CORNER CLIPPING EVENTS IN ICECUBE

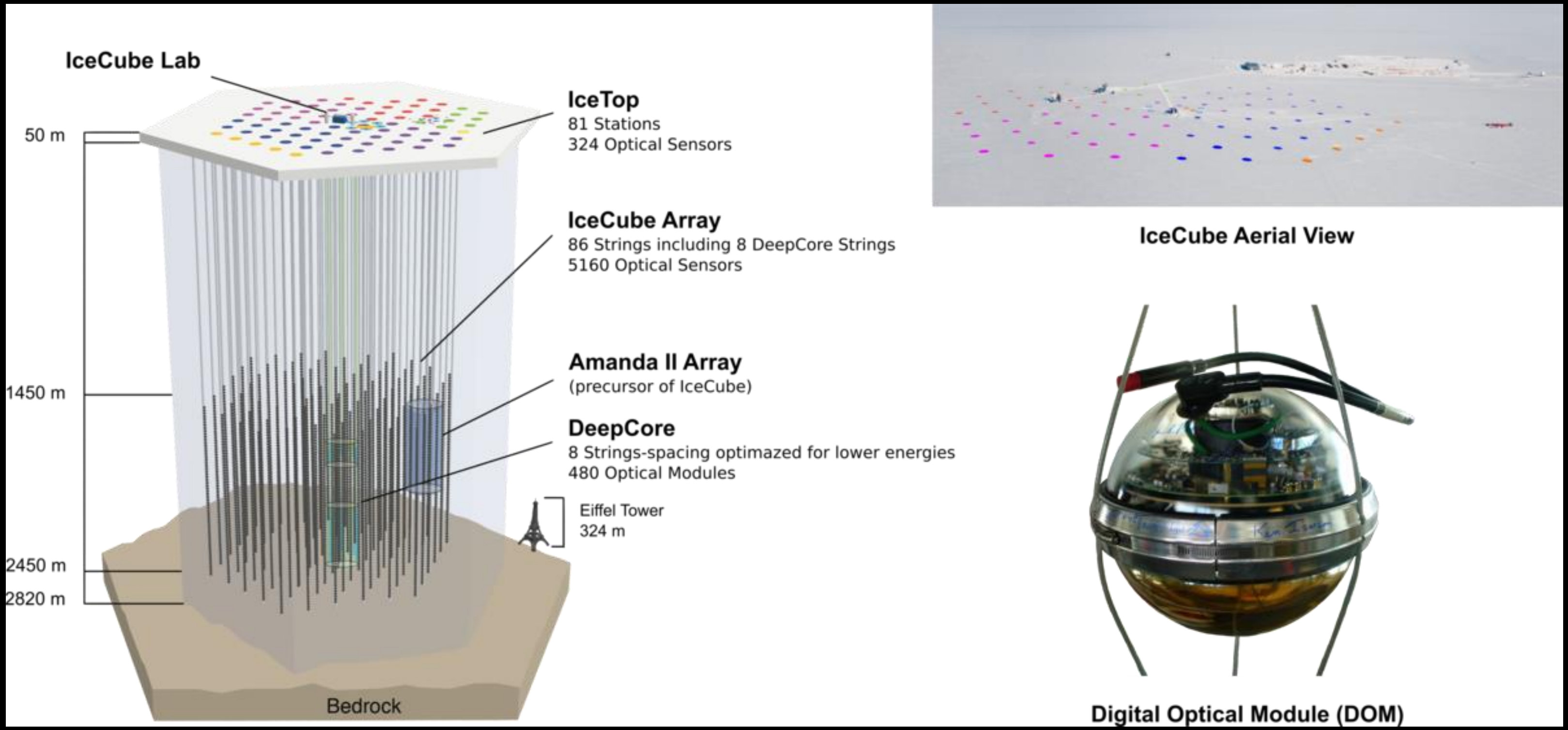
Michael Bogert
Advisor: Dr. Kurahashi

BRIEF OVERVIEW

- Background info
  - IceCube, neutrino events
- Purpose of my research
  - The problem, what I’m trying to accomplish, why it’s important
- Steps I took
- Results, discussion

ICECUBE

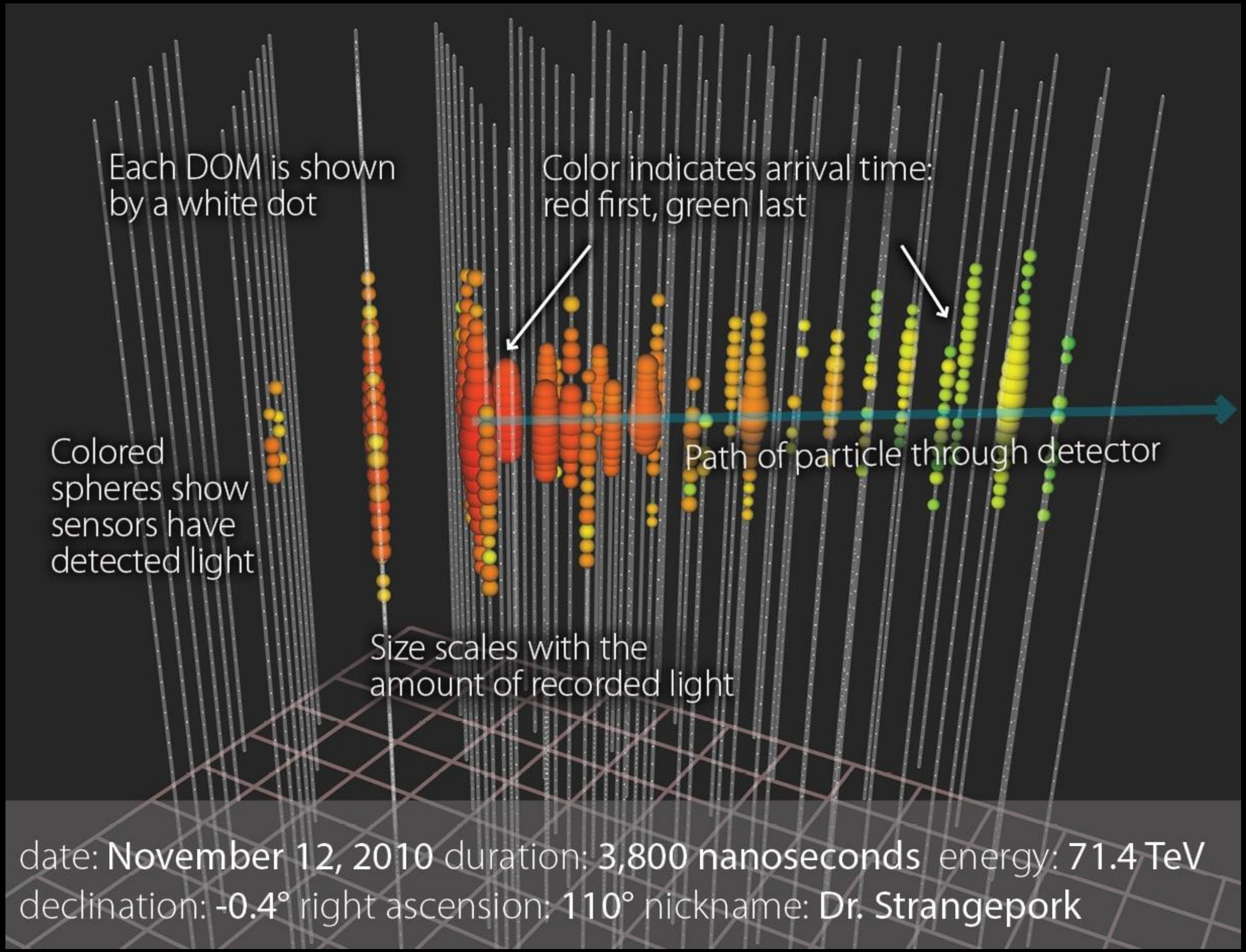
- Neutrino detection experiment
- 1 km^3 detector with 5,000 light sensors (DOMs)



Picture from https://www.iihe.ac.be/icecube

WHAT IS AN EVENT?

