

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Saat ini tersedia banyak sekali aplikasi mobile di *google play store*. Dalam *Google Play* dilengkapi dengan adanya fitur berisi ulasan dari para pengguna yang dapat digunakan untuk melihat *review* dari pengguna aplikasi. Dengan banyaknya aplikasi tersebut membuat pengguna mempunyai banyak pilihan. Untuk mengatasi hal tersebut, langkah pertama adalah mereka harus mengetahui fungsi dari aplikasi yang akan di *install*, bagaimana pendapat konsumen lain terhadap aplikasi tersebut dan memutuskan untuk menginstall atau tidak. Biasanya, ulasan pengguna berisi dua bagian, yaitu nilai *rating* dan komentar/*review* secara tekstual. Nilai *rating* merupakan evaluasi dari pengalaman pengguna menggunakan skala numerik, sedangkan komentar/*review* tekstual mampu bercerita lebih dalam.

Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Penelitian baru-baru ini menemukan hampir 50% dari pengguna internet bergantung pada rekomendasi *word-of-mouth* (opini) sebelum menggunakan suatu produk, karena *review* dari pengguna lain dapat menyediakan informasi terbaru dari produk tersebut berdasarkan perspektif pengguna-pengguna lain yang sudah menggunakan produk tersebut (Fanani, 2017).

Ada lebih dari 30.000 ribu komentar/*review* tekstual dari aplikasi yang ada di *google play store*. Hal ini sangat menantang bagi pengguna untuk membuat keputusan. Selain itu, pengembang aplikasi mengalami kesulitan dalam mencari tahu bagaimana meningkatkan kinerja aplikasi berdasarkan dari ribuan komentar tekstual. Dari permasalahan di atas, dibutuhkan sebuah sistem yang dapat mengklasifikasikan *review* ke dalam tiga kategori yaitu *positive*, *negative* dan *neutral*.

Dengan menggunakan *Machine Learning*, *review* dapat diklasifikasikan secara otomatis. *Machine learning* biasa digunakan untuk klasifikasi dan prediksi. Ada beberapa algoritma dalam *machine learning* yang dapat digunakan dalam proses klasifikasi antara lain *Naïve Bayes*, *KNN*, *J48*, *Random Forest* dan *Support Vector Machine (SVM)*.

Random forest merupakan sebuah *ensemble* (kumpulan) metode pembelajaran menggunakan pohon keputusan sebagai *base classifier* yang dibangun dan dikombinasikan, adapun beberapa aspek penting dalam metode *Random Forest* diantaranya melakukan *bootstrap sampling* untuk membangun pohon prediksi, masing-masing pohon keputusan memprediksi dengan prediktor acak dan *Random Forest* sendiri melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan cara *majority vote* untuk klasifikasi dan juga rata-rata untuk regresi (Sari, 2018).

Ada beberapa penelitian yang telah dilakukan dengan menggunakan algoritma *random forest*. Salah satunya adalah yang dilakukan oleh (Wankhede, 2017). Dalam penelitiannya mereka melakukan klasifikasi pada *review* sebuah Film. Dengan menggunakan dataset yang diperoleh dari *Times Of India* dengan dataset berjumlah 1.000 *review* yang dikelompokkan kedalam *class Positive, Negative* dan *Neutral*. Penelitian ini mengusulkan metode *preprocessing* tokenisasi, *remove stopwords*, dan *POS Tagging*. Kemudian dilakukan transformasi data dengan cara menghitung bobot *term* dari setiap kalimat, di mana *term* yang diperoleh dari kata sifat yang berasal dari proses *POS Tagging*. Selanjutnya dilakukan proses *Feature Extraction* dan *Feature Reduction*.

Penelitian selanjutnya dilakukan klasifikasi *review* produk *smartphone* di Kanada (Hegde, 2017), memanfaatkan metode *ensemble learning Random Forest* yang mengeksekusi 20-500 pohon keputusan. Penelitian ini menggunakan 28 data produk *smartphone* dengan 32 fitur. Dan mengklasifikasikan ke dalam empat kelas sentimen yaitu: *Bad Buy, Okok Buy, Good Buy, dan Best Buy*. Hasil dari penelitian tersebut memperoleh akurasi secara keseluruhan 72%.

Berdasarkan latar belakang tersebut, dalam tugas akhir ini penulis akan menggunakan algoritma *Random Forest* untuk mengklasifikasi *review* yang ada di *google play store*. Diharapkan dapat menghasilkan akurasi yang lebih baik dari penelitian sebelumnya.

1.2 Perumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dalam penelitian ini permasalahan yang akan dirumuskan diantaranya adalah:

1. Bagaimana caranya mengklasifikasikan *review* ke dalam 3 kategori *positive, negative* dan *neutral*.

2. Bagaimana tingkat akurasi yang dihasilkan dengan menggunakan algoritma *Random Forest* pada klasifikasi *sentiment review*.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Dapat mengklasifikasikan *review* ke dalam 3 kategori yaitu *positive*, *negative* dan *neutral*.
2. Mengetahui tingkat akurasi algoritma *Random Forest* dalam mengklasifikasikan *sentiment review* pada beberapa aplikasi di *google play store*.

1.4 Batasan Masalah

Batasan masalah yang ditentukan untuk menghindari perluasan pembahasan dalam penelitian ini adalah sebagai berikut:

1. Data yang akan diklasifikasi adalah data *review* dari beberapa aplikasi yang ada di *Google Play Store*.
2. Dataset diambil dari situs *kaggle.com* yang berjumlah 64.295 data.
3. Dataset yang digunakan adalah 37.427
4. Dataset yang diproses adalah data dalam bahasa *inggris*.
5. Algoritma yang digunakan untuk klasifikasi *review* adalah algoritma *Random Forest*.
6. Klasifikasi dibagi ke dalam 3 kategori yakni *positive*, *neutral* dan *negative*.
7. Sistem ini dibuat menggunakan bahasa *Python*.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini meliputi :

1. Pengumpulan Data
Mengambil, dan mengumpulkan data serta mencari dan mempelajari referensi-referensi yang berkaitan dengan *machine learning*, *Random Forest*, yang didapat dari jurnal, buku, maupun media lain seperti internet.
2. *Preprocessing*
Pada tahap ini data yang telah diambil dari *kaggle.com* diproses menjadi data yang siap diolah menggunakan algoritma *Random Forest* . Proses yang perlu disiapkan diantaranya adalah sebagai berikut :

- a) *Case Folding*, mengubah huruf menjadi huruf kecil.
 - b) *Cleaning*, menghapus *noise* seperti tanda baca yaitu titik(.), koma(,), dan lain-lain.
 - c) *Filtering*, menghapus kata yang tidak memiliki makna.
 - d) *Tokenisasi*, memecah kalimat menjadi kata.
 - e) *Stemming*, mengubah kata-kata menjadi kata dasar.
4. Analisis dan Perancangan
- Pembagian dataset dan merancang *user interface* serta melakukan perancangan model.
5. Implementasi
- Algoritma *Random Forest* diimplementasikan dengan menggunakan bahasa pemrograman *Python*.

1.6 Sistematika Penulisan

Sistematika penulisan yang dipergunakan dalam penulisan tugas akhir ini dapat diuraikan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini akan dibahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan dibahas tentang teori-teori dan konsep yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan masalahnya. Selain itu, bab ini juga memuat teoriteori dalam pelaksanaan pengumpulan dan pengolahan data serta saat melakukan penganalisaan.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini akan dibahas mengenai analisa yang dilakukan terhadap hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Pada bab ini akan membahas tentang implementasi algoritma random forest dalam melakukan klasifikasi review serta memberikan hasil klasifikasi dan akurasi.

BAB VI PENUTUP

Pada bab ini berisi tentang kesimpulan serta saran-saran yang dapat diterapkan dari hasil pengolahan data yang dapat menjadi masukan yang berguna kedepannya.

