Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan Vol. 7 No. 1. Juni 2022 | P-ISSN: 2527-9858
E-ISSN: 2548-1180

# Perbandingan Metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT)

Ronny Susetyoko[1], Wiratmoko Yuwono[2], Elly Purwantini[3], Nana Ramadijanti[4]

*[1]Program Studi Sains Data Terapan, [2,4]Program Studi Teknik Informatika, [3]Program Studi Teknik Elektronika*
*Politeknik Elektronika Negeri Surabaya*

*[1]rony@pens.ac.id, [2]moko@pens.ac.id, [3]elly@pens.ac.id, [4]nana@pens.ac.id*

*Abstrak*—Uang Kuliah Tunggal (UKT) adalah biaya yang dikenakan kepada setiap mahasiswa untuk digunakan dalam proses pembelajaran untuk program diploma dan program sarjana dari setiap jalur penerimaan yang ditetapkan oleh pemimpin perguruan tinggi negeri (PTN). Penetapan UKT masing-masing mahasiswa baru mengikuti kebijakan masing-masing PTN, tergantung ketersediaan informasi maupun target finansial berupa pendapatan negara bukan pajak (PNBP) yang tetapkan. Rumusan atau algoritma klasifikasi UKT yang digunakan tentunya akan berdampak pada distribusi dan ekspektasi rerata UKT. Tujuan dari penelitian ini adalah membandingkan kinerja beberapa metode yaitu *Random Forest*, Regresi Logistik, Naïve Bayes, dan *Multilayer Perceptron* dalam mengklasifikasi UKT. Beberapa atribut atau fitur yang digunakan dalam model adalah status rumah, penghasilan, jumlah rumah, jumlah motor, jumlah mobil, daya listrik, kepemilikan tanah, dan jumlah anak. Dataset sebanyak 873 record dibagi menjadi data *training* dan data *testing* masing-masing sebanyak 80% dan 20%. Untuk mendapatkan metode yang terbaik, dilakukan evaluasi kinerja empat metode tersebut didasarkan pada rerata akurasi, karakteristik fungsi tingkat akurasi terhadap jumlah fitur, dan nilai ekspektasi UKT. Hasil dari penelitian ini, metode *Random Forest*, Regresi Linier, dan *Multilayer Perceptron* dapat digunakan sebagai model klasifikasi UKT karena memiliki rerata akurasi lebih dari 85%. Namun dari ketiga model tersebut, *Random Forest* dapat dipilih sebagai model klasifikasi terbaik dengan rerata akurasi 97,9%. Berdasarkan karakteristik fungsi tingkat akurasi, penggunaan metode *Random Forest* tidak harus melibatkan banyak fitur dalam model. Dengan menerapkan metode tersebut, ekspektasi rerata UKT sebesar Rp. 3,833,811 dan simpangan baku Rp. 2,123,758.

*Kata kunci*—*Random Forest*, Regresi Logistik, Naive Bayes, *Multilayer Perceptron*, rerata akurasi.

*Abstract*— Single Tuition Fee (UKT) is a fee charged to each student to be used in the learning process for diploma programs and undergraduate programs from each admission path determined by the leader of state universities (PTN). The determination of UKT for each new student follows the policies of each PTN, depending on the availability of information and financial targets in the form of non-tax state income (PNBP) that are set. The UKT classification formula or algorithm used will of course have an impact on the distribution and expectations of the UKT average. The purpose of this study was to compare the performance of several methods, namely Random Forest, Logistic Regression, Naïve Bayes, and Multilayer Perceptron in classifying UKT. Some of the attributes or features used in the model are house status, income, number of houses, number of motorbikes, number of cars, electric power, land ownership, and number of children. The dataset of 873 records is divided into training data and testing data of 80% and 20%, respectively. To get the best method, the performance evaluation of the four methods was carried out based on the average accuracy, the characteristics of the function of the level of accuracy against the number of features, and the expected value of UKT. The results of this study, the method of Random Forest, Linear Regression, and Multilayer Perceptron can be used as a UKT classification model because it has an average accuracy of more than 85%. However, from the three models, Random Forest can be chosen as the best classification model with an average accuracy of 97.9%. Based on the characteristics of the function of the level of accuracy, the use of the Random Forest method does not have to involve many features in the model. By applying this method, the expected average UKT is Rp. 3,833,811 and standard deviation of Rp. 2,123,758.

*Keywords*— *Random Forest*, Logistic Regression, Naive Bayes, *Multilayer Perceptron*, average accuracy.

## I. PENDAHULUAN

Uang Kuliah Tunggal yang selanjutnya disingkat UKT adalah biaya yang dikenakan kepada setiap mahasiswa untuk digunakan dalam proses pembelajaran. Berdasarkan Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 25 Tahun 2020 Tentang Standar Satuan Biaya Operasional Pendidikan Tinggi Pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan Dan Kebudayaan, besaran UKT ditetapkan oleh pemimpin PTN bagi Mahasiswa program diploma dan program sarjana dari setiap jalur penerimaan Mahasiswa. Besaran UKT bagi Mahasiswa program diploma dan program sarjana sebagaimana dimaksud terbagi dalam beberapa kelompok, terdiri atas paling sedikit 2 (dua) kelompok: a) kelompok I dengan besaran UKT paling tinggi Rp500.000,00 (lima ratus ribu rupiah); dan b) kelompok II dengan besaran UKT paling rendah Rp501.000,00 (lima ratus satu ribu rupiah) dan paling tinggi Rp1.000.000,00 (satu juta rupiah). Penetapan besaran UKT untuk setiap kelompok berlaku sama bagi mahasiswa pada setiap jalur penerimaan. Penetapan kelompok besaran UKT dilakukan dengan mempertimbangkan kemampuan ekonomi: a) mahasiswa; b) orang tua mahasiswa; atau c) pihak lain yang membiayai mahasiswa. Penetapan kemampuan ekonomi dilakukan berdasarkan pendapatan dan jumlah tanggungan keluarga dari mahasiswa, orang tua mahasiswa, atau pihak lain yang membiayai mahasiswa [1].

Penetapan UKT masing-masing mahasiswa baru mengikuti kebijakan masing-masing PTN, tergantung ketersediaan informasi maupun target finansial berupa pendapatan negara bukan pajak (PNBP) yang tetapkan. Rumusan atau algoritma

klasifikasi UKT yang digunakan tentunya akan berdampak pada distribusi UKT dan ekspektasi rerata UKT.

Menurut Breiman dalam [2], *Random Forest* adalah salah satu jenis algoritma klasifikasi yang terdiri dari lebih satu pohon keputusan yang setiap pohon keputusan dibentuk bergantung pada nilai-nilai vektor acak sampel secara independen dan identik didistribusikan yang sama untuk semua pohon.

*Random Forest* adalah metode *machine learning* yang tidak sensitif terhadap multikolinearitas, Metode ini digunakan untuk mengklasifikasi fase tanaman padi. Dalam penelitian tersebut memberikan akurasi 0,236 jika kita menggunakan satu fitur temporal indeks vegetasi. Jika menggunakan lebih banyak fitur temporal, akurasi meningkat menjadi 0,7091. Eksistensi autokorelasi temporal antar fitur harus dipertimbangkan dalam model untuk meningkatkan akurasi klasifikasi [3].

