

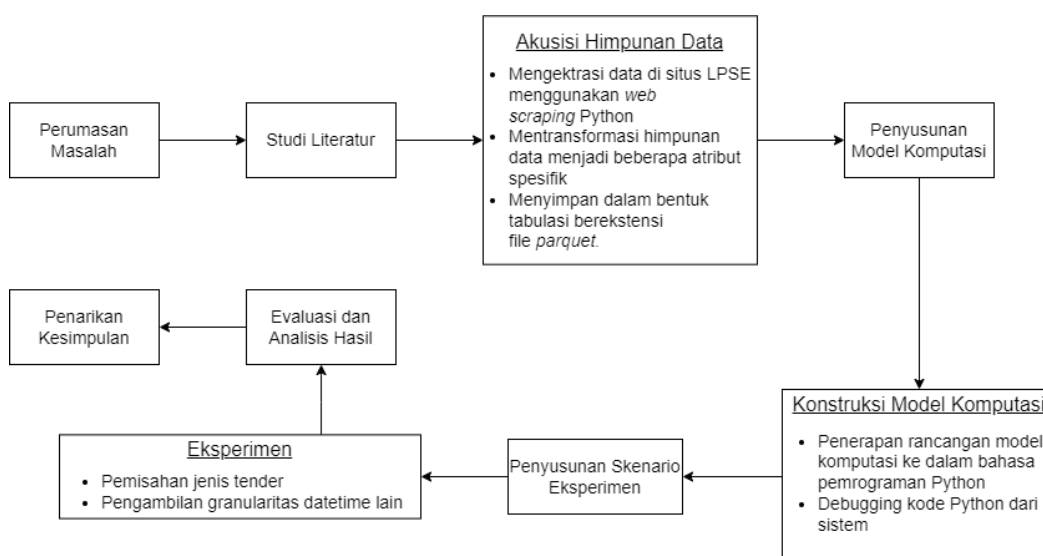
## BAB III

### METODE PENELITIAN

#### 3.1 Desain Penelitian

Penelitian yang dilaksanakan berupa penelitian dengan kategori eksperimental. Metode eksperimental adalah penelitian yang melibatkan kalibrasi variabel independen, mengontrol variabel eksternal / luar, dan juga mengevaluasi dampak variabel independen terhadap variabel dependen (Hastjarjo, 2019). Dalam memahami penelitian yang diajukan, maka peneliti menggunakan sebuah skema desain penelitian.

Desain penelitian adalah alur pengerjaan dari awal mula penelitian hingga ke bagian penarikan kesimpulan atau akhir penelitian. Desain penelitian yang dilakukan dapat dilihat pada Gambar 3.1 di bawah ini.



Gambar 3.1 Desain Penelitian

Berlandaskan desain penelitian yang diajukan, berikut beberapa penjelasan masing – masing tahapnya yang diawali dari perumusan masalah hingga ke bagian penarikan kesimpulan.

##### 1. Perumusan Masalah

Tahapan awal dari penelitian ini adalah dengan merumuskan berbagai masalah yang ada. Peneliti mengurasi beragam latar belakang masalah yang berasal dari tema penelitian yang diputuskan. Dari sini, peneliti juga menunjukkan adanya

kesenjangan antara kondisi objektif dengan penelitian terkini. Selain itu, tahapan ini memperlihatkan beberapa asumsi yang dimiliki oleh peneliti dan nantinya akan diuji menggunakan metode statistika yang dapat digunakan sebagai basis pengambilan keputusan.

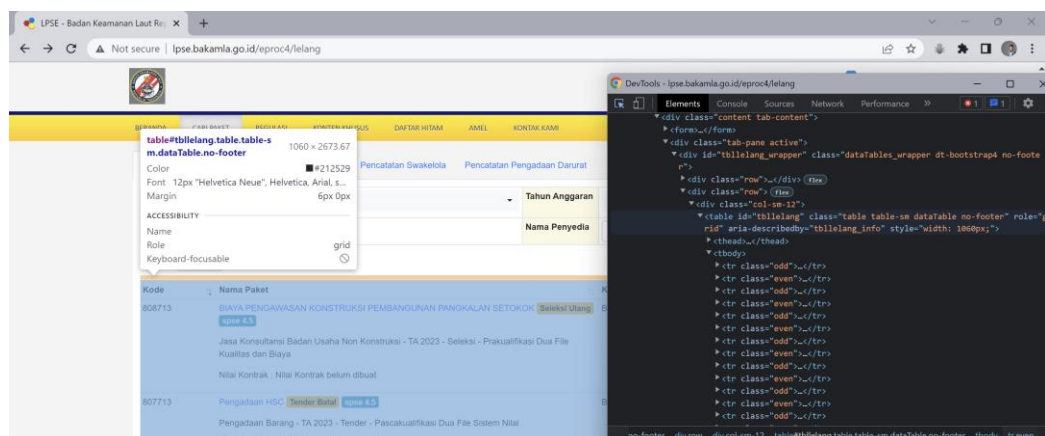
## 2. Studi Literatur

Studi literatur ialah tahapan di mana peneliti akan mereviu beragam penelitian yang telah dilakukan oleh peneliti terdahulu. Penelitian – penelitian yang direviu akan membantu peneliti dalam melaksanakan metode yang tepat untuk dilakukan. Literatur penelitian sebelumnya akan dikaji dan juga dianalisis terkait masalah, metode, data, dan solusi yang digunakan untuk disesuaikan pada penelitian ini.

Literatur yang akan dikurasi dan dipelajari ialah berkenaan dengan topik HPS, tender, regulasi pengadaan barang/jasa, prediksi, teknik *web scraping*, metode – metode prediksi seperti ARIMA dan SARIMA, metrik – metrik pengevaluasian untuk metode prediksi serta topik literatur lainnya.

## 3. Akuisisi Himpunan Data

Himpunan data yang akan digunakan merupakan data tender di situs LPSE Indonesia. Data tersebut terdapat dalam format tabel HTML, bila dilihat salah satu strukturnya pada Gambar 3.2 bawah ini. Data tersebut akan diakuisisi menggunakan sistem yang terautomasi berbasis bahasa Python. Sistem akuisisi ialah kumpulan komponen yang saling bekerja sama untuk melakukan pengumpulan, penyimpanan, pengolahan data, serta distribusi data untuk menghasilkan informasi yang bermakna dan berguna untuk proses pengambilan keputusan (Agus Dani Yudianto, 2021).



Gambar 3.2 Struktur HTML Data Tender di Situs LPSE

Daud Fernando, 2023

**PREDIKSI TENDER PADA SITUS PELELANGAN LPSE INDONESIA MENGGUNAKAN ALGORITMA SEASONAL AUTOREGRESSIVE MOVING AVERAGE (SARIMA)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Ada tiga komponen yang akan dilakukan pada tahapan akuisisi himpunan data ini, di antaranya:

a) Pengekstrakan Data di Situs LPSE

Data di situs LPSE terdiri dari beberapa halaman dan hanya tiga halaman saja yang akan dilakukan teknik *web scraping*. Halaman pertama adalah data tender keseluruhan yang disediakan oleh suatu LPSE (biasanya sesuai dengan KLPD atau bahkan letak geografisnya), seperti pada Gambar 3.2 di halaman sebelumnya. Halaman ini akan digunakan sebagai *landing page* dalam melakukan perulangan *web scraping* per tender karena formatnya dalam sebuah tabel dengan baris yang berulang lengkap juga dengan nama KLPD-nya. Halaman kedua adalah pengumuman lelang, pada halaman ini akan ada banyak sekali detail atribut dari suatu tender yang dilelang mulai dari kode tender bahkan hingga ke bagian syarat kualifikasi tender yang ada, seperti yang diperlihatkan pada Gambar 3.3 di bawah ini.

| Pengumuman             |  |  |             |
|------------------------|--|--|-------------|
| Peserta                |  |  |             |
| Hasil Evaluasi         |  |  |             |
| Pemenang               |  |  |             |
| Pemenang Berkontrak    |  |  |             |
| Kode Tender            | 792713   |  |             |
| Nama Tender            | PEKERJAAN PENGADAAN BAHAN BAKAR MINYAK HSD/BIO SOLAR INDUSTRI (NON SUBSIDI) UNTUK ASET PATROLI TAHUN ANGGARAN 2022 |  |             |
| Rencana Umum Pengadaan | Kode RUP   | Nama Paket   | Sumber Dana |
|                        | 37294537   | PEKERJAAN PENGADAAN BAHAN BAKAR MINYAK HSD/BIO SOLAR INDUSTRI (NON SUBSIDI) UNTUK ASET PATROLI TAHUN ANGGARAN 2022 | APBN        |
|                        | 37294537   | PEKERJAAN PENGADAAN BAHAN BAKAR MINYAK HSD/BIO SOLAR INDUSTRI (NON SUBSIDI) UNTUK ASET PATROLI TAHUN ANGGARAN 2022 | APBN        |
|                        | Konsolidasi ? ✖  |  |             |
| Tanggal Pembuatan      | 4 November 2022  |  |             |
| Tahap Tender Saat Ini  | Tender Sudah Selesai   |  |             |

Gambar 3.3 Halaman Pengumuman LPSE

Halaman ketiga adalah halaman hasil pelelangan alias hasil evaluasi pada Gambar 3.4 di bawah ini. Halaman ini akan menunjukkan beragam metrik performansi suatu peserta yang mengikuti tender berdasarkan data – data dari halaman peserta. Dari halaman ini data harga kontrak akan diekstraksi dengan syarat suatu peserta diberikan tanda bintang dan juga tendernya bukanlah tender ulang. Tender ini mengartikan telah selesai dan juga sudah memiliki satu pemenang penyedia agar dijadikan basis dari prediksi penelitian ini.

| Pengumuman          |   |   |   |   |   |                      |                      |                      |    |        |
|---------------------|---|---|---|---|---|----------------------|----------------------|----------------------|----|--------|
| Peserta             |   |   |   |   |   |                      |                      |                      |    |        |
| Hasil Evaluasi      |   |   |   |   |   |                      |                      |                      |    |        |
| Pemenang            |   |   |   |   |   |                      |                      |                      |    |        |
| Pemenang Berkontrak |   |   |   |   |   |                      |                      |                      |    |        |
| No                  | Nama Peserta  | K | A | T | P | PT                   | HN                   | HI                   | PK | Alasan |
| 1                   | PT. Arghaniaga Puncatunggal - 01.640.204.2-013.000          | ✓ | ✓ | ✓ | ✓ | Rp. 5.067.921.672.00 | Rp. 5.067.921.672.00 | Rp. 5.067.921.672.00 | ✓  | ★      |
| 2                   | PT. NIPSEA PAINT AND CHEMICALS - 01.001.769.7-092.000       | - | - | - | - | -                    | -                    | -                    | -  | -      |
| 3                   | PT. KASUMA AGUNG WICAKSANA - 01.395.176.9-005.000           | - | - | - | - | -                    | -                    | -                    | -  | -      |
| 4                   | PT. Transformasi Sejahtera Indonesia - 31.455.376.9-451.000 | - | - | - | - | -                    | -                    | -                    | -  | -      |
| 5                   | PT. Runinco Bumi Lestari - 90.827.460.8-721.000             | - | - | - | - | -                    | -                    | -                    | -  | -      |
| 6                   | cv. yofensa - 03.154.489.3-941.000                          | - | - | - | - | -                    | -                    | -                    | -  | -      |
| 7                   | PT. TEMPORASI INDONESIA - 31.739.178.7-071.000              | - | - | - | - | -                    | -                    | -                    | -  | -      |

A Evaluasi Administrasi   T Evaluasi Teknis   ST Skor Teknis   P Penawaran   PT Penawaran Terkoreksi   HN Hasil Negosiasi   SH Skor Harga   SA Skor Akhir   K Pembuktian Kualifikasi  
 K Evaluasi Kualifikasi   KA Skor Kualifikasi   SA Skor Pembuktian   HI Evaluasi Harga/Biaya   P Pemenang   PK Pemenang Berkontrak

Gambar 3.4 Halaman Hasil Evaluasi LPSE

Daud Fernando, 2023

**PREDIKSI TENDER PADA SITUS PELELANGAN LPSE INDONESIA MENGGUNAKAN ALGORITMA SEASONAL AUTOREGRESSIVE MOVING AVERAGE (SARIMA)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

#### b) Mentransformasi Himpunan Data menjadi Atribut Spesifik

Komponen selanjutnya adalah proses transformasi data. Karena proses akuisisi data yang dilakukan bersumber kepada situs LPSE (kemungkinan besar pemasukkan datanya dilakukan oleh manusia) memiliki tingkat *human error* yang cukup tinggi. Sehingga butuh perlakuan transformasi data juga. Selain itu, untuk mencegah data yang bersifat redundansi, akan dilakukan proses relasi antar tabel lengkap dengan nilai kardinalitas yang ada di tiap relasinya. Transformasi yang dilakukan dapat menghapus nilai *prefix* 0 dari nilai HPS dan harga kontrak, lalu melakukan perubahan *capital case* ke beberapa tipe data teks seperti jenis tender, dan bahkan merubah angka 0 dalam variabel numerik menjadi nilai 0.

#### c) Penyimpanan Himpunan Data Tertransformasi

Komponen terakhir ialah penyimpanan data yang sudah tertransformasi ke dalam file berekstensi *parquet*. Diadopsi dari Apache, file *parquet* adalah format file yang memfasilitasi penyimpanan efisien dan juga pembacaan file yang cepat. Hal tersebut dikarenakan file melakukan penyimpanan berbasis bingkahan – bingkahan kolom (Berk, 2022). Sehingga dapat dilakukan proses pengolahan data yang memenuhi kebutuhan dengan mudah dan cepat. Data yang akan dikumpulkan ialah berasal dari beberapa LPSE di seluruh Indonesia dan akan dilakukan *set operator* berupa *union* agar data yang didapatkan tidak redundansi.

#### 4. Penyusunan Model Komputasi untuk Metode Prediksi

Tahapan keempat dari desain penelitian ini ialah penyusunan model komputasi untuk memprediksikan nilai HPS dan juga margin tender hasil selisih harga kontrak tender dengan nilai HPS tender. Perencanaan model yang dibangun terdiri dari dua proses besar, yaitu proses pembuatan model komputasi menggunakan data latih dan proses kedua ialah pemvalidasian model komputasi menggunakan data uji. Selain itu model komputasi ini juga akan menerapkan metode prediksi yaitu ARIMA dan juga SARIMA.

#### 5. Konstruksi Sistem

Tahap selanjutnya ialah tahapan pengaplikasian penyusunan model komputasi ke dalam kode program berbahasa Python. Konstruksi sistem akan berdasar kepada susunan model komputasi yang telah dibangun pada tahapan sebelumnya secara komprehensif. Selain itu, di tahap ini akan ada proses *debugging* sehingga

memastikan sistem yang dibangun sudah berjalan dengan baik dan sesuai dengan metode yang ada. Penggunaan *tools* di tahap ini adalah Visual Studio Code untuk menjalankan *environment* dalam bentuk *notebook* agar hasil luaran analisis dapat dilihat per barisan kode.

#### 6. Penyusunan Skenario Eksperimen

Tahapan berikutnya ialah perancangan skema eksperimen yang akan dilaksanakan dimulai dari penggunaan himpunan data, penentuan orde tiap metode, pengujian hipotesis, dan juga melakukan analisis korelasi antara variabel dependen dengan variabel independen. Skenario ini akan digunakan untuk diimplementasikan pada tahapan selanjutnya agar memenuhi tujuan yang telah ditetapkan. Secara lengkap proses ini akan dibahas pada bagian selanjutnya yaitu bagian Analisis Data. Di mana tahapan ini akan berbasis kepada penelitian terdahulu yang dimodifikasi.

#### 7. Eksperimen

Dalam tahapan eksperimen, peneliti akan menerapkan segala skenario yang sudah disusun dalam memprediksikan suatu HPS dan margin tender pada model komputasi yang sudah dibuat sebelumnya pada tahapan konstruksi sistem. Di mana model komputasi akan menggunakan himpunan data latih dan menghasilkan nilai prediksi yang diinginkan dari setiap metode pada orde yang sudah ditetapkan agar bisa dibandingkan dengan data ujinya dengan upaya melihat seberapa baik performansi model dalam melakukan prediksi.

#### 8. Evaluasi dan Analisis Hasil

Ketika luaran nilai prediksi dari setiap skenario eksperimen dihasilkan, maka saatnya menggunakan metrik – metrik evaluasi yang sudah ditentukan. Metrik ini akan mengukur seberapa besar nilai galat yang dihasilkan model untuk memprediksikan nilai HPS dan margin tender agar dapat terlihat tingkat performansi dari model yang dibuat. Metrik ini juga akan memberikan gambaran akan skenario mana yang lebih baik digunakan dalam membangun sistem prediksi secara logis dan masuk akal, khususnya pada periode waktu apa suatu jenis tender optimal dalam melakukan prediksi. Selain itu juga akan dilakukan uji statistik untuk mendapatkan beberapa nilai yang dapat menjadi basis pengujian hipotesis

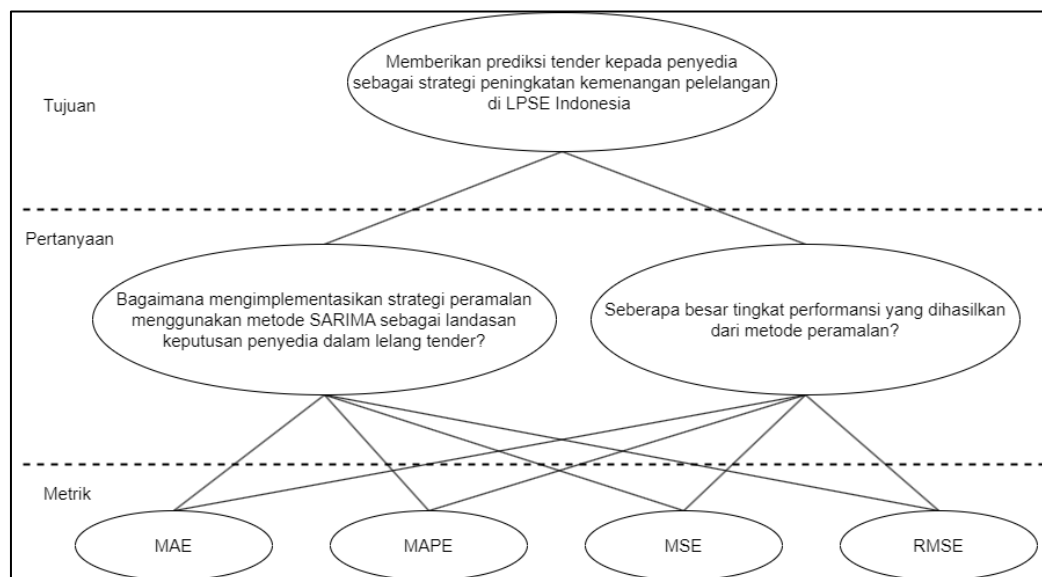
dari skenario dan eksperimen yang sudah dilakukan agar memenuhi asumsi – asumsi nilai galat hasil dari model yang sudah diciptakan.

### 9. Penarikan Kesimpulan

Tahapan terakhir dari penelitian ini ialah menetapkan suatu kesimpulan. Kesimpulan dalam perbandingan periode – periode waktu terbaik mana yang digunakan berbasis kepada hasil analisis dan juga nilai dari metrik evaluasi. Kesimpulan ini juga akan memunculkan seberapa baik performansi yang dihasilkan oleh model yang nantinya akan berpengaruh juga kepada nilai prediksi yang dihasilkan. Selain itu juga, akan diberikan interpretasi mengenai model – model terbaik yang berhasil dibuat.

### 3.2 Instrumen Penelitian

Instrumen penelitian ini dapat dilihat pada Gambar 3.5, yang merepresentasikan *Goal-Question Metric (GQM)*. Dimulai dari tujuan penelitian ini yang didefinisikan untuk diselesaikan menggunakan penelitian ini. Lalu ada pertanyaan yang dikategorisasikan menjadi dua sebagai representasi rumusan masalah penelitian. Dan terakhir adalah metrik alias metode – metode yang digunakan untuk mendapatkan nilai – nilai konkret yang bisa menjawab pertanyaan – pertanyaan sebelumnya.



Gambar 3.5 Instrumen Penelitian berupa GQM

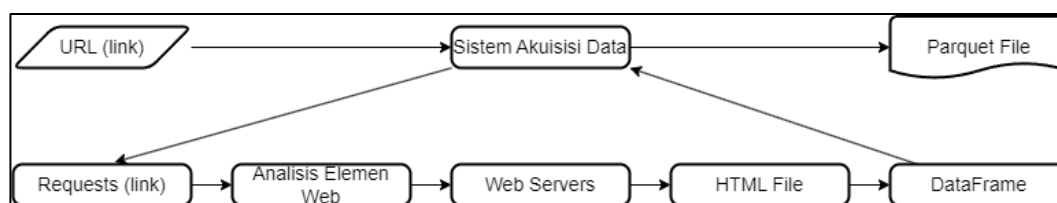
### 3.3 Prosedur Penelitian

Melansir pada desain penelitian yang tersedia di Gambar 3.1 terdapat beberapa tahapan untuk mengembangkan skenario penelitian seperti pembuatan sistem akuisisi himpunan data, konstruksi model komputasi, hingga pengaplikasian untuk eskalasi

performansi model prediksi tender berbasis eksperimen. Berikut beberapa penjelasan sekaligus skenario dan skema dari penjelasan lebih detail dari penelitian yang akan dilakukan.

### 3.3.1 Skenario Sistem Akuisisi Himpunan Data

Sistem akuisisi himpunan data merupakan serangkaian proses untuk mendapatkan data tender dari situs LPSE Indonesia yang mengimplementasikan proses *web scraping*. Merujuk pada literatur yang tervisualisasi di Gambar 2.19, dapat diaplikasikan sebuah skenario untuk penelitian ini pada Gambar 3.6 di bawah ini.



Gambar 3.6 Proses Keseluruhan Akuisisi Data

Dengan menggunakan bahasa pemrograman Python di Visual Studio Code, langkah pertama yang dilakukan adalah mendefinisikan objek *web driver* dari *selenium library*. Dari objek tersebut dapat dimulai eksekusi tahapan akuisisi data dengan mendapatkan *Uniform Resource Locator* (URL) terlebih dahulu. Kemudian URL tersebut akan dijadikan nilai masukkan dalam proses *request get* agar bisa membuka situs sesuai URL di objek *web driver*. Langkah selanjutnya dengan menganalisis elemen situs web untuk data atau spesifiknya atribut yang akan didapatkan.

Setelahnya proses *request get* alias permintaan akan ditanggapi dengan oleh *web server* dalam bentuk data tidak terstruktur. Sama halnya dengan penelitian yang dilakukan (Ristijana dkk., 2021), dalam penelitian ini menggunakan *beautifulsoup library* untuk penyimpanan sementara dari laman *HyperText Markup Language* (HTML). Selanjutnya data tersebut disimpan sementara ke dalam tabel berupa *Dataframe* yang berupa tipe data dari *library pandas*.

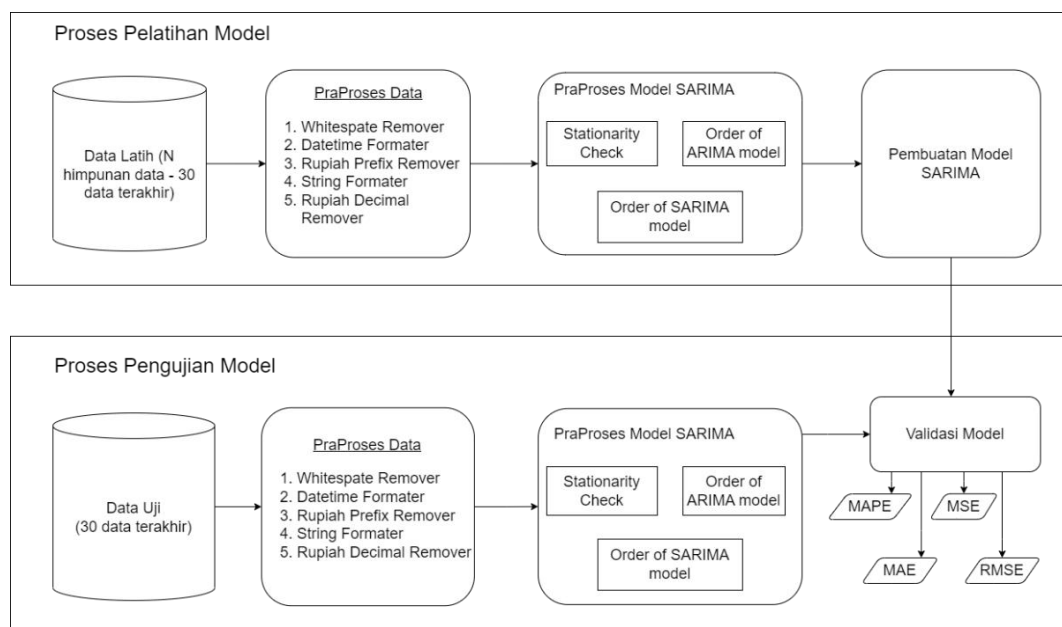
Skema penelitian ini juga akan diterapkan untuk mengambil atribut spesifik di masing – masing situs LPSE yang dapat dilihat pada kamus data terletak pada Tabel 3.1. Hasil dari data yang berhasil diakuisisi akan berbentuk dokumen dengan ekstensi *parquet* yang nantinya dilakukan proses *union data* agar nantinya

himpunan data dapat menjadi satu kesatuan dalam proses pengimplementasian model prediksi SARIMA.

### 3.3.2 Skenario Konstruksi Model Komputasi

Dalam membangun model prediksi SARIMA, dibutuhkan sebuah pedoman agar memenuhi serangkaian proses yang bersumber dari studi literatur. Oleh karenanya, dibentuk sebuah skenario konstruksi model SARIMA (model komputasi) yang dapat dilihat pada Gambar 3.7 di bawah ini. Dalam mengimplementasikan algoritma SARIMA, sistem yang dibangun akan dibagi menjadi dua mekanisme utama yaitu pada data latih dan data uji. Kedua data ini bersumber dari himpunan yang sama yakni hasil agregasi dan imputasi nilai hilang dari data tender yang berhasil diakuisisi dengan granularitas harian.

Perlu diperhatikan bahwa objek penelitian yang digunakan ialah nilai HPS dan juga margin tender, sehingga proses prediksi yang diimplementasikan akan dibedakan. Hal yang perlu diperhatikan dalam menganalisis suatu himpunan data adalah tahap prapemrosesannya. Di tahap ini, hasil dari *web scraping* akan ditransformasi ke dalam bentuk yang cocok sesuai dengan tujuan analisisnya nanti yang terdiri dari lima tahapan dan kerap diistilahkan sebagai proses *data cleaning*.



Gambar 3.7 Rancangan Model Komputasi

Setelah proses *data cleaning* selesai, serangkaian konstruksi praproses model SARIMA akan dilakukan. Diawali dengan *stationarity check* (pengujian stasioneritas). Uji stasioneritas dapat menggunakan dua jenis pengujian. Pertama,

Daud Fernando, 2023

**PREDIKSI TENDER PADA SITUS PELELANGAN LPSE INDONESIA MENGGUNAKAN ALGORITMA SEASONAL AUTOREGRESSIVE MOVING AVERAGE (SARIMA)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu



uji statistik *Augmented Dickey Fuller* (ADF) yang akan meninjau dari hasil dari rata-rata, varians, dan statistik deskriptif lainnya harus konstan dan diistilahkan sebagai data deret waktu yang *white noise*. Kedua, uji statistik *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) yang akan meninjau dari tersedia atau tidaknya akar unit dari data deret waktu yang dihasilkan. Pengujian stasioneritas ini berbasis kepada pengambilan keputusan pada hipotesis null dan bila datanya belum stasioner harus dilakukan proses *differencing* terlebih dahulu.

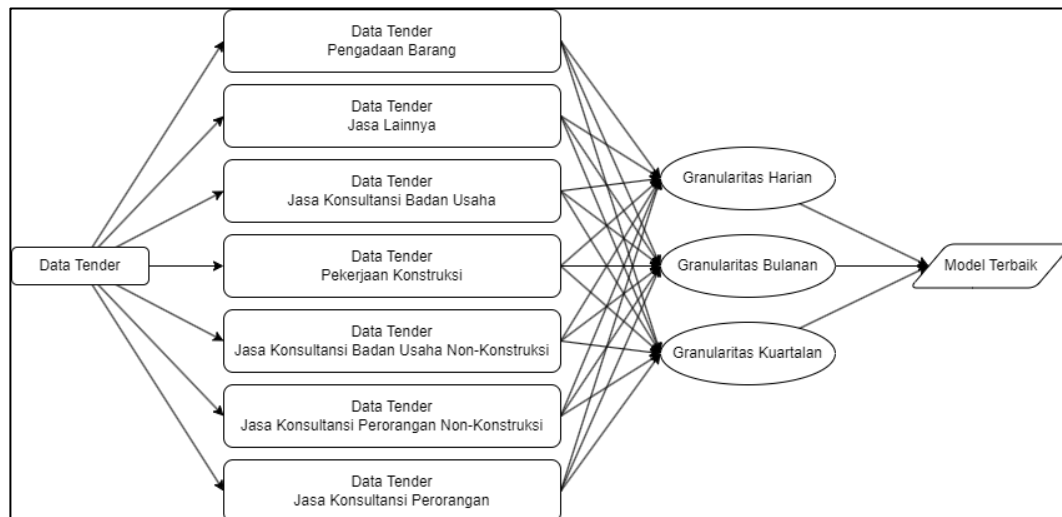
Apabila uji stasioneritas menunjukkan pemenuhan hipotesis bahwa data deret waktunya stasioner, maka melaju ke tahap penentuan ordo model ARIMA terlebih dahulu. Penentuan ordo dari model ARIMA ditentukan berdasarkan nilai autokorelasi antar lag data historis yang ada. Nilai autokorelasi ini yang akan menjadi penentu (seakan seperti bobot) dalam memprediksi nilai di masa depan. Data yang digunakan pada model ini adalah data deret waktunya langsung. Sedangkan proses selanjutnya adalah penentuan ordo model SARIMA yang menggunakan atribut musiman dari data deret waktu yang digunakan agar mempertimbangkan aspek musiman ketika modelnya SARIMA dibuat.

Dari kedua bagian data tersebut, maka terbentuklah sebuah model SARIMA yang telah berbasis literatur dan pemenuhan asumsi yang dibutuhkan. Sehingga, dapat dilakukan konstruksi model SARIMA menggunakan *library scikit-learn* yang selanjutnya masuk ke dalam tahapan validasi model yang menerapkan persamaan (16), (17), (18), dan (19) agar dapat diukur seberapa baik performansi yang dihasilkan dari model SARIMA yang dibentuk. Perlu diperhatikan kembali, model SARIMA akan dikategorisasi berdasarkan persentase nilai MAPE yang dihasilkan sesuai rujukan dari buku Lewis (1982) dan tersedia pada Tabel 2.1 di mana nilai MAPE harus berada di bawah 50% agar nilai prediksi yang dihasilkan dapat dipertanggungjawabkan dan masuk akal sebagai landasan keputusan nilai masa depan dari data historis yang digunakan.

### 3.3.3 Skenario Eskalasi Performansi Model Prediksi Berbasis Eksperimen

Dalam metode penelitian eksperimental, pengoptimalan dari hasil penelitian harus terus dilakukan dengan penyesuaian variabel independen terhadap variabel dependennya. Sehingga, harapannya, nilai dari performansi model SARIMA yang dibentuk dapat meningkat. Sebagai upaya yang dilakukan, akan dilihat terlebih

dahulu statistik deskriptif dari sebaran data agregasi granularitas harian lengkap dengan kolom bertipe data *datetime* pada himpunan datanya yang sebagai nilai masukkan untuk model komputasi SARIMA pertama. Secara keseluruhan skenario eksperimen yang digunakan dapat dilihat pada Gambar 3.8 di bawah ini.



