

LAPORAN AKHIR
PENELITIAN HIBAH KOMPETITIF LPPM
PENELITIAN DASAR STUDI DALAM NEGERI



JUDUL PENELITIAN:
PENGEMBANGAN MODEL PENGELOMPOKAN UANG KULIAH
TUNGGAL BERBASIS ALGORITMA UNSUPERVISED LEARNING

PENGUSUL:
Wiyli Yustanti, S.Si, M.Kom/0003027708

UNIVERSITAS NEGERI SURABAYA
NOPEMBER 2022

HALAMAN PENGESAHAN

PENELITIAN DASAR STUDI DALAM NEGERI

Judul Penelitian : Pengembangan Model Pengelompokan Uang
Kuliah Tunggal Berbasis Algoritma *Unsupervised Learning*

Kode>Nama Rumpun Ilmu : 458/Teknik Informatika

Bidang Fokus Penelitian : Sains dan Teknologi

Ketua Peneliti

a. Nama Lengkap : Wiyli Yustanti, S.Si, M.Kom

b. NIDN : 0003027708

c. Jabatan Fungsional : Lektor Kepala

d. Program Studi : Sistem Informasi

e. No HP : 087854407576

f. Alamat Surel (e-mail) : wilyyustanti@unesa.ac.id

Anggota Peneliti (1)

a. Nama Lengkap : -

b. NIDN : -

c. Perguruan Tinggi : -

Lama Penelitian : 1 Tahun

Usulan Penelitian Tahun : 1

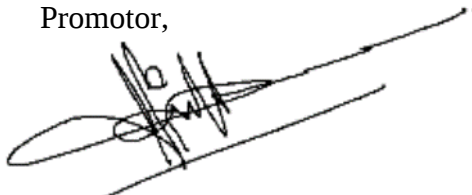
Biaya Penelitian Keseluruhan Rp. 30.000.000,00

Biaya Penelitian

- Diusulkan ke LPPM Unesa : Rp. 30.000.000,00

Mengetahui,

Promotor,



Prof. Drs. Nur Iriawan, MI.Komp, Ph.D
NIP. 196210151988031002

Ketua ,

Surabaya, 27 Nopember 2022



Wiyli Yustanti, S.Si, M.Kom
NIP. 197702032005012001

Mengetahui,
Ketua Lembaga Penelitian dan Pengabdian Masyarakat
Universitas Negeri Surabaya



Prof. Dr. Darni, M.Hum
NIP. 196509261990022001

RINGKASAN

Pendidikan tinggi sebagai bagian dari sistem pendidikan nasional memiliki peran strategis dalam mencerdaskan kehidupan bangsa dan memajukan ilmu pengetahuan dan teknologi dengan memperhatikan dan menerapkan nilai humaniora serta pembudayaan dan pemberdayaan bangsa Indonesia yang berkelanjutan. Dengan demikian untuk mewujudkan keterjangkauan dan pemerataan yang berkeadilan dalam memperoleh pendidikan tinggi yang bermutu dan relevan dengan kepentingan masyarakat demi kemajuan, kemandirian, dan kesejahteraan maka diperlukan penataan pendidikan tinggi yang terencana, terarah, dan berkelanjutan serta memperhatikan aspek demografis dan geografis. Berdasarkan tujuan tersebut, salah satu upaya sangat strategis yang telah dilakukan pemerintah Indonesia dalam melakukan penataan pendidikan tinggi di Indonesia yaitu dengan menerbitkan dan memberlakukan UU No. 12 Tahun 2012 mengenai Pendidikan Tinggi dimana salah satu bentuk implementasinya diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan Nomor: 073/2014 tentang Perubahan Atas Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 55/2013 mengenai Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri (PTN) di Lingkungan Kementerian Pendidikan Dan Kebudayaan. mengelompokkan mahasiswa berdasarkan pada kondisi sosial ekonomi mahasiswa

Penelitian terkait UKT sebenarnya sudah banyak dilakukan antara lain penelitian tentang analisis perbandingan metode data mining untuk kasus klasifikasi UKT. Dari semua penelitian yang ada terkait dengan UKT, fokus penelitiannya terletak pada akurasi hasil prediksi dalam melakukan klasifikasi kelas UKT baik dengan ukuran Akurasi, F1 Score dan ROC atau AUC. Semua metode diatas dilakukan dengan menggunakan metode supervised learning artinya memanfaatkan data UKT yang sudah berlabel kemudian dilakukan pemodelan yang bertujuan untuk memprediksi level UKT objek baru. Pada penelitian ini difokuskan untuk mencari struktur tersembunyi dari datas UKT yang kemudian dikelompokkan berdasarkan tingkat kemiripannya sehingga dapat ditemukan kelas UKT yang diharapkan dalam satu kelompok memiliki variansi yang minimum.

Kata Kunci : *Unsupervised Learning*, Data Kategorik, UKT, Akurasi, Klastering

PRAKATA

Alhamdulillah segala puji syukur kami panjatkan kepada Allah SWT yang telah memberikan kemudahan dalam melaksanakan penelitian ini. Penelitian ini merupakan eksplorasi terhadap data Uang Kuliah Tunggal (UKT) di Universitas Negeri Surabaya yang dianalisis dengan pendekatan machine learning untuk mendapatkan model yang paling tepat , baik untuk pihak mahasiswa, universitas maupun pemerintah.

Hasil penelitian ini diharapkan dapat dijadikan dasar untuk melakukan evaluasi kebijakan penentuan kelompok UKT yang selama ini digunakan dan dengan model yang dihasilkan dari penelitian ini dapat digunakan sebagai rekomendasi model yang lebih baik baik dari sisi ketepatan kelompok UKT dengan status sosial ekonomi, ketepatan terhadap proporsi UKT rendah dari pemerintah dan kesesuaian dengan pagu Universitas Negeri Surabaya.

Nopember, 2022

Peneliti

DAFTAR ISI

HALAMAN PENGESAHAN	ii
DAFTAR ISI	
RINGKASAN.....	3
PRAKATA.....	4
DAFTAR ISI.....	5
DAFTAR TABEL.....	7
DAFTAR GAMBAR.....	8
DAFTAR LAMPIRAN.....	9
BAB I.....	10
PENDAHULUAN	10
A. Latar Belakang	10
B. Rumusan Masalah	12
BAB II.....	13
TINJAUAN PUSTAKA	13
A. Algoritma Klustering.....	13
B. Evaluasi Klustering	15
C. Uang Kuliah Tunggal	19
BAB III	26
TUJUAN DAN MANFAAT PENELITIAN	26
A. Tujuan Penelitian.....	26
B. Manfaat Penelitian.....	26
BAB IV	27
METODE PENELITIAN.....	27
A. Kerangka Metode	27
1. Metode Penyelesaian Masalah Pertama	27
2. Metode Penyelesaian Masalah Kedua.....	28
3. Metode Penyelesaian Masalah Ketiga.....	28
4. Metode Penyelesaian Masalah Keempat.....	29
B. Sumber Data	30

1. Struktur Data	31
2. Variabel Penelitian	32
3. Definisi Operasional.....	35
BAB V	37
HASIL DAN PEMBAHASAN.....	37
A. Evaluasi Struktur Pengelompokan UKT Eksisting	37
B. Metode Klastering untuk Perbaikan Pengelompokan UKT	39
1) Agglomerative Clustering	39
2) BIRCH.....	40
3) DBSCAN.....	42
4) K-Means	43
5) Mini-Batch K-Means.....	44
6) Mean Shift.....	44
C. Pelabelan Kelompok UKT berbasis Klastering	46
D. Pemodelan Prediksi Kelompok UKT	47
BAB VI	51
KESIMPULAN DAN SARAN.....	51
A. Kesimpulan.....	51
B. Saran.....	51
DAFTAR PUSTAKA	52
Lampiran 1. Biodata Peneliti	54
Lampiran 2. Artikel Jurnal	60

DAFTAR TABEL

Tabel 2. 1 Indeks Kemahalan untuk Wilayah.....	24
Tabel 4.1 Jumlah Data Penelitian	30
Tabel 4.2 Struktur Variabel Penelitian	31
Tabel 4.3 Statistik Deskriptif Penetapan Biaya UKT menurut Permenristekdikti No 22 tahun 2015 (dalam ribuan).....	31
Tabel 4.4 Sebaran Jumlah Mahasiswa Berdasarkan Kelompok Uang Kuliah Tunggal (UKT) Non Bidikmisi Jalur SNMPTN dan SBMPTN Tahun 2017-2021	32
Tabel 4.5 Variabel Penelitian	32
Tabel 5.1 Hasil Rotasi PCA.....	46
Tabel 5.2 Perhitungan Rata-Rata F1- Score	48
Tabel 5.2 Uji Statistik Non Parametrik untuk Uji Beda F1-Score	49
Tabel 5.3 Perhitungan Rata-Rata Nilai AUC	49
Tabel 5.4 Uji Statistik Non Parametrik untuk Uji Beda AUC.....	49

DAFTAR GAMBAR

Gambar 2.1 Diagram Indeks Validitas Klustering.....	16
Gambar 2.2 UKT Mahasiswa pada Program Studi	22
Gambar 4.1 Kerangka Metode Penelitian.....	27
Sedangkan metode yang digunakan untuk menyelesaikan setiap masalah penelitian dapat dijelaskan dalam sub bab berikut.....	27
Gambar 4.2 Skema proses relabelling hasil klustering dalam kelompok UKT.....	29
Gambar 5.1 Scree Plot dari Dataset	37
Gambar 5.3 Scatter UKT hasil Pengelompokan dengan Metode Agglomerative	40
Gambar 5.4 Scatter UKT hasil Pengelompokan dengan Metode BIRCH.....	41
Gambar 5.5 Scatter UKT hasil Pengelompokan dengan Metode DBSCAN.....	42
Gambar 5.6 Scatter UKT hasil Pengelompokan dengan Metode K-Means	43
Gambar 5.7 Scatter UKT hasil Pengelompokan dengan Metode Mini Batch K-Means	44
Gambar 5.8 Scatter UKT hasil Pengelompokan dengan Metode Mean Shift	45
Gambar 5.9 Scatter UKT hasil Pengelompokan dengan Relabelling K-Means	47
Gambar 5.10 Perbandingan Nilai F1-Score data Asli dengan Data Relabel	48
Gambar 5.11 Perbandingan Nilai AUC data Asli dengan Data Relabel	50

DAFTAR LAMPIRAN

Lampiran 1 . Biodata Peneliti

Lampiran 2. Artikel Seminar Internasional

Lampiran 3. Artikel Jurnal Internasional

BAB I

PENDAHULUAN

A. Latar Belakang

Dalam undang-undang No 12 Tahun 2012 tentang pendidikan tinggi dijelaskan bahwa pendidikan tinggi sebagai bagian dari sistem pendidikan nasional memiliki peran strategis dalam mencerdaskan kehidupan bangsa dan memajukan ilmu pengetahuan dan teknologi dengan memperhatikan dan menerapkan nilai humaniora serta kebudayaan dan pemberdayaan bangsa Indonesia yang berkelanjutan. Dengan demikian untuk mewujudkan keterjangkauan dan pemerataan yang berkeadilan dalam memperoleh pendidikan tinggi yang bermutu dan relevan dengan kepentingan masyarakat demi kemajuan, kemandirian, dan kesejahteraan maka diperlukan penataan pendidikan tinggi yang terencana, terarah, dan berkelanjutan serta memperhatikan aspek demografis dan geografis. Berdasarkan tujuan tersebut, salah satu upaya sangat strategis yang telah dilakukan pemerintah Indonesia dalam melakukan penataan pendidikan tinggi di Indonesia yaitu dengan menerbitkan dan memberlakukan UU No. 12 Tahun 2012 mengenai Pendidikan Tinggi dimana salah satu bentuk implementasinya diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan Nomor: 073/2014 tentang Perubahan Atas Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 55/2013 mengenai Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri (PTN) di Lingkungan Kementerian Pendidikan Dan Kebudayaan. Tujuan utama dari peraturan ini adalah agar pendidikan tinggi mudah terjangkau dan merata bagi seluruh masyarakat Indonesia serta meringankan beban mahasiswa terhadap pembiayaan pendidikan.

Berpedoman pada tujuan utama pemberlakuan UKT, dalam penelitian ini telah dilakukan survey pendahuluan secara umum terkait implementasi pembayaran UKT di sebuah PTN-BLU di Surabaya. Dari data UKT yang didapatkan mulai tahun 2016 – 2019 untuk jalur SNMPTN dan SBMPTN diluar penerima bidikmisi, terdapat jumlah cukup besar dari data mahasiswa yang melakukan banding penurunan UKT karena keberatan pembayaran yang dibuktikan dengan kondisi kemampuan sosial ekonomi orang tua mereka. Data jumlah pengajuan banding penurunan UKT selama 5 tahun berkisar antara 3% sampai 10%. Dengan adanya fenomena terjadinya pengajuan penurunan UKT ini, maka bagi pihak universitas akan cukup berpengaruh terhadap

target penerimaan negara bukan pajak (PNBP) sehingga harus membuat kebijakan yang dapat menjadi solusi bagi kedua belah pihak. Data bidikmisi tidak diikuti dalam penelitian ini dikarenakan data sosial ekonomi yang masuk ke dalam database universitas tidak digunakan dalam proses penentuan UKT karena biaya kuliah sudah ditanggung pemerintah.

Sementara di sisi lain, pemerintah melalui Peraturan Menteri Riset , Teknologi dan Pendidikan Tinggi No 90 Tahun 2017 yang kemudian diperbaharui dengan Peraturan Menteri Pendidikan dan Kebudayaan No 6 Tahun 2020 tentang Penerimaan Mahasiswa Baru Program Sarjana pada PTN menyebutkan dalam pasal 9 ayat 1 bahwa PTN wajib mencari dan menjangring calon mahasiswa yang memiliki potensi akademik tinggi, tetapi kurang mampu secara ekonomi dan calon mahasiswa dari daerah terdepan, terluar, dan tertinggal untuk diterima paling sedikit 20% dari seluruh mahasiswa baru yang diterima dan tersebar pada semua Program Studi. Dengan demikian , berdasarkan kondisi diatas didapatkan tiga prinsip penting dalam pengelompokan UKT. Pertama, pengelompokan UKT harus berdasarkan tingkat kemampuan sosial ekonomi orang tua mahasiswa sehingga mahasiswa mendapatkan rasa keadilan dan aman dalam menempuh pendidikannya. Kedua, perguruan tinggi harus mencapai target pendapatan negara bukan pajak (PNBP) melalui UKT semaksimal mungkin. Selanjutnya, yang ketiga adalah bahwa PTN sebagai lembaga pemerintahan memiliki tanggung jawab sosial untuk menerima minimal 20% mahasiswa dari keluarga kurang mampu. Ketiga prinsip ini menjadi latar belakang utama dalam penelitian ini sehingga dibutuhkan sebuah pengembangan metode pengelompokan UKT yang lebih presisi dalam mengelompokkan mahasiswa berdasarkan pada kondisi sosial ekonomi mahasiswa

Penelitian terkait UKT sebenarnya sudah banyak dilakukan antara lain penelitian tentang analisis perbandingan metode data mining untuk kasus klasifikasi UKT (Utomo et al., 2018) yang membandingkan kinerja algoritma Naïve Bayes (NB), Decision Tree (DT) dan SVM. Dalam penelitian tersebut akurasi klasifikasi untuk algoritma DT lebih unggul 1% dibanding algoritma SVM. Penelitian yang lain untuk klasifikasi UKT juga ditunjukkan dengan algoritma berbeda antara lain metode NB (Anas et al., 2019; Yanifa & Candra, 2018), K-Mean atau C-Mean dan KNN (Arianti, 2017; Jazuli, 2016; Muchsin & Sudarma, 2017; Muhammad, 2016; Zurraedah, 2018, DT(Karim et al., 2017), DT an NB (Abidin et al., 2020), Simple Additive Weighting (Tahir, 2016; Testiana, 2017), SVM (Suyoga et al., 2018); Yustanti et al., 2018), TOPSIS dan AHP (Prasetyanti & Listyaningrum, 2017) dan Multi Objective Optimization on the Basis of Ratio Analysis (Rokhman et al., 2017). Dari semua penelitian yang ada terkait dengan UKT, fokus

penelitiannya terletak pada akurasi hasil prediksi dalam melakukan klasifikasi kelas UKT baik dengan ukuran Akurasi, F1 Score dan ROC atau AUC. Semua metode diatas dilakukan dengan menggunakan metode supervised learning artinya memanfaatkan data UKT yang sudah berlabel kemudian dilakukan pemodelan yang bertujuan untuk memprediksi level UKT objek baru. Pada penelitian ini difokuskan untuk mencari struktur tersembunyi dari datas UKT yang kemudian dikelompokkan berdasarkan tingkat kemiripannya sehingga dapat ditemukan kelas UKT yang diharapkan dalam satu kelompok memiliki variansi yang minimum.

B. Rumusan Masalah

Adapun rumusan masalah yang ingin diselesaikan dalam penelitian ini adalah :

1. Bagaimana kebaikan struktur pengelompokan UKT dengan metode eksisting ?
2. Bagaimana melakukan pengelompokan UKT dengan struktur yang lebih baik dengan menggunakan metode *unsupervised learning*?
3. Bagaimana melakukan pelabelan kelompok UKT yang memiliki kesesuaian dengan status social ekonomi mahasiswa?
4. Bagaimana melakukan prediksi kelompok UKT berbasis hasil algoritma *unsupervised learning*?

BAB II

TINJAUAN PUSTAKA

A. Algoritma Klastering

Pembelajaran *unsupervised learning* mengacu pada penggunaan algoritma kecerdasan buatan (AI) untuk mengidentifikasi pola dalam kumpulan data yang berisi data yang tidak diklasifikasikan atau diberi label. Dengan demikian, algoritma ini memiliki fungsi untuk mengklasifikasikan, memberi label, dan/atau mengelompokkan titik data yang terkandung dalam kumpulan data tanpa memiliki panduan eksternal apa pun dalam melakukan tugas itu. Dengan kata lain, *unsupervised learning* memungkinkan sistem untuk mengidentifikasi pola dalam kumpulan data sendiri. Dalam *unsupervised learning*, sistem AI akan mengelompokkan informasi yang tidak disortir menurut kemiripan dan perbedaan meskipun tidak ada kategori yang disediakan. Algoritma *unsupervised learning* dapat melakukan tugas pemrosesan yang lebih kompleks daripada *supervised learning*.

Banyak algoritma pengelompokan (klastering) tradisional baik hirarki maupun non-hirarki yang sebagian besar dirancang untuk data numerik dengan menggunakan beberapa cara menghitung matriks jarak untuk mengukur kemiripan (*similarity*) antara dua titik data. Penggunaan pengukuran berbasis jarak ini dapat menghasilkan hasil yang memuaskan untuk atribut numerik, akan tetapi tidak menghasilkan hasil yang memuaskan untuk kumpulan data dengan atribut kategorik. Hal ini disebabkan karena ukuran kemiripan antara objek kategorik sulit untuk diukur akibat kurang-nya keteraturan hubungan antara nilai kategorik itu sendiri (Zheng dkk., 2020). Berikut akan dijelaskan gambaran umum metode klastering yang sudah dikembangkan pada penelitian-penelitian sebelumnya.

1. Metode Hirarki

Metode ini digunakan untuk mencari struktur pengelompokan dari objek-objek. Selanjutnya hasil pengelompokannya disajikan secara hierarki atau ber-jenjang. Metode hirarki ini terdiri dari dua pendekatan yaitu:

a) Penggabungan (*Agglomerative*)

Metode ini dimulai dengan masing-masing objek diasumsikan sebagai klaster yang terpisah, dan berturut-turut menggabungkan kelompok menurut ukuran jarak. Pengelompokan dapat

berhenti ketika semua objek berada dalam satu kelompok atau pada jumlah klaster tertentu yang dipilih.

b) Pemecahan (*Devise*).

Metode ini mengikuti strategi yang berlawanan dengan cara penggabungan. Caranya dimulai dengan diasumsikan bahwa semua objek merupakan satu kelompok kemudian secara bertahap membagi kelompok menjadi yang ukuran kelompok yang lebih kecil, sampai setiap objek jatuh dalam satu klaster, atau sampai jumlah klaster yang diinginkan tercapai.

Metode hirarki yang sudah dikembangkan untuk menangani masalah klastering data kategorik adalah algoritma *Robust Clustering using linKs* (ROCK). Guha dkk. (1999) mengusulkan pendekatan baru berdasarkan konsep *link* antara objek data. Gagasan yang diusulkan membantu mengatasi masalah yang muncul dari penggunaan jarak *Euclidean* untuk data input yang merupakan pengidentifikasi dari nilai-nilai kategorik. Data yang mempunyai tingkat hubungan (*link*) tinggi akan digabungkan ke dalam satu klaster, sedang yang mempunyai tingkat hubungan (*link*) yang kecil akan dipisahkan dari klaster dimana data tersebut dikelompokkan. Kemudian tahun berikutnya dikembangkan algoritma Quick-ROCK (QROCK) oleh Dutta dkk.(2005) untuk mengatasi masalah ketika tidak ada *link* antar objek.

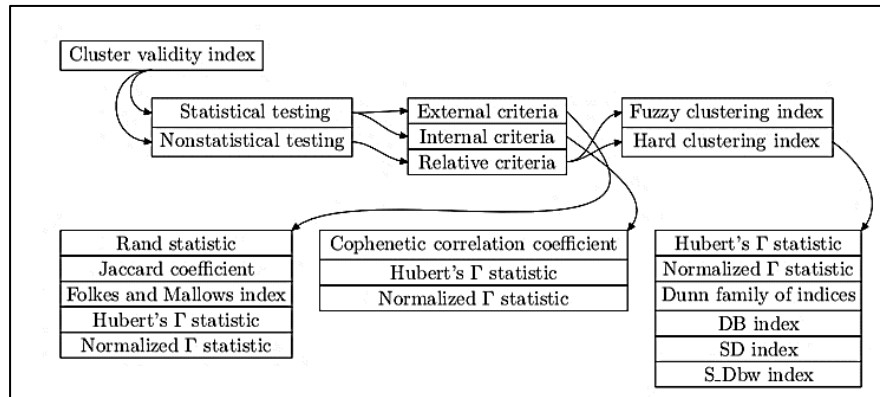
2. Metode Non-Hirarki (*partition or centroid based*)

Metode klastering berbasis partisi yang paling populer adalah *K-Means*. Metode ini merupakan pendekatan non hirarki yang dapat digunakan untuk mempartisi objek ke dalam kelompok-kelompok berdasarkan kedekatan karakteristik, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu klaster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam klaster yang lain. Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang di set dalam proses pengelompokan, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu klaster dan memaksimalkan variasi antar klaster. Jika diberikan sekumpulan objek maka algoritma *K-Means* akan mempartisi X dalam k buah klaster, setiap klaster memiliki *centroid* dari objek-objek dalam klaster tersebut. Pada tahap awal algoritma *K-Means*, dipilih secara acak k buah objek sebagai *centroid*, kemudian jarak antara objek dengan *centroid* dihitung dengan menggunakan jarak *Euclidian*, objek ditempatkan dalam klaster yang terdekat dihitung dari titik tengah klaster. *centroid* baru ditetapkan jika semua objek sudah ditempatkan dalam klaster terdekat. Proses penentuan *centroid* dan penempatan objek dalam klaster diulangi sampai nilai *centroid* konvergen (centroid dari semua

klaster tidak berubah lagi). Titik pusat yang digunakan dalam metode *K-Means* merupakan nilai rata-rata dari variabel-variabel dalam satu kelompok (Johnson & Wichern, 2007). Untuk kasus domain bertipe kategorik telah diusulkan sebuah pengembangan dari *K-Means* yang disebut dengan metode *K-Modes* (Huang, 1997). Secara umum, prinsip kerja algoritma *K-Modes* tidak berbeda dengan *K-Means*, akan tetapi terdapat tiga perbedaan dasar sebagai bentuk modifikasi dari *K-Means* yaitu menggunakan formulasi pengukuran *similarity* yang berbeda, mengganti k rata-rata (*means*) dengan k modus (*modes*), serta menggunakan metode berbasis frekuensi untuk memperbarui nilai modus pada iterasi berikutnya. Beberapa tahun kemudian, algoritma *K-Modes* banyak dikembangkan oleh peneliti lain dan salah satunya adalah algoritma *K-Distribution* yang diusulkan oleh Cai dkk.(2007). Prinsip kerja dari algoritma *K-Distribution* hampir sama dengan *K-Means* dan *K-Modes*. Jika diasumsikan bahwa $D(X_1, X_2, \dots, X_n)$ adalah kumpulan data kategorik yang terdiri dari n objek kategorik dan $A_1, A_2, \dots, A_m$ merupakan atribut kategorik dari setiap objek X , maka objek kategorik X diwakili oleh vektor $\langle a_1, a_2, \dots, a_m \rangle$, di mana a_i adalah nilai atribut A_i . Pertama-tama, dilakukan partisi semua objek menjadi subset K *nonempty* dan saling eksklusif secara acak, dan setiap subset diperlakukan sebagai sebuah klaster. Kemudian, untuk setiap objek $\langle a_1, a_2, \dots, a_m \rangle$, dihitung probabilitas gabungan masing-masing cluster $P(a_1, a_2, \dots, a_m)$ dan selanjutnya dapat ditetapkan objek tersebut akan masuk kedalam klaster yang memiliki probabilitas gabungan (*joint probability*) maksimal. Langkah-langkah tersebut diulangi sampai konvergen.

B. Evaluasi Klustering

Secara umum, ada tiga kriteria mendasar untuk menyelidiki validitas hasil dari klustering, yaitu kriteria eksternal, kriteria internal, dan kriteria relatif (Gan dkk., 2007). Dua pendekatan pertama melibatkan pengujian statistik dan komputasi, sedangkan yang ketiga yaitu kriteria relatif tidak melibatkan pengujian statistik.



Gambar 2.1 Diagram Indeks Validitas Klastering

Berdasarkan Gambar 1, dimana pengujian secara statistik dilakukan untuk kriteria eksternal dan internal, maka pada bagian ini hanya membahas dua kriteria tersebut. Konsep dasar dari pengukuran validitas klastering sebenarnya untuk mengetahui apakah data berasal dari distribusi yang random atau tidak (Halkidi dkk., 2002). Untuk menguji hipotesis tersebut dilakukan simulasi *monte carlo* yang membutuhkan prosedur komputasi yang kompleks. Oleh karena itu digunakan pendekatan eksternal dan internal sebagai alat ukur validitas klastering.

1. Kriteria Eksternal

Kriteria validitas eksternal bertujuan untuk mengukur seberapa baik hasil pengelompokan sesuai dengan pengetahuan sebelumnya tentang data. Diasumsikan bahwa informasi sebelumnya tidak dapat dihitung dari $\mathbf{X}$. Mungkin bentuk informasi eksternal yang paling umum digunakan adalah kelas (kategori) dan label kelas untuk objek yang berkaitan dengan $\mathbf{X}$. Informasi ini biasanya diperoleh melalui klasifikasi manual. Jadi pada prinsipnya kriteria eksternal adalah indeks yang dirancang untuk mengukur kesamaan antara dua partisi dimana hanya memperhitungkan distribusi titik-titik dalam kelompok yang berbeda dan tidak digunakan untuk mengukur kualitas distribusi ini (Desgraupes, 2017). Terdapat dua pendekatan yang bisa dilakukan, pertama adalah mengevaluasi struktur pengelompokan $\mathbf{S}$ yang dihasilkan, dengan membandingkannya dengan partisi independen dari data $\mathbf{T}$ yang dibangun sesuai dengan intuisi atau informasi sebelumnya tentang struktur pengelompokan dari data. Kemudian, pendekatan yang kedua adalah membandingkan matriks *proximity* $\mathbf{T}$ dengan matriks partisi $\mathbf{T}$ (Halkidi dkk., 2002). Misalkan $S = \{S_1, \dots, S_m\}$ merupakan stuktur klastering populasi dari himpunan data $\mathbf{X}$ dan

$T = \{T_1 \dots T_g\}$ menyatakan partisi dari data X . Jika sebuah pasangan titik dinyatakan dengan (x_s, x_t) maka terdapat empat kemungkinan kondisi ,yaitu:

- *SS* : Apabila kedua titik pengamatan berada pada klaster yang sama baik pada struktur klaster S maupun kelompok partisi T .
- *SD* : Apabila kedua titik pengamatan berada pada klaster yang sama dalam struktur klaster S tetapi berbeda kelompok partisi T .
- *DS* : Apabila kedua titik pengamatan berada pada klaster yang berbeda pada struktur klaster S tetapi berada pada kelompok yang sama dalam partisi T
- *DD* : Apabila kedua titik pengamatan berada pada klaster yang berbeda baik pada struktur klaster S maupun kelompok partisi T .

Diasumsikan bahwa a, b, c dan d adalah jumlah kejadian untuk setiap keadaan *SS*, *SD*, *DS* dan *DD* berturut-turut, maka $a + b + c + d = U$ merupakan jumlah maksimum dari pasangan data sehingga dapat diartikan bahwa $U = \frac{n(n-1)}{2}$ dimana n merupakan jumlah total titik pengamatan.

Selanjutnya dapat didefinisikan pengukuran indeks kemiripan antara struktur klaster S dan partisi T .

$$Rand\ statistic : \frac{a+d}{U}$$

$$Jaccard\ Coefficient : \frac{a}{a+b+c}$$

$$Folkes\ and\ Mallows\ (FM)\ index : \frac{a}{\sqrt{u_1 u_2}} = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}}$$

$$\text{dimana : } u_1 = a / (a+b) \text{ dan } u_2 = a / (a+c)$$

Ketiga indeks diatas memiliki nilai antara 0 sampai dengan 1. Semakin tinggi (mendekati 1) nilai indeksnya maka semakin mirip struktur klaster antara S dan T . Indeks kriteria eksternal yang lain adalah *Hubert's Γ statistic* dan *Normalized Γ statistic*, masing-masing dengan formulasi sebagai berikut :

$$Hubert's\ \Gamma statistic = \Gamma = \frac{1}{U} \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_{ij} Y_{ij}$$

$$\text{Normalized's } \Gamma \text{ statistic} = \hat{\Gamma} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (X_{ij} - \mu_x)(Y_{ij} - \mu_y)}{U \sigma_x \sigma_y}$$

Interpretasi dari dua indeks statistik Γ dan $\hat{\Gamma}$ adalah semakin tinggi nilainya maka semakin tinggi kemiripan antara matriks $\mathbf{X}$ dan $\mathbf{Y}$. X_{ij} dan Y_{ij} adalah elemen ke- i,j dari matriks $\mathbf{X}$, $\mathbf{Y}$ yang dibandingkan. Sedangkan μ_x, μ_y, σ_x dan σ_y berturut-turut adalah mean dan varians dari matriks $\mathbf{X}$ dan $\mathbf{Y}$. Range nilai Γ adalah $[0,1]$ sedangkan untuk $\hat{\Gamma}$ adalah $[-1,1]$.

