

Application of the Random Forest Method for UKT Classification at Politeknik Negeri Lhokseumawe

Al Khaidar^{1*}, Muhammad Arhami², Musta'inul Abdi³

^{1,2,3} Jurusan Tekniknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

Informasi Artikel

Diterima : 23 September, 2024
Revisi : 1 November 2024
Publikasi : 30 November 2024

Kata Kunci:

Klasifikasi
Random Forest
Uang Kuliah Tunggal

ABSTRAK

Klasifikasi adalah pengelompokan sistematis objek, gagasan, buku, atau benda lain ke dalam kelas tertentu berdasarkan ciri-ciri yang sama. Salah satu penerapannya adalah dalam pengelompokan Uang Kuliah Tunggal (UKT), yaitu biaya kuliah yang dibayar setiap semester atau tahun ajaran berdasarkan kemampuan ekonomi mahasiswa. Politeknik Negeri Lhokseumawe menerapkan sistem UKT sebagai metode pembayaran biaya pendidikan sejak 22 Oktober 2020, dengan biaya kuliah dibagi menjadi 8 kelompok. Namun, terdapat beberapa masalah, seperti penempatan mahasiswa kurang mampu pada kelompok UKT yang masih belum sesuai dan keterbatasan keakuratan proses pengelompokan karena dilakukan secara manual. Untuk mengatasi masalah ini, dirancang sistem klasifikasi UKT menggunakan metode *Random Forest*. *Random Forest* adalah algoritma *machine learning* yang menggabungkan beberapa pohon keputusan untuk prediksi yang lebih akurat. Pengujian metode *Random Forest* menggunakan *cross validation* menunjukkan akurasi rata-rata 95%. Evaluasi dengan *confusion matrix* menunjukkan akurasi 94%, dengan nilai *precision*, *recall*, dan *f1-score* yang bervariasi untuk setiap kelompok.

ABSTRACT

Classification is the systematic grouping of objects, ideas, books, or other items into specific classes based on similar characteristics. One of its applications is in the grouping of tuition fees, which are fees paid each semester or academic year based on the student's economic ability. However, there are several issues, such as the placement of underprivileged students into fee groups that are still not appropriate and the limited accuracy of the grouping process due to it being done manually. To address these issues, a classification system was designed using the Random Forest method. Random Forest is a machine learning algorithm that combines multiple decision trees for more accurate predictions. Testing the Random Forest method using *cross-validation* shows an average accuracy of 95%. Evaluation with a *confusion matrix* shows an accuracy of 94%, with varying values of *precision*, *recall*, and *f1-score* for each group.

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



*Penulis Koresponden

Email: alkhaidarkutablang@gmail.com

Cara sitasi IEEE:

1. A. Khaidar, M. Arhami, and M. Abdi, "Application of the Random Forest Method for UKT Classification at Politeknik Negeri Lhokseumawe," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol.4, no.2, pp. 94-103, Bulan tahun. doi: 10.30811/jaise.v4i2.6131

1. PENDAHULUAN

Klasifikasi adalah pengelompokan yang sistematis dari sejumlah obyek, gagasan atau benda-benda lain kedalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama [1]. Selain itu juga klasifikasi merupakan proses pengelompokan, artinya mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama [2]. Klasifikasi digunakan di berbagai bidang, termasuk ilmu komputer, biologi, dan ilmu sosial [3]. Klasifikasi juga dapat diterapkan dalam pengelompokan Uang Kuliah Tunggal (UKT) yaitu pengelompokan UKT yang ditanggungkan kepada setiap mahasiswa berdasarkan kemampuan ekonominya [4]. UKT atau Uang Kuliah Tunggal adalah biaya kuliah yang harus dibayar oleh mahasiswa setiap semester atau tahun ajaran. Besar UKT yang harus dibayar oleh mahasiswa tergantung pada program studi, jenjang pendidikan, dan kategori mahasiswa [5]. Politeknik Negeri Lhokseumawe merupakan perguruan tinggi negeri di lingkungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi yang berkedudukan di Kota Lhokseumawe, Provinsi Aceh [6]. Pengelompokan Uang Kuliah Tunggal di Politeknik Negeri Lhokseumawe dilakukan berdasarkan Surat Edaran yang berpedoman pada Peraturan Nomor 25 Tahun 2020 [7]. Hal ini juga mempertimbangkan beberapa peraturan terkait, yaitu Peraturan Nomor 55 Tahun 2013, perubahan pada Peraturan Nomor 55 Tahun 2013 yang tertuang dalam Peraturan Nomor 73 Tahun 2014, Peraturan Nomor 22 Tahun 2015, Peraturan Nomor 39 Tahun 2016, dan Peraturan Nomor 2 Tahun 2024. Politeknik Negeri Lhokseumawe menerapkan sistem Uang Kuliah Tunggal (UKT) pada tanggal 22 Oktober 2020 sebagai metode pembayaran biaya pendidikan. Dalam penerapan UKT di Politeknik Negeri Lhokseumawe, biaya kuliah dibagi menjadi 8 kelompok, yakni Kelompok I, Kelompok II, Kelompok III, Kelompok IV, Kelompok V, Kelompok VI, Kelompok VII dan Kelompok VIII [8].

