

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Prediksi (*Forecasting*)**

Peramalan (*forecasting*) adalah kegiatan memperkirakan atau memprediksikan sesuatu hal yang akan terjadi pada masa yang akan datang dengan rentang waktu yang relative lama. Ramalan adalah suatu kondisi yang akan diperkirakan akan terjadi pada masa yang akan datang. Untuk memprediksi, kita memerlukan data yang akurat dan berasal dari data masa lalu/lampau, sehingga dapat dilihat prospek situasi dan kondisi dimasa yang akan datang [2][3].

Pada umumnya kegunaan peramalan adalah sebagai berikut [4] :

- a. Sebagai alat bantu dalam perencanaan yang efektif dan efisien.
- b. Untuk menentukan kebutuhan sumber daya dimasa mendatang.
- c. Untuk membuat keputusan yang tepat.

Fungsi kita melakukan peramalan yaitu dapat membantu dalam pengambilan keputusan. Keputusan yang baik adalah keputusan berdasarkan pertimbangan data-data yang sudah ada juga berdasarkan sesuatu apa yang akan terjadi pada organisasi/perusahaan. Ketetapan prediksi dapat mempengaruhi baik dan tidaknya suatu. Namun perlu kita ketahui tidak selamanya peramalan itu sepenuhnya benar, sehingga yang perlu diperhatikan adalah usaha untuk memperkecil kesalahan dari ramalan tersebut [5].

#### **2.2 Time Series Model**

*Time Series* (data berkala) adalah data yang didasarkan pada urutan titik dengan jarak yang sama (mingguan, bulanan, triwulanan, dan seterusnya). Peramalan data deret waktu menyiratkan bahwa nilai masa depan diprediksi hanya dari nilai masa lalu dan variabel lain, tidak peduli seberapa berpotensi nilainya, dapat diabaikan [6].

Analisis data berkala memungkinkan kita untuk mengetahui perkembangan suatu atau beberapa kejadian serta hubungan/pengaruhnya terhadap kejadian lainnya [7]. Pola pergerakan data atau nilai-nilai variabel dapat diketahui dengan adanya data berkala, sehingga data berkala dapat dijadikan sebagai dasar untuk pembuatan keputusan, peramalan pada masa yang akan datang, serta perencanaan kegiatan dimasa yang akan datang [8].

### 2.3 Machine Learning

*Machine learning* adalah sebuah aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang. Proses pembelajaran yang dimaksud adalah suatu usaha dalam memperoleh kecerdasan yang melalui dua tahap antara lain latihan (*training*) dan pengujian (*testing*). *Machine learning* berkaitan dengan pertanyaan tentang bagaimana membangun program komputer agar meningkat secara otomatis dengan berdasar dari pengalaman.

Menurut Kusumaningrum ada empat jenis pembelajaran dalam *machine learning* yaitu [9]:

1. *Supervised Learning* yaitu pembelajaran dengan data (*input*) dan keluaran (*output*) yang sudah ditentukan.
2. *Unsupervised Learning* yaitu pembelajaran dengan data (*input*) tidak mencakup keluaran (*output*) yang ditentukan.
3. *Semi-supervised Learning* yaitu pembelajaran dengan data (*input*) mencakup beberapa keluaran (*output*) yang ditentukan.
4. *Reinforcement Learning* adalah pembelajaran dengan data (*input*) dan keluaran (*output*) yang tidak ditentukan atau dengan kata lain mesin bekerja secara otomatis dalam menentukan perilaku yang ideal sehingga dapat memaksimalkan kinerja algoritmanya.

## 2.4 Normalisasi Data

Normalisasi data adalah sebuah proses transformasi nilai untuk merubah nilai data. Normalisasi digunakan untuk menyamakan skala atribut data kedalam range yang spesifik [10]. Proses ini dilakukan karena beberapa data terdapat rentang nilai yang berbeda. Normalisasi data akan menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses [11]

$$x' = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

Persamaan di atas akan menghasilkan nilai baru  $x'$  yang memiliki rentang antara 0 sampai 1, atau dengan kata lain nilai terendah kini menjadi 0 dan nilai tertinggi menjadi 1.

## 2.5 *Least Square Support Vector Machine*

*Least Squares Support Vector Machine* adalah pengembangan dari metode Support Vector Machine, perbedaannya adalah least square support vector machine memiliki performa yang lebih baik. Least Squares Support Vector Machines merupakan formulasi ulang terhadap Support Vector Machine standar yang mengarah untuk memecahkan sistem linear [12].

Support Vector Machine pada dasarnya berhubungan dengan masalah klasifikasi dua kelas yang khas. Support Vector Machine awalnya dikembangkan oleh Boser dan Vapnik, didasarkan pada Vapnik, teori Chervonenkis dan prinsip meminimalkan risiko struktural. Dengan mencoba menemukan *trade-off* antara meminimalisir kesalahan set data *training* dan memaksimalkan margin untuk mencapai kemampuan generalisasi terbaik dan tetap menjaga kesesuaiannya. Support Vector Machine adalah menggunakan *quadratic programming*, yang menyediakan *global minima* [13].

Seiring berjalanya waktu *Support Vector Machine* berkembang menjadi *Least Square Support Vector Machine*, hal ini disebabkan untuk mengatasi *quadratic*

*programming* maka pada *Least Square Support Vector Machine* diubah menjadi menggunakan satu set persamaan linier untuk pelatihan. Perubahan ini ditujukan untuk mengurangi banyak memori maupun processor pada saat memproses data [12].

Menurut Carlos et al, kinerja *Least Square Support Vector Machine* dapat mendukung untuk data tahunan dan mengevaluasi kesalahan relative 0,91%, 1,86%, dan 0,93% terdeteksi. Ini menurut para penulis, menunjukkan bahwa *Least Square Support Vector Machine* sangat akurat untuk peramalan. Berdasarkan penelitian mereka, penulis menyimpulkan bahwa model yang baik untuk digunakan dalam hal peramalan.[3]

Pada penelitian ini penulis akan melakukan pemrosesan data untuk menghasilkan prediksi penggunaan pupuk subsidi. Dalam hal ini penulis menggunakan metode *Least Square Vector Support Machine*. Metode ini pengembangan dari metode *Support Vector Machine*, metode *Least Square Support Vector Machine* penulis gunakan karena penulis akan memproses data yang besar dan banyak secara berkala sehingga pada saat pemrosesan data membutuhkan memori dan processor yang minim tetapi dengan performa yang bagus [12].

Dalam sebuah penelitian yang dilakukan oleh Fahteem dan Ahmad menunjukkan hasil prediksi dalam memprediksi irradansi solar mampu mencegah ketidakseimbangan produksi listrik yang disebabkan oleh radiasi yang tidak pasti [14]. Berdasarkan hasil penelitian tersebut maka penulis menggunakan metode *Least Square Support Vector Machine* ini dengan harapan mampu meprediksi jatah pupuk subsidi dan yang akan mendatang sehingga dapat membantu dinas terkait dalam mempersiapkan stok pupuk subsidi.

## **2.6 Autoregressive Integrated Moving Average (ARIMA)**

Metode *Autoregressive Integrated Moving Average (ARIMA)* atau biasa disebut juga sebagai metode *Box-Jenkins* merupakan metode yang secara intensif dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970 [15].

Model AR, MA dan ARMA menggunakan asumsi bahwa data deret waktu yang dihasilkan sudah bersifat stasioner. Pada kenyataannya, data deret waktu lebih banyak bersifat tidak stasioner. *Model Autoregressive and Moving Average* tidak mampu menangani data yang tidak stasioner karena data yang tidak stasioner akan sangat sulit untuk dimodelkan dengan alasan tersebut model *ARIMA* dikembangkan. Data yang tidak stasioner menyebabkan rata-rata dan variannya berubah, dampaknya adalah dapat menyebabkan kesalahan perhitungan [16].

Analisis Data dilakukan menggunakan metode *ARIMA* dengan software statistika yaitu Eviews 11 sebagai alat bantu. Tahapan dalam menerapkan metode *ARIMA*:

1. Pemeriksaan Kestasioneran Data
2. Identifikasi model dalam *ARIMA*. Melalui plot ACF (Fungsi Autokorelasi) dan PACF (Fungsi Autokorelasi Parsial) kita dapat menentukan model *ARIMA* yang bisa digunakan dalam peramalan.
3. Penentuan Parameter p, d dan q dalam *ARIMA*.
4. Penentuan persamaan model *ARIMA* dengan melihat nilai *Akaike Info Criterion (AIC)* dan *Schwarz Bayesian Criterion (SBC)* yang terkecil.
5. Validasi peramalan.
6. Prediksi jumlah penebusan pupuk subsidi menggunakan model *ARIMA* terbaik yang telah ditetapkan.

