

BAB III

METODOLOGI PENELITIAN

3.1 Objek dan Subjek Penelitian

Objek penelitian adalah topik permasalahan untuk di investigasi dalam penelitian, pada penelitian ini objek penelitian adalah model regresi linear dan *ARIMA* yang dapat memprediksi tren *keyword SEO* beserta merekomendasikannya.

Subjek penelitian merupakan narasumber untuk dijadikan sumber data riset, pada penelitian ini subjek penelitian adalah data *keyword SEO* dari *google trend* yang berhubungan dengan kebutuhan pemasaran dalam industri *real estate* seperti *keyword* perumahan, apartmen, dan lain-lain.

3.2 Alat dan bahan penelitian

Dalam melakukan penelitian membangun model prediksi tren *keyword SEO* diperlukan alat dan bahan untuk menyelesaikan penelitian

3.2.1 Alat

Alat yang digunakan dalam pengembangan model berupa perangkat keras dan perangkat lunak.

3.2.1.1 Perangkat Keras

Perangkat keras yang digunakan pada pengembangan model prediksi tren *keyword SEO* yaitu personal computer (PC). Berikut spesifikasi PC yang digunakan untuk mendukung pengembangan model prediksi tren *keyword SEO* adalah:

1. Processor: AMD A8-7650K Radeon R7
2. VGA: RX 580 SP
3. Memori: 16GB RAM
4. SSD: 256GB

3.2.1.2 Perangkat Lunak

Perangkat lunak yang digunakan pada pengembangan model prediksi tren *keyword SEO* ini adalah:

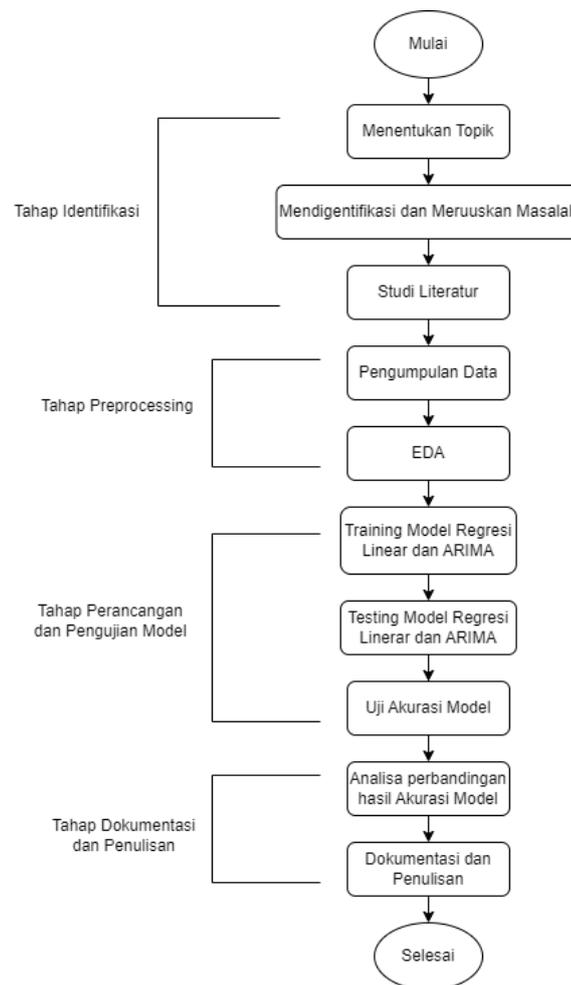
1. Windows 10 Pro 64 bit
2. Google Colab
3. Microsoft Excel
4. Google Spreadsheet

3.2.2 Bahan

Bahan yang digunakan dalam penelitian ini yaitu data *time series keyword SEO* yang diambil dari *google trend* dengan *keyword* atau kata kunci perumahan yang diambil dari bulan September 2018 hingga bulan Agustus 2023 yang menampilkan jumlah hasil pencarian kata kunci (*search keyword*) perumahan pada tiap minggunya dalam bentuk format csv.

