# Decision Tree - Kevin

The code aims to predict a continuous target velocity from finite element simulation data. The goal is to map categorical variables such as material types and configurations with time and velocity to predict the target velocity over time. Decision Trees are useful when working with both categorical and numerical data. Decision trees also perform better with less dataset rows compared to a neural net model; however, it is worse at predicting non-linearities. They also provide better transparency than a neural net. Attempting to outperform the models in the paper I opted for a decision tree regression with a grid search cross-validation for parameter tuning. The grid search was initially tested with minimum samples per split, minimum samples per leaf, max depth, and minimum weight fraction per leaf (totaling 4,851,000 fits). Searching across the best parameters the model closed in on minimum split of 7 and minimum samples of 1. Max depth was best left at none and minimum weight at 0.0 as these were optimal values. The Decision Tree algorithm recursively partitions the feature space into axis aligned regions. Each portion attempts to minimize variance and splits the data such that the predicted and actual values are minimized at each branch. Using an 80/20 split and the same random state as the Neural Net model the mean absolute error was 4.83, the mean squared error (also used as the splitter criterion) was 227.46, and the R^2 of the regression was 0.992. Most interesting is the regression function has a relatively lower y-intercept than the other models (y=1.00x+0.78). The model does not directly outperform the models outlined in the paper. The other goal was to explore tuning methods to prevent under and overfitting. It would be interesting to explore which model would perform better if the FEA simulation had a much smaller element size with many more dataset rows. The decision tree is also useful to directly compare to the random forest model explored later in the paper.

A graph with blue dots and red line

AI-generated content may be incorrect.

A diagram of a network

AI-generated content may be incorrect.

Figure X : Decision tree with depth 18 to show complexity of simulation data and transparency