**Project Proposal (Deadline 28 Jan 2021)**



The proposal should be as detailed as possible. You may include additional items in the proposal and it can be more than one page long. Any major deviation from the original project idea after confirmation of project title will result in a penalty of marks.

|  |  |
| --- | --- |
| **Project Title:** | ML-POWERED CONVERSION OF SUBJECTIVE TEXT TO OBJECTIVE TEXT |
| **Team Members:** | Wang Hengyue (M21504)  Mendoza Kieran Majel (M21504) |
| **Domain:** | Media, Literature |
| **Subject Area:** | Supervised Learning  Natural Language Processing |
| **Description:** | Often, the biases and the judgements of journalists get in the way of the facts and the truth. This project aims to first detect whether a piece of text is subjective or objective, then reframe it to eliminate the prejudices and be more factual.  This project will include the development of a model that will detect the objectivity of the text input, and then will progress to transform it to subjective text and vice-versa. |
| **Objective(s):** | * Detection of author stance * Detection of subjective or objective text * Manipulation of text from being subjective to objective and vice-versa |
| **Target Users:** | * Fake news Detection agencies - a highly subjective article is likely to contain some fallacious statements * Media companies - a quick method of turning an otherwise boring, dull chain of facts into a more engaging story * English Teachers - aid for grading essays to determine if the student is writing subjectively or objectively |
| **Project Plan and Timeline** | **Core functionalities**  29 Jan: Proposal  11 Feb: Cleaned dataset  25 Feb: Text objectivity detection  11 Mar: Text manipulation on a sentence level  **Additional (optional) functionalities**  18 Mar: Text manipulation on a paragraph  25 Mar: Writing of report, slides and video presentation |
| **Resources and Tools Use** | * sklearn [scikit-learn: machine learning in Python — scikit-learn 0 ...](https://sklearn.org/) * (unconfirmed) GPT-2: [GitHub - openai/gpt-2: Code for the paper "Language Models …](https://github.com/openai/gpt-2) * tensorflow   **Useful DataCamp Courses**   1. Introduction to Natural Language Processing in Python (Katharine Jarmul) 2. Natural Language Generation in Python (Biswanath Halder) 3. Recurrent Neural Networks for Language Modeling in Python (David Cecchini) |
| **Target Outcome and Benefits** | * Catering to the preference of different readers - some readers prefer the facts presented neatly and clearly while others may want more personal voice * Formation of an engaging story from a “skeleton” of facts collected, allowing for a “quick story”. |
| **References** | <https://www.kaggle.com/snapcrack/all-the-news>  <https://components.one/datasets/all-the-news-2-news-articles-dataset/>  <https://deepai.org/publication/automatically-neutralizing-subjective-bias-in-text> |