

## PREDIKSI PENJUALAN SEPATU MENGGUNAKAN METODE K- NEAREST NEIGHBOR

**Bagus Hardiyanto<sup>1)</sup>, Fahrur Rozi<sup>2)</sup>,**

<sup>1, 2)</sup>Pendidikan Teknologi Informasi dan Universitas Bhinneka PGRI Tulungagung  
Jalan Mayor Sujadi Timur No 7 Tulungagung, 66221  
e-mail: [bagushrd12@gmail.com](mailto:bagushrd12@gmail.com)<sup>1)</sup>, [rozi.fahrur04@gmail.com](mailto:rozi.fahrur04@gmail.com)<sup>2)</sup>

### ABSTRAK

*Prediksi penjualan sepatu untuk klasifikasi potensi pelanggan baru di toko Obral Murah dengan menggunakan metode K-Nearest Neighbor. Agar toko Obral Murah tetap menjadi toko favorit dan tidak kalah dengan pesaing pesaing baru, untuk menghindari hal tersebut maka perlu adanya prediksi penjualan untuk melihat potensi dari para pelanggan dan barang yang disukai pelanggan. Tujuan Penelitian ini dilaksanakan di sebuah toko sepatu yang bernama toko Obral Murah yang beralamat di barat lampu merah gragalan 200 meter selatan jalan Kec.Kedungwaru Kab.Tulungagung dengan penjualan sepatu. Data yang digunakan adalah seluruh data pada penjualan sepatu dari tahun 2018-2019.*

*Penelitian ini dibutuhkan studi literatur untuk dapat menyelesaikan masalah. Teori-teori mengenai prediksi potensi pelanggan baru serta metode KNN yang digunakan sebagai dasar penelitian diperoleh dari sumber-sumber atau buku- buku referensi. Sumber data adalah pembeli dengan jumlah 99 pembeli dengan perhitungan KNN. Metode yang digunakan adalah metode wawancara dan survey.*

*Untuk menggunakan training data yang berjumlah 99 data dengan menguji testing data berjumlah 25 data, maka didapatkan nilai k yang terbaik untuk memprediksi masa studi mahasiswa yaitu sebagai berikut: Untuk pembeli ke 100 yaitu nilai k yang terbaik untuk digunakan memprediksi jenis sepatu yang akan di beli adalah nilai k = 9 dengan tingkat keberhasilan 49%. Untuk pembeli ke 100 yaitu nilai k yang terbaik untuk digunakan memprediksi pembelian jenis sepatu adalah nilai k = 7 dengan tingkat keberhasilan 74%. Untuk pembeli ke 100 yaitu nilai k = 1 merupakan nilai k yang terbaik untuk digunakan memprediksi masa studi mahasiswa dengan tingkat keberhasilan 89%.*

*Kata Kunci: KNN, Prediksi, Sepatu,*

### ABSTRACT

*Shoe sales prediction for the classification of potential new customers at the Cheap Sale store using the K-Nearest Neighbor method. So that the Cheap Sale store remains a favorite store and is not inferior to new competitors, in order to avoid this, it is necessary to have sales predictions to see the potential of the customers and the goods the customers like. The purpose of this study was carried out in a shoe store called Cheap Sale which is located at the west of the red light gragalan 200 meters south of the road Kec.Kedungwaru Kab. Tulungagung with shoe sales. The data used are all data on shoe sales from 2018-2019.*

*Data sources are 99 buyers with the calculation of KNN. The method used is the interview and survey methods. The first step I came to the store with coordination first through whatsapp then then I came and saw various brands of shoes and sandals sold in the store and conducted interviews at the store with various questions and goals that I would use as research material. To use training data which amounted to 99 data by testing the testing data amounting to 25 data, then the best k value is obtained to predict the study period of students, as follows: For the 100th buyer, the best k value to be used to predict the type of shoes to buy is k = 9 with a success rate of 49%. For the 100th buyer, the best k value to use is to predict the purchase of a shoe type is a k = 7 value with a 74% success rate. For the 100th buyer, the value of k = 1 is the best k value to be used to predict the period of study of students with a success rate of 89%.*

*Keywords: KNN, Prediction, Shoes,*

## I. PENDAHULUAN

**A**danya prediksi ini maka toko Obral Murah dapat melakukan pengambilan keputusan yang tepat dalam produksinya. Selama ini prediksi penjualan sepatu ditoko Obral Murah masih mengandalkan metode hafalan dan voute dari setiap toko cabang dan hasil data penjualan pada sebelumnya. Semakin banyak pesaing perusahaan bermunculan mengharuskan toko Obral Murah untuk merencanakan strategi pejualan. Agar toko Obral Murah tetap menjadi toko favorit dan tidak kalah dengan pesaing pesaing baru, untuk menghindari hal tersebut maka perlu adanya prediksi penjualan untuk melihat potensi dari para pelanggan dan



barang yang disukai pelanggan. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban dkk. 2005). Selain itu juga ada metode *K-Nearest Neighbor* (KNN), metode *K-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data contoh yang mempunyai jarak paling dekat dengan objek tersebut. Algoritma ini hanya melakukan penyimpanan dan klasifikasi data contoh. Algoritma *K-Nearest Neighbor* memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data dengan jumlah yang cukup besar.

Agar pembahasan dalam penelitian ini bisa dapat mudah dipahami dan di mengerti maka penulis memberikan batasan masalah yang akan di teliti seperti :

1. Penelitian ini dilaksanakan di sebuah toko sepatu yang bernama toko Obral Murah yang beralamat di barat lampu merah gragalan 200 meter selatan jalan Kec.Kedungwaru Kab.Tulungagung dengan penjualan sepatu.
2. Data yang digunakan adalah seluruh data pada penjualan sepatu dari tahun 2018-2019.

## II. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam penelitian ini adalah Research and Development (RND) Penelitian ini dibutuhkan studi literatur untuk dapat menyelesaikan masalah. Teori-teori mengenai prediksi potensi pelanggan baru serta metode KNN yang digunakan sebagai dasar penelitian diperoleh dari sumber-sumber atau buku- buku referensi, dan melakukan pencarian dari internet. Setelah itu data yang telah dilakukan analisis, agar dapat di implementasikan ke dalam sebuah program. Penelitian ini merupakan penelitian teknik dengan mengacu pada pengolahan data pembelian sepatu yang sesuai dengan kebutuhan untuk menstok sepatu di Toko sepatu Obral Tulungagung. Tahapan yang dilakukan adalah memberikan masukan data setiap jenis pakan kucing disesuaikan dengan usia pembeli, kebutuhan pembeli dan kegunaan sepatu.

### a. Identifikasi Masalah

Sebelum memulai pekerjaan yang bersifat teknis, sangat diperlukan adanya komunikasi dengan *customer* demi memahami dan mencapai tujuan yang ingin dicapai. Hasil dari komunikasi tersebut adalah inisialisasi proyek, seperti menganalisis permasalahan yang dihadapi dan mengumpulkan data-data yang diperlukan, serta membantu mendefinisikan fitur dan fungsi *software*. Pengumpulan data-data tambahan bisa juga diambil dari jurnal, artikel, dan internet (Pressman, 2015).

Analisis kebutuhan *user* merupakan gambaran layanan yang diinginkan oleh pengguna terhadap aplikasi yang akan dihasilkan, antara lain:

- 1) Aplikasi dapat membantu menerjemahkan kata dengan baik.  
Pengembangan media aplikasi pengambil keputusan dengan pengolahan *data mining* pada Toko Obral Sepatu Tulungagung mengimplementasikan kebutuhan stok sepatu
- 2) Penggunaan aplikasi dapat dipahami oleh pengguna (*user friendly*)  
Pengguna pada penelitian ini adalah administrasi dan cashier Toko Obral Sepatu Tulungagung yang akan di sesuaikan dengan kemampuan administrasi dan cashier secara cara luas agar dapat dimanfaatkan secara baik.
- 3) Tampilan *interface* aplikasi menarik.  
Kebutuhan sistem adalah proses yang mampu dilakukan oleh sistem ataupun aplikasi, diantaranya proses penginputan kata, proses menerjemahkan kata, dan menampilkan hasil terjemahan. Pembuatan jadwal pengerjaan sistem bertujuan untuk mengatur agar proses pada setiap tahap pembuatan aplikasi berjalan dengan baik dan terstruktur. Selain itu juga, perlu dilakukan identifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang mampu digunakan untuk menerapkan aplikasi yang akan dibuat.

### b. Penentuan Hardware dan Software yang digunakan

Tahap berikutnya adalah tahapan perencanaan yang menjelaskan tentang estimasi tugas-tugas teknis yang akan dilakukan, resiko-resiko yang dapat terjadi, sumber daya yang diperlukan dalam membuat sistem, produk kerja yang ingin dihasilkan, penjadwalan kerja yang akan dilaksanakan, dan *tracking* proses pengerjaan sistem. Pada tahapan ini peneliti melakukan:

- 1) Menentukan Komponen-komponen *hardware* dan *software* yang mendukung teknologi *Rapid Miner*.
- 2) Spesifikasi yang harus dimiliki oleh perangkat keras yang digunakan untuk menyelesaikan penelitian ini berupa Personal Computer dengan spesifikasi:
  - a) CPU: AMD FX-6300 2,30 Ghz

- b) VGA: Nvidia GT 9600 2GB
- c) RAM 8 GB

## 1. Teknik Data Mining

Tahapan ini adalah tahap perancangan dan permodelan arsitektur sistem yang berfokus pada perancangan struktur data, arsitektur *software*, tampilan *interface*, dan algoritma program. Tujuannya untuk lebih memahami gambaran besar dari apa yang akan dikerjakan.

Algoritma k-NN bekerja berdasarkan jarak terdekat dari query instance ke training data untuk menentukan k-NN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode euclidian distance. Euclidian Distance sering digunakan untuk menghitung jarak. Euclidian Distance berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek, di bawah ini merupakan rumus Euclidian Distance:

$$\left( \sum_{k=1}^m (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}}$$

Dimana,

X<sub>ik</sub> = nilai X pada training data

X<sub>jk</sub> = nilai X pada testing data

m = batas jumlah banyaknya data

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai k yang terbaik itu tergantung pada jumlah data. Ukuran nilai k yang besar belum tentu menjadi nilai k yang terbaik begitupun juga sebaliknya.

validasi menggunakan cross validation dengan jumlah (k) dalam pengujian adalah 2, 5, 7, 10 dan serta pengujian menggunakan split validation dengan 4 kali percobaan dengan menggunakan data training 60%, 70%, 80%, dan 90%. akan dilakukan proses pengujian *cross validation* dan *split validation* dalam rapidMiner, yang dengan melakukan permodelan dengan metode *decision Tree* KNN, selanjutnya memasukkan *apply model* dan *performa.Validation* merupakan proses untuk mengevaluasi keakuratan prediksi dari model. Teknik validasi membagi data secara acak ke dalam bagian disebut dengan *cross validation*. Dalam *cross validation* jumlah lipatan atau partisi dari data dilakukan sendiri, Pada proses *cross validation* menghasilkan *confusion matrix* dan akurasi dari model .

## 2. Memasukkan Data Mining

Pada tahap ini, rancangan arsitektur sistem dan desain sistem yang telah dibuat kemudian diterjemahkan ke dalam kode bahasa pemrograman. Tujuan penjadwalan ini agar penelitian dapat berjalan dengan efektif dan selesai tepat waktu, sehingga diharapkan produk yang dihasilkan dari penelitian dapat segera selesai dan memberikan alternatif solusi atas permasalahan yang sudah di analisis.

### a. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi.

- 1) Ditulis dengan bahasa pemrograman *Java* sehingga dapat dijalankan di berbagai sistem operasi.
- 2) Proses penemuan pengetahuan dimodelkan sebagai *operator trees*
- 3) Representasi XML internal untuk memastikan format standar pertukaran data.
- 4) eksperimen.
- 5) Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
- 6) Memiliki GUI, *command line mode*, dan *Java API* yang dapat dipanggil dari program lain.

*Tools* yang digunakan dalam RapidMiner ini adalah:

- 1) *Panel operator view*
- 2) *Panel operator view* merupakan induk langkah kerja dari proses analisa
- 3) *Panel repository view*
- 4) merupakan komponen utama dalam *Design Perspective* selain *Operator View*. *View* ini dapat digunakan untuk mengelola dan menata proses Analisis menjadi proyek dan pada saat yang sama juga dapat digunakan sebagai sumber data dan yang berkaitan dengan meta data.
- 5) *Process view*
- 6) *Process View* menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung



langkah-langkah tersebut.

- 7) *Parameter view*
- 8) Digunakan untuk mengatur fungsionalitas dari beberapa operator yang membutuhkannya agar dapat dijalankan sesuai dengan proses analisa.

### 3. Pengujian Data

Tahapan *Deployment* merupakan tahapan implementasi *software* ke *customer*, pemeliharaan *software* secara berkala, perbaikan *software*, evaluasi *software*, dan pengembangan *software* berdasarkan umpan balik yang diberikan agar sistem dapat tetap berjalan dan berkembang sesuai dengan fungsinya. (Pressman, 2015:17)

#### Keterangan :

Entropy (s)

S = Himpunan Kasus

A = Atribut

N= Jumlah Partisi S

|Si| = jumlah Jumlah kasus pada partisi ke i

pi = Proposi dari Si terhadap S

Identifikasi sampel dari *data set* baca data. Kemudian menghitung *entropy (S)* dari keseluruhan atribut, kemudian menghitung *gain* tertinggi dari seluruh atribut, selanjutnya didapatkan atribut yang akan digunakan sebagai akar/ *node*. buat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang, ulangi perhitungan *Gain* sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak lagi diikuti dalam perhitungan.

### 4. Ujicoba Produk

Untuk mengetahui tingkat keakurasian potensi pelanggan baru, maka harus dibuat rancangan uji coba yang menggunakan data latih berdasarkan nilai keakurasian menggunakan rancangan uji coba. Ketika nilai *K* beserta keakurasiannya sudah terjawab maka selanjutnya adalah menghitung tingkat akurasi data dengan nilai *K* tersebut.

## III. HASIL PENELITIAN

Semua atribut yang ada pada data penjualan diantaranya field Kode barang, Nama barang, Kuant, Harga, Jumlah dan Bulan. Dari semua data yang ada akan diseleksi dan digunakan hanya 3 field yang digunakan untuk proses knowledge discovery in database (KDD). Field tersebut yaitu:

- a. Nama Barang merupakan atribut yang terdapat pada tabel data penjualan yang berisi informasi tentang nama barang yang di beli di Toko Obral Murah Tulungagung.

Tabel Nama barang Toko Obral

no	Nama Barang	Pengguna	Penggunaan
1	Fleatshoes	Usia sekolah	Sekolah
2	Kickkers	Usia sekolah	Sekolah
3	Adidas Pelon	Usia Kerja	Santai
4	Pantofel Hogy	Usia Kerja	Kerja

Pengklasifikasian tersebut digunakan untuk mempermudah peneliti dalam membuat atribut sebagai dasar perhitungan data mining dengan KNN.

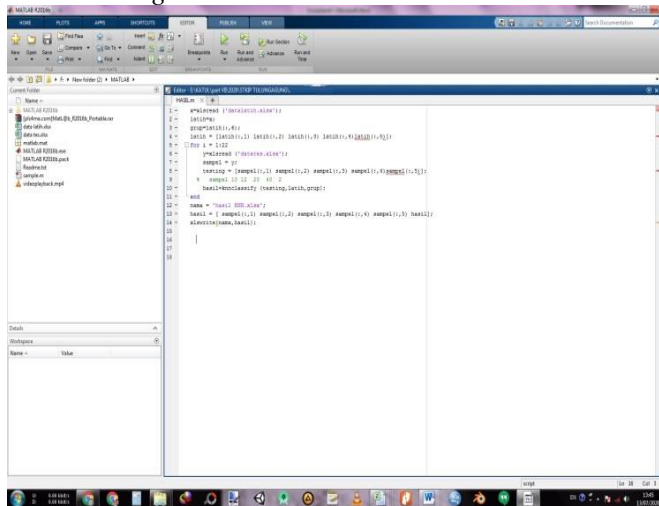
**Tabel Informasi Jumlah Transaksi 2018-2019**

no	Nama Barang	Pembelian
1	Fleatshoes	126
2	Kickkers	195
3	Adidas Pelon	132
4	Pantofel Hogy	110

Tahap ini merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu berdasarkan proses KDD secara keseluruhan. Metode yang digunakan pada penelitian ini adalah metode K-Nearest Neighbor (KNN) dimana metode ini memiliki atribut yang diinisialisasikan sebagai  $k$ , yaitu jumlah tetangga yang dijadikan acuan pada KNN, nilai  $k$  adalah bilangan bulat positif, berjumlah kecil dan ganjil.

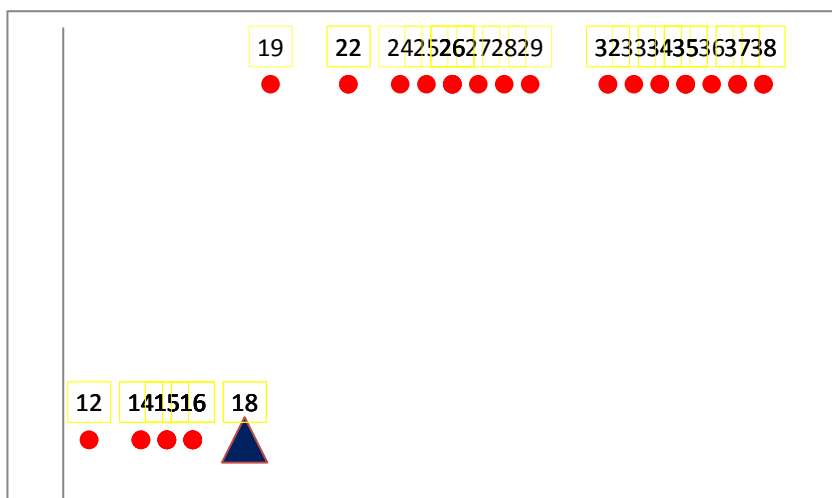
a. Menentukan nilai  $K$  dengan menggunakan Matlab.

1) Menentukan coding



2) Hasil perhitungan

Hasil perhitungan diketahui berupa perhitungan hasil analisis dari data latih sebanyak 40 pembeli dan data tes sebanyak 20 pembeli. Mengacu pada hasil di atas diketahui dari perhitungan excel dan dikuatkan perhitungan dengan menggunakan aplikasi Matlab didapatkan hasil sebagai berikut:



#### IV. KESIMPULAN

Berdasarkan pembahasan dari bab-bab sebelumnya, maka penelitian diambil beberapa simpulan: Proses Prediksi Pembelian sepatu di toko obral Tulungagung menggunakan algoritma k-Nearest Neighbor yang akan diterapkan di Jurusan Teknik Informatika. Untuk menggunakan training data yang berjumlah 99 data dengan menguji testing data berjumlah 25 data, maka didapatkan nilai k yang terbaik untuk memprediksi masa studi mahasiswa yaitu sebagai berikut: Untuk pembeli ke 100 yaitu nilai k yang terbaik untuk digunakan memprediksi jenis sepatu yang akan di beli adalah nilai  $k = 9$  dengan tingkat keberhasilan 49%. Untuk pembeli ke 100 yaitu nilai k yang terbaik untuk digunakan memprediksi pembelian jenis sepatu adalah nilai  $k = 7$  dengan tingkat keberhasilan 74%. Untuk pembeli ke 100 yaitu nilai  $k = 1$  merupakan nilai k yang terbaik untuk digunakan memprediksi masa studi mahasiswa dengan tingkat keberhasilan 89%.

#### DAFTAR PUSTAKA

##### DAFTAR PUSTAKA

- [1]Berry, Michael J.A. dan Gordon S. Linoff. 2004. *Data mining techniques for marketing, sales, customer relationship management*. Second edition. Wiley publishing, Inc.
- [2]Bramer, Max. 2007. *Principles of Data Mining*. London : Springer Gorunescu, Florin. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg : Springer
- [3]Gambetta, Windy. 2012. Pohon Keputusan (Decision Tree). Departemen Teknik Informatika. Institute Teknologi Bandung. Bandung.
- [4]Han, J., & Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco : Morgan Kauffman.
- [5]Kusrini dan Emha Taufiq Lutfi. 2009. *Algoritma Data Mining*. Yogyakarta; Andi Offset.
- [6]Kusrini & Luthfi, E.T. 2009. *Algoritma Data Mining*. Yogyakarta : Andi Publishing.
- [7]Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey : John Willey & Sons, Inc.
- [8]Liao. 2007. *Recent Advances in Data Mining of Enterprise Data : Algorithms and Application*. Singapore : World Scientific Publishing