

Single Tuition Fee Classification Using Light Gradient Boosting Machine with Confusion Matrix Analysis

Al Khaidar^{1*}, Nurdin², Fajriana³

^{1,2,3} Program Studi Magister Teknologi Informasi, Fakultas Teknik, Universitas Malikussaleh, Lhokseumawe, 24355, Indonesia

Informasi Artikel

Diterima : 13 Desember 2025
Revisi : 20 Desember 2025
Publikasi : 31 Desember 2025

Kata Kunci:

Uang Kuliah Tunggal
LightGBM
Klasifikasi
Machine Learning

ABSTRAK

Uang Kuliah Tunggal merupakan sistem pembiayaan pendidikan tinggi yang ditetapkan berdasarkan kemampuan ekonomi mahasiswa. Penetapan UKT yang masih dilakukan secara manual berpotensi menimbulkan subjektivitas dan ketidaktepatan klasifikasi. Penelitian ini bertujuan untuk mengembangkan model klasifikasi UKT berbasis data menggunakan metode Light Gradient Boosting Machine (LightGBM). Dataset yang digunakan terdiri dari 10.000 data mahasiswa Politeknik Negeri Lhokseumawe yang telah melalui tahap prapemrosesan dan transformasi fitur. Model dilatih menggunakan pembagian data latih dan data uji sebesar 80:20, serta dievaluasi menggunakan metrik akurasi, classification report, confusion matrix, dan 10-Fold Cross Validation. Hasil pengujian menunjukkan bahwa model LightGBM mencapai akurasi sebesar 98% pada data uji. Pengujian 10-Fold Cross Validation menghasilkan rata-rata akurasi sebesar 99,21% dengan standar deviasi 0,29%, yang menunjukkan stabilitas dan kemampuan generalisasi yang sangat baik. Hasil ini membuktikan bahwa LightGBM efektif dan andal untuk mendukung penetapan UKT yang lebih objektif dan berbasis data.

ABSTRACT

Single Tuition Fee is a higher education financing system determined based on students' economic capabilities. Manual UKT determination has the potential to lead to subjectivity and classification inaccuracy. This study aims to develop a data-based UKT classification model using the Light Gradient Boosting Machine (LightGBM) method. The dataset used consists of 10,000 students of the Lhokseumawe State Polytechnic that have gone through the preprocessing and feature transformation stages. The model was drilled using an 80:20 split of training data and test data, and evaluated using accuracy metrics, classification reports, confusion matrices, and 10-Fold Cross Validation. The test results show that the LightGBM model achieves 98% accuracy on the test data. 10-Fold Cross Validation testing produces an average accuracy of 99.21% with a standard deviation of 0.29%, indicating excellent stability and generalization capabilities. These results prove that LightGBM is effective and reliable in supporting more objective and data-driven UKT determination.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



*Penulis Koresponden

Email: alkhaidarkutablang@gmail.com

Cara sitasi IEEE::

A. Khaidar, N. Nurdin, F. Fajriana, "Single Tuition Fee Classification Using Light Gradient Boosting Machine with Confusion Matrix Analysis," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 4, p. 1444-1454, Desember 2025. doi: 10.30811/jaise.v5i4.847

1. PENDAHULUAN

Uang Kuliah Tunggal (UKT) merupakan sistem pembiayaan pendidikan tinggi yang diterapkan oleh perguruan tinggi negeri di Indonesia sebagai bentuk penyederhanaan biaya pendidikan bagi mahasiswa [1], [2],[3]. Melalui sistem ini, setiap mahasiswa dikenakan satu jenis biaya tetap setiap semester yang telah disesuaikan dengan kemampuan ekonomi keluarga, sehingga tidak perlu lagi membayar biaya per komponen seperti SPP, praktikum, maupun biaya penunjang lainnya. UKT diterapkan di perguruan tinggi negeri sebagai bagian dari kebijakan pemerintah Indonesia melalui Permendikbud No. 55 Tahun 2013 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal. Klasifikasi tidak hanya digunakan dalam bidang industri, bisnis, maupun kesehatan, tetapi juga memiliki peran penting dalam dunia akademik, salah satunya dalam penentuan Uang Kuliah Tunggal [4].

Klasifikasi adalah proses pengelompokan data atau objek ke dalam kategori atau kelas tertentu berdasarkan karakteristik atau fitur yang dimilikinya. Tujuan utama dari klasifikasi adalah untuk mempermudah analisis, pengambilan keputusan, atau prediksi terhadap data baru berdasarkan pola yang telah dikenali dari data sebelumnya [5],[6],[7]. Machine learning klasifikasi merupakan salah satu teknik supervised learning yang digunakan untuk memetakan input data ke dalam label atau kategori output yang telah ditentukan [7],[8]. Model klasifikasi dilatih menggunakan dataset berlabel sehingga dapat mengenali pola dan membuat prediksi terhadap data baru [9],[10].

Politeknik Negeri Lhokseumawe (PNL) merupakan salah satu perguruan tinggi negeri yang berada di bawah naungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, dan berlokasi di Kota Lhokseumawe, Provinsi Aceh. Penentuan besaran Uang Kuliah Tunggal (UKT) di PNL mengacu pada Surat Edaran internal kampus yang berlandaskan Peraturan Nomor 25 Tahun 2020 [12]. Selain itu, kebijakan tersebut turut mempertimbangkan berbagai regulasi pendukung lainnya, seperti Peraturan Nomor 55 Tahun 2013, perubahan-perubahan yang tertuang dalam Peraturan Nomor 73 Tahun 2014, Peraturan Nomor 22 Tahun 2015, Peraturan Nomor 39 Tahun 2016, serta Peraturan terbaru Nomor 2 Tahun 2024 [13]. Sistem pembayaran UKT secara resmi mulai diberlakukan di Politeknik Negeri Lhokseumawe sejak 22 Oktober 2020. Pelaksanaannya, mahasiswa dikelompokkan ke dalam delapan kategori UKT, mulai dari Kelompok I hingga Kelompok VIII, yang disesuaikan dengan kondisi ekonomi masing-masing mahasiswa.

Permasalahan ini terjadi karena penetapan UKT masih dilakukan secara manual dan bergantung pada interpretasi subjektif petugas, sehingga menimbulkan perbedaan penilaian dan ketidakkonsistenan dalam evaluasi kondisi ekonomi mahasiswa. Keterbatasan waktu dan sumber daya manusia dalam proses verifikasi juga menyebabkan pengelompokan UKT belum sepenuhnya akurat, sehingga sebagian mahasiswa kurang mampu masih berpotensi ditempatkan pada kelompok UKT yang tidak sesuai. Oleh karena itu, diperlukan solusi berbasis teknologi untuk meningkatkan akurasi dan efisiensi klasifikasi Uang Kuliah Tunggal. Metode LightGBM dipilih karena memiliki kemampuan pemrosesan data yang cepat, akurasi tinggi, serta efektif dalam menangani data berukuran besar dan kompleks.

Hasil penelitian yang dilakukan oleh Zili (2025) menunjukkan bahwa LightGBM mampu mencapai akurasi hingga 96,4% dalam klasifikasi tingkat literasi digital mahasiswa, yang menegaskan keunggulan LightGBM dalam menangani permasalahan klasifikasi di bidang pendidikan tinggi [14]. Penelitian lain oleh Abukader (2025) juga menemukan bahwa LightGBM yang dioptimasi melalui hyperparameter tuning menghasilkan nilai R^2 sekitar 0,94, yang menunjukkan kemampuan prediksi yang tinggi dan stabil pada data pendidikan [15]. Berdasarkan dua penelitian terdahulu tersebut, *Light Gradient Boosting Machine* (LightGBM) sebagai algoritma berbasis pohon keputusan dinilai unggul, efisien, dan relevan untuk diterapkan pada sistem klasifikasi berbasis data pendidikan, termasuk dalam penentuan kelompok Uang Kuliah Tunggal (UKT), karena mampu menghasilkan prediksi yang akurat, konsisten, dan mendukung pengambilan keputusan secara objektif.

Penelitian ini mampu menghasilkan sistem klasifikasi UKT yang lebih tepat dan adil, sehingga mahasiswa dari keluarga kurang mampu dapat ditempatkan pada golongan biaya sesuai kondisi ekonomi sebenarnya. Dengan demikian, keputusan penetapan UKT menjadi lebih objektif, berbasis data, dan tidak lagi bergantung pada subjektivitas manual, sekaligus meningkatkan efisiensi administrasi perguruan tinggi dalam proses verifikasi. Keterbaruan penelitian ini terletak pada penerapan pendekatan klasifikasi berbasis pembelajaran mesin yang memanfaatkan data institusional secara langsung untuk menghasilkan penetapan UKT yang lebih objektif, adil, dan transparan. Model yang dikembangkan bersifat aplikatif serta mampu mendukung

pengambilan keputusan secara efisien dan akuntabel dalam kebijakan UKT perguruan tinggi. Dampak praktisnya adalah terwujudnya kebijakan penetapan UKT yang transparan dan akuntabel, meningkatkan keadilan akses pendidikan bagi mahasiswa, serta memberikan dasar yang kuat bagi perbaikan kebijakan UKT di masa mendatang.

2. METODE

Bab ini menjelaskan metode penelitian yang digunakan sebagai dasar dalam pelaksanaan penelitian. Uraian metode mencakup tahapan pengolahan data, pemilihan dan penerapan metode *Light Gradient Boosting Machine* (LightGBM), teknik pembagian data, serta metode evaluasi kinerja model yang digunakan untuk memastikan hasil klasifikasi Uang Kuliah Tunggal diperoleh secara sistematis, terukur, dan dapat dipertanggungjawabkan secara ilmiah.

2.1 Metode LightGBM

Light Gradient Boosting Machine (LightGBM) adalah algoritma gradient boosting berbasis pohon keputusan yang dikembangkan oleh Microsoft [16]. Algoritma ini dirancang untuk memberikan performa lebih cepat, penggunaan memori lebih rendah, serta akurasi tinggi dibandingkan metode boosting tradisional seperti Gradient Boosting Decision Tree (GBDT) dan XGBoost.

LightGBM menggunakan teknik *leaf-wise growth*, yaitu membangun pohon dengan memperluas leaf yang menghasilkan penurunan loss terbesar. Berbeda dengan *level-wise growth* pada GBDT yang memperluas seluruh node pada level yang sama, *leaf-wise* membuat LightGBM lebih efisien dan mampu menangkap pola yang lebih kompleks [17], [18]. LightGBM juga menggunakan dua teknik utama untuk mempercepat perhitungan.

1. *Gradient-based One-Side Sampling* (GOSS)

Mengambil sampel dengan memprioritaskan data yang memiliki nilai gradien besar (lebih berpengaruh terhadap loss) dan mensampling sebagian data dengan gradien kecil.

2. *Exclusive Feature Bundling* (EFB)

Menggabungkan fitur-fitur sparse atau yang jarang aktif dalam satu bundle, sehingga mengurangi dimensi fitur tanpa mengurangi informasi penting.

Adapun rumus penting yang digunakan dalam LightGBM adalah sebagai berikut.

1. Objective Function (Fungsi Objektif)

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \Omega(f) \quad (1)$$

Dimana :

$l(y_i, \hat{y}_i)$ = Loss function (misal: logloss untuk klasifikasi)

$\Omega(f)$ = Regularisasi untuk mencegah overfitting

Regularisasi:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (2)$$

Dimana :

T = Jumlah daun

w_j^2 = Nilai output daun ke-j

$\gamma\lambda$ = Parameter regularisasi

2. Rumus Gain (Pemilihan Split Terbaik)

Rumus inti yang digunakan LightGBM untuk memilih *split* terbaik:

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \gamma \quad (3)$$

Dimana :

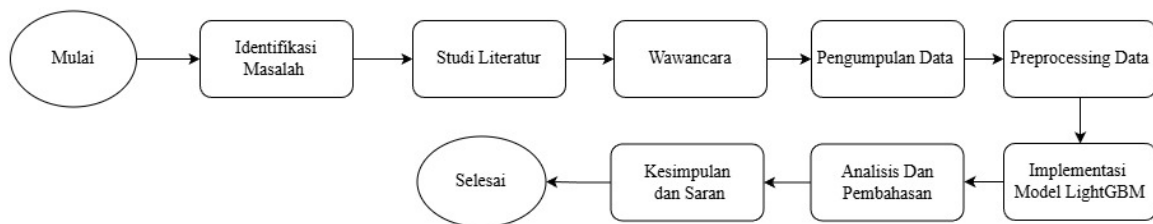
G_L^2, G_R^2 = Jumlah gradien di sisi kiri dan kanan

H_L, H_R = Jumlah hessian (gradien kedua)
 λ = Regularisasi L2
 γ = Penalti jumlah daun

Melalui kedua rumus tersebut, LightGBM mampu menentukan struktur pohon keputusan secara lebih efisien dan akurat. Fungsi objektif memastikan model tetap optimal dan tidak mengalami overfitting, sedangkan perhitungan *gain* digunakan untuk memilih pemisahan node terbaik pada setiap tahap pembentukan pohon. Kombinasi proses optimasi tersebut memungkinkan LightGBM menghasilkan model klasifikasi yang cepat, stabil, dan memiliki performa tinggi pada berbagai jenis dataset.

2.2 Tahapan Penelitian

Tahapan penelitian ini disusun untuk memberikan gambaran sistematis mengenai alur penelitian, mulai dari identifikasi masalah, pengumpulan dan pra pengolahan data, penerapan algoritma machine learning, hingga evaluasi hasil klasifikasi. Tahapan ini disusun agar penelitian berjalan terstruktur dan dapat dipertanggungjawabkan secara ilmiah. Adapun alur tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menggambarkan alur penelitian yang digunakan dalam proses perancangan dan pengembangan model klasifikasi. Penelitian dimulai dengan tahap identifikasi masalah, kemudian dilanjutkan dengan studi literatur untuk memperoleh dasar teori yang relevan. Setelah itu dilakukan wawancara untuk mendapatkan informasi pendukung, diikuti dengan pengumpulan data sebagai bahan analisis. Data yang terkumpul kemudian melalui tahap preprocessing guna memastikan kualitas dan kelayakan data sebelum digunakan. Selanjutnya dilakukan implementasi model machine learning, di mana model dibangun, diuji, dan dievaluasi. Hasil dari implementasi tersebut dianalisis secara sistematis pada tahap analisis dan pembahasan, sebelum akhirnya ditarik kesimpulan dan saran sebagai penutup penelitian. Proses berakhir setelah seluruh tahapan tersebut selesai dilaksanakan.

2.3 Variabel Penelitian

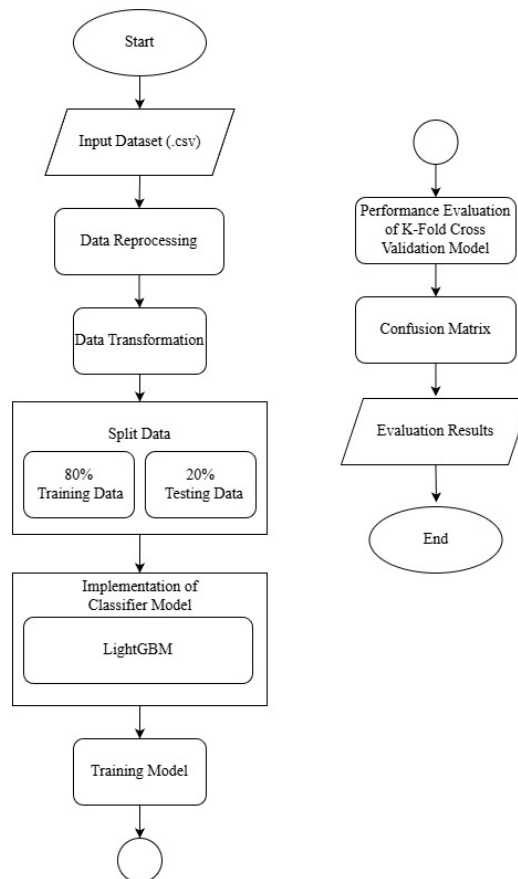
Variabel penelitian merupakan faktor-faktor yang dijadikan dasar atau acuan dalam proses analisis dan klasifikasi Uang Kuliah Tunggal (UKT) pada mahasiswa. Variabel-variabel ini diperoleh dari pengumpulan data mahasiswa yang mencakup berbagai aspek, seperti kondisi sosial, kondisi ekonomi, serta latar belakang keluarga. Dengan mempertimbangkan variabel-variabel tersebut, penelitian dapat melakukan evaluasi dan klasifikasi UKT secara lebih akurat dan relevan. Adapun variabel penelitian yang digunakan dalam penelitian ini disajikan secara rinci pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Variabel	Keterangan
1	Pekerjaan Ayah	Jenis pekerjaan ayah mahasiswa
2	Pekerjaan Ibu	Jenis pekerjaan ibu mahasiswa
3	Penghasilan Ayah	Jumlah pendapatan ayah per bulan
4	Penghasilan Ibu	Jumlah pendapatan ibu per bulan
5	Status Ayah	Kondisi keberadaan ayah
6	Status Ibu	Kondisi keberadaan ibu
7	Jumlah Tanggungan Keluarga	Banyaknya anggota keluarga yang ditanggung
8	Sumber Air	Jenis sumber air
9	Kepemilikan Rumah	Status kepemilikan rumah
10	Kondisi Rumah	Keadaan fisik rumah

