

## PREDIKSI JUMLAH PENGIRIMAN BARANG MENGGUNAKAN KOMBINASI METODE SUPPORT VECTOR REGRESSION, ALGORITMA GENETIKA DAN MULTIVARIATE ADAPTIVE REGRESSION SPLINES

Nendi<sup>1</sup>, Arief Wibowo<sup>2</sup>

<sup>1,2</sup>Universitas Budi Luhur

Email: <sup>1</sup>nendijuve@gmail.com, <sup>2</sup>arief.wibowo@budiluhur.ac.id

\*Penulis Korespondensi

(Naskah masuk: 31 Agustus 2019, diterima untuk diterbitkan: 26 November 2020)

### Abstrak

Salah satu sektor usaha di Indonesia yang telah berkembang saat ini adalah logistik. PT XYZ adalah sebuah perusahaan logistik nasional yang menyediakan jasa pengiriman barang dan melayani pengiriman ke banyak daerah di Indonesia. Sebagai perusahaan logistik berskala nasional, maka perusahaan didukung dengan armada sebanyak 2.100 unit kendaraan truk. Jumlah armada akan terus bertambah seiring dengan target yang dicanangkan perusahaan, yaitu pada tahun 2020 jumlah armada truk harus mencapai 6.000 unit truk. Saat ini strategi operasional logistik dihasilkan berdasarkan pengalaman *stakeholder*, hal ini tentu menimbulkan masalah dalam operasional perusahaan secara keseluruhan. Prediksi jumlah pengiriman barang harian dapat menjadi solusi dalam membantu perusahaan dalam merencanakan, memonitoring dan mengevaluasi strategi operasional logistik, berdasarkan kemampuan perusahaan dalam ketersediaan armada kendaraan untuk pengiriman. Studi ini mengusulkan kombinasi metode *Support Vector Regression* (SVR), algoritma genetika dan *Multivariate Adaptive Regression Splines* (MARS) untuk pemecahan masalah pada proses prediksi, termasuk dalam pemilihan data latih yang tepat. Hasil pengujian menunjukkan penggabungan tiga metode tersebut dapat menghasilkan prediksi jumlah pengiriman barang harian dengan nilai *Mean Absolute Percentage Error* (MAPE) yaitu 0,0969% dengan parameter *epsilon* ( $\epsilon$ ) 1,92172577675873E-20, *complexitas* ( $c$ ) 62 dan *gamma* ( $\gamma$ ) 1,0.

**Kata kunci:** *Prediksi, Support Vector Regression, algoritma genetika, Multivariate Adaptive Regression Splines.*

### **FREIGHT SHIPPING PREDICTION USING COMBINATION OF SUPPORT VECTOR REGRESSION, GENETIC ALGORITHM AND MULTIVARIATE ADAPTIVE SPLINES METHOD**

#### *Abstract*

*The logistics business sector has developed very rapidly in Indonesia today. PT XYZ is a national logistics company that provides freight forwarding services from one place to another. As a national-scale logistics company, the company is supported by a fleet of 2,100 trucks. The number of fleets will continue to grow in line with the target set by the company, namely in 2020 the number of truck fleets must reach 6,000 trucks. Currently the logistics operational strategy is produced based on stakeholder experience, this certainly causes problems in the company's overall operations. Prediction of the number of daily goods shipments can be a solution in helping companies in planning, monitoring and evaluating logistical operational strategies, based on the company's ability in the availability of a fleet of vehicles for shipping. This study proposes a combination of Support Vector Regression (SVR) methods, genetic algorithms and Multivariate Adaptive Regression Splines (MARS) for problem solving in the prediction process, including in the selection of appropriate training data. From the tests that have been done, the combination of the three methods can produce predictions of the number of daily shipments with values of Mean Absolute Percentage Error (MAPE) 0.0969%, epsilon ( $\epsilon$ ) 1.92172577675873E- 20, complexity ( $c$ ) 62, and gamma ( $\gamma$ ) 1.0.*

**Keywords:** *Prediction, Support Vector Regression, Genetic Algorithm, multivariate adaptive regression splines.*

## 1. PENDAHULUAN

Teknologi komputasi cerdas telah memungkinkan manusia untuk mempermudah

membuat keputusan dalam memecahkan suatu masalah. Pada bidang logistik saat ini, sebuah keputusan yang bersifat strategis tidak terlepas dari pemanfaatan kemajuan teknologi komputasi cerdas.

Saat ini pemanfaatan komputansi cerdas belum dimanfaatkan secara optimal oleh PT XYZ. Sebagai perusahaan logistik dengan jumlah kendaraan 2100 armada truk dan akan terus bertambah seiring dengan target yang dicanangkan perusahaan, dimana pada 2020 jumlah armada truk harus mencapai 6.000 armada truk. Dalam upaya meningkatkan keuntungan perusahaan, dibutuhkan sebuah analisa bisnis yang dapat digunakan untuk menganalisa, memonitoring dan mengevaluasi strategi operasional perusahaan. Salah satu analisa tersebut yaitu memprediksi jumlah order pengiriman barang untuk setiap harinya. Dimana prediksi tersebut dapat memberikan masukan strategi operasional logistik seperti apa yang akan diambil oleh perusahaan.

