

ABSTRACT

Researches confirms that social media provides good insights on what people think, feel, concern, etc. It is expected that those insight mined from Twitter data has potential to support a better decision-making, especially in public sectors. Public sector wants to know local's insight level; therefore they need to make sure they use the conversation from residents. However, the ground truth shows that tweets are mixed from the residents and tourist. This study investigates the best automatic fashion model to classify tweets posted by resident and tourist, in NTB. Indonesia. To do so, several consecutive phases were conducted. Those are pre-processing, data training, classification system, data testing, accuracy comparison, and result visualization.

First of all, a Twitter dataset, which has 700,000 tweets posted by approximately 26,000 users in Nusa Tenggara Barat, Indonesia was prepared. The dataset divided into two sets, tweets from 4,000 users for data training and 22,000 users for data testing. Then, three popular classification algorithms were applied to the datasets. There are Multinomial Naïve Bayes, Support Vector Machines and Decision Tree. After that, 7 features are created. There are Bag of Words, Normalizer location, Total Tweet, Total Day, Tweet per Day, Total Location and Location per Day.

Experiment shows that Multinomial Naïve Bayes with Bag of Words feature has 86% accuracy, while the rest of features give less than 65% accuracy. This is different with Support Vector Machines and Decision Tree results. These two algorithms produce better accuracy results excluding Bag of Words feature. It implies that Support Vector Machine and Decision Tree are more powerful when processing numerical value. However, among all classification system, Multinomial Naïve Bayes still being the most accurate algorithm for the model.

Keyword: Classification, MNB, SVM, Decision Tree, NTB, Twitter.