

BAB 1

PENDAHULUAN

1.1. Latar Belakang

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. 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 *The Global Burden of Disease* pada tahun 2017. Obesitas merupakan masalah kesehatan yang telah ada dari dulu namun masih terus dihadapi hingga sekarang. 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, 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 [1]. (fedesoriano. (Juni 2021). Body Fat Prediction Dataset. Retrieved [Date Retrieved] from <https://www.kaggle.com/fedesoriano/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. Berat badan atau massa badan dapat dengan mudah diperoleh namun untuk memperoleh volume tidak semudah itu karena tubuh manusia beragam bentuknya. Tingkat persentase lemak pada badan yang diperoleh nantinya ini berguna sebagai

informasi penting dikarenakan persentase yang berlebih dapat beresiko terkena berbagai penyakit seperti type-2 diabetes dan penyakit jantung.

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*). Kualitas algoritma yang dipilih diharapkan mampu menghasilkan hasil prediksi yang aktual, akurat dan dapat diandalkan. Penelitian ini juga diharapkan nantinya juga dapat digunakan untuk memperoleh data prediksi untuk membantu perolehan data untuk database kondisi kesehatan lainnya seperti database penyakit jantung dan lainnya.

1.2. Rumusan Masalah

Adapun rumusan masalah yaitu: Bagaimana implementasi algoritma Support Vector Machine dalam membantu pengukuran persentase lemak pada badan dengan melakukan prediksi *Machine Learning*?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Adapun tujuan dari penelitian ini adalah untuk menilai kemampuan Support Vector Machine dalam melakukan regresi serta mempersiapkan algoritma prediksi bertipe regresi dengan nilai peforma yang baik.

1.3.2. Manfaat

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.

1.4. Batasan Masalah

Dalam penelitian ini, terdapat batasan masalah dari permasalahan yang ada, yaitu :

1. Dataset terdiri dari 251 observasi dengan 15 atribut. Tidak terdapat data hilang pada database (bersih). Densitas tubuh pada dataset diukur dengan pengukuran dibawah air untuk menangkap dengan presisi bentuk tubuh subjek.
2. Algoritma yang digunakan adalah Support Vector Machine.
3. Bahasa pemrograman yang digunakan adalah bahasa Python dengan memanfaatkan beberapa *library* yang telah ada, yaitu: *pandas*, *numpy*, *seaborn*, *matplotlib*, dan *sklearn*.
4. Hasil penelitian difokuskan dengan ditentukannya angka persentase lemak pada badan sebagai data dependen (target) klasifikasi untuk menunjukkan prediksi data yang nantinya dapat digunakan sebagai informasi medis.

1.5. Keterbaruan

Penelitian sebelumnya yang memprediksi persentase lemak pada badan juga pernah dilakukan oleh penelitian dengan judul “A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat *prediction*”. Tujuan penelitian tersebut adalah untuk menginvestigasi penggunaan model *machine learning* untuk dapat dengan akurat memprediksi presentase lemak badan. Algoritma yang disarankan oleh penelitian tersebut adalah *fuzzy-weighted Gaussian kernel-based Relative Error Support Vector Machine* atau yang disingkat sebagai (RE-SVM). Design dari algoritma yang disarankan dirancang dengan operasi *fuzzy-weighted* yang mengimplementasi *fuzzy weights* pada konstrain eror dari RE-SVM tersebut untuk menangani noise pada data. Fuzzy weights juga digunakan untuk meningkatkan Gaussian-kernel mengingat pentingnya sampel yang berbeda. Hasil penelitian menyimpulkan bahwa algoritma mampu menyajikan jalur alternative jika jalur pengukuran yang mahal tidak tersedia. Peforma algoritma diukur dengan metode MAE dan RMSE dimana nilai MAE, SD, RMSE, MAC yang diperoleh adalah 4.3855, 5.7874, 5.7889, 0.9470 [1].

Penelitian berikutnya dengan judul “Using an improved relative error support vector machine for body fat prediction”. Secara ringkas peneliti tersebut menerapkan *feature selection* ditambahkan dari pendekatan baru terhadap algoritma yang diterapkan yang disimpulkan dapat meningkatkan peforma algoritma tersebut. Algoritma yang digunakan masih sama dengan yang sebelumnya namun ditambahkan bias eror baru sehingga menjadi IRE-SVM. Nilai peforma akhir secara runtun

diperoleh 4.4610, 5.6322, 5.6488, 0.9508 [2]. Dari dua penelitian diatas dapat disimpulkan penerapan fuzzy meningkatkan sedikit peforma algoritma prediksi data.

Penelitian SVM terakhir yang menjadi bahan acuan peneliti adalah “Modelling the Psychological Impact of COVID-19 in Saudi Arabia Using Machine Learning”. 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 peforma baik dengan penilaian R^2 values secara urut: 99.59%, 99.93%, dan 99.88%. Pendapatan disimpulkan menjadi salah satu faktor penting yang mempengaruhi kondisi selama masa pandemi [3].

Peneliti dalam penelitian ini, menargetkan prediksi terhadap persentase lemak pada badan dengan membangun model algoritma SVM-R yang dibangun dengan database yang stabil dan bersih. Visualisasi juga menjadi salah satu yang belum dilakukan secara maksimal oleh penelitian sebelumnya. Penelitian ini nantinya akan menerapkan tingkatan akurasi sebagai parameter penilaian akhir. Penelitian ini berfokuskan pada pengembangan dan penyederhanaan model yang tidak dilakukan ataupun terlewatkan oleh penelitian yang sebelumnya.