2. Kriteria Internal

Tujuan dari kriteria internal adalah untuk mengevaluasi struktur pengelompokan yang dihasilkan oleh suatu algoritma klustering melalui jumlah dan fitur yang diwarisi dari kumpulan data (Halkidi dkk., 2002). Untuk menerapkan kriteria internal, ada dua situasi: (a) skema pengelompokan hirarki (metode klustering hirarki) dan (b) skema pengelompokan non-hirarki (metode berbasis partisi/titik pusat). Gagasan untuk memvalidasi skema hirarki pengelompokan adalah dengan menggunakan apa yang disebut matriks *cophenetic* $\mathbf{T}_c$ dan kemudian menggunakan koefisien korelasi *cophenetic* untuk mengukur derajat kemiripan antara $\mathbf{T}_c$ dan matriks *proximity* $\mathbf{T}$. Matriks *cophenetic* $\mathbf{T}_c$ didefinisikan sedemikian rupa sehingga elemen $\mathbf{T}_c(i,j)$ mewakili tingkat kedekatan antara dua titik X_i dan X_j yang ditemukan di sekitar kluster yang sama untuk pertama kalinya. Koefisien korelasi *cophenetic* dapat didefinisikan sebagai :

$$CPCC = \frac{(1/U) \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_{ij} h_{ij} - \mu_t \mu_{tc}}{\sqrt{\left((1/U) \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_{ij}^2 - \mu_t^2 \right) \left((1/U) \sum_{i=1}^{n-1} \sum_{j=i+1}^n h_{ij}^2 - \mu_{tc}^2 \right)}}$$

Dimana :

$$U = \frac{n(n-1)}{2} \quad ; \quad \mu_t = \frac{1}{U} \sum_{i=1}^{n-1} \sum_{j=i+1}^n q_{ij} \quad ; \quad \mu_{tc} = \frac{1}{U} \sum_{i=1}^{n-1} \sum_{j=i+1}^n h_{ij}$$

Sedangkan q_{ij} dan h_{ij} merupakan elemen ke- i,j dari matriks $\mathbf{T}$ dan $\mathbf{T}_c$ berturut-turut. Nilai koefisien korelasi *cophenetic* adalah $[-1,1]$ dan semakin tinggi nilai CPCC mengindikasikan

tingkat kesamaan antara matrik T dan T_c semakin besar. Dalam penelitian ini, digunakan kriteria internal yaitu Koefisien Silhouette dari setiap titik.

$$\text{Silhouette Coefficient}(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.1)$$

di mana,

$a(i)$: Jarak rata-rata i dari semua titik lain dalam cluster.

$b(i)$: Jarak rata-rata terkecil dari titik i ke semua titik di cluster lain

Interpretasi dari hasil penghitungan Koefisien Silhouette rata-rata adalah:

- a) Jika nilai $S(i)$ mendekati 0 berarti titik tersebut berada di antara dua cluster
- b) Jika nilai $S(i)$ semakin mendekati -1, maka sebaiknya diletakkan di cluster lain
- c) Jika $S(i)$ mendekati 1, maka titik tersebut termasuk ke dalam cluster yang tepat

C. Uang Kuliah Tunggal

Penetapan kebijakan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri (PTN) di Indonesia berawal dari realita di lapangan yang menunjukkan banyaknya pungutan kepada mahasiswa selama masa studi di perguruan tinggi. Untuk itu, diperlukan suatu standar biaya di perguruan tinggi untuk memberikan kepastian pembayaran bagi mahasiswa. Biaya perkuliahan yang ditanggung oleh mahasiswa tersebut juga harus disesuaikan dengan kemampuan ekonomi dari mahasiswa.

1. Kebijakan UKT

Pada tahun 2013 pemerintah melalui Kementerian Pendidikan dan Kebudayaan mengeluarkan kebijakan Uang Kuliah Tunggal (UKT) bagi Perguruan Tinggi Negeri (PTN). Selain dilatarbelakangi faktor munculnya pungutan diluar SPP, penetapan kebijakan UKT merupakan wujud pelaksanaan amanat Pasal 88 Undang-Undang Nomor 12 Tahun 2012 tentang Perguruan Tinggi untuk menetapkan standar satuan biaya operasional bagi PTN dalam menentukan biaya yang ditanggung mahasiswa. Dalam pasal tersebut juga disebutkan bahwa biaya yang dimaksud harus disesuaikan dengan kemampuan ekonomi mahasiswa, orang tua mahasiswa, maupun pihak lain yang membiayainya. Berdasarkan hal tersebut maka, tarif UKT pada PTN

dibuat secara berjenjang dan dibagi dalam beberapa golongan. Hal tersebut berfungsi agar mahasiswa dengan keadaan ekonomi rendah selayaknya mendapat tarif UKT rendah, begitupun mahasiswa dengan yang mampu secara ekonomi, sepantasnya mendapat UKT tinggi untuk mensubsidi mahasiswa UKT rendah tersebut. Berdasarkan bahan Konferensi Pers Uang Kuliah Tunggal (UKT) Permendikbud No.55 Tahun 2013, Tanggal 23 Mei 2013, prinsip dasar dalam perhitungan Penetapan Bantuan Operasional Perguruan Tinggi Negeri (BOPTN), Biaya Kuliah Tunggal (BKT) dan Uang Kuliah Tunggal (UKT) menggunakan prinsip dasar sebagai berikut:

“Uang kuliah yang ditanggung oleh mahasiswa diusahakan semakin lama semakin kecil dengan memperhatikan masyarakat yang tidak mampu (afirmasi), subsidi silang (yang kaya mensubsidi yang miskin), dan pengendalian biaya yang tepat”

Untuk menjamin keakuratannya, maka setiap tahun kebijakan penetapan BOPTN, BKT, dan UKT akan dievaluasi dan diperbaiki disesuaikan dengan kondisi terkini. Berdasarkan Surat edaran Dirjen Dikti Nomor: 272/E1.1/KU/2013 menjelaskan bahwa kelompok UKT harus memenuhi persyaratan :

- c) Tarif UKT sebaiknya dibagi atas 5 kelompok, dari yang paling rendah (kelompok 1) sampai yang paling tinggi (kelompok 5);
- d) Tarif UKT kelompok yang paling rendah (kelompok 1) rentangnya yang bisa dijangkau oleh masyarakat tidak mampu (misal: kuli bangunan, tukang becak. misal Rp. 0,- s.d. Rp 500.000;
- e) Paling sedikit ada 5% dari total mahasiswa yang diterima membayar UKT kelompok 1;
- f) Kelompok 3 s.d. 5 masing-masing membayar UKT sesuai dengan kemampuan ekonominya, dimana kelompok 5 merupakan kelompok dengan UKT tertinggi sesuai dengan program studi masing-masing;
- g) Paling sedikit ada 5% dari total mahasiswa yang diterima membayar UKT kelompok 2 dengan rentang Rp. 500.000.- s.d. Rp. 1.000.000,-.

Kemudian pada tahun 2014 terbit permendikbud no 73 tahun 2014 tentang Perubahan Atas Peraturan Menteri Pendidikan Dan Kebudayaan Nomor 55 Tahun 2013 Tentang BKT dan UKT. Secara esensial, Permendikbud No. 73 Tahun 2014 tidaklah berbeda dengan Permendikbud sebelumnya No. 55 Tahun 2013, bahkan pada perubahan tersebut merupakan penyempurnaan untuk terus menaikkan biaya pendidikan tinggi. Perubahan terjadi dengan penetapan kelompok

UKT yang asalnya 1 sampai 5 menjadi 1 sampai 8 dengan masing-masing PTN diberi wewenang untuk menentukan besarnya biaya UKT untuk setiap kelompok.

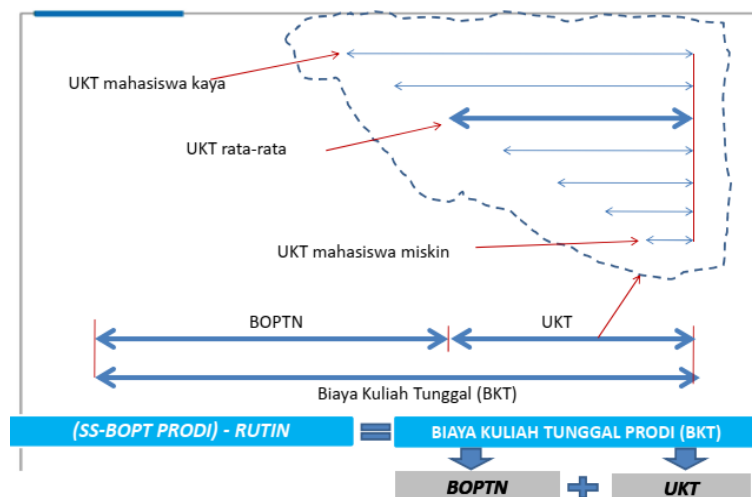
Pada tahun berikutnya, terbit Permendikbud No. 22 Tahun 2015 yang memberikan arahan tentang perbedaan besaran UKT untuk mahasiswa angkatan 2013, 2014 dan 2015 di seluruh PTN di Indonesia dengan tetap memperhatikan proporsi minimal 5% untuk masing-masing UKT kelompok 1 dan 2 yang termasuk dalam kelompok UKT golongan rendah. Kebijakan UKT selanjutnya diteruskan dibawah Kementrian Riset, Teknologi, dan Pendidikan Tinggi dalam Permenristekdikti Nomor 39 Tahun 2016 tentang Biaya Kuliah Tunggal (BKT) dan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri. Dalam peraturan tersebut dijelaskan bahwa, BKT adalah keseluruhan biaya operasional yang terkait langsung dengan proses pembelajaran mahasiswa per semester pada program studi di PTN. Sementara itu, pengertian UKT adalah sebagian BKT yang ditanggung setiap mahasiswa berdasarkan kemampuan ekonominya. Kebijakan ini diimplementasikan oleh PTN dalam bentuk penetapan tarif dan golongan UKT, penentuan penerima UKT, penetapan ulang UKT terhadap mahasiswa, serta larangan untuk tidak memungut uang pangkal dan/atau pungutan lain, dan mekanisme penyampaian laporan realisasi penerimaan UKT.

Dalam permenristekdikti No. 39 tersebut, juga disebutkan bahwa jumlah mahasiswa penerima UKT kelompok I dan kelompok II serta mahasiswa penerima bidikmisi diterapkan kepada paling sedikit 20% (dua puluh persen) dari seluruh mahasiswa baru yang diterima di setiap PTN dan tersebar pada semua Program Studi. Dalam peraturan ini juga diberikan kebebasan untuk PTN memungut biaya selain UKT untuk mahasiswa yang diterima diluar jalur SNMPTN dan SBMPTN dengan proporsi maksimal 30% dari total jumlah mahasiswa baru. Permenristekdikti No 39 tahun 2016 selanjutnya diperbarui dengan Permenristekdikti No 39 tahun 2017 ditambahkan dengan pasal tentang kewenangan PTN untuk memberikan keringanan biaya UKT apabila terjadi (1) ketidaksesuaian kemampuan ekonomi mahasiswa yang diajukan oleh mahasiswa, orang tua mahasiswa, atau pihak lain yang membiayainya; dan/atau (2) perubahan data kemampuan ekonomi mahasiswa, orang tua mahasiswa, atau pihak lain yang membiayainya. Permenristekdikti tentang UKT ini terus diperbarui dengan keluarnya Permenristekdikti No 91/M/KPT/2018 dan selanjutnya pada tahun 2019, pemerintah melalui surat edaran Menristekdikti No. B/416/M/PR.03.04/2019 mengatur pungutan uang pangkal atau pungutan lain selain UKT dengan memperhatikan keadaan

sosial ekonomi mahasiswa, orang tua mahasiswa atau pihak yang membiayainya. Selain itu untuk mahasiswa yang secara ekonomi tidak mampu, tidak boleh dikenakan uang pangkal dan tarif uang pangkal yang dikeluarkan PTN harus tetap memperhatikan prinsip kewajaran, proporsional dan berkeadilan.

2. Perhitungan Biaya UKT

Uang Kuliah Tunggal atau UKT adalah biaya kuliah yang ditanggung setiap mahasiswa setiap satu semester. Fungsi dari UKT adalah memberikan keringanan bagi mahasiswa yang kurang mampu melalui sistem subsidi silang pemerintah yang besarnya didasarkan pada kondisi ekonomi dan sosial keluarga mahasiswa. Dengan kata lain, UKT adalah sistem pembayaran yang ditujukan untuk memberikan pemerataan biaya kuliah sesuai dengan kemampuan ekonomi masing-masing orang tua atau wali mahasiswa. Pada sistem UKT, semua biaya kuliah berupa uang gedung, SPP, dana praktikum, dan biaya penunjang perkuliahan lainnya telah dilebur menjadi satu dan dibagi rata dalam delapan semester. Jika mahasiswa telah membayar UKT di satu semester, mereka tidak perlu dikenakan biaya tambahan lainnya. Sehingga, mereka hanya membayar biaya perkuliahan di setiap awal semester, yaitu UKT.



Gambar 2.2 UKT Mahasiswa pada Program Studi

Pada Gambar 2 dapat dilihat bahwa kebijakan UKT ini berimplikasi pada besaran biaya kuliah yang dibayarkan oleh setiap mahasiswa disebuah program studi akan berbeda-beda tergantung pada kemampuan ekonomi orang tua mahasiswa atau pihak yang membiayainya.

Dalam perhitungan besarnya UKT terlebih dahulu perlu dihitung biaya kuliah tunggal (BKT) yang merupakan keseluruhan biaya operasional per mahasiswa per semester pada program studi di perguruan tinggi negeri. Selanjutnya biaya kuliah tunggal digunakan sebagai dasar penetapan biaya yang dibebankan kepada mahasiswa masyarakat dan pemerintah. Uang kuliah tunggal sebagaimana dimaksud pada ayat ditetapkan berdasarkan biaya kuliah tunggal dikurangi biaya yang ditanggung oleh pemerintah. Sehingga formulasi besaran UKT adalah :

$$UKT = BKT - BOPTN$$

BOPTN merupakan bantuan biaya dari pemerintah yang diberikan pada perguruan tinggi negeri untuk membiayai kekurangan biaya operasional sebagai akibat tidak adanya kenaikan sumbangan pendidikan (SPP) di perguruan tinggi negeri. Pada awal mula munculnya kebijakan UKT tahun 2013. Perhitungan BKT dilakukan dengan formulai berikut :

$$\begin{aligned} \text{Biaya Kuliah Tunggal (BKT)} &= C \times K_1 \times K_2 \times K_3 \\ &= 5.080.000 \times K_1 \times K_2 \times K_3 \end{aligned}$$

Dalam rumus tersebut, BKT terdiri dari empat komponen yaitu :

C : Konstanta yaitu nilai rata-rata uang kuliah per semester seluruh PTN. Nilai konstanta adalah Rp5.080.000,00;

K_1 : Indeks program studi, program studi yang indeksinya tinggi antara lain kedokteran, sedangkan program studi yang indeknya standar adalah misal ilmu sosial. Pehitungan indeks program studi ditentukan oleh tiga komponen yaitu akreditasi program studi, akreditasi institusi dan akreditasi internasional.

K_2 : Indeks kualitas atau mutu Perguruan Tinggi, sebagai contoh UI dan ITB memiliki indeks mutu sebesar 1,5 , sedangkan UGM dan IPB memiliki indeks sebesar 1,3 dan untuk perguruan tinggi yang lain diberikan indeks 1.

K_3 : Indeks kemalahan wilayah tempat Perguruan Tinggi berada, misalnya PTN di Jawa, Bali dan NTB memiliki indeks kemahalan 1.00 , untuk PTN di wilayah Sumatera memiliki indeks 1.05, Kalimantan, Sulawesi, NTT memiliki indeks 1.15 dan yang berada Maluku, Papua sebesar 1.30.

Akan tetapi, formulasi yang diberlakukan pada tahun 2013 tersebut dianggap terlalu umum, dimana seharusnya indeks mutu didapat dari data empiris, bukan dari variabel mutu serta pengelompokan program studi terlalu umum. Kemudian pada tahun 2014 dilakukan update perubahan perhitungan BKT dengan memperhitungkan biaya-biaya yang berpengaruh pada komponen BKT. Cara menghitung biaya kuliah tunggal mahasiswa sebuah prodi adalah dengan menjumlahkan biaya langsung dengan biaya tidak langsung yang dibutuhkan mahasiswa dalam menjalani proses perkuliahan selama delapan semester. Diasumsikan bahwa besaran biaya tidak langsung adalah 50 % dari biaya langsung. Sehingga formulasi BKT yang terbaru menjadi :

$$\text{BKT} = \text{BL} + \text{BTL} = \text{BL} + 0,5\text{BL} = 1,5\text{BL}$$

Dari formulasi diatas, yang dimaksud biaya langsung (BL) adalah biaya operasional satuan yang terkait langsung dengan penyelenggaraan kurikulum program studi. BL dihitung secara cukup detail pada level aktivitas, yang didasari atas asumsi pemenuhan atas standar nasional perguruan tinggi (SNPT) , serta memperhatikan praktek baik (*best practices*) yang selama ini sudah berjalan. Sedangkan biaya tidak langsung (BTL) adalah biaya operasional satuan yang tidak secara langsung terkait dengan penyelenggaraan kurikulum program studi namun mutlak diperlukan dalam pengelolaan institusi pendidikan tinggi dalam rangka mendukung penyelenggaraan program studi. Kemudian formulasi BKT diatas dikoreksi dengan faktor kemahalan dan indeks kualitas.

Tabel 2. 1 Indeks Kemahalan untuk Wilayah

Wilayah	Indeks Kemahalan
Jawa, Bali, NTB	1,00
Sumatera	1,05
Kalimantan, Sulawesi, NTT	1,15
Maluku, Papua	1,30

Dengan menggunakan data indeks kemahalan didapatkan nilai BKT terbobot (BKT_w) yaitu :

$$\text{BKT}_w = \text{BKT} \times \text{indeks kemahalan}$$

Selanjutnya nilai BKT terbobot dikalikan dengan faktor kualitas program studi, dimana komponen kualitas dipengaruhi oleh kualitas institusi dan kualitas internasional yang dimiliki. Secara umum, indeks kualitas ditentukan dengan rumus sebagai berikut :

$$\text{indeks kualitas} = 1 + \text{APS} + \text{AIPT} + \text{AI}$$

dimana :

APS : Akreditasi Program Studi (APS) merupakan nilai dari akreditasi program studi dengan konversi akreditasi A diberikan nilai 0,15, akreditasi B diberikan nilai 0,1 dan akreditasi C diberikan nilai 0,05.

AIPT : Akreditasi Institusi Perguruan Tinggi (AIPT) merupakan nilai dari akreditasi perguruan tinggi dengan konversi akreditasi A diberikan nilai 0,15, akreditasi B diberikan nilai 0,1 dan akreditasi C diberikan nilai 0,05.

AI : Akreditasi Internasional (AI) merupakan nilai dari akreditasi program studi level internasional dengan konversi akreditasi A diberikan nilai 0,15, akreditasi B diberikan nilai 0,1 dan akreditasi C diberikan nilai 0,05.

Sehingga besaran BKT final (BKT_q) yang digunakan untuk menghitung UKT dapat dirumuskan sebagai perkalian antara BKT terbobot (BKT_w) dengan indeks kualitas.

$$BKT_q = BKT_w \times \text{indeks kualitas}$$

Untuk menjamin keakuratannya, maka pemerintah menghimbau agar kebijakan penetapan BOPTN, BKT, dan UKT dievaluasi dan diperbaiki setiap tahun untuk disesuaikan dengan kondisi terkini.

Penetapan kebijakan Uang Kuliah Tunggal (UKT) pada Perguruan Tinggi Negeri (PTN) di Indonesia berawal dari realita di lapangan yang menunjukkan banyaknya pungutan kepada mahasiswa selama masa studi di perguruan tinggi. Untuk itu, diperlukan suatu standar biaya di perguruan tinggi untuk memberikan kepastian pembayaran bagi mahasiswa. Biaya perkuliahan yang ditanggung oleh mahasiswa tersebut juga harus disesuaikan dengan kemampuan ekonomi dari mahasiswa.

BAB III

TUJUAN DAN MANFAAT PENELITIAN

A. Tujuan Penelitian

Adapun tujuan penelitian ini adalah :

1. Mendapatkan hasil evaluasi struktur pengelompokan UKT dari metode eksisting (kebijakan universitas)
2. Mendapatkan struktur pengelompokan UKT dengan metode unsupervised learning dari data social ekonomi mahasiswa dengan hasil yang lebih baik.
3. Mengembangkan metode pelabelan dari label acak yang dihasilkan oleh metode unsupervised learning dengan kelompok UKT riil.
4. Membandingkan hasil akurasi dari prediksi pengelompokan UKT dari data eksisting dan data hasil pelabelan.

B. Manfaat Penelitian

Sedangkan manfaat yang didapatkan dari hasil penyelesaian masalah adalah dapat digunakan sebagai alternatif metode penentuan UKT bagi mahasiswa baru dengan hasil yang lebih baik.

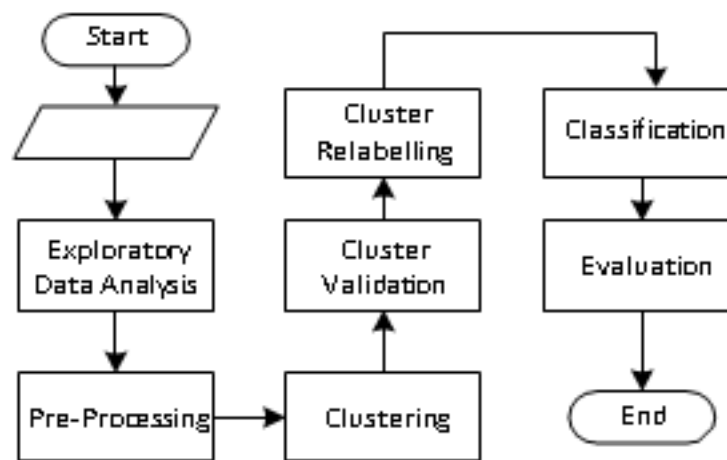
BAB IV

METODE PENELITIAN

Pada bagian ini akan diuraikan langkah-langkah penelitian untuk menyelesaikan permasalahan dalam penelitian serta mencapai tujuan penelitian. Secara umum, pendekatan teori yang dilakukan berkaitan dengan konsep *machine learning* baik *unsupervised learning* dilanjutkan dengan *supervised learning*.

A. Kerangka Metode

Secara umum tahapan yang dilakukan pada penelitian ini dapat dilihat pada Gambar 4.1.



Gambar 4.1 Kerangka Metode Penelitian

Sedangkan metode yang digunakan untuk menyelesaikan setiap masalah penelitian dapat dijelaskan dalam sub bab berikut.

1. Metode Penyelesaian Masalah Pertama

Permasalahan pertama bertujuan untuk melakukan evaluasi struktur kelompok eksisting. Dalam penelitian ini, digunakan kriteria internal yaitu Koefisien Silhouette dari setiap titik.

$$\text{Silhouette Coefficient}(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4.1)$$

di mana,

$a(i)$: Jarak rata-rata dari semua titik lain dalam cluster.

$b(i)$: Jarak rata-rata terkecil dari titik i ke semua titik di cluster lain

Interpretasi dari hasil penghitungan Koefisien Silhouette rata-rata adalah:

- a) Jika nilai $S(i)$ mendekati 0 berarti titik tersebut berada di antara dua cluster
- b) Jika nilai $S(i)$ semakin mendekati -1, maka sebaiknya diletakkan di cluster lain
- c) Jika $S(i)$ mendekati 1, maka titik tersebut termasuk ke dalam cluster yang tepat

2. Metode Penyelesaian Masalah Kedua

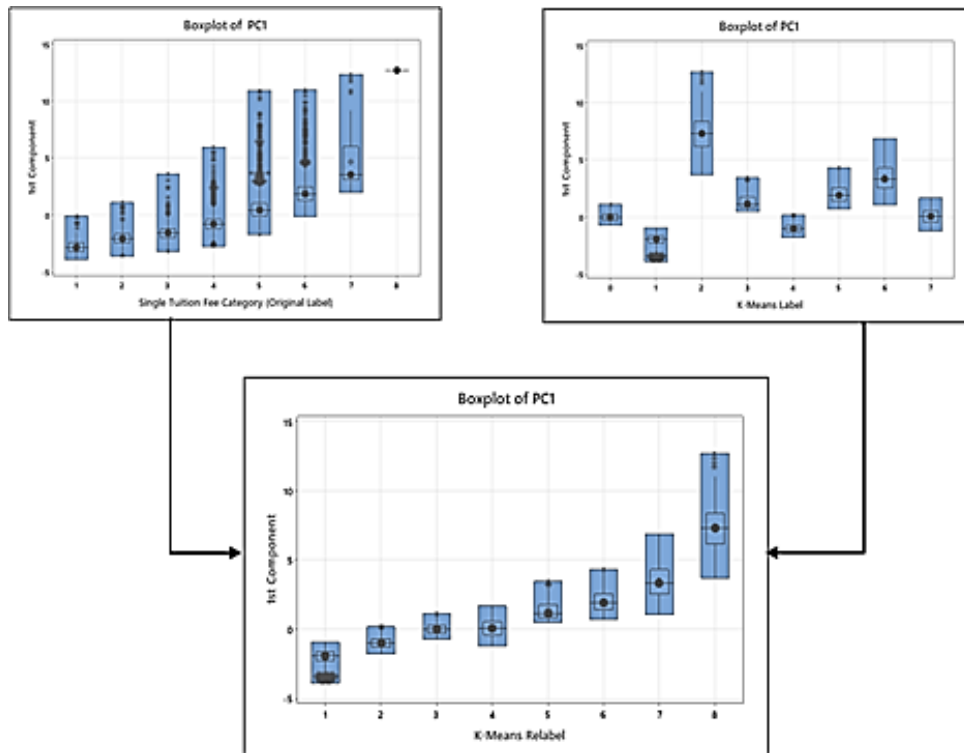
Untuk menyelesaikan permasalahan kedua, dilakukan dengan langkah-langkah berikut :

- a) Melakukan preprocessing terhadap dataset antara lain melakukan transformasi dengan menggunakan ordinal encoding, melakukan standarisasi nilai variabel input serta melakukan reduksi dimensi menggunakan Principle Component Analysis (PCA)
- b) Mengimplementasikan berbagai macam jenis algoritma unsupervised learning antara lain :
 - 1) *Agglomerative Clustering*
 - 2) *BIRCH*
 - 3) *DBSCAN*
 - 4) *K-Means*
 - 5) *Mini-Batch K-Means*
 - 6) *Mean Shift*
- c) Menghitung indeks Silhouette
- d) Merekomendasi hasil klastering yang paling baik

3. Metode Penyelesaian Masalah Ketiga

Permasalahan terkait dengan proses relabelling pada tahapan ini menggunakan pendekatan selang kepercayaan dari hasil PCA. Kerangka proses yang diusulkan dapat dilihat pada gambar 4.2. Pada proses pelabelan ulang, digunakan nilai confidence interval dari komponen PCA pertama. Langkah-langkah yang dilakukan dalam proses pelabelan ulang ini adalah sebagai berikut:

- Menghitung Confidence Interval (CI) komponen pertama PCA dengan tingkat kepercayaan 95% baik untuk label data asli maupun label data K-Means.
- Buatlah Box Plot untuk setiap CI.
- Memetakan pola Boxplot dari label random hasil klastering ke label asal.
- Perbarui label yang dihasilkan oleh algoritma klastering



Gambar 4.2 Skema proses relabelling hasil klastering dalam kelompok UKT

4. Metode Penyelesaian Masalah Keempat

Selanjutnya hasil klastering dan pelabelan ulang yang sudah diterapkan pada data pengelompokan UKT digunakan untuk pemodelan prediksi dengan menggunakan algoritma klasifikasi (*supervised learning*) dengan Langkah-langkah sebagai berikut :

- Data yang sudah ditransformasi ke numerik kemudian dibagi data *training* dan testing untuk setiap jenis dataset menggunakan pendekatan *q-fold cross validation*. Dalam percobaan ini diujicobakan nilai $q = 3, 5, 10$, karena data ujicoba berdistribusi *imbalanced class*, maka pendekatan *cross validation* yang dilakukan adalah *stratified cross validation*.

- b) Melakukan prediksi klasifikasi dengan menggunakan metode SVM *multiclass* dengan pendekatan *One-Versus-One* (OVO) dan *One-Versus-All* (OVA) sebagai perbandingan.
- c) Fungsi kernel yang diujicobakan dalam tahap ini adalah *Radial Basis Function* (RBF).
- d) Optimasi untuk mendapatkan nilai parameter terbaik menggunakan metode *GridSearch*.
- e) Melakukan evaluasi klasifikasi dengan menghitung nilai rata-rata F1 Score, dan AUC untuk setiap metode *encoding* yang berbeda sebagaimana pada tabel 3.2.
- f) Melakukan *benchmarking* terhadap hasil kinerja metode klasifikasi lain yaitu *Decision Tree* (DT), *K-Nearest Neighbour* (KNN), Regresi Logistik (LR), *Naïve Bayes* (NB), dan *Random Forest* (RF).
- g) Membuat kesimpulan metode *encoding* yang direkomendasi sebagai *pre-processing* untuk SVM *Multiclass* dilihat dari nilai F1 Score dan AUC.

