

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian yang akan digunakan untuk skripsi ini adalah data mentah laporan pengaduan yang diterima oleh Pusat Pelayanan Terpadu Pemberdayaan Perempuan dan Anak (P2TP2A) baik laporan dari pihak kepolisian, masyarakat, atau korban. Pengambilan data yang didapat berupa inisial korban dan pelaku, umur, jenis laporan, lokasi kejadian. Pengambilan data lain seperti pengeluaran perkapita, populasi dan pengangguran didapat dari badan pusat statistik pelalawan. Semua data diubah kedalam *dataset* agar sesuai format *time series*.

P2TP2A Pelalawan merupakan suatu lembaga layanan pemerintah yang beroperasi dibawah dinas pemberdayaan perempuan dan perlindungan anak Pelalawan. Pada saat penelitian ini ditulis, Ibu Eman Rianda ditunjuk sebagai kepala unit teknis P2TP2A untuk unit daerah Pelalawan. P2TP2A Pelalawan bertujuan untuk meningkatkan kesetaraan gender dan meningkatkan kualitas hidup perempuan serta perlindungan kepada seluruh anak – anak di daerah Pelalawan.

Tugas P2TP2A Pelalawan dijadikan sebagai pusat data untuk seluruh situasi mengenai perempuan dan anak – anak diberbagai aspek kehidupan, baik mengenai kesejahteraan, hukum, ekonomi dan sosial. P2TP2A juga dijadikan sebagai penengah untuk perempuan dan anak-anak yang terjerat kasus hukum di daerah Pelalawan. Oleh karena itu P2TP2A Pelalawan sebagai pusat

pengambilan data untuk penelitian ini dikarenakan fungsi P2TP2A sebagaimana yang dijelaskan diatas yaitu tempat pengumpulan data kasus hukum yang terjadi mengenai perempuan dan anak-anak di Pelalawan.

3.2 Metode Penelitian

Tabel 3. 1 Tabel Perbandingan Metode *Forecasting*

Jenis Perbedaan	Kelebihan	Kekurangan
ARIMA (Pal & Prakash, 2017) dan (Aileen, 2019)	<ul style="list-style-type: none"> - Umum digunakan pada peramalan data <i>time series</i>. - Dapat diintegrasikan pada dataset yang tidak stasioner - Pembetulan model yang cepat - Parameter yang sedikit. 	<ul style="list-style-type: none"> - Hanya memiliki input berupa data yang <i>univariate</i>,
ARIMAX (Yucesan et al., 2018) dan (Azhari & Utomo, 2018)	<ul style="list-style-type: none"> - Dapat diintegrasikan pada dataset yang tidak stasioner - Pembetulan model yang cepat - Dapat menampung input data yang <i>multivariate</i>. - Dapat memberikan pengaruh dari variabel independen yang ada pada model. 	<ul style="list-style-type: none"> - Kurang baik dalam memprediksi data non linear
VAR (Aileen, 2019)	<ul style="list-style-type: none"> - -Dapat menampung input data yang <i>multivariate</i>. - Sering digunakan pada perhitungan <i>time series</i> yang berbentuk <i>multivariate</i>. 	<ul style="list-style-type: none"> - Asumsi model yang menganggap semua variabel <i>exogenous</i> menjadikan kesesuaian dari model atau <i>goodness of fit</i> sulit untuk dinilai.

Berdasarkan tabel 3.1, metode atau model yang digunakan untuk penelitian peramalan kriminalitas menyangkut perempuan dan anak-anak di Kabupaten Pelalawan adalah menggunakan ARIMAX dan VAR. Model atau metode VAR dan ARIMAX digunakan karena kemampuannya yang dapat memiliki input *dataset* berupa *multivariate* atau *dataset* prediksi dengan variabel lebih dari satu. Pemilihan ini juga dilakukan mengingat *dataset* yang akan digunakan pada penelitian ini berupa *dataset* dengan variabel *multivariate*. Adapun penjelasan dari tiap variabel akan dijelaskan pada subbab berikutnya.

