

STOCK PREDICTION PERFORMANCE OPTIMIZATION: ENHANCING COVARIANCE MATRIX WITH KNN

Iskandar Abdul Azis Saputra¹, Muhammad Rais Sidiq², Sangaji Suryo Guritno³, Hasan Dwi Cahyono^{*4}

^{1,2,3,4}Faculty of Information Technology and Data Science, Universitas Sebelas Maret, Indonesia
Email: ¹andar.alfath@student.uns.ac.id, ²raissidiq16@student.uns.ac.id, ³sguritno16@student.uns.ac.id,
⁴hasandc@staff.uns.ac.id

(Article received: June 27, 2024; Revision: July 27, 2024; published: December 29, 2024)

Abstract

Stock price prediction is a fundamental yet complex challenge in quantitative finance. With the increasing availability of data and advancements in machine learning techniques, various models have been developed to capture intricate patterns in stock price movements. While complex neural network models such as Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs), and Transformers have shown potential in handling stock market data, they often face optimization difficulties and performance limitations, especially when data is scarce. This paper explores the use of simpler and more accessible prediction methods, specifically Linear Regression (LR) and K-Nearest Neighbors (KNN), alongside more advanced models like Temporal Spatial Transformer (TST) and a Multi-Layer Perceptron (MLP) model called Stockmixer. The NASDAQ dataset is utilized in this study, providing a comprehensive view of stock market dynamics with high variability. Results indicate that KNN, among the evaluated models, exhibits superior and more stable performance in predicting validation data compared to MLP. KNN achieved a low Mean Squared Error (MSE) at 100 epochs, and demonstrated positive Information Coefficient (IC) and Return Information Coefficient (RIC) values. Additionally, it showed high Precision at 10 (P@10) and Sharpe Ratio (SR), making it a robust choice for stock price prediction tasks. In contrast, MLP, despite its sophistication, revealed some weaknesses, particularly in the alignment between predictions and actual values. These findings offer valuable insights into the effectiveness of various models for stock price prediction and suggest that simpler models like KNN can provide competitive results compared to more complex models.

Keywords: K-Nearest Neighbours, Multi-Layer Perceptron, NASDAQ, Stock Price Prediction.

OPTIMALISASI KINERJA STOCKMIXER: MENINGKATKAN MATRIKS KOVARIANS DENGAN KNN

Abstrak

Prediksi harga saham adalah tantangan mendasar namun kompleks dalam keuangan kuantitatif. Dengan meningkatnya ketersediaan data dan kemajuan teknik pembelajaran mesin, berbagai model telah dikembangkan untuk menangkap pola rumit dalam pergerakan harga saham. Meskipun model jaringan saraf yang kompleks, seperti *Recurrent Neural Networks* (RNN), *Graph Neural Networks* (GNN), dan *Transformer*, menunjukkan potensi dalam menangani data pasar saham, mereka sering menghadapi kesulitan dalam hal optimasi dan keterbatasan kinerja, terutama ketika data terbatas. Makalah ini mengeksplorasi penggunaan metode prediksi harga saham yang lebih sederhana dan dapat diakses, yaitu *Linear Regression* (LR) dan *K-Nearest Neighbors* (KNN), serta model yang lebih canggih seperti *Temporal Spatial Transformer* (TST) dan *Multi-Layer Perceptron* (MLP) berbasis *Stockmixer*. Dataset NASDAQ digunakan dalam penelitian ini, memberikan gambaran tentang dinamika pasar saham dengan keragaman yang tinggi. Hasil eksperimen menunjukkan bahwa KNN, di antara model-model yang dievaluasi, menunjukkan performa yang lebih baik dan lebih stabil dalam memprediksi data validasi dibandingkan dengan MLP. KNN mencapai *Mean Squared Error* (MSE) yang rendah pada 100 epoch, serta menunjukkan nilai Indeks Kinerja (IC) dan *Return Information Coefficient* (RIC) yang positif. Selain itu, model ini juga menunjukkan nilai *Precision at 10* (P@10) dan *Sharpe Ratio* (SR) yang tinggi, menjadikannya pilihan yang solid untuk tugas prediksi harga saham. Sebaliknya, MLP, meskipun merupakan model yang lebih canggih, menunjukkan beberapa kelemahan, terutama dalam hubungan antara prediksi dan nilai aktual. Temuan ini memberikan wawasan penting mengenai efektivitas berbagai model dalam konteks prediksi harga saham dan menunjukkan bahwa model yang lebih sederhana seperti KNN dapat memberikan hasil yang kompetitif dibandingkan dengan model yang lebih kompleks.

