

**MACHINE LEARNING UNTUK PREDIKSI GAYA HIDUP  
BERDASARKAN SOCIOECONOMIC STATUS (SES)  
MENGGUNAKAN ALGORITMA CATBOOST  
STUDI KASUS: MAHASISWA UIN JAKARTA**



**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH  
JAKARTA  
2023 M / 1444 H**

**MACHINE LEARNING UNTUK PREDIKSI GAYA HIDUP  
BERDASARKAN SOCIOECONOMIC STATUS (SES)  
MENGGUNAKAN ALGORITMA CATBOOST  
STUDI KASUS: MAHASISWA UIN JAKARTA**



**Disusun Oleh:**

**CHALISTA PUTRI ANANDA**

**11190930000038**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH  
JAKARTA  
2023 M / 1444 H**

**PERNYATAAN**

DENGAN INI SAYA MENYATAKAN BAHWA SKRIPSI INI BENAR-BENAR HASIL KARYA SENDIRI DAN BELUM PERNAH DIAJUKAN SEBAGAI SKRIPSI ATAU KARYA ILMIAH PADA PERGURUAN TINGGI ATAU LEMBAGA MANAPUN.



## ABSTRAK

**Chalista Putri Ananda** – 11190930000038, *Machine Learning* untuk Prediksi Gaya Hidup Berdasarkan *Socioeconomic Status* (SES) Menggunakan Algoritma CatBoost Studi Kasus: Mahasiswa UIN Jakarta. Di bawah bimbingan **Suci Ratnawati, MTI** dan **Elvi Fetrina, M.IT**.

Pengguna internet tertinggi yaitu kelompok umur yang masuk dalam Generasi Z (gen Z) banyak memanfaatkan penggunaannya untuk mengikuti *influencers* di media sosial yang dapat menimbulkan suatu tren di berbagai kalangan. Perkembangan tren tersebut menyebabkan adanya perilaku konsumtif bagi para pengikutnya. Mahasiswa sebagai bagian dari gen Z yang didominasi oleh mereka yang sumber pendapatannya dari orang tua memiliki tingkat konsumtif yang berbeda dalam mengikuti tren untuk memenuhi gaya hidup dilingkungan sekitarnya. Gaya hidup yang diikuti bergantung pada kategori *socioeconomic status* (SES) yang dinilai dari biaya pengeluaran rutinnya setiap bulan. Oleh karena itu, pada penelitian ini, menggunakan teknik *machine learning* yaitu *supervised learning* untuk memprediksi gaya hidup mahasiswa dengan mempelajari fitur-fitur yang digunakan dan mengklasifikasikan berdasarkan kelasnya yaitu SES. Prediksi dengan algoritma CatBoost digunakan untuk mengetahui tingkat akurasi prediksi dan faktor-faktor yang mempengaruhi berdasarkan data input yang diberikan. Hasil prediksi menggunakan CatBoost dengan hasil akurasi 85,94% menunjukkan bahwa golongan UKT, fakultas, dan biaya pengeluaran untuk hiburan perbulan merupakan fitur terpenting berdasarkan nilai *feature importances* yang menentukan kategori SES dan mempengaruhi peningkatan gaya hidup mahasiswa/i UIN Jakarta.

Kata Kunci: *Machine Learning*, Prediksi, Gaya Hidup, SES, CatBoost

V BAB + XIII Halaman + 155 Halaman + 43 Gambar + 10 Tabel + Daftar Pustaka

## KATA PENGANTAR

Assalamualaikum Wr.Wb.

Segala puji dan syukur saya panjatkan kepada Allah SWT karena atas limpahan kasih dan rahmat-Nya, saya dapat menyelesaikan skripsi yang berjudul “*Machine learning untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status (SES) Menggunakan Algoritma CatBoost Studi Kasus: Mahasiswa UIN Jakarta*” dengan baik dan tepat waktu. Pada saat penyusunan laporan praktik kerja lapangan ini, penulis banyak mendapatkan bantuan, saran, bimbingan, dukungan, serta do'a dari berbagai pihak yang tidak dapat diukur oleh materi. Oleh karena itu, dengan segala hormat dan kerendahan hati, penulis mengucapkan banyak terimakasih kepada:

1. Bapak Husni Teja Sukmana, Ph.D selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta.
2. Ibu Dr. Qurrotul Aini, M.T selaku Ketua Program Studi Sistem Informasi Fakultas Sains dan Teknologi.
3. Ibu Suci Ratnawati, M.TI selaku dosen pembimbing I dan Ibu Elvi Fetrina, M.IT selaku dosen pembimbing II yang telah membimbing, mengarahkan, dan memberikan ilmu dan pengetahuannya dalam membimbing penulis sehingga laporan ini dapat terselesaikan.
4. Bapak Bayu Waspodo, M.M selaku dosen pembimbing akademik saya yang telah banyak membantu selama masa perkuliahan.
5. Seluruh Dosen Program Studi Sistem Informasi yang telah memberikan ilmu dan dukungan selama perkuliahan.
6. Seluruh karyawan Fakultas Sains dan Teknologi yang sudah menolong penulis dalam mengurus administrasi yang berhubungan dengan skripsi.
7. Kedua orang tua, Bapak R.P. Ali Suharto dan Ibu Yunita Kurniasih, beserta kakak-kakak dan keponakan-keponakan saya yang telah memberikan doa, semangat, dukungan, serta motivasi selama masa perkuliahan hingga selesai.

8. Teman-teman Sistem Informasi angkatan 2019, khususnya Mahasiswa Sistem Informasi Kelas B yang saling membantu dan mendukung satu sama lain.
9. Sahabat penulis Naura, Zahwa, Fahira, Nabe, Pandu, Vaza, dan Dwiki yang selama kuliah memberikan dukungan, berbagi ilmu serta bertukar pikiran baik dari hal perkuliahan maupun hal pribadi penulis.
10. Seluruh responden penelitian yang telah membantu mengisi dan menyebarkan kuesioner penelitian skripsi ini.
11. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu hingga laporan ini terselesaikan.

Dalam penyusunan skripsi ini, penulis menyadari bahwa masih ada kekurangan yang disebabkan oleh keterbatasan ilmu pengetahuan, wawasan, dan pengalaman yang penulis miliki. Penulis memohon maaf atas segala kekurangan tersebut dan membuka saran maupun kritik bagi pembaca yang dapat disampaikan melalui email chalista.putria19@mhs.uinjkt.ac.id. Akhir kata, penulis skripsi ini dan apa yang telah penulis kerjakan dapat bermanfaat bagi pembaca dan penulis sendiri. Aamiin Ya Rabbal 'Alamin.

Wassalamualaikum Wr.Wb.

Jakarta, 17 Mei 2023



Chalista Putri Ananda

## DAFTAR ISI

<b>LEMBAR PENGESAHAN UJIAN .....</b>	<b>III</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>IV</b>
<b>ABSTRAK .....</b>	<b>V</b>
<b>KATA PENGANTAR.....</b>	<b>VI</b>
<b>DAFTAR ISI.....</b>	<b>VIII</b>
<b>DAFTAR GAMBAR.....</b>	<b>XI</b>
<b>DAFTAR TABEL .....</b>	<b>XIII</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah .....	8
1.3 Rumusan Masalah .....	8
1.4 Batasan Masalah.....	9
1.5 Tujuan Penelitian.....	9
1.6 Manfaat Penelitian.....	10
1.7 Metodologi Penelitian .....	10
1.8 Sistematika Penulisan.....	11
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>14</b>
2.1 Gen Z .....	14
2.2 Prediksi .....	14
2.3 Gaya Hidup.....	15
2.3.1 Faktor yang Mempengaruhi Gaya Hidup .....	17
2.4 <i>Socioeconomic Status (SES)</i> .....	20
2.4.1 Variabel <i>Socioeconomic Status (SES)</i> .....	21
2.4.2 Klasifikasi <i>Socioeconomic Status (SES)</i> .....	24
2.5 <i>Machine Learning</i> .....	24
2.5.1 Teknik <i>Machine Learning</i> .....	25
2.6 Klasifikasi.....	27
2.6.1 Algoritma Klasifikasi.....	28
2.6.2 Boosting.....	30
2.6.3 Matriks Korelasi .....	35
2.6.4 Metode Evaluasi .....	35
2.7 Python.....	39
2.7.1 Jupyter Notebook.....	40
2.8 Metodologi Penelitian Kuantitatif .....	40
2.9 Populasi dan Sampel.....	41
2.9.1 Definisi Populasi.....	41
2.9.2 Definisi Sampel .....	41
2.9.3 Jenis-jenis Metode Pengambilan Sampel .....	41
2.10 Metode Slovin .....	45
2.10.1 Rumus Metode Slovin .....	45

2.11 Penelitian Sejenis.....	47
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>61</b>
3.1 Pendekatan Penelitian.....	61
3.2 Populasi dan Sampel Penelitian.....	61
3.3 Metode Pengumpulan Data .....	62
3.3.1 Studi Pustaka .....	62
3.3.2 Kuesioner.....	62
3.4 Instrumen Penelitian.....	63
3.5 Pengolahan Data.....	70
3.5.1 Pra-pemrosesan Data .....	71
3.5.1.1 Impor <i>Libraries</i> .....	71
3.5.1.2 Impor Kumpulan Data .....	71
3.5.1.3 Menemukan Data yang hilang ( <i>Missing value</i> ) .....	71
3.5.1.4 Split Data .....	72
3.5.2 Pemrosesan Data.....	72
3.5.2.1 Pelatihan Model .....	72
3.5.2.2 Metrik Evaluasi .....	73
3.5.2.3 Visualisasi Data .....	73
3.6 Kerangka Penelitian.....	73
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>76</b>
4.1 Impor <i>Libraries</i> .....	76
4.2 Impor Kumpulan Data.....	76
4.3 Menemukan Data yang Hilang .....	76
4.4 Seleksi Fitur.....	76
4.5 Split Data .....	78
4.6 Pelatihan Model.....	79
4.7 Hasil Pemodelan CatBoost .....	79
4.8 <i>Feature Importances</i> .....	80
4.9 Performance Metrics .....	84
4.9.1 <i>Accuracy</i> .....	84
4.9.2 <i>Precision, Recall, &amp; F1 Score</i> .....	85
4.9.3 <i>Confusion Matrix</i> .....	87
4.9.4 Area Under ROC Curve (AUC).....	89
4.10 Visualisasi Data .....	89
4.10.1 Profil Mahasiswa UIN Jakarta.....	89
4.10.2 Variabel SES.....	89
4.10.3 Kepentingan Fitur .....	92
4.10.3.1 Golongan UKT .....	92
4.10.3.2 Pekerjaan Orang Tua.....	92
4.10.3.3 Fakultas .....	93
4.10.3.4 Biaya pengeluaran untuk hiburan setiap bulan .....	94
4.10.3.5 Biaya pengeluaran rutin setiap bulan.....	95
4.10.3.6 Biaya pengeluaran untuk kebutuhan pribadi setiap bulan ( <i>Skincare</i> , kosmetik, perawatan, dll) .....	96

4.10.3.7 Biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan .....	97
4.10.3.8 Biaya pengeluaran untuk perlengkapan mandi setiap bulan .....	98
4.10.3.9 Biaya pengeluaran untuk kegiatan organisasi setiap bulan.....	99
4.10.3.10 Biaya pengeluaran untuk makanan dan minuman setiap bulan. 99	
4.10.3.11 Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan.....	100
4.10.3.12 Biaya pengeluaran untuk pulsa kuota internet setiap bulan....	101
4.10.3.13 Biaya Pengeluaran untuk Konsumsi Tembakau, Pods, ataupun Vape Setiap Bulan .....	102
<b>4.10.4 Visualisasi Data Dukung .....</b>	<b>102</b>
4.10.4.1 Sumber Keuangan yang Didapatkan.....	102
4.10.4.2 Tempat Tinggal yang Ditempati Selama Kuliah <i>Offline</i> .....	103
4.10.4.3 Transportasi yang digunakan sehari-hari .....	104
4.10.4.4 Menerima beasiswa.....	104
4.10.4.5 Platform Layanan Pesan Antar yang Sering Digunakan.....	105
4.10.4.6 <i>E-wallet</i> yang Sering Digunakan .....	106
4.10.4.7 Platform <i>E-commerce</i> yang Sering Digunakan.....	106
4.10.4.8 Metode Pembelian Makanan.....	107
4.10.4.9 Operator Seluler yang Digunakan .....	108
4.10.4.10 Jenis Pembayaran Operator Seluler .....	108
4.10.4.11 Langganan Platform Berbayar yang Diikuti .....	109
4.10.4.12 Jenis Hiburan yang Dilakukan .....	111
4.11 Interpretasi Hasil .....	112
<b>BAB V PENUTUP .....</b>	<b>117</b>
5.1 Kesimpulan.....	117
5.2 Saran .....	118
<b>DAFTAR PUSTAKA .....</b>	<b>120</b>
<b>LAMPIRAN .....</b>	<b>136</b>

## DAFTAR GAMBAR

Gambar 1.1 Tingkat Penetrasi Pengguna Internet berdasarkan Demografi Responden.....	2
Gambar 1.2 Presentase Pengguna Internet yang Mengikuti (Influencers) di Media Sosial.....	3
Gambar 1.3 Frekuensi Transaksi dalam Sebulan Berdasarkan Kelompok Umur... 3	3
Gambar 2.1 Skor Jumlah Anggota Keluarga .....	22
Gambar 2.2 Machine learning techniques and data requirements .....	26
Gambar 2.3 Iterasi Pohon pertama dan kedua pada CatBoost.....	32
Gambar 2.4 Iterasi Pohon ke-N.....	33
Gambar 2.5 Kurva AUC – ROC .....	38
Gambar 2.6 Ukuran Sampel untuk Batas-batas Kesalahan dan Jumlah Populasi yang ditetapkan .....	46
Gambar 3.1 Kerangka Penelitian .....	74
Gambar 4.1 Hasil Correlation Matrix dengan Heatmap .....	77
Gambar 4.2 CatBoost MetricVisualizer.....	80
Gambar 4.3 Confusion Matrix .....	88
Gambar 4.4 AUC-ROC Curve .....	89
Gambar 4.5 Sumber Air Minum Berdasarkan SES .....	91
Gambar 4.6 Sumber Bahan Bakar untuk Memasak Berdasarkan SES .....	91
Gambar 4.7 Hubungan SES dengan Golongan UKT .....	92
Gambar 4.8 Hubungan SES dengan Pekerjaan Orang Tua.....	93
Gambar 4.9 Hubungan SES dengan Fakultas .....	94
Gambar 4.10 Hubungan SES dengan Biaya Pengeluaran untuk Hiburan .....	95
Gambar 4.11 Hubungan SES dengan Biaya Pengeluaran Rutin.....	96
Gambar 4.12 Hubungan SES dengan Biaya Pengeluaran untuk Kebutuhan Pribadi .....	97
Gambar 4.13 Hubungan SES dengan Biaya Pengeluaran untuk Langganan Platform Berbayar .....	98
Gambar 4.14 Hubungan SES dengan Biaya Pengeluaran untuk Perlengkapan Mandi .....	98
Gambar 4.15 Hubungan SES dengan Biaya Pengeluaran untuk Organisasi .....	99
Gambar 4.16 Hubungan SES dengan Biaya Pengeluaran untuk Makan dan Minum .....	100
Gambar 4.17 Hubungan SES dengan Biaya Pengeluaran untuk ATK .....	101
Gambar 4.18 Hubungan SES dengan Biaya Pengeluaran untuk Pulsa Internet .	101
Gambar 4.19 Hubungan SES dengan Biaya Pengeluaran untuk Konsumsi Tembakau.....	102
Gambar 4.20 Hubungan SES dengan Sumber Keuangan .....	103

Gambar 4.21 Hubungan SES dengan Tempat Tinggal .....	103
Gambar 4.22 Hubungan SES dengan Transportasi yang Digunakan .....	104
Gambar 4.23 Hubungan SES dengan Penerima Beasiswa.....	105
Gambar 4.24 Hubungan SES dengan Platform Layanan Pesan Antar .....	105
Gambar 4.25 Hubungan SES dengan E-wallet yang Digunakan.....	106
Gambar 4.26 Hubungan SES dengan E-commerce yang Digunakan .....	107
Gambar 4.27 Hubungan SES dengan Metode Pembelian Makanan.....	107
Gambar 4.28 Hubungan SES dengan Operator Seluler yang Digunakan.....	108
Gambar 4.29 Hubungan SES dengan Jenis Pembayaran Operator Seluler .....	109
Gambar 4.30 Hubungan SES dengan Langganan Platform Berbayar .....	109
Gambar 4.31 Platform Berbayar yang Diikuti .....	111
Gambar 4.32 Hubungan SES dengan Jenis Hiburan.....	111
Gambar 4.33 Kegiatan Hiburan yang Dilakukan .....	112



## **DAFTAR TABEL**

Tabel 2.1 Skor Air Minum Rumah Tangga .....	22
Tabel 2.2 Skor Sumber Bahan Bakar untuk Memasak .....	23
Tabel 2.3 Kategori SES.....	24
Tabel 2.4 Confusion Matrix .....	36
Tabel 2.5 Penelitian Sejenis .....	47
Tabel 3.1 Pertanyaan Kuesioner .....	64
Tabel 5.1 Hasil Analisis Seleksi Fitur .....	78
Tabel 5.2 Feature Importances .....	81
Tabel 5.3 Precision, Recall, & F1 Score .....	85
Tabel 5.4 Biaya Pengeluaran Rutin Mahasiswa/i UIN Jakarta per Bulan .....	90





## **BAB I**

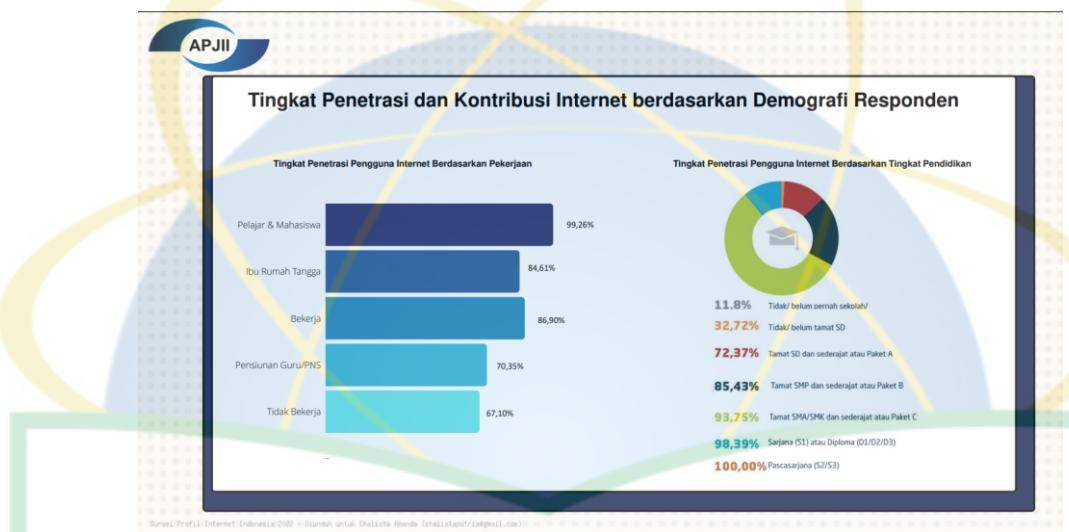
### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Pandemi Covid-19 yang dialami sejak Maret 2020 telah memberikan banyak dampak pada berbagai sektor serta aktivitas masyarakat Indonesia. Kegiatan yang biasanya dilakukan secara langsung banyak yang beralih menjadi daring atau *online*. Sebelum adanya Covid-19 pembelian barang maupun jasa secara daring sudah ada namun, belum banyak digunakan seperti sekarang. Teknologi dan digitalisasi yang berkembang pesat membawa pengaruh besar bagi kehidupan manusia saat ini dan dimasa yang akan datang sehingga, masyarakat lebih mudah memenuhi kebutuhan ataupun keinginannya baik itu barang, makanan, ataupun jasa. Selain itu, kemudahan juga dapat dirasakan oleh masyarakat dengan memilih, membeli, dan menggunakan kebutuhan serta keinginannya karena dapat dilakukan dimanapun dan kapanpun. Perkembangan tren berbagai macam produk yang semakin banyak pilihan dan fasilitas yang semakin memadai dari platform penjualan, pembelian, hingga berbagai metode pembayaran dapat memengaruhi kehidupan seseorang salah satunya pada generasi Z (gen Z).

Stillman pada bukunya yang berjudul “*Gen Z @work: How The Next Generation is Transforming The Workplace*” mengungkapkan bahwa gen Z merupakan generasi setelah generasi milenial yang lahir dari tahun 1995 sampai tahun 2012. Sedangkan menurut data Badan Pusat Statistik (BPS) hingga 31 Desember 2021, generasi Z yang berasal dari generasi kelahiran 1997-2012 mendominasi populasi di Indonesia sebesar 27,94% atau sebanyak 68.662.815 jiwa. Gen Z yang saat ini memiliki rentang usia 10-25 tahun yang didominasi oleh pelajar dan mahasiswa yang memiliki karakteristik sebagai generasi pertama pada generasi internet (Grail, 2011). Selain itu, menurut Tari Annamária pada bukunya yang berjudul Z “*generáció*” gen Z memiliki atribut “*net generation*” yang memiliki makna bahwa gen Z lahir pada saat berkembangnya era digital dan dicirikan sebagai “*Facebook-generation*”, “*digital native*” atau “*iGeneration*” (Annamária dikutip dalam Andrea et al., 2016). Hal tersebut didukung dengan penetrasi pengguna internet yang terdapat dalam profil internet Indonesia dan dirilis oleh Asosiasi

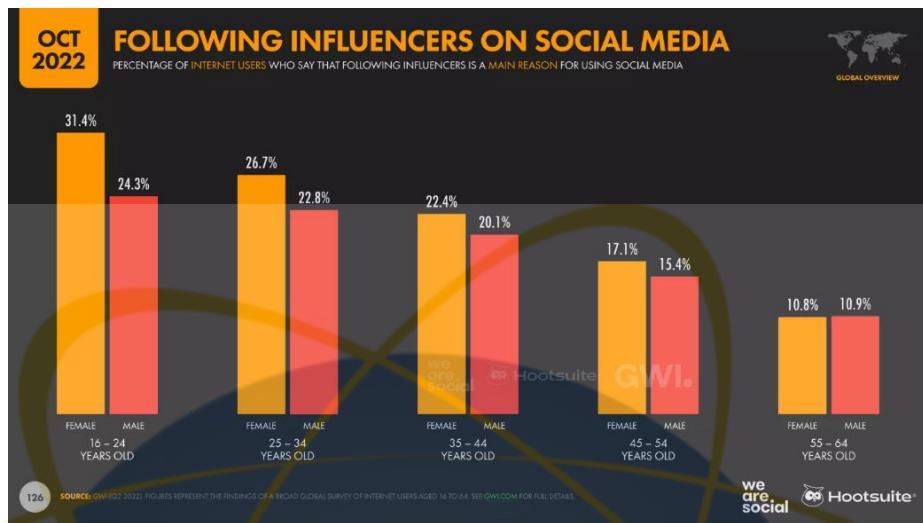
Penyelenggara Jasa Internet Indonesia (APJII) pada Juni 2022. Melalui survei yang dilakukan APJII tentang tingkat penetrasi pengguna internet tertinggi berdasarkan pekerjaan pada 7.568 responden sebagai sampel, hasilnya menunjukkan bahwa sebesar 14,65% merupakan persentase penggunaan internet yang digunakan oleh pelajar dan mahasiswa Indonesia (Asosiasi Penyelenggara Jasa Internet Indonesia, 2022).



Gambar 1.1 Tingkat Penetrasi Pengguna Internet berdasarkan Demografi Responden

Sumber: APJII, Profil Internet Indonesia 2022

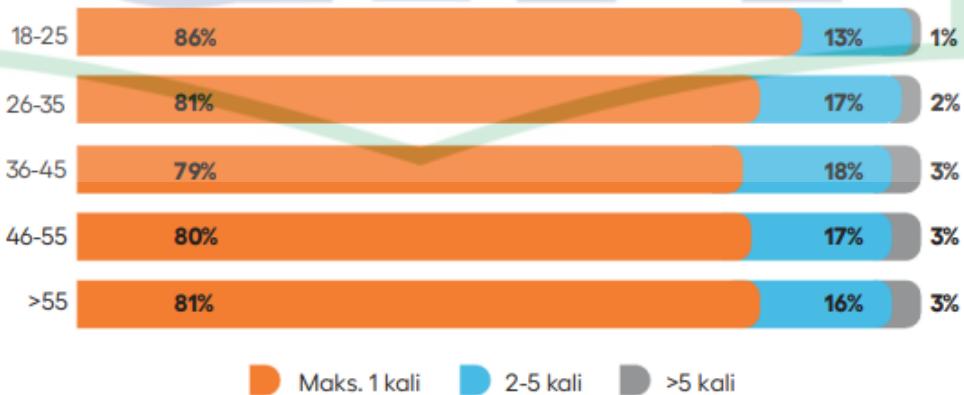
Gen Z sangat bersifat sosial karena sebagian besar generasi ini berada dalam hidup yang sedang tumbuh dan bersosialisasi secara langsung maupun memanfaatkan berbagai media digital, khususnya pada rentang usia sekolah menengah hingga perguruan tinggi (Skrovan, 2022). Gen Z kerap mengikuti tren, hal tersebut dipengaruhi oleh orang-orang yang mereka ikuti di berbagai sosial media (*influencers*) maupun lingkungan sekitar berdasarkan apa yang mereka pakai ataupun yang mereka miliki dengan memamerkan produk tersebut, hingga menyebabkan konsumen Gen Z juga menginginkan barang-barang tersebut (Fromm, 2022). Hal tersebut didukung dengan presentase pengguna internet yang mengikuti (*Influencers*) di sosial media menurut Hootsuite Oktober 2022 terbanyak yaitu kategori umur 16-24 tahun.



Gambar 1.2 Presentase Pengguna Internet yang Mengikuti (Influencers) di Media Sosial

Sumber: <https://www.hootsuite.com/resources/digital-trends-q4-update>

Mahasiswa sebagai salah satu bagian dari gen Z yang paling banyak mengakses berbagai hal melalui media sosial dalam kehidupan sehari-harinya memungkinkan bagi mereka untuk melakukan pembelian barang lebih banyak. Berdasarkan data dari *Kredivo e-commerce Behavior Report 2022*, jika dibandingkan tahun 2020, jumlah transaksi konsumen kelompok umur 18-25 tahun mengalami penurunan secara proporsi namun dengan pendapatan lebih rendah kelompok umur ini memiliki pengeluaran yang lebih besar untuk berbelanja di *e-commerce*. Gambar 1.3 menunjukkan hanya 14% kelompok konsumen paling muda (18-25 tahun) yang berbelanja secara daring lebih dari satu kali setiap bulannya.



Gambar 1.3 Frekuensi Transaksi dalam Sebulan Berdasarkan Kelompok Umur

Sumber: Kredivo, Perilaku Konsumen E-commerce Indonesia 2022

Salah satu pengaruh meningkatnya jumlah transaksi pembelian barang dapat disebabkan oleh keinginan memiliki suatu barang diluar kebutuhan karena adanya rasa takut jika tertinggal tren dan meminimalisir adanya kesenjangan sosial di lingkungan sekitar yang biasa disebut dengan *Fear of Missing Out* (FoMO). FoMO merupakan salah satu karakteristik Gen Z dimana adanya perasaan takut atau gelisah akan sesuatu hal yang sifatnya baru dan sedang terjadi seperti tren, berita, informasi, dan lainnya (Stillman, 2017).

Kondisi FoMO dapat mengakibatkan perilaku konsumtif pada seseorang dengan melakukan pembelian barang maupun jasa yang tidak sesuai dengan kebutuhan dan tanpa pikir panjang (Rasyid, 2022). Pada remaja, perilaku konsumtif dapat dipengaruhi oleh SES (pendapatan) orang tua (Sipunga & Muhammad, 2014, p. 67, dikutip dalam Zahrawati & Faraz, 2018). Menurut Basrowi (2005), ukuran yang digunakan dalam menentukan kedudukan SES seseorang di masyarakat salah satunya, yaitu ukuran kekayaan yang dapat dilihat dari bentuk rumah yang bersangkutan, mobil pribadinya, cara-caranya berpakaian serta bahan yang dipakainya, dan kebiasaannya berbelanja barang dan jasa.

SES itu sendiri merupakan cara pengelompokan kemampuan ekonomi dan status sosial dari seseorang dan keluarga berdasarkan biaya yang dikeluarkan. Biaya kebutuhan maupun keinginan yang dikeluarkan dapat diklasifikasi ke dalam SES yang biasanya dibagi menjadi 3 kategori yaitu *high SES*, *middle SES* dan *low SES*. Setiap kategori memiliki skor berbeda yang akan ditentukan berdasarkan pengeluaran dengan 4 variabel yang telah ditentukan. Variabel SES yang dapat digunakan menurut Indonesia Data yaitu, pengeluaran rutin bulanan rumah tangga, jumlah anggota rumah tangga, sumber air minum rumah tangga, sumber bahan bakar untuk memasak.

Beberapa penelitian sebelumnya menyatakan pengaruh orang tua berpengaruh signifikan terhadap perilaku konsumsi (Bahjatussaniah et al., 2015). Selain itu, Ardianti juga mengungkapkan bahwa interpretasi khalayak terhadap gaya hidup travelling yang ditampilkan melalui foto dalam akun Instagram dapat berbeda karena dipengaruhi oleh perbedaan usia, jenis kelamin, latar belakang pendidikan, SES dan aktivitas *travelling* informan. Biaya pengeluaran konsumsi

rumah tangga untuk memenuhi kebutuhan hidup dasar telah bergeser untuk kebutuhan sekunder maupun primer dapat yang dilihat dari biaya pengeluaran untuk memenuhi gaya hidup yang tumbuh di atas konsumsi makanan dan minuman seperti berkumpul di kafe maupun *coffee shop* serta *traveling* yang saat ini menjadi gaya hidup bagi masyarakat (Putri, 2019). Namun dengan adanya perbedaan gaya hidup antar generasi dapat terjadi perubahan sosial di masyarakat dan lingkungan ekonomi yang nantinya dapat menjadi peluang bagi pemasar untuk menyesuaikan produk miliknya sesuai dengan gaya hidup pasar yang akan dituju (Hana, 2019 dikutip dalam Aini & Andjarwati, 2020)

Pada saat melakukan pemasaran kepada konsumen, gaya hidup konsumen dianggap sebagai variabel psikologis yang dapat mempengaruhi proses keputusan pembeli bagi konsumen (Niosi, 2021). Menurut Khalipha Ntloko pada artikelnya yang berjudul "*Lifestyle Marketing: How to Leverage Your Consumers' Lifestyle*" dengan memahami berbagai gaya hidup konsumen tersebut, pemasar dapat mengelompokkan konsumen menurut gaya hidup mereka atau yang biasa dikenal sebagai segmentasi gaya hidup. Seorang peneliti gaya hidup terkemuka, Joseph T. Plummer juga menyatakan bahwa "...*lifestyle patterns, combines the virtues of demographics with the richness and dimensionality of psychological characteristics.... Lifestyle is used to segment the marketplace because it provides the broad, everyday view of consumers lifestyle segmentation and can generate identifiable whole persons rather than isolated fragments*" (Plummer, 1974). Sehingga, gaya hidup dapat digunakan untuk mensegmentasi pasar dan dapat menghasilkan gambaran luas yang diidentifikasi melalui gaya hidup konsumen.

Prinsip dasar segmentasi gaya hidup adalah semakin memahami gaya hidup konsumen, semakin efektif strategi pemasaran dan komunikasi kepada konsumen (Ntoloko, 2020). Pemasar juga dapat menggunakan segmentasi gaya hidup untuk menyelaraskan merek, produk, dan layanan yang dimilikinya untuk bisa berkembang sesuai dengan target pasarnya.

Segmentasi gaya hidup juga dapat membantu menciptakan pesan dari merek tersebut mengenai produk atau layanan yang dapat mempersonalisasikannya sesuai dengan setiap segmen gaya hidup (Ntoloko, 2020). Personalisasi juga merupakan hal

penting bagi pemasar agar dapat memberikan pengalaman yang dipersonalisasi kepada konsumen, bersama dengan produk dan layanan yang lebih memenuhi kebutuhan konsumen.

Selain itu, gaya hidup remaja yang memiliki tingkat penggunaan media sosial yang tinggi juga dapat menjadi peluang dalam strategi pemasaran (KomalaSari et al., 2022). Pemasar dapat melakukan setidaknya dua hal untuk melibatkan Gen Z dalam mempromosikan mereka dengan cara menjangkau mereka melalui media sosial dan menyelaraskan dengan cara berpikir mereka yang progresif (Fromm, 2022). Keselarasan dalam berbagai hal yang dilakukan oleh konsumennya dapat menunjukkan bahwa pemasar berada di tempat yang lebih dekat untuk menjangkau Gen Z dan mendapatkan loyalitas dari para konsumen (Fromm, 2022).

Dari sisi konsumen, mahasiswa saat ini yang merupakan bagian dari gen Z dapat melakukan literasi keuangan dengan memahami faktor-faktor yang mempengaruhi meningkatnya gaya hidup berdasarkan SES. Pada umumnya mahasiswa mulai menjalani masa peralihan dari pengelolaan keuangan yang diatur oleh orang tua menjadi pengelolaan keuangan secara mandiri (Yushita, 2017). Meningkatkan literasi keuangan pada mahasiswa diharapkan mereka dapat mengatur sendiri pengeluarannya dengan terencana, memiliki dan mengelola tabungannya dengan baik dan dapat mengatur konsumsi dan meminimalkan pemborosan (Rohmanto & Susanti, 2021).

Didukung oleh beberapa penelitian sebelumnya, literasi keuangan merupakan pengetahuan dasar tentang pengelolaan keuangan yang baik dan bijak yang akan mempengaruhi perilaku keuangan mahasiswa dimasa yang akan datang dengan pertimbangan yang baik terhadap keputusan pengelolaan keuangan yang bijak dan efisien (Abdurrahman & Oktapiiani, 2020). Selain itu juga, literasi keuangan dan gaya hidup hedonisme berpengaruh terhadap manajemen keuangan pribadi mahasiswa akuntansi STIE PGRI Dewantara Jombang (Pramesti, 2023). Literasi keuangan yang berpengaruh terhadap perilaku pengelolaan keuangan Mahasiswa pada Fakultas Ekonomi dan Bisnis Universitas Singaperbangsa Karawang (Sugiharti & Maula, 2019).

Kemajuan teknologi komputasi saat ini mendukung berbagai metode analisis, dengan memprediksi faktor-faktor terkait seperti pada penelitian sebelumnya yaitu, memprediksi dengan menggunakan *machine learning*. *Machine learning* mulai diterapkan dalam kehidupan sehari-hari dalam beberapa tahun terakhir mulai dari rekomendasi film, makanan, produk yang akan dibeli, hingga personalisasi teman di album foto dan berbagai situs web ((Müller & Guido, 2015). Pada hal tersebut *machine learning* dengan teknik *supervised Learning* diterapkan dengan kumpulan data yang diberikan dapat dipelajari dan dapat memprediksi dengan benar dari data yang disebut “label” dengan cara mengklasifikasikan label tersebut.

Pada penelitian sebelumnya kemungkinan gagal bayar pinjaman untuk memberi rekomendasi ke pihak bank mana pinjaman pelanggan yang disetujui dan tidak dengan mengklasifikasikan pelanggang ke dalam 4 kategori yaitu: *Excellent Credit*, *Good Credit*, *Low Credit*, dan *Bad Credit* serta verifikasi dokumen yang dibutuhkan saat pengajuan pinjaman (Anuradha & David, 2021). Prediksi kebangkrutan berdasarkan faktor keuangan (Jabeur et al., 2021), dan prediksi sewa rumah dengan menganalisis faktor karakteristik lainnya, seperti usia rumah, metode persalinan, dll (Zhang et al., 2019) menggunakan *machine learning*. Metode *machine learning* digunakan karena dapat membantu mencapai akurasi prediksi yang lebih tinggi (Park & Lee, 2022).

Berbagai studi yang telah dilakukan tersebut membuat model prediksi yang terbaik dalam mengklasifikasikan pada penelitian tersebut juga menerapkan algoritma *boosting*. Algoritma *boosting* merupakan sejumlah pengklasifikasi lemah (pengklasifikasi yang memprediksi sedikit data lebih baik daripada acak) yang digabungkan (atau ditingkatkan) untuk menghasilkan pengklasifikasi ansambel dengan tingkat kesalahan kesalahan klasifikasi umum yang unggul (Kuhn & Johnson, 2013).

*Categorical Boosting* (CatBoost) yang merupakan bagian dari *gradient boosting* dan implementasi *Gradient Boosted Decision Tree* (GBDT) di mana *decision tree* merupakan bagian dari *supervised learning* yang digunakan untuk memprediksi faktor-faktor yang berpengaruh terhadap gaya hidup mahasiswa berdasarkan SES mahasiswa UIN Jakarta mennggunakan data yang dikumpulkan

melalui kuesioner. CatBoost digunakan pada penelitian ini karena memiliki hasil kinerjanya yang baik dengan nilai akurasi paling tinggi di antara semua algoritma lainnya. Algoritma untuk prediksi gagal bayar pinjaman dengan membandingkan dua algoritma berbeda yaitu *random forest* dan *gradient boosting* dengan hasil CatBoost mencapai akurasi tertinggi diantara algoritma lainnya (Barua et al., 2021). Waktu komputasi dan penggunaan memori untuk pemrosesan data CatBoost jauh lebih sedikit dibandingkan dengan *random forest* (RF) dan *support vector machine* (SVM) dimana CatBoost memiliki peningkatan signifikan dalam akurasi, stabilitas, dan biaya komputasi jika dibandingkan dengan RF (Huang et al., 2019).

Berdasarkan dari latar belakang permasalahan yang telah dijelaskan di atas, peneliti tertarik untuk melakukan penelitian yang berjudul "**Machine Learning Untuk Prediksi Gaya Hidup Mahasiswa Berdasarkan Socioeconomic Status (SES) Menggunakan Algoritma CatBoost Studi Kasus: Mahasiswa UIN Jakarta**". Hasil yang diharapkan dari penelitian ini pada akhirnya dapat mengetahui faktor-faktor yang berpengaruh terhadap meningkatnya gaya hidup mahasiswa UIN Jakarta yang juga dapat digunakan sebagai rekomendasi riset pasar.

## 1.2 Identifikasi Masalah

Identifikasi masalah pada penelitian ini didapatkan berdasarkan latar belakang yang telah diuraikan sebelumnya. Identifikasi masalah pada penelitian ini yaitu:

1. Meningkatnya daya konsumtif mahasiswa sebagai bagian dari gen Z dengan pendapatan yang lebih kecil namun memiliki jumlah pengeluaran untuk belanja yang lebih besar yang dipengaruhi oleh FoMO.
2. Biaya *living cost* untuk memenuhi kebutuhan hidup dasar yang bergeser untuk kebutuhan sekunder maupun primer karena adanya biaya pengeluaran untuk memenuhi gaya hidup.

## 1.3 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah disampaikan, penelitian ini memiliki rumusan masalah antara lain:

1. Bagaimana hasil prediksi SES pada mahasiswa UIN Jakarta menggunakan algoritma CatBoost dalam *performance metrics*?
2. Apa saja faktor-faktor yang mempengaruhi meningkatnya gaya hidup mahasiswa berdasarkan SES?

#### 1.4 Batasan Masalah

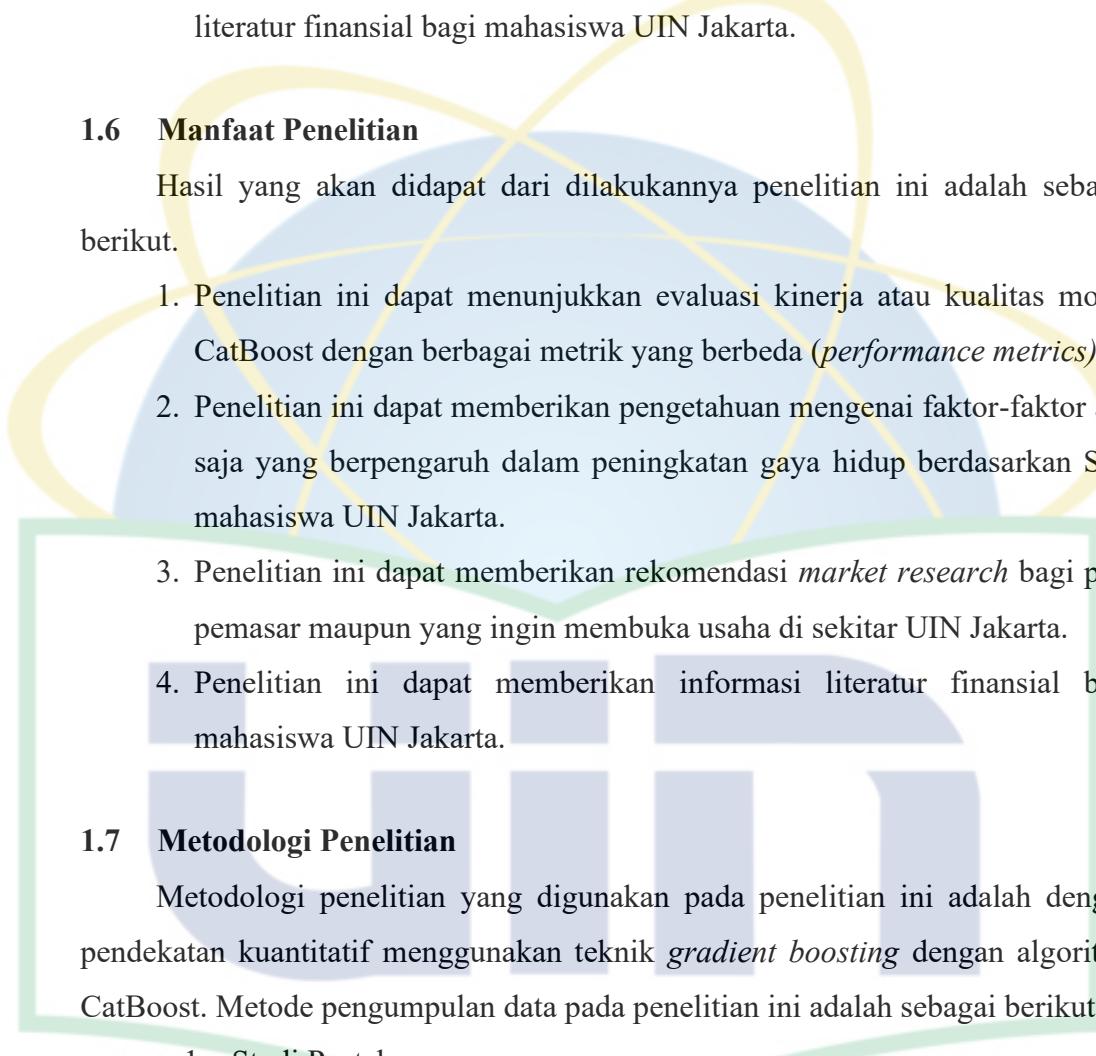
Penelitian ini memerlukan batasan masalah agar fokus pada permasalahan yang diteliti. Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Penelitian ini menggunakan 4 variabel SES yaitu pengeluaran rutin bulanan, jumlah anggota, sumber air minum, sumber bahan bakar untuk memasak tiap individu.
2. Jumlah anggota keluarga diasumsikan berjumlah 1 orang atau setiap individu yaitu mahasiswa itu sendiri.
3. Prediksi yang dilakukan dengan pemodelan menggunakan algoritma CatBoost
4. Hasil pemodelan CatBoost akan ditampilkan dalam *performance metrics* untuk memperlihatkan fitur dan akurasi keakuratan data yaitu *accuracy*, *confusion matrix*, *precision & recall*, *F1 score*, dan *Area under curve (AUC)*.
5. Penelitian ini menggunakan metode pendekatan kuantitatif dengan data yang diperoleh berdasarkan studi pustaka dan survei melalui penyebaran kuesioner secara daring (*online*) dengan sampel penelitian dibatasi hanya pada Mahasiswa aktif UIN Jakarta tahun angkatan 2019-2022 yang merupakan bagian dari gen Z.
6. Penelitian ini menggunakan bantuan *tools* Google Form, Python, Jupyter Notebook, dan Tableau.

#### 1.5 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut.

1. Penulis mengetahui hasil prediksi dengan pemodelan CatBoost dalam *performance metrics* pada SES mahasiswa/i UIN Jakarta.

- 
- Penulis dapat mengetahui prediksi faktor-faktor yang berpengaruh terhadap meningkatnya gaya hidup berdasarkan SES serta memberikan rekomendasi *market research* bagi para pemasar maupun yang ingin membuka usaha di sekitar UIN Jakarta dan dapat digunakan sebagai bahan literatur finansial bagi mahasiswa UIN Jakarta.

