

A COMPARATIVE STUDY OF ARIMA, XGBOOST, AND HYBRID ARIMA–XGBOOST APPROACHES FOR FORECASTING IT PROJECT DEMAND

ANALISIS KOMPARATIF MODEL ARIMA, XGBOOST, DAN PENDEKATAN HYBRID ARIMA-XGBOOST UNTUK PREDIKSI PERMINTAAN PROYEK IT

Muhamad Hafidz¹, Esa Fauzi²

Universitas Widyatama^{1,2}

muhamad.hafidz@widyatama.ac.id¹, esa.fauzi@widyatama.ac.id²

ABSTRACT

From the end of 2022 to the end of 2024, the software development industry has experienced a significant decline in IT project demand, leading to challenges in resource planning and business strategy formulation. This study aims to develop a predictive model for IT project demand through a comparative analysis of three approaches: the statistical ARIMA (AutoRegressive Integrated Moving Average) model, the machine learning XGBoost (Extreme Gradient Boosting) model, and a Hybrid approach that combines the strengths of both. The data were obtained from Clutch.co, a global platform that provides information and reviews on IT service providers. Model performance was evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The evaluation was conducted on a test dataset to assess each model's ability to handle previously unseen data. The results show that the XGBoost model achieved the best performance, with an MAE of 1.5574 and an RMSE of 2.3597, outperforming both ARIMA and the Hybrid ARIMA-XGBoost models.

Keyword: *IT Project Demand, Forecasting, ARIMA, XGBoost, Comparative Analysis.*

ABSTRAK

Sejak akhir tahun 2022 hingga akhir tahun 2024, industri pengembangan perangkat lunak mengalami penurunan signifikan dalam permintaan proyek IT, yang berdampak pada kesulitan dalam perencanaan sumber daya dan penyusunan strategi bisnis. Penelitian ini bertujuan untuk mengembangkan model prediksi permintaan proyek IT melalui analisis komparatif terhadap tiga pendekatan, yaitu model statistik ARIMA (AutoRegressive Integrated Moving Average), model machine learning XGBoost (Extreme Gradient Boosting), dan pendekatan Hybrid yang menggabungkan keunggulan keduanya. Data diperoleh dari Clutch.co, sebuah platform global yang menyediakan informasi dan ulasan mengenai penyedia jasa IT. Evaluasi performa masing-masing model dilakukan dengan menggunakan metrik evaluasi seperti *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Hasil penelitian yang dilakukan terhadap data test set, untuk menilai kemampuan model dalam menghadapi data yang benar-benar belum terlihat sebelumnya. Hasilnya, model XGBoost menunjukkan performa terbaik dengan MAE sebesar 1.5574 dan RMSE sebesar 2.3597, mengungguli ARIMA dan Hybrid ARIMA-XGboost.

Kata Kunci: *Permintaan Proyek IT, Peramalan, ARIMA, XGBoost, Analisis Komparatif.*

PENDAHULUAN

Industri pengembangan perangkat lunak merupakan sektor yang sangat dinamis dan bergantung pada tren teknologi serta kebutuhan pasar. Namun, sejak akhir tahun 2022 hingga akhir 2024, terjadi penurunan signifikan dalam permintaan proyek IT di berbagai perusahaan pengembang perangkat lunak. Kondisi ini menciptakan ketidakpastian dalam proses perencanaan sumber daya, alokasi anggaran, serta penyusunan strategi bisnis yang berkelanjutan.

Penurunan ini dipengaruhi oleh berbagai faktor, seperti perubahan kebutuhan klien, fluktuasi ekonomi global, serta meningkatnya persaingan di antara penyedia layanan teknologi. Dalam menghadapi situasi tersebut, kemampuan untuk memprediksi permintaan proyek IT menjadi semakin penting.

Prediksi yang akurat dapat membantu perusahaan dalam menyusun strategi pemasaran yang tepat, mengelola kapasitas tim pengembang secara efisien, serta meminimalkan risiko keterlambatan

atau kekurangan proyek. Oleh karena itu, diperlukan suatu pendekatan analitis yang mampu memproyeksikan tren permintaan di masa mendatang secara andal.

