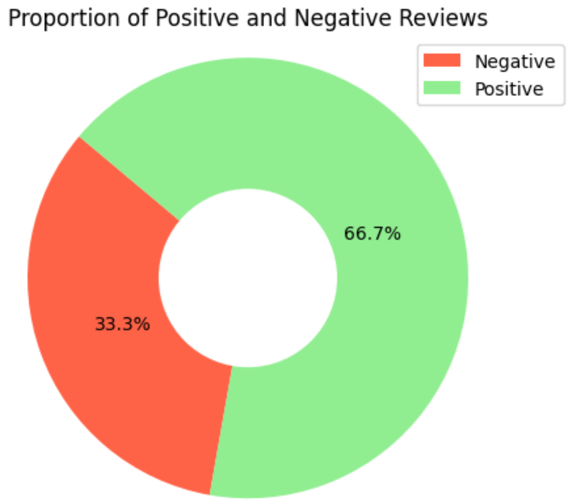


Figure 14: Proportion of Positive and Negative Sentiments

Figure 15 visually represents the portion of predicted classes (either positive or negative) among reviews, potentially derived from a classification model's predictions. Each slice of the chart corresponds to a class, showcasing the distribution of these classes. The 'Positive' class is depicted in light green, while the 'Negative' class is represented in tomato red. The percentage displayed on each slice denotes the proportion of each class in the dataset.

This visualization is essential for understanding the balance or imbalance between different classes in the dataset or model predictions. It provides an intuitive overview of the relative frequencies of predicted classes quickly. Imbalanced class distributions could impact a model's performance, and this visualization helps in identifying such discrepancies. Additionally, it aids in evaluating whether the model is biased towards any specific class, highlighting potential areas for model improvement, such as addressing class imbalances or adjusting prediction thresholds.

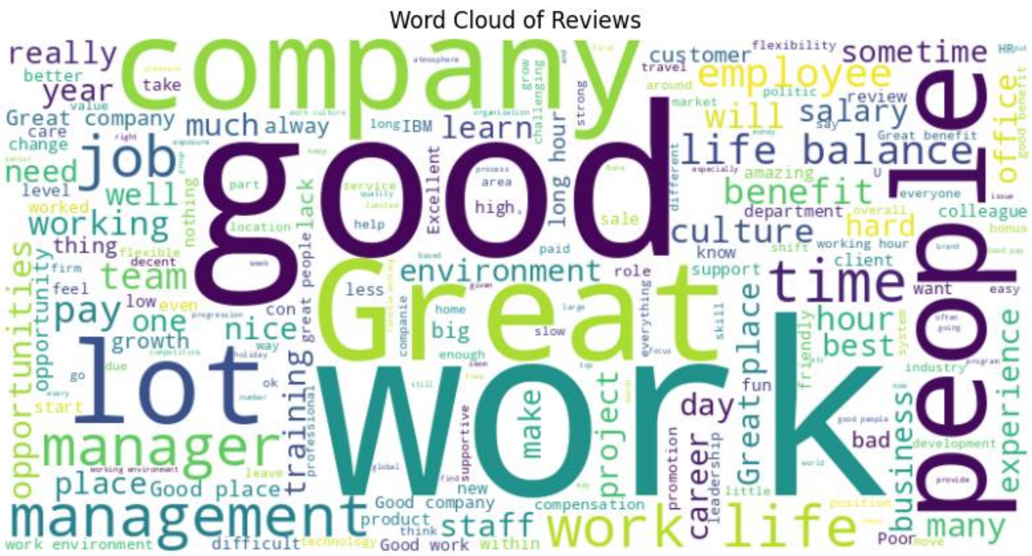

Word Cloud of Reviews

Figure 15: Word cloud visualisation of Company Reviews

Word clouds serve as visual representations highlighting the most frequently occurring words within a text or corpus, emphasizing word frequency rather than context or sentiment. However, in the context of sentiment analysis, word clouds might not capture sentiment polarity, particularly with words containing negation, such as 'not.' For instance, phrases like 'not good' might only display 'good' in the word cloud, missing the negation's effect on the sentiment. Sentiment analysis requires a solid understanding of context. It uses word usage to accurately determine sentiment. Relying solely on word frequency, as depicted in word clouds, might not fully reflect the true sentiment polarity of the text, especially when negation or contextual variations alter the meaning of words. Despite of this drawback, they were considered here because they offer a quick and intuitive overview of the most frequent terms in a text, which can be useful for initial exploratory analysis or gaining a general understanding of the main topics or themes present.

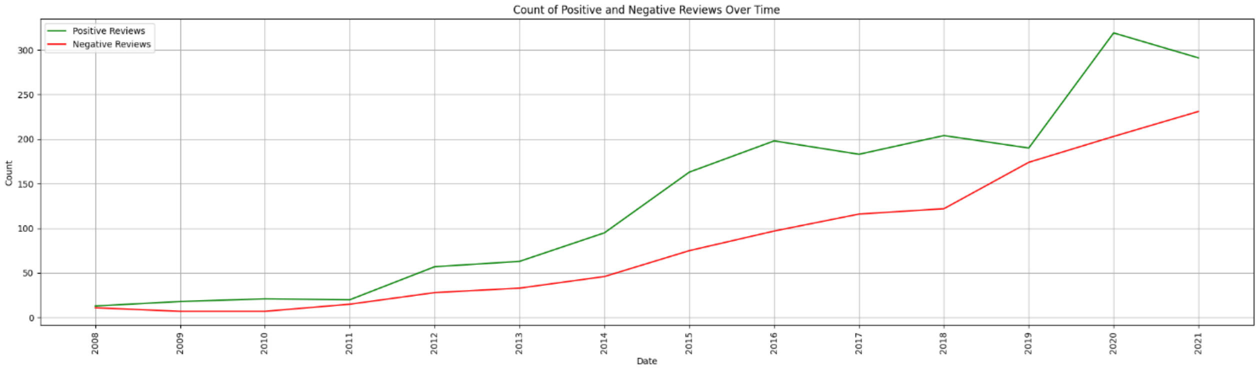


Figure 16: Review Trends: Positive and Negative Over Time

Figure 16 visualizes the count of positive and negative reviews over time. The x-axis represents time intervals based on the selected period (Annual, Monthly, or Daily), while the y-axis represents the review count. The green line indicates the count of positive reviews, whereas the red line represents the count of negative reviews.

