

Data Mining  
Assurance of Learning Report

**Implementation of Sentiment Analysis to Predict Stock Market Condition**

2401959023	Andreas Alexander	2022
2440062073	Susan Lady	2022
2440082024	Andrew Nathaniel	2022
2440037903	Tang Nyquel Gimli	2022
Table of Content		

Introduction	3
Literature Review	5
A. Stock Market	5
B. Analisis Cluster	5
C. Sentiment Analysis	6
D. Library	6
E. Explanatory Data Analytics	6
Data	7
Dataset	8
Methodology	11
Result and Analysis	13
Discussion	18
Implication	20

## Chapter I

### Introduction

Saham Merupakan Salah Satu instrumen keuangan jangka panjang yang diperdagangkan di pasar modal. Saham dapat didefinisikan sebagai tanda penyertaan atau kepemilikan seseorang atau badan usaha dalam suatu perusahaan atau perseroan terbatas. Selembar kertas yang berisi mengenai bukti kepemilikan atas perusahaan yang menerbitkan surat berharga tersebut merupakan wujud dari saham. Posisi permintaan dan penawaran atas saham yang ada di pasar modal, membuat saham memiliki harga untuk diperjualbelikan. Semakin tinggi tingkat permintaan dan penawaran terhadap lembar saham, maka harga saham pun akan tinggi dan juga sebaliknya.

Dalam berinvestasi pasti memiliki masing-masing risiko. Risiko yang dihadapi dalam saham dapat berupa risiko sistematis dan risiko yang tidak sistematis. Risiko sistematis tidak mungkin bisa dihindari karena resiko ini merupakan bentuk dari perubahan perekonomian yang tidak bisa dikendalikan oleh perusahaan. Sedangkan resiko yang tidak sistematis bisa diminumkan dengan melakukan diversifikasi. Berinvestasi dalam bentuk saham memiliki tingkat ketidakpastian yang tinggi sehingga resikonya juga tinggi. Maka perlunya Analisis yang dilakukan oleh investor-investor. Dalam melakukan analisis fundamental, analisis teknikal dan analisis sentimen. Dalam melakukan analisis, investor juga perlu melihat kondisi perusahaan itu sendiri, maupun kondisi perekonomian dimana perusahaan tersebut beroperasi sehingga dibutuhkan informasi informasi yang jelas.

Dalam analisis pergerakan harga saham terdapat analisis dimana mencakup aspek ekonomi, aspek industri, aspek perusahaan, dan banyak hal. Analisis terhadap aspek perusahaan penting untuk dilakukan oleh investor, karena analisis perusahaan ini menyangkut penilaian keadaan keuangan perusahaan, dimana dapat dilihat pendapatan atau laba yang diperoleh perusahaan. Apabila pendapatan atau laba yang diperoleh perusahaan tinggi, hal ini akan dapat membuat harga saham juga tinggi. Tingginya pendapatan atau laba yang diperoleh perusahaan membuat kepercayaan investor terhadap perusahaan akan pengembalian yang diharapkan investor.

Kepercayaan investor inilah yang akan memberikan keputusan investasi untuk membeli saham perusahaan tersebut. Tingkat pembelian saham perusahaan merupakan permintaan yang nantinya akan mempengaruhi harga saham perusahaan tersebut. Untuk menanamkan modal di suatu perusahaan para investor membutuhkan informasi yang berkaitan dengan perusahaan dimana informasi pada penelitian tersebut diambil dari sosial media yaitu twitter yang merupakan sumber data yang cukup besar dengan ribuan tweet yang dianalisis untuk setiap perusahaan oleh peneliti bahwa perusahaan-perusahaan yang memiliki potensi untuk kenaikan saham karena memiliki nilai sentiment negative yang lebih tinggi dibandingkan sentimen positif

atau potensi untuk mencapai kestabilan nilai saham karena memiliki nilai sentimen positif yang relatif sama dibandingkan nilai sentimen yang negatif.

Analisis sentimen ini merupakan proses penggunaan text analytics untuk mendapatkan berbagai sumber data dari internet dan beragam platform media sosial dengan memperoleh opini dari pengguna yang terdapat pada platform tersebut untuk mengidentifikasi opini, sikap, dan kondisi emosional seseorang. Pandangan masing-masing orang berdasarkan jenis analisis ini dapat dibagi menjadi positif dan negatif. Analisis jenis ini biasanya dilakukan dalam sebuah kalimat, kemudian diambil sentimen dari kalimat tersebut. Kata sifat memainkan peran penting dalam menentukan sentimen dari sebuah *Sentiment Analysis*. Analisis sentimen ini merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks ini yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur. Penelitian ini bertujuan untuk menunjukkan kepada masyarakat bahwa sentimen masyarakat mengenai kondisi saham suatu perusahaan dapat mempengaruhi potensi perusahaan untuk mengalami kenaikan atau penurunan nilai saham. Selain itu, penelitian ini juga bertujuan untuk mencari perusahaan yang memiliki potensi mengalami kenaikan nilai saham sebagai rekomendasi pembelian saham bagi investor.

## Chapter II

### Literature Review

#### A. Stock Market

Pasar saham (Stock Market) mengacu pada pengumpulan pasar dan bursa dimana penerbitan dan perdagangan ekuitas (saham perusahaan publik), obligasi dan jenis sekuritas lainnya terjadi, baik melalui bursa formal maupun pasar bebas. dikenal sebagai pasar ekuitas, pasar saham adalah salah satu komponen paling vital dari ekonomi pasar bebas, karena perusahaan ini menyediakan akses ke modal dengan imbalan memberi investor sepotong kepemilikan

Mengidentifikasi hubungan antara analisis sentimen berbasis Twitter dari perusahaan/indeks tertentu dan kinerja pasar jangka pendeknya menggunakan pengumpulan data tweet berskala besar. Hasil kami menunjukkan membawa dimensi negatif dan positif dari suasana hati publik hubungan sebab-akibat yang kuat dengan pergerakan harga saham/indeks individu. Kami juga telah menyelidiki berbagai fitur lain seperti fitur sentimen minggu sebelumnya mengontrol pembukaan minggu depan, nilai penutupan saham indeks untuk berbagai perusahaan teknologi dan indeks utama seperti DJIA dan NASDAQ-100. menunjukkan dibandingkan dengan pendekatan sebelumnya di daerah yang telah terbatas suasana hati publik yang sehat dan diskusi terbatas ticker saham, kami memverifikasi kinerja yang kuat dari model alternatif kami yang menangkap sentimen publik massa terhadap tertentu indeks atau perusahaan dengan cara yang dapat diskalakan dan karenanya memberdayakan investor tunggal untuk mengidealkan perbandingan relatif yang koheren.

