p-ISSN : 2301-8364, e-ISSN : 2685-6131 , available at : https://ejournal.unsrat.ac.id/index.php/informatika

Application of Single Tuition Fee Classification Based on Clustering Using PCA and K-Means

Aplikasi Klasifikasi Uang Kuliah Tunggal Berbasis Clustering Menggunakan PCA dan K-Means

Kaseger Arthur Farly
Master Program of Informatics, Postgraduate Program, Sam Ratulangi University, Manado, Indonesia
e-mails: kasegerfarly026@student.unsrat.ac.id
Received: 5 January 2026; revised: 19 January 2026; accepted: 28 January 2026

***Abstract* — The determination of the Single Tuition Fee (Uang Kuliah Tunggal/UKT) is an important policy to ensure fairness in higher education financing based on students' socio economic conditions. However, conventional UKT determination is often subjective and less effective in processing multidimensional data. This study applies an unsupervised learning approach using Principal Component Analysis (PCA) and the K-Means algorithm to classify students into UKT categories. Socio-economic data such as parents' income, number of dependents, electricity capacity, housing conditions, and supporting attributes were used in the analysis. PCA was employed to reduce data dimensionality, while K-Means was used to cluster students based on similar characteristics. The evaluation results show that two clusters provide the optimal grouping with a silhouette score of 0.7631, which were mapped into low and high UKT categories. The resulting model was then integrated into a web-based application to enable automatic UKT classification. The developed system demonstrates the potential to improve objectivity, consistency, and efficiency in the UKT determination process.**

***Key words* — Single Tuition Fee; PCA; K-Means; Clustering; Classification Application**

***Abstrak* — Penentuan Uang Kuliah Tunggal (UKT) merupakan kebijakan penting untuk menjamin keadilan pembiayaan pendidikan tinggi berdasarkan kondisi social ekonomi mahasiswa. Namun, proses penentuan UKT secara konvensional masih berpotensi menimbulkan subjektivitas dan kurang optimal dalam mengolah data multidimensi. Penelitian ini menerapkan pendekatan *unsupervised learning* menggunakan Principal Component Analysis (PCA) dan algoritma K-Means untuk mengklasifikasikan mahasiswa ke dalam kategori UKT. Data social ekonomi yang digunakan meliputi penghasilan orang tua, jumlah tanggungan, daya listrik, kondisi tempat tinggal, serta atribut pendukung lainnya. PCA digunakan untuk mereduksi dimensi data, sedangkan K-Means digunakan untuk melakukan pengelompokan mahasiswa berdasarkan kemiripan karakteristik. Hasil evaluasi menunjukkan bahwa dua cluster merupakan konfigurasi optimal dengan nilai silhouette score sebesar 0,7631, yang kemudian dipetakan ke dalam kategori UKT rendah dan UKT tinggi. Model yang dihasilkan selanjutnya diintegrasikan ke dalam aplikasi berbasis web untuk mendukung proses klasifikasi UKT secara otomatis.**

***Kata kunci* — Uang Kuliah Tunggal; PCA; K-Means; Clustering; Aplikasi Klasifikasi**

I. PENDAHULUAN

Uang Kuliah Tunggal (UKT) merupakan salah satu kebijakan strategis dalam sistem pendidikan tinggi di Indonesia yang bertujuan menjamin akses pendidikan yang adil dan berkelanjutan. Penetapan UKT didasarkan pada kondisi sosial-ekonomi mahasiswa agar beban biaya pendidikan dapat disesuaikan dengan kemampuan finansial keluarga secara proporsional [1] [2]. Oleh karena itu, akurasi dan objektivitas dalam proses pengelompokan mahasiswa ke dalam kategori UKT menjadi aspek yang sangat krusial dalam mendukung keadilan sosial di lingkungan perguruan tinggi [3].

