

PREDIKSI PENERIMA BEASISWA UANG KULIAH TUNGGAL MENGUNAKAN KOMBINASI ALGORITMA K-MEANS & DECISION TREE.

Ahmat Arifin*, Farrikh Al Zami, Edi Norsasangko

^aUniversitas Dian Nuswantoro Semarang

Jln. Imam Bonjol No.207 Pendrikan Kidul. Kota Semarang

*Email : rifian37@gmail.com

Info Artikel	Abstrak
DOI : https://doi.org/10.26751/jikoma.v5i2.2392	<p>Beasiswa merupakan dambaan setiap mahasiswa atau pelajar yang menuntut ilmu di sekolah atau perguruan tinggi, terutama bagi keluarga yang kurang mampu atau keterbatasan secara ekonomi. Beasiswa dapat meringankan beban keluarga dalam hal biaya Pendidikan atau kuliah. Pada tahun 2020 terjadi wabah Covid-19 yang melanda negara berbagai di dunia. waktu itu sedang terjadi lockdown sehingga masyarakat pada saat itu dibatasi aktifitas kegiatan sehari-hari. Berbagai sector terdampak wabah covid-19 salah satu nya adalah dunia Pendidikan dimana Pemerintah Menggelontorkan Bantuan Beasiswa Uang Kuliah Tunggal untuk menaggulangi wabah penyakit Covid-19 yang melanda Indonesia. Salah satunya di Perguruan Tinggi. Tujuan dari penelitian ini adalah memprediksi agar penyaluran beasiswa ini tepat sasaran pada Mahasiswa Universitas Muhammadiyah Kudus untuk semua program studi dengan student body Mahasiswa 2566. dimana, terdapat 637 mahasiswa yang mengajukan beasiswa dan hanya 157 mahasiswa yang berhak untuk mendapatkan beasiswa. Dimana peneliti mencoba agar penyaluran beasiswa ini tepat sasaran dengan menggunakan kombinasi klastering dan kalsifikasi yaitu Algoritma K-Means dan Decision Tree. Dimana penelitian ini mencoba mengelompokkan terlebih dahulu penerima beasiswa menggunakan K-Means. Sedangkan untuk penentuan penerima beasiswa menggunakan decision tree.dalam pengolahan data dalam penelitian ini menggunakan python.</p>
Article history: Received 2024-05-02 Revised 2024-07-18 Accepted 2024-08-13	
Kata kunci : Kombinasi Algoritma, K-Means, Disions Tree	

Abstract

Scholarships are the dream of every student or student studying at school or college, especially for families who are less fortunate or economically limited. Scholarships can lighten the family's burden in terms of education or college costs. In 2020 there was a Covid-19 outbreak that hit various countries in the world. At that time there was a lockdown so that people at that time were limited in their daily activities. Various sectors have been affected by the Covid-19 outbreak, one of which is the world of education, where the government has provided single tuition scholarship assistance to tackle the Covid-19 disease outbreak that has hit Indonesia. One of them is in Higher Education.. The aim of this research is to predict that the distribution of scholarships will be right on target for Muhammadiyah Kudus University students for all study programs with a student body of 2566 students. Where, there are 637 students who apply for scholarships and only 157 students are entitled to receive

scholarships. Where researchers try to make the distribution of scholarships right on target by using a combination of clustering and calcification, namely the K-Means and Decision Tree algorithms. Where this research tries to first group scholarship recipients using K-Means. Meanwhile, to determine scholarship recipients, we use a decision tree. The data processing in this research uses Python.

This is an open access article under the [CC BY-SA](#) license.

I. PENDAHULUAN

Pada awal tahun 2020 dunia di goncang dengan adanya wabah covid-19 yang berpengaruh pada segala sendi kehidupan, salah satu diantaranya yang sangat terdampak adalah dunia Pendidikan dimana saat itu pembelajaran dilakukan secara daring. Selain Pendidikan juga berdampak langsung pada sektor ekonomi dimana daya beli masyarakat menurun.

Untuk itu pemerintah melalui program untuk menangani dampak covid-19 di sector Pendidikan, pemerintah mengeluarkan kebijakan melalui Menteri Pendidikan dan kebudayaan RI/Kemdikbud Ristek melalui Permendikbud Nomor 10 Tahun 2020 mengeluarkan Program Indonesia Pintar (PIP) yang di prioritaskan bagi mahasiswa yang diterima perguruan tinggi dengan sasaran mahasiswa penyandang disabilitas, keluarga tidak mampu, mahasiswa yang dari panti sosial atau dengan kebutuhan khusus lainnya. Pada semester gasal tahun akademik 2020/2021 Kemdikbud melalui Puslapdik menyalurkan bantuan beasiswa UKT kepada 453.012 mahasiswa dari awal target sebelumnya yang hanya 419.605 mahasiswa dengan alokasi anggaran dari APBN yang hampir mencapai Rp. 1.007.051.316.461,00.

Universitas Muhammadiyah Kudus merupakan salah satu penerima Bantuan Uang Kuliah Tunggal /SPP dengan kuota sebanyak 157 Mahasiswa . Dengan kuota yang terbatas dan pendaftar yang melebihi kuota sebanyak 637 mahasiswa yang mengajukan bantuan UKT/SPP, maka diperlukan sebuah pendekatan untuk memberikan saran keputusan mahasiswa yang mendapatkan bantuan beasiswa dengan lebih tepat.

Beberapa pendekatan penentuan beasiswa sudah diteliti oleh banyak orang, antara lain: 1) penelitian dari Nurul Rohmawati dkk yang bertujuan untuk mengukur kinerja algoritma k-means pada pengelompokan data mahasiswa pelamar beasiswa BBM. Data pelamar tersebut dikelompokkan menjadi tiga

Kelompok berhak menerima, dipertimbangkan dan tidak berhak menerima beasiswa. Algoritma k-means digunakan untuk mengelompokkan data tersebut dan diukur kinerjanya dengan menghitung nilai purity masing- masing cluster. Penelitian ini menggunakan data dari 36 mahasiswa dan menghasilkan tiga dataset dengan format yang berbeda-beda: data atribut kodifikasi sebagian, atribut kodifikasi keseluruhan, dan atribut data asli. Hasil penelitian menunjukkan bahwa algoritma k-means lebih cocok digunakan pada dataset dengan format atribut yang dikodifikasi keseluruhan, dengan nilai purity sebesar 80.56% [1]. Meskipun mendapatkan nilai purity yang tinggi, namun dataset yang dimiliki peneliti tidak mempunyai outlier, dengan kata lain, datanya sudah bersih dari awal.

Penelitian dari Abu Salam dkk Tujuan dari penelitian ini adalah untuk merekomendasikan penerima beasiswa dengan menggunakan algoritma k-means, hasil rekomendasi berupa penempatan data pendaftar beasiswa ke masing-masing kelompok cluster yang dihasilkan. Eksperimen proses clustering dilakukan menggunakan data pendaftar beasiswa PPA dari biro kemahasiswaan udinus tahun 2016 sebanyak 44 pendaftar beasiswa PPA. Melalui seleksi atribut, k-means ini melakukan perhitungan untuk menempatkan setiap data ke cluster yang sudah ditentukan. Sebanyak 154 mahasiswa direkomendasikan

mendapatkan beasiswa PPA sedangkan 287 mahasiswa tidak mendapatkan [2]. Data hasil klasterisasi ini disandingkan dengan hasil evaluasi dari Biro kemahasiswaan, dan didapatkan akurasi 82% penerima beasiswa. Meskipun hasil cukup tinggi, namun pada penelitian Abu Salam dkk, tidak membahas apakah hasil pemilihan dari Biro, sudah sesuai dan tidak ada outlier.

Penelitian dari R. Sovia bertujuan untuk mempermudah proses pemberian beasiswa prestasi dengan menggunakan data mining untuk mengelompokkan siswa berprestasi dan membuat sistem pendukung keputusan untuk perangkingan. Algoritma K-Means digunakan untuk pengelompokan siswa berprestasi berdasarkan lima kriteria: nilai rata-rata siswa, peringkat, nilai sikap, nilai ekstrakurikuler, dan prestasi akademik atau non-akademik.

