

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian terdahulu oleh [2] yang memprediksi jumlah penumpang harian kereta api di Stasiun Surabaya Gubeng menggunakan metode Monte Carlo dengan berbagai tipe pembangkit bilangan acak (*multiplicative, mixed, dan random uniform*). Data yang digunakan mencakup 140 hari dari 16 Mei 2022 hingga 2 Oktober 2022, yang dikelompokkan berdasarkan hari dalam seminggu. Tingkat akurasi prediksi diukur dengan *Mean Absolute Percentage Error (MAPE)*, menunjukkan hasil yang bervariasi untuk setiap hari, dengan nilai MAPE tertinggi pada Senin (25.25%) dan terendah pada Minggu (2.62%). Hasil analisis menunjukkan bahwa metode Monte Carlo dapat memberikan prediksi yang cukup akurat, terutama pada hari Kamis, Sabtu, dan Minggu. Kesimpulannya, penelitian ini memberikan kontribusi dalam perbaikan layanan dan pengelolaan penumpang di Stasiun Surabaya Gubeng melalui alat prediksi yang berbasis pada ketidakpastian probabilitas yang diaplikasikan secara efisien.

Selanjutnya penelitian oleh [3] yang memprediksi jumlah penumpang yang akan menggunakan Pelabuhan Tanjung Perak selama periode Lebaran tahun 2024, menggunakan metode regresi linier sederhana untuk memahami tren dan pola yang ada. Metode penelitian yang digunakan meliputi pengumpulan data relevan tentang jumlah penumpang dari tahun-tahun sebelumnya dan penerapan teknik regresi linier untuk memodelkan hubungan antara variabel independen, seperti tanggal Lebaran dan kondisi ekonomi, dengan variabel dependen, yaitu jumlah penumpang. Hasil dari analisis menunjukkan bahwa prediksi jumlah penumpang untuk tahun 2024 menunjukkan peningkatan yang signifikan, mencerminkan tren yang berkembang akibat faktor-faktor yang telah dianalisis, yang pada gilirannya dapat digunakan oleh pengelola pelabuhan dan otoritas transportasi dalam merencanakan dan mengoptimalkan layanan selama periode liburan.

Kemudian penelitian oleh [4] untuk mengestimasi jumlah penumpang pada PT. Pinem Lau Guna Medan menggunakan metode regresi linier berganda yang

merupakan salah satu teknik dalam data mining. Peneliti memanfaatkan data jumlah hari libur, jumlah keberangkatan, dan harga tiket sebagai variabel independen untuk memprediksi jumlah penumpang sebagai variabel dependen. Data diperoleh melalui observasi langsung dan wawancara, lalu dianalisis menggunakan persamaan regresi linier berganda. Hasil analisis menunjukkan bahwa metode ini mampu menghasilkan estimasi yang mendekati data aktual dengan tingkat akurasi yang tinggi. Kesimpulannya, metode regresi linier berganda terbukti efektif untuk membantu PT. Pinem Lau Guna Medan dalam mengelola data penumpang dan meningkatkan efisiensi operasional melalui estimasi yang akurat. Implementasi lebih lanjut direkomendasikan dalam bentuk aplikasi berbasis web untuk mempercepat dan mempermudah proses estimasi jumlah penumpang.

Selanjutnya pada penelitian yang dilakukan oleh [5] membahas penerapan metode regresi linear berganda dalam data mining untuk mengestimasi jumlah penumpang pada PT Rajawali Citra Transportasi Medan. Metode ini dipilih karena mampu memanfaatkan data historis, seperti jumlah hari libur, keberangkatan, dan harga tiket, untuk membuat prediksi jumlah penumpang. Data diperoleh melalui observasi dan wawancara, kemudian dianalisis menggunakan regresi linear berganda untuk membangun persamaan prediksi. Hasil analisis menunjukkan bahwa metode ini efektif menghasilkan estimasi jumlah penumpang dengan mempertimbangkan hubungan signifikan antara variabel-variabel yang memengaruhi. Berdasarkan pengujian kasus, metode ini mampu memberikan hasil estimasi yang mendekati data aktual. Kesimpulannya, regresi linear berganda dapat diimplementasikan sebagai alat bantu yang andal dalam pengambilan keputusan operasional dan pengelolaan transportasi di PT Rajawali Citra Transportasi Medan.

Kemudian penelitian yang dilakukan oleh [6] mengevaluasi performa model regresi linier dalam memprediksi jumlah penumpang bus TransJakarta menggunakan pendekatan *Knowledge Discovery in Database* (KDD). Metode penelitian melibatkan lima tahapan utama, yaitu *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* dengan memanfaatkan algoritma regresi linier dalam perangkat lunak RapidMiner. Dataset yang digunakan terdiri dari 1232 *record* jumlah penumpang bus TransJakarta yang diambil dari *Open Data*

Jakarta, kemudian dilakukan proses pembersihan dan transformasi data nominal menjadi numerik. Hasil dari pembangunan model prediktif menunjukkan bahwa nilai *Root Mean Squared Error (RMSE)* yang diperoleh adalah 115.306,990, yang menandakan bahwa model memiliki tingkat kesalahan prediksi yang relatif kecil dan stabil. Penelitian ini menyimpulkan bahwa regresi linier dapat memberikan gambaran yang akurat terhadap tren penumpang dan dapat dimanfaatkan dalam perencanaan operasional transportasi perkotaan, khususnya dalam meningkatkan kualitas layanan dan efisiensi TransJakarta di DKI Jakarta.

2.2 Landasan Teori

2.2.1 Kereta Api

Kereta api merupakan salah satu moda transportasi darat yang memiliki peran strategis dalam sistem transportasi publik. Sebagai sarana angkutan massal, kereta api memiliki karakteristik unik yang membedakannya dari moda transportasi lainnya. Menurut Undang-Undang Republik Indonesia Nomor 23 Tahun 2007 tentang Perkeretaapian, kereta api didefinisikan sebagai sarana perkeretaapian dengan tenaga gerak berupa lokomotif atau *railbus* yang memiliki karakteristik khusus untuk mengangkut penumpang dan barang [7].

Sistem transportasi kereta api mencakup beberapa komponen penting, meliputi infrastruktur (rel, stasiun, sinyal), sarana (kereta, lokomotif), dan operasional. Efisiensi dan kapasitas angkut yang besar menjadikan kereta api pilihan transportasi yang efektif, terutama untuk perjalanan jarak menengah dan panjang. Kemampuan mengangkut penumpang dalam jumlah besar dengan konsumsi energi yang relatif rendah membuat kereta api menjadi moda transportasi yang ramah lingkungan.

2.2.2 Regresi Linier

Regresi linier memodelkan hubungan variabel independen dan variabel dependen, proses prediksi regresi linier dengan memasukkan nilai-nilai variabel independen ke persamaan regresi [8]. Regresi linier memiliki kelebihan bisa mensimulasikan hubungan linear dari dua atau lebih variabel. Pada penelitian ini, regresi linier yang akan digunakan adalah regresi linier sederhana. Regresi linier sederhana menggunakan satu variabel dependen (Y) dan satu variabel independen

(X). Persamaan matematisnya dapat ditulis sebagai berikut:

$$Y = a + bX$$

Dimana Y adalah variabel dependen, a adalah konstanta, X adalah variabel independen dan b adalah koefisien Regresi. Pada penelitian ini yang variabel dependen (Y) adalah jumlah penumpang kereta api, jumlah penumpangnya bisa diambil dari jumlah penumpang berdasarkan kelas gerbong keretanya, bisa juga jumlah penumpang kereta untuk bulan tertentu. Sedangkan yang menjadi variabel independen (X) adalah waktu, waktu yang digunakan dapat menggunakan hari, bulan atau tahun. Data yang digunakan dalam penelitian ini adalah jumlah penumpang kereta api dari tahun 2022-2024.

Rumus Mencari konstanta a :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

Dimana $\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data.

Rumus mencari koefisien b :

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Dimana $\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data.

2.2.3 Metrik Evaluasi

Penggunaan metrik evaluasi sangatlah penting dalam analisis data dan penerapan metode regresi linier untuk prediksi. Metrik evaluasi ini digunakan untuk menilai seberapa baik model yang telah dibangun mampu melakukan prediksi terhadap data baru atau yang belum pernah dilihat sebelumnya. Metrik evaluasi memberikan gambaran mengenai performa model prediksi dalam menangkap pola dan hubungan yang ada dalam data. Selain itu, metrik evaluasi juga membantu dalam membandingkan berbagai model yang berbeda sehingga dapat dipilih model yang paling efektif dan efisien.

Pada penelitian ini, beberapa metrik evaluasi yang digunakan untuk mengukur kinerja model regresi linier antara lain sebagai berikut:

1. *Mean Absolute Error (MAE)* adalah rata-rata dari selisih absolut antara nilai aktual dan nilai prediksi. *Mean Absolute Error (MAE)* memberikan gambaran tentang seberapa besar kesalahan prediksi secara keseluruhan [9]. Rumus *Mean Absolute Error (MAE)* adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

Dimana n adalah jumlah sampel dalam data, y^i adalah aktual dan \hat{y}^i adalah nilai prediksi.

2. *Mean Squared Error (MSE)* adalah rata-rata dari kuadrat selisih antara nilai aktual dan nilai prediksi. *Mean Squared Error (MSE)* memberikan bobot lebih besar pada kesalahan yang lebih besar [10]. Rumus *Mean Squared Error (MSE)* adalah sebagai berikut:

$$MSE = \sum \frac{(\hat{y} - y)^2}{n}$$

Dimana n adalah jumlah sampel dalam data, y^i adalah aktual dan \hat{y}^i adalah nilai prediksi.

3. *Root Mean Squared Error (RMSE)* adalah akar kuadrat dari *Mean Squared Error (MSE)*. *Root Mean Squared Error (RMSE)* memiliki interpretasi yang serupa dengan MSE, namun lebih mudah dipahami karena berada dalam satuan yang sama dengan data asli. Rumus *Root Mean Squared Error (RMSE)* adalah sebagai berikut:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Dimana n adalah jumlah data, i adalah urutan data pada database, y_i adalah aktual dan \hat{y}_i adalah nilai prediksi.

Metrik-metrik evaluasi ini sangat penting untuk menentukan seberapa baik model regresi linier yang diterapkan dalam memprediksi jumlah penumpang kereta api. Dengan menggunakan *Mean Absolute Error (MAE)*,

Mean Squared Error (MSE), dan *Root Mean Squared Error (RMSE)* peneliti dapat menilai keakuratan prediksi yang dibuat dan mengidentifikasi area yang perlu diperbaiki.

2.2.4 *Machine Learning*

Machine Learning adalah bagian dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan sistem yang dapat belajar dan membuat keputusan berdasarkan data. Dalam *machine learning*, komputer tidak diprogram secara eksplisit, melainkan dilatih untuk mengenali pola dari data yang diberikan [11]. Salah satu pendekatan dalam *machine learning* adalah *supervised learning*, di mana model dilatih dengan data yang memiliki label. Metode regresi linier termasuk dalam kategori ini dan sering digunakan untuk tugas prediksi nilai kontinu. Dengan pendekatan ini, *machine learning* memberikan kontribusi penting dalam pengolahan data skala besar dan pengambilan keputusan berbasis data.

Beberapa jenis pendekatan dalam *machine learning*, antara lain:

1. *Supervised Learning*

Pada jenis *Machine Learning* ini, model dilatih menggunakan data yang sudah memiliki label. Model belajar dari pasangan *input-output* untuk memprediksi nilai atau klasifikasi baru. Contohnya adalah regresi linier untuk prediksi nilai kontinu dan klasifikasi untuk pengelompokan data.

2. *Unsupervised Learning*

Dalam jenis *Machine Learning* ini, data yang digunakan tidak memiliki label. Tujuan utamanya adalah menemukan struktur atau pola tersembunyi dalam data. Contohnya adalah algoritma clustering seperti K-Means dan analisis komponen utama (PCA).

3. *Semi-Supervised Learning*

Jenis *Machine Learning* ini menggabungkan sejumlah kecil data berlabel dengan sejumlah besar data tanpa label. Cocok digunakan ketika pelabelan data membutuhkan biaya atau waktu yang besar.

4. *Reinforcement Learning*

Machine Learning jenis ini melibatkan agen yang belajar melalui interaksi dengan lingkungan. Agen menerima umpan balik berupa *reward* atau

punishment dan bertujuan memaksimalkan total *reward* dalam jangka panjang.

2.2.5 Penumpang Kereta Api

Penumpang kereta api merupakan komponen penting dalam sistem transportasi publik. Jumlah penumpang mencerminkan tingkat penggunaan dan efektivitas layanan kereta api. Dalam konteks perencanaan transportasi, data jumlah penumpang digunakan sebagai dasar dalam pengambilan keputusan operasional seperti penentuan frekuensi perjalanan, kapasitas gerbong, serta estimasi pendapatan. Jumlah penumpang dipengaruhi oleh berbagai faktor, seperti musim liburan, hari kerja, event tertentu, harga tiket, dan kondisi sosial ekonomi. Oleh karena itu, prediksi jumlah penumpang menjadi sangat krusial untuk meningkatkan kualitas dan efisiensi layanan [12].

2.2.6 Python

Python adalah bahasa pemrograman yang banyak digunakan dalam analisis data, statistika, dan pengembangan model prediktif karena sintaksisnya yang sederhana dan fleksibel. *Python* memiliki berbagai pustaka (*library*) yang mendukung kebutuhan analisis data dan machine learning, seperti *Pandas* untuk manipulasi data, *NumPy* untuk operasi numerik, *Matplotlib* dan *Seaborn* untuk visualisasi, serta *Scikit-learn* untuk membangun dan mengevaluasi model prediksi [13]. Dalam penelitian ini, *Python* digunakan sebagai alat bantu utama untuk membangun model regresi linier, mengolah data historis, serta mengevaluasi performa model secara kuantitatif.

2.2.7 Prediksi

Prediksi adalah proses memperkirakan nilai atau kejadian yang akan terjadi di masa depan berdasarkan pola atau tren dari data historis. Dalam dunia transportasi, prediksi jumlah penumpang berperan penting untuk mendukung perencanaan jangka pendek maupun jangka panjang. Dengan model prediksi yang akurat, perusahaan kereta api dapat mengoptimalkan alokasi sumber daya, menghindari kekurangan kapasitas, dan meningkatkan kepuasan pelanggan [14]. Salah satu metode statistik yang umum digunakan dalam proses ini adalah regresi linier, yang mampu memodelkan hubungan antara variabel independen dan

variabel dependen secara matematis.

2.2.8 *Visual Studio Code*

Visual Studio Code merupakan editor kode sumber yang dikembangkan oleh Microsoft dan banyak digunakan dalam pengembangan perangkat lunak karena sifatnya yang ringan, gratis, dan mendukung lintas platform. Editor ini dilengkapi dengan berbagai fitur seperti *syntax highlighting*, *debugging*, terminal terintegrasi, serta dukungan ekstensi untuk berbagai bahasa pemrograman, termasuk *Python*. Dalam konteks penelitian ini, *Visual Studio Code* digunakan sebagai lingkungan kerja utama untuk menulis, menjalankan, dan menguji kode *Python* dalam membangun model regresi linier. Kemampuannya untuk terintegrasi dengan pustaka analisis data seperti *Pandas*, *NumPy*, dan *Scikit-learn* menjadikan *Visual Studio Code* sebagai alat yang efektif dan efisien dalam proses analisis data dan pengembangan model prediktif [15].