# Deteksi PCOS pada Wanita Menggunakan Explanatory Data Analysis (EDA) dan Support Vector Machine (SVM)

**Weny Mistarika Rahmawati[1], Renovita Edelani[2]**
[1,2]Program Studi Teknik Informatika, Departemen teknik Informatika dan Komputer,
Politeknik Elektronika Negeri Surabaya
Email: [1]weny@pens.ac.id, [2]renovita@pens.ac.id

***Abstract.*** *Polycystic Ovarian Syndrome (PCOS) is a hormonal disorder that occurs in women of productive age and can cause infertility. The exact cause of PCOS is not yet known, but genetic factors and lifestyle are factors that can influence a woman to get PCOS. This study aims to detect PCOS based on data consisting of age, body mass index, testosterone levels and the number of follicles. The initial data obtained has a poor distribution or can be said to be unbalanced. Researchers conducted Explanatory Data Analysis (EDA) in the early stages by creating a scatterplot to find the correlation of each feature with the target class. The result is a menstrual regularity feature whose omission value greatly affects PCOS detection so that data is deleted at the null value in the feature. After that, Support Vector Machine (SVM) classification was carried out to separate the class diagnosed with PCOS or not. Several SVM kernels were tested to determine the best results that could be produced. Evaluation was carried out by calculating accuracy, precision, recall and f1-score on the confusion matrix formed. The results of the study showed that the polynomial kernel provided the best classification results with an accuracy of 89.62%, precision 81.08%, recall 88.23% and f1-score 84.5%. This study confirms that the combination of EDA and SVM can be used as an effective approach in supporting PCOS detection.*
***Keywords:*** *classificaton, EDA, PCOS detection, SVM*

***Abstrak.*** *Polycystic Ovarium Syndrom (PCOS) merupakan gangguan hormonal yang terjadi pada Wanita usia produktif dan bisa mengakibatkan infertilitas. PCOS sendiri belum diketahui penyebab pastinya tetapi factor genetic dan pola hidup merupakan factor yang dapat memengaruhi seorang Wanita terkena PCOS. Penelitian ini bertujuan melakukan deteksi PCOS berdasarkan data yang terdiri dari umur, indeks masa tubuh, level testosterone serta jumlah folikel. Data awal yang didapat memiliki distribusi yang tidak baik atau bisa dikatakan tidak seimbang. Peneliti melakukan Explanatory Data Analysis (EDA) pada tahap awal dengan membuat scatterplot untuk mencari korelasi setiap fitur dengan kelas target. Hasilnya ada fitur keteraturan haid yang nilai kealpaannya sangat mempengaruhi deteksi PCOS sehingga dilakukan penghapusan data pada nilai null pada fitur tersebut. Setelah itu dilakukan klasifikasi Support Vector Machine (SVM) untuk memisahkan kelas terdiagnosa PCOS atau tidak. Beberapa kernel SVM diujikan untuk mengetahui hasil terbaik yang bisa dihasilkan. Evaluasi dilakukan dengan menghitung akurasi, precision, recall dan f1-score pada confussion matrix yang terbentuk. Hasil dari penelitian menunjukkan bawa kernel polynomial memberikan hasil klasifikasi terbaik dengan akurasi sebesar 89,62%, precision 81,08%, recall 88,23% dan f1-score 84,5%. Penelitian ini mengonfirmasi bahwa kombinasi EDA dan SVM dapat digunakan sebagai pendekatan yang efektif dalam mendukung deteksi PCOS.*
***Kata Kunci:*** *deteksi PCOS, Klasifikasi, EDA, SVM*

## 1. Pendahuluan

Polycystic Ovary Syndrome atau biasa disebut PCOS merupakan gangguan endokrin yang memengaruhi sekitar 8-13% wanita usia produktf (Lizneva et al., 2016). PCOS ditandai oleh ketidakseimbangan hormon yang menyebabkan gangguan ovulasi, hiperandrogenisme, dan terbentuknya kista kecil di ovarium. Kondisi ini tidak hanya berpengaruh pada kesuburan, tetapi juga berhubungan dengan risiko penyakit metabolik seperti resistensi insulin, diabetes tipe 2, dan penyakit kardiovaskular (Teede et al., 2018). Penelitian tentang deteksi pada berbagai bidang khusunya bidang

medis saat ini masih banyak diminati. Mulai deteksi penyakit jantung, TBC, kanker dan penyakit lain masih banyak dikembangkan dengan harapan akan memberikan dampak yang signifikan pada dunia kedokteran. Data yang digunakan bisa merupakan data teks atau citra (Liantoni et al., 2021). PCOS menjadi salah satu topik yang sampai saat ini masih dikembangkan mengingat berkembangnya kemungkinan-kemungkinan fitur yang mungkin menjadi penyebab dari PCOS (Pitaloka & Kusnawi, 2023).

Dalam beberapa tahun terakhir, penerapan *Exploratory Data Analysis* (EDA) banyak digunakan dalam melakukan analisis data khususnya data medis dan telah mampu memberikan wawasan yang lebih mendalam terkait pola dan tren dalam dataset (Nur Rizki et al., 2024). EDA dilakukan sebelum data masuk ke proses utama. Dengan menggunakan EDA memungkinkan melakukan identifikasi fitur yang berkontribusi secara signifikan terhadap kelas target yang diinginkan mulai dari sisi distribusi, frekuensi korelasi dan lainnya. EDA dapat dilakukan dengan memberikan visualisasi seperti histogram atau scatter plot sehingga dapat lebih mudah untuk dilakukan analisis(Wibowo, 2022).

Support Vector Machine (SVM) sebagai salah satu algoritma klasifikasi masih menjadi algoritma andalan yang digunakan dalam menentukan kelas dari berbagai dataset mulai dari bidang pertanian, pendidikan hingga medis. Algoritma ini tidak terlalu kompleks namun mampu memberikan hasil klasifikasi yang baik khusunya untuk data teks. Tidak hanya untuk dataset yang sifatnya linear, SVM juga mampu memisahkan data yang sifatnya nonlinear dengan cara mengubah dimensi data menjadi lebih tinggi sehingga lebih mudah untuk dicari pemisahnya. Beberapa kernel dapat digunakan untuk mengubah dimensi data mulai dari polynomial, sigoid, rbf ,dll sehingga SVM mampu memberikan hasil terbaik dari pada algoritma lainnya. Meskipun begitu, beberapa penelitian mengkombinasikan SVM dengan algoritma lain untuk meningkatkan kinerja dari SVM (Subarkah et al., 2022).

Penelitian ini bertujuan untuk menerapkan EDA dalam menganalisis fitur pasien PCOS serta membangun model klasifikasi menggunakan SVM untuk mendeteksi PCOS berdasarkan fitur klinis yang signifikan. Parameter pengujian yang akan dicobakan adalah beberapa kernel SVM yaitu linear, rbf dan polynomial.Hasil akan dievaluasi menggunakan beberapa parameter yaitu akurasi, *precision*, *recall* dan *f1-score* mengunakan *confussion matrix*.

## 2. Tinjauan Pustaka

### 2.1 Polycystic Ovarium Syndrom (PCOS)

Merupakan kelainan yang mengakibatkan gangguan pada hormon reproduksi dan metabolisme pada Wanita. PCOS dapat menyebabkan menstuasi tidak teratur hingga infertilitas. Penyebab pasti dari PCOS belum diketahui secara pasti, namun penelitian menyebutkan factor genetic, gaya hidup dan lingkungan dapat menjadi salah satu factor penyebabnya. Gaya hidup yang kurang sehat seperti makan junk food, kurang istirahat hingga stress berlebihan dapat meningkatkan kemungkinan seorang Wanita mengalami PCOS (Kaur et al., 2022).