B. Sumber Data

Tujuan dari penelitian ini adalah memprediksi kelompok UKT setiap mahasiswa yang sesuai dengan kondisi sosial ekonomi mereka tanpa melanggar peraturan kebijakan pemerintah terkait proporsi jumlah mahasiswa di kelompok UKT level terendah, selain itu juga mampu memenuhi target penerimaan negara bukan pajak (PNBP) dari pembayaran UKT mahasiswa yang diterima di jalur SNMPTN dan SBMPTN. Dengan demikian perlu dilakukan kajian empiris yang menggunakan data riil. Sumber data dalam penelitian ini adalah data sosial ekonomi orang tua mahasiswa baru yang diterima di sebuah perguruan tinggi negeri BLU di Surabaya melalui jalur SNMPTN dan SBMPTN pada periode tahun penerimaan 2017 sampai 2019. Total jumlah data yang diperoleh adalah 12.372 record dengan rincian dapat dilihat seperti Tabel 4.1.

Tabel 4.1 Jumlah Data Penelitian

Tahun Masuk	Jalur SNMPTN	Jalur SBMPTN	Penerima Bidikmisi	Total
(1)	(2)	(3)	(4)	(5) = (2) + (3) – (4)
2017	1524	2631	1069	3086
2018	1547	2707	727	3527
2019	1371	2592	985	2978
Total			2781	9591
			12372	

Adapun terkait dengan kebutuhan kajian empiris, data yang digunakan khusus data mahasiswa yang membayar UKT diluar penerima program beasiswa bidikmisi. Sehingga data yang digunakan selama 3 tahun adalah 9.591 mahasiswa dengan rincian tahun 2017 sebanyak 3086 mahasiswa, tahun 2018 sebanyak 3527 mahasiswa dan tahun 2019 sebanyak 2978 mahasiswa.

1. Struktur Data

Adapun struktur data yang akan digunakan dalam penelitian ini dapat dilihat pada tabel 4.2.

Tabel 4.2 Struktur Variabel Penelitian

Mahasiswa	Variabel Penelitian					
	X_1	X_2	...	X_{12}	X_{13}	Y
1	X_{11}	X_{21}	...	X_{121}	X_{131}	y_1
2	X_{12}	X_{21}	...	X_{122}	X_{132}	y_2
3	X_{13}	X_{23}	...	X_{123}	X_{133}	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{1n}	X_{2n}	...	X_{12n}	X_{13n}	y_n

Struktur table diatas menggambarkan bahwa terdapat satu variabel respon Y yang bertipe kategorik dan 13 variabel prediktor ($X_1, X_2, \dots, X_{13}$) yang juga memiliki tipe kategorik. Variabel respon merupakan variabel target dari model klasifikasi yang terdiri dari 8 kategori yang menunjukkan kelompok UKT. Gambaran umum tentang besaran UKT yang ditetapkan berdasarkan Permenristekdikti Nomor 22 Tahun 2015 Tentang Biaya Kuliah Tunggal Dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri Di Lingkungan Kementerian Riset, Teknologi, Dan Pendidikan Tinggi dapat dilihat pada Tabel 4.3 Permenristek ini diberlakukan sejak tahun 2016 sampai sekarang.

Tabel 4.3 Statistik Deskriptif Penetapan Biaya UKT menurut Permenristekdikti No 22 tahun 2015 (dalam ribuan)

Ukuran	K1	K2	K3	K4	K5	K6	K7	K8
MEAN	500	1000	2400	3479,3	4553,7	5619,7	6680,7	7733,3
MINIMUM	500	1000	2400	3120	3840	4560	5280	6000
MAKSIMUM	500	1000	2400	4140	5880	7620	9360	11100
STDEV	0	0	0	307,7	616,2	932,8	1245,6	1569,5

Adapun sebaran jumlah mahasiswa per angkatan mulai tahun 2017 sampai 2021 berdasarkan kelompok UKT dapat dilihat pada Tabel 4.4 .

Tabel 4.4 Sebaran Jumlah Mahasiswa Berdasarkan Kelompok Uang Kuliah Tunggal (UKT)
Non Bidikmisi Jalur SNMPTN dan SBMPTN Tahun 2017-2021

Tahun	K1	K2	K3	K4	K5	K6	K7	K8	Total
2017	15	99	280	1197	1308	36	0	0	2935
2018	13	98	244	1320	1429	44	0	0	3146
2019	38	200	274	879	1958	1134	239	29	4751
2020	30	211	86	1112	964	355	35	1	2794
2021	36	72	311	1122	1201	709	39	0	3490

2. Variabel Penelitian

Adapun variabel-variabel yang digunakan dalam penelitian ini merupakan variabel-variabel yang mencerminkan karakteristik sosial ekonomi orang tua mahasiswa yang diterima disebuah perguruan tinggi negeri BLU di Surabaya mulai tahun 2017 sampai tahun 2021 melalui jalur SNMPTN dan SBMPTN diluar penerima program beasiswa bidikmisi. Variabel yang menjadi target adalah variabel kelompok UKT (Y). Analisis kalsifikasi menggunakan variabel X_1 samapai X_{13} . Sedangkan unit pengamatan dalam penelitian ini adalah data sosial ekonomi orang tua mahasiswa. Variabel-variabel tersebut disajikan pada Tabel 4.5

Tabel 4.5 Variabel Penelitian

Variabel	Keterangan	Skala Pengukuran	Tipe Data
Y	Kelompok UKT K1 = 500.000 K2 = 1.000.000 K3 = 2.400.000 K4 = 3.120.000 - 4.140.000 K5 = 3.840.000 – 5.880.000 K6 = 4.560.000 – 7.620.000 K7 = 5.280.000 – 9.360.000 K8 = 6.000.000 – 11.100.000	Ordinal	Kategorik

Variabel	Keterangan	Skala Pengukuran	Tipe Data
X ₁	Jenis Pekerjaan Ayah P0 = tidak bekerja/meninggal/sakit P1 = Pensiunan Swasta/Buruh Tani/Pabrik/Asongan/Nelayan P2 = Penisunan PNS/Guru/ABRI/ Gol I-II/ Pedagang Kecil P3 = Petani/ Penisunan PNS/Guru Gol III/Pegawai BUMN/ Nelayan Pemilik Kapal Kecil	Ordinal	Kategorik
	P4 = Dosen/PNS Gol IV/Kabag BUMN/ Pemilik Toko/Pengusaha Menengah/ Nelayan Pemilik Kapal Sedang P5 = Guru Besar/Pejabat BUMN/Pengusaha Besar/perwira		
X ₂	Jenis Pekerjaan Ibu P0 = tidak bekerja/meninggal/sakit P1= Pensiunan Swasta/Buruh Tani/Pabrik/Asongan/Nelayan P2= Penisunan PNS/Guru/ABRI/ Gol I-II/ Pedagang Kecil P3 = Petani/ Penisunan PNS/Guru Gol III/Pegawai BUMN/ Nelayan Pemilik Kapal Kecil P4 = Dosen/PNS Gol IV/Kabag BUMN/ Pemilik oko/Pengusaha Menengah/ Nelayan Pemilik Kapal Sedang 5= Guru Besar/Pejabat BUMN/ Pengusaha Besar/perwira	Ordinal	Kategorik
X ₃	Gaji Ayah G1 = 0 – 1.000.000 G2 = 1.000.000 – 2.500.000 G3 = 2.500.000 – 3.500.000 G4 = 3.500.000 – 7.500.000 G5 = > 7.500.000	Ordinal	Kategorik
X ₄	Gaji Ibu G1 = 0 – 1.000.000 G2 = 1.000.000 – 2.500.000 G3 = 2.500.000 – 3.500.000 G4 = 3.500.000 – 7.500.000	Ordinal	Kategorik

Variabel	Keterangan	Skala Pengukuran	Tipe Data
	G5 = > 7.500.000		
X ₅	Jumlah Tanggungan T1 = >=5 orang T2 = 4 orang T3 = 3 orang T4 = 2 orang T5 = 1 orang	Ordinal	Kategorik
X ₆	Pembayaran PBB B1 = < 60.000 B2 = 60.000 – 75.000 B3 = 75.000 – 90.000 B4 = 90.000 – 100.000 B5 = > 100.000	Ordinal	Kategorik
X ₇	Tagihan Listrik L1 = < 50.000 L2 = 50.000 – 75.000 L3 = 75.000 – 100.000 L4 = 90.000 – 100.000 L5 = > 100.000	Ordinal	Kategorik
X ₈	Kepemilikan Mobil M0 = Tidak Punya M1 = Punya dan tahun kendaraan lebih dari 10 tahun M2= Punya dan tahun kendaraan kurang dari 10 tahun	Ordinal	Kategorik
X ₉	Kepemilikan Motor D0 = Tidak Punya D1 = 1 Motor D2 = 2 Motor D3 = > 2 motor	Ordinal	Kategorik
X ₁₀	Sumber Penerangan S1 = NON-PLN S2 = PLN	Ordinal	Kategorik
X ₁₁	Kepemilikan Rumah R1 = Kos R2 = Menumpang/Kontrakan R3 = Milik Sendiri/Rumah Dinas	Ordinal	Kategorik
X ₁₂	Pendapatan Ibu Lainnya	Ordinal	Kategorik

Variabel	Keterangan	Skala Pengukuran	Tipe Data
	G1 = 0 - 100.000 G2 = 500.000 – 2.000.000 G3 = 2.000.000 – 5.000.000 G4 = 5.000.000 – 7.000.000 G5 = > 7.000.000		
X ₁₃	Pendapatan Ayah Lainnya G1 = 0 - 100.000 G2 = 500.000 – 2.000.000 G3 = 2.000.000 – 5.000.000 G4 = 5.000.000 – 7.000.000 G5 = > 7.000.000	Ordinal	Kategorik

3. Definisi Operasional

Adapun definisi operasil variabel penelitian adalah :

a. Kelompok UKT (Y)

Kelompok UKT adalah Kategori UKT menurut Peraturan Menteri Riset Teknologi dan Pendidikan Tinggi No 2 Tahun 2015 Lampiran III halaman 111-114 yang terdiri dari 8 kelas UKT yang dimulai dari kelas terendah (K1) yang memiliki nilai nominal 500.000 per semester dan kelas tertinggi (K8) dengan maksimal nominalnya adalah 11.100.000 per semester.

b. Pekerjaan Ayah (X₁)

Pekerjaan Ayah adalah pencaharian atau kegiatan bekerja yang dilakukan oleh ayah dan dijadikan sebagai sumber penghasilan utama untuk mendapat nafkah dalam memenuhi kebutuhan pokok keluarga.

c. Pekerjaan Ibu (X₂)

Pekerjaan Ibu adalah pencaharian atau kegiatan bekerja yang dilakukan oleh ibu dan dijadikan sebagai sumber penghasilan utama .

d. Gaji Ayah (X₃)

Gaji Ayah adalah upah kerja atau penghasilan yang diterima oleh ayah karena pekerjaan atau usaha sehingga dibayarkan dalam waktu yang tetap.

e. Gaji Ibu (X₄)

Gaji Ibu adalah upah kerja atau penghasilan yang diterima oleh ibu karena pekerjaan atau usaha sehingga dibayarkan dalam waktu yang tetap.

f. Jumlah Tanggungan Keluarga (X₅)

Jumlah tanggungan keluarga adalah banyaknya beban yang menjadi tanggungjawab kepala rumah tangga baik secara moral maupun materiil,

g. Pembayaran PBB (X_6)

Pembayaran PBB adalah rata-rata nominal pembayaran Pajak Bumi dan Bangunan (PBB) yang tertera pada Surat Pemberitahuan Pajak Terhutang (SPPT) PBB selama 3 tahun terakhir yang menjadi tanggungan untuk dibayarkan oleh kepala rumah tangga.

h. Tagihan Listrik (X_7)

Tagihan Listrik adalah rata-rata nominal biaya yang harus dibayarkan untuk kebutuhan listrik selama 3 bulan terakhir oleh kepala rumah tangga .

i. Kepemilikan Mobil (X_8)

Kepemilikan Mobil adalah jumlah mobil yang dimiliki oleh keluarga dan menjadi tanggungan kepala rumah tangga dalam perawatan operasionalnya.

Kepemilikan Motor (X_9)

j. Kepemilikan Motor adalah jumlah motoryang dimiliki oleh keluarga dan menjadi tanggungan kepala rumah tangga dalam perawatan operasionalnya.

k. Sumber Penerangan (X_{10})

Sumber penerangan adalah alat utama didalam keluarga yang dijadikan sebagai sumber utama dalam penghasilkan energi cahaya/penerangan.

l. Kepemilikan Rumah (X_{11})

Kepemilikan Rumah adalah status hunian yang saat ini ditinggali oleh keluarga

m. Pendapatan Ibu Lainnya (X_{12})

Pendapatan Ibu Lainnya adalah perolehan upah atau tambahan penghasilan diluar gaji pokok atau upah utama ibu setiap bulannya.

n. Pendapatan Ayah Lainnya (X_{13})

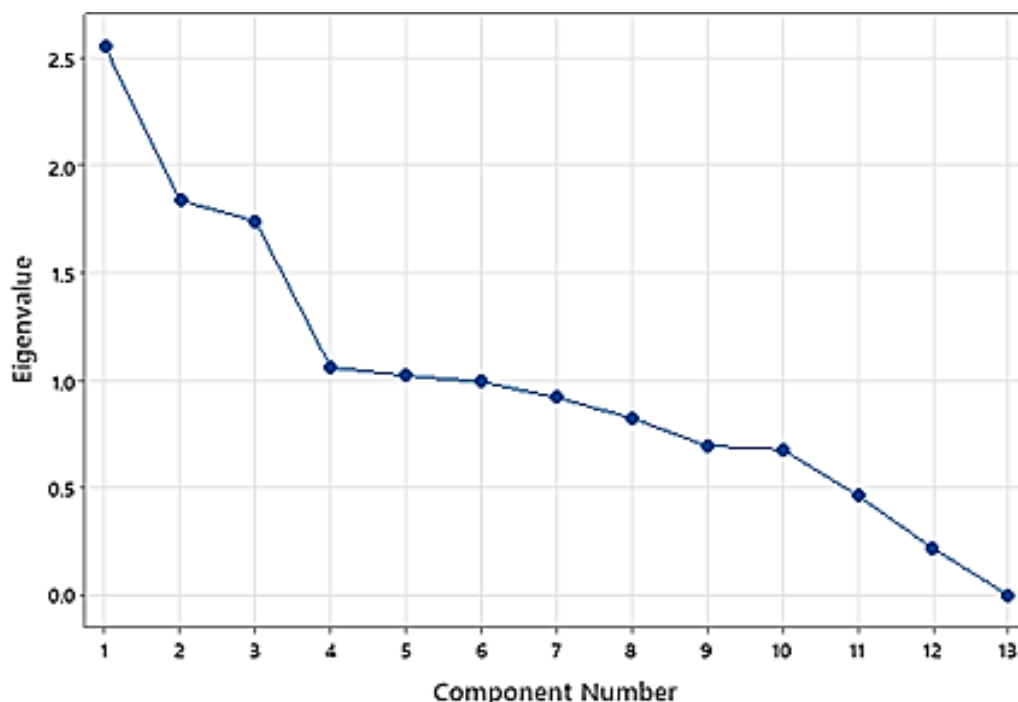
Pendapatan Ayah Lainnya adalah perolehan upah atau tambahan penghasilan diluar gaji pokok atau upah utama ayah setiap bulannya.

BAB V

HASIL DAN PEMBAHASAN

A. Evaluasi Struktur Pengelompokan UKT Eksisting

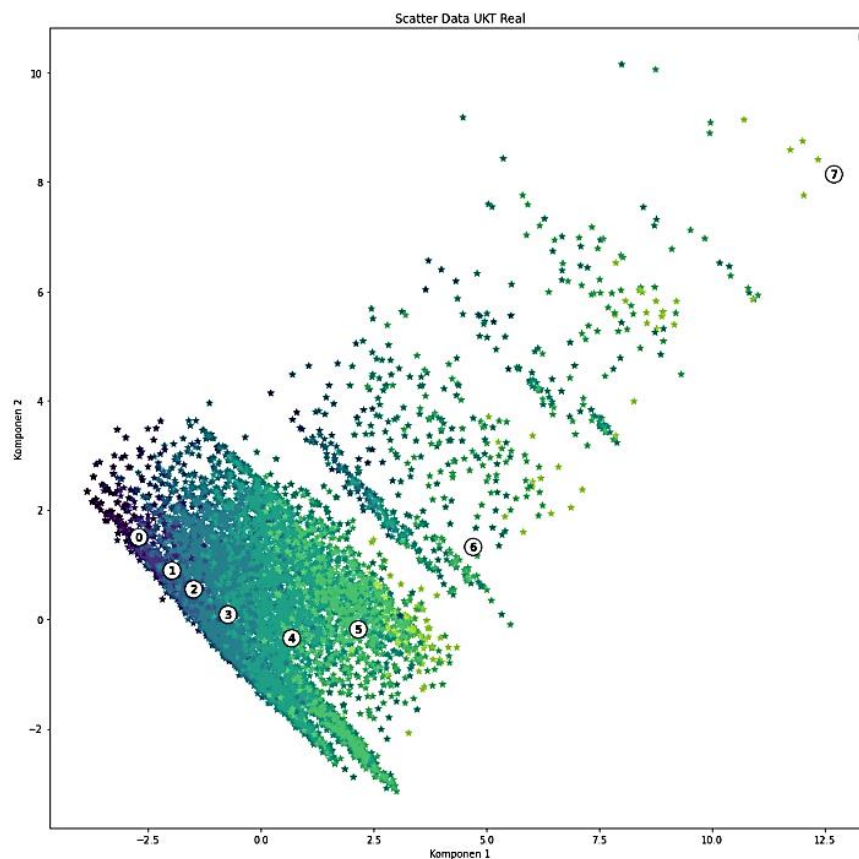
Selama fase pra-pemrosesan, data asli ditransformasikan menggunakan pengkodean ordinal dan kemudian divalidasi menggunakan uji statistik Kaiser-Meyer-Olki (KMO) dan Bartlett untuk menentukan kelayakan penerapan PCA. Bartlett's Test of Sphericity bertujuan untuk menguji hipotesis bahwa variabel-variabel tersebut tidak berkorelasi dalam populasi (H_0). Jika nilai p kecil, kita dapat mengesampingkan hipotesis nol dan setuju bahwa setidaknya ada satu pasang variabel yang berkorelasi dalam data, maka PCA direkomendasikan. Hasil perhitungan nilai Chi-Square dan p -value untuk Uji Bartlett adalah 601308.601 dan 0,000. Oleh karena itu, proses PCA dapat dilanjutkan. Sedangkan KMO adalah ukuran Sampling Adequacy (MSA), indeks yang menilai keakuratan PCA. PCA biasanya tidak disarankan jika MSA kurang dari 0,5 karena kemungkinan tidak akan ada pengurangan. Nilai MSA untuk kumpulan data kami adalah 0,5818, yang lebih besar dari 0,5. Akibatnya, menerapkan PCA ke dataset kami diharapkan menghasilkan dimensi pengurangan dan ekstraksi komponen yang relevan.



Gambar 5.1 Scree Plot dari Dataset

Berdasarkan nilai eigen pada Gambar 5.1 dapat disimpulkan bahwa variabel prediktor dari dataset yang digunakan dapat direduksi menjadi 5 komponen (nilai eigen >1). Sebelum proses PCA dilakukan, data distandarisasi terlebih dahulu. Selanjutnya kelima variabel baru dari PCA tersebut akan digunakan sebagai dataset baru untuk proses clustering. Variasi total yang diwakili oleh komponen yang diekstraksi adalah 63,17%.

Hasil reduksi dimensi menggunakan PCA dapat digunakan untuk membuat visualisasi antara komponen pertama dan kedua berdasarkan label level UKT -nya. Gambar 5.2 menunjukkan bahwa kluster antar UKT tidak terpisah, dan terjadi overlap. Jika diukur menggunakan indeks Silhouette, nilai yang diperoleh adalah -0.0274 yang berarti kluster tersebut menunjukkan bahwa suatu objek tidak berada pada kluster yang tepat. Karena nilai kebaikan kluster berdasarkan indeks Silhouette mendekati -1, maka perlu dilakukan kluster ulang agar dapat meningkatkan nilai indeks Silhouette-nya. Bentuk cluster data baru dihasilkan menggunakan algoritma klustering yang akan dijelaskan pada bagian pembahasan selanjutnya.



Gambar 5.2 Scatter UKT hasil Pengelompokan Eksisting (kebijakan universitas)

B. Metode Klustering untuk Perbaikan Pengelompokan UKT

Banyak algoritma menggunakan ukuran kesamaan atau jarak antara titik dalam ruang fitur dalam upaya untuk menemukan daerah padat dari pengamatan. Karena itu, seringkali dilakukan tahapan menyamakan skala data sebelum menggunakan algoritma pengelompokan. Beberapa algoritma pengelompokan mengharuskan untuk menentukan jumlah klaster dalam data, sedangkan yang lain memerlukan spesifikasi beberapa jarak minimum antara pengamatan di mana titik sampel dapat dianggap "dekat" atau "terhubung." Dengan demikian, analisis klaster adalah proses berulang di mana evaluasi subjektif dari klaster yang diidentifikasi diumpankan kembali ke dalam perubahan konfigurasi algoritma sampai hasil yang diinginkan atau sesuai tercapai. Pada library scikit-learn menyediakan serangkaian algoritma pengelompokan yang berbeda untuk dipilih. Berikut algoritma yang cukup populer digunakan adalah sebagai berikut:

- 1) Agglomerative Clustering
- 2) BIRCH
- 3) DBSCAN
- 4) K-Means
- 5) Mini-Batch K-Means
- 6) Mean Shift

Hasil implementasi setiap algoritma berikut perhitungan Indeks Silhouette dapat dilihat pada grafik scatter berikut ini.

1) Agglomerative Clustering

Hierarchical methods adalah teknik klustering membentuk hirarki atau berdasarkan tingkatan tertentu sehingga menyerupai struktur pohon. Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Biasanya, metode ini digunakan pada data yang jumlahnya tidak terlalu banyak dan jumlah klaster yang akan dibentuk belum diketahui. Di dalam metode hirarki, terdapat dua jenis strategi pengelompokan yaitu agglomerative dan divisive. Agglomerative (metode penggabungan) adalah strategi pengelompokan hirarki yang dimulai dengan setiap objek dalam satu cluster yang terpisah kemudian membentuk cluster yang semakin membesar. Jadi, banyaknya cluster awal adalah sama dengan banyaknya objek. Sedangkan Divisive (metode pembagian) adalah strategi pengelompokan hirarki yang dimulai dari semua objek dikelompokkan menjadi cluster tunggal kemudian dipisah sampai

setiap objek berada dalam cluster yang terpisah. Hasil klastering dengan menggunakan metode agglomerative untuk data UKT dapat dilihat pada gambar 5.3



Gambar 5.3 Scatter UKT hasil Pengelompokan dengan Metode Agglomerative

Hasil pengukuran indeks Silhouette dengan pengelompokan dengan metode Agglomerative ini adalah sebesar 0.3589, meskipun lebih tinggi dari hasil perhitungan indeks Silhouette kelompok UKT eksisting, tetapi masih dapat dioptimalkan lagi dengan mencoba mengimplementasikan algoritma klastering yang lain.

2) BIRCH

BIRCH adalah kependekan dari *Balanced Iterative Reducing and Clustering using Hierarchies* yang melibatkan pembangunan struktur pohon dari mana pusat cluster di ekstraksi.

Metode Hierarchies dengan Algoritma BIRCH pada clustering mampu untuk mengelolah data yang besar dengan memori yang kecil dan menghasilkan waktu eksekusi yang lebih singkat. Tetapi, pada algoritma ini diperlukan modifikasi nilai threshold agar menghasilkan kualitas cluster yang baik. Hasil implementasi pengelompokan UKT dengan algoritma BIRCH dapat dilihat pada Gambar 5.4.

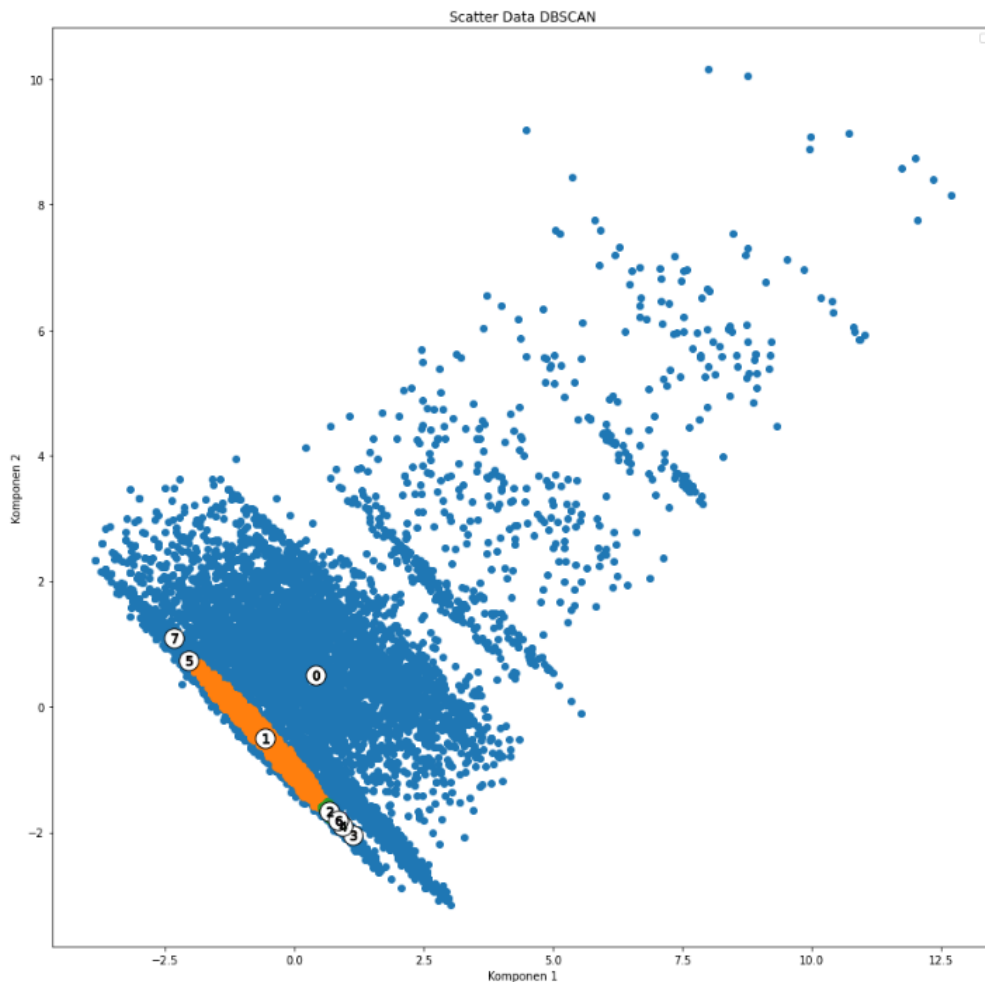


Gambar 5.4 Scatter UKT hasil Pengelompokan dengan Metode BIRCH

Hasil perhitungan indeks Silhouette dengan algoritma BIRCH menghasilkan nilai yang sama dengan metode Agglomerative yaitu 0.3589 karena pada dasarnya sama-sama dalam pendekatan metode Hirarki.

3) DBSCAN

DBSCAN adalah kependekan dari *Density-Based Spatial Clustering of Applications with Noise* yang melibatkan pencarian area kepadatan tinggi dalam domain dan memperluas area ruang fitur di sekitar mereka sebagai kluster. Hasil implementasi data UKT dengan metode DBSCAN dapat dilihat pada gambar 5.5.

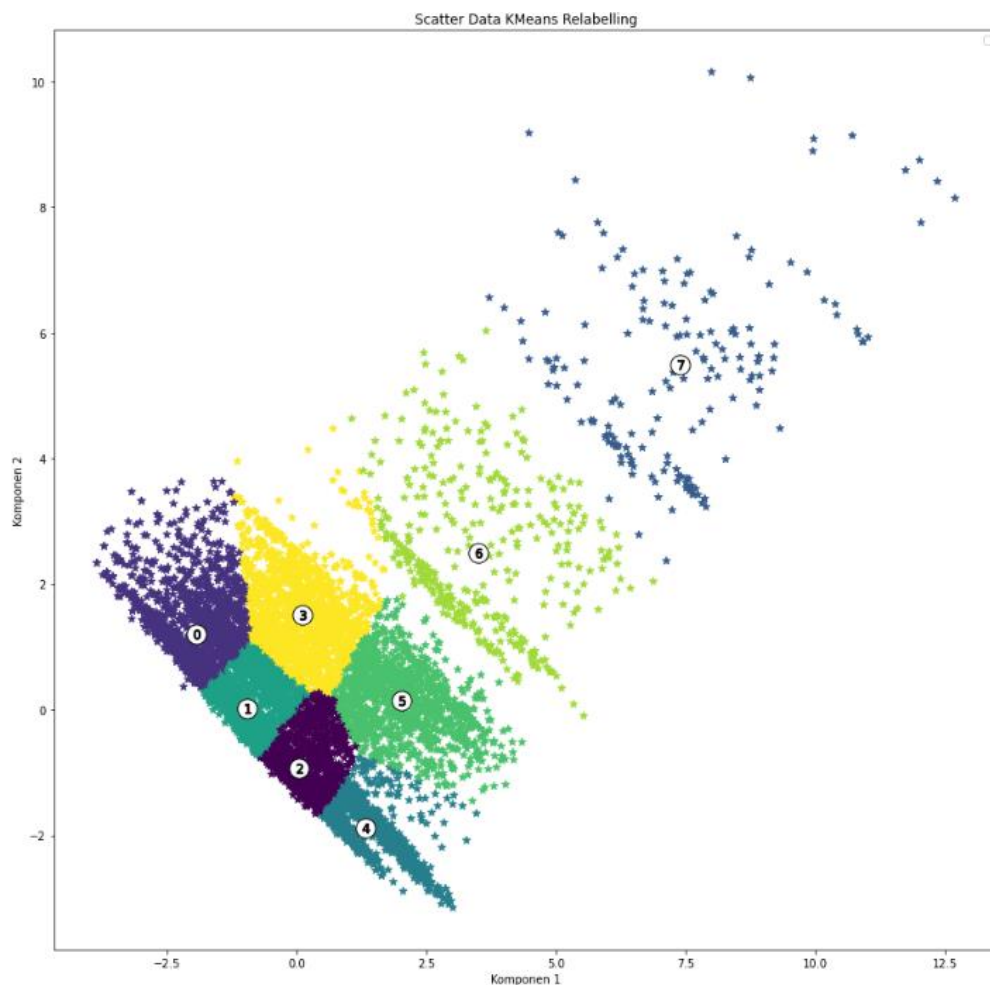


Gambar 5.5 Scatter UKT hasil Pengelompokan dengan Metode DBSCAN

Dengan menggunakan indeks Sillhouette diperoleh nilai sebesar -0.1975 apabila dipaksakan untuk membentuk 8 kluster sesuai dengan jumlah kluster yang dikehendaki pada pengelompokan UKT. Nilai indeks tersebut menunjukkan bahwa dengan metode ini, data UKT tidak dapat terkluster dengan baik dan masih terjadi overlapping.

