



## Prediksi Kesiapan Kerja Mahasiswa menggunakan Algoritma K-Means dan C4.5.

Hendri Noviyanto<sup>✉1</sup>, Bayu Mukti<sup>2</sup>

Universitas Surakarta<sup>1,2</sup>

email: [hendrinoviyantoo@gmail.com](mailto:hendrinoviyantoo@gmail.com)<sup>1</sup>, [bayu.unsa@gmail.com](mailto:bayu.unsa@gmail.com)<sup>2</sup>

Received 11 Agustus 2022, Accepted 21 September 2022, Published 30 September

### Abstrak

Kesiapan kerja mahasiswa merupakan hal yang cukup berpengaruh terhadap eksistensi sebuah Perguruan Tinggi. Waktu tunggu yang diperlukan mahasiswa dalam mendapatkan pekerjaan dapat mempengaruhi mental mahasiswa dan nilai Perguruan Tinggi dari asumsi masyarakat. Hal ini akan sangat berdampak pada minat orang tua ataupun calon mahasiswa untuk melanjutkan pendidikan di Perguruan Tinggi yang memiliki citra kurang baik. Oleh karena itu, prediksi kesiapan kerja mahasiswa sebelum lulus sangatlah diperlukan guna bahan pertimbangan Perguruan Tinggi untuk mengatasi masalah waktu tunggu kerja mahasiswa setelah lulus. Sumber data penelitian ini didapatkan dari basisdata Universitas Surakarta dengan memanfaatkan data alumni dari tracer study sebagai data train dan data mahasiswa aktif semester 6 sebagai data test. Langkah awal yang dilakukan adalah preprocess untuk menghilangkan noise yang dapat mengganggu atau mempengaruhi hasil akhir. Metode penelitian yang akan digunakan adalah mengimplementasikan algoritma K-Means dan C4.5 untuk proses pengelompokan dan prediksi. Data train yang digunakan sebanyak 150 data dan data testing sebanyak 59 data. Hasil yang didapatkan algoritma K-Means dapat mengklusterkan 143 data dengan tepat yang dibandingkan dengan data asli. Nilai kluster terbaik yang didapatkan adalah  $K = 3$ . Nilai akurasi yang didapatkan pada algoritma C4.5 sebesar 100% dan Naïve Bayes sebesar 91.53%. Hal ini membuktikan bahwa teknik klustering dapat membantu meningkatkan nilai akurasi pada teknik klasifikasi.

**Kata Kunci:** *Data Mining; Kesiapan Kerja; K-Means; C4.5; Prediksi.*

### Abstract

*Student work readiness is quite influential on the existence of a university. The waiting time required by students to get a job can affect the mentality of students and the value of higher education from the assumptions of society. This will greatly affect the interest of parents or prospective students to continue their education at universities that have a bad image. Therefore, predictions of student work readiness before graduation are needed for consideration by higher education institutions to overcome the problem of waiting time for student work after graduation. The source of this research data is obtained from the Surakarta University database by utilizing alumni data from tracer studies as train data and 6-semester active student data as test data. The initial step taken is preprocessing to eliminate noise that can interfere with or affect the final result. The research method that will be used is to implement the K-Means and C4.5 algorithms for grouping and prediction processes. The data train used is 150 data and the testing data is 59 data. The results obtained by the K-Means algorithm can cluster 143 data correctly by comparing with the original data. The best cluster value obtained is  $K = 3$ .*

*The accuracy value obtained in the C4.5 algorithm is 100% and Naïve Bayes is 91.53%. This proves that the clustering technique can help increase the accuracy value of the classification technique.*

**Keywords:** *Data Mining; K-Means; C4.5; Prediction; Working Readiness.*

✉Corresponding author

## PENDAHULUAN

Sebuah Perguruan Tinggi menyimpan banyak sekali data yang sifatnya bervariasi dan jumlahnya sangatlah besar. Sebuah data yang berukuran sangat besar jika hanya dibiarkan saja akan menjadi data yang tidak banyak manfaatnya [1]. Database mahasiswa aktif dan alumni menyimpan banyak sekali informasi yang bisa dieksplorasi menggunakan teknik pengolahan data tertentu. Data yang bisa dieksplorasi salah satunya adalah informasi tentang kesiapan kerja mahasiswa. Proses pengolahan data tersebut dikenal dengan proses *Data Mining* (DM). DM melakukan proses pengolahan data dengan mencari *Knowledge Discovery in Database* (KDD) dari serangkaian langkah yang dilakukan [2].

*Tracer Study* merupakan sebuah alat yang digunakan untuk melakukan palacakan terhadap lulusan terkait dengan informasi kompetensi pendidikan di dunia industri profesional dan menjadi salah satu syarat akreditasi. Tolak ukur keberhasilan sebuah Perguruan Tinggi adalah tesorapnya alumni di dunia industri secara cepat dan tepat [3,4]. Selain digunakan sebagai salah satu syarat akreditasi, Ditjen Dikti Kemenristekdikti [5] menggunakan *tracer study* untuk menilai relevansi pendidikan tinggi terhadap informasi kebutuhan dunia industri. Kesiapan kerja mahasiswa dalam dunia industri melibatkan beberapa faktor yang dapat mempengaruhi proses belajar untuk kesiapan kerja. Proses prediksi kesiapan kerja diharapkan mampu membantu Perguruan Tinggi dalam mengavaluasi mahasiswa.

Penelitian terkait dengan prediksi kesiapan kerja mahasiswa dilakukan dengan merujuk pada penelitian prediksi waktu tunggu alumni dalam mendapatkan pekerjaan, Pada penelitian ini dilakukan kombinasi teknik yaitu *Supervised Learning* dan *Unsupervised Learning*. Penelitian dengan teknik *Supervised Learning* yang dilakukan oleh [6,7] dengan menerapkan algoritma C4.5 untuk memprediksi masa tunggu kerja alumni. Penelitian [8] menggunakan *Naïve Bayes* dengan *Backward Elimintaion* dan terbukti dapat meningkatkan akurasi, [9,10] menggunakan ANN, [11] menggunakan *logistic regression*. Penelitian dengan teknik *Unsupervised Learning* dilakukan oleh [12][13][14][15] dengan mengimplementasikan algoritma *K-Means*.

Penelitian [16] melakukan pengelompokan lulusan pendidikan tinggi vokasi menggunakan algoritma *K-Means*. Model yang telah dibuat dievaluasi berdasarkan nilai *Davies Bouldin Indek* (DBI) sebesar 0,221 pada nilai  $k = 6$ . Proses pemodelan pengelompokan pada penelitian ini dibantu dengan perangkat lunak RapidMiner. Penelitian [17] melakukan pengelompokan potensi kelulusan mahasiswa menggunakan algoritma *K-Means*. Penelitian [12] mengimplementasikan algoritma *K-*

*Means* untuk mengelompokkan kinerja akademik mahasiswa. Proses evaluasi *clustering* di dapatkan nilai  $k = 3$  memiliki nilai *silhouette coefficient* yang paling mendekati  $S_i = 1$ . Pada penelitian ini menghasilkan sebuah *rule* bahwa penerimaan mahasiswa akan di dominasi oleh jalur reguler dan prestasi akademik.

Penelitian yang dilakukan [9] menerapkan algoritma *Artificial Neural Network* (ANN) untuk memprediksi tingkat kelulusan mahasiswa dengan perolehan akurasi sebesar 92,3%. Pada penelitian ini, dataset yang digunakan adalah data *demographic* mahasiswa yang didapatkan dari *UCL Machine Learning Repository* dengan variabel *assignment*, *tests*, dan *final exam*. Penelitian [18] menerapkan algoritma ANN menggunakan 11 variabel input dan memperoleh akurasi sebesar 84,8%. Penelitian yang dilakukan [11] menggunakan metode *Regresi Logistik* untuk mengetahui faktor apa saja yang mempengaruhi nilai Indeks Prestasi Kumulatif (IPK) mahasiswa. IPK diketahui merupakan salah satu komponen penting untuk melihat keberhasilan mahasiswa. Hasil dari penelitian ini adalah tempat tinggal memiliki pengaruh terhadap IPK menggunakan teknik Uji Parsial. Penelitian [6] melakukan klasifikasi masa tunggu alumni menggunakan algoritma C4.5 dengan fitur *forward selection* dengan nilai akurasi sebesar 80,37%. Pada penelitian ini, bertujuan untuk melihat kemampuan *fitur forward selection* yang ternyata lebih baik dibandingkan dengan C4.5 tunggal. Penelitian [7] melakukan klasifikasi kompetensi alumni menggunakan algoritma C4.5 dan didapatkan akurasi sebesar 82,86%. Penelitian [3] melakukan analisis klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) dan *Backpropagation Neural Network* (BPNN), hasilnya diperoleh nilai akurasi sebesar 83,33 (BPNN) dan 83,00% (SVM). Penelitian [8] melakukan prediksi waktu tunggu alumni menggunakan *Naiïve Bayes* dengan *Backward Elimination*. Hasil dari penelitian ini menunjukkan bahwa *Naiïve Bayes* dengan fitur seleksi *Backward Elimination* mampu meningkatkan nilai akurasi sebesar 68,52%.