Gambar 3.8 Skema Eksperimen Penelitian

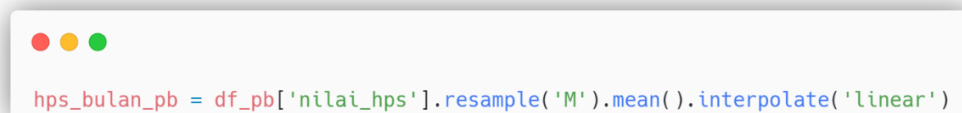
Proses pertama yang dilakukan adalah menganalisis terlebih dahulu data tender yang berhasil diakuisisi per jenis tender yang didapatkan yakni tersedia tujuh jenis tender. Kategorisasi data tender tersebut sesuai dengan jenisnya masing – masing dalam satu variabel untuk bisa diakses dalam pembentukan model komputasinya. Mekanisme yang dilakukan dalam pemisahan jenis tender dapat dilihat pada Gambar 3.9 di bawah ini.

```
df_pb = df_clean[df_clean['jenis_tender']=='Pengadaan Barang']
```

Gambar 3.9 Pemisahan Data Per Jenis Tender

Hal pertama yang perlu dilakukan dari Gambar 3.9 adalah menyeleksi nilai dari atribut himpunan data *jenis\_tender* di mana kolom dalam tabel yang menampung kategorisasi tender yang ada. Kemudian berikan nilai sesuai dengan jenis tendernya masing – masing di variabel yang berbeda – beda. Lakukan proses ini untuk semua jenis tender yang ada dan kemudian dari variabel tersebut dilakukan proses agregasi dan imputasi bertipe linear sebagai nilai masukkan ke

dalam model komputasi yang dibuat. Adapun caranya dapat dilihat pada Gambar 3.10 di bawah ini. Metode *resample* akan mengambil data per granularitas argumen yang diberikan, dalam kasus ini granularitas ‘M’ artinya bulan.

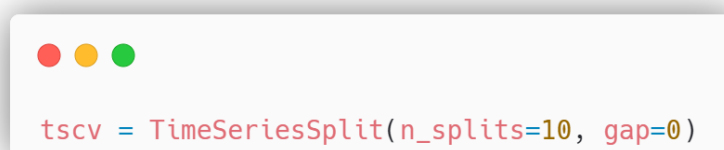


```
hps_bulan_pb = df_pb['nilai_hps'].resample('M').mean().interpolate('linear')
```

Gambar 3.10 Agregasi dan Imputasi Data Tender sebagai Input Model Komputasi

Setelahnya, berikan metode agregator yang akan digunakan yaitu rata-rata. Sehingga, metode yang dipanggil adalah *mean()* dan setelahnya ketika ada suatu tanggal nilainya hilang maka akan diimputasi berbasis teknik linear dengan metode *interpolate*. Langkah ini akan berlangsung secara repetitif per tujuh jenis tender dan di setiap jenis tender granularitasnya diberikan nilai harian, bulanan, dan kuartal. Dalam granularitas harian gunakan argumen ‘D’ dan kuartal menggunakan argumen ‘Q’.

Ketika semua model telah terbentuk berbasis skema eksperimen yang dibuat, saatnya melakukan proses model komputasi lengkap dengan proses validasi modelnya juga. Selain mekanisme 30 data terakhir untuk validasi model komputasi yang dibuat, skenario validasi lainnya adalah *time series cross validation*. Melansir Gambar 2.18 mengenai mekanisme pengerjaan validasi bertipe ini, maka dapat diimplementasikan ke dalam bahasa pemrograman Python seperti pada Gambar 3.11 di bawah ini.



```
tscv = TimeSeriesSplit(n_splits=10, gap=0)
```

Gambar 3.11 Base Code Time Series Cross Validation

Penggunaan nilai argumen *n\_splits* sebanyak 10 artinya memberikan iterasi sebanyak 10 kali secara sekuensial ketika proses validasi pada data latih dieksekusi. Hal tersebut yang memunculkan setiap iterasi mendapatkan metrik performansinya masing – masing. Nilai di setiap iterasinya kemudian dilakukan agregasi dengan penggunaan nilai rata-rata. Hal tersebut digunakan supaya

tergambar secara keseluruhan performansi model dari skenario validasi ini menggunakan keseluruhan data latihnya. Sehingga, performansi model yang dihasilkan akan mengeluarkan delapan metrik. Empat metrik untuk skenario validasi 30 data terakhir dan empat metrik lagi berbasis pada skenario *time series cross validation* dari model komputasi yang dibuat.

### 3.4 Analisis Data

Data yang akan digunakan dalam penelitian ini terdiri dari beberapa atribut sebagai berikut:

Tabel 3.1  
Kamus Himpunan Data

| Nama Atribut                 | Deskripsi Atribut   |
|------------------------------|---|
| Tanggal Pembuatan            | Tanggal ketika tender dilelangkan di sistem LPSE  |
| Jenis Tender                 | Kategori dari tender yang dilelangkan, terdapat sepuluh kategori yang tersedia                          |
| Harga Kontrak (dalam Rupiah) | Harga kontrak dari tender yang telah terdapat pemenangnya (tender selesai)                              |
| Nilai HPS (dalam Rupiah)     | Harga perkiraan sendiri suatu tender yang ditetapkan oleh penyelenggara                                 |
| Margin Tender (dalam Rupiah) | Hasil rekayasa data dengan mengurangi kolom harga kontrak dengan nilai HPS                              |
| Persentase Margin (%)        | Hasil rekayasa data untuk melihat seberapa besar persentase penurunan harga kontrak terhadap nilai HPS. |

Untuk empat atribut yang berada di Tabel 3.1 merupakan hasil dari proses sistem akuisisi data, sedangkan atribut Margin Tender dan Persentase Margin merupakan hasil dari rekayasa fitur (*feature engineering*). Tabel 3.2 memuat lima sampel himpunan data yang akan menjadi nilai *input* atau nilai masukkan dengan menggunakan metode – metode yang ada. Tabel tersebut dapat dilihat pada halaman selanjutnya.

Tabel 3.2  
Tiga Sampel Himpunan Data

| <b>Tanggal Pembuatan</b> | <b>Jenis Pengadaan</b>       | <b>Harga Kontrak (dalam Rupiah)</b> | <b>Nilai HPS (dalam Rupiah)</b> | <b>Margin Tender (dalam Rupiah)</b> | <b>Margin (%)</b> |
|--------------------------|------------------------------|-------------------------------------|---------------------------------|-------------------------------------|-------------------|
| 2022-09-08               | Pekerjaan Konstruksi         | 1904116190                          | 1939998469                      | 35882279                            | 1.8496            |
| 2021-03-04               | Pengadaan Barang             | 1782825000                          | 1798500000                      | 15675000                            | 0.8715            |
| 2019-07-12               | Jasa Konsultansi Badan Usaha | 196900000                           | 198440000                       | 1540000                             | 0.7760            |

Perlu diperhatikan kembali, data yang berhasil diakuisisi dari situs web LPSE adalah sebanyak 785.361 data tender yang memiliki rentang waktu publikasi tender sejak 2013-01-07 hingga 2022-11-30 dan atribut tersebut menjadi indeks dari himpunan data yang digunakan. Selanjutnya, akan dilakukan proses pemisahan antara nilai HPS dan juga margin tendernya. Hal ini supaya masing – masing objek penelitian tersimpan ke dalam variabel yang berbeda dan dapat dengan mudah untuk diakses ketika model komputasi perlu proses pembelajaran (*training*). Sebagai proses agregasi dari nilai HPS per granularitas Hari dapat dilihat pada Tabel 3.3 di bawah ini.

Tabel 3.3  
Data *Moving Average* Nilai HPS Tender Granularitas Harian

| <b>Tanggal Pembuatan</b> | <b>Nilai HPS Tender (dalam Rupiah)</b> |
|--------------------------|--|
| 2013-01-04               | 485287803.33                           |
| 2013-01-05               | 505567410.18                           |
| 2013-01-06               | 525847017.03                           |

Lain halnya untuk nilai margin tender, tiga cuplikan datanya dengan granularitas bulanan dapat dilihat pada Tabel 3.4 di halaman selanjutnya.

Tabel 3.4

Data *Moving Average* Nilai Margin Tender Granularitas Bulanan

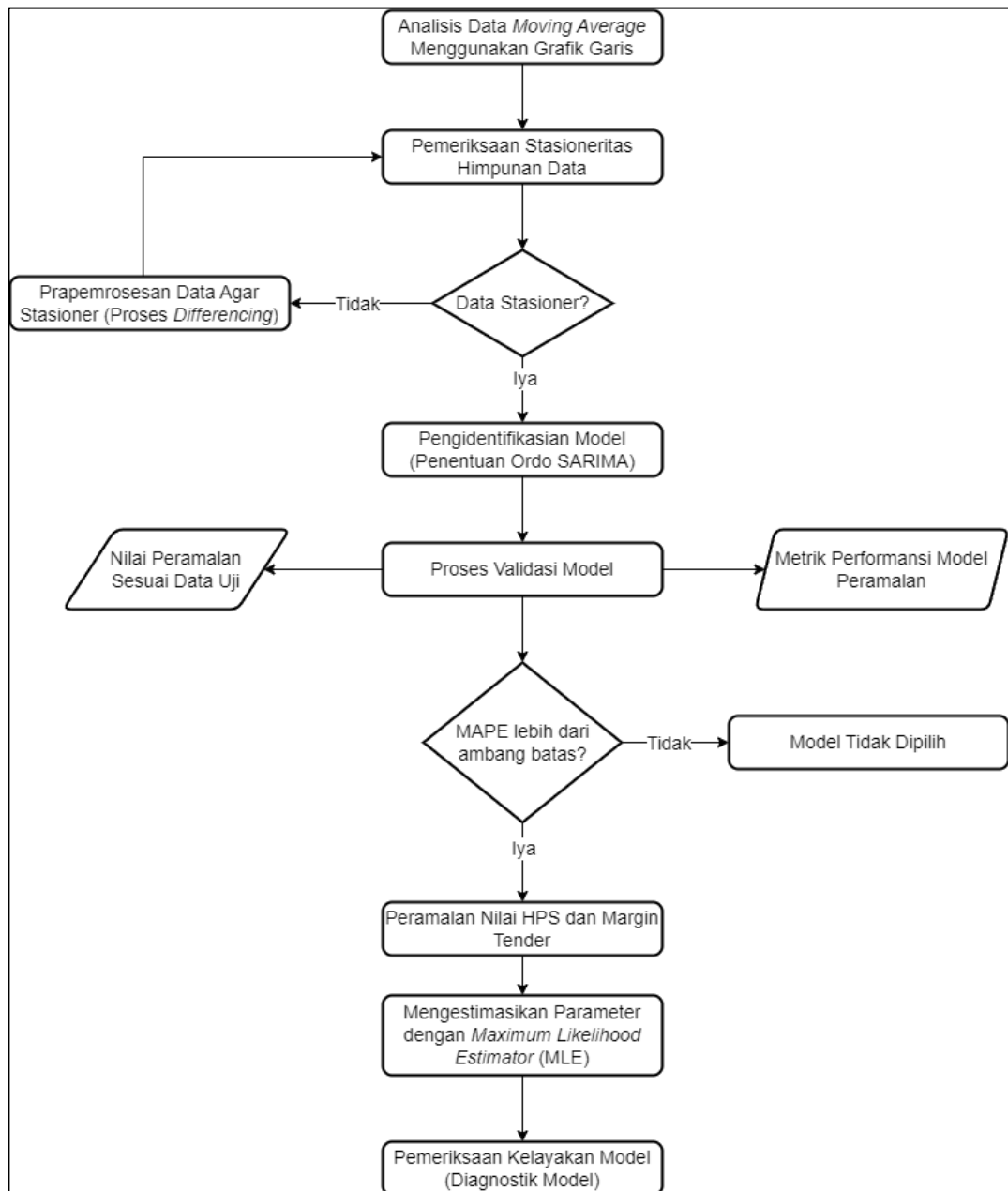
| Tanggal Pembuatan | Nilai Margin Tender (dalam Rupiah) |
|-------------------|------------------------------------|
| 2013-01-31        | 36327938.05                        |
| 2013-02-28        | 36758993.05                        |
| 2013-03-31        | 33864856.47                        |

Terlihat perbedaan yang signifikan antara Tabel 3.3 dengan Tabel 3.4 ada pada granularitas atribut Tanggal Pembuatan. Bila granularitas yang digunakan adalah harian maka unit dari atribut tanggal pembuatan berjarak satu hari antar baris data tendernya. Lain halnya untuk Tabel 3.4 yang memiliki granularitas bulanan pada nilai margin tender. Jarak antar satu tender dengan tender lainnya adalah satu bulan dan di tiap bulannya unit harian atau tanggal yang digunakan adalah tanggal terakhir dari bulan tersebut dan merupakan nilai agregasi pada periode waktu, sebagai contohnya di bulan Januari 2013, memiliki *time frame* 2013-01-01 hingga 2013-01-31. Semua nilai margin tender yang memiliki tanggal publikasi pada *time frame* tersebut akan diagregasi menggunakan nilai rata-rata yang ditampilkan pada kolom selanjutnya alias atribut nilai margin tender.

Dalam mengimplementasi algoritma SARIMA untuk membuat modelnya diperlukan sebuah skema agar memenuhi penelitian – penelitian terdahulu. Adapun skema yang dimaksud dapat dilihat pada Gambar 3.12 pada halaman selanjutnya sebagai proses analisis data hasil dari proses *web scraping* yang dilakukan. Analisis Data yang digunakan akan berlangsung beberapa tahap. Tahapan tersebut di antaranya :

1. Analisis Data *Moving Average* Menggunakan Grafik Garis

Tahapan pertama dalam menganalisis data deret waktu adalah dengan memvisualisasikannya ke dalam grafik garis. Dalam hal ini, rata-rata dalam nilai HPS dan juga margin tender per periode waktu (hari, bulan, dan kuartal) akan berada pada sumbu y dan nilai dari sumbu x adalah masing – masing periode waktunya. Grafik garis ini pula dapat menjadi pengambilan kesimpulan data historis mengenai performansi suatu tender berdasarkan atribut deret waktunya di beragam periode, apakah mengalami kenaikan atau penurunan.



Gambar 3.12 Diagram Analisis Data

## 2. Pemeriksaan Stasioneritas Himpunan Data

Stasioneritas adalah kondisi suatu nilai rata-rata dan juga nilai variansi sepanjang deret waktu yang stabil dalam himpunan datanya. Untuk mengujinya dapat dilakukan dengan uji unit *root* menggunakan *Augmented Dickey-Fuller Test* (ADF) (Fauziah dkk., 2021). Bila didapati datanya stasioner maka dilanjutkan kepada tahap selanjutnya. Namun bila tidak harus melakukan proses transformasi data terlebih dahulu, salah satunya menggunakan transformasi *differencing* yang

bergantung kepada nilai variansi datanya harus dilakukan seberapa banyak proses transformasi.

### 3. Pengidentifikasian Model (Penentuan Ordo SARIMA)

Model SARIMA dapat ditentukan melalui penentuan dari tujuh parameter utama, hal ini membuat model yang dapat digunakan ada banyak sekali kandidatnya. Untuk menyeleksi model SARIMA dengan parameter atau ordo  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$ , dan  $m$  haruslah berbasis nilai dari PACF dan juga ACF. Untuk menentukan orde pada model  $SARIMA(p, d, q)(P, D, Q)_m$  dapat melalui plot PACF milik data deret waktunya untuk orde  $p$ , plot ACF milik data deret waktunya untuk orde  $q$ , banyaknya *differencing* yang dilakukan guna mendapatkan data deret waktu yang stasioner pada orde  $d$ , dan hal yang serupa untuk orde  $P$ ,  $Q$ , dan  $D$  hanya saja ditinjau dari efek musiman dari data deret waktunya. Kemudian nilai orde  $m$  ditentukan dari rentang perulangan suatu nilai musiman yang terjadi pada data deret waktunya.

### 4. Proses Validasi Model

Validasi model dilakukan dengan membagi himpunan data menjadi dua bagian. Bagian pertama adalah data latih yang dijadikan sebagai nilai masukkan model untuk mempelajari tren data berderet waktu. Lalu bagian kedua adalah data uji, bagian ini akan menjadi landasan performansi dari suatu model ketika melakukan prediksi. Karena nilai dari prediksi akan dihitung selisihnya dengan data uji yang ada.

Penelitian ini akan membagi proses validasi ke dalam dua skenario. Skenario pertama, kuantitas data latih secara manual diambil 30 data terakhirnya untuk semua granularitas yang dipakai (hari, bulan, dan kuartal). Skenario kedua adalah penggunaan teknik *time series cross validation*. Teknik ini akan mengizinkan model membagi himpunan data ke sepuluh kelompok seperti yang ditunjukkan pada Gambar 2.18 hasil dari penelitian Wulandari, dengan asumsi mendapatkan nilai performansi yang cukup optimal ketika model melakukan prediksi terhadap data latih dan dibandingkannya dengan data uji. Setelahnya akan didapatkan nilai prediksi dan juga beberapa metrik yang mengukur performansi dari modelnya.



#### 5. Penentuan Ambang Batas MAPE Model SARIMA

Selain beberapa asumsi yang sudah dijelaskan ketika membuat model SARIMA, dalam penelitian ini akan dipadupadankan dengan ambang batas metrik MAPE. Setidaknya nilai evaluasi yang berada di bawah nilai 50% akan ditetapkan sebagai model yang masuk akal dalam melakukan prediksi (semakin rendah semakin baik). Bila model memenuhi kualifikasi tersebut, maka akan digunakan untuk melakukan prediksi dan dilakukan interpretasi modelnya. Namun, bila model berada di atas rentang tersebut maka model akan dibuang dan tidak digunakan untuk proses prediksi.

#### 6. Mengestimasi Parameter dengan *Maximum Likelihood Estimator* (MLE)

Setelah model SARIMA mendapatkan model potensial lengkap dengan parameter / ordonya. Kemudian dilakukan pengestimasi terhadap model – model tersebut untuk melihat seberapa besar kontribusi data *history* dalam melakukan prediksi. Nilai ini akan bergantung terhadap banyaknya ordo yang ditetapkan ke modelnya. Bila nilai ordo p sama dengan lima maka akan ada lima parameter dari model tersebut dan berlaku untuk ordo lainnya.

#### 7. Pemeriksaan Kelayakan Model

Kelayakan suatu model dinilai dari nilai galat atau residual yang dihasilkan (F. A. Putri dkk., 2020). Bila residual memenuhi perilaku *white noise*, maka model tersebut dapat ditetapkan sebagai model yang layak. Di mana *white noise* artinya residual tidak memiliki auto korelasi. Dalam melakukannya terdapat beberapa tahapan yang dilakukan untuk mengevaluasi nilai residual suatu model ialah dengan pengujian menggunakan Q-Ljung Box.