Research Proposal Form

Name and Surname:  
Ramon Falzon

MCAST Email:  
[ramon.falzon.a100947@mcast.edu.mt](mailto:ramon.falzon.a100947@mcast.edu.mt)

Group:  
SWD-6.2A

Research Title:  
AI for Research

Hypothesis and/or Research Questions:  
Is it possible to train a model that can deduct whether a user review is in favor or against the product?

Is it possible to train a model that can predict what rating the review would have on the product?

Outline of Key Literature:

<https://www.dphu.org/uploads/attachements/books/books_3754_0.pdf> (Source 1)  
Katharina J. Srnka and Sabine T. Koeszegi suggest that no data should be removed from the data set, unless the data set is too large; should this be the case, the selection process for the data should be at random. It also suggested that the more detailed the category scheme is, the more valid the results will be. Lastly, it is possible for categories to be hierarchal; but this is not necessary.

<https://www.researchgate.net/publication/333671063_Marketing_and_Artificial_Intelligence> (Source 2)   
Krystyna Jarek and Grzegorz Mazurek describe AI as a computer system that would require human intelligence; such tasks are: visual perception, understanding speech, making decisions and translating languages. However, a simple algorithm can classify as AI if it meets one of these criteria. It is Machine Learning that improves the AI and allows it to learn from the data given to it by a user. Deep Learning carries Machine Learning further by removing the need of manual interaction. Krystyna Jarek and Grzegorz Mazurek then show the application of AI in marketing, the applications branch into nearly every branch of marketing.

https://dl.acm.org/doi/abs/10.1145/3342827.3342829 (Source 3)  
In Xueying Wang’s paper, his method to see how positive a post was, was by pre-processing it first and then dividing the total word count of the sentence by the number of positive words. However, some of the pre-processed data ended up with no words, thus creating some results of greater than 1, which simply should be impossible when working with percentages. The pre-processing of data is effectively a self-tagging method.

<https://ieeexplore.ieee.org/abstract/document/6930158/> (Source 4)  
In a paper titled “Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis” made by Haruna Isah, Paul Trundle and Daniel Neagu, it was shown that it was possible to create an AI that observes Social Media, understand what people are talking about and then apply labels to that subject depending on what people have said about it.

Overview of Methodology:  
<https://www.kaggle.com/bittlingmayer/amazonreviews> (Dataset to be used)  
Tensorflow will be fed a tagged dataset of Amazon Reviews to have the model trained and then an application will be made where the user reviews an item and then the model gives a prediction on how much the user would rate it.

Ethical Considerations: None

References:  
<https://www.kaggle.com/bittlingmayer/amazonreviews>   
<https://www.dphu.org/uploads/attachements/books/books_3754_0.pdf>   
<https://www.researchgate.net/publication/333671063_Marketing_and_Artificial_Intelligence>  
<https://dl.acm.org/doi/abs/10.1145/3342827.3342829>  
<https://ieeexplore.ieee.org/abstract/document/6930158/>