Metode populer yang sering digunakan oleh peneliti terdahulu adalah algoritma *K-means* untuk proses pengelompokan data. Pada penelitian ini, peneliti mengusulkan teknik kombinasi metode *Unsupervised Learning* (*K-Means*) dan *Supervised Learning* (*C4.5*) untuk mencari tingkat kesiapan kerja mahasiswa. Data *Training* pada penelitian ini adalah data alumni mahasiswa yang didapatkan dari *Tracer Study*, sedangkan data *testing* adalah data mahasiswa semester 6 yang didapatkan dari database Universitas. Kombinasi dari metode ini dibuat untuk mendapatkan hasil maksimal dari proses penelitian yang dilakukan.

## METODOLOGI

### Penentuan Attribute

Pada penelitian ini menggunakan beberapa atribut untuk proses *data mining* yang telah ditentukan untuk kebutuhan analisis. Adapun daftar atribut yang digunakan pada penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Keterangan Atribut

Atribut	Variabel
Gender	X1
IPK	X2
Sertifikat Keahlian	X3
Nilai Magang	X4
Pengalaman Organisasi	X5
Kemampuan Ekonomi	X6
Kelas	Y1

Pada Tabel 1 terdapat atribut dengan variabel X1 sampai X6 merupakan data input yang akan digunakan untuk proses klusterisasi dan prediksi. Atribut kelas dengan variabel Y1 merupakan atribut yang akan digunakan untuk proses prediksi menggunakan teknik klasifikasi. Kelas disini atau biasa dikenal dengan label pada proses klasifikasi ditentukan dari hasil klustering untuk menentukan label terbaik. Atribut yang digunakan ditentukan dari beberapa hasil studi literatur [19][15] dan diskusi dengan beberapa alumni serta ketua program studi.

### Pengumpulan Data

Pada penelitian ini data alumni yang digunakan sebanyak 150 data mahasiswa Fakultas Teknik Elektro dan Informatika Universitas Surakarta yang telah lulus selama 3 tahun. Sedangkan jumlah data yang digunakan untuk testing sebesar 59 data yang didapatkan dari mahasiswa aktif.

### Analisis Data

Tahap pertama pada penelitian ini adalah mengklusterkan dataset alumni mahasiswa menggunakan algoritma K-Means. Setelah mendapatkan hasil kluster, kemudian dibandingkan dengan data nyata sehingga dapat diketahui tingkat keberhasilan proses klusterisasi sebelum data dilabeli.

*K-Means* merupakan salah satu algoritma *clustering* / pengelompokan data yang bersifat *Unsupervised Learning* yang berarti input data pada algoritma ini menerima data tanpa label kelas [20]. Karakteristik dari algoritma ini adalah sebagai berikut :

1. Memiliki  $n$  buah data.
2. Input berupa jumlah data dan jumlah *cluster*
3. Pada setiap *cluster* memiliki sebuah *centroid* yang mempresentasikan *cluster* tersebut.

Secara sederhana algoritma *K-means* berjalan sebagai berikut [17] :

1. Pilih  $K$  (nilai bebas) buah titik *centroid*.
2. Membangkitkan nilai random untuk pusat *cluster* awal (*centroid*) sebanyak  $k$ .
3. Menghitung jarak data dengan *centroid* menggunakan rumus jarak *Euclidean Distance* hingga ditemukan jarak yang paling dekat.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - \mu_j)^2} \quad (1)$$

Dimana :

$d(x, y)$  : jarak antara data x ke data y

$x_i$  : data testing ke - i

$y_i$  : data training ke - i

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid* (jarak terkecil)
5. Memperbaharui nilai *centroid* baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus.

