

LAPORAN PEMBELAJARAN MESIN



IMPLEMENTASI ALGORITMA NAIVE BAYES/LOGISTIC REGRESSION/RANDOM FOREST UNTUK MONKEYPOX

ANGGOTA

Alif Rizky Riyaninchola / 17220499

Hanif Fachrizal / 17220945

Muhamad Alvin Tribuana / 17221172

Muhammad Sandi Rahman / 17221047

**FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS BINA SARANA INFORMATIKA**

2024

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Hasil dari penelitian ini berupa metode klasifikasi *MonkeyPox* yang memiliki akurasi tinggi diantara algoritma Logistic Regression, RandomForest, dan *Naive Bayes* dengan K-Nime untuk mendiagnosa penyakit MonkeyPox.

2. Pembahasan

A. Analisa Data

Dataset ini berisi data yang dikumpulkan dari para pasien. Setelah dilakukan teknik *Table View*, maka *dataset* berisi 15909 pasien penderita MonkeyPox sedangkan 9091 pasien bukan penderita MonkeyPox. *Dataset* ini memiliki 12 *atribut* dimana 11 *atribut* merupakan *atribut input* sedangkan 1 *atribut* sebagai *output* atau *class*, dan memiliki 25000 *record*. Adapun deskripsi *atribut* seperti ditunjukkan pada tabel 1. berikut:

Tabel 1.
Deskripsi *Atribut*

Atribut	Keterangan
Patient_ID	Identifikasi Pasien
Systematic Illness	Penyakit Sistematis
Rectal Pain	Rasa sakit yang terjadi di rectum
Sore Throat	Tenggorokan merasakan sakit
Penile Oedema	Penilaian Edema
Oral Lesions	Istilah medis pada segala jenis luka, borok, perubahan jaringan dalam mulut.
Solitary Lesion	Area abnormal dalam tubuh
Swollen Tonsils	Pembengkakan jaringan lunak dibelakang tenggorokan
HIV Infec	Infeksi saluran reproduksi/ alat kelamin.
Sexually	Infeksi yang menular melalui aktivitas

Transmitted Infect	seksual.
MonkeyPox	Penyakit yang disebabkan oleh MonkeyPox yang gejalanya mirip dengan cacar.

B. Pengujian Model

Penelitian ini dilakukan dengan eksperimen pengujian pada model yang diusulkan. Kemudian dilakukan evaluasi dan validasi model untuk menghasilkan nilai *accuracy* dan AUC. Pengujian menggunakan *KNime* dengan operator ntuk mendapatkan hasil *accuracy* dan AUC pada setiap algoritma yang diuji. Evaluasi yang dilakukan adalah dengan *Confusion Matrix* dan *ROC Curve* atau *Area Under Curve* (AUC).

1. *Confusion Matrix*

a. Algoritma Naive Bayes

Tabel 2. adalah *confusion matrix* untuk algoritma Naive Bayes. Diketahui 4206 data diklasifikasi “Yes” diprediksi sesuai dengan data sebenarnya, lalu 1694 data diprediksi “No” tetapi ternyata “Yes”. Kemudian 1024 data diklasifikasi “No” diprediksi sesuai, dan 576 data diprediksi “Yes” ternyata “No”.

Tabel 2.
Model *Confusion Matrix* untuk Algoritma Naive Bayes

Accuracy: 69,73%			
	true Yes	true No	class precision
pred. Yes	4206	576	71.29%
pred. No	1694	1024	64.00%
class recall	87.95%	37.67%	

b. *Logistic Regression*

Tabel 3. adalah *confusion matrix* untuk algoritma *Logistic Regression*. Diketahui 5747 data diklasifikasi “Yes” tepat sesuai dengan data sebenarnya, lalu 2644 data diprediksi “No” tetapi ternyata “Yes”. Kemudian 913 data diklasifikasi “No” diprediksi sesuai, dan 696 data diprediksi “Yes” ternyata “No”.

Tabel 3.
Model *Confusion Matrix* untuk Metode *Logistic Regression*

Accuracy: 66.60%			
	true Yes	true No	class precision
pred. Yes	5747	696	69.49 %
pred. No	2644	913	56.74%
class recall	89.20%	25.67%	

c. Algoritma Random Forest

Tabel 4 adalah *confusion matrix* untuk algoritma Algoritma Random Forest. Diketahui 4488 data diklasifikasi “Yes” diprediksi sesuai dengan data sebenarnya, lalu 2208 data diprediksi “No” tetapi ternyata “Yes”. Kemudian 494 data diklasifikasi “No” diprediksi sesuai, dan 310 data diprediksi “Yes” ternyata “No”.

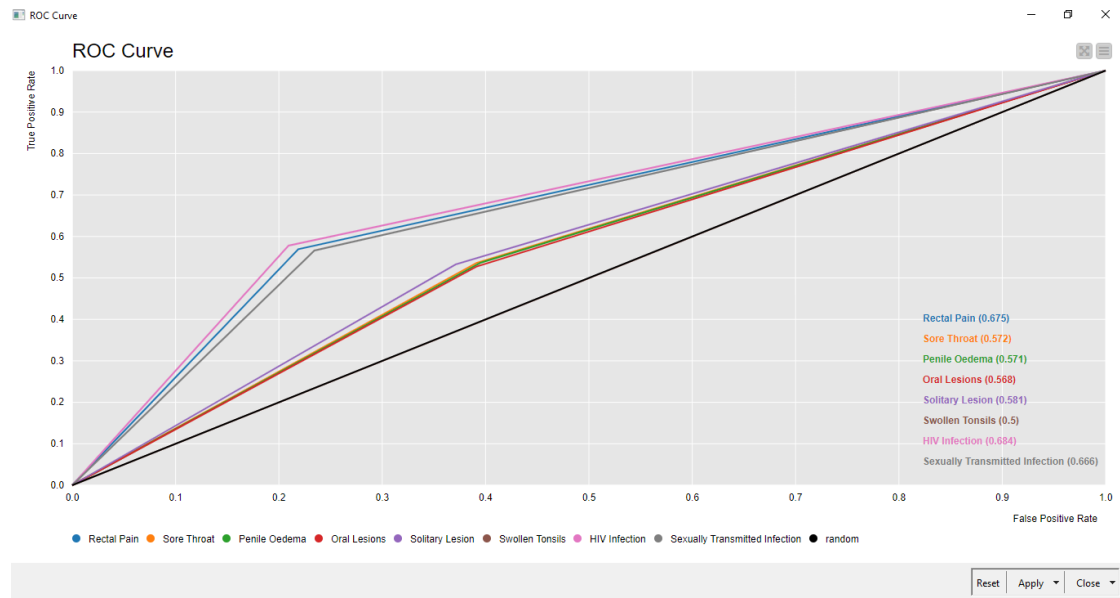
Tabel 4.
Model *Confusion Matrix* untuk Algoritma Random Forest

Accuracy: 66.43%			
	true Yes	true No	class precision
pred. Yes	4488	310	67.03%
pred. No	2208	494	61.44 %
class recall	93.54%	18.28%	

2 Kurva ROC

a. Algoritma Naïve Bayes

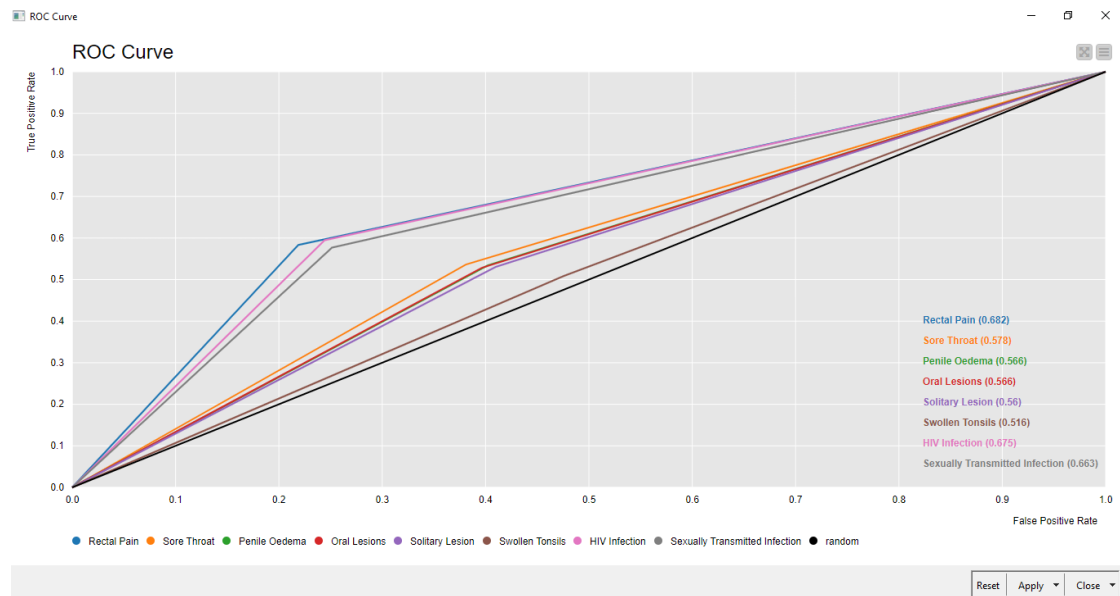
Kurva ROC untuk algoritma Naïve Bayes seperti ditunjukkan oleh gambar 1 di bawah ini.



Gambar 1. Kurva ROC dengan algoritma Naïve Bayes

b. Logistic Regression

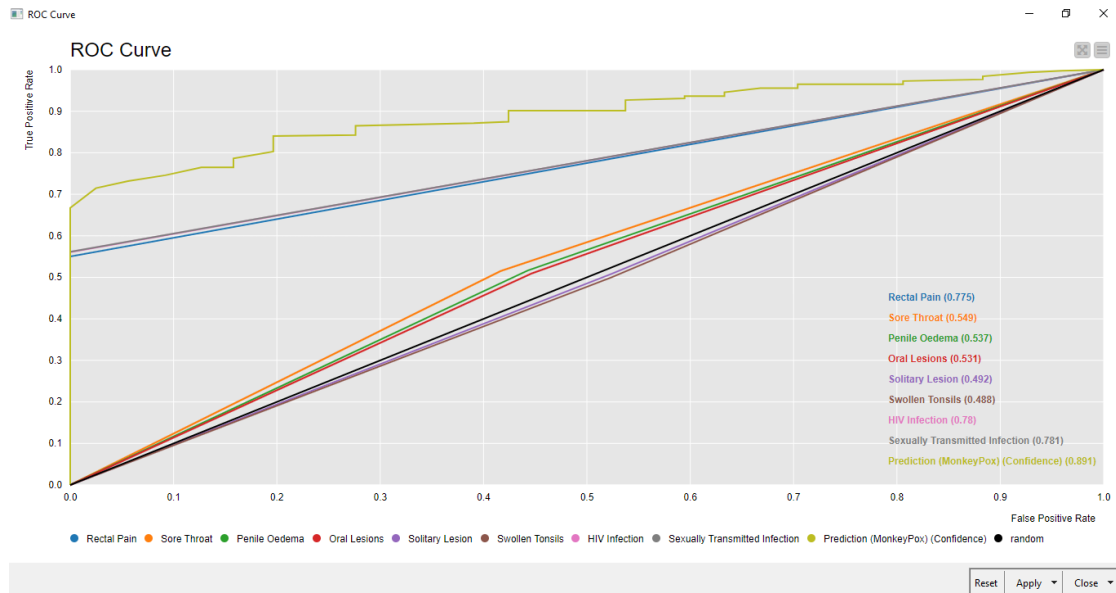
Kurva ROC untuk algoritma Logistic Regression seperti ditunjukkan oleh gambar 2 di bawah ini.



Gambar 2. Kurva ROC dengan Logistic Regression

c. Algoritma Random Forest

Kurva ROC untuk algoritma Random Forest seperti ditunjukkan oleh gambar 3 di bawah ini.



Gambar 3. Kurva ROC Metode Random Forest

Tabel 6.

Komparasi Nilai *Accuracy* dan AUC

	Naïve Bayes	Logistic Regression	Random Forest
<i>Accuracy</i>	69.73%	66.60 %	66.43%
AUC			0.891

Tabel 6 membandingkan *accuracy* dan AUC dari tiap algoritma. Terlihat bahwa nilai *accuracy* algoritma Naïve Bayes paling tinggi. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok, yaitu:

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokkan di atas dan Tabel 6 maka dapat disimpulkan bahwa model algoritma Random Forest termasuk kategori klasifikasi baik.