

**PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT GRADIENT BOOSTING MACHINE* DENGAN *OPTUNA HYPERPARAMETER OPTIMIZATION*
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)**

SKRIPSI



WARDALISAH

H011201078

**PROGRAM STUDI MATEMATIKA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT*
GRADIENT BOOSTING MACHINE DENGAN *OPTUNA*
HYPERPARAMETER OPTIMIZATION
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)**

SKRIPSI



**WARDALISAH
H011201078**

**PROGRAM STUDI MATEMATIKA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT*
GRADIENT BOOSTING MACHINE DENGAN *OPTUNA*
HYPERPARAMETER OPTIMIZATION
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Matematika Departemen Matematika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin**

UNIVERSITAS HASANUDDIN

**WARDALISAH
H011201078**

**PROGRAM STUDI MATEMATIKA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

2024

HALAMAN PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT GRADIENT BOOSTING MACHINE* DENGAN *OPTUNA HYPERPARAMETER OPTIMIZATION*
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)

Adalah benar hasil karya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, 7 Mei 2024



WARDALISAH

NIM. H011201078

PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT GRADIENT BOOSTING MACHINE* DENGAN *OPTUNA HYPERPARAMETER OPTIMIZATION*
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)

Disetujui oleh:

Pembimbing Utama


Dr. Khaeruddin, M.Sc

NIP. 19650914 199103 1 003

Pada 7 Mei 2024

HALAMAN PENGESAHAN

PERAMALAN HARGA SAHAM MENGGUNAKAN METODE *LIGHT GRADIENT BOOSTING MACHINE* DENGAN *OPTUNA HYPERPARAMETER OPTIMIZATION*
(Studi Kasus: Saham PT. Vale Indonesia Tbk.)

Disusun dan diajukan oleh

WARDALISAH

H011201078

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Matematika Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Pada Tanggal 7 Mei 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui

Pembimbing Utama

Dr. Khaeruddin, M.Sc.

NIP. 19650914 199103 1 003

Ketua Program Studi

Dr. Firman, S.Si., M.Si.

NIP. 19680429 200212 1 001

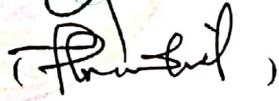
HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Wardalisah
NIM : H011201078
Program Studi : Matematika
Judul Skripsi : Peramalan Harga Saham menggunakan Metode *Light Gradient Boosting Machine* dengan *Optuna Hyperparameter Optimization* (Studi Kasus: Saham PT. Vale Indonesia Tbk.)

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Matematika Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

DEWAN PENGUJI

1. Ketua : Dr. Khaeruddin, M.Sc. ()
2. Anggota : Prof. Dr. Syamsuddin Toaha, M.Sc. ()
3. Anggota : Jusmawati Massalesse, S.Si., M.Si. ()

Ditetapkan di : Makassar

Tanggal : 7 Mei 2024



KATA PENGANTAR

Alhamdulillah, segala puji bagi Allah SWT yang selalu melimpahkan Rahmat dan Hidayah-Nya kepada penulis sehingga skripsi dengan judul “Peramalan Harga Saham menggunakan Metode *Light Gradient Boosting Machine* dengan *Optuna Hyperparameter Optimization* (Studi Kasus: Saham PT. Vale Indonesia Tbk.)” dapat terselesaikan. Shalawat dan salam semoga senantiasa kita curahkan kepada teladan kita Nabi Muhammad SAW, beserta keluarga dan para sahabat. Skripsi ini adalah salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Matematika Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis mengucapkan terima kasih khususnya kepada kedua orang tua Almarhum Bapak **Sudarman** dan Ibu **Nurhawanti** yang telah membesarkan dan mendidik penulis dengan penuh kesabaran serta selalu mencurahkan kasih sayang yang tak pernah putus, memberikan dukungan dan doa sehingga penulis dapat menyelesaikan penulisan skripsi ini. Begitu pula kepada kedua kakak penulis **Hermawanto** dan **Ikhsan Ramadhan**, serta kedua adik penulis **Ferdiansyah** dan **Zahrah** yang telah memberikan dukungan pada penulis. Terima kasih telah menjadi sumber motivasi dan kekuatan bagi penulis selama perjalanan akademik ini.

Penulis menyadari bahwa penyelesaian skripsi ini tidak lepas dari dukungan dan bantuan banyak pihak baik secara langsung maupun tidak langsung. Untuk itu pada kesempatan ini penulis mengucapkan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin.
2. Bapak **Dr. Eng. Amiruddin, S.Si., M.Si.**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.
3. Bapak **Dr. Khaeruddin., M.Sc.**, selaku pembimbing untuk segala ilmu, nasihat, dan kesabaran dalam membimbing dan mengarahkan penulis, serta bersedia meluangkan waktunya untuk mendampingi penulis sehingga skripsi ini dapat diselesaikan.
4. Ibu **Jusmawati Massalesse, S.Si. M.Si.**, selaku penguji dan Bapak **Prof. Dr Syamsuddin Toaha, M.Sc.**, selaku penguji sekaligus penasehat akademik penulis yang telah bersedia meluangkan waktunya untuk memberikan saran dan arahan kepada penulis dalam penulisan skripsi ini.

5. Bapak/Ibu Dosen Departemen Matematika yang telah membagikan ilmu dan pengalamannya, serta Staf Departemen Matematika atas segala bantuannya.
6. Sahabat-sahabat penulis **Sisilia Angraeni, Nurpadian, Rofifah Fakhriyah**, teman-teman semasa SMA penulis **Arini Aulia, Nurul Mutaharah, dan Suhra Istimalah** yang telah membantu, menemani, menyemangati, dan tempat berbagi keluh-kesah penulis selama kurang lebih 4 tahun perkuliahan.
7. Teman-teman penulis **Hilda, Asfi, Fahira, Indah, Wulan, Sulfina, Mona, dan Nurkholisah** yang telah membantu dan menjadi teman diskusi penulis selama perkuliahan dan proses penulisan skripsi.
8. Teman-teman Matematika 2020 atas segala dukungan, kebersamaan, dan kerjasamanya selama ini.
9. Serta segala pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, kritik dan saran yang membangun diharapkan oleh penulis untuk perbaikan dan pengembangan penelitian lebih lanjut. Akhir kata, semoga hasil penelitian ini dapat bermanfaat bagi kalangan akademisi, praktisi, dan semua pihak.

