BYTE PANACHE COMPETITION

FOUNDATIONS OF MODERN MACHINE LEARNING

OBJECTIVES

The following are the objectives of this project:

1. Develop a model to predict if a star is A Pulsar star (a rare type of Neutron star that produces radio emission detectable here on Earth)
2. Use Support Vector Machine for classification. (supervised ML algorithm)
3. Use kernel method, which is used to classify non-linear problem using linear classifier.

MATERIALS & METHODS

The following materials were required to complete the research:

- Google Colab
- Pulsar star csv file with the following attributes : Mean of the integrated profile, Standard deviation of the integrated profile , Excess kurtosis of the integrated profile, Skewness of the integrated profile, Mean of the DM-SNR curve, Standard deviation of the DM-SNR curve, Excess kurtosis of the DM-SNR curve, Skewness of the DM-SNR curve and Class

INTRODUCTION

Support Vector Machines (SVM) are supervised machine learning algorithms. The kernel is a method of using a linear classifier to solve a non-linear problem and is used due to a set of mathematical functions used in the SVMs to manipulate the data. In this kernel, we built a SVM classifier to classify a Pulsar star. Pulsars are a rare type of Neutron star that produces radio emission detectable here on Earth. Classification algorithms in particular are being adopted, which treat the data sets as binary classification problems. We used various kernels to build this SVM classifier.

RESULTS 2

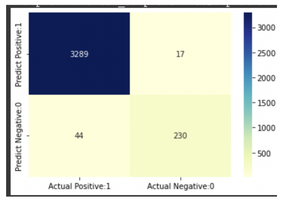

Figure 4: Confusion Matrix

	precision	recall	f1-score	support
0	0.99	0.99	0.99	3306
1	0.93	0.84	0.88	274
accuracy			0.98	3580
macro avg	0.96	0.92	0.94	3580
weighted avg	0.98	0.98	0.98	3580

Figure 5: Classification report

- In figure 4 we use confusion matrix to describe the performance of a classification model on a set of test data for which the true values are known
- In figure 5 we use classification report to measure the quality of predictions from a classification algorithm.

RESULTS 1

```
Model accuracy score with default hyperparameters: 0.9827
```

Figure 1: SVM with default hyperparameters

```
Model accuracy score with linear kernel and C=100.0 : 0.9832
```

Figure 2: Run SVM with linear kernel and C=100.0

```
Training set score: 0.9783
Test set score: 0.9830
```

Figure 3: Comparing test set and train set accuracy

We need to build highly accurate model for our data set and for that we tune the hyper parameters of our model.

- In figure 1 we use default parameters of support vector machine which is C=1.0, kernel=rbf and gamma=auto and we get an accuracy of 0.9827
- In figure 2 we use Svm with linear kernel and c=100.00 and the accuracy is 0.9832
- In figure 3 we compare accuracy of training data and test scores to check for overfitting or underfitting and as we can see there is no question of overfitting

CONCLUSION

We get maximum accuracy with rbf and linear kernel with C=100.0 and the accuracy is 0.9832. Polynomial kernel and sigmoid kernel gives a poor performance as they may be overfitting the training set. So, we can conclude that our model is doing a good job in terms of predicting the class labels. But, this is not true. Here, we have an imbalanced dataset. Accuracy is an inadequate measure for quantifying predictive performance in the imbalanced dataset problem. So, we must explore confusion matrix and classification metrics to know the underlying distribution of values and the type of errors our classifier is making which provides better guidance in selecting models.

REFERENCES

[1] Aurélién Géron. *Hands on Machine learning*. Wiley, 1st edition, 2017.

[2] Wikipedia. Support vector machines. https://en.wikipedia.org/wiki/Support-vector$_machine, March$2015.

FUTURE RESEARCH

- Using the same model on different types of astronomical objects like black holes and quasars and predicting there energy release during their outburst
- Avoid training data biased towards bright pulsars..

CONTACT INFORMATION

FMML20211049 :fmml20211049@ihub-data.iiit.ac.in
FMML20211050 :fmml20211050@ihub-data.iiit.ac.in
FMML20211052 :fmml20211052@ihub-data.iiit.ac.in