Prediction of Independent Fission Fragment Isomeric Yields using Machine Learning

Daniel Potemkin

Introduction:

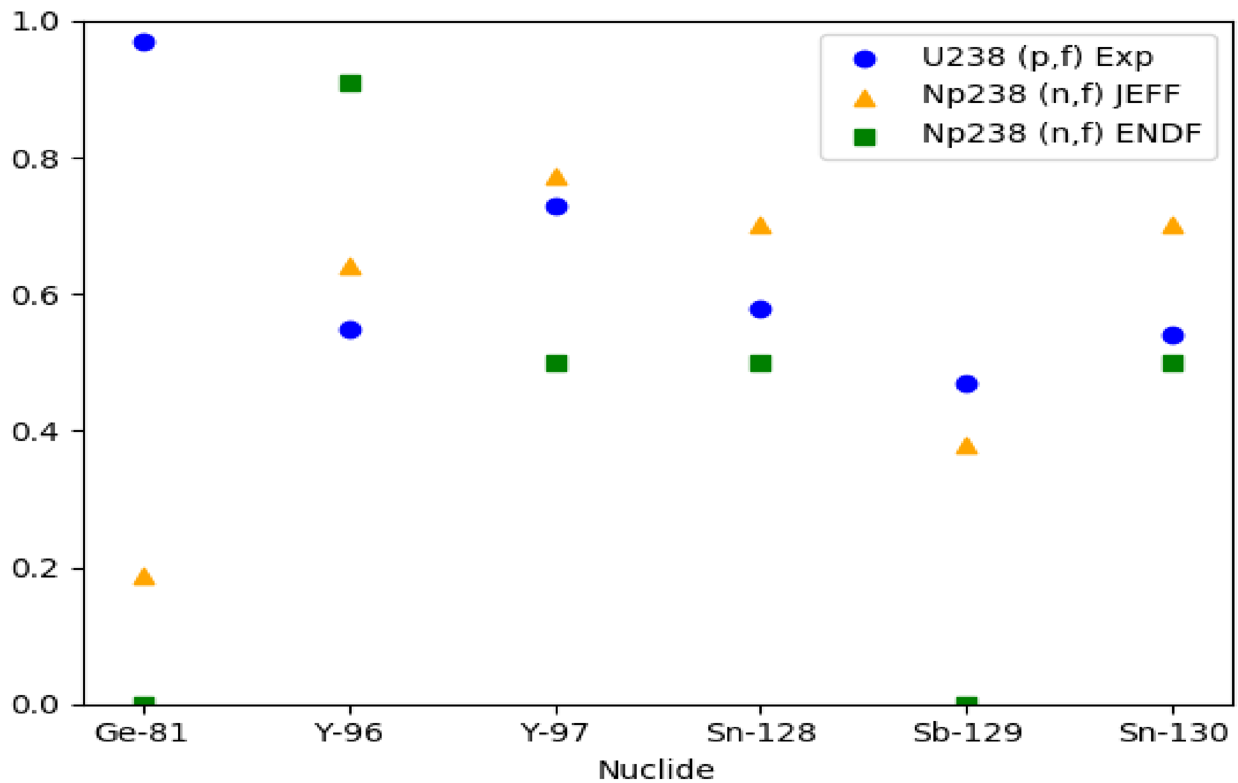


Figure 1: Comparison of ENDF and JEFF 3.3 Neptunium 238 yields, and experimental data for isomeric yield ratios of proton induced fission of Uranium-238. Both reactions result in the same compound nucleus [2][3][4] (Figure was created by author).

In the process of fission, a heavy compound nucleus breaks into two excited β-unstable fission fragments. These fragments first deexcitate through prompt sequential neutron and gamma emissions which result in independent fission yields. Figure 1 displays the independent isomeric yield ratios of the same compound nuclear system, Np-239, for several isomers. A compound nucleus is formed by the target nucleus absorbing the projectile. Isomers are excited nuclei that reach a metastable form, meaning that they remain in this excited state. These isomers are significant because they are the only source for post fission spin distribution data. Figure 1 shows that there is incomplete and inconsistent data within the ENDF and JEFF 3.3 evaluations compared to experimental U-238. Obtaining accurate data provides necessary information for determining the antineutrino spectrum. In beta decay, antineutrinos are formed as a result of the launched neutrons decaying through the weak force. The antineutrino spectrum (the number of antineutrinos per outgoing energy) has applications for nonproliferation and the understanding of the Standard Model of particle physics. Two of the largest contributors of uncertainty in the reactor antineutrino spectrum are the fission products Rubidium-92 and Yttrium-96. While

Rb-92 production has been studied in great depth, Y-96 has not [1]. Therefore, understanding the production of Yttrium-96 and its isomer will reduce the uncertainty in the antineutrino spectrum.