*Autoregressive Integrated Moving Average Model (ARIMA)* adalah model umum dari *Autoregressive Moving Average (ARMA)* yang menggabungkan proses *Autoregressive (AR)* dan *Moving Average (MA)* membentuk model komposit dari *time series*. Seperti yang ditunjukkan oleh persamaan  $ARIMA(p,d,q)$  inti dari model tersebut adalah [17] :

1. *AR: Autoregressive*. Model regresi yang menggunakan nilai ketergantungan antara pengamatan dan sejumlah pengamatan data hsitori (p).
2. *I: Integrated*. Untuk membuat deret waktu stasioner dengan mengukur perbedaan pengamatan pada waktu yang berbeda (d).

3. *MA: Moving Average*. Suatu pendekatan yang memperhitungkan ketergantungan antara pengamatan dan istilah kesalahan residual ketika model rata-rata bergerak digunakan untuk pengamatan tertinggal ( $q$ ).

Eksperimen menunjukkan bahwa model ARIMA yang dibangun mampu memprediksi harga bahan pokok dengan cukup akurat dengan rata-rata error 2.22% [18].

### 2.6.1 Uji Stasioner Menggunakan *Augmented Dickey Fuller (ADF)*

Langkah pertama dalam mengidentifikasi model  $ARIMA(p, d, q)$  yaitu melakukan uji stasioneritas. Uji stasioner bertujuan untuk mengetahui apakah data runtut waktu yang digunakan sudah stasioner atau belum. Jika data mengalami ketidakstasioneran pada level series maka perlu dilakukan differensiasi level 1 atau 2 [19]. Proses differensiasi yaitu mengurangi nilai data pada suatu periode dengan nilai data periode sebelumnya [16].

Pada penelitian ini uji stasioner didasarkan pada *Augmented Dickey Fuller (ADF)* dimana pada pengujian menggunakan *Augmented Dickey Fuller* akan menghasilkan nilai ADF, jika nilai ADF lebih besar dari nilai kritis yaitu 5% atau 0,05 maka data deret waktu bisa dikatakan stasioner [16].

### 2.6.2 Identifikasi Model ARIMA

Proses identifikasi Model ARIMA dapat menggunakan plot ACF (Fungsi Autokorelasi) dan PACF (Fungsi Autokorelasi Parsial) dengan cara melihat pola fungsi ACF dan PACF dari data yang bertujuan untuk menentukan parameter  $p$  dan  $q$  [16].

### 2.6.3 Penentuan Parameter Model ARIMA

Parameter yang akan dicari adalah  $p$ ,  $d$  dan  $q$ .  $d$  adalah nilai differensiasi, jika data stasioner pada level series maka nilai  $d$  adalah 0 dan jika data stasioner pada level 1 maka nilai  $d$  adalah 1 [16]. Untuk nilai  $p$  dan  $q$  didapat dari melihat plot ACF dan PACF. Plot ACF digunakan untuk menentukan nilai  $q$ , dapat dilihat dari *cut off pada lag* ke  $n$ . Untuk plot PACF digunakan untuk menentukan nilai  $p$ , dapat dilihat dari *cut off pada lag* ke  $n$  [20].

Jika hasil pengamatan plot ACF dan PACF menghasilkan lebih dari 1 model *ARIMA* maka untuk menentukan model terbaik berdasarkan nilai *AIC* (*Akaike's Information Criterion*) dan *SBC* (*Schwartz Bayesian Criterion*). *AIC* dan *SBC* menunjukkan nilai *error* dari model, maka semakin kecil nilai *AIC* dan *SBC* semakin bagus model tersebut [20].

## 2.7 Mean Absolute Deviation (MAD)

Tahapan pertama dalam mengevaluasi model menggunakan *Mean Absolute Deviation (MAD)*. Yaitu sebuah metode untuk melakukan evaluasi atau pengujian pada metode peramalan menggunakan jumlah kesalahan mutlak. Rumus yang digunakan untuk menghitung MAD ditunjukkan pada persamaan berikut [21]:

$$MAD = \frac{\sum_{t=1}^n |Y_t - F_t|}{n}$$

Keterangan

$X_t$  = Data aktual waktu  $t$

$F_t$  = Nilai peramalan pada waktu  $t$

$n$  = banyaknya data

## 2.8 Mean Square Error (MSE)

*Mean square error* adalah pendekatan lain untuk menguji tingkat kesalahan peramalan. Setiap kesalahan dikuadratkan. Teknik pendekatan ini menetapkan kesalahan yang besar dari peramalan karena kesalahan dikuadratkan. *MSE* adalah cara kedua untuk mengukur besarnya kesalahan peramalan secara keseluruhan. *MSE* adalah rata-rata dari selisih kuadrat antara nilai prediksi dan nilai sebenarnya. Rumus *MSE* yang digunakan ditunjukkan pada persamaan berikut [21]:

$$MSE = \sum_{t=1}^n \frac{(X_t - F_t)^2}{n}$$

Keterangan

$X_t$  = Data aktual waktu  $t$

$F_t$  = Nilai peramalan pada waktu  $t$

$n$  = banyaknya data

## 2.9 Mean Absolute Percentage Error (MAPE)

*MAPE* dihitung dengan menggunakan kesalahan absolut pada tiap periode dibagi dengan nilai observasi yang nyata untuk periode itu dan kemudian nilai tersebut dirata-ratakan [22]. *MAPE* mengindikasikan seberapa besar kesalahan dalam meramal yang dibandingkan dengan nilai nyata. *MAPE* diperoleh dengan menggunakan persamaan berikut ini [21]:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right|$$

Keterangan

$X_t$  = Data aktual waktu  $t$

$F_t$  = Nilai peramalan pada waktu  $t$

$n$  = banyaknya data

## 2.10 Kartu Petani Berjaya

Menurut website resmi Dinas TPH Provinsi Lampung, Kartu Petani Berjaya atau bisa disingkat menjadi KPB adalah suatu program yang menghubungkan semua kepentingan pertanian dengan tujuan mencapai kesejahteraan petani dan semua pihak yang terlibat dalam proses pertanian secara bersama-sama.

Sementara itu, Prof. Dr. M. Yusuf Sulfarano Barusman, M.B.A selaku Ketua Tim Program Kartu Petani Berjaya Provinsi Lampung dalam laporannya menyatakan dengan adanya KPB diharapkan dapat meningkatkan nilai tukar petani (NTP), meningkatkan keuntungan usaha tani, efisiensi usaha tani, dan peningkatan produksi



serta produktivitas pertanian. Adapun pihak yang dilibatkan dalam KPB yakni petani, distributor sarana produksi, poktan, penyuluh, kios, lembaga keuangan, pembeli, para ahli dan pemerintah daerah.

## 2.11 Pupuk

Pupuk adalah suatu bahan yang mengandung satu atau lebih unsur hara atau nutrisi bagi tanaman untuk menopang tumbuh dan berkembangnya tanaman. Sementara pupuk bersubsidi adalah pupuk yang pengadaan dan penyalurannya mendapat subsidi dari Pemerintah untuk kebutuhan petani yang dilaksanakan atas dasar program Pemerintah. Pupuk subsidi ada lima jenis yaitu ZA, Urea, SP-36, NPK dan Organik.

## 2.12 Penelitian Sebelumnya

**Tabel 2.1 Hasil Penelitian Sebelumnya**

No	Judul	Tahun	Metode	Hasil
1	<i>Forecasting Water Demand in Residential, Commercial, and Industrial Zones in Bogotá, Colombia, Using Least-Squares Support Vector Machines</i>	2016	<i>Least Squares Support Vector Machines</i>	Model LS-SVM terbukti lebih unggul dari model FNN-BP dalam hal menghitung kebutuhan air secara akurat. Dalam model LS-SVM, sebagian besar hasil untuk nilai RMSE berada di bawah 2%. Hal ini menunjukkan bahwa distribusi kesalahan antara model dan data yang dilaporkan

				<p>sangat rendah. Adapun efisiensi model (diukur sebagai prediksi akurat dari data yang dilaporkan), nilai AARE berada di bawah 8% untuk model LS-SVM, yang mencerminkan kemampuan peramalannya. Dengan demikian, lebih banyak faktor yang disertakan menghasilkan peramalan yang lebih baik.</p> <p>Akhirnya, prediksi perilaku permintaan air jelas terbukti menjadi alat yang efektif untuk perencanaan dan pengelolaan kota, karena membantu mengidentifikasi kebutuhan akan keputusan administratif untuk mengatur konsumsi berbagai strata dan penggunaan.</p>
--	--	--	--	--

2	Penerapan <i>Least Squares Support Vector Machines (LSSVM)</i> dalam Peramalan Indonesia Composite Index	2021	<i>Least Squares Support Vector Machines</i>	Hasil dari penelitian ini menunjukkan bahwa akurasi dari Metode LSSVM sangat akurat, dibuktikan dengan nilai MSE terkecil yaitu 0.00025248 pada 1200 data histori. Maka dapat disimpulkan semakin banyak data histori maka hasil peramalan akan semakin akurat. Ketika data histori diperbanyak maka nilai MSE semakin kecil.
3	<i>Forecasting Performance of LS-SVM for Nonlinear Hydrological Time Series</i>	2011	<i>Least Squares Support Vector Machines</i>	Tujuan dari makalah ini adalah untuk menguji kelayakan penggunaan LS-SVM dalam peramalan deret waktu hidrologi nonlinier dengan membandingkannya dengan metode statistik seperti MLR dan metode heuristik seperti NNBP. Dan kami akan dengan jelas memverifikasi

				<p>kinerja prediksi dari tiga model sesuai dengan tingkat linier menggunakan data kebutuhan air harian dan rata-rata arus masuk harian dari data bendungan. Hasil studi utama diringkas sebagai berikut. Akurasi dan efisiensi peramalan LS-SVM lebih unggul daripada NBNP dan MLR untuk WD dan Df. Dengan nilai RMSE 56259.7 dan nilai MAPE 0.843 %.</p>
4	<p>Prediksi Harga Bahan Pokok Nasional Jangka Pendek Menggunakan ARIMA</p>	2017		<p>Hasil percobaan pada penelitian tersebut menghasilkan nilai rata-rata error 2.22% hal ini menunjukkan model <i>ARIMA</i> mampu meramalkan harga bahan pokok dengan cukup akurat. Dengan meningkatnya horizon prediksi menyebabkan</p>

				menurunya akurasi dari beberapa bahan pokok yang diuji coba.
5	Analisis Peramalan Ihsg Dengan Time Series Modeling Arima	2020	ARIMA	Dari hasil analisis dan pembahasan yang telah diuraikan, maka dapat ditarik kesimpulan bahwa metode ARIMA dapat dipergunakan untuk meramalkan pergerakan IHSG. Model terbaik yang dipergunakan berdasarkan pengujian yang dilakukan adalah ARIMA (7,3,1). Hasil peramalan dengan model ARIMA (7,3,1) ini tidak jauh berbeda dari nilai aktual IHSG. Hal ini dibuktikan dengan nilai RMSE 30.33293, MAE 22.99950 dan MAPE 0.002615.
6	<i>A Comparison Between ARIMA, LSTM, And GRU For Time Series Forecasting</i>	2019	ARIMA	Dari hasil tersebut menunjukkan bahwa Model ARIMA memberikan yang terbaik akurasi dan waktu dengan nilai MAPE 2.76, RMSE

				<p>302.53 dan kecepatan 0.0007 Detik. Hasil ini mungkin juga sebagai akibat dari beberapa faktor. Parameter yang dipilih dan total jumlah data juga dapat mempengaruhi hasil. Jumlah data yang kami miliki untuk makalah ini relatif kecil. RNN biasanya bekerja pada kumpulan data yang lebih besar, seperti yang ditunjukkan oleh makalah penelitian sebelumnya. Fitur terpilih yang dipilih mungkin tidak cukup untuk diprediksi harga Bitcoin secara akurat</p>
--	--	--	--	---