**9980 / 01 Cambridge International Project Qualification**

**Project Proposal Form**

|  |  |  |
| --- | --- | --- |
| **Centre number** | IN693 |  |
|  |  |  |
| **Centre name** | Aditya Birla World Academy | |
|  |  |  |
| **Candidate number** | 0006 |  |
|  |  |  |
| **Candidate name** | Ayaan Amyn Mecklai | |
|  |  | |
| **Examination series (MM/YY)** | 10/22 |  |

**Research question:**

|  |
| --- |
| To what extent can Artificial Intelligence be used to help filter email spam ? |

**Briefly explain why you have chosen this question and topic.**

|  |
| --- |
| Machine Learning systems are computer programs used to input large pieces of labelled training data to make correlations and patterns and using these patterns to make predictions.  Machine Learning focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.  In today’s day and age, we use our technology to improve online communications, examples of these types are emails, messages, phone/video calls  and many others. However, spam messages may be the biggest problem faced by online users. Spam is unwanted digital communication that is sent out in bulk to many people. Most spam is often observed via email, but it is also distributed through phone calls, text messages, and social media. These spam messages are also very harmful as they can trick users(using phishing and malspam) into giving up their information.  These various types of spam are illegal and have a very big negative  impact on users. Therefore, websites such as CAPTCHA and DataDome verify user from machines which distribute the mass spam messages. While these websites are able to differentiate between the two, they are not able to identify  spam content.  This essay will try to find out whether or not deep learning can be used to differentiate between actual messages and spam messages . |

**Outline the research method(s) you will use and explain how this will help you to answer your research question.**

|  |
| --- |
| I will use secondary research based on codes and explanations by ensuring the code is checked for their intellectual rigor to explain, analyse and evaluate the method/process of finding the most optimal solution for the problem. This can be answered through the use of content-based filtering, case ban spam filtering or rule-based spam filtering. In machine learning specifically we can use Clustering and Naïve Bayes Classifiers to approach the question.    I will also try to create simple codes on python based on content filtering to differentiate between actual messages from spam messages. We can also use open-source packages like Pandas and NumPy to help filter these messages. We can filter messages by using feeding data into a piece of code which finds similarities in common emails and messages. These keywords can then be used to differentiate between actual messages and spam messages. |

**Outline the main sources of information you have identified for the project.**

|  |
| --- |
| <https://www.computerworld.com/article/2541475/spammers-establishing-use-of-artificial-intelligence.html>  <https://www.sciencedirect.com/science/article/pii/S2405844018353404>  <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.6385&rep=rep1&type=pdf>  <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.14504201146> |

**Teacher feedback**

|  |  |  |
| --- | --- | --- |
| **Teacher name** | Vandana Puri | |
|  |  |  |
| **Date of comments** | 19-10-2021 |  |

**Comments**

|  |
| --- |
| Ayaan your topic renders itself to meaningful research. However, you will need to narrow down your RQ. Work on your RQ and make it more focussed by picking up one technique out of the many mentioned above by you for spam filtering. Once you are able to narrow down the RQ you will need to look for reliable and relevant resources based on the topic. Out of the four resources mentioned the second resource is a comprehensive one. Read it carefully to help you in making your RQ focused. Other resources will not prove helpful for an in depth analysis.  Your research methodology is described clearly.  I will like to see you continue this journey with excitement and rigour after selecting a technique of spam filtering for deeper analysis. |

Approved ☐

Approved (with proviso) ☐ yes

Not approved (resubmission required) ☐