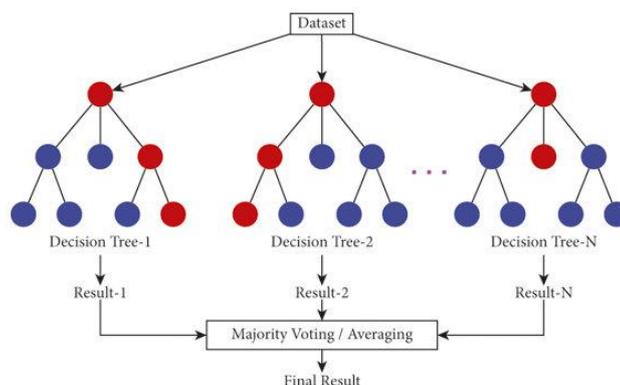
Permasalahan diidentifikasi berdasarkan hasil wawancara dengan staf akademik yaitu: Sebagian mahasiswa dengan kondisi ekonomi yang kurang mampu ditempatkan pada kelompok uang kuliah yang masih belum sesuai; Keakuratan proses pengelompokan UKT masih terbatas karena masih dilakukan secara manual.

Berdasarkan masalah yang diidentifikasi, penulis merancang sebuah sistem klasifikasi UKT menggunakan metode Random Forest. Random Forest adalah algoritma yang menggabungkan beberapa pohon keputusan untuk membuat prediksi yang lebih kuat. Setiap pohon keputusan dibuat secara acak dari data training, dan hasil akhirnya diambil melalui voting [9]. Metode ini dipilih karena Random Forest dalam sistem klasifikasi UKT kemampuannya untuk mengatasi kompleksitas data tanpa memerlukan optimasi numerik yang rumit. Keunggulan ini menjadikan Random Forest sebagai pilihan yang tepat untuk mengimplementasikan sistem klasifikasi UKT yang mudah digunakan dan dapat diandalkan dalam menghadapi data yang beragam. Oleh karena itu, penulis mengusulkan untuk merancang sebuah sistem klasifikasi UKT menggunakan algoritma *Random Forest*.

2. METODE

2.1. Metode *Random Forest*

Pada tahun 1995, tin kam ho dari bell labs memperkenalkan random forest sebagai algoritma ensemble untuk tahap data mining dengan tujuan klasifikasi dan regresi [10]. Algoritma ini menggunakan teknik bagging untuk membangun kumpulan decision tree. Beberapa kelebihan random forest, termasuk kemampuannya menghasilkan hasil klasifikasi yang baik dengan tingkat kesalahan yang lebih rendah, kemampuan menangani jumlah data yang besar, dan efektivitasnya dalam mengatasi masalah missing data [10]. Random forest beroperasi melalui dua fase esensial. Pada tahap awal, dilakukan penggabungan sejumlah n pohon keputusan untuk membentuk struktur yang kompleks dan kuat yang dikenal sebagai random forest. Selanjutnya, dalam tahap kedua, setiap pohon keputusan yang telah terbentuk pada fase sebelumnya digunakan untuk membuat prediksi [11]. Proses ini memanfaatkan kekuatan kolektif dari seluruh hutan keputusan, menghasilkan prediksi yang lebih akurat.



Gambar 1. Metode random forest

Random forest menghasilkan sejumlah pohon keputusan acak, dan kelas yang diprediksi berasal dari proses klasifikasi yang dipilih berdasarkan modus kelas yang paling banyak muncul dari pohon keputusan yang terbentuk [12]. Algoritma atau langkah-langkah dalam membangun random forest pada dataset yang terdiri dari n observasi dan p variabel penjelas (predictor) adalah sebagai berikut. Pada tahapan ini proses penyusunan model dan prediksi menggunakan random forest [13].

1. Tahapan bootstrapping

Untuk mulai membangun random forest langkah pertama yang dilakukan adalah pengambilan sampel secara acak berukuran n dari kumpulan data asli dengan pengembalian.

2. Tahapan random feature selection

Dalam tahap ini, pohon dibangun tanpa pemangkasan hingga mencapai ukuran maksimum. Proses pemilihan variabel prediktor dilakukan secara acak, dengan jumlah variabel yang dipilih (m) jauh lebih kecil dari total variabel (p). Pemilihan variabel terbaik dilakukan berdasarkan m prediktor. Berikut adalah contoh dari proses pembangunan decision tree [13].

