

# Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine

Moch. Fauzi Raja Mahendra, Nuril Lutvi Azizah dan Sumarno

Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo Kampus 2  
Jl. Raya Gelam No.250, Pagerwaja, Gelam, Kec. Candi, Kabupaten Sidoarjo, Jawa Timur 61271  
E-mail : rehanarif21@gmail.com , {nurillutviazizah, sumarno}@umsida.ac.id

## Abstrak

Cuaca ialah keseluruhan dari peristiwa yang berlangsung di atmosfer bumi selama beberapa hari. Cuaca yang terjadi dalam jangka waktu lebih lama disebut sebagai iklim. Hal yang dapat memengaruhi keadaan suatu cuaca yakni antara lain: suhu, tekanan udara, kecepatan angin, udara serta curah hujan. Kini iklim di Indonesia sendiri kadang-kadang tidak menentu. Bahkan, ketika cuaca di suatu wilayah cerah, dalam waktu yang tidak lama bisa berubah menjadi hujan atau bahkan badai. Tidak menentunya perubahan pada iklim ini dapat mengakibatkan kesulitan dalam pemrediksian suatu cuaca. Salah satu kemajuan teknologi yang dapat digunakan saat ini adalah machine learning. Machine learning adalah salah satu aplikasi yang menjadi bagian dari Artificial intelligence. Machine learning dalam bahasa Indonesia disebut juga pembelajaran mesin. Salah satu algoritma machine learning yang dapat dilakukan dalam proses klasifikasi ini adalah SVM. Kelebihan dari algoritma ini adalah komputasinya yang cepat. Klasifikasi memprediksi hujan dengan menggunakan algoritma SVM ini mendapatkan nilai dengan akurasi sebesar 83%. Tujuan di adakan penelitian ini adalah untuk menambah akurasi dalam memprediksi hujan.

**Kata kunci** : Machine learning, klasifikasi, memprediksi hujan, klasifikasi, SVM

## Pendahuluan

Cuaca ialah keseluruhan dari peristiwa yang berlangsung di atmosfer bumi selama beberapa hari. Cuaca yang terjadi dalam jangka waktu lebih lama disebut sebagai iklim. Hal yang dapat memengaruhi keadaan suatu cuaca yakni antara lain: suhu, tekanan udara, kecepatan angin, udara serta curah hujan[1].

Iklim adalah kondisi rata-rata cuaca dalam jangka waktu yang lama, sekitar 30 tahun atau lebih, dilansir dari SciJinks. Iklim oleh sistem iklim suatu wilayah. Terdekat lima komponen utama yang di gunakan dalam sistem iklim, antara lain atmosfer, Hidrosfer, kriosfer, permukaan tanah, dan biosfer[1].

Machine learning adalah salah satu aplikasi yang menjadi bagian dari Artificial intelligence. Machine learning dalam bahasa Indonesia disebut juga pembelajaran mesin. Machine Learning di peruntukkan dalam pengembangan sebuah sistem atau mesin cerdas yang dapat belajar sendiri tanpa harus di program oleh manusia secara berulang kali [2].

Penelitian yang dilakukan Sulthoni akbar, candra dewi Randy cahya dengan judul "prediksi Cuaca Kota Denpasar menggunakan algoritma

ELM dengan optimasi quantum delta particle swarm optimization" pada tahun 2021. Penelitian ini menggunakan data cuaca yang di gunakan ialah suhu, kecepatan angina, kelembapan udara[3].

Penelitian yang dilakukan Nur Rochman Darmawan pada tahun 2019 dengan judul "Prediksi Kondisi Cuaca Kota Surabaya Menggunakan Metode Artificial Neural Network" Penelitian ini Menggunakan metode Artificial Neural Network Meliputi input layer yang berisikan satu set data[4].

Penelitian yang dilakukan oleh Yusuf Rizqon Rangkuti, Muhammad Valensyah Alfansyuri dengan judul "Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam memprediksi dan menghitung tingkat akurasi data cuaca di Indonesia" pada tahun 2021 data penelitian ini dikumpulkan dari situs jejaring national centers for environmental information[5].

