

ABSTRACT

Tourist attractions encompass everything that possesses uniqueness, beauty, and value from various natural and cultural treasures, making them destinations for tourists. Information technology provides several conveniences, one of which is facilitating tourists in searching for information about a tourist attraction. Social media is a suitable platform for individuals or tourism service providers to share information with each other. The relatively low cost of using social media has led many tourism organizations to adopt it as an effective means of communicating with tourists. One of the services that can be utilized is Google My Business, provided by Google and integrated with Google Maps. This service offers information about a place and also allows visitors to provide comments or reviews.

The comments or reviews from tourists are abundant, making it time-consuming to review them one by one. Therefore, a method is needed to gather comments and analyze the data. One of the methods that can be used is sentiment analysis. Sentiment analysis is the process of automatically understanding and processing text data to extract sentiment information contained in a sentence or text in the form of opinions. This analysis aims to determine the sentiment orientation of a text related to a specific issue or object, whether it is viewed positively or negatively. In this study, multidimensional sentiment analysis is conducted on tourist attractions in the provinces of East Java, Central Java, West Java, Banten, and DKI Jakarta. The dimensions taken are the regional dimensions, namely the provinces, and the time dimension used is in monthly units. Data collection was carried out from January to March 2023 using Data Miner.

Data labeling is done by comparing two labeling methods, namely transformers and textblob. Then, a confusion matrix is used as the evaluation method. The sentiment analysis implementation is carried out using the multi-class classification algorithm Support Vector Machine (SVM) with the One-Against-All approach. Four kernels are tested in this study: Linear, RBF, Sigmoid, and Polynomial. The default parameters are used to determine the kernel, and then the kernel with the highest accuracy is chosen. Each kernel is tested using three

ratios: 70:30, 80:20, and 90:10. The results of these tests will be used to determine the best kernel type for further Grid Search process. The Grid Search process is used to find the best combination of parameters that produce an optimal model performance. The labeling with the highest accuracy is achieved by textblob using a data splitting of 90:10 (90% data for training and 10% data for testing).

To assess the performance of the model and obtain detailed information about the comparison between predicted results and actual classifications, this research utilizes the Confusion Matrix table. This table is useful for analyzing the accuracy and performance of the model in classifying data by comparing four main values: true positive, false positive, true negative, and false negative. The accuracy achieved in this study is 93%, with an average precision of 90%, recall of 87%, and an f1-score of 88%. Subsequently, data visualization will be conducted using the Power BI tool to illustrate the prediction results. Power BI is a resource for making decisions and a tool for creating reports and visualizations. It transforms data from various sources that may be difficult to understand into interactive, shareable, visual, and informative formats. By utilizing Power BI, the data of the prediction results will be displayed in the form of graphs, tables, charts, and more, providing a clear and comprehensive representation of the model's performance.

Keywords— [Sentiment Analysis, Multidimensional Analysis, Support Vector Machine, Transformer, Textblob, and Confusion Matrix]