



## Komparasi Metode SVM dan *Adaboost* untuk Klasifikasi Kanker Payudara

Ikka Ayu Elfitrianna<sup>1</sup>, Rastri Prathivi<sup>2\*</sup>

<sup>1</sup> Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang  
Jl. Soekarno Hatta. RT.7/ RW.7. Tlogosari Kulon, Ke. Pedurungan, Kota Semarang, Jawa Tengah  
50196, e-mail: ikkaayu77@gmail.com

<sup>2</sup> Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang  
Jl. Soekarno Hatta. RT.7/ RW.7. Tlogosari Kulon, Ke. Pedurungan, Kota Semarang, Jawa Tengah  
50196, e-mail: vivi@usm.ac.id

### ARTICLE INFO

#### *History of the article :*

Received 23 December 2024  
Received in revised form 23 December 2024  
Accepted 27 January 2025  
Available online 29 January 2025

#### Keywords:

*Adaboost*; SVM; Klasifikasi; Kanker Payudara

#### \* Correspondence:

Telepon:  
-  
E-mail:  
vivi@usm.ac.id

### ABSTRACT

One of the most prevalent malignancies in women and a major global cause of death is breast cancer. To determine whether a cancer is benign or malignant, early detection is essential. The usefulness of the Support Vector Machine (SVM) and Adaptive Boosting (*Adaboost*) algorithms for breast cancer classification using mammography data is compared in this study. 569 records make up the dataset, which was sourced from the Kaggle Repository and is split into 75% training data and 25% testing data. Preprocessing steps include feature and target variable creation, categorical-to-numerical conversion, data splitting, and normalization. SVM achieved an accuracy of 97%, with a precision of 98%, recall of 94%, and F1 score of 96%. *Adaboost*, on the other hand, achieved an accuracy of 96%, precision of 98%, recall of 92%, and F1 score of 95%. The results reveal that both algorithms are highly effective for breast cancer detection, with SVM marginally exceeding *Adaboost* in total performance. These findings emphasize the promise of machine learning techniques in facilitating early cancer diagnosis, hence boosting survival rates. It is advised that future research employ a wider range of datasets and investigate different classification techniques in order to improve accuracy and dependability even more.

## 1. INTRODUCTION

Kanker adalah penyakit yang dapat berkembang hampir diseluruh organ tubuh manusia. Ketika gen yang mengatur reproduksi sel tubuh rusak dan berkembang dengan tidak normal, sel akan membelah lebih banyak, yang akan menyebabkan abnormalitas siklus sel yang khas. Akibatnya, ia memiliki kemampuan untuk menyerang jaringan biologis terdekat dengannya kemudian menyebar ke jaringan tubuh lainnya. [1] Kanker menjadi salah satu penyebab kematian didunia. Meningkatnya gaya hidup masyarakat serta kebiasaan buruk yang dilakukan terus-menerus juga menjadi pemicu timbulnya penyakit ini, seperti merokok, kurangnya aktivitas fisik, dan melakukan diet “*westernized*”. [2] Faktor resiko utama dari kanker payudara adalah usia dan riwayat penyakit dari keluarga. Berdasarkan pemaparan dari Organisasi Kesehatan Dunia (WHO), faktor resiko obesitas dapat mencegah lebih dari 30% kematian yang diakibatkan oleh kanker payudara. [3] kanker payudara termasuk jenis kanker yang paling banyak diderita oleh kaum wanita. [4]

Organisasi Kesehatan Dunia (WHO) melaporkan bahwa terdapat 2,3 juta kasus kanker payudara yang didiagnosis di seluruh dunia pada tahun 2020, dengan angka kematian sebesar 685.000. Selain itu pada akhir tahun 2020 WHO juga mencatat bahwa terdapat 7,8 juta wanita yang masih hidup setelah didiagnosis menderita kanker payudara dalam lima tahun terakhir. [5] dalam kanker payudara dibagi menjadi kanker Ganas dan Jinak, pencegahan dan pengobatan dapat dilakukan dengan mengetahui jenis kanker secepatnya. (novita) [6] *Support Vector Machine (SVM)* merupakan salah satu teknik yang efektif untuk digunakan pada masalah klasifikasi. Pada awalnya metode SVM hanya digunakan untuk klasifikasi data linier, tetapi sekarang dapat digunakan juga untuk klasifikasi data non-linier dengan menggunakan trik kernel. Metode ini bekerja dengan mencari *hyperplane* dengan margin yang maksimal untuk membedakan antar kelas klasifikasi, metode SVM ini memiliki waktu komputasi yang cukup cepat. [2] algoritma *Adaboost* adalah teknik dalam data mining yang dapat digunakan untuk meningkatkan akurasi pada suatu metode dalam klasifikasi. [7] *Adaptive Boosting* yang lebih dikenal dengan sebutan *Adaboost* merupakan metode yang termasuk dalam *essamble* pada algoritma *boosting*, diterapkan untuk mendapatkan hasil yang maksimal sehingga memudahkan dalam proses analisis fitur dan klasifikasi [8], algoritma ini ditemukan oleh Yoav Freund dan Robert Schapire. [9]

Permasalahan yang muncul dalam penelitian ini adalah kurangnya akurasi untuk mendeteksi kanker payudara dengan diagnosa mammogram secara manual. Pendeteksian secara dini untuk kanker payudara diperlukan untuk dapat segera mengenali jenis kanker tersebut tergolong kanker yang ganas atau yang jinak. Metode SVM dan *Adaboost* merupakan metode yang akurasinya baik untuk mengatasi permasalahan deteksi kanker payudara menggunakan *diagnose mammogram*.

Pada penelitian yang dilakukan oleh Risha dkk (2020) [10], penggunaan metode SVM untuk [11] klasifikasi apakah terdapat *Pneumonia* pada hasil *x-ray* yang telah dilakukan terhadap pasien. Tahapan yang dilakukan sebelum melakukan klasifikasi yaitu *cropping*, *resizing*, *contrast stretching* dan *thresholding*, yang kemudian dilakukan ekstrasi dengan menggunakan GLCM yang kemudian diklasifikasi menggunakan metode SVM. dari penelitian yang dilakukan diperoleh hasil akurasi sebesar 62,66%. Pada penelitian yang dilakukan oleh Marthin Yudi (2020) [12], untuk melakukan perbandingan hasil klasifikasi curah hujan dengan menggunakan metode SVM, dengan data yang diperoleh dari BMKG Stasiun Meteorologi Maritim Tanjung Priok, Jakarta pada periode tahun 2017 sampai tahun 2018. Berdasarkan hasil dari analisis klasifikasi, diketahui bahwa metode SVM memiliki tingkat akurasi sebesar 79,45%. Pada penelitian yang dilakukan oleh Yusuf dkk (2022) [13] untuk klasifikasi jenis buah pisang berdasarkan citra warna dengan metode SVM. data yang digunakan adalah citra pisang dengan total pisang sebanyak 1256, yang diklasifikasi menjadi 2 jenis pisang yaitu, Pisang Ambon dan Pisang Lady. Hasil dari penelitian diperoleh akurasi sebesar 89,86%, yang dapat diartikan bahwa Algoritma SVM cukup baik untuk digunakan dalam mengklasifikasi jenis pisang.

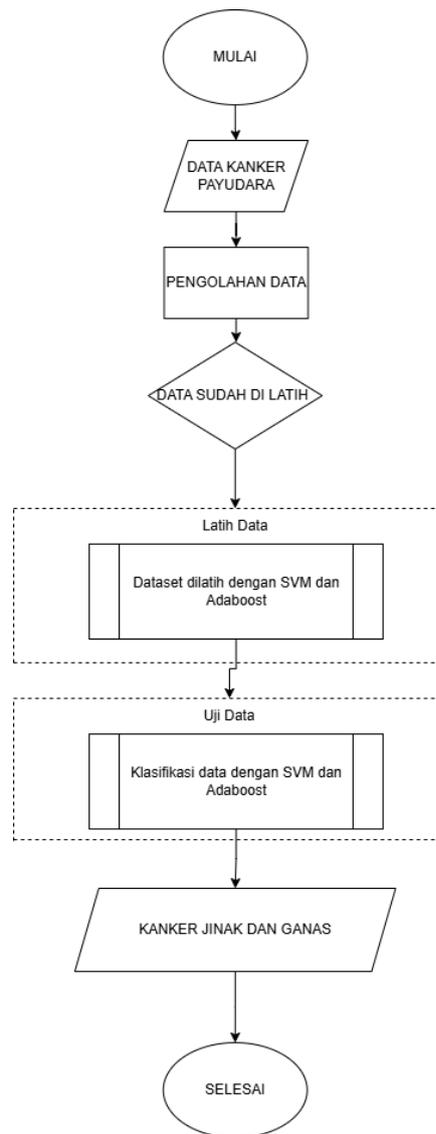
Penelitian selanjutnya yang dilakukan oleh Muchammad Arief (2019) [14], dengan judul Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Warna Menggunakan Metode SVM. Dalam penelitian ini, kamera digital atau kamera *handphone* digunakan untuk mengambil beberapa gambar buah jeruk. Kemudian sampel tersebut diubah menjadi warna *greyscale* dan kemudian dikembalikan ke warna LAB untuk mengetahui nilai warna A dari LAB dan nilai R, G dari RGB. Hasil penelitian menunjukkan akurasi sebesar 80% dari data 100 gambar buah jeruk.

Penelitian yang dilakukan oleh Laila dkk (2021) [15], tentang penggunaan metode *Adaboost* untuk melakukan klasifikasi penerima program bantuan sosial. Pada penelitian ini data yang digunakan merupakan data penduduk yang diperoleh dari kantor kelurahan setempat. Berdasarkan hasil penelitian dengan menggunakan metode *Adaboost* didapatkan hasil akurasi sebesar 95%. Selanjutnya yaitu penelitian yang dilakukan oleh Ginanjar Abdurrohman (2022) [9], tentang penggunaan metode *Adaboost* untuk klasifikasi penyakit diabetes melitus, dengan dataset yang diperoleh dari UCI Machine Learning, dengan 8 variabel kelas predictor, 1 variabel kelas target. Dari penelitian yang telah dilakukan diperoleh hasil akurasi sebesar 76,19%. Penelitian selanjutnya dilakukan oleh Agus & M Basit (2020) [16], dilakukan optimasi metode *Adaboost* untuk penyakit stroke, dengan data yang didapatkan melalui penyebaran kuisioner menggunakan gogleform yang disebar kepada pasien di beberapa puskesmas di daerah Banjarmasin. Dari penelitian tersebut didapatkan hasil akurasi sebesar 0,981% dengan melakukan data split 70/30. Penelitian untuk klaifikasi penyakit diabetes menggunakan metode CFS dan ROS dengan algoritma J48 berbasis *Adaboost* yang dilakukan oleh Dian Pramadhana(2021) [17], menggunakan dataset diabetes yang diambil dari repository public obline dari archive.ics berupa data prima dengan jumlah data 768. Dari penelitian ini dihasilkan akurasi sebesar 92,3%.

## RESEARCH METHODS

### Tahapan Penelitian

Tahapan penelitian merupakan beberapa langkah yang harus diikuti dalam suatu penelitian guna mewujudkan tujuan penelitian yang sudah ditentukan. Adapun langkah-langkah pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Penelitian

## 1. Pengumpulan Data

Dalam penelitian ini metode pengumpulan data yang digunakan untuk memperoleh sumber data adalah pengumpulan data sekunder. Data yang digunakan merupakan dataset yang diperoleh dari situs *Kaggle Repository*, sebanyak 569 data.

## 2. Pengolahan Data

Adapun teknik yang digunakan pada tahapan ini sebagai berikut;

- Preprocessing* data, adalah suatu tahap untuk mempersiapkan data yang telah diperoleh dari tahap pengumpulan data sebelumnya,
- Membuat Variabel Fitur (X) dan Variabel Target (y). Variabel (X) berisi semua kolom kecuali kolom 'id', sedangkan variabel y merupakan target yang akan diprediksi oleh model.

- c. Konversi nilai target, proses ini merupakan proses dimana nilai kategorial diubah menjadi nilai numerik agar dapat digunakan oleh model.
- d. Split Data, pada proses ini data yang dimiliki akan dibagi menjadi data pelatihan (*training*) dan data pengujian (*testing*).
- e. Normalisasi Data, tujuan dari proses ini ialah untuk menyesuaikan fitur-fitur data agar memiliki skala yang sama.

### 3. Data Training

Tahap data training merupakan tahapan untuk melatih model atau sistem, yang berfungsi untuk membantu model mempelajari pola, hubungan, dan karakteristik tertentu yang akan digunakan untuk pembuatan prediksi.

Adaboost=

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right)$$

SVM=

Terdapat beberapa tahapan dalam klasifikasi SVM, tahapannya adalah sebagai berikut:

- a. Mencari nilai *weight* ( $w$ ) dan bias ( $b$ )

$$Y_i (w \cdot x_i + b) \geq 1$$

- b. Menentukan nilai *Hyperplane*

$$f(x) = w_1 x_1 + w_2 x_2 + w_n x_n + b = 0$$

- c. Menguji dan mengklasifikasi data

$$f(x) = \text{sign}(w_1 x_1 + w_2 x_2 + w_n x_n + b)$$

### 4. Uji Data

Pada tahapan ini dilakukan pengujian dataset untuk klasifikasi dengan menggunakan metode SVM dan *Adaboost*.

## RESULTS

### 1. Data kanker payudara

Dari dataset didapatkan data sebanyak 569 data, berdasarkan data tersebut kanker payudara diklasifikasikan menjadi 2 tipe yaitu Ganas (G) dan Jinak (J).

### 2. Pengolahan Data

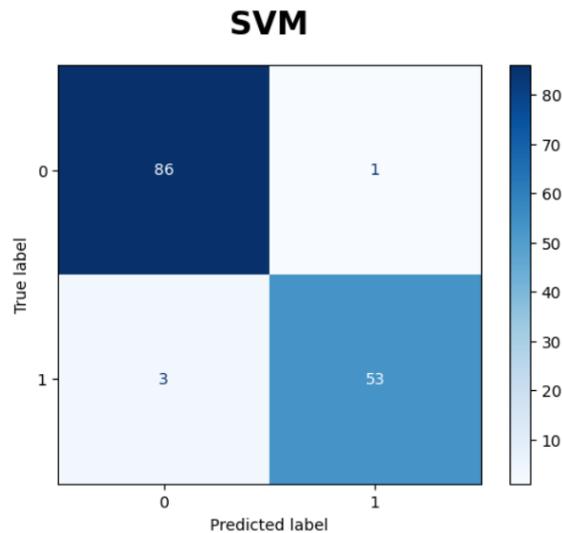
Pada *preprocessing* data, kolom diagnosis diganti menjadi 'ganas\_cancer'. Selanjutnya membuat variable fitur ( $X$ ), variable ini berisi semua kolom kecuali kolom 'id' dan 'ganas\_cancer', lalu membuat variable target ( $y$ ) dengan hasil kolom ganas\_cancer sebagai variable targetnya, dari variable target tersebut lalu diubah menjadi nilai numerik kanker ganas(G) dikonversi menjadi 1, sedangkan kanker Jinak (J) dikonversi menjadi 0.

Hasil dari dataset tersebut lalu dibagi menjadi data training dan data testing, data yang digunakan sebagai data *testing* sebesar 25% dari data atau sebanyak 142 data. Sedangkan untuk data *training* yang digunakan yaitu sisanya yaitu sebanyak 75% dari data atau sebanyak 427 data.

### 3. Klasifikasi Data dengan SVM dan Adaboost

Klasifikasi dilakukan dengan menggunakan metode SVM dan *Adaboost*, dengan data yang digunakan sebanyak 569 data. Pada penelitian ini proses klasifikasi menggunakan *library* SVM pada *Sklearn Python*, oleh karena itu klasifikasi ini merupakan *multiclass classification*.

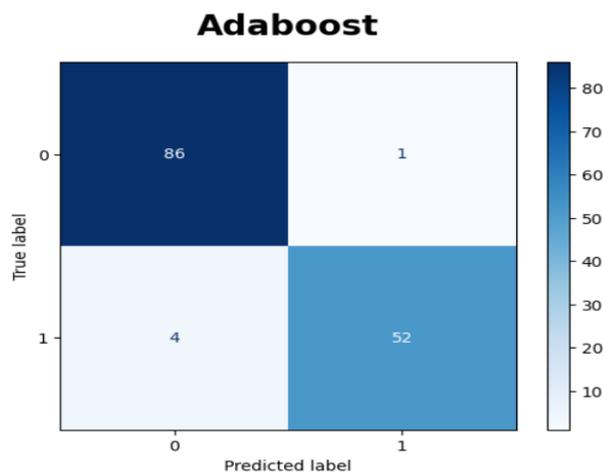
Untuk hasil dari model SVM ditunjukkan dengan *confusion matrix* pada gambar 2



Gambar 2. Confusion matrix SVM

*Confusion matrix* merupakan tabel yang digunakan untuk melakukan evaluasi kinerja klasifikasi model, dengan membandingkan *True Label* dan *Predicted Label*. Dari hasil klasifikasi menggunakan metode SVM dihasilkan rincian *True Negative* (TN)= 86, *False Positive* (FP)= 1, *False Negative* (FN)= 3, *True Positive* (TP)= 53. Akurasi model SVM terlihat tinggi karena sebagian besar prediksi benar, dengan kesalahan hanya 1 sampel salah yang diprediksi menjadi positif (FP), dan 3 sample salah diprediksi menjadi negatif (FN).

Sedangkan untuk hasil dari model *Adaboost* ditunjukkan dengan *confusion matrix* pada gambar 3.



Gambar 3. Confusion Matrix *Adaboost*

*Confusion matrix* dari hasil klasifikasi model *Adaboost* menunjukkan rincian *True Negative* (TN)= 86, *False Positive* (FP)= 1, *False Negative* (FN) = 4, *True Positive* (TP) = 52. Akurasi *True Positive* dan *True Negative* cukup tinggi yang artinya model bekerja dengan

cukup baik. Kesalahan pada model *Adaboost* 1 sample salah dan dipredikasi menjadi positif (FP), dan 4 sample salah diprediksi menjadi negatif (FN).

## DISCUSSION

### Evaluasi Metode

Pada tahapan ini setelah hasil dari tahap permodelan didapatkan, maka langkah selanjutnya adalah melakukan evaluasi model. Setelah tahapan pengujian model selesai dilakukan analisis tambahan untuk menilai hasil dari evaluasi model yang digunakan. Analisis hasil digunakan untuk mengubah *matrix Accuracy, Precision, Recall, dan F1 Score* pada data uji yang nantinya akan menghasilkan hasil yang akurat, berikut rumus untuk *Accuracy, Precision, Recall, dan F1 Score*;

1.  $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
2.  $Precision = \frac{TP}{TP+FP}$
3.  $Recal = \frac{TP}{TP+FN}$
4.  $F1\ Score = 2 \times \frac{Recall \times Precission}{Reacll+Precission}$

Selain itu pada tahap evaluasi ini juga menggunakan *Confussion Matrix*, yang merupakan tabel yang digunakan untuk menyatakan jumlah data uji yang benar dan data uji yang salah untuk mengetahui tingkat performa suatu model berdasarkan presisi dan akurasi, adapun rumus *confussion matrix* ditunjukkan pada tabel 1

Tabel 1 Confusion Matrix

Nilai Prediksi	Positif	Negatif
Positif	True Positif	False Positif
Negatif	False Negatif	True Negatif

Hasil dari evaluasi model SVM dengan *Accuracy score* sebesar 0,97 dengan hasil tersebut dapat dikatakan bahwa model bekerja dengan baik dalam klasifikasi data, dengan tingkat kesalahan yang rendah. Setelah evaluasi model dapat dilakukan perhitungan dengan *matrix* sebagai berikut:

1.  $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 0,97$
2.  $Precision = \frac{TP}{TP+FP} = 0,98$
3.  $Recal = \frac{TP}{TP+FN} = 0,94$
4.  $F1\ Score = 2 \times \frac{Recall \times Precission}{Reacll+Precission} = 0,96$

Sedangkan hasil dari penggunaan *Adaboost* mendapatkan *Accuracy score* sebesar 0,96, dari hasil yang didapatkan tersebut dapat dikatakan bahwa model bekerja dengan baik karena nilai akurasi yang dihasilkan cukup besar. Evaluasi model dapat dilakukan dengan menggunakan *matrix* sebagai berikut:

1.  $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 0,96$
2.  $Precision = \frac{TP}{TP+FP} = 0,98$
3.  $Recal = \frac{TP}{TP+FN} = 0,92$
4.  $F1\ Score = 2 \times \frac{Recall \times Precision}{Recall+Precision} = 0,95$

Hasil perhitungan dengan menggunakan metode SVM dan *Adaboost* ditunjukkan pada tabel 2.

Tabel 2 Hasil Evaluasi

Model	Accuracy	Precision	Recall	F1 Score
SVM	0,97	0,98	0,94	0,96
<i>Adaboost</i>	0,96	0,98	0,92	0,95

## CONCLUSIONS AND RECOMMENDATIONS

Hasil penelitian menunjukkan bahwa kedua teknik *Support Vector Machine* (SVM) dan *Adaptive Boosting* (*Adaboost*) efektif dalam menggunakan data mammogram untuk mengklasifikasikan kanker payudara menjadi ganas atau jinak. SVM memiliki performa terbaik dengan akurasi 97%, sedangkan *Adaboost* sedikit lebih rendah dengan akurasi 96%. Kedua algoritma juga memiliki *precision* yang tinggi (98%) dan *recall* yang baik, dengan SVM memiliki sedikit keunggulan dalam *recall* (94%) dibandingkan *Adaboost* (92%). Hasil ini menunjukkan bahwa kedua algoritma tersebut mampu mendeteksi kanker payudara dengan baik. Akibatnya, baik SVM maupun *Adaboost* dapat digunakan sebagai alat bantu yang berguna untuk mendeteksi kanker payudara lebih awal. Ini sangat penting untuk meningkatkan kesembuhan dan angka kematian akibat kanker.

Dari kesimpulan penelitian ini, rekomendasi yang diberikan untuk penelitian berikutnya adalah dengan menggunakan dataset yang lebih kaya dan beragam, juga bisa mengeksplor metode klasifikasi lain untuk membandingkan hasil dan menemukan metode yang lebih optimal.

## REFERENCES

- [1] J. KUSUMA, B. H. HAYADI, W. WANAYUMINI, and R. ROSNELLY, "Komparasi Metode Multi Layer Perceptron (MLP) dan Support Vector Machine (SVM) untuk Klasifikasi Kanker Payudara," *MIND J.*, vol. 7, no. 1, pp. 51–60, 2022, doi: 10.26760/mindjournal.v7i1.51-60.
- [2] R. Resmiati and T. Arifin, "Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination," *Sistemasi*, vol. 10, no. 2, p. 381, 2021, doi: 10.32520/stmsi.v10i2.1238.
- [3] D. Alfiani, M. P. Putri, and W. Widayanti, "Literature Study: Obesitas sebagai Faktor

- Risiko pada Kanker Payudara Triple Negative,” *Bandung Conf. Ser. Med. Sci.*, vol. 2, no. 1, 2022, doi: 10.29313/bcsms.v2i1.760.
- [4] K. Khadijah and R. Kusumaningrum, “Ensemble Classifier untuk Klasifikasi Kanker Payudara,” *It J. Res. Dev.*, vol. 4, no. 1, pp. 61–71, 2019, doi: 10.25299/itjrd.2019.vol4(1).3540.
- [5] H. Sundari, M. A. Amrustian, A. Dwi, and P. Wicaksono, “Penerapan Recursive Feature Elimination pada Support Vector Machine untuk Klasifikasi Kanker Payudara,” vol. 8798, pp. 60–65, 2024.
- [6] N. R. Muntiar and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, 2022, doi: 10.35960/ikomti.v3i1.766.
- [7] R. T. Febianto, D. Suranti, and R. T. Alinse, “Penerapan Algoritma Adaboost Dalam Mengetahui Pola Pengguna Kb Di Puskesmas Tanjung Harapan,” *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 145–155, 2024.
- [8] M. Siddik Hasibuan and H. Harahap, “Penerapan Metode Haar-Like Feature Dan Algoritma Adaboost Dalam Penentuan Klasifikasi Hama Tanaman Kopi,” *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 87–93, 2024.
- [9] G. Abdurrahman, “Jurnal Sistem dan Teknologi Informasi Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier,” *JUSTINDO (Jurnal Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 59–66, 2022.
- [10] R. A. Wati, H. Irsyad, and M. E. A. R. Rivian, “Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine,” *J. Algoritm.*, vol. 1, no. 1, pp. 21–32, 2020.
- [11] Y. Fatman, N. Khoirun Nafisah, and P. Bendoro Jembar Pambudi, “Implementasi Payment Gateway dengan Menggunakan Midtrans pada Website UMKM Geberco,” *J. KomtekInfo*, vol. 10, pp. 64–72, 2023, doi: 10.35134/komtekinfo.v10i2.364.
- [12] Marthin Luter Laia and Yudi Setyawan, “Perbandingan Hasil Klasifikasi Curah Hujan Menggunakan Metode SVM dan NBC,” *J. Stat. Ind. dan Komputasi*, vol. 5, no. 2, pp. 51–61, 2020.
- [13] Y. Amrozi, D. Yuliati, A. Susilo, N. Novianto, and R. Ramadhan, “Klasifikasi Jenis Buah Pisang Berdasarkan Citra Warna dengan Metode SVM,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 394–399, 2022, doi: 10.32736/sisfokom.v11i3.1502.
- [14] M. Arief, “Klasifikasi Kematangan Buah Jeruk Berdasarkan Fitur Warna Menggunakan Metode SVM,” *J. Ilmu Komput. dan Desain Komun. Vis.*, vol. 4, no. 1, pp. 9–16, 2019.
- [15] L. Qadrini, A. Sepperwali, and A. Aina, “Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial,” *J. Inov. Penelit.*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [16] A. Byna and M. Basit, “Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [17] D. Pramadhana, “Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost,” *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 89–98, 2021, doi: 10.29408/edumatic.v5i1.3336.