

Analisis Prediktif untuk Keputusan Bisnis : Peramalan Penjualan

Fachrul Rozi Lubis¹, Muhammad Khoiruddin Harahap^{2*}, Amir Mahmud Husein³

¹Universitas Harapan Medan, Fakultas Teknik dan Komputer, Sistem Informasi, Medan

^{2*}Politeknik Ganesha Medan, Medan

³Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika, Medan

^{1*}fachrul_rozi.unhar@harapan.ac.id

Received: xx xx xxxx

Accepted: 9 May 2019

Published: 14 May 2019



***Muhammad Khoiruddin Harahap**

Keywords: Peramalan Penjualan, Analisis Preditif, Data Science, CRISP-DM, Exploratory Data Analysis, XGBoost.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Peramalan penjualan produk adalah aspek utama dari manajemen pembelian, persediaan yang melebihi permintaan atau kekurangan akan berdampak pada manajemen pelayanan maupun secara ekominis. Makalah ini fokus mencoba menyajikan penerapan analisis prediktif dengan mengadopsi kerangka kerja Data Science (ilmu data) untuk menemukan wawasan yang berguna dalam pengambilan keputusan bisnis khususnya tentang peramalan penjualan produk di masa depan. Kerangka CRISP-DM diusulkan dengan tahapan pemahasaan bisnis, pemahaman dan persiapan data, *exploratory data analysis* (EDA) dan pemodelan. Berdasarkan hasil pengujian data penjualan yang dievaluasi berdasarkan RMSE dan MAE, algoritma XGBoost menghasilkan prediksi berada dalam 1,3% kemudian ARIMA sebesar 1.6%, masih lebih baik dibandingkan LinearRegression, RandomForest dan LSTM dengan tingkat kesalahan sebesar 1.81%, 1.97%, 2.21% pada masing-masing algoritma dari data aktual.

PENDAHULUAN

Analisis prediktif adalah penggunaan analisis data, pemodelan statistik, dan teknologi pembelajaran mesin untuk memprediksi kemungkinan hasil [1]. Analisis ini mencakup berbagai macam teknik dan teknologi statistik. Dua jenis model analitik prediktif yang paling umum adalah model klasifikasi dan regresi [2]. Model klasifikasi membantu memprediksi hasil seperti ketika komponen akan gagal [3]. Model regresi yang membantu memprediksi angka, seperti waktu rata-rata sebelum kerusakan. Analisis prediktif semakin banyak digunakan oleh berbagai organisasi di banyak sektor, seperti rumah sakit [4]–[7], ritel [8], [9], manufaktur [10], layanan keuangan dan lainnya [11].

Pada umumnya, perusahaan retail memiliki banyak data transaksi. Melalui penggunaan data historis dan pola data transaksional, risiko diidentifikasi selain mengeksplorasi peluang untuk kebutuhan masa depan [12]. penggunaan analisis prediktif, bisnis dapat secara efektif menginterpretasikan tanggal yang tersedia bagi mereka untuk keuntungan mereka. Membangun model peramalan yang berguna untuk produk permintaan yang jarang adalah suatu tantangan. Jika perkiraan lebih rendah dari permintaan sebenarnya, hal itu dapat menyebabkan keputusan pemilihan dan pengisian ulang yang buruk, dan pelanggan tidak akan bisa mendapatkan produk yang mereka inginkan saat mereka membutuhkannya [13]. Jika perkiraan lebih tinggi dari permintaan aktual, produk yang tidak terjual akan menempati rak inventaris, dan jika produk mudah rusak, produk tersebut harus dilikuidasi dengan biaya rendah untuk mencegah pembusukan [14].

Peramalan penjualan produk adalah aspek utama dari manajemen pembelian. Prakiraan sangat penting dalam menentukan tingkat persediaan dan secara akurat memperkirakan permintaan barang di masa depan telah menjadi tantangan yang berkelanjutan, terutama di industri ritel [15]. Jika barang tidak tersedia atau

ketersediaan barang lebih dari permintaan, keuntungan keseluruhan dapat dikompromikan. Akibatnya, peramalan penjualan barang dapat menjadi signifikan untuk meminimalkan tingkat kerugian. Selain itu, memastikan persediaan sesuai dengan kebutuhan dan lokasi yang tersedia memerlukan beberapa tahapan analisis yang dapat membantu dalam pengambilan keputusan bisnis di masa depan.