#### B. Analisis Cluster

Analisis cluster adalah salah satu teknik analisis statistik yang digunakan untuk meringkas data dengan cara mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu yang dimiliki masing-masing objek. Kesamaan karakteristik tersebut dinyatakan dalam ukuran jarak antar objek. Pembentukan kelompok-kelompok berdasarkan jarak, obyek yang mirip seharusnya berada dalam kelompok yang sama dan mempunyai jarak yang lebih kecil. Sebaliknya objek yang berbeda berada dalam kelompok yang berbeda dan mempunyai jarak yang lebih besar.

Pelaksanaan Millenium Development Goals (MDGs) di Indonesia telah membawa banyak perubahan yang signifikan di Indonesia. Hal ini ditunjukkan oleh perkembangan beberapa indikator sosial ekonomi yang menjadi representasi dari MDGs. Selanjutnya, mengikuti perkembangan rekomendasi PBB, Indonesia meneruskan perencanaan pembangunan berkelanjutan sebagaimana yang direkomendasikan dalam program Sustainable Development

Goals (SDGs) yang dicanangkan tahun 2015. Pelaksanaan rekomendasi SDGs dilakukan serentak di seluruh Indonesia, namun perkembangan yang dicapai belum merata di seluruh wilayah. Masih terdapat daerah-daerah yang perkembangan pelaksanaannya masih tertinggal dibandingkan wilayah-wilayah Indonesia lainnya. Penelitian ini bertujuan untuk mengidentifikasi karakteristik provinsi berdasarkan kelompok perkembangan indikator SDGs serta mencari pemecahan masalah berdasarkan karakteristik kelompok tersebut

### C. Sentiment Analysis

Sentiment analysis merupakan salah satu bidang dari Natural Language Processing (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review. Dengan bantuan sentiment analysis, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.

Menyelidiki hubungan antara konten umpan Twitter dan pergerakan pasar saham. Secara khusus, kami ingin melihat apakah, dan seberapa baik, informasi sentimen yang diambil dari umpan ini dapat digunakan untuk memprediksi pergeseran harga di masa depan. Untuk menjawab pertanyaan ini, kami membuat model, memperkirakan akurasi, dan mengujinya pada data pasar nyata menggunakan portofolio tiruan. Hasil kami menunjukkan bahwa model tersebut berhasil menghasilkan laba tambahan.

### D. Library

Library merupakan sekumpulan kode yang memiliki fungsi-fungsi tertentu dan dapat dipanggil dalam program lain. Library dibuat untuk mempermudah dalam membangun sebuah aplikasi. Dengan library programmer tidak harus membangun kode dari awal untuk suatu fungsi tertentu.

### E. Explanatory Data Analytics

Exploratory Data Analysis (EDA) adalah sebuah proses kritis dalam melakukan investigasi awal pada data dengan tujuan menemukan pola, anomali, menguji hipotesis dan dapat memeriksa asumsi dengan bantuan statistik ringkasan kemudian representasi grafis (visualisasi). EDA bagi bisnis dan praktisi data menjadi sebuah kebutuhan utama agar membantu menemukan wawasan yang tidak terduga dalam data, mengecek kualitas data yang selanjutnya proses cleansing dan processing, hingga dapat menyampaikan wawasan data pada para stakeholder.

### Chapter III

#### Data

Dalam penelitian tersebut, data yang digunakan adalah data-data berupa kondisi keuangan enam buah perusahaan dari tahun 2019 hingga tahun tanggal 2 Desember 2022. Data keuangan perusahaan-perusahaan yang digunakan berasal dari perusahaan sebagai berikut:

- Tesla (TSLA) yang merupakan perusahaan otomotif yang berfokus pada penjualan mobil listrik.
- Apple (AAPL), yaitu perusahaan yang bergerak di bidang teknologi dan umumnya dikenal dengan produknya yang bernama *iphone*.
- Amazon (AMZN) yang merupakan perusahaan yang bergerak di berbagai bidang yang berbeda-beda, seperti e-commerce, *cloud computing*, dan streaming film digital.
- Nvidia (NVDA), yaitu perusahaan yang bergerak di bidang teknologi dengan produk-produk berupa *Graphics Processing Unit* (GPU).
- Microsoft (MSFT), yaitu perusahaan yang bergerak di berbagai sektor bidang teknologi, seperti sistem operasi, memberikan lisensi mengenai produk dan jasa, *cloud computing*, dan lainnya.
- Meta (META) atau yang dikenal juga dengan nama Facebook merupakan perusahaan yang bergerak di bidang media sosial dengan produk aplikasi-aplikasi media sosial seperti *Instagram*, *Facebook*, *Whatsapp*, serta project *Virtual Reality* yang dikenal dengan istilah *Metaverse*.

Data keuangan tersebut diambil secara manual melalui metode *data scraping* dengan menggunakan bahasa pemrograman python. Data tersebut diperoleh dengan menggunakan *yfinance library* melalui fungsi *Ticker* dan ditampilkan dalam bentuk grafik dengan menggunakan fungsi *graph\_objects* yang berasal dari *plotly library* dan didapat data sebanyak 4.542 buah data dengan masing-masing 757 buah data dalam satu buah perusahaan.

Selain data mengenai kondisi keuangan perusahaan, project tersebut juga menggunakan data berupa tweet mengenai keenam perusahaan tersebut dari tanggal 3 November 2019 hingga 2 Desember 2022 yang didapat menggunakan metode *data scraping* menggunakan *snsrape.module.twitter library*. Total data yang digunakan sebanyak 90.000 tweet yang terbagi menjadi 15.000 buah tweet di setiap perusahaan tersebut.