### 2.2 Data Tidak Seimbang (*Imbalanced data)*

Merupakan kondisi dataset dimana perbandingan jumlah data yang masuk kelas target tidak seimbang. Pada dataset tidak seimbang memiliki 2 kelas berdasarkan jumlahnya yaitu kelas mayoritas dan kelas minoritas. Dataset jenis ini biasa ditemui pada data penipuan kartu kredit atau penyakit langka dimana kemungkinan terjadinya memang sangat kecil(Wijaya et al., 2024). Data tidak seimbang memerlukan penanganan khusus. Hal ini disebabkan karena apabila data tak seimbang digunakan maka akan mengakibatkan terjadi *overfitting* pada hasil klasifikasi. Penanganan yang dapat dilakukan pada data tak seimbang diantaranya yaitu melakukan Teknik resampling pada data (Matharaarachchi et al., 2024). Resampling dapat dilakukan dengan menambah jumlah data pada data yang masuk kelas minor (*over sampling*) atau dengan mengurangi data yang masuk dalam kelas mayor (*under sampling*). Beberapa contoh metode yang dapat digunakan untuk melakukan *over sampling* diantaranya adalah *Synthetic Minority Over Sampling Technique* (SMOTE) dan *Adaptive Synthetic Sampling* (ADASYN)(Pamuji & Putri, 2023). Sedangakan Teknik untuk melakukan *under sampling* diantaranya adalah *Rondom Under Sampling*, yaitu dengan menghapus data pada kelas mayoritas secara acak atau

dapat juga menggunakan metode lain seperti Tomek Link. Beberapa penelitian menggabungkan *oversampling* dan *undersampling* untuk mendapatkan hasil lebih optimal(Hidayat et al., 2021).

### 2.3 Explanatory Data Analysis (EDA)

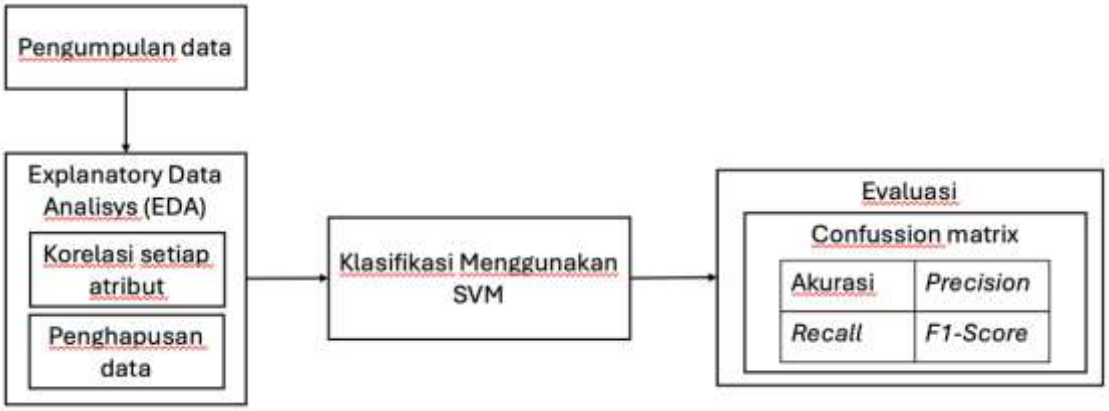
Explanatory Data Anlysis (EDA) merupakan pendekatan analisa statistic yang digunakan untuk memberikan wawasan lebih dalam mengenai pola dan tren dalam dataset. Dengan menggunakan EDA fitur-fitur yang ada dianalisa terlebih dahulu sebelum masuk ke proses utama. Analisa yang dilakukan bertujuan untuk mengenali pola, mencari anomaly, menguji asumsi, atau mencari hubungan fitur dengan kelas target yang diharapkan(Mehta et al., 2024). EDA dapat dilakukan dengan memberikan visualisasi data. Visualisasi dapat berupa histogram, scatter plot, box plot atau juga dapat berupa analisis deskriptif seperti mencari mean, median dan standar deviasi data (Dhany et al., 2023). Dengan menggunakan EDA diharapkan mampu mengoptimalkan hasil klsaifikasi dengan mendaatkan hasil evaluasi yang lebih baik.