Metode *Random Forest* juga untuk membantu perbankan mengambil keputusan atau suatu kebijakan, diantaranya adalah untuk memprediksi kelayakan kredit pinjaman secara dini untuk mengetahui nasabah yang layak atau tidak layak. Tahap pelatihan menggunakan 80% data dan pengujian menggunakan 20% data secara acak dari 1000 data. Hasil performa dari algoritma *Random Forest* tersebut yaitu memiliki tingkat akurasi sebesar 0,83 atau 83% sehingga termasuk pada kategori klasifikasi modelnya sangat bagus [4].

Penggunaan *Random Forest* juga untuk mendeteksi diabetes. Berdasarkan hasil penelitian, akurasi terbaik adalah model 1 (Min-max normalization-RF) sebesar 95,45%, disusul model 2 (Z-score normalization-RF) sebesar 95%, dan model 3 (tanpa normalisasi data-RF) dari 92%. Dari hasil tersebut dapat disimpulkan bahwa model 1 (normalisasi min-max-RF) lebih baik dari kedua model normalisasi data lainnya dan mampu meningkatkan kinerja klasifikasi *Random Forest* sebesar 95,45% [2].

*Multi Class Decision Forest Machine Learning* untuk membangun model yang dapat memprediksi dan mengevaluasi kebangkrutan industri perbankan. Pemodelan *machine learning* menggunakan lima variabel input. Secara keseluruhan, *Multi Class Decision Forest Machine Learning* mampu melatih data hubungan input-output dan perilaku pemodelan dengan baik, nilai akurasi 92%, nilai *precision* 92% dan nilai *under area curve* 90% [5].

*Decission Tree*, *Random Forest* dan *Support Vector Machine* (SVM) juga digunakan untuk mengklasifikasi kualitas red wine berdasarkan kandungan masing-masing jenis wine. Hasil penelitian tersebut, *Random Forest* mempunyai akurasi paling tinggi, yaitu 0.7468. Sedangkan *Decission Tree* mempunyai akurasi 0,7031, dan *Support Vector Machine* (SVM) mempunyai akurasi 0,65 [6].

*Classification and Regression Tree* (CART) merupakan salah satu metode klasifikasi yang populer digunakan di berbagai bidang yang mampu menghadapi berbagai kondisi data. Namun memiliki kelemahan pada prediksi pohon klasifikasi yaitu kurang stabil pada perubahan data pembelajaran yang akan menyebabkan perubahan besar pada hasil prediksi pohon klasifikasi. Untuk meningkatkan kestabilan dan akurasi prediktif digunakan CART dengan *Random Forest* untuk mengklasifikasi ketidaktepatan waktu kelulusan mahasiswa Universitas Terbuka. *Random Forest* mampu meningkatkan akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa dengan akurasi 93.23% [7].

Naïve Bayes Classifier merupakan pendekatan yang mengadopsi teorema Bayes, dengan menggabungkan pengetahuan sebelumnya dengan pengetahuan baru. Kelebihan dari metode ini adalah algoritma yang sederhana dan akurasi yang tinggi. Naïve Bayes Classifier digunakan untuk mengklasifikasikan kualitas jurnal yang biasa disebut dengan Quartile. Dengan menggunakan dataset sebanyak 1491, nilai akurasi sebesar 71,60% dan tingkat kesalahan sebesar 28,40% [8].

Metode *K-Nearest Neighbor* (KNN), *Decision Tree*, *Linear Discriminant Analysis* (LDA), *Logistic Regression*, *Support Vector Machine* (SVM), dan *Naïve Bayes* digunakan untuk mengklasifikasi citra CT Scan paru-paru. *Random Forest* menghasilkan akurasi sebesar 96,9%, diikuti oleh KNN sebesar 96,5%, Decision Tree sebesar 95,5%, dan yang paling rendah yaitu Naive Bayes sebesar 42,4% [9].

Penelitian ini mencoba membandingkan *Random Forest*, Regresi Logistik, Naïve Bayes, dan *Multilayer Perceptron* untuk klasifikasi UKT. Dataset yang digunakan adalah data Seleksi Bersama Mahasiswa Politeknik Negeri (SBMPN) pada Politeknik Elektronika Negeri Surabaya (PENS). Pada tahun-tahun sebelumnya, penetapan UKT dengan mempertimbangkan beberapa fitur seperti penghasilan, daya listrik, dan jumlah anak. Selanjutnya dilakukan perhitungan skor dengan menjumlahkan masing-masing nilai fitur dengan bobot yang ditetapkan oleh direksi sebagai pengambil keputusan. Penetapan fitur dan bobot dari tahun ke tahun tidak konsisten, tergantung pada kesepakatan. Dalam kurun waktu 4 tahun (2018 – 2021), ada 0,9% - 4,8% dari total mahasiswa mengajukan permohonan keringanan UKT dengan alasan keberatan dengan klas UKT yang ditetapkan karena permasalahan ekonomi. Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengklasifikasikan UKT menggunakan beberapa metode yaitu *Random Forest*, Regresi Logistik, Naïve Bayes, dan *Multilayer Perceptron*. Empat metode tersebut dievaluasi berdasarkan rerata akurasi, karakteristik fungsi akurasi terhadap jumlah fitur, dan nilai ekspektasi UKT.

## II. METODOLOGI PENELITIAN

Tahapan dalam penelitian ini dimulai dengan studi literatur, identifikasi dan pengumpulan data, *data preprocessing*, membagi data *training* dan data *testing*, pemodelan menggunakan metode *Random Forest*, Regresi Logistik, Naïve Bayes, dan *Multilayer Perceptron*. Selanjutnya dilakukan evaluasi model dan seleksi model klasifikasi. Tahapan tersebut dapat dilihat pada Gambar 1.

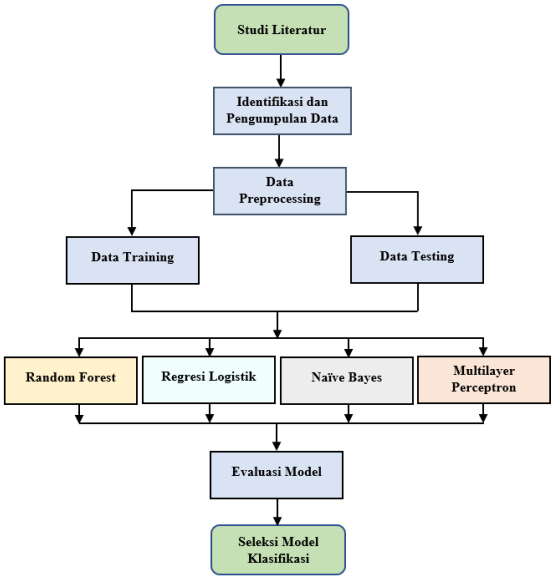
### *A. Sumber Data*

Penelitian ini menggunakan data pendaftar Seleksi Bersama Masuk Politeknik Negeri (SBMPN) yang diterima mulai tahun 2019 sampai dengan tahun 2021 pada Politeknik Elektronika Negeri Surabaya (PENS).