Prediksi atau peramalan (*forecasting*) adalah suatu perhitungan untuk meramalkan masa depan melalui pengujian keadaan dimasa lalu (Azlina Putri, 2021). Berbagai metode prediksi telah dikembangkan dalam bidang ilmu data dan statistik. Model statistik seperti ARIMA (*AutoRegressive Integrated Moving Average*) banyak digunakan dalam analisis deret waktu karena kemampuannya dalam menangkap pola historis dan tren data (Zidan Rusminto dkk., 2024). Model ARIMA tetap menjadi salah satu metode statistik yang andal dalam prediksi deret waktu (Rahman & Taqwa, 2025).

Dalam *machine learning* terdapat teknik klasifikasi yang digunakan untuk membuat prediksi dan mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan karakteristik atau ciri-ciri yang sama. Terdapat banyak metode klasifikasi dalam *machine learning* salah satunya yaitu XGBoost (Dava Maulana dkk., 2023). XGBoost (*Extreme Gradient Boosting*) ini telah banyak digunakan dalam berbagai skenario penambahan data serta kompetisi algoritma (Zhang Lingyu dkk., 2021). Selain itu, XGBoost menawarkan fleksibilitas dalam menangani data yang kompleks serta keunggulan dalam akurasi prediksi dalam *data science* (Huang dkk., 2025).

Selain itu, pendekatan *hybrid* yang menggabungkan elemen dari model statistik dan *machine learning* juga mulai banyak diterapkan untuk meningkatkan performa prediktif. *Hybrid model* dapat digunakan untuk menggabungkan model yang lebih baik dalam menangani data yang kompleks (Fikri Alfawaid dkk., 2023).

Studi lain menunjukkan bahwa pendekatan *hybrid* dapat meningkatkan akurasi prediksi permintaan dalam berbagai konteks. Seperti Studi yang

dilakukan oleh (Liang, 2024) mengembangkan model *hybrid* yang menggabungkan ARIMA dengan *Attention-based CNN-LSTM* dan XGBoost untuk prediksi harga saham di pasar AS, yang menunjukkan performa baik dalam kondisi volatilitas tinggi.

Studi lainnya yang dilakukan oleh (Atesongun & Gulsen, 2024) mengusulkan struktur *hybrid* yang mengintegrasikan ARIMA dan *Artificial Neural Network* (ANN) untuk meningkatkan akurasi prediksi pada data dengan pola kompleks.

Penelitian lainnya yang dilakukan oleh (Quiñones-Rivera dkk., 2023) membandingkan model ARMAX dengan model *hybrid* NN-ARMAX dalam konteks permintaan produk listrik, dan menemukan bahwa model *hybrid* memberikan hasil yang lebih akurat.

Studi Lainnya yang dilakukan oleh (Suddala, 2024) mengembangkan arsitektur *hybrid* ARIMA-LSTM untuk prediksi permintaan dalam rantai pasok, yang terbukti lebih unggul dibandingkan model tunggal.

Penelitian lainnya yang dilakukan oleh (Mei dkk., 2024) menerapkan model *hybrid* ARIMA-XGBoost untuk strategi penetapan harga dan pengisian ulang produk yang mudah rusak, yang menghasilkan keputusan yang lebih akurat dalam konteks tersebut.

Dalam penelitian ini, dilakukan analisis komparatif terhadap tiga pendekatan yaitu ARIMA, XGboost dan *hybrid* ARIMA-XGBoost untuk memprediksi permintaan proyek IT.

Data yang digunakan bersumber dari Clutch.co. Clutch.co adalah platform direktori B2B yang komprehensif yang menghubungkan bisnis dengan penyedia layanan yang tepat. Platform ini dirancang untuk perusahaan dari berbagai ukuran, dengan menyediakan basis data luas berisi ulasan klien yang telah diverifikasi, profil agensi yang mendetail, serta penilaian yang transparan. Keunggulan Clutch terletak pada kemampuannya membantu para pendiri startup menyaring informasi

yang tidak relevan, sehingga lebih mudah menemukan mitra terpercaya di berbagai industri seperti pengembangan perangkat lunak, pemasaran, dan desain (Sinan Sari, 2025).

Pengambilan data dari clutch.co menggunakan *Instant Data Scraper*. *Instant Data Scraper* adalah ekstensi browser yang mendeteksi struktur tabel/daftar HTML dan mengekstrak data ke dalam file Excel atau CSV tanpa perlu *scripting*. *Instant Data Scraper* menggunakan AI (*Artificial Intelligence*) untuk memprediksi data mana yang paling relevan pada halaman HTML dan memungkinkan penyimpanannya ke dalam file Excel atau CSV. Alat ini tidak memerlukan skrip khusus untuk setiap situs web, melainkan menggunakan analisis AI heuristik terhadap struktur HTML untuk mendeteksi data yang akan diekstrak. Jika prediksinya kurang memuaskan, pengguna dapat menyesuaikan pilihan secara manual untuk mendapatkan akurasi yang lebih tinggi. Teknologi *scraping* seperti ini jauh lebih praktis karena tidak membutuhkan pustaka skrip *scraping* buatan pengguna yang besar, yang sering kali berisi versi usang dan redundan. Ini berarti metode *scraping* bekerja sama baiknya pada situs kecil dan kurang dikenal, maupun pada raksasa global seperti Amazon. Selain itu, pengguna tidak perlu memiliki keahlian dalam pemrograman (Paulius, 2025).

Mean Absolute Error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. *Mean absolute error* digunakan untuk mengukur rata-rata nilai *absolute* dari kesalahan antara nilai yang diprediksi dengan nilai yang sebenarnya (Parera & Yulian, 2024).

Root Mean Squared Error (RMSE) adalah salah satu metrik evaluasi yang digunakan untuk mengukur rata-rata perbedaan antara nilai prediksi dan nilai actual dalam bentuk kuadrat, yang dimana semakin rendah nilai RMSE maka semakin akurat model algoritma tersebut

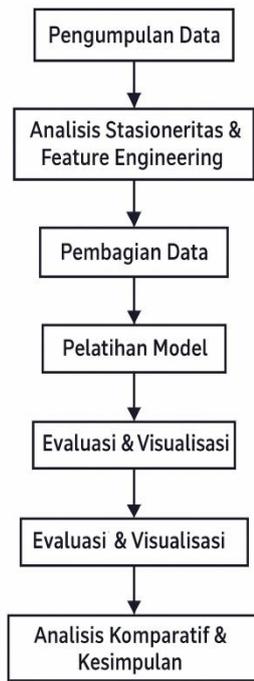
dalam memprediksi (Aji & Idifitriani, 2024).

Model-model prediktif yang digunakan akan dievaluasi menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) untuk menentukan model dengan performa terbaik.

Penelitian ini diharapkan dapat memberikan kontribusi praktis dalam mendukung pengambilan keputusan strategis, terutama dalam konteks perencanaan proyek dan pengembangan bisnis di sektor teknologi informasi. Dengan adanya model prediksi yang andal, perusahaan dapat lebih siap dalam mengantisipasi dinamika pasar dan menjaga keberlanjutan operasional secara lebih terukur.

METODE

Penelitian ini bersifat kuantitatif dengan metode komparatif analisis untuk membandingkan kinerja tiga model prediksi deret waktu: ARIMA, XGBoost, dan *Hybrid* ARIMA-XGBoost. Pendekatan ini meliputi serangkaian tahapan sistematis yang digambarkan dalam kerangka kerja penelitian seperti yang ditunjukkan pada Gambar 1. Setiap tahapan dirancang agar hasil prediksi objektif, *reproducible*, dan dapat diinterpretasi.



Gambar 1. Tahapan Penelitian

Desain dan Alur Proses Penelitian

Desain penelitian diuraikan menjadi delapan tahapan utama. Berikut deskripsi setiap proses:

1. Pengumpulan Data

Data dikumpulkan dari platform Clutch.co menggunakan ekstensi *Instant Data Scraper*, yang secara otomatis mengekstraksi informasi terstruktur dari halaman web dan menyimpannya dalam format CSV. Hasil keluaran terdiri atas file CSV yang memuat lima atribut utama, yakni:

- `project_start_date` (tanggal dimulainya proyek)
- `service_type` (jenis layanan yang diberikan)
- `client_profile` (profil klien)
- `rating` (penilaian numerik dari klien)

Dengan format ini, data siap dianalisis lebih lanjut untuk keperluan pengolahan deret waktu, pemodelan prediksi, atau visualisasi tren layanan dan kepuasan klien.

