Rafael Lacerda

rafaelplmelo@gmail.com
Toronto, ON, Canada
(647) 878 9650

Overview

I tailor ML / DL / NLP research to fit your business. Proven results in 3 years of experience.

My contributions:

- Research and development of ML and NLP systems.
- Automation and machine learning to achieve business goals.
- Infrastructure required to host machine learning applications in the cloud.
- Communicating research to stakeholders of all skill levels.

Education

MSc. in Applied Computing *Sep 2016 — Dec 2017*

University of Toronto, *Department of Computer Science*

- Courses of note: Natural Language Processing, Computational Neuroscience, Probabilistic Learning and Reasoning.
- Recipient of 30k Mitacs scholarship (2x terms).
- Natural Language & Deep Learning final thesis: Question Answering in the Legal Domain.

BSc. in Economics *Aug 2011 — Jul 2014*

Universidade Presbiteriana Mackenzie, *Department of Applied Social Sciences*

- Econometrics final thesis
  - The Impact of Government Intervention in the Sugarcane Industry.
  - Modeled the effect of interventions on uncertainty using Real Options analysis.
- Previously a student of Computer Science at the same university:
  - Transitioned into Economics to focus on modeling human activity. Since then I have used my CS skills as tools to assist in modeling and automation.

Work Experience

Data Scientist (Applied Research) *Aug 2019 — Present*

Scribd Inc., Toronto, ON

- Developed a document feature pipeline to enable classification of millions of long-form documents.
- Specialized in the field of entity linking to create a novel entity linker with very few compromises.
- Created a novel entity extractor: higher precision than open-source alternatives as well as the previously contracted API.
- Clustering of non-uniform time series in Fourier space.
- Coordinated migration of the applied research team to a cloud-based ML serving platform.
- Involved in: weekly ML / deep learning / NLP update; keyword extraction; topic classification.

Machine Learning Engineer *May 2017 — Feb 2019*

Ross Intelligence Inc., Toronto, ON

- Developed a state-of-the-art legal research system that surpassed IBM Watson's performance, leveraging deep learning and NLP.

- Coordinated new research that further improved the research system, by creating specialized models to extract entities in documents.
- NLP projects include information retrieval, document ranking, summarization, topic modeling, document classification and data augmentation in the legal domain.
- Deep Learning tasks included designing networks from scratch, customizing existing networks, training models and word embeddings, model selection and productionizing models.
- Built fast and scalable NLP and Deep Learning pipelines.
- Sped up models for production by using caching and precomputation.
- Developed statistical tests with human assessments to ensure model superiority before deployment into production.
- Used Python, Tensorflow, Keras, Solr/Lucene, Pandas, Spacy, NLTK, Scikit-Learn, MongoDB, Redis, Docker, Travis CI, Kubernetes, AWS.

Technical Skills

- **Languages & Frameworks:** Python, R, Tensorflow, Keras, Numpy, Pandas, Scikit-learn, Spacy, NLTK, Matplotlib.
- **Databases, Caching:** SQL, NoSQL, MongoDB, Redis, Solr.
- **Cloud computing:** AWS ecosystem, Microservices architecture, REST APIs.
- **Continuous Integration, Continuous Delivery**: Github, Travis CI, Docker, Kubernetes.
- **Machine Learning:** Regression, Classification, Clustering, Dimensionality Reduction, Model Validation, Model Selection, Time series clustering in Fourier space.
- **Deep Learning:** NN, CNN, RNN, LSTM, GRU, Autoencoders.
- **Natural Language Processing:** Embeddings, Information Retrieval, Natural Language Understanding, Question Answering, Learning to Rank, Summarization, Entity Extraction, Entity Linking.

Prior domain knowledge in industries

- Legal Research & Case Law
- Finance & Treasury
- Real Estate Investment
- Customer Service Management
- Publishing

Other Projects

- Computational graph framework to chain video filters. Used OpenCV, Python, Numpy (2019).
  - Tensorflow-like API to chain image filters that operate spatial and temporal dimensions.
- Emotion based movie recommendation system. Used Python, BeautifulSoup, Numpy (2016).
  - Scraped movie scripts annotated with Ekman emotions throughout the timeline. Recommendations through similarity of emotional progression.
- 2 month immersive bootcamp covering a wide array of data science theory and practice.
- BSc. Economics thesis on a time series model to measure the effects of uncertainty brought upon by Federal intervention on oil policy. Used the Gretl stats package for the regression model (2014).
- 6 month internship in real estate management (OR Investimentos) on financial modeling. Automated processes using VBA (2014).
- 6 month internship at a hedge fund (Kondor Invest, risk office) clearing daily trades and automating processes. Used Python, SQL and VBA (2012).