LAPORAN AKHIR

KLASIFIKASI PROFIL RISIKO INVESTASI MAHASISWA BERDASARKAN TUJUAN DAN TOLERANSI RISIKO

KELOMPOK: SD-A2 - KELOMPOK 4

Raden Bagus Rifai Kacanegara	**162112133047**
Muhammad Aqeela Addimaysqi	**162112133050**
Gede Nayaka Baswara	**162112133065**
Richard Hasannain Mongide	**162112133093**
Andi Rafi Fajar Wally	**162112133097**

TEAM-BASED PROJECT

MATA KULIAH DATA MINING I

PROGRAM STUDI TEKNOLOGI SAINS DATA

FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN

UNIVERSITAS AIRLANGGA

2023

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB II

BAB III

BAB IV

BAB I
PENDAHULUAN

1.1. Latar Belakang

Saat ini investasi merupakan kegiatan penting dalam mendorong perkembangan perekonomian. Investasi kerap dikaitkan dengan penanaman modal yang dilakukan oleh seorang yang disebut investor. Jika kita mengkaji definisi investasi itu sendiri merupakan penempatan sejumlah dana pada saat ini dengan harapan menghasilkan keuntungan di masa depan. Pada umumnya individu melakukan investasi karena ingin mendapatkan keuntungan atau tingkat pengembalian yang cukup tinggi atau sesuai dengan apa yang diharapkan (Bakhri, 2018). Investasi saham di pasar modal merupakan salah satu bentuk investasi yang dikenal masyarakat luas dan pasar modal berperan penting dalam meningkatkan perekonomian suatu negara (Pajar & Pustikaningsih, 2017). Dengan maraknya tren investasi yang dilakukan anak muda jika ditelisik terdapat faktor penting yang melatarbelakangi, yaitu kaum milenial yang mulai ingin membiasakan diri untuk dapat mengontrol sikap keuangan (financial attitudes) mereka sejak muda. Mereka mulai memikirkan kehidupan masa depannya. Karena masih muda, mereka sering terpacu adrenalinnya saat menghadapi risiko dan menganggap bahwa mereka memiliki kesempatan untuk mencoba lagi.

Lalu terdapat toleransi risiko investasi, Toleransi risiko investasi didefinisikan sebagai suatu proses penentuan tingkat toleransi seorang individu menggunakan alat yang efektif melalui metode Psikometri (Roszkowski, Davey, dan Grable, 2005) dalam Cheng *et al*. (2009). faktor-faktor yang mempengaruhi profil risiko investasi mahasiswa berdasarkan tujuan dan toleransi risiko mereka serta mengklasifikasikan ke dalam profil risiko investasi yang konservatif/moderat atau agresif adalah penting untuk memahami preferensi dan karakteristik individu dalam hal investasi. Selain itu, toleransi risiko individu juga memainkan peran kunci dalam menentukan profil risiko investasi. Toleransi risiko mencerminkan sejauh mana seseorang bersedia untuk mengambil risiko dalam investasi. Faktor-faktor seperti tingkat pengetahuan dan pemahaman tentang investasi, tingkat pendapatan dan stabilitas keuangan, serta pengalaman sebelumnya dalam melakukan investasi dapat mempengaruhi tingkat kenyamanan dalam mengambil risiko investasi.

Secara garis besar, terdapat beberapa macam klasifikasi profil risiko mahasiswa dalam berinvestasi dengan tujuan keuangan masa depan yang berbeda-beda, dimana untuk memenuhi tujuan keuangan tersebut diperlukan perencanaan investasi yang matang. Menurut Hanafi (2009), Risiko bisa didefinisikan dengan berbagai cara yaitu bisa kejadian yang merugikan. Profil risiko merupakan hal awal yang seharusnya diketahui seorang investor untuk mendapatkan jenis investasi yang cocok untuk diri investor terutama yang menyangkut dengan harta pribadi. Penting bagi investor untuk menyesuaikan dengan profil risikonya, Secara umum ada 3 jenis profil risiko investor yaitu Konservatif, Moderat dan Agresif. Yang pertama ada Konservatif: Investor yang menginginkan kestabilan dan kepastian. Profil konservatif tidak menyukai risiko sehingga investor cenderung melakukan investasi seperti properti, deposito dan asuransi. Yang kedua Moderat: Investor tipe ini dapat mentolerir risiko menengah dibandingkan profil konservatif dalam menghadapi berbagai risiko dengan harapan mendapatkan *return* yang lebih seimbang. Biasanya profil berinvestasi pada reksadana. Yang

ketiga yaitu Agresif: Profil risiko ini merupakan profil risiko dengan toleransi risiko yang paling tinggi. Jika investor dengan profil risiko agresif berarti dia merupakan *risk taker*. Dengan profil ini umumnya investor menginginkan produk investasi dengan *return* yang tinggi namun risiko yang tinggi tersebut diak menjadi masalah bagi investor yang agresif. Kesimpulan yang kami dapat dari ketiga profil ini ternyata belum ada klasifikasi profil risiko mahasiswa secara khusus yang dapat membantu mahasiswa berinvestasi secara aman.

1.2. Rumusan Masalah

1. Bagaimana karakteristik mahasiswa dengan profil risiko Konservatif/Moderat?
2. Apa variabel-variabel yang berpengaruh dalam profil risiko investasi mahasiswa?
3. Apa model yang tepat untuk melakukan klasifikasi profil risiko investasi mahasiswa?

1.3. Tujuan Penelitian

1. Mengetahui karakteristik mahasiswa dengan profil risiko Konservatif/Moderat.
2. Menentukan variabel-variabel yang berpengaruh dalam profil risiko investasi mahasiswa.
3. Mengetahui model yang tepat untuk melakukan klasifikasi profil risiko investasi mahasiswa.

1.4. Batasan Penelitian

1. Penelitian ini hanya terfokus pada mahasiswa sebagai objek penelitian.
2. Variabel-variabel yang digunakan hanya terbatas pada tujuan berinvestasi dan tingkat toleransi risiko mahasiswa.
3. Model klasifikasi profil risiko yang ditemukan hanya dapat diterapkan untuk mahasiswa.

BAB II
TINJAUAN PUSTAKA

2.1. Klasifikasi

Klasifikasi merupakan cara mengelompokkan hal berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi. Klasifikasi sendiri dapat digunakan pada banyak cara baik secara manual maupun dengan bantuan teknologi. Klasifikasi manual merupakan klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, dibagi menjadi beberapa algoritma yaitu Naive Bayes, Support Vector Machine, Decision Tree, Fuzzy dan Jaringan saraf tiruan (WIbawa et al., 2018).

Kemudian klasifikasi dibagi sesuai dengan target variabel kategori, contoh pendapatan dapat dipisahkan menjadi pendapatan tinggi, sedang, dan rendah (Mardi, 2017). Contoh lain nya dalam bisnis dan penelitian adalah :

a. Menentukan apakah sebuah transaksi merupakan transaksi yang curang atau bukan.
b. Memperkirakan apakah suatu ajuan hipotek oleh customer merupakan suatu kredit yang baik atau kurang baik
c. Mendiagnosis penyakit pasien untuk mendapatkan termasuk penyakit jenis apa.

2.2. Profil Risiko Investasi

Risiko investasi adalah sebuah informasi yang harus diterima para investor agar bisa menerima hasil investasi yang berupa imbal hasil. Risiko investasi juga dibagi menjadi 2, yaitu risiko sistematis dan risiko tidak sistematis. Risiko sistematis merupakan risiko yang erat dengan pergerakan dan perubahan pada keadaan pasar yang disebabkan oleh antisipasi investor terhadap imbal hasil yang diinginkan. Risiko tidak sistematis merupakan risiko yang disebabkan oleh spesifikasi masing-masing perusahaan (Putri et al., 2017).

Risiko investasi dalam bahasa umum nya adalah potensi kerugian dari hasil investasi yang tidak sesuai dengan harapan para investor. Tapi bila para investor dapat melakukan manajemen risiko investasi maka risiko investasi dapat dikontrol dengan baik selama risiko tersebut masih dalam batas toleransi (Nabila & Safri, 2022). Karena itu bila investor memiliki pengalaman yang cukup dalam melakukan investasi akan cenderung memilih return besar walaupun risiko investasi yang akan didapat juga lebih besar. Namun dengan pengalaman mereka dapat lebih memastikan dan memperhatikan setiap langkah investasi yang dibuat (Mandagie et al., 2020).

2.3. Mahasiswa

Mahasiswa dapat dikatakan sebagai individu yang sedang menuntut ilmu di tingkat perguruan tinggi. Mahasiswa juga suatu kewajiban yang diselesaikan dalam jangka waktu yang secepat mungkin atau sesingkat mungkin. Mereka juga dipercayai memiliki tingkat kecerdasan yang tinggi dalam hal berpikir, imajinasi dan perencanaan dalam bertindak kritis (Ambarwati et al., 2017).

Fungsi dasar dari mahasiswa sendiri adalah bergumul dengan ilmu pengetahuan dan memberikan perubahan agar menjadi lebih baik dengan intelektualitas yang mereka miliki selama mereka hidup (Papilaya & Huliselan, 2016). Ide dan pemikiran cerdas mereka dapat merubah paradigma yang berkembang dalam suatu ruang lingkup dan membuatnya lebih terarah sesuai kepentingan bersama. Karena mahasiswa mempunyai peran penting dalam masyarakat yaitu sebagai agen of change, social control, iron stock dan moral force (Cahyono, 2019).

2.4. Tujuan dan Toleran Investasi

Toleransi Investasi didefinisikan sebagai jumlah maksimal ketidakpastian bahwa seseorang bersedia menerima risiko ketika membuat keputusan dalam hal keuangan, mencapai ke hampir setiap bagian dari kehidupan ekonomi dan juga sosial. Toleransi Investasi sendiri menjadi faktor yang penting untuk menjadi pertimbangan dalam menentukan preferensi investasinya (Putri et al., 2017).

Para investor akan berperilaku berbeda saat menghadapi risiko saat melakukan investasi, mereka tidak boleh bertindak gegabah ataupun terlalu santai dalam menghadapi risiko yang muncul dalam waktu yang tak tertentu. Karena setiap risiko dapat membuat mereka sangat merugi, sehingga toleransi investasi akan berpengaruh terhadap keputusan investasi mereka (Mandagie et al., 2020). Tujuan dari toleransi investasi sendiri adalah agar investor dapat mengetahui kemampuanya sendiri untuk menoleransi risiko dalam berinvestasi untuk meminimalkan kerugian, mengetahui risiko dari investasi pilihannya, dan dapat mengetahui mana jenis investasi yang cocok dengan kepribadian mereka dan menghindari yang kurang sesuai (Nabila & Safri, 2022).

2.5. Model Klasifikasi Machine Learning

2.5.1. Logistic Regression

Regresi Logistik (logistic regression) merupakan bagian dari analisis regresi yang digunakan saat variabel dependen (respon) merupakan variabel dikotomi. Variabel dikotomi adalah dua nilai yang mewakili kemunculan atau tidak adanya sebuah kejadian yang biasanya diberi angka 1 (ya) dan 0 (tidak). Berbeda dengan regresi linier biasa, regresi logistik tidak menghipotesis hubungan antara variabel independen dan dependen secara linier (Dewi, 2016).

Cara untuk menginterpretasi koefisien variabel prediksi yakni dengan menggunakan metode odds ratio. Jika nilai odds ratio kurang dari 1, maka antara variabel prediksi dan respon terdapat hubungan negatif setiap kali perubahan nilai variabel prediksi (X). Kemudian bila nilai yang didapatkan lebih dari 1, maka antara variabel prediksi dan respon terdapat hubungan positif setiap kali perubahan nilai variabel prediksi (X) (Bimantara & Dina, 2018).

2.5.2. Decision Tree

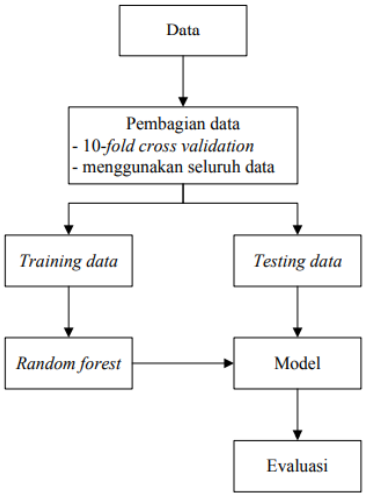
Decision tree merupakan metode yang populer pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan auran dan tentunya lebih mudah dipahami dengan bahasa alami (Achmad & Slamat, 2012). Pengertian yang lebih ilmiah dari decision tree adalah sebuah

diagram air yang berbentuk seperti struktur pohon yang mana setiap interval node menyatakan pengujian terhadap suatu atribut, setiap cabang memberikan output dari hasil uji tersebut dan leaf node menyatakan kelas-kelas atau distribusi kelas.

Pada node teratas disebut sebagai root node atau node akar. Sebuah root node akan punya beberapa edge keluar tetapi tidak punya edge masuk, internal node akan mempunyai satu edge masuk dan beberapa edge keluar, sedangkan leaf node hanya akan memiliki satu edge masuk tanpa punya edge keluar. Pada intinya decision tree bertujuan untuk mengklasifikasikan suatu sampel data yang belum diketahui/memiliki kelasnya ke dalam kelas yang tersedia (Qadrini et al., 2021).

2.5.3. Random Forest

Random forest adalah metode yang dipakai untuk melakukan klasifikasi dan regresi. Metode ini juga dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap node dilakukan secara acak. Metode ini dipakai untuk membangun pohon keputusan yang terdiri oleh root node, internal node, dan leaf node dengan memakai atribut dan data secara random sesuai ketentuan. Pohon keputusan sendiri dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan nilai gain (Siburian & Mulyana, 2018).



Gambar 2. 1 Random Forest

Metode ini termasuk sebuah ensemble (kumpulan) dikarenakan metode pembelajaran yang menggunakan pohon sebagai base classifier yang dibangun dan dikombinasikan. ada 3 aspek penting dalam metode random forest yaitu (Primajaya & Sari, 2018) :

1. Melakukan bootstrap sampling untuk membangun pohon prediksi.
2. Masing - masing pohon keputusan memprediksi dengan prediktor acak.
3. Lalu random forest melakukan prediksi dengan mengkombinasikan hasil dari setiap pohon keputusan dengan majority vote untuk klasifikasi.

2.5.4. K-Nearest Neighbor (KKN)

K-Nearest Neighbor (K-NN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek. Keakuratan K-NN ini sangat dipengaruhi oleh ada atau tidaknya faktor-faktor yang tidak relevan atau bila bobot fitur tersebut tidak sama dengan relevansinya terhadap klasifikasi. K-NN juga contoh dari teknik lazy learning, yaitu sebuah teknik yang menunggu sampai pertanyaan datang agar sama dengan data training (Dewi, 2016).

Konsep dasar dari K-Nearest Neighbor adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya. Nilai dari jarak antara data uji dengan data latih dibariskan dari nilai terendah. Proses tersebut dilakukan untuk memilih jarak minimum sebanyak K buah. Perhitungan dilakukan dengan persamaan sebagai berikut (Nasution et al., 2019) :

$$d(x,y)=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}$$

Gambar 2. 2 Rumus Euclidean

2.5.5. XGBoost

Metode XGBoost adalah sebuah metode yang ditemukan oleh seseorang Friedman. Metode ini adalah pengembangan atau versi advenced dari algoritma GDBT (Gradient Boosting Decision Tree). XGBoost merupakan salah satu library machine learning yang dapat berfungsi untuk prediksi atau pengklasifikasikan berbasis pohon keputusan. Model ini ini memiliki optimasi 10 kali lebih cepat dibandingkan pendahulu GBM lainnya. kemudian untuk keakuratan nilai hasil klasifikasi tergantung oleh parameter yang dipakai (Givari et al., 2022).

No	Parameter	Rentang Nilai
1	*Max_depth*	5
2	*seed*	7
3	*Test_size*	0.35
4	*Feature*	1-9
5	*Learning rate*	0.05

Gambar 2. 3 XGBoost

XGBoost sendiri sebuah algoritma yang dapat membangun boosted trees secara efisien dan beroperasi secara paralel. Metode ini juga termasuk metode ensemble, yang mana pada dalam pohon regresi node bagian dalam mewakili nilai untuk test atribut dan leaf nodess dengan skor mewakili keputusan (Karo, 2020).

BAB III
METODOLOGI

3.1. Metode Pengambilan Sampel

3.1.1 Populasi

Populasi adalah kelompok orang atau objek yang memiliki ciri-ciri atau karakteristik tertentu yang diinginkan oleh peneliti untuk dipelajari dan dianalisis sehingga dapat diambil kesimpulan yang relevan (Garaika & Darmanah, 2019). Dalam penelitian ini, kami menggunakan populasi seluruh Mahasiswa Teknologi Sains Data Universitas Airlangga (Angkatan 2020, 2021, dan 2022) sebagai populasi dalam penelitian ini.

3.1.2 Sampel

Sampel adalah sebagian kecil dari populasi yang memiliki karakteristik yang sama dengan populasi secara keseluruhan. Sampel sering digunakan dalam penelitian sebagai representasi dari populasi yang lebih besar, terutama jika populasi tersebut sangat besar sehingga sulit untuk mempelajari seluruh anggotanya. Dari seluruh mahasiswa Teknologi Sains Data Universitas Airlangga, kami menetapkan sampel dengan ketentuan sebagai berikut:

a. Teknik Pengambilan Sampel

   Di dalam penelitian ini, kami menggunakan teknik pengambilan sampel berupa Non Probability Sampling dengan jenis Convenience Sampling. Dalam sampel non-probabilitas, individu dipilih berdasarkan kriteria non-acak dan tidak semua individu memiliki peluang yang sama untuk dipilih. Convenience Sampling dipilih karena lebih mudah dan murah diakses, tetapi memiliki risiko bias pengambilan sampel yang lebih tinggi.

b. Jumlah Sampel

   Dalam penentuan jumlah sampel atau ukuran sampel agar dapat merepresentasikan populasi, kami menggunakan metode slovin untuk mengetahui jumlah sampel yang dibutuhkan. Metode slovin dapat ditulis secara matematis sebagai berikut:

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

   Keterangan:

   N = Jumlah populasi (seluruh mahasiswa Teknologi Sains Data Universitas Airlangga atau sebanyak 321 mahasiswa)

   e = *margin of error* dalam penelitian ini ditetapkan sebesar 0,01.

   Dengan penghitungan menggunakan rumus di atas dengan jumlah seluruh mahasiswa Teknologi Sains Data Universitas Airlangga sebanyak 321 mahasiswa, kami mendapatkan jumlah sampel minimal sebanyak 76 mahasiswa.

   Untuk melakukan pengambilan sampel, kami telah menyebarluaskan Google Form kepada seluruh mahasiswa Teknologi Sains Data Universitas Airlangga.

3.2. Variabel Penelitian

Untuk mengetahui profil risiko investasi mahasiswa berdasarkan tujuan dan toleransi risikonya, kita perlu menggunakan variabel-variabel berikut:

A. Variabel dependen

Variabel dependen atau biasa disebut variabel respon yang akan digunakan dalam penelitian ini adalah variabel Profil Risiko Investasi dari setiap mahasiswa Teknologi Sains Data Universitas Airlangga.

B. Variabel independen

Untuk dapat mengklasifikasikan profil risiko investasi mahasiswa, kami menggunakan variabel-variabel sebagai berikut:

a. Jenis Kelamin

Variabel ini merupakan variabel kategorik, berisikan 2 pilihan (Laki-laki atau Perempuan), dan berfungsi untuk mengetahui apakah jenis kelamin berpengaruh terhadap profil risiko investasi.

b. Tujuan Investasi

Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (Jangka panjang, Jangka pendek, dan Meningkatkan kekayaan), dan berfungsi untuk mengetahui tujuan investasi responden serta mengetahui pengaruhnya terhadap variabel respon.

c. Lama Investasi

Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (<5 Tahun, 1-5 Tahun, atau >5 tahun), dan berfungsi untuk mengetahui preferensi jangka investasi responden serta mengetahui pengaruhnya terhadap variabel respon.

d. Dana Investasi

Variabel ini merupakan variabel numerik dan berfungsi untuk mengetahui dana yang disiapkan oleh responden untuk melakukan investasi serta mengetahui pengaruhnya terhadap variabel respon.

e. Pemahaman Investasi

Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (Kurang Baik, Cukup Baik, dan Sangat Baik) dan berfungsi untuk mengetahui tingkat pemahaman terkait konsep investasi serta mengetahui pengaruhnya terhadap variabel respon.

f. Pendapatan dan Stabilitas Keuangan

Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (Rendah, Menengah, dan Tinggi) dan berfungsi untuk mengetahui tingkat stabilitas keuangan responden serta mengetahui pengaruhnya terhadap variabel respon.

g. Pengalaman Investasi

Variabel ini merupakan variabel kategorik, berisikan 2 pilihan (Ya dan Tidak), berfungsi untuk mengetahui apakah responden memiliki pengalaman investasi sebelumnya serta mengetahui pengaruhnya terhadap variabel respon.

h. Frekuensi Riset terkait Investasi
Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (Jarang, Kadang-kadang, dan Selalu), berfungsi untuk mengetahui kesiapan responden sebelum melakukan investasi serta mengetahui pengaruhnya terhadap variabel respon.

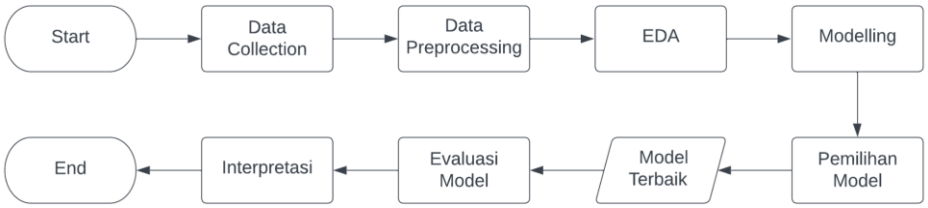
i. Respon Terhadap Volatilitas

Variabel ini merupakan variabel kategorik, berisikan 3 pilihan (Tenang, Khawatir, dan Panik), berfungsi untuk mengetahui bagaimana respon dari responden terhadap perubahan nilai investasi serta mengetahui pengaruhnya terhadap variabel respon.

j. Instrumen Investasi Favorit

Variabel ini merupakan variabel kategorik, berisikan 5 pilihan (Saham, Properti, Reksa Dana, Obligasi, dan Lainnya), berfungsi untuk mengetahui instrumen favorit dari responden serta mengetahui pengaruhnya terhadap variabel respon.

3.3. Diagram Alur Penelitian



Gambar 3. 1 Flowchart Penelitian

3.4. Data Preprocessing

Berikut adalah proses dari *data preprocessing* yang telah dilakukan:

1. *Import library* yang dibutuhkan

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
```

Gambar 3. 2 Import Library

2. Memanggil *file* data mentah yang kita pakai dalam *project* ini

```
df = pd.read_csv('data-raw.csv', sep=',')
df
```

	Timestamp	Nama Mahasiswa	NIM	Jenis Kelamin	Angkatan	Tujuan utama anda dalam melakukan investasi adalah untuk (pilih salah satu opsi)	Berapa lama anda berencana untuk menanamkan investasi anda?	Berapa jumlah dana yang bersedia anda investasikan? (Contoh : 100000)	Seberapa baik anda memahami konsep investasi dan instrumen-instrumen investasi?	Bagaimana tingkat pendapatan dan stabilitas keuangan anda?	Sejauh mana anda bersedia mengambil risiko dalam investasi	Apakah Anda memiliki pengalaman sebelumnya dalam melakukan investasi?
0	5/26/2023 5:27:21	muhammad yahya	164221005	Laki - Laki	2022	Mempersiapkan masa depan jangka panjang	Lebih dari 5 tahun	100000	Cukup Baik	Menengah	Sangat Konservatif	TIDAKK

Gambar 3. 3 File Raw

3. Mengganti nama kolom dalam data frame menjadi lebih singkat

```
df = df.rename(columns={
    'Nama Mahasiswa': 'Nama',
    'Jenis Kelamin': 'Gender',
    'Angkatan': 'Tahun',
    'Tujuan utama anda dalam melakukan investasi adalah untuk (pilih salah satu opsi)': 'Tujuan',
    'Berapa lama anda berencana untuk menanamkan investasi anda?': 'Lama_Investasi',
    'Berapa jumlah dana yang bersedia anda investasikan? (Contoh : 100000)': 'Dana_Investasi',
    'Sejauh mana anda bersedia mengambil risiko dalam investasi': 'Tingkat_Resiko',
    'Seberapa baik anda memahami konsep investasi dan instrumen-instrumen investasi?': 'Pemahaman_Investasi',
    'Bagaimana tingkat pendapatan dan stabilitas keuangan anda?': 'Pendapatan_dan_Stabilitas_Keuangan',
    'Apakah Anda memiliki pengalaman sebelumnya dalam melakukan investasi?': 'Pengalaman_Investasi',
    'Seberapa sering Anda melakukan riset dan analisis sebelum melakukan investasi?': 'Frekuensi_Riset',
    'Bagaimana Anda menghadapi volatilitas pasar dan fluktuasi nilai investasi Anda?': 'Respon_Terhadap_Volatilitas',
    'Jenis instrumen investasi mana yang paling Anda minati?': 'Instrumen_Favorit',
    'Apakah Anda cenderung memiliki profil risiko investasi yang konservatif/moderat atau agresif': 'Profil_Risiko'
})
```

Gambar 3. 4 Nama DataFrame

Mengganti nama-nama kolom dalam data frame menggunakan metode rename dari pandas. Di sini, setiap kolom diberi nama baru dengan menggunakan kamus yang diberikan. Misalnya, kolom 'Nama Mahasiswa' diganti menjadi 'Nama', 'Jenis Kelamin' diganti menjadi 'Gender', dan seterusnya. Lalu kita cek berapa ukuran data frame dan kolom yang sudah diganti.

```
df.shape
```

```
(118, 16)
```

```
df.columns
```

```
Index(['Timestamp', 'Nama Mahasiswa', 'NIM', 'Jenis Kelamin', 'Angkatan',
       'Tujuan utama anda dalam melakukan investasi adalah untuk (pilih salah satu opsi)',
       'Berapa lama anda berencana untuk menanamkan investasi anda?',
       'Berapa jumlah dana yang bersedia anda investasikan? (Contoh : 100000)',
       'Seberapa baik anda memahami konsep investasi dan instrumen-instrumen investasi?',
       'Bagaimana tingkat pendapatan dan stabilitas keuangan anda?',
       'Sejauh mana anda bersedia mengambil risiko dalam investasi',
       'Apakah Anda memiliki pengalaman sebelumnya dalam melakukan investasi?',
       'Seberapa sering Anda melakukan riset dan analisis sebelum melakukan investasi?',
       'Bagaimana Anda menghadapi volatilitas pasar dan fluktuasi nilai investasi Anda?',
       'Jenis instrumen investasi mana yang paling Anda minati?',
       'Apakah Anda cenderung memiliki profil risiko investasi yang konservatif/moderat atau agresif'],
      dtype='object')
```

Gambar 3. 5 Ubah nama Data Frame

Terlihat bahwa data frame tersebut berisi 118 baris dan 16 kolom.

4. Menghapus kolom yang tidak relevan

Menghapus kolom 'Timestamp, Nama, NIM, dan Tahun' dari data frame menggunakan metode drop dari pandas. Parameter axis=1 menunjukkan bahwa yang dihapus adalah kolom, bukan baris.

5. Cek apakah ada *missing value* (NaN) dalam setiap kolom

```
df.isna().sum()
```

```
Gender                                0
Tujuan                                0
Lama_Investasi                        0
Dana_Investasi                        0
Pemahaman_Investasi                   0
Pendapatan_dan_Stabilitas_Keuangan    0
Pengalaman_Investasi                  0
Frekuensi_Riset                       0
Respon_Terhadap_Volatilitas           0
Instrumen_Favorit                     0
Profil_Risiko                         0
dtype: int64
```

Gambar 3. 6 Melihat Jumlah Missing Value

Menggunakan metode isna() untuk mengidentifikasi nilai yang hilang (NaN) dalam data frame, dan kemudian menggunakan metode sum() untuk menghitung jumlah nilai yang hilang dalam setiap kolom. Pada data ini kita tidak menemukan adanya *missing value* atau nilai yang hilang.

6. Menangani duplikat

```
duplicate_rows = df[df.duplicated()]
print("Jumlah duplikasi berdasarkan semua kolom:", len(duplicate_rows))
```

```
Jumlah duplikasi berdasarkan semua kolom: 1
```

```
df = df.drop_duplicates()
df
```

	Gender	Tujuan	Lama_Investasi	Dana_Investasi	Pemahaman_Investasi	Pendapatan_dan_Stabilitas_Keuangan	Pengalaman_Investasi	Frekuensi
0	Laki-Laki	Mempersiapkan masa depan jangka panjang	Lebih dari 5 tahun	100000	Cukup Baik	Menengah	TIDAKK	
1	Laki-Laki	Mempersiapkan masa depan jangka panjang	Lebih dari 5 tahun	100000	Kurang baik	Menengah	TIDAKK	Kadang -

Gambar 3. 7 Melihat Duplicated Values

Menggunakan metode duplicated() pada data frame untuk mengidentifikasi baris-baris yang merupakan duplikat, dan kemudian menyimpannya dalam variabel duplicate_rows. Lalu dengan fungsi len() kita hitung jumlah baris yang merupakan duplikat, dan kemudian mencetak jumlahnya. Dari hasil tersebut, ada satu baris yang duplikat. Setelah mengetahui ada baris yang duplikat, kita hapus baris tersebut dengan metode drop_duplicates().

7. Pembagian kolom kategorik dan numerik

Untuk memudahkan modelling dan EDA, kita dapat melakukan pengelompokan dari variabel kategorik dan numerik dengan menggunakan modul df.unique sebagai beriku:

```
col = list(df.columns)
categorical_features = []
numerical_features = []
for i in col:
    if len(df[i].unique()) > 6:
        numerical_features.append(i)
    else:
        categorical_features.append(i)

print('Data Kategorik :',*categorical_features)
print('Data Numerik :',*numerical_features)
```

```
Data Kategorik : Gender Tujuan Lama_Investasi Pemahaman_Investasi Pendapatan_dan_Stabilitas_Keuangan Pengalaman_Investasi Frekuensi_Riset Respon_Terhadap_Volatilitas Instrumen_Favorit Profil_Risiko
Data Numerik : Dana_Investasi
```

Gambar 3. 8 Pemisahan Variabel Kategorik dan Numerik

Terlihat bahwa variabel kategorik tersimpan di ‘categorical_features’ dan variabel numerik tersimpan di ‘numerical_features’.

8. Deteksi dan Penanganan Outlier
   - Menghitung statistik deskriptif dari kolom 'Dana_Investasi' sebelum melakukan penanganan outlier

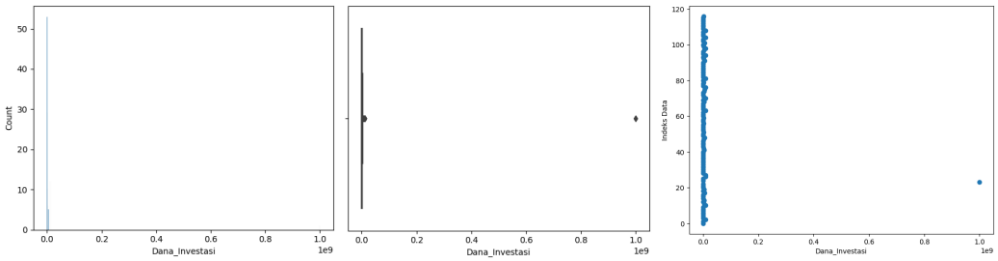
```
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
Dana_Investasi	118.00	10417889.82	91924339.87	0.00	200000.00	1000000.00	2000000.00	1000000000.00

Gambar 3. 9 Statistika Deskriptif Variabel Numerik

Menggunakan metode describe().T pada data frame untuk menghitung statistik deskriptif dari kolom ‘Dana_Investasi’ sebelum penanganan outlier. Terlihat bahwa dana investasi memiliki 118 baris, lalu nilai meannya Rp10.417.889.82, standar deviasi sebesar Rp91.924.339.87, nilai terkecilnya Rp0, quartile pertama sebesar Rp200.000, quartile kedua sebesar Rp1.000.000, quartile ketiga sebesar Rp2.000.000, dan nilai terbesarnya adalah Rp1.000.000.000.

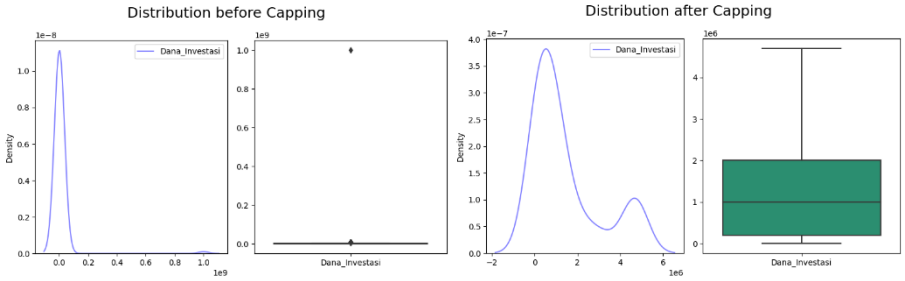
   - Handling Outlier



Gambar 3. 10 Visualisasi Outlier

Pertama kita memvisualisasikan data untuk cek apakah ada outlier atau tidak pada kolom 'Dana_Investasi'. Terlihat bahwa ada outlier pada kolom tersebut.

Selanjutnya, kita melakukan capping atau pembatasan pada outlier dalam kolom "Dana_Investasi" dengan menggunakan metode IQR. Dimana nilai atas outlier akan diganti dengan ‘max_limit’ dan nilai bawah outlier akan diganti dengan ‘min_limit’.



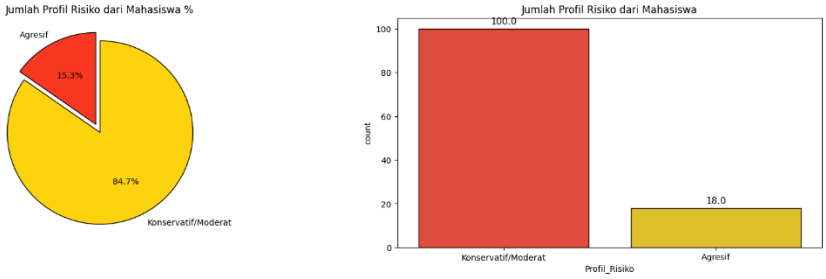
Gambar 3. 11 Visualisasi setelah *Handling Outlier*

Dari gambar di atas, terlihat bahwa data yang memiliki outlier sudah di gantikan dengan nilai dari IQR, sehingga tidak terlihat outlier yang sangat signifikan. Sebagai gantinya, distribusi dari variabel “Dana_Investasi” terlihat bimodal atau memiliki dua puncak.

BAB IV
HASIL DAN PEMBAHASAN

4.1. Exploratory Data Analysis (EDA)

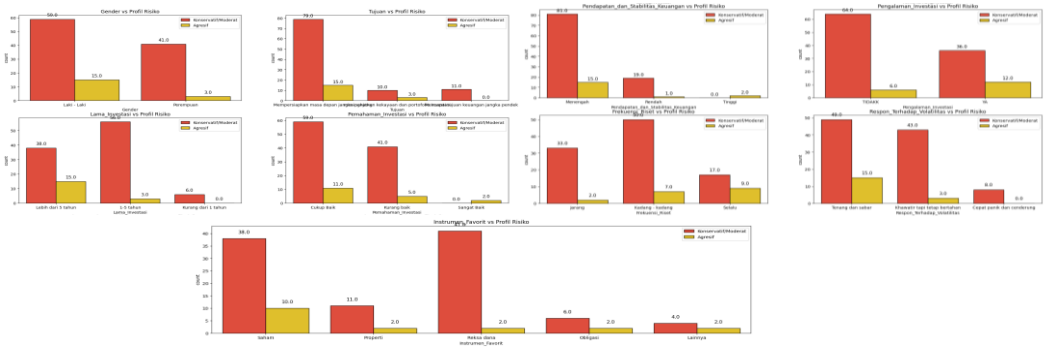
4.1.1. Target Visualisation Distribution



Gambar 4. 1 Visualisasi Target Variabel

Dari hasil visualisasi diatas dapat kita lihat terjadinya ketidakseimbangan (*imbalance*) karena pemilihan kategori agresif hanya sebesar 15.3% sementara 84.7% ada pada kategori Konservatif/Moderat. Hal ini dapat mempengaruhi analisis dan pemodelan, di mana kategori mayoritas "Konservatif/Moderat" akan memiliki pengaruh lebih besar daripada kategori minoritas "Agresif".

4.1.2. Countplot Variabel Kategorik Vs Profil Risiko



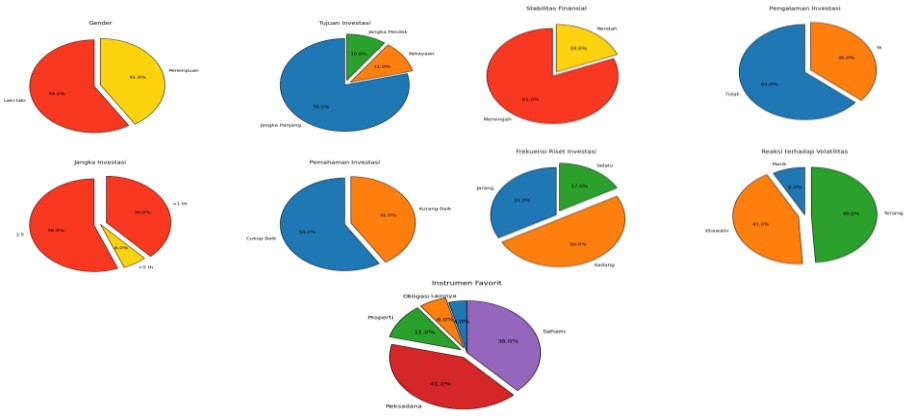
Gambar 4. 2 Visualisasi Variabel Kategorik

Setiap countplot tersebut menunjukkan jumlah pengamatan dalam setiap kategori variabel, dibagi berdasarkan Profil Risiko yaitu "Konservatif/moderat" dan "Agresif". Berikut merupakan penjelasan dari tiap countplot diatas.

- Gender Vs Profil Risiko: Populasi laki-laki dan perempuan cenderung memiliki profil risiko "Konservatif/moderat" lebih banyak daripada "Agresif".
- Tujuan Vs Profil Risiko : Populasi yang berinvestasi untuk mempersiapkan masa depan jangka panjang, meningkatkan kekayaan dan portofolio investasi, serta mencapai tujuan keuangan jangka pendek, cenderung memiliki profil risiko "Konservatif/moderat".

- Lama_Investasi Vs Profil Risiko : Populasi dengan rencana investasi lebih dari 5 tahun, 1 sampai 5 tahun, dan kurang dari 1 tahun cenderung memiliki profil risiko "Konservatif/moderat".
- Pemahaman_Investasi Vs Profil Risiko : Populasi yang memiliki pemahaman konsep investasi dan instrumen-instrumen investasi cukup baik dan kurang baik cenderung memiliki profil risiko “Konservatif/moderat” sedangkan yang memiliki tingkat pemahaman konsep sangat baik cenderung memiliki profil risiko “Agresif”.
- Pendapatan_dan_Stabilitas_Keuangan Vs Profil Risiko : Populasi yang memiliki tingkat pendapatan dan stabilitas keuangan menengah dan rendah cenderung memiliki profil risiko “Konservatif/moderat” sedangkan yang memiliki tingkat pendapatan dan stabilitas keuangan tinggi cenderung memiliki profil risiko “Agresif”.
- Pengalaman_Investasi Vs Profil Risiko : Populasi yang memiliki pengalaman sebelumnya dalam melakukan investasi dan yang tidak memiliki pengalaman juga cenderung memiliki profil risiko “Konservatif/moderat”.
- Frekuensi_Riset : Populasi yang selalu, kadang-kadang, dan jarang melakukan riset dan analisis sebelum melakukan investasi cenderung memiliki profil risiko “Konservatif/moderat” daripada “Agresif”.
- Respon_Terhadap_Volatilitas Vs Profil Risiko : Populasi yang menghadapi volatilitas pasar dan fluktuasi nilai investasi dengan tenang dan sabar, khawatir tapi tetap bertahan, dan cepat panik dan cenderung memiliki profil risiko “Konservatif/moderat”.
- Instrumen_Favorit Vs Profil Risiko : Populasi yang lebih memilih instrumen investasi saham, properti, reksadana, obligasi dan lainya cenderung memiliki profil risiko “Konservatif/moderat”.

4.1.3. Pie Chart Variabel Kategorik terhadap Konservatif/Moderat



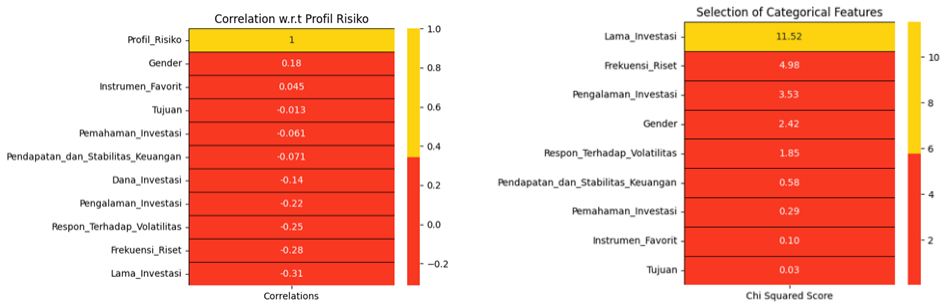
Gambar 4. 3 Visualisasi Variabel Kategorik di Konservatif/Moderat

Setiap pie chart diatas menunjukkan proporsi persentase dari setiap kategori variabel, khususnya untuk kelompok dengan profil risiko yang "Konservatif/Moderat". Berikut merupakan penjelasan dari tiap pie chart tersebut.

- Gender : Dari semua kelompok, laki-laki memiliki persentase lebih besar dari perempuan yaitu sebesar 59%.
- Tujuan : Dari semua kelompok, yang memiliki tujuan jangka panjang memiliki persentase lebih besar dari semuanya yaitu sebesar 79%.
- Lama_Investasi : Dari semua kelompok, yang memiliki rencana investasi lebih dari 1 sampai 5 tahun memiliki persentase lebih besar dari semuanya yaitu sebesar 56%.
- Pemahaman_Investasi : Dari semua kelompok, yang memilikii pemahaman yang cukup baik memiliki persentase lebih besar yaitu sebesar 59% sedangkan yang memiliki pemahaman kurang baik memiliki persentase sebesar 41%.
- Pendapatan_dan_Stabilitas_Keuangan : Dari semua kelompok, tingkat menengah memiliki persentase lebih besar yaitu sebesar 81% sedangkan yang memiliki tingkat rendah sebesar 19%.
- Pengalaman_Investasi : Dari semua kelompok, yang tidak memiliki pengalaman sebelumnya dalam melakukan investasi memiliki persentase lebih besar yaitu sebesar 64% dan yang memiliki pengalaman memiliki persentase sebesar 36%.
- Frekuensi_Riset : Dari semua kelompok, yang kadang-kadang melakukan riset dan analisis sebelum melakukan investasi memiliki persentase yang paling besar yaitu sebesar 50%.
- Respon_Terhadap_Volatilitas : Dari semua kelompok, yang menghadapi volatilitas pasar dan fluktuasi nilai investasi dengan tenang dan sabar memiliki persentase yang paling besar yaitu sebesar 49%.
- Instrumen_Favorit : Dari semua kelompok, yang lebih memilih instrumen investasi reksadana memiliki persentase yang paling besar yaitu sebesar 41%.

4.1.5. Feature Selection

Disini, kita melakukan analisis dari uji korelasi dan uji chi-squared terhadap variabel target dalam dataset.



Gambar 4. 4 Visualisasi Analisis Feature

Berdasarkan analisis kami, terdapat keyakinan bahwa dari 5 fitur teratas yang menunjukkan korelasi dan signifikansi yang kuat terhadap variabel target 'Profil_Risiko' yaitu variabel 'Lama_Investasi', 'Frekuensi_Riset', 'Pengalaman_Investasi', 'Gender', dan 'Dana_Investasi', fitur-fitur tersebut sangat penting dalam memprediksi profil risiko. Dengan mempertimbangkan faktor-faktor tersebut, kita dapat mengidentifikasi variabel-variabel yang paling berpengaruh dalam menentukan profil risiko mahasiswa. Dengan menggunakan fitur-fitur tersebut, kita dapat mengoptimalkan model prediksi profil risiko dan memberikan rekomendasi yang lebih akurat.

4.2. Pembuatan Model

4.2.1 Data Splitting atau Train-Test Set

Setelah melakukan feature selection, kami membagi dataset menjadi dua bagian, yaitu data training untuk melatih model dan data testing untuk menguji model. Pembagian data dilakukan dengan ukuran 30% untuk data testing (36 baris) dan 70% untuk data training (82 baris). Selain itu, parameter yang digunakan adalah random state dengan nilai 2..

4.2.2 Melakukan Modelling

Dalam pembuatan model, kami menggunakan library ‘sklearn’ dan ‘xgboost’. Untuk mempermudah pembuatan model dan evaluasi, kami membuat fungsi untuk melakukan pembuatan model dan evaluasi model sebagai berikut:

```
def model(classifier):

    classifier.fit(x_train,y_train)
    prediction = classifier.predict(x_test)
    cv = RepeatedStratifiedKFold(n_splits = 10,n_repeats = 3,random_state = 1)
    print("Accuracy : ",'{0:.2%}'.format(accuracy_score(y_test,prediction)))
    print("Cross Validation Score : ",'{0:.2%}'.format(cross_val_score(classifier,x_train,y_train,cv = cv,scoring = 'roc_auc').mean()))
    print("ROC_AUC Score : ",'{0:.2%}'.format(roc_auc_score(y_test,prediction)))
    RocCurveDisplay.from_estimator(classifier, x_test,y_test)
    plt.title('ROC_AUC_Plot')
    plt.show()
```

Gambar 4. 5 Fungsi model

```
def model_evaluation(classifier):

    # Confusion Matrix
    cm = confusion_matrix(y_test,classifier.predict(x_test))
    names = ['True Neg','False Pos','False Neg','True Pos']
    counts = [value for value in cm.flatten()]
    percentages = ['{0:.2%}'.format(value) for value in cm.flatten()/np.sum(cm)]
    labels = [f'{v1}\n{v2}\n{v3}' for v1, v2, v3 in zip(names,counts,percentages)]
    labels = np.asarray(labels).reshape(2,2)
    sns.heatmap(cm,annot = labels,cmap = colors,fmt ='')

    # Classification Report
    print(classification_report(y_test,classifier.predict(x_test)))
```

Gambar 4. 6 Fungsi model_evaluation

Fungsi ‘model’ disini mencakup mulai dari pelatihan model, prediksi model, pengujian Cross-Validation, pengujian ROC_AUC Score, dan menampilkan visualisasi dari ROC Curve. Sedangkan fungsi ‘model_evaluation’ berisikan confusion matrix, dan classification report dari classifier yang digunakan.

Selanjutnya, kita akan melakukan *model building* dengan algoritma sebagai berikut:

4.2.2.1. Logistic Regression

a. Model Fitting

```
classifier_lr = LogisticRegression(random_state = 0,C=10,penalty= 'l2')
model(classifier_lr)
```

Gambar 4. 7 Model Fitting Logistic Regression

Disini, kita menggunakan metode Logistic Regression dengan parameter random state = 0, C=10, dan penalty= 'l2'. Setelah melakukan fungsi di atas, didapatkan hasil sebagai berikut:

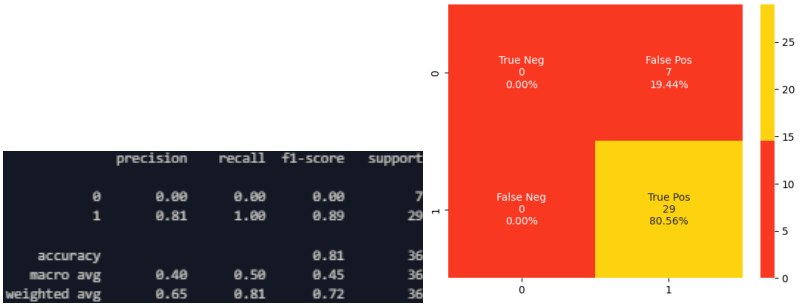
- Accuracy : 80.56%
- Cross Validation Score : 24.17%
- ROC_AUC Score : 50.00%

b. Model Evaluation

```
model_evaluation(classifier_lr)
```

Gambar 4. 8 Model Evaluation Logistic Regression

Setelah melakukan fungsi di atas, didapatkan hasil evaluasi model sebagai berikut:



Gambar 4. 9 Confusion Matrix Logistic Regression

4.2.2.2 Decision Tree

a. Model Fitting

```
classifier_dt = DecisionTreeClassifier(random_state = 1000,max_depth =
4,min_samples_leaf = 1)
model(classifier_dt)
```

Gambar 4. 10 Model Fitting Decision Tree

Disini, kita menggunakan metode Decision Tree dengan parameter random state = 1000, max_depth=4, dan min_samples_leaf=1. Setelah melakukan fungsi di atas, didapatkan hasil sebagai berikut:

- Accuracy : 69.44%
- Cross Validation Score : 67.77%
- ROC_AUC Score : 59.36%

b. Model Evaluation

```
model_evaluation(classifier_dt)
```

Gambar 4. 11 Model Evaluation Decision Tree

Setelah melakukan fungsi di atas, didapatkan hasil evaluasi model sebagai berikut:

Gambar 4. 12 Confusion Matrix Decision Tree

4.2.2.3. Random Forest

a. Model Fitting

```
classifier_rf = RandomForestClassifier(max_depth = 4,random_state = 0)
model(classifier_rf)
```

Gambar 4. 13 Model Random Forest

Disini, kita menggunakan metode Random Forest dengan parameter max_depth=4, dan random_state=0. Setelah melakukan fungsi di atas, didapatkan hasil sebagai berikut:

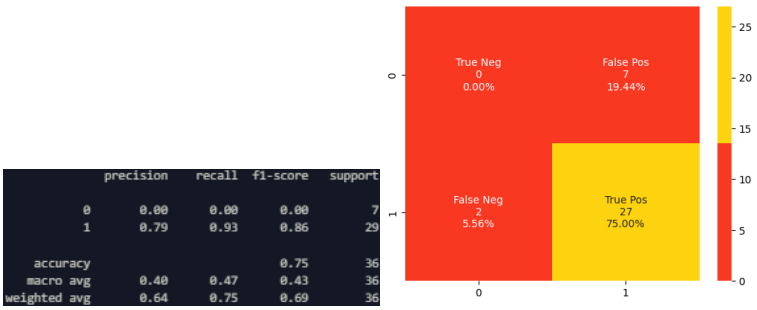
- Accuracy : 75.00%
- Cross Validation Score : 82.08%
- ROC_AUC Score : 46.55%

b. Model Evaluation

```
model_evaluation(classifier_rf)
```

Gambar 4. 14 Model Evaluation Random Forest

Setelah melakukan fungsi di atas, didapatkan hasil evaluasi model sebagai berikut:



Gambar 4. 15 Confusion Matrix Random Forest

4.2.2.4. K-Nearest Neighbour

a. Model Fitting

```
classifier_knn = KNeighborsClassifier(leaf_size = 1, n_neighbors = 3,p = 1)
model(classifier_knn)
```

Gambar 4. 16 Model KNN

Disini, kita menggunakan metode K-Nearest Neighbour dengan parameter leaf_size=1, n_neighbors=3, dan p=1. Setelah melakukan fungsi di atas, didapatkan hasil sebagai berikut:

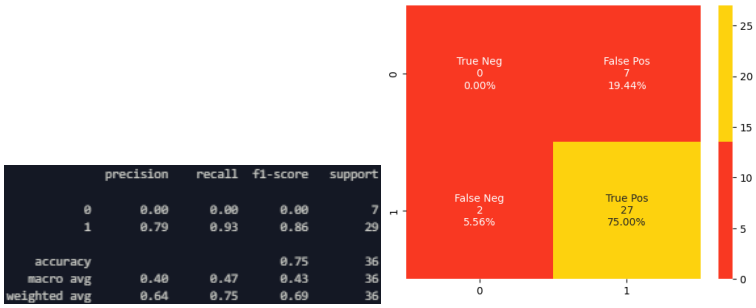
- Accuracy : 75.00%
- Cross Validation Score : 80.86%
- ROC_AUC Score : 46.55%

b. Model Evaluation

```
model_evaluation(classifier_knn)
```

Gambar 4. 17 Model Evaluation KNN

Setelah melakukan fungsi di atas, didapatkan hasil evaluasi model sebagai berikut:



	precision	recall	f1-score	support
0	0.00	0.00	0.00	7
1	0.79	0.93	0.86	29
accuracy			0.75	36
macro avg	0.40	0.47	0.43	36
weighted avg	0.64	0.75	0.69	36

Gambar 4. 18 Confusion Matrix KNN

4.2.2.5. XGBoost

a. Model Fitting

```
classifier_xgb = XGBClassifier(use_label_encoder=False, eval_metric='mlogloss')
model(classifier_xgb)
```

Gambar 4. 19 Model XGBoost

Disini, kita menggunakan metode XGBoost dengan parameter use_label_encoder=False, dan eval_metric= 'mlogloss'. Setelah melakukan fungsi di atas, didapatkan hasil sebagai berikut:

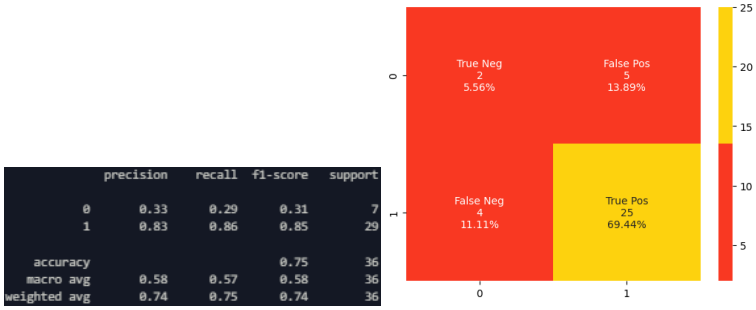
- Accuracy : 75.00%
- Cross Validation Score : 83.10%
- ROC_AUC Score : 57.39%

b. Model Evaluation

```
model_evaluation(classifier_xgb)
```

Gambar 4. 20 Model Evaluation XGBoost

Setelah melakukan fungsi di atas, didapatkan hasil evaluasi model sebagai berikut:



	precision	recall	f1-score	support
0	0.33	0.29	0.31	7
1	0.83	0.86	0.85	29
accuracy			0.75	36
macro avg	0.58	0.57	0.58	36
weighted avg	0.74	0.75	0.74	36

Gambar 4. 21 Confusion Matrix XGBoost

4.3. Pemilihan Model Terbaik

Setelah melakukan berbagai model dengan berbagai algoritma yang tersedia, kita dapat menentukan model terbaik dengan metric-metric yang dapat digunakan sebagai perbandingan antar model. Berikut adalah perbandingan setiap model dengan metric-metricnya:

Nama Classifier	Accuracy	Precision	Recall	F1-Score	CV Score	ROC_AUC Score
Logistic Regression	0.81	0.81	1.00	0.89	0.24	0.50
Decision Tree	0.69	0.85	0.76	0.80	0.68	0.59
Random Forest	0.75	0.79	0.93	0.86	0.82	0.47
KNN	0.75	0.79	0.93	0.86	0.81	0.47
XGBoost	0.75	0.83	0.86	0.85	0.83	0.57

Tabel 4. 1 Tabel Pemilihan Model

Berdasarkan tabel di atas, dan dengan kondisi variabel target kami yang memiliki imbalance, maka metric yang paling tepat untuk digunakan adalah **Recall** dan **CV-Score**, karena metric tersebut baik untuk *imbalance data* seperti yang ada di dataset kami. Sehingga, model terbaik adalah dari Random Forest dengan Recall dan CV Score sebesar 0.93 dan 0.82.

BAB V
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan analisa dari setiap variabel yang sudah dipaparkan di bagian sebelumnya, kita dapat menyimpulkan bahwa untuk kelas Konservatif/Moderat, kebanyakan responden memiliki jenis kelamin Laki-laki, memiliki tujuan jangka panjang dibanding jangka pendek, memiliki tingkat pemahaman finansial yang cukup baik, memiliki tingkat finansial yang menengah, tidak memiliki pengalaman investasi, terkadang melakukan riset terlebih dahulu sebelum melaksanakan kegiatan investasi, memiliki reaksi tenang apabila terjadi volatilitas terhadap harga suatu instrumen investasi, dan yang terakhir memiliki instrumen favorit dengan urutan berikut: Reksa Dana, Saham, Properti, Obligasi, dan Lainnya.

Dalam menentukan variabel-variabel yang berpengaruh dalam profil risiko investasi, kita perlu melakukan uji chi-squared dan analisis korelasi terhadap variabel target. Dari kedua analisis yang telah dilakukan, kita dapat memilih 5 variabel dengan rincian sebagai berikut, yaitu: 'Lama_Investasi', 'Frekuensi_Riset', 'Pengalaman_Investasi', 'Gender', 'Dana_Investasi'. Meskipun tidak termasuk dalam 5 tertinggi di analisis korelasi, variabel 'Dana_Investasi' juga terpilih karena merupakan satu-satunya variabel numerik di dataset ini.

Model yang tepat atau terbaik dapat dipilih dengan penilaian metric Recall dan CV-Score karena data kami memiliki *imbalance* pada variabel target. Berdasarkan Tabel 4.1, model terbaik dalam penelitian ini adalah model Random Forest dengan Recall dan CV Score sebesar 0.93 dan 0.82.

5.2. Saran

Adapun saran yang dapat kami berikan untuk penelitian ini agar penelitian selanjutnya lebih efisien adalah:

1. Melakukan pengambilan sampel dengan ruang lingkup yang lebih besar.
2. Dapat lebih siap untuk struktur dalam melakukan penelitian.
3. Diharapkan atau disarankan untuk menambahkan beberapa uji/algoritma klasifikasi yang lebih lanjut dari penelitian ini.

DAFTAR PUSTAKA

Achmad, B. D. M. & Slamat, F., 2012. Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree. *Jurnal IPTEK*, 16(1), p. 17.

Ambarwati, P. D., Pinilih, S. S. & Astuti, R. T., 2017. Gambaran Tingkat Stres Mahasiswa. *Jurnal Keperawatan*, 5(1), p. 40.

Bimantara, A. & Dina, T. A., 2018. Klasifikasi Web Berbahaya Menggunakan Metode Logistic Regression. *Computer Science and ICT*, 4(1), p. 173.

Cahyono, H., 2019. Peran Mahasiswa di Masyarakat. *Jurnal Pengabdian Masyarakat*, 1(1), p. 32.

Dewati, A. A. & M., 2021. DETERMINAN MINA MAHASISWA BERINVESTASI PADA PASAR MODAL. *Jurnal Universitas Islam Indonesia*, 21 06, 4(1), pp. 45 - 60.

Dewi, S., 2016. KOMPARASI 5 METODE ALGORITMA KLASIFIKASI DATA MINING PADA PREDIKSI KEBERHASILAN PEMASARAN PRODUK LAYANAN PERBANKAN. *Jurnal Techno Nusa Mandiri*, 8(1), p. 60.

Givari, M. R., Sulaeman, M. R. & Umaidah, Y., 2022. Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *JURNAL NUANSA INFORMATIKA*, 16(1).

Karo, I. M., 2020. Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan. *Journal of Software Engineering, Information and Communication Technology*, 1(1), p. 10.

KRISTIANTO, Y., 2015. ANALISIS WEALTH MANAGEMENT, PERLINDUNGAN DAN PELESTARIAN ASET PADA LEMBAGA KEUSKUPAN PANGKALPINANG. *Jurnal Universitas Atma Jaya Yogyakarta*, 17 12, 1(1), pp. 95 - 110.

Mandagie, Y. R., Febrianti, M. & Fujianti, L., 2020. ANALISIS PENGARUH LITERASI KEUANGAN, PENGALAMAN INVESTASI DAN TOLERANSI RISIKO TERHADAP KEPUTUSAN INVESTASI. *Jurnal Universitas Pancasila*, 1(1), pp. 35-47.

Mardi, Y., 2017. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), pp. 213-219.

Nabila, V. & S., 2022. PENGARUH LITERASI KEUANGAN DAN TOLERANSI RISIKO TERHADAP KEPUTUSAN INVESTASI TABUNGAN EMAS (STUDI KASUS NASABAH DI PT PEGADAIAN (PERSERO) CABANG KRAMAT JATI). *JIMA*, 2(1), p. 32.

Nasution, D. A., Khotimah, H. H. & Chamidah, N., 2019. PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN. *(Journal of Computer Engineering System and Science,* 4(1).

Papilaya, J. O. & Huliselan, N., 2016. Identifikasi Gaya Belajar Mahasiswa. *Jurnal Psikologi Undip,* 15(1), pp. 56-63.

P., B. & H., 2017. Pengaruh Faktor Kepribadian terhadap Toleransi Risiko Keputusan Investasi Saham. *JUrnal Sains dan Seni,* 6(1), p. 7.

Primajaya, A. & Sari, B. N., 2018. Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM),* 1(1), p. 27.

Qadrini, L., Seppewali, A. & Aina, A., 2021. DECISION TREE DAN ADABOOST PADA KLASIFIKASI PENERIMA PROGRAM BANTUAN SOSIAL. *Jurnal Inovasi Pendidikan,* 2(7), p. 1959.

S., 2020. KEAMANAN INFORMASI DAN MANAJEMEN RISIKO PERPUSTAKAAN. *Jurnal Perpustakaan Universitas Islam Indonesia,* 3(1), pp. 105 - 120.

Siburian, V. W. & Mulyana, I. E., 2018. Prediksi Harga Ponsel Menggunakan Metode Random Forest. *Computer Science and ICT,* 4(1), p. 144.

WIbawa, A. P., Purnama, M. G., Akbar, M. F. & Dwiyanto, F. A., 2018. Metode-metode klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi,* 3(1), p. 134.

LAMPIRAN

Link Code:

- EDA & Modelling

https://colab.research.google.com/drive/1eyAjXS6t1DCbKoKLjDluuO-dML97ySN_?usp=sharing

- Data Prepocessing

https://colab.research.google.com/drive/1vjNJStIPX3d7Gpxex04BgJmXIrE2UTN0?usp=sharing

Link Kuisioner:

https://docs.google.com/forms/d/e/1FAIpQLSf1uPVF8M2i6JFJJEnL0Xp9-TE_QTTjIf8USaf0WfGPpUqrlA/viewform?usp=sharing