

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Salah satu organ tubuh yang cukup vital namun sering diabaikan oleh banyak orang adalah ginjal. Ginjal memiliki fungsi penting dalam tubuh manusia sebagai penyaring darah dari produk hasil metabolisme sehingga apabila kesehatan ginjal terganggu dan tidak dapat melakukan fungsinya tersebut dengan baik, maka akan besar kemungkinan manusia tersebut mengalami Penyakit Ginjal Kronis (PGK) [1].

Saat ini, PGK secara global mengalami peningkatan dan pada tahun 2010 menjadi penyebab kematian di peringkat ke-18 [1]. Masalah kesehatan ginjal ini terus berkembang dan terbukti dengan hasil laporan *World Health Organization* (WHO) pada tahun 2016 yang menyatakan bahwa penyakit ginjal merupakan salah satu dari sepuluh penyebab utama kematian di negara berpenghasilan tinggi di dunia [2]. Di Indonesia sendiri, diperkirakan hampir mencapai 70.000 pasien yang mengalami PGK [3]. Perawatan penyakit ginjal di Indonesia menempati urutan kedua dalam pembiayaan terbesar dari BPJS Kesehatan di bawah penyakit jantung [4]. Sehingga, masih diperlukan teknik perawatan yang lebih mutakhir dibandingkan dengan yang ada saat ini untuk dapat membantu perawatan pasien dengan lebih efektif [5].

*Precision medicine* adalah salah satu metode pengobatan modern yang bertujuan untuk mengembangkan strategi pencegahan dan pengobatan berdasarkan lingkungan, gaya hidup, genetik, serta keragaman fenotipe suatu individu [6]. Salah satu strategi yang dimaksud adalah dengan melakukan analisis hubungan antara genotipe (kode genetik) dan fenotipe (karakteristik yang dapat diamati) dari suatu individu dengan suatu penyakit tertentu [7]. *Precision medicine* dapat digunakan oleh tenaga medis untuk mengidentifikasi tingkat stadium penyakit ginjal kronis pasien dengan menilai fungsi kerja ginjal. Namun, pengaplikasian penelitian langsung pada manusia terhalang oleh Etika, Legalitas, dan Implikasi sosial (ELSI). Manusia dan hewan mamalia memiliki sistem dengan kompleksitas yang hampir sama. Salah satu contoh hewannya adalah tikus yang paling umum dijadikan model penelitian [8]. Ilmuwan memodifikasi genetik pada tikus dengan memanipulasi genom tikus dan fungsi gen sehingga ginjal tikus dapat digunakan untuk pembelajaran biologi ginjal. Ginjal tikus pun secara morfologis mirip dengan

ginjal manusia [9].

Adapun bentuk variasi genetik yang paling umum ditemukan dalam tubuh manusia adalah *Single Nucleotide Polymorphism* (SNP) [10]. SNP adalah perbedaan susunan basa nukleotida tunggal pada genom suatu individu yang menyebabkan adanya variasi genetik dalam suatu populasi [11]. Selain itu, SNP juga dapat digunakan sebagai *biomarker*. Terdapat penelitian yang telah menunjukkan bahwa SNP dalam urutan pengaturan ACE2 dan TMPRSS2 dapat memengaruhi ekspresi dan stabilitas mRNA. SNP juga berperan bersama dengan gen lain dalam menentukan metabolisme suatu obat [12]. Berbagai metode digunakan untuk mencari asosiasi antara genom SNP dengan fenotipe penyakit tertentu karena melalui asosiasi tersebut dapat menghasilkan informasi SNP mana saja yang secara signifikan memengaruhi fenotipe yang sedang diamati [13].

Untuk mengidentifikasi analisis asosiasi PGK, diperlukan pengumpulan data SNP berdasarkan variasi genetik penderitanya. Analisis asosiasi menggunakan data SNP memiliki tantangan sendiri dikarenakan dimensi data yang besar dan jumlah sampel yang sedikit, sehingga rentan akan masalah *curse of dimensionality* [14]. Untuk mengolah data dengan banyak dimensi, diperlukan pengurangan jumlah dimensi tanpa kehilangan banyak informasi penting. Salah satu metode yaitu dengan seleksi fitur atau ekstraksi fitur atau keduanya [15].

Sebelumnya, telah dilakukan beberapa penelitian sejenis terkait asosiasi SNP dengan fenotipe untuk penyakit diabetes melitus tipe 2 dengan berbagai metode seperti *Random Forest* [16], *Stepwise Regression* [17], *Support Vector Regression* (SVR) dan *Genetic Algorithm* (GA) [18], *Genetic Algorithm* (GA) dan *CatBoost* [19], *Gradient Boosting* [20], serta *Elastic Net* [21]. *Stepwise Regression* merupakan metode pemilihan model dengan seleksi fitur bertahap, namun jumlah variabel prediktor yang dimasukkan dalam model harus dipilih secara manual, sehingga kesalahan saat pemilihan variabel dapat terjadi. Selain itu, metode *Stepwise Regression* tidak selalu menghasilkan model terbaik dengan multikolinearitas minimal [22]. Metode SVR adalah teori yang diadaptasi dari teori *machine learning* yang ditujukan guna memecahkan masalah klasifikasi, namun tidak mampu menangani data dengan jumlah dimensi yang sangat besar [23] [24].

