

A graph with blue squares

Description automatically generated

CLASSIFYING THE STATE

The classification of the State in the random forest takes the coordinates as the greater feature importance. It is expected because the state is a place in the US. The coordinates are followed by the ‘Install\_MW” which is due to the energy generation by State; every state has a different amount of energy generation (e.g. Texas is the greatest energy generatior in the US).

Solar ranks in 4th place, which relates to the number of solar installations in the US (the greatest)

Natural gas is 5th place. It represents the major energy contributor

A graph of a bar graph

Description automatically generated with medium confidence

On the other hand, when predicting the Primary source based on the features, Solar energy ranks as the greatest feature importance (supported in the number of powerplants, and it is followed by by natural gas and Hydro (WHY?)

Despite the coordinates, the model doesn’t have much difference in its prediction (differs to the state prediction