

## 1. Pendahuluan

### Latar Belakang

Dengan menggunakan fitur *microblogging* Twitter, pengguna dapat memposting tweet singkat dengan karakter terbatas yang mengungkapkan pemikiran dan pendapat mereka terkait suatu hal [1]. Transportasi terbaru di Indonesia, kereta cepat bernama Whoosh menjadi salah satu hal yang ditanggapi pengguna Twitter. Whoosh melakukan perjalanan dari Jakarta ke Bandung dan sebaliknya. Kereta cepat memiliki kemajuan teknologi dibandingkan kereta konvensional yang mampu mempersingkat waktu tempuh, khususnya untuk perjalanan jarak menengah [2]. Namun, transportasi ini menimbulkan banyak pendapat dari masyarakat. Menurut Dhea salah satu warga Bandung, transportasi kereta cepat harus memiliki harga yang terjangkau dan fasilitas yang sepadan dengan yang ditawarkan [3]. Transportasi massal yang terjangkau dan efisien sangat penting bagi kelangsungan hidup masyarakat [4]. Oleh karena itu, untuk memudahkan pemahaman pebisnis atau perusahaan terhadap opini masyarakat mengenai perbaikan pelayanan di masa depan, maka diperlukan analisis sentimen pada media sosial untuk mengetahui opini pengguna terhadap transportasi kereta cepat.

Banyak penelitian yang membahas analisis sentimen menggunakan metode klasifikasi dan berbagai metode ekstraksi fitur. *Word2Vec* dapat menghasilkan nilai *f1-score* yang tinggi pada kategorisasi berita Indonesia menggunakan algoritma *Naïve Bayes* dengan rata-rata nilai *f1-score* sebesar 90% [5]. Pada penelitian analisis sentimen *e-wallet* menggunakan *Naïve Bayes* dan *N-Gram*, penggunaan 2 gram membuat data yang diuji lebih baik dari sebelumnya. Diperoleh akurasi 94,90% dengan algoritma *Naïve Bayes* untuk OVO dan akurasi 94,70% untuk DANA [6]. Pada penelitian analisis sentimen menggunakan *Random Forest* dan *Word2Vec*, hasil akurasi menggunakan *Word2Vec* sebesar 86,87%, lebih tinggi dibandingkan menggunakan *Bag-of-Word* dan *TF-IDF* dengan akurasi sebesar 83,49% dan 84,49% [7].

Penerapan klasifikasi pembelajaran mesin *Naïve Bayes* pada analisis sentimen memperoleh akurasi yang cukup baik dibandingkan *Random Forest* dan *SVM*. Algoritma yang diusulkan memiliki kinerja lebih baik untuk tweet bervolume lebih tinggi [8]. *Naïve Bayes* memberikan hasil akurasi sebesar 80,9% dibandingkan *KNN* 75,58% dan *SVM* 63,88% pada analisis sentimen Calon Presiden Republik Indonesia Tahun 2019 pada Twitter [9].

*Naïve Bayes* memiliki beberapa model dan masing-masing model memiliki karakteristiknya masing-masing. Pengklasifikasi *Multinomial* dan *Bernoulli Naïve Bayes* cocok untuk data diskrit. Bedanya, ketika mengklasifikasikan dokumen uji, *Multinomial Naïve Bayes* bekerja dengan melacak beberapa kejadian, sedangkan *Bernoulli Naïve Bayes* menggunakan informasi kejadian biner, mengabaikan jumlah kejadian [10]. *Complement Naïve Bayes* cocok untuk kumpulan data yang tidak seimbang dan pengklasifikasi ini meningkatkan akurasi klasifikasi dengan menggunakan data dari semua kelas kecuali kelas yang menjadi fokusnya [11]. Ketika berhadapan dengan data kontinu, umumnya diasumsikan bahwa nilai kontinu setiap kelas didistribusikan menggunakan distribusi *gaussian*. Misalnya, atribut kontinu  $x$  ada di set pelatihan. Rata-rata dan varians  $x$  dihitung untuk setiap kelas setelah data dibagi menjadi segmen-segmen berdasarkan kelas [12].

Dari riset-riset sebelumnya, *Naïve Bayes* merupakan salah satu metode klasifikasi yang cukup baik dalam melakukan klasifikasi sentimen di antara metode lainnya dan *Word2Vec* dapat meningkatkan hasil performansi pada algoritma klasifikasi. Dalam penelitian ini akan dilakukan analisis sentimen pengguna kereta cepat pada sosial media Twitter dengan *Word2Vec* dan *Naïve Bayes* sebagai metode klasifikasi. Dalam penelitian ini juga akan dilakukan perbandingan model-model *Naïve Bayes* untuk mengetahui model peluang metode *Naïve Bayes* yang terbaik untuk analisis sentimen pengguna kereta cepat di media sosial Twitter dengan menggunakan ekstraksi fitur *Word2vec*. Secara bersamaan, metode ekstraksi fitur *Word2vec* dipilih karena *Word2Vec* dapat digunakan untuk memperbaiki kinerja model dan meningkatkan akurasi klasifikasi sentimen.

### Topik dan Batasannya

Penelitian ini dilakukan untuk mengklasifikasi sentimen pengguna terhadap transportasi kereta cepat pada sosial media Twitter. Model-model *Naïve Bayes* (*Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, *Gaussian Naïve Bayes*, dan *Complement Naïve Bayes*) serta model arsitektur *Word2Vec* (*CBOW* dan *Skip-Gram*) digunakan dalam penelitian ini sebagai metode. Data yang digunakan dalam penelitian ini berupa data tweet yang berkaitan dengan transportasi kereta cepat. Jenis klasifikasi sentimen dibagi menjadi dua, yaitu sentimen positif dan sentimen negatif.

### Tujuan

Penelitian ini bertujuan untuk mengetahui opini publik mengenai transportasi kereta cepat pada sosial media Twitter. Tujuan selanjutnya yaitu mengetahui model peluang metode *Naïve Bayes* yang terbaik serta model arsitektur *Word2Vec* yang lebih unggul. Perbandingan hasil performansi antara model-model *Naïve Bayes* dan model arsitektur *Word2Vec* juga dilakukan pada penelitian ini.

### Organisasi Tulisan

Penelitian ini diawali dengan pembahasan studi terkait yang berkaitan dengan *machine learning classifier*, analisis sentimen, *Naïve Bayes*, dan *Word2Vec*. Kemudian pembahasan metodologi mengenai sistem yang dibangun pada penelitian ini. Selanjutnya evaluasi dari hasil penelitian dan ditutup dengan pemberian kesimpulan.