

Prediction of Shipment Delivery Delays in Courier Services Using the K-Nearest Neighbor (K-NN) Algorithm

[Prediksi Keterlambatan Pengiriman Barang Jasa Ekspedisi Menggunakan Algoritma K-Nearest Neighbor (K-NN)]

Mochamad Bagus Ichwantoro¹⁾, Indah Apriliana Sari Wulandari ^{*,2)}

^{1,2)}Program Studi Teknik Industri, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: indahapriliana@umsida.ac.id

Abstract. *Lion Parcel is a logistics service company in Indonesia operating under the Lion Air Group, providing both domestic and international parcel delivery services through several shipping products, namely BOSSPACK, REGPACK, JAGOPACK, and BIGPACK. Delivery delays at Lion Parcel occur due to the wide variety of shipping products offered, the extensive delivery coverage area, and the absence of a data-driven delay prediction system. This study aims to determine the optimal value of k to be used in predicting delivery delays. The research employs the K-Nearest Neighbor (K-NN) algorithm as the prediction method. The results indicate that the optimal value of k for predicting delivery delays is $k = 5$, achieving an accuracy rate of 95.78%. Furthermore, the model demonstrated excellent predictive performance, with a confusion matrix accuracy score of 1.00.*

Keywords - *K-Nearest Neighbor (K-NN); delivery delay prediction; cross-validation; logistics*

Abstrak. *Lion Parcel merupakan perusahaan jasa logistik di Indonesia yang berada di bawah naungan Lion Air Group, yang menyediakan layanan pengiriman paket baik domestik maupun internasional dengan produk pengiriman BOSSPACK, REGPACK, JAGOPACK, dan BIGPACK. Keterlambatan pengiriman yang terjadi di Lion Parcel terjadi karena banyaknya produk yang ditawarkan dan jangkauan pengiriman yang sangat luas serta belum ada sistem prediksi keterlambatan berbasis data. Penelitian ini bertujuan untuk menentukan nilai k terbaik yang akan digunakan untuk melakukan prediksi keterlambatan pengiriman. Metode penelitian ini menggunakan K-Nearest Neighbor (K-NN). Hasil penelitian ini didapatkan nilai k terbaik untuk melakukan prediksi adalah $k = 5$ dengan nilai akurasi 95,78%. Model juga mampu melakukan prediksi secara tepat dengan hasil akurasi confusion matrix sama dengan 1.00.*

Kata Kunci – *K-Nearest Neighbor; prediksiketerlambatan pengiriman; cross-validation; logistik*

I. PENDAHULUAN

Ketepatan waktu pengiriman merupakan salah satu faktor utama yang menentukan kepercayaan dan kepuasan pelanggan terhadap jasa ekspedisi[1]. Dalam praktiknya, keterlambatan pengiriman masih sering terjadi karena berbagai faktor seperti jarak pengiriman, cuaca, volume paket, dan kendala operasional di lapangan. Permasalahan ini menyebabkan perusahaan ekspedisi kesulitan menjaga tingkat pelayanan yang konsisten, terutama dalam menghadapi lonjakan permintaan akibat pertumbuhan *e-commerce* [2].

Lion Parcel merupakan perusahaan jasa logistik di Indonesia yang berada di bawah naungan *Lion Air Group*, yang menyediakan layanan pengiriman paket baik domestik maupun internasional. Perusahaan ini didukung oleh jaringan distribusi yang luas serta pemanfaatan armada udara *Lion Air Group* guna menunjang efisiensi dan kecepatan pengiriman antarwilayah. Dalam operasionalnya, Lion Parcel menawarkan berbagai jenis layanan dengan karakteristik yang berbeda, seperti *BOSSPACK* (layanan ekspres), *REGPACK* (layanan reguler), *JAGOPACK* (layanan ekonomis), dan *BIGPACK* (pengiriman barang berukuran besar).

Perbedaan karakteristik antar layanan tersebut menyebabkan variasi dalam estimasi waktu pengiriman, cakupan distribusi, serta tingkat risiko keterlambatan pada masing-masing jenis layanan.

Lion Parcel Mangundiprojo 2 Buduran merupakan perusahaan layanan pengiriman paket dari Lion Parcel yang berlokasi di Mangundiprojo, Buduran, Sidoarjo. Lion Parcel menawarkan beberapa layanan pengiriman, yaitu *BOSSPACK* untuk pengiriman cepat, *REGPACK* untuk pengiriman reguler, *JAGOPACK* untuk pengiriman regular dengan harga ekonomis, dan *BIGPACK* untuk pengiriman paket besar minimal 10kg.

Tabel 1. Estimasi Pengiriman Produk Layanan

Produk Layanan	Estimasi Pengiriman
<i>BIGPACK</i>	1-2 hari
<i>BOSSPACK</i>	2-3 hari
<i>JAGOPACK</i>	6-9 hari

REGPACK

3-7 hari

Estimasi pengiriman pada Tabel 1. tidak dihitung pada hari Minggu dan Hari Libur Nasional. Banyaknya layanan yang ditawarkan membuat masih adanya risiko keterlambatan pengiriman. Hal tersebut juga dipengaruhi oleh Lion Parcel yang menyediakan pengiriman lintas pulau dan seluruh Indonesia.