2.4 Perancangan Metode LightGBM

Perancangan metode Light Gradient Boosting Machine (LightGBM) dalam penelitian ini dilakukan sebagai pendekatan klasifikasi berbasis pembelajaran mesin untuk menentukan kelompok Uang Kuliah Tunggal (UKT) secara lebih objektif dan akurat. Tahapan perancangan mencakup proses pengolahan data, pembentukan dan pelatihan model, serta evaluasi kinerja menggunakan metrik yang relevan, sehingga model yang dihasilkan diharapkan mampu memberikan hasil klasifikasi yang konsisten dan dapat mendukung pengambilan keputusan penetapan UKT berbasis data. Adapun perancangan dapat dilihat pada Gambar 2.



Gambar 2. Perancangan Metode LightGBM

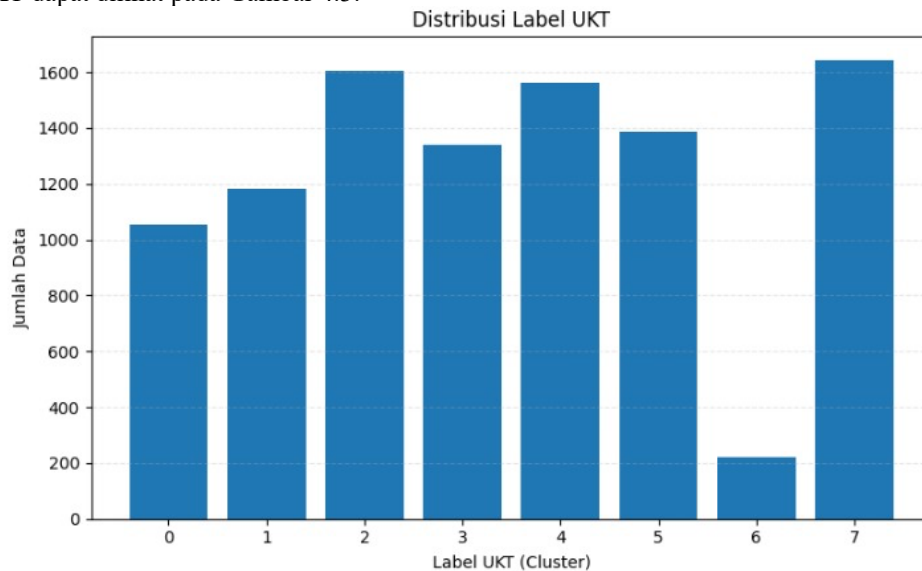
Gambar 2 menunjukkan alur perancangan dan implementasi metode *Light Gradient Boosting Machine* (LightGBM) untuk klasifikasi UKT yang dimulai dari proses input dataset dalam format *.csv*, kemudian dilanjutkan dengan tahap *data preprocessing* dan *data transformation* untuk menyiapkan data agar layak digunakan dalam pemodelan. Selanjutnya, data dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Data pelatihan digunakan pada tahap implementasi dan pelatihan model klasifikasi menggunakan algoritma LightGBM. Setelah model terbentuk, dilakukan evaluasi kinerja menggunakan metode K-Fold Cross Validation untuk memastikan kestabilan dan keandalan model, kemudian hasil evaluasi dianalisis menggunakan *confusion matrix* guna mengukur akurasi dan kesalahan klasifikasi. Tahapan akhir menghasilkan evaluasi performa model yang menjadi dasar penilaian keberhasilan sistem klasifikasi UKT.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian dan pembahasan yang diperoleh dari penerapan metode *Light Gradient Boosting Machine* (LightGBM) dalam klasifikasi Uang Kuliah Tunggal (UKT). Pembahasan difokuskan pada analisis karakteristik data, kinerja model berdasarkan metrik evaluasi yang digunakan, serta interpretasi hasil pengujian untuk menilai tingkat akurasi, stabilitas, dan keandalan model dalam mengklasifikasikan kelompok UKT secara objektif dan berbasis data.

3.1 Distribusi Kelas

Distribusi label UKT digunakan untuk melihat sebaran jumlah mahasiswa pada setiap kategori UKT. Analisis ini penting untuk memahami karakteristik data, terutama terkait keseimbangan antar kelas. Ketidakseimbangan jumlah data pada tiap kategori dapat memengaruhi kinerja model, karena algoritma klasifikasi cenderung bekerja lebih baik pada kelas dominan dibandingkan kelas minoritas. Adapun distribusi label UKT dapat dilihat pada Gambar 4.3.



Gambar 3. Distribusi Kelompok UKT

Gambar 3 menunjukkan distribusi label UKT pada dataset, yang memperlihatkan variasi jumlah data pada setiap klaster. Klaster dengan jumlah data terbesar adalah klaster 2, 4, dan 7, masing-masing dengan 1.604, 1.564, dan 1.645 entri. Hal ini menunjukkan bahwa ketiga klaster tersebut memiliki representasi yang lebih dominan dalam keseluruhan dataset. Sementara itu, klaster 0, 1, 3, dan 5 memiliki jumlah data yang relatif seimbang, yaitu berkisar antara 1.054 hingga 1.389 entri, sehingga tetap memberikan kontribusi signifikan terhadap komposisi data secara keseluruhan. Sebaliknya, klaster 6 memiliki jumlah data paling sedikit, yaitu hanya 222 entri. Perbedaan jumlah antara klaster-klaster tersebut menggambarkan kecenderungan karakteristik data mahasiswa yang lebih terkonsentrasi pada kategori tertentu dibandingkan kategori lainnya. Pemahaman terhadap distribusi ini penting sebagai dasar untuk mengetahui representasi setiap label dalam dataset, sehingga hasil analisis dan pemodelan dapat diinterpretasikan secara proporsional sesuai dengan komposisi data yang tersedia.

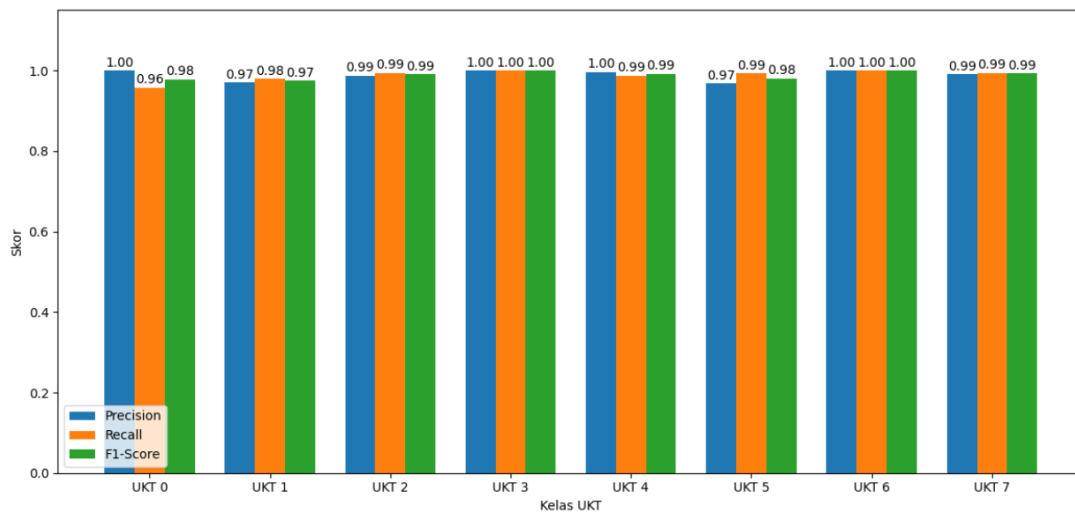
3.2 Model LightGBM

Hasil pengujian model *Light Gradient Boosting Machine* (LightGBM) yang digunakan dalam proses klasifikasi Uang Kuliah Tunggal. Hasil yang ditampilkan mencakup kinerja model berdasarkan metrik evaluasi yang digunakan, sehingga dapat diketahui tingkat akurasi dan keandalan model dalam mengklasifikasikan data mahasiswa. Adapun hasil pengujian model LightGBM dapat dilihat pada Tabel 2.

Tabel 2. Classification Report Confusion Matrix

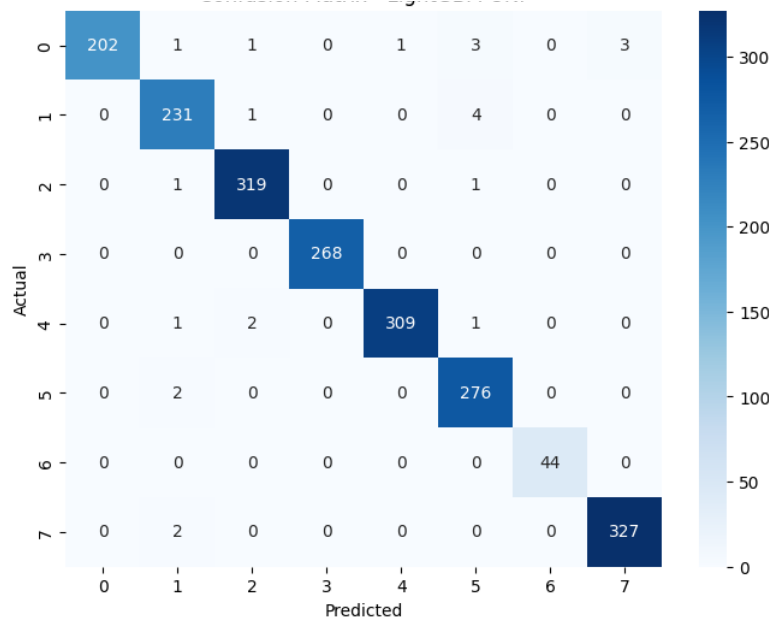
Class	Precision	Recal	F1-Score	Support
1	1.00	0.96	0.98	211
2	0.97	0.98	0.97	236
3	0.99	0.99	0.99	321
4	1.00	1.00	1.00	268
5	1.00	0.99	0.99	313
6	0.97	0.99	0.98	278
7	1.00	1.00	1.00	44
8	0.99	0.99	0.99	329
Accuracy			0.98	2000

