

PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI PENJUALAN SEPEDA MOTOR TERLARIS PADA PT DAYA ANUGRAH MANDIRI

Rozimin¹
Rahmat Fauzi²

¹Program Studi Teknik Informatika Universitas Putera Batam

² Dosen Program Studi Teknik Informatika Universitas Putera Batam

email: pb160210207@upbatam.ac.id

ABSTRACT

Motorcycles are an option for everyone as a means of transportation because they are cheap and can be used for a long time. The increase in the price of consumer goods or daily necessities such as the increase in fuel prices, even on religious holidays (eid), or often Also known as fluctuation (seasonal), the problem that often occurs in making sales plans is when sales predictions are made. If the company is too small, the company will run out of goods and vice versa. The purpose of this research is that it is hoped that this research can help companies to make decisions in providing stock. Processing of motorcycle sales data as much as 170 and 3 attributes that exist in the data selection using the K-Nearest Neighbor algorithm method produces predictions of sales of Honda motorcycles with the metric type more in demand by consumers than the sport and CUB types. This study uses the K-Nearest Neighbord algorithm with an accuracy value of 97.65%.

Keywords: K-Nearest Neighbord algorithm; Data Mining; Prediction.

PENDAHULUAN

Sepeda motor menjadi sebuah pilihan bagi setiap orang sebagai alat transportasi dikarenakan harga yang murah dan dapat digunakan dalam waktu yang cukup lama. di indonesia industri sepeda motor memiliki respon yang terbilang positif dan faktor yang mempengaruhinya yaitu kemudahan dalam memperoleh sepeda motor dengan cara pembayaran melalui kredit ataupun Kes/kontan.

Honda adalah salah satu perusahaan terbesar yang berada di indonesia yang bergerak dalam bidang transportasi roda dua yang cukup terkenal karena harga yang di tawarkan

murah, irit bahan bakar dan tentunya kualitas dari produk honda itu sendiri.

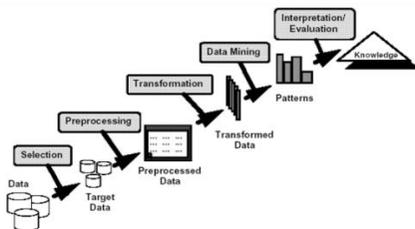
Permasalahan yang sering terjadi dalam membuat rencana penjualan yaitu apabila prediksi penjualan yang dibuat perusahaan terlalu kecil maka perusahaan akan habis ketersediaan barang. biaya produksi akan naik jika prediksi penjualan dibuat terlalu besar, sehingga investasi akan menjadi tidak tepat. permasalahan yang sering juga terjadi pada PT Daya Anugrah Mandiri dalam penjualan motor diantaranya adalah kenaikan harga barang konsumsi atau kebutuhan sehari-hari seperti kenaikan

harga bahan bakar minyak dan bahkan pada hari-hari besar keagamaan (lebaran), atau sering disebut karena faktor *fluktuasi* (musiman), ini sangat-sangat mempengaruhi daya beli masyarakat khususnya pada penjualan sepeda motor.

PT Daya Anugrah Mandiri membutuhkan prediksi atau peramalan penjualan untuk mengetahui hasil penjualan sepeda motor beberapa tahun ke depan berdasarkan data penjualan yang dimiliki dari beberapa tahun sebelumnya. analisis ini dibutuhkan sebagai dasar pengambilan keputusan oleh perusahaan agar tidak terjadi penyuplaian barang secara berlebihan.

2.1. Knowledge Discovery in Database (KDD)

Data mining merupakan sebuah tahapan proses dari rangkaian *Knowledge Discovery in Database (KDD)* yang berhubungan dengan penemuan teknik integrasi, ilmiah, interpretasi serta visualisasi pola dari jumlah data.



Gambar 2.1 Tahapan *Knowledge Discovery in Database*
Sumber: (Peneliti 2022)

2.2. Data Mining

Data Mining sendiri sering disebut juga sebagai (KDD) *Knowledge Discovery in Database* merupakan kegiatan yang meliputi pengumpulan data historis yang digunakan untuk menemukan keteraturan pada hubungan pola dalam data set dengan berukuran yang besar (Handoko, 2016).

2.3. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) adalah salah satu metode data mining yang masuk dalam algoritma klasifikasi. menurut (Harrington, 2012), algoritma K-

NN memiliki beberapa kelebihan, di antaranya akurasi tinggi, *insentif* terhadap *outlier* dan dugaan terhadap data tidak ada. akan tetapi algoritma ini juga mempunyai kekurangan di antaranya yaitu menentukan optimal nilai K, membutuhkan banyak memori dan komputasi yang mahal. di bawah ini adalah rumus *Euclidean distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Rumus 3.4.1. *Euclidean distance*.

di mana:

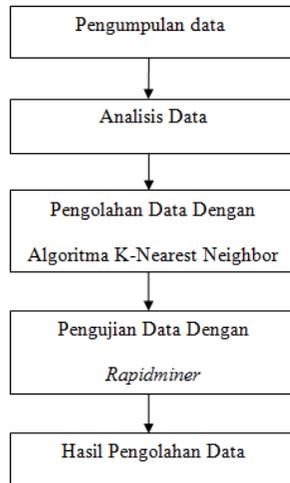
- 1) $d(x, y)$ adalah antara jarak data x dan data y .
- 2) x_i yaitu data testing ke- i
- 3) y_i yaitu data training ke- i

Berikut ini adalah langkah-langkah algoritma K-NN:

- 1) menentukan parameter nilai K (dimana nilai K akan dipilih dengan manual).
- 2) untuk menghitung antara jarak data testing dan data training
- 3) mengurutkan data training dengan berdasarkan nilai jarak yang terkecil.
- 4) untuk menetapkan kelas, di mana data testing akan dipilih berdasarkan nilai k dengan jumlah terbanyak.

METODE PENELITIAN

Penelitian ini dilakukan pada salah satu diler motor yang ada di batam, lebih tepatnya di Jl. Tiban IV, Tiban Riau Bertuah Tahap 1 Blok A No 4. desain penelitian merupakan tahapan atau proses yang digunakan untuk merencanakan suatu penelitian. desain penelitian ini bertujuan untuk mempermudah peneliti dalam melakukan penelitian. Berikut ini adalah langkah-langkah dalam desain penelitian:



Gambar 3.1 Desain Penelitian.
Sumber: (Peneliti, 2022)

Di bawah ini adalah penjelasan tahapan dari desain penelitian di atas, sebagai berikut.

1. Mengumpulkan Data Pengumpulan data yang dilakukan oleh peneliti menggunakan beberapa teknik yaitu, observasi langsung ke PT Daya Anugrah Mandiri dan studi pustaka.
2. Analisis Data Pada tahapan ini data penjualan sepeda motor yang diperoleh dari PT Daya Anugrah Mandiri dari tahun 2018, 2019 sampai dengan 2020 akan dianalisis menggunakan operasional variabel yaitu harga dan tipe sepeda motor untuk dijadikan data masukan dan diolah untuk mendapatkan hasil dari data penjualan sepeda motor.
3. Pengolahan data dengan Algoritma *K-Nearest Neighbor* Tahapan berikutnya adalah pengolahan data dan data yang telah diperoleh dari analisa dan diolah kedalam metode Algoritma *K-Nearest Neighbor* untuk melakukan klasifikasi data pada kelas tertentu.
4. Pengujian data dengan *Rapidminer* Setelah data diolah menggunakan metode *K-Nearest Neighbor* kemudian

data diuji menggunakan *software Rapidminer* untuk mendapatkan hasil yang selaras dengan perhitungan secara manual.

5. Hasil, Setelah pengujian dilakukan menggunakan *software Rapidminer* maka akan mendapatkan hasil dan ini merupakan akhir dari proses penelitian yang dilakukan oleh peneliti.

HASIL DAN PEMBAHASAN

4.1 Analisis Data

Untuk penelitian ini, data yang digunakan adalah penjualan data produk sepeda motor 3 tahun terakhir yaitu dari tahun 2018, 2019 dan 2020 dari PT Daya Anugrah Mandiri. data tersebut berupa data mentah yang berjumlah 1.043 yang kemudian akan diseleksi dan diolah untuk memprediksi penjualan produk terlaris dengan menggunakan aribut nama motor, tipe, kuantiti dan bulan.

Tabel 1. Data *Training* Penjualan Vario

| No | Bulan 1 | Bulan 2 | Bulan 3 | Bulan 4 | Bulan 5 | Target |
|----|---------|---------|---------|---------|---------|--------|
| 1 | 1 | 5 | 7 | 6 | 5 | 5 |
| 2 | 5 | 7 | 6 | 5 | 5 | 11 |
| 3 | 7 | 6 | 5 | 5 | 11 | 5 |
| 4 | 6 | 5 | 5 | 11 | 5 | 3 |
| 5 | 5 | 5 | 11 | 5 | 3 | 5 |
| 6 | 5 | 11 | 5 | 3 | 5 | 8 |
| 7 | 11 | 5 | 3 | 5 | 8 | 4 |
| 8 | 5 | 3 | 5 | 8 | 4 | 5 |
| 9 | 3 | 5 | 8 | 4 | 5 | 1 |
| 10 | 5 | 8 | 4 | 5 | 1 | 2 |
| 11 | 8 | 4 | 5 | 1 | 2 | 0 |
| 12 | 4 | 5 | 1 | 2 | 0 | 4 |
| 13 | 5 | 1 | 2 | 0 | 4 | 3 |
| 14 | 1 | 2 | 0 | 4 | 3 | 3 |
| 15 | 2 | 0 | 4 | 3 | 3 | 2 |
| 16 | 0 | 4 | 3 | 3 | 2 | 4 |
| 17 | 4 | 3 | 3 | 2 | 4 | 8 |
| 18 | 3 | 3 | 2 | 4 | 8 | 2 |
| 19 | 3 | 2 | 4 | 8 | 2 | 6 |

(Sumber : Data Penelitian, 2022)

Pada tabel 1 merupakan data latih (*training*) yang digunakan sebagai penelitian dan hanya mengambil data 2 tahun sebelumnya yaitu pada data penjualan tahun 2018 dan 2019. data uji (*training*) terbagi menjadi dua kelompok yaitu data masukan (*input*) dan data target, yang mana data masukan (*input*) adalah data penjualan pada bulan ke satu sampai dengan bulan ke lima sedangkan bulan ke 6 digunakan untuk data target.

Pada tabel 2 data yang digunakan sebagai data *testing* (uji) merupakan data penjualan sepeda motor dari bulan januari 2020 sampai dengan bulan desember 2020 untuk

mendapatkan hasil prediksi penjualan pada tahun 2021.

Tabel 2. Data *Testing* Penjualan Honda Vario

| Bulan 1 | Bulan 2 | Bulan 3 | Bulan 4 | Bulan 5 |
|---------|---------|---------|---------|---------|
| 4 | 8 | 2 | 6 | 5 |
| 8 | 2 | 6 | 5 | 8 |
| 2 | 6 | 5 | 8 | 5 |
| 6 | 5 | 8 | 5 | 4 |
| 5 | 8 | 5 | 4 | 14 |
| 8 | 5 | 4 | 14 | 7 |
| 5 | 4 | 14 | 7 | 10 |
| 4 | 14 | 7 | 10 | 4 |
| 14 | 7 | 10 | 4 | 11 |
| 7 | 10 | 4 | 11 | 5 |
| 10 | 4 | 11 | 5 | 5 |
| 4 | 11 | 5 | 5 | 5 |

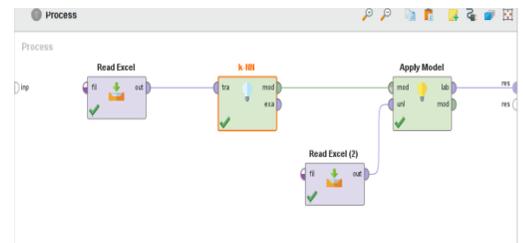
(Sumber : Data Penelitian, 2022)

Pada tabel 2 data yang digunakan sebagai data *testing* (uji) merupakan data penjualan sepeda motor dari bulan januari tahun 2020 sampai dengan bulan desember tahun 2020 untuk mendapatkan hasil prediksi penjualan pada tahun 2021.

Mengurutkan data dari jarak yang terdekat sampai jarak yang terjauh. maka didapatkan sebagai berikut: $d_9 = 2,64575$, $d_1 = 4,47213$, $d_2 = 5,65685$, $d_{17} = 5,83095$, $d_5 = 6,08276$, $d_{15} = 6,48074$, $d_{16} = 6,78232$, $d_{18} = 7,61577$, $d_8 = 7,68114$, $d_{11} = 7,87400$, $d_{13} = 7,87400$, $d_6 = 8,30662$, $d_3 = 8,48528$, $d_{19} = 8,54400$, $d_{10} = 8,66025$, $d_{14} = 9,05538$, $d_{12} = 9,48683$, $d_4 = 10,39230$, $d_7 = 10,95445$.

Menentukan kelompok dari data hasil uji dengan berdasarkan label yang terbanyak dari nilai K dengan tetangga terdekat, sebab nilai K adalah 3 maka dapat diambil dengan 3 jarak yang terkecil yaitu: d_9 , d_1 serta d_2 .

Pengujian algoritma K-Nearest Neighbor untuk data penjualan motor menggunakan *software RapidMiner* versi 9.9. Read excel 1 sebagai data *training* (latih) dan read excel 2 sebagai data (uji). dapat dilihat pada gambar 1.



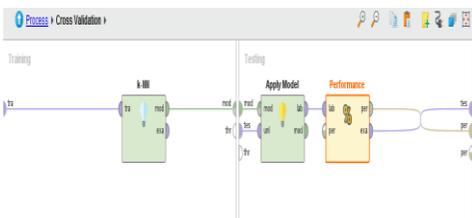
Gambar 1. Susunan Operator K-NN (Sumber : Peneliti 2022)

Setelah kesemua operator terhubung antara satu sama lain selanjutnya klik *icon play* (F11) yang berbentuk segi tiga dan berwarna biru maka akan muncul tab *result history* yang berisikan sebuah data prediksi dari semua data yang memenuhi nilai $K=3$. maka didapatkan prediksi penjualan seperti gambar

| Row No. | prediction(T... | Bulan-1 | Bulan-2 | Bulan-3 | Bulan-4 | Bulan-5 |
|---------|-----------------|---------|---------|---------|---------|---------|
| 1 | 7.025 | 4 | 8 | 2 | 6 | 5 |
| 2 | 4.658 | 8 | 2 | 6 | 5 | 8 |
| 3 | 6.881 | 2 | 6 | 5 | 8 | 5 |
| 4 | 5.709 | 6 | 5 | 8 | 5 | 4 |
| 5 | 5.837 | 5 | 8 | 5 | 4 | 14 |
| 6 | 3.920 | 8 | 5 | 4 | 14 | 7 |
| 7 | 3.661 | 5 | 4 | 14 | 7 | 10 |
| 8 | 7.032 | 4 | 14 | 7 | 10 | 4 |
| 9 | 6.441 | 14 | 7 | 10 | 4 | 11 |
| 10 | 5.251 | 7 | 10 | 4 | 11 | 5 |
| 11 | 5.639 | 10 | 4 | 11 | 5 | 5 |
| 12 | 7.290 | 4 | 11 | 5 | 5 | 5 |

Gambar 2. Prediksi Penjualan Honda Vario
(Sumber : Peneliti 2022)

Gambar di atas merupakan hasil prediksi penjualan motor vario di mana penjualan terendah berada pada bulan juni dan juli dengan nilai prediksi 3 dan penjualan tertinggi terdapat pada bulan Januari, gustus dan desember dengan prediksi nilai 7.



Gambar 3. Cross Validason
(Sumber : Peneliti 2022)

Gambar di atas merupakan Cross Validason untuk mencari nilai akurasi terhadap data.

● Table View ○ Plot View

accuracy: 97.65% +/- 4.11% (micro average: 97.65%)

| | true Matic | true Sport | true CUB |
|--------------|------------|------------|----------|
| pred. Matic | 100 | 1 | 0 |
| pred. Sport | 1 | 50 | 2 |
| pred. CUB | 0 | 0 | 16 |
| class recall | 99.01% | 98.04% | 88.89% |

Gambar 4. Performance.
(Sumber : Peneliti 2022)

Gambar di atas adalah gambar akurasi data yang didapatkan dengan nilai 97,65%.

Kesimpulan

Pada bab sebelumnya telah dibahas mengenai pengolahan dan analisis pada data maka dari itu dapat diambil beberapa kesimpulan diantaranya:

1. Pada perhitungan manual dengan menggunakan rumus *Euclidean distance* yang diambil dengan nilai 3 tetangga terdekat untuk mengetahui jarak antara data *training* dan *testing* maka menghasilkan model prediksi tertinggi yang di dapat adalah honda beat dengan nilai prediksi 22.
2. Pengolahan data penjualan motor sebanyak 170 dan 3 atribut yang ada pada data *selection* menggunakan metode algoritma *K-Nearest Neighbor* menghasilkan prediksi penjualan sepeda motor honda dengan *type metic* lebih diminati oleh konsumen dibandingkan dengan *type sport* dan CUB.
3. Pengujian yang dilakukan dengan menggunakan *software* atau perangkat lunak *Rapidminer* cukup efektif karena hasil yang didapatkan sama halnya dengan perhitungan manual. karena sebab itu, pengujian dengan *Rapidminer* cukup membantu untuk melakukan pengujian.
4. Penelitian ini dilakukan dengan menggunakan algoritma *k-nearest neighbor* berdasarkan perhitungan menggunakan data mining dengan teknik klasifikasi maka didapatkan hasil prediksi penjualan sepeda motor honda dengan *type metic, sport* dan *type CUB* dengan nilai akurasi terhadap klasifikasi penjualan sebesar 97,65

DAFTAR PUSTAKA

Enriko, Suryanegara, M., & Gunawan, D. (1843). Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient ' s Health Parameters. 8(12).

Esme, & Kiran, M. S. (2018). Prediction of Football Match Outcomes Based on Bookmaker Odds by Using k-Nearest Neighbor Algorithm. 8(1). <https://doi.org/10.18178/ijmlc.2018.8.1.658>

Ghaeini-hessaroeiyeh, Kekeringan, P., Standar, I. P., Terdekat, M. T., Model, E., & Farahmand, C. (2017). Itu standar Kemudian itu probabilitas kumulatif diubah menjadi a Z-. 5(2), 319–325.

Issn. (2018). PENGELOMPOKKAN DATA MINING PADA JUMLAH PENUMPANG DI BANDARA HANG NADIM Koko Handoko , Lido Sabda Lesmana. 02.

Penerima, Mahasiswa, B., & Stmik, D. I. (n.d.). PENERAPAN METODE K-NEAREST NEIGHBOR PADA APLIKASI. 19–25.

Print, Wajhillah, R., Ubaidallah, I. H., & Bahri, S. (2019). InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan Analisis Kelayakan Kredit Berbasis Algoritma K-Nearst Neighboar (Studi Kasus : Koperasi AKU). 1, 0–4.

Suntoro. (2019). DATA MINING Algoritma dan Implementasi dengan Pemrograman PHP (p. 180). PT Elex Media KOMputindo. <https://doi.org/978-602-9881-2>

Vulandari, Retno. (2017). DATA MINING Teori dan Aplikasi Rapidminer (p. 124). Gava Media.

Worong, Sompie, S. R. U. A., Jacobus, A., Elektro, T., Sam, U., Manado, R., & Manado, J. K. B. (2020). Implementasi K-Means dan K-NN pada Pengklasifikasian Citra Bunga. 15(3), 217–222.

| | |
|---|---|
|  | <p>Rozimin adalah mahasiswa Prodi Teknik Informatika Universitas Putera Batam.</p> |
|  | <p>Rahmat Fauzi merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam, Penulis banyak berkecimpung di bidang Teknologi Informasi.</p> |