Jurnal Indonesia : Manajemen Informatika dan Komunikasi
www.journal.amikindonesia.ac.id/jimik/ Vol 4 No 1, Januari (2023)
E-ISSN: 2723-7079, P-ISSN: 2776-8074

K-NEAREST NEIGHBOR MENGGUNAKAN *FEATURE SELECTION BACKWARD ELIMINATION* UNTUK PREDIKSI JUMLAH PERMINTAAN DARAH PADA PMMI KOTA GORONTALO

Yulianti Lasena [1], Sunarto Taliki [2], Mohamad Efendi Lasulika [3], Andi Bode [4*]

[1,2,3,4*] Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo, Kota Gorontalo, Provinsi Gorontalo, Indonesia.

Email: yuliantylasena86@gmail.com [1], atotaliki@gmail.com [2], fendidsn.ui@gmail.com [3], andibode22@gmail.com [4*]

Histori Artikel:
Dikirim 11 Desember 2022; *Diterima dalam bentuk revisi* 25 Desember 2022; *Diterima* 15 Januari 2023; *Diterbitkan* 25 Januari 2023.

Abstrak

Pentingnya ketersediaan darah di PMI, maka diharapkan PMI untuk selalu menjaga jumlah persediaan darah untuk memenuhi kebutuhan transfusi darah. Prediksi persediaan darah sangat diperlukan untuk mengatasi permasalahan terkait dengan persediaan stok darah pada PMI Kota Gorontalo. Penerapan prediksi jumlah permintaan darah dengan Algoritma K-Nearest Neighbor dapat dilakukan untuk mengatasi permasalahan yang ada. K-NN merupakan Algoritma non-parametrik yang dapat digunakan untuk klasifikasi dan regresi. Beberapa dekade belakangan ini digunakan dalam kasus prediksi, namun Algoritma K-NN lebih baik jika diterapkan feature selection dalam menyeleksi fitur yang tidak relevan terhadap model, feature selection yang digunakan pada penelitian ini yaitu Backward Selection. penelitian ini untuk mengetahui nilai error dalam memprediksi jumlah permintaan darah pada PMI di Kota Gorontalo. Sedangkan yang menjadi Tujuan Penelitian ini untuk mencari nilai error dari Algoritma K-Nearest Neighbor dan Feature Selection yang dapat dijadikan acuan bagi pihak PMI dalam mengambil kebijakan untuk melakukan berbagai upaya untuk menjaga stok darah di masa mendatang.

Kata Kunci: Prediksi; K-Nearest Neighbor; Backward Elimination; Stok Darah.

Abstract

The importance of the availability of blood at PMI, it is expected that PMI always maintains the amount of blood supply to meet the need for blood transfusions. Prediction of blood supply is needed to overcome problems related to bloodstock supply at PMI Gorontalo. The application of predicting the number of blood requests with the K-Nearest Neighbor Algorithm can be done to overcome the existing problems. K-NN is a non-parametric algorithm that can be used for classification and regression. The last few decades have been used in prediction cases, but the K-NN algorithm is better if feature selection is applied in selecting features that are not relevant to the model, the feature selection used in this study is Backward Selection. This study aims to determine the error value in predicting the number of requests for blood at the PMI in Gorontalo City. Meanwhile, the purpose of this research is to find the error value of the K-Nearest Neighbor Algorithm and Feature Selection which can be used as a reference for PMI in making policies to make various efforts to maintainbloodstockk in the future.

Keyword: Prediction; K-Nearest Neighbor; Backward Elimination; Blood Stock.

