

# ESMAEIL (ISHMAEL) REZAEI

Immediately Available | Green Card Holder

+1 (857) 376-9504

erezaei@umassd.edu

411 Allen st, New Bedford, MA-02740

<https://ishmaelrezaei.github.io>

<https://github.com/ishmaelrezaei>

<http://www.linkedin.com/in/esmaeil-rezaei>

## EDUCATION

### Ph.D. in Machine Learning and Data Science

University of Massachusetts, Dartmouth, MA, US

Relevant Courses: Machine Learning & AI, Numerical Linear Algebra, Numerical Optimization, Advanced Data Mining

Thesis: HYBRID - A Dimensional Reduction Algorithm for Big Data

Sep 2021 - Jun 2024

GPA 3.95

### M.S. in Data Science

University of Massachusetts, Dartmouth, MA, US

Relevant Courses: Database Design, Numerical Linear Algebra, Machine Learning, Advanced Mathematical Statistics

Thesis: OPGPR: Online Parametric Gaussian Process Regression

Sep 2021 - May 2024

GPA 4.0

### M.S. in Industrial Engineering

University of Science and Technology of Mazandaran, Behshahr

Relevant Courses: Operations Research, Supply Chain Management, Economic, Quantitative Methods, Management (MIS)

Thesis: Bayesian Theorem for Maintenance Scheduling Reliability Optimization of Correlated Complex Systems

Sep 2013 - Sep 2015

GPA 3.61

### B.S. in Mathematics

Velayat University, Sistan & Baluchestan

Relevant Courses: Operations Research, Statistics and Probability, Numerical Analysis, Algebra, Differential Equations

Jan 2009 - Jul 2012

GPA 3.05

## SKILLS

**Programming Languages:** Python • SQL • R • SAS • MATLAB • C++

**Libraries:** OpenAI, LangChain, Scikit-Learn • PyTorch • TensorFlow • Keras • Pyspark • Matplotlib • Plotly • SciPy

**Big Data and Cloud Technologies:** AWS (SageMaker) • Hadoop

**Data Analysis and Visualization Tools:** Power BI • Tableau • Excel • PowerPoint • SPSS

**Database Management:** NoSQL (MongoDB) • Relational (Microsoft SQL Server, SQLite)

**Other skills:** Regression • Random Forests • Support Vector Machines • Bagging • Boosting • Decision Trees • Data Mining

• Stochastic Processes • Predictive Analytics • Statistical Modeling • Clustering Optimization • Large Language Models (LLMs)

• Computer Vision • Git • GAN • CNN • ETL

## WORK EXPERIENCE

### Data Manager

Care New England

RI, US

Jun 2024 - present

- Handling the project of **clinical data migration** from REDCap to Ripple Science.
- Creating data pipelines** for clinical data and performing extract, transform, and load (ETL).
- Performing data analysis, cleaning, exploratory data analysis (EDA), and hypothesis testing.
- Mentoring staffs to learn the registry protocols, and work with both Ripple Science and REDCap.
- Skills: Python, R, SQL, Airflow, ETL, Airflow, Snowflake, Hevo, Excel, PowerPoint

### Data Scientist Research Assistant

University of Massachusetts

MA, US

Sep 2021 - Jun 2024

#### Research Assistant:

- Achieved 94% accuracy in creating an optimized a driver assist system** to detect aggressive drivers for **US Department of Transportation**. Did feature engineering, formulation, optimization, simulation, ETL. Built the model in Python and used C and Delphi to extract, load, and transform data by simulator.
- Developed an optimized model addressing the scalability** of predictive models for big data, achieving high accuracy for noisy datasets while reducing processing time and costs.
- Achieved 7700% improvement in PCA speed for large data and solved** the scalability problem. The model successfully aggregated with regression models and predicted utility demand of Commonwealth Edison Inc.
- Performed EDA and data cleaning for 180 million data points** from the **Michelin** company using PySpark, Polars, and SQL, resulting in weekly presentations and collaboration with MIT faculty for the next phase of the project.

#### Teaching Assistant:

- Taught programming** of engineering problems in MATLAB to three classes of graduate students.

### Data Scientist

Behin Tadbir Paradise

Yazd

Feb 2020 - Jul 2021

### Founder and Manager

Sharifnagar

Tehran

Nov 2015 - Feb 2020

- **Successfully managed and led a team** of 60 professionals in developing and operating an e-commerce website.
- **Generated significant revenue** by selling high-quality tutorials in Data Science, Engineering, and Mathematics.
- Implemented effective strategies to optimize website performance and user experience, resulting in increased customer satisfaction and repeat business.
- Developed and maintained strong relationships with clients, ensuring alignment with business objectives and customer needs.

## PROJECTS

### LLM GenAI to extract information from PDFs and links

[Link](#)

- **Successfully built a LLM to extract information from multiple links and PDF files.** Utilized libraries such as LangChain, Streamlit, and OpenAI to design this Q&A system.

### Parameter Tuning and ML Model Selection (Fraud Detection/House Price Prediction)

[Link 1](#), [Link 2](#), [Link 3](#)

- **Achieved 98% accuracy in credit card fraud detection** using machine learning algorithms, **and 0.06 MSE** in house price prediction, showcasing expertise in regression and classification techniques. Employed an ensemble method by stacking the selected algorithms among variety of regression/classification algorithms.

### Diabetic Retinopathy and Pneumonia Diagnosis by (CNN, PyTorch, TensorFlow, keras)

[Link 1](#), [Link 2](#)

- **Achieved 96% accuracy by employing a CNN pipeline in PyTorch** for diabetic retinopathy prediction, demonstrating expertise in deep learning, medical imaging analysis, data-driven healthcare solutions, and adept problem-solving skills ([Link 1](#)). Additionally, completed a similar project for pneumonia diagnosis using Keras, achieving 82% accuracy and 86% recall ([Link 2](#)).

### Healthcare Exploratory Data Analysis in R

[Link](#)

- Performed exploratory data analysis and data cleaning (data wrangling) on a synthetic healthcare dataset using **R Studio**. Leveraged **ggplot** and tidyverse for comprehensive data analysis and employed statistical analysis techniques to derive insights.

### Pharmacy Database Design

[Link](#)

- Developed a Pharmacy Database, including ER diagram and relational schema design. Implemented the database utilizing MS SQL for relational data storage, MongoDB for NoSQL data management, designed a user-friendly GUI for frontend interaction.

### EDA Using Pyspark and Polars

[Link](#)

- Successfully managed a vast dataset comprising 514M observations, employing advanced techniques with both **Pyspark** and **Polars** libraries for two comprehensive EDA projects, driving actionable insights and informed decision-making.

## PRESENTATIONS

### University of Illinois Urbana-Champaign

*Chicago, IL, US, 2024*

Presented 'Online Parametric Gaussian Process Regression' at the ASCE Engineering Mechanics Institute, University of Illinois Urbana-Champaign, demonstrating advanced research in predictive modeling

### Georgia Institute of Technology

*Atlanta, GA, US, 2023*

Presented 'Handling High-dimensional Data to Predicting Utility Demand for ComEd' at the ASCE Engineering Mechanics Institute, Georgia Institute of Technology, demonstrating advanced research in predictive modeling for saleable data.

### Johns Hopkins University

*Baltimore, MD, US, 2022*

Presented 'HYBRID: A Dimensional Reduction Algorithm for Big Data' at the ASCE Engineering Mechanics Institute, Johns Hopkins University, demonstrating advanced research in machine learning.

## AWARDS

Fully funded with a tuition waiver for the entire duration of both PhD and MS degrees, supported by:

- Michelin North America, Inc
- US Department of Transportation
- National Science Foundation

## CERTIFICATIONS

*Deep Learning: Convolutional Neural Networks in Python*

[Link](#)

*AWS SageMaker Practical for Beginners | Build 6 Projects*

[Link](#)

*Generative Adversarial Networks (GANs) A-Z*

[Link](#)

*SQL Essential Training*

[Link](#)

*Data Visualization in R with ggplot2*

[Link](#)

*R for Data Science: Analysis and Visualization*

[Link](#)

*Tableau for Data Scientists*

[Link](#)

*Python for Data Visualization*

[Link](#)

*Generative AI with Large Language Models (LLMs) | In progress*

[Link](#)