## 1.6 Manfaat Penelitian

Hasil yang akan didapat dari dilakukannya penelitian ini adalah sebagai berikut.

- Penelitian ini dapat menunjukkan evaluasi kinerja atau kualitas model CatBoost dengan berbagai metrik yang berbeda (*performance metrics*).
- Penelitian ini dapat memberikan pengetahuan mengenai faktor-faktor apa saja yang berpengaruh dalam peningkatan gaya hidup berdasarkan SES mahasiswa UIN Jakarta.
- Penelitian ini dapat memberikan rekomendasi *market research* bagi para pemasar maupun yang ingin membuka usaha di sekitar UIN Jakarta.
- Penelitian ini dapat memberikan informasi literatur finansial bagi mahasiswa UIN Jakarta.

## 1.7 Metodologi Penelitian

Metodologi penelitian yang digunakan pada penelitian ini adalah dengan pendekatan kuantitatif menggunakan teknik *gradient boosting* dengan algoritma CatBoost. Metode pengumpulan data pada penelitian ini adalah sebagai berikut.

### 1. Studi Pustaka

Penulis mengidentifikasi dan menganalisis dokumen-dokumen yang memuat informasi yang berkaitan dengan penelitian yang dilakukan. Dokumen-dokumen ini bersumber dari buku maupun *website* yang berkaitan dengan topik dalam penelitian ini. Daftar buku, *website*, dan sumber lain terlampir di dalam daftar pustaka.

## 2. Persiapan Instrumen Penelitian

Penulis melakukan pembuatan kuesioner yang mengacu pada 4 variable SES di Indonesia. Penulis juga melakukan pengujian kuesioner untuk menguji keefektifan instrumen survei (kuesioner) sebagai alat komunikasi antara peneliti dan responden.

## 3. Penyebaran Kuesioner

Penulis menyebarluaskan kuesioner secara daring (*online*) dengan menentukan jumlah dan karakteristik sampel responden yang sesuai dengan batas dan kriteria penelitian yaitu mahasiswa UIN Jakarta sebagai populasi.

## 4. Pengolahan Data

Penulis melakukan perhitungan *data testing* dan *data training* yang digunakan sebagai parameter pemodelan algoritma Catboost dengan data yang dibagi menjadi 20% *data testing* dan 80% *data training*.

## 5. Interpretasi Hasil

Penulis melakukan interpretasi atau penjabaran terhadap hasil yang diperoleh dari pengolahan data yang dilakukan pada penelitian dalam *performance metrics* yang terdiri dari *accuracy*, *confusion matrix*, *precision & recall*, *F1 score*, dan *Area under ROC curve(AUC)*.

## 6. Rekomendasi

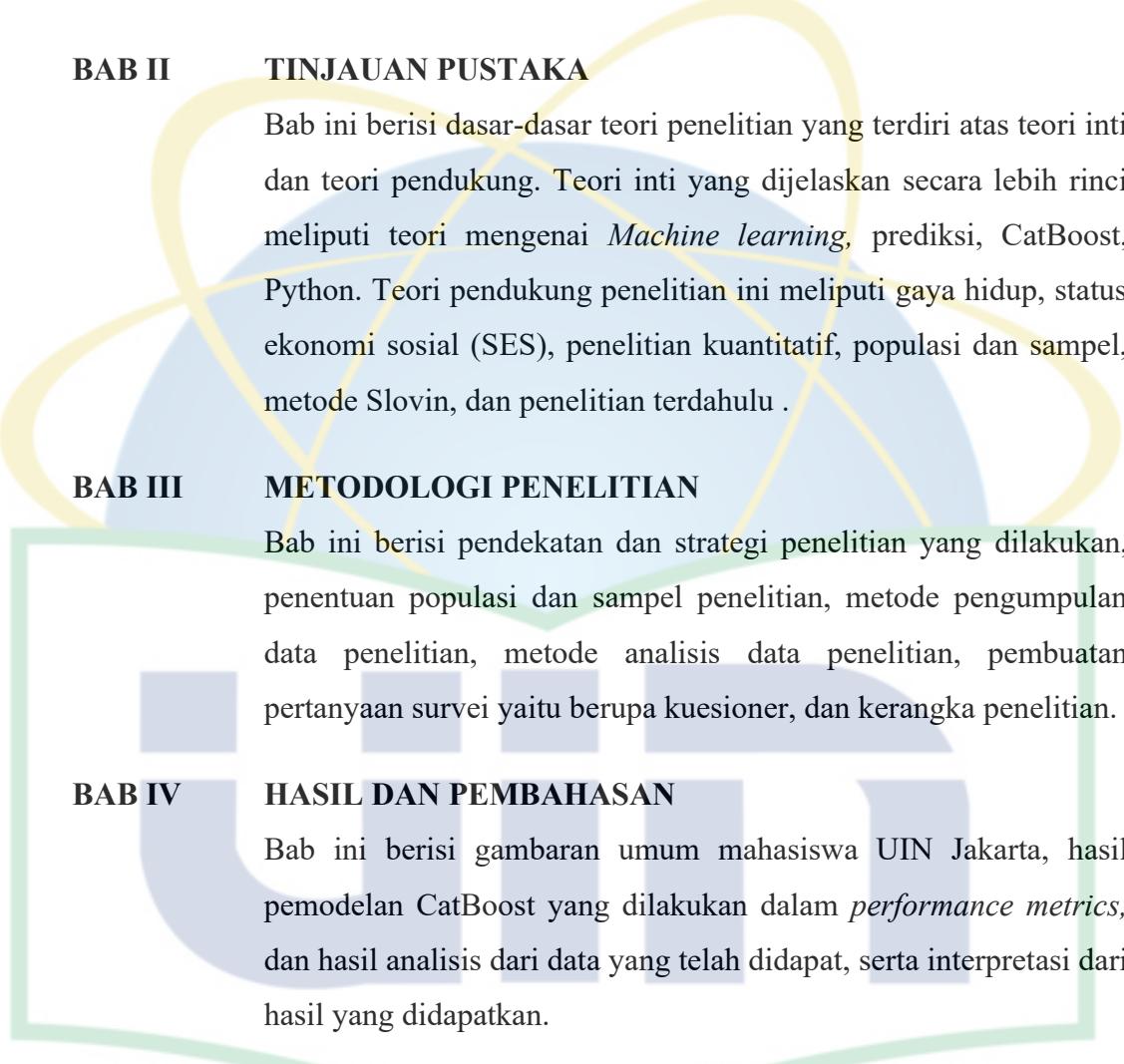
Penulis memberikan hasil prediksi data dan faktor yang mempengaruhi gaya hidup mahasiswa sesuai dengan akurasi yang didapatkan berdasarkan hasil penelitian.

## 1.8 Sistematika Penulisan

Sistematika penulisan pada penelitian ini terdiri dari lima struktur bab. Gambaran umum pokok pembahasan yang dibahas pada tiap-tiap bab tersebut adalah sebagai berikut.

### BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang masalah gaya hidup yang meningkat dengan adanya perkembangan tren dan mode



berdasarkan SES. Latar belakang pada penelitian ini menghasilkan identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, hipotesis penelitian, dan sistematika penulisan penelitian.

## BAB II

### TINJAUAN PUSTAKA

Bab ini berisi dasar-dasar teori penelitian yang terdiri atas teori inti dan teori pendukung. Teori inti yang dijelaskan secara lebih rinci meliputi teori mengenai *Machine learning*, prediksi, CatBoost, Python. Teori pendukung penelitian ini meliputi gaya hidup, status ekonomi sosial (SES), penelitian kuantitatif, populasi dan sampel, metode Slovin, dan penelitian terdahulu .

## BAB III

### METODOLOGI PENELITIAN

Bab ini berisi pendekatan dan strategi penelitian yang dilakukan, penentuan populasi dan sampel penelitian, metode pengumpulan data penelitian, metode analisis data penelitian, pembuatan pertanyaan survei yaitu berupa kuesioner, dan kerangka penelitian.

## BAB IV

### HASIL DAN PEMBAHASAN

Bab ini berisi gambaran umum mahasiswa UIN Jakarta, hasil pemodelan CatBoost yang dilakukan dalam *performance metrics*, dan hasil analisis dari data yang telah didapat, serta interpretasi dari hasil yang didapatkan.

## BAB V

### PENUTUP

Bab ini berisi kesimpulan dan saran atau rekomendasi atas penelitian yang telah dilakukan.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Gen Z**

Generasi Z (gen Z) merupakan generasi setelah generasi milenial yang lahir dari tahun 1995 sampai tahun 2012 (Stillman & Stillman, 2017). Gen Z saat ini memiliki rentang usia 11-26 tahun yang didominasi oleh pelajar dan mahasiswa memiliki karakteristik sebagai generasi pertama pada generasi internet (Grail, 2011).

Selain itu, menurut Tari Annamária pada bukunya yang berjudul Z “generáció” gen Z memiliki atribut *“net generation”* yang memiliki makna bahwa gen Z lahir pada saat berkembangnya era digital dan dicirikan sebagai *“Facebook-generation”*, *“digital native”* atau *“iGeneration”* (Annamária dikutip dalam Andrea et al., 2016). Hellen Chou P. menjelaskan bahwa Generasi Z merupakan generasi muda yang tumbuh dan berkembang dengan sebuah ketergantungan yang besar pada teknologi digital.

Stillman & Stillman mengungkapkan bahwa Generasi Z tidak hanya ditentukan oleh kurun waktu lahir dan perkembangan teknologi saja, namun generasi Z terbentuk oleh peristiwa dan kondisi yang dialami seperti keberagaman yang semakin menguat, ekonomi yang mengalami resesi, merebaknya terorisme, kondisi lingkungan hidup yang memburuk, kedulian terhadap politik, dan munculnya selebritas-selebritas internet sebagai patron

#### **2.2 Prediksi**

Pengertian prediksi sama dengan ramalan atau perkiraan. Menurut Kamus Besar Bahasa Indonesia (KBBI), prediksi merupakan hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang menggunakan data masa lalu. Prediksi menunjukkan yang akan terjadi pada suatu keadaan tertentu dan merupakan *input* bagi proses perencanaan dan pengambilan keputusan di masa yang akan datang.

Peramalan atau prediksi memerlukan pengambilan data historis dan menerapkannya ke masa depan (Heizer & Render, 2011). Prediksi dapat

memperkirakan beberapa kebutuhan pada periode yang datang meliputi kebutuhan dalam ukuran kuantitas (jumlah), kualitas (mutu), waktu dan lokasi yang dibutuhkan dalam rangka memenuhi permintaan barang atau jasa (Nasution, 2003).

Menurut Wantono, prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil (Wantono, 2014). Menurut Herdianto (2013), prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin dengan apa yang akan terjadi.

### 2.3 Gaya Hidup

Gaya hidup dalam KBBI merupakan pola tingkah laku sekelompok manusia yang dilakukan pada kehidupan sehari-hari mereka dalam lingkungan masyarakat.

*Lifestyle* (gaya hidup) merupakan perilaku individu dan juga *customer behavior* yang diwujudkan dalam bentuk aktivitas, minat, serta pandangan individu dalam mengaktualisasikan kepribadiannya karena adanya pengaruh interaksi dengan lingkungan sekitarnya (Widjaja, 2013:39).

Menurut Philip Kotler dan Kevin Lane Keller (2016:187) “*A lifestyle is a person pattern of life as expressed in activities, interests, and opinions. It portrays the whole person interacting with his or her environment.*”. Gaya Hidup adalah pola hidup seseorang yang diungkapkan dalam kegiatan, minat, dan pendapat, serta menggambarkan seseorang saat berinteraksi dengan lingkungannya.

Gaya hidup suatu masyarakat akan berbeda dengan masyarakat yang lainnya. Masa ke masa gaya hidup suatu individu dan kelompok masyarakat tertentu akan bergerak secara dinamis. Gaya hidup pada dasarnya merupakan suatu perilaku yang mencerminkan masalah sebenarnya yang terdapat di dalam alam pikir pelanggan yang cenderung berbaur dengan berbagai hal yang terkait dengan masalah emosi dan psikologis konsumen (Nugroho, 2010: 77).

Sutisna dalam Heru Suprihadi (2017) mendefinisikan gaya hidup secara luas sebagai cara hidup yang diidentifikasi oleh bagaimana orang lain menghabiskan

waktu mereka (aktivitas) dilihat dari pekerjaan, hobi, belanja, olahraga, dan kegiatan sosial serta interest (minat) terdiri dari makanan, mode, keluarga, rekreasi dan juga opinion (pendapat) terdiri dari mengenai diri mereka sendiri, masalah-masalah sosial, bisnis, dan produk.

Setiawan dan Handojo (2018: 3) mengemukakan bahwa gaya hidup mencakup sesuatu yang lebih dari sekedar kelas sosial ataupun kepribadian seseorang. Gaya hidup dapat mempengaruhi perilaku seseorang yang juga menentukan pola konsumsi serta pilihan-pilihannya sebagaimana orang tersebut menggunakan waktu dan uangnya (Saputra, 2020).

Konsumen termotivasi dalam melakukan pembelian karena unsur dan dorongan kebutuhan yang muncul karena *lifestyle*. Terdapat empat kategori yang menjadi motif dalam proses pembelian oleh konsumen karena *lifestyle* (Widjaja, 2013:43), yaitu:

1. *Utilitarian purchase* (pembelian produk bermanfaat)

Konsumen membelanjakan produk ini, dalam kondisi tidak sangat mendesak membutuhkan, tetapi memberikan keyakinan bahwa produk/jasa yang dibelinya akan meningkatkan kehidupan yang lebih baik atau lebih mudah.

2. *Indulgences* (kesukaan/memanjakan diri)

Individu mencoba untuk hidup menikmati sedikit kemewahan tanpa banyak menambah pengorbanan dari pengeluarannya. Gratifikasi dari produk/jasa ini terletak pada faktor emosional.

3. *Lifestyle luxuries* (gaya hidup mewah)

*Lifestyle luxuries* menawarkan manfaat dan kegunaan bagi konsumen berupa peningkatan *prestige, image, dan superior quality* dari suatu merek. Dalam hal ini, peranan merek menjadi gratifikasi konsumen untuk membeli produk atau jasa.

#### 4. *Aspirational luxuries* (hasrat kemewahan)

Seiring dengan *indulgences*, *aspirational luxuries* akan memuaskan konsumen dari aspek kebutuhan emosionalnya. Melalui pembelian, konsumen dapat mengekspresikan dirinya, sistem nilai, minat dan hasratnya. Kepuasan muncul dari emosi lebih besar daripada kepuasan pemenuhan kebutuhan praktis atau fungsional.

##### 2.3.1 Faktor yang Mempengaruhi Gaya Hidup

Menurut Donni Juni Priansa (2017:190) banyak faktor yang mempengaruhi gaya hidup konsumen, namun secara umum faktor tersebut dibagi menjadi dua, yaitu faktor internal dan faktor eksternal. Adaupun faktor-faktor tersebut antara lain:

###### 1. Faktor Internal Konsumen

Faktor internal konsumen terdiri dari sikap, pengalaman dan pengamatan, kepribadian, konsep diri, motif, dan persepsi dengan penjelasan sebagai berikut:

###### a. Sikap

Sikap merupakan kondisi jiwa yang merupakan refleksi dari pengetahuan dan cara berpikir konsumen untuk memberikan respon terhadap suatu objek yang diorganisasi melalui pengalaman dan mempengaruhi secara langsung pada perilaku yang ditampilkannya. Kondisi tersebut sangat dipengaruhi oleh tradisi, kebiasaan, kebudayaan, serta lingkungan sosialnya.

###### b. Pengalaman dan Pengamatan

Pengalaman dan pengamatan merupakan hal yang saling erat terkait. Pengalaman dapat mempengaruhi pengamatan sosial dalam tingkah laku, pengalaman dapat diperoleh dari semua tingkah laku dan perbuatan konsumen di masa lampau serta dapat dipelajari melalui interaksi dengan orang lain yang selanjutnya menghasilkan pengalaman. Hasil dari pengalaman sosial tersebut dapat membentuk terhadap suatu objek.

c. Kepribadian

Kepribadian merupakan konfigurasi karakteristik dari individu konsumen dan cara berperilaku yang menentukan perbedaan perilaku dari setiap individu.

d. Konsep diri

Konsep diri erat kaitannya dengan citra merek dari produk yang dikonsumsi. Bagaimana konsumen secara individu memandang tentang dirinya akan sangat mempengaruhi minatnya terhadap suatu objek. Konsep diri merupakan inti dari pola kepribadian yang akan menentukan perilaku individu dalam menghadapi permasalahan hidupnya, karena konsep diri merupakan *frame of reference* yang menjadi awal timbulnya perilaku yang ditampilkan oleh konsumen.

e. Motif

Perilaku individu muncul karena adanya motif kebutuhan dan keinginan yang menyertainya. Konsumen membutuhkan dan menginginkan untuk merasa aman serta memiliki akan prestise lebih besar maka akan membentuk gaya hidup yang cenderung mengarah kepada gaya hidup hedonis.

f. Persepsi

Persepsi merupakan proses dimana konsumen memilih, mengatur, dan menginterpretasikan informasi yang diterimanya untuk membentuk suatu gambar tertentu atas informasi tersebut.

## 2. Faktor Eksternal

Faktor eksternal yang mempengaruhi gaya hidup konsumen terdiri dari kelompok referensi, keluarga, kelas sosial, dan kebudayaan. Masing-masing diuraikan sebagai berikut:

a. Kelompok Referensi

Kelompok referensi merupakan kelompok yang memberikan pengaruh langsung atau tidak langsung terhadap sikap dan perilaku konsumen. Kelompok yang memberikan pengaruh langsung adalah kelompok dimana konsumen tersebut menjadi anggotanya dan saling berinteraksi, sedangkan kelompok yang memberi pengaruh tidak langsung adalah kelompok dimana konsumen tidak menjadi anggota didalam kelompok tersebut. Pengaruh-pengaruh tersebut akan menghadapkan konsumen pada perilaku dan gaya hidup tertentu.

b. Keluarga

Keluarga memegang peranan terbesar dan terlama dalam pembentukan sikap dan perilaku konsumen. Hal ini karena pola asuh orang tua akan membentuk kebiasaan anak yang secara tidak langsung mempengaruhi pola hidupnya.

c. Kelas Sosial

Kelas sosial merupakan kelompok yang relatif homogen dan bertahan lama dalam sebuah masyarakat, yang tersusun dalam sebuah urutan jenjang, dimana para anggota dalam setiap jenjang tersebut memiliki nilai, minat, dan tingkah laku yang sama. Ada dua unsur pokok dalam sistem sosial pembagian kelas dalam masyarakat, yaitu kedudukan (status) dan peranan. Kedudukan sosial artinya tempat seseorang dalam lingkungan pergaulan, prestise hak-haknya serta kewajibannya. Kedudukan sosial ini dapat dicapai oleh seseorang dengan usaha yang sengaja maupun diperoleh karena kelahiran. Peranan merupakan aspek yang dinamis dari kedudukan. Apabila individu melaksanakan hak dan kewajibannya sesuai dengan kedudukannya maka ia menjalankan suatu peranan.

d. Kebudayaan

Kebudayaan yang meliputi pengetahuan, kepercayaan, kesenian, moral, hukum, adat istiadat, dan kebiasaan-kebiasaan yang diperoleh konsumen

sebagai individu yang merupakan bagian dari anggota masyarakat. Kebudayaan terdiri dari segala sesuatu yang dipelajari dari pola-pola perilaku yang normatif, meliputi ciri-ciri pola pikir, merasakan dan bertindak.

#### 2.4 Socioeconomic Status (SES)

Indonesia Data mendefinisikan SES sebagai cara mengelompokkan individu maupun sebuah keluarga berdasarkan kemampuan ekonomi dan status sosialnya. Menurut American Psychological Association, SES adalah posisi seseorang atau kelompok dalam skala sosial ekonomi, yang ditentukan oleh kombinasi faktor sosial dan ekonomi seperti pendapatan, jumlah dan jenis pendidikan, jenis dan prestise pekerjaan, tempat tinggal, dan dalam beberapa masyarakat atau bagian masyarakat—asal etnis atau latar belakang agama.

Selain itu, *National Center for Education Statistics* (ED) juga mengartikan SES secara luas sebagai akses seseorang ke sumber daya keuangan, sosial, budaya, dan sumber daya manusia. Secara tradisional SES siswa telah memasukkan, sebagai komponen, pencapaian pendidikan orang tua, status pekerjaan orang tua, dan pendapatan rumah tangga atau keluarga, dengan penyesuaian yang sesuai untuk komposisi rumah tangga atau keluarga. Ukuran SES yang diperluas dapat mencakup ukuran rumah tangga tambahan, lingkungan sekitar, dan sumber daya sekolah.

Pemeriksaan status sosial ekonomi sering mengungkapkan ketidaksetaraan dalam akses ke sumber daya, serta isu-isu yang berkaitan dengan hak istimewa, kekuasaan, dan kontrol. Dalam riset pemasaran (*marketing research*) SES digunakan sebagai cara untuk mengelompokkan dan mengetahui individu maupun rumah tangga berdasarkan kemampuan ekonomi dan status sosialnya kedalam klasifikasi SES.

Klasifikasi SES terdiri dari 3 segmen utama yakni *high* SES, *middle* SES dan *low* SES. SES dapat diukur dari beberapa variabel tergantung lembaga atau institusi yang melakukan riset. Perusahaan riset pasar yang populer mengadakan pengukuran SES di Indonesia adalah Nielsen. Selain dari Nielsen, Indonesia Data,

Talenta Data Indonesia (TDI) juga mengeluarkan laporan bisnis yang dapat mengetahui klasifikasi SES masyarakat Indonesia.

#### **2.4.1 Variabel Socioeconomic Status (SES)**

Variabel SES menurut Indonesia Data 2022 sebagai berikut:

##### A. Pengeluaran Rutin Bulanan Rumah Tangga

Pengeluaran rutin bulanan rumah tangga adalah seberapa besar pengeluaran individu maupun rumah tangga untuk kebutuhan sehari-hari. Kebutuhan yang dimaksud antara lain, makanan dan minuman, pakaian, kosmetik, *toiletries, personal care*, obat-obatan, transportasi, listrik, air, bahan bakar untuk masak, biaya bulanan sekolah, biaya produksi bulanan, dan pengeluaran rutin lainnya. Indikator-indikator tersebut dapat disesuaikan berdasarkan kebutuhan riset dilakukan.

Terdapat hal yang tidak termasuk dalam indikator variabel ini yakni pengeluaran yang sifatnya dalam jumlah besar maupun tahunan seperti, pembayaran cicilan, pembelian barang elektronik, pembelian mobil/motor, kontrakan tahunan, modal usaha, pembelian barang mewah, nonton konser/acara, bioskop hingga pertandingan olahraga, dan liburan/jalan-jalan.

Dari variabel pengeluaran rutin bulanan rumah tangga yang telah dijelaskan, nantinya akan diperoleh skor (kuantifikasi) yang mana skor tersebut akan dihitung/dikombinasikan dengan jumlah anggota rumah tangga.

##### B. Jumlah Anggota Rumah Tangga

Seperti yang telah disebutkan sebelumnya, skor dari pengeluaran rutin bulanan akan dikombinasikan dengan jumlah anggota rumah tangga. Untuk tingkat pengeluaran yang sama, semakin banyak anggota rumah tangga, maka skor akan semakin rendah. Berikut merupakan gambar tabel skor pengeluaran rutin bulanan dan jumlah anggota rumah tangga.

Pengeluaran rutin bulanan	Jumlah anggota rumah tangga																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
IDR/Rupiah																				
S/d 300.000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
300.001 - 400.000	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
400.001 - 500.000	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
500.001 - 600.000	6	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
600.001 - 700.000	6	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
700.001 - 750.000	6	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
750.001 - 800.000	7	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
800.001 - 900.000	7	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
900.001 - 1.000.000	7	5	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.000.001 - 1.250.000	7	6	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.250.001 - 1.500.000	8	6	4	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.500.001 - 1.750.000	8	7	6	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.750.001 - 2.000.000	8	7	6	5	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
2.000.001 - 2.250.000	8	7	6	6	4	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1
2.250.001 - 2.500.000	8	7	7	6	5	3	3	2	2	1	1	1	1	1	1	1	1	1	1	1
2.500.001 - 2.750.000	9	8	7	6	5	4	3	2	2	2	2	1	1	1	1	1	1	1	1	1
2.750.001 - 3.000.000	9	8	7	6	6	5	4	3	2	2	2	1	1	1	1	1	1	1	1	1
3.000.001 - 3.250.000	9	8	7	7	6	5	4	3	2	2	2	1	1	1	1	1	1	1	1	1
3.250.001 - 3.500.000	9	8	7	7	6	6	5	4	3	3	2	2	2	2	1	1	1	1	1	1
3.500.001 - 4.000.000	9	8	7	7	6	6	5	4	4	3	3	2	2	2	2	1	1	1	1	1
4.000.001 - 4.500.000	9	8	8	7	7	6	6	5	4	3	3	2	2	2	2	2	1	1	1	1
4.500.001 - 5.000.000	9	8	8	7	7	6	5	5	4	3	3	2	2	2	2	2	1	1	1	1
5.000.001 - 6.000.000	9	9	8	8	8	7	6	5	5	4	3	3	2	2	2	2	2	2	2	2
6.000.001 - 7.000.000	9	9	8	8	8	7	6	6	6	5	4	3	3	3	3	3	3	3	3	2
7.000.001 - 8.000.000	9	9	8	8	8	8	7	6	6	6	5	4	4	4	4	4	4	4	4	4
8.000.001 - 9.000.000	9	9	9	8	8	8	7	7	7	6	6	5	5	5	5	5	5	5	5	5
9.000.001 - 10.000.000	9	9	9	9	8	8	8	7	7	7	6	6	6	6	6	6	6	6	6	6
10.000.001 - 12.500.000	9	9	9	9	9	8	8	8	8	7	7	7	7	7	7	7	7	7	7	7
12.500.001 - 15.000.000	9	9	9	9	9	9	8	8	8	8	7	7	7	7	7	7	7	7	6	6
Lebih dari Rp 15.000.000	9	9	9	9	9	9	9	8	8	8	8	7	7	7	7	7	7	7	6	6

Gambar 2.1 Skor Jumlah Anggota Keluarga

Sumber: Data Olahan Indonesia Data.

### C. Sumber Air Minum Rumah Tangga

Selain variabel pengeluaran rutin bulanan dan jumlah anggota rumah tangga yang telah dijelaskan di atas, terdapat variabel lainnya yaitu sumber air minum rumah tangga. Sumber air minum yang paling sering digunakan di rumah tangga. Adapun skor untuk variabel ini dapat dilihat pada tabel berikut:

Tabel 2.1 Skor Air Minum Rumah Tangga

Sumber: Indonesia Data

Sumber Air Minum	Skor
Air minum dalam kemasan botol/galon bermerek	4
Air minum dalam galon isi ulang	3
Air PAM	3
Air keran non-PAM	2
Air sumur bor/pompa	2
Air sumur tampung	1
Air sumur yang tidak ditampung	1
Mata air yang ditampung	0

Sumber Air Minum	Skor
Mata air yang tidak ditampung	0
Air sungai	0
Air hujan	0
Lainnya, sebutkan: _____	0

D. Sumber Bahan Bakar Untuk Memasak

Sumber bahan bakar untuk memasak yang dimaksud adalah sumber bahan bakar yang paling sering digunakan untuk memasak di rumah tangga. Adapun skor untuk variabel ini dapat dilihat pada tabel berikut:

Tabel 2.2 Skor Sumber Bahan Bakar untuk Memasak  
Sumber: Indonesia Data

Sumber Bahan Bakar Masak	Skor
Kompor listrik	5
Gas Elpiji – 12 Kg/ Gas 7 Kg/Gas 5 Kg	4
Gas 3 Kg	2
Gas kota/Alam	4
Minyak tanah	2
Kayu	1
Arang	1
Bricket	1
Lainnya (tolong sebutkan.....)	1
Tidak Memasak	
Jika Pengeluaran Rumah Tangga < 2,5 juta per bulan	2
Jika Pengeluaran Rumah Tangga 2,5 – 5,0 juta per bulan	3
Jika Pengeluaran Rumah Tangga > 5,0 juta per bulan	4

Dari penjelasan 4 variabel di atas (Variabel A, Variabel B, Variabel C dan Variabel D) skor dari individu maupun rumah tangga yang diperoleh, diverifikasi dan dianalisa dengan desain riset yang telah

ditetapkan. Hasil olahan dari skor ini yang menjadi tolok ukur dalam menentukan kelas atau kategori SES individu maupun rumah tangga untuk suatu wilayah.

Setiap lembaga riset dapat menentukan sendiri variabel, metode dan indikator yang digunakan untuk menentukan kategori SES. Dalam riset Indonesia Data, TDI menggunakan metode ini untuk menentukan kategori SES. Adapun indikator yang kami gunakan adalah SES A, SES B, SES C dan SES D/E.

#### 2.4.2 Klasifikasi *Socioeconomic Status* (SES)

Setiap variabel tersebut memiliki skor tersendiri. Setiap skor di masing-masing variabel akan dijumlahkan sehingga total skor akan dikonversikan menjadi klasifikasi SES. Semakin besar skornya maka status atau kelas SES akan semakin tinggi. Tabel berikut memperlihatkan jangkauan skor (*range score*) untuk setiap kelas SES menurut Indonesia Data.

Tabel 2.3 Kategori SES  
Sumber: Indonesia Data

Score	SES	Category
15-18	SES A	Upper 1
13-14	SES B	Upper 2
11-12	SES C1	Middle 1
8-10	SES C2	Middle 2
5-7	SES D	Lower 1
2-4	SES E	Lower 2

#### 2.5 Machine Learning

*Machine learning* (ML) adalah bidang ilmu komputer yang mempelajari algoritma dan teknik untuk mengotomatisasi solusi untuk masalah kompleks yang sulit diprogram menggunakan metode pemrograman konvensional (Rebala, 2019). Selain itu, ML juga dapat diartikan sebagai cara untuk mengekstraksi pengetahuan dari data yang mana pemrograman komputer untuk mengoptimalkan kriteria kinerja dengan menggunakan data contoh atau pengalaman masa lalu.

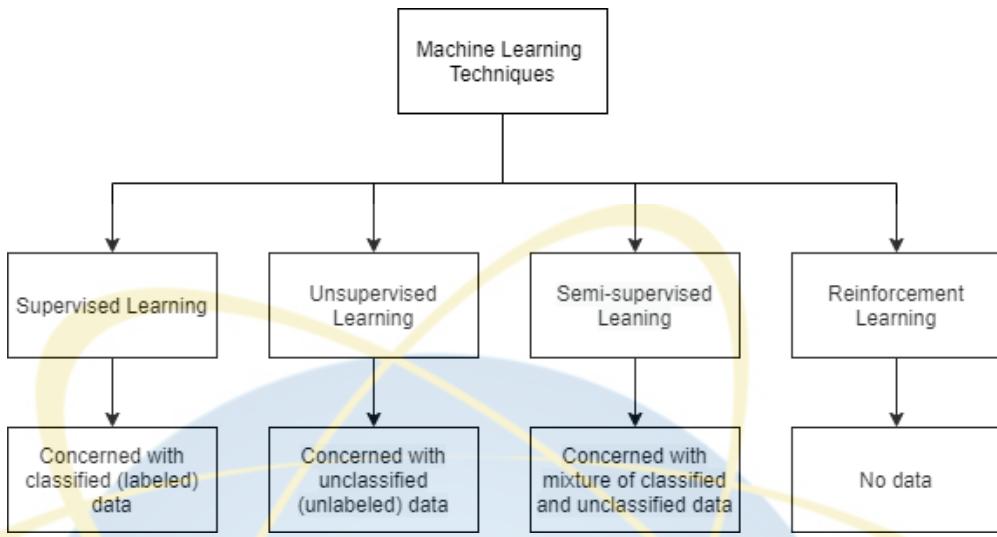
ML bertujuan untuk menghasilkan ekspresi klasifikasi yang cukup sederhana untuk dipahami dengan mudah oleh manusia. ML harus memiliki cukup data untuk meniru penalaran manusia yang dapat memberikan wawasan ke dalam proses pengambilan keputusan. Seperti pendekatan statistik, latar belakang pengetahuan dapat dieksplorasi dalam pengembangan, tetapi pengoperasian diasumsikan tanpa campur tangan manusia (IV et al., 1996).

ML biasanya mengacu pada perubahan dalam sistem yang melakukan tugas yang terkait dengan kecerdasan buatan (AI). Tugas-tugas tersebut melibatkan pengenalan, perencanaan, kontrol robot, prediksi, dll (Nilsson, 2005). Penerapan metode *machine learning* dalam beberapa tahun terakhir menjadi umum dalam kehidupan sehari-hari. Mulai dari rekomendasi otomatis film mana yang akan ditonton, hingga makanan apa yang akan dipesan atau produk mana yang akan dibeli, hingga radio online yang dipersonalisasi dan mengenali teman di album foto pada *smartphone*, banyak situs web dan perangkat modern memiliki algoritma ML sebagai intinya. Saat melihat situs web yang kompleks seperti Facebook, Amazon, atau Netflix, kemungkinan besar setiap bagian situs berisi beberapa model ML (Müller & Guido, 2015).

ML juga mengeksplorasi dan mengembangkan model matematika dan algoritma untuk belajar dari data. Paradigmanya berfokus pada tujuan klasifikasi dan terdiri dari pemodelan pemetaan optimal antara domain data dan kumpulan pengetahuan dan mengembangkan algoritma pembelajaran. Klasifikasi ini juga disebut *supervised learning*, yang membutuhkan kumpulan data pelatihan (berlabel), kumpulan data validasi, dan kumpulan data uji. Kumpulan data pelatihan digunakan untuk menemukan parameter optimal dari suatu model. Kumpulan data validasi juga dapat menghindari *overfitting model*, dan digunakan untuk menentukan keakuratan model (Suthaharan, 2016).

### 2.5.1 Teknik *Machine Learning*

Berdasarkan metode dan cara pembelajarannya, teknik *machine learning* dibagi menjadi empat jenis, yaitu: *Supervised Machine Learning*, *Unsupervised Machine Learning*, *Semi-Supervised Machine Learning*, *Reinforcement Learning*.



Gambar 2.2 Machine learning techniques and data requirements

Sumber: (Mitchell M.T., 2006 dalam Makovskaja,2018)

### 1. *Supervised Learning*

Dalam jenis sistem ML ini, data yang dimasukkan ke dalam algoritma, dengan solusi yang diinginkan, disebut sebagai "label". Algoritma *supervised learning* harus mempelajari karakteristik utama dalam setiap titik data dalam kumpulan data untuk menentukan jawabannya. Jadi, saat titik data baru diberikan ke algoritma, berdasarkan karakteristik kunci, algoritma harus dapat memprediksi hasil/jawaban yang benar.

### 2. *Unsupervised Learning*

Dalam jenis sistem ML ini, data yang dimasukkan ke dalam algoritma, dengan solusi yang diinginkan, disebut sebagai "tidak berlabel". Masalah *unsupervised learning* yaitu mencoba menemukan struktur tersembunyi dalam data. Pengelompokan, yang melibatkan penemuan kategori dari data, adalah bentuk dasar dari *unsupervised learning*, dan di mana telah mencapai banyak keberhasilan. Algoritma ini akan mengidentifikasi cluster atau kelompok item yang mirip atau kesamaan item baru dengan grup yang sudah ada, dll (Neapolitan & Jiang, 2018).

### 3. *Semi-supervised Learning*

*Semi-supervised learning* berada di antara pembelajaran yang diawasi dan tidak diawasi. Di sini, mesin diberi kumpulan data yang besar, di mana hanya beberapa titik data yang diberi label. Algoritma akan menggunakan teknik pengelompokan ini untuk mengidentifikasi grup dalam kumpulan data yang diberikan dan menggunakan beberapa titik data berlabel dalam setiap grup untuk memberikan label ke titik data lain dalam kluster/grup yang sama. Salah satu manfaat paling menonjol dari teknik ini adalah tidak perlu menghabiskan banyak waktu dan tenaga untuk melabeli setiap titik data, yang bisa menjadi proses yang sangat manual (Rebala et al., 2019).

### 4. *Reinforcement Learning*

*Reinforcement learning* adalah paradigma pembelajaran yang berkaitan dengan belajar untuk mengontrol sistem sehingga memaksimalkan ukuran kinerja numerik yang mengungkapkan tujuan jangka panjang. *Reinforcement learning* berbeda dari *supervised learning* dimana *reinforcement learning* hanya sebagian umpan balik yang diberikan kepada pelajar tentang prediksi pelajar. Selanjutnya, prediksi mungkin memiliki efek jangka panjang melalui pengaruh keadaan masa depan dari sistem yang dikendalikan. Jadi, waktu memainkan peran khusus. Tujuan dalam *reinforcement learning* adalah untuk mengembangkan algoritma pembelajaran yang efisien, serta untuk memahami kelebihan dan kekurangan algoritma. *Reinforcement learning* sangat menarik karena banyaknya aplikasi praktis yang dapat digunakan untuk mengatasi, mulai dari masalah dalam kecerdasan buatan hingga riset operasi atau teknik kontrol.

## 2.6 Klasifikasi

Menurut Kamus Besar Bahasa Indonesia (KBBI), klasifikasi adalah penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar

yang ditetapkan. Klasifikasi pada ML merupakan metode dari *supervised learning*. Klasifikasi bertujuan adalah untuk membuat model yang dapat memprediksi label yang benar dari data input yang diberikan. Dalam klasifikasi, model sepenuhnya dilatih menggunakan data pelatihan, dan kemudian dievaluasi pada data uji sebelum digunakan untuk melakukan prediksi pada data baru yang tidak terlihat. Klasifikasi terkadang dipisahkan menjadi klasifikasi biner, yang merupakan kasus khusus untuk membedakan tepat antara dua kelas, dan klasifikasi multikelas, yaitu klasifikasi antara lebih dari dua kelas. Klasifikasi biner sebagai upaya untuk menjawab pertanyaan ya/tidak. Mengklasifikasikan email sebagai spam atau bukan spam adalah contoh masalah klasifikasi biner.

### 2.6.1 Algoritma Klasifikasi

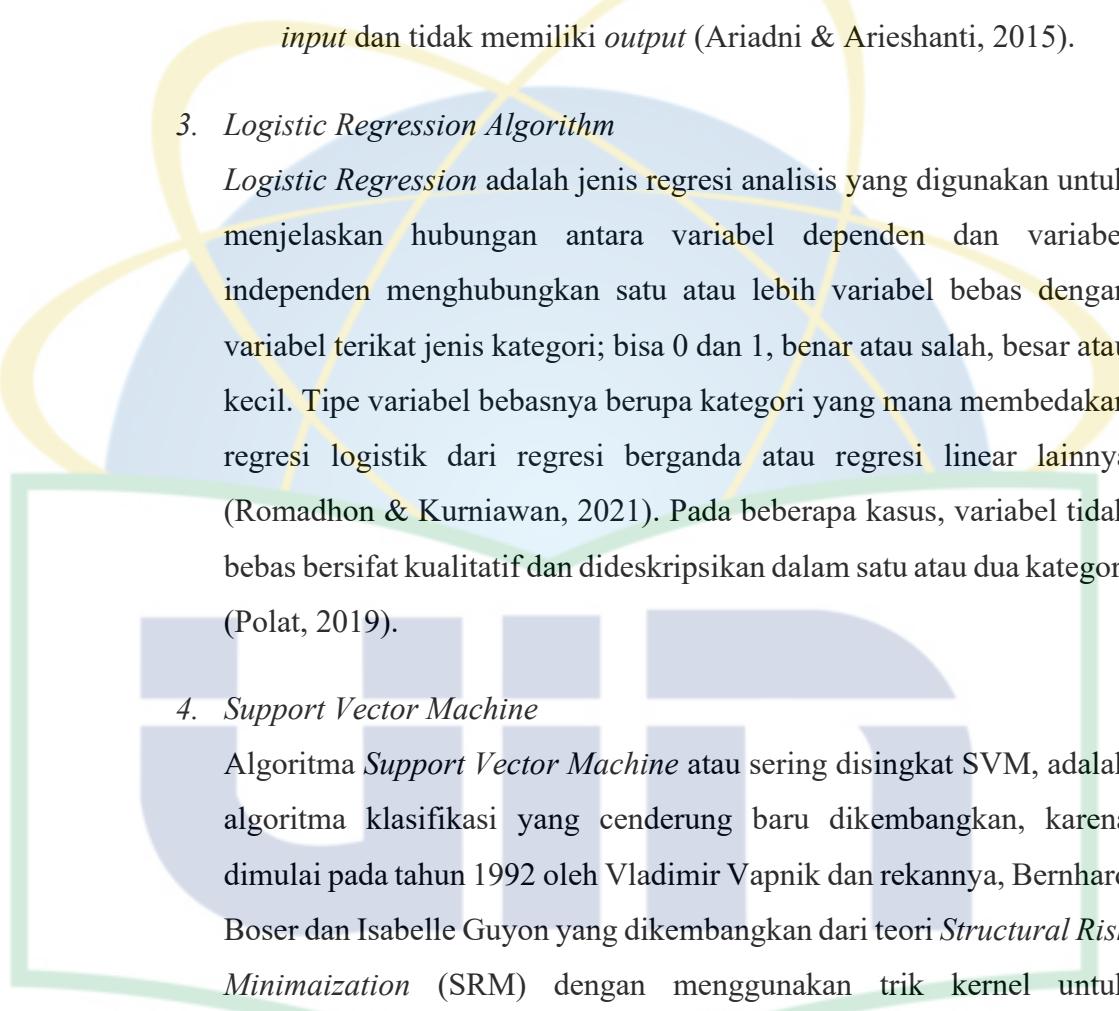
#### 1. Random Forest Algorithm

Random Forest adalah salah satu jenis algoritma klasifikasi yang terdiri dari lebih satu pohon keputusan yang setiap pohon keputusan dibentuk bergantung pada nilai-nilai vector acak sampel secara independen dan identik didistribusikan yang sama untuk semua pohon (Breiman, 2011). Random Forest masuk ke dalam kelompok Supervised Learning yang dikembangkan oleh Leo Breiman.

#### 2. Decision Tree Algorithm

*Decision tree* merupakan metode klasifikasi yang menggunakan representasi struktur pohon berisi alternatif untuk pemecahan masalah. Peranan pohon keputusan yaitu *decision support tool* yang membantu manusia dalam mengambil keputusan (Ariadni & Arieshanti, 2015). Pada dasarnya *decision tree* mengambil keputusan dari yang rumit menjadi lebih sederhana dengan konsep mengubah data menjadi suatu pohon keputusan dan aturan-aturan keputusan (*rule*). *Decision tree* menggunakan metode pembelajaran supervised learning, yaitu data akan diklasifikasikan berdasarkan data training atau data latih yang ada.

*Decision tree* terbagi menjadi 3 buah bagian, diantaranya:

- 
- 1) *Root Node*, node ini merupakan node yang terletak paling atas dari suatu algoritma pohon keputusan (Wijaya, 2019)
  - 2) *Internal Node*, node ini merupakan percabangan, dimana hanya terdapat satu input serta dua output.
  - 3) *Leaf Node*, node ini merupakan node akhir yang memiliki satu *input* dan tidak memiliki *output* (Ariadni & Arieshanti, 2015).
3. *Logistic Regression Algorithm*
- Logistic Regression* adalah jenis regresi analisis yang digunakan untuk menjelaskan hubungan antara variabel dependen dan variabel independen menghubungkan satu atau lebih variabel bebas dengan variabel terikat jenis kategori; bisa 0 dan 1, benar atau salah, besar atau kecil. Tipe variabel bebasnya berupa kategori yang mana membedakan regresi logistik dari regresi berganda atau regresi linear lainnya (Romadhon & Kurniawan, 2021). Pada beberapa kasus, variabel tidak bebas bersifat kualitatif dan dideskripsikan dalam satu atau dua kategori (Polat, 2019).

4. *Support Vector Machine*
- Algoritma *Support Vector Machine* atau sering disingkat SVM, adalah algoritma klasifikasi yang cenderung baru dikembangkan, karena dimulai pada tahun 1992 oleh Vladimir Vapnik dan rekannya, Bernhard Boser dan Isabelle Guyon yang dikembangkan dari teori *Structural Risk Minimization* (SRM) dengan menggunakan trik kernel untuk memetakan sampel pelatihan dari ruang input ke ruang fitur dimensi tinggi (Li et al., 2008). SVM adalah sebuah algoritma yang bekerja menggunakan pemetaan nonlinear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam dimensi yang baru, kemudian akan mencari linear optimal pemisah *hyperplane* (yaitu, “*decision boundary*” yang memisahkan tupel dari satu kelas dengan kelas lainnya) (Han et al., 2022).

## 2.6.2 Boosting

Pada awal 1990-an algoritma *boosting* muncul (Schapire 1990; Freund 1995; Schapire 1999), yang dipengaruhi oleh teori pembelajaran (Valiant 1984; Kearns dan Valiant 1989), di mana sejumlah pengklasifikasi lemah (pengklasifikasi yang memprediksi sedikit lebih baik daripada acak) digabungkan (atau ditingkatkan) untuk menghasilkan pengklasifikasi ansambel dengan tingkat kesalahan klasifikasi umum yang unggul (Kuhn & Johnson, 2013).

*Boosting* dirancang untuk masalah klasifikasi tetapi juga dapat diperluas secara menguntungkan ke regresi juga (Hastie et al, 2008). *Boosting* mengasumsikan ketersediaan algoritma pembelajaran dasar atau lemah yang, diberikan contoh pelatihan berlabel dan menghasilkan pengklasifikasi dasar atau lemah.

Berikut merupakan beberapa algoritma *boosting* yang biasa digunakan dalam *machine learning*. Beberapa yang paling populer antara lain:

### 1. AdaBoost (Adaptive Boosting)

Pendekatan *boosting* yang digunakan dalam *machine learning*. Freund dan Schapire berkolaborasi untuk menghasilkan algoritma AdaBoost (Schapire 1999). AdaBoost yang memberikan implementasi praktis dari konsep Kerns dan Valiant untuk meningkatkan pembelajar yang lemah menjadi pembelajar yang kuat (Kearns dan Valiant 1989). Bobot yang dialokasikan kembali ke setiap instans, dengan bobot yang lebih tinggi diterapkan ke instans yang diidentifikasi secara tidak tepat yang juga disebut *adaptive boosting*. Dalam *supervised learning*, *boosting* digunakan untuk mengurangi bias dan variasi. Hal ini didasarkan pada gagasan belajar berturut-turut (Chadaga et al., 2021).

### 2. Gradient Boosting

Algoritma *boosting gradient boosting* merupakan cara yang dilakukan oleh setiap prediktor dalam peningkatan gradien untuk mengoreksi kesalahan pendahulunya. Tidak seperti Adaboost, bobot instans pelatihan tidak disesuaikan; sebagai gantinya, setiap prediktor dilatih menggunakan

kesalahan residual pendahulu sebagai label. Penyusutan adalah karakteristik penting untuk dipertimbangkan saat menggunakan pendekatan ini. Setiap pohon dalam ansambel dikalikan dengan laju pembelajaran dan berkisar dari 0 hingga 1. Pada setiap iterasi, prediksi pohon dalam ansambel berkurang (Chadaga et al., 2021).

### 3. XGBoost (*Extreme Gradient Boosting*)

Varian dari *gradient boosting* yang menggunakan model yang lebih teratur untuk mencegah *overfitting*. Implementasi kerangka *gradient boostung* yang efisien dan terukur oleh (Friedman, 2001; Friedman et al., 2000). XGBoost ini mencakup pemecah model linier yang efisien dan algoritma pembelajaran pohon dan dirancang agar sangat skalabel dan dapat menangani kumpulan data yang sangat besar. XGBoost juga mendukung berbagai fungsi tujuan, termasuk regresi, klasifikasi, dan pemeringkatan (Chen & He, 2014).

### 4. LightGBM (Light Gradient Boosting Machine)

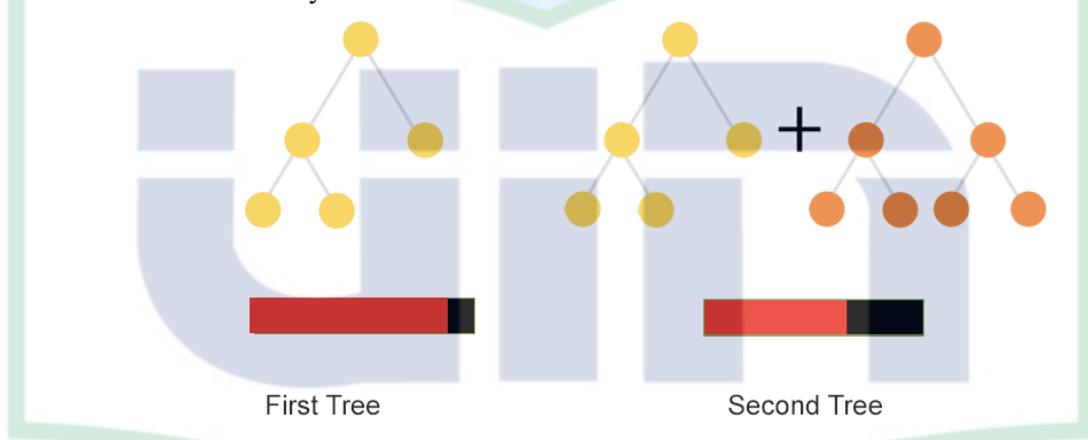
LightGBM adalah kerangka peningkatan berdasarkan algoritma pohon keputusan yang dirilis oleh Microsoft pada tahun 2017. LightGBM dan XGBoost keduanya mendukung aritmatika paralel, tetapi LightGBM lebih kuat daripada model XGBoost sebelumnya, dengan kecepatan pelatihan yang cepat dan penggunaan memori yang lebih sedikit, yang dapat mengurangi komunikasi biaya pembelajaran paralel. LightGBM terutama ditampilkan oleh algoritma pohon keputusan berdasarkan pengambilan sampel satu sisi berbasis gradien (GOSS), bundel fitur eksklusif (EFB), dan strategi pertumbuhan histogram dan daun-bijaksana dengan batas kedalaman (Wang & Wang, 2020).

### 5. CatBoost (*Categorical Boosting*)

Algoritma *machine learning* ini seperti *gradient boosting* dan *decision tree* memberikan hasil yang paling akurat. CatBoost digunakan untuk menangani fitur kategorikal dan juga lebih cepat dibandingkan dengan

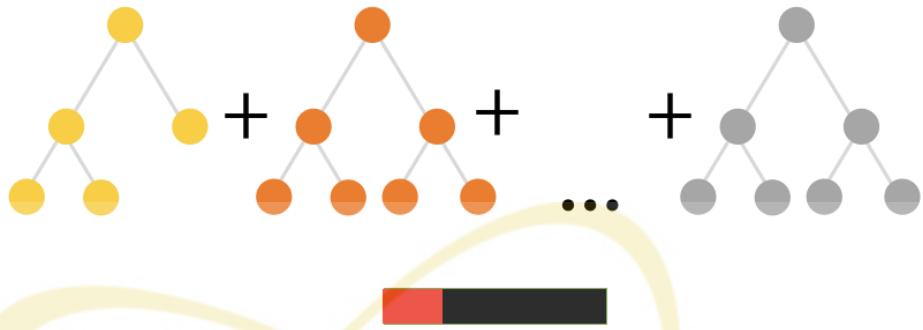
algoritma penguat lainnya karena mengimplementasikan pohon simetris. Dengan perubahan data dalam kumpulan data dari waktu ke waktu, algoritma CatBoost adalah algoritma yang paling kompeten, bekerja dengan baik dengan kumpulan data besar dan memiliki persyaratan latensi rendah (Barua et al., 2021).

CatBoost yang merupakan implementasi dari *Gradient Boosting on Decision Trees* (GBDT) memiliki kombinasi *gradient boosting* dengan pohon keputusan memberikan hasil canggih di banyak aplikasi dengan data terstruktur. Pengembang CatBoost menggunakan teknik tersebut untuk membuat model ansambel dengan cara berulang. Pada iterasi pertama, algoritma mempelajari pohon pertama untuk mengurangi kesalahan pelatihan, ditunjukkan pada gambar sebelah kiri pada gambar 2.3. Model ini biasanya memiliki kesalahan yang signifikan, sehingga bukan ide yang baik untuk membangun pohon yang sangat besar dalam meningkatkan karena terlalu menyesuaikan data.



Gambar 2.3 Iterasi Pohon pertama dan kedua pada CatBoost

Gambar sebelah kanan pada gambar 2.3 menunjukkan iterasi kedua, di mana algoritma mempelajari satu pohon lagi untuk mengurangi kesalahan yang dibuat oleh pohon pertama. Algoritma mengulangi prosedur ini hingga membangun mode kualitas yang layak, seperti yang dapat dilihat pada gambar 2.4:



Gambar 2.4 Iterasi Pohon ke-N

Pendekatan umum untuk klasifikasi pada CatBoost dengan *Logloss* dan tugas pemeringkatan umumnya mengimplementasikan beberapa variasi *LambdaRank*. Setiap langkah *Gradient Boosting* menggabungkan dua langkah:

1. Menghitung gradien dari fungsi kerugian yang ingin kami optimalkan untuk setiap objek masukan
2. Mempelajari pohon keputusan yang memprediksi gradien dari fungsi kerugian.

GBDT pada CatBoost membutuhkan pohon keputusan yang sesuai secara iteratif. Pohon keputusan klasifikasi yang digunakan mempelajari dengan cara serakah; mengharuskan menghitung semua kemungkinan pemisahan fitur (nilai fitur kurang dari beberapa nilai yang ditentukan sebelumnya) dari semua fitur dalam data, lalu memilih fitur yang meningkatkan fungsi kerugian dengan nilai terbesar.

Setelah belahan pertama dipilih, belahan berikutnya dalam pohon akan dipilih dengan cara belahan pertama diperbaiki dan belahan berikutnya dipilih dengan memberikan belahan pertama. Operasi ini diulang sampai seluruh pohon dibangun. Skema pembelajaran CatBoost pada dasarnya bersifat mendalam dengan beberapa penyederhanaan, yang diperoleh dari jenis pohon keputusan. Pemilihan *oblivious trees* memiliki beberapa keunggulan dibandingkan dengan pohon klasik:

1. Skema pemasangan sederhana
2. Efisien untuk diimplementasikan pada CPU
3. Kemampuan untuk membuat aplikasi model yang sangat cepat

Struktur pohon yang digunakan di CatBoost juga berfungsi sebagai regularisasi, sehingga dapat memberikan manfaat yang berkualitas untuk banyak tugas. Algoritma pembelajaran pohon keputusan klasik bersifat komputasi intensif. Untuk menemukan pemisahan berikutnya, perlu mengevaluasi jumlah fitur kali jumlah pengamatan untuk kondisi pemisahan yang berbeda. Hal ini menyebabkan sejumlah besar kemungkinan pemisahan untuk kumpulan data besar menggunakan input berkelanjutan dan, dalam banyak kasus, juga menyebabkan overfitting. Peningkatan memungkinkan CatBoost secara signifikan mengurangi jumlah pemisahan yang perlu dipertimbangkan. Sehingga, dapat membuat perkiraan kasar untuk fitur masukan.

CatBoost adalah versi khusus GBDT yang dengan sempurna memecahkan masalah dengan fitur yang dipesan sambil juga mendukung fitur kategorikal yang merupakan variabel yang dapat diambil salah satu dari sejumlah nilai (kategori) yang terbatas dan biasanya tetap. Fitur kategoris merupakan secara efisien adalah salah satu tantangan terbesar dalam ML. Teknik yang paling banyak digunakan untuk menangani prediktor kategori adalah *one-hot-encoding*. Fitur asli dihapus dan variabel biner baru ditambahkan untuk setiap kategori.

Cara lain untuk menangani fitur kategori pada CatBoost yang sering kali memberikan kualitas superior dibandingkan dengan *hashing* dan/atau *one-hot-encoding* adalah dengan menggunakan apa yang disebut teknik *label-encoding* yang mengubah kategori diskrit menjadi fitur numerik. *Label-encoding* juga dapat digunakan dalam sistem pembelajaran online. Oleh karena itu, pohon keputusan bekerja dengan baik dengan input numerik, CatBoost harus mengkonversi faktor kategori menjadi numerik. Sehingga fitur kategorikal diganti dengan beberapa statistik, yang dihitung dari label (CatBoost, 2018).

### 2.6.3 Matriks Korelasi

Pemilihan fitur dalam *machine learning* merupakan proses pemilihan subset fitur atau variabel yang relevan dari dataset asli untuk digunakan sebagai input model. Tujuan dari pemilihan fitur adalah untuk meningkatkan performa model, mengurangi kompleksitas model, dan meningkatkan interpretabilitas (Kuhn & Johnson, 2019).

Matriks korelasi adalah alat yang umum digunakan dalam pemilihan fitur untuk mengidentifikasi fitur yang sangat berkorelasi. Fitur yang sangat berkorelasi adalah fitur yang memiliki hubungan linier yang kuat satu sama lain, dan dapat menyebabkan *overfitting* dan menurunkan performa model. Mengidentifikasi dan menghapus fitur yang sangat berkorelasi dapat meningkatkan generalisasi model dan mengurangi risiko *overfitting* (Liu & Motoda, 2007).

### 2.6.4 Metode Evaluasi

Pada klasifikasi, kategori atau kelas data diidentifikasi berdasarkan data pelatihan. Model belajar dari *dataset* yang diberikan dan kemudian mengklasifikasikan data baru ke dalam kelas atau grup berdasarkan pelatihan. Itu memprediksi label kelas sebagai output, seperti *Ya atau Tidak, 0 atau 1, Spam atau Bukan Spam*, dll. Untuk mengevaluasi kinerja model klasifikasi, metrik yang berbeda digunakan, dan beberapa di antaranya adalah sebagai berikut:

#### 1. *Confusion Matrix*

Performa pada model bisa dilihat dari hasil analisis evaluasinya dengan memakai *confusion matrix*. *Confusion matriks* menggunakan perhitungan nilai *accuracy, recall, precision* dan *f1 score* dari sebuah model ML (Ghoneim, 2019). Pada perhitungan *confusion matriks*, terdapat empat kondisi kasus seperti yang ditunjukkan di bawah ini:

Tabel 2.4 *Confusion Matrix*

Sumber: [towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62](https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62)

		Actual	
		1	0
Predict	1	<i>True Positives (TP)</i>	<i>False Positives (FP)</i>
	0	<i>False Negatives (FN)</i>	<i>True Negatives (TN)</i>

- *True Positive (TP)*: Kondisi ketika hasil prediksi benar dan realitanya ternyata benar.
- *True Negative (TN)*: Kondisi ketika hasil prediksi benar dan realitanya memang benar salah.
- *False Positive (FP)*: Kondisi ketika hasil prediksi salah dan realitanya ternyata benar.
- *False Negative (FN)*: Kondisi ketika hasil prediksi salah dan realitanya memang salah.

## 2. Accuracy

Metrik akurasi adalah salah satu metrik klasifikasi yang paling sederhana untuk diterapkan, dan dapat ditentukan sebagai jumlah prediksi yang benar terhadap jumlah total prediksi. Metrik Akurasi digunakan pada saat kelas variabel target dalam data kira-kira seimbang. Berikut formulasi metrik akurasi:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad 2.1$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad 2.2$$

### 3. *Precision & Recall*

*Precision* dan *Recall* sangat penting untuk pengambilan informasi, di mana kelas positif lebih penting dibandingkan dengan kelas negatif. Metrik presisi digunakan untuk mengatasi keterbatasan Akurasi. *Precision* digunakan untuk menentukan proporsi prediksi positif yang sebenarnya benar. Ini dapat dihitung sebagai TP atau FP. Nilai presisi terletak antara 0 dan 1.

$$Precision = \frac{TP}{TP + FP} \quad 2.3$$

Sedangkan recall merupakan nilai atau berapa persen yang diperkirakan positif dari total positif bertujuan untuk menghitung proporsi positif aktual yang salah diidentifikasi. Ini dapat dihitung sebagai TP atau prediksi yang sebenarnya benar terhadap jumlah total positif, baik diprediksi dengan benar sebagai TP atau FN.

$$Recall = \frac{TP}{TP + FN} \quad 2.4$$

### 4. *F1 score*

*F1 score* adalah metrik untuk mengevaluasi model klasifikasi biner berdasarkan prediksi yang dibuat untuk kelas positif. Ini dihitung dengan bantuan *Precision* dan *Recall*. *F1 score* adalah jenis skor tunggal yang mewakili *Precision* dan *Recall* yang dapat dihitung sebagai rata-rata harmonik dari *Precision* dan *Recall*, dengan memberikan bobot yang sama untuk masing-masingnya. *F1 score* digunakan karena dapat bekerja dengan baik pada kumpulan data yang tidak seimbang. Rumus untuk menghitung skor F1 diberikan di bawah ini:

$$F1 Score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad 2.5$$

### 5. *Area under ROC curve (AUC)*.

Kurva AUC (*Area Under Curve*) - ROC (*Receiver Operating Characteristic*) digunakan untuk memvisualisasikan kinerja model klasifikasi ini merupakan salah satu metrik penting untuk mengevaluasi kinerja model klasifikasi. ROC

mewakili grafik untuk menunjukkan kinerja model klasifikasi pada tingkat ambang yang berbeda. Kurva diplot antara dua parameter, yaitu:

- *True Positive Rate (TPR)*
- *False Positive Rate (FPR)*

TPR adalah sinonim untuk *Recall*, karenanya dapat dihitung sebagai:

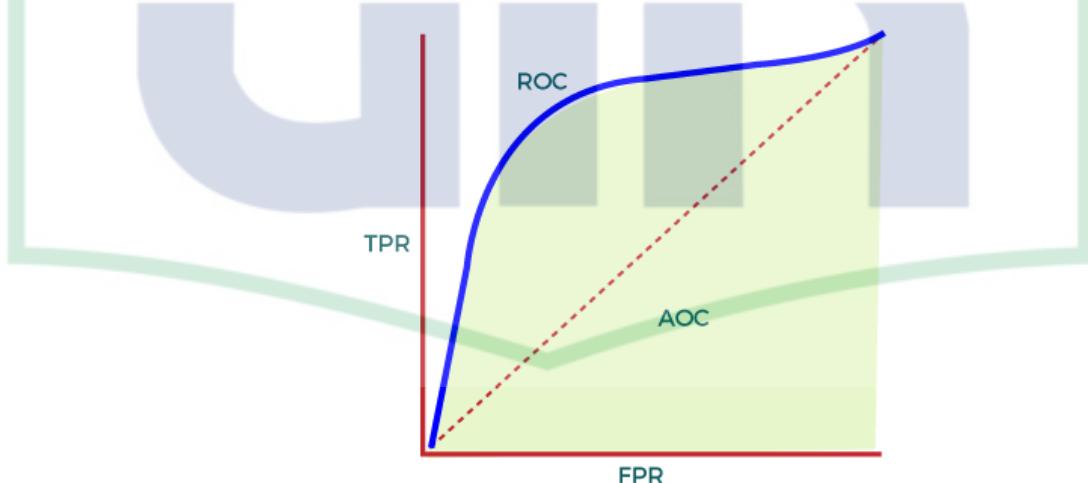
$$TPR = \frac{TP}{TP + FN} \quad 2.6$$

FPR atau Tingkat Positif Palsu dapat dihitung sebagai:

$$FPR = \frac{FP}{FP + TN} \quad 2.7$$

Untuk menghitung nilai pada setiap titik dalam kurva ROC, kita dapat mengevaluasi model regresi logistik berkali-kali dengan ambang klasifikasi yang berbeda, tetapi ini tidak akan jauh lebih efisien. Satu metode efisien digunakan dikenal sebagai AUC.

AUC: Area di bawah kurva ROC. AUC dikenal dengan area di bawah kurva ROC. Seperti namanya, AUC menghitung area dua dimensi di bawah seluruh kurva ROC, seperti yang ditunjukkan gambar di bawah ini:



Gambar 2.5 Kurva AUC – ROC

Sumber: [www.javatpoint.com/auc-roc-curve-in-machine-learning](http://www.javatpoint.com/auc-roc-curve-in-machine-learning)

AUC menghitung kinerja di semua ambang batas dan memberikan ukuran agregat. Nilai AUC berkisar antara 0 sampai 1. Artinya model dengan prediksi

100% salah akan memiliki AUC 0.0, sedangkan model dengan prediksi 100% benar akan memiliki AUC 1.0.

AUC harus digunakan untuk mengukur seberapa baik prediksi diurutkan daripada nilai absolutnya. Selain itu, ini mengukur kualitas prediksi model tanpa mempertimbangkan ambang klasifikasi.

Hasil perhitungan ditampilkan menggunakan kurva ROC atau AUC. Kurva ROC ini memiliki tingkat nilai diagnosa sebagai berikut (Gorunescu, 2011):

- a. 0.90 - 1.00 = klasifikasi sangat baik (*excellent classification*).
- b. 0.80 - 0.90 = klasifikasi baik (*good classification*).
- c. 0.70 - 0.80 = klasifikasi cukup (*fair classification*).
- d. 0.60 - 0.70 = klasifikasi buruk (*poor classification*).
- e. 0.50 - 0.60 = klasifikasi salah (*failure*).

## 2.7 Python

Python adalah bahasa pemrograman tingkat tinggi yang ditafsirkan, berorientasi objek, dengan semantik dinamis. Struktur data bawaan tingkat tinggi, dikombinasikan dengan pengetikan dinamis dan pengikatan dinamis, membuatnya sangat menarik untuk *Rapid Application Development*, serta untuk digunakan sebagai bahasa skrip atau lem untuk menghubungkan komponen yang ada bersama-sama. Sintaks Python yang sederhana dan mudah dipelajari menekankan keterbacaan dan karenanya mengurangi biaya pemeliharaan program. Python mendukung modul dan paket, yang mendorong modularitas program dan penggunaan kembali kode. Interpreter Python dan perpustakaan standar yang luas tersedia dalam bentuk sumber atau biner tanpa biaya untuk semua platform utama, dan dapat didistribusikan secara bebas.

Python memiliki perpustakaan untuk pemuatan data, visualisasi, statistik, pemrosesan bahasa alami, pemrosesan gambar, dan sebagainya. Salah satu keuntungan utama menggunakan Python adalah kemampuan untuk berinteraksi langsung dengan kode, menggunakan terminal atau alat lain seperti Jupyter Notebook. ML dan analisis data pada dasarnya merupakan sebuah proses iteratif, di mana data mendorong analisis. Pada proses ini untuk

memiliki alat yang memungkinkan iterasi cepat dan interaksi yang mudah. Sebagai bahasa pemrograman tujuan umum, Python juga memungkinkan pembuatan antarmuka pengguna grafis (GUI) dan layanan web yang kompleks, dan untuk integrasi ke dalam sistem yang ada.

### 2.7.1 Jupyter Notebook

Jupyter Notebook adalah lingkungan interaktif untuk menjalankan kode di browser, yang digunakan untuk menganalisis data eksplorasi serta banyak digunakan oleh para ilmuwan data. Jupyter Notebook juga memudahkan untuk memasukkan kode, teks, dan gambar (Müller & Guido, 2015).

## 2.8 Metodologi Penelitian Kuantitatif

Menurut (Sugiyono, 2019) metode penelitian kuantitatif merupakan metode yang didasarkan pada filosofi positivis, yang digunakan untuk mensurvei populasi atau sampel tertentu yang kemudian dilakukan pengujian asumsi yang telah ditetapkan, mengumpulkan data dengan menggunakan alat penelitian, dan menganalisis data kuantitatif.

Penelitian kuantitatif mementingkan adanya varibel-variabel sebagai objek penelitian dan harus didefinisikan dalam bentuk operasional untuk tiap variabelnya. Tujuan yang ingin dicapai adalah menguji teori, membangun fakta, menunjukkan hubungan dan pengaruh serta perbandingan antar-variabel, memberikan deskripsi statistik, menafsir, dan meramalkan hasilnya (Siregar, 2014).

Metode kuantitatif bertujuan untuk menggambarkan fenomena atau gejala sosial dengan cara menganalisi bagaimana fenomena yang terjadi di masyarakat saling berhubungan satu sama lain (Sudaryono, 2017). Analisis dilakukan dengan menggunakan statistik, tabel atau bagan dan membahas cara mereka memperlihatkan keterkaitan terhadap hipotesis (Neuman, 2003).

Menurut V. Wiratna Sujarweni (2014:39) penelitian kuantitatif adalah jenis penelitian yang menghasilkan penemuan-penemuan yang dapat dicapai (diperoleh) dengan menggunakan prosedur-prosedur statistik atau cara lain dari kuantifikasi (pengukuran). Penelitian kuantitatif nantinya menghasilkan informasi yang lebih

terukur. Hal ini karena ada data yang dijadikan landasan untuk menghasilkan informasi yang lebih terukur.

## 2.9 Populasi dan Sampel

### 2.9.1 Definisi Populasi

Populasi ialah sekelompok orang, kejadian, atau hal – hal yang menarik untuk diteliti (Sekaran dan Bougie, 2017). Menurut Sugiyono (2019:126), populasi adalah wilayah generalisasi yang terdiri atas suatu objek atau subjek yang memiliki kuantitas, karakteristik, serta kualitas tertentu yang telah ditetapkan peneliti untuk diteliti dan dilakukan penarikan kesimpulannya.

### 2.9.2 Definisi Sampel

Menurut Sekaran dan Bougie (2017), sampel merupakan bagian dari populasi dan terdiri dari banyak anggota yang dipilih sebagai sampel. Menurut Sugiyono (2019:127), sampel adalah sebagian dari populasi dengan ciri-ciri tertentu. Di mana, tidak semua anggota populasi dapat diteliti dikarenakan waktu yang terbatas, tenaga dan dana. Sehingga, peneliti menggunakan sampel yang diambil dari populasi dengan ciri-ciri sampel yang telah ditetapkan oleh peneliti. Pada prinsipnya, sampel yang dipilih untuk dipelajari dapat menggambarkan suatu generalisasi terhadap populasi. Bila sampel tidak representatif, maka peneliti tidak dapat menyimpulkan sesuai dengan kenyataan atau membuat kesimpulan yang salah.

### 2.9.3 Jenis-jenis Metode Pengambilan Sampel

#### 1. *Probability Sampling*

*Probability sampling* adalah teknik sampling yang memberikan peluang yang sama bagi setiap unsur (anggota) populasi untuk dipilih menjadi anggota sampel (Sugiyono, 2018:81). Pemilihan *probability sampling* ini sangat direkomendasikan pada penelitian

kuantitatif. *Probability sampling* memiliki 4 teknik sampling yang dapat digunakan antara lain:

1) *Simple Random Sampling*

*Simple random sampling* adalah teknik pemilihan sampel dari populasi yang dilakukan secara acak tanpa memperhatikan strata yang ada dalam populasi itu. Keuntungannya ialah anggota sampel mudah dan cepat diperoleh, sedangkan kelemahannya ialah rendahnya jaminan mengenai sampel yang terpilih dapat bersifat representatif atau dapat mewakili populasi yang dituju.

2) *Proportionate Stratified Random Sampling*

*Proportionate stratified random sampling* adalah teknik *sampling* yang digunakan apabila populasi mempunyai anggota/unsur yang tidak homogen dan berstrata secara proporsional. Peneliti harus mengetahui bahwa dalam populasi ada strata, kelas, lapisan, atau ras. Keuntungan menggunakan cara ini ialah anggota sampel yang diambil lebih representatif. Kelemahannya ialah lebih banyak memerlukan usaha pengenalan terhadap karakteristik populasi.

3) *Disproportionatestratif Random Sampling*

*Disproportionatestratif Random Sampling* adalah teknik *sampling* jika populasi berstrata atau bertingkat tetapi kurang proposisional.

4) *Area Cluster Sampling*

*Area Cluster Sampling* adalah teknik *sampling* daerah yang digunakan untuk menentukan sampel bila objek yang akan diteliti atau sumber data sangat luas, misalnya daerah, provinsi, kabupaten, kecamatan dan seterusnya. Keuntungan

menggunakan teknik ini adalah dapat mengambil populasi besar yang tersebar di daerah, pelaksanaannya lebih mudah dan murah dibandingkan teknik lainnya. Kelemahannya ialah jumlah individu dalam setiap pilihan tidak sama, ada kemungkinan penduduk satu daerah berpindah ke daerah lain tanpa sepengetahuan peneliti, sehingga penduduk tersebut mungkin menjadi anggota rangkap sampel penelitian.

## 2. *Non-Probability Sampling*

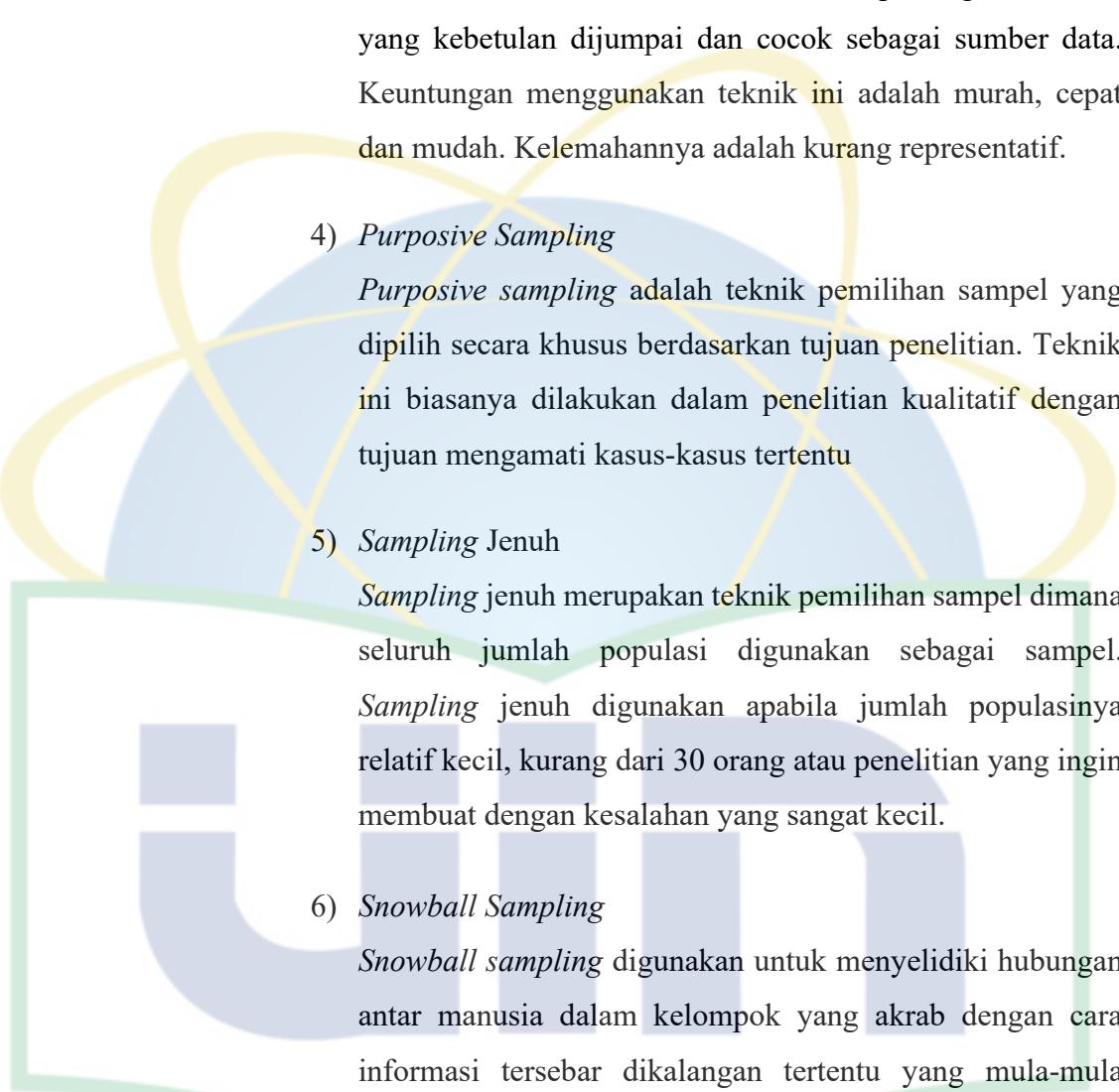
*Non probability sampling* adalah teknik yang tidak memberi peluang/kesempatan yang sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel (Sugiyono, 2018). Teknik ini dapat dilakukan dengan mudah dalam waktu yang sangat singkat. Namun kelemahan teknik ini adalah hasilnya tidak dapat diterima dan berlaku bagi seluruh populasi, karena sebagian besar dari populasi tidak dilibatkan dalam penelitian. Dalam teknik *non probability sampling* ini ada 6 macam teknik memilih sampel yaitu:

### 1) *Systematic Sampling*

*Systematic Sampling* adalah teknik pengambilan sampel yang dipilih berdasarkan urutan dari anggota populasi yang telah diberi nomor urut. Keuntungan dari teknik ini adalah mudah dalam pelaksanaannya dan cepat diselesaikan. Kelemahannya ialah populasi yang berada di antara yang kesekian dan kesekian dikecualikan, sehingga cara ini tidak sebaik *random sampling*.

### 2) *Quota Sampling*

*Quota Sampling* adalah teknik untuk menentukan sampel pada suatu tingkat yang dipilih sampai dengan jumlah tertentu (kuota) yang dibutuhkan.



### 3) *Incidental Sampling*

*Incidental Sampling* adalah teknik pemilihan sampel berdasarkan kebetulan dilakukan terhadap orang atau benda yang kebetulan dijumpai dan cocok sebagai sumber data. Keuntungan menggunakan teknik ini adalah murah, cepat dan mudah. Kelemahannya adalah kurang representatif.

### 4) *Purposive Sampling*

*Purposive sampling* adalah teknik pemilihan sampel yang dipilih secara khusus berdasarkan tujuan penelitian. Teknik ini biasanya dilakukan dalam penelitian kualitatif dengan tujuan mengamati kasus-kasus tertentu

### 5) *Sampling Jenuh*

*Sampling jenuh* merupakan teknik pemilihan sampel dimana seluruh jumlah populasi digunakan sebagai sampel. *Sampling jenuh* digunakan apabila jumlah populasinya relatif kecil, kurang dari 30 orang atau penelitian yang ingin membuat dengan kesalahan yang sangat kecil.

### 6) *Snowball Sampling*

*Snowball sampling* digunakan untuk menyelidiki hubungan antar manusia dalam kelompok yang akrab dengan cara informasi tersebar dikalangan tertentu yang mula-mula jumlahnya kecil, kemudian membesar. Keuntungan dari snowball sampling adalah adanya peningkatan kecenderungan menempatkan karakteristik- karakteristik yang diinginkan dalam populasi. Kelemahannya adalah penanganannya sulit sekali dikendalikan jika jumlah sampel melebihi 100 orang.

## 2.10 Metode Slovin

Rumus Slovin adalah sebuah rumus atau formula untuk menghitung jumlah sampel minimal apabila perilaku dari sebuah populasi tidak diketahui secara pasti. Rumus ini diperkenalkan pertama kali oleh seorang Ilmuwan Matematis yang bernama Slovin, pada tahun 1960 (Ghozali, 2011). Pengertian rumus slovin menurut Sugiyono (2017) adalah suatu rumus yang digunakan untuk mencari besaran sampel yang dinilai mampu mewakili keseluruhan populasi.

Rumus Slovin digunakan untuk sebuah penelitian pada suatu objek tertentu dalam jumlah populasi yang besar dan peneliti tidak tahu berapa proporsi populasi yang ideal untuk mewakili. Sehingga digunakanlah metode Slovin untuk meneliti pada sebuah sampel dari populasi objek yang besar tersebut.

Rumus Slovin termasuk ke dalam *simple random sampling* yang mana setiap individu mempunyai kesempatan yang sama untuk dijadikan sampel. Tujuan utama dari rumus slovin ini adalah untuk mencari estimasi jumlah populasi. Estimasi di sini maksudnya adalah proporsi populasi, bukan rata-rata dari populasi. Nantinya, nilai estimasi itu akan dibuat dalam bentuk persentase.

Dalam penggunaan rumus Slovin ini, hal yang pertama kali harus kita lakukan, yaitu menetapkan Taraf Keyakinan atau *Confidence Level* (...%) terhadap hasil kebenaran, atau Taraf Signifikansi Toleransi Kesalahan (0,...) yang akan terjadi. Taraf keyakinan/*Confidence Level* 95% akan kebenaran hasil (maksudnya yaitu yakin bahwa penelitian yang kalian lakukan 95% benar) dan Taraf Signifikansi 0,05 (memastikan hanya 5% saja kesalahan yang akan terjadi) (Nalendra, et al., 2021).

### 2.10.1 Rumus Metode Slovin

Berikut merupakan rumus Slovin yang dapat digunakan untuk menentukan minimal sampel yang dibutuhkan berdasarkan jumlah populasi yang dipilih:

$$n = \frac{N}{1 + Ne^2} \quad 2.8$$

Keterangan:

n= sample minimum

N=sampel populasi

e= presentase batas toleransi

Rumus Slovin ditentukan oleh nilai tingkat kesalahan, semakin besar batas kesalahan yang digunakan semakin sedikit jumlah sampel yang diambil (Riyanto & Hatmawan, 2020). Ukuran batas toleransi atau batas kesalahan yang dapat digunakan pada ukuran populasi dapat dilihat seperti pada gambar berikut:

Populasi	Batas-batas Kesalahan					
	± 1%	± 2%	± 3%	± 4%	± 5%	± 10%
500	*	*	*	*	222	83
1500	*	*	638	441	316	94
2500	*	1250	769	500	345	96
3000	*	1364	811	517	353	97
4000	*	1538	870	541	364	98
5000	*	1667	909	556	370	98
6000	*	1765	938	566	375	98
7000	*	1842	959	574	378	99
8000	*	1905	976	580	381	99
9000	*	1957	989	584	383	99
10000	5000	2000	1000	588	385	99
50000	8333	2381	1087	617	387	100

Gambar 2.6 Ukuran Sampel untuk Batas-batas Kesalahan dan Jumlah

Populasi yang ditetapkan

Sumber: Umar, 2002

## 2.11 Penelitian Sejenis

Tabel 2.5 Penelitian Sejenis

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
1.	Pengaruh Gaya Hidup Terhadap Willingness to Subscribe Pada Aplikasi Streaming Film Online Netflix di Kalangan Generasi Z Kota Surabaya	Ahmad Nanang Hawari	2023	Menganalisis pengaruh gaya hidup terhadap willingness to subscribe aplikasi Netflix di kalangan generasi Z yang merupakan pengguna media baru terbanyak di Kota Surabaya.	Analisis regresi	Semakin banyak generasi Z yang memiliki gaya hidup digital dengan menggunakan gadget yang mereka miliki untuk mempermudah kegiatan sehari-hari, maka willingness to subscribe pada sebuah produk digital seperti Netflix akan semakin tinggi dan disebabkan karena semakin banyak generasi Z di kota Surabaya yang meninggalkan cara lama menonton film dengan pergi ke bioskop dan lebih memilih untuk beralih pada menonton film

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
						melalui gadget mereka dengan aplikasi Netflix
2.	Pengaruh Gaya Hidup Hedonis terhadap Perilaku Konsumtif di Kalangan Mahasiswa	Mira Nurazijah <sup>1</sup> , Syaipia Lailla Nur Fitriani <sup>2</sup> , Tin Rustini <sup>3</sup>	2023	Untuk mengetahui bagaimana tingkat gaya hedonis yang menyebabkan perilaku konsumtif.	Mixed method antara pedekatan kuantitatif dan kualitatif	Responden selalu ingin mendapatkan apa yang diinginkan, tetapi bukan berarti suka mengikuti gaya yang sedang trend. Kemudian adanya suatu produk baru yang diiklankan, responden mengaku tidak mengikuti dan harus memberi hal tersebut. Kegiatan berbelanja dan bermain dalam satu minggu juga, responden masih dalam batas yang wajar dan tidak berlebihan.
3.	Pengaruh Gengsi, Diskon, Dan Literasi Keuangan	Khairani Marpaung, Tri Inda Fadhila Rahma	2023	Mengetahui pengaruh gengsi terhadap perilaku konsumeris me generasi Z pada	Path coefficients dan	Gengsi, diskon, dan literasi keuangan berpengaruh terhadap perilaku konsumerisme generasi Z di kalangan mahasiswa

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	Terhadap Perilaku Konsumerisme Generasi Z Di Kalangan Mahasiswa Manajemen			mahasiswa manajemen,untuk mengetahui pengaruh diskon terhadap perilaku konsumerisme generasi Z pada mahasiswa manajemen, untuk mengetahui pengaruh literasi keuangan terhadap perilaku konsumerisme generasi Z di kalangan mahasiswa manajemen.	bootstrapping smartPLS 3	manajemen. Maka dari itu gengsi, diskon, dan literasi keuangan dapat meningkatkan perilaku konsumerisme generasi Z di kalangan mahasiswa manajemen.
4.	A Comparative Study of	Annisarahmi Nur Aini	2023	Mempelajari klasifikasi multi-kelas	CatBoost dan DRF	CatBoost dan DRF dapat dengan tepat memecahkan masalah

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	CatBoost and Double Random Forest for Multi - class Classification	Aldania, Agus Mohamad Soleh, Khairil Anwar Notodiputro		menggunakan CatBoost, sebuah metode yang dikembangkan dengan gradient boosting dan double random forest (DRF), pengembangan RF yang baik digunakan ketika model RF yang dihasilkan underfitting.		klasifikasi multi-kelas pada jarak jauh, diukur dengan skor akurasi seimbang 100%. Pada jarak menengah, CatBoost dan DRF menghasilkan skor akurasi seimbang masing-masing sebesar 99,25% dan 97,54%, sedangkan pada jarak rendah 32,37% dan 23,97%. Dalam studi empiris, performa CatBoost mengungguli DRF sebesar 4,27%.
5.	Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2	Asier Uribeetxebarria, Ander Castellón and Ana Aizpurua	2023	Memperkirakan hasil gandum secara akurat untuk pengambilan keputusan yang tepat dalam pertanian presisi (PA) dan	CatBoost Algorithm	Model yang dikembangkan untuk memperkirakan hasil menggunakan informasi dari satelit S1 dan S2 menunjukkan hasil yang lebih baik daripada analisis korelasi. Di antara model

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	with CatBoost Algorithm			meningkatkan pengelolaan tanaman.		yang dievaluasi, CatBoost, yang masih relatif kurang dimanfaatkan di bidang pertanian, memberikan hasil terbaik. Menggabungkan gambar dari S1 dan S2 secara substansial meningkatkan prediksi, memberikan tingkat presisi yang cukup untuk mempertimbangkan peta hasil untuk penyesuaian pemupukan.
6.	Faktor-Faktor yang Mempengaruhi Perilaku Konsumtif Mahasiswa Pendidikan Ekonomi	Yeni Yanti Harita, Gusnardi, M. Yogi Riyantama Isjoni	2022	Mengetahui pengaruh media sosial dan gaya hidup terhadap perilaku konsumtif mahasiswa Pendidikan Ekonomi Universitas Riau.	Teknik analisis data menggunakan analisis regresi linear berganda	Media sosial dan gaya hidup menentukan seseorang dalam berperilaku konsumtif. Semakin tinggi tingkat penggunaan media sosial dan tingkat gaya hidup seseorang maka akan semakin tinggi pula tingkta perilaku konsumtifnya. Hal ini dikarenakan

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
						seseorang yang aktif menggunakan media sosial cenderung meningkatkan gaya hidup yang mewah mengikuti orang kebanyakan.
7.	Using Boosted Machine Learning to Predict Suicidal Ideation by Socioeconomic Status among Adolescents	Hwanjin Park dan Kounseok Lee	2022	Menggunakan teknik <i>machine learning</i> untuk mengidentifikasi faktor risiko ide bunuh diri di kalangan remaja dan memahami hubungan antara faktor risiko ini	Synthetic minority over-sampling technique dan XGBoost	Pada kelompok SES tinggi, prestasi akademik merupakan faktor risiko utama untuk ide bunuh diri, sedangkan pada kelompok SES rendah, hanya faktor emosional seperti stres dan kecemasan yang berkontribusi secara signifikan terhadap ide bunuh diri. SES berperan penting dalam kesehatan mental remaja.

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
				dan status sosial ekonomi (SES)		Perbaikan SES pada masa remaja dapat mengatasi emosi negatif mereka dan mengurangi risiko bunuh diri.
8.	CatBoost — An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance	Abhisht Joshi, Pranay Saggar, Rajat Jain, Moolchand Sharma, Deepak Gupta and Ashish Khanna	2021	pemeriksaan model untuk menentukan apa sebenarnya fitur utama yang berkontribusi dan berdampak signifikan pada kinerja siswa dan kemungkinan menentukan keberhasilan siswa.	<i>Ensemble machine learning model</i>	CatBoost— <i>ensemble machine learning model</i> — lebih unggul daripada model <i>machine learning</i> dengan akurasi 92,27%. CatBoost mampu menunjukkan modelnya dapat diandalkan dengan penggunaan optimasi smote dan hyperparameter, yang terbukti menjadi metode dan pendekatan yang berharga. Serta fitur tambahan yang juga signifikan.
9.	Md-Pred: A Multidimensional Hybrid	Xinyuan Tian, Bingqin Pan,	2021	Prediksi pembatalan pesanan hotel untuk mengoptimalkan	Evaluasi multidimensi	Model evaluasi multi-perspektif menangkap perbedaan dan koneksi antar fitur dengan lebih

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	Prediction Model Based on Machine Learning for Hotel Booking Cancellation Prediction	Liping Bai and Deyun Mo		keakuratan prediksi dalam meningkatkan nilai analisis bisnis selanjutnya dan optimisasi operasional.	hybrid Md-Pred	baik, yang memungkinkan penggunaan informasi dan data yang lebih besar. Pada saat yang sama, logika klasifikasi yang jelas dapat memberi manajer dasar untuk penilaian dan dukungan keputusan, menjadikan tingkat akurasi bukan hanya nilai numerik, tetapi alat yang ampuh untuk memecahkan masalah industri.
10.	Research Application of Ensemble Machine Learning Methods to the Problem of	Denis Parfenov, Larisa Kuznetsova, Natalia Yanishevskaya, Irina Bolodurina,	2021	Mengidentifikasi serangan DDoS sedang diselesaikan. Dengan sebuah solusi yang diusulkan berdasarkan perluasan serangkaian	Gradient Boosting, AdaBoost, CatBoost	Dalam kasus klasifikasi biner, pengklasifikasi terbaik dari yang dipertimbangkan adalah CatBoost dengan akurasi 99,3%, yang rata-rata 0,3% lebih tinggi dari algoritma yang ada. Pada klasifikasi multikelas, algoritma

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	Multiclass Classification of DDoS Attacks Identification	Arthur Zhigalov, Leonid Legashev		fitur yang secara tradisional digunakan untuk mengidentifikasi serangan pada jaringan menggunakan algoritma hashing khusus untuk masing-masing blok file konfigurasi perangkat dalam jaringan perangkat yang dipertimbangkan.		CatBoost pada kumpulan fitur yang menggunakan hashing data dari perangkat jaringan juga menunjukkan kinerja terbaik, dengan tingkat akurasi 97%, setidaknya 3,9% lebih baik dari pengklasifikasi serupa.
11.	Developing a New Hourly Forest Fire Risk Index Based on	Yoojin Kang, Eunna Jang, Jungho Im, Chungeun	2020	Untuk mengetahui risiko kebakaran hutan suatu wilayah, untuk mempersiapkan	Catboost	faktor aksesibilitas berperan sangat penting dalam pemodelan risiko kebakaran hutan dimana sebagian besar kebakaran hutan

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	Catboost in South Korea	Kwon, and Sungyong Kim		dan menanggapi peristiwa bencana secara tepat dengan mengembangkan <i>hourly forest fire risk index</i> (HFRI) per jam dengan resolusi spasial 1 km menggunakan faktor aksesibilitas, bahan bakar, waktu, dan cuaca berdasarkan <i>machine learning</i> Catboost di Korea Selatan		disebabkan oleh faktor antropogenik.
12.	Analysis and classification of heart rate using	B. Dhananjay, J. Sivaraman	2020	Mengembangkan model <i>machine learning</i> untuk	Catboost	Akurasi, sensitivitas, presisi, dan nilai F1 model Catboost lebih baik dibandingkan model ET dan RC.

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	CatBoost feature ranking model			mengklasifikasikan kondisi Sinus Rhythm (SR), Sinus Tachycardia (ST), dan Atrial Tachycardia (AT) dari ECG.		Berdasarkan kepentingan fitur menyimpulkan bahwa gelombang P ( $\mu$ V), PRI (ms), dan Atrial rate (PPI (ms)) dapat secara akurat mengklasifikasikan sinyal SR, ST, dan AT ECG. Waktu komputasi model <i>machine learning</i> berbasis Catboost minimal karena menggunakan sistem inferensi berbasis pohon simetris. Algoritma penguat yang ada di pengklasifikasi Catboost dapat meminimalkan masalah pemasangan berlebih.
13.	Discrimination of Excessive Exhaust Emissions of	Xihong Fei, Yi Fang, Qiang Ling	2020	Untuk membedakan emisi gas buang kendaraan yang berlebihan,	Catboost	Model diskriminasi berdasarkan algoritma CatBoost lebih unggul daripada algoritma klasifikasi pembelajaran mesin tradisional,

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
	Vehicles based on Catboost Algorithm			mendeteksi secara akurat konsentrasi emisi gas buang CO dan HC emisi berlebih melalui area pengukuran.		yang dapat mencapai akurasi diskriminasi yang lebih tinggi dimana algoritma CatBoost memiliki metode yang paling akurat untuk membedakan emisi gas buang kendaraan yang berlebihan
14.	On linking socioeconomic status to consumer loyalty behaviour	Christopher J. White, Eudora Tong	2019	Mengeksplorasi hubungan antara SES dan loyalitas dan setelah mengadaptasi urutan motivasi dengan mengetahui bagaimana perilaku loyalitas terhadap seni pertunjukan terbentuk.	Structural Equation Modeling (SEM) dalam AMOS 24	Kepuasan kebutuhan psikologis dasar atau <i>basic psychological need</i> (BPN) memainkan peran penting dalam mentransfer efek SES ke loyalitas untuk demografi penting ini melalui sikap utilitarian dan hedonis.

No.	Judul	Nama Penulis	Tahun	Tujuan	Metode	Hasil
15.	Pengaruh Status Sosial Ekonomi Orang Tua Dan Kelompok Teman Sebaya Terhadap Perilaku Konsumsi Siswa	Yayah Bahjatussaniah, Nuraini, dan Achmadi	2015	Untuk mengetahui pengaruh status sosial ekonomi orang tua terhadap perilaku konsumsi siswa, pengaruh kelompok teman sebaya terhadap perilaku konsumsi siswa, dan pengaruh status sosial ekonomi orang tua dan kelompok teman sebaya secara bersama-sama terhadap perilaku konsumsi siswa	explanatory research	Status sosial ekonomi orang tua siswa dan kelompok teman sebaya berpengaruh. Apabila status sosial ekonomi orang tua dan kelompok teman sebaya meningkat, maka perilaku konsumsi siswa juga akan meningkat.



## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Pendekatan Penelitian**

Pada penelitian ini menggunakan metode penelitian kuantitatif. Penelitian kuantitatif adalah suatu metode penelitian yang menggambarkan fenomena sesuai fakta yang ada, yang berlangsung pada saat ini atau pada sebelumnya dan juga mencoba menganalisis kebenarannya melalui data yang diperoleh. Model analisis menggunakan CatBoost dan dimuat dalam *performance metrics* yang terdiri dari 5 bagian yaitu *confusion matrix, accuracy, precision & recall, f1 score, area under ROC curva (AUC)*. Pengumpulan informasi yang dilakukan adalah dengan data SES mahasiswa UIN Jakarta yang diperoleh melalui penyebaran kuesioner secara *online*.

#### **3.2 Populasi dan Sampel Penelitian**

Sampel penelitian merupakan bagian dari populasi yang diambil sebagai sumber data dan mewakili seluruh populasi. Pada penelitian ini sampel akan diambil dari total populasi mahasiswa UIN Jakarta tahun angkatan 2019-2022 yang masuk kedalam kategori umur gen Z dengan jumlah mahasiswa/i nya sebanyak 25.607 orang yang diperoleh dari Pusat Teknologi Informasi dan Pangkalan Data (Pustipanda) UIN Jakarta. Sampel yang digunakan pada penelitian ini yaitu dengan menggunakan metode *proportionate stratified random sampling* dan besarnya sampel dalam penelitian ini ditentukan menggunakan rumus slovin dengan batas toleransi kesalahan 5% sebagai berikut:

$$n = \frac{N}{1 + Ne^2} \quad 3.1$$

Keterangan:

n= Jumlah sampel

N= Jumlah populasi

e= Batas toleransi kesalahan

Dengan rumus Slovin tersebut maka jumlah sampel dalam penelitian ini sebagai berikut:

$$n = \frac{25607}{1 + 25067 \times 0,05^2}$$

$$n = \frac{25607}{1 + 25067 \times 0,05^2}$$

$$n = \frac{25607}{1 + 64,0175}$$

$$n = \frac{25607}{65,0175}$$

$$n = 393,8478 \approx 394 \approx 400$$

Maka, berdasar perhitungan dengan menggunakan metode Slovin diatas, didapatkan sampel yang menjadi responden pada penelitian sebanyak 394 orang yang dibulatkan oleh peneliti menjadi 400 orang mahasiswa dari aktif UIN Jakarta tahun angkatan 2019-2022.

### 3.3 Metode Pengumpulan Data

#### 3.3.1 Studi Pustaka

Studi pustaka merupakan salah satu cara yang dilakukan dalam penelitian ini untuk menghimpun informasi yang relevan dengan *machine learning*, metodologi penelitian, SES, gaya hidup, serta CatBoost. Informasi terkait diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, buku tahunan, ensiklopedia, dan sumber-sumber tertulis baik tercetak maupun elektronik lainnya. Metode studi pustaka ini digunakan sebagai bahan tambahan untuk membantu mengarahkan konsep penulis dalam penyusunan skripsi prediksi gaya hidup berdasarkan SES.

#### 3.3.2 Kuesioner

Pengumpulan data akan dilakukan dengan cara mengirim pertanyaan untuk diisi sendiri oleh responden yaitu menyebar kuesioner yang berisi pertanyaan-pertanyaan meliputi pengeluaran rutin bulanan mahasiswa UIN Jakarta. Kuesioner digunakan untuk mendapatkan informasi yang dibutuhkan dan

juga mendukung penelitian. Kuesioner yang digunakan dalam penelitian ini adalah kuesioner berdasarkan klasifikasi SES mahasiswa UIN Jakarta. Menentukan klasifikasi SES akan diukur berdasarkan variabel dan skor pada setiap rentang biaya pengeluaran setiap individu. Kemudian skor tersebut dijadikan sebagai target untuk menyusun item-item instrumen yang dapat berupa pernyataan atau pertanyaan.

Pertanyaan turunan dari biaya pengeluaran rutin bulanan merupakan pertanyaan yang didapatkan dari daftar pertanyaan biaya pengeluaran rutin yang oleh Indonesia Data 2022 antara lain makanan dan minuman, pakaian, kosmetik, *toiletries, personal care*, obat-obatan, transportasi, listrik, air, bahan bakar untuk masak, biaya bulanan sekolah, biaya produksi bulanan, dan pengeluaran rutin lainnya. Selain itu pertanyaan tersebut juga berdasarkan kuesioner penelitian sejenis terdahulu yang terdiri dari biaya pengeluaran konsumsi makanan, tembakau, dan non-makanan serta rentang biaya yang digunakan merupakan rata-rata biaya yang dihasilkan dari hasil kuesioner yaitu biaya pengeluaran konsumsi mahasiswa (Tama, 2014).

### 3.4 Instrumen Penelitian

Penulis membuat kuesioner dengan mengacu pada variabel SES yang digunakan dalam penelitian ini. Kuesioner penelitian terdiri atas 10 pertanyaan mengenai profil responden yang nantinya menghasilkan data demografis mengenai profil responden. Kuesioner penelitian ini memiliki 3 pertanyaan variabel inti SES yang setiap variabelnya memiliki skor sesuai kondisi responden dan dijumlah untuk mengetahui kategori SES setiap responden. Variabel jumlah anggota keluarga tidak ditanyakan karena seluruh responden dihitung 1 anggota yaitu mahasiswa atau responden itu sendiri. Variabel biaya pengeluaran rutin bulanan memiliki 11 pertanyaan turunan serta 13 pertanyaan pendukung. Maka, total pertanyaan pada kuesioner penelitian tersebut adalah 37 pertanyaan sebagai berikut.

Tabel 3.1 Pertanyaan Kuesioner

No	Pertanyaan	Jawaban
<b>Profil Responden</b>		
1.	Nomor Telepon	(Jawaban Singkat)
2.	Kewarganegaraan	<ul style="list-style-type: none"> <li>- Laki-laki</li> <li>- Perempuan</li> </ul>
3.	Apakah Anda menerima beasiswa?	<ul style="list-style-type: none"> <li>- Ya</li> <li>- Tidak</li> </ul>
4.	Fakultas	<ul style="list-style-type: none"> <li>- Ilmu Tarbiyah dan Keguruan</li> <li>- Adab dan Humaniora</li> <li>- Ushuluddin</li> <li>- Syariah dan Hukum</li> <li>- Ilmu Dakwah dan Ilmu Komunikasi</li> <li>- Dirasat Islamiyah</li> <li>- Psikologi</li> <li>- Ekonomi dan Bisnis</li> <li>- Sains dan Teknologi</li> <li>- Ilmu Kesehatan</li> <li>- Ilmu Sosial dan Ilmu Politik</li> <li>- Kedokteran</li> </ul>
5.	Jurusan	<ul style="list-style-type: none"> <li>- Pendidikan Agama Islam</li> <li>- Pendidikan Bahasa Arab</li> <li>- Pendidikan Bahasa Inggris</li> <li>- Pendidikan IPS</li> <li>- Pendidikan Matematika</li> <li>- Pendidikan Biologi</li> <li>- Pendidikan Fisika</li> <li>- Pendidikan Kimia</li> <li>- Manajemen Pendidikan</li> <li>- Pendidikan Bahasa dan Sastra Indonesia</li> <li>- Pendidikan Guru MI/SD</li> <li>- Pendidikan Islam Anak Usia Dini (PIAUD)</li> <li>- Bahasa dan Sastra Arab</li> <li>- Sastra Inggris</li> <li>- Sejarah dan Kebudayaan Islam</li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>- Tarjamah (Bahasa Arab)</li> <li>- Ilmu Perpustakaan</li> <li>- Studi Agama Agama</li> <li>- Ilmu Al-Quran dan Tafsir</li> <li>- Ilmu Hadis</li> <li>- Aqidah dan Filsafat Islam</li> <li>- Ilmu Tasawuf</li> <li>- Perbandingan Mazhab</li> <li>- Hukum Keluarga Islam (Akhwal Syakhsiyah)</li> <li>- Hukum Tata Negara (Siyasah)</li> <li>- Hukum Pidana Islam (Jinayah)</li> <li>- Hukum Ekonomi Syariah (Muamalat)</li> <li>- Ilmu Hukum</li> <li>- Komunikasi dan Penyiaran Islam</li> <li>- Bimbingan Penyuluhan Islam</li> <li>- Manajemen Dakwah</li> <li>- Pengembangan Masyarakat Islam Kesejahteraan Sosial</li> <li>- Jurnalistik</li> <li>- Dirasat Islamiyah</li> <li>- Psikologi</li> <li>- Manajemen</li> <li>- Akuntansi</li> <li>- Ekonomi Pembangunan</li> <li>- Perbankan Syariah</li> <li>- Ekonomi Syariah</li> <li>- Teknik Informatika</li> <li>- Agribisnis</li> <li>- Sistem Informasi</li> <li>- Matematika</li> <li>- Biologi</li> <li>- Kimia</li> <li>- Fisika</li> <li>- Teknik Pertambangan</li> <li>- Kesehatan Masyarakat</li> <li>- Farmasi</li> <li>- Ilmu Keperawatan</li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>- Sosiologi</li> <li>- Ilmu Politik</li> <li>- Ilmu Hubungan Internasional</li> <li>- Kedokteran</li> </ul>
6.	Tahun Angkatan	<ul style="list-style-type: none"> <li>- 2019</li> <li>- 2020</li> <li>- 2021</li> <li>- 2022</li> </ul>
7.	Golongan UKT	<ul style="list-style-type: none"> <li>- Golongan 1</li> <li>- Golongan 2</li> <li>- Golongan 3</li> <li>- Golongan 4</li> <li>- Golongan 5</li> <li>- Golongan 6</li> <li>- Golongan 7</li> </ul>
8.	Dimanakah tempat tinggal yang ditempati selama kuliah <i>offline</i> ?	<ul style="list-style-type: none"> <li>- Rumah</li> <li>- Kost</li> <li>- Apartemen</li> <li>Lainnya:</li> </ul>
9.	Darimanakah sumber keuangan yang Anda dapatkan untuk kebutuhan sehari-hari?	<ul style="list-style-type: none"> <li>- Orang Tua</li> <li>- Bekerja</li> <li>- Magang</li> <li>- Memiliki Usaha</li> </ul>
10.	Pekerjaan Orang Tua	<ul style="list-style-type: none"> <li>- PNS</li> <li>- Karyawan Swasta</li> <li>- Wiraswasta</li> <li>- POLRI/TNI</li> <li>- Lain-lain</li> </ul>
<b>Biaya Pengeluaran Rutin Bulanan</b>		
11.	Berapakah biaya pengeluaran Anda setiap Bulan?	<ul style="list-style-type: none"> <li>- <math>\geq</math> Rp7.500.000</li> <li>- Rp5.000.001 - Rp7.500.000</li> <li>- Rp3.000.000 - Rp5.000.000</li> <li>- Rp2.000.001 - Rp3.000.000</li> <li>- Rp1.500.001 - Rp2.000.000</li> <li>- Rp1.000.001 - Rp1.500.000</li> <li>- <math>\leq</math> Rp1.000.000</li> </ul>

<b>No</b>	<b>Pertanyaan</b>	<b>Jawaban</b>
12.	Berapakah pengeluaran pulsa untuk kebutuhan kuota internet Anda setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}500.001</math></li> <li>- <math>\text{Rp}250.001 - \text{Rp}500.000</math></li> <li>- <math>\text{Rp}150.001 - \text{Rp}250.000</math></li> <li>- <math>\text{Rp}100.001 - \text{Rp}150.000</math></li> <li>- <math>\leq \text{Rp}100.000</math></li> </ul>
13.	Berapakah biaya pengeluaran Anda untuk makanan dan minuman setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}1.500.001</math></li> <li>- <math>\text{Rp}1.000.001 - \text{Rp}1.500.000</math></li> <li>- <math>\text{Rp}800.001 - \text{Rp}1.000.000</math></li> <li>- <math>\text{Rp}500.001 - \text{Rp}800.000</math></li> <li>- <math>\leq \text{Rp}500.000</math></li> </ul>
14.	Berapakah pengeluaran biaya transportasi Anda setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}500.001</math></li> <li>- <math>\text{Rp}250.001 - \text{Rp}500.000</math></li> <li>- <math>\text{Rp}150.001 - \text{Rp}250.000</math></li> <li>- <math>\text{Rp}100.001 - \text{Rp}150.000</math></li> <li>- <math>\leq \text{Rp}100.000</math></li> </ul>
15.	Berapakah biaya pengeluaran Anda untuk kebutuhan pribadi setiap bulan? (Skincare, kosmetik, perawatan, dll)	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}500.001</math></li> <li>- <math>\text{Rp}250.001 - \text{Rp}500.000</math></li> <li>- <math>\text{Rp}150.001 - \text{Rp}250.000</math></li> <li>- <math>\text{Rp}100.001 - \text{Rp}150.000</math></li> <li>- <math>\leq \text{Rp}100.000</math></li> </ul>
16.	Berapakah biaya pengeluaran Anda untuk membeli perlengkapan mandi setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}500.001</math></li> <li>- <math>\text{Rp}250.001 - \text{Rp}500.000</math></li> <li>- <math>\text{Rp}150.001 - \text{Rp}250.000</math></li> <li>- <math>\text{Rp}100.001 - \text{Rp}150.000</math></li> <li>- <math>\leq \text{Rp}100.000</math></li> </ul>
17.	Berapakah biaya yang Anda keluarkan untuk kegiatan organisasi setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}100.001</math></li> <li>- <math>\text{Rp}70.001 - \text{Rp}. 100.000</math></li> <li>- <math>\text{Rp}50.001 - \text{Rp}. 70.000</math></li> <li>- <math>\text{Rp}30.001 - \text{Rp}50.000</math></li> <li>- <math>\leq \text{Rp}30.000</math></li> </ul>
18.	Berapakah biaya yang Anda keluarkan untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}100.001</math></li> <li>- <math>\text{Rp}70.001 - \text{Rp}. 100.000</math></li> <li>- <math>\text{Rp}50.001 - \text{Rp}. 70.000</math></li> <li>- <math>\text{Rp}30.001 - \text{Rp}50.000</math></li> <li>- <math>\leq \text{Rp}30.000</math></li> </ul>
19.	Berapakah biaya yang Anda keluarkan untuk berlangganan platform hiburan berbayar setiap bulan?	<ul style="list-style-type: none"> <li>- <math>\geq \text{Rp}500.001</math></li> <li>- <math>\text{Rp}250.001 - \text{Rp}500.000</math></li> <li>- <math>\text{Rp}150.001 - \text{Rp}250.000</math></li> <li>- <math>\text{Rp}100.001 - \text{Rp}150.000</math></li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>- ≤ Rp100.000</li> <li>- Tidak Berlangganan</li> </ul>
20.	Berapakah biaya yang Anda keluarkan untuk hiburan setiap bulan?	<ul style="list-style-type: none"> <li>- ≥ Rp500.001</li> <li>- Rp250.001 - Rp500.000</li> <li>- Rp150.001 - Rp250.000</li> <li>- Rp100.001 - Rp150.000</li> <li>- ≤ Rp100.000</li> </ul>
21.	Berapakah biaya pengeluaran konsumsi Tembakau, Pods, ataupun Vape Anda setiap Bulan?	<ul style="list-style-type: none"> <li>- ≥ Rp500.001</li> <li>- Rp250.001 - Rp500.000</li> <li>- Rp150.001 - Rp250.000</li> <li>- Rp100.001 - Rp150.000</li> <li>- ≤ Rp100.000</li> </ul>
<b>Sumber Air Minum</b>		
22.	Apakah sumber air minum yang anda konsumsi?	<ul style="list-style-type: none"> <li>- Air minum dalam kemasan botol/galon bermerek</li> <li>- Air minum dalam galon isi ulang</li> <li>- Air PAM</li> <li>- Air keran non-PAM</li> <li>- Air sumur bor/pompa</li> <li>- Air sumur tampung</li> <li>- Air sumur yang tidak ditampung</li> <li>- Mata air yang ditampung</li> <li>- Mata air yang tidak ditampung</li> <li>- Air sungai</li> <li>- Air hujan</li> </ul>
<b>Sumber Bahan Bakar</b>		
23.	Apakah sumber bahan bakar yang Anda gunakan saat memasak?	<ul style="list-style-type: none"> <li>- Kompor listrik</li> <li>- Gas Elpiji – 12 Kg/ Gas 7 Kg/Gas 5 Kg</li> <li>- Gas 3 Kg</li> <li>- Gas kota/Alam</li> <li>- Minyak tanah</li> <li>- Kayu</li> <li>- Arang</li> <li>- Bricket</li> <li>- Tidak Memasak</li> </ul>
<b>Pertanyaan Pendukung</b>		

No	Pertanyaan	Jawaban
24.	Apakah operator seluler yang anda gunakan?	<ul style="list-style-type: none"> <li>- Telkomsel</li> <li>- Simpati</li> <li>- Kartu As</li> <li>- by.U</li> <li>- IM3</li> <li>- 3</li> <li>- XL</li> <li>- Axis</li> <li>- Smartfren</li> </ul>
25.	Jenis pembayaran apakah yang digunakan pada operator seluler Anda?	<ul style="list-style-type: none"> <li>- Prabayar</li> <li>- Pascabayar</li> </ul>
26.	Apakah jenis transportasi apa yang Anda gunakan sehari-hari?	<ul style="list-style-type: none"> <li>- Motor</li> <li>- Mobil</li> <li>- Kereta</li> <li>- Transjakarta</li> <li>- Transportasi Online</li> <li>- Berjalan Kaki</li> </ul>
27.	Jika Anda ingin makan, cara apa yang sering Anda lakukan untuk mendapatkan makanan tersebut?	<ul style="list-style-type: none"> <li>- Membeli secara langsung</li> <li>- Menggunakan layanan pesan antar</li> <li>- Memasak</li> </ul>
28.	Jika Anda menggunakan layanan pesan antar, platform layanan pesan antar apa yang paling sering Anda gunakan?	<ul style="list-style-type: none"> <li>- Grabfood</li> <li>- GoFood</li> <li>- ShopeeFood</li> </ul>
29.	Apakah E-wallet yang sering Anda gunakan?	<ul style="list-style-type: none"> <li>- OVO</li> <li>- Gopay</li> <li>- Dana</li> <li>- Lainnya:</li> </ul>
30.	Seberapa sering Anda menggunakan platform layanan pesan antar makanan dalam sebulan ?	<ul style="list-style-type: none"> <li>- 1-5 kali</li> <li>- 6-10 kali</li> <li>- &gt; 10 kali</li> </ul>
31.	Apakah Anda berbelanja kebutuhan bulanan secara langsung dengan mengunjungi minimarket atau supermarket terdekat?	<ul style="list-style-type: none"> <li>- Ya</li> <li>- Tidak</li> </ul>

No	Pertanyaan	Jawaban
32.	Jika berbelanja secara online, manakah platform <i>e-commerce</i> yang paling sering Anda gunakan?	<ul style="list-style-type: none"> <li>- Shopee</li> <li>- Tokopedia</li> <li>- Lazada</li> <li>- Blibli</li> <li>- TikTok Shop</li> <li>- Lainnya:</li> </ul>
33.	Apakah Anda mengikuti organisasi?	<ul style="list-style-type: none"> <li>- Ya</li> <li>- Tidak</li> </ul>
34.	Apakah Anda berlangganan platform hiburan berbayar?	<ul style="list-style-type: none"> <li>- Ya</li> <li>- Tidak</li> </ul>
35.	Apa saja platform berbayar yang Anda ikuti?	<ul style="list-style-type: none"> <li>- Netflix</li> <li>- Viu</li> <li>- Disney+ Hotstar</li> <li>- Wetv</li> <li>- Vidio</li> <li>- Spotify</li> <li>- Apple Music</li> <li>- Tidak Berlangganan</li> </ul>
36.	Jenis hiburan apa yang anda lakukan setiap bulan?	<ul style="list-style-type: none"> <li>- Olahraga</li> <li>- Nonton Bioskop</li> <li>- Nongkrong</li> <li>- Lainnya:</li> </ul>
37.	Apakah Anda mengkonsumsi Tembakau atau Rokok Elektrik (Pod atau Vape) ?	<ul style="list-style-type: none"> <li>- Ya</li> <li>- Tidak</li> </ul>

### 3.5 Pengolahan Data

Proses pengolahan data memiliki langkah-langkah secara terperinci yang diadaptasi dari langkah-langkah atau tutorial yang dibuat oleh CatBoost.ai (catboost, 2023). Penyesuaian langkah-langkah pemodelan dibuat karena tipe data yang dimiliki yaitu *multiclass* berdasarkan 4 kategori atau kelas yang dimiliki oleh SES sehingga terdapat beberapa perbedaan dengan langkah-langkah awal milik CatBoos.ai.

### **3.5.1 Pra-pemrosesan Data**

Pada tahap awal dilakukan pra-pemrosesan data untuk membersihkan data dan membuatnya cocok dengan model *machine learning* yang dibuat, serta dapat meningkatkan akurasi dan efisiensi model *pembelajaran machine learning*. Pra-pemrosesan data terdiri dari beberapa tahap yaitu, mendapatkan kumpulan data, mengimpor *library*, mengimpor kumpulan data, menemukan data yang hilang, mengkodekan data kategorikal, memisahkan dataset menjadi set pelatihan (*training*) dan pengujian (*testing*), dan penskalaan fitur.

#### **3.5.1.1 Impor Libraries**

Pada proses klasifikasi akan menggunakan *library* seaborn untuk memeriksa data sebagai petak berpasangan untuk setiap variabel. Kemudian, terdapat *Scikit Learn* (Sk-learn) yang terdiri dari beberapa data umum dan pustaka matematika Python antara lain, *library Numerical Python* (NumPy) merupakan *library* python yang digunakan untuk melakukan *array*, Pandas untuk menganalisis, membersihkan, menjelajahi, dan memanipulasi data, serta Matplotlib yang digunakan untuk memplot angka-angka berdefinisi tinggi seperti diagram lingkaran, histogram, *scatterplot*, grafik, dan lain-lain. Selain itu juga yang terakhir terdapat CatBoost sebagai algoritma klasifikasi *machine learning* yang menggunakan *gradient boosting* pada *decision tree*.

#### **3.5.1.2 Impor Kumpulan Data**

Data hasil kuesioner yang sudah dikumpulkan akan di impor ke Python untuk diolah dengan menggunakan berbagai *library* yang sudah di impor untuk menghasilkan output yang diinginkan nantinya.

#### **3.5.1.3 Menemukan Data yang hilang (*Missing value*)**

Dalam melakukan pra-pemrosesan data terdapat pembersihan data, yaitu dengan memproses data yang akan menemukan dan memperbaiki data yang tidak akurat atau terdapat nilai-nilai yang hilang dalam kumpulan data.

#### **3.5.1.4 Split Data**

Pada penelitian ini split data menggunakan rasio 80/20 atau 80% data *training* dan 20% data *testing*. Menurut Wong et al pemisahan 80/20 memberikan kinerja yang baik untuk evaluasi model (Wong et al, 2005). Selain itu penggunaan rasio 80/20 membantu mencegah overfitting (Zayid & Akay, 2013). Hal tersebut juga didukung oleh Li et al yang menyatakan bahwa dengan menggunakan rasio 80/20 memungkinkan hasil evaluasi performa model lebih akurat (Li et al, 2013)

a. Data *Training*

Data *training* yang digunakan pada penelitian ini dengan komposisi 80%. Proses pembentukan data *training* yaitu dengan cara menggunakan komposisi yang telah ditentukan dari jumlah seluruh data. Sehingga data *training* yang terbentuk antara lain komposisi 80% diperoleh 320 data *training*. Berdasarkan data *training* yang diperoleh kemudian digunakan untuk menentukan pemodelan CatBoost.

b. Data *Testing*

Pada penelitian ini data *testing* merupakan sisa komposisi data yang telah digunakan pada data *training*. Komposisi yang digunakan antara lain 20%. Data *testing* yang diperoleh 20% yaitu 80 data *testing*. Hasil data *testing* digunakan untuk menentukan pemodelan CatBoost yang terbaik dengan tingkat akurasi.

### **3.5.2 Pemrosesan Data**

#### **3.5.2.1 Pelatihan Model**

Pelatihan model *machine learning* adalah proses di mana algoritma CatBoost diberi data pelatihan yang dapat dipelajari. Pelatihan model bekerja untuk menyesuaikan bobot dan bias terbaik ke algoritma untuk meminimalkan fungsi kerugian atau *loss function* pada rentang prediksi. Fungsi kerugian menentukan cara mengoptimalkan algoritma *machine learning*.

### **3.5.2.2 Metrik Evaluasi**

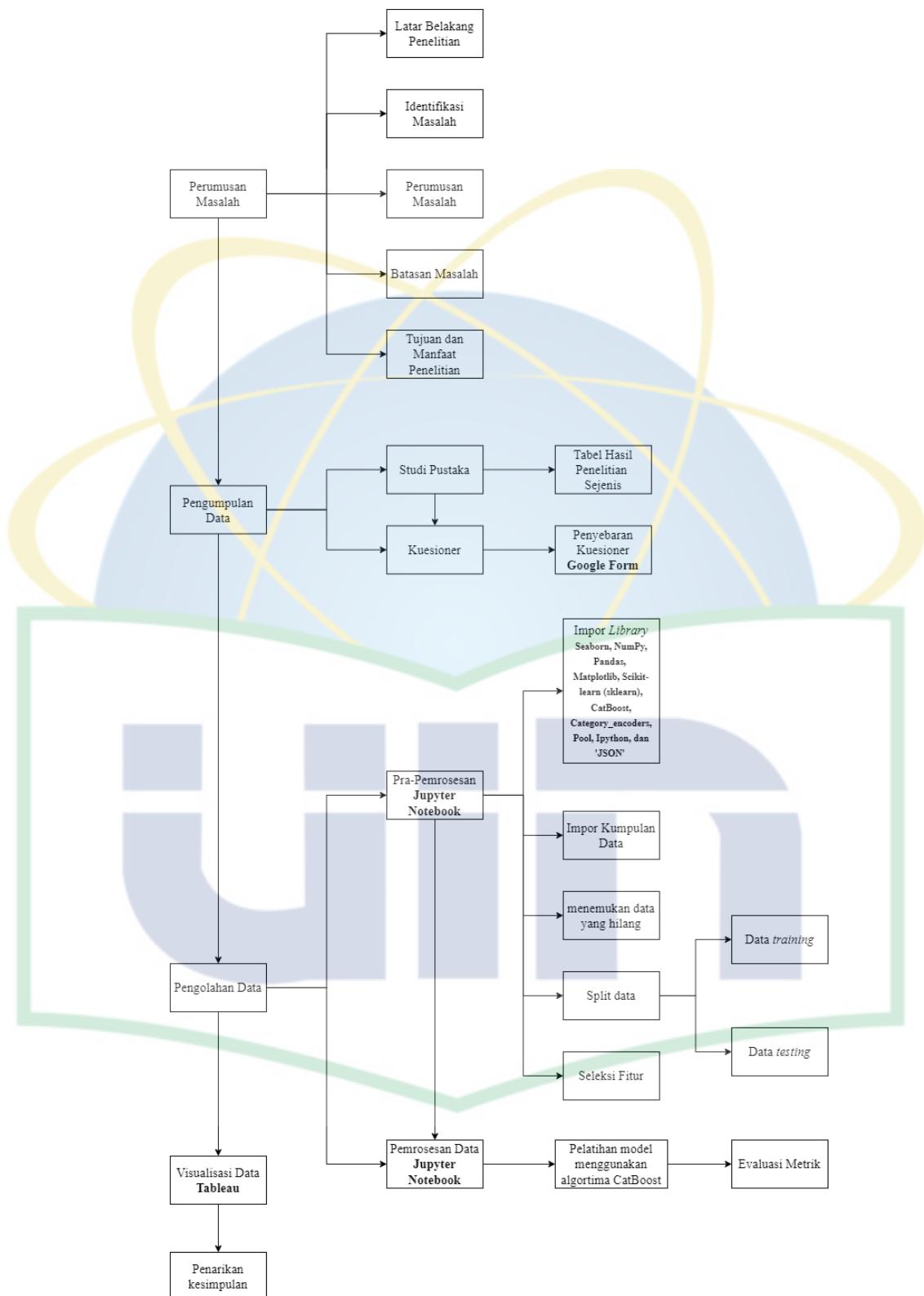
Setelah melatih model dapat dilakukan evaluasi kinerjanya pada data testing dan dapat melihat fitur apa yang paling berpengaruh dengan merepresentasikan dalam *performance metrics* antara lain *Confusion Matrix*, *Accuracy*, *Precision & Recall*, *F1 Score*, dan *Area Under ROC Curve (AUC)*.

### **3.5.2.3 Visualisasi Data**

Pada visualisasi data yang sudah dikumpulkan dari hasil kuesioner untuk setiap variabel dan juga tingkat kepentingan fitur dalam pemodelan dengan menggunakan Tableau. Visualisasi data yang dibuat untuk menjelaskan lebih rinci mengenai hubungan fitur dengan setiap kategori SES yang diprediksi. Data dukung juga divisualisasikan yang menjelaskan keterlibatan platform yang sering digunakan hingga kegiatan yang sering dilakukan yang mendukung fitur-fitur yang sudah dipilih saat pemodelan.

## **3.6 Kerangka Penelitian**

Kerangka Penelitian atau alur penelitian merupakan rangkaian proses dalam penelitian untuk mencapai suatu tujuan. Berikut merupakan alur penelitian yang dilakukan pada penelitian ini:



Gambar 3.1 Kerangka Penelitian



## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Impor Libraries**

Sebelum melakukan pemodelan dibutuhkan beberapa pustaka (*libraries*) dan paket Python yang mendukung pemodelan CatBoost ketika dijalankan. Adapun *library* dan paket Python yang digunakan antara lain, Seaborn, NumPy, Pandas, Matplotlib, Scikit-learn (*sklearn*), CatBoost, Category\_encoders, Pool yaitu kelas di CatBoost, Ipython, dan 'JSON' yaitu kelas di Ipython.

#### **4.2 Impor Kumpulan Data**

Data yang telah dikumpulkan melalui kuesioner di impor ke dalam jupyter notebook sebagai dataset yang akan diprediksi menggunakan pemodelan CatBoost. Data yang diimpor berbentuk *file csv* atau *text file*.

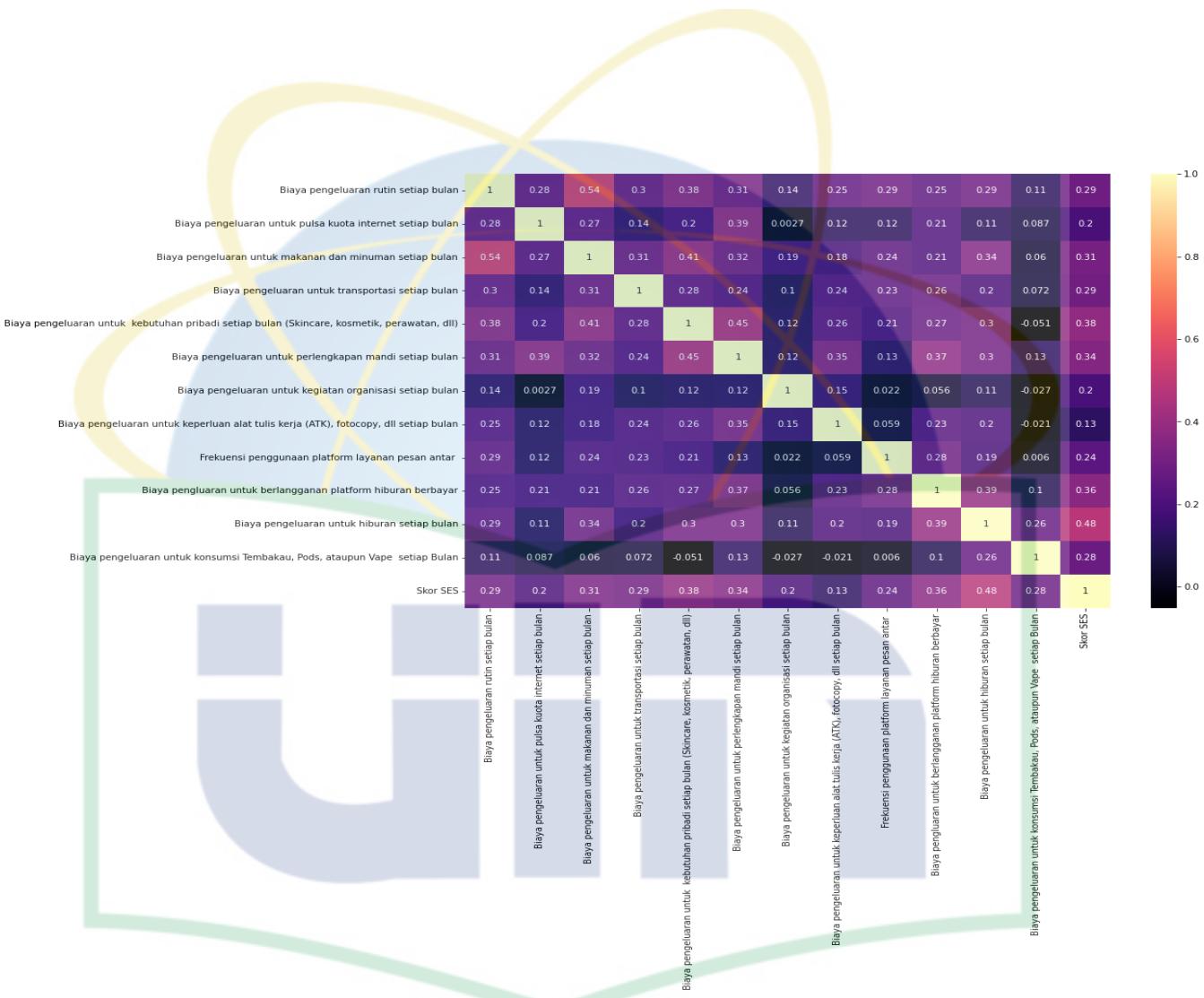
#### **4.3 Menemukan Data yang Hilang**

Pengecekan data dari kumpulan data yang telah diimpor juga dilakukan untuk menghindari adanya *missing value* atau data yang hilang dari dataset tersebut. Penggunaan *library* Pandas digunakan untuk mendapatkan informasi ada atau tidaknya data yang hilang dari dataset yang telah diimpor tersebut.

#### **4.4 Seleksi Fitur**

Membuat pemodelan untuk memprediksi SES berdasarkan gaya hidupnya dibutuhkan fitur-fitur yang dapat mempengaruhi dan menentukan kategori SES.

Seleksi fitur dilakukan dengan cara melakukan analisis korelasi data menggunakan heatmap untuk menentukan fitur yang relevan dengan output yaitu Skor SES. Hasil Seleksi fitur nantinya adalah yang akan menjadi variabel 'cat\_features' dalam set pelatihan di mana ditentukan dengan memilih kolom yang bukan tipe data numerik. Namun, variabel 'numerik' juga akan digunakan untuk memilih fitur numerik.



Gambar 4.1 Hasil *Correlation Matrix* dengan *Heatmap*

Pada gambar 4.1 dapat dilihat pada baris paling bawah, jika, fitur atau kotak yang berwarna ungu muda atau ungu maka fitur tersebut memiliki hubungan yang relevan dengan Skor SES dan sebaliknya jika terdapat fitur atau kotak yang berwarna hitam maka fitur tersebut kurang relevan dengan Skor SES. Berikut merupakan hasil analisis korelasi data dengan target Skor SES:

Tabel 4.1 Hasil Analisis Seleksi Fitur

No	Fitur	Keterangan
1.	Biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan	Relevan
2.	Biaya Pengeluaran Rutin Setiap Bulan	
3.	Biaya pengeluaran untuk hiburan setiap bulan	Relevan
4.	Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll)	Relevan
5.	Biaya pengeluaran untuk kegiatan organisasi setiap bulan	Relevan
6.	Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan	Relevan
7.	Biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap bulan	Relevan
8.	Biaya pengeluaran untuk makanan dan minuman setiap bulan	Relevan
9.	Biaya pengeluaran untuk perlengkapan mandi setiap bulan	Relevan
10.	Biaya pengeluaran untuk pulsa kuota internet setiap bulan	Relevan
11.	Frekuensi penggunaan platform layanan pesan antar	Relevan

Terdapat penyesuaian dari korelasi fitur terhadap target SES dengan mengurangi salah satu fitur yaitu frekuensi penggunaan platform layanan pesan antar yang tidak termasuk dalam atribut variabel SES melainkan hanya sebagai data dukung. Selain itu, dari hasil korelasi fitur sebagai fitur yang digunakan pada pemodelan dalam penelitian ini juga terdapat fitur tambahan dari data primer setiap responden yaitu fitur golongan UKT, fakultas, dan pekerjaan orang tua.

#### 4.5 Split Data

Pada proses split data dengan menggunakan fungsi 'train\_test\_split' dari sklearn, fungsi tersebut akan membagi data pelatihan menjadi set pelatihan (*training*) dan validasi (*testing*). Ukuran set validasi diatur ke 0,2 (20%) atau dataset

yang telah diimpor dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20%.

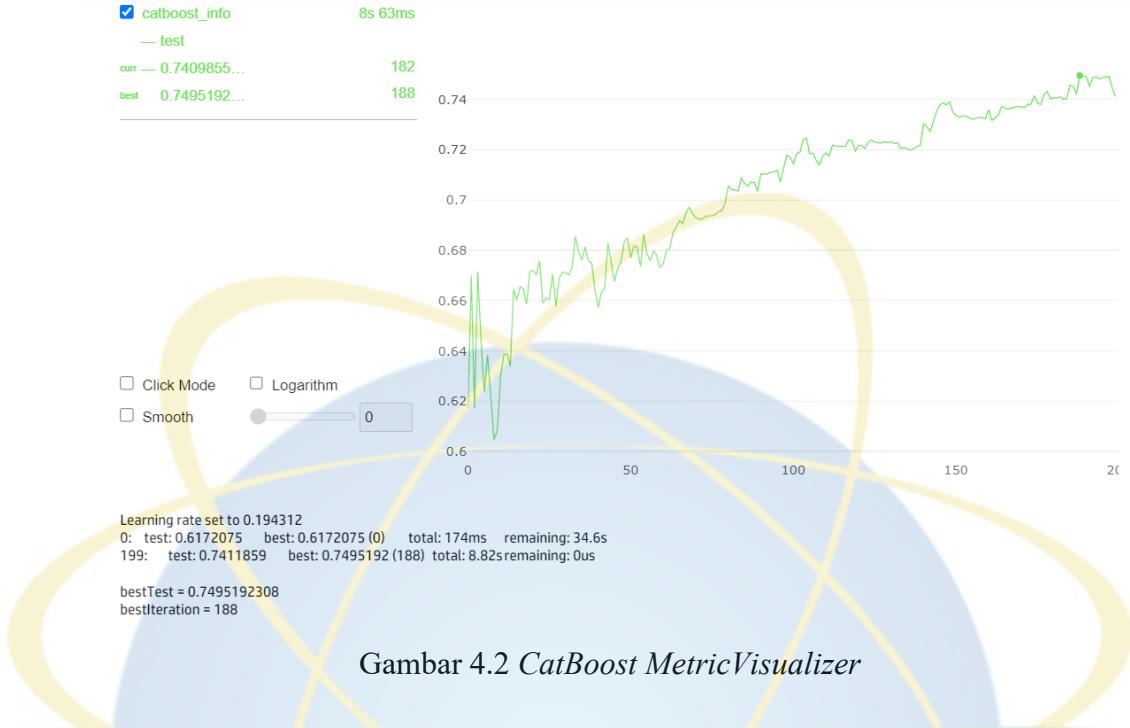
#### 4.6 Pelatihan Model

Pelatihan model CatBoost dilakukan setelah tahap pra pemrosesan data selesai dan dataset telah siap untuk dilakukan data *training* menggunakan pemodelan CatBoost. Model klasifikasi CatBoost dibuat dan dilihat menggunakan hyperparameter yang telah ditentukan.

