

# PREDIKSI KINERJA MAHASISWA DALAM PERKULIAHAN DARING BERBASIS *E-LEARNING* MENGGUNAKAN METODE *LOGISTIC REGRESSION*

Agustya Nanda Pratiwi<sup>1</sup>, Taghfirul Azhima Yoga Siswa<sup>2</sup>  
Universitas Muhammadiyah Kalimantan Timur<sup>1</sup>  
Jalan Jl. Ir. H. Juanda No.15 Samarinda  
Sur-el : [ndaaprtwii@gmail.com](mailto:ndaaprtwii@gmail.com)<sup>1</sup>, [tay758@umkt.ac.id](mailto:tay758@umkt.ac.id)<sup>2</sup>

---

**Abstract :** *There are many problems that occur in the online learning process, one of which is the difficulty of students in understanding the material well. Various efforts have been declared by lecturers to support online learning, starting from direct material explanations through OpenLearning, Zoom, and Google Meet media. To find out whether the student's performance in this online lecture is good or not. Prediction of student performance in online lectures is used as one of the supports for evaluation decisions at the University of Muhammadiyah, East Kalimantan. The purpose of this study is to determine indicators, implement and evaluate the performance of the Logistic Regression algorithm using the confusion matrix to see student performance in online lectures. The number of data used in this study was 2663 data on odd semester citizenship courses in 2020/2021 and 2021/2022. . The results of the Logistic Regression algorithm using 80% training data sharing and 20% testing data obtained an accuracy value of 91.66%.*

**Keywords:** *Prediction, Student Performance, Online Lectures, Logistic Regression, Accuracy*

**Abstrak :** *Banyak masalah yang terjadi dalam proses pembelajaran daring, salah satunya adalah kesulitan mahasiswa dalam memahami materi dengan baik. Berbagai upaya telah dilakukan dosen untuk mendukung pembelajaran secara daring, mulai dari penjelasan materi langsung melalui media OpenLearning, Zoom, dan Google Meet. Untuk mengetahui apakah kinerja mahasiswa dalam perkuliahan daring ini baik atau tidak. Prediksi kinerja mahasiswa dalam perkuliahan daring digunakan sebagai salah satu penunjang keputusan evaluasi pada Universitas Muhammadiyah Kalimantan Timur. Tujuan penelitian ini ialah menentukan indikator, mengimplementasi dan mengevaluasi kinerja algoritma Logistic Regression menggunakan Confusion Matrix untuk melihat kinerja mahasiswa dalam perkuliahan daring. Jumlah data yang digunakan pada penelitian ini sebanyak 2663 data pada mata kuliah kewarganegaraan semester ganjil pada tahun 2020/2021 dan 2021/2022. Hasil algoritma Logistic Regression dengan menggunakan pembagian data training 80% dan 20% data testing diperoleh nilai akurasi sebesar 91.66%.*

**Kata kunci:** *Prediksi, Kinerja Mahasiswa, Perkuliahan Daring, Logistic Regression, Akurasi.*

---

## 1. PENDAHULUAN

Pada saat ini negara Indonesia sedang dilanda pandemi yang mengakibatkan dunia pendidikan masih menerapkan perkuliahan daring. Salah satu perguruan tinggi yang menerapkan pembelajaran daring adalah Universitas Muhammadiyah Kalimantan Timur *Prediksi Kinerja Mahasiswa Dalam Perkuliahan Daring berbasis E learning ... ..*  
(Taghfirul Azhima Yoga Siswa, Agustya Nanda Pratiwi)

(UMKT). Perkuliahan daring menyebabkan dosen di UMKT hanya dapat menjelaskan materi melalui media *E-Learning* sehingga dosen tidak mengetahui apakah mahasiswa memahami dan mengerti dengan materi yang telah dijelaskan. Sehingga perlukan metode tentang prediksi Kinerja Mahasiswa dalam perkuliahan daring untuk membantu dosen dan kampus mengetahui

