



IMPLEMENTASI DATA MINING PADA KETEPATAN PENGIRIMAN BARANG DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS

Arif Rinaldi Dikananda^{1)*}, Nurjana Adi Wijaya²⁾, Mulyawan³⁾, Ahmad Faqih⁴⁾

^{1,3)} Rekayasa Perangkat Lunak, STMIK IKMI Cirebon, Kota Cirebon, Jawa Barat

^{2,4)} Manajemen Informatika, STMIK IKMI Cirebon, Kota Cirebon, Jawa Barat

email: rinaldi21crb@gmail.com¹⁾, nurjanaadiwijaya612@gmail.com²⁾,
mulyawan00@gmail.com³⁾, ahmadfaqih367@gmail.com⁴⁾

Abstrak

Perkembangan teknologi digital *E-commers* semakin meningkat, layanan pengiriman barang *online shop* dibutuhkan untuk menunjang kebutuhan sehari-hari. *Delivery man* bertugas untuk mengirim barang kepada pelanggan, dengan adanya aplikasi pengiriman barang dapat memantau keterlambatan dalam pengiriman barang ke pelanggan sehingga memperoleh data keterlambatan pengiriman barang atau disebut *Over SLA*. Namun dalam mengolah data masih menggunakan manual dengan *Microsoft Excel* sehingga kurang dalam memberikan informasi yang lebih akurat seperti akurasi tingkat ketepatan pengiriman barang, pengelompokan data keterlambatan dalam pengiriman barang. Metode yang digunakan dalam penelitian ini dengan memanfaatkan data mining menggunakan algoritma *K-Nearest Neighbors* atau KNN untuk melakukan klasifikasi atau pengelompokan data keterlambatan pengiriman barang. Metode ini digunakan dalam data mining menggunakan aplikasi machine learning Rapidminer. Penelitian ini bertujuan Untuk mengklasifikasikan data pengiriman barang dan pengelompokan data ketepatan waktu pengiriman agar data bisa diolah dengan baik sehingga menghasilkan informasi mengenai ketepatan pengiriman barang *Delivery man*, agar lebih efektif dan lebih cepat dalam menyajikan data dan mengelompokkan data.

Kata Kunci : Data Mining, Klasifikasi, K-Nearest Neighbors (KNN).

Abstract

PTThe development of digital technology E-commers is increasing, online shop goods delivery services are needed to support daily needs. The delivery man is in charge of sending goods to customers, with the delivery of goods applications can monitor delays in delivery of goods to customers so as to obtain data on delays in delivery of goods or called Over SLA. However, in processing the data, they still use manuals with Microsoft Excel so that they are lacking in providing more accurate information such as the accuracy of the accuracy of the delivery of goods, grouping of data on delays in the delivery of goods. The method used in this study by utilizing data mining using the K-Nearest Neighbors or KNN algorithm to classify or group data on delays in shipping goods. This method is used in data mining using Rapidminer machine learning applications. This study aims to classify data on delivery of goods and grouping data on timeliness of delivery so that data can be processed properly so as to produce information about the accuracy of delivery of goods by delivery man, to be more effective and faster in presenting data and classifying data.

Keywords: Data mining, Classification, K-Nearest Neighbor (KNN).



PENDAHULUAN

Perkembangan teknologi informasi semakin maju disegala bidang salah satunya dibidang dunia usaha dan bisnis E-Commerce. Belanja online merupakan transaksi jual beli secara online melalui sebuah aplikasi online shop. Pelayanan jasa pengiriman barang dibutuhkan untuk menunjang aktifitas online shop. Pengiriman barang dilakukan oleh *deliveryman* melalui sebuah aplikasi online sehingga dihasilkan data ketepatan pengiriman barang. Data tersebut menjadi sebuah acuan untuk penilaian ketepatan pengiriman barang, sehingga data mining digunakan untuk memanfaatkan data dengan jumlah yang banyak kemudian diolah menjadi informasi penting. Untuk membantu penilaian ketepatan pengiriman barang maka dilakukan dengan cara mengimplementasikan data mining dengan mengklasifikasikan data dan mengukur tingkat akurasi ketepatan[1].

