

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada dasarnya, kehamilan merupakan suatu proses yang mengalami perubahan fisik dan emosional yang cukup signifikan oleh seorang ibu. Selama masa kehamilan, terdapat beberapa faktor yang dapat meningkatkan risiko terjadinya komplikasi baik pada ibu maupun janin. Faktor-faktor tersebut di antaranya adalah usia ibu yang terlalu muda atau terlalu tua, obesitas, riwayat penyakit seperti diabetes atau hipertensi, serta riwayat kehamilan yang tidak normal sebelumnya. Faktor usia ibu bila mengandung pada usia diatas 35 tahun dapat meningkatkan risiko terjadinya kelainan kromosom pada janin, seperti sindrom *Down* [1]. Faktor lain yaitu obesitas pada ibu hamil dapat meningkatkan risiko komplikasi kehamilan, termasuk diabetes gestasional dan tekanan darah tinggi [2]. Menurut data pada tahun 2018, sekitar 76% dari kematian ibu terjadi selama fase persalinan dan pasca persalinan, dengan 24% terjadi selama kehamilan, 36% selama persalinan, dan 40% pasca persalinan [3]. Persentase tersebut menunjukkan pentingnya pengembangan sistem deteksi dini terkait risiko kehamilan.

Untuk menentukan level risiko kehamilan pada seorang ibu, biasanya dilakukan dengan mengumpulkan data tentang kondisi fisik dan emosional ibu tersebut selama masa kehamilan. Data tersebut kemudian diolah dengan menggunakan metode klustering dan klasifikasi yang dapat membantu dalam mengidentifikasi tingkat risiko kehamilan pada setiap individu. Dataset terkait penentuan level risiko kehamilan yang digunakan pada penelitian ini didapatkan dari UCI *Machine Learning repository*.

Penelitian terdahulu terkait identifikasi level risiko kehamilan pada seorang ibu menggunakan metode klasifikasi telah umum dilakukan, seperti pada artikel jurnal “*Model for Predicting Risk Levels in Maternal Healthcare*” [4]. Penelitian ini berfokus pada pengembangan model prediktif untuk menilai tingkat risiko dalam

perawatan kesehatan ibu. Tujuan utamanya adalah untuk mengidentifikasi faktor risiko yang mempengaruhi kesehatan ibu selama masa kehamilan dan melahirkan, sehingga dapat memberikan perawatan yang lebih baik dan mengurangi angka kematian ibu. Data yang ada kemudian dianalisis menggunakan teknik-teknik statistik dan pemodelan matematis untuk mengidentifikasi faktor-faktor yang berhubungan dengan risiko kesehatan ibu. Data tersebut termasuk parameter vital seperti tekanan darah, denyut nadi, dan suhu tubuh dan sebagainya [4]. Pada penelitian tersebut hanya melibatkan penggunaan algoritma klasifikasi saja, sedangkan pada penelitian yang sedang dilakukan melibatkan penggunaan algoritma klustering yang hasilnya digunakan pada algoritma klasifikasi.

Pada penelitian sebelumnya digunakan algoritma KNN, Neural Network, Random Forest, serta AdaBoost untuk membangun model prediksi risiko tingkat kehamilan. Selain itu, pada penelitian juga digunakan Python sebagai bahasa pemrograman dan Jupyter Notebook sebagai platform pengembangan. Penelitian tersebut menghasilkan model prediksi risiko tingkat kehamilan dengan teknik klasifikasi. Model tersebut dapat membantu dokter dalam memprediksi risiko kehamilan dan memberikan rekomendasi pengobatan yang tepat. Hasil penelitian menunjukkan bahwa model yang dikembangkan dapat memprediksi tingkat risiko dalam perawatan kesehatan ibu dengan akurasi yang cukup baik sebesar 86% pada model terbaik yang dihasilkan, yaitu pada algoritma AdaBoost, sedangkan pada algoritma dengan performa terendah ada pada algoritma KNN dengan akurasi sebesar 64%. Beberapa faktor risiko utama yang diidentifikasi meliputi usia ibu, riwayat kelahiran sebelumnya, status gizi, dan adanya komplikasi medis [4].

