Klasifikasi Kendaraan Roda Empat Berbasis KNN

Ahmad Nouvel
AMIK Bina Sarana Informatika Purwokerto
ahmad.avl@bsi.ac.id

Abstrac *- For the classification of the best car is not something easy, because the choice of each other has advantages and disadvantages of each. This paper discusses the decision to choose the best car alternative. So far, the probability of choice is determined more by the intuition and subjectivity of decision makers, who tend to be biased in view of human cognitive limitations. To solve this problem the author uses the K Nearest Neighbor (KNN) method which is proven by the Weka tool, and is applied using matlab. The results of this experiment are that the amount of data as much as 14 has an accuracy rate of 78.57% and an RMSE of 0.23, while the amount of data of 1728 has an accuracy rate of 95.78%, an RMSE of 0.19 and ROC area 0.99. Shows the greater the amount of data the higher the accuracy level.*
Keyword :KNN,RMSE,ROC,Matlab

Abstraksi -Untuk klasifikasi mobil terbaik bukanlah sesuatu yang mudah, karena pilihan satu dengan yang lainnya mempunyai kelebihan dan kekurangan masing-masing. Makalah ini membahas mengenai pengambilan keputusan untuk memilih alternatif mobil terbaik. Selama ini besar probabilitas pilihan ditentukan lebih banyak dengan intuisi dan subyektifitas pengambil keputusan, yang cenderung bias mengingat keterbatsan kognitif manusia. Untuk memecahkan masalah ini penulis menggunakan metode *K Nearest Neighbour(KNN) yang dibuktikan dengan tool weka,* dan diaplikasikasikan menggunakan matlab. Hasil dari experiment ini adalah bahwa dengan jumlah data sebanyak 14 memiliki tingkat accuracy 78,57% dan RMSE 0,23, sedangkan pada jumlah data sebanyak 1728 memiliki tingkat accuracy mencapai 95,78%, RMSE 0,19 dan ROC area 0,99. Menunjukan semakin besar jumlah data semakin tinggi tingkat accuraynya.
Kata kunci: ***KNN, RMSE,ROC,Matlab***

I. PENDAHULUAN

Memilih kendaraan yang terbaik dan teraman merupakan hal yang sangat penting, mobil adalah salah satu kendaraan yang memiliki tingkat keamanan yang memadai dibanding roda dua.

Pengembangan mobil sampai sekarang ini semakin bersaing, oleh karena itu perlu konsumen mengetahui mana mobil yang masuk dalam kategori mobil yang paling baik

Sistim pendukung keputusan yang tepat perlu kiranya ada suatu pendekatan ilmiah yang digunakan untuk memilih mobil yang ditawarkan

Penelitian ini menggunakan pendekatan klasifikasi data mining dengan metode KNN, yang diaplikasikan untuk user memakai program Matlab.

Sehingga nanti dari hasil metode KNN dapat menyimpulkan mobil mana yang masuk klasifikasi yang diinginkan.

II. TINJAUAN PUSTAKA

1. Mobil

Dalam Peraturan Pemerintah Republik Indonesia Nomor 44 Tahun 1993:

a. Kendaraan bermotor adalah setiap kendaraan yang digerakkan oleh peralatan mekanik berupa mesin selain kendaraan yang berjalan di atas rel.
b. Mobil penumpang adalah kendaraan bermotor beroda empat yang dilengkapi sebanyak-banyaknya 8 (delapan) tempat duduk, tidak termasuk tempat duduk pengemudi, baik dengan maupun tanpa perlengkapan pengangkutan bagasi.

