Alexandre **Tytgat**

Machine Learning Engineer

1332 Genval, Belgium

(+32) 497 10 80 48 | alextytgat@gmail.com | atytgat | alexandre-tytgat

Skills

Programming Languages ***Experienced:*** Python, R. ***Familiar:*** SQL, Java, Matlab, SAS, C++, AMPL.

ML Libraries Scikit-Learn, Pytorch, Caret, NumPy, Pandas, spaCy, NLTK, Gensim, Doc2Vec, Seaborn, Matplotlib, JAGS.

Tools Git, Anaconda, Jupyter Notebook, Excel, Google Colab.

Languages ***Native:*** French. ***Fluent:*** English. ***Beginner:*** Dutch, Spanish.

Education

MSc in Data Science, *with distinction*

UCLouvain, Belgium

Specialization in numerical methods and optimization. *09/2019 - 09/2021*

- Relevant courses: databases (SQL, NoSQL), machine learning (supervised learning, unsupervised learning, reinforcement learning, deep learning, CNN, NLP), data visualization, Bayesian statistics.
- Student representative for the academic year 2020-2021.
- Thesis: Study of a new interpretation of the learning process of AI. Obtained the grade of 16/20. *Online access*

MSc in Physics

UCLouvain, Belgium

Specialization in statistical and mathematical physics. *09/2018 - 09/2019*

- Dropped out after a year to study data science.

BSc in Physics, *with distinction*

UCLouvain, Belgium

Specialization in mathematical physics. *09/2015 - 09/2019*

- Student representative for the academic year 2018-2019.
- Thesis: A new approach to dark matter as a quantum interaction phenomena. Obtained the grade of 18/20.

Projects

Air quality prediction

GitHub link

Goal : Predict the concentration of fine particles in the air of Beijing using meteorological data collected by the Beijing Municipal Environmental Monitoring Center. *Nov. 2019*

- Data pre-processing: transformed the time and the wind direction features in a way that takes into account their cyclical nature.
- Selected the five most informative features about the target variable to train the model on, as evaluated by the mutual information.
- Trained a one layer neural network on the resulting dataset, and performed a 5-fold cross validation scheme to select the best performing model.

T-cells Identification

GitHub link

Goal : Use hand annotated datasets of gene expression to train machine learning algorithms in order to automatically classify new single-cell datasets. *May. 2020*

- Carried out the data pre-processing, and applied PCA to reduce the number of features from >23.000 to less than a hundred.
- Reduced the bias in the dataset using the SMOTE method to artificially balance the classes.
- Performed 5-fold cross validation over several models (randomforest, XGBoost, etc.), and selected the six best performing ones. A voting system was employed to make the final prediction, resulting in a balanced accuracy of 0.78 on the test set.

Political Comments Classifier

GitHub link

Goal : Predict the political affiliation of users from their Reddit comments using modern NLP techniques. *Nov. 2020*

- Performed the necessary pre-processing of the comments using Gensim and NTLK.
- Created one BERT embedding of the comments and trained a Doc2Vec model to create a second embedding.
- For each embedding, a randomforest was trained to identify the political affiliation of the comments. The BERT model obtained an accuracy of 0.75 on the test set.

Extracurricular

Lab tutor on wave mechanics

UCLouvain, Belgium

02/2019 - 05/2019

- Prepared lectures to introduce students to the theory and to the experimental apparatus.
- Monitored and guided the students to successfully realize the desired observations.
- Charged with grading the lab reports and took the initiative to provide individual feedback after each session.