**Rafael Lacerda**

[rafael@lacerda.ch](mailto:rafael@lacerda.ch)

Toronto, ON, Canada

**Overview**

I tailor ML / DL / NLP research to fit your business. Proven results in 4 years of experience in industry.

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 Scienceat 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*

* Research in taxonomy: category tree generation, document classification.
* Developed embeddings for: Entities; Long Documents (content-based & usage-based).
* 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.
* 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 papers update; keyphrase extraction.
* Used Python, PyTorch, Scikit-Learn, Spark, Spacy, ElasticSearch, MLFlow, Databricks, AWS.

**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.
* Reduced inference time in production by ~70% through 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.

**Other analytical roles**

* Treasury Analyst in Real Estate Investment (Jul 2014 — Apr 2016)
* Head of Customer Service in a Rocket Internet venture (Jan 2013 — Nov 2013)

**Technical Skills**

* **Languages & Frameworks:** Python, R, PyTorch, Tensorflow, Keras, Numpy, Pandas, Scikit-learn, Spacy, NLTK, Matplotlib.
* **Databases, Caching:** SQL, NoSQL, MongoDB, Redis, Solr, ElasticSearch.
* **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, Dynamic Time Warping.
* **Deep Learning:** NN, CNN, RNN, LSTM, GRU, Autoencoders, Transformers.
* **Natural Language Processing:** Embeddings, Information Retrieval, Natural Language Understanding, Question Answering, Learning to Rank, Summarization, Entity Extraction, Entity Linking, BERT.

**Prior domain knowledge in industries**

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

**Other Projects**

* Computational graph framework to chain video filters. Used OpenCV, Python, Numpy (2019).
  + Tensorflow-like API to chain image filters that operate in 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.