3.3 Diagram alir penelitian

Pada tahap ini akan menjelaskan tahapan-tahapan dalam pembuatan model prediksi tren *keyword SEO*. Diagram alir metodologi yang digunakan pada penelitian ini yaitu studi literatur penelitian, analisis kebutuhan penelitian, pengumpulan data penelitian, perancangan model, implementasi model dan penarikan kesimpulan. Diagram alir penelitian untuk menggambarkan tahapan-tahapan pengerjaan terdapat pada Gambar 3.1.



Gambar 3. 1 Tahapan-Tahapan Penelitian

3.3.1 Tahap Identifikasi

1. Mengidentifikasi dan Merumuskan Masalah

Tahap mengidentifikasi dan merumuskan masalah yang terjadi dalam bisnis untuk memprediksi terutama dalam bidang *SEO* khususnya dalam kata kunci perumahan.

2. Studi Literatur

Proses studi literatur pada penelitian ini untuk mempelajari literatur dari beberapa model dan bidang yang berhubungan dalam pembuatan model prediksi tren *keyword SEO* menggunakan metode regresi linear dan *ARIMA*, meliputi:

a. *Search engine optimization (SEO)*

b. Metode regresi linear

c. Metode *ARIMA*

Literatur tersebut didapat dari buku, jurnal, e-book, dan penelitian sebelumnya.

3.3.2 Tahap *Preprocessing*

1. Pengumpulan Data

Pada tahap ini, data diambil dengan melakukan pengunduhan file .CSV dari google *trend* pada alamat *website* berikut <https://trends.google.co.id/trends/>, Pengunduhan dilakukan dengan melakukan sorting terhadap *keyword* mana yang berpotensi untuk industri *real estate* dan juga seberapa lama jagka waktu data yang dibutuhkan.

2. EDA (*Exploratory Data Analysis*)

Dalam tahap ini dilakukan eksplor data yang bertujuan untuk mengetahui karakteristik data yang akan dianalisis. Berikut tahap EDA yang dilakukan pada penelitian ini sebagai berikut:

a. *Loading Data*

Hal pertama yang dilakukan EDA adalah memuat dataset perumahan.csv dan melihat 5 data teratas dari dataset tersebut.

	Minggu	perumahan: (Indonesia)
0	2018-09-16	71
1	2018-09-23	75
2	2018-09-30	68
3	2018-10-07	72
4	2018-10-14	74

Gambar 3. 2 Hasil *Loading Data*

b. Informasi Dasar Data

Melihat info dasar atau deskripsi dari data perumahan.csv dengan menggunakan perintah `df.info()`. Data yang

diperlihatkan berupa tipe data, jumlah kolom, jumlah baris. Dapat dilihat pada Gambar 3.3.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Minggu                260 non-null    datetime64[ns]
1   perumahan: (Indonesia) 260 non-null    int64
dtypes: datetime64[ns](1), int64(1)
memory usage: 4.2 KB
```

Gambar 3. 3 Melihat Info dasar Dataset

c. Mencari Nilai Kosong (*Null*)

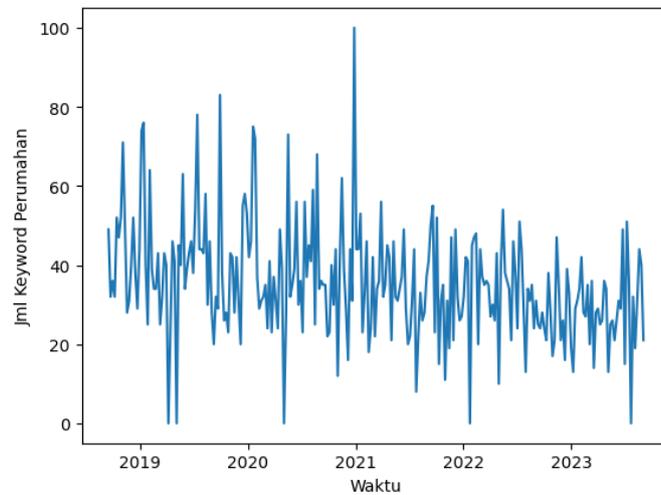
Dataset yang digunakan pada penelitian ini tidak ada nilai yang berisi NULL. Dapat dilihat pada Gambar 3.4.

