

**PREDIKSI HARGA SAHAM MENGGUNAKAN LIGHT
GRADIENT BOOSTING MACHINE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

YANUAR NUR KHOLIK

18.12.0974

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**PREDIKSI HARGA SAHAM MENGGUNAKAN LIGHT
GRADIENT BOOSTING MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

YANUAR NUR KHOLIK

18.12.0974

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

**PREDIKSI HARGA SAHAM MENGGUNAKAN LIGHT GRADIENT
BOOSTING MACHINE**

yang disusun dan diajukan oleh

Yanuar Nur Kholik

18.12.0974

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Juni 2022

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

HALAMAN PENGESAHAN
SKRIPSI
PREDIKSI HARGA SAHAM MENGGUNAKAN LIGHT GRADIENT
MACHINE

yang disusun dan diajukan oleh

Yanuar Nur Kholik

18.12.0974

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Asro Nasiri, Drs, M.Kom.
NIK. 190302152

Norhikmah, M.Kom.
NIK. 190302245

Anggit Dwi Hartanto, M.Kom.
NIK. 190302163

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Yanuar Nur Kholik
NIM : 18.12.0974

Menyatakan bahwa Skripsi dengan judul berikut:

Prediksi Harga Saham Menggunakan Light Gradient Boosting Machine

Dosen Pembimbing: **Anggit Dwi Hartanto, M.Kom.**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 Agustus 2022

Yang Menyatakan,



Yanuar Nur Kholik

KATA PENGANTAR

Segala puji bagi Allah SWT karena atas nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Dalam proses penyelesaian skripsi ini banyak pihak yang telah memberikan bantuan, dukungan, saran dan kritik yang telah penulis terima, oleh karena itu dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Allah SWT, yang telah memberikan pertolongan dan kekuatan dalam proses pembuatan skripsi ini.
2. Kedua orangtua penulis, Bapak Masnan dan Ibu Ratna Andriyani yang selalu mendoakan dan selalu memberikan dukungan kepada penulis dalam menyelesaikan skripsi ini.
3. Bapak Anggit Dwi Hartanto, M.Kom selaku dosen pembimbing skripsi, atas waktu yang telah diberikan untuk membimbing, mengarahkan, memberi dukungan, motivasi serta memberikan masukan kepada penulis dalam pengerjaan skripsi ini hingga akhir.
4. Seluruh dosen Fakultas Ilmu Komputer Universitas Amikom Yogyakarta atas segala ilmu pengetahuan yang diberikan kepada penulis.
5. Segenap dosen Jurusan Sistem Informasi yang telah memberikan bimbingan dan ilmu kepada penulis selama masa studi.
6. Seluruh rekan-rekan studi yang tidak dapat disebutkan satu persatu, terima kasih atas segala kebaikan yang diberikan kepada penulis.

Berbagai kekurangan dan kesalahan mungkin pembaca temukan dalam penulisan skripsi ini, untuk itu kritik dan saran sangat diharapkan untuk perbaikan yang akan datang. Semoga apa yang menjadi kekurangan bisa disempurnakan dan semoga skripsi ini dapat bermanfaat dan menambah wawasan serta pengetahuan bagi para pembaca.

Yogyakarta, 18 September 2022

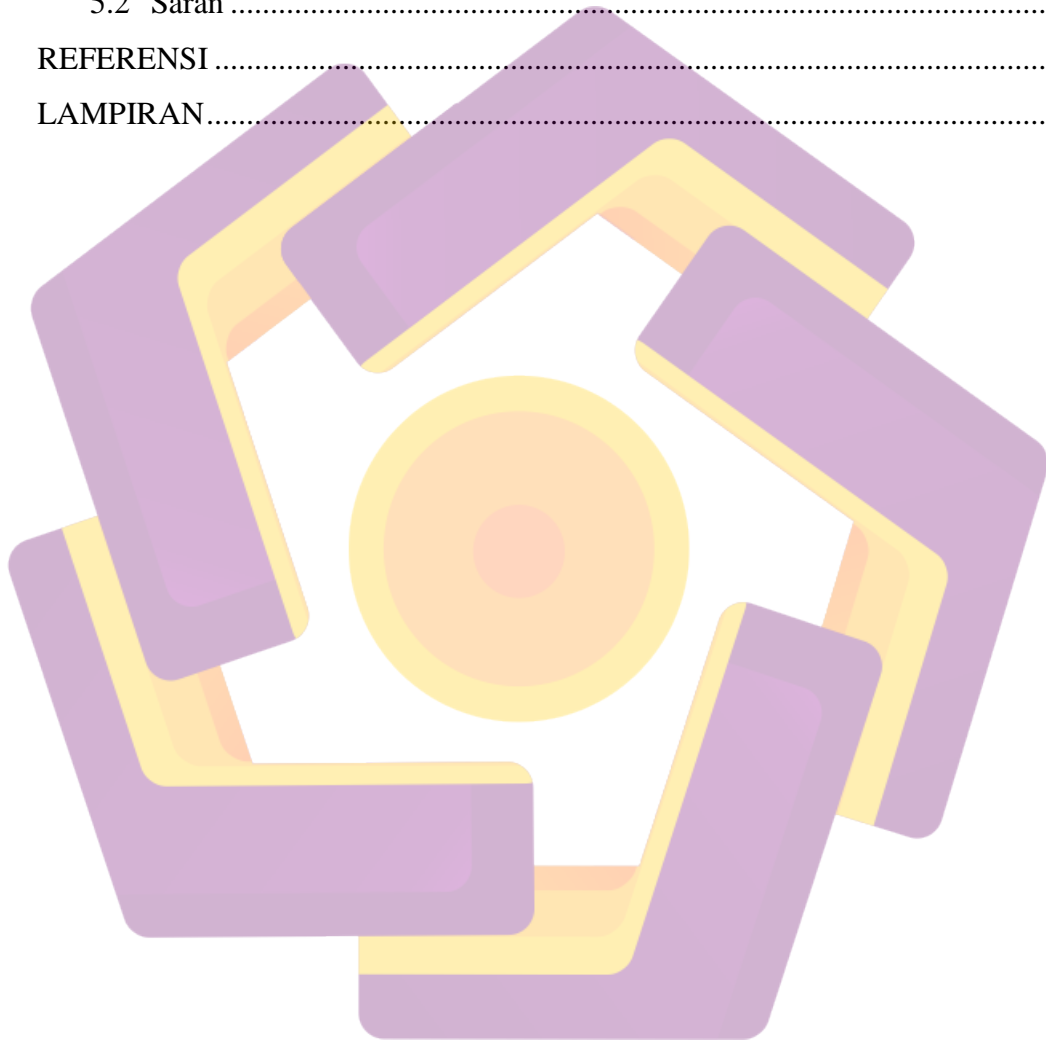
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	13
2.2.1. Prediksi	13
2.2.2. Regresi	14
2.2.3. Boosting	16
2.2.4. Exploratory data analysis (EDA).....	17
2.2.5. Preprocessing	18
2.2.6. Data cleaning	18

2.2.7.	Feature engineering atau transformasi data	19
2.2.8.	Data stasioner	20
2.2.9.	Hyperparameter tuning	20
2.2.10.	Cross validation.....	21
2.2.11.	Evaluasi	22
2.2.12.	Pengujian asumsi.....	23
2.2.13.	Pengujian data dan metode.....	23
2.2.14.	Pengujian replikasi langsung.....	24
2.2.15.	Penyajian hasil evaluasi	24
BAB III METODE PENELITIAN		25
3.1	Alur Penelitian	25
3.1.1.	Prosedur penelitian.....	25
3.1.2.	Parameter penelitian.....	26
3.1.3.	Metode analisis	27
3.1.4.	Penjabaran model.....	28
3.1.4.1.	Pengumpulan data	30
3.1.4.2.	Exploratory data analysis	32
3.1.4.3.	Preprocessing	32
3.1.4.4.	Data splitting	34
3.1.4.5.	Hyperparameter tuning.....	35
3.1.4.6.	Evaluasi	35
3.2.	Alat dan Bahan Penelitian.....	36
3.2.1.	Data penelitian	36
3.2.2.	Data penelitian	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1.	Dataset	38
4.2.	Preprocessing	39
4.2.1.	Null value.....	40
4.2.2.	Data stasioner	40
4.2.3.	Outlier handling	45
4.3.	Data Splitting	51

4.4. Hyperparameter tuning	53
4.5. Eksperimen	57
4.6. Evaluasi.....	60
BAB V PENUTUP	64
5.1 Kesimpulan	64
5.2 Saran	64
REFERENSI	66
LAMPIRAN.....	70



DAFTAR TABEL

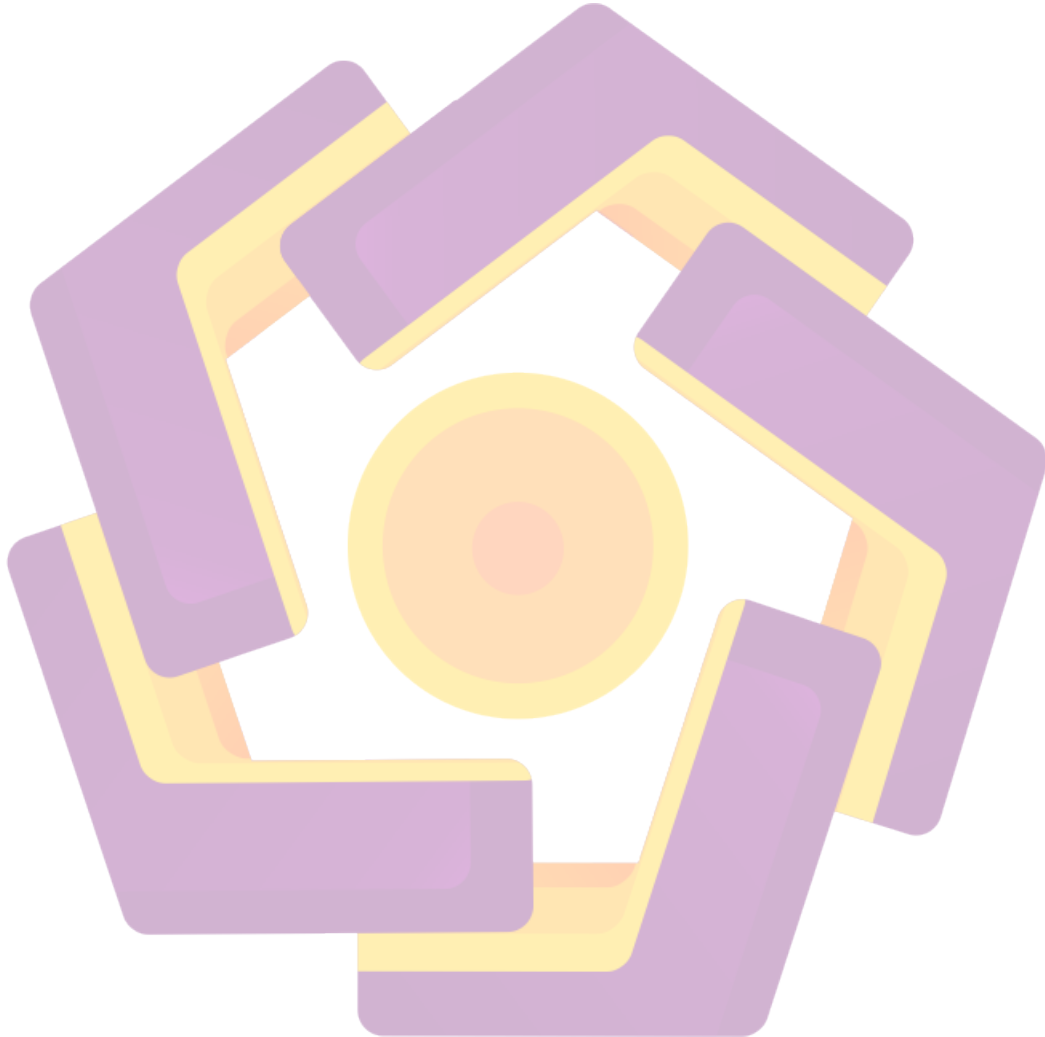
Tabel 2.1 Keaslian penelitian.....	10
Tabel 3.1 Parameter peramalan.....	26
Tabel 3.2 Contoh data deret waktu	31
Tabel 4.1 Dataset	38
Tabel 4.2 Perbandingan transformasi data.....	40
Tabel 4.3 Perbandingan Transformasi Data.....	44
Tabel 4.4 Tabel Hipotesis KPSS dan ADF.....	44
Tabel 4.5 Tabel critical value KPSS dan ADF	44
Tabel 4.6 Perbandingan antara Tukey's Method dengan ThymeBoost.....	48
Tabel 4.7 Perbandingan kurtosis dan skewness	49
Tabel 4.8 Tabel Pembagian Data	53
Tabel 4.9 Grid search cross validation.....	54
Tabel 4.10 Perbandingan metode deteksi outlier	58
Tabel 4.11 Perbandingan algoritma default LGBM	59
Tabel 4.12 Perbandingan model peramalan alternatif	60
Tabel 4.13 Peramalan dengan rentang waktu berbeda.....	61

DAFTAR GAMBAR

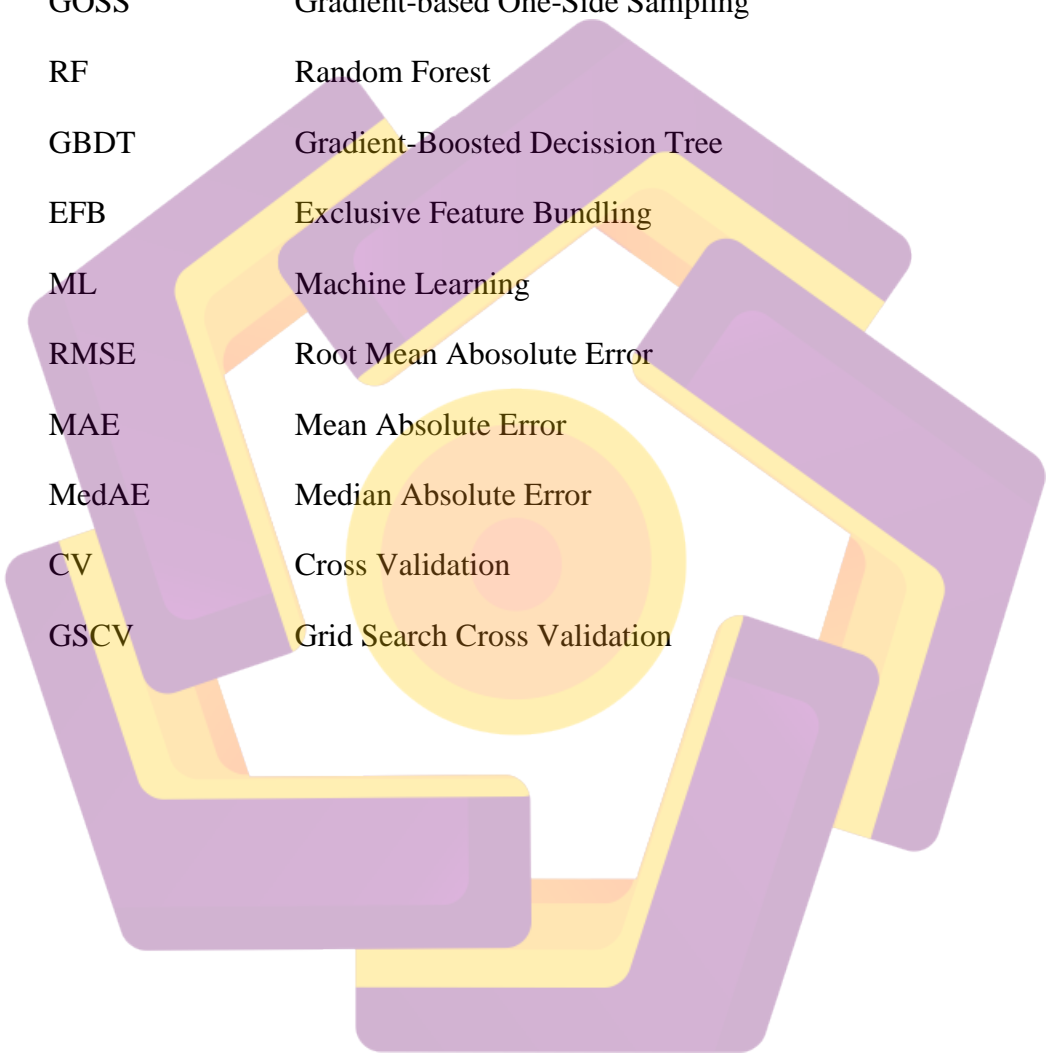
Gambar 2.1 Grafik tren dan seasonalitas	14
Gambar 2.2 Grafik regresi linear	15
Gambar 2.3 Cross validation.....	21
Gambar 3.1 Pseudocode algoritma GOSS (kiri) dan algoritma EFB (kanan)	30
Gambar 3.2 Bagan alur mode peramalan.....	30
Gambar 3.3 Alur data pulling yfinance.....	31
Gambar 3.4 Contoh plot visualisasi data (distplot).....	33
Gambar 3.5 Grid Search Cross Validation	35
Gambar 4.1 Data non-stasioner (atas) dan data stasioner (bawah)	42
Gambar 4.2 <i>Ilustrasi boxplot</i>	46
Gambar 4.3 Boxplot data transformasi (atas) dan data outlier dihapus (bawah)...	50
Gambar 4.4 Densitas persebaran data sesudah transformasi sebelum penanganan outlier (atas), dan sesudah penanganan outlier (bawah).....	51
Gambar 4.5 Perbandingan skor GSCV	55
Gambar 4.6 Perbandingan metode deteksi outlier	59
Gambar 4.7 Perbandingan algoritma default LGBM.....	60
Gambar 4.8 Perbandingan model peramalan alternatif.....	61
Gambar 4.9 Pengujian berdasarkan rentang waktu.....	63

DAFTAR LAMPIRAN

Lampiran 1. Ranking <i>hyperparameter</i> GSCV	70
--	----

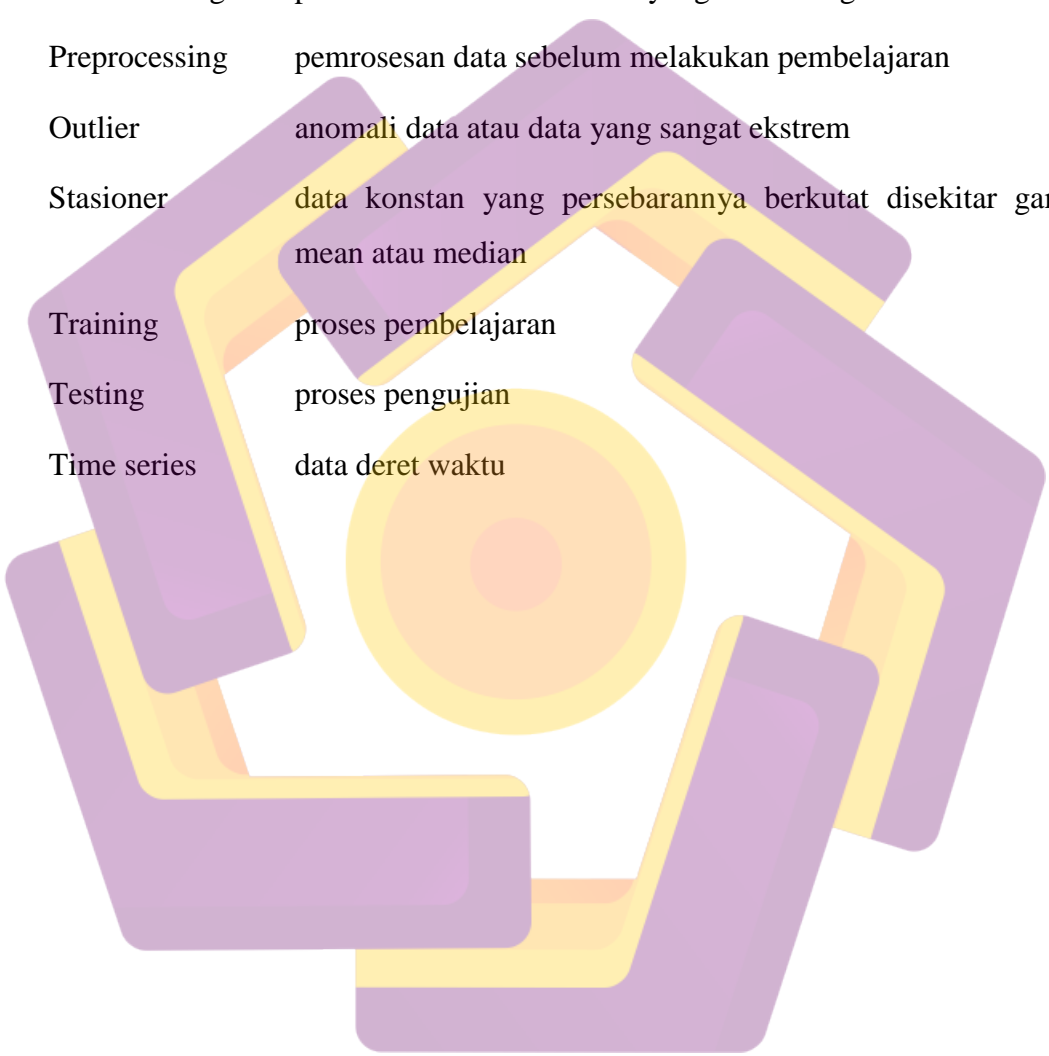


DAFTAR SINGKATAN



LGBM	Light Gradient Boosting Machine
DART	Dropouts meet Multiple Additive Regression Trees
GOSS	Gradient-based One-Side Sampling
RF	Random Forest
GBDT	Gradient-Boosted Decision Tree
EFB	Exclusive Feature Bundling
ML	Machine Learning
RMSE	Root Mean Absolute Error
MAE	Mean Absolute Error
MedAE	Median Absolute Error
CV	Cross Validation
GSCV	Grid Search Cross Validation

DAFTAR ISTILAH



Hyperparameter	konfigurasi proses training yang berada di luar model dan nilainya tidak dapat diperkirakan dari data
Data cleaning	pembersihan data dari fitur yang tidak diinginkan
Preprocessing	pemrosesan data sebelum melakukan pembelajaran
Outlier	anomali data atau data yang sangat ekstrem
Stasioner	data konstan yang persebarannya berkutat disekitar garis mean atau median
Training	proses pembelajaran
Testing	proses pengujian
Time series	data deret waktu

INTISARI

Pada akhir tahun 2019 dunia digoncangkan dengan menyebarnya COVID-19. Hal ini menyebabkan fluktuasi harga saham sepanjang tahun 2020 sehingga banyak orang takut dalam berinvestasi saham karena faktor psikologis. Investasi adalah salah satu cara untuk bertahan dari peristiwa seperti COVID-19.

Untuk itu peneliti membangun peramalan harga saham menggunakan data Yahoo Finance. Light Gradient Boosting Machine (LightGBM) adalah open source framework gradient boosting untuk machine learning. LightGBM memiliki tingkat akurasi yang tinggi dan efisien. Hal ini terbukti dalam kompetisi M5 Forecasting, LightGBM mampu meraih peringkat pertama melawan model peramalan deret waktu lainnya. Dalam penelitian ini, peneliti menggunakan model ansembel LightGBM untuk melakukan peramalan deret waktu harga saham.

Peneliti menggunakan *Grid Search Cross Validation* (GSCV) untuk menentukan hyperparameter. Hasil dari peramalan data deret waktu harga saham menunjukkan bahwa LightGBM bisa bersaing bahkan mampu mengungguli model peramalan berbasis boosting, seperti XGBoost, AdaBoost, dan CatBoost.

Kata kunci: LightGBM, peramalan, saham, boosting, grid search

ABSTRACT

At the end of 2019 the world was shaken by the spread of COVID-19. This causes stock price fluctuations throughout 2020 so that many people are afraid to invest in stocks because of psychological factors. Investing is one way to survive events like COVID-19.

For this reason, researchers build stock price forecasts using Yahoo Finance data. Light Gradient Boosting Machine (LightGBM) is an open source gradient boosting framework for machine learning. LightGBM has a high level of accuracy and efficiency. This is evident in the M5 Forecasting competition, LightGBM is able to achieve first place against other time series forecasting models. In this study, researchers used the LightGBM ensemble model to forecast the time series of stock prices.

Author used Grid Search Cross Validation (GSCV) to determine the hyperparameters. The results of stock price time series forecasting show that LightGBM can compete and even outperform boosting-based forecasting models, such as XGBoost, AdaBoost, and CatBoost.

Keyword: *LightGBM, forecasting, stock, boosting, grid search*