Selama periode pengiriman tahun 2025 mulai dari bulan Januari hingga Desember, contoh sample data pengiriman diambil pada bulan Mei 2025 yang disajikan pada tabel 2 berikut.

Tabel 2. Histori Pengiriman Periode Januari-Desember 2025

No	Tanggal Booking	Alamat Pengirim	Kota Tujuan	Nama Produk	Tanggal Terima	Lama Pengiriman (Hari)	Status Keterlambatan
1	02/01/2025	POS LION PARCEL MANGUNDIPROJO DUA	SUB-SURABAYA	BOSSPACK	03/01/2025	1	On Time
2	02/01/2025	POS LION PARCEL MANGUNDIPROJO DUA	ENE-ENDE	REGPACK	10/01/2025	8	Late
3	02/01/2025	POS LION PARCEL MANGUNDIPROJO DUA	MLG-MALANG	REGPACK	03/01/2025	1	On Time
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
10646	06/12/2025	POS LION PARCEL MANGUNDIPROJO DUA	DJJ-JAYAPURA	REGPACK	13/12/2025	7	On Time
10647	06/12/2025	POS LION PARCEL MANGUNDIPROJO DUA	DEP-DEPOK	JAGOPACK	07/12/2025	1	On Time
10648	06/12/2025	POS LION PARCEL MANGUNDIPROJO DUA	MWI-MELAWI	JAGOPACK	16/12/2025	10	Late

Keterlambatan pengiriman tidak hanya berdampak pada efisiensi operasional, tetapi juga pada reputasi perusahaan dan loyalitas pelanggan[2]. Saat ini, sebagian besar laporan keterlambatan bersifat evaluatif, artinya baru disusun setelah pengiriman selesai. Padahal, jika perusahaan memiliki sistem yang mampu memprediksi potensi keterlambatan sejak awal, langkah antisipatif seperti penyesuaian rute, jadwal, atau pemberitahuan kepada pelanggan dapat dilakukan lebih cepat[3].

Keterlambatan pengiriman yang terjadi di Lion Parcel terjadi karena tidak ada sistem prediksi keterlambatan berbasis data. Prediksi keterlambatan sebagai pendukung pengambilan keputusan operasional. Pola yang dihasilkan dalam analisa prediksi keterlambatan diharapkan dapat membantu mengoptimasi manajemen rute, sistem distribusi, atau hal-hal untuk tindakan preventif dan manajemen risiko yang lebih baik. Dalam sudut pandang pelanggan, mengingat terkait potensi keterlambatan juga berpengaruh terhadap kepuasan pelanggan. Hal tersebut dapat digunakan oleh sistem agar estimasi waktu yang diberikan tidak hanya berdasarkan data sebelumnya, tapi juga memperhatikan waktu keterlambatan yang pernah terjadi.

Dalam beberapa penelitian sebelumnya, metode prediksi berbasis *machine learning* terbukti efektif untuk mengklasifikasi dan memprediksi kejadian berbasis data historis[4]. Salah satu algoritma yang populer adalah K-Nearest Neighbor (K-NN), karena memiliki keunggulan dalam kesederhanaan implementasi dan performa klasifikasi yang stabil pada dataset berskala kecil hingga menengah [2], [5]. K-NN bekerja dengan membandingkan kemiripan antara data baru dan data historis, sehingga dapat memprediksi kemungkinan sebuah pengiriman akan “tepat waktu” atau “terlambat”[6].

Metode K-NN juga telah diterapkan dalam berbagai bidang prediktif, seperti memprediksi keterlambatan pembayaran [7], kelulusan mahasiswa[8], dan keterlambatan pengumpulan tugas[5]. Hasil-hasil penelitian tersebut menunjukkan bahwa pendekatan berbasis kedekatan data (*proximity-based learning*) mampu memberikan akurasi yang tinggi dalam memetakan pola keterlambatan. Berdasarkan keberhasilan tersebut, penelitian ini mengadaptasi pendekatan K-NN untuk konteks ekspedisi, dengan tujuan memprediksi keterlambatan pengiriman berdasarkan data historis operasional.

Dengan demikian, penelitian ini diharapkan dapat menghasilkan model prediksi yang mampu membantu perusahaan ekspedisi dalam mengidentifikasi potensi keterlambatan lebih dini. Selain itu, hasil penelitian juga dapat menjadi acuan untuk pengembangan sistem pendukung keputusan berbasis data di sektor logistik[2], [9].

Tujuan penelitian : (1) Mengetahui nilai k dengan tingkat akurasi tinggi dalam prediksi keterlambatan pengiriman barang di masa mendatang dengan algoritma K-NN, (2) Mengetahui seberapa besar tingkat akurasi dan *error* dalam prediksi keterlambatan pengiriman barang di masa mendatang dengan algoritma K-NN.

II. METODE

A. Waktu dan Tempat Penelitian

Penelitian ini dilakukan selama dua belas bulan pada bulan Januari 2025 sampai dengan bulan Oktober 2025 yang berlokasi di Lion Parcel Mangundiprojo 2 yang terletak di Buduran, Sidoarjo, Jawa Timur.

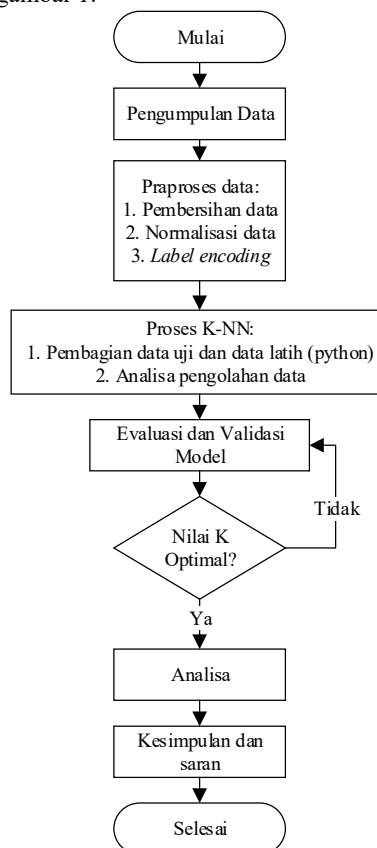
B. Pengumpulan Data

Obeservasi dilakukan secara langsung di Lion Parcel Mangundiprojo 2 Buduran selama 6 bulan mengenai situasi kondisi dan kegiatan yang sedang berlangsung, lalu dilakukan pencatatan. Selain itu observasi juga dilakukan secara tidak langsung melalui laporan pengiriman sepanjang tahun 2025. Jumlah data pengiriman periode Januari – Desember 2025 adalah 10.648 kali pengiriman.

Wawancara dilakukan kepada narasumber yang terkait yaitu, petugas operasional (kurir) sejumlah 5 orang dan staf gudang atau bagian penerimaan barang sejumlah 5 orang. Wawancara pada kurir untuk mengetahui penyebab keterlambatan pengiriman berdasarkan kondisi di lapangan serta staf gudang untuk mengetahui proses pencatatan waktu serta sortir yang mempengaruhi keterlambatan pengiriman. Wawancara dilakukan tatap muka dan tanya jawab mengenai permasalahan yang diangkat pada penelitian ini.

C. Alur Penelitian

Dalam pelaksanaan penelitian ini terdiri dari beberapa proses yang diawali dengan identifikasi permasalahan yang terjadi pada perusahaan, observasi secara langsung di lapangan, melakukan studi literatur yaitu mencari referensi dari buku, jurnal, dan penelitian terdahulu yang sejenis, merumuskan masalah yang ada, mengumpulkan data yang diperlukan, praproses data, pengolahan data, dan evaluasi model hingga mendapatkan hasil akhir penelitian. Adapun diagram alir penelitian disajikan pada gambar 1.



Gambar 1. Diagram Alir Penelitian

Berikut ini adalah tahapan pengolahan data yang dilakukan dalam penelitian ini:

Pengumpulan data dilakukan peneliti melalui studi pustaka dan studi lapangan. Studi lapangan melalui wawancara kepada narasumber untuk menghasilkan data primer pada penelitian ini serta pengamatan data histori pengiriman sepanjang tahun 2025. Studi pustaka yaitu mengumpulkan data pendukung melalui buku dan jurnal yang berkaitan dengan permasalahan yang akan diangkat pada penelitian ini. Data histori pengiriman sepanjang tahun 2025 memuat

tanggal kirim, estimasi paket tiba, tanggal aktual saat paket diterima, jenis layanan yang digunakan, alamat tujuan, dan jumlah paket. Hal tersebut digunakan untuk mengetahui jumlah paket yang dikirim tiap jenis layanan.

Praproses data dilakukan untuk menyiapkan data yang akan diolah menggunakan metode K-NN. Tahap praproses data ada 3, yaitu pembersihan data, normalisasi data, dan pemberian *label encoding*.

1. Pembersihan data merupakan proses penghapusan data yang tidak lengkap, duplikat, atau tidak relevan. Hal tersebut dilakukan agar data lebih akurat dan konsisten[5]. Data yang sudah lengkap akan dihitung lama waktu delay pengiriman sehingga dapat diketahui persentase keterlambatan tiap jenis layanan pengiriman. Data akan dianalisa secara deskriptif dan akan dipilih berdasarkan 1 jenis layanan dengan persentase keterlambatan tertinggi.
2. Normalisasi data dilakukan pada atribut numerik menggunakan *min-max normalization*. Atribut numerik yang dilakukan normalisasi yaitu jarak tempuh. Tujuan normalisasi data adalah untuk menghindari dominasi atribut skala besar [10]. Metode normalisasi ini dipilih karena metode ini mentransormasikan secara linear data sebelum dan sesudah proses sehingga menghasilkan keseimbangan [11].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Sumber: [11], [12], [13]

3. *Label encoding* merupakan proses perubahan variabel kategorikal menjadi variabel numerik. Tujuannya adalah agar data dapat terbaca oleh dan diproses oleh computer sehingga lebih mudah dalam evaluasi model[14]. Variabel kategorikal yang diubah ke dalam kategori numerik yaitu: Tepat waktu = 0, Terlambat = 1

Pengiriman dikatakan terlambat jika paket dalam status *cancel* lebih dari estimasi yang diberikan dan waktu tiba sebenarnya melebihi dari waktu tiba aktual.