2. Data Mining

Untuk menghasilkan informasi dan pengetahuan yang berguna dari suatu basis data yang besar diperlukan proses penggalian data yang disebut data mining sehingga ditemukan pola pola dan relasi yang tersembunyi dalam sejumlah data yang besar tersebut dengan tujuan melakukan klasifikasi, estimasi, prediksi, asosiasi, deskripsi dan visualisasi(Han dan Kamber, 2001)

3. KNN

Dasar Algoritma K Nearest Neaghbour (Brammer,2007):

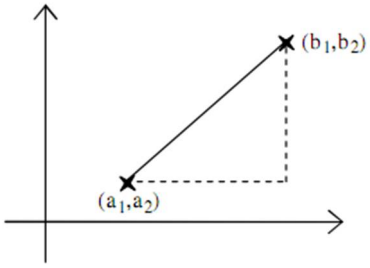
a. Temukan k pelatihan yang paling dekat dengan data yang tidak diketahui classnya.
b. Ambil yang paling sering terjadi untuk klasifikasi dari sebanyak k pelatihan.

K Nearest Neaghbour terutama digunakan ketika semua atribut bernilai kontinue,meskipun dapat dimodifikasi untuk menangani atribut kategorikal.

Dalam atribut continue untuk jarak terdekat dipakai rumus jarak Euclidean antara titik (a1,

a2, ..., an) dan (b1, b2, ..., bn) dalam ruang n-dimensi adalah generalisasi dari dua hasil ini. Jarak Euclidean diberikan oleh rumus :

$\sqrt{(a1 - b1)^2 + (a2 - b2)^2 + ... + (an - bn)^2}$



(Brammer, 2007)

Gambar 1. contoh jarak euclidean

Salah satu kelemahan dari pendekatan K NN untuk klasifikasi adalah bahwa tidak ada cara yang sepenuhnya memuaskan berurusan dengan atribut kategoris. Salah satu kemungkinannya adalah untuk menjawab bahwa selisih diantara dua nilai atribut yang identik adalah nol dan bahwa selisih diantara dua nilai berbeda adalah 1. Contoh efektif untuk atribut warna, misal merah - merah = 0, merah - biru = 1, biru - hijau = 1.

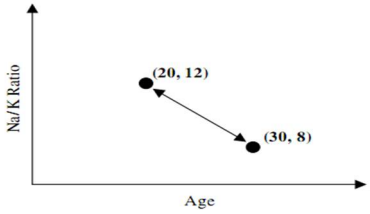
Fungsi jarak yang paling umum adalah jarak Euclidean, yang merupakan cara biasa di mana manusia berpikir jarak di dunia nyata:

$$d_{\text{Euclidean}}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$$

di mana x = x1, x2, ..., xm, dan y = y1, y2, ..., ym mewakili nilai atribut m dari dua catatan. Misalnya, pasien A adalah x1 = 20 tahun dan memiliki rasio Na / K dari x2 = 12, sedangkan pasien B adalah y1 = 30 tahun dan memiliki rasio Na / K dari y2 = 8. Kemudian jarak Euclidean antara titik-titik ini, seperti yang ditunjukkan pada Gambar 2.2, adalah

$$d_{\text{Euclidean}}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2} = \sqrt{(20-30)^2 + (12-8)^2}$$

$$= \sqrt{100 + 16} = 10.77$$



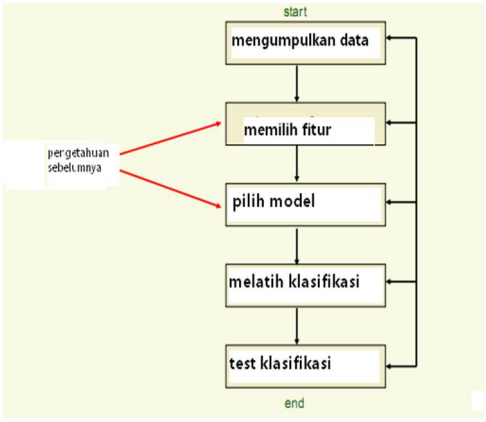
(Larose.2005)

Gambar 2.2 jarak Euclidean

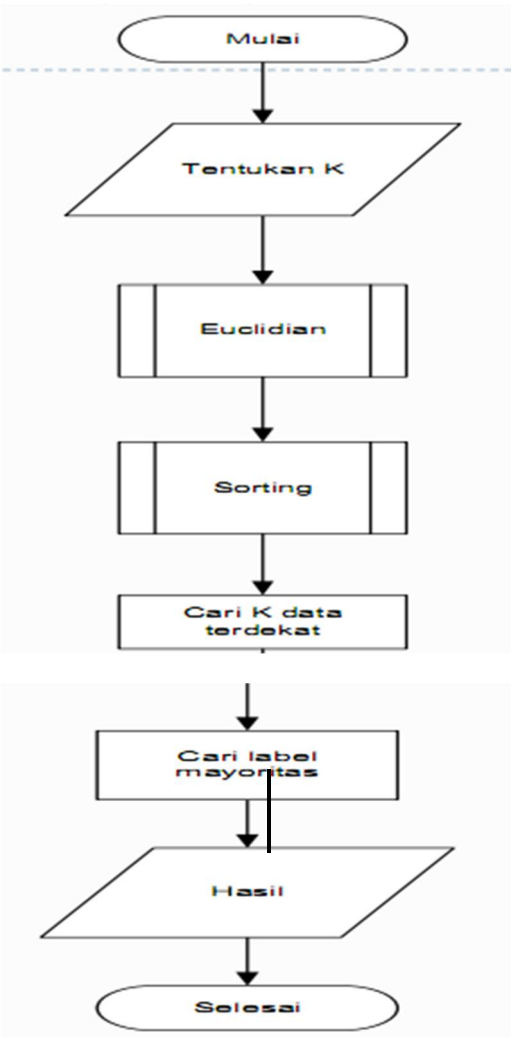
II. Metode penelitian

Data dalam hal ini menggunakan data repository uci machine learning.

Merancang sebuah sistem klasifikasi?



Algoritma K-Nearest Neaghbour :



Sumber : Astrid Darmawan (2012)

Gambar 3.2.Agoritma K nearest Neighbour

IV. Hasil dan Pembahasan

1. Perancangan :

Aplikasi data mining yang dibuat terdiri dari dua data, yaitu:

a. Data Testing :
   1. Harga Mobil
   2. Harga Maintenance
   3. Pintu
   4. Muatan Orang
   5. Besar Bagasi
   6. Keamanan

b. Data Training :
   1. Harga Mobil
   2. Harga Maintenance
   3. Pintu
   4. Muatan Orang
   5. Besar Bagasi
   6. Keamanan
   7. Kelayakan (kategori Peringkat)

2. Penjelasan hasil penelitian

Kasus : Klasifikasi kelayakan mobil

Data traning :

Harga_mobil	Harga_maint	Pintu	Orang	Bagasi	Keamanan	Kelayakan
sangat_tinggi	sangat_tinggi	3	2	sedang	rendah	tidak_baik
sangat_tinggi	sangat_tinggi	3	2	sedang	sedang	tidak_baik
sangat_tinggi	sangat_tinggi	5	6	besar	tinggi	tidak_baik
sangat_tinggi	tinggi	3	2	besar	sedang	tidak_baik
sangat_tinggi	tinggi	3	2	besar	tinggi	tidak_baik
tinggi	sedang	5	6	besar	tinggi	kurang_baik
tinggi	rendah	5	6	kecil	sedang	tidak_baik
tinggi	rendah	5	6	besar	tinggi	kurang_baik
sedang	sangat_tinggi	2	6	besar	tinggi	kurang_baik
sedang	sedang	4	6	besar	sedang	kurang_baik
sedang	sedang	4	6	besar	tinggi	Sgt baik
sedang	rendah	5	6	sedang	sedang	baik
sedang	rendah	5	6	sedang	tinggi	Sgt baik
sedang	rendah	5	6	besar	rendah	tidak_baik
sedang	rendah	5	6	besar	sedang	baik
sedang	rendah	5	6	besar	tinggi	Sgt baik
rendah	sangat_tinggi	5	4	besar	tinggi	kurang_baik
rendah	sangat_tinggi	5	6	kecil	rendah	tidak_baik

Label class

Data testing :

Harga_mobil	Harga_maint	Pintu	Orang	Bagasi	Keamanan	Kelayaka
tinggi	sangat_tinggi	2	2	kecil	rendah	?

Label class

Proses :

- Parameter yang dipakai adalah K=11
- Menghitung kuadrat jarak Euclidean (query instance) masing-masing objek terhadap sampel data atau training sample yang diberikan dengan menggunakan rumus:

Jarak Auclidean

$$d_{\text{Euclidean}}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2}$$

Contoh : kasus data, Perhitungan jarak antara data baru dengan data sample :

1.Jarak :

=((sangat tinggi-tinggi)^2 +(sangat tinggi-sangat tinggi)^2+(3-2)^2+(2-2)^2+(sedang-kecil)^2+(rendah-rendah)^2)$^{1/2}$

=1,732 dan seterusnya sampai semua data sample.

Tabel 4.1.
Tabel Perhitungan jarak Auclidean

No	Atribut							
	Hrg Mobil	Hrg maint	Pintu	Orang	Bagasi	Keamanan	Kelayakan	Jarak
1	Sgt tinggi	Sgt tinggi	3	2	sedang	rendah	Tdk baik	1,732
2	tinggi	sedang	5	6	besar	tinggi	Krg baik	5,292
3	Sgt tinggi	Sgt tinggi	3	4	besar	tinggi	Tdk baik	5,292
4	sedang	Sgt tinggi	2	6	besar	tinggi	Krg baik	4,359
5	Sgt tinggi	tinggi	3	2	besar	sedang	Tdk baik	2,236
6	Sgt tinggi	Sgt tinggi	3	2	sedang	sedang	Tdk baik	2
7	rendah	tinggi	3	2	sedang	sedang	Tdk baik	2,236
8	Sgt tinggi	Sgt tinggi	5	2	besar	tinggi	Tdk baik	3,464
9	tinggi	tinggi	3	6	kecil	rendah	Tdk baik	4,243
10	sedang	sedang	4	6	besar	kecil	Sgt baik	4,899
11	sedang	rendah	5	6	sedang	sedang	Baik	5,395
12	tinggi	tinggi	3	4	sedang	sedang	Tdk baik	2,828
13	rendah	tinggi	4	2	kecil	sedang	Tdk baik	2,646
14	rendah	Sgt tinggi	5	4	besar	tinggi	Krg baik	4
dst	-------	-------	------	------	---------	---------	------	-----

Sumber : Pengolahan (2015)

Dari tabel 4.1 Kemudian mengurutkan objek-objek tersebut sebanyak 11 nomer ke dalam kelompok yang mempunyai jarak Euclid terkecil

Cari mayoritas kategori kelayakan terbanyak

Kategori kelayakan Tdk baik sebanyak :8, kelayakan Krg baik 2, Sgt baik 1.

Tabel 4.2
. Hasil uji validitas sistem

inst#,	actual	predicted,	keterangan
1	Tdk baik	Tdk baik	T
2	Krg bai	Krg bai	T
3	Tdk baik	Tdk baik	T
4	Krg baik	Krg baik	T
5	Tdk baik	Tdk baik	T
6	Tdk baik	Tdk baik	T
7	Tdk baik	Tdk baik	T
8	Tdk baik	Tdk baik	T
9	Tdk baik	Tdk baik	T
10	Sgt baik	Krg bai	F
11	Baik	Tdk baik	F
12	Tdk baik	Tdk baik	T
13	Tdk baik	Tdk baik	T
14	Krg baik	Tdk baik	F

Keterangan:

T = True. Terjadi apabila hasil sistem sama dengan data sampel.

F = False. Terjadi apabila hasil sistem berbeda dengan data sampel.
Berdasarkan pengujian validitas yang dilakukan maka diperoleh:

Kinerja KNN=$\frac{Banyaknya\ \boldsymbol{pengujian}\ bernilai\ benar}{Banyaknya\ data\ sampel}$ ×100%
=$\frac{11}{14}$×100%=78,57%

Hasil dapat dilihat dengan menggunakan tool weka :

Correctly Classified Instances	11	78.5714 %
Incorrectly Classified Instances	3	21.4286 %
Kappa statistic	0.5484	
Mean absolute error	0.1262	
Root mean squared error	0.2285	
Relative absolute error	51.17 %	
Root relative squared error	67.7714 %	
Total Number of Instances	14	

Untuk data jumlah 1728 didapat hasil accuracy memuaskan yaitu :

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1655               95.7755 %
Incorrectly Classified Instances        73                4.2245 %
Kappa statistic                          0.9051
Mean absolute error                      0.1123
Root mean squared error                  0.1942
Relative absolute error                 49.0156 %
Root relative squared error             57.4461 %
Total Number of Instances             1728

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.033      0.986     1         0.993      1        tidak_baik
                 0.943     0.016      0.945     0.943     0.944      0.996    kurang_baik
                 0.908     0.003      0.922     0.908     0.915      1        Sgt  baik
                 0.739     0.002      0.944     0.739     0.829      0.997    baik
Weighted Avg.    0.973     0.027      0.973     0.973     0.973      0.999
```

IV.1 Aplikasi dengan matlab :

Gambar 4.1.Hasil pengolahan 2015

Dengan menggunakan aplikasi matlab terlihat pada gambar 4.1 bahwa metode KNN dapat memprediksi Kelayakan mobil

V KESIMPULAN

1. Sistem ini dapat dijadikan sebagai alat bantu untuk menentukan kelayakan mobil. 2. K-NN dapat digunakan untuk menentukan kelayakan mobil menurut parameter kondisi fisik dari mobil tersebut.
3. Aplikasi data mining ini dapat memprediksi dengan menggunakan 1 data mobil atau 1 database.
4. Untuk menggunakan data training yang berjumlah 14 data dengan jumlah k=3 didapat nilai accuracy 78%.
5.untuk data training yang berjumlah 1728 data dengan k=11 didapat nilai accuracy 95.78%.
6.nilai kappa statistic dan precission mendekati nilai 1, yang artinya bahwa metode KNN dapat digunakan untuk klasifikasi dengan memuaskan
7.nilai ROC area juga mendekati 1 artinya sistim ini cukup akurat.
8.semakin besar jumlah data training sistim akan semakin akurat

VI DAFTAR PUSTAKA

Larose (2005). *Discovering Knowledge In Data* Central Connecticut State University. United States of America

Brammer (2007). *Principles of Data Mining*. Digital Professor of Information Technology, University of Portsmouth, UK.

Witten.(2011). *Data Mining Practical Machine Learning Tools and Techniques* Library of Congress Cataloging-in-Publication Data..

Cunningham and Jane Delany.(2007). *k-Nearest Neighbour Classifiers,* Dublin Institute of Technology.

Tedy Rismawan, Ardhitya Wiedha Irawan, Wahyu Prabowo, Sri Kusumadewi.(2008). *sistem pendukung keputusan berbasis pocket pc sebagai penentu status gizi menggunakan metode knn (k-nearest neighbor)*. Fakultas Teknologi Industri, Universitas Islam Indonesia

Alexander Hinneburg et all. (2000). *What is the nearest neighbor in high dimensional spaces?*. Proceedings of the 26th VLDB Conference, Cairo, Egypt

Astrid Darmawan (2012). *pembuatan aplikasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma k-nearest neighborhood*. fakultas teknik dan ilmu computer universitas komputer Indonesia.

Olga Kudoyan. (2010). *The incremental benefits of the nearest neighbor forecast of u.s. energy commodity prices*: Thesis Texas A&M University