Dalam makalah ini, analisis prediktif diusulkan untuk membuat model peramalan menggunakan algoritma pembelajaran mesin untuk membantu dalam menentukan keputusan bisnis dalam meningkatkan perkiraan penjualan produk yang akurat. Model yang diusulkan secara khusus ditargetkan untuk mendukung pembelian di masa depan. Sebuah model berdasarkan data toko nyata dikembangkan untuk memvalidasi penggunaan berbagai algoritma pembelajaran mesin. Dalam studi kasus, metode regresi berganda dibandingkan. Metode berdampak pada perkiraan ketersediaan produk di toko untuk memastikan mereka memiliki produk yang cukup pada waktu yang tepat. Hasil temuan ini berguna untuk menciptakan proses pengambilan keputusan yang berjalan lancar bagi pengguna dan menghasilkan wawasan baru yang mengarah pada tindakan yang lebih baik.

TINJAUAN LITERATUR

Peramalan memiliki dampak yang signifikan pada proses pengambilan keputusan bisnis, memberikan pemahaman yang lebih baik tentang posisi keuangan untuk mengelola tenaga kerja, meningkatkan manajemen rantai pasokan di masa depan. Peramalan memungkinkan membantu perencanaan inventaris, harga kompetitif, dan strategi promosi tepat waktu [16], [17]. Peramalan penjualan dapat membantu proses memahami siklus hidup bisnis dalam hal penjualan dan dampak sasaran produk jangka pendek seperti pertumbuhan, stabilitas, penurunan, promosi maupun harga [18].

Menurut penelitian yang dilakukan oleh [17], mereka mengusulkan pendekatan *Convolutional Neural Network* (CNN) untuk melakukan peramalan penjualan di e-commerce. Penelitian ini dilakukan untuk memecahkan masalah yang teridentifikasi batasan yang merupakan metode yang memerlukan rekayasa fitur manual kasus per kasus untuk skenario tertentu yang sulit, memakan waktu dan membutuhkan banyak pengetahuan ahli. Dari hasil pengujian, model ARIMA memiliki nilai rata-rata tertinggi, namun algoritma CNN dapat secara otomatis mengekstrak fitur yang efektif dan melakukan peramalan penjualan menggunakan fitur yang diekstraksi. Selanjutnya, Penggunaan pendekatan regresi untuk peramalan penjualan lebih baik dibandingkan dengan metode deret waktu [11].

BAHAN DAN METODE

Pada bagian ini, kami uraikan bahan dan metode yang digunakan untuk membangun model peramalan penjualan produk, yaitu:

1. Himpunan Data

Dalam penelitian ini, kami fokus mempelajari analisis prediktif dalam pengambilan keputusan bisnis dengan menggunakan model pembelajaran mesin untuk peramalan deret waktu penjualan. Untuk kebutuhan analisis ini, maka kami menggunakan kumpulan dataset penjualan toko yang bersumber dari kompetisi ilmu data kaggle yang dapat dilihat pada lampiran. Kumpulan dataset ini memiliki data catatan transaksi penjualan toko selama 5 tahun dari tahun 2013 sampai dengan 2018. Data ini memiliki 913.000 bari data transaksi penjualan, 10 toko berbeda dan 50 jenis item produk berbeda. Lebih rinci statistik data terlihat pada tabel 1 dan grafik distribusi data.

2. Metode yang diusulkan

Metode yang diusulkan dalam penelitian kami gunakan teknik dan kerangka kerja ilmu data yang berbeda untuk membuat perkiraan permintaan yang akurat. Dalam setiap tugas ilmu data, saya menggunakan CRISP-DM untuk mengikuti semua proses yang diperlukan selama mengerjakan proyek, yaitu:

a) Pemahaman bisnis

Analisis bisnis dalam pembahasan makalah ini adalah bagaimana memprediksi volume penjualan semua item di semua toko pada kuartal pertama tahun 2018. Semua model diprediksi dalam 2% dari penjualan rata-rata bulanan untuk prediksi 12 bulan ke depan. Jadi, masalah peramalan deret waktu cenderung tidak stasioner dimana data penjualan mungkin memiliki tren dan pola musiman. Tren, musim dan bahkan seluruh distribusi data dapat berubah, tergantung pada periode waktu. Selain itu, kami juga berurusan dengan deret waktu multivariat. Penjualan barang di satu toko atau satu barang di toko yang berbeda dapat dikorelasikan. Dalam masalah ini kami harus menggunakannya agar sesuai dengan model yang diterapkan agar menghasilkan peramalan yang akurat.

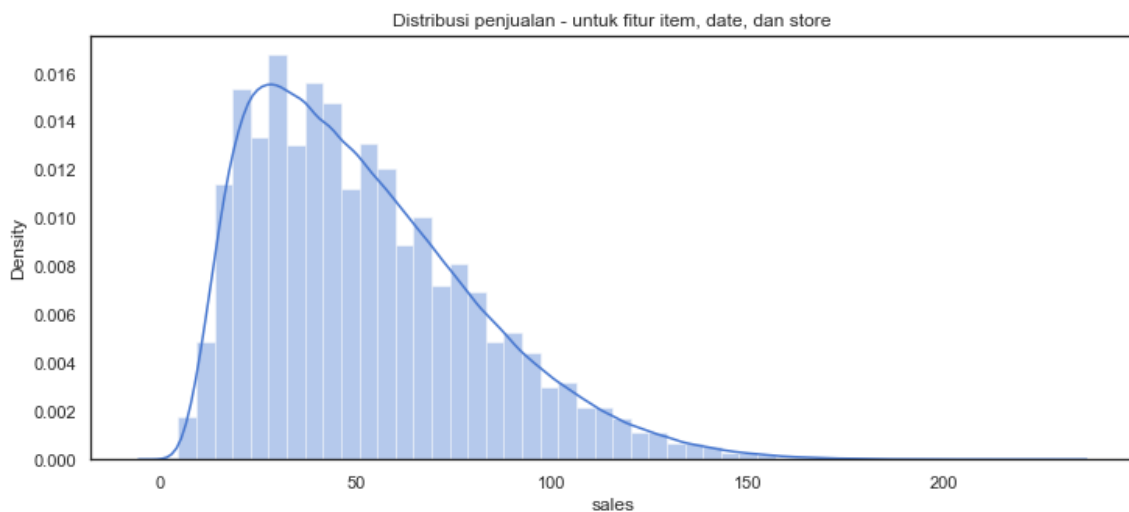
b) Pemahaman dan Persiapan Data

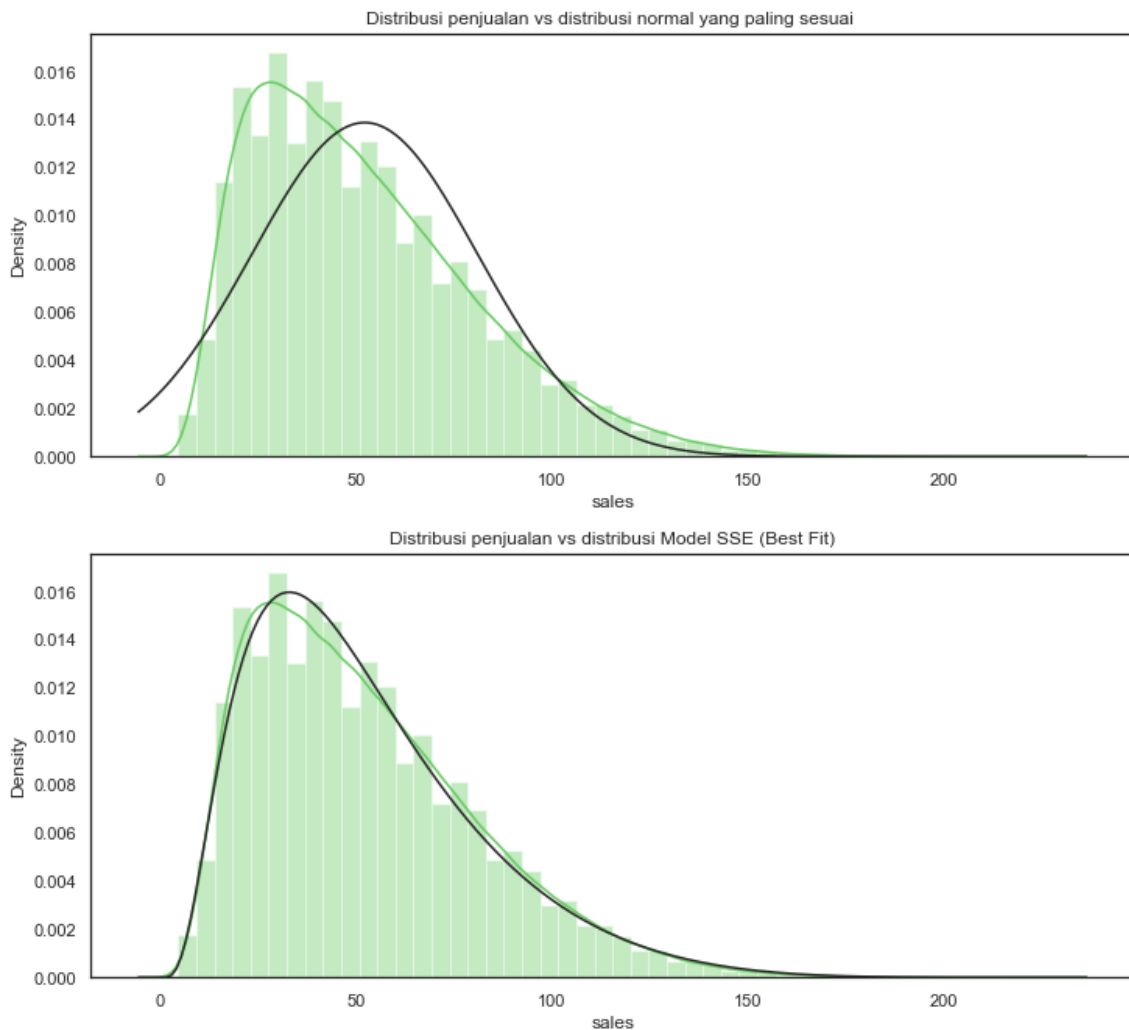
Data diberikan dalam format csv. Dan kumpulan dataset latihan memiliki kurang dari satu juta baris dan ukuran lebih kecil, jadi tidak diperlukan teknologi data besar yang canggih untuk pengolahan data. Kami juga menggunakan prediksi model yang telah dilatih sebelumnya untuk keperluan pengujian model dengan tujuan mempercepat waktu eksekusi.

Nilai penjualan diberikan sebagai bilangan bulat, jadi satu unit mungkin sesuai dengan satu item. Untuk menghindari pembulatan angka yang tidak disengaja di masa mendatang, kami mengubah penjualan menjadi menjadi type data *float*. Untuk pengujian, kami membulatkan nilai penjualan yang diprediksi. Semua fitur data tidak memiliki data yang hilang sehingga tidak memerlukan proses pembersihan data dan terdapat 3 fitur penting untuk kebutuhan peralaman yaitu: store, item dan sales.

Tabel 1 Statistik Data

	store	item	sales
count	913000.000000	913000.000000	913000.000000
mean	5.500000	25.500000	52.250287
std	2.872283	14.430878	28.801144
min	1.000000	1.000000	0.000000
25%	3.000000	13.000000	30.000000
50%	5.500000	25.500000	47.000000
75%	8.000000	38.000000	70.000000
max	10.000000	50.000000	231.000000





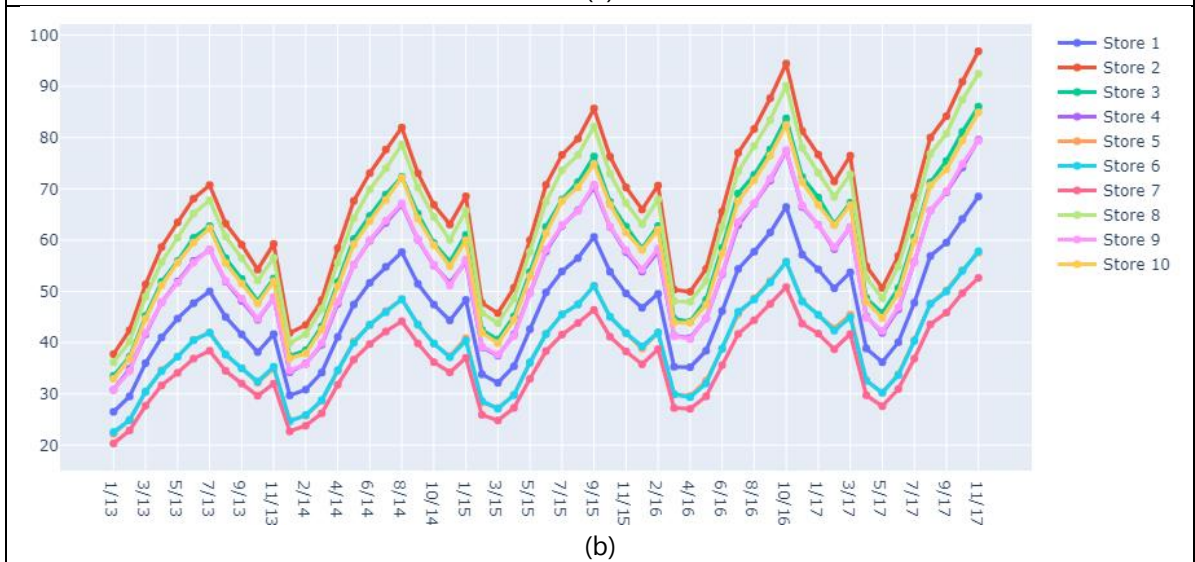
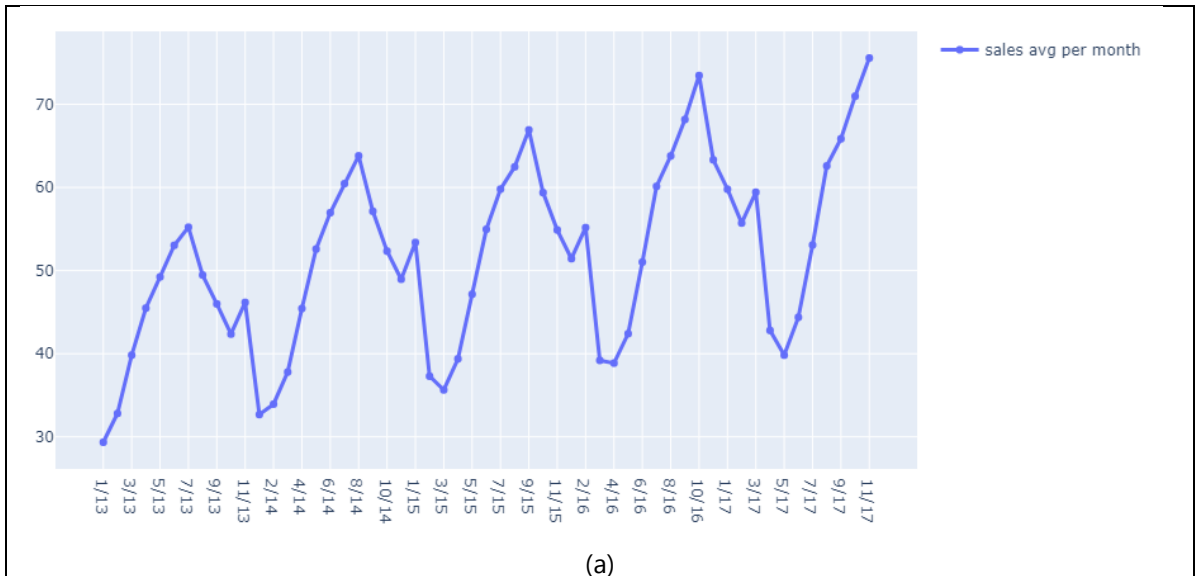
Gambar 1 (a) Grafik distribusi fitur, (b) Perbandingan distribusi penjualan vs distribusi normal data, (c) distribusi terbaik berdasarkan *sum of square error* (SSE)

Pada gambar 1 plot (b) di atas, garis hijau mewakili distribusi penjualan, sedangkan garis hitam adalah distribusi normal terbaik yang dapat di sesuaikan dengan data. Nilai p menunjukkan bahwa hipotesis nol dapat ditolak dan oleh karena itu data ini tidak sesuai dengan distribusi normal, maka kami melakukan distribusi normal data dengan pendekatan *sum of square error* (SSE) yang ditunjukkan pada plot (c) dimana data sudah berhasil di distribusikan dengan baik.

c) Exploratory Data Analysis

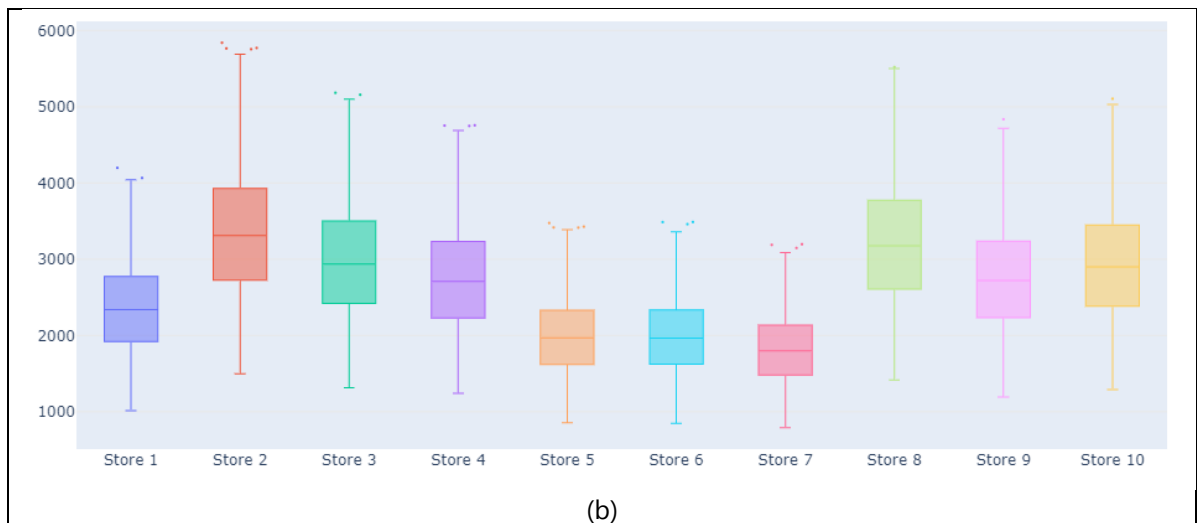
Exploratory Data Analysis (EDA) merupakan salah satu tahapan untuk memberikan pemahaman data dalam bentuk visualisasi sebelum dilakukan pemodelan data. Dalam tahapan ini, kami mengusulkan analisis prediktif untuk lebih memahami kumpulan data transaksi penjualan yang disajikan dalam bentuk visualisasi dan hasilnya di tunjukkan pada gambar 2.

Analisis prediktif dapat membantu untuk memahami parameter input utama dari kumpulan data. Dalam kumpulan data, analisis nilai transaksi penjualan untuk masing-masing toko merupakan bagian penting yang perlu dilakukan dalam pekerjaan ini mulai dari nilai rata-rata transaksi setiap bulan, rata-rata total transaksi untuk masing-masing toko. Berdasarkan hasil pengamatan bahwa deret waktu memiliki pola musiman, seperti penjualan selalu rendah di awal tahun dan tinggi di tengah tahun dan sekali lagi rendah di puncak tahun. Setelah memplot penjualan untuk setiap toko, kami melihat bahwa jumlah penjualan maksimum diamati oleh Toko 1, diikuti oleh Toko 8. Jumlah penjualan terkecil diamati oleh Toko 7



Gambar 2 Distribusi transaksi penjualan (a) rata-rata bulan, (b) rata-rata per toko





Gambar 3 Distribusi Nilai Penjualan (a) Total Per toko, (b) Box Plot

d) Peramalan Penjualan

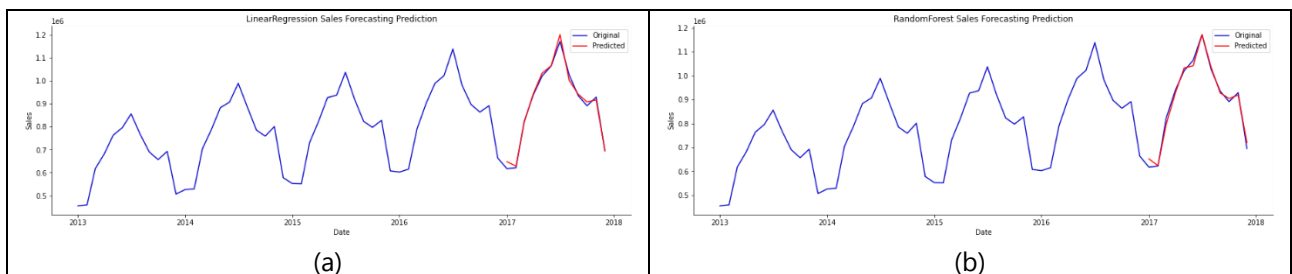
Pada tahapan ini diusulkan lima algoritma berbeda yaitu ARIMA, Random Forest (RF), Linear Regression (LR), XGBoost Classifier dan *Long short-term memory* (LSTM). Semua hasil peramalan algoritma yang diusulkan akan dievaluasi berdasarkan *root mean squared error* (RMSE), *mean absolute error* (MEA).

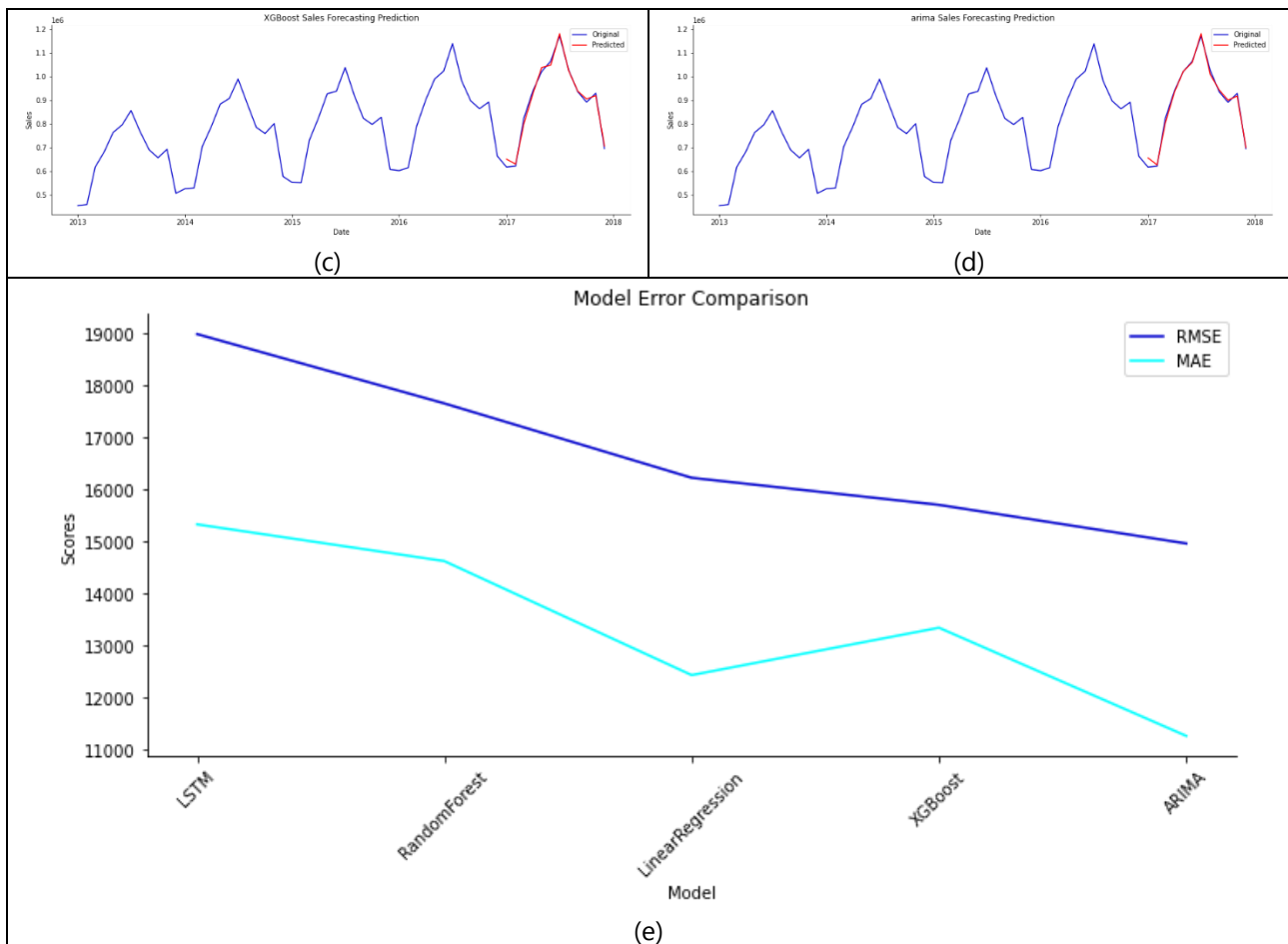
HASIL PENELITIAN

Hasil penelitian ini dibangun menggunakan bahasa pemrograman python 3 dengan Jupyter Notebook dan beberapa pustaka untuk visualisasi maupun pengolahan data seperti numpy, pandas, matplotlib, seaborn. Berdasarkan hasil pengujian terhadap lima model yang diusulkan, algoritma XGBoost menghasilkan tingkat kesalahan paling rendah dengan nilai MAE sebesar 1.3% dari nilai transaksi penjualan aktual. Lebih rinci dapat dilihat pada tabel 1 dan gambar 4 hasil prediksi dari masing-masing algoritma.

Tabel 2 Hasil Prediksi

Algoritma	RMSE	MAE	R2
LSTM	18979.7381	15327.16667	0.987289068
RandomForest	17648.48278	14621.75	0.989009645
LinearRegression	16221.04079	12433	0.990715588
XGBoost	15701.00336	13342.66667	0.991301351
ARIMA	14959.89347	11265.33575	0.983564288





Gambar 4 Hasil Prediksi (a) LinearRegression, (b) Random Forest, (c) XGBoost, (d) ARIMA, (e) Komparasi Kesalahan Metode

Diskusi

Penelitian ini mencoba menyajikan penerapan analisis prediktif untuk meramalkan data transaksi penjualan untuk bulan berikutnya. Dari hasil pengamatan kami, algoritma XGBoost menghasilkan 1.3% hasil prediksi dari data aktual. Algoritma XGBoost merupakan kerangka Gradient Boosting yang dirancang khusus agar selalu efisien, fleksibel dan kuat, namun metode ini tidak dapat menangani fitur kategoris dengan sendirinya, dan hanya menerima nilai numerik yang mirip dengan Random Forest sehingga perlu beberapa tahapan yang dilakukan agar menghasilkan akurasi yang baik seperti pelabelan. XGboost terbagi hingga **max_depth** yang ditentukan hyperparameter. Selanjutnya pengaturan hyper-parameter untuk meningkatkan hasil prediksi sangat di perlukan untuk penelitian di masa depan..

KESIMPULAN

Penelitian ini fokus menyajikan penerapan analisis prediktif untuk membuat model peramalan menggunakan algoritma pembelajaran mesin berbeda yaitu ARIMA, Random Forest (RF), Linear Regression (LR), XGBoost Classifier dan *Long short-term memory* (LSTM). Semua hasil peramalan model yang diusulkan berguna untuk membantu dalam menentukan keputusan bisnis. Berdasarkan hasil pengujian, algoritma XGBoost menghasilkan prediksi berada dalam 1,3% dari yang sebenarnya. ARIMA sebesar 1.6% masih lebih baik dibandingkan RandomForest, LinearRegression dan LSTM. Hasil penelitian masih dapat ditingkatkan dengan

menggunakan metode Rolling Mean untuk mengurangi kesalahan peramalan akan dipertimbangkan di masa depan.

Supplementary Materials (optional)

Sumber dataset tersedia di <https://www.kaggle.com/blacktile/starbucks-app-customer-reward-program-data>

Kontribusi Penulis

Semua Penulis memiliki kontribusi yang sama dalam makalah ini Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Konflik kepentingan

Para penulis menyatakan tidak ada konflik kepentingan.

REFERENCES

- [1] J. P. U. Cadavid, S. Lamouri, and B. Grabot, "Trends in machine learning applied to demand & sales forecasting: A review," *ILS 2018 - Inf. Syst. Logist. Supply Chain. Proc.*, pp. 307–316, 2018.
- [2] J. Huikku, T. Hyvönen, and J. Järvinen, "The role of a predictive analytics project initiator in the integration of financial and operational forecasts," *Balt. J. Manag.*, vol. 12, no. 4, pp. 427–446, Sep. 2017, doi: 10.1108/BJM-05-2017-0164.
- [3] J. Li, Y. W. Liu, Y. Jia, Y. Ren, and J. Nanduri, "Predictive modeling with delayed information: A case study in e-commerce transaction fraud control," *arXiv*, pp. 1–12, 2018.
- [4] D. S. Siringo-Ringo, R. B. D. Tambunan, D. Yulizar, T. A. Daulay, and A. M. Husein, "Application of Data Mining for Optimal Drug Inventory in a Hospital," *Sinkron*, vol. 4, no. 1, p. 207, 2019, doi: 10.33395/sinkron.v4i1.10236.
- [5] A. M. Husein, M. Arsyah, S. Sinaga, and H. Syahputa, "Generative Adversarial Networks Time Series Models to Forecast Medicine Daily Sales in Hospital," *Sinkron*, vol. 3, no. 2, p. 112, 2019, doi: 10.33395/sinkron.v3i2.10044.
- [6] A. M. Husein, A. M. Simarmata, M. Harahap, S. Aisyah, and A. Dharma, "Implementation ANFIS Method for Prediction Needs Drug-based Population Diseases and Patient," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*, 2019, pp. 1–5.
- [7] A. M. Husein, M. Harahap, S. Aisyah, W. Purba, and A. Muhazir, "The implementation of two stages clustering (k-means clustering and adaptive neuro fuzzy inference system) for prediction of medicine need based on medical data," *J. Phys. Conf. Ser.*, vol. 978, no. 1, p. 12019, 2018, doi: 10.1088/1742-6596/978/1/012019.
- [8] A. Krishna, V. Akhilesh, A. Aich, and C. Hegde, "Sales-forecasting of Retail Stores using Machine Learning Techniques," *Proc. 2018 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2018*, pp. 160–166, 2018, doi: 10.1109/CSITSS.2018.8768765.
- [9] R. K. Vashishtha, S. Sethuraman, V. Burman, A. R. Sekar, R. Kumar, and S. Ramanan, "Product age based demand forecast model for fashion retail," *arXiv*, 2020.
- [10] C. P. P. Maibang, A. M. Husein, and others, "Prediksi Jumlah Produksi Palm Oil Menggunakan Fuzzy Inference System Mamdani," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 2, p. 19, 2019, doi: 10.34012/jutikomp.v2i2.528.
- [11] B. M. Pavlyshenko, "Machine-learning models for sales time series forecasting," *Data*, vol. 4, no. 1, pp. 1–11, 2019, doi: 10.3390/data4010015.
- [12] S. Hanumanthasrastry and Prasadababu, "Analysis & Prediction of Sales Data in Sap-Erp System Using Clustering Algorithms," *Int. J. Comput. Sci. Inf. Technol.*, vol. 1, no. 4, pp. 95–109, 2013, doi: 10.5121/ijcsity.2013.1407.
- [13] M. Harahap, A. M. Husein, S. Aisyah, F. R. Lubis, and B. A. Wijaya, "Mining association rule based on the

- diseases population for recommendation of medicine need," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, p. 12017, 2018, doi: 10.1088/1742-6596/1007/1/012017.
- [14] R. Fildes, S. Ma, and S. Kolassa, "Retail forecasting: Research and practice," *Int. J. Forecast.*, no. xxxx, 2019, doi: 10.1016/j.ijforecast.2019.06.004.
- [15] Y. Lingxian, K. Jiaqing, and W. Shihuai, "Online retail sales prediction with integrated framework of K-mean and neural network," *ACM Int. Conf. Proceeding Ser.*, pp. 115–118, 2019, doi: 10.1145/3345035.3345048.
- [16] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman, "Sales demand forecast in E-commerce using a long short-term memory neural network methodology," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11955 LNCS, pp. 462–474, 2019, doi: 10.1007/978-3-030-36718-3_39.
- [17] K. Zhao and C. Wang, "Sales forecast in E-commerce using convolutional neural network," *arXiv*, 2017.
- [18] S. Ma and R. Fildes, "Forecasting third-party mobile payments with implications for customer flow prediction," *Int. J. Forecast.*, vol. 36, no. 3, pp. 739–760, 2020, doi: 10.1016/j.ijforecast.2019.08.012.