4) K-Means

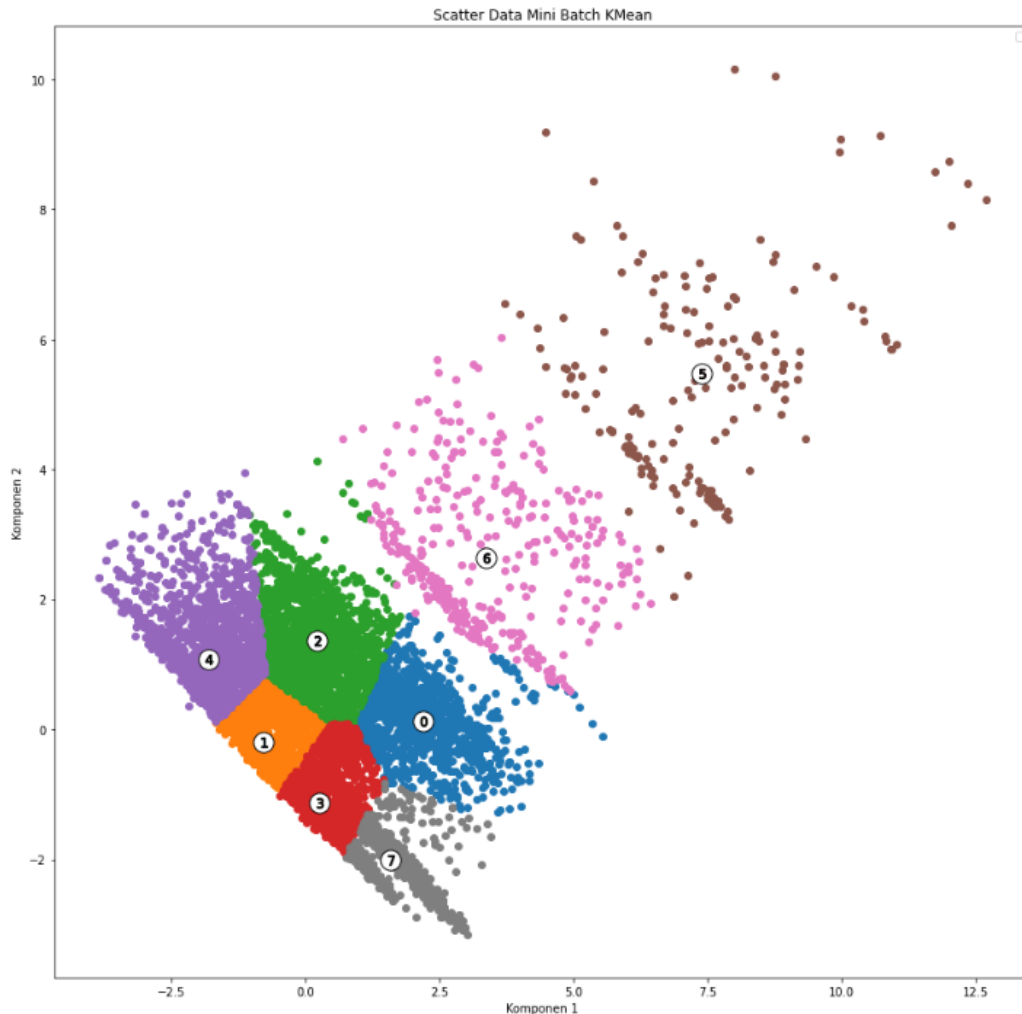
K-means clustering adalah salah satu algoritma analisis klaster non hirarki. Tujuan dari K-means clustering, seperti metode klaster lainnya, adalah untuk mendapatkan kelompok data dengan memaksimalkan kesamaan karakteristik dalam klaster dan memaksimalkan perbedaan antar klaster. Algoritma K-means mengelompokkan data berdasarkan jarak antara data terhadap titik centroid klaster yang didapatkan melalui proses berulang. Analisis perlu menentukan jumlah K sebagai input algoritma. Hasil implementasi pengelompokan data UKT dapat dilihat pada Gambar 5.6. Indeks Sillhouette dari hasil klastering K-Means adalah sebesar 0.4190, mengindikasikan tingkat pengelompokan yang baik dan lebih tinggi nilainya dibandingkan metode sebelumnya.



Gambar 5.6 Scatter UKT hasil Pengelompokan dengan Metode K-Means

5) Mini-Batch K-Means

Mini-Batch K-Means adalah versi modifikasi dari k-means yang membuat pembaruan pada centroid cluster menggunakan sampel mini-batch daripada seluruh dataset, yang dapat membuatnya lebih cepat untuk dataset besar, dan mungkin lebih kuat terhadap gangguan statistik. Implementasi pada data UKT yang memiliki jumlah sampel tidak termasuk dalam Big Data maka hasilnya menjadi sama dengan hasil K-Means. Plot scatter data UKT berbasis mini batch k-means.

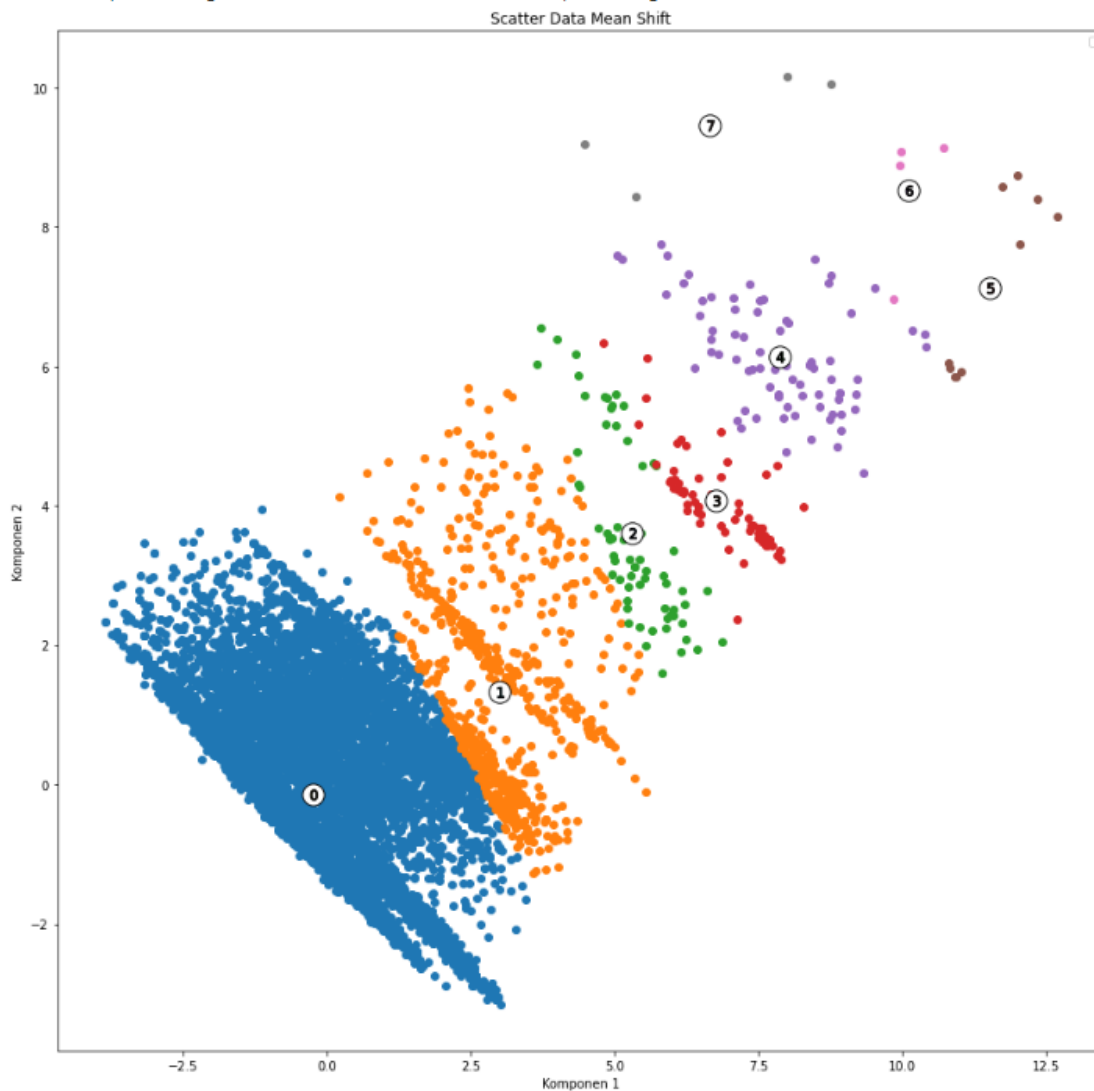


Gambar 5.7 Scatter UKT hasil Pengelompokan dengan Metode Mini Batch K-Means

6) Mean Shift

Algoritma ini memiliki konsep pengelompokan dengan pergeseran rata-rata yang melibatkan adaptasi centroid berdasarkan kepadatan titik sampel di ruang fitur. Tidak seperti pengelompokan K-means, yang tidak membutuhkan asumsi apa pun; maka itu K-Means

merupakan algoritma non parametrik. Sedangkan algoritma mean-shift pada dasarnya menetapkan titik data ke cluster secara iteratif dengan menggeser menunjuk ke arah kepadatan tertinggi dari titik data yaitu cluster centroid. Perbedaan antara algoritma K-Means dan Mean-Shift adalah bahwa nanti tidak perlu untuk menentukan jumlah cluster terlebih dahulu karena jumlah cluster akan menjadi ditentukan oleh algoritma w.r.t data. Pada kasus pengelompokan UKT, diregulasi parameter nya sehingga dapat menghasilkan 8 klaster.



Gambar 5.8 Scatter UKT hasil Pengelompokan dengan Metode Mean Shift

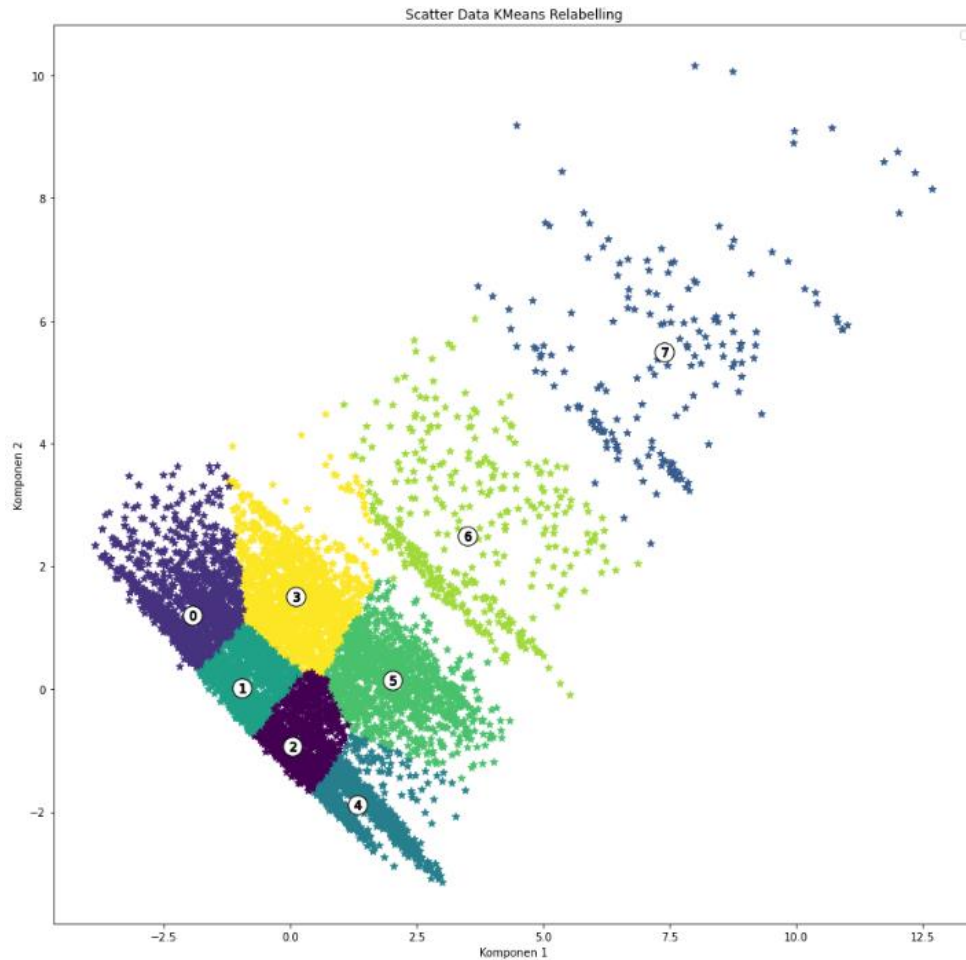
C. Pelabelan Kelompok UKT berbasis Klastering

Dengan melihat hasil klastering yang sudah dijelaskan sebelumnya, didapatkan bahwa pengelompokan data UKT yang paling baik adalah dengan menggunakan K-Means atau Mini Batch K-Means. Dengan demikian, data ini selanjutnya akan digunakan untuk proses selanjutnya yaitu pelabelan ulang berdasarkan nilai konfidensi PCA pada data UKT riil. Adapapun hasil pelabelan dapat dilihat pada gambar 5.9.

Tabel 5.1 Hasil Rotasi PCA

Variabel	1 st Component	2 nd Componen
X1	0.0609	0.5278
X2	0.0682	-0.0312
X3	0.0659	0.7815
X4	0.0954	0.0697
X5	-0.0400	-0.2243
X6	0.0359	0.2769
X7	0.1356	0.4277
X8	0.0370	0.3381
X9	-0.0018	-0.0051
X10	0.0284	0.1591
X11	0.0665	0.4540
X12	0.9994	0.0347
X13	0.9994	0.0347

Proses ini berbasis pada nilai variabel yang dominan pada hasil rotasi PCA. Jika hasil PCA dirotasi menggunakan metode varimax untuk melihat variabel apa saja yang dominan pada masing-masing komponen maka hasilnya dapat dilihat pada Tabel 5.1. Berdasarkan tabel 5.1 tersebut yang komponen pertama didominasi oleh variabel X12 dan X13 yang merupakan pendapatan orang tua selain gaji pokok. Box Plot pada Gambar 4.2 menunjukkan bahwa semakin tinggi tingkat UKT, semakin besar pendapatan ayah dan ibu. Sebaliknya, jenis pekerjaan ayah dan upah pokok ayah mendominasi faktor kedua.



Gambar 5.9 Scatter UKT hasil Pengelompokan dengan Relabelling K-Means

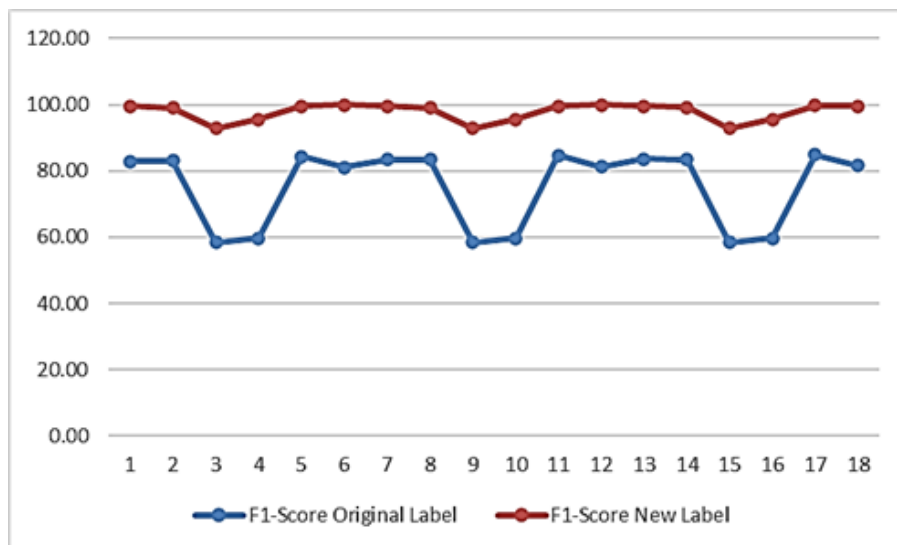
D. Pemodelan Prediksi Kelompok UKT

Hasil perhitungan rata-rata F1-Score dan AUC masing-masing dapat dilihat pada Tabel 5.2 dan Tabel 5.4. Secara umum terdapat perbedaan hasil pengukuran F1-Score untuk data dengan label asli dan data dengan label baru. Untuk mendukung apakah perbedaan ini signifikan, perlu dilakukan uji statistik untuk dua sampel berpasangan. Pemilihan statistik uji yang sesuai harus diperiksa pada asumsi distribusi normalitas hasil pengukuran untuk F1-Score dan AUC. Uji normalitas dilakukan melalui uji Kolmogorov-Smirnov, dan didapatkan data F1-Score dan AUC tidak berdistribusi normal. Dengan demikian, uji beda antara F1-Score dan nilai AUC sebelum dan sesudah pelabelan menggunakan pendekatan statistik nonparametrik.

Tabel 5.2 Perhitungan Rata-Rata F1- Score

Dataset (<i>p</i>)	fold (<i>q</i>)	Classification Algorithm (<i>r</i>)					
		<i>DT</i>	<i>KNN</i>	<i>LR</i>	<i>NB</i>	<i>RF</i>	<i>SVM</i>
Original Labeled Data	3	82.98	83.11	58.48	59.65	84.29	81.03
	5	83.41	83.41	58.35	59.64	84.69	81.22
	10	83.63	83.49	58.47	59.69	84.85	81.58
Relabeled Data	3	99.50	98.98	92.82	95.54	99.65	99.85
	5	99.51	99.11	92.85	95.50	99.66	99.85
	10	99.56	99.15	92.93	95.56	99.73	99.60

Jika data pada Tabel 5,2 disajikan dalam bentuk grafik pada Gambar 5.10 , dapat dilihat bahwa rata-rata akurasi F1-Score pada dataset dengan label baru memiliki nilai yang lebih tinggi dibandingkan pada dataset dengan label aslinya.



Gambar 5.10 Perbandingan Nilai F1-Score data Asli dengan Data Relabel

Uji statistik Wilcoxon sampel berpasangan dengan hipotesis nol bahwa tidak ada perbedaan antara F1-Score dataset dengan label asli dan label baru dilakukan untuk mendukung kesimpulan bahwa hasil F1 Score rata-rata pada dataset dari label baru secara signifikan lebih baik.

Tabel 5.2 Uji Statistik Non Parametrik untuk Uji Beda F1-Score

Wilcoxon Signed Test Statistics ^a		Conclusion
	F1-Score New Label – F1-Score Original Label	
Z	-3.724 ^b	
Asymp. Sig. (2-tailed)	0.000	H ₀ Rejected
a. Wilcoxon Signed Ranks Test		
b. Based on negative ranks.		

Tabel 5.2 menunjukkan bahwa p-value kurang dari 5%. Disimpulkan bahwa data mendukung secara statistik bahwa hasil dataset F1-Score dengan label baru berbeda (lebih baik) dengan dataset dengan label aslinya.

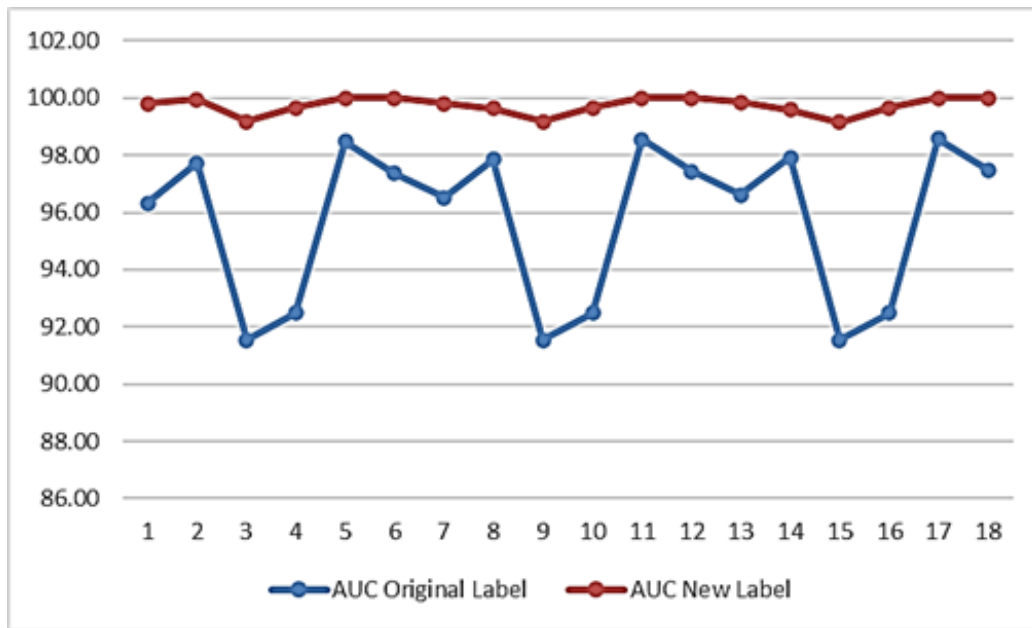
Tabel 5.3 Perhitungan Rata-Rata Nilai AUC

Dataset (<i>p</i>)	fold (<i>q</i>)	Classification Algorithm (<i>r</i>)					
		<i>DT</i>	<i>KNN</i>	<i>LR</i>	<i>NB</i>	<i>RF</i>	<i>SVM</i>
Original Labeled Data	3	96.33	97.73	91.54	92.49	98.48	97.38
	5	96.51	97.84	91.54	92.48	98.55	97.44
	10	96.62	97.92	91.55	92.48	98.59	97.49
Relabeled Data	3	99.81	99.96	99.18	99.66	100.0	100.0
	5	99.81	99.63	99.17	99.66	100.0	100.0
	10	99.84	99.59	99.16	99.66	100.0	100.0

Sama halnya dengan F1-Score, untuk hasil pengukuran AUC antara data asli dengan data berlabel baru pada Tabel 5.43 dilakukan uji beda dengan hasil seperti terlihat pada Tab 5.4. Hasil pengujian menunjukkan perbedaan yang signifikan antara nilai AUC sebelum dan sesudah pelabelan ulang. Perbedaan ini dapat dilihat secara visual pada Gambar 5.11 bahwa nilai AUC pada data label saat ini lebih tinggi dari data aslinya.

Tabel 5.4 Uji Statistik Non Parametrik untuk Uji Beda AUC

Wilcoxon Signed Test Statistics ^a		Conclusion
	AUC New Label – AUC Original Label	
Z	-3.724 ^b	
Asymp. Sig. (2-tailed)	0.000	H ₀ Rejected
a. Wilcoxon Signed Ranks Test		
b. Based on negative ranks.		



Gambar 5.11 Perbandingan Nilai AUC data Asli dengan Data Relabel

BAB VI

KESIMPULAN DAN SARAN

A. Kesimpulan

Hasil penelitian ini menyimpulkan bahwa (1) kualitas klaster data awal berdasarkan indeks Silhouette menunjukkan banyak objek yang tidak berada pada posisi klaster yang tepat, sehingga perlu dilakukan klaster ulang, (2) kualitas klaster terbaik dari pengelompokan UKT diperoleh dengan menggunakan algoritma K-Mean atau Mini Batch K-Mean dengan indeks Silhouette mendekati 1, dan (3) proses pelabelan dari hasil cluster menggunakan K-Means dapat dilakukan dengan pemetaan label K-Means menggunakan nilai CI dari hasil PCA terutama pada komponen pertama, (4) percobaan pemodelan klasifikasi untuk data sebelum dan sesudah pelabelan menunjukkan perbedaan yang signifikan, dan hasil akurasi jauh lebih baik.

B. Saran

Perlu dilakukan eksperimen lebih lanjut untuk implementasi pengelompokan berbasis unsupervised learning seperti dengan algoritma OPTICS, Spectral dan Mixture Gaussian.

DAFTAR PUSTAKA

- Abidin, T. F., Rizal, S., Iqbalsyah, T. M., & Wahyudi, R. (2020). Decision tree classifier for university single rate tuition fee system. *International Journal of Business Intelligence and Data Mining*, 17(2), 258–271. <https://doi.org/10.1504/IJBIDM.2020.108764>
- Anas, Tempola, F., & Khairan, A. (2019). Hybrid fuzzy dan Naive Bayes Dalam Penentuan Status UKT. *PROtek*, 06(1), 2–7.
- Arianti, K. (2017). *Klasifikasi Mahasiswa Baru IPB Berdasarkan Besaran Uang Kuliah Tunggal Menggunakan Metode Fuzzy K-Nearest Neighbor*. IPB.
- Cai, Z., Wang, D., & Jiang, L. (2007). K-Distributions : A New Algorithm for Clustering Categorical Data. *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, 4682, 436–437.
- Desgraupes, B. (2017). Clustering Indices. *Quest - Lab Modal'X*, 1(November), 34.
- Dutta, M., Mahanta, A. K., & Pujari, A. K. (2005). QROCK: A quick version of the ROCK algorithm for clustering of categorical data. *Pattern Recognition Letters*, 26(15), 2364–2373. <https://doi.org/10.1016/j.patrec.2005.04.008>
- Gan, G., Ma, C., & Wu, J. (2007). Data Clustering : Theory, Algorithms and Applications. In *ASA-SIAM Series on Statistics and Applied Probability*. American Statistical Association and the Society for Industrial and Applied Mathematics. 10. <https://doi.org/10.1017/CBO9781107415324.004>
- Guha, S., Rastogi, R., & Shim, K. (1999). ROCK: A Robust Clustering Algorithm for Categorical. In *International Conference on Data Engineering* (pp. 512–521).
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2002). Cluster Validity Methods: Part I. *SIGMOD Rec.*, 31(2), 40–45. <https://doi.org/10.1145/565117.565124>
- Huang, Z. (1997). A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. In *In Research Issues on Data Mining and Knowledge Discovery* (pp. 1–8).
- Jazuli, M. (2016). *Uang Kuliah Tunggal Menggunakan Metode K-Means Clustering*. UIN Maulana Malik Ibrahim.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis - International Edition* (6th ed.). Pearson, Prentice Hall.
- Karim, B., Sentinuwo, S. R., & Sambul, A. M. (2017). Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan Data Mining. *E-Journal Teknik Informatika*, 11(1).
- Muchsin, A. K., & Sudarma, M. (2017). Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggal Mahasiswa Baru. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 6(3), 175. <https://doi.org/10.24843/lkjiti.2015.v06.i03.p04>
- Muhammad, F. (2016). *Pengembangan Sistem Penentuan Uang Kuliah Tunggal Dengan Metode*

Fuzzy C-Means. Universitas Islam Negeri Maulana Malik Ibrahim.

- Prasetyanti, D. N., & Listyaningrum, R. (2017). Kaji Banding Metode TOPSIS, Saw Dan AHP-TOPSIS Guna Menentukan UKT Mahasiswa Baru Di Politeknik Negeri Cilacap. *Jurnal Infotekmes*, 8(1), 9–14. <http://www.infotekmesin.org/ojs/index.php/infotekmesin/article/view/46>
- Rokhman, S., Rozi, I. F., & Asmara, R. A. (2017). Pengembangan sistem penunjang keputusan penentuan UKT mahasiswa dengan menggunakan metode Moora studi kasus Politeknik Negeri Malang. *Jurnal Informatika Polinema*, 3, 36–42.
- Suyoga, I. G. S., Kencana, I. P. E. N., & Sukarsa, I. K. G. (2018). Penggolongan Uang Kuliah Tunggal Menggunakan Support Vector Machine. *E-Jurnal Matematika*, 6(4), 220. <https://doi.org/10.24843/mtk.2017.v06.i04.p169>
- Tahir, I. (2016). Model Pengambilan Keputusan Penentuan Uang Kuliah Tunggal (UKT) Pada Perguruan Tinggi Negeri (Studi Kasus: Universitas Sembilanbelas November Kolaka). *Speed*, 8(2), 1–9.
- Testiana, G. (2017). Pemanfaatan Metode Simple Additive Weighting (SAW) untuk Penentuan Penerima UKT Kelompok1. *Proceeding Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)* 9.
- Utomo, M. N. Y., Permanasari, A. E., Tungadi, E., & Syamsuddin, I. (2018). Determining single tuition fee of higher education in Indonesia: A comparative analysis of data mining classification algorithms. *Proceedings of 2017 4th International Conference on New Media Studies, CONMEDIA 2017, 2018-Janua*, 113–117. <https://doi.org/10.1109/CONMEDIA.2017.8266041>
- Yanifa, Y. S., & Candra, F. (2018). Sistem Pendukung Keputusan Penentuan Golongan UKT dengan Metode Naive Bayes. *Jom FTeknik*, 5(2), 110–117.
- Yustanti, W., Anistyasari, Y., & Imah, E. M. (2018). Determining student's single tuition fee category using correlation based feature selection and support vector machine. *Proceeding International Conference on Advanced Computer Science and Information Systems, ICACISIS 2017, January*, 172–176. <https://doi.org/10.1109/ICACISIS.2017.8355029>
- Zheng, Q., Diao, X., Cao, J., Liu, Y., Li, H., Yao, J., Chang, C., & Lv, G. (2020). From Whole to Part: Reference-Based Representation for Clustering Categorical Data. *IEEE Transactions on Neural Networks and Learning Systems*, 31(3), 927–937. <https://doi.org/10.1109/tnnls.2019.2911118>
- Zurraedah. (2018). *Evaluasi klasifikasi kategori pembayaran SPP di Universitas Islam Negeri Alauddin Makassar (UINAM) tahun akademik 2017/2018 menggunakan k-means clustering*. Universitas Islam Negeri Alauddin Makassar (UINAM).

