

**PENGEMBANGAN MODEL FEEDFORWARD NEURAL
NETWORK UNTUK PREDIKSI MUTASI DATA SEKUENSIAL
SITUS CLEAVAGE SARS-COV-2**

SKRIPSI

**OLEH
NATASYAH ASVIANA
NIM. 210601110077**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**PENGEMBANGAN MODEL FEEDFORWARD NEURAL
NETWORK UNTUK PREDIKSI MUTASI DATA SEKUENSIAL
SITUS CLEAVAGE SARS-COV-2**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**OLEH
Natasyah Asviana
NIM. 210601110077**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**PENGEMBANGAN MODEL FEEDFORWARD NEURAL
NETWORK UNTUK PREDIKSI MUTASI DATA SEKUENSIAL
SITUS CLEAVAGE SARS-COV-2**

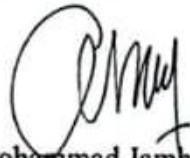
SKRIPSI

**Oleh
Natasyah Asviana
NIM.210601110077**

Telah Disetujui untuk Diuji

Malang, 12 Desember 2025

Dosen Pembimbing I



Dr. Mohammad Jamhuri, M.Si.
NIP.19810502 2005011004

Dosen Pembimbing II



Erna Herawati, M.Pd.
NIPPPK. 19760723 2023212006



Mengetahui,
Ketua Program Studi Matematika

Dr. Fachrud Rozi, M.Si
NIP. 19800527 2008011012

**PENGEMBANGAN MODEL FEEDFORWARD NEURAL
NETWORK UNTUK PREDIKSI MUTASI DATA SEKUENSIAL
SITUS CLEAVAGE SARS-COV-2**

SKRIPSI

OLEH:
Natasyah Asviana
NIM. 210601110077

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)

Tanggal 23 Desember 2025

Ketua Penguji : Ari Kusumastuti, M.Pd, M.Si.

Anggota Penguji 1 : Juhari, M.Si

Anggota Penguji 2 : Dr. Mohammad Jamhuri, M.Si

Anggota Penguji 3 : Erna Herawati, M.Pd

Mengetahui,
Ketua Program Studi Matematika

Dr. Fachrud Rozi, M.Si
NIM. 19800527 2008011012



PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Natasyah Asviana

NIM : 210601110077

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Pengembangan Model Feedforward Neural Network
Untuk Prediksi Mutasi Situs Cleavage SARS-CoV-2

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri. Bukan merupakan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada dftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 16 Desember 2025
Yang Membuat Pernyataan



Natasyah Asviana
NIM.210601110077

MOTO

"Mustahil Allah membawamu sejauh ini hanya untuk gagal. Karena di balik ujianmu yang berat, ada kesuksesan yang kelak Allah berikan"

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."
(QS. Al-Baqarah: 286)

PERSEMBAHAN

Dengan penuh rasa syukur, skripsi ini saya persembahkan kepada:

Ayah Asrukin dan Ibu Roi Khanah

Kepada Ayah, yang keteguhan hatinya terus menuntunku melangkah, dan kepada Ibuku yang telah kembali kepada Sang Pemilik Hidup namun tetap hidup dalam doa dan ingatan terbaikku. Terima kasih atas doa-doa yang kalian tanamkan sejak aku belum mengenal dunia, atas pengorbanan yang tak pernah meminta balas, dan atas restu yang menjadi jembatan hingga aku sampai pada titik ini. Meski Ibu tak lagi hadir secara nyata, aku percaya doanya tetap mengiringi dari tempat yang lebih damai, sementara Ayah terus menjadi penyangga kuat di bumi. Semoga karya sederhana ini menjadi bisik bakti untuk Ibu di surga dan menjadi kebanggaan lembut bagi Ayah yang selalu mendoakan di sini.

Kakak Devi Novita Ariani, Rico Ferdiansyah, dan Adik Sintya Ramadani

Untuk saudara-saudaraku yang selalu memberi dukungan dengan cara masing-masing. Terima kasih atas bantuan, perhatian, dan kebersamaan yang membuat perjalanan ini terasa jauh lebih ringan.

Diriku sendiri, Natasyah asviana

Terima kasih untuk diriku yang tetap bertahan meski lelah, terus melangkah meski ragu, dan tetap memberi yang terbaik walau sering tertatih. Terima kasih telah percaya bahwa setiap proses akan selesai pada waktunya.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahirabbil'alamiin. Segala puji bagi Allah Swt. atas segala rahmat-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam tercurah kepada baginda Nabi Muhammad SAW. yang membawa manusia menuju jalan kebenaran agama Islam dan kita harapkan syafaatnya kelak.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika pada Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Ibrahim Malang. Dalam proses penyusunannya, penulis memperoleh doa, arahan, dan bimbingan dari berbagai pihak, terutama kedua orang tua. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. H. Agus Mulyono, M. Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrur Rozi, M.Si, selaku Ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Mohammad Jamhuri, M.Si., selaku Dosen Pembimbing I, atas kesabaran, motivasi, serta ilmu yang beliau berikan sepanjang proses penyusunan proposal hingga penulisan skripsi ini.
5. Erna Herawati, M.Pd., selaku Dosen Pembimbing II, atas bantuan, tuntunan, dan nasihat berharga yang beliau curahkan selama proses penulisan skripsi ini.
6. Ari Kusumastuti, M.Si., M.pd., selaku Ketua Penguji dalam ujian skripsi ini, atas masukan, koreksi, dan arahan yang berperan penting dalam penyempurnaan penelitian ini.
7. Juhari, M.Si., selaku Anggota Penguji I, yang dengan penuh perhatian telah memberikan saran, masukan, dan arahan yang berharga bagi penyempurnaan skripsi ini.
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim, atas ilmu dan pengajaran yang telah diberikan selama masa studi.

9. Orang tua dan seluruh keluarga yang telah bersusah payah dalam memenuhi segala kebutuhan, arahan, pengorbanan serta dengan iringan do'anya sehingga penulis dapat menyelesaikan proposal skripsi ini.
10. Teman-teman TEOREMA yang telah menjadi rekan selama masa perkuliahan. Terima kasih atas kebersamaan dan bantuannya.
11. Serta semua pihak yang telah memberikan bantuan demi terselesainya skripsi ini.

Semoga Allah SWT. membalas segala kebaikan atas bantuan dan bimbingan yang diberikan. Semoga skripsi ini bermanfaat bagi penulis dan semua yang membacanya. Aamiin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Malang, 16 Desember 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
ABSTRAK	xv
ABSTRACT	xvi
مستخلص البحث.....	xvii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah	6
1.6 Definisi Istilah	6
BAB II KAJIAN PUSTAKA.....	9
2.1. Struktur SARS-CoV-2	9
2.2. Struktur Protein Spike SARS-CoV-2.....	11
2.3. Situs Furin <i>Cleavage</i> pada Virus SARS-CoV-2	12
2.4. Multiple Sequence Alignment	14
2.5. Prediksi Data Sekuensial Berbasis <i>Neural Network</i>	16
2.6. <i>Feedforward Neural Network</i> (FFNN)	18
2.7. Mean Squared Error (MSE).....	22
2.8. Mean Absolute Error (MAE).....	23
2.9. Metrik Evaluasi Model	23
2.10. Integrasi Keislaman Prediksi Mutasi dengan Al-Qur'an & Hadits	25
2.11. Kajian Prediksi Mutasi Virus dengan Teori Pendukung	28
BAB III METODE PENELITIAN	30
3.1. Jenis Penelitian	30
3.2. Data dan Sumber Data.....	30
3.3. Teknik Pengumpulan Data	31
3.4. Teknik Analisis Data	32
BAB IV HASIL DAN PEMBAHASAN.....	37
4.1 Penerapan Model FFNN Pada Prediksi Mutasi	37
4.1.1 .Deskripsi Data Penelitian.....	37
4.1.2 Pra-Pemrosesan Data	41
4.1.3 Pengaturan Data	44

4.1.4 Hiperparameter Pelatihan dan Penerapan Model FFNN	45
4.1.5 Hasil Pelatihan dan Konvergensi Loss	47
4.1.6 Hasil Prediksi Dataset	49
4.2 Performa Kinerja Model FFNN pada Data <i>Furin Cleavage Stie</i>	50
4.3 Perbandingan dengan Penelitian Sebelumnya	58
4.4 Integrasi Keislaman dalam Analisis Prediksi Mutasi	60
BAB V PENUTUP	62
5.1 Kesimpulan	62
5.2 Saran	62
DAFTAR PUSTAKA	63
LAMPIRAN	68
RIWAYAT HIDUP	83

DAFTAR TABEL

Tabel 4.1	Cuplikan Data Pertama Sebelum Proses Penyelesaian	41
Tabel 4.2	Cuplikan Data Pertama Setelah Proses Penyelesaian di Sekuens.....	42
Tabel 4.3	Cuplikan Data Sekuens Pertama Hasil Konversi Asam Amino ke Representasi Numerik	43
Tabel 4.4	Skema Pembagian Dataset SARS-CoV-2	45
Tabel 4.5	Hiperparameter pada Proses Pelatihan Model	45
Tabel 4.6	Eksperimen Optimizer pada Model FFNN	46
Tabel 4.7	Implementasi Model FFNN pada Dataset SARS-CoV-2.....	46
Tabel 4.8	Hasil Prediksi Model FFNN terhadap Data Uji	49
Tabel 4.9	Evaluasi Kinerja Model FFNN pada Data Uji	50
Tabel 4.10	Perbandingan Model Baseline dengan FFNN pada Data Uji.....	52
Tabel 4.11	Hasil Prediksi Model pada Sampel Data Mutasi SARS-CoV-2	55
Tabel 4.12	Akurasi dan MAE Per-Posisi pada Region FCS	56
Tabel 4.13	Peringkat Posisi Berdasarkan Error Rate (Tertinggi di Atas)	56
Tabel 4.14	Misprediksi Teratas pada Posisi Panas.....	57

DAFTAR GAMBAR

Gambar 2.1	Struktur SARS-CoV-2	10
Gambar 2.2	Furin Cleavage site	13
Gambar 3.1	Flowchart Teknik Analisis Data	36
Gambar 4.1	Tren Jumlah Sekuens SARS-CoV-2 Per Bulan.....	37
Gambar 4.2	Distribusi Jumlah 'X' Per Sekuens SARS-CoV-2	38
Gambar 4.3	Histrogram Jumlah Duplikasi Sekuens SARS-CoV-2	38
Gambar 4.4	Distrubi Karakter Ambigu Per Sekuens SARS-CoV-2	39
Gambar 4.5	Heatmap Frekuensi Asam Amino per Posisi pada <i>Region</i> FCS...	39
Gambar 4.6	Nilai Entropi Asam Amino pada Setiap Posisi FCS	40
Gambar 4.7	Distribusi Jumlah Per-Panjang sekuens Furin SARS-CoV-2.....	42
Gambar 4.9	Kurva <i>History</i> MSE Model FFNN Selama Proses Pelatihan	48
Gambar 4.10	Kurva <i>History</i> MAE Model FFNN Selama Pelatihan.....	48
Gambar 4.11	Akurasi Prediksi FFNN pada Setiap Posisi FCS	53
Gambar 4.12	<i>Heatmap Tingkat Error Posisional</i> pada Prediksi Asam Amino .	54

DAFTAR LAMPIRAN

Lampiran 1	Akses Dataset Sekuensial dan Notebook Kaggle Prediksi Mutasi..	68
Lampiran 2	<i>Script</i> python Model FFNN Prediksi Mutasi.....	68

ABSTRAK

Asviana, Natasyah. 2025. **Pengembangan Model Feedforward Neural Network Untuk Prediksi Mutasi Data Sekuensial Situs Cleavage Sars-Cov-2**. Skripsi. Jurusan Matematika Fakultas Sains Dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (1) Dr. Mohammad Jamhuri, M.Si. (2) Erna Herawati, M.Pd.

Kata kunci: SARS-CoV-2, *furin cleavage site*, *Feedforward Neural Network*, prediksi mutasi, data sekuensial.

Mutasi pada protein spike SARS-CoV-2, khususnya pada wilayah *furin cleavage site* (FCS), berperan penting dalam meningkatkan infektivitas virus dan memengaruhi dinamika penyebaran varian. Oleh karena itu, diperlukan pendekatan prediksi mutasi yang terfokus dan efisien untuk mendukung pemantauan evolusi virus. Penelitian ini bertujuan mengembangkan model prediksi mutasi data sekuensial pada situs cleavage SARS-CoV-2 menggunakan *Feedforward Neural Network* (FFNN) serta mengevaluasi kinerjanya. Data berupa sekuens protein spike varian Omicron (B.1.1.529) diperoleh dari basis data NCBI Virus, kemudian diproses melalui tahap penyalarsan sekuens, ekstraksi wilayah FCS, windowing, dan representasi numerik asam amino. Model FFNN dilatih menggunakan pembagian data latih, validasi, dan uji, serta dievaluasi dengan metrik Mean Squared Error (MSE), Mean Absolute Error (MAE), akurasi, presisi, *recall*, dan *F1_score*. Hasil penelitian menunjukkan bahwa model FFNN mencapai akurasi prediksi posisional rata-rata sebesar 82% pada wilayah FCS, dengan nilai MAE yang relatif rendah serta *konvergensi loss* yang stabil selama pelatihan. Model juga mampu mengidentifikasi posisi dengan tingkat kesalahan tinggi yang berpotensi menjadi hotspot mutasi. Dengan demikian, FFNN terbukti efektif dalam memodelkan pola mutasi sekuensial pada situs cleavage dan berpotensi digunakan sebagai alat bantu analisis evolusi SARS-CoV-2.

ABSTRACT

Asviana, Natasyah. 2025. **Development of a Feedforward Neural Network Model for Predicting Sequential Mutation at the SARS-CoV-2 Cleavage**. Thesis. Department of Mathematics, Faculty of Science and Tecnology. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisors: (1) Dr. Mohammad Jamhuri, M.Si. (2) Erna Herawati, M.Pd.

Keywords: SARS-CoV-2, furin cleavage site, Feedforward Neural Network, mutation prediction, sequential data.

Mutations in the SARS-CoV-2 spike protein, particularly within the furin cleavage site (FCS), play a crucial role in increasing viral infectivity and influencing the evolutionary dynamics of emerging variants. Therefore, a focused and efficient mutation prediction approach is required to support viral evolution surveillance. This study aims to develop a Feedforward Neural Network (FFNN) model for predicting sequential mutation patterns at the SARS-CoV-2 cleavage site and to evaluate its performance. The dataset consists of spike protein sequences of the Omicron variant (B.1.1.529) obtained from the NCBI Virus database and processed through sequence alignment, FCS region extraction, windowing, and numerical encoding of amino acids. The FFNN model was trained using separated training, validation, and testing datasets and evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), accuracy, precision, recall, and F1_score metrics. The results show that the proposed FFNN achieved an average positional prediction accuracy of 82% within the FCS region, with relatively low MAE values and stable loss convergence during training. Furthermore, the model successfully identified positions with high prediction error that potentially represent mutation hotspots. These findings indicate that FFNN is effective in modeling sequential mutation patterns at the cleavage site and has potential applications as a computational tool for SARS-CoV-2 evolutionary analysis.

مستخلص البحث

أصفينا، ناتاشا. ٢٠٢٥. تطوير نموذج شبكة عصبية تغذية أمامية للتنبؤ بالتحورات المتسلسلة للبيانات في موقع انقسام *SARS-CoV-2*. البحث العلمي. قسم الرياضيات، كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرف: (١) الدكتور محمد جمهوري، الماجستير. (٢) إيرنا هيراواقي، الماجستير.

الكلمات الأساسية: *SARS-CoV-2*، موقع انقسام الفورين، شبكة عصبية تغذية أمامية، توقع الطفرات، البيانات المتسلسلة.

لعبت الطفرات في بروتين السنبلة لفيروس *SARS-CoV-2*، ولا سيما في منطقة موقع القطع بالفورين (*FCS*)، دوراً مهماً في زيادة قابلية العدوى والتأثير في ديناميكيات انتشار المتحورات. لذلك، تبرز الحاجة إلى منهجية تنبؤ بالطفرات تكون مركزة وفعالة لدعم مراقبة تطور الفيروس. هدف هذا البحث إلى تطوير نموذج للتنبؤ بطفرات البيانات التسلسلية في موقع القطع لفيروس *SARS-CoV-2* باستخدام الشبكة العصبية ذات التغذية الأمامية، إضافةً إلى تقييم أدائه. اعتمدت الدراسة على بيانات تسلسلات بروتين السنبلة لمتحور أوميكرون (*B.1.1.529*) المأخوذة من قاعدة بيانات *NCBI Virus*، والتي جرى معالجتها من خلال محاذاة التسلسلات، واستخراج منطقة *FCS*، وتطبيق تقنية النوافذ، وتحويل الأحماض الأمينية إلى تمثيل عددي. تم تدريب نموذج *FFNN* باستخدام تقسيم البيانات إلى مجموعات تدريب وتحقق واختبار، وتقييمه باستخدام مقاييس متوسط مربع الخطأ (*MSE*)، ومتوسط الخطأ المطلق (*MAE*)، والدقة، والإحكام، والاستدعاء، ومعامل *F1*. أظهرت النتائج أن نموذج *FFNN* حقق متوسط دقة تنبؤ موضوعية بلغ ٨٢٪ في منطقة *FCS*، مع قيم *MAE* منخفضة نسبياً واستقرار تقارب دالة الخطأ أثناء التدريب. كما تمكن النموذج من تحديد مواضع ذات معدلات خطأ مرتفعة يُحتمل أن تمثل بؤراً ساخنة للطفرات. وبناءً على ذلك، يثبت نموذج *FFNN* فعاليته في نمذجة أنماط الطفرات التسلسلية في موقع القطع، مع إمكانية استخدامه كأداة مساعدة في تحليل تطور فيروس *SARS-CoV-2*.

BAB I

PENDAHULUAN

1.1 iLatar Belakang

SARS-CoV-2 dikenal sebagai virus penyebab pandemi COVID-19 yang mulai terdeteksi pada akhir tahun 2019 (Holmes et al., 2021). Mutasi pada protein spike dari virus SARS-CoV-2 menjadi faktor utama yang memengaruhi penularan, efektivitas vaksin, dan kemampuan virus beradaptasi dengan sistem imun (Harvey et al., 2021). Salah satu bagian paling kritis dari protein spike tersebut adalah *Furin Cleavage Site* (FCS), dimana pemotongan protein spike oleh enzim furin pada situs S1/S2 sangat penting untuk fusi antar sel dan masuknya virus ke dalam sel inang (Hoffmann et al., 2020). Proses ini meningkatkan infektivitas virus serta berkontribusi pada penyebaran infeksi (Hoffmann et al., 2020). Oleh karena itu, prediksi mutasi pada situs ini sangat penting untuk mendukung sistem pemantauan evolusi virus dan pengembangan strategi kesehatan masyarakat.

Upaya prediksi mutasi genom SARS-CoV-2 memerlukan data yang bersumber dari repositori terpercaya. Data genom SARS-CoV-2 diambil dari repositori NCBI Virus dengan menerapkan filter berdasarkan gen target, kualitas sekuens, metadata asal sampel, dan jenis varian untuk memperoleh data yang relevan dan valid. NCBI Virus menyediakan sekuens genetik dalam FASTA dan *nucleotide record* hasil pengumpulan global yang dapat digunakan untuk analisis bioinformatika (Brister et al., 2015). Basis data ini telah dimanfaatkan secara luas dalam ribuan publikasi ilmiah yang meneliti mutasi, filogenetik, dan dinamika genom (Brister et al., 2015). Dengan data yang terstandar serta terverifikasi,

proses analisis prediktif dapat dilakukan secara lebih akurat dan terarah. Selain itu, data genomik bersifat sekuensial karena tersusun atas urutan nukleotida yang saling berkaitan, sehingga perubahan basa mengikuti pola tertentu yang dapat ditelusuri melalui sejarah mutasinya.

Feedforward Neural Network (FFNN) diterapkan untuk menganalisis data sekuensial dan telah terbukti efektif melalui berbagai penelitian sebelumnya. Seperti Yuliandar et al., (2012) menggunakan FFNN untuk memprediksi data deret waktu nilai kurs Dolar Australia terhadap Rupiah. Model jaringan saraf tiruan ini memiliki arsitektur sederhana, umumnya terdiri dari satu lapisan tersembunyi (*hidden layer*), dan dapat dimanfaatkan untuk prediksi pada deret waktu (Azam et al., 2018). Dalam *Feedforward Neural Network*, data bergerak hanya dalam satu arah (*feed-forward*) dari lapisan input menuju lapisan output tanpa umpan balik, sehingga ini mampu mengenali pola kompleks dan menghasilkan prediksi berdasarkan data historis (Septiana & Br Bangun, 2023). Dengan demikian, *Feedforward Neural Network* dapat diterapkan dalam pemrosesan data sekuensial.

Berbagai pendekatan telah dikembangkan untuk memahami mutasi pada *furin cleavage site* (FCS). Tian et al. (2012) memperkenalkan PiTou, alat prediksi berbasis *hidden Markov model* (HMM) dan skor biologis yang berhasil mendeteksi keberadaan *furin cleavage site* dengan sensitivitas dan spesifisitas tinggi. Namun, PiTou tidak dirancang untuk memprediksi mutasi, melainkan hanya mendeteksi situs pemotongan pada urutan protein yang tetap. Gu, C (2020) juga mengembangkan *FindFur*, alat prediksi berbasis *profile hidden Markov model* (HMM) dan *logistic regression* untuk mendeteksi situs *furin cleavage* pada

glikoprotein virus berselubung. Hasil penelitian menunjukkan performa baik, tetapi pendekatan Gu tidak diarahkan untuk memprediksi mutasi, melainkan hanya mendeteksi keberadaan motif *furin cleavage site*.

Sementara itu, Maher dan Yousif (2023) mengembangkan sebuah model prediktif berbasis *meachine learning*, khususnya *neural network* dan fitur epidemiologis, untuk memprediksi mutasi SARS-CoV-2 yang berpotensi menyebar secara luas. Model ini menggunakan data genomik, evolusi, imunologi, serta data spasial dan temporal dari penyebaran mutasi, lalu mengintegrasikannya dalam kerangka prediksi mutasi yang berpotensi menjadi bagian dari varian yang diwaspadai. Di sisi lain, Saldivar-espinoza et al. (2022) menggunakan Artificial Neural Network (ANN) untuk memprediksi *recurrent mutations* dalam genom SARS-CoV-2. Model ini mampu mengidentifikasi mutasi berulang berdasarkan fitur-fitur seperti struktur sekunder RNA dan reaktivitas kimia. Walaupun hasil prediksi cukup menjanjikan, penelitian ini tidak secara khusus menyasar situs *furin cleavage*, dan lebih menyoroti mutasi global.

Kemajuan lebih lanjut ditunjukkan oleh Jamhuri et al., (2025) dalam disertasinya, yang mengembangkan pendekatan optimasi Gauss-Newton terdistribusi untuk meningkatkan efisiensi pelatihan model deep learning dalam memprediksi mutasi spike protein SARS-CoV-2. Ia menggunakan kombinasi model MLP, CNN, dan LSTM, dan menguji akurasi prediksi mutasi pada berbagai varian termasuk Alpha, Beta, Delta, Gamma, dan Omicron. Hasilnya menunjukkan bahwa model tersebut mampu mencapai akurasi tinggi hingga 99% pada sebagian besar varian. Namun, varian Omicron, yang memiliki kompleksitas mutasi paling tinggi dan data terbatas, menyebabkan penurunan akurasi sebesar 1–

2%, yang menunjukkan tantangan nyata dalam menangani virus yang terus berkembang. Di sisi lain, model yang digunakan tergolong kompleks dan tidak secara eksplisit ditujukan untuk prediksi mutasi di situs *cleavage*.

Dari tinjauan studi tersebut, dapat diketahui bahwa meskipun telah dilakukan banyak upaya untuk memodelkan mutasi SARS-CoV-2, masih terdapat kekosongan penelitian, yaitu belum adanya pendekatan secara khusus ditujukan untuk memprediksi mutasi pada *furin cleavage site* menggunakan *Feedforward Neural Network* (FFNN). Kekurangan ini mengindikasikan perlunya metode yang lebih terarah dalam menganalisis mutasi pada bagian penting dari virus tersebut.

Mengingat pentingnya peran *furin cleavage site* dalam infektivitas virus dan implikasinya terhadap efektivitas vaksin, penelitian ini bertujuan untuk mengembangkan sistem prediksi data sekuensial berbasis *Feedforward Neural Network* (FFNN) dalam menganalisis mutasi situs *cleavage* pada virus SARS-CoV-2. Model ini diharapkan mampu menangkap pola-pola mutasi yang signifikan secara biologis dengan efisien, serta memberikan kontribusi pada pengembangan sistem pemantauan evolusi virus yang adaptif dan berkelanjutan.

Pengembangan ilmu pengetahuan merupakan perintah Islam agar umatnya mencari dan mengembangkan ilmu demi kemaslahatan. Hal ini ditegaskan Allah SWT dalam QS.Al-Mujadilah ayat 11:

وَإِذَا قِيلَ انْشُزُوا فَانْشُزُوا يَرَفَعُ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ ۖ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ۚ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ ۝۱۱

“..Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat. Allah melihat pada apa yang kamu kerjakan.” (QS.Al-Mujadilah: 11)

Ayat ini menegaskan bahwa ilmu pengetahuan memiliki kedudukan yang sangat mulia dan menjadi sarana meninggikan derajat manusia di sisi Allah SWT. pengembangan ilmu, termasuk penelitian mutasi virus SARS-CoV-2, bernilai ibadah ketika diarahkan untuk kemaslahatan umat. Dalam konteks ini, pengembangan model prediksi berbasis *Feedforward Neural Network* bukan hanya kontribusi terhadap keilmuan modern, tetapi juga bentuk aktualisasi nilai Islam yang mendorong pemanfaatan akal demi menjaga jiwa manusia.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka rumusan masalah sebagai berikut:

1. Bagaimana penerapan model prediksi mutasi situs *cleavage* dari virus SARS CoV-2 menggunakan *Feedforward Neural Network* (FFNN)?
2. Bagaimana *performance* model yang dikembangkan dalam memprediksi mutasi situs *cleavage* dari virus SARS CoV-2?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini sebagai berikut:

1. Untuk mengembangkan model prediksi mutasi situs *cleavage* dari virus SARS CoV-2 menggunakan *Feedforward Neural Network* (FFNN).
2. Untuk mengetahui *performance* model yang dikembangkan dalam memprediksi mutasi situs *cleavage* dari virus SARS CoV-2.

1.4 Manfaat Penelitian

Dalam penelitian ini diharapkan memiliki manfaat sebagai berikut:

1. Manfaat teoritis

Manfaat teoritis penelitian ini meliputi hal berikut:

- a. Memberikan pengembangan model prediksi mutasi situs cleavage SARS-CoV-2 berbasis *Feedforward Neural Network* dan memperluas kajian mengenai kemampuan FFNN dalam mengolah data sekuensial biologis.
- b. Memperkaya literatur terkait analisis akurasi dan kemampuan pembelajaran model pada data sekuensial virus.

2. Manfaat Praktis

Manfaat praktis penelitian ini meliputi hal berikut:

- a. Model prediksi berfungsi sebagai alat bantu analisis untuk lembaga riset dan instansi kesehatan dalam mendeteksi potensi mutasi atau varian baru virus SARS-CoV-2.
- b. Hasil penelitian mendukung pengambilan keputusan yang lebih cepat dan akurat dalam pengembangan vaksin, pengawasan epidemiologi, serta penentuan strategi respon medis terhadap penyakit menular.

1.5 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka dilakukan pembatasan permasalahan yaitu penelitian ini hanya menggunakan data sekuens genom dari varian SARS-CoV-2 Omicron (B.1.1.529) yang diperoleh dari basis data publik NCBI. Analisis difokuskan pada mutasi yang terjadi pada wilayah furin *cleavage* site (FCS) pada protein spike, karena bagian ini memiliki peran penting dalam proses infeksi virus.

1.6 Definisi Istilah

Dalam penelitian ini digunakan beberapa istilah untuk menghindari perbedaan makna dan menyamakan pemahaman terhadap tema serta arah penelitian. Definisi istilah pada penelitian ini yaitu sebagai berikut:

1. Mutasi

Perubahan pada urutan nukleotida dalam genom organisme yang dapat memengaruhi struktur dan fungsi protein yang dihasilkan.

2. Prediksi

Proses memperkirakan nilai atau kejadian di masa depan berdasarkan data historis.

3. SARS-CoV-2

Virus penyebab COVID-19 yang pertama kali diidentifikasi akhir 2019.

4. Data Sekuensial

Data yang disusun dalam urutan tertentu di mana urutan tersebut penting untuk analisis.

5. *Artificial Neural Network*

Suatu model komputasional yang terinspirasi dari struktur jaringan saraf biologis, terdiri dari lapisan neuron yang saling terhubung untuk memproses data dan melakukan prediksi.

6. FASTA

Format file standar untuk menyimpan sekuens nukleotida atau protein dalam bentuk teks, yang digunakan dalam bioinformatika untuk analisis genomik.

7. *Machine Learning*

Cabang kecerdasan buatan yang melibatkan algoritma yang belajar dari data untuk membuat prediksi atau keputusan.

8. NCBI Virus

Repositori basis data genomik yang dikelola oleh National Center for Biotechnology Information (NCBI), menyediakan sekuens virus global

dalam format FASTA untuk analisis bioinformatika.

9. Protein Spike

Protein permukaan SARS-CoV-2 yang berperan dalam pengikatan dan masuknya virus ke sel inang.

10. Residu

Satu unit penyusun sekuens, misalnya satu asam amino dalam sekuens protein atau satu nukleotida dalam DNA/RNA.

11. *Furin Cleavage Site*

lokasi spesifik pada protein virus yang harus dipotong oleh enzim protease untuk meningkatkan infeksi atau perakitan partikel virus yang matang.

12. Windowing

Teknik membagi data berurutan menjadi potongan dengan panjang tertentu untuk dianalisis atau diprediksi.

13. *Lag* (Jendela)

Jarak waktu atau urutan antara data sebelumnya dengan data yang diprediksi.

BAB II

KAJIAN PUSTAKA

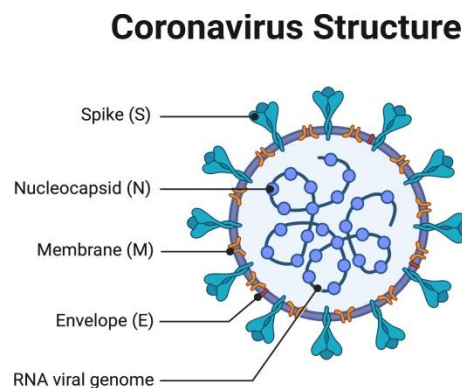
Subbab ini memuat landasan teoritis yang digunakan dalam penelitian, mencakup struktur biologis SARS-CoV-2, karakteristik protein Spike, dan konsep situs furin cleavage. Selain itu dibahas prinsip prediksi data sekuensial dengan *neural network*, arsitektur *Feedforward Neural Network*, serta metrik evaluasi yang digunakan dalam menilai kualitas model.

2.1. Struktur SARS-CoV-2

SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) adalah jenis virus corona baru yang pertama kali diidentifikasi di Wuhan, Tiongkok, pada akhir tahun 2019 dan memiliki kemiripan genetik dengan virus penyebab SARS pada tahun 2002 (Holmes et al., 2021). Penamaan tersebut diberikan oleh Komite Internasional Taksonomi Virus (ICTV). Virus ini adalah virus yang menyebabkan penyakit pernafasan pada manusia dan ditularkan dari hewan ke manusia dalam bentuk mutasi. Dalam perkembangannya, SARS-CoV-2 menghasilkan sejumlah varian yang diklasifikasikan sebagai *Variants of Concern* (VOC), seperti Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), serta Omicron yang awalnya diklasifikasikan sebagai B.1.1.529 dan kemudian berkembang menjadi berbagai *sublineage* BA (Park et al., 2023).

SARS-CoV-2 dikenal sebagai virus RNA yang beruntai tunggal positif (*positive-sense single-stranded RNA virus*), yang artinya materi genetiknya dapat langsung digunakan sebagai cetakan untuk membuat protein virus setelah memasuki inang (Naqvi et al., 2020). RNA tersebut dilindungi oleh protein neukleokapsid (N) yang berfungsi menjaga stabilitas dan mendukung proses

replikasi. Partikel inti virus ini selanjutnya dikelilingi oleh selubung membran luar yang terbuat dari lipid (*lipid envelope*) dengan protein yang disisipkan. Selubung ini sangat penting dalam menjaga integritas virus, namun juga menjadi titik lemah karena mudah rusak oleh senyawa lipid seperti larutan hidrogen peroksida, pemutih hipoklorit, deterjen dan sabun (V'kovski et al., 2021). Inilah alasan mengapa mencuci tangan dengan sabun menjadi salah satu langkah pencegahan yang efektif terhadap penyebaran virus SARS-CoV-2.



Gambar 2.1 Struktur SARS-CoV-2
(Sumber: www.biophysics.org)

Pada Gambar 2.1 merupakan Ilustrasi struktur SARS-CoV-2. Selubung membran luar inti virus berasal dari sel tempat virus terakhir kali dirakit tetapi dimodifikasi untuk mengandung protein virus tertentu, termasuk protein spike (S), membrane (M), dan envelope (E). Protein spike merupakan komponen kunci dalam proses infeksi karena mampu mengenali dan berikatan dengan reseptor *Angiotensin Converting Enzyme 2* (ACE2) di permukaan sel manusia, terutama di saluran pernapasan (Yan et al., 2020). Protein E dan M juga berperan penting dalam siklus hidup virus. Protein E terlibat dalam proses perakitan dan pelepasan virus baru dari dalam sel, sementara protein protein M membantu

mempertahankan bentuk dan struktur virus secara keseluruhan. Interaksi antar-protein ini sangat penting dalam membentuk virion yang infeksi.

2.2. Struktur Protein Spike SARS-CoV-2

Protein spike (S) merupakan glikoprotein tipe I yang terletak di permukaan virus SARS-CoV-2. Struktur ini berbentuk seperti paku dan menjadi ciri khas dari virus corona yang tampak seperti mahkota di bawah mikroskop elektron. Protein spike memiliki peran penting dalam proses masuknya virus ke dalam sel inang dengan cara menempel pada reseptor sel manusia (Walls et al., 2020).

Protein spike dari SARS-CoV-2 adalah homotrimer, yang setiap rantainya terdiri dari 1273 residu (Sinha et al., 2023). Protein spike memiliki dua subunit utama, yaitu S1 (14-685 residu) dan S2 (686-1273 residu) (Huang et al., 2020), Subunit S1 mengandung *receptor-binding domain* (RBD) untuk mengenali dan menempel pada reseptor ACE2 pada permukaan sel manusia (Yan et al., 2020). Proses pengikatan ini menjadi tahap awal dalam infeksi, karena menentukan keberhasilan virus menempel pada sel target.

Subunit S2 bertanggung jawab dalam mediasi fusi membran virus dengan membran sel inang. Bagian ini terdiri atas *fusion peptide* (FP), *heptad repeat 1* (HR1), *heptad repeat 2* (HR2), serta domain transmembran (TM) dan sitoplasmik (CT). Setelah subunit S1 berikatan dengan ACE2, S2 mengalami perubahan konformasi besar yang memungkinkan FP menyisip ke dalam membran sel target. HR1 dan HR2 kemudian berinteraksi membentuk struktur *six-helix bundle* (6-HB) yang menarik membran virus dan membran sel inang menjadi sangat dekat hingga akhirnya menyatu (Huang et al., 2020).

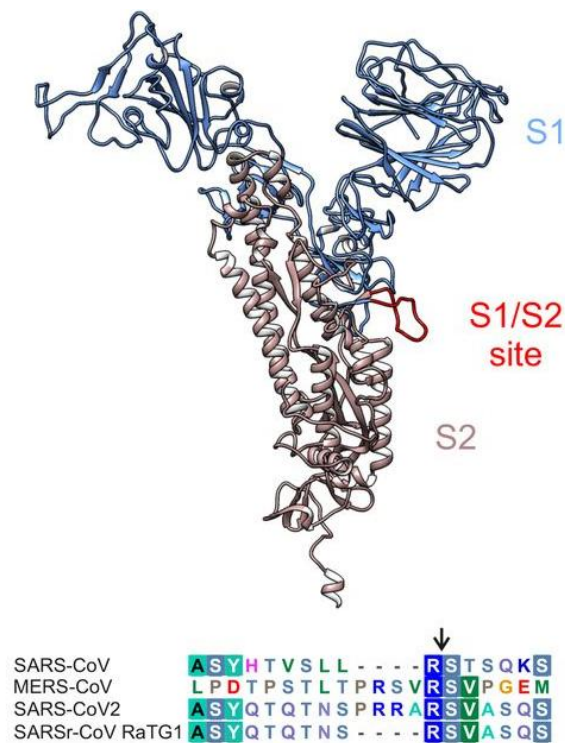
Aktivasi protein spike membutuhkan proses pemotongan atau *cleavage* oleh protease sel inang pada dua titik utama, yaitu S1/S2 dan S2'. Enzim seperti furin, TMPRSS2, dan cathepsin L berperan dalam memediasi pemotongan ini. Secara khusus, SARS-CoV-2 memiliki *furin cleavage site* unik pada batas S1/S2 yang tidak ditemukan pada SARS-CoV, sehingga meningkatkan efisiensi pemotongan oleh protease inang. Hal ini diduga menjadi salah satu faktor yang membuat SARS-CoV-2 memiliki tingkat penularan lebih tinggi dibandingkan virus corona sebelumnya (Huang et al., 2020).

Protein spike tersusun atas rantai asam amino yang direpresentasikan sebagai deretan karakter tunggal dalam alfabet protein. Representasi ini menggunakan dua puluh asam amino standar yang masing-masing memiliki kode satu huruf, yaitu alanin (A), arginin (R), asparagin (N), aspartat (D), sistein (C), glutamin (Q), glutamat (E), glisin (G), histidin (H), isoleusin (I), leusin (L), lisin (K), metionin (M), fenilalanin (F), prolin (P), serin (S), treonin (T), triptofan (W), tirosin (Y), dan valin (V). Sekuens protein diperlakukan sebagai rangkaian simbol, sehingga pola residu dapat dianalisis secara komputasional. Pendekatan ini memungkinkan identifikasi perubahan huruf dalam sekuens yang merepresentasikan substitusi asam amino, sehingga sangat relevan dalam memetakan variasi dan mutasi pada protein spike SARS-CoV-2 (Motomura et al., 2012).

2.3. Situs Furin *Cleavage* pada Virus SARS-CoV-2

Situs furin *cleavage* adalah lokasi spesifik pada protein virus yang harus dipotong oleh enzim protease untuk meningkatkan infeksi atau perakitan partikel virus yang matang. Pada SARS-CoV-2, situs *cleavage* utama terdapat di protein

spike, yang digunakan virus untuk berikatan dengan resepto ACE2 di sel manusia. Keberadaan *furin cleavage site* (FCS) pada SARS-CoV-2 membuat virus lebih mudah masuk ke dalam sel dibandingkan SARS-CoV versi tahun 2002 (Chan & Zhan, 2022).



Gambar 2.2 Furin Cleavage site
(Sumber: www.nature.org)

Pada Gambar 2.2, karakteristik unik SARS-CoV-2 dibandingkan virus corona lainnya adalah keberadaan *situs furin cleavage* pada batas antara subunit S1 dan S2. Situs S1/S2 (ditandai merah) adalah lokasi spesifik dimana spike dipotong oleh enzim seperti furin. Keberadaan situs ini mempermudah pemrosesan dan aktivasi protein spike di berbagai jaringan tubuh manusia, sehingga meningkatkan kemampuan penyebaran virus serta memperluas tropisme jaringan. Pemotongan ini penting untuk mengungkapkan bagian peptida fusi, sehingga memungkinkan terjadinya perubahan konformasi protein spike menuju

struktural *post-fusion*. Proses ini memfasilitasi fusi membran virus dengan membran sel inang. Tanpa aktivasi di kedua situs ini, protein spike tidak dapat menjalankan fungsi fusi secara optimal (Hoffmann et al., 2020)..

Secara spesifik, posisi *furin cleavage site* (FSC) pada protein spike SARS-CoV-2 terletak di batasi subunit S1/S2, tepatnya pada residu asam amino 680-685 dengan urutan motif PRRAR↓SV, di mana tanda panah (↓) menunjukkan titik pemotongan oleh enzim furin. Sisipan empat asam amino PRRA pada posisi ini merupakan ciri khas SARS-CoV-2 yang tidak ditemukan pada SARS-CoV klasik (Örd et al., 2020). Karena mutasi pada situs ini dapat memengaruhi infektivitas, virulensi, dan penyebaran virus, FCS menjadi fokus penting dalam studi evolusi, pengembangan vaksin, dan pemodelan prediksi berbasis sekuens.

2.4. Multiple Sequence Alignment

Multiple Sequence Alignment (MSA) merupakan proses penyelarasan lebih dari dua sekuens biologis untuk mengidentifikasi posisi homolog yang memiliki keterkaitan evolusioner. MSA dijelaskan sebagai pengembangan dari *pairwise alignment*, namun dengan kompleksitas matematis yang jauh lebih tinggi sehingga perhitungan optimal tidak dapat dilakukan menggunakan *dynamic programming* biasa. Pada *pairwise alignment*, skor penyelarasan dua sekuens $A = a_1, \dots, a_m$ dan $B = b_1, \dots, b_n$ dihitung menggunakan formulasi rekursif:

$$S_{i,j} = \max \begin{cases} S_{i-1,j-1} + s(a_i, b_j) \\ S_{i,j-1} + s(-, b_j) \\ S_{i-1,j} + s(a_i, -) \end{cases}$$

Dengan fungsi skor $s(a, b) = 1$ untuk *match*, -1 untuk *mismatch*, dan -1 untuk *gap*. Karena perumusan ini hanya efisien untuk dua sekuens, MSA menggunakan pendekatan heuristik seperti *progressive alignment* yang dimulai

dari *pairwise* untuk membentuk *distance matrix*, kemudian membuat *guide tree* menggunakan metode UPGMA atau Neighbor-Joining. *Guide tree* tersebut menentukan urutan penyelarasan, dan proses selanjutnya melibatkan *profile alignment* ketika sekuens baru diselaraskan dengan kelompok sekuens yang telah tersusun. (Singh, 2015).

MAFFT merupakan salah satu algoritma MSA yang efisien dan akurat, dikembangkan oleh Katoh sejak 2002 dan terus disempurnakan. Pada proses *alignment*, MAFFT menambahkan gap (“–”) ketika terdapat sisipan atau penghapusan, sehingga panjang seluruh sekuens menjadi seragam dan residu dapat sejajar pada kolom yang sama. Penambahan gap berfungsi sebagai penanda insersi dan delesi tanpa mengubah urutan residu asli, sehingga posisi residu homolog dapat sejajar pada kolom yang sama (Katoh, 2020). Sebagai ilustrasi, misalkan lima sekuens awal;

Sequens 1 : T Q V H R C E

Sequens 2 : T Q V H C E

Sequens 3 : S Q V H N D E

Sequens 4 : Q V H R D E

Sequens 5 : T Q V H R C E

Sekuens tersebut memiliki panjang berbeda sehingga posisi homolog tidak langsung terlihat. Hasil penyelarasan menggunakan algoritma MSA sebagai berikut:

Sequens 1 : T Q V H R C E

Sequens 2 : T Q V H – C E

Sequens 3 : S Q V H N D E

Sequens 4 : – Q V H R D E

Sequens 5 : T Q V H R C E

Penyesuaian ini memungkinkan setiap residu menempati kolom yang sesuai dengan posisi homolognya, sehingga pola kemiripan maupun perbedaan antar sekuens dapat diidentifikasi secara lebih akurat. Dengan demikian, MSA tidak hanya berfungsi untuk menata sekuens dalam format yang sejajar, tetapi juga menyediakan fondasi matematis untuk analisis homologi, inferensi hubungan evolusioner, serta identifikasi residu konservatif dalam data biologis.

2.5. Prediksi Data Sekuensial Berbasis *Neural Network*

Neural network merupakan pendekatan populer dalam bidang kecerdasan buatan yang terinspirasi dari cara kerja sistem saraf manusia (Lecun et al., 2015). Jaringan ini tersusun atas neuron buatan yang saling terhubung dalam beberapa lapisan, dan digunakan untuk mempelajari serta mengenali pola dari data masukan. Dalam konteks prediksi, neural network telah terbukti efektif, terutama pada data yang kompleks dan tidak bersifat linier (Lecun et al., 2015).

