

Sentiment Analysis of Comments Using K-Nearest Neighbor on Anteraja Expedition Service

ANALISIS SENTIMEN TERHADAP KOMENTAR DENGAN *K-NEAREST NEIGHBOR* PADA PELAYANAN JASA EKSPEDISI ANTERAJA

I Putu Okky Widiana Putra¹, Gerson Feoh^{2*}, I Made Dwi Ardiada³

^{1,2,3}Prodi Teknik Informatika, Universitas Dhyana Pura, Bali, Indonesia

(*) Corresponding Author: Gerson.feoh@undhirabali.ac.id

Article info

Keywords:

Sentiment analysis,
Anteraja Expedition,
KNN

Abstract

Anteraja expedition services provide a variety of delivery options to meet customer needs, and some customers provide their own opinions to influence the views of others. From the analysis of the Twitter timeline, the keyword "Anteraja expedition" has become a trend, and there are many positive and negative opinions that are interesting to study further. The purpose of this research is to identify the results of sentiment analysis, accuracy, precision, and recall from two data comparisons, namely 60:40 and 70:30, using the K-Nearest Neighbor (KNN) algorithm. The selection of the KNN algorithm is done because KNN has a more accurate ability to handle training data. The results of this classification with the KNN algorithm after testing get an accuracy of 56.76%, negative precision of 66.00%, positive precision of 37.50%, negative recall of 66.00%, and positive recall of 42.86% from the 60:40 comparison while in the 70:30 comparison get an accuracy of 60.71%, negative precision 68.83%, positive precision 42.86%, negative recall 68.83% and positive recall 37.50%. The results showed that there were 66.5% negative opinions and 33.5% positive opinions, this is because it can harm the community in sending packages.

Kata kunci:

Analisis sentimen,
Ekspedisi Anteraja,
KNN

Abstrak

Pelayanan jasa ekspedisi Anteraja menyediakan berbagai pilihan pengiriman untuk memenuhi kebutuhan pelanggan, dan beberapa pelanggan memberikan pendapat mereka sendiri untuk mempengaruhi pandangan orang lain. Dari hasil analisis *timeline Twitter*, terlihat bahwa kata kunci "ekspedisi Anteraja" telah menjadi tren, dan muncul banyak opini positif maupun negatif yang menarik untuk diteliti lebih lanjut. Tujuan dari penelitian ini adalah untuk mengidentifikasi hasil analisis sentimen, *accuracy*, *precision* dan *recall* dari dua perbandingan data, yaitu 60:40 dan 70:30, menggunakan algoritma K-Nearest Neighbor (KNN). Pemilihan algoritma KNN dilakukan karena KNN memiliki kemampuan yang lebih akurat dalam menangani data *training*. Hasil ini klasifikasi dengan algoritma KNN setelah dilakukan pengujian mendapatkan *accuracy* 56.76%, *precision* negatif 66.00%, *precision*

positif 37.50%, *recall* negatif 66.00% dan *recall* positif 42.86% dari perbandingan 60:40 sedangkan pada perbandingan 70:30 mendapatkan *accuracy* 60.71%, *precision* negatif 68.83%, *precision* positif 42.86%, *recall* negatif 68.83% dan *recall* positif 37.50%. Hasil penelitian menunjukan terdapat 66.5% opini negatif dan 33,5% opini positif hal ini dikarenakan dapat merugikan masyarakat dalam melakukan pengiriman paket.