Interpreting this graph allows us to understand the temporal distribution and trends of positive and negative sentiments within the reviews dataset. It assists in identifying patterns or fluctuations in sentiment over time, offering insights into potential shifts in employer opinions or satisfaction levels across different time periods. For instance, an increasing trend in negative reviews over specific dates or periods might indicate issues or concerns that emerged during those times, prompting further investigation or action.

Such visualizations are crucial in sentiment analysis as they provide a clear representation of sentiment dynamics, enabling businesses or analysts to track changes, recognize patterns, and make decisions built upon the temporal aspect in sentiment fluctuations within the dataset.

Topic Modelling:

In the context of the analysis conducted, the selection of NMF over LDA for topic modeling was primarily driven by the data's structured nature and the specific requirements of the analysis goals. NMF, with its ability to capture inherent sparsity in datasets with structured information, was preferred as it aligns well with such data characteristics. Moreover, the inherent simplicity and relatively faster computation of NMF, particularly beneficial for scalability with larger datasets, contributed to its selection. The interpretability factor was also a consideration, acknowledging that although LDA may often yield more interpretable topics, manually assessing interpretability revealed that NMF's results were better aligned with the domain expectations based on prior domain knowledge.

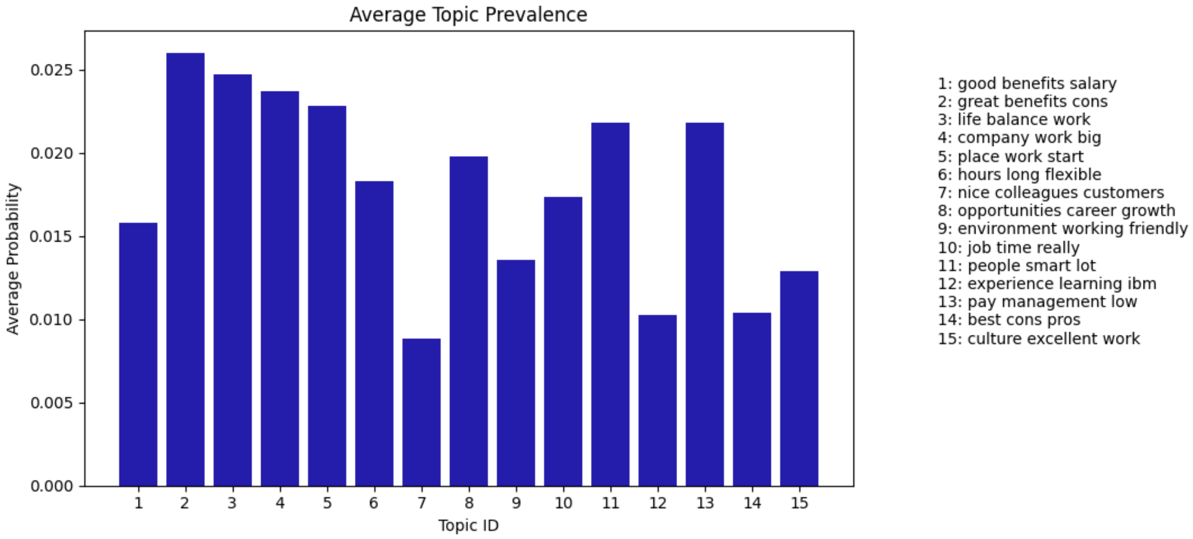


Figure 17: Insights from Topic Modelling

Figure 17 portrays the average probability of each topic derived from NMF, where each topic is represented by its respective ID and the most significant words characterizing that particular topic. This approach provides insights into prevalent themes within the dataset, aiding in understanding the underlying topics and their prevalence within the corpus.

