

# 1. PENDAHULUAN

## 1.1. *State Of The Art*

Selama beberapa tahun terakhir penderita diabetes terus meningkat secara signifikan. Berdasarkan data yang didapatkan dari *International Diabetes Federation* (IDF) menginformasikan bahwa terdapat 415 juta orang yang menderita diabetes di tahun 2015 di seluruh dunia ada jumlah diprediksikan akan terus meningkat hingga lebih dari 640 juta pengidap diabetes di tahun 2040. Diabetes memiliki dampak yang fatal jika tidak tertangani dengan baik sejak dini. Beberapa efek samping dari diabetes yaitu hipertensi, kolesterol, masalah kesehatan otak, penyakit gusi dan gangguan pendengaran [1].

Di Indonesia menurut tim jurnalisme data harian kompas, diperkirakan akan terjadi peningkatan beban biaya penanganan diabetes di masa mendatang. Peserta Jaminan Kesehatan Nasional (JKN) juga diproyeksikan meningkat dari 6,9 juta orang pada tahun 2022 menjadi 10,2 juta orang pada tahun 2045. Selain itu, terdapat perkiraan bahwa biaya rata-rata per peserta dalam penanganan diabetes selama lima tahun terakhir (2018-2022) akan mengalami peningkatan dari Rp 912.538 menjadi Rp 2.309.599 dengan mempertimbangkan inflasi rata-rata selama 10 tahun terakhir sebesar 4,12 persen per tahun. Dengan demikian, total biaya penanganan diabetes untuk 10,2 juta peserta pada tahun 2045 diperkirakan mencapai Rp 23,59 triliun. Angka ini menunjukkan kenaikan yang signifikan dibandingkan dengan biaya saat ini sebesar Rp 6,3 triliun. Padahal 59% total dari penderita diabetes di Indonesia saat ini menjadi pasien BPJS Kesehatan [2]. angka tersebut tentu akan mempengaruhi keuangan dari BPJS Kesehatan yang merupakan bagian dari instansi kesehatan di negara Indonesia.

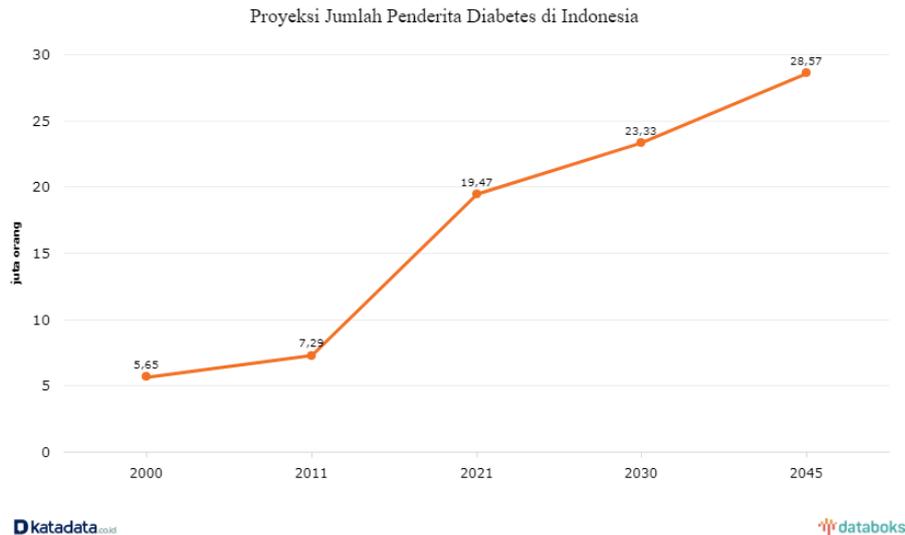
Untuk mengatasi masalah tersebut maka dapat dilakukan tindakan deteksi dini untuk menangani kejadian diabetes sebelum menjadi semakin fatal. Deteksi atau prediksi diabetes dapat digunakan berbagai macam metode, salah satu pendekatan yang dapat digunakan dari bidang ilmu sistem informasi adalah dengan mengimplementasikan *machine learning* atau *data mining* [5-8]. Namun beberapa metode yang telah dilakukan oleh penelitian sebelumnya belum menunjukkan hasil yang baik dan langkah langkah dalam setiap fasenya tidak terdokumentasi dengan baik.

