

PREDIKSI INTERAKSI SENYAWA-PROTEIN UNTUK *DRUG REPURPOSING* ANTI COVID-19 MENGGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*

BELLA ANGGITA SAFITRI



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**

@Hak cipta milik IPB University

IPB University



IPB University
Bogor, Indonesia

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul Prediksi Interaksi Senyawa-Protein untuk *Drug Repurposing* Anti COVID-19 Menggunakan Metode *Convolutional Neural Network* adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apapun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2021

Bella Anggita Safitri
NIM G64170059

Hak Cipta Dilindungi Undang-undang
©Hak Cipta Milik IPB University

IPB University





ABSTRAK

BELLA ANGGITA SAFITRI. Prediksi Interaksi Senyawa-Protein untuk *Drug Repurposing* Anti COVID-19 Menggunakan Metode *Convolutional Neural Network*. Dibimbing oleh SONY HARTONO WIJAYA.

COVID-19 menyebabkan masalah kesehatan seperti demam, batuk kering, gangguan pernapasan, dan bahkan kematian. Penemuan obat secara tradisional memerlukan banyak sumber daya, sehingga pendekatan komputasional menjadi salah satu pendekatan yang efisien untuk *screening* senyawa potensial melalui prediksi interaksi senyawa-protein. Model *deep learning* yang digunakan pada penelitian ini adalah *Convolutional Neural Network* (CNN). Hasil pemodelan CNN dibandingkan dengan model *Support Vector Machine* dan *Naive Bayes* dengan representasi fitur protein *Amino Acid Composition* (AAC) dan *Dipeptide Composition* (DC). Selain itu, juga diamati pengaruh penggunaan seleksi fitur pada model. Selanjutnya, kinerja dari metode untuk memprediksi interaksi senyawa dan protein diukur dengan menggunakan akurasi, *precision*, *recall*, *F-measure*, dan AUROC. Hasil penelitian menunjukkan bahwa pemodelan dengan representasi fitur protein DC lebih baik dibandingkan dengan AAC. Pemodelan interaksi senyawa-protein menggunakan *PubChem fingerprint* sebagai representasi senyawa dan DC sebagai representasi protein pada CNN dengan seleksi fitur ANOVA menghasilkan kinerja terbaik dengan nilai akurasi sebesar 0.9475, *recall* 0.9687, *precision* 0.9679, *F-measure* 0.9683, dan AUROC 0.9751.

Kata Kunci: *convolutional neural network*, COVID-19, *drug repurposing*, *drug target interaction*

ABSTRACT

BELLA ANGGITA SAFITRI. Prediction of Compound-Protein Interaction for Drug Repurposing Anti COVID-19 Using Convolutional Neural Network Method. Supervised by SONY HARTONO WIJAYA.

COVID-19 is a disease that causes health problems. Traditional drug discovery requires many resources. Thus, the computational approach is one of the approaches that can be employed to screen potential compounds through the prediction of compound-protein interactions. The deep learning model used in this study is Convolutional Neural Network (CNN). The results of the CNN model were compared to Support Vector Machine (SVM) and Naive Bayes (NB) with representations of proteins using Amino Acid Composition (AAC) and Dipeptide Composition (DC). We also examined the effect of the feature selection approach using ANOVA. The results were evaluated in terms of accuracy, precision, recall, F-measure, and AUROC. Results showed that modeling with a representation of DC protein features was better than AAC. Prediction of compound-protein interaction modeling using PubChem fingerprint as a compound representation and DC as protein representation on CNN using ANOVA feature selection resulted in the best performance with an accuracy value of 0.9475, recall 0.9687, precision 0.9679, F-measure 0.9683, and AUROC 0.9751.

Keywords: *convolutional neural network*, COVID-19, *drug repurposing*, *drug target interaction*



@Hak cipta milik IPB University

IPB University

© Hak Cipta milik IPB, tahun 2021
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



**PREDIKSI INTERAKSI SENYAWA-PROTEIN UNTUK *DRUG
REPURPOSING* ANTI COVID-19 MENGGUNAKAN METODE
*CONVOLUTIONAL NEURAL NETWORK***

BELLA ANGGITA SAFITRI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Skripsi :
Medria Kusuma Dewi Hardhienata, SKom PhD
Dr Eng Annisa, SKom MKom

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.





Judul Skripsi: Prediksi Interaksi Senyawa-Protein untuk *Drug Repurposing* Anti COVID-19 Menggunakan Metode *Convolutional Neural Network*

Nama : Bella Anggita Safitri
NIM : G64170059

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Dr Sony Hartono Wijaya, SKom MKom



Diketahui oleh

Ketua Departemen Ilmu Komputer:

Dr Sony Hartono Wijaya, SKom MKom
NIP 198108092008121002



Tanggal Ujian:
16 Juli 2021

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Oktober 2020 sampai bulan Juli 2021 ini ialah *Drug Repurposing* Anti COVID-19, dengan judul “Prediksi Interaksi Senyawa-Protein untuk Drug Repurposing Anti COVID-19 Menggunakan *Convolutional Neural Network*”. Terima kasih penulis ucapkan kepada pihak-pihak yang telah berjasa dalam penyelesaian tugas akhir ini, antara lain:

1. Orangtua dan keluarga yang selalu hadir di hati penulis dan memberi doa, semangat, dukungan, kasih sayang, dan motivasi bagi penulis sehingga penelitian ini dapat diselesaikan.
2. Bapak Dr Sony Hartono Wijaya, SKom MKom selaku dosen pembimbing yang telah menyediakan waktu dan tenaga untuk memberikan bimbingan dan saran dari awal penelitian dilakukan hingga penelitian selesai.
3. Ibu Medria Kusuma Dewi Hardhienata, SKom PhD dan Ibu Dr Eng Annisa, SKom MKom selaku dosen penguji skripsi.
4. Muhammad Kamal Nasution, SKom MKom dan Lidya Dwi Utami yang telah membantu selama pengumpulan data.
5. Annisa Widia Astuti, Dina Fadhila, Faldi Sulistiawan dan Fathiya yang telah banyak membantu selama proses pembuatan skripsi.
6. Departemen Ilmu Komputer, dosen dan staf yang telah banyak membantu selama masa perkuliahan.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2021

Bella Anggita Safitri



DAFTAR ISI

DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	viii
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	3
Tujuan Penelitian	3
Manfaat Penelitian	3
Ruang Lingkup Penelitian	3
REVIEW PUSTAKA	4
Drug Repurposing	4
One Way ANOVA F-Test	4
Convolutional Neural Network	5
METODE	7
Tahapan Penelitian	7
Data Penelitian	7
Praproses Data	7
Membuat Model Prediksi	9
Pengujian dan Evaluasi Hasil	11
Lingkungan Pengembangan	12
HASIL DAN PEMBAHASAN	12
Praproses Data	12
Pembuatan Model Prediksi	14
Tuning parameter dan hasil model prediksi SVM	15
Pengujian dan Evaluasi hasil model prediksi SVM	15
Tuning parameter dan hasil model prediksi NB	17
Pengujian dan Evaluasi hasil model prediksi NB	17
Tuning parameter dan hasil model prediksi CNN	19
Pengujian dan Evaluasi hasil model prediksi CNN	20
Evaluasi model classifier	22
SIMPULAN DAN SARAN	25
Simpulan	25
Saran	25
DAFTAR PUSTAKA	26
LAMPIRAN	28
RIWAYAT HIDUP	30

Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



DAFTAR TABEL

1	Perhitungan ANOVA	4
2	Confusion matrix	11
3	Statistik data penelitian	12
4	Dataset penelitian	14
5	Tuning parameter SVM	15
6	Hasil Tuning parameter SVM	15
7	Hasil <i>confusion matrix</i> dan metrik SVM	16
8	Tuning parameter NB	17
9	Hasil Tuning parameter NB	17
10	Hasil confusion matrix dan metrik NB	18
11	Tuning parameter CNN	19
12	Hasil Tuning parameter CNN	20
13	Hasil confusion matrix dan metrik CNN	20
14	Perbandingan hasil evaluasi classifier dataset 1	22
15	Perbandingan hasil evaluasi classifier dataset 2	22
16	Perbandingan <i>running time classifier</i> (detik)	24

DAFTAR GAMBAR

1	Contoh arsitektur 1D CNN (Kiranyaz et al. 2019)	5
2	<i>Forward dan backpropagation</i> pada hidden CNN layer (Kiranyaz et al. 2019)	6
3	Tahapan penelitian	7
4	Tahapan praproses data	8
5	Ilustrasi pengambilan fitur senyawa	8
6	Ilustrasi pengambilan fitur protein	9
7	Ilustrasi baris dataset penggabungan <i>feature vector</i>	9
8	Ilustrasi <i>drug discovery</i> menggunakan SVM	10
9	Arsitektur CNN yang akan digunakan (Zhao et al. 2018)	11
10	Perbandingan kelas positif dan negatif setelah sampling	13
11	Scatter plot penentuan <i>k-feature</i> terbaik ANOVA	14
12	ROC SVM menggunakan dataset 1	16
13	ROC SVM menggunakan dataset 2	17
14	ROC NB menggunakan dataset 1	19
15	ROC NB menggunakan dataset 2	19
16	ROC CNN menggunakan dataset 1	21
17	ROC CNN menggunakan dataset 2	21
18	Perbandingan ROC antar model dataset 1	23
19	Perbandingan ROC antar model dataset 2	23

DAFTAR LAMPIRAN

Implementasi seleksi fitur	29
Pemodelan <i>Convolutional Neural Network</i>	29