- Neutrino interaction recorded by detector
- Events are viewed in steamshovel, a visualization software

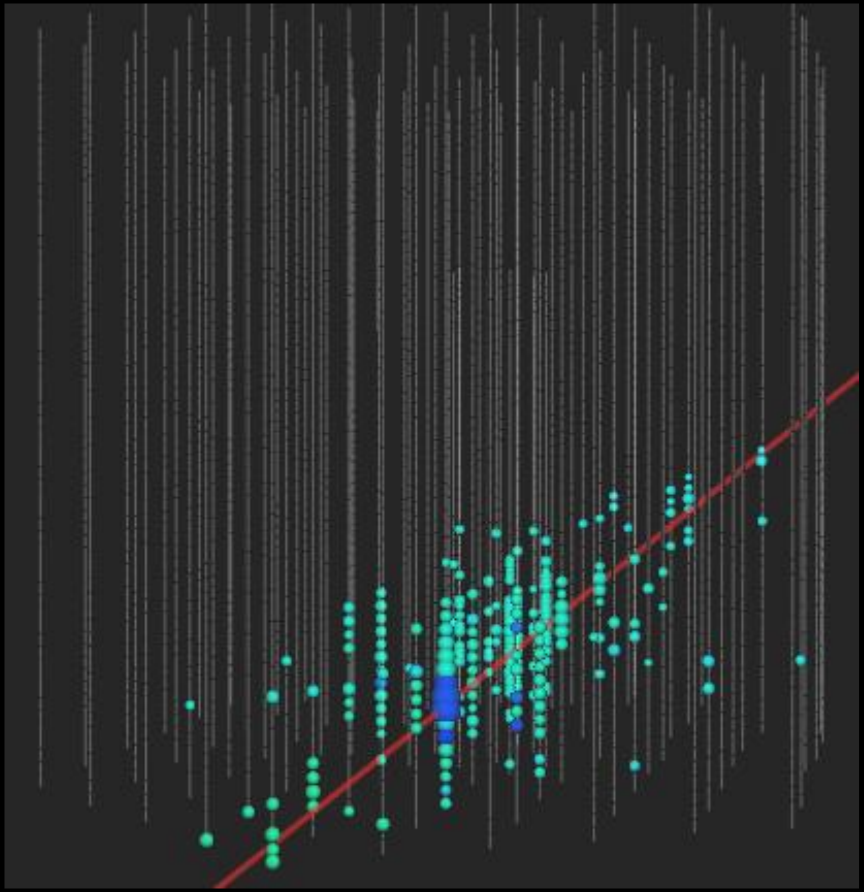


Picture from https://icecube.wisc.edu/science/icecube/

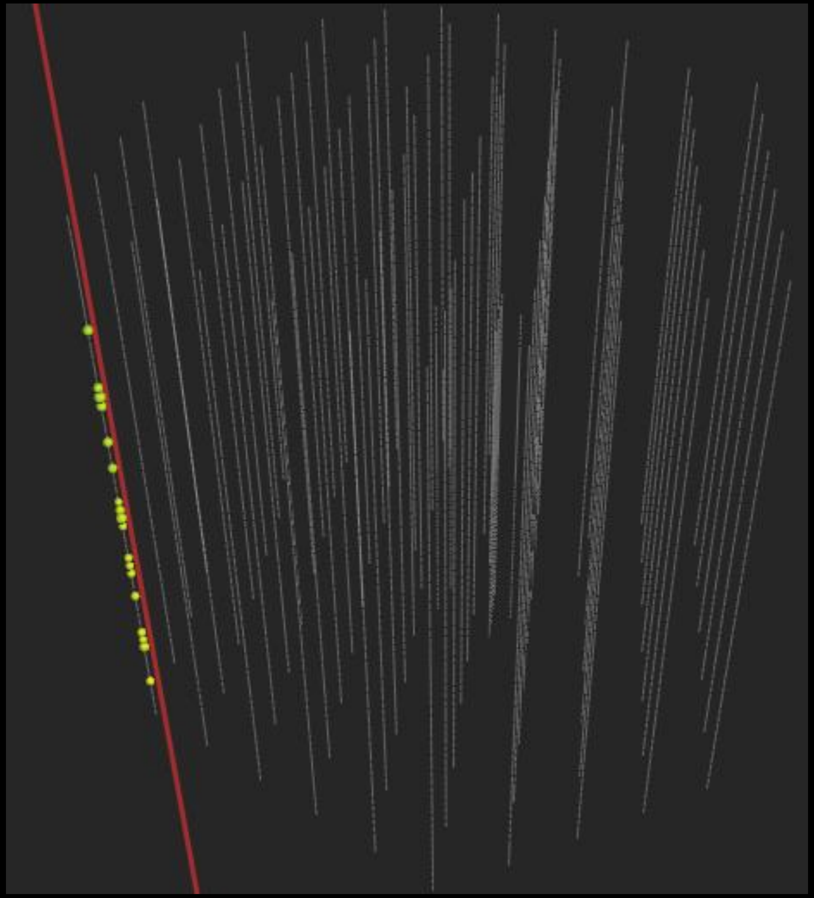
EVENT CATEGORIES

- Two categories: Normal events and corner clipping events
- Reconstruction of a corner clipper can't be trusted

