

# Felix Drinkall

DOCTORAL CANDIDATE · UNIVERSITY OF OXFORD

Department of Engineering Science, University of Oxford, Eagle House, Walton Well Road, Oxford, OX2 6ED

✉ felix.drinkall@eng.ox.ac.uk | 🌐 www.felixdrinkall.com | 📱 felix-drinkall-645b0111b

Doctoral Candidate at the Oxford-Man Institute of Quantitative Finance, University of Oxford, and Professional Athlete for GB Rowing.

## Education

### University of Oxford

Oxford, United Kingdom

DPHIL IN NATURAL LANGUAGE PROCESSING AND QUANTITATIVE FINANCE

October 2021 - Present

- Member of the Oxford-Man Institute of Quantitative Finance and Oxford e-Research Centre.
- Researching the intersection between **Natural Language Processing** and **Time Series Forecasting**.
- Organising committee of the **Conference on NLP for Social Data Sciences (SoDaS)** (previously the Conference on NLP for Economic and Financial Modelling).

MENG IN ENGINEERING, ENTREPRENEURSHIP AND MANAGEMENT

September 2017 - June 2021

- **First-Class Honours**
- **Recipient of the Vincent's Ferside Scholarship** for academic and sporting success at Oxford University.
- **84% in my Masters' Thesis:** "Forecasting COVID-19 caseloads using social media posts" - Edgell Sheppee Prize 2021 for the **best thesis from an Engineering, Entrepreneurship and Management student**
- Courses: Entrepreneurship and Management, Machine Learning, Machine Vision and Robotics, Medical Imaging and Informatics, Advanced Probability, Systems and Perturbation Methods and Advanced Mathematics.

## Work Experience

### Stanford University

Oxford, UK

NATURAL LANGUAGE PROCESSING TUTOR

May 2022 - Present

- I provide tutoring and guidance to visiting Stanford University students participating in an exchange program at Oxford. My role includes instructing them in an Introduction to Natural Language Processing course and supervising the completion of their thesis.

### GB Rowing Team

Reading, UK

PROFESSIONAL ATHLETE

November 2017 - Present

- Throughout my university studies I have had to balance the obligations and responsibilities of a professional athlete. This has made me **able to balance a tough workload and work under pressure** whilst maintaining peak performance in and out of the classroom.

### Rahn+Bodmer

Zurich, Switzerland

QUANTITATIVE INVESTMENT ANALYST INTERN

July 2020 - August 2020

- I **created a full-stack portfolio optimising tool using Python** within a Django framework. This tool was able to optimise a portfolio for different objectives and constraints using the historical data from the assets that they own.
- During the internship I was also given insight into the trading department, the development of investment strategy and fundamental research.
- On completion, I presented the tool to the financial analysis team and several of the partners.

## Papers

### Forecasting COVID-19 Caseloads Using Unsupervised Embedding Clusters of Social

Seattle, USA

#### Media Posts

NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (NAACL)

July 2022

We present a novel approach incorporating transformer-based language models into infectious disease modelling. Text-derived features are quantified by tracking high-density clusters of sentence-level representations of Reddit posts within specific US states' COVID-19 subreddits. We benchmark these clustered embedding features against features extracted from other high-quality datasets. In a threshold-classification task, we show that they outperform all other feature types at predicting upward trend signals, a significant result for infectious disease modelling in areas where epidemiological data is unreliable. Subsequently, in a time-series forecasting task we fully utilise the predictive power of the caseload and compare the relative strengths of using different supplementary datasets as covariate feature sets in a transformer-based time-series model.

This paper evaluates the ability to predict COVID-19 caseloads in local areas using the text of geographically specific subreddits, in conjunction with other features. The problem is constructed as a binary classification task on whether the caseload change exceeds a threshold or not. We find that including Reddit features, alongside other informative resources, improves the models' performance in predicting COVID-19 cases. On top of this, we show that exclusive use of Reddit features can act as a strong alternative data source for predicting a short-term rise in caseload due to its strong performance and the fact that it is readily available and updates instantaneously.

## Skills and achievements

---

### Technical

- Proficient in **Python, MATLAB, Excel, LaTeX and Solidworks CAD.**
- Basic knowledge of **C++ and SQL.**

### Other

- **Junior World Rowing Champion from 2017**
- **U23 World Champion from 2019 and 2021.**
- In the Oxford Blue Boats that competed in the Boat Race in 2018, 2019 ad 2021.
- Winner of the National Pro Corda Chamber Music competition, playing at the Wigmore Hall with my brass band.
- In my school play Jerusalem (written by Jez Butterworth) I performed to 400 people each night.
- Training to become a Pommelier.