Proses K-NN diawali dengan pembagian data uji dan data latih. Pembagian data set ini bertujuan untuk melakukan pengujian keandalan dalam memprediksi keterlambatan pengiriman[15]. Algoritma K-NN dengan melakukan klasifikasi untuk memprediksi apakah pengiriman akan “tepat waktu” atau “terlambat” berdasarkan prinsip kedekatan antar data uji dengan data latih[8]. Algoritma K-NN dapat mengolah data dalam jumlah besar dan mampu memberikan tingkat akurasi yang tinggi melalui proses membandingkan kedekatan antara data baru dengan data-data sebelumnya. Algoritma KNN bergantung pada nilai k yang tepat dalam proses klasifikasi[16].

Penentuan dataset atau data yang akan dibagi menjadi 80% data uji dan 20% data latih. *K-fold cross validation* merupakan salah satu metode statistik yang diimplementasikan untuk mengevaluasi performansi dari model atau algoritma yang telah dirancang[17]. Pada penelitian ini digunakan *5-fold cross validation* dalam mengevaluasi kinerja model. *Five-fold cross validation* merupakan metode yang paling direkomendasikan untuk estimasi kemampuan model dalam generalisasi data. Pada *5-fold cross validation* data dibagi menjadi 5-fold dengan ukuran sama, dimana 4-fold akan digunakan sebagai data latih dan 1-fold digunakan sebagai data uji [11].

Google Colab adalah sebuah ide untuk pemrograman *Python* dimana pemrosesan akan dilakukan oleh server Google yang memiliki perangkat keras dengan performa yang tinggi[14]. Dari sisi perangkat lunak, *Google Colab* telah menyediakan hampir sebagian besar pustaka (*library*) yang dibutuhkan. Penelitian ini menggunakan *Google Colab* untuk membantu mengolah data menggunakan metode K-NN.

Evaluasi model dilakukan dengan memperhatikan *confusion matrix* hingga didapatkan hasil dengan nilai K paling optimum[18]. Untuk setiap nilai K yang diuji akan dilakukan *Cross-Validation* hingga didapatkan data akurasi stabil dan nilai *error* rendah[19]. Penentuan nilai K menggunakan bilangan ganjil yaitu 1, 3, 5, 7, dan 9 [11].

Penelitian yang dilakukan Wardhani, dkk dengan dataset 2,126 nilai k yang besar cenderung menunjukkan penurunan akurasi. Pada penelitian tersebut akurasi tertinggi pada nilai k = 1 sampai 3. Nilai K yang lebih besar dari 3 menunjukkan penurunan akurasi yang signifikan, yang mengindikasikan bahwa nilai K yang lebih tinggi membuat model kurang efektif[20].

Analisa merupakan tahap penelitian untuk melakukan identifikasi hasil berdasarkan uji yang telah dilakukan. Analisa dilakukan untuk menganalisa hubungan keterlambatan apakah sudah memenuhi tujuan penelitian. Kesimpulan dan saran juga diambil berdasarkan hasil yang telah dianalisa.

III. HASIL DAN PEMBAHASAN

A. Praproses Data

Data yang didapatkan berdasarkan database Lion Parcel Mangundiprojo 2 yaitu terdapat 10.648 kali pengiriman selama periode Januari – Desember 2025. Praproses data bertujuan untuk memastikan bahwa data yang didapatkan dapat digunakan untuk melakukan pemodelan. Pada tahap ini dilakukan pembersihan data yang hilang atau *error*. Keterlambatan pengiriman dihitung berdasarkan durasi waktu pengiriman dan batas atas estimasi pengiriman yang ditawarkan. Jika durasi waktu pengiriman lebih dari estimasi pengiriman, maka pengiriman tersebut dikategorikan terlambat. Data kategorik dikonversi menjadi data numerik menggunakan *label encoding* agar dapat diproses algoritma. Terlambat = 1, On Time = 0.

Tabel 3. Jumlah keterlambatan pengiriman

<u>status_pengiriman count</u>	
0	10099
1	549

Sejumlah 549 pengiriman terdeteksi sebagai pengiriman yang terlambat, disajikan dalam tabel 3. Hasil perhitungan kontribusi keterlambatan tiap produk pengiriman disajikan pada tabel 4.

Tabel 4. Persentase keterlambatan pengiriman

<u>Nama Produk</u>	<u>Number of Late Deliveries</u>	<u>Total Shipments</u>	<u>Percentage Contribution (%)</u>
REGPACK	234	5212	42.62
BOSSPACK	159	1119	28.96
JAGOPACK	106	4107	19.31
BIGPACK	50	210	9.11

Berdasarkan hasil pengolahan persentase kontribusi keterlambatan pengiriman menggunakan Google Collab, diketahui bahwa produk pengiriman *REGPACK* memberikan kontribusi keterlambatan tertinggi sebesar 42,62%. Hal tersebut linear dengan jumlah pengiriman *REGPACK* dengan total pengiriman tertinggi yaitu 5.212 pengiriman.

B. Implelementasi K-NN

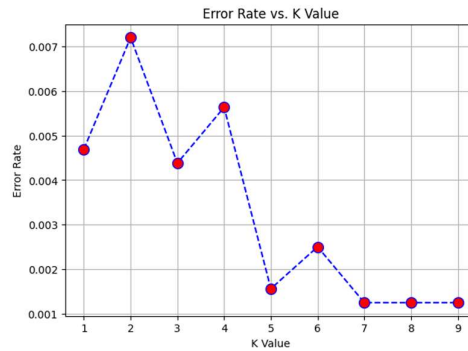
Dataset dilakukan normalisasi menggunakan *MinMax Normalization* untuk menghindari bias data atau data terbesar dan terkecil mendominasi pengolahan data, hasil ditampilkan pada tabel 5. Data dibagi menjadi 80% data uji dan 20% data latih. Pembagian dataset bertujuan untuk menguji performa setiap metode dalam mengklasifikasikan pengiriman yang mengalami keterlambatan atau tidak[15].

Tabel 5. Preview hasil normalisasi

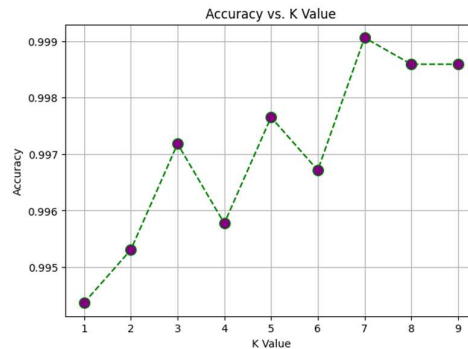
<u>Estimasi_lama_pengiriman lama_pengiriman</u>		
0	0.166667	0.030303
1	0.666667	0.090909
2	0.666667	0.030303
3	0.666667	0.060606
4	0.666667	0.030303
5	0.666667	0.060606
6	0.166667	0.030303
7	0.666667	0.030303
8	0.666667	0.030303
9	0.666667	0.030303

Berdasarkan hasil normalisasi *Min-Max Normalization*, variabel estimasi lama pengiriman dan lama pengiriman telah ditransformasikan ke dalam rentang 0 hingga 1. Proses normalisasi bertujuan untuk menyamakan skala antar variabel sehingga tidak ada dominasi salah satu variabel pada perhitungan K-NN. Oleh karena itu, model dapat melakukan proses klasifikasi secara lebih objektif tanpa dipengaruhi jarak salah satu variabel.

Nilai k yang diuji yaitu pada rentang nilai 1-10. Nilai k yang kecil cenderung menghasilkan model yang lebih akurat namun lebih sensitif terhadap noise, sedangkan k besar menghasilkan model yang lebih stabil namun berpotensi kehilangan spesifisitas [21]. Pada proses implementasi K-NN, akurasi dan *error rate* menjadi penentu nilai k terbaik dan mengevaluasi model. K terbaik yaitu nilai k dengan akurasi tertinggi dan *error rate* terendah. Hasil pengujian nilai k ditampilkan pada gambar 2 dan gambar 3.



Gambar 2. Hasil *error rate* pada k (split 80:20)



Gambar 3. Hasil akurasi pada k (80:20)

Grafik *error rate* dengan nilai *error* terendah terdapat pada nilai $k = 7$ hal tersebut didukung oleh hasil akurasi tertinggi pada nilai $k = 7$. Hasil menunjukkan nilai akurasi 99,91% dengan *precision* 100%, recall 98,29% dan F1 score 99,14%

C. Evaluasi dan Validasi Model

Evaluasi model menggunakan metode *k-fold cross validation*. *K-fold cross validation* digunakan untuk melakukan evaluasi performa dari model yang dibuat [17]. Penelitian ini menggunakan nilai *k-fold* = 5. Data set akan dibagi menjadi 5 bagian dengan ukuran yang sama dan akan diuji secara menyilang. Skema *5-fold validation* ditampilkan pada tabel 6.

Tabel 6. Skema *5-fold cross validation*

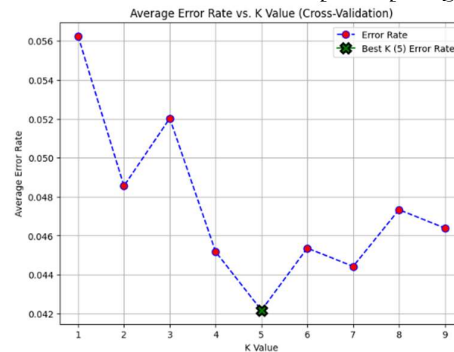
<i>kFold</i>	<i>Cross Validation</i>				
1	Uji	Latih	Latih	Latih	Latih
2	Latih	Uji	Latih	Latih	Latih
3	Latih	Latih	Uji	Latih	Latih
4	Latih	Latih	Latih	Uji	Latih
5	Latih	Latih	Latih	Latih	Uji

Nilai k yang diuji menggunakan rentang 1-10 dengan hasil evaluasi pada setiap k disajikan pada tabel 7.

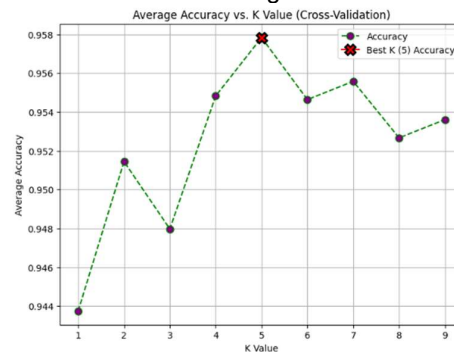
Tabel 7. Hasil kinerja menggunakan *cross validation*

Nilai K	Akurasi	Precision	Recall	F-1 score
1	94,37%	93,22%	94,37%	93,50%
2	95,14%	93,78%	95,14%	93,54%
3	94,80%	94,97%	94,80%	94,55%
4	95,48%	95,31%	95,48%	94,47%
5	95,78%	95,80%	95,78%	94,10%
6	95,46%	95,57%	95,46%	93,72%
7	95,56%	95,58%	95,56%	93,93%
8	95,27%	95,49%	95,27%	93,30%
9	95,36%	95,48%	95,36%	93,51%
10	95,17%	95,41%	95,17%	93,10%

Nilai k terbaik dengan akurasi tertinggi sebesar 95,78% dan *error rate* didapatkan pada nilai $k = 5$. Hasil pengujian nilai k menggunakan *cross validation* menunjukkan hasil yang lebih stabil. Hal tersebut disebabkan oleh *cross validation* yang dapat meminimalkan risiko *overfitting* dengan cara melatih dan menguji model melalui 5 model kombinasi data. Grafik hasil validasi model *cross validation* ditampilkan pada gambar 4 dan gambar 5.



Gambar 4. Error rate dengan cross validation

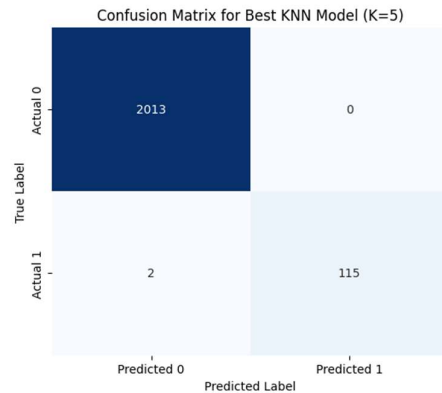


Gambar 5. Accuracy dengan cross validation

Hasil pengujian nilai k menggunakan *cross validation* menunjukkan nilai *error rate* turun secara signifikan, hasil tersebut lebih stabil dibandingkan pengujian nilai k hanya dengan pembagian data latih dan data uji 80 : 20. Hal tersebut disebabkan oleh jumlah data yang besar memungkinkan terjadinya bias jika data hanya dibagi menjadi satu bagian data uji dan data latih. Penggunaan *cross validation* membuat kombinasi data sejumlah *kfold* sehingga kombinasi data uji dan data latih dapat lebih merata.

Nilai $k = 1$ menunjukkan nilai akurasi terendah dengan persentase akurasi 94,37% dan naik secara tajam untuk $k = 2$ dengan nilai akurasi 95,14% dan turun pada nilai $k = 3$ dengan akurasi 94,80%. Nilai $k = 4$ dan $k = 5$ mengalami kenaikan secara konsisten dengan akurasi tertinggi pada nilai $k = 5$ yaitu 95,78%. Pada nilai k lebih dari 5 menunjukkan fluktuasi naik turun dengan kecenderungan menurun.

Nilai $k = 5$ menunjukkan bahwa K-NN memberikan performa terbaik dalam memprediksi keterlambatan pengiriman dengan mempertimbangkan 5 tetangga terdekat. Nilai k menunjukkan penurunan secara bertahap untuk nilai k lebih dari 5. Hal tersebut menunjukkan bahwa ketika k lebih besar, model menjadi kurang sensitif terhadap pola dalam mengidentifikasi keterlambatan pengiriman. Pemilihan nilai k penting untuk memastikan akurasi yang tinggi dalam melakukan prediksi keterlambatan pengiriman secara akurat sehingga dapat membantu manajemen dalam membuat mitigasi sebelum terjadinya keterlambatan pengiriman.



Gambar 6. Confussion matrix

	precision	recall	f1-score	support
On Time (0)	1.00	1.00	1.00	2013
Delayed (1)	1.00	0.98	0.99	117
accuracy			1.00	2130
macro avg	1.00	0.99	1.00	2130
weighted avg	1.00	1.00	1.00	2130

Gambar 7. Classification report confusion matrix

Gambar 6 menampilkan hasil evaluasi menggunakan *confussion matrix*. Evaluasi menggunakan *confussion matrix* bertujuan untuk menilai kinerja model dalam melakukan prediksi [2]. *Confussion matrix* menunjukkan jumlah prediksi benar dan salah untuk setiap kelas secara baik. Model berhasil memprediksi 2013 dengan benar untuk pengirim tepat waktu (*true positive*) dan 115 dengan benar untuk pengiriman terlambat (*true negative*). Namun, model memprediksi 2 pengiriman terlambat yang secara aktual tepat waktu (*false negative*) dan tidak ada data tepat waktu yang diklasifikasikan sebagai terlambat (*false positive*).

PENJELASAN GAMBAR 7

Berdasarkan hasil evaluasi menggunakan *confussion matrix* menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data dari kelas 1 dan cukup akurat dalam memprediksi kelas 0. Tidak adanya *false positive* menunjukkan model tidak salah dalam memprediksi data dari kelas 0 sebagai kelas 1. Hasil *precision* pada *classification report* yang menunjukkan nilai 1 menunjukkan bahwa model mampu melakukan klasifikasi dengan tepat.

D. Prediksi Keterlambatan Pengiriman

Model K-NN dengan nilai $k = 5$ merupakan model terbaik yang dapat digunakan untuk melakukan prediksi status pengiriman yang akan datang. Hasil tersebut berdasarkan analisa hasil persentase akurasi menggunakan *5-fold cross validation* dan evaluasi *confussion matrix*. Berdasarkan model tersebut, didapatkan hasil prediksi keterlambatan pada tabel 8.

Tabel 8. Persentase prediksi Keterlambatan

	Nama Produk	Delay Percentage
0	BIGPACK	20.476190
1	BOSSPACK	13.672922
3	REGPACK	4.451266
2	JAGOPACK	2.532262

Persentase keterlambatan tertinggi pada jenis pengiriman *BIGPACK* dengan persentase sebesar 20,47%, persentase keterlambatan tertinggi kedua yaitu produk *BOSSPACK* diprediksi mengalami keterlambatan sebesar 13,67%. Pada posisi ketiga produk pengiriman *REGPACK* diprediksi mengalami keterlambatan sebesar 4,45% dan *JAGOPACK* dengan persentase 2,53%.

Tabel 9. Perbandingan keterlambatan pengiriman aktual dan hasil prediksi

Nama Produk	On Time (Actual)	Delayed (Actual)	On Time (Predicted)	Delayed (Predicted)
<i>BIGPACK</i>	160	50	167	43
<i>BOSSPACK</i>	960	159	966	153
<i>JAGOPACK</i>	4001	106	4003	104
<i>REGPACK</i>	4978	234	4980	232

PENJELASAN TABEL 9

Berdasarkan perbandingan antara data aktual dan hasil prediksi menggunakan algoritma *K-Nearest Neighbor* (K-NN), model menunjukkan kemampuan prediksi yang sangat baik pada seluruh jenis layanan pengiriman. Selisih antara data aktual dan prediksi relatif kecil, yaitu 7 pengiriman pada *BIGPACK*, 6 pengiriman pada *BOSSPACK*, serta masing-masing 2 pengiriman pada *JAGOPACK* dan *REGPACK*. Secara keseluruhan, dari total 10.648 data pengiriman, model memprediksi 10.116 pengiriman tepat waktu dan 532 pengiriman terlambat, sedangkan data aktual menunjukkan 10.099 pengiriman tepat waktu dan 549 pengiriman terlambat. Perbedaan yang hanya mencapai 17 pengiriman atau sekitar 0,16% dari total data menunjukkan bahwa model K-NN mampu mempelajari pola keterlambatan pengiriman dengan baik sehingga hasil prediksi yang diperoleh sangat mendekati kondisi aktual. Hal tersebut didukung dengan hasil nilai *error rate* pada gambar 5 sebesar 0,042 atau 4,2%.

Pengiriman *BIGPACK* (260) memiliki frekuensi total paling rendah dibandingkan pengiriman menggunakan *REGPACK* (5446) yang menjadi produk pengiriman paling banyak dipilih. Meskipun secara kuantitas pengiriman *REGPACK* yang terlambat lebih banyak dibandingkan *BIGPACK*, namun secara kualitas pengiriman menggunakan *BIGPACK* diprediksi akan terjadi keterlambatan tertinggi sehingga memerlukan *improvement*.

V. SIMPULAN

Nilai *k* dengan tingkat akurasi tertinggi dalam memprediksi keterlambatan pengiriman di Lion Parcel adalah *k* = 5. Hasil akurasi nilai *k* = 5 adalah 95,78%, *precision* 95,80%, *recall* 95,78%, dan *F-1 score* 94,10%. Akurasi hasil prediksi dibuktikan dengan *confussion matrix* yang menunjukkan bahwa model sudah dapat memberikan prediksi secara tepat, hal tersebut didukung oleh tidak adanya *false positive* menunjukkan model tidak salah dalam memprediksi data dari kelas 0 sebagai kelas 1.

Optimasi nilai *k* diharapkan dapat meningkatkan keandalan sistem pendeteksian dini keterlambatan pengiriman paket dalam berbagai jenis layanan produk pengiriman. Penelitian ini memiliki keterbatasan tidak adanya jarak sebagai variabel penentu keterlambatan. Penelitian selanjutnya disarankan untuk menambahkan variabel jarak untuk mengukur performa model yang lebih komprehensif pada tiap keterlambatan produk pengiriman sehingga dapat membantu manajemen memberikan evaluasi yang lebih tepat sasaran.

REFERENSI

- [1] M. F. Ardiansyah dan I. A. S. Wulandari, "Penentuan Dalam Pemilihan Ekspedisi Menggunakan Metode Analytic Processes Network dan Dynamic Programming," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, hlm. 347–356, 2024.
- [2] A. H. Arayadiba, E. K. Nurnawati, N. Rahmawati, dan I. Setiawan, "Prediksi Tingkat Risiko Keterlambatan Pengiriman Barang Menggunakan Algoritma Decision Tree," *J. Scr.*, vol. 13, no. 01, hlm. 51–57, 2025.
- [3] L. E. Rendhiva, M. D. N. A. Saleh, R. K. D. Pramukti, dan D. Hartanti, "Implementasi Regresi Linier Untuk Waktu Pengiriman Barang," *Pros. Semin. Nas. Teknol. Inf. Dan Bisnis SENATIB*, vol. 1, no. 1, hlm. 600–605, 2023.
- [4] S. Anggraini dan E. Aryanny, "Optimizing Spare Part Delivery Routes Using Ant Colony Optimization," *Acad. Open*, vol. 10, no. 2, hlm. 6–19, Agu 2025, doi: 10.21070/acopen.10.2025.11477.
- [5] N. A. Hardiansyah, Z. D. Zahrani, dan E. Rivani, "Prediksi Tingkat Keterlambatan Pengumpulan Tugas Mahasiswa Berdasarkan Aktivitas Perkuliahan Menggunakan Algoritma K-Nearest Neighbor (KNN)," *J. Rekayasa Teknol. Nusa Putra*, vol. 11, no. 2, hlm. 103–110, 2025.
- [6] D. T. Huandito, Sri Lestanti, dan Filda Febrinita, "Implementation of K-Nearest Neighbors for Prediction of Motorcycle Service Waiting Times in Develop Tech," *J. Artif. Intell. Eng. Appl. JAIEA*, vol. 4, no. 1, hlm. 312–317, Okt 2024, doi: 10.59934/jaiea.v4i1.627.
- [7] M. R. Akhmad dan T. A. Y. Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, hlm. 185–192, Jul 2022, doi: 10.35889/progresif.v18i2.921.

