

University of Westminster

# Project Report Monkey Pox Testing Analysis

Data Mining and Machine Learning (7BUIS008W)

Coursework 1

B Gayan Kalinga w2027429 20231057 [gayan.20231057@iit.ac.lk](mailto:gayan.20231057@iit.ac.lk)

**Table of Contents**

[Project Report Monkey Pox Testing Analysis 1](#_Toc163416563)

[Task 01: Domain Understanding Classification 1](#_Toc163416564)

[Task 02: Producing Your Experimental Design 1](#_Toc163416565)

[Task 03: Cleaning and Transforming the Data 2](#_Toc163416566)

[Task 04: Create Predictive Classification Models 3](#_Toc163416567)

[Task 05: How Good Is the Model 4](#_Toc163416568)

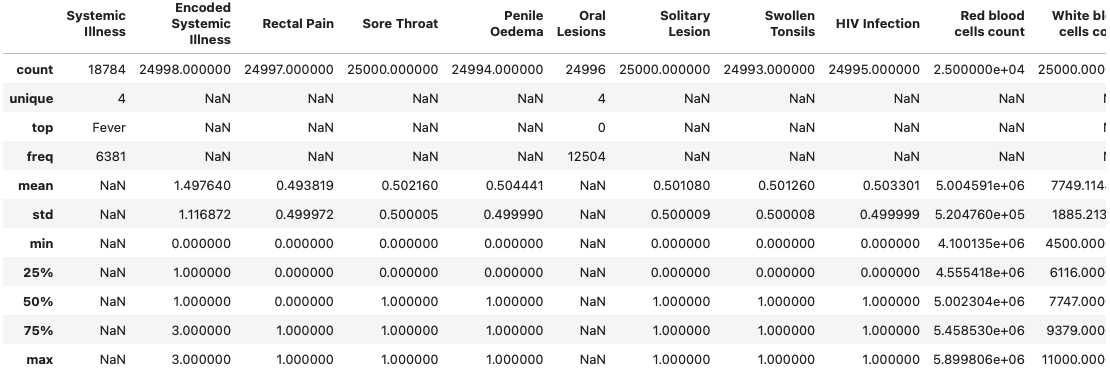
# Task 01: Domain Understanding Classification

Mark the variables logically applicable in the classification modeling of MPOX

|  |  |
| --- | --- |
| Attribute Name | Retain or drop |
| Test ID | Drop |
| Systemic Illness | Retain |
| Sore Throat | Retain |
| Rectal Pain | Retain |
| Penile Oedema | Retain |
| Oral Lesions | Retain |
| Solitary Lesion | Retain |
| Swollen Tonsils | Retain |
| HIV Infection | Retain |
| Red blood cells | Retain |
| White blood cells | Retain |
| Home Ownership | Drop |
| Age | Retain |
| Month of Birth | Drop |
| Health Insurance | Retain |
| Sexually Transmitted Infection | Retain |
| MPOX | Retain |

# Task 02: Producing Your Experimental Design

Basic Statistical Descriptions



Measurement Scale Types

A screenshot of a computer program

Description automatically generated

Distribution of the Result

A bar graph with blue and orange squares

Description automatically generated

# Task 03: Cleaning and Transforming the Data

A: Issues identified in the retained data set and possible variables

|  |  |  |
| --- | --- | --- |
| Dataset or Variable Issue | Name of the Variable | Issue Description |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

B: Solution to mitigate issues found with justification for using that solution

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset or Variable Issue | Name of the Variable | The Issue | Solution | Justification |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

C: Outputs (Before & After)

# Task 04: Create Predictive Classification Models

A: Classification algorithm details

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm Name | Type of Algorithm | Learnable Parameters | Possible Hyper – Parameters | Python package source code to call the algorithm |
| LR |  |  |  |  |
| DT |  |  |  |  |
| KNN |  |  |  |  |
| SVM (RBF) |  |  |  |  |
| NB |  |  |  |  |

B: Data shape function output

Justify the training-test split ratio and provide an in-text reference.

Code line from the source code.

# Task 05: How Good Is the Model

A: test confusion matrix for each trained model (output screenshots)

B: Five different classification evaluation metrics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric Name | “USE” or “DO NOT USE” | Justification in relation to the success criteria | Model Name | Metric Score |
| Accuracy |  |  | LR |  |
| DT |  |
| KNN |  |
| SVM (RBF) |  |
| NB |  |
| Recall |  |  | LR |  |
| DT |  |
| KNN |  |
| SVM (RBF) |  |
| NB |  |
| Precision |  |  | LR |  |
| DT |  |
| KNN |  |
| SVM (RBF) |  |
| NB |  |
| F-Measure |  |  | LR |  |
| DT |  |
| KNN |  |
| SVM (RBF) |  |
| NB |  |
| AUC-ROC |  |  | LR |  |
| DT |  |
| KNN |  |
| SVM (RBF) |  |
| NB |  |

C: Best classification model/s based on “USED” performance metrics score and describe how the model satisfies the needs of health care professionals

D:

E:

F: