Analisis Sentimen Destinasi Wisata Kuliner di Twitter Menggunakan TF-IDF dan Complement Naïve Bayes pada Dataset Tidak Seimbang

Fakhrana Kurnia Sutrisno[1], Jondri[2], Kemas Muslim Lhaksmana[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung

[1]fakhranakurnia@student.telkomuniversity.ac.id, [2]jondri@telkomuniversity.ac.id, [3]kemasmuslim@telkomuniversity.ac.id

Abstract

Public's opinion on a culinary tourism destination is very helpful for the owners and the tourists. Therefore, sentiment analysis was performed on culinary tourism destinations in Bandung, namely Cuanki Serayu and Sate DJ. The process begins with crawling the data from Twitter and manually labelling it as positive, neutral, and negative. The data that has been labeled are preprocessed and undersampled on the unbalanced data. Dataset is divided into train data and test data with a ratio of 70:30. Data training was performed using the Complement Naïve Bayes method with TF-IDF feature extraction. From the test results, the greatest f1-score value is 0.80 from the data that has been oversampled.

Keywords: sentiment analysis, oversampling, TF-IDF, Complement Naïve Bayes, f1-score