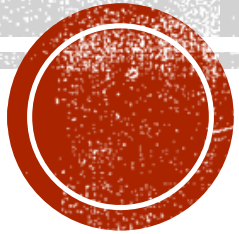
DATA 606 CAPSTONE IN DATA SCIENCE

A VIEW ON HCV DATA

SRIHARSHA DAPARTI

UX46295

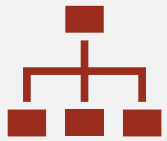
Instructor : DR. ERGUN SIMSEK



LITERATURE REVIEW

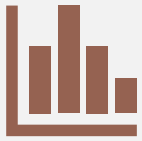
- The research on how to tackle the presented problem and the quest for similar problems and the methodologies used to solve those problems was challenging. The primary paper1 that used this data set was thoroughly examined along with the purpose of the research, approach and methodologies. The paper talks about coming up with 'Rules' which are combinations of two or more symptoms observed. The intention of this project is to take a more broadened approach. In this paper, the expected findings are the study of each individual symptom correlated with the test results.
- Many other works mentioning health data were also examined and a classification approach is decided to be the best way to tackle this question.

DATA CLEANING AND WORK DONE



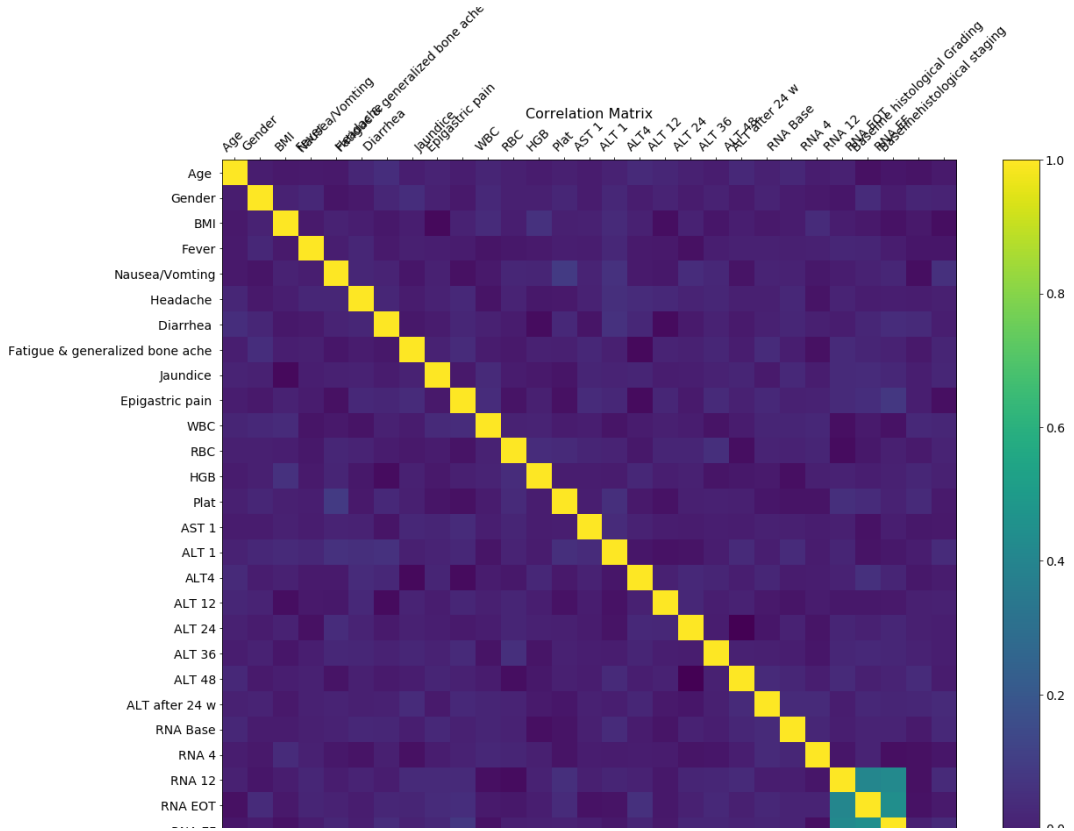
At the end of about three weeks from the project decision, a repetitive and considerable amount of cleaning and viewing has been done on the data set for it to be fit for further analysis.

Discretization was a major chunk of the problem, Discretizing the attributes of the data set one by one, due to their varied nature has been an issue.



The correlation between different attribute was visualized using a color coded Correlation Matrix plotted using ‘matplotlib’ package

EARLY RESULTS



- This correlation graph has shown that RNA levels at the end of 12 weeks and at the End of Treatment has shown to be playing an important role.

REFERENCES

1. Nasr, M., El-Bahnasy, K., Hamdy, M., & Kamal, S. M. (2017). A novel model based on non invasive methods for prediction of liver fibrosis. 2017 13th International Computer Engineering Conference (ICENCO).
2. Kelman, C. W., Bass, A. J., & Holman, C. D. J. (2002). Research use of linked health data - a best practice protocol. Australian and New Zealand Journal of Public Health, 26(3), 251–255.
3. Patton, G. C., Coffey, C., Sawyer, S. M., Viner, R. M., Haller, D. M., Bose, K., … Mathers, C. D. (2009). Global patterns of mortality in young people: a systematic analysis of population health data. The Lancet, 374(9693), 881–892.