

Abstrak

Antonius Angga Kurniawan. 92316008

IMPLEMENTATION OF TEXT CLASSIFICATION ON SENTIMENT ANALYSIS USING SVM AND NAÏVE BAYES IN APACHE SPARK ENVIRONMENT.

Tesis Magister, Information System Management, Information System Software, Universitas Gunadarma, 2018.

Kata Kunci : Big Data, Apache Spark, Spark MLlib, Spark SQL, Text Classification, SVM, Naïve Bayes, Sentiment Analysis

(xiv+ 64+ lampiran)

Big Data memiliki volume dan variasi data yang besar sehingga tidak dapat diproses secara tradisional. Metode yang biasa digunakan di Big Data adalah penambangan teks. Dalam penambangan teks, diperlukan teknik klasifikasi teks. Beberapa teknologi yang dapat menangani pemrosesan data skala besar dan klasifikasi teks adalah Hadoop, Weka, dan Apache Flink. Namun teknologi tersebut masih memiliki kelemahan dalam pengolahan datanya terutama pada iterative latency dan waktu pengerjaan yang dibutuhkan masih kurang cepat. Apache Spark adalah kerangka kerja komputasi terdistribusi di mana proses berjalan dalam memori dan cocok untuk pemrosesan data skala besar, pembelajaran mesin, dan klasifikasi teks.

Penelitian ini dilakukan untuk mengimplementasikan klasifikasi teks pada analisis sentimen menggunakan Support Vector Machines (SVM) dan Naïve Bayes di Apache Spark dan membandingkan kedua algoritma tersebut. Performa klasifikasi diperoleh menggunakan metrik evaluasi yang diukur dengan akurasi, presisi, recall, f-measure, dan kurva roc. Objek yang digunakan dalam penelitian ini adalah analisis sentimen berdasarkan review pengguna dari aplikasi Blackberry Messenger (BBM). Hasil yang diperoleh menyatakan bahwa proses klasifikasi teks di lingkungan Spark relatif cepat. Hasil penelitian juga menunjukkan bahwa algoritma SVM lebih baik daripada Naive Bayes dalam hal akurasi. Sedangkan algoritma Naïve Bayes lebih baik dari segi kecepatan.

Referensi (2002-2018)