PENDAHULUAN

Teknologi informasi banyak digunakan dalam berbagai bidang, salah satunya adalah jasa ekspedisi dan pelayaran. Barang yang digunakan untuk mengirim barang atau jual beli barang, baik online maupun *offline*. Di zaman dengan digitalisasi, perusahaan jasa pengiriman telah berperan sentral dalam pendistribusian barang ke konsumen. Internet memfasilitasi kegiatan jual beli dan pengiriman barang dalam perdagangan elektronik. Anggota masyarakat dapat memilih dari layanan ekspedisi yang tersedia contohnya AnterAja. Ekspedisi Anteraja adalah perusahaan ekspedisi yang menawarkan layanan pengiriman barang atau paket melalui berbagai jalur pengiriman. Tujuan utama dari perusahaan ini adalah untuk membantu konsumen mengirimkan barang mereka dengan cepat, aman dan efisien. Jasa pengiriman memiliki begitu banyak manfaat sehingga tidak dapat dipisahkan. Ada kekurangan yang dimiliki semua ini pelayaran dan merupakan salah satu faktor yang mempengaruhi. Kerugian dari jasa kurir adalah pendapat pelanggan menjelaskan pendapat mereka sendiri mempengaruhi pandangan orang lain dengan kecenderungan dan preferensi tertentu terhadap jasa ekspedisi (Salam, Zeniarja and Khasanah, 2018). KNN memiliki kelebihan salah satunya adalah kemampuannya untuk menangani data *Training* dengan lebih akurat yang dapat meningkatkan keakuratan hasil klasifikasi.

Twitter merupakan salah satu media sosial yang didirikan oleh Jack Dorsey yang umumnya berfungsi untuk mengirimkan pesan yang disebut dengan kicauan atau *tweet*. Penggunaan Twitter umumnya diakses para pengguna di Indonesia mencapai 59% dan menduduki peringkat ke-5 media sosial yang sering digunakan pada tahun 2020. Dari data yang ada menjadikan *platform* Twitter menjadi media sosial yang cukup memiliki pengaruh bagi para pengguna di Indonesia (Krisdiyanto, 2021). Media sosial Twitter sebagai sumber data, fokus pada cakupan wilayah Indonesia. Metode pengumpulan data yang digunakan adalah *Crawling* menggunakan *Tools* Jupyter Notebook untuk mencari dan mengumpulkan *tweet* dari pengguna. Data yang dikumpulkan dikhususkan pada topik "ekspedisi anterja" sebagai kata kunci pencarian. Setelah data terkumpul, analisis dilakukan menggunakan algoritma K-Nearest Neighbor untuk memahami pendapat publik pengguna di Twitter terkait topik tersebut. Penelitian ini untuk menghasilkan suatu model yang tepat dalam pelayanan ekspedisi anteraja dan kemudian bisa menjadi evaluasi untuk kedepannya.

Dalam analisis sentimen yang penulis buat langkah pertama yang dilakukan adalah pengumpulan dataset yang berasal dari media sosial Twitter, pengumpulan data menggunakan *Tools* Jupyter Notebook untuk melakukan *Crawling* data menggunakan kata kunci "Anteraja" dengan kriteria tanggal dimulai dari tanggal 1 Oktober 2020 hingga 31 Desember 2021, dengan limit 10.000 data, menggunakan bahasa Indonesia, setelah mendapatkan 9886 data *tweet*. Proses dilanjutkan dengan seleksi data dan memberikan label pada data *tweet* berdasarkan emotikon, setelah diberikan label data *tweet* ini akan di *import* ke dalam *RapidMiner* dan masuk pada tahap *Pre-Processing* dalam *RapidMiner*.

Tahap berikutnya merupakan klasifikasi KNN yaitu pembuatan data *training* dan data *testing* dimana data *tweet* akan dibagi menjadi 60% dan 70% untuk data *training* dan 40% dan 30% untuk data *testing* dan kemudian data *training* dan *testing* ini akan di *import* ke dalam *RapidMiner* untuk proses klasifikasi dengan metode KNN yang akan menghasilkan *Prediction* positif dan *Prediction* negatif. Proses selanjutnya pengujian data *training* dan *testing* yang telah dibuat sebelumnya, data akan di *split* dalam proses pengujian ini dimana pengujian pertama menggunakan *split* data 60:40 dengan 60% data *training* dan 40% data *testing* dari data *tweet* keseluruhan kemudian pengujian kedua menggunakan *split* data 70:30 dengan 70% data *training* dan 30% data *testing* dari data *tweet* keseluruhan dan akan keluar hasil *confusion matrix* dimana pada proses ini menghasilkan *accuracy*, *precision* dan *recall*. Sehingga tujuan penelitian sebagai berikut yang pertama untuk mengetahui hasil analisis sentimen terhadap pelayanan ekspedisi anteraja di media sosial Twitter menggunakan algoritma *K-Nearest Neighbor*, yang kedua untuk menghitung tingkat *accuracy*, *precision*, dan *recall* dari perbandingan data 60:40 dan 70:30 dengan menggunakan algoritma KNN terhadap analisis sentimen pelayanan ekspedisi Anteraja berdasarkan opini media sosial Twitter.

METODE

Penelitian ini dimulai dengan mengumpulkan data atau *Crawling* data dari Twitter menggunakan kata kunci “Ekspedisi Anteraja” pada tahap berikutnya, seleksi data yaitu memisahkan data yang akan digunakan dan tidak digunakan kemudian penulis memberikan label berdasarkan emotikon pada dataset dan melanjutkan ke tahapan *Pre-Processing* dan TF-IDF yang bertujuan memberikan bobot pada setiap kata dalam dokumen, pada tahap *Pre-Processing* terdapat lima proses, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Tujuan dari lima tahap tersebut untuk mengubah dokumen mentah menjadi dokumen representatif yang siap untuk diproses, tahap berikutnya klasifikasi yaitu proses mengklasifikasikan data menggunakan algoritma KNN. Pada tahap terakhir, pengujian dilakukan untuk mendapatkan hasil *accuracy*, presisi, dan *recall* dengan menggunakan *confusion matrix*

HASIL DAN PEMBAHASAN

Berdasarkan hasil pengujian pertama menggunakan perbandingan 60:40 dan perbandingan kedua menggunakan perbandingan 70:30 diketahui:

Pengujian dengan data 60:40

Hasil pengujian menggunakan proses *split* dengan 60% data *training*, yang menghasilkan 222 data, dan 40% data *testing*, yang menghasilkan 148 data, dari total 370 data

accuracy: 56.76%

	true Positif	true Negatif	class precision
pred. Positif	18	30	37.50%
pred. Negatif	34	66	66.00%
class recall	34.62%	68.75%	

Gambar 1 Hasil *Confusion matrix* dengan data *testing* 40%

Hasil dari pengujian 70.30 dengan proses split 70% pada data training yaitu berjumlah 259 dan 30% pada data testing berjumlah 111 dari total data 370.

accuracy: 60.71%

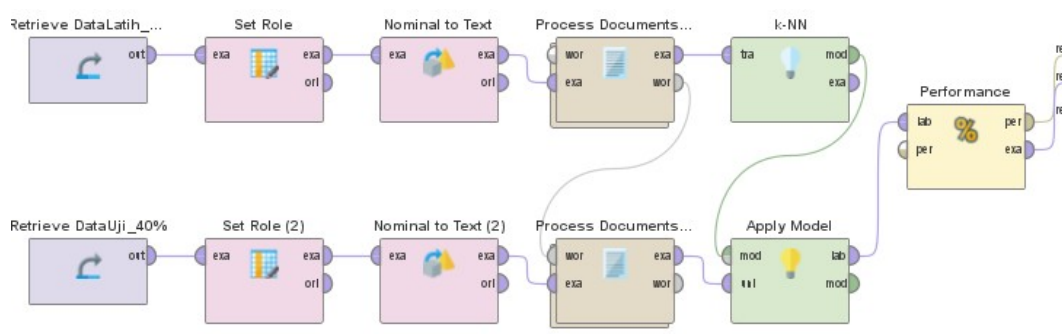
	true Negatif	true Positif	class precision
pred. Negatif	53	24	68.83%
pred. Positif	20	15	42.86%
class recall	72.60%	38.46%	

