Abstract

Antonius Angga Kurniawan. **92316008**
IMPLEMENTATION OF TEXT CLASSIFICATION ON SENTIMENT ANALYSIS USING SVM AND NAÏVE BAYES IN APACHE SPARK ENVIRONMENT.
Master Thesis, Information System Management, Information System Software, Gunadarma University, 2018.
Keyword : Big Data, Apache Spark, Spark MLlib, Spark SQL, Text Classification, SVM, Naïve Bayes, Sentiment Analysis

(xiv+ 64+ appendix)

Big Data has a large volume and variety of data. Thus, Big Data cannot be processed manually or using traditional tools. The method commonly used in Big Data is text mining or opinion mining. In text mining or opinion mining, text classification techniques are needed. Some technologies that can handle large-scale data processing and text classification are Hadoop, Weka, and Apache Flink. However, these technologies still have weaknesses in data processing, especially in iterative latency and the processing time required is still less fast. Along with the development of technology, a new framework emerged, namely Apache Spark. It is a distributed computing framework where the process runs in memory and suitable for large-scale data processing, machine learning, and text classification.

This research was conducted to implement the text classification on sentiment analysis using Support Vector Machines (SVM) and Naïve Bayes in the Apache Spark environment and to compare the two algorithms. Classification performance is obtained using evaluation metrics measured by accuracy, precision, recall, f-measure, and the roc curve. The object used in this study is sentiment analysis based on user reviews from the Blackberry Messenger (BBM) application. The results obtained state that the text classification process in the Spark environment is relatively fast. The results also show that the SVM algorithm is better than Naive Bayes in terms of accuracy. While the Naïve Bayes algorithm is better in terms of speed.
References (2002-2018)