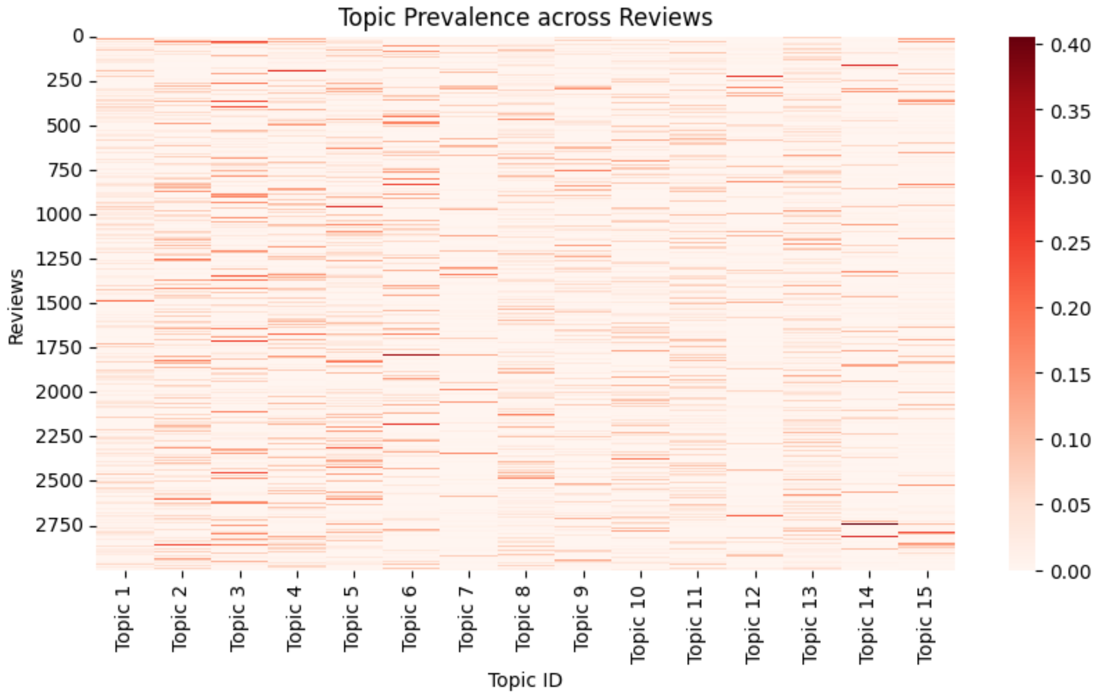


Figure 18: Heatmap of the topic prevalence

In this visualization process, the aim was to represent the prevalence of topics across the reviews. The topic probabilities generated by Non-Negative Matrix Factorization (NMF) were used for constructing a heatmap. Each row in the heatmap denoting a review and each column corresponding to a specific topic identified by NMF. The intensity of colour in each cell represents the prevalence or probability of a particular topic within a review. The heatmap provides a comprehensive overview, allowing easy identification of the dominant topics within individual reviews and their distribution across the dataset.

Aspect-based Sentiment Analysis:

Figure 19 represents the frequency distribution of various entity labels extracted from the processed sentences using a Natural Language Processing (NLP) model, likely SpaCy. Each bar on the chart corresponds to a specific named entity label, such as "PERSON," "ORG" (organization), or "DATE." The height of each bar indicates the frequency or count of occurrences for each respective entity label found within the analyzed text corpus. This visualization helps in understanding the prevalence or occurrence frequency of different types of named entities in the dataset, providing insights into the types of information or entities present in the text data.

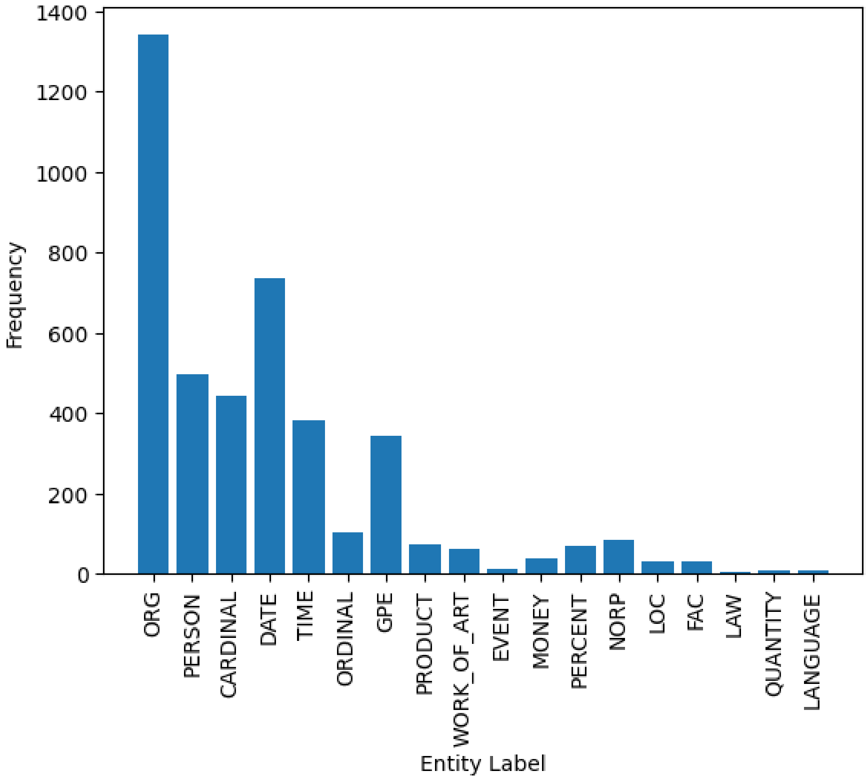
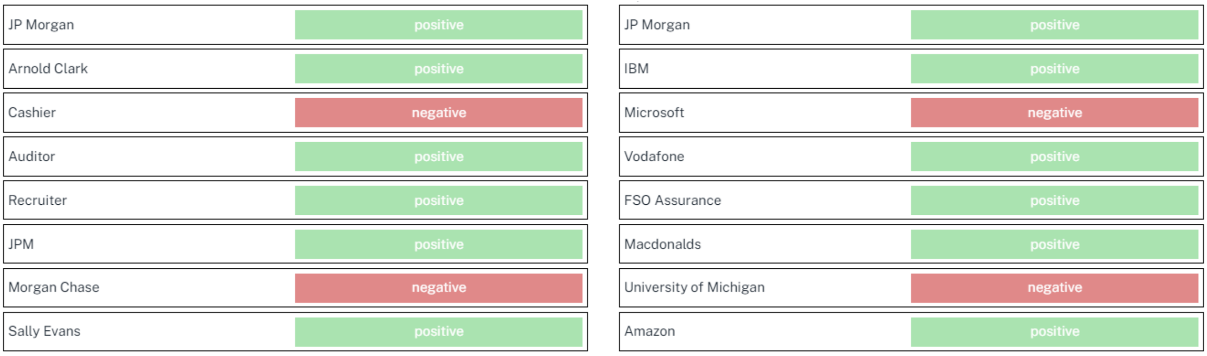


Figure 19: Frequencies of the identified named entities in the data

The dataset consisted of reviews or texts, and for each sentence in the dataset, the spaCy model identified the entities and assigned sentiments to each entity using VADER sentiment analysis. The sentiment analysis generated sentiment scores for positive and negative sentiments. These scores were aggregated and visualized using color-coded HTML tables, illustrating sentiment trends for various entities. This analysis aimed to demonstrate the capability of identifying entities and their associated sentiments without requiring domain-specific expertise. The output served as a demonstration for higher-level management or clients through a web-based interface.



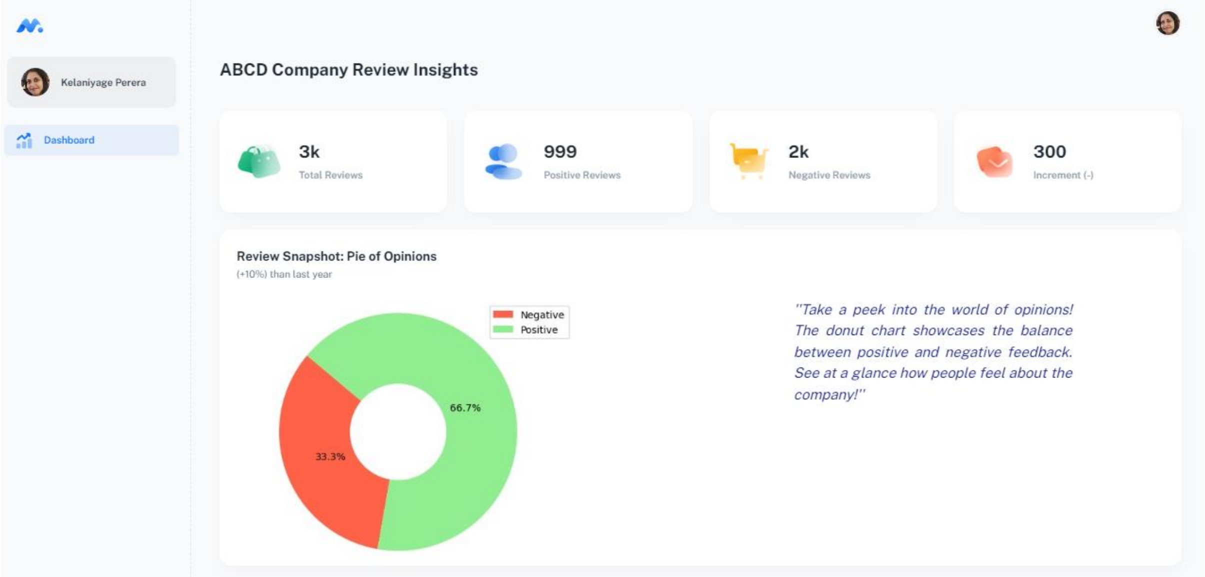
JP Morgan	positive
Arnold Clark	positive
Cashier	negative
Auditor	positive
Recruiter	positive
JPM	positive
Morgan Chase	negative
Sally Evans	positive

JP Morgan	positive
IBM	positive
Microsoft	negative
Vodafone	positive
FSO Assurance	positive
Macdonalds	positive
University of Michigan	negative
Amazon	positive

Figure 20:Positive and negative sentiments related to person and organisational entities

Prototype Web application:

This approach can be utilised for commercial implementation by developing a comprehensive web application. However, a prototype version was developed using ReactJS and deployed on an AWS S3 bucket. The application is publicly accessible and published for demonstration purpose via the URL: http://project-uog.s3-website.eu-north-1.amazonaws.com/. Figure 21-26 shows the screenshots captured from the above prototype web interfaces.


Kelaniyage Perera

Dashboard

ABCD Company Review Insights

3k
Total Reviews

999
Positive Reviews

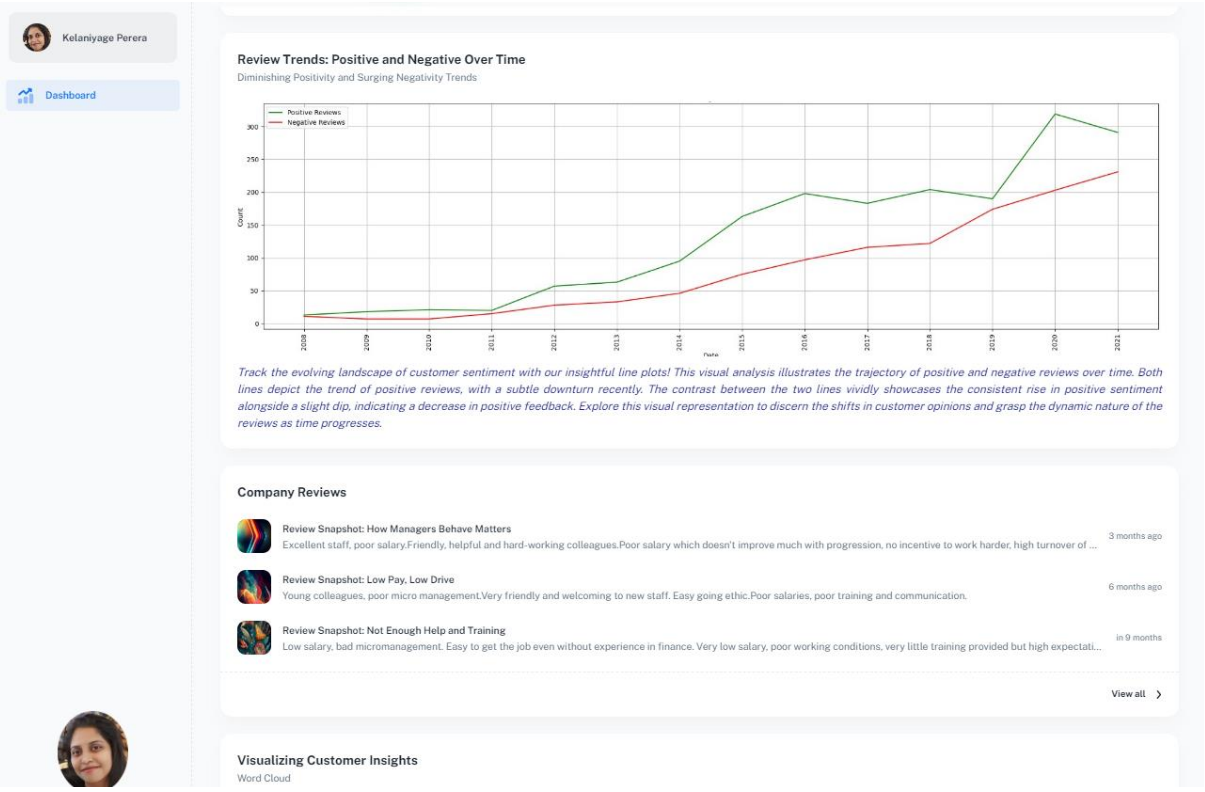
2k
Negative Reviews

300
Increment (-)

Review Snapshot: Pie of Opinions
(+10%) than last year

"Take a peek into the world of opinions! The donut chart showcases the balance between positive and negative feedback. See at a glance how people feel about the company!"

Figure 21: Web interface_donut plot


Kelaniyage Perera

Dashboard

Review Trends: Positive and Negative Over Time
Diminishing Positivity and Surging Negativity Trends

Track the evolving landscape of customer sentiment with our insightful line plots! This visual analysis illustrates the trajectory of positive and negative reviews over time. Both lines depict the trend of positive reviews, with a subtle downturn recently. The contrast between the two lines vividly showcases the consistent rise in positive sentiment alongside a slight dip, indicating a decrease in positive feedback. Explore this visual representation to discern the shifts in customer opinions and grasp the dynamic nature of the reviews as time progresses.

Company Reviews

Review Snapshot: How Managers Behave Matters
Excellent staff, poor salary.Friendly, helpful and hard-working colleagues.Poor salary which doesn't improve much with progression, no incentive to work harder, high turnover of ...
3 months ago

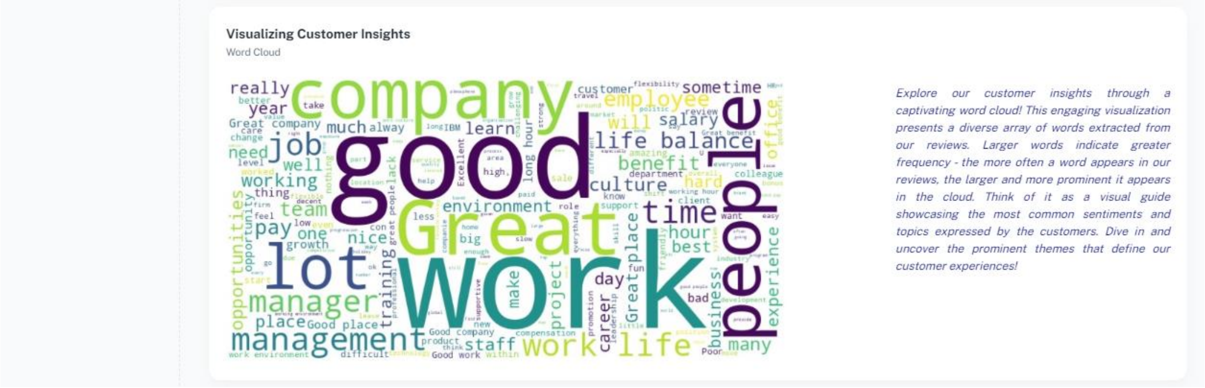
Review Snapshot: Low Pay, Low Drive
Young colleagues, poor micro management.Very friendly and welcoming to new staff. Easy going ethic.Poor salaries, poor training and communication.
6 months ago

Review Snapshot: Not Enough Help and Training
Low salary, bad micromanagement. Easy to get the job even without experience in finance. Very low salary, poor working conditions, very little training provided but high expectati...
in 9 months

View all ›

Visualizing Customer Insights
Word Cloud

Figure 22: Web interface_sample reviwes

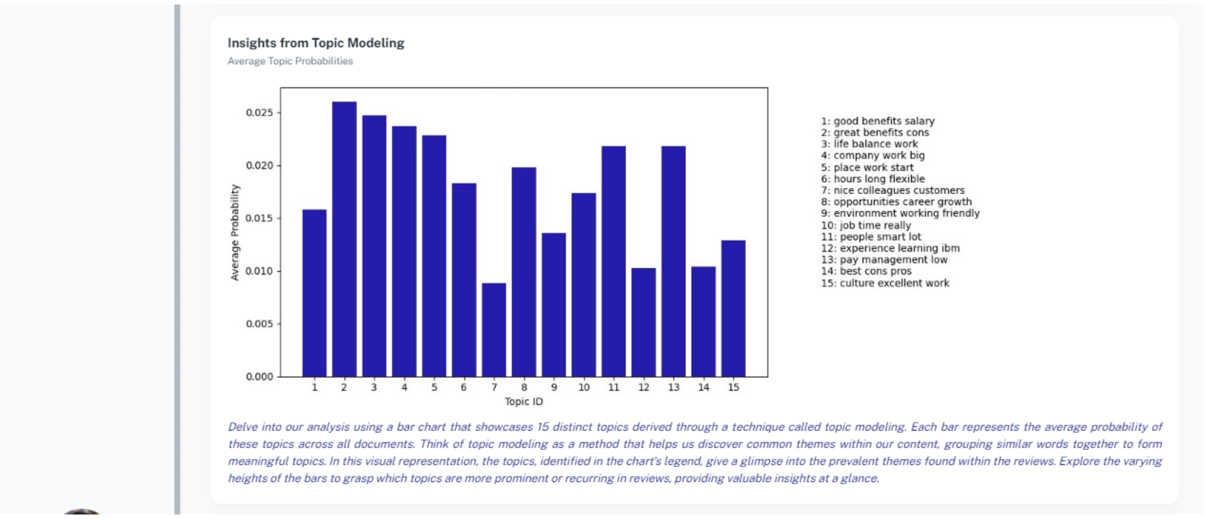


Visualizing Customer Insights

Word Cloud

Explore our customer insights through a captivating word cloud! This engaging visualization presents a diverse array of words extracted from our reviews. Larger words indicate greater frequency - the more often a word appears in our reviews, the larger and more prominent it appears in the cloud. Think of it as a visual guide showcasing the most common sentiments and topics expressed by the customers. Dive in and uncover the prominent themes that define our customer experiences!