Dalam praktiknya, proses penentuan UKT masih sering dilakukan melalui penilaian administratif dan evaluasi manual terhadap data sosial-ekonomi mahasiswa. Pendekatan tersebut memiliki sejumlah keterbatasan, antara lain tingginya potensi subjektivitas, inkonsistensi antar penilai, serta keterbatasan dalam mengolah data yang bersifat multidimensi [4]. Data sosial ekonomi mahasiswa umumnya terdiri dari berbagai variabel yang saling berkorelasi, seperti penghasilan orang tua, jumlah tanggungan keluarga, kepemilikan aset, daya listrik, dan karakteristik tempat tinggal [5] [6]. Kompleksitas data ini menyulitkan proses penilaian apabila hanya mengandalkan aturan sederhana atau penilaian konvensional, sehingga berpotensi menghasilkan klasifikasi UKT yang kurang tepat sasaran[7].

Seiring dengan berkembangnya teknologi pengolahan data, pendekatan berbasis machine learning menjadi alternatif yang menjanjikan dalam mendukung proses klasifikasi UKT secara objektif [8]. Metode unsupervised learning, seperti Principal Component Analysis (PCA) dan K-Means clustering, mampu mengolah data sosial-ekonomi berdimensi tinggi dan mengungkap pola laten yang tidak mudah diamati secara langsung. PCA berperan dalam mereduksi dimensi data dan mengatasi multikolinearitas antar variabel, sedangkan K-Means digunakan untuk mengelompokkan mahasiswa berdasarkan kemiripan karakteristik sosial-ekonomi [9]. Namun demikian, hasil analisis berbasis machine learning akan memiliki dampak yang terbatas apabila tidak diintegrasikan ke dalam sistem atau aplikasi yang dapat digunakan secara operasional.

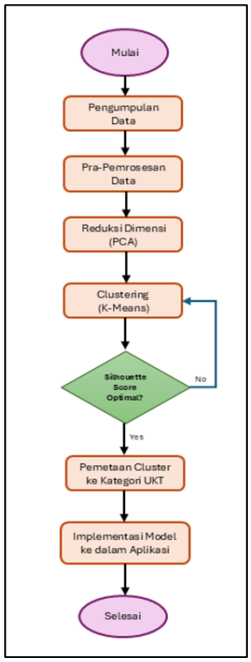
Permasalahan utama yang sering muncul adalah kesenjangan antara hasil analisis data dan implementasinya dalam proses administrasi UKT. Model analisis yang baik tidak secara otomatis menjamin kemudahan penggunaan oleh pengelola perguruan tinggi apabila hasilnya tidak disajikan dalam bentuk yang praktis, mudah dipahami, dan siap digunakan. Oleh karena itu, diperlukan pengembangan aplikasi klasifikasi UKT yang

mampu menerjemahkan hasil pemodelan data menjadi keluaran yang bersifat aplikatif, seperti kategori UKT per mahasiswa, visualisasi distribusi UKT, serta data hasil klasifikasi yang terstruktur [10].

Penelitian ini berfokus pada pengembangan aplikasi prediksi atau klasifikasi UKT mahasiswa dengan memanfaatkan metode PCA dan algoritma K-Means. Aplikasi yang dikembangkan dirancang untuk mengolah data sosial ekonomi mahasiswa secara otomatis dan menghasilkan kategori UKT yang dapat langsung digunakan dalam proses pengambilan keputusan. Dengan mengintegrasikan teknik unsupervised learning ke dalam sebuah sistem aplikasi, penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan objektivitas, konsistensi, dan efisiensi penetapan UKT, serta mendukung penerapan kebijakan pendidikan tinggi yang lebih adil dan berbasis data.

II. METODE

Pada penelitian ini, data yang digunakan merupakan data sosial ekonomi mahasiswa yang digunakan sebagai dasar penentuan UKT. Data yang telah dikumpulkan kemudian diproses melalui tahap data preprocessing untuk memastikan data berada dalam kondisi siap olah, termasuk penyesuaian format dan skala data. Selanjutnya, proses modelling dilakukan dengan membangun model klasifikasi UKT berbasis teknik analisis data, yang kemudian dievaluasi pada tahap model evaluation untuk memastikan kinerja model yang optimal. Apabila hasil evaluasi belum memenuhi kriteria, proses pemodelan dilakukan kembali hingga diperoleh model yang sesuai. Model terbaik selanjutnya diterapkan pada tahap deployment dalam bentuk aplikasi klasifikasi UKT. Rancangan alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Penelitian ini menggunakan data sosial ekonomi mahasiswa yang diperoleh dari Unit Pelayanan Akademik TIK Universitas Sam Ratulangi sebagai data internal universitas yang dikumpulkan dalam proses administrasi akademik. Dataset memuat informasi mahasiswa yang mencakup atribut ekonomi, demografi, dan fasilitas pendukung, seperti penghasilan orang tua, jumlah tanggungan keluarga, daya listrik rumah tangga, kondisi dan status tempat tinggal, serta informasi akademik dasar. Atribut-atribut tersebut dipilih karena memiliki keterkaitan langsung dengan kemampuan ekonomi mahasiswa dan digunakan sebagai indikator dalam penentuan kategori UKT. Data disusun dalam bentuk dataset terstruktur dan digunakan sebagai dasar dalam proses analisis, sehingga kelengkapan dan relevansi data sangat berpengaruh terhadap kualitas hasil klasifikasi UKT yang dihasilkan.

