

## 1. Pendahuluan

### Latar Belakang

Diabetes merupakan penyakit kronis yang terjadi karena gangguan metabolik dalam tubuh yang ditandai dengan kadar gula berlebihan. Riset Kesehatan Dasar (Riskesdas) meneliti bahwa kasus diabetes di Indonesia terus meningkat dari 6,9% pada tahun 2013 menjadi 8,5% pada tahun 2018 [1]. Selain itu, Riskesdas menyatakan bahwa pada tahun 2013 dan 2018 pengidap diabetes pada perempuan lebih tinggi daripada laki-laki [1]. Bertambahnya pengidap diabetes dikhawatirkan meningkat tiap tahunnya, selain itu penyakit ini berbahaya dalam tubuh yang menyebabkan banyak komplikasi hingga kematian. Maka dari itu, dengan adanya perkembangan teknologi seperti *machine learning* sangat diperlukan untuk mendiagnosis penyakit diabetes secara akurat.

Untuk mendiagnosis penyakit diabetes, beberapa peneliti melakukan penelitian dengan menggunakan bermacam-macam metode algoritma klasifikasi dan *feature selection* untuk mengoptimalkan hasil akurasi. Penelitian Ghosh dkk. Menggunakan metode *Minimal Redundancy Maximal Relevance* (MRMR) serta beberapa algoritma klasifikasi seperti *Gradient Boosting*, *Support Vector Machine*, *AdaBoost*, dan *Random Forest* [2]. Hasilnya algoritma *Random Forest* dengan MRMR dan penggunaan semua fitur mendapat akurasi terbaik dan skor sama yaitu 99,35%. Penelitian Zou dkk. meneliti dengan metode *Principal Component Analysis* (PCA) dan MRMR dan pemilihan fitur lainnya serta menggunakan beberapa algoritma seperti *Neural Network*, *J48 Decision Tree*, *Random Forest* [3]. Hasilnya akurasi dari dataset Luzhou 0,8084% dan *Pima Indian Diabetes* 0,7721%, performansi terbaik dari dua dataset tersebut dengan menggunakan semua fitur dan algoritma *Random Forest*. Penelitian Qiao dkk. Meneliti efek pengobatan diabetes dengan metode *Decision Tree* CART (*Classification And Regression Tree*) dan PCA, performansi CART dengan PCA meningkat menjadi 73% [4].

Berdasarkan Penelitian terkait sebelumnya, penggunaan *Feature selection* digunakan untuk mengurangi dimensi dan meminimalisasi penggunaan atribut atau fitur pada data yang digunakan. Dalam penelitian ini membangun suatu sistem klasifikasi dengan metode pemilihan fitur MRMR guna untuk mengurangi dimensi disertai algoritma *Random Forest* dan *Decision Tree* CART yang ditujukan pada penderita penyakit diabetes, dataset yang digunakan dari Gula Karya Medika yang terdapat pada penelitian sebelumnya sebagai bahan penelitian [5]. Penelitian ini bertujuan untuk mengetahui performa algoritma *Random Forest* dan *Decision Tree* CART dengan MRMR.

### Topik dan Batasannya

Topik yang dibahas pada penelitian ini terkait perbandingan hasil prediksi dengan suatu metode yang dapat membantu diagnosis penyakit diabetes seseorang berdasarkan data Gula Karya Medika. Penelitian dilakukan agar dapat berkontribusi di bidang Kesehatan dalam mendiagnosis penyakit diabetes lebih awal. Batasan dari penelitian ini adalah data yang digunakan dari dataset Gula Karya Medika berjumlah 470 baris dan 6 atribut (*Glucose*, *Gender*, *Blood Pressure*, *BMI*, *Class*).

### Tujuan

Tujuan dari penelitian ini yaitu menganalisis hasil prediksi penyakit diabetes dengan mengimplementasikan metode klasifikasi *Random Forest* dan *Decision Tree* CART, serta metode *Feature Selection* MRMR. Beberapa metode tersebut digunakan untuk mengoptimalkan kinerja performansi hasil klasifikasi.

### Organisasi Tulisan

Bagian bab dua membahas studi terkait yang berisi penjelasan teori pendukung yang berkaitan dengan penelitian. Bab tiga berisi tentang rancangan sistem yang dibangun, rancangan yang menjelaskan proses arah penelitian dari awal sampai akhir. Bab empat berisi tentang evaluasi hasil dan analisis dari penelitian yang telah dilakukan. Bab lima berisi tentang kesimpulan dari hasil penelitian.