Selanjutnya, terdapat juga dataset berupa *review* terhadap *Amazon* yang digunakan untuk membangun model dengan tujuan analisis sentimen. Adapun data ini berbentuk data tabel *excel* (.csv) yang memiliki kolom 'Label' dengan tipe data *integer* dan kolom 'Review' dengan tipe data *string*. Kolom label berisikan nilai 1 dan 2, dimana label 1 menandakan *review* negatif dan 2 menandakan *review* positif. Pada analisis ini digunakan 50 ribu baris (*row*) *review* untuk dataset *train* dan *test*. Berikut adalah link dari dataset *review* yang kami gunakan:

[https://drive.google.com/drive/folders/1\\_NvM\\_8-N8Y\\_RMgbo\\_KnTjQU3T3wxWo1p?usp=sharing](https://drive.google.com/drive/folders/1_NvM_8-N8Y_RMgbo_KnTjQU3T3wxWo1p?usp=sharing)



## Chapter IV

### Dataset

Dari data - data yang telah didapatkan sebelumnya, dibentuk dataset untuk melakukan proses Explanatory Data Analytics (EDA). Dataset dibentuk dengan mengambil stock history dari setiap perusahaan yang telah disebutkan sebelumnya. Adapun stock history yang diambil merupakan stock tiap perusahaan dari 3 tahun terakhir.

Dataset yang didapatkan divisualisasikan menggunakan candlestick diagram. Jenis diagram ini biasa digunakan untuk menentukan pergerakan harga berdasarkan pola - pola yang ada. Pada diagram ini digunakan bentuk bar untuk merepresentasikan poin - poin dari harga di hari tersebut. Poin - poin dari harga tersebut adalah open, close, high, dan low. High dan low merupakan harga maksimal dan minimal pada waktu tertentu. Sedangkan open adalah harga memulai stock dan close adalah harga pada saat dagang (stock) selesai.

Berikut merupakan diagram stock history yang didapatkan dari masing - masing perusahaan:

#### 1. Tesla (TSLA)



#### 2. Apple (AAPL)



### 3. Amazon (AMZN)



### 4. Nvidia (NVDA)



### 5. Microsoft (MSFT)



## 6. Meta (META)



Dari diagram diatas, terdapat bar yang berwarna merah dan hijau. Bar yang berwarna merah menandakan bahwa harga closing lebih rendah dibandingkan dengan hari sebelumnya. Lalu, bar yang berwarna hijau menandakan bahwa harga closing lebih tinggi dibandingkan hari sebelumnya. Berdasarkan grafik tersebut, diketahui bahwa yang mengalami kenaikan beberapa bulan terakhir adalah perusahaan Nvidia, perusahaan yang mengalami penurunan adalah Meta, Amazon, dan Tesla, serta Microsoft dan Apple memiliki grafik yang cukup stabil di beberapa bulan terakhir ini. Mayoritas perusahaan mengalami kenaikan nilai saham sejak Januari 2020 dan terdapat juga beberapa perusahaan yang mengalami penurunan nilai saham, seperti Meta, Tesla, dan Amazon.

## Chapter V

### Methodology

*Sentiment analysis project* tersebut dibuat dengan menggunakan bahasa pemrograman *Python* dan *Google Colaboratory* sebagai tool untuk menjalankan bahasa pemrograman tersebut. Analisis tersebut dimulai dengan mengunduh (*installing*) beberapa *library* tambahan, seperti *yfinance* untuk mengambil data berupa kondisi saham secara *real time*, *snsrape* untuk mengambil dataset berupa tweet, serta *sklearn* untuk membuat model sentiment analysis tersebut.

Setelah mengimpor fungsi-fungsi yang akan digunakan, proses dilanjutkan dengan menentukan 6 buah perusahaan yang akan dianalisis, yaitu Tesla, Apple, Amazon, Nvidia, Microsoft, and Meta. Masing-masing perusahaan akan diambil kondisi saham dari tahun 2019 hingga tahun 2022 dengan fungsi *company\_ticker.history*. Setelah didapat data saham tersebut, data ditampilkan ke dalam grafik dengan menggunakan fungsi *figure* dari *graph\_object library*. Berdasarkan grafik tersebut, kami melakukan *explanatory data analytic* (EDA) untuk melihat baik buruknya kondisi saham setiap perusahaan selama 3 tahun terakhir.

Setelah dilakukan *explanatory data analytics*, proses dilanjutkan dengan memproses data-data yang didapat dari kaggle berupa review amazon sejumlah 50000 data testing dan 50000 data training. Data-data tersebut diproses dengan fungsi-fungsi dari library NLTK dengan menghilangkan beberapa tanda baca seperti titik dan menghilangkan imbuhan.

Sebelum membuat model, dilakukan *preprocessing* atau pembersihan data - data yang digunakan. Pembersihan data dilakukan dengan bantuan dari *library python*, yaitu NLTK (*Natural Language Toolkit*). Proses pembersihan data yang dilakukan adalah:

a) Menghilangkan *Stopwords*

*Stopwords* adalah kata - kata yang tidak memiliki arti sesungguhnya, dan hanya digunakan untuk pelengkap dalam sebuah kalimat. Contoh dari beberapa *stopwords* dalam bahasa Inggris adalah *the, a, an, is, are*, dan sebagainya. Untuk membuat training model lebih efisien, maka kata - kata yang tidak bermakna ini dapat dihilangkan, dan hanya memfokuskan training pada kata - kata yang memiliki makna.

b) Memecah kalimat menjadi kata - kata

Data - data review yang berbentuk kalimat akan dipecah menjadi kata per kata. Hal ini dilakukan agar setiap kata dalam kalimat tersebut dapat dilakukan analisis lebih lanjut lagi. Selain itu, memisahkan kata per kata akan memudahkan penetapan artinya.

### c) *Stemming*

*Stemming* merupakan proses untuk mengubah data kembali ke dalam bentuk basisnya. Contohnya pada kata *buying*, ketika dilakukan *stemming* akan menjadi *buy*. Hal ini dilakukan agar arti dari sebuah kata dapat menjadi lebih pasti, tidak ada keambiguan akibat adanya *prefix*, atau *suffix*.