Adapun penelitian yang dilakukan oleh amril mutoi siregar, tukino sultan faisal, ilham kudori dengan judul 'klasifikasi Untuk Prediksi Cuaca menggunakan Esemble Learning' pada tahun 2020. Penelitian ini menggunakan model confusion matrix untuk menghitung accuracy dan MSE untuk menghitung clasfication yang eror [6].

Cuaca merupakan keadaan udara pada suatu

waktu yang singkat dan tempat yang sempit, kondisi cuaca di suatu tempat dapat di tentukan oleh beberapa faktor, Seperti suhu udara, kelembapan udara,intensif cahaya matahari[7].

Machine learning ialah Suatu mengenai algoritma yang mempelajari suatu hal dalam melaksanakan sejumlah hal tertentu yang secara otomatis dilakukan manusia untuk menyelesaikan bermacam – macam tugas yang ada, ataupun memprediksi simpulan baru yang akurat[8].

Phyton ialah bahasa yang sangat terkenal diantara pengembangan machine, data scientist ataupun data miner, alasannya dikarenakan bahasa ini terlalu mudah serta gratis, di samping itu phyton juga bahasa pemrograman yang interpretative dengan banyak kegunaan, tidak sama dengan bahasa lainnya yang sulit dibaca serta di pahami, bahasa inilebih menekankan pada keterbacaan kode supaya memudahkan pemahaman sintak[9].

Support Vector Machine (SVM) ialah satu dari berbagai metode yang dipergunakan dalam proses pengklasifikasian ataupun regresi, Metode ini mampu melakukan pengklasifikasian ataupun regresi, Metode ini juga mampu melakukan pengklasifikasian masalah secara linier[10].

Metode SVM memiliki sifat non-parametrik dimana umumnya dimanfaatkan untuk mengklasifikasikan data dna mengolah citra , tingkat keakuratan metode ini diambil dari parameter serta kernel, pengguna bisa menentukan parameter di mana pada kernel tiap parameteranya akan mempunyai dampak yang sangat berbeda[11].

Support Vector Machine (SVM) secara konseptual adalah mesin linear yang dilengkapi dengan fitur special[12], dan didasarkan pada metode minimalisasi resiko struktural[13], serta teori pembelajaran statistic. Svm telah banyak di teliti dalam komunitas data mining dan pembelajaran mesin (Machine Learning) selama sepuluh tahun terakhir[14].

Konsep SVM bisa dijelaskan sebagai sebuah cara ataupun metode pencari hyperlane paling baik di mana difungsikan untuk memisahkan dua buah ckass pada input space[15].

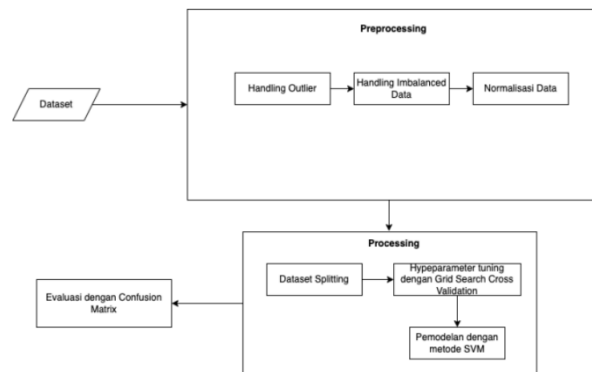
Metode yang bisa dipergunakan dalam proses pemrediksian salah satunya yaitu support vector machine. Metode ini ialah tehnik dengan basis machine learning di mana masih jarang di gunakan dalam proses pemrediksian, baik dipergunakan pada kasus klasifikasi ataupun regresi, serta sangat populer akhir-akhir ini. Support vector machine menjadi salah satu metode yang bisa di gunakan dalam penyelesaian bermacam masalah di antaranya yaitu permasalahan prediksi. Dalam penyelesaian suatu masalah, metode ini dapat mengatasi masalah non linier dengan fungsi kernel yang menjadikan metode ini biasa dipergunakan sebagai peramal time series.

Berdasarkan latar belakang yang diuraikan di atas, maka peneliti tertarik untuk membahas dan mengkaji lebih dalam mengenai penelitian yang berkaitan dengan Prediksi cuaca dengan judul, “Im-

plementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine”.

## Metode Penelitian

Jelaskan Tahapan penelitian merupakan gambaran umum terkait alur penelitian yang akan dilakukan dalam pengerjaan penelitian ini dari awal hingga akhir. Tahapan yang dilakukan dalam penelitian ini dapat di paparkan melalui diagram alir seperti pada gambar 1. Seluruh tahapan ini dikerjakan menggunakan Bahasa pemrograman phython. Dengan rindian untuk tahapan preprocessing hingga pemodelan menggunakan software jupyter notebook.



Gambar 1: Gambaran tahapan penelitian

## Data

Data yang digunakan pada penelitian ini diambil dari website kaggle yang terdiri dari 1461 record dan diambil dari tahun 2012 hingga tahun 2015. Dataset ini terdiri dari 5 indikator. Berikut merupakan rincian indikator atau attribut dari dataset yang akan digunakan.

Tabel 1: Pengujian Blackbox sistem penyewaan

Attribut	Keterangan
Date	Tanggal
Precipitation	Presipitasi
Temp_max	Temperature maksimum
Temp_min	Temperatur minimum
Wind	Kecepatan angin
Weather	Cuaca

## Preprocessing

Data mentah tidak bisa dipergunakan langsung oleh sistem. Dengan begitu, perlu dilakukan beber-

apa preprocessing untuk sedikit melakukan modifikasi data supaya kualitas data yang dipergunakan meningkat. Preprocessing dalam penelitian dilakukan sebagai pembersihan data sebelum dilakukannya modeling.

## Handling Outlier

Handling outlier menggunakan teknik Z-score. Z-score dapat dipergunakan untuk membantu penentuan apakah suatu data memiliki nilai ekstrem, ataupun outlier. Data outlier ialah data dengan jauh dari rata-rata. Aturan umumnya ialah Z-score bernilai kurang dari  $-3$  ataupun lebih dari  $+3$  menjelaskan bahwasanya nilai data bernilai ekstrem. Dengan demikian data dengan nilai lebih dari batas bawah serta batas atas akan dihapuskan.

## Handling Imbalanced

Data Tahap berikutnya yakni handling imbalanced data. Imbalanced data atau data yang tidak seimbang merupakan suatu kondisi dimana sebuah himpunan data terdapat satu kelas yang jumlahnya lebih banyak dibanding kelas lainnya. Pada penelitian ini, peneliti menggunakan Teknik oversampling dengan SMOTE. Berikut merupakan kode sumber dari handling imbalanced data dengan SMOTE.

## Normalisasi Data

Tahap selanjutnya yaitu normalisasi data. Hal ini berguna untuk menghilangkan redundansi data dan menstandarisasi informasi pada data. Peneliti menggunakan metode standard scaler. Standard Scaler adalah teknik pemrosesan data yang umum digunakan dalam analisis data dan pembelajaran mesin untuk mentransformasi fitur-fitur numerik agar memiliki rata-rata nol dan simpangan baku satu. Tujuan utamanya adalah untuk mengubah distribusi data sehingga fitur-fitur memiliki skala yang serupa, sehingga algoritma pembelajaran mesin dapat bekerja lebih efisien

## Processing

Pada sub bab ini akan dijelaskan mengenai proses pengolahan data sehingga mendapatkan hasil yang diharapkan seperti yang telah dijelaskan sebelumnya.

## Klasifikasi Dengan SVM

Tahapan pertama dari pengklasifikasian menggunakan SVM ialah membuka data yang sudah diekstraksi fitur ke dalam jupyter notebook dengan library pandas. Saat data telah berhasil dimuat, selanjutnya data dibagi antara data X dan data y di mana data X ialah kolom fitur sedangkan data y ialah kolom target.

Sesudah dibagi antara data X dengan y, berikutnya dibagi antara data train dengan data test pada data X memakai modul scikit-learn yakni `train_test_split`. Besaran pembagian data yakni 70% untuk data train sedangkan 30% untuk data test.

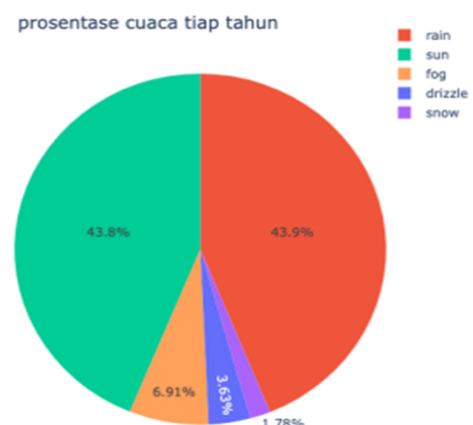
Setelah dilakukan pembagian dataset, selanjutnya adalah tahapan modelling. Untuk memperoleh parameter teroptimal dalam kasus ini, peneliti memanfaatkan Teknik GridSearchCV yang mana Teknik ini mampu melakukan pencarian parameter optimal dari algoritma yang dipergunakan untuk kasus yang tengah dilakukan analisa. GridSearchCV ialah bagian dari modul scikit-learn di mana ditujukan secara otomatis serta sistematis memvalidasi beberapa model dan setiap hyperparameter. Saat proses running GridSearchCV selesai, akan diperoleh model disertai score test dan score train.

## Evaluasi

Tahap ini dipergunakan sebagai pengukur performa dari model machine learning yang telah dibuat. terdapat tiga metrik evaluasi yang dapat digunakan yaitu precision, recall dan confusion matrix. Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada confusion matrix. Dengan adanya perhitungan dari confusion matrix maka dapat diperoleh accuracy, precision, recall dan f1 score.

## Hasil dan Pembahasan

Hasil Data Data yang digunakan pada penelitian ini merupakan data cuaca dari tahun 2012 sampai dengan 2015. Gambar 1 merupakan prosentase cuaca setiap tahunnya. Dari gambar tersebut dapat diketahui bahwa musim cuaca cerah (sun) dan hujan (rain) memiliki prosentase yang hampir sama besarnya dibanding cuaca kabut (fog), salju (snow) dan gerimis (drizzle).

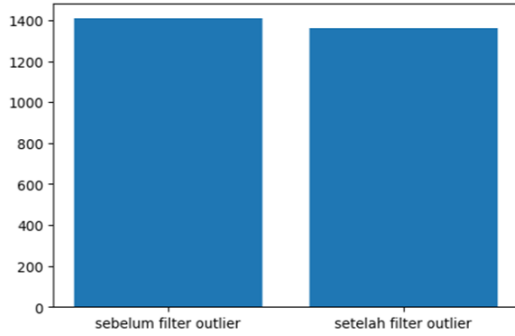


Gambar 2: Prosentase cuaca tiap tahun

## Preprocessing

### Handling Outlier

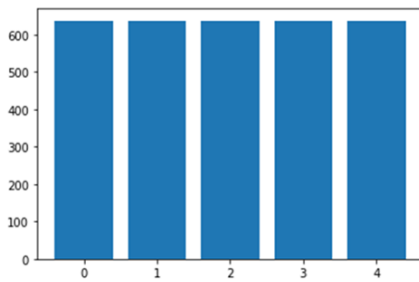
Seperti yang telah dijelaskan pada bab sebelumnya, penelitian ini melakukan filterisasi outlier atau nilai menyimpang terlebih dahulu dengan menggunakan Teknik Z-score. Gambar 2 merupakan hasil visualisasi data sebelum dilakukan outlier dan setelah dilakukan filterisasi outlier.



Gambar 3: filter outlier

### Handling Imbalanced

Data Setelah tahap filter outlier telah selesai, selanjutnya dilakukan handling imbalanced data. Hal ini dikarenakan jumlah kelas yang tidak seimbang. Dapat dilihat pada gambar 1 bahwa jumlah kelas sun dan rain jauh lebih banyak dibanding jumlah kelas lainnya. Peneliti menggunakan Teknik oversampling dengan SMOTE. Berikut merupakan hasil dari oversampling tersebut.