### *B. Data Preprocessing*

Dataset asli sebanyak 873 record dengan 31 atribut. Sebelum digunakan untuk pemodelan dilakukan pembersihan data dan dipilih sebanyak 8 atribut sebagai fitur atau variabel prediktor. Selanjutnya dilakukan kodifikasi fitur yang digunakan dan standarisasi variabel penghasilan. Dataset yang

telah siap dimodelkan dibagi menjadi dua, yaitu data *training* dan data *testing* dengan perbandingan 80%:20%.




Gambar 1. Metodologi Penelitian

## C. *Variabel Penelitian*

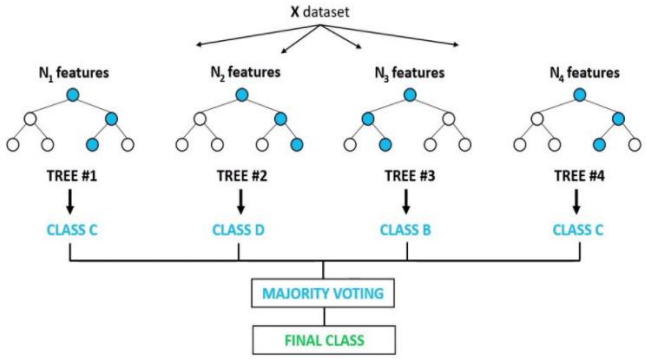
Beberapa fitur yang digunakan variabel model klasifikasi UKT adalah penghasilan, status rumah, jumlah rumah, jumlah motor, jumlah mobil, daya listrik, kepemilikan tanah dapat dilihat pada Tabel I.

TABEL I
VARIABEL PENELITIAN

| Variabel | Tipe Variabel [Skala Pengukuran] | Keterangan |
|---|---|---|
| Klas UKT | Integer [ordinal] | 1 : Rp. 500.000<br>2 : Rp. 1.000.000<br>3 : Rp. 3.000.000<br>4 : Rp. 4.500.000<br>5 : Rp. 5.500.000<br>6 : Rp. 6.500.000<br>7 : Rp. 7.500.000 |
| Fitur/Variabel Prediktor | | |
| Status Rumah | integer [kategorik] | 1 : milik sendiri<br>2 : kontrak/sewa<br>3 : menumpang |
| Penghasilan | float | penghasilan orangtua (Rp.) |
| Rumah | integer | jumlah rumah |
| Motor | integer | jumlah motor |
| Mobil | integer | jumlah mobil |
| Listrik | integer | daya listrik rumah<br>1 : 450 Watt<br>2 : 900 Watt<br>3 : 1300 Watt atau lebih |
| Tanah | boolean | Kepemilikan tanah<br>0 : Tidak<br>1 : Ya |
| Anak | integer | Jumlah anak |

## D. *Metode Klasifikasi*

*1) Random Forest:* Random Forest adalah salah satu dari sekian banyak metode *ensemble* yang dikembangkan oleh Leo Breiman pada tahun 2001. *Random Forest* merupakan pengembangan metode *Classification and Regression Tree* (CART) dengan menerapkan metode *Boostrap Aggregating* (*Bagging*), dan *Random Feature Selection*. Metode *Random Forest* memiliki beberapa kelebihan antara lain, menghasilkan hasil klasifikasi yang baik, menghasilkan *error* yang lebih rendah, secara efisien dapat mengatasi data training dengan jumlah data yang sangat besar. Metode Random Forest menghasilkan satu set pohon acak [7]. Kelas yang dihasilkan berasal dari proses klasifikasi yang dipilih dari kelas yang paling banyak (modus) yang dihasilkan oleh pohon keputusan yang ada. Gambar 2 adalah skema kinerja algoritma *Random Forest*.




Gambar 2. Skema Kinerja Algoritma Random Forest
Sumber : (medium.com)

*2) Regresi Logistik:* Regresi logistik adalah metode yang dapat digunakan dalam mencari hubungan variabel respon yang bersifat dikotomus (bersakala ordinal atau nominal) atau polikotomus (memiliki skala nominal atau ordinal dengan lebih dari 2 kategori). Ketika variabel dependen memiliki skala yang bersifat polikotomus atau multinomial maka dapat digunakan regresi logistik multinomial [10]. Analisis regresi logistik ordinal merupakan salah satu metode statistik yang digunakan untuk menganalisa hubungan antara variabel respon dan variabel prediktor. Variabel respon pada regresi logistik ordinal bersifat polikotomus dengan skala ordinal. Model yang digunakan untuk regresi logistik ordinal adalah model logit [11]. Model tersebut adalah model logit kumulatif, pada model ini terdapat sifat ordinal dari respon *Y* yang dituangkan dalam peluang kumulatif. Model logit kumulatif merupakan model yang didapatkan dengan cara membandingkan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori respon ke-*j* pada *p* variabel prediktor yang dinyatakan dalam vektor **x** dengan peluang lebih besar daripada ketegori respon ke-*j*, $P(Y > j|\mathbf{x})$ [10].

*3) Naïve Bayes: Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu klas. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decesion tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasai dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [12].

*4) Multilayer Perceptron: Multilayer perceptron* (MLP) merupakan algoritma yang mengadopsi cara kerja jaringan

saraf pada makhluk hidup. MLP merupakan topologi yang paling umum dalam jaringan syaraf tiruan, di mana masing-masing *perceptron* terhubung membentuk beberapa lapisan/layer. Sebuah MLP memiliki lapisan masukan (input layer), minimal 1 lapisan *hidden layer* dan lapisan output. MLP menggunakan pembelajaran propagasi balik (*backpropagation*) yang harus dilakukan dalam metode ini yaitu inisialisasi (*initialization*), aktivasi (*activation*), pelatihan bobot (*weight training*), dan iterasi (*iteration*). Pada langkah inisialisasi, nilai awal bobot dan ambang batas (*threshold*) ditentukan secara acak namun dalam batasan tertentu. Pada tahapan aktivasi, diberikan masukan dan nilai keluaran yang diharapkan (*desired output*). Proses penyesuaian bobot terjadi pada tahap pelatihan bobot, nilai luaran sebenarnya (*actual output*) dibandingkan dengan *desired output* dan dilakukan penyesuaian bobot. Langkah kedua dan ketiga diulangi sampai dengan tercapai kondisi yang ditentukan [13].

### *E. Evaluasi Model*

Empat metode yang digunakan dibandingkan kinerjanya berdasarkan rerata akurasi, karakteristik fungsi tingkat akurasi berdasarkan jumlah fitur yang digunakan, serta ekspektasi UKT dari model terbaik.

*1) Rerata Akurasi:* Rerata akurasi rata-rata tingkat akurasi yang diperoleh dengan tiga kali pemodelan dengan inisialisasi fungsi random yang berbeda

*2) Fungsi Tingkat Akurasi:* Fungsi tingkat akurasi adalah pemodelan regresi polynomial tingkat akurasi terhadap jumlah fitur.

