

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Komparasi atau perbandingan antara algoritma - algoritma merupakan langkah penting dalam bidang ilmu komputer dan *data science*. Para peneliti melakukan komparasi algoritma dengan beberapa alasan seperti performa dan efektifitas. Kasus terbaik efisiensi algoritma adalah algoritma yang bekerja lebih cepat diantara semua kemungkinan input dengan ukuran yang sama pada tiap - tiap algoritma (Levitin, 2012). Melalui komparasi algoritma, kita dapat mengevaluasi dan membandingkan performa serta efektivitas dari berbagai pendekatan atau teknik yang digunakan dalam pemrosesan data atau pemecahan masalah. Dengan mengetahui kekuatan dan kelemahan masing-masing algoritma, kita dapat memilih algoritma yang paling cocok untuk tugas atau masalah tertentu. Selain itu juga, komparasi algoritma memungkinkan kita untuk melakukan generalisasi dan membuat kesimpulan yang lebih luas tentang performa algoritma dalam konteks yang lebih umum. Dengan mengumpulkan dan menganalisis data dari berbagai studi komparatif, kita dapat memperoleh pemahaman yang lebih mendalam tentang kekuatan dan kelemahan algoritma-algoritma tertentu, serta menerapkan temuan tersebut dalam konteks yang lebih luas.

Algoritma adalah sebuah urutan instruksi dasar yang eksplisit, tepat, tidak ambigu, dan dapat dieksekusi secara mekanis, biasanya dimaksudkan untuk mencapai tujuan tertentu (Erickson, 2019). Studi tentang algoritma mendefinisikan suatu algoritma sebagai urutan instruksi yang tidak ambigu untuk memecahkan masalah, dan digunakan untuk mendapatkan output yang diperlukan untuk setiap input yang sah dalam waktu yang terbatas (Kleinberg & Tardos, 2006). Penelitian terkini dalam bidang algoritma telah menghasilkan berbagai teknik yang inovatif dan canggih untuk menangani tantangan dalam analisis data. Untuk menggali informasi berharga dari data tersebut, diperlukan algoritma-algoritma yang efisien dan efektif dalam analisis data. Namun, meskipun ada banyak algoritma yang tersedia, setiap algoritma memiliki

kelebihan dan keterbatasan masing-masing. Oleh karena itu, penting untuk melakukan tinjauan terkini tentang algoritma-algoritma yang ada saat ini guna memahami keunggulan, kelemahan, dan aplikasi yang tepat untuk setiap algoritma. Sehingga, perlu bagi seorang pengguna algoritma dalam mengetahui karakteristik tiap - tiap algoritma. Salah satu hal yang dapat dilakukan yaitu dengan melakukan komparasi antar algoritma, sehingga pengguna dapat mengetahui secara langsung dan teknis mengenai perbedaan antar algoritma.

Dalam penelitian ini, perbandingan operasi dan hasil akan diuji menggunakan dua algoritma terpilih yaitu regresi linear dan *k - nearest neighbors*. Diungkapkan bahwa tiap - tiap algoritma memiliki karakter dan cara kerjanya tersendiri, namun pada prosesnya beberapa algoritma memiliki proses yang beririsan satu sama lain (Kabanda, 2019). Pemilihan kedua algoritma tersebut tidak semata - mata hanya karena kelebihan dan kekurangan dari algoritma tersebut, tetapi terdapat sebuah dasar yang menjelaskan bahwa kedua algoritma tersebut memiliki cara kerja yang sangat berbeda satu sama lain. Kedua algoritma tersebut dinilai cocok untuk mengolah data yang bersifat kecil. Algoritma regresi linear dan *k - nearest neighbors* menggunakan teknik pendekatan yang berbeda. Regresi linear bertujuan untuk menemukan hubungan linear terbaik antara variabel independen dan variabel dependen (Khairiati et al., 2022). Sementara itu, *k-nearest neighbors* menggunakan nilai-nilai tetangga terdekat untuk memprediksi (Dwi Selvy Wisdayani, 2019). Maka dari itu, karakteristik yang berbeda satu sama lain, kedua algoritma ini dipilih sebagai pertimbangan algoritma mana yang dapat dijadikan sebuah model prediksi yang lebih baik dari satu sama lainnya.

