

Perbandingan Algoritma K-Nearest Neighbor Dengan Decision Tree Dalam Memprediksi Penjualan Makanan Hewan Peliharaan Di Petshop Dore Vet Clinic

Dina Meilida Meliala¹, Penda Hasugian²

Teknik Informatika STMIK Pelita Nusantara

Jl. Iskandar Muda No. 1 Medan 20154 INDONESIA

¹dina.meilida@gmail.com, ²penda.hasugian@gmail.com

INTISARI

Memprediksi penjualan sangat penting dalam kemajuan sebuah usaha, terutama dalam penjualan barang yang memiliki tanggal kadaluarsa seperti makanan hewan peliharaan. Ada beberapa algoritma yang digunakan untuk menginformasikan prediksi harga penjualan salah satunya algoritma K-Nearest Neighbor dan algoritma Decision Tree. Dengan metode K-nn, dihasilkan kondisi dari 30 data, 6 data diklasifikasikan terlaris sesuai dengan prediksi yang dilakukan dengan metode k-nn, 3 data dari 6 data diprediksi terlaris ternyata tidak terlaris, (data urutan 1, 2, 6). 24 data diprediksi tidak terlaris ternyata 10 data sebelumnya diklasifikasikan terlaris (data urutan 22, 5, 16, 26, 28, 19, 17, 20, 23, 24). Dengan metode decision tree algoritma C45, diketahui dari 30 data, merek purina terlaris, ada 4 data daripada royal canin (false negative). Hasil tingkat akurasi decision tree algoritma c45, diketahui true terlaris = 17, false tidak terlaris = 4. False terlaris = 16, true tidak terlaris = 13. Akurasi decision tree algoritma c45 = 83%.

Kata kunci— algoritma K-NN, Algoritma C45, Data Mining, Pohon Keputusan, Prediksi

ABSTRACT

Predicting sales is very important in the progress of a business, especially in the sale of goods that have an expiration date such as pet food. There are several algorithms used to inform sales price predictions, one of which is the K-Nearest Neighbor algorithm and the Decision Tree algorithm. With the K-nn method, conditions are generated from 30 data, 6 data are classified as best-selling according to the predictions made by the k-nn method, 3 data from 6 data are predicted to be the best-sellers in fact, (data order 1, 2, 6). The 24 data predicted not best-selling turned out to be the 10 previously classified bestsellers (data sequences 22, 5, 16, 26, 28, 19, 17, 20, 23, 24). With the C45 decision tree algorithm method, it is known that from 30 data, the best-selling purina brand, there are 4 data than royal canin (false negative). The result of the accuracy level of the decision tree algorithm is c45, it is known that best-selling true = 17, false not best-selling = 4. Best-selling false = 16, best-selling true not = 13. Accuracy of decision tree algorithm c45 = 83%.

Keywords—K-NN algorithm, C45 algorithm, Data Mining, Decision Tree, Prediction

I. PENDAHULUAN

Masa kadaluarsa terpanjang untuk makanan hewan peliharaan biasanya hanya selama 2 tahun. Sangat rentan jika sebuah usaha memiliki stok yang banyak jika usaha tersebut tidak dapat memprediksi berapa banyak makanan tersebut akan terjual.

Perilaku pembeli dalam memenuhi kebutuhan hewan peliharaannya tidak dapat ditebak tidak ada rutinitas belanja makanan sementara petshop setiap bulan melakukan pembelian barang untuk persediaan penjualan. Pembelian item makanan hewan dilakukan berdasarkan informasi merek dagang, tekstur makanan, rasa yang bervariasi, gizi dan vitamin yang terkandung didalamnya

menjadikan para pembeli tidak selalu membeli merek dagang dan varian yang sama disetiap pembeliannya tergantung kebutuhan masing-masing pembeli untuk hewan peliharaan mereka. Faktor-faktor seperti ini mengakibatkan permintaan para konsumen/pembeli tidak sesuai dengan produk yang tersedia.