### 2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu algoritma *machine learning* yang digunakan untuk melakukan klasifikasi atau pemisahan kelas dengan cara mencari *hyperplane* terbaik dalam memisahkan dua kelas. Yang disebut sebagai *hyperplane* terbaik disini adalah *hyperplane* yang memiliki margin terbesar, yaitu jarak terjauh dari data titik terdekat di kedua kelasnya. *Hyperplane* dapat berupa garis, bidang atau hiperbidang(Alfian et al., 2022). Tidak hanya mampu memisahkan data secara linear, SVM memungkinkan memisahkan kelas dengan mengubah data ke dimensi lebih tinggi sebelum dipisahkan. SVM menggunakan kernel untuk mengubah data ke dimensi lebih tinggi dengan harapan dapat memudahkan pemisahan data menjadi lebih baik. Beberapa kernel yang sering digunakan adalah kernel Radial Basis Function (RBF), kernel polynomial, kernel sigmoid dll(Wahyuni & Kusumodestoni, 2024). Persebaran data menjadi penentu kernel mana yang baik digunakan untuk memisahkan kelas target.

## 3. Metode Penelitian

Pada bagian ini dijelaskan alur dari penelitian ini mulai dari pengumpulan data, preprosesing, klasifikasi hingga evaluasi. Penelitian ini secara garis besar menggunakan metodologi yang dapat dilihat pada Gambar 1.




**Gambar 1 Diagram Alur Penelitian**

Sebagaimana yang dapat dilihat pada Gambar 1, tahap pertama pada penelitian ini adalah tahap pengumpulan data dimana peneliti mendapatkan data dari www.kaggle.com dalam format csv. Data memiliki 5 fitur dan 1 kelas target. Fitur fitur pada dataset yang digunakan dijelaskan pada tabel 1. Sementara target kelasnya adalah PCOS Diagnosis yang berisi nilai yes dan no yang menandakan apakah seseorang terdeteksi PCOS atau tidak. Data berjumlah 1000 baris dan memiliki distribusi data yang tidak seimbang. Data kelas No mencapai 800 data sedangkan untuk data yang masuk kelas yes (PCOS) ada 200 data sehingga data tersebut masuk sebagai *imbalanced data*.

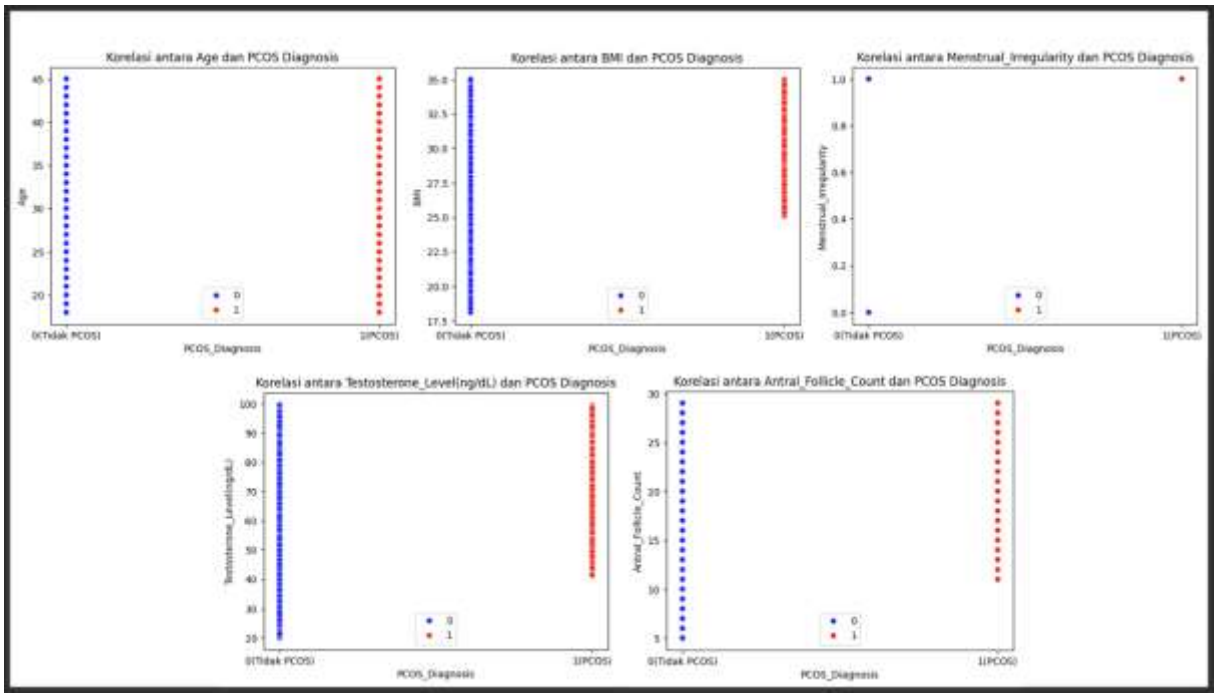
 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

**Tabel 1 Dataset PCOS yang Digunakan**

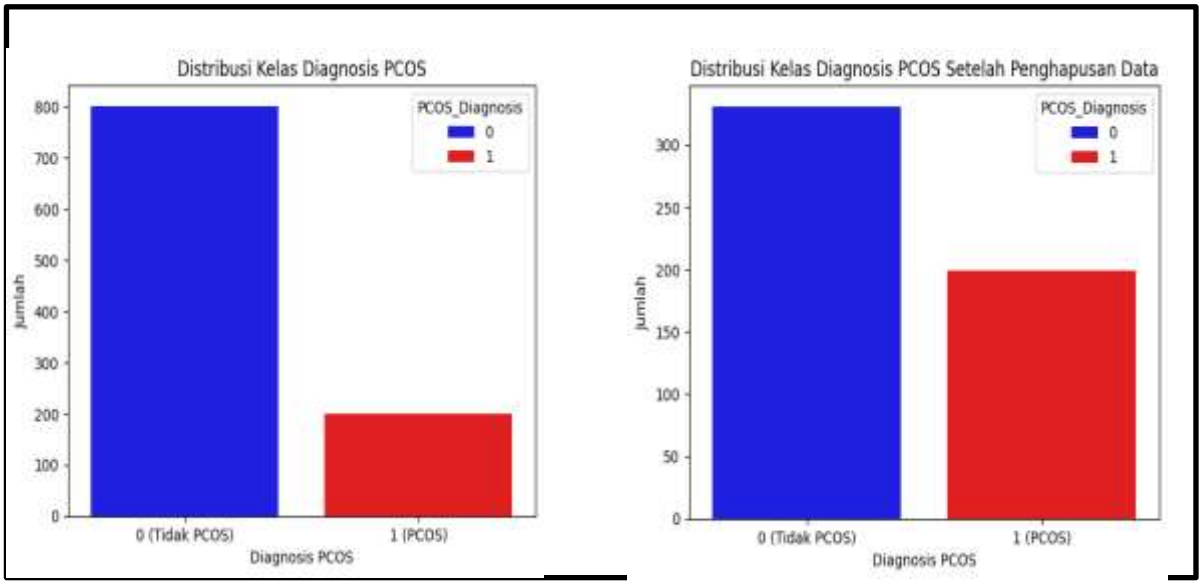
| Nama fitur | Keterangan |
|---|---|
| Age | Umur pasien 18-45 tahun |
| BMI | Indeks masa tubuh |
| Menstrual_Irregularity | Biner apakah seorang wanita mengalami haid tidak teratur (Nilai 0 jika teratur, dan 1 jika tidak teratur) |
| Testosterone_Level (ng/dL) | Level testosterone pada darah pasien |
| Antral_Follicle_Count | Jumlah folikel antral yang terdeteksi selama USG |

Sebelum melakukan penanganan untuk data tidak seimbang, peneliti melakukan EDA dengan membuat scatter plot untuk melihat hubungan setiap fitur yang ada dengan kelas target. Hasilnya dapat dilihat pada Gambar 2. Dari Gambar 2 dapat dilihat bahwa pada fitur *menstrual_irregularity* memiliki pola yang cukup kuat untuk kelas terdiagnosa PCOS. Semua data yang terdiagnosa PCOS selalu memiliki siklus menstruasi yang tidak teratur. Atau dalam dataset ini dinyatakan dengan nilai 1 pada fitur *menstrual_irregularity*. Sedangkan tidak semua yang bernilai 1 pada fitur *menstrual_irregularity* terdiagnosa sebagai PCOS. Berdasarkan data tersebut, peneliti melakukan proses penghapusan data pada semua data yang bernilai 0 pada fitur *menstrual_irregularity*. Hasil penghapusan data ini menyebabkan nilai pada atribut *menstrual_irregularity* hanya memiliki 1 nilai, sehingga keberadaannya pada pemodelan dapat diabaikan. Hasil distribusi kelas sebelum dan sesudah dilakukan penghapusan data dapat dilihat pada Gambar 3.

Berdasarkan Gambar 3, saat ini distribusi jumlah kelas PCOS dan tidak PCOS lebih seimbang walaupun masih terdapat perbedaan jumlah tetapi masih dapat dikategorikan dataset yang seimbang. Langkah berikutnya adalah memisahkan antara data training dan data testing dengan data testing sebanyak 20% dari jumlah data secara random. Klasifikasi yang digunakan adalah klasifikasi menggunakan SVM. Peneliti mengujikan beberapa kernel SVM yaitu kernal linear, rbf dan polynomial sebagai parameter pengujian. Pada kernel RBF menggunakan gamma =0,1 dan pada kernel polynomial menggunakan derajad=3.




**Gambar 2 Korelasi Setiap Fitur Terhadapa Kelas Target**




**Gambar 3 Distribusi Jumlah Kelas Sebelum dan Setelah Penghapusan Data**

## 4. Hasil Analisa

Percobaan yang dilakukan melibatkan 3 kernel SVM yang masing-masing akan diuji cobakan pada dataset. SVM dengan kernel linear, rbf dan polynomial akan diujikan dan dianalisa hasilnya. Hasil yang diamati akan meliputi hasil berupa akurasi, *presicion, recall dan F1-score* dengan menggunakan *confussion matrix*. Rumus untuk mendapatkan masing-masing nilai tersebut adalah sebagai berikut:

Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan seluruh data yang diuji. Untuk mendapatkan akurasi dapat menggunakan persamaan (1).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

*Precision* mengukur seberapa banyak prediksi positif yang benar di antara semua prediksi positif. Untuk mendapatkan nilai precision dapat menggunakan persamaan (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

*Recall* digunakan untuk mengukur sensitivitas yaitu seberapa banyak data positif yang berhasil diprediksi dengan benar. Recall dapat dihitung menggunakan persamaan (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

*F1-Score* adalah rata-rata harmonis dari Precision dan Recall. Persamaan (4) merupakan persamaan untuk mendapatkan nilai F1-score.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

Di mana: TP = *True Positives* (jumlah prediksi positif yang benar); TN = *True Negatives* (jumlah prediksi negatif yang benar); FP = *False Positives* (jumlah prediksi positif yang salah); FN = *False Negatives* (jumlah prediksi negatif yang salah)

Hasil pengujian yang dihasilkan dapat dilihat pada tabel 2. Pada tabel 2 disajikan hasil akurasi, *presicion*, *recall* dan *F1-score* hasil klasifikasi menggunakan SVM dengan basis kernel linear, rbf dan polynomial. Berdasarkan Tabel 2 dapat dilihat bahwa model terbaik dicapai oleh SVM yang menggunakan kernel polynomial. Hal ini berarti data yang digunakan tidak terpisah secara linear dan memiliki pola interaksi antar fitur yang tidak terlalu kompleks yang bisa diselesaikan dengan fungsi polynomial. Model SVM dengan kernel polynomial ini rentan terhadap o*verfitting* apabila pemilihan derajadnya terlalu tinggi dan memerlukan waktu dan memori lebih tinggi dari sisi komputasi sehingga model ini mungkin kurang cocok untuk data yang memiliki fitur besar. Data PCOS yang digunakan pada penelitian ini memiliki jumlah fitur yang masih bisa ditangani dengan baik oleh SVM kernel polynomial.

**Tabel 2 Hasil Pengukuran Kinerja SVM**

| **Kernel SVM** | **Akurasi** | ***Precision*** | ***Recall*** | ***F1-score*** |
|---|---|---|---|---|
| Linear | 82,07% | 69,23% | 79,41% | 73,97% |
| RBF | 87,73% | 76,92% | 88,23% | 82,19% |
| Polynomial | 89,62% | 81,08% | 88,23% | 84,5% |

## Kesimpulan

Pada penelitian ini telah dilakukan deteksi PCOS dengan dataset yang diperoleh dari www.kaggle.com yang memiliki 5 fitur dan satu kelas target. Dataset awal merupakan dataset tidak seimbang dengan jumlah distribusi data 200 dan 800 untuk setiap kelas terdeteksi PCOS dan tidak. Dataset terlebih dahulu dilakukan Explanatory Data Analysis (EDA) dengan mencari korelasi setiap fitur yang ada terhadap kelas target dengan menggunakan *scatter plot*. Hasilnya dapat diketahui bahwa ada satu fitur yaitu menstrual_irregularity yang nilainya memiliki korelasi yang kuat dengan kelas target. Peneliti melakukan penghapusan data yang memiliki nilai yang tidak berkolerasi terhadap target. Setelah penghapusan data dilakukan jumlah distribusi data menjadi seimbang sehingga tidak perlu dilakukan penanganan khusus data tidak seimbang. Hasil dataset dilakukan klasifikasi menggunakan algoritma SVM dengan beberapa kernel yaitu linear, rbf dan polynomial. Setelah dilakukan evaluasi, model terbaik didapatkan oleh SVM kernel polynomial dengan akruasi 89,62%, *precision* 81,08%, *recall* 88,23% dan *F1-score* sebesar 84,5%.

## Referensi

Alfian, G., Syafrudin, M., Fahrurrozi, I., Fitriyani, N. L., Atmaji, F. T. D., Widodo, T., Bahiyah, N., Benes, F., & Rhee, J. (2022). Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. *Computers 2022, Vol. 11, Page 136*, *11*(9), 136. https://doi.org/10.3390/COMPUTERS11090136

Dhany, H. W., Sutarman, S., & Izhari, F. (2023). Exploratory Data Analysis (EDA) methods for healthcare classification. *Journal of Intelligent Decision Support System (IDSS)*, *6*(4), 209–215. https://doi.org/10.35335/IDSS.V6I4.165

Hidayat, W., Ardiansyah, M., & Setyanto, A. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Edumatic: Jurnal Pendidikan Informatika*, *5*(1), 11–20. https://doi.org/10.29408/EDUMATIC.V5I1.3125

Kaur, R., Kumar, R., & Gupta, M. (2022). Food Image-based diet recommendation framework to overcome PCOS problem in women using deep convolutional neural network. *Computers and Electrical Engineering*, *103*, 108298. https://doi.org/10.1016/J.COMPELECENG.2022.108298

Liantoni, F., Rozi, N. F., Indriyani, T., Rahmawati, W. M., & Hapsari, R. K. (2021). Gradient based ant spread modification on ant colony optimization method for retinal blood vessel edge detection. *IOP Conference Series: Materials Science and Engineering*, *1010*(1), 012021. https://doi.org/10.1088/1757-899X/1010/1/012021

Lizneva, D., Suturina, L., Walker, W., Brakta, S., Gavrilova-Jordan, L., & Azziz, R. (2016). Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertility and Sterility*, *106*(1), 6–15. https://doi.org/10.1016/J.FERTNSTERT.2016.05.003

Matharaarachchi, S., Domaratzki, M., & Muthukumarana, S. (2024). Enhancing SMOTE for imbalanced data with abnormal minority instances. *Machine Learning with Applications*, *18*, 100597. https://doi.org/10.1016/J.MLWA.2024.100597

Mehta, V., Batra, N., Poonam, Goyal, S., Kaur, A., Dudekula, K. V., & Victor, G. J. (2024). Machine Learning based Exploratory Data Analysis (EDA) and Diagnosis of Chronic Kidney Disease (CKD). *EAI Endorsed Transactions on Pervasive Health and Technology*, *10*(1), 1. https://doi.org/10.4108/EETPHT.10.5512

Nur Rizki, I., Lia Puspita, M., Prayoga, D., Qomarul Huda, M., & Syarif Hidayatullah Jakarta, U. (2024). IMPLEMENTASI EXPLORATORY DATA ANALYSIS UNTUK ANALISIS DAN VISUALISASI DATA PENDERITA STROKE KALIMANTAN SELATAN MENGGUNAKAN PLATFORM TABLEAU. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(1), 2830–7062. https://doi.org/10.23960/JITET.V12I1.3856

Pamuji, F. Y., & Putri, S. D. A. (2023). Komparasi metode SMOTE dan ADASYN untuk penanganan data tidak seimbang MultiClass. *Jurnal Informatika Polinema*, *9*(3), 331–338. https://doi.org/10.33795/JIP.V9I3.1330

Pitaloka, N. T., & Kusnawi, K. (2023). PCOS DISEASE CLASSIFICATION USING FEATURE SELECTION RFECV AND EDA WITH KNN ALGORITHM METHOD. *Jurnal Teknik Informatika (Jutif)*, *4*(4), 693–701. https://doi.org/10.52436/1.JUTIF.2023.4.4.831

Subarkah, M. A., Rahmawati, W. M., Wardhana, S. R., Hapsari, R. K., Informatika, T., Adhi, T., & Surabaya, T. (2022). Analisis Sentimen Terhadap Video Ulasan Produk Menggunakan Metode Support Vector Machine Dengan Sequential Minimal Optimization. *KERNEL: Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika*, *3*(2), 17–24. https://doi.org/10.31284/J.KERNEL.2022.V3I2.4039

Teede, H. J., Misso, M. L., Costello, M. F., Dokras, A., Laven, J., Moran, L., Piltonen, T., Norman, R. J., Andersen, M., Azziz, R., Balen, A., Baye, E., Boyle, J., Brennan, L., Broekmans, F., Dabadghao, P., Devoto, L., Dewailly, D., Downes, L., … Yildiz, B. O. (2018). Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Human Reproduction*, *33*(9), 1602–1618. https://doi.org/10.1093/HUMREP/DEY256

Wahyuni, S. D., & Kusumodestoni, R. H. (2024). Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting. *Bulletin of Information Technology (BIT)*, *5*(2), 56–64. https://doi.org/10.47065/BIT.V5I2.1247

Wibowo, A. (2022). Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(3), 2292–2304. https://doi.org/10.35957/jatisi.v9i3.2737

Wijaya, M. G., Pinaringgi, M. F., Zakiyyah, A. Y., & Meiliana. (2024). Comparative Analysis of Machine Learning Algorithms and Data Balancing Techniques for Credit Card Fraud Detection. *Procedia Computer Science*, *245*(C), 677–688. https://doi.org/10.1016/J.PROCS.2024.10.294

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)