

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini dilakukan dalam pendekatan kuantitatif. Metode penelitian kuantitatif merujuk kepada filosofi positivisme dan digunakan untuk menyelidiki populasi dan sampel tertentu dengan tujuan menguji hipotesis yang telah ditetapkan (Sugiyono, 2009). Metode kuantitatif juga disebut demikian karena data penelitian berbentuk angka dan analisis menggunakan statistik. Proses penelitian kuantitatif bersifat linier, yaitu alur penelitian berjalan secara berurutan dan tidak mengalami pengulangan.

3.2 Prosedur Penelitian

Prosedur penelitian diacu sebagai serangkaian langkah atau tahapan sistematis yang diikuti dalam pelaksanaan suatu penelitian (Sugiyono, 2009). Tujuan dari prosedur penelitian ini adalah untuk memastikan bahwa penelitian dilakukan dengan metode yang kredibel, terstruktur, dan menghasilkan informasi yang akurat dan bermakna.

3.2.1 Hipotesis

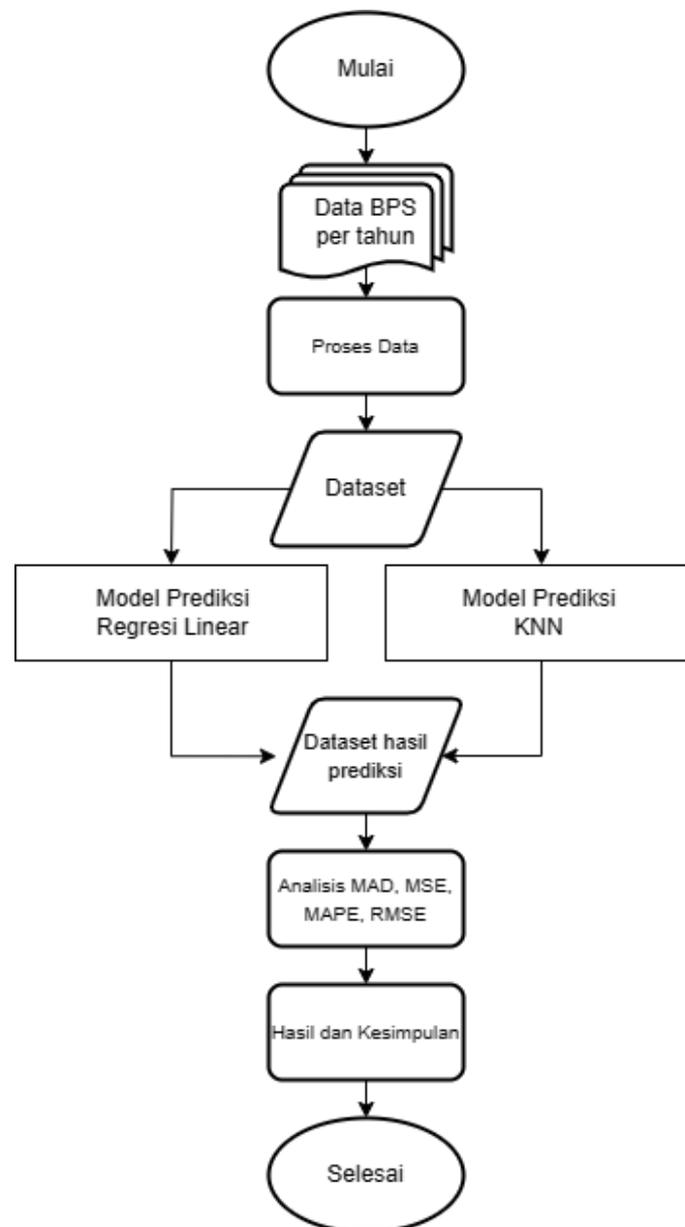
Jawaban sementara terhadap rumusan masalah yang dikemukakan merupakan bagian dari hipotesis (Sugiyono, 2009). Dalam penelitian kuantitatif, perlu adanya hipotesis yang berlandaskan pada suatu teori yang valid. Dalam penelitian ini, beberapa hipotesis yang dikemukakan adalah sebagai berikut:

1. Jumlah pelanggan telekomunikasi akan bertambah pesat pada tahun 2030.
2. Perubahan pelanggan yang dibuat oleh model regresi linear akan lebih tinggi atau memiliki *gap* yang lebih besar dibandingkan *k - nearest neighbors*, karena regresi linear memiliki pola linear kenaikan yang tetap.
3. Algoritma regresi linear bekerja lebih baik dari algoritma *k - nearest neighbors* dikarenakan fungsi utamanya sebagai algoritma regresi,

sedangkan *k - nearest neighbors* memiliki fungsi utama sebagai algoritma klasifikasi.

