

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan *data mining* sangatlah berkembang dengan sangat cepat (Pramanik et al., 2020). *Data mining* adalah proses ekstraksi untuk menemukan *pattern* dan menggunakan informasi tersebut untuk membuat model yang dapat diprediksi (Durairaj & Ranjani, 2013). Banyak sekali kegunaan *data mining* yang bisa kita dapatkan, seperti memprediksi tren pada sebuah bisnis, sebagai pembuat keputusan (*decision-making*) yang kritis serta mengidentifikasi pola-pola yang belum pernah ditemukan sebelumnya. Area *data mining* mencakup seperti klasifikasi, prediksi, estimasi, analisis, *clustering*, deskripsi dan asosiasi (Koh & Tan, 2005).

Prediksi yaitu prakiraan sebuah nilai yang belum diketahui dan untuk masa yang akan datang (*Arti kata prediksi - Kamus Besar Bahasa Indonesia (KBBI) Online, 2020*). Prediksi dalam *data mining* sudah digunakan dalam banyak bidang seperti ekonomi untuk *fraud detection*, dibidang pendidikan serta di area kesehatan. Prediksi di area kesehatan (*health-care*) sudah banyak dilakukan seperti untuk mendeteksi penyakit-penyakit dan berguna untuk membantu tenaga medis untuk memberikan tindakan medis. Selain itu, jika suatu penyakit terlambat terdeteksi maka biaya rumah sakit yang dikeluarkan juga akan lebih besar pula (Gamadarenda & Waspada, 2020).

Penyakit yang sering terjadi pada salah satu organ tubuh manusia adalah organ ginjal. Ginjal merupakan salah satu organ yang paling penting di tubuh

manusia, fungsi ginjal yaitu menjaga komposisi darah dengan mencegah limbah yang menumpuk dan berfungsi untuk mengendalikan keseimbangan cairan dalam tubuh (*Your Kidneys & How They Work / NIDDK*, 2020). Penyakit Ginjal Kronis atau biasa disingkat dengan PGK adalah penyakit yang mengenai organ ginjal yang dapat disebabkan oleh berbagai kondisi seperti diabetes dan tekanan darah yang tinggi (*Stages of Chronic Kidney Disease (CKD) - American Kidney Fund (AKF)*, n.d.). Jika dibiarkan dalam jangka panjang, maka akan menyebabkan kerusakan pada ginjal dan fungsi ginjal juga akan perlahan menurun sehingga menjadi gagal ginjal (Gamadarenda & Waspada, 2020).

PGK sering disebut dengan “*silent disease*” karena orang yang mengidap penyakit ginjal kronis terlihat sama seperti orang biasa sehingga kita perlu memprediksi tingkatan pada pasien yang terkena penyakit ginjal kronis sehingga dapat mencegah menjadi semakin parah (*Kidney Disease / Lab Tests Online*, 2020). Penyakit ginjal kronis rawan terdeteksi saat pasiennya sudah ditahap yang sangat parah sehingga prediksi tingkatan pada PGK dapat mendukung untuk penentuan jenis perawatan yang perlu dilakukan (Rady & Anwar, 2019).

Mengacu pada *website (Stages of Chronic Kidney Disease (CKD) - American Kidney Fund (AKF)*, n.d.), terdapat 5 tahapan dalam kerusakan ginjal. Mulai dari kerusakan ginjal yang ringan pada tahapan pertama hingga kerusakan parah sampai gagal ginjal ditahapan paling terakhir. Pada tahap awal penyakit ginjal, ginjal masih dapat menyaring limbah dari darah tetapi pada tahap selanjutnya, ginjal harus bekerja lebih keras untuk membuang limbah dan mungkin akan sampai berhenti bekerja sama sekali. Cara dokter mengukur seberapa baik

ginjal menyaring limbah dari darah Anda adalah dengan *estimated glomerular filtration rate* (eGFR) yang merupakan angka yang didasarkan pada tes darah untuk kreatinin (Welch, 2013).

Oleh maka dari itu, PGK dapat dicegah, ditanggulangi dan kemungkinan mendapatkan terapi yang efektif akan lebih besar jika sudah diketahui sejak awal. Ketika PGK terlambat terdeteksi maka peluang sembuh juga akan semakin kecil sehingga dibutuhkan tenaga medis yang lebih ahli dalam penanganannya serta memerlukan biaya yang lebih besar dalam pengobatannya dengan peluang penyembuhan yang semakin kecil. Selain itu, atribut yang diperlukan untuk melakukan tes kesehatan untuk PGK juga cukup banyak sehingga membutuhkan biaya yang lumayan mahal (Gamadarenda & Waspada, 2020). Sudah banyak penelitian yang dilakukan untuk memprediksi PGK menggunakan *data mining*. Menurut (Gopika & Vanitha, 2017) ada sebanyak 7 penelitian yang membahas tentang penelitian dengan topik yang sama tetapi masih sedikit penelitian yang memprediksi PGK menggunakan salah satu teknik pre-processing yaitu *feature selection*. *Feature selection* atau yang biasa disebut dengan *variable selection* adalah teknik *preprocessing* data yang banyak digunakan dalam *data mining* yang pada dasarnya digunakan untuk reduksi data dengan menghilangkan atribut yang tidak signifikan dan berlebihan dari *dataset* manapun (Jain & Singh, 2018). Penggunaan *feature selection* dalam penelitian ini juga berguna untuk mengurangi atribut-atribut sehingga dapat mengurangi biaya tes laboratorium yang tidak diperlukan sehingga dapat menghemat biaya pengobatan untuk PGK (Gamadarenda & Waspada, 2020).

