BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 *Grand Theory, Middle Theory,* dan *Applied Theory*

Landasan teoritis penelitian ini menggunakan perspektif *Data-Driven Decision Making* (DDDM) sebagai *Grand Theory*. DDDM merupakan pendekatan pengambilan keputusan yang menempatkan data dan analisis sebagai dasar dalam menentukan tindakan organisasi. Dalam perspektif ini, data tidak hanya berfungsi sebagai catatan dari aktivitas operasional, tetapi dapat diolah menjadi informasi yang mendukung proses pengambilan keputusan secara lebih objektif. Penelitian Brynjolfsson, Hitt, dan Kim menunjukkan bahwa penerapan pengambilan keputusan berbasis data dan *business analytics* berkaitan dengan peningkatan produktivitas dan kinerja perusahaan [7]. Dalam konteks usaha, pemanfaatan teknologi digital juga dapat mendukung pengelolaan informasi dan meningkatkan kinerja keuangan UMKM [8].

Berdasarkan perspektif tersebut, penelitian ini menempatkan data transaksi historis Barbershop Azkara sebagai sumber informasi yang dapat diolah untuk menghasilkan prediksi pendapatan. Penerapan DDDM dalam penelitian ini tidak dimaksudkan untuk menggantikan pertimbangan manajerial, tetapi untuk menyediakan informasi berbasis data yang dapat digunakan sebagai bahan pendukung dalam proses evaluasi dan pengambilan keputusan bisnis.

Sebagai *Middle Theory*, penelitian ini menggunakan konsep *Predictive Analytics*. *Predictive Analytics* berkaitan dengan pemanfaatan data historis, metode

UNIVERSITAS NUSA MANDIRI

statistik, dan teknik *machine learning* untuk mengidentifikasi pola pada data serta menghasilkan prediksi terhadap nilai atau kondisi yang akan datang [9], [10]. Konsep ini menjadi penghubung antara prinsip pengambilan keputusan berbasis data dengan penerapan model prediksi pada data transaksi. Melalui pendekatan tersebut, data historis tidak hanya digunakan untuk mengetahui kondisi yang telah terjadi, tetapi juga dapat dimanfaatkan untuk memperoleh estimasi terhadap kondisi yang akan datang.

Dalam penelitian ini, *Predictive Analytics* diterapkan pada data transaksi historis yang diperoleh dari sistem *Point of Sales* (POS) Barbershop Azkara. Data transaksi yang terdokumentasi secara digital dapat menjadi sumber data untuk membangun model prediksi pendapatan. Penelitian terdahulu juga menunjukkan pemanfaatan algoritma *machine learning*, termasuk *Linear Regression* dan *Random Forest*, untuk melakukan prediksi penjualan [3], sedangkan *Random Forest Regression* telah diterapkan untuk prediksi pendapatan penjualan [4]. Dengan demikian, *Predictive Analytics* menjadi landasan yang menghubungkan data transaksi dengan proses pemodelan dan prediksi pendapatan dalam penelitian ini.

Pada tingkat *Applied Theory*, penelitian ini menerapkan algoritma *Supervised Learning*, khususnya *Linear Regression* dan *Random Forest Regression*, untuk memprediksi pendapatan harian. *Linear Regression* merupakan metode regresi yang digunakan untuk memodelkan hubungan antara variabel prediktor dan variabel respons sehingga dapat digunakan untuk menghasilkan estimasi berdasarkan pola hubungan yang terdapat dalam data [2]. Metode ini digunakan sebagai salah satu model dalam penelitian karena struktur pemodelannya relatif sederhana dan dapat digunakan untuk mengestimasi variabel dengan nilai kontinu, termasuk pendapatan.

UNIVERSITAS NUSA MANDIRI

Algoritma kedua yang digunakan adalah *Random Forest Regression***.** *Random Forest* merupakan metode *ensemble learning* yang membangun sejumlah *decision tree* dan menggabungkan hasil prediksi dari pohon-pohon tersebut untuk menghasilkan prediksi akhir [11]. Berbeda dengan *Linear Regression* yang memodelkan hubungan linear, *Random Forest* dapat digunakan untuk menangkap hubungan yang lebih kompleks dalam data. Penelitian terdahulu menunjukkan bahwa *Random Forest* dapat diterapkan dalam permasalahan prediksi penjualan maupun pendapatan [3], [4]. Oleh karena itu, penggunaan kedua algoritma tersebut memungkinkan penelitian melakukan perbandingan kinerja model pada data pendapatan harian Barbershop Azkara.

Kinerja kedua model selanjutnya dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (R^2). Penggunaan beberapa metrik evaluasi dilakukan untuk memperoleh gambaran yang lebih menyeluruh mengenai tingkat kesalahan prediksi dan kemampuan model dalam menjelaskan variasi data. Model dengan kinerja terbaik kemudian dapat digunakan untuk menghasilkan estimasi pendapatan harian berdasarkan pola yang terdapat pada data historis.

Ketiga lapisan teori tersebut memiliki hubungan yang saling melengkapi dalam penelitian ini. DDDM memberikan landasan konseptual mengenai pentingnya pemanfaatan data sebagai dasar pendukung pengambilan keputusan. Predictive Analytics menjelaskan pendekatan yang digunakan untuk mengolah data historis menjadi informasi prediktif. Selanjutnya, *Linear Regression* dan *Random Forest Regression* menjadi bentuk penerapan metode *machine learning* untuk menghasilkan prediksi pendapatan harian.

UNIVERSITAS NUSA MANDIRI

Integrasi tersebut menempatkan data transaksi Barbershop Azkara sebagai sumber informasi yang diolah melalui tahapan analisis dan pemodelan untuk menghasilkan prediksi pendapatan. Hasil prediksi selanjutnya digunakan sebagai informasi pendukung dalam evaluasi dan perencanaan operasional bisnis. Dengan demikian, penelitian ini tidak hanya berfokus pada perbandingan algoritma, tetapi juga menerapkan prinsip pengambilan keputusan berbasis data melalui pemanfaatan informasi prediktif yang dihasilkan dari data transaksi historis.

2.1.2 Kajian Variabel Independen: Algoritma *Linear Regression* dan *Random Forest*

Algoritma *Linear Regression* didefinisikan secara konseptual sebagai metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dengan satu atau lebih variabel independen melalui estimasi koefisien linear. Dalam bidang sains data, algoritma ini sering digunakan sebagai model *baseline* karena memiliki struktur yang relatif sederhana serta mampu memberikan interpretasi yang jelas terhadap hubungan antara variabel prediktor dan variabel respons. *Linear Regression* sesuai digunakan ketika hubungan antarvariabel dapat direpresentasikan secara linear, tetapi kemampuan model dapat menjadi terbatas ketika data memiliki pola hubungan yang lebih kompleks atau bersifat non-linear [2], [9].

Dimensi utama dari *Linear Regression* terletak pada asumsi linearitas, normalitas, dan homoskedastisitas yang perlu diperhatikan dalam proses pemodelan untuk menghasilkan estimasi yang valid [2]. Namun, dalam praktiknya, data pendapatan harian dapat menunjukkan pola yang lebih kompleks dan bersifat non-linear akibat adanya variasi waktu maupun perubahan perilaku konsumen. Kondisi tersebut dapat menjadi keterbatasan bagi *Linear Regression* ketika hubungan antara

UNIVERSITAS NUSA MANDIRI

variabel prediktor dan variabel respons tidak dapat direpresentasikan secara linear [2], [9]. Indikator performa yang digunakan untuk mengukur akurasi algoritma ini meliputi *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), yang memberikan gambaran mengenai besarnya deviasi antara nilai prediksi dengan nilai aktual. Selain itu, penelitian ini menggunakan koefisien determinasi (R^2) sebagai metrik tambahan untuk mengevaluasi kemampuan model dalam menjelaskan variasi pada data yang diprediksi.

Di sisi lain, *Random Forest Regression* merupakan algoritma *ensemble learning* yang bekerja dengan membangun sekumpulan pohon keputusan (*decision trees*) selama fase pelatihan. Algoritma ini mengatasi kelemahan pohon keputusan tunggal yang cenderung mengalami *overfitting* dengan membangun dan menggabungkan sejumlah pohon yang dilatih menggunakan proses pengacakan terhadap data dan fitur, kemudian menghasilkan prediksi berdasarkan agregasi hasil dari seluruh pohon [11]. Pendekatan tersebut memungkinkan *Random Forest* menghasilkan model yang lebih stabil dan mampu menangkap hubungan yang kompleks dalam data dibandingkan penggunaan satu pohon keputusan.

Faktor yang memengaruhi performa *Random Forest* antara lain jumlah pohon (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*), yang termasuk dalam parameter yang dapat disesuaikan selama proses pemodelan [1]. Penelitian terdahulu menunjukkan bahwa *Random Forest* dapat memberikan kinerja yang baik dibandingkan model linear pada permasalahan prediksi yang memiliki hubungan antarvariabel dan interaksi fitur yang kompleks [3], [12]. Teori pendukung *Bootstrap Aggregating* (Bagging) pada *Random Forest* memungkinkan model menggabungkan hasil dari sejumlah model dasar yang dibangun melalui proses pengambilan sampel secara bootstrap sehingga dapat meningkatkan stabilitas prediksi [11]. Karakteristik

UNIVERSITAS NUSA MANDIRI

tersebut menjadikan *Random Forest* sebagai salah satu kandidat model yang relevan untuk digunakan dalam memprediksi pendapatan harian pada bisnis multi-cabang yang memiliki pola transaksi dan karakteristik operasional yang berbeda pada setiap lokasi.

Hubungan antara kedua algoritma ini dalam penelitian adalah sebagai subjek komparasi untuk menentukan efektivitas pemodelan pada data Barbershop Azkara. *Linear Regression* diuji untuk melihat sejauh mana hubungan linear dapat menjelaskan pola pendapatan, sementara *Random Forest* digunakan untuk menangkap hubungan yang lebih kompleks yang mungkin tidak dapat direpresentasikan secara optimal oleh model linear [9], [11]. Sintesis kritis menunjukkan bahwa pemilihan algoritma perlu disesuaikan dengan karakteristik data dan pola hubungan antarvariabel. Oleh karena itu, penelitian ini membandingkan *Linear Regression* dan *Random Forest Regression* untuk mengetahui algoritma yang memberikan kinerja prediksi terbaik pada data pendapatan harian Barbershop Azkara [12].

Penelitian terdahulu telah menerapkan dan membandingkan *Linear Regression* dan *Random Forest* pada berbagai permasalahan prediksi untuk mengetahui model yang memberikan kinerja terbaik sesuai dengan karakteristik data [3], [12]. Dalam konteks prediksi pendapatan, pemilihan algoritma perlu mempertimbangkan karakteristik data dan fitur yang digunakan dalam proses pemodelan. Posisi variabel independen dalam penelitian ini adalah sebagai variabel prediktor yang digunakan untuk membantu model dalam menghasilkan estimasi pendapatan pada periode penelitian.

Secara operasional, implementasi kedua algoritma didahului oleh tahap pra-pemrosesan data, termasuk pemeriksaan data yang hilang, data duplikat,

UNIVERSITAS NUSA MANDIRI

ketidakkonsistenan data, serta transformasi fitur yang diperlukan sesuai dengan karakteristik data [9], [13]. Tahap pra-pemrosesan menjadi bagian penting karena kualitas data yang digunakan dapat memengaruhi proses pembentukan model dan hasil prediksi. Oleh karena itu, evaluasi model dilakukan secara sistematis untuk memastikan bahwa hasil prediksi yang diperoleh dapat merepresentasikan kondisi data penelitian.

Salah satu perbedaan karakteristik antara *Random Forest* dan *Linear Regression* terletak pada tingkat interpretabilitas model. *Random Forest* yang terdiri atas sejumlah pohon keputusan memiliki struktur yang lebih kompleks sehingga proses pembentukan prediksinya relatif lebih sulit diinterpretasikan secara langsung. Sebaliknya, *Linear Regression* memiliki struktur model yang lebih sederhana sehingga hubungan antara variabel prediktor dan variabel respons dapat dijelaskan melalui koefisien regresi [9], [11]. Perbedaan karakteristik tersebut menjadi salah satu pertimbangan dalam melakukan perbandingan kedua algoritma, karena pemilihan model tidak hanya berkaitan dengan tingkat akurasi, tetapi juga dengan kemudahan interpretasi hasil model.

Sintesis dari kajian kedua algoritma menunjukkan bahwa komparasi antara *Linear Regression* dan *Random Forest* tidak hanya bertujuan untuk memperoleh nilai kesalahan prediksi terkecil, tetapi juga untuk mengetahui karakteristik model yang paling sesuai dengan data pendapatan harian Barbershop Azkara. Dalam konteks penelitian ini, model yang terpilih diharapkan mampu menghasilkan estimasi yang mendekati nilai aktual sehingga dapat digunakan sebagai informasi pendukung dalam proses pengambilan keputusan.

Keberhasilan model prediksi juga dipengaruhi oleh pemilihan dan representasi fitur yang digunakan dalam proses pemodelan. Fitur seperti hari dalam

minggu, waktu transaksi, lokasi cabang, maupun informasi lain yang tersedia pada data transaksi dapat digunakan sebagai variabel prediktor apabila memiliki relevansi terhadap pendapatan yang diprediksi [9], [13]. Dengan demikian, variabel independen dalam penelitian ini diposisikan sebagai instrumen analitis yang akan diuji melalui kedua algoritma untuk mengetahui kemampuannya dalam merepresentasikan pola pendapatan harian.

2.1.3 Kajian Variabel Dependen: Pendapatan Bulan (September 2026)

Pendapatan harian secara konseptual dapat didefinisikan sebagai jumlah pendapatan yang diperoleh dari aktivitas operasional bisnis dalam periode satu hari. Dalam konteks industri jasa seperti barbershop, pendapatan harian dapat dipengaruhi oleh jumlah pelanggan, jenis layanan yang dipilih, harga layanan, serta pola transaksi yang terjadi pada periode tertentu. Variasi pada faktor-faktor tersebut dapat menyebabkan perubahan jumlah pendapatan dari satu hari ke hari lainnya. Oleh karena itu, pendapatan harian menjadi salah satu indikator penting dalam mengevaluasi kondisi operasional dan kinerja bisnis. Dalam penelitian ini, pendapatan harian digunakan sebagai variabel dependen yang menjadi nilai target dalam proses pemodelan prediksi berdasarkan data transaksi historis Barbershop Azkara.

Dimensi pendapatan harian dalam penelitian ini mencakup nilai nominal transaksi, frekuensi transaksi, dan nilai rata-rata transaksi (*average transaction value*) yang dapat digunakan untuk menggambarkan karakteristik pendapatan selama periode pengamatan. Data transaksi yang tercatat melalui sistem Point of Sales (POS) memberikan informasi historis yang dapat digunakan untuk mengidentifikasi pola pendapatan harian. Dengan adanya data yang terstruktur,

UNIVERSITAS NUSA MANDIRI

pendapatan yang sebelumnya hanya berfungsi sebagai catatan operasional dapat dimanfaatkan sebagai variabel target dalam proses analisis dan pemodelan prediktif.

Indikator pengukuran pendapatan dalam penelitian ini difokuskan pada data aktual periode september 2026. Nilai pendapatan pada setiap tanggal digunakan sebagai dasar untuk melihat variasi pendapatan harian selama periode pengamatan. Perbedaan nilai pendapatan antarhari menunjukkan adanya fluktuasi yang perlu dipertimbangkan dalam proses pemodelan. Dengan demikian, periode september 2026 digunakan sebagai fokus pengamatan untuk mengetahui kemampuan model dalam menghasilkan estimasi pendapatan berdasarkan karakteristik data transaksi yang tersedia.

Fluktuasi pendapatan harian dapat terjadi karena adanya perbedaan jumlah transaksi, jenis layanan, waktu transaksi, maupun karakteristik pelanggan pada setiap periode. Kondisi tersebut menyebabkan pendapatan tidak selalu memiliki nilai yang sama dari satu hari ke hari lainnya. Oleh karena itu, pemodelan pendapatan membutuhkan pendekatan yang mampu memanfaatkan pola yang terdapat pada data historis sehingga perubahan nilai pendapatan dapat dipelajari secara sistematis.

Penelitian terdahulu menunjukkan bahwa algoritma *Linear Regression* dan *Random Forest* dapat diterapkan dalam permasalahan prediksi penjualan dan pendapatan. Falah et al. menerapkan *Linear Regression* dan *Random Forest* untuk memprediksi penjualan, sedangkan Kurniawan et al. menerapkan *Random Forest Regression* dalam memprediksi pendapatan penjualan [3], [4]. Hasil penelitian tersebut menunjukkan bahwa data historis transaksi dapat dimanfaatkan sebagai dasar dalam membangun model prediksi yang mendukung analisis kinerja penjualan maupun pendapatan. Dalam konteks Barbershop Azkara, pendapatan harian menjadi

UNIVERSITAS NUSA MANDIRI

variabel dependen yang akan diprediksi untuk menghasilkan informasi yang dapat mendukung perencanaan operasional usaha.

Posisi pendapatan harian sebagai variabel dependen menempatkannya sebagai nilai hasil yang menjadi fokus utama dalam proses pemodelan. Seluruh variabel independen yang digunakan dalam penelitian diarahkan untuk menjelaskan pola perubahan pada nilai pendapatan tersebut. Oleh karena itu, keberhasilan model tidak hanya dilihat dari kemampuannya menghasilkan prediksi, tetapi juga dari tingkat kedekatan antara nilai prediksi dan nilai aktual yang diukur menggunakan metrik evaluasi seperti MAE, RMSE, dan R^2.

Dalam proses analisis, karakteristik data pendapatan perlu diperhatikan untuk mengidentifikasi pola, variasi, serta kemungkinan adanya nilai yang tidak biasa. Pemeriksaan terhadap data historis menjadi bagian penting sebelum model dibangun karena kualitas data dapat memengaruhi hasil pemodelan. Analisis data eksploratori (*Exploratory Data Analysis*/EDA) digunakan untuk memahami distribusi data dan pola pendapatan sebelum memasuki tahap pemodelan menggunakan *Linear Regression* dan *Random Forest Regression*.

Sintesis antara variabel dependen dan variabel independen dalam penelitian ini terletak pada kemampuan algoritma untuk memanfaatkan informasi yang terdapat pada data historis dalam menghasilkan estimasi pendapatan. Penggunaan model regresi yang tepat diharapkan dapat memberikan gambaran mengenai pola pendapatan selama periode pengamatan sehingga hasil prediksi dapat digunakan sebagai informasi pendukung bagi manajemen Barbershop Azkara dalam melakukan perencanaan dan evaluasi operasional.

Secara keseluruhan, pendapatan harian dalam penelitian ini tidak hanya diposisikan sebagai angka hasil transaksi, tetapi sebagai variabel dependen yang

menjadi tolok ukur dalam mengevaluasi kinerja model prediksi. Fokus pada periode september 2026 digunakan untuk menguji kemampuan *Linear Regression* dan *Random Forest Regression* dalam memprediksi nilai pendapatan berdasarkan data transaksi yang tersedia. Dengan demikian, hasil prediksi yang diperoleh dapat menjadi dasar untuk menentukan model yang memiliki kinerja paling sesuai dengan karakteristik data pendapatan harian Barbershop Azkara.

2.2 Penelitian Terkait

Penyusunan landasan teoretis dalam penelitian ini didukung oleh serangkaian studi relevan yang telah memvalidasi penggunaan algoritma *machine learning* dalam konteks prediksi. Tabel berikut merangkum kontribusi ilmiah dari berbagai peneliti yang menjadi acuan dalam membedah performa *Linear Regression* dan *Random Forest Regression* dapat dilihat pada Tabel II.1.

Tabel II.1 Penelitian Terdahulu

NO	PENULIS	TAHUN	JUDUL PENELITIAN	METODOLOGI	HASIL PENELITIAN
1	Moh Adi Kurniawan, Gutti Zaidan Syauqi, Mia Safriyanti, Fadila Ullul Azmie, Arif Setiawan	2025	Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma *Random Forest* Regression	Penelitian kuantitatif dengan pendekatan eksperimen menggunakan algoritma *Random Forest* Regressor. Evaluasi model menggunakan MAE dan R^2 Score.	*Random Forest* menghasilkan performa prediksi yang sangat baik dengan **MAE sebesar 9.587,48** dan **R^2 sebesar 0,9998**, sehingga efektif digunakan untuk prediksi pendapatan penjualan serta mendukung pengambilan keputusan manajemen stok dan promosi.
2	Doni Saputra, Dio	2025	Perbandingan Metode *Random*	Menggunakan pendekatan	*Random Forest* memperoleh

NO	PENULIS	TAHUN	JUDUL PENELITIAN	METODOLOGI	HASIL PENELITIAN
	Raihan Trinadi, Desti Agustina, Pujianto		*Forest, Linier Regression*, SVM Untuk Memprediksi Harga Beras Premium	*supervised learning* dengan validasi silang (*cross validation*). Membandingkan *Random Forest, Linear Regression*, dan SVM menggunakan evaluasi MAE.	nilai **MAE paling kecil**, sehingga menjadi algoritma terbaik dibandingkan *Linear Regression* dan SVM dalam memprediksi harga beras premium.
3	Diah Pita Loka, Lena Magdalena, Mesi Febima	2025	Perancangan Sistem Prediksi Pendapatan BRT Menggunakan Penerapan Hybrid *Random Forest* dan LSTM	Pendekatan *hybrid Machine Learning*. *Random Forest* digunakan untuk *feature selection*, sedangkan LSTM digunakan untuk prediksi *time series*. Evaluasi menggunakan MAE, MSE, RMSE, dan R^2.	Model hybrid meningkatkan akurasi prediksi dengan **MAE 8,90%**, **RMSE 11,59%**, dan **R^2 75,97%**, sehingga lebih baik dibandingkan metode konvensional.
4	Muthia Rahmah, Kanaya Ramadanti, Imelda Fransiska Aulia	2026	Perbandingan Algoritma Regresi dalam Memprediksi Penjualan Berdasarkan Indikator Sosial Ekonomi Kabupaten Cirebon (2010–2023)	Membandingkan SVR, *Gradient Boosting Regressor* (GBR), *Random Forest Regressor* (RFR), dan XGBoost. Evaluasi menggunakan MAE, RMSE, dan R^2.	**Gradient Boosting Regressor (GBR)** memberikan hasil terbaik dengan **MAE 127,98**, **RMSE 185,63**, dan **R^2 0,94**, mengungguli *Random Forest*, SVR, dan XGBoost.
5	Marju Gopinda	2026	Studi Perbandingan Algoritma *Machine Learning* untuk Prediksi Penjualan Retail	Studi komparatif menggunakan *Linear Regression, Random Forest,* SVR, ANN, Gradient Boosting, dan XGBoost. Evaluasi	Model ***Gradient Boosting*** dan **XGBoost** menghasilkan error paling rendah dan stabilitas prediksi terbaik. *Random Forest* dan SVR lebih

UNIVERSITAS NUSA MANDIRI

NO	PENULIS	TAHUN	JUDUL PENELITIAN	METODOLOGI	HASIL PENELITIAN
				menggunakan RMSE, MAE, dan R^2.	baik dibanding *Linear Regression* pada data non-linear.
6	Algies Rifkha Fadillah, Mohamad Nurkamal Fauzan, Nuraini Siti Fathonah	2024	Analisis Perbandingan *Linear Regression* dan *Random Forest Regression* untuk Prediksi Batas Kredit: Pendekatan Optimasi Hyperparameter	Membandingkan *Linear Regression* dan *Random Forest Regression* dengan optimasi hyperparameter menggunakan Optuna. Evaluasi menggunakan RMSE dan MAE.	*Random Forest Regression* memberikan hasil terbaik dengan **RMSE 7,90%** dan **MAE 4,72%**, sehingga lebih akurat dibandingkan *Linear Regression* dalam prediksi batas kredit.
7	Rafli Adiguna Saputra	2024	Perbandingan Metode Regresi dan Jaringan Saraf Tiruan Backpropagation (Prediksi Harga Makanan Laut)	Penelitian menggunakan data time series harga makanan laut. Membandingkan *Linear Regression* dan Jaringan Saraf Tiruan Backpropagation.	*Linear Regression* memperoleh rata-rata akurasi **95,26%**, sedangkan Backpropagation mencapai **96,50%**, sehingga JST lebih unggul dalam prediksi harga makanan laut.
8	Wilsen Grivin Mokodaser, Hartiny Pop Koapaha, Stenly Ibrahim Adam	2025	Model *Random Forest* Data Historis Multivariat untuk Prediksi Pendapatan Asuransi	Penelitian menggunakan ***Random Forest Regression*** pada data historis multivariat. Tahapan penelitian meliputi *preprocessing data* (*missing value* dan *outlier*), pemodelan *Random Forest*, evaluasi menggunakan **MAE, MSE, dan R^2 Score**,	Model *Random Forest* mampu memberikan prediksi pendapatan asuransi dengan performa yang baik. Hasil evaluasi menunjukkan **MAE sebesar ±25.139.426**, **MSE sebesar $2{,}9815 \times 10^{15}$**, dan **R^2 Score sebesar 0,85**, yang berarti model dapat menjelaskan

NO	PENULIS	TAHUN	JUDUL PENELITIAN	METODOLOGI	HASIL PENELITIAN
				implementasi model, ***5-Fold Cross Validation***, serta visualisasi melalui dashboard interaktif Gradio.	**85% variasi pendapatan**. Hasil *cross validation* memperoleh rata-rata **R^2 sebesar 0,673**, sehingga model dinilai cukup stabil dan layak digunakan untuk prediksi pendapatan berdasarkan data historis multivariat.
9	Doni Saputra, Dio Raihan Trinadi, Desti Agustina, Pujianto	2025	Perbandingan Metode *Random Forest, Linier Regression*, SVM untuk Memprediksi Harga Beras Premium	*Supervised Learning, Random Forest, Linear Regression,* SVM, *Cross Validation*, evaluasi MAE	*Random Forest* memperoleh nilai MAE paling kecil sehingga menjadi algoritma yang paling akurat dibandingkan *Linear Regression* dan SVM dalam memprediksi harga beras premium.
10	Miftahul Jannah, Wanayumini	2026	Implementasi Regresi Linier Berganda untuk Prediksi Penjualan Produk *Home Brand* pada Outlet Farmasi	*Multiple Linear Regression*, data transaksi harian, evaluasi MAE, RMSE, MAPE, dan R^2	Model menghasilkan performa prediksi yang baik dengan nilai MAPE sebesar 12,21% dan R^2 sebesar 0,9068 serta berhasil diimplementasikan pada sistem berbasis web.
11	Moh. Adi Kurniawan, dkk.	2025	Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma	*Random Forest Regression*	*Random Forest* mampu menghasilkan prediksi pendapatan penjualan dengan tingkat

NO	PENULIS	TAHUN	JUDUL PENELITIAN	METODOLOGI	HASIL PENELITIAN
			Random Forest Regression		akurasi yang baik dan layak digunakan sebagai pendukung pengambilan keputusan.
12	Diah Pita Loka, L. Magdalena, M. Febima	2025	Perancangan Sistem Prediksi Pendapatan BRT Menggunakan Hybrid *Random Forest* dan LSTM	Hybrid *Random Forest* dan LSTM	Kombinasi *Random Forest* dan LSTM meningkatkan akurasi prediksi dibandingkan penggunaan satu algoritma saja.
13	Marju Gopinda	2026	Studi Perbandingan Algoritma *Machine Learning* untuk Prediksi Penjualan Retail	*Linear Regression*, *Random Forest,* Decision Tree, XGBoost	*Random Forest* memberikan performa yang lebih stabil dibandingkan model regresi sederhana pada data penjualan retail.
14	Algies Rifkha Fadillah, dkk.	2025	Analisis Perbandingan *Linear Regression* dan *Random Forest Regression* untuk Prediksi Batas Kredit: Pendekatan Optimasi *Hyperparameter*	*Linear Regression*, *Random Forest Regression*, *Hyperparameter Tuning*	Optimasi hyperparameter meningkatkan performa *Random Forest* sehingga menghasilkan prediksi yang lebih akurat dibandingkan model dasar.
15	Géron, A.	2025	*Hands-On Machine Learning with Scikit-Learn*, Keras & *TensorFlow* (3rd Edition)	*Machine Learning, Hyperparameter Tuning*	Menjelaskan implementasi *Random Forest*, GridSearchCV, RandomizedSearchCV, dan evaluasi model menggunakan Python.

Studi terdahulu menunjukkan bahwa *Linear Regression* dan *Random Forest* telah banyak digunakan dalam berbagai permasalahan prediksi, termasuk prediksi

penjualan, pendapatan, harga, dan risiko. *Random Forest* memiliki kemampuan untuk menangani hubungan yang kompleks antarvariabel melalui pembentukan sejumlah pohon keputusan dan proses agregasi hasil prediksi [3], [4], [11]. Kemampuan tersebut menjadikan *Random Forest* sebagai salah satu algoritma yang relevan untuk permasalahan prediksi dengan karakteristik data yang beragam. Namun, kinerja *Random Forest* tetap bergantung pada karakteristik data dan parameter model yang digunakan dalam proses pemodelan.

Linear Regression tetap digunakan dalam berbagai penelitian sebagai model pembanding karena struktur modelnya relatif sederhana dan hubungan antara variabel prediktor dengan variabel target dapat diinterpretasikan melalui koefisien regresi [2], [9]. Penelitian terdahulu menunjukkan bahwa *Linear Regression* dapat digunakan untuk memprediksi nilai kontinu, termasuk penjualan dan pendapatan. Sementara itu, *Random Forest* dapat menjadi alternatif ketika pola hubungan pada data tidak dapat direpresentasikan secara optimal oleh model linear . Perbedaan kinerja kedua algoritma tersebut menunjukkan bahwa pemilihan model perlu disesuaikan dengan karakteristik dataset dan tujuan prediksi yang dilakukan.

Keterbatasan dari penelitian terdahulu terlihat dari perbedaan objek, karakteristik data, dan pendekatan pemodelan yang digunakan. Beberapa penelitian berfokus pada prediksi penjualan atau harga pada satu jenis objek, sedangkan penelitian lainnya menggunakan pendekatan hybrid atau membandingkan beberapa algoritma sekaligus [9], [11], [13]. Selain itu, penelitian mengenai optimasi *hyperparameter* menunjukkan bahwa konfigurasi parameter dapat memengaruhi kinerja *Random Forest*, sehingga proses penyetelan parameter menjadi salah satu aspek yang perlu diperhatikan dalam pembangunan model [7], [8]. Kondisi tersebut menjadi dasar bagi penelitian ini untuk melakukan perbandingan *Linear Regression*

UNIVERSITAS NUSA MANDIRI

dan *Random Forest Regression* dengan mempertimbangkan karakteristik data transaksi pada bisnis jasa multi-cabang.

Research gap dalam penelitian ini terletak pada masih terbatasnya penelitian yang secara spesifik membandingkan *Linear Regression* dan *Random Forest Regression* untuk memprediksi pendapatan harian pada bisnis jasa barbershop dengan memanfaatkan data transaksi dari sistem *Point of Sales* (POS). Penelitian terdahulu yang tersedia lebih banyak membahas prediksi penjualan, harga, atau pendapatan pada objek seperti retail, parfum, transportasi, dan komoditas pangan [3], [4], [14], [15]. Dengan demikian, penerapan kedua algoritma tersebut pada data pendapatan harian Barbershop Azkara dengan karakteristik usaha multi-cabang menjadi konteks yang berbeda dari penelitian sebelumnya.

Posisi penelitian ini dalam State of the Art (SOTA) adalah melakukan komparasi antara *Linear Regression* dan *Random Forest Regression* menggunakan data transaksi aktual dari sistem *Point of Sales* (POS) pada Barbershop Azkara yang memiliki beberapa cabang. Penelitian ini tidak hanya membandingkan kinerja kedua algoritma berdasarkan nilai evaluasi, tetapi juga menerapkan tahapan *preprocessing*, *feature engineering*, serta optimasi hyperparameter pada *Random Forest* untuk memperoleh konfigurasi model yang sesuai dengan karakteristik data. Fokus pada pendapatan harian periode maret-agustus 2026 digunakan sebagai ruang lingkup pengujian dalam penelitian. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi berupa informasi mengenai algoritma yang lebih sesuai untuk melakukan prediksi pendapatan harian pada konteks bisnis jasa barbershop multi-cabang.

UNIVERSITAS NUSA MANDIRI

2.3 Tinjauan Organisasi/Objek Penelitian

2.3.1 Gambaran Umum Organisasi

Barbershop Azkara adalah sebuah bisnis yang menyediakan layanan perawatan dan pemotongan rambut pria. Sebagai perusahaan jasa, organisasi ini utama melakukan pekerjaan dengan memberikan layanan kepada pelanggan, seperti pemotongan rambut dan perawatan rambut berbagai jenis. Dalam berjalannya operasionalnya, setiap transaksi pelayanan menghasilkan data yang mencakup waktu transaksi, jenis layanan, tarif, jumlah pelanggan, serta pendapatan yang diperoleh perusahaan. Data tersebut menjadi acuan penting bagi manajemen dalam menilai kinerja operasional serta kondisi keuangan perusahaan.

2.3.2 Kegiatan Operasional Organisasi

Operasional pada usaha barbershop secara umum dimulai ketika pelanggan memilih layanan yang diinginkan, kemudian barber memberikan pelayanan sesuai kebutuhan pelanggan. Setelah proses pelayanan selesai, pelanggan melakukan pembayaran dan seluruh transaksi dicatat ke dalam sistem *Point of Sales* (POS). Sistem tersebut menyimpan data transaksi secara otomatis sehingga memudahkan proses pencatatan penjualan, pengelolaan inventaris, serta penyusunan laporan keuangan perusahaan. Penerapan sistem POS juga mampu meminimalkan kesalahan pencatatan yang sering terjadi pada proses manual dan membantu penyediaan informasi operasional secara lebih cepat dan akurat.

2.3.3 Sistem Informasi Organisasi

Perkembangan teknologi informasi mendorong banyak usaha jasa, termasuk barbershop, untuk mengubah proses pencatatan transaksi dari sistem manual menjadi

sistem digital. Penggunaan sistem *Point of Sales* (POS) memungkinkan seluruh transaksi tersimpan dalam basis data yang terintegrasi sehingga mempermudah proses monitoring pendapatan, pengelolaan stok, serta penyusunan laporan operasional. Selain meningkatkan efisiensi kerja, sistem POS juga menghasilkan data historis yang dapat dimanfaatkan sebagai sumber informasi dalam proses analisis data dan pengambilan keputusan berbasis data (*data-driven decision making*).

2.3.4 Hubungan Organisasi dengan Penelitian

Data transaksi yang dihasilkan dari sistem *Point of Sales* merupakan sumber data utama dalam penelitian ini. Informasi seperti tanggal transaksi, jumlah pelanggan, tarif layanan, cabang, dan total pendapatan harian digunakan sebagai variabel dalam proses pembangunan model *machine learning*. Dengan memanfaatkan data historis tersebut, penelitian ini membandingkan algoritma *Linear Regression* dan *Random Forest Regression* untuk mengetahui model yang memiliki kemampuan terbaik dalam memprediksi pendapatan harian sehingga dapat mendukung proses pengambilan keputusan manajemen berdasarkan data yang tersedia.

UNIVERSITAS
NUSA MANDIRI