

BAB II

METODE PENELITIAN

2.1 Objek Penelitian

Penelitian ini memiliki fokus khusus pada data pencetakan Kartu Tanda Penduduk (KTP) Disdukcapil Kecamatan Samarinda Utara pada bulan Agustus 2023 yang dikumpulkan selama pelaksanaan Program Kerja Lapangan (PKL). Penelitian ini bertujuan untuk mengembangkan sebuah model prediksi waktu pencetakan KTP dengan menerapkan Metode Regresi Linier Berganda.

2.2 Alat dan Bahan

Dalam proses penelitian ini, terdapat alat dan bahan yang digunakan untuk mendukung pelaksanaan kegiatan penelitian. Adapun alat dan bahan tersebut adalah sebagai berikut:

2.2.1 Data Pencetakan KTP

Pengumpulan data dilakukan dengan mengumpulkan data historis atau riwayat proses pencetakan KTP oleh Disdukcapil kecamatan samarinda utara yang berlokasi di Jalan Gunung Kapur, Lempake, Kec. Samarinda Utara, Kota Samarinda. Data ini mencakup pengumpulan data, verifikasi informasi, hingga proses pencetakan kartu KTP.

2.2.2 Python

Python adalah bahasa pemrograman yang banyak digunakan oleh pengembang untuk pengembangan aplikasi web, pengembangan perangkat lunak, *Data Science*, dan *Machine Learning* (AWS 2024, diakses 10 Maret 2024). Karena gratis dan mudah dipelajari, bahasa ini dapat digunakan di banyak platform, dan cocok dengan semua jenis sistem dan meningkatkan kecepatan pengembangan.

2.2.3 Google Colab

Google Colaboratory adalah dokumen yang dapat dieksekusi dan digunakan untuk menyimpan, menulis, serta berbagi program yang telah dibuat melalui Google Drive (Andre Oliver 2022, diakses 10 Maret 2024). *Software* ini pada dasarnya mirip dengan Jupyter Notebook gratis berbentuk *cloud* yang dapat digunakan dengan browser seperti Firefox dan Chrome. Penggunaanya dapat menjalankan kode Python tanpa perlu melakukan proses instalasi dan konfigurasi lainnya.

2.2.4 Regresi Linier

Regresi linier merupakan suatu Metode yang berfungsi untuk memeriksa korelasi antara variabel independen dan variabel dependen (Mardiani et al., 2023). Dalam konteks regresi linear sederhana, kita mempertimbangkan hubungan antara satu variabel independen (X) dan satu variabel dependen (Y), tujuannya adalah untuk membangun model dapat memprediksi nilai Y berdasarkan nilai X (Ayuni & Fitriyah, 2019). Persamaan dasar Metode Regresi Linier adalah sebagai berikut:

$$Y = a + bX \quad (2.1)$$

Dimana:

Y = Variabel Terikat

a = Konstanta (*intercept*)

b = Koefisien Regresi

X = Variabel Bebas

Jika terdapat masalah prediksi yang melibatkan dua atau lebih variabel bebas, Regresi Linier Berganda digunakan untuk mengatasi hal tersebut (Susanti & Sussolaikah, 2022). Persamaan Metode Regresi Linier Berganda adalah sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (2.2)$$

Dimana:

Y = Variabel Terikat

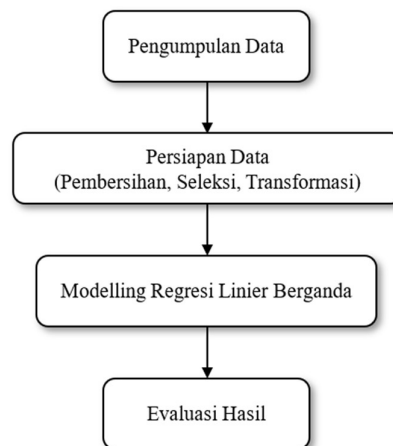
a = Konstanta (*intercept*)

b = Koefisien Regresi

X = Variabel Bebas

2.3 Tahapan Penelitian

Penelitian ini akan melibatkan 4 tahapan, dengan rincian tahapan pada gambar 2.1 sebagai berikut:



Gambar 2.1 Tahapan Penelitian

2.3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data pencetakan Kartu Tanda Penduduk (KTP) pada bulan Agustus 2023 sebanyak 350 data pencetakan yang dikumpulkan selama pelaksanaan

Program Kerja Lapangan (PKL) di Disdukcapil Kecamatan Lempake. Data tersebut memiliki beberapa atribut seperti KK, NIK, Nama, Umur, Jenis Kelamin, RT, Kelurahan, Keterangan, Tanggal Pengajuan, Waktu Pengajuan, serta Waktu Selesai sebagai label. Keterangan atribut dapat dilihat pada Tabel 2.1.

Tabel 2.1 Atribut Data Pencetakan KTP

NO	Atribut	Keterangan
1	KK	Nomor Kartu Keluarga
2	NIK	Nomor Induk Keluarga
3	Nama	Nama Pemohon
4	Umur	Umur Pemohon
5	Jenis Kelamin	Jenis Kelamin
6	RT	RT Pemohon
7	Kelurahan	Kelurahan Pemohon
8	Keterangan	Tujuan Pencetakan KTP
9	Tanggal Pengajuan	Tanggal Pengajuan
10	Waktu Pengajuan	Waktu Pengajuan
11	Waktu Selesai	Label/Waktu Selesai Pencetakan

Adapun contoh data yang digunakan dalam penelitian ini disajikan dalam bentuk Tabel 2.2 sebagai berikut.

Tabel 2.2 Data Pencetakan KTP

No	KK	NIK	Nama	Umur	Jenis Kelamin	RT	Kelurahan	Keterangan	Tanggal Pengajuan	Waktu Pengajuan	Waktu Selesai
1	6411012...	6472032...	THOMAS ...	42	L	12	SPJB	Pindahan	08/08/2023	08:05	08:10
2	6472052...	6472051...	SYAHRUL ...	20	L	6	SPJU	PRR	08/08/2023	08:32	08:48
3	6472052...	6472052...	ROHMAN...	17	L	15	TM	PRR	08/08/2023	09:19	09:31
4	6472051...	6472050...	MUHAMMAD...	17	L	29	SPJS	PRR	08/08/2023	09:47	10:03
5	6472052...	6472052...	LENDI ...	21	L	35	SPJT	Hilang	08/08/2023	10:10	10:16
...
346	6472050...	6472055...	PHEMA...	17	P	48	SPJT	PRR	08/30/23	09:03	09:27
347	6472053...	1210196...	MIKA...	17	P	29	TM	PRR	08/30/23	09:08	09:31
348	6472051...	6472052...	WINARKO	40	L	1	LPK	Hilang	08/30/23	09:13	09:21
349	6472052...	6472056...	RINA...	37	P	42	LPK	Hilang	08/30/23	09:19	09:26
350	3508182...	3508184...	NINGSIH	34	P	1	SPJU	Pindahan	08/30/23	09:26	09:40

2.3.2 Persiapan Data

Pada tahapan ini data yang telah didapat akan melewati beberapa persiapan seperti pembersihan data, seleksi data serta transformasi data untuk membersihkan data dari bentuk awalnya yang tidak terstruktur menjadi struktur yang lebih terorganisir untuk memudahkan proses prediksi. Berikut adalah langkah-langkahnya:

1) Pembersihan Data

Pembersihan data bertujuan untuk membersihkan data yang berpotensi mengurangi akurasi model seperti *Missing Value* (nilai yang hilang pada dataset) dengan cara menghapus baris/kolom data yang tidak memiliki nilai.

2) Seleksi Data

Pada tahapan ini akan dilakukan penghapusan atribut-atribut yang tidak relevan pada data seperti KK, NIK, Nama serta Tanggal Pengajuan sehingga data yang digunakan menjadi lebih efisien dan menghasilkan model yang optimal dalam melakukan prediksi.

3) Transformasi Data

Langkah ini dilakukan dengan tujuan mengubah jenis data kategorik menjadi bentuk numerik menggunakan teknik *Label Encoding*. Proses ini sangat penting karena pemodelan regresi linear hanya dapat menangani data numerik. Sebagai contoh, data mengalami transformasi dengan menggantikan setiap nilai dalam suatu kolom dengan angka berurutan seperti 0, 1, 2 dan seterusnya.

2.3.3 Modelling Regresi Linier Berganda

Karena proses prediksi melibatkan lebih dari 1 atribut maka Metode Regresi Linear Berganda digunakan untuk membangun model prediksi waktu pencetakan KTP dalam penelitian ini. Adapun tahapan dalam membuat model sebagai berikut:

1) *Import* Data

Import data merupakan langkah awal dalam proses pemodelan, berfungsi untuk menampilkan dataset yang telah dikumpulkan sebelumnya kedalam bentuk DataFrame.

2) Korelasi Data

Proses ini bertujuan untuk mencari hubungan antara atribut dengan label. Tujuannya adalah untuk mengidentifikasi atribut mana yang berpengaruh lebih signifikan dalam model prediksi ini. Korelasi ditentukan dengan memeriksa nilai positif terbesar dari atribut yang berkaitan dengan label data.

3) Split Data

Penelitian ini memanfaatkan teknik Split Data untuk mendapatkan hasil akurasi yang optimal dari proses dataset. Pada tahapan ini dilakukan pengujian terhadap beberapa rasio perbandingan untuk mendapatkan rasio perbandingan dataset yang terbaik.

4) Prediksi Regresi Linier Berganda

Proses ini akan menerapkan Metode Regresi Linier Berganda untuk memprediksi waktu pencetakan KTP berdasarkan atribut- atribut yang ada pada data.

5) Hasil

Hasil prediksi Regresi Linier Berganda nantinya akan berupa menit yang kemudian akan dikonversi ke dalam bentuk jam.

2.3.4 Evaluasi Hasil

Pada penelitian ini Metode *Mean Absolute Error* (MAE) digunakan untuk menghitung rata-rata perbedaan absolut antara nilai prediksi dan nilai (Rusdy et al., 2022). MAE merupakan salah satu Metode pengujian yang digunakan untuk mengevaluasi hasil akurasi model prediksi (Alita et al., 2021). Nilai MAE menunjukkan rata-rata kesalahan absolut antara nilai sebenarnya dan hasil prediksi (Li et al., 2021). Persamaan MAE dijelaskan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.3)$$

Dimana:

\hat{y}_i = Nilai Hasil Prediksi

y_i = Nilai Sebenarnya

n = Jumlah Data

Selain itu, Koefisien determinasi atau R² juga digunakan untuk mengetahui seberapa pengaruh variable bebas (X) terhadap variabel terikat (Y), Semakin mendekati nilai 1 maka variabel bebas (X) dapat memiliki hubungan linier yang kuat terhadap variabel terikat (Y) (Yuliara, 2016). Persamaan untuk menghitung nilai R² adalah sebagai berikut:

$$r^2 = \frac{(b_1 \Sigma X_1 Y) + (b_2 \Sigma X_2 Y) + \dots + (b_n \Sigma X_n Y)}{\Sigma Y^2} \quad (2.4)$$

Dimana:

Y = Variabel Terikat

b = Koefisien Regresi

X = Variabel Bebas