Algoritma K-Means merupakan salah satu algoritma klustering yang digunakan untuk mengelompokkan data berdasarkan kesamaan fitur-fiturnya. Setelah proses klustering selesai, maka tahap selanjutnya adalah melakukan klasifikasi terhadap setiap kluster yang telah dibentuk oleh algoritma K-Means. Algoritma klasifikasi dibutuhkan untuk membaca hasil klustering karena klustering hanya mengelompokkan data ke dalam kelompok atau kluster berdasarkan kemiripan yang dimiliki, tanpa memberikan label atau informasi kategori apa pun pada kluster

tersebut. Algoritma klasifikasi membantu mengidentifikasi pola atau karakteristik dari setiap kluster dan memberikan label atau kategori pada kluster tersebut berdasarkan pola yang dipelajari. Dengan menggunakan algoritma klasifikasi, kita dapat mengklasifikasikan data baru ke dalam kluster yang ada atau mengidentifikasi kluster baru yang belum diketahui sebelumnya [5].

Untuk melakukan klasifikasi tersebut, terdapat beberapa algoritma klasifikasi yang dapat digunakan, di antaranya adalah Naive Bayes, SVM (Support Vector Machine), dan KNN (K-Nearest Neighbors). Masing-masing algoritma tersebut memiliki keunggulan dan kelemahan masing-masing. Naive Bayes, SVM, dan KNN adalah tiga algoritma klasifikasi yang sering digunakan dalam analisis data. Ketiga algoritma ini juga dapat digunakan untuk mengklasifikasi hasil klustering dengan hasil yang akurat dan efektif [6]. Oleh karena itu, ketiga algoritma ini digunakan dalam perbandingan algoritma klasifikasi untuk mengetahui algoritma mana yang paling sesuai digunakan dalam menentukan level risiko kehamilan pada seorang ibu berdasarkan hasil klustering yang dilakukan. Analisis perbandingan algoritma klasifikasi tersebut dilakukan dengan mengukur tingkat akurasi, presisi, *recall*, dan *f1-score* dari masing-masing algoritma. Dengan demikian, akan diketahui algoritma klasifikasi mana yang paling baik dalam menentukan level risiko kehamilan pada seorang ibu berdasarkan hasil klustering yang dilakukan. Tingkat akurasi, presisi, *recall*, dan *f1-score* adalah metrik evaluasi yang umum digunakan dalam membandingkan hasil prediksi algoritma dalam berbagai bidang, termasuk dalam *machine learning*, *data mining*, dan analisis prediktif. Akurasi digunakan untuk mengukur seberapa baik algoritma dapat mengklasifikasikan data dengan benar secara keseluruhan, sedangkan presisi mengukur seberapa baik algoritma dapat mengidentifikasi kelas positif dengan benar [7]. *Recall* mengukur seberapa baik algoritma dapat menemukan semua kelas positif yang relevan, sedangkan *f1-score* adalah *harmonic mean* dari presisi dan *recall*, dan digunakan untuk memperoleh pemahaman yang lebih komprehensif tentang kinerja algoritma [8].

Beberapa penelitian telah menggunakan teknik klustering untuk memperbaiki akurasi klasifikasi dalam berbagai bidang. Salah satunya terdapat pada penelitian menggunakan teknik *semi-supervised clustering* untuk meningkatkan akurasi klasifikasi serangan DDoS pada jaringan. Hasil penelitian menunjukkan bahwa teknik *clustering semi-supervised* mampu meningkatkan akurasi klasifikasi dibandingkan dengan metode *supervised learning* tradisional [9].

Penelitian terkait risiko kehamilan telah dilakukan dengan menggunakan klasifikasi risiko World Health Organization yang dimodifikasi (mWHO) pada 2742 wanita hamil di negara-negara maju dan berkembang. Hasil penelitian menunjukkan bahwa klasifikasi mWHO hanya memiliki kinerja moderat dalam membedakan wanita yang mengalami dan tidak mengalami risiko selama kehamilan dengan tingkat akurasi sebesar 71.1% [10]. Oleh karena itu, diperlukan penelitian lanjutan untuk mengembangkan dan meningkatkan performa model yang telah dibuat.

Penelitian lain juga pernah dilakukan dengan menggunakan metode K-means *clustering* dan pendekatan klasifikasi untuk mengembangkan sistem prediksi diabetes yang akurat. Fitur-fitur yang paling signifikan dalam memprediksi diabetes adalah usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah puasa, dan riwayat keluarga. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan memiliki akurasi sebesar 91.15% [11].