Berdasarkan Tabel 2 hasil *classification report* menunjukkan bahwa model LightGBM memiliki kinerja yang sangat baik dalam mengklasifikasikan data UKT. Secara keseluruhan, model mencapai akurasi sebesar 98% dari total 2.000 data uji, yang mengindikasikan tingkat ketepatan prediksi yang tinggi. Nilai *precision*, *recall*, dan *F1-score* pada setiap kelas UKT berada pada rentang 0,97 hingga 1,00, menunjukkan bahwa model mampu mengklasifikasikan masing-masing kelas secara konsisten dan seimbang. Kelas 4 dan 7 memperoleh performa sempurna dengan nilai *precision*, *recall*, dan *F1-score* sebesar 1,00, yang menandakan tidak adanya kesalahan klasifikasi pada kelas tersebut. Sementara itu, kelas lain juga menunjukkan tingkat kesalahan yang sangat rendah, sehingga dapat disimpulkan bahwa model LightGBM efektif dan andal dalam melakukan klasifikasi UKT secara akurat dan merata pada seluruh kelompok.



Gambar 4. Grafik Confusion Matrix

Gambar 4 menunjukkan perbandingan nilai *precision*, *recall*, dan *F1-score* untuk setiap kelas UKT hasil klasifikasi menggunakan metode LightGBM. Secara umum, seluruh kelas UKT memiliki nilai kinerja yang sangat tinggi dengan rentang skor antara 0,96 hingga 1,00, yang menandakan bahwa model mampu melakukan klasifikasi secara akurat dan konsisten pada masing-masing kelas. Beberapa kelas, seperti UKT 3, UKT 4, dan UKT 6, bahkan mencapai nilai sempurna pada ketiga metrik evaluasi, menunjukkan tidak adanya kesalahan klasifikasi pada kelas tersebut. Sementara itu, kelas UKT lainnya juga memperlihatkan keseimbangan yang baik antara *precision* dan *recall*, sehingga menghasilkan *F1-score* yang tinggi. Hasil ini menegaskan bahwa model LightGBM memiliki performa yang stabil dan andal dalam mengklasifikasikan kelompok UKT secara merata.

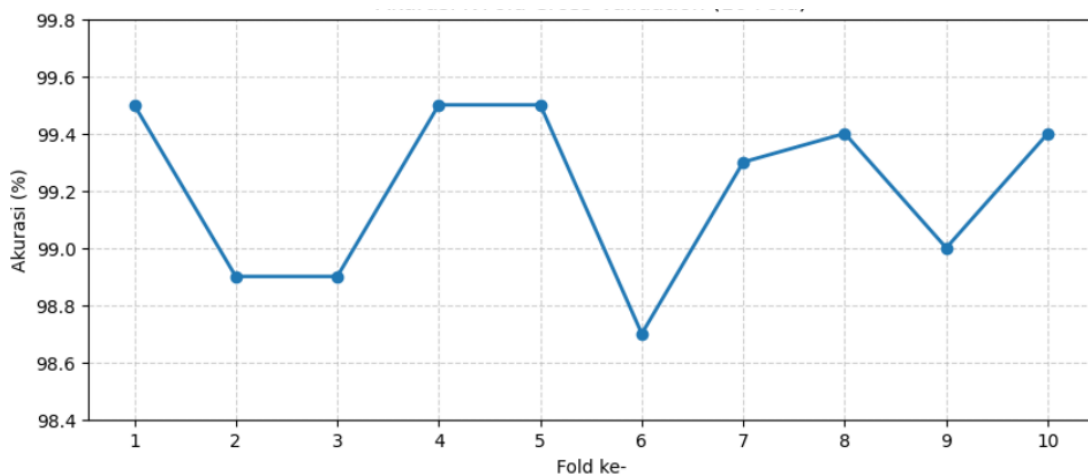


Gambar 5. Confusion Matrix LightGBM

Gambar 5 *confusion matrix* yang ditampilkan, model LightGBM menunjukkan kinerja klasifikasi yang sangat baik dengan jumlah prediksi benar yang dominan pada diagonal utama. Pada kelas 0, sebanyak 202 data berhasil diklasifikasikan dengan benar, dengan hanya 9 data yang salah klasifikasi ke kelas lain. Kelas 1 memiliki 231 prediksi benar dari total 236 data, sedangkan kelas 2 menunjukkan 319 prediksi benar dari 321 data. Seluruh data pada kelas 3 berhasil diklasifikasikan dengan benar sebanyak 268 data tanpa kesalahan. Kelas 4 memiliki 309 prediksi benar dari 313 data, kelas 5 sebanyak 276 prediksi benar dari 278 data, dan kelas 6 menunjukkan klasifikasi sempurna dengan 44 data seluruhnya terprediksi dengan benar. Pada kelas 7, sebanyak 327 data berhasil diklasifikasikan dengan tepat dari total 329 data. Jumlah kesalahan klasifikasi yang relatif kecil dan tersebar pada kelas yang berdekatan menunjukkan bahwa model LightGBM memiliki tingkat akurasi dan keandalan yang tinggi dalam mengklasifikasikan kelompok UKT.

