

PENERAPAN MACHINE LEARNING UNTUK MENGATASI KETIMPANGAN DATA DALAM MENENTUKAN KLASIFIKASI UANG KULIAH TUNGGAL (UKT)

Nurul Tarizya Syam, Irmawati, Zawiyah Saharuna

Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Indonesia

Info Artikel

Riwayat artikel:

Received, (1 Juni 2024)

Revised, (15 Juni 2024)

Accepted, (22 Juni 2024)

Kata kunci:

UKT;

Klasifikasi,

Imbalance Data;

Smote;

ABSTRACT

Single Tuition Fee or UKT is a tuition fee borne by students every semester. Payment is made every time students enter a new semester while studying at tertiary institutions. One of the state universities in Indonesia that has implemented the UKT payment system is Politeknik Negeri Ujung Pandang. Based on observations, the purchase of UKT is still done manually, so it has the potential to produce decisions that are not on target. This study was classified based on the amount of single student tuition fees using the smote method and without smote. The algorithm used in classifying is SVM, Decision Tree, Random Forest. The data used is 985 UKT data in 2021. Based on experiments that have been carried out with the Random Forest algorithm, it has the best performance compared to the SVM and Decision Tree algorithms. The proportion of results obtained before being hit is accuracy of 84.75%, precision of 79.22%, recall of 81.15% and F1 score of 80.17%. Whereas after applying smote it increased with an accuracy proportion of 98.9%. So it can be interpreted that the best algorithm used in classification is the Random Forest algorithm by applying smote.

ABSTRAK

Uang Kuliah Tunggal atau UKT adalah biaya kuliah yang dibayarkan oleh mahasiswa setiap semesternya. Pembayaran dilakukan setiap kali mahasiswa memasuki semester baru selama menempuh pendidikan di perguruan tinggi. Salah satu perguruan tinggi negeri di Indonesia yang telah menerapkan sistem pembayaran UKT adalah Politeknik Negeri Ujung Pandang. Berdasarkan observasi, penentuan UKT masih dilakukan secara manual, sehingga berpotensi menghasilkan keputusan yang tidak tepat sasaran. Penelitian ini dilakukan klasifikasi berdasarkan besaran uang kuliah tunggal mahasiswa menggunakan metode smote dan tanpa smote. Algoritma yang digunakan dalam melakukan klasifikasi adalah SVM, *Decision Tree*, *Random Forest*. Data yang digunakan sebanyak 985 data UKT pada tahun 2021. Berdasarkan eksperimen yang telah dilakukan dengan algoritma *Random Forest* memiliki performa yang paling baik di banding dengan algoritma SVM dan *Decision Tree*. Hasil persentase yang diperoleh sebelum dilakukan smote yaitu akurasi sebesar 84,75%, presisi 79,22%, *Recall* 81,15% dan F1 Score 80,17%. Sedangkan setelah menerapkan smote meningkat dengan persentase akurasi sebesar 98,9%. Maka dapat disimpulkan bahwa algoritma yang paling baik digunakan dalam klasifikasi adalah algoritma *Random Forest* dengan menerapkan smote.

Penulis Korespondensi:

Irmawati

Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Jl. Perintis Kemerdekaan KM. 10, Makassar, 90245

Email: irmawati@poliupg.ac.id

1. PENDAHULUAN

Uang Kuliah Tunggal (UKT) merupakan biaya pendidikan mahasiswa yang ditetapkan selama satu semester, dan sudah mendapat pengurangan biaya melalui subsidi pemerintah sehingga tidak ada lagi pungutan biaya di luar biaya UKT [1]. Menurut Permendikbud Nomor 55 Tahun 2013, penetapan biaya UKT tersebut

disesuaikan dengan status ekonomi masing-masing mahasiswa [2]. Salah satu cara yang paling umum bagi kampus untuk dapat mengetahui kondisi perekonomian mahasiswa adalah dengan melakukan wawancara bersama calon mahasiswanya.