#### 4.7 Hasil Pemodelan CatBoost

Hasil nilai "*Best Test*" pemodelan CatBoost yang dibuat sebesar 0,7495. Nilai "*Best Test*" menunjukkan performa terbaik yang dicapai oleh model CatBoost pada metrik evaluasi tertentu selama proses pelatihan. Metrik evaluasi terdiri dari *accuracy*, *F1-score*, *precision & recall*, *confusion matrix*, dan *AUC*.

Hasil nilai "*Best Iteration*" pemodelan CatBoost yang dibuat menunjukkan iterasi atau putaran selama pelatihan dengan kinerja terbaik yaitu pada iterasi ke 188. Dalam peningkatan gradien, algoritma secara iteratif menambahkan data *testing* yang lemah (biasanya pohon keputusan) ke dalam model, dengan setiap data *testing* yang baru untuk mengoreksi kesalahan sebelumnya. Gambar 4.2 merupakan gambar grafik iterasi yang dilakukan oleh model yang telah dijelaskan.



#### 4.8 Feature Importances

Menentukan akurasi SES dengan menggunakan pemodelan CatBoost dibutuhkan fitur-fitur yang dipilih dari data yang telah dikumpulkan. Pada pemodelan CatBoost ini dipilih 14 fitur-fitur yang dipilih. Setiap fitur yang dipilih memiliki skor kepentingan yang berbeda. Skor kepentingan fitur digunakan untuk memahami fitur mana yang paling relevan dengan prediksi model dan dapat digunakan untuk memandu pemilihan fitur dan upaya pengoptimalan model.

*Feature ID* adalah indeks atau nama fitur dalam kumpulan data dan *importances* merupakan skor yang menunjukkan kepentingan relatif fitur dalam memprediksi variabel target. Semakin tinggi skornya, semakin penting fitur tersebut. Pada penelitian ini tingkat kepentingan fitur dibagi berdasarkan peringkat menjadi tingkat kepentingan berdasarkan rentang skor atau persentil dengan empat tingkat kepentingan yaitu sangat tinggi, tinggi, sedang, dan rendah (Molnar, 2019). Batas fitur setiap tingkatan memiliki proporsi yang berbeda antara lain, skor di 10% teratas sangat tinggi, fitur di 25% tinggi, fitur di 50% teratas sedang, dan sisanya rendah (Li et al, 2014).

Berikut merupakan *output* skor kepentingan fitur pada empat belas fitur yang digunakan dalam pemodelan:

Tabel 4.2 *Feature Importances*

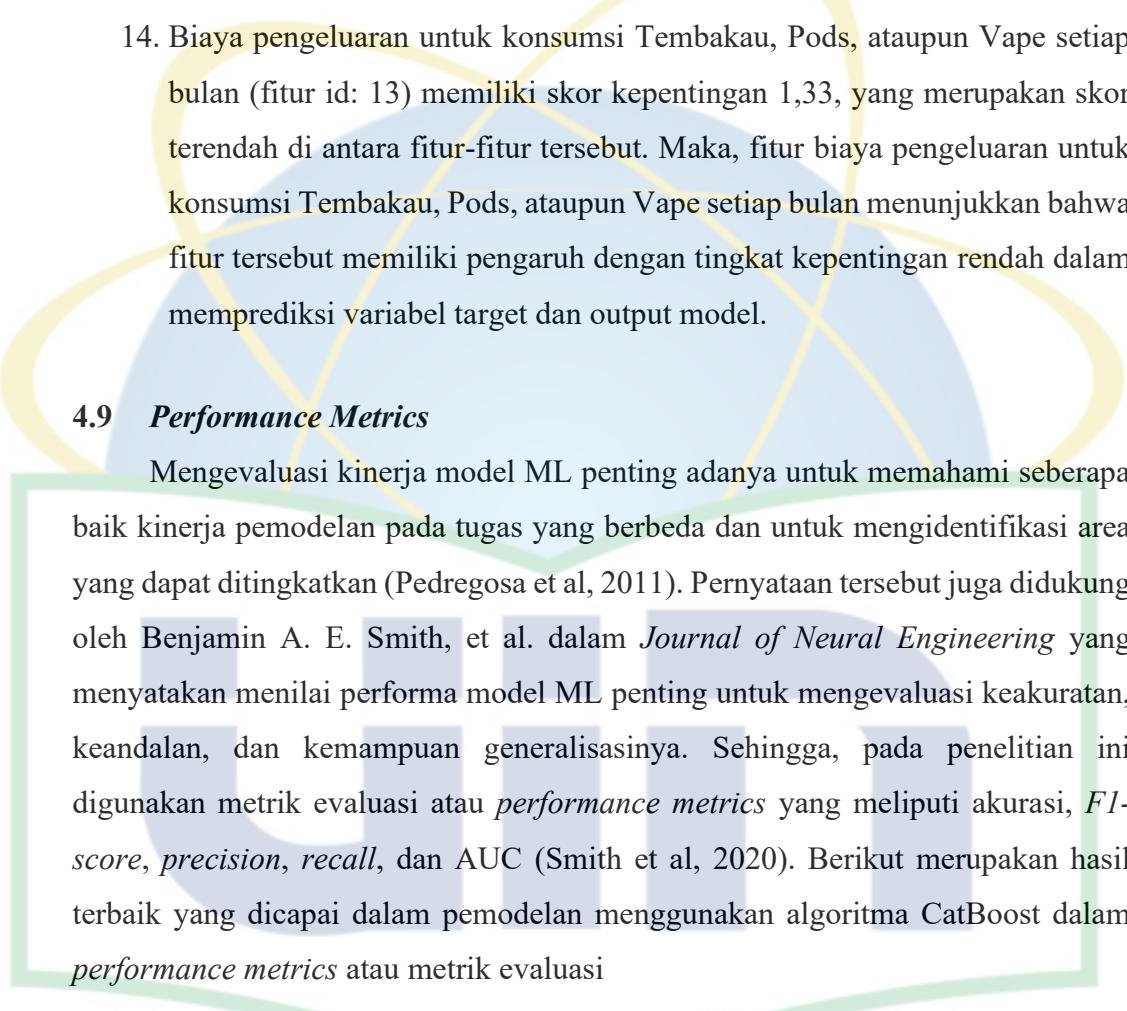
<i>Feature Id</i>	<i>Feature</i>	<i>Importances</i>
0	Golongan UKT	21,56
1	Pekerjaan Orang Tua	17,14
2	Fakultas	11,36
3	Biaya pengeluaran untuk hiburan setiap bulan	10,62
4	Biaya pengeluaran rutin setiap bulan	6,05
5	Biaya pengeluaran untuk transportasi setiap bulan	5,27
6	Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll)	4,80
7	Biaya pengeluaran untuk berlangganan platform hiburan berbayar	4,71
8	Biaya pengeluaran untuk perlengkapan mandi setiap bulan	4,07
9	Biaya pengeluaran untuk kegiatan organisasi setiap bulan	3,72
10	Biaya pengeluaran untuk makanan dan minuman setiap bulan	3,39
11	Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan	3,37
12	Biaya pengeluaran untuk pulsa kuota internet setiap bulan	2,61
13	Biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap bulan	1,33

1. Golongan UKT (id fitur: 0) memiliki skor kepentingan 21,56, yang menunjukkan bahwa fitur tersebut merupakan fitur yang paling penting

dalam memprediksi variabel target dengan tingkat kepentingan sangat tinggi terhadap output model.

2. Pekerjaan Orang Tua (id fitur: 1) memiliki skor kepentingan 17,44, yang merupakan skor tertinggi kedua di antara fitur-fitur tersebut. Fitur pekerjaan orang tua menunjukkan bahwa fitur tersebut memiliki tingkat kepentingan tinggi dalam memprediksi variabel target.
3. Fakultas (fitur id: 2) memiliki skor kepentingan 11,36, yang merupakan skor tertinggi ketiga di antara fitur-fitur tersebut. Fitur fakultas menunjukkan bahwa fitur tersebut memiliki tingkat kepentingan tinggi pada output model.
4. Biaya pengeluaran untuk hiburan setiap bulan (fitur id: 3) memiliki skor kepentingan 10,62, yang merupakan skor tertinggi keempat di antara fitur-fitur tersebut. Fitur Biaya pengeluaran untuk hiburan setiap bulan menunjukkan bahwa fitur tersebut memiliki tingkat kepentingan tinggi dalam memprediksi variabel target.
5. Biaya pengeluaran rutin setiap bulan (fitur id: 4) memiliki skor kepentingan 6,05, yang merupakan skor tertinggi kelima di antara fitur-fitur tersebut. Fitur biaya pengeluaran rutin setiap bulan menunjukkan bahwa fitur tersebut memiliki tingkat kepentingan tinggi dalam memprediksi variabel target.
6. Biaya pengeluaran untuk transportasi setiap bulan (fitur id: 5) memiliki skor kepentingan 5,27, yang merupakan skor tertinggi keenam di antara fitur-fitur tersebut. Fitur Biaya pengeluaran untuk transportasi setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.
7. Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (*Skincare*, kosmetik, perawatan, dll) (fitur id: 6) memiliki skor kepentingan 4,80, yang merupakan skor tertinggi ketujuh di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk kebutuhan pribadi setiap bulan (*Skincare*, kosmetik, perawatan, dll) menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.

8. Biaya pengeluaran untuk berlangganan platform hiburan berbayar (fitur id: 7) memiliki skor kepentingan 4,71, yang merupakan skor tertinggi kedelapan di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk berlangganan platform hiburan berbayar menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.
9. Biaya pengeluaran untuk perlengkapan mandi setiap bulan (fitur id: 8) memiliki skor kepentingan 4,07, yang merupakan skor tertinggi kesembilan di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk perlengkapan mandi setiap bulan setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.
10. Biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan (fitur id: 9) memiliki skor kepentingan 3,34, yang merupakan skor tertinggi kesepuluh di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.
11. Biaya pengeluaran untuk makanan dan minuman setiap bulan (fitur id: 10) memiliki skor kepentingan 3,39, yang merupakan skor tertinggi kesebelas di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk makanan dan minuman setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.
12. Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), photocopy, dll setiap bulan (fitur id: 11) memiliki skor kepentingan 3,37, yang merupakan skor tertinggi kedua belas di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk keperluan alat tulis kerja (ATK), photocopy, dll setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan sedang dalam memprediksi variabel target dan output model.

- 
13. Biaya pengeluaran untuk pulsa kuota internet setiap bulan (fitur id: 12) memiliki skor kepentingan 2,61, yang merupakan skor ketiga terendah di antara fitur-fitur tersebut. Fitur biaya pengeluaran untuk pulsa kuota internet setiap bulan menunjukkan bahwa fitur memiliki pengaruh dengan tingkat kepentingan rendah dalam memprediksi variabel target dan output model.
  14. Biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap bulan (fitur id: 13) memiliki skor kepentingan 1,33, yang merupakan skor terendah di antara fitur-fitur tersebut. Maka, fitur biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap bulan menunjukkan bahwa fitur tersebut memiliki pengaruh dengan tingkat kepentingan rendah dalam memprediksi variabel target dan output model.

#### 4.9 *Performance Metrics*

Mengevaluasi kinerja model ML penting adanya untuk memahami seberapa baik kinerja pemodelan pada tugas yang berbeda dan untuk mengidentifikasi area yang dapat ditingkatkan (Pedregosa et al, 2011). Pernyataan tersebut juga didukung oleh Benjamin A. E. Smith, et al. dalam *Journal of Neural Engineering* yang menyatakan menilai performa model ML penting untuk mengevaluasi keakuratan, keandalan, dan kemampuan generalisasinya. Sehingga, pada penelitian ini digunakan metrik evaluasi atau *performance metrics* yang meliputi akurasi, *F1-score*, *precision*, *recall*, dan AUC (Smith et al, 2020). Berikut merupakan hasil terbaik yang dicapai dalam pemodelan menggunakan algoritma CatBoost dalam *performance metrics* atau metrik evaluasi

##### 4.9.1 *Accuracy*

Hasil akurasi SES menggunakan pemodelan CatBoost didapatkan nilai keakuratan dari hasil data training sebesar 0,8625 yang menjelaskan bahwa pengklasifikasi CatBoost dapat memprediksi label kelas dengan benar sebesar 85,94% dari data input. Dengan harapan jika model diberi set data baru, diharapkan dapat memprediksi dengan benar label kelas sekitar 85,94% dari instance dalam set tersebut.

Penting untuk diperhatikan bahwa akurasi tidak selalu merupakan metrik terbaik yang digunakan untuk mengevaluasi performa model, terutama saat menangani kumpulan data yang tidak seimbang atau saat berbagai jenis kesalahan klasifikasi memiliki biaya yang berbeda. Dalam kasus seperti itu, metrik lain seperti presisi, daya ingat, dan skor F1 dapat memberikan evaluasi performa model yang lebih bermuansa.

#### 4.9.2 *Precision, Recall, & F1 Score*

Berikut merupakan hasil dari data training pemodelan CatBoost dalam matrik *precision*, *recall*, dan F1-score yang terdapat pada tabel 4.3.

Tabel 4.3 *Precision, Recall, & F1 Score*

<i>CLASS</i>	<i>PRECISION</i>	<i>RECALL</i>	<i>F1-SCORE</i>	<i>SUPPORT</i>
SES A	0,9643	0,8308	0,8926	65
SES B	0,8205	0,9412	0,8767	102
SES C1	0,8375	0,8171	0,8272	82
SES C2	1,0	0,4286	0,6	7

- a. Kelas berlabel "SES A".

*Precision* (0,9643): Pada saat model memprediksi sebuah contoh data dari "SES A" hasilnya menunjukkan proporsi prediksi positif atau kondisi hasil prediksi benar dan realitanya ternyata benar (TP) dengan nilai *precision* sebesar 0,9643.

*Recall* (0,8308): Berdasarkan semua contoh data yang benar-benar positif atau milik "SES A", dalam hal ini model memprediksi dengan benar prediksi positif dengan nilai *recall* sebesar 83,08% dari total positif.

*F1-score* (0,8926): Rata-rata harmonik dari *precision* dan *recall* dengan menggabungkan kedua ukuran menjadi satu skor atau *F1-score* untuk "SES A" memiliki nilai sebesar 0,8926. Nilai *F1-score* dalam pemodelan yang dilakukan dapat secara akurat memprediksi kejadian "SES A" dengan tingkat *precision* dan *recall* yang cukup tinggi.

*Support* (65): Jumlah contoh dalam kumpulan data yang termasuk dalam kelas "SES A". Pada proses pemodelan ini, terdapat 65 contoh yang digunakan dari dalam kumpulan data milik "SES A".

b. Kelas berlabel "SES B".

*Precision* (0,8205): Pada saat model memprediksi sebuah contoh data dari "SES B" hasilnya menunjukkan proporsi prediksi positif atau kondisi hasil prediksi benar dan realitanya ternyata benar (TP) dengan nilai *precision* sebesar 0,8205.

*Recall* (0,9412): Berdasarkan semua contoh data yang benar-benar positif atau milik "SES B", dalam hal ini model memprediksi dengan benar prediksi positif dengan nilai *recall* sebesar 94,12% dari total positif.

*F1-score* (0,8767): Rata-rata harmonik dari *precision* dan *recall* dengan menggabungkan kedua ukuran menjadi satu skor atau *F1-score* untuk "SES B" memiliki nilai sebesar 0,8767. Nilai *F1-score* dalam pemodelan yang dilakukan dapat secara akurat memprediksi kejadian "SES B" dengan tingkat *precision* dan *recall* yang cukup tinggi.

*Support* (102): Jumlah jumlah contoh dalam kumpulan data yang termasuk dalam kelas "SES B". Pada proses pemodelan ini, terdapat 102 contoh yang digunakan dari dalam kumpulan data milik "SES B".

c. Kelas berlabel "SES C1".

*Precision* (0,8375): Pada saat model memprediksi sebuah contoh data dari "SES C1" hasilnya menunjukkan proporsi prediksi positif atau kondisi hasil prediksi benar dan realitanya ternyata benar (TP) dengan nilai *precision* sebesar 0,8375.

*Recall* (0,8171): Berdasarkan semua contoh data yang benar-benar positif atau milik "SES C1", dalam hal ini model memprediksi dengan benar prediksi positif dengan nilai *recall* sebesar 81,71% dari total positif.

*F1-score* (0,8272): Rata-rata harmonik dari *precision* dan *recall* dengan menggabungkan kedua ukuran menjadi satu skor atau *F1-score* untuk "SES C1" memiliki nilai sebesar 0,8272. Nilai *F1-score* dalam pemodelan yang

dilakukan dapat secara akurat memprediksi kejadian "SES C1" dengan tingkat *precision* dan *recall* yang cukup tinggi.

*Support* (82): Jumlah jumlah contoh dalam kumpulan data yang termasuk dalam kelas "SES C1". Pada proses pemodelan ini, terdapat 82 contoh yang digunakan dari dalam kumpulan data milik "SES C1".

- d. Kelas berlabel "SES C2".

*Precision* (1,0): Pada saat model memprediksi sebuah contoh data dari "SES C2" hasilnya menunjukkan proporsi prediksi positif atau kondisi hasil prediksi benar dan realitanya ternyata benar (TP) dengan nilai *precision* sebesar 1,0.

*Recall* (0,4286): Berdasarkan semua contoh data yang benar-benar positif atau milik "SES C2", dalam hal ini model memprediksi dengan benar prediksi positif dengan nilai *recall* sebesar 42,86% dari total positif.

*F1-score* (0,6): Rata-rata harmonik dari *precision* dan *recall* dengan menggabungkan kedua ukuran menjadi satu skor atau *F1-score* untuk "SES C2" memiliki nilai sebesar 0,6. Nilai *F1-score* dalam pemodelan yang dilakukan dapat secara akurat memprediksi kejadian "SES C2" dengan tingkat *precision* dan *recall* yang cukup rendah.

*Support* (7): Jumlah jumlah contoh dalam kumpulan data yang termasuk dalam kelas "SES C2". Pada proses pemodelan ini, terdapat 7 contoh yang digunakan dari dalam kumpulan data milik "SES C2".

#### 4.9.3 *Confusion Matrix*

Menginterpretasikan kinerja model dengan menggunakan *confusion matrix* dalam memahami seberapa baik performa model dan dapat membantu meningkatkan performa model dengan menggabungkan beberapa metrik yang dapat dihitung antara lain, *accuracy*, *precision & recall*, dan *F1-score*. *Confusion matrix* dari pemodelan CatBoost yang dilakukan menghasilkan kondisi prediksi dan realita yang dijelaskan dalam TP, FP, TN, dan FN yang terdiri empat kelas sebagai berikut:

- Gambar 4.3 menunjukkan bahwa hasil prediksi benar dan realitanya benar (TP) pada SES A berjumlah 54, hasil prediksi salah dan realitanya benar (FP) berjumlah 2, hasil prediksi salah dan realitanya salah (FN) berjumlah 11, dan hasil prediksi benar dan realitanya salah berjumlah 189 pada SES A.
- Pada SES B hasil prediksi benar dan realitanya benar (TP) berjumlah 96, hasil prediksi salah dan realitanya benar (FP) berjumlah 21, hasil prediksi salah dan realitanya salah (FN) berjumlah 6, dan hasil prediksi benar dan realitanya salah berjumlah 133 pada SES B.
- Hasil prediksi benar dan realitanya benar (TP) pada SES C1 berjumlah 67, hasil prediksi salah dan realitanya benar (FP) berjumlah 13, hasil prediksi salah dan realitanya salah (FN) berjumlah 15, dan hasil prediksi benar dan realitanya salah berjumlah 161 pada SES C1.
- Sedangkan pada hasil prediksi benar dan realitanya benar (TP) SES C2 berjumlah 3, hasil prediksi salah dan realitanya benar (FP) berjumlah 0, hasil prediksi salah dan realitanya salah (FN) berjumlah 4, dan hasil prediksi benar dan realitanya salah berjumlah 249 pada SES C2.

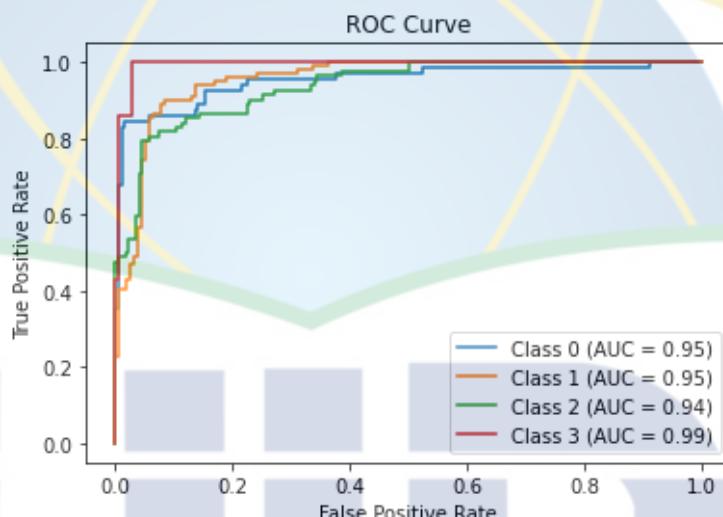
		Confusion Matrix				
		SES A	SES B	SES C1	SES C2	
Predicted	SES A	54	6	5	0	-80
	SES B	1	96	5	0	-60
SES C1	SES A	1	14	67	0	-40
	SES B	0	1	3	3	-20
SES C2	SES A	0	1	3	3	-0
	SES B	0	1	3	3	-0

Gambar 4.3 *Confusion Matrix*

#### 4.9.4 Area Under ROC Curve (AUC)

*Area Under ROC Curve (AUC)* pada gambar 4.4 menunjukkan hasil dari pemodelan CatBoost sebagai bagian dari metrik evaluasi kinerja model yang digunakan berdasarkan kelas yang terdapat pada keseluruhan data yaitu terbagi ke dalam 4 kelas sebagai berikut:

- *Class 0* atau SES A memiliki tingkat nilai 0.95 (klasifikasi sangat baik),
- *Class 1* atau SES B memiliki tingkat nilai 0.95 (klasifikasi sangat baik),
- *Class 2* atau SES C1 memiliki tingkat nilai 0.94 (klasifikasi sangat baik),
- *Class 3* atau SES C2 memiliki tingkat nilai 0.99 (klasifikasi sangat baik).



Gambar 4.4 AUC-ROC Curve

### 4.10 Visualisasi Data

#### 4.10.1 Profil Mahasiswa UIN Jakarta

Kuesioner yang mendapatkan 400 responden pada penelitian ini yang terdiri dari 250 berjenis kelamin perempuan dan 150 laki-laki mewakili dari seluruh mahasiswa/i UIN Jakarta yang terdiri dari 12 fakultas dan 52 jurusan.

#### 4.10.2 Variabel SES

Berdasarkan jumlah skor dari 3 variabel SES didapatkan 4 kategori SES yaitu SES A, SES B, SES C1, dan SES C2. SES D dan SES E tidak terdapat pada hasil pengolahan data penelitian ini karena seluruh data dan jumlah skor yang

didapatkan oleh masing-masing individu tidak ada yang memiliki rentang skor SES D dan SES E. Berikut merupakan visualisasi variable SES pada setiap kategori SES:

A. Biaya Pengeluaran Rutin Setiap Bulan

Berdasarkan total skor SES yang dinilai dari 3 variabel SES yaitu biaya pengeluaran rutin, sumber air minum, dan sumber bahan bakar untuk memasak dari 400 responden yang mewakili mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 didapatkan bahwa mahasiswa/i UIN Jakarta terbagi menjadi 4 kategori yaitu SES A, B, C1, dan C2. Rata - rata pengeluaran rutin yang dikategorikan sesuai skor SES nya mahasiswa/i UIN Jakarta yang termasuk ke dalam kategori SES A memiliki biaya pengeluaran rutin lebih dari Rp1.500.001, SES B Rp1.200.001 - Rp1.500.000, SES C1 Rp700.000 - Rp1.200.000, dan SES C2  $\leq$  Rp700.000 yang dapat dilihat pada tabel 4.4.

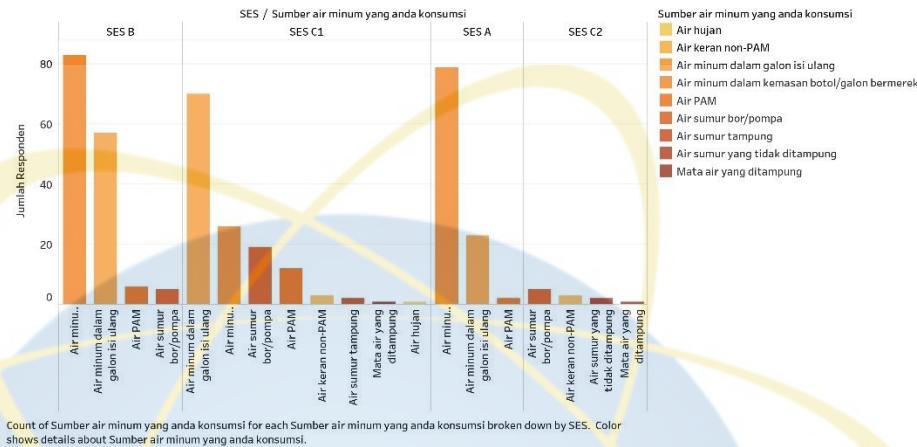
Tabel 4.4 Biaya Pengeluaran  
Rutin Mahasiswa/i UIN Jakarta per Bulan

Kategori SES	Biaya Pengeluaran Rutin Mahasiswa/i UIN Jakarta per Bulan
A	$\geq$ Rp1.500.001
B	Rp1.200.001 - Rp1.500.000
C1	Rp700.000 - Rp1.200.000
C2	$\leq$ Rp700.000

B. Sumber Air Minum

Pada salah satu variabel SES yaitu sumber air minum yang terdapat pada gambar 4.5 dapat dilihat bahwa SES A dan B didominasi oleh mereka yang mengkonsumsi air minum dalam kemasan botol/galon bermerek. Namun, kategori SES C1 lebih banyak yang mengkonsumsi air minum dalam galong isi ulang dan pada SES C2 paling banyak mengkonsumsi dari air sumur bor/pompa. Hal tersebut telah menggambarkan secara sesuai tingkat SES mereka berdasarkan air minum yang dikonsumsi.

Variabel SES Sumber Air Minum

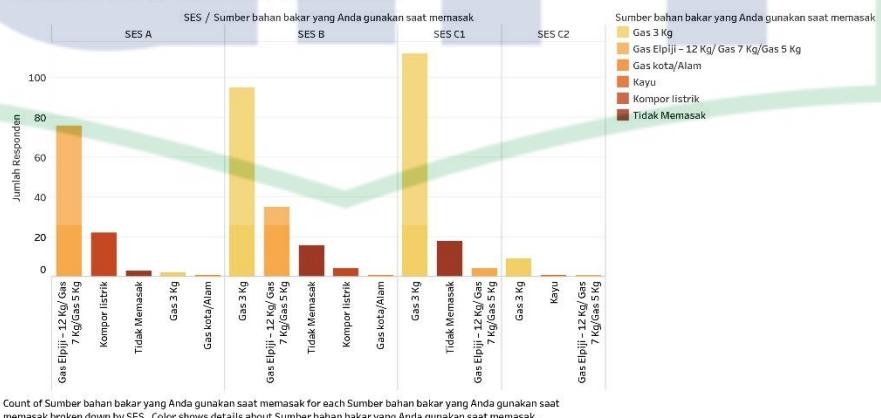


Gambar 4.5 Sumber Air Minum Berdasarkan SES

### C. Sumber Bahan Bakar untuk Memasak

Variabel SES sumber bahan bakar untuk memasak yang divisualisasikan pada gambar 4.6, dapat dilihat bahwa SES A lebih banyak yang menggunakan gas elpsi – Gas 12Kg/ Gas 7Kg/ Gas 5Kg yang merupakan sumber bahan bakar untuk memasak yang memiliki skor tertinggi. Pada kategori SES B, C1, dan C2 lebih banyak yang menggunakan Gas 3 Kg.

Variabel SES Sumber Bahan Bakar untuk Memasak

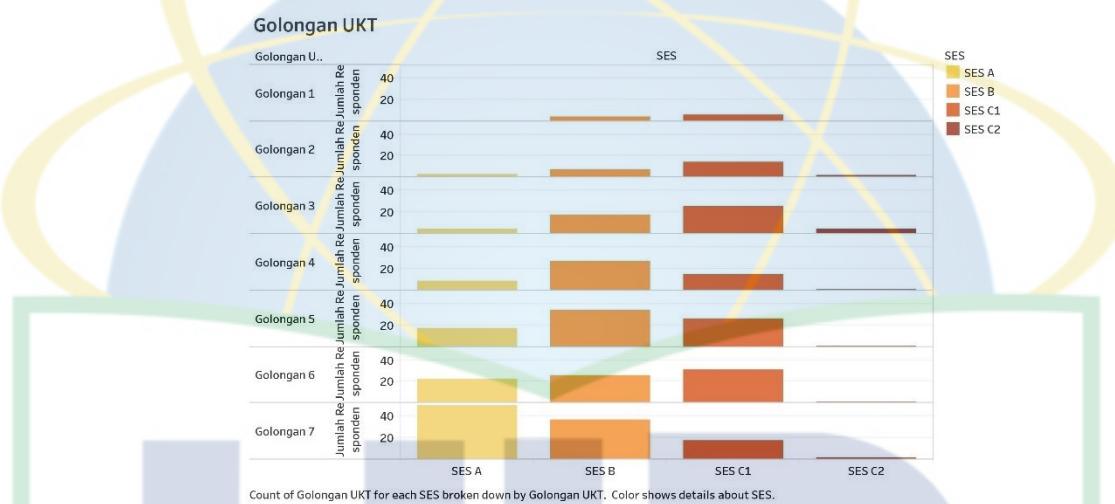


Gambar 4.6 Sumber Bahan Bakar untuk Memasak Berdasarkan SES

### 4.10.3 Kepentingan Fitur

#### 4.10.3.1 Golongan UKT

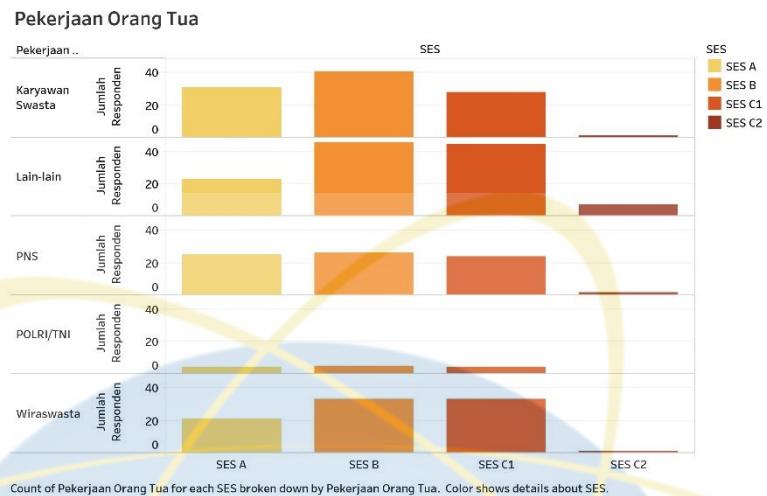
Berdasarkan kepentingan fitur yang paling penting yaitu Golongan UKT pada gambar 4.7, dapat dilihat bahwa mahasiswa/i yang termasuk SES A paling banyak berada di UKT golongan 7. SES B didominasi oleh mahasiswa/i yang memiliki UKT di golongan 7 dan golongan 5. Pada SES C1 mahasiswa/i UIN Jakarta paling banyak memiliki UKT 3,5, dan 6. Sedangkan pada SES C2 paling banyak mahasiswa/i memiliki UKT di golongan 3.



Gambar 4.7 Hubungan SES dengan Golongan UKT

#### 4.10.3.2 Pekerjaan Orang Tua

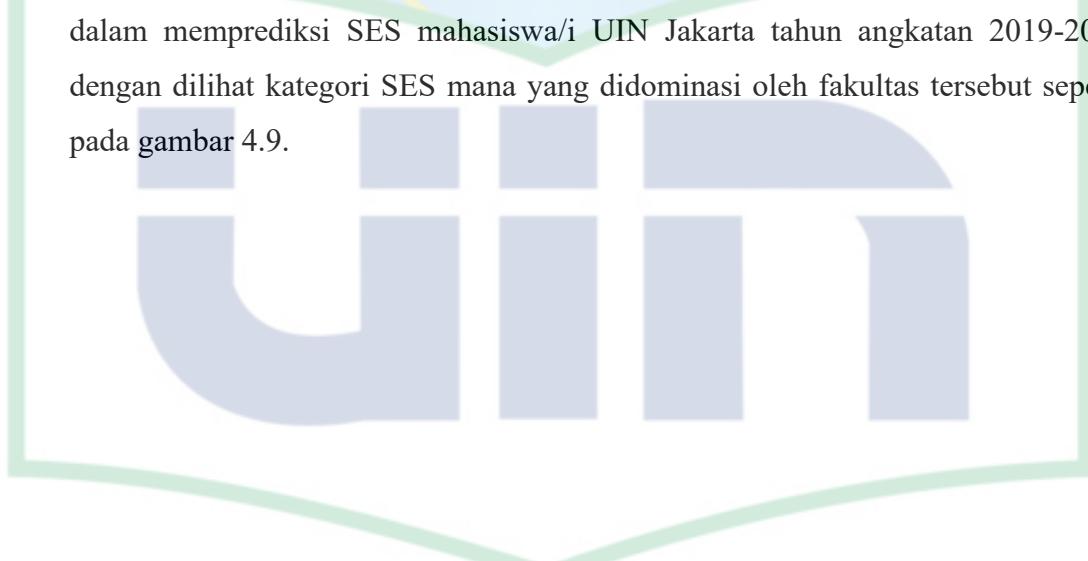
Menurut data yang telah dihimpun sebanyak 400 responden pada penelitian ini yang mewakili seluruh mahasiswa/i UIN Jakarta, dapat dilihat pada gambar 4.8 bahwa pekerjaan orang tua mahasiswa/i UIN Jakarta didominasi oleh pekerjaan lainnya diluar pekerjaan yang tersedia atau tercantum pada kuesioner, lalu diikuti oleh karyawan swasta, wiraswasta, PNS, dan TNI/Polri.

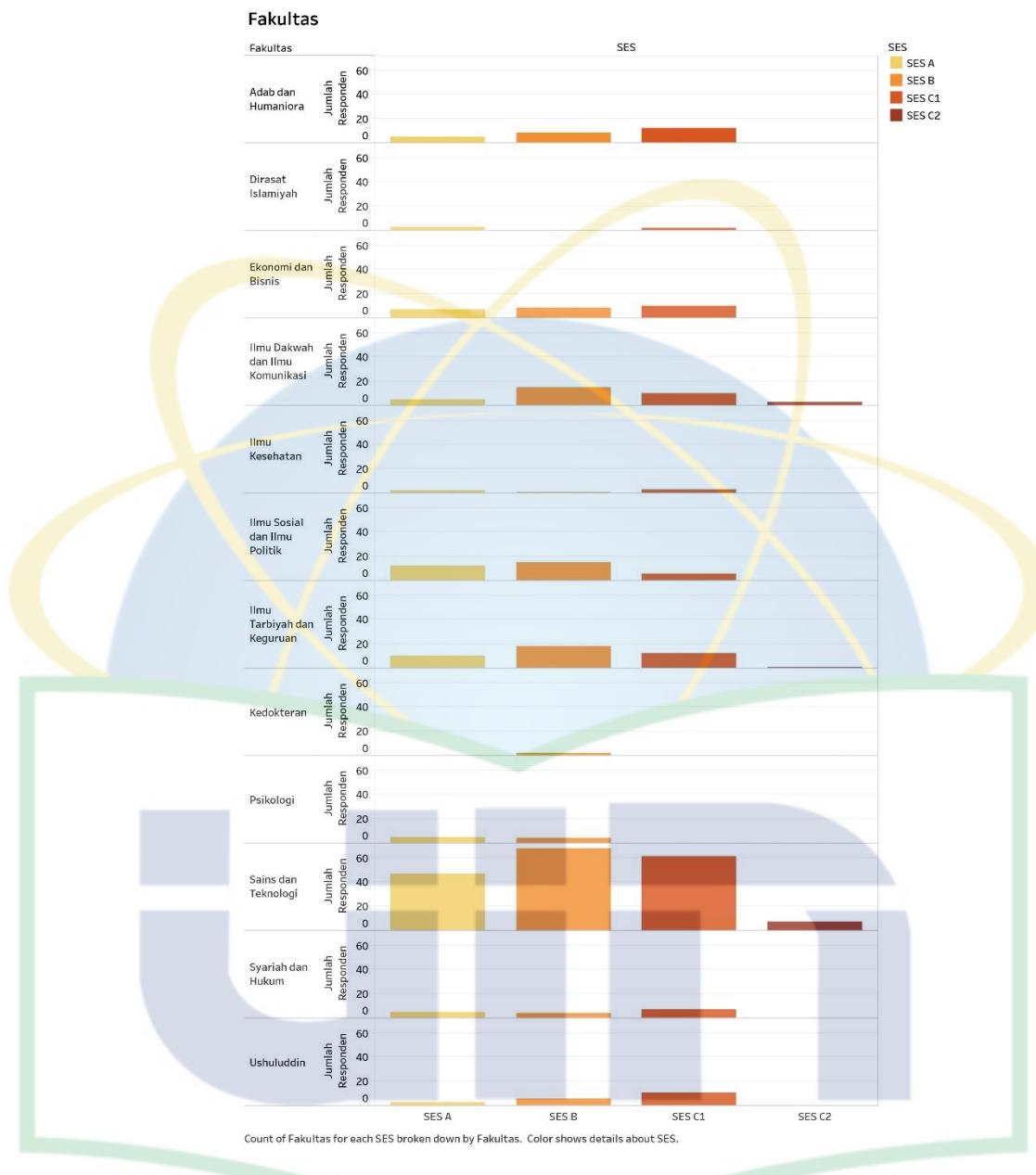


Gambar 4.8 Hubungan SES dengan Pekerjaan Orang Tua

#### 4.10.3.3 Fakultas

UIN Jakarta yang terdiri dari 12 fakultas telah diwakilkan oleh 400 responden yang didapatkan. Fakultas-fakultas tersebut memiliki peran penting dalam memprediksi SES mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 dengan dilihat kategori SES mana yang didominasi oleh fakultas tersebut seperti pada gambar 4.9.



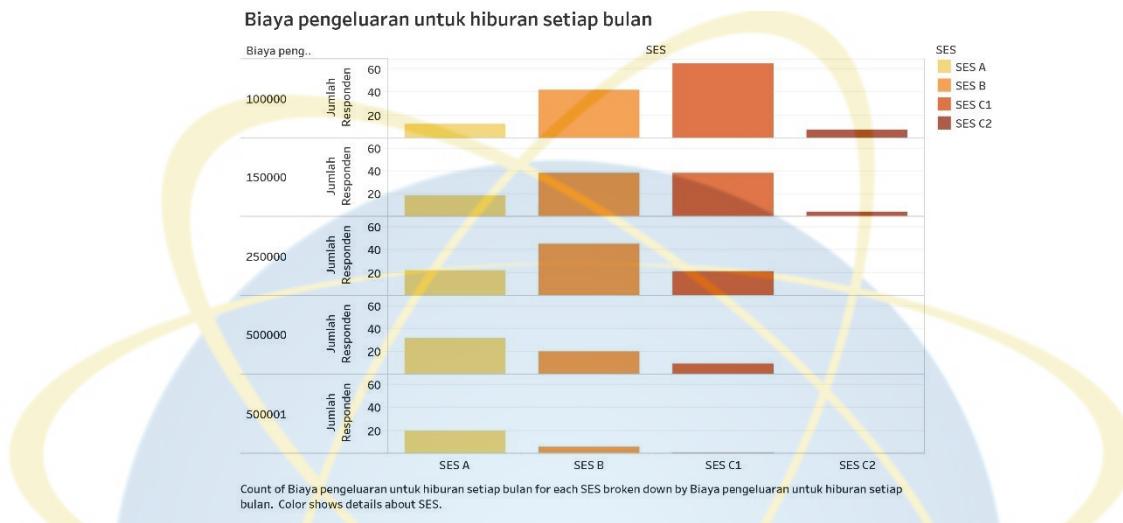


Gambar 4.9 Hubungan SES dengan Fakultas

#### 4.10.3.4 Biaya pengeluaran untuk hiburan setiap bulan

Biaya pengeluaran hiburan berdasarkan kategori SES dibawah pada setiap kategori mewakili besar kecilnya biaya yang mereka keluarkan. Pada SES A paling banyak pengeluaran terbesar untuk hiburan yang dikeluarkan yaitu Rp.250.000 – Rp.500.000. Sedangkan pada SES C2 yang merupakan kategori SES terendah biaya yang dikeluarkan paling banyak kurang dari Rp.100.000 dan paling besar

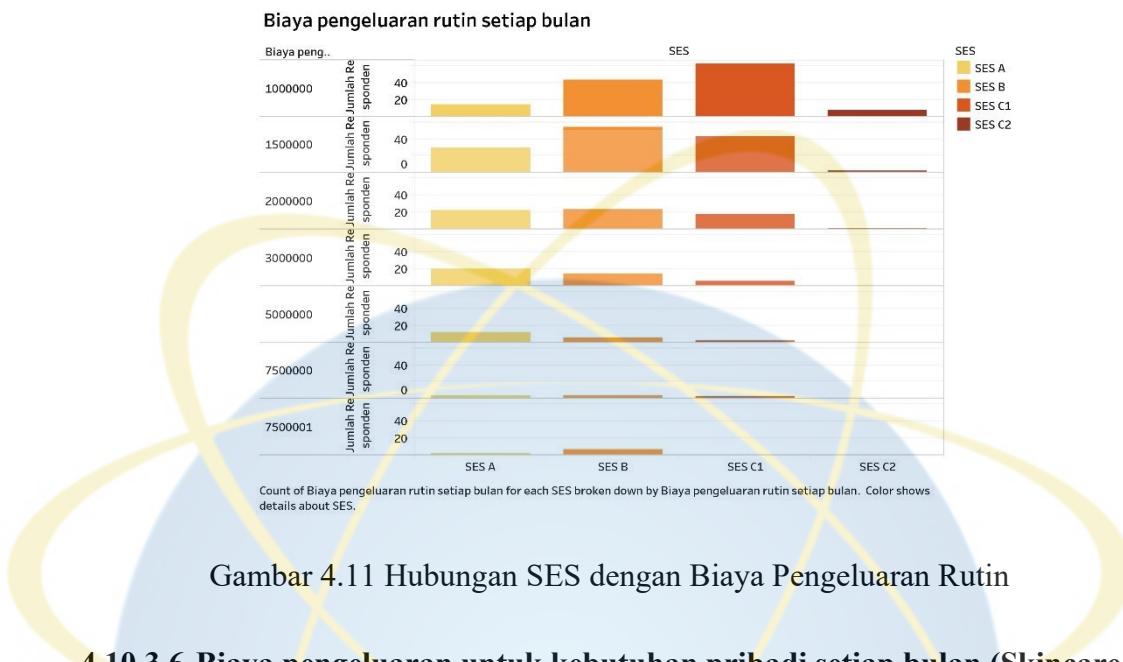
Rp.100.000. Maka biaya pengeluaran untuk hiburan setiap bulan memiliki pengaruh besar terhadap gaya hidup dan biaya pengeluaran yang paling mempengaruhi biaya pengeluaran rutin mahasiswa/i UIN Jakarta.



