

ABSTRACT

Abstract - A tourist attraction is a place or location that attracts people to visit and enjoy the place with its natural and cultural beauty. Nowadays information technology is growing and social media is increasingly used by many people to communicate with many people or share their experiences and include in searching for tourist attractions. Now social media is not only used by individuals but tourism organizations to share information. One that allows visitors to comment or review about a tourist attraction on Google My Business.

In Google My Business, the facility is from Google and already connected to Google Maps. This service provides information about a place, provides ratings and provides facilities to comment or comment on places you have visited. Comments or reviews from visitors are overwhelming, so it takes a very long time to comment one by one. Therefore, sentiment analysis is needed to group comments that contain positive, neutral and negative reviews. Sentiment analysis is the process of analyzing judgments, sentiment, views, behaviors related to a particular product, service, organization, object of activity and so on.

Sentiment analysis helps in understanding whether reviews or comments given by visitors regarding a particular topic are positive, neutral and negative. The study used multidimensional sentiment analysis on place and time for tourist attractions in East Java Province, Central Java, West Java, Banten and DKI Jakarta with the time range of January, February, and March 2023 using the K-Nearest Neighbor algorithm (KNN). Data collection is assisted with DataMiner tools, in this research using a library that can perform automatic labeling. The data will be labelled with two Python libraries for the Natural Language Processing (NLP) process, namely transformer labelling and textblob labelling, and compared at the ratio 70:30, 80:20 and 90:10.

Implement sentiment analysis with multi-class classification algorithm K-Nearest Neighbor (KNN). This method performs calculations by searching for the nearest neighbor, when receiving the new data that is to be classified. By searching for

the K value and calculating the Euclidean or Manhattan distance. This method uses GridSearch to combine good parameters and produce accurate and optimal model performance. The most accurate outcomes of this test are transformers with a ratio of 80:20 (80% data training and 20% data testing).

In this study, to be able to evaluate the performance of the model and gain a more detailed understanding of the results of the comparison of classifications carried out by the library (prediction) with the actual classification results using a confusion matrix. The confusion matrix table is useful for analyzing the performance of models in classifying data by comparing 4 values: true positive, false positive, true negative, and false negative.

In this study, the models made had an accuracy of 78% with an average precision of 65%, recall of 49%, and f1-score of 51%. Next, the results of the prediction are visualized using Power BI. Power BI is a tool used to visualize data, such as displaying total data, filtering neutral negative positive divisions, pie charts, bar diagrams, maps, and tables, and then applying them in engaging and interactive visualizations. Data information is displayed in various forms of graphs, charts, tables, cards, multi-row cards, pie cards, stacked bar cards, maps to facilitate understanding, decision making and make an easy-to-understand view.

Keywords - [Sentiment analysis, Multidimensional Sentiment Analysis, Transformers, TextBlob, K-Nearest Neighbor, Confusion Matrix, Power BI]