JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 4, Oktober 2021, Page 1580-1584
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i4.3347

Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine

Nur Ghaniaviyanto Ramadhan[1,*], Azka Khoirunnisa[2]

[1] Fakultas Informatika, Prodi Rekayasa Perangkat Lunak, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia
[2] Fakultas Informatika, Prodi Informatika, Universitas Telkom, Bandung, Indonesia
Email: [1,*]ghani@ittelkom-pwt.ac.id, [2]azkhoi@telkomuniversity.ac.id
Email Penulis Korespondensi: ghani@ittelkom-pwt.ac.id

Abstrak−Malaria adalah penyakit yang mengancam jiwa, disebabkan oleh parasit yang ditularkan ke manusia melalui gigitan nyamuk Anopheles betina yang terinfeksi. Pada tahun 2019, diperkirakan terdapat 229 juta kasus malaria di seluruh dunia dan jumlah kematian mencapai 409.000. Daerah yang paling sering terjangkit penyakit malaria menurut WHO ada pada wilayah Afrika. Penyakit malaria dapat di deteksi sebelumnya dengan menggunakan informasi yang ada pada data pasien dan menerapkan teknik machine learning. Pada penelitian ini bertujuan untuk deteksi dan klasifikasi penyakit malaria berat berdasarkan histori pemeriksaan data pasien dengan menggunakan metode Support Vector Machine (SVM) dengan sebelumnya dilakukan teknik normalisasi menggunakan min-max pada dataset dan dilakukan teknik cross validation dengan beberapa eksperimen nilai K terhadap hasil. Penelitian ini juga membandingkan metode Support Vector Machine dengan Naïve Bayes (NB) yang dimana hasil akurasi model SVM lebih unggul daripada Naïve Bayes dengan gap akurasi rata-rata 25%. Akurasi yang dihasilkan dengan penerapan metode usulan sebesar 92.3%.

Kata Kunci: Malaria, Klasifikasi; Support Vector Machine; Min-Max; K Cross Validation

Abstract−Malaria is a life-threatening disease, caused by a parasite that is transmitted to humans through the bite of an infected female Anopheles mosquito. In 2019, there were an estimated 229 million cases of malaria worldwide and the death toll reached 409,000. The area most frequently affected by malaria, according to WHO, is the African region. Malaria can be detected beforehand by using the information inpatient data and applying machine learning techniques. This study aims to detect and classify severe malaria based on the history of examining patient data using the Support Vector Machine (SVM) method with a normalization technique using min-max on the dataset and a cross-validation technique with several experiments on the K value of the results. This study also compares the Support Vector Machine method with Naïve Bayes (NB) where the accuracy of the SVM model is superior to Nave Bayes with an average accuracy gap of 25%. The accuracy generated by the application of the proposed method is 92.3%.

Keywords: Malaria; Classification; Support Vector Machine; Min-Max; K Cross Validation

1. PENDAHULUAN

Malaria adalah penyakit yang mengancam jiwa, disebabkan oleh parasit yang ditularkan ke manusia melalui gigitan nyamuk Anopheles betina yang terinfeksi [1]. Pada tahun 2019, diperkirakan terdapat 229 juta kasus malaria di seluruh dunia dan jumlah kematian mencapai 409.000 [1]. Anak-anak di bawah 5 tahun adalah kelompok yang paling rentan terkena malaria [1]. Daerah yang paling sering terjangkit penyakit malaria menurut WHO ada pada wilayah Afrika [1]. Penyakit malaria dapat di deteksi sebelumnya dengan menggunakan informasi yang ada pada data pasien dan menerapkan teknik machine learning. Berbagai penelitian telah dilakukan terkait deteksi penyakit malaria dengan menggunakan berbagai teknik machine learning yang ada seperti Classification and Regression Tree, Naïve Bayes, Artificial Neural Network, Random Forest, bahkan Multi-Layer Perceptron.

Pada penelitian [2] membuat sistem prediksi untuk mengklasifikasikan penyakit malaria berat menggunakan metode Classification and Regression Tree (CART) dan probabilitas komplikasi malaria menggunakan metode Naïve Bayes dengan akurasi tertinggi yang dihasilkan 81.2%. Study [3] membahas tentang riwayat dan gejala pasien malaria dianggap sebagai input data, lalu sistem menganalisis data tersebut dan memprediksi hasilnya. Metode yang digunakan yaitu Artificial Neural Network dengan MLP (Multi-Layer Perceptron) digunakan bersama Back-Propagation, akurasi yang dihasilkan sebesar 85% [3]. Penelitian [4] membahas tentang membangun model machine learning untuk diagnosis penyakit malaria berdasarkan informasi dari pasien, dataset yang digunakan berasal dari abstrak laporan kasus penyakit non-parasit (kanker, Alzheimer, penyakit rheumatoid, dan diabetes) yang diterbitkan dari 1956 to 2019 by PubMed [5] dan dataset yang sesuai dengan 56 laporan penyakit parasit yang disediakan oleh Pusat Pengendalian dan Pencegahan Penyakit (CDC) [6]. Pada study lain [7] mengusulkan model berbasis pembelajaran mesin untuk klasifikasi malaria menggunakan variabilitas iklim di enam negara Afrika Sub-Sahara selama periode dua puluh delapan tahun.

Pada paper [8] bertujuan untuk membandingkan kinerja prediktif peta prevalensi yang dihasilkan menggunakan model Bayesian Decision Network (BDN) dan model regresi logistik bertingkat dalam hal akurasi prediksi risiko spasial malaria. Penelitian [8] menggunakan dataset yang dikumpulkan dari 77 desa yang dipilih secara acak untuk menentukan hubungan prevalensi Plasmodium falciparum dan Plasmodium vivax dengan curah hujan, suhu, ketinggian, kemiringan (aspek medan), peningkatan indeks vegetasi dan jarak ke pantai. Study [9] menyelidiki pengaruh faktor iklim terhadap kejadian malaria di distrik Sundargarh, Odisha, India. Model yang digunakan yaitu MLP dan J48, serta membandingkan penerapan teknik testing yang digunakan (fold cross-validation, percentile split, and supplied test) dengan akurasi yang dihasilkan 71% [9]. Paper lain [10]

Submitted: **07/10/2021**; Accepted: **26/10/2021**; Published: **26/10/2021**

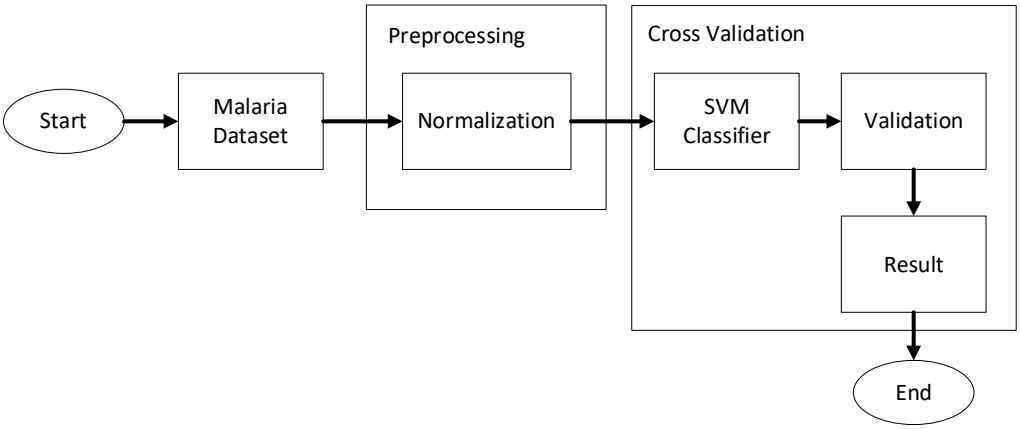
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 4, Oktober 2021, Page 1580-1584
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i4.3347

mengusulkan kerangka kerja untuk memprediksi endemik malaria di Nigeria dengan menerapkan algoritma Long Short-Term Memory (LSTM). Penelitian [11] mengusulkan kerangka kerja baru yang terukur untuk memprediksi kejadian malaria di lokasi geografis tertentu, dataset yang digunakan berasal dari India dan menerapkan teknik klasifikasi LSTM.

Berdasarkan paparan dari beberapa penelitian yang telah dilakukan maka pada penelitian ini bertujuan untuk deteksi dan klasifikasi penyakit malaria berat berdasarkan histori pemeriksaan data pasien dengan menggunakan metode Support Vector Machine (SVM), penerapan teknik normalisasi min-max dan teknik cross validation.

2. METODOLOGI PENELITIAN

Pada gambar 1 merupakan sistem diagram usulan pada penelitian ini.



Gambar 1. Diagram Proses Usulan

2.1 Dataset

Penelitian ini menggunakan dataset penyakit malaria yang berasal dari Nigeria [12]. Total baris pada dataset ini sebanyak 337 pasien. Tabel 1 merupakan karakteristik dataset yang digunakan. Untuk class pada dataset yaitu fitur severe_maleria.

Tabel 1. Karakteristik Dataset

No	Nama Fitur	Tipe
1	age	Numerik
2	sex	Binary (0,1)
3	fever	Binary (0,1)
4	cold	Binary (0,1)
5	rigor	Binary (0,1)
6	fatigue	Binary (0,1)
7	headace	Binary (0,1)
8	bitter_tongue	Binary (0,1)
9	vomitting	Binary (0,1)
10	diarrhea	Binary (0,1)
11	Convulsion	Binary (0,1)
12	Anemia	Binary (0,1)
13	jundice	Binary (0,1)
14	cocacola_urine	Binary (0,1)
15	hypoglycemia	Binary (0,1)
16	prostraction	Binary (0,1)
17	hyperpyrexia	Binary (0,1)
18	severe_maleria	Binary (0,1)

2.2 Normalisasi

Normalisasi adalah Normalisasi adalah teknik penskalaan atau pemetaan teknik atau tahap pra-pemrosesan [13]. Teknik normalisasi ada berbagai macam seperti, min-max, Z-score, dan feature scaling. Pada penelitian ini menerapkan teknik normalisasi min-max. Min-Max Normalization adalah metode normalisasi yang sering digunakan untuk mengatasi permasalahan nilai antar fitur yang memiliki jarak terlampau jauh [14]. Normalisasi dilakukan pada fitur age hal tersebut dikarenakan pada fitur lainnya nilai berbentuk 0 dan 1. Sehingga normalisasi

Submitted: **07/10/2021**; Accepted: **26/10/2021**; Published: **26/10/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 4, Oktober 2021, Page 1580-1584
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i4.3347

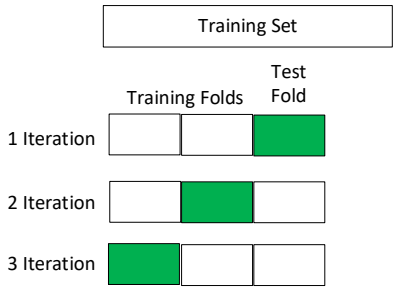
dilakukan pada fitur age guna mendapatkan nilai yang berada di antara 0 dan 1. Formula yang digunakan yaitu (1) [15].

$$N^* = \frac{N - \min(n)}{\max(n) - \min(n)} \quad (1)$$

Dimana nilai N^* adalah hasil normalisasi data. *N* yaitu data yang belum dinormalisasi. *min(n)* merupakan nilai minimum dari semua data dan *max(n)* adalah nilai maksimum dari semua data.

2.3 Cross Validation

Cross validation merupakan sebuah prosedur acak yang membagi kumpulan data menjadi K yang terputus-putus dengan ukuran yang kira-kira sama, dan setiap lipatan digunakan secara bergantian untuk menguji model yang diinduksi dari lipatan K-1 lainnya oleh x algoritma klasifikasi [16]. Figure 2 merupakan contoh gambar proses terjadinya cross validation.



Gambar 2. Proses Cross Validation

2.4 Support Vector Machine

Support Vector Machine (SVM) merupakan bagian penting dari teori pembelajaran mesin. SVM sangat efisien untuk banyak aplikasi dalam sains dan teknik, terutama untuk masalah klasifikasi (pengenalan pola) [17]. Ide klasifikasi SVM dapat digambarkan sebagai berikut: misalkan ada *m* sampel pengamatan (set pelatihan), (*xi, yi*), i = 1, 2, ...,m di mana:

$$x_i^T = (xi1, \dots xid) \in R^d \quad (2)$$

Dimana x_i^T adalah fitur d-dimensi dari sampel *i* dan *y* ∈ *{-1, +1}* adalah kelas label berkodenya. Jika sampel *xi* ditugaskan ke kelas positif, maka *yi* adalah +1, dan jika ditugaskan ke kelas negatif, maka *yi* adalah -1. Set pelatihan ini dapat dipisahkan oleh hyperplane $w^T xi + b = 0$, di mana *w* adalah vektor bobot dan *b* adalah bias. Persamaan hyperplanes marginal H1 dan H2 dapat dilihat pada (3) dan (4).

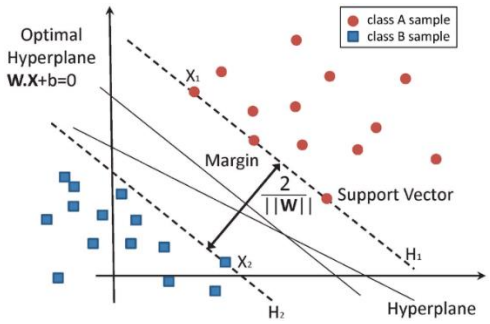
$$H_1: (w^T xi + b) = 1 \quad (3)$$

$$H_2: (w^T xi + b) = -1 \quad (4)$$

Jadi, titik-titik yang diklasifikasikan dengan benar memenuhi pertidaksamaan (5).

$$y_i: (w^T xi + b) \geq 1 \quad (5)$$

Untuk *xi*, i = 1,2,…,m. Jarak antara *hyperplanes* marginal yaitu sama dengan $\frac{2}{||w||}$. Sample pelatihan apa saja yang jatuh pada *hyperplanes* H1 atau H2, sisi yang mendefinisikan margin adalah vektor pendukung, seperti yang ditunjukkan pada gambar 3.



Gambar 3. Klasifikasi Support Vector Machine

Submitted: **07/10/2021**; Accepted: **26/10/2021**; Published: **26/10/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 4, Oktober 2021, Page 1580-1584
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i4.3347

Fungsi kernel yang umum digunakan pada *Support Vector Machine* adalah kernel Linear, Radial Basis Function (RBF), dan Polynomial [18]. Fungsi kernel dan parameter yang digunakan dalam analisis SVM sangat mempengaruhi akurasi yang akan dihasilkan [19].

3. HASIL DAN PEMBAHASAN

Pada tahap ini akan dilakukan eksperimen untuk mendapatkan hasil akurasi terbaik. Eksperimen yang dilakukan berupa penerapan beberapa nilai K pada cross validation dan penerapan normalization. Untuk perhitungan akurasi dapat menggunakan formula (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Pada tabel 2 merupakan bentuk confusion matrix secara umum.

Tabel 2. Confussion Matrix

	Aktual Benar	Aktual Salah
Prediksi Benar	TP	FP
Prediksi Salah	FN	TN

True Positif (TP) merupakan jumlah data malaria yang diprediksi benar dan actual nya benar. False Positif (FP) merupakan jumlah data malaria yang diprediksi benar akan tetapi actual sebenarnya salah. True Negatif (TN) merupakan jumlah data malaria yang diprediksi salah dan sebenarnya bukan malaria. False Negatif (FN) merupakan jumlah data malaria yang diprediksi salah akan tetapi sebenarnya malaria.

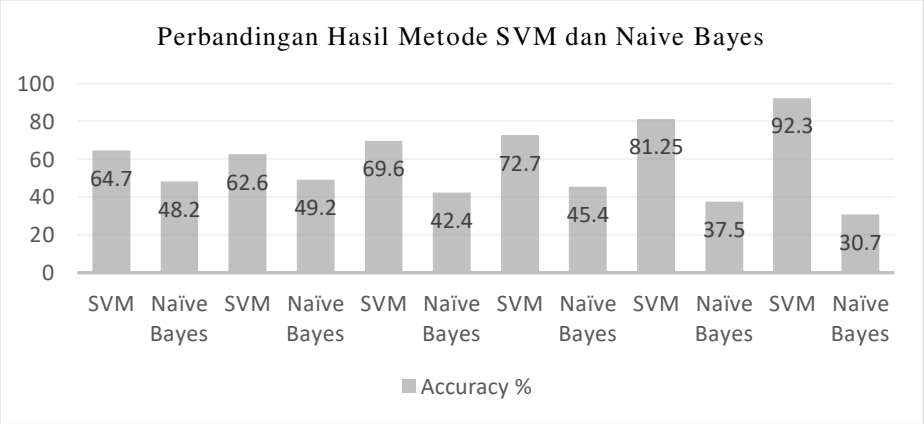
Tabel 3 merupakan hasil eksperimen yang dihasilkan pada penelitian ini dengan menggunakan model SVM.

Tabel 3. Hasil Klasifikasi SVM

K-Fold	Normalisasi	Akurasi %
-	-	64.7
5	Min-Max	62.6
10	Min-Max	69.6
15	Min-Max	72.7
20	Min-Max	81.25
25	Min-Max	92.3

Berdasarkan tabel di atas dapat dilihat bahwa penerapan normalisasi min-max maupun cross validation memiliki pengaruh pada kasus penelitian ini. Hasil yang didapatkan jika tanpa dilakukan normalisasi min-max pada preprocessing dan tanpa dilakukan teknik cross validation hanya sebesar 69.4%. Sebaliknya, jika diterapkan teknik cross validation dan normalisasi min-max hasil yang didapatkan mengalami peningkatan. Nilai K pada cross validation juga memiliki pengaruh, jika nilai K semakin besar maka nilai akurasi yang dihasilkan semakin besar. Pemilihan nilai K yang dimulai dengan nilai 5 pada penelitian ini menunjukkan bahwa nilai K pada cross validation akan bekerja efektif saat lebih dari sama dengan 5.

Penelitian ini juga melakukan perbandingan metode SVM dengan metode lain. Pada Gambar 4 merupakan perbandingan menggunakan metode Naïve Bayes.



Gambar 4. SVM vs Naive Bayes

Terlihat pada gambar 4 di atas bahwa metode usulan SVM lebih unggul dibandingkan dengan metode Naïve Bayes (NB). Metode SVM unggul pada seluruh eksperimen yang dilakukan dengan gap akurasi yang jauh dengan

Submitted: **07/10/2021**; Accepted: **26/10/2021**; Published: **26/10/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 4, Oktober 2021, Page 1580-1584
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i4.3347

rata-rata 25%. Hal tersebut terjadi karena pada metode SVM memiliki fungsi kernel yang dapat bekerja dengan baik pada kategori data biner.

4. KESIMPULAN

Pada penelitian ini berhasil melakukan deteksi dan klasifikasi penyakit malaria berat berdasarkan data histori pasien. Model SVM mampu menghasilkan akurai tertinggi 92.3% dengan menerapkan teknik cross validation dan normalisasi min-max. Penelitian ini membuktikan juga bahwa model SVM juga lebih unggul daripada model Naïve Bayes. Nilai K pada teknik cross validation memiliki pengaruh terhadap hasil akurasi. Untuk penelitian selanjutnya dapat dilakukan penerapan teknik normalisasi dan metode machine learning lainnya.

REFERENCES

[1] WHO, 2021. “Key Fact Malaria”. Tersedia [https://www.who.int/news-room/fact-sheets/detail/malaria] diakses 8 Agustus 2021.
[2] Irmanita, Rachmadania, Sri Suryani Prasetiyowati, and Yuliant Sibaroni. "Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 5.1 (2021): 10-16.
[3] Parveen, Rahila, et al. "Prediction of malaria using artificial neural network." *Int J Comput Sci Netw Secur* 17.12 (2017): 79-86.
[4] Lee, You Won, Jae Woo Choi, and Eun-Hee Shin. "Machine learning model for predicting malaria using clinical information." *Computers in Biology and Medicine* 129 (2021): 104151.
[5] P.J. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J. de Hoon, Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics, 25 (2009) 1422-1423
[6] C.f.D.C.a. Prevention, DPDx - Laboratory Identification of Parasites of Public Health Concern 2020.
[7] Nkiruka, Odu, Rajesh Prasad, and Onime Clement. "Prediction of malaria incidence using climate variability and machine learning." *Informatics in Medicine Unlocked* 22 (2021): 100508.
[8] Cleary, Eimear, et al. "Spatial prediction of malaria prevalence in Papua New Guinea: a comparison of Bayesian decision network and multivariate regression modelling approaches for improved accuracy in prevalence prediction." *Malaria Journal* 20.1 (2021): 1-16.
[9] Mohapatra, Pallavi, et al. "Determining suitable machine learning classifier technique for prediction of malaria incidents attributed to climate of Odisha." *International Journal of Environmental Health Research* (2021): 1-17.
[10] Awotunde, Joseph Bamidele, et al. "Prediction of malaria fever using long-short-term memory and big data." *International Conference on Information and Communication Technology and Applications*. Springer, Cham, 2020.
[11] Santosh, Thakur, Dharavath Ramesh, and Damodar Reddy. "LSTM based prediction of malaria abundances using big data." *Computers in Biology and Medicine* 124 (2020): 103859.
[12] Adeboye, Nureni Olawale, Olawale Victor Abimbola, and Sakinat Oluwabukola Folorunso. "Malaria patients in Nigeria: Data exploration approach." *Data in brief* 28 (2020): 104997.
[13] Shalabi, L.A., Z. Shaaban and B. Kasasbeh, Data Mining: A Preprocessing Engine, J. Comput. Sci., 2: 735-739, 2006
[14] Larose, Daniel T., and Chantal D. Larose. *Discovering knowledge in data: an introduction to data mining*. Vol. 4. John Wiley & Sons, 2014.
[15] Wijayanti, Ratna Ayu, Muh Tanzil Furqon, and Sigit Adinugroho. "Penerapan Algoritme Support Vector Machine Terhadap Klasifikasi Tingkat Risiko Pasien Gagal Ginjal." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* 2548 (2018): 964X.
[16] Wong, Tzu-Tsung. "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation." *Pattern Recognition* 48.9 (2015): 2839-2846.
[17] Wu, Qiang, and Ding-Xuan Zhou. "Analysis of support vector machine classification." *Journal of Computational Analysis & Applications* 8.2 (2006).
[18] Sari, Esa Anindika, et al. "Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia dengan Support Vector Machine." *JIKO (Jurnal Informatika dan Komputer)* 3.3 (2020): 188-195.
[19] Feta, Neneng Rachmalia, and Asep Rahmat Ginanjar. "Komparasi Fungsi Kernel Metode Support Vector Machine Untuk Pemodelan Klasifikasi Terhadap Penyakit Tanaman Kedelai." *BRITech, Jurnal Ilmiah Ilmu Komputer, Sains dan Teknologi Terapan* 1.1 (2019): 33-39.