Gambar 4: hasil oversampling

Normalisasi Data Agar range nilai dari setiap kolom sama, peneliti melakukan normalisasi data menggunakan Teknik standard scaler. Berikut merupakan hasil dari normalisasi data.

```
array([[ 0.24731035, -0.59517437, -0.429742,  1.38452078],
       [ 0.88203072, -0.29897175,  0.33808472,  0.48078619],
       [-0.09271841,  0.2261147,  0.88934493,  0.48078619],
       ...,
       [-0.50075294,  1.72059152,  1.20435076, -0.79950448],
       [-0.50075294,  0.60309984,  0.98778425,  0.33016376],
       [-0.31940426, -0.44707306, -0.64630851, -0.4229484 ]])
```

Gambar 5: hasil normalisasi

## Processing

Dataset dibagi menjadi dua bagian yaitu data train dan data test. Prosentase masing-masing dari kedua pembagian data tersebut adalah 70% untuk data train dan 30% untuk data test. Setelah dilakukan pembagian data, selanjutnya dilakukan pemodelan menggunakan metode SVM dengan list parameter yang ada pada gambar 5.

```
# train the model on train set
from sklearn.model_selection import GridSearchCV

# defining parameter range
param_grid = {'C': [0.1, 1, 10, 100, 1000],
              'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
              'kernel': ['rbf']}

grid = GridSearchCV(SVC(), param_grid, refit = True, verbose = 3)

# fitting the model for grid search
grid.fit(X_train, y_train)
```

Gambar 6: Parameter tuning

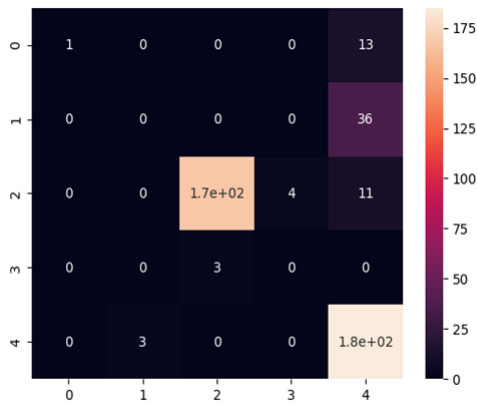
Dari parameter pada gambar 5 di atas, selanjutnya dilakukan tuning dengan grid search cross validation guna mendapatkan parameter terbaik. Gambar 6 merupakan hasil dari tuning parameter dimana pada gambar tersebut diketahui bahwa parameter terbaik pada penelitian ini adalah C=1000 dan gamma = 1 dengan menggunakan kernel rbf. Dari parameter tersebut dihasilkan skor train 86% dan skor test 83%.

```
[32] grid.best_estimator_
SVC
SVC(C=1000, gamma=1)
```

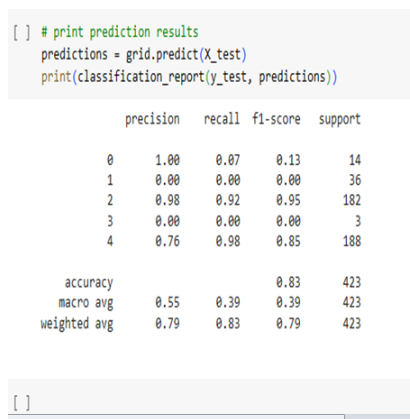
Gambar 7: Parameter tuning

## Evaluasi

Pada tahap evaluasi, peneliti menggunakan confusion matrix. Dapat dilihat pada gambar 6 bahwa nilai akurasi yang dihasilkan pada penelitian ini adalah 83%. sedangkan untuk nilai precisionnya adalah 79% dan F1 Score adalah 79%



Gambar 8: confusion matrix



Gambar 9: classification report

## Penutup

Peneleitian ini telah berhasil melakukan klasifikasi mulai dari preprocessing sampai evaluasi. Berdasarkan penelitian yang telah dilakukan system dapat melakukan klasifikasi prediksi hujan dengan skor akurasi sebesar 83% dan precision yang dihasilkan 0.79% sehingga jika di lihat dari gap antara skor akurasi dan precision model tidak mengalami overfitting maupun underfitting. Dengan kata lain penelitian ini menghasilkan model yang fit dan mendapatkan hasil pengujian sesuai dengan realisasi yang di harapkan. Kelebihan dari metode SVM ini mampu menghasilkan model klasifikasi yang sangat baik.

## Daftar Pustaka

[1] A Novandya dan Isni Oktria, "Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi", *J. Format Vol.6 No.2* (2017) : 98.

[2] F. R. Lumbanraja, R. S. Sani, D. Kurniawan dan A. R. Irawati, "Implementasi Metode Sup-

port Vector Machine Dalam Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung", *J. Komputasi*, vol. 7, no. 2, doi: 10.23960/komputasi.v7i2.2426, 2019.

[3] A. Sulthoni Akbar, C. Dewi dan R. C. Wihandika, "Prediksi Cuaca Kota Denpasar menggunakan Algoritma ELM dengan Optimasi Quantum Delta Particle Swarm Optimization", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 1126–1135, 2021.

[4] Nur Rochman Darmawan, "Prediksi Kondisi Cuaca Kota Surabaya Menggunakan Metode Artificial Neural Network Prediction of Surabaya City Weather Conditions Using Artificial Neural Network Method", *Int. J. Online Biomed. Eng.*, vol. 15, no. 8, 2019.

[5] M. Y. R. Rangkuti, M. V. Alfansyuri dan W. Gunawan, "Penerapan Algoritma K-Nearest Neighbor (Knn) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia", *Hexag. J. Tek. dan Sains*, vol. 2, no. 2, pp. 11–16, doi: 10.36761/hexagon.v2i2.1082, 2021.

[6] A. M. Siregar, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning", *Petir*, vol. 13, no. 2, pp. 138–147, doi: 10.33322/petir.v13i2.998, 2020.

[7] Broto Poernomo dan Rina Dewi Indah Sari, "Penerapan Data Mining Untuk Prakiraan Cuaca Di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (Id3)", *Jouticla*, vol. 2, no. 2, pp. 101–108, doi: 10.30736/jti.v2i2.68, 2017.

[8] W. Wahyudi, W. Purnamasari S, A. Hidayat dan M. M. Fakhri, "Penerapan Machine Learning Pada Mikrokontroler Arduino Mega PRO MINI ATmega2560-16AU", *J. Embed. Syst. Secur. Intell. Syst.*, vol. 3, no. 1, p. 30, doi: 10.26858/jessi.v3i1.33370, 2022.

[9] Ibnu Baqil Id, "Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python", *Journal of Engineering, Computing & Architecture*, vol. 11, no. 4, pp. 82-86, 2021.

[10] D. I. Pushpita Anna Octaviani dan Yuciana Wilandari, "Penerapan Metode SVM Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang", *J. Gaussian*, vol. 3, no. 8, pp. 811–820, 2014.

[11] R. R. P. Putri, M. T. Furqon dan B. Rahayudi, "Implementasi Metode JST - Backpropagation untuk Klasifikasi Rumah Layak Huni (Studi Kasus Desa Kidul Kecamatan Tumpang Kabupaten Malang)", *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. Vol. 2 No., no. 10, pp. 3360–3365, 2018.

- [12] E. Sutoyo, I. T. R. Yanto, R. R. Saedudin, and T. Herawan, "A soft set-based co-occurrence for clustering web user transactions", *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 3, 2017.
- [13] A.W. Yudhistira, B Agus, Kaslani dan R. Nining, "Analisa Klasifikasi Menggunakan Algoritma Decision Tree Pada Data Log Firewall" *Jurnal SIM*, Vol.9 No.3, Desember 2021
- [14] R. Windawati dan H. D. Koeswanti, "Pengembangan Game Edukasi Berbasis Android untuk Meningkatkan hasil Belajar Siswa di Sekolah Dasar", *J. Basicedu*, vol. 5, no. 2, pp. 1027–1038, doi: 10.31004/basicedu.v5i2.835, 2021.
- [15] Honakan, Adiwijaya dan S. A. Faraby, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification Indonesian news", *E-Proceeding of Engineering, Universitas Telkom*, Vol 8, No 8, Hal 26-35. 2018.