Jennifer Ruiz

DSC 680

Professor Williams

Project Three Presentation Q&A

1. How did you settle on your methodology?
   1. The goal of the analysis was determining the factors that best predict increased risk of malignancy. This involved classifying tumors based on their characteristics. I chose to use a random forest model as well as a gradient boost model in combination with feature decomposition. The idea was to test which model performs best.
2. If you had to do the project again, what would you do differently?
   1. If I were to approach with project with the information I have learned, I would have extended the analysis to include the age and ethnicity of the patient. I believe these factors would provide additional insights that could be important in diagnosis and treatment.
3. How can the insight from this project be applied?
   1. The insight gained from this project can aid in the understanding the most important tumor features in predicting malignancy.
4. Who was your target audience?
   1. The target audience for this project are physicians, researchers, and breast cancer focused organizations.
5. What was your methodology for visualization?
   1. My approach to visualization was to use it as a tool to aid in understanding the story within the data. For this project, I chose to exclusively use Python for creating visualizations. I employed a variety of libraries and techniques to provide insight and value to the project.
6. What is one area of weakness in your analysis?
   1. The biggest weakness in the analysis is
7. What are some alternate methods to approach is problem?
   1. I tailored my approach to classification and prediction. However, this could have been approached using an ensemble method or strictly as a regression problem.
8. What have you learned from this process?
   1. This project strengthened my understanding and comfort levels with feature decomposition. It has allowed me to use new modeling methods that I have had little exposure to previously.
9. What are some areas for follow up analysis?
   1. Follow up analysis should patient focused factors and medical history for a more insightful picture.
10. What other data sources could be added to strengthen future analyses?
    1. Data sources regarding patient outcomes such as survival, tumor re-occurrence, and metastatic growth could also be considered.