

**ASOSIASI SINGLE NUCLEOTIDE POLYMORPHISM DAN FENOTIPE  
PADA PENYAKIT GINJAL KRONIS MENGGUNAKAN METODE  
ELASTIC NET**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Maureen Zerlina Oktaviani**

**00000036850**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

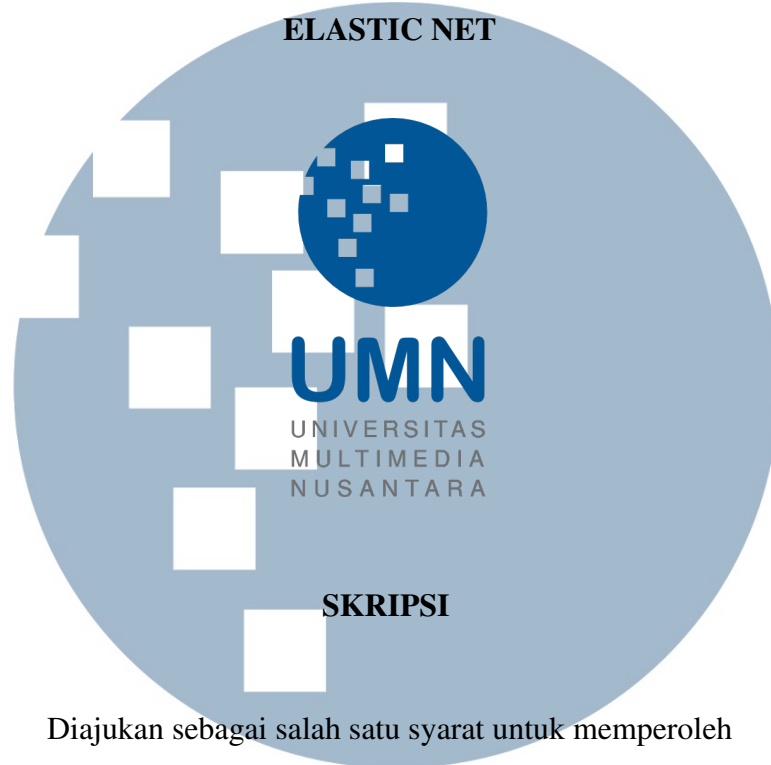
**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

**ASOSIASI SINGLE NUCLEOTIDE POLYMORPHISM DAN FENOTIPE  
PADA PENYAKIT GINJAL KRONIS MENGGUNAKAN METODE  
ELASTIC NET**



**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh  
Gelar Sarjana Komputer (S.Kom.)

**Maureen Zerlina Oktaviani**

**00000036850**

**UMN**

**UNIVERSITAS**

**MULTIMEDIA**

**NUSANTARA**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA**

**TANGERANG**

**2023**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Maureen Zerlina Oktaviani  
Nomor Induk Mahasiswa : 00000036850  
Program Studi : Informatika

Skripsi dengan judul:

**Asosiasi Single Nucleotide Polymorphism dan Fenotipe Pada Penyakit Ginjal Kronis Menggunakan Metode Elastic Net**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 6 Juni 2023



(Maureen Zerlina Oktaviani)

## HALAMAN PENGESAHAN

Skripsi dengan judul

### **ASOSIASI SINGLE NUCLEOTIDE POLYMORPHISM DAN FENOTIPE PADA PENYAKIT GINJAL KRONIS MENGGUNAKAN METODE ELASTIC NET**

oleh

Nama : Maureen Zerlina Oktaviani  
NIM : 00000036850  
Program Studi : Informatika  
Fakultas : Fakultas Teknik dan Informatika

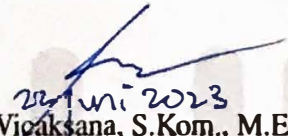
Telah diujikan pada hari Selasa, 13 Juni 2023

