

Implementasi Sistem Prediksi Saham Real-Time dengan Integrasi Yahoo Finance API dan Machine Learning di Google Colab: Analisis dan Evaluasi Performa Model

IMPLEMENTATION OF A REAL-TIME STOCK PREDICTION SYSTEM WITH INTEGRATION OF YAHOO FINANCE API AND MACHINE LEARNING IN GOOGLE COLAB: ANALYSIS AND EVALUATION OF MODEL PERFORMANCE.

Muhammad Faiq Abdi^{*1}, Yonhendri ²

^{1,2} Jurusan Teknik Informatika, Sekolah Tinggi Ilmu Komputer Muhammadiyah Batam
Jln. Prof. Dr. Hamka No. 03, Kelurahan Kibing, Kecamatan Batu Aji,
Kota Batam, Kepulauan Riau 29424, Indonesia
e-mail: abdifaiq9@gmail.com, ^{*2} yonhendri@gmail.com

Abstrak

Masalah prediksi harga saham merupakan topik penting dalam dunia keuangan, di mana akurasi prediksi sangat diperlukan untuk pengambilan keputusan investasi yang lebih baik. Penelitian ini mengembangkan sistem prediksi harga saham real-time dengan integrasi Yahoo Finance API dan algoritma Machine Learning yang dijalankan di Google Colab. Sistem ini memungkinkan pengambilan data pasar saham secara langsung dari Yahoo Finance untuk menganalisis pola data dan menghasilkan prediksi harga saham yang lebih akurat. Metode yang digunakan meliputi pengumpulan data historis dan real-time dari Yahoo Finance API, preprocessing data, pelatihan model menggunakan Long Short Term Memory (LSTM), validasi model dengan K-Fold Cross Validation, dan evaluasi kinerja menggunakan berbagai metrik standar. Hasil evaluasi menunjukkan bahwa model LSTM memiliki Mean Squared Error (MSE) sebesar 73325,26 dan Mean Absolute Error (MAE) sebesar 266,18, yang menunjukkan akurasi tinggi dalam prediksi harga saham. Sistem ini memberikan kontribusi signifikan dalam meningkatkan akurasi prediksi harga saham dan membantu investor membuat keputusan yang lebih baik berdasarkan data real-time.

Kata kunci — Prediksi saham, sistem *real-time*, *Yahoo Finance API*, *machine learning*, *Google Colab*

Abstract

The issue of stock price prediction is a critical topic in the financial world, where prediction accuracy is essential for making better investment decisions. This research develops a real-time stock price prediction system by integrating the Yahoo Finance API and machine learning algorithms executed on the Google Colab platform. This system allows for direct retrieval of stock market data from Yahoo Finance to analyze data patterns and generate more accurate stock price predictions. The methods used include collecting historical and real-time data from the Yahoo Finance API, data preprocessing, training models using Long Short-Term Memory (LSTM), validating the model with K-Fold Cross Validation, and evaluating performance using various standard metrics. The development and implementation of the model are carried out on Google Colab. The results show that the LSTM model can provide stock price predictions with high accuracy, achieving a Mean Squared Error (MSE) of 73325.26 and a Mean Absolute Error (MAE) of 266.18. This system significantly contributes to improving stock price prediction accuracy and helps investors make better decisions based on real-time data.

Kata kunci — *Stock prediction*, *real-time system*, *Yahoo Finance API*, *machine learning*, *Google Colab*

1. PENDAHULUAN

Dalam era digital yang semakin berkembang pesat, prediksi harga saham menjadi topik yang sangat penting bagi investor dalam pengambilan keputusan investasi di pasar modal. Peran teknologi memiliki peran penting dalam mengembangkan sistem prediksi

yang lebih akurat dan efisien. Penelitian ini mengimplementasikan sistem prediksi saham *real-time* menggunakan Yahoo Finance API dan teknologi *Machine Learning* melalui Google Colab. Algoritma *Deep Learning*, khususnya *Long Short Term Memory* (LSTM), digunakan untuk memprediksi harga saham dengan akurasi lebih tinggi dibandingkan metode tradisional.