Berdasarkan observasi yang dilakukan di kampus Politeknik Negeri Ujung Pandang (PNUP), bahwa penentuan UKT masih dilakukan secara manual dan berulang-ulang setiap tahunnya. Hal ini berakibat pada proses yang tidak efisien, dan berpotensi tidak konsistennya penetapan UKT terhadap mahasiswa, serta membutuhkan waktu yang lama bagi kampus untuk menyesuaikan data mahasiswa baru ke kelompok UKT yang sesuai. Selain itu dalam kurun waktu 4 tahun (2018 – 2021), terdapat sekitar 0,9% - 4,8% dari total mahasiswa mengajukan permohonan keringanan UKT dengan alasan keberatan dengan kelas UKT yang ditetapkan karena permasalahan ekonomi. Di sisi lain terdapat penelitian yang mengangkat isu tersebut dan mengklasifikasikan UKT menggunakan perbandingan beberapa metode klasifikasi Machine Learning, diantaranya Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron. Klasifikasi adalah pendekatan supervised learning dengan tujuan untuk menemukan model atau fungsi yang mampu membedakan data berdasarkan fitur/kelas pada data [3].

Klasifikasi dapat diterapkan dalam berbagai aspek sehingga seiring berjalannya waktu metode klasifikasi cukup banyak dikembangkan, namun terdapat permasalahan yang sering ditemui dalam klasifikasi yaitu masalah ketidakseimbangan data (*Imbalance Data*). *Imbalance Data* merupakan suatu keadaan dimana distribusi kelas data tidak seimbang, jumlah kelas data (*instance*) yang satu lebih sedikit atau lebih banyak dibandingkan dengan jumlah kelas data lainnya [4]. Kelompok kelas data yang lebih sedikit dikenal dengan kelompok minoritas (*minority*), kelompok kelas data yang lainnya disebut dengan kelompok mayoritas (*majority*). Data yang akan diperiksa akan cukup banyak dan akan diklasifikasikan menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyelesaikan masalah ketidakseimbangan kelas pada dataset UKT.

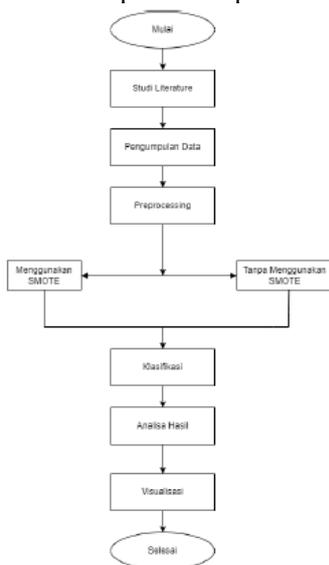
Ketidakeimbangan dataset pada *machine learning* adalah masalah yang serius. Dataset yang tidak seimbang menyebabkan misleading atau kesesatan dalam hasil klasifikasi dimana data kelas minoritas sering diklasifikasikan sebagai kelas mayoritas [5]. Penerapan algoritma klasifikasi tanpa memperhatikan keseimbangan kelas mengakibatkan prediksi yang baik bagi kelas mayoritas dan kelas minoritas diabaikan. Apabila algoritma klasifikasi di implementasikan langsung terhadap dataset yang *imbalance* maka akan mengalami penurunan performa [6].

Penelitian yang dilakukan oleh Wenhao et al [7] dengan judul “*An Improved Oversampling Algorithm Based on the Samples Selection Strategy for Classifying Imbalance Data*” yang bertujuan untuk meningkatkan metode *Oversampling* terhadap pengklasifikasian data tidak seimbang. Hasil penelitian ini menunjukkan bahwa algoritma *Oversampling* yang ditingkatkan menunjukkan kinerja yang baik ada klasifikasi data tidak seimbang.

Berdasarkan latar belakang yang telah dikemukakan, maka akan dilakukan penelitian untuk penanganan *imbalance data* pada klasifikasi uang kuliah tunggal (UKT) dengan berbasis *machine learning*, dengan menggunakan metode SMOTE. Hasil klasifikasi uang kuliah tunggal (UKT) akan mengkonfirmasi peningkatan kinerja model dengan menggunakan metode SMOTE untuk klasifikasi UKT mahasiswa.

