

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung koroner merupakan sebuah penyakit yang dikarenakan terlalu sedikit aktifitas fisik yang dilakukan dan pola makan yang kurang baik. Satu dari banyaknya penyebab penyakit jantung adalah pola hidup dalam kehidupan sehari - hari seperti merokok, mengkonsumsi makanan berlemak , stress, kerja berlebihan, gula darah, obesitas, aktivitas fisik, dan lain sebagainya. Penyakit jantung paling sering dijumpai pada usia 65–74 tahun (3,6%) kemudian diikuti usia 75 tahun ke atas (3,2%), usia 55–64 tahun (2,1%), dan usia 35–44 tahun (1,3%)(Anakonda Selviana et al., n.d.). Sejumlah faktor dapat menyebabkan terjadinya penyakit jantung koroner, termasuk merokok, pola hidup yang buruk, tingginya kadar gula darah dalam tubuh, kurangnya aktivitas fisik, obesitas, dan kondisi ekonomi. Menurut beberapa studi, gangguan metabolisme seperti diabetes mellitus diidentifikasi sebagai faktor utama yang berpotensi memicu penyakit jantung koroner. Penyebab diabetes mellitus pada tubuh tidak hanya disebabkan oleh hiperglikemi secara langsung, melainkan juga oleh kemunculan komplikasi, salah satunya yang berhubungan dengan jantung. Pasien diabetes, terutama tipe II, mempunyai probabilitas lebih tinggi terkena penyakit jantung koroner. Pada diabetes mellitus tipe II, pankreas masih menghasilkan insulin secara anatomi, namun tidak seefektif pankreas pada individu tanpa diabetes mellitus tipe II. Ini dapat menyebabkan resistensi insulin seiring waktu. Individu dengan risiko tinggi untuk mengembangkan diabetes mellitus perlu mendapatkan perhatian khusus, termasuk penyuluhan pencegahan terkait komplikasi seperti penyakit jantung kronis, untuk menghindari risiko munculnya penyakit tersebut. (Koroner et al., n.d.).

Penyakit jantung sendiri merupakan penyakit yang diketahui melalui diagnosa. Untuk melakukan prediksi penyakit jantung diperlukan beberapa tes. Karena kurangnya staf medis yang dapat memperkirakan kemungkinan prediksi yang salah (Pouriyeh et al., 2017).. Melakukan diagnosa sedini mungkin terhadap penyakit jantung merupakan sesuatu yang sangat sulit karena kekurangan tenaga medis yang ahli serta beberapa sumber daya yang diperlukan untuk melakukan diagnosis dan penanganan yang tepat bagi pasien yang memiliki masalah jantung (Ghwanmeh et al., 2013).. Evaluasi didasarkan prediksi yang akurat dari risiko terjadinya gagal jantung akan sangat membantu pasien dalam mencegah terjadinya serangan jantung yang parah dan mengurangi angka terjadinya gagal jantung (Al-Shayea, 2011). Salah satu cara yang paling efektif dalam mendiagnosis dan memprediksi adanya penyakit jantung adalah dengan menggunakan algoritma *machine learning* (Ghosh et al., 2021). *Machine learning* sendiri mampu dalam mengatasi kerumitan dalam mendiagnosis penyakit jantung dengan adanya model prediksi pembelajaran mesin *logistic regression* (LR), *Support Vector Machine* (SVM), *decision tree* (DT), *k-nearest neighbour* (KNN), *Naive Bayes* (NB), *AdaBoost* (AB), jaringan saraf tiruan (JST), *fuzzy logic* (FL) (Haq et al., 2018) (Nugraha, n.d.).

Lewat jurnal oleh (Jefri Junifer Pangaribuan et al., 2021) yang meneliti tentang " MENDETEKSI PENYAKIT JANTUNG MENGGUNAKAN *MACHINE LEARNING* DENGAN ALGORITMA *LOGISTIC REGRESSION* " penulis tertarik dalam membuat penelitian tentang prediksi penyakit jantung menggunakan pembelajaran mesin, kemudian peneliti kesulitan dalam menentukan penggunaan metode yang akan digunakan untuk melakukan prediksi (Junifer Pangaribuan & Tanjaya, 2021), lalu peneliti menemukan sebuah jurnal oleh (Ilias Tougui et al., 2020) yang berjudul " *Heart disease classification using data mining*

tools and machine learning techniques " pada jurnal ini, dilakukan perbandingan antara enam metode, yaitu *Naïve Bayes*, *Logistic Regression*, *KNN*, *Random Forest*, *SVM*, dan *ANN*. Penelitian ini menggunakan enam tools yang berbeda, yakni *Knime*, *Sci-Kit Learn*, *Weka*, *Matlab*, *RapidMiner*, dan *Orange*. Setiap menguji metode mempergunakan keenam buah tools tersebut. Hasil penelitian menyatakan dimana metode *Logistic Regression* dan *Naïve Bayes* mendapatkan nilai akurasi yang stabil dan cukup baik. Dari jurnal tersebut jumlah metode yang akan digunakan dipersempit menjadi dua pilihan (Tougui et al., 2020). sembari mencari jurnal yang berhubungan dengan topik yang sama, peneliti menemukan bahwa dengan menggunakan metode *feature selection* dapat meningkatkan nilai akurasi metode klasifikasi dan disini pengguna menemukan jurnal oleh (XuKui Li et al., 2020) yang berjudul "*Building Auto-Encoder Intrusion Detection System based on random forest feature selection*" yang dimana jurnal ini menggunakan *random forest* sebagai metode *feature selection*nya yang dimana biasanya *random forest* adalah metode yang biasanya digunakan untuk metode klasifikasi (Li et al., 2020). dari jurnal tersebut penulis memutuskan untuk melakukan perbandingan antara dua *feature selection* yakni *random forest* dan *particle swarm optimization* yang dimana *particle swarm optimization* merupakan metode seleksi fitur yang paling sering ditemui (Khourdifi & Bahaj, 2019). penggunaan *feature selection* ini akan diterapkan pada klasifikasi *machine learning* yang menggunakan dua metode yakni *logistic regression* dan *naive bayes* karena sudah disebutkan sebelumnya kedua metode klasifikasi tersebut adalah klasifikasi yang memiliki nilai akurasi terbaik.

Pada penelitian ini penulis memiliki keinginan membuat dan tertarik untuk mengembangkan penelitian sebuah klasifikasi menggunakan algoritma *Logistic Regression* dan *Naive Bayes* yang di optimasi menggunakan *Feature Selection Random Forest* Dan

Particle Swarm Optimization (PSO) dengan judul " Perbandingan Metode Klasifikasi *Logistic Regession* Dan *Naive Bayes* Yang Di Optimasi Menggunakan *Feature Selection Random Forest* Dan *Particle Swarm Optimization (PSO)* Pada Prediksi Penyakit Jantung " Penelitian ini mengimplementasikan *feature selection* dengan tujuan akan meningkatkan kerja prediksi, mengurangi waktu komputasi, dan untuk mendapatkan pemahaman yang lebih baik mengenai data dalam *machine learning* atau *pattern recognition applications* (Chandrashekar & Sahin, 2014). Pada penelitian ini penulis ingin melihat hasil perbandingan klasifikasi *Logistic Regession* dan *Naive Bayes* menggunakan *feacture selection Random Forest* dan *Particle Swarm Optimization (PSO)* pada prediksi penyakit jantung bagi penderita diabetes dengan melihat model klasifikasi yang dihasilkan setelah melakukan seleksi fitur apakah dapat menghasilkan peningkatan akurasi klasifikasi dan menghasilkan model klasifikasi yang memiliki tingkat akurasi yang lebih baik atau tidak

1.2 Rumusan Masalah

Berdasarkan masalah yang disampaikan oleh penulis maka penting untuk mengkaji atau mengangkat lebih detail mengenai Perbandingan Klasifikasi *Logistic Regession* dan *Naive Bayes* menggunakan *Feacture Selection Random Forest* dan *Particle Swarm Optimization (PSO)* Pada Prediksi Penyakit Jantung Bagi Penderita Diabetes, Berkaitan dengan hal tersebut terdapat permasalahan yang peneliti angkat yakni :

- 1) Bagaimana perbandingan metode Klasifikasi *Logistic Regession* Dan *Naive Bayes* yang di optimasi menggunakan *feature selection Random Forest* Dan *Particle Swarm Optimization (PSO)* pada prediksi penyakit jantung

- 2) Apakah Implementasi *feature selection Random Forest* dapat menghasilkan tingkat akurasi yang lebih baik dibandingkan tingkat akurasi pada *Particle Swarm Optimization* (PSO) ?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dicantumkan mengenai Perbandingan Klasifikasi *Logistic Regression* dan *Naive Bayes* menggunakan *Feature Selection Random Forest* dan *Particle Swarm Optimization* (PSO) Pada Prediksi Penyakit Jantung Bagi Penderita Diabetes, berkaitan dengan hal tersebut berikut tujuan yang peneliti angkat yakni :

- 1) Untuk mendapatkan hasil perbandingan metode Klasifikasi *Logistic Regression* Dan *Naive Bayes* yang di optimasi menggunakan *feature selection Random Forest* Dan *Particle Swarm Optimization* (PSO) pada prediksi penyakit jantung
- 2) Untuk mengetahui Apakah Implementasi *feature selection Random Forest* dapat menghasilkan tingkat akurasi yang lebih baik dibandingkan tingkat akurasi pada *Particle Swarm Optimization* (PSO)

1.4 Batasan Penelitian

Mendapatkan hasil *output* berupa cara Perbandingan Klasifikasi *Logistic Regression* dan *Naive Bayes* menggunakan *Feature Selection Random Forest* dan *Particle Swarm Optimization* (PSO) Pada Prediksi Pengidap Penyakit Jantung Bagi Penderita Diabetes

- 1) Dataset yang akan digunakan merupakan dataset yang di dapat melalui *Kaggle*
- 2) Bahasa pemrograman yang akan digunakan nanti adalah bahasa pemrograman *python*
- 3) Metode klasifikasi yang digunakan nanti adalah Klasifikasi *Logistic Regression* dan *Naive Bayes*

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah untuk mendapatkan nilai Perbandingan Klasifikasi *Logistic Regression* dan *Naive Bayes* menggunakan *Feature Selection Random Forest* dan *Particle Swarm Optimization* (PSO) Pada Prediksi Penyakit Jantung , dengan tujuan untuk membantu dalam mempercepat prediksi suatu penyakit