Pembuatan model prediksi dari sebuah algoritma dapat dibangun menggunakan bantuan dari *machine learning*. Penggunaan *machine learning* pada sebuah model prediksi melibatkan keterkaitannya dengan kemampuan komputasi dan pondasi matematika. Jika kita menggunakan perbandingan, kita bisa memandang dasar-dasar matematika sebagai akal pikiran dan kemampuan komputasi sebagai tenaga serta sistem yang menggerakkannya. Kedua elemen ini saling berinteraksi dan saling mendukung satu sama lain. Kemampuan komputasi mengacu pada keterampilan dalam menggunakan bahasa pemrograman atau alat

berbasis komputer khusus untuk memproses data (Ibnu Daqiqil Id, 2021). Salah satu bahasa pemrograman yang biasa digunakan adalah *python*. *Python* sering digunakan karena bahasa ini mudah, gratis, dan interpretatif multiguna. Selain itu juga, *python* dapat digunakan secara *cross-platform*, mendukung multi-paradigma, dan memiliki banyak *library* dan *package manager*. Selain itu juga, penggunaan *machine learning* sebagai model prediksi dapat diunggulkan dalam masalah efisiensi pengerjaan model.

Dalam pengujian model prediksi pada masing - masing algoritma, kasus yang digunakan untuk mengukur akurasi model tersebut menggunakan kasus pelanggan telekomunikasi. Hal tersebut melihat isu telekomunikasi yang sedang ramai terjadi di Indonesia saat ini, yaitu tentang korupsi pengadaan *Base Transceiver Station* atau BTS. Hal tersebut akan berdampak pada kualitas telekomunikasi dalam beberapa tahun kedepan, dimana akan terjadi kemungkinan kenaikan jumlah pelanggan tetapi tidak disertai infrastruktur BTS yang memadai. Selain itu, telekomunikasi juga menjadi suatu bidang yang menjadi pendukung kehidupan sehari - hari pada masa ini. Dengan memprediksi jumlah pelanggan telekomunikasi, diharapkan sektor - sektor lainnya yang berhubungan dengan telekomunikasi dapat menyesuaikan terhadap jumlah yang diprediksi pada beberapa tahun kedepan.

Penelitian ini menggunakan beberapa matriks evaluasi sebagai pengukuran keakuratan model prediksi dalam menghasilkan data prediksi dengan pertimbangan data aktual. Evaluasi model merujuk pada tahapan untuk mengukur kinerja model yang telah dibentuk selama proses pelatihan (Ibnu Daqiqil Id, 2021). Terdapat berbagai cara untuk menilai evaluasi model, yang umumnya tergantung pada jenis tugas yang dilakukan serta informasi yang ingin diperoleh. Sebagai ilustrasi, dalam situasi regresi dan klasifikasi, diperlukan metode yang berbeda karena penilaian keberhasilan model regresi dan klasifikasi memiliki ukuran yang berbeda. Pada klasifikasi, contoh matriks evaluasi model yang cocok digunakan adalah *precision*, *recall*, dan *accuration* karena jenis tersebut mengkategorikan nilai positif dan nilai negatif dari data aktual dan data prediksi. Sedangkan untuk kasus regresi, model evaluasi seperti *mean absolute deviation*, *root mean square error*, dan *mean absolute percentage error* lebih cocok

digunakan karena matriks ini mengukur nilai rata - rata hasil prediksi terhadap nilai data aktual. Berdasarkan penjelasan pada kalimat sebelumnya, penelitian ini menggunakan matriks evaluasi *mean absolute deviation* (MAD), *mean square error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE) untuk pengujian model prediksi yang dirancang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang ditulis sebelumnya, rumusan masalah yang didapatkan adalah sebagai berikut :

1. Bagaimana rancangan model prediksi yang dibuat dengan *python* menggunakan algoritma regresi linear dan *k - nearest neighbors*?
2. Berapa perubahan jumlah pelanggan telekomunikasi berdasarkan hasil prediksi dari algoritma regresi linear dan *k-nearest neighbors*?
3. Bagaimana perbandingan kinerja model prediksi dari algoritma regresi linear dan algoritma *k-nearest neighbors* dalam kasus jumlah pelanggan telekomunikasi?

## 1.3 Tujuan Penelitian

Berdasarkan dengan rumusan masalah serta batasan masalah, penelitian ini memiliki beberapa tujuan seperti berikut :

1. Menggunakan algoritma regresi linear dan *k - nearest neighbors* dalam membangun model prediksi jumlah pelanggan telekomunikasi.
2. Mengetahui perubahan jumlah pelanggan yang ditunjukkan pada hasil dari model prediksi algoritma regresi linear dan *k - nearest neighbors*.
3. Menganalisa perbandingan kinerja model prediksi antara regresi linear dan *k - nearest neighbors* dengan pengukuran empat matriks evaluasi *mean absolute deviation* (MAD), *mean square error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE).

#### 1.4 Batasan Masalah

Berdasarkan rumusan masalah yang telah disebutkan, terdapat beberapa batasan masalah pada penelitian ini yang diantaranya :

1. Terdapat banyak jenis algoritma machine learning, pada penelitian ini algoritma yang digunakan adalah regresi linear dan *k-nearest neighbors*.
2. Perbandingan kedua algoritma akan diuji menggunakan matriks evaluasi yang digunakan untuk model regresi *mean absolute deviation* (MAD), *mean square error* (MSE), *root mean square error* (RMSE), dan *mean absolute percentage error* (MAPE).
3. Terdapat banyak sekali data yang tersedia di internet, pada penelitian ini akan menggunakan data resmi dari Badan Pusat Statistik (BPS) tentang Statistik Telekomunikasi Indonesia
4. Data objektif prediksi dapat diatur hingga seberapa jauh prediksi bisa dilakukan. Pada penelitian ini, data prediksi diobjektifkan berada di antara tahun 2022 hingga 2030, hal tersebut untuk membuat data prediksi yang cukup akurat melihat dari banyaknya sumber data yang digunakan dan disesuaikan dengan algoritma yang digunakan.
5. Terdapat banyak *software* dan bahasa pemrograman coding program yang tersedia, pada penelitian kali ini akan menggunakan *Python* dan *Google Collab*

#### 1.5 Manfaat Penelitian

Terdapat banyak manfaat pada penelitian ini, dengan begitu penulis merincikan manfaat - manfaat tersebut sebagai berikut:

##### 1.5.1 Manfaat Teoritis

Berdasarkan teoritis, penelitian ini bermanfaat sebagai wawasan dalam hal pelanggan telekomunikasi di Indonesia serta penggunaan *machine learning* sebagai metode peramalan yang dapat diandalkan. Selain itu juga, penelitian ini diharapkan dapat menjadi referensi oleh peneliti lainnya dalam melakukan penelitian sejenis.

### 1.5.2 Manfaat Praktis

Manfaat secara praktis penelitian ini dapat bermanfaat sebagai berikut :

1. Bagi penulis

Manfaat praktis bagi penulis secara pribadi dapat menambah pengalaman dalam pengerjaan project sejenis dan lebih mendalami terkait algoritma regresi linear dan *k - nearest neighbors* itu sendiri.

2. Bagi pengembangan ilmu

Manfaat praktis untuk pengembangan ilmu pada penelitian ini diharapkan dapat menjadi wawasan tambahan dalam aspek - aspek lainnya yang dapat dikaitkan dengan pengetahuan dan karakteristik algoritma regresi linear dan *k- nearest neighbors*.

3. Bagi perusahaan

Manfaat praktis bagi perusahaan diharapkan dengan adanya penelitian ini, pihak perusahaan dapat mempersiapkan hal - hal yang diperlukan di masa depan untuk kesiapan infrastruktur telekomunikasi yang lebih memadai.

### 1.6 Struktur Organisasi Skripsi

Berdasarkan kepada Keputusan Rektor UPI (Universitas Pendidikan Indonesia) mengenai kriteria pembuatan karya tulis ilmiah UPI untuk tahun akademik 2021, peraturan tersebut menjadi acuan penulis pada penyusunan skripsi ini. Struktur organisasi skripsi proposal tugas akhir ini memiliki beberapa bagian sebagai berikut :

1. BAB I : Pendahuluan

Beberapa hal yang dibahas dalam bab ini diantaranya adalah latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan struktur organisasi skripsi.

2. BAB II : Kajian Pustaka

Hal yang dibahas dalam bab ini adalah hal - hal yang berkaitan dengan ilmu atau teori yang mendukung atau berkaitan dengan penelitian ini.

3. BAB III : Metode Penelitian

Bab ini berisikan tentang karakteristik penelitian yang akan dilakukan itu sendiri seperti desain penelitian, populasi dan sampel, instrumen penelitian, prosedur penelitian yang meliputi alur penelitian, metode pengumpulan data, data penelitian, dan analisis data.

#### 4. BAB IV : Temuan dan Pembahasan

Hal yang dibahas dalam bab ini yaitu mengenai temuan - temuan penelitian yang didapatkan berdasarkan hasil pengolahan dan analisis data dan menghasilkan sebuah output yang beragam dalam bentuk dan pembahasannya untuk menjawab penelitian yang telah dirumuskan sebelumnya.

#### 5. BAB V : Kesimpulan, Implikasi, dan Rekomendasi

Bab ini berisikan tentang kesimpulan, implikasi, dan rekomendasi yang disajikan dalam suatu gagasan atau penafsiran peneliti terhadap hasil temuan yang didapatkan dari proses penelitian, dan juga pengajuan beberapa gagasan yang bertujuan untuk dapat dimanfaatkan dari hasil penelitian ini.