*3) Ekspektasi UKT:* Ekspektasi UKT adalah nilai ekspektasi rata-rata dan ekspektasi simpangan baku besaran UKT setelah dilakukan klasifikasi.UKT.

### *F. Seleksi Model Klasifikasi*

Model klasifikasi yang dipilih adalah model yang memiliki rerata akurasi paling tinggi dan mempunyai karakteristik fungsi dengan jumlah fitur yang lebih sedikit.

## III. HASIL DAN PEMBAHASAN

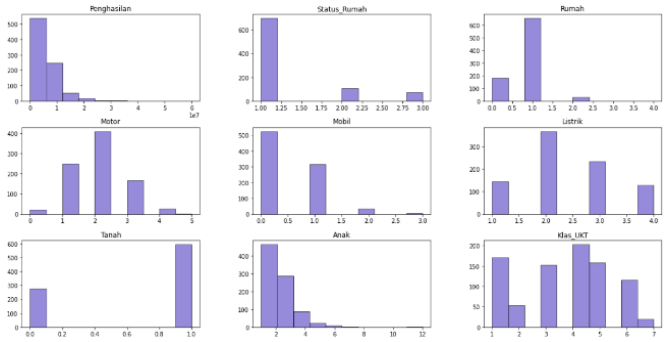
### *A. Eksplorasi Data*

Tujuan dari eksplorasi data adalah melakukan analisis awal sebelum melakukan pemodelan klasifikasi. Dari 8 fitur yang digunakan. Ada 5 fitur non kategorikal, yaitu penghasilan, jumlah rumah, jumlah motor, jumlah mobil, dan jumlah anak. Berdasarkan informasi dari Tabel II, rerata penghasilan orang tua sebesar Rp. 6,16 juta dengan simpangan baku sebesar Rp. 6,18 juta. Fitur penghasilan cenderung berdistribusi Eksponensial (kemiringan positif). Status rumah sebagian besar milik sendiri (79,4%), dengan jumlah rumah sebagian besar masih satu unit (74,8%). Jumlah kepemilikan sepeda motor antara 0 – 5 unit, namun paling banyak adalah 2 unit.

TABEL II
STATISTIK DESRIPTIF VARIABEL PREDIKTOR

| | Penghasilan | Rumah | Motor | Mobil | Listrik |
|---|---|---|---|---|---|
| count | 8.730000e+02 | 873.000000 | 873.000000 | 873.000000 | 873.000000 |
| mean | 6.164197e+06 | 0.838488 | 1.920962 | 0.447881 | 2.399771 |
| std | 6.183685e+06 | 0.504824 | 0.840180 | 0.592265 | 0.929560 |
| min | 0.000000e+00 | 0.000000 | 0.000000 | 0.000000 | 1.000000 |
| 25% | 2.500000e+06 | 1.000000 | 1.000000 | 0.000000 | 2.000000 |
| 50% | 4.500000e+06 | 1.000000 | 2.000000 | 0.000000 | 2.000000 |
| 75% | 8.000000e+06 | 1.000000 | 2.000000 | 1.000000 | 3.000000 |
| max | 6.000000e+07 | 4.000000 | 5.000000 | 3.000000 | 4.000000 |

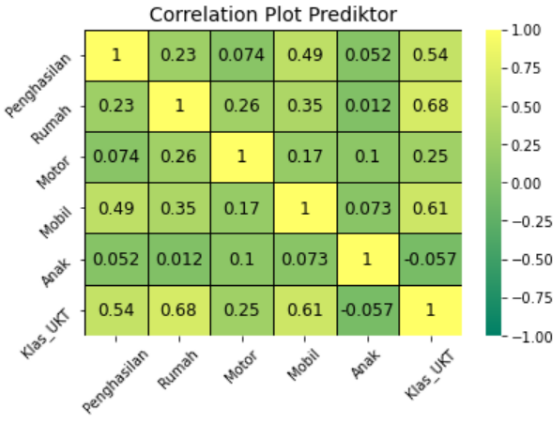
Kepemilikan mobil antara 0 – 3 unit, dengan rincian sebanyak 59,9% tidak mempunyai mobil dan sebanyak 35,9% mempunyai mobil satu unit. Daya listrik paling banyak pada daya 900 watt (42%), kemudian 1300 watt (26,8%). Detail distribusi masing-masing fitur dapat dilihat histogram pada Gambar 3.




Gambar 3. Histogram Variabel Prediktor

Matriks korelasi 5 fitur non kategorik terhadap klas UKT dapat dilihat heatmap pada Gambar 4. Dari gambar tersebut, fitur jumlah rumah mempunyai koefisien korelasi terbesar, yaitu 0,68, disusul fitur jumlah mobil dengan korelasi sebesar 0,61, dan fitur penghasilan sebesr 0,54. Berikutnya jumlah motor dan jumlah anak, masing-masing mempunyai korelasi sebesar 0,25 dan -0,057.

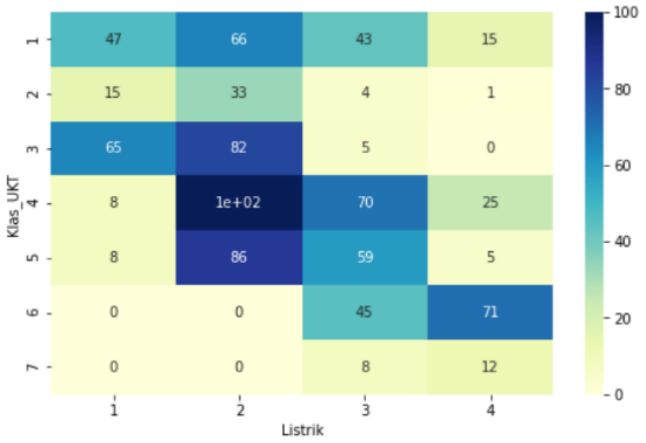
Untuk beberapa fitur lainnya dengan tipe data kategorik, dibuat tabel kontingensi. Sebagian besar status rumah milik sendiri. Sedangkan mereka dengan status rumah kontrak/sewa atau menumpang, cenderung berada pada klas UKT-1 (Rp. 500.000), yaitu sebanyak 100 mahasiswa dengan status kontrak/sewa dan 70 mahasiswa dengan status menumpang.




Gambar 4. Matriks Korelasi Variabel Prediktor

Ada beberapa dari mereka yang kontrak atau sewa atau menumpang berada pada klas UKT-5, UKT-6, dan UKT-7, kemungkinan orang tuanya mutasi kerja, sering berpindah-pindah atau memang berpenghasilan tinggi namun belum mempunyai rumah sendiri. Hasil uji asosiasi didapatkan nilai Chi-Square sebesar 822.09 dan p-value 3.03e-168, sehingga disimpulkan bahwa dengan ada hubungan yang signifikan status rumah terhadap klas UKT.

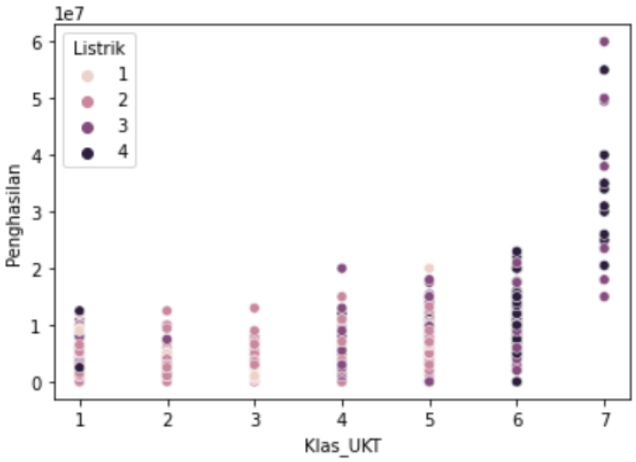
*Heatmap* kontingensi daya listrik dan kelas UKT dapat dilihat pada Gambar 5.




Gambar 5. Heatmap Kontingensi Daya Listrik dan UKT

Daya listrik 900 watt didominasi pada klas UKT-1 sampai dengan UKT-4. Sedangkan daya listrik 1300 watt atau di atas 1300 watt Sebagian besar pada klas UKT-4 sampai dengan UKT-7. Nilai Chi-Square = 514,25 dan p-value = 1,05e-97 sehingga dapat disimpilkan ada hubungan yang signifikan antara daya listrik dan klas UKT. Mereka dengan klas UKT-6 atau UKT-7 sebagian besar menggunakan daya listrik 1300 watt dan lebih dari 1300 watt.

Demikian juga dengan kepemilikan tanah, nilai Chi-Square = 75,05 dan p-value = 3,75e-14. Sehingga disimpulkan bahwa ada hubungan yang signifikan antara kepemilikan tanah dengan klas UKT. Nilai korelasi dan hubungan asosiasi fitur-fitur ini yang akan dijadikan acuan pada tahap pemodelan. Scatter plot antara klas UKT dan penghasilan dengan faktor daya listrik dapat dilihat pada Gambar 6.




Gambar 6. Scatter Plot Klas UKT dan Penghasilan

*B. Pemodelan Klasifikasi*

Pada tahap pemodelan, dilakukan pemodelan dengan variasi jumlah fitur secara *backward*, yaitu dari 8 fitur sampai dengan 1 fitur. Urutan penghapusan fitur didasarkan pada nilai korelasi yang paling rendah atau kekuatan hubungan asosiasi dari fitur-fitur yang digunakan. Berdasarkan urutan, penghapusan fitur dimulai dari fitur jumlah anak, status kepemilikan tanah, jumlah sepeda motor, daya listrik, penghasilan, jumlah mobil, dan terakhir jumlah rumah. Masing-masing metode dilakukan pemodelan sebanyak 3 kali dengan pembangkitan bilangan random (*random state*) yang berbeda.

Tabel III adalah matriks konfusi metode *Random Forest* untuk data *training* dengan tingkat akurasi sebesar 100%. Dari 698 record, ada sebanyak 130 terklasifikasi pada UKT-1 (18,6%), UKT-2 sebanyak 38 (5,4%), UKT-3 sebanyak 118 (16,9%), UKT-4 sebanyak 172 (24,6%), UKT-5 sebanyak 131 (18,8%), UKT-6 sebanyak 93 (13,3%), dan UKT-7 sebanyak 16 (2,3%).

TABEL III
MATRIKS KONFUSI METODE RANDOM FOREST (DATA TRAINING)

| | | **Prediksi Klas UKT** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Aktual Klas UKT** | 1 | 130 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 0 | 38 | 0 | 0 | 0 | 0 | 0 |
| | 3 | 0 | 0 | 118 | 0 | 0 | 0 | 0 |
| | 4 | 0 | 0 | 0 | 172 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 0 | 0 | 131 | 0 | 0 |
| | 6 | 0 | 0 | 0 | 0 | 0 | 93 | 0 |
| | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 16 |

Sedangkan Tabel IV adalah matriks konfusi metode *Random Forest* untuk data *testing* dengan tingkat akurasi sebesar 94,3%. Dari 175 record, ada sebanyak 41 terklasifikasi pada UKT-1 (23,4%), UKT-2 sebanyak 11 (6,3%), UKT-3 sebanyak 34 (19,4%), UKT-4 sebanyak 31 (17,7%), UKT-5 sebanyak 24 (13,7%), UKT-6 sebanyak 22 (12,6%), dan UKT-7 sebanyak 2 (1,1%).

TABEL IV
MATRIKS KONFUSI METODE RANDOM FOREST (DATA TESTING)

| | | Prediksi Klas UKT | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Aktual Klas UKT | 1 | 41 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 1 | 11 | 0 | 2 | 1 | 0 | 0 |
| | 3 | 0 | 0 | 34 | 0 | 0 | 0 | 0 |
| | 4 | 0 | 0 | 0 | 31 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 0 | 0 | 24 | 3 | 0 |
| | 6 | 1 | 0 | 0 | 0 | 0 | 22 | 0 |
| | 7 | 0 | 0 | 0 | 1 | 0 | 1 | 2 |

Berikutnya Tabel V adalah matriks konfusi metode MLP untuk data *training* dengan tingkat akurasi yang sangat tinggi juga yaitu sebesar 96,9%. Dari 699 record, ada sebanyak 129 terklasifikasi pada UKT-1 (18,5%), UKT-2 sebanyak 27 (3,9%), UKT-3 sebanyak 116 (16,6%), UKT-4 sebanyak 172 (24,6%), UKT-5 sebanyak 131 (18,7%), UKT-6 sebanyak 92 (13,2%), dan UKT-7 sebanyak 11 (1,6%).

TABEL V
MATRIKS KONFUSI METODE MLP (DATA TRAINING)

| | | Prediksi Klas UKT | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Aktual Klas UKT | 1 | 129 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 2 | 0 | 27 | 5 | 5 | 1 | 0 | 0 |
| | 3 | 0 | 0 | 116 | 0 | 1 | 1 | 0 |
| | 4 | 0 | 0 | 0 | 172 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 0 | 0 | 131 | 0 | 0 |
| | 6 | 1 | 0 | 0 | 1 | 0 | 92 | 0 |
| | 7 | 0 | 0 | 0 | 0 | 2 | 3 | 11 |

Sedangkan Tabel VI adalah matriks konfusi metode MLP untuk data *testing* dengan tingkat akurasi sebesar 94,6%. Dari 175 record, ada sebanyak 41 terklasifikasi pada UKT-1 (23,4%), UKT-2 sebanyak 9 (5,1%), UKT-3 sebanyak 34 (19,4%), UKT-4 sebanyak 31 (17,7%), UKT-5 sebanyak 26 (14,9%), UKT-6 sebanyak 23 (13,1%), dan UKT-7 sebanyak 2 (1,1%).

