

Analisis dan Visualisasi Data Saham Menggunakan PySpark

Disusun untuk memenuhi Tugas Kelompok

Mata Kuliah Big Data



Oleh:

Abid Husein	312110031
Afif Firmansyah	312110232
Ahmad Syukron	312110056
Iman Setiawan	312110219

KELAS TI.21.A.1

PROGRAM STUDI TEKNIK INFORMATIKA

UNIVERSITAS PELITA BANGSA

KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa, karena atas limpahan rahmatnya penyusun dapat menyelesaikan makalah ini tepat waktu tanpa ada halangan yang berarti dan sesuai dengan harapan.

Kami menyadari bahwa dalam penyusunan makalah ini masih banyak kekurangan karena keterbatasan kami. Maka dari itu penyusun sangat mengharapkan kritik dan saran untuk menyempurnakan makalah ini. Semoga apa yang ditulis dapat bermanfaat bagi semua pihak yang membutuhkan.

DAFTAR ISI

COVER	1
KATA PENGANTAR	2
DAFTAR ISI	3
BAB I : PENDAHULUAN	
1.1 LATAR BELAKANG	5
1.2 RUMUSAN MASALAH	6
1.3 TUJUAN	6
1.4 BATASAN MASALAH	7
1.5 RELEVANSI	7
BAB II : TINJAUAN PUSTAKA	
2.1 PENGANTAR BIG DATA	8
2.2 APACHE SPARK	8
2.3 VISUALISASI DATA	8
2.4 PENGOLAHAN DAN NORMALISASI DATA	9
BAB III : METODE PENELITIAN	
3.1 DESAIN PENELITIAN	10
3.2 SUMBER DATA	10
3.3 PROSES PEMROSESAN DATA	10
3.4 ANALISIS RETURN SAHAM	11
3.5 ANALISIS KORELASI	11
3.6 VISUALISASI DATA	11
3.7 EVALUASI KINERJA SAHAM	11
BAB IV : HASIL DAN ANALISA	
4.1 DESKRIPSI DATA	12
4.2 PEMROSESAN DATA	12

4.3 NORMALISASI DATA SAHAM	13
4.4 ANALISIS RETURN SAHAM	14
4.5 ANALISIS KORELASI	15
4.6 VISUALISASI DATA	15
4.7 EVALUASI KINERJA SAHAM	16
4.8 TEMUAN UTAMA	16
BAB V : KESIMPULAN	
5.1 RINGKASAN HASIL PENELITIAN	17
5.2 IMPLIKASI PRAKTIS	18
5.3 BATASAN PENELITIAN	18
DAFTAR PUSTAKA	19

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saham merupakan tanda bukti penyertaan modal dalam suatu perusahaan atau perseroan. Jadi ketika kita membeli saham suatu perusahaan sama saja kita memiliki perusahaan tersebut dan klaim atas pendapatan serta asset perusahaan tersebut. Perdagangan saham terjadi di Bursa Efek Indonesia yaitu tempat berlangsungnya perdagangan efek di pasar sekunder. Untuk dapat melakukan transaksi, investor harus menjadi nasabah di perusahaan efek tertentu. Setelah menjadi nasabah di perusahaan efek, investor bisa memilih saham perusahaan apa yang akan dibeli oleh investor. Ada 2 macam analisa yang biasa dipakai dalam perdagangan saham yaitu analisa fundamental dan analisa teknikal. Analisa fundamental adalah analisa mengenai kondisi perusahaan yang mencakup pendapatan, pengeluaran, produk, serta strategi pemasaran perusahaan. Analisa teknikal adalah salah satu analisis atau metode pendekatan yang mengevaluasi pergerakan suatu harga saham, valas, kontrak berjangka, indeks dan instrumen keuangan lainnya. Analisa fundamental digunakan oleh investor untuk memilih perusahaan mana yang sahamnya akan dibeli. Analisa teknikal digunakan untuk meramalkan harga di masa yang akan datang berdasarkan pola pergerakan harga di masa lampau. Analisa teknikal dipakai oleh investor untuk menentukan kapan harus membeli dan kapan harus menjual sahamnya.