Prediksi data sekuensial, yakni data yang bergantung pada urutan seperti teks, sinyal biologis, atau urutan genom, membutuhkan pendekatan khusus karena hubungan antar elemen sangat ditentukan oleh posisi dalam rangkaian. Untuk itu, dikembangkan arsitektur *neural network* yang dirancang untuk memahami urutan, seperti *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), dan *Gated Recurrent Unit* (GRU) (Greff et al., 2017). Model-model ini memungkinkan jaringan menyimpan informasi dari langkah-langkah sebelumnya dalam suatu urutan, sehingga sangat cocok untuk tugas seperti analisis sekuens genom atau prediksi mutasi virus.

Dalam konteks biologi molekuler dan bioinformatika, neural network telah digunakan untuk memprediksi interaksi protein, mengenali motif pengikatan asam nukleat, serta mendeteksi kemungkinan mutasi pada urutan virus. Sebagai contoh, Alipanahi et al. (2015) menggunakan deep learning untuk memprediksi spesifisitas pengikatan protein terhadap DNA dan RNA secara akurat. Pendekatan ini dapat diterapkan untuk mempelajari situs penting dalam protein virus seperti *cleavage site*, yang berperan dalam infektivitas dan penyebaran virus.

Pemahaman representasi data biologis dalam konteks prediksi ini memerlukan pendekatan yang mampu mengubah sekuens protein menjadi bentuk numerik agar dapat diolah oleh model *neural network*. Pendekatan ini sejalan dengan penjelasan yang diberikan oleh Jason Brownlee (2020) dalam bukunya, yang menekankan pentingnya proses encoding sekuens menjadi nilai numerik sebelum digunakan sebagai input dalam model pembelajaran mesin. Misalkan terdapat sekuens *furin cleavage* dari beberapa virus yang direpresentasikan dalam bentuk matriks:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \\ h_{41} & h_{42} & h_{43} \\ h_{51} & h_{52} & h_{53} \end{bmatrix} \quad (1)$$

Langkah pertama melakukan *windowing* terhadap matriks H dengan ukuran jendela (*lag*) sebesar 2. Ini berarti setiap pasangan baris berturutan akan diambil sebagai input, dan baris berikutnya akan menjadi targetnya. Diperoleh *window* sebagai berikut:

$$\mathbf{x}'_1 = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{bmatrix} \quad y_1 = [h_{31} \quad h_{32} \quad h_{33}] \quad (2)$$

$$\mathbf{x}'_2 = \begin{bmatrix} h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad y_2 = [h_{41} \quad h_{42} \quad h_{43}] \quad (3)$$

$$\mathbf{x}'_3 = \begin{bmatrix} h_{31} & h_{32} & h_{33} \\ h_{41} & h_{42} & h_{43} \end{bmatrix} \quad y_3 = [h_{51} \quad h_{52} \quad h_{53}] \quad (4)$$

Langkah berikutnya adalah *flattening*, yaitu meratakan masing-masing *window* input menjadi satu vektor baris agar dapat digunakan sebagai *input* pada model pembelajaran mesin. Menghasilkan:

$$\mathbf{x}_1 = [h_{11} \quad h_{12} \quad h_{13} \quad h_{21} \quad h_{22} \quad h_{23}] \quad (5)$$

$$\mathbf{x}_2 = [h_{21} \quad h_{22} \quad h_{23} \quad h_{31} \quad h_{32} \quad h_{33}] \quad (6)$$

$$\mathbf{x}_3 = [h_{31} \quad h_{32} \quad h_{33} \quad h_{41} \quad h_{42} \quad h_{43}] \quad (7)$$

Selanjutnya, menggabungkan input hasil *flattening* dan target menjadi pasangan data. Sehingga didapatkan:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{21} & h_{22} & h_{23} \\ h_{21} & h_{22} & h_{23} & h_{31} & h_{32} & h_{33} \\ h_{31} & h_{32} & h_{33} & h_{41} & h_{42} & h_{43} \end{bmatrix} \quad (8)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} h_{31} & h_{32} & h_{33} \\ h_{41} & h_{42} & h_{43} \\ h_{51} & h_{52} & h_{53} \end{bmatrix} \quad (9)$$

Maka, hasil akhir data tersebut yaitu;

$$D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), (\mathbf{x}_3, \mathbf{y}_3)\} \quad (10)$$

Dengan terbentuknya pasangan data $(\mathbf{x}_i, \mathbf{y}_i)$, maka struktur data sekuensial selanjutnya dapat digunakan dalam proses model prediksi berbasis *neural network*.

2.6. Feedforward Neural Network (FFNN)

Feedforward Neural Network (FFNN) merupakan salah satu arsitektur paling dasar dalam jaringan saraf tiruan (*Artificial Neural Network/ANN*), di mana aliran informasi hanya bergerak maju dari input ke output, tanpa sirkulasi (*loop*)

atau umpan balik (Malik et al., 2013). Jaringan ini biasanya terdiri dari tiga lapisan utama, yaitu *input layer*, *hidden layer* dan *output layer*. *Input layer* bertugas menerima data awal, *hidden layer* memproses data melalui kombinasi linear dan fungsi aktivasi non-linear, dan *output layer* memberikan hasil akhir berupa prediksi atau klasifikasi (Malik et al., 2013).

Setiap neuron dalam FFNN bekerja dengan prinsip yang sederhana namun kuat. Neuron menerima input berupa vektor fitur

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

Misalkan $n = 80$ pada vektor input $\mathbf{x}$, maka;

$$\mathbf{x} = (x_1 \quad x_2 \quad \cdots \quad x_{80}) \quad (11)$$

Dan bobot sinaptik dinyatakan dalam vektor baris W , dan bias dinotasikan dengan

$\mathbf{b}$. Dinotasikan sebagai berikut:

$$W = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,80} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,80} \\ \vdots & \vdots & \ddots & \vdots \\ w_{256,1} & w_{256,2} & \cdots & w_{256,80} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{256} \end{pmatrix} \quad (12)$$

Vektor input $\mathbf{x}$ terlebih dahulu ditranspose untuk menyesuaikan dimensi dengan matriks bobot, sehingga proses matriks dapat dilakukan secara matematis sesuai kaidah. Dilanjutkan dengan langkah pertama yaitu operasi linear pada hidden layer pertama, operasi ini menghasilkan:

$$\begin{aligned} A &= W\mathbf{x}^T + \mathbf{b} \\ &= \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,80} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,80} \\ \vdots & \vdots & \ddots & \vdots \\ w_{256,1} & w_{256,2} & \cdots & w_{256,80} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{80} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{256} \end{pmatrix} \\ &= \begin{pmatrix} w_{1,1}x_1 + w_{1,2}x_2 + \cdots + w_{1,80}x_{80} + b_1 \\ w_{2,1}x_1 + w_{2,2}x_2 + \cdots + w_{2,80}x_{80} + b_2 \\ \vdots \\ w_{256,1}x_1 + w_{256,2}x_2 + \cdots + w_{256,80}x_{80} + b_{256} \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{256} \end{pmatrix} \quad (13)$$

Kemudian dilewatkan ke fungsi aktivasi non-linear, fungsi aktivasi σ diterapkan setiap elemen untuk menghasilkan *output non-linear* dari hidden layer pertama.

Output ini menjadi input bagi lapisan berikutnya, maka didapatkan;

$$\begin{aligned} F &= \sigma(A) \\ &= \begin{pmatrix} \sigma(a_1) \\ \sigma(a_2) \\ \vdots \\ \sigma(a_{256}) \end{pmatrix} \\ &= \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{256} \end{pmatrix} \end{aligned} \quad (14)$$

Langkah selanjutnya yaitu operasi linear pada hidden layer kedua, misalkan bobot untuk lapisan kedua dinyatakan dengan:

$$V = \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,256} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,256} \\ \vdots & \vdots & \ddots & \vdots \\ v_{128,1} & v_{128,2} & \cdots & v_{128,256} \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{128} \end{pmatrix} \quad (15)$$

Sehingga, operasi ini menghasilkan:

$$\begin{aligned} B &= VF + \mathbf{c} \\ &= \begin{pmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,256} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,256} \\ \vdots & \vdots & \ddots & \vdots \\ v_{128,1} & v_{128,2} & \cdots & v_{128,256} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{256} \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{128} \end{pmatrix} \\ &= \begin{pmatrix} v_{1,1}f_1 + v_{1,2}f_2 + \cdots + v_{1,256}f_{256} + c_1 \\ v_{2,1}f_1 + v_{2,2}f_2 + \cdots + v_{2,256}f_{256} + c_2 \\ \vdots \\ v_{128,1}f_1 + v_{128,2}f_2 + \cdots + v_{128,256}f_{256} + c_{128} \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{128} \end{pmatrix} \quad (16)$$

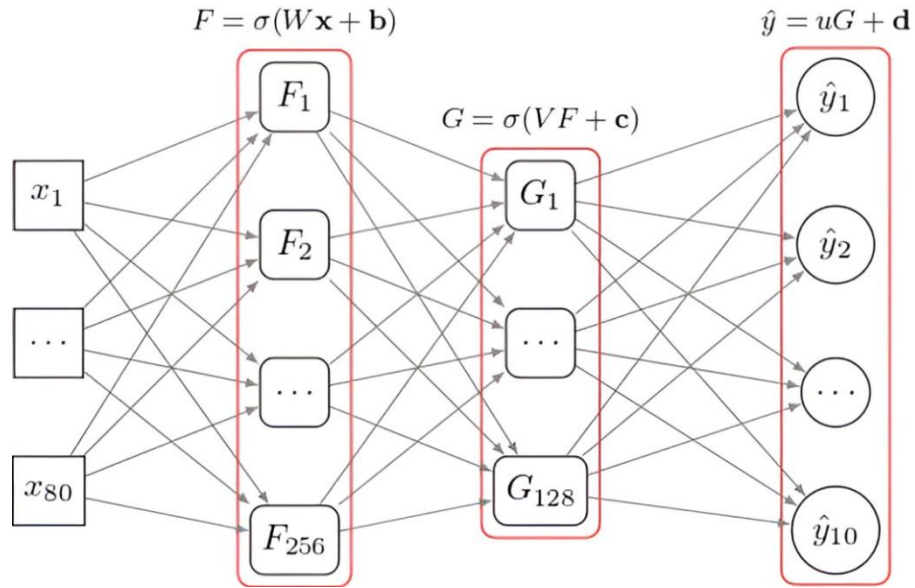
Kemudian dilewatkan ke fungsi aktivasi non-linear, maka didapatkan:

$$\begin{aligned} G &= \sigma(B) \\ &= \begin{pmatrix} \sigma(e_1) \\ \sigma(e_2) \\ \vdots \\ \sigma(e_{128}) \end{pmatrix} \\ &= \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{128} \end{pmatrix} \end{aligned} \quad (17)$$

Langkah terakhir, vektor hasil aktivasi G dikombinasikan vektor bobot output u dengan bias $\mathbf{d}$ untuk menghasilkan output akhir prediksi $\hat{y}$, sehingga diperoleh:

$$\begin{aligned} \hat{y} &= uG + \mathbf{d} \\ &= \begin{pmatrix} u_{1,1} & u_{1,2} & \dots & u_{1,128} \\ u_{2,1} & u_{2,2} & \dots & u_{2,128} \\ \vdots & \vdots & \ddots & \vdots \\ u_{10,1} & u_{10,2} & \dots & u_{10,128} \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{128} \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_{10} \end{pmatrix} \\ &= \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_{10} \end{pmatrix} \end{aligned} \quad (18)$$

Representasi dari arsitektur FFNN, dapat dilihat pada Gambar 2.4 berikut.



Gambar 2.3. Arsitektur *Feedforward Neural Network*

Gambar 2.4 tersebut merupakan representasi arsitektur FFNN dengan struktur berlapis yang terdiri dari satu lapisan input, dua lapisan tersembunyi (*hidden layer*), dan satu lapisan *output*. Gambar ini secara utuh menggambarkan alur transformasi data dari input mentah hingga menjadi output akhir.

2.7. Mean Squared Error (MSE)

Mean Squared Error (MSE) adalah metrik yang digunakan untuk mengukur kualitas hasil prediksi model dengan menghitung rata-rata dari kuadrat selisih antara nilai prediksi dan nilai aktual. Nilai MSE yang lebih kecil menunjukkan bahwa kesalahan prediksi model relatif rendah, sehingga model dianggap memiliki performa yang baik (Nugroho et al., 2024). Secara matematis, MSE dituliskan sebagai

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Pada formula tersebut, n adalah jumlah data, y_i merupakan nilai aktual, dan $\hat{y}_i$ merupakan nilai prediksi. Nugroho (2024) menyatakan bahwa MSE digunakan

sebagai fungsi kerugian (*loss function*) saat melatih model karena memberikan ukuran yang jelas terhadap kesalahan prediksi, serta sensitif terhadap error yang besar sehingga membantu proses konvergensi model secara lebih stabil dibandingkan MAE.

2.8. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah indikator yang digunakan untuk mengevaluasi kinerja model dengan cara menghitung rata-rata besarnya selisih antara nilai yang diprediksi dan nilai aktual. Nilai MAE yang semakin kecil menunjukkan prediksi model semakin mendekati data sebenarnya, sehingga tingkat akurasi semakin baik (Nugroho et al., 2024). Hubungan matematis MAE ditulis sebagai berikut:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Berbeda dengan Mean Squared Error (MSE) yang menghitung rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, MAE hanya mengambil nilai absolut dari selisih tersebut tanpa mengkuadratkan. Pada MAE, arah kesalahan tidak diperhitungkan karena yang diukur hanya besar selisih antara nilai prediksi dan nilai aktual. Baik kesalahan bernilai positif maupun negatif akan memiliki kontribusi yang sama terhadap nilai MAE (Nugroho et al., 2024).

2.9. Metrik Evaluasi Model

Evaluasi dan metrik performa merupakan aspek penting dalam menilai dan membandingkan suatu model. Melalui metrik ini, diperoleh gambaran mengenai seberapa baik model mampu menyelesaikan tugas yang diberikan, seperti proses klasifikasi atau segmentasi (Sathyanarayanan & Tantri, 2024). Menurut

Sathyanarayanan & Tantri (2024), beberapa indikator evaluasi yang umum digunakan dalam penilaian antara lain sebagai berikut:

1. Akurasi

Akurasi merupakan metrik evaluasi yang paling sederhana dan paling umum digunakan dalam klasifikasi. Metrik ini mengukur proporsi prediksi yang benar baik prediksi positif maupun negatif dari keseluruhan data uji. Secara matematis, akurasi dihitung dengan formula berikut:

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Di mana:

TP: True Positive (prediksi benar kelas positif)

TN : True Negative (Prediksi benar kelas negatif)

FP : False Positive (prediksi salah positif)

FN : False Negative (Prediksi salah negatif)

Meskipun akurasi mampu menunjukkan performa model secara keseluruhan, metrik ini berpotensi menimbulkan bias apabila data tidak seimbang atau proporsi tiap kelas tidak merata.

2. Presisi (*Precision*)

Presisi mengukur tingkat ketepatan model dalam memprediksi kelas positif. Nilai presisi menunjukkan seberapa banyak prediksi yang benar-benar positif dibandingkan dengan seluruh prediksi positif yang dihasilkan model. Formula presisi sebagai berikut:

$$\text{Presisi} = \frac{TP}{(TP + FP)}$$

3. Recall

Recall mengukur sejauh mana kemampuan model dalam menemukan seluruh data positif yang sebenarnya ada pada dataset. *Recall* tinggi menunjukkan bahwa model mampu menangkap sebagian besar kelas positif. Formula *Recall* sebagai berikut:

$$Recall = \frac{TP}{(TP + FN)}$$

4. F1_Score

F1_score merupakan ukuran gabungan antara presisi dan recall. Nilai ini dihitung dengan rata-rata harmonis dari presisi dan recall untuk memberikan keseimbangan antara keduanya. Formula F1_score sebagai berikut:

$$F1_Score = \frac{2 \times (\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})}$$

2.10. Integrasi Keislaman Prediksi Mutasi dengan Al-Qur'an & Hadits

Al-Qur'an tidak secara langsung membahas mengenai mutasi virus ataupun teknologi prediktif seperti *neural network*, namun terdapat ayat-ayat yang menekankan pentingnya menjaga kesehatan, keselamatan jiwa, dan penggunaan akal untuk menyelesaikan berbagai persoalan, termasuk penyakit menular. Dalam Islam, tubuh manusia adalah titipan dari Allah SWT. yang harus dirawat dan dijaga dengan baik. Hal ini sejalan dengan usaha para ilmuwan dalam memprediksi penyebaran penyakit melalui pendekatan berbasis teknologi dan sains. Seperti disebutkan dalam QS.Al-Baqarah ayat 195:

وَأَنْفَقُوا فِي سَبِيلِ اللَّهِ وَلَا تُلْقُوا بِأَيْدِيكُمْ إِلَى التَّهْلُكَةِ ۚ وَأَحْسِنُوا ۚ إِنَّ اللَّهَ يُحِبُّ
 الْمُحْسِنِينَ ١٩٥

“Dan infakkanlah (hartamu) di jalan Allah, dan janganlah kamu jatuhkan (diri sendiri) ke dalam kebinasaan dengan tangan sendiri, dan berbuat baiklah. Sungguh, Allah menyukai orang-orang yang berbuat baik.” (QS.Al-Baqarah: 195).

Ayat ini menjadi dasar anjuran untuk melakukan upaya perlindungan dari bahaya, termasuk wabah penyakit seperti COVID-19.

Menurut Tafsir Al-Muyassar, ayat tersebut menekankan bahwa setiap Muslim harus menjaga dirinya dan masyarakat dari hal-hal yang membahayakan (Mashudi, 2020). Allah SWT melarang perbuatan yang membawa pada at-tahlukah (التَّهْلُكَةُ), yaitu segala bentuk kebinasaan atau kehancuran, baik berupa kelalaian terhadap kewajiban, abai terhadap kesehatan, maupun membiarkan kerusakan menyebar. Dalam konteks ini, pengembangan model prediksi mutasi virus menjadi upaya preventif untuk membantu mengantisipasi munculnya varian berbahaya secara lebih cepat.

Upaya pencegahan tersebut juga ditegaskan dalam sebuah hadis Rasulullah SAW yang diriwayatkan dari Abu Hurairah RA:

حَدَّثَنَا عَبْدُ اللَّهِ بْنُ يُوسُفَ، قَالَ أَخْبَرَنَا مَالِكٌ، عَنْ أَبِي الزِّنَادِ، عَنِ الْأَعْرَجِ، عَنْ أَبِي هُرَيْرَةَ، أَنَّ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ: "وَإِذَا اسْتَيْقَظَ أَحَدُكُمْ مِنْ نَوْمِهِ فَلْيَغْسِلْ يَدَهُ قَبْلَ أَنْ يُدْخِلَهَا فِي وَضُوئِهِ، فَإِنْ أَحَدُكُمْ لَا يَدْرِي أَيْنَ بَاتَتْ يَدُهُ" (رواه البخاري ١٦٢).

“Abdullah bin Yusuf telah menceritakan kepada kami. Ia berkata: Malik telah mengabarkan kepada kami, dari Abu az-Zinad, dari al-A‘raj, dari Abu Hurairah, bahwa Rasulullah ﷺ bersabda:

‘Dan apabila salah seorang di antara kalian bangun dari tidurnya, maka hendaklah ia membasuh tangannya sebelum memasukkannya ke dalam air wudhunya, karena salah seorang dari kalian tidak mengetahui di mana tangannya bermalam.’” (HR. Bukhari no. 162).

Hadis ini menekankan bahwa menjaga kebersihan, baik jasmani maupun rohani, merupakan bagian penting dari keimanan seorang Muslim. Dalam konteks

modern, termasuk pada masa pandemi COVID-19, menjaga kebersihan seperti mencuci tangan, menjaga lingkungan tetap steril, dan menghindari tempat yang kotor merupakan langkah nyata untuk menghindari penularan virus. Dengan demikian, perintah menjaga kebersihan bukan sekadar tuntutan kesehatan, melainkan juga kewajiban agama yang memiliki nilai spiritual tinggi.

Selain itu, Rasulullah SAW juga memberikan arahan dalam menghadapi wabah dengan Hadits berikut:

حَدَّثَنَا عَبْدُ الْعَزِيزِ بْنُ عَبْدِ اللَّهِ، قَالَ حَدَّثَنِي مَالِكٌ، عَنْ مُحَمَّدِ بْنِ الْمُنْكَدِرِ، وَعَنْ أَبِي النَّضْرِ، مَوْلَى عُمَرَ بْنِ عَبْدِ اللَّهِ عَنْ عَامِرِ بْنِ سَعْدٍ بْنِ أَبِي وَقَّاصٍ، عَنْ أَبِيهِ، أَنَّهُ سَمِعَهُ يُسْأَلُ، أُسَامَةَ بْنَ زَيْدٍ مَاذَا سَمِعْتَ مِنْ رَسُولِ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ فِي الطَّاعُونَ فَقَالَ أُسَامَةُ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ " الطَّاعُونَ رَجَسٌ أُرْسِلَ عَلَى طَائِفَةٍ مِنْ بَنِي إِسْرَائِيلَ أَوْ عَلَى مَنْ كَانَ قَبْلَكُمْ، فَإِذَا سَمِعْتُمْ بِهِ بِأَرْضٍ فَلَا تَقْدُمُوا عَلَيْهِ، وَإِذَا وَقَعَ بِأَرْضٍ وَأَنْتُمْ بِهَا فَلَا تَخْرُجُوا فِرَارًا مِنْهُ" (رواه البخاري ٣٤٧٣).

"Abdul Aziz bin Abdullah telah menceritakan kepada kami. Ia berkata: Malik telah menceritakan kepadaku, dari Muhammad bin al-Munkadir, dan dari Abu an-Nadhr, maula 'Umar bin 'Ubaidillah, dari 'Amir bin Sa'd bin Abi Waqqash, dari ayahnya (Sa'd bin Abi Waqqash), bahwa ia mendengar Sa'd bertanya kepada Usamah bin Zaid:

'Apa yang engkau dengar dari Rasulullah ﷺ tentang tha'un (wabah)?'

Maka Usamah menjawab: Rasulullah ﷺ bersabda:

'Tha'un adalah suatu kotoran (azab) yang dikirimkan kepada suatu kelompok dari Bani Israil atau kepada orang-orang sebelum kalian. Apabila kalian mendengar tha'un terjadi di suatu negeri maka janganlah kalian memasukinya. Dan apabila tha'un terjadi di negeri tempat kalian berada, maka janganlah kalian keluar darinya untuk melarikan diri darinya.'"(HR. Bukhari no. 3473).

Hadis ini menunjukkan bahwa prinsip karantina telah diajarkan Rasulullah SAW sejak 14 abad lalu, dan tetap relevan dalam penanganan pandemi modern karena sejalan dengan praktik pembatasan mobilitas dan isolasi wilayah untuk mencegah penyebaran COVID-19. Prediksi mutasi virus dapat dipandang sebagai ikhtiar manusia untuk lebih cepat mengambil langkah perlindungan sebelum kerusakan meluas. Integrasi antara Al-Qur'an, hadis, dan sains modern mencerminkan semangat Islam yang menekankan pentingnya menjaga jiwa,

kesehatan, dan memanfaatkan ilmu demi kemaslahatan umat. Hal ini sejalan dengan firman Allah SWT dalam QS. Al-Mujadilah ayat 11:

وَإِذَا قِيلَ انْشُرُوا فَاَنْشُرُوا يَرْفَعُ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ ۖ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ ۚ
وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ ﴿١١﴾

“..Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat. Allah melihat pada apa yang kamu kerjakan.” (QS.Al-Mujadilah: 11)

Ayat ini menegaskan bahwa ilmu bukan hanya bernilai duniawi, tetapi juga menjadi sarana memperoleh kemuliaan di sisi Allah SWT. Karena itu, penelitian mutasi SARS-CoV-2 melalui model prediksi berbasis Feed Forward Neural Network dapat dipandang sebagai ibadah ketika ditujukan untuk mencegah kerusakan dan menjaga kehidupan.

2.11. Kajian Prediksi Mutasi Virus dengan Teori Pendukung

Prediksi mutasi virus, khususnya pada SARS-CoV-2 memerlukan pemahaman dasar mengenai struktur dan karakteristik biologis dari virus itu sendiri. SARS-CoV-2 memiliki empat protein struktural utama, yaitu; Spike (S), Envelope (E), Membrane (M), dan Nukleokapsid (N). Di antara keempatnya, protein spike berperan penting karena bertanggung jawab dalam proses masuknya virus ke dalam sel inang melalui pengikatan dengan reseptor ACE2. Pada protein spike ini terdapat *cleavage site*, yaitu titik pemotongan oleh enzim inang furin. Pemotongan tersebut menentukan aktivitas spike dan mempengaruhi kemampuan virus menginfeksi sel. Mutasi di sekitar situs *cleavage* telah terbukti meningkatkan infektivitas dan kemudahan penularan, sehingga prediksi mutasi di area ini sangat penting untuk memahami evolusi virus dan mangantisipasi lonjakan kasus.

Dalam konteks data, urutan genetik virus dapat direpresentasikan sebagai data sekuensial. Oleh karena itu, pendekatan *machining learning* yang sesuai adalah menggunakan model yang mampu menangani urutan, salah satunya adalah jaringan saraf tiruan (*artificial neural network/ANN*). Salah satu arsitektur dasar yang dapat digunakan adalah *Feedforward Neural Network* (FFNN), yang bekerja dengan menghubungkan lapisan input ke lapisan output melalui satu atau lebih lapisan tersembunyi untuk mempelajari pola dalam data sekuensial.

BAB III METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif karena semua variabel penelitian dinyatakan dalam angka dan dianalisis menggunakan teknik statistik serta pemodelan matematis.

3.2. Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data yang diperoleh dari sumber yang telah ada sebelumnya dan tersedia secara publik. Data diakses melalui NCBI Virus Database (National Center for Biotechnology Information) yang menyediakan informasi sekuens virus secara global, termasuk SARS-CoV-2. Fokus utama dalam penelitian ini adalah sekuens *Surface Glycoprotein* (spike protein) dari virus SARS-CoV-2, khususnya pada situs cleavage yang berperan penting dalam proses pematangan protein dan infektivitas virus. Data diunduh dalam format FASTA.

Proses penyaringan data mencakup seleksi berdasarkan *host* yaitu manusia untuk memastikan relevansi biologis, kelengkapan sekuens untuk menjaga integritas data, penghapusan sekuens dengan karakter ambigu untuk menghindari kesalahan analisis, serta pemilihan sekuens yang tergolong dalam *Pango lineage* Omicron (B.1.1529) untuk memfokuskan studi pada varian yang menjadi perhatian. Setelah menerapkan proses penyaringan di NCBI diperoleh total 1.914 sekuens spike protein. Dataset ini digunakan untuk mendeteksi mutasi di sekitar situs cleavage dan menganalisis perbedaan pola sekuens antarvarian.

3.3. Teknik Pengumpulan Data

Teknik pengambilan data dilakukan dengan langkah-langkah berikut:

1. Akses basis data

Akses situs NCBI Virus untuk memperoleh data sekuen yang relevan dengan objek kajian. Situs NCBI Virus dapat di akses pada link berikut;
<https://url-shortener.me/2TLG>

2. Navigasi ke visual data dashboard

Pada halaman utama, pilih menu *visual data dashboard*, kemudian klik opsi *All Nucleotides* untuk menampilkan seluruh data sekuens virus dalam bentuk nukleotida.

3. Penyaringan berdasarkan jenis virus

Pada bagian *Refine Results*, dilakukan penyaringan berdasarkan jenis virus dengan memilih opsi *Virus/Taxonomy*, kemudian memasukkan *Severe acute respiratory syndrome coronavirus 2* (TaxID: 2697049) untuk memfokuskan data pada SARS-CoV-2.

4. Penyaringan berdasarkan protein target

Aktifkan filter *Has Protein* pada panel *Refine Results*, kemudian ketik *Surface glycoprotein* untuk memperoleh data terkait Spike protein.

5. Penyaringan berdasarkan varian (*lineage*)

Masih pada panel *Refine Results*, pilih opsi *Pango Lineage* dan masukkan kode *lineage* varian Omicron yaitu B.1.1.529 guna menyaring sekuens berdasarkan varian spesifik. Setelah itu, buka kolom protein.

6. Penyaringan berdasarkan *host* (*Human*)

Pada panel *Refine Results*, dilakukan penyaringan berdasarkan *host* dengan

memilih opsi *Host*, kemudian mengetik dan memilih *Homo sapiens* (*human*), TaxID: 9606. Langkah ini memastikan bahwa seluruh data virus yang diperoleh berasal dari isolat manusia, bukan dari hewan atau sumber *host* lainnya.

7. Penyaringan berdasarkan karakter ambigu

Aktifkan filter untuk *Ambiguous Characters* dan tetapkan maksimum 0 untuk memastikan sekuens yang diperoleh tidak mengandung huruf atau simbol yang tidak jelas, sehingga mengurangi kesalahan pada analisis mutasi.

8. Pengunduhan data

Setelah data disaring, sekuens diunduh dalam format FASTA. Proses pengunduhan dilakukan dengan memilih opsi *Download All Results*, kemudian memilih protein dan mengklik *Download All Records*. Selanjutnya digunakan *Build Custom* untuk menentukan informasi yang ingin disertakan yaitu *Accession*, *GenBank Title*, dan *Collection Date*.

3.4. Teknik Analisis Data

Analisis dalam penelitian ini dilakukan menggunakan Python sebagai bahasa pemrograman utama. Adapun teknik analisis data adalah sebagai berikut:

1. Pra-pemrosesan data

- a. Penyelarasan sekuens dilakukan menggunakan MAFFT dengan data sekuens protein Spike varian Omicron dan sekuens referensi Wuhan (YP_009724390.1) sebagai berkas rujukan. Proses dilakukan melalui opsi *Align long sequences to a short MSA* agar sekuens panjang dari varian Omicron diselaraskan mengikuti struktur sekuens referensi, sehingga

posisi konservatif tetap terjaga, dengan mengunggah sekuens referensi pada bagian *Reference alignment file* dan sekuens Omicron pada *Additional sequences file*. Setelah itu, penyelarasan dijalankan dengan menekan tombol Submit, dan hasil alignment diperoleh dalam format FASTA untuk analisis lebih lanjut.

- b. Ekstraksi wilayah lokal di sekitar *furin cleavage site* (motif PRRAR) dilakukan secara terotomatisasi menggunakan pustaka Biopython, dengan menetapkan sejumlah asam amino sebelum dan sesudahnya.
- c. Hasil ekstraksi kemudian disusun ke dalam struktur DataFrame menggunakan pustaka pandas dan diurutkan berdasarkan informasi tanggal yang diperoleh dari header file FASTA.
- d. Setiap asam amino dikonversi menjadi vektor integer melalui *integer encoding* sebagai bentuk representasi numerik untuk input model.

2. Pengaturan Data

Dataset yang telah direpresentasikan secara numerik kemudian disusun menjadi matriks fitur, di mana setiap baris menjadi satu sampel dan setiap kolom merepresentasikan posisi residu dalam sekuens. Proses pengaturan ini mengikuti alur sebagai berikut:

a. *Windowing*

Pembentukan *window* sesuai konsep pada Persamaan (2) – (4), yaitu mengambil elemen sekuens secara berurutan sebagai input dan elemen berikutnya sebagai target. Dengan *lag* 8, delapan residu digunakan untuk memprediksi satu residu selanjutnya.

b. *Flattening*

Setelah *window* terbentuk, setiap *window* diratakan menjadi vektor satu dimensi mengikuti Persamaan (5) – (7). Pada penelitian ini, *window* berisi 8 residu disusun menjadi vektor panjang 8, sehingga setiap sampel memiliki format numerik yang sesuai untuk pemrosesan oleh jaringan saraf.

c. Pembentukan Input dan Target

Tahap selanjutnya menggabungkan vektor hasil *flattening* menjadi matriks input dan seluruh target ke dalam matriks keluaran, sesuai konsep pada Persamaan (8) dan (9). Dengan penyusunan vertikal ini terbentuk matriks X dan Y, sehingga seluruh pasangan data dapat diproses secara serempak oleh model.

3. Pemisahan dataset

Dataset dibagi menjadi tiga bagian, yaitu data latih (70%), data validasi (15%) dan data uji (15%). Pembagian ini bertujuan untuk memastikan proses pelatihan, validasi, dan evaluasi dilakukan secara terpisah sehingga penilaian kinerja model berlangsung obyektif.

4. Desain arsitektur FFNN

Model *Feedforward Neural Network* (FFNN) dikembangkan menggunakan framework TensorFlow pada platform Kaggle. Arsitektur jaringan terdiri dari tiga lapisan Dense dengan satu lapisan Dropout yang ditempatkan setelah lapisan pertama. Rinciannya adalah sebagai berikut:

a. Lapisan pertama (*Dense*, 256 unit)

Lapisan ini berperan sebagai pemrosesan awal untuk mengekstraksi fitur,

dengan menerapkan operasi linear sesuai Persamaan (11) – (13) dan mengalirkan hasilnya ke fungsi aktivasi pada Persamaan (14).

b. Lapisan *Dropout* (0,1)

Dropout ditempatkan setelah lapisan pertama untuk mengurangi risiko *overfitting* dengan menonaktifkan sebagian neuron secara acak selama pelatihan.

c. Lapisan kedua (*Dense*, 128 unit)

Lapisan ini menyempurnakan representasi fitur sebelum menuju lapisan output. *Dropout* setelah lapisan pertama mengurangi *overfitting* dengan menonaktifkan sebagian neuron secara acak. Operasi linear sesuai Persamaan (15) – (16) kemudian diaktifkan kembali mengikuti Persamaan (17) untuk memperdalam ekstraksi pola.

d. Lapisan output (*Dense*, 10 unit)

Memproyeksikan fitur lapisan kedua ke 10 unit keluaran menggunakan operasi linear akhir sebagaimana dinyatakan pada Persamaan (18) tanpa aktivasi tambahan.

Pada tahap pelatihan menggunakan fungsi kerugian Mean Squared Error (MSE), Mean Absolute Error (MAE) dan algoritma optimasi Adam.

5. Prediksi mutasi

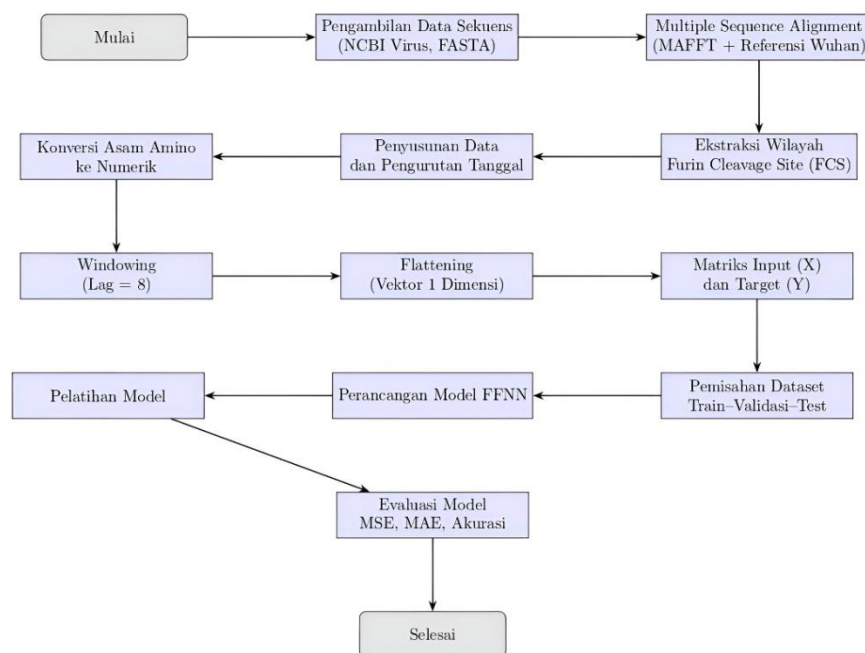
Setelah model dilatih dan dievaluasi, model digunakan untuk memprediksi potensi mutasi pada sekuens spike protein baru. Input berupa sekuens yang telah melalui pra-pemrosesan, sedangkan output yang dihasilkan berupa nilai numerik yang merepresentasikan kemungkinan mutasi di sekitar FCS. Nilai keluaran tersebut kemudian dikonversi kembali ke huruf

asam amino melalui proses *decoding*, sehingga prediksi akhir tampil dalam bentuk sekuens biologis yang mudah diinterpretasikan.

6. Evaluasi model

Model tersebut dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1_score*. Akurasi menunjukkan ketepatan prediksi secara keseluruhan, presisi menggambarkan proporsi prediksi positif yang benar, *recall* menunjukkan kemampuan model dalam menemukan seluruh kasus mutasi, sedangkan *F1_score* menyeimbangkan presisi dan *recall* melalui rata-rata harmonik. Selain itu, evaluasi dilakukan dengan membandingkan kinerja FFNN terhadap model baseline sebagai tolok ukur objektif, serta dilanjutkan dengan analisis kesalahan dan identifikasi hotspot mutasi untuk memahami pola prediksi dan keterbatasan model.

Untuk mempermudah pemahaman terhadap tahapan penelitian ini, alur teknik analisis data pada disajikan dalam bentuk flowchart berikut.



Gambar 3 1 Flowchart Teknik Analisis Data

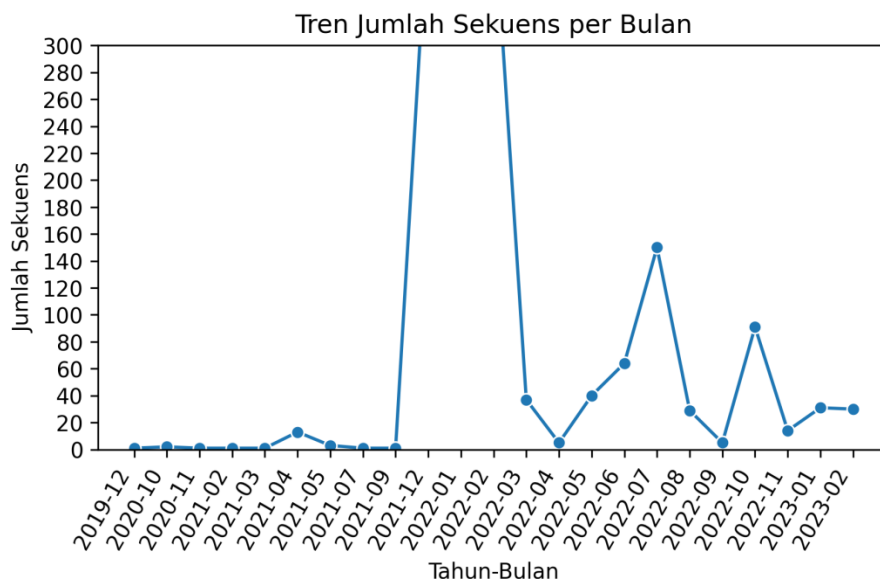
BAB IV HASIL DAN PEMBAHASAN

4.1 Penerapan Model FFNN Pada Prediksi Mutasi

Pada bagian ini dibahas penerapan model *Feedforward Neural Network* (FFNN) dalam melakukan prediksi mutasi. Model FFNN digunakan untuk mempelajari pola perubahan data sekuensial berdasarkan informasi historis sehingga mampu menghasilkan prediksi mutasi pada langkah berikutnya.

4.1.1 Deskripsi Data Penelitian

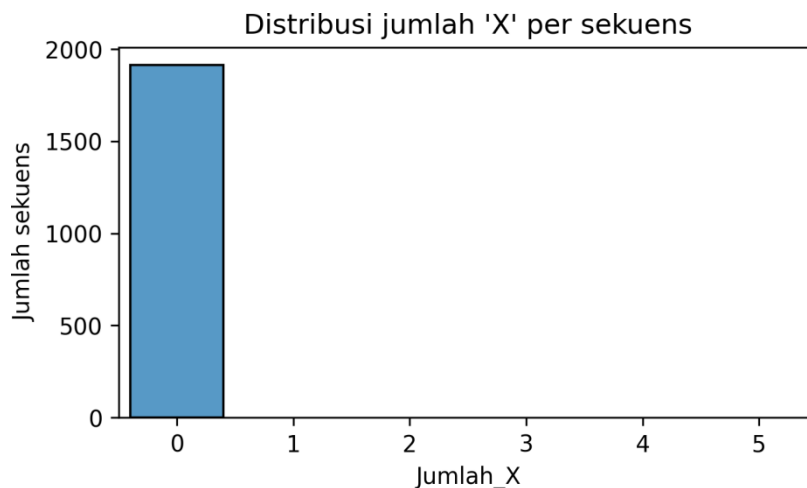
Setiap sekuens dilengkapi dengan metadata yang mencakup tanggal isolasi dan identitas varian. Dari keseluruhan data, sebanyak 1.914 sekuens memiliki informasi waktu isolasi yang lengkap. Rentang waktu pengumpulan data mencakup Desember 2019 hingga Februari 2023.



Gambar 4.1 Tren Jumlah Sekuens SARS-CoV-2 Per Bulan

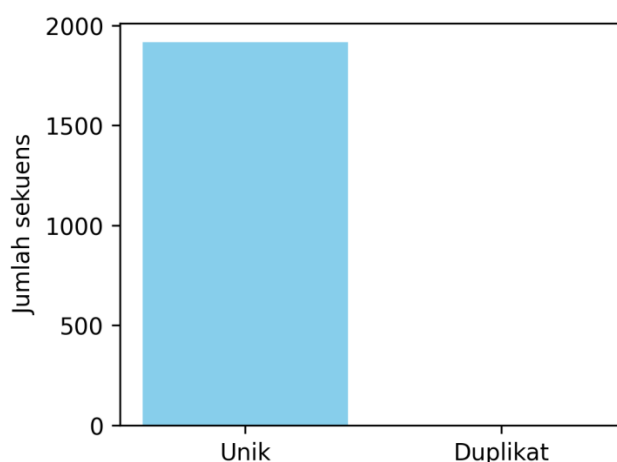
Gambar 4.1 menunjukkan distribusi jumlah sekuens berdasarkan waktu isolasi. Pola tersebut merepresentasikan fase awal kemunculan virus hingga masa dominasi varian Omicron yang meningkat tajam pada akhir 2021 hingga

awal 2022 serta menggambarkan tren pengumpulan data yang mengikuti dinamika surveilans genomik global terhadap SARS-CoV-2.

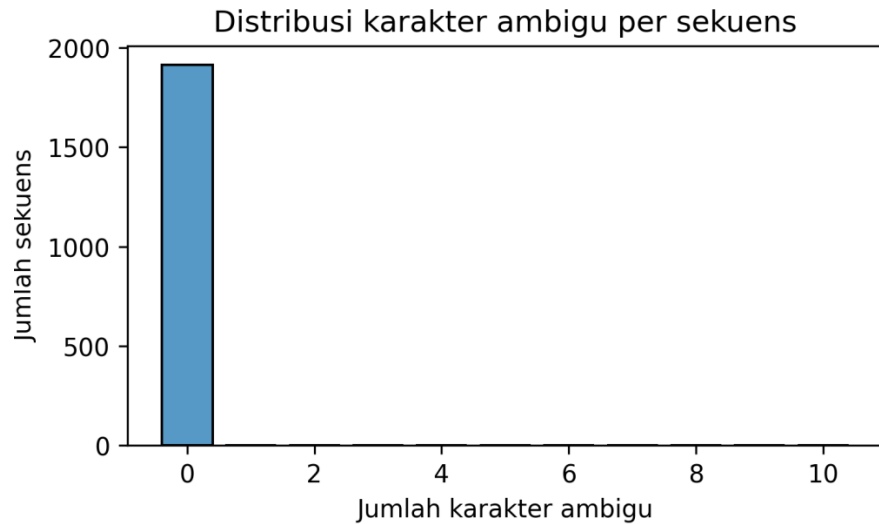


Gambar 4.2 Distribusi Jumlah 'X' Per Sekuens SARS-CoV-2

Berdasarkan pemeriksaan awal, seluruh sekuens protein menunjukkan kualitas yang sangat baik, ditandai dengan tidak ditemukannya residu tidak teridentifikasi ('X') pada keseluruhan dataset. Hal ini tampak pada Gambar 4.2, di mana distribusi jumlah 'X' hanya berada pada nilai nol, menandakan bahwa setiap sekuens memiliki tingkat kelengkapan yang optimal dan layak digunakan untuk tahap analisis berikutnya.

