



Komparasi Algoritma Data Mining Menggunakan Forward Selection pada Prediksi Harga Jagung

Mohamad Efendi Lasulika¹, Andi Bode²

Teknik Informatika, Universitas Ichsan Gorontalo

Email: fendilasulika123@gmail.com¹, andibode22@gmail.com²

Abstrak

Meningkatnya volume produksi jagung yang diperdagangkan dan kecenderungan harga jagung dipasaran menuntut kebijakan pemerintah dalam mengendalikan stabilisasi harga jagung. Sehingga pemerintah kesulitan dalam menentukan atau memprediksi harga komoditi yang akan datang, namun pada penelitian ini hanya fokus pada beberapa algoritma klasifikasi untuk mengetahui algoritma apakah yang mempunyai tingkat akurasi tertinggi dalam hal prediksi harga jagung sehingga dapat digunakan dalam melakukan prediksi harga beberapa hari kedepannya. Dalam data mining ada beberapa algoritma klasifikasi yang dapat digunakan, seperti K-Neural Network, Artificial Neural Network, Naïve Bayes, Regresi Linear, C-45 dll, namun pada penelitian ini hanya fokus pada dua metode saja yaitu Naïve Bayes dan K- Neural Network. Berdasarkan hasil penelitian yang dilakukan metode K-nn merupakan metode yang sangat bagus atau baik dalam melakukan prediksi ataupun klasifikasi, hala 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 forward selection, sementara itu untuk naïve bayes metode ini juga merupakan metode terbaik dalam melakukan prediksi ataupun klasifikasi, akan tetapi naïve bayes mempunyai beberapa kekurangan apabila digunakan untuk type data univariate ataupun numerical. Penambahan forward selection kepada pengolahan data dapat membantu menghasilkan akurasi yang baik pula. Walaupun tanpa forward selection K-NN dan Naïve bayes merupakan metode komputasi yang sangat baik dalam prediksi ataupun klasifikasi.

Kata kunci: Harga Jagung, Komparasi Metode, K-Neural Network, Naïve bayes, forward selection

Abstract

Failure mode risk analysis is an effort that can be implemented to detect any risk in which it influences failure mode particularly production mode that will effect to system failure. The analysis will be used as a reference to overcome the emerge risk later and it also can be used to improve or minimize the failure before the working system reduces. The concept of failure risk analysis with conventional FMEA method (Failure Mode Effect Analysis) and Fuzzy FMEA that is implemented in 5 (five) stages of production processes of purifying water got 20 failure mode. The determination of priority order of process failure improvement is by calculating the value of RPN with conventional FMEA method and summing up the value of FRPN with Fuzzy FMEA method. There is significant different between the results of the both method as a result it reduces the bias in getting priority improvement information that should be done. The findings are expected to give improvement in real situation of a company.

Keywords: Risk, Process of Purifying Water, FMEA, Fuzzy

A. PENDAHULUAN

Provinsi Gorontalo adalah salah satu provinsi yang telah banyak megimpor produksi jagung ke berbagai daerah maupun ke luar daerah, mengingat mayoritas penduduk yang ada di Provinsi Gorontalo adalah petani jagung. Akan tetapi kebutuhan komoditas jagung yang digunakan untuk pangan, pakan dan bahan baku untuk industri terus meningkat sehingga untuk memenuhi kebutuhan tersebut sebagian dipenuhi dari jagung impor, Semakin meningkatnya volume impor jagung dimana harga impor tersebut relative rendah dibandingkan harga jagung di pasaran disamping itu kualitas harus produkpun harus terjamin. Meningkatnya volume produksi jagung yang diperdagangkan dan kecenderungan harga jagung dipasaran menuntut kebijakan pemerintah dalam mengendalikan stabilasi harga jagung, sehingga pemerintah kesulitan dalam menentukan harga jagung dipasaran apakah naik atau turun (Rachman, 2013).

