

Mini Project on 'Detecting whether the person is wearing masks or not. '

Vaishnavi Katiyar*, Hiteshi Singh[†],

*Indian Institute of Technology, Jodhpur
Pattern Recognition and Machine Learning (CSL2050)

Professor: Richa Singh
`richa@iitj.ac.in`

Introduction

We used the 'Self-built-masked-face-recognition-dataset' data. The database contains 90,000 face images without masks, 2203 face images with masks, and 525 subjects.

As the unmasked dataset was large so we extracted 5557 images of the unmasked dataset and 2142 images of masked dataset were dealt as it is. So here we need to predict the face without mask and with mask. Therefore we converted the images into pixel arrays and now we have the new dataset in the form of csv files. Finally we labelled the dataset, unmasked samples as 1 and masked dataset samples 0..



Type of Classifiers

Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks. Random forests is a supervised learning algorithm. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting.

Gaussian Naive Bayes

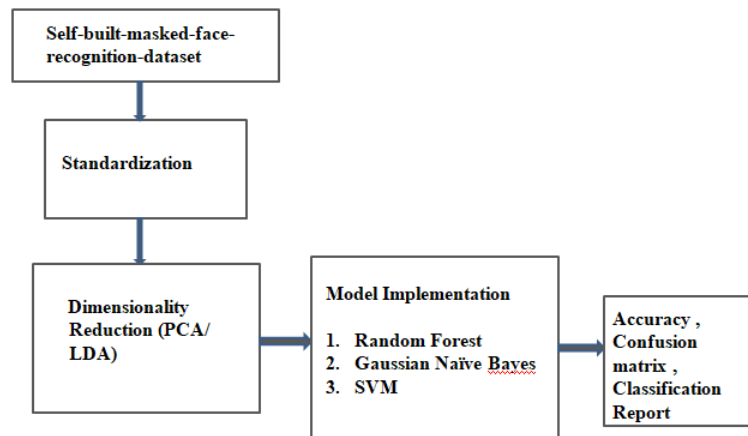
A Gaussian Naive Bayes algorithm is a special type of NB algorithm. It's specifically used when the features have continuous values. It's also assumed that all the features are following a gaussian distribution i.e, normal distribution.

Support vector Machine(SVM)

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes.

General Pipeline

A common pipeline was followed for all the three classifiers used. And it is given as follows:

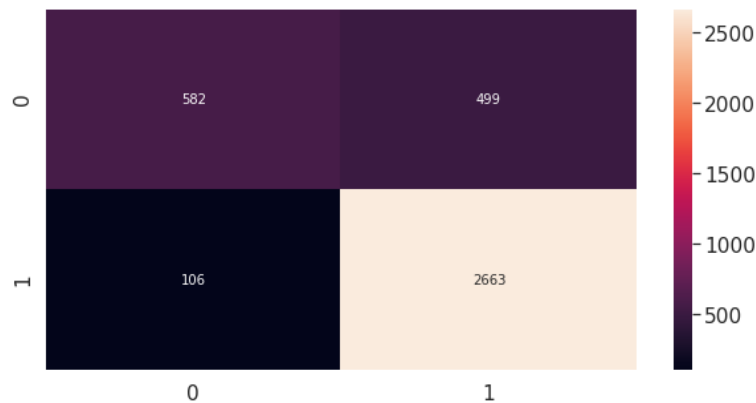


Random Forest

The Random Forest was implemented using sklearn .Number of trees used in this ensemble method are 200 i.e number of estimators= 200 and maximum depth=2.

The model was implemented before and after applying PCA on the dataset .

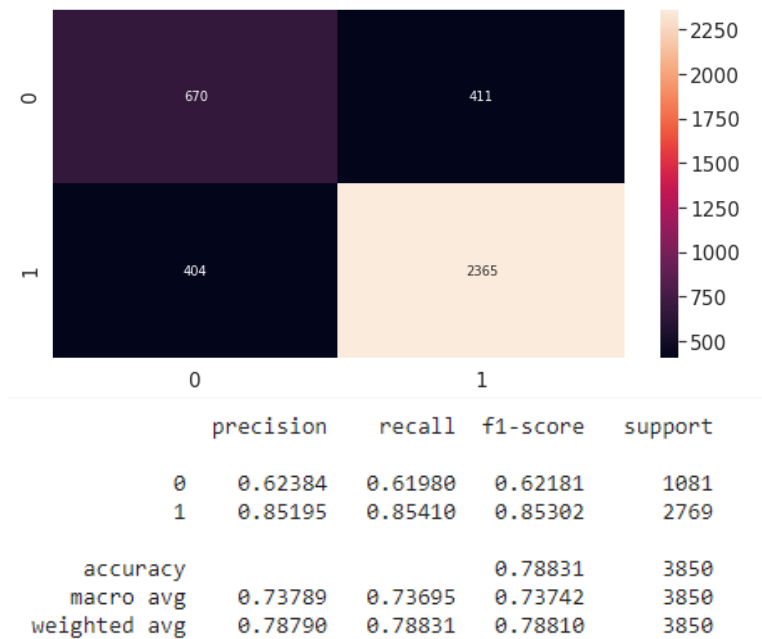
Accuracy reported was highest in the case when PCA was not applied



	precision	recall	f1-score	support
0	0.84593	0.53839	0.65800	1081
1	0.84219	0.96172	0.89799	2769
accuracy			0.84286	3850
macro avg	0.84406	0.75005	0.77800	3850
weighted avg	0.84324	0.84286	0.83061	3850

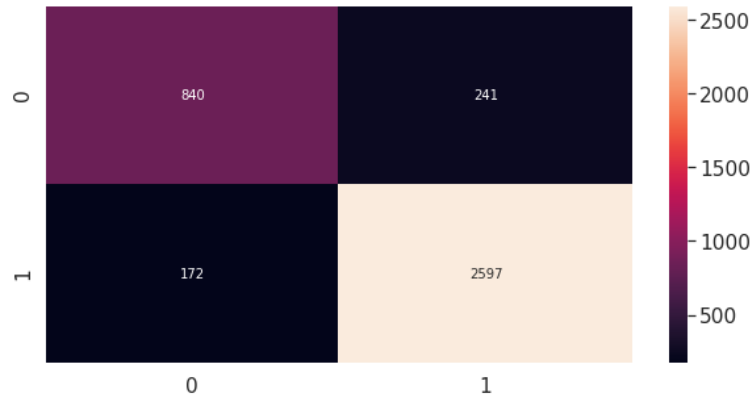
Gaussian Naive Bayes

We implemented the naive bayes algorithm using the Gaussian NB classifier with the parameters = priors=None, smoothing=1e-09 and computed the accuracy on the dataset which is not normalised. Finally we obtained the accuracy = 0.78



Support Vector Machine (SVM)

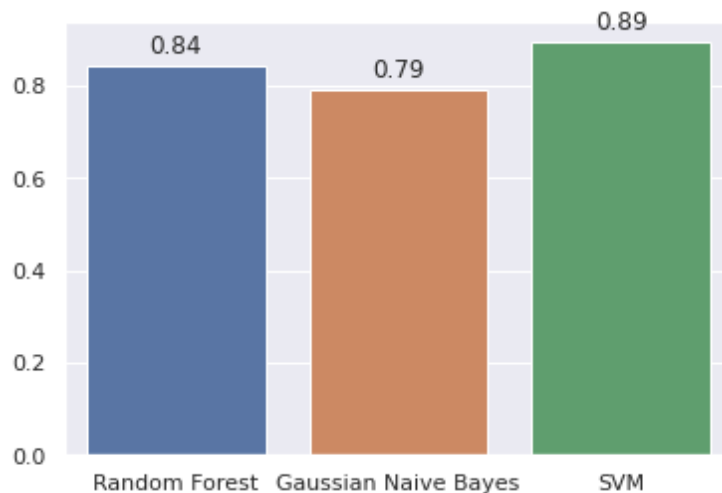
We implemented the SVM algorithm using the library from sklearn and using the parameters C=1, kernel = 'rbf', tol=0.001, degree=3. We applied first SVM on the dataset created by us (by converting images to pixels array) and then applied on the normalised dataset of pixel values and compared the performance.



	precision	recall	f1-score	support
0	0.83004	0.77706	0.80268	1081
1	0.91508	0.93788	0.92634	2769
accuracy			0.89273	3850
macro avg	0.87256	0.85747	0.86451	3850
weighted avg	0.89120	0.89273	0.89162	3850

Conclusion

After implementing 3 different models with a general pipeline, it was observed that Support Vector Machine (SVM) applied on normalised dataset performed best with 88.9% accuracy, Random Forest on PCA implemented data gave 84.2% accuracy and Gaussian NB with 78.8% accuracy. They lack the adaptive ability to handle performance. But nowadays, there are more advanced and optimizable models like Support Vector Machines etc. with more control on Hyper-Parameters.



References

- https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html
- <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
- <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- https://drive.google.com/file/d/1EG_A3kRwaPn15AFUGmEaQXatKZNxrWxH/view

Contribution

Hiteshi Singh(B19EE039):- Data preprocessing , Random Forest classifier and Report
Vaishnavi Katiyar (B19EE087):- Gaussian Naive Bayes , Support Vector Machine (SVM) classifiers and Report