

# Daniel Petrov

daniel.s.petrov@gmail.com

linkedin.com/in/danielspetrov

github.com/daniel-petrov

## Education

### Trinity Hall, University of Cambridge

MPhil in Data Intensive Science

Cambridge, UK

October 2023–August 2024

- Awarded **Cambridge Masters & Trinity Hall Studentship**.

### University of St Andrews

BSc (Joint Honours) Computer Science and Statistics: **First class**

St Andrews, UK

September 2019–June 2023

- Achieved **Deans' List** (average first class in all modules across the academic year) in all four years of degree.
- Received letter of **academic excellence** in Maths and Medal for **highest performance in Third level Statistics**.
- Completed two dissertations in Computer Science and Statistics achieving 90% and 83%, respectively.
- **Relevant Modules**: Machine Learning, Artificial Intelligence, Language and Computation, Bayesian Inference, Time Series Analysis, Markov Chains and Processes, Spatial Statistics, Multivariate Calculus, Software Engineering Project.

## Work Experience

### Equinor

Data Science Intern

Natural Language Processing

Machine Learning

Python

London, UK

June 2023–August 2023

- Joined **Knowledge AI team** to automate the production of news article summarisation reports.
- Heavily **NLP-based**, including tasks in **named entity recognition**, textual vectorisation, sentence similarity, summarisation, **bias detection**, document-based clustering, and **question and answering chatbot**.
- Made use of models found on **HuggingFace**, an **open-source** machine learning community, and Kubeflow notebooks.
- Environment: Windows 11, Python 3.8 (pandas, numpy, sentence-transformers, spacy, nltk, dbias), Kubeflow.

### Tripadvisor

Software Engineering Intern

Java

Typescript

SQL

Oxford, UK

June 2022–September 2022

- Worked in **Payment Services Engineering** team on system migration from in-house data centre to **Amazon Web Services (AWS)**.
- Developed, tested and deployed an **automatic backup process** for customer subscriptions and payments history data stored in **DynamoDB** and **RDS** instances to **AWS Backup Vaults** created within multiple AWS accounts to increase security and failover in cases of AWS account failures.
- Deployed the backup process as a **cloud-based microservice** for a common usage by other company teams.
- Built **API endpoints** for **import/export** of subscriptions data from a local database to AWS DynamoDB instances.

### AgSpace Agriculture Ltd.

Data Science Intern

Python

Jupyter Notebook

SQL

Remote, UK

June 2021–July 2021

- Built a **fully working prototype** that implements automatic agriculture field boundary detection using **geospatial data** taken from Sentinel, Planet and Mapbox.
- Enhanced the prototype's image recognition with elements of **Machine Learning (CNN)**.
- Experienced **agile** development in the workplace using **Azure DevOps**.
- Environment: Anaconda Python 3.7, Jupyter Notebooks, PostgreSQL, AWS (SageMaker, DataWrangler, S3 Buckets and Cloudwatch Logs), JetBrains PyCharm, MacOS and AWS Linux.

### Vodafone Group Plc.

Data Scientist Assistant

Python

R

Madrid, Spain

April 2018

- Visualised spatial data of a mobile network for the analysis of the catchment areas of retail chains.

### Vodafone Group Plc.

Vodafone Business Group Program Management Office Analyst

Newbury, UK

August 2017

- Worked in the Group Business Products PMO team on visualisation of large data sets used in a dynamic dashboard built in Microsoft Power BI.

## Projects

---

- **Stacshack Hackathon, March 2023** [Jupyter Notebook](#) [Python](#) [CVZone](#) 
  - In a group of two, created an application that allows the user to **map any hand gesture to a series of computer controls** (e.g. mouse movement, click, commands, etc.), making sure to keep the application **highly customisable**.
  - **Won the general category of the AI-themed hackathon.**
- **Hack the Bubble (Hackathon), October 2022** [Jupyter Notebook](#) [Python](#) [NumPy](#) [scikit-learn](#) 
  - Carried out an individual **data science project** to answer the question “What makes a good sci-fi movie?” within an 8 hour time frame.
  - Made use of a **5000 movie database** from kaggle, and scikit-learn to preprocess and make conclusions on the **significance of features for a successful science fiction movie.**
- **Webscraper to query surf conditions for a beach nearby** [BeautifulSoup](#) [Facebook](#)
  - Built a webscraper to automatically and sequentially check the wave conditions for a beach nearby.
  - Enhanced the process with a notification from Facebook messenger for when I wake up.
- **Neural Network to predict heart disease from scratch in Python** [Jupyter Notebook](#) [NumPy](#) [Pandas](#) [scikit-learn](#) 
  - Constructed a **binary classification** neural network to predict whether heart disease would be present.
  - Constructed a **multi-classification** neural network to label two-feature data.
- **Stacshack Hackathon, March 2021** [Python](#) 
  - Developed a website which produced artificially made pick-up lines generated by the **GPT-2 deep neural network** trained on a web-scraped dataset.
  - Won “Hackiest Hack” category from Hackathons UK.
- **Minimax algorithm in Connect Four** [Python](#) 
  - Users play the Connect Four game against the computer which utilises the **minimax** algorithm to make decisions.
  - Made use of **alpha-beta pruning** to halve the time taken in decision-making.
- **Machine Learning, Stanford University, Coursera, June - August 2020** [Jupyter Notebook](#) 
  - Learnt Multivariate Linear and Non-Linear Regression, Classification, Regularisation, Neural Networks, Support Vector Machines, K-Means Clustering, Dimension Reduction and Anomaly Detection
- **Stacshack Hackathon, March 2020** [Java](#)
  - Developed a 2D, top-down, single player game in a group of four over 24 hours.
  - The game contained four biomes (meadow, desert, snow and river) and was enhanced with dialogues, character interactions and player animations.

## Skills

---

- **Programming Languages:** **Python/Java** (Proficient, 3+ years experience), **C/R** (Familiar, 1-2 years experience) **SQL** (Working Knowledge, <6 months experience)
- **Tools and Frameworks:** Git, numpy, pandas, scikit-learn, matplotlib, seaborn, plotly, cufflinks, BeautifulSoup, shapely
- **Languages:** **English** (native), **Russian** (intermediate), **French** (beginner)