2. METODE

Prosedur penelitian secara keseluruhan dapat dilihat pada Gambar 1. sebagai berikut :



Gambar 1. Prosedur Penelitian

2.1. Studi Literature

Studi literatur merupakan kegiatan yang dilakukan untuk memperoleh dan mengumpulkan data serta referensi yang tepat dalam penelitian dengan cara mempelajari, membaca, dan mencatat literatur dari beberapa buku, jurnal maupun dari internet baik berupa teori, laporan maupun penemuan sebelumnya yang berkaitan dengan rumusan masalah penelitian.

2.2. Pengumpulan Data

Penelitian ini menggunakan dataset dari data primer yang merupakan data UKT mahasiswa Kampus Politeknik Negeri Ujung Pandang tahun angkatan 2021 yang didapatkan dari Kasubag Akademik Politeknik Negeri Ujung Pandang dengan format .xlsx. Dataset ini mengambil data asli yang diperoleh dari Bulan Juni 2021 hingga bulan Agustus 2021 dengan jumlah total 985 baris data. Atribut yang didapat dari data pada penelitian ini yaitu NIM, penghasilan orang tua, pendapatan orang tua, jumlah tanggungan, daftar kipk, biaya pbb, biaya tagihan listrik perbulan, status kepemilikan rumah, kepemilikan lahan, daya listrik.

2.3. Preprocessing

Data preprocessing dibagi menjadi beberapa langkah, yaitu *cleaning data*, *normalization data*.



Gambar 2. *Preprocessing*

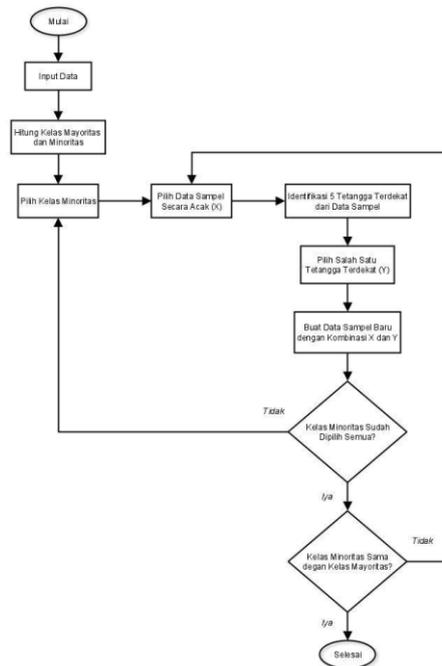
Data preprocessing merupakan sekumpulan teknik yang diterapkan pada database untuk menghapus *noise*, *missing value*, dan data yang tidak konsisten. Data preprocessing ini digunakan karena dalam data *realtime database* seringkali tidak lengkap dan tidak konsisten sehingga mengakibatkan hasil data mining tidak tepat dan kurang akurat. Oleh karena itu, untuk meningkatkan kualitas data yang akan dianalisis, perlu dilakukan langkah-langkah preprocessing data. Data yang baru saja dikumpulkan kemungkinan besar memiliki banyak bagian yang tidak relevan bahkan ada bagian yang hilang. Oleh karena itu perlu adanya proses pembersihan data atau biasa dikenal dengan *data cleaning*. Hal yang bisa diatasi menggunakan data cleaning adalah penanganan *missing value* dan *noise*. *Missing value* merupakan kondisi dimana adanya data yang hilang atau tidak lengkap di dalam database. Cara untuk mengatasi *missing value* adalah dengan mengabaikan *row* atau baris data tersebut lalu mengisi *missing value* tersebut.

Proses selanjutnya akan dilakukan *data transformation* untuk mengubah data dalam bentuk yang sesuai dalam proses *data mining*. Teknik yang dilakukan untuk *data transformation* adalah *normalization*, pemilihan atribut, dan *discretization*. *Normalization* dilakukan untuk menskalakan nilai data dalam rentang nilai tertentu, misalnya -1 sampai 1 atau 0 sampai 1. Teknik kedua adalah pemilihan atribut. Pemilihan atribut merupakan proses pemilihan atribut yang diberikan untuk proses *data mining*. Terakhir adalah teknik *discretization*, teknik ini dilakukan untuk mengganti *row value* pada atribut numerik dengan nilai interval. Setelah dilakukan transformasi data, data yang digunakan akan melalui tahap *reduction* untuk meningkatkan efisiensi penyimpanan atau memori.