Lampiran 1. Biodata Peneliti

A. Identitas Diri

1	Nama Lengkap (dengan gelar)	Wiyli Yustanti, S.Si,M.Kom
2	Jenis Kelamin	P
3	Jabatan Fungsional	Lektor Kepala
4	NIP/NIK/Identitas lainnya	197702032005012001
5	NIDN	0003027708
6	Tempat dan Tanggal Lahir	Mojokerto, 3 Februari 1977
7	E-mail	wiyllyustanti@unesa.ac.id
9	Nomor Telepon/HP	087854407576
10	Alamat Kantor	Kampus Unesa Ketintang Ged. A10
11	Nomor Telepon/Faks	-
13. Mata Kuliah yg Diampu	1.	Sistem Basis Data
	2.	Manajemen Basis Data
	3.	Pemrograman Basis Data
	4.	Sistem Pendukung Keputusan Bisnis
	5.	Statistika dan Probabilitas
	6.	Statistika

B. Riwayat Pendidikan

	S-1	S-2	S-3
Nama Perguruan Tinggi	ITS	ITS	ITS
Bidang Ilmu	Statistika	Teknik Informatika	Ilmu Statistik
Tahun Masuk-Lulus	1995 – 1999	2001 - 2004	2019 -
Judul Skripsi/Tesis/Disertasi	<i>Study of Multivariate Signed Rank Test for Testing K Mean Vector</i>	<i>Time Series Forecasting using Combination of Wavelet Transformation and Artificial Neural Network Methods</i>	-
Nama Pembimbing/Promotor	Dr. Purhadi M.Si	Prof. Arief Djunaidi MSc,PhD	Prof. Nur Iriawan, M.Comp, PhD

C. Pengalaman Penelitian Dalam 5 Tahun Terakhir

(Bukan Skripsi, Tesis, maupun Disertasi)

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber*	Jml (Juta) Rp
1	2014	Penentuan Sindrom Penyakit Pada Traditional Chinese Medicine (Tcm) Dengan Menggunakan <i>Expert System</i>	HIBAH BERSAING	59 Juta
2	2015	Pengembangan Buku Saku Elektronik Sebagai Media Pembelajaran Teknologi Seluler di Perguruan Tinggi	HIBAH BERSAING	59 Juta
3	2016	Strategi Pengembangan <i>Repository System</i> Universitas Negeri Surabaya	PNBP FT	7,5 Juta
4	2017	Pengembangan Tata Kelola PPTI Unesa dengan menggunakan Indeks KAMI	DESENTRALISASI	112,5 juta
5	2017	Validasi Prediktif Penerimaan Mahasiswa Baru jalur SNMPTN Universitas Negeri Surabaya	PNBP Universitas	50 juta
6	2017	Pengembangan Sistem Kinerja Pegawai (SKP) Universitas Negeri Surabaya	PNBP Universitas	50 juta
7	2018	Pengembangan Tata Kelola PPTI Unesa dengan menggunakan Indeks KAMI Lanjutan	DESENTRALISASI	130 juta
8	2019	Pengembangan Media Pembelajaran Pengenalan Alat Music Berbasis <i>Augmented Virtual Reality</i>	PNBP FT	15 Juta

D. Pengalaman Pengabdian Kepada Masyarakat dalam 5 Tahun Terakhir

No.	Tahun	Judul Pengabdian Kepada Masyarakat	Pendanaan	
			Sumber*	Jml (Juta) Rp
1	2015	Pelatihan Pemrograman Robotic Untuk LKS	Jurusan TE	7,5 Juta

2	2016	Pelatihan <i>Digital Mind Mapping</i> untuk Guru SD Lab Unesa	Jurusan TI	7,5 Juta
3	2017	Peningkatan Kompetensi Guru TIK melalui workshop implementasi web sebagai portofolio guru MGMP TIK dan Prakarya SMP Negeri se Kabupaten Mojokerto	Jurusan TI	10 Juta
4	2018	Workshop Pembuatan Animasi Game Menggunakan Unity di SMA Wachid Hasyim 2 Ngelom Sidoarjo	Jurusan TI	10 Juta
5	2019	Pelatihan <i>Digital Mind Mapping</i> bagi Guru SMPN 2 Magetan untuk meningkatkan Kompetensi Paedagogik	Jurusan TI	10 Juta

E. Publikasi Artikel Ilmiah Dalam Jurnal alam 5 Tahun Terakhir

No.	Judul Artikel Ilmiah	Nama Jurnal	Volume/
1	<i>Computer Simulation Based on K-Means as Learning Media to Determine Syndrome of Disease in Traditional Chinese Medicine</i>	<i>Journal of Convergence Information Technology</i>	Volume 10 No.5, September 2015 hal. 79-88, ISSN: 1975-9320 (print), 2233-9299 (online)
2	Sistem Pendukung Keputusan Penerimaan Siswa Baru Menggunakan Metode SAW (Studi Kasus: SMK IPIEMS Surabaya)	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 5 No.2 Tahun 2016, hal. 143 - 151 ISSN: 2252-5157
3	Rancang Bangun Aplikasi Monitoring Tumbuh Kembang Balita Berbasis Android	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 6 No.1 Tahun 2016 hal 1-6, ISSN: 2252-5157
4	Rancang Bangun E-Voting Berbasis Website di Universitas Negeri Surabaya	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 6 No.1 Tahun 2016 hal 72-81, ISSN: 2252-5157

5	Sistem Informasi Sasaran Kerja Pegawai (SKP) Universitas Negeri Surabaya	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 6 No.1 Tahun 2016 hal 117-125, ISSN: 2252-5157
6	Aplikasi Forum Komunikasi Pada Universitas Negeri Surabaya Berbasis Android Menggunakan Mongoddb	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 6 No. 1Tahun 2016 hal 127-133, ISSN: 2252-5157
7	Perbandingan Penggunaan NoSQL Mongoddb Dan MySQL Pada Basis Data Forum Komunikasi	Jurnal Manajemen Informatika (JMI) Fakultas Teknik Universitas Negeri Surabaya	Volume 6 No. 1Tahun 2016 hal 134-142, ISSN: 2252-5157
8	Perancangan Model Enterprise Architecture untuk Bidang Non Akademik pada Perguruan Tinggi Negeri X dengan Menggunakan Pendekatan Framework TOGAF ADM	<i>Journal of Information Engineering and Educational Technology</i> (JIEET)	Volume 3 No.1 Tahun 2019, hal 31-38 , ISSN: 2549-869X
9	Strategi Identifikasi Resiko Keamanan Informasi Dengan Kerangka Kerja ISO 27005: 2018	<i>Journal of Information Engineering and Educational Technology</i> (JIEET)	Volume 3 No.2 Edisi Desember 2019, hal 51-56 , ISSN: 2549-869X
10	Analisis Pemilihan Aplikasi Opensource ERP terhadap UKM Menggunakan Metode Kombinasi ANP dan PROMETHEE	<i>Journal of Informatics and Computer Science</i> (JINACS)	Volume 1 No.3 Tahun 2020, hal 122-127 , ISSN: 2686- 2220

F. Pemakalah Seminar Ilmiah (*Oral Presentation*) dalam 5 Tahun Terakhir

No	Nama Pertemuan Ilmiah / Seminar	Judul Artikel Ilmiah	Waktu dan Tempat
1	Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat LPPM 2015	Pengembangan Buku Saku Elektronik Sebagai Media Pembelajaran Teknologi Seluler di Perguruan	November 2015 Surabaya

2	<i>The 2nd Annual Applied Science and Engineering Conference (AASEC) 2017</i>	<i>A Polychoric Correlation to Identify the Principle Component in Classifying Single Tuition Fee Capabilities on the Students Socio-Economic Database</i>	24 Agustus 2017, Bandung
3	<i>The Consortium of Asia-Pacific Education Universities (CAPEU) 2017</i>	<i>An Analysis of Indonesia's Information Security Index: A Case Study in a Public University</i>	22-23 Mei 2017, Surabaya
4	<i>International Conference on Advanced Computer Science and Information Systems (ICACSIS) 2017</i>	<i>Determining student s single tuition fee category using correlation based feature selection and support vector machine</i>	28-29 Okt 2017 Jakarta
5	Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat LPPM 2018	Analisis Tingkat Kesiapan Dan Kematangan Implementasi Iso 27001: 2013 Menggunakan Indeks Keamanan Informasi 3: 2015 Pada UPT. PPTI Universitas Negeri Surabaya	November 2018 Surabaya

G. Karya Buku dalam 5 Tahun Terakhir

No	Judul Buku	Tahun	Jumlah Halaman	Penerbit
1	Pengembangan E-Materi Virtual Learning ISBN : -	2016	250	Unesa Press
2	Keamanan Sistem Informasi ISBN :978-602-5815-29-4	2018	276	Zifatama Jawa

H. Perolehan HKI dalam 5–10 Tahun Terakhir

No.	Judul/Tema HKI	Tahun	Jenis	Nomor P/ID
1				

I. Pengalaman Merumuskan Kebijakan Publik/Rekayasa Sosial Lainnya dalam 5 Tahun Terakhir

No.	Judul/Tema/Jenis Rekayasa Sosial Lainnya yang Telah Diterapkan	Tahun	Tempat Penerapan	Respon Masyarakat
1				

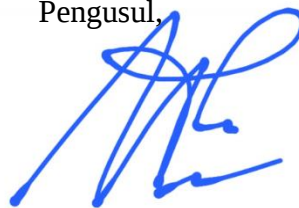
J. Penghargaan dalam 10 tahun Terakhir (dari pemerintah, asosiasi atau institusi lainnya)

No.	Jenis Penghargaan	Institusi Pemberi Penghargaan	Tahun
1			

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Hibah/Pengabdian Masyarakat.

Surabaya, 1 September 2022

Pengusul,



Wiyli Yustanti, S.Si,M.Kom

NIP. 197702032005012001

Lampiran 2. Artikel Jurnal

Categorical Encoder Based Performance Comparison in Pre-Processing Imbalanced Multiclass Classification

Wiyli Yustanti^{1,2}, Nur Iriawan¹, Irhamah¹

¹Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Informatics, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

Keywords:

Categorical Encoding

Classification

Multiclass

Imbalanced

Performance Analysis

ABSTRACT

The distribution of response variables (class), predictor variable data type, classification algorithm, number of predictor variables, number of samples, validation, and evaluation model are some of the characteristics that might influence a classification model's performance results. This study's contribution is to offer suggestions for categorical predictor variable encoding techniques that produce the most significant performance outcomes for multiclass classification problems with imbalanced class distributions. We selected four open-source data sets with imbalanced class distributions and categorical predictors. The methods of Ordinal, Nominal, Dirichlet, Frequency, Target, Leave One Out, One-Hot, Dummy, Binary, and Hashing Encoder are used. Oversampling and stratified k-fold cross-validation techniques solve the imbalance issue. We use the Grid-Search technique to find the best hyperparameters. The F1-Score and AUC follow. Then, the F1-Score and AUC with OVO and OVA approach for multiclass were evaluated to choose the optimal model. In all datasets with 10-fold stratified cross-validation and 95% to 99% accuracy for each dataset, the results showed that SVM outperformed the DT, KNN, NB, RL, and RF algorithms. Probability-based or binary encodings, such as Target, Dirichlet, Dummy, One-Hot, or Binary, are excellent for situations with a minor class proportion of less than 3%. The Nominal or Ordinal encoder is preferred for data with a minor class proportion of more than 3%.

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



Corresponding Author:

Nur Iriawan

Department of Statistics, Faculty of Science and Data Analytics

Institut Teknologi Sepuluh Nopember

Sukolilo Campus, Surabaya, Indonesia

Email: nur_i@statistika.its.ac.id

1. INTRODUCTION

Classification algorithms are included in the supervised learning method in the context of machine learning. Supervised learning is machine learning with a function that maps input to output based on examples of input-output pair data [1]. This function results from learning a series of labeled training datasets [2]. This training data consists of vector input objects and the desired output value (target). The algorithm used in supervised learning will analyze the training data to produce the classifier function used to map new input data (testing data) to the correct output data. Therefore, the main goal of this machine learning is to get the optimal scenario that can correctly define the class label for the new object. Many factors affect the performance results of object classification, including data types (numeric, categorical, or mixed), data dimensions, sample size, and feature engineering. The feature engineering process could include handling missing values, data transformation, dimension reduction, feature selection, outlier detection, and resampling with under or oversampling if the class distribution is unequal. The classification performance is also influenced by the model selected, the hyperparameter optimization approach, the model validation, and the evaluation techniques. The

studies related to the performance comparison of classification algorithms can be found in Osisanwo et al. [3], Bisgin et al. [4], Shafri et al. [5], and Zheng et al. [6]. The results of those studies concluded that SVM showed better accuracy than other algorithms. Furthermore, related categorical features for prediction models are also found in Heru et al. [7] and Azmi et al. [8]. This research is focused on the pre-processing stage when the input data type is categoric and the response variable (target) has more than two classes with unequal proportions. The main problem to be investigated is the categorical data transformation method that can increase the prediction performance of object classification with the highest scores of F1-Scores and AUC, especially for multiclass cases with unbalanced distributions. Experiments will be applied to six classification algorithms and ten types of encoders.

2. RELATED WORKS

2.1 Categorical Encoding

The quality of the data and the amount of helpful information are key factors in determining how well machine learning algorithms perform the learning process. Therefore, it is crucial to correctly transform categorical variables before entering such data into machine learning algorithms. It is well known that most classification methods in machine learning require predictor variables of the numeric type. Therefore, this section will briefly discuss several ways that can be used to perform the transformation or encoding from categoric to numeric data. According to Hancock [9] there are three data transformation models from categoric to numeric, such as the determined, algorithmic, and automatic techniques. Firstly, the determined approach is a way to transform categorical data by converting categorical data to numeric vectors with low computational complexity. Examples of this technique include ordinal, nominal, target, leave one out, hashing, frequency, [10] binary, dummy, one hot [11],[12] and Dirichlet [13],[14],[15]. Second, the algorithmic technique is a categorical encoding method that requires a large computational process, such as Latent Dirichlet Allocation (LDA) for document data [15] and Enabling Deep Learning for Generic Data Classification (EDLT) for tabular data [16]. Lastly, the third is an automatic encoding technique that combines data representation search methods into the machine learning process. Automated methods are more attractive because they are more general-purpose than algorithmic techniques. For example, the algorithm Word2Vec [17]. In this study, ten types of deterministic encoding techniques will be used, namely:

- 1) *Nominal Encoding (NE)*. In Nominal encoding, an integer is selected for each value of a categorical variable regardless of order. For example, if the dataset has a categorical variable whose value is set to {"female", "male"}, then the nominal label can simply use {0, 1} which means 0 for "female" and 1 for "male".
- 2) *Ordinal Encoding (OE)*. Ordinal encoding has the same concept as nominal encoding, only in labeling the order of the integer values that are mapped. For example, the education level variable has a value of {"Junior High School", "Senior High School", "Diploma", "Undergraduate", "Postgraduate"} transformed into a value of {0,1,2,3,4}.
- 3) *Target Encoding (TE)*. In this encoding, the probability value of each predictors' category level is calculated based on the class (target variable). Furthermore, the probability value (0 to 1) is used to replace the calculated categorical data.
- 4) *Frequency Encoding (FE)*. Frequency Coding is an encoding technique that encodes the value of a categorical feature to its frequency. After finding the frequency value for each category can be normalized or standardized in 0 to 1.
- 5) *Dirichlet Encoding (DRE)*. Conjugate Bayesian Method (CBM) has been implemented by Slakey et al. [12] by using the prior distribution function is Dirichlet with parameter α , as in (1), and the likelihood function of the data is assumed to be multinomial distribution ($p_1, p_2, \dots, p_k$) for multiclass. So that the posterior distribution function produces a Dirichlet distribution with parameters:

$$\alpha^* = \alpha + \sum_{i=1}^n y_i \quad \text{where } \alpha, \alpha^* \in \mathbb{R}^k \quad (1)$$

s.t

$$p(\theta_{mv} | y) \propto L(\theta_{mv} | y) p(\theta_{mv})$$

$L(\theta_{mv} | y)$ is a function of the likelihood of each value v in categorical predictor m against the target y , and θ_{mv} is the distribution parameter for each categorical value, then $p(\theta_{mv}) \sim \text{Dirichlet}(\alpha)$ with $\alpha = \frac{1}{k}$, k is number of classes.

- 6) *Leave One Out Encoding (LE)*. The idea is the concept of k-fold encoding to compute the target variable's average for all data containing the same value for the categorical feature variable. The average value can be obtained if the target data type is numeric, but if the target data type is categorical, then use the probability value.
- 7) *One-Hot Encoding (OHE)*. A new variable will be created if the predictor variable is nominal (no order). Each category is mapped with a binary variable containing either 0 or 1. In this case, 0 represents absence, and 1 represents level presence. The number of features will increase to N features if there are N categories in a variable.
- 8) *Hashing Encoding (HE)*. This approach is suitable for variables that have many categorical levels. Many types of hash functions map random-size data to fixed-size data in a numeric hash. The initial idea of the hashing method is to change a string of data in a code. That is intended to keep the data secret and runs in one direction, meaning the original data value cannot be seen, except by matching the data between hashed characters and other hashed characters.
- 9) *Binary Encoding (BE)*. Binary encoding is a combination of Hash encoding and One-Hot encoding. In this case, the categorical features are first converted to numeric using the Ordinal Encoder. Then the number is converted to a binary number and divided into different columns.
- 10) *Dummy Encoding (DE)*. Dummy coding scheme is like one-hot encoding. In the case of One-Hot encoding, N categories in a variable use N binary variable. The Dummy encoding is slightly improved over One-Hot encoding because its N-1 features represent N labels/categories.

2.2 Multiclass Classification

Several algorithms can be applied to the case of multiclass classification. Some of them are Decision Tree, Naive Bayes, K-Nearest Neighbor, Random Forest, Logistic Regression, and Support Vector Machine. A brief explanation of each algorithm is as follows:

- 1) *Decision Tree (DT)*. A decision tree consists of nodes and branches. Nodes can be divided into root nodes (primary nodes in the tree), decision nodes (conditionally sub-nodes), and leaf nodes (no longer branching nodes). Because the decision tree follows an if-then-else structure, each node uses an independent variable to divide into two or more branches. For categorical variables, the categories are used to determine the segregation of nodes, and for continuous variables, the algorithm generates several thresholds that act as decision rules [18]
- 2) *Naïve Bayes (NB)*. NB Classifier is a classification method that is rooted in Bayes theorem. The main character is a strong assumption of independence from each event. According to Olson [19], each decision class is determined by the probability that the decision class is true. The probabilities involved in producing the final estimate as to the sum of the frequencies from the decision table.
- 3) *K-Nearest Neighbor (KNN)*. KNN is a classification algorithm based on k nearest neighbors, and k is the number of nearest neighbors. The most common nearest neighbor search technique is using the distance formula. Distance formulas can use Euclidean, Hamming, Manhattan, or Minkowski.
- 4) *Random Forest (RF)*. Random forest is a combination of several models of decision trees to make one model. The more DT used the better accuracy. The decision of classification is taken based on the voting results of the formed tree.
- 5) *Logistic Regression (LR)*. Logistic regression is a model to classify objects based on probability thresholds. For example, in the case of a binary class, if the probability value is more than 0.5, it will be rounded to 1, which means that the response classification is in the event class. If the probability value is less than or equal to 0.5, it will be rounded to 0, meaning the response classification is in the non-event class. For multiclass classification, the LR model is known as multinomial logistic regression. It formed a separate binary logistic regression model for each response (class) category dummy variable. For example, if it has k class categories, it will produce a k-1 binary LR model. Each model is a probability of the class compared to the reference class.
- 6) *Support Vector Machine (SVM)*. The initial concept of SVM was to support binary classification and separate data points into two classes. So, the same principle can be used for multiclass classification by breaking the multiclass problem into several binary classification problems. Two approaches are often used in handling multiclass cases for SVM, namely the One-vs-One (OVO) and One-vs-All (OVA). *One versus One*. In the OVO approach, it takes a hyperplane to separate any two classes by ignoring the data points of the other classes. This means the split only takes into account the data points of the two classes in each classifier function. The classification function for a new object is:

$$\hat{f}(\mathbf{x}_{new}) = \text{sign}(\mathbf{x}_{new}^T \hat{\mathbf{w}}^{rs} + \hat{b}^{rs} - (1 + \xi^{rs})) \quad (2)$$

where,

$$\hat{\mathbf{w}}^{rs} = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i, \quad b^{rs} = \frac{1}{n_{sv}} \left(\sum_{i=1}^{n_{sv}} \frac{1}{y_i} - (\mathbf{x}_{new}^T \hat{\mathbf{w}}^{rs}) \right)$$

By using a kernel trick to map the original predictor variable to a higher dimension then optimization function in (2) can be defined as (3):

$$\min_{\mathbf{w}^{rs}, b^{rs}, \xi^{rs}} \frac{1}{2} (\mathbf{w}^{rs})^T \mathbf{w}^{rs} + C \sum_i \xi_i^{rs} \quad (3)$$

s.t

$$(\mathbf{w}^{rs})^T \phi(\mathbf{x}_i) + b^{rs} \geq 1 - \xi_i^{rs}, \quad y_i = r$$

$$(\mathbf{w}^{rs})^T \phi(\mathbf{x}_i) + b^{rs} \leq -(1 - \xi_i^{rs}), \quad y_i = s$$

The decision of an i -th object to enter a class uses a voting strategy.

One versus All. In the OVA approach, a strategy is used to create a hyperplane to separate classes and others at once. This means the separation takes into account all data points and then divides them into two groups, namely a group for class data points and a group for all other class data points. a new observation can be classified using the following classification measures:

$$\hat{f}(\mathbf{x}_{new}) = \text{sign}(\mathbf{x}_{new}^T \hat{\mathbf{w}}^r + \hat{b}^r - (1 + \xi^r)) \quad (4)$$

where,

$$\hat{\mathbf{w}}^r = \sum_{i=1}^n \hat{\alpha}_i y_i \mathbf{x}_i, \quad b^r = \frac{1}{n_{sv}} \left(\sum_{i=1}^{n_{sv}} \frac{1}{y_i} - (\mathbf{x}_{new}^T \hat{\mathbf{w}}^r) \right)$$

Where $\mathbf{x}_i$ is the support vector, $\mathbf{x}_{new}$ is the classified data while α_i is the Lagrange multiplier, b^r is the bias, and n_{sv} is the number of support vectors. For non-linear separable cases, the kernel trick $\mathbf{x}_i \rightarrow \phi(\mathbf{x}_i)$ can be obtained.

$$\min_{\mathbf{w}^r, b^r, \xi^r} \frac{1}{2} (\mathbf{w}^r)^T \mathbf{w}^r + C \sum_i \xi_i^r \quad (5)$$

has the following constraints:

$$(\mathbf{w}^r)^T \phi(\mathbf{x}_i) + b^r \geq 1 - \xi_i^r, \quad y_i = r$$

$$(\mathbf{w}^r)^T \phi(\mathbf{x}_i) + b^r \leq -(1 - \xi_i^r), \quad y_i \neq r$$

$$\xi_i^r \geq 0$$

The decision of an i -th object into a class r can be predicted by using the highest confidence score (**cs**) of the input vector $\mathbf{x}_i$ for each of the parameter vectors $\mathbf{w}^r$ and b^r as formulated in (6).

$$\mathbf{cs} = f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + \mathbf{b} - (1 + \xi) \quad (6)$$

2.3 Handling Imbalanced Class

Class imbalance is a common problem in machine learning classification, with a disproportionate ratio in each class. It can be found in medical diagnostics, spam filtering, fraud detection, emotion classification, and more. For example, fraud detection in banking, the rare occurrence possible is only 1%. Most machine learning algorithms work unsuccessfully with unbalanced datasets. Several techniques [20],[21],[22] that can be used to overcome classification performance on unbalanced data are:

- 1) *Evaluation Metrics.* Applying inappropriate evaluation metrics to models on unbalanced data can be misleading. If a model with outstanding accuracy, for example, 99.8% in the majority class, and 0% in the minority class, it cannot provide valuable information to predict rare events. What is needed is the model's ability to predict rare (minority) events. Thus, the accuracy is not appropriate to be used for evaluation measures on unbalanced datasets. In this case, other alternatives of model evaluation metrics can be used is F1-Score [23] with the formula as in (7).

$$\text{F1-Score} = \frac{2 \times \text{AP} \times \text{AS}}{\text{AP} + \text{AS}} \quad (7)$$

where,

AP : average precision

AS : average sensitivity (recall)

- 2) *Resampling Dataset*. Another way is to make the dataset balanced by resampling[24], and [25]. Two approaches that can be taken are the undersampling and oversampling techniques. Under-sampling is done by reducing the amount of data from the majority class. On the other hand, oversampling is done by increasing the amount of data from the minority class.
- 3) *Stratified k-Fold Cross-Validation*. The stratified k -fold cross-validation (SCV) is an extension of the cross-validation technique. It is usually used for imbalanced classification problems. This method keeps the class ratio in the k -fold as the ratio in the original data set.

3 METHOD

In general, the methodological framework for problem-solving can be described in Fig 1. As shown in Figure 1, this study uses public data with classification cases and predictors of categorical type and the number of classes is not balanced. The treatment applied for the transformation of categorical data into numeric using 10 encoding methods that already exist. Datasets that already have numeric type predictors are divided into training and testing datasets using the k -fold cross validation approach with folds of 3,5 and 10. The modeling process is carried out using 6 types of classification algorithms and evaluating their performance using the F1-Score and AUC values.

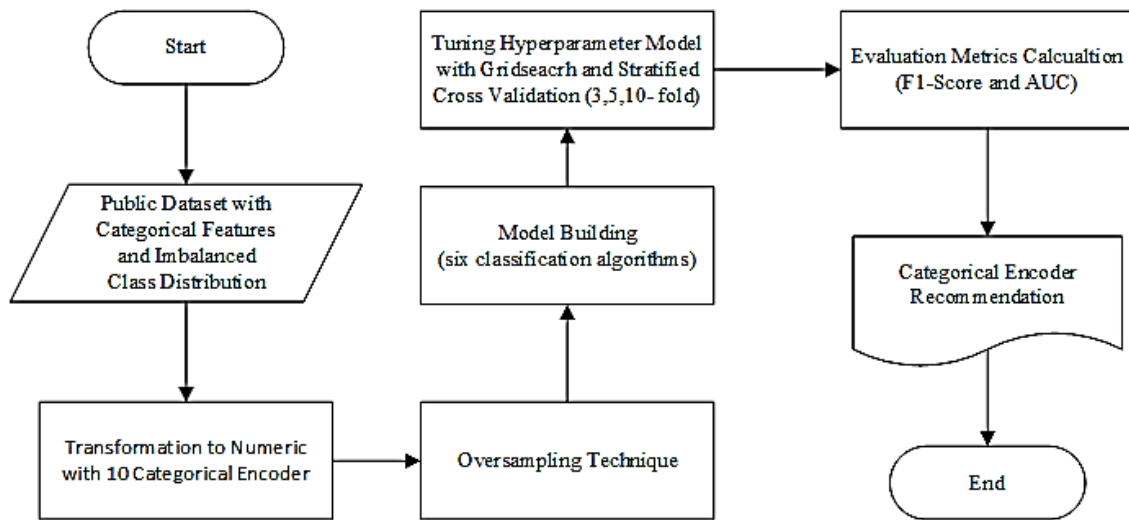


Figure 1. The research methodological framework

3.1 Dataset

The data used in this study consisted of 4 public datasets taken from the UCI Machine Learning Repository, namely car evaluation, nursery, lymphography, and balance scale data. The structure and characteristics of each dataset can be seen in Table I.

Table 1. Public Dataset with Imbalanced Class

Dataset	Features Type	Features Number	Class Category	N
Car Evaluation	Categoric	6	4	1728
Lymphography	Categoric	18	4	148
Nursery	Categoric	8	5	12960
Balance Scale	Categoric	4	3	625

The proportion of each class in all datasets can be explained visually through a histogram, as shown in Figure 2. Based on the histogram in Fig.2, the smallest proportions for the car evaluation, lymphography, balance scale, and nursery datasets were 3.76%, 1.35%, 7.84%, and 0.02%, respectively. It can be said that Nursery, Lymphography, Car Evaluation and Balance Scale dataset have severe imbalanced class with different proportion of minority class.

3.2 Pre-Processing

Data pre-processing is the first step to converting raw data collected from various sources into cleaner information that can be used for further processing. A common problem handled at the pre-processing stage is missing values, noise data, and inconsistent data.

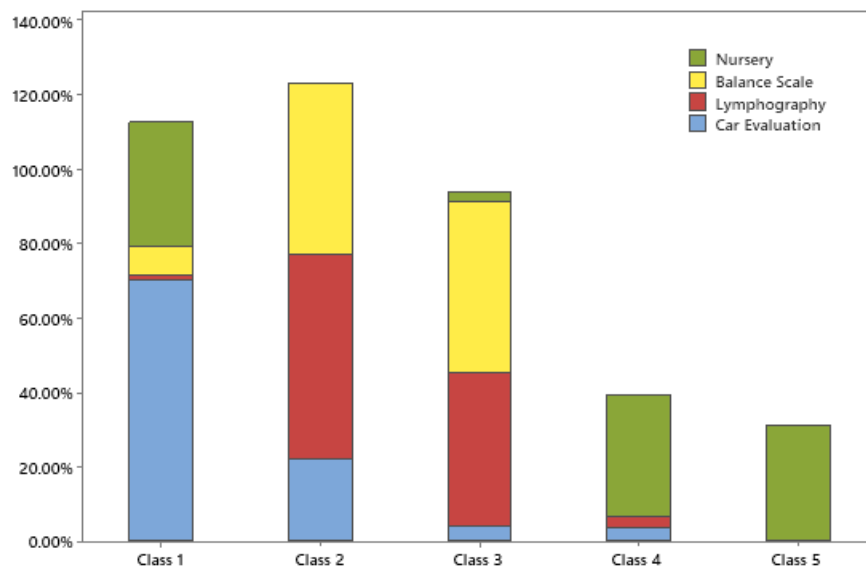


Figure. 1. Imbalance distribution of class variable in datasets

The method used to overcome missing values and noise is called data cleaning. After data cleaning, then continued to the transformation stage. The data will be converted into a form needed by the analytical method. For this study, it is necessary to transform categorical data to numeric. Data conversion is tried with several categorical encoders. Table 2 describes the encoding method used at this stage.

Table 2. The Number of Features Before and After Encoding

Categorical Encoding	Features Before Encoding				Features After Encoding			
	Car	Lym	Nur	Bal	Car	Lym	Nur	Bal
Ordinal	6	18	8	4	6	18	8	4
Nominal	6	18	8	4	6	18	8	4
Target	6	18	8	4	6	18	8	4
Frequency	6	18	8	4	6	18	8	4
Dirichlet	6	18	8	4	24	71	41	12
Leave One Out	6	18	8	4	6	18	8	4
Binary	6	18	8	4	18	41	24	16
Hashing	6	18	8	4	8	8	8	8
Dummy	6	18	8	4	15	41	19	16
One Hot	6	18	8	4	21	59	27	20

3.3 Model Building

Several multiclass classification models are applied using six algorithms, namely DT, KNN, NB, LR, RF, and SVM. In the hyperparameter optimization process for each model, the Grid Search method is used to get the best performance with range of parameters setting as in Table 3.

Table 3. Range Value of Hyperparameter Tuning using Grid Search

Classification Algorithm	Parameters Range
Decision Tree	Criterion: ['gini', 'entropy'], Min Samples Leaf: [0.1, 1, 10, 100, 500] Max Depth: [1, 10, 100, 500], Min Samples Split: [1, 10, 100, 500]
K Nearest Neighbour	N- Neighbors: [1, 10, 100, 500], Weights: ['uniform', 'distance'] Metric: ['euclidean', 'manhattan']
Naïve Bayes	Priors: [None], Var Smoothing: np. logspace (0, -9, num=100)
Logistic Regression	Penalty: ['l1', 'l2'], C: [1.0, 0.5, 0.1], Solver: ['liblinear']
Random Forest	Min Samples Leaf: [1, 10, 100, 500], Max Depth: [1, 10, 100, 500] Min Samples Split: [1, 10, 100, 500]

Classification Algorithm	Parameters Range
Support Vector Machine	Kernel: ['rbf'], C= [0.1, 1, 10, 100], Gamma = [1, 0.1, 0.01] Decision Function Shape: ['ovo','ovr']

The experiment was carried out using the Intel Core i7 Gen 7 computer specifications with 64 GB RAM. In this experiment, no observations were made regarding the running time and memory used for each categorical data encoding method in each classification algorithm. This is because the main objective of this research is to select a categorical data encoding method that can contribute to increasing the accuracy (F1 Score) of the classification algorithm performance, especially in cases of multiclass imbalance with categorical features.

4 RESULTS AND DISCUSSION

4.1 Performance Analysis based Classification Algorithm

In general, the performance of multiclass classification improved on training-testing validation by 10-fold over the F1-Score measure. To find out which algorithm has superior performance, it can be seen in Table 4.

Table 4. Descriptive Statistics Value F1-Score Classification of Multiclass Imbalanced Dataset based on Type of Classification Algorithm with 10-fold

Classification Algorithms	N	Minimum	Maximum	Mean	Std. Deviation
DT	40	0.423	0.941	0.769	0.159
KNN	40	0.426	0.947	0.801	0.137
LR	40	0.452	0.978	0.821	0.134
NB	40	0.466	0.934	0.799	0.129
RF	40	0.444	0.944	0.779	0.146
SVM	40	0.392	0.995	0.897	0.143
Valid N (listwise)	40				

Table 4 shows that SVM has the highest average F1-Score of 0.897 compared to other classification algorithms. To strengthen this conclusion, statistical tests were carried out for more than two independent samples from the calculation results of the average F1-Score in the six classification algorithms. In the analysis of the results of each experiment, it was found that the validation with 10 folds got the highest performance value. The Kruskal Wallis Test approach, which is a nonparametric test as an alternative to the One Way Anova test, was carried out because the assumption of normality of the data was not met. This Kruskal Wallis test is based on rank which aims to determine whether there are statistically significant differences between two or more groups of independent variables. The test decision shows the significance value is less than 5%. It means that there is a significant difference from the average value of the F1-Score based on the type of classification algorithm used.

4.2 Performance Analysis based Categorical Encoder

Furthermore, an analysis will be carried out on the effect of the type of transformation of categorical variables on classification performance. The data used are 4 datasets with 6 types of classification algorithms. Thus, for each type of transformation has a sample of 24 data. Descriptive statistics of the distribution of data can be seen in Table 5.

Table 5. Descriptive Statistics Value F1-Score Classification of Multiclass Imbalanced Dataset based on Type of Categorical Encoder with 10-fold

Categorical Encoder	N	Minimum	Maximum	Mean	Std. Deviation
OE	24	0.679	0.995	0.879	0.080
NE	24	0.636	0.995	0.857	0.102
FE	24	0.392	0.942	0.659	0.204
DRE	24	0.675	0.987	0.880	0.073
TE	24	0.701	0.995	0.879	0.078
DE	24	0.512	0.989	0.800	0.146
HE	24	0.493	0.868	0.643	0.103
OHE	24	0.609	0.993	0.848	0.103
BE	24	0.656	0.995	0.852	0.100
LE	24	0.523	0.995	0.812	0.161
Valid N (listwise)	24				

The results of the statistical calculation of the non-parametric test concluded that there were differences in treatment with different types of categorical variable transformation methods for the F1-Score with a significance value of less than 5%. Based on the ranking results of the average F1-Score value, it is found that the three types of categorical variable transformations that rank at the top are Ordinal Encoding, Dirichlet Encoding and Target Encoding. The choice of this type of transformation can be related to the proportion of the level of imbalance in the dataset class used. To support the selection of the appropriate transformation method, Fig. 3 is a comparison chart of the three methods above using different datasets for the SVM algorithm. it can be concluded that for the distribution of class data that has the smallest proportion value of less than 3%, the transformation method with the Dirichlet function approach gives the best effect on the results of the F1-Score measurement. Meanwhile, for class data whose smallest proportion is more than 3%, the transformation method with Ordinal Encoding gives the best prediction performance.

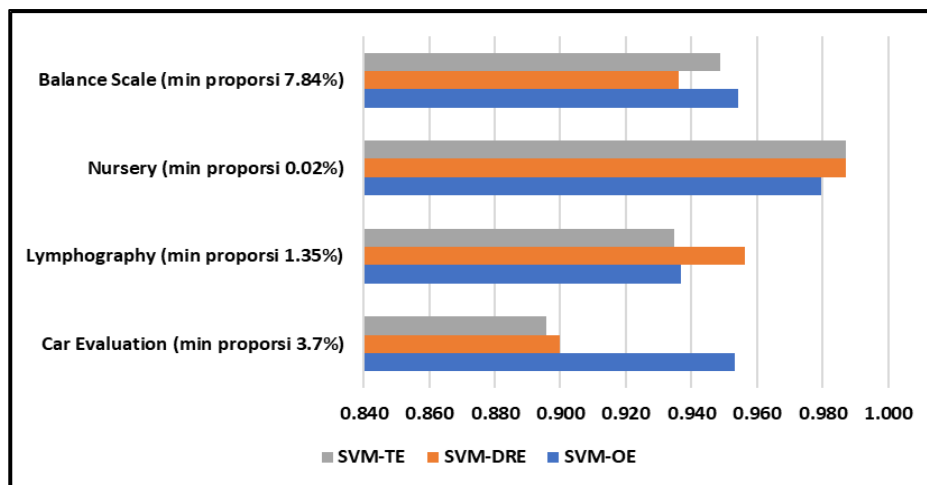


Figure 2. Comparison of Average F1-Score Values for OE, DRE, and TE on the Benchmarking Dataset

4.3 Performance Analysis based Multiclass Decision Strategies

In the results of the comparative analysis of classification algorithms, it is concluded that SVM has a higher performance than other algorithms. The transformation method with the Dirichlet function or number labels in order (ordinal) can be used based on the analysis of the results of the comparison of the categorical encoding method. Thus, in this section only analyze the AUC value with OVO and OVA approaches for all transformation methods.

Table 6. Calculation of Mean Rank on AUC Data for OVO and OVA Strategies with 10-Fold

Decision Strategies		N	Mean Rank	Sum of Ranks
AUC	OVO	40	39.23	1569
	OVA	40	41.78	1671
	Total	80		

The decision of statistical testing regarding whether there is a difference between the use of OVO and OVA strategies in the multiclass classification for imbalanced data using the AUC measure, it was found that there was no significant difference for the use of OVO and OVA strategies. However, the results of ranking the AUC performance values between the OVO and OVA strategies in table 7 show that the OVA ranking is higher than the OVO ranking, thus the OVA strategy is recommended.

5 CONCLUSION

The conclusion from the results of this study is that there are three groups of encoding methods recommended as transformation techniques for categorical to numerical features. The first group is called the encoding label, which consists of the Ordinal and Nominal Encoder. Secondly, based on conditional probability calculation between class and categorical variable levels, namely Target and Dirichlet encoder. The third group is the transformation method by increasing the number of features using binary concepts, namely Dummy, One-Hot, and Binary Encoder. The three groups produced excellent predictive performance. For datasets with severe imbalanced levels, it is recommended to use probability-based or binary-based encoding.

REFERENCES

- [1] S. J. Russell and P. Norvig, *Artificial Intelligence A Modern Approach*, vol. 48, no. 1. Pearson Education, 2021. [Online]. Available: www.pearsonglobal editions.com
- [2] M. Mohri, A. Rostamizadeh, and A. Talwaker, *Foundation Machine Learning*, 2nd Editio. London: The MIT Press, 2018.
- [3] Osisanwo, Akinsola, Awodele, Hinmikaiye, Olakanmi, and Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128–138, 2017, doi: 10.14445/22312803/ijctt-v48p126.
- [4] H. Bisgin *et al.*, "Comparing SVM and ANN based Machine Learning Methods for Species Identification of Food Contaminating Beetles," *Sci Rep*, vol. 8, no. 1, pp. 1–12, 2018, doi: 10.1038/s41598-018-24926-7.
- [5] H. Z. M. Shafri and F. S. H. Ramie, "A comparison of support vector machine and decision tree classifications using satellite data of Langkawi Island," *Information Technology Journal*, vol. 8, no. 1, pp. 64–70, 2009. doi: 10.3923/itj.2009.64.70.
- [6] Z. Zheng, Y. Cai, Y. Yang, and Y. Li, "Sparse weighted naive bayes classifier for efficient classification of categorical data," *Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018*, pp. 691–696, 2018, doi: 10.1109/DSC.2018.00110.
- [7] H. C. Rustamaji, O. S. Simanjuntak, S. F. Luhrie, B. Yuwono, and Juwairiah, "Categorical Data Classification based on Fuzzy K-Nearest Neighbor Approach," *Proceeding - 2019 5th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Cyber Physical System, ICSITech 2019*, pp. 171–175, 2019, doi: 10.1109/ICSITech46713.2019.8987477.
- [8] M. Azmi, G. C. Runger, and A. Berrado, *Interpretable regularized class association rules algorithm for classification in a categorical data space*, vol. 483. Elsevier Inc., 2019. doi: 10.1016/j.ins.2019.01.047.
- [9] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *J Big Data*, 2020, doi: 10.1186/s40537-020-00305-w.
- [10] Y. Wang and W. Xu, "Leveraging Deep Learning with LDA-based Text Analytics to Detect Automobile Insurance Fraud," *Decis Support Syst*, vol. 105, pp. 87–95, 2018, doi: 10.1016/j.dss.2017.11.001.
- [11] O. E. Ogundijo, D. He, and L. Parida, "Performance evaluation of different encoding strategies for quantitative genetic trait prediction," *2015 IEEE 5th International Conference on Computational Advances in Bio and Medical Sciences, ICCABS 2015*, pp. 1–6, 2015, doi: 10.1109/ICCABS.2015.7344715.
- [12] F. Pargent, "A Benchmark Experiment on How to Encode Categorical Features in Predictive Modeling," 2019.
- [13] A. Slakey, D. Salas, and Y. Schamroth, "Encoding categorical variables with conjugate bayesian models for wework lead scoring engine," *ArXiv*, pp. 1–15, 2019.
- [14] N. Lee and J. M. Kim, "Conversion of Categorical Variables into Numerical Variables via Bayesian Network Classifiers for Binary Classifications," *Comput Stat Data Anal*, vol. 54, no. 5, pp. 1247–1265, 2010, doi: 10.1016/j.csda.2009.11.003.
- [15] H. Li, R. Yuan, W. Peng, Y. Liu, and H. Z. Huang, "Bayesian Inference of Weibull Distribution Based on Probability Encoding Method," *ICQR2MSE 2011 - Proceedings of 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, no. 1, pp. 365–369, 2011, doi: 10.1109/ICQR2MSE.2011.5976632.
- [16] H. Han, X. Zhu, and Y. Li, "EDLT: Enabling Deep Learning for Generic Data Classification," *Proceedings - IEEE International Conference on Data Mining, ICDM*, vol. 2018-Novem, pp. 147–156, 2018, doi: 10.1109/ICDM.2018.00030.
- [17] Z. Yin and Y. Shen, "On the Dimensionality of Word Embedding," *NIPS Conference*, no. NeurIPS, 2018.
- [18] S. Raschka, D. Julian, and J. Hearty, *Python: Deeper Insights into Machine Learning*, 1st ed. Birmingham: Packt Publishing All, 2016.
- [19] D. L. Olson and D. Delen, *Advanced data mining techniques [electronic resource]*. Springer, 2008.
- [20] R. Luo *et al.*, "Feature Learning with a Divergence-Encouraging Autoencoder for Imbalanced Data Classification," *IEEE Access*, vol. 6, pp. 70197–70211, 2018, doi: 10.1109/ACCESS.2018.2879221.
- [21] R. Ghorbani and R. Ghousi, "Comparing Different Resampling Methods in Predicting Student's Performance Using Machine Learning Techniques," *IEEE Access*, vol. 8, 2020.
- [22] T. Al-shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, 2021, doi: 10.3390/e23101258.
- [23] B. S. Arkok and A. M. Zeki, "Classification of Quranic topics based on imbalanced classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, p. 678, May 2021, doi: 10.11591/ijeecs.v22.i2.pp678-687.
- [24] A. K. Hamoud *et al.*, "A prediction model based machine learning algorithms with feature selection approaches over imbalanced dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 2, pp. 1105–1116, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp1105-1116.
- [25] Hartono, E. Ongko, and Y. Risyani, "Combining feature selection and hybrid approach redefinition in handling class imbalance and overlapping for multi-class imbalanced," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, pp. 1513–1522, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1513-1522.

Cluster Relabelling-based Classification for Single Tuition Fee Determination in Public University

Wiyli Yustanti^{‡*}, Nur Iriawan[‡], Irhamah[‡]

[‡]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember

^{*}Department of Informatics, Faculty of Engineering, Universitas Negeri Surabaya
Surabaya, Indonesia

wiyliyustanti@unesa.ac.id, nur_i@statistika.its.ac.id, irhamah@statistika.its.ac.id

Abstract— Public universities in Indonesia administer yearly entrance examinations via two pathways: the achievement path and the computer-based exam. Prospective students who pass the selection process are subsequently provided with tuition fee information based on their socioeconomic status. This study gathered data from the State University of Surabaya utilizing thirteen socio-economic characteristics as predictors of the tuition fee group (8 categories). The results of the data exploration indicate that the distribution of data based on tuition fee groupings is not distinct. Consequently, it is important to relabel the cluster with a more appropriate grouping structure. The clustering procedure is carried out using the K-Mean algorithm and relabeling is accomplished using confidence interval mapping of PCA components between the original label and the new label. Using six classification algorithms, namely DT, KNN, NB, LR, RF, and SVM, it was discovered that there was a considerable improvement in the prediction accuracy of the original data in comparison to the labeling data.

Keywords—K-means, clustering, relabeling, classification, dimension reduction, PCA, Tuition Fee

I. INTRODUCTION

The fact that students have to pay various fees while obtaining a higher education was a primary consideration process that led to the implementation of Indonesia's Single Tuition Fee (STF) policy at Public Universities (PU). Consequently, a standard tuition fee is required to ensure students have the financial means to pay. The student's ability to pay must also be considered when determining tuition costs. The Ministry of Education and Culture established the Single Tuition Fee policy for Public Universities in 2013. In addition to being motivated by the emergence of non-tuition fees, the determination of the STF policy reflects the implementation of the mandate of Article 88 of Law No. 12 of 2012 concerning Higher Education to establish a standard unit of operational costs in determining the costs borne by students. In addition, the article stipulates that the costs in question must be proportional to the economic capabilities of students, parents, and other parties who fund them. Based on this, STF rates at PU are established in phases and separated into numerous groups. The inference is that students with poor economic circumstances should have low STF rates. In contrast, economically able students should receive high STF rates to assist those with low STF rates. These classification tasks belong to the field of unsupervised learning from a scientific perspective. The K-Means clustering algorithm is a method that is frequently utilized.

Theoretically, a good cluster is a cluster that has high homogeneity (similarity) within the cluster and high heterogeneity (difference) between clusters. By having a good group structure, STF group predictions for new students are expected to be more accurate. All state

universities have adopted a variety of socioeconomic indicators to identify the STF category and used different mechanisms to formulate it. The previous study [1] identified variables that impacted students' socioeconomic circumstances, 16 of the 43 initial variables substantially affected how students' STF was calculated. Furthermore, the STF group will be modelled using a feature selection method based on the SVM algorithm in the following study [2] and has an accuracy rate of 81% with the F1-Score metric. Several researchers in the following year developed a decision-making system based on the fuzzy-c means algorithm [3], [4] and [5]. A decision tree-based algorithm in determining the STF group is also applied through comparison experiments with the J48, ID3 and Naive Bayes algorithms. This study [6] concluded that J48 had better accuracy than the ID3 and Naive Bayes algorithms, with an accuracy value using the F1-Score of 91.1%. Before conducting supervised learning-based modelling, the study's findings neglect the conditions of variety both inside and across clusters. Every study supposes that the outcomes of grouping with the university formulations are enough and sufficient for use as training and testing materials to forecast the STF level of new students. Therefore, the objectives of this study are to: (1) assess the quality of the existing data STF grouping results obtained using the university formula; (2) implement the STF grouping labeling process based on the cluster goodness index; and (3) conduct experiments for classification algorithm-based modeling to compare the accuracy before and after the labeling process.

II. RELATED WORKS

This section will briefly explain the basic theory of dimension reduction with Principle Component Analysis (PCA), K-Means algorithm, validation and evaluation of clustering results. Furthermore, the existing methods from related works about the labelling process in clustering and the classification algorithm-based clustering will be explained at the end of this part.

A. Principle Component Analysis

Principal component analysis is one of the most widely used strategies for dimensionality reduction. The central concept of principal component analysis is that it generates new features and projects the original data onto these features, which are a linear combination of the original features, with the goal of maximizing the total variation in the data, that is retaining as much information as possible. The processes for calculating the Principal Components are as follows:

- 1) Standardize variables
- 2) Calculating the Covariance Matrix in order to determine Correlation.

- 3) Determine the Principal Components by calculating the Eigenvalues and Eigenvectors of the Covariance Matrix.
- 4) Determine which Principal Components to retain for further research based on differences in Components utilizing a Scree Plot.
- 5) Transform data along the Principal Components axis.

PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the most significant variance by some scalar projection of the data lies on the first coordinate (termed the first principal component), the second largest variance on the second coordinate [7].

B. K-Means Algorithm

K-means, originally used in signal processing, is a vector quantization technique known as clustering to divide a set of n observations into k clusters, with each observation belonging to the cluster with the closest mean (also known as the cluster centroid or cluster centers). The procedure of the K Means algorithm as follows [8]:

- 1) Determine a random cluster center and the number of clusters (k).
- 2) Determine the distance from each data point to the cluster's center.
- 3) Assign the data to clusters with the least distances.
- 4) Determine the cluster center.
- 5) Keep going back and forth between steps 2 and 4 until no more data is moving to the other clusters.

C. Cluster Validation and Evaluation

The silhouette value [9] is one of metric to evaluate how closely related an object is to its cluster (separation). An object is said to be well matched to its cluster and poorly matched to nearby clusters if the silhouette value $s(i)$ is high, which ranges from 1 to +1. The clustering design is suitable if most items have high values. There may be many or few clusters in the clustering configuration if a large number of points have low or negative values. Any distance measurement, including the Manhattan or Euclidean distance, can be used to determine the silhouette's index. Assumed a point i in the C_i cluster has a value of $a(i)$ which is the average value of the distance from that point to all points in the same cluster, then equation (1) shows the formula $a(i)$.

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (1)$$

Where $d(i, j)$ is the distance between the i -th and j -th point in the same cluster (C_i) and $|C_i|$ is the number of points in the C_i . The smaller value of $a(i)$ can be interpreted as a better cluster assignment. In contrast, $b(i)$ is the average distance between a point i and every other point outside the C_i cluster determined by the equation (2):

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|} \sum_{j \in C_J, i \neq j} d(i, j) \quad (2)$$

With equation (1) and (2), then form the $s(i)$ equation (3),

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ if } |C_i| > 1 \quad (3)$$

D. Cluster Labelling

In the case of text grouping analysis, cluster labeling issues are frequently encountered. The clarity labeling requires selection a human-readable descriptive label for the cluster produced by the document clustering method; ordinary clustering algorithms typically do not produce such labels. The cluster labeling algorithm looks at the document's contents by cluster to identify labels that summarize each cluster's subject and set them apart from one another. Research related to cluster labeling can be found at [10] it suggested approach examines interconnection paths between crucial locations separating various cluster contours and analyzes the topology of the function representing the Support Vector Clustering (SVC) cluster outlines. This method makes it possible to identify distinct clusters and link each point to its appropriate cluster. This study is connected to research findings of [11] about labelling for SVC. In the other study [12], a mechanism referred to as Maximal Resemblance Data Labeling (MARDL). It is proposed to assign each unlabeled data point to the correct cluster based on a new categorical grouping, called the N-Nodeset Importance Representative (NNIR). The NNIR represents the cluster with the importance of attribute value combinations. The cluster labeling approach that will be proposed in this study aims to improve the cluster structure so that it can be differentiated between clusters. Consequently, the labeling must be connected to the original data label because the data comes from the results of poor clustering towards better clustering.

E. Classification based Clustering

Labelled input data is required for modelling to predict a group from a set of items. Research [13],[14], [15],[16][17] contains previous studies demonstrating the labelling outcomes employed in the classification modelling procedure. Cluster-based data relabeling (CBDR) is a strategy introduced in the latest recent study [18] that enables linear classifiers to operate successfully on nonlinear data. CBDR divides the data set into multiple class-specific clusters that do not overlap and relabels the data by cluster. On the relabelled data, a linear classifier can be employed to seek cluster-based linear decision boundaries as opposed to class-based ones. Extensive trials have showed that CBDR can greatly improve the performance of linear classifiers and even outperform their nonlinear counterparts. CBDR has also been found to increase the classification performance of nonlinear classifiers, according to additional experiments. In both instances, the most substantial outperformance was observed with skewed data. In this study, the results of labeling in the clustering process will also be evaluated on many classification models, including Decision Tree (DT), K-Nearest Neighbor (KNN), Logical Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). It is expected that a substantial difference in the accuracy of the prediction results can be produced between the data with the original label and the new label based on the proposed method.

III. METHODOLOGY

This section will explain the research framework to be carried out, the dataset and the experimental design used.

A. Research Framework

This research begins by collecting a dataset on single tuition fee and then proceeds with data exploration. The results of data exploration will provide enlightenment regarding the treatment at the preprocessing stage. In accordance with Fig.1, after the pre-processing phase, cluster the data and calculate the silhouette index between the original group structure and the clustered group structure. The clustering procedure is carried out by the K-Means algorithm.

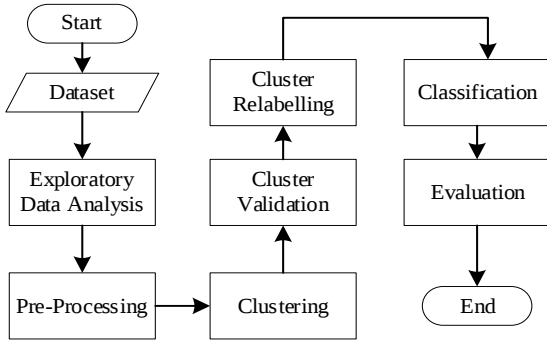


Fig. 1. Research Methodology

The results of clustering generate random labels for each cluster. Random labels can lead to misinterpretation if not addressed. Therefore, it is necessary to map the random label of the K-Means results into the original label (associated with the level of tuition fees). Utilizing the 95 % confidence level value of the first component as a result of dimension reduction in the preprocessing phase is the implemented approach. During the pre-processing phase, the original data were transformed using ordinal encoding and then validated for feasibility using The Kaiser-Meyer-Olki (KMO) and Bartlett statistical test to determine the viability of applying PCA. Bartlett's Test of Sphericity aims to test the hypothesis that the variables are uncorrelated within the population (H_0). If the p -value is small, we can rule out the null hypothesis and concur that there is at least one pair of correlated variables in the data, hence PCA is recommended. The results of the calculation of Chi-Square value and p -value for the Bartlett Test are 601308.601 and 0.000. Therefore, the PCA process can be continued. Meanwhile, the KMO is a measure of Sampling Adequacy (MSA), an index that assesses the accuracy of PCA. PCA is typically not advised if the MSA is less than 0.5 because there is not likely to be any reduction. The MSA value for our dataset is 0.5818, which is greater than 0.5. Consequently, applying PCA to our dataset is expected to result a reduction dimension and the extraction of relevant components.

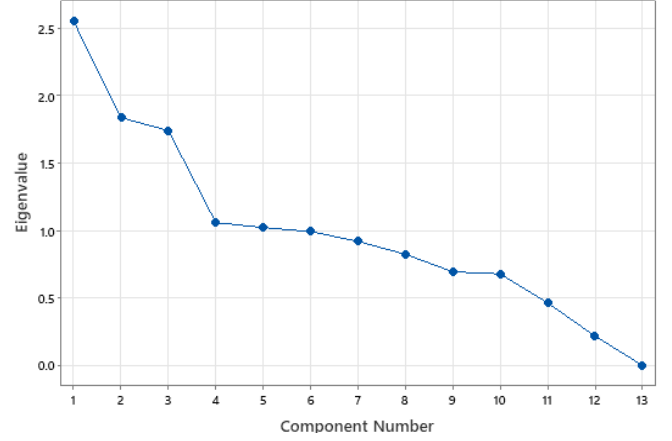


Fig. 2. Scree Plot of Dataset

Based on the eigenvalues in Fig. 2, it can be concluded that the predictor variable from the dataset used can be reduced to 5 components (eigenvalue >1). Before the PCA process is carried out, the data is standardized first. Furthermore, the five new variables from the PCA will be used as a new dataset for the clustering process. The total variation represented by the extracted components is 63.17%.

B. Dataset

The dataset used is single tuition fee data from 15875 students at the State University of Surabaya with categories 1 to 8 for the last five years, starting from 2017 to 2021. Category 1 has the lowest tuition fees, and category 8 has the highest tuition fee. The tuition fee referred to here is the amount of student payments each semester. For each expense category throughout a five-year period, Table 1 displays the distribution of students. It demonstrates how unequally distributed the student population in the group.

TABLE I. SINGLE TUITION FEE DATASET

Year	Single Tuition Fee Category							
	1	2	3	4	5	6	7	8
2017	20	103	294	1266	1367	36	0	0
2018	21	126	298	1481	1557	44	0	0
2019	54	240	306	830	999	501	48	0
2020	30	211	86	1112	964	355	35	1
2021	36	72	311	1122	1201	709	39	0
Total	161	752	1295	5811	6088	1645	122	1
Grand total								15875

The variables used in this study are variables that reflect the socio-economic characteristics of students' parents. The target variable is the tuition fee group variable (Y), and the predictor variables X_1 to X_{13} are socio-economic indicators which can be seen in Table II.

TABLE II. RESEARCH VARIABLES

Variable	Value	Type	Description
Y	1,2,3,4,5,6,7,8	Ordinal	Single Tuition Fee Level
X_1	1,2,3,4,5,6	Ordinal	Father's Occupation
X_2	1,2,3,4,5,6	Ordinal	Mother's Occupation
X_3	1,2,3,4,5	Ordinal	Father's Salary
X_4	1,2,3,4,5	Ordinal	Mother's Salary
X_5	1,2,3,4,5	Ordinal	Number of Dependent
X_6	1,2,3,4,5	Ordinal	Building Land Tax Billing
X_7	1,5,6	Ordinal	Number of Cars
X_8	1,4,5,6	Ordinal	Number of Motorbikes

X_9	1,2	Ordinal	Electricity Source
X_{10}	1,2,3	Ordinal	Residential Status
X_{11}	1,2,3,4,5	Ordinal	Electricity Billing
X_{12}	1,2,3,4	Ordinal	Mother's other income
X_{13}	1,2,3,4	Ordinal	Father's other income

C. Experiment Setup

An experimental design was carried out by utilizing a re-labelling dataset based on the PCA component to accomplish the third study goal. Table III shows the experimental design structure.

TABLE III. EXPERIMENT DESIGN

Dataset (p)	q -fold (q)	Classification Algorithm (r)					
		DT	KNN	LR	NB	RF	SVM
Original Labeled Data	3	m_{111}	...	...	...	...	m_{166}
	5	m_{121}	...	...	...	...	m_{126}
	10	m_{131}	...	m_{pqr}	...	...	m_{136}
Relabeled Data	3	m_{211}	...	...	...	...	m_{216}
	5	m_{221}	...	...	...	...	m_{226}
	10	m_{231}	...	...	...	...	m_{236}

As in Table III there is a m_{pqr} element, where index p is the type of dataset used, $p=1,2$, and index q is the fold level in the cross-validation process, where $q=3,5,10$, while r is the index of the classification algorithm where $r = 1,2,3,4,5,6$. In this study, m is a metric used to measure the goodness of classification predictions in the form of F1-Score or AUC values. The experiment was conducted using Jupyter-lab on a computer with an Intel Core i7 6th generation processor and 64 Gb of RAM. In hyperparameter optimization, the Grid Search approach is used with the following parameter ranges in Table IV.

TABLE IV. TUNING HYPERPARAMETER RANGE

Algorithm	Parameter	Range Value
DT	criterion	['gini', 'entropy']
	min_samples_leaf	[1, 10, 100, 500]
	min_samples_split	[1, 10, 100, 500]
KNN	n_neighbors	[1, 10, 100, 500]
	weights	['uniform', 'distance']
	metric	['euclidean', 'manhattan']
LR	penalty	['l1', 'l2']
	C	[0.1, 1, 10, 100]
	F1	[1.0, 0.5, 0.1]
	solver	['liblinear']
NB	prior	None
	smoothing	np.logspace(0,-9, num=100)
RF	min_samples_leaf	[1, 10, 100, 500]
	max_depth	[1, 10, 100, 500]
	min_samples_split	[1, 10, 100, 500]
SVM	kernel	['linear', 'rbf']
	C	[1, 0.1, 0.01]
	Gamma	[0.1, 1, 10, 100]
	Decision function shape	['ovo', 'ovr']

IV. RESULT AND DISCUSSION

This section will explain the results of the problem-solving to achieve the expected goals. The first goal is carried out through data exploration activities from STF data, the second is through a relabelling process using the first component of PCA, and the third is achieved by experimenting with six classification algorithms using the original dataset and the labelling result dataset.

A. Exploratory Data Analysis

The result of dimension reduction using PCA can be used to create visualizations between the first and second components based on their STF level labels. Fig. 3 shows that the clusters between STFs are not separated, and overlap occurs. If measured using the Silhouette index, a value of -0.054 is obtained, which means that the cluster indicate that an object is not in the right cluster. Because the value of the cluster goodness based on the Silhouette index is close to -1, it is necessary to re-cluster it so that it can increase the value of its Silhouette index. A new data cluster form is generated using the K-Means algorithm, as shown in Fig.4. The location of objects between clusters in Fig.4 is visually clearly separated. The problem is that the label given by K-Means is a random number with no meaning as an STF level. Therefore, it is necessary to do a strategy to map the random label given by K-Means with the original label. If the Silhouette index is calculated from the K-Means results, a value of 0.591(close to 1) is obtained, which means that the object is in the right cluster.

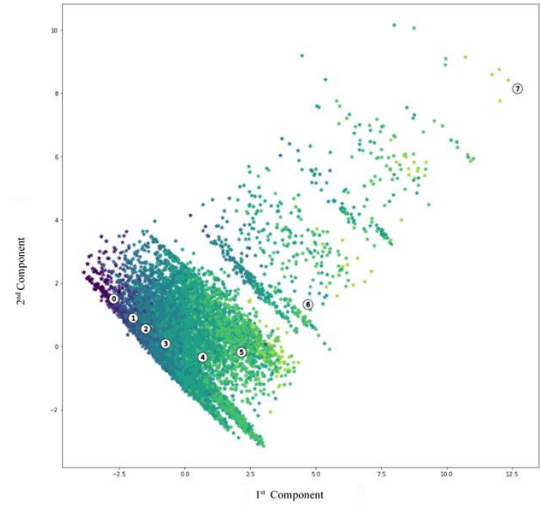


Fig. 3. Scatter Plot of Original Dataset based on PCA

The number of clusters (k) used in the K-Means algorithm is eight because there are eight tuition fees categories applied in Surabaya State University.

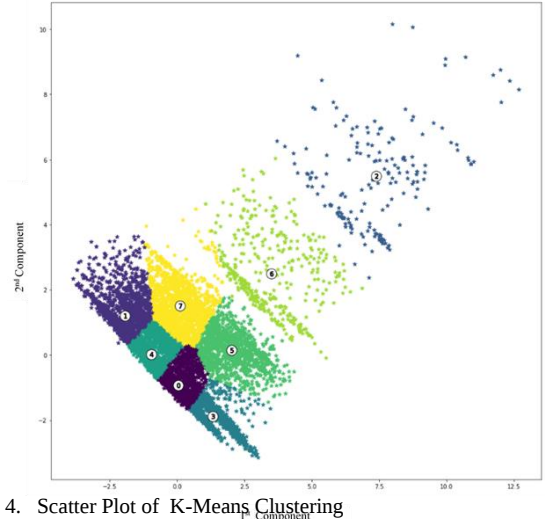


Fig. 4. Scatter Plot of K-Means Clustering

B. Relabelling Process

In the relabelling process, used the confidence interval value of the first PCA component. The steps involved in this relabelling process are as follows:

- 1) Calculating the Confidence Interval (CI) of the first component of PCA with a 95% confidence level for both the original data label and the K-Means data label.
- 2) Make a Box Plot for each CI.
- 3) Mapping Boxplot pattern from K-Mean label to origin label.
- 4) Update the K-mean label.

If the PCA results are rotated using the varimax method to see what variables are dominant in each component, the results can be seen in Table V. According to table V, the first component is dominated by variables X_{12} and X_{13} , which represent the parent's income in addition to the basic salary.

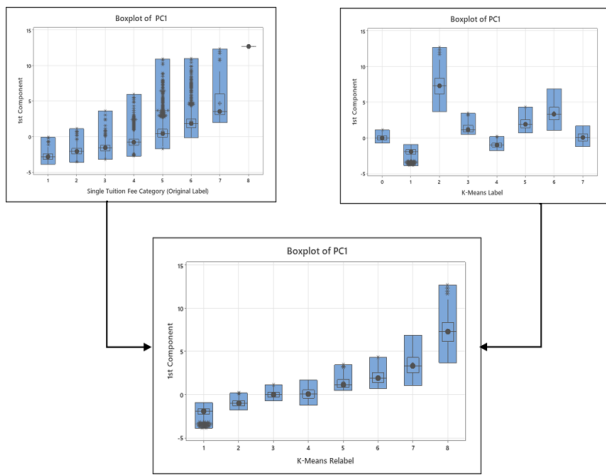


Fig. 5. Relabelling Cluster Scheme

The box plot in Fig.5 reveals that the higher the STF level, the greater the father's and mother's income. In contrast, the type of the father's occupation and the father's basic wage dominate the second factor.

TABLE V. VARIMAX ROTATION OF PCA

Variable	1 st Component	2 nd Componen
X_1	0.0609	0.5278
X_2	0.0682	-0.0312
X_3	0.0659	0.7815
X_4	0.0954	0.0697
X_5	-0.0400	-0.2243
X_6	0.0359	0.2769
X_7	0.1356	0.4277
X_8	0.0370	0.3381
X_9	-0.0018	-0.0051
X_{10}	0.0284	0.1591
X_{11}	0.0665	0.4540
X_{12}	0.9994	0.0347
X_{13}	0.9994	0.0347

C. Classification Modelling

By using the experimental design as shown in Table III, the calculation results for the average of F1-Score and AUC can be shown in Table VI and Table VII, respectively. In general, there are differences in results between the F1-Score measurements for data with the original label and data

with the new label. To support whether this difference is significant, it is necessary to conduct statistical tests for two paired samples. The selection of the appropriate test statistic must be checked at the normality distribution assumption of the measurement results for both the F1-Score and AUC. The normality test was carried out through the Kolmogorov-Smirnov test, and it was found that the F1-Score and AUC data were not normally distributed. Thus, the difference test between the F1-Score and AUC values before and after labelling used a nonparametric statistical approach.

TABLE VI. F1-SCORE FROM CLASSIFICATION MODEL

Dataset (p)	fold (q)	Classification Algorithm (r)					
		DT	KNN	LR	NB	RF	SVM
Original Labeled Data	3	82.98	83.11	58.48	59.65	84.29	81.03
	5	83.41	83.41	58.35	59.64	84.69	81.22
	10	83.63	83.49	58.47	59.69	84.85	81.58
Relabeled Data	3	99.50	98.98	92.82	95.54	99.65	99.85
	5	99.51	99.11	92.85	95.50	99.66	99.85
	10	99.56	99.15	92.93	95.56	99.73	99.60

If the data in Table VI is presented in graphical form in Fig.6, it can be seen that the average accuracy of the F1-Score in the dataset with the new label has a higher value than in the dataset with the original label.

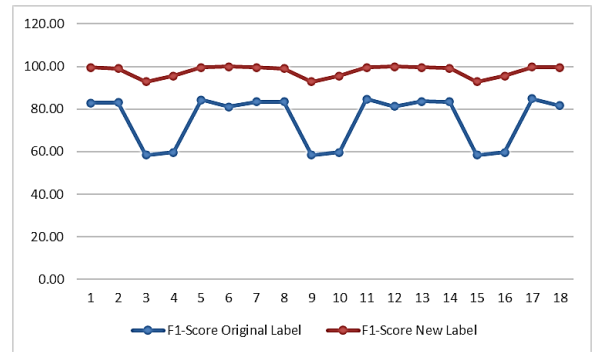


Fig. 6. F1-Score between Original Label and New Label dataset

The Wilcoxon statistical test of paired samples with the null hypothesis that there is no difference between the F1-Score of the dataset with the original label and the new label is carried out to support the conclusion that the results of the F1 Score average on the dataset of the new label are significantly better.

TABLE VII. WILCOXON SIGNED TEST OF F1-SCORE

Wilcoxon Signed Test Statistics ^a		Conclusion
	F1-Score New Label – F1-Score Original Label	
Z	-3.724 ^b	
Asymp. Sig. (2-tailed)	0.000	H ₀ Rejected

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Table VII shows that the p-value is less than 5%. It is concluded that the data support statistically that the results of the F1-Score dataset with the new label are different (better) from the dataset with the original label. As with the F1-Score, for the results of the AUC measurement between the original data and the data with the new label in Table VIII, a different test was carried out with the results as shown in Table IX.

TABLE VIII. AUC FROM CLASSIFICATION MODEL

Dataset (p)	fold (q)	Classification Algorithm (r)					
		DT	KNN	LR	NB	RF	SVM
Original Labeled Data	3	96.33	97.73	91.54	92.49	98.48	97.38
	5	96.51	97.84	91.54	92.48	98.55	97.44
	10	96.62	97.92	91.55	92.48	98.59	97.49
Relabeled Data	3	99.81	99.96	99.18	99.66	100.0	100.0
	5	99.81	99.63	99.17	99.66	100.0	100.0
	10	99.84	99.59	99.16	99.66	100.0	100.0

TABLE IX. WILCOXON SIGNED TEST OF AUC

Wilcoxon Signed Test Statistics ^a		Conclusion
	AUC New Label – AUC Original Label	
Z	-3.724 ^b	
Asymp. Sig. (2-tailed)	0.000	H ₀ Rejected

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

The test results show a significant difference between the AUC values before and after relabelling. This difference can be seen visually in Fig.8 that the AUC value in the current label data is higher than the original data.

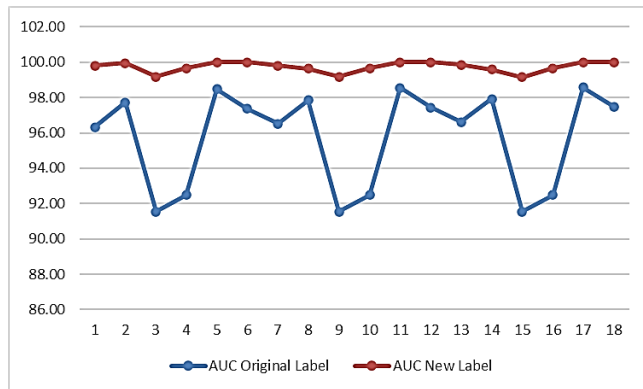


Fig. 7. AUC between Original Label and New Label dataset

V. CONCLUSION

The results of this study conclude that (1) the quality of the initial data cluster based on the Silhouette index shows that many objects are not in the right cluster position, so it needs to be re-clustered, and (2) the labelling process of the cluster results using K-Means can be done by mapping K-Means labels using CI values from PCA results, especially in the first component, (3) classification modelling experiments for data before and after labelling showed significant differences, and the accuracy results were much better.

REFERENCES

- [1] W. Yustanti and Y. Anistyasari, "A Polychoric Correlation to Identify the Principle Component in Classifying Single Tuition Fee Capabilities on the Students Socio-Economic Database," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 288, no. 1, p. 012150. doi: 10.1088/1757-899X/288/1/012150.
- [2] W. Yustanti, Y. Anistyasari, and E. M. Imah, "Determining student's single tuition fee category using correlation based feature selection and support vector machine," *Proceeding International Conference on Advanced Computer Science and Information Systems, ICACSIS 2017*, vol. January, pp. 172–176, 2018, doi: 10.1109/ICACSIS.2017.8355029.
- [3] Indrawati, Anwar, and N. Amalia, "Determination System of Single Tuition Group Using a Combination of Fuzzy C-Means Clustering and Simple Additive Weighting Methods," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 536, no. 1. doi: 10.1088/1757-899X/536/1/012148.
- [4] H. Syahputra, Sutrisno, and S. Gultom, "Decision Support System for Determining the Single Tuition Group (UKT) in State University of Medan Using Fuzzy C-Means," in *Journal of Physics: Conference Series*, Mar. 2020, vol. 1462, no. 1. doi: 10.1088/1742-6596/1462/1/012071.
- [5] A. W. Sugiyarto, R. Pamungkas, A. R. Rasjava, and A. M. Abadi, "Fuzzy Multi Attribute Decision Making (FMADM) Implementation for Classifying Student's Single Tuition Fee (UKT) Based on Android Applications," in *Journal of Physics: Conference Series*, Dec. 2019, vol. 1397, no. 1. doi: 10.1088/1742-6596/1397/1/012061.
- [6] T. F. Abidin, S. Rizal, T. M. Iqbalsyah, and R. Wahyudi, "Decision tree classifier for university single rate tuition fee system," *International Journal of Business Intelligence and Data Mining*, vol. 17, no. 2, pp. 258–271, 2020, doi: 10.1504/IJBIDM.2020.108764.
- [7] I. T. Jolliffe, *Principal Component Analysis*, Second Edition. New York: Springer, 2002.
- [8] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms and Applications*. American Statistical Association and the Society for Industrial and Applied Mathematics. 10, 2007. doi: 10.1017/CBO9781107415324.004.
- [9] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," 1987.
- [10] V. D'Orangeville, M. A. Mayers, M. E. Monga, and M. S. Wang, "Efficient cluster labeling for support vector clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 11, pp. 2494–2506, 2013, doi: 10.1109/TKDE.2012.190.
- [11] J. Lee and D. Lee, "An improved cluster labeling method for support vector clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 461–464, Mar. 2005, doi: 10.1109/TPAMI.2005.47.
- [12] H. L. Chen, K. T. Chuang, and M. S. Chen, "On data labeling for clustering categorical data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 11, pp. 1458–1471, 2008, doi: 10.1109/TKDE.2008.81.
- [13] A. A. Klaib, A. A. Milad, and M. A. Algaet, "A New Approach for Labelling XML Data," in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2021*, Sep. 2021, pp. 603–607. doi: 10.1109/3ICT53449.2021.9581352.
- [14] W. C. Sleeman IV et al., "A Machine Learning method for relabeling arbitrary DICOM structure sets to TG-263 defined labels," *Journal of Biomedical Informatics*, vol. 109, Sep. 2020, doi: 10.1016/j.jbi.2020.103527.
- [15] M. Sperrin, T. Jaki, and E. Wit, "Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models," *Statistics and Computing*, vol. 20, no. 3, pp. 357–366, 2010, doi: 10.1007/s11222-009-9129-8.
- [16] Z. Li, J. Li, Y. Liao, S. Wen, and J. Tang, "Labeling clusters from both linguistic and statistical perspectives: A hybrid approach," *Knowledge-Based Systems*, vol. 76, pp. 219–227, Mar. 2015, doi: 10.1016/j.knsys.2014.12.019.
- [17] R. Kusumaningrum and Farikhin, "An Automatic Labeling of K-means Clusters based on Chi-Square Value," in *Journal of Physics: Conference Series*, Mar. 2017, vol. 801, no. 1. doi: 10.1088/1742-6596/801/1/012071.
- [18] H. Wan, H. Wang, B. Scotney, J. Liu, and X. Wei, "Cluster-based Data Relabelling for Classification," *Information Sciences SSRN*, pp. 1–32, Jul. 2022, [Online]. Available: <https://ssrn.com/abstract=4169883>



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI
UNIVERSITAS NEGERI SURABAYA
Kampus Lidah, Jalan Lidah Wetan Unesa, Surabaya 60213
Telepon 031-99421834, 99421835, Faksimil : 031-99424002
Laman : www.unesa.ac.id

KEPUTUSAN REKTOR UNIVERSITAS NEGERI SURABAYA
NOMOR 632/UN38/HK/PP/2022

TENTANG

PENETAPAN PENERIMA PENELITIAN DASAR PERCEPATAN STUDI DOKTOR (LPPM)
TAHUN 2022

REKTOR UNIVERSITAS NEGERI SURABAYA,

- Menimbang : a. bahwa berdasarkan hasil seleksi desk evaluasi dan pemaparan proposal penelitian yang dilakukan oleh panitia seleksi, telah ditetapkan Penerima Penelitian Dasar Percepatan Studi Dalam Negeri dan Luar Negeri (LPPM) Tahun 2022;
- b. bahwa berdasarkan pertimbangan sebagaimana dimaksud dalam huruf a, perlu menetapkan Keputusan Rektor Universitas Negeri Surabaya Tentang Penetapan Penerima Penelitian Dasar Percepatan Studi Dalam Negeri dan Luar Negeri (LPPM) Tahun 2022;
- Mengingat : 1. Peraturan Pemerintah Republik Indonesia Nomor 37 Tahun 2009 tentang Dosen (Lembaran Negara Republik Indonesia Tahun 2009 Nomor 76, Tambahan Lembaran Negara Republik Indonesia Nomor 5007);
2. Peraturan Pemerintah RI Nomor 4 Tahun 2014 tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi (Lembaran Negara Republik Indonesia Tahun 2014 Nomor 16, Tambahan Lembaran Negara Republik Indonesia Nomor 5500);
3. Peraturan Menteri Keuangan RI Nomor 92/PMK.05/2011 tentang Rencana Bisnis dan Anggaran Serta Pelaksanaan Anggaran Badan Layanan Umum (Berita Negara Republik Indonesia Tahun 2011 Nomor 363);
4. Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi RI Nomor 15 Tahun 2016 tentang Organisasi dan Tata Kerja Universitas Negeri Surabaya (Berita Negara Republik Indonesia Tahun 2015 Nomor 889);
5. Peraturan Menteri Riset, Teknologi dan Pendidikan Tinggi RI Nomor 79 Tahun 2017 tentang Statuta Universitas Negeri Surabaya (Berita Negara Republik Indonesia Tahun 2017 Nomor 1858);

6. Keputusan Menteri Keuangan RI Nomor 50/KMK.05/2009 tentang Penetapan Universitas Negeri Surabaya Pada Departemen Pendidikan Nasional sebagai Instansi Pemerintah yang menerapkan Pengelolaan Keuangan Badan Layanan Umum;
7. Keputusan Menteri Riset, Teknologi dan Pendidikan Tinggi RI Nomor 461/M/KPT.KP/2018 tentang Pemberhentian dan Pengangkatan Rektor Universitas Negeri Surabaya Periode Tahun 2018-2022;

MEMUTUSKAN :

- Menetapkan : KEPUTUSAN REKTOR UNIVERSITAS NEGERI SURABAYA TENTANG PENETAPAN PENERIMA PENELITIAN DASAR PERCEPATAN STUDI DALAM NEGERI DAN LUAR NEGERI (LPPM) TAHUN 2022.
- KESATU : Menetapkan Penerima Penelitian Dasar Percepatan Studi Dalam Negeri dan Luar Negeri (LPPM) Tahun 2022 sebagaimana tercantum dalam Lampiran I dan II yang merupakan bagian tidak terpisahkan dari Keputusan Rektor ini.
- KEDUA : Dalam melaksanakan tugasnya sebagai Penerima Penelitian Dasar Percepatan Studi Dalam Negeri dan Luar Negeri (LPPM) Tahun 2022, wajib berpedoman pada ketentuan yang berlaku.
- KETIGA : Keputusan Rektor ini mulai berlaku sejak tanggal ditetapkan sampai dengan tanggal 30 November 2022.

Ditetapkan di Surabaya
pada tanggal 15 Juni 2022
REKTOR UNIVERSITAS NEGERI
SURABAYA,

ttd

NURHASAN
NIP 196304291990021001

Salinan sesuai dengan aslinya.
Kepala Biro Umum dan Keuangan,



SULAKSONO
NIP 196504091987011001

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
1	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	MODEL INTENS BERWIRUSAHA TERHADAP PENGAMBILAN KEPUTUSAN KARIR FRESH GRADUATE UNESA	Erta, S.E., M.M	0029078709	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
2	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	PENGARUH LATIHAN STABILISASI DINAMIC LEG PRESS DAN SINGLE STRAIGHT LEG DEADLIFT DENGAN MEDICINE BALL TERHADAP KEKUATAN OTOT TUNGKAI DAN KESEIMBANGAN	Panji Bana, M.Pd.	0017049006	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
3	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Jasmani, Kesehatan & Rekreasi S1	Korelasi Flat Foot Terhadap Keseimbangan Pada Atlet Cabang Olahraga Senam	Irma Febriyanti, S.Or., M.Kes.	0007028105	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
4	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Jasmani, Kesehatan & Rekreasi S1	Analisis Kemampuan TPACK (Technological Pedagogical Content Knowledge) pada Guru PJOK	Faridha Nurhayati, S.Pd., M.Kes.	0026068101	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
5	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Biologi S1	PENGEMBANGAN MODEL SPATIAL AUGMENTED PROBLEM BASED LEARNING UNTUK MELATIHKAN BERPIKIR SPASIAL MAHASISWA CALON GURU BIOLOGI	Ahmad Bashri, S.Pd., M.Si.	0707128202	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
6	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Sistem Informasi S1	Pendekatan Agglomerative Hierarchical Clustering Dalam Klasterisasi Pola Interaksi Mahasiswa Untuk Menentukan Efektivitas Pembelajaran Daring	I Kadek Dwi Nuryana, S.T., M.Kom.	0014048107	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
7	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Fisika S1	Pengembangan Lembar Kerja Mahasiswa Berbasis Argumentasi untuk Meningkatkan Berpikir Kritis pada Matakuliah Ilmu Pengetahuan Bumi dan Antariksa	Setyo Admoko, S.Pd., M.Pd.	0014127603	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
8	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Sistem Informasi S1	Pengembangan Smart Farming Berbasis IoT	Aries Dwi Indriyanti, S.Kom., M.Kom.	0012048006	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
9	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Jasmani, Kesehatan & Rekreasi S1	Pengembangan Kemandirian Belajar Calon Guru PJOK Melalui Model Peer-Tutoring	Fifukha Dwi Khory, S.Pd., M.Pd.	0025018201	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
10	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	Pengembangan Dilemmas Stories Sebagai Suplemen Bahan Ajar Pembelajaran Matematika Sekolah Dasar	Delia Indrawati, S.Pd., M.Pd.	0011128701	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
11	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Jasmani, Kesehatan & Rekreasi S1	PENGEMBANGAN MULTIMEDIA PEMBELAJARAN SEPAKBOLA BERBASIS ANDROID	Mochamad Ridwan, S.Pd., M.Pd.	0017028703	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
12	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	MANFAAT PENGULURAN OTOT, MASASE OLAHRAHA DAN MASASE-ULUR TERHADAP PENURUNAN ASAM LAKTAT DALAM DARAH BAGI PEMAIN SEPAK BOLA DIKLAT SEPAK BOLA PELAJAR NUSANTARA SURABAYA	Drs. Fatkur Rohman Kafrawi, M.Pd.	0019126708	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
13	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Tata Rias S1	TRAINING NEEDS ANALYSIS PRODUKSI ECO COSMETICS UNTUK PEREMPUAN DESA DI PACET MOJOKERTO	Sri Dwiyanti, S.Pd., M.PSDM.	0006027901	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
14	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	Dimensi Guru Tentang Nilai-Nilai Kewarganegaraan Dalam Upaya Mencegah Perundungan di Sekolah Dasar	Vicky Dwi Wicaksono, S.Pd., M.Pd.	0030038901	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
15	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Hukum S1	URGensi PEMBENTUKAN DAN PENGATURAN PENGADILAN KHUSUS AGRARIA DALAM PENYELESAIAN SENGKETA TANAH	Tamsil, S.H., M.H.	0003046209	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
16	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Teknologi Pendidikan S1	Pengembangan Model Inclusive Seamless Learning Berbasis Sistem “Among” di Perguruan Tinggi	Citra Fitri Kholidya, S.Pd., M.Pd.	0016058802	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
17	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Tata Busana S1	PENGEMBANGAN INSTRUMEN PERFORMANCE ASSESSMENT UNTUK MENGUKUR TECHNOPEURSHIP INTENTION PADA SISWA SMK TATA BUSANA	Imami Arum Tri Rahayu, S.Pd., M.Pd.	0701128101	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
18	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	DELIBERASI KEWARGANEGARAAN LINGKUNGAN DI SEKOLAH PENYELENGGARA PROGRAM ADIWiyATA	Ganes Gunansyah, S.Pd., M.Pd.	0029018005	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
19	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Bimbingan Dan Konseling S1	PENGEMBANGAN MODEL KONSELING KELUARGA UNTUK MENGURANGI ADIKSI GAME ONLINE	Wiryo Nuryono, S.Pd., M.Pd.	0012028601	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
20	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	ANALISIS KEBUTUHAN BAHAN AJAR IPS LITERASI FINANSIAL BERBASIS KEARIFAN LOKAL UNTUK SISWA SEKOLAH DASAR	Putri Rachmadyanti, S.Pd., M.Pd.	0002068902	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
21	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Akuntansi S1	Pengaruh Sistem Remunerasi terhadap Kinerja Dosen dengan Variabel Mediasi Kepuasan Kerja di Universitas Negeri Surabaya	Drs. Eko Wahjudi, M.Si.	0003066704	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
22	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Bimbingan Dan Konseling S1	Pengembangan Panduan Bimbingan Kelompok Bermuatan Nilai Budi Priyayi dan Budi Saudagar Bagi Siswa SMP di Surabaya	Denok Setiawati, M.Pd., Kons.	0002098101	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
23	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Hukum S1	Problematika Hukum Pengelolaan dan Pemanfaatan Sumber Daya Agaria (Hutan) Bagi Kepentingan Masyarakat Hukum adat	Indri Fogar Susilowati, S.H., M.H.	0014117201	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
24	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Luar Sekolah S1	ORGANISASI BELAJAR SEBAGAI INSTITUSI PEMBERDAYAAN MASYARAKAT	Heryanto Susilo, S.Pd., M.Pd.	0013058106	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
25	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa dan Sastra Indonesia S1	IMPLEMENTASI ASPEK PLURIKULTURAL DALAM PEMBELAJARAN BIPA	Prima Vidya Asteria, S.Pd., M.Pd.	0009108901	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
26	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Bisnis Digital S1	EKSPLORASI FAKTOR INTERNAL DAN EKSTERNAL ORGANISASI YANG MEMPENGARUHI INNOVATION TERHADAP ENTERPRENEURSHIP: STUDI PADA UMKM DI INDONESIA	Hujjatullah Fazlurrahman, S.E., MBA.	0723108603	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
27	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Bimbingan Dan Konseling S1	DINAMIKA MOTIVASI BERPRESTASI SISWA SMA DI KOTA SURABAYA	Bambang Dibyo Wiyono, S.Pd., M.Pd.	0030128704	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
28	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Administrasi Negara S1	Implementasi Program Merdeka Belajar Kampus Merdeka (Studi Pada KKNT-Proyek Desa di Fakultas Ilmu dan Hukum Unesa)	Indah Prabawati, S.Sos., M.Si.	0029077404	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
29	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Sains S1	Pengembangan Instrumen Untuk Menentukan Karakteristik Aktivitas STEM Calon Guru IPA	Tutut Nurita, S.Pd., M.Pd.	0028058202	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
30	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknologi Informasi S1	DISTRIBUTED NEWS CRAWLER	I Gusti Lanang Putra Eka Prisma, S.Kom., M.Kom.	0025038013	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
31	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Bimbingan Dan Konseling S1	IDENTIFIKASI PSYCHOLOGICAL WELLBEING SANTRI DI PONDOK PESANTREN MIFTAHUL ULUM JABUNG MALANG	Ari Khusumadewi, S.Pd., M.Pd.	0015068601	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
32	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Teknik Mesin D4	DESAIN DAN MODELLING FEMORAL STEM HIP ARTHOPLASTY SESUAI DENGAN ANATOMI RAS MONGOLOID (ORANG INDONESIA)	Diah Wulandari, S.T., M.T.	0005037804	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
33	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	Korelasi Panjang Tungkai dan Daya Ledak Tungkai Terhadap Kemampuan Sprint Pada Siswa SDN 236 Gresik	dr. Ananda Perwira Bakti, M.Kes.	0005068502	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
34	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Bisnis Digital S1	Manajemen Sumber Daya Manusia Hijau Sebagai Penentu Kinerja Lingkungan Dan Kinerja Hijau Karyawan: Peran Moderator Persepsi Iklim Kerja Hijau Karyawan	Hafid Kholidi Hadi, S.E., M.SM.	0013038701	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
35	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	PERAN PERSONALITY TRAITS PADA NIAT BELI ULANG MELALUI CO-CREATION DAN TRUST PADA PLATFORM E-COMMERCE	Widyastuti, S.Si., M.Si.	0020127509	III/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
36	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Administrasi Negara S1	PENINGKATAN KAPASITAS APARATUR DESA MELALUI RPL- DESA DALAM PERSPEKTIF INTERGOVERNMENTAL NETWORK	Muhammad Farid Ma'ruf, S.Sos., M.AP.	0030057606	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
37	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Jerman S1	ANALISIS HASIL TERJEMAHAN VERBA TRANSITIF DALAM AUSZUG ARTIKEL SKRIPSI 2021 MAHASISWA BAHASA JERMAN FBS UNESA	Drs. Suwarno Imam Samsul, M.Pd.	0026046801	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
38	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Teknik Mesin D4	Proses Pembuatan Serat Kulit Ari Kacang Kedelai	Arya Mahendra Sakti, S.T., M.T.	0009027903	IV/a	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
39	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Sastra Inggris S1	humor dalam ceramah agama: kajian sosiopragmatik	Fithriyah Inda Nur Abida, S.S., M.Pd.	0030048205	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
40	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Administrasi Perkantoran S1	Analisis Perilaku Administrasi Tenaga Kependidikan Dalam Pemanfaatan E-Office Sebagai Perwujudan Paperless Office	Brilliant Rosy, S.Pd., M.Pd.	0026058703	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
41	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Tata Niaga S1	ANALISIS KOMPARATIF STRATEGI MOTIVASI BELAJAR MAHASISWA DI JAWA TIMUR	Novi Marlina, S.Pd., M.Si.	0004098208	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
42	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	Peran Marketing Campaign dalam membentuk Engagement Behaviour	Monika Tiarawati, S.E., M.M.	0023127806	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
43	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	ANALISIS PEMBELAJARAN STEAM DI SEKOLAH DASAR DALAM PEMBELAJARAN TEMATIK MUATAN MAPEL IPA	Farida Istianah, S.Pd., M.Pd.	0018048306	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
44	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Bisnis Digital S1	Model Kebijakan Dividen Dengan Mediasi Profitabilitas Dan Moderasi Likuiditas Pada Perusahaan Agrikultur	Achmad Kautsar, S.E., M.M.	0015128901	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
45	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Seni Musik S1	PENCIPTAAN KARYA MUSIK AREK SEBAGAI WUJUD REVITALISASI NILAI EGALITER DALAM BUDAYA AREK	Joko Winarko, S.Sn., M.Sn.	0026037604	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
46	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	PERFORMANCE DI MEDIASI INOVASI PADA PENGARUH ADAPTIVE CAPABILITY DAN NETWORK CAPABILITY DENGAN RESOURCE FLEXIBILITY SEBAGAI PEMODERASI	Tias Andarini Indarwati, S.E., M.M.	0724097702	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
47	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Bimbingan Dan Konseling S1	Pengukuran nilai Validitas dan Reliabilitas Career Decision Making Self efficacy Scale-Sort Form	Evi Winingsih, S.Pd., M.Pd.	0018048902	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
48	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	PENGEMBANGAN LEMBAR KERJA MAHASISWA DENGAN PENDEKATAN MATEMATIKA REALSTIK BERBASIS PEMECAHAN MASALAH	Ika Rahmawati, S.Si., M.Pd.	0026038701	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
49	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Ekonomi S1	PENGEMBANGAN INSTRUMEN PENILAIAN HASIL BELAJAR RANAH AFEKTIF DALAM PEMBELAJARAN EKONOMI	Dhiah Fitrayati, S.Pd., M.E.	0007118201	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
50	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Pendidikan Sejarah S1	Pengembangan perangkat pembelajaran sejarah asia selatan berbasi blended history learning	Septina Alrianingrum, S.S., M.Pd.	0011097203	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
51	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Pendidikan Pancasila dan Kewarganegaraan S1	PENGEMBANGAN PERANGKAT PEMBELAJARAN PENDIDIKAN PANCASILA BERBASIS ONLINE COLLABORATIVE LEARNING	Listyaningsih, S.Pd., M.Pd.	0020027505	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
52	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknik Mesin S1	Analisis Kemampuan Menyelesaikan Masalah dan Kecerdasan Kinestetik Terhadap Kompetensi Kelistrikan Kendaraan Ringan Pada Siswa SMK Negeri di Surabaya	Heru Arizal, S.Pd., M.M., M.Pd.	0026078508	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
53	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Pendidikan Anak Usia Dini S1	SISTEM HOLISTIK INTEGRATIF PAUD BERBASIS SINGLE SIGN ON (SSO) SEBAGAI OPTIMALISASI LAYANAN PENDIDIKAN TERHADAP PENINGKATAN KEMAMPUAN LITERASI SAINS ANAK USIA DINI	Nur Ika Sari Rakhmawati, S.Pd., M.Pd.	0026088801	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
54	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknik Bangunan S1	PENGARUH INTELLIGENCE QUOTIENT TERHADAP KEMAMPUAN BERPIKIR, SIKAP KERJA, DAN KEPERIBADIAN TENAGA KEPENDIDIKAN DI UNIVERSITAS NEGERI SURABAYA	Suprpto, S.Pd., M.T.	0002046906	IV/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
55	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknik Bangunan S1	Pengaruh Kemampuan Berkomunikasi dan Kemampuan Bekerja Tim Terhadap Kesiapan Kerja Siswa SMK Bidang Keahlian Desain Pemodelan dan Informasi Bangunan di Kota Surabaya	Wahyu Dwi Mulyono, S.Pd., M.Pd.	0002068907	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
56	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Administrasi Negara D4	KEPEMIMPINAN DESA PEREMPUAN BERBASIS KEARIFAN LOKAL DI JAWA TIMUR	Yuni Lestari, S.AP., M.AP.	0013068501	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
57	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Administrasi Negara D4	Analisis Implementasi Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi Republik Indonesia Nomor 1 Tahun 2020 Tentang Pedoman Analisis Jabatan dan Analisis Beban Kerja di Kota Surabaya.	Gading Gamaputra, S.AP., MPA.	0016068702	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
58	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Administrasi Negara S1	Penerapan Sistem Informasi Pelayanan Perizinan Terpadu "Sippadu 2.0" melalui Prespektif Future of Government dalam menopang Public Service Innovation Successes oleh Pemerintah Kabupaten Sidoarjo	Trenda Aktiva Oktariyanda, S.AP., M.AP.	0025108901	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
59	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknik Bangunan S1	PENGARUH HARDSKILLS DAN KARAKTER TERHADAP KOMPETENSI LULUSAN SEKOLAH MENENGAH KEJUURUAN	Drs. H. Soeparno, M.T.	0001116506	IV/a	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
60	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Pendidikan Pancasila dan Kewarganegaraan S1	IMPLIKASI HUKUM PENGATURAN PENGAKUAN AGAMA BAGI PENGANUT AGAMA BAHAI DI KOTA SURABAYA	Iman Pasu Marganda Hadiarto Purba, S.H., M.H.	0019098501	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
61	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Jasmani, Kesehatan & Rekreasi S1	Pengaruh Metode Latihan High Intensity Interval Training (HIIT) Terhadap Peningkatan Kekuatan Otot, Kecepatan Tendangan, dan Daya Ledak Otot Atlet Karate	Afifan Yulfadinata, S.Pd., M.Pd.	0027068801	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
62	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Kimia S1	Pengembangan Instrumen Argumentasi Jawaban untuk Mengeksplorasi Konsepsi Awal Mahasiswa Pada Materi Kinetika Kimia	Bertha Yonata, S.Pd., M.Pd.	0022068201	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
63	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Ekonomi S1	ANALISIS TINGKAT LITERASI EKONOMI MAHASISWA PRODI PENDIDIKAN EKONOMI FEB UNESA	Riza Yonisa Kurniawan, S.Pd., M.Pd.	0031018601	III/d	S2		Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
64	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Pendidikan Geografi S2	Profil Metakognisi Mahasiswa dalam Memecahkan Masalah Geografi Sosial Ditinjau dari Perspektif Gender dan Gaya Kognitif Field Independent (FI) dan Field Dependent (FD)	Dr. Bambang Sigit Widodo, M.Pd.	0003037309	III/c	S3	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
65	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Kepelatihan Olahraga S1	IMPLEMENTASI MODEL CIPP SEBAGAI EVALUASI PROGRAM PEMBINAAN PRESTASI PADA CABANG OLAHRAGA PETANQUE DI PROVINSI JAWA TIMUR	Abdul Hafidz, S.Pd., M.Pd.	0011027701	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
66	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Seni Drama, Tari dan Musik S1	TRANSFORMASI KESENIAN SANDUR SEBAGAI MEDIA PENDIDIKAN KULTURAL	Arif Hidayat, S.Sn., M.Pd.	0025086908	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
67	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Administrasi Perkantoran S1	PENGARUH PENDIDIKAN KEWIRAUUSAHAAN DAN CAPACITY BUILDING TERHADAP PEMBENTUKAN KARAKTER WIRAUUSAHA MAHASISWA PAP UNESA	Durinda Puspasari, S.Pd., M.Pd.	0003018402	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
68	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Sains S1	Validitas Model Pembelajaran "ISOCC " Dengan Pendekatan Heutagogik Untuk Melatihkan Keterampilan Literasi informasi dan Argumentasi Kimia Mahasiswa Dalam Menumbuhkan Learning Skill di Abad 21	Siti Nurul Hidayati, S.Pd., M.Pd.	0014087504	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
69	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	TINGKAT AKTIVITAS FISIK DAN PARTISIPASI OLAHRAGA MASYARAKAT SERTA PENYEBARANNYA DI JAWA TIMUR	Hijrin Fithroni, S.Or., M.Pd.	0725088703	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
70	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Administrasi Perkantoran S1	PENGARUH FAMILY ENVIRONMENT DAN PENDIDIKAN KEWIRAUUSAHAAN TERHADAP SELF EFFICACY MAHASISWA PRODI PENDIDIKAN ADMINISTRASI PERKANTORAN UNESA	Durinta Puspasari, S.Pd., M.Pd.	0003018403	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
71	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Desain Grafis D4	PENGEMBANGAN INSTRUMEN PENILAIAN VISUAL LITERACY SKILLS UNTUK MAHASISWA DESAIN GRAFIS	Asidigisanti Surya Patria, S.T., M.Pd.	0019077703	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
72	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Sastra Inggris S1	Buku Anak Menggali Kesetaraan: Kritik Sastra Anak Gender	Diana Budi Darma, S.S., M.Pd.	0015056902	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
73	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Administrasi Perkantoran S1	Analisis Tingkat Administrative Literacy Mahasiswa S1 Pendidikan Administrasi Perkantoran	Triesninda Pahlevi, S.Pd., M.Pd.	0010118603	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
74	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Pendidikan Teknik Mesin S1	Analisis Kecerdasan Spasial, Kecerdasan Adversiti, terhadap Kreativitas Siswa SMK Jurusan Teknik Mesin Program Keahlian Teknik Pemesinan	Ali Hasbi Ramadani, S.Pd., M.Pd.	0720048904	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
75	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Teknologi Pendidikan S1	Pengembangan Virtual Learning Mata Kuliah Pengembangan Media Video di Prodi S1 Teknologi Pendidikan Universitas Negeri Surabaya	Khusnul Khotimah, S.Pd., M.Pd.	0004067904	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
76	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Manajemen Pendidikan S1	Implementasi Manajemen Perguruan Tinggi Inklusif di Universitas Negeri Surabaya	Supriyanto, S.Pd., M.Pd.	0014048601	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
77	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Sistem Informasi S1	PENGEMBANGAN MODEL PENGELOMPOKAN UANG KULIAH TUNGGAL BERBASIS ALGORITMA UNSUPERVISED LEARNING	Wiyli Yustanti, S.Si., M.Kom.	0003027708	IV/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
78	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Ekonomi Islam S1	Meta Analisis : Determinan Dana Pihak Ketiga Pada Bank Syariah Di Indonesia	Rachma Indrarini, S.EI., M.SEI.	0015019002	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
79	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Sastra Inggris S1	Menggunakan Kerangka ORIM untuk Mengungkap Keterlibatan Guru Bahasa Inggris sebagai Sponsor Literasi pada Sekolah Menengah Pertama di Surabaya	Drs. Much Koiri, M.Si.	0024036504	IV/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
80	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Luar Sekolah S1	LITERASI DIGITAL MASYARAKAT MARGINAL SUB-URBAN	Rivo Nugroho, S.Pd., M.Pd.	0005048107	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
81	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Hukum S1	KONSTRUKSI HUKUM INTERNASIONAL TERKAIT DISTINCTION PRINCIPLE DALAM KONFLIK BERSENJATA INTERNASIONAL	Elisabeth Septin Puspoayu, S.H., M.H.	0017098801	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
82	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Hukum S1	Keabsahan Surat Pernyataan Sebagai Pengalihan Hak Cipta Pada Publikasi Artikel Ilmiah Dalam Sebuah Jurnal	Budi Hermono, S.H., M.H.	0019038002	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
83	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Ekonomi Islam S1	IMPLEMENTASI SHARIA MARKETING MIX DAN ISLAMIC WORK ETHIC TERHADAP LOYALITAS MUZAKKI PADA LEMBAGA AMIL ZAKAT DI JAWA TIMUR	Khusnul Fikriyah, S.E., M.SEI.	0018088705	III/b	S2		Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
84	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Manajemen Informatika D4	Perbandingan Metode untuk mengetahui Dampak Kompetensi Soft Skills dalam Memprediksi Lulusan Mendapatkan Pekerjaan	Salamun Rohman Nudin, S.Kom., M.Kom.	0002118203	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
85	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	Perbandingan kadar hemoglobin dan myoglobin, pada mencit berdasarkan waktu tempuh lari saat melakukan tes ketahanan di atas treadmill	Dita Yuliastrid, S.Si., M.Kes.	0025077405	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
86	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Sains S1	Pendekatan Engineering Design Process dalam Pembelajaran Daring untuk Mahasiswa Calon Guru IPA	An Nuril Maulida Fauziah, S.Pd., M.Pd.	0001058503	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
87	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Sekolah Dasar S1	Pengembangan Buku Dongeng Karakter (DOKAR) Untuk Acuan dalam Program Penguatan Karakter di Sekolah Dasar	Hendrik Pandu Paksi, S.Pd., M.Pd.	0031058405	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
88	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Kimia S1	Pengembangan Lembar Kerja Mahasiswa Terintegrasi Kritis Kreatif untuk Meningkatkan Creative Problem Solving	Rusmini, S.Pd., M.Si.	0012067905	IV/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
89	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	Pengaruh Intervensi Glukosa terhadap Level Kortisol, Serotonin dan Performa Atlet Wanita Fase Luteal Siklus Menstruasi	Anna Noordia, S.TP., M.Kes.	0001117608	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
90	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	Mediasi Kepuasan dalam Hubungan Citra Destinasi dan Persepsi Resiko terhadap Loyalitas Wisatawan di Era New Normal COVID 19	Dra. Hj. Anik Lestari Andjarwati, M.M.	0005026306	IV/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
91	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Akuntansi S1	PENGARUH CORPORATE SOCIAL RESPONSIBILITY TERHADAP PENCIPTAAN NILAI BAGI STAKEHOLDER	Susi Handayani, S.E., Ak., M.Ak., CA.	0016097602	IV/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
92	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Administrasi Negara S1	Analisis Kesiapan Fakultas Ilmu Sosial Universitas Negeri Surabaya dan Hukum Menuju Perguruan Tinggi Negeri Berbadan Hukum (PTN-BH)	Dra. Meirinawati, M.AP.	0021056804	IV/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
93	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Kepelatihan Olahraga S1	Perbandingan Kombinasi Latihan Pliometrik Dan Latihan Resistan Untuk Meningkatkan Power, Kecepatan Dan Kelincahan	David Agus Prianto, S.Pd., M.Pd.	0815088101	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
94	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FT	Teknik Elektro S1	PENGEMBANGAN INSTRUMEN PENILAIAN UNTUK MEMAHAMI KONSEP SENSOR DAN AKTUATOR BAGI MAHASISWA TEKNIK ELEKTRO DI TINJAU DARI PENGARUH BERPIKIR KRITIS DAN KOLABORASI	Farid Baskoro, S.T., M.T.	0023058603	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
95	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	PENINGKATAN KEMAMPUAN BERPIKIR KRITIS DAN PEMAHAMAN NILAI-NILAI MORAL MELALUI PEMBACAAN TEKS NARASI UNTUK SISWA SEKOLAH MENENGAH PERTAMA	Nur Fauzia, S.S., M.Pd.	0021107804	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
96	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Tata Niaga S1	PENGEMBANGAN LEMBAR KERJA MAHASISWA (LKM) BERBASIS KETERAMPILAN PEMBELAJARAN ABAD 21 PADA MATA KULIAH KEWIRAUSAHAAN	Renny Dwijayanti, S.Pd., M.Pd.	0029068601	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
97	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Sains S1	STRATEGI REPRESENTASI VISUAL-SIMBOLIK DALAM MELATIHKAN KETERAMPILAN BERPIKIR KRITIS MAHASISWA PADA MATERI RANGKAIAN LISTRIK	Laily Rosdiana, S.Pd., M.Pd.	0029058202	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
98	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Kimia S1	Lembar Kerja Remediasi Conceptual Change Termodifikasi Tiga Pertanyaan Keilmuan dan Metakognitif untuk Mereduksi Miskonsepsi Calon Guru Kimia di Materi Keseimbangan Kimia	Dian Novita, S.T., M.Pd.	0019117409	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
99	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Ekonomi Islam S1	MINAT GENERASI Z UNTUK BERKONTRIBUSI DALAM E- CASH WAQF	Clarashinta Canggih, S.E., CFP.	0004098702	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
100	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Seni Drama, Tari dan Musik S1	PENGEMBANGAN INSTRUMEN PENGUKURAN ASPEK PSIKOMOTOR BIDANG SENI TEATER	Welly Suryandoko, S.Pd., M.Pd.	0025038801	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
101	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Manajemen Pendidikan S1	Pengembangan Digital Library Fakultas Ilmu Pendidikan berbasis SLIMS 9 Bulian	Mohammad Syahidul Haq, S.Pd., M.Pd.	0009048801	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
102	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Ekonomi S1	PEMAHAMAN DASAR EKONOMI MAHASISWA PENDIDIKAN EKONOMI DI SURABAYA	Albrian Fiky Prakoso, S.Pd., M.Pd.	0015118901	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
103	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Tata Niaga S1	Determinan Entrepreneurial Ecosystem Terhadap Entrepreneurial Self-Efficacy Mahasiswa Pendidikan Bisnis Di Pulau Jawa	Raya Sulistyowati, S.Pd., M.Pd.	0015058004	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
104	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	Penerapan Strategi Afektif dalam Praktek Mengajar Bahasa Inggris: Tantangan dan Solusi dimasa Post-Pandemic Covid-19	Ririn Pusparini, S.Pd., M.Pd.	0021057604	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
105	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	Pengembangan Model Pembelajaran Online-Project Based Learning Untuk Menuumbuhkan Keterampilan Berpikir Kritis Mahasiswa Calon Guru Bahasa Inggris	Arik Susanti, S.Pd., M.Pd.	0005027803	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
106	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	PENINGKATAN KETERAMPILAN PEMIKIRAN KRITIS MAHASISWA BAHASA INGGRIS MELALUI MATAKULIAH MEMBACA KRITIS	Henny Dwi Iswati, S.S., M.Pd.	0711107501	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
107	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Pendidikan Guru Pendidikan Anak Usia Dini S1	Pengembangan Game Augmented Reality Fonik Alfabeth Untuk Mengembangkan Literasi Siswa	Kartika Rinakit Adhe, S.Pd., M.Pd.	0015069001	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
108	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	APAKAH SELF CONTROL MERUPAKAN PENDORONG PERILAKU DAN SIKAP KEUANGAN KAUM MILENIAL?	Trias Madanika Kusumaningrum, S.E., S.Pd., M.M.	0024068801	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
109	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Kepelatihan Olahraga D4	ANALISIS TRAINING LOAD SUB ELIT ATLET DISABILITAS	Kunjung Ashadi, S.Pd., M.Fis., AIFO.	0008098104	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
110	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Manajemen S1	PERAN MEDIASI FINANCIAL CONSTRAIN PADA PENGARUH GREEN INNOVATION DAN OWNERSHIP STRUCTURE TERHADAP CORPORATE FINANCIAL PERFORMANCE	Yuyun Isbanah, S.E., M.SM.	0028128601	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
111	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Tata Boga D4	Studi Karakteristik Keju Merah Peram dengan Monascus purpureus sebagai Produk Pangan Baru yang Sehat Ditinjau dari Sifat Fisik, Kimia dan Sifat Organoleptik	Lilis Sulandari, S.Pt., M.P.	0020027407	IV/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
112	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	Pembelajaran Otonom Dalam Kelas Online Bahasa Inggris Berbasis Project Pada Konteks Pendidikan Tinggi	Nur Chakim, S.Pd., M.Pd.	0024077704	III/c	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
113	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Kepeleatihan Olahraga S1	PENGARUH SPORT MASSAGE TERHADAP PERUBAHAN HEART RATE, TEKANAN DARAH PADA ATLET SETELAH OLAHRAGA	Aghus Sifaq, S.Or., M.Pd.	0707088402	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
114	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FISH	Ilmu Hukum S1	KONSEP GANTI KERUGIAN PADA KESEPAKATAN PENYELESAIAN PERKARA PIDANA MELALUI ALTERNATIF PENYELESAIAN SENGKETA (ADR)	Emmilia Rusdiana, S.H., M.H.	0019067901	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
115	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	Menumbuhkan Keterampilan Berfikir Tinggi Melalui Praktik Literasi Digital Bagi Guru Bahasa Inggris Peserta Program Profesi Guru	Fauris Zuhri, S.Pd., M.Hum.	0025096703	IV/a	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
116	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Akuntansi S1	PENGEMBANGAN E-BOOK AKUNTANSI PEMERINTAHAN DESA UNTUK SMK	Rochmawati, S.Pd., M.Ak.	0003058013	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
117	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Pendidikan Kimia S1	PENGEMBANGAN INSTRUMEN PENILAIAN KETERAMPILAN BERPIKIR KREATIF BERORIENTASI COLLABORATIVE PROBLEM BASED LEARNING PADA MATERI LAJU REAKSI	Rusly Hidayah, S.Si., M.Pd.	0025098105	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
118	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa dan Sastra Jawa S1	PENDIDIKAN ETIKA DALAM SERAT MADUBASA II	Octo Dendy Andriyanto, S.Pd., M.Pd.	0026078901	III/b	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00

Skema Penelitian Dasar Studi Dalam Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
119	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIP	Manajemen Pendidikan S1	Pengaruh Inovasi Manajemen terhadap Perilaku Pembelajaran Digital pada Program Sekolah Penggerak se-Kota Surabaya	Syunu Trihantoyo, S.Pd., M.Pd.	0013088703	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
120	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FBS	Pendidikan Bahasa Jepang S1	KORELASI ANTARA EFIKASI DIRI GURU BAHASA JEPANG DENGAN PRAKTIK PENILAIAN HASIL BELAJAR	Amira Agustin Kocimaheni, S.Pd., M.Pd.	0007087803	III/d	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
121	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	VOKASI	Teknik Listrik D4	Optimalisasi Voltage Droop Control Dengan Beberapa Sumber Photovoltaic Berbasis Metode Aquila Optimizer Pada Sistem Microgrid DC	Widi Aribowo, S.T., M.T.	0023048005	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
122	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Ekonomi S1	Pengembangan Perkuliahan Ekonomi Mikro dengan Inseri Ekonomi Perilaku	Retno Mustika Dewi, S.Pd., M.Pd.	0024088501	III/c	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
123	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FMIPA	Kimia S1	Bifunctional Katalis Ni/ZSM-5 dan Aplikasinya Pada Reaksi Deoksigenasi Minyak Kemiri Sunan untuk Produksi Biojetfuel	Dina Kartika Maharani, S.Si., M.Sc.	0006068204	IV/a	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
124	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FIO	Pendidikan Kepelatihan Olahraga S1	PENGEMBANGAN PROGRAM MOBILE LEARNING BERBASIS VIRTUAL REALITY UNTUK IMPLEMENTASI KURIKULUM PENDIDIKAN JASMANI OLAHRAGA DAN KESEHATAN DI SEKOLAH DASAR	Mohammad Faruk, S.Pd., M.Kes.	0015018105	III/d	S2	L	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
125	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Pendidikan Tata Niaga S1	Analisis Pengaruh Kualitas Pelayanan Dan Kualitas Pembelajaran Terhadap Kepuasan Mahasiswa Unesa	Hapsari Shinta Citra Puspita Dewi, S.E., M.M.	0007048906	III/b	S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
126	2022 Penelitian Dasar Studi Dalam Negeri (LPPM)	FEB	Bisnis Digital S1	INTEGRASI GAMEFUL EXPERIENCE (GAMEX) DAN GAMEFUL HEURISTIC DESIGN DALAM MENINGKATKAN SELF BRAND CONNECTION DAN PURCHASE INTENTION DI E-COMMERCE	Ika Diyah Candra Arifah, S.E., M.Com.	0006118505		S2	P	Juni - November	Rp30,000,000.00	Rp21,000,000.00	Rp9,000,000.00
TOTAL											Rp3,780,000,000.00	Rp2,646,000,000.00	Rp1,134,000,000.00

Salinan sesuai dengan aslinya.
Kepala Biro Umum dan Keuangan,

SULAKSANA
NIP. 196504091987011001

Ditetapkan di : Surabaya
Pada tanggal 15 Juni 2022
REKTOR UNIVERSITAS NEGERI SURABAYA,

ttd

NURHASAN
NIP. 196304291990021001

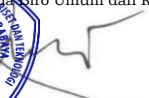
Skema Penelitian Dasar Studi Luar Negeri

No.	Skema Penelitian	Fakultas (Ketua)	Jurusan (Ketua)	Judul Penelitian	Nama Tim Peneliti	NIDN	Gol.	Pend.	L/P	Jangka Waktu	Dana Disetujui	Dana 70%	Dana 30%
1	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FT	Teknik Elektro S1	Disain Sistem Kontrol Berhierarchy untuk Meningkatkan Kebutuhan Pembagian Daya dan Perbaikan Tegangan pada Sistem DC Microgrid	Rifqi Firmansyah, S.T., M.T.	0704038901	III/b	S2	L	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
2	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FISH	Ilmu Administrasi Negara S1	Ekonomi Kampung Kota di Surabaya	Tauran, S.Sos., M.Soc.Sc.	0013047602	III/d	S2	L	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
3	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FIP	Psikologi S1	Pengajaran Sejarah berdasarkan Kurikulum 2006 (KTSP): Video analisis	Ira Darmawanti, S.Psi., M.Psi.	0017077304	III/d	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
4	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FEB	Manajemen S1	PENGARUH TEKNOLOGI, ORGANISASI, DAN LINGKUNGAN (TOE FRAMEWORK) DALAM PENGGUNAAN MEDIA SOSIAL TERHADAP KEMAMPUAN BRANDING USAHA MIKRO KECIL (UMK) MAKANAN	Rosa Prafitri Juniarti, S.E., M.S.M.	0027068803	III/b	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
5	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FBS	Pendidikan Bahasa Inggris S1	Investigasi Pengaruh Keyakinan, Nilai dan Sikap Guru Dalam Jabatan terhadap Praktik Professional untuk Meningkatkan Efektivitas Program PPG	Silfia Asningtias, S.Pd., M.TESOL.	0719068004	III/b	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
6	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FBS	Sastra Inggris S1	Subjektivitas, Tubuh Perempuan 'Pengungsi', dan Ruang Pendidikan dalam Karya Sastra Asia Amerika serta Kontribusinya dalam Mata Kuliah American Literature di Prodi Sastra Inggris Unesa	Hujala Rika Ayu, S.S., M.A.	0014128104	III/b	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
7	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FEB	Ekonomi S1	Faktor internal dan eksternal yang berpengaruh pada kinerja dan nilai perusahaan: studi pada sahaam Syariah di Indonesia	Choirul Nikmah, S.AB., M.AB.	0016108701	III/b	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
8	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FIO	Ilmu Keolahragaan S1	Evaluasi Kondisi Fisik Atlet Putri Tim Nasional Floorball Indonesia di Kejuaraan AOFC 2022	Luthfi Abdil Khuddus, S.Pd., M.Pd.	0010058701	III/b	S2	L	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
9	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FIP	Pendidikan Luar Biasa S1	FAKTOR-FAKTOR YANG MEMPREDIKSI SIKAP GURU TERHADAP PENDIDIKAN INKLUSIF: STUDI METODE CAMPURAN	Khofidotur Rofiah, S.Pd., M.Pd.	0010038901	III/b	S2	P	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
10	2022 Penelitian Dasar Studi Luar Negeri (LPPM)	FBS	Pendidikan Bahasa dan Sastra Indonesia S1	Kritik Sastra Cyber: Analisis Komputasional pada Karya Kritik Sastra Terpilih di Internet	Mohammad Rokib, S.S., M.A.	0001098403	III/b	S2	L	Juni - November	Rp50,000,000.00	Rp35,000,000.00	Rp15,000,000.00
TOTAL											Rp500,000,000.00	Rp350,000,000.00	Rp150,000,000.00

Ditetapkan di : Surabaya
Pada tanggal 15 Juni 2022
REKTOR UNIVERSITAS NEGERI SURABAYA,

ttd

NURHASAN
NIP. 196304291990021001

Sesuai dengan aslinya.
Biro Umum dan Keuangan,

SULAKSONO
NIP. 196504091987011001