

## PENERAPAN ALGORITMA *DECISION TREE* ID3 UNTUK PREDIKSI KELULUSAN MAHASISWA JENJANG PENDIDIKAN D3 DI FAKULTAS TEKNIK UNIVERSITAS PANDANARAN

Abdul Rohman<sup>1</sup>, Anief Rufiyanto<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Elektronika, Fakultas Teknik, Universitas Pandanaran  
email: <sup>1</sup>abdulrohman@unpand.ac.id, <sup>2</sup>aniefrufiyanto@gmail.com

### Abstrak

Data mahasiswa menjadi hal yang sangat penting untuk mengambil suatu keputusan, jika data tersebut dianalisa dengan menggunakan *data mining*. Setiap kumpulan dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi perguruan tinggi. Pada perguruan tinggi, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap pada pengambilan suatu keputusan. *Data mining* dengan algoritma *Decision Tree* ID3 dapat digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola kelulusan mahasiswa.

Banyak penelitian tentang penerapan data mining untuk memprediksi kelulusan mahasiswa dengan menggunakan algoritma *Decision Tree*, dengan data mahasiswa reguler dan mayoritas statusnya belum bekerja. Sedangkan dalam penelitian ini mahasiswa Universitas Pandanaran, memiliki data mahasiswa kelas reguler dan mahasiswa kelas karyawan, dan kebanyakan statusnya sudah bekerja.

Tahapan yang dilakukan dalam penelitian ini yaitu; (1) pengumpulan data mahasiswa universitas pandanaran, (2) mengolah data awal mahasiswa (3) pengolahan data selanjutnya mahasiswa dengan menggunakan klasifikasi data mining algoritma *Decision Tree* ID3 (4) eksperimen dan pengujian algoritma (5) evaluasi dan validasi hasil (8) Menghasilkan Pola/Model Kelulusan Mahasiswa yang dapat dimanfaatkan untuk sebuah keputusan diperguruan tinggi.

Eksperimen dan pengujian algoritma *Decision Tree* ID3 terhadap data kelulusan mahasiswa menghasilkan nilai akurasi 73,19% dengan nilai AUC 0,806.

Kata Kunci:

*Data Mining*, *Decision Tree*, ID3, Mahasiswa

### PENDAHULUAN

Data mahasiswa menjadi hal yang sangat penting untuk mengambil suatu keputusan, jika data tersebut dianalisa dengan menggunakan *data mining*. Setiap kumpulan atau gudang data dapat memberikan pengetahuan penting yang menjadi informasi yang sangat berharga bagi perguruan tinggi. Pada perguruan tinggi, suatu sistem informasi dapat digunakan untuk memperoleh informasi yang menunjang setiap pada pengambilan suatu keputusan [1].

*Data mining* dengan algoritma *Decision Tree* banyak digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola kelulusan mahasiswa, karena banyak kelebihan dengan algoritma yang lainnya. Kelebihan Algoritma *Decision Tree* dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan

dapat menangani atribut bertipe diskret dan numerik [2].

Banyak penelitian tentang implementasi *data mining* untuk memprediksi kelulusan mahasiswa dengan menggunakan algoritma *Decision Tree*, dengan data mahasiswa reguler dan mayoritas statusnya belum bekerja [3], [4], [5], [6], [7], Sedangkan dalam penelitian ini mahasiswa Universitas Pandanaran terutama jenjang pendidikan D3 di Fakultas Teknik, memiliki data mahasiswa kelas reguler dan mahasiswa kelas karyawan dan kebanyakan statusnya sudah bekerja. Maka dengan itu diperlukan penelitian untuk memprediksi kelulusan mahasiswa.

### KAJIAN PUSTAKA

#### a. Kelulusan Mahasiswa

Mahasiswa sering disebut kelompok masyarakat yang memiliki ciri intelektualitas yang lebih luas dibandingkan dengan kelompok usia mereka yang bukan mahasiswa ataupun

kelompok usia lain yang dibawah mereka. Dengan intelektualitasnya mahasiswa akan mampu menghadapi dan mencari permasalahan secara sistematis yang nantinya diterapkan dalam kehidupan sehari-hari agar bisa bersaing dalam dunia kerja[8].

Kelulusan mahasiwa adalah hal yang penting diperhatikan, karena persentase jumlah kelulusan mempengaruhi penilaian pemerintah serta mempengaruhi status akreditasi program studi [9]. Faktor-faktor yang dapat mempengaruhi kelulusan mahasiswa antara lain adalah nilai akhir SMA, Indeks Prestasi Semester (IPS), gaji orang tua dan pekerjaan orang tua [6]. Indeks prestasi sering digunakan sebagai indikator penilaian akademik, banyak perguruan tinggi memberi standar minimum yang sulit di peroleh mahasiswa [10]. Adapun variabel yang dapat digunakan dalam prediksi kelulusan mahasiswa seperti umur, status pernikahan, jumlah saudara [11]. Pada penelitian ini parameter yang digunakan adalah nama jurusan, usia, jenis kelamin, status pekerjaan dan indeks prestasi semester satu sampai dengan indeks prestasi semester empat.

**b. Klasifikasi Data Mining**

Data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data [12]. Data mining akan memecahkan masalah dengan menganalisis data yang telah ada dalam basis data. Data mining, sering juga disebut *knowledge discovery in database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan pola keteraturan, pola hubungan dalam set data berukuran besar [13].

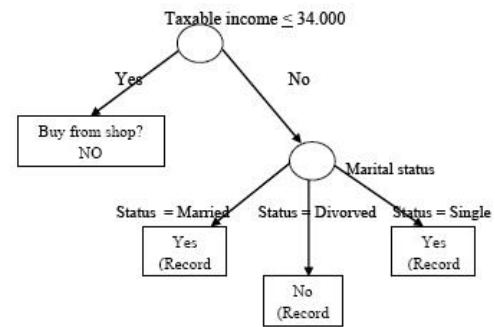
Dalam penelitian ini akan memanfaatkan *data mining* untuk mengklasifikasi data mahasiswa dengan jenjang pendidikan D3 di Fakultas Teknik Universitas Pandanaran.

**c. Algoritma Decision Tree ID3**

Algoritma ID3 merupakan singkatan dari *iterative Dichotomiser 3* atau *Induction of Decision Tree*. Algoritma ID3 dicetuskan pertama kali pada akhir decade 70-an oleh Ross Quinian. Algoritma ID3 bertujuan untuk mendapatkan pohon keputusan sederhana yang terbaik, yang konsisten dari sekumpulan data. Mekanisme konstruksi dilakukan secara *top-*

*down* dengan memilih atribut mana yang harus diperiksa dapat *root* dari *decision tree* [14].

*Decision Tree* atau pohon keputusan adalah model prediksi menggunakan struktur pohon atau hirarki dengan mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.



Gambar 1. Contoh Konsep Keputusan Sederhana[15]

Algoritma ID3 merupakan sebuah metode yang digunakan untuk membangkitkan pohon keputusan Secara ringkas, langkah kerja Algoritma ID3 dapat digambarkan sebagai berikut [16]:

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-]$$

Dimana :

- S adalah ruang (data) sample yang digunakan untuk training.
- P+ adalah jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.

$$Entropy(S,A) = Entropy(s) \sum \frac{S_v}{S} Entropy(S_v)$$

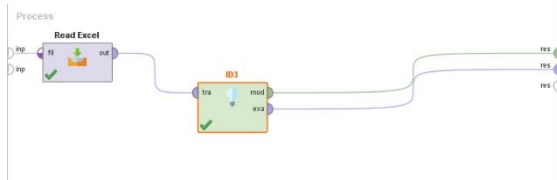
Dimana:

- S adalah ruang (data) sample yang digunakan untuk training.
- A adalah atribut
- V adalah suatu nilai yang mungkin untuk atribut A
- Nilai(A) adalah himpunan yang mungkin atribut A
- Sv adalah Jumlah sampel untuk nilai V
- S adalah jumlah seluruh sampel data
- Entropy (Sv) adalah entropy untuk sampel yang memiliki nilai V