Gambar 2 Hasil *Confusion matrix* Dengan Data Testing 30%

Pembahasan

Pada penelitian ini pengumpulan data dilakukan dengan cara *Crawling* data dengan kata kunci “Ekspedisi Anteraja”, menggunakan bahasa Indonesia dengan kriteria tanggal dimulai dari tanggal 1 Oktober 2020 hingga 31 Desember 2021, dengan limit 10.000 data, dilanjutkan pada tahapan seleksi data Twitter yang dilakukan sebanyak dua kali yang menghasilkan 370 data *tweet*, 370 data tersebut akan diberi diberikan label secara manual berdasarkan emotikan dan disimpan dalam format file CSV, file ini sudah siap untuk diproses dalam *RapidMiner*. 370 data *tweet* tersebut di *import* ke dalam *RapidMiner* dan dilanjutkan pada tahapan *Pre-Processing*. Pada tahap *Pre-Processing* terbagi atas lima. Yang pertama *cleaning* data *tweet*, guna menghilangkan URL, *mentions*, simbol, RT dan *hashtag*. Yang kedua *case folding*, tahapan *case folding* pada *Pre-Processing* guna menyamaratakan huruf menjadi *lowercase*. Dilanjutkan pada tahap ketiga yaitu *tokenizing*, guna untuk merubah kalimat menjadi kata per kata. Yang keempat *stopword*, guna untuk menghapus kata yang terdapat dalam kamus *stopword*, dalam penelitian ini jumlah kata dalam kamus *stopword* berjumlah 1.704 kata. Lalu tahap yang terakhir adalah *stemming* yang berfungsi untuk mengubah kata berimbuhan menjadi kata dasar, jumlah kata dasar yang digunakan pada penelitian ini berjumlah 1.380 kata.

Klasifikasi *K-Nearest Neighbors* merupakan tahapan selanjutnya, dilakukan pembagian data menjadi *training* dan *testing* dengan perbandingan 60:40 dan perbandingan 70:30 dari jumlah keseluruhan data yaitu 370.



Gambar 3 Tahapan Klasifikasi KNN Pada *RapidMiner*

Tahapan klasifikasi KNN pada data *testing* dengan menggunakan *RapidMiner* dimana proses awalnya dilakukan dengan menarik data *training* dan data *testing* yang sudah di *import* sebelumnya, dihubungkan ke operator *set role*, operator ini berguna untuk menentukan atribut sentimen dan target role sebagai label yang terdapat di dalamnya, selanjutnya dihubungkan ke operator *nominal to text* untuk kedua set data, operator ini digunakan untuk mengubah nilai nominal menjadi teks, berikutnya dihubungkan ke operator *Pre-Processing* dimana pada proses ini terdapat empat tahap yang saling berkaitan yaitu, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* yang sudah dibuat sebelumnya, setelah proses *Pre-Processing* selesai data *training* dihubungkan dengan metode KNN dan tahap selanjutnya KNN dihubungkan dengan operator *apply model* dan *Pre-Processing* pada data *testing* dihubungkan dengan operator *apply model*, yang terakhir *apply model* dihubungkan dengan operator *performance* untuk menampilkan hasil klasifikasi knn dan hasil *accuracy*, *precision* dan *recall* dari data *testing*.

Hasil dari klasifikasi KNN dengan data *testing* 40% yang berjumlah 112 data hasil dari label sentimen positif yang diberi label secara manual berjumlah 50 data sedangkan yang negatif berjumlah 96 data. Data dengan prediksi sentimen KNN menggunakan K2 menghasilkan data positif sebanyak 58 sedangkan data dengan prediksi negatif berjumlah 90. Hasil dari klasifikasi KNN dengan data *testing* 30% yang berjumlah 112 data, Hasil dari label sentimen positif yang diberi label secara manual berjumlah 39 data sedangkan yang negatif berjumlah 73 data. Data dengan prediksi sentimen KNN menggunakan K2 menghasilkan data positif sebanyak 39 sedangkan data dengan prediksi negatif berjumlah 73. Tahapan selanjutnya yaitu pengujian antara Labeling manual dengan hasil prediksi sentimen dari metode KNN dimana pada tahap ini akan menghitung banyaknya data *testing* yang sentimennya benar (*True* positif atau *True* negatif) dengan data *testing* hasil prediksi (*Prediction* positif atau *Prediction* negatif) dari hasil tersebut bisa menghasilkan *accuracy*, *precision*, dan *recall* dan tahap ini akan dilanjutkan pada tahapan pengujian. Pada *RapidMiner* pengujian *confusion matrix* menggunakan operator *performance* untuk menampilkan hasil klasifikasi KNN dan hasil *accuracy*, *precision* dan *recall*.

SIMPULAN

Berdasarkan analisis sentimen yang telah dilakukan dengan topik pelayanan Ekspedisi Anteraja pada media sosial twitter menggunakan algoritma KNN dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan analisis sentimen data komentar pengguna Twitter terhadap layanan jasa ekspedisi Anteraja, menggunakan algoritma KNN, ditemukan bahwa terdapat banyak opini layanan ekspedisi tersebut di media sosial Twitter. Opini-opini ini dapat berdampak merugikan masyarakat dalam melakukan pengiriman paket. Di sisi lain, ada juga beberapa pengguna Twitter yang merasa tidak terbebani dengan pelayanan ekspedisi Anteraja. Hasil klasifikasi menunjukkan bahwa sentimen negatif mendominasi dibandingkan sentimen positif. Secara kuantitatif, terdapat 66,5% opini yang bersifat negatif dan hanya 33,5% opini

yang bersifat positif. Berdasarkan perhitungan tingkat *accuracy*, *precision* dan *recall* pada data 60:40 menghasilkan *accuracy* 56,76%, *precision* negatif 66,00%, *precision* positif 35,50%, *recall* negatif 66,00% dan *recall* positif 37,50% sedangkan pada data 70:30 menghasilkan *accuracy* 60,71%, *precision* negatif 68,83%, *precision* positif 42,86%, *recall* negatif 68,83%, dan *recall* positif 37,50%.

2. Berdasarkan perhitungan tingkat akurasi yang dilakukan, penelitian ini menggunakan algoritma KNN dengan dua perbandingan data *training* dan data *testing*, yaitu 60:40 dan 70:30. Pada perbandingan 60:40, proses *split* 60% pada data *training* menghasilkan 222 data, dan 40% data *testing* menghasilkan 148 data dari total 370 data. Hasil menunjukkan tingkat *accuracy* sebesar 56,76% *accuracy* ini menjelaskan tentang seberapa akurat klasifikasi KNN. Tingkat akurasi mencakup *precision* negatif sebesar 66,00%, *precision* positif sebesar 37,50% *precision* negatif dan positif ini menjelaskan akurasi antar data yang diminta dengan hasil dari prediksi KNN, *recall* negatif sebesar 66,00%, dan *recall* positif sebesar 37,50% *recall* menjelaskan tentang keberhasilan model, Sementara itu pada perbandingan 70:30, proses *split* 70% pada data *training* menghasilkan 259 data, dan 30% pada data *testing* menghasilkan 111 data dari total 370 data. Dalam kondisi ini, tingkat *accuracy* meningkat menjadi 60,71% *accuracy* ini menjelaskan tentang seberapa akurat klasifikasi KNN. Tingkat akurasi mencakup *precision* negatif sebesar 68,83%, *precision* positif sebesar 42,86% *precision* negatif dan positif ini menjelaskan akurasi antar data yang diminta dengan hasil dari prediksi KNN, *recall* negatif sebesar 68,83%, dan *recall* positif sebesar 37,50% *recall* menjelaskan tentang keberhasilan model.

DAFTAR PUSTAKA

- Krisdiyanto, T. (2021) 'Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers', *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 7(1), p. 32. Available at: <https://doi.org/10.24014/coreit.v7i1.12945>.
- Salam, A., Zeniarja, J. and Khasanah, R.S.U. (2018) 'Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)', *Prosiding SINTAK*, pp. 480–486. Available at: <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/6659>.