Selain itu, beberapa penelitian lain seperti "MRI brain tumour segmentation using hybrid clustering and classification by back propagation algorithm" [12], "Prediction of heart disease by clustering and classification techniques" [13], "A New Intelligent Medical Decision Support System Based on Enhanced Hierarchical Clustering and Random Decision Forest for the Classification of Alcoholic Liver Damage, Primary Hepatoma, Liver Cirrhosis, and Cholelithiasis" [14], dan "Unsupervised classification of eclipsing binary light curves through k-medoids clustering" [15] juga menggunakan teknik klustering untuk memperbaiki akurasi klasifikasi dalam bidang yang berbeda.

Pada penelitian ini, akan dibuat model prediksi level risiko kehamilan menggunakan teknik klustering dan klasifikasi secara berurutan, yang berguna untuk menentukan apakah seorang ibu berada pada level risiko rendah, menengah, ataupun tinggi pada masa kehamilannya. Perbedaan penelitian yang telah ada [4] dengan penelitian ini terdapat pada penggunaan metode klustering terlebih dahulu yaitu algoritma K-Means sebelum dilakukan klasifikasi menggunakan algoritma Naive Bayes, SVM (*Support Vector Machine*), dan KNN (*K-Nearest Neighbors*) dalam penentuan level risiko kehamilan pada seorang ibu. Hal tersebut dilakukan dengan tujuan untuk mengetahui apakah algoritma klustering dapat meningkatkan akurasi model algoritma klasifikasi.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan sebelumnya, permasalahan yang akan dibahas pada penelitian yang dilakukan ialah sebagai berikut.

1. Apakah penerapan klustering menggunakan algoritma K-Means dapat mempengaruhi akurasi dari algoritma Naïve Bayes, SVM, dan KNN dalam mengklasifikasi level risiko kehamilan pada seorang ibu?
2. Algoritma pemodelan manakah yang paling sesuai untuk digunakan mengklasifikasi level risiko kehamilan pada seorang ibu berdasarkan hasil akurasi?
3. Bagaimana pembuatan prototipe implementasi hasil pemodelan untuk memprediksi level risiko kehamilan pada seorang ibu?

## 1.3 Batasan Masalah

Dalam penentuan level risiko kehamilan pada seorang ibu, terdapat beberapa Batasan masalah yang ditetapkan. Batasan masalah tersebut ialah sebagai berikut.

- a. Dataset yang digunakan berupa data Excel yang didapatkan dari UCI *Machine Learning repository*.
- b. Data yang diolah berupa data numerik.
- c. Data yang digunakan adalah data informasi seperti usia, *upper value* tekanan darah, *lower value* tekanan darah, kadar glukosa darah, suhu tubuh,

detak jantung normal, dan hasil penentuan tingkat risiko kehamilan pada seorang ibu.

- d. Data yang digunakan berjumlah 1.014 baris, berupa data wanita hamil di salah satu rumah sakit yang berada di Bangladesh.
- e. Proses klustering hanya menggunakan algoritma K-Means, yang dilanjutkan dengan proses klasifikasi menggunakan algoritma Naïve Bayes, SVM, dan KNN.

## **1.4 Tujuan dan Manfaat Penelitian**

### **1.4.1 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah ditentukan, tujuan penelitian pada penelitian yang dilakukan ialah sebagai berikut.

1. Menganalisis pengaruh penerapan klustering menggunakan algoritma K-Means terhadap akurasi klasifikasi level risiko kehamilan oleh algoritma Naïve Bayes, SVM, dan KNN.
2. Menganalisis dan membandingkan tingkat akurasi berbagai algoritma pemodelan untuk mengidentifikasi algoritma yang paling sesuai untuk mengklasifikasi level risiko kehamilan pada seorang ibu.
3. Membuat prototipe yang mampu memprediksi level risiko kehamilan pada seorang ibu berdasarkan hasil pemodelan.

### **1.4.2 Manfaat Penelitian**

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Dapat memberikan informasi mengenai hasil kombinasi algoritma K-Means yang diklasifikasi dengan algoritma Naïve Bayes, SVM, dan KNN dalam menentukan level risiko kehamilan pada seorang ibu.
2. Memberikan rujukan yang dapat digunakan untuk pengembangan penelitian lain khususnya terkait pengembangan kombinasi algoritma klustering dan klasifikasi pada sistem prediksi level risiko kehamilan.
3. Menambah wawasan mengenai kombinasi algoritma klustering dan klasifikasi yang digunakan dalam penelitian ini, yaitu algoritma K-

Means, Support Vector Machine, Naïve Bayes dan K-Nearest Neighbour.

