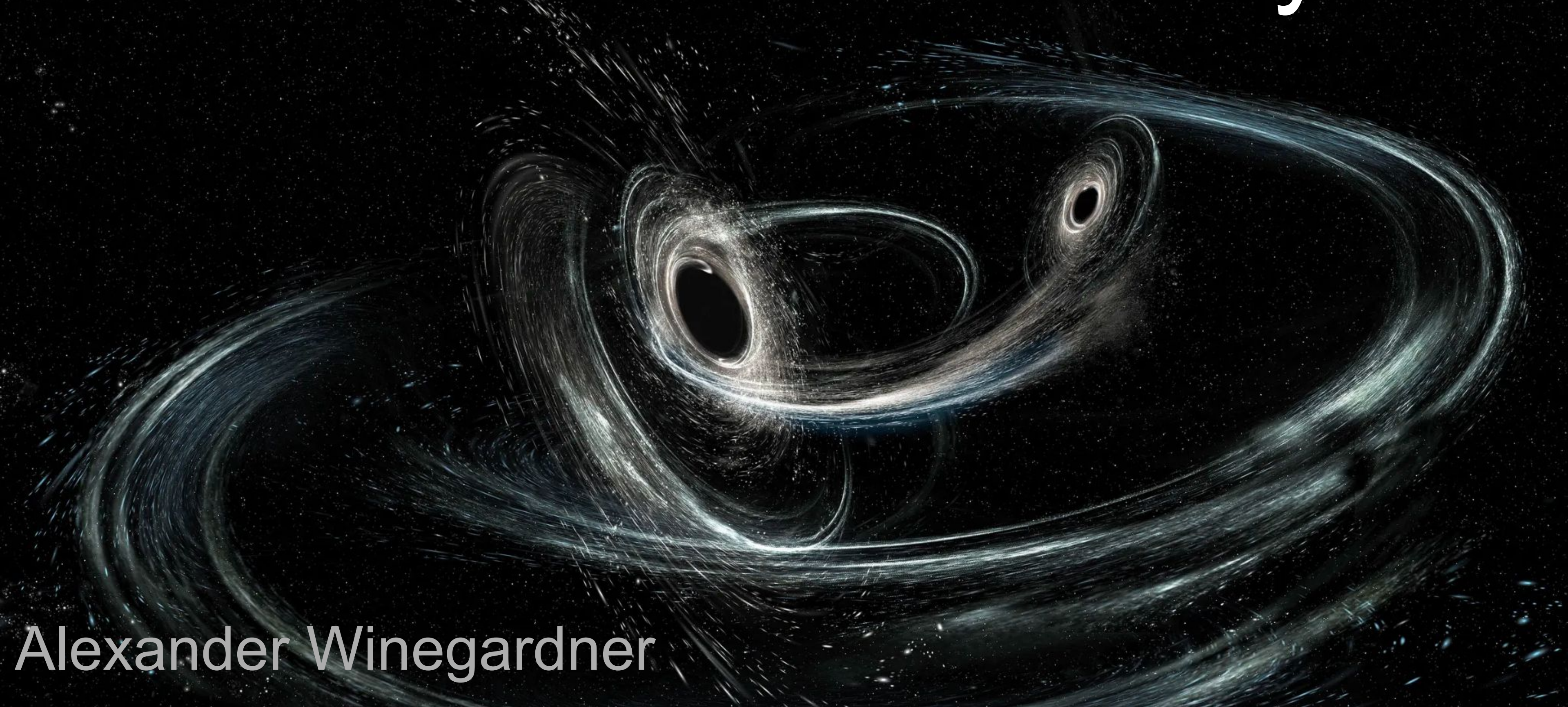
Gravitational Wave Analysis



Alexander Winegardner

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

Goal

- Build a machine learning model that can 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
- 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 relating to the signal and physical properties of the system