2.4. Metode SMOTE

Metode SMOTE yang digunakan untuk melakukan proses *oversampling* pada dataset, teknik *oversampling* yang diimplementasikan berfungsi untuk memperbanyak data pada kelas data minoritas agar data menjadi seimbang. Jumlah data pada kelas minoritas diperbanyak sehingga memiliki data yang serupa dengan kelas mayoritas. Metode SMOTE yang diimplementasikan dimulai dari menghitung jumlah data pada kelas mayoritas atau kelas yang paling banyak datanya. Jumlah ini akan menjadi batas pembuatan data baru pada kelas-kelas minoritas. Setelah itu, sampel data acak dipilih dari data kelas minoritas tersebut. Kemudian, sejumlah lima data *nearest neighbors* dari data terpilih tersebut ditentukan dengan menggunakan metode SMOTE memanfaatkan

algoritma *K-Nearest Neighbors* (KNN) untuk menentukan data baru yang ditambahkan pada dataset, berikut adalah *flowchart* yang digunakan:



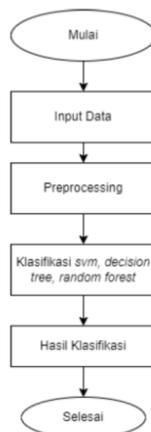
Gambar 3. Alur SMOTE

2.5 Klasifikasi

Secara umum pada tahap ini data akan melewati dua tahapan proses, yakni *data training* dan *data testing*. *Data training* bertujuan untuk melatih data yang akan dianalisa oleh mesin dan akan memberikan validasi data. Dataset yang dilatih terlebih dahulu akan menghasilkan keseimbangan pada masalah *overfitting* dan *underfitting*. Sedangkan pada proses *data testing* dataset akan diuji untuk dilihat tingkat akurasi atau bisa juga dilihat performanya sesuai model yang kita pilih.

Pada proses klasifikasi ini, hasil akurasi bergantung pada algoritma klasifikasi yang digunakan. Tiap model atau algoritma klasifikasi memiliki beragam variabel yang berbeda, sehingga akan mempengaruhi performansi tiap-tiap algoritma yang digunakan. Penelitian ini akan menggunakan tiga model algoritma klasifikasi, yaitu *support vector machine* (SVM), *decision tree* dan *random forest*.

Penelitian kali ini akan menggunakan pembagian *dataset* menggunakan metode deterministik/*holdout*, yaitu dengan menentukan rasio pembagian dari kedua dataset tersebut. Hasil penelitian [10] menyatakan dalam studi yang dilakukannya tentang efek dari *training data* pada model klasifikasi menunjukkan bahwa *error* terkecil didapatkan pada jumlah *training data* 79,09 sampai 86,19 persen. Oleh karena itu, pada penelitian ini digunakan 80 persen dari keseluruhan dataset untuk *training data* dan sisanya 20 persen data digunakan untuk *testing data* yang dapat menghasilkan tingkat kinerja klasifikasi. Gambaran alur proses klasifikasi dapat dilihat pada Gambar 4.



Gambar 1. Klasifikasi

2.6 Analisa Hasil

Model algoritma klasifikasi yang telah diuji melalui rangkaian *data testing* akan dilakukan evaluasi kinerja klasifikasi. Ukuran kinerja klasifikasi yang digunakan dalam penelitian ini akan memperhatikan *confusion matrix*. *Confusion matrix* adalah alat yang berguna untuk menganalisis seberapa baik atau seberapa akurat metode klasifikasi dapat mengenali objek pengamatan dari kelas yang berbeda [11]. Setelah hasil dari pengujian telah diketahui, maka akan dilakukan analisis perbandingan dari hasil semua algoritma yang telah diuji untuk melihat tingkat akurasi yang paling optimal untuk memecahkan masalah dari kasus yang telah disebutkan dan akan dilakukan perbandingan dengan penelitian sebelumnya.

2.7 Visualisasi

Pada tahapan visualisasi ini tujuannya untuk menampilkan visual sistem yang dapat berinteraksi dengan pengguna sehingga mempermudah pengguna dalam menjalankan sistem. Hasil klasifikasi dalam menentukan *class ukt* akan ditampilkan pada framework web yakni *streamlit*. Pada penelitian ini memanfaatkan *streamlit* yang berbahasa *python* dan bersifat *open source* dalam membangun sebuah website dengan model *machine learning* ataupun *data science*.

Streamlit juga menawarkan kemudahan bagi pengembang untuk mendeploy aplikasinya, baik itu *web* ataupun *mobile*. Untuk melakukan proses *deploy* pada *streamlit*, terlebih dahulu melakukan *push* kode program ke github, kemudian mendaftarkan link github ke dalam *streamlit*. Hasil dari tahapan ini adalah berupa klasifikasi UKT menggunakan model *machine learning* dengan performa terbaik.

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data Mahasiswa pendaftar Seleksi Bersama Masuk Politeknik Negeri Ujung Pandang (SBMPN) yang diterima tahun 2021 pada Politeknik Negeri Ujung Pandang (PNUP). Data ini diperoleh dari Sub-Bagian Akademik Kampus Politeknik Negeri Ujung Pandang dengan referensi data permohonan pengurangan Uang Kuliah Tunggal (UKT) yang memiliki tipe data excel dengan ukuran data sebesar 1104 baris dan 34 kolom.

Tabel 1. Informasi Dataset

Fitur	Type
<i>timestamp</i>	Numerikal
<i>no_pendaftaran</i>	Numerikal
<i>nama</i>	Kategorikal
<i>program_studi</i>	Kategorikal
<i>no_hp</i>	Numerikal
<i>pekerjaan_ayah</i>	Kategorikal
<i>pekerjaan_ibu</i>	Kategorikal
<i>pendapatan_ayah</i>	Numerikal
<i>pendapatan_ibu</i>	Numerikal
<i>jumlah_tanggungan</i>	Numerikal
<i>pembayaran_listrik</i>	Numerikal
<i>pembayaran_pbb</i>	Numerikal
<i>pembayaran_PKB</i>	Numerikal
<i>nama_pewawancara</i>	Kategorikal
<i>daftar_KIPK</i>	Numerikal
<i>Tanggungan</i>	Numerikal
<i>ukt_gaji</i>	Numerikal
<i>ukt_tanggungan</i>	Numerikal
<i>get_ukt</i>	Numerikal
<i>ukt_min</i>	Numerikal

3.2. Preprocessing

3.2.1 Data Cleaning

Proses ini dilakukan untuk membersihkan data dari *missing value* atau data yang bernilai null. Pada dataset ini tidak banyak data yang memiliki *missing value*. Untuk meningkatkan efisiensi kerja, dilakukan pembersihan data dengan menghapus data yang kosong tersebut.

	A	B	C	D	E	F	G
1	program	pendapat	pekerjaan	pendapat	jumlah	ta	pendapat
2	452	3000000	20	2000000	3	202000	
3	452	3500000	3	900000	2	300000	
4	452	5000000	1	1000000	3	100000	
5	452	7500000	9	0	4	764391	
6	452	3000000	8	0	3	352242	
7	452	4707400	15	0	4	627723	
8	452	4500000	16	800000	2	100000	
9	452	2400000	20	0	3	100000	
10	452	4285800	15	0	4	255824	
11	452	1250000	2	150000	3	100000	
12	452	2000000	20	0	3	50000	
13	452	1000000	2	0	6	31000	
14	351	3621783	10	0	3	97931	
15	351	3000000	9	0	3	700000	

Gambar 5. Hasil Data setelah *Cleaning*

3.2.1 Data Normalization

Data yang diperoleh setelah diunduh berbentuk file dengan format .xlsx yang masih tidak beraturan, data yang seharusnya terdiri dari beberapa kolom masih terlihat menyatu dan harus dipisahkan menjadi beberapa kolom agar terlihat fitur-fitur yang akan digunakan.

