

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Diabetes merupakan penyakit serius yang terjadi ketika pankreas tidak menghasilkan cukup Insulin sebagai hormon yang mengatur gula darah, atau glukosa, atau ketika tubuh tidak dapat menggunakan insulin secara efektif, serta dapat berdampak pada kualitas kehidupan manusia. (Alam,2019), Penyakit diabetes dialami oleh beberapa masyarakat di dunia, secara statistik global pada tahun 2013 mengungkapkan sekitar 382 juta orang menderita penyakit diabetes di dunia (WHO,2016). Pada riset kesehatan dasar pada tahun 2018 menunjukan prevalensi diabetes melitus di Indonesia terdapat peningkatan sebesar 8.5% hal ini dapat diperkirakan dari karena gaya hidup masyarakat (Kemenkes,2020). Diabetes serta komplikasinya dapat membawa kerugian ekonomi yang besar bagi penderita, dalam diagnosa penyakit diabetes seseorang perlu melakukan serangkaian test terhadap tubuh orang tersebut kemudian menunggu hasil test terkait untuk dilakukan keputusan terhadap apakah mengidap diabetes atau tidak serta jenis diabetes yang dimilikinya, pengobatan dalam jadwal berkala merupakan hal penting bagi pengidap diabetes untuk mengatur pola hidup selama masih memiliki penyakit diabetes dengan diat sampai dengan konsumsi obat. (Alam,2019). Berdasarkan permasalahan tersebut dibutuhkan sebuah sistem untuk memprediksi atau diagnosa penyakit diabetes lebih awal sehingga dapat teratasi lebih awal serta perawatan selama menderita penyakit diabetes, menurut (McKinsey,2016) dengan menggunakan teknologi kesehatan dapat membantu dokter mendiagnosa serta membantu dalam perawatan penyakit yang dialami seperti penjadwalan perawatan.

Beberapa peneliti melakukan penelitian terkait bagaimana peran teknologi dalam penanganan penyakit diabetes oleh (Andrea,2019) melakukan penelitian terkait dampak dari aplikasi perawatan diabetes berbasis selular dalam membantu pengobatan, hasil penelitian ini menunjukan bahwa aplikasi seluler efektif untuk pasien diabetes membantu dalam konsultasi dan kontrol terkait penyakit diabetes,selanjutnya oleh (Thongyoo,dkk,2020) melakukan penelitian terkait pengembangan sistem *Chatbot* rekomendasi makanan untuk penderita diabete berbasis pada aplikasi obrolan atau *Messaging* yaitu Line terdapat rekomendasi

makanan, koleksi data, informasi nutrisi, dan saran pengujian menunjukkan dapat membantu penderita diabetes dalam memberikan saran dengan mudah dan cepat, diusulkan sistem diagnosa dan prediksi penyakit diabetes menggunakan *Machine Learning* atau pembelajaran mesin oleh beberapa peneliti di dunia seperti oleh (Raihan.dkk,2020) melakukan penelitian untuk klasifikasi penyakit diabetes menggunakan algoritma Random Forest dan dataset dengan 340 data. Hasil pengujian algoritma Random Forest menghasilkan akurasi sebesar 75.30%, selanjutnya penelitian terkait deteksi awal penyakit diabetes menggunakan seleksi fitur optimal oleh (Sneha,dkk.2019) menggunakan algoritma Naive Bayes dan algoritma Random Forest. Penelitian ini menggunakan dataset terdiri atas 2500 data, Hasil pengujian menunjukan algoritma Random Forest menghasilkan akurasi sebesar 75.39% dan algoritma klasifikasi Naive Bayes menghasilkan akurasi sebesar 73.48%. Penelitian sebelumnya sistem klasifikasi diterapkan secara *Model-Based* untuk klasifikasi penyakit diabetes.

Penelitian ini membangun aplikasi bernama diabetis berbasis Android dan menganalisa terkait performa serta akurasi metode *Term Frequency Inverse Document Frequency* dengan algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetis menerapkan komputasi secara *on device machine learning* sehingga perhitungan dalam lingkup aplikasi tanpa interaksi koneksi internet penerapan ini memiliki keunggulan dapat digunakan secara *offline* atau tanpa koneksi internet. Metode *Term Frequency Inverse Document Frequency* dalam fitur *chat* untuk pengkategorian kalimat aplikasi kemudian algoritma Naive Bayes dan algoritma Random Forest sebagai klasifikasi penyakit diabetes berdasarkan hasil ekstraksi fitur kalimat masukan pengguna sebelumnya dengan riwayat kesehatan yang diberikan, terdapat *preprocessing* untuk pembersihan kalimat selanjutnya proses pengkategorian dengan metode *Term Frequency Inverse Document Frequency* apakah kalimat ini termasuk klasifikasi atau sebuah perintah, jika merupakan kalimat untuk klasifikasi diabetes menggunakan algoritma Naive Bayes dan algoritma Random Forest serta terdapat perintah informasi diabetes serta mengatur kegiatan perawatan diabetes. Metode *Term Frequency Inverse Document Frequency* digunakan karena merupakan teknik untuk kuantitatif dalam menghitung kumpulan kata pada sebuah dokumen dengan mencari kata input yang relevan dengan dataset sebelumnya, (Kadhim.dkk,2019) melakukan perbandingan antara metode pembobotan yaitu BM25 dan *Term*

*Frequency Inverse Document Frequency* pengujian menunjukkan *Term Frequency Inverse Document Frequency* menghasilkan akurasi sebesar 90 % dibandingkan BM 25 89% serta metode ini dapat menghindari terjadinya *over estimation* and *underestimation* dari panjang sebuah teks(Jo.t,2019).Penelitian ini menggunakan algoritma Naive Bayes dengan menggunakan nilai probabilitas untuk kasus klasifikasi dengan efisien dan cepat (Irwanyah,2021), algoritma Naive Bayes memiliki waktu komputasi lebih cepat sebesar 1.5 detik dibandingkan algoritma Support Vector Machine sebesar 2.5 detik(Onah.dkk,2019). Algoritma Random Forest dalam menyelesaikan masalah klasifikasi dengan membangun pohon keputusan dalam mengambil suara mayoritas dari setiap pohon, algoritma ini dapat mengurangi *overfitting* karena nilai akhir didapatkan dari suara mayoritas (Irwanyah,2021). Algoritma Random Forest cukup efektif dalam klasifikasi karena menggunakan prinsip *Law of Large Numbers* atau hukum bilangan besar sehingga dapat menghindari terjadi *overfitting* (Breiman, L.2001).Penelitian ini menggunakan Pima Diabetes dataset diakusisi dari situs dataset terbuka UCI *Machine Learning* terdapat 768 data memiliki 9 fitur dataset serta dua class yaitu terkena diabetes sebagai 1 dan tidak terkena diabetes sebagai 0. Dataset digunakan untuk proses pelatihan serta pengujian perbandingan algoritma Naive Bayes dan algoritma Random Forest. Hasil penelitian untuk mengetahui performa, akurasi, konsumsi daya terkait penggunaan CPU dan memori dalam proses algoritma dan metode berlangsung, kinerja, dan waktu komputasi metode *Term Frequency Inverse Document Frequency* serta perbandingan antara algoritma Naive Bayes dengan algoritma Random Forest dalam perangkat ponsel cerdas berbasis Anroid dengan spesifikasi rendah maupun tinggi dalam aplikasi prediksi diabetes menerapkan fitur *chat* dalam aplikasi menerapkan *on device Machine Learning* perhitungan algoritma terjadi hanya dalam lingkup aplikasi.

## **1.2 Rumusan masalah**

Berdasarkan latar belakang diatas, maka penulis mengidentifikasi beberapa rumusan masalah yakni sebagai berikut :

1. Bagaimana penerapan metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetic berbasis Android ?

2. Bagaimana perbandingan akurasi dan performa algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi berbasis Android berdasarkan akurasi, performa, dan konsumsi sumber daya proses ?
3. Bagaimana akurasi dan performa metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetis berbasis Android ?
4. Bagaimana waktu komputasi dan konsumsi sumber daya dibutuhkan selama proses metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetis berbasis Android ?

### **1.3 Batasan Masalah**

Mengingat banyaknya pengembangan yang dapat dilakukan dalam topik penelitian ini, adapun batasan-batasan masalah dalam penelitian ini sebagai berikut :

1. Dataset yang digunakan merupakan data sekunder yang di ambil secara tidak langsung pada website *open-source* data UCI Machine learning dataset ini terdiri atas 768 data dengan 8 atribut seperti jumlah glukosa, berat badan, umur, ketebalan kulit, nilai pedigree, jumlah insulin, tekanan darah, dan label class terdapat 2 yaitu ya dan tidak termasuk diabetes .

### **1.4 Tujuan Penelitian**

Tujuan yang ingin dicapai dalam penelitian ini pada uraian sebagai berikut :

1. Mengetahui penerapan metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetis berbasis Android.
2. Mengetahui perbandingan akurasi dan performa algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi berbasis Android berdasarkan akurasi, performa, dan konsumsi sumber daya proses.

3. Mengetahui akurasi dan performa metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetic berbasis Android.
4. Mengetahui waktu komputasi dan konsumsi sumber daya dibutuhkan selama proses metode *Term Frequency Inverse Document Frequency* serta algoritma Naive Bayes dan algoritma Random Forest dalam aplikasi diabetic berbasis Android.

### **1.5 Manfaat Penelitian**

Terdapat beberapa manfaat penelitian bagi masyarakat dan pembaca sebagai berikut ini:

1. Penelitian ini diharapkan dapat membantu masyarakat dalam diagnosa awal penyakit diabetes serta perawatan diabetes.
2. Penelitian ini diharapkan dapat memberikan bagaimana implementasi metode *Term Frequency Inverse Document Frequency* dan algoritma Naive Bayes dan algoritma Random Forest dalam penerapan dalam aplikasi mobile Android.
3. Penelitian ini diharapkan dapat memberikan wawasan terkait performa, kaurasi waktu komputasi dan konsumsi sumber daya yang dibutuhkan oleh metode *Term Frequency Inverse Document Frequency* dan algoritma Naive Bayes dan algoritma Random Forest dalam penerapan dalam aplikasi mobile Android.