Saat ini prediksi jumlah pengiriman harian masih berdasarkan pengalaman *stakeholder* sehingga mengakibatkan prediksi tidak sesuai dengan aktual. Target yang ditetapkan oleh perusahaan masih belum terpenuhi oleh target pengiriman saat ini dikarenakan jumlah prediksi yang tidak sesuai dengan aktual. Dengan adanya prediksi jumlah order pengiriman diharapkan dapat membantu memperbaiki strategi operasional logistik perusahaan. Strategi operasional yang ada saat ini adalah menentukan jumlah armada truk yang *ready* untuk dapat memenuhi jumlah order dari pelanggan setiap harinya, menentukan jumlah modal uang jalan untuk operasional logistik dan meminimalisir kerusakan armada truk. Dengan menggunakan teknik dalam data mining, penelitian ini berusaha membantu PT XYZ dalam memprediksi jumlah pengiriman barang sehingga membantu perusahaan dalam menentukan strategi operasional apa yang akan diambil sehingga dapat meningkatkan keuntungan perusahaan.

Berdasarkan data pengiriman barang yang sudah ada sebelumnya, prediksi adalah salah satu poin yang harus diperhitungan di masa yang akan datang. Ketika melakukan prediksi diperlukan penerapan sebuah metode komputasi cerdas yang mempunyai tingkat akurasi yang tinggi sehingga menghasilkan suatu prediksi yang akurat. Terdapat banyak teknik *soft computing* dalam melakukan kegiatan prediksi, antara lain yaitu *Neural Network Backpropagation* (BPN), *Support Vector Regression* (SVR), *Fuzzy Time Series*, regresi linier berganda, dan *Autoregressive Integrated Moving Average* (ARIMA). SVR merupakan metode pengembangan dari *Support Vector Machine* (SVM) untuk kasus regresi. Pada tahun 2015 Desy, Diah dan Abdul Mukid menggunakan metode SVR untuk memprediksi jumlah tamu hotel di kabupaten Demak. Dalam penelitiannya dihasilkan prediksi dengan nilai MAPE yaitu 11,6220%. Semakin kecil nilai MAPE berarti semakin baik nilai akurasi prediksi yang dihasilkan (Anggrainingsih, Aprianto and Sihwi, 2015). Penelitian lain yang menggunakan SVR yaitu dilakukan oleh mimin, rekyan dan budi pada tahun 2018 untuk memprediksi jumlah pengunjung pariwisata di kabupaten Jember. Pada penelitiannya

dihasilkan nilai MAPE yaitu sebesar 6,98% (Raharyani, Putri and Setiawan, 2018).

SVR bekerja sangat baik jika diaplikasikan pada kasus *nonlinier* dengan menambahkan fungsi kernel. Tetapi SVR mempunyai sensitifitas dalam menentukan parameter masukan untuk menghasilkan hasil prediksi yang akurat. Oleh sebab itu diperlukan sebuah algoritma yang digunakan untuk menentukan parameter optimal. Algoritma Genetika (GA) adalah salah satu metode optimasi yang dapat digunakan. Masalah optimasi dalam bidang komputer *science* dapat dipecahkan menggunakan metode tersebut dengan tingkat kesuksesan yang baik (Mu'asyaroh and Mahmudy, 2016). Penelitian yang dilakukan oleh Yuan dalam memprediksi volume penjualan menggunakan SVR dengan GA, menunjukkan bahwa SVR-GA dapat memberikan hasil *error rate* yang lebih baik dibandingkan SVR, *Artificial Neural Network* (ANN), BPN dan *Least Mean Square* (LMS). Nilai MAPE yang dihasilkan dari penggabungan SVR dan GA yaitu 30,20% (Azzahra, Setiawan and Adikara, 2018).

Permasalahan lain yang sering terjadi pada saat melakukan *forecasting*/peramalan yaitu penentuan data yang tepat untuk data latih. Oleh karena itu perlu dilakukan pemilihan variabel/data yang tepat atau sering disebut dengan *feature selection*. Chi-Jie Lu (2014) telah melakukan penggabungan teknik *feature selection Multivariate Adaptive Regression Splines* (MARS) dan SVR untuk memprediksi penjualan peralatan komputer di Taiwan. Dalam penelitiannya dihasilkan kesimpulan bahwa penggabungan MARS dan SVR dapat menghasilkan nilai prediksi dengan tingkat akurasi yang lebih baik apabila dibandingkan dengan ARIMA, SVR, MARS dan GA+SVR. Tingkat akurasi yang dihasilkan dari MARS dan SVR yaitu 15,22 % (Lu, 2014).

Atas dasar penelitian-penelitian tersebut maka pada studi ini dilakukan penggabungan tiga metode yaitu metode MARS, SVR dan algoritma genetika. Metode MARS digunakan untuk mengatasi permasalahan pemilihan data latih yang tepat. SVR digunakan sebagai metode untuk melakukan prediksi jumlah pengiriman barang harian serta algoritma genetika digunakan untuk mengatasi permasalahan sensitifitas dalam menentukan parameter masukan pada SVR.

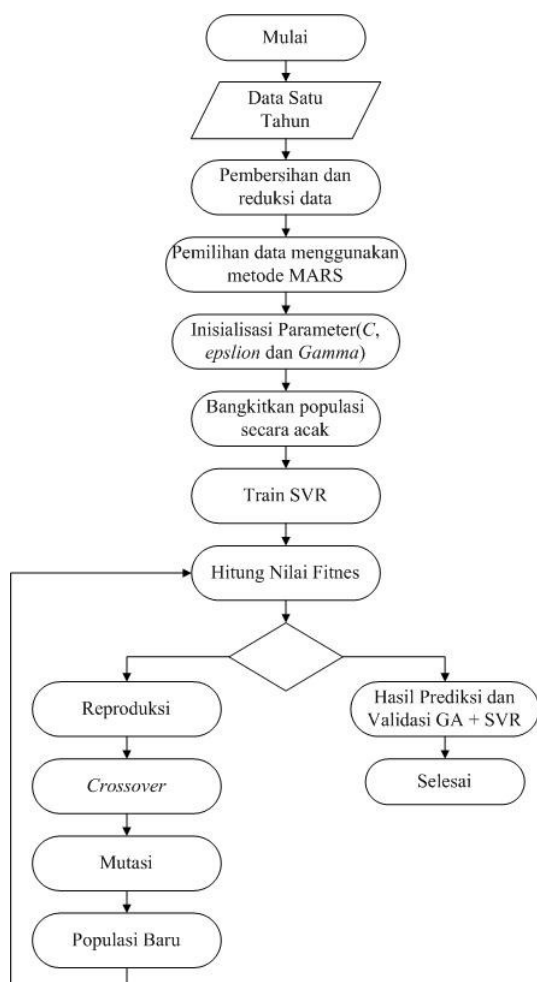
## 2. METODE PENELITIAN

Pada penelitian ini menggunakan metode penelitian deskriptif. Penelitian deskriptif mempunyai sifat fokus pada penyelesaian masalah yang ada pada masa sekarang. Metode deskriptif bertujuan untuk memecahkan permasalahan yang ada pada sebuah organisasi.

Untuk pengumpulan data, penelitian ini melakukan metode studi pustaka, wawancara dan observasi. Observasi dilakukan dengan cara mengamati langsung dan mengambil data yang

dibutuhkan untuk mendukung penelitian di tempat objek yang diteliti yaitu PT XYZ. Wawancara dilakukan dengan manajer IT, manajer marketing, general manajer operasional, manajer finance dan direktur utama PT. Seino Indomobil Logistics. Tahap studi pustaka berupa pengumpulan sumber referensi, seperti jurnal, buku dan internet yang terkait dengan topik data mining.

Teknik analisis kualitatif dipilih untuk digunakan pada penelitian ini. Teknik perancangan pada penelitian ini yaitu teknik perancangan CRISP-DM (*Cross Industry Standart Process for Data Mining*) dimana teknik ini telah banyak digunakan dalam penelitian-penelitian data mining. Teknik Pengujian akan mengfokuskan hasil dari pengujian proses data mining yang telah dimodelkan. Pengujian akurasi dari proses prediksi akan diuji menggunakan *Mean Absolute Percentage Error* (MAPE). Berikut adalah metode yang diusulkan pada penelitian ini.



Gambar 1. Metode usulan

### 3. KERANGKA TEORI

#### 3.1 Prediksi

Studi terdahulu telah melakukan analisis prediksi. Prediksi merupakan bentuk peramalan

tentang apa yang akan terjadi di masa mendatang dalam jangka pendek (Surtiningsih, Furqon and Adinugroho, 2018). Pada umumnya dua jenis prediksi terdiri dari prediksi kualitatif dan prediksi kuantitatif. Prediksi kualitatif adalah prediksi berdasarkan dari pengalaman empiris, intuisi pengambilan keputusan dan emosi manusia yang bersifat subjektif. Prediksi kuantitatif adalah jenis prediksi yang bersifat objektif, menggunakan jenis data aktual yang diolah menggunakan metode tertentu.

#### 3.2 Support Vector Regression

Support Vector Regression (SVR) merupakan model Support Vector Machine (SVM), model ini diperkenalkan oleh Vapnik (1995) pada kasus regresi. Konsep SVM didasari oleh risk minimization (meminimalisasi resiko), atau lebih tepatnya yaitu mengestimasi suatu fungsi dengan usaha meminimalkan batas atas dari generalization error sehingga SVR dapat mengatasi masalah overfitting. Tujuan dari SVR ini adalah untuk memetakan vektor input menggunakan fungsi kernel ke dimensi yang lebih tinggi. Dalam pemecahan kasus regresi, algoritma sekuensial untuk SVR mampu memberikan solusi dengan waktu komputasi yang lebih cepat dan lebih optimal apabila metode SVR konvensional sebagai pembandingnya (Azzahra, Setiawan and Adikara, 2018).

##### Tahap Pertama

Tentukan parameter-parameter SVR yang akan digunakan yaitu  $\gamma$  ( $\gamma$ ), kompleksitas ( $C$ ),  $\epsilon$  ( $\epsilon$ ), serta jumlah iterasi maksimum. Selain itu, inisialisasi  $\alpha_i$  dan  $\alpha_j^*$  (untuk inisialisasi awal, menggunakan nilai 0).

##### Tahap Kedua

Menciptakan matriks *hessian*.

$$R_{ij} = (x, x_i) + \lambda^2 \text{ untuk } i, j = 1, 2, \dots, l \quad (1)$$

$R_{ij}$  = matriks *hessian* baris ke- $i$  dan kolom ke- $j$

$(x, x_i)$  = fungsi kernel

$l$  = jumlah data

$\lambda$  = variabel skalar

##### Tahap Ketiga

Melakukan proses *sequential learning* :

1. Menghitung nilai *error* ( $E$ ).

$$E_i = y_i - \sum_{j=1}^l (\alpha_j^* - \alpha_j) R_{ij} \quad (2)$$

$E_i$  = nilai *error* ke- $i$

$y_i$  = nilai aktual

$\alpha_j^*$ ,  $\alpha_j$  = *lagrange multipliers*

$R_{ij}$  = matriks *hessian* baris ke- $i$  kolom ke- $j$

2. Menghitung  $\delta\alpha_i^*$  dan  $\delta\alpha_i$ .

$$\delta\alpha_i^* = \min \{ \max [(E_i - \epsilon), -\alpha_i^*], C - \alpha_i^* \} \quad (3)$$

$$\delta\alpha_i = \min \{ \max [(-E_i - \varepsilon), -\alpha_i], C - \alpha_i \}$$

$\gamma$  = nilai *learning rate*

$E_i$  = nilai *error* ke- $i$

$\varepsilon$  = nilai untuk kerugian

$C$  = nilai untuk kompleksitas

$\alpha_i^*$ , = *lagrange multipliers*

3. Hitung *lagrange multipliers*.

$$\alpha_i^* = \alpha_i^* + \delta\alpha_i^* \quad (4)$$

$$\alpha_i = \alpha_i + \delta\alpha_i$$

$\alpha_i^*$ ,  $\alpha_i$  = *lagrange multipliers*

#### Tahap Keempat

Dilakukan proses iterasi *sequential learning* sampai iterasi mencapai jumlah maksimum atau kondisi berhenti.

$$\max(|\delta\alpha_i^*|) < \varepsilon \text{ dan } \max(|\delta\alpha_i|) < \varepsilon$$

#### Tahap Kelima

Apabila data sudah memenuhi persyaratan ( $\alpha_i^* - \alpha_i$ ) tidak sama dengan 0, maka dapat disebut sebagai *support vector*.