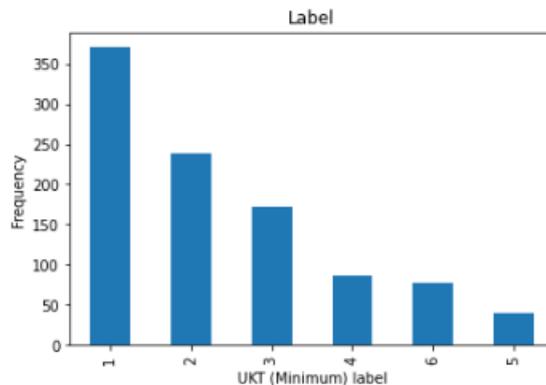
	A	B	C	D	E	F
1	program	pendapat	pekerjaan	pendapat	jumlah	ta
2	452	3000000	20	2000000	3	202000
3	452	3500000	3	900000	2	300000
4	452	5000000	1	1000000	3	100000
5	452	7500000	9	0	4	764391
6	452	3000000	8	0	3	352242
7	452	4707400	15	0	4	627723
8	452	4500000	16	800000	2	100000
9	452	2400000	20	0	3	100000
10	452	4285800	15	0	4	255824
11	452	1250000	2	150000	3	100000
12	452	2000000	20	0	3	50000
13	452	1000000	2	0	6	31000
14	351	3621783	10	0	3	97931
15	351	3000000	9	0	3	700000

Gambar 6. Hasil Data *Normalization*

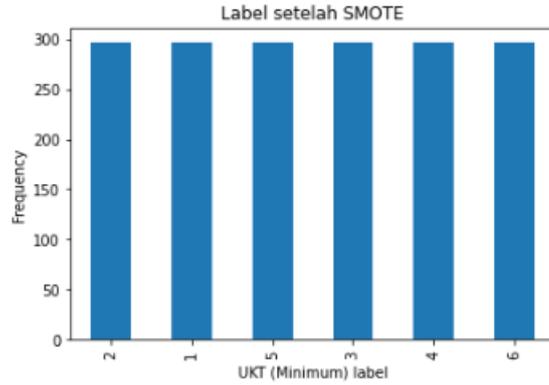
3.3. SMOTE

Hal pertama yang dilakukan setelah *import* dataset menggunakan *google colab* adalah menentukan kelas yang akan dilakukan *oversampling*. Pada penelitian kali ini fitur *class* memiliki kelas *imbalance* dimana *class* yang memiliki 6 *value* merupakan data minoritas, Setelah menentukan *class* akan dilanjutkan proses implementasi metode SMOTE pada data yang tidak seimbang.

Data kelas minoritas setelah *oversampling* menggunakan metode SMOTE memiliki presentasi jumlah data yang lebih baik yakni 50,60% dari total keseluruhan jumlah data. Hal itu terlihat pada Tabel 4.3 yang menunjukkan presentasi data sebelum menggunakan metode SMOTE yang menunjukkan jarak antara kelas minoritas dan mayoritas adalah 76,4%. Namun, setelah dilakukan *oversampling* jumlah perbedaan tersebut berhasil dipangkas menjadi hanya terpaat 1,50% . Data minoritas telah memiliki *instance* atau baris baru sebanyak 1776 baris data.



Gambar 7. Distribusi Fitur Class Sebelum SMOTE



Gambar 8. Distribusi Fitur Class Setelah SMOTE

3.4. Proses Klasifikasi

Pada Tabel 2 menunjukkan bahwa tiga dari lima model klasifikasi menyentuh angka akurasi yang cukup baik yakni di atas 90%, bahkan untuk model klasifikasi *random forest* memiliki nilai akurasi dengan 96,6% mendekati nilai sempurna. Hal tersebut merupakan peningkatan yang signifikan setelah dilakukannya proses penyeimbangan data menggunakan metode SMOTE.

Tabel 2. Performa Model Menggunakan metode SMOTE

Model	Akurasi	F1 Score	Precision	Recall
SVM	0.71	0.71	0.737	0.71
Decision Tree	0.81	0.807	0.808	0.808
Random Forest	0.901	0.901	0.902	0.901

3.5. Visualisasi dengan Streamlit

Dalam menampilkan hasil klasifikasi dilakukan dengan membuat sistem berbasis website. Sistem ini Tabel 1. Pengujian 5 *user* awal ditujukan sebagai wadah untuk menampilkan informasi/analisis secara visual mengenai pengelompokan *class ukt* berdasarkan label dalam rangka menentukan strategi untuk meningkatkan produktivitas *site*. Pembuatan sistem ini menggunakan streamlit sebagai *framework* yang berbahasa python. Sistem ini bersifat *public* dengan tujuan untuk memudahkan *user* dalam mengakses website.



Gambar 9. Tampilan Visualisasi Streamlit

4. KESIMPULAN

Berdasarkan hasil analisis penelitian yang dilakukan yaitu penanganan *imbalance* data menggunakan metode *smote* untuk melakukan *klasifikasi* dengan algoritma yang tepat, dapat disimpulkan bahwa :

1. Pada penelitian ini ini metode *Smote* mampu menangani masalah ketidakseimbangan (*imbalance*) data dengan melakukan proses sintesis data terdekat untuk memunculkan data baru hingga data menjadi seimbang.

2. Pada penelitian ini telah dilakukan perbandingan tiga model klasifikasi *Machine Learning* yaitu SVM, *Decision Tree* dan *Random Forest* untuk melakukan klasifikasi *ukt*. Model *Random Forest* mempunyai performasi yang signifikan dalam mengklasifikasikan *UKT* dengan rerata akurasi sebesar 89,9% Peningkatan yang signifikan pada akurasi *Random Forest* dengan nilai presentase peningkatan sebesar 14,43%.
3. Berhasil membuat sistem menggunakan Streamlit Python untuk menampilkan visualisasi chart berdasarkan hasil analisis yang diperoleh, dengan sumber data yang terhubung pada *Google Colab*

5. REFERENCES

- [1] R. M. Firzatullah, "Menggunakan Sistem Pendukung Keputusan Penentuan Uang Kuliah Tunggal Universitas XYZ Menggunakan Algoritma Backpropagation," *Petir*, vol. 14, no. 2, pp. 170–180, 2021, doi: 10.33322/petir.v14i2.996.
- [2] Y. D. Retnoningsih and A. Marom, "Analisis Kebijakan Penyelenggaraan Pendidikan Berbasis Uang Kuliah Tunggal Bagi Perguruan Tinggi Negeri Fakultas Ilmu Sosial Dan Ilmu Politik Universitas Diponegoro Semarang Jawa Tengah," *J. Econ. Educ.*, vol. 6, no. 2, 2016.
- [3] D. Wanto, Anjar, *Data Mining: Algoritma dan Implementasi - Books*. 2020.
- [4] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *Int. J. Adv. Soft Comput. its Appl.*, vol. 7, no. 3, 2015.
- [5] M. Mustaqim, B. Warsito, and B. Surarso, "Kombinasi Synthetic Minority Oversampling Technique (SMOTE) dan Neural Network Backpropagation untuk menangani data tidak seimbang pada prediksi pemakaian alat kontrasepsi implan," *Regist. J. Ilm. Teknol. Sist. Inf.*, 2019, doi: 10.26594/register.v5i2.1705.
- [6] M. Sulistiyono, Y. Pristyanto, S. Adi, and G. Gumelar, "Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi," *Sistemasi*, vol. 10, no. 2, p. 445, 2021, doi: 10.32520/stmsi.v10i2.1303.
- [7] W. Xie, G. Liang, Z. Dong, B. Tan, and B. Zhang, "An Improved Oversampling Algorithm Based on the Samples' Selection Strategy for Classifying Imbalanced Data," *Math. Probl. Eng.*, 2019, doi: 10.1155/2019/3526539.