Gambar 4.10 Hubungan SES dengan Biaya Pengeluaran untuk Hiburan

#### 4.10.3.5 Biaya pengeluaran rutin setiap bulan

Pada kategori SES A dan SES B oleh mahasiswa/i UIN Jakarta yang memiliki biaya pengeluaran rutin yang dikeluarkan setiap bulannya sebesar Rp. 1.000.000-1.500.000 sedangkan pada kategori SES C1 dan C2 di dominasi oleh mahasiswa/i yang memiliki biaya pengeluaran rutin kurang dari 1.000.000 setiap bulannya. Pada gambar 4.11 memvisualisasikan biaya pengeluaran rutin yang dikeluarkan oleh mahasiswa/i UIN Jakarta berdasarkan kategori SES nya.

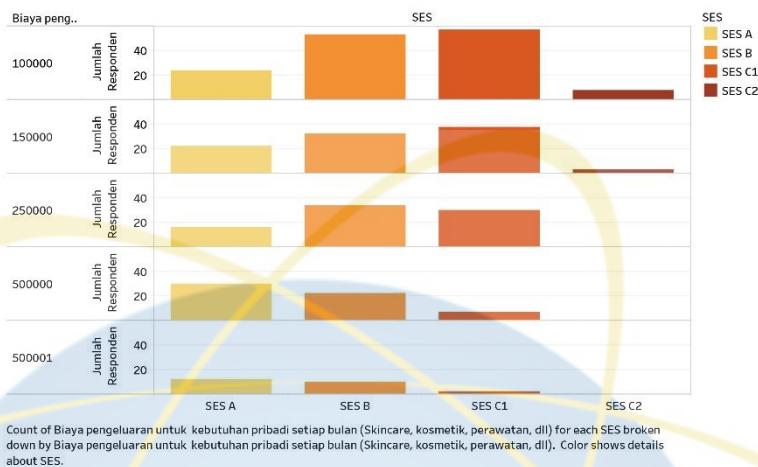


Gambar 4.11 Hubungan SES dengan Biaya Pengeluaran Rutin

#### 4.10.3.6 Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll)

Pada biaya pengeluaran kebutuhan pribadi pada setiap kategori, biaya yang dikeluarkan kurang dari Rp.100.000 namun tetap hasil penghimpunan data pada pertanyaan ini bervariasi. Dapat dilihat bahwa golongan SES A juga cukup banyak yang mengeluarkan biaya kisaran Rp.500.000, SES B Rp.250.000, SES C1 Rp.150.000, dan SES C2 Rp.150.000 namun pada SES C2 tidak terlalu signifikan.

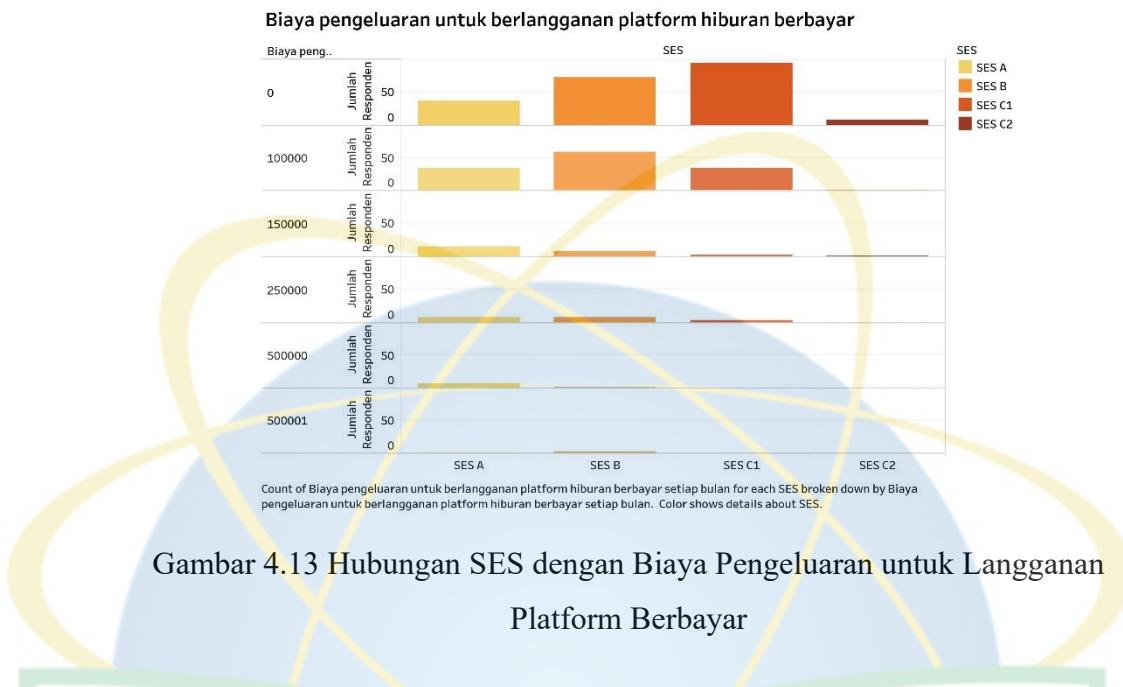
**Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll)**



**Gambar 4.12 Hubungan SES dengan Biaya Pengeluaran untuk Kebutuhan Pribadi**

#### **4.10.3.7 Biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan**

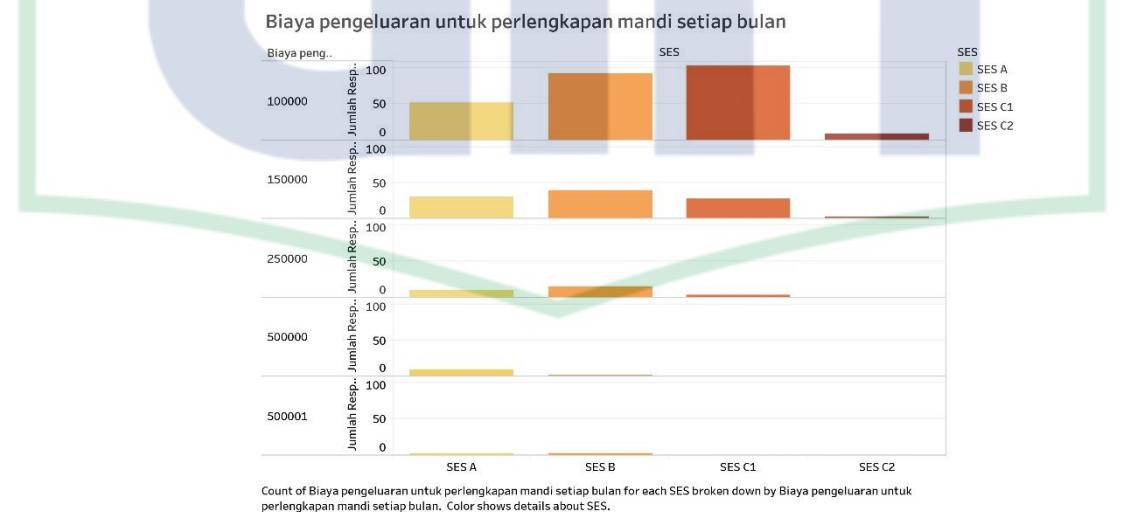
Biaya pengeluaran untuk berlangganan platform hiburan berbayar setiap bulan berdasarkan kategori SES pada setiap kategori didominasi oleh mahasiswa/i yang tidak berlangganan platform hiburan berbayar. Adapun mereka yang berlangganan paling banyak mengelurkan biaya kurang dari Rp.100.000 setiap bulannya pada masing-masing kategori SES.



Gambar 4.13 Hubungan SES dengan Biaya Pengeluaran untuk Langganan Platform Berbayar

#### 4.10.3.8 Biaya pengeluaran untuk perlengkapan mandi setiap bulan

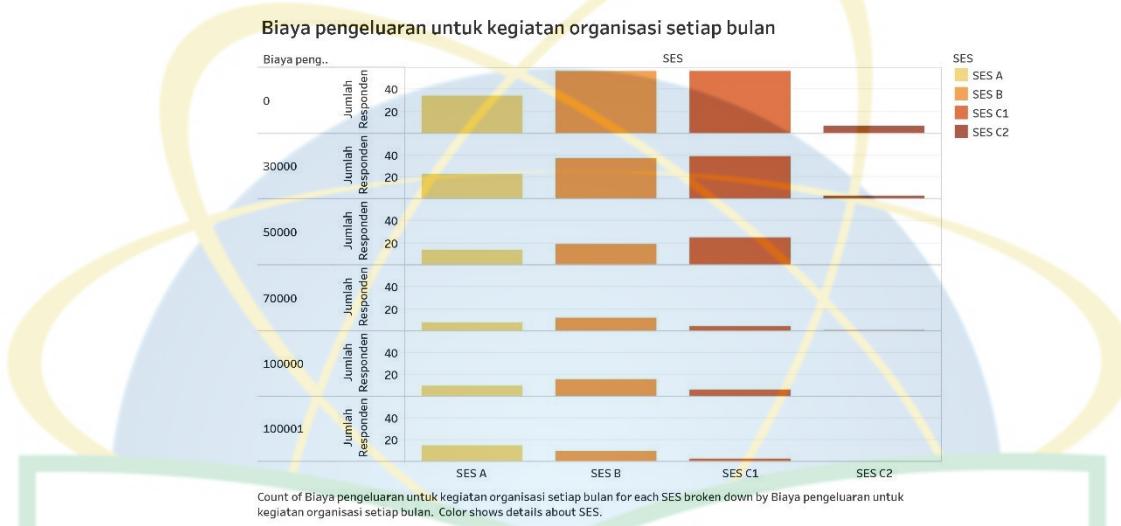
Pada biaya pengeluaran alat mandi didominasi mahasiswa/i yang mengeluarkan biaya kurang dari Rp.100.000 untuk biaya perlengkapan mandi. Hal tersebut bisa disebabkan karena kebanyakan dari mereka tinggal dirumah pribadi sehingga tidak perlu mengeluarkan banyak biaya untuk perlengkapan mandi.



Gambar 4.14 Hubungan SES dengan Biaya Pengeluaran untuk Perlengkapan Mandi

#### 4.10.3.9 Biaya pengeluaran untuk kegiatan organisasi setiap bulan

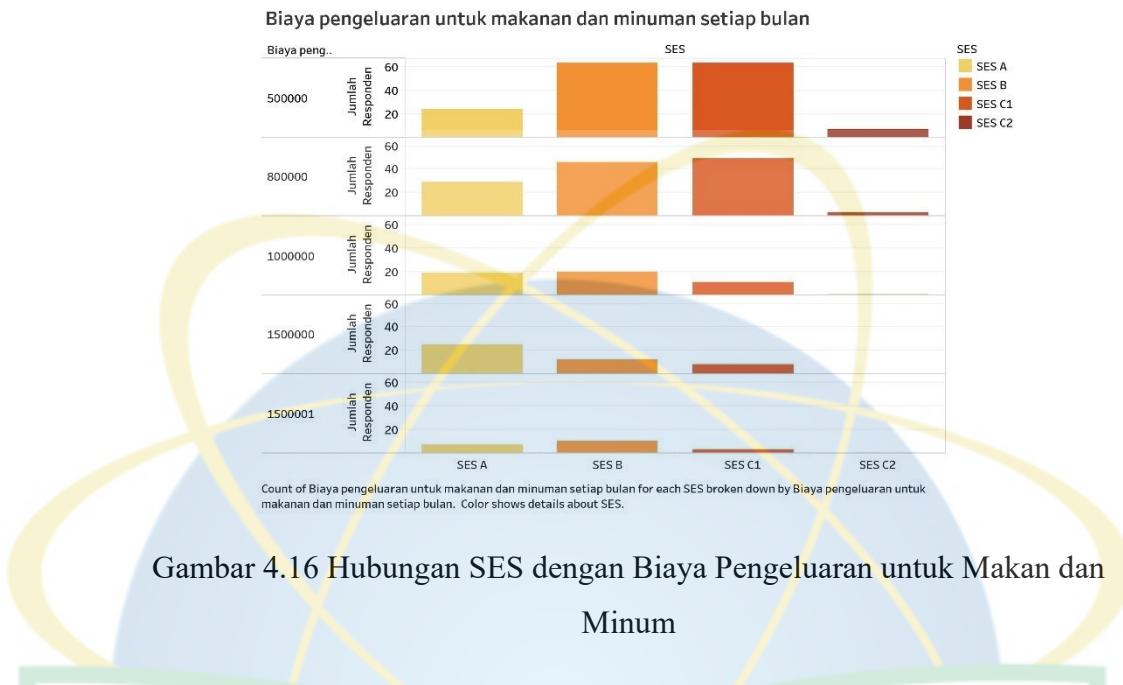
Pada data yang ada pada gambar 4.15 dibawah dapat terlihat mahasiswa pada setiap kategori kebanyakan mengeluarkan biaya Rp.0 untuk organisasi. Hal tersebut juga dapat disimpulkan kebanyakan dari mereka tidak mengikuti organisasi.



Gambar 4.15 Hubungan SES dengan Biaya Pengeluaran untuk Organisasi

#### 4.10.3.10 Biaya pengeluaran untuk makanan dan minuman setiap bulan

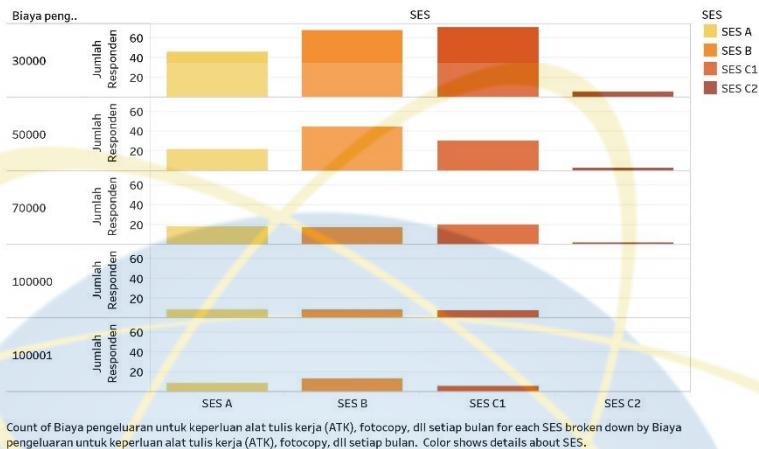
Biaya pengeluaran pokok untuk kebutuhan makan dan minum setiap bulan pada kategori SES A, SES B, SES C1, SES C2 kebanyakan kurang dari Rp.500.000 dan biaya pengeluaran terbesar mahasiswa/i untuk makan dan minum setiap bulannya pada gambar 4.16 dibawah yaitu lebih dari Rp.1.500.000.



#### 4.10.3.11 Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan

Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan kebanyakan kurang dari Rp.30.000 perbulannya. Namun, beberapa data juga menunjukkan bahwa beberapa mahasiswa/i mengeluarkan lebih dari Rp.30.000 untuk keperluan tersebut.

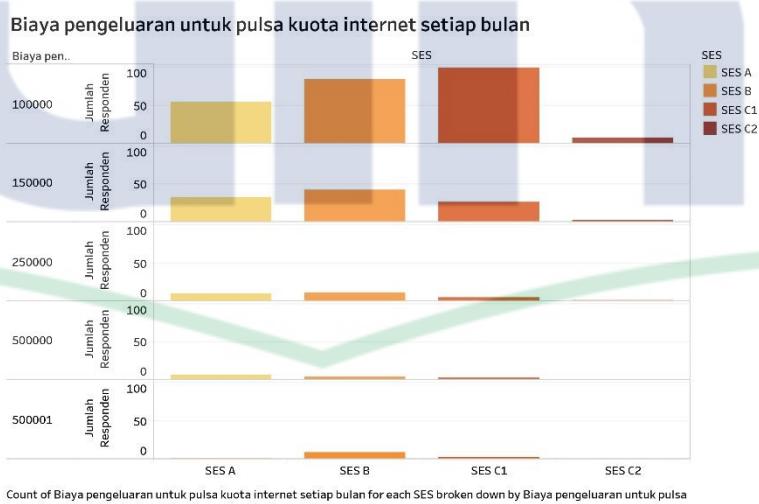
**Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan**



Gambar 4.17 Hubungan SES dengan Biaya Pengeluaran untuk ATK

#### 4.10.3.12 Biaya pengeluaran untuk pulsa kuota internet setiap bulan

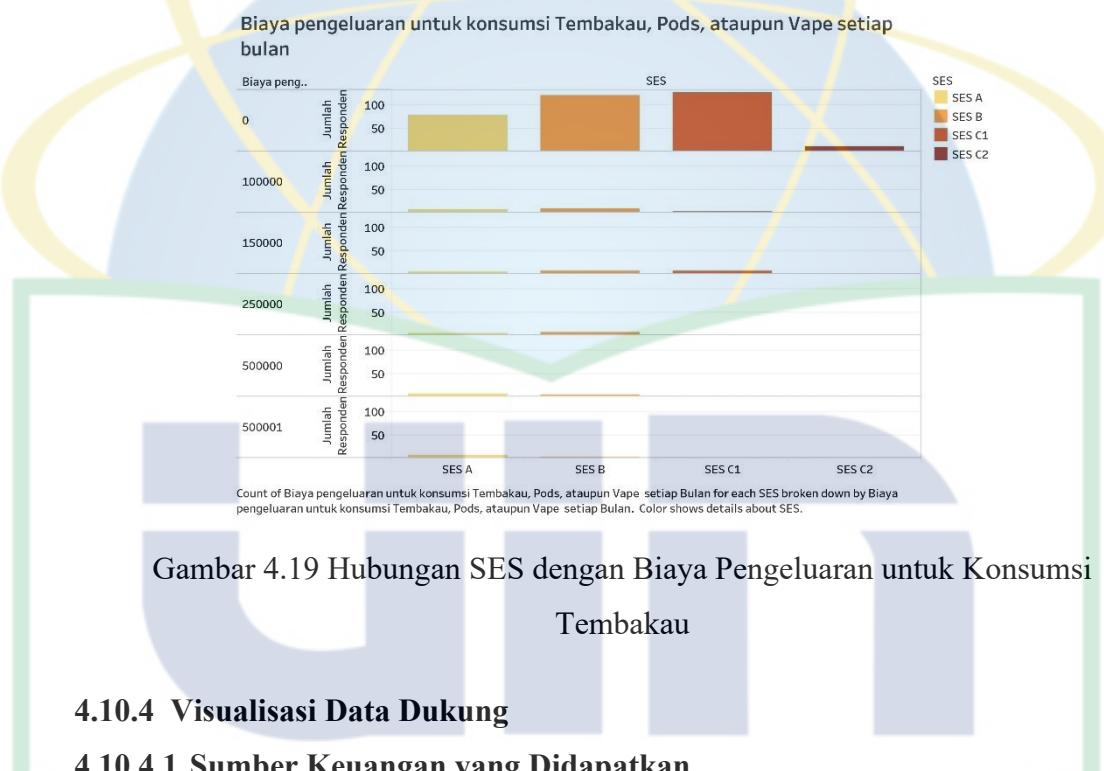
Biaya pengeluaran untuk pulsa kuota internet setiap bulan pada kategori SES A, SES B, SES C1, SES C2 paling banyak mengeluarkan biaya kurang dari Rp.100.000 data pada gambar 4.18. Berdasarkan pengeluaran biaya untuk pulsa kuota internet ini menyesuaikan kebutuhan masing-masing mahasiswa/i.



Gambar 4.18 Hubungan SES dengan Biaya Pengeluaran untuk Pulsa Internet

#### **4.10.3.13 Biaya Pengeluaran untuk Konsumsi Tembakau, Pods, ataupun Vape Setiap Bulan**

Berdasarkan dari 400 responden pada penelitian ini yang mewakili seluruh mahasiswa/i UIN Jakarta pada setiap kategori SES mengeluarkan biaya Rp.0 untuk keperluan konsumsi Tembakau, Pods, ataupun Vape setiap bulan atau tidak mengkonsumsi barang-barang tersebut. Sehingga bagi mereka yang mengkonsumsi tembakau, pods, ataupun vape pada setiap bulannya mengeluarkan biaya Rp.100.000 sampai dengan lebih dari Rp.500.000 dari setiap kategori SES.

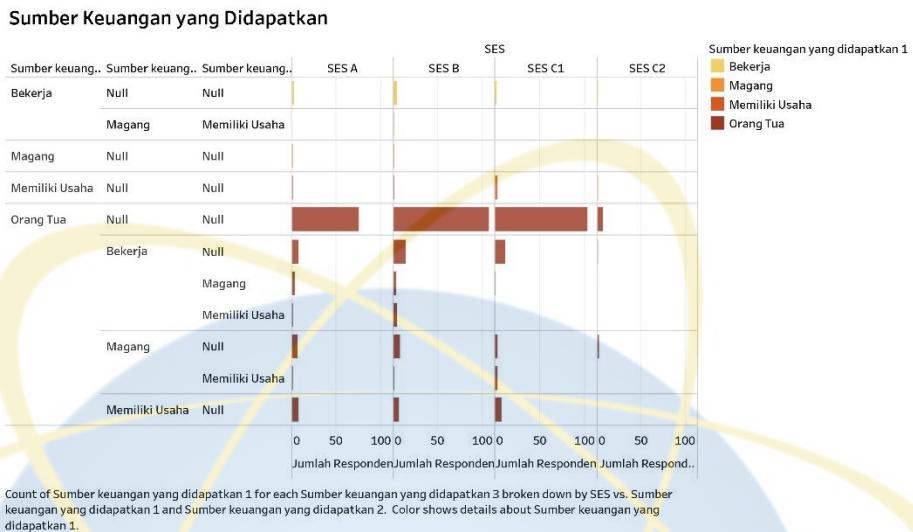


Gambar 4.19 Hubungan SES dengan Biaya Pengeluaran untuk Konsumsi Tembakau

#### **4.10.4 Visualisasi Data Dukung**

##### **4.10.4.1 Sumber Keuangan yang Didapatkan**

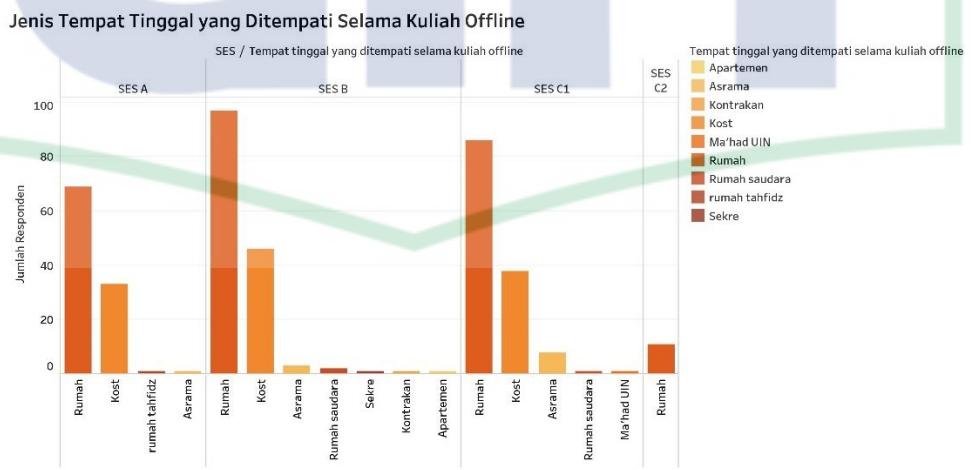
Pada setiap kategori SES mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 paling banyak mendapatkan uang yang berasal dari orang tua. Namun, juga ada beberapa dari mereka yang bekerja, magang, maupun memiliki usaha sendiri untuk mendapatkan uang dan memenuhi kebutuhan mereka setiap bulannya.



Gambar 4.20 Hubungan SES dengan Sumber Keuangan

#### 4.10.4.2 Tempat Tinggal yang Ditempati Selama Kuliah Offline

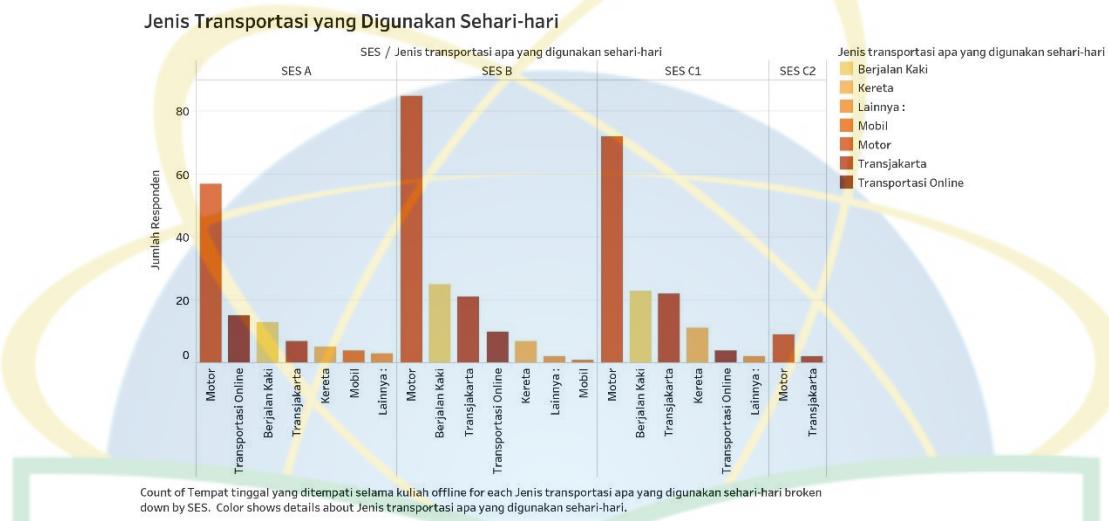
Tempat tinggal yang dipilih selama kuliah *offline* berdasarkan kategori lebih banyak yang bertempat tinggal dirumah. Pada kategori SES B, SES C1, SES C1 juga banyak yang memilih untuk bertempat tinggal di kost, asrama, dsb yang ditempati selama kuliah *offline*. Maka dari 400 responden tersebut dapat dikatakan bahwa mahasiswa/i UIN Jakarta lebih banyak yang berdomisili di Jabodetabek sehingga bertempat tinggal dirumah masing-masing.



Gambar 4.21 Hubungan SES dengan Tempat Tinggal

#### 4.10.4.3 Transportasi yang digunakan sehari-hari

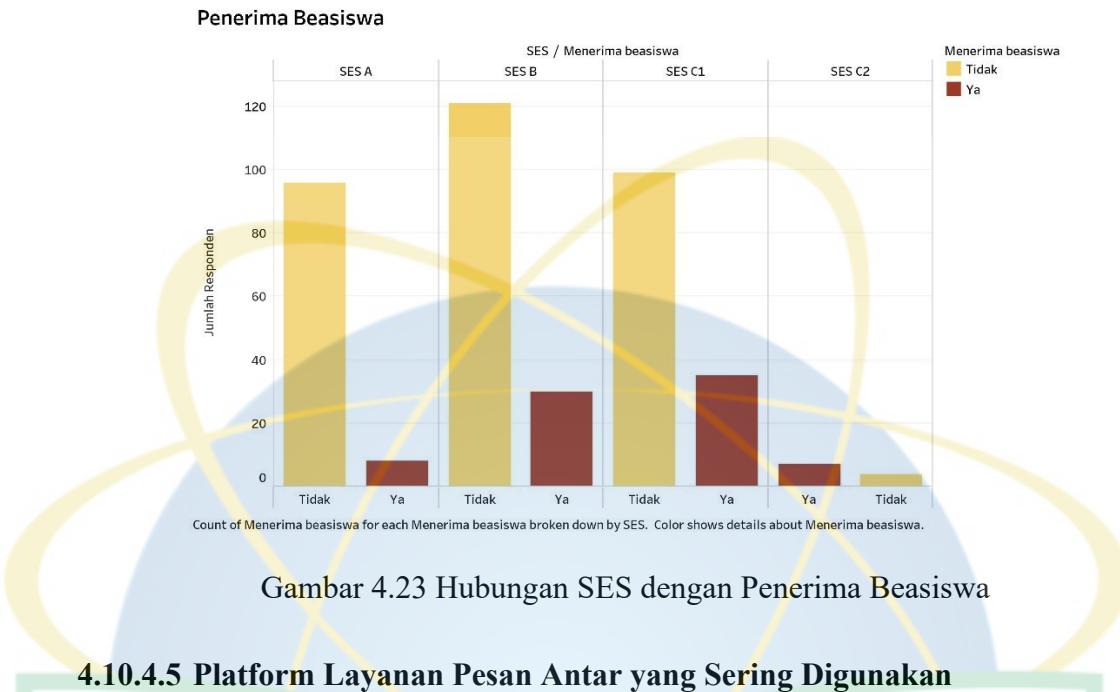
Pada data yang ditunjukkan pada gambar 4.22 menunjukkan bahwa pada semua kategori SES kebanyakan mahasiswa menggunakan sepeda motor sebagai transportasi sehari-hari. Namun juga tidak sedikit yang menggunakan transportasi lainnya seperti berjalan kaki, kereta, mobil, transjakarta, dan transportasi online.



Gambar 4.22 Hubungan SES dengan Transportasi yang Digunakan

#### 4.10.4.4 Menerima beasiswa

Menurut data visualisasi pada gambar 4.23 pada kategori SES A, B, dan C1 lebih banyak yang tidak menerima beasiswa, sementara pada kategori SES C2 penerima beasiswa lebih banyak dibanding yang tidak menerima.



Gambar 4.23 Hubungan SES dengan Penerima Beasiswa

#### 4.10.4.5 Platform Layanan Pesan Antar yang Sering Digunakan

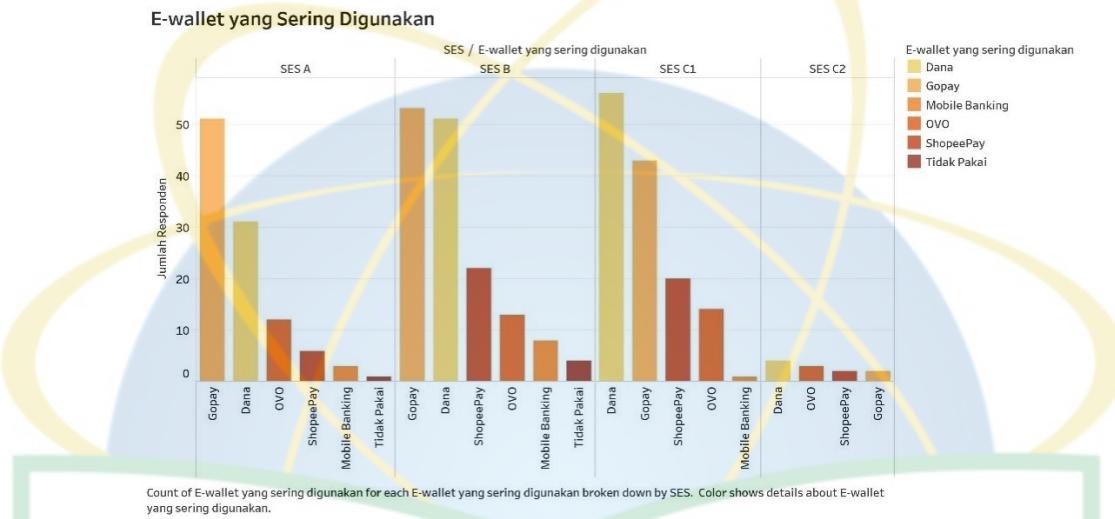
Selain itu juga dapat dilihat pada gambar 4.24 yang menunjukkan bahwa dari berbagai kategori SES yang ada memiliki kesamaan dimana GoFood merupakan platform layanan pesan antar yang paling sering digunakan dikalangan mahasiswa/i UIN Jakarta tahun angkatan 2019-2022. Disusul oleh ShopeeFood pada peringkat kedua dan yang terakhir GrabFood.



Gambar 4.24 Hubungan SES dengan Platform Layanan Pesan Antar

#### 4.10.4.6 *E-wallet* yang Sering Digunakan

Dana merupakan salah satu *e-wallet* yang paling banyak digunakan oleh mahasiswa/i UIN Jakarta dari semua kategori SES. Pada gambar 4.25 dapat dilihat bahwa Gopay sebagai *e-wallet* juga memiliki pengguna yang cukup banyak untuk melakukan transaksi melalui *e-wallet* tersebut.

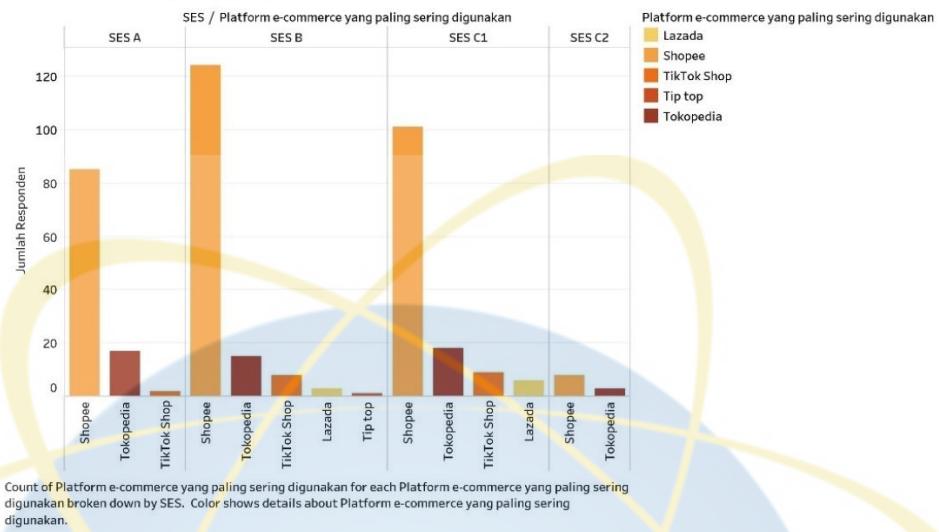


Gambar 4.25 Hubungan SES dengan *E-wallet* yang Digunakan

#### 4.10.4.7 Platform *E-commerce* yang Sering Digunakan

Pada penggunaan platform *e-commerce* pada keempat kategori SES, Shopee merupakan *e-commerce* yang paling banyak digunakan oleh mahasiswa/i UIN Jakarta tahun angkatan 2019-2022. Kemudian, juga terdapat Tokopedia yang menjadi peringkat kedua sebagai platform *e-commerce* yang paling sering digunakan.

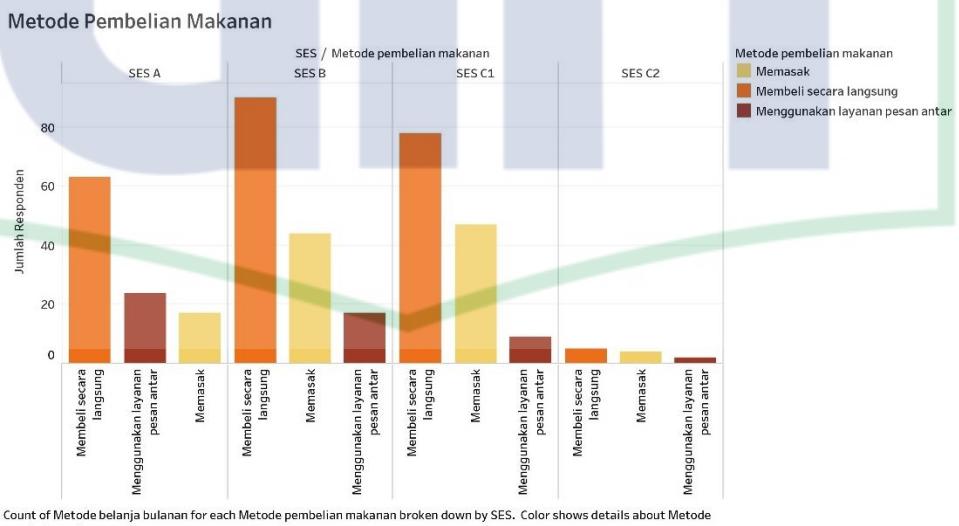
#### E-commerce yang Sering Digunakan



Gambar 4.26 Hubungan SES dengan *E-commerce* yang Digunakan

#### 4.10.4.8 Metode Pembelian Makanan

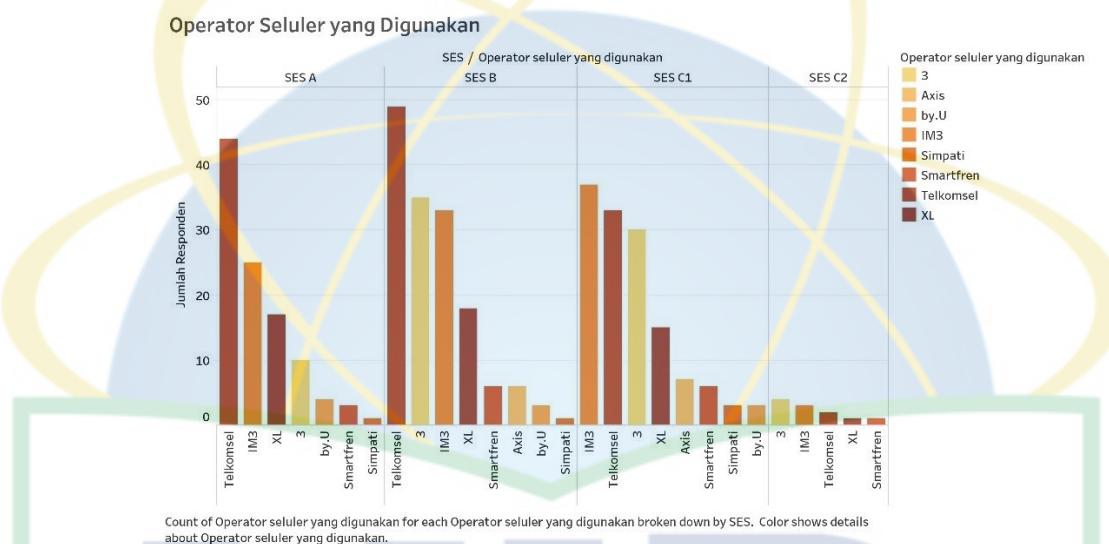
Pada setiap kategori membeli makan secara langsung banyak dilakukan oleh mahasiswa/i UIN Jakarta tahun angkatan 2019-2022. Namun, pada kategori SES A membeli makan menggunakan layanan pesan antar juga banyak digunakan pada kategori ini yang mana pada ketiga kategori lainnya lebih memilih memasak daripada menggunakan layanan pesan antar.



Gambar 4.27 Hubungan SES dengan Metode Pembelian Makanan

#### 4.10.4.9 Operator Seluler yang Digunakan

Secara keseluruhan pada kategori SES A, B, C1, dan C2 operator seluler Telkomsel, IM3, 3, dan XL merupakan operator seluler yang paling banyak digunakan. Pada kategori SES A dan B Telkomsel paling banyak penggunananya, SES C1 lebih banyak yang menggunakan IM3, sedangkan SES C2 3 yang paling banyak digunakan.

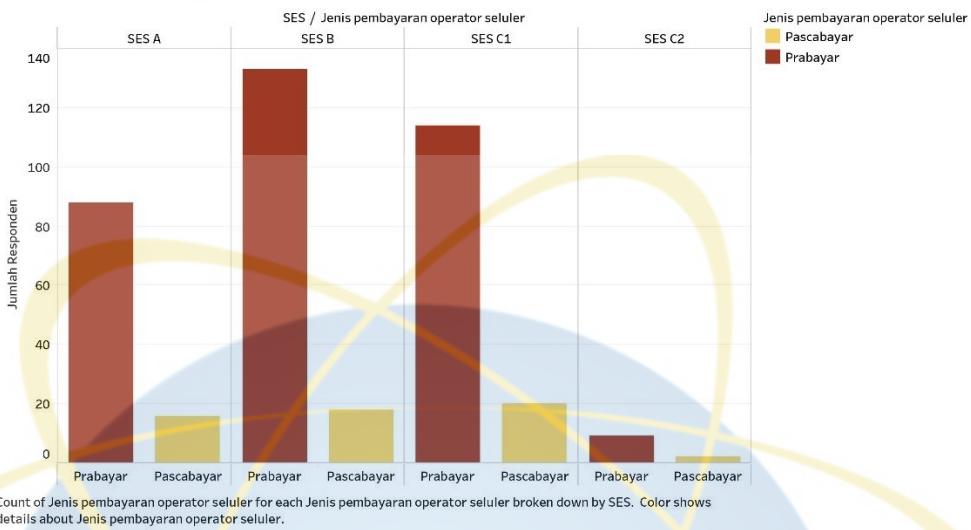


Gambar 4.28 Hubungan SES dengan Operator Seluler yang Digunakan

#### 4.10.4.10 Jenis Pembayaran Operator Seluler

Jenis pembayaran yang dilakukan pada operator seluler pun pada semua kategori SES lebih memilih dengan melakukan isi ulang pulsa atau prabayar dibandingkan dengan pascabayar. Hal tersebut merupakan salah satu cara dimana para mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 membayar atau mengisi pulsa sesuai kebutuhan mereka dibandingkan membayar dengan nominal tetap atau lebih dengan pembayaran setelah pemakaian.

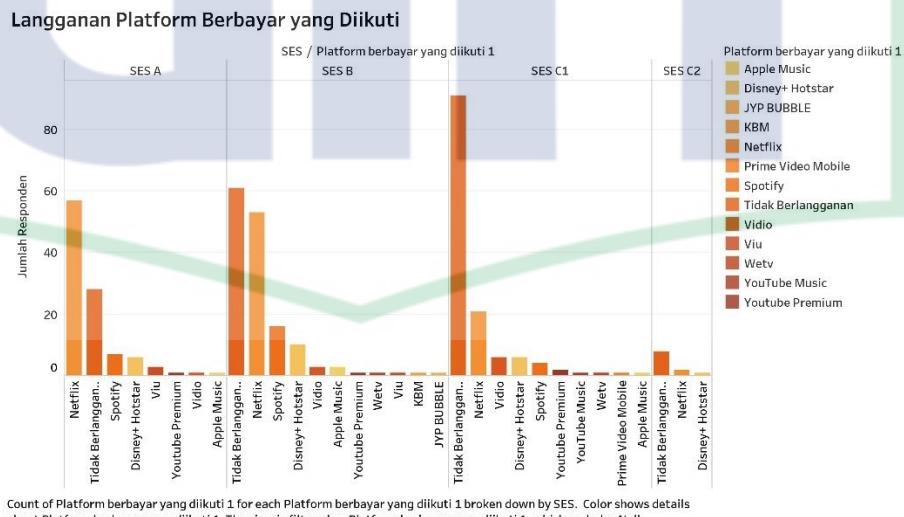
Jenis Pembayaran Operator Seluler



Gambar 4.29 Hubungan SES dengan Jenis Pembayaran Operator Seluler

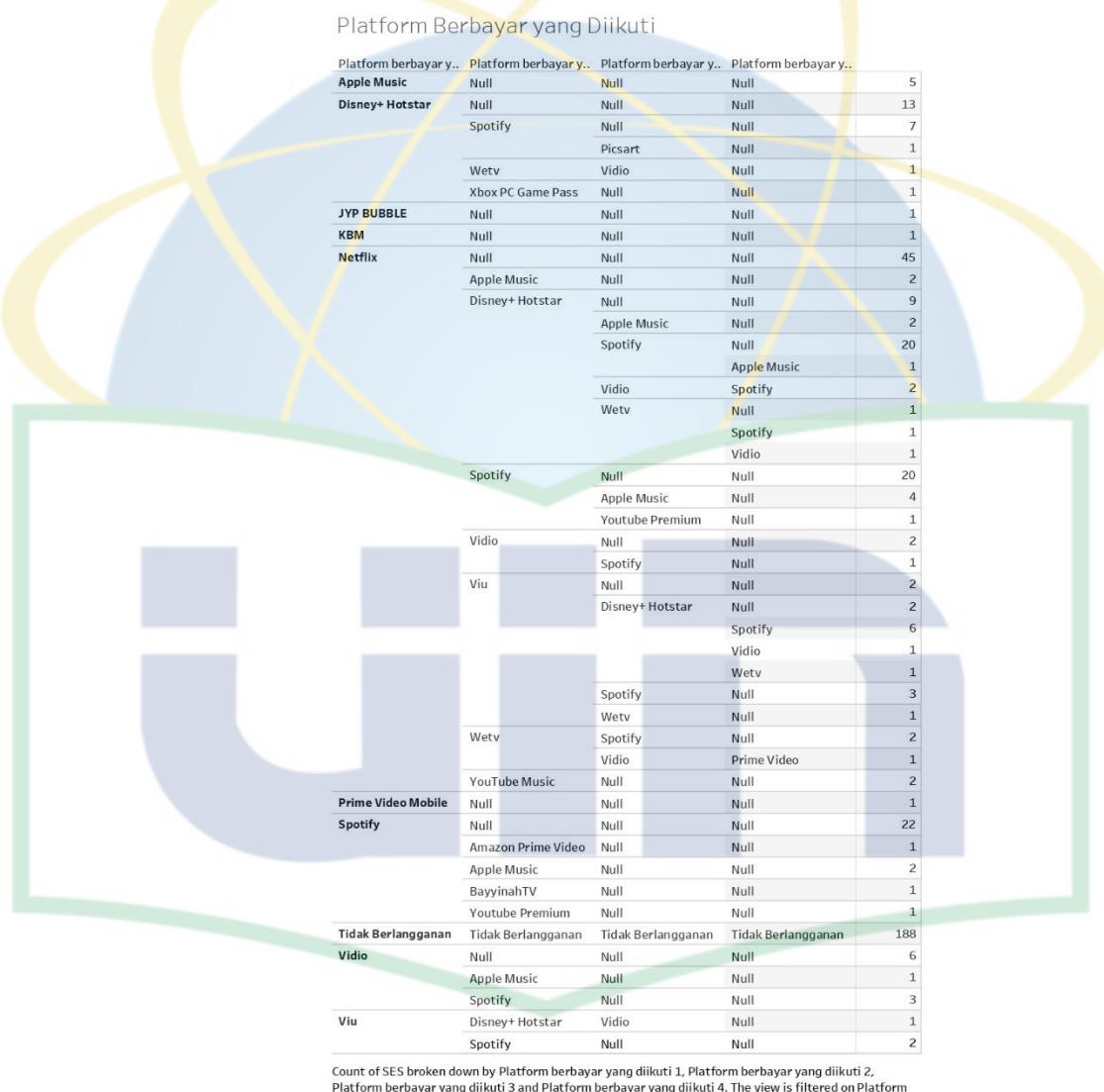
#### 4.10.4.11 Langganan Platform Berbayar yang Diikuti

Berdasarkan 400 responden lebih banyak yang tidak berlangganan daripada yang berlangganan platform berbayar. Namun, bagi mereka yang berlangganan platform berbayar, pada setiap kategori SES, Netflix merupakan platform berbayar yang paling banyak diikuti oleh mahasiswa/i UIN Jakarta tahun angkatan 2019-2022



Gambar 4.30 Hubungan SES dengan Langganan Platform Berbayar

Berdasarkan visualisasi data dukung ini juga dapat dilihat beberapa kombinasi platform berbayar yang diikuti seperti pada gambar 4.31. Mahasiswa/i yang berlangganan Netflix saja paling banyak diikuti. Kemudian, yang hanya berlangganan spotify saja dan pada urutan ketiga paling banyak dari mereka mengikuti Netflix; Disney+ Hotstar; Spotify, Disney Hotstar+; Spotify, dan juga yang hanya berlangganan Spotify saja.