efektivitas pembelajaran daring ini berjalan dengan baik atau tidak. Berbagai penelitian yang telah dilakukan mengenai kinerja mahasiswa dalam perkuliahan daring diantaranya, Penelitian yang dilakukan oleh Dewantara & Nurgiansah [1] tentang “Efektivitas Pembelajaran Daring di Masa Pandemi COVID 19 Bagi Mahasiswa Universitas PGRI Yogyakarta”. Responden dari penelitian ini berjumlah 1.000 mahasiswa. Hasil penelitian menunjukkan bahwa 79% mahasiswa menginginkan pembelajaran secara tatap muka, sedangkan hanya 1% saja mahasiswa yang menginginkan pembelajaran daring, sehingga dapat disimpulkan bahwa pembelajaran daring secara terus menerus selama masa pandemi ini sangat tidak efektif. Kemudian oleh Djusar et al. [2] tentang “Analisa Efektifitas Pembelajaran Daring Saat Pandemi Covid-19 Pada Fakultas Ilmu Komputer Universitas Lancang Kuning”. Objek penelitian ini adalah 63 orang mahasiswa angkatan 2018 dan 2019 yang aktif dan mengikuti pembelajaran 2 secara daring menggunakan kuisioner dengan 26 pertanyaan. Dari 6 indikator efektivitas yang di ukur menggunakan rumus persentase, maka didapatkan hasil untuk semua indikator dikatakan efektif. Batas efektif yang ditentukan adalah sebesar 50%. Hasil dari pengukuran indikator, semuanya memiliki nilai lebih dari 50%. Berdasarkan hasil pengukuran, maka dinyatakan bahwa pembelajaran daring saat pandemi covid-19 pada Fakultas Ilmu Komputer Universitas Lancang Kuning dikatakan efektif dengan rata-rata hasil pengukuran di atas 50%.

Beberapa penelitian sebelumnya terkait prediksi dan algoritma *Logistic Regression* sudah

pernah dilakukan diantaranya, Prasetyo et al., [6] tentang “Komparasi Algoritma *Logistic Regression* dan *Random Forest* Pada Prediksi *Cacat Software*”. Hasil akurasi yang didapatkan Algoritma *Random Forest* menghasilkan akurasi dengan presentase 95.582% sedangkan algoritma *Logistic Regression* menunjukkan akurasi 91,4%.

Peneliti Handayani [4] tentang komparasi *Support Vector Machine*, *Logistic Regression*, dan *Artificial Neural Network* dalam prediksi penyakit jantung. Dari hasil penelitian didapatkan hasil akurasi tertinggi pada metode algoritma *Logistic Regression* sebesar 86% menggunakan pembagian data 80:20.

Peneliti Goldblatt [3] tentang analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. Algoritma *Logistic Regression* menunjukkan hasil bahwa algoritma tersebut paling dominan terhadap algoritma yang lain meskipun nilai *accuracy* tidak menunjukkan hasil yang tertinggi yaitu 81,64% dan nilai AUC 0,933.

Peneliti Santosa & Artanto [9] tentang Prediksi loyalitas pelanggan telekomunikasi menggunakan *logistic regression* dengan seleksi fitur *particle swarm optimization*. Dari hasil penelitian yang di peroleh delapan model regresi logistik, yaitu model A-H. Nilai akurasi tertinggi untuk jalur prestasi ialah model F (dataset 2008-2013) dengan akurasi prediksi 73,73%. Hasil akurasi tertinggi untuk jalur non prestasi ialah dengan model D (dataset 2008- 2011) dan E (dataset 2008-2012) dengan akurasi prediksi 56,76%.

Peneliti Putra & Azhar [7] tentang Perbandingan Model *Logistic Regression* dan *Artificial Neural Network* pada Prediksi Pembatalan Hotel. Dengan nilai *accuracy* sebesar 79.77%, nilai *precision* 85.86% dan nilai *recall* 55.07%.

Peneliti Gunawan et al., [6] tentang Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode *Grid Search* pada Algoritma *Logistic Regression*. Model algoritma yang di gunakan ialah *Logistic Regression*. Diperoleh model *Logistic Regression* dengan *Grid Search* pada *Classification Report* memiliki rata-rata akurasi model sekitar 79% dan akurasi data check sebesar 83,33%.

Peneliti Rini [8] tentang Perbandingan Algoritma *Random Forest Classifier*, *Support Vector Machine* dan *Logistic Regression Classifier* Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News). Nilai performa *Sensitivity* terbaik dari semua skenario percobaan *Logistic Regression Classifier* dengan persentase 99,96 %.

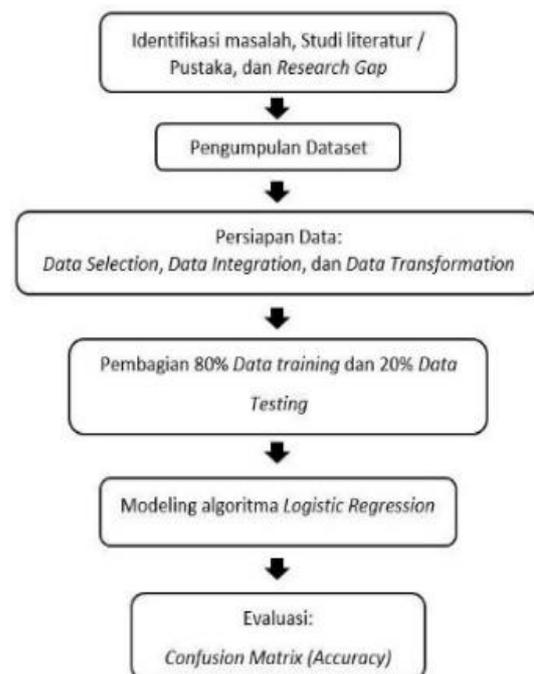
Peneliti Vinarti & Anggraeni [10] tentang Identifikasi Faktor Prediksi Diagnosis Tingkat Keganasan Kanker Payudara Metode *Stepwise Binary Logistic Regression*. akurasi tertinggi didapatkan oleh penggalan data campuran sebesar 98.4%.

Peneliti Kinoto et al., [5] tentang Prediksi Employee Churn Dengan Uplift Modeling Menggunakan Algoritma *Logistic Regression*. Hasil prediksi pada penelitian ini mendapat akurasi sebesar 64,40%.

Berdasarkan permasalahan dan beberapa penelitian yang diuraikan di atas, maka dalam penelitian ini, *Logistic Regression* di gunakan untuk meneliti Prediksi kinerja mahasiswa dalam perkuliahan daring berbasis *e-learning*.

## 2. METODOLOGI PENELITIAN

Penelitian ini mempunyai alur seperti pada gambar 1.



Gambar 1. Alur Penelitian

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Penelitian ini menggunakan data pada mata kuliah Kewarganegaraan seluruh program studi yang diperoleh melalui platform OpenLearning dan data nilai akhir yang diperoleh dari Bagian Akademik Administrasi (BAA). Data yang digunakan mulai dari tahun 2020/2021 dan 2021/2022 semester ganjil

dengan jumlah data sebanyak 2.663. Data nilai akhir yang diperoleh terdiri dari 5 atribut diantaranya nomor induk mahasiswa (NIM), nama, nilai akhir, bobot dan simbol. Sedangkan data OpenLearning terdiri dari 17 atribut diantaranya yaitu *Profile Name, Learner Name, Learner Email, Enrolment ID, Institution Membership Id, Enrolment Date, Completion Date, Time Spent On Course, Progress, %Course Completed, Certificate, Comments, Kudos, Enrolment Cost, Tugas, Uts Dan Quiz*.

### 3.2 Data Preparation

#### 3.2.1 Data Selection

Tahap ini dilakukan setelah data telah terkumpul, kemudian akan dipilih atribut yang akan digunakan lalu diseleksi. Hal ini dilakukan agar bisa meningkatkan akurasi dan efisiensi algoritma yang akan diterapkan. Atribut pada data BAA UMKT yang tidak diperlukan atau dihapus adalah nama dan bobot, sedangkan atribut yang tidak diperlukan atau dihapus pada data OpenLearning adalah *Profile name, Learner name, Learner email, Enrolment ID, Institution Membership ID, Enrolment date, Completion date, Progres*, dan *Certificate ID*. Kemudian melihat korelasi antara atribut (nim, *Time Spent On Course, Course Completed, Comments, Kudos, Program Studi, Tugas, Uts, Quiz*, dan Nilai Akhir) terhadap kelas target / target class (simbol).



Gambar 2. Atribut Korelasi

Hasil pencarian korelasi antara atribut terhadap kelas target dapat dilihat pada gambar 2. Hasil pencarian korelasi menunjukkan bahwa Atribut Nim, *Comments*, dan *Kudos* memiliki nilai korelasi yang sangat rendah terhadap kelas target. Atribut yang dipilih adalah *Time Spent On Course = 0.15, Course Completed = 0.32, Tugas = 0.35, Uts = 0.26, Quiz = 0.22*.

#### 3.2.2 Data Integration

Tahapan ini dilakukan untuk menggabungkan data pada nilai akhir dan data OpenLearning. Atribut *Time Spent On Course, Course Completed*, tugas, UTS, quiz dan simbol merupakan data yang digabungkan.

#### 3.2.3 Data Transformation

Tahap ini dilakukan untuk mengubah isi data atau tipe data. Data yang akan di transformasi yaitu data OpenLearning pada atribut *time spent on course* yang merupakan lama waktu mahasiswa menyelesaikan kursus. Hasil transformasi dapat dilihat pada tabel 1.

**Tabel 1. Hasil Transformasi Atribut Time Spent On Course**

<i>Time Spent On Course</i> (Sebelum Di Transformasi)	<i>Time Spent on Course</i> (Setelah Di Transformasi)
11 Hrs 27 Mins	687
6 Hrs 52 Mins	412
6 Hrs 31 Mins	391
:	:
2 Hrs 39 Mins	155

Kemudian untuk data pada atribut simbol merupakan atribut yang akan digunakan sebagai label target pada pemodelan. Hasil transformasi dapat dilihat pada tabel 2.

**Tabel 2 Hasil Transformasi Atribut Simbol**

<i>Simbol</i>	<i>Simbol(Kategorik)</i>
<b>A</b>	BAIK
<b>AB</b>	BAIK
<b>B</b>	BAIK
<b>BC</b>	BURUK
<b>C</b>	BURUK
<b>D</b>	BURUK
<b>E</b>	BURUK

**3.2.4 Data Reduction dan Data Cleaning**

Tahapan ini dilakukan untuk menghindari dataset yang tidak seimbang dengan mengurangi

data mayoritas. Data mayoritas merupakan data kelas BAIK yang berjumlah 2536 sehingga dikurangi sejumlah data minoritas BURUK yang berjumlah 127 data. Jadi total data kelas baik berjumlah 127 yang diambil secara random. Setelah itu dilakukan tahapan *Data Cleaning* terhadap data yang tidak konsisten. Oleh karena itu pada tahapan *Reduction* ini dilakukan proses pembersihan data yang tidak konsisten. Sehingga jumlah data yang digunakan pada penelitian ini adalah 178 data.

**3.3 Modelling**

**3.3.1 Dataset**

Data yang digunakan berjumlah 178 data yang kemudian dilakukan pembagian data uji 80:20. Setelah itu diperoleh hasil data training berjumlah 142 data dan data testing sebanyak 36 data.

**Tabel 3. Dataset**

<i>No</i>	<i>Time Spent on Course</i>	<i>Course Completed</i>	<i>Tugas</i>	<i>Quiz</i>	<i>UTS</i>	<i>Simbol</i>
1.	0	0	0	0	0	BURUK
2.	0	0	0	0	0	BURUK
3.	0	0	0	0	0	BURUK
4.	2	0	2	38	0	BURUK
5.	4	5	0	0	0	BURUK
:	:	:	:	:	:	:
178.	304	81	80.2	70	60	BAIK

**Tabel 4. Data Training**

<i>No</i>	<i>Time Spenton Course</i>	<i>Course Completed</i>	<i>Tugas</i>	<i>UTS</i>	<i>Quiz</i>	<i>Simbol</i>
1.	443	82	82,6	52	66	BAIK
2.	1347	82	86,3	86	80	BAIK
3.	880	98	76,7	86	71	BAIK
4.	280	99	85	86	63	BAIK
5.	73	77	0	0	0	BURUK
:	:	:	:	:	:	:
142.	328	81	80	66	81	BAIK

**Tabel 5. Data Testing**

No	TimeSpenton Course	Course Completed	Tugas	UTS	Quiz	Simbol
1.	143	89	19	0	0	BURUK
2.	561	97	75,5	78	66	BAIK
3.	187	85	41,9	22	0	BURUK
4.	151	86	22,2	62	53	BURUK
5.	799	97	35	86	60	BAIK
:	:	:	:	:	:	:
36.	120	83	27	14	0	BURUK

**3.3.2 Logistic Regression.**

Setelah data training dan data testing terbagi, kemudian dilanjutkan dengan mencari nilai konstanta dari *Time Spent on Course*, *Course Completed*, UTS, Tugas, Quiz menggunakan perhitungan dengan SPSS. Didapatkan nilai konstanta dari *Time Spent on Course* sebesar 0.005, *Course Completed* sebesar -0.085, UTS sebesar -0.0118, Tugas sebesar -0.148, Quiz sebesar -0.092 dan untuk konsanta utamanya sebesar 21.752.

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for Exp(B)	
							Lower	Upper
Step 1 <sup>a</sup>								
time_spent_on_course	.005	.005	1.176	1	.278	1.005	.996	1.014
course_completed	-.085	.105	.654	1	.419	.919	.749	1.128
tugas	-.148	.063	5.476	1	.019	.862	.761	.976
uts	-.118	.061	3.713	1	.054	.888	.788	1.002
quiz	-.092	.042	4.851	1	.028	.913	.841	.990
Constant	21.752	11.881	3.346	1	.067	2796330435		

a. Variable(s) entered on step 1: time\_spent\_on\_course, course\_completed, tugas, uts, quiz.

**Gambar 3. Nilai Konstanta dari SPSS**

Selanjutnya untuk pembuatan model *Logistic Regression* yang digunakan untuk

mencari nilai probabilitas dengan rumus sebagai berikut:

$$P = \frac{e^{\beta_0 + \beta_1 X + \epsilon_i}}{1 + e^{\beta_0 + \beta_1 X + \epsilon_i}} \tag{1}$$

1. Diprediksi BURUK karena nilai probabilitas yang diperoleh adalah 0.43 karena probabilitasnya lebih dari  $\leq 0.5$ .
2. Diprediksi BAIK karena nilai probabilitas yang diperoleh adalah 0.62 karena nilai probabilitasnya lebih dari  $\geq 0.5$ .

Estimasi probabilitas data testing 1diperoleh 0.43. Nilai estimasi probabilitas sebesar 0,43 ( $\leq 0,5$ ), maka diprediksi masuk ke dalam kelompok yang dinyatakan BURUK. Kemudian untuk data testing 2 diperoleh nilai estimasi probabilitas sebesar 0.62 ( $\geq 0.5$ ), maka diprediksi masuk kedalam kelompok yang dinyatakan BAIK.

$$P = \frac{e^{21.752+0.005(383)+-0.085(79)+-0.148(23.7)+-0.118(54)+-0.092(74)}}{1+e^{21.752+0.005(383)+-0.085(79)+-0.148(23.7)+-0.118(54)+-0.092(74)}} = 0.43 \tag{2}$$

$$P = \frac{e^{21.752+0.005(383)+-0.085(79)+-0.148(23.7)+-0.118(54)+-0.092(74)}}{1+e^{21.752+0.005(383)+-0.085(79)+-0.148(23.7)+-0.118(54)+-0.092(74)}} = 0.6 \tag{3}$$

**Tabel 6. Hasil Perhitungan Probabilitas**

<i>Time Spent On Course</i>	<i>Course Completed</i>	<i>Tugas uts</i>	<i>Quiz</i>	<i>Simbol</i>	<i>Predict</i>
143	89	19	0	BURUK	BURUK
561	97	75,5	78	BAIK	BURUK
187	85	41,9	22	BURUK	BURUK
151	86	22,2	62	BURUK	BAIK
799	97	35	86	BAIK	BAIK

Dengan menghitung probabilitas diperoleh hasil klasifikasi untuk metode *Logistic Regression* dapat dilihat pada tabel 6.

**Tabel 7. Hasil Perhitungan Probabilitas(Lanjutan)**

:	:	:	:	:	:	:
383	79	23.7	52	74	BAIK	BAIK

### 3.3.3 Evaluasi Model

Dari pengujian data diperoleh hasil evaluasi dan hasil *accuracy algoritma Logistic Regression* pada pembagian data 80:20. Dimana pembagian data ini memiliki *true positive* 15 data, *true negative* 18 data, *false positive* 1 data, dan *false negative* 2 data. Dapat dilihat pada tabel 7.

**Tabel 7. Confusion Matrix**

<i>Actual</i>	<i>Predict</i>	
	Lulus (1)	Tidak Lulus (0)
Lulus (1)	15	1
Tidak Lulus (0)	2	18

Kemudian dilakukan perhitungan dari hasil pembagian data dan diperoleh hasil perhitungan *Confusion Matrix* dengan nilai akurasi sebesar 91%.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$= \frac{15+18}{15+18+1+2} \times 100\%$$

$$= \frac{33}{36} = 0.91$$

$$= 0.91 \times 100\%$$

$$= 91\%$$

(5)

*Prediksi Kinerja Mahasiswa Dalam Perkuliahan Daring berbasis E learning ... ..*  
(Taghfirul Azhima Yoga Siswa, Agustya Nanda Pratiwi)

## 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Prediksi kinerja mahasiswa dalam perkuliahan daring berbasis e-learning dapat dilakukan dengan *atribut time spent on course, course completed, tugas, quiz, dan uts.*
2. Implementasi algoritma *logistic regression* telah diterapkan pada pemodelan prediksi kinerja mahasiswa dalam perkuliahan daring berbasis *e-learning* diperoleh ketepatan model sebesar 91.66%.

## DAFTAR PUSTAKA

- [1] Dewantara, J. A., & Nurgiansah, T. H. (2021). Efektivitas Pembelajaran Daring di Masa Pandemi COVID 19 Bagi Mahasiswa Universitas PGRI Yogyakarta. *Jurnal Basicedu*, 5(1), 367–375.
- [2] Djusar, S., Sadar, M., & Asril, E. (2021). Analisa Efektifitas Pembelajaran Daring Saat Pandemi Covid-19 Pada Fakultas Ilmu Komputer Universitas Lancang Kuning. *Jurnal Sistem Informasi Dan Manajemen (JURSIMA)*, 9(2), 12–20.
- [3] Goldblatt, D. (1994). Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non Aktif. *Seminars in Neurology*, 14(1), 241–249.
- [4] Handayani, F. (2021). Komparasi Support

- Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(3), 329.
- [5] Kinoto, J., Damanik, J. L., Situmorang, E. T. S., Siregar, J., & Harahap, M. (2020). Prediksi Employee Churn Dengan Uplift Modeling Menggunakan Algoritma Logistic Regression. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 3(2), 503-508.
- [6] Prasetyo, R., Nawawi, I., Fauzi, A., Nusa Mandiri, U. (2021). Komparasi Algoritma Logistic Regression dan Random Forest pada Prediksi Cacat Software. *Jurnal Teknik Informatika Unika St. Thomas (JTIUST)*, 06, 2657–1501.
- [7] Putra, M. S. T., & Azhar, Y. (2021). Perbandingan Model Logistic Regression dan Artificial Neural Network pada Prediksi Pembatalan Hotel. *Jurnal Informatika Sunan Kalijaga (JISKA)*, 6(1), 29–37.
- [8] Rini, D. P. (2021). Perbandingan Algoritma Random Forest Classifier, Support Vector Machine dan Logistic Regression Classifier Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News). *Jurnal Media Informatika Budidarma*, 5, 1720–1728.
- [9] Santosa, S., & Artanto, F. A. (2015). Prediksi Loyalitas Pelanggan Telekomunikasi Menggunakan Logistic Regression Dengan Seleksi Fitur Particle Swarm Optimization. *Jurnal Teknologi Informasi*, 11(April), 90–99.
- [10] Vinarti, R. A., & Anggraeni, W. (2014). Identifikasi Faktor Prediksi Diagnosis Tingkat Keganasan Kanker Payudara Metode Stepwise Binary Logistic Regression. *Jurnal Informatika*, 12(2), 70-76.