Saat ini teknologi komputasi big data dilakukan secara terdistribusi dengan bantuan engine komputasi terdistribusi yang menggabungkan beberapa komputer untuk melakukan tugas pengolahan data. Salah satu contoh engine komputasi terdistribusi adalah Apache Spark. Apache Spark memiliki kemampuan yang tidak dimiliki oleh engine-engine lainnya yaitu untuk mengolah data secara realtime. Apache Spark juga memiliki kecepatan yang tinggi dibandingkan dengan engine

lainnya serta mendukung komputasi untuk kecerdasan buatan atau Artificial Intelligence.

Analisis pasar saham memiliki peranan yang sangat penting dalam dunia keuangan. Para investor sering kali bergantung pada berbagai metode dan teknik untuk memahami pergerakan harga saham dan mengidentifikasi peluang investasi. Dalam konteks ini, penggunaan teknologi seperti Apache Spark dan Python dapat mempermudah analisis besar dataset, dan visualisasi data dengan bantuan pustaka seperti Plotly, Seaborn, dan Matplotlib.

1.2 Rumusan Masalah

Dalam menghadapi volatilitas pasar saham, investor sering dihadapkan pada tantangan untuk memahami pergerakan harga saham secara efektif. Analisis historis dan perhitungan tingkat pengembalian dapat memberikan wawasan yang penting, namun pengelolaan data dan pemrosesan informasi yang besar sering menjadi hambatan. Oleh karena itu, penelitian ini bertujuan untuk memanfaatkan kekuatan teknologi Big Data dan analisis data menggunakan PySpark untuk menyediakan alat analisis saham yang efisien.

1.3 Tujuan

Tujuan dari penelitian ini adalah mengembangkan alat analisis saham menggunakan Apache Spark dan Python. Melalui integrasi teknologi ini, penelitian bertujuan untuk memberikan solusi untuk permasalahan pengelolaan data besar dalam analisis saham, serta menyediakan visualisasi yang interaktif dan informatif untuk membantu investor dalam pengambilan keputusan.

1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan, antara lain:

- Fokus pada analisis harga saham menggunakan data historis.
- Analisis hanya mencakup sejumlah saham tertentu.
- Tidak mempertimbangkan faktor eksternal yang dapat memengaruhi pasar saham.

1.5 Relevansi

Penelitian ini relevan dalam konteks pengembangan alat analisis saham yang dapat memberikan informasi yang akurat dan mudah diinterpretasi bagi investor. Penggunaan Apache Spark memungkinkan pemrosesan data yang efisien, sedangkan pustaka visualisasi seperti Plotly, Seaborn, dan Matplotlib memberikan alat yang kuat untuk menyajikan informasi secara menarik dan informatif. Hasil penelitian diharapkan dapat memberikan kontribusi pada pemahaman pasar saham dan membantu investor dalam mengambil keputusan investasi yang lebih baik.

Dengan demikian, penelitian ini akan membahas implementasi kode PySpark untuk analisis saham dan visualisasi data, serta memberikan wawasan tentang bagaimana teknologi ini dapat menjadi solusi untuk tantangan dalam analisis pasar saham.

BAB II

TINJAUAN PUSTAKA

2.1 Pengantar Big Data

Big Data merujuk pada pengelolaan dan analisis dataset yang sangat besar dan kompleks yang tidak dapat dengan mudah diatasi oleh sistem konvensional. Beberapa karakteristik utama Big Data melibatkan Volume (jumlah data), Velocity (kecepatan data diterima dan diproses), Variety (beragam jenis data), Variability (fluktuasi data), dan Value (nilai informasi yang dapat diperoleh).

2.2 Apache Spark

Apache Spark adalah sebuah framework open-source untuk pemrosesan data yang dirancang untuk kecepatan dan skalabilitas. PySpark, sebagai antarmuka Python untuk Spark, memungkinkan pengguna untuk menggunakan kekuatan Spark menggunakan bahasa pemrograman Python. Dengan kemampuan pemrosesan data secara paralel dan distribusi, Spark menjadi pilihan yang populer dalam aplikasi Big Data. Pustaka PySpark yang berbasis Python memungkinkan integrasi dengan ekosistem Python untuk analisis data dan pemodelan.

2.3 Visualisasi Data menggunakan Matplotlib, Seaborn, dan Plotly

Matplotlib, Library visualisasi data yang menyediakan berbagai jenis plot seperti line plot, bar plot, dan histogram. Digunakan dalam kode untuk membuat plot matriks korelasi dan histogram.



Plotly, Library visualisasi data interaktif yang memungkinkan pembuatan plot interaktif dengan mudah. Digunakan dalam kode untuk membuat plot interaktif garis.



2.4 Pengolahan dan Normalisasi Data dengan PySpark

Pengolahan data melibatkan langkah-langkah seperti pengisian nilai yang hilang dan normalisasi data. Dalam kode ini, pengisian nilai yang hilang dilakukan dengan mengisi nilai rata-rata dan menggunakan fungsi interpolasi. Normalisasi dilakukan untuk membandingkan kinerja saham dengan menghitung nilai relatif terhadap harga awal.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan analisis data menggunakan Apache Spark dan PySpark untuk pemrosesan data besar dalam lingkungan distribusi. Visualisasi data dilakukan dengan menggunakan Matplotlib, Seaborn, dan Plotly. Metode analisis melibatkan pemrosesan, pengolahan, dan interpretasi data saham dari berbagai perusahaan dalam sebuah portofolio.

3.2 Sumber Data

Data saham yang digunakan berasal dari beberapa dataset CSV yang menyimpan informasi harga penutupan harian dari saham-saham tertentu. Dataset utama berisi indeks IHSG dan beberapa saham perusahaan, yang digunakan untuk analisis portofolio.

3.3 Proses Pemrosesan Data

Pemrosesan data dilakukan dengan menggunakan Apache Spark dan PySpark. Langkah-langkah pemrosesan melibatkan pengurutan data berdasarkan tanggal, penanganan nilai yang hilang dengan menggunakan metode pengisian nilai rata-rata dan interpolasi, serta normalisasi data untuk membandingkan kinerja saham-saham dalam portofolio.

3.4 Analisis Return Saham

Return Harian, Perhitungan return harian dilakukan untuk setiap saham dalam portofolio menggunakan metode persentase perubahan.

Return Keseluruhan, Menghitung return keseluruhan dari portofolio saham selama periode waktu tertentu.

3.5 Analisis Korelasi

Matriks korelasi dibuat untuk mengevaluasi hubungan antar saham-saham dalam portofolio. Korelasi memberikan wawasan tentang sejauh mana pergerakan harga saham-saham tersebut berhubungan satu sama lain.

3.6 Visualisasi Data

Line Plot Interaktif, Menggunakan Plotly Express untuk membuat line plot interaktif yang menunjukkan pergerakan harga saham-saham dalam portofolio.

Heatmap Korelasi, Menggunakan Seaborn untuk membuat heatmap korelasi untuk memvisualisasikan hubungan antar saham.

Histogram dan Distribusi Plot, Matplotlib dan Plotly Figure Factory digunakan untuk membuat histogram dan distribusi plot return harian saham-saham.

3.7 Evaluasi Kinerja Saham

Kinerja saham dievaluasi melalui analisis return harian, return keseluruhan, dan korelasi antar saham. Analisis ini memberikan pemahaman tentang sejauh mana saham-saham dalam portofolio berkontribusi terhadap kinerja keseluruhan.