2. Pra-pemrosesan Data

Proses pra-pemrosesan data dimulai dengan tahap *cleaning*, di mana duplikat

dihilangkan untuk menjaga keunikan setiap catatan, nilai hilang diatasi melalui metode imputasi atau interpolasi sesuai karakteristik variabel, dan *outlier* difilter menggunakan rentang interkuartil (IQR) demi mencegah distorsi analisis.

Selanjutnya, pada tahap transformasi, kolom `project_start_date` dikonversi ke format `datetime` agar memungkinkan operasi waktu, sedangkan variabel kategori (`service_type`) diubah menjadi format numerik, baik melalui *one-hot encoding* untuk mempertahankan independensi antar kategori, maupun *label encoding* jika urutan kategori diperlukan. Terakhir, tahap normalisasi dilakukan dengan menerapkan `StandardScaler` pada seluruh fitur numerik, sehingga setiap variabel memiliki *mean* nol dan *varians* satu, memastikan kestabilan dan kesetaraan skala dalam pemodelan selanjutnya.

3. Analisis Stasioneritas & Feature Engineering

Analisis Stasioneritas dilakukan menggunakan Uji *Augmented Dickey-Fuller* (ADF) terhadap setiap variabel deret waktu untuk mengidentifikasi apakah data bersifat stasioner atau tidak. Jika hasil uji menunjukkan non-stasioneritas, maka proses differencing dilakukan secara bertahap (nilai `d`) hingga deret waktu mencapai kondisi stasioner, yang merupakan prasyarat penting dalam pemodelan ARIMA dan beberapa algoritma deret waktu lainnya.

Feature Engineering mencakup beberapa teknik penting untuk menangkap pola *temporal* dan kontekstual. Pertama, dibuat lag features dari lag₁ hingga lag₁₂, yang merepresentasikan nilai data pada bulan-bulan sebelumnya dan membantu model mengenali pola musiman maupun jangka pendek. Kedua, dihitung *rolling statistics* berupa rata-rata (*mean*) dan simpangan baku (*standard deviation*) pada jendela waktu 3, 6, dan 12 bulan untuk mengungkap dinamika lokal dalam data. Selain itu, variabel eksternal seperti

service_type, rating, dan client_profile diikutsertakan sebagai faktor tambahan.

Variabel-variabel ini kemudian dikodekan menggunakan teknik *One-Hot Encoding* atau *Label Encoding*, tergantung pada kebutuhan algoritma, untuk memastikan dapat digunakan dalam model prediktif. Dengan kombinasi antara stasioneritas yang terjamin dan fitur yang direkayasa secara informatif, data menjadi lebih siap untuk proses pelatihan model prediksi yang akurat dan andal.

4. Pembagian Data

Pembagian data dilakukan menggunakan skema *time-based split*, yang mempertahankan urutan kronologis data sesuai dengan sifat alami deret waktu. Metode ini dipilih agar model hanya dilatih pada data historis dan dievaluasi pada periode yang benar-benar belum terlihat sebelumnya, sehingga menghindari data leakage dan mencerminkan kondisi prediksi nyata di dunia nyata.

Dataset dibagi menjadi tiga bagian utama:

- *Training set* mencakup data dari Januari 2019 hingga Desember 2022 (sekitar 70% dari keseluruhan data). Bagian ini digunakan untuk melatih model dan menangkap pola historis dalam data.
- *Validation set* mencakup periode Januari 2023 hingga Juni 2023 (15%), digunakan untuk menyetel parameter model dan melakukan *tuning* terhadap *hyperparameter* secara objektif tanpa menyentuh data uji.
- *Testing set* terdiri dari data Juli 2023 hingga Desember 2024 (15%), berfungsi untuk mengevaluasi performa akhir model pada data yang sepenuhnya belum digunakan selama pelatihan maupun validasi.

Dengan pembagian berdasarkan waktu ini, model yang dikembangkan diuji ketangguhannya dalam memprediksi tren masa depan berdasarkan pola masa lalu yang realistis.

5. Pelatihan Model

Pelatihan model dilakukan untuk tiga pendekatan: ARIMA, XGBoost, dan *Hybrid ARIMA-XGBoost*.

Untuk ARIMA, parameter (p , d , q) ditentukan menggunakan plot ACF dan PACF, lalu disempurnakan melalui *grid search*. Model ini dilatih menggunakan data dari *training set*.

Model XGBoost dilatih dengan input berupa *lag features* dan variabel eksternal seperti service_type, rating, dan client_profile. Proses tuning dilakukan untuk parameter seperti learning_rate, max_depth, dan n_estimators.

Pada model *hybrid ARIMA-XGBoost*, ARIMA digunakan untuk membuat prediksi awal. Selisih antara nilai aktual dan prediksi ARIMA (*residual*) kemudian digunakan sebagai target untuk melatih XGBoost. Hasil akhir diperoleh dengan menjumlahkan prediksi ARIMA dan prediksi residual dari XGBoost.

6. Validasi Silang

Validasi model dilakukan menggunakan metode *rolling-origin cross-validation*, yang dirancang khusus untuk data deret waktu agar urutan temporal tetap terjaga.

Prosedurnya dimulai dengan menentukan jumlah *fold* (k). Untuk setiap *fold* ke- i , model dilatih pada data hingga periode ke- i dan diuji pada periode ke- $i+1$. Proses ini diulang hingga seluruh *fold* tercakup.

Setelah semua iterasi selesai, metrik evaluasi seperti *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) dari masing-masing *fold* diakumulasi dan dirata-ratakan untuk memberikan gambaran performa model yang lebih stabil dan terpercaya di berbagai kondisi waktu.

7. Evaluasi dan Visualisasi

Evaluasi model dilakukan menggunakan dua metrik utama, yaitu *Mean Absolute Error* (MAE) dan *Root*

Mean Squared Error (RMSE) yang lebih sensitif terhadap kesalahan besar.

Untuk membantu interpretasi hasil, dilakukan visualisasi berupa plot deret waktu yang membandingkan nilai aktual dan prediksi dari masing-masing model, serta grafik perbandingan nilai *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) untuk menilai performa relatif antar model secara visual dan intuitif.

HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil analisis yang diperoleh dari penerapan model prediksi terhadap data permintaan proyek IT. Proses evaluasi dilakukan terhadap tiga pendekatan model, yaitu ARIMA, XGBoost, dan *hybrid* ARIMA–XGBoost. Evaluasi dilakukan dalam tiga tahap utama, yakni validasi awal, validasi silang dengan teknik *rolling-origin cross-validation* (ROCV), serta pengujian akhir menggunakan data test set.

Hasil Model Validasi Awal

Pada tahap validasi awal, model dilatih menggunakan data hingga Desember 2022 dan dievaluasi terhadap periode Januari hingga Juni 2023. Evaluasi dilakukan dengan mengukur *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) sebagai indikator utama akurasi.

Tabel 1. Hasil Validasi Model pada Validation Set

Model	MAE	RMSE
ARIMA	1.8531	2.6730
XGBoost	2.1921	2.6706
Hybrid ARIMA-XGBoost	4.1636	5.1549

Hasil menunjukkan bahwa model ARIMA menghasilkan MAE sebesar 1.8531 dan RMSE sebesar 2.6730, sedikit lebih baik dibandingkan model XGBoost yang mencatat MAE sebesar 2.1921 dan RMSE 2.6706. Sebaliknya, pendekatan hybrid menunjukkan performa terburuk dengan MAE sebesar 4.1636 dan RMSE 5.1549.

