



Prediksi Waktu Kedatangan Pelanggan Servis Kendaraan Bermotor Berdasarkan Data Historis menggunakan *Support Vector Machine*

Benni Agung Nugroho^{#1}, Andika Kurnia Adi Pradana^{#2}, Ellya Nurfarida^{#3}

[#]Manajemen Informatika, PSDKU Politeknik Negeri Malang di Kota Kediri
Jl. Lingkar Maskumambang No.1, Sukorame, Kota Kediri, Jawa Timur

¹benni.agung@polinema.ac.id

²andika.kurnia@polinema.ac.id

³ellya.nurfarida@polinema.ac.id

Abstrak— Dealer kendaraan perlu menjaga hubungan baik dengan pelanggan sehingga inti bisnis dealer dapat berlanjut dan berkembang. Salah satu strategi yang digunakan adalah memprediksi kapan pelanggan akan berkunjung lagi untuk servis kendaraan (layanan perawatan atau perbaikan kendaraan) berdasarkan analisis data riwayat kunjungan pelanggan. Dengan hasil prediksi berupa hari kedatangan pelanggan dimasa depan maka dealer kendaraan dapat mengingatkan pelanggan tentang kapan waktunya servis kendaraan. *Support vector machine* (SVM) adalah sebuah model pembelajaran mesin (*machine learning*) yang menggunakan *hyperplane* dan *support-vector* untuk memisahkan kelas dalam suatu ruang dimensi secara optimal sehingga sesuai untuk digunakan dalam pemecahan masalah prediksi waktu kedatangan pelanggan. SVM diimplementasikan untuk memprediksi kapan pelanggan akan datang lagi dimasa depan untuk perbaikan atau perawatan kendaraan. Hasil menunjukkan bahwa, dengan pemilihan metode yang tepat, SVM dapat memprediksi waktu kedatangan pelanggan dengan tingkat akurasi mencapai 92.5% berdasarkan validasi *K-Fold cross-validation* pada data latih dan mencapai rata-rata 97.33% untuk pengukuran nilai presisi, akurasi dan *recall* pada data uji

Kata kunci— *Support Vector*, SVM, Klasifikasi, Prediksi, *Hyperplane*, Kedatangan Pelanggan

I. PENDAHULUAN

Authorized dealer kendaraan adalah perusahaan yang berperan sebagai mitra distributor kendaraan dari produsen ke konsumen atau yang disebut sebagai agen penjualan. Inti bisnis dari dealer kendaraan selain jual-beli kendaraan dan suku cadang, juga melayani jasa servis berupa perawatan dan perbaikan kendaraan[1]. Jasa servis memberikan keuntungan yang besar bagi dealer kendaraan dikarenakan perawatan kendaraan perlu dilakukan secara rutin agar kondisi kendaraan selalu prima. Pelanggan akan secara berkala datang ke dealer kendaraan yang menyediakan jasa pemeliharaan yang memuaskan. Oleh sebab itu, dealer

kendaraan perlu untuk menjaga hubungan yang baik dengan pelanggan agar inti bisnis dapat terus berjalan dan ditingkatkan. Untuk menjaga hubungan yang baik dengan pelanggan maka dealer kendaraan perlu menerapkan strategi dalam mengelola hubungan dengan pelanggan, disebut juga dengan *customer relationship management* (CRM)[2]. CRM melakukan analisis data terhadap riwayat pelanggan dengan perusahaan untuk meningkatkan hubungan bisnis dengan pelanggan dengan tujuan menjaga agar pelanggan tetap loyal terhadap perusahaan sehingga dapat meningkatkan profit/keuntungan yang didapat perusahaan[3]. Salah satu strategi dalam CRM adalah memprediksi kapan pelanggan akan datang lagi untuk melakukan servis kendaraan. Dengan mengetahui waktu perkiraan kapan pelanggan akan datang untuk servis kendaraan maka dealer dapat mengingatkan pelanggan untuk melakukan servis kendaraan sebagai bentuk kepedulian dealer kepada pelanggan.

Support vector Machine (SVM) merupakan salah satu model dalam *machine learning* yang dapat digunakan untuk klasifikasi suatu data menjadi beberapa kelas dengan konsep yang secara matematis lebih matang dibanding *artificial neural network* (ANN)[4]. Penelitian mengenai SVM untuk prediksi telah banyak dilakukan, diantaranya, penelitian tentang SVM dan *genetic algorithm* (GA) untuk memprediksi waktu kedatangan bus menggunakan variabel input berupa waktu, panjang jalan, kecepatan bus, cuaca dan kepadatan jalan raya[5]. penelitian tentang prediksi beban listrik pada hari berikutnya di sektor bangunan menggunakan SVM menggunakan perangkat sub-metering[6], prediksi apakah seorang wanita menderita kanker payudara berdasarkan riwayat perawatan dan gejala-gejala yang muncul[7], dan prediksi tentang pergerakan harga saham pada bursa efek[8] berdasarkan data *time series*

Berdasarkan pada penelitian tentang SVM yang dilakukan oleh peneliti terdahulu[5][6][7][8] yang menunjukkan SVM dapat memberikan hasil prediksi yang

lebih baik dibanding metode lain, maka pada penelitian ini diusulkan pendekatan penyelesaian masalah menggunakan SVM untuk memprediksi kapan pelanggan akan datang lagi ke dealer untuk perbaikan atau perawatan kendaraan berdasarkan data historis pelanggan dimana didalam data historis tersebut dapat diekstrak berbagai macam fitur yang dapat digunakan untuk membuat model prediksi kedatangan pelanggan. Penelitian yang dilakukan diharapkan memberikan kontribusi pada bidang *machine learning* terutama pada metode SVM untuk mendukung CRM pada dealer kendaraan dimana kontribusi dalam penelitian ini adalah :

1. Penentuan fitur yang digunakan pada metode SVM untuk memprediksi kedatangan pelanggan ke dealer kendaraan berdasarkan data historis pelanggan.
2. Penentuan kernel yang sesuai pada metode SVM berdasarkan hasil pengujian *K-Fold cross-validation* dan berdasarkan nilai akurasi, presisi dan *recall* untuk melakukan prediksi kedatangan pelanggan
3. Penentuan model SVM yang bisa digunakan untuk memprediksi kedatangan pelanggan pada dealer kendaraan berdasarkan data historis pelanggan
4. Hasil penelitian menjadi salah satu metode alternatif dalam CRM dalam meningkatkan hubungan pelanggan dan perusahaan untuk meningkatkan keuntungan perusahaan

II. TINJAUAN PUSTAKA

A. Support Vector Machine

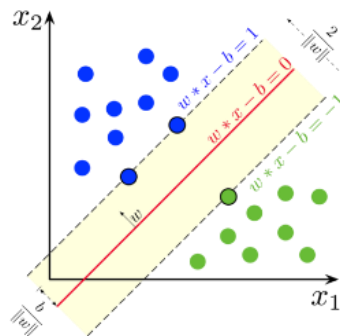
Didalam *machine learning*, *support vector machine* (SVM)[9][10] adalah model algoritma pembelajaran terawasi yang digunakan untuk analisa data untuk kebutuhan klasifikasi dan regresi yang dikembangkan di AT&T Bell Laboratory bersama dengan koleganya[11], SVM merupakan salah satu metode terbaik untuk melakukan prediksi yang berdasarkan kerangka kerja pembelajaran statistik yang dikembangkan oleh Vapnik and Chervonenkis.

B. Linear SVM

SVM merupakan salah satu algoritma terbaik untuk penyelesaian masalah klasifikasi. SVM membagi kelas data menggunakan garis pemisah atau *hyperplane*[12] yang secara umum dinyatakan dalam persamaan 1

$$w \cdot x + b = 0 \tag{1}$$

Hyperplane dalam SVM terdiri dari satu garis pemisah utama dan dua garis pendukung seperti ditunjukkan oleh Gambar 1. Dengan adanya dua garis pendukung inilah yang membuat SVM lebih baik dibanding dengan algoritma pembelajaran lainnya[13] dalam memisahkan kelas. Untuk mendapatkan garis pemisah utama pada SVM maka yang dilakukan pertama adalah mencari dua garis pendukung antara dua kelas menggunakan persamaan 2 dan 3.



Gambar 1. *Hyperplane* pada SVM yang memisahkan dua kelas

$$x_i \cdot w + b \geq 1, \text{ jika } y_i = 1 \tag{2}$$

$$x_i \cdot w + b \leq -1, \text{ jika } y_i = -1 \tag{3}$$

dimana b adalah besaran skalar, w merupakan *normal vector*, dan x_i merupakan vektor data yang bersinggungan dengan garis pendukung yang disebut dengan *support vector*. Garis pemisah optimum didapatkan dengan memaksimalkan margin antara dua garis pendukung dengan menyelesaikan masalah optimasi menggunakan persamaan 4 dan 5, yaitu dengan mencari titik minimal dari persamaan 2 dan 3[14]

$$\min(w) = \frac{1}{2} \|w\|^2 \tag{4}$$

dengan batasan

$$y_i(x_i \cdot w + b) - 1 \geq 0 \tag{5}$$

Masalah optimasi pada persamaan (4) dan (5) dapat diselesaikan menggunakan *Lagrange multiplier*[15]. Setelah masalah optimasi berhasil diselesaikan maka kelas data x_i dapat ditentukan berdasarkan persamaan (6)

$$f(x_d) = \sum_{i=1}^{ns} a_i y_i x_i x_d + b \tag{6}$$

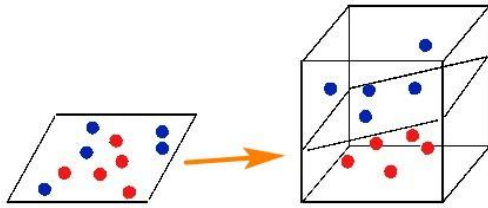
Dimana x_i adalah *support vector*, ns =urutan *support vector*, x_d adalah data yang akan diklasifikasi dan a_i adalah *Lagrange multiplier*

C. Non-linear SVM

Persamaan (1) sampai dengan (6) tersebut dapat digunakan apabila data terpisah secara linear. Pada permasalahan di dunia nyata, himpunan data pada umumnya terpisah secara non-linear. Untuk menyelesaikan permasalahan non linear ini maka dapat digunakan *kernel trick* dimana himpunan data dipetakan pada ruang dimensi yang lebih tinggi[16] dengan menggunakan persamaan 7

$$\Phi: R^p \rightarrow R^q \text{ dengan } p < q \tag{7}$$

sehingga garis pemisah / *hyperplane* dibentuk dalam ruang vektor yang baru (Gambar 2)



Gambar 2. Mapping data ke dimensi yang lebih tinggi

Tetapi untuk menentukan formula transformasi Φ itu sendiri cukup sulit. Untuk itu, perhitungan dot product dapat digantikan dengan fungsi kernel $K(x_i, x_j)$ yang mendefinisikan fungsi transformasi Φ tersebut yang disebut dengan *kernel trick*[17] yang ditunjukkan dalam persamaan 8

$$K(x_i, x_j) = \Phi_i(x_i) \cdot \Phi_j(x_j) \tag{8}$$

Dengan menggunakan *kernel trick* maka perlu mengetahui fungsi kernel yang dipakai untuk menentukan *support vector*. Pada umumnya terdapat 4 jenis fungsi kernel yang dapat digunakan[18][19], yaitu :

1) Kernel linear, persamaan 9

$$K(x, x_k) = x_k^T x \tag{9}$$

2) Kernel polynomial, persamaan 10

$$K(x, x_k) = (x_k^T x + 1)^d \tag{10}$$

3) Kernel sigmoid, persamaan 11

$$K(x, x_k) = \tanh[\kappa x_k^T x + \Phi] \tag{11}$$

4) Kernel radial basis function, persamaan 12

$$K(x, x_k) = \exp\{-\|x - x_k\|_2^2 / \sigma^2\} \tag{12}$$

Setelah menentukan kernel yang akan dipakai, maka proses klasifikasi dengan menggunakan kernel dapat dilakukan dengan menggunakan persamaan 13:

$$\begin{aligned} f(\Phi(x)) &= w \cdot \Phi(x) + b \\ &= \sum_{i=1}^{ns} a_i y_i \Phi(x) \cdot \Phi(x_i) + b \\ &= \sum_{i=1}^{ns} a_i y_i K(x, x_i) + b \end{aligned} \tag{13}$$

dimana *SV* adalah data pada himpunan data latih yang terpilih sebagai *support vector*

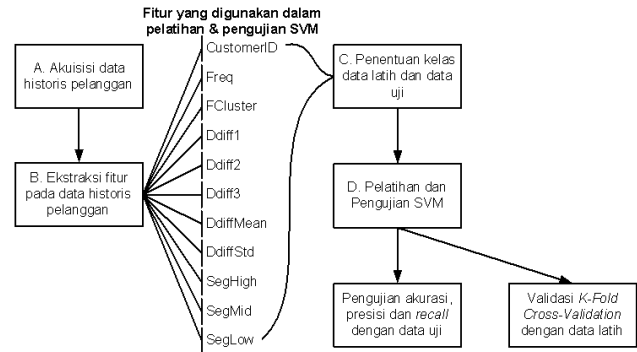
III. METODE PENELITIAN

Tahapan metode penelitian yang dilakukan untuk memprediksi kedatangan pelanggan dimasa depan ditunjukkan oleh Gambar 3 dan dapat dijelaskan sebagai berikut :

A. Akuisisi Data

Data yang digunakan untuk proses pelatihan dan uji coba didapatkan dari dealer kendaraan yang berisi informasi tentang ID pelanggan dan tanggal kedatangannya. Data yang digunakan adalah data 6 tahun yang berjumlah 1044 data dengan 150 pelanggan unik dari tanggal awal servis

pada tanggal 8-Februari-2014 sampai dengan tanggal 31-Mei-2020, Tabel 1 menunjukkan contoh data tersebut



Gambar 3. Tahapan metode penelitian

TABEL I
CONTOH RAW DATA UNTUK PELATIHAN DAN PENGUJIAN

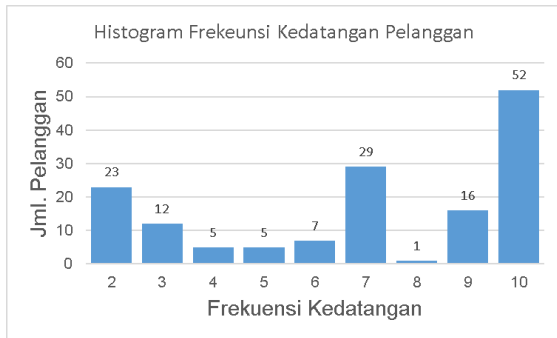
1	CustomerID	TglService
2	414	31/05/2020
3	614	19/05/2020
4	501	18/05/2020
5	101	07/05/2020

B. Ekstraksi Fitur

Sebelum dapat dilakukan proses pelatihan ataupun uji coba, maka perlu ditentukan fitur-fitur yang digunakan dalam proses pelatihan ataupun uji coba[20][21]. Fitur yang akan digunakan untuk proses pelatihan ataupun uji coba haruslah fitur yang informatif yang dapat memberikan informasi bagi pengguna. Didalam proses pelatihan ini digunakan sebanyak 11 fitur yang ditunjukkan oleh Tabel 2.

TABEL II
FITUR UNTUK PROSES PELATIHAN DAN PENGUJIAN

No	Nama Fitur	Deskripsi
1	CustomerID	ID customer
2	Freq	Menghitung frekuensi kedatangan setiap pelanggan
3	FCluster	Segmentasi pelanggan berdasarkan nilai frekuensi. Berdasarkan pada persebaran nilai <i>Frequency</i> pada histogram Gambar 4 maka <i>FCluster</i> dibagi ke dalam 6 cluster
4	Ddiff1	Mencari selisih hari antara servis terakhir pelanggan (<i>t</i>) dengan servis sebelumnya (<i>t-1</i>)
5	Ddiff2	Mencari selisih hari antara (<i>t-1</i>) pelanggan dengan servis sebelumnya (<i>t-2</i>)
6	Ddiff3	Mencari selisih hari antara (<i>t-2</i>) pelanggan dengan servis sebelumnya (<i>t-3</i>)
7	DdiffMean	Menghitung nilai mean/rata-rata selisih waktu kunjungan tiap pelanggan
8	DdiffStd	Menghitung nilai standar deviasi selisih waktu kunjungan tiap pelanggan
9	SegHigh	Segmentasi pelanggan kedalam kategori ini bila masuk dalam $FCluster > 3$
10	SegMed	Segmentasi pelanggan kedalam kategori ini bila masuk dalam $1 < FCluster \leq 3$
11	SegLow	Segmentasi pelanggan kedalam kategori ini bila masuk dalam $FCluster \leq 1$



Gambar 4. Grafik histogram frekuensi kedatangan pelanggan

Setelah melalui ekstraksi fitur, maka didapatkan 115 data pelanggan unik yang dapat digunakan dalam proses pelatihan dan pengujian. Hal ini dikarenakan, untuk memprediksi waktu kedatangan pelanggan di masa depan dibutuhkan data *time series*[22] dimana untuk membentuk data *time series* minimal terdapat 4 data kunjungan pelanggan ke dealer agar pelanggan tersebut memiliki data historis[23] yang cukup untuk membentuk *time series*. Data *time series* diekstrak ke dalam fitur *Ddiff1*, *Ddiff2* dan *Ddiff3* yang menghitung waktu selisih antar kedatangan pelanggan dengan kedatangan berikutnya.

C. Penentuan Kelas pada Data Latih dan Data Uji

Setelah proses ekstraksi fitur, maka langkah selanjutnya adalah menentukan kelas pada setiap data latih ataupun data uji[24] yang memenuhi kriteria, yaitu yang fiturnya lengkap. Kelas kedatangan pelanggan dibagi menjadi 5, yaitu kelas 0 sampai kelas 4 dengan penentuan kelas berdasarkan nilai *DdiffMean*. Untuk penjelasan setiap kelas dapat dilihat pada Tabel 3

TABEL III
DESKRIPSI SETIAP KELAS KEDATANGAN PELANGGAN

No	Kelas	Deskripsi
1	0	Pelanggan akan datang lagi setelah ≤ 90 hari dari tanggal servis terakhir
2	1	Pelanggan akan datang lagi setelah ≤ 180 hari dari tanggal servis terakhir
3	2	Pelanggan akan datang lagi setelah ≤ 270 hari dari tanggal servis terakhir
4	3	Pelanggan akan datang lagi setelah ≤ 360 hari dari tanggal servis terakhir
5	4	Pelanggan akan datang lagi setelah ≥ 360 hari dari tanggal servis terakhir

Setelah setiap data ditentukan kelasnya, maka setiap data tersebut siap untuk digunakan sebagai data latih maupun data uji. Untuk contoh data latih yang dapat digunakan dalam proses pelatihan dapat dilihat pada Gambar 5.

CustomerID	Freq	FCluster	Ddiff	Ddiff2	Ddiff3	DdiffMean	DdiffStd	Segment_High	Segment_Low	Segment_Mid	Kelas	
0	410	4	1	416.0	717.0	1052.0	350.666667	59.079043	0	1	0	3
1	510	4	1	358.0	680.0	988.0	329.333333	25.794056	0	1	0	3
2	210	4	1	323.0	660.0	994.0	331.333333	7.371115	0	1	0	3
3	110	4	1	345.0	666.0	993.0	331.000000	12.489996	0	1	0	3
4	610	4	1	278.0	682.0	994.0	331.333333	65.186911	0	1	0	3
5	501	7	3	260.0	419.0	712.0	292.000000	186.719040	0	0	1	3
6	101	7	3	211.0	439.0	678.0	283.000000	177.488028	0	0	1	3
7	401	7	3	184.0	446.0	678.0	283.833333	184.861480	0	0	1	3
8	601	7	3	189.0	436.0	689.0	279.333333	177.171856	0	0	1	3
9	208	7	3	257.0	655.0	1115.0	314.500000	113.279742	0	0	1	3

Gambar 5. Contoh data latih model SVM

D. Pelatihan dan Pengujian SVM

Proses pelatihan dijalankan sebanyak 4 kali dengan setiap pelatihan menggunakan kernel SVM yang berbeda. Untuk kernel yang digunakan sesuai penjelasan mengenai algoritma SVM maka akan digunakan kernel linear, sigmoid, polynomial dan RBF. Setiap model yang dihasilkan dari proses pelatihan akan divalidasi tingkat akurasinya menggunakan teknik validasi *K-Fold cross-validation*[25]. Sedangkan untuk pengujian dengan data uji (data yang tidak digunakan dalam proses pelatihan) maka setiap model akan dinilai tingkat akurasi, presisi dan *recall*-nya

IV. HASIL DAN PEMBAHASAN

Dataset terdiri dari 115 buah data, sekitar 70% (80 buah data) dipilih secara acak untuk digunakan sebagai data latih dalam proses pelatihan dan sisanya 30% (35 buah data) digunakan sebagai data uji dalam proses pengujian model SVM.

Proses pelatihan melibatkan 4 buah kernel SVM yang akan dibandingkan, yaitu kernel linear, polinomial, RBF dan sigmoid. Sedangkan untuk menilai akurasi setiap kernel SVM digunakan metode *K-Fold cross-validation* dengan membagi himpunan data latih menjadi 10 subhimpunan secara acak. Dari subhimpunan tersebut setiap subhimpunan akan berkesempatan menjadi data testing ($1/k$) dan subhimpunan lainnya menjadi data latih ($(k-1)$). Berdasarkan dari *K-Fold cross-validation* maka didapatkan akurasi setiap kernel SVM dalam melakukan prediksi seperti ditunjukkan oleh Tabel 4.

TABEL IV
HASIL UJI KERNEL SVM DENGAN VALIDASI *K-FOLD CROSS-VALIDATION*

Kernel SVM	Akurasi rata-rata (%)	Standar Deviasi (%)
Linear	92.50	8.29
RBF	60.00	14.57
Polinomial	88.75	8.75
Sigmoid	60.00	14.57

Berdasarkan dari hasil di Tabel 4, dapat dilihat bahwa setiap kelas dapat dipisahkan secara linear yang memberikan batasan-batasan setiap kelas dengan lebih akurat dibanding dengan pemisah non-linear. Kernel linear memberikan nilai akurasi prediksi yang paling tinggi yaitu 92.5% bila dibanding kernel lainnya dan memiliki standar deviasi yang paling kecil dibanding yang lain, yaitu 8.29% yang menandakan margin *error* paling kecil dibanding menggunakan kernel SVM yang lain.

- Control for Chemical Engineers*, 2013.
- [16] “Separating Hyperplane Theorem,” in *Encyclopedia of Operations Research and Management Science*, 2013.
- [17] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, “An introduction to kernel-based learning algorithms,” *IEEE Transactions on Neural Networks*. 2001.
- [18] C. C. Chang and C. J. Lin, “LIBSVM: A Library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, 2011.
- [19] D. Meyer, “Support vector machines: the interface to libsvm in package e1071,” ... *Syst. their ...*, 2014.
- [20] N. K. Verma and A. Salour, “Feature extraction,” in *Studies in Systems, Decision and Control*, 2020.
- [21] J. Kittler, “Feature selection and extraction.,” *Handb. pattern Recognit. image Process.*, 1986.
- [22] N. Sapankevych and R. Sankar, “Time series prediction using support vector machines: A survey,” *IEEE Comput. Intell. Mag.*, 2009.
- [23] X. Ma, Y. Tian, C. Luo, and Y. Zhang, “Predicting Future Visitors of Restaurants Using Big Data,” *Proc. - Int. Conf. Mach. Learn. Cybern.*, vol. 1, no. May, pp. 269–274, 2018.
- [24] D. R. S. R Manikandan, “Machine Learning Algorithms for Classification,” *Int. J. Acad. Res. Dev.*, 2018.
- [25] D. Berrar, “Cross-validation,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2018.
- [26] A. Tharwat, “Classification assessment methods,” *Appl. Comput. Informatics*, 2018.