**Comprehensive Summary of comparing 8 different LLMs with HR data**

**Muhammad Arslan, Dallas Diaz, Muhammad Tayyab, Ramya Thambabattula**

Introduction In this report, we delve into the application of Natural Language Processing (NLP) techniques, specifically utilizing advanced Language Models (LLMs) such as BERT and XLNet, to analyze text data related to resumes, perform sentiment analysis, topic modeling, and evaluate model performance. The aim is to extract meaningful insights from textual data and improve understanding across various NLP tasks.

BERT for Resume Analysis For the resume analysis task, BERT (Bidirectional Encoder Representations from Transformers) was employed as a robust LLM. By processing the data from "processed\_data.csv," BERT provided valuable insights into the skills required for different job roles. It successfully identified the skills necessary for specific positions, aiding in understanding job requirements and skill sets.

A close-up of words

Description automatically generated

XLNet Comparison Additionally, XLNet, another powerful LLM, was explored for comparison purposes. Unlike BERT, XLNet utilizes a permutation language modeling (PLM) objective, considering all possible permutations of words in a sentence for bidirectional context understanding. Despite this difference in approach, the coding structure between BERT and XLNet remained similar, requiring minimal modifications. The results obtained from XLNet were comparable to those of BERT, indicating limited variation in output due to the static nature of the text and the size of the dataset.

A graph of a number of jobs

Description automatically generated

Visual Insights Visualizations played a crucial role in presenting the findings effectively. Graphs and visuals were generated to showcase the skills needed across various positions. A bar graph, for instance, depicted the number of skills required for each role, offering insights for individuals planning their career paths based on skill acquisition needs.

Model Performance Analysis was an in-depth analysis of model performance was conducted using confusion matrices and loss trend graphs. The confusion matrix revealed biases in classification, particularly regarding employee layoffs, where all instances were incorrectly classified into a single class. Strategies such as class balancing and model optimization were suggested to improve performance. Moreover, discrepancies between training and validation loss trends indicated overfitting, highlighting the need for strategies like regularization and adjusting model complexity to enhance generalization.

A blue square with white text

Description automatically generated

Sentiment Analysis Functions To delve into sentiment analysis, functions were defined to extract sentiment using both BERT and RoBERTa models. The sentiment analysis results were applied to the DataFrame, and a visual comparison of sentiment distribution was presented. The conclusion highlighted RoBERTa's superior performance in sentiment analysis due to its simpler categorization system and accurate sentiment interpretation compared to BERT. Upon examining the sentiment analysis results provided by BERT and RoBERTa models, RoBERTa appears to perform better overall. RoBERTa's simpler categorization system of "Negative," "Neutral," and "Positive" aligns more accurately with the general tone of the texts, particularly when handling straightforward and neutral sentiments. BERT, while offering a more nuanced sentiment scale from "1 star" to "5 stars," occasionally misinterprets the sentiment, leading to overestimations or underestimations. Consequently, for broader, generalized sentiment analysis, RoBERTa emerges as the more reliable and accurate choice, especially in scenarios requiring clear categorization of sentiments.

A graph with blue bars

Description automatically generatedA screenshot of a graph

Description automatically generated

LDA and LLM Analysis Topic modeling using Latent Dirichlet Allocation (LDA) was employed on preprocessed layoff stories data, generating topic labels and analyzing word importance. Additionally, LLM analysis was conducted on layoff stories, evaluating performance metrics such as Precision, Recall, F1 score, and ROC curve for GPT-3.5 and GPT-4. The use of encoder-decoder with TensorFlow-Keras and Levenshtein distance for similarity score calculation further enriched the analysis and evaluation metrics. The results shown were precision, recall and an F1 score for the model which all showed to be .80 percent. The model in both situations was able to understand and train well enough to be able to be used in a real life application.

In conclusion, this report demonstrates the comprehensive application of NLP techniques, including LLMs like BERT and XLNet, for diverse tasks such as resume analysis, sentiment analysis, topic modeling, and model performance evaluation. Through visualizations and rigorous analysis, valuable insights were extracted, paving the way for enhanced understanding and decision-making in NLP-related endeavors.Top of Form

A screenshot of a graph

Description automatically generatedBottom of Form