

MODEL PREDIKSI OBESITAS DENGAN MENGGUNAKAN SUPPORT VECTOR MACHINE

Sherly, Delima Sitanggang
Sistem Informasi, FTIK, Universitas Prima Indonesia
Jalan Sampul, Medan
E-mail : delimasitanggang@unprimdn.ac.id

ABSTRAK- Obesitas atau kelebihan berat badan merupakan kondisi dimana adanya abnormalitas maupun lemak berlebih pada individu yang berperan sebagai salah satu faktor penyakit yang mengancam kesehatan seseorang. Menurut WHO, data dari tahun 1975 hingga 2016 tingkat obesitas pada anak dan remaja dengan umur 5 sampai 19 tahun terus meningkat hingga lebih dari empat kali lipat dari 4% menjadi 18%. Pada masa sekarang, obesitas tidak hanya menjadi masalah pada negara yang memiliki pendapatan perkapita tinggi, negara berkembang dengan pendapatan perkapita rendah menengah juga mengalami peningkatan jumlah obesitas dengan tingkat peningkatan 30% lebih tinggi dari negara maju. Pada penelitian ini, berfokus pada memprediksi tingkat persentase lemak pada badan menggunakan Support Vector Machine. Data target yang akan diprediksi adalah 'BodyFat' dengan mengacuhkan 'Density' karena pengukuran persentase 'BodyFat' diambil dari nilai 'Density'. Model prediksi ini dibangun untuk mempermudah proses penemuan tingkat densitas dari tubuh manusia dikarenakan prosesnya pengambilan datanya yang tidak mudah. Tujuan dari penelitian ini adalah untuk menilai kemampuan Support Vector Machine dalam melakukan regresi serta mempersiapkan algoritma prediksi bertipe regresi dengan nilai performa yang baik. Manfaat dari penelitian ini adalah untuk memperoleh model prediksi data yang dapat membantu memprediksi nilai persentase lemak pada badan sehingga dapat digunakan untuk kelengkapan data serta penyajian informasi tanpa perlu memperhatikan faktor bentuk badan yang beragam. Proses implementasi algoritma SVR dapat dengan baik melakukan regresi dengan tingkat akurasi akhir 71.80% dan MSE 17.76. Sistem prediksi yang dihasilkan dengan algoritma mampu membantu dalam penentuan otomatis persentase lemak pada badan tanpa perlu pengukuran densitas badan yang memerlukan pengukuran dalam air dikarenakan volume tubuh manusia yang beragam dan bervariasi. Persentase lemak pada badan merupakan informasi yang penting baik untuk keperluan diagnosa maupun sebagai informasi peringatan yang dikarenakan apabila persentase berlebih dapat menyebabkan penyakit beresiko tinggi seperti type-2 diabetes dan penyakit jantung lainnya.

Kata kunci : Prediksi, Obesitas, *Support Vector Machine* (SVM)

1. PENDAHULUAN

Obesitas atau kelebihan berat badan merupakan kondisi dimana adanya abnormalitas maupun lemak berlebih pada individu yang berperan sebagai salah satu faktor penyakit yang mengancam kesehatan seseorang [1,2]. Standarisasi yang diakui WHO dalam limitasi pengukuran dinilai dengan satuan BMI (*Body Mass Index*) yang mana angka 25 keatas dikategorikan sebagai kelebihan berat badan dan 30 keatas dinilai sebagai obesitas. Tingkat kematian (mortality rate) dari kematian yang disebabkan kelebihan berat badan atau obesitas dikisar sebesar 4 juta orang per tahunnya seperti yang dikutip dari [3].

Pada penelitian ini, peneliti memutuskan untuk memprediksi tingkat persentase lemak pada badan menggunakan *Support Vector Machine*. Data yang dipakai untuk melatih algoritma *Support Vector Machine* yang dipakai diambil dari data cek medis yang diperoleh dari Dr. A. Garth [4]. (fedesoriano. (Juni 2021). *Body Fat Prediction Dataset*).

Dataset yang dipakai berasal dari repositori *Kaggle*, yaitu dataset "Body Fat Prediction Dataset". Data target yang akan diprediksi adalah 'BodyFat' dengan mengacuhkan 'Density'

karena pengukuran persentase 'BodyFat' diambil dari nilai 'Density'. Model prediksi ini dibangun untuk mempermudah proses penemuan tingkat densitas dari tubuh manusia dikarenakan prosesnya pengambilan datanya yang tidak mudah. Pemilihan algoritma *Support Vector Machine* dikarenakan algoritma tersebut adalah salah satu algoritma *Machine Learning* yang cukup dinamis dalam implementasinya selain dapat dipakai untuk klasifikasi *Support Vector Machine* juga memungkinkan prediksi terhadap model data regresi (SVM Regression) [5,6]. Kualitas algoritma yang dipilih diharapkan mampu menghasilkan hasil prediksi yang aktual, akurat dan dapat diandalkan.

Hasil penelitian menyimpulkan bahwa algoritma mampu menyajikan jalur alternative jika jalur pengukuran yang mahal tidak tersedia. Performa algoritma diukur dengan metode MAE dan RMSE dimana nilai MAE, SD, RMSE, MAC yang diperoleh adalah 4.3855, 5.7874, 5.7889, 0.9470 [4].

Penelitian berikutnya dengan judul "Using an improved relative error support vector machine for body fat prediction" [7]. Secara ringkas peneliti tersebut menerapkan feature selection ditambahkan dari pendekatan baru terhadap algoritma yang

diterapkan yang disimpulkan dapat meningkatkan performa algoritma tersebut. Algoritma yang digunakan masih sama dengan yang sebelumnya namun ditambahkan bias error baru sehingga menjadi IRE-SVM. Nilai akhir secara runtun 4 diperoleh 4.4610, 5.6322, 5.6488, 0.9508.

Penelitian SVM terakhir yang menjadibahan acuan peneliti adalah “Modelling the Psychological Impact of COVID-19 in Saudi Arabia Using Machine Learning” [8]. Penelitian tersebut mengamati mengenai faktor perilaku, sikap dan psikologi masyarakat selama masa pandemi di Saudi Arabia. Questionnaire dengan 320 responden dijadikan sebagai data inputnya. Hasil penelitian menunjukkan SVMR mendapatkan performa baik dengan penilaian R2 values secara urut: 99.59%, 99.93%, dan 99.88%. Pendapat disimpulkan menjadi salah satu faktor penting yang mempengaruhi kondisi selama masa pandemi.

2. ISI PENELITIAN

2.1 Sub Bab

Adapun jenis penelitian yang dilakukan dalam penelitian ini yaitu penelitian eksperimental kuantitatif. Penelitian eksperimental adalah penelitian yang dilakukan untuk mengetahui akibat yang ditimbulkan dari suatu kondisi yang disusun secara sengaja oleh peneliti. Penelitian kuantitatif adalah suatu proses menemukan pengetahuan yang menggunakan data berupa angka sebagai alat menganalisis keterangan mengenai apa yang ingin diketahui. Eksperimen dilakukan dengan menggunakan algoritma SVM untuk melakukan prediksi regresi terhadap data persentase lemak badan.

2.2 Tempat dan Waktu

Tempat pelaksanaan penelitian berada di Universitas Prima Indonesia (UNPRI) yang berlokasi di Jl. Sekip Jl. Sikambing No. simpang, Sei Putih Tim. I, Kec. Medan Petisah, Kota Medan, Sumatera Utara 20111. Waktu penelitian dilaksanakan terhitung mulai dari tanggal 28 Desember 2021 hingga selesai.

2.3 Prosedur Kerja

Metode yang digunakan untuk menyusun prosedur kerja laporan akhir ini adalah metode waterfall ditambah dengan visualisasi diagram alur penelitian. Tahapan-tahapan yang diperlukan adalah:

1. Pengumpulan Data

Pada tahapan ini, metode pengumpulan data yang digunakan mencakup:

- Metode Analisa Data Kuantitatif, dimana dataset yang akan terlebih dahulu dilakukan analisa untuk menentukan target klasifikasi terhadapnya demi keperluan penyajian data hasil analisa.

- Metode Studi Literatur, dimana akan dikumpulkan informasi dan bahan yang diperlukan dalam penelitian ini, yang diperoleh dari berbagai sumber baik dari jurnal ilmiah maupun artikel.

2. Preprocessing

Dataset terdapat 15 kolom dengan keseluruhan 251 entri data dimana semua datanya merupakan hasil pengukuran pada tiap bagian tubuh subjek. Semua data entri bersih dan tidak terdeteksi adanya data hilang. Data kemudian dibagi menjadi data dependen dan independent dengan pengecualian kolom ‘Density’ karena merupakan kalkulasi terhadap ‘BodyFat’ yang merupakan data independent /target.

3. Analisa

Setelah preprocessing data, analisa dapat dimulai dengan melakukan penjabaran terhadap fitur data pada database. Fitur pada database terdiri dari:

- Age: umur pasien [tahun]
- Weight [lbs]
- Height [cm]
- Neck circumference: Lingkar Leher [cm]
- Chest circumference: Linkar Dada [cm]
- Abdomen 2 circumference [cm]
- Hip circumference [cm]
- Thigh circumference [cm]
- Knee circumference [cm]
- Ankle circumference [cm]
- Biceps (extended) circumference [cm]
- Forearm circumference [cm]
- Wrist circumference [cm]
- Percent body fat: persentase lemak badan (Target Regresi)

4. Implementasi

Data dependen dan independent yang telah melewati proses preprocessing kemudian dilakukan split data untuk data latih dan data uji dengan perbandingan 0.33.

5. Pengujian

Pada penelitian ini tipe SVM yang diimplementasi adalah SVR karena dataset bertipe regresi. Parameter yang digunakan dalam SVR yang digunakan adalah:

- 'kernel': “linear”,
- 'C': 100,
- 'gamma': “auto”,

Diagram alur penelitian dapat dilihat pada Gambar 2.1 dibawah. Diagram alur menggambarkan langkah proses penelitian yang dilakukan pada penelitian ini.

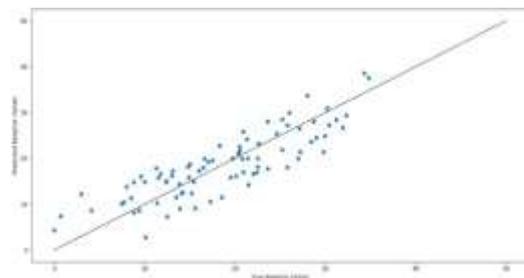


Gambar 2.1. Diagram Alur Penelitian

3. HASIL DAN PEMBAHASAN

Normalisasi dataset menghasilkan 251 entri data dengan 13 fitur dan satu target data 'BodyFat'. Imputasi tidak diperlukan dikarenakan tidak adanya data hilang dalam dataset. Data dependen pada penelitian ini merupakan regresi yang pada artinya data dependen bersifat numerik dengan hasil bertingkat. Algoritma SVM dapat melakukan regresi dalam tipe SVR (Support Vector Machine for regression) sehingga diimplementasikan terhadap dataset dan diukur performanya dengan pengukuran akurasi R-squared (R^2) dan perbandingan grafik prediksi. SVR berbeda dari model regresi yang sederhana, dimana pada regresi sederhana diupayakan untuk meminimalisir error rate sedangkan SVR diupayakan untuk memuat error pada batas tertentu. Singkatnya SVR mempunyai dua hyperplane atau garis bayang diantara garis prediksi yang berjarak tertentu dan menentukan data diantara kedua garis tersebut dan garis tengah prediksi tersebut termasuk kedalam hasil prediksi [9].

Pembagian dataset menjadi data latih dan data uji menghasilkan 168 entri data latih dan 84 entri data uji. Pengukuran regresi pada umumnya juga memanfaatkan Mean Squared Error (MSE). MSE menghitung jumlah error pada model regresi dengan menggunakan rata-rata hasil kuadrat perbedaan antara data asli dengan data 8 prediksi [10]. MSE pada penelitian ini mendapatkan hasil 17.76 dimana merupakan hasil yang cukup baik dimana model akan semakin sempurna jika angka hasil mendekati nol.



Gambar 2.2 Hasil Regresi SVR Terhadap Dataset

Akurasi R^2 score algoritma diatas mendapatkan hasil sebesar 71.80%. Gambar 2.2 menunjukkan grafik persebaran antara data asli dan data prediksi. Semakin setiap titik mendekati garis diagonal maka semakin akurat hasil prediksi. Untuk SVR walau beberapa titik data tidak menyentuh garis diagonal utama tetap akan dicantumkan kedalam hasil selama tidak melewati batas garis hyperplane. Berdasarkan data grafik yang diperoleh, dapat disimpulkan bahwa model yang dipakai dapat dengan tepat mengikuti garis lurus pada diagram perbandingan data prediksi dengan data yang sebenarnya. Grafik dibuat sedemikian untuk memvisualisasi kestabilan maupun ketimpangan kemampuan model dalam melakukan prediksi.

Beberapa data menunjukkan ketimpangan ekstreme namun kelompok data tersebut sangat minim jumlahnya. Model yang dibangun juga berhasil dibangun dengan rata-rata waktu latih sebesar 1 menit yang tergolong unggul dibandingkan model serupa lainnya. Hasil latih juga berhasil mendapatkan hasil MSE optimal dengan angka 17.76. Optimisasi secara penentuan kelompok data untuk memastikan keseimbangan antara data latih dan data uji dilakukan dengan pengacakan data aktif, sehingga pengelompokan data dilakukan secara acak untuk mencegah kebiasaan terhadap data yang dipilih. Teknik serupa yang dilakukan pada dataset yang sama mendapatkan hasil tertinggi 69.5% dengan metode lasso regression dengan rata-rata hasil algoritma berada pada kisaran 60-70%. SVR yang digunakan dapat menghasilkan akurasi dari penilai R^2 sebesar 71.80% dinilai telah dapat mengungguli hasil algoritma yang diterapkan pada dataset sebelumnya. Hasil analisa diharapkan dapat digunakan sebagai alternatif dalam pengukuran persentase lemak pada badan yang terotomatisasi, dapat diandalkan dan akurat.

4. PENUTUP

Kesimpulan

1. Proses implementasi algoritma SVR dapat dengan baik melakukan regresi dengan tingkat akurasi akhir 71.80% dan MSE 17.76.

2. Sistem prediksi yang dihasilkan dengan algoritma mampu membantu dalam penentuan otomatis persentase lemak pada badan tanpa perlu pengukuran densitas badan yang memerlukan pengukuran dalam air dikarenakan volume tubuh manusia yang beragam dan bervariasi.

3. Persentase lemak pada badan merupakan informasi yang penting baik untuk keperluan diagnosa maupun sebagai informasi peringatan yang dikarenakan apabila persentase berlebih dapat menyebabkan penyakit beresiko tinggi seperti type-2 diabetes dan penyakit jantung lainnya.

SARAN

1. Pada penelitian ini tidak dilakukan optimasi parameter secara otomatis melalui teknik cross validation dan grid search dikarenakan hasil akhir telah dinilai mencukupi dibandingkan dengan penelitian sebelumnya terhadap dataset yang sama.

2. Penelitian berikutnya juga dapat mencoba mencari cara baru untuk memvisualisasi data dan menyampaikan informasi yang didapat dengan lebih baik dan dilakukan dengan hasil observasi baru terutama pada visualisasi algoritma dan hasil regresi agar informasi atau hasil yang didapatkan dapat terlihat semakin jelas.

DAFTAR PUSTAKA

- [1] OECD and World Health Organization, "Health at a glance: Asia/pacific 2020," Health at a Glance: Asia/Pacific, 2020.
- [2] S.-Z. Jiang, W. Lu, X.-F. Zong, H.-Y. Ruan, and Y. Liu, "Obesity and hypertension," *Experimental and Therapeutic Medicine*, vol. 12, no. 4, pp. 2395–2399, 2016.
- [3] S. L. James, Et. All, "Estimating global injuries morbidity and mortality: Methods and data used in the global burden of disease 2017 study," *Injury Prevention*, vol. 26, no. Supp 1, pp. i125–i153, 2020.
- [4] Fedesoriano. (Juni 2021). Body Fat PredictionDataset. Retrieved [29 December] from,<https://www.kaggle.com/fedesoriano/body-fat-prediction-dataset>).
- [5] D. A. Pisner and D. M. Schnyer, "Support Vector Machine," *Machine Learning*, pp. 101–121, 2020.
- [6] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support Vector Machine Classification: Applications, challenges and Trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [7] R. Chiong, Z. Fan, Z. Hu, and F. Chiong, "Using an improved relative error support vector machine for body fat prediction," *Comput. Methods Programs Biomed.*, vol. 198, p.105749, 2021.
- [8] M. A. Aleid, K. A. Z. Alyamani, M. Rahmouni, T. H. H. Aldhyani, N. Alsharif, and M. Y. Alzahrani, "Modelling the Psychological Impact of COVID-19 in Saudi Arabia Using Machine Learning," *Comput. Mater. Contin.*, vol. 67, no. 2, 2021, doi: 10.32604/cmc.2021.014873.
- [9] D. Parbat and M. Chakraborty, "A python based support vector regression model for prediction of COVID19 cases in India," *Chaos, Solitons & Fractals*, vol. 138, p. 109942, 2020.
- [10] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than smape, Mae, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, 2021.