



**PREDIKSI STRUKTUR SEKUNDER PROTEIN DENGAN ALGORITMA
SUPPORT VECTOR MACHINE MENGGUNAKAN METODE
RANDOMSEARCHCV**

SKRIPSI

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh
SORAYA IHDATI ANNISA
NIM. 1811016320014**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2025**

SKRIPSI

PREDIKSI STRUKTUR SEKUNDER PROTEIN DENGAN ALGORITMA SUPPORT VECTOR MACHINE MENGGUNAKAN METODE RANDOMIZEDSEARCHCV

Oleh :

SORAYA HIDATI ANNISA

NIM. 1811016320014

Telah dipertahankan di depan Dosen Penguji pada tanggal 3 Juli 2025 :

Susunan Penguji :

Pembimbing I



Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001

Penguji I



Muhammad Itqan Mazdadi, S.Kom, M.Kom
NIP. 199006122019031013

Pembimbing II



Dodon Turianto Nugrahadi, S.Kom, M.Eng
NIP. 198001122009121002

Penguji II



Friska Abadi, S.Kom, M.Kom
NIP. 198809132023211010

Banjarbaru, 4 Juli 2025

Koordinator Program Studi Ilmu Komputer




Pratiwi, S.Kom, M.Kom
198704212012122003

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 4 Juli 2025

Yang Menyatakan,



Soraya Ihdati Annisa
NIM. 1811016320014

ABSTRAK

PREDIKSI STRUKTUR SEKUNDER PROTEIN DENGAN ALGORITMA SUPPORT VECTOR MACHINE MENGGUNAKAN METODE RANDOMSEARCHCV (Oleh: Soraya Ihdati Annisa; Pembimbing: Irwan Budiman, S.T, M.Kom dan Dodon Turianto Nugrahadi, M.Eng., 2025; 67 halaman)

Bioinformatika merupakan suatu cabang ilmu yang menggabungkan antara ilmu biologi dan ilmu komputasi, salah satu permasalahan di bidang bioinformatika yaitu prediksi struktur sekunder protein. Prediksi struktur sekunder protein bertujuan untuk mencari fungsi protein dengan cara mengklasifikasikan urutan struktur primer protein ke dalam urutan struktur sekunder protein. Support Vector Machine (SVM) merupakan salah satu algoritma untuk menyelesaikan masalah klasifikasi. Dalam penelitian ini, penulis menggunakan algoritma Support Vector Machine untuk mengklasifikasikan urutan struktur primer protein dan urutan struktur sekunder protein. Selain itu, penulis juga menggunakan metode *RandomizedSearchCV* sebagai hyperparameter untuk meningkatkan hasil akurasi dari model yang ada. Dengan menggunakan empat jenis kernel, yaitu kernel Linier, Kernel Polynomial, Kernel RBF, dan Kernel Sigmoid. Hasil akhir menunjukkan bahwa penerapan algoritma SVM dengan *RandomizedSearchCV* pada setiap kernel memiliki akurasi yang beragam. Pada rasio pembagian data sebesar 60% data latih dan 40% data uji kernel Linier, kernel polynomial, kernel RBF, dan kernel Sigmoid berturut-turut menghasilkan nilai akurasi sebesar 59%, 76,9%,44,8%, dan42,9%. Pada rasio pembagian data sebesar 70% data latih dan 30% data uji kernel Linier, kernel polynomial, kernel RBF, dan kernel Sigmoid berturut-turut menghasilkan nilai akurasi sebesar 59,2%, 78,9%, 44,7%, dan 42,9%. Pada rasio pembagian data sebesar 80% data latih dan 20 data uji kernel Linier, kernel polynomial, kernel RBF, dan kernel Sigmoid berturut-turut menghasilkan nilai akurasi sebesar 59,2%, 80,2%, 45,5%, dan 41,9%. Hasil akurasi paling tinggi didapatkan dengan penggunaan SVM dengan *RandomizedSearchCV* pada kernel Polynomial memiliki akurasi yang paling tinggi yakni sebesar 76,9%, 79,8%, dan 80,2%.

Keywords : Prediksi Struktur Sekunder Protein, SVM, Sliding Window, *RandomizedSearchCV*, kernel, classification report.