3.3 K-Fold Cross Validation

Teknik K-Fold Cross Validation bertujuan untuk mengevaluasi kinerja model secara lebih komprehensif dengan membagi dataset ke dalam beberapa bagian (*fold*) yang digunakan secara bergantian sebagai data latih dan data uji. Metode ini membantu mengurangi bias evaluasi serta memastikan bahwa model memiliki kemampuan generalisasi yang baik. Pada penelitian ini digunakan 10 iterasi (*10-Fold Cross Validation*) sehingga setiap subset data memperoleh peran yang sama dalam proses pelatihan dan pengujian. Hasil pengujian menggunakan K-Fold Cross Validation dapat dilihat pada Gambar 6.



Gambar 6. Hasil Pengujian 10 Iterasi Validation

Gambar 6 menunjukkan hasil pengujian menggunakan metode 10-Fold Cross Validation, model LightGBM menunjukkan kinerja yang sangat konsisten dengan nilai akurasi pada setiap fold berada pada rentang 98,7% hingga 99,5%. Rata-rata akurasi yang diperoleh sebesar 99,21% dengan nilai standar deviasi 0,29%, yang mengindikasikan bahwa variasi kinerja model antar fold relatif kecil. Hal ini menunjukkan bahwa model memiliki tingkat stabilitas dan generalisasi yang sangat baik terhadap data yang berbeda, sehingga dapat disimpulkan bahwa metode LightGBM mampu menghasilkan performa klasifikasi UKT yang andal dan konsisten serta layak diterapkan dalam sistem pendukung keputusan penetapan UKT berbasis data.

3.4 Perhitungan LightGBM

Perhitungan model *Light Gradient Boosting Machine* (LightGBM) pada penelitian ini dilakukan melalui tahapan pelatihan menggunakan data latih, pembentukan pohon keputusan secara bertahap berdasarkan prinsip *gradient boosting*, serta pengoptimalan fungsi objektif untuk meminimalkan nilai kesalahan klasifikasi. Seluruh proses perhitungan dilakukan secara iteratif dengan memanfaatkan parameter yang telah ditentukan guna menghasilkan model klasifikasi UKT yang optimal. Adapun perhitungan LightGBM dapat dilihat pada Tabel 3,4,5,6 dan 7.

Tabel 3. Data dan Nilai Numerik

Fitur	Nilai Asli	Nilai Numerik
Pekerjaan Ayah	Lainnya	7
Penghasilan Ayah	Rp2.750.001–Rp3.000.000	12
Status Ayah	Bercerai	1
Pekerjaan Ibu	Petani	1
Penghasilan Ibu	Rp750.001–Rp1.000.000	4
Status Ibu	Hidup	1
Jumlah Tanggungan	20 Orang	20
Kepemilikan Rumah	Menumpang Tanpa Izin	1
Sumber Air	PDAM	2
Kondisi Rumah	Buruk	0

Tabel 3 ini menyajikan data awal responden yang telah melalui proses transformasi dari data kategorikal menjadi nilai numerik. Setiap fitur sosial ekonomi, seperti pekerjaan dan penghasilan orang tua, status keluarga, jumlah tanggungan, serta kondisi tempat tinggal, dikonversi ke dalam bentuk numerik agar dapat diproses oleh model LightGBM pada tahap pelatihan dan pengujian.

Tabel 4. Probabilitas Prior dan Log-Odds Awal

Class	Rumus Probabilitas	P	$F_0 = \ln(P)$
Cluster 1	1 / 3	0,333	-1,099
Cluster 3	1 / 3	0,333	-1,099
Cluster 5	1 / 3	0,333	-1,099

Tabel 4 menunjukkan perhitungan probabilitas awal (*prior probability*) untuk setiap kelas cluster yang diasumsikan memiliki peluang yang sama, yaitu 0,333. Nilai probabilitas tersebut kemudian ditransformasikan ke dalam bentuk log-odds sebagai nilai prediksi awal (F_0) yang menjadi dasar pembentukan pohon keputusan pada iterasi pertama LightGBM.

Tabel 5. Update Nilai Prediksi Setelah Iterasi ke-1

Class	Output Leaf	Rumus Update	F_1
Cluster 1	-1,50	$-1,099 + (0,1 \times -1,50)$	-1,249
Cluster 3	2,00	$-1,099 + (0,1 \times 2,00)$	-0,899
Cluster 5	-1,25	$-1,099 + (0,1 \times -1,25)$	-1,224

Tabel 5 menggambarkan proses pembaruan nilai prediksi setelah iterasi pertama, di mana output daun (*leaf output*) dari pohon keputusan dikalikan dengan *learning rate* dan ditambahkan ke nilai log-odds awal. Hasil perhitungan ini menghasilkan nilai prediksi baru (F_1) untuk masing-masing cluster sebagai bagian dari mekanisme *gradient boosting*.

Tabel 6. Probabilitas Akhir (Softmax)

Class	$\exp(F_1)$	Rumus Softmax	Probabilitas
Cluster 1	0,287	$0,287 / 0,988$	0,290
Cluster 3	0,407	$0,407 / 0,988$	0,412
Cluster 5	0,294	$0,294 / 0,988$	0,298

Tabel 6 menampilkan hasil konversi nilai prediksi F_1 ke dalam bentuk probabilitas menggunakan fungsi *softmax*. Proses ini bertujuan untuk menormalisasi nilai prediksi sehingga total probabilitas seluruh kelas bernilai satu, sekaligus menunjukkan tingkat keyakinan model terhadap masing-masing cluster.

Tabel 7. Hasil Prediksi Akhir

Nama	Probabilitas Tertinggi	Prediksi	Aktual
Al Khaidar	0,412	Cluster 3	Cluster 3

Tabel 7 menunjukkan hasil akhir klasifikasi UKT berdasarkan probabilitas tertinggi yang dihasilkan oleh model LightGBM. Berdasarkan perhitungan, Cluster 3 memiliki probabilitas terbesar sebesar 0,412 sehingga ditetapkan sebagai hasil prediksi, yang sesuai dengan data aktual, menunjukkan bahwa model mampu melakukan klasifikasi dengan tepat pada contoh data tersebut.

4. KESIMPULAN

Berdasarkan hasil analisis dan pengujian yang telah dilakukan, dapat disimpulkan bahwa metode Light Gradient Boosting Machine (LightGBM) mampu memberikan kinerja yang sangat baik dalam mengklasifikasikan kelompok Uang Kuliah Tunggal (UKT). Distribusi kelas UKT menunjukkan adanya variasi jumlah data antar klaster, namun model tetap mampu menangani perbedaan tersebut dengan baik tanpa menunjukkan penurunan kinerja yang signifikan pada kelas minoritas. Hasil evaluasi menggunakan classification report dan confusion matrix menunjukkan tingkat akurasi yang tinggi, yaitu sebesar 98%, dengan nilai precision, recall, dan F1-score yang konsisten pada seluruh kelas UKT. Selain itu, pengujian menggunakan 10-Fold Cross Validation menghasilkan rata-rata akurasi sebesar 99,21% dengan standar deviasi yang rendah, menandakan bahwa model memiliki stabilitas dan kemampuan generalisasi yang sangat baik. Dengan demikian, LightGBM dapat dinyatakan sebagai metode yang efektif, andal, dan layak digunakan sebagai dasar dalam sistem pendukung keputusan untuk penetapan UKT berbasis data mahasiswa.

REFERENSI

- [1] A. Khaidar, M. Arhami, and M. Abdi, "Application of the Random Forest Method for UKT Classification at Politeknik Negeri Lhokseumawe," *J. Artif. Intell. Softw. Eng.*, vol. 4, no. 2, pp. 94–103, 2024.
- [2] S. Kurnia and A. Khaidar, "Perbandingan Metode Machine Learning Menggunakan Metode Support Vector Machine Dan Artificial Neural Network Dalam Memprediksi Serangan Jantung," *J. Inform. Kaputama (JIK)*, vol. 9, no. 2, pp. 87–94, 2025.
- [3] R. Susetyoko, W. Yuwono, and dkk, *Perbandingan metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron pada klasifikasi Uang Kuliah Tunggal (UKT)*. Infomedia: Jurnal Ilmu Computer & Teknologi Informasi, 2022.
- [4] E. B. Setiawan, "Rekonstruksi kebijakan hukum kenaikan uang kuliah tunggal berdasarkan tanggung jawab konstitusional negara," *Lex Renaiss.*, vol. 9, no. 2, pp. 123–145, 2024, doi: 10.20885/jlr.vol9.iss2.art5.
- [5] K. Taha, "A Comprehensive Survey of Text Classification Techniques," *Expert Syst. Appl.*, vol. 202, pp. 117–134, 2024.
- [6] M. Chen and Z. Liu, "Predicting performance of students by optimizing tree components of random forest using genetic algorithm," *Heliyon*, vol. 10, no. 12, p. e32570, 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e32570>.
- [7] N. Nurdin, "Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor," *J. Inform. Dan Tek. Elektro Terap.*, vol. 12, no. 2, 2025.
- [8] V. Sheth, "A Comparative Analysis of Machine Learning Algorithms for Classification Tasks," *Procedia Comput. Sci.*, vol. 187, pp. 1159–1166, 2022.
- [9] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput Sci*, vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00592-x.
- [10] U. I. Akpan, "Review of Classification Algorithms with Changing Inter-Class Distances," *J. King Saud*, 2021.
- [11] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer. Behav Ther," 2021. doi: 10.1016/j.beth.2020.05.002.
- [12] K. P. Kebudayaan Republik Indonesia, "Peraturan Menteri Pendidikan dan Kebudayaan Nomor 2 Tahun 2024 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan," 2024, *Jakarta*.
- [13] K. P. Kebudayaan Republik Indonesia, "Peraturan Menteri Pendidikan dan Kebudayaan Nomor 25 Tahun 2020 tentang Standar Satuan Biaya Operasional Pendidikan Tinggi pada Perguruan Tinggi

- Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan,” 2020, *Jakarta*. [Online]. Available: <https://peraturan.bpk.go.id/Details/163756/permendikbud-no-25-tahun-2020>
- [14] A. H. A. Zili and S. A. Amellia, “Predicting digital literacy levels in higher education: A LightGBM model integrating feature selection for improved accuracy,” *G-Tech J. Teknol. Terap.*, vol. 9, no. 4, 2025.
- [15] A. Abukader, A. Alzubi, O. R. Adegboye, M. D. P. I. Zili, A. H. A., and S. A. Amellia, “Intelligent system for student performance prediction: Educational data mining approach using metaheuristic-optimized LightGBM with SHAP-based learning analytics,” *Appl. Sci.*, vol. 15, no. 20, p. 10875, 2025.
- [16] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, “A comparative analysis of gradient boosting algorithms,” *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, 2021.
- [17] Y. Liu, Y. Wang, J. Zhang, and L. Wang, “An improved LightGBM model for classification of imbalanced data,” *IEEE Access*, vol. 8, pp. 13294–13305, 2021.
- [18] S. Zhang, Y. Wang, and X. Liu, “LightGBM-based prediction model for educational data mining,” *J. Intell. Fuzzy Syst.*, vol. 41, no. 3, pp. 3895–3906, 2020.