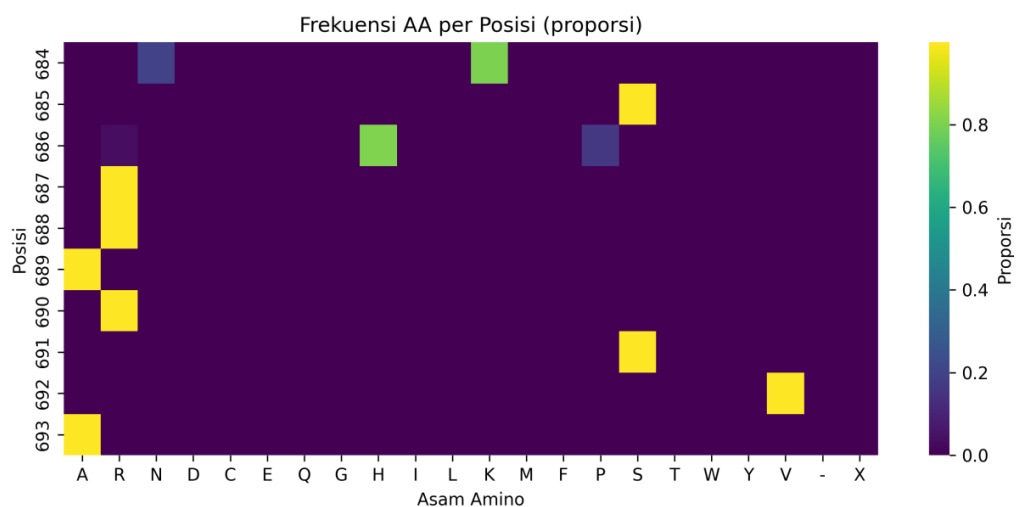


Gambar 4. 3 Histrogram Jumlah Duplikasi Sekuens SARS-CoV-2



Gambar 4. 4 Distrubi Karakter Ambigu Per Sekuens SARS-CoV-2

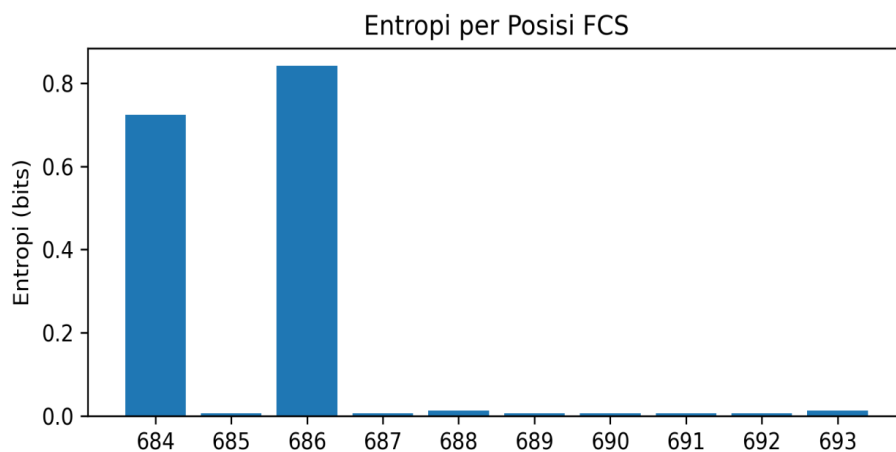
Selain pemeriksaan metadata, dilakukan pula pemeriksaan terhadap keberadaan sekuens duplikat dan karakter ambigu. Hasil pemeriksaan sekuens duplikat ditunjukkan pada Gambar 4.3, di mana seluruh 1.915 sekuens teridentifikasi sebagai unik tanpa adanya baris duplikat. Selanjutnya, pemeriksaan karakter ambigu yang meliputi tanda hubung (–) dan titik (.) ditampilkan pada Gambar 4.4, yang menunjukkan bahwa seluruh sekuens tidak mengandung karakter ambigu. Kondisi ini mengindikasikan bahwa dataset telah bersih dan konsisten.



Gambar 4.5 Heatmap Frekuensi Asam Amino per Posisi pada *Region* FCS

Setelah penyelarasan menggunakan MAFFT, area FCS yang dianalisis mencakup 10 residu (posisi 684–693 terhadap referensi Wuhan) dan dievaluasi berdasarkan distribusi frekuensi serta variabilitas setiap posisi. Hasil visualisasi pada Gambar 4.5 menunjukkan heatmap frekuensi kemunculan asam amino pada setiap posisi. Sebagian besar posisi, hanya memiliki satu residu dominan dengan proporsi mendekati 1,0, yang menunjukkan sifatnya yang sangat konservatif. Sebaliknya, posisi 687 dan 689 menampilkan variasi frekuensi yang lebih tinggi dengan lebih dari satu residu muncul dalam proporsi signifikan, sehingga dapat diidentifikasi sebagai hotspot mutasi.

Analisis ini diperkuat oleh hasil perhitungan entropi pada Gambar 4.6. Nilai entropi yang tinggi menunjukkan variabilitas yang besar, sedangkan nilai mendekati 0 mencerminkan stabilitas residu. Sesuai hasil visualisasi, posisi 684 (0,72) dan posisi 686 (0,84) memiliki entropi tertinggi dan dapat diidentifikasi sebagai hotspot mutasi. Sebaliknya, posisi lainnya seperti 687–693 memiliki entropi mendekati nol, menandakan bahwa residu pada posisi tersebut sangat stabil. Temuan ini menegaskan bahwa variasi pada region FCS hanya terkonsentrasi pada sebagian kecil posisi.



Gambar 4.6 Nilai Entropi Asam Amino pada Setiap Posisi FCS

4.1.2 Pra-Pemrosesan Data

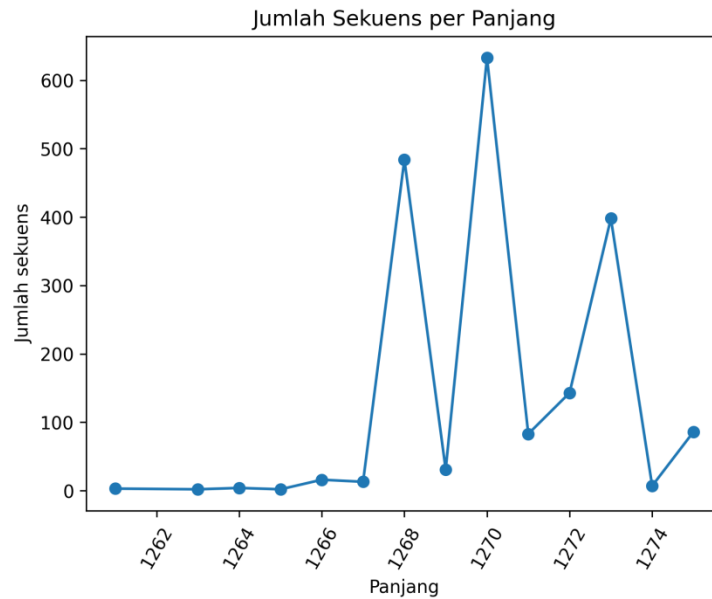
Data sekuens protein spike varian Omicron dan sekuens referensi Wuhan (YP_009724390.1) diproses dan diselaraskan menggunakan MAFFT untuk memastikan setiap residu berada pada posisi yang konsisten.

Tabel 4.1 Cuplikan Data Pertama Sebelum Proses Penyelarasan

	<i>ID</i>	<i>Sequence</i>	<i>Length</i>
0	YP_009724390.1	...QTQTNSPRRARSVASQSIIAYTM...	1273
1	XKD21093.1	...QTQTKSHRRARSVASQSIIAYTM..	1273
2	XKB59856.1	...QTNSHRRARSVASQSIIAYTMSL...	1271
3	XKB59892.1	...QTQTKSHRRARSVASQSIIAYTM..	1273
4	XKB61116.1	...SHRRARSVASQSIIAYTMSLGAE...	1268
...	...	...	...
1914	QUA30928.1	...TNSHRRARSVASQSIIAYTMSLG...	1048

Cuplikan data sebelum proses penyelarasan ditampilkan pada Tabel 4.1, yang menunjukkan bahwa setiap sekuens memiliki panjang yang berbeda. Untuk melihat data secara lebih lengkap, seluruh detailnya disajikan pada bagian Lampiran 1. Variasi panjang ini membuat proses penyelarasan (*Multiple Sequence Alignment/MSA*) diperlukan untuk menyamakan posisi residu, sehingga setiap kolom merepresentasikan posisi yang homolog dan analisis mutasi dapat dilakukan secara konsisten.

Gambar 4.7 menampilkan distribusi jumlah sekuens berdasarkan panjangnya. Terlihat bahwa sebagian besar sekuens berada pada rentang 1268–1274, dengan panjang 1270 dan 1268 sebagai kelompok yang paling dominan. Pola ini menunjukkan bahwa meskipun terdapat sedikit variasi panjang, mayoritas sekuens tetap berada pada kisaran panjang yang relatif konsisten.



Gambar 4.7 Distribusi Jumlah Per-Panjang sekuens Furin SARS-CoV-2

Selanjutnya melakukan proses penyelarasan menggunakan MAFFT untuk menyamakan posisi residu pada seluruh sekuens. Setiap sekuens kemudian disejajarkan sehingga gap (“–”) ditambahkan pada posisi tertentu untuk menyesuaikan perbedaan panjang. Dengan cara ini, residu yang homolog berada pada kolom yang sama, mutasi dapat diidentifikasi secara tepat, dan seluruh sekuens siap digunakan untuk analisis lanjutan.

Tabel 4.2 Cuplikan Data Pertama Setelah Proses Penyelarasan di Sekuens

	<i>ID</i>		<i>Sequence</i>	<i>Length</i>
0	YP_009724390.1	2019-12	...TQTNSP RR ARS V ASQS...	1279
1	XKD21093.1	2022-07-20	...TQT K SH RR ARS V ASQS...	1279
2	XKB59856.1	2023-01-01	...TQTNSH RR ARS V ASQS...	1279
3	XKB59892.1	2023-01-01	...TQT K SH RR ARS V ASQS...	1279
4	XKB61116.1	2023-01-01	...TQTNSH RR ARS V ASQS...	1279
...	...		...	...
1914	QUA30928.1	2021-02-24	...TQTNSH RR ARS V ASQS...	1279

Tabel 4.2 menampilkan hasil penyelarasan sekuens protein Spike yang memuat identitas sampel, sekuens hasil *alignment*, serta panjang masing-masing sekuens. Seluruh sampel memiliki panjang yang konsisten, yaitu 1279 residu, yang menunjukkan bahwa proses penyelarasan berhasil mempertahankan struktur sekuens. Selanjutnya, sekuens diurutkan berdasarkan tanggal pengambilan sampel untuk merepresentasikan urutan temporal kemunculan mutasi, kemudian dilakukan ekstraksi residu pada posisi FCS yang hasilnya disajikan pada Tabel 4.3.

Tabel 4.3 Cuplikan Data Sekuens Pertama Hasil Konversi Asam Amino ke Representasi Numerik

	Tanggal	FCS_Sequence	Seq_int
0	2019-12-01	NSPRRARSVA	[3, 16, 15, 2, 2, 1, 2, 16, 20, 1]
1	2020-10-20	NSPRRARSVA	[3, 16, 15, 2, 2, 1, 2, 16, 20, 1]
2	2020-10-26	KSHRRARSVA	[12, 16, 9, 2, 2, 1, 2, 16, 20, 1]
3	2020-11-23	KSHRRARSVA	[12, 16, 9, 2, 2, 1, 2, 16, 20, 1]
4	2021-02-24	NSHRRARSVA	[3, 16, 9, 2, 2, 1, 2, 16, 20, 1]
...	...	...	...
1914	2023-02-13	NSPRRARSVA	[3, 16, 15, 2, 2, 1, 2, 16, 20, 1]

Tabel 4.3 menampilkan penyelarasan region furin cleavage site (FCS) pada residu 684–693 sehingga variasi residu dapat diamati secara jelas. Sekuens referensi Wuhan menunjukkan pola NSPRRARSV, sedangkan varian Omicron memperlihatkan substitusi khas pada wilayah tersebut. Sekuens kemudian dikonversi ke bentuk numerik agar dapat diproses oleh model machine learning, dengan pemetaan asam amino sebagai berikut: A=1, R=2, N=3, D=4, C=5, E=6, Q=7, G=8, H=9, I=10, L=11, K=12, M=13, F=14, P=15, S=16, T=17, W=18,

$Y=19$, dan $V=20$, sedangkan karakter gap (“-”) dan residu ambigu (“X”) direpresentasikan dengan nilai 0.

4.1.3 Pengaturan Data

Berdasarkan Tabel 4.1 pada bagian *sequence integer*, data disusun menjadi matriks fitur, kemudian diproses menggunakan windowing dengan lag 8, di mana delapan residu berurutan digunakan sebagai input untuk memprediksi residu ke-9 sebagai target. Bentuk matriks hasil konversi ditunjukkan berikut.

$$H = \begin{bmatrix} 3 & 16 & 15 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \\ 3 & 16 & 15 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \\ 12 & 16 & 9 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \\ 12 & 16 & 9 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \\ 3 & 16 & 9 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 3 & 16 & 15 & 2 & 2 & 1 & 2 & 16 & 20 & 1 \end{bmatrix}$$

Proses pengolahan data dimulai dengan menerapkan *windowing* pada matriks H menggunakan *lag* 8, di mana delapan residu berurutan digunakan sebagai masukan dan residu berikutnya sebagai target. Setiap hasil windowing kemudian diratakan melalui *flattening* menjadi vektor baris yang disusun ke dalam matriks X , sementara targetnya dihimpun ke dalam matriks Y . Penggunaan lag 8 bertujuan menangkap konteks lokal dari pola perubahan residu berdasarkan informasi historis yang cukup, sehingga pasangan data (x_i, y_i) yang merepresentasikan hubungan antara konteks residu sebelumnya dan residu yang diprediksi, dan pasangan inilah yang digunakan dalam tahap pelatihan model.

Tabel 4.4 Skema Pembagian Dataset SARS-CoV-2

Lag	n_seq	Dimensi		Train	Validasi	Test
		Input	Output			
8	1915	80	10	1334	286	287

Tabel 4.4 menunjukkan bahwa pemisahan dataset diperoleh 1.334 sekuens yang siap digunakan sebagai data pembelajaran. Dengan lag sebesar 8, setiap sampel menggunakan delapan residu sebagai input untuk memprediksi residu berikutnya sebagai output. Konversi ke representasi numerik menghasilkan 80 fitur input dan 10 fitur output pada setiap sampel. Dari total sekuens tersebut terbentuk 1.915 *window* data, yang kemudian dibagi secara berurutan menjadi 1.334 data pelatihan (70%), 286 data validasi (15%), dan 287 data pengujian (15%), sehingga kontinuitas perubahan mutasi tetap terjaga dan risiko data leakage dapat dihindari.

4.1.4 Hiperparameter Pelatihan dan Penerapan Model FFNN

Pada tahap ini ditetapkan hiperparameter untuk mengatur proses pelatihan model. Hiperparameter tersebut dirangkum pada Tabel 4.5 berikut.

Tabel 4.5 Hiperparameter pada Proses Pelatihan Model

Optimizer	Learning rate	Batch Size	Epoch maks	Patience (Early Stopping)	Dropout
Adam	0.001	32	200	10	0.1
SGDM	0.010	32	200	10	0.1
RMSprop	0.001	32	200	10	0.1

Berdasarkan hasil eksperimen perbandingan optimizer yang ditunjukkan pada Tabel 4.6, optimizer Adam menunjukkan kinerja terbaik dibandingkan SGDM dan RMSprop. Adam menghasilkan nilai kesalahan yang lebih rendah

baik pada data latih maupun data validasi, dengan MSE validasi sebesar 0.006152 dan MAE validasi sebesar 0.030855. Meskipun SGDM memiliki selisih kesalahan (*gap*) yang relatif kecil antara data latih dan validasi, nilai MAE validasinya masih lebih tinggi, yaitu sebesar 0.040673. Sementara itu, RMSprop menunjukkan perbaikan dibandingkan SGDM dengan MAE validasi sebesar 0.033883, namun tetap berada di atas Adam. Secara keseluruhan, Adam dipilih sebagai *optimizer* terbaik karena mampu mencapai kesalahan prediksi yang lebih rendah dengan tingkat generalisasi yang tetap terjaga, sehingga lebih efektif dalam memodelkan pola mutasi sekuens yang bersifat tidak stasioner.

Tabel 4.6 Eksperimen Optimizer pada Model FFNN

Optimizer	MSE			MAE		
	Train	Val	Gap	Train	Val	Gap
Adam	0.004966	0.006152	0.001186	0.026089	0.030855	0.004767
SGDM	0.005596	0.006373	0.000778	0.037874	0.040673	0.002799
RMSprop	0.005095	0.005972	0.000877	0.029478	0.033883	0.004405

Arsitektur model FFNN yang digunakan ditunjukkan pada Tabel 4.7, yang terdiri atas dua lapisan dense tersembunyi dengan masing-masing 256 dan 128 neuron, diikuti oleh satu lapisan dense keluaran dengan 10 neuron. Jumlah total parameter yang digunakan dalam model ini adalah 54.922 parameter, yang menunjukkan bahwa model memiliki kompleksitas yang relatif moderat sehingga tetap efisien secara komputasi namun cukup representatif untuk menangkap pola mutasi antarresidu pada dataset SARS-CoV-2.

Tabel 4.7 Implementasi Model FFNN pada Dataset SARS-CoV-2

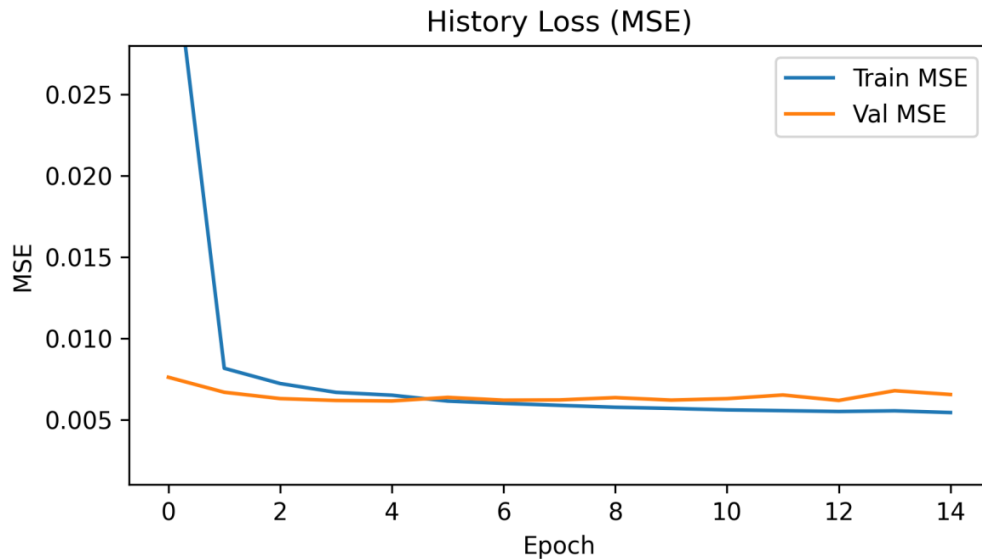
Layer	Type	Output Shape	Params
1	Dense	(None, 256)	20736
2	Dropout	(None, 256)	0

3	Dense	(None, 128)	32896
4	Dense	(None, 10)	1290

4.1.5 Hasil Pelatihan dan Konvergensi Loss

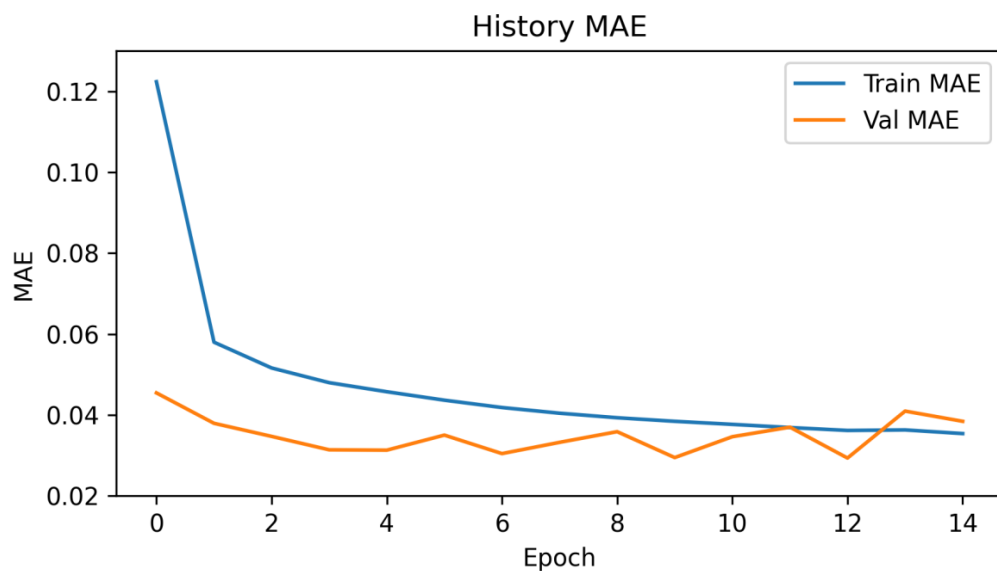
Pada penelitian ini, model dilatih selama 30 epoch dengan menggunakan mekanisme *early stopping* untuk mencegah *overfitting*. Berdasarkan hasil pelatihan, nilai *loss* (MSE) pada data pelatihan menurun sangat cepat dari nilai awal yang tinggi menuju sekitar 0.005–0.006, kemudian stabil hingga akhir proses pelatihan. Sementara itu, nilai MSE pada data validasi juga stabil pada kisaran 0.006–0.007, tanpa menunjukkan tren peningkatan yang signifikan. Nilai MAE pelatihan turun secara bertahap dari sekitar 0.12 menuju 0.036, sedangkan MAE validasi berada di kisaran 0.03–0.04 dengan fluktuasi ringan. Pola ini menunjukkan bahwa model belajar secara efektif dengan risiko *overfitting* yang rendah.

Berdasarkan karakteristik kurva pembelajaran, Barinov et al., (2023) menegaskan bahwa *overfitting* terjadi apabila *training loss* terus menurun secara konsisten, tetapi *validation loss* berhenti menurun atau justru mulai meningkat, sehingga *generalization gap* membesar dari waktu ke waktu. Sebaliknya, kondisi ideal ditunjukkan ketika kedua kurva *loss* bergerak menuju keadaan stabil dengan *generalization gap* yang minimal atau mendekati nol. Pada hasil penelitian ini, *validation loss* tidak menunjukkan tren kenaikan yang berkelanjutan dan *generalization gap* tidak membesar dari *epoch* ke *epoch*, sehingga model dapat disimpulkan berada dalam kondisi pelatihan yang stabil dan tidak mengalami *overfitting*.