```
df.isnull().sum()
Minggu                0
perumahan: (Indonesia) 0
dtype: int64
```

Gambar 3. 4 Melihat Dataset yang terdapat Nilai *Null*

d. Visualisasi Data

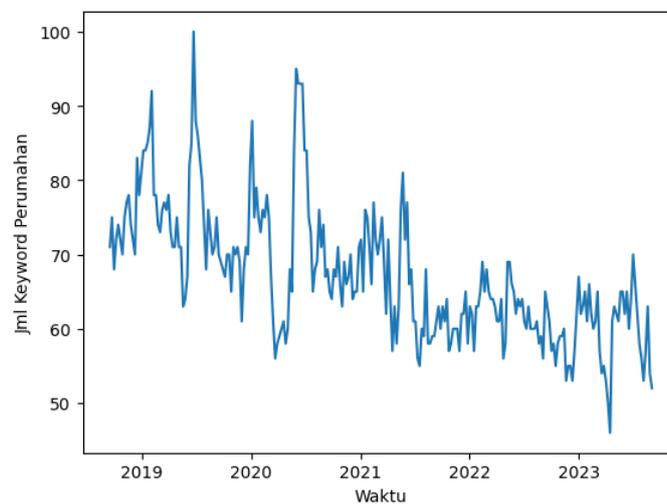
Dataset kemudian dilakukan plot untuk menampilkan pola dan tren pada dataset. Pada proses ini juga akan dilakukan analisa untuk melihat pola dan tren yang cocok untuk model regresi linear dan *ARIMA*



Gambar 3. 5 Hasil plot *rolling mean*.

e. Interpolasi Data

Dapat dilihat pada hasil visualisasi plot data bahwa ada banyak outlier atau data yang nilai nya terpaut jauh sehingga membuat data menjadi tidak seimbang, maka dari itu dilakukan interpolasi data agar dapat meminimalisir ketidak seimbangan dalam data dengan melakukan interpolasi.



Gambar 3. 6 Hasil interpolasi data

f. Merubah fomate data pada kolom minggu

Model regresi linear dan *ARIMA* membutuhkan format yang sesuai untuk dapat bekerja dengan baik. Pada penelitian ini

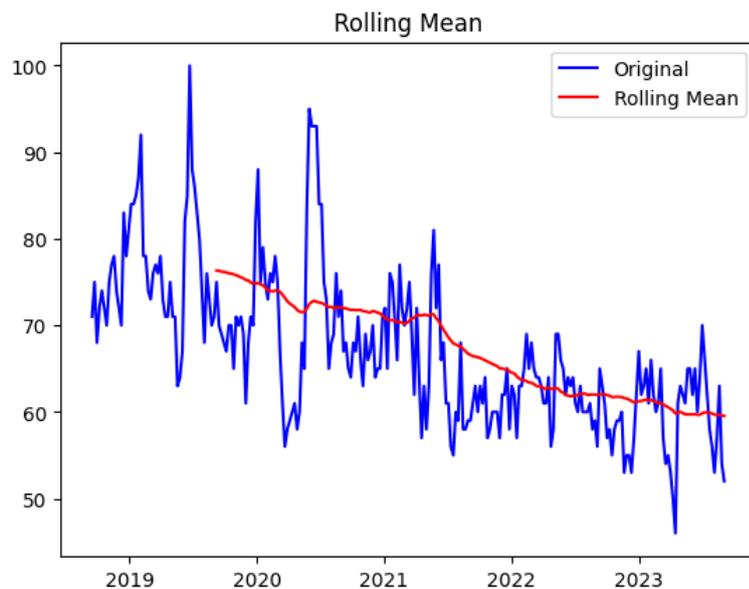
kolom pada dataset minggu perlu diubah (*parse*) agar model dapat bekerja dengan baik.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 260 entries, 0 to 259
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Minggu                260 non-null   datetime64[ns]
1   perumahan: (Indonesia) 260 non-null   float64
dtypes: datetime64[ns](1), float64(1)
memory usage: 4.2 KB
```

Gambar 3. 7 Hasil *parse* data menjadi *date time*

g. Melihat arah *tren* dataset

Pada proses ini dilakukan plot dengan *rolling mean* untuk melihat arah pergerakan rata-rata dari data yang dapat disimpulkan sebagai tren.



Gambar 3. 8 Hasil plot *rolling mean*.

3.3.3 Tahap Perancangan dan Pengujian Model

1. Training Model Regresi Linear dan *ARIMA*

Hasil dari tahap EDA selanjutnya dilakukan proses training menggunakan model Regresi Linear dan *ARIMA* agar dapat melakukan peramalan *time series* dengan baik. Namun sebelum itu akan dilakukan 2 test untuk membuat model *ARIMA* dapat bekerja dengan baik., Kemudian dilakukan pembagian dataset *train* dan *test* dibagi menjadi 80% *train* dan 20% *test*. Berikut adalah tahapan perancangan dan pengujian model

a. Analisa ADF (*Augmented Dickey Fuller*)

Pada proses ini dataset akan diuji terlebih dahulu apakah dataset bersifat stasioner. Pengujian dilakukan menggunakan ADF dengan cara melihat apakah *dataset* rata-ratanya dan varians tetap konstan sepanjang waktu. Hasil pengujian ADF ini akan menghasilkan nilai *p-value* yang akan menjadi acuan data stasioner atau tidak.

b. Analisa ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*)

Pada proses ini pengujian ACF dan PACF dilakukan dengan tujuan agar mendapatkan hyper parameter *p* dan *q* dengan melihat hubungan antara *lags* yaitu hubungan setiap baris data yang disebut *lags* terhadap nilai masa lalu yang mempengaruhi nilai masa depan pada dataset. Hasil dari uji analisa ACF dan PACF ini adalah nilai perkiraan untuk hyper parameter *p* dan *q*.

c. Analisa Model Regresi Linear

Proses ini akan melibatkan library *sklearn* untuk menggunakan model regresi linear yang telah disediakan oleh library tersebut, pada tahap ini juga akan dilakukan pembagian dataset untuk *train* dan *test* dengan komparasi sebesar 80% dari jumlah dataset dan 20% dari jumlah dataset.

d. Analisa Model *ARIMA*

Proses ini akan melibatkan library *statsmodel* yang menyediakan banyak alat untuk membangun model *ARIMA*.

Proses ini pada dasarnya hampir sama dengan pembuatan model regresi linear, namun yang membedakannya adalah proses analisa ADF, ACF, dan PACF yang menjadi fundamental bagi model *ARIMA*. Hal ini karena rangkaian test tersebut akan menentukan apakah dataset dapat digunakan oleh model *ARIMA* dan juga memberikan rekomendasi untuk hyper parameter model *ARIMA*.

2. Uji Akurasi Model

a. Uji Akurasi RMSE

Uji akurasi menggunakan *RMSE* diujikan kepada metode regresi linear dan *ARIMA*. *RMSE* digunakan untuk pengujian berapa banyak error yang terjadi pada sebuah model. Hasil uji ini menghasilkan nilai dalam bentuk *decimal*

b. Uji Akurasi MAPE

Uji akurasi menggunakan *MAPE* diujikan kepada metode regresi linear dan *ARIMA*. *MAPE* digunakan untuk pengujian berapa banyak error yang terjadi pada sebuah model. Hasil uji ini menghasilkan nilai dalam bentuk persentase.

3.3.4 Tahap Dokumentasi dan Penulisan

Pada tahap terakhir yaitu dokumentasi dan penulisan untuk tugas akhir dan publikasi jurnal ilmiah dengan mengikuti aturan yang berlaku.