**IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING DAN SUPPORT VECTOR MACHINE UNTUK MENGETAHUI FAKTOR PEMICU DIABETES MELITUS DI**

## **Nama Tidak Ditemukan**

UNIVERSITAS MUHAMMADIYAH MAKASSAR 2024

e-mail: xxxx@xxxx.xxx\*

***Abstract***

*Keywords: : Algorithm Implementation, K-Means Algorithm,, SVM Algorithm, Lappae Health Center, Diabetes Mellitus*

**Abstrak**

Kata Kunci: : Implementasi Algoritma, Algoritma K-Means, Algoritma SVM, Puskesmas Lappae, Diabetes Melitus

**1. Pendahuluan**

Diabetes mellitus merupakan suatu kondisi jangka panjang yang ditunjukkan oleh kadar gula darah tinggi. Ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat menggunakan insulin secara efektif, kondisi ini dikenal sebagai diabetes mellitus. Salah satu efek paling umum dari diabetes mellitus yang tidak terkontrol adalah hiperglikemia atau peningkatan gula darah, yang secara bertahap dapat mengancam banyak sistem tubuh, terutama saraf dan pembuluh darah (WHO, 2019).

Diabetes mellitus (DM) merupakan penyakit kronis yang terjadi ketika tubuh tidak dapat menghasilkan cukup insulin (hormon yang mengontrol gula darah atau glukosa) karena gangguan pankreas, atau tubuh tidak dapat secara efisien memanfaatkan insulin yang diproduksi (Arania dkk, 2021). Di Indonesia, diabetes mellitus menjadi salah satu masalah kesehatan yang semakin meningkat. Puskesmas Lappae di Kabupaten Sinjai, Sulawesi Selatan, merupakan salah satu fasilitas kesehatan yang menghadapi banyak tantangan dalam menemukan dan mengelola pasien diabetes. Dengan jumlah penderita diabetes yang meningkat di wilayah ini, pendekatan yang lebih efektif diperlukan untuk mengidentifikasi dan mengelompokkan pasien berdasarkan faktor-faktor risiko yang mereka miliki.

Algoritma K-Means merupakan salah satu teknik klaster non-hirarki yang bertujuan untuk mengelompokkan variabel ke dalam kelas-kelas berdasarkan kemiripan. Teknik ini membagi data yang tersedia ke dalam satu atau lebih kelompok, sehingga data yang memiliki karakteristik serupa dikelompokkan dalam satu klaster (Wahyudi dkk., 2020). Dalam konteks diabetes mellitus, algoritma ini dapat digunakan untuk mengelompokkan pasien berdasarkan

1

faktor-faktor risiko seperti usia, indeks massa tubuh, dan kadar gula darah. Pengelompokan ini dapat membantu dalam memahami kelompok pasien yang memiliki risiko tinggi terkena diabetes mellitus.

Di sisi lain, classifier biner seperti Support Vector Machine digunakan untuk melakukan klasifikasi dengan membagi data menjadi dua kelas menggunakan hyperplane. SVM bekerja dengan mengubah ruang input asli menjadi ruang fitur yang lebih besar untuk memaksimalkan pemisahan antar kelas (Herwijayanti, 2019). Implementasi SVM dapat menghasilkan prediksi yang lebih akurat mengenai faktor-faktor yang meningkatkan risiko seseorang terkena diabetes mellitus.

Berdasarkan pernyataan di atas, penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering dan SVM guna mengetahui faktor-faktor pemicu diabetes mellitus di Puskesmas Lappae. Dengan mengidentifikasi faktor-faktor ini, diharapkan dapat memberikan wawasan yang lebih mendalam untuk upaya pencegahan dan pengelolaan penyakit diabetes mellitus yang lebih efektif. Penggunaan teknologi analisis data ini juga mendukung pengambilan keputusan berbasis data bagi tenaga kesehatan di Puskesmas Lappae.

Berdasarkan pada latar belakang masalah di atas, maka rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan pasien berdasarkan faktor- faktor risiko yang terkait dengan Diabetes Melitus

2. Bagaimana cara mengimplementasikan metode Support vector machine untuk melakukan klasifikasi risiko Diabetes Melitus serta mengukur tingkat akurasi SVM dalam melakukan klasifikasi risiko Diabetes Melitus di

Puskesmas Lappae.

Berdasarkan pada latar belakang masalah, rumusan masalah di atas, maka tujuan yang ingin dicapai dari penelitian ini yaitu :

1. Mengimplementasikan algoritma K-Means Clustering dalam mengelompokkan pasien dengan risiko Diabetes Melitus tinggi dan rendah di Puskesmas Lappae

2. Mengimplementasikan algoritma SVM dalam mengklasifikasikan pasien dengan risiko Diabetes Melitus tinggi dan rendah di Puskesmas Lappae beserta tingkat akurasinya.

Berdasarkan pada uraian latar belakang , rumusan masalah dan tujuan penelitian di atas, maka manfaat penelitian ini dibagi menjadi bebera aspek sebagai berikut :

Mengimplementasikan penggunaan algoritma Clustering K-Meansdan Support vector machine dalam analisis data medis. Selain itu, memberikan referensitambahan untuk penelitian-penelitian selanjutnya yang berminat mengembangkan metode analisis data khususnya yang berkaitan dengan penyakit diabetes melitus.

1. Bagi Peneliti

2. Peneliti akan mendapatkan pengalaman praktisdalam menerapkan algoritma Clustering K-Means Dan Support vector machine di dunia nyata.

3. Peneliti akan memperoleh keterampilan yang lebih baik dalam preprocessing data, analisis data, dan evaluasi model prediktif serta dapat mengubahmetodologi mereka untuk memenuhi kebutuhan dan masalah yang muncul selama penelitian.

tambahan untuk penelitian-penelitian selanjutnya yang berminat mengembangkan metode analisis data khususnya yang berkaitan dengan penyakit diabetes melitus.

1. Bagi Universitas

2. Universitas dapat membantu meningkatkan keterampilan dan pengetahuan dalam analisis data dan klasifikasi menggunakan algoritma K-Means Clustering Dan Support vector machine (SVM)

3. Memberikan kontribusi pada pengetahuan dan pemahaman tentang analisis data, klasifikasi, yang merupakan aset bagi universitas untuk mendukung pengajaran, penelitian, dan layanan masyarakat.

Dari analisis rumusan masalah di atas dapat dirumuskan beberapa batasan masalah yaitu :

1. Lokasi Penelitian di Puskesmas Lappae.

2. Rekam medis pasien termasuk variabel: usia, jenis kelamin,berat badan,hasil dan status indeks massa tubuh (BMI),

kadar gula darah(gula darah puasa dan gula darah sewaktu),tekanan darah, status tekanan darah.

1. Data pasien yang dikumpulkan dari periode tertentu untuk relevansi dan aktualitas yang terbatas pada data yang tersedia di Puskesmas Lappae

2. K-MeansClustering: Untuk mengelompokkan pasien berdasarkan faktor risiko DM.

3. Support vector machine (SVM): Untuk klasifikasi dan prediksi risiko

DM

1. Analisis Hasil: Mengidentifikasi faktor risiko utama berdasarkan hasil Clustering dan prediksi SVM.

Untuk memberikan gambaran umum dari seluruh penulisan ini, Adapun sistematika penulisan yaitu:

**2. Metode Penelitian**

Penelitian ini dilakukan di PuskesmasLappae Kecamatan Tellulimpoe, Kabupaten Sinjai, Adapun pelaksanaan penelitian ini dilakukan pada bulan mei 2024 sampai semua proses pengumpulan data selesai.

Adapun alat dan bahan yang akan digunakan dalampenelitian ini adalah sebagai berikut :

1. Laptop Asus Tuf

2. HP

3. Excel

4. Visual Studio Code

5. Python (jupyter notebook)

6. Google collab

Perancangan sistem adalah bagian penting dari pembangunan sistem karena menguraikan bagaimana suatu sistem dibangun dari tahap perencanaan hingga tahap pembuatan fungsi-fungsi yang diperlukan untuk mengoperasikannya. Tujuan dari perancangan sistem adalah untuk menentukan apakah sistem yang akan dibangun akanmenghasilkan hasil yang diinginkan.

Mengumpulkan data pasien dari Puskesmas Lappaeyang mencakup berbagai variabel seperti usia, jeniskelamin, berat badan, tinggi badan, indeks massa tubuh(BMI), riwayat keluarga, tingkat gula darah, tekanan darah

1. Data Cleaning dan Feature Selection: Menghapus atau memperbaiki data yang hilang, tidak valid, atau anomali, serta Memilih fitur-fitur yang paling relevan untuk analisis, seperti usia, BMI, kadar gula darah

2. Data Normalization: Menormalkan nilai-nilai dalam dataset agar memiliki skala yang serupa.

3. Implementasi Algoritma K-MeansClustering

Menggunakan algoritma K-Means Clustering untuk mengelompokkan pasien berdasarkan karakteristik yang relevan. Menentukan jumlah cluster (k) yang optimal menggunakan metode seperti Elbow Method atau Silhouette

Score. Menganalisis hasil Clustering untuk mengidentifikasi kelompok- kelompok pasien dengan risiko DM yang berbeda.

1. Penentuan Jumlah Cluster: Menentukan jumlah cluster (k) yang optimal menggunakan metode seperti Elbow Method atau Silhouette Score.

2. Pengelompokan Data: Mengelompokkan pasien ke dalam k cluster berdasarkan kesamaan karakteristik.

3. Evaluasi Clustering: Menggunakan metrik seperti Sum of Squared Errors (SSE) atau Silhouette Coefficient untuk mengevaluasi kualitas pengelompokan.

4. Interpretasi Cluster: Menganalisis setiap cluster untuk memahami ciri-ciri pasien dalam setiap kelompok.

5. Implementasi Algoritma Support vector machine

Melatih model SVM menggunakan data pelatihan untuk membangun model prediktif yang dapat mengklasifikasikan risiko DM. Menggunakan kernel yang sesuai (linear, polynomial, atau RBF) untuk mendapatkan hasil terbaik.

1. Pembagian Data: Memisahkan data menjadi training set dan testing set.

2. Training Model: Melatih model SVM menggunakan trainingset.

3. Hyperparameter Tuning: Menyesuaikan parameter model (seperti C, gamma, dan kernel) untuk meningkatkan kinerja model.

Menganalisis hasil Clustering dan prediksi untuk mengidentifikasi faktor- faktor utama yang memicu Diabetes Melitus. Membandingkan hasil prediksi SVM dengan metode tradisional atau algoritma lain untuk menilaikeakuratan dan efektivitasnya.

Setelah pengujian selesai, kesimpulan dapat ditarik berdasarkan hasil evaluasi. Dengan perancangan sistem ini,diharapkan implementasi algoritma K-Means Clustering danSVM dapat membantu Puskesmas Lappae dalam mengidentifikasi dan mengelola faktor risiko Diabetes Melitus dengan lebih efektif.

Flowchart di atas menggambarkan Proses Clustering melibatkan beberapa langkah utama. Pertama, tentukan jumlahcluster yang diinginkan (k). Setelah itu, tentukan posisi awal dari centroid untuk masing-masing cluster. Selanjutnya,periksa apakah nilai centroid telah berubah. Jika nilai centroid berubah, hitung jarak setiap data dari centroid . Jika tidak, proses Clustering selesai. Langkah berikutnya adalah mengelompokkan data berdasarkan jarak terdekat ke centroid yang baru dihitung. Proses ini diulangi hingga nilai centroid tidak lagi berubah, menandakan bahwa cluster telah stabil

Dalam perancangan sistem atau diagram system yang akan dibuat yaitu sebagai berikut :

Pada diagram di atas dapat dijelaskan bahwa alur proses analisis data yang melibatkan beberapa tahap utama. Dimulai dengan pengumpulandata dari berbagai sumber, data tersebut kemudian diproses melalui tahap preprocessing yang mencakup pembersihan, pengisian nilai yanghilang, dan normalisasi. Selanjutnya, data yang telah diproses diterapkan pada algoritma K-Means Clustering untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Hasil Clustering ini dianalisis untuk memahami pola yang terbentuk dalam setiap klaster. Tahap berikutnya adalah implementasi Support vector machine (SVM), yang menggunakan hasil Clustering untuk membangun model klasifikasi. Setelah model SVM dibangun, dilakukan evaluasi dan validasi untuk memastikan kinerja dan kehandalan model. Proses ini diakhiri dengan menghasilkan output berupa laporan, visualisasi data, atau prediksi yang dihasilkan oleh model, yang kemudian dapat digunakan untuk pengambilan keputusanatau analisis lebih lanjut.

Implementasi K-Means Clustering dimulai dengan menentukan jumlah cluster (k) yang optimal menggunakan metodeseperti Elbow Method atau Silhouette Score. Hasil Clustering dievaluasi dengan mengukur seberapa baik data dikelompokkan menggunakan metrik seperti SSE (Sum of Squared Errors) atau Silhouette Coefficient, dan cluster yang dihasilkan dianalisis untukmemahami karakteristik kelompok pasien berdasarkan faktor risiko yang teridentifikasi. Selanjutnya, algoritma SVM diimplementasikan untuk klasifikasi risiko DM dengan melatih model menggunakan training set, menyesuaikan parameter model untuk mendapatkan kinerja terbaik, dan menguji model pada testing set. Kinerja model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score, serta validasi silang dilakukan untuk memastikan model tidak overfitting dan memilikigeneralisasi yang baik.

Hasil pengujian dan analisis kemudian divisualisasikandalam bentuk grafik dan tabel yang mudah dipahami, seperti diagram sebaran untuk hasil Clustering.Laporan hasil analisis disusun untuk memberikan wawasan kepada tenaga kesehatan di Puskesmas Lappae, mencakup deskripsi hasil Clustering dan interpretasi setiapcluster serta analisis hasil prediksi SVM dan identifikasi faktor risiko utama yang mempengaruhi Diabetes Melitus. Akhirnya, hasil analisis diimplementasikan dalam bentuk dashboard interaktif bagi tenaga kesehatan untuk mengakses hasil analisis danlaporan secara real-time, sehingga membantu dalam pengambilan keputusan klinis. Tahapan pengujian teknik ini diharapkan dapat memberikan hasil yang akurat dan bermanfaat dalam mengidentifikasi faktor pemicu Diabetes Melitus, membantu Puskesmas Lappae dalam pengelolaan dan pencegahan penyakit tersebut secara lebih efektif.

Analisis data merupakan serangkaian metode yang digunakan untuk menganalisis data, dan tujuan dari analisis data adalah untukmengubah data mentah menjadi wawasan atau informasi yangbermanfaat untuk proses pengambilan keputusan bisnis. Wawasanyang ditemukan dalam data terdiri dari fakta, angka, dan metrik. Analisis data kualitatif dan kuantitatif adalah dua metode utama; keduanya dapat digunakan secara terpisah atau bersama-sama. Tujuannya jelas: membantu perusahaan membuat keputusan berbasis data.

(Latifatunnisa, H, 2022) Proses analisis data dilakukanpada penelitian ini adalah sebagai berikut:

Memastikan bahwa model tidak overfittingdan memiliki generalisasi yang baik.

Menggunakan grafik dan tabel untuk memvisualisasikan hasil Clustering dan prediksi SVM.

Menganalisis dan menginterpretasi hasil Clustering dan prediksi SVM untuk mengidentifikasi faktor-faktor pemicu Diabetes Melitus.

Menyusun laporan hasil analisis yang mencakup deskripsi hasil Clustering, interpretasi setiap cluster, analisis hasil prediksi SVM, dan rekomendasi untuk tindakan selanjutnya.

**3. Hasil dan diskusi**

**4. Kesimpulan**

Membahas tentang penarikan kesimpulan dari hasil pembahasan serta saran penelitian.

1. Implementasi Algoritma K-Means dapat digunakan untuk mengelompokkan data pasien berdasarkan karakteristik yang mirip. Proses ini melibatkan pemilihan kolom numerik yang relevan, pembersihan data, dan standarisasi menggunakan StandardScaler. Clustering dilakukan dengan menentukan jumlah kluster yang diinginkan, dalam hal ini sebanyak dua kluster yaitu kelompok beresiko tinggi dengan jumlah data yaitu 584 data pasien deteksi dini beresiko tinggi, dan kelompok beresiko rendah dengan jumlah data yaitu 456 data pasien deteksi dini beresiko rendah. Hasil dari Clustering ini diperkuat dengan penjelasan narasumber

2. Implementasi Algoritma Support vector machine dapat digunakan untuk klasifikasi dan prediksi risiko Diabetes Melitus . Model SVM dilatih menggunakan training set dan diuji pada testing set. Untuk akurasinya Evaluasi kinerja model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score . Dari hasil prediksi dan penginputan hasil prediksi dapat disimpulkan bahwa evaluasi hasil prediksi menunjukkan model SVM memiliki kinerja yang baik dengan akurasi sebesar 93.75%. Dengan jumlah prediksi benar sebanyak 195 dari 208 data yang diujikan, sehingga dapat dipersentasekan (195:208) x 100 % = 93,75 % , nilai ini sesuai dengan nilai hasil prediksi yang telah dibahas sebelumnya pada bagian pembahasan hasil prediksi implementasi algoritma SVM.

Pengujian hasil Clustering dan prediksi SVM dilakukan dengan menggunakan berbagai metrik evaluasi. Hasil analisis divisualisasikan dalam bentuk grafik dan tabel untuk memudahkan pemahaman.

Mendorong penelitian lanjutan untuk mengeksplorasi metode baru dalam analisis dan prediksi risiko diabetes melitus jika di masa mendatang terdapat metode yang lebih efektif dan efisien dalam melakukan klasifikasi dan prediksi pada studi kasus ini.

**Referensi**

[1] American Heart Association. (2016). Understanding blood pressure readings.

[2] Retrieved from /Understanding-Blood-Pressure- Readings\_UCM\_301764\_Article.jsp

[3] Arania, R., Triwahyuni, T., Esfandiari, F., & Nugraha, F. R. (2021). Hubungan antara usia, jenis kelamin, dan tingkat pendidikan dengan kejadian diabetes mellitus di Klinik Mardi Waluyo Lampung Tengah. Jurnal Medika Malahayati, 5(3), 146-153.

[4] Fahrudin, T. M., & Kamisutara, M. (2021). Classification of Toddler Nutritional Status Based on Antrophometric Index and Feature Discrimination using Support vector machine Hyperparameter Tuning. Ijconsist Journals, 2(02), 60-65.

[5] Fauziah, S. N. (2023). Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Klasifikasi Data Pada Diabetes Prediction Dataset (Doctoral Dissertation,Universitas Siliwangi).

[6] Febrianti, A. (2020). Penerapan metode K-Means Clustering dan Support vector machine SVM dalam Identifikasi Api pada Citra Warna digital (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).

[7] Hasanah, S. (2018). Propensity Score Matching Menggunakan Support Vector Machine Pada Kasus Diabetes Melitus (DM) Tipe 2 (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[8] Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2(1), 306-312.

[9] Kemenkes RI. Infodatin (2020) Diabetes Melitus Pusat Data dan Informasi Kementerian Kesehatan RI. Jakarta: Kementerian Kesehatan Republik Indonesia; 2020.

[10] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87-90).

[11] Kusumah, Elsandy Wirahadi, Chandranegara, Didih Rizki, & Nuryasin, Ilyas. (2023). Pengembangan Sistem Informasi Geografis Pemetaan daerah Prioritas Tujuan Vaksin Berbasis Website. Jurnal Impresi Indonesia, 2(3), 236–245

[12] Latifatunnisa, H. (2022). Teknik Analisis Data: Jenis, Contoh, dan Langkah- langkahnya.Retrieved from Revoupedia: https://revou. co/panduan- teknis/teknik-analisis-data.

[13] Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2017). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 2(4), 1704–1708

[14] Riandari, F., Sihotang, H. T., Tarigan, T. and Rafli, M. . (2022) “Classification of Book Types Using the Support vector machine Method”, Jurnal Mantik, 6(1), pp. 43-49. Available at: http://iocscience.org/ejournal/index.php/mantik/article/view/2132 (Accessed: 26May2024).

[15] Wahyudi, M., Masitha, Saragih, R., &Solikhun. (2020). Data Mining:Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering OlehMochamad Wahyudi, Masi.

[16] Wang, Y., Li, D., & Wang, Y. (2019). Realization of remote sensing image segmentation based on K-means clustering. IOP Conference Series: Materials Science and Engineering, 490(1), 072008.

[17] World Health Organization. (2019). Diagnosis and classification of diabetes mellitus (pp. 1–49). Geneva: WHO.

[18] Wulandari, C., Aviani, T. H. B., & Saputra, R. (2024). Penerapan Algortima Support vector machine Untuk Prediksi Tingkat Kelulusan SiswaSMA. Resolusi: Rekayasa Teknik Informatika dan Informasi, 4(4), 397- 407.

[19] Zakir, A. (2022). Implementasi Algoritma K-MeansUntuk Clustering Judul Skripsi Universitas Harapan Medan. Jurnal Media Informatika, 4(1), 40-47.