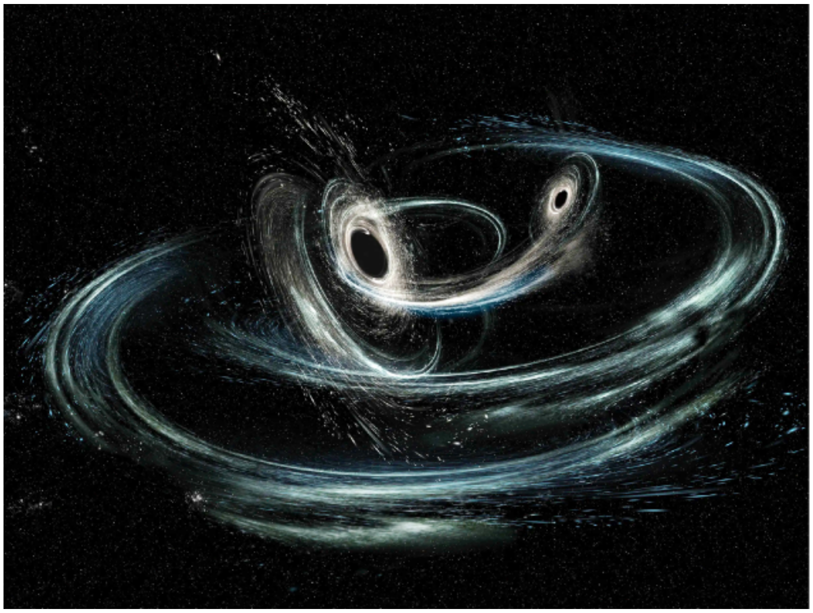
Alexander Winegardner

Gravitational Wave Analysis Report



Problem Statement

Gravitational wave astronomy has the potential to uncover many mysteries of the universe. Current Earth-based observatories are sensitive enough to detect signals that originate from merging pairs of compact objects, such as black holes and neutron stars. While each signal can take significant amounts of time and computational resources to fully analyze, not all will have a high probability of being from a real astrophysical event. A solution is needed that can serve as an early screening mechanism to filter out potentially uninteresting signals, while giving greater priority to those that might come from real astrophysical sources. In this report, I will outline the steps I took to explore the relevant data, as well as present a machine learning model built to classify which detections are likely to have 99% or greater probability of being a real astrophysical event (before having to do full parameter estimations).

Data

The data for this project was provided by the Gravitational Wave Open Science Center. Specifically, we used the Gravitational-wave Transient Catalog (GWTC) as our main source. The GWTC is the cumulative set of all confidently-detected events from multiple data releases, and is maintained by the LIGO/Virgo/KAGRA collaboration. The original dataset contained 93 rows for each event and 43 columns for various features such as: the masses of each black hole/ neutron star, total mass of the system, final mass, spin components, luminosity distance, redshift, network matched filter SNR (signal-to-noise ratio), probability of astrophysical origin, and more, as well as several columns containing the error bars for each measurement.

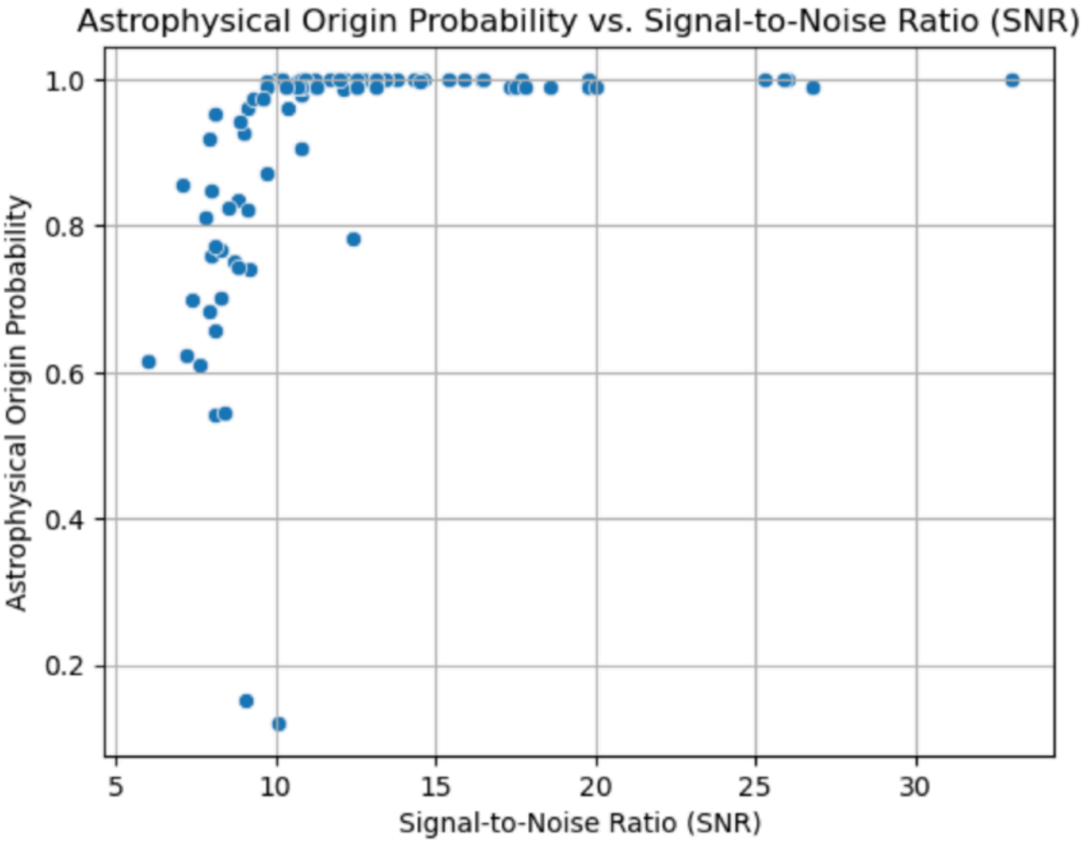
Data Wrangling

For our data wrangling steps, since we are interested in building a model that can classify which detections are likely to be from a real astrophysical event before parameter estimations are done, we cannot use any features that are the results of the parameter estimations themselves. This includes the majority of the features in our dataset and especially those which describe the actual physical properties of the binary systems in question. Because of this, we are left with only a few features that are acquired directly from the signal during early stages of a detection.

The obvious place to start was by looking at the SNR (signal-to-noise-ratio), which is a measure of how distinguishable a signal is from the background noise, and see how far that alone could take us. We were thus able to drop the majority of the columns in our dataset and keep only: the common name for each event, the value of the SNR, and probability of astrophysical origin.

Exploratory Data Analysis

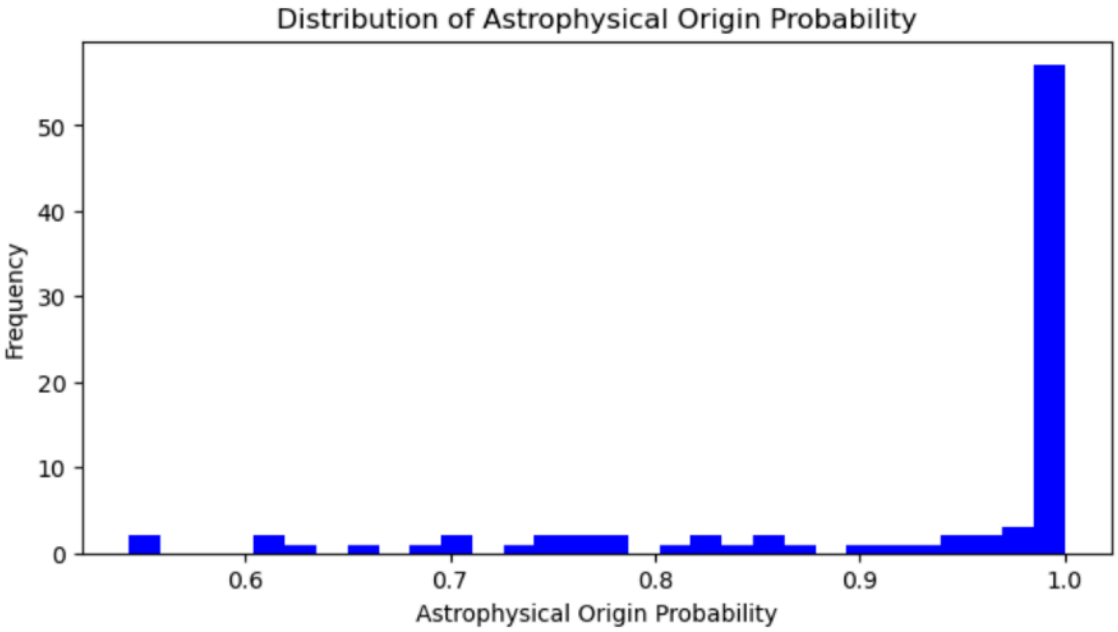
In the EDA phase of this project, we simply created a plot to visualize the astrophysical origin probability versus the signal-to-noise ratio:



It was immediately apparent that there is a very steep drop-off in the astrophysical origin probability for events with an SNR less than ~10. While it was expected that the higher the SNR, the higher the astrophysical probability, it was surprising to see such a striking cutoff point. Instead of having to just guess or "eyeball" which exact value of SNR would be the optimal decision boundary to classify high astrophysical origin probability events, this plot gave us the intuition that a machine learning model would be better suited to the task.

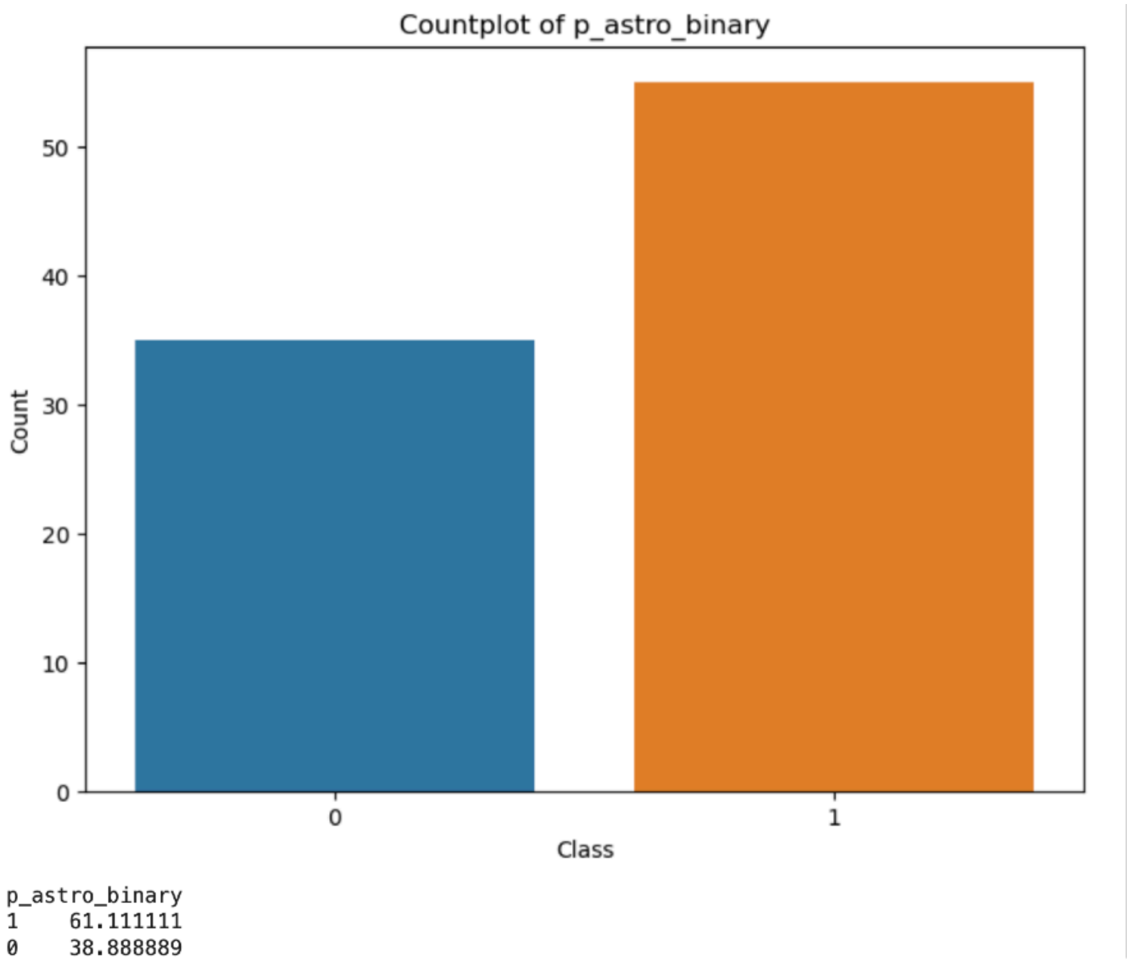
Modeling

In this phase of the project, we set out to create three different machine learning models and then compare which one worked best. The three types of models we chose were: logistic regression, decision tree, and support vector machine (SVM). However, before we could actually begin to build the models, we first had to do a bit of data pre-processing. This included binarizing our target variable (p_astro) based on a threshold value. In order to choose our threshold value, we did some summary statistics on our data and also created a few more visualizations:



Given that we found the 50% percentile value for p_astro was 0.99, and by what we can also see from the plot, it looks like about half the values are >= 0.99 and the other half < 0.99, so we chose that as the threshold.

We also made a countplot of our binarized variables to get a better sense of the balance:

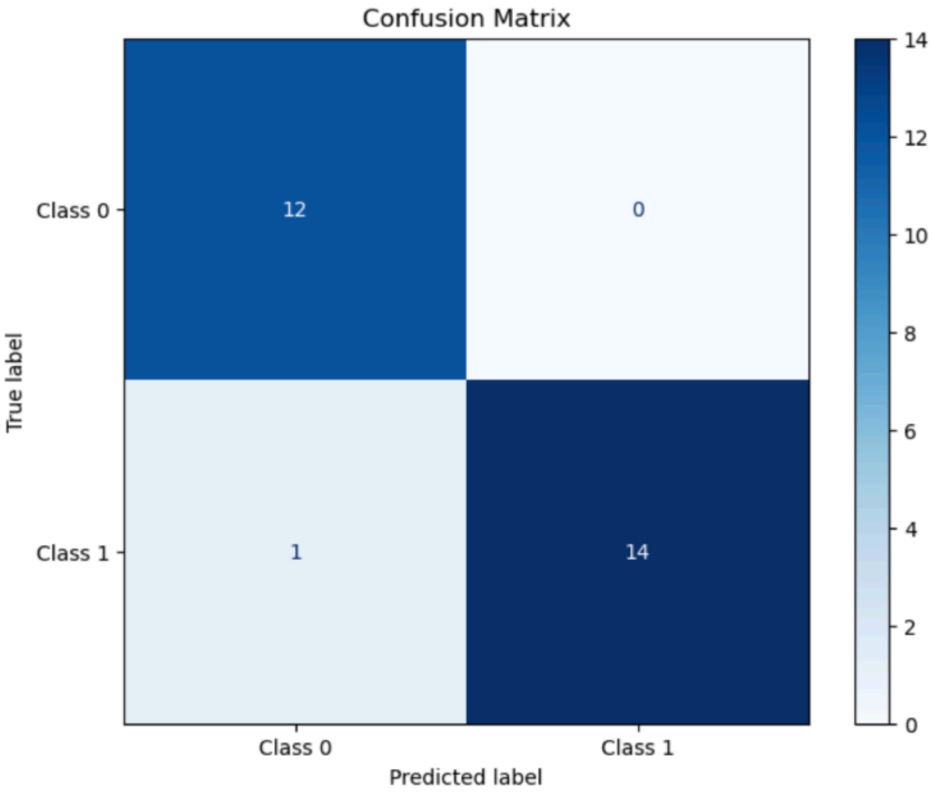


```
p_astro_binary
1    61.111111
0    38.888889
```

While not a perfectly balanced dataset (with a split of about 60/40), it was still good enough to continue.

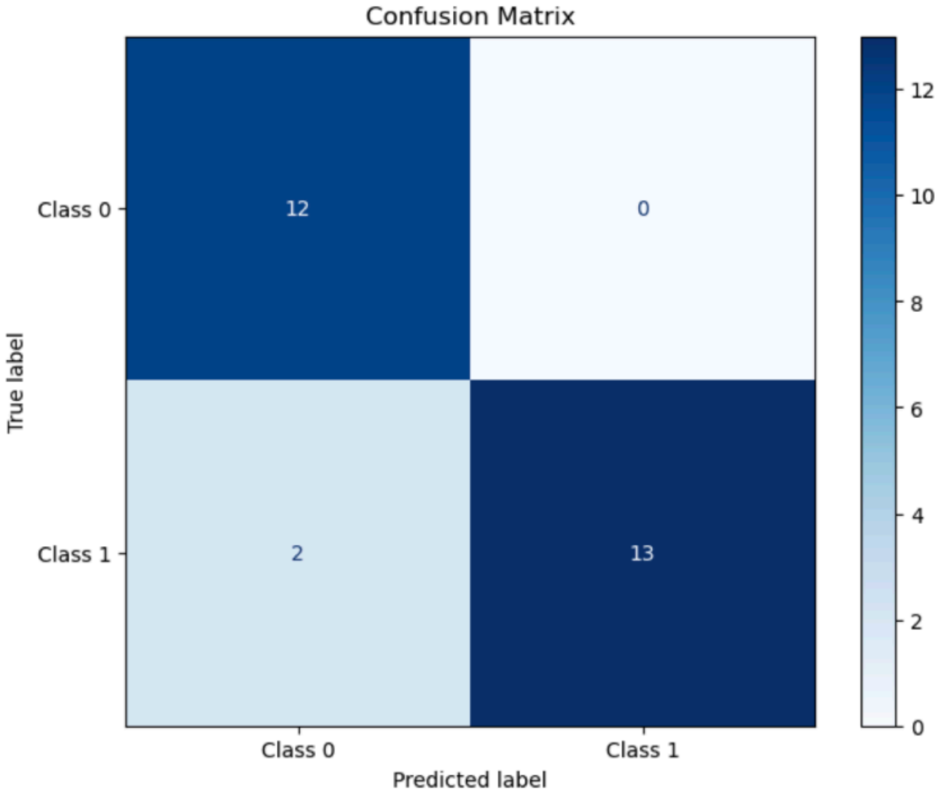
Our next step was to split our pre-processed data into train and test subsets. We took the standard approach of training on 70% of the data while testing on the remaining 30%. We then fit the data to each of our three machine learning models and made predictions. Below are the results of the confusion matrix for each model:

Logistic Regression

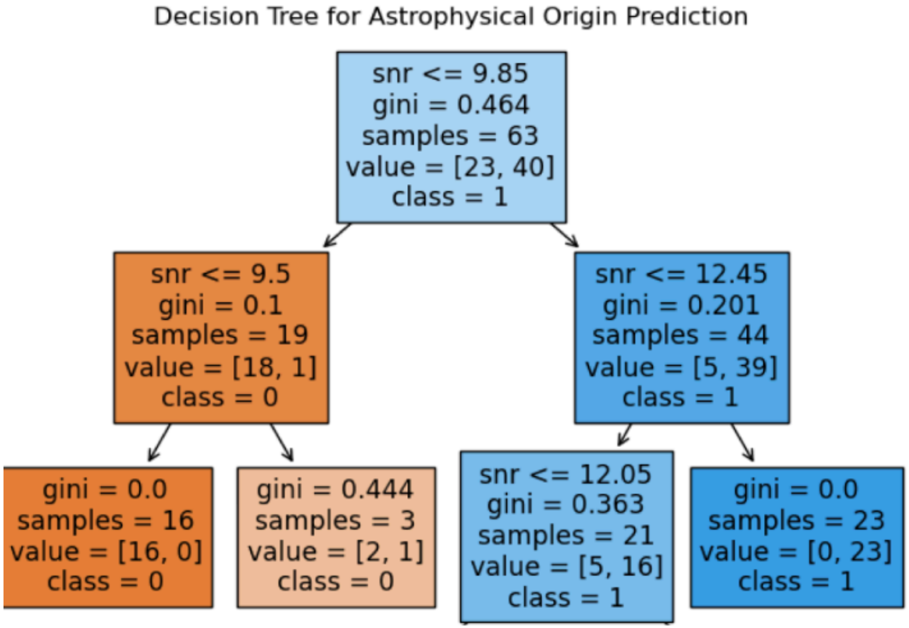


- Above we can visualize the confusion matrix and see that only one of the labels were predicted incorrectly. This was for a class 1 label (p_astro >= 0.99) being mistaken as a class 0 label (p_astro < 0.99). This probably happened to an event that had its SNR value very close to the decision boundary.
- For class 0, a precision of 0.92 means that 92% of the predicted labels were classified correctly (12 out of a total of 13 predictions), while a recall score of 1.00 means that 100% of the true labels were classified correctly (12 out of 12).
- For class 1, a precision of 1.00 means that 100% of the predicted labels were classified correctly (14 out of 14 predictions), while a recall score of 0.93 means that 93% of the true labels were classified correctly (14 out of a total of 15).

Decision Tree

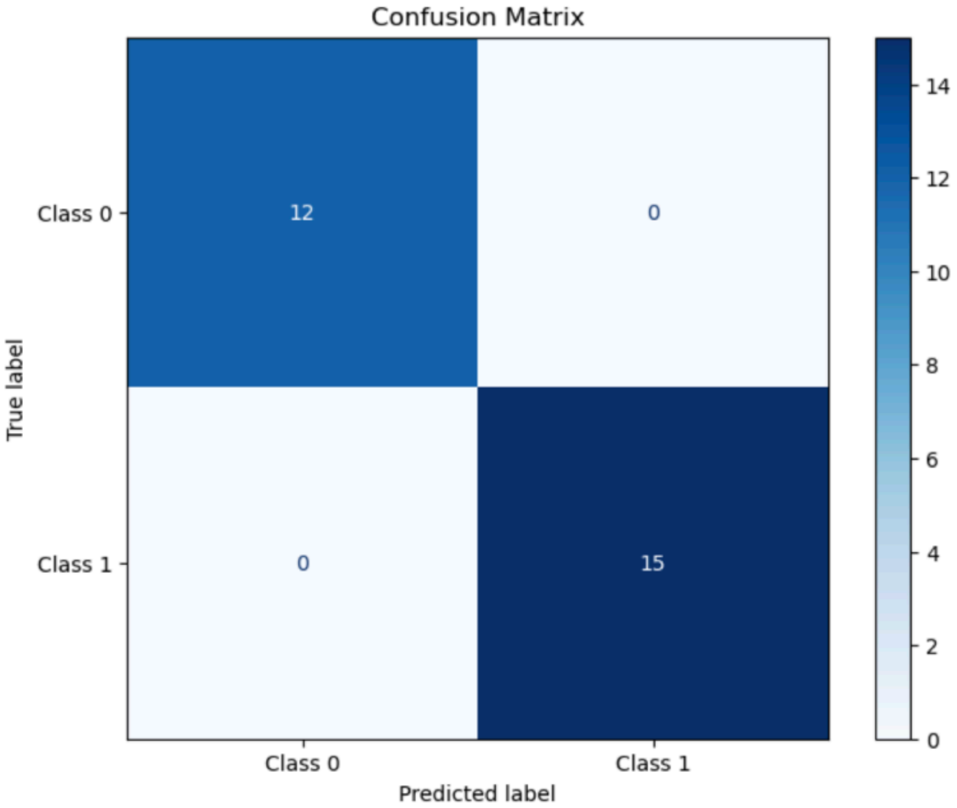


Our decision tree model performed slightly worse, this time incorrectly predicting two labels. While decision trees are prone to overfit (which might be the reason the performance got worse), they are at least more interpretable. Below I include a graph showing the first few branches of how the decision tree model reached its results:



It is interesting to see that it picked out an SNR of 9.85 as its first split, which is close to the value of ~10 we could see from the plot we made earlier during the EDA phase.

Support Vector Machine (SVM)



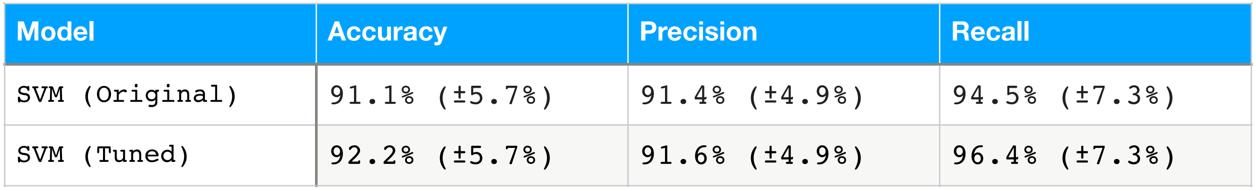
This time we got a perfect score – but sometimes that may be cause for concern. In order to account for this, we performed cross-validation on all our models to create a more representative expectation of how they would perform on new, unseen data:

Model	Accuracy	Precision	Recall
Logistic Regression	88.9% (±5.0%)	91.2% (±5.3%)	90.9% (±5.7%)
Decision Tree	86.7% (±2.7%)	90.0% (±5.6%)	89.1% (±10.6%)
SVM	91.1% (±5.7%)	91.4% (±4.9%)	94.5% (±7.3%)

This seems more realistic. Before we might have just happened to have been making predictions on a "lucky" test set. However, it is still apparent that the SVM model performs better across all key metrics.

Hyperparameter Tuning

After selecting the SVM as the best model, we continued to improve it through hyperparameter tuning. We used GridSearchCV to scan the optimal values for our model, searching over set values for the regularization parameter, kernel coefficient, and kernel type. Since our dataset was small, we were able to tune the model in just over one minute. Once we identified the best parameters for the model, we built a new version and then compared it to the original:



Model	Accuracy	Precision	Recall
SVM (Original)	91.1% (±5.7%)	91.4% (±4.9%)	94.5% (±7.3%)
SVM (Tuned)	92.2% (±5.7%)	91.6% (±4.9%)	96.4% (±7.3%)

We can see a clear improvement across all key metrics. Particularly, we notice the tuned model has gained an extra percent for accuracy and almost two percent for recall (while very slightly improving precision).

Conclusion

We have successfully built a prototype for a machine learning model that helps us accomplish our goal of predicting whether a signal will have a very high astrophysical origin probability (>= 99%) based just on its SNR.

Our highest score was for recall which in this case is what we want – it means that the vast majority (96.4% on average) of labels with p_astro >= 99% have been correctly classified. The lower precision (91.6% on average) means that some labels with p_astro <= 99% were predicted to have p_astro >= 99%. However, even for these cases, it is probable they still had pretty high p_astro values nonetheless (it is much more likely for events with p_astro between 90-98% to be misclassified than for events with p_astro between 50-80%, for example).

It is still to be seen if implementing a model like this before proceeding with the full parameter estimations is actually useful or practical, especially now while gravitational wave detections are relatively few. But as gravitational wave detectors become more sensitive and are able to pick up signals at a much higher rate, future observation runs might need to prioritize which signals to estimate parameters for first.

Future Research

While this project has served as a demonstration of what could be possible by making classifications based just on SNR, if there is ever serious interest in pursuing it further, there are several things which can be done to improve:

- The first is to simply use more data once it is available. The more data the better!
- The second is to consider even more models, such as gradient boosting machines (GBM) / XGBoost, and to perform an even more exhaustive set of hyperparameter tuning for all models.

- The third is to determine if there are more features which are not dependent on the results of full parameter estimations, such as false alarm rates (FAR), which could be used in conjunction with SNRs for early screening.

Implementing the steps above could help build an even better model for handling new, unseen data. However, it is important to always remember the limitations of the model. When determining whether a new signal is worth further analysis, several other factors are taken into account, and a machine learning model alone should not be expected to authoritatively classify everything correctly. To be useful, having a human expert in the loop for validation may be necessary as well as including other analysis methods. If successful, the model could potentially help create a catalog of very high confidence events in a more efficient manner.

Acknowledgments

- R. Abbott et al. (LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration), “Open data from the third observing run of LIGO, Virgo, KAGRA and GEO”, ApJS 267 29 (2023) -- INSPIRE
- R. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), “Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo”, SoftwareX 13 (2021) 100658 -- INSPIRE

“This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo

Collaboration, and KAGRA. This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.”

Image Credit

The cover image for this report is adopted from art by Aurore Simonnet at Sonoma State.