1. Pendahuluan

PMI (Palang Merah Indonesia) merupakan sebuah organisasi himpunan nasional di Indonesia yang bergerak dalam bidang social kemanusiaan dan kesehatan. Sebagai organisasi besar, PMI mempunyai banyak unit kerja atau divisi yang masing-masing membutuhkan barang dalam pekerjaannya. Bagian logistik UTD (Unit Transfusi Darah) merupakan basis penyimpanan darah yang dibutuhkan. Mengingat pentingnya ketersediaan atau stok akan darah di PMI, maka diharapkan PMI untuk konsisten menjaga jumlah persediaan darah untuk memenuhi kebutuhan transfusi darah. Darah menjadi kebutuhan penting setiap hari yang harus dipenuhi oleh rumah sakit pada setiap daerah. Kebutuhan darah yang cukup besar, terkadang tidak bisa dipenuhi oleh Palang Merah Indonesia (PMI) pada waktu tertentu [1]. Dari permasalahan yang ada maka dibutuhkan sebuah metode yang mampu melakukan prediksi. Data mining merupakan metode untuk menemukan pengetahuan (knowledge) dalam suatu tumpukan data berdimensi tinggi. Dalam Data mining terbagi menjadi beberapa kelompok berdasarkan tugas masing-masing yaitu: Deskripsi, Estimasi, Prediksi, Klasifikasi, Clustering, dan Asosiasi [2].

Algoritma K-Nearest Neighbor merupakan salah satu pendekatan yang digunakan dalam pengklasifikasian, konsep dasar algoritma K-Nearest Neighbor yaitu mencari jarak terdekat diantara data yang terevaluasi dengan k tetangga (neighbor) paling dekat dengan data uji. KNearest Neighbor membandingkan data uji dan data training serta mencari pola data template yang paling mendekati dengan data uji. K-Nearest Neighbor memiliki kelebihan terhadap training data set dengan banyak noisy dan efektif terhadap jumlah data training tinggi/besar, namun K-Nearest Neighbor memiliki kekurangan dalam penentuan nilai k dan untuk pemelihan atribut terbaik. Feature Selection digunakan dalam pemilihan subjek dari fitur asli dengan mengeliminasi fitur yang tidak relevan. Backward Elimination metode yang memiliki fungsi untuk pengoptimalan kinerja suatu model dengan pemilihan mundur. Atribut atau variabel yang tidak relevan yang tidak berpengaruh atau tidak signifikan dalam model dihapus dari model [3]. Untuk mengatasi permasalahan yang ada penulis mengunakan feature selection untuk meningkatkan kinerja atau nilai error pada algoritma K-Nearest Neighbor dalam memprediksi jumlah Permintaan Darah pada PMI Kota Gorontalo.

Adapun yang menjadi masalah pada penelitian ini yaitu pihak PMI (Palang Merah Indonesia) perlu mengetahui Jumlah Permintaan Darah di masa mendatang. Sementara itu yang menjadi tujuan penelitian ini menghasilkan model prediksi menggunakan Algoritma *K-Nearest Neighbor* dan *Backward Elimination* untuk mencari nilai *error* terkecil dalam prediksi jumlah permintaan darah.

2. Metode Penelitian

Berdasarkan yang telah penjelasan di atas, maka objek penelitiannya adalah penerapan algoritma K-Nearest Neighbor menggunakan feature Selection Backward Elimination untuk memprediksi Jumlah Permintaan Darah. Adapun data pada Penelitian ini diambil pada PMI (Palang Merah Indonesia) Kota GorontaloAlgoritma K-Nearest Neighbor merupakan salah satu metode algoritma terbimbing yang terbagi menjadi dua jenis yaitu supervised learning dengan unsupervised learning. Algoritma pembelajaran terawasi bertujuan untuk mendapatkan pola baru, sedangkan tanpa pengawasan bertujuan untuk mempertahankan pola dalam data. Akurasi algoritma K-Nearest Neighbor ditentukan oleh ada dan tidak adanya data yang tidak relevan atau ketika bobot fitur sesuai dengan relevansinya. Algoritma K-Nearest Neighbor adalah salah satu metode analisis klasifikasi, tetapi dalam beberapa dekade terakhir metode K-Nearest Neighbor juga telah digunakan untuk prediksi. Pada data latih, carilah jarak terpendek antara data yang akan dievaluasi dan k tetangga yang terdekat dengannya. Ruangan ini dibagi menjadi beberapa kelompok berdasarkan klasifikasi data latih. Suatu titik pada titik ini ditandai dengan kelas c, dimana jika kelas c merupakan klasifikasi yang sangat sering dijumpai pada k tetangga terdekat titik tersebut. Untuk sistem kerja K-Nearest Neighbor, data latih diproyeksikan ke lokasi multidimensi, dimana karakteristik data dari

Jurnal Indonesia : Manajemen Informatika dan Komunikasi, Volume 4 No 1, Januari (2023), pp. 303-309
DOI : https://doi.org/10.35870/jimik.v4i1.172

masingmasing dimensi direpresentasikan [4]. Model persamaan Algorima *K-Nearest Neighbor* seperti dibawah ini:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (1)$$

Keterangan :
X : Sampel data
Y : Data uji
D : Jarak
x' : Perkiraan atau estimasi K = Jumlah tetangga terdeka ($\_x'$) = Tetangga terdekat yi = Output tetangga terdekat

$$fk(x)^2 = x = \frac{k}{1} \sum_{i \in N} k(x') \quad (2)$$

Keterangan :
x' : Perkiraan atau estimasi
K : jumlah tetanga terdekat
($\_x'$) : Tetangga terdekat
yi : Output tetangga terdekat

Adapun model atau alur pada penelitian ini yaitu Keterangan: Yang pertama adalah pengumpulan data kemudian Data set yang diperoleh seterusnya akan di preprocessing, setelah melakukan preprocessing maka selanjutnya data dibagi menjadi dua buah data set, yakni data training dan data testing, setelah melakukan pembagian data maka langkah seterusnya yakni masuk dalam tahap proses algoritma K-NN, kemudian K-NN Menggunakan Backward Elimination, Backward Elimination digunakan untuk mengeluarkan variable yang tidak signifikan, Backward Elimination memiliki fungsi untuk mengoptimalkan kinerja suatu model dengan sistem kerja pemilihan mundur, setelah itu maka dilakukan evaluasi model terbaik yang akan digunakan untuk memperoleh hasil prediksi. Pada dalam proses algoritma K-NN dan K-NN - Backward Elimination dilakukan percobaanpercobaan dengan menganti-ganti nilai ketetanggaan atau nilai k pada setiap perobaan data training mulai dari 1 sampai dengan 7 periode, hal ini guna untuk mendapatkan model yang baik sehingga diperoleh nilai RMSE yang lebih kecil.

3. Hasil dan Pembahasan

Pada tahap ini peneliti melakukan normalisasi data, ini dilakukan untuk mengelompokkan data ke dalam skala atau jangkauan tertentu sehingga dapat mempermudah dalam melakukan pengelohan data. Data yang di dapatkan berupa data univariate sehingga perlu untuk merubah menjadi multivariate agar menjadi beberapa variable independent. Pengolahan ini dilakukan menggunakan Microsoft excel dengan persamaan rumus sebagai berikut :

$$New\ Data = \frac{(Data - Min)\ x\ (Newmax - Newmin)}{(Max - Min) + Newmin}$$

Keterangan :
Data : Variabel Harga
Min : nilai terkecil dari variabel harga : 371
Max : nilai tertinggi dari variabel data : 9012
Newmax : 1
Newmin : 0

Jurnal Indonesia : Manajemen Informatika dan Komunikasi, Volume 4 No 1, Januari (2023), pp. 303-309
DOI : https://doi.org/10.35870/jimik.v4i1.172

Tabel 1. Hasil Normalisasi Data

Minggu		Hasil Normalisasi
Minggu Ke	1	0,229051543
Minggu Ke	2	0,270418474
Minggu Ke	3	0,293589327
Minggu Ke	4	0,278693274
Minggu Ke	5	0,278933533
Minggu Ke	6	0,265726359
Minggu Ke	7	0,292536427
Minggu Ke	8	0,298189578
Minggu Ke	9	0,336221151
Minggu Ke	10	0,355632658
Minggu Ke	11	0,31290897
Minggu Ke	12	0,345888039
Minggu Ke	13	0,324370734
Minggu Ke	14	0,358678293
Minggu Ke	15	0,378563252
Minggu Ke	16	0,398427011
Minggu Ke	17	0,44412567
Minggu Ke	18	0,45733991
Minggu Ke	19	0,425908391
Minggu Ke	20	0,404864536
……..	….	…….
Minggu Ke	293	0,483450401

Setelah di lakukan perubahan data normalisasi langkah selanjutnya yaitu merubah data menjadi Univariat hal ini dilakukan untuk mendapatkan hasil yang lebih baik dari pengolahan data, data univariat terdiri dari beberapa periode yaitu 1 periode, 4 periode dan 6 Periode, Adapun contoh data univariat dapat di lihat pada table di bawah ini :

Tabel 2. Data Univariat 6 Periode

Xt-5	Xt-4	Xt-3	Xt-2	Xt-1	Xt
0,265726	0,278934	0,278693	0,293589	0,270418	0,229051543
0,292536	0,265726	0,278934	0,278693	0,293589	0,270418474
0,29819	0,292536	0,265726	0,278934	0,278693	0,293589327
0,336221	0,29819	0,292536	0,265726	0,278934	0,278693274
0,355633	0,336221	0,29819	0,292536	0,265726	0,278933533
0,312909	0,355633	0,336221	0,29819	0,292536	0,265726359
0,345888	0,312909	0,355633	0,336221	0,29819	0,292536427
0,324371	0,345888	0,312909	0,355633	0,336221	0,298189578
0,358678	0,324371	0,345888	0,312909	0,355633	0,336221151
0,378563	0,358678	0,324371	0,345888	0,312909	0,355632658
0,398427	0,378563	0,358678	0,324371	0,345888	0,31290897
0,444126	0,398427	0,378563	0,358678	0,324371	0,345888039
0,45734	0,444126	0,398427	0,378563	0,358678	0,324370734
0,425908	0,45734	0,444126	0,398427	0,378563	0,358678293
…	….	…	…	…	…
0,48345	0,509978	0,498573	0,4425	0,511554	0,573823085

Jurnal Indonesia : Manajemen Informatika dan Komunikasi, Volume 4 No 1, Januari (2023), pp. 303-309
DOI : https://doi.org/10.35870/jimik.v4i1.172

3.1 Hasil Pengujian K-NN

Pada Tahap awal data di ujicoba dengan menggunakan algoritma K-NN, pengujian ini dilakukan uji coba dengan menggunakan beberapa nilai K untuk mencari hasil akurasi yang baik, serta mencoba dari beberapa data periode yang telah di tentukan.

Tabel 3. Hasil Uji 2 Periode

Parameter K	RMSE
3	0,064
5	0,059
7	0,056

Dari hasil table di atas dapat di lihat bahwa parameter K = 7 Merupakan yang terbaik, dengan data set 1 periode ini dimana semakin tinggi nilai K pada K-nn akan mendapatkan hasil yang lebih baik lagi.

Tabel 4. Hasil Uji 4 Periode

Parameter K	RMSE
3	0,070
5	0,067
7	0,066

Sementara dari Data 4 Periode dapat di lihat pada table di atas hasilnya pun tidak terlalu berbeda jauh dari masing2 nilai Parameter K.

Tabel 5. Hasil Uji 6 Periode

Parameter K	RMSE
3	0,072
5	0,068
7	0,068

Pada Hasil uji selanjutnya dengan data 6 Periode dimana parameter K 7 dan 5 mendapatkan nilai atau hasil yang sama. Dari hasil pengujian-pengujian melakukan KNN dapat di lihat bahwa semakin tinggi nilai Parameter K maka akan menghasilkan nilai RMSE yang baik, sementara denga adanya pembagian periode pada data akan berpengaruh juga terhadap hasil RMSE yang di hasilkan. Setelah melakukan pengujian menggunakan K-nn dapat disimpulkan bahwa parameter K 7 pada dua periode menghasilkan nilai RMSE terbaik yaitu 0,056.

3.2 Hasil Pengujian Menggunakan K-NN dan *Backward Elimination*

Setelah pengujian menggunakan kedua algoritma di atas selanjutnya mencoba menambahkan forward selection ke dalam pengolahan data untuk melihat apakah hasil yang di hasilkan sama baiknya atau malah mendapatkan nilai rata-rata.

Tabel 6. Hasil uji K-NN dan *Backward Elimination*

Parameter K	RMSE
3	0,064
5	0,059
7	0,056

Dari table di atas dapat dilihat bahwa dengan menggunakan *Backward Elimination* terlihat sama atau tidak berbeda jauh dengan tanpa menggunakan *Backward Elimination*, sebab tanpa menggunakan *Backward Elimination* algoritma K-NN sangat baik dalam prediksi ataupun klasifikasi.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi, Volume 4 No 1, Januari (2023), pp. 303-309
DOI : https://doi.org/10.35870/jimik.v4i1.172

Pengujian di atas juga telah di lakukan ujicoba dengan beberapa data periode di mulai dari periode 2, 4 dan 6 rata-rata mendapatkan nilai RMSE 0,05.

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan metode K-nn merupakan metode yang sangat bagus atau baik dalam melakukan prediksi ataupun klasifikasi, hal ini dapat dilihat dari hasil RMSE yang di hasilkan yaitu 0,05, metode ini mampu menghasilkan nilai terbaik walaupun tanpa adanya penambahan metode lain seperti *Backward Elimination*. Saran Sebaiknya melakukan uji coba menggunakan forward selection dalam penelitian ini agar bisa menjadi acuan atau pembelajaran selanjutnya sebab dengan menambahkan metode ini dan dapat menghasilkan akurasi ataupun nilai RMSE yang sangat baik.

5. Daftar Pustaka

[1] Rachman, B., 2003. Dinamika harga dan perdagangan komoditas jagung. *Soca: Jurnal Sosial Ekonomi Pertanian*, *3*(1), p.1-15.

[2] Lasulika, M.E., 2019. Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel. *ILKOM Jurnal Ilmiah*, *11*(1), pp.11-16. DOI: https://doi.org/10.33096/ilkom.v11i1.408.11-16.

[3] Lasulika, M.E., 2017. Prediksi Harga Komoditi Jagung Menggunakan K-Nn Dan Particle Swarm Optimazation Sebagai Fitur Seleksi. *ILKOM Jurnal Ilmiah*, *9*(3), pp.233-238. DOI: https://doi.org/10.33096/ilkom.v9i3.148.233-238.

A. Bode, A., 2017. K-nearest neighbor dengan feature selection menggunakan backward elimination untuk prediksi harga komoditi kopi arabika. *ILKOM Jurnal Ilmiah*, *9*(2), pp.188-195. https://doi.org/10.33096/ilkom.v9i2.139.188-195.

[4] Rasyidi, M.A., 2017. Prediksi Harga Bahan Pokok Nasional Jangka Pendek Menggunakan ARIMA. *Journal of Information Systems Engineering and Business Intelligence*, *3*(2), p.107. DOI: https://doi.org/10.20473/jisebi.3.2.107-112.

[5] Sari, Y., 2017. Prediksi harga emas menggunakan metode neural network backpropagation algoritma conjugate gradient. *Jurnal Eltikom*, *1*(2), pp. 64–70. DOI: https://doi.org/10.31961/eltikom.v1i2.21.

[6] Sya'baniyah Pangesti, C.S. and Rismawan, T., Aplikasi Prediksi Harga Sembako Menggunakan Metode Box-Jenkins Berbasis Website. *Coding Jurnal Komputer dan Aplikasi*, *6*(3), pp. 139-149. DOI: http://dx.doi.org/10.26418/coding.v6i3.29045.

[7] Rahma, I.N. and Setiadi, T., 2014. Penerapan Data Mining Untuk Memprediksi Jumlah Penumpang Bus Trans Jogja Menggunakan Time Series Data. *Jurnal Sarjana Teknik Informatika, 2*(3), pp. 161–171. DOI: http://dx.doi.org/10.12928/jstie.v2i3.2886.

[8] Yustanti, W., 2012. Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah. *Jurnal Matematika, Statistika dan Komputasi*, *9*(1), pp.57-68. DOI: https://doi.org/10.20956/jmsk.v9i1.3399.

[9] Pratiwi, R.W. and Nugroho, Y.S., 2016. Prediksi Rating Film Menggunakan Metode Naïve Bayes. *DutaCom*, *12*(1), pp.91-108.

[10] Nanja, M. and Purwanto, P., 2015. Metode k-nearest neighbor berbasis forward selection untuk prediksi harga komoditi lada. *Pseudocode*, *2*(1), pp.53-64. DOI: https://doi.org/10.33369/pseudocode.2.1.53-64.

[11] Drajana, I.C.R., 2017. Metode support vector machine dan forward selection prediksi pembayaran pembelian bahan baku kopra. *ILKOM Jurnal Ilmiah*, *9*(2), pp.116-123. DOI: https://doi.org/10.33096/ilkom.v9i2.134.116-123.