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x^i \quad (2)$$

Dimana :

$\mu_k$  : titik *centroid* dari cluster ke-k

$N_k$  : banyaknya data pada cluster ke -k

$x^i$  : data ke-I pada cluster ke-k

6. Ulangi langkah 3 sampai 5 sampai nilai dari titik *centroid* tidak lagi berubah.

Tahap kedua setelah proses melakukan klusterisasi dataset alumni, data tersebut dilabeli dan digunakan sebagai data training. Data tersebut digunakan untuk membantu mengklasifikasikan data mahasiswa pada semester 6. Proses prediksi menggunakan algoritma C4.5.

Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini memiliki input berupa training sample. Training sampel merupakan contoh data yang akan dimanfaatkan untuk membangun sebuah tree yang telah diuji kebenarannya [21]

## HASIL DAN PEMBAHASAN

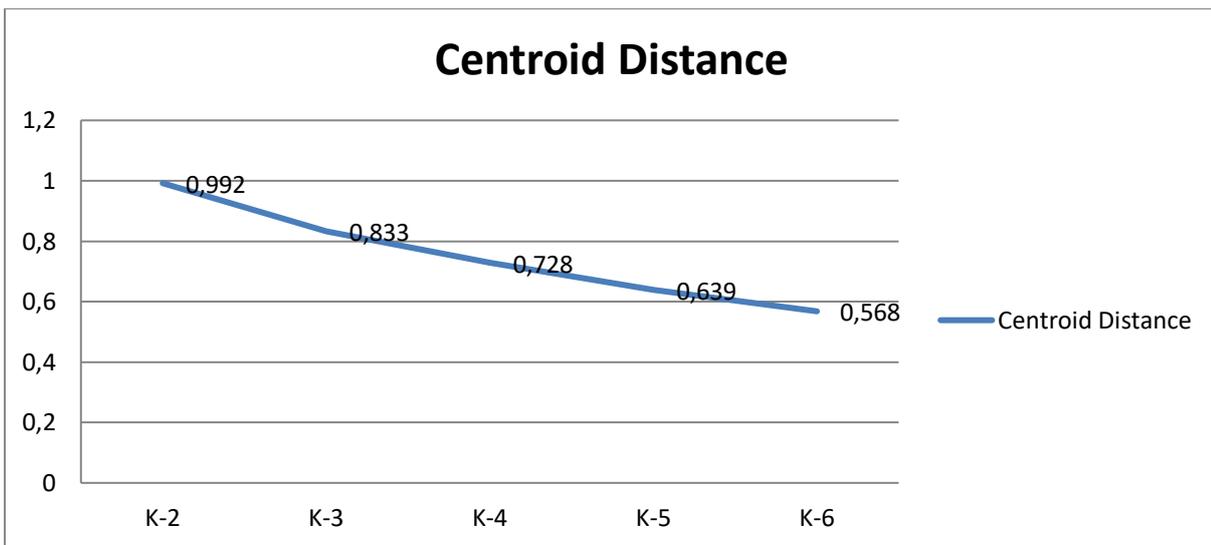
Pada penelitian ini proses klusterisasi dan klasifikasi dibantu oleh software RapidMiner untuk prediksi kesiapan kerja mahasiswa. Proses klustering menggunakan algoritma K-Means, kemudian setelah mendapatkan label dilanjutkan dengan proses prediksi menggunakan algoritma C4.5. Skema pengujian proses klustering menggunakan K dari 2-6 dan melihat K terbaik, setelah K didapatkan maka data hasil kluster dibandingkan dengan data nyata sebelum dilabeli. Data training dan data testing pada penelitian ini sudah ditentukan jumlahnya yaitu 150 (train) dan 59 (testing). Dataset yang digunakan pada penelitian ini adalah data yang berasal dari database Universitas Surakarta dengan jumlah atribut sebanyak 6 attribute.

Dalam penelitian ini, terdapat beberapa skenario pengujian, seperti proses komparasi antara metode prediksi yang digunakan. Skenario pengujian pertama adalah proses klustering untuk mengelompokkan data ke dalam karakteristik yang sama. Skenario pengujian kedua adalah mengklasifikasikan data

menggunakan algoritma C4.5 yang dikomparasikan dengan algoritma naïve bayes untuk membandingkan nilai akurasi yang didapatkan.

### Pengujian menggunakan algoritma K-Means

Hasil pengujian menggunakan algoritma K-Means dengan melihat Centroid Distance dapat dilihat pada Gambar 1.



Gambar 1. Nilai Pengujian K-Means

Pada Gambar 1 diatas, jumlah K terbaik diperoleh pada K = 3 dengan nilai 0.833 metode yang digunakan untuk mengetahui K terbaik adalah Elbow Method yaitu dengan melihat garis yang paling bengkok. Jumlah K ini menjadi rujukan untuk pelabelan data yang telah diklusterkan.

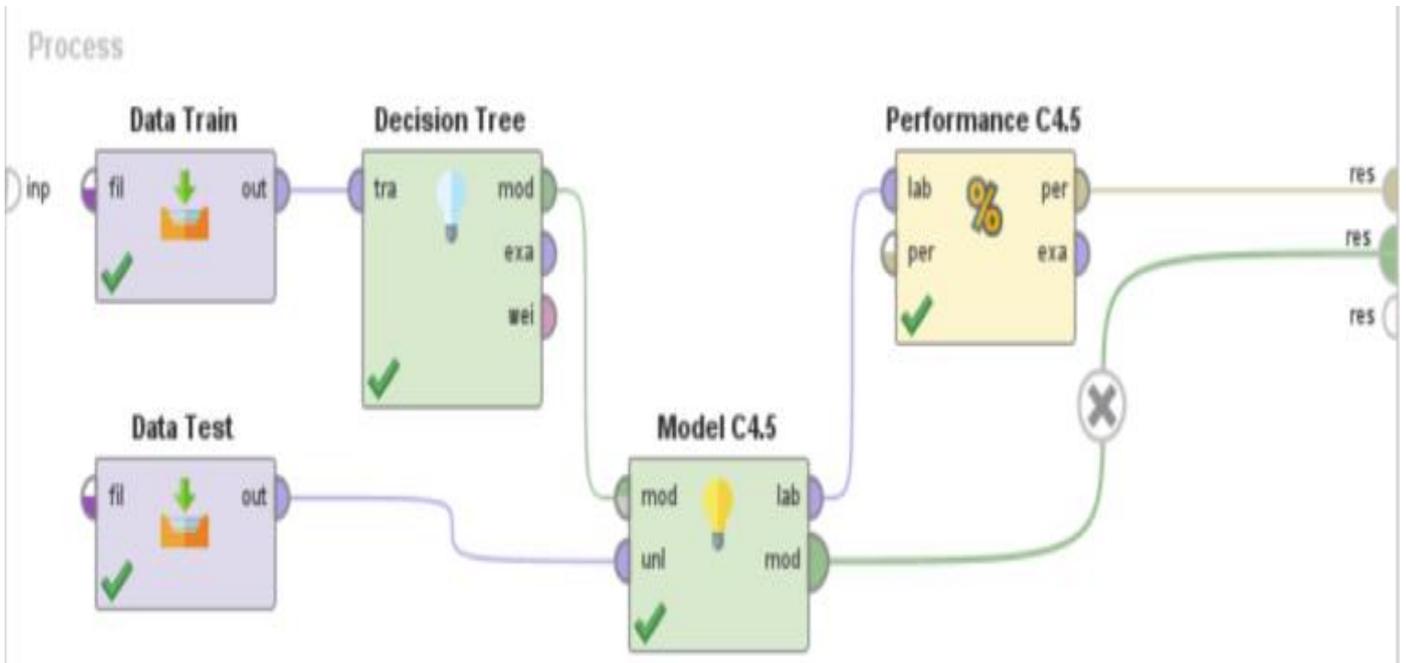
Tabel 2 Perbandingan Data Hasil Kluster

	Data Asli	Data Hasil Kluster
K-3	150	143
Gagal Mengkluster		7

Pada Tabel 2 Perbandingan data hasil kluster dengan data yang sesungguhnya memiliki selisih 7 data yang tidak dapat diklusterkan dengan tepat. Hal ini dipengaruhi dari beberapa faktor, faktor tersebut bisa berasal dari input data yang memiliki karakter nilai sama atau memiliki kesamaan input dengan variabel yang berbeda.

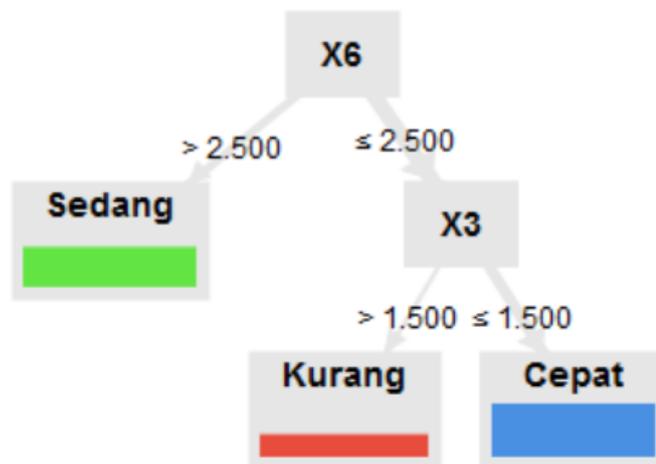
### Pengujian proses klasifikasi

Pengujian kedua adalah proses klasifikasi menggunakan algoritma C4.5 yang dikomparasikan dengan metode lain seperti Naïve Bayes.



Gambar 2. Rancangan Sistem Pembelajaran

Pada Gambar 2, merupakan rancangan sistem pada rapidminer yang menggambarkan proses prediksi menggunakan algoritma C4.5. Hasilnya memiliki nilai akurasi yang cukup bagus yaitu 100% yang dapat dilihat pada tabel 2 dibawah ini. Tree yang dibentuk menampilkan attribut yang cukup berpengaruh terhadap kesiapan kerja dari mahasiswa, yaitu attribute X6 sebagai root dan X3 attribut penting dibawahnya (lihat gambar 3). Attribut tersebut adalah keadaan ekonomi dan kepemilikan sertifikat keahlian (lihat tabel 1).



Gambar 3. Root Hasil Decision Tree

Dari Gambar 3 dapat dilihat, komponen terpenting dalam kesiapan kerja mahasiswa adalah Variabel X6 yang mewakili latar belakang ekonomi dan X3 yang mewakili kepemilikan sertifikat keahlian. Latar belakang ekonomi menjadi root atau variabel yang sangat penting karena cukup berpengaruh terhadap kesejahteraan dan ketenangan dalam menuntut ilmu. Faktor lainnya adalah sertifikat keahlian, dari hasil diskusi dengan beberapa subject penelitian (alumni), sertifikat ini sangat penting untuk menunjang keilmuan seorang mahasiswa.

Tabel 2. Nilai Akurasi Hasil Proses Klasifikasi

	Algoritma	
	C4.5	Naïve Bayes
Dataset	100%	91,53%

Pada Tabel 2 diatas, dapat dilihat nilai akurasi yang didapatkan pada masing-masing algoritma. Klasifikasi yang ditugaskan pada algoritma C4.5 mendapatkan nilai akurasi sebesar 100%, dan nilai akurasi pada Naïve Bayes sebesar 91,53%.Melihat hasil tersebut, teknik klustering yang dipadukan dengan teknik prediksi sangat bermanfaat sekali. Hal ini dapat meminimalisir kesalahan label dan meningkatkan akurasi yang didapatkan algoritma pengklasifikasi.

## SIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan maka dapat disimpulkan bahwa proses klustering menggunakan algoritma K-Means cukup baik dan akurat dengan nilai  $K = 3$ , meskipun di data yang berbeda akan berbeda pula  $K$  optimalnya. Hasil kluster dibandingkan dengan data nyata pada dataset alumni dan ditemukan tingkat kemiripan data yang cukup besar dengan selisih hanya 7 data yang gagal diklusterkan. Proses klasifikasi menggunakan data alumni sebagai data train dan data mahasiswa semester 6 sebagai data testing. Tingkat akurasi yang didapatkan sebesar 100% untuk algoritma C4.5 dan 91,53% untuk algoritma Naïve Bayes. Penelitian ini mengkombinasikan teknik klustering dan prediksi untuk meningkatkan hasil.

