

ARMAN HAGHANIFAR

 arman.haghanifar@gmail.com

 Toronto, ON

 416-389-7049

 [Linkedin](#)

 [Github](#)

 [Google Scholar](#)

WORK EXPERIENCE

Data Scientist

Gore Mutual Insurance

-  Nov 2024 – Present
-  Toronto, ON (Hybrid)
- Built and optimized Databricks ETL pipeline integrating >20M records from Unity Catalog & Azure Data Lake, consolidated fact/dim tables into OBT at endorsement granularity, engineered 52 new external features (expanding into 426 columns), reduced runtime by 24%, and improved reconciliation accuracy (e.g., earned exposure error reduced from -4.38% to 0.06%) using **PySpark, Python, and DBSQL**.
 - Built Databricks feature store, expanding dataset into 875 engineered features, incorporating multi-level stratified splits, imputations, downsamplings, and feature engineering; optimized pipeline runtime from 4.5 hrs to <1 hr to support downstream loss cost modeling.
 - Developed CatBoost regression frequency & severity models with tuned Tweedie loss functions using **MLflow**. Improved performance by incorporating engineered/external features and hyperparameter tuning, reducing mean Poisson deviance by 27% and mean Gamma deviance by 22% vs. benchmark. Designed comprehensive evaluation framework including lift charts, QQ plots, SHAP/PDP explainability, and advanced diagnostics (risk profiling, simplicity bias detection, KL divergence, etc.) to ensure robust, interpretable, and business-aligned loss cost predictions.
 - Collaborated with Actuaries to develop unconstrained CatBoost-based loss cost model for commercial auto (IRCA), utilizing 6 validated IRCA-specific features and leveraging aggressive regularization to overcome overfitting, benchmarked performance against technical premiums and main model. Demonstrated improved lift and risk segmentation through double-lift chart and Lorenz curve, with 4.8% increase in Gini index, indicating gains over actuarial baseline.

AI Engineer

Bluesense AI

-  Jan 2024 – May 2024
-  Vancouver, BC (Remote)
- Developed a robust Face Mesh algorithm for accurate facial landmark detection and expression optimization using **Python, OpenCV, MediaPipe, and DeepFace** packages.
 - Implemented Enhanced Super-Resolution GAN (ESRGAN) model to upscale face image resolution, improving clinical hyper-pigmentation diagnosis accuracy by 7.5%.
 - Collaborated with a team of 5-10 freelancers to research and apply state-of-the-art methods for skin disease detection.

Data Scientist II

GoDaddy Inc

-  Mar 2022 – May 2023
-  Toronto, ON (Remote)
- Ranking optimization of the main domain search engine results using multimodal RNN-CNN model. Using **AWS S3, Athena, SageMaker, API Gateway and Lambda** for EDA, training, and inference.
 - Increased training-set size for specific user path by approximately 2.5x, eliminated correlating features and added new ones by feature engineering. Improved NDCG score on top-10 recommendations by approximately 30% compared to previous model.

Senior Research Assistant

University of Saskatchewan

-  Oct 2019 – Apr 2022
-  Saskatoon, SK (Hybrid)
- Developed end-to-end pipeline for collecting/labeling radiographs and dental caries classification via ensemble transfer learning
 - Automated segmentation of teeth in dental radiography using image processing techniques and genetic algorithm
 - Performed Visual screening of COVID-19 pneumonia in a large curated database of frontal chest x-rays using deep learning

TECHNICAL SKILLS

Programming Languages:		Python Libraries:		Other Tools:	
Python	●●●●●	TensorFlow 2.x/Keras	●●●●●	Git	●●●●●
Postgres/MySQL	●●●●●	PyTorch	●●●●●	Docker	●●●●●
MATLAB	●●●●●	Scikit-Learn, MLib	●●●●●	PowerBI	●●●●●
C/C++	●●●●●	Numpy, Scipy, Pandas	●●●●●	Amazon AWS	●●●●●
Java	●●●●●	PySpark	●●●●●	MS Azure	●●●●●
Bash Script	●●●●●	OpenCV, Scikit-Image	●●●●●	Apache Airflow/Kafka	●●●●●

SELECTED PROJECTS

Academic Recommender System [GitHub](#)

- Developed a semantic similarity-based academic recommender using word embeddings to match students and professors, based on Gensim's word2vec and transformers like BERT for generating vectors
- Implemented a local Flask API for testing the system, allowing users to input strings and receive the top n most similar results

Multi-Objective Optimization using epsilon-constraint method with Pyomo [\[GitHub\]](#)

- Optimized removal efficiencies and operating cost based on variables initially calculated via response surface methodology
- Utilized Pyomo package with Interpoint Optimizer as the backend, Github Actions as CI/CD platform, and PyLint as code analyzer

Localization, Segmentation, and Classification of Scars in Swine Carcasses

- Implemented carcass detection using YOLO v4 and tracking it through the video. Utilized a Siamese network to find best view angles
- Developed Segmentation of scars via U-net and classification via a feed-forward neural network
- Utilized TensorFlow Lite for edge deployment, reducing inference time by 29% and enabling near real-time processing a laptop system connected to the camera

Canine Disease Diagnosis based on Chemical Tests [\[GitHub\]](#)

- Conducted preprocessing and analysis of chemical tests (blood-work data) performed at the College of Vet. by PDS Inc.
- Applied XGBoost and MLP for multi-label and multi-class classification, and visualized results and feature importance via Shapley
- Implemented a web-based application using Flask with API for integration with EHR systems, enabling veterinarians to input test results and receive diagnostic predictions, significantly reducing data entry time

COVID-19 Pneumonia Screening in Frontal Chest X-rays [\[GitHub\]](#)

- Transfer Learning based on CheXNet and visualization of the results using Grad-CAM and LIME to evaluate localization
- Compilation of the largest curated dataset of frontal chest x-rays from COVID-19 patients

Electric Vehicles Fueling Data Analysis with Machine Learning [\[GitHub\]](#)

- Comparing ML algorithms in terms of consumption rate deviation classification and driving range estimation
- Building a dataset of fueling records by crawling data from SpritMonitor website using Selenium

EDUCATION

MSc in Electrical & Computer Engineering - Biomedical University of Saskatchewan (USask)

📅 Sep 2019 – Feb 2022

📍 Saskatoon, SK

- Research Area: Medical Image Processing, Application of Deep Learning/ Machine Learning in Healthcare

BSc in Computer Engineering Iran University of Science and Technology (IUST)

📅 Sep 2013 – Mar 2018

📍 Tehran, Iran

- Research Area: Machine Learning, Fuzzy Logic, Evolutionary Algorithm, Computer-Aided Diagnosis

RECENT MOOCS

LLMs: Application through Production EdX

📅 May 2024

Building Recommendation Engines in Python DataCamp

📅 Oct 2023

Build Machine Learning Pipeline using Dataiku Coursera

📅 Jun 2023

Software Engineering for Data Scientists DataCamp

📅 Apr 2022

Advanced SQL Kaggle

📅 Dec 2021

Introduction to MLOps Coursera

📅 Jul 2021