BAB IV

HASIL DAN ANALISA

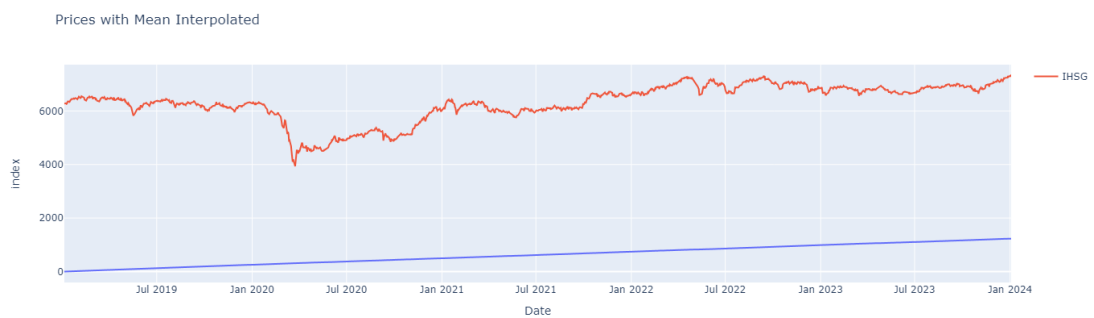
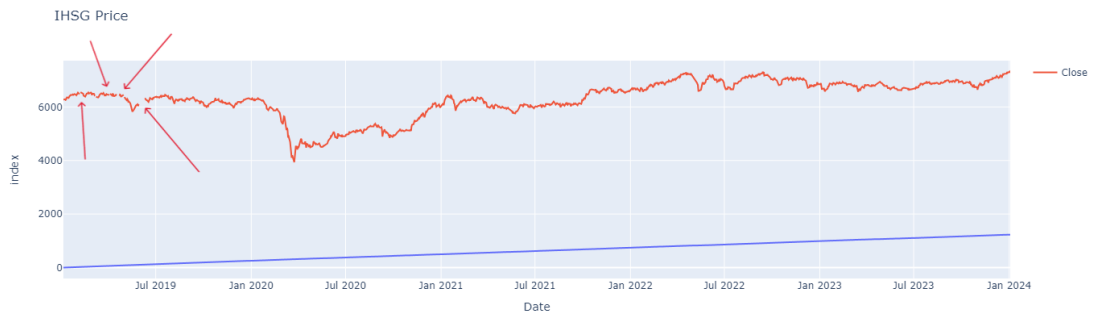
4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini terdiri dari harga penutupan harian dari beberapa saham yang membentuk portofolio, seperti IHSG, BBCA, BBNI, ASII, UNVR, ICBP, INKP, dan TLKM. Data ini mencakup periode tertentu dan diurutkan berdasarkan tanggal untuk memastikan urutan waktu yang benar.



4.2 Pemrosesan Data

Proses pemrosesan data melibatkan langkah-langkah pengurutan data, penanganan nilai yang hilang, dan normalisasi data. Data diurutkan berdasarkan tanggal untuk memastikan kronologi yang benar. Nilai yang hilang diisi dengan menggunakan metode pengisian nilai rata-rata dan interpolasi. Normalisasi data dilakukan untuk membandingkan kinerja saham-saham dalam portofolio dengan menghitung nilai relatif terhadap harga awal.



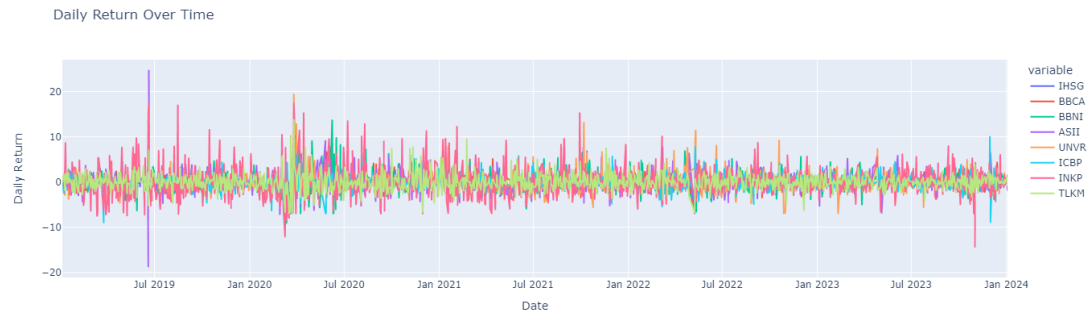
4.3 Normalisasi Data Saham

Normalisasi data saham membantu para investor dan analis untuk mendapatkan pemahaman yang lebih baik tentang perbandingan kinerja saham-saham dalam suatu periode. Dengan menyamakan titik awal, analisis return saham menjadi lebih objektif dan memfasilitasi evaluasi yang lebih akurat terhadap perubahan nilai saham dari waktu ke waktu.