Data mining merupakan proses menggunakan teknik statistika, matematika, kecerdasan buatan dan machine learning untuk mengolah *database* yang besar menjadi informasi yang bermanfaat dan pengetahuan. *Clustering* merupakan teknik proses pengumpulan data kedalam sejumlah grup sehingga masing masing grup mempunyai persamaan antar satu dengan yang lain. Metode pengumpulan data yang dilakukan dengan cara observasi atau pengumpulan data, studi pustaka, dan analisis data[2].

Penelitian dengan metode *K-Nearest Neighbor* sebelumnya pernah dilakukan oleh berberapa peneliti, salah satunya [3] dengan judul “ Implementasi Metode *K-Nearest Neighbor* Dalam Pencarian Lokasi Rumah Makan “ dengan hasil penelitian Metode KNN telah berhasil digunakan untuk mengelompokkan data rumah makan

berupa: 1) Jarak rumah makan dengan pengguna. 2) Nama rumah makan yang dicari oleh pengguna. 3) Menu makanan yang dicari oleh pengguna. 4) Harga dari menu yang dicari oleh pengguna.

Lalu penelitian tentang ketepatan waktu pengiriman sebelumnya pernah diteliti salah satunya oleh [4] dengan judul “ Pengaruh Harga dan Ketepatan Waktu Pengiriman Terhadap Kepuasan Konsumen Pada J&T Express di Kota Medan “ dengan hasil penelitian Secara simultan harga dan ketepatan waktu pengiriman berpengaruh positif dan signifikan terhadap kepuasan pelanggan sebesar 58,7% dan sisanya 41,3 %.

Fenomena yang terjadi dilapangan dalam pengiriman barang terjadi yang namanya over SLA atau keterlambatan dalam pengiriman barang sehingga mempengaruhi kinerja dari *delivery man*, karena pelayanan yang kurang maksimal. Dari data keterlambatan pengiriman tersebut menjadi acuan untuk penilaian *deliveryman* dalam pengiriman barang, dengan dibuatkan data mining bertujuan agar bisa mengelompokkan deliman yang terlambat dalam pengiriman barang dan mengukur tingkat akurasi ketepatan pengiriman barang, dan juga bisa mengevaluasi kinerja dari *delivery man*.

Data mining merupakan proses penggalian informasi dan berfungsi dari set data besar yang melibatkan konsep interdisipliner yang relatif baru yang melibatkan analisis dan penemuan pengetahuan dari database dan menggunakan pendekatan multi-sisi yang mencakup analisis statistik, visualisasi data, penemuan pengetahuan, pengenaaan pola, dan manajemen basis data [5].



METODE PENELITIAN

Metode penelitian dari data mining mempunyai beberapa model proses yang digunakan untuk pengarahannya pelaksanaan data mining, model proses yang biasa digunakan yaitu *Knowledge Discovery Databases* (KDD), CRISP-DM, SEMMA. Pada penyusunan tugas akhir ini memakai model *Knowledge Discovery Databases* (KDD).



Gambar 1. Tahapan KDD

Dari gambar diatas adalah langkah langkah pada tahapan KDD yang akan dijelaskan sebagai berikut :

1. *Data Selection*

Selection (seleksi/pemilihan) data dari merupakan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery Database* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Data Preprocessing*

Proses *Preprocessing* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau

informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal. Proses ini dilakukan untuk pembersihan bagian data yang hilang atau tidak lengkap didalam database dengan melakukan penyortiran database.

3. *Data Transformation*

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses *Data Mining*. proses ini dilakukan guna Mengubah data ke dalam bentuk yang sesuai dalam data mining dengan memberikan atribut pada data mining agar dapat mengubah skala pengukuran data asli kedalam bentuk lain.

4. *Data Mining*

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan dengan menambang informasi penting dari suatu data, mencari anomali, pola, dan korelasi dalam dataset untuk memprediksi data yang ada pada *database*, kemudian melakukan penerapan algoritma atau metode pencarian pengetahuan kedalam pola tersebut.

5. *Interpretation/Evaluation*

Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses *Data Mining* Pola informasi. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak



memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

6. *Knowledge*

Penggunaan pengetahuan yang ditemukan dari proses KDD, dimana memutuskan apa yang akan dilakukan dengan pengetahuan dihasilkan.

7. Metode Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan populer adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk interpretasi.

A. Algoritma K-Nearest Neighbors

Algoritma KNN merupakan salah satu metode yang digunakan dalam pengelompokan data. Prinsip kerja KNN adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbor*) terdekatnya dalam data pelatihan (Rismawan, dkk 2008). Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada klasifikasi vektor yang sama dihitung untuk data pembelajaran. Pada klasifikasi vektor yang sama dihitung untuk data test

(yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data pembelajaran dihitung, dan sejumlah k buah yang paling dekat diambil. Pada penerapannya KNN memiliki langkah-langkah dalam penenerapannya sebagai berikut:

1. Menentukan parameter k (jumlah tetangga yang dekat).
2. Menghitung kuadrat jarak eucliden objek terhadap data training yang diberikan.
3. Mengurutkan hasil perhitungan langkah nomer 2 secara ascending.
4. Mengumpulkan kategori Y (Klasifikasi nearest neighbor berdasarkan nilai k).
5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksikan kategori objek.

Nilai k yang terbaik ditentukan pada data, secara umumnya nilai k yang tinggi akan mengurangi efek noise pada pengklasifikasian data, tetapi membuat batasan setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dipilih dengan optimasi parameter, misalnya menggunakan *cross-validation*. Kasus khusus dimana klasifikasi diprediksi berdasarkan data pembelajaran paling dekat (dengan kata lain, $k=1$). Sesuai dengan prinsip kerja KNN yaitu mencari jarak terdekat antara data yang akan dievaluasi dengan k-tetangga terdekatnya dalam data pelatihan.

Metode KNN memiliki kelebihan dan kekurangan, kelebihan dari metode KNN antara lain, robust terhadap data yang memiliki noisy, efektif jika data training berjumlah banyak dan KNN dapat memproses klasifikasi di sisi server. Sedangkan kekurangan dari KNN antara



lain, perlu menunjukkan parameter k (jumlah tetangga terdekat), nilai komputasi yang tinggi karena memerlukan perhitungan jarak dari nilai baru ke semua data yang ada di data training, serta biaya komputasi yang cukup tinggi karena diperlukan perhitungan jarak dari tiap sampel uji pada keseluruhan sampel latih.

HASIL DAN PEMBAHASAN

A. Data Selection

Pada tahap tugas akhir ini peneliti melakukan proses pemilihan data yang relevan untuk analisis yang akan diharapkan kemudian memberikan keterangan penilaian data pengiriman barang dengan kategori sangat baik sampai kurang baik

	A	B	C	D	E	F
252	Ekmal Lubis	CRB	150.000	14.000	90.670	BAK
253	ANDIYANTO	CRB	202.000	19.000	90.590	BAK
254	IQBAL MUHAMMAD	CRB	241.000	23.000	90.460	BAK
255	Dodi Gunawan	CRB	111.000	11.000	90.090	BAK
256	Jean tumawan s.	CRB	128.000	13.000	89.840	BAK
257	AFAN NOER AZIEZ	CRB	117.000	12.000	89.740	BAK
258	Ridho aya Pang	CRB	97.000	10.000	89.690	BAK
259	TINO SUPRIATNO	CRB	122.000	13.000	89.340	CUKUP BAK
260	Muhamad Imam	CRB	224.000	25.000	88.840	CUKUP BAK
261	AQHMAD SAFRUL	CRB	256.000	29.000	88.670	CUKUP BAK
262	YASIR RAHMAN	CRB	141.000	16.000	88.650	CUKUP BAK
263	MOH ANIP PRAS.	CRB	260.000	30.000	88.460	CUKUP BAK
264	MUHAMMAD RAFI	CRB	86.000	10.000	88.370	CUKUP BAK
265	MUHAMMAD RIF.	CRB	137.000	16.000	88.320	CUKUP BAK

Gambar 2. Data Selection

B. Data Transformation

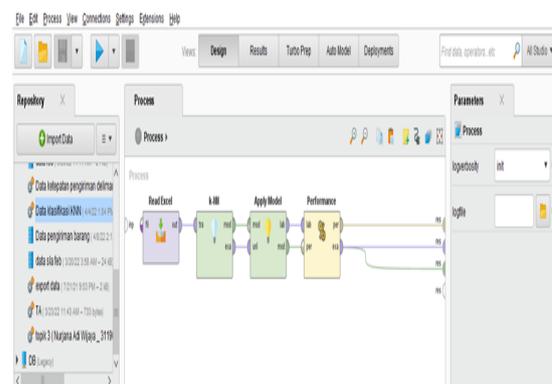
Pada tahap tugas akhir ini peneliti mengubah data ke dalam bentuk yang sesuai dalam data mining dengan memberikan atribut pada data mining agar dapat mengubah skala pengukuran data asli kedalam bentuk lain. Atribut yang digunakan yaitu polynominal, integer dan real. Kemudian mengatur role label dari atribut penilaian dan id pada atribut delivery man.

