

Individual Contribution Statement – Benjamin Hermans – Group 10

As requested, to accompany the group report, the following paragraphs outline my contribution to the project, what my achievements were and what I have learned from the course.

My substantial contributions to the project began with much of the initial research and wider reading into the topic area of our project, finding example papers who perform similar experiments on either the same data or different datasets to share with the group. These were then taken forward, and even further research was performed by various members of the group surrounding these papers. This is where we initially encountered methods of Multiple Imputation and other concepts that we would go on to incorporate in our process. From here I wrote a substantial amount of the underlying code for our project, including encoding the data into the numerical format we could use for exploration and modelling. I produced code to plot the distributions of each variable and correlations between pairs of variables for exploration purposes, prior to feature selection. I also researched and implemented all the model evaluation that was used for our experiments.

After this, I wrote a significant amount of the initial report, including the introduction, research, exploration, conclusion and future work sections, due to my lighter workload at the time. I also contributed to the modelling section – working alongside another group member. As a group we then built upon this and refined what I had written, incorporating the sections that were outstanding, which were written by those group members who had implemented the subject material themselves (and thus had the best understanding of it). Finally, I produced most of the figures included in the report, even those that related to an area for which the code was unfamiliar to me.

My achievements, therefore, are the pieces of code that were used by everyone throughout the project. This code was integral to the process and written in such a way that each member could understand and use it effectively, reducing time wasted by my colleagues focussing on getting their code to integrate in with the rest of the project. On top of this, I have achieved a lot by my further reading and research, as well as writing substantial parts of the report - gaining a strong understanding of the data science process, especially in the medical domain.

I have learned about the entire data science process in this course and have experienced first-hand the intrigues and complexities that this includes. Although the course was substantially marred by the unforeseen events of the time which I feel have negatively impacted my group experience, I have learned about a vitally important real world application which one day I hope I may experience in the workplace.