Fenomena ini mengindikasikan bahwa pola linier dalam data masih cukup dominan, sehingga ARIMA mampu memodelkannya dengan baik. Namun, performa XGBoost yang hampir setara dengan ARIMA menandakan adanya komponen non-linier yang mulai signifikan. Di sisi lain, model hybrid yang menggabungkan prediksi ARIMA dan residual dari XGBoost belum berhasil mengoptimalkan keunggulan kedua pendekatan, kemungkinan karena *overfitting* pada proses pemodelan residual.

Hasil Validasi Silang

Untuk mengevaluasi kestabilan performa model dalam kondisi waktu yang berbeda, dilakukan validasi silang dengan metode *rolling-origin cross-validation*. Evaluasi ini memberikan wawasan lebih komprehensif terhadap kemampuan generalisasi model.

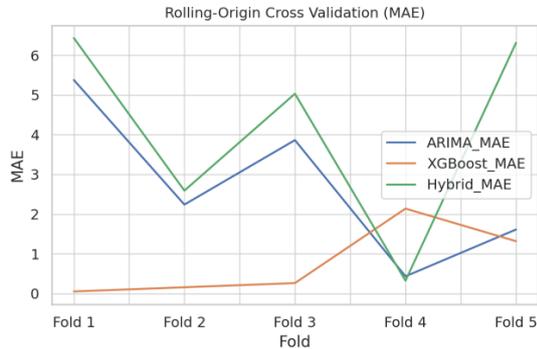
Tabel 2. MAE per Fold pada ROCV

Fold	ARIMA MAE	XGBoost MAE	Hybrid MAE
1	5.3787	0.0526	6.4351
2	2.2398	0.1584	2.5921
3	3.8637	0.2631	5.0319
4	0.4338	2.1388	0.3234
5	1.6123	1.3180	6.3147

Model XGBoost menunjukkan performa paling stabil dan akurat dengan MAE sangat rendah pada sebagian besar fold (misalnya, MAE 0.0526 pada fold pertama). Rata-rata performanya lebih unggul dibandingkan ARIMA dan *hybrid*, yang masing-masing menunjukkan fluktuasi signifikan antar *fold*. Hal ini menunjukkan bahwa XGBoost mampu beradaptasi lebih baik terhadap variasi data dan menangkap pola kompleks yang mungkin diabaikan oleh model statistik tradisional.

Temuan menarik terdapat pada fold keempat, di mana model ARIMA dan *hybrid* menunjukkan MAE yang sangat rendah (masing-masing 0.4338 dan 0.3234), sedangkan XGBoost mencatat

MAE lebih tinggi sebesar 2.1388. Ini menunjukkan bahwa dalam situasi tertentu, model linier atau kombinasi linier-nonlinier dapat mengungguli model machine learning murni, terutama jika pola data bersifat lebih konsisten dan historis.



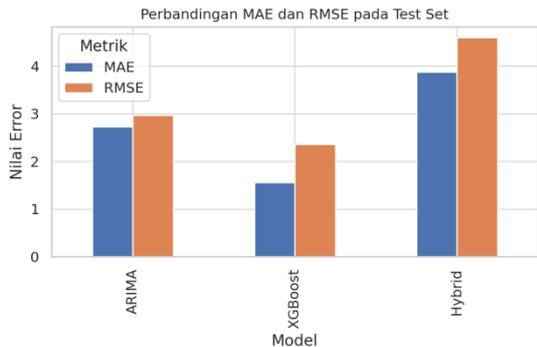
Gambar 1. Grafik Rolling-Origini Cross Validation MAE

Hasil Model Test Set

Tabel 3. Hasil Model Test Set

Model	MAE	RMSE
ARIMA	1.8531	2.6730
XGBoost	2.1921	2.6706
Hybrid ARIMA-XGBoost	4.1636	5.1549

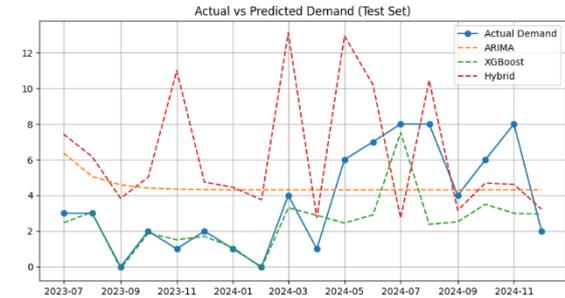
Evaluasi akhir dilakukan terhadap data test set (Juli 2023 – Desember 2024), untuk menilai kemampuan model dalam menghadapi data yang benar-benar belum terlihat sebelumnya. Hasilnya, model XGBoost menunjukkan performa terbaik dengan MAE sebesar 1.5574 dan RMSE sebesar 2.3597, mengungguli ARIMA dan Hybrid ARIMA-XGboost.



Gambar 2. Grafik Perbandingan MAE dan RMSE pada Test Set

Hasil ini mengonfirmasi bahwa model XGBoost tidak hanya unggul dalam proses validasi, tetapi juga mampu

memberikan generalisasi yang lebih baik terhadap data masa depan. Keunggulan ini memperkuat posisi XGBoost sebagai model yang lebih adaptif dan fleksibel terhadap karakteristik deret waktu yang kompleks dan non-linier.



Gambar 3. Grafik plot aktual vs Prediksi Permintaan Proyek IT

Grafik ini menampilkan perbandingan antara nilai aktual permintaan proyek IT dengan hasil prediksi dari tiga model: ARIMA, XGBoost, dan Hybrid ARIMA-XGBoost selama periode Juli 2023 hingga Desember 2024 (test set).

Garis biru dengan titik bulat menunjukkan nilai permintaan aktual setiap bulan. Kita dapat melihat bahwa permintaan bersifat fluktuatif, dengan tren meningkat mulai awal 2024.

Garis oranye putus-putus (ARIMA) cenderung datar dan tidak banyak mengikuti perubahan tren permintaan aktual. Hal ini menunjukkan bahwa model ARIMA tidak cukup adaptif terhadap variasi data terbaru, karena fokus utamanya adalah pola linier dan historis.

Garis hijau putus-putus (XGBoost) lebih mampu mengikuti pola fluktuasi aktual. Model ini secara konsisten memprediksi dengan nilai yang lebih dekat ke permintaan nyata, terutama di paruh kedua tahun 2024. Ini menandakan bahwa XGBoost menangkap pola musiman dan non-linier lebih baik.

Garis merah putus-putus (Hybrid) menunjukkan ketidakkonsistenan dan deviasi besar terhadap nilai aktual. Prediksi dari model hybrid mengalami lonjakan dan penurunan tajam yang tidak sesuai dengan pola data sebenarnya.

Hal ini mengindikasikan bahwa kombinasi residual ARIMA dengan XGBoost dalam model hybrid belum optimal, bahkan cenderung menghasilkan prediksi yang tidak stabil.

SIMPULAN

Penelitian ini mengungkapkan bahwa model *Extreme Gradient Boosting* (XGBoost) menunjukkan performa terbaik dalam memprediksi permintaan proyek teknologi informasi, melampaui akurasi model ARIMA maupun pendekatan hybrid ARIMA-XGBoost. Kemampuannya dalam menangkap pola *nonlinier* yang kompleks menjadikan XGBoost lebih unggul dalam menghadapi dinamika pasar yang cepat berubah.

Model ARIMA tetap relevan untuk data berpola linier, sementara pendekatan *hybrid* yang menggabungkan ARIMA dan XGBoost bertujuan memanfaatkan keunggulan keduanya. Namun, hasil yang diperoleh menunjukkan bahwa performa hybrid belum optimal. Hal ini kemungkinan disebabkan oleh kurangnya penyetelan parameter yang tepat atau strategi integrasi model yang belum maksimal. Dengan demikian, potensi peningkatan akurasi pada model *hybrid* masih terbuka, terutama melalui teknik *tuning hyperparameter* dan pendekatan optimasi yang lebih cermat.

Temuan ini diperkuat melalui *rolling-origin cross-validation*, yang menunjukkan stabilitas performa XGBoost di berbagai horizon waktu. Dengan demikian, model *machine learning* seperti XGBoost terbukti lebih adaptif dan akurat untuk prediksi permintaan proyek IT dalam konteks pasar yang fluktuatif. Hasil ini memberikan kontribusi praktis bagi pengambilan keputusan strategis serta referensi metodologis bagi penelitian lanjutan di bidang peramalan deret waktu.

