

# Matteo Esposito

JUNIOR DATA SCIENTIST & MILA MASTERS STUDENT

☎ 514-922-6789 | ✉ [matteoesposito97@gmail.com](mailto:matteoesposito97@gmail.com) | 🏠 [www.matteoesposito.com](http://www.matteoesposito.com) | 📷 [matteo-esposito](#) | 📺 [matteo-esposito](#)

## Skills

<b>Programming</b>	Python, SQL, Java, R, SAS, Bash, Git, $\LaTeX$
<b>Libraries</b>	Pandas, Numpy, Scikit-Learn, Pytorch, TorchAudio, SpeechBrain, PyYAML, Shap, HuggingFace, Transformers, Nltk, OpenCV
<b>Technologies</b>	Google Colab, Google Cloud Compute Engine, Apache Airflow, Github, Gitlab, Linux, Jira, Agile
<b>Languages</b>	English (Fluent), French (Fluent), Italian (Intermediate)

## Relevant Work Experience

### Intact Financial Corporation

*Montreal, QC*

DATA SCIENTIST I

*May 2021 - Present*

- Piloted and presented various proof-of-concepts for expert system implementations in **Python** with Experta and the **Java** Drools framework.
- Maintained a Java **expert system**, managed dependencies and builds with **Maven**, wrote unit tests, monitored **CI/CD pipelines** in Gitlab.
- Created **data preparation pipelines** leveraging **pandas, numpy, SAS and SQL** with **Oracle databases** and automated them using **Apache Airflow**.
- Trained supervised learning models for numerous added functionalities on web quoting service (tailored-web experience).
- Led recurring Agile meetings with the team and stakeholders for project management.
- Coached and supervised 2 data science interns over 8 months.

### Aviva Canada

*Montreal, QC*

DATA ANALYST CO-OP

*May 2019 - Apr. 2020*

- Evaluated 3rd party geolocation/telematic data source for added value of new features in pricing regression models.
- Developed and maintained an **internal python library** used for actuarial calculations.
- Integrated library with modeling pipeline for premium validation and replaced legacy process. (4x improvement in mean run-time)

### Aviva Canada

*Montreal, QC*

DATA ANALYST CO-OP

*Jan. 2018 - Aug. 2018*

- Built predictive models using **scikit-learn** to model frequency and severity of personal auto insurance claims.
- Reduced error of **variable imputation** methods in Ontario pricing databases by 20% using **multi-class classifiers**.
- Created a database of at-risk-clients in SAS & trained a **binary classifier** to predict risk of default with 85% out-of-sample accuracy for the optimization of the current risk/client selection process.

## Projects

### Open Source Contribution to SpeechBrain (Concurrent Speaker Counting)

ADDITIONAL PROJECT IN COLLABORATION WITH PROF. MIRCO RAVANELLI

- Create modeling pipeline (data cleaning, augmentation, processing and modeling using **PyTorch**) for XVector and ECAPA TDNN model architectures to identify number of concurrent speakers in an audio sample.
- Synthesize overlapping speech dataset (0-5 speakers) from the **LibriSpeech** audio dataset.
- <https://speechbrain.github.io/>

### Open Source Contribution to SpeechBrain (Automatic Speech Recognition)

PART OF IFT6759 - ADVANCED PROJECTS IN MACHINE LEARNING

- Implemented Facebook AI's (fairseq) **wav2vec2.0 self-supervised learning** framework in the SpeechBrain library.
- Used **nvidia-smi** for GPU management and interaction.
- Reproduced state-of-the-art phoneme error rates on automatic speech recognition downstream tasks.

## Education

### University of Montreal/MILA

*Montreal, QC*

MASTERS OF SCIENCE - COMPUTER SCIENCE, MACHINE LEARNING SPECIALIZATION

*Sept. 2020 - Jan. 2023 (Expected)*

- Professional Masters, GPA: 3.7, completing while working full time as a data scientist at Intact Insurance.

### Concordia University (Graduation with Distinction)

*Montreal, QC*

BACHELOR OF SCIENCE - ACTUARIAL MATHEMATICS CO-OP WITH MINOR IN COMPUTER SCIENCE

*Sept. 2016 - May 2020*

- Completed **4 full time, 2 part time co-op internships & 3 actuarial exams** during studies.
- Member of the Co-op Institute, **Golden Key International Honour Society** (Top 15% of students) & Data Innovation Society.