

## Implementasi Algoritme *Support Vector Machine* dan Fitur Seleksi MRMR untuk Prediksi Glikosilasi

<sup>1</sup>Favorisen Rosyking Lumbanraja, <sup>2</sup>Naurah Nazhifah, <sup>3</sup>Dewi Asiah Shofiana dan <sup>4</sup>Akmal Junaidi

<sup>1,2,3,4</sup> Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung  
Jalan Soemantri Brojonegoro No.1 Gedung Meneng, Bandar Lampung, Provinsi Lampung, Indonesia  
e-mail: <sup>1</sup>[favorisen.lumbanraja@fmipa.unila.ac.id](mailto:favorisen.lumbanraja@fmipa.unila.ac.id), <sup>2</sup>[naurahnazhifah1309@gmail.com](mailto:naurahnazhifah1309@gmail.com), <sup>3</sup>[dewi.asiah@fmipa.unila.ac.id](mailto:dewi.asiah@fmipa.unila.ac.id),  
<sup>4</sup>[akmal.junaidi@fmipa.unila.ac.id](mailto:akmal.junaidi@fmipa.unila.ac.id)

---

**Abstract** — During the protein formation process, there are post-translational modifications that provide additional properties and produce new R groups in the polypeptide chain. One of the results of post-translational modifications is glycosylation. Glycosylation reactions occur between protein and glucose at high concentrations. There are 3 categories of glycosylation found in the human body, namely N-glycosylation, O-glycosylation and C-glycosylation. To determine the functional role of glycosylation, that is by predicting the substrate of each glycosylation site. A computational approach is a way to predict the glycosylation site, using the Support Vector Machine (SVM) algorithm. In this study there are 2 types of data, namely independent data and benchmark data. The features used are feature extraction and feature selection using Maximum Redundancy Minimum Relevance (MRMR) of 25, 50 and 75 columns. SVM classification test using 5-fold cross validation. The highest accuracy result lies in the use of the 75 column MRMR selection feature. In Independent N data, the greatest accuracy lies in the sigmoid kernel with a causation value of 86.66%, while for independent C data, the accuracy is 87.5% in the sigmoid kernel and for independent O data, the largest accuracy is 89.31% which is in the RBF kernel. For benchmark N data, the highest accuracy is 70.54% in the RBF kernel, for benchmark C data the greatest accuracy lies in the RBF kernel with a value of 95.06% and for benchmak O data it is in the RBF kernel with the greatest accuracy, which is 92.64%.

**Keywords:** Glycosylation; Maximum Redundancy Minimum Relevance; Post Translation Modification; Support Vector Machine.

---

### 1. PENDAHULUAN

Pada tubuh manusia terdapat ribuan jenis sel yang masing-masing selnya mengandung ribuan protein. Kumpulan protein-protein pada sel dalam waktu tertentu disebut dengan proteome [1]. Pada proses penentuan urutan protein banyak mengalami modifikasi pasca-translasi. Modifikasi ini memberikan sifat tambahan yang menghasilkan gugus R baru pada rantai polipeptida salah satunya adalah glikosilasi [2].

Glikosilasi adalah proses penggabungan glukosa pada struktur protein yang terjadi pada retikulum endoplasma dan badan golgi [3]. Modifikasi pascatranslasi atau PTM pada proses seluler memainkan peran yang sangat besar, karena sebagian organisme eukariotik mengalami glikosilasi [4]. Sebagian besar protein pada modifikasi pasca-translasi ini berperan penting untuk sintesis dan pergantian protein, metabolisme nitrogen, siklus sel, dormansi dan lain-lain [5]. Salah satu penyebab terjadinya modifikasi pada protein ini adalah pembebanan glukosa melalui reaksi glikosilasi [6].