4.4 Analisis Return Saham

Return Harian, Perhitungan return harian dilakukan untuk setiap saham dalam portofolio. Hasilnya menunjukkan bagaimana perubahan harga harian dari masing-masing saham.

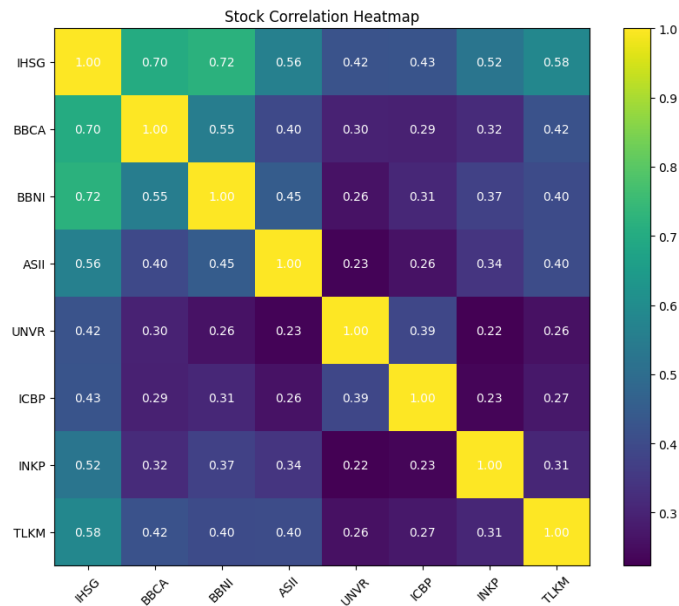


Return Keseluruhan, Menghitung return keseluruhan dari portofolio saham selama periode waktu tertentu. Hal ini memberikan gambaran tentang kinerja portofolio secara keseluruhan.

- IHSG 16.009950924786832%
- BBKA 79.63496637848222%
- BBNI 22.636103151862464%
- ASII -32.93413173652695%
- UNVR -63.07531380753137%
- ICBP 0.2358490566037652%
- INKP -24.498886414253896%
- TLKM 6.738544474393526%

4.5 Analisis Korelasi

Matriks korelasi dibuat untuk mengevaluasi hubungan antar saham-saham dalam portofolio. Nilai korelasi dapat berkisar antara -1 hingga 1, dengan nilai positif menunjukkan korelasi positif dan nilai negatif menunjukkan korelasi negatif.



Menganalisis korelasi saham dengan IHSX dapat membantu investor memahami sejauh mana pergerakan IHSX memengaruhi pergerakan harga saham individu. Korelasi positif menunjukkan bahwa saham cenderung bergerak seiring dengan IHSX, sementara korelasi negatif menunjukkan hubungan sebaliknya. Hal ini dapat menjadi panduan untuk diversifikasi portofolio dan manajemen risiko.

4.6 Visualisasi Data

Line Plot Interaktif: Menggunakan Plotly Express, line plot interaktif dibuat untuk memvisualisasikan pergerakan harga saham-saham dalam portofolio.

Heatmap Korelasi: Dengan menggunakan Seaborn, heatmap korelasi dibuat untuk memberikan gambaran visual tentang hubungan antar saham.

Histogram dan Distribusi Plot, Matplotlib dan Plotly Figure Factory digunakan untuk membuat histogram dan distribusi plot return harian saham-saham.

