DS 740

Final Project

**Introduction:**

Over the past 20 years, the U.S. Army Corps of Engineers Rock Island District Navigation and Operations section has been tracking dredging events within the Mississippi River and Illinois Waterway. Dredging and the maintenance of a navigable channel is one of the key missions of the Rock Island District. One of the challenges with meeting the demand for dredging, is that dredging is often a reactive activity that occurs following a survey that identifies an area of shoaling. To best allocate resources towards the efficient operation of dredge plants, a better understanding of dredge material volumes along the river is crucial. The purpose of this project is to see if river gage forecasts can be used to predict dredge volumes within the river.

The data used includes USACE Rock Island Districts internal dredging and survey datasets, coupled with historic river gauge readings at the date of the survey as well as 7 days prior and 14 days prior to a given survey. After cleaning, the dataset for this project contained 6,533 observations and 159 variables. Due to the highly dimensional nature of the dataset, PCA was used to first explore the variables followed by xgBoost to develop a model to predict dredging quantity volume from river gage forecasts. These forecast windows are often the best data available for dredge coordinators to base dredging and survey decisions prior to a grounding.

**Data Preparation and Cleaning:**

To prepare the data for analysis, missing values were interpolated for the river gauge readings. In order to keep the interpolation window small, an average interpolation was only completed if an empty reading was between two observed readings. Any other NAs were omitted since it was harder to interpolate between multiple-day gaps as the median would not have been reflective of the state of the river. In addition to interpolation, the survey and dredge data was filtered for the intended window of analysis from April 1999 through September 2021 to match the resolution of the gage readings. Additional filters were used to ensure the data was for the Illinois River and the Mississippi River as well as had pools within Rock Island District, the data also was filtered for routine surveys as well as routine channel dredging as that was the focus of the project. Once the data was cleaned, gage data was joined to the survey and dredging data. An additional join was completed for observations from 7 days prior to the survey and 14 days prior to the survey to represent a forecast window. While methods like xgBoost can handle gaps in the data, PCA cannot which is why the values were interpolated. After doing some final checks for NAs, a “Season” function was created to classify the data into seasons to support interpretation.

**Principal Components Analysis**

Map, scatter chart

Description automatically generated PCA was used to see if there were any discernable clusters within the variables of the data. With over 100 gages across two river systems, we hypothesized that there would be a potential cluster between the two river systems. There were no obvious clusters between the observations based on the full dataset and when visualized across the different metrics of “Season”, “Pool”, “River”, and “Dredge Type” (Fig. 1). There does seem to be some form of grouping along PC1 of “Dredge” to “No Dredging” points, however the observations appear to be randomly distributed as there still is some overlap within the PCA space. The loadings do seem to have a split between the Mississippi River Gages and the Illinois Water Way Gages which could explain PC2, however; since PC2 only captures 13% of the cumulative variation, this isn’t a very strong signal within the data.

Fig 1.

After seeing the split between the Mississippi River and Illinois Water Way Vectors, I decided to split my data along those river systems and see if we get a better separation of the clusters. (Fig 2)

**xgBoost**

**Conclusion:**

Main-channel river gage observations and 7 day and 14 day forecasts are poor predictors of dredging volume on their own. This approach was the first of many in an ongoing project to use machine learning and deep learning models to predict dredge material volume within the Mississippi River and Illinois Waterway. Future work will focus on bringing in additional predictors, like tributary river gages, which can capture incoming sediment loads to the system, and measures of river geomorphology to better constrain and develop a predictive model.