B. Pra-Pemrosesan Data

Data yang telah dikumpulkan selanjutnya melalui tahap pra-pemrosesan untuk memastikan data berada dalam kondisi yang layak dianalisis. Tahap ini meliputi penanganan nilai hilang, pengubahan data kategorikal menjadi bentuk numerik, serta proses normalisasi atau standardisasi data agar seluruh variabel memiliki skala yang sebanding. Pra-pemrosesan dilakukan untuk mengurangi potensi bias akibat perbedaan skala antar variabel dan meningkatkan stabilitas kinerja algoritma yang digunakan [11] [12]n konsisten, proses pemodelan dapat berjalan secara lebih optimal.

C. Reduksi Dimensi Menggunakan PCA

Setelah data siap, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA). PCA bertujuan untuk menyederhanakan data berdimensi tinggi dengan cara mempertahankan variansi utama data ke dalam sejumlah komponen utama [13]. Proses ini membantu mengurangi kompleksitas data, mengatasi multikolinearitas antar variabel, serta meningkatkan efisiensi komputasi pada tahap clustering [14]. Hasil dari PCA berupa representasi data dalam ruang dimensi yang lebih rendah, namun tetap merepresentasikan karakteristik utama kondisi sosial-ekonomi mahasiswa.

D. Clustering Menggunakan Algoritma K-Means

Data hasil reduksi dimensi kemudian digunakan sebagai masukan pada proses clustering menggunakan algoritma K-Means. Algoritma ini bertujuan untuk mengelompokkan mahasiswa ke dalam beberapa cluster berdasarkan kemiripan karakteristik sosial-ekonomi. K-Means dipilih karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam menangani dataset berukuran besar. Proses clustering dilakukan secara iteratif hingga pusat cluster mencapai kondisi yang stabil [15] [16].

E. Evaluasi Jumlah Cluster Menggunakan Silhouette Score

Untuk memastikan kualitas pengelompokan, dilakukan evaluasi jumlah cluster menggunakan silhouette score [17]. Metode ini digunakan untuk mengukur tingkat kedekatan data dalam satu cluster dan tingkat pemisahan antar cluster. Nilai

silhouette score yang tinggi menunjukkan bahwa hasil clustering memiliki struktur yang baik. Jika nilai yang diperoleh belum optimal, maka proses clustering diulang dengan jumlah cluster yang berbeda hingga diperoleh hasil terbaik. Tahap ini memastikan bahwa jumlah cluster yang digunakan benar-benar merepresentasikan struktur data secara optimal [18].

F. Pemetaan Cluster ke Kategori UKT

Cluster yang telah terbentuk selanjutnya dianalisis karakteristiknya untuk dilakukan pemetaan ke dalam kategori UKT. Pemetaan ini didasarkan pada profil social ekonomi masing-masing cluster, seperti tingkat penghasilan orang tua dan kondisi tempat tinggal. Cluster dengan karakteristik ekonomi yang lebih rendah dipetakan sebagai kategori UKT rendah, sedangkan cluster dengan kondisi ekonomi yang lebih tinggi dipetakan sebagai kategori UKT yang lebih tinggi. Tahap ini berperan penting dalam memberikan interpretasi yang bermakna terhadap hasil clustering [19]

