



## ***Support Vector Machine Menggunakan Forward Selection*** **untuk Prediksi Penjualan Obat**

**Andi Bode**

Fakultas Ilmu Komputer, Jurusan Teknik Informatika, Universitas Ichsan Gorontalo

Email: andibode22@gmail.com

### **Abstrak**

Perkembangan industri kesehatan terutama farmasi mengalami peningkatan yang pesat. Bersamaan dengan meningkatnya dunia industri farmasi maka informasi produk menjadi masukan bagi perusahaan. Informasi penjualan obat dan informasi persediaan obat. Obat merupakan produk yang dihasilkan dari bahan yang berasal dari tumbuhan, mineral, binatang serta obat syntetis. Prediksi jumlah penjualan di masa yang akan datang bertujuan untuk mengendalikan jumlah stok produk yang ada, sehingga kelebihan stok produk atau kekurangan dapat diminimalkan. Hasil prediksi penjualan akurat maka, pemenuhan permintaan konsumen dapat dipenuhi tepat waktu. Penelitian ini dilakukan untuk memprediksi penjualan obat. Rekapitulasi pesediaan obat adalah permasalahan yang sering dijumpai oleh pihak apotek. Dalam mempermudah prediksi persediaan obat diperiode yang akan datang menggunakan data penjualan obat pada periode sebelumnya. Metode time series sering digunakan sebagai metode peramalan, data time series untuk mengetahui bentuk pola dimasa lalu untuk mengetahui nilai dimasa mendatang. Metode Support Vector Machine tidak bekerja dengan akurat ketika memiliki fitur tidak relevan, karena tidak semua fitur diperlukan. Metode SVM memberikan kinerja yang efektif, jika fitur yang tidak relevan dihapus. Seleksi fitur bekerja secara langsung mengeliminasi fitur serta memilih fitur yang benar memberikan informasi. Pemilahan fitur akan meningkatkan efisiensi. Pada eksperimen algoritma Support Vector Machine dihasilkan model terbaik yang dilihat berdasarkan nilai error terkecil yaitu 0.135 dengan variabel periode 4, validation Shuffled Sampling 10 dan type kernel Polynomial. Kemudian pada eksperimen algoritma Support Vector Machine menggunakan fitur seleksi Forward Selection dihasilkan model terbaik yang dilihat berdasarkan nilai error terkecil yaitu

0.133 dengan variabel periode 8, validation Shuffled Sampling 10 dan type kernel Polynomial.

**Kata kunci:** *Support Vector Machine, Forward Selection, Data Mining, Penjualan Obat*

### **Abstract**

*The development of the health industry, especially pharmaceuticals, has experienced a rapid increase. Along with the increasing world of the pharmaceutical industry, product information is input for the company. Drug sales information and drug supply information. Medication is a product produced from ingredients derived from plants, minerals, animals and syntetic drugs. The prediction of the number of sales in the future aims to control the amount of stock of existing products, so that the excess stock of products or deficiencies can be minimized. The results of accurate sales predictions, fulfillment of consumer demand can be met on time. This research was conducted to predict drug sales. Recapitulation of drug treatment is a problem that is often encountered by pharmacies. In facilitating the prediction of drug availability in the period to come using drug sales data in the previous period. The time series method is often used as a forecasting method, time series data to determine the shape of past patterns to find out the value in the future. The Support Vector Machine method does not work accurately when it has irrelevant features, because not all features are needed. The SVM method provides effective performance, if irrelevant features are deleted. Selection of working features directly eliminates features and selects the correct features that provide information. Sorting features will increase efficiency. In the experiment Support Vector Machine algorithm is produced the best model which is seen based on the smallest error value of 0.135 with period 4 variables, Shuffled Sampling 10 validation and Polynomial kernel type. Then in the Support Vector Machine algorithm experiment using the Forward Selection selection feature produced the best model which is seen based on the smallest error value of 0.133 with period 8 variables, Shuffled Sampling 10 validation and Polynomial kernel type.*

**Keywords:** *Support Vector Machine, Forward Selection, Data Mining, Drug Sales*

## **A. PENDAHULUAN**

Obat merupakan obat yang umum dihasilkan dari bahan-bahan yang berasal dari tumbuh-tumbuhan, mineral, binatang serta obat syntetis. Obat berfungsi sebagai obat untuk jenis penyakit apabila tepat dalam penggunaannya yang disertai dengan dosis, waktu serta tanggal yang tepat (Nangi, j., Indrianti, S.H., Pramono, B., 2018). Perkembangan industri kesehatan terutama farmasi mengalami peningkatan yang pesat. Dalam

hal ini dapat dilihat dari bermunculannya apotek-apotek farmasi. Bersamaan dengan meningkatnya dunia industri farmasi maka informasi produk-produk menjadi masukan bagi perusahaan. Salah satu yaitu informasi penjualan obat dan informasi persediaan obat. Informasi berapa banyak persediaan obat merupakan hal yang sangat penting karena hal ini berkaitan dengan berapa banyak penjualan yang terjadi atau target pasar yang akan dicapai dalam kurun waktu tertentu. Prediksi penjualan adalah kegiatan untuk menapsirkan besarnya jumlah penjualan produk oleh distributor pada wilayah pemasaran dan periode waktu tertentu. Prediksi jumlah penjualan produk di masa yang akan datang bertujuan untuk mengendalikan jumlah stok produk yang ada, sehingga kelebihan stok produk atau kekurangan dapat diminimalkan. Ketika hasil prediksi penjualan akurat maka, pemenuhan permintaan konsumen dapat dipenuhi tepat waktu dengan demikian kerjasama tetap berkesinambungan sehingga dapat mempertahankan konsistensi konsumen (Gustriansyah, R., 2017).

Metode *time series* sering digunakan sebagai metode peramalan, data *time series* untuk mengetahui bentuk pola dimasa lalu untuk mengetahui nilai di masa mendatang. Metode *time series* memiliki kemampuan prediksi yang tinggi, dengan melihat nilai di masa lalu sehingga nilai masa mendatang dapat diketahui (Sularno, 2014). Metode *Support Vector Machine* adalah metode yang berlandaskan pada teori pembelajaran *statistic* yang memberikan hasil lebih baik dibanding metode lain. *Support Vector Machine* bekerja dengan baik terhadap data yang memiliki berdimensi tinggi dengan menggunakan teknik kernel. Kelas metode kernel yang berakar pada teori belajar statistik. Kernel berfungsi sebagai dasar pembelajaran semua algoritma, tujuan umum

algoritma ini mengatasi masalah fungsi kernel tertentu, karena mesin *linear* hanya dapat mengklasifikasi data dalam *linear* ruang fitur terpisah. Parafungsi kernel untuk mendorong sebuah ruang fitur oleh implisit pemetaan (Drajana, 2017).

Metode *Support Vector Machine* tidak bekerja dengan akurat ketika memiliki fitur yang tidak relevan, karena tidak semua fitur diperlukan. Metode *Support Vector Machine* memberikan kinerja yang efektif, jika fitur yang tidak relevan dihapus. Seleksi fitur bekerja secara langsung mengeliminasi fitur serta memilih fitur yang benar memberikan informasi. Pemilahan fitur akan meningkatkan efisiensi. Tujuan utama seleksi fitur adalah pencarian sebuah data yang relevan. Fitur yang dianggap tidak relevan atau fitur berlebihan sangat mempengaruhi terhadap hasil (Poonguzhali, E., Vijayalakshmi, M., Shamily, K., Priyadarshini, V., 2014). Penelitian ini memiliki tujuan untuk menghasilkan model prediksi jumlah penjualan obat dengan tingkat *error* lebih kecil dengan *Feature Selection* menggunakan *Forward Selection* pada *Support Vector Machine*.

## **B. METODE**

*Praprocessing* data univariat *time series* numerik data harian, kemudian data di ubah ke *ascending* ke *descending* proses ini dilakukan dengan memanfaatkan toolbar *sort & filter* yang ada pada *Microsoft Office Excel*, setelah itu data dinormalisasi. Pada basis data, proses normalisasi memiliki tujuan untuk mengurangi resiko terjadinya anomaly data serta tidak konsistensi data. Normalisasi database memiliki tujuan utama yaitu untuk memperoleh data yang berdimensi kecil namun tetap dapat

mewakili data asli dan tidak kehilangan karakteristik data tersebut (Drajana, 2017).

Persamaan normalisasi dapat dilihat sebagai berikut.

$$N = \frac{(X - \min)}{(max - \min)} \quad (1)$$

Dimana:

**N** = Normalisasi

**x** = Data

**Min** = Data Minimum

**Max** = Data Maksimum

Setelah data melewati tahap normalisasi, jika untuk mengembalikan ukuran data yang sudah ternormalisasi ke ukuran data asli adalah dengan cara denormalisasi dataset.

Persamaan sebagai berikut:

$$D = Y (max - min) + min \quad (2)$$

Dimana:

**D** = Denormalisasi

**Y** = Hasil Keluaran Dari Pelatihan

**Min** = Data Minimum

**Max** = Data Maksimum

Penelitian ini melalui beberapa tahapan. Pada tahap ke-1 data diubah dari *ascending* ke *descending*. Tahap ke-2 melakukan normalisasi data. Tahap ke-3 data diubah dari *univariate* ke *multivariate*. Tahap ke-4 menentukan parameter *Support Vector Machine* dan *forward selection*. Penentuan parameter *Support Vector Machine* melewati beberapa pengujian, seperti penentuan jumlah variabel *input independent*, variabel periode *xt-i* serta pemilihan *type kernel*, uji coba ini digunakan untuk

menemukan model terbaik dengan melihat tingkat nilai *root mean square error* (RSME) terkecil. Tahap eksperimen digunakan beberapa sampel data set dalam melakukan uji coba, kemudian dilanjutkan dengan penetapan parameter sebelum memulai proses uji coba nilai parameter yang digunakan dalam proses uji coba sama dengan nilai training yaitu nilai variabel input atau variabel periode  $xt-i$  dari 1 – 10, *validation Shuffled sampling* 10 dengan *type kernel polynomial* kemudian dilakukan proses testing untuk mengetahui hasil *root mean square error*.

Sumber data set pada penelitian ini di ambil dari PT. EXA MEDICA Tangerang Selatan Banten pada 01 Mei 2018. Data yang dikumpulkan adalah data kuantitatif harian *time series* univariat. Jenis data ini adalah data Penjualan Obat dari tahun 2011 – 2013. Sesuai dengan ketentuan jam kerja yang ada di perusahaan bahwa dataset yang ada hari libur dan hari minggu tidak terjadi transaksi pertahunnya. Dataset yang diambil dari variabel penjualan obat memiliki 863 *record* yang berasal dari setiap pelanggan.

Data *mining* adalah proses menjalankan satu atau lebih teknik *machine learning* dan bertujuan menganalisis dan mengekstraksi pengetahuan secara otomatis (Widayun, H., Nasution, D.S., Silalahi, N., Mesran., 2017). Pengertian lainnya adalah *induction-based learning* dimana proses pembentukan definisi-definisi konsep umumnya yang dilakukan dengan mengobservasi contoh spesifik dari konsep yang dipelajari. *Knowledge Discovery in Databases* (KDD) yaitu penerapan dari metode saintifik yang ada pada data mining. Data mining juga bisa diartikan dengan satu langkah dari proses *Knowledge Discovery In Database* (Muktamar, A.B., Setiawan, A.N., Adji, B.T., 2015).

Metode *Forward Selection* adalah metode seleksi maju, termasuk dalam algoritma pencarian paling sederhana. *Forward Selection* didasarkan pada model *Regresi Linear*. *Forward Selection* merupakan teknik untuk mereduksi dimensi data serta menghilangkan atribut yang tidak relevan. Metode model ini diawali dengan nol *variable*, selanjutnya *variable* dimasukan *one by one* hingga kriterianya terpenuhi, jadi prosedur *Forward Selection* diawali dengan tidak ada variabel didalam model (Astuti, P.A., Sudiby, U., Kurniawan, W.A., Rahayu, Y., 2018).

Metode *Support Vector Machine* merupakan metode yang berdasar pada teori pembelajaran *statistic* dengan tujuan memberikan hasil lebih baik dibandingkan metode lain. *Support Vector Machine* bekerja baik terhadap data yang memiliki dimensi tinggi dengan memanfaatkan teknik kernel. Kernel memiliki fungsi sebagai dasar pembelajaran semua algoritma. Secara umum algoritma ini mengatasi kendala fungsi kernel tertentu, karena mesin *linear* hanya dapat mengklasifikasi data dalam *linear* ruang fitur terpisah. Fungsi kernel mendorong sebuah ruang fitur oleh implisit pemetaan (Drajana, 2017).

Proses pembelajaran *Support Vector Machine* yaitu menentukan *support vector*, mengetahui fungsi kernel yang digunakan, serta tidak perlu mengetahui fungsi *non-linear*.

Persamaan *Support Vector Machine* yaitu:

$$(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + \mathbf{b} \quad (3)$$

Dimana:

$\mathbf{b}$  = Bias

$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D)^T$  = Variabel Input

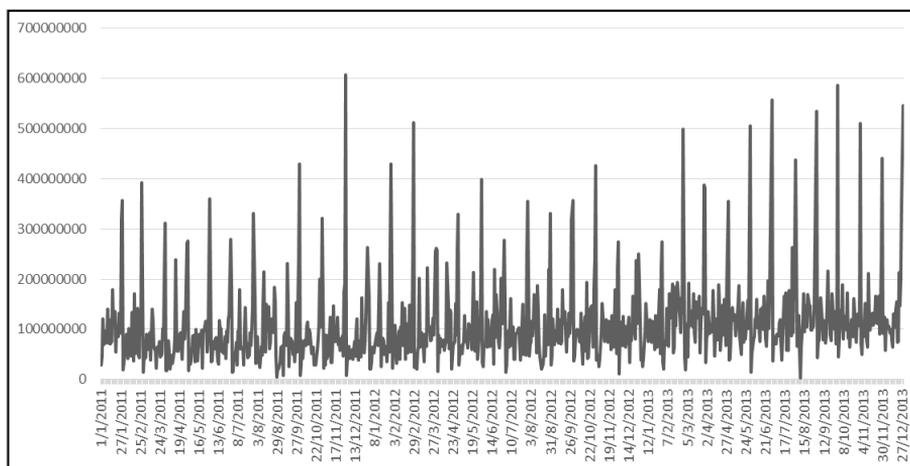
$\mathbf{w} = (\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_D)^T$  = Parameter Bobot

$\phi(\mathbf{x})$  = Fungsi Transformasi fitur

Alat dan bahan yang digunakan dalam penelitian ini berupa *hardware*, *software* dan ATK. *Hardware* meliputi, PC (Laptop), buku-buku literatur, polpen dan kertas. *Software* berupa *tools RapidMiner* untuk pengolahan data mencari nilai *error* terkecil, *Microsoft Office Word* 2016 untuk penyusunan naskah, *Microsoft Office Excel* 2016 untuk merekap data, *ascending* ke *descending*, dan mengubah data *univariate* ke data *multivariate*.

### C. HASIL DAN PEMBAHASAN

Tujuan dari penelitian ini untuk menerapkan algoritma *Support Vector Machine* menggunakan *Forward Selection* pada data Penjualan Obat dengan tujuan mencari model yang terbaik dalam memprediksi *time series* jumlah penjualan obat.



Gambar 1. Grafik Data Penjualan Obat Tahun 2011-2013

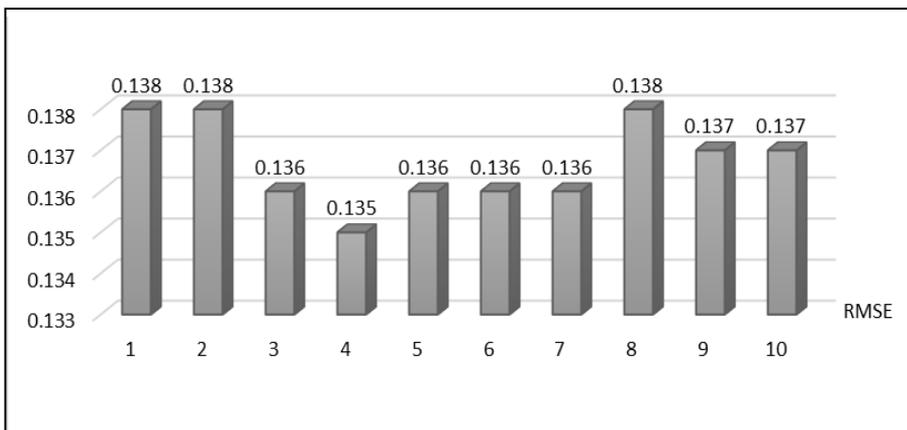
Gambar 1 menggambarkan bahwa terjadi fluktuasi penjualan obat sehingga terjadi masalah dalam persediaan jumlah stok obat. Grafik diatas menjelaskan bahwa data tersebut termasuk dalam data *nonlinear*. Data

yang bentuknya bukan dalam kategori kelas *linear* dalam implementasinya memanfaatkan pendekatan *kernel*.

**Tabel 1. Nilai RMSE Support Vector Machine**

Variabel Periode	Validation Shuffled Sampling	Type Kenel	RMSE
1	10	Polynimial	0.138
2	10	Polynimial	0.138
3	10	Polynimial	0.136
<b>4</b>	<b>10</b>	<b>Polynimial</b>	<b>0.135</b>
5	10	Polynimial	0.136
6	10	Polynimial	0.136
7	10	Polynimial	0.136
8	10	Polynimial	0.138
9	10	Polynimial	0.137
10	10	Polynimial	0.137

Sesuai dengan table 1 hasil eksperimen uji coba penjualan obat dengan pemilihan model menggunakan algoritma *support vector machine* (SVM) mulai dari 1 sampe 10 variabel periode untuk data penjualan obat menggunakan *validation Shuffled Sampling* 10 dan *type kernel polynomial* yang diambil berdasarkan tingkat nilai *root mean square error* (RMSE) paling terkecil. Maka ditemukannya model terbaik yang dilihatkan berdasarkan tingkat nilai error terkecil yang bernilai 0,135 dengan jumlah variabel periode 4.



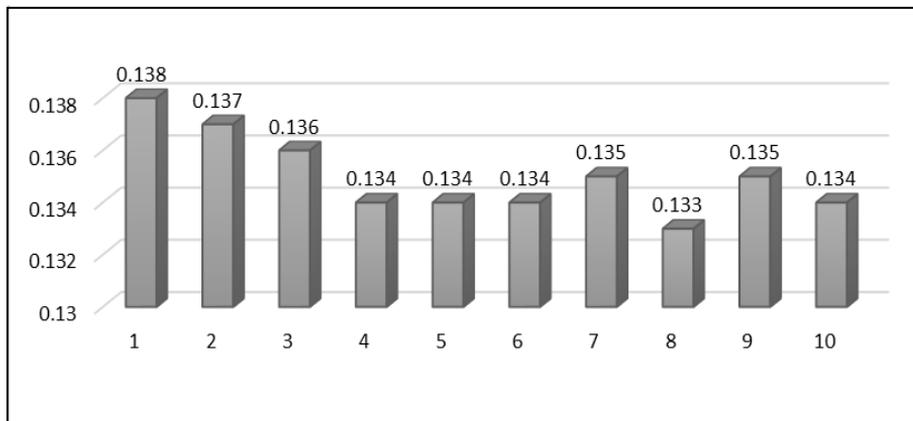
**Gambar 2. Grafik Model Algoritma Support Vector Machine**

Sesuai dengan gambar 2 yaitu grafik model yang paling baik untuk data penjualan obat, model algoritma *support vector machine* (SVM) menggambarkan bahwa model terbaik terdapat pada variabel periode 4, dengan *Validation Shuffled Sampling* 10 dan *type kernel polynomial* yaitu tingkat RMSE sebesar 0,135.

**Tabel 2. Nilai RMSE Support Vector Machine & Forward Selection**

Variabel Periode	Validation Shuffled Sampling	Type Kenel	RMSE Forward Selection
1	10	Polynimial	0.138
2	10	Polynimial	0.137
3	10	Polynimial	0.136
4	10	Polynimial	0.134
5	10	Polynimial	0.134
6	10	Polynimial	0.134
7	10	Polynimial	0.135
<b>8</b>	<b>10</b>	<b>Polynimial</b>	<b>0.133</b>
9	10	Polynimial	0.135
10	10	Polynimial	0.134

Sesuai dengan table 2 rangkuman hasil eksperimen uji coba penjualan obat dengan pemilihan model menggunakan algoritma SVM dan *forward selection* mulai dari 1 sampai 10 variabel periode untuk data penjualan obat menggunakan *validation Shuffled Sampling* 10 dan *type kernel polynomial* yang diambil berdasarkan tingkat nilai *root mean square error* (RMSE) paling terkecil. Maka ditemukannya model terbaik yang dilihat berdasarkan tingkat nilai error terkecil yang bernilai 0,133 dengan jumlah variabel periode 8.



**Gambar 3. Grafik Model Algoritma Support Vector Machine & Forward Selection**

Sesuai dengan gambar 3 yaitu grafik perbandingan model yang paling baik untuk data penjualan obat, SVM dan *forward selection* menggambarkan bahwa model terbaik algoritma SVM terdapat pada variabel periode 8, dengan *Validation Shuffled Sampling* 10 dan *type kernel polynomial* yaitu tingkat RMSE sebesar 0,133.

Prediksi jumlah penjualan obat berdasarkan data penjualan algoritma *Support Vector Machine* menggunakan *Feature Selection* telah selesai dilakukan. Pada eksperimen menggunakan algoritma *Support Vector Machine* dihasilkan model terbaik yang dilihat berdasarkan nilai

*error* terkecil yaitu 0.135 dengan variabel periode 4, validation *Shuffled Sampling* 10 dan *type* kernel *Polynomial*. Kemudian pada eksperimen algoritma *Support Vector Machine* menggunakan fitur seleksi *Forward Selection* dihasilkan model terbaik yang dilihat berdasarkan nilai *error* terkecil yaitu 0.133 dengan variabel periode 8, validation *Shuffled Sampling* 10 dan *type* kernel *Polynomial*. Sesuai hasil eksperimen yang ada, menjelaskan bahwa *feature selection* yaitu *Forward Selection* menghasilkan performa yang lebih baik.

Prediksi jumlah penjualan produk dimasa yang akan datang bertujuan untuk mengendalikan jumlah stok produk yang ada, sehingga kelebihan stok produk atau kekurangan dapat diminimalkan. Manfaat penelitian ini Ketika hasil prediksi penjualan akurat maka, pemenuhan permintaan konsumen dapat dipenuhi tepat waktu dengan demikian kerjasama tetap berkesinambungan sehingga dapat mempertahankan konsistensi konsumen.

Kelebihan dari penelitian ini adalah penelitian ini menghasilkan hasil eksperimen nilai *error* lebih kecil dibandingkan dengan peneliti sebelumnya. Kelemahan dari penelitian ini terletak pada data, karena data yang ada memiliki banyak *record* dan kekosongan nilai pada hari libur ataupun tanggal merah. Untuk peneliti selanjutnya agar bisa meneliti dan menghasilkan nilai *error* yang lebih kecil dan lebih lagi bekerja keras pada pengolahan data.

## **D. PENUTUP**

### **Simpulan dan Saran**

Prediksi jumlah penjualan obat menggunakan algoritma *support Vector Machine* dengan *Feature Selection* menggunakan *Forward*

*Selection* telah berhasil dilakukan. Melihat hasil eksperimen yang dilakukan, menjelaskan seleksi fitur *Forward Selection* meningkatkan performa yang lebih baik dibandingkan dengan algoritma *Support Vector Machine* tanpa menggunakan *feature selection*. Saran dalam penelitian ini untuk penulis atau peneliti selanjutnya untuk menggunakan algoritma lain bertujuan menghasilkan tingkat *error* lebih kecil.

## DAFTAR PUSTAKA

- Astuti P. A, Sudibyo U, Kurniawan W. A, Rahayu Y. 2018. Algoritma Naive Bayes Dengan Fitur Seleksi Untuk Mengetahui Hubungan Variabel Nilai Dan Latar Belakang Pendidikan, Universitas Dian Nuswantoro, *Jurnal SIMETRIS*, Vol. 9 No. 1, April, ISSN: 2252-4983.
- Drajana, Ivo C. R. 2017. Metode *Support Vector Machine* Dan *Forward Selection* Prediksi Pembayaran Pembelian Bahan Baku Kopra, Universitas Ichsan Gorontalo, *ILKOM Jurnal Ilmiah*, Volume 9 Nomor 2 Agustus, ISSN: 2087-1716.
- Gustriansyah, R. 2017. Analisis Metode *Single Exponential Smoothing* Dengan *Brown Exponential Smoothing* Pada Studi Kasus Memprediksi Kuantiti Penjualan Produk Farmasidi Apotek. *Seminar Nasional Teknologi Informasi dan Multimedia*. STMIK AMIKOM Yogyakarta. Februari. ISSN: 2302-3805.
- Muktamar A.B, Setiawan A.N, Adji B.T. 2015. Pembobotan Korelasi pada Naive Bayes Classifier, STMIK AMIKOM Yogyakarta, *Seminar Nasional Teknologi Informasi dan Multimedia*, Februari, ISSN : 2302-3805.
- Nangi, J., Indrianti, S.H., & Pramono, B. 2018. Peramalan Persediaan Obat Menggunakan Metode Triple Exponential Smoothing (tes) (Studi Kasus: Instalasi Farmasi RSUD Kab. Muna). *Jurnal semanTIK*. Vol.4. No.1. Jan-Jun. ISSN: 2502-8928. pp. 135-142.

- Poonguzhali. E, Vijayalakshmi. M, Shamily. K, Priyadarshini. V. 2014. Hybrid feature selection algorithm for high dimensional database, Pondicherry university india, *International Journal Of Engineering Trends And Technology (IJETT)* – Volume 8 Number 9.
- Sularno, A. 2014. Prediksi Nilai Saham Menggunakan Pemograman Genetika Dan Pemograman Ekspresi Gen, Universitas Gunadarma Depok. Indonesia.
- Widayun H, Nasution D.S, Silalahi N, Mesran. 2017. Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5, *Media Informatika Budidarma*, Vol 1, No 2, Juni, ISSN: 2548-8368, Hal 32-37.