Setelah itu, proses dilanjut dengan membuat model yang terbagi menjadi 3 buah tahap, diantaranya adalah:

#### A. Model Creation

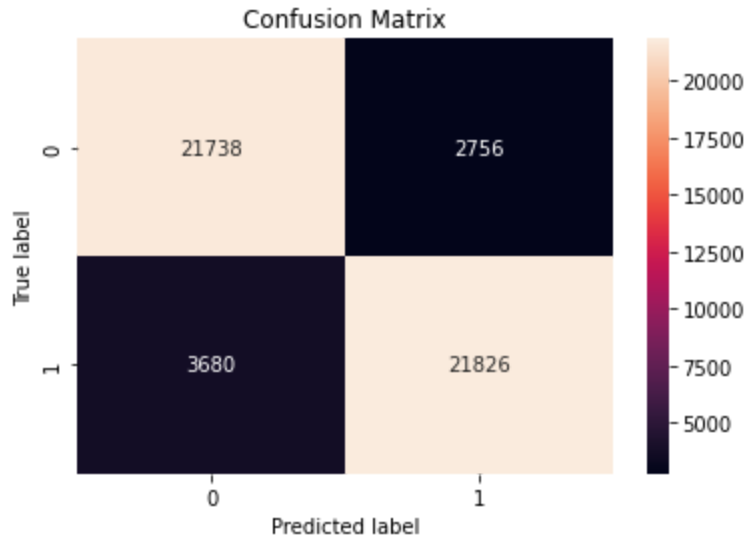
Pada proses model creation, kami membuat model dengan menggunakan pendekatan *supervised learning* (Naive Bayes), yang sesuai untuk tugas klasifikasi teks yang tujuannya adalah untuk memprediksi label (misalnya, sentimen positif atau negatif) berdasarkan teks. Alasan penggunaan metode Naive Bayes adalah karena Naive Bayes merupakan salah satu algoritma yang populer digunakan untuk tugas klasifikasi teks, relatif mudah diterapkan, dan seringkali dapat mencapai kinerja yang baik.

*Algoritma Multinomial Naive Bayes* tersebut diimplementasikan menggunakan fungsi dari library *scikit-learn* untuk mengklasifikasikan ulasan sebagai positif atau negatif berdasarkan teksnya. Hal ini dilakukan dengan terlebih dahulu mengubah teks menjadi fitur numerik menggunakan *CountVectorizer*, kemudian melatih model menggunakan data pelatihan, dan terakhir membuat prediksi pada data testing.

#### B. Model Evaluation

Setelah dilakukan training dan testing pada model, hasil prediksi dari model tersebut dievaluasi dalam bentuk confusion matrix dan menampilkan akurasi, presisi, f1-score, dan recall dari model tersebut.

Berdasarkan hasil yang telah diperoleh, kami mengevaluasi apakah nilai akurasi yang telah diperoleh sudah mencapai tingkat yang ditetapkan sebagai standar minimal, yaitu sebesar 85%. Hasilnya, model yang telah kami buat telah mencapai standar minimal tersebut dengan nilai akurasi sebesar 87,128% dengan hasil confusion matrix sebagai berikut.



### C. Model Deployment

Setelah dilakukan proses pembuatan model serta evaluasi, model tersebut akan digunakan untuk menganalisis dataset yang telah diambil dari twitter dengan metode *data scraping* yang dikonversikan menjadi dataframe dan diproses. Hasil dari analisis tersebut akan divisualisasikan dalam bentuk bar chart dan kami mencari perusahaan yang berpotensi mengalami kenaikan nilai saham sebagai rekomendasi perusahaan untuk dibeli nilai sahamnya. Selain itu, kami juga menentukan perusahaan yang lebih baik dihindari untuk dibeli nilai sahamnya.

Note: untuk script code dapat diakses pada link berikut.

[https://colab.research.google.com/drive/1Wcg\\_2SK83GBaxxygLWOEb4OwPIceMS41](https://colab.research.google.com/drive/1Wcg_2SK83GBaxxygLWOEb4OwPIceMS41)

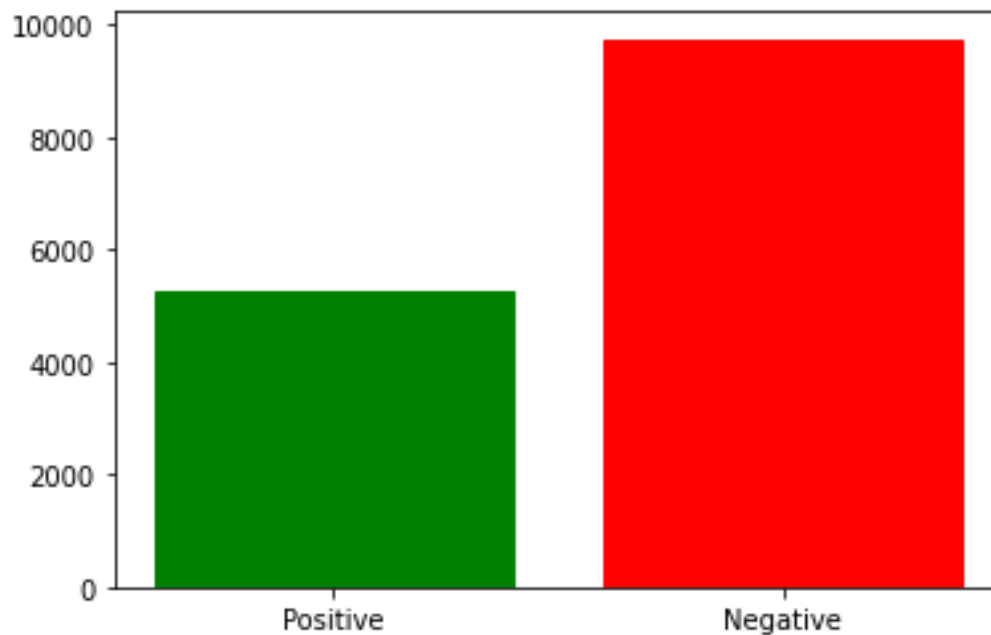
## Chapter VI

### Result and Analysis

Setelah dilakukan *sentiment analysis*, didapat hasil berupa dua buah kategori tweet, yakni positif dan negatif dengan masing-masing jumlah sebagai berikut.

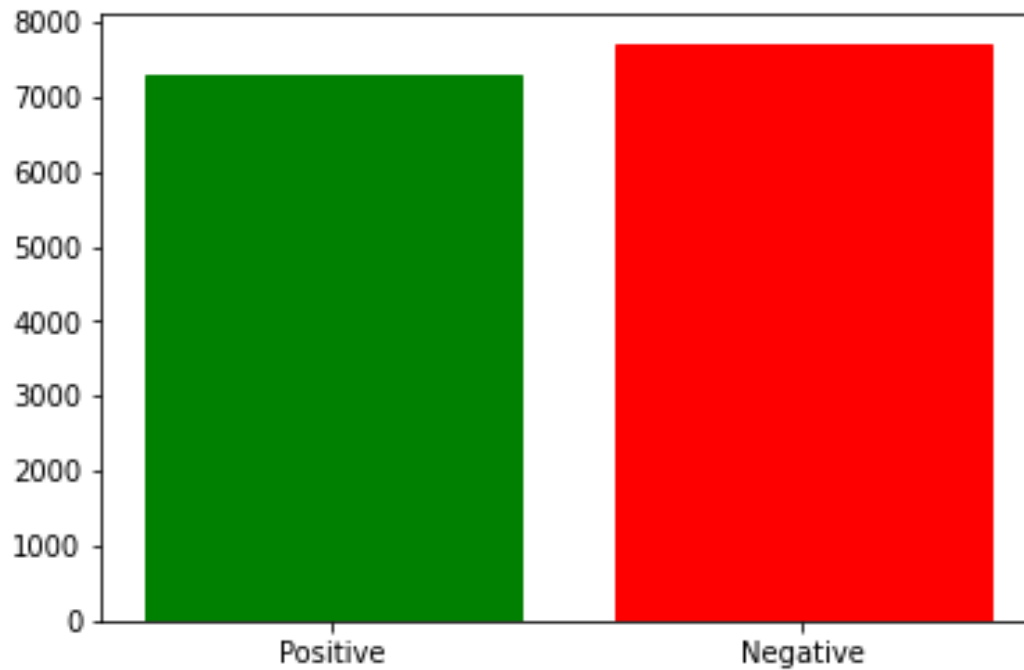
- Apple (AAPL)

Perusahaan Apple memiliki hasil analisis berupa 5248 buah tweet yang termasuk dalam golongan positif dan 9753 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.

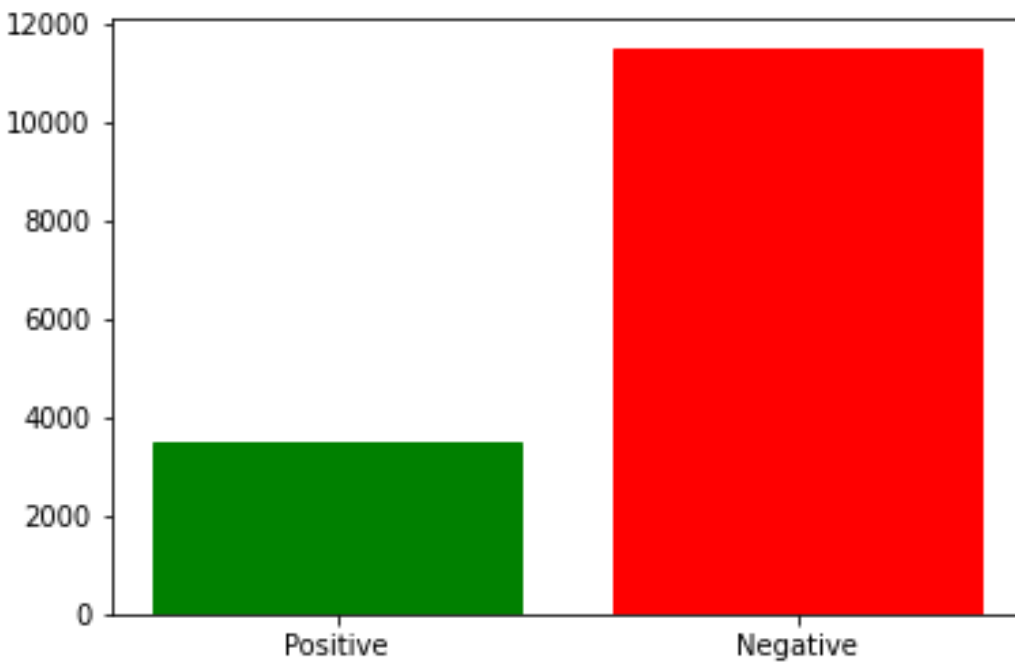


- Tesla (TSLA)

Perusahaan Tesla memiliki hasil analisis berupa 7288 buah tweet yang termasuk dalam golongan positif dan 7713 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.



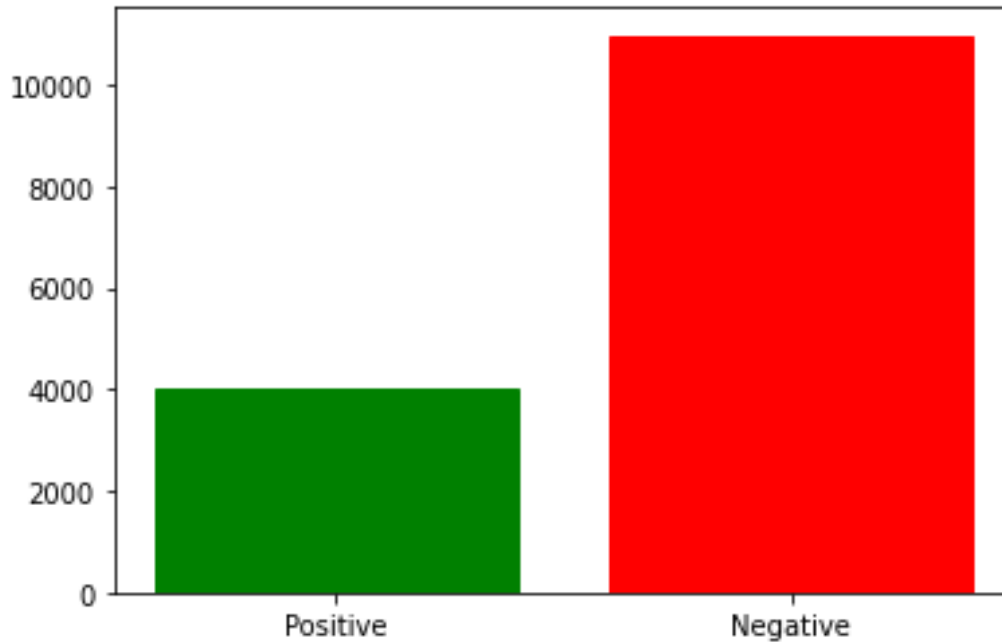
- Amazon (AMZN)  
Perusahaan Amazon memiliki hasil analisis berupa 3501 buah tweet yang termasuk dalam golongan positif dan 11500 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.



