Project Description Document

1- Introduction

This project explores two types of datasets, image-based and numeric, to train machine learning models for specific predictive tasks. The goal is to compare the performance of different models in terms of accuracy, precision, recall, and AUC. Key visualizations like loss curves, ROC curves, and confusion matrices are presented to evaluate the models.

2- Datasets

2.1- Image Dataset

- **Dataset Name :** `Cell Images For Detecting Malaria`
- **Number of Classes :** `2 (Parasitized, Uninfected)`
- **Total Number of Samples**: 10,000
- **Image Size**: 64x64 pixels
- **Data Split**:
    - Training: 8000 samples
    - Validation: 1000 samples
    - Testing: 1000 samples

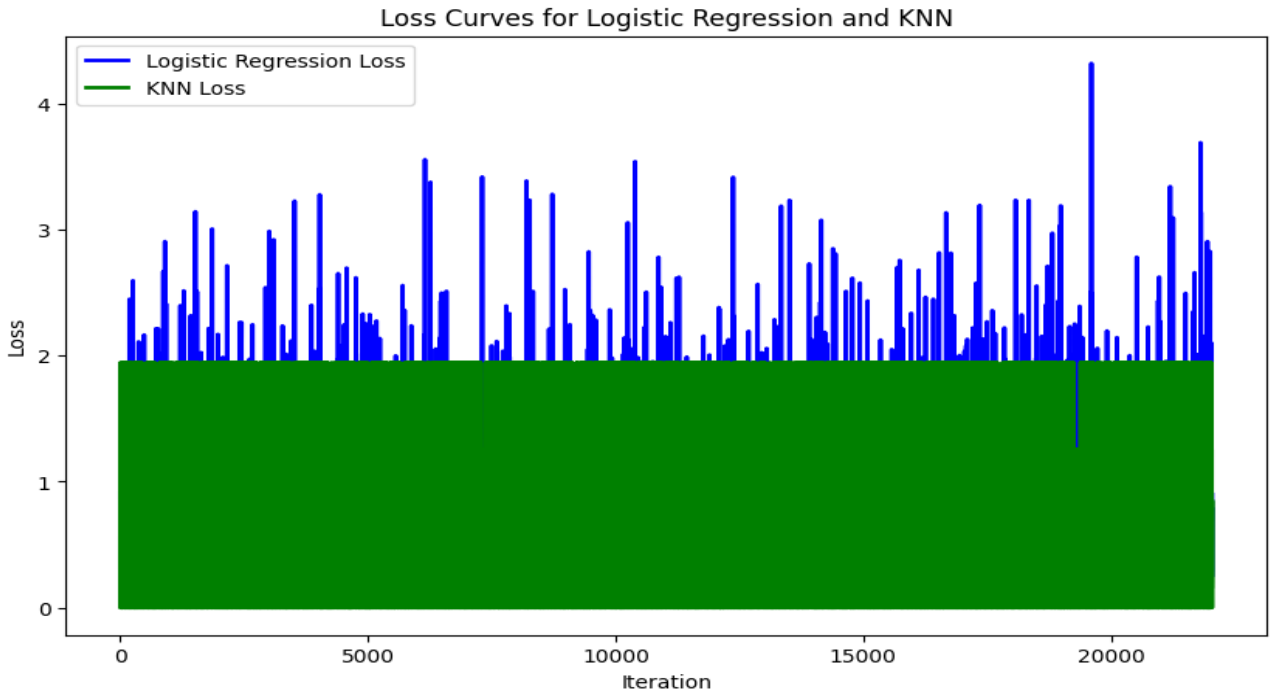
2.2- Numeric Dataset

- **Dataset Name**: `Used Cars For Sale in Egypt`
- **Number of Classes :** `0 (No Classes Specified)`
- **Total Number of Samples**: 500
- **Data Split**:
    - Training: 350 samples
    - Validation: 75 samples
    - Testing: 75 samples

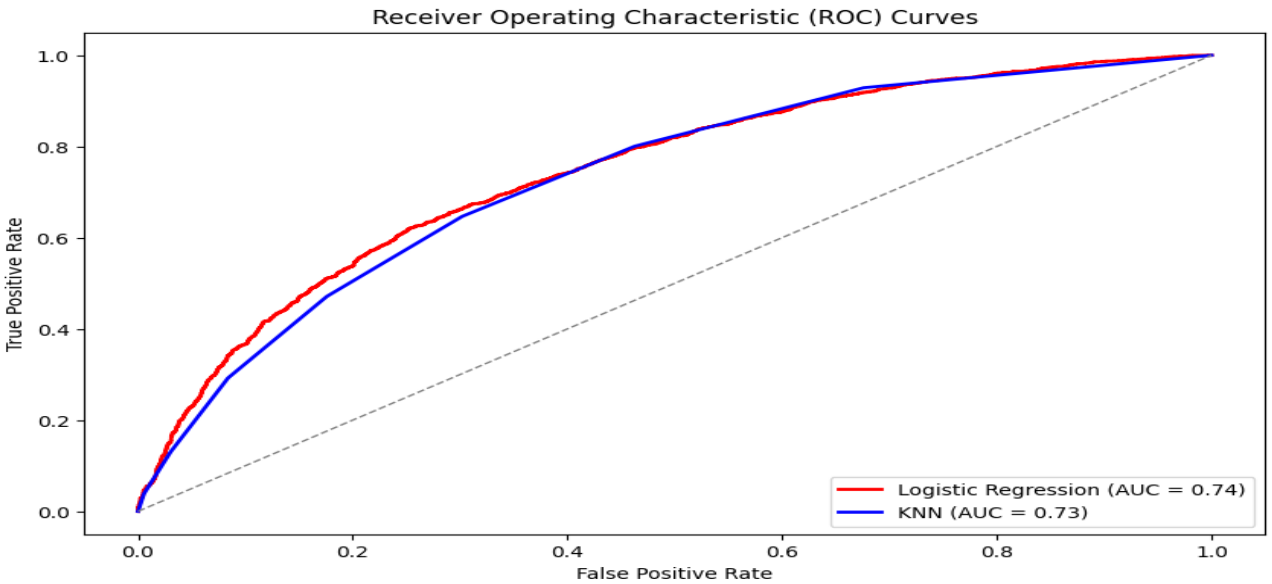
3- Models

3.1- Image Dataset Models

- **Loss Curve:**

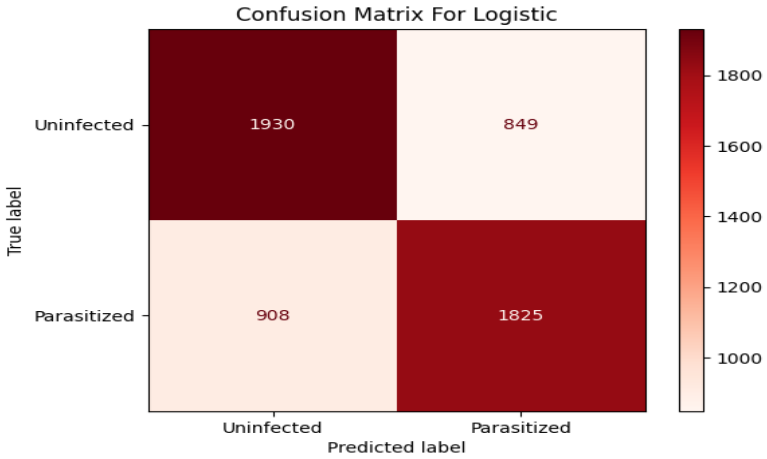


- **ROC Curve and AUC Plot:**



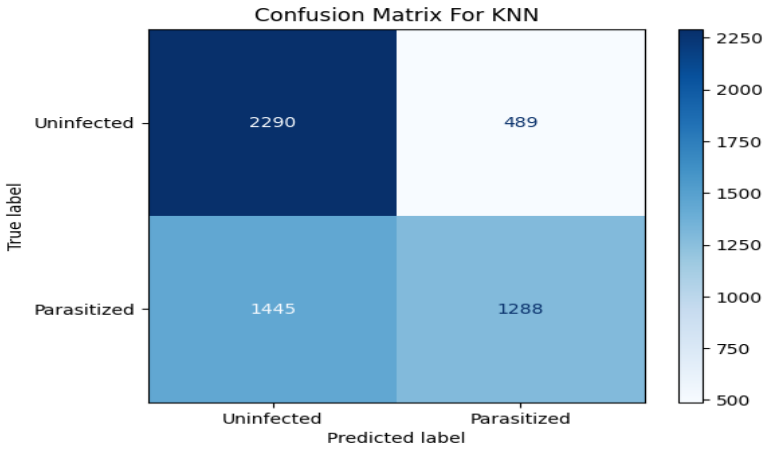
3.1.1- Logistic Regression Model

- **Accuracy**: 0.6812409288824384
- **Precision**: 0.6824981301421092
- **Recall**: 0.667764361507501
- **Confusion Matrix**: [[1930 849] [908 1825]]
- **Confusion Matrix Display**:



3.1.2- KNN Model

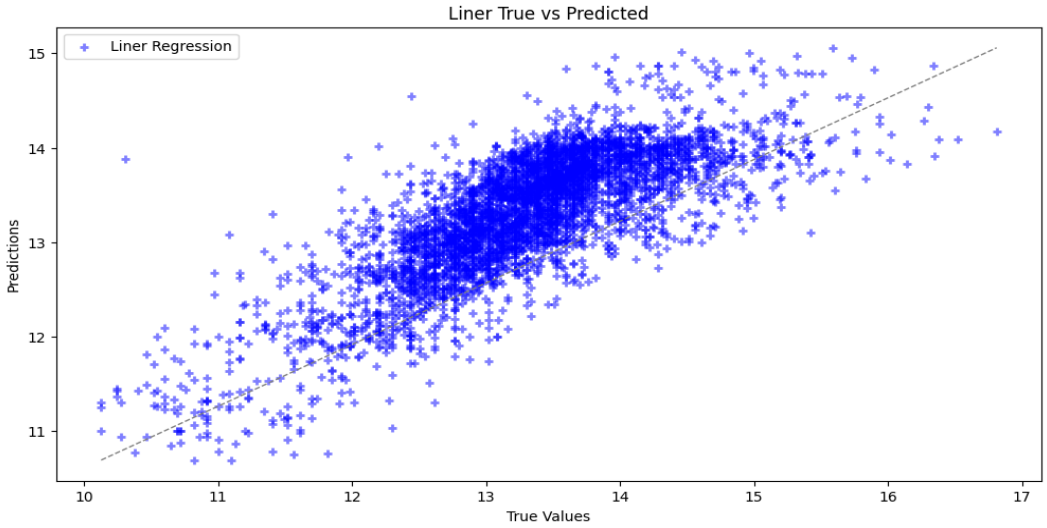
- **Accuracy**: 0.6491291727140783
- **Precision**: 0.7248171074845244
- **Recall**: 0.4712769849981705
- **Confusion Matrix**: [[2290 489] [1445 1288]]
- **Confusion Matrix Display**:



3.2- Numeric Dataset Models

3.2.1- Linear Regression Model

- **MAE**: 0.40426621312152944
- **R2 Score**: 0.6117641574183054
- **Scatter Plot:**



3.2.2- KNN Model

- **MAE**: 0.18835639313724023
- **R2 Score**: 0.8788546268754598
- **Scatter Plot:**