Machine learning is a tool in data analysis which enables a user to extract hidden relationships within data for classification or regression based tasks [6]. There are many forms of machine learning such as neural networks, classification algorithms, and reinforcement learning algorithms. These different forms of machine learning enable users to find new insights into data that were otherwise unable to be seen. This poses the following research question: Is there a hidden relationship in the features of independent isomeric fission yields that can be used to create a model? Analyzing the production of independent isomeric fission yields, there are multiple observables which could influence the yields. It is known that the spin and nuclear structure of the fission fragments and the fissioning nucleus dramatically influence fission yields in a way that is poorly understood. Therefore, it is hypothesized that a regression model should be able to extract the correlations to accurately predict the yield ratios.

Methodology:

Target Atomic Number	Target Atomic Mass	Compound Atomic Number	Compound Atomic Mass	Projectile

Incident Energy (MeV)	Fission Fragment Atomic Number	Fission Fragment Atomic Mass	Calculation Type

Figure 2: The list of observed parameters associated with each Isomeric Fission Yield Ratio. Each isomeric fission yield ratio contained these 9 features for the model to use to predict isomeric ratios (Created by author).

EXFOR, a nuclear database that contains every experimental nuclear reaction since the Manhattan Project, was used as the database to compile isomeric fission yield data for the machine learning model. Every isomeric ratio was recorded along with the set of features illustrated in Figure 2. These features were chosen because of the logical influence they would have on the isomeric fission yields. Calculation type is the method for which the isomeric ratio for the specific data entry was calculated with, such as the metastable over the ground state or vice versa. The calculation types were assigned a numeric value to reduce computational complexity in the machine learning model. Every entry was compiled together into a comma separated values file, or CSV, to easily extract in python. A python reader was constructed to extract the information from the CSV into a python data structure created to enable easy manipulation of the data conglomerate.

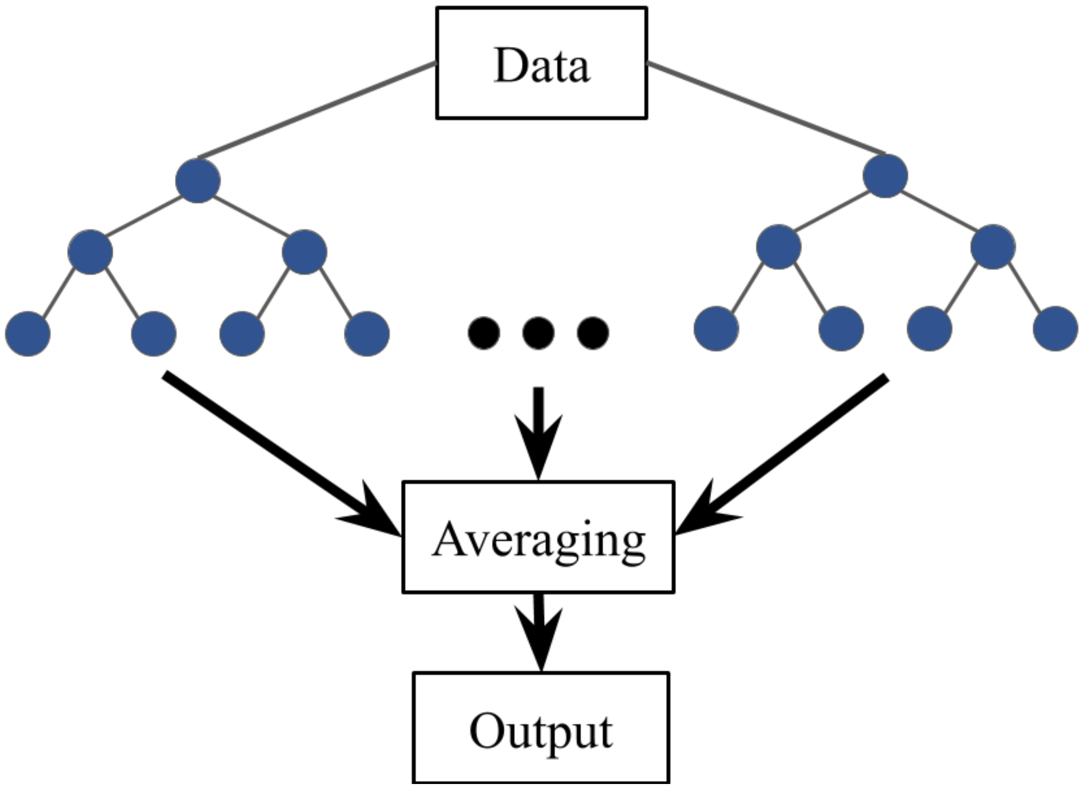
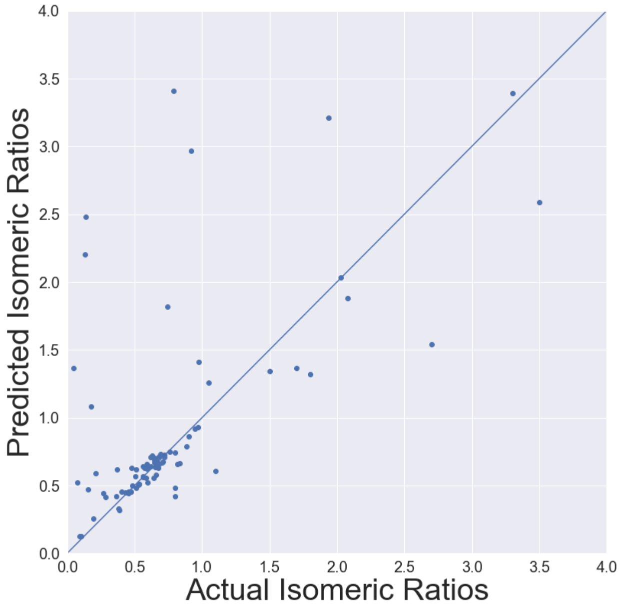