Makassar, 7 Mei 2024

Penulis,



Wardalisah

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMI**

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan dibawah ini:

Nama : Wardalisah
NIM : H011201078
Program Studi : Matematika
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

demii pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

“Peramalan Harga Saham menggunakan Metode *Light Gradient Boosting Machine* dengan *Optuna Hyperparameter Optimization* (Studi Kasus: Saham PT. Vale Indonesia Tbk.)”

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada 7 Mei 2024

Yang menyatakan



Wardalisah

ABSTRAK

Wardalisah. H011201078. “Peramalan Harga Saham menggunakan Metode *Light Gradient Boosting Machine* dengan *Optuna Hyperparameter Optimization* (Studi Kasus: Saham PT. Vale Indonesia Tbk.)” dibimbing oleh **Khaeruddin** sebagai Pembimbing Utama.

Investasi saham merupakan salah satu jenis investasi yang sangat diminati oleh investor. Pergerakan harga saham yang fluktuatif menjadi masalah bagi para investor karena sulit menentukan kapan waktu yang tepat melakukan investasi agar memperoleh keuntungan yang besar dan mengurangi risiko kerugian. Salah satu solusi dari permasalahan tersebut adalah melakukan peramalan pergerakan harga saham menggunakan metode *Light Gradient Boosting Machine* (LightGBM). Pada penelitian ini model LightGBM dibangun dengan menggunakan 4 *features*, yakni harga *open*, *high*, *low*, dan *volume*, serta harga *close* sebagai *target* dengan proporsi data latih dan data uji yakni 80%:20%. Model terbaik pada penelitian ini adalah model LightGBM dengan optimasi parameter menggunakan *Optuna Hyperparameter Optimization*. Hasil penelitian ini adalah model LightGBM mampu meramalkan harga saham dengan sangat baik berdasarkan nilai RMSE yang diperoleh sebesar 21,54 dan MAPE sebesar 0,09%.

Kata Kunci:

Saham, Peramalan, LightGBM, *Optuna Hyperparameter Optimization*.

ABSTRACT

Wardalisah. H011201078. “Stock Price Forecasting using Light Gradient Boosting Machine Method with Optuna Hyperparameter Optimization (Case Study: PT. Vale Indonesia Tbk. Stock)” survived by **Khaeruddin** as the Supervisor.

Stock investment is one type of investment that is very attractive to investors. The fluctuating stock price movements are a problem for investors because it's difficult to determine the right time to invest in order to obtain large profits and minimize the risk of loss. One solution to the problem is forecasting stock price using the Light Gradient Boosting Machine (LightGBM) method. In this study, the LightGBM model was built using 4 features: open, high, low, and volume as well as close price as target with the proportion of training data and test data is 80%:20%. The best model in this study is the LightGBM model with parameter optimization using Optuna Hyperparameter Optimization. The result of this study show that the LightGBM model is able to forecast stock prices very well based on the obtained RMSE value is 21,54 and MAPE of 0,09%.

Keywords:

Stock price, Forecasting, LightGBM, Optuna Hyperparameter Optimization.

DAFTAR ISI

HALAMAN PERNYATAAN KEOTENTIKAN iii

HALAMAN PENGESAHAN v

HALAMAN PENGESAHAN vi

KATA PENGANTAR.....vii

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMI ix

ABSTRAK x

ABSTRACT xi

DAFTAR ISI xii

DAFTAR TABEL xv

DAFTAR GAMBAR xvi

DAFTAR LAMPIRAN xvii

BAB I PENDAHULUAN 1

 1.1 Latar Belakang..... 1

 1.2 Rumusan Masalah..... 3

 1.3 Batasan Masalah 3

 1.4 Tujuan Penelitian 4

 1.5 Manfaat Penelitian..... 4

BAB II TINJAUAN PUSTAKA..... 5

 2.1 Saham..... 5

 2.2 Peramalan..... 6

 2.3 *Gradient Boosting Decision Tree* (GBDT) 6

 2.4 *Light Gradient Boosting Machine* (LightGBM)..... 7

 2.5 *Gradient Based One-Side Sampling* (GOSS)..... 9

2.6	<i>Exclusive Feature Bundling (EFB)</i>	11
2.7	<i>Hyperparameter Tuning</i>	11
2.8	<i>Optuna Hyperparameter Optimization</i>	12
2.9	<i>Cross Validation</i>	12
2.10	Metode Evaluasi Kinerja Model.....	13
2.10.1	Mean Absolute Percentage Error (MAPE).....	14
2.10.2	Root Mean Square Error (RMSE).....	14
BAB III METODOLOGI PENELITIAN.....		15
3.1	Jenis dan Sumber Data.....	15
3.2	Identifikasi Variabel.....	16
3.2.1	Variabel Masukan.....	16
3.2.2	Variabel Keluaran.....	16
3.3	Metode Analisis Data.....	16
3.3.1	Identifikasi Masalah.....	16
3.3.2	Studi Literatur.....	16
3.3.3	Pengumpulan Data.....	16
3.3.4	Preprocessing Data.....	16
3.3.5	Split Data.....	17
3.3.6	Pengimplementasian Algoritma.....	17
3.3.7	Pemilihan Model Terbaik.....	18
3.3.8	Evaluasi Model.....	18
3.3.9	Analisis Hasil.....	18
3.3.10	Menarik Kesimpulan.....	19
3.4	Alur Kerja Penelitian.....	19
BAB IV HASIL DAN PEMBAHASAN.....		20
4.1	Deskripsi Data.....	20

4.2	<i>Preprocessing</i> Data	21
4.2.1	Reduce Memory	22
4.2.2	Data Transformation	22
4.2.3	Input Missing Value	24
4.3	Data <i>Split</i>	25
4.4	Pengimplementasian Algoritma.....	26
4.4.1	Pemodelan LightGBM	26
4.4.2	Training Model	27
4.4.3	Hyperparameter Tuning	28
4.5	Hasil Prediksi Harga Saham.....	30
4.6	Evaluasi Model.....	33
BAB V KESIMPULAN DAN SARAN.....		35
5.1	Kesimpulan.....	35
5.2	Saran	35
DAFTAR PUSTAKA.....		36
LAMPIRAN		38

DAFTAR TABEL

Tabel 4. 1 Data Harga Saham.....	20
Tabel 4. 2 Hasil Optimasi Penggunaan Memori Dataset.....	22
Tabel 4. 3 Missing Value setiap Kolom.....	24
Tabel 4. 4 Statistika Deskriptif.....	25
Tabel 4. 5 Jenis Hyperparameter yang Digunakan.....	26
Tabel 4. 6 Hasil Hyperparameter Tuning dengan Optuna	29
Tabel 4. 7 Nilai RMSE dan MAPE selama Proses Training.....	30
Tabel 4. 8 Nilai RMSE dan MAPE setiap Rasio.....	31
Tabel 4. 9 Hasil Evaluasi Model	34

DAFTAR GAMBAR

Gambar 2. 1 Teknik Pembelahan Leaf-wise Tree Growth.....	9
Gambar 2. 2 Metode Cross Validation.....	13
Gambar 3. 1 Data Harga Saham INCO.JK	15
Gambar 3. 2 Alur Kerja Penelitian	19
Gambar 4. 1 Grafik Data Harga Saham (a) Open, (b) High, (c) Low, (d) Close, dan (e) Volume.....	21
Gambar 4. 2 Dataset Sebelum Proses Transformation	23
Gambar 4. 3 Dataset Setelah Proses Transformation	24
Gambar 4. 4 Pembagian Data Latih dan Data Uji.....	26
Gambar 4. 5 Hasil Kinerja Model selama Proses Training.....	27
Gambar 4. 6 Grafik Hasil Prediksi Data Latih.....	28
Gambar 4. 7 Grafik Hasil Prediksi Data Latih dengan Optuna	29
Gambar 4. 8 Grafik Hasil Prediksi dan Data Uji	31
Gambar 4. 9 Fitur Signifikan selama Proses Peramalan dengan Data Uji	32
Gambar 4. 10 Struktur Pohon Keputusan	33

DAFTAR LAMPIRAN

Lampiran 1 : Dataset Harga Saham PT. Vale Indonesia Tbk.	38
Lampiran 2 : Source Code Python.....	38

BAB I

PENDAHULUAN

1.1 Latar Belakang

Investasi adalah menempatkan modal atau dana pada suatu aset dengan harapan memberikan hasil atau peningkatan nilai di masa yang akan datang. Salah satu jenis investasi yang banyak diminati saat ini adalah investasi saham. Hermuningsih (2019) mendeskripsikan saham sebagai salah satu surat berharga yang diperdagangkan di pasar modal dan bersifat kepemilikan. Investasi saham banyak diminati oleh para investor karena dapat memperoleh keuntungan yang besar dibandingkan dengan jenis investasi lainnya, walaupun memiliki risiko kerugian yang besar dalam waktu singkat. Risiko kerugian dapat diminimalkan dengan melakukan analisis secara fundamental terhadap pergerakan harga saham. Selain itu, membuat model peramalan pergerakan saham juga dapat membantu dalam mengurangi risiko kerugian dan menambah potensi keuntungan karena mampu dijadikan sebagai acuan perencanaan yang lebih efektif dan efisien untuk mengetahui pergerakan harga saham selanjutnya.

Pergerakan harga saham sulit untuk diprediksi perubahannya karena harga saham dipengaruhi oleh beberapa faktor fundamental, teknikal, dan sentimen yang harus dipertimbangkan dalam meramalkan harga saham selanjutnya. Harga saham adalah data *time series* atau deret waktu yang selalu mengalami fluktuasi yaitu dapat naik dan turun, sehingga dalam meramalkan harga saham dapat menggunakan metode peramalan *time series* seperti metode tradisional *Autoregressive Integrated Moving Average* (ARIMA), metode *deep learning* seperti *Long Short-Term Memory* (LSTM) dan metode *gradient boosting* seperti *Extreme Gradient Boosting* (XGBoost). Namun, Tussifah (2022) menemukan bahwa metode tradisional memiliki nilai akurasi yang lebih rendah jika dibandingkan dengan metode *machine learning* dalam peramalan harga saham.

Seiring dengan berkembangnya ilmu pengetahuan, banyak peneliti yang mengusulkan metode untuk memprediksi pergerakan harga saham berdasarkan data historis secara otomatis. Yang dkk. (2021) menemukan bahwa metode *Gradient Boosting* dapat digunakan untuk memprediksi harga saham dan memiliki akurasi yang lebih tinggi dibandingkan dengan metode *Neural Network*. *Extreme Gradient*

Boosting (XGBoost) dan *Light Gradient Boosting Machine* (LightGBM) merupakan implementasi dari *Gradient Boosting Decision Tree* (GBDT), namun metode LightGBM memiliki performa yang lebih baik jika dibandingkan dengan metode XGBoost (Rufo dkk., 2021). Metode XGBoost apabila digunakan dalam peramalan dengan data yang besar masih rentan mengalami *overfitting*, yaitu keadaan saat suatu model mempelajari data latih terlalu baik, sehingga apabila dilakukan *testing* dengan menggunakan data yang berbeda dapat mengurangi akurasi. Sedangkan metode LightGBM mampu menghindari hal tersebut. Selain itu, dalam aspek kecepatan proses *training*, LightGBM lebih unggul dibandingkan dengan metode XGBoost. LightGBM adalah metode ensemble yang dikembangkan oleh Microsoft dan memiliki kelebihan dalam kecepatan *training* dan lebih efisien, penggunaan memori yang lebih rendah, memiliki tingkat akurasi yang baik, dan mampu mengatasi kesalahan pada data ukuran besar (Ke dkk., 2017). Bahkan, pada kompetisi M5 *Forecasting-Accuracy*, LightGBM lebih unggul dibandingkan metode lainnya dan efektif dalam memproses data yang berkorelasi dan bersifat eksogen serta mampu meminimalisir kesalahan model peramalan.

Adapun penelitian sebelumnya terkait penerapan metode LightGBM pernah dilakukan Sun dkk. (2020) dengan menerapkan metode ini pada peramalan tren harga mata uang kripto. Hasil penelitian tersebut menyimpulkan bahwa metode LightGBM lebih baik dibandingkan dengan metode lain dan efektif dalam membantu investor dalam membuat portofolio yang tepat dan mengurangi risiko kerugian.

Performa dari suatu algoritma atau model sangat bergantung pada pemilihan *hyperparameter*, maka dari itu diperlukan alternatif dalam mempermudah pemilihan *hyperparameter* tersebut. *Grid Search Cross Validation* (GSCV) dan *Random Search* adalah dua metode yang paling umum digunakan untuk melakukan penentuan *hyperparameter*. Namun, GSCV dan *Random Search* memerlukan biaya komputasi yang mahal dan hanya efektif digunakan untuk data kecil (Srinivas dan Katarya, 2022). Metode lain yang lebih efisien untuk penentuan *hyperparameter* adalah metode *Optuna Hyperparameter Optimization*. Metode ini dirancang khusus untuk mengatasi permasalahan *hyperparameter* pada tugas *machine learning*

dengan fokus pada kesederhanaan dan efisiensi yang lebih baik dibandingkan dengan metode GSCV dan *Random Search* (Akiba dkk., 2019).

PT. Vale Indonesia Tbk. merupakan salah satu perusahaan pertambangan terkemuka di Indonesia yang menambang nikel laterit untuk menghasilkan produk akhir berupa nikel dalam matte. Nikel merupakan penghantar listrik dan panas yang cukup baik, sehingga bisa digunakan sebagai bahan baku dalam pembuatan baterai kendaraan listrik. Zaidan dan Garinan (2021) menyatakan bahwa pada tahun 2030 mendatang, sebanyak 20% dari total konsumsi nikel global diperkirakan akan digunakan untuk baterai mobil listrik maupun baterai sistem penyimpanan energi. Sebagai salah satu produsen nikel di Indonesia, PT. Vale Indonesia Tbk. memiliki peluang baru yang signifikan pada masa transisi menuju kendaraan listrik. Oleh karena itu, harga saham PT. Vale Indonesia Tbk. menjadi salah satu harga saham yang menarik untuk diprediksi.

Berdasarkan latar belakang di atas, maka peneliti tertarik untuk melakukan penelitian dengan judul **“Peramalan Harga Saham menggunakan Metode *Light Gradient Boosting Machine* dengan *Optuna Hyperparameter Optimization* (Studi Kasus: Saham PT. Vale Indonesia Tbk.)”**.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini sebagai berikut:

1. Bagaimana akurasi hasil prediksi model LightGBM dalam memprediksi harga saham?
2. Bagaimana *hyperparameter tuning* dengan *Optuna Hyperparameter Optimization* dalam meningkatkan kinerja model LightGBM untuk peramalan harga saham?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini sebagai berikut:

1. Penelitian ini menggunakan algoritma LightGBM dalam membangun model peramalan.
2. *Hyperparameter tuning* menggunakan metode *Optuna Hyperparameter Optimization*.
3. *Tools* bahasa pemrograman yang digunakan dalam penelitian ini adalah *Python*.

4. Data yang digunakan dalam penelitian ini adalah data harga saham PT. Vale Indonesia Tbk. yang diambil dari Januari 2018 sampai Desember 2023.
5. Data input yang digunakan adalah data historis harga saham satu hari sebelumnya yang meliputi harga *low*, *open*, *close*, *high*, dan *volume* untuk meramalkan data harga penutupan saham satu hari berikutnya.
6. Parameter yang digunakan untuk penentuan akurasi hasil peramalan harga saham adalah evaluasi model menggunakan *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini sebagai berikut:

1. Memperoleh hasil akurasi kinerja model peramalan dengan algoritma LightGBM dalam memprediksi harga saham.
2. Menganalisis pengaruh *hyperparameter tuning* dengan *Optuna Hyperparameter Optimization* dalam meningkatkan kinerja model LightGBM.

1.5 Manfaat Penelitian

Penelitian ini bermanfaat dalam pengembangan pengetahuan khususnya dalam pengimplementasian metode LightGBM dalam peramalan harga saham dan memberikan model proyeksi terhadap harga saham bagi para investor dalam melakukan investasi saham.

BAB II

TINJAUAN PUSTAKA

2.1 Saham

Saham menurut Fauzi (2019) adalah salah satu instrumen pasar keuangan atau surat berharga yang menunjukkan kepemilikan seseorang atas suatu perusahaan maupun badan usaha yang menanamkan modal pada suatu perusahaan asal. Semakin besar saham yang dimiliki, maka semakin besar kekuasaannya di perusahaan tersebut. Saham merupakan salah satu jenis investasi yang banyak diminati oleh para investor. Harga saham selalu mengalami fluktuasi sehingga mampu memberikan keuntungan yang besar sekaligus mampu memberikan kerugian yang besar dalam waktu yang singkat, sehingga tingkat resikonya lebih tinggi jika dibandingkan dengan jenis investasi lainnya.

Pergerakan saham biasanya bersifat non linier dan non stasioner karena dipengaruhi oleh banyak faktor, sehingga pergerakannya sangat kompleks dan sulit untuk ditebak. Dalam menganalisis bagaimana pergerakan harga saham terdapat tiga faktor yang perlu dipertimbangkan yaitu faktor teknikal, faktor fundamental, dan faktor sentimen. Faktor teknikal adalah pergerakan yang mengamati bagaimana harga saham di masa lalu, faktor fundamental adalah teknik analisis pendekatan secara bisnis yang terjadi, dan faktor sentiment adalah pergerakan harga saham yang dipengaruhi oleh faktor-faktor bisnis dan pelaku operasional bisnis (Suyudi dkk., 2019).

Data harga saham dapat diamati berdasarkan harga *high price*, *low price*, *open price*, *close price*, *volume*, dan *adjusted close*. Harga *low* dan *high* adalah harga pencapaian terendah dan tertinggi setiap harinya. Harga *open* dan *close* adalah harga pembukaan dan penutupan setiap harinya. *Volume* adalah banyaknya pertukaran yang diperdagangkan selama beberapa periode waktu. *Adjusted close* adalah harga penutupan setelah penyesuaian untuk semua pembagian saham dan pembagian dividen yang berlaku.

Secara umum perubahan harga saham dipengaruhi oleh beberapa faktor internal dan faktor eksternal suatu perusahaan. Faktor yang berhubungan dengan kinerja perusahaan disebut faktor internal, seperti perubahan harga, ekspansi, struktur modal, tingkat laba perusahaan, kebijakan dividen, dan lainnya. Sedangkan

faktor eksternal yang mempengaruhi harga saham yaitu berhubungan dengan kebijakan perekonomian suatu negara seperti inflasi, suku bunga, kurs nilai tukar, dan kebijakan pemerintah (Sukartaatmadja dkk., 2023).

2.2 Peramalan

Peramalan adalah suatu proses menganalisis pola dari data masa lalu dan sekarang untuk memperkirakan tentang sesuatu yang mungkin terjadi di masa depan. Hasil peramalan tidak selalu memberikan jawaban yang sesuai dengan yang akan terjadi, namun akan selalu diusahakan untuk memberikan jawaban sedekat mungkin dengan apa yang akan terjadi sehingga mampu meminimalisir kesalahan di masa depan (Herdianto, 2013).

Makridakis dkk. (1999) mendeskripsikan metode peramalan menjadi dua kategori utama, yaitu peramalan kuantitatif dan peramalan kualitatif. Peramalan kuantitatif, yaitu peramalan yang menggunakan data-data kuantitatif yang terdapat di masa lalu yang diperoleh dari pengamatan nilai-nilai sebelumnya sedangkan peramalan kualitatif, yaitu peramalan yang menggunakan data kualitatif yang diambil dari masa lalu, seperti intuisi pengambilan keputusan, emosi, pengalaman pribadi, dan sistem nilai. Hasil dari peramalan kualitatif berdasarkan pengamatan kejadian di masa lalu digabungkan dengan pemikiran dari penyusunnya.

2.3 *Gradient Boosting Decision Tree (GBDT)*

Gradient Boosting Decision Tree (GBDT) merupakan algoritma *machine learning* yang banyak digunakan karena memiliki tingkat efisiensi dan akurasi yang tinggi, serta kemampuan interpretasinya dalam menangani masalah klasifikasi, prediksi, penentuan peringkat dan beberapa tugas *machine learning* lainnya. GBDT adalah model *ensemble* dari pohon keputusan yang dilatih secara berurutan. Pada setiap iterasi, GBDT mempelajari pohon keputusan dengan menyesuaikan gradien negatif atau *residual errors* (Friedman, 2001).

Bagian yang paling memakan waktu dalam proses *training* pohon keputusan adalah menentukan titik-titik pemisah terbaik. Salah satu algoritma yang paling populer untuk menentukan titik pisah adalah algoritma *pre-sorted*, yaitu algoritma yang menghitung semua titik pisah yang mungkin pada nilai fitur yang telah diurutkan sebelumnya. Algoritma ini sederhana dan dapat menemukan titik pisah yang optimal, namun tidak efisien dalam hal kecepatan pelatihan dan konsumsi

memori. Algoritma lain yang populer adalah *histogram-based algorithm*. Algoritma ini memasukkan nilai fitur kontinu ke dalam *discrete bins* dan menggunakan *bins* tersebut untuk membuat histogram fitur selama pelatihan (Ke dkk., 2017).

GBDT menggunakan pohon keputusan untuk mempelajari fungsi dari ruang input X^s ke ruang gradien G . Misalkan terdapat satu set data *instance* $X = \{x_1, x_2, \dots, x_n\}$, dimana setiap x_i adalah vektor dengan dimensi s dalam ruang X^s . Dalam setiap iterasi peningkatan gradien, gradien negatif dari *loss function* terhadap output dari model dinotasikan sebagai $\{g_1, g_2, \dots, g_n\}$. Model pohon keputusan membagi setiap *node* pada fitur yang paling informatif (*information gain* terbesar). Perolehan informasi biasanya diukur dengan variansi setelah pemisahan. Misalkan O adalah dataset pelatihan pada sebuah *fixed node* dari pohon keputusan. Variansi *gain* dari pemisahan fitur j pada titik d untuk *node* ini didefinisikan oleh persamaan (2.1) berikut (Ke dkk., 2017):

$$V_{j|O}(d) = \frac{1}{n_O} \left(\frac{\left(\sum_{\{x_i \in O: x_{ij} \leq d\}} g_i \right)^2}{n_{l|O}^j(d)} + \frac{\left(\sum_{\{x_i \in O: x_{ij} > d\}} g_i \right)^2}{n_{r|O}^j(d)} \right) \quad (2.1)$$

dimana:

$V_{j|O}(d)$: Variansi *gain* untuk fitur j dalam himpunan O pada titik d

n_O : Jumlah titik data pada himpunan O

g_i : Gradien negatif dari *loss function*

$n_{l|O}^j(d)$: Jumlah titik data pada himpunan O yang memiliki nilai fitur $j \leq d$

$n_{r|O}^j(d)$: Jumlah titik data pada himpunan O yang memiliki nilai fitur $j > d$.

2.4 *Light Gradient Boosting Machine (LightGBM)*

Light Gradient Boosting Machine (LightGBM) merupakan algoritma yang dirancang oleh *Microsoft Research Asia* berdasarkan kerangka *Gradient Boosting Decision Tree* (GBDT) (Ke dkk., 2017). LightGBM merupakan salah satu model *Gradient Boosting* yang dapat digunakan dalam penentuan peringkat, klasifikasi, regresi, dan beberapa tugas *machine learning* lainnya dengan cepat, terdistribusi dan berkinerja tinggi berdasarkan algoritma pohon keputusan. Keunggulan dari metode LightGBM dibandingkan dengan metode GBDT lainnya yaitu, kecepatan pelatihan yang lebih cepat, efisiensi yang lebih tinggi, penggunaan memori yang

lebih rendah, tingkat akurasi yang lebih baik, kemampuan dalam menangani data dengan skala yang besar, dan didukung oleh pembelajaran paralel dan GPU (Rufodkk., 2021).

LightGBM mengimplementasikan algoritma GBDT dengan menambahkan dua teknik baru yaitu *Gradient Based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB) untuk meningkatkan efisiensi dan skalabilitas GBDT. Algoritma LightGBM berkinerja dengan baik dalam kompetisi *machine learning* karena mampu menangani berbagai jenis data, hubungan, distribusi, dan keragaman *hyperparameter* yang dapat disempurnakan.

Zhang (2020) dalam penelitiannya mengasumsikan bahwa terdapat himpunan data $N = \{1, 2, \dots, n\}$ dan model LightGBM yang memiliki pohon $T = \{1, 2, \dots, t\}$, setelah iterasi t kali, maka prediksi akhir sama dengan jumlah pertama $(1 - t)th$ dan tth . Proses iterasi digambarkan oleh persamaan (2.2) berikut:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_i(x_i) \quad (2.2)$$

dimana:

- $\hat{y}_i^{(t)}$: Nilai prediksi pada iterasi tth
- $\hat{y}_i^{(t-1)}$: Model pohon yang dihasilkan sebelumnya
- $f_i(x_i)$: Model baru yang dibangun.

Berdasarkan persamaan (2.2) diperoleh bahwa prediksi baru dihasilkan oleh residu dan prediksi sebelumnya. Dalam mengurangi kompleksitas model dan meningkatkan kegunaan pada dataset, dapat menggunakan proses latihan dan regularisasi yang digambarkan oleh persamaan (2.3) dan (2.4) berikut (Rizky dkk., 2022):

$$\begin{cases} \hat{y}_i^{(0)} = 0 \\ \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \end{cases} \quad (2.3)$$

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \sum_{t=1}^T \Omega(f_t) \quad (2.4)$$

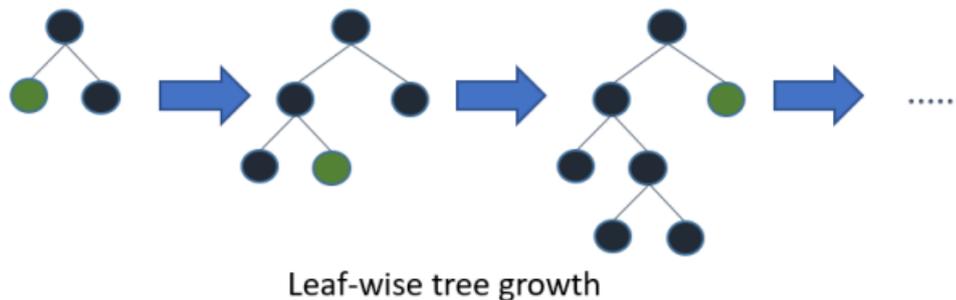
$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$$

dimana:

- y_i : Nilai aktual

- $\hat{y}_i^{(t)}$: Nilai prediksi
 $\mathcal{L}^{(t)}$: Jumlah *loss* pada iterasi t
 $\sum l$: Jumlah *loss* antara y_i dan $\hat{y}_i^{(t)}$
 $\Omega(f_t)$: Nilai regularisasi
 T : Jumlah daun
 ω : Bobot daun
 γ dan λ : koefisien dengan nilai $\gamma = 0$ dan $\lambda = 1$

Karakteristik yang membedakan LightGBM dengan model *gradient boosting* lainnya, yaitu LightGBM menggunakan teknik pembelahan *leaf-wise growth strategy* yang akan membatasi kedalaman dari model LightGBM dengan mencari *leaf-node* dengan nilai *splitting gains* terbesar, lalu memecah *node* tersebut, dan meneruskan proses untuk *node* yang baru. Selain itu, untuk mencegah model LightGBM tumbuh terlalu dalam, level tambahan tidak lagi digunakan untuk meningkatkan *purity level* atau tingkat kemurnian dari model. Hal ini juga akan membuat sistem akan mengkonsumsi daya komputasi yang jauh lebih sedikit dan dapat menghindari *overfitting* (Zhang, 2020).



Gambar 2. 1 Teknik Pembelahan *Leaf-wise Tree Growth*

2.5 Gradient Based One-Side Sampling (GOSS)

Gradient Based One-Side Sampling (GOSS) adalah teknik optimalisasi pada algoritma LightGBM untuk menghilangkan data dengan gradien kecil (Deshtiara, 2023). Teknik ini dikembangkan untuk mengurangi waktu komputasi dan penggunaan memori pada model data besar (Ke dkk., 2017). GOSS dapat memungkinkan model untuk lebih fokus pada data *instance* yang penting dengan mempertahankan sebagian kecil data *instance* yang memiliki gradien lemah (*weak gradient*). Data *instance* dengan gradien lemah akan dihapus sebagian secara acak

sehingga jumlah data *instance* dapat berkurang secara signifikan tanpa mengurangi performa model. Hal ini juga dapat membantu dalam membuat model yang lebih cepat dan efisien pada data yang besar (Febriantoro dkk., 2023).

GOSS menyimpan semua data *instance* yang memiliki gradien besar lalu melakukan *sampling* secara acak pada data *instance* yang memiliki gradien kecil. Algoritma GOSS menggunakan pengali konstan untuk data *instance* dengan gradien kecil ketika menghitung nilai *gain* agar mampu mengkompensasi pengaruh terhadap distribusi data. Secara umum, GOSS mengurutkan data *instance* berdasarkan nilai sampel gradien dalam urutan menurun, lalu menyimpan data urutan teratas $a \times 100\%$ dari total data *instance* dan mendapatkan subset A , kemudian untuk himpunan B terdiri dari $(1 - a) \times 100\%$ dari total data *instance* dengan gradien yang lebih kecil. Kemudian, data *instance* dengan gradien kecil akan diperkuat dengan konstanta $\frac{1-a}{b}$. Langkah terakhir yaitu *instance* dibagi sesuai dengan estimasi variansi *gain* $\tilde{V}_j(d)$ dari subhimpunan $A \cup B$, atau didefinisikan oleh persamaan (2.5) berikut (Ke dkk., 2017):

$$\tilde{V}_j(d) = \frac{1}{n} \left(\frac{(\sum_{\{x_i \in A_l\}} g_i + \frac{1-a}{b} \sum_{\{x_i \in B_l\}} g_i)^2}{n_l^j(d)} + \frac{(\sum_{\{x_i \in A_r\}} g_i + \frac{1-a}{b} \sum_{\{x_i \in B_r\}} g_i)^2}{n_r^j(d)} \right), \quad (2.5)$$

dimana:

- $\tilde{V}_j(d)$: Variansi *gain* untuk fitur j pada titik d
- n : Jumlah titik data pada himpunan A dan B
- g_i : Gradien negatif dari *loss function*
- A_l : Subhimpunan dari A yang terdiri dari titik-titik dengan nilai fitur $j \leq d$
- A_r : Subhimpunan dari A yang terdiri dari titik-titik dengan nilai fitur $j > d$
- B_l : Subhimpunan dari B yang terdiri dari titik-titik dengan nilai fitur $j \leq d$
- B_r : Subhimpunan dari B yang terdiri dari titik-titik dengan nilai fitur $j > d$
- $n_l^j(d)$: Jumlah titik data yang memiliki nilai fitur $j \leq d$
- $n_r^j(d)$: Jumlah titik data yang memiliki nilai fitur $j > d$

Dengan demikian, dalam GOSS akan menggunakan estimasi $\tilde{V}_j(d)$ pada subset *instance* yang lebih kecil, sehingga biaya komputasi dapat dikurangi secara signifikan dan model yang digunakan dapat lebih fokus pada *instance* yang penting tanpa banyak mengubah distribusi data asli.

2.6 *Exclusive Feature Bundling (EFB)*

Exclusive Feature Bundling (EFB) merupakan teknik pengoptimalan pada algoritma LightGBM untuk mengurangi biaya komputasi dan memori pada model yang memiliki banyak fitur (Febriantoro, 2023). EFB digunakan untuk menemukan kelompok fitur yang berkorelasi tinggi dan membangun fitur baru yang merupakan gabungan atau *bundling* dari fitur-fitur tersebut. Cara kerja EFB adalah dengan menggunakan *Correlation-based Feature Selection (CFS)* untuk melakukan eksplorasi dalam menemukan pasangan-pasangan fitur yang memiliki korelasi yang tinggi, kemudian menggunakan teknik *binning* untuk menggabungkan pasangan-pasangan fitur tersebut untuk membentuk fitur baru (Ke dkk., 2017).

Teknik *binning* dilakukan dengan mengubah setiap nilai pada pasangan-pasangan fitur menjadi label *bin*. Selanjutnya, gabungan dari pasangan-pasangan fitur tersebut dibangun dengan melakukan pengkodean label *bin* menjadi variabel baru pada model untuk menghasilkan fitur baru. Fitur baru ini selanjutnya digunakan sebagai input pada algoritma LightGBM. EFB dapat mengurangi dimensi data yang besar, menghilangkan fitur-fitur dengan korelasi yang rendah, serta membangun fitur baru yang lebih baik, dan meningkatkan performa model dengan menggunakan cara tersebut. Khususnya pada data dengan jumlah fitur yang sangat banyak, EFB mampu meningkatkan kecepatan dan efisiensi dari model LightGBM (Ke dkk., 2017).

EFB mampu menggabungkan banyak fitur eksklusif ke fitur yang jauh lebih sedikit sehingga dapat menghindari komputasi yang tidak perlu untuk nilai fitur nol secara efektif. Pengoptimalan algoritma histogram sebenarnya juga mampu mengabaikan nilai fitur nol dengan menggunakan tabel untuk merekam data setiap fitur dengan nilai bukan nol. Namun, metode ini membutuhkan memori tambahan dan biaya komputasi yang lebih besar dibandingkan dengan algoritma EFB (Ke dkk., 2017).

2.7 *Hyperparameter Tuning*

Hyperparameter adalah variabel yang menentukan bagaimana suatu model dilatih untuk mendapatkan performa yang baik. *Hyperparameter Tuning* merupakan proses mencari nilai optimal dari *hyperparameter* suatu model *machine learning* untuk memperbaiki performa model. *Hyperparameter Tuning* sangat

berperan penting dalam model dengan kasus struktur yang kompleks (Yu dan Zhu, 2020). Algoritma *machine learning* seperti *gradient boosting*, *random forest*, dan *neural networks* untuk regresi dan klasifikasi melibatkan banyak *hyperparameter* yang harus ditetapkan sebelum menjalankannya (Probst dkk., 2019). *Hyperparameter Tuning* hanya dapat diterapkan ketika ukuran datanya besar, memiliki banyak *hyperparameter* yang kompleks, dan model dengan struktur yang dirancang dengan baik menyiratkan bahwa *hyperparameter* harus disetel ke rentang yang ketat dan perlu ketepatan.

2.8 *Optuna Hyperparameter Optimization*

Optuna adalah suatu *framework* optimasi *hyperparameter* yang dirancang khusus untuk menemukan kombinasi *hyperparameter* yang optimal untuk model *machine learning* secara otomatis dan efisien. *Optuna* menggunakan kode API gaya *define-by-run*, sehingga lebih mudah dipahami dengan kode yang sebenarnya, dan memungkinkan pengguna untuk membangun ruang pencarian secara dinamis. Penggunaan *Optuna* untuk model *machine learning* seperti LightGBM adalah dengan merumuskan optimasi *hyperparameter* sebagai proses meminimalkan atau memaksimalkan *objective function* atau fungsi objektif yang mengambil sekumpulan *hyperparameter* sebagai *input* dan nilai validasi sebagai *output* (Akiba dkk., 2019).

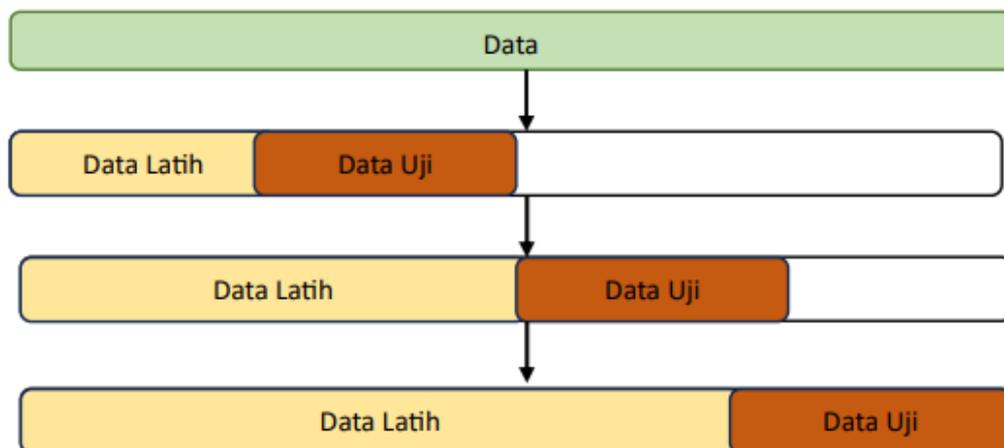
Optuna secara bertahap membangun fungsi objektif melalui interaksi dengan data uji. *Optuna* menggunakan teknik optimasi *Bayesian* dengan algoritma *Tree-structured Parzen Estimator* (TPE) untuk mencari kombinasi *hyperparameter* yang memberikan hasil terbaik dalam konteks fungsi objektif yang diberikan. Pendekatan ini memungkinkan *Optuna* memodelkan fungsi tujuan secara berulang dan menemukan nilai *hyperparameter* yang paling optimal. Kelebihan dari *Optuna* adalah *Optuna* memiliki arsitektur yang ringan dan serbaguna, penentuan ruang pencarian menggunakan sintaks *Python*, algoritma optimasi yang efisien, paralelasi yang mudah, serta visualisasi yang cepat (Akiba dkk., 2019).

2.9 *Cross Validation*

Cross Validation atau validasi silang adalah salah satu metode *resampling* data yang digunakan untuk memperhitungkan kesalahan prediksi model yang sebenarnya dan mencegah terjadinya *overfitting*. *Cross Validation* membagi *dataset*

secara silang dengan menjadikan dua bagian dengan satu bagian dijadikan set data latih dan bagian lain dijadikan set data uji yang kemudian akan dilakukan proses evaluasi kinerja model pada data uji dengan pelatihan model pada data latih. Evaluasi kinerja model deret waktu dapat menggunakan metode *cross validation* secara bergulir (Prihanditya, 2020).

Cara kerja dari metode *cross validation* adalah dengan menggunakan sebagian kecil data awal sebagai data latih kemudian memeriksa akurasi pada bagian setelahnya yang digunakan sebagai peramalan. Data yang sebelumnya digunakan sebagai peramalan kemudian digabungkan dengan data latih sebelumnya dan digunakan kembali sebagai pelatihan selanjutnya, serta diperiksa kembali akurasi pada sebagian data setelahnya. Langkah-langkah tersebut dilakukan berulang-ulang sampai pada data terakhir. Hasil evaluasi model diperoleh dari rata-rata setiap perhitungan akurasi peramalan (Prihanditya, 2020).



Gambar 2. 2 Metode *Cross Validation*

2.10 Metode Evaluasi Kinerja Model

Evaluasi kinerja model dalam metode *machine learning* penting untuk dilakukan agar dapat diukur tingkat akurasi kinerja suatu model yang telah digunakan. Pada penelitian ini, dalam mengukur kinerja model yang telah dibuat akan menggunakan metode evaluasi *Mean Absolute Percentage Error* (MAPE) untuk melihat presentase *error* data hasil peramalan dan *Root Mean Square Error* (RMSE) untuk melihat tingkat kesalahan prediksi berdasarkan selisih antara nilai aktual dan nilai prediksi. Terdapat beberapa metrik evaluasi yang dapat digunakan

dalam mengevaluasi kinerja suatu model, namun MAPE dan RMSE sudah cukup untuk memastikan akurasi dari hasil peramalan yang diperoleh.

2.10.1 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 aktual dan nilai prediksi yang menunjukkan persentase *error* data. Berikut rumus dari MAPE:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (2.6)$$

Keterangan:

- n : Jumlah data
- y_i : Nilai aktual ke $- i$
- \hat{y}_i : Nilai hasil prediksi ke $- i$

Nilai evaluasi MAPE memiliki kriteria sebagai berikut (Siringoringo, 2021):

MAPE < 10% : Kemampuan prediksi sangat baik.

10% ≤ MAPE < 20% : Kemampuan prediksi baik.

20% ≤ MAPE < 50% : Kemampuan prediksi cukup.

MAPE ≥ 50% : Kemampuan prediksi buruk.

2.10.2 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) merupakan ukuran akurasi dari suatu prediksi dengan memberikan nilai akar kuadrat dari rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi. RMSE menunjukkan besarnya tingkat kesalahan prediksi, dimana semakin kecil nilai RMSE maka semakin baik performa prediksi dari model. Berikut rumus RMSE

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.7)$$

Keterangan:

- n : Jumlah data
- y_i : Nilai aktual ke $- i$
- \hat{y}_i : Nilai hasil prediksi ke $- i$