Figure 23: Web interface_wordcloud



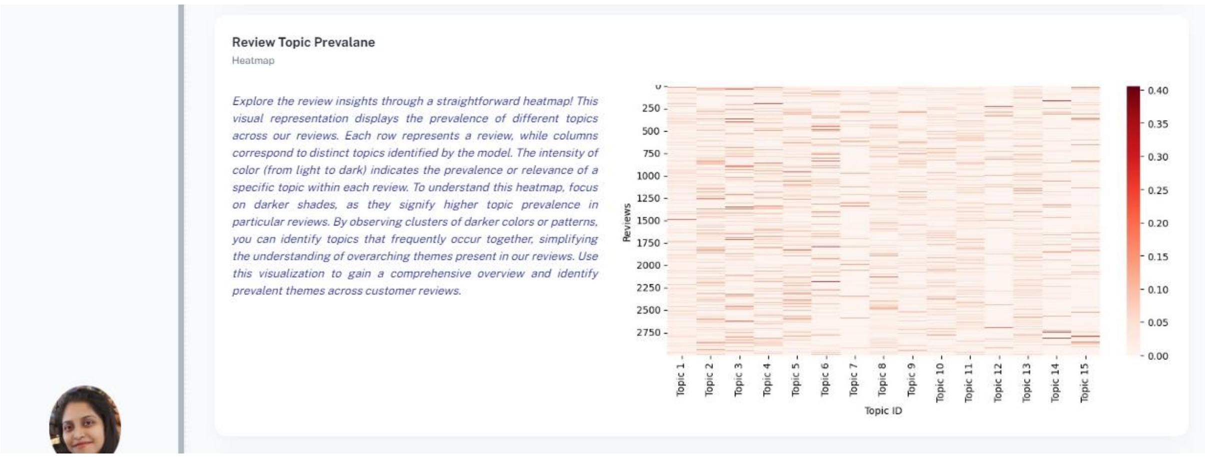
Insights from Topic Modeling

Average Topic Probabilities

1: good benefits salary
2: great benefits cons
3: life balance work
4: company work big
5: place work start
6: hours long flexible
7: nice colleagues customers
8: opportunities career growth
9: environment working friendly
10: job time really
11: people smart lot
12: experience learning ibm
13: pay management low
14: best cons pros
15: culture excellent work

Delve into our analysis using a bar chart that showcases 15 distinct topics derived through a technique called topic modeling. Each bar represents the average probability of these topics across all documents. Think of topic modeling as a method that helps us discover common themes within our content, grouping similar words together to form meaningful topics. In this visual representation, the topics, identified in the chart's legend, give a glimpse into the prevalent themes found within the reviews. Explore the varying heights of the bars to grasp which topics are more prominent or recurring in reviews, providing valuable insights at a glance.

Figure 24: Web interface_topic prevalence

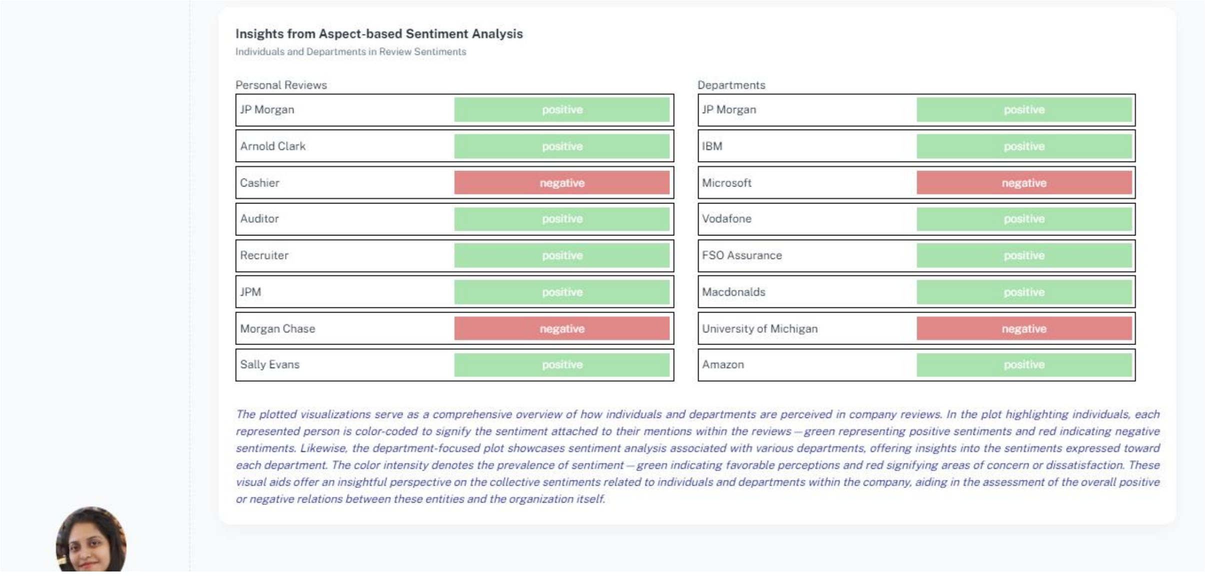


Review Topic Prevalane

Heatmap

Explore the review insights through a straightforward heatmap! This visual representation displays the prevalence of different topics across our reviews. Each row represents a review, while columns correspond to distinct topics identified by the model. The intensity of color (from light to dark) indicates the prevalence or relevance of a specific topic within each review. To understand this heatmap, focus on darker shades, as they signify higher topic prevalence in particular reviews. By observing clusters of darker colors or patterns, you can identify topics that frequently occur together, simplifying the understanding of overarching themes present in our reviews. Use this visualization to gain a comprehensive overview and identify prevalent themes across customer reviews.

