In deciding what to write about this week, I wanted to gain a better understand and explain ensembles in the world of data science and machine learning. When you initially think of ensembles, you imagine a band consisting of multiple musicians coming together to create one sound rather than the sounds that come from simply one of the instruments. With this is mind, how does that look with data science? Author Evan Lutins explains, “Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.” Rather than using only one model to fit a data set unto to, you will be utilizing various models together to investigate a question you are currently presented with at the time. By using various models at the same time, we can then determine certain results from one model that could play a factor in another’s results. Author Lutins discusses how one could use a decision tree to answer certain questions first and dependent on that answer will employ a completely different model. The key take away that he mentions is how we no longer have to rely on using model for a particular situation, but that we can group various models to help us. You do not have to make the difficult situation of choosing one model and hoping that it works at the time.

In exploring the meaning of ensembles within machine learning, I came across various types of ensemble models. An interesting one is Bagging or BootStrap Aggregating. For this ensemble model, the process if performed as followed:

1 – Multiple bootstrapped subsamples are pulled from a sample data set

2 – A decision tree is formed on each of the bootstrapped samples

3 – An algorithm is used to aggregate over the decision tree to form the most efficient predictor

After think through this process, it sounds a like a team of data scientists are working together by individually completing their own tasks for it all to come together and create the best result. While one data scientist can complete this process, it is important to understand that at time it is maybe better to employ more than one to ensure that you obtain the most accurate form. But, like with every algorithm that is created for modeling, there are some advantages and disadvantages. With advantages, the big one is that the prediction results are going to be more accurate overall since you have various models coming together to prevent more error and bias. A second advantage is that ensemble models can be used to capture linear as well as non-linear relationships in different sets of data. (Ramzai, 2019) Now, moving unto the cons of bringing these models together, the first is that with more and more models; it can become much harder to understand the final result. Since it is harder to understand, it will be more difficult for business insights to drawn from them. Since I provided two advantages, it is only fair that I also explain another disadvantage and that it is going to take much more time and computation for ensemble model meaning that it is not something that we can rely on for immediate results based on constant input of data. (Ramzai, 2019)

Looking at the utilization of ensemble models, it truly shows how far machine learning and models can expand when they are united to understand and predict accurate results.

References:

Lutins, E. (2017, August 02). Ensemble Methods in Machine Learning: What are They and Why Use Them? Retrieved July 13, 2020, from <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f>

Ramzai, J. (2019, March 06). Simple guide for ensemble learning methods. Retrieved July 13, 2020, from <https://towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2>