

PERFORMANSI KLASIFIKASI KELAS SISWA MENGGUNAKAN NAÏVE BAYES DAN DECISION TREE

Siti Mujilahwati¹, Febrianti Widyahastuti², Nur Nafi'iyah³

^{1,2}Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan
moedjee@gmail.com, febriantiwidyahastuti@gmail.com, mynaff26@gmail.com

ABSTRAK

Pembagian kelas di dalam pendidikan bertujuan untuk mengetahui kemampuan intelektual siswa. Di dunia pendidikan telah mengenal beberapa pembagian kelas mulai dari kelas reguler, kelas unggulan, dan kelas akselerasi. Beberapa jenis kelas tersebut untuk membagi kelas-kelas siswa berdasarkan kemampuan intelektualnya. Kelas reguler merupakan kelas untuk siswa yang mempunyai kemampuan rata-rata, kelas unggulan merupakan kelas untuk siswa yang di atas rata-rata, sedangkan kelas akselerasi merupakan kelas untuk siswa yang jenius. Dengan kondisi tersebut maka penelitian ini akan melakukan klasifikasi kelas siswa dalam pembelajaran di dunia pendidikan. Tujuan penelitian ini, yaitu: untuk mengetahui hasil klasifikasi pembagian kelas siswa menggunakan naïve bayes dan decision tree. Tahapan penelitian ini, yaitu: mengumpulkan data dari sampel lembaga pendidikan, selanjutnya menentukan variabel dan data set, tahapan terakhir mengklasifikasikan kelas siswa. Klasifikasi kelas siswa dalam penelitian ini terdapat 3 kelas, yaitu: kelas Low, kelas Middle, dan kelas High. Data yang digunakan dalam penelitian ini sebanyak 480 baris dan 12 atribut, dengan total kelas Low 127 baris, kelas Middle 211 baris, dan kelas High sebanyak 142 baris. Hasil klasifikasi dari penelitian ini menunjukkan bahwa akurasi dari Nive Bayes 58%, akurasi klasifikasi algoritma ID3 60%, akurasi klasifikasi algoritma C4.5 62%, sedangkan akurasi klasifikasi algoritma CART 58%. Sehingga dapat disimpulkan bahwa performansi akurasi dari keempat algoritma yang paling baik, yaitu algoritma C4.5 sebesar 62%.

Kata Kunci: *performansi, klasifikasi, naïve bayes, decision tree*

1. PENDAHULUAN

Perkembangan teknologi dan informasi mempengaruhi dunia pendidikan, terutama dalam menerima dan menangkap informasi. Di beberapa negara berkembang cara belajar mengajar mulai terus diperbaiki, dari hal belajar online, tes online, serta pembagian kelas belajar mengajar.

Beberapa tahun terakhir muncul model belajar online dengan menggunakan aplikasi atau website online (e-learning). Dan juga berkembang ujian online dengan istilah CBT (Computer Based Test). Muncul juga pembagian kelas dalam proses belajar mengajar. Hal itu untuk membagi kelas belajar sesuai kemampuan siswa. Mulai dari kelas reguler, kelas unggulan dan kelas akselerasi. Pembagian kelas sesuai dengan kemampuan intelektual siswa, kelas reguler untuk siswa yang mempunyai kemampuan rata-rata, kelas unggulan untuk siswa yang mempunyai kemampuan di atas rata-rata, dan kelas akselerasi untuk siswa dengan kemampuan jenius.

Dari kondisi di dunia pendidikan yang semakin berkembang, maka penelitian ini bermaksud untuk mengklasifikasi kelas belajar siswa sesuai kemampuan intelektualnya. Tujuan penelitian ini, yaitu: untuk mengklasifikasikan kelas belajar siswa menggunakan algoritma naïve bayes dan decision tree. Agar dapat membantu dunia pendidikan dalam membagi kelas belajar sesuai kemampuan intelektual siswa.

Dalam penelitian ini pengklasifikasian menggunakan algoritma naïve bayes dan decision tree. Menurut Andrew Mc Callum, Kamal Nigam bahwa Naïve Bayes klassifier adalah model yang paling sederhana dari beberapa klassifier, di mana dalam mengasumsikan semua atribut saling independen satu sama lain dalam konteks kelasnya.

Dalam penelitian P. Bhargavi, S. Jyothi, 2009 bahwa Naïve Bayes classfier dapat bekerja dengan cepat dan ikremental sehingga dapat menangani atribut yang diskrit dan berkelanjutan. Naïve bayes juga memiliki kinerja yang sangat baik dalam menangani data di kehidupan nyata dan dapat mengambil keputusan sangat baik.

Algoritma decision tree terdiri dari ID3, C4.5, CART. Di mana Mardi mengatakan bahwa klasifikasi dalam data mining dapat dilakukan dengan menggunakan algoritma C4.5. Dengan algoritma C4.5, akan didapatkan sebuah pohon keputusan yang mudah dipahami dan mudah dimengerti.

Dalam penelitian Fadlan Amirudin, Eneng Tita Tosida, Irma Anggraeni, Algoritma CART ini menggunakan pendekatan nonparametic yang tidak membutuhkan asumsi distribusi, lalu akan mengidentifikasi variable secara otomatis yang berpengaruh dan mereduksi kompleksitas data, mudah dalam mengatasi data outlier, dan mudah dalam interpretasi. Untuk itu diterapkan algoritma

CART dengan tujuan mampu menghasilkan klasifikasi berdasarkan informasi data yang ada.

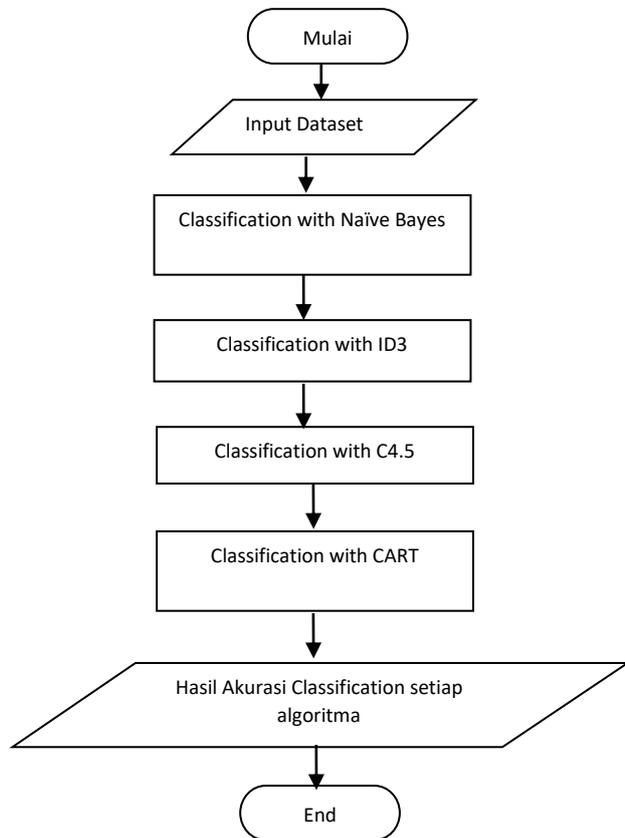
Di penelitian Rani, 2015 dengan metode keputusan decision tree menggunakan algoritma C4.5 diharapkan proses penggalian informasi lebih cepat dan optimal dengan kapasitas data yang lebih besar, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan.

Dari beberapa penelitian terdahulu, sehingga peneliti akan menggunakan algoritma Naïve Bayes, ID3, C4.5 dan CART dalam mengklasifikasi kelas siswa saat proses belajar mengajar.

Dalam penelitian ini terdapat pendahuluan, tinjauan pustaka, metodologi penelitian, hasil, kesimpulan dan

2. METODE

Alur dari penelitian ini seperti dalam Gambar 2.



Gambar 2. Alur Penelitian

Penelitian ini diharapkan menghasilkan nilai akurasi yang terbaik dari masing-masing algoritma. Algoritma yang digunakan dalam klasifikasi, yaitu; Naïve bayes, ID3, C4.5, dan CART. Data yang digunakan dalam penelitian sebanyak 480 baris, dengan 12 atribut dan terdiri 3 kelas, yaitu: kelas Low, Middle, dan High.

Data sebelumnya dicari nilai korelasi setiap variabel. Selanjutnya dilakukan ujicoba klasifikasi dari setiap algoritma. Proses ujicoba klasifikasi dataset menggunakan aplikasi weka, dan data disimpan dalam bentuk arff.

3. PEMBAHASAN

Data yang akan dibaca dalam aplikasi weka disimpan dengan format arff, seperti Gambar 3.

No.	NationalITy	PlaceofBirth	StageID	GradeID	Semester	raisedhands	VisITedR
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nor
1	KW	KuwaIT	lowerle...	G-04	F	Low	Low
2	KW	KuwaIT	lowerle...	G-04	F	Low	Low
3	KW	KuwaIT	lowerle...	G-04	F	Low	Low
4	KW	KuwaIT	lowerle...	G-04	F	Low	Low
5	KW	KuwaIT	lowerle...	G-04	F	Low	Low
6	KW	KuwaIT	lowerle...	G-04	F	Low	Low
7	KW	KuwaIT	Middle...	G-07	F	Low	Low
8	KW	KuwaIT	Middle...	G-07	F	High	Low
9	KW	KuwaIT	Middle...	G-07	F	Low	Low
10	KW	KuwaIT	Middle...	G-07	F	High	High
11	KW	KuwaIT	Middle...	G-07	F	High	High
12	KW	KuwaIT	Middle...	G-07	F	Low	Low
13	KW	KuwaIT	lowerle...	G-04	F	Low	Low
14	lebanon	lebanon	Middle...	G-08	F	Low	Low
15	KW	KuwaIT	Middle...	G-08	F	High	High
16	KW	KuwaIT	Middle...	G-06	F	Low	Low
17	KW	KuwaIT	Middle...	G-07	F	Low	Low
18	KW	KuwaIT	Middle...	G-07	F	High	Low
19	KW	KuwaIT	Middle...	G-07	F	High	Low
20	KW	KuwaIT	Middle...	G-07	F	High	Low
21	KW	KuwaIT	Middle...	G-07	F	High	High
22	KW	KuwaIT	Middle...	G-07	F	Low	Low
23	KW	KuwaIT	Middle...	G-07	F	Low	Low

Gambar 3. Dataset Classification

Selanjutnya melakukan proses ujicoba klasifikasi dengan algoritma Naïve Bayes, seperti Gambar 4.

```

Classifier output
Correctly Classified Instances      277          57.7083 %
Incorrectly Classified Instances    203          42.2917 %
Kappa statistic                    0.3674
Mean absolute error                 0.2937
Root mean squared error            0.4349
Relative absolute error             67.8376 %
Root relative squared error        93.4786 %
Total Number of Instances          480

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area
              -----  -----  -
              0.398   0.264   0.542     0.398   0.459     0.63
              0.811   0.176   0.624     0.811   0.705     0.901
              0.634   0.207   0.563     0.634   0.596     0.825
Weighted Avg.  0.577   0.224   0.57      0.577   0.565     0.76

=== Confusion Matrix ===
  a  b  c  <-- classified as
 84 59 68 | a = M
 22 103 2 | b = L
 49  3 90 | c = H
    
```

Gambar 4. Hasil Ujicoba Klasifikasi Naïve Bayes

Selanjutnya data hasil ujicoba klasifikasi algoritma ID3 seperti Gambar 5.

- Produk dengan Menggunakan Metode Algoritma Apriori. *Seminar Nasional Teknologi Informasi dan Multimedia* (pp. 19-23). Yogyakarta: STIMIK AMIKOM.
- K. Srinivas, B. Kavitha Rani, A. Gourdhan. (2010). Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. *International Journal on Computer Science and Engineering Vol.2, No.2*, 250-255.
- M. Mahmoodi, A. Mirzazadeh. (2014). A New Analysis of Failure Models and Effects by Fuzzy Todim with using Fuzzy Time Function. *International Journal of Fuzzy Logic System Vol.4, No.2*, 7-21.
- Mardi, Y. (n.d.). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Endik Informatika Vol.2, No.2*, 213-219.
- Nafi'iyah, N. (2015). Algoritma CART dalam Penentuan Pohon Keputusan Sertifikasi Guru. *Jurnal SPIRIT Vol7, No.2*.
- Nur Nafi'iyah, Retno Wardhani. (2018). Analisa Fuzzy C45 dalam Mengklasifikasi Jenis Kelamin Manusia dari Fitur Citra Panoramik Gigi Kaninus. *SENIATI* (pp. 160-166). Malang: ITN.
- Nur Suriati Jamil, Nor Adzlan Jamaludin, Nurazzah Abdul Rahman, Nora Shida Sabari. (2011). Implementation of Vector-Space Online Document Retrieval System using Open Source Technology . *Conference on Open System* (pp. 395-399). Malaysia: IEEE.
- P. Bhargavi, S. Jyothi. (2009). Applying Naive Bayes Data Mining Techniques for Classification of Agricultural Land Soils. *International Journal of Computer Science and Network Security Vol.9, No.8*, 117-122.
- Ramadhanuz A Djamal, Warih Maharani, Angelina Prima Kurniati. (2010). Analisis dan Implementasi Metode Item-based Clustering Hybrid pada Recommender System. *Konferensi Nasional Sistem & Informatika*, (pp. 216-222). Bali.
- Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma C4.5 sebagai Dasar Pemberian Kredit. *Jurnal KomTekInfo Vol. 2, No. 2*, 33-38.
- Sellappan Palaniappan, Rafiah Awang. (2008). Intelligent Heart Disease Prediction System using Data Mining Techniques. *IEEE*, 108-115.
- Shelly Gupta, Dharminder Kumar, Anand Sharma. (2011). Data Mining Classification Techniques Applied for Breast Cancer Diagnosis and Prognosis. *Indian Journal of Computer Science and Engineering Vol.2, No.2*, 188-195.
- V. Krishnaiah, G. Narsimha, N. Subhash Chandra. (2013). Diagnosis of Lung Cancer Prediction System using Data Mining Classification Techniques. *International Journal of Computer Science and Information Technologies Vol.4, No.1*, 39-45.
- Wendi Warasta, Zaki Parasti. (n.d.). Implementasi Algoritma Apriori untuk Menganalisa Pola Pembelin Produk pada Data Transaksi Penjualan. 1-6.
- Yahya Al-Ashmoery, Rochdi Messoussi. (2015). Learning Analysis System for Assessing Students Performance Quality and Text Mining in Online Commication. *IEEE*.
- Yi-Chung Hu, Ruey-Shun Chen, Gwo-Hshiung Tzeng. (2003). Finding Fuzzy Classification Rules using Data Mining Techniques. *Pattern Recognition Letters Vol. 24*, 509-519.
- Zulhanif. (2015). Algoritma AdaBoost dalam Pengklasifikasian. *Seminar Nasional Matematika & Pendidikan Matematika* (pp. 559-569). UMS.