Gambar 4.8 Kurva *History* MSE Model FFNN Selama Proses Pelatihan

Gambar 4.9 memperlihatkan bahwa kurva MSE pelatihan turun tajam hingga sekitar epoch ke-2 sampai ke-3, kemudian mendatar. Kurva validasi bergerak stabil dekat dengan kurva pelatihan, menunjukkan bahwa performa model pada data yang belum pernah dilihat tetap konsisten.



Gambar 4.9 Kurva *History* MAE Model FFNN Selama Pelatihan

Gambar 4.10 menunjukkan bahwa MAE pelatihan terus menurun, sedangkan MAE validasi mengalami sedikit fluktuasi, tetapi tetap berada pada

kisaran yang rendah dan tidak membentuk pola peningkatan yang mengarah pada *overfitting*. Secara keseluruhan, kedua grafik menunjukkan bahwa model FFNN mampu mengikuti pola data dengan baik, di mana hasil prediksi mendekati nilai aktual meskipun terdapat sedikit variabilitas pada beberapa titik.

4.1.6 Hasil Prediksi Dataset

Berdasarkan hasil prediksi pada Tabel 4.8, model FFNN menghasilkan urutan nilai numerik yang setelah dibulatkan dan dikonversi kembali menjadi residu asam amino membentuk sekuens “IPHRRARPYA”.

Tabel 4.8 Hasil Prediksi Model FFNN terhadap Data Uji

Posisi	Nilai Prediksi	Nilai Dibulatkan	Huruf Asam Amino
1	9.905124	10	I
2	15.187288	15	P
3	8.686368	9	H
4	1.620337	2	R
5	1.785047	2	R
6	1.352342	1	A
7	1.873368	2	R
8	15.358913	15	P
9	19.318659	19	Y
10	1.224479	1	A

Hasil prediksi model FFNN membentuk sekuens IPHRRARPYA, yang diperoleh dari konversi nilai numerik ke residu asam amino sesuai pemetaan yang digunakan. Urutan ini menunjukkan bahwa model tidak menghasilkan prediksi secara acak, karena kemunculan residu bermuatan seperti R (Arginin) dan H (Histidin) masih mendominasi bagian tengah sekuens. Dominasi Arginin selaras dengan karakteristik biologis FCS yang umumnya bermuatan positif.

Komposisi residu pada bagian tengah urutan, yaitu H–R–R–A–R, memperlihatkan pengelompokan residu bermuatan dan residu kecil yang sering muncul pada wilayah cleavage dengan tingkat variasi sedang. Pola ini mengindikasikan bahwa model mampu menangkap hubungan antarposisi residu yang relatif stabil berdasarkan data latih, sehingga menghasilkan sekuens yang konsisten dengan kecenderungan mutasi alami pada sebagian besar sekuens.

Di sisi lain, kemunculan residu hidrofobik seperti I (Isoleusin) pada awal urutan dan Y (Tirosin) pada akhir urutan menunjukkan adanya pergeseran sifat residu dibandingkan pola yang umum ditemukan pada data latih. Pergeseran ini menggambarkan kemungkinan arah mutasi alternatif pada beberapa posisi tertentu, sementara secara keseluruhan prediksi model tetap mempertahankan relevansi biologis dan dapat digunakan sebagai referensi awal dalam mengamati dinamika mutasi di sekitar wilayah FCS.

4.2 Performa Kinerja Model FFNN pada Data *Furin Cleavage Site*

Evaluasi dilakukan dengan menguji model pada data uji untuk mengukur kemampuan generalisasi prediksi mutasi pada *region* FCS. Hasil pengujian ditunjukkan pada Tabel 4.9, yang memuat metrik kinerja model meliputi Mean Squared Error (MSE), Mean Absolute Error (MAE), serta tingkat akurasi baik pada level posisional maupun sekuen penuh.

Tabel 4.9 Evaluasi Kinerja Model FFNN pada Data Uji

MSE (test)	MAE		Akurasi (%)	
	Normalized	Skala Asli	Posisional	Sekuen Penuh
0.004398	0.025283	0.505663	84.66899	0.0

Berdasarkan Tabel 4.9, model FFNN menghasilkan nilai MSE sebesar 0.004398 dan MAE sebesar 0.025283 pada skala normalisasi, yang menunjukkan bahwa selisih numerik antara prediksi dan nilai sebenarnya relatif rendah. Pada skala aslinya, MAE mencapai 0.505663, yang mencerminkan adanya deviasi rata-rata sekitar setengah unit integer encoding, masih wajar untuk model sederhana seperti FFNN. Akurasi posisional sebesar 84.66899% menunjukkan bahwa model mampu memprediksi residu yang benar pada sebagian besar posisi. Namun, akurasi sekuens penuh sebesar 0% mengindikasikan bahwa model belum mampu menghasilkan satu rangkaian sekuens lengkap tanpa kesalahan pada salah satu posisinya. Dengan demikian, meskipun performa posisional cukup baik, konsistensi prediksi sekuens utuh masih menjadi tantangan utama.

Evaluasi posisional juga dilakukan menggunakan metrik *precision*, *recall*, dan *F1_score* dengan rata-rata makro untuk memberikan gambaran yang lebih seimbang terhadap seluruh kelas mutasi. Nilai *precision* sebesar 0.3996 menunjukkan proporsi prediksi positif yang benar pada seluruh kelas. Nilai *recall* sebesar 0.3883 menggambarkan kemampuan model dalam mengenali seluruh kelas aktual. Sementara itu, *F1_score* sebesar 0.3938 mencerminkan keseimbangan antara *precision* dan *recall*, sehingga memberikan gambaran umum bahwa performa klasifikasi posisional model berada pada tingkat moderat tanpa dominasi kelas tertentu.

4.2.1 Model Baseline

Penilaian efektivitas model FFNN dilakukan melalui perbandingan kinerja dengan dua model baseline, yaitu metode persistence dan moving average. Perbandingan ini bertujuan memberikan gambaran objektif mengenai

kemampuan FFNN dalam memprediksi mutasi, baik dari sisi kesalahan prediksi maupun tingkat akurasi, sebagaimana disajikan pada Tabel 4.10.

Tabel 4.10 Perbandingan Model Baseline dengan FFNN pada Data Uji

Model	MAE		Akurasi (%)	
	<i>Normalized</i>	<i>Skala Asli</i>	Posisional	Sekuen Penuh
<i>Persistence</i>	0.019181	0.383624	94.912892	72.822300
<i>Moving Average</i>	0.019129	0.382578	92.404181	59.930314
<i>FFNN</i>	0.025283	0.505663	84.668990	0.000000

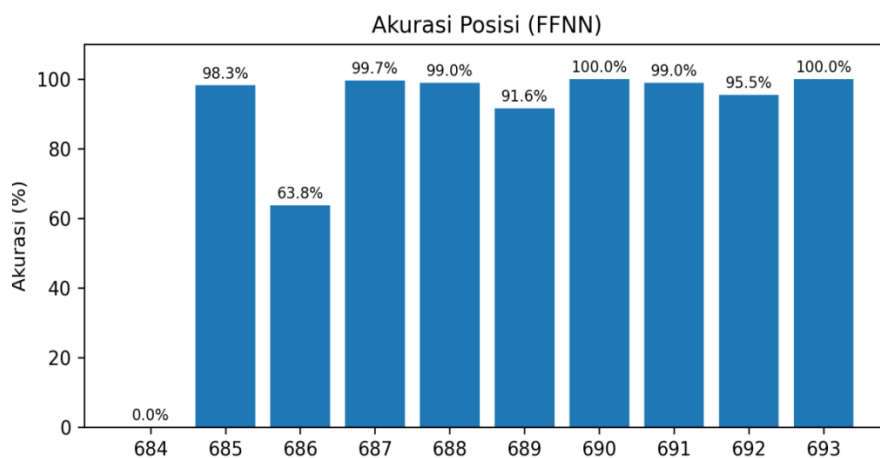
Tabel 4.10 menunjukkan bahwa performa FFNN masih berada di bawah dua model baseline. Metode persistence memberikan hasil terbaik dengan MAE sebesar 0.019181, akurasi posisional 94.912892%, dan akurasi sekuens penuh 72.822300%. Metode ini bekerja dengan menetapkan semua nilai prediksi sama dengan pengamatan terakhir pada deret waktu (Kiriakidou, 2025). Moving average juga menunjukkan kinerja yang cukup baik dengan MAE 0.019129 dan akurasi posisional 92.404181%, meskipun akurasi sekuens penuhnya lebih rendah. Metode moving average mampu menangkap pola lokal melalui perataan dua sekuens terakhir, namun sifat perataannya dapat menghilangkan variasi diskrit yang penting dalam prediksi sekuens penuh (Taufiqillah et al., 2024).

Sementara itu, model FFNN menghasilkan MAE 0.025283, akurasi posisional 84.668990%, dan akurasi sekuens penuh 0%, menunjukkan bahwa model belum mampu melampaui baseline. Oleh karena itu, penelitian ini menegaskan perlunya pengembangan lebih lanjut, baik melalui peningkatan arsitektur, penambahan konteks sekuens yang lebih luas, maupun penggunaan model nonlinier yang lebih dalam agar prediksi mutasi pada region FCS dapat dilakukan dengan lebih akurat.

Meskipun demikian, hasil ini masih belum melampaui performa model dasar. Model persistensi memberikan akurasi posisional tertinggi, sedangkan model moving average juga menunjukkan hasil yang lebih baik dalam mengenali pola global dibandingkan FFNN. Kondisi ini menunjukkan bahwa FFNN hanya mampu mempelajari pola mutasi secara lokal dan belum dapat menangkap dependensi jangka panjang antarposisi dalam sekuens protein. Keterbatasan tersebut muncul karena arsitektur *feedforward* tidak mempertimbangkan informasi kontekstual antarresidu. Oleh karena itu, penggunaan arsitektur yang lebih kontekstual seperti RNN atau model berbasis Transformer menjadi langkah penting agar hubungan spasial dan pola evolusioner dapat dipelajari dengan lebih menyeluruh.

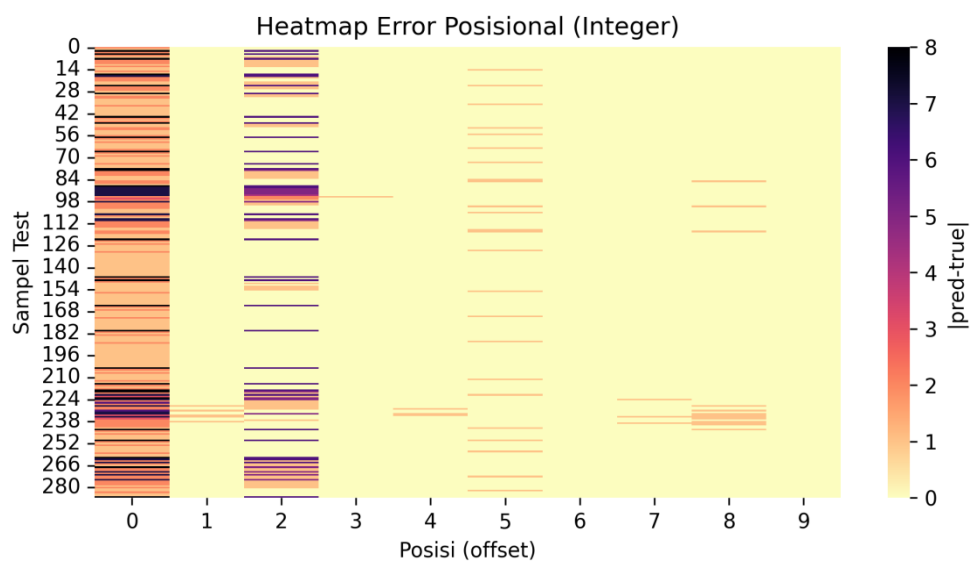
4.1.7 Analisis Kesalahan dan Identifikasi Hotspot Mutasi

Pemahaman terhadap performa model dalam memprediksi mutasi pada setiap titik asam amino diperoleh melalui evaluasi akurasi per posisi. Visualisasi disajikan pada Gambar 4.11 menunjukkan tingkat keberhasilan model FFNN dalam memprediksi setiap posisi individual.



Gambar 4.10 Akurasi Prediksi FFNN pada Setiap Posisi FCS

Berdasarkan Gambar 4.11, akurasi model bervariasi antarposisi. Posisi 685, 687, 688, 690, 691, dan 693 mencapai akurasi yang sangat tinggi (98–100%), menunjukkan bahwa pola mutasi pada posisi-posisi ini relatif stabil dan mudah dipelajari oleh model. Posisi 692 juga memiliki akurasi tinggi, yaitu 95.47%. Sebaliknya, dua posisi menunjukkan akurasi yang jauh lebih rendah, yaitu 684 (0.00%) dan 686 (63.76%), mengindikasikan bahwa model mengalami kesulitan dalam memprediksi residu pada titik tersebut. Posisi 689 memiliki akurasi menengah sebesar 91.64%. Secara keseluruhan, mayoritas posisi dapat diprediksi dengan baik, meskipun terdapat beberapa titik yang cukup variatif sehingga menurunkan kinerja model secara lokal.



Gambar 4.11 *Heatmap Tingkat Error Posisional* pada Prediksi Asam Amino

Hasil tersebut diperkuat oleh visualisasi *heatmap error* pada Gambar 4.12, yang menunjukkan akumulasi kesalahan prediksi terbesar pada posisi offset 0 dan offset 2. Warna yang lebih pekat pada kedua posisi tersebut mencerminkan perbedaan nilai prediksi terhadap nilai aktual yang lebih besar dibandingkan dengan posisi lainnya.

Hasil prediksi yang ditampilkan pada Tabel 4.11 menunjukkan perbandingan antara urutan asam amino aktual (*True*) dan hasil prediksi model (*Pred*) pada sepuluh sampel data uji. Nilai match per posisi berada pada kisaran 80–90%, yang mengindikasikan bahwa sebagian besar residu berhasil diprediksi dengan benar. Kesalahan prediksi umumnya terjadi pada residu awal urutan, sebagaimana terlihat pada perbedaan karakter pertama atau kedua antara urutan *True* dan *Pred*, sementara residu di bagian tengah hingga akhir urutan cenderung konsisten.

Tabel 4.11 Hasil Prediksi Model pada Sampel Data Mutasi SARS-CoV-2

	Sampel	True	Pred	Match per posisi (%)
0	1	KSHRRARSVA	ISHRRARSVA	90
1	2	KSHRRARSVA	LSHRRARSVA	90
2	3	NSPRRARSVA	LSHRRARSVA	80
3	4	KSHRRARSVA	ISHRRARSVA	90
4	5	NSPRRARSVA	LSHRRARSVA	80
5	6	KSHRRARSVA	ISHRRARSVA	90
6	7	KSHRRARSVA	LSIRRARSVA	80
7	8	NSPRRARSVA	LSHRRARSVA	80
8	9	KSHRRARSVA	ISIRRARSVA	80
9	10	KSHRRARSVA	ISIRRARSVA	80

Temuan ini konsisten dengan analisis per-posisi pada Tabel 4.12, yang menunjukkan bahwa posisi dengan entropi rendah, seperti posisi 685, 687, dan 690, memiliki akurasi posisional tinggi (di atas 98%) dan nilai MAE yang kecil. Sebaliknya, posisi dengan entropi tinggi menunjukkan penurunan akurasi dan peningkatan kesalahan prediksi. Hal ini mengindikasikan bahwa kinerja prediksi model sangat dipengaruhi oleh tingkat stabilitas mutasi, di mana posisi yang

lebih konservatif lebih mudah dipelajari oleh model dibandingkan posisi dengan variasi mutasi yang tinggi.

Tabel 4.12 Akurasi dan MAE Per-Posisi pada Region FCS

	Posisi	Akurasi Posisisonal (%)	MAE (Skala Integer)	Entropi
0	684	0.00	2.463	0.724
1	685	98.26	0.235	0.006
2	686	63.76	1.234	0.842
3	687	99.65	0.143	0.006
4	688	98.95	0.118	0.013
5	689	91.64	0.313	0.006
6	690	100.00	0.125	0.006
7	691	98.95	0.126	0.006
8	692	95.47	0.218	0.006
9	693	100.00	0.082	0.013

Selanjutnya, analisis *error rate* per posisi dilakukan guna mengidentifikasi distribusi kesalahan prediksi dan menentukan titik-titik yang paling sulit dipelajari oleh model, sebagaimana ditunjukkan pada tabel 4.13.

Tabel 4.13 Peringkat Posisi Berdasarkan Error Rate (Tertinggi di Atas)

	Posisi	Error rate (%)	err (int)			Entropi	AA Unik (True)
			Mean	Median	Max		
0	684	100.00	2.519	2.0	8	0.724	2
1	686	36.24	1.146	0.0	6	0.842	2
2	689	8.36	0.084	0.0	1	0.006	1
3	692	4.53	0.045	0.0	1	0.006	1
4	685	1.74	0.017	0.0	1	0.006	1
5	688	1.05	0.010	0.0	1	0.013	1
6	691	1.05	0.010	0.0	1	0.006	1
7	687	0.35	0.003	0.0	1	0.006	1

8	690	0.00	0.000	0.0	0	0.006	1
9	693	0.00	0.000	0.0	0	0.013	1

Tabel 4.13 memperjelas distribusi kesalahan dengan mengurutkan posisi berdasarkan *error rate*. Posisi 684 menempati peringkat kesalahan tertinggi dengan *error rate* 100%, diikuti posisi 686 (36.24%). Kedua posisi ini juga memiliki nilai $|err|$ rata-rata tertinggi dan jumlah residu unik lebih banyak, sehingga menunjukkan bahwa posisi ini merupakan hotspot mutasi yang menyebabkan prediksi model tidak stabil. Posisi lainnya, seperti 687, 688, 689, 691, 692, dan 693, memiliki *error rate* yang sangat rendah (0–8.36%) dan entropi dekat nol, sehingga dianggap stabil dan mudah dipelajari oleh model.

Tabel 4.14 Misprediksi Teratas pada Posisi Panas

Posisi	<i>Error rate (%)</i>	<i>Mean err (int)</i>	<i>Top-5 mispred (true→pred)</i>
684	100.0	2.519	K→L (141); K→I (90); N→L (28); N→I (22); K→H (4)

Tabel 4.13 memperlihatkan pola misprediksi pada posisi dengan *error rate* tertinggi, yaitu 684. Pada posisi ini, model menunjukkan *error rate* 100% dengan deviasi rata-rata 2.519 pada skala integer. Substitusi yang paling dominan adalah K→L, K→I, N→L, N→I, dan K→H. Keragaman substitusi ini mengonfirmasi bahwa posisi 684 merupakan mutation hotspot yang memiliki dinamika residu kompleks. Tingginya *error* menunjukkan bahwa model belum mampu menangkap variasi biologis yang terjadi pada titik tersebut. Secara keseluruhan, evaluasi posisional menunjukkan bahwa performa FFNN sangat dipengaruhi oleh stabilitas mutasi pada tiap posisi. Posisi yang konservatif dapat

diprediksi dengan akurasi mendekati sempurna, sedangkan posisi dengan keragaman mutasi tinggi terutama pada posisi 684 dan 686 menjadi sumber utama kesalahan model.

4.3 Perbandingan dengan Penelitian Sebelumnya

Penelitian Mustafa et al. (2022) menunjukkan bahwa daerah *furin cleavage site* (FCS) pada posisi 678–686 cenderung mengalami mutasi yang meningkatkan muatan positif, terutama melalui residu seperti K dan R. Mutasi dominan seperti N679K dan P681R meningkatkan efisiensi pemotongan dan memperkuat replikasi virus. Bila dibandingkan dengan hasil penelitian ini, bagian tengah dari sekuens prediksi yaitu H–R–R–A–R menunjukkan kesesuaian karena tiga posisi menghasilkan Arginin (R) yang selaras dengan pola peningkatan muatan positif yang dilaporkan oleh Mustafa et al. Kecocokan pada bagian ini menandakan bahwa penelitian ini juga menemukan kecenderungan penguatan muatan pada inti FCS, seperti yang dicatat dalam penelitian tersebut.

Berbeda dengan itu, Magazine et al. (2022) melaporkan bahwa mutasi pada FCS, khususnya pada rentang residu 680–689, menjadi salah satu perubahan paling signifikan selama evolusi SARS-CoV-2. Mereka menekankan bahwa mutasi P681R meningkatkan efisiensi pemotongan S1/S2 oleh furin, sehingga memperkuat infektivitas virus, sementara P681H tidak memberikan peningkatan signifikan secara mandiri. Dalam penelitian ini, pola prediksi menghasilkan dominasi residu bermuatan positif, terutama Arginin (R), pada inti FCS. Dominasi residu R tersebut selaras dengan mekanisme yang dijelaskan Magazine et al., karena keberadaan residu bermuatan positif di sekitar P681 diketahui berpotensi meningkatkan kecenderungan pemotongan furin. Dengan demikian, hasil prediksi

penelitian ini menunjukkan arah mutasi yang konsisten dengan pola penguatan cleavage yang dijelaskan dalam penelitian mereka.

Sementara itu, penelitian Tang et al. (2023) menunjukkan bahwa varian Omicron membawa tiga mutasi utama di sekitar *furin cleavage site*, yakni P681H, H655Y, dan N679K, yang menyebabkan penurunan efisiensi pemotongan S1/S2 serta berkurangnya pemanfaatan TMPRSS2 dalam proses aktivasi spike. Kondisi ini menghasilkan aktivitas fusi membran yang lebih lemah dan berkontribusi pada patogenesis Omicron yang relatif lebih rendah dibandingkan varian sebelumnya, termasuk Delta. Hasil penelitian ini memperlihatkan pola residu pada awal dan akhir urutan prediksi yang cenderung menghasilkan konfigurasi yang kurang mendukung pemotongan optimal, seperti munculnya residu I, P, dan Y. Meskipun sifat residu-residu tersebut tidak dibahas secara langsung oleh Tang et al. (2023), kecenderungan pola yang muncul pada model menunjukkan konsistensi biologis dengan mekanisme pelemahan cleavage dan penurunan fusi membran yang diidentifikasi pada varian Omicron.

Secara keseluruhan, perbandingan dengan ketiga penelitian menunjukkan bahwa hasil penelitian ini menghasilkan pola yang bercampur. Bagian tengah urutan prediksi selaras dengan temuan Mustafa et al. dan sebagian Magazine et al., karena menggambarkan kecenderungan peningkatan muatan positif di inti FCS. Sebaliknya, bagian awal dan akhir urutan tidak mengikuti pola tersebut dan justru lebih mendekati karakteristik mutasi pada Omicron yang dijelaskan oleh Magazine et al. Dengan demikian, penelitian ini menggambarkan arah mutasi yang tidak sepenuhnya mengarah pada peningkatan efisiensi pemotongan,

melainkan menunjukkan kecenderungan perubahan sifat biologis yang lebih stabil dan kurang bermuatan.