## 1.5 Sistematika Penulisan

Sistematika penulisan dari suatu penelitian harus terorganisir dengan baik agar mudah dipahami oleh pembaca. Berikut ini adalah sistematika penulisan yang digunakan dalam penulisan Skripsi ini.

### BAB 1 PENDAHULUAN

Bagian ini berisi latar belakang masalah yang mendasari dilakukannya penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan. Selain itu, dijelaskan juga secara singkat mengenai penentuan level risiko kehamilan pada seorang ibu, pemanfaatan algoritma klustering K-Means, serta algoritma klasifikasi Naïve Bayes, SVM, dan KNN pada bagian latar belakang.

### BAB II LANDASAN TEORI

Bagian ini berisi teori yang relevan dengan masalah yang diteliti, diantaranya mengenai klasifikasi, algoritma klasifikasi Naïve Bayes, SVM, dan KNN, serta algoritma klustering K-Means. Selain itu, juga dijelaskan mengenai konsep-konsep yang terkait dengan penelitian, seperti akurasi, presisi, *recall*, dan *f1-score*.

### BAB III METODOLOGI PENELITIAN

Bagian ini berisi deskripsi mengenai metode yang digunakan dalam penelitian, termasuk jenis penelitian, sampel yang digunakan, cara pengumpulan data, serta cara analisis data yang akan dilakukan.

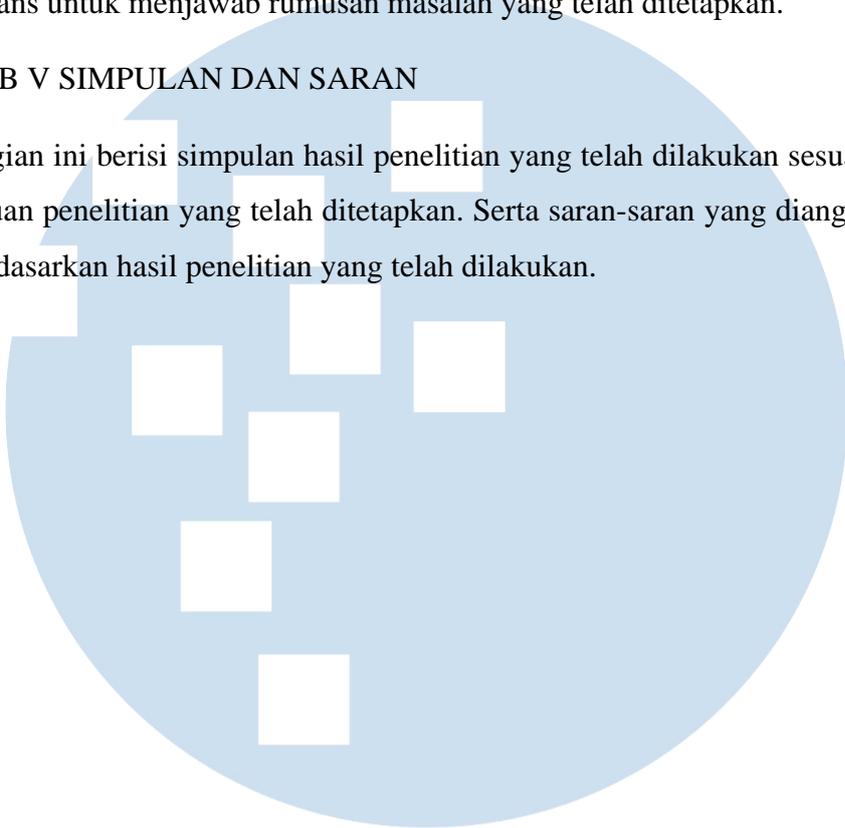
### BAB IV ANALISIS DAN HASIL PENELITIAN

Bagian ini berisi hasil-hasil yang diperoleh dari penelitian, diantaranya mengenai performa masing-masing algoritma klasifikasi Naive Bayes, SVM,

dan KNN yang telah dikombinasikan dengan hasil klustering algoritma K-Means untuk menjawab rumusan masalah yang telah ditetapkan.

## BAB V SIMPULAN DAN SARAN

Bagian ini berisi simpulan hasil penelitian yang telah dilakukan sesuai dengan tujuan penelitian yang telah ditetapkan. Serta saran-saran yang dianggap perlu berdasarkan hasil penelitian yang telah dilakukan.



# UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA