

BAB 1 PENDAHULUAN

1.1. Latar Belakang

YouTube merupakan salah satu *platform* yang biasa digunakan untuk memberikan kebebasan dalam menyampaikan pendapat dan membagikan momen ke seluruh dunia dalam bentuk video dan audio. Pada tahun 2020 belakangan ini YouTube mengalami kenaikan jumlah penonton akibat pandemi virus *Covid-19*. Kebijakan *lockdown* di beberapa daerah di Indonesia membuat tidak sedikit orang beralih profesi menjadi *content creator* di YouTube. Selain meningkatnya *content creator* di *platform* YouTube, jumlah penonton pada *platform* YouTube juga meningkat.

Content creator atau yang biasa disebut sebagai *Vlogger* memiliki nama masing-masing untuk setiap kategori. *Beauty Vlogger* merupakan salah satu konten kreator yang membuat konten di bidang kecantikan. *Beauty Vlogger* biasa mengunggah konten seperti tips perawatan wajah, cara *makeup*, serta ulasan mengenai produk-produk kecantikan sedetil mungkin. Video ulasan produk kecantikan pun semakin banyak beredar di *platform* YouTube. Selain semakin banyaknya *beauty vlogger*, jumlah pencarian kata kunci pada kategori *Beauty and Fitness* di YouTube juga meningkat sebanyak 25% dari lima tahun terakhir (*Google Trends, 2021*).

Vlog sebagai bukti perkembangan media digital, turut membuka cara baru untuk menghasilkan *PR coverage* (aktifitas untuk mempromosikan barang pada media *online* maupun *offline*) dan sangat mampu diandalkan untuk memberikan pengaruh (*influence*) serta menarik perhatian dari sejumlah konsumen (Mariezka, Haufar, & Yustikasari, 2018). Keberadaan *Beauty Vlogger* dapat dimanfaatkan sehingga dapat memberikan manfaat dan keuntungan bagi para produsen produk kecantikan. Tanpa perlu memberikan imbalan, produk kecantikan akan diulas secara rinci oleh para *Beauty Vlogger*. Para *beauty vlogger* biasanya akan mengulas produk kecantikan mulai dari harga, kemasan, kekentalan krim, kesan pertama yang dirasakan ketika menggunakan produk, hingga hasil pemakaian produk setelah beberapa hari. Hal ini tentu saja akan menjadi pertimbangan bagi orang-orang yang menonton ulasan tersebut apakah akan membeli dan menggunakan produk kecantikan tersebut atau tidak. Besarnya pengaruh *Beauty Vlogger* pada penjualan produk kecantikan membuat para produsen produk kecantikan patut mengetahui sentimen dari isi video produk kecantikan.

Machine Learning dapat digunakan untuk mengetahui sentimen dari video ulasan produk kecantikan. *Machine Learning* (ML) atau pembelajaran mesin merupakan pendekatan dalam kecerdasan buatan yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi (Bhandary, 2019). *Machine Learning* biasa digunakan untuk melakukan klasifikasi dan prediksi. Pelatihan diperlukan oleh *Machine Learning* untuk memprediksi dan mengklasifikasikan data dengan baik. Klasifikasi adalah metode dalam *Machine Learning* yang digunakan oleh mesin untuk memilah atau mengklasifikasikan objek berdasarkan ciri tertentu sebagaimana manusia mencoba membedakan benda satu dengan yang lainnya (Bhandary, 2019). Prediksi digunakan untuk menerka keluaran dari suatu data masukan berdasarkan data yang sudah dipelajari dalam *training* (Bhandary, 2019). Dengan bantuan *Machine Learning*, data ulasan produk kecantikan dapat diberikan klasifikasi sentimennya secara otomatis. Terdapat enam algoritma *Machine Learning* yang biasa digunakan untuk menganalisis sentimen, algoritma tersebut adalah algoritma *Logistic Regression*, *K-nearest Neighbors*, *Support Vector Machine* (SVM), *Naïve Bayes*, *Decision Tree*, dan *Random Forest Classification*.

Random Forest adalah kombinasi dari masing-masing *tree* yang baik kemudian dikombinasikan ke dalam suatu model (Ahmad, Maqsood, Khan, & Jehad, 2012). *Random Forest Classification* terbentuk dari banyak *Decision Tree* dengan tujuan untuk mengklasifikasikan data. Model akan dilatih dengan cara membentuk beberapa *Decision Tree*. Masing-masing *Decision Tree* akan menghasilkan satu hasil klasifikasi data. Seluruh hasil dari setiap *tree* kemudian akan dihitung dan diambil hasil yang paling banyak. Sistem *voting* tersebut membuat algoritma *Random Forest Classification* memiliki nilai akurasi yang tinggi.

Berdasarkan penelitian analisis sentimen pada isi video yang telah dilakukan sebelumnya oleh Unnathi Bandhary pada tahun 2019, Bandhary mmeneliti tentang pendeteksian ujaran kebencian pada video menggunakan *Machine Learning*. Pada penelitiannya, Bandhary melakukan klasifikasi video ujaran kebencian menggunakan empat buah algoritma, yaitu algoritma Multinomial Naïve Bayes, Linear SVM, Random Forest, dan RNN. Berdasarkan penelitiannya tersebut, *Random Forest Classifier* menghasilkan hasil klasifikasi pada isi video dengan akurasi yang jauh lebih baik dibandingkan dengan algoritma lain.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah di atas, masalah pokok yang akan dibahas adalah sebagai berikut :

- Apakah *Random Forest Classifier* dapat digunakan untuk mengklasifikasikan video ulasan produk kecantikan ke dalam 5 *class*.
- Seberapa baik performa algoritma *Random Forest Classification* pada analisis sentimen video ulasan produk kecantikan.

1.3. Tujuan dan Manfaat Penelitian

Tujuan dan manfaat dari penelitian ini adalah sebagai berikut :

- Mengetahui performa algoritma *Random Forest Classification* dalam mengklasifikasikan video ulasan produk kecantikan ke dalam 5 *class rating*.
- Membantu produsen produk kecantikan dalam mendapatkan *rating* produk dari video ulasan produk kecantikan.

1.4. Ruang Lingkup Penelitian

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih baik, maka penelitian akan dibatasi pada ruang lingkup sebagai berikut :

- Video yang digunakan berdurasi minimal satu menit dan maksimal dua belas menit lima puluh sembilan detik.
- Video yang digunakan adalah video dengan bahasa Indonesia.
- Produk kecantikan yang akan dibahas terbatas pada produk perawatan wajah seperti pembersih wajah, *toner*, *serum*, dan krim wajah.
- Video yang digunakan hanya memuat ulasan satu merek dagang.
- Video yang digunakan hanya memuat satu jenis produk perawatan wajah.
- Bahasa pemrograman yang digunakan adalah *Python*.
- Data akan diklasifikasikan menjadi lima *class rating*, yaitu *class 1* dengan ulasan positif (*rating 5*), *class 2* dengan ulasan agak positif (*rating 4*), *class 3* dengan ulasan netral (*rating 3*), *class 4* dengan ulasan agak negatif (*rating 2*), *class 5* dengan ulasan negatif (*rating 1*).

1.5. Metodologi Penelitian

Metodologi penelitian yang digunakan dalam tugas akhir ini adalah sebagai berikut:

a) Studi Pustaka

Pada tahap ini dilakukan pengumpulan referensi yang diperlukan dalam penelitian. Hal ini dilakukan untuk mengumpulkan informasi yang dibutuhkan dalam penulisan penelitian ini. Referensi dapat berupa buku, jurnal, artikel, atau situs internet yang berhubungan dengan analisis sentimen video ulasan produk kecantikan menggunakan *Random Forest Classifier*.

b) Analisis dan Perancangan

Pada tahap ini hal-hal yang dibutuhkan untuk penelitian akan dianalisis berdasarkan ruang lingkup penelitian, kemudian dirancang sehingga menghasilkan model yang baik.

c) Implementasi

Pada tahap ini, pengumpulan dan pembentukan dataset serta implementasi model dilakukan berdasarkan hasil analisis dan perancangan menggunakan bahasa pemrograman *Python*.

d) Pengujian

Pada tahap ini model yang telah dirancang dan diimplementasikan akan diuji coba apakah model dapat digunakan untuk mengklasifikasikan video ulasan produk kecantikan menggunakan *Random Forest Classifier* dengan baik.

e) Dokumentasi

Pada tahap ini penelitian yang telah dilakukan akan didokumentasikan mulai dari tahap analisis dan perancangan sampai tahap pengujian dalam bentuk laporan tugas akhir.

1.6. Sistematika Penelitian

Secara garis besar penulisan laporan tugas akhir ini terbagi menjadi beberapa bagian. Bagian-bagian laporan tugas akhir ini disusun sehingga menghasilkan susunan laporan seperti berikut :

- BAB 1 : berisi Pendahuluan penelitian yang terdiri dari latar belakang, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, metodologi penelitian yang digunakan, serta sistematika penulisan penelitian.
- BAB 2 : berisi Landasan Teori penelitian yang terdiri dari teori dasar *Natural Language Processing*, *Machine Learning*, *Random Forest Classifier* dan *Confusion Matrix*.
- BAB 3 : berisi Analisis dan Perancangan yang meliputi analisis model, analisis masalah, analisis kebutuhan, dan perancangan model analisis sentimen.

- BAB 4 : berisi Implementasi dan Pengujian yang terdiri dari pengimplementasian lingkungan eksperimen, dataset, model analisis sentimen, pengujian model, serta analisis yang terbagi menjadi dua yaitu analisis studi kasus dan analisis model.
- BAB 5 : mencantumkan tentang Penutup yang berisi simpulan dari pembahasan analisis sentimen pada video ulasan produk kecantikan menggunakan *Random Forest* dan saran untuk pengembangannya.
- Daftar referensi mencantumkan beberapa acuan referensi/pustaka yang digunakan untuk membuat laporan ini.