- [8] R. H. Hidayatullah, R. Sanjaya, M. J. Alazami, dan R. Samsinar, "Perbandingan Algoritma K-Nearest Neighbor dan Random Forest dalam Prediksi Kelulusan Mahasiswa Universitas Muhammadiyah Jakarta," *HUBISINTEK 2025*, vol. 1, no. 1, hlm. 329–335, 2025.
- [9] N. Hidayat dan T. Wahyudi, "Prediksi Jasa Pengiriman Barang Top Trend Logistik Menggunakan Algoritma Regresi Linear pada PT. XNH: Prediction of Top Trend Logistics Goods Delivery Services Using Linear Regression Algorithm at PT. XNH," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, hlm. 1448–1455, Sep 2024, doi: 10.57152/malcom.v4i4.1536.
- [10] M. Surur, N. A. Santoso, dan B. A. Santoso, "Klasifikasi Keterlambatan Pembayaran SPP Santri Menggunakan Algoritma K-Nearest Neighbor di Pesantren Al Fajar Tegal," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, hlm. 873–883, Agu 2025, doi: 10.31004/riggs.v4i3.2117.
- [11] P. P. Allorerung, A. Erna, M. Bagussahrir, dan S. Alam, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA J. Inform. Sunan Kalijaga*, vol. 9, no. 3, hlm. 178–191, Sep 2024, doi: 10.14421/jiska.2024.9.3.178-191.
- [12] M. Sholeh, D. Andayati, dan Rr. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor dengan Normalisasi untuk Prediksi Penyakit Diabetes," *TeKa J. Teknol. Inf. Dan Komun.*, vol. 12, no. 02, hlm. 77–87, Okt 2022, doi: 10.36342/teika.v12i02.2911.
- [13] A. J. Sidauruk, J. S. Sirait, dan S. Sipayung, "Penerapan Normalisasi Data pada Angkatan Kerja Indonesia Bulan Februari 2025 Berdasarkan Kelompok Umur," *JDMIS J. Data Min. Inf. Syst.*, vol. 4, no. 1, hlm. 33–39, 2025.
- [14] M. Guntara dan F. D. Astuti, "Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran," *JIKO J. Inform. Dan Komput.*, vol. 9, no. 2, hlm. 352–360, Jun 2025, doi: 10.26798/jiko.v9i2.1605.
- [15] A. J. Winata, "Prediksi Keterlambatan Sebuah Maskapai Menggunakan Metode K-Nearest Neighbor," *J. Ilmu Comput. Dan Sist. Inf.*, vol. 13, no. 1, hlm. 1–6, 2025.
- [16] I. G. A. M. S. Yasa, E. M. Dharma, dan N. W. Utami, "Penerapan Klasifikasi Voting KNN dan Naive Bayes untuk Klasifikasi Diabetes Melitus Tipe II," *J. INOVTEK POLBENG*, vol. 10, no. 3, hlm. 1867–1876, 2024.
- [17] W. Wijiyanto, A. I. Pradana, S. Sopingi, dan V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *J. Algoritma*, vol. 21, no. 1, hlm. 239–248, Mei 2024, doi: 10.33364/algoritma/v.21-1.1618.
- [18] M. A. Rachman dan T. Sukmono, "Machine Learning Predicts Truck Breakdowns in Indonesia with 83% Accuracy," *Indones. J. Innov. Stud.*, vol. 25, no. 3, hlm. 1–17, Jun 2024, doi: 10.21070/ijins.v25i3.1156.
- [19] A. D. L. Nenohaifeto dan Y. Rada, "Klasifikasi Masa Studi Mahasiswa Program Studi Teknik Informatika Menggunakan Algoritma K-Nearest Neighbor," *JTIF J. Inov. WIRA WACANA*, vol. 04, no. 02, hlm. 20–29, 2025.
- [20] A. K. Wardhani, R. I. Sudra, E. Nugraha, dan A. N. Putri, "Optimasi Nilai K pada Algoritma K-Nearest Neighbor untuk Klasifikasi Kesehatan Janin," *J. Comput. Sci. Technol.*, vol. 5, no. 1, hlm. 44–51, 2025.
- [21] M. A. Munajad, A. Ridwan, dan T. G. Pratama, "Pengembangan Sistem Rekomendasi Musik dengan K-Means dan KNN Berbasis Cosine Similarity," *Sainteks*, vol. 22, no. 2, hlm. 153–165, Okt 2025, doi: 10.30595/sainteks.v22i2.27815.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.