	Delivery man	Cabang	On Time	Over SLA	PctSLA	Penilaian
	polynominal	polynominal	integer	integer	real	polynominal/label
1	PANDI WJAYA	CRB	467	0	100.000	SANGAT BAK
2	AGUS SAE PUR.	CRB	260	0	100.000	SANGAT BAK
3	RAHMAT	CRB	238	0	100.000	SANGAT BAK
4	MUHAMMAD APR.	CRB	233	0	100.000	SANGAT BAK
5	MUHAMMAD MAJ.	CRB	225	0	100.000	SANGAT BAK
6	HUSNI ABOLAH	CRB	201	0	100.000	SANGAT BAK
7	JAYAMAN	CRB	179	0	100.000	SANGAT BAK
8	OM IBROHIM	CRB	174	0	100.000	SANGAT BAK
9	ABDUL WAHID	CRB	173	0	100.000	SANGAT BAK
10	GORJI SARPUJIN	CRB	128	0	100.000	SANGAT BAK
11	IRFAN MAULANA	CRB	118	0	100.000	SANGAT BAK
12	Tatang Tarsuli	CRB	101	0	100.000	SANGAT BAK

Gambar 3. Data Transformation

C. Data Mining

Pada tahap tugas akhir ini peneliti menjalankan proses design menggunakan perangkat lunak rapid miner dan merapkan model klasifikasi dan akurasi performance dengan menggunakan metode K-Nearest Neighbors.

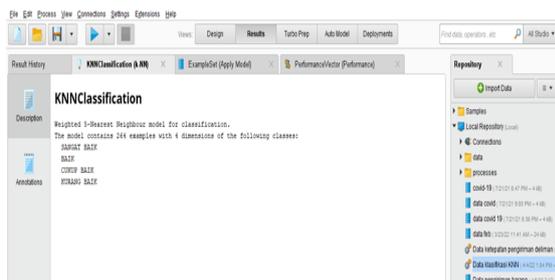


Gambar 4. Proses Data Mining

D. PEMBAHASAN

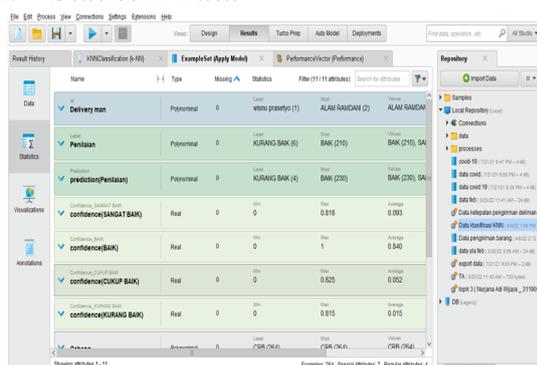
1. Klasifikasi Data

Hasil jumlah klasifikasi data dengan menggunakan K-NN terdapat 264 sampel data dengan 4 klasifikasi yang dikategorikan yaitu sangat baik, baik, cukup baik dan kurang baik.



Gambar 5. Klasifikasi Data

2. Statistik Data



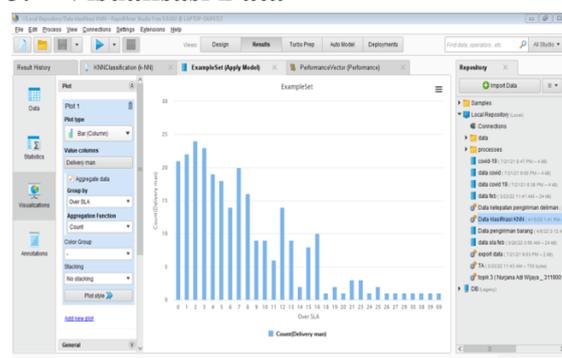
Gambar 6. Statistik Data

Dari statistik data diatas dapat diinterpretasikan sebagai berikut :

