* Please submit any Python code used as a separate file in addition to a text document (Word) which provides answers to the below problems. Writing should be in complete sentences with enough detail to demonstrate a thoughtful consideration of what is asked

The included .csv files contains information about the naturally occurring elements. The goal of this homework is to build a random forest model to predict the melting point of an element based on the other provided properties.

1. Split the database, setting aside 10% for testing and train a random forest model on the other 90% to predict melting temperature. What are the RMSE of the model on the training dataset and the test dataset respectively? What inputs were most important for determining the melting point? Does this make physical sense?
2. Generate a series of random forest models increasing the number of trees from 1 to 100. How do the training and test RMSE error vary as the number of trees changes? Plot your results. What trends to you see? Why do you see the trends that you do? What is the best number of trees to use? Why? (Note: you do not need to create a forest for every number between 1 and 100, but you will need to consider that the same number of trees might produce different results depending on the random seed).
3. For the optimum number of trees found in part 2, perform a similar study varying the maximum tree depth between 1 and 20. Plot your results, and explain the trends that you see. What is the optimal maximum depth to use?
4. Use the RandomizedSearchCV function to explore the same space for number of trees and maximum tree depth (1-100 and 1-20 respectively). Use 5-fold cross validation. What parameters gave the best result? Do they agree with your results for parts 2 and 3?
5. How did the train and test RMSE change between the initial random forest (part 1) and the optimized one? How significant is this improvement?