LSTM, sebagai jenis Recurrent Neural Network (RNN), mampu menangani data berurutan dan mengatasi fluktuasi harga saham. Integrasi dengan Yahoo Finance API memungkinkan akses data saham real-time, yang kritis untuk analisis prediktif yang akurat [1,14]. Penelitian oleh Yen dan Siahaan [11] mengusulkan sistem prediksi berbasis website menggunakan algoritma hibrid ARIMA-LSTM, sementara penelitian saat ini mengimplementasikan pendekatan machine learning melalui platform Google Colab dengan integrasi Yahoo Finance API. Perbandingan metodologi menunjukkan perbedaan signifikan dalam strategi prediksi:

Tabel 1. Evaluasi MSE dan MAE

Aspek	Jurnal (ARIMA-LSTM Hybrid)	Penelitian Saat Ini
Algoritma	Hybrid ARIMA-LSTM	LSTM dan Random Forest
Platform	Berbasis Website	Google Colab
Sumber Data	Tidak Spesifik	Yahoo Finance API
Fokus Utama	Prediksi Statis	Prediksi Real-Time

Keunggulan utama penelitian ini terletak pada integrasi langsung dengan sumber data real-time melalui Yahoo Finance API, pemanfaatan komputasi *cloud* Google Colab, implementasi multi-algoritma (*LSTM* dan *Random Forest*), dan fokus pada prediksi *real-time* yang responsif. Dibandingkan dengan pendekatan hybrid ARIMA-LSTM pada jurnal sebelumnya, penelitian ini menawarkan fleksibilitas yang lebih tinggi dalam analisis data, akses langsung ke data pasar terkini, kemampuan komputasi yang lebih canggih, dan pendekatan *machine learning* yang lebih modern.

Tujuan utama penelitian adalah mengembangkan sistem prediksi saham yang tidak hanya akurat, tetapi juga responsif terhadap perubahan pasar real-time. Dengan memanfaatkan kekuatan machine learning dan integrasi API, penelitian ini bertujuan memberikan alat analisis yang lebih powerful bagi investor dalam mengambil keputusan investasi. Kontribusi inovatif penelitian mencakup metode prediksi multi-algoritma, integrasi sumber data *real-time*, aksesibilitas komputasi *cloud*, dan pendekatan fleksibel dalam analisis saham

2. METODE PENELITIAN

Penelitian ini menggunakan metode yang meliputi pengumpulan data historis dan real-time melalui *Yahoo Finance API*, dilanjutkan dengan tahap *preprocessing data*, pembangunan model menggunakan *Long Short Term Memory* (LSTM), validasi model dengan *K-Fold Cross Validation*, dan evaluasi kinerja menggunakan berbagai metrik standar.

Pengumpulan Data

Pengumpulan data harga saham dilakukan melalui Yahoo Finance API, yang menyediakan akses ke data historis serta data real-time dari berbagai pasar saham di seluruh dunia. Dengan menggunakan Yahoo Finance API, peneliti dapat memperoleh data yang komprehensif dan relevan untuk analisis prediksi harga saham, termasuk informasi tentang harga pembukaan, penutupan, tertinggi, terendah, dan volume perdagangan. Keunggulan dari penggunaan API ini adalah kemampuannya untuk memberikan data yang akurat dan terkini, yang sangat penting dalam konteks analisis pasar yang dinamis. Selain itu, Yahoo Finance API telah banyak digunakan dalam penelitian sebelumnya, yang menunjukkan kredibilitas dan keandalannya dalam menyediakan data pasar. Dengan akses yang mudah dan cepat ke informasi yang diperlukan, peneliti dapat melakukan analisis yang lebih mendalam dan menghasilkan model prediksi yang lebih akurat [1][2].