Dalam pembentukan pohon keputusan, langkah pertama yang dilakukan adalah mengukur tingkat ketidakmurnian atribut dan nilai information gain. Untuk menghitung tingkat ketidakmurnian, kita menggunakan nilai entropy, yang dapat dihitung dengan menggunakan persamaan (1) untuk satu atribut atau persamaan (2) untuk dua atribut menggunakan tabel frekuensi. Selanjutnya, nilai information gain dapat ditentukan dengan menggunakan persamaan (3).

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (1)$$

Keterangan

- S = Himpunan dataset
- C = Jumlah kelas
- P_i = Probabilitas frekuensi kelas ke- i dalam dataset

$$Entropy(T, X) = \sum_{c \in X} P(c) E(c) \quad (2)$$

Keterangan

- (T, X) = Atribut T dan Atribut X
- $P(c)$ = Probabilitas kelas atribut
- $E(c)$ = Nilai Entropy kelas atribut

$$Gain(A) = Entropy(S) - \sum_{i=1}^k P(c) E(c) \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)$$

Keterangan

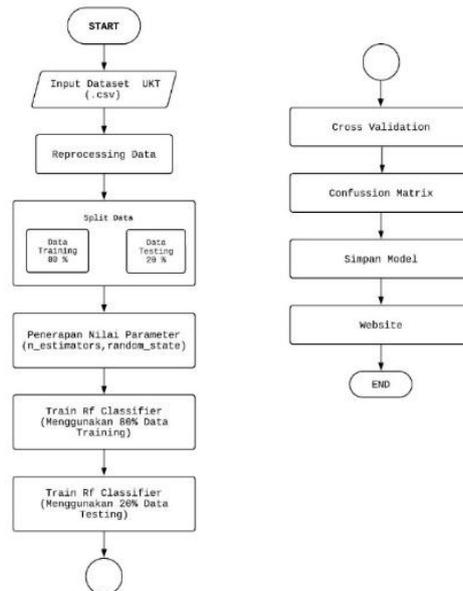
- S = Himpunan dataset
- A = Atribut
- $|S_i|$ = Jumlah sampel untuk nilai i
- $|S|$ = Jumlah seluruh sampel data
- $Entropy(S_i)$ = Entropy untuk sampel yang memiliki nilai i

2.2. Kebutuhan data

Data yang digunakan dalam penelitian ini merupakan dataset dari empat tahun terakhir. Data uji klasifikasi uang kuliah tunggal mencakup variabel pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, status ayah, status ibu, jumlah kendaraan roda empat (mobil), jumlah kendaraan roda dua (motor), jumlah tanggungan, kepemilikan rumah, watt listrik, dan kondisi rumah. Pengumpulan data dilakukan secara primer melalui wawancara langsung dengan staf akademik di politeknik negeri lhokseumawe.

2.3. Perancangan metode

Berikut merupakan gambaran flowchart dari implementasi metode random forest dapat dilihat pada gambar 2.

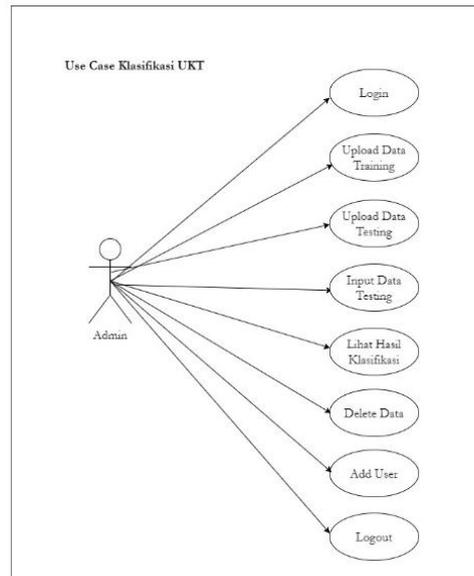


Gambar 2. Flowchart metode random forest

Gambar 2 menjelaskan penelitian ini menggunakan dataset yang berisi data uang kuliah tunggal (UKT) mahasiswa di politeknik negeri lhokseumawe. Langkah awal adalah reprocessing data, yaitu membersihkan entri yang tidak lengkap atau null serta melakukan normalisasi data. Ini bertujuan memastikan data yang digunakan memiliki kualitas baik. Setelah itu, data dibagi melalui proses split data menjadi 80% data training dan 20% data testing. Data training dipakai untuk melatih model, sedangkan data testing digunakan untuk menguji performa model. Selanjutnya, dilakukan penerapan nilai parameter seperti jumlah pohon ($n_estimators$) dan $random_state$. Hal ini penting untuk mengoptimalkan kinerja model. Model random forest kemudian dilatih melalui tahap train rf classifier, di mana model mempelajari pola dari data. Setelah pelatihan selesai, tahap test rf classifier dilakukan untuk menguji prediksi model dengan data uji. Untuk menghindari *overfitting* dan memastikan generalisasi model, digunakan metode cross validation, yang membagi dataset menjadi beberapa subset. evaluasi model dilakukan dengan confusion matrix, yang menghasilkan metrik seperti akurasi, presisi, recall, spesifisitas, dan f1-score. Setelah model dilatih, dilakukan simpan model untuk menyimpan konfigurasi model, sehingga dapat digunakan tanpa pelatihan ulang. Akhirnya, model terbaik akan diterapkan pada web klasifikasi ukt, di mana pengguna dapat melakukan klasifikasi ukt mahasiswa menggunakan model yang telah dilatih.