Pukul 08.00 s/s 10.00 dan dinyatakan

**LULUS**

Dengan susunan penguji sebagai berikut

Ketua Sidang

  
(Arya Wijaksana, S.Kom., M.Eng.Sc.,  
OCA, CEH)  
NIDN: 0315109103

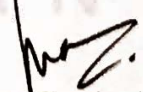
Penguji

  
(Marlinda Vasty Overbeek, M.Kom.)  
NIDN: 0818038501

Pembimbing

  
(Angga Aditya Perdana, M.Kom.)  
NIDN: 0407128901

Ketua Program Studi Informatika,

  
(Marlinda Vasty Overbeek, M.Kom.)  
NIDN: 0818038501

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

Nama : Maureen Zerlina Oktaviani  
NIM : 00000036850  
Program Studi : Informatika  
Fakultas : Teknik dan Informatika  
Jenis Karya : Skripsi

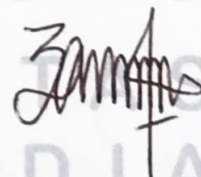
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

### ASOSIASI SINGLE NUCLEOTIDE POLYMORPHISM DAN FENOTIPE PADA PENYAKIT GINJAL KRONIS MENGGUNAKAN METODE ELASTIC NET

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 6 Juni 2023

Yang menyatakan



Maureen Zerlina Oktaviani

**Halaman Persembahan / Motto**

”CAST ALL YOUR ANXIETY ON HIM BECAUSE HE CARES FOR YOU ”

1 PETER 5:7 (NIV)

UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

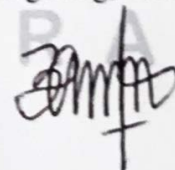
## KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: *Asosiasi Single Nucleotide Polymorphism dan Fenotipe Pada Penyakit Ginjal Kronis Menggunakan Metode Elastic Net* dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Marlinda Vasty Overbeek, M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Angga Aditya Permana, M.Kom., sebagai Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.
6. Ananda S, Bintang R, Jacinda Gaby dan teman-teman yang lain yang telah memberikan saya kepercayaan untuk menyelesaikan pendidikan sarjana ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 6 Juni 2023



Maureen Zerlina Oktaviani

**ASOSIASI SINGLE NUCLEOTIDE POLYMORPHISM DAN FENOTIPE  
PADA PENYAKIT GINJAL KRONIS MENGGUNAKAN METODE  
ELASTIC NET**

Maureen Zerlina Oktaviani

**ABSTRAK**

Ginjal merupakan salah satu organ penting dalam tubuh manusia yang berperan besar dalam penyaringan darah pada sistem metabolisme tubuh sehingga apabila rusak dapat mengakibatkan Penyakit Ginjal Kronis (PGK). Untuk meminimalisir masalah kesehatan PGK ini, dibutuhkan suatu pengobatan yang lebih mutakhir dengan pengobatan yang sudah ada saat ini, salah satunya adalah dengan *precision medicine* yang dikembangkan berdasarkan lingkungan, gaya hidup, genetik, serta keragaman fenotipe suatu individu. *Single Nucleotide Polymorphism* (SNP) dapat digunakan sebagai bentuk variasi genetik yang dapat diamati kaitannya dengan fenotipe PGK. Untuk menganalisis asosiasi antara SNP dan fenotipe pada PGK diperlukan bantuan algoritma *machine learning* yang dapat mengolah data dengan dimensi yang banyak, salah satunya dengan menggunakan metode *Elastic Net*. Pada penelitian ini, akan digunakan metode *Elastic Net* untuk mengasosiasikan SNP dengan fenotipe untuk PGK karena metode *Elastic Net* mampu mengatasi masalah *curse of dimensionality*. Berdasarkan hasil yang didapatkan pada penelitian ini, menunjukkan metode *Elastic Net* dapat menyeleksi 88 SNP signifikan dengan nilai MAE yang rendah yaitu 0.755 dan koefisien determinasi ( $R^2$ ) yang baik yaitu 0.999. Adapun SNP signifikan yang terpilih dengan model *Elastic Net* antara lain Rhd, F2, Colla1, Nos3, F5, Gypc, F8, F10, Bcam, Rbp4, Plg, Thpo, Myh9, Akt1, Fgb, dan F7.

