

KOMPARASI NEURAL NETWORK DAN SUPPORT VECTOR MACHINE UNTUK DATA TIME SERIES DAN NON-TIME SERIES

Suamanda Ika Novichasari¹, Restu Rakhmawati²

SI Teknologi Informasi, Universitas Tidar

Jl. Kapten Suparman No.39, Potrobangsari, Kec. Magelang Utara, Kota Magelang, Jawa Tengah 56116

¹manda@untidar.ac.id

²resturakhma@untidar.ac.id

Abstrak— *Neural Network dan Support Vector Machine merupakan metode datamining yang sering digunakan. Penelitian ini bertujuan untuk mengetahui performa dari Neural Network dan Support Vector Machine yang diterapkan pada data time series dan non-time series. Sehingga terlihat perbedaan dan keunggulan dari kedua metode tersebut. Data yang digunakan merupakan dataset publik, "Australian Credit Approval dan Polar Ice Data". Untuk tahap validasi model menggunakan 10fold cross-validation dan proses evaluasi model menggunakan Root Mean Square Error (RMSE). Hasil percobaan membuktikan bahwa pada data time series SVM lebih unggul dari NN dilihat dari kinerja dan waktu eksekusinya, sedangkan pada data non-time series NN lebih unggul. Hasil akhir evaluasi percobaan data time series berbanding terbalik dengan hasil percobaan data non-time series..*

Kata kunci— *Time series, Non-time series, Neural Network, Support Vector Machine, klasifikasi kelayakan kredit, Prediksi Polar Es.*

I. PENDAHULUAN

Pembiayaan perekonomian nasional sangat dipengaruhi oleh kredit dalam perbankan, hal tersebut juga berfungsi sebagai penggerak pertumbuhan ekonomi [1]. Penilaian kelayakan kredit bank bukan merupakan hal yang mudah, karena bagi bank sangat penting untuk membedakan antara UKM baik dan buruk dalam hal kredit melalui proses penilaian kredit tersebut [2]. Likuiditas yang semakin rendah dalam kondisi keuangan yang sangat lemah dengan jumlah kredit bermasalah yang semakin besar dapat menyebabkan bank kesulitan untuk membiayai kegiatan usahanya [3]. Seperti yang tercantum dalam Gatra News, Lembaga Penjamin Simpanan (LPS) menyatakan bahwa sejak delapan tahun terakhir ini, dari 2006 hingga 2013, sebanyak 55 bank dilikuidasi, yang terdiri dari Bank Perkreditan Rakyat (BPR) dan 1 bank umum, serta menyelamatkan Bank Century yang kini bersulih nama menjadi Bank Mutiara [4]. Oleh sebab itu proses penilaian kelayakan kredit sangat penting dilakukan untuk memperkecil jumlah kredit bermasalah sehingga kelangsungan hidup sebuah bank tetap terjaga. Analisa resiko kredit yang dilakukan adalah dengan mengklasifikasikan data menjadi dua kelas, yaitu kredit

baik dan kredit buruk. Sehingga tepat menggunakan teknik klasifikasi data mining. Metode yang paling populer digunakan untuk teknik klasifikasi adalah Neural Net (NN), Support Vector Machines (SVM), dan lain lain [5].

Mencairnya es di kutub laut Artik dapat dijadikan ukuran terhadap pemanasan global di bumi [6]. Berkurangnya luas daerah es di laut Artik yang merupakan dampak dari pemanasan global mengakibatkan semakin tinggi permukaan air laut dan perubahan iklim yang ekstrim. Oleh karena itu prediksi luas daerah es di kutub penting dilakukan sehingga dapat digunakan untuk dasar dari peningkatan pencegahan pemanasan global [7].

Neural Network (NN) dan Support Vector Machine (SVM) merupakan algoritma yang dapat digunakan untuk klasifikasi dan prediksi. SVM memiliki kelemahan membutuhkan waktu eksekusi yang lama sehingga kurang efisien jika digunakan pada data yang berukuran besar [8]. NN dan SVM terbukti menghasilkan kinerja yang sangat baik untuk masalah klasifikasi terutama untuk kelayakan kredit [9]. Beberapa peneliti telah menggunakan metode NN dan SVM untuk klasifikasi dan prediksi [10][11][12][13][14].

Hadianto, N. dkk, pada tahun 2019 menerapkan metode NN pada dataset publik dari <http://www.kaggle.com> dengan jumlah data 5000 observasi mengenai perilaku nasabah pada bank dalam periode tertentu. Percobaan dilakukan dengan algoritma backpropagation dalam Neural Network menggunakan struktur 12 – 15 – 8 – 1 , training cycles sejumlah 500, learning rate 0.01 dan momentum 0.1 menghasilkan nilai akurasi = 98.24%, dengan AUC sebesar =0,979 [10].

Ramdhani, Y., & Mubarak, A. pada tahun 2019 menerapkan SVM untuk prediksi harga saham PT. Antm.JK. Hasil penelitian menunjukkan nilai terendah RMSE dari algoritma SVM dengan nilai 22.662. Kemudian dilakukan optimasi parameter dengan menggunakan algoritma genetika sehingga nilai terendah RMSE didapat sebesar 10.495 [11].

Nugroho, B. A., dkk pada tahun (2021) menggunakan algoritma SVM untuk memprediksi waktu kedatangan pelanggan untuk servis kendaraan pada dealer kendaraan. Hasilnya dengan metode pengujian Cross Validation,

Confusion Matrix dan Kurva ROC terbukti SVM memiliki akurasi sebesar 97.33% [12].

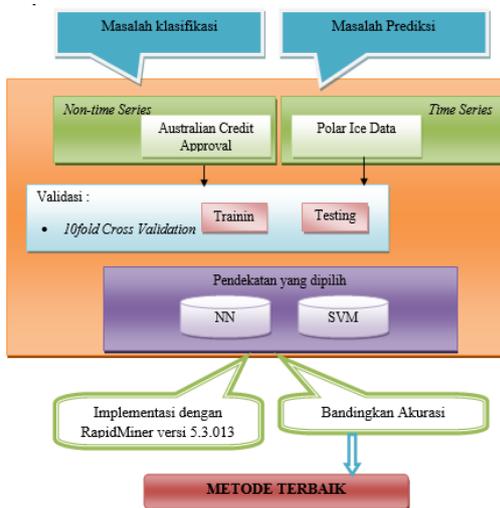
Purnamaswari, A. A. A., dkk pada tahun 2022 menerapkan metode NN dan metode SVM untuk memprediksi volume sampah pada TPA Suwung. Hasil dari peramalan volume sampah menggunakan Metode Backpropagation mencapai tingkat kesalahan terkecil yaitu 0.048 sedangkan peramalan volume sampah menggunakan Metode SVM mencapai tingkat kesalahan yaitu 0.108. [13].

Herdiansah, A., dkk pada tahun 2021 menerapkan metode Back Propagation dari Neural Networks untuk mengklasifikasikan citra daun herbal berdasarkan ekstraksi ciri dan bentuk. Hasil percobaan dari penelitian ini menunjukkan hasil akurasi rata-rata dari daun jambu biji, sirih, papaya, kunyit sebesar 88,75% [14].

II. METODE PENELITIAN

A. Metode yang diusulkan

NN dan SVM sudah terbukti dapat di terapkan untuk masalah klasifikasi dan prediksi. Namun SVM memiliki kelemahan yaitu lamanya waktu yang dibutuhkan, sehingga kurang tepat digunakan untuk data berukuran besar. Data set non-times series yang digunakan pada penelitian ini memiliki ukuran yang cukup besar namun berbeda dengan data time series.



Gambar 1. Metode penelitian

Penelitian ini akan membuktikan pada data non-time series, metode NN akan mengungguli SVM, sedangkan pada data time series akan terjadi sebaliknya.

B. Dataset

Data non-time series yang digunakan pada penelitian ini berasal dari University of California, Irvine (UCI) Machine Learning dengan judul Australian Credit Approval. Data ini lebih menjurus untuk aplikasi kartu kredit. Disediakan oleh See Quinlan, data ini berjumlah

1000 record dan terdiri dari 14 atribut bebas, dengan 6 atribut bertipe numerik dan 8 bertipe kategorikal dan 1 atribut terikat bertipe biner (0-1). Atribut yang bertipe kategorikal diubah menjadi numerik seperti berikut ini :

TABEL I
Konversi Australian Credit Approval Dataset

Atribut	Awal	Konversi	Tipe
A1:	a,b	0,1	Kategorikal
A2:			Kontinu
A3:			Kontinu
A4:	p,g,gg	1,2,3	Kategorikal
A5:	ff,d,i,k,j,aa,m,c,w, e, q, r,cc, x	1, 2,3,4,5, 6,7,8,9,10,11,12,13,14	Kategorikal
A6:	ff,dd,j,bb,v,n.o,h,z	1, 2,3, 4,5,6,7,8,9	Kategorikal
A7:			Kontinu
A8:	t, f	1, 0	Kategorikal
A9:	t, f	1, 0	Kategorikal
A10:			Kontinu
A11:	t, f	1, 0	Kategorikal
A12:	s, g, p	1, 2, 3	Kategorikal
A13:			Kontinu
A14:			Kontinu
A15:	1, 2	1, 0	(class attribute)

Data ini juga pernah digunakan sebagai dataset penelitian oleh See Quinlan pada tahun 1987 dan 1992.

Untuk data time series digunakan dataset dari website new.censusatschool.org.nz berjudul "Polar Ice Data". Data berisi jumlah luas daerah yang tertutup es di laut Artik dan Antartika. Periode data ini adalah bulanan dari bulan Januari tahun 1990 sampai bulan Maret tahun 2011 berjumlah 255 record. Data yang dipakai dalam laporan ini hanya untuk laut Artic saja, sehingga data ini menjadi bertipe univariate time series. Kemudian diciptakan 4 variabel baru yaitu y-1, y-2, y-3, dan y-4 seperti terlihat dalam contoh data dibawah ini :

TABEL II
Konversi Polar Ice Data

Y	Y t-1	Y t-2	Y t-3	Y t-4
12.72	13.33	13.44	12.16	10.84
13.33	13.44	12.16	10.84	9.12
13.44	12.16	10.84	9.12	6.44
12.16	10.84	9.12	6.44	4.92

10.84	9.12	6.44	4.92	4.5
9.12	6.44	4.92	4.5	6.67
6.44	4.92	4.5	6.67	9.58
4.92	4.5	6.67	9.58	11.4

Y adalah data es asli laut Artik, y t-1 adalah data es di 1 tahun sebelumnya, y t-2 data es di 2 tahun sebelumnya, y t-3 untuk 3 tahun sebelumnya dan y t-4 untuk 4 tahun sebelumnya. Dari 255 record, di pakai 251 record saja untuk menghindari missing value.

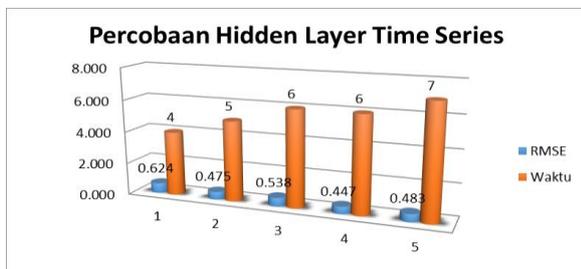
III. HASIL DAN PEMBAHASAN

Tujuan utama penelitian ini adalah untuk mengetahui nilai akurasi dari algoritma NN dan SVM menggunakan Australian credit Approval dan Polar Ice Data. Kemudian membandingkan kedua algoritma tersebut sehingga dapat diperoleh salah satu algoritma yang terbaik.

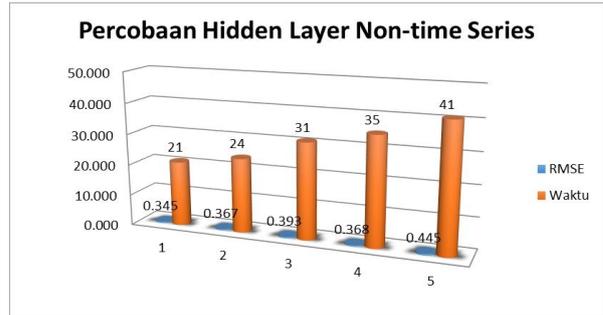
Desain model NN dapat dilihat pada gambar 2. Retrieve berfungsi untuk memasukan dataset ke dalam RapidMiner. Validation menggunakan tenfold cross-validation. Didalam validation terdapat dua kolom, training dan testing. Didalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu Neural Net yang merupakan implementasi dari Back Propagation Neural Network, sedangkan di dalam kolom testing terdapat Apply Model untuk menjalankan algoritma/model NN dan Performance untuk mengukur performa dari model NN tersebut. Desain model SVM sama seperti model NN namun pada kolom training didalam validation berisi Support Vector Machine. Data yang digunakan sudah bertipe numerik jadi tidak perlu fungsi tambahan untuk mengubah atribut nominal ke numerik atau yang lainnya. Karena metode NN dan SVM hanya bisa digunakan untuk atribut bertipe numerik.

A. Percobaan Neural Net

Percobaan pertama dilakukan untuk mencari hidden layer terbaik. Dari 5 kali percobaan hidden layer, untuk data time series, hidden layer yang di pilih adalah 2, sedangkan untuk non-time series yang dipilih adalah 1. Hasil percobaan dapat di lihat pada gambar grafik di bawah ini.

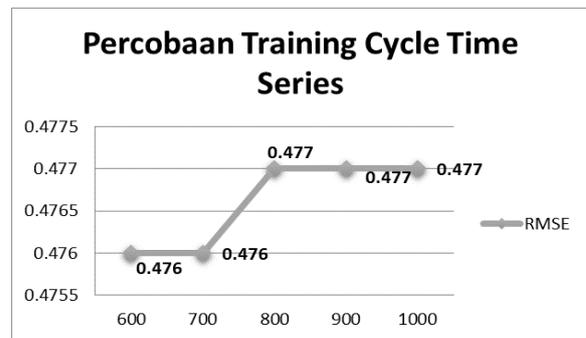


Gambar 4. Grafik hasil hidden layer time series

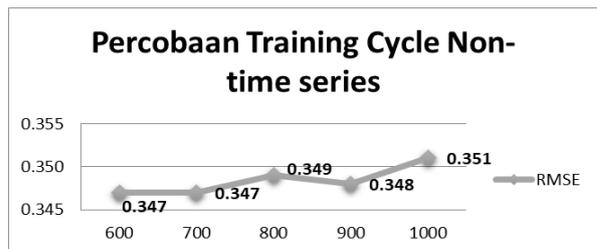


Gambar 5. Grafik hasil hidden layer non-time series

Setelah ditentukan jumlah hidden layer, kemudian di lakukan percobaan untuk menemukan training cycle untuk performa terbaik. Training cycle adalah jumlah iterasi yang dilakukan untuk mendapatkan bobot maksimal. Percobaan dilakukan dengan angka 600, 700, 800, 900, dan 1000. Untuk data time series dan non-time series sama-sama mendapatkan training cycle terbaik pada jumlah 600 dan 700, namun berdasarkan waktu eksekusi yang lebih singkat maka di pilih 600.

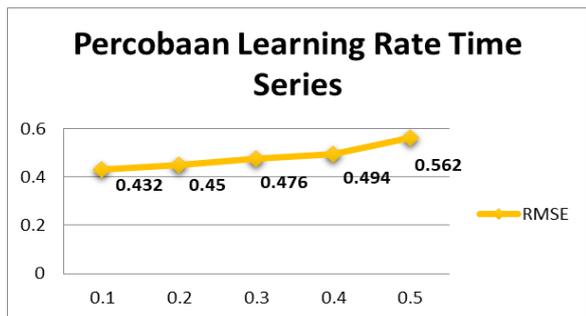


Gambar 6. Grafik hasil training cycle time series

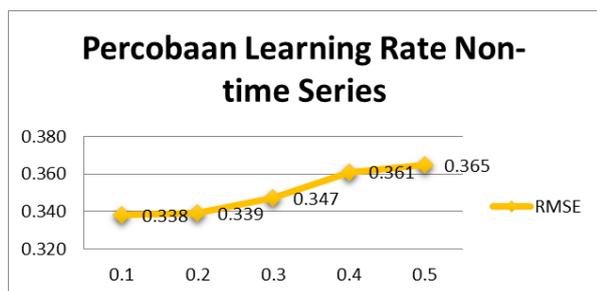


Gambar 7. Grafik hasil training cycle non-time series

Setelah itu dilakukan percobaan untuk menentukan Learning Rate yang menunjukkan berapa kali bobot di ubah dalam setiap putaran. Dari rentang 0,1 sampai 0,5, untuk data time series didapatkan nilai 0,1 dan non-time series sama-sama di dapatkan learning rate sebesar 0,1 .

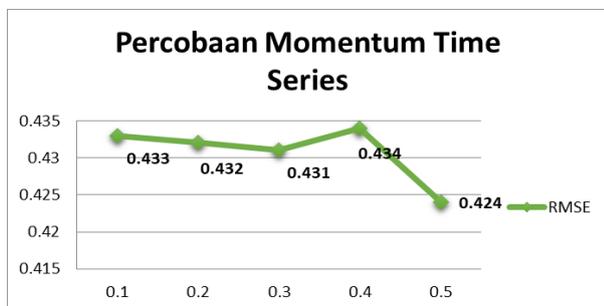


Gambar 8. Grafik hasil Learning rate time series

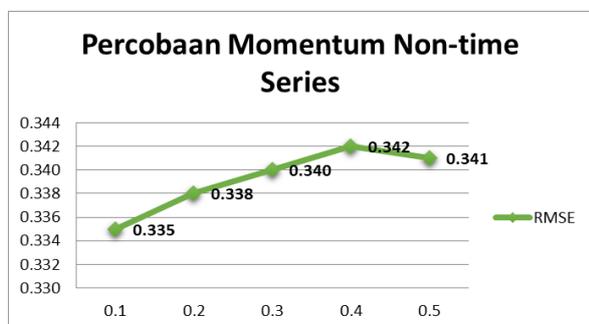


Gambar 9. Grafik hasil Learning Rate non-time series

Terakhir, dilakukan percobaan untuk menentukan momentum untuk mencegah adanya lokal maksimum. Dari percobaan 0,1 sampai 0,5 nilai momentum yang dilakukan, untuk data time series didapatkan nilai momentum 0,5 sedangkan untuk non-time series 0,1.



Gambar 10. Grafik hasil momentum time series



Gambar 11. Grafik hasil momentum non-time series

Hasil dari semua percobaan adalah untuk data time series dan non-time series dapat dilihat dari tabel dibawah ini.

TABEL III
Hasil Percobaan Time Series

Time Series					
hidden layer	Training cycle	Learning Rate	Momentum	RMSE	waktu
2	600	0,1	0,5	0,424	5

TABEL IV
Hasil Percobaan Non-Time Series

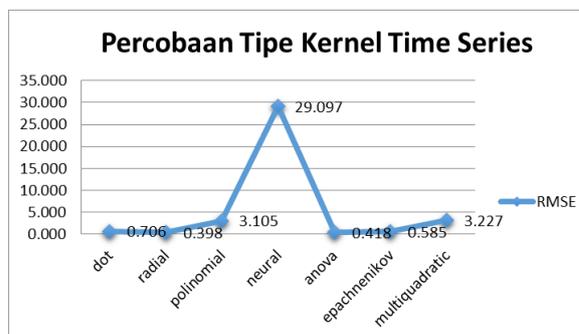
Non-time Series					
hidden layer	Training cycle	Learning Rate	Momentum	RMSE	waktu
1	600	0,1	0,1	0,335	23

B. Percobaan SVM

Tidak seperti percobaan NN, Percobaan SVM hanya dilakukan 1 kali untuk menentukan tipe kernel yang terbaik. Untuk data time series didapatkan tipe anova, sedangkan untuk non-time series didapatkan tipe anova. Seperti terlihat pada tabel dan gambar grafik dibawah ini.

TABEL V
Hasil Percobaan Kernel Time Series

Time Series		
Kernel	RMSE	waktu
dot	0,706	1
radial	0,398	1
polinomial	3,105	0
neural	29,097	0
anova	0,418	2
epachnenikov	0,585	1
multiquadratic	3,227	1

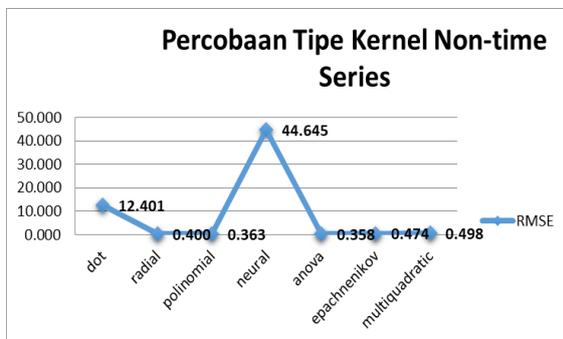


Gambar 12. Grafil Percobaan kernel Time series

TABEL VI

Hasil Percobaan Kernel Non-Time Series

Non-time Series		
Kernel	RMSE	waktu
dot	12,401	12
radial	0,400	7
polinomial	0,363	13
neural	44,645	10
anova	0,358	177
epachnenikov	0,474	4
multiquadratic	0,498	2



Gambar 13. Grafil Percobaan kernel Non-Time series

Definisi dari kernel-kernel diatas adalah sebagai berikut :

- **dot**: merupakan definisi dari $k(x,y)=x*y$.
- **radial**: definisi dari $\exp(-\gamma \|x-y\|^2)$ dimana γ adalah gamma, yang merupakan spesifikasi dari parameter kernel gamma.
- **polynomial**: merupakan definisi dari $k(x,y)=(x*y+1)^d$ dimana d adalah degree of polynomial dari parameter kernel degree. Kernel polynomial cocok digunakan jika semua data training sudah di normalisasi.
- **neural**: definisi dari 2 layer neural net $\tanh(ax*y+b)$ dimana a adalah alpha and b adalah nilai konstan.
- **anova**: merupakan definisi dari meningkatnya kekuatan d dari penjumlahan $\exp(-\gamma (x-y))$ dimana γ adalah gamma and d adalah degree dari parameter kernel gamma dan kernel degree.
- **epachnenikov**: adalah function $(3/4)(1-u^2)$ untuk u diantara -1 and 1 and nol untuk u diluar range tersebut.
- **multiquadric**: adalah definisi dari akar kuadrat dari $\|x-y\|^2 + c^2$.

IV. KESIMPULAN

Penelitian ini menggunakan algoritma NN dan SVM menggunakan dataset publik time-series dan non-time series. Penelitian ini berfokus pada penerapan algoritma NN dan SVM dalam teknik klasifikasi data mining untuk kredit aproval dan teknik prediksi untuk data es di laut Artik. Validasi model menggunakan 10fold cross-validation dan evaluasi model menggunakan Root Mean Square Error (RMSE).

Hasil penelitian membuktikan bahwa pada data time series SVM lebih unggul dari NN dilihat dari kinerja dan waktu eksekusinya, sedangkan pada data non-time series NN lebih unggul.

Untuk penelitian selanjutnya sebaiknya menggunakan lebih banyak algoritma untuk dibandingkan seperti Decision Tree, K-nn, atau yang lainnya. Atau menggunakan banyak dataset yang berbeda dan menambahkan metode pengukuran lain untuk mempermudah dalam menentukan algoritma yang terbaik.

UCAPAN TERIMA KASIH / ACKNOWLEDGMENT

Penulis mengucapkan terima kasih Universitas Tidar yang telah memberikan kesempatan kepada penulis untuk berkarya.

REFERENSI

- [1] Utari D, G.A., Arimurti, T., dan Kurniati, I.K (2012). "Pertumbuhan Kredit Optimal," Buletin Ekonomi Moneter dan Perbankan. Oktober 2012. 3-36.
- [2] Guidici, P., and Figini, S (2009). Applied Data Mining for Business and Industry. 2nd ed. United Kingdom: A John Wiley And Sons, Ltd., Publication.
- [3] Triandaru, S., dan Budisantoso, T (2009). Bank dan Lembaga Keuangan Lain. Edisi kedua. Jakarta : Salemba Empat.
- [4] <http://www.gatra.com/ekonomi-1/44068-lps-selama-8-tahun,-55-bank-dilikuidasi.html>, diakses tanggal 26 Maret 2014.
- [5] Gorunescu, F. (2011). Data Mining Concepts, Models And Techniques. Verlag Berlin Heidelberg: Springer.
- [6] Polyakov, I. V., Timokhov, L. A., Alexeev, V. A., Bacon, S., Dmitrenko, I. A., Fortier, L., ... & Toole, J. (2010). Arctic Ocean warming contributes to reduced polar ice cap. Journal of Physical Oceanography, 40(12), 2743-2756.
- [7] http://id.wikipedia.org/wiki/Pemanasan_global, diakses tanggal 20 November 2014
- [8] Wu, Xindong and Kumar, Vipin. (2009). The Top Ten Algorithms in Data Mining. Boca Raton: CRC Press.
- [9] Setiono, R., Baesens, B., & Martens, D. (2012). Rule Extraction from Neural Networks and Support Vector Machines for Credit Scoring. In Data Mining: Foundations and Intelligent Paradigms (pp. 299-320). Springer Berlin Heidelberg.
- [10] Hadiano, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network. Jurnal Pilar Nusa Mandiri, 15(2), 163-170.
- [11] Ramdhani, Y., & Mubarak, A. (2019). Analisis time series prediksi penutupan harga saham Antm. Jk dengan algoritma SVM model regresi. Jurnal Responsif: Riset Sains dan Informatika, 1(1), 77-82.
- [12] Nugroho, B. A., Pradana, A. K. A., & Nurfarida, E. (2021). Prediksi Waktu Kedatangan Pelanggan Servis Kendaraan Bermotor Berdasarkan Data Historis menggunakan Support Vector Machine. JEPIN (Jurnal Edukasi dan Penelitian Informatika), 7(1), 25-30.
- [13] Purnamaswari, A. A. A., Putra, I. K. G. D., Putra, S., & Made, I. (2022). Komparasi Metode Neural Network Backpropagation dan Support Vector Machines dalam Prediksi Volume Sampah TPA Suwung. Jurnal Ilmiah Teknologi dan Komputer, 3(1), 853-861.

[14] Herdiansah, A., Borman, R. I., Nurnaningsih, D., Sinlae, A. A. J., & Al Hakim, R. R. (2022). Klasifikasi Citra Daun Herbal Dengan Menggunakan Backpropagation Neural Networks Berdasarkan

Ekstraksi Ciri Bentuk. JURIKOM (Jurnal Riset Komputer), 9(2), 388-395.