### Platform Berbayar yang Diikuti

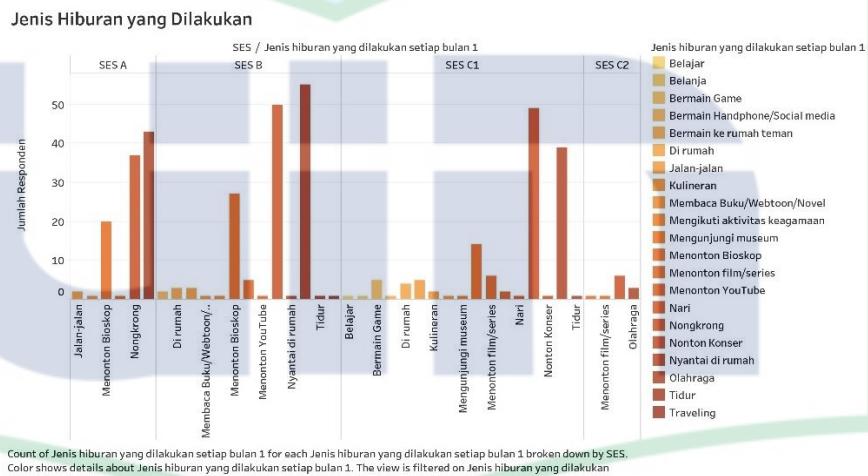
Platform berbayar y..	Platform berbayar y..	Platform berbayar y..	Platform berbayar y..	
Viu	Wetv	Vidio	Spotify	1
<b>Wetv</b>	Spotify	Null	Null	1
		Apple Music	Null	1
YouTube Music	Null	Null	Null	1
Youtube Premium	Null	Null	Null	4

Count of SES broken down by Platform berbayar yang diikuti 1, Platform berbayar yang diikuti 2, Platform berbayar yang diikuti 3 and Platform berbayar yang diikuti 4. The view is filtered on Platform berbayar yang diikuti 1, which excludes Null.

Gambar 4.31 Platform Berbayar yang Diikuti

#### 4.10.4.12 Jenis Hiburan yang Dilakukan

Terdapat 3 kegiatan hiburan yang paling banyak dilakukan oleh mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 yaitu nongkrong, olahraga, dan menonton bioskop. Keempat kategori SES tersebut memiliki tingkatan kegiatan yang sama yaitu nongkorong merupakan kegiatan yang paling banyak dilakukan oleh seluruh responden, yang kedua terdapat olahraga, dan yang terakhir yaitu menonton bioskop.



Gambar 4.32 Hubungan SES dengan Jenis Hiburan

Berdasarkan visualisasi data dukung hiburan yang dilakukan juga dapat dilihat beberapa kombinasi kegiatan yang dilakukan seperti pada gambar 4.33. Mahasiswa/i yang nongkrong paling banyak dilakukan. Kemudian, yang olahraga saja maupun olahraga dan nongkrong pada urutan kedua. Ketiga yaitu mereka yang melakukan olahraga, menonton bioskop dan nongkrong. Selain itu juga banyak dari

mereka yang melakukan menonton bioskop dan nongkrong, maupun mengunjungi museum.

#### Hiburan yang Dilakukan

Jenis hiburan yang dilakukan..	Jenis hiburan yang dilakukan..	Jenis hiburan yang dilakukan..	
<b>Belajar</b>	Null	Null	1
<b>Belanja</b>	Null	Null	1
<b>Berman Game</b>	Null	Null	4
	Membaca Buku/Webtoon/N..	Menonton film/series	1
<b>Berman Handphone/Social media</b>	Null	Null	1
	Menonton televisi	Menonton film/series	1
<b>Bermain ke rumah teman</b>	Null	Null	1
<b>Di rumah</b>	Null	Null	6
	Jalan-jalan	Null	1
<b>Jalan-jalan</b>	Null	Null	6
	Kulineran	Null	2
	Mengunjungi Event Jepang	Null	1
	Menonton Youtube	Null	1
<b>Kulineran</b>	Null	Null	2
<b>Membaca Buku/Webtoon/N..</b>	Null	Null	3
<b>Mengikuti aktivitas keagam..</b>	Null	Null	1
<b>Mengunjungi museum</b>	Null	Null	1
<b>Menonton Bioskop</b>	Null	Null	23
	Bermain Game	Null	1
	Fangirling	Null	1
	Kajian Islam	Null	1
	Kulineran	Jalan-jalan	1
	Nongkrong	Null	34
		Kulineran	1
<b>Menonton film/series</b>	Null	Null	13
<b>Menonton YouTube</b>	Null	Null	2
	Membaca Buku/Webtoon/N..	Null	1
<b>Nari</b>	Embroidery	Null	1
<b>Nongkrong</b>	Null	Null	127
	Bermain Game	Menonton film/series	1
	Jalan-jalan	Null	3
	Kulineran	Null	1
	Membaca Buku/Webtoon/N..	Bermain Handphone/Social ..	1
	Mendengarkan Musik	Null	1
	Menonton film/series	Null	5
		Jalan-jalan	1
	Menonton YouTube	Null	1
	Traveling	Null	1
<b>Nonton Konser</b>	Null	Null	1
<b>Nyantai di rumah</b>	Null	Null	1
<b>Olahraga</b>	Null	Null	42
	Bermain Handphone/Social ..	Null	1
	Bermain ke rumah teman	Null	1
	Kulineran	Null	1

#### Hiburan yang Dilakukan

Jenis hiburan yang dilakukan	Jenis hiburan yang dilakukan	Jenis hiburan yang dilakukan	
Olahraga	Main dengan wewehan pelihara..	Null	1
	Membaca Buku/Webtoon/N..	Null	1
	Mengunjungi perpustakaan	Null	1
	Menonton Bioskop	Null	4
		Bерmain Game	1
		Bерmain ke rumah teman	1
		Mendengarkan Musik	1
		Nongkrong	39
	Menonton film/series	Null	2
	Nongkrong	Нет	42
Tidur	Tidur	Menonton Youtube	1
Творческое	Null	Null	2
	Null	Null	3

Gambar 4.33 Kegiatan Hiburan yang Dilakukan

#### 4.11 Interpretasi Hasil

Berdasarkan hasil prediksi gaya hidup mahasiswa/i UIN Jakarta menurut kategori SES nya, dengan menggunakan algoritma CatBoost diketahui bahwa golongan UKT sangat menentukan gaya hidup setiap mahasiswa/i UIN Jakarta berdasarkan tingkat kategori SES setiap mahasiswa/i. Hal ini juga dapat membuktikan bahwa golongan UKT yang didapatkan oleh setiap mahasiswa/i UIN Jakarta sudah sesuai dengan kemampuan ekonomi mereka. Golongan UKT yang berbeda-beda dari setiap mahasiswa/i UIN Jakarta juga dapat dilihat berdasarkan pekerjaan orang tuanya, yang mana pekerjaan tersebut juga berkaitan dengan

penghasilan orang tuanya. Sehingga, pekerjaan orang tua berpengaruh dan menentukan dimana tingkat SES mahasiswa/i tersebut.

Selain itu, setiap fakultas di UIN Jakarta juga memiliki peranan penting terhadap tingkat SES, yang mana terdapat kategori SES tertentu yang mendominasi di fakultas tersebut. Dengan kata lain, tingkat SES mahasiswa/i UIN Jakarta dipengaruhi oleh lingkungannya yaitu fakultas mereka masing-masing.

Selain dilihat dari golongan UKT, pekerjaan orang tua, dan fakultas, yang merupakan data primer dari setiap mahasiswa/i UIN Jakarta, dapat dilihat melalui biaya pengeluaran rutin yang dikeluarkan setiap bulannya sebagai data sekunder yang merupakan salah satu dari variabel SES. Biaya pengeluaran yang dikeluarkan setiap bulan ini merupakan faktor terpenting yang mempengaruhi gaya hidup mahasiswa/i UIN Jakarta sesuai dengan tingkat SES mereka.

Biaya pengeluaran untuk hiburan setiap bulan merupakan biaya pengeluaran yang paling mempengaruhi hasil prediksi SES mahasiswa/i UIN Jakarta dibandingkan biaya pengeluaran rutin secara keseluruhan maupun biaya kebutuhan primer lainnya. Hal tersebut menunjukkan bahwa biaya pengeluaran yang dikeluarkan oleh setiap mahasiswa/i untuk melakukan sebuah kegiatan sebagai hiburan setiap bulannya memiliki pengaruh yang signifikan berdasarkan besarnya biaya yaitu sesuai dengan tingkat SES masing-masing.

Kegiatan yang dilakukan oleh mahasiswa/i UIN Jakarta untuk hiburan juga dapat dilihat pada visualisasi data dukung yang mana paling banyak dari mereka memilih untuk nongkrong, olahraga, dan menonton bioskop sebagai kegiatan hiburannya. Hal tersebut menunjukkan keterkaitan dari gaya hidup seseorang yang dinilai berdasarkan pola hidupnya saat melakukan kegiatan dan berinteraksi di lingkungannya yang mana didukung oleh pendapat dari Philip Kotler dan Kevin Lane Keller (2016:87) dalam bukunya yang berjudul *Marketing Management*.

Biaya pengeluaran untuk transportasi setiap bulan yang dikeluarkan oleh mahasiswa/i UIN Jakarta berdasarkan kategorinya memiliki pengaruh dan perbedaan yang cukup signifikan pada setiap kategori SES. Pengaruh besar atau kecilnya biaya pengeluaran transportasi ini juga bergantung dengan tempat tinggal yang mereka tempati dan juga jenis transportasi yang mereka gunakan.

Selanjutnya yaitu biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll) yang merupakan salah satu bagian dari biaya untuk kebutuhan primer yang berpengaruh dalam melakukan prediksi kategori SES, namun biaya pengeluaran ini juga dapat dikatakan sebagai pengeluaran kebutuhan sekunder. Biaya pengeluaran ini termasuk kebutuhan primer dan sekunder karena responden yang di dominasi oleh perempuan dan di dukung oleh pendapat Jeff Fromm dalam artikelnya yang berjudul "*As Gen Z's Buying Power Grows, Businesses Must Adapt Their Market*", yang menjelaskan bahwa berkembangnya berbagai produk yang cepat dapat menumbuhkan sifat konsumtif dengan memiliki barang atau produk yang digunakan dari *influencers* maupun lingkungan sekitarnya. Hal tersebut juga didukung oleh (Rasyid, 2022) yang menyatakan bahwa perilaku konsumtif dapat diakibatkan oleh kondisi FoMO dimana mereka membeli barang yang tidak sesuai dengan kebutuhan.

Biaya pengeluaran untuk berlangganan platform hiburan berbayar juga merupakan biaya kebutuhan sekunder lebih berpengaruh daripada biaya kebutuhan primer perlengkapan mandi setiap bulan dan biaya pengeluaran untuk makanan dan minuman setiap bulan. Meskipun, lebih banyak yang tidak berlangganan platform hiburan berbayar, mahasiswa/i UIN Jakarta mengeluarkan biaya yang memiliki perbedaan yang cukup signifikan berdasarkan kategori SES nya. Biaya pengeluaran platform hiburan berbayar yang dibayarkan memiliki pengaruh lebih ini juga didukung oleh beberapa mahasiswa/i UIN Jakarta yang berlangganan lebih dari satu platform sehingga harus membayar dengan biaya yang lebih besar.

Biaya pengeluaran untuk makanan dan minuman setiap bulan sebagai kebutuhan primer memiliki pengaruh yang lebih rendah dibanding kebutuhan sekunder yang sudah dijelaskan sebelumnya. Biaya pengeluaran ini dapat dipengaruhi oleh tempat tinggal mahasiswa/i UIN Jakarta yang di dominasi bertempat tinggal dirumah. Oleh karena itu, perbedaan biaya pengeluaran untuk makan dan minum yang tidak signifikan bedanya menyebabkan biaya ini tidak memiliki pengaruh yang besar saat prediksi SES mahasiswa/i UIN Jakarta.

Biaya pengeluaran untuk kegiatan organisasi, biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll, biaya pengeluaran untuk pulsa kuota

internet, biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap bulan merupakan kebutuhan skunder yang mana hal tersebut menjelaskan bahwa tidak semua mahasiswa/i UIN Jakarta mengikuti organisasi, menggunakan operator dan kebutuhan kuota yang sama, serta tidak semua dari mereka yang mengkonsumsi tembakau maupun rokok elektrik. Ketiga biaya pengeluaran ini tidak berpengaruh banyak terhadap prediksi SES karena kepentingan biaya pengeluaran ketiga hal tersebut bukanlah biaya yang pasti dikeluarkan oleh mereka setiap bulannya atau bisa dikurangi.

Adapun biaya kebutuhan primer meliputi biaya pengeluaran untuk makan dan minum, transportasi, perlengkapan mandi, dan pulsa kuota internet setiap bulan. Sedangkan biaya kebutuhan sekunder antara lain, biaya pengeluaran untuk hiburan, kebutuhan pribadi setiap bulan (*Skincare*, kosmetik, perawatan, dll), platform hiburan berbayar, kegiatan organisasi, keperluan alat tulis kerja (ATK), fotocopy, dll, dan konsumsi tembakau, pods, ataupun vape setiap bulan.

Berdasarkan penjelasan diatas dapat diketahui kelas sosial yang merupakan faktor eksternal dari gaya hidup ini dipengaruhi oleh tingkat ekonomi dan pola konsumsi mahasiswa/i UIN Jakarta sesuai dengan tingkat SES masing - masing. Pernyataan tersebut juga didukung oleh (Saputra, 2020) yang mengungkapkan bahwa pola konsumsi serta pilihannya dalam menggunakan waktu dan uangnya yang mempengaruhi gaya hidup mereka.

Oleh karena itu, hasil penelitian ini dapat digunakan sebagai bahan literasi keuangan bagi mahasiswa/i UIN Jakarta agar lebih baik mengatur biaya pengeluarannya dengan meminimalkan pembelian produk yang tidak sesuai kebutuhan dengan mementingkan keinginan untuk mengikuti tren. Selain itu juga, bagi pemasar di lingkungan sekitar UIN Jakarta dapat menyesuaikan produk berdasarkan personalisasi mahasiswa/i UIN Jakarta sesuai tingkat ekonomi target pasar yang diinginkan. Hasil penelitian ini juga selaras dengan hasil penelitian (Putri, 2019) yang menyatakan bahwa biaya pengeluaran konsumsi rumah tangga untuk memenuhi kebutuhan hidup telah bergeser untuk kebutuhan sekunder yang dilihat dari biaya pengeluaran mengikuti keinginan dan dari kondisi lingkungan sekitar.



## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil pembahasan yang telah dipaparkan sebelumnya, maka dapat ditarik kesimpulan yaitu sebagai berikut:

1. Penelitian ini berhasil melakukan prediksi kategori SES dengan menggunakan pemodelan algoritma CatBoost dengan hasil akurasi 85,94%.
2. Pemodelan CatBoost secara akurat memprediksi 4 kelas SES yaitu SES A, SES B, SES C1 dengan tingkat *precision* dan *recall* yang cukup tinggi berdasarkan F1-score dan SES C2 yang cukup rendah.
3. Kurva *Area Under ROC Curve (AUC)* sebagai metrik evaluasi kinerja model memberikan hasil klasifikasi sangat baik dengan memperoleh nilai nilai SES A (0.95), SES B (0.95), SES C1 (0.94), dan SES C2 (0.99).
4. Berdasarkan tingkat kepentingannya, golongan UKT, pekerjaan orang tua, dan fakultas dengan tingkat kepentingan tertinggi pada saat memprediksi kategori SES dari 14 fitur yang telah ditentukan berdasarkan hasil korelasi dengan skor SES.
5. Penelitian ini berhasil mengetahui faktor-faktor yang mempengaruhi meningkatnya gaya hidup mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 berdasarkan hasil prediksi SES menggunakan pemodelan algoritma CatBoost. Diketahui bahwa biaya pengeluaran yang mempengaruhi meningkatnya gaya hidup mahasiswa/i UIN Jakarta tahun angkatan 2019-2022 adalah biaya pengeluaran untuk hiburan yang biaya pengeluarannya tergantung pada tingkat kategori SES masing-masing.
6. Mahasiswa/i dapat menekan daya konsumtif yang berlebih sesuai dengan tingkat kemampuan ekonominya dengan meminimalisir pembelian barang atau jasa dengan tidak selalu mengikuti tren yang dilakukan hanya untuk memenuhi keinginan tanpa memikirkan kepentingan dari fungsi barang yang dibeli dalam jangka panjang.
7. Mahasiswa dapat lebih memperhatikan pengelolaan keuangannya yang mana dari hasil penelitian terindikasi bahwa biaya pengeluaran sekunder

yang lebih besar menjadi penyebab meningkatnya gaya hidup mahasiswa/i UIN Jakarta tahun angkatan 2019-2022.

8. Para pemasar dapat menyesuaikan harga berdasarkan biaya pengeluaran rata-rata sehingga dapat dilakukan segmentasi gaya hidup dan personalisasi mahasiswa/i UIN Jakarta.
9. Berdasarkan tiga fitur yang dihasilkan yaitu golongan UKT, pekerjaan orang tua, dan fakultas dari penelitian, didapatkan bahwa kampus dapat memprediksi SES mahasiswa/i nya sesuai dengan nilai tingkat kepentingan fitur yang diperoleh.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk peneliti ingin melakukan penelitian selanjutnya, yaitu:

1. Melanjutkan penelitian ini yaitu menggunakan metode pemodelan yang berbeda ataupun *ensemble learning* untuk mendapatkan hasil akurasi yang lebih baik.
2. Penelitian selanjutnya dapat membuat pertanyaan pendukung yang lebih dikembangkan sebagai fitur-fitur penilaian yang mendukung pemodelan serta hasil prediksi.



## DAFTAR PUSTAKA

- Abdullahi A. Ibrahim, Raheem L. Ridwan, Muhammed M. Muhammed, Rabiat O. Abdulaziz and Ganiyu A. Saheed, “Comparison of the CatBoost Classifier with other Machine Learning Methods” International Journal of Advanced Computer Science and Applications(IJACSA), 11(11), 2020.
- Abdurrahman, S. W., & Oktapiani, S. (2020). Pengaruh Literasi Keuangan dan Lingkungan Sosial Terhadap Perilaku Keuangan Mahasiswa Universitas Teknologi Sumbawa. *Jurnal Ekonomi Dan Bisnis Indonesia*, 05(02), 50–55.
- Adamker, G., Holzer, T., Karakis, I., Amitay, M., Anis, E., Singer, S. R., & Barnett-Itzhaki, Z. (2018). Prediction of Shigellosis outcomes in Israel using machine learning classifiers. *Epidemiology and Infection*, 146(11), 1445–1451.
- Ahyar, H., Maret, U. S., Andriani, H., Sukmana, D. J., Mada, U. G., Hardani, S.Pd., M. S., Nur Hikmatul Auliya, G. C. B., Helmina Andriani, M. S., Fardani, R. A., Ustiawaty, J., Utami, E. F., Sukmana, D. J., & Istiqomah, R. R. (2020). *Buku Metode Penelitian Kualitatif & Kuantitatif* (Issue March).
- Aini, E. N., & Andjarwati, A. L. (2020). Pengaruh Gaya Hidup Konsumtif dan Kualitas Produk Terhadap Keputusan Pembelian. *BISNIS : Jurnal Bisnis Dan Manajemen Islam*, 8(1), 17.
- Aldania, A. N. A., Soleh, A. M., & Notodiputro, K. A. (2023). A Comparative Study of CatBoost and Double Random Forest for Multi-class Classification. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(1), 129-137.
- Andrea, B., Gabriella, H. C., & Tímea, J. (2016). Y and Z generations at workplaces. *Journal of Competitiveness*, 8(3), 90–106.  
<https://doi.org/10.7441/joc.2016.03.06>
- Angela, V., & Paramita, E. L. (2020). Pengaruh Lifestyle Dan Kualitas Produk Terhadap Keputusan Impulse Buying Konsumen Shopee Generasi Z. *Jurnal Ekobis : Ekonomi Bisnis & Manajemen*, 10(2), 248–262.

- Anggraeni, E., & Setiaji, K. (2018). Pengaruh Media Sosial dan Status Ekonomi Orang Tua Terhadap Perilaku Konsumtif Mahasiswa. *Economic Education Analysis Journal*, 7(1), 172–180.
- Anuradha, P., & David, V. K. (2021). Feature Selection and Prediction of Heart diseases using Gradient Boosting Algorithms. In *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021* (pp. 711–717).
- Ardianti, R. (2016). Interpretasi Khalayak Terhadap Gaya Hidup Travelling yang Ditampilkan Melalui Foto dalam Akun Instagram indotravellers. co (*Doctoral dissertation, Universitas Airlangga*).
- Asosiasi Penyelenggara Jasa Internet Indonesia. (2022). Profil Internet Indonesia 2022. *Apji.or.Od, June*, 10. apji.or.id
- Badan Pusat Statistik (BPS). (2019). Data Sosial Ekonomi. *Badan Pusat Statistik Nasional Indonesia*, 5(1), 1–8.
- Barua, S., Gavandi, D., Sangle, P., Shinde, L., & Ramteke, J. (2021). Swindle: Predicting the Probability of Loan Defaults using CatBoost Algorithm. In *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021* (pp. 1710–1715).  
<https://doi.org/10.1109/ICCMC51019.2021.9418277>
- Basrowi. (2005 : 63). Achieved Status. Diambil kembali dari uin-malang.ac.id:  
<http://etheses.uin-malang.ac.id/600/6/10410177%20Bab%202.pdf>
- Breiman, L., 2011. Random Forests. Netherlands: Kluwer Academic Publishers.
- catboost. (2023). *tutorials/python\_tutorial.ipynb* at master · catboost/tutorials. GitHub.  
[https://github.com/catboost/tutorials/blob/master/python\\_tutorial.ipynb](https://github.com/catboost/tutorials/blob/master/python_tutorial.ipynb)
- Catboost.ai (2018). *CatBoost Enables Fast Gradient Boosting on Decision Trees Using GPUs*. <https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus> diakses pada 11 Mei 2023 pukul 20.00

- Chadaga, K., Prabhu, S., Umakanth, S., Bhat, V. K., Sampathila, N., Chadaga, R. P., & Prakasha, K. K. (2021). COVID-19 Mortality Prediction among Patients Using Epidemiological Parameters: An Ensemble Machine Learning Approach. *Engineered Science*, 16, 221–233.  
<https://doi.org/10.30919/es8d579>
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 1-4.
- Deloitte. (2022). Striving for balance, advocating for change. *The Deloitte Global 2022 Gen Z & Millennial Survey*, 1–40.
- Dhananjay, B., & Sivaraman, J. (2021). Analysis and classification of heart rate using CatBoost feature ranking model. *Biomedical Signal Processing and Control*, 68(April), 102610. <https://doi.org/10.1016/j.bspc.2021.102610>
- Dhananjay, B., & Sivaraman, J. (2021). Analysis and classification of heart rate using CatBoost feature ranking model. *Biomedical Signal Processing and Control*, 68, 102610.
- Donni Juni Priansa.2017.Perilaku Konsumen dalam persaingan bisnis kontemporer. Bandung: alfabet
- Ethem, A. (2010). Introduction machine Learning. In រាជ្យវិទ្យាការណ៍ហាវិមាគីសភាខ័ណ្ឌនៅខេត្ត (Vol. 4, Issue 1). <https://www.ptonline.com/articles/how-to-get-better-mfi-results>
- Fei, X., Fang, Y., & Ling, Q. (2020). Discrimination of Excessive Exhaust Emissions of Vehicles based on Catboost Algorithm. In *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020* (pp. 4396–4401). <https://doi.org/10.1109/CCDC49329.2020.9164224>
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.

- Friedman J, Hastie T, Tibshirani R, et al. (2000). "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)." *The annals of statistics*, 28(2), 337–407.
- Friedman JH (2001). "Greedy function approximation: a gradient boosting machine." *Annals of Statistics*, pp. 1189–1232.
- Fromm, J. (2022, October 12). As Gen Z's Buying Power Grows, Businesses Must Adapt Their Marketing. *Forbes*.  
<https://www.forbes.com/sites/jefffromm/2022/07/20/as-gen-zs-buying-power-grows-businesses-must-adapt-their-marketing/>, diakses pada 26 Desember 2022 pukul 13.30.
- Ghoneim, S. 2019. 'Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?'.
- Ghozali, I. (2011). *Multivariate Dengan Program IBM SPSS 19*. Badan Penerbit Universitas Diponegoro.
- Gorunescu, Florin (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer
- Grail Research. 2011. Consumers of Tomorrow Insights and Observations About Generation Z. [http://www.integreon.com/pdf/Blog/Consumers\\_of\\_Tomorrow\\_Insights\\_and\\_Observations\\_About\\_Generation\\_Z\\_246.pdf](http://www.integreon.com/pdf/Blog/Consumers_of_Tomorrow_Insights_and_Observations_About_Generation_Z_246.pdf)
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Hana, K. F. (2019). Minat Beli Online Generasi Milenial: Pengaruh Kepercayaan dan Kualitas Layanan. *BISNIS : Jurnal Bisnis Dan Manajemen Islam*, 7(2), 203.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

- Hawari, A. N. (2023). *Pengaruh Gaya Hidup Terhadap Willingness To Subscribe Pada Applikasi Streaming Film Online Netflix di Kalangan Generasi Z Kota Surabaya Danang Tandyonomanu Abstrak*. 06, 13–19.
- Haynos, A. F., Wang, S. B., Lipson, S., Peterson, C. B., Mitchell, J. E., Halmi, K. A., Agras, W. S., & Crow, S. J. (2021). Machine learning enhances prediction of illness course: A longitudinal study in eating disorders. *Psychological Medicine*, 51(8), 1392–1402.
- Heizer, Jay dan Barry Render., 2011. Operations Management, Buku 1 edisi ke sembilan. Salemba empat : Jakarta.
- Hellen Chou Pratama. Cyber Smart Parenting. (Bandung: Visi Anugerah Indonesia, 2012), hlm. 20.
- Herdianto. 2013, Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan Backpropagation. Medan: Universitas Sumatera Utara
- Hertinjung, W. S., & Karyani, U. (2015). Profil Perilaku dan Korban Bullying di Sekolah Dasar. *The 2nd University Research Coloquium, 2011*, 173–180.
- Hootsuite Inc. (2022). *Digital Trends & Social Media Statistics 2022 October Update*. Digital Trends & Social Media Statistics 2022 October Update; Hootsuite.
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., & Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029–1041.
- Indonesia Data. (2022, March 15). Indonesia Data.
- <https://indonesiadata.id/mengenal-istilah-ses-atau-status-sosial-ekonomi/>, diakses pada 1 Februari 2023 pukul 10.30.
- Indonesia Data. (2022, March 16). Variabel SES dan Analisis SES.
- <https://indonesiadata.id/variabel-dan-analisis-ses/>, diakses pada 1 Februari 2023 pukul 11.30.
- IV, J. F. E., Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1996). Machine Learning, Neural, and Statistical Classification. *Journal of the American Statistical Association*, 91(433), 436. <https://doi.org/10.2307/2291432>

- J. Setiadi, Nugroho, Perilaku Konsumen, (Jakarta: PT Kencana, 2010), 77-79.
- Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166, 120658.
- Jabeur, S. Ben, Gharib, C., Mefteh-Wali, S., & Arfi, W. Ben. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166(October 2020), 120658.
- Jindal, N., Liu, B., & Lim, E.-P. (2016). An empirical study of data splitting for machine learning. In Proceedings of the 2016 IEEE International Conference on Big Data (pp. 505-514). IEEE.
- Joshi, A., Saggar, P., Jain, R., Sharma, M., Gupta, D., & Khanna, A. (2021). CatBoost—An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance. *Advances in Data Science and Adaptive Analysis*, 13(03n04), 2141002.
- Joshi, A., Saggar, P., Jain, R., Sharma, M., Gupta, D., & Khanna, A. (2021). CatBoost—An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance. *Advances in Data Science and Adaptive Analysis*, 13(03n04), 2141002.
- Kang, Y., Jang, E., Im, J., Kwon, C., & Kim, S. (2020). Developing a new hourly forest fire risk index based on catboost in South Korea. *Applied Sciences (Switzerland)*, 10(22), 1–21. <https://doi.org/10.3390/app10228213>
- Kearns M, Valiant L (1989). “Cryptographic Limitations on Learning Boolean Formulae and Finite Automata.” In “Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing,” .
- Komalasari, Y., Dewi, S. W. K., Sulastriningsih, R. D., Firmansyah, R., Mauliana, P., Hunaifi, N., & Wiguna, W. (2022). Pelatihan Segmentasi Pemasaran Digital Pasca Pandemi Covid-19 di Tjendana Food Point. *Jumat Informatika: Jurnal Pengabdian Masyarakat*, 3(2), 66–71.  
[https://doi.org/10.32764/abdimas\\_if.v3i2.2868](https://doi.org/10.32764/abdimas_if.v3i2.2868)
- Kondeti, P. K., Ravi, K., Mutheneni, S. R., Kadiri, M. R., Kumaraswamy, S., Vadlamani, R., & Upadhyayula, S. M. (2019). Applications of machine

- learning techniques to predict filariasis using socio-economic factors. *Epidemiology and Infection*, 147, e260.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, p. 13). New York: Springer.
- Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. CRC Press.
- Li, H., Chen, Y., Zhang, C., Li, J., Huang, Y., & Zhao, Y. (2014). A novel incremental feature selection approach for effective dimension reduction in text classification. *Data Mining and Knowledge Discovery*, 28(4), 1008-1031. doi: 10.1007/s10618-013-0325-9.
- Li, X., Wang, L., & Sung, E. (2008). AdaBoost with SVM-based component classifiers. *Engineering Applications of Artificial Intelligence*, 21(5), 785–795. <https://doi.org/10.1016/j.engappai.2007.07.001>
- Liu, H., & Motoda, H. (Eds.). (2007). Feature selection for knowledge discovery and data mining (Vol. 454). Springer Science & Business Media.
- Luthfianto, D., & Suprihadi, H. (2017). Pengaruh kualitas layanan dan gaya hidup terhadap keputusan pembelian cafe jalan Korea. *Jurnal Ilmu dan Riset Manajemen (JIRM)*, 6(2).
- Makovskaja, N. (2018). Classification of High Vegetation in an Urban environment: A Performance Comparison of Machine Learning Methods in a LiDAR dataset.
- Manalu, M. (2015). Korelasi Status Sosial Ekonomi Keluarga Terhadap Gaya Hidup Remaja Di Kota Pekanbaru. *Pengaruh Harga Diskon Dan Persepsi Produk Terhadap Nilai Belanja Serta Perilaku Pembelian Konsumen*, 7(9), 27–44.
- Mao, D., & Edwards, J. R. (2003). Simulation of chemically-reacting gas-solid flowfields using a preconditioning strategy. In *16th AIAA Computational Fluid Dynamics Conference*. <https://doi.org/10.2514/6.2003-4110>
- Marpaung, K., & Rahma, T. I. F. (2023). Pengaruh Gengsi, Diskon, dan Literasi Keuangan Terhadap Perilaku Konsumerisme Generasi Z di Kalangan Mahasiswa Manajemen. *Cakrawala Repository IMWI*, 6(1), 8-23.

- Mengenal Istilah SES - Indonesia Data.* (2022, March 15). Indonesia Data.  
<https://indonesiadat.id/mengenal-istilah-ses-atau-status-sosial-ekonomi/>
- Molnar, C. (2019). Interpretable machine learning: A brief history, state-of-the-art and challenges. arXiv preprint arXiv:1901.04592.
- Molnar, C. (2019). Interpretable machine learning: A brief history, state-of-the-art and challenges. arXiv preprint arXiv:1901.04592.
- Müller, A. C., & Guido, S. (2015). Introduction to Machine Learning with Python and Scikit-Learn. In *O'Reilly Media, Inc.*  
<http://kukuruku.co/hub/python/introduction-to-machine-learning-with-python-andscikit-learn>
- Nalendra, A. R. A., Rosalinah, Y., Priadi, A., Subroti, I., Rahayuningsih, R., Lestari, R., Kusamandari, S., Yuliasari, R., Astuti, D., Latumahina, J., Purnomo, M. W., & Zede, V. A. (2021). Statistika Seri Dasar dengan SPSS. In *Penerbit Media Sains Indonesia.*
- Nasution, Hakim Arman., 2003. Perencanaan dan Pengendalian Produksi, Edisi Pertama. Guna Widya : Jakarta.
- National Center for Education Statistics (ED). (2012). *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation: Recommendations to the National Center for Education Statistics.* ERIC Clearinghouse.
- Neapolitan, R. E., & Jiang, X. (2018). Artificial Intelligence With an Introduction to Machine Learning. In *CRC Press Taylor & Francis Group.*
- Nilsson, N. J. (2005). INTRODUCTION TO MACHINE LEARNING AN  
EARLY DRAFT OF A PROPOSED TEXTBOOK Department of Computer  
Science. *Machine Learning*, 56(2), 387–399.  
<http://www.ncbi.nlm.nih.gov/pubmed/21172442>
- Niosi, A. (2021, June 25). *Lifestyle and Psychographics.* Opentextbc.ca;  
BCcampus. <https://opentextbc.ca/introconsumerbehaviour/chapter/lifestyle-and-psychographics/>, diakses pada 25 Januari 2023 pukul 17.10.
- Ntloko, K. (2020, October 28). *Lifestyle Marketing: How To Leverage Your Consumers' Lifestyle.* Meltwater; Meltwater.

- <https://www.meltwater.com/en/blog/lifestyle-marketing>, diakses pada 25 Januari 2023 pukul 16.30.
- Nugroho, J. Setiadi. Perilaku Konsumen. Jakarta: PT Kencana, 2010.
- Nurazijah, Mira;Fitriani, Syaipia Lailla Nur; Rustini, T. (2023). Pengaruh gaya hidup hedonis terhadap perilaku konsumtif mahasiswa. *Journal on Education*, 05(02), 2345–2352. <https://eprints.umm.ac.id/23776/>
- Parfenov, D., Kuznetsova, L., Yanishevskaya, N., Bolodurina, I., Zhigalov, A., & Legashev, L. (2020). Research application of ensemble machine learning methods to the problem of multiclass classification of DDoS attacks identification. In *2020 International Conference Engineering and Telecommunication, En and T 2020*.  
<https://doi.org/10.1109/EnT50437.2020.9431255>
- Park, H., & Lee, K. (2022). Using Boosted Machine Learning to Predict Suicidal Ideation by Socioeconomic Status among Adolescents. *Journal of Personalized Medicine*, 12(9). <https://doi.org/10.3390/jpm12091357>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2
- Philip Kotler and Kevin Lane Keller. 2016. Marketing Management, 15th Edition New Jersey: Pearson Prentice Hall, Inc
- Plummer, J.T. (1974, January). The concept and application of lifestyle segmentation. *Journal of Marketing*, 38(1), 33-37.
- Pohan, M., Azhar, M. E., Purnama, N. I., & Jasin, H. (2022). Model Perilaku Konsumtif Mahasiswa Fakultas Ekonomi Dan Bisnis Universitas Swasta Di Kota Medan. *Owner*, 6(2), 1498–1508.
- Polat, K. (2019). Freezing of gait (FoG) detection using logistic regression in Parkinson's disease from acceleration signals. *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, 1–4. <https://doi.org/10.1109/EBBT.2019.8742042>
- Pramesti, L. R. (2023). *PENGARUH LITERASI KEUANGAN DAN GAYA HIDUP HEDONISME TERHADAP MANAJEMEN KEUANGAN PRIBADI*

- MAHASISWA PRODI AKUNTANSI STIE PGRI DEWANTARA JOMBANG* (Doctoral dissertation, STIE PGRI Dewantara Jombang).
- Putri, S. M., Mahardayani, I. H., & Ahyani, L. N. (2022). Perilaku Konsumtif Produk Fashion Ditinjau dari Gaya Hidup Hedonis dan Kepribadian Ekstrovert pada Wanita Dewasa Awal. *Jurnal Psikologi Perseptual*, 7(1), 120.
- Putri, Y. A. (2019). Pengaruh bauran promosi dan gaya hidup terhadap minat beli pada kedai kopi serasi. *Jurnal Manajemen Strategi Dan Aplikasi Bisnis*, 2(2), 137–144.
- R. Ariadni and I. Arieshanti, “Implementasi Metode Pohon Keputusan untuk Klasifikasi Data dengan Nilai Fitur yang Tidak Pasti,” ResearchGate, no. June, pp. 3–5, 2015.
- Rachmadani, S. I. (2022). Tinjauan Peran Moderasi Jenis Kelamin Atas Pengaruh Status Sosial Ekonomi Orang Tua Dan Literasi Ekonomi Terhadap Perilaku Konsumtif. *Jurnal Apresiasi Ekonomi*, 10(2), 159–171.
- Raschka, S. (2020). Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing.
- Raschka, S., & Mirjalili, V. (2017). Python machine learning. Packt Publishing Ltd.
- Rasyid, M. B. A. (2022). PENGARUH PENGGUNAAN DEBIT CARD DAN E-MONEY TERHADAP PERILAKU KONSUMTIF MAHASISWA PRODI EKONOMI ISLAM FAKULTAS EKONOMIKA DAN BISNIS UNESA. *Jurnal Ekonomika dan Bisnis Islam*, 5(2), 157-186.
- Ratih Ariadni, I. A. (2015). *Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data*. June.
- Rebala, G., Ravi, A., Churiwala, S. (2019). An Introduction to Machine Learning. Germany: Springer International Publishing.
- Riyanto, S., & Hatmawan, A. A. (2020). *Metode riset penelitian kuantitatif penelitian di bidang manajemen, teknik, pendidikan dan eksperimen*. Deepublish.

- Rohmanto, F., & Susanti, A. (2021). Pengaruh Literasi Keuangan, Lifestyle Hedonis, Dan Sikap Keuangan Pribadi Terhadap Perilaku Keuangan Mahasiswa. *Ecobisma (Jurnal Ekonomi, Bisnis Dan Manajemen)*, 8(1), 40–48. <https://doi.org/10.36987/ecobi.v8i1.2057>
- Romadhon, M. R., & Kurniawan, F. (2021). A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, 41–44. <https://doi.org/10.1109/EIConCIT50028.2021.9431845>
- Romadloniyah, & Setiaji. (2020). Pengaruh Status Sosial Ekonomi Orang Tua, Konformitas, Dan Literasi Keuangan Terhadap Perilaku Konsumtif Dalam Perspektif Gender. *Eeaj*, 9(1), 50–64.
- Sahli, H. (2020). An Introduction to Machine Learning. In *TORUS 1 - Toward an Open Resource Using Services: Cloud Computing for Environmental Data*. <https://doi.org/10.1002/9781119720492.ch7>
- Saputra, A. (2020). Pengaruh Gaya Hidup Terhadap Keputusan Pembelian Produk Fashion Pada Toko Online “SHOPEE”(Studi Pada Mahasiswa Perbankan Syari’ah Angkatan 2016 IAIN Kediri) (Doctoral dissertation, IAIN Kediri).
- Sastrawati, N. (2020). Konsumtivisme Dan Status Sosial Ekonomi Masyarakat. *El-Iqthisadi : Jurnal Hukum Ekonomi Syariah Fakultas Syariah Dan Hukum*, 2(1), 17.
- Schapire R (1990). “The Strength of Weak Learnability.” *Machine Learning*, 45, 197–227.
- Schapire YFR (1999). “Adaptive Game Playing Using Multiplicative Weights.” *Games and Economic Behavior*, 29, 79–103.
- Sekaran, U., & Bougie, R. (2017). Metode Penelitian Untuk Bisnis (Edisi ke 6). Salemba Empat. Setiawan, Kezia M dan Handojo Djoko. 2018. Pengaruh Gaya Hidup, Brand Image, dan Word of Mouth terhadap Keputusan Pembelian Oppo Smartphone. Dalam *Jurnal Administrasi Bisnis*, Vol 7, No 4, Halaman 1-7.