Figure 25: Web interface_heatmap



Insights from Aspect-based Sentiment Analysis

Individuals and Departments in Review Sentiments

Personal Reviews

JP Morgan	positive
Arnold Clark	positive
Cashier	negative
Auditor	positive
Recruiter	positive
JPM	positive
Morgan Chase	negative
Sally Evans	positive

Departments

JP Morgan	positive
IBM	positive
Microsoft	negative
Vodafone	positive
FSO Assurance	positive
Macdonalds	positive
University of Michigan	negative
Amazon	positive

The plotted visualizations serve as a comprehensive overview of how individuals and departments are perceived in company reviews. In the plot highlighting individuals, each represented person is color-coded to signify the sentiment attached to their mentions within the reviews – green representing positive sentiments and red indicating negative sentiments. Likewise, the department-focused plot showcases sentiment analysis associated with various departments, offering insights into the sentiments expressed toward each department. The color intensity denotes the prevalence of sentiment – green indicating favorable perceptions and red signifying areas of concern or dissatisfaction. These visual aids offer an insightful perspective on the collective sentiments related to individuals and departments within the company, aiding in the assessment of the overall positive or negative relations between these entities and the organization itself.

Figure 26: Web interface_ABSA

CHAPTER 6 – FUTURE WORK

Web Application Improvement with Python-Integrated Backend:

Enhancing the prototype web application involves integrating a Python backend to leverage advanced NLP models for real-time analysis. This integration enables seamless communication between the frontend and backend, allowing for dynamic analysis of text inputs. Implementing features such as sentiment analysis, entity recognition, and more interactive visualizations can significantly augment user experience and utility.

Fine-Tuning Strategies:

Exploring and implementing more sophisticated fine-tuning strategies for transformer-based models, particularly for sentiment analysis, can lead to improved performance. Techniques like differential learning rates, discriminative fine-tuning, or gradual unfreezing of layers can help the model focus on specific aspects and adapt to the dataset more effectively.

Transfer Learning and Model Enhancement:

Leveraging transfer learning by initializing models like BERT with pre-trained embeddings from Word2Vec, Glove or FastText could potentially enhance model performance. Fine-tuning these models on the Glassdoor dataset may improve their ability to capture domain-specific nuances and context, leading to more accurate sentiment analysis.

Expansion to Other Domains:

Generalizing the sentiment analysis model beyond company reviews to other domains, such as product reviews, news articles, social media posts, etc., widens its applicability. Adapting the model to diverse datasets and domains requires careful consideration of domain-specific features and language nuances.

Model Generalization Using Multiple Datasets:

Testing the model's generalization across various datasets from different sources and domains is crucial. Evaluating its performance on diverse data sources will assess its robustness and applicability across a broader spectrum of textual data, enhancing its reliability and utility.

Research into Multimodal Approaches:

Exploring multimodal approaches by incorporating not only textual but also visual or other data modalities, such as images or audio, can provide a more comprehensive understanding of sentiment. Integrating these multiple modalities might enhance the model's ability to capture nuanced sentiments present in diverse sources.

Deployment for Real-Time Analysis:

Developing mechanisms for deploying the sentiment analysis model for real-time analysis in various applications, such as social media monitoring tools, customer feedback systems, or market sentiment trackers, can extend its practical utility in decision-making processes.

Expansion to Multilingual Models:

Considering the expansion of current models from English-only training to multilingual capabilities is paramount. Training and adapting models to multilingual datasets can enhance their applicability and usability across diverse languages and global contexts, broadening their scope and impact.

Ethical Considerations and Bias Mitigation:

Conducting extensive research on mitigating biases in sentiment analysis models is crucial. Investigating ethical implications, fairness, and biases inherent in the dataset or model predictions and implementing strategies to minimize biases is imperative for responsible and equitable AI deployment.

Conclusions

The project, aiming for an extensive analysis into sentiment analysis using the Glassdoor dataset, comprising a wealth of employee reviews, successfully achieved all its objectives. The study aimed to extract valuable insights into company cultures, market sentiments, and employee satisfaction. Beginning with a comprehensive dataset of 838,566 records across 18 columns, the study meticulously curated and pre-processed a representative subset, ensuring data quality and uniformity. Employing advanced NLP techniques, the project undertook tokenization, lemmatization, and feature extraction, enabling the transformation of textual data into structured representations for analysis. Five state-of-the-art transformer-based language models, including BERT, DistilBERT, RoBERTa, DeBERTa, and XLNet, were trained and evaluated for sentiment classification. Among these models, XLNet emerged as the top performer, showcasing an impressive F1-score of 0.83, demonstrating superior precision and recall in identifying sentiments.

This study significantly impacts the field of sentiment analysis and NLP by examining into the finer aspects of company reviews and market sentiments. By leveraging NLP techniques and transformer-based models, the study not only validated the effectiveness of these models in sentiment classification but also highlighted the importance of contextual understanding in sentiment analysis. Topic modelling was explored using LDA and NMF offered insights into prevalent themes within the dataset, aiding in understanding underlying topics and their prevalence.

Despite the considerable achievements, the project acknowledges several limitations and avenues for future exploration. One of the limitations includes the reliance on a specific dataset, primarily focusing on employee reviews, which might limit the model's applicability across diverse domains. Future work involves expanding the model's generalization to various domains beyond company reviews, necessitating adaptations to different linguistic patterns and context-specific features. Furthermore, the incorporation of multimodal approaches, ethical considerations, and bias mitigation strategies are essential directions for ensuring the model's fairness, interpretability, and ethical deployment. Additionally, incorporating multilingual capabilities is an opportunity for further research to enhance the model's adaptability and usability across various languages.

The project also suggests refining the web application by integrating a Python backend, allowing real-time analysis, and enhancing the user interface for more intuitive interaction.

In conclusion, the results of this project highlight the significant of transformer-based models when applied to sentiment analysis, offering comprehensive insights into employee sentiments, company cultures, and market dynamics. The advancements made in model performance, topic modelling, and prototype application development lay the groundwork for future research and practical applications across various domains, propelling the field of sentiment analysis towards more accurate, ethical, and widely applicable NLP solutions.

----------------------------End-------------------------------

List of References

Adoma, A.F., Henry, N.M., & Chen, W. (2020). Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition. In 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 117-121). IEEE.

Asur, S., & Huberman, B.A. (2010). Predicting the future with social media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 492-499). IEEE Computer Society.

Bermingham, A., & Smeaton, A.F. (2011). On using Twitter to monitor political sentiment and predict election results. In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011) (pp. 2-10).

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(Jan), 993-1022.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Chen, Q., Zhu, X., Ling, Z.H., Wei, S., Jiang, H., & Inkpen, D. (2019). BERT for Aspect-Based Sentiment Analysis: Survey, Insights, and Challenges. arXiv preprint arXiv:1910.00883.

Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

Filbeck, G., & Preece, D. (2003). Shareholder value, stakeholder management, and social issues: What's the bottom line? Managerial Finance, 29(11), 10-26.

Ghazzawi, A., & Alharbi, B. (2019). Analysis of customer complaints data using data mining techniques. Procedia Computer Science, 163, 62-69.

Glassdoor: About Us. (2016). Retrieved from [https://www.glassdoor.co.uk/index.htm]

Guiso, L., Sapienza, P., & Zingales, L. (2014). The value of corporate culture. Journal of Financial Economics, 117(1), 60-76.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.

Joshi, M., Singh, V., & Chaudhary, S. (2020). Sentiment analysis using pre-trained language models. In Advances in Computational Sciences and Technology (pp. 1269-1275). Springer.

Khan, S. U., & Haq, T. U. (2022). Towards efficient sentiment analysis of short social media texts using DistilBERT and convolutional neural networks. International Journal of Advanced Computer Science and Applications, 13(3), 389-400.

Lane, H., Howard, C., & Hapke, H.M. (2019). Natural language processing in action: Understanding, analyzing, and generating text with Python. USA: Manning Publications. ISBN: 9781617294631

Lee, D., & Seung, H.S. (2000). Algorithms for non-negative matrix factorization. Advances in Neural Information Processing Systems, 13.

Liu, H., Li, J., Chen, C., & Wang, J. (2018). Aspect Extraction and Sentiment Classification for Aspect-Based Sentiment Analysis: A Survey. Information Fusion, 41, 146-167.

Luo, N., Zhou, Y., & Shon, J. (2016). Employee satisfaction and corporate performance: Mining employee reviews on glassdoor.com. Management Science, 62(8), 2208-2235.

Mekala, R.R., Irfan, A., Groen, E.C., Porter, A., & Lindvall, M. (2021, September). Classifying user requirements from online feedback in small dataset environments using deep learning. In 2021 IEEE 29th International requirements engineering conference (RE) (pp. 139-149). IEEE.

Moniz, J. (2015). Unveiling corporate value through employee satisfaction: A textual analysis of Glassdoor reviews. Journal of Corporate Valuation, 8(2), 201-218.

O'Reilly, C.A., Chatman, J., & Caldwell, D.F. (2014). People and organizational culture: A profile comparison approach to assessing person-organization fit. Academy of Management Journal, 34(3), 487-516.

Pontiki, M., et al. (2014). Semeval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014).

Rubera, G., & Kirca, A.H. (2012). Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. Journal of Marketing, 76(3), 130-147.

Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on Twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

Schneider, B., Hanges, P.J., Smith, D.B., & Salvaggio, A.N. (2003). Which comes first: Employee attitudes or organizational financial and market performance? Journal of Applied Psychology, 88(5), 836-851.

Sun, C., Xu, P., Yang, N., Lv, M., & Zhou, M. (2021). RoBERTa pretraining for long document understanding. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 8451-8462). Association for Computational Linguistics. doi:10.18653/v1/2021.emnlp-main.660

Symitsi, E., Stamolampros, P., Daskalakis, G., & Korfiatis, N. (2018). Employee satisfaction and corporate performance in the UK. Socially Responsible Investment eJournal, 13(4), 124-145.

V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Vaswani, A., et al. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems (NeurIPS 2017).

Wang, C.J., Tsai, M.F., Liu, T., & Chang, C.T. (2013, October). Financial sentiment analysis for risk prediction. In Proceedings of the Sixth International Joint Conference on Natural Language Processing (pp. 802-808).

Wang, Y., Huang, M., Zhao, L., & Zhu, X. (2016). Attention-Based LSTM for Aspect-Level Sentiment Classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Wolf, T., Rush, A., Shazeer, N., Chylak, J., Sumers, M., & McClelland, J. (2019). DistilBERT, distilling BERT for efficient inference. arXiv preprint arXiv:1905.04905.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., LeCun, Y., & Prenger, R. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In Advances in Neural Information Processing Systems (pp. 9577-9587).