



Perbandingan Algoritma Machine Learning Untuk klasifikasi Amenorrhea

Indah Lestari¹, Muhamad Akbar^{2*}, Bunga Intan³

^{1,2,3} Fakultas Ilmu Teknik, Universitas Bina Insan, Lubuklinggau, Indonesia
Email: ¹1902020032@mhs.univbinainsan.ac.id, ²muhamad.akbar@univbinainsan.ac.id,
³bungaintan@univbinainsan.ac.i

Abstrak

Penelitian ini bertujuan untuk mengatasi masalah kurangnya penelitian yang dilakukan pada algoritma perbandingan dalam mengklasifikasikan Amenore. Banyak orang yang mengabaikan atau bahkan tidak memahami Amenore karena informasi yang terbatas mengenai masalah ini. Dalam penelitian ini, dilakukan perbandingan antara empat algoritma, yaitu k-NN, SVM, Decision Tree, dan Naïve Bayes, untuk mengklasifikasikan Amenore dengan menggunakan data dari salah satu repositori data Google. Hasil penelitian menunjukkan bahwa algoritma k-NN memberikan nilai akurasi tertinggi, yaitu 0,99. Temuan ini dapat membantu dalam pengembangan metode diagnosa Amenore yang lebih akurat dan efektif.

Kata Kunci: Amenorrhea, k-NN,Svm, Decision Tree, Naïve Bayes, dataset, repository

1. PENDAHULUAN

Perkembangan teknologi yang sangat luas pada era modern sekarang. Sangat membantu manusia dalam bidang apapun baik dari segi pendidikan,kesehatan,sosial dan masih banyak lagi. Dengan kemajuan teknologi yang begitu pesat didunia ini tentunya banyak sekali orang atau hampir semua orang yang memanfaatkan teknologi. Didalam teknologi itu sudah diterapkan ilmu-ilmu yang telah dirancang sehingga bisa mengelola suatu Salah satu cabang ilmu yang bisa dimanfaat dalam membantu manusia memecahkan masalah yaitu *machine learning*.

Machine learning adalah mesin yang dikembangkan untuk bisa belajar dengan sendiri tanpa arahan dari penggunanya. Didalam *machine learning* terdapat banyak algoritma atau metode yang dapat di gunakan



dalam membantu pekerjaan manusia. Dengan adanya algoritma *machine learning* masalah. kita dapat membandingkan antara satu algoritma dengan algoritma lainnya agar kita dapat menemukan algoritma yang lebih akurat.

Dengan *machine learning* bisa membantu untuk mengklasifikasi masalah, masalah yang sering diabaikan atau bahkan belum banyak orang yang mengetahuinya terutama seorang wanita pada saat ini yaitu masalah pada siklus menstruasi atau haid yang tidak teratur. Karena belum banyak orang yang melakukan penjelasan mengenai siklus menstruasi itu sendiri. Menstruasi atau haid merupakan proses kematangan seksual bagi seorang wanita. Menstruasi juga dapat didefinisikan sebagai proses keluarnya darah dari endometrium yang terjadi secara rutin melalui vagina sebagai proses pembersihan Rahim terhadap pembuluh darah, kelenjar-kelenjar dan sel-sel yang tidak terpakai karena tidak adanya pembuahan atau kehamilan.

Wanita akan mengalami masa pubertas dengan mulai mengalami masa menstruasi atau haid. Menstruasi normalnya pada usia 10-15 tahun namun menstruasi juga bisa terjadi lebih awal dari usia normalnya dan bisa juga terlambat dari usia normalnya. Secara umum, menstruasi berlangsung sekali sebulan sampai wanita berusia 45-50 tahun. Masa haid normal bisa terjadi rata-rata 28 hari dan berkisar sekitar 21-35 hari panjang masa haid pada wanita itu bervariasi antara wanita satu dengan lainnya itu berbeda masa haidnya tergantung kondisi setiap wanita.

Siklus menstruasi atau haid setelah hari pertama sampai menstruasi berikutnya terjadi selama 28 hari. Rata-rata siklus menstruasi yang normal itu terjadi sekitar 21-35 hari sedangkan yang termasuk gangguan siklus menstruasi seperti *polimenorrhea* (<20hari), *oligomenorrhea* (>35 hari), dan *Amenorrhea*(>3 bulan)[1]. Waktu menstruasi biasanya antara 3-5 hari, ada yang 1-2 hari diikuti darah sedikit-sedikit kemudian ada yang 7-8 hari. Pada setiap wanita biasanya lama menstruasi itu tetap.jumlah darah yang keluar rata-rata kurang lebih 16 cc, bila lebih dari 80 cc bersifat patologik

Tidak semua wanita mengalami kelancaran menstruasi dengan normal. Ada beberapa wanita yang mengeluh dengan siklus menstruasi yang tidak lancar. Ketidaklancaran menstruasi bisa di sebabkan oleh banyak faktor penyebabnya salah satunya adalah *Amenorrhea* yang masih sering

diabaikan atau bahkan tidak sedikit orang yang belum mengenal istilah *Amenorrhoea* ini. *Amenorrhoea* sendiri terbagi menjadi 2 yaitu *Amenorrhoea* primer yaitu terjadi pada wanita usia 15 tahun yang tidak kunjung mengalami menstruasi walaupun sudah menunjukkan tanda pubertas Dan *Amenorrhoea* sekunder yaitu terjadi pada wanita yang sebelumnya sudah mengalami siklus menstruasi yang normal akan tetapi kemudian tidak menstruasi selama 3 bulan atau lebih secara berturut.

World Health Organization (WHO) perkiraan *Amenorea* pada remaja adalah 10-15%, sedangkan di negara maju seperti: Belanda, persentase *amenorrhoe* cukup besar yaitu 13%. Angka kejadian *amenorea* di Indonesia cukup tinggi. Menurut survei yang dilakukan oleh Departemen Kesehatan pada beberapa sekolah di Indonesia pada tahun 2008. Hasilnya 17.665 remaja putri 6.855 yang mengalami masalah dengan menstruasinya (40%)[2]. Kurangnya pengetahuan seseorang terutama wanita pada permasalahan ini penulis tertarik untuk melakukan penelitian terhadap masalah pada *Amenorrhoea* yang dapat menyebabkan siklus menstruasi atau haid yang tidak teratur atau tidak normal, menggunakan atau memanfaatkan *machine learning*. Karena dengan *machine learning* hasil yang akan diperoleh lebih akurat.

2. METODE

2.1. *Machine Learning*

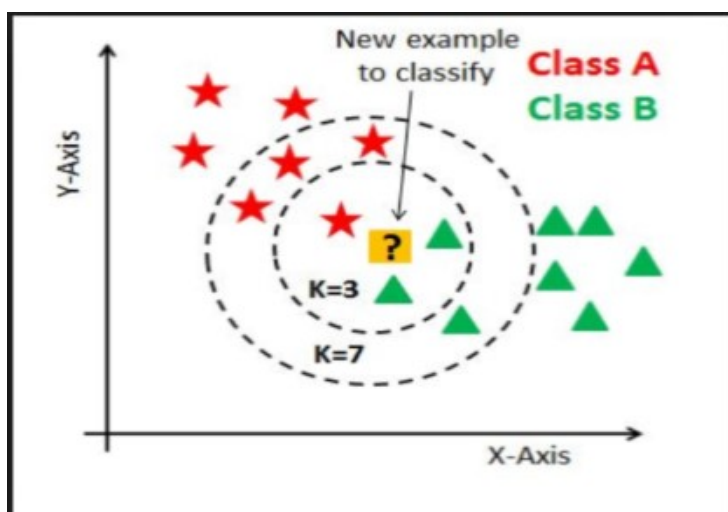
Machine learning adalah bidang dari *artificial intelligence* atau kecerdasan berkaitan dengan pembuatan metode yang dapat deprogram dan dipelajari dari data sebelumnya. Pembelajaran mesin, pengenalan pola, dan penambangan data sering mengacu pada hal yang sama. Bidang probabilitas dan statistik dan tergantung pengoptimalan juga. Dalam penambangan data pembelajaran mesin menjadi alat analisis.[4].

2.2. *K-Nearest Neighbors(k-NN)*

Algoritma klasifikasi pembelajaran mesin langsung dikenal sebagai *K-Nearest Neighbors (k-NN)*. Karena algoritma *k-NN* tidak menggunakan model saat mengklasifikasikan data, maka dikategorikan sebagai algoritma *lazy learner*. Sebagai algoritma klasifikasi, *k-NN* bertujuan untuk memprediksi kategori dari suatu kumpulan data berlabel $S = \{x_i, t_i\}_{i=1}^N$ [7].

Secara umum, algoritma k -NN terdiri dari langkah-langkah sebagai berikut.

- a) Menetapkan sebuah bilangan bulat sebagai nilai k .
- b) Mencari k buah sampel dari data input yang memiliki jarak paling dekat k -nn dengan data baru
- c) Hitung frekuensi setiap kelas dari k -nn tersebut
- d) Data baru diberi label sesuai kelas yang memiliki frekuensi paling tinggi dari k -nn



Gambar 2.1 Klasifikasi menggunakan k -NN

2.3. Decision Tree

Decision tree adalah model pembelajaran *machine learning* data berdasarkan himpunan data $S = \{(x_i, t_i)\}_{i=1}^N$ dimana x_i adalah *variable* data dan t_i adalah *variable* target $t_i \in K$, K adalah himpunan kategori data. *decision tree* paling mudah dipahami jika dibandingkan dengan model pembelajaran mesin lainnya. Dengan membagi data secara rekursif untuk menghasilkan sejumlah pohon klasifikasi hierarkis, *decision tree* melakukan pembelajaran bertahap. model *decision tree* memulai proses pembelajarannya dengan menghitung skor ketidakmurnian (1) dari semua data untuk memilih atribut untuk akar pohon keputusan. Dengan menggunakan subset yang relevan dengan subpohon tersebut, prosedur ini dilakukan secara *rekursif* untuk memilih atribut dari *subtree*[7].

2.4. Support Vecktor Machine

Persamaan berikut dapat digunakan untuk menggambarkan model SVM, yang merupakan model klasifikasi atau *Machine learning*.

$$y(xi) = wtxi + w0. \quad (1)$$

Dimana $y(x_i)$ berfungsi sebagai prediksi dari t_i , w adalah *vector weight* (parameter model), x_i adalah *variable* data, dan w_0 adalah bias.

Untuk mengklasifikasi data dalam ruang tempatnya berada, model SVM bertujuan untuk memprediksi *hyperplane* dalam dimensi m . Subruang dari bidang dua dimensi biasanya disebut sebagai *hyperplane* [7].

2.5. Naïve Bayes

Naïve Bayes adalah metode yang tidak memiliki aturan apapun. Itu melihat frekuensi setiap klasifikasi dalam data pelatihan dan menggunakan bidang matematika yang disebut teori probabilitas untuk mengetahui klasifikasi yang paling mungkin. Klasifikasi statistik yang dikenal sebagai klasifikasi *Naïve Bayes* dapat digunakan untuk memprediksi kemungkinan menjadi anggota suatu kelas. Kemampuan klasifikasi bayes sebanding dengan pohon keputusan dan jaringan saraf. Rumus yang dapat digunakan untuk menghitung probabilitas setiap kemungkinan yang terjadi disediakan oleh Algoritma *Naïve Bayes*, yang menggabungkan probabilitas sebelumnya dengan kemungkinan kondisi. Jenis umum hipotesis bayes sebagai berikut [8].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Dimana:

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik.

$P(H|X)$: Perbandingan hipotesis H berdasar kondisi X

2.6. Amenorrhea

Kita harus menyadari bahwa *Amenorrhea* adalah masalah yang signifikan terutama pada remaja atau wanita. *Amenorrhea* adalah kondisi tidak adanya menstruasi selama setidaknya 3 bulan berkelanjutan atau lebih. Terdapat 2 jenis *Amenorrhea* yaitu *Amenorrhea* primer dan *Amenorrhea*

telah diselesaikan didalam *Business understanding*.

c) Data requirements

Tahapan ketiga didalam metodologi data *science* ini adalah data *requirements*. Pemilihan pendekatan analisis menentukan menentukan data *requirements* atau data apa saja yang dibutuhkan agar permasalahan dapat terjawab.

d) Data collection

Setelah menentukan data *requirements*, data *scientist* mulai mengidentifikasi dan mengumpulkan data yang relevan dengan domain masalah. Tahap ini merupakan tahap ke empat di metodologi *data science*. Adakalanya dalam tahap ini, data yang kita butuhkan tidak tersedia. Maka saat menemukan kendala seperti ini, sangat mungkin bagi kita untuk merevisi kembali data *requirement* dan memutuskan apakah akan mengumpulkan lebih banyak atau lebih sedikit data.

e) Data understanding

Dalam proses ini, kita mencoba mempelajari data dan melihat kualitas data yang kita miliki, apakah sudah cukup baik untuk proses modeling atau belum. Pada tahap kelima dari metodologi data science ini, kita akan mengecek apakah ada *missing values*, data yang *imbalanced*, *outlier*, salah format, dan sebagainya yang harus diperbaiki dulu.

f) Data preparation

Pada tahapan data preparation terdiri dari Semua kegiatan atau proses yang dilakukan untuk membangun *dataset* yang akan digunakan dalam tahap pemodelan, termasuk membersihkan data, menggabungkan data, dan mengubah data menjadi variable yang lebih berguna. Agar dapat diproses secara efektif pada tahap pemodelan, data harus harus dipersiapkan dengan baik dengan membersihkannya dari *missing values*, *invalid values*, dan data duplikat serta memastikan bahwa seluruh data telah memiliki format yang benar. *Feature engineering* juga merupakan bagian dari data preparation. *Feature engineering* adalah proses transformasi data menjadi fitur-fitur yang lebih representatif dan dalam membantu menyelesaikan masalah lebih baik. Fitur-fitur didalam data sangat penting untuk model prediktif dan akan berdampak pada hasil

yang ingin dicapai Data preparation merupakan tahapan yang paling memakan banyak waktu. Sebagian data scientist mengatakan tahap ini bisa menghabiskan sekitar 70% atau bahkan 80% dari keseluruhan proses dalam projek *data science*.

g) Modeling

Modeling atau pemodelan adalah tahap dalam metodologi *data science* dimana *data scientist* membuat model untuk menjawab permasalahan. Pemodelan data berfokus pada mengembangkan model, baik itu model deskriptif atau prediktif. Proses pemodelan untuk model prediktif menggunakan data training. Data training adalah kumpulan data dimana hasil labelnya telah diketahui. Data training bertindak sebagai pengukur apakah model tersebut perlu penyesuaian lagi atau tidak. Proses ini sifatnya iterative atau bisa diulang hingga mendapatkan hasil yang baik.

h) Evaluation

Setelah proses pemodelan, tahap selanjutnya adalah mengevaluasi kualitas model dan mengujinya apakah dapat mengatasi permasalahan bisnis dengan tepat. Evaluasi model dapat memiliki dua fase yaitu, fase *diagnostic measures* dan fase *statistical significance testing*. *Diagnostic measures* digunakan untuk memastikan bahwa data yang digunakan telah ditangani dan diinterpretasikan dengan benar didalam model.

2.9. Metode Pengujian dan pengolahan data

Pada metode pengujian dan pengolahan data ini penulis melakukan pengujian dengan menggunakan tahapan-tahapan pada perhitungan *confusion matrix*. untuk menghitung nilai akurasi dibutuhkan dengan rumus dibawah ini.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Keterangan

TP : True Positive

TN : True Negative

FP : False Positive

3. HASIL DAN PEMBAHASAN

Pada penelitian ini setelah melalui tahapan pengujian dan analisa terhadap 4 algoritma yaitu k-NN,SVM,Decision Tree,dan Naïve Bayes maka hasil yang diperoleh dibawah ini.

a) Pengujian perhitungan akurasi pada algoritma k-NN

K-NN :

In [32]:

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(X_train, y_train)
print('Accuracy of K-NN classifier on training set: {:.2f}'
      .format(knn.score(X_train, y_train)))
print('Accuracy of K-NN classifier on test set: {:.2f}'
      .format(knn.score(X_test, y_test)))
```

Accuracy of K-NN classifier on training set: 0.99
Accuracy of K-NN classifier on test set: 0.99

Gambar 3 Hasi algoritma k-NN

Setelah melakukan pengujian menggunakan bahasa pemrograman phyton dan perhitungan confusion matrix dengan menggunakan dataset 510 data dengan algoritma k-NN maka hasil nilai akurasi nya sebesar 0.99.

b) Pengujian perhitungan akurasi pada algoritma SVM

```
from sklearn.svm import SVC
svm = SVC()
svm.fit(X_train, y_train)
print('Accuracy of SVM classifier on training set: {:.2f}'
      .format(svm.score(X_train, y_train)))
print('Accuracy of SVM classifier on test set: {:.2f}'
      .format(svm.score(X_test, y_test)))
```

Accuracy of SVM classifier on training set: 0.83
Accuracy of SVM classifier on test set: 0.79

Gambar 4. Hasil algoritma SVM

Setelah melakukan pengujian menggunakan bahasa pemrograman phyton dan perhitungan confusion matrix dengan menggunakan dataset

510 data dengan algoritma SVM maka hasil nilai akurasi nya sebesar 0.79

c) Pengujian perhitungan akurasi pada algoritma Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier().fit(X_train, y_train)
print('Accuracy of Decision Tree classifier on training set: {:.2f}'
      .format(clf.score(X_train, y_train)))
print('Accuracy of Decision Tree classifier on test set: {:.2f}'
      .format(clf.score(X_test, y_test)))
```

Accuracy of Decision Tree classifier on training set: 1.00
Accuracy of Decision Tree classifier on test set: 0.98

Gambar 5. Hasil algoritma decision Tree

Setelah melakukan pengujian menggunakan bahasa pemrograman python dan perhitungan confusion matrix dengan menggunakan dataset 510 data dengan algoritma decision tree maka hasil nilai akurasi nya sebesar 0.98

d) Pengujian perhitungan akurasi pada algoritma Naïve Bayes

```
from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
gnb.fit(X_train, y_train)
print('Accuracy of GNB classifier on training set: {:.2f}'
      .format(gnb.score(X_train, y_train)))
print('Accuracy of GNB classifier on test set: {:.2f}'
      .format(gnb.score(X_test, y_test)))
```

Accuracy of GNB classifier on training set: 0.99
Accuracy of GNB classifier on test set: 0.98

Gambar 6. Hasil algoritma Naïve Bayes

Setelah melakukan pengujian menggunakan bahasa pemrograman python dan perhitungan confusion matrix dengan menggunakan dataset 510 data dengan algoritma Naïve Bayes maka hasil nilai akurasi nya sebesar 0.98.

5. KESIMPULAN

Berdasarkan hasil dari perbandingan algoritma k-NN, SVM, Decision Tree, dan Naïve Bayes untuk melakukan klasifikasi terhadap 510 data dengan 2 jenis kelas kategori yaitu berbahaya atau tidak berbahaya (1

atau 2), hasilnya menunjukkan bahwa algoritma k-NN lebih baik digunakan dibandingkan algoritma pembandingan lainnya, hasil akurasi k-NN yaitu sebesar 0.99 sedangkan nilai akurasi pada algoritma Decision Tree yaitu 0.98, disusul dengan Naïve bayes yang nilai akurasinya juga 0.98 dan yang paling rendah nilai akurasinya yaitu algoritma SVM, yang akurasi hanya bernilai 0.79, maka dapat disimpulkan dari perbandingan ke 4 algoritma pada penelitian ini, menunjukkan bahwa algoritma k-NN lebih baik.

DAFTAR PUSTAKA

- [1] L. Sitoayu, D. A. Pertiwi, and E. Y. Mulyani, "Kecukupan zat gizi makro, status gizi, stres, dan siklus menstruasi pada remaja," *J. Gizi Klin. Indones.*, vol. 13, no. 3, p. 121, 2017, doi: 10.22146/ijcn.17867.
- [2] C. Magdalena and T. Bolon, "Gambaran Faktor – Faktor Penyebab Terjadinya Amenorea Pada Remaja Putri Di Desa Helvetia Timur Kecamatan Helvetia Medan Periode April-Juni Tahun 2012," *J. Ilm. Keperawatan*, vol. 1, no. 1, pp. 16–20, 2015.
- [3] A. Isroqmi, "Kemampuan Mahasiswa Memahami Logika Pemrograman Komputer Melalui Algoritma," *Nabla Dewantara J. Pendidik. Mat.*, vol. 2, no. 2, pp. 59–73, 2017.
- [1] L. Sitoayu, D. A. Pertiwi, and E. Y. Mulyani, "Kecukupan zat gizi makro, status gizi, stres, dan siklus menstruasi pada remaja," *J. Gizi Klin. Indones.*, vol. 13, no. 3, p. 121, 2017, doi: 10.22146/ijcn.17867.
- [2] C. Magdalena and T. Bolon, "Gambaran Faktor – Faktor Penyebab Terjadinya Amenorea Pada Remaja Putri Di Desa Helvetia Timur Kecamatan Helvetia Medan Periode April-Juni Tahun 2012," *J. Ilm. Keperawatan*, vol. 1, no. 1, pp. 16–20, 2015.
- [3] A. Isroqmi, "Kemampuan Mahasiswa Memahami Logika Pemrograman Komputer Melalui Algoritma," *Nabla Dewantara J. Pendidik. Mat.*, vol. 2, no. 2, pp. 59–73, 2017.
- [4] Dwi Retnosari, "Sistem Aplikasi Data Mining Untuk Menampilkan," *J. Integr. Sist. Ind. UMJ*, vol. 1, no. 2, pp. 13–20, 2014.
- [5] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 2, p. 126, 2016, doi: 10.35314/isi.v1i2.131.
- [6] Sitti Rohmah and Y. P. Rusady, "Hubungan Status Gizi Dengan Kejadian Amenorea Pada Siswi SMP 1 Pandewangi," *J. Satuan Bakti Bidan Untuk Negeri (Sakti Bidadari)*, vol. 4, no. 2, 2021.
- [7] D. Y. Heryadi and M. C. Teguh Wahyono, *machine learning konsep*

- dan implementasi*, Cetakan 1. Yogyakarta: GAVA MEDIA, 2020.
- [8] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 20–28, 2018.
- [9] A. N. Syahrudin and T. Kurniawan, "Input dan Output pada Bahasa Pemrograman Python," *J. Dasar Pemrograman Python STMIK*, no. June 2018, pp. 1–7, 2018.
- [10] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.
- [11] A. M. Argina, "Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [12] S. AULIA, S. HADIYOSO, and D. N. RAMADAN, "Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 3, no. 1, p. 75, 2015, doi: 10.26760/elkomika.v3i1.75.
- [13] M. R. Yuliansyah, M. B, and A. Franz, "Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda," *Adopsi Teknol. dan Sist. Informasi(ATASI)*, vol. 1, no. 1, pp. 8–20, 2022.
- [14] S. Wiyono, "Perbandingan Algoritma Machine Learning SVM dan Decision Tree untuk Prediksi Keaktifan Mahasiswa," *Sinkron*, vol. 3, no. 1, pp. 105–108, 2018.
- [15] T. A.M and A. Yaqin, "Perbandingan Algoritma Naïve Bayes, K-Nearest Neighbors dan Random Forest untuk Klasifikasi Sentimen Terhadap BPJS Kesehatan pada Media Twitter," *InComTech J. Telekomun. dan Komput.*, vol. 12, no. 1, p. 01, 2022, doi: 10.22441/incomtech.v12i1.13642.
- [16] S. Shedriko, "Perbandingan Algoritma SVM dan KNN dalam Mengklasifikasi Kelulusan Mahasiswa pada Suatu Mata Kuliah," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 2, p. 115, 2021, doi: 10.30998/string.v6i2.9160.
- [17] A. Delaney, "Burden of Rare Variants in Hypothalamic Amenorrhea-Supplemental data", 2020, doi: 10.6084/m9.figshare.12730694.v3.