

**IMPLEMENTASI ALGORITME *SUPPORT VECTOR REGRESSION*
DAN PARTICLE SWARM OPTIMIZATION
DALAM PERAMALAN PENJUALAN**

TESIS

Disusun sebagai salah satu syarat untuk
memperoleh gelar Magister Komputer
dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh :

RAHMAWATI

NPM : 2018210038



**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

**IMPLEMENTASI ALGORITME *SUPPORT VECTOR REGRESSION*
DAN PARTICLE SWARM OPTIMIZATION
DALAM PERAMALAN PENJUALAN**

Oleh :

RAHMAWATI

NPM : 2018210038

Bandung, 27 Januari 2020

Menyetujui,

Dr. Hery Heryanto, S.Kom., M.Kom
Pembimbing

**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

“Hai orang-orang yang beriman, mintalah pertolongan (kepada Allah) dengan sabar dan (mengerjakan) shalat, sesungguhnya Allah beserta orang-orang yang sabar.”
(Al-Baqarah:153)

Dipersembahkan untuk:

*Ibunda dan Ayahanda Tercinta,
Saudara kandung teh Iyam, teh Ipa, dan Alvy*

ABSTRAK

IMPLEMENTASI ALGORITME *SUPPORT VECTOR REGRESSION* DAN *PARTICLE SWARM OPTIMIZATION* DALAM PERAMALAN PENJUALAN

Oleh :

Nama : Rahmawati

NPM : 2018210038

Peramalan penjualan merupakan salah satu elemen penting dalam perencanaan strategi bisnis yang memiliki kemungkinan terjadi di masa depan. Berdasarkan data dari *statista.com* angka penjualan di bisnis *fashion* akan terus mengalami kenaikan secara signifikan. Hal ini menjadi sebuah tantangan bagi *brand* ZXC yang merupakan perusahaan retail di bidang *fashion* untuk bisa mengembangkan strategi bisnis perusahaan, salah satunya dengan teknik peramalan. Tetapi bagaimana meningkatkan ketepatan proses peramalan saat ini masih menjadi tanda tanya besar bagi perusahaan, karena peramalan sering kali rentan mengalami kegagalan atau jauh dengan data aktual karena dipengaruhi penjualan yang sangat fluktuatif. Banyak metode yang dapat digunakan untuk membuat peramalan, dimulai dari metode statistik tradisional sampai metode atau algoritme untuk *data mining* modern seperti *machine learning* yang terkomputasi. Namun metode statistik tradisional seperti *exponential smoothing*, ARIMA dan *regression* memiliki akurasi lebih rendah dibandingkan dengan metode pembelajaran mesin seperti *support vector regression* (SVR), *long short-term memory* (LSTM) dan *artificial neural network* (ANN). Oleh karena itu, pembelajaran mesin sangat berpotensi dalam mencapai kualitas peramalan yang baik.

Penelitian ini menggunakan data penjualan *brand* ZXC dan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai tahapan pemecahan masalah untuk *data mining*. Untuk teknik peramalan, algoritme *Support Vector Regression* (SVR) dipilih karena merupakan salah satu teknik peramalan yang dikategorikan baik dibandingkan dengan teknik yang lainnya. Ini telah dibuktikan pada penelitian-penelitian sebelumnya. Untuk mencapai kualitas peramalan yang baik, algoritme *Particle Swarm Optimization* (PSO) digunakan yang akan diintegrasikan untuk optimasi atribut agar akurasi peramalan dapat lebih baik.

Berdasarkan hasil pengujian terhadap data penjualan dengan menggunakan SVR-PSO, didapatkan SVR-PSO menghasilkan *Root Mean Squared Error* (RMSE) atau tingkat kesalahan sebesar 9,40. Nilai tersebut menjadi lebih baik dibandingkan dengan hanya menggunakan SVR saja yang mempunyai nilai RMSE sebesar 35,83. Nilai RMSE tersebut dapat dikatakan baik karena tingkat kesalahan yang dihasilkan semakin mengecil dan mendekati angka 0. Maka dapat dikatakan kombinasi algoritme SVR-PSO memiliki hasil optimal untuk peramalan.

Kata Kunci : *peramalan penjualan, support vector regression, particle swarm optimization*

ABSTRACT

IMPLEMENTATION OF SUPPORT VECTOR REGRESSION ALGORITHM AND PARTICLE SWARM OPTIMIZATION IN SALES FORECASTING

Author :

Name : Rahmawati

NPM : 2018210038

The sales forecasting is important to provides a guide to long-term strategic planning in a business in possibility detections in the future. In addition, the sales is on eof element in strategic planning of business the has possibility in the future as well. Based in the statista.com the number of sales in fashion business will increase significantly. This case is a challenge for brand of ZXC as a retail in fashion sectors to develop the strategy of business by forecasting. Unfortunately, the main problem for this company regarding the strategy to increase the accuracy of forecasting, because the forecasting is often prone to calculating failure from actual data that impacted by fluctuative sales. There several methods to creates the forecasting, begin from traditional statistical methods or algorithms for modern data mining such as computational machine learning. Nevertheless, the traditional statistical methods such as *exponential smoothing*, ARIMA and *regression* has lower accuracy than machine learning methods like *support vector regression* (SVR), *long short-term memory* (LSTM) and *artificial neural network* (ANN). Therefore, the machine learning has an opportunity to reach the best quality of forecasting.

In this research used sales data of ZXC brands and *Cross-Industry Standard Process for Data Mining* (CRISP-DM) to solve the problem of data mining. The *Support Vector Regression* (SVR) is the forecasting technique that catagorized as good than another. This cases indicated in previous research to provides the accuracy forecasting we required algorithm such as *Particle Swarm Optimization* (PSO) to optimalize the data and impact on the quality of forecasting accuracy.

Based on the results of testing of sales data using SVR-PSO, obtained SVR-PSO produces a *Root Mean Squared Error* (RMSE) value of 9,40. the RMSE value is better than using only the SVR which has a RMSE value of 35,83. The RMSE value can be said to be good because the resulting error rate is close to 0. So it can be said the combination of SVR-PSO algorithm has optimal results for forecasting.

Keywords : *Forecasting, Sales Forecasting, Support Vector Regression, Particle Swarm Optimization*

KATA PENGANTAR

Dengan memanjatkan Puji Syukur Kehadirat Allah SWT atas segala Rahmat dan Karunianya pada penulis, akhirnya penulis dapat menyelesaikan Tesis yang berjudul “Implementasi Algoritme *Support Vector Regression* (SVR) dan *Particel Swarm Optimization* (PSO) dalam Peramalan Penjualan”. Penulis lebih khusus mengucapkan terima kasih kepada Bapak Dr. Hery Heryanto, S.Kom., M.Kom atas segala bimbingan dan motivasinya sehingga penulis tetap semangat dalam menyelesaikan segala permasalahan yang dihadapi dan akhirnya tesis ini bisa diselesaikan dengan baik.

Penulis menyadari bahwa tesis dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak, oleh karena itu penulis juga ingin menyampaikan apresiasi yang setinggi-tingginya kepada:

1. Bapak Roby Chandra, S.Kom. selaku Manager perusahaan *brand* ZXC yang telah memberikan kesempatan kepada penulis untuk melaksanakan penelitian di perusahaan *brand* ZXC;
2. Seluruh *staff* perusahaan *brand* ZXC bekerja sama dalam proses pengumpulan data dan wawancara yang tidak dapat disebutkan satu persatu penulis ucapkan terima kasih atas partisipasinya;
3. Seluruh Civitas Akademika STMIK LIKMI yang telah banyak membantu kelancaran aktifitas perkuliahan;
4. Rekan-rekan seperjuangan program Pascasarjana (S2) angkatan 2018, dan kawan-kawan yang tidak bisa disebutkan semuanya, terima kasih motivasi dan kebersamaan selama melaksanakan perkuliahan;
5. Bapak Fikri Fahru Roji selaku mentor diskusi dalam penyusunan Tesis ini, penulis mengucapkan terima kasih.
6. Ayahanda Endang Hayat dan Ibunda Ai Fatimah yang tak mengenal lelah memperjuangkan dan mendidik serta memberikan dukungan baik moril maupun materil dari mulai penulis dilahirkan hingga saat ini, semoga apa yang mereka telah

lakukan menjadi amalan yang diterima di sisi Allah S.W.T. Saudara-saudaraku yang tetap mendukung dalam pengerjaan tesis, penulis ucapkan terima kasih.

Ucapan terima kasih, disampaikan kepada semua pihak yang tidak dapat disebutkan namanya satu-persatu, yang telah banyak memberikan dukungan, semoga Allah S.W.T. memberikan rahmat, hidayah-Nya serta menggantinya dengan yang lebih baik dan sempurna. Amin. "*Manusia merupakan tempat khilaf dan lupa*", demikian juga dengan penelitian ini, masih banyak kekurangan dan kekeliruan, untuk itu kritik serta saran yang sifatnya membangun, sangat diharapkan. Terima kasih.

Bandung, 27 Januari 2020

Rahmawati

DAFTAR ISI

	Halaman
LEMBAR PENGESAHAN	i
LEMBAR PERSEMBAHAN	ii
ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR DAN ILUSTRASI	ix
DAFTAR TABEL	ix
DAFTAR RUMUS	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Ruang Lingkup Penelitian	4
1.5 Sistematika Penelitian	7
BAB II LANDASAN TEORI	8
2.1 Peramalan (<i>forecasting</i>)	8
2.2.1 Metode Peramalan	8
2.2.2 Mengukur Akurasi Peramalan	11
2.2 <i>Data Mining</i>	11
2.3 Metodologi Data Mining	12

2.4	<i>Machine Learning</i>	15
2.5	<i>Support Vector Regression (SVR)</i>	15
2.6	<i>Particle Swarm Optimization (PSO)</i>	21
2.7	Optimasi Parameter Pencarian <i>Grid</i>	24
2.8	<i>K-Fold Cross-Validation</i>	25
2.9	Penjualan	26
2.10	Penelitian Terkait	28
BAB III OBJEK DAN METODOLOGI PENELITIAN		31
3.1	Penjualan	31
3.2	Metodologi Penelitian.....	33
BAB IV PEMBAHASAN DAN HASIL EKSPERIMEN.....		36
4.1	Pemahaman Bisnis (<i>business understanding</i>).....	36
4.2	Pemahaman Data (<i>data understanding</i>).....	37
4.3	Persiapan Data (<i>data preparation</i>).....	43
4.4	Pemodelan (<i>modelling</i>)	44
4.5	Evaluasi (<i>evaluation</i>).....	54
BAB V KESIMPULAN DAN SARAN		56
5.1	Kesimpulan.....	56
5.2	Saran.....	56
DAFTAR PUSTAKA		58
LAMPIRAN		60

DAFTAR GAMBAR DAN ILUSTRASI

	Halaman
Gambar 1. 1 Pertumbuhan Pendapatan Segmen Fashion (sumber : statista.com).....	2
Gambar 1. 2 Metodologi penelitian yang digunakan	5
Gambar 2. 1 Proses CRISP-DM (sumber : Larose, 2014:24)	13
Gambar 2. 2 Ilustrasi Support Vector Machine (SVM).....	16
Gambar 2. 3 Ilustrasi Support Vector Regression (SVR)	17
Gambar 2. 4 Parameter SVR menggunakan pencarian grid	25
Gambar 2. 5 Contoh Pembagian 10-fold cross-validation (Berrar, 2019:2)	26
Gambar 3. 1 Konsep Sistem Penjualan.....	31
Gambar 3. 2 Tampilan Sales Cycle	32
Gambar 3. 3 Metodologi Penelitian.....	34
Gambar 4. 1 Bar plot atribut PRODUK terhadap atribut QTOTAL	41
Gambar 4. 2 Bar plot atribut PWKL terhadap atribut QTOTAL.....	41
Gambar 4. 3 Bar plot atribut TG_BPK terhadap atribut QTOTAL.....	42
Gambar 4. 4 Bar plot sebaran data atribut PWKL terhadap atribut PRODUK	42
Gambar 4. 5 Grafik pengujian rentang pencarian nilai parameter C	46
Gambar 4. 6 Grafik pengujian rentang pencarian nilai parameter ε	46
Gambar 4. 7 Grafik pengujian rentang nilai parameter γ	47
Gambar 4. 8 Grafik pengujian model SVR dengan optimasi parameter pencarian grid (pertahun).....	48
Gambar 4. 9 Grafik pengujian model SVR dengan optimasi parameter pencarian grid (perbulan).....	48

Gambar 4. 10 Grafik pengujian model SVR tanpa optimasi parameter pencarian grid (pertahun).....	49
Gambar 4. 11 Grafik pengujian model SVR tanpa optimasi parameter pencarian grid (perbulan).....	49
Gambar 4. 12 Grafik perbedaan performa hasil pengujian parameter population size dan maximum number of generation	50
Gambar 4. 13 Hasil optimasi pembobotan atribut.....	51
Gambar 4. 14 Grafik hasil pengujian	52
Gambar 4. 15 Grafik penjualan history dan peramalan perproduk 2020	53

DAFTAR TABEL

	Halaman
Tabel 2. 1 Penelitian Terkait	28
Tabel 4. 1 Penjelasan atribut pada data penjualan	38
Tabel 4. 2 Penjelasan atribut pada data master toko	39
Tabel 4. 3 Analisis statistik pada atribut penjualan tipe data nominal	39
Tabel 4. 4 Analisis statistik pada atribut penjualan tipe data numerik	40
Tabel 4. 5 Pemilihan data	43
Tabel 4. 6 Penggabungan data.....	44
Tabel 4. 7 Nilai parameter SVR	47
Tabel 4. 8 Hasil pengujian parameter population size dan maximum number of generation	49
Tabel 4. 9 Hasil optimasi pembobotan atribut.....	50
Tabel 4. 10 Hasil pengujian data baru	52
Tabel 4. 11 Peramalan penjualan perproduk januari-juni 2020	53
Tabel 4. 12 Hasil peramalan penjualan perwilayah perproduk Januari-juni 2020	53
Tabel 4. 13 Hasil peramalan penjualan berdasarkan tipe cust Januari-juni 2020	53
Tabel 4. 14 Hasil peramalan penjualan berdasarkan tipe toko Januari-juni 2020	54

DAFTAR RUMUS

	Halaman
2. 1 <i>Root Mean Squared Error</i> (RMSE)	11
2. 2 Fungsi linear nilai prediksi.....	18
2. 3 Persamaan kuadrat	18
2. 4 <i>Slack Variable</i>	19
2. 5 <i>Lagrange Multiplier</i>	19
2. 6 <i>Support Vector Regression</i>	20
2. 7 Prinsip <i>Karush Kuhn Tucker</i> (KKT) 1	20
2. 8 <i>Karush Kuhn Tucker</i> (KTT) 2	20
2. 9 Pemetaan fitur dimensi lebih tinggi	20
2. 10 Fungsi kernel.....	20
2. 11 Fungsi kernel gaussian RBF	21
2. 12 Persamaan <i>Particle Swarm Optimization</i>	21
2. 13 Persamaan posisi <i>particle swarm optimization</i>	21
2. 14 Persamaan kecepatan <i>particle swarm optimization</i>	21

DAFTAR LAMPIRAN

	Halaman
Lampiran A Hasil Pemodelan Dengan Tools Rapidminer	60
Lampiran B Dataset Penjualan (sebelum transformasi)	62
Lampiran C Dataset Penjualan (setelah transformasi)	68
Lampiran D Hasil Pengujian Optimasi Parameter	70
Lampiran E Hasil Pengujian PSO	71
Lampiran F Hasil Pengujian SVR-PSO.....	72

BAB I

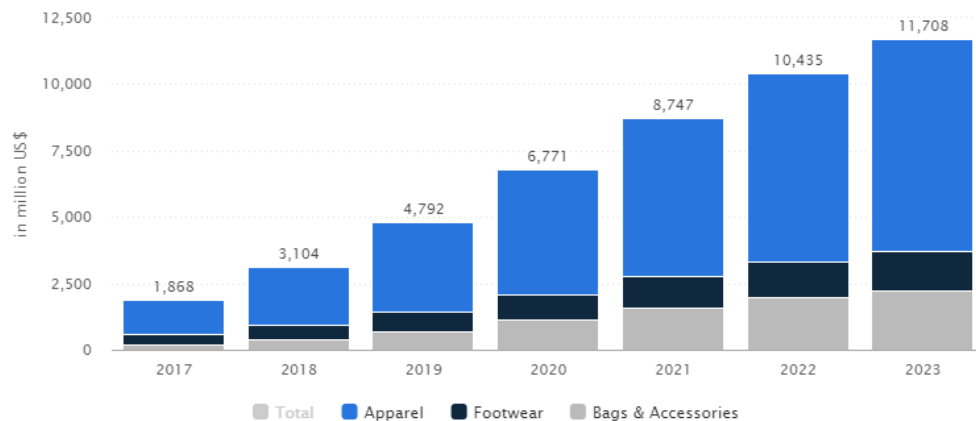
PENDAHULUAN

1.1 Latar Belakang

Tujuan utama dari peramalan (*sales forecasting*) adalah untuk menemukan informasi yang memiliki kemungkinan terjadi di masa depan. Menurut (Heizer, 2017:114) peramalan adalah salah satu bidang seni dan keilmuan matematis untuk memprediksikan peristiwa di masa depan dengan melibatkan data historis. Dengan menggunakan teknik peramalan, pengguna dapat menganalisis dan menemukan pengetahuan baru untuk membuat strategi, keputusan yang tepat dan merencanakan berbagai sumber daya dengan efektif dan efisien. Teknik peramalan telah banyak digunakan dalam berbagai bidang salah satunya adalah bidang bisnis.

Pada saat ini bisnis industri *fashion* di Indonesia berkembang dengan sangat pesat. Fenomena tersebut sejalan dengan pemenuhan gaya hidup yang menganggap pakaian bukan lagi sekedar alat untuk menutupi tubuh saja, tetapi sebagai identitas *life style* seseorang.

Menurut data Statista pada gambar 1.1, angka pendapatan di segmen *fashion* Indonesia berjumlah US \$3.104 juta pada tahun 2018, di mana angka tersebut menjadi pertumbuhan baik dengan melihat angka ditahun sebelumnya yang sebesar US \$1.868 juta (statista.com, Accessed 1 September 2019). Selain meninjau tren ditahun 2017 dan 2018, Statista memprediksikan pendapatan segmen *fashion* dari tahun 2019-2023 yang akan mengalami kenaikan secara signifikan.



Gambar 1. 1
Pertumbuhan Pendapatan Segmen *Fashion* (sumber : statista.com)

Mempertimbangkan data pada gambar 1.1, peramalan bisnis menjadi sangat penting bagi perusahaan guna memenuhi permintaan pasar yang mengalami kenaikan pada tahun-tahun berikutnya. Tetapi bagaimana meningkatkan ketepatan proses peramalan saat ini masih menjadi tanda tanya besar bagi perusahaan (Narkevičius, 2016:4). Untuk data yang mengandung pola tren, musiman atau yang dapat disebut dengan data *time series*, sering kali rentan mengalami kegagalan atau peramalan yang jauh dengan data aktual dikarenakan pengaruh penjualan yang sangat fluktuatif. Banyak metode yang dapat digunakan untuk membuat peramalan, dimulai dari metode statistik tradisional sampai metode atau algoritme untuk *data mining* modern seperti *machine learning* yang terkomputasi. Namun metode statistik tradisional seperti *exponential smoothing*, ARIMA dan *regression* memiliki akurasi lebih rendah dibandingkan dengan metode pembelajaran mesin seperti *support vector regression* (SVR), *long short-term memory* (LSTM) dan *artificial neural network* (ANN) (Namini, 2018:1400) (Zhang, 2014:15) (Kandanand, 2012:6). Oleh karena itu, pembelajaran mesin sangat berpeluang dalam mencapai kualitas peramalan yang baik.

Support Vector Regression (SVR) merupakan salah satu pembelajaran mesin populer dalam membuat peramalan dengan kemampuannya mengatasi *overfitting* yang merupakan masalah dasar terkait data yang befluktuatif (Rifqi, 2018:3333). SVR telah banyak diimplementasikan dalam kasus peramalan. Penelitian yang dilakukan oleh (Rifqi,

2018:3340) menggunakan algoritme SVR untuk peramalan permintaan darah PMI kota Malang, menghasilkan bahwa nilai rata-rata kesalahan (*average error*) pengujian berkisar 3,8%, yang di mana jika hasil rata-rata kesalahan yang dihasilkan <10% dapat dikategorikan baik untuk peramalan. Penelitian lainnya (Hasan, 2015:554) (Lu, 2011:74) (Wen, 2014:354) menunjukkan jika algoritme SVR memiliki kemampuan generalisasi dalam peramalan yang baik. Algoritme SVR juga menjadi algoritme dengan rata-rata kesalahan yang paling rendah dalam peramalan dibandingkan dengan algoritme pembelajaran mesin lainnya seperti *Artificial Neural Network* (Anandhi, 2013:375) (Kandananond, 2012:6) dan *decision tree* (Wen, 2014:351).

Kehandalan performa algoritme SVR sangat ditentukan dengan pemilihan parameter yang digunakan (Kridanto, 2015:302). Tidak ada aturan umum dalam pemilihan parameter-parameter tersebut. Optimasi atribut dilakukan untuk memberikan atribut yang optimal. *Particle Swarm Optimization* (PSO) adalah salah satu algoritme yang dikembangkan oleh Eberhart dan Kennedy untuk optimasi berbasis populasi. Algoritma PSO tidak hanya memiliki kemampuan pencarian global yang kuat, tetapi dapat juga memecahkan masalah *over-fitting* (Lu, 2011:71). Selain itu, dalam kasus peramalan penjualan, algoritma PSO memberikan efisiensi dan akurasi yang lebih unggul dibandingkan dengan *genetic algorithm* (Lu, 2011:74).

Dalam penelitian ini akan dilakukan pencarian suatu model yang bisa melakukan peramalan jumlah penjualan dengan menggunakan teknik *Data Mining*. Berdasarkan fungsi dari dua algoritme yang telah dibahas sebelumnya, penelitian ini akan menggunakan algoritme SVR dalam membuat model peramalan yang akan dikombinasikan dengan algoritme PSO sebagai optimasi atribut yang digunakan. Model yang dihasilkan dituangkan dalam bentuk visualisasi data yang diharapkan bisa digunakan sebagai sumber informasi bagi pihak eksekutif dalam melakukan pengambilan keputusan. Pengambilan keputusan yang diambil dapat berupa penentuan target untuk operasional, penentuan strategi marketing, atau sebagai acuan produksi dalam menyiapkan produk yang sesuai dengan kebutuhan.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang, dapat dirumuskan masalah yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana mengkombinasikan algoritme *Support Vector Regression* (SVR) dan algoritme *Particle Swarm Optimization* (PSO) dalam peramalan penjualan?
2. Berapa tingkat akurasi peramalan dengan algoritme *Support Vector Regression* (SVR) dan algoritme *Particle Swarm Optimization* (PSO) dalam peramalan penjualan?

1.3 Tujuan

Berdasarkan rumusan masalah maka tujuan dari penelitian ini adalah :

1. Melakukan kombinasi algoritme *Support Vector Regression* (SVR) dan algoritme *Particle Swarm Optimization* (PSO) dalam peramalan penjualan.
2. Mengetahui tingkat akurasi peramalan dengan algoritme *Support Vector Regression* (SVR) dan algoritme *Particle Swarm Optimization* (PSO).

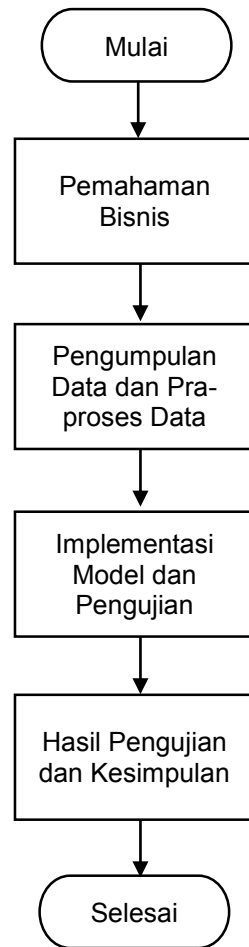
1.4 Ruang Lingkup Penelitian

Ruang lingkup yang akan dibahas dari penelitian ini dilakukan pada data penjualan periode tahun 2016-2018 dan algoritme yang digunakan adalah *Support Vector Regression* (SVR) dan *Particle Swarm Optimization* (PSO).

1.5 Metodologi Penelitian

Metodologi penelitian yang digunakan pada penelitian ini ditampilkan pada gambar 1.2. Berdasarkan gambar 1.2 terdapat 4 fase terkait penelitian yang dilakukan. Fase pertama dalam penelitian dimulai dari pemahaman bisnis. Dalam fase pemahaman bisnis, studi literatur, observasi dan wawancara dilakukan agar analisa masalah dan tujuan penelitian dapat diinisialisasikan secara jelas. Setelah analisa masalah dan tujuan penelitian diinisialisasikan secara jelas selanjutnya fase pengumpulan data dan pra-proses data. Pengumpulan data disertai dengan mengeksplorasi data dilakukan untuk mengetahui secara lebih dalam tentang struktur data yang akan diproses dan mengkonfirmasi lebih

lanjut lagi terkait tujuan penelitian sebelumnya. Tahap pra-proses data dilakukan agar data dapat diproses pada tahap pemodelan. Fase implementasi model dan pengujian merupakan fase selanjutnya untuk dilakukannya pengujian terkait hasil yang ditunjukkan antara data dan algortime yang dipakai. Tahap selanjutnya hasil pengujian dan kesimpulan yang merupakan tahap terakhir sebagai hasil eksperimen dan penarikan kesimpulan terkait penelitian.



Gambar 1. 2
Metodologi penelitian yang digunakan

Fase-fase berikut penjelasan terkait metodologi penelitian yang digunakan adalah sebagai berikut:

1. Pemahaman Bisnis

Dalam fase pemahaman bisnis terdapat kegiatan-kegiatan yang dilakukan agar pemahaman dalam bisnis dapat dipahami. Kegiatan pemahaman bisnis adalah sebagai berikut :

a. Studi Literatur

Pencarian dan pengumpulan referensi yang relevan dengan topik penelitian dilakukan, sehingga sesuai dengan rumusan masalah dan tujuan yang akan diteliti dalam karya tulis ini. Kegiatan ini mencakup studi untuk menggali informasi tentang analisis masalah, *data mining*, metodologi, penelitian-penelitian yang terkait dengan algoritme *support vector regression* atau algoritme *particle swarm optimization* dan perancangan model.

b. Observasi

Observasi dilakukan dengan cara melakukan kunjungan langsung pada perusahaan. Hal ini dilakukan agar kegiatan-kegiatan yang menjadi dasar penelitian dapat dirasakan secara langsung.

c. Wawancara

Wawancara dilakukan dengan berkomunikasi langsung dengan para pemangku kepentingan perusahaan. Hal ini dilakukan agar tujuan bisnis yang ingin dicapai perusahaan dapat ditangkap dengan baik.

Berdasarkan kegiatan tersebut diperoleh tujuan dari bisnis berdasarkan permasalahan bisnis, penilaian terhadap situasi yang ada dan tujuan dari *data mining*.

2. Pengumpulan Data dan Pra-proses Data

Pengumpulan data dan pra-proses data sebelum memasuki tahap *data mining*. Tahap ini dilakukan dari pengumpulan data, deskripsi data, eksplorasi data, pemilihan data, pembersihan data, transformasi data dan penggabungan data.

3. Implementasi Model dan Pengujian

Tahap ini dilakukan berbagai eksperimen terhadap *dataset* untuk mendapatkan parameter dan model terbaik dalam memprediksi jumlah interaksi yang terjadi. Dilakukan juga pengujian akurasi untuk setiap model yang telah dihasilkan.

4. Hasil Akhir dan Kesimpulan

Tahap ini dilakukan penarikan kesimpulan terhadap hasil dari seluruh penelitian yang telah dilaksanakan. Diberikan juga masukan atau saran untuk penelitian lanjutan.

1.6 Sistematika Penelitian

Bab I Pendahuluan. Bab ini menjelaskan tentang permasalahan dalam penelitian, perumusan dari topik permasalahan, tujuan, ruang lingkup dari penelitian yang dilakukan dan sistematika penulisan.

Bab II Landasan Teori. Bab ini berisi penjelasan mengenai teori-teori *data mining*, peramalan, algoritme yang akan dipakai untuk penelitian, teori penjualan dan penelitian-penelitian terkait dengan penelitian ini.

Bab III Objek dan Metodologi Penelitian. Bab ini berisi penjelasan mengenai objek penelitian dan metodologi penelitian yang akan digunakan.

Bab IV Pembahasan dan Hasil Eksperimen. Bab ini berisi pembahasan mengenai tahapan-tahapan dalam eksperimen yang sesuai dengan metodologi penelitian yang dipakai, dimulai dengan menentukan tujuan yang ingin dicapai perusahaan dan kaitannya dengan tujuan *data mining*, selanjutnya dilakukan pra proses data sampai kepada pengujian model *data mining* yang kemudian akan dievaluasi dan menjadi hasil eksperimen.

Bab V Kesimpulan. Bab ini menjelaskan tentang kesimpulan yang didapatkan dari hasil penelitian dan saran agar penelitian dapat dikembangkan kembali.

BAB II

LANDASAN TEORI

2.1 Peramalan (*forecasting*)

Tujuan utama peramalan (*forecasting*) adalah menghasilkan prediksi yang akurat. Pada dunia bisnis, ketika peramalan terlalu rendah berdampak pada kurangnya *inventory*. Begitu pun sebaliknya ketika peramalan terlalu tinggi berdampak pada *inventory* yang berlebihan dan memakan *budget* yang tidak efektif. Peramalan (*forecasting*) merupakan seni dan ilmu untuk memprediksi peristiwa masa depan. Peramalan mungkin melibatkan pengambilan data historis (seperti rekap penjualan masa lalu) dan memproyeksikannya ke masa depan dengan model matematika (Heizer, 2017:108).

Dari definisi di atas bisa disimpulkan bahwa peramalan sangat penting bagi perusahaan dalam mencapai tujuan perusahaan, karena peramalan berdampak pada keputusan manajemen ketika melakukan perencanaan ke depannya. Perencanaan yang efektif baik dalam jangka pendek maupun jangka panjang tergantung pada perkiraan permintaan untuk produk-produk perusahaan.

Terdapat 7 langkah dasar dalam menentukan peramalan (Heizer, 2017:110) :

1. Tentukan tujuan peramalan.
2. Pilih item yang akan diramalkan.
3. Tentukan horizon waktu peramalan.
4. Pilih metode peramalan.
5. Kumpulkan data yang diperlukan untuk membuat peramalan.
6. Buat peramalan.
7. Validasi dan terapkan hasilnya.

2.2.1 Metode Peramalan

Secara umum ada dua metode untuk membuat peramalan, yaitu metode kuantitatif dan metode kualitatif. Metode model matematika yang bergantung pada data historis atau *variable* asosiatif digunakan oleh metode kuantitatif. Sedangkan metode kualitatif

menggabungkan faktor-faktor seperti emosi, intuisi, pengalaman pribadi dan nilai sistem dalam mencapai peramalan. Berikut penjelasan tentang kedua metode tersebut :

1. Metode Kualitatif

Ada empat teknik metode kualitatif menurut (Heizer, 2017:111) :

a. Juri pendapat eksekutif

Metode ini menggabungkan dan merata-ratakan beberapa pandangan eksekutif mengenai keputusan atau perkiraan yang lebih spesifik atau mengumpulkan pendapat beberapa sekelompok ahli atau manajer tingkat tinggi untuk membentuk perkiraan kelompok permintaan. Dalam praktiknya para eksekutif dikumpulkan, misalkan dari bidang keuangan, sumber daya manusia, pemasaran, pembelian, dan produksi, yang memberikan informasi latar belakang, pengalaman dan pendapat kepada direksi.

b. Metode Delphi

Ada tiga jenis kelompok dalam metode ini, yaitu: pembuat keputusan, personel staf, dan responden. Pembuat keputusan biasanya terdiri dari 5 hingga 10 anggota ahli yang akan membuat perkiraan aktual. Personel staf membantu pembuat keputusan dengan menyiapkan, mendistribusikan, mengumpulkan, dan merangkum serangkaian kuesioner dan hasil survei. Responden adalah sekelompok orang, sering kali berlokasi di tempat yang berbeda, yang penilaiannya dinilai. Sebelum ramalan dibuat, kelompok tersebut memberikan saran kepada pembuat keputusan.

c. Komposit tenaga penjualan

Dalam pendekatan ini, setiap tenaga penjualan memperkirakan penjualan apa yang kemungkinan terjadi di wilayahnya. Prakiraan ini kemudian ditinjau untuk memastikan bahwa perkiraan tersebut realistis. Kemudian mereka digabungkan pada tingkat yang lebih global untuk mencapai perkiraan keseluruhan.

d. Survei pasar

Survei pasar memungkinkan untuk meminta saran dari pelanggan atau calon pelanggan mengenai rencana pembelian yang akan dilakukan di masa depan.

Hal itu dapat membantu dalam meningkatkan desain produk dan perencanaan untuk produk baru.

2. Metode Kuantitatif

Secara keseluruhan metode ini menggunakan data historis. Ada lima metode yang terbagi ke dalam dua kategori, yaitu model *time series* dan model asosiasi.

Tujuan utama *time series* adalah memberikan pengaturan statistik untuk menggambarkan karakter data yang tampaknya berfluktuasi secara acak dari waktu ke waktu. Waktu yang digunakan dapat berupa menit, jam, hari, minggu, bulan, tahun dan sebagainya. *Time series* merupakan serangkaian data pengamatan yang berasal dari suatu sumber tetap dan terjadi berdasarkan indeks waktu t secara beruntun dengan interval waktu yang tetap (Astuti, 2013:1). Secara umum, *time series* dipengaruhi oleh empat komponen utama yang dapat dipisahkan dari data yang sedang diamati. Komponen-komponen tersebut antara lain, tren, variasi siklus (*cyclical*), variasi musiman (*seasonal*), dan variasi acak (*irregular*) (Adhikari, 2013:12). Penjelasan singkat mengenai tren adalah gerakan yang menunjukkan suatu kecenderungan umum dari *time series* apakah terjadi kecenderungan naik, turun atau konstan. Sebagai contoh, seri yang berhubungan dengan angka kematian atau minat konsumen, seri yang berkaitan dengan pertumbuhan pada suatu populasi. Variasi siklus (*cyclical*) di dalam *time series* menjelaskan perubahan jangka menengah yang disebabkan oleh keadaan yang berulang dalam siklus. Sebagian besar rangkaian dalam bidang ekonomi dan perdagangan menunjukkan semacam variasi siklus, dan untuk menilai variasi siklus tersebut biasanya membutuhkan waktu dua tahun atau lebih. Fluktuasi dalam satu musim disebut variasi musiman (*seasonal*). Hal ini merupakan faktor penting bagi manajer ataupun pengusaha untuk membuat rencana masa depan yang tepat. Contoh dari variasi musiman adalah keadaan iklim atau cuaca, musim libur panjang, hari raya keagamaan dan lain lain. Terakhir adalah variasi acak (*irregular*), merupakan suatu gerakan yang disebabkan oleh kondisi yang tidak terduga, contoh perang antar suku, gempa bumi, banjir dan lain-lain.

2.2.2 Mengukur Akurasi Peramalan

Mengukur akurasi peramalan bisa dilakukan dengan membandingkan nilai perkiraan dengan nilai aktual atau nilai yang diamati. *Root Mean Squared Error* (RMSE) merupakan salah satu cara untuk mengevaluasi model regresi dengan mengukur tingkat akurasi hasil perkiraan suatu model. RMSE dihitung dengan mengkuadratkan error (prediksi – observasi) dibagi dengan jumlah data (= rata-rata), lalu diakarkan.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}} \quad (2.1)$$

Nilai RMSE rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai obeservasinya. RMSE menghitung seberapa berbedanya seperangkat nilai. Semakin kecil nilai RMSE, semakin dekat nilai yang diprediksi dan diamati.

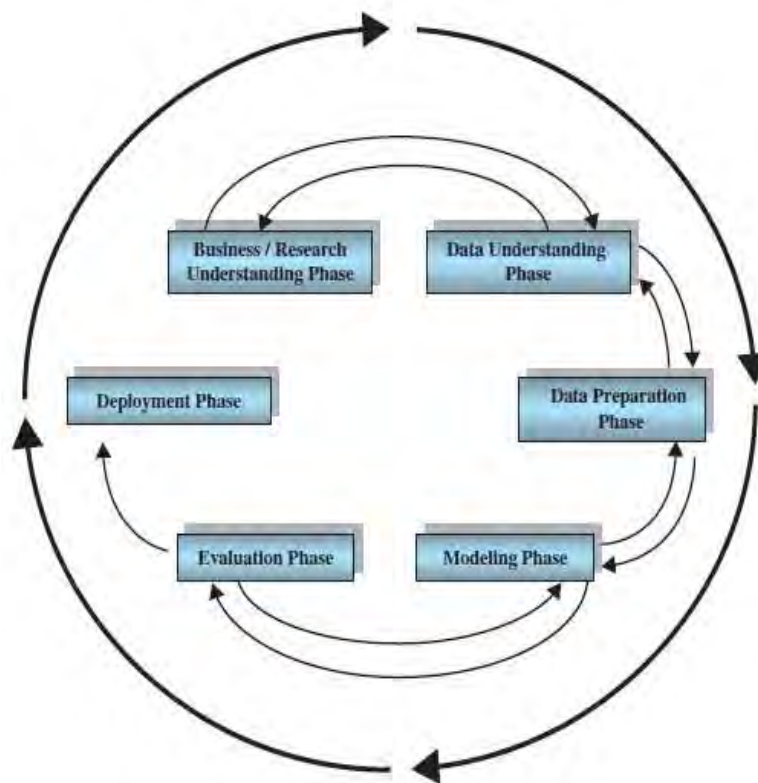
2.2 Data Mining

Cabang ilmu *data mining* muncul dari pertemuan ilmu statistik dan pembelajaran mesin (*machine learning*) yang menyediakan teknologi untuk membantu menganalisis dan memahami informasi yang terkandung dalam basis data (Palmer, 2011:373). *Data mining* dapat dikatakan sebagai hasil evolusi alami dari teknologi informasi. Secara garis besar *data mining* merupakan proses menemukan pola atau tren yang berguna di dalam *dataset*. Sumber data bisa mencakup *data base*, *data warehouse*, *web*, repositori informasi lainnya, atau data yang dialirkan ke sistem secara dinamis (Han, 2012:8). *Data mining* juga didefinisikan sebagai suatu proses untuk menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan data yang berukuran besar dengan menggunakan teknik pengenalan pola seperti teknik statistika dan matematika (Larose, 2014:1). Selain itu definisi lainnya tentang *data mining* adalah serangkaian analisis dengan tujuan untuk menemukan pola dari sekumpulan data yang mempunyai makna yang tersembunyi (Wicaksana, 2013:42).

Berdasarkan beberapa definisi sebelumnya, dapat disimpulkan *data mining* menggunakan kombinasi dari basis pengetahuan eksplisit, dan keterampilan analisis yang canggih, untuk mengungkap tren dan pola tersembunyi. *Data mining* digunakan untuk memecahkan berbagai masalah bisnis, seperti deteksi kecurangan, *market basket analysis*, dan analisis *churn* pelanggan. Umumnya, perusahaan menggunakan alat *data mining* pada volume besar data terstruktur, seperti basis data manajemen hubungan pelanggan (CRM) atau *inventory* suku cadang pesawat. Tujuan dari *data mining* adalah untuk menjelaskan dan memahami data. *Data mining* tidak dimaksudkan untuk membuat prediksi atau mendukung hipotesis (Hurwitz, 2018:12).

2.3 Metodologi Data Mining

Konsep dasar *data mining* adalah memproses data mentah menjadi suatu pengetahuan baru. Bidang *data mining* dalam beberapa tahun terakhir telah terjadi pertumbuhan yang cukup besar. Beberapa upaya dilakukan untuk menghasilkan beberapa proses yang kemudian dijadikan standar proses dalam pengembangan *data mining*. *Cross-Industry Standard Process for Data Mining* (CRISP-DM) merupakan Salah satu proses yang telah dijadikan standar. CRISP-DM adalah suatu metodologi *data mining* yang dikembangkan melalui upaya konsorsium yang awalnya disusun oleh Daimler Chrysler, SPSS, dan NCR. Berikut 6 tahap proses dalam CRISP-DM yang dapat dilihat pada gambar 2.1 (Larose, 2014:24).



Gambar 2. 1
Proses CRISP-DM (sumber : Larose, 2014:24)

Proses CRISP-DM yang ditunjukkan pada Gambar 2.1 dapat dijelaskan sebagai berikut:

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)

Fase ini berorientasi pada pemahaman tujuan dan permasalahan dalam perspektif bisnis. Mendefinisikan masalah dan memahami tujuan bisnis sangat penting agar dapat digeneralisasikan ke dalam ilmu *data mining* sebagai upaya pemecahan masalah dalam mencapai tujuan bisnis.

2. Fase Pemahaman Data (*Data Understanding Phase*)

Selama fase ini, data dikumpulkan dan dilakukan proses integrasi jika data berasal lebih dari satu sumber. Selanjutnya dilakukan eksplorasi untuk mendapatkan keakraban dengan data dan mempelajarinya, termasuk bentuk, konten dan struktur. Pengetahuan dan pemahaman tentang fitur numerik dan properti data (misalkan data kategorikal versus data kontinu) akan menjadi penting selama proses persiapan data dan penting untuk pemilihan alat statistik atau algoritma yang

digunakan selama fase pemodelan. Kesimpulannya, dengan fase ini merupakan pemandu untuk proses analisis selanjutnya, seperti tahap persiapan, pemodelan, mengevaluasi hasil dan menyiapkan output/laporan.

3. Fase Persiapan Data (*Data Preparation Phase*)

Mencakup semua kegiatan untuk membangun *dataset* akhir dari data mentah. Tugas persiapan data kemungkinan akan dilakukan beberapa kali dan tidak dalam urutan yang ditentukan. Tugas ini meliputi tabel, catatan dan seleksi atribut serta transformasi dan pembersihan data untuk fase pemodelan.

4. Fase Pemodelan (*Modeling Phase*)

Terdapat 4 tahapan dalam fase pemodelan :

- a. Memilih dan mengaplikasikan teknik pemodelan data yang sesuai
- b. Lakukan kalibrasi model untuk optimasi hasil
- c. Jika memungkinkan, gunakan beberapa teknik berbeda untuk permasalahan yang sama.
- d. Kembali ke fase persiapan data jika diperlukan untuk menjadikan data ke dalam bentuk tertentu.

5. Fase Evaluasi (*Evaluation Phase*)

Selanjutnya mengevaluasi satu atau lebih model untuk menetapkan model yang digunakan telah relevan dengan tujuan bisnis. Selanjutnya menentukan apakah terdapat permasalahan penting dari bisnis yang tidak tertangani dengan baik. Selanjutnya mengambil keputusan berkaitan dengan penggunaan hasil *data mining*.

6. Fase Penyebaran (*Deployment Phase*)

Pada fase penyebaran dari hasil data yang diperoleh, perlu adanya pengelompokan data dan dirangkai agar dapat menyesuaikan dengan kebutuhan dan dapat melakukan proses *data mining* secara berulang. Bentuk dari fase ini dapat berupa pembuatan laporan atau penerapan proses *data mining* pada departemen atau divisi lain.

2.4 Machine Learning

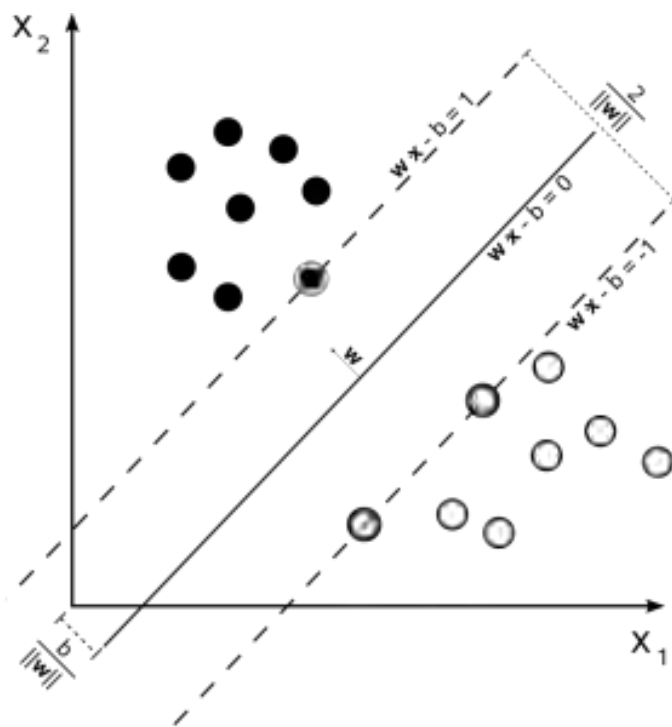
Machine learning (ML) adalah bentuk dari *Artificial Intelligence* (AI) yang memungkinkan suatu sistem untuk belajar secara mandiri dari data daripada melalui pemrograman eksplisit. ML erat kaitannya dengan performa atau kinerja dari pada arti dari pengetahuan / *knowledge* yang ditemukan. ML dapat diuji dengan mengamati perilaku saat ini dan membandingkannya dengan perilaku masa lalu (Witten, 2011:7). Dalam beberapa tahun terakhir, banyak perusahaan menerapkan metode ML dan mengalami perkembangan. Dengan model ML yang tepat, perusahaan memiliki kemampuan untuk terus memprediksi perubahan dalam bisnis sehingga mereka dapat memprediksi apa yang akan terjadi selanjutnya. Karena data terus ditambahkan, model ML memastikan bahwa solusi diperbarui secara konstan (Hurwitz, 2018:4).

Ciri khas dari ML adalah adanya proses *training*. Dengan kata lain, ML membutuhkan data untuk dipelajari yang disebut sebagai data *training*. Klasifikasi merupakan metode dalam ML yang dipakai oleh mesin untuk mengklasifikasikan obyek berdasarkan ciri tertentu sebagaimana manusia mencoba membedakan benda satu dengan yang lain. Sedangkan prediksi atau regresi digunakan oleh mesin dalam menerka *output* dari suatu data *input* berdasarkan data yang sudah dipelajari. Metode ML yang paling populer yaitu *Decision Making System*, *Support Vector Machine* (SVM) dan *Neural Network*.

2.5 Support Vector Regression (SVR)

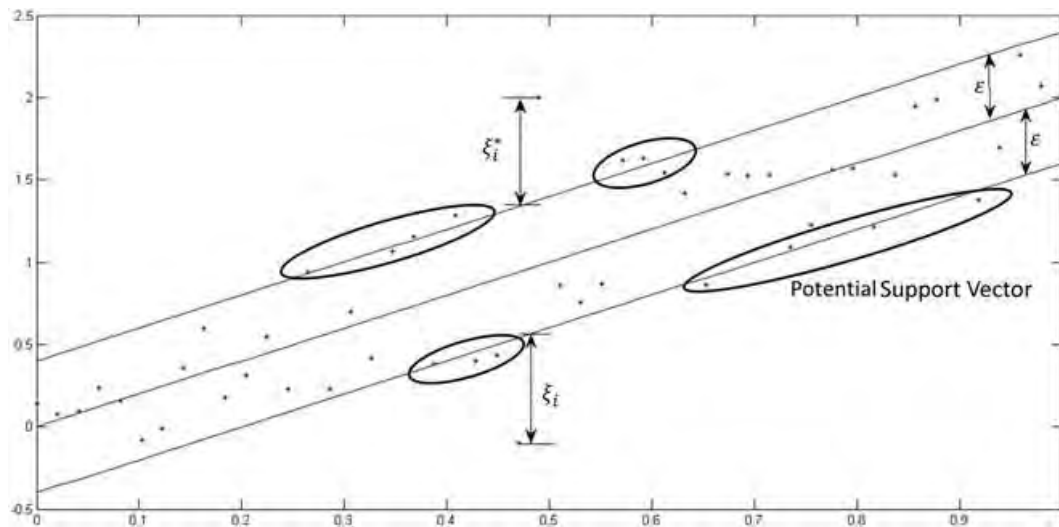
Support Vector Regression (SVR) merupakan adaptasi dari paradigma klasifikasi berdasarkan statistik / pembelajaran mesin yang baru diperkenalkan yaitu, *Support Vector Machine* (SVM) (Hsieh, 2011:2). Berawal dari pembelajaran statistik atau teori Vapnik-Chervonenkis, SVM merupakan salah satu pembelajaran mesin populer yang dapat memberikan solusi baik dalam menggeneralisasikan data yang belum terlihat. Konsep SVM dapat digeneralisasikan agar dapat diterapkan pada kasus regresi. Seperti halnya SVM, SVR bertujuan untuk menciptakan *hyperplane* dan *support vector* dengan penggunaan fungsi kernel untuk menentukan nilai prediksi. *Hyperplane* merupakan garis pembatas yang memisahkan antara dua kelas yang merupakan ciri umum dari konsep

klasifikasi. Dalam menciptakan *hyperplane*, *support vector* diperlukan untuk mengetahui garis atas yang merupakan batas atas dan garis bawah yang merupakan batas bawah dari masing-masing kelas. *Support vector* merupakan istilah untuk merepresentasikan suatu titik-titik atau data poin yang menjadi calon pembatas agar semua data poin bisa masuk ke dalam zona atau kelas tertentu, sehingga *hyperplane* dapat terbentuk. *Hyperplane* terbaik adalah menciptakan pembagian paling optimal dengan *hyperplane* yang terletak ditengah-tengah antara dua set objek dalam dua kelas. Mencari *hyperplane* terbaik ekuivalen dengan memaksimalkan jarak atau margin yang memisahkan dua kelas tersebut. Margin antara *hyperplane* dengan garis batas atas ataupun garis batas bawah adalah sama. Margin tersebut dikenal dengan sebutan epsilon (ϵ). Singkatnya, konsep SVM adalah bagaimana memisahkan dua set objek dalam dua kelas dengan memaksimalkan margin epsilon antara *hyperplane* dengan batas atas dan batas bawah yang ditandai dengan *support vector*, untuk menciptakan pembagian yang optimal. Pembagian optimal adalah kondisi dimana semua data poin masuk ke dalam salah satu dari dua zona kelas dengan tingkat kepastian yang presisi (margin epsilon yang paling besar). Ilustrasi dari konsep SVM ditunjukkan pada gambar 2.2.



Gambar 2. 2
Ilustrasi *Support Vector Machine* (SVM) (sumber : Awad, 2015:48)

Konsep SVR tidak berbeda jauh dengan konsep SVM. Perbedaannya adalah jika SVM bagaimana menciptakan *hyperplane* untuk memisahkan dua set objek dalam dua zona kelas dengan memaksimalkan margin epsilon, sedangkan konsep SVR bagaimana memasukkan semua set data ke dalam satu zona dengan meminimalkan margin epsilon dan tetap menciptakan *hyperplane*. Meskipun tidak sepopuler SVM, SVR terbukti menjadi alat efektif dalam prediksi fungsi nilai riil. Pada gambar 2.3 menampilkan ilustrasi dari SVR.



Gambar 2. 3
Ilustrasi *Support Vector Regression* (SVR) (sumber : Awad, 2015:67)

Berdasarkan gambar 2.3, menunjukkan sebuah *hyperplane* yang diapit oleh dua garis batas atas dan garis batas bawah. Terdapat beberapa titik atau data point yang dilingkari, dan titik tersebut merupakan potensial *support vectors*. Artinya titik-titik tersebut merupakan calon pembatas atas dan pembatas bawah, sehingga semua titik dapat masuk ke satu zona dengan tetap sebisa mungkin meminimasi margin epsilon (ϵ). Hal tersebut yang membuat SVR mampu mengatasi *overfitting* (Awad, 2015:67). Jika divisualisasikan, garis *hyperplane* sebisa mungkin melewati semua titik-titik data. Sebagai pendekatan pembelajaran terawasi (*supervised learning*), SVR melatih menggunakan fungsi kerugian simetris (*symmetrical loss function*) yang sama-sama memberikan pinalti untuk kesalahan prediksi tinggi dan rendah. Dengan menggunakan pendekatan ϵ -tidak sensitif (ϵ -insensitive) Vapnik, tabung fleksibel radius minimal dibentuk secara simetris di sekitar fungsi yang diperkirakan, sehingga nilai absolut kesalahan yang kurang dari ambang tertentu ϵ diabaikan baik di atas maupun di bawah prediksi. Dengan cara ini, titik-titik di

luar tabung diberikan pinalti, tetapi titik-titik lain yang berada di dalam tabung, baik di atas atau di bawah fungsi, tidak menerima penalti. Salah satu keuntungan utama SVR adalah bahwa kompleksitas komputasinya tidak tergantung pada dimensi ruang input. Selain itu, ia memiliki kemampuan generalisasi yang sangat baik, dengan akurasi prediksi yang tinggi.

Penjelasan lebih jelas tentang proses SVR dirincikan sebagai berikut. Diberikan data $\{(x_i, y_i)\}_{i=1}^N$, dimana x_i adalah vektor *input* dari data, y_i adalah nilai label dari data pelatihan, dan N adalah jumlah data pelatihan. Fungsi linear dari data tersebut dapat dilihat pada persamaan (2.2):

$$g(x_i) = (w \cdot x_i) + b \quad (2.2)$$

dimana :

$g(x_i)$ = nilai yang diprediksi

w = nilai bobot

b = nilai bias

SVR merumuskan masalah perkiraan fungsi ini sebagai masalah optimasi yang berupaya menemukan suatu garis pemisah atau yang disebut *hyperplane* terbaik, sambil meminimalkan kesalahan prediksi, yaitu jarak antara hasil yang diprediksi ($g(x_i)$) dan yang diinginkan (Awad, 2015:68). Hal ini dapat dilakukan dengan membentuk optimasi persamaan kuadrat pada persamaan :

minimalkan :

$$Z = \frac{1}{2} ||w^2|| \quad (2.3)$$

dengan batasan :

$$y_i - w(x_i) - b \leq \varepsilon$$

$$w(x_i) - y_i + b \leq \varepsilon$$

Kendala yang terjadi adalah untuk meminimalkan kesalahan antara fungsi nilai prediksi untuk input yang diberikan dan output aktual. SVR mengadopsi fungsi kerugian yang tidak sensitif (*insensitive loss function*), memberikan pinalti untuk nilai prediksi yang

jauh dari output yang diinginkan. Nilai ε menentukan lebar tabung, nilai yang lebih kecil menunjukkan toleransi yang lebih rendah untuk kesalahan prediksi dan juga mempengaruhi jumlah *support vectors*. Pada gambar 4.1 menunjukkan jika nilai ε berkurang, garis batas digeser ke dalam. Oleh karena itu, lebih banyak titik data di sekitar batas, yang menunjukkan lebih banyak *support vectors*. Demikian pula, peningkatan nilai ε akan menghasilkan lebih sedikit poin di sekitar batas. Dalam kasus ini, diasumsikan bahwa semua titik berada pada rentang $(x) \pm \varepsilon$, kondisi demikian disebut sebagai permasalahan yang *feasible*. Sedangkan pada kenyataannya tidak selalu demikian, ada kemungkinan beberapa titik berada di luar area tersebut (lihat gambar 4.1). Untuk mengatasi permasalahan ini, dapat ditambahkan *slack variable* yaitu ξ_i, ξ_i^* sehingga menghasilkan persamaan:

meminimalkan :

$$Z = \frac{1}{2} \|w^2\| + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2.4)$$

dengan batasan :

$$y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i$$

$$(w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Nilai C merupakan nilai penalti yang diberikan jika hasil prediksi memiliki kesalahan melebihi ε . Persamaan optimasi kuadratik pada persamaan (2.4) dapat diselesaikan menggunakan fungsi *Lagrange* sehingga didapatkan nilai bobot yang baru pada persamaan (2.5).

$$w = \sum_{i=1}^l (a_i - a_i^*) x^i \quad (2.5)$$

dimana :

$$a_i, a_i^* = \text{Lagrange Multiplier}$$

Dengan menggunakan nilai bobot pada persamaan (2.5), persamaan (2.2) dapat diubah menjadi persamaan (2.6).

$$g(x_i) = (w \cdot x_i) + b = \sum_{i=1}^l (a_i - a_i^*) < x_i, x > + b \quad (2.6)$$

Di mana x_i merupakan titik paling informatif dalam SVR atau biasa disebut *Support Vector* (SV) yang mewakili keseluruhan informasi dari model SVR. Support vector dalam SVR terletak di luar tabung elektronik. Semakin kecil nilai

Dalam SVR, vektor dukungan adalah titik-titik yang terletak di luar tabung elektronik. Semakin kecil nilai ϵ , semakin banyak poin yang terletak di luar tabung dan karenanya semakin besar jumlah vektor dukungan. Dengan ALB-SVR ϵ -tube dipotong menjadi dua, dan ϵ -terikat yang lebih rendah dijatuhkan. Oleh karena itu, untuk parameter g dan ϵ yang sama, lebih banyak titik terletak di luar tabung, dan ada lebih banyak vektor dukungan. Ini berarti bahwa jumlah vektor dukungan lebih besar untuk ALB-SVR daripada untuk SVR. Peningkatan jumlah vektor dukungan ini menunjukkan bahwa menggunakan ALB SVR memiliki beberapa efek negatif pada kompleksitas fungsi estimasi. Meskipun persentase set data kesalahan relatif lebih tinggi (5,06 persen), ini dapat diterima, karena tujuan utamanya adalah untuk mengurangi jumlah perkiraan yang lebih rendah dan ini dicapai.

Nilai b dapat dihitung menggunakan prinsip Karush Kuhn Tucker (KKT) sebagai berikut :

$$b = y_i - \langle w^* - x_i \rangle - \epsilon, \quad 0 \leq a_i \leq C \quad (2.7)$$

$$b^* = y_i - \langle w^* - x_i \rangle + \epsilon, \quad 0 \leq a_i^* \leq C \quad (2.8)$$

Untuk mengatasi permasalahan non-linear, fungsi *kernel* dapat diterapkan dalam SVR untuk memetakan data pelatihan ke dalam ruang fitur dengan dimensi yang lebih tinggi menggunakan persamaan (2.9) dan fungsi *kernel* pada persamaan (2.10).

$$\varphi : R^n \rightarrow R^{n+h}; x \rightarrow \varphi(x) \quad (2.9)$$

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (2.10)$$

Salah satu fungsi *kernel* yang sering digunakan dalam SVR adalah *kernel* Gaussian RBF. *Kernel* tersebut dapat memetakan data pelatihan ke dalam ruang fitur dan handal

dalam menangani permasalahan non-linear. Fungsi *kernel* Gaussian RBF dapat dilihat pada persamaan (2.11).

$$K(x_i, x_j) = \exp(-\frac{1}{2\sigma^2} ||x_i - x_j||^2) \quad (2.11)$$

Namun kehandalan performa algoritme SVR sangat ditentukan dengan pemilihan parameter yang digunakan (Kridanto, 2015:302). Dalam penggunaan fungsi *kernel* linear dalam SVR, terdapat 3 parameter yang harus ditentukan, yaitu: nilai C sebagai nilai penalti, nilai ε sebagai nilai toleransi kesalahan, dan nilai γ yang menyatakan lebar atau jarak satu titik ke titik lainnya yang dianggap serupa dari fungsi Gaussian RBF.

2.6 Particle Swarm Optimization (PSO)

Algoritme *Particle Swarm Optimization* (PSO) adalah teknik komputasi stokastik yang secara iteratif mengoptimalkan solusi alternatif suatu masalah sampai mencapai sasaran *fitness* atau kualitas (Awad, 2015:113). Algoritme PSO pertama kali diperkenalkan oleh Dr. Kennedy dan Dr. Eberhart pada tahun 1995 yang terinspirasi oleh perilaku sosial dari pergerakan sekawanan burung atau ikan (*bird flocking* atau *fish schooling*).

Algoritme PSO pada dasarnya belajar dari aktivitas atau perilaku hewan untuk menyelesaikan masalah optimasi. Dalam PSO, setiap anggota populasi disebut *particle* dan populasi disebut *swarm*. Dimulai dengan populasi yang diinisialisasi secara acak dan bergerak ke arah yang dipilih secara acak, setiap partikel melewati ruang pencarian dan mengingat posisi terbaik sebelum dirinya dan tetangganya (*neighbors*). Partikel dari suatu kawanan mengkomunikasikan posisi yang baik satu sama lain (*pbest*) serta secara dinamis menyesuaikan posisi dan kecepatan mereka sendiri yang berasal dari posisi terbaik dari semua partikel (*gbest*). Dan pada akhirnya, semua partikel cenderung terbang menuju posisi yang lebih baik dan selama proses pencarian sampai gerombolan bergerak mendekati fungsi *fitness* yang optimal. Persamaan PSO adalah sebagai berikut :

$$v_i^{k+1} = v_i^k + c_1 r_1 (pbest_i^k - x_i^k) + c_2 r_2 (gbest_i^k - x_i^k) \quad (2.12)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2.13)$$

$$v_i^{k+1} = v_i^k + c_1 r_1 (pbest_i^k - x_i^k) + c_2 r_2 (gbest_i^k - x_i^k) \quad (2.14)$$

Dimana :

x_i^k = posisi partikel i untuk iterasi k

v_i^k = kecepatan partikel i untuk iterasi k

c_1, c_2 = koefisien pembobotan

r_1, r_2 = angka acak antara 0 dan 1

Berikut merupakan tahapan dari algoritme PSO :

1. Inisialisasi kawanan (*swarm*) dengan mengalokasikan posisi acak x_i^0 untuk setiap partikel i dari kerumunan yang dibatasi oleh ruang masalah.
2. Evaluasi *fitness* untuk setiap partikel i relatif terhadap posisi saat ini.
3. Bandingkan *fitness* partikel i dengan $pbest_i^{k-1}$, jika *fitness* saat ini lebih besar dari $pbest$, maka tetapkan nilai $pbest$ $pbest_i^k$ ke nilai *fitness* saat ini.
4. Pilih partikel j dengan *fitness* terbaik $pbest_i^k$. Tandai *fitness* $gbest_i^k$ ini.
5. Evaluasi posisi baru x_i^{k+1} partikel i ,
6. Evaluasi kecepatan baru v_i^{k+1} partikel i ,

Beberapa parameter yang sangat mempengaruhi kinerja *PSO*, sebagian parameter diantaranya mempunyai nilai yang berdampak besar dan yang lainnya mempunyai nilai yang berdampak kecil bahkan tidak ada sama sekali. Parameter dasar *PSO* meliputi ukuran kerumunan (*swarm size*) atau jumlah partikel, jumlah iterasi, kecepatan dan koefisien percepatan. Selain itu, *PSO* juga dipengaruhi oleh bobot inersia (*inertia weight*), *velocity clamping*, dan *velocity constriction*.

1. Jumlah Kerumunan (*Swarm Size*)

Parameter jumlah kerumunan (*swarm size*) atau jumlah populasi (*population size*) sangat mempengaruhi performa *PSO*. Semakin banyak partikel maka performa algoritma semakin lambat, terlalu sedikit partikel maka mendorong partikel terjebak dalam optimasi lokal. Namun, ketika dimensi masalah meningkat, maka ukuran kerumunanpun harus ditingkatkan.

Jumlah kerumunan adalah jumlah partikel n dalam gerombolan. Kerumunan besar menghasilkan bagian yang lebih besar dari ruang pencarian yang akan dicakup per

iterasi. Sejumlah besar partikel dapat mengurangi jumlah iterasi yang diperlukan untuk mendapatkan hasil optimasi yang baik. Sebaliknya, sejumlah besar partikel meningkatkan kompleksitas komputasi per iterasi, dan lebih banyak memakan waktu.

2. Jumlah Iterasi

Jumlah iterasi atau *maximum number of generation* merupakan pengulangan suatu proses untuk menghasilkan urutan hasil. Jumlah iterasi yang terlalu rendah dapat menghentikan proses pencarian sebelum waktunya, sementara iterasi yang terlalu besar memiliki konsekuensi dari kompleksitas komputasi tambahan yang tidak perlu dan lebih banyak waktu yang dibutuhkan.

3. Kecepatan (*Velocity*)

Karena nilai kecepatan sangat penting untuk memperbarui kecepatan partikel, ada tiga istilah kecepatan partikel dalam persamaan (4) dan (5):

- a. Istilah v_i^t disebut komponen yang menyediakan memori dari arah pencarian sebelumnya yang berarti pergerakan di masa lalu. Komponen ini direpresentasikan sebagai momentum yang mencegah untuk secara drastis mengubah arah partikel dan bias ke arah saat ini.
- b. Istilah $c_1 r_1 (P_{best} - x_t)$ disebut komponen kognitif yang mengukur kinerja partikel relatif terhadap kinerja sebelumnya. Komponen ini terlihat seperti memori individual dari posisi yang terbaik untuk partikel. Efek komponen kognitif mewakili kecenderungan individu untuk kembali ke posisi yang paling optimal sebelumnya. Komponen kognitif disebut sebagai nostalgia partikel.
- c. Istilah $c_2 r_2 (G_{best} - x_t)$ untuk *PSO* G_{best} disebut komponen sosial yang mengukur kinerja partikel relatif terhadap sekelompok partikel atau *neighbors*. Efek komponen sosial adalah bahwa setiap partikel terbang menuju posisi terbaik yang ditemukan oleh lingkungan partikel.

4. Koefisien Percepatan (*Acceleration coefficients*)

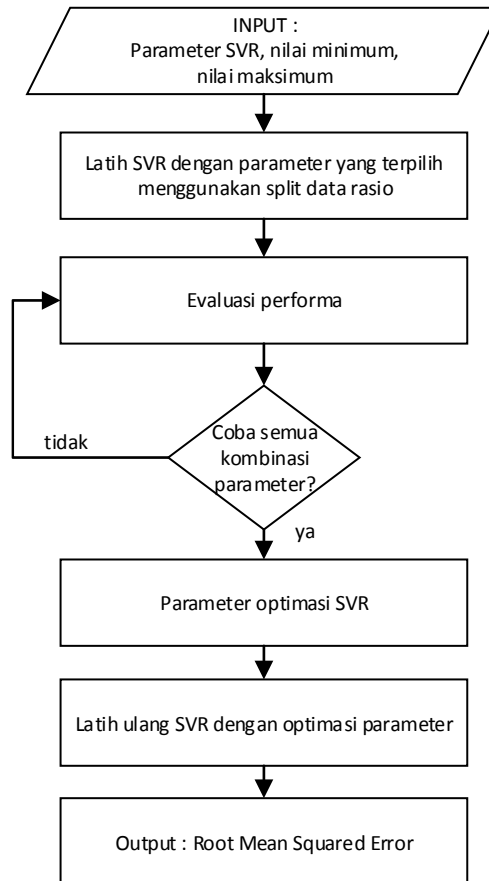
Koefisien percepatan mewakili bobot dari istilah akselerasi stokastik yang menarik partikel ke arah P_{best} dan g_{best} . Jika nilai konstanta ini terlalu tinggi, partikel-partikel bergerak tiba-tiba dan risiko terjebak dalam optima palsu meningkat. Sebaliknya, jika

nilainya terlalu rendah, partikel-partikel bergerak terlalu lambat, upaya komputasi meningkat secara signifikan dan kemungkinan algoritma tidak dapat bertemu.

Koefisien percepatan c_1 dan c_2 , bersama-sama dengan nilai random r_1 dan r_2 , mempertahankan pengaruh stokastik komponen kognitif dan sosial dari kecepatan partikel masing-masing. Konstanta c_1 mengekspresikan seberapa besar keyakinan suatu partikel dalam dirinya sendiri (meningkatkan daya tarik P_{best}), sementara c_2 mengekspresikan seberapa besar keyakinan suatu partikel terhadap *neighbors* (meningkatkan daya tarik G_{best}).

2.7 Optimasi Parameter Pencarian *Grid*

Pencarian *grid* mengoptimalkan parameter SVR (C , ε , γ , dll.) Tujuannya adalah untuk mengidentifikasi kombinasi parameter-hiper yang baik sehingga model dapat memprediksi data yang tidak diketahui secara akurat. Pencarian *grid* yaitu salah satu Algoritme umum yang sering digunakan untuk estimasi parameter, dengan prinsip kerjanya dengan menentukan beberapa nilai parameter pada rentang tertentu, kemudian memilih parameter pada nilai terbaik pada rentang tersebut dan melakukan pencarian berulang pada *grid* (rentang nilai) yang lebih kecil, dst. Fungsi pencarian *grid* mengeluarkan nilai parameter terbaik yang dibutuhkan saat pembentukan model (tahap pelatihan) menggunakan kernel RBF. Parameter untuk kernel RBF adalah cost (c) dan gamma (γ). Selain mengeluarkan nilai parameter terbaik, fungsi ini juga mengeluarkan tingkat kesalahan atau akurasi.

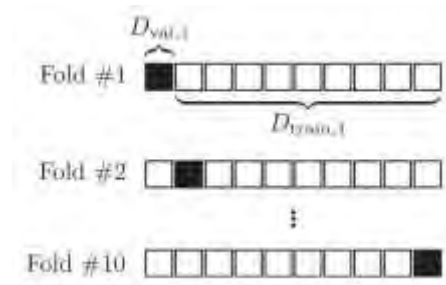


Gambar 2. 4
Parameter SVR menggunakan pencarian *grid* (sumber : Syarif, 2016:1508)

2.8 K-Fold Cross-Validation

Cross-validation adalah salah satu metode resampling data yang paling banyak digunakan untuk memperkirakan kesalahan prediksi sebenarnya dari model dan untuk menyetel parameter model (Berrar, 2019:1). *K-Fold cross-validation* banyak yang menggunakan untuk melatih model. Dalam *k-fold cross-validation*, set pembelajaran yang tersedia dipartisi ke dalam subset k *disjoint* dengan ukuran yang kira-kira sama. *Fold* mengacu pada jumlah himpunan bagian yang dihasilkan. Partisi ini dilakukan dengan cara mengambil sampel secara acak dari perangkat pembelajaran tanpa penggantian. Model dilatih dengan menggunakan himpunan bagian $k - 1$, yang, bersama-sama, mewakili set pelatihan. Kemudian, model diterapkan ke subset yang tersisa, yang dilambangkan sebagai set validasi, dan kinerja diukur. Prosedur ini diulangi sampai masing-masing himpunan bagian k telah berfungsi sebagai set validasi. Secara umum, k bernilai 10 dan

dianjurkan untuk memperkirakan nilai akurasi (jika daya komputasi memungkinkan bisa lebih) karena bias yang relatif rendah dan varians (Han, 2012:370)



Gambar 2. 5
Contoh Pembagian 10-fold cross-validation (sumber : Berrar, 2019:2)

2.9 Penjualan

Aktivitas penjualan merupakan pendapatan utama perusahaan karena jika aktivitas penjualan produk ataupun jasa tidak dikelola dengan baik maka secara langsung dapat merugikan perusahaan. Hal ini bisa disebabkan karena tidak tercapainya sasaran penjualan dan pendapatan pun akan berkurang. Penjualan adalah sebuah taktik yang mengintegrasikan perusahaan, pembeli dan hubungan yang terjadi antara keduanya untuk menciptakan hubungan jangka panjang melalui produk atau jasa perusahaan (Kertajaya, 2006:13).

(Swastha, 2004:403) mendefinisikan:

“Penjualan adalah ilmu dan seni mempengaruhi pribadi yang dilakukan oleh penjual untuk mengajak orang lain agar bersedia membeli barang dan jasa yang ditawarkan.”

Dari kedua definisi tersebut dapat disimpulkan bahwa penjualan merupakan serangkaian aktivitas yang melibatkan penjual dan pembeli untuk melakukan transaksi barang atau jasa, berdasarkan harga yang telah disepakati oleh kedua belah pihak dalam kegiatan tersebut.

Tujuan umum penulain yaitu (Swastha, 2004:404):

1. Mencapai volume penjualan.
2. Mendapatkan laba tertentu.
3. Menunjang pertumbuhan usaha.

Jenis-jenis penjualan dapat dibedakan menjadi 3 (tiga), yaitu:

1. Penjualan tunai

Pengertian penjualan tunai adalah penjualan yang dilakukan ketika terjadinya penyerahan barang dari penjual kepada pembeli atau ketika terjadinya transaksi jual beli.

2. Penjualan kredit

Definisi dari penjualan kredit adalah penjualan yang pembayarannya dilakukan dalam waktu tertentu setelah penyerahan barang yang dijual kepada pembeli.

3. Penjualan konsinyasi

Penjualan Konsinyasi adalah penjualan barang dengan cara menitipkan barang tersebut kepada pihak lain untuk di perjualbelikan bagi kepentingan pihak yang menitipkan barang atau kepentingan kedua belah pihak. Sistem penjualan konsinyasi banyak digunakan dalam perusahaan menimbang banyaknya keuntungan yang didapatkan.

Dalam praktiknya, penjualan dipengaruhi oleh beberapa faktor. Faktor-faktor yang mempengaruhi penjualan (Swastha, 2002:129-131), yaitu:

1. Kondisi dan kemampuan penjual

Keyakinan yang dimiliki oleh penjual merupakan hal yang penting untuk pembeli agar sasaran yang telah ditetapkan dapat tercapai. Keyakinan tersebut berasal dari pemahaman penjual dalam beberapa aspek, yaitu pemahaman dalam jenis karakteristik produk, harga produk, syarat penjualan misalnya pembayaran, pengantaran dan pelayanan jual dan garansi.

2. Kondisi pasar

Kondisi pasar yang menjadi sorotan adalah jenis pasar kelompok pembeli, segmen pasar, daya beli, frekuensi pembelian, keinginan dan kebutuhan

3. Modal

Untuk memperkenalkan barang kepada pembelinya maka penjual memerlukan sarana serta usaha seperti: alat transportasi, tempat peragaan baik di dalam perusahaan maupun di luar perusahaan, usaha promosi, dan sebagainya. Dan semua

ini hanya dapat dilakukan oleh penjual apabila memiliki sejumlah modal yang diperlukan untuk melaksanakan maksud tersebut.

4. Kondisi organisasi penjualan

Biasanya di dalam perusahaan besar masalah penjualan ditangani oleh bagian penjualan yang dipegang orang-orang tertentu/ahli dibidang penjualan. Lain halnya perusahaan kecil, masalah penjualan ditangani oleh orang yang juga melakukan fungsi-fungsi lain.

5. Faktor lain

Faktor-faktor lain, seperti periklanan, peragaan, kampanye pemberian hadiah, sering mempengaruhi penjualan.

2.10 Penelitian Terkait

Terdapat beberapa studi literatur yang mendukung terbentuknya penelitian ini.

Beberapa penelitian terkait disajikan dalam tabel 2.1.

Tabel 2. 1
Penelitian Terkait

No.	Tahun	Nama Peneliti	Judul Penelitian	Metode dan Hasil Penelitian
1	2018	M.R. Rifqi, Budi Darma, S, Fitra A. Bachtiar	<i>Support Vector Regression</i> Untuk Peramalan Permintaan Darah: Studi Kasus Unit Transfusi Darah Cabang – PMI Kota Malang	Metode yang digunakan penelitian ini adalah SVR dan kernel RBF. Data yang digunakan merupakan jumlah permintaan darah dari tahun 2010 sampai tahun 2015, yang dimana data tahun 2010 sampai tahun 2014 digunakan sebagai data <i>training</i> , sedangkan data tahun 2015 digunakan sebagai data <i>testing</i> . Hasil dari penelitian tersebut adalah algoritme SVR dapat dikategorikan baik karena menghasilkan nilai minimum untuk test akurasi <i>Mean Absolute Percentage Error</i> (MAPE) yakni 3,89%. Namun nilai parameter sangat mempengaruhi kinerja SVR. Jika parameter sigma dan lambda bernilai terlalu besar, maka akan rawan

No.	Tahun	Nama Peneliti	Judul Penelitian	Metode dan Hasil Penelitian
				menghasilkan nilai MAPE yang tinggi
2	2015	Nasimul Hasan, Nayan C. Nath, Risul I. Israel	<i>A Support Vector Regression Model for Forecasting Rainfall</i>	Metode yang digunakan penelitian ini adalah SVR dan data time series. Data yang digunakan adalah data <i>time series</i> curah hujan selama 6 tahun (tahun 2008 sampai tahun 2014) digunakan. Hasil model yang diusulkan dapat menghasilkan prediksi maksimal 99,92%.
3	2010	Chi-Jie Lu, Tian-Shyug Lee, Chia-Mei Lian	<i>Sales Forecasting of IT Products using A Hybrid Multivariate Adaptive Regression Splines(MARS) and SVR Model</i>	Metode yang digunakan adalah kombinasi dari <i>Multivariate Adaptive Regression Splines</i> (MARS) dan SVR. Data yang dijadikan bahan penelitian adalah data penjualan mingguan dari tiga produk TI, yaitu <i>Note Book</i> , monitor LCD dan <i>Mother Board</i> dari bulan januari 2006 hingga agustus 2009. Hasil penelitian adalah model MARS-SVR dapat menjadi alternatif yang baik untuk peramalan dengan kesalahan peramalan paling rendah dibanding dengan kinerja MARS dan SVR tunggal yakni rata-rata 0.07%.
4	2013	A. Kridanto, J. L. Buliali,	<i>Metode Hibrida FCM dan PSO-SVR Untuk Prediksi Data Arus Lalu Lintas</i>	Metode yang digunakan dipenelitian ini adalah <i>Fuzzy C-Means</i> (FCM), SVR dan PSO. Dataset berisi data lalu lintas dalam kurun waktu 1 bulan dengan interval waktu pengamatan 15 menit sehingga terdapat 96 data pengamatan dalam 1 hari. Hasil penelitian adalah Metode FCM-PSO-SVR mampu memberikan performa prediksi dengan kesalahan di bawah 4% untuk semua lokasi yang diuji.
5	2011	Xiaoyong Lu, Xiaomeng Geng.	<i>Car Sales Volume Prediction Based on Particle Swarm Optimization Algorithm and Support Vector Regression</i>	Metode yang digunakan penelitian PSO, <i>Genetic Algorithm</i> (GA) dan SVR. Data yang digunakan adalah data penjualan mobil periode 2003-2008 di taiwan. Hasil penelitian adalah akurasi model PSO-SVR

No.	Tahun	Nama Peneliti	Judul Penelitian	Metode dan Hasil Penelitian
				lebih unggul dibandingkan dengan GA-VSR.

Berdasarkan penjelasan penelitian-penelitian terkait, penelitian ini bermaksud melakukan kombinasi antara algoritme PSO dengan SVR untuk kasus penjualan bisnis retail. Algoritme SVR dipilih karena mampu meramalkan dengan baik. Sedangkan Algoritme PSO dipilih untuk optimasi atribut agar ketepatan/akurasi peramalan meningkat.

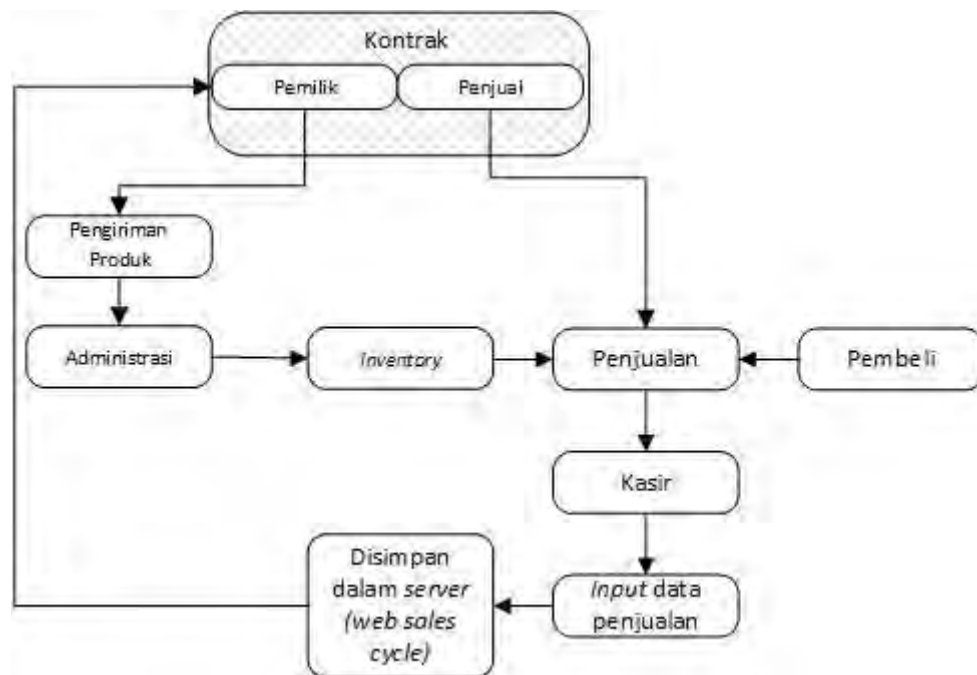
BAB III

OBJEK DAN METODOLOGI PENELITIAN

3.1 Penjualan

Pada penelitian ini dilakukan pada data penjualan perusahaan *brand* ZXC. *Brand* ZXC merupakan salah satu sub *brand* di dalam naungan PT. ZYX yang merupakan perusahaan manufaktur dalam bidang bisnis retail pakaian yang didirikan pada tahun 1970. Sedangkan *brand* ZXC sendiri mulai berdiri pada tahun 2009.

Gambar 3.1 merupakan alur atau sistem dari penjualan *brand* ZXC. Sistem penjualan *brand* ZXC merupakan suatu konsep di mana pihak pemilik produk mempunyai perjanjian dengan pihak lain untuk menjualkan produknya dengan perhitungan komisi tertentu. Konsep tersebut merupakan salah satu cara untuk lebih memperluas jangkauan pasar yang sudah mapan dan keuntungan dapat meningkat karena harga jual masih dikendalikan oleh pihak pemilik.



Gambar 3. 1
Konsep Sistem Penjualan

Data hasil penjualan di *input* ke sistem *Sales Cycle* yang akan disimpan secara otomatis pada *server* perusahaan. Sistem *sales Cycle* merupakan *web* perusahaan yang berisikan informasi seputar aktivitas perusahaan yang disimpan dalam bentuk data, meliputi data produk, data pelanggan/toko, target *marketing brand*, target *marketing* operasional, perintah pengeluaran produk (PPP), transaksi penjualan, *monitoring* stok, dan data gudang jadi. Berikut merupakan tampilan dari web *sales cycle* yang dapat dilihat pada gambar 3.2



Gambar 3. 2
Tampilan *Sales Cycle*

Dalam penggunaannya, data penjualan dijadikan bahan membuat peramalan untuk dapat menentukan target *marketing* sekaligus sebagai bahan tim produksi dalam merencanakan *inventory* yang sesuai dengan kebutuhan. Metode peramalan sederhana merupakan metode peramalan yang diterapkan selama ini oleh perusahaan untuk menganalisis data yang bersifat kuantitatif. Metode tersebut digunakan untuk merata-ratakan data penjualan tiga bulan terakhir, dikombinasikan dengan metode kualitatif yang berdasarkan intuisi, pengetahuan dan keahlian individu dalam meramalkan kebutuhan barang. Analisis dilakukan untuk 6 bulan terakhir dan satu tahun terakhir, tidak hanya data penjualan tiga bulan terakhir. Namun metode tersebut tidak dapat dipakai ketika menghadapi data yang bersifat non-linier, seperti demografi *customer* atau tipe toko di berbagai wilayah tertentu.

3.2 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah dasar metodologi CRISP-DM yang telah dimodifikasi berupa keterangan tambahan yang sesuai dengan kebutuhan penelitian. Berikut desain metodologi penelitian yang digambarkan pada gambar 3.3. Gambar 3.3 dapat dijelaskan dalam beberapa poin sebagai berikut :

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)

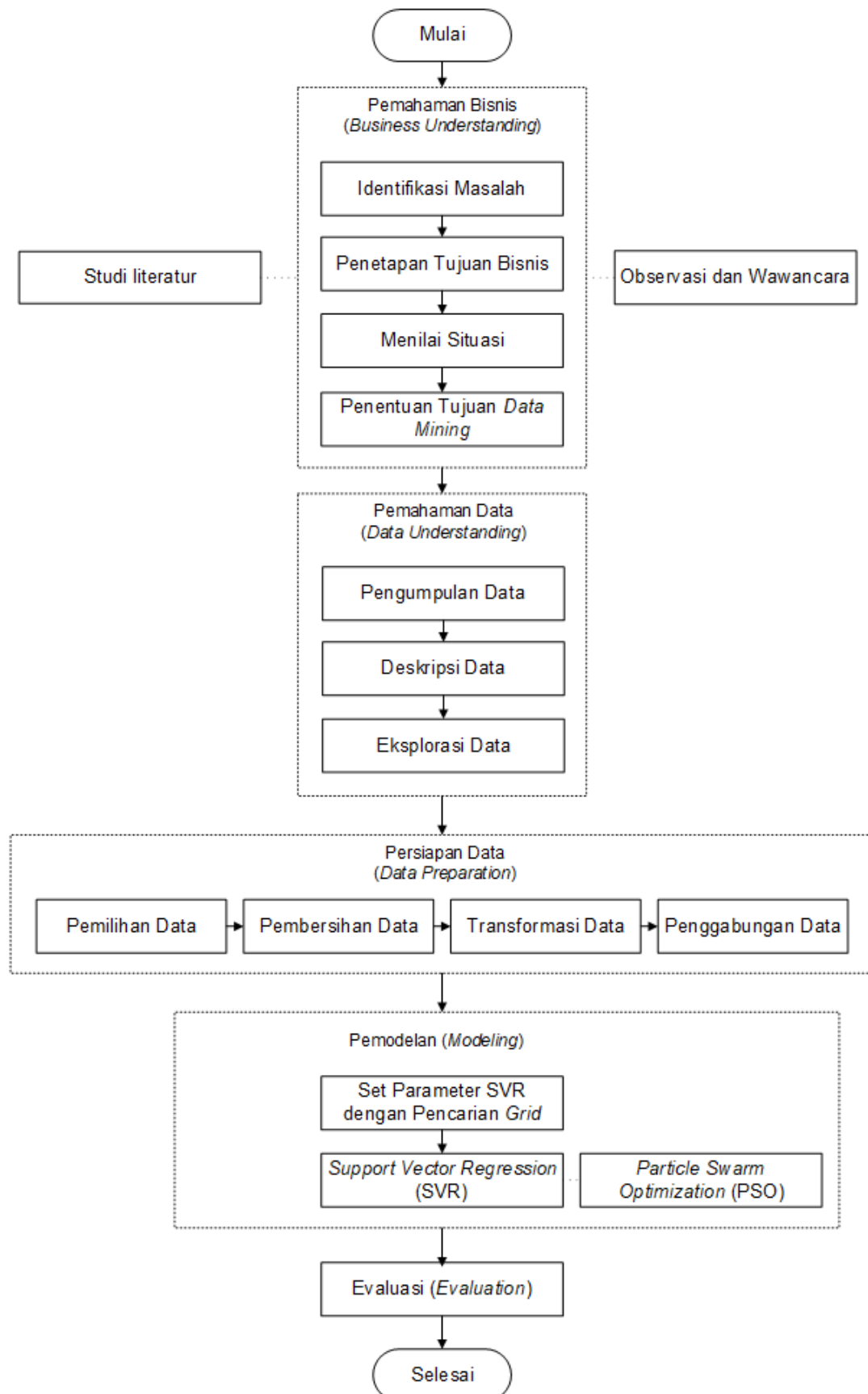
Tahapan pemahaman bisnis berfokus pada identifikasi masalah dan pemahaman tujuan dari perspektif bisnis. Tahap ini dilakukan observasi langsung dan wawancara kepada pihak *stackholder* untuk mendapatkan informasi mengenai perusahaan. Tahap ini penting dilakukan guna mengetahui kebutuhan perusahaan dan menyelaraskannya pada ilmu *data mining*.

2. Fase Pemahaman Data (*Data Understanding Phase*)

Fase ini dimulai dengan mengumpulkan data. Apabila data berasal lebih dari satu sumber data, maka dilakukan proses integrasi. Untuk *brand ZXC* pengumpulan data dilakukan dengan mengakses *Sales Cycle* yang merupakan *web* perusahaan. Setelah pengumpulan data selesai, lakukan eksplorasi data sebagai upaya untuk membiasakan data dan mengetahui secara lebih mendalam tentang tipe data, struktur atribut serta mengkonfirmasi apakah data terdistribusi seperti yang diharapkan, atau apakah terjadi penyimpangan tak terduga yang perlu ditangani pada tahap selanjutnya.

3. Fase Persiapan Data (*Data Preparation Phase*)

Setelah pengumpulan data selesai, dilakukan persiapan data dimulai dari seleksi data, pembersihan data, transformasi data dan integrasi data. Seleksi data dilakukan untuk memilih data yang terlihat relevan dengan objek penelitian. Selanjutnya pembersihan data dilakukan guna membersihkan data dari *noise*, data yang hilang atau data yang tidak sesuai dengan format awal. Setelah data diseleksi dan dibersihkan, data ditransformasikan dan diintegrasikan ke dalam bentuk baru untuk diproses pada tahapan selanjutnya.



Gambar 3. 3
Metodologi Penelitian

4. Fase Pemodelan (*Modeling Phase*)

Pada fase ini mulai dibentuk pemodelan dengan algoritme *Support Vector Regression* (SVR). Lakukan pengujian terhadap dataset dengan pengoptimalan parameter pencarian *grid* untuk mendapatkan parameter yang optimal. Kemudian dilakukan prosesi pemodelan data dengan teknik *k-fold cross validation* untuk menguji akurasi pemodelan secara langsung. Terakhir integrasikan algoritme Particle Swarm Optimization (PSO) untuk optimasi atribut agar akurasi peramalan dapat maksimal.

5. Fase Evaluasi (*Evaluation Phase*)

Selanjutnya dilakukan evaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk menetapkan apakah model yang digunakan telah relevan dengan tujuan bisnis di fase awal. Selanjutnya menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.

BAB IV

PEMBAHASAN DAN HASIL EKSPERIMEN

Bab ini berisikan pembahasan tentang analisis-analisis yang dilakukan dalam penelitian dengan metodologi CRISP-DM yang telah dibahas pada bab sebelumnya dan telah disesuaikan dengan kebutuhan penelitian. Metodologi ini mempunyai tahapan pemahaman bisnis (*business understanding*), pemahaman data dan persiapan data (*data understanding and preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan terakhir dokumentasi.

4.1 Pemahaman Bisnis (*business understanding*)

Berikut beberapa tugas dari fase pemahaman bisnis CRISP-DM ini, diantaranya :

1. Identifikasi Tujuan Bisnis (*Determine Business Objective*)

Perusahaan *brand* ZXC secara garis besar memiliki tujuan dalam proses bisnisnya yaitu mengembangkan berbagai model dan kerangka kerja untuk membantu dalam pengambilan keputusan strategis dalam konteks lingkungan yang kompleks dan dinamika persaingan. Salah satu manifestasi dari tugas tersebut adalah dengan melakukan peramalan untuk membantu pihak manajer dalam menentukan target dan meningkatkan penjualan.

2. Menilai Situasi (*Assess the Situation*)

Tugas dari penilaian situasi yaitu untuk merinci semua fakta yang ada kebutuhan, kendala, asumsi dan faktor yang harus dipertimbangkan dalam proses *Data Mining*.

Berikut beberapa rincian dari fase menilai situasi :

- a. Data yang digunakan sudah memiliki izin dari pihak yang terkait
- b. Pihak perusahaan *brand* ZXC hanya membutuhkan data peramalan penjualan dengan hasil perbulan.
- c. Risiko dan kemungkinan yang mungkin terjadi pada penelitian ini adalah akurasi hasil dari modeling yang akan dibentuk. Dikarenakan algoritma SVR sangat ditentukan oleh parameter yang sesuai terjadi kemungkinan risiko kualitas data prediksi yang dihasilkan.

3. Menentukan Tujuan Data Mining (*Determine the Data Mining Goals*)

Setelah mengetahui tujuan bisnis dan situasi yang ada di lingkungan perusahaan *brand ZXC*, maka tujuan *data mining* adalah membuat model perhitungan yang dapat meramalkan data yang bersifat non-linier dengan algoritme PSO-SVR untuk menggali pengetahuan dan pola tentang:

- a. Peramalan Kuantitas Penjualan (*Quantity Sales Forecasting*), untuk memberikan masukan tentang penentuan target operasional/*marketing* perusahaan, mengalokasikan sumber daya secara efektif dan efisien sekaligus sebagai bahan perencanaan untuk tim produksi dalam menentukan jumlah *inventory* agar sesuai dengan kebutuhan tim *marketing*.
- b. Menentukan pemodelan yang sesuai dalam mengoptimalkan hasil dari peramalan.

4.2 Pemahaman Data (*data understanding*)

1. Mengumpulkan Data Awal (*Collect the Data*)

Data yang digunakan dalam penelitian ini yaitu data penjualan periode tahun 2016 sampai 2018 yang didapat dengan mengakses sistem *sales cycle*. Selain data penjualan *sales cycle*, master toko dibutuhkan untuk nantinya dapat diintegrasikan berdasarkan kebutuhan penelitian. Data penjualan yang akan digunakan berupa *file excel* (format *.xls).

2. Mendeskripsikan Data (*Describe the Data*)

Data penjualan dari *brand ZXC* selama tiga periode memiliki jumlah 845111 *record* dengan 41 atribut. 9 atribut menerangkan transaksi seperti kode transaksi, waktu transaksi, harga yang didapatkan perusahaan setelah potongan acara, *user input* dan keterangan. 2 atribut menerangkan profil toko. 6 atribut menerangkan produk. dan 24 atribut menerangkan jumlah penjualan (pcs) berikut detail ukuran (*size* yang terjual). Masing-masing atribut dijelaskan dalam tabel 4.1.

Tabel 4. 1
Penjelasan atribut pada data penjualan

No	Atribut	Tipe Data	Keterangan
1	BPK	Nominal	Bukti penjualan konsinyasi
2	TG_BPK	Date	Tanggal terjadinya transaksi
3	PWKL	Nominal	Kode perwakilan
4	CUST	Nominal	Kode toko
5	KET	Nominal	Keterangan tambahan seputar acara dan penjualan
6	INUSER	Nominal	ID user untuk mengakses <i>web sales cycle</i>
7	INTIME	Date	Tgl input data transaksi penjualan pada <i>web sales cycle</i> .
8	IDBARCODE	Numeric	Kode barcode
9	NO_KW	Nominal	ID produk
10	WARNA	Nominal	ID warna produk
11	MERK	Nominal	ID brand
12	PRODUK	Nominal	ID produk global
13	HGJ	Numeric	Harga jual toko
14	QTOTAL	Numeric	Jumlah produk (no_kw) yang terjual
15	SZ24	Numeric	Ukuran produk (no_kw) terkecil yang terjual
16	SZ25	Numeric	Ukuran produk (no_kw) yang terjual
17	SZ26	Numeric	Ukuran produk (no_kw) yang terjual
18	SZ27	Numeric	Ukuran produk (no_kw) yang terjual
19	SZ28	Numeric	Ukuran produk (no_kw) yang terjual
20	SZ29	Numeric	Ukuran produk (no_kw) yang terjual
21	SZ30	Numeric	Ukuran produk (no_kw) yang terjual
22	SZ31	Numeric	Ukuran produk (no_kw) yang terjual
23	SZ32	Numeric	Ukuran produk (no_kw) yang terjual
24	SZ33	Numeric	Ukuran produk (no_kw) yang terjual
25	SZ34	Numeric	Ukuran produk (no_kw) yang terjual
26	SZ35	Numeric	Ukuran produk (no_kw) yang terjual
27	SZ36	Numeric	Ukuran produk (no_kw) yang terjual
28	SZ37	Numeric	Ukuran produk (no_kw) yang terjual
29	SZ38	Numeric	Ukuran produk (no_kw) yang terjual
30	SZ39	Numeric	Ukuran produk (no_kw) yang terjual
31	SZ40	Numeric	Ukuran produk (no_kw) yang terjual
32	SZ41	Numeric	Ukuran produk (no_kw) yang terjual
33	SZ42	Numeric	Ukuran produk (no_kw) yang terjual
34	SZ43	Numeric	Ukuran produk (no_kw) yang terjual
35	SZ44	Numeric	Ukuran produk (no_kw) yang terjual
36	SZ45	Numeric	Ukuran produk (no_kw) yang terjual
37	SZ46	Numeric	Ukuran produk (no_kw) terbesar yang terjual
38	TOTBRUTO	Numeric	Total harga kotor produk pertransaksi
39	POTONGAN	Numeric	Potongan margin toko
40	HRGKASIR	Numeric	Total harga yang dibayar dikasir
41	TOTNETTO	Numeric	Total harga bersih yang diterima perusahaan

Selain data penjualan, master toko dibutuhkan untuk memenuhi tujuan perusahaan dan tujuan penelitian. Master toko dibutuhkan untuk mengetahui wilayah, tipe toko dan

tipe pelanggan untuk kepentingan demografi. Berikut penjelasan masing-masing atribut dalam ketiga master tersebut.

Tabel 4. 2
Penjelasan atribut pada data master toko

No	Atribut	Tipe Data	Keterangan
1	CUST	Nominal	ID toko yang akan dihubungkan dengan data penjualan
2	REGION	Nominal	Tempat toko berdasarkan wilayah.
3	STORETYPE		Tipe toko berdasarkan volume jual toko. Dikelompokkan menjadi 5 kelompok, yaitu : A+, A, B, C, D
4	CUSTTYPE	Nominal	Tipe <i>customer</i> berdasarkan hasil observasi dan kuisisioner perusahaan yang kemudian dikelompokkan menjadi 3 kelompok, yaitu : Basic, Fashion dan keduanya (Fas/Bas).

Semua data tersebut akan diintegrasikan dan disesuaikan dengan kebutuhan penelitian yang akan dijelaskan pada fase persiapan data (*data preparation*).

3. Mengeksplorasi Data (*Explore Data*)

Data penjualan perusahaan *brand* ZXC terdiri dari data nominal dan data numerik. Data nominal memungkinkan untuk memberikan gambaran secara umum. Namun data numerik memberikan wawasan lebih banyak tentang hasil dan demografi penjualan pada periode 2016-2018. Beberapa temuan dasar dari eksplorasi data dapat dibagi menjadi 2, yaitu analisis statistik dan visualisasi yang dijelaskan sebagai berikut:

a. Analisis statistik

Tujuan dari analisis statistik adalah untuk mendeskripsikan dan menerangkan data yang telah dikumpulkan tanpa adanya kesimpulan secara umum. Beberapa analisis deskriptif untuk masing-masing atribut dijelaskan pada tabel 4.3 dan tabel 4.4.

Tabel 4. 3
Analisis statistik pada atribut penjualan tipe data nominal

No	Atribut	Missing Value	Distinct	Most Value	Least Value
1	BPK	0	300679	2DC160410548	2IM181200992
2	PWKL	0	14	DB	IK
3	CUST	0	341	06.024	30.281
4	KET	560295	2721	web-(263446)	webpot 100 ribu
5	INUSER	616690	111	0026	W078
6	NO_KW	0	2477	ECI6005F	LKR6190D

No	Atribut	Missing Value	Distinct	Most Value	Least Value
7	WARNA	624384	116	43	1
8	MERK	0	1	G	G
9	PRODUK	0	4	BC	AJ

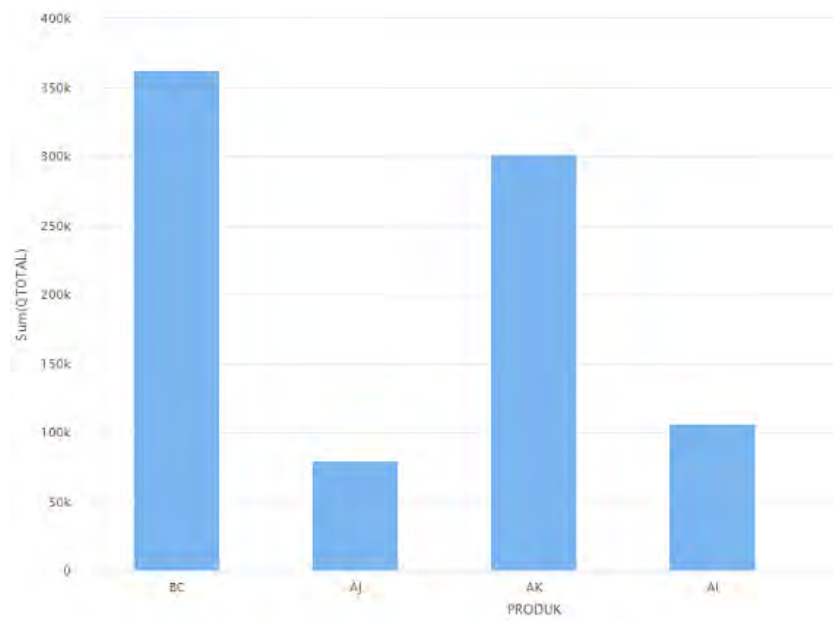
Tabel 4. 4
Analisis statistik pada atribut penjualan tipe data numerik

No	Atribut	Missing Value	Min	Max	Average
1	HGJ	0	1	45	274458.977
2	IDBARCODE	0	6347	121655	92650.333
3	QTOTAL	0	1	45	1.149
4	SZ24	0	0	19	0.183
5	SZ25	0	0	19	0.342
6	SZ26	0	0	14	0.327
7	SZ27	0	0	20	0.194
8	SZ28	0	0	6	0.016
9	SZ29	0	0	20	0.009
10	SZ30	0	0	4	0.009
11	SZ31	0	0	20	0.014
12	SZ32	0	0	19	0.014
13	SZ33	0	0	6	0.013
14	SZ34	0	0	5	0.013
15	SZ35	0	0	5	0.007
16	SZ36	0	0	14	0.008
17	SZ37	0	0	1	0.000
18	SZ38	0	0	2	0.000
19	SZ39	0	0	0	0
20	SZ40	0	0	0	0
21	SZ41	0	0	0	0
22	SZ42	0	0	0	0
23	SZ43	0	0	0	0
24	SZ44	0	0	0	0
25	SZ45	0	0	0	0
26	SZ46	0	0	0	0
27	TOTBRUTO	226900	1	8158300	319997.428
28	POTONGAN	226927	0	3850000	11122.962
29	HRGKASIR	226927	1	5710810	246382.252
30	TOTNETTO	226900	1	6000000	182663.166

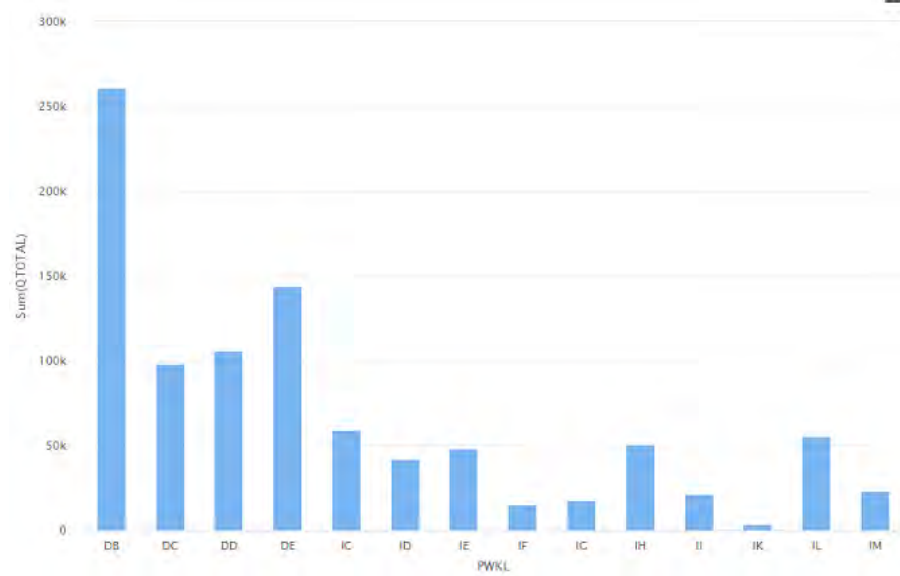
Untuk atribut TGL_BPK dan atribut INTIME yang merupakan tipe data *date*, keduanya memiliki nilai tanggal terlama 1-1-2016 dan tanggal terlama 31-12-2018.

b. Visualisasi

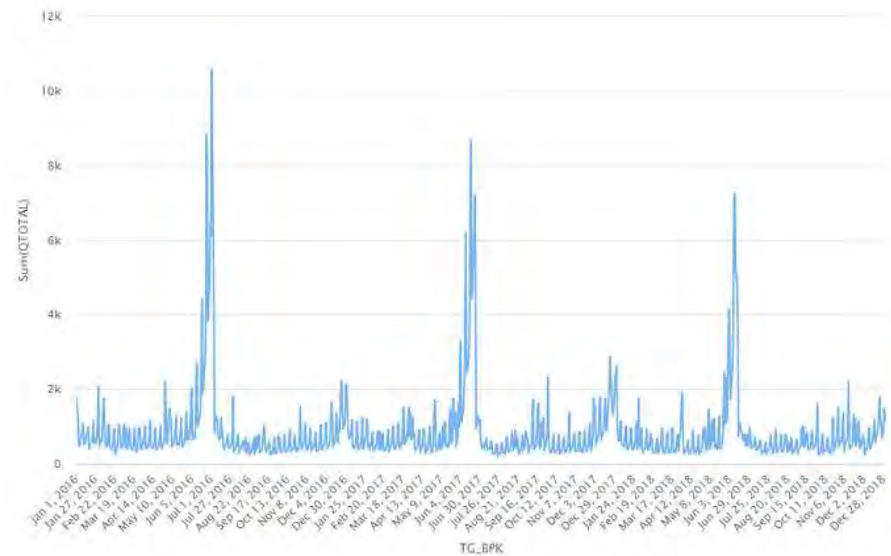
Visualisasi data dilakukan terhadap beberapa atribut yang mempunyai kemungkinan dalam penjualan. Berikut merupakan data visualisasi atribut yang didapat, yaitu:



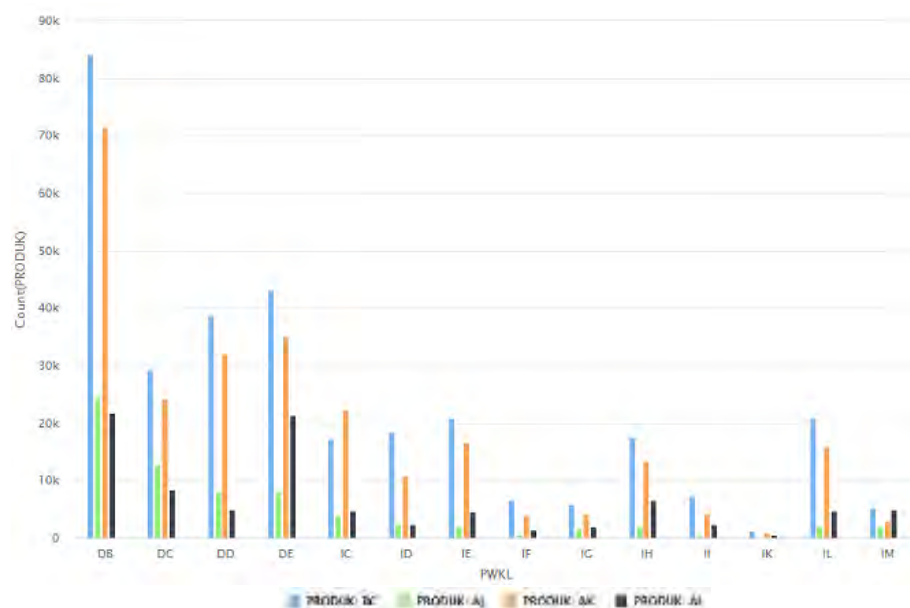
Gambar 4. 1
Bar plot atribut PRODUK terhadap atribut QTOTAL



Gambar 4. 2
Bar plot atribut PWKL terhadap atribut QTOTAL



Gambar 4. 3
Bar plot atribut TG_BPK terhadap atribut QTOTAL



Gambar 4. 4
Bar plot sebaran data atribut PWKL terhadap atribut PRODUK

Berdasarkan gambar 4.1 dari atribut PRODUK terhadap atribut QTOTAL menunjukkan bahwa penjualan tertinggi terdapat pada produk BC dan penjualan terendah ada pada produk AJ. gambar 4.2 menunjukkan atribut PWKL terhadap QTOTAL dapat dilihat pwkl tertinggi ada pada DB dan yang terendah ada pada pwkl IK. Gambar 4.3 menunjukkan fluktuasi yang terus berubah tiap bulannya. Dari gambar 4.3 terlihat lonjakan penjualan

tertinggi terdapat pada antara bulan juni dan bulan juli. Selanjutnya berdasarkan gambar 4.4 menunjukkan atribut PWKL terhadap atribut PRODUK dengan produk penjualan tertinggi yaitu BC berada pada pwkl DB.

4.3 Persiapan Data (*data preparation*)

1. Pemilihan Data

Dalam tinjauan data penjualan dengan pakar perusahaan, disepakati bahwa atribut yang usang dan atribut dengan jumlah nilai hilang sangat tinggi dapat dibuang. 36 atribut dalam data penjualan diidentifikasi tidak terlalu relevan dengan hasil akhir penelitian, dan karena itu dikeluarkan dari pemrosesan lebih lanjut. 5 atribut yang tersisa (1 atribut numerik, dan 4 atribut nominal). Untuk atribut dari hasil pemilihan data adalah sebagai berikut :

Tabel 4. 5
Pemilihan data

No	Atribut	Tipe Data	Keterangan data
1	TG BPK	Date	Data penjualan
2	PWKL	Nominal	Data penjualan
3	CUST	Nominal	Data penjualan
4	NO_KW	Nominal	Data penjualan
5	QTOTAL	Numeric	Data penjualan

2. Pembersihan Data

Pemeriksaan yang lebih rinci menghasilkan penemuan sejumlah besar nilai yang hilang dan nilai yang salah seperti nol, spasi , #N/A, dll. nilai tersebut. Seperti yang telah dilampirkan pada tabel 4.3 dan 4.4, di antara 5 atribut tersebut tidak ada data yang hilang ataupun data yang salah.

3. Transformasi Data

Tahap transformasi dilakukan untuk mengubah sejumlah nilai atau atribut yang disesuaikan dengan kebutuhan penelitian. Data ditransformasi menjadi data perbulan dan pertahun, diikuti dengan atribut-atribut penunjang lainnya seperti master toko untuk mendapatkan data demografi yang merupakan salah satu faktor besar dalam mempengaruhi penjualan. Selanjutnya, transformasi dilakukan pada nilai atribut dari

nominal ke numerik untuk pemrosesan modeling lebih lanjut. Seperti yang disarankan oleh penelitian sebelumnya (Yu, 2013:1058), dilakukannya transformasi data dengan mengubah atribut nominal PWKL dan STORETYPE antara 0 dan 1, dan atribut nominal CUSTTYPE dengan angka skala 0-1, terkecuali atribut tipe numerik.

4. Penggabungan Data

Tahap selanjutnya setelah data ditransformasikan, data tersebut dianalisis dan diproses pada proses selanjutnya. Untuk hasil dari penggabungan data dan sekaligus data yang akan digunakan dalam penelitian ini ditampilkan pada tabel 4.6.

Tabel 4. 6
Penggabungan data

No	Atribut	Tipe Data	Keterangan	Value
1	YEARMONTH	Nominal	Penjualan dalam satuan waktu tahun dan bulan	
2	PWKL	Nominal	Tempat toko berdasarkan wilayah	0 dan 1
3	STORETYPE	Nominal	Tipe toko berdasarkan volume jual toko.	0 dan 1
4	CUSTTYPE	Nominal	Tipe <i>customer</i> berdasarkan hasil observasi dan kuisisioner perusahaan	0-1
5	AI	Numeric	Kode produk <i>t-shirt</i>	
6	AJ	Numeric	Kode produk <i>jacket</i>	
7	AK	Numeric	Kode produk kemeja	
8	BC	Numeric	Kode produk celana denim	
9	QTOTAL	Numeric	Total penjualan	

4.4 Pemodelan (*modelling*)

Tahap pemodelan merupakan tahap keempat dari kerangka kerja CRISP-DM. Tahap ini memiliki tugas untuk pemilihan teknik pemodelan, pembuatan model, serta pengujian model. Pemodelan menggunakan *tools Rapidminer 9.5* dan untuk pengujian akurasi digunakan perhitungan *Root Mean Squared Error (RMSE)*.

1. Teknik pemodelan

Data dipilih berdasarkan hasil pencarian yang relevan antara objek penelitian dengan penelitian sebelumnya yang terkait. Dipilih algoritma *Support Vector Regression* sebagai algoritma peramalan dengan teknik regresi. Kinerja SVR sangat dipengaruhi oleh pengaturan parameter *complement (C)* sebagai batas nilai penalti toleransi kesalahan, nilai *epsilon (ε)* yang mengatur batas kesalahan fungsi dan nilai *gamma (γ)*

yang menyatakan lebar dari fungsi kernel RBF. Tidak ada aturan umum dalam pemilihan parameter-parameter tersebut. Penentuan nilai biasanya berdasarkan metode *try-error* (atau disebut *cross validation*). Di sisi lain, metode berdasarkan keahlian sangat beresiko untuk pengguna yang tidak ahli. Untuk mengurangi kemungkinan kelemahan tersebut, metode pemilihan parameter diusulkan dalam penelitian ini. Metode pencarian *grid* merupakan salah satu metode yang cukup baik dalam menemukan kombinasi nilai parameter terbaik (Syarif, 2016:1508).

2. Analisis pengujian model

Sebelum memasuki tahap pengujian model, dalam kinerjanya, SVR sangat tergantung pada parameter yang diberikan. Maka dari itu, terlebih dahulu dilakukan pengujian untuk menentukan parameter yang sesuai. Pengujian tersebut diantaranya menentukan nilai C , nilai ε dan nilai γ . Untuk setiap pengujian diulang sebanyak 10 kali dengan menggunakan pencarian *grid* dan *tenfold cross validation* untuk mendapatkan parameter dengan akurasi yang baik. Dalam pengujian digunakan parameter awal sebagai berikut (Rusmalawati, 2018:1835) :

C : 1-200

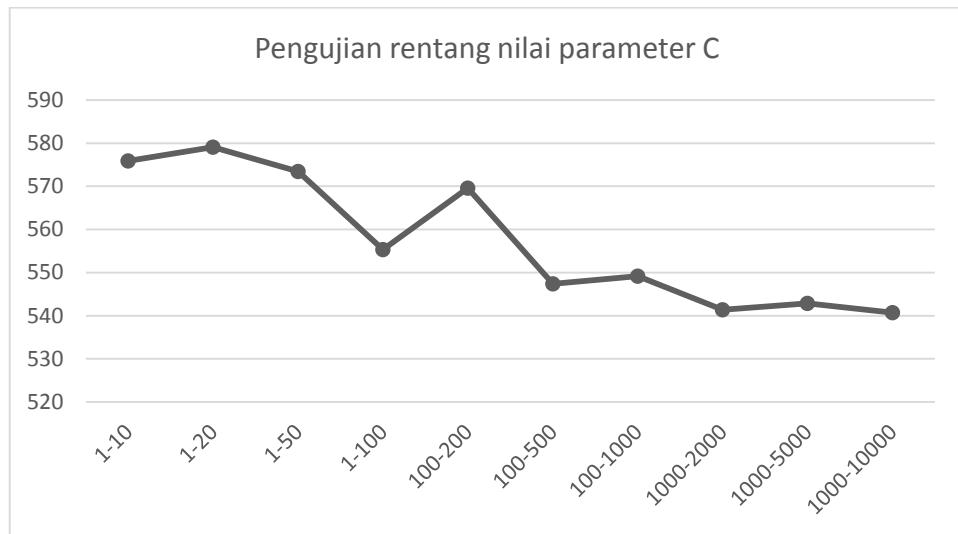
ε : 0,00001-0,1

γ : 0,1-1

a. Pengujian rentang pencarian nilai parameter *complement* (C)

Parameter C merupakan penalti toleransi yang diberikan terhadap kesalahan peramalan. Jika nilai C semakin tinggi, artinya model peramalan yang dibentuk semakin tidak mentoleransi kesalahan sehingga dapat disimpulkan nilai tersebut memberikan peramalan yang baik (Rifqi, 2018:3338). Dapat dilihat pada grafik pengujian pada gambar 4.5, rentang nilai 1000-2000 mulai terjadi konvergensi terhadap rentang nilai yang pada akhirnya rentang nilai 1000-10000 dipilih sebagai rentang terbaik untuk parameter C karena menghasilkan nilai tingkat kesalahan atau RMSE terkecil. Pernyataan dari penelitian terkait sebelumnya tentang semakin besar nilai C maka dapat disimpulkan nilai tersebut memberikan

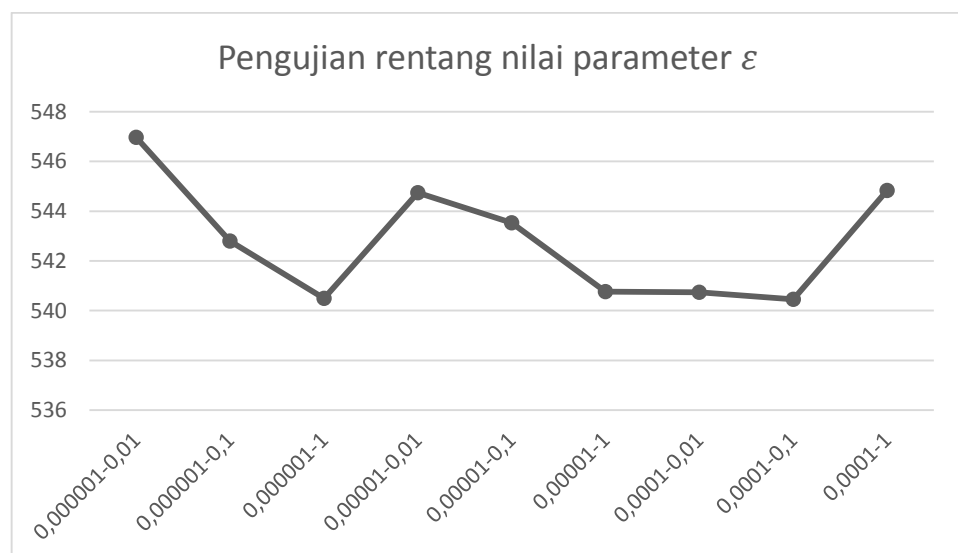
peramalan yang baik dapat dibuktikan dengan nilai dari hasil pengujian yang mempunyai RMSE terendah pada gambar 4.5.



Gambar 4. 5
Grafik pengujian rentang pencarian nilai parameter C

b. Pengujian rentang pencarian nilai parameter ϵ (ϵ)

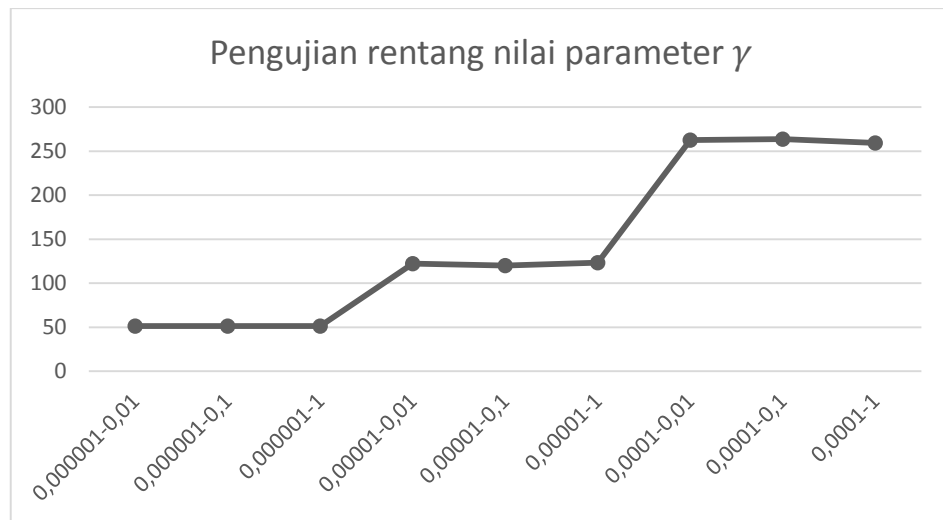
Parameter ϵ mempunyai fungsi untuk mengatur batas kesalahan dari nilai fungsi. Jika nilai fungsi melebihi wilayah *error* maka akan dikenai pinalti sebesar nilai parameter C . Pada gambar 4.6 merupakan hasil dari pengujian yang didapatkan RMSE terendah ada pada rentang 0.0001-0.1.



Gambar 4. 6
Grafik pengujian rentang pencarian nilai parameter ϵ

c. Pengujian rentang pencarian nilai parameter γ (γ)

Parameter γ mempunyai fungsi sebagai jarak maksimum yang disesuaikan untuk mendeteksi seberapa mirip dua data input yang dianggap serupa. Berdasarkan gambar 4.7 rentang parameter γ yang mempunyai nilai RMSE rendah adalah 0.000001-0.01.



Gambar 4. 7
Grafik pengujian rentang nilai parameter γ

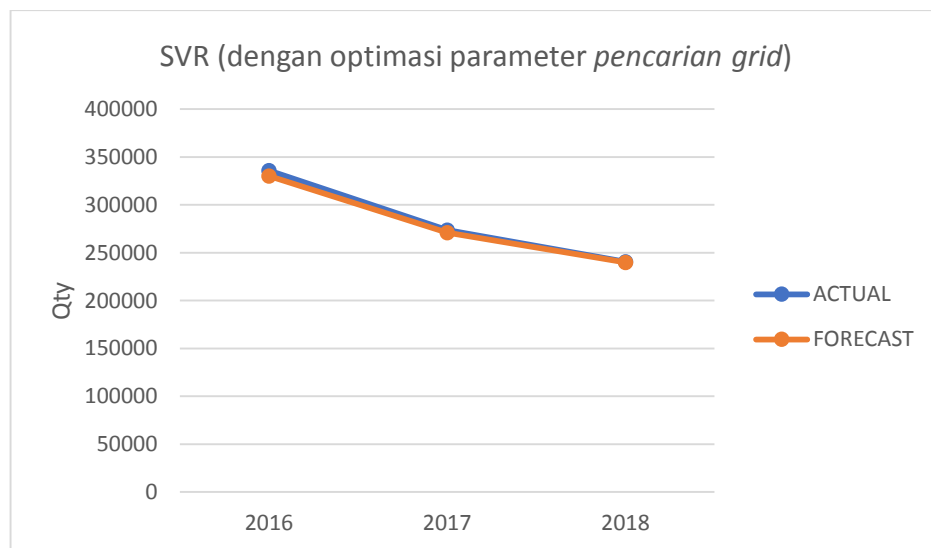
Dari beberapa hasil pengujian parameter maka dapat ditetapkan rentang nilai parameter yang akan digunakan dalam tahap selanjutnya adalah $C = 1000-10000$, $\varepsilon = 0.0001-0.1$ dan $\gamma = 0.000001-0.01$. Rentang nilai tersebut akan diuji kembali menggunakan *dataset* yang sama dan metode yang sama, yaitu pencarian *grid* dan *tenfold cross-validation*. Maka hasil yang didapat untuk nilai parameter terbaik dari pencarian *grid* adalah sebagai berikut :

Tabel 4. 7
Nilai parameter SVR

C	ε	γ
7300	0.02008	0.000001

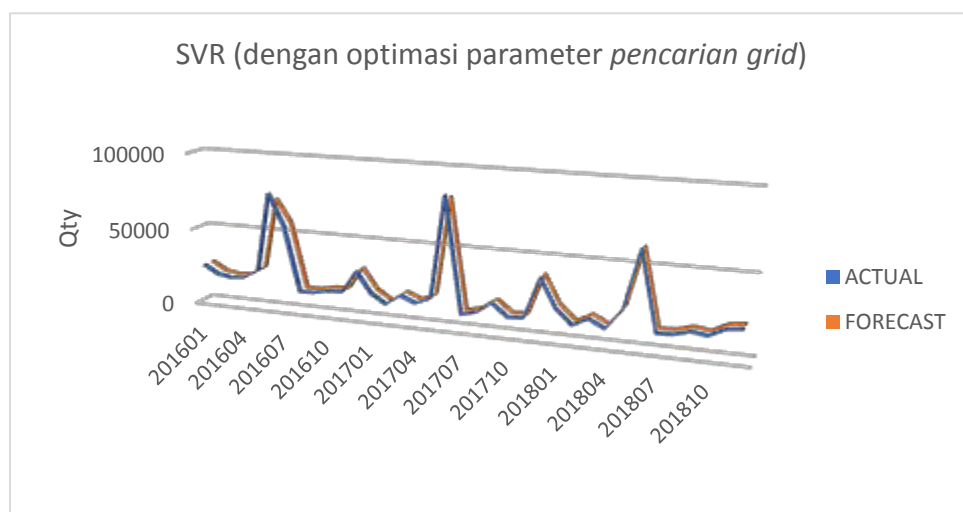
Setelah menemukan nilai parameter yang dapat dilihat pada pada tabel 4.7, pengujian model dilanjutkan untuk meramalkan penjualan. Berikut merupakan hasil dari pengujian *dataset* terhadap model SVR dalam meramalkan penjualan *brand ZXC*. Berdasarkan pengujian algoritma SVR dengan optimasi parameter pencarian *grid* diperoleh hasil RMSE 35,83. Untuk melihat seberapa efektif hasil parameter yang dilakukan, maka dilakukan

pengujian dengan parameter tanpa dioptimalkan. Pada tools *rapidminer* angka *default* untuk semua parameter SVR adalah $C = 0.0$, $\varepsilon = 0.001$ dan $\gamma = 0.0$. Hasil pengujian didapatkan hasil RMSE sebesar 303,63. Dengan melihat hasil RMSE dari kedua pengujian tersebut dapat disimpulkan bahwa optimasi parameter pencarian *grid* terbukti efektif dalam mencari parameter optimal untuk SVR, dengan hasil RMSE atau tingkat kesalahan peramalan yang semakin mengecil. Grafik pengujian model SVR dengan optimasi parameter pencarian *grid* dapat dilihat pada gambar 4.8 dan gambar 4.9. sedangkan untuk grafik pengujian model SVR tanpa optimasi parameter dapat dilihat pada gambar 4.10 dan gambar 4.11.



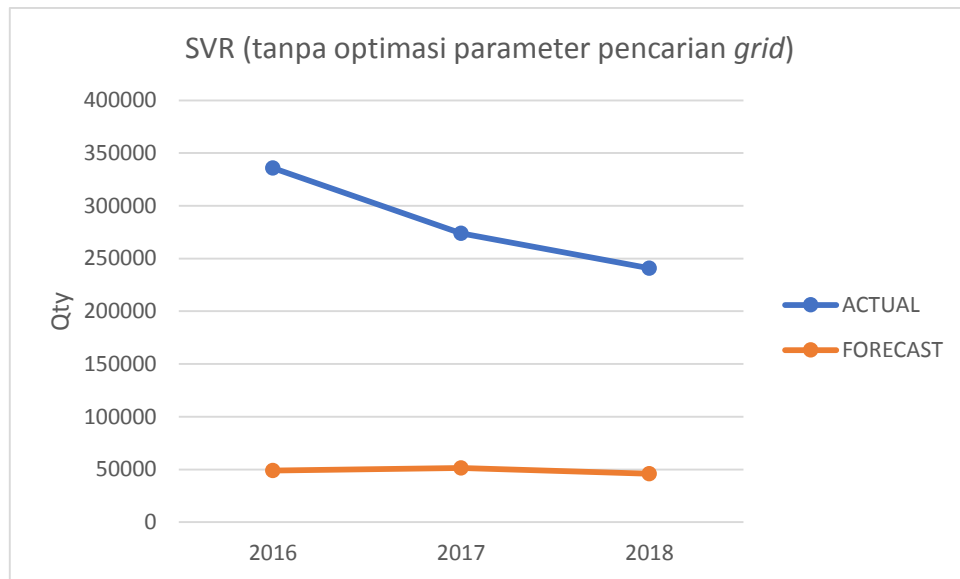
Gambar 4. 8

Grafik pengujian model SVR dengan optimasi parameter pencarian *grid* (pertahun)



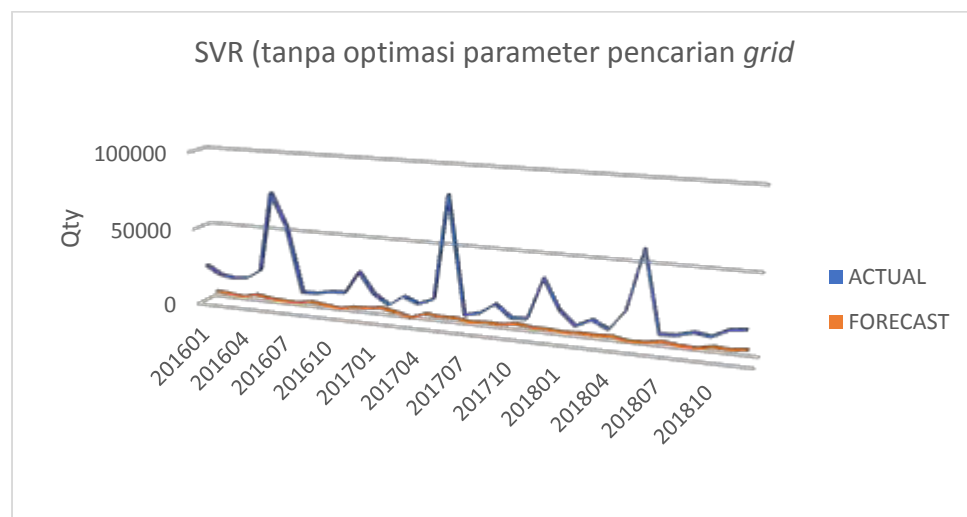
Gambar 4. 9

Grafik pengujian model SVR dengan optimasi parameter pencarian *grid* (perbulan)



Gambar 4. 10

Grafik pengujian model SVR tanpa optimasi parameter pencarian *grid* (pertahun)



Gambar 4. 11

Grafik pengujian model SVR tanpa optimasi parameter pencarian *grid* (perbulan)

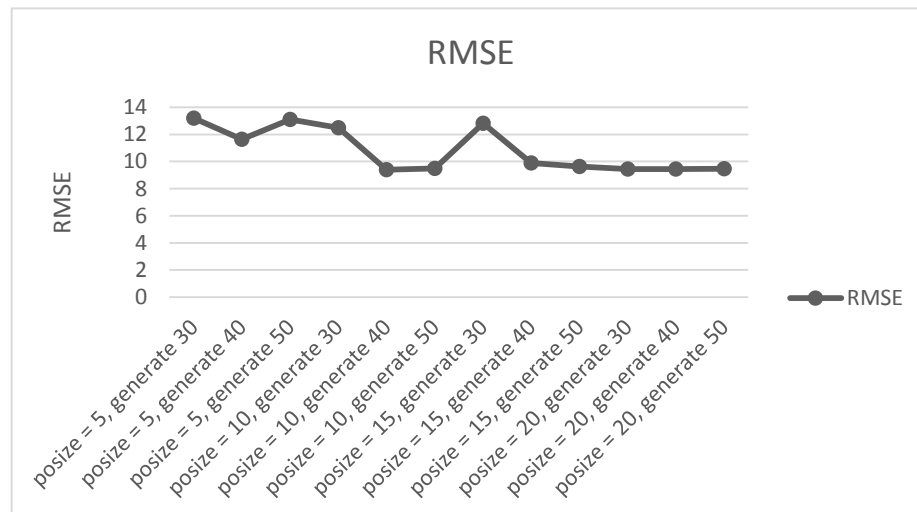
Pengujian selanjutnya melakukan proses PSO kepada model SVR yang telah dibentuk. Pertama, dilakukan pengujian pada nilai parameter *population size* dan *maximum number of generation* yang mempunyai nilai *default* pada *rapidminer* 5 dan 30. Selanjutnya dilakukan kembali uji coba dengan mengubah nilai-nilai untuk parameter PSO. Pada tabel 4.8 merupakan hasil dari pengujian yang dilakukan.

Tabel 4. 8

Hasil pengujian parameter *population size* dan *maximum number of generation*

Parameter PSO	RMSE
Population size = 5, max generate = 30	13,19
Population size = 5, max generate = 40	11,63
Population size = 5, max generate = 50	13,10

Parameter PSO	RMSE
Population size = 10, max generate = 30	12,48
Population size = 10, max generate = 40	9,40
Population size = 10, max generate = 50	9,49
Population size = 15, max generate = 30	12,82
Population size = 15, max generate = 40	9,9
Population size = 15, max generate = 50	9,63
Population size = 20, max generate = 30	9,44
Population size = 20, max generate = 40	9,45
Population size = 20, max generate = 50	9,47



Gambar 4. 12

Grafik perbedaan performa hasil pengujian parameter *population size* dan *maximum number of generation*

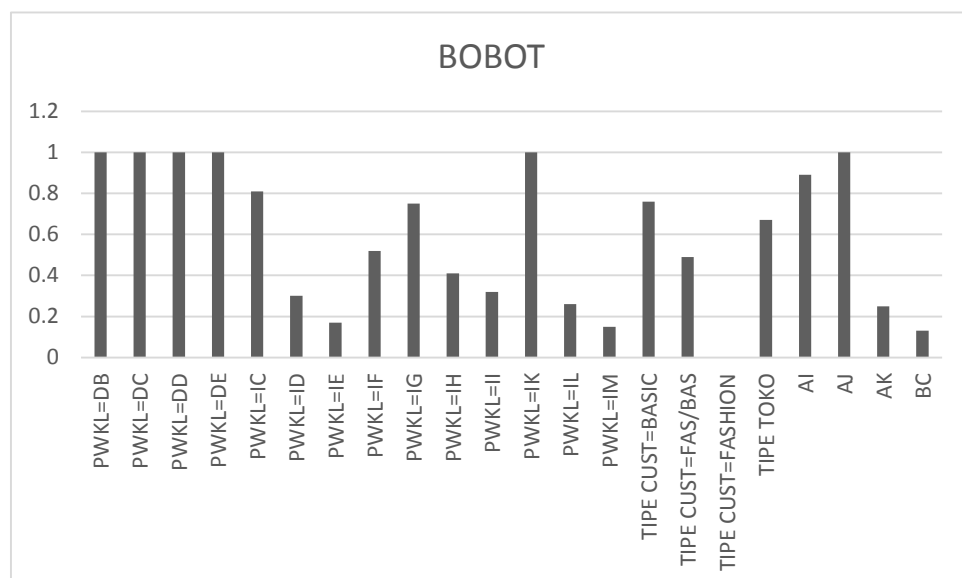
Berdasarkan tabel 4.8 pada parameter *population size* dengan nilai 10 dan parameter *maximum number of generation* dengan nilai 40 telah terjadi konvergen yang artinya performa akurasi telah maksimal. Dengan demikian dapat diketahui SVR-PSO menghasilkan *error rate* terbaik pada saat *population size* 10 dan *maximum number of generation* 40 dengan hasil prediksi RMSE 9,40.

Selanjutnya pada tabel 4.9 menunjukkan hasil bobot (*weight*) atribut prediktor yang telah dioptimalisasikan PSO dengan *error rate* terbaik. Hasil ini dimaksudkan untuk mengetahui atribut mana yang paling berpengaruh terhadap atribut target.

Tabel 4. 9
Hasil optimasi pembobotan atribut

Atribut	Bobot
PWKL=DB	1
PWKL=DC	1
PWKL=DD	1
PWKL=DE	1
PWKL=IC	0,81
PWKL=ID	0,3

Atribut	Bobot
PWKL=IE	0,17
PWKL=IF	0,52
PWKL=IG	0,75
PWKL=IH	0,41
PWKL=II	0,32
PWKL=IK	1
PWKL=IL	0,26
PWKL=IM	0,15
TIPE CUST=BASIC	0,76
TIPE CUST=FAS/BAS	0,49
TIPE CUST=FASHION	0
TIPE TOKO	0,67
AI	0,89
AJ	1
AK	0,25
BC	0,13



Gambar 4. 13
Hasil optimasi pembobotan atribut

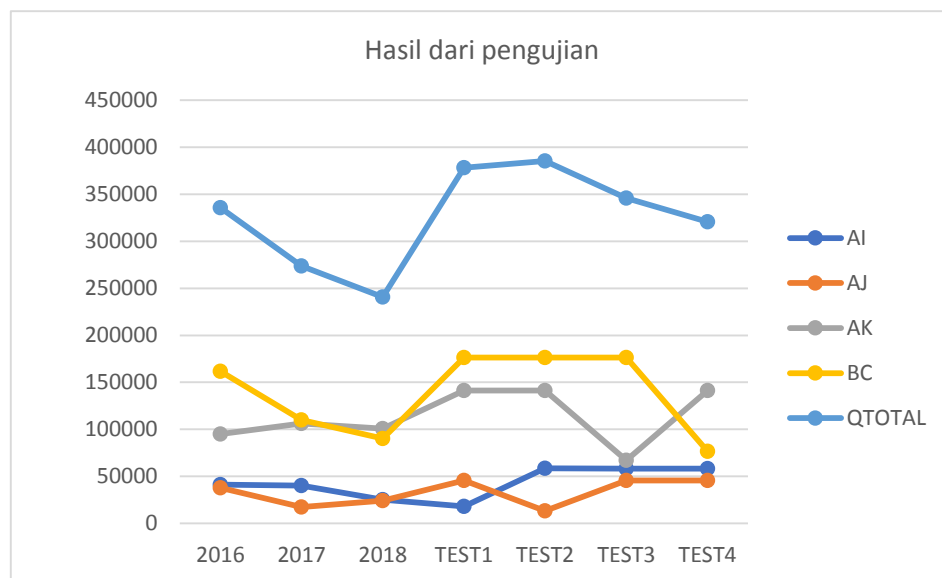
Berdasarkan tabel 4.9 terdapat 22 atribut prediktor yang digunakan. Hasil dari algoritma PSO didapatkan 1 atribut yang menghasilkan bobot 0, yaitu TIPE CUST=FASHION.. Atribut tersebut dimungkinkan tidak akan mempengaruhi hasil akurasi jika dihilangkan.

Pengujian dilanjutkan dengan mengujikan data baru sebagai nilai-nilai prediktor, agar pemodelan dapat digunakan untuk meramalkan penjualan di masa depan. Nilai-nilai prediktor didapatkan dari rentang nilai penjualan perproduk pada tahun 2016-2018. Terdapat 4 *dataset* berbeda, dimana setiap *dataset* meminimalkan nilai pada rentang salah satu produk dan yang lainnya memaksimalkan nilai pada rentang produk. Tujuan dibentuk 4 *dataset* tersebut adalah untuk melihat kombinasi dari 4 produk yang paling

mempengaruhi terhadap QTOTAL peramalan penjualan. Gambar 4.10 merupakan hasil dari pengujian 4 *dataset* berbeda.

Tabel 4. 10
Hasil pengujian data baru

Years	AI	AJ	AK	BC	QTOTAL
2016	41237	37854	94850	161842	335783
2017	40226	17318	106280	109997	273821
2018	25218	24356	100749	90369	240692
TEST1	18115	45614	141311	176424	378228
TEST2	58620	13247	141311	176424	385500
TEST3	58220	45614	66992	176424	345798
TEST4	58220	45614	141311	76565	320657

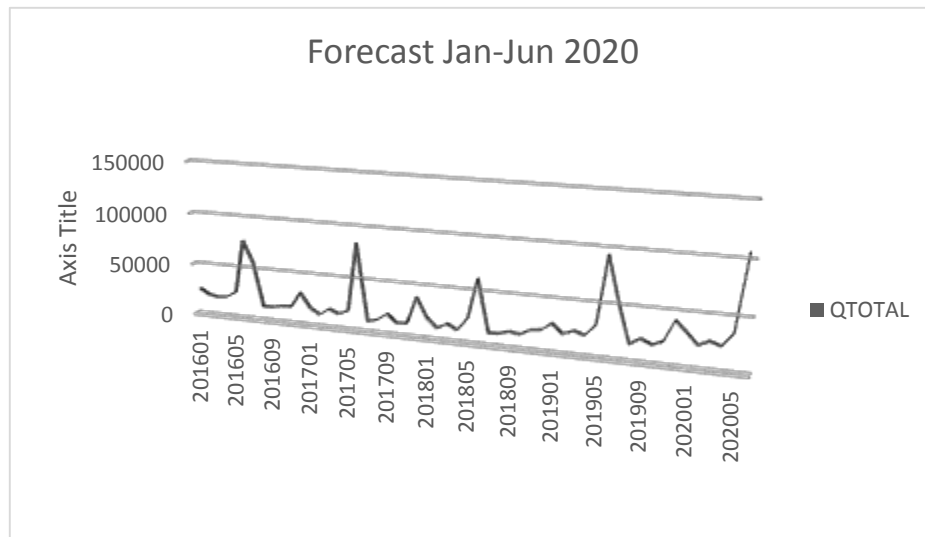


Gambar 4. 14
Grafik hasil pengujian

Berdasarkan tabel 4.10, dari 4 *dataset* yang dilakukan uji coba, *dataset* TEST2 menghasilkan QTOTAL yang paling tinggi. Dengan demikian kombinasi produk yang paling optimal didapatkan pada *dataset* TEST2. Kombinasi produk pada *dataset* TEST2 adalah meminimalkan nilai pada rentang produk AJ dan memaksimalkan nilai pada rentang produk AI, AK dan BC. Grafik 4.14 menunjukkan gambaran jelas tentang produk yang diramalkan akan terjadi penurunan dan kenaikan. Tahap selanjutnya, dilakukan percobaan untuk meramalkan 6 bulan ke depan untuk tahun 2020 dengan kombinasi produk pada *dataset* TEST 2. Pada tabel 4.11 menunjukkan hasil peramalan penjualan 6 bulan kedepan pada tahun 2020. Selain perproduk, hasil dari peramalan juga ditampilkan berdasarkan wilayah, tipe cust dan tipe toko.

Tabel 4. 11
Peramalan penjualan perproduk januari-juni 2020

Month	AI	AJ	AK	BC	Forecast
Jan	6023	899	11731	13510	32161
Feb	4431	808	6515	10072	21829
Mar	4599	1081	10273	10567	26524
Apr	5348	730	6954	10090	23120
Mei	5874	663	13123	15613	35275
Jun	16597	1614	42169	54177	106412



Gambar 4. 15
Grafik penjualan *history* dan peramalan perproduk 2020

Tabel 4. 12
Hasil peramalan penjualan perwilayah perproduk Januari-juni 2020

PWKL	AI	AJ	AK	BC	FORECAST
DB	11335	1993	24295	32521	62515
DC	3701	1204	9538	11648	26125
DD	2611	733	11726	13176	28244
DE	10364	604	13335	15723	39481
IC	2192	335	8003	6264	16759
ID	1234	153	3702	7735	12840
IE	1828	126	5670	6672	14296
IF	755	7	1374	2273	4409
IG	1041	90	1503	1863	4497
IH	2954	165	4317	5734	13152
II	1008	56	1072	2074	4211
IK	264	8	354	657	1283
IL	1969	165	5124	6377	13643
IM	1616	156	752	1312	3833

Tabel 4. 13
Hasil peramalan penjualan berdasarkan tipe cust Januari-juni 2020

TIPE CUST	AI	AJ	AK	BC	FORECAST
BASIC	23347	3541	57401	58712	135443
FAS/BAS	15331	2096	31362	52196	100376

Tabel 4. 14
Hasil peramalan penjualan berdasarkan tipe toko Januari-juni 2020

TIPE TOKO	AI	AJ	AK	BC	FORECAST
0,2	0	0	0	0	0
0,4	7418	1107	18908	20223	46922
0,6	17575	2400	39415	46697	99238
0,8	16167	2133	30109	44836	92649
1	1712	155	2333	2273	6479

4.5 Evaluasi (*evaluation*)

Berdasarkan beberapa kali pengujian yang telah dilakukan pada sub bab 4.4 pemodelan, terdapat beberapa hal yang menjadi evaluasi penelitian Pertama, dalam menentukan model peramalan penjualan yang baik adalah dengan melihat nilai tingkat kesalahan yang dihasilkan model pada *dataset* yang diberikan. *Root Mean Squared Error* (RMSE) merupakan salah satu ukuran dalam menentukan akurasi peramalan. Semakin kecil nilai RMSE (nilai RMSE mendekati angka nol) maka semakin kecil pula tingkat kesalahan yang dihasilkan pada data *testing* terhadap data *training*. Dengan hasil demikian maka peramalan yang dibuat dapat dikatakan baik.

Kedua, kinerja *Support Vector Regression* (SVR) sangat ditentukan dengan parameter yang diberikan. Hal tersebut dapat dilihat dari hasil pengujian *dataset* pada gambar 4.8 dan gambar 4.9 yang diperoleh hasil RMSE sebesar 35,83. Hasil tersebut menunjukkan hasil baik karena semakin kecil nilai RMSE akan semakin akurat peramalan karena tingkat kesalahan yang terbentuk semakin mengecil. Hal ini berbanding terbalik dengan hasil pengujian tanpa optimasi parameter pada gambar 4.10 dan gambar 4.11 yang diperoleh nilai RMSE sebesar 303,63. Maka dari itu pernyataan kridanti perihal kehandalan performa algoritme SVR sangat ditentukan dengan pemilihan parameter yang digunakan (Kridanto, 2015:302) terbukti benar.

Ketiga, pengujian menambahkan algortime *Particle Swarm Optimization* (PSO) dengan SVR dilakukan dengan mengubah nilai pada *population size* dan *maximum number of generation* yang merupakan salah satu parameter yang mempengaruhi kinerja algortime PSO. *Population size* merupakan jumlah partikel n dalam gerombolan. Gerombolan terbentuk karena dorongan keseragaman dari masing-masing partikel.

Sedangkan *maximum number of generation* merupakan jumlah pengulangan (iterasi) yang dilakukan untuk menghasilkan urutan hasil. Dari kedua parameter tersebut memiliki sifat semakin tinggi nilai populasi dan iterasi yang diberikan maka akan semakin memakan waktu dalam eksekusi. Sebaliknya jika nilai kedua parameter tersebut terlalu rendah kemungkinan kompleksitas partikel tidak maksimal dikarenakan iterasi rendah menghentikan proses pencarian sebelum waktunya. Pengujian dilakukan dengan mengubah nilai pada kedua parameter tersebut, pengamatan berdasarkan pengujian pertama dengan memberikan nilai *population size* 5 dan *maximum number of generation* 30 menghasilkan lamanya eksekusi 18 menit 50 detik dengan RMSE 13,19. Sedangkan dengan nilai *population size* 20 dan *maximum number of generation* 50 diperoleh lama eksekusi 43 menit 10 detik dengan RMSE 9,47. Dari hasil pengamatan tersebut dapat dikatakan semakin tinggi nilai parameter yang diberikan maka semakin memerlukan waktu eksekusi. Pengamatan lainnya didapatkan ketika nilai *population size* 10 dan *maximum number of generation* 40 telah terjadi konvergen yang artinya performa akurasi telah maksimal. Dengan demikian dapat diketahui SVR-PSO menghasilkan tingkat kesalahan terbaik pada saat *population size* 10 dan *maximum number of generation* 40 dengan hasil prediksi RMSE 9,40 dan lama eksekusi 25 menit 25 detik. Baik nilai RMSE menjadi lebih kecil terhadap tingkat kesalahan dibandingkan dengan nilai yang dihasilkan model SVR saja, maka dapat dikatakan SVR-PSO memiliki hasil paling optimal untuk peramalan. Dari pembobotan atribut diperoleh hasil 1 atribut mempunyai bobot 0.

Keempat, berdasarkan tabel 4.10 dengan melakukan uji coba terhadap 4 *dataset* berbeda mendapatkan hasil kombinasi produk yang paling optimal pada *dataset* TEST2. Kombinasi produk pada *dataset* TEST2 adalah meminimalkan nilai pada rentang produk AJ dan memaksimalkan nilai pada rentang produk AI, AK dan BC. Pertumbuhan maksimal untuk produk didapatkan dari produk AK dengan pertumbuhan 49% dari penjualan 2016 yang merupakan penjualan total tertinggi dari data training.

Keempat, percobaan selanjutnya dilakukan untuk meramalkan penjualan pada 6 bulan ke depan, periode Januari-juni 2020. Hasil dari peramalan dapat digunakan sebagai salah satu pertimbangan keputusan untuk menentukan target marketing ataupun target produksi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pembahasan yang telah dijelaskan sesuai dengan tahapan penelitian pada masing-masing bab sebelumnya, maka dapat diambil beberapa kesimpulan:

1. Algoritme *Support Vector Regression* (SVR) dapat dikombinasikan dengan algoritme *Particle Swarm Optimization* (PSO) dalam melakukan peramalan penjualan.
2. Hasil dari mengkombinasikan algoritme SVR-PSO didapatkan hasil akurasi peramalan menjadi lebih baik, dimana nilai RMSE yang dihasilkan SVR tanpa PSO sebesar 35,83, sedangkan hasil kombinasi dari SVR-PSO didapatkan hasil prediksi RMSE sebesar 9,40.
3. Pemodelan SVR-PSO dapat digunakan untuk meramalkan penjualan. Berdasarkan hasil pengujian dalam meramalkan penjualan produk dapat dilihat bahwa penjualan dari berbagai produk akan mengalami kenaikan dan penurunan pada tahun-tahun berikutnya. Hal-hal tersebut dapat dijadikan pengetahuan atau sumber informasi bagi pihak eksekutif dalam mengambil keputusan. Hasil dari peramalan tersebut juga dapat dipakai sebagai acuan pihak produksi dalam mempersiapkan *inventory* yang sesuai.

5.2 Saran

Berikut beberapa saran yang disampaikan agar hasil penelitian ini dapat bermanfaat dan dapat lebih baik pada penelitian selanjutnya, adapun saran-saran tersebut adalah:

1. Menambahkan atribut lebih banyak dan jumlah data yang lebih besar, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi, dikarenakan konsep dari mesin pembelajaran adalah semakin banyak data yang diolah maka semakin banyak mesin dapat belajar dan menemukan pola yang lebih efektif.

2. Menggunakan metode lainnya seperti *deep learning* untuk penanganan kapabilitas lebih kompleks untuk mempelajari, mencerna, dan mengklasifikasikan dalam data penjualan.

DAFTAR PUSTAKA

- Adhikari K., R., & R.K., A. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *ArXiv Preprint ArXiv:1302.6613*.
- Anandhi, V. (2013). COMPARISON OF THE FORECASTING TECHNIQUES – ARIMA, ANN AND SVM - A REVIEW. *IJCET*, 370–376.
- Astuti, A. I. (2013). *Pemodelan Runtun Waktu ARIMAX dengan Efek Variasi Kalender*. 1–4.
- Awad, M., & Khanna, R. (2015). Efficient learning machines: Theories, concepts, and applications for engineers and system designers. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*.
- Berrar, D. (2019). Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology*, 542–545.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concept and Techniques* (3rd Editio). Elsevier Inc.
- Hasan, N., Nath, N. C., & Rasel, R. I. (2015). A Support Vector Regression Model for Forecasting Rainfall. *IEEE*, (Eict).
- Heizer, J., Render, B., & Munson, C. (2017). *T Welf T H Edit Ion Oper At Ions*.
- Hsieh, H., & Lee, T. (2011). *A Hybrid Particle Swarm Optimization and Support Vector Regression Model for Financial Time Series Forecasting*. 2(2), 48–56.
- Hurwitz, J., & Kirsch, D. (2018). Machine Learning for Dummies. In *Journal of the American Society for Information Science* (IBM Limite, Vol. 35). United States of America: John Wiley & Sons, Inc.
- Kandananond, K. (2012). *International Journal of Engineering Business Management A Comparison of Various Forecasting Methods for Autocorrelated Time Series Regular Paper*. 4, 1–6.
- Kridanto, A., & Buliali, J. L. (2015). Metode Hibrida FCM dan PSO-SVR untuk Prediksi Data Arus Lalu Lintas. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(3), 302–311.
- Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining* (2nd editio).
- Lu, X., & Geng, X. (2011). Car Sales Volume Prediction Based on Particle Swarm Optimization Algorithm and Support Vector Regression. *IEEE*, (1), 71–74.
- Narkevičius, R., & Šeškauskis, Z. (2016). *Sales Forecasting Management*. University of Borås.
- Palmer, A., Jimenez, R., & Gervill, E. (2011). Data Mining: Machine Learning and Statistical Techniques. *ResearchGate*, 3(May 2014).
- Rifqi, M. R., Setiawan, B. D., & Bachtiar, F. A. (2018). *Support Vector Regression Untuk*

Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang – PMI Kota Malang. 2(10), 3332–3342.

Rusmalawati, V., & Furqon, M. T. (2018). Peramalan Harga Saham Menggunakan Metode Support Vector Regression (SVR) Dengan Particle Swarm Optimization (PSO). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(5), 1980–1990.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1394–1401.

Syarif, I., Prugel-bennett, A., & Wills, G. (2016). SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA*, 14(4), 1502–1509.

Wen, Q., Mu, W., Sun, L., Hua, S., & Zhou, Z. (2014). Daily Sales Forecasting for Grapes by Support Vector Machine. *IFIP*, 351–360.

Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Miner*. jakarta.

Witten, H. I., Frank, E., & Hall, A. M. (2011). Data Mining Practical Machine Learning Tools and Techniques 3th Edition. In *Atom Probe Tomography: Put Theory Into Practice* (3th ed.). United States of America: Elsevier Inc.

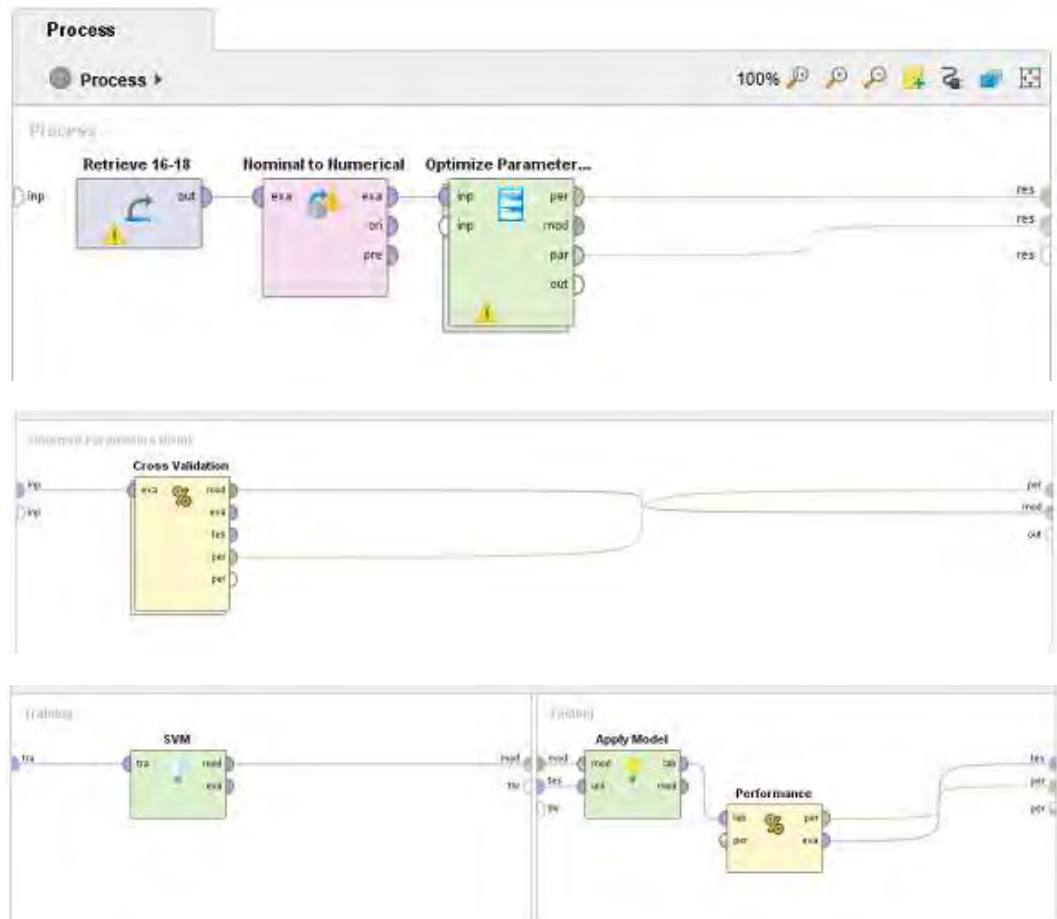
Yu, X., Qi, Z., & Zhao, Y. (2013). Support vector regression for newspaper/magazine sales forecasting. *Procedia Computer Science*, 17, 1055–1062. <https://doi.org/10.1016/j.procs.2013.05.134>

Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). *Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data.* 9(2), 1–16.

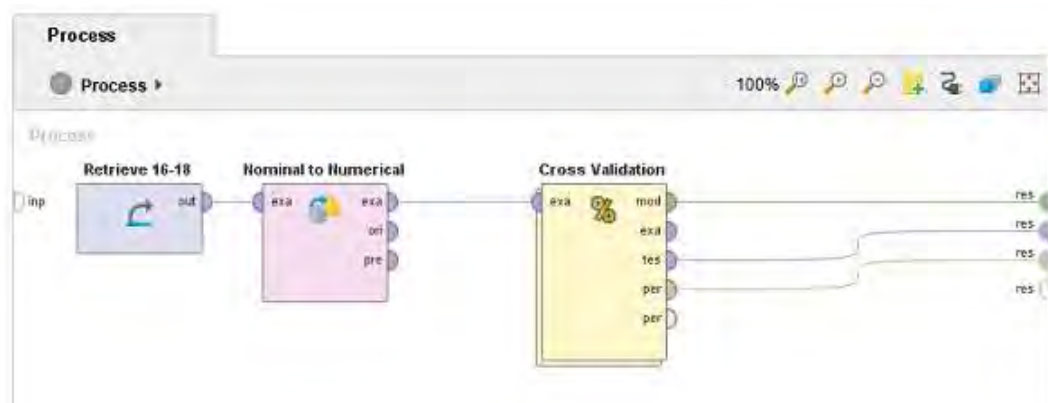
LAMPIRAN

Lampiran A Hasil Pemodelan Dengan *Tools Rapidminer*

Pemodelan pencarian parameter SVR yang paling optimal dengan menggunakan optimisasi parameter pencarian *grid*

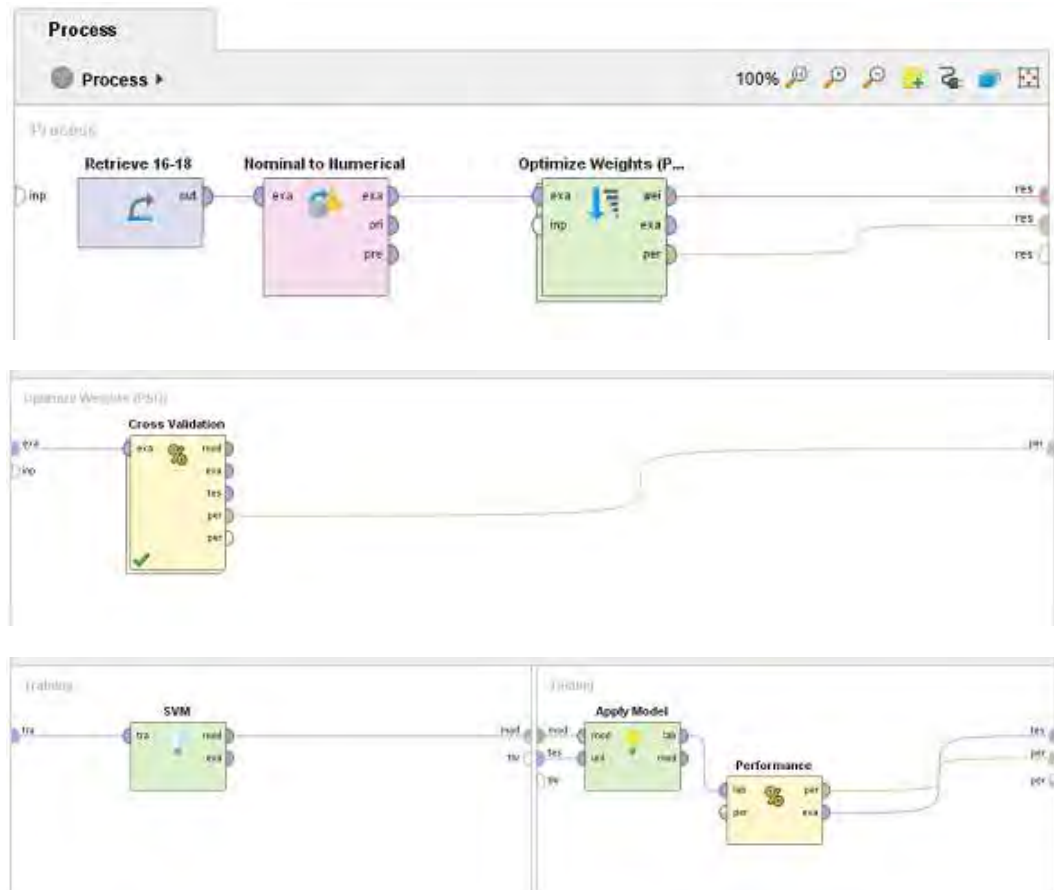


Pemodelan SVR





Pemodelan SVR-PSO



Lampiran B *Dataset* Penjualan (sebelum transformasi)

No.	BPK	TG_BPK	PWKL	CUST	KET	INUSER	INTIME	IDBRCODE	NO_KW	WARNA	MERK	PRODUK	HGJ	QTOTAL
1	2DB160100359	1/Jan/16	DB	09.363	web-				ECC6024F	02	G	BC	309900	1
2	2DB160100362	2/Jan/16	DB	09.363	web-				ECC6024F	02	G	BC	309900	1
3	2DB160100473	1/Jan/16	DB	10.697	web-girl pot 100rb				ECC6026F	02	G	BC	309900	3
4	2DB160100581	2/Jan/16	DB	10.697	web-girl pot 100rb				ECC6026F	02	G	BC	309900	1
5	2DB160101345	1/Jan/16	DB	10.118	web-				ECC6026F	02	G	BC	309900	1
6	2DB160101421	3/Jan/16	DB	10.697	web-girl pot 100rb				ECC6026F	02	G	BC	309900	1
7	2DB160101864	3/Jan/16	DB	10.358	web-				ECC6024F	02	G	BC	309900	1
8	2DB160102061	2/Jan/16	DB	09.445	web-				ECC6026F	02	G	BC	309900	1
9	2DB160102173	1/Jan/16	DB	30.139	web-				EML5409K	01	G	BC	129900	1
10	2DB160102312	2/Jan/16	DB	10.118	web-				ECC6026F	02	G	BC	309900	1
11	2DB160103133	3/Jan/16	DB	10.131	web-				ECC6026F	02	G	BC	309900	2
12	2DB160104432	5/Jan/16	DB	09.017	web-pot 100				ECC6024F	02	G	BC	309900	1
13	2DB160106495	9/Jan/16	DB	30.138	web-				EML5409K	01	G	BC	129900	1
14	2DB160107028	8/Jan/16	DB	30.108	web-				ECC6026F	02	G	BC	309900	1
15	2DB160107275	9/Jan/16	DB	09.363	web-				ECC6024F	02	G	BC	309900	1
16	2DB160107376	9/Jan/16	DB	10.327	web-				ECC6026F	02	G	BC	309900	1
17	2DB160107541	10/Jan/16	DB	09.017	web-pot 100				ECC6026F	02	G	BC	309900	1
18	2DB160108188	8/Jan/16	DB	10.416	web-				ECC6026F	02	G	BC	309900	1
19	2DB160108207	8/Jan/16	DB	30.139	web-				ECC6024F	02	G	BC	309900	1
20	2DB160108538	10/Jan/16	DB	10.118	web-				ECC6026F	02	G	BC	309900	2
21	2DB160111205	14/Jan/16	DB	10.118	web-				ECC6026F	02	G	BC	309900	1
22	2DB160113116	16/Jan/16	DB	09.363	web-				ECC6026F	02	G	BC	309900	1
23	2DB160113162	16/Jan/16	DB	08.057	web-				ECC6026F	02	G	BC	309900	1
24	2DB160113200	16/Jan/16	DB	10.128	web-				ECC6024F	02	G	BC	309900	1
25	2DB160114381	16/Jan/16	DB	10.118	web-				ECC6026F	02	G	BC	309900	1

No.	BPK	TG BPK	PWKL	CUST	KET	INUSER	INTIME	IDBRCODE	NO_KW	WARNA	MERK	PRODUK	HGJ	QTOTAL
26	2DB160115062	18/Jan/16	DB	10.327	web-				ECC6026F	02	G	BC	309900	1
27	2DB160115462	18/Jan/16	DB	10.114	20%				ECC6024F	02	G	BC	309900	1
28	2DB160115631	18/Jan/16	DB	09.363	web-				ECC6024F	02	G	BC	309900	1
29	2DB160115663	19/Jan/16	DB	09.037	sp				ECC6026F	02	G	BC	309900	1
30	2DB160116443	20/Jan/16	DB	09.051	web-				ECC6024F	02	G	BC	309900	1
31	2DB160117803	22/Jan/16	DB	09.017	web-				GBCBC00272F	02A	G	BC	279900	1
32	2DB160117854	22/Jan/16	DB	10.416	web-				GBCBC00272F	02A	G	BC	279900	1
33	2DB160117947	22/Jan/16	DB	09.103					EML5409K	01	G	BC	129900	1
34	2DB160118690	22/Jan/16	DB	10.114	30%				GBCBC00272F	02A	G	BC	279900	1
	...	...	...	...	...	...	...	...	...	...	...	...	...	...
737062	2IM181200883	31/Dec/18	IM	26.007				116261	GAIAI00462J	02Z	G	AK	349900	1
737063	2IM181200883	31/Dec/18	IM	26.007				117527	GAIAB00308P	06E	G	AI	209900	1
737064	2IM181200884	31/Dec/18	IM	26.007				024131	ECC6026F	02	G	BC	269900	1
737065	2IM181200947	28/Dec/18	IM	26.013				024131	ECC6026F	02	G	BC	269900	1
737066	2IM181200947	28/Dec/18	IM	26.013				115234	GBCBP00522F	08A	G	BC	269900	1
737067	2IM181200947	28/Dec/18	IM	26.013				115437	GBCB100401F	17A	G	BC	269900	1
737068	2IM181200947	28/Dec/18	IM	26.013				119596	GBCBP00673F	16A	G	BC	269900	1
737069	2IM181200947	28/Dec/18	IM	26.013				120611	GBCBP00104F	16A	G	BC	269900	2
737070	2IM181200949	31/Dec/18	IM	26.013				024129	ECC6024F	02	G	BC	269900	1
737071	2IM181200949	31/Dec/18	IM	26.013				109133	GBCBP00586F	14A	G	BC	269900	1
737072	2IM181200949	31/Dec/18	IM	26.013				113420	GBCBS00003F	01A	G	BC	269900	1
737073	2IM181200949	31/Dec/18	IM	26.013				115234	GBCBP00522F	08A	G	BC	269900	1
737074	2IM181200949	31/Dec/18	IM	26.013				115437	GBCB100401F	17A	G	BC	269900	1
737075	2IM181200949	31/Dec/18	IM	26.013				115649	GBCBC00014F	14A	G	BC	269900	2
737076	2IM181200949	31/Dec/18	IM	26.013				115703	GBCBC00017F	01A	G	BC	269900	2
737077	2IM181200950	28/Dec/18	IM	26.013				117320	GAKAD00467J	11A	G	AK	349900	1
737078	2IM181200951	31/Dec/18	IM	26.013				105507	GAKAD00359P	02H	G	AK	319900	1
737079	2IM181200951	31/Dec/18	IM	26.013				117321	GAKAR00468J	11A	G	AK	349900	1
737080	2IM181200951	31/Dec/18	IM	26.013				117527	GAIAB00308P	06E	G	AI	209900	2
737081	2IM181200951	31/Dec/18	IM	26.013				118742	GAKAD00475P	02H	G	AK	349900	1
737082	2IM181200951	31/Dec/18	IM	26.013				119187	GAKAD00484J	02H	G	AK	349900	1
737083	2IM181200990	26/Dec/18	IM	26.007				107376	GAIAI00272P	01A	G	AI	179900	1
737084	2IM181200991	22/Dec/18	IM	26.007				107376	GAIAI00272P	01A	G	AI	179900	1
737085	2IM181200992	25/Dec/18	IM	26.007				107376	GAIAI00272P	01A	G	AI	179900	1
737086	2IM181100269	23/Nov/18	DB	09.494				112546	GAJAR00023J	16A	G	AJ	169000	1

[illegible]

No	SZ38	SZ39	SZ40	SZ41	SZ42	SZ43	SZ44	SZ45	SZ46	TOTBRUTO	POTONGAN	HRGKASIR	TOTNETTO
31													
32													
33													
34													
	...	...	...	...	...	...	...	...	...	...	...	...	...
737062	0	0	0	0	0	0	0	0	0	349900	0	244930	189747
737063	0	0	0	0	0	0	0	0	0	209900	0	146930	113342
737064	0	0	0	0	0	0	0	0	0	269900	0	269900	215920
737065	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737066	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737067	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737068	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737069	0	0	0	0	0	0	0	0	0	539800	0	539800	388656
737070	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737071	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737072	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737073	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737074	0	0	0	0	0	0	0	0	0	269900	0	269900	194328
737075	0	0	0	0	0	0	0	0	0	539800	0	539800	388656
737076	0	0	0	0	0	0	0	0	0	539800	0	539800	388656
737077	0	0	0	0	0	0	0	0	0	349900	0	279920	201542
737078	0	0	0	0	0	0	0	0	0	319900	0	223930	161230
737079	0	0	0	0	0	0	0	0	0	349900	0	244930	176350
737080	0	0	0	0	0	0	0	0	0	419800	0	293860	211579
737081	0	0	0	0	0	0	0	0	0	349900	0	244930	176350
737082	0	0	0	0	0	0	0	0	0	349900	0	244930	176350
737083	0	0	0	0	0	0	0	0	0	179900	0	179900	100744
737084	0	0	0	0	0	0	0	0	0	179900	0	179900	88151
737085	0	0	0	0	0	0	0	0	0	179900	0	179900	88151
737086	0	0	0	0	0	0	0	0	0	169000			169000

Lampiran C *Dataset* Penjualan (setelah transformasi)

YEARMONTH	PWKL	TIPE CUST	TIPE TOKO	AI	AJ	AK	BC	QTOTAL
201601	DB	BASIC	0,4	244	191	297	541	1273
201601	DB	FAS/BAS	0,4	14	43	55	98	210
201601	DB	BASIC	0,6	382	516	766	1366	3030
201601	DB	FAS/BAS	0,6					
201601	DB	BASIC	0,8	223	382	403	431	1439
201601	DB	FAS/BAS	0,8	165	205	225	647	1242
201601	DB	FASHION	0,8	96	61	192	202	551
201601	DB	BASIC	1	84	176	98	198	556
201601	DC	BASIC	0,4	19	92	64	98	273
201601	DC	BASIC	0,6	150	224	196	229	799
201601	DC	FAS/BAS	0,6	5	14	22	18	59
201601	DC	BASIC	0,8	21	28	7	28	84
201601	DC	FAS/BAS	0,8	65	246	121	249	681
201601	DD	BASIC	0,4	13	46	65	69	193
201601	DD	FAS/BAS	0,4	12	14	14	54	94
201601	DD	BASIC	0,6	110	172	361	481	1124
201601	DD	FAS/BAS	0,6	39	68	163	337	607
201601	DD	BASIC	0,8	28	11	15	61	115
201601	DD	FAS/BAS	0,8	27	106	134	327	594
201601	DE	BASIC	0,4	132	20	99	151	402
201601	DE	FAS/BAS	0,4	102	22	77	92	293
201601	DE	BASIC	0,6	128	57	223	188	596
201601	DE	FAS/BAS	0,6	21	33	42	43	139
201601	DE	FASHION	0,6	110	13	28	23	174
201601	DE	BASIC	0,8	69	34	192	219	514
201601	DE	FAS/BAS	0,8	341	207	362	301	1211
201601	DE	FASHION	0,8	261	66	76	37	440
201601	DE	FASHION	1	168	20	24	25	237
201601	IC	BASIC	0,4	55	41	127	86	309
201601	IC	BASIC	0,6	102	75	247	211	635

YEARMONTH	PWKL	TIPE CUST	TIPE TOKO	AI	AJ	AK	BC	QTOTAL
...	...	...	...	...	...	...	...	...
201812	ID	BASIC	0,6	19	0	87	29	135
201812	ID	FAS/BAS	0,6	22	8	47	67	144
201812	ID	FAS/BAS	0,8	68	54	171	156	449
201812	IE	BASIC	0,4	44	24	230	128	426
201812	IE	BASIC	0,6	35	5	61	26	127
201812	IE	FAS/BAS	0,6	23	10	55	51	139
201812	IE	FAS/BAS	0,8	5	27	49	92	173
201812	IF	BASIC	0,4	20	0	8	9	37
201812	IF	FAS/BAS	0,4	2	0	40	11	53
201812	IF	BASIC	0,6	9	0	9	0	18
201812	IF	FAS/BAS	0,6	33	0	41	28	102
201812	IF	FAS/BAS	0,8	39	12	108	160	319
201812	IG	BASIC	0,4	15	19	33	59	126
201812	IG	BASIC	0,6	37	11	19	20	87
201812	IG	BASIC	0,8	28	6	17	23	74
201812	IG	FAS/BAS	0,8	26	0	59	42	127
201812	IH	BASIC	0,4	0	1	58	34	93
201812	IH	BASIC	0,6	42	10	111	41	204
201812	IH	FAS/BAS	0,6	120	42	297	123	582
201812	IH	FAS/BAS	0,8	37	56	112	156	361
201812	II	BASIC	0,4	56	0	91	34	181
201812	II	FAS/BAS	0,8	142	16	347	303	808
201812	IL	BASIC	0,4	1	5	18	10	34
201812	IL	BASIC	0,6	0	1	104	50	155
201812	IL	FAS/BAS	0,6	180	22	164	214	580
201812	IL	BASIC	0,8	5	2	42	15	64
201812	IL	FAS/BAS	0,8	26	7	584	358	975
201812	IM	BASIC	0,4	0	0	0	3	3
201812	IM	BASIC	0,6	66	14	74	49	203
201812	IM	FAS/BAS	0,6	22	11	55	115	203

Lampiran D Hasil Pengujian Optimasi Parameter

PARAMETER C

RENTANG	RMSE
1-10	575,899
1-20	579,1
1-50	573,443
1-100	555,352
100-200	569,583
100-500	547,418
100-1000	549,139
1000-2000	541,403
1000-5000	542,879
1000-10000	540,735

PARAMETER EPSILON

RENTANG	RMSE
0,000001-0,01	546,971
0,000001-0,1	542,797
0,000001-1	540,49
0,00001-0,01	544,743
0,00001-0,1	543,533
0,00001-1	540,76
0,0001-0,01	540,734
0,0001-0,1	540,45
0,0001-1	544,836

PARAMETER GAMMA

RENTANG	RMSE
0,000001-0,01	51,24
0,000001-0,1	51,24
0,000001-1	51,24
0,00001-0,01	122,169
0,00001-0,1	120,105
0,00001-1	123,463
0,0001-0,01	262,653
0,0001-0,1	263,635
0,0001-1	259,365

Lampiran E Hasil Pengujian PSO

ATRIBUT PREDIKTOR	P=5 G=30	P=5 G=40	P=5 G=50	P=10 G=30	P=10 G=40	P=10 G=50	P=15 G=30	P=15 G=40	P=15 G=50	P=20 G=30	P=20 G=40	P=20 G=50
PWKL=DB	1	0,87	0,83	0,99	1	1	0	0,42	1	0,43	1	1
PWKL=DC	1	1	1	1	1	1	0	1	0	0	0,13	1
PWKL=DD	0,33	0,3	0,36	0,68	1	0	0	1	1	0,2	1	0
PWKL=DE	1	1	1	0,37	1	1	0,92	0	0	0,68	0,6	1
PWKL=IC	1	1	1	0,03	0,81	0	1	0,82	0	1	0,2	0
PWKL=ID	0	0	0	0	0,3	0,63	0,63	0,97	0,05	0,55	1	0,51
PWKL=IE	1	1	0,63	0	0,17	0	0,11	1	0,82	0,001	0,11	0
PWKL=IF	0	1	0,66	0	0,52	0,75	0,56	0,78	0,99	0	0,54	0
PWKL=IG	0,76	0	0,65	1	0,75	0,68	0,93	0	0	1	0	1
PWKL=IH	1	1	1	0,85	0,41	0	0,27	0,91	0	0	0,6	0
PWKL=II	0	0	0	0	0,32	0	0,92	0,06	1	0,47	0,48	0,96
PWKL=IK	0,75	0	0,38	0	1	1	1	0,66	0	0,79	0,008	0
PWKL=IL	0,02	0,14	0,46	0,17	0,26	0,32	0	1	0	0,32	1	0,14
PWKL=IM	0	0	0	0	0,15	0	1	0,006	1	0,023	0,02	0
TIPE CUST=BASIC	0,7	0,67	0,55	0,42	0,76	0,89	0,13	0,98	0	0,2	0	1
TIPE CUST=FAS/BAS	0,73	1	0,93	0,21	0,49	0	0,85	1	0	0,035	0	0,38
TIPE CUST=FASHION	0,26	0,03	0	0,39	0	0,1	0,08	1	0,14	0	1	0,71
TIPE TOKO	0,13	0	0	0,08	0,67	0	0,68	0,41	0	0	0,18	0,1
AI	0,18	0,2	0,2	0,18	0,89	0,19	0,52	0,44	0,22	0,11	0,21	0,35
AJ	1	1	1	0,37	1	0,13	1	0,12	0,19	0,041	0,26	0,25
AK	0,3	0,3	0,34	0,36	0,25	0,23	0,26	0,18	0,26	0,1	0,17	0,28
BC	0,1	0,12	0,13	0,07	0,13	0,15	0,13	0,15	0,13	0,034	0,19	0,09

Lampiran F Hasil Pengujian SVR-PSO

YEARMONTH	AI	AJ	AK	BC	QTOTAL	FORECAST
201601	4411	3923	6454	10735	25523	25513
201602	3246	3784	4575	8522	20127	20125
201603	2961	3239	4121	8286	18607	18616
201604	3333	3116	4545	8645	19639	19640
201605	4150	2686	6295	12390	25521	25541
201606	5454	4973	22140	44263	76830	71502
201607	6127	5357	16769	28330	56583	56437
201608	2229	2107	3792	5919	14047	14055
201609	2222	1665	4550	5927	14364	14370
201610	1889	2241	4838	7421	16389	16396
201611	1805	2039	5451	7553	16848	16857
201612	3410	2724	11320	13851	31305	31311
201701	2329	1346	6512	7724	17911	17917
201702	1489	1265	4134	4991	11879	11884
201703	2486	1735	7542	6727	18490	18494
201704	2789	1266	5249	5366	14670	14673
201705	2308	985	8060	7488	18841	18846
201706	12464	2595	31985	37459	84503	82015
201707	1335	820	4223	4318	10696	10702
201708	1179	1066	5642	5318	13205	13211
201709	1137	971	9374	8345	19827	19833
201710	1478	1195	4900	4581	12154	12160
201711	2282	1122	4781	4624	12809	12816
201712	8950	2952	13878	13056	38836	38898
201801	2096	2233	9048	7099	20476	20487
201802	1300	1434	4123	4677	11534	11543
201803	2233	1962	6552	5838	16585	16594
201804	1896	1282	4418	4046	11642	11649
201805	2366	1375	10157	10090	23988	23997
201806	4320	2823	31525	23269	61937	61163
201807	1469	1521	5024	4187	12201	12210
201808	1473	2208	4406	4647	12734	12742
201809	2417	3448	4003	5721	15589	15601
201810	1225	1854	5538	5574	14191	14199
201811	1005	1984	7816	8475	19280	19290
201812	3418	2232	8139	6746	20535	20538
201901	4603	1122	10137	11737		27603
201902	3386	1003	5632	8750		18778
201903	3511	1343	8879	9176		22915
201904	4088	906	6010	8763		19772
201905	4489	826	11341	13565		30227
201906	12676	2007	36445	47062		94091
201907	6195	736	16953	28375		52200
201908	2418	865	6185	6641		16117
201909	2973	842	9521	8876		22210
201910	2363	1007	6550	7964		17889
201911	2665	929	8535	9719		21854
201912	9253	1661	15123	15796		41844
202001	6023	899	11731	13510		32161
202002	4431	808	6515	10072		21829
202003	4599	1081	10273	10567		26524
202004	5348	730	6954	10090		23120
202005	5874	663	13123	15613		35275

YEARMONTH	AI	AJ	AK	BC	QTOTAL	FORECAST
202006	16597	1614	42169	54177		106412

PWKL	2019 (TEST2)				
	AI	AJ	AK	BC	FORECAST
DB	14288	4728	38043	48635	101607
DC	5276	2708	14131	17026	39166
DD	3219	1822	17732	21321	44102
DE	14190	1368	18881	22794	57118
IC	3068	745	12560	10064	26413
ID	1655	328	5867	11449	19297
IE	2574	256	8723	10150	21705
IG	1391	198	2485	3062	7136
IH	4034	316	7108	9501	20950
IK	381	19	568	931	1899
IL	2920	351	9011	11163	23450

PWKL	2020 (JAN-JUN)				
	AI	AJ	AK	BC	FORECAST
DB	11335	1993	24295	32521	62515
DC	3701	1204	9538	11648	26125
DD	2611	733	11726	13176	28244
DE	10364	604	13335	15723	39481
IC	2192	335	8003	6264	16759
ID	1234	153	3702	7735	12840
IE	1828	126	5670	6672	14296
IG	1041	90	1503	1863	4497
IH	2954	165	4317	5734	13152
IK	264	8	354	657	1283
IL	1969	165	5124	6377	13643

TIPE CUST	AI	AJ	AK	BC	FORECAST
BASIC	23347	3541	57401	58712	135443
FAS/BAS	15331	2096	31362	52196	100376

TIPE TOKO	AI	AJ	AK	BC	FORECAST
0,2	0	0	0	0	0
0,4	7418	1107	18908	20223	46922
0,6	17575	2400	39415	46697	99238
0,8	16167	2133	30109	44836	92649
1	1712	155	2333	2273	6479