



King Fahd University of Petroleum and Minerals

ICS 485 - Machine Learning

Term: 201

Sec 01

Term Project

Utilizing Machine Learning Algorithms to Produce Face Mask Detection Models

For

Dr. Andri Mirzal

Team: IDDQD-Lite	
Ahmad Alsaqqa	201541790
Mohammed Alghamdi	201670960

1 CONTENTS

2	Introduction.....	3
3	Datasets.....	3
4	Methods	4
5	Simulation Results.....	6
6	References.....	8

2 INTRODUCTION

Wearing a face mask helps in preventing the spread of COVID-19. People not taking precautions causes major damage to themselves, other people around them, and eventually their country's economy, having a program that can detect whether people are wearing a face mask or not could save people's lives. This influenced us to utilize the opportunity of the machine learning term project to experiment and develop models that could classify masked faces and unmasked faces. Our goal is to explore different machine learning algorithms, using different datasets, to train a model that shall be used to predict masked faces from non-masked faces.

3 DATASETS

We have collected our datasets from different sources. We give large credit for, and not limited to, Kaggle.com for hosting such a wide array of useful datasets. For this project, we have used 5 datasets to diversify our sources of information and develop the best models possible. Our datasets used are as listed below:-

Dataset Name	URL	Status	With Mask Images	Without Mask Images
Dataset01	https://www.kaggle.com/sumansid/facemask-dataset	Applied	208	131
Dataset02	https://www.kaggle.com/omkargurav/face-mask-dataset	Applied	3725	3828
Dataset03	https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset	Applied	5883	5909
Dataset04	https://www.kaggle.com/dhruvmak/face-mask-detection	Applied	220	220

Dataset05	https://www.kaggle.com/prithwirajmitra/covid-face-mask-detection-dataset	Applied	503	503
Dataset06	https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset (1) Real-World Masked Face Recognition Dataset	Discarded	5000	90000
Dataset07	https://www.kaggle.com/wobotintelligence/face-mask-detection-dataset	Discarded	1698	0
Dataset08	https://github.com/prajnasb/observations/tree/master/experiments	Discarded	690	686
Dataset09	https://github.com/balajisrinivas/Face-Mask-Detection	Discarded	1915	1918

4 METHODS

In order to produce the best prediction model for face masks detection, a number of algorithms must be experimented with to find the best algorithm for our problem. For the sakes of this project, we have considered the following algorithms to train our and produce models: -

1. **Support Vector Machine.**
2. **Naive Bayes.**
3. **Logistic regression.**
4. **K-Nearest Neighbor.**
5. **Perceptron.**

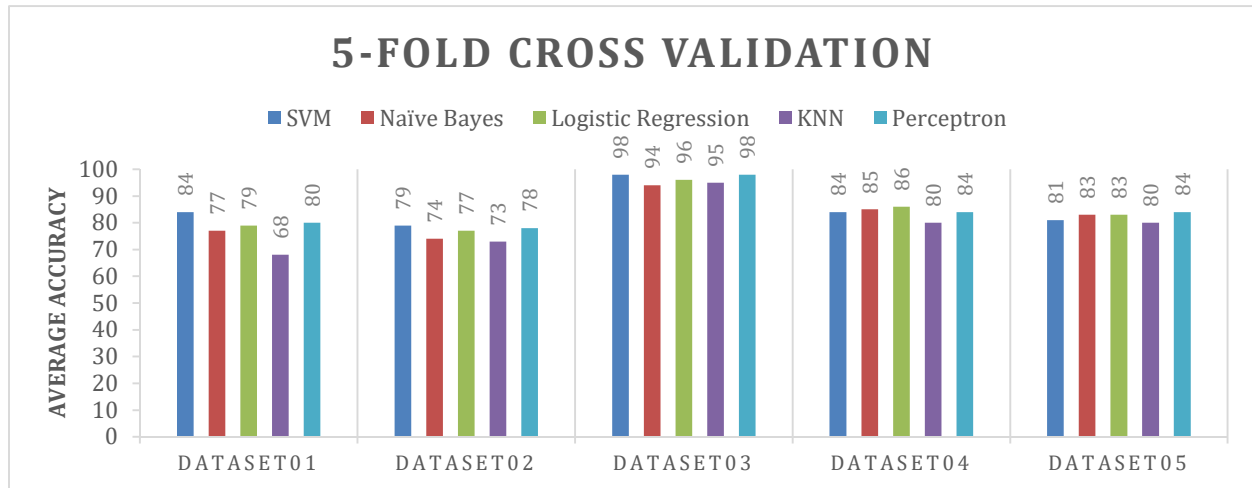
Cross-validation with 5-fold was used to determine the best model. The models were trained using all the five datasets so 25 models (5 datasets * 5 algorithms) with 5-fold cross validation were trained to determine the best five models. After that, for each dataset a model was chosen as the best model. Moreover, from the five best models a model was chosen as the best model out of the five models. As a result, the model can be considered as the best model that was trained with the best dataset. More on the results and the best model in section 4.

Our problem is a classification problem with only 2 classes. Based on this fact, we have utilized confusion-matrix evaluation method to evaluate our predicted labels and compare them to the actual labels from the testing data. Finally, we could generate the 5 performance measures we utilized to evaluate each algorithm. The performance measure are as follows: -

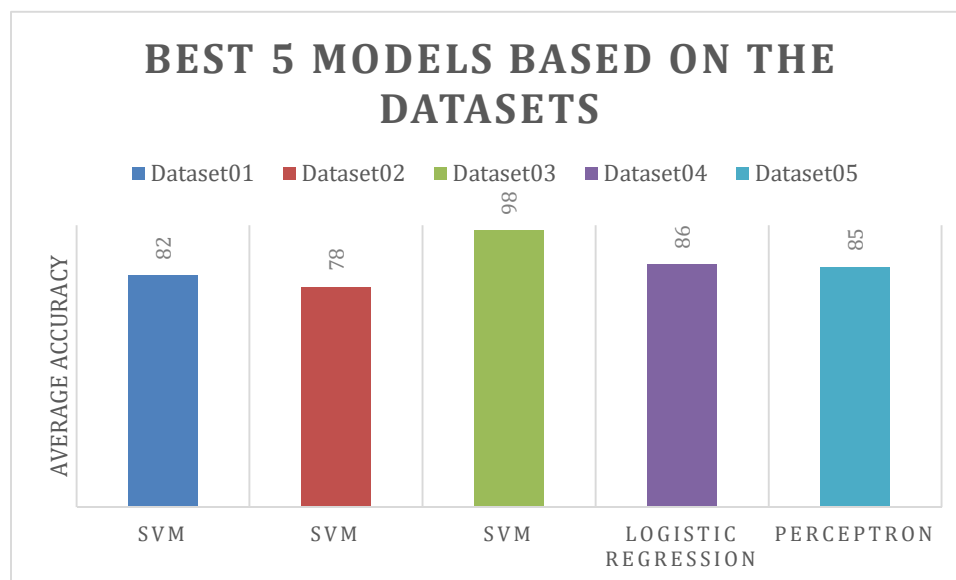
1. Testing Accuracy.
2. F1 score.
3. Sensitivity.
4. Precision.
5. Recall.

5 SIMULATION RESULTS

We have run the algorithms we have chosen using each dataset and recorded the training accuracies. The results are as follows:-



Based on that, we chose the algorithm with the highest training accuracy per each dataset and ran our testing routine. The testing accuracy for each dataset are calculated and the results were as follows: -



The rest of the performance measures are as follows: -

	Algorithm	F1 Score	Sensitivity	Precision	Recall
Dataset01	SVM	0.7	0.57	0.91	0.57
Dataset02	SVM	0.79	0.81	0.77	0.81
Dayaset03	SVM	0.98	0.99	0.89	0.99
Dataset04	Logistic Regression	0.87	0.87	0.87	0.87
Dataset05	Perceptron	0.85	0.85	0.85	0.85

As demonstrated above, using SVM on Dataset03 generated the best overall results. Therefore, we have decided to use the model generated to implement a Telegram based bot that receives a picture of a face and predicts whether the face in the picture is wearing a mask or not. The results of this implementation were accurate.

6 REFERENCES

- Brownlee, J. (2019, August 13). How to Implement Resampling Methods From Scratch In Python. Retrieved November 29, 2020, from <https://machinelearningmastery.com/implement-resampling-methods-scratch-python/>
- Brownlee, J. (2019, August 13). How To Implement The Perceptron Algorithm From Scratch In Python. Retrieved November 25, 2020, from <https://machinelearningmastery.com/implement-perceptron-algorithm-scratch-python/>
- Brownlee, J. (2019, October 24). Naive Bayes Classifier From Scratch in Python. Retrieved November 28, 2020, from <https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/>
- Brownlee, J. (2019, December 11). How To Implement Logistic Regression From Scratch in Python. Retrieved November 29, 2020, from <https://machinelearningmastery.com/implement-logistic-regression-stochastic-gradient-descent-scratch-python>
- Brownlee, J. (2020, February 23). Develop k-Nearest Neighbors in Python From Scratch. Retrieved November 29, 2020, from <https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>
- Abbassi, Q. (2020, April 01). SVM From Scratch-Python. Retrieved December 1, 2020, from <https://towardsdatascience.com/svm-implementation-from-scratch-python-2db2fc52e5c2?gi=d214d6fc8999>
- Histogram of Oriented Gradients¶. (n.d.). Retrieved December 1, 2020, from https://scikit-image.org/docs/dev/auto_examples/features_detection/plot_hog.html
- Jeppbautista. (2018, October 09). Logistic Regression from scratch - Python. Retrieved December 3, 2020, from <https://www.kaggle.com/jeppbautista/logistic-regression-from-scratch-python>
- Loeber. 2020. ML from scratch. <https://github.com/python-engineer/MLfromscratch>. (2020)
- Persson. 2020. Machine Learning Collection. <https://github.com/aladdinpersson/Machine-Learning-Collection>. (2020)
- Golubin. 2020. Machine learning algorithms. <https://github.com/rushter/MlAlgorithms>. (2020)
- Linder-Norén. 2020. Machine Learning From Scratch. <https://github.com/eriklindernoren/ML-From-Scratch>. (2020)
- Bourgin. 2020. Numpy-ml. <https://github.com/ddbourgin/numpy-ml>. (2020)
- Sentdex. 2017. https://www.youtube.com/playlist?list=PLQVvva0QuDfKTOs3Keq_kaG2P55YRn5v. (2020)