**d. Kerangka Pemikiran**

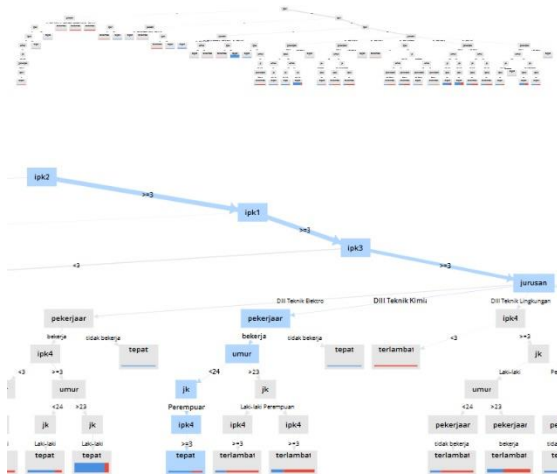
Kerangka pemikiran yang dilakukan adalah:





Gambar 4. Proses Olah Data dengan *Decision Tree ID3*

Sehingga menghasilkan model pohon keputusan yaitu sebagai berikut:



Gambar 5. Model Pohon Keputusan Kelulusan Mahasiswa

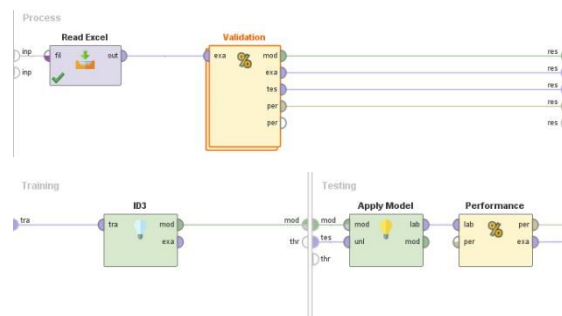
Tujuan utama dari menganalisis data dengan menggunakan algoritma *Decision Tree* ini adalah ingin mendapatkan rule [12], yang akan dimanfaatkan untuk pengambilan keputusan pada data baru. Adapun rule yang didapat adalah:

1. Jika  $IPK2 < 3$  dan Jurusan Elektronika dan  $ipk3 < 3$  dan  $ipk4 < 3$  dan umur  $> 23$  dan sudah bekerja dan  $ipk1 < 3$  maka mahasiswa lulus tepat waktu
2. Jika  $IPK2 < 3$  dan Jurusan Elektronika dan  $ipk3 < 3$  dan  $ipk4 \geq 3$  maka mahasiswa lulus tepat waktu
3. Jika  $IPK2 < 3$  dan Jurusan Elektronika dan  $ipk3 \geq 3$  maka mahasiswa lulus tepat waktu
4. Jika  $IPK2 < 3$  dan Jurusan Kimia  $ipk3 \geq 3$  maka mahasiswa lulus terlambat
5. Jika  $IPK2 < 3$  dan Jurusan Lingkungan maka mahasiswa lulus terlambat
6. Jika  $IPK2 < 3$  dan Jurusan Mesin maka mahasiswa lulus terlambat
7. Jika  $IPK2 < 3$  dan Jurusan Sipil dan  $ipk1 < 3$  maka mahasiswa lulus terlambat
8. Jika  $IPK2 < 3$  dan Jurusan Sipil dan  $ipk1 \geq 3$  maka mahasiswa lulus terlambat
9. Jika  $IPK2 \geq 3$  dan  $ipk1 < 3$  dan Jurusan Elektronika maka mahasiswa lulus tepat waktu
10. Jika  $IPK2 \geq 3$  dan  $ipk1 < 3$  dan Jurusan Kimia maka mahasiswa lulus terlambat
11. Jika  $IPK2 \geq 3$  dan  $ipk1 < 3$  dan Jurusan Sipil dan  $ipk3 < 3$  maka mahasiswa lulus terlambat
12. Jika  $IPK2 \geq 3$  dan  $ipk1 < 3$  dan Jurusan Sipil dan  $ipk3 \geq 3$  maka mahasiswa lulus tepat waktu
13. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan umur  $< 24$  maka mahasiswa lulus tepat waktu
14. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan umur  $> 23$  maka mahasiswa lulus terlambat
15. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan sipil dan  $ipk4 < 3$  maka mahasiswa lulus tepat waktu
16. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan sipil dan  $ipk4 \geq 3$  dan umur  $> 23$  dan jenis kelamin laki-laki dan sudah bekerja maka mahasiswa lulus terlambat
17. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan sudah bekerja dan  $ipk4 < 3$  dan umur  $> 23$  dan jenis kelamin laki-laki maka mahasiswa lulus tepat waktu
18. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan sudah bekerja dan  $ipk4 \geq 3$  dan umur  $< 24$  dan jenis kelamin laki-laki maka mahasiswa lulus tepat waktu
19. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan sudah bekerja dan  $ipk4 \geq 3$  dan umur  $> 23$  dan jenis kelamin laki-laki maka mahasiswa lulus tepat waktu
20. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan elektro dan tidak bekerja maka mahasiswa lulus tepat waktu
21. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan kimia dan sudah bekerja dan umur  $< 24$  dan jenis kelamin perempuan dan  $ipk4 \geq 3$  maka mahasiswa lulus tepat waktu
22. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan kimia dan sudah bekerja dan umur  $> 23$  dan jenis kelamin laki-laki dan  $ipk4 \geq 3$  maka mahasiswa lulus terlambat
23. Jika  $IPK \geq 3$  dan  $ipk1 \geq 3$  dan  $ipk3 < 3$  dan jurusan kimia dan sudah bekerja dan umur  $> 23$  dan jenis kelamin perempuan

- dan ipk4  $\geq 3$  maka mahasiswa lulus terlambat
24. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan kimia dan tidak bekerja maka mahasiswa lulus tepat waktu
  25. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan lingkungan dan ipk4  $< 3$  maka mahasiswa lulus terlambat
  26. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan lingkungan dan ipk4  $\geq 3$  dan jenis kelamin laki-laki dan umur  $< 24$  dan tidak bekerja maka mahasiswa lulus terlambat
  27. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan lingkungan dan ipk4  $\geq 3$  dan jenis kelamin laki-laki dan umur  $> 23$  dan sudah bekerja maka mahasiswa lulus terlambat
  28. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan lingkungan dan ipk4  $\geq 3$  dan jenis kelamin perempuan dan umur  $< 24$  dan tidak bekerja maka mahasiswa lulus tepat waktu
  29. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan lingkungan dan ipk4  $\geq 3$  dan jenis kelamin perempuan dan umur  $> 23$  dan sudah bekerja maka mahasiswa lulus terlambat
  30. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan mesin dan sudah bekerja dan umur  $< 24$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus tepat waktu
  31. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan mesin dan sudah bekerja dan umur  $> 24$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus tepat waktu
  32. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan mesin dan tidak bekerja dan umur  $< 24$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus terlambat
  33. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan mesin dan tidak bekerja dan umur  $> 23$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus terlambat
  34. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan sipil dan sudah bekerja dan umur  $< 24$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus tepat waktu
  35. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan sipil dan sudah bekerja dan umur  $< 24$  dan jenis kelamin perempuan maka mahasiswa lulus tepat waktu

36. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan sipil dan sudah bekerja dan umur  $> 23$  dan jenis kelamin laki-laki dan ipk4  $\geq 3$  maka mahasiswa lulus tepat waktu
37. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan sipil dan sudah bekerja dan umur  $> 23$  dan jenis kelamin perempuan dan ipk4  $\geq 3$  maka mahasiswa lulus tepat waktu
38. Jika IPK  $\geq 3$  dan ipk1  $\geq 3$  dan ipk3  $< 3$  dan jurusan sipil dan tidak bekerja maka mahasiswa lulus tepat waktu

Setelah mendapatkan model dan rule, langkah selanjutnya pengujian algoritma *Decision tree* terhadap data kelulusan mahasiswa dengan *K-Fold Cross Validation*. Dalam pengujian *K-Fold Cross Validation* algoritma *Decision Tree ID3*, peneliti menggunakan 10 kali percobaan dengan sampling type Stratified (bertingkat-tingkat).



Gambar 6. Proses Pengujian

Hasil dari proses pengujian diatas menghasilkan *confision matrix* sebagai berikut:

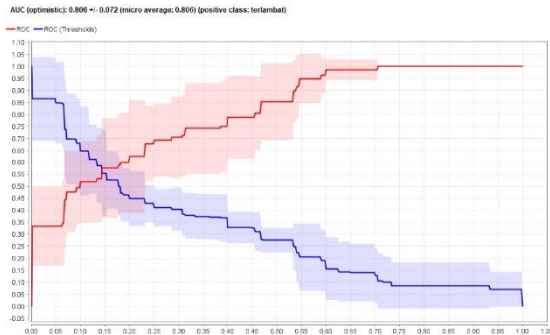
accuracy: 73.19% +/- 9.72% (micro average: 73.19%)

	true tepat	true terlambat	class precision
pred. tepat	130	42	75.58%
pred. terlambat	21	42	66.67%
class recall	86.09%	50.00%	

Gambar 7. *confision matrix*

Jumlah *True Positive* (TP) adalah 130 *record* diklasifikasikan sebagai TEPAT terpilih dan *False Negative* (FN) sebanyak 42 *record* diklasifikasikan sebagai TEPAT terpilih tetapi TERLAMBAT terpilih. Berikutnya 42 *record* untuk *True Negative* (TN) diklasifikasikan sebagai TERLAMBAT terpilih, dan 21 *record* *False Positive* (FP) diklasifikasikan sebagai TERLAMBAT terpilih ternyata TEPAT. Sehingga nilai akurasi adalah 73,19%.

Dan dari *confision matrix* diatas juga menghasilkan *performance* keakurasian dengan nilai AUC (*Area Under Curve*) sebesar 0.806 dengan nilai akurasi Baik.



Gambar 8. Grafik confusion matrix

### KESIMPULAN

Dalam penelitian ini dilakukan pengolahan data kelulusan mahasiswa Universitas Pandanaran dengan menggunakan algoritma *Decision Tree* sehingga menghasilkan model dan rule. Dan dari hasil model kelulusan mahasiswa tersebut dilakukan evaluasi algoritma sehingga menghasilkan nilai akurasi 73,19 % dengan nilai AUC 0,806 dan termasuk klasifikasi data Baik.

### UCAPAN TERIMA KASIH

Kami sangat berterima kasih kepada Kemenristek-Dikti atau Kementerian Riset dan Pendidikan Tinggi Indonesia (DPRM-DIKTI) yang membiayai penelitian.

### REFERENSI

[1] Y. Asriningtias *et al.*, “Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa,” *J. Inform.*, vol. 8, no. 1, pp. 837–848, 2014.

[2] M. Ridwan, H. Suyono, and M. Sarosa, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier,” *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013.

[3] D. Himawan, “Aplikasi Data Mining Menggunakan Algoritma ID3 Untuk Mengklasifikasi Kelulusan Mahasiswa Pada Universitas Dian Nuswantoro Semarang,” *Fak. Ilmu Komput.*, 2011.

[4] Muhammad Sholeh, “Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X,” *Snast*, no. November, pp. 211–216, 2014.

[5] H. Romadhona, Agus; suprapedi; himawan, “Prediksi Kelulusan Tepat Waktu Mahasiswa Stmik-Ymi,” *J.*

*Teknol. Inf.*, vol. 13, no. 1, pp. 69–83, 2017.

[6] E. Marselina Silvia Suhartinah, “Graduation Prediction Of Gunadarma University Students Using Algorithm And Naive Bayes C4.5 Algorithm,” *Fac. Ind. Technol. Gunadarma Univ.*, 2010.

[7] E. S. Siska Haryati, Aji Sudarsono, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.

[8] Azwar, *Penyusunan Skala Psikologi*. Yogyakarta: Pustaka Pelajar, 2014.

[9] S. T. Karamouzis and A. Vrettos, “An Artificial Neural Network for Predicting Student Graduation Outcomes,” *Lect. Notes Eng. Comput. Sci.*, vol. 2173, no. 1, pp. 991–994, 2008.

[10] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, “Application of k Means Clustering algorithm for prediction of Students Academic Performance,” vol. 7, pp. 292–295, 2010.

[11] J. Yingkuachat, P. Praneetpolgrang, and B. Kijisirikul, “An Application of the Probabilistic Model to the Prediction of Student Graduation Using Bayesian Belief Networks,” *Electr. Eng. Comput. Telecommun. Inf. Technol. Assoc. Thail. (ECTI Thailand)*, pp. 63–71, 2007.

[12] I. H. Witten, “Data Mining Data Mining Complications: Overfitting Statistical modeling One attribute does all the work?,” 2007.

[13] Budi Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.

[14] R. T. Vulandari, *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media, 2017.

[15] F. Gorunescu, *Data Mining: Concepts, Models and Techniques (Intelligent Systems Reference Library)*. 2011.

[16] I. R. Munthe and V. Sihombing, “Klasifikasi Algoritma Iterative Dichotomizer ( ID3 ) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer,” *Jutikomp*, vol. 1, no. 2, pp. 27–34, 2018.