Berikut merupakan beberapa contoh penelitian prediksi PGK dengan implementasinya yang menggunakan berbagai macam algoritma. Yang pertama menggunakan algoritma K-NN (*k-nearest neighbour*) yang hanya memprediksi PGK mendapatkan akurasi sebesar 99.25% (Gamadarena & Waspada, 2020). Yang kedua, ada penelitian untuk memprediksi PGK menggunakan teknik *5-fold cross-validation* yang menggunakan algoritma *logistic regression* dengan hasil akurasi sebesar 96.55%, *decision tree* dengan hasil akurasi sebesar 94.8% dan K-NN dengan hasil akurasi sebesar 98.1% (Charleonnann et al., 2017). Yang ketiga, ada penelitian yang menyarankan untuk memprediksi PGK dengan algoritma yang bersifat klasifikasi yaitu *Decision tree* dan K-NN serta algoritma yang bersifat untuk memprediksi sesuatu seperti *Logistic regression* (Thiyagaraj & Suseendran, 2019). Selan itu, ada juga penelitian yang menyatakan bahwa menggunakan teknik *feature selection* dapat meningkatkan akurasi algoritma dari prediksi suatu penyakit kronis (Jain & Singh, 2018) dan ada juga penelitian yang menyatakan bahwa salah satu metode dari *feature selection* yaitu *forward selection* dapat meningkatkan akurasi dari prediksi penyakit jantung, yang pada awalnya akurasi algoritma K-NN tersebut hanya sebesar 73.44% setelah menggunakan metode *forward selection* akurasi meningkat sebesar 5.22% dan menjadi 78.66% (Zeniarta et al., 2020). Berikut merupakan tabel 1.1 yang merupakan perbandingan antara penelitian sebelumnya.

Tabel 1.1. Tabel Perbandingan Penelitian Sebelumnya

<i>Author/ year</i>	Algoritma						Feature Selection	
	DT	K-NN	LR	RF	ANN	SVM	<i>Forward Selection</i>	<i>Backward Elimination</i>

Tabel 1.1. Tabel Perbandingan Penelitian Sebelumnya

(Gamadar enda & Waspada, 2020)	-	99.2%	-	-	-	-		
(Charleon nan et al., 2017)	94.8%	98.1%	96.5%	-	-	-		-
(Hermawa nti & Safriando no, 2016)	-	96%	-	-	-	-		-
(Rady & Anwar, 2019)	-	-	-	-	-	60%	-	-
(Kunwar et al., 2016)	-	99%	-	-	72.7%	-	-	-
(Arifin, 2020)	92.2%	-	-	-	-	-		-
(Rianto & Iswari, 2017)	91.5%	-	-	-	-	-	-	-
(Herliawa n et al., 2020)	-	-	-	76%	-	-	-	

Pada tabel 1.1 diatas, menunjukkan bahwa penelitian yang menggunakan metode *forward selection* dan *backward elimination* memiliki peningkatan performa akurasi yang tinggi serta penggunaan algoritma K-NN, *decision tree* serta *logistic regression* juga memberikan performa akurasi dan memiliki tingkat akurasi yang lebih dari 90% daripada akurasi algoritma lain yang cenderung dibawah 77%. Sehingga telah terpilihnya algoritma K-NN, *decision tree* serta *logistic regression* sebagai algoritma yang akan digunakan untuk memprediksi PGK dan juga berfokus untuk melengkapi penelitian sebelumnya menggunakan teknik *feature selection* dan akan dibandingkan hasil akurasi. Perbandingan hasil akurasi ini dilakukan

untuk mengetahui teknik *feature selection* yang mana dan cocok untuk masing-masing algoritma tersebut.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan diatas maka peneliti dapat menjadikan rumusan masalah yang akan diselesaikan pada penelitian ini untuk melakukan prediksi terhadap gejala PGK (Penyakit Ginjal Kronis) menggunakan algoritma *data mining* dengan rumusan masalah sebagai berikut:

- a. Bagaimana membuat model algoritma K-NN, *Decision tree* dan *Logistic regression* menggunakan teknik *feature selection* yaitu metode *forward selection* dan *backward elimination* dalam memprediksi PGK?
- b. Bagaimana hasil akurasi algoritma K-NN, *Decision tree* dan *Logistic regression* menggunakan teknik *feature selection* yaitu metode *forward selection* dan *backward elimination* dalam memprediksi PGK ?

1.3. Batasan Masalah

Berikut merupakan batasan-batasan masalah yang akan dibahas dalam penelitian ini:

- a. Penelitian ini hanya membahas hasil perbandingan akurasi prediksi pada PGK pada algoritma K-NN, *Decision tree* dan *Logistic regression* serta teknik *feature selection* yang digunakan .
- b. Penelitian ini hanya membuat model prediksi pada PGK menggunakan algoritma K-NN, *Decision tree* dan *Logistic regression* .
- c. Penelitian ini tidak membahas tahap *deployment* pada tahapan *CRISP-DM*.

1.4. Tujuan Penelitian

Penelitian ini bertujuan adalah untuk melakukan perbandingan performa akurasi prediksi dari algoritma *K-NN*, *Decision tree* dan *Logistic Regression* yang menggunakan teknik *feature selection*. Data yang digunakan pada penelitian ini adalah data set mengenai penyakit ginjal kronis yang dikumpulkan di Apollo Hospitals di India yang didapatkan secara cuma-cuma dari *website* The UCI Learning.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Sebagai bahan referensi atau acuan untuk selanjutnya dapat digunakan sebagai rancang bangun aplikasi dengan algoritma *K-NN*, *Decision tree* dan *Logistic regression* menggunakan teknik *feature selection* serta dapat digunakan dan diterapkan sebagai aplikasi prediksi PGK.
- b. menambah wawasan mengenai penggunaan beberapa algoritma pada proses *data mining* dalam hal melakukan prediksi PGK.