Penelitian ini hanya fokus pada beberapa algoritma klasifikasi untuk mengetahui algoritma apakah yang mempunyai tingkat akurasi

tertinggi dalam hal prediksi harga jagung. Dalam data mining ada beberapa algoritma klasifikasi yang dapat digunakan, seperti K-Neural Network, Artificial Neural Network, Naïve Bayes, Regresi Linear, C-45 dll, namun pada penelitian ini hanya fokus pada dua metode saja yaitu Naïve Bayes dan K- Neural Network. Metode tersebut merupakan metode yang terbaik, dimana pada penelitian yang telah pernah dilakukan dengan menggunakan tipe data multivariat mendapatkan Naïve Bayes merupakan metode yang akurat dengan memiliki nilai accuracy tertinggi yaitu 96% sehingga metode ini termasuk dalam kategori klasifikasi terbaik (Lasulika, 2019). Sementara itu K-NN juga termasuk dalam kategori algoritma klasifikasi terbaik, pada penelitian sebelumnya dalam memprediksi harga jagung menggunakan particle swarm optimization didapatkan hasil akurasi sebesar 98,7 % (Lasulika, 2017), pada beberapa penelitian yang telah dilakukan dimana k-nn menggunakan backward elimination dapat memperkecil nilai RMSE yang dihasilkan dibandingkan dengan K-nn tanpa *backward elimination* (Bode, 2017).

Adapun pada penelitian ini akan di uji coba kedua metode tersebut dengan menggunakan forward selection untuk mengetahui hasil akurasi tertinggi dari Naïve Bayes, K-NN, K-nn Menggunakan Forward selection dan Naïve Bayes Menggunakan forward selection, forward selection di tujukan untuk meningkatkan kinerja dari kedua metode tersebut. Adapun yang menjadi Rumusan Masalah pada penelitian ini bagaimana hasil akurasi dari penerapan metode KNN, Naïve bayes serta K-nn Menggunakan *Forward Selection* dan *Naïve Bayes* menggunakan Forward Selection dalam memprediksi Harga Jagung. Tujuan Penelitian ini yaitu untuk memperoleh akurasi yang terbaik dari kedua metode tersebut sehingga dapat diimplementasikan dalam hal memprediksi harga jagung.

Sementara itu Manfaat dari penelitian ini agar bisa mengetahui algoritma mana yang bisa menghasilkan nilai akurasi terbaik dengan menggunakan forward selection atau tanpa Forward Selection.

Ada beberapa penelitian terkait dengan penelitian ini antara lain yaitu penelitian yang di lakukan oleh Mohammad Arif Rasyidi pada tahun 2017 yang berjudul Prediksi Harga Bahan Pokok Nasional Jangka Pendek Menggunakan ARIMA Penelitian ini bertujuan untuk mengatasi permasalahan fluktuasi harga bahan pokok yang tidak terkendali sehingga menyebabkan kerugian bagi konsumen maupun produsen, langkah yang diambil yaitu dengan membuat prediksi harga menggunakan ARIMA, hasil eksperimen menunjukkan bahwa model arima yang dihasilkan mampu memprediksi harga dengan tingkat error rata-rata sebesar 2,22 % (Rasyidi, 2017). Selanjutnya penelitian yang di lakukan Mohammad Guntur dkk yaitu tentang Prediksi Harga Emas dengan menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko Pada penelitian algoritma naïve bayes dipercaya mampu menghasilkan sebuah klasifikasi yang digunakan untuk memprediksi harga emas yang bisa membantu pengambil keputusan dalam menetapkan apakah harus menjual atau membeli emas, berdasarkan hasil perhitungan metode ini mampu digunakan untuk memprediksi harga emas untuk 14 hari kedepan dengan pengujian sebanyak 16 Data dan tingkat akurasi sebesar 75 % (Sari, 2017). Selanjutnya penelitian yang di lakukan Sy'baniyah Pengesti, dkk pada tahun 2018 yaitu tentang Aplikasi Prediksi Harga Sembako Menggunakan Metode *Box-Jenkins* Berbasis Website Pada penelitian ini telah dibuat aplikasi prediksi harga sembako menggunakan metode *BoxJenkins* ARIMA berbasis website, dalam penelitian ini Metode yang digunakan

dapat menghasilkan rata-rata 90 % akurasi yang dihasilkan (S. Pangesti & Rismawan, 2017).

Saat ini Kemajuan Ilmu pengetahuan dan teknologi meningkatkan pengertian mengenai berbagai aspek lingkungan dan mengakibatkan banyak peristiwa yang dapat diramalkan. Kecenderungan untuk dapat meramalkan peristiwa secara lebih tepat, khususnya dalam bidang ekonomi akan memberikan dasar yang baik bagi perencanaan (Soepome, 2014).

B. METODE

1. Metode K-Neural Network

Algoritma K-NN merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat objek. Algoritma ini memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar. Sedangkan kelemahan pada Algoritma ini adalah penentuan nilai K yaitu jumlah tetang terdekat, dekat atau jauhnya tetangga terdekat biasanya dihitung dengan berdasarkan jarak Euclidean dengan Rumus Persamaan dan langkah-langkah di bawah ini (Yustanti, 2012).

- a. Menentukan Nilai Parameter K.
- b. menghitung jarak (Euclidean Distance)

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

Keterangan :

X = sampel data

Y = data uji

D = Jarak

- c. Mengurutkan hasil jarak dan menentukan tetangga terdekat berdasarkan nilai K
- d. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru Sementara untuk mencari nilai prediksi k-NN dapat dihitung dengan persamaan

$$Y = \frac{1}{K} \sum_{i=1}^k y_i \quad (2)$$

Y = Perkiraan

K= jumlah tetangga terdekat

Y_i = output Nilai K

2. Naïve Bayes

Metode Naïve Bayes adalah sebuah pengelompokan statistic yang bisa dipakai untuk memprediksi probabilitas anggota suatu class, selain itu naïve bayes juga mempunyai akurasi yang sangat akurat, berikut persamaan naïve bayes (Wijayatun & Sulisty, 2016).

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (3)$$

Keterangan :

y = Data dengan class yang belum diketahui

x = hipotesis data y

P(x | y) = probabilitas hipotesis x berdasarkan kondisi y

P(x) = probabilitas hipotesis x

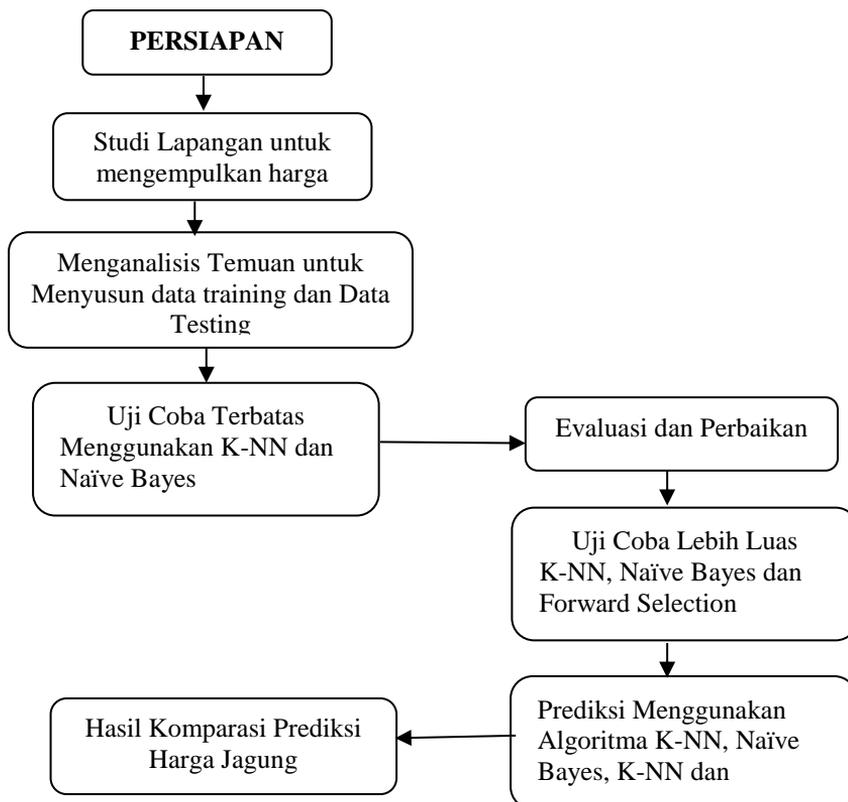
P(y | x) = probabilitas y berdasarkan kondisi pada hipotesis x

P(y) = probabilitas dari y

3. Forward Selection

Forward Selection adalah salah satu prosedur yang bertujuan menambah variabel dikendalikan satu persatu kedalam persamaan yang didasarkan pada alpha tertentu untuk masuknya (Nanja, 2015). Dalam pemilihan variabel dilakukan dengan cara pemilihan mundur yaitu menguji semua variabel kemudian menghapus variabel yang dianggap tidak relevan (Drajana, 2017).

4. Alur Penelitian



Gambar 1. Alur Penelitian

C. HASIL DAN PEMBAHASAN

1. Pengolahan Data

Tahap pengolahan data peneliti melakukan normalisasi data, ini dilakukan untuk mengelompokkan data ke dalam skala atau jangkauan tertentu sehingga dapat mempermudah dalam melakukan pengolahan 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.

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

Keterangan:

Data = Variabel Harga

Min = nilai terkecil dari variabel harga : 371

Max = nilai tertinggi dari variabel data : 9012

Newmax = 1

Newmin = 0

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

2. 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 masing-masing 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. Pengujian Naïve Bayes

Tahap selanjutnya yaitu ujicoba menggunakan Algoritma Naïve Bayes, Algoritma ini tidak akan bisa mengolah data apabila type data numerical, maka dari itu peneliti mencoba merubah type data dari numerical ke polynominal, adapun hasil yang di dapatkan yaitu dapat di lihat pada table di bawah ini.

Tabel 6. Hasil Uji Metode Naïve Bayes

Periode Data	RMSE
2	0,493
4	0,058

Dari hasil uji coba yang di lakukan ternyata naïve bayes sangat tidak cocok digunakan apabila type angka berupa angka, jadi perlu dilakukan perubahan data set numerical ke dalam bentuk data polynominal atau text. Di sini hasil terbaik yaitu terdapat pada 4 periode dengan nilai RMSE 0,058.

4. Hasil Pengujian Menggunakan K-NN Dan Forward Selection

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 7 Hasil uji K-NN dan Forward Selection

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

Dari table di atas dapat dilihat bahwa dengan menggunakan forward selection terlihat sama atau tidak berbeda jauh dengan tanpa menggunakan forward selection, sebab tanpa menggunakan forward selection algoritma K-NN sangat baik dalam prediksi ataupun klasifikasi. Pengujian di atas juga telah dilakukan dengan beberapa data periode di mulai dari periode 2, 4 dan 6 rata-rata mendapatkan nilai RMSE 0,05. Proses pengolahan data menggunakan Naïve bayes dan Forward selection tidak dapat dilakukan karena memang naïve bayes tidak dapat dilakukan dengan menambahkan forward selection apabila type datanya berupa Numerical, sehingga perlu adanya perubahan data agar bisa dilakukan pengolahan data. Penggunaan naïve bayes dan forward selection sendiri sudah banyak dilakukan dengan type data text atau selain numerical, seperti prediksi klasifikasi atau lainnya.

5. Hasil Komparasi Prediksi Harga Jagung

Setelah melakukan beberapa uji coba pengujian maka dapat disimpulkan hasil perbandingan dari kedua metode tersebut dapat dilihat pada tabel di bawah ini.

Tabel 8. Perbandingan Metode Berdasarkan AUC

Metode	RMSE
K-NN	0,059
Naïve Bayes	0,058
Knn, Naïve Bayes dan Forward Selection	0,056

Dari Gambar di Atas dapat dilihat dengan menambahkan forward selection bisa menghasilkan nilai RMSE terbaik, namun tanpa adanya forward selection ke dua metode ini baik K-nn dan Naïve bayes merupakan metode yang bagus dalam melakukan prediksi ataupun klasifikasi. Hal ini dapat dilihat dari hasil RMSE yang di hasilkan yaitu rata-rata 0,05. Sementara itu untuk metode naïve bayes harus melihat data yang akan digunakan, sebab naïve bayes akan lebih baik apabila type data bukan berupa numerical atau dengan type data univariate.

D. PENUTUP

Simpulan dan Saran

Berdasarkan hasil penelitian yang dilakukan metode K-nn merupakan metode yang sangat bagus atau baik dalam melakukan prediksi ataupun klasifikasi, hala 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 forward selection, sementara itu untuk naïve bayes metode ini juga merupakan metode terbaik dalam melakukan prediksi ataupun klasifikasi, akan tetapi naïve bayes mempunyai beberapa kekurangan apabila digunakan untuk type data univariate ataupun numerical. Saran saya Sebaiknya dalam melakukan penelitian sebelum memilih ke dua metode ini disarankan untuk memperhatikan data yang akan di gunakan sebab ke dua metode ini mempunyai kelebihan dan kekurangan pada saat pengolahan data, adanya penambahan forward selection dalam penelitian ini bisa menjadi acuan atau pembelajaran selanjutnya sebab dengan menambahkan metode ini dapat menghasilkan akurasi ataupun nilai RMSE yang sangat baik.

DAFTAR PUSTAKA

- Bode, A. (2017). K-Nearest Neighbor dengan Feature Selection Menggunakan Backward Elimination untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, 9 (2): 188-195.
- Drajana, I. C. (2017). Metode Support Vector Machine dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku. *ILKOM Jurnal Ilmiah*, 9 (2): 116-123.
- Lasulika, M. E. (2017). Prediksi Harga Komoditi Jagung Menggunakan K-NN dan Particle Swarm Optimazation. *ILKOM Jurnal Ilmiah*, 9 (3): 233-238.
- Lasulika, M. E. (2019). Komparasi Naive Bayes, Support Vector Machine dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi. *ILKOM Jurnal Ilmiah*, 11 (1): 11-16.
- Nanja, M. dan Purwanto. (2015). Metode K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Harga Komoditi Lada. *PSEUDOCODE*, 2 (1): 53-64.
- Rachman, B. (2013). Dinamika Harga dan Perdagangan Komoditas Jagung. *SOCA : Jurnal Sosial Ekonomi Pertanian in Indonesian*, 3 (1): 1-15.
- Rasyidi, M. A. (2017). Prediksi Harga Bahan Pokok Nasional Jangka Pendek Menggunakan Arima. *Journal of Information Systems Engineering and Business Intelligence*, 3 (2): 107-112.
- Pangesti, S., Suhery, C., dan Rismawan, T. (2017). Aplikasi Prediksi Harga Sembako Menggunakan Box- Jenkins Berbasis Website. *Coding Jurnal Komputer dan Aplikasi*, 6 (3): 64-70.
- Sari, Y. (2017). Prediksi Harga Emas Menggunakan Metode Neural Network Backropagation Algoritma Conjugate Gradient. *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi dan Komputer*, 1 (2): 64-70.
- Soepome, P. (2014). Penerapan Data Mining Untuk Memprediksi Jumlah

Penumpang Bus Trans Jogja. 161-171.

Wijayatun, R., & Sulisty, Y. (2016). Algoritma K-Nearest Neighbor Untuk Memprediksi Harga Jual Tanah. 60-63.

Yustanti, W. (2012). Algoritma K-Nearest Neighbor Untuk Memprediksi Harga Jual Tanah. 57-68.