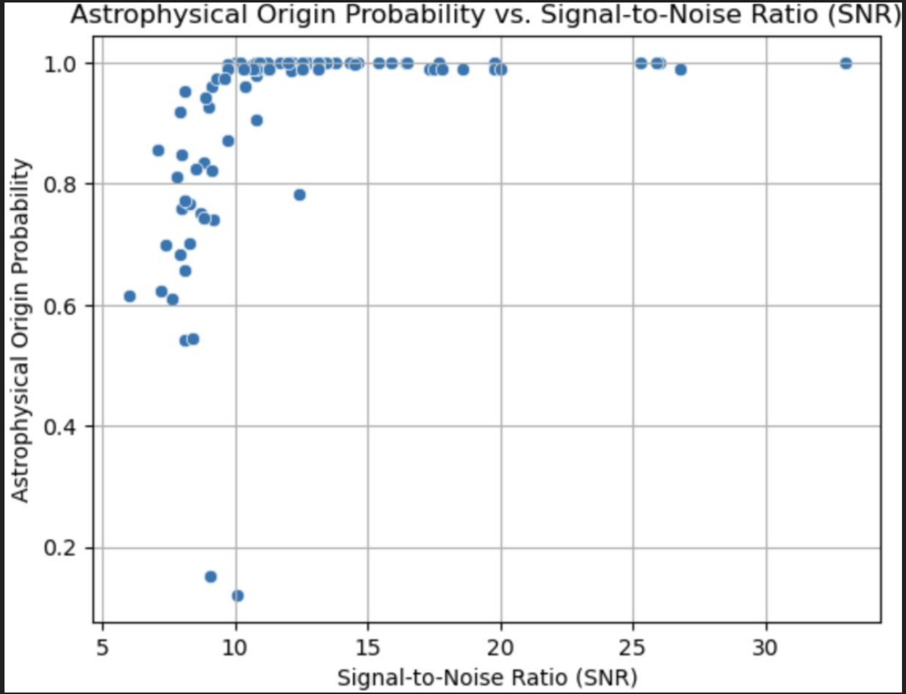
Data Wrangling

- We were not interested in any features that were the results of the parameter estimations themselves

- Only a few features remained that were acquired directly from the signal during early stages of a detection

- We chose the the SNR (signal-to-noise-ratio) as our main feature, and astrophysical origin probability as our target variable

Exploratory Data Analysis

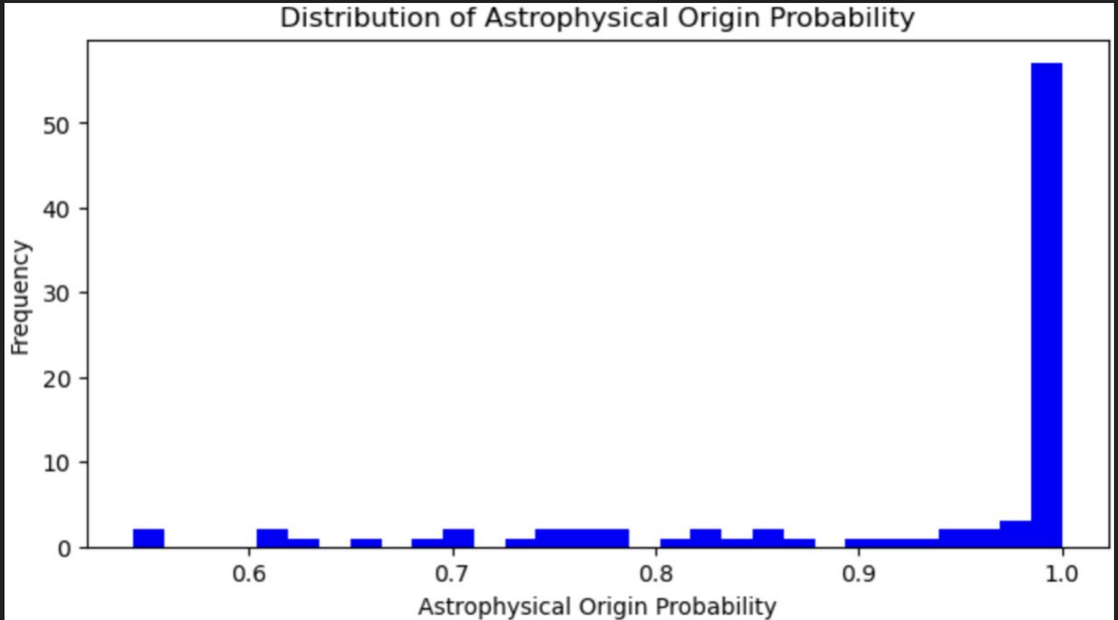


Very steep drop-off in the astrophysical origin probability for events with an SNR less than ~10

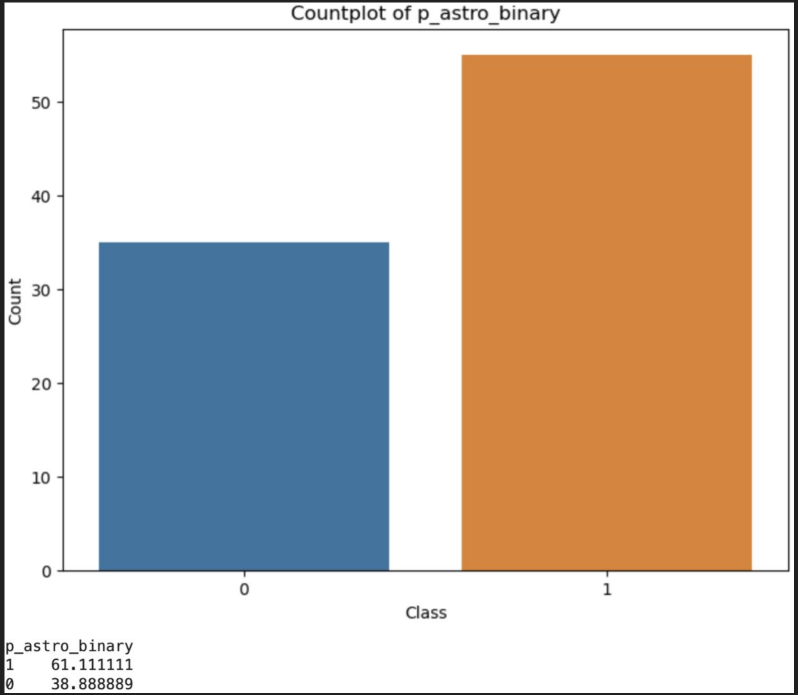
Data Pre-processing

- We first had to binarize our target variable (p_astro) by choosing a threshold
- Summary statistics: 50% value for p_astro was 0.99

Histogram:



Data Pre-processing



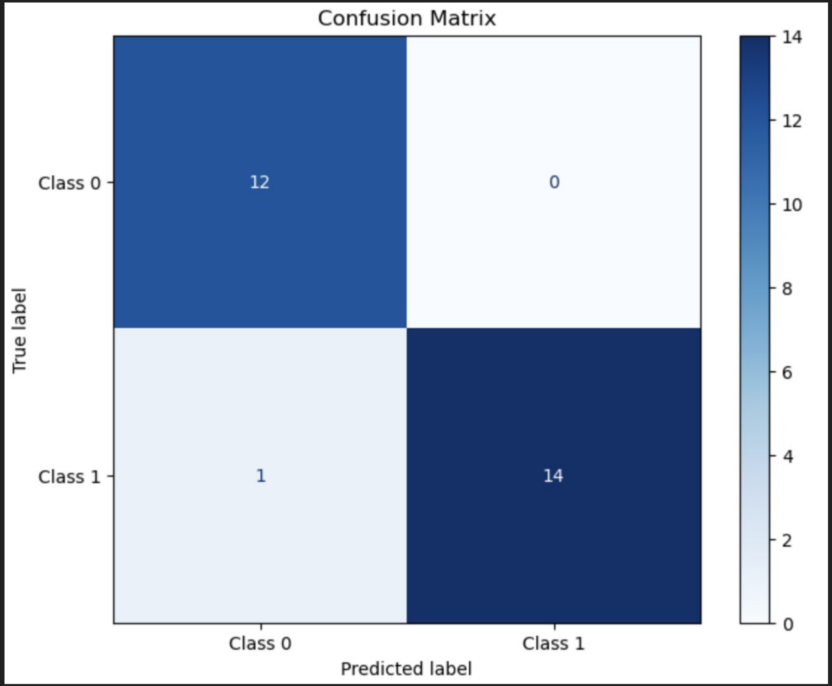
- Class 1: p_astro >= 0.99
- Class 0: p_astro < 0.99

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

Modeling

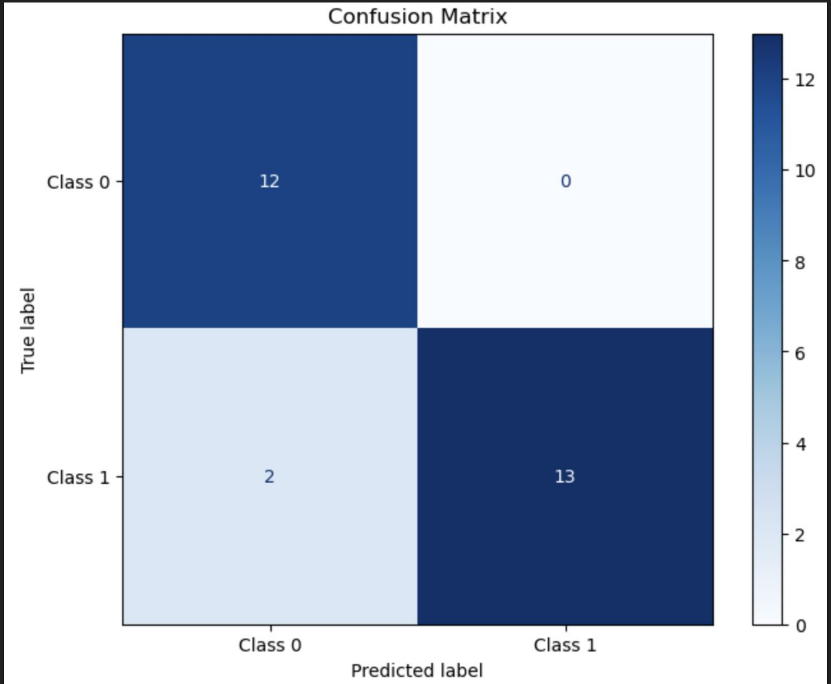
- Three types of models were chosen:
- Logistic regression, decision tree, and support vector machine (SVM)
- Trained on 70% of the data and tested on the remaining 30%
- We then fit the data to each of our models and made predictions

Logistic Regression



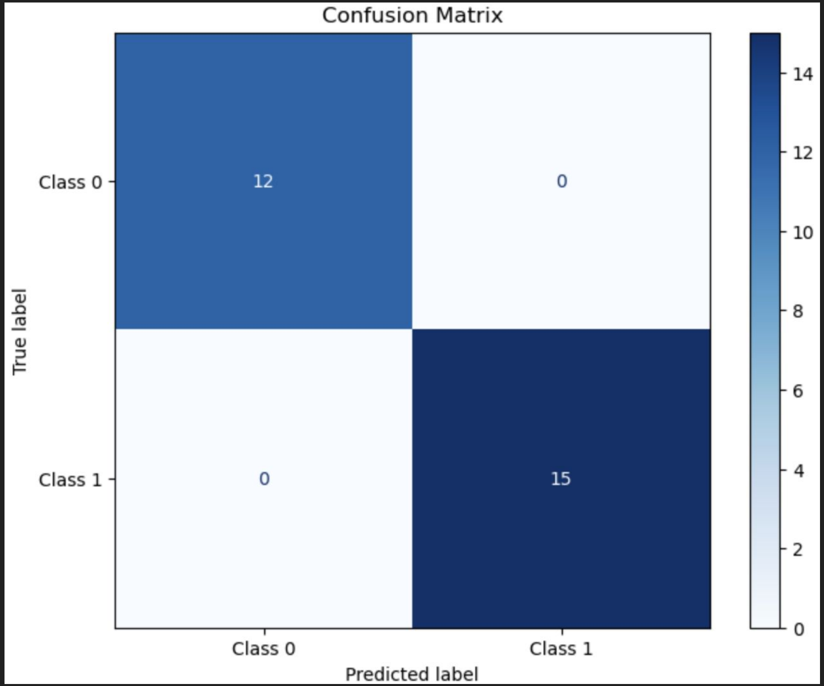
- Only one of the labels was predicted incorrectly
- This was for a class 1 label (p_astro >= 0.99) being mistaken as a class 0 label (p_astro < 0.99)

Decision Tree



- Our decision tree model performed slightly worse, this time incorrectly predicting two labels
- Decision trees are prone to overfit

SVM



- This time we got a perfect score – but sometimes that may be cause for concern

Cross-Validation

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%)

It is clear that the SVM model performs the best

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
- Once we identified the best parameters, 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%)

Conclusion

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

- Our highest score was for recall (96.4% on average)

- Our lowest score was for precision (91.6% on average)

- As gravitational wave detectors become more sensitive and are able to pick up signals at a much higher rate, future observation runs might benefit from a model that can prioritize which signals to estimate parameters for first

Future Research (3 Ideas)

- The first is to simply use more data once it’s 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

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 presentation is adopted from art by Aurore Simonnet at Sonoma State