TABEL VI
MATRIKS KONFUSI METODE MLP (DATA TESTING)

| | | Prediksi Klas UKT | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Aktual Klas UKT | 1 | 41 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 1 | 9 | 0 | 3 | 2 | 0 | 0 |
| | 3 | 0 | 0 | 34 | 0 | 0 | 0 | 0 |
| | 4 | 0 | 0 | 0 | 31 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 0 | 0 | 26 | 1 | 0 |
| | 6 | 0 | 0 | 0 | 0 | 0 | 23 | 0 |
| | 7 | 0 | 0 | 0 | 0 | 0 | 2 | 2 |

### *C. Evaluasi Model Klasifikasi*

Rerata tingkat akurasi masing-masing model dengan variasi jumlah fitur pada data training dapat dilihat pada Tabel VII.

TABEL VII
RERATA TINGKAT AKURASI (DATA *TRAINING*)

| Jumlah Fitur | Metode Klasifikasi | | | |
|---|---|---|---|---|
| | Random Forest | Regresi Logistik | Naïve Bayes | Multilayer Perceptron |
| **8** | 100.00% | 90.59% | 68.91% | 96.80% |
| **7** | 97.76% | 89.21% | 68.15% | 91.12% |
| **6** | 94.70% | 82.43% | 66.43% | 85.77% |
| **5** | 88.49% | 74.64% | 59.93% | 76.22% |
| **4** | 79.23% | 63.66% | 52.29% | 68.34% |
| **3** | 62.46% | 61.17% | 47.80% | 61.27% |
| **2** | 45.08% | 44.65% | 38.87% | 44.75% |
| **1** | 42.98% | 42.98% | 42.98% | 42.98% |

Berdasarkan informasi Tabel VII, secara umum metode *Random Forest* memiliki tingkat akurasi paling tinggi. Akurasi tertinggi berikutnya adalah metode *Multilayer Perceptron*, kemudian metode regresi logistik, dan metode Naïve Bayes. Pada model dengan 8 fitur, tingkat akurasi metode Random Forest sebesar 100%, disusul metode *Multilayer Perceptron* dengan akurasi sebesar 96,80%, dan Regresi Logistik sebesar 90,59%. Metode Naïve Bayes hanya mempunyai akurasi sebesar 68,91%.

Penghapusan jumlah fitur cenderung menurunkan tingkat akurasi. Pada metode *Random Forest* dan *Multilayer Perceptron* dengan 5 fitur masih mempunyai akurasi lebih dari 75%. Pada metode *Random Forest*, penurunan tingkat akurasi yang sangat besar pada saat penghapusan 5 – 6 fitur. Kinerja keempat metode tersebut cenderung sama ketika hanya melibatkan dua fitur (status dan jumlah rumah), yaitu antara 38,87 – 45,08%. Pada data testing, rerata tingkat akurasi keempat metode dapat dilihat pada Tabel VIII.

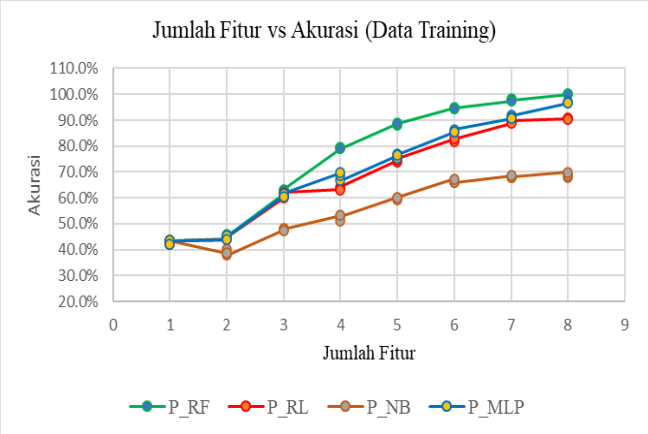
TABEL VIII
RERATA TINGKAT AKURASI (DATA *TESTING*)

| Fitur | Metode Klasifikasi | | | |
|---|---|---|---|---|
| | RF | RL | GNB | MLP |
| **8** | 95.81% | 88.76% | 67.81% | 94.29% |
| **7** | 85.14% | 87.81% | 66.67% | 89.52% |
| **6** | 80.00% | 81.71% | 65.52% | 84.19% |
| **5** | 74.48% | 74.86% | 61.52% | 76.57% |
| **4** | 65.90% | 61.52% | 51.81% | 65.52% |
| **3** | 60.19% | 60.00% | 47.24% | 60.00% |
| **2** | 44.76% | 44.76% | 42.67% | 44.76% |
| **1** | 42.29% | 42.29% | 42.29% | 42.29% |

Metode *Random Forest* mempunyai tingkat akurasi tertinggi ketika menggunakan 8 fitur, yaitu sebesar 95,81%. Akurasi tertinggi kedua dan ketiga adalah *Multilayer Perceptron* dan Regresi Logistik dengan tingkat akurasi masing-masing sebesar 94,29% dan 88,76%. Berikutnya

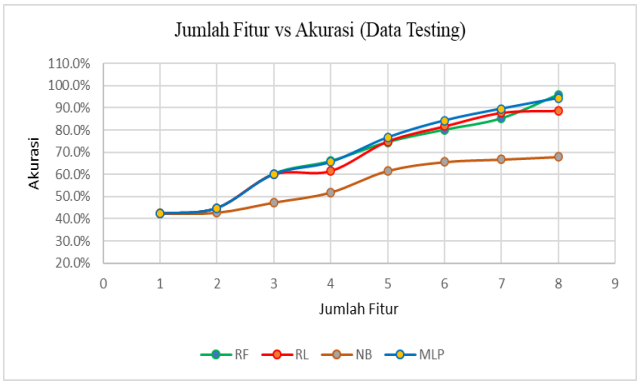
Multilayer Perceptron mempunyai tingkat akurasi tertinggi ketika menggunakan 5 – 7 fitur. Dan pada saat menggunakan 2 – 4 fitur, *Random Forest*, Regresi Logistik, dan *Multilayer Perceptron* mempunyai tingkat akurasi yang relatif sama. Sedangkan metode Naïve Bayes mempunyai kinerja paling rendah dibandingkan tiga metode lainnya.

Secara visual, hubungan antara jumlah fitur dengan tingkat akurasi data training dapat dilihat pada Gambar 7. Metode *Random Forest*, Regresi Linier, dan *Multilayer Perceptron* mempunyai tingkat akurasi yang hampir sama pada model dengan 1 – 3 fitur.Namun pada pemodelan 4 – 8 fitur, kinerja *Random Forest* lebih unggul dibandingkan metode Regresi Linier dan *Multilayer Perceptron* yang cenderung mempunyai kinerja yang sama. Sedangkan metode Naïve Bayes, tingkat akurasinya cenderung rendah dibandingkan metode lainnya.



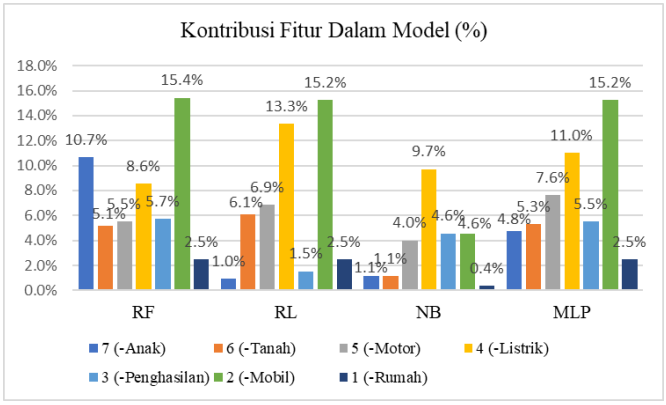

Gambar 7. Hubungan Jumlah Fitur vs. Tingkat Akurasi (Data Training)

Gambar 8 adalah hubungan antara jumlah fitur dengan tingkat akurasi data testing. Tingkat akurasi metode *Random Forest*, Regresi Linier, dan *Multilayer Perceptron* mempunyai tingkat akurasi yang cenderung sama pada model dengan 1 – 8 fitur. Namun pada metode Naïve Bayes, hampir tidak ada peningkatan akurasi pada pemodelan dari 7 fitur ke 8 fitur.




Gambar 8. Hubungan Jumlah Fitur vs. Tingkat Akurasi (Data Training)

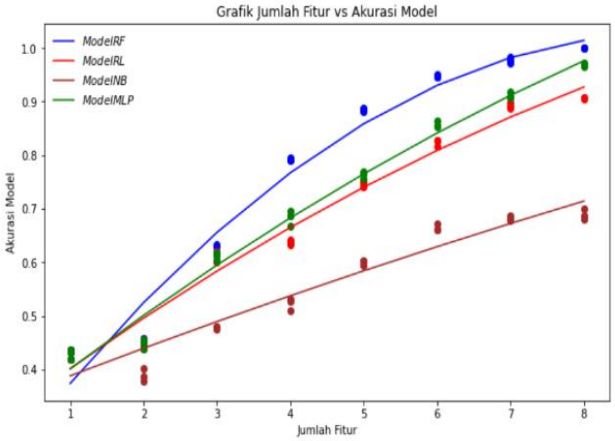
Gambar 9 adalah proporsi atau kontribusi masing-masing fiturdalam peningkatan akurasi model dapat dihitung dari penurunan tingkat akurasi setelah fitur tersebut dihilangkan. Berdasarkan tiga model terbaik, fitur jumlah mobil mempunyai kontribusi terbesar dalam meningkatkan akurasi model, yaitu di atas 15%. Fitur yang mempunyai kontribusi terbesar kedua adalah daya listrik, yaitu antara 8,6% - 13,3%.




Gambar 9. Persentase Kontribusi Fitur (Variabel Prediktor)

Pada model Random Forest. Kontribusi kedua adalah jumlah anak (10,7%), baru disusul daya listrik (8,6%), penghasilan (5,7%), jumlah motor (5,5%), dan kepemilikan tanah (5,1%). Fitur jumlah rumah memiliki kontribusi paling kecil, yaitu 2,5%.

Secara umum tingkat akurasi terhadap jumlah fitur keempat metode tersebut dapat dilihat pada Gambar 10. Keempat model tersebut didekati dengan persamaan polinomial derajat dua. Model Random Forest cenderung mempunyai tingkat akurasi yang lebih tinggi daripada model lainnya setiap variasi jumlah fitur. Model berikutnya adalah Multilayer Perceptron dan Regresi Logistik.




Gambar 10. Pendekatan Model Sebagai Fungsi Polinomial Jumlah Fitur

Detail metrik kinerja beberapa model tersebut dapat dilihat pada Tabel IX. Secara umum masing-masing model mempunyai total variasi $R^2$ antara 92,7% - 98,4%. Total variasi yang dijelaskan oleh komponen linier lebih dominan dibandingkan komponen kuadratik. Pada model *Random Forest* total variasi komponen linier sebesar 92,61%. Berikutnya model Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron masing-masing sebesar 96,76%, 92,59%, dan 97,87%.

TABEL IX
TOTAL VARIASI DAN FUNGSI POLINOMIAL MODEL

| | S | $R^2$ | SS Linier | SS Kuadratik | SS Reg |
|---|---|---|---|---|---|
| **RF** | 0.0407 | 96.9% | 1.0546 | 0.0494 | 1.1388 |
| **RL** | 0.0299 | 97.4% | 0.7106 | 0.0051 | 0.7344 |
| **NB** | 0.0320 | 92.7% | 0.2736 | 0.0003 | 0.2955 |
| **MLP** | 0.0259 | 98.4% | 0.8501 | 0.0043 | 0.8686 |

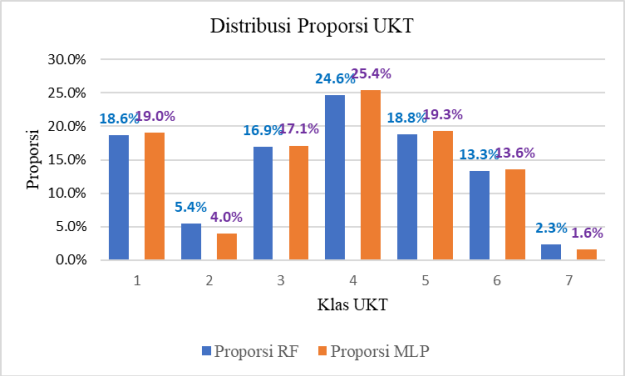
RF = 0.2031 + 0.1806 Fitur - 0.009904 $Fitur^2$
RL = 0.3010 + 0.1037 Fitur - 0.003181 $Fitur^2$
GNB = 0.3349 + 0.05386 Fitur - 0.000807 $Fitur^2$
MLP = 0.2952 + 0.1087 Fitur - 0.002951 $Fitur^2$

Hasil prediksi tingkat akurasi dengan pendekatan persamaan polinomial untuk masing-masing metode dapat dilihat pada Tabel X. Dengan 5 fitur, metode *Random Forest* diprediksi mempunyai akurasi sebesar 85,9%. Sedangkan metode Regresi Linier, Naïve Bayes, dan *Multilayer Perceptron* masing-masing sebesar 74,0%, 58,4%, dan 76,5%. Metode *Random Forest* diprediksi mempunyai tingkat akurasi sebesar 100% dengan menggunakan 8 fitur dalam model. Sedangkan *Multilayer Perceptron* diprediksi mempunyai akurasi 100% dengan menggunakan 9 fitur, Regresi Logistik sebanyak 10 fitur, dan Naïve Bayes sebanyak 17 fitur.

TABEL X
PREDIKSI RERATA AKURASI

| Jumlah Fitur | Model Polinomial | | | |
|---|---|---|---|---|
| | RF | RL | NB | MLP |
| **5** | 85.9% | 74.0% | 58.4% | 76.5% |
| **6** | 93.0% | 80.9% | 62.9% | 84.1% |
| **7** | 98.2% | 87.1% | 67.2% | 91.2% |
| **8** | **100.0%** | 92.7% | 71.4% | 97.6% |
| **9** | 100.0% | 97.7% | 75.4% | **100.0%** |
| **10** | 100.0% | **100.0%** | 79.3% | 100.0% |
| **11** | 100.0% | 100.0% | 83.0% | 100.0% |
| **12** | 100.0% | 100.0% | 86.5% | 100.0% |
| **13** | 100.0% | 100.0% | 89.9% | 100.0% |
| **14** | 100.0% | 100.0% | 93.1% | 100.0% |
| **15** | 100.0% | 100.0% | 96.1% | 100.0% |
| **16** | 100.0% | 100.0% | 99.0% | 100.0% |
| **17** | 100.0% | 100.0% | **100.0%*** | 100.0% |

Berdasarkan empat metode yang digunakan, Gambar 11 adalah perbandingan metode *Random Forest* dan *Multilayer Perceptron* berdasarkan distribusi proporsi pada masing-masing klas UKT. Kedua metode tersebut cenderung mempunyai proporsi yang sama untuk setiap kelas UKT. Proporsi yang paling tinggi pada klas UKT 4 atau Rp. 4.500.000 dengan proporsi 24,6 – 25,4%. Proporsi terbesar berikutnya pada klas 5 atau Rp. 5.500.000 dengan proporsi 18.8 – 19,3%.




Gambar 11. Distribusi Proporsi UKT (Random Forest dan MLP)

Tabel XI adalah perbandingan nilai ekspektasi UKT metode *Random Forest* dan *Multilayer Perceptron*. Dari perhitungan tersebut, nilai ekspektasi atau rerata UKT metode *Random Forest* sebesar Rp. 3.833.811 dengan simpangan baku Rp. 2,123,758. Sedangkan rerata UKT metode *Multilayer Perceptron* sebesar Rp. 3.856.195 dengan simpangan baku Rp. 2,093,933. Kedua metode tersebut tidak mempunyai perbedaan yang terlalu besar. Nilai ekspektasi ini dapat menjadi dasar prediksi perhitungan sumber PNBP sekaligus acuan dan pertimbangan dalam menentukan proporsi di masing-masing klas UKT berdasarkan kebijakan pemerintah maupun internal.

TABEL XI
EKSPEKTASI RERATA DAN SIMPANGAN BAKU UKT

| Kelas UKT | UKT (Rp. Jt) | % RF | E(RF) | %MLP | E(MLP) |
|---|---|---|---|---|---|
| **1** | 0.5 | 18.6% | 93,123 | 19.0% | 95,133 |
| **2** | 1.0 | 5.4% | 54,441 | 4.0% | 39,823 |
| **3** | 3.0 | 16.9% | 507,163 | 17.1% | 513,274 |
| **4** | 4.5 | 24.6% | 1,108,883 | 25.4% | 1,141,593 |
| **5** | 5.5 | 18.8% | 1,032,235 | 19.3% | 1,062,684 |
| **6** | 6.5 | 13.3% | 866,046 | 13.6% | 882,006 |
| **7** | 7.5 | 2.3% | 171,920 | 1.6% | 121,681 |
| **Total** | | 100.0% | | 100.0% | |
| **Ekspektasi Rerata UKT (Rp.)** | | | **3,833,811** | | **3,856,195** |
| **Ekspektasi Simp. Baku UKT (Rp.)** | | | **2,123,758** | | **2,093,933** |

## IV. KESIMPULAN

Hasil penelitian ini, metode *Random Forest* mempunyai rerata akurasi tertinggi, yaitu sebesar 97,9%. Sedangkan *Multilayer Perceptron* tertinggi kedua, yaitu dengan rerata akurasi 95,54%, kemudian Regresi Logistik dengan rerata akurasi sebesar 89,68%, dan Naïve Bayes sebesar 68,36%. Dengan pendekatan model regresi polinomial, jumlah fitur berpengaruh secara signifikan terhadap tingkat akurasi, baik secara linier maupun kuadratik. Setiap penambahan 1 fitur dalam model *Random Forest*, ada kecenderungan kenaikan rerata akurasi sebesar 18,06 %. Sedangkan penambahan 1 fitur pada *Multilayer Perceptron* dan Regresi logistik cenderung meningkatkan rerata akurasi masing-masing sebesar 10,87% dan 10,37%.

Berdasarkan hasil tersebut, pada dasarnya metode *Random Forest*, Regresi Linier, dan *Multilayer Perceptron* dapat digunakan sebagai model klasifikasi UKT. Ketiga metode

tersebut mempunyai rerata akurasi yang tinggi. Seleksi metode yang digunakan sebaiknya tidak hanya berdasarkan rerata akurasi maupun karakteristik fungsi rerata akurasi. Dalam penerapannya juga dapat mempertimbangkan waktu komputasi, distribusi UKT, dan nilai ekspektasi UKT dari metode yang dipilih. Model klasifikasi yang dipilih dapat digunakan untuk membangun sistem aplikasi penetapan UKT atau dibenamkan pada sistem informasi yang sudah ada.

## REFERENSI

[1] M. Pendidikan, D. A. N. Kebudayaan, and R. Indonesia, "ht ps:/ publik2 .blogspot.com/2020/ 7/permendikbud-nomor-25-tahun-2020.html," 2020.

[2] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.

[3] D. W. Triscowati, B. Sartono, A. Kurnia, D. Dirgahayu, and A. W. Wijayanto, "Classification of Rice-Plant Growth Phase Using Supervised Random Forest Method Based on Landsat-8 Multitemporal Data," *Int. J. Remote Sens. Earth Sci.*, vol. 16, no. 2, p. 187, 2020, doi: 10.30536/j.ijreses.2019.v16.a3217.

[4] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.

[5] B. Siswoyo, "MultiClass Decision Forest Machine Learning Artificial Intelligence," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 1–7, 2020, doi: 10.30871/jaic.v4i1.1155.

[6] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.

[7] I. K. P. Suniantara, "Analisis Random Forest Pada Klasifikasi Cart Universitas Terbuka Analysis of Random Forest In Inaccuracies CART Classification of Terbuka University Student Graduates," vol. 13, no. 3, pp. 179–186, 2019.

[8] A. P. Wibawa *et al.*, "Naïve Bayes Classifier for Journal Quartile Classification," *Int. J. Recent Contrib. from Eng. Sci. IT*, vol. 7, no. 2, p. 91, 2019, doi: 10.3991/ijes.v7i2.10659.

[9] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.

[10] A. Agresti, "3Rd-Ed-Alan_Agresti_Categorical_Data_Analysis.Pdf," *International encyclopedia of statistical science*, vol. 47, no. 4. pp. 755–758, 2013.

[11] A. J. Scott, D. W. Hosmer, and S. Lemeshow, *Applied Logistic Regression.*, vol. 47, no. 4. 1991.

[12] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.

[13] I. N. Purnama, "Perbandingan Klasifikasi Website Secara Otomatis Menggunakan Metode Multilayer Perceptron dan Naive Bayes," *J. Sist. Komput. dan Inform.*, vol. 2, no. 2, pp. 155–161, 2021, doi: 10.30865/json.v2i2.2703.