*Elastic Net* telah dipertimbangkan dan dipilih sebagai metode yang digunakan dalam penelitian ini karena merupakan metode yang jumlah parameternya tidak terbatas oleh jumlah sampel data sebagaimana regresi linear konvensional [21]. Metode *Elastic Net* mengadaptasi *Ridge Regression* dalam memprediksi dan *Lasso Regression* dalam menyeleksi fitur [25].

Berdasarkan uraian di atas, maka penelitian ini akan menggunakan metode *Elastic Net* dalam mengasosiasikan SNP dengan fenotipe PGK. Sejauh ini belum ada penelitian terkait mengenai topik yang dibahas pada penelitian ini sehingga tidak diketahui bagaimana hasil penelitian dengan metode *Elastic Net*. Maka, diharapkan metode *Elastic Net* dapat membantu untuk mengasosiasikan SNP dengan fenotipe PGK dengan nilai *Mean Absolute Error* (MAE) yang rendah dan koefisien determinasi ( $R^2$ ) yang baik.

## 1.2 Rumusan Masalah

Adapun berikut merupakan rumusan masalah penelitian ini berdasarkan latar belakang yang telah dijelaskan:

1. Bagaimana mencari asosiasi SNP terhadap fenotipe pada PGK menggunakan metode Elastic Net?
2. Bagaimana nilai koefisien determinasi yang didapatkan dari metode Elastic Net dalam memprediksi nilai fenotipe PGK dengan menggunakan genotipe SNP?
3. Bagaimana nilai mean absolute error yang didapatkan dari metode Elastic Net dalam mengasosiasikan SNP dan fenotipe pada PGK?

## 1.3 Batasan Permasalahan

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan merupakan dataset SNP tikus sebanyak 31 *strain* (kelompok organisme yang terdapat dalam spesies yang sama) tikus.
2. Hanya memperhatikan fenotipe *blood albumin amount*, tanpa melibatkan faktor interaksi antar SNP dan faktor lingkungan.

## 1.4 Tujuan Penelitian

Berikut merupakan tujuan dari pelaksanaan penelitian ini:

1. Menghasilkan model Elastic Net yang mampu memprediksi nilai fenotipe PGK.

2. Mengetahui nilai koefisien determinasi dari metode Elastic Net dalam memprediksi nilai fenotipe PGK dengan menggunakan genotipe SNP.
3. Mengetahui nilai mean absolute error dari metode Elastic Net dalam mengasosiasikan SNP dan fenotipe pada PGK.

### 1.5 Manfaat Penelitian

Dengan dilaksanakannya penelitian ini, manfaat yang didapatkan sebagai berikut:

1. Peneliti dapat mengetahui implementasi penggunaan metode *Elastic Net* dalam dunia kesehatan, khususnya untuk *precision medicine*.
2. Penelitian ini dapat menjadi referensi salah satu contoh pendekatan metode *machine learning* untuk kebutuhan *precision medicine* di masa mendatang.
3. Mengetahui performa metode *Elastic Net* dalam mengasosiasikan SNP dengan fenotipe PGK.

### 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN  
Pada bab pendahuluan berisikan latar belakang masalah, rumusan masalah, batasan permasalahan serta tujuan dan manfaat penelitian ini.
- Bab 2 LANDASAN TEORI  
Dalam bab landasan teori berisikan teori-teori yang telah dikumpulkan dari berbagai sumber, meliputi teori tentang Penyakit Ginjal Kronis (PGK), *Single Nucleotide Polymorphism* (SNP), hingga *Elastic Net*.
- Bab 3 METODOLOGI PENELITIAN  
Bab metodologi dan penelitian memaparkan tahapan serta metode yang digunakan dalam penelitian ini, seperti pengumpulan data, pra-proses data, pemodelan dengan Elastic Net, pengujian dan evaluasi, dan validasi hasil seleksi SNP.

- Bab 4 HASIL DAN DISKUSI

Pada bab ini akan berisikan mengenai pembahasan detail keseluruhan hasil penelitian, seperti koding dan tabel, yang telah dilaksanakan dengan menggunakan metode Elastic Net.

- Bab 5 KESIMPULAN DAN SARAN

Bab terakhir akan berisikan kesimpulan berupa tujuan dari seluruh pelaksanaan penelitian, disertai dengan beberapa informasi tambahan yang berupa saran untuk penelitian terkait berikutnya.