#### Tahap Keenam

Melakukan pengujian.

$$(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i)(K(x, x_i) + \lambda^2) \quad (5)$$

$\alpha_i^*$ ,  $\alpha_i$  = *lagrange multipliers*

$(x, x_i)$  = fungsi *kernel*

$\lambda$  = variabel skalar

$l$  = banyak data

#### Tahap Ketujuh

Menghitung *error rate* menggunakan *Mean Absolute Percentage Error* (MAPE).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \times 100\% \quad (6)$$

$n$  = Jumlah pengiriman barang

$y_i$  = Nilai data latih

$y'_i$  = Nilai hasil prediksi

$i$  = Indeks

### 3.3 Algoritma Genetika

Algoritma Genetika merupakan suatu metode optimasi untuk mencari solusi yang optimal dari suatu permasalahan. Ada beberapa tahap pada algoritma ini, antara lain representasi kromosom, *crossover*, mutasi dan seleksi. Tahap pertama yang dilakukan algoritma genetika ini adalah pembuatan solusi alternatif (populasi). Setiap solusi mewakili sebagai individual atau kromosom. Pada tahun 1960 dan 1970 algoritma ini diciptakan oleh John Holland. Algoritma genetika menggunakan analogi alami berdasarkan seleksi alam (Noersasongko *et al.*, 2016).

### 3.4 Multivariate Adaptive Regression Splines

*Multivariate Adaptive Regression Splines* (MARS) adalah sebuah model yang diperkenalkan oleh Jerome H. Friedman pada tahun 1991 sebagai metode untuk mengotomatisasi model prediktif yang

akurat. Metode ini mampu memecahkan permasalahan diskontinuitas dan data berdimensi tinggi, yang memiliki variabel  $3 \leq n \leq 20$ . MARS mampu memecahkan prediksi variabel respon dengan akurat, serta menghasilkan model yang kontinu dalam knot berdasarkan nilai GCV terkecil (Friedman, 1991). MARS menggunakan pendekatan model regresi multivariate nonparametrik antara variabel respon dan beberapa variabel prediktor pada *piecewise* regresi (Sita and Otok, 2014).

*Piecewise* regresi adalah regresi yang memiliki sifat tersegmen. Jika keseluruhan data tidak bisa dijelaskan oleh suatu garis regresi dan beberapa garis regresi akan digunakan menjelaskan perubahan seluruh data dari variabel independen. Perubahan pola tersebut disebut dengan knot. Pada tahap pengujian dilakukan dengan cara melakukan kontrol terhadap jumlah derajat bebas (DF) untuk optimalisasi knot. Untuk menjelaskan hubungan antara variabel respon dan variabel prediktor diperlukan Basis fungsi (Tehupuring, Soeharsono and Hadi, 2015)

Berikut adalah model MARS :

$$f(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \prod_{k=1}^{k_m} (S_{km}(X_{v(k,m)} - t_{km}))^+, \quad (7)$$

dimana :

$\alpha_0$  = konstanta regresi pada fungsi basis

$\alpha_m$  = koefisien dari fungsi basis ke- $m$ ,  $m = 1, \dots, M$

$M$  = maksimum fungsi basis

$k_m$  = derajat interaksi

$S_{km} = \begin{cases} +1, & \text{jika knot terletak dikanan subregion} \\ -1, & \text{jika knot terletak dikiri subregion} \end{cases}$

$X_{v(k,m)}$  = variabel prediktor

$t_{km}$  = nilai *knots* dari variabel prediktor  $X_v(k, m)$ .

Dengan nilai GCV sebagai berikut.

$$GCV(M) = \frac{1/N \sum_{i=1}^N [y_i - f_M(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \quad (8)$$

Dimana :

$M$  = jumlah fungsi basis

$x_i$  = variabel prediktor

$y_i$  = variabel respon

$N$  = banyaknya pengamatan

## 4. HASIL DAN PEMBAHASAN

### 4.1. Data yang digunakan

Data yang akan digunakan untuk prediksi pengiriman harian adalah data pengiriman barang satu tahun terakhir. Data ini diperoleh dari sistem logistik yang digunakan oleh PT XYZ. Pada penelitian ini jumlah data yang digunakan yaitu 233 data pengiriman barang harian. Kemudian ada dua atribut yang digunakan, yang pertama adalah data yang akan digunakan untuk pemilihan atribut/data menggunakan metode MARS dan yang kedua adalah data yang akan digunakan untuk memprediksi jumlah pengiriman barang harian menggunakan metode SVR dan algoritma genetika.

Terdapat tiga atribut yang digunakan untuk pemilihan atribut/data menggunakan metode MARS, atribut tersebut yaitu tanggal, jumlah pengiriman dan hari. Atribut hari dipilih sebagai variabel respon dikarenakan hari diduga memiliki pengaruh terhadap jumlah pengiriman barang. Kemudian atribut yang digunakan untuk prediksi jumlah pengiriman barang menggunakan SVR dan algoritma genetika yaitu tanggal dan jumlah pengiriman harian.