- Nvidia (NVDA)

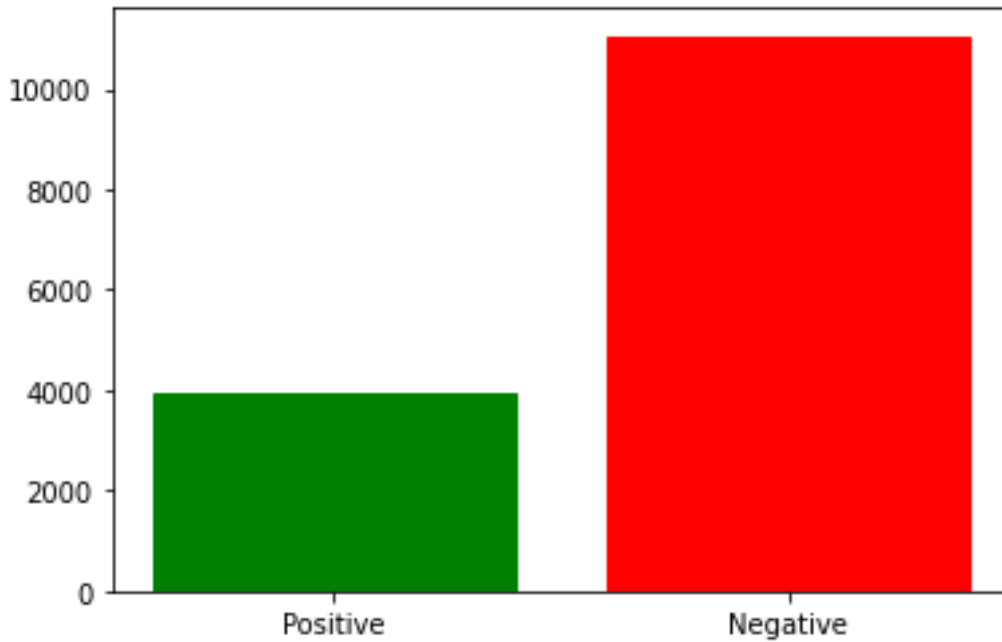


Perusahaan Nvidia memiliki hasil analisis berupa 4022 buah tweet yang termasuk dalam golongan positif dan 10979 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.



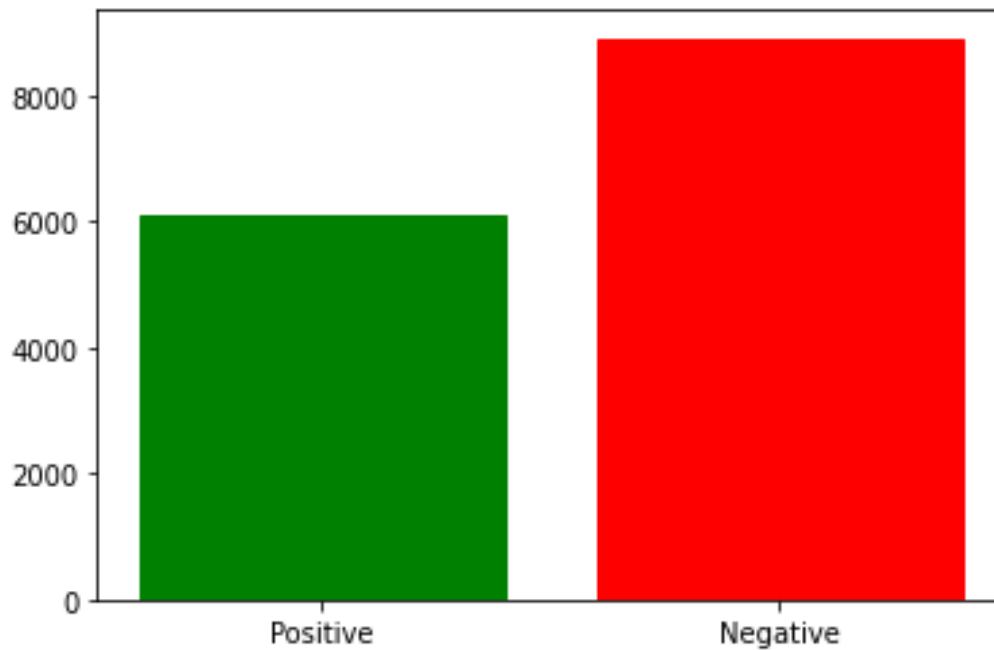
- Microsoft (MSFT)

Perusahaan Microsoft memiliki hasil analisis berupa 3943 buah tweet yang termasuk dalam golongan positif dan 11058 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.



- Meta (META)

Perusahaan Meta memiliki hasil analisis berupa 6098 buah tweet yang termasuk dalam golongan positif dan 8903 buah tweet yang termasuk dalam golongan negatif yang ditampilkan dalam *bar chart* sebagai berikut.



Berdasarkan hasil *sentiment analysis* tersebut, perusahaan Apple, Amazon, Nvidia, Meta, dan Microsoft memiliki jumlah tweet negatif lebih banyak dibandingkan dengan tweet positif. Hal ini menandakan bahwa perusahaan-perusahaan tersebut diprediksi tidak berpotensi mengalami kenaikan nilai saham. Di lain sisi, perusahaan Tesla memiliki jumlah tweet positif yang mendekati seimbang dibandingkan negatif dan diprediksi bahwa perusahaan tersebut berpotensi mengalami kestabilan nilai saham.

## Chapter VII

### Discussion

Dari hasil yang telah divisualisasikan diatas, dapat dikatakan bahwa perusahaan Tesla memiliki kemungkinan terkecil untuk mengalami penurunan saham. Hal ini dapat kita lihat dalam bar chart diatas yang menggambarkan perbandingan komentar negatif dan positif. Terlihat bahwa perbandingan komentar negatif dan positif paling hampir seimbang dimiliki oleh perusahaan Tesla.

Perusahaan APPL, AMZN, META, MSFT, NVDA memiliki jumlah komentar negatif yang lebih tinggi dibandingkan komentar positifnya. Dimana perbandingan jumlah komentar negatif dan positif terbilang cukup jauh. Sedangkan untuk perusahaan TSLA, walaupun jumlah komentar negatif masih lebih tinggi dibandingkan dengan komentar positifnya, rasio perbandingan tersebut tidak berbeda jauh dibandingkan perusahaan-perusahaan yang lain.

Perusahaan APPL memiliki potensi yang minim. Produk - produk APPL memang menarik, desainnya memiliki khasnya yang sulit dilawan oleh perusahaan lain yang sejenis, fitur - fiturnya sudah tidak perlu diragukan, terutama kualitas dari kameranya. Namun, harga yang ditawarkan cenderung tinggi, dan masing - masing dari produk APPL sangat bergantung satu sama lain (gadgetnya hanya bisa terkoneksi dengan gadget dengan merek yang sama, dan tidak fleksibel).

Perusahaan AMZN juga tidak jauh beda dengan APPL untuk perbandingan komentar positif dan negatifnya. AMZN juga memiliki masalah dimana penggunaannya tergolong kompleks, keuntungan tidak terjamin, kompetisi yang tinggi. Hal ini menyebabkan potensi nya dalam market tidak begitu besar.

Lalu, perusahaan lain yang memiliki nilai negatif yang jauh lebih tinggi adalah MSFT, META, dan NVDA. Masalah dari perusahaan MSFT adalah, walaupun beberapa dari produknya bisa dikatakan terjangkau, tetapi sistem operasi MSFT seringkali terdapat masalah dan tidak jarang user memiliki masalah pada sistem auto - update dari MSFT. Masalah dari perusahaan META adalah dari sisi periklanannya, kompetisi, belum bisa digunakan secara luas (terbatas dengan fasilitas). Masalah dari perusahaan NVDA adalah produk seri terbaru mereka dijual dengan harga yang tidak masuk akal. Dimana produk baru ini pun lebih berorientasi kepada tampilan luar produk, dibandingkan dengan kegunaan utamanya. Dari masalah-masalah tersebut ketiga perusahaan tersebut memiliki potensi yang rendah untuk mengalami peningkatan harga saham di masa mendatang.

## Chapter VIII

### Conclusion

*Sentiment Analysis* digunakan untuk mengidentifikasi opini, sikap, dan kondisi emosional seseorang. Pandangan masing-masing orang berdasarkan jenis analisis ini dapat dibagi menjadi positif dan negatif. Analisis jenis ini biasanya dilakukan dalam sebuah kalimat, kemudian diambil sentimen dari kalimat tersebut. Kata sifat memainkan peran penting dalam menentukan sentimen dari sebuah *Sentiment Analysis*.

Penelitian dilakukan dengan mengambil data saham dari 6 buah perusahaan, yaitu Twitter, Apple, Amazon, Nvidia, Microsoft, dan Meta. berdasarkan data tersebut, EDA dilakukan peneliti agar mendapat kondisi saham selama 3 Tahun terakhir. Kemudian, peneliti mengambil data tweet tentang kondisi keuangan perusahaan dari 3 November 2019 hingga 2 Desember 2022. Data tersebut melalui proses *data cleaning* menggunakan fungsi dari NLTK. Pada proses tersebut, beberapa tanda baca seperti titik dan imbuhan dihilangkan. Setelah diproses, data tersebut akan masuk ke tahap Sentiment Analysis pertama, yaitu Model Creation, dimana kita menggunakan library scikit-learn untuk mengklasifikasi data menjadi positif dan negatif. Setelah itu model melalui proses Model Evaluation, dimana akurasi, presisi, f1-score, dan recall ditampilkan dalam bentuk confusion matrix. Tahap terakhir merupakan Model Deployment, dimana model yang sudah dievaluasi digunakan untuk menganalisis dataset yang telah diambil dari twitter. Hasil tersebut kemudian divisualisasikan dalam bentuk bar chart.

Twitter merupakan sumber data yang cukup besar, dengan informasi diambil dari tweet yang dilakukan setiap orang. Ribuan tweet yang dianalisis untuk setiap perusahaan oleh peneliti menunjukkan bahwa Apple, Amazon, Nvidia, Meta, dan Microsoft tidak memiliki potensi untuk kenaikan saham karena memiliki nilai sentimen negatif yang lebih tinggi dibandingkan sentimen positif, sedangkan Tesla merupakan satu-satunya perusahaan yang memiliki kemungkinan paling kecil terjadinya penurunan nilai saham karena memiliki nilai sentimen positif yang hampir sama dibandingkan nilai sentimen yang negatif.

## Chapter IX

### Implication

Studi tersebut memiliki keterlibatan dalam menentukan perusahaan-perusahaan yang berpotensi mengalami kenaikan nilai saham (*stock market*) dan perusahaan-perusahaan yang tidak berpotensi mengalami kenaikan nilai saham. Penelitian ini dapat membantu para pemula untuk menentukan perusahaan untuk diinvestasikan dengan tujuan untuk mendapat keuntungan melalui investasi dengan instrumen saham. Selain itu, studi ini juga terlibat dalam menunjukkan kepada masyarakat mengenai pengaruh sentimen terhadap kondisi saham sebuah perusahaan.