Normal event



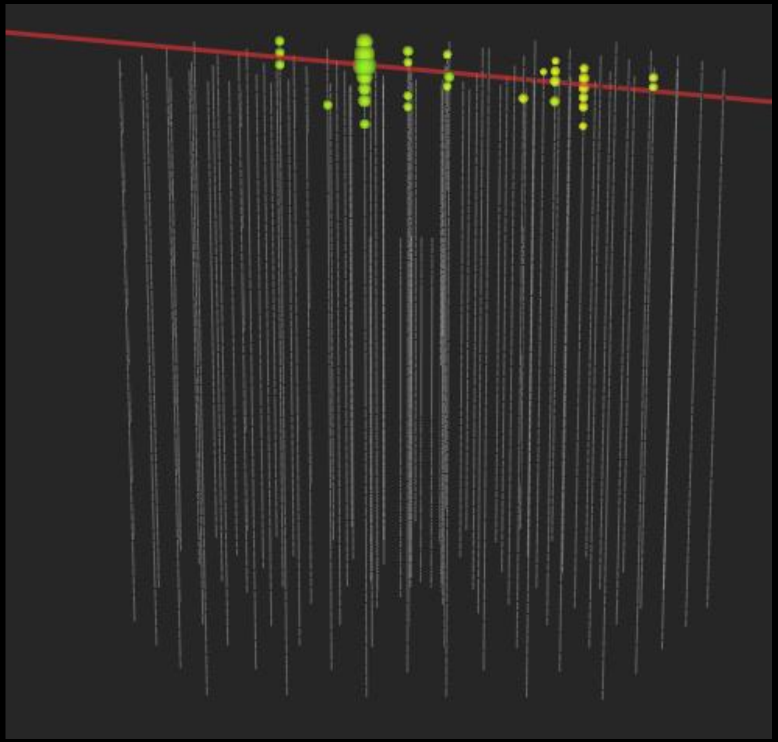
Corner clipper



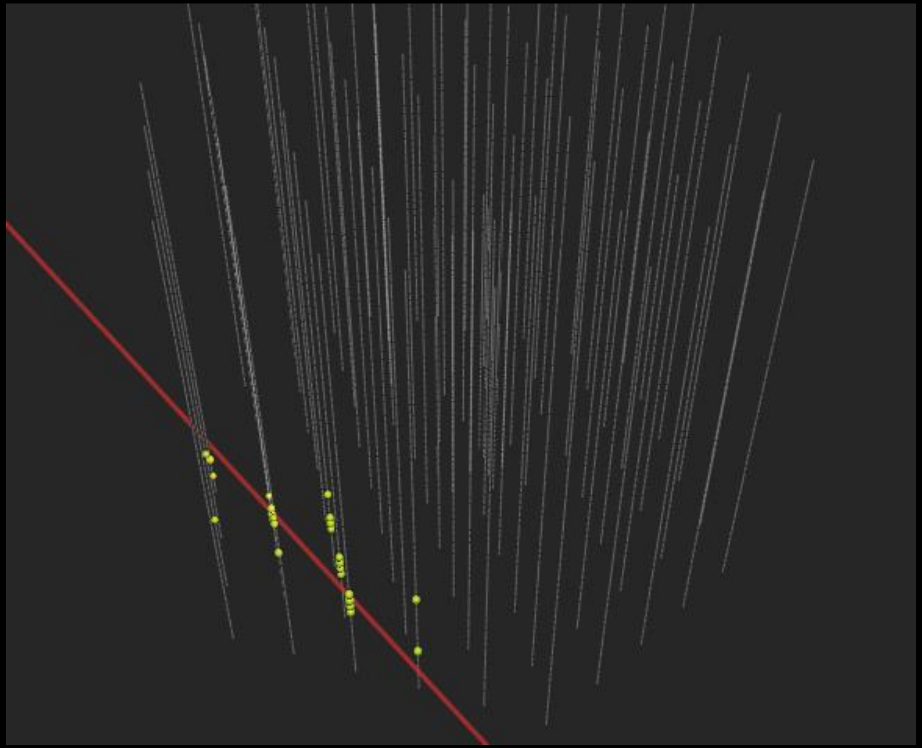
MULTIPLE TYPES OF CLIPPERS

- Corner clippers don't just have to be on 1 or 2 strings
  - Can clip the top/bottom of detector
  - Can cut through a side

Clip through top of detector

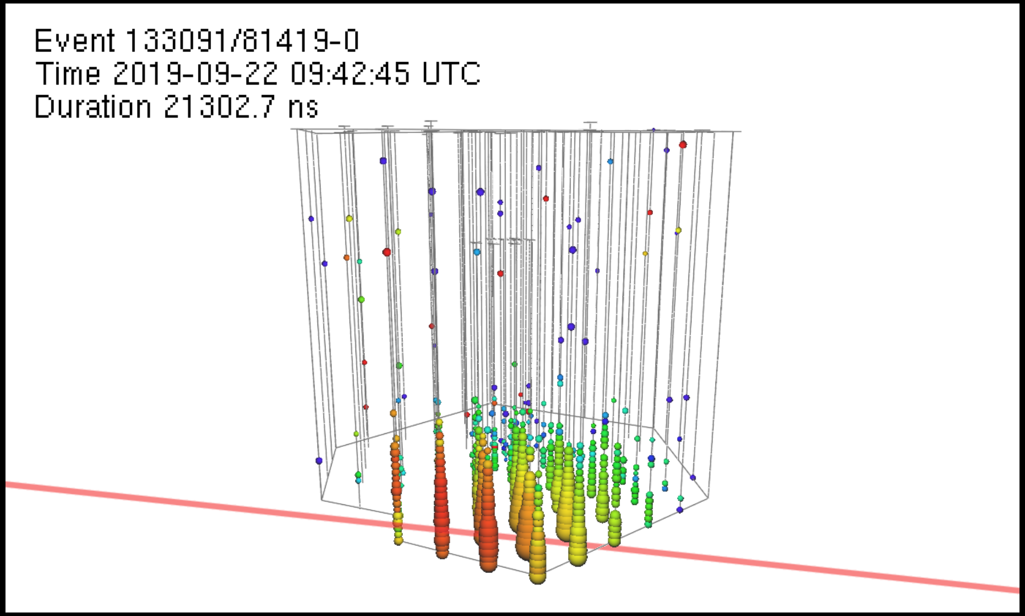


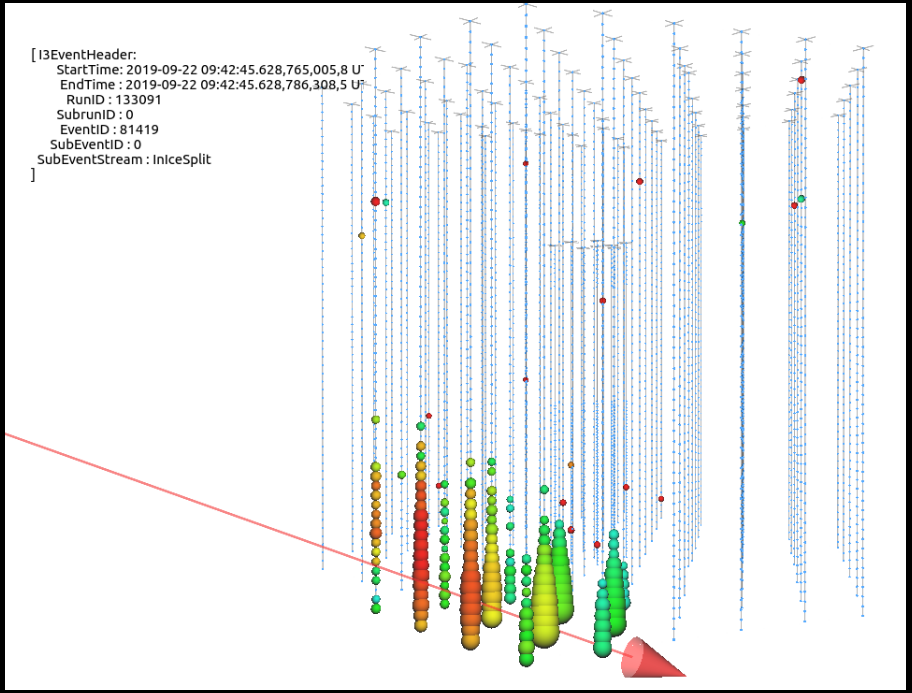
Clip through multiple strings



THE PROBLEM

- Detector constantly recording events (10^{11} events per year)
- No method to filter out corner clippers
  - Easily identifiable by eye, but difficult to define algorithmically
- Alerts are sent out for interesting events (> ~330 TeV)
  - Sent out automatically before visual confirmation, may be retracted



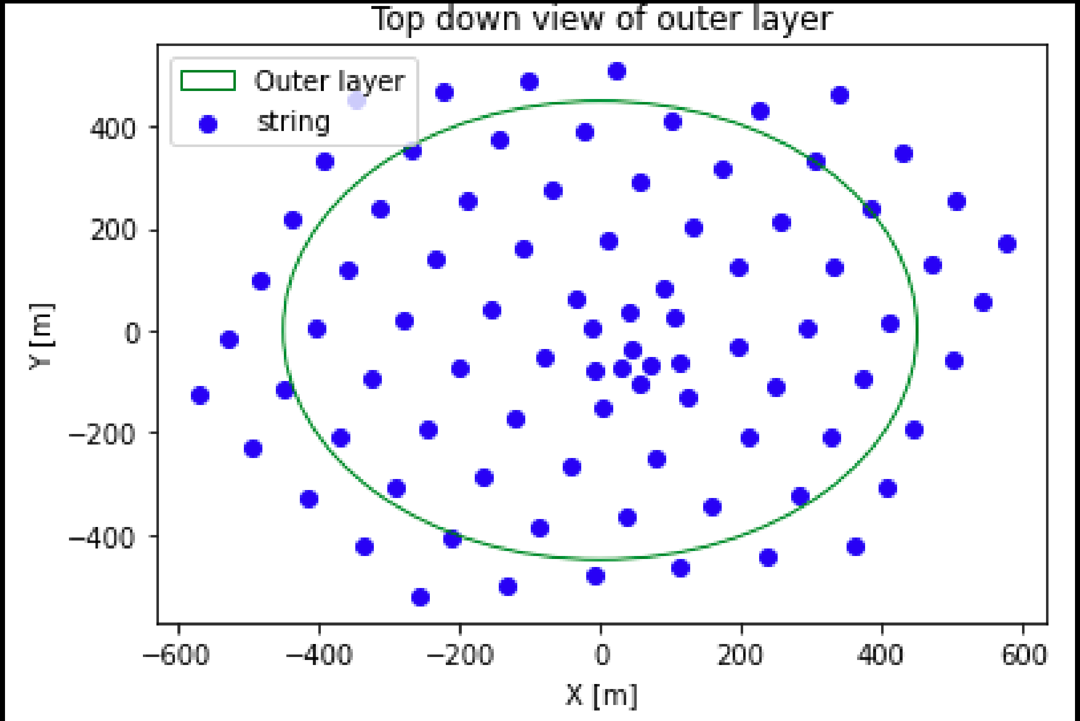


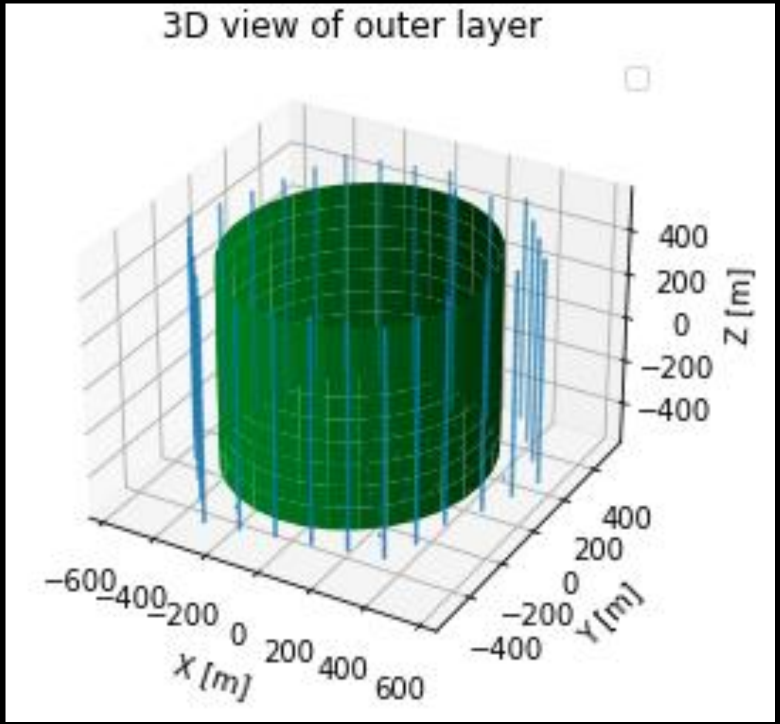
GOAL/WHY IT MATTERS

- Goals
  - Come up with a reliable way to classify events (normal or corner clipper)
  - Use machine learning to further improve classification
- Why it matters
  - Significant number of corner clippers (too many to go through by eye)
  - Reconstructing corner clippers wastes time/resources

INITIAL STEPS

- Get familiar with corner clippers
  - Come up with common features all corner clippers should share





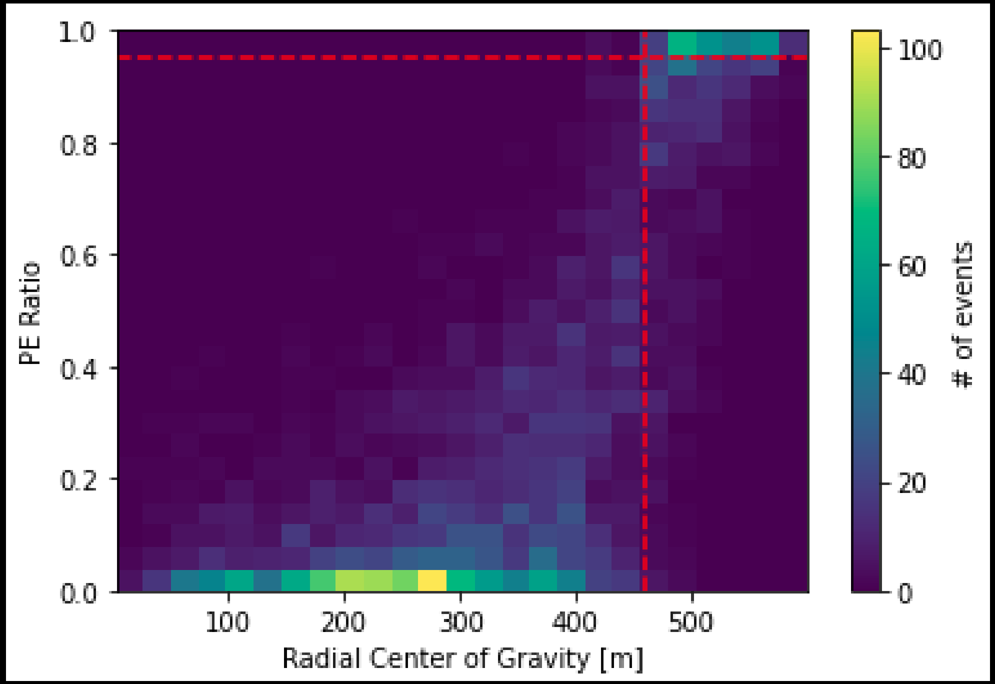
COMMON FEATURES

- Decided all corner clipping events should have a "radial center of gravity" far from the center[1]

$$COG(\rho) = \frac{\sum_{i=0}^{\#DOMs} q_i \, \rho_i}{\sum_{i=0}^{\#DOMs} q_i}$$

where q_i = # of photoelectrons (PEs) at the ith DOM, ρ_i is the radial distance

- Should also have high ratio of $\frac{\#\ Outer\ PEs}{\#\ Total\ PEs}$



STARTING POINT

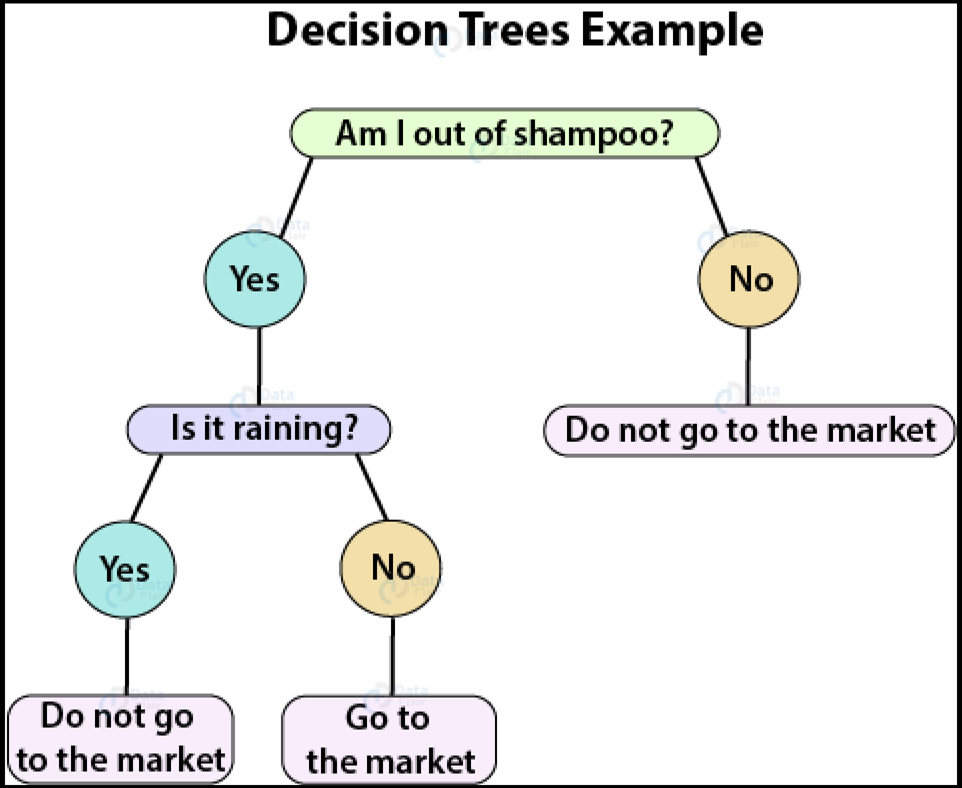
- Hypothesized cluster of events in top-right corner of 2D histogram were corner clippers – they were

- Could now use $COG(\rho)$ and PE ratio as features in machine learning algorithms

- Secondary features added
  - $COG(z)$ - vertical center of gravity
  - Total PEs – Corner clippers should have low # of total PEs

MACHINE LEARNING ALGORITHMS

- Machine learning algorithms are methods that will predict results based on input data[2]
- 4 categories: unsupervised learning, supervised learning, semi-supervised learning, and reinforcement



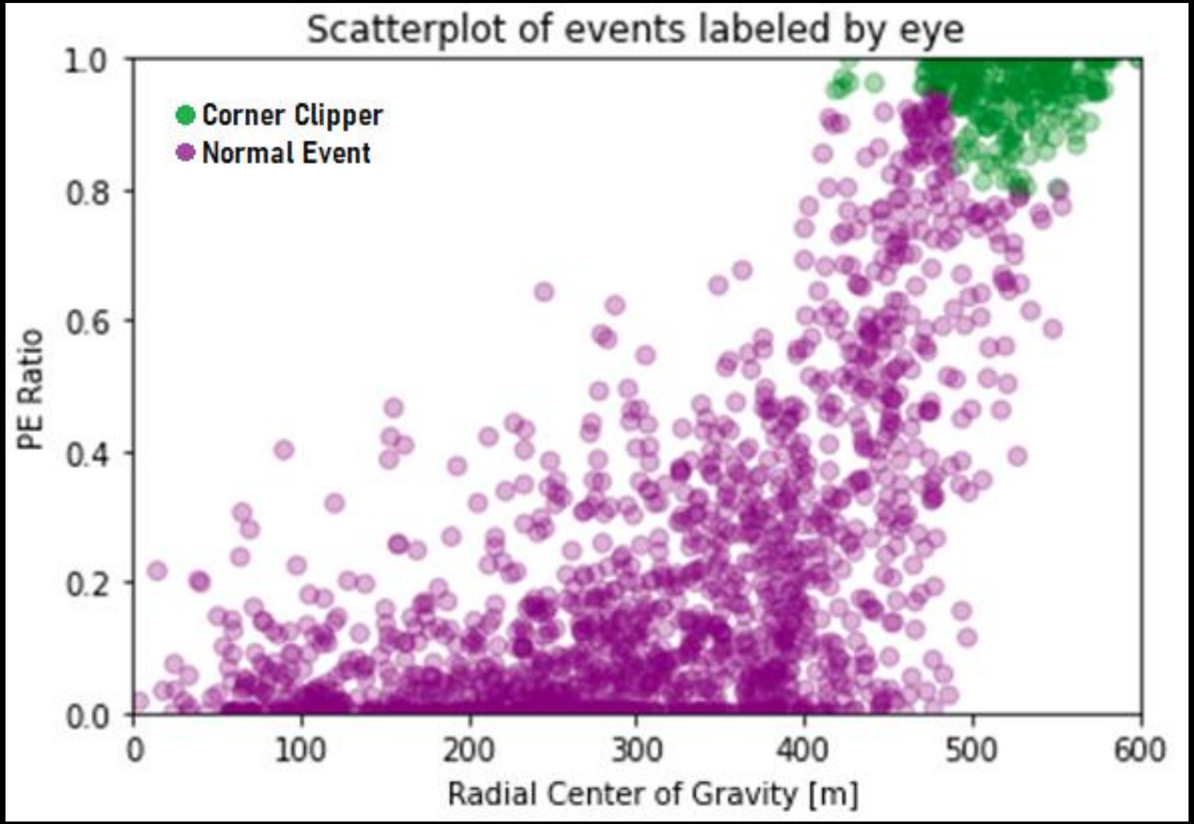
Picture from https://data-flair.training/blogs/r-decision-trees/

SUPERVISED LEARNING

- Learn by example (data is labeled)
  - Split data into training & test set: practice on training then predict on test
- Binary classification is a type of supervised learning that will label data as positive (1) or negative (0)
- Corner clipper = 1, normal event = 0

CLASSIFICATION METHODS

- Tried multiple classification algorithms after labeling data (3,517 events, 737 clippers → 21% corner clippers)
  - K-Nearest Neighbors
  - Decision Trees
  - Random Forests



- Made scatter plot to see locations
  - Green = corner clipper
  - Purple = normal

WHAT MAKES AN ALGORITHM GOOD?

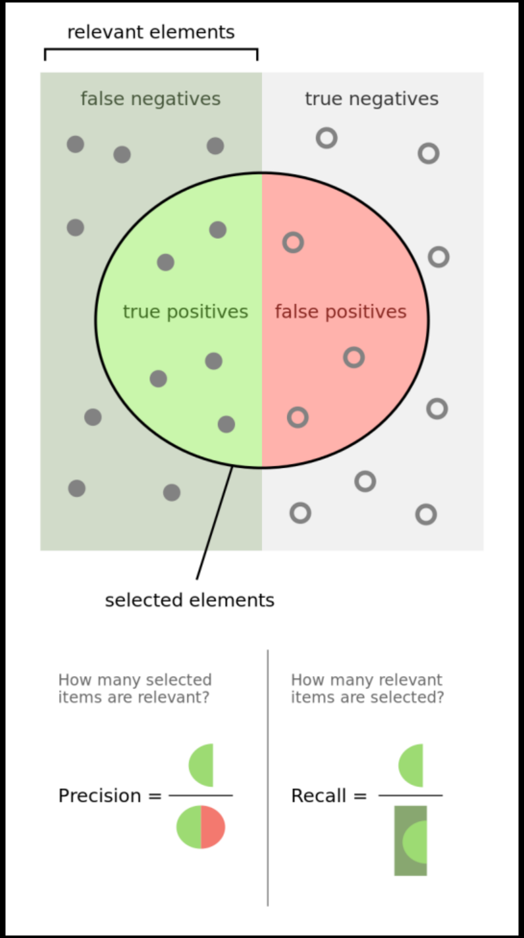
	Corner clipper (true label)	Normal event (true label)
Corner clipper (prediction)	True Positive	False positive
Normal event (prediction)	False Negative	True Negative

- Look at Completeness (recall) and Precision

$$Completeness = \frac{\#True\ positives}{\#True\ positives + \#False\ Negatives}$$

$$Precision = \frac{\#True\ positives}{\#True\ positives + \#False\ Positives}$$

- Prioritize completeness > precision (want low #FN)



Picture from https://en.wikipedia.org/wiki/F-score

ALGORITHM COMPARISONS

- Tried multiple algorithms
- Found the best by comparing precision & completeness values

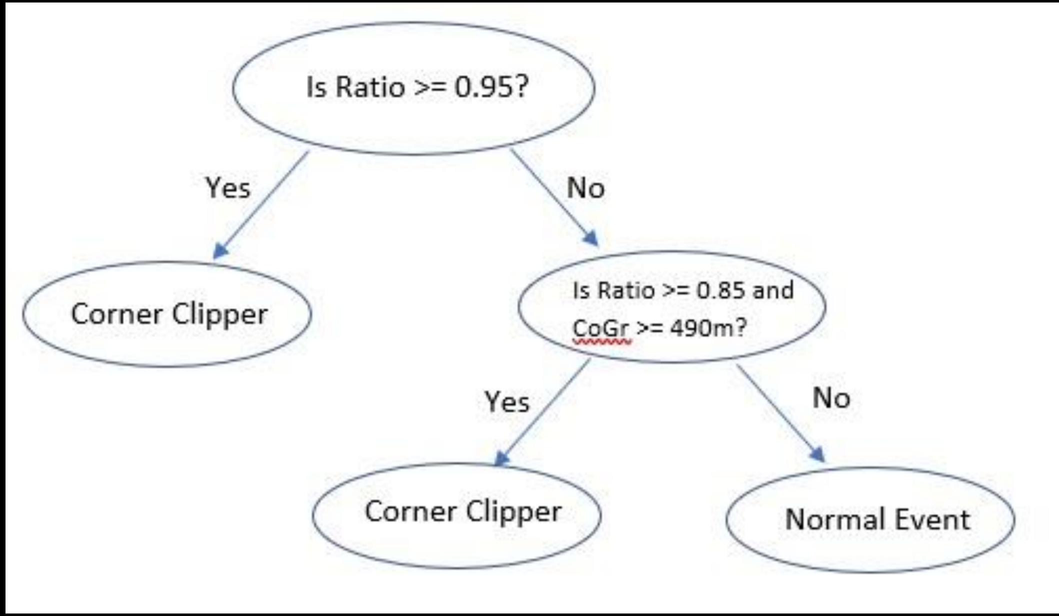
	Hard-Cut	K-Nearest Neighbors	Gaussian Naïve Bayes	Decision Tree	Decision Tree w/ Bagging	Random Forest	MPLRegression	Support Vector Machine
Precision	0.99	0.8	0.72	0.97	0.94	0.99	0.94	0.9
Completeness	0.67	0.47	.06	0.93	0.87	0.96	0.86	0.8

- Random Forest performed best

Info about the different algorithms can be found at https://scikit-learn.org/stable/supervised_learning.html and https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html

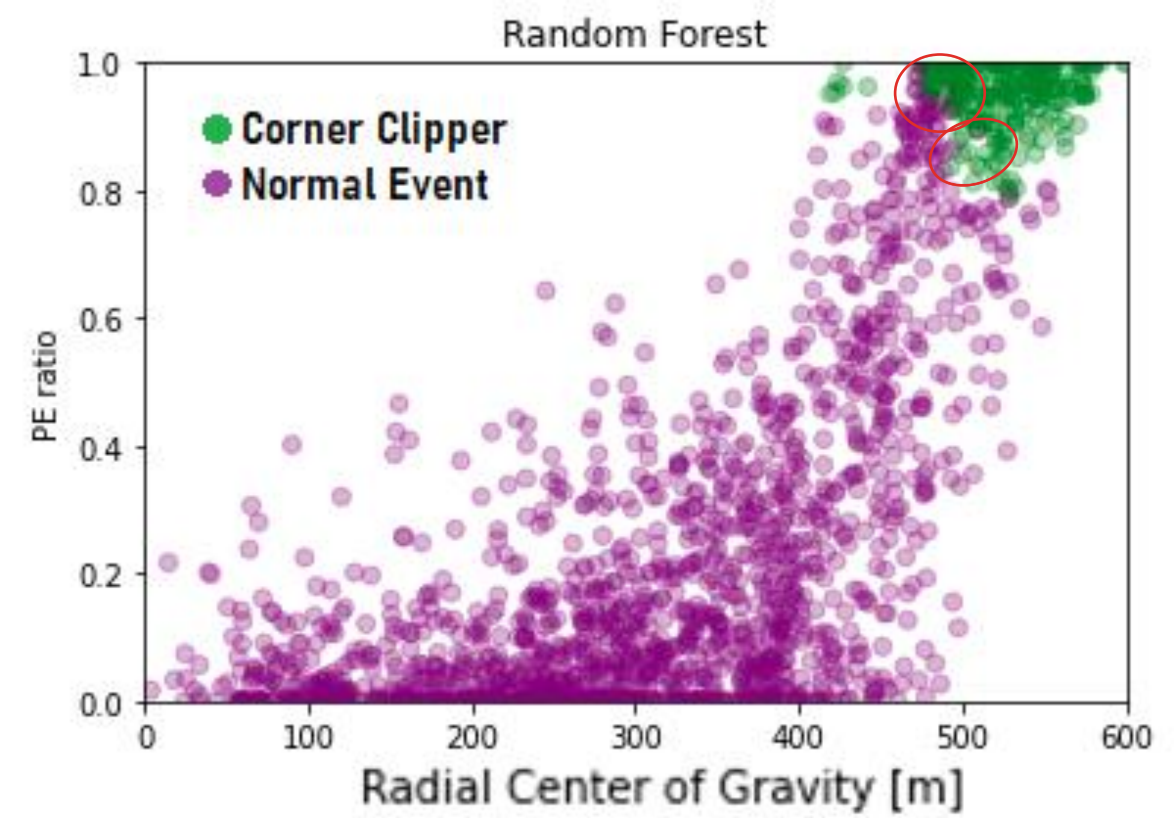
RANDOM FOREST

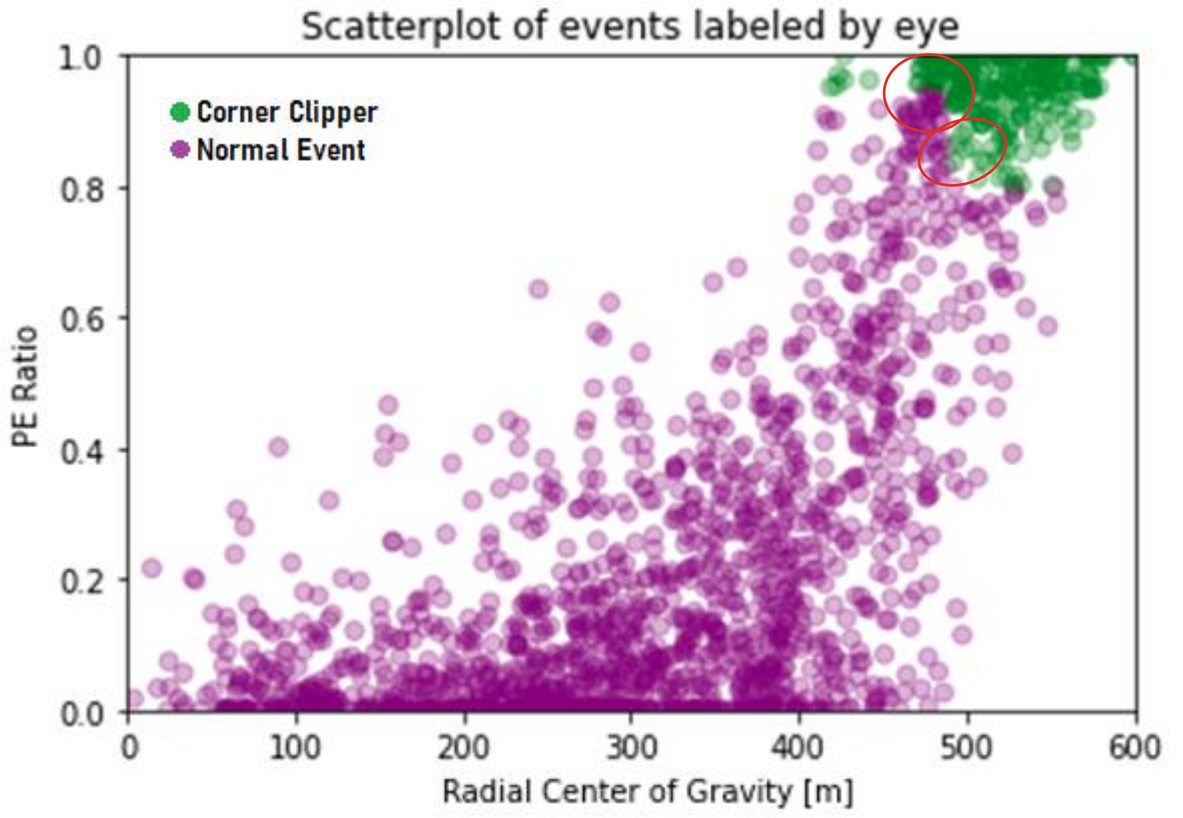
- Based off Decision Trees
- Creates multiple trees (makes a forest) w/ varying attributes and depth
- Final tree is made by combining results of all trees



CONCLUSIONS

- Algorithm works very well (1,759 test events , only 14 mislabeled → 99.2% correct)



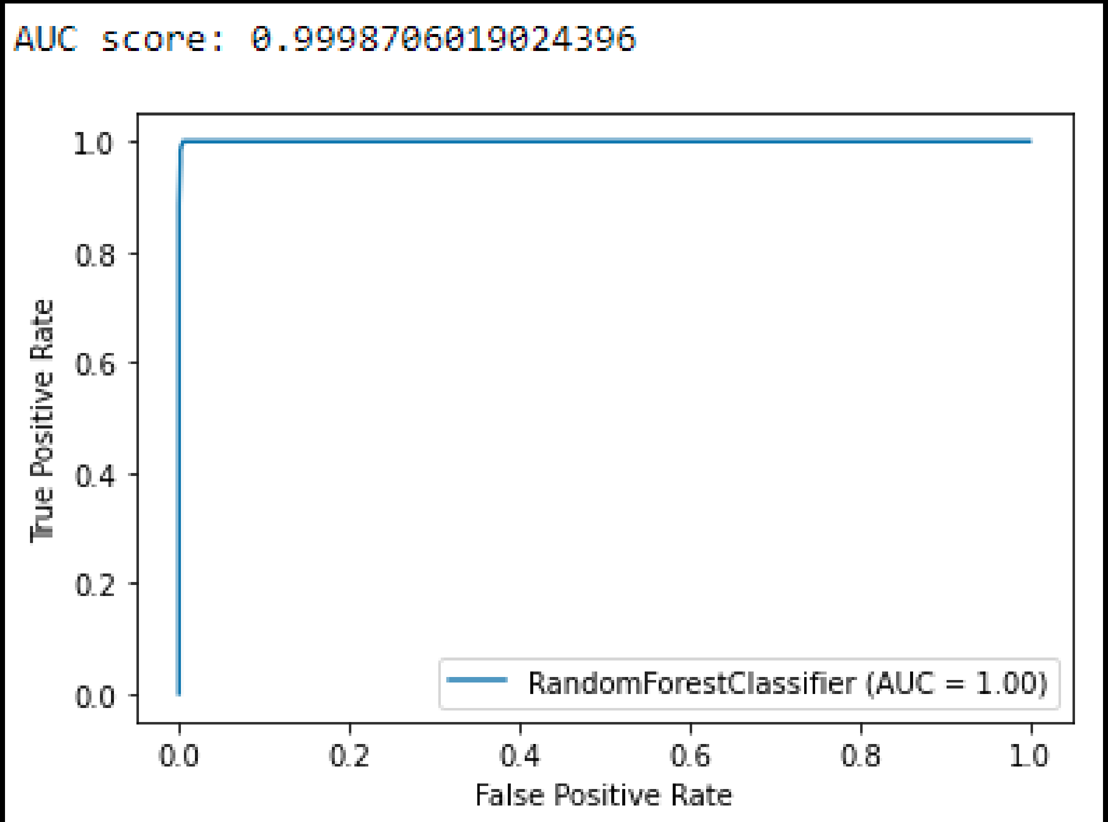


CONCLUSIONS 2

- AUROC Curve
  - Closer the value is to 1, the better
- Measures degree of separability[3]
- Very close to 1, another good sign

$$TPR = \frac{\#TP}{\#TP + \#FN}$$

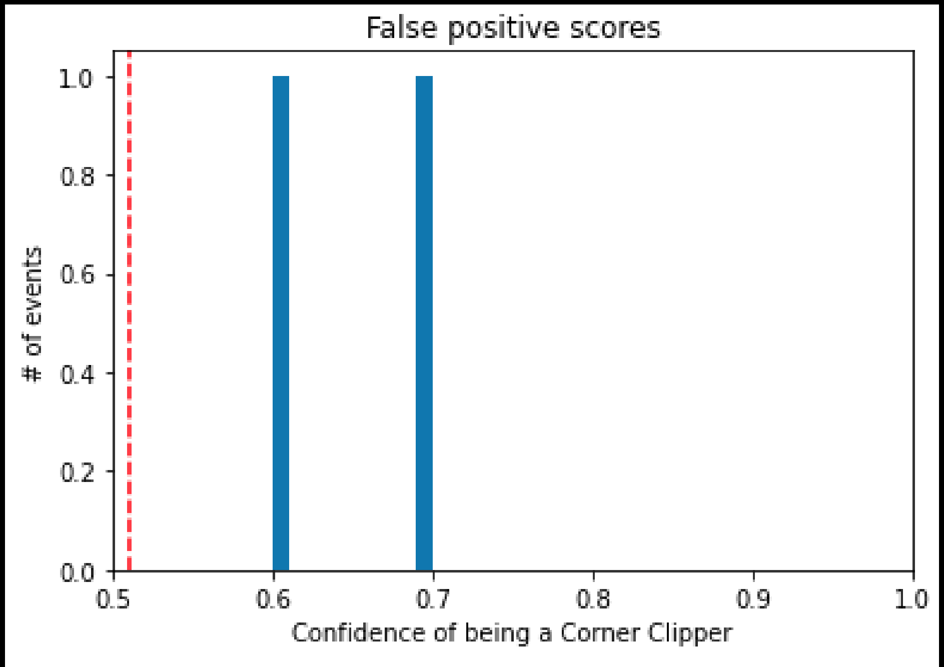
$$FPR = \frac{\#FP}{\#TN + FP}$$

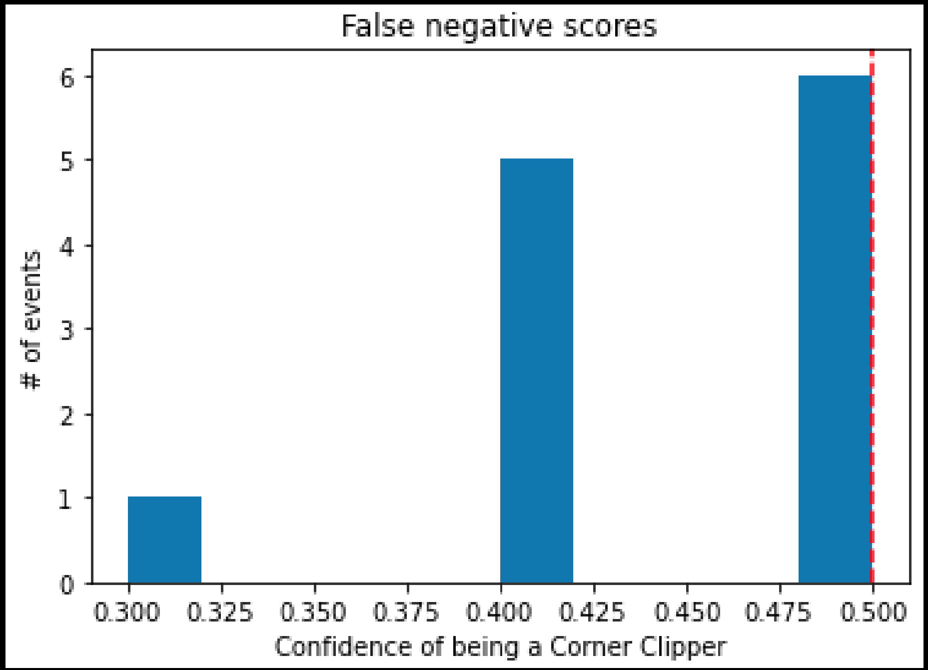


INCORRECT EVENTS

- How incorrect are they?
  - What type of event is hard to classify?

	Corner clipper (true label)	Normal event (true label)
Corner clipper (prediction)	True Positive	False positive
Normal event (prediction)	False Negative	True Negative

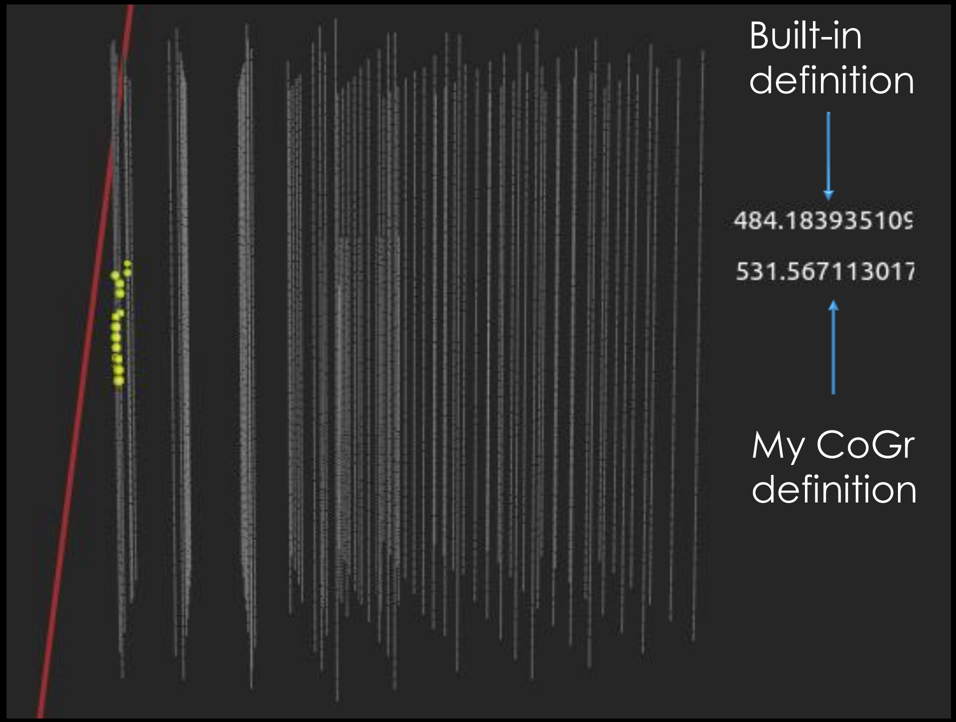




EXAMPLE OF A FALSE NEGATIVE

- Event is a corner clipper, but not labeled as one

- 2 different CoGr's could be the issue



FUTURE IMPROVEMENTS

- Overall, the method to identify corner clippers works very well
  - Can still be improved (fix CoGr issue)
- Replace Random Forest w/ faster algorithm
- Test on more/different event topologies

ACKNOWLEDGEMENTS

A HUGE Thank You to Dr. Kurahashi and her grad students Mike C, Mike K, Luna, and Steve!

You guys are the best!!

SOURCES

[1] CARVER, Tessa. Time Integrated searches for Astrophysical Neutrino Sources using the IceCube Detector and Gender in Physics studies for the Genera Project. Université de Genève. Thèse, 2019. doi: 10.13097/archive-ouverte/unige:120924 https://archive-ouverte.unige.ch/unige:120924

[2] Wakefield, K. (n.d.). A guide to the types of machine learning algorithms. Retrieved March 09, 2021, from https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html#:~:text=At%20its%20most%20basic%2C%20machine,values%20within%20an%20acceptable%20range.&text=There%20are%20four%20types%20of,%2Dsupervised%2C%20unsupervised%20and%20reinforcement.

[3] Narkhede, S. (2021, January 14). *Understanding AUC - ROC Curve*. Medium. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

Questions?