

Dear Hiring Team,

I am a recent graduate from University College London with a Masters of Engineering in Computer Science with a specialisation in Deep Learning, Reinforcement Learning and Natural Language Processing. I am coming out of university with a passion for Data Science and new technologies.

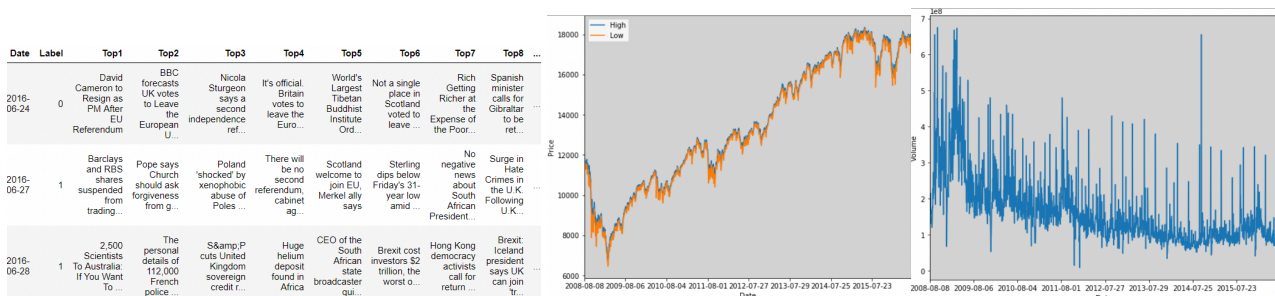
I am a trustworthy and dedicated individual always looking for a challenge and an opportunity to learn more about data science. I thrive in collaborative environments. I am outgoing and extraverted and always share my ideas with the people I work with. I believe that getting feedback is the best way to deliver innovative products.

I have valuable communication skills. Through my years as a teacher, I have learned to speak confidently in front of audiences and explain complex ideas to people with less technical knowledge than me. This helps me translate technological jargon to non-tech people more easily and be more business focused. Overall, I enjoy spending time explaining my ideas and point of views.

In this cover letter, I have summarised my recent data science academic and professional experiences with links to the research papers and some screenshots.

Predicting DJIA Stock price and Volume

During my master (2018-2019), I led a project which involved looking at Reddit News articles in order to predict changes in price and volume of the DJIA with two other students. The idea was to give a daily trader an edge over other traders by using Reddit news from the previous day to predict what would happen in the following day. Screenshots below show a picture of the dataset, a graph of the DJIA price over 8 years and lastly a graph of the volume traded of DJIA changing over time.



A lot of research had been done on predicting the changes in price for the next day but much less on predicting volume traded on the next day. Our hypothesis was that news would predict changes in volume more effectively than changes in price. We regarded both problems as classification tasks and used the XGBoost algorithm and the LSTM neural network to predict volume and price changes. The results of this research confirmed our hypothesis although we would not advise a trader to take it for granted. We received a first class grade for this research.

Link to research paper: https://rdumon.github.io/NLP_report.pdf

Quantitative Study of Differentiable Neural Computer (DNC)

For my master thesis, I studied the Differentiable Neural Computer (DNC) by Alex Graves et al. from Google DeepMind. This research was supervised by Professor John Shawe Taylor - head of Computer Science at UCL.

The DNC is a memory augmented neural network. It consists of two main units: a controller unit and a memory unit. Training a DNC means that you are training the controller (neural network) to use the memory unit effectively. My research was divided into 2 parts. First, I looked at the history behind the model and also explained the neuroscience theories which were used to build the model. Secondly, I looked at the difference in performance between the DNC and the BrsDNC (an augmented DNC) on the bAbI Natural Language Processing dataset. The bAbI dataset consists of twenty synthetic question answering tasks. The idea behind the dataset is that "an intelligent dialogue agent" should have a set of basic tasks it needs to answer correctly in order to be considered worthy of being called intelligent. Through a set of simple

induction, deduction, reasoning tasks and more, the authors argue that this dataset puts together the simplest question answering capabilities that any dialogue agent should possess. A human would make zero mistakes solving these tasks. They also propose this dataset to enable researchers to benchmark their neural network architectures against each other and identify their failures.

To look at the difference in performance between the two models, I first carried out an ablation study. This ablation study helped me identify that the normalisation in the BrsDNC was the main reason for the improvement in performance of the model compared to the DNC. I then tried to evaluate the limitations of the BrsDNC by hypothesising that it would perform worst than the DNC on task which involved sequential memory pattern recognitions. To test my hypothesis, I had to developed two new synthetic datasets. Screenshot below shows examples of the type of questions feature in each of the respective datasets I built.

Example Task 21: Listening to one specific person

Focus on Yann.
Yann is saying that Daniel went back to the kitchen.
Winona is saying that Mary and Sandra journeyed to the garden.
John is saying that Daniel and Mary went back to the garden.
Brian is saying that then he moved to the bedroom.
What did Yann say ? Daniel went back to the kitchen

Example Task 22: Location Reasoning

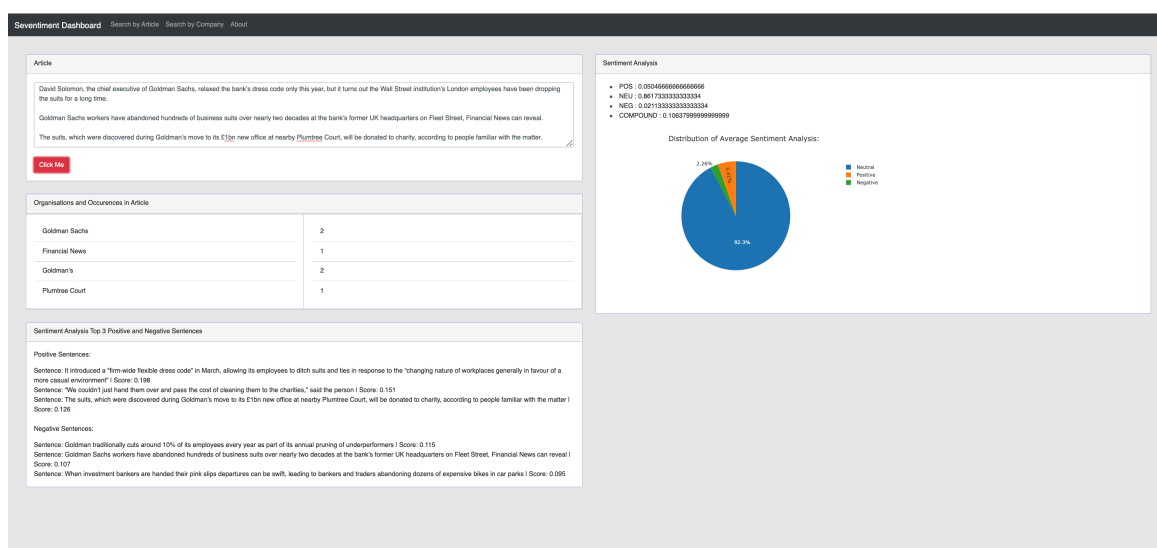
Greg travelled to Madrid , Rome , Sarajevo , Dublin and London.
Emily travelled to Bern , Berlin , Kiev , Prague and Paris.
Daniel travelled to Lisbon , Amsterdam , Belgrade , London and Copenhagen.
Where did Emily travel to before Prague? Kiev

These tasks are synthetic in that any researchers wanting to use them in the future would be able to run a Python Script to recreate these datasets. I was rewarded by obtaining a first class grade for this research. Through this research, I have proven that I can understand complex machine learning models and work with state of the art research.