Data yang akan diambil mencakup harga penutupan harian saham, volume perdagangan, dan informasi terkait lainnya. Proses pengambilan data dilakukan menggunakan pustaka *yfinance* dalam *Python*, yang memudahkan akses dan pengunduhan data dari *Yahoo Finance*. Dengan menggunakan *yfinance*, peneliti dapat mengunduh data dalam format *Data Frame* yang mudah untuk dianalisis dan diproses lebih lanjut [3]. Periode pengumpulan data mencakup rentang waktu 1 Januari 2018 hingga 1 Januari 2023. Rentang waktu ini dipilih untuk menganalisis tren harga saham dalam jangka waktu yang cukup panjang, sehingga dapat memberikan wawasan yang lebih baik mengenai pola pergerakan harga dan faktor-faktor yang mempengaruhi harga saham [4].

Pemrosesan Data

Setelah proses pengunduhan data, langkah selanjutnya adalah melakukan pembersihan dan pra-pemrosesan data. Tahap ini mencakup penghapusan nilai yang hilang, normalisasi data menggunakan metode *Min-Max Scaling*, dan pembentukan *dataset* untuk pelatihan model. Proses ini memiliki peran penting dalam memastikan model yang dibangun dapat berfungsi dengan baik dan menghasilkan prediksi yang akurat [5].

Data yang telah dibersihkan akan dinormalisasi menggunakan metode *Min-Max Scaler*. Tahap normalisasi bertujuan untuk memastikan semua fitur berada dalam rentang skala yang sama, yang sangat penting untuk meningkatkan kinerja model *machine learning*. Melalui proses normalisasi, model akan lebih mudah untuk memahami pola dalam data tanpa terpengaruh oleh skala yang berbeda-beda [5].

Setelah proses normalisasi data selesai, *dataset* akan dibagi menjadi dua bagian: data latih (70%) dan data uji (30%). Pembagian ini dilakukan untuk mengevaluasi kinerja model secara akurat, dimana data latih dimanfaatkan untuk melatih model dan data uji digunakan untuk menguji kemampuan model setelah pelatihan selesai. Pembagian data membantu dalam memastikan model memiliki kemampuan generalisasi yang baik dan melakukan prediksi akurat pada data baru yang belum pernah diproses sebelumnya [6].

Pembangunan Model

Pembangunan model dimulai dengan pemilihan algoritma. *Long Short-Term Memory* (LSTM) dipilih untuk membangun model prediksi harga saham, karena kemampuannya dalam menangani data *time-series* dan kemampuannya dalam menangkap pola jangka panjang dalam data. LSTM adalah jenis jaringan saraf tiruan yang dirancang untuk memproses urutan data dengan cara yang mempertimbangkan konteks temporal [7].

Random Forest dipilih untuk membangun model prediksi harga saham karena kemampuannya dalam menangani data kompleks dan non-linier. Sebagai algoritma ensemble learning, Random Forest mampu mengekstraksi pola tersembunyi dalam data *time-series* keuangan melalui agregasi keputusan multiple decision trees [12].

Model LSTM akan dirancang dengan beberapa lapisan tersembunyi dan jumlah neuron yang dioptimalkan untuk mencapai keseimbangan antara *overfitting* dan *underfitting*. *Overfitting* terjadi ketika model terlalu sesuai dengan data latih sehingga tidak mampu menggeneralisasi pada data baru, sedangkan *underfitting* terjadi ketika model terlalu sederhana dan tidak mampu menangkap pola dalam data latih. Dengan pemilihan arsitektur yang tepat, model diharapkan memiliki performa yang baik dalam memprediksi harga saham [7].

Model Random Forest dirancang untuk mengoptimalkan prediksi harga saham melalui pengaturan parameter yang tepat, mencegah *overfitting* dan *underfitting* dengan strategi kontrol jumlah pohon, kedalaman, dan pemilihan fitur representatif, bertujuan menghasilkan model dengan performa prediksi tinggi dan kemampuan generalisasi yang baik [12-13].

Pelatihan Model

Model akan dilatih menggunakan *dataset* latih dengan parameter yang disesuaikan seperti *batch size*, jumlah *epoch*, dan fungsi aktivasi yang sesuai. *Batch size* menentukan jumlah sampel yang diproses sebelum model diperbarui, sedangkan *epoch* adalah jumlah iterasi pelatihan melalui seluruh dataset. Fungsi aktivasi membantu model untuk belajar pola non-linear dalam data [8].

Untuk mencegah *overfitting*, berbagai teknik optimasi akan diterapkan, termasuk *Early Stopping* dan *Regularization*. *Early Stopping* berfungsi untuk menghentikan pelatihan model ketika performa pada data uji mulai menurun, sedangkan *Regularization* dapat menambahkan penalti pada kompleksitas model untuk mencegah model menjadi terlalu rumit [8].

Evaluasi Model

Kinerja model akan dievaluasi menggunakan metrik *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). MSE bertujuan mengukur rata-rata kuadrat kesalahan prediksi, sedangkan MAE untuk mengukur rata-rata kesalahan absolut. Kedua metrik ini memberikan pandangan yang komprehensif tentang seberapa baik model dapat memprediksi harga saham [15].

Validasi silang (*cross-validation*) juga akan dilakukan untuk memastikan bahwa model memiliki generalisasi yang baik pada data yang belum pernah dilihat sebelumnya. Dalam validasi silang, *dataset* dibagi menjadi beberapa *subset*, dan model dilatih dan diuji pada *subset* yang berbeda-beda untuk memastikan bahwa model dapat menangani variasi dalam data [7].

Implementasi

Model yang telah dilatih dan divalidasi akan diimplementasikan dalam sistem lokal yang dapat melakukan prediksi harga saham secara *real-time*. Prediksi ini akan dilakukan dengan menjalankan *script Python* di *Google Colab* untuk memanfaatkan kekuatan komputasi dan kemudahan akses [9]. Pengujian model dilakukan untuk memastikan bahwa semua fungsi bekerja dengan baik dan prediksi yang dihasilkan adalah akurat. Pengujian juga akan mencakup analisis hasil prediksi untuk memastikan kesesuaian dengan kondisi pasar saat ini [10].

3. HASIL DAN PEMBAHASAN

Hasil dan Pelatihan Model

Setelah melakukan pelatihan model menggunakan data historis harga saham Tesla dari tahun 2018 hingga 2023, model *Long Short Term Memory* (LSTM) dan *RandomForest* menunjukkan performa yang baik dalam memprediksi harga saham. Hasil pelatihan menunjukkan bahwa kedua model berhasil belajar pola harga saham dengan cukup akurat. Gambar 1 menunjukkan tren harga saham Tesla selama periode pelatihan dan prediksi.

Analisis Grafik Harga Saham Tesla

Gambar 1 menampilkan tren harga saham Tesla yang kompleks dan dinamis selama periode 2018-2023. Periode pertama (2018-2020) ditandai dengan harga saham yang stabil berkisar antara \$20-\$50, menggambarkan pertumbuhan lambat dan menjadi fase persiapan sebelum akselerasi. Fase kedua (2020-2021) menunjukkan pertumbuhan eksponensial yang mengagumkan, dengan harga saham melonjak dari \$50 menjadi \$250. Periode ini bertepatan dengan pandemi COVID-19 dan mengindikasikan resiliensi Tesla dalam menghadapi krisis global.

Fase volatilitas tinggi (2021-2022) mencapai puncaknya dengan harga saham mencapai \$400, ditandai oleh fluktuasi ekstrem yang mencerminkan ketidakpastian pasar dan sentimen investor yang tinggi. Periode ini merupakan masa spekulasi dan ekspektasi yang tinggi bagi perusahaan. Fase terakhir (2022-2023) menunjukkan koreksi signifikan, dengan harga turun dari \$400 menjadi \$150-\$200. Penurunan ini mengindikasikan konsolidasi pasar dan kemungkinan dipengaruhi oleh faktor makroekonomi serta dinamika internal perusahaan.

Kompleksitas tren saham Tesla membutuhkan model machine learning canggih. LSTM dan Random Forest terbukti mampu menangkap pola non-linier dan memiliki kemampuan adaptasi yang tinggi terhadap perubahan cepat. Kedua model berhasil mengekstraksi pola kompleks dari data historis, menunjukkan potensi signifikan dalam prediksi harga saham yang dinamis dan tidak pasti.