Kata kunci: *K-Nearest Neighbors, Multi-Layer Perceptron, NASDAQ, Prediksi Harga Saham.*

1. PENDAHULUAN

Salah satu topik yang paling menarik untuk dibicarakan dalam kaitannya dengan investasi jangka panjang dan jangka pendek adalah upaya prediksi, terutama dalam konteks prediksi harga saham [1]. Peramalan harga saham sangat penting dalam keuangan kuantitatif dan strategi investasi. Keberhasilan dalam meramalkan harga saham dapat memberikan keuntungan finansial yang signifikan dan memungkinkan investor untuk membuat keputusan yang lebih baik. Sehingga, ramalan saham ini sangat penting untuk mengetahui perkiraan harga saham di masa depan dan mencegah investor mengalami kerugian yang signifikan [2]. Namun, pergerakan harga saham yang saling terkait serta volatilitas dan dinamika kompleks pasar saham menimbulkan tantangan signifikan dalam mencapai akurasi prediksi yang tinggi.

Peramalan harga saham sering diformulasikan sebagai masalah deret waktu *multivariate* karena adanya hubungan yang kompleks antara berbagai faktor yang mempengaruhi harga saham. Faktor-faktor ini meliputi indikator teknis seperti *moving averages*, volume perdagangan, serta faktor fundamental seperti pendapatan perusahaan dan kondisi ekonomi global [3]. Hubungan temporal antara harga saham pada waktu-waktu yang berbeda juga memainkan peran penting dalam peramalan ini [4].

Dalam upaya menemukan pola dalam data saham, berbagai teknik pembelajaran mesin tradisional telah digunakan. Pohon keputusan adalah salah satu metode yang populer karena kemampuannya untuk menangkap hubungan non-linear dalam data. *Support vector machine* (SVM) digunakan untuk menemukan *hyperplane* yang memisahkan kelas dalam data yang tidak linear, sementara *k-nearest neighbors* (KNN) adalah metode yang sederhana namun efektif yang mengklasifikasikan titik data berdasarkan kedekatannya dengan titik-titik data lain dalam ruang fitur [5].

Dengan munculnya pembelajaran mendalam, metode yang lebih canggih diperkenalkan untuk menangani kompleksitas data saham. *Recurrent Neural Networks* (RNN) mampu memodelkan hubungan temporal dalam data saham dengan mempertimbangkan urutan data secara eksplisit [6]. *Graph Neural Networks* (GNN) dapat menangkap korelasi antara saham yang berbeda dengan memodelkan pasar saham sebagai graf, di mana *node* mewakili saham dan *edge* mewakili hubungan antar saham [7]. *Transformer* yang awalnya dikembangkan untuk pemrosesan bahasa alami, juga telah digunakan dalam peramalan harga saham karena kemampuannya untuk memodelkan hubungan jangka

panjang dalam data dengan diperkenalkannya *Temporal Spatial Transformer* (TST) [8].

Meskipun memiliki kekuatan ekspresif yang tinggi, model-model kompleks ini memiliki beberapa kelemahan. Mereka rentan terhadap *overfitting*, terutama karena ukuran *dataset* saham yang terbatas. Model ini juga sulit dioptimalkan dan mungkin mengalami bias induktif yang tidak akurat. Misalnya, GNN mungkin mengasumsikan kehalusan antara saham yang terkait, yang tidak selalu benar. Oleh karena itu, terdapat kebutuhan untuk model yang lebih sederhana yang mudah dioptimalkan dan mampu melakukan generalisasi dengan baik.

Dari beberapa model dan metode yang sudah dipaparkan, sudah dilakukan beberapa penelitian berkaitan dengan prediksi harga saham. Pada penelitian yang dilakukan oleh [9] mengenai prediksi harga saham menggunakan 7 fitur, yaitu data harga paling rendah, harga paling tinggi, harga buka, harga tutup, volume, rata-rata dan perubahan yang diperoleh dalam suatu perusahaan. Dari penelitian tersebut, diperoleh hasil akurasi untuk data latih sebesar 94.16% dan data uji sebesar 55.26%. Sementara itu, penelitian oleh [10] membahas mengenai prediksi harga saham menggunakan metode KNN pada data saham PT Bank Central Asia Tbk. Hasil yang diperoleh pada penelitian tersebut yaitu tingkat akurasi sebesar 61.79% pada prediksi harga saham menggunakan metode KNN.

Pada penelitian ini, kami mencoba mencari model terbaik untuk prediksi harga saham, diantaranya *Linear Regression* (LR), KNN, TST, dan metode berbasis *Multi-Layer Perceptron* (MLP) yang bernama *StockMixer*. Penelitian ini bertujuan untuk membandingkan performa berbagai model prediksi harga saham, termasuk LR [11], KNN [12], TST, dan *StockMixer* (SM) [13]. Khususnya, penelitian ini akan mengevaluasi model mana yang paling optimal dalam memprediksi harga saham NASDAQ berdasarkan metrik evaluasi seperti *Mean Squared Error* (MSE), *Information Coefficient* (IC), *Rank Information Coefficient* (RIC), *Precision at 10* ($p@10$), dan *Sharpe Ratio* (SR).

2. METODE PENELITIAN

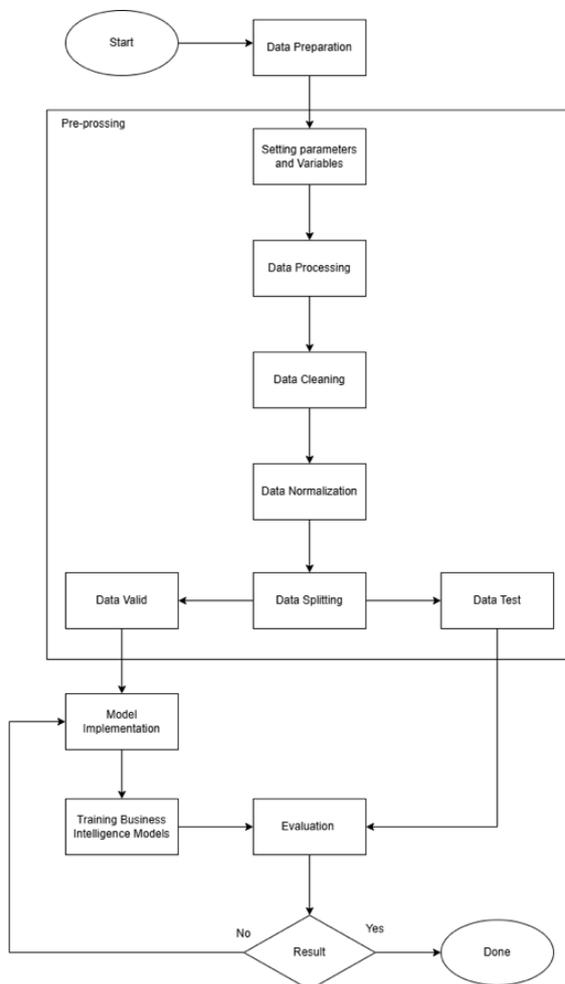
Penelitian kami dimulai dengan *pre-processing* (persiapan data) yaitu proses pengumpulan dan pembersihan data dari *noise* (gangguan) atau *outlier*. Selanjutnya, data dipersiapkan lebih lanjut melalui pengaturan parameter dan variabel serta normalisasi. Setelah itu, data divalidasi dan diuji sebelum model diterapkan. Model dilatih dan dievaluasi, dan jika hasil memuaskan, penelitian dianggap selesai. Gambar 1 menginformasikan pendekatan sistematis dalam mengolah dan menganalisis data yang penting untuk keabsahan dan kehandalan temuan penelitian.

2.1. Data Preparation dan Pre-processing

Tahap ini melibatkan persiapan data (*data preparation*) sebelum analisis lebih lanjut (*pre-processing*). Aktivitas termasuk membersihkan data dari *noise* dan mengisi nilai yang hilang. Selain itu, tahapan ini juga melibatkan pengumpulan, penggabungan, dan penyusunan data agar siap untuk analisis.

2.2. Model Implementation dan Evaluation

Penerapan model (*model implementation*) atau melibatkan pelatihan model agar memahami pola dalam data. Tahap penting untuk mengukur performa model dan hasil penelitian. Evaluasi (*evaluation*) melibatkan pengukuran akurasi dan kesesuaian hasil dengan tujuan awal.



Gambar 1. Alur Penelitian

2.3. Dataset

Dataset NASDAQ yang digunakan dalam penelitian ini merupakan kumpulan data yang dikenal luas dan komprehensif berasal dari pasar saham NASDAQ. *Dataset* ini meliputi berbagai informasi terkait saham, termasuk informasi harga penutupan, volume perdagangan, dan fitur-fitur tambahan yang relevan untuk analisis pasar. Untuk memastikan

konsistensi dan performa hasil yang didapatkan, penelitian ini akan membandingkan menggunakan data yang diambil dari sumber yang sama seperti pada penelitian [14].

2.4. Model dan Statistic Analysis

Penggunaan dan penerapan model *machine learning* serta teknik analisis statistik digunakan untuk memahami dan memproses data. Tujuan utama dari sub-bab ini adalah untuk memberikan pemahaman yang komprehensif tentang bagaimana berbagai model *machine learning* dan metode statistik digunakan dalam analisis data, termasuk cara memilih, melatih, dan mengevaluasi model untuk berbagai aplikasi. *Machine learning* melibatkan algoritma yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa pemrograman eksplisit. Analisis statistik melibatkan teknik untuk mengumpulkan, menganalisis, dan menarik kesimpulan dari data [15].

2.4.1. K-Nearest Neighbors (KNN)

Algoritma KNN digunakan untuk klasifikasi dan regresi dengan memprediksi nilai berdasarkan tetangga terdekat dalam ruang fitur. Penjelasan mencakup bagaimana memilih nilai K dan metode pengukuran jarak seperti *Euclidean distance*. Dengan menggunakan jarak terdekat, metode KNN ini tidak memerlukan asumsi tertentu (*non-parametric*) [16].

2.4.2. Multi-Layer Perceptron (MLP)

MLP adalah algoritma yang terdiri dari beberapa lapisan neuron, yang dapat menangani data non-linear. MLP juga memiliki fungsi aktivasi, dan teknik pelatihan seperti *backpropagation*. MLP merupakan algoritma yang mirip dengan jaringan syaraf tiruan manusia [17].

2.4.3. Time Series Trees (TST) dan Linear Regression (LR)

Penjelasan tentang time series trees (TST) adalah model berbasis decision tree yang dapat digunakan untuk analisis deret waktu (time series) dengan mendekomposisi data deret waktu menjadi fitur-fitur yang relevan. Selain itu, *model linear regression* (LR) digunakan untuk klasifikasi biner dengan mengestimasi probabilitas. Penjelasan mencakup interpretasi koefisien dan evaluasi model menggunakan metrik seperti akurasi dan ROC-AUC[18].

3. HASIL DAN PEMBAHASAN

Penelitian dilakukan dengan melakukan perbandingan antar model, di sini kami menggunakan perbandingan pada 15 *epoch* dan 100 *epoch*. Penelitian 15 *epoch* dilakukan untuk mencari model terbaik dari model yang akan diujikan yaitu LR, KNN, TST dan SM. Pada kasus 100 *epoch* dilakukan

setelah melakukan uji coba pada 15 *epoch*, pada 100 *epoch* dilakukan untuk mendapatkan perbandingan antara KNN dan *SM*, KNN dipilih karena pada percobaan 15 *epoch* mendapatkan hasil yang paling baik berdasarkan matriks *kovarian*.

3.1. Data Preparation

Data NASDAQ yang digunakan dalam penelitian ini didapatkan dari penelitian sebelumnya [14], [15]. Penelitian ini memfokuskan pada *dataset* NASDAQ yang uji pada semua model untuk mendapatkan perbandingan model terbaik. *Dataset* NASDAQ yang diambil hanya hingga kolom ke-915 dan mengambil semua fitur. *Dataset* dibagi menjadi empat *file* dengan kegunaan masing – masing, yaitu *end of day* (*eod*) data, *ground truth* (*gt*) data, *mask* data, dan *price* data. Hal ini dilakukan untuk membantu pada proses pengelolaan, pemrosesan, dan analisis data. Harga penutupan saham diambil dari data indeks terakhir pada setiap fitur dan menginisialisasi data. Nilai *gt* dihitung sebagai persentase perubahan harga penutup dengan langkah waktu tertentu serta melakukan *looping* untuk menghitung perubahan harga pada setiap saham dan setiap waktu. Harga penutupan saham diperoleh dari data indeks terakhir pada setiap fitur dan diinisialisasi dengan nilai 1, menandakan bahwa data tersebut valid. Nilai *ground truth* dihitung sebagai persentase perubahan harga penutupan pada interval waktu tertentu, dengan pendekatan *looping* untuk mengukur perubahan harga setiap saham pada setiap periode waktu.

3.2. Data Preprocessing

Data pre-processing adalah serangkaian langkah yang diambil untuk mengubah data mentah menjadi format yang siap digunakan untuk analisis atau model *machine learning*. Tahap ini sangat penting untuk memastikan kualitas data yang akan digunakan. Langkah – langkah utama dalam pemrosesan data dimulai sebagai berikut: .

3.2.1. Setting Parameters dan Variables

Dalam penelitian ini, berbagai parameter dan variabel telah diatur untuk memastikan konfigurasi yang tepat dalam analisis pasar saham. Jumlah saham yang dianalisis adalah 1026, ditentukan oleh *stock_num*. Panjang periode pengamatan untuk data historis adalah 16 hari. Indeks validasi dan pengujian ditetapkan pada 756 dan 1008 (*valid_index* dan *test_index*). Model menggunakan 5 fitur (*fea_num*) dan mencakup 20 pasar berbeda (*market_num*). Kecepatan pembelajaran untuk optimasi ditentukan oleh *learning_rate* (0.001), dengan koefisien regulasi (*alpha*) sebesar 0.1 dan faktor skala (*scale_factor*) sebesar 3. Fungsi aktivasi yang digunakan dalam model adalah GELU (*activation*)..

3.2.2. Data Cleaning

Langkah ini melibatkan identifikasi dan penanganan *missing values* atau data yang tidak lengkap [17]. Pendekatan yang umum digunakan termasuk penghapusan data yang tidak lengkap. Selain itu, pendekatan ini juga biasa disebut sebagai imputasi nilai-nilai yang hilang.

Tabel 1. Hasil matriks *kovarian* pada *train set*

Model	MSE (10^{-2})	IC (10^{-2})	RIC (10^{-1})	P@10 (10^{-1})	SR
LR	0.05	1.08	-0.004	0.001	-0.06
KNN	0.09	2.71	2.46	5.26	2.43
TST	2450	-1.95	-2.23	5.31	-3.62
SM	0.141	-2.34	-1.85	5.31	0.39

3.2.3. Data Normalization

Untuk memastikan konsistensi dan stabilitas pelatihan model, data perlu dinormalisasi. Normalisasi data mengubah skala fitur sehingga berada dalam rentang tertentu. Skala ini biasanya antara 0 dan 1 yang membantu dalam mempercepat konvergensi model dan meningkatkan performa prediksi [18].

3.2.4. Data Splitting

Data yang telah diproses kemudian dibagi menjadi tiga bagian yaitu *train*, *validation*, dan *test set*. *Train set* digunakan untuk melatih model. Sementara, *validation set* digunakan untuk *tuning* parameter model, dan *test set* digunakan untuk mengevaluasi performa akhir model [19]. Pemrosesan data dilakukan dengan menggunakan *library scikit-learn* yang menyediakan berbagai fungsi dan utilitas untuk keperluan *pre-processing* data, dengan menggunakan beberapa fungsi *Imputer* dapat mengatasi *missing values* dengan teknik imputasi, seperti penggantian nilai yang hilang dengan *mean*, median, atau mode dari kolom terkait. Selain itu, *MinMaxScaler* digunakan untuk normalisasi fitur-fitur data agar berada dalam rentang tertentu. Selanjutnya, *train_test_split* berfungsi untuk membagi *dataset* menjadi *train*, *validation*, dan *test set*.

3.2.4.1. Validation Set

Validation set, di sisi lain, digunakan untuk mengevaluasi performa model selama proses pelatihan. Data ini tidak digunakan untuk melatih model, melainkan untuk memantau apakah model tersebut mengalami *overfitting* atau tidak. Evaluasi performa pada data validasi dilakukan pada setiap *epoch*, yang bertujuan untuk menentukan parameter terbaik bagi model. Dalam kode, evaluasi ini dilakukan melalui fungsi *validate* (*valid_index*, *test_index*) yang menghasilkan nilai *val_loss*, *val_reg_loss*, *val_rank_loss*, dan *val_perf*.

3.2.4.2. Test set

Test set digunakan untuk mengevaluasi performa akhir model setelah proses pelatihan selesai. Data ini memberikan gambaran tentang bagaimana model akan bekerja pada data yang benar-benar baru dan tidak terlihat sebelumnya. Evaluasi performa pada *test set* juga dilakukan melalui fungsi *validate(test_index, trade_dates)*, yang menghasilkan nilai *test_loss*, *test_reg_loss*, *test_rank_loss*, dan *test_perf*.

3.3. Model Implementation - Training Business Intelligence Models and Evaluation

Dataset historis harga saham NASDAQ yang terdiri dari N data titik, masing-masing data terdiri dari fitur-fitur dan label, tujuan penelitian ini adalah meminimalkan MSE antara harga saham yang diprediksi dengan harga saham aktual:

Tabel 2. Hasil matriks kovarian pada *validation set*

Model	MSE (10^{-4})	IC (10^{-3})	RIC (10^{-2})	P@10 (10^{-1})	SR (10^{-1})
LR	3.82	6.04	-0.01	0.03	-0.60
KNN	8.14	-1.54	-2.57	5.22	-27.70
TST	5730.00	2.34	3.64	4.48	-80.28
MLP	12.60	-4.87	-6.07	5.37	-6.20

Tabel 3. Hasil matriks kovarian pada *validation set* dengan 100 *epoch*

Model	MSE (10^{-3})	IC (10^{-2})	RIC (10^{-1})	P@10 (10^{-1})	SR (10^{-1})
MLP	1.18	-2.16	-1.59	5.52	8.89
KNN	0.91	2.33	2.21	5.26	24.4

Tabel 4. Hasil matriks kovarian pada *test set* dengan 100 *epoch*

Model	MSE (10^{-4})	IC (10^{-3})	RIC (10^{-2})	P@10 (10^{-1})	SR (10^{-1})
SM[14]	9.85	-7.16	-8.58	5.37	1.93
KNN	8.19	-1.09	-1.82	5.15	-5.10

Berdasarkan

Tabel 1, untuk model LR dan KNN menunjukkan MSE yang rendah, akan tetapi nilai dari IC dan RIC mendekati nol dan negatif menunjukkan hubungan yang lemah atau tidak ada antara prediksi model dan nilai sebenarnya, nilai pada KNN masih lebih baik jika dibandingkan dengan LR. TST menunjukkan bahwa performa yang buruk dengan MSE yang tinggi dan nilai evaluasi lainnya negatif, hal ini memungkinkan bahwa model ini tidak cocok untuk *dataset* ini. Hasil evaluasi dari MLP menunjukkan MSE yang rendah dan performa yang relatif baik pada P@10 dan SR, tetapi dengan nilai IC dan RIC yang negatif. Hasil negatif ini menunjukkan hubungan yang lemah antara prediksi dan nilai sebenarnya. Pada kasus pengujian pada data validasi menunjukkan KNN memiliki performa terbaik dalam validasi, diikuti oleh MLP.

3.4. Model Evaluation

Dari

Tabel 2, menunjukkan bahwa LR memiliki nilai MSE yang paling rendah, menunjukkan kesalahan

prediksi yang kecil. Namun, nilai IC dan RIC yang mendekati nol dan SR yang negatif menunjukkan bahwa model ini memiliki hubungan yang lemah antara prediksi dan nilai sebenarnya serta tidak menguntungkan berdasarkan rasio risiko. KNN pada data uji mengalami *overfitting* ditunjukkan dari nilai IC dan RIC yang negatif serta SR yang negatif. Dari hasil tersebut, LR menunjukkan performa terbaik dalam hal MSE pada data uji, namun semua model memiliki kelemahan dalam hal hubungan antara prediksi dan nilai sebenarnya serta keuntungan berdasarkan rasio risiko.

3.5. Results

Pada analisis didapatkan hasil evaluasi lain seperti yang sudah dijelaskan sebelumnya perbandingan ini diambil dengan melakukan pengujian dengan 100 *epoch* dan 15 *epoch*. Nilai yang ditunjukkan sebelumnya adalah hasil pada pelatihan 15 *epoch*, berikut adalah hasil komparasi pada 100 *epoch* yang dilakukan pada model KNN dan MLP, berdasarkan hasil validasi dan hasil pengujian terbaik.

Mengacu pada Tabel 3, dapat diketahui bahwa KNN memiliki performa yang lebih baik daripada MLP dalam memprediksi data validasi. KNN menunjukkan MSE yang lebih rendah, hubungan yang lebih kuat antara prediksi dan nilai sebenarnya (IC dan RIC positif), serta tingkat presisi yang tinggi (P@10 tinggi) dan keuntungan yang tinggi berdasarkan rasio risiko (SR tinggi). Pada MLP menunjukkan bahwa MSE yang rendah, memiliki nilai IC dan RIC yang negatif, menunjukkan bahwa hubungan antara prediksi dan nilai sebenarnya cenderung lemah. Namun, model ini masih mampu memprediksi dengan baik pada posisi atas (P@10 tinggi) dan menghasilkan keuntungan berdasarkan rasio risiko yang positif (SR positif). Dengan demikian, untuk prediksi pada data validasi, KNN menjadi pilihan yang lebih baik berdasarkan performa yang lebih konsisten dan tingkat keuntungan yang lebih tinggi.

Penelitian yang kami lakukan dibandingkan dengan penelitian terdahulu seperti yang ditampilkan pada Tabel 4. Untuk *source code* dari penelitian kami dapat diakses pada [github https://github.com/Azoss/Business_Intelligence_project](https://github.com/Azoss/Business_Intelligence_project). SM memiliki MSE yang sedikit lebih tinggi dibandingkan KNN, tetapi masih menunjukkan kemampuan prediksi yang baik dengan P@10 yang tinggi. Namun, hubungan antara prediksi dan nilai sebenarnya tidak terlalu kuat (IC dan RIC negatif). KNN menunjukkan MSE yang sedikit lebih rendah, menunjukkan kesalahan prediksi yang lebih rendah daripada MLP pada data uji. Meskipun demikian, performa prediksi pada posisi atas (P@10) sedikit lebih rendah dan SR negatif menunjukkan bahwa model ini tidak menguntungkan secara finansial berdasarkan rasio risiko. Dengan demikian, KNN menunjukkan performa yang lebih stabil dan sedikit

lebih baik dalam memprediksi data uji dibandingkan MLP dalam kasus ini.

4. DISKUSI

Penelitian ini menyoroti bahwa performa model prediksi sangat dipengaruhi oleh berbagai faktor seperti jenis model, jumlah *train epoch*, dan karakteristik *dataset*. Meskipun LR dan KNN memiliki nilai MSE yang rendah, kelemahan dalam hal hubungan antara prediksi dan nilai sebenarnya serta rasio risiko menunjukkan bahwa pemilihan model tidak bisa hanya didasarkan pada MSE saja.

Meskipun pada penerapan metode dengan KNN menunjukkan tanda-tanda *overfitting* pada data uji, menghasilkan performa yang paling konsisten dan menguntungkan pada data validasi setelah 100 *epoch*. Sebaliknya, TST ternyata tidak cocok untuk *dataset* ini, menekankan pentingnya memilih model yang sesuai dengan karakteristik *dataset*. MLP menunjukkan performa yang hampir setara dengan KNN dalam beberapa metrik, namun hubungan yang lemah antara prediksi dan nilai sebenarnya mengurangi keandalan model ini. Ini menunjukkan bahwa meskipun model *neural networks* seperti MLP dapat memberikan hasil yang baik dalam beberapa aspek, tantangan tetap ada dalam hal *interpretabilitas* dan hubungan yang kuat dengan data sebenarnya.

Penelitian ini membandingkan hasil eksperimen dengan penelitian terdahulu yang menerapkan metode *Multi Layers Processing* (MLP) dalam analisis prediksi pasar saham NASDAQ. Pada hasil dari penelitian terdahulu menunjukkan bahwa MLP dapat secara efektif memodelkan pola kompleks dalam data pasar saham untuk prediksi harga. Dalam penelitian kami, kami menerapkan KNN untuk meningkatkan nilai *matrik kovarian* pada SM. Hasil eksperimen kami menunjukkan peningkatan nilai *matrik kovarian* terutama pada SR mengalami peningkatan yang cukup signifikan. Hasil dari perbandingan SM KNN dengan SM dibandingkan untuk melihat dimana peningkatan yang dihasilkan dengan model pada penelitian kami dengan penelitian sebelumnya seperti yang ditunjukkan pada Tabel 5.

Tabel 5 menunjukkan perbandingan antara SM dan SM yang dioptimalkan dengan metode KNN. Perbandingan ini memberikan wawasan tentang bagaimana peningkatan teknik dapat mempengaruhi performa model dalam analisis prediksi pasar saham. Selain itu, perbandingan ini sejalan dengan penelitian terdahulu yang dilaporkan di [14], dimana MLP menunjukkan kemampuan untuk memodelkan pola kompleks dalam data pasar saham dengan hasil yang konsisten dalam prediksi harga. Namun, hasil dari penelitian kami menunjukkan bahwa optimasi KNN dapat memberikan peningkatan dalam nilai *matrik kovarian*, terutama pada metrik SR, dibandingkan dengan model yang tidak di optimalisasi.

5. KESIMPULAN

Penelitian ini membandingkan performa beberapa model prediksi, yaitu LR, KNN, TST, dan MLP, pada *dataset* tertentu. Uji coba dilakukan dalam dua tahap: 15 *epoch* untuk mencari model terbaik dan 100 *epoch* untuk perbandingan antara model KNN dan MLP. Hasil penelitian menunjukkan bahwa LR memiliki nilai MSE yang rendah. Di samping itu, LR juga memiliki nilai IC dan RIC mendekati nol atau negatif, menunjukkan hubungan yang lemah antara prediksi dan nilai sebenarnya. SR negatif pada LR menunjukkan model ini tidak menguntungkan berdasarkan rasio risiko.

Dengan 15 *epoch*, KNN menunjukkan hasil terbaik berdasarkan matriks kovarian. Pada 100 *epoch*, KNN menunjukkan performa yang baik dengan MSE yang rendah, IC dan RIC positif, serta P@10 dan SR yang tinggi. Selain itu, TST mengalami *overfitting* pada data uji dengan nilai IC, RIC, dan SR yang negatif.

Tabel 5. Hasil perbandingan dengan penelitian terdahulu

Model	IC	RIC	P@10	SR
SM	0.043	0.501	0.545	1.465
KNN	0.0259	0.232	0.526	2.63

TST menunjukkan performa yang buruk dengan MSE tinggi dan nilai evaluasi lainnya negatif. Berdasarkan percobaan yang telah kami lakukan tersebut, TST menunjukkan bahwa model ini tidak cocok untuk *dataset* ini. Sehingga, investigasi lebih lanjut diperlukan agar dapat meningkatkan performa TST pada *dataset* ini. MLP menunjukkan MSE yang rendah dan performa yang relatif baik pada P@10 dan SR. Nilai IC dan RIC negatif pada performa MLP menunjukkan hubungan yang lemah antara prediksi dan nilai sebenarnya. Pada 100 *epoch*, MLP memiliki MSE yang sedikit lebih tinggi dibandingkan KNN, tetapi masih menunjukkan kemampuan prediksi yang baik dengan P@10 yang tinggi dan SR positif. Secara keseluruhan, KNN menunjukkan performa yang lebih baik dan lebih stabil dalam memprediksi data validasi dibandingkan dengan MLP, meskipun kedua model memiliki kelemahan dalam hal hubungan antara prediksi dan nilai sebenarnya.

6. UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah mendukung penelitian ini. Ucapan terima kasih kami sampaikan kepada pembimbing dan mentor atas bimbingan dan dukungannya, rekan peneliti atas kerja sama yang luar biasa, keluarga dan sahabat atas dukungan moral yang tak ternilai, serta institusi dan organisasi yang menyediakan fasilitas dan sumber daya. Terima kasih juga kepada para responden dan partisipan yang telah berkontribusi dengan data dan informasi penting. Semoga hasil penelitian ini memberikan manfaat yang berarti bagi pengembangan ilmu pengetahuan dan aplikasi praktis. Terima kasih.

DAFTAR PUSTAKA

- [1] A. O. Indarso and A. B. Pangaribuan, "Penggunaan Metode Multilayer Perceptron Pada Prediksi Indeks Saham LQ45," *Informatik : Jurnal Ilmu Komputer*, vol. 17, no. 1, p. 38, May 2021, doi: 10.52958/iftk.v17i1.2225.
- [2] RR. K. N. Sari, W. Sutisna, M. J. M. Wororomi, and V. R. Tjahjono, "Komparasi Model Gerak Brown Geometrik Termodifikasi dan Model Kecerdasan Buatan untuk Prediksi Harga Saham Sektor Kesehatan di Indonesia," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 2, Oct. 2023, doi: 10.23887/jstundiksha.v12i2.48960.
- [3] W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain of Informatics Journal*, vol. 5, no. 2, p. 45, Sep. 2020, doi: 10.21111/fij.v5i2.4449.
- [4] W. Y. Rusyida and V. Y. Pratama, "Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA," *Square : Journal of Mathematics and Mathematics Education*, vol. 2, no. 1, p. 73, Apr. 2020, doi: 10.21580/square.2020.2.1.5626.
- [5] E. Eka Patriya, "IMPLEMENTASI SUPPORT VECTOR MACHINE PADA PREDIKSI HARGA SAHAM GABUNGAN (IHSG)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 24–38, Apr. 2020, doi: 10.35760/tr.2020.v25i1.2571.
- [6] I. Akil and I. Chaidir, "Prediksi Harga Saham Twitter Dengan Long Short-Term Memory Recurrent Neural Network," *INTI Nusa Mandiri*, vol. 17, no. 1, pp. 1–7, Aug. 2022, doi: 10.33480/inti.v17i1.3277.
- [7] D. Matsunaga, T. Suzumura, and T. Takahashi, "Exploring Graph Neural Networks for Stock Market Predictions with Rolling Window Analysis," Sep. 2019, [Online]. Available: <http://arxiv.org/abs/1909.10660>
- [8] I. Idham, M. Ghudafa Taufik Akbar, S. Panggabean, and M. Noor, "Perbandingan Prediksi Harga Saham Dengan Menggunakan LSTM GRU Dengan Transformer," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, no. 1, pp. 44–47, Jan. 2022, doi: 10.30591/smartcomp.v11i1.3185.
- [9] M. A. D. Suyudi, E. C. Djamal, and A. Maspupah, "Prediksi Harga Saham menggunakan Metode Recurrent Neural Network," in *Seminar Nasional Aplikasi Teknologi Informatika (SNATI)*, 2019.
- [10] E. R. Tauran, "PREDIKSI HARGA SAHAM PT BANK CENTRAL ASIA TBK BERDASARKAN DATA DARI BURSA EFEK INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBORS (KNN)," *TeIKa*, vol. 11, no. 2, pp. 123–129, Oct. 2021, doi: 10.36342/teika.v11i2.2609.
- [11] P. C. Hartono and A. D. Widiatoro, "Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner," *Journal of Computer and Information Systems Ampera*, vol. 5, no. 3, 2024, doi: 10.51519/journalcisa.v5i3.481.
- [12] V.SARALA and G. N. V. P. BHUSHAN, "Stock Market Trend Prediction Using K-Nearest Neighbor (KNN)," *Journal of Engineering Sciences*, vol. 13, no. 8, pp. 249–256, 2022, doi: 10.15433/JES.2022.V13I08.43P.39.
- [13] E. Fitri and D. Riana, "ANALISA PERBANDINGAN MODEL PREDICTION DALAM PREDIKSI HARGA SAHAM MENGGUNAKAN METODE LINEAR REGRESSION, RANDOM FOREST REGRESSION DAN MULTILAYER PERCEPTRON," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 6, no. 1, pp. 69–78, Apr. 2022, doi: 10.46880/jmika.Vol6No1.pp69-78.
- [14] J. Fan and Y. Shen, "StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, pp. 8389–8397, Mar. 2024, doi: 10.1609/aaai.v38i8.28681.
- [15] X. Wei and Y. Li, "Fuzzy alternating automata over distributive lattices," *Inf Sci (N Y)*, vol. 425, pp. 34–47, Jan. 2018, doi: 10.1016/j.ins.2017.10.015.
- [16] Z. Zhu, "Research on Parameter Optimization in Collaborative Filtering Algorithm," *Communications and Network*, vol. 10, no. 03, pp. 105–116, 2018, doi: 10.4236/cn.2018.103009.
- [17] "Books received," *Brittonia*, vol. 36, no. 3, pp. 222–222, Jul. 1984, doi: 10.1007/BF02812122.
- [18] J. Lines and A. Bagnall, "Time series classification with ensembles of elastic distance measures," *Data Min Knowl Discov*, vol. 29, no. 3, pp. 565–592, May 2015, doi: 10.1007/s10618-014-0361-2.