UNIVERSITY
MUHAMMADIYAH MALANG

POLITICAL SENTIMENT ANALYSIS 2023/2024

tenkaslave@gmail.com

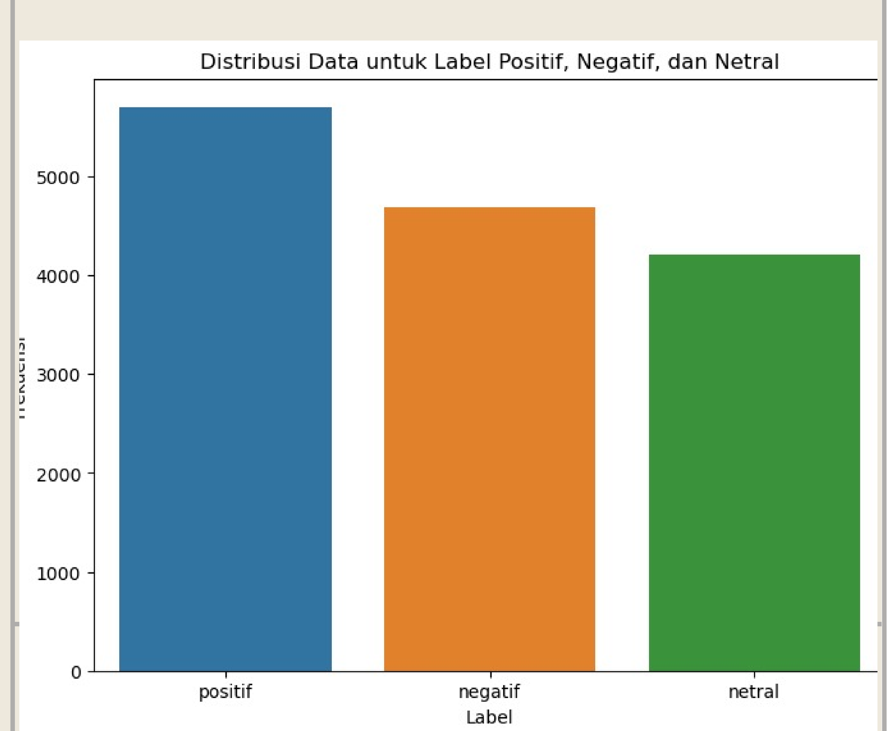
Aldi Oktafianto
Muhammad Teguh Prasetyo
Vicky sandy Laksana Putra

introduction

This research was conducted in the Indonesian political context ahead of the general election, where social media, especially Twitter, has become an important platform for voicing political opinions and views. The main aim of this research is to understand people's perceptions and sentiments regarding political issues through Twitter data analysis. We use keywords that are relevant in the Indonesian political context, such as "Politik", "Pemilu", "DPR", and others, to collect a dataset that reflects current political discussions.

Objective

The target of this research is to assess and compare the effectiveness of each model in analyzing and classifying sentiment in political data. By understanding which models are the most accurate, we hope to provide recommendations for sentiment analysis methods that are most suitable for similar data, especially in the Indonesian political context.



From the distribution table data above, positive values get numbers above 5,000, while negative values reach around 4,500. For the neutral category, there is a score of 4,200. These results indicate a picture of political sentiment data in the dataset, where opinions that tend to be positive have the highest frequency, followed by negative and neutral opinions respectively.

analysis method

*Natural Language Processing

*Classification

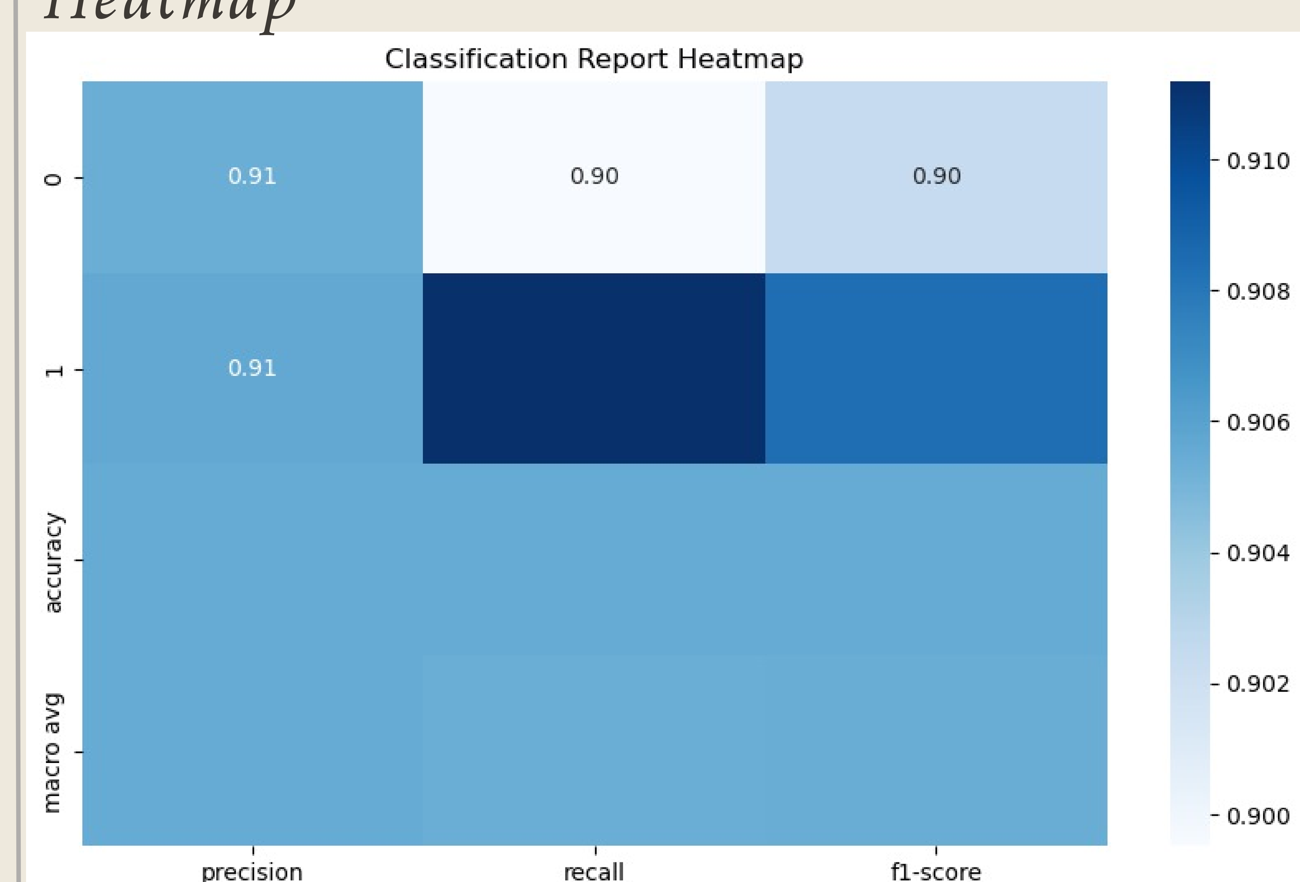
*Big Data Analysis Techniques

The gathered data was analyzed using three different approaches: a classic SVM-based model, a pre-trained IndoBERT model adapted from BERT for the Indonesian language, and a Deep Learning approach utilizing LSTM. These two methods were chosen due to their different text comprehension representations: SVM with a classical statistical approach, IndoBERT with a deep understanding of language, and LSTM with its ability to model data sequentiality. Data labeling was carried out manually with high precision, categorizing tweets into positive, neutral, and negative categories based on their sentiment content.

Datasets

The data analyzed consists of tweets collected using various political keywords. The data distribution based on sentiment labels is as follows: 5,705 tweets were categorized as positive, 4,696 as negative, and 4,214 as neutral. These labels are coded with numbers, where 0 represents neutral, 1 for positive, and 2 for negative. This distribution shows the active involvement of Twitter users in discussions related to politics, with positive sentiment being the most dominant.

Heatmap



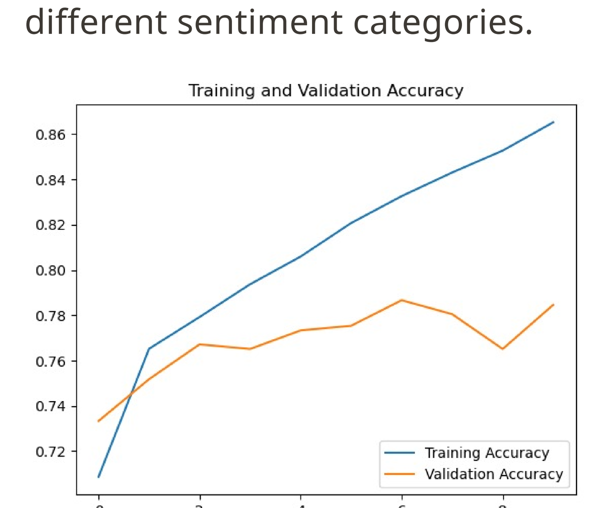
Result

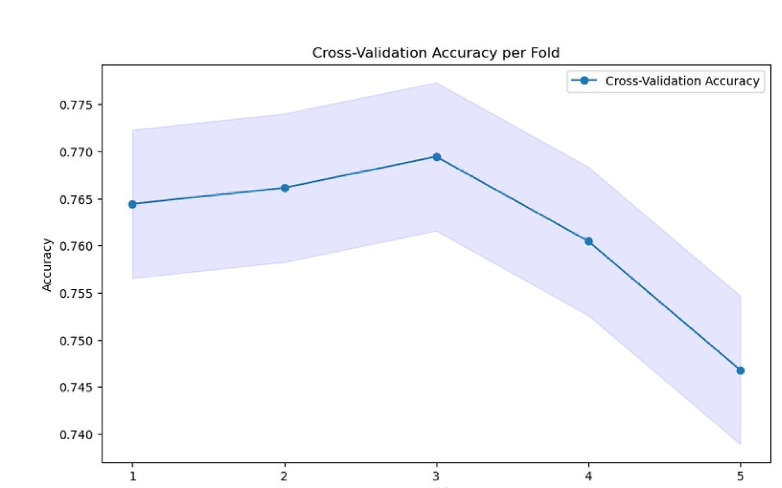
	Precission	Recall	f1-score	Support
Clasic (SVM)	0. 0.72 1. 0.79 2. 0.77 M.avg 0.76 W.avg 0.76	0. 0.72 1. 0.81 2. 0.74 M.avg 0.76 W.avg 0.76	0. 0.72 1. 0.80 2. 0.75 Acc. 0.76 M.avg 0.76 W.avg 0.76	0. 287 1. 376 2. 312 Acc. 975 M.avg 975 W.avg 975
Deeplearning (LSTM)	0. 0.85 1. 0.90 2. 0.86 M.avg 0.87 W.avg 0.87	0. 0.87 1. 0.88 2. 0.86 M.avg 0.87 W.avg 0.87	0. 0.86 1. 0.89 2. 0.86 Acc. 0.87 M.avg 0.87 W.avg 0.87	0. 290 1. 372 2. 313 Acc. 975 M.avg 975 W.avg 975
Pre-Traine (INDOBERT)	0. 0.69 1. 0.76 2. 0.73 M.avg 0.73 W.avg 0.74	0. 0.76 1. 0.73 2. 0.72 M.avg 0.74 W.avg 0.74	0. 0.73 1. 0.76 2. 0.72 Acc. 0.74 M.avg 0.73 W.avg 0.74	0. 282 1. 390 2. 303 Acc. 975 M.avg 975 W.avg 975

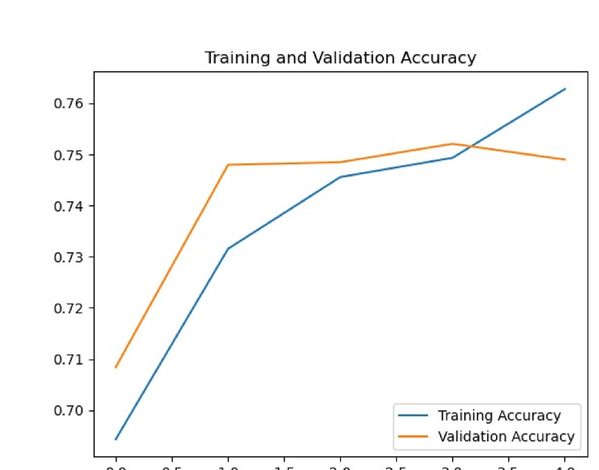
The Deep Learning model with LSTM showed the best performance with 87% accuracy, followed by IndoBERT with 74%, and SVM with 76%. The LSTM model excels at understanding language nuances and context in tweets, which is reflected in its high f1 scores across all sentiment categories. Although IndoBERT offers deep language understanding, this model still lags behind LSTM in this context. SVM, despite having lower accuracy, remains relevant as a comparison method. These results pave the way for further applications in understanding political dynamics through social media, especially in helping political stakeholders to understand and respond to the dynamics of public opinion.

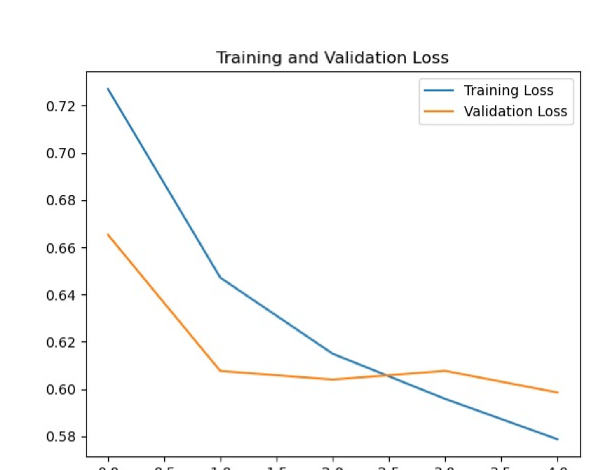
Analysis

Use of machine learning algorithms such as Support Vector Machines (SVM) to train models that can classify text into different sentiment categories.









Conclusion

Political sentiment analysis highlights efforts to analyze people's opinions, attitudes and emotional reactions regarding political issues using data analysis techniques. This approach allows us to unravel and understand hidden patterns in texts from various sources such as social media, news articles, surveys or online discussions.