## References

- [1] Ahmad, "Pengertian Resesi Ekonomi: Penyebab, Dampak Dan Solusi," *Gramedia Literasi*, 02-Jul-2022. [Online]. Available: <https://www.gramedia.com/literasi/resesi-ekonomi/>. [Accessed: 15-Dec-2022]
- [2] "Bab I Pendahuluan A. Latar Belakang - eprints.ums.ac.id." [Online]. Available: [http://eprints.ums.ac.id/29219/2/04.\\_BAB\\_I.pdf](http://eprints.ums.ac.id/29219/2/04._BAB_I.pdf). [Accessed: 15-Dec-2022]
- [3] B. Gapo and B. Gapo, "Nvidia GeForce RTX 4000 series release date, Price, Specs," *GPU Mag*, 14-Dec-2022. [Online]. Available: <https://www.gpumag.com/nvidia-geforce-rtx-4000-series/>. [Accessed: 18-Dec-2022].
- [4] C. Mitchell, "Understanding a candlestick chart," Investopedia, 08-Jul-2022. [Online]. Available: <https://www.investopedia.com/trading/candlestick-charting-what-is-it/>. [Accessed: 18-Dec-2022]
- [5] D. Burrows, "Tesla stock slumps on demand concerns," Kiplinger.com, 07-Dec-2022. [Online]. Available: <https://www.kiplinger.com/investing/stocks/tesla-stock-slumps-on-demand-concerns>. [Accessed: 18-Dec-2022]
- [6] Fool, T.M. (2018) The top three Amazon challenges, and how to solve them, Adlucent. Available at: <https://www.adlucent.com/resources/blog/the-top-three-amazon-challenges-and-how-to-solve-them/> (Accessed: December 18, 2022).
- [7] I. Adeosun, "What exactly is a library in programming?," *Treehouse*, 10-Aug-2019. [Online]. Available: <https://teamtreehouse.com/community/what-exactly-is-a-library-in-programming>. [Accessed: 15-Dec-2022]
- [8] Id, "ID," *Amazon*, 1980. [Online]. Available: <https://aws.amazon.com/id/what-is/sentiment-analysis/>. [Accessed: 15-Dec-2022]
- M. Lewis, "5 big problems @ Microsoft," Medium, 22-Jul-2017. [Online]. Available: <https://wtfmitchel.medium.com/5-big-problems-microsoft-f6b2733eb2a4>. [Accessed: 18-Dec-2022]
- [9] N. S. Dewi, "Non it belajar exploratory data analysis, SOLUSI data analysis Lebih Dalam," *www.digation.id*. [Online]. Available: <https://www.digation.id/read/017978/non-it-belajar-exploratory-data-analysis-solusi-data->

analysis-lebih-dalam#:~:text=Exploratory%20Data%20Analysis%20(EDA)%20adalah,ke mudian%20representasi%20grafis%20(visualisasi). [Accessed: 15-Dec-2022]

[10] O. S. T. ost@lincoln.ac.uk, “Analyzing stock market movements using Twitter sentiment analysis,” *Analyzing stock market movements using Twitter sentiment analysis - The*, 01-Jan-1970. [Online]. Available: <http://eprints.lincoln.ac.uk/id/eprint/11274/>. [Accessed: 15-Dec-2022]

[11] “Pemetaan Masalah Pembangunan Berkelanjutan Dan Pertumbuhan Ekonomi ...” [Online]. Available: [https://www.researchgate.net/profile/Pardomuan-Sihombing/publication/344202783\\_Pemetaan\\_Masalah\\_Pembangunan\\_Berkelanjutan\\_dan\\_Pertumbuhan\\_Ekonomi\\_Inklusif\\_di\\_Indonesia\\_Implementasi\\_Analisis\\_Kluster/links/5f5b53ff4585154dbbc8996d/Pemetaan-Masalah-Pembangunan-Berkelanjutan-dan-Pertumbuhan-Ekonomi-Inklusif-di-Indonesia-Implementasi-Analisis-Kluster.pdf](https://www.researchgate.net/profile/Pardomuan-Sihombing/publication/344202783_Pemetaan_Masalah_Pembangunan_Berkelanjutan_dan_Pertumbuhan_Ekonomi_Inklusif_di_Indonesia_Implementasi_Analisis_Kluster/links/5f5b53ff4585154dbbc8996d/Pemetaan-Masalah-Pembangunan-Berkelanjutan-dan-Pertumbuhan-Ekonomi-Inklusif-di-Indonesia-Implementasi-Analisis-Kluster.pdf). [Accessed: 15-Dec-2022]

[12] “Pemetaan Masalah Pembangunan Berkelanjutan Dan Pertumbuhan Ekonomi ...” [Online]. Available: [https://www.researchgate.net/profile/Pardomuan-Sihombing/publication/344202783\\_Pemetaan\\_Masalah\\_Pembangunan\\_Berkelanjutan\\_dan\\_Pertumbuhan\\_Ekonomi\\_Inklusif\\_di\\_Indonesia\\_Implementasi\\_Analisis\\_Kluster/links/5f5b53ff4585154dbbc8996d/Pemetaan-Masalah-Pembangunan-Berkelanjutan-dan-Pertumbuhan-Ekonomi-Inklusif-di-Indonesia-Implementasi-Analisis-Kluster.pdf](https://www.researchgate.net/profile/Pardomuan-Sihombing/publication/344202783_Pemetaan_Masalah_Pembangunan_Berkelanjutan_dan_Pertumbuhan_Ekonomi_Inklusif_di_Indonesia_Implementasi_Analisis_Kluster/links/5f5b53ff4585154dbbc8996d/Pemetaan-Masalah-Pembangunan-Berkelanjutan-dan-Pertumbuhan-Ekonomi-Inklusif-di-Indonesia-Implementasi-Analisis-Kluster.pdf). [Accessed: 15-Dec-2022]

[13] “Pengertian Library Dalam Dunia Pemrograman,” *CODING RAKITAN*. [Online]. Available: <https://codingrakitan.blogspot.com/2019/10/pengertian-library-dalam-dunia.html>. [Accessed: 15-Dec-2022]

[14] “Sentiment analysis of Twitter feeds for the ... - stanford university.” [Online]. Available: <https://cs229.stanford.edu/proj2011/ChenLazer-SentimentAnalysisOfTwitterFeedsForThePredictionOfStockMarketMovement.pdf>. [Accessed: 15-Dec-2022]

[15] S. Smarter, “SWOT analysis of Apple,” StudySmarter US, 2020. [Online]. Available: <https://www.studysmarter.us/explanations/business-studies/business-case-studies/swot-analysis-of-apple/>. [Accessed: 18-Dec-2022]

[16] “Stock market,” *Corporate Finance Institute*, 29-Nov-2022. [Online]. Available: <https://corporatefinanceinstitute.com/resources/wealth-management/stock-market/>. [Accessed: 15-Dec-2022]

- [17] V. Patel, M. Lancaster, A. Artashyan, A. A. Shanto, and Y. Hua, "Meta is experiencing serious problems due to the metaverse," Gizchina.com, 29-Oct-2022. [Online]. Available: <https://www.gizchina.com/2022/10/29/meta-is-experiencing-serious-problems-due-to-the-metaverse/>. [Accessed: 18-Dec-2022]
- [18] Y. E. A. Sunarso, "Analisis cluster Dan Aplikasinya," *CORE*. [Online]. Available: <https://core.ac.uk/display/158342821>. [Accessed: 15-Dec-2022].