



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam pengembangan piranti lunak, kebutuhan pengguna merupakan suatu komponen yang sangat dibutuhkan oleh pengembang piranti lunak. Untuk memenuhi kebutuhan pengguna tersebut, dikenal istilah *Requirement Engineering*. *Requirement Engineering* adalah cabang dari *Software Engineering* yang mengurus masalah yang berhubungan dengan tujuan, fungsi, dan batasan-batasan pada sistem *software* (Wahono, 2006). *Requirement Engineering* membantu *Software Engineer* untuk memberikan pengertian yang lebih baik terhadap masalah yang hendak dipecahkan (Waseso, 2008). Salah satu cara dari *Requirement Engineering* adalah dengan menggunakan *user feedback*.

User feedback didefinisikan sebagai seluruh informasi yang didapatkan dari pengguna tentang apakah mereka puas atau tidak dengan produk atau jasa yang digunakan (Ceban, 2018). Pengguna akan memberikan *feedback* apabila merasa puas ataupun terjadi kendala-kendala yang terjadi selama menggunakan *software*. *User feedback* dapat memberikan informasi yang dapat membantu pengembang untuk melakukan perbaikan ataupun pengembangan terhadap *software*. Meskipun *user feedback* memiliki manfaat dalam pengembangan *software*, terkadang terdapat beberapa tantangan bagi pengembang seperti adanya *feedback* yang merupakan *spam*.

Spam adalah aktivitas mengirimkan pesan kepada orang lain dengan menggunakan perangkat elektronik secara terus-menerus dan dengan jumlah yang masif tanpa dikehendaki oleh penerimanya (Maxmonroe, n.d.). Biasanya, *spam* dikirim secara acak dan masif dan menyasar siapapun targetnya. Pesan *spam* selalu memiliki karakteristik yang hampir sama, sehingga membedakannya tidak terlalu sulit.

Pada *user feedback*, *spam* lebih mengarah kepada *inappropriate feedback* (ulasan yang tidak tepat). Contohnya pada aplikasi *online travel agent*, terdapat banyak *feedback* yang bukan merupakan ulasan, seperti kalimat hinaan, ataupun kalimat pertanyaan seperti “Apakah bisa untuk memesan tiket kereta dari kota x ke kota x tidak ya?”. Ulasan yang tidak tepat dianggap *spam* karena merugikan bagi *developer* dari aplikasi terkait yang menggunakan *feedback* untuk melakukan perbaikan atau perkembangan pada aplikasi. Oleh karena itu dilakukan klasifikasi *spam* pada *user feedback* agar dapat membedakan *feedback* spam dan non-spam. Salah satu metode klasifikasi yang dapat digunakan adalah Multinomial Naive Bayes.

Multinomial Naive Bayes merupakan sebuah improvisasi dari algoritma *naive bayes*. *Naive bayes* sendiri merupakan sebuah metode klasifikasi yang menggunakan teorema bayes dimana setiap fitur dianggap independen dari fitur lainnya (Gandhi, 2018). Algoritma tersebut memiliki performa yang bagus dalam *text classification*. Menurut penelitian dari (Matwin & Sazonova, 2012), algoritma Multinomial Naive Bayes tidak hanya memiliki akurasi yang lebih baik dari Support Vector Machine, namun juga MNB berjalan 6x lebih cepat dari algoritma SVM. Pada penelitian yang

dilakukan oleh (Kalokasari, et al., 2017), algoritma *Multinomial Naive Bayes* berhasil mendapatkan akurasi 89,58% pada sistem klasifikasi surat keluar.

Untuk melakukan *text classification*, perlu untuk mengubah bentuk data yang sebelumnya berbentuk teks menjadi bentuk vektor agar bisa dimengerti oleh sistem. Salah satu algoritma yang dapat digunakan adalah TF-IDF. TF-IDF adalah suatu algoritma yang berdasarkan nilai statistik menunjukkan kemunculan suatu kata di dalam dokumen (Wijaya & Santoso, 2016). Menurut penelitian (Arora, 2017), TF-IDF berhasil memiliki performa yang lebih baik dibandingkan dengan algoritma Doc2Vec dengan akurasi sebesar 95% pada Logistic Regression dan 73,62% pada Naive Bayes.

Dalam penelitian ini algoritma Multinomial Naive Bayes dan TF-IDF akan digunakan untuk melakukan klasifikasi teks pada user feedback untuk membedakan antara *spam* dan *non-spam*.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan metode klasifikasi Multinomial Naive Bayes dalam sistem *spam filtering* pada *user feedback*?
2. Bagaimana performa dari metode klasifikasi Multinomial Naive Bayes dalam sistem *spam filtering* pada *user feedback*?

1.3 Batasan Masalah

Dalam penelitian ini, untuk mendapatkan hasil penelitian yang akurat, maka penulis merumuskan batasan masalah berikut:

1. Penelitian ini menggunakan *dataset* berupa *user feedback* pada aplikasi Android.
2. *Dataset* yang digunakan berbahasa Indonesia.
3. *Dataset* yang digunakan merupakan *feedback* dari aplikasi tiket.com.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui bagaimana mengimplementasikan metode klasifikasi Multinomial Naive Bayes untuk melakukan *spam filtering* pada *user feedback* dan mengukur performa dari algoritma Multinomial Naive Bayes pada sistem *spam filtering* yang telah dibuat.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini dapat dijabarkan sebagai berikut.

1. Membantu *developer* dari aplikasi terkait untuk memahami pandangan pengguna terhadap aplikasi tersebut.
2. Penelitian ini juga diharapkan dapat membantu penelitian lainnya terkait *requirement engineering*.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi latar belakang pemilihan judul skripsi “Implementasi Algoritma Multinomial Naive Bayes Untuk Spam Filtering pada User Feedback”, rumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II LANDASAN TEORI

Bab ini berisi dasar-dasar teori yang digunakan dalam penelitian ini. Teori-teori yang digunakan dalam penelitian ini antara lain *tokenizing*, *filtering*, *stemming*, *TF/IDF*, *Multinomial Naive Bayes*, serta *confusion matrix*.

BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini berisi metodologi penelitian yang dilakukan dan perancangan sistem yang terdiri dari *flowchart* dan rancangan antarmuka.

BAB IV IMPLEMENTASI DAN ANALISIS

Bab ini berisi pembahasan mengenai implementasi dari sistem yang telah dibuat serta hasil uji coba dari skenario uji terhadap algoritma *Multinomial Naive Bayes* untuk *spam filtering* pada *user feedback*.

BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian dan uji coba yang telah dilakukan serta saran untuk penelitian selanjutnya berdasarkan sistem yang telah dibuat.