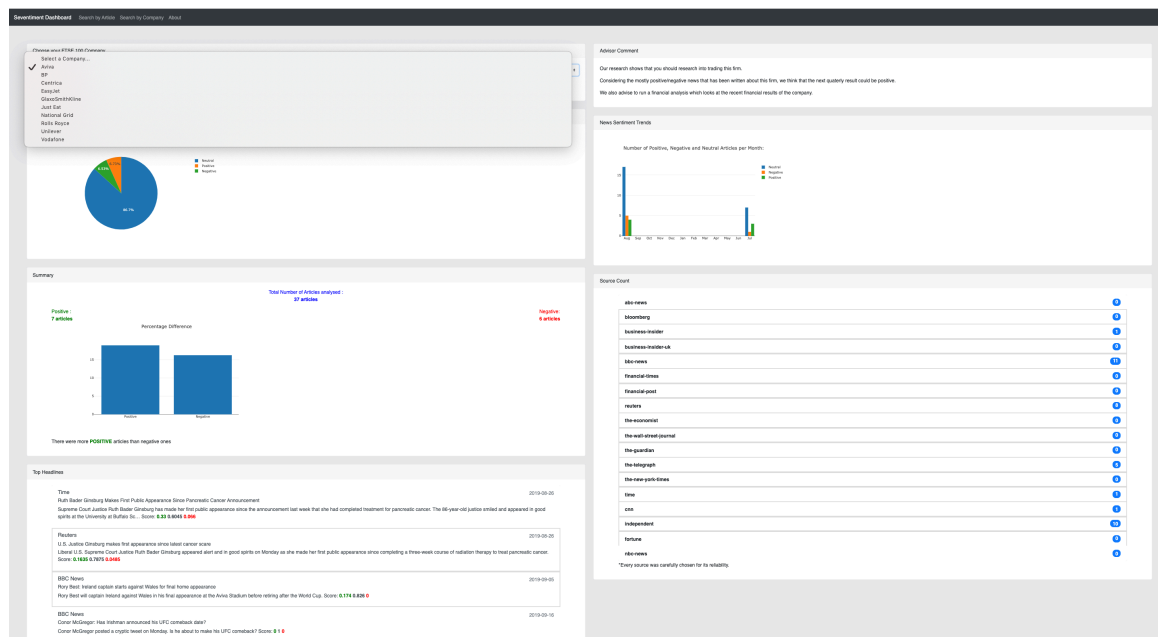
Link to research paper : <https://rdumon.github.io/finalproject.pdf>

Sentiment Analysis Project

During my internship at Deutsche Bank, in a team of four we developed a dashboard to analyse FTSE 100 companies for a trader in the business side using ReactJS for visualisation and Python Flask on the Back-End. My team used Sentiment Analysis models to build different graphs that would enable traders to look at the news in a new way. The tool had two interfaces. First, a sentiment analyser where the user can copy & paste an article and get an entity analysis of the article, a sentiment score, and a ranking of the most meaningful sentences (Positive and Negative) in the article.



The second interface is used to analyse the historical news about all the companies in the FTSE100. The user selects a company and on clicking submit receives a thorough analysis of the news articles published about the company over the last year. The interface consists of overall score analysis over the last year, a monthly analysis (number of positive vs negative vs neutral articles) in the last year, a top five headlines feature and also a count from source. (note: the original database belonged to Deutsche bank here the image shows dummy data).



Overall, our platform was incredibly successful at displaying useful information and was selected as the best intern project. The project served as a proof of concept for real life application which is now being used by traders.

Open Street Maps Research

In my last year of Bachelor (2017-2018), I was involved in a data science group project led Professor Licia Capra on the Open Street Map (OSM) Dataset. OSM is the second biggest crowd platform in the world after Wikipedia making this project a very fun real life experience to have worked on. The goal of our research was to look at the impact of big data imports from corporations/companies on the activity of individual contributors in 7 major European cities. For example, the picture below shows a screenshot of bus stop imports from Transport For London (TFL) in London.



In this research, we asked three major questions: would individual users decrease their amount of contributions after imports, would users change their object of contributions (ex: from restaurants to park

benches) and would contributors change the type of contributions they make (update more than create). Our group was awarded a first for this project.

Link to research paper : https://rdumon.github.io/Group_Report.pdf

Looking forward to hearing back from you,

Romain Dumon