G. Implementasi Model ke dalam Aplikasi

Tahap akhir penelitian adalah implementasi model klasifikasi UKT ke dalam aplikasi. Model yang telah terbentuk, termasuk tahapan pra-pemrosesan, PCA, dan clustering, diintegrasikan ke dalam sistem aplikasi sehingga dapat digunakan untuk mengolah data mahasiswa secara otomatis. Aplikasi ini menyajikan hasil klasifikasi UKT dalam bentuk tabel dan visualisasi, serta menyediakan keluaran yang dapat digunakan sebagai alat bantu pengambilan keputusan. Dengan implementasi ini, hasil penelitian tidak hanya bersifat konseptual, tetapi juga aplikatif dan siap digunakan dalam konteks nyata.

III. HASIL DAN PEMBAHASAN

A. Hasil Pra-pemrosesan Data

Tabel I menyajikan contoh data mentah sebelum dilakukan tahap pra-pemrosesan. Data yang ditampilkan merupakan sebagian kecil dari dataset keseluruhan untuk menggambarkan kondisi awal data sebagaimana diperoleh dari sumber administrasi. Atribut yang ditampilkan meliputi penghasilan ayah, penghasilan ibu, daya listrik rumah tangga, jumlah tanggungan keluarga, serta luas tanah tempat tinggal. Dari tabel tersebut terlihat bahwa setiap atribut memiliki skala dan satuan yang berbeda, misalnya penghasilan dinyatakan dalam satuan tertentu, daya listrik dalam satuan watt, serta luas tanah dalam rentang kategori.

TABEL I
CONTOH DATA MENTAH SEBELUM PRA-PEMROSESAN DATA

Penghasilan Ayah	Penghasilan Ibu	Daya Listrik	Jumlah Tanggungan	Luas Tanah
10.0	1.0	1300	4	100 – 200 m²
10.0	9.0	2200	4	50 – 99 m²
1.0	1.0	900	5	50 – 99 m²
1.0	1.0	900	6	25 – 50 m²
9.0	10.0	2200	5	50 – 99 m²

TABEL II
CONTOH DATA SETELAH PRA-PEMROSESAN DATA

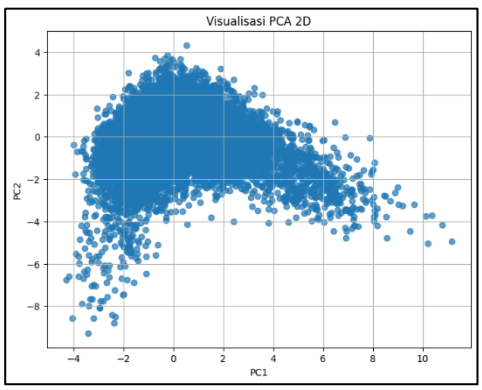
Penghasilan Ayah	Penghasilan Ibu	Daya Listrik	Jumlah Tanggungan	Luas Tanah
2.58	-0.53	0.20	-0.10	-1.17
2.58	3.06	0.98	-0.10	0.10
-0.74	-0.53	-0.14	0.42	0.74
-0.74	-0.53	-0.14	0.95	-0.53
2.21	3.51	0.98	0.42	0.10

Kemudian pada Tabel II menyajikan contoh data sosial-ekonomi mahasiswa setelah melalui tahap pra-pemrosesan, khususnya proses standardisasi data. Nilai-nilai pada tabel ini merupakan hasil transformasi dari data mentah sehingga setiap atribut berada pada skala yang seragam dengan rata-rata mendekati nol. Nilai positif menunjukkan bahwa suatu atribut berada di atas nilai rata-rata populasi, sedangkan nilai negatif menunjukkan kondisi di bawah rata-rata.

Sebagai contoh, nilai penghasilan ayah dan ibu yang bernilai positif mengindikasikan tingkat pendapatan yang relatif lebih tinggi dibandingkan rata-rata, sementara nilai negatif pada atribut luas tanah dan jumlah tanggungan menunjukkan kondisi yang relatif lebih rendah. Proses standardisasi ini dilakukan untuk menghilangkan pengaruh perbedaan skala antar atribut dan memastikan bahwa tidak ada variabel tertentu yang mendominasi proses analisis. Dengan data yang telah distandarisasi, tahapan selanjutnya seperti reduksi dimensi menggunakan PCA dan clustering menggunakan K-Means dapat dilakukan secara lebih optimal dan stabil.

B. Hasil Reduksi Dimensi Menggunakan PCA

Visualisasi pada Gambar 2 menunjukkan hasil reduksi dimensi menggunakan Principal Component Analysis (PCA) ke dalam dua komponen utama, yaitu PC1 dan PC2. Proyeksi ini dilakukan untuk menyederhanakan struktur data sosial-ekonomi mahasiswa yang semula berdimensi tinggi menjadi bentuk yang lebih mudah dianalisis secara visual. Sebaran titik-titik data pada plot memperlihatkan pola distribusi mahasiswa berdasarkan karakteristik sosial-ekonominya, yang kemudian digunakan sebagai dasar dalam proses clustering. Dengan memanfaatkan PCA, informasi penting dari variabel-variabel awal tetap dipertahankan secara maksimal, sehingga visualisasi ini tidak hanya membantu dalam eksplorasi data, tetapi juga memperkuat interpretasi hasil klasifikasi UKT secara lebih efektif.



Gambar 2. Visualisasi PCA 2D

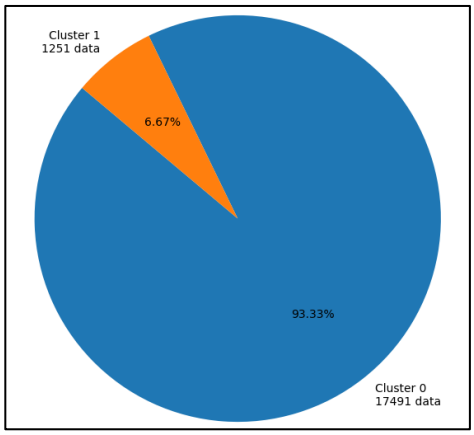
C. *Hasil Klasifikasi Mahasiswa Menggunakan K-Means*

Berdasarkan hasil evaluasi menggunakan Silhouette Score, nilai optimal diperoleh saat jumlah klaster (k) adalah 2, dengan skor tertinggi sebesar 0.7631. Nilai ini menunjukkan bahwa pemisahan antar cluster paling jelas dan kompak dibandingkan konfigurasi lainnya. Ketika jumlah klaster ditambah, skor cenderung menurun, menandakan penurunan kualitas segmentasi. Yakni dimana pada k=3 dan k=4, skor masing-masing adalah 0.6710 dan 0.6895, sementara pada k=5 hingga k=9, skor berada di bawah 0.57. Hal ini mengindikasikan bahwa pembentukan dua kelompok merupakan konfigurasi paling stabil dan representatif untuk data yang digunakan.

D. *Distribusi Hasil Clustering dan Pemetaan ke Kategori UKT*

Visualisasi pada Gambar 3 menunjukkan distribusi data mahasiswa berdasarkan hasil clustering menggunakan algoritma K-Means. Dari total data yang dianalisis, sebanyak 93.33% mahasiswa tergolong dalam Cluster 0, sementara sisanya 6.67% masuk dalam Cluster 1. Proporsi yang timpang ini mengindikasikan bahwa mayoritas mahasiswa memiliki karakteristik sosial-ekonomi yang serupa dan cenderung membentuk satu kelompok dominan. Ketimpangan distribusi ini juga menjadi indikasi awal bahwa terdapat satu kelompok besar dengan kondisi ekonomi relatif homogen, serta satu kelompok kecil yang berbeda secara signifikan.

Tabel II memperjelas hasil clustering dengan mengaitkan masing-masing cluster terhadap kategori UKT. Cluster 0 dikaitkan dengan kategori UKT rendah, mencakup 17.491 mahasiswa, sedangkan Cluster 1 mewakili kategori UKT tinggi dengan jumlah 1.251 mahasiswa. Hasil ini menunjukkan bahwa proses klasifikasi berhasil memisahkan mahasiswa berdasarkan kemampuan ekonomi secara konsisten dengan kebijakan UKT yang berlaku. Kesesuaian antara hasil clustering dan kategori UKT memperkuat validitas pendekatan yang digunakan, serta menunjukkan bahwa atribut sosial ekonomi yang dipilih mampu merepresentasikan kondisi riil mahasiswa secara efektif.



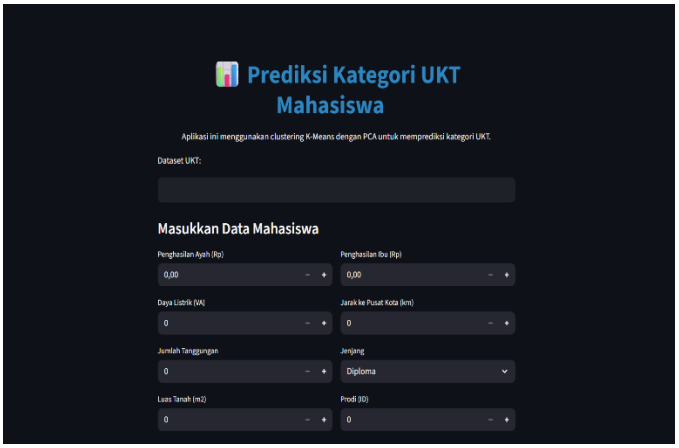
Gambar 3. Visualisasi Distribusi Data Berdasarkan Cluster

TABEL II
CONTOH DATA SETELAH PRA-PEMROSESAN DATA

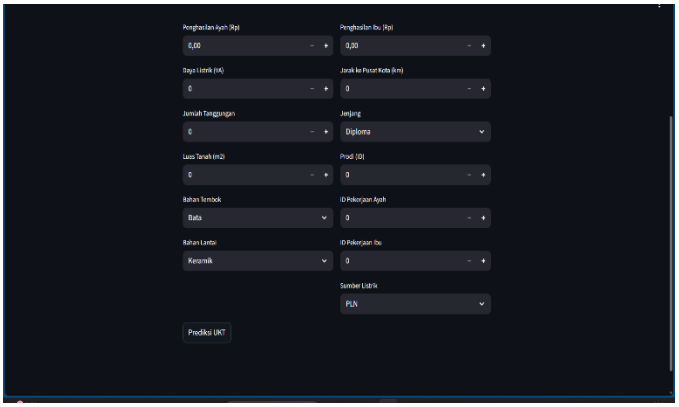
Cluster	Kategori UKT	Jumlah Mahasiswa
0	UKT Rendah	17.491
1	UKT Tinggi	1.251

E. *Hasil Implementasi Model ke dalam Aplikasi*

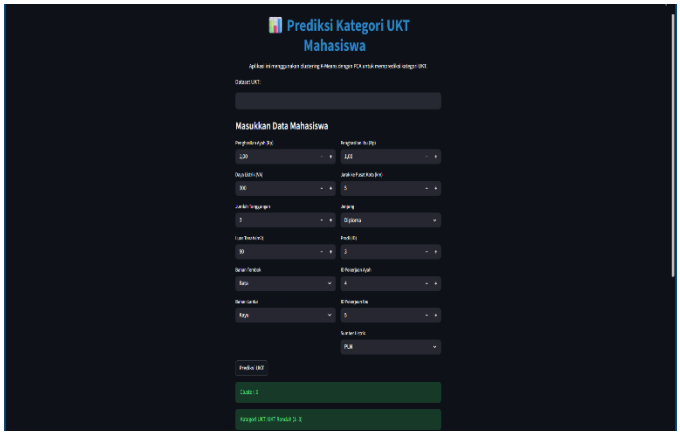
Gambar 4 menampilkan tampilan awal halaman aplikasi prediksi kategori UKT mahasiswa. Aplikasi ini dirancang sebagai antarmuka web yang sederhana dan informatif, dengan judul utama yang menjelaskan bahwa sistem menggunakan metode *clustering* K-Means yang telah dipadukan dengan PCA. Kemudian, pada Gambar 5 menunjukkan bagian formulir input data yang menjadi inti dari proses prediksi. Di sini, pengguna diminta untuk mengisi sejumlah variabel penting seperti penghasilan orang tua, daya listrik rumah, jumlah tanggungan keluarga, jarak ke pusat kota, jenis bahan bangunan, serta informasi akademik seperti jenjang pendidikan dan program studi. Setelah semua data diisi, pengguna dapat menekan tombol "Prediksi UKT" untuk mengetahui kategori UKT yang sesuai. Formulir ini merepresentasikan implementasi langsung dari model klasifikasi yang telah dibangun sebelumnya.



Gambar 4. Tampilan Awal Aplikasi Prediksi Kategori UKT Mahasiswa



Gambar 5. Tampilan Aplikasi Untuk Input Data



Gambar 6. Tampilan Aplikasi Berhasil Prediksi Kategori UKT

Selanjutnya, pada Gambar 6 menampilkan hasil akhir dari proses prediksi kategori UKT mahasiswa berdasarkan data yang telah diinput sebelumnya. Setelah pengguna mengisi seluruh atribut sosial-ekonomi dan akademik, sistem secara otomatis memproses data melalui tahapan reduksi dimensi menggunakan PCA dan pengelompokan dengan algoritma K-Means. Hasil prediksi ditampilkan secara langsung di halaman aplikasi, menunjukkan bahwa mahasiswa tersebut termasuk dalam Cluster 0, yang dikategorikan sebagai UKT Rendah. Tampilan ini memberikan umpan balik yang jelas dan cepat kepada pengguna, sehingga mereka dapat mengetahui estimasi kategori UKT berdasarkan kondisi yang mereka masukkan. Selain itu, hasil ini juga memperlihatkan bahwa sistem berhasil mengintegrasikan model klasifikasi ke dalam antarmuka web secara fungsional.

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan Principal Component Analysis (PCA) dan algoritma K-Means clustering mampu mengelompokkan mahasiswa berdasarkan kondisi sosial-ekonomi secara objektif dan sistematis sebagai dasar penentuan kategoriUang UKT. Tahap pra-pemrosesan dan standardisasi data berhasil mengatasi perbedaan skala antar variabel, sementara reduksi dimensi menggunakan PCA mampu menyederhanakan data tanpa menghilangkan informasi utama. Evaluasi menggunakan silhouette score menunjukkan bahwa dua cluster merupakan konfigurasi optimal, yang selanjutnya dapat dipetakan secara konsisten ke dalam kategori UKT rendah dan UKT tinggi. Implementasi model ke dalam aplikasi berbasis web membuktikan bahwa hasil pemodelan tidak hanya bersifat konseptual, tetapi juga aplikatif dan siap digunakan sebagai alat bantu pengambilan keputusan dalam penetapan UKT yang lebih adil, konsisten, dan berbasis data.

Penelitian selanjutnya dapat mengembangkan model dengan menambahkan atribut sosial-ekonomi yang lebih beragam agar klasifikasi UKT dapat merepresentasikan kondisi mahasiswa secara lebih komprehensif. Selain itu, eksplorasi metode clustering atau pendekatan pembelajaran lain sebagai pembanding dapat dilakukan untuk meningkatkan kualitas dan robustitas hasil klasifikasi. Dari sisi implementasi, aplikasi yang telah dikembangkan dapat ditingkatkan dengan fitur visualisasi analisis yang lebih mendalam serta integrasi langsung dengan sistem informasi akademik universitas, sehingga proses penentuan UKT dapat dilakukan secara otomatis, real-time, dan terintegrasi dengan sistem administrasi resmi perguruan tinggi.

V. KUTIPAN

[1] N. Rahmadani and S. Royal, “IMPLEMENTASI METODE K-MEANS CLUSTERING TUNGGAKAN UANG KULIAH PADA STMIK ROYAL KISARAN,” 2022. [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[2] N. Shabir Ramadhan, W. Nadia Nursafitri, T. Informasi dan Komputer, S. Informasi, and P. Negeri Subang, “INVENTOR: Jurnal Inovasi dan Tren Pendidikan Teknologi Informasi Analisis Klasifikasi UKT Mahasiswa Berdasarkan Tingkat Penghasilan Orang Tua Menggunakan Algoritma C4.5,” vol. 2, no. 1, pp. 3026–3190, 2024, doi: 10.37630/inventor.v1i3.1248.

[3] K. Fahriya and W. Yustanti, “Optimalisasi Jumlah Klaster Uang Kuliah Tunggal pada Data Sosial Ekonomi Mahasiswa,” *JEISBI*, vol. 02, p. 2021.

[4] A. Priyo and J. Heikal, “PT. Media Akademik Publisher IMPLEMENTATION OF K-MEANS CLUSTERING ALGORITHM FOR SEGMENTATION OF PROSPECTIVE STUDENTS FOR CORPORATE ACADEMY SCHOLARSHIP,” vol. 3, no. 6, p. 42, 2025, doi: 10.62281.

[5] Haris Kurniawan, Sarjon Defit, and Sumijan, “Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal,” *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80–89, Dec. 2020, doi: 10.52158/jacost.v1i2.102.

[6] M. R. Kusnaidi, T. Gulo, and S. Aripin, “Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal,” *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 330–338, Sep. 2022, doi: 10.47065/josyc.v3i4.2112.

[7] N. Nur Rahmawati and M. Ivan Ariful Fathoni, “PENENTUAN PENERIMA KIP KULIAH MAHASISWA S1 UNUGIRI MENGGUNAKAN FUZZY C-MEANS CLUSTERING,” *Jurnal Pendidikan Matematika dan Matematika*, vol. 6, no. 2, pp. 121–130, 2022, doi: 10.36526/tr.v%vi%i.2197.

[8] “38 Rosalia Hadi-Analysis of Application Design for Discounting Suggestions with the K-Nearest Neighbor Method in the.” [Online]. Available: https://ejournal.unsrat.ac.id/index.php/elekdankom

[9] S. Bandyopadhyay, S. S. Thakur, and J. K. Mandal, “Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society,” *Innov Syst Softw Eng*, vol. 17, no. 1, pp. 45–52, Mar. 2021, doi: 10.1007/s11334-020-00372-5.

[10] Wenny Tarisa Oktaviany, Fitri Insani, Alwis Nazir, and Pizaini, “Pengelompokan Wilayah Bencana Banjir di Indonesia Menggunakan Algoritma K-Means,” *Bulletin of Computer Science Research*, vol. 5, no. 4, pp. 542–553, Jun. 2025, doi: 10.47065/bulletincsr.v5i4.608.

[11] I. Syafrinal and E. L. Febrianti, "PENERAPAN ALGORITMA K-MEANS PADA APLIKASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN (STUDI KASUS: ZAHRA MART)," 2023.

[12] A. Amato and V. Di Lecce, "Data preprocessing impact on machine learning algorithm performance," *Open Computer Science*, vol. 13, no. 1, Jan. 2023, doi: 10.1515/comp-2022-0278.

[13] T. Muslim Yaqoob and E. Fadhil Abdullah, "Preprocessing andClustering Techniques forUncovering Demographic Insights: A Study ofBook Preferences Using PCA And K-Means," *Al-Furat Journal of Innovations in Electronics and Computer Engineering*, vol. 4, no. 1, pp. 1–14, Mar. 2025, doi: 10.46649/fjiece.v4.1.1a.25.3.2025.

[14] M. Afrizal Kesuma, C. Auzia Nugraha, O. Indi Cahyani, and M. Wati, "Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means Oleh : Cellia Auzia Nugraha Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means," *JUKI : Jurnal Komputer dan Informatika*, vol. 7, 2025.

[15] W. R. Nurputra, W. Witanti, and A. Komarudin, "Principal Component Analysis (PCA) Untuk Meningkatkan Hasil Klasterisasi Penjualan Video Game Menggunakan Algoritma K-Means".

[16] A. Moubayed, M. Injadat, A. Shami, and H. Lutfiyya, "Student Engagement Level in an e-Learning Environment: Clustering Using K-means," *American Journal of Distance Education*, vol. 34, no. 2, pp. 137–156, Apr. 2020, doi: 10.1080/08923647.2020.1696140.

[17] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS," *Jurnal SIMETRIS*, vol. 11, no. 2, 2020.

[18] M. Syahrul, A. Y. Kustiyahningsih, E. Rahmanita, B. Dwi, S. Muhammad, and I. Firmansyah, "Application of the DBSCAN Algorithm in MSME Clustering using the Silhouette Coefficient Method," 2024.

[19] A. Atira and B. Nurina Sari, "Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokkan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 17, pp. 76–86, 2023, doi: 10.5281/zenodo.8282638.

[20] D. Hediyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro."