



Visualisasi Data Finansial Investasi : Berdasarkan *Age* (Umur) & *Gender* (Jenis Kelamin) dalam memilih *Debentures* (Obligasi) Melalui pendekatan Data Science Menggunakan Google Colab.

Dimas Wahyu Pangestu

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Jojok Dwiridotjahjono

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: 20042010196@student.upnjatim.ac.id

Abstract: *In this digital era, investment and the stock market are quite interesting topics to discuss in the financial world. Developments in technology, information and communication have made it possible for everyone to collect and access financial data in large, wide and varied volumes of data or commonly referred to as "Big Data". In addition, advances in technology have given rise to artificial intelligence and machine learning, which are very helpful in descriptive analysis of investments and the stock market. Sophisticated algorithms can be used to identify patterns, trends and correlations that are not visible manually in financial data. This study aims to analyze the finance dataset obtained from the Kaggle platform using Google Colaboratory and using 4 methods, namely use case analysis, K-NN, K-Means and Decision Tree..*

Keywords: *Data science, Algorithm, Financial, Data analysis*

Abstrak: Pada era digital ini, investasi dan pasar saham merupakan salah satu topik yang cukup menarik untuk dibahas dalam dunia keuangan. Perkembangan teknologi, informasi, dan komunikasi telah memungkinkan semua orang untuk mengumpulkan serta mengakses data finansial dalam volume data yang besar, luas, dan beragam atau biasa disebut dengan “*Big Data*”. Selain itu, dengan kemajuan dalam teknologi telah memunculkan kecerdasan buatan (*artificial intelligence*) dan pembelajaran mesin (*machine learning*), yang sangat membantu dalam analisis deskriptif investasi dan pasar saham. Algoritma canggih dapat digunakan untuk mengidentifikasi pola, tren, dan korelasi yang tidak terlihat secara manual dalam data finansial. Penelitian ini bertujuan untuk menganalisis *dataset finance* yang diperoleh dari platform *Kaggle* dengan menggunakan *Google Colaboratory* dan menggunakan 4 metode yaitu analisis *use case*, *K-NN*, *K-Means* dan *Decision Tree*.

Kata kunci: Data sains, Algoritma, Finansial, Data analisis

LATAR BELAKANG

Pada era digital saat ini, data memiliki peran yang sangat penting dalam melakukan analisis penjualan, tren, perilaku masyarakat, korelasi antar variabel, regresi dll. Namun, perlu ditekankan bahwa data yang tidak bisa diolah dengan baik akan mengakibatkan dampak yang merugikan dan akan menjadi bumerang bagi para peneliti dalam mengambil suatu keputusan. Meskipun demikian, data akan menjadi aset yang sangat berharga apa bila diolah dengan benar. Data yang besar dan kompleks atau biasa disebut dengan “*Big Data*” Menjadi salah satu tantangan tersendiri bagi para peneliti dalam mengumpulkan data mentah dan kemudian diolah menjadi data berharga. Menurut Hizriansyah,dkk (2023) “Penggunaan data menjadi sangat penting untuk monitoring dan mendukung proses pengambilan keputusan masyarakat yang semakin kompleks”.

Data science memiliki beberapa pendekatan untuk menganalisis suatu data maupun permasalahan yang ada, beberapa diantaranya seperti, analisis *use case*, K-NN, K-Means dan *Decision Tree*, keempat pendekatan tersebut merupakan suatu metode analisis yang biasa digunakan dalam data sains yang bertujuan untuk melakukan visualisasi data, *clustering*, regresi, korelasi antar variabel dan salah satu yang mendasari pengambilan keputusan.

Pada era teknologi yang sudah semakin berkembang pesat, tidak hanya data dan data science saja yang menarik perhatian, namun perkembangan program dan software juga menjadi hal yang menarik untuk dibahas, salah satunya Google Colab dan *Artificial Intelligent (AI)* ChatGpt yang memberikan dampak yang cukup besar dalam dunia pengolahan data. Dengan menggunakan Google Colab dan Chatgpt para peneliti maupun professional akan semakin terbantu dengan fitur-fitur yang disajikan, salah satunya Google Colab yang memberikan tempat atau *space* untuk para peneliti dalam melakukan analisis data dan ChatGpt yang memberikan bantuan untuk memecahkan *problem* yang dihadapi saat melakukan analisis. Menurut Peter Norvig dan Stuart Russell (2011), AI adalah “studi tentang bagaimana membuat komputer melakukan hal-hal yang pada saat ini orang lebih baik dalam melakukannya”. Mereka juga menyatakan bahwa AI melibatkan penggunaan teknologi untuk membuat mesin yang cerdas dan mampu belajar dari pengalaman.

Pendekatan data science dalam visualisasi dapat memudahkan para peneliti dalam menganalisis sebuah data yang kompleks, dengan adanya visualisasi bahkan orang awam pun bisa membaca dan memahami data dengan mudah, oleh karena itu visualisasi data memiliki peranan penting dalam melakukan analisis sebuah data maupun untuk menunjukkan data kepada customer ataupun pimpinan. Menurut Mc Cormick, (1987) mendefinisikan visualisasi sebagai metode penggunaan komputer untuk mentransformasikan simbol menjadi geometrik

dan memungkinkan peneliti dalam hal mengamati simulasi komputasi yang dapat memperkaya proses penemuan ilmiah sehingga dapat mengembangkan pemahaman yang lebih dalam dan tak terduga. Dalam penelitian ini, akan membahas mengenai visualisasi data finansial berdasarkan *age* (umur) dan *gender* (jenis kelamin) yang berfokus dalam pemilihan *debentures* (obligasi) menggunakan google colab menggunakan beberapa metode seperti analisis *use case*, K-NN, K-Means dan *Decision Tree*.

METODE PENELITIAN

Penelitian deskriptif ini menggunakan metode analisis berbasis data dan dataset diperoleh dari www.kaggle.com yang merupakan sebuah platform untuk mengumpulkan dataset *real* maupun data *dummy*. Dataset yang akan dianalisis berisikan data investor dan jenis investasi yang dipilih oleh para investor yang nantinya dataset tersebut akan diolah menggunakan Google Colab dan Program Python. Pada analisis ini penulis menggunakan 4 metode yaitu analisis *use case*, K-NN, K-Means dan *Decision Tree*. Menurut The Liang Gie (1989:26), “Analisis adalah segenap rangkaian perubahan pikiran yang menelaah sesuatu secara mendalam, terutama ketika mempelajari bagian-bagian dari suatu kebulatan untuk mengetahui ciri dari masing-masing bagian serta hubungan dan peranannya dalam keseluruhan yang bulat itu.”. Sedangkan data menurut The Liang Gie Mendefinisikannya dalam 2 pengertian, yaitu:

- Data merupakan hal, peristiwa atau kenyataan apapun yang mengandung sesuatu pengetahuan untuk dijadikan sebagai dasar untuk penyusunan keterangan, pembuatan kesimpulan atau penerapan keputusan.
- Data merupakan sebuah ibarat mentah yang melalui pengolahan tertentu lalu menjadi keterangan (informasi).



Gambar 1. Tahapan Desain dan Metode Penelitian

1. Data Collecting

Data yang kami peroleh berasal dari *www.kaggle.com* yaitu berupa file csv (Comma-Separated Values) dengan judul *Finance_data.csv*. Pengumpulan data, dalam konteks teknologi informasi dan ilmu komputer, merujuk pada proses mengumpulkan informasi atau fakta-fakta yang relevan dari berbagai sumber. *Data collecting* (pengumpulan data) adalah langkah awal dalam analisis data dan penelitian yang memungkinkan untuk mengumpulkan informasi yang diperlukan untuk menjawab pertanyaan atau mencapai tujuan tertentu.

2. Pemodelan Algoritma dan Implementasi

Pemodelan algoritma melibatkan proses merancang dan menggambarkan langkah-langkah yang diperlukan untuk menyelesaikan tugas atau masalah tertentu. Ini melibatkan pemikiran kritis dan analisis terhadap masalah yang akan dianalisis. Dalam implementasi dataset tersebut menggunakan Google Colab, penulis menggunakan beberapa algoritma yang mencakup KNN (K-Nearest Neighbors), K-Means, dan Pohon Keputusan (Decision Tree). Dalam hal ini, penulis menggunakan Google Colab sebagai lingkungan yang memfasilitasi pemodelan dan implementasi algoritma yang terkait dengan dataset tersebut. Algoritma KNN (K-Nearest Neighbors) digunakan untuk masalah klasifikasi dan regresi.

