

1. Pendahuluan

1.1. Latar Belakang

Dalam berwisata, banyak hal yang dapat dilakukan. Beberapa contohnya adalah wisata bahari, wisata budaya, wisata Pendidikan, wisata alam, dan wisata kuliner. Wisata kuliner adalah wisata dengan objek tujuannya adalah kuliner. Wisata kuliner memiliki tujuan mencicipi atau mencoba kuliner yang terkenal di daerah yang dituju. Salah satu contoh daerahnya adalah Kota Bandung. Kota Bandung memiliki banyak destinasi wisata kuliner, beberapa contohnya adalah Sate DJ dan Cuanki Serayu. Kedua destinasi wisata kuliner tersebut ramai didatangi pengunjung. Tidak sedikit orang yang menyebut kedua destinasi wisata kuliner tersebut pada media sosial Twitter.

Twitter merupakan layanan *micro blog* yang memiliki aspek jejaring sosial. Pengguna Twitter dapat mengirim pesan secara instan yang disebut dengan *tweet* dengan batasan maksimal 140 karakter dalam satu *tweet*. Selain itu, pengguna Twitter dapat melihat *tweet* mengenai topik-topik tertentu [1].

Di Indonesia, Twitter menjadi media sosial ketiga yang paling banyak digunakan [2]. Banyak orang yang memanfaatkan Twitter untuk bercerita mengenai kesehariannya maupun mengeluarkan opini pribadi mereka. Contohnya adalah opini mengenai destinasi wisata kuliner, yaitu Sate DJ dan Cuanki Serayu.

Opini bersifat subjektif dan berisi mengenai sentimen terhadap suatu topik. Analisis sentimen merupakan analisis sentimen yang ada pada sebuah opini pada sebuah topik atau permasalahan [3]. Dari analisis sentimen dapat diketahui apakah opini tersebut bernilai positif, negatif, atau netral. Analisis sentimen sudah banyak dilakukan dalam penelitian, contohnya adalah analisis sentimen terhadap calon gubernur DKI Jakarta 2017 dengan menggunakan metode Naïve Bayes dan Support Vector Machine [4]. Dalam penelitian tersebut, nilai akurasi menggunakan metode klasifikasi Naïve Bayes memiliki nilai tertinggi. Oleh karena itu, pada penelitian ini analisis sentimen dilakukan menggunakan ekstraksi fitur TF-IDF dan *Naïve Bayes Classifier*. Complement Naïve Bayes digunakan karena memiliki hasil yang baik dibandingkan metode Naïve Bayes lainnya seperti pada penelitian yang telah dilakukan A. Anagaw dan Y. L. Chang [5]. Pada penelitian tersebut dilakukan perbandingan metode Complement Naïve Bayes dengan metode Naïve Bayes yang lainnya. Hasil dari penelitian tersebut menunjukkan bahwa Complement Naïve Bayes memiliki hasil yang baik pada data yang seimbang maupun tidak seimbang.

Metode TF-IDF untuk ekstraksi fitur dipilih karena TF-IDF menghitung bobot kata yang muncul pada dokumen. Selain itu, metode TF-IDF mudah dan efisien untuk membantu analisis sentimen. Evaluasi model dilakukan menggunakan Grid Search Cross Validation.

1.2. Topik dan Batasannya

Topik yang dibahas dalam penelitian ini adalah pengaruh *oversampling* dan performa yang dihasilkan dari *oversampling* dataset pada data yang tidak seimbang. Batasan dari penelitian ini adalah dataset diambil dari Twitter dengan kata kunci berupa nama destinasi wisata kuliner terkenal di Kota Bandung, yaitu “sate dj” dan “cuanki serayu”. Pengambilan data Twitter dilakukan menggunakan bahasa pemrograman Python dengan library Tweepy. Dataset yang diperoleh sejumlah 5754 data.

1.3. Tujuan

Tujuan dilakukannya penelitian ini adalah untuk melakukan analisis sentimen opini masyarakat terhadap destinasi wisata kuliner pada media sosial Twitter dengan membandingkan pengaruh dan performa dari skenario-skenario *oversampling* pada dataset yang tidak seimbang menggunakan metode TF-IDF dan Complement Naïve Bayes.

1.4. Organisasi Tulisan

Setelah Bagian 1 yang berisi pendahuluan, jurnal ini dilanjutkan dengan Bagian 2 yang berisi studi terkait jurnal ini. Selanjutnya adalah Bagian 3 yang menjelaskan sistem yang dibangun dalam penelitian ini. Pada Bagian 4 dijelaskan hasil dan evaluasi hasil pengujian dari sistem yang dibangun. Setelah itu, dijelaskan kesimpulan dari penelitian pada Bagian 5.