Gambar 1. Tren Harga Saham Tesla Selama Periode Pelatihan Dan Prediksi

Metrik Evaluasi

Tabel 2. Evaluasi MSE dan MAE

Model	MSE	MAE
LSTM	73325,258586	266,184550
Random Forest	76498,734178	270,855634

Hasil evaluasi menunjukkan bahwa model LSTM dan *Random Forest* berhasil memprediksi harga saham dengan akurasi yang baik. Tabel 2 menunjukkan bahwa model LSTM memiliki nilai MSE 73325,26 yang sedikit lebih rendah dibandingkan *Random Forest* yang mencapai 76498,73. Nilai MSE yang lebih rendah menunjukkan bahwa model memiliki kesalahan prediksi yang lebih kecil, sehingga LSTM lebih unggul dalam hal ini. Sementara itu, LSTM juga memiliki nilai MAE lebih kecil daripada *Random Forest* yaitu sebesar 266,18 (Tabel 2). Ini menandakan bahwa rata-rata kesalahan absolut pada prediksi yang dihasilkan oleh LSTM lebih rendah, yang berarti konsistensi hasil prediksi dari model lebih baik.

Tabel 3. Persentase Keberhasilan

Nama Model	Persentase Keberhasilan (%)
LSTM	93,292301%
Random Forest	94,771736%

Tabel 3 menunjukkan persentase keberhasilan model. *Random Forest* memiliki tingkat keberhasilan sedikit lebih besar yaitu 93,29%, dibandingkan LSTM yang memiliki tingkat

keberhasilan 94,77%. Meskipun LSTM memiliki tingkat keberhasilan yang lebih rendah, namun perbedaannya relatif kecil. Sehingga dapat disimpulkan bahwa kedua model memiliki tingkat performa yang sebanding.

Diskusi

Berdasarkan hasil evaluasi, LSTM maupun *Random Forest* dapat bekerja dengan cukup akurat dalam memprediksi harga saham. LSTM memiliki nilai MSE dan MAE lebih tinggi dan lebih unggul dalam menangani kesalahan prediksi dibandingkan *Random Forest*. Sedangkan *Random Forest* memiliki persentase keberhasilan lebih tinggi dibandingkan LSTM, ini menunjukkan *Random Forest* dapat memberikan prediksi yang lebih akurat pada beberapa kondisi tertentu.

Namun, prediksi harga saham selalu memiliki tingkat ketidakpastian karena dipengaruhi oleh berbagai faktor eksternal. Selain itu, model ini dapat ditingkatkan dengan menambahkan fitur. Oleh karena itu, untuk dapat meningkatkan performa model, fitur tambahan seperti volume perdagangan, indeks pasar, atau data fundamental lainnya dapat ditambahkan ke dalam model.

4. KESIMPULAN

Model LSTM dan Random Forest menghadapi keterbatasan signifikan dalam prediksi harga saham. Ketergantungan tinggi pada data historis, risiko *overfitting*, dan kompleksitas komputasional menjadi tantangan utama dalam mengembangkan model prediksi yang akurat dan *reliable*. Penelitian ini memberikan kontribusi signifikan dalam pengembangan metode prediksi saham berbasis *machine learning*. Kedua model menunjukkan akurasi di atas 93%, dengan Random Forest sedikit unggul, membuka peluang pengembangan lebih lanjut dalam prediksi keuangan. Temuan penelitian menunjukkan potensi machine learning sebagai alat *powerful* dalam memahami dinamika pasar saham yang kompleks. Rekomendasi bagi penelitian untuk kasus serupa dalam penelitian ini dimasa mendatang disarankan untuk menguji dataset saham yang lebih luas, mengintegrasikan sumber data eksternal, mengoptimasi arsitektur model, dan mengembangkan teknik regularisasi lanjutan.