1. Data id delivery man dengan type data polynominal terdapat 264 data
2. Data label penilaian dengan type data polynominal terdapat 264 data dengan rincian 6 data kurang baik, 210 data baik, 20 data cukup baik dan 28 sangat baik
3. Data prediksi penilaian dengan type data polynominal terdapat 264 data dengan rincian 4 data kurang baik, 230 data baik, 19 data sangat baik dan 11 data cukup baik
4. Corfidance “Sangat baik” nilai terkecil 0, terbesar 0.818, rata rata 0.093
5. Corfidance “Baik” nilai terkecil 0, terbesar 1, rata rata 0.840

6. Corfidance “Cukup baik” nilai terkecil 0, terbesar 0.825, rata rata 0.052
7. Corfidance “Kurang baik” nilai terkecil 0, terbesar 0.815, rata rata 0.015
8. On time nilai terkecil 4, terbesar 467, rata rata 170.966
9. Over SLA nilai terkecil 0, terbesar 69, rata rata 7.678
10. Pct SLA nilai terkecil 50, terbesar 100, rata rata 94.886

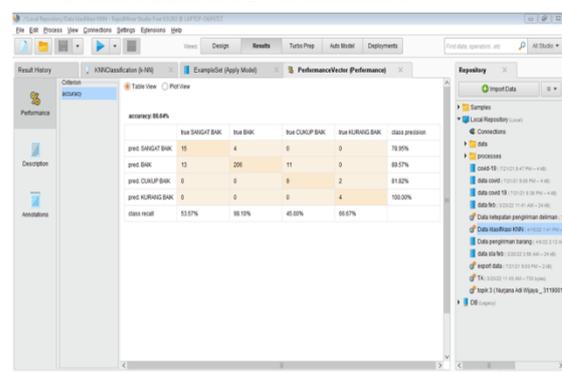
3. Visualisasi Data



Gambar 7 Visualisasi Data

Dari hasil visualisasi diatas menghasilkan exsample set apply model data.

4. Nilai Akurasi Performance



Gambar 8 Nilai Akurasi Performance



Hasil akurasi performa yaitu 88,64% dengan rincian class precision prediksi sangat baik 78,95%, prediksi baik 89,57%, prediksi cukup baik 81,82%, dan prediksi kurang baik 100%. Rincian class recall true sangat baik 53,57 %, true baik 98,10%, true cukup baik 45,00%, dan true kurang baik 66,67%. Bagian ini menyajikan hasil penelitian. Hasil penelitian dapat dilengkapi dengan tabel, grafik (gambar), dan/atau bagan. Bagian pembahasan memaparkan hasil pengolahan data, menginterpretasikan penemuan secara logis, mengaitkan dengan sumber rujukan yang relevan.

SIMPULAN

Berdasarkan pembahasan diatas dapat disimpulkan bahwa metode *K-Nearest Neighbors* dapat digunakan untuk mengolah data ketepatan pengiriman barang dengan menghasilkan klasifikasi data dan nilai *performance* prediksi ketepatan pengiriman barang sehingga dapat mempermudah untuk penilaian terhadap *delivery man* dalam pengiriman barang.

Berdasarkan hasil implementasi KNN pada data pengiriman barang didapatkan 4 klasifikasi penilaian *delivery man*, yaitu sangat baik, baik, cukup baik dan kurang baik dan nilai akurasi *performance* prediksi ketepatan pengiriman barang yaitu 88,64 %.

DAFTAR PUSTAKA

[1] E. S. Han and A. goleman, daniel; boyatzis, Richard; Mckee, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Tata Letak Merk Produk Berdasarkan Jenis Produk Di Swalayan Larisso Ambulu-Jember," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

- [2] K. Handoko and L. S. Lesmana, "Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering," *Snistek*, no. 1, pp. 97–102, 2018.
- [3] B. S. Aji, "IMPLEMENTASI METODE K - Nearest Neighbor," *Skripsi UPT Univ. Negri Semarang*, 2019.
- [4] A. R. Nasution, "Pengiriman Terhadap Kepuasan Konsumen Pada J&T Express di Kota Medan (Studi Pada Cabang Besar / Drop Point Krakatau Medan)," pp. 1–100, 2020.
- [5] I. A. Nikmatun, U. Diponegoro, I. Waspada, and U. Diponegoro, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 10, no. 2, pp. 421–432, 2019.