

1. Pendahuluan

1.1 Latar Belakang

Diabetes merupakan salah satu penyakit yang disebabkan karena gula darah di dalam tubuh yang tinggi atau melampaui batas normal. Berdasarkan data laporan Kementerian Kesehatan RI tahun 2018 melalui *conference* Suara Dunia Perangi Diabetes, Indonesia merupakan negara peringkat keenam sebagai penderita diabetes terbanyak di dunia dengan jumlah penderita diabetes usia 20-79 tahun sekitar 10,3 juta [1]. Riset Kesehatan Dasar juga menyebutkan penderita diabetes di Indonesia mengalami peningkatan, yaitu semula dari tahun 2013 sebesar 6,9% menjadi 8,5% di tahun 2018 dengan perkiraan jumlah penderita lebih dari 16 juta orang [1]. Penyakit diabetes dapat mengakibatkan komplikasi penyakit, yang tentunya sangat berbahaya terhadap penderita diabetes. Oleh karena itu, sangat diperlukannya suatu teknologi yang dapat mendeteksi penyakit diabetes dengan tingkat analisis yang akurat, sehingga penyakit diabetes dapat ditangani lebih awal untuk mengurangi jumlah penderita, kecacatan, dan kematian.

Beberapa tahun terakhir, penelitian terhadap penyakit diabetes sudah dilakukan dengan menggunakan berbagai macam metode klasifikasi untuk mendeteksi diabetes. Berikut beberapa penelitian yang terkait dengan pengujian data diabetes. Manimaran dan Vanita [2] mengusulkan metode *Decision Tree* dalam melakukan klasifikasi penyakit diabetes. Peneliti juga melakukan *preprocessing* data dan transformasi data seperti mengganti nilai yang hilang dan menormalisasikan data untuk meningkatkan hasil dan efisiensi dalam penambangan data. Pada penelitiannya, peneliti menggunakan *cross validation* untuk membagi data ke dalam dua bagian data latih dan data uji dengan perbandingan 70:30. Untuk mengevaluasi model yang sudah dibangun peneliti menggunakan *confusion matrix* dengan hasil *accuracy* yang diperoleh dari data asli tanpa *cross validation* sebesar 83,5937% dan setelah menggunakan *cross validation* hasil *accuracy* klasifikasi didapatkan sebesar 85,0163%. Selanjutnya pada tahun 2020 Diniyal Amru Agatsa [3], membangun model klasifikasi pasien pengidap diabetes menggunakan metode *Support Vector Machine* pada data diabetes dan validasi model menggunakan *K-Fold Cross Validation* untuk membagi data menjadi k bagian dengan hasil akurasi yang diperoleh sebesar 77,92%. Selanjutnya Indrayanti tahun 2017 [4], peneliti menggunakan KNN sebagai metode klasifikasi untuk mengklasifikasi penyakit diabetes melitus, dengan hasil akurasi yang diperoleh sebesar 75,14% dengan nilai $k=13$ merupakan nilai k yang paling optimal. Selanjutnya Januar Adi Putra tahun 2016 [5], peneliti melakukan klasifikasi penyakit diabetes menggunakan metode penggabungan SVM dengan KNN, dengan hasil akurasi yang diperoleh sebesar 92%. Selanjutnya pada tahun 2014 Ankita Parashar [6], mendeteksi penyakit diabetes dengan LDA-*Support Vector Machine* dan *Feed Forward Neural Network* sebagai metode klasifikasi dan LDA sebagai reduksi dimensi, dengan hasil akurasi LDA-SVM yang diperoleh sebesar 75,65%. Selanjutnya pada tahun 2012 Chamidah [21] mengusulkan pengujian terkait Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain untuk klasifikasi data kanker payudara. Peneliti melakukan analisis terhadap metode normalisasi mana yang efektif untuk kasus klasifikasi menggunakan algoritma pelatihan BPGD/AG. Pada penelitiannya, peneliti melakukan evaluasi dari eksperimen dengan mengambil rata-rata epoch dan akurasi dari seratus pengujian. Hasil yang diperoleh peneliti adalah decimal scaling 40 untuk epoch dan 96 % untuk akurasi, sigmoid 584 untuk epoch dan 76,91 untuk akurasi, softmax 49 untuk epoch dan 96,19 % untuk akurasi, min-max 21 untuk epoch dan 96,86 % untuk akurasi, statistical column 45 untuk epoch dan 95,91 untuk akurasi, dan z-score 38 untuk epoch dan 95,68 untuk akurasi. Dari hasil yang diperoleh bahwa min-max memiliki rata-rata epoch paling rendah yang membuktikan bahwa metode min-max memiliki waktu yang paling singkat untuk mencapai konvergensi dan hasil akurasi yang paling tinggi yang membuktikan bahwa min-max normalisasi dapat meningkatkan hasil akurasi dibandingkan dengan metode lainnya.

Dari beberapa penelitian yang sudah dilakukan, penelitian kali ini membangun suatu sistem klasifikasi untuk menganalisis dan memprediksi apakah seseorang sebagai penderita diabetes atau tidak berdasarkan dataset Gula Karya Medika. Pada tahap *preprocessing* dilakukan transformasi data dengan menghapus nilai yang hilang dan normalisasi data dengan tujuan membuat nilai setiap atribut berada pada rentang yang sama, dengan harapan dapat meningkatkan hasil dan efisiensi dari klasifikasi [2]. Selain itu, peneliti juga menggunakan *cross validation* untuk memisahkan *data training* dengan *data testing* yang diharapkan dapat meningkatkan hasil akurasi dari model. Penelitian ini mengusulkan tiga buah model, dua model dengan normalisasi data yaitu: *min-max normalization*, *z-score normalization*, dan satu model tanpa normalisasi data. Pada tahap klasifikasi dibandingkan hasil akurasi yang diperoleh *random forest* dari *min-max normalization*, *z-score normalization*, dan satu model tanpa normalisasi data. Pada tahap klasifikasi dibandingkan hasil akurasi yang diperoleh *Random Forest* dari *Min-max normalization*, *Z-score normalization*, dan tanpa normalisasi data untuk mengetahui metode normalisasi data mana yang lebih optimal dan akurat dalam meningkatkan performansi klasifikasi penyakit diabetes. Dalam penelitian ini menggunakan beberapa skenario pengujian seperti jumlah tree pada *random forest*, jumlah K pada *cross validation*, model normalisasi data yang digunakan dengan menghitung akurasi, *precision*, dan *recall*.

1.2 Topik dan Batasannya

Topik yang dibahas dan dianalisis pada penelitian ini adalah memprediksi apakah seseorang sebagai penderita diabetes atau tidak berdasarkan *dataset* Gula Karya Medika, agar dikemudian hari dapat dimanfaatkan untuk membantu bidang Kesehatan dalam mendeteksi penyakit diabetes dengan tingkat analisis yang akurat sehingga penyakit diabetes dapat ditangani lebih awal. Batasan pada penelitian ini adalah data yang digunakan untuk analisis merupakan dataset dari Gula Karya Medika, tidak membandingkan algoritma klasifikasi *Random Forest* dengan algoritma klasifikasi lainnya dalam penelitian, dan data yang digunakan dibagi menjadi dua, yaitu *data training* dan *data testing*.

1.3 Tujuan

Tujuan dari penelitian ini yaitu menganalisis deteksi penyakit diabetes dapat dilakukan dengan sistem klasifikasi menggunakan random forest dan menganalisis metode normalisasi data (*min-max normalization* dan *z-score normalization*) mampu meningkatkan performansi hasil klasifikasi *Random Forest*.

1.4 Organisasi Tulisan

Bab dua membahas studi terkait yang berisi teori pendukung yang berkaitan dalam penelitian. Bab tiga membahas rancangan sistem yang dibangun, yang berisi rancangan dari awal penelitian hingga akhir penelitian dan metode-metode yang digunakan. Bab empat membahas evaluasi yang berisi hasil analisis dan pembahasan dari penelitian. Bab lima menjelaskan kesimpulan dari hasil penelitian yang diperoleh dan saran untuk penelitian berikutnya.