2.4. Use case diagram

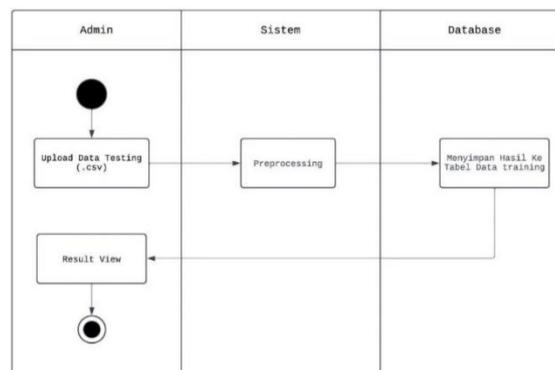
Use case diagram pada sistem klasifikasi ukt menggunakan metode random forest digunakan untuk menggambarkan interaksi antara aktor-aktor dan fungsionalitas utama sistem. Use case tersebut mencakup aksi-aksi seperti login, upload data training, upload data testing, lihat data klasifikasi, add users dan logout yang dapat dilakukan oleh pengguna pada sistem. Adapun rancangan use case pada penelitian ini dapat dilihat pada gambar 3.



Gambar 3. use case diagram

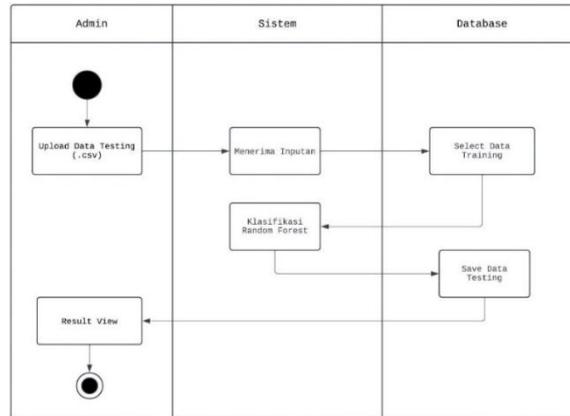
2.5. Activity Diagram

Activity diagram adalah diagram yang memodelkan aliran aktivitas pada sistem. Adapaun *activity diagram* pada sistem ditunjukkan pada gambar 4.



Gambar 4 activity diagram upload data training

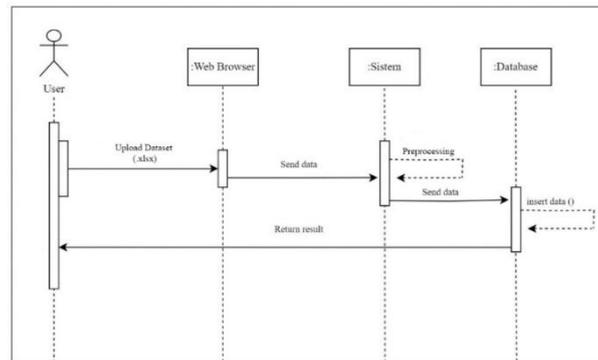
Gambar 4 menggambarkan proses upload data training oleh pengguna dimulai dengan pengguna mengunggah file .csv ke dalam sistem, yang kemudian sistem melakukan preprocessing. Selanjutnya, data training yang telah diunggah disimpan dalam sebuah tabel di dalam database. Setelah proses penyimpanan selesai, admin dapat mengakses hasil dari data training dengan meminta sistem untuk menampilkan dari tabel data training yang telah disimpan sebelumnya. Gambar 5 menggambarkan proses upload data testing dalam format .csv oleh admin ke dalam sistem, yang kemudian sistem menerima inputan tersebut. Setelah itu, sistem melakukan seleksi terhadap data training yang telah tersimpan dalam database untuk melakukan proses klasifikasi menggunakan metode random forest. Hasil dari proses klasifikasi tersebut kemudian disimpan dalam database dalam sebuah tabel yang disebut sebagai tabel data testing. Selanjutnya, admin dapat melihat hasil dari klasifikasi yang dilakukan oleh sistem.



Gambar 5 activity diagram upload data testing

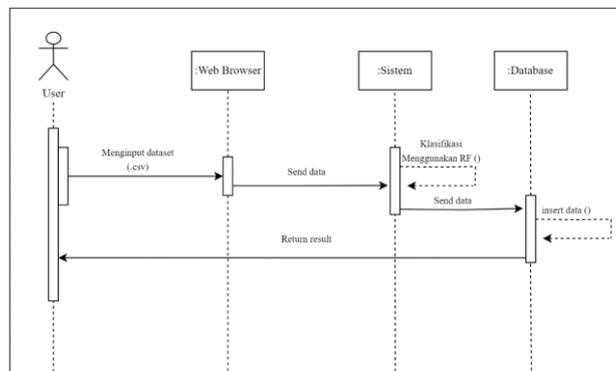
2.6. Sequence diagram

Sequence diagram menggambarkan urutan waktu dari aksi dan pesan dalam sistem. Setiap pesan menandai komunikasi antar objek. Sequence diagram sistem klasifikasi ukt menggunakan metode random forest di politeknik negeri lhokseumawe.



Gambar 6. Sequence diagram upload data training

Gambar 6 menjelaskan alur tahapan upload data training yang akan dilakukan oleh pengguna pada sistem. Proses dimulai ketika pengguna memilih untuk mengunggah data training dalam format .xlsx melalui antarmuka pengguna yang disediakan oleh situs web. Setelah pengguna mengunggah data, sistem akan melakukan preprocessing data tersebut dan mengirimkannya ke basis data yang terhubung. Di dalam database, data training akan disimpan secara aman dan terstruktur sesuai dengan skema yang telah ditentukan sebelumnya. Setelah data berhasil disimpan, pengguna dapat dengan mudah mengaksesnya kembali melalui antarmuka pengguna untuk melihatnya.



Gambar 7. Sequence diagram upload data testing

Gambar 7 menggambarkan dengan rinci alur tahapan unggah data testing yang dilakukan oleh pengguna pada sistem. Proses dimulai saat pengguna mengunggah data dalam format .xlsx ke dalam website yang disediakan. Setelah data diunggah, situs web akan meneruskan data tersebut ke sistem. Di dalam sistem,

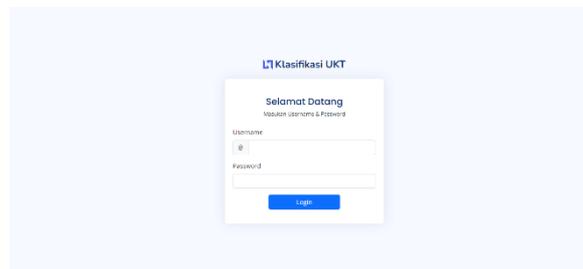
proses klasifikasi dilakukan menggunakan metode random forest terhadap data yang diunggah. Setelah proses klasifikasi selesai, hasil klasifikasi akan disimpan dalam database. Selanjutnya, pengguna dapat mengakses hasil klasifikasi tersebut melalui antarmuka pengguna yang disediakan oleh sistem untuk melihat dan menganalisisnya sesuai kebutuhan.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil implementasi metode pada klasifikasi UKT menggunakan Metode Random Forest serta hasil dan pengujian sistem yang telah dibangun. Hasil penelitian dan pengujian sistem ini akan dijelaskan secara lebih rinci sebagai berikut.

3.1. Halaman *Login*

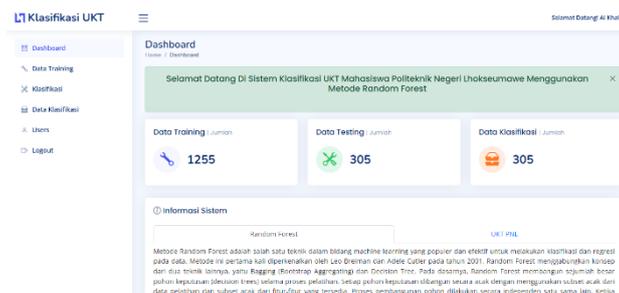
Halaman Login merupakan langkah krusial dalam sebuah sistem untuk menjaga keamanan dan aksesibilitas antara pengguna dan sistem. Hanya pengguna yang telah terdaftar sebelumnya dalam sistem yang dapat melakukan login. Proses ini memastikan bahwa hanya pengguna yang memiliki izin yang dapat mengakses data dan fitur dalam sistem. Halaman login dapat diakses oleh pengguna sebelum memasuki sistem dengan memasukkan username dan kata sandi yang telah terdaftar. Proses ini mencakup verifikasi kredensial pengguna untuk memastikan keabsahan dan otoritas akses yang dimiliki. Hasil implementasi halaman login pada sistem ini dapat dilihat pada Gambar 8.



Gambar 8. Tampilan Halaman Login

3.2. Halaman *Dashboard*

Halaman dashboard merupakan halaman utama atau halaman awal yang akan ditemui oleh pengguna setelah berhasil melakukan login. Halaman ini berfungsi untuk memberikan gambaran umum tentang berbagai fitur dan fungsi yang dapat diakses oleh pengguna dalam sistem. Hasil implementasi halaman dashboard pada sistem ini dapat dilihat pada Gambar 9.



Gambar 9. Tampilan Halaman Dashboard

3.3. Halaman *Data Training*

Halaman Data Training merupakan bagian dari sistem untuk menampilkan dan mengelola data yang digunakan dalam proses pelatihan. Pada halaman ini, pengguna dapat melihat, upload, dan menghapus data yang akan digunakan untuk melatih model. Implementasi halaman Data Training pada sistem ini dapat dilihat pada Gambar 10.

No #	Nama Mahasiswa	Pekerjaan Ayah	Penghasilan Ayah	Status Ayah	Pekerjaan Ibu	Penghasilan Ibu
1	T. KHAYANI	Petani	Rp. 500.001 - Rp. 750.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
2	Fendi Fyzzan	Lainnya	Rp. 2.250.001 - Rp. 2.500.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
3	REZA PRIBADI	TN / POLRI	Rp. 4.000.001 - Rp. 4.250.000	Hidup	Petani	Tidak Berpenghasilan
4	HIRI ANEKA MALI ANA	TIDAK BEKERJA	Tidak Berpenghasilan	Wafat	Lainnya	Rp. 750.001 - Rp. 1.000.000
5	ALI HA KHASAN	Lainnya	Rp. 750.001 - Rp. 1.000.000	Wafat	Lainnya	Rp. 500.001 - Rp. 750.000
6	MURAHMAD ALRAN	Lainnya	Rp. 750.001 - Rp. 1.000.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan

Gambar 10. Tampilan Halaman Data Training

3.4. Halaman Klasifikasi

Halaman Klasifikasi merupakan bagian dari sistem yang menyediakan antarmuka untuk melakukan proses klasifikasi atau prediksi berdasarkan model yang telah dilatih sebelumnya. Pengguna dapat memasukkan data yang ingin diklasifikasikan dan sistem akan memberikan prediksi berdasarkan model yang telah disiapkan. Implementasi halaman klasifikasi pada sistem ini dapat dilihat pada Gambar 11.

No #	Nama Mahasiswa	Pekerjaan Ayah	Penghasilan Ayah	Status Ayah	Pekerjaan Ibu	Penghasilan Ibu
1	Adira Elysa	Lainnya	Rp. 3.500.001 - Rp. 3.750.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
2	Rayhan Ardianto	Lainnya	Rp. 4.750.001 - Rp. 5.000.000	Hidup	Lainnya	Rp. 250.001 - Rp. 500.000
3	Zahira Farhana	Peg. Swasta	Rp. 3.750.001 - Rp. 4.000.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
4	Fitri Ramadhani	Peg. Swasta	Rp. 3.750.001 - Rp. 4.000.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
5	Ayra Anindita	PHS	Rp. 4.750.001 - Rp. 5.000.000	Hidup	TIDAK BEKERJA	Tidak Berpenghasilan
6	Rizki Pratama	Lainnya	Rp. 4.000.001 - Rp. 4.250.000	Hidup	Lainnya	Rp. 2.000.001 - Rp. 2.250.000

Gambar 11. Halaman Klasifikasi

3.5. Halaman Hasil Klasifikasi

Halaman Data Klasifikasi merupakan bagian dari sistem yang menampilkan hasil dari proses klasifikasi data yang telah dimasukkan oleh pengguna. Pada halaman ini, pengguna dapat melihat prediksi atau klasifikasi yang dihasilkan oleh model berdasarkan data yang telah dimasukkan sebelumnya. Informasi yang disajikan meliputi label atau kategori yang diprediksi untuk setiap data, yaitu kelompok Uang Kuliah Tunggal (UKT) dan besaran uang kuliah yang harus dibayar. Besaran uang kuliah tunggal ditentukan berdasarkan kelompok UKT dan program studi mahasiswa. Sebagai contoh, jika seorang mahasiswa bernama Al Khaidar ditentukan masuk ke dalam kelompok UKT 5 dengan program studi Teknik Informatika, maka besaran UKT yang harus dibayar adalah sebesar Rp 4.250.000. Implementasi dari halaman data klasifikasi pada sistem ini dapat dilihat pada Gambar 12 berikut:

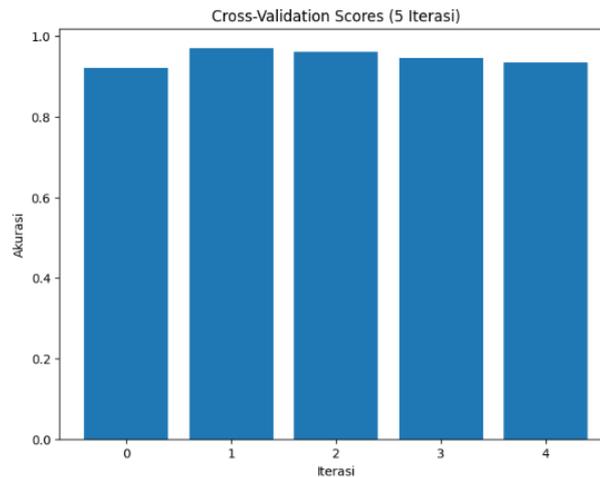
No #	Nama Mahasiswa	Jurusan	Prodi	Nama UKT	Jumlah UKT
1	Adira Elysa	Teknik Sipil	Teknologi Konstruksi Bangunan Air	Kelompok 8	Rp. 5.000.000,00
2	Rayhan Ardianto	Teknik Sipil	Teknologi Konstruksi Jalan dan Jembatan	Kelompok 8	Rp. 6.000.000,00
3	Zahira Farhana	Teknik Sipil	Teknologi Konstruksi Bangunan Gedung	Kelompok 8	Rp. 6.000.000,00
4	Fitri Ramadhani	Teknik Kimia	Teknologi Konstruksi Jalan dan Jembatan	Kelompok 8	Rp. 6.000.000,00
5	Ayra Anindita	Teknik Kimia	Teknologi Pengolahan Minyak dan Gas	Kelompok 8	Rp. 6.000.000,00
6	Rizki Pratama	Teknik Kimia	Teknologi Rekayasa Kimia Industri	Kelompok 8	Rp. 7.000.000,00
7	Saibabti Putri	Teknik Kimia	Teknologi Kimia	Kelompok 8	Rp. 6.000.000,00
8	Daffa Nugroho	Teknik Mesin	Teknologi Mesin	Kelompok 8	Rp. 7.000.000,00
9	Zahra Amalia	Teknik Mesin	Teknologi Industri	Kelompok 8	Rp. 6.000.000,00
10	Silwa Aditya	Teknik Mesin	Teknologi Rekayasa Manufaktur	Kelompok 8	Rp. 7.000.000,00

Gambar 12. Halaman Hasil Klasifikasi

3.6. Cross Validation dan Confusion Matrix

Gambar 13 menunjukkan hasil pengujian kinerja metode *Random Forest* menggunakan *cross validation*. Terdapat 5 iterasi pengujian yang sebelumnya telah dijelaskan yang mana setiap iterasi menghasilkan akurasi masing-masing berdasarkan baris data yang diuji, secara keseluruhan diperoleh akurasi rata-rata sebesar 95%. Sedangkan pada Gambar 14 merupakan hasil akurasi pembelajaran model menggunakan metode *Random Forest* dan hasil evaluasi model menggunakan metode *confusion matrix*.

Akurasi model: 94.42%
 Cross Validation Scores: [0.96015936 0.96812749 0.95219124 0.93625498 0.94820717]
 Rata-rata Cross Validation Score: 0.952988047808765



Gambar 13. Cross Validation

Gambar 14 menunjukkan hasil evaluasi metode random forest menggunakan confusion matrix memperoleh akurasi sebesar 94% dengan nilai precision: kelompok satu 96%, kelompok dua 91%, kelompok tiga 98%, kelompok empat 94%, kelompok lima 93%, kelompok enam 97%, kelompok tujuh 89% dan kelompok delapan 100%. Nilai recall : kelompok satu 85%, kelompok dua 83%, kelompok tiga 94%, kelompok empat 100%, kelompok lima 100%, kelompok enam 95%, kelompok tujuh 95% dan kelompok delapan 90%. Nilai f1-score : kelompok satu 90%, kelompok dua 87%, kelompok tiga 96%, kelompok empat 97%, kelompok lima 96%, kelompok enam 96%, kelompok tujuh 92% dan kelompok delapan 95%. Jadi kesimpulannya metode Random Forest menunjukkan kinerja yang sangat baik.

Classification Report:				
	precision	recall	f1-score	support
0	0.96	0.85	0.90	26
1	0.91	0.83	0.87	12
2	0.98	0.94	0.96	53
3	0.94	1.00	0.97	17
4	0.93	1.00	0.96	52
5	0.97	0.95	0.96	40
6	0.89	0.95	0.92	41
7	1.00	0.90	0.95	10
accuracy			0.94	251
macro avg	0.95	0.93	0.94	251
weighted avg	0.95	0.94	0.94	251

Gambar 14. Evaluasi Model Menggunakan Confusion Matrix

3.7. Hold-Out Cross Validation

Hold-Out Cross Validation adalah teknik evaluasi umum dalam machine learning untuk memastikan model mampu generalisasi dengan baik terhadap data baru. Dalam skenario ini dilakukan eksplorasi menggunakan 9 model untuk mencari proporsi optimal antara data training dan testing. Tujuan utamanya adalah menggabungkan akurasi model, efisiensi penggunaan data, dan sensitivitas terhadap perubahan dalam komposisi data saat training dan testing, serta memahami interaksi ini terhadap performa keseluruhan model. Tabel 1 hasil akurasi dari 9 model dalam tabel tersebut menunjukkan bahwa proporsi antara data training dan testing mempengaruhi performa model secara signifikan. Model-model dengan proporsi data training yang tinggi, seperti Model 1 (90% Train, 10% Test) dan Model 2 (80% Train, 20% Test), cenderung menghasilkan akurasi yang tinggi (98% dan 94% secara berturut-turut), karena memiliki akses yang lebih besar terhadap data pembelajaran yang relevan. Namun, ketika proporsi data training semakin rendah, seperti pada Model 8 (20% Train, 80% Test) dan Model 9 (10% Train, 90% Test), akurasi model cenderung menurun (90% untuk kedua model tersebut), karena model memiliki akses terbatas untuk mempelajari pola yang penting dari data *training*. Proporsi yang seimbang antara data training dan testing, seperti pada Model 5 (50% Train, 50% Test) yang memiliki akurasi 94%, menunjukkan bahwa model dapat menggeneralisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya.

Tabel 1. Hold-out cross validation

Skenario	Train	Test	Accuracy Model
1	90%	10%	98%
2	80%	20%	94%
3	70%	30%	94%
4	60%	40%	93%
5	50%	50%	94%
6	40%	60%	93%
7	30%	70%	93%
8	20%	80%	90%
9	10%	90%	90%

4. KESIMPULAN

Adapun kesimpulan setelah melalui tahapan perancangan dan pengujian klasifikasi UKT sebagai berikut. Penerapan Metode *Random Forest* pada sistem diawali dengan pelabelan data dilakukan dengan metode K-Means ke dalam delapan kluster, diikuti *preprocessing* dan pembobotan data menggunakan metode *mapping*. Data kemudian dibagi menjadi data *training* dan data *testing*. Model *Random Forest* dilatih dengan *RandomForestClassifier* menggunakan 100 *estimator*. Hasil pengujian *black box* pada sistem memiliki persentase kelulusan sebesar 94,7% dan persentase ketidakkelulusan sebesar 5,3%. Hasil akurasi penerapan metode *random forest* menggunakan confusion matrix diperoleh akurasi sebesar 94%. Jadi kesimpulannya metode *Random Forest* menunjukkan kinerja yang sangat baik.

REFERENSI

- [1] T. P. Hamakonda and J. N. B. Tairas, Pengantar Klasifikasi Persepuluhan Dewey / oleh Towa P. Jakarta: Gunung Mulia, 2006.
- [2] Sulisty-Basuki, Pengantar ilmu perpustakaan / Sulisty Basuki. Jakarta: Gramedia Pustaka Utama, 1993.
- [3] P. G. Cole and L. Chan, Teaching principles and practice, 2nd ed. New York: Prentice Hall, 1994.
- [4] B. Karim and S. Sentinuwo, "Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan Data Mining," J. Tek. Inform., vol. 11, no. 1, 2017.
- [5] M. Ardiansyah, T. Suharto, and A. S. Farid, "Upaya Penanganan Uang Kuliah Tunggal (UKT) Bermasalah bagi Mahasiswa yang tidak Mampu pada Perguruan Tinggi," JIIP-Jurnal Ilm. Ilmu Pendidik., vol. 5, no. 10, pp. 4432–4441, 2022.
- [6] Kemenristek-Dikti, "Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 37 Tahun 2018 Tentang Statuta Politeknik Negeri Lhokseumawe." 2018.
- [7] Kemendikbud RI, "Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 25 Tahun 2020," Kementeri. Pendidik. dan Kebud. RI, pp. 1–76, 2020.
- [8] Kepmen, Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri Di Lingkungan Kementerian Riset, Teknologi. Dan Pendidikan Tinggi Tahun Angkatan, 2018.
- [9] F. A. Kurniawan, A. Kurniati, and dkk, "Analisis dan implementasi random forest dan classification dan regression tree (cart) untuk klasifikasi pada misuse intrusion detection system," in IT Telkom, Program Studi Teknik Informatika, Skripsi, Bandung: IT Telkom, 2011.
- [10] T. K. Ho, "Random decision forests," in Dalam Proceedings of 3rd international conference on document analysis and recognition, 1995, pp. 278–282.
- [11] L. Breiman and A. Cutler, "Manual–Setting Up, Using, and Understanding Random Forest V4.0." 2001.
- [12] K. A. Sambodo, M. I. Rahayu, N. Indriasari, and M. Natsir, "Klasifikasi Hutan-Non Hutan Data Alos Palsar Menggunakan Metode Random Forest," in Prosiding Seminar Nasional Penginderaan Jauh 2014, LAPAN, 2014, pp. 120–127.
- [13] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5–32, 2003.