

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (K-NN) DAN
DECISION TREE PADA ANALISIS SENTIMEN ULASAN PENGGUNA
APLIKASI EXPEDISI JNE DI *GOOGLE PLAY STORE***

SKRIPSI

Diajukan sebagai salah satu syarat untuk mendapatkan
gelar Sarjana Teknik di Program Studi Sistem Telekomunikasi



Oleh
Nova Nurul Putri
1909291

**PROGRAM STUDI SISTEM TELEKOMUNIKASI
KAMPUS UPI DI PURWAKARTA
UNIVERSITAS PENDIDIKAN INDONESIA
2023**

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (K-NN) DAN
DECISION TREE PADA ANALISIS SENTIMEN ULASAN PENGGUNA
APLIKASI EXPEDISI JNE DI *GOOGLE PLAY STORE***

**Oleh
Nova Nurul Putri**

**Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat
memperoleh Sarjana Teknik pada Program Studi telekomunikaasi**

**© Nova Nurul Putri
Universitas Pendidikan Indonesia
Agustus 2023**

**Hak cipta dilindungi undang-undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
Dengan dicetak ulang, difotocopi, atau cara lainnya tanpa izin penulis**

LEMBAR PENGESAHAN

NOVA NURUL PUTRI

**PERBANDINGAN METODE K-NEAREST NEIGHBOR (K-NN) DAN
DECISION TREE PADA ANALISIS SENTIMEN ULASAN PENGGUNA
APLIKASI EXPEDISI JNE DI GOOGLE PLAY STORE**

**Disetujui dan disahkan oleh pembimbing :
Pembimbing I**



**Hafiyyan Putra Pratama, S.ST, M.T.
NIP. 920190219921224101**

Pembimbing II



**Ahmad Fauzi, S.Si., M.T
NIP. 920171219820915101**

**Mengetahui,
Ketua Program Studi Sistem Telekomunikasi**



**Galura Muhammad Suranegara, S.Pd., M.T.
NIP. 929190219920111101**

PERNYATAAN

Saya menyatakan bahwa skripsi yang berjudul “Perbandingan Metode *K-Nearest Neighbor* (K-NN) Dan *Decision Tree* Pada Analisis Sentimen Ulasan Pengguna Aplikasi Expedisi Jne Di *Google Play Store*” ini beserta seluruh isinya adalah benar karya saya sendiri. Tidak ada bagian didalamnya yang merupakan plagiat dari karya orang lain dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung resiko/detikansi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini.

Purwakarta, Agustus 2023

Nova Nurul Putri
NIM 1909291

KATA PENGANTAR

Puji Syukur kehadiran Allah SWT, atas limpahanrahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Performa Virtual Machine pada Platform Amazon Web Services, Google Cloud Platform, dan Microsoft Azure”. Hal ini untuk memenuhi salah satu syarat untuk menyelesaikan Pendidikan S1 Program Studi Sistem Telekomunikasi di Universitas Pendidikan Indonesia Kampus Purwakarta.

Penulis telah berusaha semaksimal mungkin dalam penyusunan skripsi ini, namun kesempurnaan hanya milik Allah SWT, sehingga penulis sadari bahwa karya ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang membangun sangat dibutuhkan oleh penulis.

Purwakarta, Agustus 2023

Nova Nurul Putri
NIM 1909291

ABSTRAK

Perkembangan teknologi internet telah mengubah lanskap bisnis dengan mendorong pertumbuhan *e-commerce*. *E-commerce*, sebagai *platform* penjualan produk dan layanan secara online, sangat tergantung pada industri logistik, khususnya layanan ekspedisi. Di Indonesia, layanan ekspedisi seperti JNE, J&T, dan Pos Indonesia memainkan peran penting dalam pengiriman produk dari penjual ke pelanggan. Namun, keberhasilan layanan ekspedisi juga tergantung pada kualitas dan kecepatan pengiriman, yang tercermin dalam ulasan pelanggan di *platform* seperti *Google Play*. Penelitian ini menganalisis ulasan pelanggan terkait layanan ekspedisi JNE menggunakan metode klasifikasi seperti *K-Nearest Neighbors* (KNN) dan *Decision Tree*. Data latih dan data uji dipisahkan untuk evaluasi model, dan performa diukur dengan akurasi, presisi, *recall*, dan *F1-Score*. Pada hasil evaluasi menggunakan KNN, ditemukan bahwa akurasi model berada pada kisaran 87,2% hingga 87,6% dengan metrik jarak *Minkowski*, *Euclidean*, dan *Manhattan*. Pada hasil evaluasi menggunakan *Decision Tree*, akurasi model berada pada kisaran 90,4% hingga 90,7% dengan kriteria *Gini*, *Entropy*, dan *log_loss*. Selain itu, visualisasi data dilakukan untuk mengetahui kata-kata yang paling sering muncul dalam ulasan positif dan negatif. Kata-kata seperti "lambat," "buruk," dan "pengiriman" sering muncul dalam ulasan negatif. Dari hasil perbandingan, ditemukan bahwa hasil akurasi pada model *Decision Tree* cenderung memberikan performa yang lebih baik dibanding KNN dalam mengklasifikasikan ulasan pengguna terkait layanan ekspedisi JNE. Secara keseluruhan, penelitian ini memberikan wawasan tentang bagaimana metode klasifikasi dapat digunakan untuk menganalisis ulasan pelanggan dan memahami sentimen terkait layanan ekspedisi.

Kata Kunci : JNE, Analisis Sentimen, *K-Nearest Neighbor* (K-NN), *Decision Tree*

ABSTRACT

The development of internet technology has changed the business landscape by driving the growth of e-commerce. E-commerce, as a platform for selling products and services online, is highly dependent on the logistics industry, especially shipping services. In Indonesia, forwarding services such as JNE, J&T and Pos Indonesia play an important role in delivering products from sellers to customers. However, the success of delivery services also depends on the quality and speed of delivery, which is reflected in customer reviews on platforms such as Google Play. This study analyzes customer reviews regarding JNE courier services using classification methods such as K-Nearest Neighbors (KNN) and Decision Tree. Training data and test data are separated for model evaluation, and performance is measured by accuracy, precision, recall, and F1-Score. In the evaluation results using KNN, it was found that the accuracy of the model was in the range of 87.2% to 87.6% with the Minkowski, Euclidean, and Manhattan distance metrics. In the evaluation results using the Decision Tree, the model accuracy is in the range of 90.4% to 90.7% with the Gini, Entropy, and log_loss criteria. In addition, data visualization was carried out to find out the words that appear most often in positive and negative reviews. Words like "slow," "poor," and "delivery" frequently appear in negative reviews, while words like "good," "service," and "convenient" appear in positive reviews. From the results of the comparison, it was found that the Decision Tree model tends to provide better performance than KNN in classifying user reviews regarding JNE expedition services. Overall, this study provides insight into how classification methods can be used to analyze customer reviews and understand sentiment regarding courier services.

Keyword : JNE, Sentiment Analysis, K-Nearest Neighbor (KNN), Decision Tree

UCAPAN TERIMAKASIH

Alhamdulillahirobbil'alamin, Puji Syukur penulis panjatkan atas kehadiran Allah SWT karena dengan Rahmat untuk mencapai jenjang tugas akhir ini dapat diselesaikan. Tidak lupa sholawat serta salam kepada nabi besar Muhammad SAW.

Adapun tujuan utama penulisan skripsi ini adalah memenuhi salah satu syarat dalam menempuh Sidang Skripsi Teknik Sistem Telekomunikasi Universitas Pendidikan Indonesia. Dalam penelitian skripsi ini, penulis mendapatkan ilmu, dukungan dan saran dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih kepada :

- 1 Bapak Prof. Turmudi, M.Ed., M.Sc., Ph.D selaku Direktur Universitas Pendidikan Indonesia Kampus Daerah Purwakartayang telah memberikan kesempatan penulis untuk dapat mempelajari
- 2 Bapak Galura Muhammad Suranegara, M.T. selaku Kepala Prodi Sistem Telekomunikasi Universitas Pendidikan Indonesia yang telah membantu penulis dalam proses bimbingan perkuliahan.
- 3 Bapak Hafiyyan Putra Pratama, S.ST., M.T selaku dosen pembimbing I dan Bapak Ahmad Fauzi, S.Si.,MT. selaku dosen pembimbing II Tugas Akhir program studi Sistem Telekomunikasi.
- 4 Ibu Endah Setyowati, M.T. selaku dosen pebimbing akademik yang telah membantu saya selama masa perkuliahan.
- 5 Bapak Ahmad Fauzi, S.Si.,MT. selaku dosen yang pernah menjabat sebagai Kepala Prodi Sistem Telekomunikasi terimakasih sudah membimbing penulis selama kurang lebih 3 tahun pada masa perkuliahan.
- 6 Bapak Lili Sujana dan Ibu Tuti Kartuti selaku orang tua penulis serta kaka Novan Akbar Putra dan Risa Sekar Rahmatia yang selalu memberikan kasih sayang dan support kepada penulis hingga saat ini.
- 7 Coach Angel Metanosa Afinda, S.Kom, yang senantiasa berbagi ilmu dan menyempatkan waktu untuk membantu penulis ini.
- 8 Teman-teman seperjuangan Sistem Telekomunikasi Universitas Pendidikan Indonesia khususnya angkatan 2019

DAFTAR ISI

LEMBAR PENGESAHAN	III
PERNYATAAN	IV
KATA PENGANTAR	V
ABSTRAK	VI
ABSTRACT	VII
UCAPAN TERIMAKASIH	VIII
DAFTAR ISI	IX
DAFTAR GAMBAR	XI
DAFTAR TABEL	XII
DAFTAR LAMPIRAN	XIII
BAB I	1
Pendahuluan	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan penelitian	5
1.5 Manfaat Penelitian	6
1. Manfaat teoritis	6
2. Manfaat praktis	6
BAB II	7
Kajian Teori	7
2.1 Jasa Ekspedisi JNE (Jalur Nugraha Ekakurir)	7
2.2 Google Play	8
2.3 Analisis Sentimen	9
2.3.1 Klasifikasi Text	9
2.3.2. Algoritma <i>K-Nearest Neighbor</i>	10
2.3.3. Algoritma <i>Decision Tree Classifier C4.5</i>	14
2.3.4 <i>Web Scraping</i>	16
2.3.5 <i>Confusion Matrix</i>	16
2.4 Text Mining	18
2.5 <i>Text Preprocessing</i>	19
2.6 TF-IDF	20
2.7 <i>Machine Learning</i>	21
2.8 <i>Google Colaboratory</i>	23
2.9 Penelitian Relevan	25
BAB III	27
METODE PENELITIAN	27
3.1 Jenis Penelitian	27
3.2 Populasi dan Sampel	27
3.3 Variabel dan Definisi Operasional	27
3.6 Teknik Pengumpulan Data	28
3.7 Teknik Analisis Data	29
3.8 Pembagian Data Uji dan Data Latih	29
3.9 Analisis Kebutuhan	30
3.11 Alur Penelitian	32
BAB IV	35

Temuan dan Pembahasan	35
4.1. Pengambilan Data	35
4.2. Pengolahan Data	37
4.2.1. Analisis Deskriptif	37
4.2.2 <i>Preprocessing Data</i>	38
4.2.3 Memberi Label Pada Kelas Sentimen	40
4.2.4. Pembagian Data Latih dan Data Uji	41
4.3. Klasifikasi menggunakan metode <i>K-nearest Neighbor</i>	42
4.4. Klasifikasi menggunakan metode Decision Tree	45
4.5. Hasil penggabungan performa metode K-NN dan Decision Tree	46
4.6. Visualisasi Data <i>Wordcloud</i>	47
4.6.1 Ulasan Negatif	48
BAB V	49
SIMPULAN, IMPLIKASI, DAN REKOMDASI	49
5.1 Simpulan	49
5.2 Implikasi	50
5.3 Rekomendasi	50
DAFTAR PUSTAKA	52
LAMPIRAN	56
Riwayat Hidup	62

DAFTAR GAMBAR

Gambar 1.1 Rating PosAja! di <i>Google Play</i>	1
Gambar 1.2 Rating <i>J&T Exprees</i> di <i>Google Play</i>	2
Gambar 1.3 Rating <i>My JNE</i> di <i>Google Play</i>	3
Gambar 1.4 Top Brand jasa kurir	3
Gambar 3.1 <i>Flowchart</i> pengumpulan data	28
Gambar 3.2 <i>Flowchart</i> metode K-NN.....	30
Gambar 3.3 <i>Flowchart</i> metode <i>decision tree</i>	31
Gambar 3.4 Alur penelitian	32
Gambar 4.1 proses <i>scraping</i> data melalui <i>Gcolab</i>	35
Gambar 4.3 Data ulasan sebelum diedit	36
Gambar 4.4 Data ulasan sesudah diedit.....	36
Gambar 4.5 Grafik ulasan tahun 2021-2022	37
Gambar 4.6 Jumlah persentase bintang 1-5	37
Gambar 4.7 Jumlah persentase kelas	41
Gambar 4.8 digram akurasi KNN dengan pengujian 80%:20%.....	43
Gambar 4.9 diagram akurasi KNN dengan pengujian 70%:30%.....	43
Gambar 4.10 hasil perbandingan performa 1	46
Gambar 4.11 hasil perbandingan 2	47
Gambar 4.12 <i>Wordcloud</i> visualilasi negatif.....	48

DAFTAR TABEL

Tabel 2.1 confusion Matrix	17
Tabel 4.1 sebelum dan sesudah <i>Cleaning Data</i>	38
Tabel 4.2 sebelum dan sesudah proses <i>Tokenizing</i>	39
Tabel 4.3 <i>Filtering</i>	40
Tabel 4.4 Pelabelan kelas	41
Tabel 4.5 Pembagian data uji dan data latih 1	42
Tabel 4.6 Pembagian data uji dan data latih 2	42
Tabel 4.7 Hasil <i>splitting</i> dengan beberapa skema KNN	44
Tabel 4.8 Hasil <i>splitting</i> dengan beberapa skema <i>Decision Tree</i>	45

DAFTAR LAMPIRAN

Lampiran 1	Pengambilan Data menggunakan <i>Google Colab</i>	56
Lampiran 2	<i>Data Preprocessing</i>	57
Lampiran 3	Split Data dan TF-IDF	59
Lampiran 4	Klasifikasi Metode K-NN	60
Lampiran 5	K-NN dengan <i>Confusion Matrix</i>	61

DAFTAR PUSTAKA

- Anonim. (2015). *'Profil Perusahaan Sejarah JNE'*. [Online]. Diakses dari web : <https://www.jne.co.id/id/perusahaan/profil-perusahaan/sejarah-milestone>
- Anonim. (2020). *'Top Brand Index. Retrieved from'*. [Online]. Diakses dari web : https://www.topbrandaward.com/top-brand-index/?tbi_find=jne.
- Bilal, M. dkk. (2016). *'Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN Classification Techniques'*. *Journal of King Saud University - Computer and Information Sciences*, 28(3), 330–344. Doi : <http://dx.doi.org/10.1016/j.jksuci.2015.11.003>
- Chrisbiyanto, A. (2018). *Perusahaan Ekspedisi Berperan Penting Dorong Pertumbuhan UKM*. *SINDONEWS*. [Online]. Diakses dari web : <https://ekbis.sindonews.com/read/1317290/34/perusahaan-ekspedisi-berperanpenting-dorong-pertumbuhan-ukm-1530201209>
- Devita, R. N. dkk (2018). *'Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia'*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(4), 427. Doi : <https://doi.org/10.25126/jtiik.201854773>
- DqLab. (2022). *'Pahami Metode Decision Tree Sebagai Algoritma Data Science'* [Online]. Diakses dari web : <https://dqlab.id/pahami-metode-decision-tree-sebagai-algoritma-data-science>
- Edi, I. & Eka, P.C. (2018). *'Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji'*. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 3, 1–13. Doi : <https://doi.org/10.30645/jurasik.v3i0.60>
- Eska, J. (2018). *'Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5'*. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 2(2), 9–13. Doi : <https://doi.org/10.31227/osf.io/x6svc>

- Fanani. (2017). *Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*. (Skripsi). Sekolah Pascasarjana, UGM Yogyakarta.
- Febriyanti, A. (2020). '*Analisis Sentimen Persepsi Pengguna JNE Menggunakan Algoritma Naïve Bayes Classifier*'. (Skripsi). Sekolah Pascasarjana, Universitas Islam Indonesia.
- Feldman, R., & Sanger, J. (2007). '*The text mining handbook : Advanced approaches in analyzing unstructured data*'. Cambridge, England : Cambridge University Press.
- Giovani, A. P. dkk. (2020). '*Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi*'. *Jurnal Teknoinfo*, 14(2), 115. Doi : <https://doi.org/10.33365/jti.v14i2.679>
- Gorunescu, F. (2011). '*Data Mining*'. Springer Berlin Heidelberg : University of Craiova
- Guo, G. dkk. (2001). '*Using K-NN Model-based Approach for Automatic Text Categorization*'. Northern Ireland : School of Computing and Mathematics
- Indriati, I., & Ridok, A. (2016). '*Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (Nwknn)*'. *Journal of Environmental Engineering and Sustainable Technology*, 3(1), 23–32. Doi : <https://doi.org/10.21776/ub.jeest.2016.003.01.4>
- Jumeilah, Fithri Selva. '*Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes*'. *Jurnal Sistem Informasi Bisnis* 8, no. 1 (30 April 2018): 92–98. Doi : 10.21456/vol8iss1pp92-98. (ISSN : 2502-2377, 2088-3587)
- Kale, V. (2020). '*Enterprise Performance Intelligence and Decision Patterns*'. *Jurnal Ilmiah Informatika (JIF)*, 8(1). Doi : <https://doi.org/10.4324/9781351228428>

- Khustul F, K. (2023). 'Mengulik Sejarah Pos Indonesia, dari Batavia hingga Seluruh Indonesia'. [Online]. Diakses melalui web : <https://www.goodnewsfromindonesia.id/2023/01/26/sejarah-pos-indonesia>
- Kuncahyo, S. N. (2019). 'Confusion Matrix Untuk Evaluasi Model Pada Supervised Learning'. [Online]. Diakses dari : <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>
- Lidya, S. K. (2014). 'Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) Dan K-Nearest Neighbor (K-NN)'. (Tesis). Sekolah Pascasarjana, Universitas Sumatera Utara, Medan
- Limbong, J. J. A. dkk. (2022). 'Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor'. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) 9(2), 347–56. Doi : <https://doi.org/10.25126>.
- Mohri. (2012). 'Foundations of Machine Learning'. Cambridge: MIT Press
- Musthofa Galih Pradana. (2020). 'Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining'. Jurnal Ilmiah Informatika (JIF), 08(01), 39–43. Doi : <https://doi.org/10.4324/9781351228428>
- Nasrullah, A. H. (2021). 'Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris. Jurnal Ilmiah Ilmu Komputer'. 7(2), 45–51. Doi : <https://doi.org/10.35329/jiik.v7i2.203>
- Nishom, M. (2019). 'Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square'. Jurnal Informatika: Jurnal Pengembangan IT, 4(1), 20–24. Doi : <https://doi.org/10.30591/jpit.v4i1.1253>
- Nitami, M. T., & Februariyanti, H. (2022). 'Analisis Sentimen Ulasan Ekspedisi J&T Express Menggunakan Algoritma Naïve Bayes'. Jurnal Manajemen

Informatika & Sistem Informasi, 5(1), 1–29. Doi :
<https://doi.org/10.36595/misi.v5i1>

- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2, 1-135. DOI: 10.1561/15000000011
- Pamungkas, F. S., & Kharisudin, I. (2021). '*Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter*'. PRISMA, *Prosiding Seminar Nasional Matematika*, 4, 628–634.
- Permatasari, P. A., Linawati, L., & Jasa, L. (2021). '*Survei Tentang Analisis Sentimen Pada Media Sosial*'. *Majalah Ilmiah Teknologi Elektro*, 20(2), 177–186. Doi : <https://doi.org/10.24843/MITE.2021.v20i02.P01>
- Putri, R. E., & Rahmawati, R. (2014). '*Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012*'. *JURNAL GAUSSIAN*, 3(4), 831–838.
- Rivki, M., & Bachtiar, A. M. (2017). '*Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia*'. *Jurnal Sistem Informasi*, 13(1), 31-37. Doi : <https://doi.org/10.21609/jsi.v13i1.500>
- Ramadhani, T. G. (2015). *Analisis Sentimen Menggunakan Metode Naive Bayes Classifier dengan Model Dokumen Bernoulli dan Support Vector Machine*. (Skripsi). Sekolah Pascasarjana, Universitas Gadjah Mada.
- Ramadhani, I. (2020). '*Klasifikasi Keluhan Terhadap Jasa Ekspedisi Barang Jne Menggunakan K-Nearest Neighbor*' (Skripsi). Sekolah Pascasarjana, Universitas Islam Negeri Sultan Syarif Kasim Baru.
- Cyber, R. (2020). '*Menilik Sejarah Google Play dari Awal Rilis*'. [Online]. Diakses dari web : <https://www.rahmancyber.net/2020/03/apa-itu-google-play.html>