DAFTAR PUSTAKA

- [1] K. A. Widiarto and E. W. Pamungkas, "Implementasi Algoritma LSTM untuk Prediksi Harga Saham pada Situs Yahoo Finance," *Universitas Muhammadiyah Surakarta*, 2024. [Online]. Available: <https://eprints.ums.ac.id/124206/1/Naskah%20Publikasi.pdf>.
- [2] C. A. F. Patra, "Sistem Analisa Harga Saham Menggunakan Algoritma Long Short Term Memory (LSTM)," *Sekolah Tinggi Manajemen Informatika & Komputer – LIKMI Bandung*, 2024. [Online]. Available: <https://library.liksi.ac.id/show/153/pdf>.
- [3] N. P. N. Kusuma et al., "Prediksi Harga Saham Blue Chip Menggunakan Algoritma Long Short-Term Memory (LSTM)," *Universitas Primakara*, 2024. [Online]. Available: <https://ejournalfb.ukdw.ac.id/index.php/jrak/article/download/6/6/31>.
- [4] A. Rosyd et al., "Penerapan Metode Long Short Term Memory (LSTM) dalam Memprediksi Harga Saham PT Bank Central Asia," *STMIK IKMI Cirebon*, 2024. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/download/8440/5012/>.
- [5] K. Kwanda et al., "Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website," *Universitas Tarumanagara*, 2024. [Online]. Available: <https://jurnal.ranahresearch.com/index.php/R2J/article/download/1255/1045>.
- [6] I. H. Witten, E. Frank, and M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, 2011. [Online]. Available: <https://www.elsevier.com/books/data-mining/witten/978-0-12-374856-0>
- [7] A. Agusta, I. Ernawati, and A. Muliawati, "Prediksi Pergerakan Harga Saham Pada Sektor Farmasi Menggunakan Algoritma Long Short-Term Memory," *Jurnal Ilmu Komputer dan Informasi*, vol. 17, no. 2, pp. 3651, 2024. [Online]. Tersedia: <https://dx.doi.org/10.52958/iftk.v17i2.3651>

- [8] Bhowmick, A., Rahman, A., & Rahman, R. M. (2019). Performance Analysis of Different Recurrent Neural Network Architectures and Classical Statistical Model for Financial Forecasting: A Case Study on Dhaka Stock Exchange (pp. 277–286). https://doi.org/10.1007/978-3-030-19810-7_27
- [9] A. G. Aleyusta, "LONG SHORT TERM MEMORY (VAR-LSTM) PADA PERAMALAN," *Digilib Unila*, 2023. [Online]. Tersedia: <http://digilib.unila.ac.id/82155/3/3.%20SKRIPSI%20TANPA%20BAB%20PEMBAHASAN%20-%20A.%20GILANG%20ALEYUSTA%20SAVADA.pdf>
- [10] K. Mauludin, "Model Prediksi Kecepatan Kurir dalam Pengiriman Surat Kabar," Eprints Unpak, 2023. [Online]. Tersedia: <https://eprints.unpak.ac.id/8115/1/Laporan%20-%20SKRIPSI%20-%2006512003%20-%20Kriti%20Mauludin.pdf>
- [11] YENG, Handyc; SIAHAAN, Mangapul. Perancangan Sistem Prediksi Harga Saham Berbasis Website Menggunakan Algoritma Hybrid (ARIMA-LSTM). *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, [S.I.], v. 13, n. 1, p. 61-72, apr. 2024. ISSN 2685-0893. Available at: <<https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/view/1620>>. Date accessed: 27 Mar. 2025. doi:<http://dx.doi.org/10.35889/jutisi.v13i1.1620>.
- [12] Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). DOI: <https://doi.org/10.1023/A:1010933404324>
- [13] M. Ridwan and A. Muzakir, "Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM," *Teknomatika*, vol. 12, no. 02, pp. 209-218, Sep. 2022, <https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/604>
- [14] F. Darnis, A. Rahman, dan Y. Ansori, "Akurasi Model Algoritma Recurrent Neural Network Untuk Prediksi Saham Syariah," *Teknomatika*, vol. 15, no. 01, pp. 17-24, Apr. 2025, <https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/690>
- [15] A. Haris, "Sistem Pendukung Keputusan Pemberian Bonus Dosen Tetap Memanfaatkan Support Vector Regression," *Teknomatika*, vol. 8, no. 1, pp. , Jul. 2018. <https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/101>