#### **4.2. Pemilihan atribut/data menggunakan metode MARS**

Pada tahap pemilihan data latih menggunakan metode MARS dilakukan tiga kali percobaan yaitu dengan data per dua bulan, per tiga bulan dan per empat bulan. Alasan pemilihan data menjadi tiga bagian karena PT XYZ melakukan evaluasi bisnis setiap enam bulan, empat bulan dan tiga bulan. Hal ini juga sesuai dengan tujuan penelitian untuk mengetahui prediksi pengiriman barang pada tiga periode waktu tersebut.

##### **4.2.1. Pemilihan atribut/data per dua bulan**

Percobaan pertama yaitu pemilihan atribut/data per dua bulan terdapat enam variabel prediktor dan satu variabel respon. Variabel prediktor yaitu data pengiriman data satu tahun yang dibagi kedalam enam atribut, dimana satu atribut mewakili jumlah pengiriman barang harian selama dua bulan. Data-data tersebut yaitu data periode November 2017 sampai dengan Desember 2017, Januari 2018 sampai dengan Februari 2018 dan seterusnya. Sehingga mendapatkan enam atribut yang akan dijadikan variabel prediktor. Variabel-variabel tersebut yaitu bulanPertama, bulanKedua, bulanKetiga, bulanKeempat, bulanKelima dan bulanKeenam. Kemudian variabel respon yang digunakan yaitu nama hari, pemilihan hari sebagai variabel respon dikarenakan hari diduga mempengaruhi jumlah pengiriman.

Dari hasil pemilihan atribut/data menggunakan data per dua bulan menggunakan metode MARS, dapat disimpulkan bahwa variabel prediktor bulanPertama mempunyai tingkat kepentingan tertinggi dengan tingkat kepentingan 100% dan nilai GCV yaitu 1.48553. Nilai GCV terkecil dihasilkan dari kombinasi nilai BF 18, MI 1 dan MO 2. Ini berarti apabila kita menggunakan pemodelan MARS untuk pemilihan atribut/data menggunakan enam atribut per dua bulan, data pengiriman barang periode 1 November 2017 sampai dengan 31 Desember 2017 mempunyai tingkat kepentingan terbaik bagi perusahaan.

##### **4.2.2. Pemilihan atribut/data per tiga bulan**

Percobaan kedua yaitu pemilihan data per tiga bulan terdapat empat variabel prediktor dan satu variabel respon. variabel prediktor tersebut yaitu data pengiriman data satu tahun yang dibagi kedalam empat atribut, dimana satu atribut mewakili jumlah

pengiriman barang harian selama tiga bulan. Data-data tersebut yaitu data periode November 2017 sampai dengan Januari 2018, Februari 2018 sampai dengan April 2018, Mei 2018 sampai dengan Juli 2018 dan Agustus 2018 sampai dengan Oktober 2018. Sehingga mendapatkan empat atribut yang akan dijadikan variabel prediktor. Variabel-variabel tersebut yaitu bulanPertama, bulanKedua, bulanKetiga dan bulanKeempat. Kemudian variabel respon yang digunakan yaitu nama hari.

Dari hasil pemilihan atribut/data menggunakan data per tiga bulan menggunakan metode MARS, dapat disimpulkan bahwa variabel prediktor bulanPertama mempunyai tingkat kepentingan tertinggi dengan tingkat kepentingan 100% dengan nilai GCV terkecil yaitu 1.67358. Nilai GCV terkecil dihasilkan dari kombinasi nilai BF 12, MI 1 dan MO 2. Ini berarti apabila kita menggunakan pemodelan MARS untuk pemilihan atribut/data dengan MARS menggunakan empat atribut per tiga bulan, data pengiriman barang periode 1 November 2017 sampai dengan 31 Januari 2018 mempunyai tingkat kepentingan terbaik.

##### **4.2.3. Pemilihan atribut/data per empat bulan**

Percobaan ketiga yaitu pemilihan atribut/data per empat bulan menggunakan metode MARS. Pada pemilihan atribut/data per empat bulan terdapat tiga variabel prediktor dan satu variabel respon. Variabel prediktor tersebut yaitu data pengiriman data satu tahun yang dibagi kedalam tiga atribut, dimana satu atribut mewakili jumlah pengiriman barang harian selama empat bulan. Data-data tersebut yaitu data periode November 2017 sampai dengan Februari 2018, Maret 2018 sampai dengan Juni 2018 dan Juli 2018 sampai dengan Oktober 2018. Sehingga mendapatkan tiga atribut yang akan dijadikan variabel prediktor. Variabel-variabel tersebut yaitu bulanPertama, bulanKedua dan bulanKetiga. Kemudian variabel respon yang digunakan yaitu nama hari.

Dari hasil pemilihan atribut/data menggunakan data per empat bulan menggunakan metode MARS, dapat disimpulkan bahwa variabel prediktor bulanPertama mempunyai tingkat kepentingan tertinggi dengan tingkat kepentingan 100% dengan nilai GCV terkecil yaitu 1.58904. Nilai GCV terkecil dihasilkan dari kombinasi nilai BF 6, MI 1 dan MO 0. Ini berarti apabila kita menggunakan pemodelan MARS untuk pemilihan atribut/data menggunakan tiga atribut per empat bulan, data pengiriman barang periode 1 November 2018 sampai dengan 28 Februari 2018 mempunyai tingkat kepentingan yang paling baik.

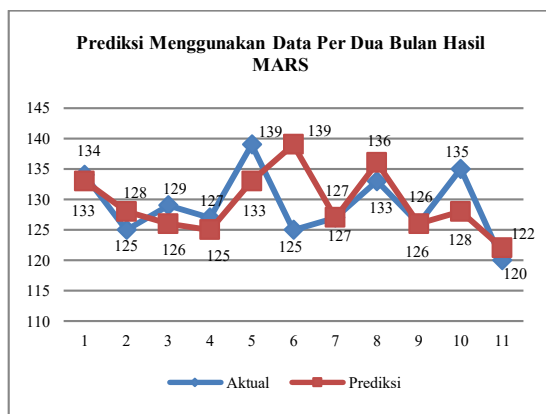
#### **4.3. Kombinasi Metode SVR dan Algoritma Genetika Dengan Data Per Dua Bulan**

Percobaan yang pertama yaitu prediksi pengiriman barang harian menggunakan SVR kernel linier yang akan dioptimasi nilai parameternya

dengan algoritma genetika. Parameter-parameter tersebut yaitu *complexitas* ( $C$ ), *epsilon* ( $\epsilon$ ) dan *gamma* ( $\gamma$ ). Pada percobaan ini data pelatihan yang digunakan yaitu 80 persen dan data pengujian 20 persen menggunakan data hasil pemilihan data per dua bulan sesuai dengan hasil pemilihan data menggunakan metode MARS. Data tersebut yaitu periode 1 November 2017 sampai dengan 31 Desember 2017.

Range nilai parameter yang akan dibangkitkan pada proses genetika untuk parameter *complexitas* adalah 1 sampai dengan 100, *epsilon*  $10^{-11}$  sampai dengan  $10^{-1}$  dan *gamma*  $10^{-11}$  sampai dengan  $10^{-1}$  dan jumlah iterasi 100. Dari nilai range yang telah ditentukan kemudian akan dibangkitkan 100 populasi awal dari kombinasi ketiga parameter dengan nilai parameter  $C$ ,  $\epsilon$  dan *gamma* secara *random*, selanjutnya akan dilakukan proses training SVR.

Dari hasil percobaan prediksi pengiriman barang harian menggunakan atribut/data hasil pemilihan per dua bulan yaitu periode 1 November 2017 sampai dengan 31 Desember 2017 menggunakan metode SVR kernel linier dan algoritma genetika dengan pengoptimalan parameter  $C$ , *epsilon* dan *gamma*. Nilai MAPE yang dihasilkan yaitu 0.0969% dengan parameter terbaik yaitu  $C = 62$ , *epsilon* =  $1.92172577675873E-20$  dan *gamma* = 1. Berikut adalah grafik hasil prediksi menggunakan data pengujian. Sumbu x mempresentasikan nilai prediksi jumlah pengiriman dan sumbu y mempresentasikan urutan hari kerja.



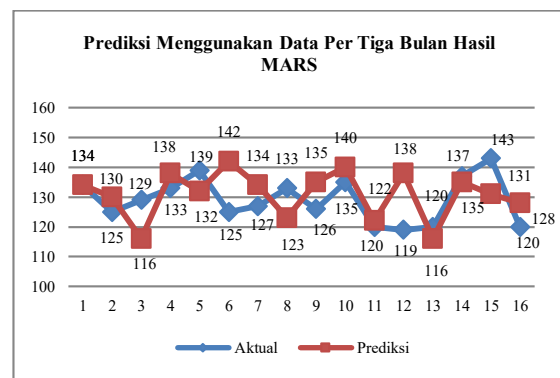
Gambar 2. Prediksi menggunakan data hasil pemilihan atribut/data per dua bulan

#### 4.4. Kombinasi Metode SVR dan Algoritma Genetika Dengan Data Per Tiga Bulan

Percobaan yang kedua yaitu prediksi pengiriman barang harian menggunakan SVR kernel linier yang akan dioptimasi nilai parameternya dengan algoritma genetika. Parameter-parameter tersebut yaitu *complexitas* ( $C$ ), *epsilon* ( $\epsilon$ ) dan *gamma* ( $\gamma$ ). Pada percobaan ini data pelatihan yang digunakan yaitu 80 persen dan data pengujian 20 persen menggunakan data hasil pemilihan data per tiga bulan sesuai dengan hasil pemilihan data menggunakan metode MARS. Data tersebut yaitu periode 1 November 2017 sampai dengan 31 Januari 2018.

Range nilai parameter yang akan dibangkitkan pada proses genetika untuk parameter *complexitas* adalah 1 sampai dengan 100, *epsilon*  $10^{-11}$  sampai dengan  $10^{-1}$ , *gamma*  $10^{-11}$  sampai dengan  $10^{-1}$  dan jumlah iterasi 100. Dari nilai range yang telah ditentukan kemudian akan dibangkitkan 100 populasi awal dari kombinasi ketiga parameter dengan nilai parameter  $C$ ,  $\epsilon$  dan *gamma* secara *random*, selanjutnya akan dilakukan proses training SVR.

Dari hasil percobaan prediksi pengiriman barang harian menggunakan atribut/data hasil pemilihan per tiga bulan yaitu periode 1 November 2017 sampai dengan 31 Januari 2018. Penerapan metode SVR linier dengan pengoptimalan parameter  $C$ , *epsilon* dan *gamma* menggunakan algoritma genetika menghasilkan nilai MAPE sebesar 1.6375% dengan parameter terbaik yaitu  $C = 65.5$ , *epsilon* =  $1.0503782010872E-20$  dan *gamma* = 1. Berikut adalah grafik hasil prediksi menggunakan data pengujian. Sumbu x mempresentasikan nilai prediksi jumlah pengiriman dan sumbu y mempresentasikan urutan hari kerja.



Gambar 3. Prediksi menggunakan data hasil pemilihan atribut/data per tiga bulan

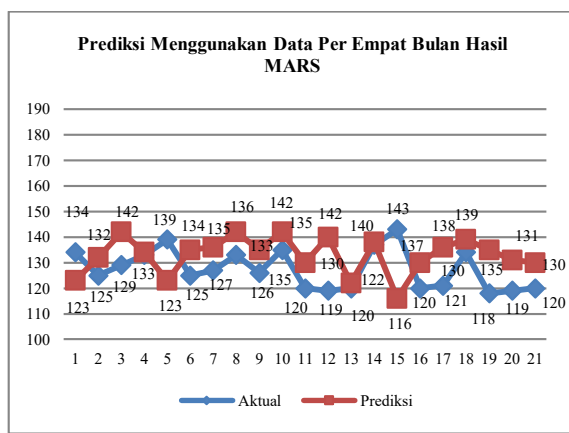
#### 4.5. Kombinasi Metode SVR dan Algoritma Genetika Dengan Data Per Empat Bulan

Percobaan yang ketiga yaitu prediksi pengiriman barang harian menggunakan SVR kernel linier yang akan dioptimasi nilai parameternya dengan algoritma genetika. Parameter-parameter tersebut yaitu *complexitas* ( $C$ ), *epsilon* ( $\epsilon$ ) dan *gamma* ( $\gamma$ ). Pada percobaan ini data pelatihan yang digunakan yaitu 80 persen dan data pengujian 20 persen menggunakan data hasil pemilihan data per tiga bulan sesuai dengan hasil pemilihan data menggunakan metode MARS. Data tersebut yaitu periode 1 November 2017 sampai dengan 28 Februari 2018.

Range nilai parameter yang akan dibangkitkan pada proses genetika untuk parameter *complexitas* adalah 1 sampai dengan 100, *epsilon*  $10^{-11}$  sampai dengan  $10^{-1}$ , *gamma*  $10^{-11}$  sampai dengan  $10^{-1}$  dan jumlah iterasi 100. Dari nilai range yang telah ditentukan kemudian akan dibangkitkan 100 populasi awal dari kombinasi ketiga parameter dengan nilai

parameter  $C$ ,  $\epsilon$  dan  $\gamma$  secara *random*, selanjutnya akan dilakukan proses *training* SVR.

Dari hasil percobaan prediksi pengiriman barang harian menggunakan atribut/data hasil pemilihan per empat bulan yaitu periode 1 November 2017 sampai dengan 28 Februari 2018. Penerapan metode SVR kernel linier dengan pengoptimasian parameter  $C$ ,  $\epsilon$  dan  $\gamma$  menggunakan algoritma genetika menghasilkan nilai MAPE sebesar 4.6510% dengan parameter terbaik yaitu  $C = 67.5$ ,  $\epsilon = 7.5449870527294E-21$  dan  $\gamma = 0.992753623$ . Berikut adalah grafik hasil prediksi menggunakan data pengujian. Sumbu x mempresentasikan nilai prediksi jumlah pengiriman dan sumbu y mempresentasikan urutan hari kerja.



Gambar 4. Prediksi menggunakan data hasil pemilihan atribut/data per empat bulan

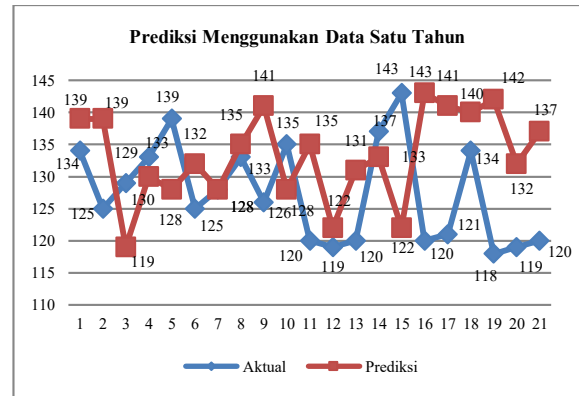
#### 4.6. Kombinasi Metode SVR dan Algoritma Genetika Dengan Data 12 bulan

Percobaan yang terakhir yaitu prediksi pengiriman barang harian menggunakan SVR kernel linier yang akan dioptimasi nilai parameternya dengan algoritma genetika. Parameter-parameter tersebut yaitu *complexitas* ( $C$ ), *epsilon* ( $\epsilon$ ) dan *gamma* ( $\gamma$ ). Pada penelitian ini data pelatihan yang digunakan yaitu 80 persen dan data pengujian 20 persen tanpa menggunakan metode untuk pemilihan atribut/data. Data tersebut yaitu periode 1 November 2017 sampai dengan 30 Oktober 2018.

*Range* nilai parameter yang akan dibangkitkan pada proses genetika untuk parameter *complexitas* adalah 1 sampai dengan 100, *epsilon*  $10^{-11}$  sampai dengan  $10^{-1}$ , *gamma*  $10^{-11}$  sampai dengan  $10^{-1}$  dan jumlah iterasi 100. Dari nilai *range* yang telah ditentukan kemudian akan dibangkitkan 100 populasi awal dari kombinasi ketiga parameter dengan nilai parameter  $C$ ,  $\epsilon$  dan  $\gamma$  secara *random*, selanjutnya akan dilakukan proses *training* SVR.

Dari hasil percobaan prediksi pengiriman barang harian menggunakan data satu tahun terakhir yaitu periode 1 November 2017 sampai dengan 30 Oktober 2018. Menggunakan SVR kernel linier dengan pengoptimasian parameter  $C$ ,  $\epsilon$  dan  $\gamma$  menggunakan algoritma genetika

menghasilkan nilai prediksi sebesar 134 pengiriman barang dan MAPE sebesar 4.9431% dengan parameter terbaik yaitu  $C = 61.5$ ,  $\epsilon = 2.100746876434E-20$  dan  $\gamma = 0.9921875$ . Berikut adalah grafik hasil prediksi menggunakan data pengujian. Sumbu x mempresentasikan nilai prediksi jumlah pengiriman dan sumbu y mempresentasikan urutan hari kerja.



Gambar 5. Prediksi menggunakan data satu tahun tanpa pemilihan atribut/data

#### 4.7. Evaluasi Model

Dari hasil empat kali percobaan prediksi jumlah pengiriman barang harian menggunakan data hasil pemilihan atribut/data menggunakan MARS per dua bulan, per tiga bulan, per empat bulan dan tanpa menggunakan pemilihan atribut/data.

Penambahan pemilihan atribut/data menggunakan metode MARS pada metode SVR dan algoritma genetika dapat meningkatkan akurasi hasil prediksi jumlah pengiriman harian. Hasil evaluasi metode terlihat pada Tabel 1.

Tabel 1. Evaluasi metode

SVR-GA-MARS	Model Pemilihan MARS	Data	MAPE
	MARS Per Dua Bulan	1 November 2017 s/d 31 Desember 2017	0.0969%
	MARS Per Tiga Bulan	1 November 2017 s/d 31 Januari 2018	1.6375%
	MARS Per Empat Bulan	1 November 2017 s/d 28 Februari 2018	4.6511%
SVR-GA	1 November 2017 - 31 Oktober 2018	4.9431%	SVR-GA

Pada Tabel 1 terlihat bahwa nilai MAPE yang dihasilkan apabila hanya menggunakan SVR dan algoritma genetika adalah sebesar 4.9431%. Perbaikan nilai akurasi MAPE diperbaiki dengan

menambahkan metode pemilihan atribut/data menggunakan MARS. Hal tersebut dapat dilihat bahwa apabila prediksi menggunakan data hasil pemilihan data per empat bulan maka nilai MAPE yang dihasilkan yaitu 4.6511%, untuk data hasil per tiga bulan yaitu 1.6375% dan untuk data hasil per dua bulan yaitu 0.0969%.

Untuk data per dua bulan nilai MAPE yang dihasilkan menjadi yang terbaik dikarenakan periode data antara data per dua bulan memiliki deviasi terkecil dibandingkan dengan periode data per empat bulan atau per tiga bulan. Dengan demikian nilai MAPE terbaik yang dihasilkan dari penggabungan metode MARS, SVR dan algoritma genetika adalah pada pemilihan data per dua bulan.

## 5. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, penggabungan metode *Multivariate Adaptive Regression Splines* untuk pemilihan atribut/data, *Support Vector Regression* untuk prediksi jumlah pengiriman barang harian dan algoritma genetika untuk optimasi parameter *Support Vector Regression* mampu menghasilkan prediksi pengiriman barang harian dengan tingkat akurasi yang tinggi. Hal ini dibuktikan dari nilai MAPE yang dihasilkan dari penggabungan ketiga metode tersebut yaitu sebesar 0.0969%. Nilai MAPE terbaik dihasilkan menggunakan data pengiriman barang periode 1 November 2017 s/d 31 Desember 2017 sesuai dengan pemilihan atribut/data per dua bulan menggunakan metode MARS. Dengan kombinasi parameter terbaik yaitu  $C = 62$ ,  $\epsilon = 1.92172577675873E-20$  dan  $\gamma = 1$ .

## DAFTAR PUSTAKA

- ANGGRAININGSIH, R., APRIANTO, G. R. AND SIHWI, S. W. (2015) 'Time Series Forecasting Using Exponential Smoothing To Predict The Number of Website Visitor of Sebelas Maret University', in *Int. Conference on Information Technology, Computer and Electrical Engineering*, pp. 14–19.
- AZZAHRA, M., SETIAWAN, B. D. AND ADIKARA, P. P. (2018) 'Optimasi Parameter Support Vector Regression Dengan Algoritma Genetika Untuk Prediksi Harga Emas', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), pp. 273–281.
- LU, C. (2014) 'Sales Forecasting of Computer Products Based on Variable Selection Scheme and Support Vector Regression', *Neurocomputing*, 128, pp. 491–499. doi: 10.1016/j.neucom.2013.08.012.
- MU'ASYAROH, F. L. AND MAHMUDY, W. F. (2016) 'Implementasi Algoritma Genetika Dalam Optimasi Model AHP', *Jurnal Teknologi Informasi dan Ilmu Komputer*, 3(4), pp. 226–237.
- NOERSASONGKO, E. ET AL. (2016) 'A Tourism Arrival Forecasting Using Genetic Algorithm based Neural Network', *Indian Journal of Science and Technology*, 9 (4), pp. 1–5.
- RAHARYANI, M. P., PUTRI, R. R. M. AND SETIAWAN, B. D. (2018) 'Implementasi Algoritma Support Vector Regression Pada Prediksi Jumlah Pengunjung Pariwisata', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(4), pp. 1501–1509.
- SITA, E. D. A. A. AND OTOK, B. W. (2014) 'Pendekatan Multivariate Adaptive Regression SPLINES (MARS) pada Pemodelan Penduduk Miskin di Indonesia Tahun 2008-2012', in *Prosiding Seminar Nasional Matematika Universitas Jember*, pp. 175–191.
- SURTININGSIH, L., FURQON, M. T. AND ADINUGROHO, S. (2018) 'Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(8), pp. 2578–2586.
- TEHUPURING, B. C., SOEHARSONO, S. AND HADI, S. (2015) 'Penerapan Multivariate Adaptive Regression Spline sebagai Alat untuk Pemodelan Pertumbuhan Ayam Broiler', *ACTA Veterinaria Indonesia*, 3(1), pp. 23–28.