- SEO Management. (2022, July 13). *Rumus Slovin: Pengertian, Notasi, dan Contoh Soal*. Sampoerna Academy.  
<https://www.sampoernacademy.sch.id/id/rumus-slovin/>, diakses pada 3 Februari 2023 pukul 14.00
- Setiawan dan Handojo (2018: 3) Setiawan, Kezia M dan Handojo Djoko. 2018. Pengaruh Gaya Hidup, Brand Image, dan Word of Mouth terhadap Keputusan Pembelian Oppo Smartphone. Dalam Jurnal Administrasi Bisnis, Vol 7, No 4, Halaman 1-7.
- Siregar, Syofian. 2014. Statistik Parametrik untuk Penelitian Kuantitatif (dilengkapi dengan perhitungan manual dan aplikasi SPSS versi 17). Jakarta : Bumi Aksara
- Smith, B. A. E., Ahsan, S., & Menon, C. (2020). An introduction to machine learning evaluation metrics and their application. *Journal of Neural Engineering*, 17(5), 051001. doi: 10.1088/1741-2552/aba936
- Sri Utami Ulfa Rahma, Eri Bukhari, & Eri Teguh Prasetyo. (2022). *Pengaruh Literasi Keuangan, Pendapatan Dan Gaya Hidup Terhadap Perilaku Konsumtif Belanja Online Pada Masa Pandemi Covid-19*, 18(1).
- Stillman, David pengarang; Stillman, Jonah pengarang. (2017). *Gen Z @ work : how the next generation is transforming the workplace / David Stillman and Jonah Stillman*. New York, NY ; Copyright © 2017 by DAS Creative LLC: Harper Business, an imprint of Harper Collins Publishers,.
- Sudaryono. 2017. Metode penelitian. Depok: Raja Grafindo Persada
- Sugiharti, H., & Maula, K. A. (2019). Pengaruh Literasi Keuangan Terhadap Perilaku Pengelolaan Keuangan Mahasiswa. *Accounthink : Journal of Accounting and Finance*, 4(2), 804–818.  
<https://doi.org/10.35706/acc.v4i2.2208>
- Sugiyono (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta
- Sugiyono. 2018. Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta

- Sujarweni, V. Wiratna. 2014. Metode Penelitian: Lengkap, Praktis, dan Mudah Dipahami. Yogyakarta: Pustaka Baru Press.
- Suparyanto dan Rosad (2015). (2020). 済無No Title No Title. *Suparyanto Dan Rosad (2015, 5(3), 248–253.*
- Suthaharan, S. (2015). Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. United States: Springer US.
- Sutisna. 2002. Perilaku Konsumen dan Komunikasi Pemasaran. Remaja Rosdakarya. Bandung.
- Tama, R. T. (2014). Pengeluaran konsumsi mahasiswa program studi Pendidikan ekonomi fakultas ekonomi Universitas negeri Yogyakarta. *Universitas Negeri Yogyakarta.*
- Tari Annamária. (2011). Z generáció. Budapest: *Tericum Kiadó Kft.*
- Tian, X., Pan, B., Bai, L., & Mo, D. (2023). Md-Pred: A Multidimensional Hybrid Prediction Model Based on Machine Learning for Hotel Booking Cancellation Prediction. *International Journal of Pattern Recognition and Artificial Intelligence*, 2351009, 1–21.  
<https://doi.org/10.1142/s0218001423510096>
- Umar, H. (2002). Metode riset bisnis: panduan mahasiswa untuk melaksanakan riset dilengkapi contoh proposal dan hasil riset bidang manajemen dan akuntansi. Indonesia: Gramedia Pustaka Utama.
- Uribeetxebarria, A., Castellón, A., & Aizpurua, A. (2023). Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2 with CatBoost Algorithm. *Remote Sensing*, 15(6), 1640.
- V Wiratna Sujarweni. 2014. Metodologi Penelitian Bisnis & Ekonomi. Yogyakarta: Pustaka Baru.
- Valiant L (1984). “A Theory of the Learnable.” *Communications of the ACM*, 27, 1134–1142.
- Wang, Y., & Wang, T. (2020). Application of improved LightGBM model in blood glucose prediction. *Applied Sciences (Switzerland)*, 10(9).  
<https://doi.org/10.3390/app10093227>

- Wantono, S. (2014). Prediksi Penyelesaian Studi Mahasiswa Baru dengan Metode Fuzzy Tsukamoto (Study Kasus di Universitas Muhammadiyah Gresik) (Doctoral dissertation, Universitas Muhammadiyah Gresik).
- White, C. J., & Tong, E. (2019). On linking socioeconomic status to consumer loyalty behaviour. *Journal of Retailing and Consumer Services*, 50(February), 60–65. <https://doi.org/10.1016/j.jretconser.2019.05.001>
- Widjaja, B. T. 2013. Lifestyle Marketing. Jakarta: Gramedia Pustaka Utama.
- Wijaya, J. (2019). *Implementasi Algoritma Pohon Keputusan CART untuk Menentukan Klasifikasi Data Evaluasi Mobil*.
- Wong, T. T., & Lam, H. K. (2005). A method for determining attribute weights in multiple criteria decision making based on decision matrix and degree of membership. *Expert Systems with Applications*, 29(2), 385-393. doi: 10.1016/j.eswa.2005.04.016.
- Yayah Bahjatussaniah, Nuraini, A. (2015). Pengaruh status sosial ekonomi orang tua dan kelompok teman sebaya terhadap perilaku konsumsi siswa. *Jurnal Pendidikan Ekonomi*, Vol 2 No 3, 1–16.
- Yushita, A. N. (2017). Pentingnya Literasi Keuangan Bagi Pengelolaan Keuangan Pribadi. Nominal, Barometer Riset Akuntansi Dan Manajemen, 6(1). <https://doi.org/10.21831/nominal.v6i1.14330>
- Zahrawati, F., & Faraz, N. J. (2018). Pengaruh kultur sekolah, konsep diri, dan status sosial ekonomi orang tua terhadap perilaku konsumtif siswa. *Harmoni Sosial: Jurnal Pendidikan IPS*, 4(2), 131–141.
- Zakia, A., Adisti, A. A., & Asmarani, A. (2022). Faktor-faktor yang mempengaruhi kelas sosial gaya hidup daya beli dan tingkat konsumsi. *Jurnal Ilmu Manajemen Terapan*, 3(5), 449–457.
- Zayid, E. I. M., & Akay, M. F. (2013). Predicting the performance measures of a message-passing multiprocessor architecture using artificial neural networks. *Neural Computing and Applications*, 23, 2481-2491.
- Zhang, K., Shen, L. C., & Liu, N. (2019). House rent prediction based on joint model. In *ACM International Conference Proceeding Series* (pp. 507–511).

Zhao, X., Li, G., Xie, H., Hu, L., Sun, W., Yang, S., ... & Takahashi, H. (2021).  
The Prediction of Day-to-Day Occurrence of Low Latitude Ionospheric  
Strong Scintillation Using Gradient Boosting Algorithm. *Space  
Weather*, 19(12), e2021SW002884.





## LAMPIRAN

### Surat Pembimbing Skripsi

**KEMENTERIAN AGAMA  
UIN SYARIF HIDAYATULLAH JAKARTA  
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Ir. H. Juanda No. 95 Ciputat 15412 Indonesia  
Telp. (62-21) 7493606, 7493547 Fax. (62-21) 7493315

Website : fst.uinjkt.ac.id  
Email : fst@uinjkt.ac.id

Nomor : B - 0824E/F9 / KM.01 /02/2023  
Lampiran : -  
Perihal : Pembimbing Skripsi  
Kepada Yth.  
1. Suci Ratnawati MTI  
2. Elvi Fetrina M.IT.

*Assalamualaikum, Wr Wb*

Dengan ini diharapkan kesediaan Saudara untuk menjadi pembimbing I/II/  
(Materi/Teknis)\* penulisan skripsi mahasiswa:

Nama	:	CHALISTA PUTRI ANANDA
NIM	:	11190930000038
Program Studi	:	Sistem Informasi
Judul Skripsi	:	MACHINE LEARNING UNTUK PREDIKSI GAYA HIDUP BERDASARKAN SOCIOECONOMIC STATUS (SES) MENGGUNAKAN ALGORITMA CATBOOST STUDI KASUS: MAHASISWA UIN JAKARTA

Judul tersebut telah disetujui oleh Program Studi bersangkutan pada tanggal dengan  
outline, abstraksi dan daftar pustaka terlampir. Bimbingan skripsi ini diharapkan selesai  
dalam waktu 6 (enam) bulan setelah ditandatanganinya surat penunjukan pembimbing skripsi

Apabila terjadi perubahan terkait dengan skripsi tersebut selama proses pembimbingan,  
harap segera melaporkan kepada Program Studi bersangkutan.

Demikian atas kesediaan Saudara, kami ucapan terima kasih.

*Wassalamu'alaikum Wr.Wb*

Jakarta, 14 Februari 2023

a.n Dekan  
**Muji Deken Bid. Akademik**


Cek Keaslian Surat di <https://e-letter.fst.uinjkt.ac.id/scan>

## Surat Permohonan Penelitian



### KEMENTERIAN AGAMA UIN SYARIF HIDAYATULLAH JAKARTA FAKULTAS SAINS DAN TEKNOLOGI

Jl. Ir. H. Juanda No. 95 Ciputat 15412 Indonesia  
Telp. (62-21) 7493606, 7493547 Fax. (62-21) 7493315

Website : fst.uinjkt.ac.id  
Email : fst@uinjkt.ac.id

Nomor : B - 0817/F9 / KM. 01 /02/2023  
Lampiran : -  
Hal : Permohonan Riset

Jakarta, 14 Februari 2023

Kepada Yth.  
Reza Alamsyah Divisi Support  
PUSTIPANDA UIN Syarif Hidayatullah  
Di  
Tempat  
*Assalamualaikum, Wr Wb*

Dengan hormat kami sampaikan bahwa:

Nama	: CHALISTA PUTRI ANANDA
Tempat/Tanggal Lahir	: JAKARTA / 27 September 2001
NIM	: 11190930000038
Semester	: 7
Program Studi	: Sistem Informasi
Alamat	: JL. H. SEKUL NO.3
Telp/HP	: 087776624451

adalah benar yang bersangkutan mahasiswa Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta dan bermaksud melakukan penelitian/riset di instansi yang Bapak/Ibu pimpin, yang sedang dalam penyelesaian skripsi dengan judul skripsi:

#### "MACHINE LEARNING UNTUK PREDIKSI GAYA HIDUP BERDASARKAN SOCIEONOMIC STATUS (SES) MENGGUNAKAN ALGORITMA CATBOOST STUDI KASUS: MAHASISWA UIN JAKARTA"

Untuk itu, kami mohon kesediaannya untuk memberikan kesempatan kepada mahasiswa tersebut dalam melaksanakan penelitian Data/riset di instansi/perusahaan yang Bapak/Ibu pimpin.

Demikian, atas perhatian dan kerjasamanya kami ucapan terima kasih.

*Wassalamu'alaikum, Wr Wb*



Cek Keaslian Surat di <https://e-letter.fst.uinjkt.ac.id/scan>

## Kuesioner



**Syarif Hidayatullah Jakarta**

### MACHINE LEARNING UNTUK PREDIKSI GAYA HIDUP BERDASARKAN SOCIOECONOMIC STATUS (SES) MENGGUNAKAN ALGORITMA CATBOOST STUDI KASUS: MAHASISWA UIN JAKARTA

Assalamualaikum Wr. Wb.  
Halo teman-teman mahasiswa UIN Syarif Hidayatullah Jakarta 😊

Perkenalkan, saya Chalista Putri Ananda, mahasiswi Sistem Informasi UIN Syarif Hidayatullah Jakarta. Saat ini saya sedang melakukan penelitian skripsi dengan judul *Machine Learning untuk Prediksi Gaya Hidup Berdasarkan Socioeconomic Status (SES)* Menggunakan Algoritma CatBoost Studi Kasus: Mahasiswa UIN Jakarta.

Pada penelitian ini saya membutuhkan responden dengan kriteria Mahasiswa aktif UIN Jakarta tahun 2019 - 2022. Apabila Anda memenuhi kriteria di atas, mohon kesediaannya untuk berpartisipasi mengisi kuesioner ini dengan meluangkan waktu 5-10 menit.

Informasi yang dikirimkan dalam kuesioner ini hanya akan digunakan untuk penelitian dan tidak akan dibagikan maupun digunakan untuk alasan lain.

Sebelumnya, saya mengucapkan terima kasih banyak 😊  
Wassalamualaikum Wr Wb 🌟

chalistaputri@gmail.com Switch account 

 Not shared

[Next](#) [Clear form](#)

This content is neither created nor endorsed by Google. [Report Abuse](#) · [Terms of Service](#) · [Privacy Policy](#)

Google Forms

### Pertanyaan Kuesioner

No	Pertanyaan	Jawaban
1.	Nomor Telepon	(Jawaban Singkat)
2.	Kewarganegaraan	<ul style="list-style-type: none"> <li>– Laki-laki</li> <li>– Perempuan</li> </ul>
3.	Apakah Anda menerima beasiswa?	<ul style="list-style-type: none"> <li>– Ya</li> <li>– Tidak</li> </ul>
4.	Fakultas	<ul style="list-style-type: none"> <li>– Ilmu Tarbiyah dan Keguruan</li> <li>– Adab dan Humaniora</li> <li>– Ushuluddin</li> <li>– Syariah dan Hukum</li> <li>– Ilmu Dakwah dan Ilmu Komunikasi</li> <li>– Dirasat Islamiyah</li> <li>– Psikologi</li> <li>– Ekonomi dan Bisnis</li> <li>– Sains dan Teknologi</li> <li>– Ilmu Kesehatan</li> <li>– Ilmu Sosial dan Ilmu Politik</li> <li>– Kedokteran</li> </ul>
5.	Jurusan	<ul style="list-style-type: none"> <li>– Pendidikan Agama Islam</li> <li>– Pendidikan Bahasa Arab</li> <li>– Pendidikan Bahasa Inggris</li> <li>– Pendidikan IPS</li> <li>– Pendidikan Matematika</li> <li>– Pendidikan Biologi</li> <li>– Pendidikan Fisika</li> <li>– Pendidikan Kimia</li> <li>– Manajemen Pendidikan</li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>– Pendidikan Bahasa dan Sastra Indonesia</li> <li>– Pendidikan Guru MI/SD</li> <li>– Pendidikan Islam Anak Usia Dini (PIAUD)</li> <li>– Bahasa dan Sastra Arab</li> <li>– Sastra Inggris</li> <li>– Sejarah dan Kebudayaan Islam</li> <li>– Tarjamah (Bahasa Arab)</li> <li>– Ilmu Perpustakaan</li> <li>– Studi Agama Agama</li> <li>– Ilmu Al-Quran dan Tafsir</li> <li>– Ilmu Hadis</li> <li>– Aqidah dan Filsafat Islam</li> <li>– Ilmu Tasawuf</li> <li>– Perbandingan Mazhab</li> <li>– Hukum Keluarga Islam (Akhwal Syakhsiyah)</li> <li>– Hukum Tata Negara (Siyasah)</li> <li>– Hukum Pidana Islam (Jinayah)</li> <li>– Hukum Ekonomi Syariah (Muamalat)</li> <li>– Ilmu Hukum</li> <li>– Komunikasi dan Penyiaran Islam</li> <li>– Bimbingan Penyuluhan Islam</li> <li>– Manajemen Dakwah</li> <li>– Pengembangan Masyarakat Islam Kesejahteraan Sosial</li> <li>– Jurnalistik</li> <li>– Dirasat Islamiyah</li> <li>– Psikologi</li> <li>– Manajemen</li> <li>– Akuntansi</li> <li>– Ekonomi Pembangunan</li> <li>– Perbankan Syariah</li> <li>– Ekonomi Syariah</li> <li>– Teknik Informatika</li> <li>– Agribisnis</li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>– Sistem Informasi</li> <li>– Matematika</li> <li>– Biologi</li> <li>– Kimia</li> <li>– Fisika</li> <li>– Teknik Pertambangan</li> <li>– Kesehatan Masyarakat</li> <li>– Farmasi</li> <li>– Ilmu Keperawatan</li> <li>– Sosiologi</li> <li>– Ilmu Politik</li> <li>– Ilmu Hubungan Internasional</li> <li>Kedokteran</li> </ul>
6.	Tahun Angkatan	<ul style="list-style-type: none"> <li>– 2019</li> <li>– 2020</li> <li>– 2021</li> <li>– 2022</li> </ul>
7.	Golongan UKT	<ul style="list-style-type: none"> <li>– Golongan 1</li> <li>– Golongan 2</li> <li>– Golongan 3</li> <li>– Golongan 4</li> <li>– Golongan 5</li> <li>– Golongan 6</li> <li>– Golongan 7</li> </ul>
8.	Dimanakah tempat tinggal yang ditempati selama kuliah <i>offline</i> ?	<ul style="list-style-type: none"> <li>– Rumah</li> <li>– Kost</li> <li>– Apartemen</li> <li>Lainnya:</li> </ul>
9.	Berapakah biaya pengeluaran Anda setiap Bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp7.500.000</li> <li>– Rp5.000.001 - Rp7.500.000</li> <li>– Rp3.000.000 - Rp5.000.000</li> <li>– Rp2.000.001 - Rp3.000.000</li> <li>– Rp1.500.001 - Rp2.000.000</li> <li>– Rp1.000.001 - Rp1.500.000</li> <li>– <math>\leq</math> Rp1.000.000</li> </ul>
10.	Darimanakah sumber keuangan yang Anda	<ul style="list-style-type: none"> <li>– Orang Tua</li> <li>– Bekerja</li> </ul>

No	Pertanyaan	Jawaban
	dapatkan untuk kebutuhan sehari-hari?	<ul style="list-style-type: none"> <li>– Magang</li> <li>– Memiliki Usaha</li> </ul>
11.	Pekerjaan Orang Tua	<ul style="list-style-type: none"> <li>– PNS</li> <li>– Karyawan Swasta</li> <li>– Wiraswasta</li> <li>– POLRI/TNI</li> <li>– Lain-lain</li> </ul>
12.	Apakah operator seluler yang anda gunakan?	<ul style="list-style-type: none"> <li>– Telkomsel</li> <li>– Simpati</li> <li>– Kartu As</li> <li>– by.U</li> <li>– IM3</li> <li>– 3</li> <li>– XL</li> <li>– Axis</li> <li>– Smartfren</li> </ul>
13.	Jenis pembayaran apakah yang digunakan pada operator seluler Anda?	<ul style="list-style-type: none"> <li>– Prabayar</li> <li>– Pascabayar</li> </ul>
14.	Berapakah pengeluaran pulsa untuk kebutuhan kuota internet Anda setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>
15.	Apakah E-wallet yang sering Anda gunakan?	<ul style="list-style-type: none"> <li>– OVO</li> <li>– Gopay</li> <li>– Dana</li> <li>– Lainnya:</li> </ul>
<b>Pengeluaran Rutin Bulanan</b>		
16.	Berapakah biaya pengeluaran Anda untuk makanan dan minuman setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp1.500.001</li> <li>– Rp1.000.001 - Rp1.500.000</li> <li>– Rp800.001 - Rp1.000.000</li> <li>– Rp500.001 - Rp800.000</li> <li>– <math>\leq</math> Rp500.000</li> </ul>

No	Pertanyaan	Jawaban
17.	Apakah jenis transportasi apa yang Anda gunakan sehari-hari?	<ul style="list-style-type: none"> <li>– Motor</li> <li>– Mobil</li> <li>– Kereta</li> <li>– Transjakarta</li> <li>– Transportasi Online</li> <li>– Berjalan Kaki</li> </ul>
18.	Berapakah pengeluaran biaya transportasi Anda setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>
19.	Jika Anda ingin makan, cara apa yang sering Anda lakukan untuk mendapatkan makanan tersebut?	<ul style="list-style-type: none"> <li>– Membeli secara langsung</li> <li>– Menggunakan layanan pesan antar</li> <li>– Memasak</li> </ul>
20.	Jika Anda menggunakan layanan pesan antar, platform layanan pesan antar apa yang paling sering Anda gunakan?	<ul style="list-style-type: none"> <li>– Grabfood</li> <li>– GoFood</li> <li>– ShopeeFood</li> </ul>
21.	Seberapa sering Anda menggunakan platform layanan pesan antar makanan dalam sebulan ?	<ul style="list-style-type: none"> <li>– 1-5 kali</li> <li>– 6-10 kali</li> <li>– <math>&gt;</math> 10 kali</li> </ul>
22.	Berapakah biaya pengeluaran Anda untuk kebutuhan pribadi setiap bulan? (Skincare, kosmetik, perawatan, dll)	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>
23.	Berapakah biaya pengeluaran Anda untuk membeli perlengkapan mandi setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>

No	Pertanyaan	Jawaban
24.	Apakah Anda berbelanja kebutuhan bulanan secara langsung dengan mengunjungi minimarket atau supermarket terdekat?	<ul style="list-style-type: none"> <li>– Ya</li> <li>– Tidak</li> </ul>
25.	Jika berbelanja secara online, manakah platform <i>e-commerce</i> yang paling sering Anda gunakan?	<ul style="list-style-type: none"> <li>– Shopee</li> <li>– Tokopedia</li> <li>– Lazada</li> <li>– Blibli</li> <li>– TikTok Shop</li> <li>– Lainnya:</li> </ul>
<b>Biaya Pengeluaran untuk Keperluan Perkuliahan dan Organisasi</b>		
26.	Apakah Anda mengikuti organisasi?	<ul style="list-style-type: none"> <li>– Ya</li> <li>– Tidak</li> </ul>
27.	Berapakah biaya yang Anda keluarkan untuk kegiatan organisasi setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp 100.001</li> <li>– Rp 70.001 - Rp. 100.000</li> <li>– Rp 50.001 - Rp. 70.000</li> <li>– Rp 30.001 - Rp 50.000</li> <li>– <math>\leq</math> Rp 30.000</li> </ul>
28.	Berapakah biaya yang Anda keluarkan untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp 100.001</li> <li>– Rp 70.001 - Rp. 100.000</li> <li>– Rp 50.001 - Rp. 70.000</li> <li>– Rp 30.001 - Rp 50.000</li> <li>– <math>\leq</math> Rp 30.000</li> </ul>
<b>Biaya Pengeluaran untuk Hiburan</b>		
29.	Apakah Anda berlangganan platform hiburan berbayar?	<ul style="list-style-type: none"> <li>– Ya</li> <li>– Tidak</li> </ul>
30.	Apa saja platform berbayar yang Anda ikuti?	<ul style="list-style-type: none"> <li>– Netflix</li> <li>– Viu</li> <li>– Disney+ Hotstar</li> <li>– Wetv</li> <li>– Vidio</li> <li>– Spotify</li> <li>– Apple Music</li> </ul>

<b>No</b>	<b>Pertanyaan</b>	<b>Jawaban</b>
		<ul style="list-style-type: none"> <li>– Tidak Berlangganan</li> </ul>
31.	Berapakah biaya yang Anda keluarkan untuk berlangganan platform hiburan berbayar setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> <li>– Tidak Berlangganan</li> </ul>
32.	Jenis hiburan apa yang anda lakukan setiap bulan?	<ul style="list-style-type: none"> <li>– Olahraga</li> <li>– Nonton Bioskop</li> <li>– Nongkrong</li> <li>– Lainnya:</li> </ul>
33.	Berapakah biaya yang Anda keluarkan untuk hiburan setiap bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>
	<b>Sumber Air Minum</b>	
34.	Apakah sumber air minum yang anda konsumsi?	<ul style="list-style-type: none"> <li>– Air minum dalam kemasan botol/galon bermerek</li> <li>– Air minum dalam galon isi ulang</li> <li>– Air PAM</li> <li>– Air keran non-PAM</li> <li>– Air sumur bor/pompa</li> <li>– Air sumur tumpung</li> <li>– Air sumur yang tidak ditampung</li> <li>– Mata air yang ditampung</li> <li>– Mata air yang tidak ditampung</li> <li>– Air sungai</li> <li>– Air hujan</li> </ul>
	<b>Sumber Bahan Bakar</b>	
35.	Apakah sumber bahan bakar yang Anda gunakan saat memasak?	<ul style="list-style-type: none"> <li>– Kompor listrik</li> <li>– Gas Elpiji – 12 Kg/ Gas 7 Kg/Gas 5 Kg</li> <li>– Gas 3 Kg</li> <li>– Gas kota/Alam</li> <li>– Minyak tanah</li> </ul>

No	Pertanyaan	Jawaban
		<ul style="list-style-type: none"> <li>– Kayu</li> <li>– Arang</li> <li>– Bricket</li> <li>– Tidak Memasak</li> </ul>
	<b>Konsumsi Tembakau</b>	
36.	Apakah Anda mengkonsumsi Tembakau atau Rokok Elektrik (Pod atau Vape) ?	<ul style="list-style-type: none"> <li>– Ya</li> <li>– Tidak</li> </ul>
	<b>Biaya Pengeluaran Konsumsi Tembakau dan Rokok Elektrik</b>	
37.	Berapakah biaya pengeluaran konsumsi Tembakau, Pods, ataupun Vape Anda setiap Bulan?	<ul style="list-style-type: none"> <li>– <math>\geq</math> Rp500.001</li> <li>– Rp250.001 - Rp500.000</li> <li>– Rp150.001 - Rp250.000</li> <li>– Rp100.001 - Rp150.000</li> <li>– <math>\leq</math> Rp100.000</li> </ul>

## Coding Pemodelan CatBoost

In [1]:	<pre>import seaborn as sns import numpy as np import pandas as pd import matplotlib.pyplot as plt import matplotlib.pylab as plt import shap</pre>																																																																				
In [2]:	<pre># Load your data into a pandas dataframe df = pd.read_csv('D:\LULUS APRIL\Book\DATASETKORELASI.csv')</pre>																																																																				
In [3]:	<pre>df=pd.DataFrame(df)</pre>																																																																				
In [4]:	<pre>df</pre>																																																																				
Out[4]:	<table border="1"><thead><tr><th></th><th>Kewarganegaraan</th><th>Jenis Kelamin</th><th>Menerima beasiswa</th><th>Fakultas</th><th>Jurusan</th><th>Golongan UKT</th><th>Tempat tinggal yang ditempati selama kuliah offline</th><th>Pekerjaan Orang Tua</th><th>Transportasi yang digunakan sehari-hari</th><th>Biaya pengeluaran rutin setiap bulan</th><th>...</th><th>Frekuensi penggunaan platform layanan pesan antar</th><th>Berlangganan platform hiburan berbayar</th><th>Biaya pengeluaran untuk berlangganan platform hiburan berbayar</th><th>Biaya pengeluaran untuk hiburan setiap bulan</th><th>Sunyarik</th></tr></thead><tbody><tr><td>0</td><td>WNI</td><td>Laki-laki</td><td>Ya</td><td>Adab dan Humaniora</td><td>Ilmu</td><td>Golongan 2</td><td>Rumah</td><td>Lain-lain</td><td>Transjakarta</td><td>1000000</td><td>...</td><td>5</td><td>Tidak</td><td>0</td><td>100000</td><td>Air</td></tr><tr><td>1</td><td>WNI</td><td>Perempuan</td><td>Tidak</td><td>Psikologi</td><td>Psikologi</td><td>Golongan 7</td><td>Kost</td><td>PNS</td><td>Berjalan Kaki</td><td>3000000</td><td>...</td><td>10</td><td>Ya</td><td>100000</td><td>500000</td><td>Air</td></tr><tr><td>2</td><td>WNI</td><td>Perempuan</td><td>Tidak</td><td>Adab dan Humaniora</td><td>Ilmu</td><td>Golongan 5</td><td>Rumah</td><td>Karyawan Swasta</td><td>Motor</td><td>2000000</td><td>...</td><td>5</td><td>Ya</td><td>100000</td><td>250000</td><td>Air</td></tr></tbody></table>		Kewarganegaraan	Jenis Kelamin	Menerima beasiswa	Fakultas	Jurusan	Golongan UKT	Tempat tinggal yang ditempati selama kuliah offline	Pekerjaan Orang Tua	Transportasi yang digunakan sehari-hari	Biaya pengeluaran rutin setiap bulan	...	Frekuensi penggunaan platform layanan pesan antar	Berlangganan platform hiburan berbayar	Biaya pengeluaran untuk berlangganan platform hiburan berbayar	Biaya pengeluaran untuk hiburan setiap bulan	Sunyarik	0	WNI	Laki-laki	Ya	Adab dan Humaniora	Ilmu	Golongan 2	Rumah	Lain-lain	Transjakarta	1000000	...	5	Tidak	0	100000	Air	1	WNI	Perempuan	Tidak	Psikologi	Psikologi	Golongan 7	Kost	PNS	Berjalan Kaki	3000000	...	10	Ya	100000	500000	Air	2	WNI	Perempuan	Tidak	Adab dan Humaniora	Ilmu	Golongan 5	Rumah	Karyawan Swasta	Motor	2000000	...	5	Ya	100000	250000	Air
	Kewarganegaraan	Jenis Kelamin	Menerima beasiswa	Fakultas	Jurusan	Golongan UKT	Tempat tinggal yang ditempati selama kuliah offline	Pekerjaan Orang Tua	Transportasi yang digunakan sehari-hari	Biaya pengeluaran rutin setiap bulan	...	Frekuensi penggunaan platform layanan pesan antar	Berlangganan platform hiburan berbayar	Biaya pengeluaran untuk berlangganan platform hiburan berbayar	Biaya pengeluaran untuk hiburan setiap bulan	Sunyarik																																																					
0	WNI	Laki-laki	Ya	Adab dan Humaniora	Ilmu	Golongan 2	Rumah	Lain-lain	Transjakarta	1000000	...	5	Tidak	0	100000	Air																																																					
1	WNI	Perempuan	Tidak	Psikologi	Psikologi	Golongan 7	Kost	PNS	Berjalan Kaki	3000000	...	10	Ya	100000	500000	Air																																																					
2	WNI	Perempuan	Tidak	Adab dan Humaniora	Ilmu	Golongan 5	Rumah	Karyawan Swasta	Motor	2000000	...	5	Ya	100000	250000	Air																																																					

	Kewarganegaraan	Jenis Kelamin	Menerima beasiswa	Fakultas	Jurusan	Golongan UKT	Tempat tinggal yang ditempati selama kuliah offline	Pekerjaan Orang Tua	Transportasi yang digunakan sehari-hari	Biaya pengeluaran rutin setiap bulan	...	Frekuensi penggunaan platform pesan antar	Berlangganan platform hiburan berbayar	Biaya pengeluaran untuk berlangganan platform hiburan berbayar	Biaya pengeluaran untuk hiburan setiap bulan	Biaya pengeluaran untuk hiburan berbayar	Sun yar kc
3	WNI	Perempuan	Ya	Ilmu Sosial dan Ilmu Politik	Ilmu Hubungan Internasional	Golongan 1	Rumah	Karyawan Swasta	Motor	1500000	...	5	Tidak	0	250000	k botc bi	Air
4	WNI	Perempuan	Tidak	Sains dan Teknologi	Fisika	Golongan 7	Rumah	Wiraswasta	Transjakarta	2000000	...	5	Ya	100000	250000	k botc bi	Air
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
395	WNI	Perempuan	Tidak	Ekonomi dan Bisnis	Akuntansi	Golongan 5	Rumah	Lain-lain	Motor	1000000	...	5	Tidak	0	150000	Ai bor	Air
396	WNI	Perempuan	Tidak	Psikologi	Psikologi	Golongan 6	Rumah	Wiraswasta	Motor	1500000	...	5	Tidak	0	250000	;	Air
397	WNI	Perempuan	Tidak	Psikologi	Psikologi	Golongan 6	Rumah	Lain-lain	Motor	1000000	...	5	Tidak	0	100000	;	Air
398	WNI	Laki-laki	Tidak	Ekonomi dan Bisnis	Manajemen	Golongan 6	Rumah	Karyawan Swasta	Motor	2000000	...	10	Ya	100000	500000	k botc bi	Air
399	WNI	Perempuan	Tidak	Sains dan Teknologi	Matematika	Golongan 7	Rumah	Lain-lain	Transjakarta	1500000	...	10	Tidak	0	150000	k botc bi	Air

400 rows × 32 columns

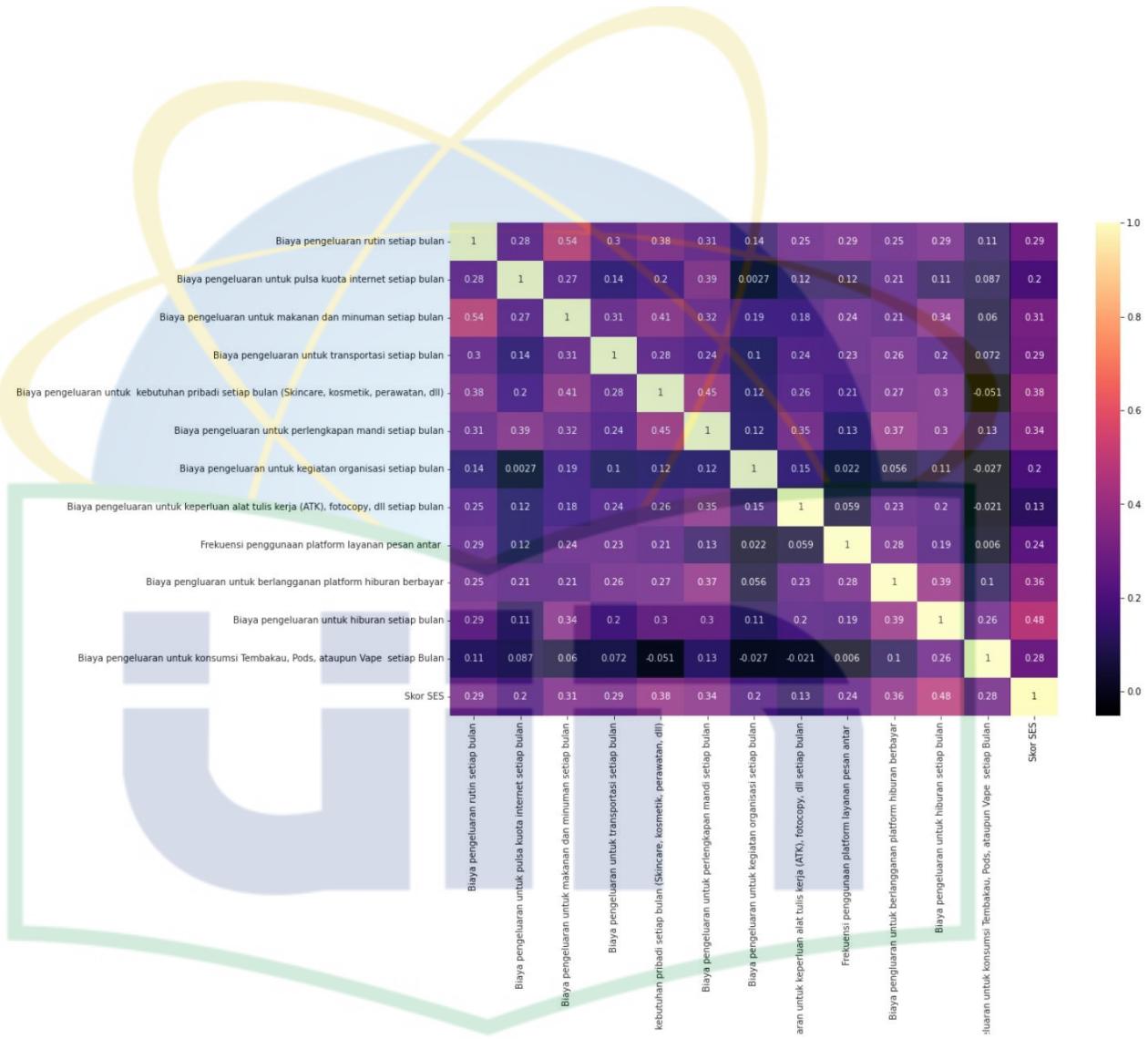


In [5]:

```
#setting dataset
X = df[['Biaya pengeluaran untuk pulsa kuota internet setiap bulan',
        'Biaya pengeluaran untuk makanan dan minuman setiap bulan',
        'Biaya pengeluaran untuk kebutuhan pribadi setiap bulan (Skincare, kosmetik, perawatan, dll)',
        'Biaya pengeluaran untuk perlengkapan mandi setiap bulan',
        'Biaya pengeluaran untuk kegiatan organisasi setiap bulan',
        'Biaya pengeluaran untuk keperluan alat tulis kerja (ATK), fotocopy, dll setiap bulan',
        'Biaya pengeluaran untuk berlangganan platform hiburan berbayar',
        'Biaya pengeluaran untuk hiburan setiap bulan',
        'Biaya pengeluaran untuk konsumsi Tembakau, Pods, ataupun Vape setiap Bulan']]
values #variabel input
y = df[['SES']]values #variabel target/output

#proses perbandingan matriks korelasi untuk setiap atribut
corrmat = df.corr()
top_corr_features = corrmat.index

#visualisasi dengan heatmap
plt.figure(figsize=(15,10))
sns.heatmap(df[top_corr_features].corr(), annot=True, cmap="magma")
plt.show()
```



```
In [6]: # create training and testing vars
y1=df.SES
df_train,df_test,y_df_train,y_df_test = train_test_split(df,y1,test_size=0.2)
drop_list=['Kewarganegaraan',
'Jenis Kelamin',
'Jurusan',
'Pembayaran operator seluler',
'E-wallet yang sering digunakan',
'Platform layanan pesan antar apa yang paling sering digunakan',
'Platform e-commerce yang paling sering digunakan',
'Organisasi',
'Frekuensi penggunaan platform layanan pesan antar',
'Berlangganan platform hiburan berbayar',
'Sumber air minum yang anda konsumsi',
'Sumber bahan bakar yang Anda gunakan saat memasak',
'Konsumsi Tembakau atau Rokok Elektrik (Pod atau Vape)', 'SES', 'Skor SES',
'Transportasi yang digunakan sehari-hari', 'Tempat tinggal yang ditempati selama kuliah offline', 'Menerima beasiswa']
X=df_train.drop(drop_list,axis=1)
X_test = df_test.drop(drop_list, axis=1)
y = df_train.SES
SEED = 1
X_train,X_valid,y_train,y_valid = train_test_split(X, y, test_size=0.2, random_state=SEED)
numerics = [int64, int32, float64, float16, float32, float64]
cat_features=X_train.select_dtypes(exclude=numerics).columns.tolist()
```

```
In [7]: cat_features
Out[7]: ['Fakultas', 'Golongan UKT', 'Pekerjaan Orang Tua']
```

```
In [8]: from catboost import Pool
train_data = Pool(data=X_train,
                  label=y_train,
                  cat_features=cat_features
                 )
valid_data = Pool(data=X_valid,
                  label=y_valid,
                  cat_features=cat_features
                 )
```

```
In [9]: params={loss_function:'MultiClass',
            'eval_metric':AUC,
            'iterations':200,
            'early_stopping_rounds': 200,
            'verbose':200,
```

```
'random_seed': SEED
}

In [10]:
# Create a CatBoostClassifier model
cat_model = CatBoostClassifier(**params)

# Fit the model to the training data
cat_model.fit(train_data, # instead of X_train,y_train
              eval_set=valid_data, # instead of (X_valid,y_valid)
              use_best_model=True,
              plot=True,
              )

Learning rate set to 0.194312
0: test: 0.6172075 best: 0.6172075 (0) total: 174ms remaining: 34.6s
199: test: 0.7411859 best: 0.7495192 {188} total: 8.82s remaining: 0us
bestTest = 0.7495192308
bestIteration = 188

Shrink model to first 189 iterations.
<catboost.core.CatBoostClassifier at 0x19e94810670>

Out[10]:
```

```
In [11]:
predictions = cat_model.predict(X_train)

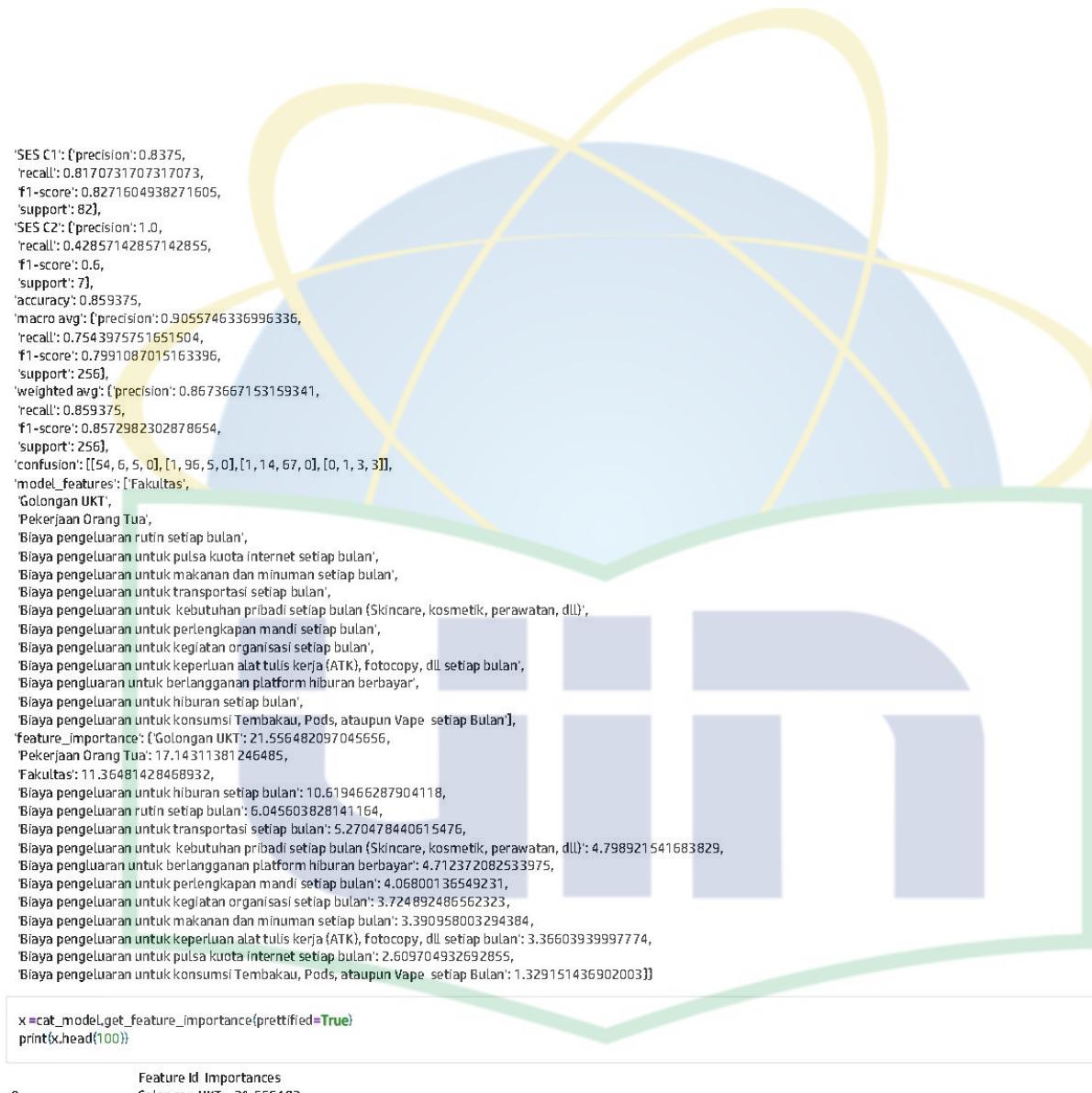
report_dict = {}
report_dict = classification_report(y_train, predictions, output_dict=True)
# print(predictions)
report_dict['accuracy'] = accuracy_score(y_train, predictions)
report_dict['confusion'] = confusion_matrix(y_train, predictions).tolist()
report_dict['model_features'] = X_train.columns.to_list()

importances = cat_model.get_feature_importance()
# std = np.std([tree.feature_importances_ for tree in model.estimators_], axis=0).tolist()
indices = np.argsort(importances)[::-1]
report_dict['feature_importance'] = {X_train.columns[idx]: float(importances[idx]) for idx in indices}

from IPython.display import JSON
JSON(report_dict)

report_dict
```

```
Out[11]: {'SES A': {'precision': 0.9642857142857143,
'recall': 0.8307692307692308,
'f1-score': 0.8925619834710744,
'support': 65},
'SES B': {'precision': 0.8205128205128205,
'recall': 0.9411764705882353,
'f1-score': 0.8767123287671232,
'support': 102},
```



```

1 Pekerjaan Orang Tua 17.143114
2 Fakultas 11.364814
3 Biaya pengeluaran untuk hiburan setiap bulan 10.619466
4 Biaya pengeluaran rutin setiap bulan 6.045604
5 Biaya pengeluaran untuk transportasi setiap bulan 5.270478
6 Biaya pengeluaran untuk kebutuhan pribadi set... 4.798922
7 Biaya pengeluaran untuk berlangganan platform h... 4.712372
8 Biaya pengeluaran untuk perlengkapan mandi set... 4.068001
9 Biaya pengeluaran untuk kegiatan organisasi se... 3.724892
10 Biaya pengeluaran untuk makanan dan minuman se... 3.390958
11 Biaya pengeluaran untuk keperluan alat tulis k... 3.366039
12 Biaya pengeluaran untuk pulsa kuota internet s... 2.609705
13 Biaya pengeluaran untuk konsumsi Tembakau, Pod...

```

In [13]:

```

predictions = cat_model.predict(X_train)
cm = confusion_matrix(y_train, predictions)

sns.heatmap(cm, annot=True, cmap='Blues', fmt='g')

plt.xlabel('Actual')
plt.ylabel('Predicted')
plt.title('Confusion Matrix')
plt.xticks([0.5, 1.5, 2.5, 3.5],['SES A', 'SES B', 'SES C1', 'SES C2'])
plt.yticks([0.5, 1.5, 2.5, 3.5],['SES A', 'SES B', 'SES C1', 'SES C2'])
plt.show()

```



In [14]:

```

# Compute the probabilities for each class
probas = cat_model.predict_proba(X_train)

# Initialize the ROC curve dictionary
roc_dict = {}

# Loop through each class
for i in range(len(cat_model.classes_)):

```

```

# Extract the probability scores for the current class
y_prob = probas[:, 1]

# Create a binary target variable for the current class
y_true = (y_train == cat_model.classes_[1]).astype(int)

# Compute the ROC curve and AUC score for the current class
fpr, tpr, threshold = roc_curve(y_true, y_prob)
roc_auc = auc(fpr, tpr)

# Add the ROC curve and AUC score to the dictionary
roc_dict[f"Class_{i}"] = {"fpr": fpr.tolist(), "tpr": tpr.tolist(), "threshold": threshold.tolist(), "roc_auc": roc_auc}

# Add the ROC curve dictionary to the report dictionary
report_dict["roc_curve"] = roc_dict

```

```

In [15]:
# Loop through each class and plot the ROC curve
for i in range(len(cat_model.classes_)):
    # Get the fpr, tpr, and threshold values for the current class
    fpr = roc_dict[f"Class_{i}"]["fpr"]
    tpr = roc_dict[f"Class_{i}"]["tpr"]
    threshold = roc_dict[f"Class_{i}"]["threshold"]

    # Plot the ROC curve
    plt.plot(fpr, tpr, label=f"Class {i} (AUC = {roc_dict[f'Class_{i}']['roc_auc']:.2f})")

# Add labels and title to the plot
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend()
plt.show()

```