**Kata kunci:** Asosiasi, *Elastic Net*, fenotipe, Penyakit Ginjal Kronis (PGK), *Single Nucleotide Polymorphism* (SNP)

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



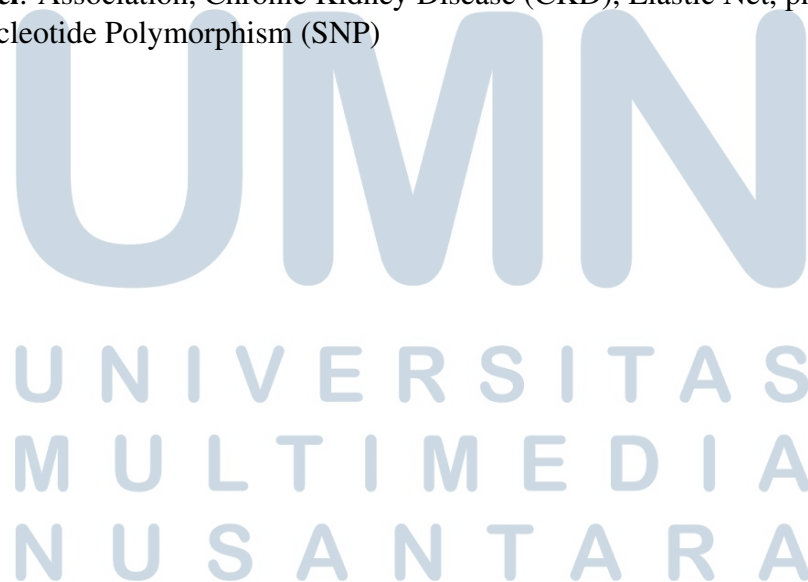
# Association of Single Nucleotide Polymorphism and Phenotype in Chronic Kidney Disease using Elastic Net Method

Maureen Zerlina Oktaviani

## ABSTRACT

Kidney is one of the most important organ in human body that has a big role in blood filtering of body's metabolic system, so if it is failing can cause Chronic Kidney Disease (CKD). To minimize this CKD health problem, it needs an up to date medication, one of them is with precision medicine that developed by the environment, lifestyle, genetics, also genotype variety of an individual. Single Nucleotide Polymorphism (SNP) can be used as a genetic variation that can be observed with PGK phenotype. To analyze the associations of SNP and phenotype on PGK, it needs assistance from a machine learning algorithm that can process data with many dimensions, one of them is the Elastic Net method. In this research, Elastic Net will be used as a method to associate SNP with phenotype for CKD because Elastic Net method can overcome curse of dimensionality. Based on the result which obtained in this research, it shows that the Elastic Net method can select 88 significant SNP with a low MAE score that is 0.755 with a good explained variance ( $R^2$ ) which is 0.999. Also, there are significant SNP that were chosen with Elastic Net method such as Rhd, F2, Col1a1, Nos3, F5, Gypc, F8, F10, Bcam, Rbp4, Plg, Thpo, Myh9, Akt1, Fgb, and F7.

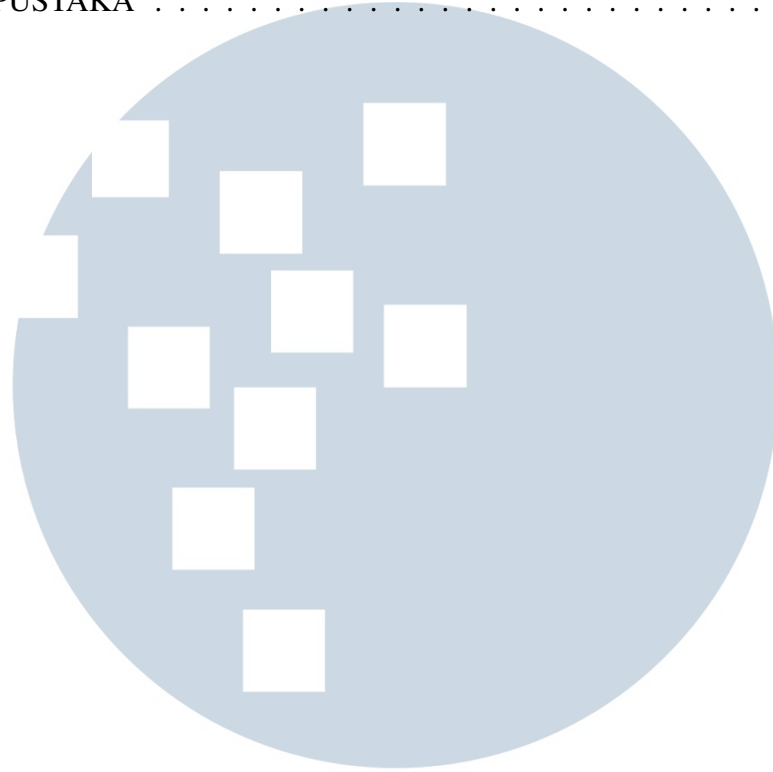
**Kata kunci:** Association, Chronic Kidney Disease (CKD), Elastic Net, phenotype, Single Nucleotide Polymorphism (SNP)



## DAFTAR ISI

HALAMAN JUDUL . . . . .	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT . . . . .	ii
HALAMAN PENGESAHAN . . . . .	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH . . . . .	iv
HALAMAN PERSEMBAHAN/MOTO . . . . .	v
KATA PENGANTAR . . . . .	vi
ABSTRAK . . . . .	vii
ABSTRACT . . . . .	viii
DAFTAR ISI . . . . .	ix
DAFTAR GAMBAR . . . . .	xi
DAFTAR TABEL . . . . .	xii
DAFTAR KODE . . . . .	xiii
DAFTAR LAMPIRAN . . . . .	xiv
BAB 1 PENDAHULUAN . . . . .	1
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	3
1.3 Batasan Permasalahan . . . . .	3
1.4 Tujuan Penelitian . . . . .	3
1.5 Manfaat Penelitian . . . . .	4
1.6 Sistematika Penulisan . . . . .	4
BAB 2 LANDASAN TEORI . . . . .	6
2.1 Precision Medicine . . . . .	6
2.2 Penyakit Ginjal Kronis (PGK) . . . . .	6
2.3 Genotipe dan Fenotipe . . . . .	7
2.4 Single Nucleotide Polymorphism (SNP) . . . . .	7
2.5 Elastic Net . . . . .	8
2.6 Mean Absolute Error . . . . .	8
2.7 Explained Variance ( $R^2$ ) . . . . .	9
BAB 3 METODOLOGI PENELITIAN . . . . .	10
3.1 Tahapan Penelitian . . . . .	10
3.2 Pengumpulan Data . . . . .	10
3.3 Praproses Data . . . . .	12
3.4 Pemodelan . . . . .	13
3.5 Pengujian dan Evaluasi Model . . . . .	14
3.6 Validasi Hasil Seleksi SNP . . . . .	15
BAB 4 HASIL DAN DISKUSI . . . . .	16
4.1 Spesifikasi Sistem . . . . .	16
4.2 Pengumpulan Data . . . . .	16
4.2.1 Data Kandidat Gen . . . . .	16
4.2.2 Data Fenotipe . . . . .	17
4.2.3 Data SNP . . . . .	18
4.3 Praproses Data . . . . .	19
4.3.1 Data Fenotipe . . . . .	19
4.3.2 Data SNP . . . . .	22
4.4 Pemodelan . . . . .	23
4.5 Pengujian dan Evaluasi Model . . . . .	26
4.6 Validasi Hasil Seleksi SNP . . . . .	27
BAB 5 SIMPULAN DAN SARAN . . . . .	31

5.1	Simpulan	31
5.2	Saran	31
DAFTAR PUSTAKA		32



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

Gambar 3.1	Alur tahapan penelitian . . . . .	10
Gambar 3.2	Tampilan <i>home website GeneCards</i> . . . . .	11
Gambar 3.3	Tampilan <i>home website The Mouse Phenome Database (MPD)</i> . . . . .	12
Gambar 3.4	Gen SNP sebelum diubah dan sesudah diubah ke dalam <i>binary string</i> . . . . .	13
Gambar 3.5	Cara kerja teknik LOOCV . . . . .	14
Gambar 4.1	Sebaran <i>relevance score</i> berdasarkan kategori . . . . .	17
Gambar 4.2	Sebaran SNP berdasarkan kromosom dan sebaran SNP berdasarkan jenis varian . . . . .	19
Gambar 4.3	SNP paling berpengaruh dengan dua arah positif dan negatif	26



## DAFTAR TABEL

Tabel 4.1	Data fenotipe terkait blood albumin amount . . . . .	18
Tabel 4.2	Data SNP dari 114 kandidat protein . . . . .	18
Tabel 4.3	Data nilai albumin untuk setiap strain yang telah distrandarisasi . . . . .	21
Tabel 4.4	Data SNP sebelum dilakukan pengkodean . . . . .	22
Tabel 4.5	Data SNP setelah dilakukan pengkodean . . . . .	23
Tabel 4.6	Konfigurasi <i>alpha</i> dan <i>ll_ratio</i> untuk pencarian <i>hyperparameter</i> . . . . .	24
Tabel 4.7	Hasil pengujian berisi perbandingan . . . . .	26
Tabel 4.8	SNP dengan nilai <i> signifikansi </i> tertinggi . . . . .	27
Tabel 4.9	Validasi SNP seleksi berdasarkan studi literatur . . . . .	28



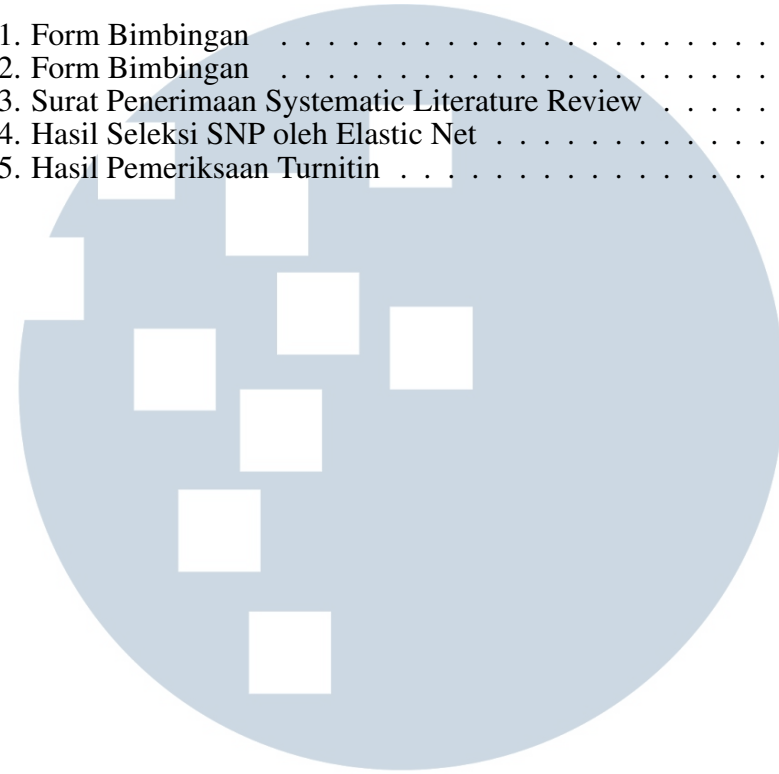
## DAFTAR KODE

4.1	<i>Import library</i> . . . . .	19
4.2	Menghilangkan <i>strain</i> yang terdapat <i>missing value</i> . . . . .	20
4.3	Standarisasi nilai rata-rata albumin . . . . .	20
4.4	<i>Transpose dataframe</i> . . . . .	22
4.5	Menghilangkan genotipe H . . . . .	22
4.6	Mengubah genotipe menjadi format <i>binary string</i> . . . . .	23
4.7	Pencarian hyperparameter terbaik dengan GridSearchCV . . . . .	23
4.8	alpha dan l1_ratio terbaik . . . . .	24
4.9	alpha dan l1_ratio terbaik . . . . .	25
4.10	Menunjukkan nilai MAE dan $R^2$ . . . . .	27



## DAFTAR LAMPIRAN

Lampiran 1. Form Bimbingan . . . . .	37
Lampiran 2. Form Bimbingan . . . . .	39
Lampiran 3. Surat Penerimaan Systematic Literature Review . . . . .	40
Lampiran 4. Hasil Seleksi SNP oleh Elastic Net . . . . .	41
Lampiran 5. Hasil Pemeriksaan Turnitin . . . . .	44



**UMMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA