Jennifer Ruiz

DSC 680

Professor Williams

December 20, 2020

Project One Presentation Q&A

1. How did you settle on your methodology?
   1. This was a classification problem, which contained both binary and categorical data. A random forest model provides the versatility to handle classification tasks with mixed data types, without the need for extensive transformation.
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 find additional data regarding cardiovascular disease to incorporate into the analysis. With this being topic of great importance to public health, more insight is valuable to treatment and prevention.
3. How can the insight from this project be applied?
   1. The insight gained from this project can aid in the understanding of the factors contributing to cardiovascular disease. By providing education about lesser known controllable risk factors, individuals can work with their medical team on better prevention.
4. Who was your target audience?
   1. The target audience for this project is the medical community and researchers who can further explore the insights gained.
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. I chose to use Power Bi for creating the visuals for EDA. I wanted to quickly show trends or areas of interest within the data.
6. What is one area of weakness in your analysis?
   1. The biggest weakness in the analysis is that many data points were from individuals who had cardiovascular disease with a non-reversible heart defect. This makes it more difficult to gain insight about risk factors no associated with the defect.
7. What are some alternate methods to approach is problem?
   1. This problem could have been approached using a regression model or a neural network model. Both methods would have required extensive data transformation and produce convoluted outcomes that are difficult to interpret.
8. What have you learned from this process?
   1. This project strengthened my understanding of random forest models and how to interpret their results. I also strengthened by EDA skillset with multiple applications including R and Power Bi.
9. What are some areas for follow up analysis?
   1. Follow up analysis should be conducted on factors such as fasting blood sugar rates, heart defects, and blood vessel involvement.
10. What other data sources could be added to strengthen future analyses?
    1. Future analysis should focus on factors such as ethnic background to explore differences in risk factors among different groups. This will help shape prevention campaigns and target those at most risk for cardiovascular disease.