3. Visualisasi dan uji akurasi

Visualisasi data adalah proses mewakili informasi dan pola yang terkandung dalam data secara visual melalui grafik, diagram, peta, atau metode lainnya. “Visualisasi data diperlukan untuk membuat penerima data memahami dengan baik hasil dari pengolahan data” (Windi (2021)). Tujuannya adalah untuk mempermudah pemahaman dan analisis data dengan cara yang lebih intuitif dan dapat dipahami. Beberapa jenis visualisasi data umum termasuk diagram batang, grafik garis, diagram lingkaran, heatmap, scatter plot, dan peta. Setelah dilakukan visualisasi, ditentukanlah akurasi tahu manakah algoritma yang terbaik dalam dataset *Finance_data.csv*.

4. Analisis dan Kesimpulan

Pada tahap terakhir melibatkan hasil dari visualisasi dan pengujian data untuk menarik kesimpulan. Dalam konteks data finansial investasi, kesimpulan yang dapat ditarik seperti terjadinya korelasi antara *age* (umur) dan *gender* (jenis kelamin) dalam pemilihan *debentures* (obligasi) yang artinya umur dan jenis kelamin berpengaruh terhadap pemilihan obligasi.

HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset yang digunakan merupakan data finansial investasi dan pasar saham yang disimpan dalam format CSV (Comma-Separated Values). Dataset ini berisi tentang umur, jenis kelamin para investor, pemilihan jalur investasi oleh investor dan banyaknya jumlah investasi para investor, salah satu jalur investasi yang dipilih yaitu obligasi.

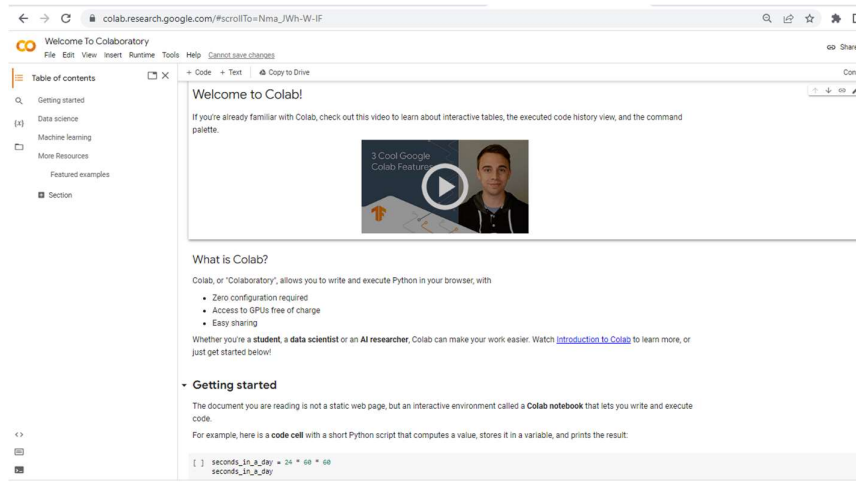
Dataset ini terdiri dari 41 baris data dan 24 kolom data. Penulis akan mengambil 10 kolom pertama dari dataset sebagai berikut:

Tabel 1. Dataset yang digunakan

No	Kolom
1	Gender
2	Age
3	Investment Avenues
4	Mutual_Funds
5	Equity_Market
6	Debentures
7	Government_Bonds
8	Fixed_Deposits
9	PPF
10	Gold

Pada analisis deskriptif ini penulis melakukan analisis menggunakan 4 metode yaitu analisis *use case*, K-NN, K-Means dan *Decision Tree* yang terdapat tahapan-tahapan dalam melakukan analisis diantaranya *data collecting*, *data pre-processing*, *data mining*, dan *data visualization*. Dalam analisis ini penulis menggunakan beberapa tools/software yaitu:

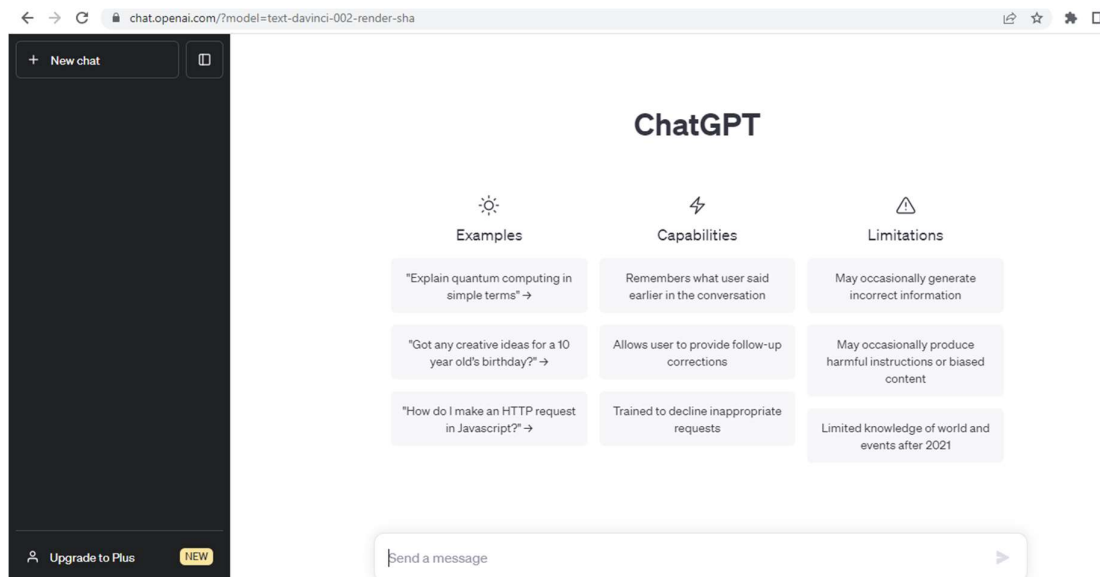
1. Google Colab



Gambar 2. Google colab interface

Google colab atau Colaboratory (“Colab” singkatnya) adalah alat analisis data dan pembelajaran mesin atau *machine learning* yang memungkinkan penulis menggabungkan kode Python yang dapat dieksekusi dan teks lengkap dengan grafik, gambar, HTML, LaTeX, dan lainnya menjadi satu dokumen yang disimpan di Google Drive.

2. ChatGpt



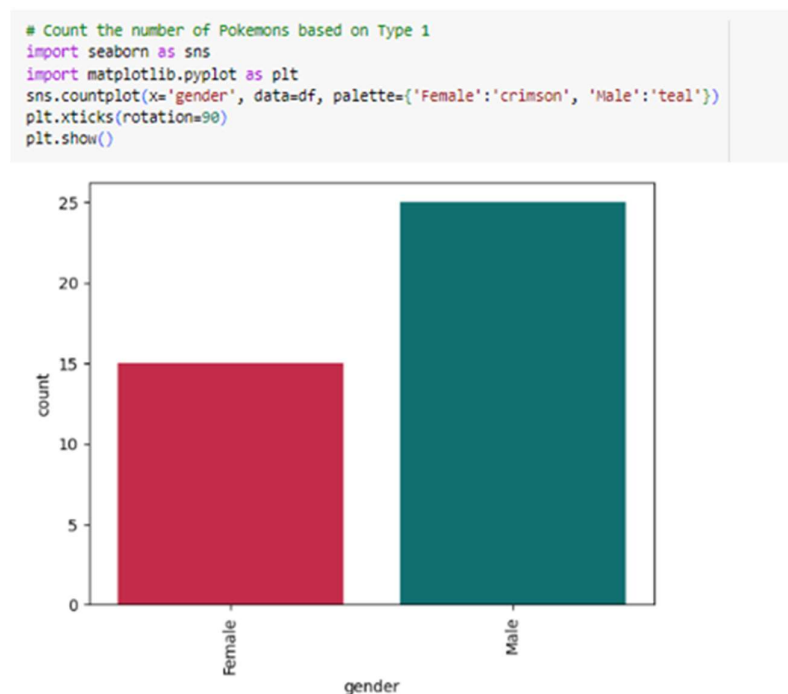
Gambar 3. ChatGpt Interface

Chat gpt merupakan kecerdasan buatan atau *artificial intelligence* atau AI yang dibuat oleh perusahaan AI bernama Open AI. ChatGPT adalah model saudara dari InstructGPT, yang dilatih untuk mengikuti instruksi dalam sebuah permintaan dan memberikan respons yang detail.

Metode analisis

1. Analisis *use case*

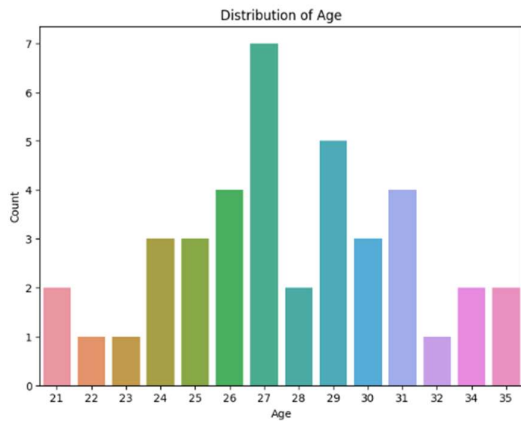
Analisis *use case* merupakan analisis yang bertujuan untuk memudahkan penulis dalam melakukan analisis. Dalam analisis *use case* akan berisikan gambaran umum mengenai data atau dataset yang digunakan yang didalamnya terdapat visualisasi data yang diperlukan untuk melakukan analisis, berikut hasil dari analisis *use case*:



Gambar 4. Visualisasi perbandingan jumlah *gender*

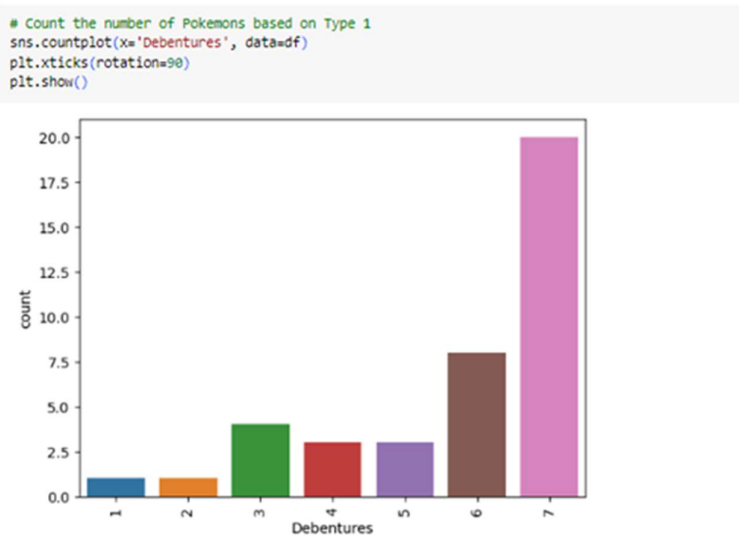
Visualisasi diatas menggunakan bentuk bar chart, menunjukkan perbandingan Jumlah *gender*. Pada visualisasi diatas menunjukkan *gender Male* (laki-laki) lebih banyak daripada jumlah *gender Female* (Perempuan) dengan perbandingan 25:15.

```
fig, axes = plt.subplots(1, 1, figsize=(8, 6))
sns.countplot(x='age', data=df, ax=axes)
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Distribution of Age')
plt.show()
```



Gambar 5. Visualisasi umur para investor

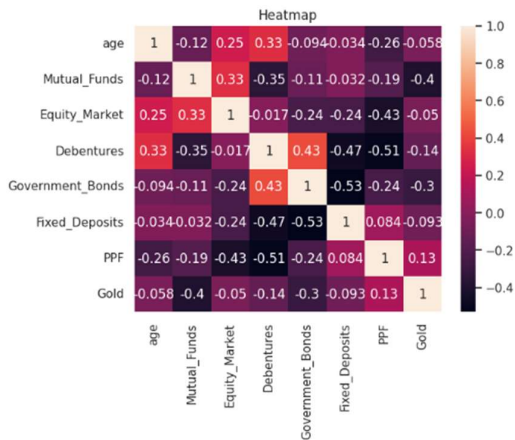
Visualisasi diatas menggunakan bentuk bar chart, menunjukkan bahwa umur dari para investor yaitu 21-35 tahun, dan investor kebanyakan berumur 27 tahun dengan jumlah 7 investor.



Gambar 6. Visualisasi jumlah debentures yang dimiliki investor

Visualisasi diatas menunjukkan banyaknya debentures yang dimiliki oleh investor, dengan range jumlah 1-7, angka 7 menunjukkan jumlah debentures terbanyak yang dimiliki investor. Visualisasi diatas juga menunjukkan bahwa investor yang memiliki jumlah debentures 7 sebanyak 20 investor.


```
#Heatmap of Correlation Matrix
corr_matrix = df.corr()
sns.heatmap(corr_matrix, annot=True)
plt.title('Heatmap')
plt.show()
```



Gambar 7. Visualisasi korelasi antar atribut

Visualisasi diatas menggunakan bentuk heatmap yang menunjukkan nilai dari korelasi antar attribute, semakin gelap warna pada nilai, maka semakin tidak adanya korelasi (negatif), begitu juga sebaliknya. Visualisasi diatas menunjukkan adanya korelasi positif antara *age* (umur) dengan *debentures*, nilai korelasi menunjukkan nilai positif dengan nilai 0.33.

2. K-NN (K-Nearest Neighbors)

KNN (K-Nearest Neighbors) adalah salah satu algoritma yang digunakan dalam *machine learning* untuk masalah klasifikasi dan regresi. Algoritma KNN bekerja dengan mencari k-nearest neighbors atau tetangga terdekat dari titik data yang ingin diprediksi berdasarkan jarak Euclidean atau metrik lainnya. Pada dasarnya, algoritma KNN bekerja dengan menghitung jarak antara titik data yang ingin diprediksi dengan semua titik data lain dalam dataset. Kemudian, k-nearest neighbors (k tetangga terdekat) dari titik data tersebut diidentifikasi berdasarkan jarak yang dihitung. Dalam kasus klasifikasi, kelas mayoritas dari tetangga terdekat ini akan digunakan untuk memprediksi kelas dari titik data yang ingin diprediksi. Dalam kasus regresi, nilai rata-rata atau median dari tetangga terdekat ini akan digunakan sebagai prediksi nilai kontinu. Berikut hasil dari analisis KNN:

```
# Menghitung akurasi model
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", accuracy)
```

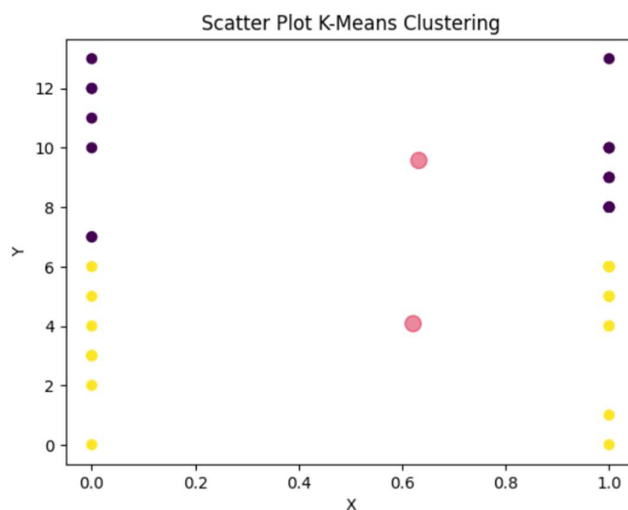
Akurasi: 0.875

Pada metode KNN yang sudah penulis lakukan menghasilkan akurasi dari atribut *gender* dan *age* sebesar 0.875 atau 87.5% merupakan nilai yang cukup tinggi dan menunjukkan adanya regresi antara *gender* dan *age* dalam preferensi pemilihan *debentures* berdasarkan data yang dianalisis.

3. K-Means

K-Means adalah salah satu algoritma yang paling umum digunakan dalam analisis data dan pengelompokan. Tujuan utama dari algoritma K-Means adalah membagi sekumpulan data menjadi beberapa kelompok berdasarkan kesamaan atributnya. Proses pengelompokan ini dilakukan dengan mencari pusat kelompok yang optimal, yang dikenal sebagai centroid. Algoritma ini menggunakan pendekatan iteratif, dimulai dengan menginisialisasi pusat kelompok secara acak dan kemudian mengupdate posisi centroid berulang kali hingga konvergen. Berikut hasil dari analisis K-Means:

```
[ ] # Scatter plot data dengan label kluster
plt.scatter(data[:, 0], data[:, 1], c=labels, cmap='viridis')
plt.scatter(centroids[:, 0], centroids[:, 1], c='crimson', s=100, alpha=0.5)
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Scatter Plot K-Means Clustering')
plt.show()
```



Gambar 8. Visualisasi K-Means

Visualisasi diatas dengan menampilkan hasil clustering menggunakan metode K-Means. Secara keseluruhan, menggambar scatter plot untuk data yang telah dikelompokkan dengan menggunakan algoritma K-Means. Titik-titik data diberi warna berdasarkan label atau kelas yang dihasilkan oleh algoritma, sementara centroid ditampilkan dalam warna yang berbeda. Titik pusat cluster diberi warna crimson, titik pusat cluster ini merepresentasikan titik pusat dari setiap cluster yang telah ditentukan oleh model K-Means clustering. Titik pusat cluster memvisualisasikan lokasi dan distribusi dari pusat cluster dalam konteks data yang telah diproses.

4. *Decision Tree*

Decision Tree, atau yang dikenal juga sebagai Pohon Keputusan, adalah salah satu algoritma pembelajaran mesin yang populer dan mudah dipahami. Algoritma ini digunakan untuk mengambil keputusan berdasarkan serangkaian aturan dan atribut yang ada dalam data. Decision Tree menggambarkan alur keputusan dalam bentuk pohon, di mana setiap node dalam pohon mewakili suatu atribut, cabang-cabang mewakili kemungkinan nilai atribut, dan daun-daun mewakili hasil keputusan atau prediksi. Pada saat membangun pohon keputusan, algoritma akan mencari atribut yang paling informatif atau signifikan untuk membagi data menjadi kelompok yang homogen. Untuk mengukur tingkat kehomogenan, digunakan metrik seperti indeks Gini atau entropi. Atribut yang memiliki nilai pemisah yang lebih tinggi akan menjadi pilihan yang lebih baik untuk membagi data. Proses ini dilakukan secara rekursif hingga mencapai kondisi berhenti, seperti mencapai tingkat kehomogenan tertentu atau ketika tidak ada atribut lagi yang tersedia untuk dibagi. Berikut hasil dari analisis *Decision tree*:

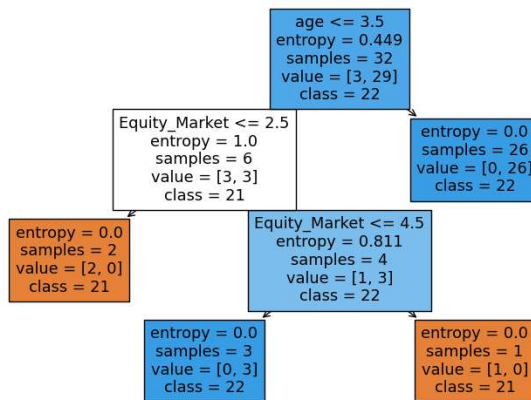
Decision Tree, atau yang dikenal juga sebagai Pohon Keputusan, adalah salah satu algoritma pembelajaran mesin yang populer dan mudah dipahami. Algoritma ini digunakan untuk mengambil keputusan berdasarkan serangkaian aturan dan atribut yang ada dalam data. Decision Tree menggambarkan alur keputusan dalam bentuk pohon, di mana setiap node dalam pohon mewakili suatu atribut, cabang-cabang mewakili kemungkinan nilai atribut, dan daun-daun mewakili hasil keputusan atau prediksi. Pada saat membangun pohon keputusan, algoritma akan mencari atribut yang paling informatif atau signifikan untuk membagi data menjadi kelompok yang homogen. Untuk mengukur tingkat kehomogenan, digunakan metrik seperti indeks Gini atau entropi. Atribut yang memiliki nilai pemisah yang lebih tinggi akan menjadi pilihan yang lebih baik untuk membagi data. Proses ini dilakukan secara rekursif hingga mencapai kondisi berhenti, seperti mencapai tingkat kehomogenan tertentu atau ketika tidak ada atribut lagi yang tersedia untuk dibagi.

```
# Menghitung akurasi model
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", accuracy)
```

Akurasi: 0.875

Pada metode Decision Tree yang sudah dilakukan menghasilkan nilai akurasi dari atribut *gender* dan *age* sebesar 0.875 atau 87.5%, merupakan nilai yang cukup tinggi dan sama dengan akurasi yang didapat pada analisis KNN. Akurasi adalah ukuran seberapa baik model atau sistem dapat memprediksi dengan benar pada data yang diberikan. Tingkat akurasi yang lebih tinggi menunjukkan bahwa model memiliki kinerja yang lebih baik dalam membuat prediksi yang benar.

```
# Plot pohon keputusan
plt.figure(figsize=(10, 6))
plot_tree(c45, feature_names=X.columns, class_names=class_names.astype(str), filled=True)
plt.show()
```



Gambar 9. Visualisasi pohon keputusan

Visualisasi diatas menunjukkan bentuk dari pohon keputusan yang terdiri dari simpul (*node*) dan cabang (*branch*). Setiap simpul pada pohon keputusan mewakili suatu aturan atau pemisahan pada data. Cabang-cabang yang keluar dari simpul menggambarkan kemungkinan hasil atau perpecahan berdasarkan fitur-fitur yang ada dalam data.

Pada setiap simpul, beberapa informasi yang dapat dianalisis antara lain :

- Fitur yang digunakan: Pada setiap simpul, akan ditunjukkan fitur yang digunakan untuk memisahkan data pada tingkat tersebut. Misalnya, “*gender*” atau “*age*”.

- b. Kriteria pemisahan: Nilai atau rentang nilai yang digunakan untuk memisahkan data pada tingkat tersebut. Misalnya, “*age* <= 30” atau “*gender* = *Female*”.
- c. Jumlah sampel: Jumlah sampel yang masuk ke simpul tersebut.
- d. Jumlah sampel per kelas: Persebaran sampel dalam kelas target pada simpul tersebut.
- e. Nilai kelas mayoritas: Kelas target mayoritas pada simpul tersebut.
- f. Kelas prediksi: Kelas target yang diprediksi oleh simpul tersebut jika sampel masuk ke simpul tersebut.

Visualisasi pohon keputusan, menunjukkan bahwa, model mengambil keputusan berdasarkan fitur-fitur yang ada dalam data, bagaimana perpecahan dilakukan pada setiap simpul dan bagaimana prediksi kelas target dibuat pada simpul-simpul daun (leaf node).

Dengan demikian tingkat akurasi yang tinggi lebih dari 50% yaitu sebesar 87,5% penulis menarik kesimpulan bahwa usia (*age*) dapat mempengaruhi kebijakan investasi dan profil risiko investor dalam pasar saham.

KESIMPULAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan Adapun beberapa kesimpulan yang didapatkan oleh penulis diantaranya:

- 1) Google Colab merupakan *tools* yang sangat efektif dan berguna dalam membuat visualisasi data, juga dalam melakukan beberapa analisis algoritma dengan menggunakan data finansial investasi dan pasar saham.
- 2) Visualisasi pada Dataset dengan menggunakan Google Colab berperan penting dalam menggambarkan data secara komprehensif, memberikan gambaran data yang mudah untuk dibaca dan dipahami.
- 3) Melalui visualisasi data dengan menggunakan Google Colab, informasi yang diperoleh dari dataset sangat mudah dibaca dan mudah untuk dipahami, contohnya seperti informasi mengenai jumlah variable dan korelasi antar variabel.
- 4) Google Colab sangat membantu dalam proses menganalisis data, visualisasi data dan melakukan pengolahan data.

Dengan menggunakan Google Colab tentu akan sangat membantu para peneliti maupun profesional khususnya dalam bidang data science untuk melakukan visualisasi data, pengolahan data serta pengambilan keputusan.

DAFTAR REFERENSI

- Hartama,Dedy (2018). “Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data” dalam *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK) Vol.3*. Pematangsiantar.
<https://ejournal.bsi.ac.id/ejournal/index.php/khatulistiwa/article/download/9593/4876>
- Hizriansyah, dkk.(2023). “Perancangan Model Dashboard Untuk Pelaporan dan Visualisasi Data Kesehatan Sebagai Sistem Monitoring di Dinas Kesehatan Gunungkidul” dalam *Journal of Information Systems for Public Health, Vol. 8, No. 1* (hlm1-9). Yogyakarta.
<https://jurnal.ugm.ac.id/jisph/article/download/72260/36320>
- Holden, Stein T & Mesfin Tilahun (2022). “Are risk preferences explaining gender differences in investment behavior?” dalam *Journal of Behavioral and Experimental Economics 101* (hlm 1-18). Norway.
<https://www.sciencedirect.com/science/article/pii/S2214804322001203>
- Irmayani,Windi (2021). “Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Naïve Bayes” dalam *Jurnal Khatulistiwa Informatika vol.9* (hlm 68 - 72). Pontianak.
<https://ejournal.bsi.ac.id/ejournal/index.php/khatulistiwa/article/download/9593/4876>
- Nainggolan,Dicky R. M. (2017). “Data Science, Big Data, And Predictive Analytics:A Platform For Cyberspace Security Intelligence Sains Data, Big Data, Dan Analisis Prediktif: Sebuah Landasan Untuk Kecerdasan Keamanan Siber” dalam *Jurnal Pertahanan & Bela Negara vol.7* (hlm. 37-54). Defense Management Study Center, Indonesia Defense University.
<https://jurnal.idu.ac.id/index.php/JPBH/article/download/192/94>
- Syaripul,Nadiar Ahmad dan Adam Mukharil Bachtiar (2016). “Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi Dki Jakarta: Topik Ekonomi Dan Keuangan Daerah” dalam *Jurnal Sistem Informasi (Journal of Information Systems) 2/12* (hlm 82-89). Bandung.
<https://jurnal.idu.ac.id/index.php/JPBH/article/download/192/9>