4.7 Evaluasi Kinerja Saham

Evaluasi kinerja saham dilakukan melalui analisis return harian, return keseluruhan, dan korelasi antar saham. Analisis ini memberikan wawasan tentang sejauh mana kontribusi masing-masing saham terhadap kinerja keseluruhan portofolio.

4.8 Temuan Utama

- Hasil return harian memberikan pemahaman tentang volatilitas saham-saham dalam portofolio.
- Return keseluruhan memberikan gambaran tentang kinerja keseluruhan portofolio selama periode tertentu.
- Matriks korelasi membantu mengidentifikasi hubungan antar saham dengan IHSG sebagai acuan untuk mempermudah mengelola saham, memberikan wawasan penting untuk diversifikasi portofolio.
- Visualisasi data memberikan pemahaman yang lebih baik tentang tren dan distribusi kinerja saham-saham.

BAB V

KESIMPULAN

5.1 Ringkasan Hasil Penelitian

Penelitian ini bertujuan untuk menganalisis kinerja saham-saham dalam portofolio menggunakan Apache Spark dan PySpark untuk pemrosesan data, serta Matplotlib, Seaborn, dan Plotly untuk visualisasi data. Berikut adalah ringkasan hasil penelitian:

- **Return Harian**

Analisis return harian menggunakan Plotly Express menunjukkan pergerakan harga saham-saham dalam portofolio selama periode tertentu. Volatilitas dan tren harian masing-masing saham dapat diidentifikasi dengan jelas.

- **Return Keseluruhan**

Perhitungan return keseluruhan memberikan gambaran tentang performa portofolio saham dalam jangka waktu tertentu. Hasil ini memberikan wawasan holistik tentang profitabilitas investasi dalam saham-saham tertentu.

- **Matriks Korelasi**

Matriks korelasi dengan heatmap Seaborn memberikan visualisasi yang jelas tentang hubungan antar saham dalam portofolio. Analisis ini dapat membantu dalam strategi diversifikasi untuk mengurangi risiko.

- **Visualisasi Data**

Visualisasi data, termasuk histogram return harian dan distribusi plot, memberikan wawasan tambahan tentang distribusi frekuensi return harian saham-saham dalam portofolio. Hal ini dapat membantu dalam identifikasi pola dan karakteristik pergerakan harga.

- **Evaluasi Kinerja Saham**

Evaluasi kinerja saham melalui analisis return harian, return keseluruhan, dan korelasi membantu dalam mengevaluasi kontribusi masing-masing saham terhadap performa portofolio secara keseluruhan. Ini memberikan pemahaman lebih dalam tentang kinerja investasi.

5.2 Implikasi Praktis

Hasil penelitian ini memiliki implikasi praktis bagi para investor dan pengambil keputusan di pasar keuangan. Analisis return harian dan keseluruhan dapat digunakan untuk memahami kinerja historis saham dan membuat keputusan investasi yang lebih terinformasi. Korelasi antar saham dapat membantu dalam membentuk strategi diversifikasi yang efektif.

5.3 Batasan Penelitian

Batasan penelitian ini juga mencakup fokus pada dataset yang terbatas pada faktor-faktor eksternal yang dapat mempengaruhi harga saham. Adanya variabel tambahan dapat meningkatkan kompleksitas analisis dan memberikan wawasan lebih lanjut terhadap dinamika pasar saham. Namun, untuk menjaga fokus penelitian, penelitian ini akan membatasi diri pada faktor-faktor yang terdapat dalam dataset dan tidak akan mencakup faktor eksternal seperti peristiwa geopolitik, perubahan kebijakan pemerintah, atau peristiwa lain yang dapat berdampak signifikan pada harga saham. Penggunaan dataset yang lebih luas dan inklusif dapat menjadi area penelitian masa depan yang potensial untuk mendapatkan pemahaman yang lebih komprehensif tentang perilaku pasar saham.

DAFTAR PUSTAKA

<https://finance.yahoo.com>

<https://matplotlib.org>

<https://plotly.com>

<https://seaborn.pydata.org>