## DAFTAR PUSTAKA

- [1] H. Noviyanto and B. Mukti, "Period Study Accuracy Prediction using Sequential Minimal Optimization Algorithm," *Sink. J. dan Penelit. Tek. Inform.*, vol. 5, no. 1, pp. 164-169, 2020, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/10621>.
- [2] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [3] D. I. Purnama, R. L. Islami, L. Sari, and P. R. Sihombing, "Analisis Klasifikasi Data

- Tracer Study Dengan Support Vector Machine Dan Neural Network," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 2, pp. 46–52, 2021, doi: 10.47970/siskom-kb.v4i2.191.
- [4] I. M. B. Adnyana, "Implementasi Naïve Bayes Untuk Memprediksi Waktu Tunggu Alumni Dalam Memperoleh Pekerjaan," *Semin. Nas. Teknol. Komput. Sains*, vol. 1, no. 1, pp. 131–134, 2020, [Online]. Available: <http://prosiding.seminar-id.com/index.php/sainteks/article/view/418>.
- [5] Ristekdikti, *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi*. Jakarta: Ristekdikti, 2016.
- [6] F. Rezkika, B. N. Sari, and A. Susilo, "Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5," *Progresif J. Ilm. Komput.*, vol. Vol. 17, pp. 95–106, 2021.
- [7] R. Cahyaningtyas, L. Luqman, and Y. I. H., "Klasifikasi Kompetensi Alumni Berdasarkan Masa Tunggu Alumni untuk Mendapatkan Pekerjaan Menggunakan Metode Agoritma C4.5," *Kilat*, vol. 9, no. 2, pp. 297–310, 2020, [Online]. Available: <https://stt-pln.e-journal.id/kilat/article/view/950>.
- [8] A. Widhiantoyo, "Penerapan Algoritma Naïve Bayes Dengan Backward Elimination Untuk Prediksi Waktu Tunggu Alumni Mendapatkan Pekerjaan," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 3, pp. 145–151, 2021, doi: 10.33387/jiko.v4i3.3272.
- [9] A. J. Kehinde, A. E. Adeniyi, and R. O. Ogundokun, "Prediction of Students' performance with Artificial Neural Network using Demographic Traits," pp. 1–10.
- [10] V. K. Pal and V. K. K. Bhatt, "Performance prediction for post graduate students using artificial neural network," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7, pp. 446–454, 2019.
- [11] Y. Tampil, H. Komaliq, and Y. Langi, "Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado," *d'CARTESIAN*, vol. 6, no. 2, p. 56, 2017, doi: 10.35799/dc.6.2.2017.17023.
- [12] F. N. R. F. Aziz, B. D. Setiawan, and I. Arwani, "Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 6, pp. 2243–2251, 2018.
- [13] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.
- [14] S. Suryadi, "Penerapan Metode Clustering K-Means Untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi," *J. Inform.*, vol. 6, no. 1, pp. 52–72, 2019, doi: 10.36987/informatika.v6i1.738.
- [15] Y. Amri, "Metode k-means untuk clustering mahasiswa berdasarkan nilai akademik," vol. XV, no. 02, 2021.
- [16] D. R. Utari and F. Sofiani, "Klasterisasi Lulusan Pendidikan Tinggi Vokasi Berbasis Hasil Studi Pelacakan Menggunakan Algoritma K-Means," *J. Serasi*, vol. 19, no. September, 2021, [Online]. Available: <https://journal.budiluhur.ac.id/index.php/serasi/article/view/1543%0Ahttps://journal.budiluhur.ac.id/index.php/serasi/article/download/1543/1034>.
- [17] V. Virtusena, A. Johar, and A. Wijanarko, "Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan Algoritma K-Means," vol. 9, no. 2, 2021.
- [18] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Appl. Sci.*, vol. 1, no. 9, pp. 1–10, 2019, doi: 10.1007/s42452-019-0884-7.
- [19] B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1,

- p. 258, 2021, doi: 10.30865/mib.v5i1.2670.
- [20] W. Ian H and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2 Edition. San Fransisco: Elsevier Inc., 2005.
- [21] S. Yunita and Nurahman, "Metode Klasifikasi Data Mining Algoritma C4.5 dan Part untuk Prediksi Waktu Kelulusan Mahasiswa di Universitas Darwan Ali," *Teknol. infotek J. Inform. dan Teknol.*, vol. 3, no. 9, pp. 1-7, 2020.