4.4 Integrasi Keislaman dalam Analisis Prediksi Mutasi

Hasil penelitian mengenai prediksi mutasi pada situs cleavage SARS-CoV-2 memberikan gambaran bahwa perubahan pada suatu sistem biologis tidak terjadi secara acak sepenuhnya, tetapi mengikuti pola tertentu yang dapat dipelajari. Misalnya, pada posisi 687 ditemukan pola pergantian residu seperti $K \rightarrow I$, $N \rightarrow L$, $N \rightarrow I$, dan $K \rightarrow H$, yang menunjukkan arah perubahan sifat asam amino dari bermuatan menjadi lebih netral atau hidrofobik. Pola seperti ini menggambarkan bahwa dinamika mutasi berjalan melalui mekanisme yang teratur, bukan sekadar peristiwa acak tanpa struktur. Perspektif ini sangat sesuai dengan konsep Islam yang menegaskan bahwa segala sesuatu di alam ini berlangsung dengan ketentuan dan aturan yang tetap. Hal ini tercermin dalam firman Allah:

وَيَرْزُقُهُ مِنْ حَيْثُ لَا يَحْتَسِبُ ۚ وَمَنْ يَتَوَكَّلْ عَلَى اللَّهِ فَهُوَ حَسْبُهُ ۚ إِنَّ اللَّهَ بَالِغُ أَمْرِهِ ۚ قَدْ جَعَلَ اللَّهُ لِكُلِّ شَيْءٍ قَدْرًا ۝٣

“dan menganugerahkan kepadanya rezeki dari arah yang tidak dia duga. Siapa yang bertawakal kepada Allah, niscaya Allah akan mencukupkan (keperluan)-nya. Sesungguhnya Allahlah yang menuntaskan urusan-Nya. Sungguh, Allah telah membuat ketentuan bagi setiap sesuatu” (QS.At-Thalaq: 3).

Ayat ini memberikan pemahaman bahwa proses biologis, termasuk mutasi virus, berada dalam bingkai ketetapan ilmiah yang telah Allah SWT tetapkan. Ketika model FFNN mampu memprediksi pola perubahan, maka hal ini menunjukkan bahwa fenomena mutasi memiliki keteraturan yang dapat diselidiki karena Allah SWT memang menciptakan alam dengan hukum-hukum yang dapat dipahami manusia melalui ilmu pengetahuan.

Integrasi antara hasil penelitian dan perspektif keislaman juga terlihat dari bagaimana manusia diberi kemampuan untuk mengembangkan ilmu guna memahami fenomena alam, termasuk penyakit dan penyebarannya. Dalam konteks prediksi mutasi, penggunaan model komputasi seperti FFNN merupakan bentuk ikhtiar ilmiah manusia untuk mencegah, mengantisipasi, atau mempelajari karakter virus yang terus berubah. Hal ini selaras dengan perintah Nabi Muhammad:

حَدَّثَنَا مُحَمَّدُ بْنُ الْمُثَنَّى، حَدَّثَنَا أَبُو أَحْمَدَ الزُّبَيْرِيُّ، حَدَّثَنَا عُمَرُ بْنُ سَعِيدٍ بْنُ أَبِي حُسَيْنٍ، قَالَ حَدَّثَنِي عَطَاءُ بْنُ أَبِي رِبَاحٍ، عَنْ أَبِي هُرَيْرَةَ رَضِيَ اللَّهُ عَنْهُ عَنِ النَّبِيِّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ " مَا أَنْزَلَ اللَّهُ دَاءً إِلَّا أَنْزَلَ لَهُ شِفَاءً " (رَوَاهُ الْبُخَارِيُّ ٥٦٧٨).

“Telah menceritakan kepada kami Muhammad bin Al-Mutsanna, telah menceritakan kepada kami Abu Ahmad Al-Zubairi, telah menceritakan kepada kami ‘Umar bin Sa’id bin Abi Husain, ia berkata: ‘Ata’ bin Abi Rabah telah menceritakan kepadaku, dari Abu Hurairah –radhiyallahu ‘anhū– dari Nabi ﷺ, beliau bersabda: ‘Allah tidak menurunkan suatu penyakit kecuali Dia juga menurunkan obatnya.’” (HR. Bukhari no. 5678).

Hadis tentang anjuran berobat menunjukkan bahwa mempelajari mutasi, termasuk menganalisis perubahan residu dan memprediksi sekuens, merupakan bentuk ikhtiar ilmiah manusia dalam menghadapi penyakit. Dengan memahami pola mutasi, peneliti dapat mengantisipasi kemunculan varian baru lebih awal. Adanya hotspot mutasi seperti posisi 687 menggambarkan sunnatullah, yaitu hukum perubahan yang terus berlangsung sehingga manusia perlu mengambil pelajaran darinya. Integrasi keislaman dalam penelitian ini menegaskan bahwa ilmu pengetahuan menjadi sarana untuk memahami ketetapan Allah SWT dalam ciptaan-Nya sekaligus menjalankan amanah menjaga kesehatan masyarakat.

BAB V PENUTUP

5.1 Kesimpulan

Kesimpulan penelitian ini merangkum temuan utama terkait pengembangan dan evaluasi model prediksi mutasi pada region FCS SARS-CoV-2, yang disajikan sebagai berikut.

1. Penelitian ini berhasil membangun model FFNN yang mampu mempelajari pola mutasi *region* FCS protein Spike SARS-CoV-2, khususnya pada tingkat posisional. Model menunjukkan performa yang kuat dengan akurasi 83,64% dan *error* rendah ($MSE = 0,006178$; $MAE = 0,0291$). Posisi residu yang bersifat konservatif dapat diprediksi secara konsisten, menunjukkan bahwa FFNN efektif dalam menangkap pola lokal pada sekuens protein.
2. Model menunjukkan kinerja yang baik pada level residu, tetapi belum akurat dalam memprediksi sekuens penuh, terutama pada posisi yang sangat bervariasi seperti residu 687 dan 689. Oleh karena itu, diperlukan pengembangan arsitektur serta penambahan konteks sekuens yang lebih luas agar prediksi menjadi lebih stabil.

5.2 Saran

Beberapa saran dapat diajukan untuk pengembangan penelitian selanjutnya.

1. Penelitian selanjutnya disarankan menggunakan model berbasis sekuens, seperti klasifikasi multilabel, untuk akurasi prediksi sekuens penuh.
2. Penelitian selanjutnya disarankan menambahkan konteks sekuens yang lebih luas, agar prediksi pada posisi residu yang sangat bervariasi menjadi lebih stabil dan relevan secara biologis.

DAFTAR PUSTAKA

- Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). *Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning*. *Nature Biotechnology*, 33(8), 831–838. <https://doi.org/10.1038/nbt.3300>
- Al-Bukhari. (n.d.). *Sahih al-Bukhari, Hadis No. 162 (Kitab Wudhu')*. Diakses dari <https://sunnah.com/bukhari:162>
- Al-Bukhari. (n.d.). *Sahih al-Bukhari, Hadis No. 3473 (Kitab Al-Anbiya')*. Diakses dari <https://sunnah.com/bukhari:3473>
- Al-Bukhari. (n.d.). *Sahih al-Bukhari, Hadis No. 5678 (Kitab Al-Tibb)*. Diakses dari <https://sunnah.com/bukhari:5678>
- Azam, D. F., Ratnawati, D. E., & Adikara, P. P. (2018). Prediksi Harga Emas Batang Menggunakan Feed Forward Neural Network Dengan Algoritme Genetika. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(8), 2317–2322.
- Barinov, R., Gai, V., Kuznetsov, G., & Golubenko, V. (2023). *Automatic Evaluation of Neural Network Training Results*.
- Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral Genomes resource. *Nucleic Acids Research*, 43(D1), D571–D577. <https://doi.org/10.1093/nar/gku1207>
- Brownlee, J. (2020). *Introduction to Time Series Forecasting with Python*.
- Chan, Y. A., & Zhan, S. H. (2022). The Emergence of the Spike Furin Cleavage Site in SARS-CoV-2. *Molecular Biology and Evolution*, 39(1), 1–7. <https://doi.org/10.1093/molbev/msab327>
- Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Gu, C. (2020). *FindFur : A Tool for Predicting Furin Cleavage Sites of Viral Envelope Substrates*.
- Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., et al. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews*

- Microbiology*, 19(7), 409–424. <https://doi.org/10.1038/s41579-021-00573-0>
- Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Molecular Cell*, 78(4), 779–784.e5. <https://doi.org/10.1016/j.molcel.2020.04.022>
- Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., et al. (2021). The origins of SARS-CoV-2: A critical review. *Cell*, 184(19), 4848–4856. <https://doi.org/10.1016/j.cell.2021.08.017>
- Huang, Y., Yang, C., Xu, X. feng, Xu, W., & Liu, S. wen. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9), 1141–1149. <https://doi.org/10.1038/s41401-020-0485-4>
- Jamhuri, M., Irawan, M. I., Mukhlash, I., Iqbal, M., & Puspaningsih, N. N. T. (2025). Neural networks optimization via Gauss–Newton based QR factorization on SARS-CoV-2 variant classification. *Systems and Soft Computing*, 7(February), 200195. <https://doi.org/10.1016/j.sasc.2025.200195>
- Katoh, K. (2020). *Multiple Sequence Alignment*.
- Kiriakidou, P. (2025). *Hierarchical Time Series Forecasting for Electricity Demand. February*.
- King, J., Kosinski-Collins, M., & Sundberg, E. (2020). Coronavirus Structure, Vaccine and Development. Biophysical Society. Diakses <https://www.biophysics.org/blog/coronavirus-structure-vaccine-and-therapydevelopment>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Magazine, N., Zhang, T., Wu, Y., Mcgee, M. C., Veggiani, G., & Huang, W. (2022). *Mutations and Evolution of the SARS-CoV-2 Spike Protein*. 1–11.
- Maher, N., & Yousif, S. A. (2023). An automated machine learning model for diagnosing COVID-19 infection. *IAES International Journal of Artificial Intelligence*, 12(3), 1360–1369.
- Malik, N., Akhtar, N., Rohilla, H., & Sharma, P. (2013). *Journal Of Harmonized Research (JOHR) Review article*. 1(2), 120–124.

- Mashudi, K. (2020). *Telaah Tafsir al-Muyassar (Jilid 1 : Juz 1 5) (1)*. Malang: Intelegensia Media
- Motomura, K., Fujita, T., Tsutsumi, M., Kikuzato, S., Nakamura, M., & Otaki, J. M. (2012). *Word Decoding of Protein Amino Acid Sequences with Availability Analysis: A Linguistic Approach*. 7(11), 1–15. <https://doi.org/10.1371/journal.pone.0050039>
- Mustafa, Z., Kalbacher, H., & Burster, T. (2022). *Occurrence of a novel cleavage site for cathepsin G adjacent to the polybasic sequence within the proteolytically sensitive activation loop of the SARS-CoV-2 Omicron variant : The amino acid substitution N679K and P681H of the spike protein*. 2, 1–11. <https://doi.org/10.1371/journal.pone.0264723>
- Naqvi, A. A. T., Fatima, K., Mohammad, T., et al. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ’ s public news and information. *BBA - Molecular Basis of Disease*, January, 1–17.
- Nugroho, A. A., Haris, M., Studi, P., Komputer, I., Informasi, F. T., & Mandiri, U. N. (2024). *Meningkatkan Kualitas Data Pada Peramalan Curah Hujan*. 7(2), 301–311.
- Örd, M., Faustova, I., & Loog, M. (2020). The sequence at Spike S1 / S2 site enables cleavage by furin and phospho - regulation in SARS - CoV2 but not in SARS - CoV1 or MERS - CoV. *Scientific Reports*, 1–10. <https://doi.org/10.1038/s41598-020-74101-0>
- Quran NU Online. (n.d.). Surat Al-Baqarah Ayat 195: Arab, Latin, Terjemah dan Tafsir Lengkap. Diakses dari <https://quran.nu.or.id/al-baqarah/195>
- Quran NU Online. (n.d.). Surat Al-Mujadilah Ayat 11: Arab, Latin, Terjemah dan Tafsir Lengkap. Diakses dari <https://quran.nu.or.id/al-mujadilah/11>
- Quran NU Online. (n.d.). Surat At-Thalaq Ayat 3: Arab, Latin, Terjemah dan Tafsir Lengkap. Diakses dari <https://quran.nu.or.id/at-thalaq/3>
- Park, S. B., Khan, M., Chiliveri, S. C., et al (2023). *SARS-CoV-2 omicron variants harbor spike protein mutations responsible for their attenuated fusogenic phenotype*. 2023, 1–14. <https://doi.org/10.1038/s42003-023-04923-x>

- Saldivar-espinoza, B., Macip, G., Garcia-segura, P., et al. (2022). *Prediction of Recurrent Mutations in SARS-CoV-2.pdf*.
- Sathyanarayanan, S., & Tantri, B. R. (2024). *Confusion Matrix-Based Performance Evaluation Metrics*. 27(4).
- Septiana, D., & Br Bangun, M. (2023). Perbandingan Antara Feed Forward Neural Network (Ffnn) Dan Seasonal Autoregressive Integrated Moving Average (Sarima) Untuk Peramalan Data Runtun Waktu Terapan. *Jurnal Deli Sains Informatika*, 2(2).
- Singh, G. B. (2015). *Modeling and Optimization in Science and Technologies Fundamentals of Bioinformatics and Computational Biology Methods and Exercises in MATLAB*.
- Sinha, A., Sangeet, S., & Roy, S. (2023). Evolution of Sequence and Structure of SARS-CoV-2 Spike Protein: A Dynamic Perspective. *ACS Omega*, 8(26), 23283–23304. <https://doi.org/10.1021/acsomega.3c00944>
- Tang, H., Shao, Y., Huang, Y., Qiao, S., An, J., et al. (2023). *Evolutionary characteristics of SARS-CoV-2 Omicron subvariants adapted to the host. March, 2022–2024*. <https://doi.org/10.1038/s41392-023-01449-w>
- Taufiqillah, R., Aulia, R. N., Dewi, P., Heikal, J., & Jakarta, S. (2024). *Analisis Evaluasi Efektivitas Metode Forecasting Moving Avarage Naive Approach , Simple Moving Average , Exponential Smoothing terhadap Supply Chain Management PT XYZ*. 4, 4477–4492.
- Tian, S., Huajun, W., & Wu, J. (2012). Computational prediction of furin cleavage sites by a hybrid method and understanding mechanism underlying diseases. *Scientific Reports*, 2. <https://doi.org/10.1038/srep00261>
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19(3), 155–170. <https://doi.org/10.1038/s41579-020-00468-6>
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., et al. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*,

367(6485), 1444–1448. <https://doi.org/10.1126/science.abb2762>

Yuliandar, D., Warsito, B., & Yasin, H. (2012). Pelatihan Feed Forward Neural Network Menggunakan Algoritma Genetika Dengan Metode Seleksi Turnamen Untuk Data Time Series. *Jurnal Gaussian*, 1, 65–72.

LAMPIRAN

Lampiran 1 Akses Dataset Sekuensial dan Notebook Kaggle Prediksi Mutasi



Dataset Sekuensial



Notebook Kaggle Prediksi Mutasi

Lampiran 2 Script python Model FFNN Prediksi Mutasi

```
pip install biopython

import os
from datetime import datetime
from typing import Optional, Tuple

import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

# Retina figure (notebook-friendly)
try:
    from IPython.display import display
    from IPython import get_ipython
    get_ipython().run_line_magic('config',
    "InlineBackend.figure_format = 'retina'")
except Exception:
    pass
matplotlib.rcParams['figure.dpi'] = 200

from Bio import SeqIO

# TensorFlow / Keras
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
```

1. Konfigurasi umum

```
RNG_SEED = 123
TF_SEED = 123
np.random.seed(RNG_SEED)
tf.random.set_seed(TF_SEED)

# Opsional: agar GPU tidak all-allocate memori
```

```

try:
    gpus = tf.config.list_physical_devices('GPU')
    if gpus:
        for gpu in gpus:
            tf.config.experimental.set_memory_growth(gpu, True)
except Exception as e:
    print(f"[WARN] Could not set memory growth: {e}")

# Path & region
FASTA_PATH = "/kaggle/input/skripsi/Data_Sequens.fasta"
START_POS = 684
END_POS = 694
# Windowing & split
LAG = 8
SPLIT_TRAIN = 0.70
SPLIT_VAL = 0.15 # sisa otomatis jadi test

# Mapping asam amino → integer
AA_TO_INT = {
    'A':1, 'R':2, 'N':3, 'D':4, 'C':5, 'E':6, 'Q':7, 'G':8,
    'H':9, 'I':10,
    'L':11, 'K':12, 'M':13, 'F':14, 'P':15, 'S':16, 'T':17,
    'W':18, 'Y':19, 'V':20,
    '-':0, 'X':0
}
INT_TO_AA = {v:k for k,v in AA_TO_INT.items()}
AA_ALPHABET = list(AA_TO_INT.keys()) # 21 simbol
AA_MAX = 20.0 # untuk skala 0..1

```

2. Helper functions

```

def parse_date_safe(date_str: Optional[str]) -> Optional[datetime]:
    if date_str is None:
        return None
    for fmt in ("%Y-%m-%d", "%Y-%m", "%Y"):
        try:
            return datetime.strptime(date_str.strip(), fmt)
        except ValueError:
            continue
    return None

def seq_to_int(seq: str) -> list:
    return [AA_TO_INT.get(aa, 0) for aa in seq]

def ints_to_aa_string(int_row: np.ndarray) -> str:
    return "".join(INT_TO_AA.get(int(x), "?") for x in int_row)

def make_windows_matrix(H: np.ndarray, lag: int = 2) -> Tuple[np.ndarray, np.ndarray]:
    X, y = [], []
    for i in range(len(H) - lag):
        win = H[i:i+lag, :]
        tgt = H[i+lag, :]

```

```

        X.append(win.flatten())
        y.append(tgt)
    return np.array(X), np.array(y)

def entropy_from_counts(counts: np.ndarray) -> float:
    """counts: array non-negatif; entropi basis 2."""
    total = counts.sum()
    if total <= 0:
        return 0.0
    p = counts / total
    p = p[p > 0]
    return float(-(p * np.log2(p)).sum())

```

3. Baca FASTA & ekstrak region FCS

```

FASTA_WUHAN_MSA = "/kaggle/input/skripsi/Data_Wuhan.fasta"

print(">> Menampilkan data sekuens sebelum MSA\n")

records = []

# Area FCS
fcs_start = 684 - 10
fcs_end = 687 + 10

def extract_fcs(seq):
    start = max(0, fcs_start)
    end = min(len(seq), fcs_end)
    return "... " + seq[start:end] + " ..."

for rec in SeqIO.parse(FASTA_WUHAN_MSA, "fasta"):
    seq_str = str(rec.seq)
    records.append({
        "ID": rec.id,
        "Sequence": extract_fcs(seq_str),
        "Length": len(seq_str)
    })

df_wuhan_msa = pd.DataFrame(records)

# Tampilkan tabel
display(df_wuhan_msa)

# =====BEFORE MSA=====
FASTA_BEFORE_MSA = "/kaggle/input/skripsi/Sebelum_MSA.fasta"

print(">> Menampilkan data sekuens sebelum MSA\n")

records = []

# Area FCS
fcs_start = 684 - 10
fcs_end = 687 + 10

def extract_fcs(seq):

```

```

        start = max(0, fcs_start)
        end = min(len(seq), fcs_end)
        return "..." + seq[start:end] + "..."

for rec in SeqIO.parse(FASTA_BEFORE_MSA, "fasta"):
    seq_str = str(rec.seq)
    records.append({
        "ID": rec.id,
        "Sequence": extract_fcs(seq_str),
        "Length": len(seq_str)
    })

df_before_msa = pd.DataFrame(records)

# Tampilkan tabel
display(df_before_msa)

# =====PLOT JUMLAH SEKUENS PER PANJANG=====
df_len = df_before_msa[(df_before_msa["Length"] >= 1260) &
                        (df_before_msa["Length"] <= 1280)]

count_len = df_len["Length"].value_counts().sort_index()

plt.plot(count_len.index, count_len.values, marker="o")
plt.xticks(rotation=60)
plt.xlabel("Length (bp)")
plt.ylabel("Count")
plt.title("Jumlah Sekuens per Panjang")
plt.savefig("Jumlah sekuen per panjang", dpi=300,
bbox_inches="tight")
plt.show()

# =====AFTER MSA=====
FASTA_AFTER_MSA = "/kaggle/input/skripsi/Data_Sequens.fasta"

print(">> Menampilkan data sekuens area FCS setelah MSA\n")

records = []

# Area FCS
fcs_start = 684 - 10
fcs_end = 687 + 10

def extract_fcs(seq):
    start = max(0, fcs_start)
    end = min(len(seq), fcs_end)
    return "..." + seq[start:end] + "..."

for rec in SeqIO.parse(FASTA_AFTER_MSA, "fasta"):
    seq_str = str(rec.seq)
    records.append({
        "ID": rec.id,
        "Sequence": extract_fcs(seq_str),
        "Length": len(seq_str)
    })

```

```

    })

df_after_msa = pd.DataFrame(records)

# Tampilkan tabel
display(df_after_msa)

# =====CEK DUPLIKAT=====
# Hitung duplikat berdasarkan ID
duplikat_id = df_after_msa.duplicated(subset="ID")

# Hitung jumlah duplikat dan unik
counts = duplikat_id.value_counts() # True = duplikat, False =
unik

# Buat plot
plt.figure(figsize=(4,3))
plt.bar(["Unik", "Duplikat"], [counts.get(False,0),
counts.get(True,0)], color=["skyblue", "salmon"])
plt.ylabel("Jumlah baris sequens")
plt.tight_layout()
plt.savefig("duplikat", dpi=300, bbox_inches="tight")
plt.show()

# =====CEK AMBIGU KARAKTER=====
df["Jumlah_Ambigu_Gap"] = df["Region_684_694"].apply(
    lambda s: s.count("-") + s.count(".")
)

df["Ada_Ambigu_Gap"] = df["Jumlah_Ambigu_Gap"] > 0

print(df["Ada_Ambigu_Gap"].value_counts())
print(df["Jumlah_Ambigu_Gap"].value_counts().sort_index())

plt.figure(figsize=(5.2, 3.2))
sns.histplot(
    data=df,
    x="Jumlah_Ambigu_Gap",
    bins=range(df["Jumlah_Ambigu_Gap"].max() + 2),
    discrete=True,
    shrink=0.8
)

plt.title("Distribusi karakter ambigu per sekuens")
plt.xlabel("Jumlah karakter ambigu")
plt.ylabel("Jumlah sekuens")
plt.tight_layout()
plt.savefig("Gambar_Distribusi_Karakter_Ambigu.png", dpi=300,
bbox_inches="tight")
plt.show()

# =====CEK SEKUENS TIDAK TERIDENTIFIKASI=====
df["Jumlah X"] = df["Region 684 694"].apply(lambda s:

```

```

s.count("X"))

plt.figure(figsize=(5.2, 3.2))
sns.histplot(data=df, x="Jumlah_X", bins=6, discrete=True, shrink=0.8)

plt.xlim(-0.5, 5.5)
plt.xticks(range(0, 6))

plt.title("Distribusi jumlah 'X' per sekuens")
plt.tight_layout()
plt.savefig("Gambar_1_Distribusi_Jumlah_X.png", dpi=300,
bbox_inches="tight")
plt.show()

# =====PLOT JUMLAH DATA PER TANGGAL=====
if len(df_clean) == 0:
    raise ValueError("Tidak ada sekuens tersisa setelah menghapus 'X'.")

# Tren waktu (opsional): jumlah per bulan
print("\n>> Gambar 2. Tren jumlah sekuens per bulan (data dengan tanggal)")
df_with_date = df_clean[df_clean["Date"].notna()].copy()

if len(df_with_date) > 0:
    df_with_date["YearMonth"] =
df_with_date["Date"].dt.to_period("M").astype(str)
    per_month =
df_with_date.groupby("YearMonth").size().reset_index(name="Count"
)

    plt.figure(figsize=(6.8, 4.0))
    sns.lineplot(data=per_month, x="YearMonth", y="Count",
marker="o")

    # ✓Set batas Y hingga 300
    plt.ylim(0, 300)
    plt.yticks(range(0, 301, 20))

    plt.xticks(rotation=60, ha="right")
    plt.tight_layout()
    plt.savefig("Tren Jumlah", dpi=300, bbox_inches="tight")
    plt.show()
else:
    print("(Tidak ada metadata tanggal; plot dilewati.)")

```

4. Statistik deskriptif FCS (frekuensi AA per posisi, entropi)

```

# =====DISTRIBUSI ASAM AMINO=====
H_str = np.array(df_clean["Region_684_694"].tolist()) # (n_seq,)
of strings

```

```

seq_len = END_POS - START_POS

# Frekuensi: posisi (rows) x AA (columns)
freq_mat = np.zeros((seq_len, len(AA_ALPHABET)), dtype=int)
aa_index = {aa:i for i, aa in enumerate(AA_ALPHABET)}
for row in H_str:
    for pos, aa in enumerate(row):
        freq_mat[pos, aa_index.get(aa, aa_index['-'])] += 1

freq_df = pd.DataFrame(freq_mat, columns=AA_ALPHABET)
freq_df.insert(0, "Posisi", list(range(START_POS, END_POS)))
print("\n>> Tabel 2. Distribusi Asam Amino per Posisi
(frekuensi)")
pd.DataFrame(freq_df).to_csv("/kaggle/working/Tabel_2.csv",
index=False)
display(freq_df)

# =====Entropi per posisi=====
entropies = np.apply_along_axis(entropy_from_counts, axis=1,
arr=freq_mat)
desc_df = pd.DataFrame({
    "Posisi": list(range(START_POS, END_POS)),
    "Entropi": entropies
})
print("\n>> Tabel 3. Entropi per posisi FCS")
display(desc_df)

print("\n>> Gambar 4. Entropi per posisi")
plt.figure(figsize=(6.4,3.0))
plt.bar(desc_df["Posisi"].astype(str), desc_df["Entropi"])
plt.xticks(rotation=0); plt.ylabel("Entropi (bits)");
plt.title("Entropi per Posisi FCS")
plt.tight_layout();
plt.savefig("Entropi per Posisi FCS", dpi=300)
plt.show()

# =====HEATMAP FREKUENSI=====
print("\n>> Gambar 3. Heatmap frekuensi AA per posisi")
plt.figure(figsize=(9, 4.3))
sns.heatmap(freq_mat / freq_mat.sum(axis=1, keepdims=True),
cmap="viridis",
cbar_kws={"label": "Proporsi"},
xticklabels=AA_ALPHABET,
yticklabels=list(range(START_POS, END_POS)))
plt.xlabel("Asam Amino"); plt.ylabel("Posisi");
plt.title("Frekuensi AA per Posisi (proporsi)")
plt.tight_layout();
plt.savefig("Frekuensi_AA_per_Posisi_(proporsi)", dpi=300,
bbox_inches="tight")
plt.show()

```

5. Mapping ke integer & windowing

```
# Urutkan berdasarkan kolom Date
```

```

df_clean = df_clean.sort_values(by="Date").reset_index(drop=True)

# Konversi sekuens menjadi integer
df_clean["Seq_Int"] =
df_clean["Region_684_694"].apply(seq_to_int)

# Buat array sekuens
H = np.array(df_clean["Seq_Int"].tolist()) # shape: (n_seq,
seq_len)

# Cek apakah data cukup untuk windowing
if H.shape[0] <= LAG:
    raise ValueError(f"Data tidak cukup untuk windowing.
n_seq={H.shape[0]} harus > LAG={LAG}.")

# Buat matriks input-output berdasarkan lag
X, y = make_windows_matrix(H, lag=LAG)

# Scaling
X_scaled = X / AA_MAX
y_scaled = y / AA_MAX

# Split berurutan 70/15/15
n = len(X_scaled)
n_train = int(SPLIT_TRAIN * n)
n_val = int(SPLIT_VAL * n)

X_train = X_scaled[:n_train]
y_train = y_scaled[:n_train]
X_val = X_scaled[n_train:n_train+n_val]
y_val = y_scaled[n_train:n_train+n_val]
X_test = X_scaled[n_train+n_val:]
y_test = y_scaled[n_train+n_val:]

# Ringkasan split
print("\n>> Tabel 4. Ringkasan Windowing & Split")
split_df = pd.DataFrame([
    "Lag": LAG,
    "n_seq (setelah X-removed)": H.shape[0],
    "Dimensi input": X.shape[1],
    "Dimensi output": y.shape[1],
    "n_train": len(X_train),
    "n_val": len(X_val),
    "n_test": len(X_test),
])
display(split_df)

print("\n>> Contoh sekuens setelah sorting dan konversi
integer:")
display(df_clean[["Date", "Region_684_694", "Seq_Int"]])

```

6. Model FFNN

```
def build_ffnn(input_dim: int, output_dim: int) -> keras.Model:
    model = keras.Sequential([
        layers.Input(shape=(input_dim,)),
        layers.Dense(256, activation="relu"),
        layers.Dropout(0.10),
        layers.Dense(128, activation="relu"),
        layers.Dense(output_dim, activation="linear"),
    ])
    model.compile(
        optimizer=keras.optimizers.Adam(learning_rate=1e-3),
        loss="mse",
        metrics=[keras.metrics.MeanAbsoluteError(name="mae")]
    )
    return model

model = build_ffnn(X_train.shape[1], y_train.shape[1])

model.summary()
```

7. Baseline Model

```
def baseline_persistence_predict(X_matrix: np.ndarray, lag: int,
                                seq_len: int) -> np.ndarray:
    return X_matrix[:, (lag-1)*seq_len : lag*seq_len]

y_pred_base_scaled = baseline_persistence_predict(X_test, LAG,
                                                    seq_len)
y_pred_base_int = np rint(y_pred_base_scaled *
                           AA_MAX).astype(int).clip(0, 20)
y_true_int_all = np rint(y_test * AA_MAX).astype(int).clip(0,
                                                            20)

acc_pos_base = np.mean(y_pred_base_int == y_true_int_all)
acc_row_base = np.mean(np.all(y_pred_base_int == y_true_int_all,
                               axis=1))
mae_base = np.mean(np.abs(y_pred_base_scaled - y_test))

base_df = pd.DataFrame([
    {"Baseline": "Persistence",
     "MAE (normalized)": mae_base,
     "MAE (skala asli)": mae_base * AA_MAX,
     "Akurasi posisional (%)": acc_pos_base * 100.0,
     "Akurasi sekuens penuh (%)": acc_row_base * 100.0
    }])

def baseline_moving_average_predict(X_matrix: np.ndarray, lag:
int, seq_len: int) -> np.ndarray:
    start = (lag-2) * seq_len
    end = lag * seq_len
    block = X_matrix[:, start:end]
    block = block.reshape(X_matrix.shape[0], 2, seq_len)
    return block.mean(axis=1)
```

```

y_pred_ma_scaled = baseline_moving_average_predict(X_test, LAG,
seq_len)
y_pred_ma_int = np rint(y_pred_ma_scaled *
AA_MAX).astype(int).clip(0, 20)

acc_pos_ma = np.mean(y_pred_ma_int == y_true_int_all)
acc_row_ma = np.mean(np.all(y_pred_ma_int == y_true_int_all,
axis=1))
mae_ma      = np.mean(np.abs(y_pred_ma_scaled - y_test))

print("\n>> Tabel 5. Kinerja Baseline")
base_df2 = pd.DataFrame([{
    "Baseline": "Moving Average",
    "MAE (normalized)": mae_ma,
    "MAE (skala asli)": mae_ma * AA_MAX,
    "Akurasi posisional (%)": acc_pos_ma * 100.0,
    "Akurasi sekuens penuh (%)": acc_row_ma * 100.0
}])
# =====Tabel arsitektur sederhana=====
print("\n>> Tabel 6. Arsitektur Model FFNN")
layers_info = []
total_params = 0
for i, lyr in enumerate(model.layers, 1):
    params = lyr.count_params()
    total_params += params
    layers_info.append({
        "Layer #": i,
        "Tipe": lyr.__class__.__name__,
        "Output shape": str(lyr.output.shape),
        "Params": params
    })
display(pd.DataFrame(layers_info))
print(f"Total params: {total_params:,}")

tabel_baseline = pd.concat([base_df, base_df2],
ignore_index=True)
display(tabel_baseline)

# =====Hiperparameter=====
print("\n>> Tabel 7. Hiperparameter Pelatihan")
hp_df = pd.DataFrame([
    {
        "Optimizer": "Adam",
        "Learning rate": 1e-3,
        "Batch size": 32,
        "Epoch maks": 200,
        "Patience (EarlyStopping)": 10,
        "Dropout": 0.10
    },
    {
        "Optimizer": "SGD (Momentum)",

```

```

        "Learning rate": 1e-2,
        "Batch size": 32,
        "Epoch maks": 200,
        "Patience (EarlyStopping)": 10,
        "Dropout": 0.10
    },
    {
        "Optimizer": "RMSprop",
        "Learning rate": 1e-3,
        "Batch size": 32,
        "Epoch maks": 200,
        "Patience (EarlyStopping)": 10,
        "Dropout": 0.10
    }
])

display(hp_df)

# =====Latih=====
callbacks = [
    keras.callbacks.EarlyStopping(monitor="val_loss",
patience=10, restore_best_weights=True)
]
history = model.fit(
    X_train, y_train,
    validation_data=(X_val, y_val),
    epochs=30,
    batch_size=32,
    verbose=1,
    callbacks=callbacks
)
# =====Kurva pelatihan=====
print("\n>> Gambar 5. Kurva pelatihan (MSE)")
plt.figure(figsize=(6.0,3.5))
plt.plot(history.history["loss"], label="Train MSE")
plt.plot(history.history["val_loss"], label="Val MSE")
plt.xlabel("Epoch"); plt.ylabel("MSE"); plt.title("History Loss
(MSE)")
plt.ylim(0.001, 0.028) # batas sumbu-y
plt.legend(); plt.tight_layout()
plt.savefig("History_Loss_MSE.png", dpi=300)
plt.show()

```

8. Prediksi

```

# =====PREDIKSI DATA SAMPEL=====
show_k = min(10, len(X_test))
rows = []
for i in range(show_k):
    true_str = ints_to_aa_string(y_true_int_all[i])
    pred_str = ints_to_aa_string(y_pred_int[i])

```

```

        match_ratio = np.mean(y_true_int_all[i] == y_pred_int[i]) *
100.0
        rows.append({"Sampel": i+1, "TRUE": true_str, "PRED":
pred_str, "Match per posisi (%)": round(match_ratio, 2)})

print("\n>> Tabel 11. Contoh Prediksi (AA)")
display(pd.DataFrame(rows))

print("\nSelesai 🍀")

# =====PREDIKSI SEKUENS BERIKUTNYA=====
def make_window_last(H, lag, n_features):
    X_last = H[-lag:, :].flatten().astype(np.float32)

    # Penyesuaian panjang fitur
    if X_last.shape[0] != n_features:
        X_last = X_last[:n_features]
        if X_last.shape[0] < n_features:
            X_last = np.pad(X_last, (0, n_features -
X_last.shape[0]), 'constant')
    return X_last.reshape(1, -1)

lag = 8
n_features = X_train.shape[1]
# Buat input terakhir untuk prediksi
X_last = make_window_last(H, lag, n_features)

# Lakukan prediksi
y_next_pred = model(X_last).numpy()

# Data hasil prediksi
pred_float = y_next_pred.flatten()
pred_int = np.round(pred_float).astype(int)
pred_aa = [INT_TO_AA.get(i, '?') for i in pred_int]

# Buat tabel hasil dalam bentuk DataFrame Pandas
tabel_prediksi = pd.DataFrame({
    'Posisi': range(1, len(pred_float) + 1),
    'Nilai Prediksi (Float)': pred_float,
    'Nilai Dibulatkan (Int)': pred_int,
    'Huruf Asam Amino': pred_aa
})

# Tampilkan tabel
tabel_prediksi.style.set_properties(**{
    'text-align': 'center'
}).set_table_styles([dict(selector='th', props=[('text-align',
'center')])])

```

9. Evaluasi Model FFNN

```

test_loss, test_mae = model.evaluate(X_test, y_test, verbose=0)

y_pred_scaled = model.predict(X_test)

```

```

y_pred_int = np rint(y_pred_scaled *
AA_MAX).astype(int).clip(0, 20)

acc_positional = np.mean(y_pred_int == y_true_int_all) * 100.0
acc_sequence = np.mean(np.all(y_pred_int == y_true_int_all,
axis=1)) * 100.0

print("\n>> Tabel 8. Kinerja Model (Test)")
model_perf_df = pd.DataFrame([
    "MSE (test)": test_loss,
    "MAE (normalized)": test_mae,
    "MAE (skala asli)": test_mae * AA_MAX,
    "Akurasi posisional (%)": acc_positional,
    "Akurasi sekuens penuh (%)": acc_sequence
])
display(model_perf_df)

# =====EVALUASI PRECISION, RECALL, F1-SCORE=====
from sklearn.metrics import precision_score, recall_score,
f1_score

# Flatten data untuk menghitung precision / recall / F1
y_true_flat = y_true_int_all.flatten()
y_pred_flat = y_pred_int.flatten()

precision = precision_score(y_true_flat, y_pred_flat,
average="macro", zero_division=0)
recall = recall_score(y_true_flat, y_pred_flat,
average="macro", zero_division=0)
f1 = f1_score(y_true_flat, y_pred_flat, average="macro",
zero_division=0)

print("\n>> Tambahan Metrik Klasifikasi (Posisional)")
print(f"Precision (macro): {precision:.4f}")
print(f"Recall (macro): {recall:.4f}")
print(f"F1-score (macro): {f1:.4f}")

# =====Perbandingan dengan baseline=====
print("\n>> Tabel 9. Perbandingan Baseline vs FFNN (Test)")
compare_df = pd.DataFrame([
    {"Model": "Persistence",
    "MAE (normalized)": mae_base,
    "MAE (skala asli)": mae_base * AA_MAX,
    "Akurasi posisional (%)": acc_pos_base * 100.0,
    "Akurasi sekuens penuh (%)": acc_row_base * 100.0},
    {"Model": "Moving Average",
    "MAE (normalized)": mae_ma,
    "MAE (skala asli)": mae_ma * AA_MAX,
    "Akurasi posisional (%)": acc_pos_ma * 100.0,
    "Akurasi sekuens penuh (%)": acc_row_ma * 100.0},
    {"Model": "FFNN",
    "MAE (normalized)": test_mae,
    "MAE (skala asli)": test_mae * AA_MAX,
    "Akurasi posisional (%)": acc_positional ,

```

```

        "Akurasi sekuens penuh (%)": acc_sequence },
    ])
display(compare_df)

```

10. Analisis posisional (akurasi & error)

```

acc_per_pos = np.mean(y_pred_int == y_true_int_all, axis=0)
# akurasi per posisi
mae_per_pos = np.mean(np.abs(y_pred_scaled - y_test), axis=0) *
AA_MAX      # MAE integer per posisi
pos_axis    = np.arange(START_POS, END_POS)

print("\n>> Tabel 10. Akurasi & MAE per Posisi")
pos_df = pd.DataFrame({
    "Posisi": pos_axis,
    "Akurasi posisional (%)": np.round(acc_per_pos * 100.0, 2),
    "MAE (skala asli)": np.round(mae_per_pos, 3),
    "Entropi": np.round(entropies, 3)
})
display(pos_df)

print("\n>> Gambar 7. Barplot Akurasi per Posisi")

plt.figure(figsize=(7.0, 3.5))
bars = plt.bar(pos_df["Posisi"].astype(str), acc_per_pos * 100.0)

plt.ylim(0, 110)
plt.ylabel("Akurasi (%)")
plt.title("Akurasi Posisi (FFNN)")

# Tambahkan keterangan nilai di atas tiap batang
for bar in bars:
    height = bar.get_height()
    plt.text(
        bar.get_x() + bar.get_width()/2,  # posisi x di tengah
        height + 1,                        # posisi y sedikit di
        f"{height:.1f}%",                  # teks, misal "92.5%"
        ha='center', va='bottom', fontsize=8
    )

plt.tight_layout()
plt.savefig("Akurasi_Posisi_FFNN.png", dpi=300)
plt.show()

# =====HEATMAP ERROR=====
print("\n>> Gambar 8. Heatmap Error |pred-true| (skala integer)")
err_int = np.abs(y_pred_int - y_true_int_all) # (n_test,
seq_len)
plt.figure(figsize=(7.2, 4.0))
sns.heatmap(err_int, cmap="magma_r", cbar_kws={"label": "|pred-
true|"})

```

```
plt.xlabel("Posisi (offset)"); plt.ylabel("Sampel Test")
plt.title("Heatmap Error Posisional (Integer)")
plt.tight_layout();
plt.savefig("Heatmap Error Posisional (Integer)", dpi=300)
plt.show()
```

RIWAYAT HIDUP



Penulis bernama Natasyah Asviana, yang akrab dipanggil Natasyah, lahir di Mojokerto pada tanggal 11 Mei 2003. Sebagai anak ketiga dari empat bersaudara, penulis lahir dari pasangan Bapak Asrukin dan Ibu Roi khanah. Penulis bertempat tinggal di Desa Jatilangkung, Kecamatan Pungging, Kabupaten Mojokerto. Pendidikan dasar ditempuh di SD IT Al-Anwar Modopuro-Mojosari, kemudian melanjutkan ke SMP IT Mambaul-Ulum, dan menamatkan pendidikan menengah atas di SMAN 1 Kutorejo pada tahun 2021. Pada tahun yang sama, penulis melanjutkan studi ke jenjang pendidikan tinggi pada Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Selama masa studi, penulis menjadi santri di Lembaga Tinggi Pesantren Luhur Malang, dan aktif sebagai pengurus Dewan Majelis di Departemen Penelitian dan Pengembangan LTPLM. Di luar kegiatan akademik, penulis juga bekerja sebagai *freelance* tutor Bimbel Science Society semenjak tahun 2024. Apabila terdapat pertanyaan, saran, maupun kritik terkait penelitian ini, penulis dapat dihubungi melalui *e-mail*: natasyahasviana@gmail.com



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Natasyah Asviana
NIM : 210601110077
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Pengembangan Model Feedforward Neural Network Untuk
Prediksi Mutasi Data Sekuensial Situs Cleavage Sars-Cov-2
Pembimbing I : Dr. Mohammad Jamhuri, M.Si.
Pembimbing II : Erna Herawati, M.Pd.

No	Tanggal	Hal	Tanda Tangan
1.	13 Maret 2025	Konsultasi Topik dan Data	1.
2.	20 Maret 2025	Konsultasi Bab I, II, dan III	2.
3.	21 April 2025	Konsultasi Bab I, II, dan III	3.
4.	30 April 2025	Konsultasi Bab I, II, dan III	4.
5.	10 Juni 2025	Konsultasi Bab I, II, dan III	5.
6.	20 Juni 2025	ACC Bab I, II, dan III	6.
7.	25 Juni 2025	Konsultasi Kajian Agama Bab I dan II	7.
8.	3 Juli 2025	Konsultasi Kajian Agama Bab I dan II	8.
9.	24 September 2025	ACC Kajian Agama Bab I dan II	9.
10.	24 September 2025	ACC Seminar Proposal	10.
11.	22 Oktober 2025	Konsultasi Revisi Seminar Proposal	11.
12.	29 Oktober 2025	Konsultasi Bab IV dan V	12.
13.	2 November 2025	Konsultasi Bab IV dan V	13.
14.	10 November 2025	Konsultasi Bab IV dan V	14.
15.	18 November 2025	Konsultasi Bab IV dan V	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
 Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	19 November 2025	ACC Bab IV dan V	16. <i>On</i>
17.	20 November 2025	Konsultasi Kajian Agama Bab IV	17. <i>Ral</i>
18.	23 November 2025	Konsultasi Kajian Agama Bab IV	18. <i>Ral</i>
19.	24 November 2025	ACC Kajian Agama Bab IV	19. <i>Ral</i>
20.	25 November 2025	ACC Seminar Hasil	20. <i>On</i>
21.	28 November 2025	ACC Seminar Hasil lanjutan	21. <i>On</i>
22.	5 Desember 2025	Konsultasi Revisi Seminar Hasil	22. <i>On</i>
23.	16 Desember 2025	Sidang Skripsi	23. <i>On</i>
24.	23 Desember 2025	ACC Keseluruhan	24. <i>On</i>

Malang, 23 Desember 2025

Mengetahui,

Ketua Program Studi Matematika

Dr. Fachrud Rozi, M.Si

NIP. 19800527 200801 1 012