Alaa Abdelgawad

📧 alaa.abdelgawad@mail.mcgill.ca ☎ (514) 999-5868 in LinkedIn: lnkd

Professional Summary

Accomplished Data Scientist with an MSc and 3+ years of experience in NLP, fraud detection, and product analytics. Proven expertise in developing scalable data processing pipelines, leveraging LLMs, transfer learning, and advanced ML techniques. Skilled in end-to-end project execution, mentoring teams, and delivering data-driven solutions that drive significant business impact.

Skills

Programming languages: Python, SQL, PySpark, R, Matlab
Data Visualization: Power BI, Streamlit, Matplotlib, Plotly
Cloud & Data management: Snowflake, DVC, MLFlow, Airflow, Git/Github, Azure, AWS(Sagemaker, S3)
AI frameworks: TensorFlow, PyTorch, Keras, Hugging Face

Education

McGill University   **Montreal, CA**
Computational Neuroscience MSc 3.83 GPA   2019–2021
PUBLICATION: Abdelgawad, A. et al. "Predicting longitudinal brain atrophy in Parkinson's disease using a Susceptible-Infected-Removed agent-based model. Network Neuroscience 2023

University of Arkansas   **Arkansas, U.S.**
Biomedical Engineering BSc 3.94 GPA (summa cum lauda)   2015–2019

Related Experience

Intact Financial Corporation   ***Toronto, CA***
Data Scientist II   Dec 2022– present

– Led the development of a product analytics tool leveraging a GPT-4-based LLM for text classification post-OCR processing with Textract, achieving a 60% reduction in manual annotation time

– Implemented CLIP with zero-shot learning for image classification, automatically categorizing insurance claim images and reducing manual processing time by 30%

– Utilized transfer learning, particularly BERT-based models, to train few-shot learning models for classification and semantic tasks of medical notes, resulting in a 15% improvement in accuracy.

– Built a classification pipeline, integrating MLFlow for model tracking and result analysis, The model had a significant impact on the business, resulting in $700k savings.

– Collaborated with business analysts and SMEs to develop a detailed annotation plan outlining annotation guidelines, data sampling strategy, and quality control measures with 95% inter-annotator agreement.

Data Scientist I   Sep 2021–Dec 2022

– Developed a product-focused Catboost model for fraud claim classification, increasing accuracy by 10% over baseline, which significantly improved product decision-making.

– Built ETL data pipelines using Airflow and Snowflake, with a focus on MLOps practices and cloud integration (AWS), leading to a 30% reduction in data retrieval time.

– Created interactive visualizations and dashboards using PowerBI and Streamlit to meet KPIs and analytics needs.

– Conducted extensive feature preprocessing/engineering using imputation, encoding, scaling, Recursive feature elimination, PCA, Shap techniques.

FluidAI   ***Waterloo, CA***
Data Scientist Intern   Jan–April 2021

– Developed an HTTP API wrapper in Python to fetch clinical trial data and implemented data ingestion pipeline from various sources such as Azure, and unstructured documents, ensuring seamless data integration.

– Built supervised machine learning pipeline to classify surgical leaks based on processed time-series biosensor data.

McGill University **Montreal, CA**
Research Assistant Sept 2019–Oct 2021
- Developed an agent-based SIR model to predict the progression of Parkinson's Disease (PD).
- Applied multivariate statistical methods such as GLM, Linear mixed effect model and weighted z-scores to study PD progression longitudinally; Image-Processed and analyzed MRI data to quantify deformation in PD.

Coding Instructor Jan-May 2021
- Led 2 weekly coding workshops of 12 students each; Organized full-day data analytics hackathon for 30 students.
- Taught intro to data analytics class using python main libraries such as NumPy, Pandas, Matplotlib, Seaborn, Sklearn.

AI4Good Lab **Montreal, CA**
Data Science Mentor *June–Aug. 2022*
- Mentored a team of 5 students in developing smart chatbot for mental health from ideation to implementation.
- Conducted weekly sessions to guide students to advance their career in data science academically and professionally.