3.3 Objek Penelitian

Metode penelitian *time series analysis* ini akan dilakukan dengan menggunakan algoritma *Auto Regressive Integrated Moving Average with Exogenous variable* atau yang bisa disebut dengan ARIMAX. Pemilihan metode ARIMAX digunakan karena adanya penggunaan variabel independen lain seperti jumlah populasi, pengangguran, dan jumlah pengeluaran perkapita. Berdasarkan penjelasan pada bab kedua mengenai penelitian selanjutnya, ARIMAX mampu memprediksi frekuensi kriminalitas pencurian motor di kota Yogyakarta dengan jumlah persebaran motor sebagai variabel independen. ARIMAX ini nantinya akan dibandingkan dengan algoritma *Vector Auto Regressive (VAR)* sebagaimana yang dilakukan percobaan selanjutnya pada bab dua VAR digunakan untuk memprediksi frekuensi kebutuhan pangan dengan beberapa variabel independen sebagai penentu dari nilai kebutuhan pangan di Indonesia.

3.4 Variabel Penelitian

Dalam penelitian ini, digunakan dua tipe variabel. Perbedaan antara dua variabel ini akan dijelaskan di subbab selanjutnya.

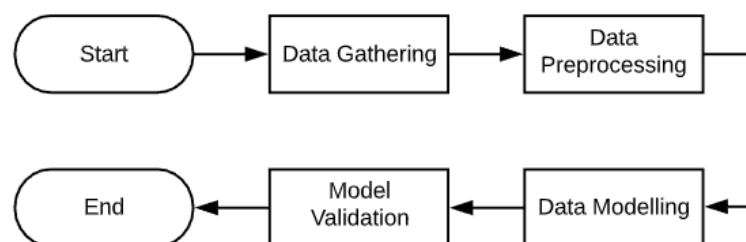
3.4.1. Variabel Dependen

Variabel pertama yang akan dijelaskan ialah variabel dependen. dependen adalah variabel yang dipengaruhi oleh variabel independen. Variabel independen pada penelitian ini adalah jumlah frekuensi kasus kriminalitas yang terjadi tiap bulannya di kabupaten Pelalawan.

3.4.2. Variabel Independen

Variabel kedua yang akan dijelaskan ialah variabel independen. Independen adalah variabel yang mempengaruhi variabel dependen. Variabel dependen pada penelitian ini adalah input berupa jumlah populasi warga(jiwa), pengangguran(jiwa), dan pengeluaran per kapita(PPK).

3.5 Alur Penelitian



Gambar 3. 1 Alur Penelitian

Penelitian ini akan mengikuti tahap seperti pada Gambar 3.1 dimana data yang sudah dikumpulkan akan memasuki tahap *preprocessing* untuk dibentuk

format *time series* lalu dibentuk model ARIMAX dan VAR dan hasilnya akan divalidasi. Hasil validasi akan dibandingkan untuk melihat algoritma terbaik untuk prediksi kriminalitas di kabupaten Pelalawan.

3.5.1. Data Gathering

Data laporan kriminalitas diambil dari dinas P2TP2A dari rentang tahun 2015 – 2019. Pengambilan data ada yang sudah berbentuk digital dan juga berbentuk pencatatan manual. Adapun data pengeluaran dan pertumbuhan penduduk serta jumlah pengangguran bisa didapatkan di badan pusat statistik Pelalawan di rentang tahun yang sama. Data akan disatukan kedalam satu *dataset*.

Data diambil dengan mendapat persetujuan dari pihak ketua teknis P2TP2A Pelalawan. Pengambilan data langsung diambil dari komputer pusat yang berada di kantor P2T2A Pelalawan. Tidak hanya berupa data digital, pengambilan data yang masih tercatat secara manual juga diambil untuk diubah ke rekam digital. Beberapa kolom data seperti nama akan disamarkan dalam bentuk inisial sebagai tindakan dari perlindungan privasi data. Pengambilan data *exogenous* diambil dari badan statistik pelalawan, baik yang tercatat melalui *website* maupun tertulis sesuai buku terbitan badan statistik pelalawan.

3.5.2. Data Preprocessing

Data yang didapatkan akan memasuki tahap *preprocessing*, mengingat data secara general tidak bisa langsung diberikan kedalam teknik pengolahan mesin (Roy et al., 2018). Tahap paling awal yaitu membersihkan data dari

missing values. Pembersihan ini dilakukan mengingat *missing values* merupakan masalah yang cukup sering dijumpai pada tahap *preprocessing* (Roy et al., 2018). Proses ini akan diberlakukan dengan dua metode diantaranya *discarding* dan *mean imputer* (Roy et al., 2018). *Discarding* digunakan bila *missing values* memiliki persentase yang kecil dari jumlah seluruh *dataset* dan *mean imputer* diterapkan bila persentase *missing values* berjumlah besar. Cara ini dilakukan dengan memasukkan nilai yang kosong dengan jumlah *mean* dari total nilai variabel atau kolom. Hal ini berguna untuk melengkapi nilai kosong pada *dataset* (Little & Donald, 2014).

Data yang sudah dikumpulkan dibuat keformat *dataset* terpisah. *Dataset* dibuat dengan format *time series* untuk masing-masing fitur. Pengelompokan data dibuat untuk setiap bulan bagi setiap *dataset*. *Dataset* frekuensi laporan dan variabel input akan di cek jika *value*, *mean*, *variance* berada pada nilai konstan (stasioner). Pengujian pertama dilakukan dengan visualisasi grafik untuk melihat objek observasi terhadap waktu. Pengecekan selanjutnya dilakukan dengan fungsi autokorelasi. Jika belum stasioner maka diperlukan proses differensiasi beberapa kali untuk membuat data menjadi stasioner (Anggraeni et al., 2017).

3.5.3. Data Modeling

Pembuatan model dilakukan dengan melakukan pembagian *dataset* menjadi dua bagian yaitu data *training* dan data *test* dengan pembagian sebagai berikut :

- Sebesar 80% *dataset* sebagai data *training*.

- Sebesar 20% *dataset* sebagai data *test*.

Pemodelan dilakukan dengan menggunakan algoritma ARIMAX karena dapat dari pada ARIMA. Hal ini seperti yang dijelaskan di landasan teori bahwa ARIMAX lebih cocok untuk *men-handle multivariate variable independent* lebih baik daripada ARIMA. Hasil pemodelan akan digunakan pada data *test* untuk melakukan forecasting yang akan menghasilkan prediksi frekuensi kejadian kriminalitas. Hasil *forecasting* juga dilakukan dengan *vector autoregression* untuk dibandingkan dengan model ARIMAX pada tahap evaluasi. *Dataset Time series* nantinya akan divisualisasikan dengan menggunakan *software* Tableau.

3.5.4. Model Validation

Penentuan algoritma terbaik untuk memprediksi frekuensi *error* akan ditentukan dengan menghitung *Root Mean Squared Error*(RMSE) dari setiap hasil prediksi algoritma pada *dataset*. Dengan menentukan RMSE terkecil kita dapat menentukan performa terbaik diantara kedua algoritma(Catlett et al., 2019). Seperti yang dijelaskan di bab ke-dua, *Mean Absolute Error*(MAE) juga akan digunakan untuk menentukan model terbaik antara ARIMAX dan VAR.

3.6 Tools Yang Digunakan

Pada penelitian ini, digunakan beberapa *tools* yang ditujukan untuk membantu proyek penelitian. *Tool* ini digunakan untuk keperluan berbeda untuk setiap alur penelitian. Terdapat tiga *tool* utama yang akan digunakan

untuk penelitian ini yaitu Python dan Tableau. Python ditujukan untuk membuat model dan Tableau ditujukan untuk membuat dashboard

3.6.1. Python

Pembentukan model dan prediksi dilakukan dengan platform Python. Salah satu kelebihan Python adalah sifatnya yang open source menjadikannya memiliki banyak *library* package yang membantu penggunaannya untuk melakukan pekerjaan machine learning. Python memiliki *library* yang dapat melakukan *modeling* untuk ARIMAX dan VAR yang bernama statsmodel. Adapun software *modeling* seperti Rapid Miner tidak dipilih dikarenakan tidak memiliki fungsi *modeling* ARIMAX untuk versi saat ini.

Pada penggunaan Python terdapat beberapa *library package* yang akan digunakan dan diinstalasikan pada environment Python. Beberapa diantaranya adalah :

1. Pandas.dataframe

Library ini akan membantu menyimpan variabel data table pada Python. Data table ini disebut sebagai *dataframe* yang dimana data *time series* nanti akan di assign kedalam variabel tersebut.

2. Numpy.array

Library ini dapat membantu menyimpan data *array* yang nantinya dapat setiap kolom tertentu dapat diambil dan di-assign kedalam variabel *numpy* untuk diolah.

3. Matplotlib

library ini nantinya akan mencakup beberapa *command* yang dapat digunakan untuk membuat *plotting* pada *environment* Python.

4. Sm.tsa.arima

library ini dapat menyediakan beberapa *command* mengenai model ARIMA. Salah satu parameter yang dapat diberikan adalah parameter *exogenous* yang dimana penggunaan ARIMA pada model tersebut menggunakan variabel independen atau dapat disebut sebagai ARIMAX.

5. Sm.tsa.api

Library ini berfungsi untuk menyediakan model *Vector Autoregressive*(VAR). Parameter yang dapat diisi pada model ini adalah orde dari *auto regressive* yang didapat dari tes PACF.

3.6.2. Tableau

Pembuatan visualisasi untuk kebutuhan exploratory data analysis pada *dataset* akan menggunakan software Tableau. Visualisasi yang dilakukan adalah berupa density heatmapping. Seperti yang dijelaskan di bab dua, Tableau memiliki berbagai fitur yang diantaranya mampu memberikan visualisasi pada *dataset spatiotemporal* dengan cara yang lebih mudah.

Tableau juga dapat membuat berbagai model grafik yang dapat membantu pihak P2TP2A Pelalawan untuk melihat trend dari jumlah laporan yang terjadi

di Pelalawan. Tableau memiliki beberapa macam visual seperti *bar plot*, *area chart*, *packed bubbles*, *line graph*, dan *map visualization*.



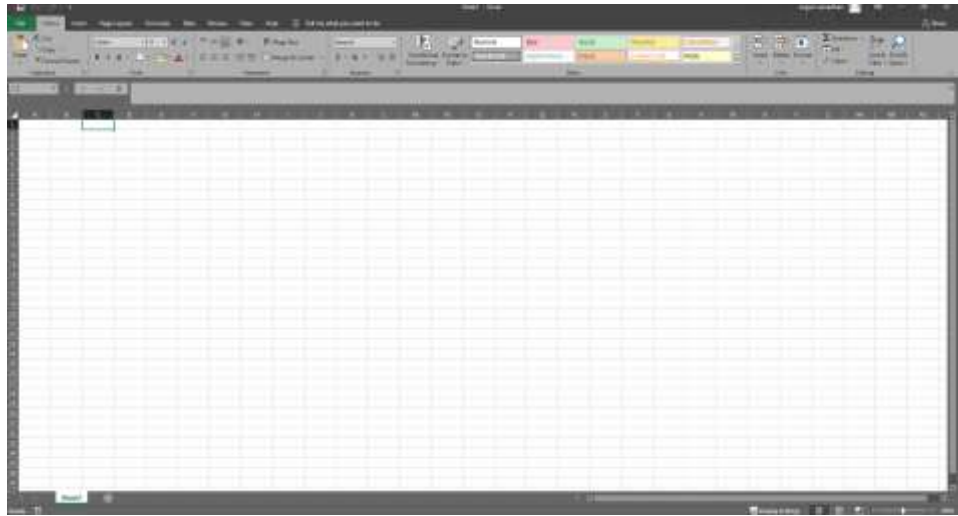
Gambar 3. 2 Pilihan Visualisasi Tableau

Tableau juga dapat menerima beberapa format *dataset* yang dapat digunakan salah satunya adalah Excel. Maka pada penelitian ini akan digunakan aplikasi Excel untuk membuat *dataset* untuk pembuatan *dashboard* Tableau. Excel akan dijelaskan disubbab selanjutnya.

3.6.3. Microsoft Excel

Microsoft Excel merupakan salah satu produk yang dikeluarkan oleh perusahaan Microsoft. Pada penelitian ini, Excel berfungsi sebagai penyusun format *dataset*. *Dataset* ini nantinya akan diolah untuk dua tujuan dari penelitian ini yakni pembuatan model dan *dashboard*. Excel sendiri dapat membantu untuk menyesuaikan pembuatan kolom dan nama tiap kolom yang akan digunakan.

Excel dapat membantu untuk menentukan jenis tipe untuk tiap kolom dengan bantuan menu yang ada aplikasinya. Setelah membuat format yang dibutuhkan untuk membuat *dataset* pada kedua proses, *dataset* dapat disimpan dalam bentuk format *xlsx* untuk Tableau dan dalam bentuk *comma seperated values*. Berikut pada gambar 3.3 merupakan tampilan dari Excel.



Gambar 3. 3 Tampilan Aplikasi Microsoft Excel