p-ISSN : 2406-7733
e-ISSN : 2597-9922

OPTIMASI SELEKSI FITUR PREDIKSI SAHAM MENGGUNAKAN *QUANTUM INSPIRED METAHEURISTIC* DAN LIGHTGBM

Rizky Wahyudi[1], Fauzan Azima Lubis[2], Windy Aulia[3], Adidtya Perdana[4]
Ilmu Komputer, Matematika Dan Ilmu Pengetahuan Alam, Universitas Negeri Medan
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara
E-mail: *rizkyw23@mhs.unimed.ac.id[1], azimafauzan018@gmail.com[2], windy.4231250021@mhs.unimed.ac.id[3], adidtya@unimed.ac.id[4]

Abstrak - Penelitian ini bertujuan untuk meningkatkan akurasi prediksi return harian saham perbankan Indonesia melalui penerapan metode seleksi fitur berbasis metaheuristik QIMA (*Quantum-Inspired Multi-Objective Algorithm*) pada model LightGBM. Permasalahan utama yang diangkat meliputi tingginya multikolinearitas antar-fitur teknikal, rendahnya validitas prediksi model *baseline*, serta mahalnya biaya komputasi saat menggunakan fitur penuh. Penelitian ini menggunakan 38 indikator teknikal yang kemudian direduksi menggunakan QIMA untuk memperoleh subset fitur optimal yang memaksimalkan akurasi sekaligus meminimalkan kompleksitas model. Hasil eksperimen menunjukkan bahwa model baseline mengalami *overfitting* dengan nilai R^2 negatif pada sebagian besar saham. Penerapan QIMA berhasil memperbaiki kinerja model menjadi positif pada BBCA (0,0006) dan BBNI (0,0173), serta menstabilkan performa BBRI dan BMRI. Selain itu, jumlah fitur berhasil direduksi sebesar 68,42% dengan waktu pelatihan yang lebih cepat hingga 6–7 kali lipat. Analisis fitur terpilih mengungkap bahwa volatilitas jangka pendek, return berbasis momentum, serta lag harga dan volume merupakan indikator yang paling relevan dalam memengaruhi pergerakan harga saham jangka pendek. Secara keseluruhan, studi ini membuktikan bahwa QIMA mampu meningkatkan validitas prediksi, efisiensi komputasi, dan interpretabilitas model dalam konteks pasar saham Indonesia.

Kata Kunci: LightGBM, Metaheuristik, QIMA, *Return* Saham, Seleksi Fitur

I. PENDAHULUAN

Perkembangan teknologi komputasi dan meningkatnya ketersediaan data finansial telah membuka peluang besar bagi penelitian prediksi pasar saham dengan pendekatan machine learning. Pasar saham dikenal bersifat non-linear, dinamis, dan dipengaruhi oleh banyak variabel seperti harga, volume, volatilitas, serta faktor eksternal lainnya sehingga model linier tradisional seperti ARIMA dan GARCH seringkali gagal menangkap kompleksitas tersebut. Literatur *review* terbaru menunjukkan bahwa model machine learning semakin diperlukan untuk menggantikan atau melengkapi metode tradisional agar bisa menangkap pola yang lebih dalam dan interaksi antar variabel yang lebih kompleks (Saberironaghi et al., 2025).

Di Indonesia, sektor perbankan merupakan penggerak utama Indeks Harga Saham Gabungan (IHSG) dengan kapitalisasi pasar terbesar. Saham-saham berkapitalisasi besar (*Big Caps*) seperti BBCA, BBRI, BMRI, dan BBNI sering dianggap sebagai proxy kondisi ekonomi nasional dan memiliki likuiditas tinggi. Namun, tingginya likuiditas ini juga membuat pergerakan harganya sangat sensitif terhadap sentimen global dan arus dana asing, menciptakan pola volatilitas harian yang sulit diprediksi (Tabash et al., 2024). Fenomena ini menuntut adanya model prediksi yang tidak hanya mengandalkan data historis linier, tetapi juga mampu memilah sinyal teknikal yang valid di tengah fluktuasi pasar yang *noisy*. Oleh karena itu, studi kasus pada empat emiten ini sangat relevan untuk menguji ketangguhan algoritma seleksi fitur.

Dalam beberapa tahun terakhir, metode ensemble learning berbasis pohon keputusan seperti LightGBM telah menunjukkan performa yang sangat menjanjikan dalam konteks data finansial, termasuk prediksi harga saham maupun volatilitas karena kemampuannya menangani data berdimensi tinggi, interaksi non-linear, dan pelatihan yang relatif efisien. Sebagai contoh, penelitian "*Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model*" menunjukkan bahwa LightGBM mampu bersaing dan bahkan mengungguli model seperti XGBoost dan CatBoost untuk data pasar saham di Indonesia (Hartanto et al., 2023). Selain itu, sebuah studi "*Stock Volatility Prediction Based on 1D-CNN and LightGBM*" berhasil menurunkan *error* dibanding LightGBM saja, dengan memanfaatkan fitur-teknikal dan arsitektur hibrid (Dan et al., 2022). Namun demikian, keberhasilan LightGBM sangat bergantung pada pemilihan fitur (*feature engineering/selection*) serta penyetelan parameter (*hyperparameter tuning*) yang tepat yang mana kedua aspek ini yang sering menjadi tantangan dalam praktik.

Walaupun optimasi fitur dan hyperparameter telah banyak dilakukan menggunakan metode metaheuristik seperti *Genetic Algorithm* (GA) dan *Particle Swarm Optimization* (PSO), masih terdapat gap penelitian terkait penerapan algoritma yang lebih canggih seperti pendekatan kuantum-inspired

p-ISSN : 2406-7733
e-ISSN : 2597-9922

(misalnya *Quantum-Inspired Metaheuristic Algorithm*, QIMA) khususnya dalam domain keuangan. Misalnya, studi "*Stock Complex Networks based on the GA-LightGBM model*" menunjukkan bahwa kombinasi GA dengan LightGBM memberi hasil state-of-the-art dalam optimasi hyperparameter (Huang et al., 2025). Karena studi kuantum-inspired pada prediksi saham masih sangat terbatas, penelitian ini bertujuan untuk mengembangkan model prediksi pasar saham berbasis LightGBM yang dioptimasi menggunakan algoritma metaheuristik kuantum-inspired (QIMA) untuk meningkatkan akurasi dan efisiensi proses seleksi fitur dan parameter, sehingga mengisi gap penelitian yang ada.

II. TINJAUAN PUSTAKA

Studi mengenai prediksi pasar saham telah bergeser dari model statistik linier menuju pendekatan machine learning yang mampu menangkap volatilitas non-linear dan ketidakpastian pasar akibat guncangan eksternal (Kayani et al., 2024; Tabash et al., 2024). Meskipun model tradisional memiliki dasar teori yang kuat, tinjauan literatur terbaru menegaskan bahwa teknik Deep Learning dan Machine Learning semakin diperlukan untuk menangkap pola kompleks yang gagal dideteksi oleh metode konvensional (Saberironaghi et al., 2025). Pendekatan hibrida yang menggabungkan indikator teknikal dengan algoritma cerdas terbukti meningkatkan akurasi prediksi jangka pendek secara signifikan (Liang, Xinyue, 2024).

Dalam ranah machine learning, metode ensemble learning seperti Gradient Boosting telah mendominasi literatur terkini karena ketahanannya terhadap noise. Light Gradient Boosting Machine (LightGBM) yang diperkenalkan oleh Ke et al. (2017) terbukti unggul dalam efisiensi komputasi dan skalabilitas dibandingkan pendahulunya, XGBoost. Studi empiris oleh Hartanto et al. (2023) dan Dan et al. (2022) mengonfirmasi bahwa LightGBM menawarkan performa prediktif yang superior pada data saham bervolume tinggi.

Keunggulan performa LightGBM tersebut tidak terlepas dari dua mekanisme novel yang diperkenalkannya, yaitu Gradient-based One-Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB). GOSS bekerja dengan asumsi bahwa sampel data dengan gradien kesalahan yang besar membawa informasi yang lebih krusial untuk proses pembelajaran. Oleh karena itu, mekanisme ini mempertahankan seluruh instans dengan gradien besar dan melakukan down-sampling secara acak pada instans dengan gradien kecil, sehingga model dapat fokus mempelajari pola data yang sulit diprediksi tanpa mengubah distribusi data asli secara signifikan. Di sisi lain, EFB menangani masalah sparsity pada data berdimensi tinggi dengan mengelompokkan fitur-fitur yang saling eksklusif (jarang bernilai bukan nol secara bersamaan) ke dalam satu bundle fitur tunggal. Hal ini mereduksi kompleksitas pemindaian fitur secara drastis, yang menjadikan LightGBM sangat efisien untuk data pasar saham yang memiliki *noise* tinggi dan dimensi fitur yang besar (Ke et al., 2017). Namun, efektivitas model ini sangat bergantung pada kualitas fitur input. Penggunaan fitur teknikal yang berlebihan tanpa seleksi sering memicu *overfitting* dan menurunkan kemampuan generalisasi model pada data out-of-sample (Khairunisa & Hendikawati, 2024; Majid et al., 2024).

Untuk mengatasi tantangan seleksi fitur, algoritma metaheuristik telah banyak diterapkan. Penelitian Fitria dan Widayanti (2024) menunjukkan bahwa integrasi algoritma optimasi dalam pemodelan prediktif secara konsisten menghasilkan error yang lebih rendah dibanding model tunggal. Meski demikian, metaheuristik klasik seperti *Genetic Algorithm* (GA) dan *Particle Swarm Optimization* (PSO) sering kali terjebak pada optimum lokal dan sangat sensitif terhadap inisialisasi parameter (Buulolo & Wau, 2024; Ferrouhi & Bouabdallaoui, 2024). Abdulloh et al. (2024) dan Xiong et al. (2024) juga mencatat bahwa meskipun metode hibrida (seperti GA-LightGBM) memberikan hasil yang baik, biaya komputasi yang dibutuhkan masih relatif tinggi.

Sebagai solusi atas keterbatasan tersebut, algoritma *Quantum-Inspired Metaheuristic* (QiM) muncul sebagai pendekatan *state-of-the-art*. QiM mengadopsi prinsip superposisi dan rotasi Q-bit untuk meningkatkan diversitas populasi dan mempercepat konvergensi (Dey et al., 2023; Sandeep & Sood, 2024). Studi terbaru menunjukkan bahwa varian algoritma berbasis kuantum mampu mencapai solusi global jauh lebih cepat daripada metaheuristik konvensional pada masalah optimasi kompleks (Husam, 2025; Khan & Hussain, 2024). Mengingat penerapan optimasi berbasis kuantum pada LightGBM (seperti pada studi Huang et al. (2025)) masih terbatas, penelitian ini mengisi celah tersebut dengan mengusulkan integrasi QIMA untuk seleksi fitur yang lebih efisien dan akurat.

III. METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan tujuan untuk menganalisis dan membandingkan performa model *baseline Light Gradient Boosting Machine* (LightGBM) dengan model usulan *Quantum-Inspired Metaheuristic* LightGBM (QIMA-LightGBM) dalam proses seleksi fitur pada prediksi harga saham. Pendekatan kuantitatif dipilih karena metode ini memungkinkan pengukuran performa

p-ISSN : 2406-7733
e-ISSN : 2597-9922

model secara objektif melalui indikator statistik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2).

Model baseline (LightGBM) dijalankan dengan parameter hasil GridSearchCV, sedangkan model usulan (QIMA-LightGBM) menggunakan algoritma *Quantum-Inspired Metaheuristic Algorithm* (QIMA) untuk mencari kombinasi fitur optimal sebelum pelatihan model. Seluruh tahapan eksperimen dilakukan menggunakan Python 3.10 di Google Colab dengan library utama: *pandas, numpy, scikit-learn, dan lightgbm.*

Sumber Data

Data yang digunakan dalam penelitian ini berasal dari dataset sekunder berjudul "Dataset Saham Bank Indonesia 2014–2024" yang diperoleh dari platform Kaggle. Dataset ini memuat data transaksi harian saham dari empat bank besar di Indonesia, yaitu Bank Rakyat Indonesia (BRI), Bank Negara Indonesia (BNI), Bank Central Asia (BCA), dan Bank Mandiri, selama periode 2014 hingga 2024. Tabel 1 berikut menunjukkan deskripsi fitur dan tipe data yang digunakan dalam penelitian:

Tabel 1. Dataset Saham Bank Indonesia 2014–2024

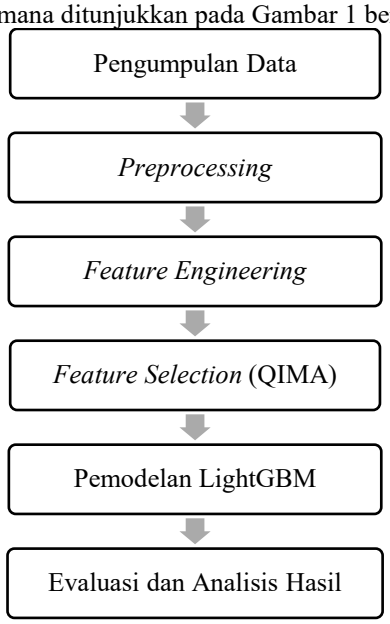
Nama Fitur	Deskripsi	Tipe Data
Date	Tanggal transaksi harian	datetime
Adj Close	Harga penutupan yang telah disesuaikan terhadap dividen atau aksi korporasi lainnya (*Adjusted Close*)	float64
Close	Harga penutupan saham pada akhir hari perdagangan	float64
High	Harga tertinggi saham dalam satu hari perdagangan	float64
Low	Harga terendah saham dalam satu hari perdagangan	float64
Open	Harga pembukaan saham pada awal hari perdagangan	float64
Volume	Jumlah transaksi saham yang terjadi dalam satu hari	int64

Data ini dipilih karena memiliki periode waktu yang panjang (10 tahun), frekuensi pencatatan yang konsisten (harian), serta mencakup beberapa emiten perbankan besar yang merepresentasikan pergerakan pasar saham di sektor keuangan Indonesia secara komprehensif.

Selain itu, dataset ini cocok digunakan untuk pengujian algoritma optimasi dan prediksi harga saham, seperti *Quantum-Inspired Metaheuristic* (QIMA), karena karakteristik datanya yang kompleks, fluktuatif, dan bersifat non-linear. Seluruh data diunduh dalam format CSV, kemudian diolah menggunakan Python (Google Colab) melalui tahapan pembersihan data (*data cleaning*), rekayasa fitur (*feature engineering*), dan pembagian data latih–uji (*train–test split*).

Tahapan Penelitian

Penelitian ini terdiri dari enam tahap utama sebagaimana ditunjukkan pada Gambar 1 berikut



Gambar 1. Diagram Alir Penelitian

a) Pengumpulan dan Pra-pemrosesan Data

Penelitian ini menggunakan data sekunder berupa riwayat transaksi harian saham perbankan (BBCA, BBNI, BBRI, BMRI) periode 2014–2024. Setelah tahap pengumpulan, data mentah melalui serangkaian proses pra-pemrosesan untuk menjamin kualitas input model. Tahap ini diawali dengan konversi tipe data harga dan volume menjadi format numerik serta pengurutan data secara kronologis berdasarkan waktu. Penanganan *missing values* dilakukan melalui metode imputasi median, diikuti dengan penghapusan baris duplikat untuk menjaga konsistensi. Terakhir, seluruh fitur numerik dinormalisasi menggunakan teknik *MinMax Scaler* ke rentang [0, 1] guna mencegah bias komputasi akibat perbedaan skala antar-variabel, khususnya pada fitur volume yang memiliki rentang nilai jauh lebih besar dibandingkan harga.

b) Rekayasa dan Seleksi Fitur Berbasis QIMA

Guna menangkap pola non-linear dan dinamika pasar, penelitian ini memperluas ruang fitur dari data harga mentah menjadi indikator teknikal yang lebih informatif. Fitur utama yang dikonstruksi adalah *Log Return* yang bertujuan menstandarisasi data harga dan menangkap autokorelasi. Perhitungan *Log Return* (R_t) dilakukan menggunakan Persamaan (1):

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1)$$

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Dimana P_t adalah harga penutupan pada waktu t dan P_{t-1} adalah harga pada periode sebelumnya.
Mengingat banyaknya fitur yang dihasilkan berpotensi menyebabkan *curse of dimensionality*, algoritma *Quantum-Inspired Metaheuristic* (QIMA) diterapkan untuk menyeleksi subset fitur yang paling optimal. Dalam mekanisme ini, setiap fitur direpresentasikan sebagai Q-bit $[\alpha, \beta]$ yang menunjukkan probabilitas fitur tersebut untuk dipilih. Algoritma berevolusi mencari kombinasi fitur yang meminimalkan fungsi *fitness* (objektif). Fungsi ini dirancang untuk menyeimbangkan akurasi model dan efisiensi dimensi, sebagaimana dirumuskan Persamaan (2) berikut:

$$Fitness = w_1 \cdot (1 - R^2_{val}) + w_2 \cdot (\frac{N_{selected}}{N_{total}}) \qquad (2)$$

Dimana w_1 dan w_2 merupakan bobot penyeimbang, R^2_{val} adalah koefisien determinasi pada data validasi, sedangkan $N_{selected}$ dan N_{total} merepresentasikan rasio fitur yang digunakan. Evaluasi internal dalam proses seleksi ini menggunakan *regressor* LightGBM ringan untuk mempercepat konvergensi pencarian solusi tanpa mengorbankan kualitas seleksi secara signifikan.

```
Input: Dataset Saham (Fitur
F1...F38), Populasi N, Iterasi Max
T
Output: Subset Fitur Optimal
(S_best)

1: Inisialisasi Populasi Q-bit Q =
{q1, q2, ..., qN}
2: Konversi Q-bit ke solusi biner B
(observasi)
3: Hitung Fitness awal untuk setiap
individu:
4:    Latih LightGBM dengan fitur
terpilih B[i]
5:    Hitung R2_val pada data
validasi
6:    Fitness[i] = w1 * (1 -
R2_val) + w2 * (Rasio_Fitur)
7: Tentukan solusi terbaik global
(Best_Sol)
8: While t < T do:
9:    For setiap individu i dalam
Populasi:
10:      Update Q-bit dengan
Gerbang Rotasi Kuantum:
11:      [α', β'] = Rotasi([α, β],
Best_Sol)
12:      Konversi Q-bit baru ke
biner B_new
13:      Evaluasi Fitness(B_new)
14:      If Fitness(B_new) <
Fitness(Best_Sol) then
15:         Update Best_Sol = B_new
16:      End If
17:   End For
18:   t = t + 1
19: End While
20: Return Best_Sol
```

Algoritma 1. Prosedur Seleksi Fitur Menggunakan *Quantum-Inspired Metaheuristic Algorithm* (QIMA)

Pseudocode di atas menggambarkan mekanisme evolusi QIMA dalam menemukan kombinasi fitur terbaik. Proses dimulai dengan Inisialisasi (Baris 1), di mana setiap fitur direpresentasikan sebagai Q-bit yang berada dalam posisi superposisi $[\alpha, \beta]$. Artinya, pada tahap awal, setiap fitur memiliki peluang yang sama untuk terpilih atau tidak.

Selanjutnya, pada tahap Observasi (Baris 2), status kuantum tersebut "diukur" dan dikonversi menjadi biner (0 atau 1). Nilai '1' berarti fitur tersebut dipilih untuk dimasukkan ke dalam model prediksi, sedangkan '0' berarti fitur dibuang. Kombinasi fitur yang terpilih kemudian dievaluasi kualitasnya menggunakan fungsi *Fitness* (Baris 3-6), yang menghitung seberapa akurat model LightGBM saat menggunakan fitur-fitur tersebut, sekaligus memberi penalti jika jumlah fitur terlalu banyak.

Inti dari kecerdasan algoritma ini terletak pada Gerbang Rotasi Kuantum (Baris 10-11). Jika ditemukan solusi baru yang lebih baik dari solusi terbaik saat ini (*Best_Sol*), maka Q-bit individu lain akan "dirotasi" mendekati posisi solusi terbaik tersebut. Rotasi ini memperbarui probabilitas setiap fitur secara bertahap: fitur yang terbukti bagus akan memiliki peluang lebih besar untuk terpilih di iterasi berikutnya, sementara fitur yang buruk akan perlahan dieliminasi. Proses ini berulang hingga mencapai iterasi maksimum, menghasilkan subset fitur yang paling optimal untuk prediksi.

c) Pemodelan Prediksi dan Evaluasi Kinerja

Model prediksi akhir dibangun menggunakan algoritma *Light Gradient Boosting Machine* (LightGBM) dengan fungsi objektif regresi. Mengingat karakteristik data time series yang memiliki ketergantungan temporal, skema validasi menggunakan pembagian kronologis *hold-out* dimana 80% data awal dialokasikan sebagai data latih dan 20% sisanya sebagai data uji, tanpa proses pengacakan (*shuffling*). Konfigurasi model difokuskan pada generalisasi dengan menerapkan regularisasi L1 dan L2, serta mekanisme *early stopping* yang akan menghentikan pelatihan jika tidak terjadi penurunan *error* dalam 50 iterasi berturut-turut.

Sebelum pelatihan final, dilakukan optimasi hyperparameter menggunakan *Grid Search Cross-Validation* (GridSearchCV) dengan skema validasi silang runtun waktu (*Time Series Split*) sebanyak 3 lipatan. Hal ini bertujuan untuk mencari konfigurasi parameter terbaik yang menyeimbangkan antara bias dan *variance*. Ruang pencarian parameter yang diuji meliputi kompleksitas pohon (*max depth, num leaves*), laju pembelajaran (*learning rate*), serta parameter regularisasi untuk mencegah *overfitting*.

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Rincian ruang parameter yang dieksplorasi disajikan pada Tabel 2.

Tabel 2. Ruang Pencarian Hyperparameter LightGBM

Parameter	Deskripsi	Ruang Pencarian (*Grid*)
n_estimators	Jumlah pohon keputusan yang dibangun	[300, 500]
max_depth	Kedalaman maksimum setiap pohon	[3, 5, 7]
learning_rate	Bobot kontribusi setiap pohon (*shrinkage*)	[0.01, 0.05]
num_leaves	Jumlah daun maksimum dalam satu pohon	[15, 31]
reg_alpha (L1)	Regularisasi L1 pada bobot (*lasso*)	[0.1, 0.5]
reg_lambda (L2)	Regularisasi L2 pada bobot (*ridge*)	[0.1, 0.5]
subsample	Rasio sampel data untuk setiap pohon	0.8

Pemilihan parameter regularisasi L1 (*reg_alpha*) dan L2 (*reg_lambda*) menjadi krusial dalam penelitian ini mengingat banyaknya fitur yang berkorelasi tinggi pada data saham. Regularisasi L1 berfungsi sebagai selektor fitur implisit dengan mendorong bobot fitur yang kurang relevan menjadi nol, sedangkan L2 membatasi besaran bobot agar model tetap stabil terhadap perubahan data input yang kecil.

Kinerja model kemudian dievaluasi secara komprehensif menggunakan kombinasi metrik statistik dan analisis visual. Validitas prediksi diukur menggunakan *Mean Absolute Error* (MAE) untuk mengetahui rata-rata penyimpangan absolut sesuai Persamaan (3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

di mana n = Jumlah data observasi. y_i= Nilai aktual harga saham pada periode ke-i. Dan $\hat{y}_i$= Nilai prediksi harga saham pada periode ke-i. (Hartanto et al., 2023).

Selanjutnya yaitu *Root Mean Squared Error* (RMSE) untuk mendeteksi sensitivitas model terhadap kesalahan ekstrem (*outliers*) (Khairunisa & Hendikawati, 2024). Persamaan RMSE adalah:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

Selain itu, Koefisien Determinasi R^2 digunakan untuk mengukur seberapa baik variabel independen (fitur hasil seleksi QIMA) mampu menjelaskan variabilitas variabel dependen (harga saham). Nilai R^2 berkisar antara ∞ hingga 1, di mana nilai yang mendekati 1 menunjukkan kecocokan model yang sempurna, sedangkan nilai negatif menunjukkan performa model lebih buruk daripada sekadar memprediksi nilai rata-rata data (Abdulloh et al., 2024). Persamaan R^2 yaitu :

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

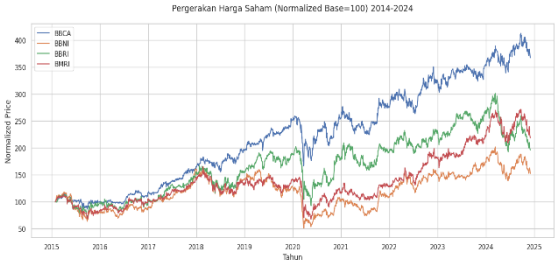
di mana $\bar{y}$ adalah nilai rata-rata dari data aktual. Untuk memastikan keandalan model, analisis distribusi residual dilakukan guna memeriksa kenormalan error, dilengkapi dengan analisis *feature importance* untuk mengidentifikasi variabel dominan yang memengaruhi keputusan prediksi model.

IV. HASIL DAN PEMBAHASAN

Deskripsi Data dan Eksplorasi Awal

Penelitian ini menggunakan data historis harian dari empat emiten perbankan dengan kapitalisasi pasar terbesar di Bursa Efek Indonesia, yaitu PT Bank Central Asia Tbk (BBCA), PT Bank Negara Indonesia Tbk (BBNI), PT Bank Rakyat Indonesia Tbk (BBRI), dan PT Bank Mandiri Tbk (BMRI). Periode pengamatan mencakup data selama 10 tahun, mulai dari tahun 2014 hingga 2024.

Sebelum dilakukan pemodelan, data mentah yang terdiri dari fitur *Open, High, Low, Close*, dan *Volume* telah melalui tahap pra-pemrosesan yang meliputi pembersihan nilai kosong (*missing values*) dan normalisasi menggunakan *MinMax Scaler*. Untuk menjaga integritas eksperimen dan mencegah kebocoran data (*data leakage*), proses normalisasi dilakukan dengan menyesuaikan parameter scaler hanya pada data latih, yang kemudian diterapkan pada data uji. Visualisasi pergerakan harga penutupan (*Close Price*) untuk keempat saham selama periode observasi dapat dilihat pada Gambar 2.

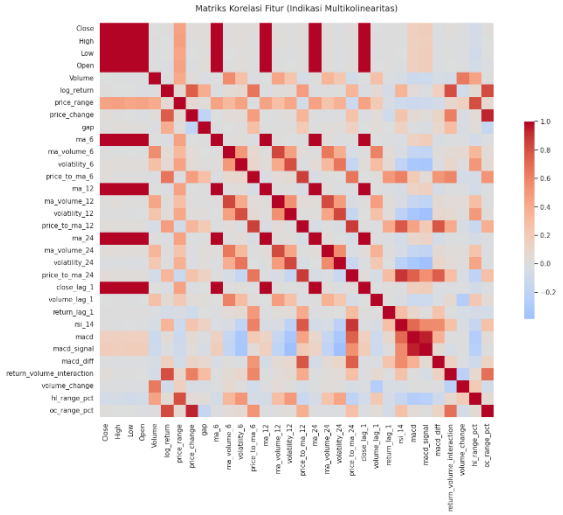


Gambar 2. Pergeseran Harga Penutup Saham BBCA, BBni, BBRI, dan BMRI (2014-2024)

Selanjutnya, analisis korelasi awal dilakukan untuk mengidentifikasi hubungan antar-variabel. Sebagaimana terlihat pada heatmap korelasi (Gambar 3), terlihat adanya kelompok fitur dengan korelasi tinggi, terutama pada indikator teknikal

p-ISSN : 2406-7733
e-ISSN : 2597-9922

yang berasal dari periode perhitungan serupa, seperti *Moving Average*, volatilitas, dan *price-to-MA*. Selain itu, komponen harga (*Open, High, Low, Close*) juga menunjukkan korelasi kuat satu sama lain. Pola ini mengindikasikan adanya multikolinearitas yang cukup signifikan, sehingga proses seleksi fitur menjadi langkah penting untuk mengurangi redundansi dan menjaga kinerja model tetap optimal.

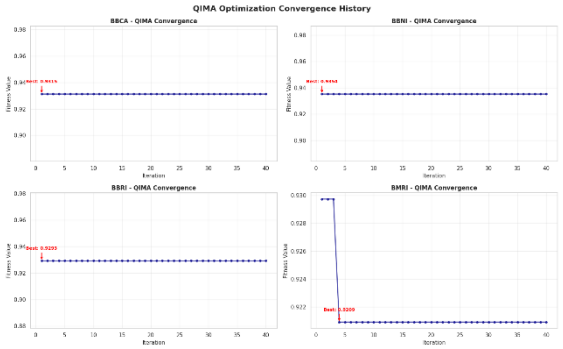


Gambar 3. Matriks Korelasi Fitur Awal Sebelum Seleksi (Menunjukkan Multikolineritas)

Hasil Optimasi Seleksi Fitur dengan Qima

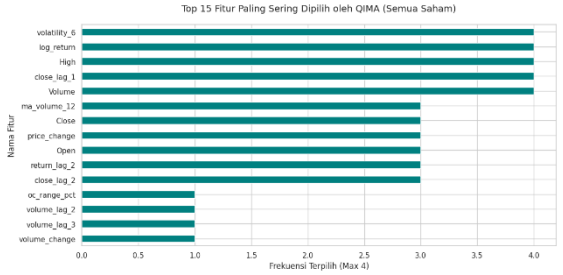
Tahap rekayasa fitur (*feature engineering*) awal menghasilkan total 38 fitur. Algoritma *Quantum-Inspired Metaheuristic* (QIMA) kemudian diterapkan untuk menyeleksi subset fitur yang paling optimal guna mereduksi dimensi data.

Berdasarkan hasil eksperimen, QIMA mampu mereduksi jumlah fitur secara signifikan, yaitu rata-rata sebesar 68,42%. Dari 38 fitur awal, algoritma hanya mempertahankan sekitar 11–15 fitur yang dinilai paling relevan. Proses optimasi ini terlihat pada grafik konvergensi di Gambar 4, di mana nilai fitness langsung mencapai titik terbaik pada iterasi awal dan kemudian tetap stabil. Pola ini menunjukkan bahwa QIMA dapat menemukan solusi optimal dengan cepat dan mempertahankannya secara konsisten sepanjang iterasi.



Gambar 4. Grafik Konvergensi Nilai *Fitness* QIMA Selama Proses Optimasi

Analisis lebih lanjut pada fitur yang terpilih (Gambar 5) menunjukkan bahwa algoritma cenderung mempertahankan variabel yang bersifat momentum jangka pendek dan indikator berbasis volatilitas, seperti *volatility_6, log_return*, dan beberapa fitur lag harga serta volume. Selain itu, komponen harga dasar (*Open, High, Close*) juga tetap muncul sebagai fitur penting. Pola ini menunjukkan bahwa QIMA lebih memprioritaskan fitur yang memberikan sinyal perubahan harga jangka pendek, sekaligus mengabaikan fitur-fitur yang dianggap redundan atau kurang informatif.



Gambar 5. Tingkat Kepentingan Fitur (*Feature Importance*) Hasil Seleksi Qima

Evaluasi Kinerja Model Prediksi

Evaluasi dilakukan dengan membandingkan performa dua model skenario: (1) Model *Baseline*, yaitu LightGBM yang menggunakan seluruh 38 fitur tanpa seleksi; dan (2) Model Usulan (QIMA-LightGBM), yaitu LightGBM yang dilatih hanya menggunakan fitur hasil seleksi QIMA. Ringkasan hasil pengujian pada data uji (*testing set*) disajikan dalam Tabel 3 berikut ini.

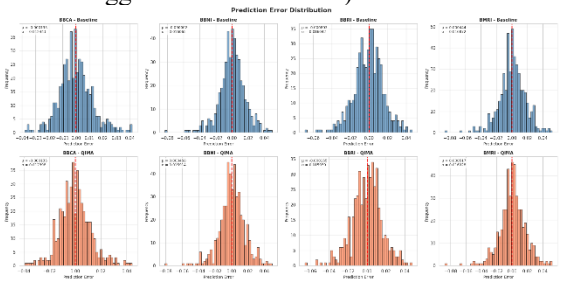
Tabel 3. Perbandingan Kinerja Model *Baseline* dan QIMA-LightGBM

Stock	Model	MAE	RMSE	R^2	Feat.	Time (s)
BBCA	Baseline	0.009382	0.012443	-0.006591	38	181.39
	QIMA	0.009347	0.012398	0.000634	11	28.28
BBNI	Baseline	0.011750	0.016065	-0.000650	38	233.35

p-ISSN : 2406-7733
e-ISSN : 2597-9922

	QIMA	0.011694	0.015920	0.017327	11	29.50
BBRI	Baseline	0.012266	0.016062	-0.023593	38	195.64
	QIMA	0.012284	0.015939	-0.008034	11	23.49
BMRI	Baseline	0.012466	0.016878	0.010759	38	217.71
	QIMA	0.012460	0.016936	0.003942	14	25.55

Analisis distribusi error pada Gambar 6 menunjukkan pola kurva lonceng yang simetris dengan puncak berada di sekitar nilai nol. Hal ini mengindikasikan bahwa residu prediksi terdistribusi secara normal dan model tidak memiliki bias sistematis (misalnya, cenderung selalu memprediksi terlalu tinggi atau terlalu rendah).



Gambar 6. Grafik Histogram Error

Ketiadaan *skewness* yang ekstrem membuktikan bahwa QIMA berhasil memilih fitur yang menghasilkan model regresi yang robust dan tidak bias.

Pembahasan

Berdasarkan hasil eksperimen yang tertera pada Tabel 3, terdapat tiga temuan utama yang menjadi fokus analisis, yaitu peningkatan validitas prediksi, efisiensi komputasi, dan efektivitas reduksi dimensi.

a) Hasil Akurasi dan Stabilitas Model

Eksperimen menunjukkan bahwa Model *Baseline* gagal melakukan generalisasi pada tiga dari empat saham, terlihat dari nilai R^2 negatif pada BBCA (-0.0066), BBNI (-0.00065), dan BBRI (-0.0236). Nilai negatif ini menunjukkan bahwa model dengan 38 fitur bukan hanya gagal menangkap pola prediksi, tetapi bahkan berkinerja lebih buruk daripada sekadar memprediksi nilai rata-rata. Fenomena tersebut mengindikasikan adanya *overfitting* yang kemungkinan dipicu oleh multikolinearitas tinggi antar-fitur, sebagaimana terlihat pada heatmap korelasi (Gambar 3), sehingga model lebih banyak menangkap *noise* daripada sinyal yang relevan.

Penerapan seleksi fitur menggunakan QIMA kemudian memperbaiki validitas prediksi. Pada BBCA dan BBNI, nilai R^2 yang semula negatif berubah menjadi positif (masing-masing 0.00063 dan 0.0173). Meski secara absolut kecil, performa tersebut tetap bermakna secara ekonomi mengingat prediksi return harian memiliki rasio *signal-to-noise* yang sangat rendah. Temuan ini sejalan dengan penjelasan Goto & Yamada (2025) bahwa model regresi untuk aset keuangan sering menghasilkan R^2 *out-of-sample* negatif akibat tantangan generalisasi. Dengan demikian, kemampuan QIMA-LightGBM mengubah R^2 negatif menjadi positif pada dua saham tersebut menunjukkan bahwa seleksi fitur berbasis kuantum berhasil memisahkan sinyal informasi yang relevan dari *noise* pasar, Hal ini didukung oleh penelitian Fitria dan Widayanti (2024) yang membuktikan bahwa penerapan algoritma optimasi metaheuristik dalam metode peramalan mampu menghasilkan tingkat akurasi yang lebih tinggi dan error yang lebih rendah dibandingkan pendekatan konvensional, terutama dalam situasi prediksi saham yang sangat sensitif.

Adapun pada BBRI dan BMRI, QIMA memang tidak meningkatkan nilai R^2, namun tetap mampu menurunkan *error* (MAE dan RMSE) serta menghasilkan model yang lebih stabil dibanding *Baseline*. Tidak adanya peningkatan signifikan pada dua saham tersebut mengindikasikan bahwa karakter pergerakan harganya lebih acak (*noisier*), sehingga reduksi fitur tidak selalu memberikan dampak langsung pada metrik deterministik seperti R^2.

Hasil ini menunjukkan bahwa model QIMA mampu menyaring fitur yang benar-benar relevan, mengurangi efek *overfitting* pada data *out-of-sample*, dan memberikan performa prediksi yang valid dalam konteks pasar keuangan yang sarat noise, sejalan dengan literatur modern yang menekankan pentingnya pemisahan sinyal dan noise dalam modelling keuangan.

b) Analisis Efesiensi Komputasi

Selain meningkatkan validitas prediksi, QIMA memberikan manfaat komputasi yang sangat signifikan. Berdasarkan hasil eksperimen, waktu pelatihan rata-rata menurun dari sekitar 205 detik pada model Baseline menjadi sekitar 26 detik pada model QIMA–LightGBM. Ini setara dengan percepatan komputasi ~7,8 kali lebih cepat atau penghematan waktu sekitar 87%.

Peningkatan efisiensi ini terjadi karena QIMA berhasil mereduksi fitur dari 38 menjadi hanya 11–14 fitur, atau setara dengan reduksi sekitar 68%. Grafik konvergensi (Gambar 4.3) memperlihatkan bahwa algoritma dapat menemukan solusi optimal dalam kurang dari 40 iterasi, yang menandakan proses eksplorasi ruang solusi berlangsung secara cepat dan stabil.

Efisiensi ini memiliki implikasi praktis yang besar, khususnya pada aplikasi yang memerlukan pembaruan model secara berkala atau sistem yang

p-ISSN : 2406-7733
e-ISSN : 2597-9922

menuntut respons cepat, misalnya *intraday analysis* atau *low-latency trading system*.

Tabel 4. Perbandingan Hasil Penelitian dengan Studi Terkait

Peneliti & Tahun	Metode	Objek Studi	Hasil Utama
Hartanto et al. (2023)	LightGBM (Semua Fitur)	Saham LQ45	MAE rendah, tetapi komputasi berat pada dataset besar.
Huang et al. (2025)	GA-LightGBM	Pasar China	GA efektif untuk optimasi parameter, namun konvergensi relatif lambat.
Penelitian Ini	QIMA-LightGBM	Bank Indonesia	Akurasi stabil ($R^2 > 0$), reduksi fitur 68%, dan komputasi ±7× lebih cepat.

Tabel 4 menunjukkan bahwa meskipun penelitian sebelumnya telah berhasil mengimplementasikan LightGBM pada konteks pasar saham, pendekatan yang diusulkan melalui QIMA memberikan keunggulan yang lebih menonjol khususnya pada efisiensi dimensi dan kecepatan konvergensi. Dengan menekan jumlah fitur secara signifikan dan meningkatkan efisiensi komputasi, metode ini berhasil mengatasi keterbatasan pada teknik optimasi evolusioner seperti GA yang cenderung membutuhkan waktu lebih panjang untuk mencapai solusi optimal.

c) Interpretasi Fitur Terpilih

Analisis fitur yang paling sering dipertahankan QIMA (Gambar 5) memberikan gambaran mengenai karakteristik sinyal jangka pendek pada saham perbankan Indonesia. Fitur yang paling sering dipilih mencakup: indikator volatilitas jangka pendek (mis. *volatility_6*), variabel berbasis momentum (*log_return, price_change*), fitur *lag* harga dan volume (mis. *close_lag_1, return_lag_2, volume_lag_2*), serta komponen harga utama (*Open, High, Close*). Sebaliknya, banyak indikator yang berasal dari perhitungan periode panjang, seperti *Moving Average* 12–24 periode dan beberapa fitur turunan lainnya, cenderung dieliminasi.

Pola ini menunjukkan bahwa dinamika harga saham perbankan lebih dipengaruhi oleh sinyal jangka pendek, yakni momentum harga, volatilitas, dan perubahan volume ketimbang indikator tren jangka panjang yang sifatnya lagging. Dengan kata lain, QIMA tidak memilih fitur secara acak, tetapi mengidentifikasi variabel yang paling informatif untuk memprediksi fluktuasi harian.

V. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini menunjukkan bahwa penerapan metode QIMA pada model LightGBM mampu menjawab seluruh rumusan masalah dengan efektif. Kesimpulan yang diperoleh adalah:

1. Seleksi fitur berbasis metaheuristik terbukti mampu mengurangi multikolinearitas dan *noise* secara signifikan dengan mereduksi fitur hingga 68,42% tanpa menurunkan kualitas sinyal informasi.
2. Performa model meningkat, di mana R^2 yang sebelumnya negatif pada model *Baseline* berubah menjadi positif pada BBCA (0.0006) dan BBNI (0.0173), serta mengalami stabilisasi pada BBRI dan BMRI.
3. QIMA memberikan efisiensi komputasi yang signifikan dengan mempercepat waktu pelatihan hingga 6–7 kali lipat dibandingkan model tanpa seleksi fitur.

Saran

Penelitian selanjutnya dapat diarahkan pada perluasan cakupan model dan optimasi teknik seleksi fitur untuk meningkatkan akurasi prediksi. Salah satu opsi adalah menerapkan QIMA pada dataset lintas-sektor guna melihat konsistensi performa pada kondisi pasar yang lebih beragam. Selain itu, peneliti dapat membandingkan QIMA dengan metaheuristik lain seperti PSO, GA, atau HHO untuk mengetahui apakah terdapat metode yang mampu menghasilkan kombinasi fitur yang lebih stabil. Penelitian lanjutan juga dapat memasukkan variabel makroekonomi atau sentimen berita untuk memperkaya sumber informasi yang digunakan model. Terakhir, implementasi model secara real-time serta pengujian pada strategi trading berbasis backtesting dapat memberikan gambaran yang lebih komprehensif mengenai kelayakan metode ini di lingkungan pasar yang dinamis.

DAFTAR PUSTAKA

Abdulloh, F. F., Aminuddin, A., Rahardi, M., & Harianto, F. J. (2024). Performance Analysis of Machine Learning Algorithms Using the Ensemble Method in Predicting the Impact of Inflation on Indonesia ' s Economic Growth. *JOIV : International Journal on Informatics Visualization*, *8*(December), 2316–2323. https://doi.org/10.62527/joiv.8.4.2567

Buulolo, K., & Wau, K. (2024). Performance Comparison of Metaheuristic Optimization Algorithms in Solving Production Scheduling Problems. *Jurnal ICT : Information and Communication Technologies Journal*, *15*(2), 122–132. https://doi.org/10.35335/jict.v15i2.197

Dan, N., Li, Y., Nie, Z., & Li, Y. (2022). Stock

p-ISSN : 2406-7733
e-ISSN : 2597-9922

Volatility Prediction Based on 1D-CNN and LightGBM. *Conference ICIDC*, 1–6. https://doi.org/10.4108/eai.17-6-2022.2322733

Dey, A., Bhattacharyya, S., Dey, S., Konar, D., Platos, J., Snasel, V., Mrsic, L., & Pal, P. (2023). A Review of Quantum-Inspired Metaheuristic Algorithms for Automatic Clustering. *Mathematic*, *11*, 1–44. https://doi.org/10.3390/math11092018

Ferrouhi, E. M., & Bouabdallaoui, I. (2024). A comparative study of ensemble learning algorithms for high-frequency trading. *Scientific African*, *24*(February), e02161. https://doi.org/10.1016/j.sciaf.2024.e02161

Fitria, V. A., & Widayanti, L. (2024). Enhancing Accuracy in Stock Price Prediction: The Power of Optimization Algorithms. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *23*(2 SE-Articles), 405–418. https://doi.org/10.30812/matrik.v23i2.3785

Goto, S., & Yamada, T. (2025). Adding Noise to Reduce Noise : A Counter-Intuitive Approach to Stock Return Prediction. *The Review of Financial Studies (Working Paper)*. https://ies.keio.ac.jp/upload/20251014econo_S.Goto_WP.pdf

Hartanto, A. D., Kholik, Y. N., & Pristyanto, Y. (2023). Stock Price Time Series Data Forecasting Using the Light Gradient Boosting Machine (LightGBM) Model. *JOIV : International Journal on Informatics Visualization*, *7*(4), 2270–2279. https://doi.org/10.62527/joiv.7.4.1740

Huang, C., Cai, Y., Cao, J., & Deng, Y. (2025). Stock complex networks based on the GA-LightGBM model: The prediction of firm performance. *Information Sciences*, *700*, 121824. https://doi.org/https://doi.org/10.1016/j.ins.2024.121824

Husam, M. (2025). Quantum-Inspired Optimization Algorithms for Large-Scale Computational Mechanics Problems in Structural Design. *International Journal for Computational Methods in Engineering and Mechanics (IJCMEM)*, *2*(2), 1–7.

Kayani, U. N., Faruk, A., Khan, M., & Khan, M. (2024). Heliyon Unleashing the pandemic volatility : A glimpse into the stock market performance of developed economies during COVID-19. *Heliyon*, *10*(4), e25202. https://doi.org/10.1016/j.heliyon.2024.e25202

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

Khairunisa, N., & Hendikawati, P. (2024). Long Short-Term Memory and Gated Recurrent Unit Modeling for Stock Price Forecasting. *Jurnal Matematika Statistika Dan Komputasi*, *21*(1), 321–333. https://doi.org/10.20956/j.v21i1.35930

Khan, A. A., & Hussain, S. (2024). A Quantum-Inspired Predator – Prey Algorithm for Real-parameter Optimization. *Algorithm*, *17*(33), 1–20. https://doi.org/10.3390/a17010033

Liang, Xinyue. (2024). Stock Market Prediction with RNN-LSTM and GA-LSTM. *SHS Web Conf.*, *196*, 2006. https://doi.org/10.1051/shsconf/202419602006

Majid, M. A., Saputri, P. D., & Soehardjoepri. (2024). Stock Market Index Prediction using Bi-directional Long Short- Term Memory. *Journal of Applied Informatics and Computing (JAIC)*, *8*(1), 55–61. https://doi.org/10.30871/jaic.v8i1.7195

Saberironaghi, M., Ren, J., & Saberironaghi, A. (2025). Stock Market Prediction Using Machine Learning and Deep Learning Techniques : A Review. *AppliedMath*, *5*(3), 1–36. https://doi.org/10.3390/appliedmath5030076

Sandeep, P., & Sood, K. (2024). Scientometric analysis of quantum - inspired metaheuristic algorithms. *Artificial Intelligence Review*, *57*(2), 1–30. https://doi.org/10.1007/s10462-023-10659-1

Tabash, M. I., Chalissery, N., Nishad, T. M., & Al-absy, M. S. M. (2024). Market Shocks and Stock Volatility : Evidence from Emerging and Developed Markets. *International Journal of Financial Studies*, *12*(1), 1–18. https://doi.org/10.3390/ijfs12010002

Xiong, X., Wang, A., He, J., & Wang, C. (2024). Application of LightGBM hybrid model based on TPE algorithm optimization in sleep apnea detection. *Frontiers in Neuroscience*, *18*(1324933), 1–17. https://doi.org/10.3389/fnins.2024.1324933