Dalam penelitian ini penelitian mengembangkan algoritma dengan tujuan untuk meningkatkan performa algoritma yang dalam hal ini adalah akurasi prediksi. Pengembangan tersebut akan dilakukan dengan memodifikasi langkah langkah yang ada pada setiap algoritma dan melakukan komparasi antara performa dari algoritma terdahulu dengan algoritma yang telah di modifikasi.

## **1.2. Latar Belakang Penelitian**

Diabetes merupakan salah satu penyakit yang paling banyak menimpa masyarakat saat ini. Diabetes adalah ketidaksempurnaan metabolik yang ditandai dengan adanya hiperglikemia yang dihasilkan dari kesalahan sekresi insulin, aksi insulin, atau keduanya. Hiperglikemia kronis diabetes seringkali mengakibatkan dengan pada organ tubuh yang berbeda beda, utamanya mata, ginjal, saraf, jantung, dan pembuluh darah [7]. Diabetes secara luas juga dianggap sebagai epidemi yang muncul yang memiliki dampak kumulatif pada hampir setiap negara di dunia ini [8]. Pada tahun 2021 organisasi *International Diabetes Federation* (IDF) menginformasikan bahwa terdapat 537 juta orang dewasa dengan umur 20 - 79 tahun mengidap diabetes. Artinya terdapat 1 dari 10 orang hidup dengan diabetes di seluruh dunia. Diabetes juga menyebabkan 6,7 juta kematian atau 1 kematian setiap 5 detik. Fokus di negara Indonesia merupakan negara dengan pengidap diabetes urutan 5 di dunia, dengan jumlah pengidap diabetes mencapai 19,47 juta penduduk. Artinya tingkat prevalensi diabetes di Indonesia mencapai 10.6% [9]. Pada gambar dibawah menjelaskan bahwa menurut *International Diabetes Federation* (IDF) memroyeksikan bahwa jumlah penderita diabetes di Indonesia pada tahun 2045 bisa mencapai 28,57 juta kasus. Itu artinya meningkat sebanyak 47% jika dibandingkan dengan kasus pada tahun 2021.

Diabetes memiliki dampak yang fatal jika tidak tertangani dengan baik sejak dini. Beberapa efek samping dari diabetes yaitu hipertensi, kolesterol, masalah kesehatan otak, penyakit gusi dan gangguan pendengaran [1]. Selain itu penderita diabetes juga salah satu yang menjadi beban dalam jaminan kesehatan yang dikelola oleh BPJS Kesehatan. Pada tahun 2045 diproyeksikan beban penanganan diabetes pada BPJS Kesehatan akan meningkat mencapai 10.22 triliun sampai dengan 23.59 triliun rupiah.



Gambar 1 Proyeksi Jumlah Penderita Diabetes di Indonesia  
 Proyeksi Jumlah Penderita Diabetes di Indonesia

Diabetes memberikan dampak yang buruk baik bagi individu maupun ekonomi negara secara umum. Oleh karenanya untuk menghindari hal buruk tersebut maka pencegahan adalah kunci utama. Dengan mengetahui kemungkinan terjadinya diabetes pada individu. Maka individu tersebut dapat melakukan tindakan preventif seperti mengatur pola makan, mengurangi konsumsi gula, mengurangi konsumsi minyak, dan lain sebagainya.

Untuk memprediksi kemungkinan terjadinya diabetes dapat dilakukan dengan berbagai macam teknik. Selama beberapa dekade terakhir sudah banyak peneliti yang berupaya untuk memberikan solusi prediksi diabetes dengan menggunakan pendekatan teknologi. Salah satunya adalah dengan menggunakan teknik *data mining*. *Data mining* merupakan salah satu teknik yang dapat digunakan untuk memprediksi kejadian diabetes berdasarkan data yang ada. Beberapa penelitian terdahulu yang sudah dilakukan memprediksi diabetes yaitu penelitian berjudul “*Prediction of Diabetes using Classification Algorithms*” yang ditulis oleh Deepti Sisodia dan Dilip Singh Sisodia. Penelitian tersebut menggunakan algoritma Naïve Bayes, Support Vector Machine, dan Decision Tree. Menghasilkan akurasi sejumlah 76.3%, 65.1%, dan 73.82% untuk masing masing algoritma [3]. Penelitian lain berjudul “*Diabetes Prediction using Machine Learning Algorithms*” yang ditulis oleh Aishwarya Mujumdar dan Dr. Vaidehi V dengan menggunakan beberapa algoritma lain seperti Random Forest dan K Nearest Neighbours

mendapatkan akurasi yang sama yaitu 72% untuk kedua algoritma tersebut [5].

Seiring dengan berjalannya waktu tentu algoritma juga terus berkembang. Akan hadir banyak algoritma klasifikasi baru yang dapat digunakan untuk metode prediksi kejadian diabetes. Seiring dengan perkembangan pengetahuan tentu juga akan berbanding lurus dengan perkembangan algoritma yang semakin hari semakin bertambah. Algoritma Penelitian terdahulu mayoritas hanya membandingkan antara algoritma satu dengan algoritma lainnya untuk mendapatkan yang terbaik. Akibatnya adalah pengimplementasian metode prediksi terbatas hanya dengan menggunakan salah satu algoritma saja. Sehingga menutup kemungkinan akan adanya hasil prediksi dari algoritma yang lain. Padahal tiap tiap algoritma masing masing memiliki kelebihan yang tentu bisa menjadi dijadikan pertimbangan dalam menentukan hasil prediksi.

Jika tiap tiap algoritma prediksi dianalogikan sebagai konsultan, dan studi kasus dianggap sebagai sebuah bisnis. Maka dengan menggunakan beberapa macam algoritma artinya bisnis tersebut menggunakan beberapa konsultan untuk menghadapi studi kasus yang sedang dihadapi. Dan suatu bisnis dapat mengambil tindakan terbaik berdasarkan mayoritas saran yang diberikan oleh masing masing konsultan tersebut. Penelitian ini bertujuan untuk melakukan pengembangan algoritma klasifikasi dengan menggabungkan beberapa algoritma klasifikasi dengan menggunakan metode *ensemble method*. Dengan menggunakan metode ini maka hasil yang didapatkan akan menjadi lebih komprehensif karena mempertimbangkan berbagai macam algoritma data mining yang mempunyai kelebihannya masing masing.

### 1.3. Rumusan Masalah

Diabetes merupakan salah satu penyakit kronis yang belum bisa diatasi di seluruh dunia. Menurut data yang didapat dari organisasi *International Diabetes Federation* (IDF) mengatakan bahwa, pada tahun 2021 terdapat 537 juta orang dewasa dengan umur 20 - 79 tahun yang menjadi pengidap diabetes [9]. Diagnosis dini dan prediksi yang akurat dalam kasus diabetes merupakan hal yang sangat penting. Komplikasi yang serius dapat terjadi jika diabetes tidak terdiagnosis atau tidak dikelola dengan baik. Beberapa komplikasi yang mungkin terjadi termasuk penyakit jantung, gagal ginjal, kerusakan saraf, gangguan penglihatan, dan masalah sirkulasi. Memiliki metode prediksi yang akurat dapat membantu mengurangi risiko komplikasi dan meningkatkan kualitas hidup pasien dengan diabetes. Diagnosis dini memungkinkan pasien untuk segera memulai pengobatan yang tepat dan mengadopsi perubahan gaya hidup yang sehat. Prediksi yang akurat juga memainkan peran penting dalam pengelolaan penyakit ini, karena memungkinkan dokter dan pasien untuk memonitor perkembangan penyakit, mengidentifikasi risiko komplikasi yang lebih tinggi, dan mengambil tindakan pencegahan yang tepat.

Algoritma prediksi diabetes saat ini masih dihadapkan pada tantangan dalam mencapai tingkat akurasi yang optimal. Faktor-faktor seperti kompleksitas data pasien, interaksi antarfitur, dan salah satu faktor yang paling utama adalah penggunaan metode pemrosesan data dapat mempengaruhi performa algoritma secara signifikan. Pemilihan metode pemrosesan data yang tepat menjadi faktor penting dalam meningkatkan performa algoritma. Penggunaan metode yang tidak sesuai dapat menghasilkan informasi yang tidak relevan atau bahkan mengabaikan faktor-faktor penting dalam prediksi diabetes. Berdasarkan konteks yang telah dijabarkan sebelumnya maka penelitian ini akan mengkaji metode untuk mengembangkan algoritma prediksi diagnosa kejadian diabetes dengan menggunakan teknik *ensemble method* dengan tujuan untuk meningkatkan performa algoritma yang dalam hal ini adalah akurasi. Dengan harapan dapat memberikan kontribusi dalam pengelolaan diabetes secara lebih efektif dan mengurangi dampak negatif yang terjadi akibat dengan kondisi ini baik di Indonesia maupun di dunia.

#### **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah disampaikan di bagian sebelumnya maka ditentukan tujuan dari penelitian ini yaitu :

- Mengembangkan algoritma *data mining* untuk prediksi kejadian diabetes.
- Membandingkan performa akurasi antara algoritma terdahulu dengan algoritma yang akan dikembangkan.

#### **1.5. Pertanyaan Penelitian**

Berdasarkan tujuan penelitian yang telah di sampaikan pada bagian sebelumnya maka dapat ditentukan pertanyaan dari penelitian ini yaitu :

- Bagaimana cara yang dapat dilakukan untuk mengembangkan algoritma prediksi diabetes?
- Bagaimana hasil akurasi dari algoritma terdahulu dengan algoritma *ensemble method* yang dikembangkan?

#### **1.6. Lingkup Penelitian**

Berikut merupakan ruang lingkup dengan batasan dan asumsi yang akan dibahas pada penelitian ini:

1. Dataset yang digunakan merupakan dataset pasien diabetes yang bersumber dari *Pima Indians Diabetes, National Institute of Diabetes and Digestive and Kidney Diseases, United States*.
2. Metode yang digunakan untuk eksperimen pada sistem prediksi ini adalah algoritma dasar yang belum mengalami perkembangan apapun.
3. Performa yang digunakan untuk menilai algoritma dalam hal ini adalah akurasi.
4. Penelitian ini terbatas sampai dengan tahap evaluasi untuk mendapatkan hasil dari algoritma terdahulu dan algoritma yang dikembangkan.

## 1.7. Signifikansi Studi

Metode prediksi diabetes dengan menggunakan *data mining* adalah salah satu metode yang dapat digunakan untuk dapat mengurangi tingkat kejadian yang ada di dunia. Metode ini bisa membantu *prevalensi* kemungkinan terjadinya diabetes terhadap suatu individu. Sehingga individu tersebut bisa mendapatkan deteksi kemungkinan terjadi diabetes dengan lebih mudah dan dan bisa memberikan *treatment* dengan lebih cepat untuk menghindari kemungkinan terjadinya hal yang fatal.

Saat ini penelitian terkait dengan prediksi diabetes cenderung hanya mengimplementasikan satu algoritma saja atau bisa juga disebut sebagai algoritma tunggal. Beberapa penelitian ada yang menggunakan metode *ensemble method* atau menggabungkan beberapa algoritma, namun tidak memberikan bobot pada masing masing algoritma sehingga terdapat kemungkinan terjadinya ketidakadilan dalam memberikan bobot untuk algoritma yang memiliki performa baik dan algoritma yang memiliki performa yang kurang baik.

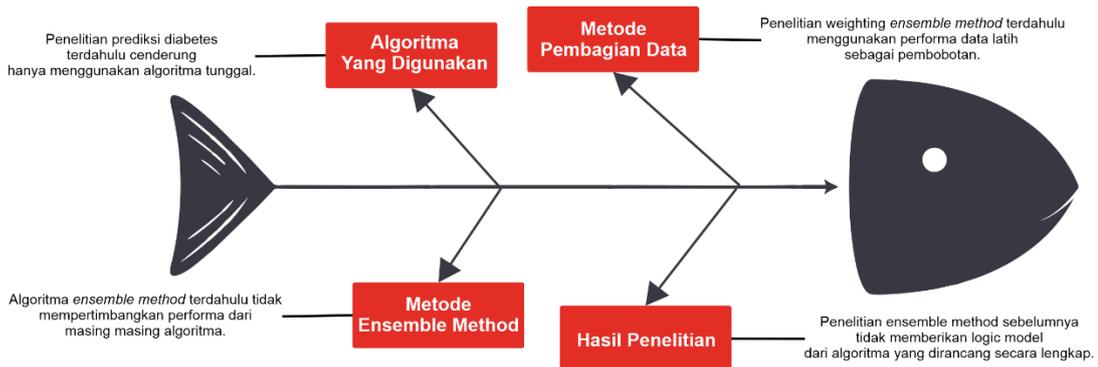
Penelitian ini memberikan signifikansi studi terkait dengan implementasi data mining untuk prediksi kasus diabetes, yaitu

- mengusulkan sebuah metode pemilihan algoritma prediksi sebelum digabungkan dengan menggunakan *ensemble method*
- mengusulkan sebuah metode pembobotan kepada tiap tiap algoritma yang nantinya akan digabungkan dengan menggunakan *ensemble method*
- merancang suatu *logic model* yang mudah dipahami sehingga dapat lebih mudah dalam diimplementasikan nantinya.

Penelitian juga dapat memberikan manfaat kepada instansi kesehatan nasional yang menangani kejadian diabetes dalam hal ini adalah pihak rumah sakit untuk dapat memprediksi kejadian diabetes dengan lebih cepat sehingga bisa memprioritaskan penanganan kepada pasien yang memiliki persentase kemungkinan terjadinya diabetes lebih tinggi terlebih dahulu, untuk nantinya dapat dilakukan tindakan preventif yang bisa mengurangi dampak buruk yang bisa terjadi.

## 1.8. Kesenjangan Penelitian

*Gap analysis* adalah pertanyaan atau masalah penelitian yang belum dijawab dengan tepat atau sama sekali belum terjawab dalam bidang studi tertentu [10]. Salah satu metode dari *gap analysis* adalah *fish bone diagram*, pada penelitian ini peneliti menggunakan Fish Bone Diagram sebagai metode untuk menganalisa kesenjangan dalam penelitian.



Gambar 2 Fish Bone Diagram Gap Analysis

Dari gambar diatas dapat diambil informasi bahwa terdapat empat poin utama yang bisa diidentifikasi sebagai gap yang berkaitan dengan topik dalam penelitian ini. Yang pertama adalah mengenai algoritma prediksi yang digunakan pada penelitian sebelumnya. Penelitian sebelumnya cenderung hanya menggunakan satu algoritma saja atau algoritma tunggal, sehingga menutup kemungkinan untuk adanya alternatif hasil yang lebih baik dari algoritma yang lain. Sedangkan seiring dengan berjalannya waktu tentu algoritma juga akan semakin bertambah banyak, dan menjadi tidak produktif jika nantinya hanya memanfaatkan satu algoritma saja, padahal ada kelebihan dari algoritma yang lain.

Yang kedua yaitu metode *ensemble method* yang digunakan dalam penelitian sebelumnya tidak memberikan pembobotan pada masing-masing algoritma yang ada, sehingga dikhawatirkan akan terjadi bias karena algoritma yang memiliki performa baik dan yang memiliki performa buruk memiliki bobot yang sama.

Yang ketiga adalah metode pembagian data, pada penelitian *weighting ensemble method* sebelumnya mayoritas tidak menjelaskan secara detail terkait dengan metode pembagian datanya. Selain itu ada juga yang menggunakan hasil performa dari data *train* sebagai pembobotan tetapi diujikan lagi dengan

menggunakan data *train* yang sama, sehingga terdapat bias pada algoritma tersebut.

Yang terakhir yaitu penelitian sebelumnya tidak secara detail menjelaskan proses perancangan *ensemble method* dari mulai tahapan *data cleansing* sampai dengan metode pembuatan *model*. Sehingga bisa menyulitkan peneliti lain yang akan melanjutkan atau mengembangkan hasil penelitian tersebut. Selain itu juga akan menyulitkan dari praktisi jika ada yang mengimplementasikan model tersebut.

Dari beberapa poin diatas dapat maka dapat diambil informasi bahwa terdapat gap yaitu ketidaksesuaian metode prediksi diabetes terdahulu karena cenderung hanya menggunakan algoritma tunggal, beberapa menggunakan algoritma ensemble namun tidak mempertimbangkan pembobotan dalam tiap algoritma penyusun, menggunakan akurasi dari data latih sebagai pembobotan, dan tidak memberikan dokumentasi yang lengkap.

## 1.9. Rasionalisasi Penelitian

Metode prediksi diabetes dengan menggunakan *data mining* sudah banyak diteliti oleh banyak peneliti. Berikut merupakan beberapa penelitian terbaru terkait dengan metode prediksi diabetes dengan menggunakan pendekatan *data mining*:

Tabel 3 Rasionalisasi dari penelitian terdahulu

Penulis	Judul	Tahun	Deskripsi
B. Shamreen Ahamed , Meenakshi S. Arya, dan Auxilia Osvin V. Nancy	Diabetes Mellitus Disease Prediction Using Machine Learning Classifiers with Oversampling and Feature Augmentation [11]	2022	Penelitian ini mengusulkan metode prediksi penyakit diabetes mellitus berbasis <i>machine learning</i> dengan membandingkan beberapa algoritma yaitu RF, GBM, dan LGBM.
Salliah Shafi Bhat , Venkatesan Selvam, Gufran Ahmad Ansari , Mohd Dilshad	Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir:	2022	Penelitian ini mengusulkan metode prediksi diabetes dengan membandingkan

Ansari , dan Md Habibur Rahman	A Case Study of District Bandipora [12]		beberapa algoritma yaitu <i>Random Forest (RF)</i> , <i>Multi-Layer Perceptron (MLP)</i> , <i>Support Vector Machine (SVM)</i> , <i>Gradient Boost (GB)</i> , <i>Decision Tree (DT)</i> , dan <i>Logistic Regression (LR)</i>
Rashi Rastogi dan Mamta Bansal	Diabetes prediction model using data mining techniques [13]	2022	Penelitian ini mengusulkan metode prediksi diabetes dengan membandingkan beberapa algoritma yaitu <i>Random Forest</i> , <i>Support Vector Machine (SVM)</i> , <i>Logistic Regression</i> , dan <i>Naive Bayes</i>

Berdasarkan tabel diatas dapat diambil informasi bahwa masih sedikit penelitian terkait dengan prediksi diabetes yang menggunakan *ensemble method* dan memberikan pembobotan kepada masing masing algoritma penyusun. Sehingga rasionalisasi dari penelitian ini yaitu :

- Penelitian ini dapat membantu pihak instansi kesehatan nasional yang dalam hal ini adalah rumah sakit dalam memprediksi diabetes secara lebih cepat dan lebih akurat.
- Penelitian ini menekankan pentingnya pembobotan untuk masing masing algoritma penyusun dalam *ensemble method* agar tidak terjadi bias dalam pengambilan kesimpulan.
- Penelitian ini memberikan *novelty* yaitu akurasi yang lebih tinggi dibandingkan dengan algoritma tunggal dan *logic model* yang terdokumentasi sehingga lebih mudah dipahami.

## **1.10. Sistematika Laporan**

Sistematika penulisan laporan tugas akhir ini terdiri dari beberapa bab yang meliputi:

### **BAB I PENDAHULUAN**

Pada bab ini berisi uraian mengenai latar belakang dilakukannya penelitian dengan batasan-batasan yang telah ditentukan sebelumnya serta solusi apa yang akan diberikan melalui penelitian.

### **BAB II TINJAUAN LITERATUR**

Pada bab ini membahas tentang teori-teori relevan yang dapat digunakan untuk mendukung penelitian.

### **BAB III METODOLOGI PENELITIAN**

Pada bab ini berisikan penjelasan terkait model konseptual dari penelitian yang kemudian menentukan bagaimana penelitian dilakukan berdasarkan metodologi yang dipilih. Sebelumnya dilakukan penjelasan mengenai metodologi penelitian. Metodologi penelitian mencakup sistematika dari penelitian yang dilakukan.

### **BAB IV PEMBAHASAN**

Bab ini berisikan hasil analisis dari data yang telah dikumpulkan dan model yang telah dirancang. Dalam bab ini penulis akan menjabarkan tahapan analisis yang telah dilakukan.

### **BAB V ANALISA DATA**

Bab ini berisikan proses pembangunan model dan tahapan analisa data dari hasil evaluasi model yang telah di bangun.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini berisikan pernyataan kesimpulan dan rekomendasi yang berasal dari analisis hasil penelitian yang telah dilaksanakan.