Juan Antonio Ramirez-Orta

Halifax, Nova Scotia, Canada

Contact: ramirez.orta.juan@gmail.com | [Homepage](#) | [GitHub](#) | [LinkedIn](#) | +1-902-943-4744

Personal Statement:

I am a Computer Science PhD with over 4 years of experience in Data Science, Data Engineering and Big Data. I am passionate about Mathematics, Machine Learning and Natural Language Processing, but what I really enjoy is to develop end-to-end projects that create insight from large collections of documents.

Education:

- **PhD in Computer Science, Dalhousie University**
  - *January 2020 - January 2024*
  - GPA: 4.13/4.3
  - Area: Machine Learning and Natural Language Processing
- **MSc in Mathematics, National Autonomous University of Mexico**
  - *January 2014 – January 2016*
  - GPA: 9.4/10
  - Area: Topology and Complex Analysis
- **BSc in Mathematics, National Autonomous University of Mexico**
  - *August 2009 – July 2013*
  - GPA: 9.8/10, Summa Cum Laude
  - Area: Geometry and Differential Equations

Work Experience:

- **Senior Data Scientist at Royal Bank of Canada**
  - *March 2024 - Present*
  - Test, validate and document Artificial Intelligence models from all across the bank, focusing on the deployment and testing of Large Language Models.
- **Director at Halifax Chess Club (Volunteering)**
  - *June 2022 - Present*
  - Monitor the club media, deliver lectures, organize tournaments, look for sponsors and manage the finances of the club.
- **Data Scientist at Dalhousie University (Contract Full-time)**
  - *January 2024 - June 2024 (6 months)*
  - Designed and implemented the front- and back-end of the Early Diagnosis Program in collaboration with the Nova Scotia Health Authority and developed the current implementation of the Explainable Hierarchical Binary Classifier.
- **Teaching Assistant at Dalhousie University (Part-time)**
  - *April 2021 - December 2021 (9 months)*
  - Elaborated fully-automated exam questions using R and Python to prevent cheating during COVID-19 in the courses of Theory of Computer Science and Foundations of Machine Learning.
- **Data Engineer at Banco Azteca**
  - *January 2019 - December 2019 (1 year)*
  - Applied NLP to optimize the client data curation process of the bank resulting in an speed improvement of over 15%, and created automated reports of the customer databases using Python, R, SQL and Hadoop.
- **Senior Data Scientist at INFOTEC**
  - *May 2018 - January 2019 (9 months)*
  - Developed and implemented an algorithm to index the publications of over two million scientists supported by the National Council of Science and Technology using NLP, Python and Linux. This algorithm is still currently in use by the National Archive of Science.

- **Graduate Statistics Professor at INFOTEC (Contract Full-time)**
  - *January 2018 - January 2019   (1 year)*
  - Developed teaching materials, delivered lectures and evaluated groups of over 30 students of the Master's Degree in Data Science.
- **Data Scientist at INFOTEC**
  - *November 2016 - November 2017   (1 year)*
  - Developed automated indicators to monitor 300,000 databases of the Mexican Open Data Portal to enable their use by Mexican citizens using R and Python.

Publications:
- *ROUGE-SciQFS: A ROUGE-based Method to Automatically Create Datasets for Scientific Query-Focused Summarization.* **COLING 2025**.
- *QuOTeS: Query-Oriented Technical Summarization.* **ICDAR 2023**.
- *Post-OCR Document Correction with Large Ensembles of Character Sequence-to-Sequence Models.* **AAAI 2022**.
- *A Web Platform for Collaborative Semi-Automatic OCR Post-Processing.* **AGRANDA 2021**.
- *Unsupervised Document Summarization using Pre-Trained Sentence Embeddings and Graph Centrality.* **Scholarly Document Processing Workshop at NAACL 2021**.
- *Detection of Phonetically Similar Words in the Spanish of Central Mexico.* **Spanish Journal of Applied Linguistics 2020**.

Open-Source Projects:
- [Pytorch Beam Search](): A lightweight implementation of Beam Search for sequence models in PyTorch.
- [FIONA](): A topically-aware, searchable and accessible system for the NASA Technical Report Server.

Languages:
- Spanish: 100% (Native)
- English: 7.5 in IELTS-Academic

Awards and Hackathons:
- **Global Finalist Honorable Mention**, [NASA International Space Apps Challenge](), 2022.
- **First Place**, [NASA Atlantic Canada Space Apps Challenge](), 2022.
- **First Place**, [DETESTS Shared Task-2 at IberLEF](), 2022.
- **Best Presented Proposal**, [ResearchNS Student Challenge](), Canada, 2022.
- **Gold Medal**, Mexico City Amateur Chess Championship, 2018.
- **Bronze Medal**, VIII International Logic Olympiad, Mexico, 2011.
- **Bronze Medal**, II International Junior Science Olympiad, Indonesia, 2005.