ABSTRACT

PREDICTION OF SECONDARY PROTEIN STRUCTURE USING THE SUPPORT VECTOR MACHINE ALGORITHM WITH THE RANDOMSEARCHCV METHOD (By: Soraya Ihdati Annisa; Supervisor : Irwan Budiman, S.T, M.Kom and Dodon Turianto Nugrahadi, M.Eng., 2025; 66 pages)

Bioinformatics is a branch of science that combines biology and computer science. One of the problems in bioinformatics is predicting protein secondary structures. Predicting protein secondary structures aims to find protein functions by classifying protein primary structure sequences into protein secondary structure sequences. Support Vector Machine (SVM) is one of the algorithms used to solve classification problems. In this study, the author used the Support Vector Machine algorithm to classify protein primary structure sequences and protein secondary structure sequences. Additionally, the author also used the RandomizedSearchCV method as a hyperparameter to improve the accuracy of the existing model. Four types of kernels were used: Linear Kernel, Polynomial Kernel, RBF Kernel, and Sigmoid Kernel. The final results showed that the application of the SVM algorithm with RandomizedSearchCV on each kernel yielded varying levels of accuracy. At a data split ratio of 60% training data and 40% test data, the Linear kernel, Polynomial kernel, RBF kernel, and Sigmoid kernel yielded accuracy values of 59%, 76.9%, 44.8%, and 42.9%, respectively. At a data split ratio of 70% training data and 30% test data, the Linear kernel, polynomial kernel, RBF kernel, and Sigmoid kernel yielded accuracy values of 59.2%, 78.9%, 44.7%, and 42.9%, respectively. At a data split ratio of 80% training data and 20% test data, the Linear kernel, Polynomial kernel, RBF kernel, and Sigmoid kernel yielded accuracy values of 59.2%, 80.2%, 45.5%, and 41.9%, respectively. The highest accuracy results were obtained using SVM with RandomizedSearchCV on the polynomial kernel, which had the highest accuracy of 76.9%, 79.8%, and 80.2%.

Keywords: Protein Secondary Structure Prediction, SVM, Sliding Window, RandomizedSearchCV, Kernel, Classification report.

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Prediksi Struktur Sekunder Protein Dengan Algoritma Support Vector Machine Menggunakan Metode RandomizedSearchCV” untuk memenuhi syarat dalam menyelesaikan Pendidikan S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam kepada Nabi Muhammad SAW, keluarga, para sahabat dan pengikut beliau hingga yaumul qiyamah.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak- pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga terutama orang tua saya yang selalu memberikan bantuan, semangat, doa, dan dukungan dalam proses menyelesaikan skripsi ini.
2. Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing utama dan dosen pembimbing akademik yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Dodon Turianto Nugrahadi, S.Kom, M.Eng selaku dosen pembimbing pendamping yang turut membantu demi kelancaran dalam penyelesaian skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM atas bantuan dan izin beliau skripsi ini dapat diselesaikan
5. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom selaku dosen ketua penguji yang telah memberikan masukan-masukkan dalam skripsi ini.
6. Bapak Friska Abadi, S,Kom, M.Kom selaku dosen penguji yang telah memberikan masukan-masukkan dalam skripsi ini.
7. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.

8. Teman-teman dan kerabat khususnya teman-teman Ilmu Komputer 2018, yang sudah ikut serta dalam memberikan dukungan, memberikan doa serta membantu mengerjakan penelitian ini.
9. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini. Penulis berharap skripsi ini dapat bermanfaat memberikan sumbangan pikiran dan ilmu pengetahuan baik bagi penulis sendiri maupun bagi pembaca.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna, oleh karena itu penulis mengharapkan kritik dan saran yang membangun guna kesempurnaan skripsi ini untuk perbaikan dikemudian hari.

Banjarbaru, 4 Juli 2025



Soraya Ihdati Annisa

NIM. 1811016320014

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
SKRIPSI.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	2
1.3 TUJUAN PENELITIAN.....	2
1.4 MANFAAT PENELITIAN	3
1.5 BATASAN MASALAH.....	3
BAB II TINJAUAN PUSTAKA	4
2.1 KAJIAN TERDAHULU	4
2.2 LANDASAN TEORI.....	8
2.2.1 BIOINFORMATIKA	8
2.2.2 PROTEIN.....	10
2.2.3 DATA MINING.....	11
2.2.4 SUPPORT VECTOR MACHINE.....	12
2.2.5 RANDOMIZEDSEARCHCV.....	16

2.2.6	SLIDING WINDOW	18
2.2.7	CLASSIFICATION REPORT	19
BAB III METODE PENELITIAN.....		21
3.1	BAHAN PENELITIAN	21
3.2	ALAT PENELITIAN	22
3.3	ALUR PENELITIAN.....	23
BAB IV HASIL DAN PEMBAHASAN.....		26
4.1	HASIL	26
4.1.1	Pengumpulan Data	26
4.1.2	Preprocessing	27
4.1.3	Ekstraksi Fitur	32
4.1.4	Data Training dan Testing	33
4.1.5	Pemodelan SVM dan RandomizedSearchCV	34
4.1.6	Evaluasi.....	35
4.2	PEMBAHASAN	35
BAB V PENUTUP.....		65
5.1	Kesimpulan	65
5.2	Saran.....	65
DAFTAR PUSTAKA		66

DAFTAR GAMBAR

Gambar 1. Prediksi Struktur Sekunder Protein	11
Gambar 2. Ilustrasi SVM untuk Data Linier	13
Gambar 3. Ilustrasi Linier Kernel	14
Gambar 4. Ilustrasi Non-Linear Kernel	14
Gambar 5. Ilustrasi Kernel linier.....	15
Gambar 6. Ilustrasi Kernel Sigmoid.....	15
Gambar 7. Ilustrasi Kernel RBF.....	16
Gambar 8. Ilustrasi Random Search.....	16
Gambar 9. Ilustrasi Cross-Valodation	17
Gambar 10. One-Hot-Encoding pada Asam Amino.....	19
Gambar 11. Alur Kerja Penelitian	23
Gambar 12. Tampilan Dataset RS126	26
Gambar 13. Hasil Pemisahan Struktur Primer dan Struktur Sekunder	26
Gambar 14. Data Tidak Seimbang	27
Gambar 15. Sampel Jumlah Karakter Sama	28
Gambar 16. Hasil Split.....	29
Gambar 17. Hasil Konversi Struktur Primer.....	29
Gambar 18. Hasil Representasi Struktur Sekunder.....	30
Gambar 19. Vektor Baru Setelah Penjumlahan.....	30
Gambar 20. Pembentukan Struktur Primer menjadi Struktur Sekunder.....	31
Gambar 21. Tampilan y_Label Sebelum Flatten.....	31
Gambar 22. Tampilan y_Label setelah Flatten.....	31
Gambar 23. Tampilan Sebelum Sliding Window	32
Gambar 24. Hasil Sliding Window	32
Gambar 25. Confusion Matrix Pada Kernel Linier dengan Rasio 70% dan 30%	38
Gambar 26. Confusion Matrix Pada Kernel Polynomial dengan Rasio 70% dan 30%	40
Gambar 27. Confusion Matrix Pada Kernel RBF dengan Rasio 70% dan 30%... ..	43
Gambar 28. Confusion Matrix Pada Kernel Sigmoid Rasio 70% dan 30%.....	45
Gambar 29. Confusion Matrix Pada Kernel Linier dengan Rasio 60% dan 40%. ..	48

Gambar 30. Confusion Matrix Pada Kernel Polynomial dengan Rasio 60% dan 40%	50
Gambar 31. Confusion Matrix Pada Kernel RBF dengan Rasio 60% dan 40%...	52
Gambar 32. Confusion Matrix Pada Kernel Sigmoid dengan Rasio 60% dan 40%	54
Gambar 33. Confusion Matrix Pada Kernel Linier dengan Rasio 80% dan 20%.	56
Gambar 34. Confusion Matrix Pada Kernel Polynomial dengan Rasio 80% dan 20%	58
Gambar 35. Confusion Matrix Pada Kernel RBF dengan Rasio 80% dan 20%...	60
Gambar 36. Confusion Matrix Pada Kernel Sigmoid dengan Rasio 80% dan 20%	63
Gambar 37. Grafik Akurasi SVM dengan RandomizedSearchCV pada Kernel...	64

DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perancangan Penelitian	8
Tabel 3. Simbol dan Singkatan Asam Amino.....	10
Tabel 4. Sampel Dataset RS126	21
Tabel 5. Representasi 20 Jenis Asam Amino dalam bentuk numerik	21
Tabel 6. Klasifikasi 20 Jenis Asam Amino Ke Dalam Kelas Target	22
Tabel 7. Jumlah Seluruh Sekuens.....	28
Tabel 8. Representasi Struktur Sekunder	30
Tabel 9. Distribusi y_label Rasio 60% dan 40%.....	33
Tabel 10. Distribusi y_label dengan Rasio 70% dan 30%	33
Tabel 11. Distribusi y_label dengan Rasio 80% dan 20%	34
Tabel 12. Kombinasi Nilai Parameter Kernel Linier pada Rasio 70% dan 30% ..	36
Tabel 13. Hasil Clasification Report dengan Kernel Linier pada Rasio 70% dan 30%	37
Tabel 14. Kombinasi Nilai Parameter Kernel Polynomial pada Rasio 70% dan 30%	39
Tabel 15. Hasil Clasification Report dengan Kernel Polynomial pada Rasio 70% dan 30%.....	40
Tabel 16. Kombinasi Nilai Parameter Kernel RBF pada Rasio 70% dan 30%.....	41
Tabel 17. Hasil Clasification Report dengan Kernel RBF pada Rasio 70% dan 30%	42
Tabel 18. Kombinasi Nilai Parameter Kernel Sigmoid pada Rasio 70% dan 30%	44
Tabel 19. Hasil Clasification Report dengan Kernel Sigmoid pada Rasio 70% dan 30%	45
Tabel 20. Kombinasi Nilai Parameter Kernel Linier pada Rasio 60% dan 40% ..	46
Tabel 21. Hasil Clasification Report dengan Kernel Linier pada Rasio 60% dan 40%	47
Tabel 22. Kombinasi Nilai Parameter Kernel Polynomial pada Rasio 60% dan 40%	48

Tabel 23. Hasil Clasification Report dengan Kernel Polynomial pada Rasio 60% dan 40%.....	50
Tabel 24. Kombinasi Nilai Parameter Kernel RBF pada Rasio 60% dan 40%.....	50
Tabel 25. Hasil Clasification Report dengan Kernel RBF pada Rasio 60% dan 40%	51
Tabel 26. Kombinasi Nilai Parameter Kernel Sigmoid pada Rasio 60% dan 40%	52
Tabel 27. Hasil Clasification Report dengan Kernel Sigmoid pada Rasio 60% dan 40%	53
Tabel 28. Kombinasi Nilai Parameter Kernel Linier pada Rasio 80% dan 20% ..	54
Tabel 29. Hasil Clasification Report dengan Kernel Linier pada Rasio 80% dan 20%	56
Tabel 30. Kombinasi Nilai Parameter Kernel Polynomial pada Rasio 80% dan 20%	56
Tabel 31. Hasil Clasification Report dengan Kernel Polynomial pada Rasio 80% dan 20%.....	58
Tabel 32. Kombinasi Nilai Parameter Kernel RBF pada Rasio 80% dan 20%.....	58
Tabel 33. Hasil Clasification Report dengan Kernel RBF pada Rasio 80% dan 20%	60
Tabel 34. Kombinasi Nilai Parameter Kernel Sigmoid pada Rasio 80% dan 20%	61
Tabel 35. Hasil Clasification Report dengan Kernel Sigmoid pada Rasio 80% dan 20%	62
Tabel 36. Perbandingan Hasil Clasification Report setiap Kernel	63

DAFTAR LAMPIRAN

- Lampiran 1. Source Code library pada Python
- Lampiran 2. Source Code untuk membaca File dataset
- Lampiran 3. Source Code untuk memeriksa Panjang sekuens
- Lampiran 4. Source Code Untuk Split Sekuens
- Lampiran 5. Source Code untuk One-Hot-Encoding pada Sekuens Primer
- Lampiran 6. Source Code Untuk Label-Encoding Pada Sekuens Sekunder
- Lampiran 7. Source Code Untuk Membuat Graph Input Sebagai Masukan
- Lampiran 8. Source Code untuk Label Target
- Lampiran 9. Source Code untuk Sliding Window
- Lampiran 10. Source Code untuk Standarisasi dan Pembagian Data
- Lampiran 11. Source Code untuk Pemodelan dan Evaluasi