Reaksi glikosilasi merupakan reaksi antara gugus amina protein dengan gugus aldehida yang dapat membentuk produk reaktif, yang kemudian dapat memodifikasi protein [6]. Reaksi ini terjadi antara protein dan glukosa pada saat konsentrasi tinggi. Glikosilasi dapat dilihat menjadi empat kategori yaitu N-glikosilasi, O-glikosilasi, C-glikosilasi, dan glikosil-fosfatidylinositol [7], tetapi pada manusia glikosilasi paling umum terjadi pada residu N-glikosilasi, O-glikosilasi dan C-glikosilasi.

N-glikosilasi adalah suatu proses yang melibatkan banyak protein dimodifikasi pada semua domain kehidupan [8]. N-glikosilasi merupakan glycan yang mengikat kelompok amino asparagine di retikulum endoplasma (ER) [9]. Pada residu treonin O-glikosilasi adalah proses monosakarida yang berikatan dengan gugus hidroksil Serin (Ser) atau Treonin (Thr) di retikulum endoplasma (ER), badan golgi, sitosol, dan nucleus [10], sedangkan C-glikosilasi merupakan reaksi pembentukan ikatan karbon yang berada di pusat anomerik karbohidrat [11].

Untuk memahami mekanisme dan peran fungsional glikosilasi lainnya adalah dengan cara mengidentifikasi substratnya dari situs glikosilasi yang sesuai. Namun, untuk mengetahuinya masih merupakan suatu proses yang memerlukan laboratorium. Pendekatan komputasi merupakan salah satu cara untuk memprediksi ikatan pada sekuens protein yaitu menggunakan fitur seleksi Minimum Redundancy Maximal Relevance (MRMR) dan algoritme Support Vector Machine(SVM).

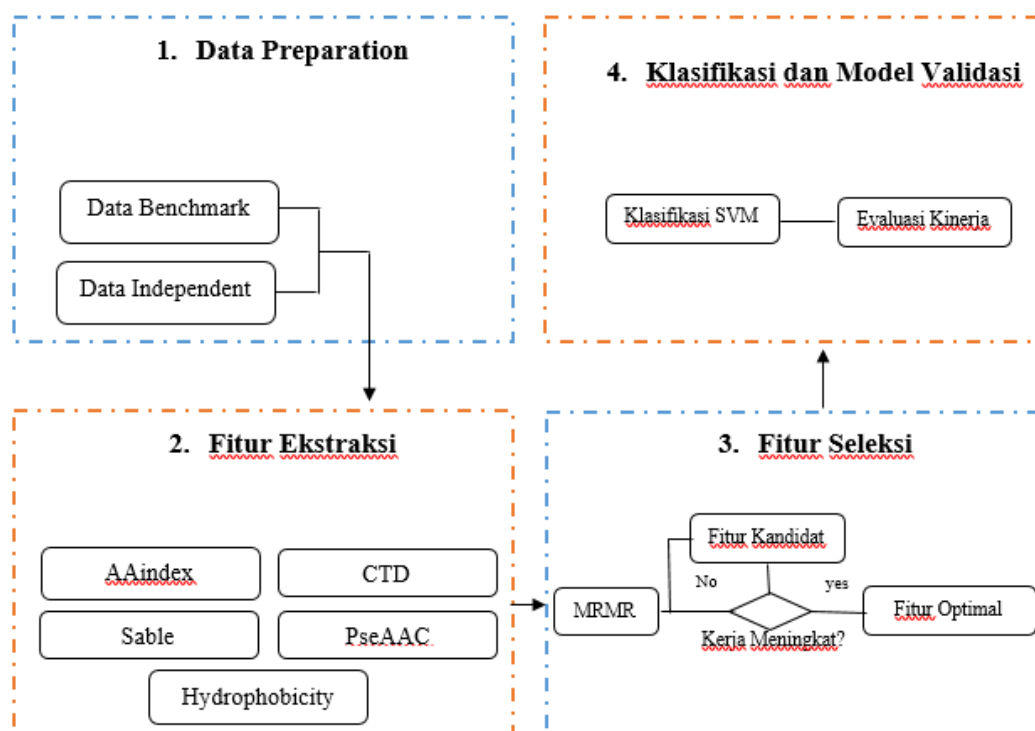
Minimum Redundancy Maximal Relevance (MRMR) adalah suatu metode pemilihan fitur yang dapat menemukan kumpulan fitur yang relevan dan tidak redundan dengan cara menggunakan mutual information (MI) [12]. Support Vector Machine (SVM) merupakan metode perhitungan yang baik dalam mendapatkan hasil klasifikasi dengan tingkat akurasi tinggi. Support Vector Machine (SVM) merupakan suatu metode yang bekerja dengan cara mendefinisikan suatu kelas dan mencari hyperplane dengan margin terbesar. Metode ini diketahui telah banyak dilakukan untuk prediksi dan pengklasifikasian [13].

Pada penelitian sebelumnya prediksi glikosilasi pada GlycoMine telah dilakukan menggunakan algoritmerandom forest yang menghasilkan bahwa GlycoMine dapat memberikan pendekatan yang kuat dan mengungguli alat yang sudah ada seperti GPP, EnsembleGly, dan NetNGlyc [14]. Metode Support Vector Machine (SVM) pernah digunakan dalam penelitian sebelumnya untuk Implementasi support vector machine (SVM) dan random forest (RF) pada diagnosis kanker payudara. Dari hasil yang didapat metode random forest (RF) yang tingkat akurasinya lebih tinggi dari metode SVM dengan tingkat akurasi sebesar 94,52% [15].

Berdasarkan permasalahan yang telah dijelaskan dan penelitian sebelumnya, pada penelitian ini bertujuan untuk melihat kemampuan SVM dengan melakukan prediksi glikosilasi pada N-, C- dan O- pada proteom manusia menggunakan fitur seleksi *minimum redundancy maximal relevance* dan algoritme *support vector machine*.

## 2. METODOLOGI PENELITIAN

Alur pengerjaan penelitian penggunaan algoritme *support vector machine* dan fitur seleksi MRMR untuk prediksi glikosilasi dapat dilihat pada Gambar 1.



Gambar 1 Alur pengerjaan penelitian prediksi glikosilasi pada n-, c- dan o- dalam proteom manusia menggunakan algoritme *support vector machine*.

## 2.1. Persiapan Data

Data glikosilasi dari situs N-, C- dan O- didapat dari *database* Uniprot. Data tersebut terdiri dari positif negatif dataset *benchmark* dan dataset *independent*. Data glikosilasi dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Dataset *independent* glikosilasi

Data	+	-	Total
N	83	166	249
C	13	28	41
O	129	258	387

Tabel 2. Dataset *benchmark* glikosilasi

Data	+	-	Total
N	333	667	1000
C	55	108	163
O	520	1018	1538

## 2.2. Fitur Ekstraksi

Ekstraksi fitur adalah suatu proses yang penting dalam mengklasifikasikan suatu objek untuk menemukan pemetaan fitur yang asli kedalam fitur yang baru. Hal ini dilakukan untuk meningkatkan tingkat akurasi hasil pengklasifikasian suatu objek [16]. Ekstraksi fitur yang digunakan yaitu *Solvent Accessibilities* (Sable) yang berfungsi sebagai prediksi fitur struktural protein, seperti struktur sekunder dan aksesibilitas pelarut yang relative. AAindex adalah *database* indeks numerik yang mewakili berbagai fisikokimia dan biokimia sifat asam amino dan pasangan asam amino. *Composition, Transition, And Distribution* (CTD) fitur yang berfungsi untuk memprediksi lokasi protein yang akan

ditargetkan. *Hydrophobicity* adalah sifat fisik suatu senyawa yang memungkinkan untuk ditolak oleh massa air, dan *Pseudo Amino Acid Composition* (PseAAC) adalah metode yang berfungsi sebagai alat untuk merumuskan urutan biologis dengan vektor, tetapi masih menyimpan informasi urutan-urutannya.

### 2.3. Fitur Seleksi

Tahap seleksi fitur berguna untuk mengurangi suatu dimensi data, menghilangkan data yang tidak akurat dan meningkatkan tingkat akurasi data tersebut [17]. Jenis fitur seleksi yang digunakan pada penelitian ini adalah *Minimum Redundancy Maximal Relevance* (MRMR). MRMR merupakan seleksi fitur yang memilih fitur dengan korelasi tertinggi dengan kelas (*output*) dan korelasi terendah. Pemilihan fitur menggunakan MRMR sebanyak 25, 50 dan 75 kolom.

### 2.4. Pemodelan SVM

Pemodelan klasifikasi menggunakan metode *Support Vector Machine* (SVM) dimana pemodelan SVM menggunakan empat Kernel yaitu *linear*, *Polynomial*, RBF dan *Sigmoid*. Penjelasan mengenai kernel SVM dapat dilihat pada Tabel 3.

Tabel 3. *Kernel support vector machine*

Jenis Kernel	Definisi
<i>Linear</i>	$K(\bar{X}_i, \bar{X}_j) = (\bar{X}_i \cdot \bar{X}_j)$
<i>Polynomial</i>	$K(\bar{X}_i, \bar{X}_j) = (\bar{X}_i \cdot \bar{X}_j + 1)^p$
Gaussian	$K(\bar{X}_i, \bar{X}_j) = \exp\left(-\frac{\ \bar{x}_i - \bar{x}_j\ ^2}{2\sigma^2}\right)$
Sigmoid	$K(\bar{X}_i, \bar{X}_j) = \tanh(\alpha \bar{x}_i \cdot \bar{x}_j + \beta)$

### 2.5. Pengujian Klasifikasi

Pengujian klasifikasi pada penelitian ini menggunakan *Confusion matrix*. *Confusion matrix* memiliki perhitungan matriks yang digunakan untuk mengukur performa dari predictor [18]. Penjelasan tentang *confusion matrix* dapat dilihat pada Tabel 4.

Tabel 4. *Confusion matrix*

<i>Predicted</i>	<i>Actual</i>	
	<i>Non-event</i>	<i>Event</i>
<i>Non-event</i>	TN	FN
<i>Event</i>	FP	TP

*Confusion matrix* memiliki hasil statistik untuk penilaian kinerja. Hasil tersebut diantaranya adalah akurasi (ACC), sensitivitas (SE), spesifisitas (SP), dan *matthew correlation coefficient* (MCC). Berikut adalah rumus hasil statistik untuk penilaian kinerja yang dapat dilihat dari persamaan 1 sampai persamaan 4.

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \quad (1)$$

$$SN = \frac{TP}{TP+FN} * 100\% \quad (2)$$

$$SP = \frac{TN}{TN+FP} * 100\% \quad (3)$$

$$MCC = \frac{TP*TN-FP*FN}{(TP+FN)(TP+FP)(TN+FP)(TN+FN)} * 100\% \quad (4)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Fitur Ekstraksi

Pada tahap ini fitur ekstraksi yang digunakan sebanyak 5 fitur yang menghasilkan jumlah dimensi yang berbeda-beda. Jumlah dimensi dapat dilihat pada Tabel 5.

Tabel 5. Dimensi ekstraksi fitur

Fitur Ekstraksi	Dimensi	%
CTD	21	23,33%
AAindex	15	16,67%
Hydrophobicity	15	16,67%
PseAAC	24	26,66%
Sable	15	16,67%
<b>Jumlah</b>	<b>90</b>	<b>100%</b>

Jumlah dimensi fitur ekstraksi paling banyak terdapat pada *Pseudo Amino Acid Composition* (PseAAC) sebanyak 24 dimensi.

#### 3.2. Fitur Seleksi

Pemilihan fitur menggunakan MRMR dengan pemilihan sebanyak 25, 50 dan 75 kolom. Pemilihan fitur ini memiliki konsep yaitu memilih fitur dengan nilai korelasi tinggi dan korelasi terendah.

#### 3.3. Hasil Klasifikasi

Data glikosilasi di klasifikasi dengan model algoritma *support vector machine* dengan menggunakan empat kernel, yaitu linear, *polynomial*, RBF dan *sigmoid*. Setiap model SVM yang menggunakan fitur seleksi MRMR memberikan nilai yang berbeda-beda. Berikut hasil dari pemodelan SVM yang menggunakan fitur seleksi MRMR dan tanpa menggunakan fitur seleksi MRMR pada data *independent* yang dapat dilihat pada Tabel 6,7 dan 8.

Tabel 6. *Independent N*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>N (Tanpa MRMR)</b>	RBF	60,89	6,58	93,05	0,45
<b>N MRMR 25</b>	Polynomial	67,5	12,21	61,15	11,89
<b>N MRMR 50</b>	Linear	84,16	79,30	88,37	66,49
<b>N MRMR 75</b>	<b>Sigmoid</b>	<b>86,66</b>	<b>68,45</b>	<b>95,35</b>	<b>69,10</b>

Tabel 7. *Independent C*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>C (Tanpa MRMR)</b>	Polynomial	82,50	86,66	81,06	64,90
<b>C MRMR 25</b>	RBF	82,5	45,47	97,71	55,44
<b>C MRMR 50</b>	Linear	80,83	80,56	81,71	61,12
<b>C MRMR 75</b>	<b>Sigmoid</b>	<b>87,5</b>	<b>57,14</b>	<b>100</b>	<b>57,99</b>

Tabel 8. *Independent O*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>O (Tanpa MRMR)</b>	Polynomial	79,31	78,89	79,70	0,45
<b>O MRMR 25</b>	RBF	85,38	72,43	92,56	11,89
<b>O MRMR 50</b>	<b>RBF</b>	<b>89,31</b>	<b>75,98</b>	<b>96,23</b>	<b>66,49</b>
<b>O MRMR 75</b>	RBF	88,72	76,94	95,25	69,10

Berdasarkan Tabel 6, 7 dan 8 pemodelan dengan menggunakan SVM dan fitur seleksi MRMR pada data *independent* memiliki nilai akurasi lebih besar dibandingkan dengan pemodelan tanpa menggunakan fitur seleksi MRMR.

Selanjutnya adalah pemodelan SVM yang menggunakan fitur seleksi MRMR dan tanpa menggunakan fitur seleksi MRMR pada data *benchmark* yang dapat dilihat pada Tabel 9,10 dan 11.

Tabel 9. *Benchmark N*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>N (Tanpa MRMR)</b>	Linear	66,01	45,95	78,35	25,66
<b>N MRMR 25</b>	RBF	65,83	27,89	88,90	21,17
<b>N MRMR 50</b>	RBF	70,05	41,20	87,30	32,61
<b>N MRMR 75</b>	<b>RBF</b>	<b>70,54</b>	<b>42,43</b>	<b>87,36</b>	<b>33,96</b>

Tabel 10. *Benchmark C*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>C (Tanpa MRMR)</b>	Polynomial	80	82,64	81,03	62,60
<b>C MRMR 25</b>	Sigmoid	91,52	89,38	92,86	82,61
<b>C MRMR 50</b>	RBF	94,11	88,89	97,32	87,71
<b>C MRMR 75</b>	<b>RBF</b>	<b>95,06</b>	<b>92,01</b>	<b>96,68</b>	<b>89,27</b>

Tabel 11. *Benchmark O*

Jenis Data	Kernel	ACC	SN	SP	MCC
<b>O (Tanpa MRMR)</b>	Linear	88,524	85,132	90,39	75,066
<b>O MRMR 25</b>	RBF	87,022	77,594	92,286	71,412
<b>O MRMR 50</b>	Sigmoid	91,904	86,58	94,878	82,292
<b>O MRMR 75</b>	<b>RBF</b>	<b>92,64</b>	<b>86,31</b>	<b>96,178</b>	<b>83,88</b>

Berdasarkan Tabel 9, 10 dan 11 pemodelan yang menggunakan algoritme SVM dan fitur seleksi MRMR pada data *benchmark* memiliki nilai akurasi lebih besar dibandingkan dengan pemodelan tanpa menggunakan fitur seleksi MRMR.

### 3.4. Perbandingan Penelitian Sebelumnya

Hasil penelitian yang sudah didapatkan, selanjutnya dibandingkan dengan penelitian sebelumnya yaitu pada penelitian Li et al. (2015) [14] yang menggunakan klasifikasi *random forest*. Perbandingan ini dilakukan untuk mendapatkan kesimpulan penggunaan algoritme klasifikasi *support vector machine* untuk memprediksi glikosilasi pada tubuh manusia. Berikut perbandingan hasil pengujian data glikosilasi yang menggunakan algoritme SVM dan fitur seleksi MRMR sebanyak 75 kolom pada data *independent* yang dapat dilihat pada Tabel 12 hingga Tabel 14.

Tabel 12. Perbandingan data *independent* N

Algoritme	ACC	SN	SP	MCC
<b>SVM (Tanpa MRMR)</b>	60,89	6,58	93,05	0,45
<b>SVM (MRMR 75)</b>	86,66	68,45	95,35	69,10
<b>RF MRMR 500 (Li et al., 2015)</b>	95,58	95,18	95,78	90,17

Tabel 13. Perbandingan data *independent* C

Algoritme	ACC	SN	SP	MCC
<b>SVM (Tanpa MRMR)</b>	82,50	86,66	81,06	64,90
<b>SVM (MRMR 75)</b>	87,5	57,14	100	57,99
<b>RF MRMR 500 (Li et al., 2015)</b>	100	100	100	100

Tabel 14. Perbandingan data *independent* O

Algoritme	ACC	SN	SP	MCC
<b>SVM (Tanpa MRMR)</b>	79,31	78,89	79,70	57,16
<b>SVM (MRMR 75)</b>	88,72	76,94	95,25	74,95
<b>RF MRMR 500 (Li et al., 2015)</b>	90,70	86,05	93,02	79,07

Pada Tabel 12 hingga Tabel 14 menunjukkan bahwa klasifikasi *random forest* dengan menggunakan fitur seleksi MRMR memiliki nilai akurasi lebih besar dibandingkan dengan penggunaan klasifikasi *support vector machine* yang menggunakan fitur seleksi MRMR 75 kolom.

Selanjutnya yaitu perbandingan hasil pengujian data glikosilasi yang menggunakan algoritme SVM dan fitur seleksi MRMR sebanyak 75 kolom pada data *benchmark* yang dapat dilihat pada Tabel 15 hingga Tabel 17.

Tabel 15. Perbandingan data *benchmark* N

Algoritme	ACC	SN	SP	MCC
<b>SVM (Tanpa MRMR)</b>	66,01	45,95	78,35	25,66
<b>SVM (MRMR 75)</b>	70,54	42,43	87,36	33,96
<b>RF MRMR 500 (Li et al., 2015)</b>	95	93,39	95,8	88,8

Tabel 16. Perbandingan data *benchmark C*

Algoritme	ACC	SN	SP	MCC
SVM (Tanpa MRMR)	80,00	82,64	81,03	62,60
SVM (MRMR 75)	95,06	92,01	96,68	89,27
RF MRMR 500 (Li et al., 2015)	100	100	100	100

Tabel 17. Perbandingan data *benchmark O*

Algoritme	ACC	SN	SP	MCC
SVM (Tanpa MRMR)	88,52	85,13	90,39	75,06
SVM (MRMR 75)	92,64	86,31	96,17	83,88
RF MRMR 500(Li et al., 2015)	86,61	84,23	87,82	70,78

Sama seperti pada data data independent. Pada Tabel 15 hingga Tabel 17 menunjukkan bahwa klasifikasi *random forest* dengan menggunakan fitur seleksi MRMR memiliki nilai akurasi lebih besar dibandingkan dengan penggunaan klasifikasi *support vector machine* yang menggunakan fitur seleksi MRMR 75 kolom.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan prediksi glikosilasi menggunakan fitur seleksi MRMR memiliki hasil yang sangat bagus dibandingkan tidak menggunakan fitur seleksi MRMR untuk data *benchmark* dan data *independent*. Perbandingan hasil prediksi glikosilasi menggunakan algoritme support vector machine dengan fitur seleksi MRMR dengan 75 fitur jika dibandingkan dengan penelitian (Li et al., 2015) yang menggunakan fitur seleksi sebanyak 500 fitur hasilnya memperlihatkan margin akurasi dari keduanya tidak terlalu jauh.

#### DAFTAR PUSTAKA

- [1] R. K. Murray, D. K. Granner & V. W. Rodwell, *Biokimia Harper Edisi 27*, Buku Kedokteran EGC, 2012.
- [2] V. W. Rodwell, D. A. Bender, K. M. Botham, P. J. Kennely & P. A. Weil, *Harper's Illustrated Biochemistry (31st Edition)*. New York: McGraw-Hill Education, 2018.
- [3] Y. Mazola, G. Chinae, & A. Musacchio, "Integrating bioinformatics tools to handle glycosylation," *PLoS Computational Biology*, 7 (12), 2011.
- [4] M. Mann & O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, 21, pp. 255-261, 2003.
- [5] B. Macek, K. Forchhammer, J. Hardouin, E. Weber-Ban, C. Grangeasse & I. Mijakovic, "Protein post-translational modifications in bacteria," *Nature Reviews Microbiology*, 17 (11), pp. 651-664, 2019.
- [6] E. Suhartono, B. Setiawan, Mashuri, M. Juniarti, I. Kamilah & Haudhiya, "Modifikasi Protein Akibat Pembebanan Glukosa dengan Model Reaksi Glikosilasi Non Enzimatik," *Jurnal Kedokteran dan Kesehatan*, Vol 8 No. 1, 2008.
- [7] K. Ohtsubo & J. D. Marth, "Glycosylation in Cellular Mechanisms of Health and Disease," *Cell*, 126 (5), pp. 855-867, 2006.
- [8] F. Schwarz & M. Aebi, "Mechanisms And Principles Of N-Linked Protein Glycosylation," *Current Opinion in Structural Biology*, Vol. 21, pp. 576-582, 2011.

- [9] M. Aebi, "N-Linked Protein Glycosylation In The ER," *Biochimica et Biophysica Acta - Molecular Cell Research*, 1833 (11), pp. 2430-2437, 2013.
- [10] J. Peter-Katalinić, "Methods In Enzymology: O-Glycosylation Of Proteins," *Methods in Enzymology*, 405, pp. 139-171, 2005.
- [11] T. Nishikawa, M. Adachi & M. Isobe, "C-Glycosylation," *Glycoscience*, pp. 755-811, 2008.
- [12] N. De Jay, S. Papillon-Cavanagh, C. Olsen, N. El-Hachem, G. Bontempi & B. Haibe-Kains, "Mrmre: An R Package For Parallelized Mrmr Ensemble Feature Selection," *Bioinformatics*, 29 (18), 2013.
- [13] N. Prakash, "Emerging Trends of Face Recognition: A Review," *Ti Journals*, Vol 12 (2), pp. 242-250, 2012.
- [14] F. Li, C. Li, M. Wang, G. I. Webb, Y. Zhang, J. C. Whisstock & J. Song, "Glycomine: A Machine Learning-Based Approach For Predicting N-, C-And O-Linked Glycosylation In The Human Proteome," *Bioinformatics*, Vol 31 (9), pp. 1411-1419, 2015.
- [15] H. Aliady, N. J. Tuasikal & E. Widodo, "Implementasi Support Vector Machine (SVM) dan Random Forest pada Diagnosis Kanker Payudara," *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA)*, 278-285, 2018.
- [16] L. Guo, D. Rivero, J. Dorado, C. R. Munteanu & A. Pazos, "Automatic Feature Extraction Using Genetic Programming: An Application To Epileptic EEG Classification," *Expert System with Applications*, Vol 38 (8), pp. 10425-10436, 2011.
- [17] L. Yu & H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," Department of Computer Science & Engineering, Arizona State University, 2003.
- [18] M. Kuhn & K. Johnson, *Applied predictive modeling*. Berlin: Springer, 2013.