Capstone Project 2 Ideas

Brexit Sentiment Analysis

Idea:

The 'Brexit vote' has lead the UK to begin negotiations to leave the European Union. Given that that only 72% of the population voted at all and only 51.89% voted to leave has made this an even more controversial topic. The invocation of Article 50 (the formal legal process by which the UK will leave the EU) could potentially be reversed/withdrawn on a legal basis. With this in mind, by studying the sentiment of the UK people since the Brexit vote till the present day, we could get an idea of how much of the population still wants to leave the European Union, given that they are more aware of what repercussions that brings.

By doing so, we would answer potentially answer the question if the UK still wants to leave the EU or not.

Data sources: social media and online crawling. Twitter, Facebook, LinkedIn and news sources online.

Either choose 1 or a few, or all, depending on difficulty and time consumption of crawling/using APIs.

For twitter, the library that works with the twitter api is: https://pypi.org/project/twitter/

limits: https://developer.twitter.com/en/docs/basics/rate-limits

https://apps.twitter.com/app/15324212

http://socialmedia-class.org/twittertutorial.html

Building a Question Answering System

Idea:

Question Answering is a field of Machine Learning/NLP where a system can answer questions posed by humans in natural language. Clearly as field it is very useful, and having the ability to create question answering systems given a set of documents/corpus is a valuable skill.

The aim of this project will be to create such a system for the popular Question Answering dataset- SQuAD (Stanford Question Answering Dataset). There is a leader board for the best performing models, and a test set so evaluation of the project, so it is possible to compare your model against the best models in the world.

Dataset: https://rajpurkar.github.io/SQuAD-explorer/

Article on a project related to this dataset: https://towardsdatascience.com/nlp-building-a-question-answering-model-ed0529a68c54

Wiki article: https://en.wikipedia.org/wiki/Question_answering

Document Summarisation

Idea:

GDPR (General Data Protection Regulation) is a regulation in EU law on data protection and privacy for all individuals within the European Union (EU) and the European Economic Area (EEA) and was implemented on 25th May 2018. It defines new rights that individuals hold with regards to the use of their data by companies and institutions, while enforcing, by law, how personal information can be used by the companies themselves. This is of course an oversimplification, and the full details can be found at: https://gdpr-info.eu/. However, as we know, individuals seldom read the Terms and Conditions when signing up for a product or service, so it is likely that a large proportion of individuals won't be aware of their new rights regarding the use of their data. This brings us to the idea and the client of this capstone. By summarising the contents of the GDPR, we will make it easier and quicker for individuals to understand the new laws, which is highly relevant in the 'data age'.

Data: we will need to scrape GDPR from this page: https://gdpr-info.eu/

Client: an 'EU citizen'.