K-nearest Neighbors atau KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (nearest neighbors). Dengan k merupakan banyaknya tetangga terdekat [1]. *Decision tree* atau pohon keputusan adalah alat pendukung

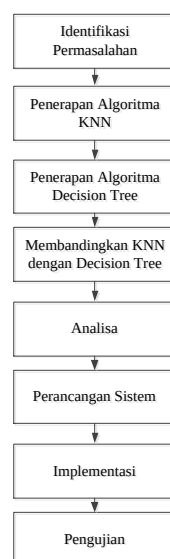
keputusan yang menggunakan model keputusan yang berbentuk seperti pohon [2]. *Decision tree* memetakan berbagai alternatif yang mungkin untuk mengatasi suatu masalah, dan terdapat juga faktor-faktor kemungkinan yang dapat mempengaruhi alternatif tersebut beserta estimasi akhirnya jika memilih alternatif yang ada [3].

Penelitian ini membandingkan penerapan algoritma *K-Nearest Neighbor* dengan algoritma *Decision Tree* dalam memprediksi penjualan jenis makanan hewan peliharaan. Tujuan melakukan perbandingan dua algoritma adalah untuk memastikan item barang yang paling laris dibeli para konsumen. Stok barang yang menumpuk mengakibatkan makanan hewan mengalami kadaluarsa sebab petshop belum dapat memprediksi item terlaris atau tidak terlaris.

Penelitian berjudul Perbandingan *K-Nearest Neighbor* dan *Decision Tree* untuk Penentuan Risiko Kredit Kepemilikan Mobil [4]. Jurnal Informatika Universitas Pamulang, menjelaskan kedua metode memiliki kekurangan dan kelebihan masing-masing dalam memprediksi suatu masalah, terutama dalam memprediksi penjualan [5].

II. METODOLOGI PENELITIAN

Metode penelitian pada penelitian ini adalah pendekatan deskriptif analitis. Penelitian deskriptif analitis merupakan suatu metode yang berfungsi untuk mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data atau sampel yang telah terkumpul sebagaimana adanya tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum. Metode penelitian, gambar 1.



Gambar 1. Metode Penelitian

Tahapan penelitian :

1. Mengidentifikasi Permasalahan

Permasalahan dalam penelitian ini adalah membangun aplikasi prediksi penjualan untuk mempermudah pemesanan barang kepada Supplier menggunakan metode Algoritma *K-Nearest Neighbor* dan *Decision Tree* dan membandingkan hasil kedua algoritma data mining.

2. Penerapan Algoritma KNN

Metode yang digunakan pada penelitian ini adalah metode *K-Nearest Neighbor* (KNN), 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 [6].

Tahapan metode KNN [7] [8]:

- 1) Penentuan nilai *k*. Penentuan nilai *k* yang digunakan dalam klasifikasi tidak memiliki aturan yang baku, namun pada penelitian ini nilai *k* yang digunakan adalah 3, 5, 7, 9, 11, 13, 15.
- 2) Perhitungan jarak antar data training dan data uji (test). Teknik perhitungan jarak yang digunakan Jarak *Euclidean Distance*.
- 3) Pengurutan data hasil perhitungan.
- 4) Menentukan kelompok data hasil uji berdasarkan label mayoritas dari *k* tetangga terdekat.

3. Penerapan Algoritma *Decision Tree*

Tahapan dalam *Knowledge Discovery in Database* (KDD), untuk menghasilkan informasi sesuai dengan urutan yang sudah ditentukan, berikut tahapan-tahapannya [9]:

1) *Data Selection*

Data hasil seleksi akan di gunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2) *Preprocessing*

Sebelum proses data mining dapat dilaksanakan perlu dilakukan tahap preprocessing, pada tahap ini akan dilakukan proses integrasi data untuk penggabungan data dari database yang berbeda, selanjutnya dilakukan data cleaning untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu mining.

3) *Integrasi Data*

Tahap ini adalah proses penggabungan data dari berbagai database yang berbeda, sehingga data tersebut saling berintegrasi. Data integrasi dilakukan

pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik.

4) Data Cleaning

Tahap ini adalah tahap awal dari proses KDD. Pada tahapan ini data yang tidak relevan, *missing value*, dan *radudant* harus dibersihkan.

5) Transformation

Tahapan Transformation merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining.

4. Penerapan algoritma C4.5 dan pohon keputusan merupakan dua model yang tidak dapat terpisahkan. Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam data mining adalah [10]:

- 1) Pilih atribut sebagai simpul akar.
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang.
- 4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.
- 5) Pemilihan atribut sebagai simpul, baik akar (root) atau simpul internal didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada.

5. Membandingkan KNN dengan algoritma *Decision Tree*, dalam penelitian ini perbandingan kedua algoritman menerapkan metode *Confusion Matrix*.

5. Analisis

Setelah pengumpulan data selesai dilakukan maka tahap berikutnya adalah melakukan analisis terhadap data.

III. HASIL DAN PEMBAHASAN

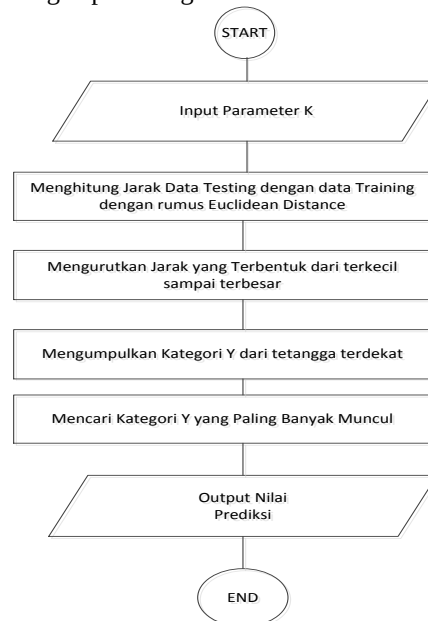
Data training dengan atribut ukuran dan harga akan diklasifikasikan tergolong terlaris atau tidak terlaris.

TABEL I.
DATA TRAINING

No	Ukuran	Harga	Keterangan
1	12	71	Tidak Terlaris
2	12	71	Tidak Terlaris
3	12	31	Tidak Terlaris
4	25	250	Tidak Terlaris
5	20	310	Tidak Terlaris
6	70	1030	Terlaris
7	20	299	Tidak Terlaris
8	20	299	Tidak Terlaris
9	15	222	Tidak Terlaris
10	20	335	Tidak Terlaris
11	20	314	Tidak Terlaris
12	15	216	Tidak Terlaris
13	15	220	Tidak Terlaris
14	10	30	Terlaris
15	25	426	Tidak Terlaris

16	20	292	Terlaris
17	1	25	Terlaris
18	2	300	Terlaris
19	15	220	Tidak Terlaris
20	0,85	21	Terlaris
21	43	50	Terlaris
22	200	1122	Terlaris
23	0,85	20	Terlaris
24	0,85	20	Terlaris
25	0,85	20	Terlaris
26	20	303	Tidak Terlaris
27	20	303	Terlaris
28	20	308	Tidak Terlaris
29	43	50	Terlaris
30	40	85	Tidak Terlaris

Bagan perhitungan K-NN:



Gambar 2. Tahapan Algoritma KNN

Parameter K sebagai banyaknya jumlah tetangga terdekat dengan objek baru. jumlah tetangga terdekat $X_1 = 55$, $Y_1 = 100$. Berdasarkan perhitungan K-NN menggunakan *Euclidean Distance*, diperoleh kategori terlaris :

1. FRISKIES ADULT MEATY GRILLS, Ukuran = 1200, Harga = 71.000.
2. FRISKIES KITTEN DISCOVERIES , Ukuran = 1200, Harga = 71.000.
3. RC DOG HYPOALLERGENIC, Ukuran = RC DOG HYPOALLERGENIC Harga = 1030000.
4. RC RENAL DOG CAN, Ukuran = 430 gr Harga = 50.000,-
5. RC GASTRO INTESTINALRC GASTRO INTESTINAL DOG DOG ukuran 430 dan harga Rp. 85.000,-
6. RC CAT FIBRE RESPON ukuran 400, harga = Rp. 85000,-.

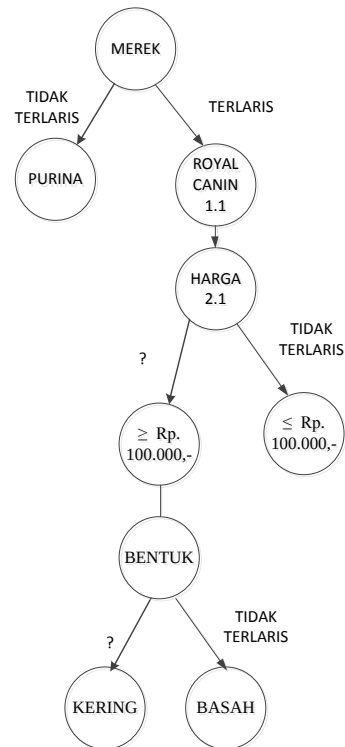
Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Pohon keputusan dapat digunakan untuk mengeksplorasi data yaitu dengan menemukan hubungan tersembunyi antara jumlah calon variable input dengan sebuah variabel target.

Tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5, mempersiapkan data training. Data yang diolah, sebanyak 30 data. Dari data yang ada, kolom yang diambil sebagai atribut/variable keputusan adalah survive. sedangkan kolom yang diambil variabel (atribut) penentuan dalam pembentukan pohon keputusan adalah Nama Makanan, Merek, Ukuran, Bentuk, Harga. Diperuntukkan, Keterangan.

Langkah awal algoritma C4.5 adalah mencari nilai entropy. Setelah semua perhitungan entropy pada masing-masing subset atribut selesai, kemudian lakukan perhitungan nilai gain. Selanjutnya perhitungan entropy pada tiap atribut berdasarkan pada jumlah kasus per subset atribut, yaitu :

- Perhitungan nilai subset pada atribut Merek Purina $[4, 0] = 0$ Royal Canin $[14, 12] = 0,151$.
- Perhitungan nilai subset pada atribut Ukuran ≥ 5000 $[1,1] = 0,150$; ≤ 5000 $[17, 11] = 0,157$.
- Perhitungan nilai subset pada atribut Bentuk Kering $[17, 4] = 0,208$; Basah $[0,9] = 0,301$.
- Perhitungan nilai subset pada atribut Harga $\leq$ Rp. 100.000,- $[13, 4] = 0,192$; $\leq$ Rp. 100.000,- $[4,9] = 0,172$.
- Perhitungan nilai subset pada atribut Diperuntukkan Hewan anak $[6, 2] = 0,188$; Hewan Anak $[11, 11] = 0,131$.

Dari hasil pada perhitungan Node 1 dapat diketahui bahwa atribut dengan Gain tertinggi adalah MEREK, sebesar 0.284. dengan demikian MEREK sebagai node akar. Ada dua nilai dari atribut MEREK yaitu Purina dan Royal Canin. Nilai atribut Merek Purina sudah mengklasifikasikan kasus menjadi satu yaitu keputusan “TERLARIS atau TIDAK TERLARIS, Gambar 3.



Gambar 3. Pohon Keputusan Node 2.1.1

TABEL II.
SUMBER DATA NODE 2.1.1

Merek	Ukuran	Bentuk	Harga Rp.	Diperuntukkan
Royal Canin	7.000	Kering	1.030.000	Dog Adult
Royal Canin	2.000	Kering	292.000	Cat Adult
Royal Canin	20.000	Kering	1.122.000	Dog Adult
Royal Canin	2.000	Kering	303.000	Kitten

Hasil perhitungan berdasarkan node 2.1.1 maka “Semua kasus sudah masuk dapat kelas”, Sehingga pohon keputusan di atas merupakan pohon keputusan terakhir yang terbentuk. Berdasarkan pohon keputusan terakhir maka didapat rule dengan aturan berikut :

- Merek PURINA Terlaris daripada Merek Royal Canin
- Bentuk Kering Terlaris daripada Bentuk Basah
- Harga $\geq$ Rp. 100.000,- Terlaris daripada harga Rp. 100.000,-
- Ukuran ≤ 5000 terlaris daripada Ukuran ≥ 5000
- Diperuntukkan untuk Hewan Dewasa terlaris daripada Anak Hewan.

Hasil akhir perhitungan memiliki gain tertinggi untuk dibuatkan node berikutnya. Gain tertinggi adalah bentuk, sebesar 0,301 dan harga sebesar 0,301.

Metode K-NN, 3 data dari 6 data diprediksi terlaris ternyata tidak terlaris, (data urutan 1, 2, 6). 24 data diprediksi tidak terlaris ternyata 10 data sebelumnya diklasifikasikan terlaris (data urutan 22, 5, 16, 26, 28, 19, 17, 20, 23, 24). Dengan metode *Decision Tree* Algoritma C45, diketahui dari 30 data, merek Purina Terlaris, ada 4 data daripada Royal Canin (false negative).

IV. KESIMPULAN

Kesimpulan dari hasil penelitian yaitu Penerapan *Decision Tree* algoritma C45 dapat dipergunakan untuk membantu memprediksi penjualan makanan hewan peliharaan di Petshop Dore Vet Clinic. Hasil Tingkat Akurasi *Decision Tree* Algoritma C45, Diketahui True terlaris = 17, False Tidak Terlaris = 4. false terlaris = 16, True Tidak Terlaris = 13. Akurasi *Decision Tree* Algoritma C45 = 83%.

REFERENSI

- [1] M. Astuti, F. D., & Guntara, "Analisis Performa Algoritma K-NN Dan C4.5 Pada Klasifikasi Data Penduduk Miskin," *Jurti*, vol. 2, no. 2, pp. 135–142, 2018.
- [2] A. S. Febriarini and E. Z. Astuti, "Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang," *Eksplora Inform.*, vol. 8, no. 2, pp. 95–103, 2019.
- [3] M. E. Lasulika, "Prediksi Harga Komoditi Jagung Menggunakan K-NN Dan Particle Swarm Optimazation Sebagai Fitur Seleksi," *Ilk. J. Ilm.*, vol. 9, no. 3, pp. 233–238, 2017.
- [4] I. Oktanisa and A. A. Supianto, "Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank a Comparison of Classification Techniques in Data Mining for," *Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 567–576, 2018.
- [5] J. S. Parapat and A. S. Sinaga, "Data Mining Algoritma C4.5 Pada Klasifikasi Kredit Koperasi Simpan Pinjam," *J. Ilmu Tek. Elektro Komput. dan Inform.*, vol. 4, no. 2, pp. 144–154, 2018.
- [6] Sanjani, H. Fahmi, and A. Sindar, "Implementasi Data Mining Penjualan Produk Pakaian Dengan Algoritma Apriori,"
- [7] Y. A. Setianto, K. Kusriani, and H. Henderi, "Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur," *Creat. Inf. Technol. J.*, vol. 5, no. 3, p. 232, 2019.
- [8] K. Siti Maesaroh, "Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining," *Unmul*, vol. 7, no. 2, pp. 25–30, 2017.
- [9] S. Al Syahdan and A. Sindar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018.
- [10] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naive Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.