Figure 3: Visual illustration of the Random Forest Architecture and more specifically, how the random forest executes its predictions based on its decision trees and averaging process (Created by author).

Multiple machine learning models were attempted such as a backpropagation neural network and a Gaussian process regression, however, the random forest regressor provided the most accurate results. This model, as displayed in Figure 4, is a composition of decision trees which average the prediction. The training data being inputted undergoes a process called bagging where random samples of the data is fed into each decision tree. Then, the decision trees train on the random samples given. Decision trees are logical operators which does a multitude of comparisons within the data of the features in order to produce a prediction. Once training is complete, a set of features can be inputted into the model to be given to each decision tree to make a prediction. The predictions are then averaged to produce a final output. The random forest model used to predict isomeric yields was written in python using the scikit learn library.



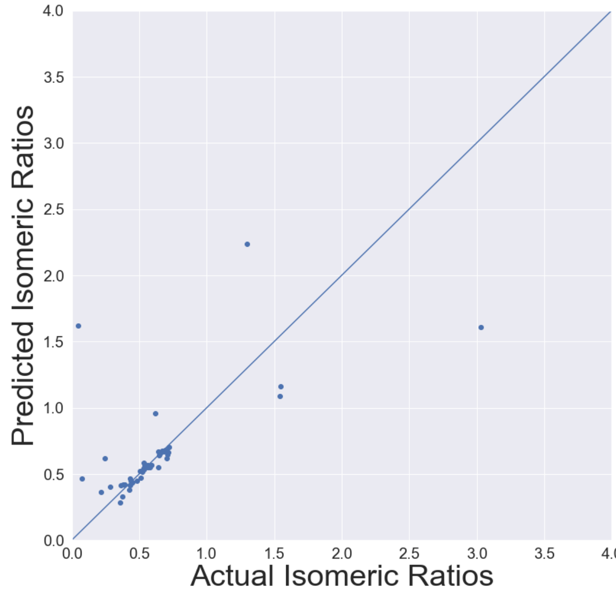


Figure 4: Comparison between unfiltered results for the Random Forest (left) and filtered results for low energy neutrons (right). There is an increase in accuracy of the model when restricting data specific to only low energy neutrons since its a single physical process that the model needs to abide to (Created by author).

The data entries compiled from EXFOR were filtered due to very specific entries containing observed parameters, or features, that aren't frequent within the data compilation and abide by different physical processes. These few entries were fission reactions that used heavy ion projectiles such as Fluorine and Carbon atoms as well as lighter projectiles such alpha particles and deuterium. Since these entries were in incredibly small amounts, they were filtered out of the final dataset since, to the model, these entries would not be justified. In comparison, although low energy and high energy neutron induced reactions abide by different processes, there are relatively large amounts of data for each type of reaction and therefore, the model would recognize this difference. Figure 4 displays an extreme example of why data filtration was done. To ensure a balance between accuracy and range, the dataset was limited to neutron, proton, and gamma induced fission reactions.

A machine learning model is generally trained on a subset of the available data called training data and its accuracy is tested on another subset of data called the prediction data. To create training and prediction data, the entries compiled were split randomly so that 70% of data entries were used as training data and 30% of data entries were used for prediction evaluations. The isomeric ratios, or labels, were separated respectively to the separation of their corresponding features. The training data was then used to train the random forest and each decision tree through the bagging method previously discussed. Once trained, the random forest model was given the prediction to formulate its own evaluations on each set of features. To evaluate the accuracy of the predictions, a histogram was created by subtracting the predictions from the actual labels to depict a distribution of error. A scatterplot was also created by plotting points with coordinates of the labels for their x values and random forest predictions for their y values. The mean absolute error was also calculated as well as the standard deviation for said metric.

Results:

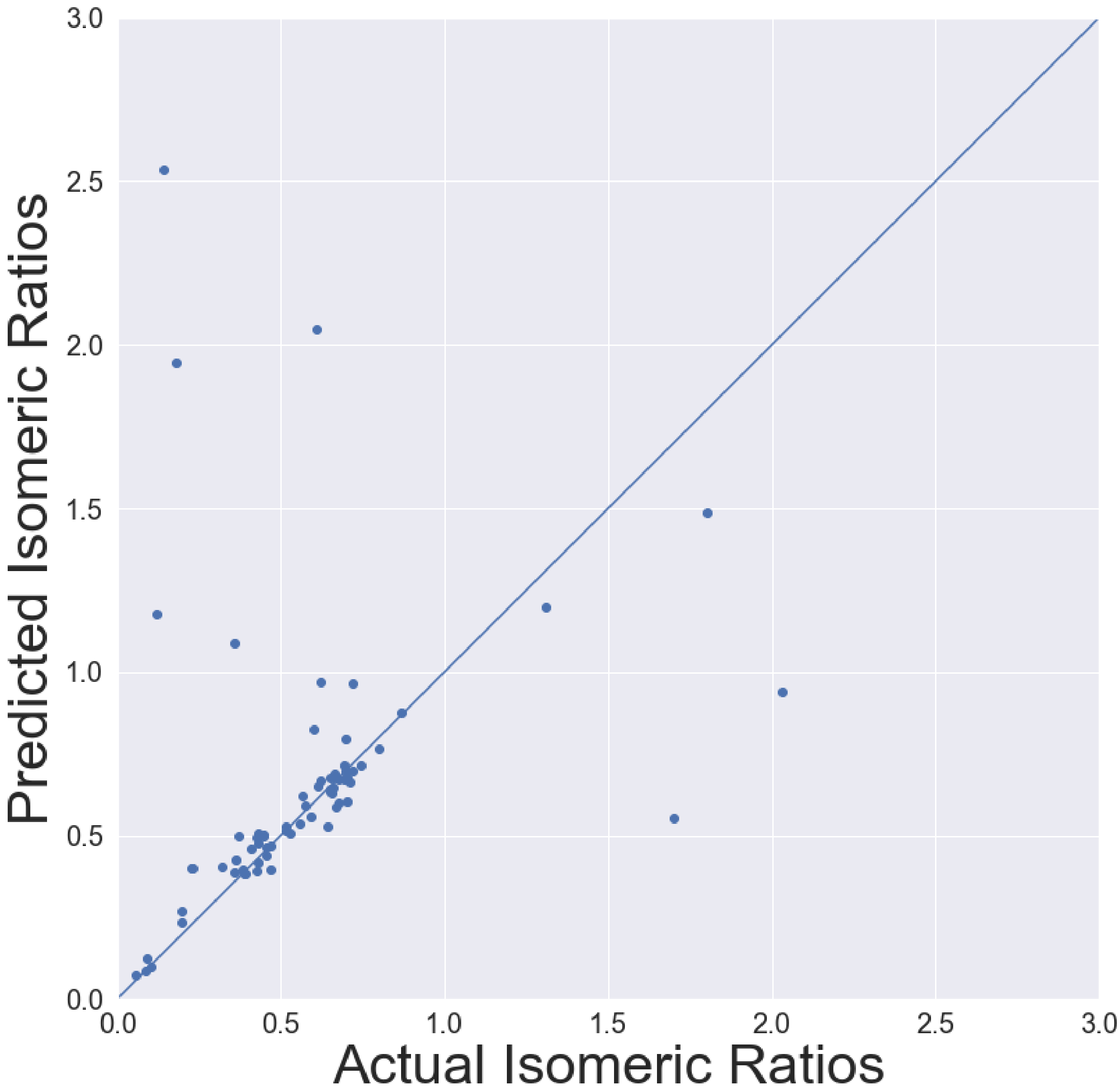


Figure 5: Graph illustrating comparison between delegated testing data and the model predictions on the given features for the testing data. Data closer to the $y = x$ line are more accurate (Created by author).

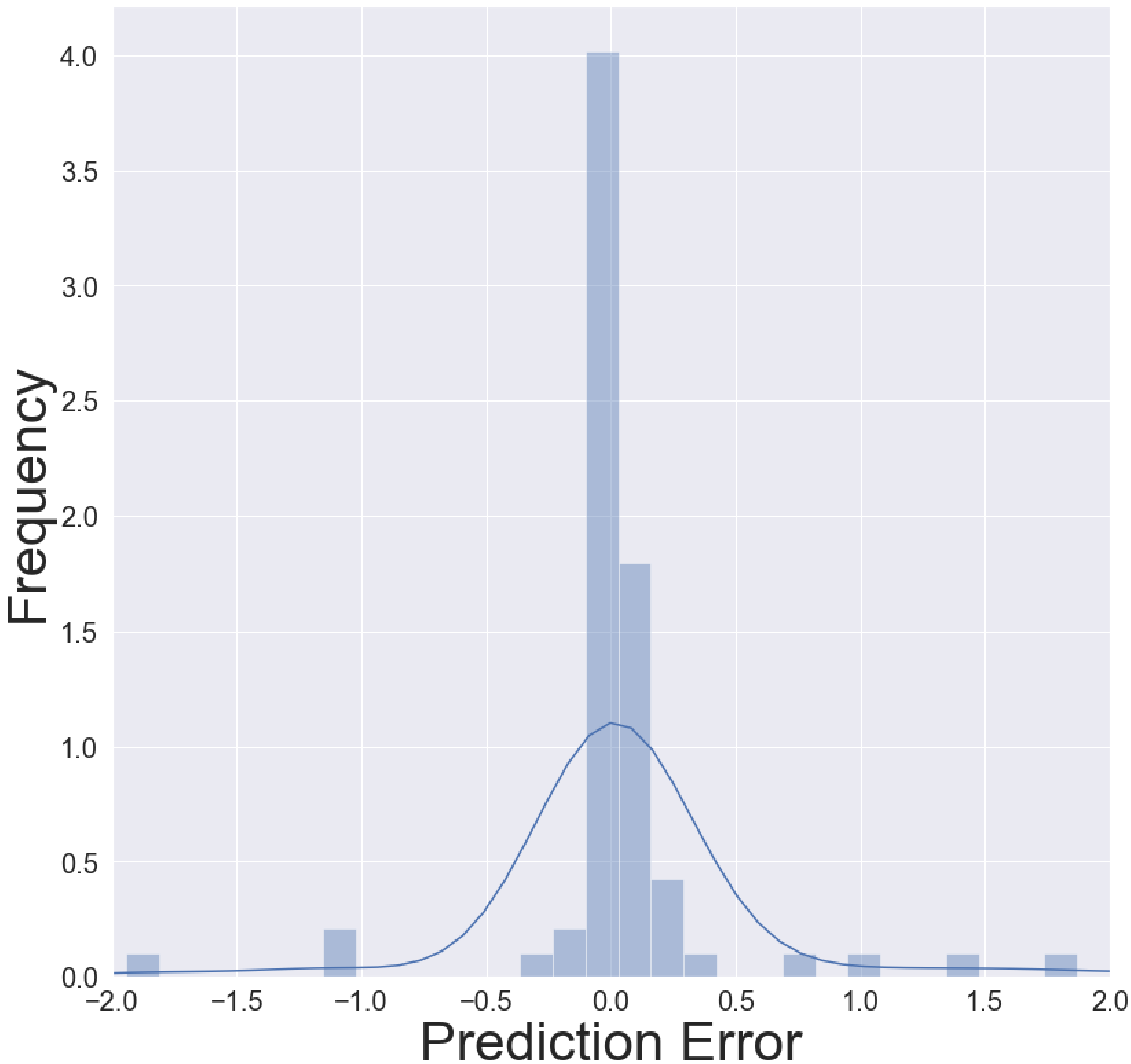


Figure 6: Histogram depicting the frequency of error of the predictions of the Random Forest model. A distribution of error is displayed over the histogram to illustrate the Gaussian distribution of error (Created by author).

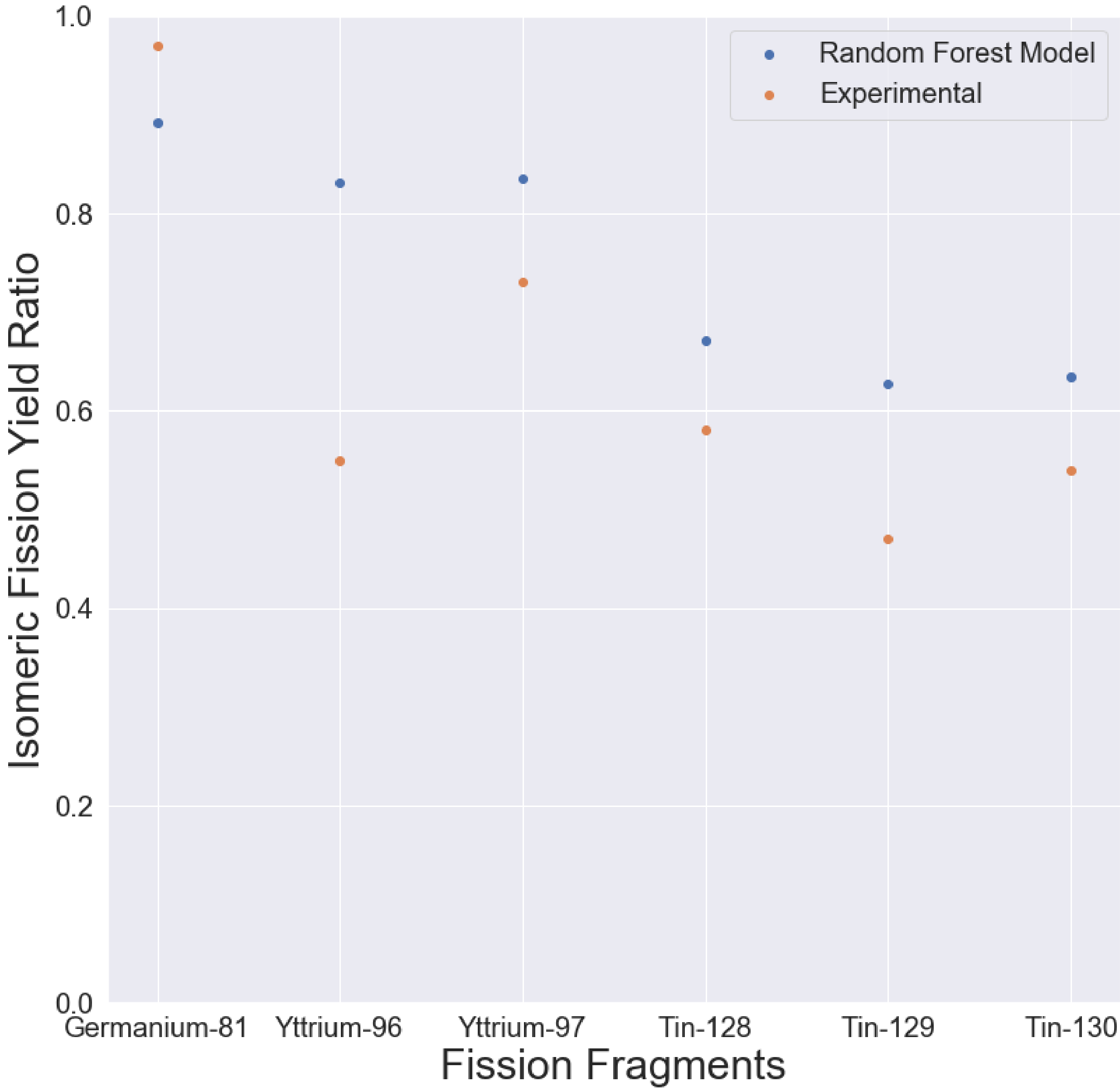


Figure 7: Comparison between experimental values of isomeric fission yield ratios from proton induced fission of natural Uranium with an incident energy of 25 MeV (Created by author).

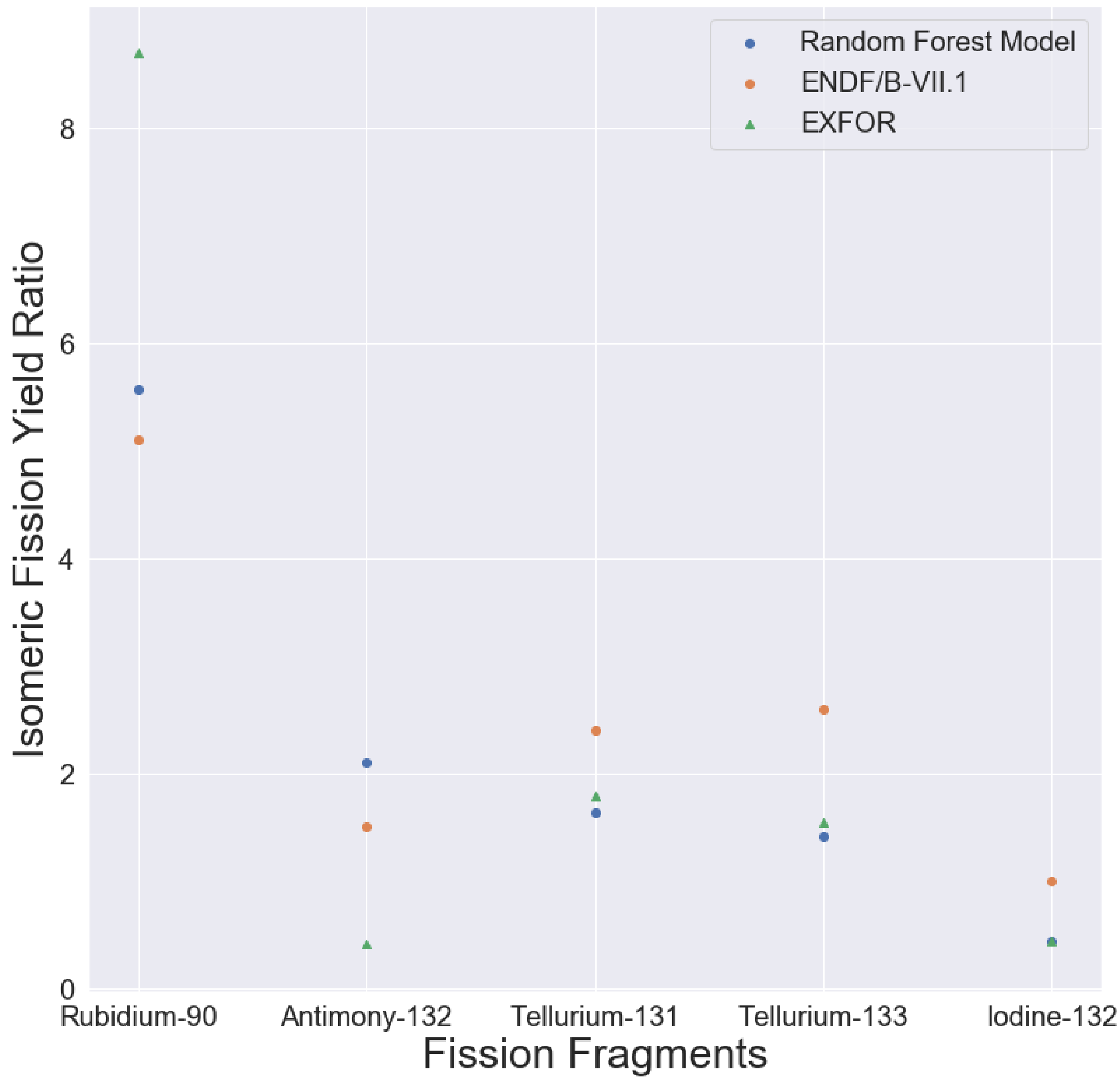


Figure 8: Comparison between ENDF evaluations and random forest model predictions of specific EXFOR data entries of neutron induced Uranium-235 fission at .0235 eV. The isomeric fission yield ratio for Rb-90 is metastable over total. The isomeric fission yield ratio for the rest of the comparisons are metastable over ground state (Created by author).

Figure 4 and 5 display the accuracy of the random forest model based on the delegated prediction data entries from EXFOR. Figure 4 demonstrates how close the different predictions are compared to their respective labels based on the distance from the y = x line. Figure 5 illustrates the frequency of error within the predictions of the model, their deviations from the actual labels. A Gaussian distribution centered at zero is superimposed on Figure 5 to show the distribution of error. Figure 7 provides a comparison between the experimental values of isomeric fission yield ratios in proton induced fission of Uranium-238 at 25 MeV. Figure 8 provides a comparison between the most modern ENDF evaluation, ENDF VII, the EXFOR data used in the model, and the model predictions.

Discussion:

Analyzing Figures 5 and 6, it is clear that for the most part, the random forest model can often accurately predict the isomeric yield ratio. There are, however, some inconsistencies in the predictions based on outliers that result from overestimations or underestimations. There are two possible reasons for the outliers. The first reason is that the specific EXFOR yields are inaccurate which would mean that the model could not predict the inaccuracies themselves. The second reason is that these specific outliers contain features that are not frequently observed in the training data. For example, the random forest model has lower accuracy when dealing with higher incident energies, a reflection of the fact that there aren't many data entries with high incident energies. In both cases, prediction error can be reduced by either producing more isomeric fission yield data through experimentation or reevaluating existing isomeric fission yield data, both of which refine the training dataset for the model to improve accuracy and consistency in prediction. Furthermore, it is important to note that the model currently treats predictions as final deterministic values rather than distributions. In future revisions of the model, implementing Bayesian methods into the model will enable predictions to be distributions which can reflect how certain the model is with the predictions relative to the features. For example, there would be a high confidence interval for low energy neutron induced fission as there is a lot of data for this situation while high energy gamma induced fission would produce low confidence interval to reflect the sparsity in the data entries for this situation.

Figure 7 and 8 display comparisons of the random forest model to experimental values and ENDF evaluations of isomeric fission yields. Again, there are some inconsistencies within the data due to the issues previously mentioned. However, the model is mostly accurate and in certain cases outperforms the ENDF evaluations relative to the EXFOR data entries used. In one case, Iodine-132 of Figure 7, the random forest model predictions is almost exactly the same as the EXFOR data.

Conclusion:

There is clearly great potential in the random forest model to provide highly accurate and consistent predictions as data entries are refined and increase in amount. Coming back to the research question, the question proposed was: Is there a hidden relationship in the features of independent isomeric fission yields that can be used to create a model? It was hypothesized that the features of the isomeric fission fragment yields should heavily influence the isomeric yields and therefore, using machine learning tactics, hidden relationships within the features can be extracted to create relatively accurate predictions within the data. Based on the results of the random forest model, it is clear that the hypothesis is supported. Aside from some inconsistencies due to sparsity or inaccuracies within the data, there are relationships that the random forest extracted which resulted in accurate predictions.

References:

1. Sonzogni, A. A., Johnson, T. D., & McCutchan, E. A. (2015). Nuclear structure insights into reactor antineutrino spectra. Physical Review C,91(1). doi:10.1103/physrevc.91.011301
2. Rakopoulos, V., Lantz, M., Al-Adili, A., *et. al.* (2017). Measurements of isomeric yield ratios of fission products from proton-induced fission on natU and 232Th via direct ion counting. EPJ Web of Conferences,146, 04054. doi:10.1051/epjconf/201714604054
3. Chadwick, M., Herman, M., Obložinský, P., Dunn, M., Danon, Y., Kahler, A., … Young, P. (2011). ENDF/B-VII.1 Nuclear Data for Science and Technology: Cross Sections, Covariances, Fission Product Yields and Decay Data. Nuclear Data Sheets, 112(12), 2887–2996. doi: 10.1016/j.nds.2011.11.002
4. (2017). Retrieved from https://www.oecd-nea.org/dbdata/jeff/jeff33/index.html
5. Machine Learning: What it is and why it matters. (n.d.). Retrieved from https://www.sas.com/en_us/insights/analytics/machine-learning.html.
6. Chakure, A. (2020, January 10). Random Forest and its Implementation. Retrieved from https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f