

IMPLEMENTASI ALGORITMA C4.5 MENGGUNAKAN PYTHON UNTUK KLASIFIKASI KEPUASAN KONSUMEN

Medy Wisnu Prihatmono¹⁾ Alviva Felicia Watratan²⁾

¹⁾Sistem Informasi, STMIK Profesional Makassar

²⁾Sistem Informasi, STMIK Profesional Makassar

Email : medywisnuprihatmonoprofesional@gmail.com, vinawatratan@gmail.com

Abstrak

Kepuasan konsumen merupakan tingkat dimana anggapan terhadap produk sesuai dengan harapan para konsumen. Harapan konsumen umumnya merupakan prakiraan atau keyakinan konsumen tentang apa yang akan diterimanya bila telah membeli atau mengkonsumsi suatu produk. Kenyataannya, apa yang bisa memuaskan konsumen di satu situasi mungkin tidak bisa memuaskan konsumen yang sama di lain situasi. Ditambah dengan sistem keluhan konsumen yang hanya disediakan contact person oleh pihak restoran. Dengan demikian, kepuasan konsumen sangat sulit diketahui, karena tidak semua konsumen menghubungi contact person yang disediakan. Penelitian dilakukan pada rumah makan pallubasasrigala kota makassar dengan menebarkan angket kuisioner sebanyak 100 data.. Kemudian diolah dengan klasifikasi Decision Tree menggunakan algoritma C4.5. Dari proses klasifikasi akan menghasilkan beberapa aturan yang menyebabkan konsumen merasa puas dan tidak puas, serta seberapa besar konsumen merasa puas maupun tidak puas. Hasil yang diperoleh pada penelitian ini adalah Sistem ini dapat mengklasifikasikan konsumen merasa puas apabila pelayanan dirasa lumayan, product dirasa lumayan, fasilitas dirasa baik dan pelayanan dirasa ramah, dengan persentasi akurasi sebanyak 70 %.

Kata Kunci : *decision tree, algoritma C4.5, konsumen, produk,*

A. PENDAHULUAN

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelakunya harus selalu memikirkan cara-cara untuk terus survive dan jika mungkin dapat mengembangkan skala bisnisnya. Ditambah persaingan dalam bidang makanan minuman sangat berkembang pesat seiring dengan bertambahnya inovasi dan kreasi baru para pebisnis. Oleh karena itu para pebisnis banyak melakukan cara atau strategi yang lebih

baik agar dapat memenuhi kebutuhan konsumen serta memberikan kepuasan secara maksimal dan melayani dengan sebaiknya. Karena tujuan dari suatu bisnis adalah untuk menciptakan rasa puas kepada konsumen. Salah satu tindakan yang harus dilakukan untuk meraih keunggulan bersaing tersebut adalah fokus terhadap konsumen dengan memberikan sebuah pengalaman yang dapat memberikan kepuasan bagi konsumen.

Konsumen yang puas adalah konsumen yang akan berbagi kepuasan dengan produsen. Bahkan, konsumen yang puas akan berbagi rasa dan pengalaman dengan konsumen lain. Oleh karena itu, baik konsumen maupun produsen akan sama-sama diuntungkan apabila kepuasan terjadi. Kenyataannya, apa yang bisa memuaskan konsumen di satu situasi mungkin tidak bisa memuaskan konsumen yang sama di lain situasi. Ditambah dengan sistem keluhan konsumen yang hanya disediakan contact person oleh pihak restoran. Dengan demikian, kepuasan konsumen sangat sulit diketahui, karena tidak semua konsumen menghubungi contact person yang disediakan. Maka dari itu, adanya penelitian mengenai analisis kepuasan konsumen di Rumah Makan Pallubasa srigala tersebut diperlukan. Pada Rumah Makan Pallubasa srigala, terdapat beberapa permasalahan yang kerap muncul mengenai kepuasan konsumen. Rumah makan sulit mendapatkan informasi-informasi mengenai tingkat kepuasan konsumen yang telah berkunjung. Karena sebab itu, penulis membuat penelitian dengan judul “implementasi algoritma c4.5 menggunakan python untuk klasifikasi kepuasan konsumen”.

B. METODE PENELITIAN

Data mining

Data mining telah banyak menarik perhatian di dunia sistem informasi dan di masyarakat secara keseluruhan dalam beberapa tahun ini, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analisis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi (Han & Kamber, 2012)

Data mining adalah proses menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika (Larose & Larose, 2014) Ada beberapa teknik yang dimiliki data mining berdasarkan tugas yang bisa dilakukan, antara lain:

1. Deskripsi.

Para peneliti biasanya mencoba menemukan cara mendeskripsikan pola dan trend yang tersembunyi dalam data.

2. Estimasi.

Estimasi hampir sama dengan klasifikasi, kecuali variable tujuan yang lebih kearah numeric dari pada kategori.

3. Prediksi.

Prediksi memiliki beberapa kemiripan dengan estimasi dan klasifikasi. Hanya saja jika prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).

4. Klasifikasi.

Dalam klasifikasi variable, tujuan bersifat kategori. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi pendapatan sedang, dan pendapatan rendah.

5. Clustering

Clustering lebih condong kearah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

6. Asosiasi.

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada suatu waktu.

Klasifikasi

Menurut Han dan Kamber (2001) data classification memiliki dua tahap proses. Tahap pertama adalah membangun suatu model yang berdasarkan serangkaian data class, yang disebut learned model. Model tersebut dibangun dengan menganalisa database tuple. Setiap tuple diasumsikan

menjadi predefined class yang ditentukan oleh satu atribut yang disebut class label attribute. Akibat terdapat class label maka tahap ini juga dikenal dengan supervised learning. Berbeda dengan unsupervised learning atau dikenal dengan clustering.

Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membangun sebuah pohon keputusan (decision tree) dari data. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang juga merupakan algoritma untuk membangun sebuah pohon keputusan. Algoritma C4.5

secara rekursif mengunjungi tiap simpul keputusan, memilih percabangan optimal, sampai tidak ada cabang lagi yang mungkin dihasilkan. (Larose, D.T. 2005)

Algoritma C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya yaitu ID3. Algoritma C4.5 dan ID3 model yang tak terpisahkan, karena membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5 Diakhir tahun 1980-an, J. Ross Quinlan seorang peneliti di bidang mesin pembelajaran mengembangkan sebuah model pohon keputusan yang dinamakan ID3. Ada beberapa tahapan

dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu:

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang sudah pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas- kelas tertentu.

2. Menghitung akar pohon. Akar akan diambil dari atribut yang akan dipilih, dengan cara menghitung nilai gain dari masing- masing atribut, nilai gain yang paling tinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. (Ilham, S.M.R. & Purwanto. 2015)

Untuk menghitung nilai entropy digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan:

S : Himpunan Kasus

n : Jumlah Partisi S

pi : Proporsi dari Si terhadap S

Kemudian setelah nilai entropy pada masing-masing atribut sudah diperoleh maka hitung nilai gain dengan menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi atribut A

|Si| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S

Python

Python adalah bahasa pemrograman model skrip (scripting language) yang berorientasi obyek. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi. Python merupakan bahasa pemrograman yang freeware atau perangkat bebas dalam arti sebenarnya, tidak ada batasan dalam penyalinannya atau mendistribusikannya. Lengkap dengan source codenya, debugger dan profiler, antarmuka yang terkandung di dalamnya untuk pelayanan antarmuka, fungsi sistem, GUI (antarmuka pengguna grafis), dan basis datanya. (Triasanti, 2010). Beberapa fitur yang dimiliki Python adalah:

1. Memiliki kepustakaan yang luas; dalam distribusi Python telah disediakan modul-modul.

2. Memiliki tata bahasa yang jernih dan mudah dipelajari.

3. Memiliki aturan layout kode sumber yang memudahkan pengecekan, pembacaan kembali dan penulisan ulang kode sumber.

4. Berorientasi obyek.

5. Dapat dibangun dengan bahasa Python maupun C/C++.

C. HASIL DAN PEMBAHASAN

Pada penelitian yang dilakukan dengan menyebarkan kuisiner terhadap para pelanggan di rumah makan pallu basa srigala, sebanyak 100 data kuisiner, yang dimana Selanjutnya data kuisiner di sederhanakan menjadi tiga kategori yaitu produk, fasilitas dan pelayanan agar pengklasifikasiannya, yang dimana menggunakan libraries dari python untuk data science antara lain pandas, scikit-learn, dan pydotplus yang berfungsi sebagai pencetak grafik atau visualisasi berdasarkan inputan data dengan menggunakan algoritma C4.5 yang termasuk dalam bagian dari supervised learning. Lebih jelasnya seperti terlihat pada Gambar 1 di bawah ini.

```
# Medy C4.5 PalluBasaSrigala
In [25]: import pandas as pd
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
import pydotplus
from IPython.display import Image

atribut = ['Produk', 'Fasilitas', 'Pelayanan', 'Klasifikasi']
pbc = pd.read_csv('rumahan.csv', header=0, names=atribut)
pbc.head()

Out[25]:
```

	Produk	Fasilitas	Pelayanan	Klasifikasi
0	Baik	Ya	Ramah	Puas
1	Baik	Ya	Kurang	Puas
2	Baik	Ya	Ramah	Puas
3	Lumayan	Ya	Ramah	Tidak Puas
4	Lumayan	Ya	Kurang	Tidak Puas

Gambar 1. Menampilkan dataset

Setelah data kuisiner yang berjumlah 100 ditampilkan, selanjutnya dataset

tersebut dirubah dari string ke biner dengan tujuan menggunakan fungsi dummies untuk memecahkan data pada python untuk memproses pengklasifikasian dari dataset diatas. Untuk lebih jelasnya seperti terlihat pada Gambar 2 di bawah ini.

```
In [26]: srigala = pd.get_dummies(pbc[['Produk', 'Fasilitas', 'Pelayanan']])
srigala.head()

Out[26]:
```

	Produk_Baik	Produk_Kurang	Produk_Lumayan	Fasilitas_Tidak	Fasilitas_Ya	Pelayanan_Kurang	Pelayanan_Lumayan	Pelayanan_Ramah
0	1	0	0	0	1	0	0	1
1	1	0	0	0	1	1	0	0
2	1	0	0	0	1	0	0	1
3	0	0	1	0	1	0	0	1
4	0	0	1	0	1	1	0	0

Gambar 2. Hasil Dummies Biner

Kemudian langkah selanjutnya adalah membuat dataset tersebut menjadi dua bagian yaitu dataset training dan dataset testing yang dimana komposisi yang dipakai adalah untuk data training sebesar 70 % dan data testing sebesar 30%. Untuk Lebih jelasnya seperti terlihat pada Gambar 3 di bawah ini.

```
In [6]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

Gambar 3. Setdata Training dan Testing Selanjutnya dengan menggunakan fungsi libraries Scikit-learn dari Python bertujuan untuk memanggil fungsi decision tree, sekaligus menampilkan hasil jumlah persentasi prediksi sebesar 70 % , Untuk Lebih jelasnya seperti terlihat pada Gambar 4 di bawah ini.

```

In [7]: # Create Decision Tree classifier object
clf = DecisionTreeClassifier()

# Train Decision Tree Classifier
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)

In [9]: print("Hasil Accuracy:",metrics.accuracy_score(y_test, y_pred))

Hasil Accuracy: 0.7012987012987013

```

Gambar 4. Model decision tree dan Hasil Prediksi yang diperoleh.

Proses Terakhir adalah membuat diagram tree, berdasarkan 100 dataset yang dipakai dengan menggunakan pydotplus serta memanggil fungsi decisiontreeclassifier, yang berfungsi untuk membuat pohon keputusan Node 1 (root node). Untuk Lebih jelasnya seperti terlihat pada Gambar 5 di bawah ini

```

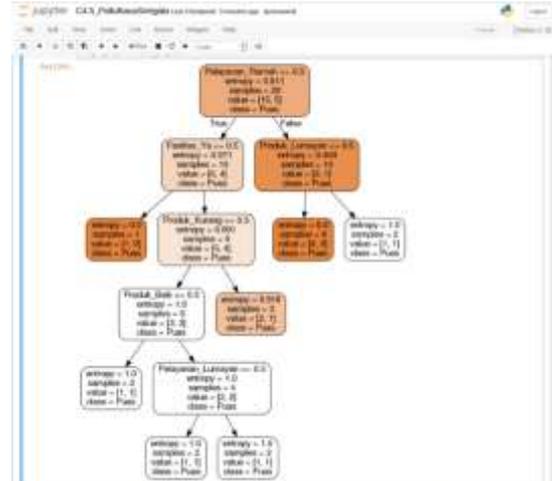
In [21]: hasil = tree.DecisionTreeClassifier(criterion='entropy')
hasil_train = hasil.fit(sragala, pls['Klasifikasi'])

In [24]: dot_data = tree.export_graphviz(hasil_train, out_file=None, feature_names=list(sragala.columns.values),
class_names=['Pusi', 'TidakPusi'], rounded=True, filled=True)
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_graph())

```

Gambar 5. Membuat pohon keputusan

Pada akhirnya diperoleh hasil gain tertinggi adalah "pelayanan = ramah = Puas". Untuk node ke dua dan seterusnya dapat dilihat pada gambar 6 dibawah ini



Gambar 6. Prediksi Pohon Keputusan

D. KESIMPULAN DAN SARAN

Dari hasil dan pembahasan penelitian yang telah dilakukan dapat disimpulkan bahwa berdasarkan dari 100 record dataset yang digunakan dapat disimpulkan bahwa

1. Sistem ini dapat mengklasifikasikan konsumen merasa tidak puas apabila pelayanan dirasa kurang, fasilitas dirasa kurang dan pelayanan ramah, product dirasa kurang, fasilitas dirasa baik dan pelayanan dirasa ramah.
2. Sistem ini dapat mengklasifikasikan konsumen merasa puas apabila pelayanan dirasa lumayan, product dirasa lumayan, fasilitas dirasa baik dan pelayanan dirasa ramah.

REFERENSI

[1] Han, J., & Kember, M. (2012). Data Mining Concepts & Techniques, Simon Fraser University Academic Press, US.

- [2] Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data*. <https://doi.org/10.1002/>
- [3] Triasanti, Dina. 2001. *Konsep Dasar Phyton*. Jakarta.
- [4] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data mining*.
- [5] *Wiley Interscience. Ney Jersey. 9781118874059.*
- [6] Ilham, S.M.R. & Purwanto. 2015. *Implementasi Data Mining Menggunakan Algoritma C4.5 untuk Prediksi Kepuasan Pelanggan Taksi Kosti*. Tesis. Semarang:
- Jurusan Teknik Informatika,
Fakultas Ilmu Komputer,
Universitas Dian Nuswantoro
Semarang.