DAFTAR PUSTAKA

Azlina Putri, A. (2021). Resolusi: Rekayasa Teknik Informatika dan Informasi Penerapan Data Mining

Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus: PT. Central Brastagi Utama). *Media Online*, 1(6), 354–361.

Zidan Rusminto, M., Adi Wibowo, S., & Santi Wahyuni, F. (2024). Peramalan Harga Saham Menggunakan Metode Arima (*Autoregressive Integrated Moving Average*) Time Series. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2).

Rahman Abdul Aziz, A., & Taqwa Prasetyaningrum, P. (2025). Analisis Perbandingan Model ARIMA dan *Exponential Smoothing* dalam Meramalkan Harga Penutupan Saham. *Technology and Science (BITS)*, 7(1).

Dava Maulana, M., Id Hadiana, A., Rakhmat Umbara Informatika, F., Jenderal Achmad Yani Cimahi Jl Terusan Jend Sudirman, U., Cimahi Sel, K., Cimahi, K., & Barat, J. (2023). Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 5).

Zhang, L., Bian, W., Qu, W., Tuo, L., & Wang, Y. (2021). Time series forecast of sales volume based on XGBoost. *Journal of Physics: Conference Series*, 1873(1).

Huang, X., Zhuang, X., Tian, F., Niu, Z., Chen, Y., Zhou, Q., & Yuan, C. (2025). A Hybrid ARIMA-LSTM-XGBoost Model with Linear Regression Stacking for Transformer Oil Temperature Prediction. *Energies*, 18(6).

Fikri Alfawaid, M., Prianggono, J., Studi, P. S., Kepolisian, I., & Tinggi Ilmu Kepolisian, S. (2023). Penerapan Metode *Multiple Machine Learning (Hybrid Model)* Untuk Mendeteksi Link Phishing Sebagai Upaya Preventif Dalam Meminimalisir Korban Pencurian Data. *Jurnal Ilmu Kepolisian*, 17.

- Liang, L. (2024). *ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction in the US stock market. SHS Web of Conferences*, 196, 02001.
- Atesongun, A., & Gulsen, M. (2024). *A Hybrid Forecasting Structure Based on Arima and Artificial Neural Network Models. Applied Sciences (Switzerland)*, 14(16).
- Quiñones-Rivera, H., Rubiano-Ovalle, O., & Alfonso-Morales, W. (2023). *Demand Forecasting Using a Hybrid Model Based on Artificial Neural Networks: A Study Case on Electrical Products. Journal of Industrial Engineering and Management*, 16(2), 363–381.
- Suddala, S. (2024). *Dynamic Demand Forecasting In Supply Chains using Hybrid Arima-Lstm Architectures. International Journal of Advanced Research*, 12(10), 1167–1171.
- Mei, Y., Yang, Y., Chen, H., Shi, W., & Su, X. (2024). *ARIMA-Xgboost Based Pricing and Replenishment Strategy for Perishable Goods. In Highlights in Science, Engineering and Technology MMACS (Vol. 2024)*.
- Sinan Sari. (2025). Clutch. diakses pada 19 juni 2025, Pukul 15.00 dari <https://www.horizon-labs.co/resources/clutch>.
- Paulius Jonaitis. (2025). *Instant Data Scraper Overview*. diakses pada 19 juni 2025, Pukul 15.30 dari <https://chromewebstore.google.com/detail/instant-data-scraper/ofaokhiedipichpaobibbnahkdoiiah>
- Parera Agrippina, A., & Yulian Pamuji, F. (2024). *Komparasi Peramalan Penerimaan Siswa Baru Menggunakan Metode Exponential Smoothing. Jurnal PROSISKO*, 11 No.1. Maret 2024.
- Aji, M. B., & Idifitriani, F. (2024). *Analisis Perbandingan RMSE Algoritma Machine Learning dalam Memprediksi Harga Saham. ICOM, Jurnal Informatika Dan Teknologi Komputer*, Vol. 05 No. 02, 113–119.