

BAB I PENDAHULUAN

1.1 Latar Belakang

Pasar saham telah menjadi salah satu instrumen investasi paling populer di Indonesia, terutama di era globalisasi dan digitalisasi. Instrumen ini tidak hanya menarik perhatian masyarakat, tetapi juga menjadi cerminan perkembangan ekonomi nasional. Saham banyak dipilih oleh investor sebagai alat investasi karena dapat memberikan tingkat keuntungan yang menarik. Namun, banyak investor masih ragu untuk berinvestasi karena fluktuasi indeks harga saham yang sangat cepat, sehingga seringkali tidak sesuai dengan harapan mereka. Indeks saham adalah ukuran statistik yang mencerminkan keseluruhan pergerakan harga atas sekumpulan saham yang dipilih berdasarkan kriteria dan metodologi tertentu, serta dievaluasi secara berkala. Indeks saham berfungsi sebagai alat untuk mengukur sentimen pasar atau kepercayaan investor. Perubahan nilai yang tercermin dalam suatu indeks dapat menjadi indikator yang merefleksikan opini kolektif dari pelaku pasar. Oleh karena itu, prediksi atau peramalan harga saham menjadi alat bantu penting dalam perencanaan yang lebih efektif dan efisien, sehingga investor dapat dengan mudah mengukur sentimen pasar (Sofi et al., 2021).

Dalam perekonomian Indonesia, sektor perbankan memiliki peran strategis sebagai motor penggerak fungsi intermediasi keuangan sekaligus pendukung program pemerintah, seperti inklusi keuangan dan stabilitas ekonomi. Sektor ini mendominasi kapitalisasi pasar saham Indonesia, khususnya melalui keberadaan perusahaan-perusahaan Badan Usaha Milik Negara (BUMN). Saham BUMN di sektor perbankan kerap digunakan sebagai indikator kesehatan ekonomi nasional karena mencerminkan dinamika pasar keuangan secara signifikan. Perubahan harga saham BUMN di sektor perbankan tidak hanya memengaruhi sentimen pasar domestik tetapi juga mencerminkan ekspektasi terhadap kinerja ekonomi secara keseluruhan (Rohyati et al., 2024).

Prediksi harga saham telah menjadi perhatian utama dalam dunia investasi karena memberikan panduan yang berguna bagi investor dalam mengambil keputusan. Berbagai metode telah dikembangkan untuk melakukan prediksi, di antaranya pendekatan tradisional dan pendekatan *machine learning*. Metode tradisional mengacu pada pendekatan statistik klasik yang menggunakan hubungan matematis sederhana antara variabel untuk membuat prediksi. Contohnya adalah regresi linear, yang memodelkan hubungan linier antara satu atau lebih variabel independen dengan variabel dependen. Meskipun metode ini sering digunakan, dalam interpretasi, Kong et al., (2025) menemukan bahwa memiliki nilai akurasi yang lebih rendah jika dibandingkan dengan *machine learning* dalam peramalan harga saham. *Machine Learning*, di sisi lain, adalah pendekatan inovatif yang mencakup berbagai algoritma dan alat yang digunakan untuk serangkaian tugas pemrosesan data. Pendekatan



ini memiliki kemampuan untuk menangkap pola data yang lebih kompleks dan non-linear, menjadikannya alat yang efektif dalam prediksi harga saham. Dalam beberapa tahun terakhir, *machine learning* telah berkembang pesat dan memasuki berbagai disiplin ilmu, termasuk analisis pasar saham (G.Carleo et al., 2019).

Salah satu pendekatan yang cukup sering diterapkan yakni regresi. *Regression* adalah teknik yang digunakan untuk dua teori, yaitu analisis *regression* yang biasanya digunakan untuk *forecasting* dan *prediction*, dan analisis yang digunakan untuk menentukan hubungan sebab-akibat antara variabel independen dan variabel dependen. Model *regression* sendiri hanya menunjukkan hubungan antara variabel dependen dan kumpulan dataset tetap dari variabel yang berbeda (Sofi et al., 2021). Berbagai penelitian telah menggunakan algoritma regresi linear untuk prediksi. Thabibi & Supriyanto (2023) menggunakan model ini untuk memprediksi harga saham Telkom, Indosat, dan XL, dimana diperoleh bahwa model *multiple linear regression* memiliki rata-rata *error* yang lebih kecil dibandingkan model *decision tree regression*. Penelitian lain oleh Marwa dan rekan-rekannya membandingkan metode recurrent neural network (RRN) dengan *linear regression* untuk prediksi harga saham Netflix. Hasilnya, metode *linear regression* mencapai lebih baik dibandingkan metode RNN yang menunjukkan bahwa regresi linear cukup efektif dalam memodelkan hubungan linear antara variabel-variabel tersebut. Akan tetapi, metode ini mengalami keterbatasan dalam menangani pola hubungan yang lebih kompleks, seperti hubungan non-linier dan interaksi antar variabel yang lebih rumit.

Seiring dengan perkembangan teknologi, metode *machine learning*, seperti *Extreme Gradient Boosting* (XGBoost), mulai digunakan untuk memprediksi harga saham. XGBoost adalah algoritma berbasis ensemble yang dirancang untuk menangkap pola-pola kompleks dalam data dengan memanfaatkan kekuatan *decision tree* secara iteratif. Beno Jange (2021) membandingkan efektivitas XGBoost dengan metode Prophet dalam memprediksi harga saham Bank Central Asia (BCA) dan menemukan bahwa XGBoost mampu menghasilkan tingkat akurasi yang lebih. Ahmad Syahreza, dan rekan-rekannya (2024) juga membandingkan kinerja model prediksi cuaca terhadap model *random forest*, *support vector regression* dan XGBoost. Dan hasilnya, XGBoost memiliki performa yang sangat baik. Hal ini menunjukkan potensi XGBoost sebagai metode yang kompetitif untuk memodelkan data dengan pola yang kompleks (Septiana Rizky et al., 2022). Meskipun demikian, XGBoost juga memiliki kekurangan, seperti kebutuhan komputasi yang lebih besar dan risiko overfitting jika data tidak ditangani dengan benar.

Pasar saham Indonesia sebagai pasar berkembang menghadirkan tantangan tersendiri, seperti volatilitas tinggi yang dipengaruhi oleh kondisi makroekonomi, dan dinamika pasar global (Putri et al., 2025). Dalam sektor moneter, seperti perubahan suku bunga acuan, memiliki dampak terhadap kinerja saham BUMN.

Sebagai latar belakang dan penelitian sebelumnya, peneliti tertarik untuk melakukan penelitian yang berjudul **“Perbandingan Kinerja Metode Regresi Gradient Boosting (XGBoost) dalam Prediksi Harga Saham**



BUMN Sektor Perbankan". Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan, baik secara akademis maupun praktis. Secara akademis, penelitian ini akan memperkaya literatur mengenai perbandingan metode *machine learning* dan statistik dalam konteks pasar saham Indonesia yang memiliki volatilitas tinggi. Secara praktis, penelitian ini diharapkan dapat memberikan panduan yang lebih baik bagi investor, regulator, dan pelaku pasar dalam memilih metode prediksi yang lebih akurat dan efisien untuk menghadapi tantangan pasar yang dinamis. Penelitian ini juga diharapkan dapat mengisi gap yang ada dalam literatur, mengingat masih terbatasnya studi yang membahas perbandingan langsung antara regresi linear dan XGBoost dalam konteks prediksi harga saham BUMN sektor perbankan Indonesia.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana perbandingan performa metode Regresi Linear dan *Extreme Gradient Boosting* (XGBoost) dalam memprediksi harga saham Badan Usaha Milik Negara (BUMN) di sektor perbankan.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan performa metode Regresi Linear dan *Extreme Gradient Boosting* (XGBoost) dalam memprediksi harga saham Badan Usaha Milik Negara (BUMN) di sektor perbankan. Perbandingan dilakukan berdasarkan nilai *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) guna menilai tingkat kesalahan prediksi dari masing-masing metode.

1.4 Landasan Teori

1.4.1 Saham

Saham adalah instrumen keuangan yang menunjukkan kepemilikan atas suatu perusahaan. Dalam konteks pasar modal, saham memberikan hak kepada pemegangnya untuk mendapatkan dividen dan berpartisipasi dalam keputusan perusahaan. Menurut (Intan Sari, 2019), saham mencerminkan nilai perusahaan yang dapat dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi makro, kebijakan moneter, kinerja internal perusahaan, dan sentimen pasar. Faktor-faktor ini menjadikan harga saham sangat dinamis, berfluktuasi berdasarkan perkembangan ekonomi lokal maupun global.

1.4.2 Pasar Saham



1 tempat di mana barang dagangan dalam bentuk saham
 1am adalah instrumen pasar yang populer di masyarakat.
 1ai dilakukan melalui perusahaan investasi yang menyediakan
 di pasar saham juga merupakan sumber pembiayaan bagi
 menjual sebagian kepemilikan perusahaan kepada publik.

Dengan pemahaman ini, pasar saham memberikan peluang untuk mengalihkan kepemilikan dari pemegang uang (investor) kepada perusahaan atau individu lain yang membutuhkan modal untuk berinvestasi (Andini et al., 2024).

1.4.3 Perusahaan Badan Usaha Milik Negara (BUMN)

Perusahaan Badan Usaha Milik Negara (BUMN) adalah perusahaan yang kepemilikannya sebagian atau seluruhnya dimiliki oleh pemerintah berdasarkan UU No. 19 Tahun 2003 tentang BUMN. BUMN sektor perbankan memiliki peran strategis sebagai lembaga intermediasi keuangan sekaligus sumber pendapatan negara melalui dividen yang dibagikan secara rutin. Di pasar modal, saham bank BUMN menarik minat investor karena dianggap memiliki tingkat keamanan dan kepastian dividen yang tinggi. Saat ini, terdapat empat bank BUMN yang terdaftar di Bursa Efek Indonesia dengan kapitalisasi pasar yang signifikan. Peningkatan jumlah investor pada saham bank BUMN berdampak pada aktivitas pasar modal dan mendukung stabilitas industri perbankan, menjadikannya sektor yang vital dalam perekonomian nasional (Wiarta et al., 2020).

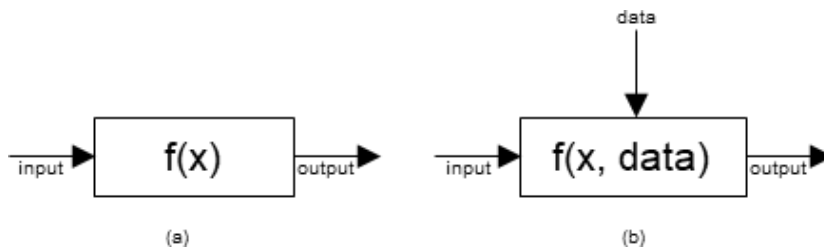
1.4.4 Prediksi Harga Saham

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Zulfani et al., 2024). Prediksi harga indeks saham merupakan salah satu tugas yang paling menantang bagi para profesional di industri keuangan dan skala terkait lainnya karena fitur volatilitas (Kafil, 2019).

1.4.5 Machine Learning

Machine Learning merupakan aplikasi dari disiplin ilmu kecerdasan buatan (*Artificial Intelligence*) yang menggunakan teknik statistika sehingga dapat menghasilkan suatu model yang otomatis dari sekumpulan data, agar dilakukan analisa oleh komputer untuk “belajar” (Iqbal Baihaqi et al., 2023).





Sumber: (Iqbal Baihaqi et al., 2023)

<http://jurnal.jcosbida.com/index.php/jubitek/article/view/244>

Gambar 1. Proses *Machine Learning*

1.4.6 Uji Asumsi Klasik

Sebelum menggunakan model regresi linear untuk analisis dan prediksi, penting untuk memastikan bahwa model memenuhi asumsi-asumsi klasik yang diperlukan agar estimasi yang dihasilkan valid dan dapat diandalkan (Mar'atush Sholihah et al., 2023).

1.4.6.1 Uji Linearitas

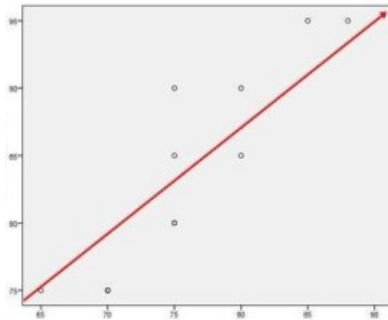
Uji linearitas digunakan untuk mengidentifikasi apakah hubungan antara dua variabel bersifat linier. Model regresi memenuhi asumsi linearitas jika terdapat hubungan garis lurus antara variabel independen dan dependen. Salah satu metode untuk mendeteksi linearitas adalah melalui grafik *scatter plot*. Hubungan linier antara variabel independen dan dependen dapat bersifat positif (searah) atau negatif (tidak searah). Hubungan positif berarti peningkatan pada variabel independen akan diikuti oleh peningkatan variabel dependen. Sebaliknya, hubungan negatif berarti peningkatan pada variabel independen menyebabkan penurunan pada variabel dependen, atau sebaliknya (Afni et al., 2022). Adapun kriteria pengujian dalam uji linearitas sebagai berikut:

H_0 = Tidak terdapat hubungan linier antara X dan Y .

H_1 = Terdapat hubungan linier antara X dan Y .

Gambar 2 menunjukkan contoh *scatter plot* dengan hubungan linear positif antara variabel independen dan dependen.





Sumber: [spssindonesia.com](https://www.spssindonesia.com/2019/05/uji-linearitas-grafik-scatter-plot-spss.html), 2014 <https://www.spssindonesia.com/2019/05/uji-linearitas-grafik-scatter-plot-spss.html>

Gambar 2. Grafik Scatter Plot Positif

Berdasarkan scatter plot di atas, titik-titik data membentuk pola garis lurus dari kiri bawah ke kanan atas, menunjukkan hubungan linier positif antara variabel independen (X) dan dependen (Y). Artinya, jika variabel independen meningkat, variabel dependen juga akan meningkat.

1.4.6.2 Uji Normalitas

Model regresi yang baik adalah model yang memiliki residual yang berdistribusi normal. Uji normalitas dilakukan menggunakan Uji *Jarque-Bera* (JB test) yang didasarkan pada nilai residual. Penilaian distribusi residual dilakukan berdasarkan nilai probabilitas (*p-value*) pada tingkat signifikansi 5% ($\alpha = 0,05$). Apabila *p-value* > 0,05, maka residual memenuhi asumsi normalitas. Sebaliknya, jika *p-value* ≤ 0,05, maka asumsi normalitas tidak terpenuhi (Winalza & Alfarisi, 2021). Uji normalitas *Jarque-Bera* (JB test) dilakukan dengan kriteria pengujian sebagai berikut:

H_0 = Residual berdistribusi normal.

H_1 = Residual tidak berdistribusi normal.

1.4.6.3 Uji Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji apakah model regresi ditemukan adanya korelasi antara variabel bebas. Model regresi yang baik seharusnya tidak terjadi korelasi di antara variabel independen. Untuk mendeteksi ada atau tidaknya multikolinearitas didalam model regresi dapat dilihat dari nilai toleran dan VIF (*Variance Inflation Factor*). Kedua ukuran ini menunjukkan setiap variabel independen manakah yang dijelaskan oleh variabel independen lainnya. Jika nilai $VIF \leq 10$ yang menunjukkan bahwa tidak terjadi gejala multikolinearitas. Oleh karena itu hasil uji multikolinearitas menunjukkan bahwa tidak terjadi gejala multikolinearitas (Atha & Nur Afdella, 2022). Hipotesis yang diuji:

H_0 = Tidak ada korelasi antar variabel independen (tidak terjadi multikolinearitas)

H_1 = Terdapat korelasi antar variabel independen (terjadi multikolinearitas)



H_1 = Terdapat korelasi antar variabel independen (terjadi multikolinearitas).

1.4.6.4 Uji Heteroskedastisitas

Uji Heteroskedastisitas terjadi ketika varians dari residual tidak konstan pada setiap tingkat observasi. Dalam penelitian regresi, yang diharapkan adalah kondisi homoskedastisitas, yaitu ketika residual memiliki varians yang konstan. Untuk mengetahui ada atau tidaknya gejala heteroskedastisitas dalam model regresi, dilakukan uji heteroskedastisitas dengan metode *Breush-Pagan-Godfrey*. Model dinyatakan memenuhi asumsi homoskedastisitas apabila nilai probabilitas (*p-value*) masing-masing variabel lebih besar dari tingkat signifikansi 10% ($\alpha = 0,10$) (Macpal, 2017). Adapun kriteria pengujian asumsi heteroskedastisitas sebagai berikut:

H_0 = Tidak terjadi heteroskedastisitas (variens residual konstan).

H_1 = Terjadi heteroskedastisitas (variens residual tidak konstan).

1.4.6.5 Uji Autokorelasi

Uji autokorelasi bertujuan menguji apakah dalam model regresi linear ada korelasi antara nilai residual pada periode t dengan nilai residual pada periode $t - 1$ (sebelumnya) (Macpal, 2017). Untuk mendeteksi autokorelasi, digunakan uji *Durbin-Watson*. Nilai statistik *Durbin-Watson* berada dalam rentang 0 hingga 4, di mana nilai mendekati 2 mengindikasikan tidak adanya autokorelasi. Nilai yang mendekati 0 menunjukkan autokorelasi positif, sedangkan mendekati 4 menunjukkan autokorelasi negatif. Adapun kriteria pengujian uji autokorelasi *Durbin-Watson* sebagai berikut:

H_0 = Tidak terdapat autokorelasi pada residual.

H_1 = Terdapat autokorelasi pada residual.

1.4.7 Regresi Linear

Regresi linear adalah metode statistika yang digunakan untuk memodelkan hubungan antara satu variabel dependen (terikat) dan satu variabel independen (prediktor). Tujuan utama dari regresi linear adalah untuk menemukan hubungan linier antara kedua variabel tersebut sehingga model dapat digunakan untuk memprediksi nilai variabel dependen berdasarkan variabel independen (Santoso et al., 2023). Persamaan dasar regresi linear sederhana adalah:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Dengan:



dependen (terikat)
 ependen (prediktor)
 ilai Y saat $X = 0$)
 gresi (kemiringan)
 esalahan model (residual)

1.4.7.1 Regresi Linear Berganda

Regresi linear berganda adalah pengembangan dari regresi linear yang digunakan untuk memodelkan hubungan antara satu variabel dependen dan lebih dari satu variabel independen (Nuriyah et al., 2023). Tujuan utama regresi linear berganda adalah untuk memprediksi nilai variabel dependen berdasarkan kombinasi beberapa variabel independen serta untuk memahami bagaimana masing-masing variabel independen mempengaruhi variabel dependen.

Persamaan umum untuk regresi linear berganda adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2)$$

Dengan:

Y	= variabel dependen (terikat)
$X_1, X_2, \dots, X_k$	= variabel independen (prediktor)
β_0	= <i>intercept</i> (nilai Y saat $X = 0$)
$\beta_1, \beta_2, \dots, \beta_k$	= koefisien regresi untuk masing-masing X_i
ε	= galat atau kesalahan model (residual)

1.4.8 Ensemble Learning

Ensemble Learning adalah sekumpulan pengklasifikasi yang dilatih secara individual dimana hasil prediksi dari setiap pengklasifikasi akan digabungkan dengan beberapa cara seperti rata-rata, pemungutan suara terbanyak dan lain-lain untuk menghasilkan prediksi akhir. *Ensamble Learning* digunakan untuk meningkatkan kinerja model atau menangani kinerja model yang buruk. Teknik *ensemble* yang umum digunakan yaitu *bagging* dan *boosting* (Yulianti et al., 2022).

1.4.8.1 Boosting

Boosting yang di presentasikan oleh Robert Schapire pada tahun 1998, adalah salah satu strategi pengumpulan yang dapat bekerja pada beberapa hasil pengelompokan yang lemah sehingga mereka dapat berubah menjadi hasil yang kuat. Teknik klasifikasi adalah strategi perhitungan pembelajaran yang dibuat berdasarkan beberapa karakterisasi atau model perkiraan yang kemudian digunakan untuk mengelompokkan informasi baru dengan mempertimbangkan beban yang diantisipasi yang dibuat sebelumnya. Konsep *ensemble* dengan *boosting* bekerja melalui persiapan kumpulan model secara berurutan dan kemudian menggabungkan seluruh model untuk membuat harapan, model berikutnya mendapatkan keuntungan sebelumnya (Zhang & Gong, 2020).



dient Boosting (XGBoost)

dient Boosting (XGBoost) adalah pengembangan dari algoritma yang diperkenalkan oleh Dr. Tianqi Chen pada tahun 2014. yang untuk menyelesaikan berbagai masalah seperti regresi,

klasifikasi, dan perankingan. Prinsip dasar XGBoost adalah pembelajaran iteratif dengan menurunkan *loss function* untuk meningkatkan akurasi prediksi, sekaligus menerapkan regularisasi untuk mengontrol kompleksitas model dan menghindari *overfitting* (Rachmi, 2020).

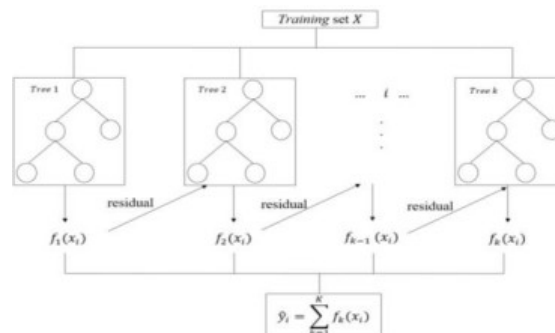
Salah satu keunggulan XGBoost adalah kemampuannya membangun pohon regresi yang lebih teratur, sehingga meningkatkan kinerja model. Hasil akhir prediksi diperoleh dari penjumlahan kontribusi seluruh pohon regresi yang dibangun. Rumus prediksi dituliskan sebagai:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (3)$$

Dengan:

- $\hat{y}_i$ = output pada *tree K*
- f_k = fungsi *decision tree* ke- k
- x_i = fungsi yang dilatih
- K = jumlah *tree* yang dibuat

Proses algoritma XGBoost dapat dijelaskan melalui skema pada Gambar 3.



Sumber: Astutiningsih et al., 2023

<https://journal.itk.ac.id/index.php/sjt/article/download/1031/544>

Gambar 3. Skema Algoritma XGBoost

Pada skema tersebut, bobot setiap pohon diperbaharui pada setiap iterasi untuk memperbaiki kesalahan prediksi sebelumnya. Selain itu, XGBoost menggunakan fungsi objektif yang menggabungkan dua komponen utama: *loss function* ($l(y_i, \hat{y}_i)$) mengukur kesalahan prediksi, dan regularisasi ($\Omega(f_k)$) mengontrol kompleksitas model agar pohon tidak menjadi terlalu dalam dan menghindari *overfitting* (Yulianti). Fungsi objektif XGBoost dirumuskan sebagai:

$$O_b^j(t) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$



$O_j(t)$ = fungsi objektif untuk pohon ke- j pada iterasi ke- t

$l(y_i, \hat{y}_i)$ = kesalahan prediksi antara nilai prediksi $\hat{y}_i$ dan nilai target y_i

$\Omega(f_k)$ = fungsi regularisasi yang mengontrol kompleksitas model

1.4.7.3 Overfitting

Overfitting merupakan kondisi ketika model pembelajaran mesin terlalu menyesuaikan diri terhadap data latih sehingga kehilangan kemampuan untuk melakukan generalisasi terhadap data baru. Model yang mengalami *overfitting* akan menghasilkan kesalahan prediksi yang sangat kecil pada data latih, tetapi memiliki performa buruk pada data uji. Kondisi ini umumnya disebabkan oleh kompleksitas model yang terlalu tinggi, kurangnya regularisasi, atau jumlah data latih yang terbatas (Yulianti et al., 2022).

Dalam konteks XGBoost, *overfitting* dapat terjadi ketika jumlah pohon terlalu banyak, kedalaman pohon terlalu besar, atau nilai learning rate terlalu tinggi. Oleh karena itu, penggunaan regularisasi seperti L1 dan L2 menjadi penting dalam menekan kompleksitas model dan menghindari *overfitting*. Salah satu tanda *overfitting* adalah kesenjangan performa antara data latih dan data uji yang signifikan (Iqbal Baihaqi et al., 2023).

1.4.9 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan dengan tujuan untuk membandingkan berbagai model statistik, mengidentifikasi kelemahan dan kekuatan model yang digunakan, serta memvalidasi hasil analisis statistik yang diperoleh. Dalam prediksi, proses evaluasi model ini penting untuk mengukur seberapa akurat model yang telah dibuat sehingga mampu memberikan prediksi yang akurat. Adapun evaluasi model prediksi yang akan dilihat untuk menentukan metode yang terbaik adalah RMSE, dan MAPE.

1.4.9.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) adalah metrik yang digunakan untuk mengukur akurasi suatu prediksi dengan menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. RMSE merepresentasikan tingkat kesalahan, di mana nilai RMSE yang lebih kecil menunjukkan performa prediksi model yang lebih baik (Wardalisah, 2024). Berikut adalah rumus RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$



a
ke - i
rediksi ke - i

1.4.9.2 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan ukuran akurasi dari suatu prediksi dengan memberikan nilai rata-rata kesalahan absolut antara nilai actual dan nilai prediksi yang menunjukkan persentase *error* data. Hasil prediksi semakin akurat jika nilai MAPE semakin kecil. *Range* MAPE ditunjukkan pada Tabel 1 (Astutiningsih et al., 2023).

Tabel 1. Range nilai MAPE

Range MAPE	Keterangan
< 10%	Akurat
10% - 20%	Baik
20% - 50%	Cukup
> 50	Tidak Akurat

Sumber: Astutiningsih et al., 2023

Rumus *MAPE* dituliskan pada persamaan 6.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (6)$$

Dengan:

n = Jumlah data

y_i = Nilai aktual ke – i

$\hat{y}_i$ = Nilai hasil prediksi ke – i



BAB II METODOLOGI PENELITIAN

2.1 Pendekatan dan Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis deskriptif-komparatif. Pendekatan kuantitatif dipilih karena sesuai untuk mengolah dan menganalisis data numerik secara objektif, khususnya data harga saham dari dua perusahaan BUMN sektor perbankan.

Jenis deskriptif digunakan untuk menggambarkan karakteristik pergerakan data saham selama periode observasi, sedangkan jenis komparatif bertujuan membandingkan akurasi dua metode prediksi, yaitu Regresi Linear dan *Extreme Gradient Boosting* (XGBoost). Regresi Linear mempresentasikan metode statistik klasik berbasis hubungan linier, sedangkan XGBoost dipilih karena kemampuannya menangani data non-linear dan kompleks secara efisien.

2.2 Waktu dan Tempat Penelitian

Penelitian dilaksanakan mulai bulan Oktober 2024 dengan lokasi kerja berbasis daring melalui platform *Google Colaboratory*. Data diperoleh dari situs <https://www.investing.com> mencakup harga saham BBNi dan BBTN untuk periode 1 Januari 2022 hingga 1 Oktober 2024.

2.3 Objek Penelitian

Objek penelitian ini berupa harga saham harian dari PT Bank Negara Indonesia (BBNI) dan PT Bank Tabungan Negara (BBTN). Variabel yang dianalisis meliputi harga penutupan (*price*), harga pembukaan (*open*), harga tertinggi (*high*), dan harga terendah (*low*), dengan total 661 hari perdagangan aktif.

2.4 Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder berupa data historis harga saham harian. Sumber data diperoleh dari situs *Investing.com*, yang menyediakan informasi pasar saham secara lengkap dan kredibel.

Proses pengolahan dan pemodelan dilakukan menggunakan bahasa pemrograman Python melalui platform *Google Colaboratory*, sementara pengujian asumsi klasik regresi linear dilakukan menggunakan perangkat lunak *EViews* untuk mendukung analisis statistik yang lebih rinci.

2.5 Metode Pengumpulan Data



Penelitian dilakukan melalui studi literatur dan pengunduhan dataset dari situs literatur dilakukan dengan menelusuri referensi dari buku, jurnal, penelitian terkait yang relevan dengan algoritma Regresi Linear dan *Extreme Gradient Boosting* (XGBoost). Dataset yang diperoleh digunakan untuk proses analisis dan pemodelan prediksi.

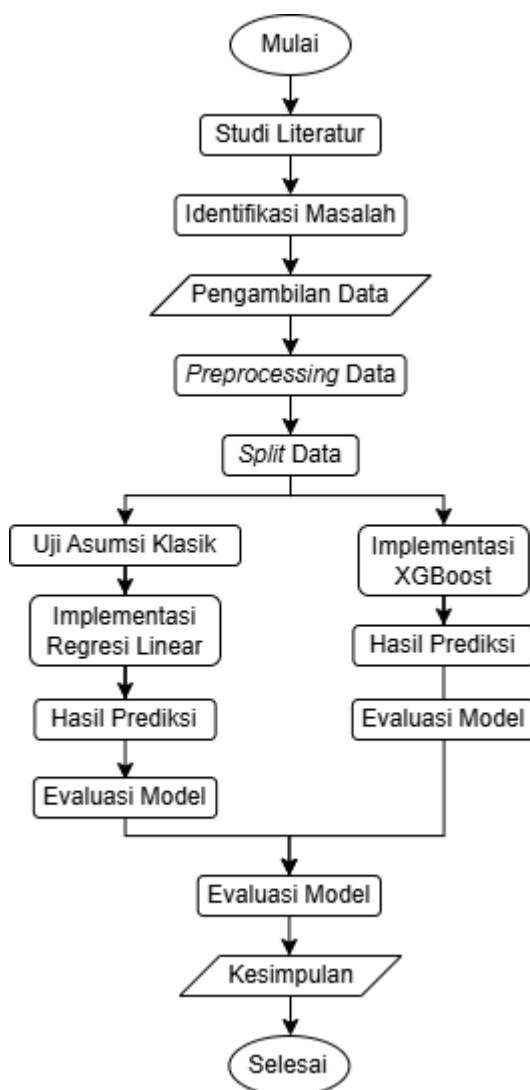
2.6 Metode Analisis Data

Tahapan analisis data yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengumpulkan data harga saham harian BBNI, dan BBTN dari *Investing.com* untuk periode 1 Januari 2022 hingga 1 Oktober 2024.
2. Memilih variabel yang relevan, yaitu: *price*, *open*, *high*, *low* dan tanggal.
3. Melakukan *date parsing* untuk memastikan format tanggal sesuai untuk analisis waktu.
4. Mengurutkan data secara kronologis berdasarkan tanggal (*ascending*).
5. Melakukan *feature engineering* dengan membuat fitur tambahan berupa *lagging data* (lag 1, lag 2) pada masing-masing variabel harga untuk menangkap hubungan temporal antar data.
6. Membagi data menjadi *training set* (80%) dan *test set* (20%) untuk keperluan pelatihan dan pengujian model.
7. Melakukan uji asumsi klasik hanya pada model regresi linear, mencakup: uji linearitas, normalitas, multikolinearitas, heteroskedastisitas, dan autokorelasi menggunakan EViews.
8. Mengimplementasikan model Regresi Linear model XGBoost dengan pengaturan *default*.
9. Mengevaluasi performa kedua model menggunakan metrik RMSE (*Root Mean Square Error*) dan MAPE (*Mean Absolute Percentage Error*).
10. Menganalisis hasil evaluasi untuk membandingkan keakuratan kedua metode berdasarkan nilai RMSE dan MAPE.
11. Menarik kesimpulan dari hasil analisis untuk menentukan metode prediksi yang lebih efektif, serta memberikan rekomendasi yang relevan.



2.7 Alur kerja



Gambar 4. Alur kerja





Optimized using
trial version
www.balesio.com