Searching for supersymetric Higgs bosons in high-energy physics with machine learning

Aspasia Pallikaridou *

Aristotle University of Thessaloniki
School of Physics
MSc Computational Physics
Professor: Th. Diakonidis

1 Introduction

This assignment is about the implementation of some machine learning techniques for choosing the events in LHC that are connected to supersymetry Higgs bosons. According to supersymetry, Higgs sector consists five bosons. On the other hand, according to the standard mondel, only one higgs boson exists.

The particular analysis is for heavy Higgs bosons, which decay into:

$$H^0 \to W^+ + W^- + h$$

$$W^\pm \to l^\pm + v, \;\; h \to b + b$$

The exercise is based on the article *Searching for exotic particles in high-energy physics with deep learning* by P. Baldi, P. Sadowski and D. Whiteson. The data are separated in classification, low level and high level quantities.

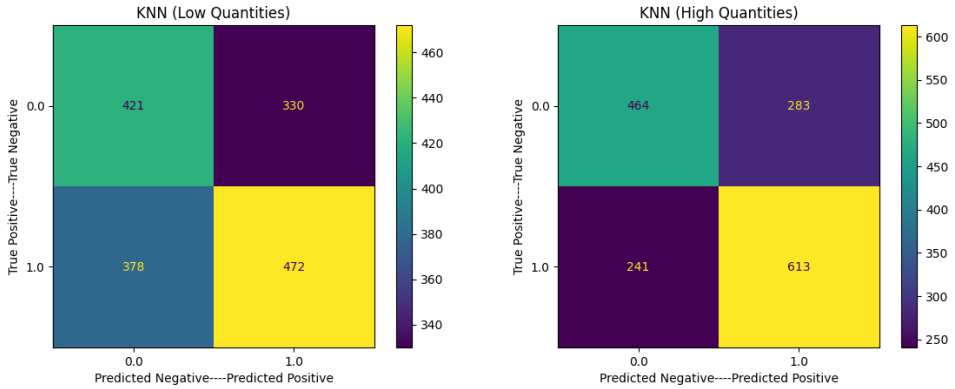
Two supervised machine learning methods where used, KNN classification, and Random Forest. Additionally, an analysis has been done with an artificial neural network technique that uses the tensorflow library. All the three methods were implement separately for low and high quantities in a Jupiter Notebook.

2 Supervised machine learning: KNN classification

In the **KNN** classification with the low level quantities 10 neighbors were used. The reason is that for either 5 or 4 or 6 there was at least one non diagonal element greater than one diagonal in the confusion matrix 1a. The accuracy was 0.5578. In the classification with the high level quantities five neighbours were used (1b). The accuracy was higher than before, 0.6727.

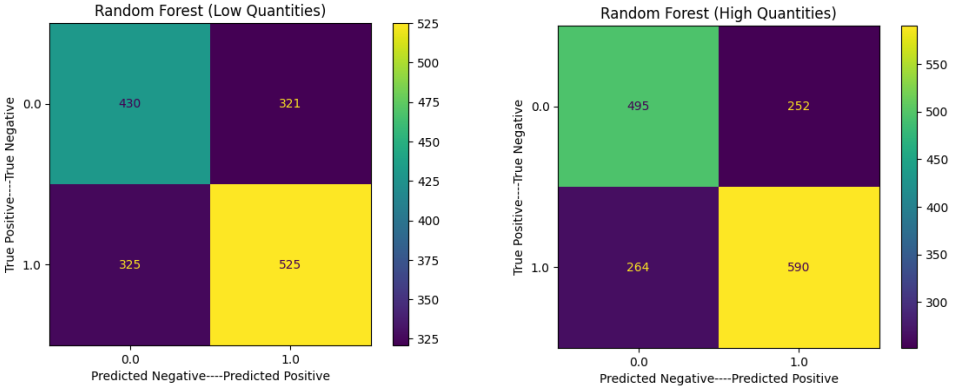
In the **Random Forest** classifications with low level and high level quantities, the accuracy was respectively 0.5965 and 0.6777. As a result, random forest analysis is more precise than KNN. Also, both analyses are more accurate when they depend on the high level quantities. This can be clearly noticed in the ROC curve 3.

*apallika@physics.auth.gr



(a) (b)

Figure 1: Confusion Matrices: KNN classification of higgs bosons' data with low level and high level quantities



(a) (b)

Figure 2: Confusion Matrices: Random Forest classification of higgs bosons' data with low level and high level quantities

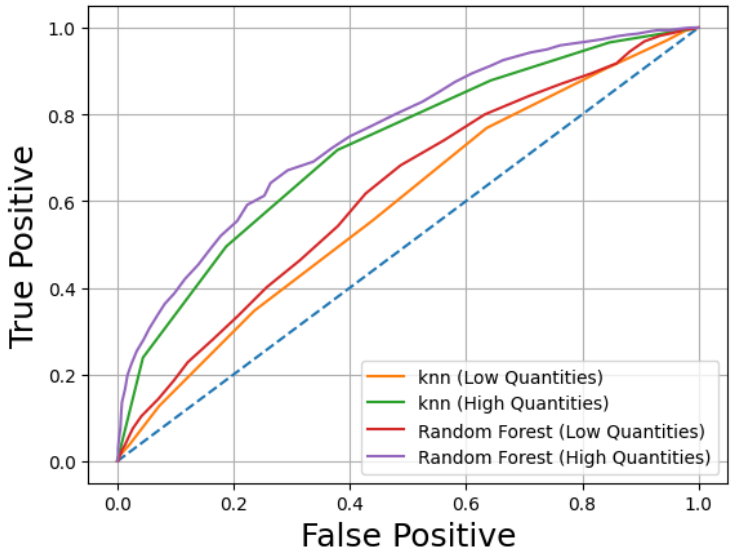


Figure 3: ROC curve for KNN and Random Forest classifications with low level and high level quantities

3 ANN: Tensorflow

The analysis of the low level quantities with tensorflow had accuracy equal to 0.5653. Furthermore, the analysis of the high level quantities with tensorflow had accuracy equal to 0.6852. Again, higher accuracy was accomplished with high level quantities.

4 Conclusion

In conclusion, Artificial Neural Network technique were almost as accurate as the two supervised machine machine learning methods. The difference between the methods is not significant. Finally, high level quantities lead to more precise results, hence they are more related to higgs bosons.

References

[1] *Searching for exotic particles in high-energy physics with deep learning* by P. Baldi, P. Sadowski and D. Whiteson

[2] Courses notes by Th. Diakonidis