Penelitian ini dilakukan di SMAN 2 Payakumbuh dengan menggunakan 20 data siswa yang mengajukan beasiswa. Hasil penelitian menunjukkan bahwa dengan menggabungkan algoritma K-Means dan metode Simple Additive Weighting (SAW), proses pengelompokkan dan perangkingan siswa berprestasi menjadi lebih cepat dan lebih tepat sasaran. Dari 20 data siswa yang mengajukan beasiswa, 10 siswa berhasil dikelompokkan sebagai siswa berprestasi dan hanya 6 siswa yang berhak menerima beasiswa [3]. Kekurangan dari penelitian ini adalah tidak memasukkan unsur pendapatan orang tua.

dalam penelitian algoritma K-Means [4] melakukan pengelompokan data (clustering) data siswa untuk menentukan apakah akan berhasil atau tidak, serta bagaimana algoritma C4.5 melakukan prediksi terhadap siswa yang telah dikelompokkan. Dimana pada penelitian tersebut 100 mahasiswa. 32 Mahasiswa Tidak menerima Basiswa sedangkan 68 Mahasiswa menerima beasiswa. diprediksi bahwa 9 % yang akan mendapatkan beasiswa dan 91 % tidak layak mendapat beasiswa. Meskipun membahas gaji bapak ibu, namun Kekurangan pada penelitian tersebut adalah tidak melibatkan pekerjaan orang tua, dimana kadang

pekerjaan orang tua mungkin mempunyai resiko dalam pembayaran uang sekolah.

Dari pembahasan penelitian diatas, dapat disimpulkan bahwa beberapa penelitian terkait tidak membahas: 1) outlier pada data, 2) pendapatan orang tua, 3) pekerjaan orang tua

Oleh karena itu, pada penelitian ini, penulis menggunakan dua pendekatan sekuensial, yaitu K-Means dan Decision Tree. Pendekatan klasterisasi K-Means digunakan untuk mengelompokkan data sehingga mempermudah deteksi outlier, kemudian hasil dari Klasterisasi yang sudah bersih, dimasukkan ke Decision Tree untuk mendapatkan pohon keputusan yang tepat. Untuk dataset yang digunakan pada penelitian ini mempunyai fitur-fitur antara lain: 1) Program Studi, 2) Semester, 3) IPK, 4) Angkatan, 5) Pekerjaan Ayah, 6) Pekerjaan Ibu, 7) Penghasilan Kotor Ayah, 8) Penghasilan Kotor Ibu, 9) Jumlah anggota keluarga.

II. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini yaitu kombinasi algoritma K-Means untuk mengklasterisasi Penerima Beasiswa UKT dan untuk penentuan diterima atau tidaknya menggunakan klasifikasi Decision Tree. Kombinasi kedua algoritma tersebut untuk mendapatkan dataset yang valid dan reliabel untuk menentukan Penerima Beasiswa UKT. Oleh karena itu penelitian ini akan membutuhkan alat dan bahan sebagai berikut :

Pengumpulan data didapat dari data pengajuan beasiswa Uang Kuliah Tunggal di Biro Administrasi Akademik Universitas Muhammadiyah Kudus sebanyak 637 Mahasiswa yang mengajukan sedangkan kuota hanya 157. Dan dengan kriteria mahasiswa yang kurang mampu yang terdampak dari pandemi covid-19. Setelah mendapatkan data mentah selanjutnya di preprocessing karena data yang di peroleh masih tercampur antara ordinal dan tekstual. Dalam proses preprocessing data set yang baru dengan mengklasterkan menggunakan algoritma k-means dan setelah mendapatkan

klister baru baru munggunakn Decession tree untuk penentuan yang mendapatkan beasiswa.

III. HASIL DAN PEMBAHASAN

Pengumpulan data pada penelitian ini di dadapatkan dari Bagian Administrasi Kemahasiswaan Universitas Muhammadiyah Kudus. Data yang didapat dari Bagian Administarsi Kemahasiswaan Universitas Muhammadiyah Kudus Pada Tahun 2020. Dengan jumlah data 637 record data. Sedangkan data yang berhak dan lolos mendapatkan beasiswa sebesar 157 Mahasiswa.

Sebelumnya didapat data set dari mahasiswa yang mengisi pengajuan beasiswa UKT dengan Mengisi link Google Foarm. dengan 18 fitur yang terdiri dari NIM, NIK, Kode Perguruan Tinggi, Semester, Nomor NIK, Nama Mahasiswa, IPK, Angkatan, Alamat, No HP, Pekerjaan Ayah, Pekerjaan Ibu, Penghasilan Kotor Ayah, Penghasilan Kotor Ibu, Jumlah Anggota Keluarga, Surat Pernyataan dan Surat Keterangan Tidak Mampu.

A. Pengolahan Data Set Awal

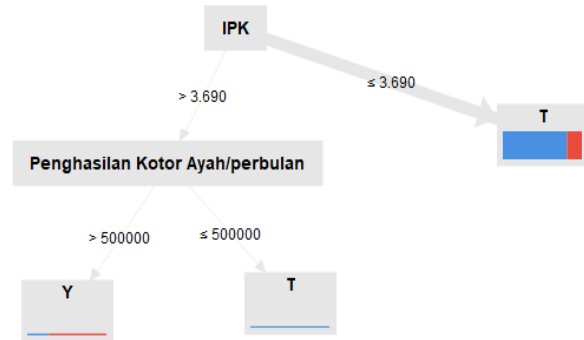
Setelah mendapatkan data set apa adanya peneliti mengolah data tersebut menggunakan rapidminer. Mencoba untuk mengetahui apakah seleksi yang dilakukan secara manual yang dilakukan oleh team seleksi apakah sudah belum tepat apa belum. Dari pengolahan data set tersebut di dapatkan bahawa Accuracy sebesar 80%, recall sebesar 49 % dan Precission sebesar 40.26 %. Dimana prediksi yang dilakukan secara manual oleh team masih bias Ketika dilakukan pengolahan menggunakan Decision Tree di Rapidminer

Dari pengolahan data set awal yang dilakukan klasifikasi menggunakan decision Tree diatas menunjukkan bahwa berdasarkan $IPK \leq 3.69$ tidak mendapatkan beasiswa, tetapi itu masih terdapat eror,

sedangkan $IPK > 3.69$ dan penghasilan Ayah perbulan yang lebih $>5.000.000$ berhak mendapatkan beasiswa jelas ini menyalahi aturan dan data setnya kurang tepat dan masi terdapat bias dalam dataset tersebut.

	true T	true Y	class precision
pred. T	124	30	80.52%
pred. Y	1	0	0.00%
class recall	99.20%	0.00%	

Diagram 1. Performance DT



B. Pengolahan Data Set Lanjutan

Setelah mendapatkan hasil dan menguji coba dataset awal yang di dapatkan dari bagian adminstrasi kemahasiswaan ternyata terdapat data yang kurang tepat dan valid dan masih bias, karena masih dilakukan secara manual oleh team,. Kemudian dilakukan perbandingan pengolahan data set awal tersebut dengan menggunakan machine learning supaya bisa memberikan masukan dan bahan pertimbangan kepada team seleksi untuk tahun berikutnya agar tepat sasaran kepada penerima beasiswa.

Melihat data set yang ada maka tahap selanjutnya dilakukan preprocessing data set tersebut agar data set tersebut valid dan dapat digunakan untuk memberikan pertimbangan keputusan salam penentuan penerima beasiswa UKT. Preprocessing data set yang sudah ada akan kita preprocessing terlebih dahulu sebagaimana dibawah ini :

Dari dataset tersebut di dapatkan bahwa untuk program studi yang paling banyak mengajukan beasiswa adalah Program Studi S1 Keperawatan. Karena memang mahasiswa keperawatan jumlah masih banyak di bandingkan dengan Program studi lainnya

Selanjutnya adalah Preprocessing data pekerjaan ayah dimana pekerjaan ayah yang paling banyak adalah wiraswasta dan pekerjaan ibu dari table distribusi yang paling banyak pekerjaan adalah ibu rumah tangga.

dimana untuk mengetahui apakah ayah dan ibu nya wafat atau masih hidup ini digunakan untuk menentukan jumlah anggota keluarga yang di tanggung. Dimana jumlah anggota keluarga yang di tanggung ditambah ayah dan ibu yang masih hidup.

Gaji orangtua adalah penghasilan kotor ayah dan ibu di gabungkan menjadi satu menjadi penghasilan orang tua. Dimana menurut indeks kemiskinan setiap orang yang mengkonsumsi di bawah Rp. 32.757,4 perhari dikatakan miskin menurut bank Dunia (World Bank).

Biaya hidup perbulan adalah anggota keluarga di kalikan biaya konsumsi dikali 30 hari. Sisa penghasilan orangtua adalah gaji orangtua dikurangi biaya hidup. Sisa penghasilan orangtua yang minus kurang dari sama dengan 0 dikatakan miskin.

Selanjutnya adalah IPK kita kelompokkan menurut Peraturan Menteri Pendidikan & Kebudayaan Tinggi No 3 Tahun 2020 Standard Nasional Pendidikan Tinggi di Pasal 27. Predikat Memuaskan dengan rentang IPK 2.76 – 3.00, Predikat Sangat Memuaskan dengan Rentang IPK 3.01 – 3.50 dan Predikat Cumlaude dengan Rentang IPK 3.51 – 4.00. Dari pengelompokan tersebut di dapatkan label IPK dimana Mahasiswa yang mempunyai IPK dengan Predikat Sangat Memuaskan Paling Banyak sedangkan untuk IPK dengan Predikat Cumlaude Jumlah nya lebih sedikit dibanding dengan IPK Predikat Memuaskan.

C. Klustering.

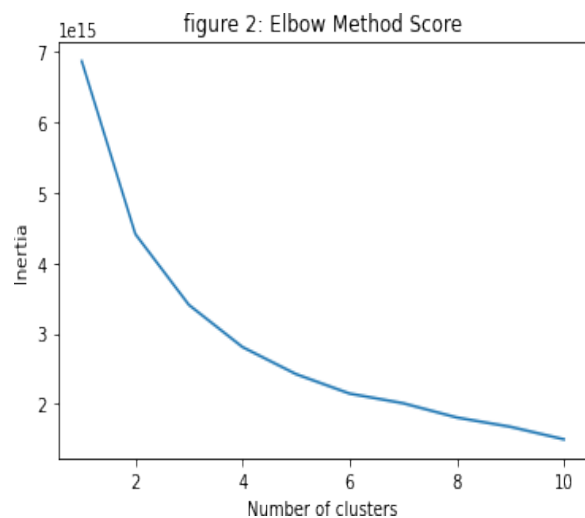
Setelah mendapatkan dataset yang sudah bersih atau dataset yang sudah di preprocessing sebelumnya maka Langkah selanjut ada melakukan clustering. Maka didapatkan data set sebagai berikut. Dari table korealsi tersebut didapatkan bahwa label ipk ada hubungan yang kuat dengan IPK. Selain itu gaji orang tua juga ada hubungan korelasi yang kuat dengan penghasilan kotor ayah dan penghasilan kotor ibu.

Melihat dari korelasi tersebut beberapa fitur akan di drop atau di hapus karena tidak ada hubungan korelasi yang kuat.

Diantaranya fitur yang di hapus adalah NIM, Label, Angkatan (Tahun Masuk) dan Alamat.

Setelah beberapa fitur di drop atau di hapus maka di dapatkan fitur yang terbaru adalah sebagai berikut : Program Studi, Semester, IPK, Pekerjaan Ayah, Pekerjaan Ibu, Penghasilan Kotor Ayah/perbulan, Penghasilan Kotor Ibu/perbulan, Jumlah Anggota Keluarga Yang ditanggung, Ayah Hidup, Ibu Hidup, anggota keluarga, gaji ortu, biaya hidup, sisa penghasilan, status miskin, label_IPK.

Setelah dataset final tersebut selanjutnya adalah di lakukan klusterisasi dengan menggunakan K-Means apa tidak. Ternyata bisa dilakukan klusterisasi dengan menggunakan K-Means dengan pengecekan menggunakan Hopkins Measure dengan rata-rata 0.97. selanjut nya menentukan jumlah kluster menggunakan metode elbow[11] dan silhouette Analisa untuk menemukan nomor kluster dalam data set, maka di dapatkan sebagaimana berikut sesuai dengn gambar di bawah ini



Setelah didapatkan 2 kluster Langkah selanjutnya dalam penelitian ini adalah menguji dataset awal yang didapatkan di bagian administrasi akademik. Dengan menggunakan 2 kluster tersebut digambarkan 0 dan 1. Dimana kluster 1 dilambangkan 0 dalam data set tersebut, yang mendapatkan Beasiswa atau yang berlabel Y dimasukan dalam kluster 0. Sedangkan kluster ke 2 dilambangkan dengan angka 1 dimana dataset awal yg di dapatkan yang berlabel T dimasukan kedalam cluster 2.

Klsuter 0 dan 1 jumlah datanya lebih banyak yang berlabel T dibandingkan yang belabel Y ini dikarenakan dataset nya dilakukan secara manual oleh team seleksi. Dari gambar 14 diatas bahwa untuk kluster 0 dengan status miskin 0 atau False. Dan untuk kluster 1 status miskin 1 atau true. Jadi dari tabel diatas di putuskan bahwa yang termasuk dalam kluster 1 adalah kelompok yang diusulkan dapat beasiswa karena miskin

Jika dilihat dari penghasilan kotor ayah dan ibu kluster 1 yang berwarna abu- abu penghasilannya rendah dan masuk dalam kategori miskin, tampak pada gambar 15, yang menggambarkan penghasilan orangtua

penghasilan kotor ayah perbulan sebelum digabungkan dengan penghasilan kotor ibu, Nampak jelas bahwa kluster 1 yang berwarna abu lebih rendah dibandingkan yang berwarna merah kluster 0.

penghasilan kotor ayah perbulan sebelum digabungkan dengan penghasilan kotor ibu, Nampak jelas bahwa kluster 1 yang berwarna abu lebih rendah dibandingkan yang berwarna merah kluster 0

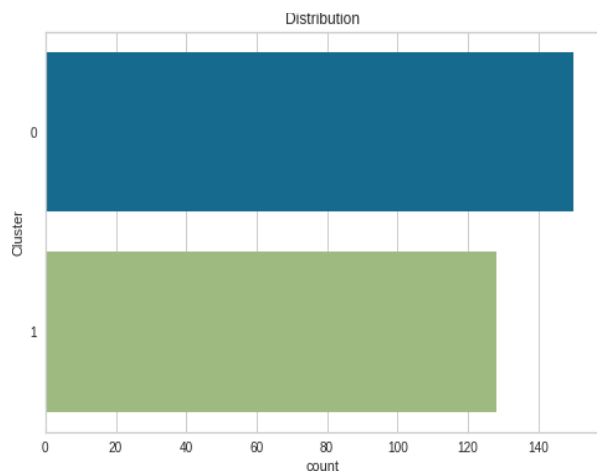
Dan jika dilihat berdasarkan IPK yang sudah dikonversi kedalam 3 kategori, Memuaskan, Sangat Memuaskan dan Cumlaude dari cluster 1 dan hampir imbang dan lebih tinggi sedikit dibandingkan dengan cluster 0. Bahwa ipk mahasiswa yang termasuk dalma kluster 1 atau yang termasuk miskin ipk nya lebih banyak yang cumlaud jika dilihat secara tabel diatas

Setelah dilakukan pembersihan data set awal menggunakan K-Means Maka di dapatkan dataset bersih beasiswa dimana yang termasuk dalam Kluster 1 dan berlabel Y dan dataset bersih tidak beasiswa yang ktermasuk dalam kluster 0 dan belabel T.

Dari hasil pengujian dimana nilai dataset awal 80 % dan setelah dilakukan testing dan pengujian menjadi 100 %. Pada penelitian adalah menguji dataset apakah data yang digunakan sudah valid atau belum.

Setalah dilakukan klusterisasi menggunakan K-Means Langkah selanjutnya adalah klasifikasi menggunakan Decision Tree.

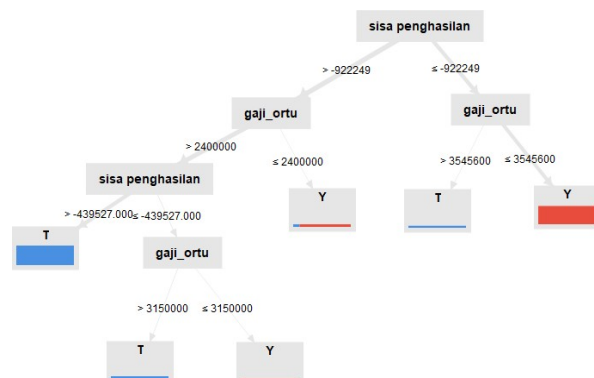
Klasifikasi dengan menggunakan Decision



DTree dimana fitur yang terbaik dan dihasilkan adalah fitur sisa penghasilan dan gaji orangtua. Dari klasifikasi tersebut bahwa yang mendapatkan beasiswa adalah apabila sisa penghasilan orangtua ≤ -922249 dan gaji orang tua ≤ 3545600 dan yang tidak mendapatkan beasiswa adalah sisa penghasilan orangtua ≤ -922249 dan gaji orang tua ≥ 3545600

accuracy: 100.00%

	true T	true Y	class precision
pred. T	30	0	100.00%
pred. Y	0	26	100.00%
class recall	100.00%	100.00%	



IV. KESIMPULAN

Dataset yang kurang lengkap dan kurang reliabel sering kali membuat bias dalam pengambilan keputusan atau dalam memahami bisnis understanding nya. Pada penelitian ini peneliti mengkombinasikan Klusterisasi dan Klasifikasi dimana kedua nya adalah metode yang berbeda. Dimana

masih jarang penelitian yang mengkombinasikan dua hal tersebut. Dari hasil pengukuran pengujian pada penelitian ini di dapatkan hasil yang maksimal bisa mencapai 100%.

Maka dari itu, kesimpulan yang didapatkan adalah metode gabungan K-Means sebagai deteksi outlier dan Decision Tree sebagai Klasifikasi dapat digunakan untuk mendapatkan hasil optimal penentuan penerima bantuan Beasiswa Bantuan Uang Kuliah Tunggal dengan tepat sasaran..

UCAPAN TERIMA KASIH

Saya Mengucapkan terimakasih sebanyak – banyak nya kepada Bapak Rektor Universitas Muhammadiyah Kudus Periode 2018 – 2023 beserta Rekan – Rekan Kerja dan tak lupa kepada keluarga yag sudah penuh dengan setia mendukung saya.

DAFTAR PUSTAKA

mohamad jajuli nurul rohmawati, sofi defiyanti, “Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa,” *Jitter 2015*, vol. I, no. 2, pp. 62–68, 2015.

- A. Salam, D. Adiatma, and J. Zeniarja, “Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- R. Sovia, E. P. W. Mandala, and S. Mardhiah, “Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 181, 2020, doi: 10.26418/jp.v6i2.37759.
- M. Mardison, S. Defit, and S. Alturky, “Prediction of Scholarship Recipients Using Hybrid Data Mining Method with Combination of K-Means and C4.5 Algorithms,” *Int. J. Artif. Intell. Res.*, vol. 5, no. 2, pp. 168–179, 2021, doi: 10.29099/ijair.v5i2.224.

- D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, “The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, 2018, doi: 10.1088/1757-899X/434/1/012039.
- P. M. Kellstedt and G. D. Whitten, *Data Mining: Concepts and Techniques: Concepts and Techniques*. 2018.
- X. Gao and S. Wu, “Hierarchical Clustering Algorithm for Binary Data Based on Cosine Similarity,” *8th Int. Conf. Logist. Informatics Serv. Sci. LISS 2018 - Proceeding*, no. August 2018, 2018, doi: 10.1109/LISS.2018.8593222.
- X. D. Wang, R. C. Chen, F. Yan, Z. Q. Zeng, and C. Q. Hong, “Fast Adaptive K-Means Subspace Clustering for High-Dimensional Data,” *IEEE Access*, vol. 7, pp. 42639–42651, 2019, doi: 10.1109/ACCESS.2019.2907043.
- P. Wang and X. Chen, “Three-Way Ensemble Clustering for Incomplete Data,” *IEEE Access*, vol. 8, pp. 91855–91864, 2020, doi: 10.1109/ACCESS.2020.2994380.
- P. C. Ncr *et al.*, “Crisp-Dm,” *SPSS inc*, vol. 78, pp. 1–78, 2000, [Online]. Available: <http://www.crisp-dm.org/CRISPWP-0800.pdf>.
- P. Bholowalia and A. Kumar, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 975–8887, 2014.
- Y. Pratama, M. Pasaribu, J. Nababan, D. Sihombing, and D. Gultom, “Selection of Scholarship Recipient by Implementing Genetic Algorithm and Fuzzy Logic,” *J. Phys. Conf. Ser.*, vol. 1933, no. 1, 2021, doi: 10.1088/1742-6596/1933/1/012069.
- Okfalisa, R. Fitriani, and Y. Vitriani, “The comparison of linear regression method and k-nearest neighbors in scholarship recipient,” *Proc. - 2018 IEEE/ACIS 19th Int. Conf. Softw. Eng. Artif. Intell. Netw.*

Parallel/Distributed Comput. SNPD
2018, pp. 194–199, 2018, doi:
10.1109/SNPD.2018.8441068.

Han, J., Kamber, M., & Pei, J. (2011). Data Mining Concepts and Techniques Third Edition. Waltham: Elsevier Inc

Gorunescu, F. (2011). Data Mining Concepts, Models and Techniques. Berlin: Springer