3.2.2 Alur Pengolahan Data

Dalam pengerjaan penelitian ini, perlu sebuah gambaran tentang bagaimana penelitian ini akan dilakukan. Untuk memudahkan pengamatan maka alur pengolahan data dapat dilihat pada gambar 3.1.



Gambar 3.1 Alur pengolahan data penelitian

1. Pengumpulan dan pemilihan variabel

Data yang digunakan dalam penelitian ini menggunakan data Statistik Telekomunikasi Indonesia yang diterbitkan hampir tiap tahun oleh Badan Pusat Statistik. Data yang diambil adalah data pelanggan menurut penyelenggaraan jaringan yang dibagi kedalam dua jenis jaringan yaitu *wired telecommunication* dan *wireless mobile telecommunication* dari tahun 2013 hingga keluaran terbarunya yaitu tahun 2021.

2. *Preprocessing* data

Setelah pemilihan variabel data dilakukan, data yang didapatkan dari Statistik Telekomunikasi Indonesia dari tahun 2013 sampai 2021 diambil dan dipindahkan menjadi satu ke dalam sebuah dataset *google spreadsheet* yang akan digunakan sebagai acuan data aktual dalam penelitian ini.

3. Prediksi Algoritma

Data aktual yang telah didapatkan sebelumnya kemudian diproses menggunakan *python* dengan bantuan *library numpy* untuk memulai proses prediksi. Proses ini dibagi menjadi dua yaitu proses prediksi menggunakan algoritma regresi linear dan proses prediksi menggunakan algoritma *k - nearest neighbors*. Setelah proses ini maka akan menghasilkan sebuah data prediksi yang akan dianalisis lebih lanjut.

4. Analisis evaluasi model

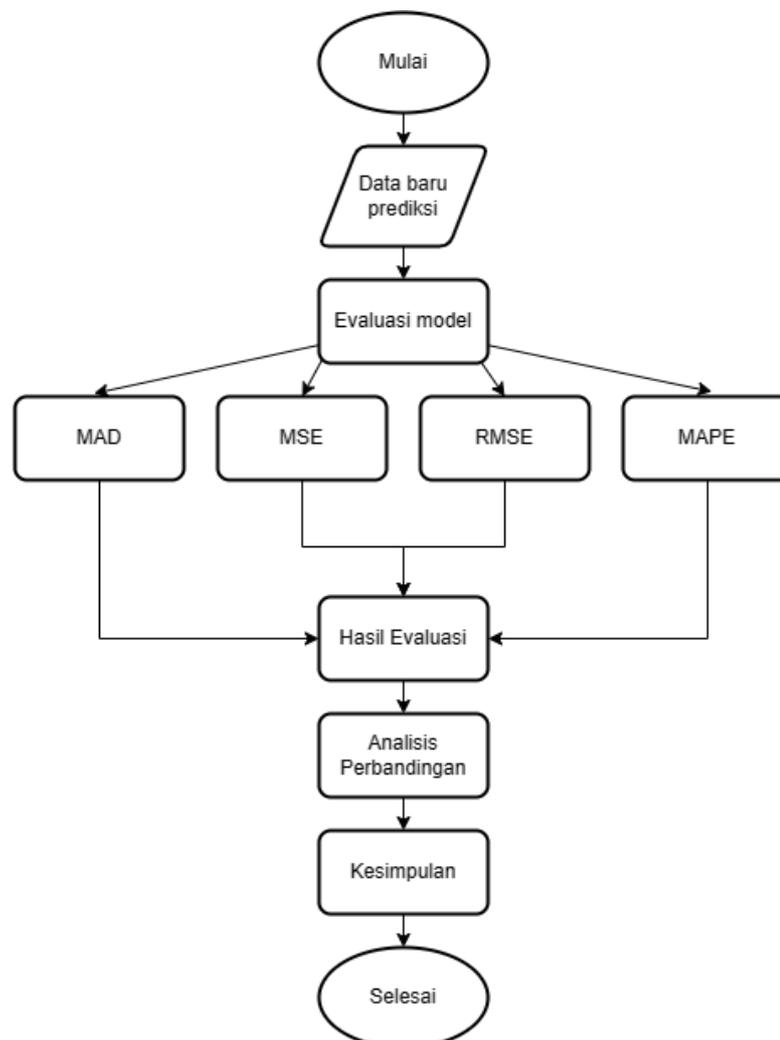
Data hasil prediksi sebelumnya kemudian dianalisis menggunakan metode *Measure Absolute Deviation* (MAD), *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Nilai dari analisis tersebut kemudian dibandingkan untuk dicari mana algoritma yang lebih efisien dan benar.

5. Kesimpulan dan pembahasan

Setelah melakukan analisis menggunakan matriks evaluasi untuk model diatas, maka data prediksi dapat disimpulkan dan dibahas mengenai perbedaan masing - masing kinerja algoritma yang digunakan.

3.2.3 Alur Evaluasi Data

Dalam pemodelan prediksi atau *forecasting* menggunakan algoritma dalam machine learning, model dari algoritma yang telah dibuat perlu diperiksa dan dievaluasi untuk mencari tahu seberapa efektif model algoritma tersebut dalam melakukan prediksi data baru. Pada penelitian ini, evaluasi model prediksi atau *forecasting* pada algoritma dilakukan menggunakan empat jenis model evaluasi. Model - model tersebut yang diantaranya *mean absolute deviation* (MAD), *mean square error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE). Untuk memudahkan pengamatan maka alur evaluasi data dapat dilihat pada gambar 3.2.



Gambar 3.2. Alur pengolahan evaluasi matriks

1. Data prediksi baru

Setelah model algoritma regresi linear dan *k - nearest neighbors* menjalankan proses prediksi, maka data prediksi akan ditampilkan sebagai *output* dan siap untuk dievaluasi.

2. Evaluasi model

Evaluasi model pada penelitian ini menggunakan empat matriks evaluasi yang diantaranya *mean absolute deviation* (MAD), *mean square error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE). Hasil dari evaluasi ini akan menampilkan nilai evaluasi yang nantinya akan dibandingkan dan dianalisis.

3. Kesimpulan

Setelah nilai evaluasi dibandingkan dan dianalisis, selanjutnya model analisis yang paling baik dalam model prediksi atau *forecasting* sudah dapat ditentukan dan dibuat kesimpulannya.

3.3 Metode Pengumpulan Data

Dalam penelitian ini, pengumpulan data penelitian didapatkan dari beberapa tahap untuk mencari dan menjawab permasalahan.

1. Studi Literatur

Pengumpulan data dengan metode studi literatur didapatkan dengan cara mengumpulkan data yang berasal dari beberapa sumber bacaan seperti buku, jurnal nasional, jurnal internasional, dan internet yang relevan dengan penelitian yang dilakukan.

2. Diskusi

Diskusi yang berkaitan dengan topik dan tema skripsi yang sudah dilakukan dengan dosen pembimbing mengenai jenis dan sumber data yang akan diuji pada penelitian ini.

3. Data Collecting

Data Collecting merupakan proses pengambilan data dari sumber yang telah didapatkan dengan cara pencarian sumber data yang dapat digunakan dalam penelitian ini.

3.4 Data Penelitian

Data yang digunakan adalah data yang dikeluarkan oleh Badan Pusat Statistik yang membahas tentang Statistik Telekomunikasi Indonesia dari tahun 2009 hingga tahun 2021. Data populasi yang diambil dari Statistik Telekomunikasi Indonesia yang merupakan jenis data *time-series* tentang jumlah pelanggan telekomunikasi yang dibedakan dari penyelenggaraan jaringan telekomunikasinya. Data sampel penelitian ini merupakan data pelanggan telekomunikasi menurut penyelenggaraan jaringan yang dibagi menjadi dua data *time-series* yang merupakan data pelanggan *wired telecommunication*, dan data pelanggan *wireless mobile telecommunication*. Data yang digunakan berjumlah sembilan sampel dari masing - masing variabel (*wired telecommunication* dan *wireless mobile telecommunication*).

Tabel 3.1. Data pelanggan telekomunikasi menurut BPS

No.	Tahun	Jumlah Pelanggan Wired Telecommunication	Jumlah Pelanggan Wireless Mobile Telecommunication
1	2013	1.008.562	313.226.914
2	2014	9.885.971	325.582.891
3	2015	10.378.037	338.948.340
4	2016	10.752.912	385.573.398
5	2017	11.053.303	435.193.605
6	2018	8.341.933	319.434.605
7	2019	9.513.751	341.277.549
8	2020	9.307.494	355.620.388
9	2021	9.019.476	365.872.608

3.5 Analisis Data

Proses pencarian dan pengorganisasian data secara sistematis yang berasal dari hasil wawancara, catatan lapangan, dan dokumentasi dalam teknik analisis data melibatkan penggolongan data ke dalam kategori, penguraian dalam kelompok tertentu, sintesis, pembentukan pola, pemilihan objek yang signifikan untuk studi, serta penyusunan kesimpulan agar dapat dimengerti oleh diri sendiri maupun orang lain (Sugiyono, 2009). Dalam penelitian ini, teknik analisis data diterapkan melalui pengujian menggunakan empat matriks evaluasi, yaitu *mean absolute deviation* (MAD), *mean square error* (MSE), *mean absolute percentage error* (MAPE), dan *root mean square error* (RMSE). MSE merepresentasikan rata-rata perbedaan kuadrat antara nilai yang diprediksi dan nilai yang diamati, sementara RMSE merupakan akar kuadrat dari MSE, dan MAPE mewakili rata-rata perbedaan persentase absolut antara nilai yang diprediksi dan nilai aktual (Sugiyono, 2009).

3.5.1 Mean Absolute Deviation

Mean Absolute Deviation (MAD) adalah pengukuran statistik yang digunakan untuk sejauh mana titik-titik data dalam suatu kumpulan data memiliki perbedaan dari *mean* (rata-rata) kumpulan data tersebut (Elsayed, 2015). Nilai pada MAD didefinisikan baik atau efisien apabila nilai yang ditampilkan semakin kecil nilainya. Cara kerja dari jenis evaluasi MAD adalah dengan cara mengukur sebaran data disekitar nilai rata - rata (mean) yang akan berguna terutama ketika terdapat nilai - nilai yang jauh atau *outliers* pada sebaran data. Dalam perumusannya sendiri, MAD memiliki rumus persamaan dibawah ini (5).

$$MAD = \frac{\sum |X_i - \text{Mean}|}{n} \quad (5)$$

- X_i adalah titik - titik pada data individual.
- Mean merupakan nilai rata-rata dari suatu kumpulan data.
- $|X_i - \text{Mean}|$ menunjukkan nilai selisih absolut antara titik data dan nilai rata-rata.
- n menunjukkan jumlah total data yang ada

3.5.2 Mean Square Error

Mean Squared Error (MSE) adalah metrik statistik yang banyak digunakan untuk mengevaluasi kinerja model, terutama dalam tugas regresi dan peramalan. Metrik ini mengukur rata-rata kuadrat selisih antara data prediksi dan data aktual (Hodson et al., 2021). MSE bekerja dengan cara menghitung nilai *mean* atau rata-rata dari kuadrat selisih antara nilai prediksi dengan nilai data asli, dan MSE dikatakan baik apabila nilai rendah atau mendekati 0. Pendekatan dengan metode ini mengatur tentang kesalahan peramalan yang besar dan mengakibatkan muncul nilai ekstrim atau *outlier* yang muncul akibat kesalahan - kesalahan itu dikuadratkan. Secara perumusan, metode *mean square error* dituliskan seperti dibawah ini (6).

$$MSE = \Sigma (X_i - X_o)^2 / n \quad (6)$$

- X_i menunjukkan nilai prediksi dari suatu model.
- X_o menunjukkan nilai sebenarnya atau data aktual.
- Σ merupakan tanda sigma yang menyatakan penjumlahan dari seluruh data points.
- $(X_i - X_o)^2$ merujuk pada perhitungan kuadrat dari selisih antara nilai prediksi dan nilai sebenarnya.
- n adalah jumlah total data yang ada.

3.5.3 Root Mean Square Error

Root Mean Square Error atau RMSE adalah bentuk lanjutan dari *mean square error*. RMSE merupakan metrik yang digunakan untuk mengevaluasi akurasi model prediksi, terutama dalam tugas regresi dan peramalan (Yuan, 2022). Secara definisi, RMSE menyatakan sebagai penjumlahan kuadrat error atau selisih antara nilai sebenarnya atau nilai data aktual dengan nilai prediksi kemudian membagi jumlah tersebut dengan banyaknya waktu data *forecasting* kemudian ditarik nilai akarnya. RMSE menyatakan bahwa semakin kecil nilai evaluasi maka semakin baik model yang dibuat. Secara perumusan, RMSE dapat dituliskan seperti dibawah ini (7).

$$\text{RMSE} = \sqrt{(\text{MSE})} = \sqrt{(\Sigma (X_i - X_o)^2 / n)} \quad (7)$$

- X_i menunjukkan nilai prediksi dari suatu model.
- X_o menunjukkan nilai sebenarnya atau data aktual.
- Σ merupakan tanda sigma yang menyatakan penjumlahan dari seluruh data points.
- $(X_i - X_o)^2$ merujuk pada perhitungan kuadrat dari selisih antara nilai prediksi dan nilai sebenarnya.
- n adalah jumlah total data yang ada.

3.5.4 Mean Absolute Percentage Error

Mean Absolute Percentage Error atau MAPE memiliki arti sebagai rata-rata kesalahan persentase mutlak. Secara definisi, MAPE memiliki arti sebagai nilai absolut dari persentase kesalahan data terhadap rata-rata. Penggunaan *Mean Absolute Percentage Error* dalam mengevaluasi hasil peramalan memungkinkan kita untuk menilai tingkat akurasi antara angka peramalan dan nilai aktual (Nabillah & Ranggadara, 2020). Dalam metode evaluasi ini, semakin kecil nilai evaluasi yang dihasilkan, semakin baik model regresi yang dibuat. Secara matematis, MAPE dirumuskan dibawah ini (8).

$$\text{MAPE} = (\Sigma |(X_i - X_o) / X_o| / n) * 100 \quad (8)$$

- X_i menunjukkan nilai prediksi dari suatu model
- X_o menunjukkan nilai sebenarnya atau nilai pada sumber data
- $|(X_i - X_o) / X_o|$ merujuk untuk menghitung kesalahan persentase mutlak untuk setiap titik data.
- n adalah jumlah total titik data yang ada.

3.6 Alat Penelitian

Dalam mengerjakan proses penelitian, diperlukan perangkat yang memadai untuk mendukung pengerjaan penelitian ini. Terdapat dua jenis alat dalam penelitian ini yang diantaranya perangkat lunak dan perangkat keras.

3.6.1 Perangkat Lunak

Untuk memberikan pemahaman yang lebih komprehensif tentang penelitian ini, diperlukan penguraian rinci terkait kebutuhan perangkat lunak yang dibutuhkan. Dalam upaya mencapai tujuan penelitian dengan efektif, perangkat lunak yang mampu mengelola, menganalisis, dan memvisualisasikan data secara akurat menjadi suatu keharusan. Disamping itu, dukungan terhadap penggunaan metode analisis yang sesuai dengan fokus penelitian juga menjadi hal yang penting.

Adapun perangkat lunak yang digunakan adalah sebagai berikut :

1. *Windows 10 Pro Operating System*
2. *Browser (Microsoft Edge)*
3. *Google Sheets*
4. *Google Colaboratory (Collab)*

3.6.2 Perangkat Keras

Memahami kebutuhan infrastruktur perangkat keras dalam konteks ini adalah langkah awal yang penting. Kehadiran perangkat keras yang sesuai dan memadai memiliki peran yang tak tergantikan dalam menjalankan berbagai tahapan penelitian dengan lancar dan efisien. Mulai dari pengumpulan data hingga pengolahan, serta pelaksanaan analisis mendalam, kesesuaian dan kinerja perangkat keras akan berdampak signifikan terhadap kualitas dan ketepatan hasil penelitian.

Adapun perangkat keras yang digunakan adalah laptop dengan spesifikasi:

1. Processor Intel(R) Core(TM) i3-3110M CPU @ 2.40GHz 2.40 GHz
2. RAM 4,00 GB
3. *64-bit Operating System, x64-based processor*