Javad Ebadi | Data Engineer and Data Scientist

Telus, Calgary, AB, Canada
+1(xxx)-xxx-6994 • javad@javadebadi.com • javadebadi.com
https://github.com/javadebadi

Skills

Proficient: Python, SQL, Git, GitHub, Data Science, Machine Learning, Django, Pandas, Matplotlib, Docker, SQLAlchemy, Django Rest Framework, Statistics, Linear Algebra, scikit-learn, Neo4j (Graph Database), Cypher, Apache Airflow, AWS Cloud, Amazon Relational Database Services (RDS), Amazon S3, Amazon Lambda, Amazon EC2, AWS Identity and Access Management (IAM), Amazon Cognito, AWS Serverless Application Model (SAM), Amazon DynamoDB

Moderate: C++, PL/SQL, Deep Learning, HTML, CSS, Tableau, JavaScript, Amazon Comprehend

Familiar: Mathematica, Tensorflow, React, Google Cloud Platform, Accounting, Algorithmic Trading, Amazon Rekognition

Work Experiences

TELUS Calgary, AB, Canada

Data Analyst and Data Engineer - 911 Mandatory Support Oct. 2022 - Present

I joined TELUS as a data analyst and data engineer and helped to improve performance, observability, security, and privacy on data collection and processing using both on-prem and cloud solutions

- Optimize data pipeline codes for faster performance using appropriate data structures and algorithms
- Collaborate on developing codes and increasing data security and data privacy in the team
- Research on using new tools such as ELK and fit them into our organization's architecture
- Server and database administration
- Develop automation codes using Python, Bash, and Selenium to replace the manual data flow processes
- Develop data mining codes to extract information from log files for automation and audit purposes
- Develop management reports using Tableau

Technolgoies: Python, SQL, PL/SQL, Git, Bash, Docker, Tableau, Oracle, Airflow, Java, Splunk, AWS Cloud, Google Cloud Platform

Human Knowledge Graph Calgary, AB, Canada

Solutions Architect and Developer - Calgary, AB Sep. 2022 - Present

I am the Founder and Solution Architect of the Human Knowledge Graph (Knaph) project. I build this project to practice different AWS skills that I don't have opportunity to use in my official job. Furthermore, this is a very ambitious and long-term project to create an interconnected content of human knowledge.

- Research and develop content
- Use AWS Identity and Access Management (IAM) to manage accounts
- Implement AWS Cognito for Customer Identity and Access Management (CIAM)
- Design a static website for the project using Amazon S3 for hosting and AWS CloudFront for caching
- Use Amazon Certificate Manager (ACM) to issue certificates for secure (HTTPS) connections
- Develop serverless codes using AWS Serverless Application Model (SAM), AWS API Gateway, AWS Lambda functions as the runtime, and AWS DynamoDB tables for database
- Develop CloudFormation and SAM templates for DevOps operation
- Use GitHub hooks, AWS CodeCommit and CodePipeline for CI/CD
- Use AWS CloudTrail to store events for audit and security purposes and implement AWS Athena to analyze data
- Use Django to develop a static website generator system considering SEO optimization best practices

Technolgoies: Django, Research, Docker, Amazon Relational Database Services (RDS), Amazon Lambda, Amazon EC2, AWS Identity and Access Management (IAM), Amazon Cognito, AWS Serverless Application Model (SAM), Amazon DynamoDB, Amazon CloudFront, AWS CloudFormation, AWS CodeCommit, AWS CodePipeline, AWS CloudTrail, Amazon Certificate Manager (ACM)

Mercator AI Calgary, AB, Canada

Data Engineer - Data Engineering Sep. 2022 - Dec. 2022

Mercator AI is an exciting startup with the goal to revolutionize construction using data and AI. I joined Mercator AI as a data engineer to develop data ingestion pipelines.

- Develop ETL Pipelines using Python and Bash working with AWS S3
- Develop codes to extract and clean data from the following sources: APIs, Web pages, PDF Files, Spreadsheets
- Research on available solutions for different problems in the startup
- Use Apache Airflow to orchestrate data ingestion pipelines

Technolgoies: Python, SQL, Git, Django Rest Framework, Bash, Docker, Amazon S3, Apache Airflow, AWS Cloud, Web Scraping, Amazon Relational Database Services (RDS), Amazon Lambda, Amazon EC2

Middle East Bank Tehran, Iran

Junior Python API Developer and Data Scientist - Systems Analysis Nov. 2020 - Sep. 2022

I was a backend developer and a data scientist who used technologies such as Python, Django, Oracle, and Machine Learning to do research on financial subjects and develop and deploy financial apps.

- Design data and application architecture for Business Financial Management (BFM) system, a service offered by Bank to its small business customers
- Build ETL pipelines and End-to-End machine learning pipelines to train on new data and deploy it to production
- Develop and deploy data model for bank governance using Neo4j graph database
- Design, implement and evaluate different machine learning models and optimize hyperparameters
- Sophisticated implementations of financial metrics such as NPV, IRR, and XIRR
- Use Natural Language Processing (NLP) techniques to classify customers' transactions based on the description of transactions
- Develop portfolio dashboard and other advanced reports using Pandas and Oracle PL/SQL
- Develop Interbank Transactions application for the Treasury Management System

Technolgoies: Python, SQL, PL/SQL, Git, Machine Learning, Data Science, Django, Angular, Research, Presentation, Django Rest Framework, SQLAlchemy, scikit-learn, Oracle, SSRS, Neo4j (Graph Database), Cypher

Institute for Research in Fundamental Sciences Tehran, Iran

PhD Researcher - CERN Group Feb. 2017 - Aug. 2020

I was involved in the CMS experiment of CERN at IPM. I used the 2016 data to write and debug my code. The code is written in a way that can be easily updated according to continuous changes in the data analysis criteria.

- Read and transform CMS data using the High-Performance Computing infrastructure of CERN
- Develop a data analysis code using C++, Python, and Bash to define physics objects and apply selection rules to particles using 2016 data
- Hypothesis (A/B) testing using the inferential statistics methods
- Develop machine learning models to create better classifiers

Technolgoies: Python, Linux, C++, Git, Data Science, Statistics, Machine Learning, Deep Neural Networks, Research, LATEX, root-cern, CMS grid, Bash

Teaching Experiences

Mentoring Program: Become a Python Web Developer with Django. [videos][material]	2023
Data Science Series 1: An Introduction to Python for Scientists. [videos][material]	2021
Git and GitHub for Scientists. [videos][material]	2021
An Introduction to Programming with Python.	2019
Introduction to ROOT-CERN Data Analysis Framework. [videos][material]	2018
Introduction to MadGraph Software for Monte Carlo Simulation of Particles Collision. [videos]	2018

Research Experiences

Reality as Resistance using Higgs as a Case Study 2020 - 2021

I was involved in philosophy of science project. There are criteria to determine what is real in science. For example,

robustness is one criterion that states that if one can engage with a physical object in various different ways, then we can say that is real. However, there are examples such as the Higgs boson that is produced only at colliders and therefore is not robust. In this project, we try to give a new criterion for reality such that the Higgs could be considered as real. Our idea was reality as resistance. Higgs is real since it is reproducible and resisted to be excluded.
Technolgoies: Research

Axion-like Particles at LHC [link] 2018 - 2019

Axions and axion-like particles are well-motivated particles in many beyond the Standard Model scenarios. We suggested new probe channels to study these models at the HL-LHC. Our results showed that LHC searches for axions are complementary to low-energy experiments. Most of the constraints on ALP come from low-energy experiments, however, we show that there is a region that was not constrained by the low-energy experiment and we could constrain that region by LHC data.
Technolgoies: C++, Git, Research, LATEX, Presentation, Monte Carlo Simulation, MadGraph, root-cern, Delphes, Pythia, Statistics

Constrain the Top Quark Coupling to U Quarks in FCNC Model [link] 2017 - 2018

Top quark FCNC interactions are useful tools to constrain new physics models. We used Monte Carlo simulation to generate a sample of events for various couplings of top FCNC models. We did realistic showering and hadronization on generated data and then simulated detector effects to prepare a dataset that can be compared with real data from LHC. We used statistical inference to set bounds on parameters of the top FCNC models. Our results improved the previous bounds on u-t coupling by 20%.
Technolgoies: Data Science, Monte Carlo Simulation, MadGraph, root-cern, Delphes, Pythia, Statistics

A Review on Dark Matter Literature [link] 2016 - 2017

I thoroughly reviewed Dark Matter (DM) physics literature, read over 30 papers and books. This review was a part of the Ph.D. qualification exam. I reviewed theoretical topics such as DM production in the early Universe, DM detection methods, and calculated analytical formulas related to them. The final output of this activity was my first 80 pages document written in English.
Technolgoies: Research, LATEX, Presentation

Calculate Hawking Radiation of Black Holes by Solving General Relativity Equations 2015 - 2015

As an MSc student, I was working on the theoretical physics of black holes for my thesis. The problem was about the horizons of a black hole that was radiating. I have used xAct and Mathematica to do some tensor calculations and to solve General Relativity equations. I was computing the rate of Hawking radiation for various vacuums of black holes.
Technolgoies: Mathematica

Projects

Python Wrapper for INSPIREHEP API (pyinspirehep package) Open Source

homepage: https://pypi.org/project/pyinspirehep/ Jan. 2022 - Feb. 2022

Inspirehep.net is a platform which contains scholarly data about high energy physicists. The Inspirehep has provided an API to get its data. We have developed a Python package to fetch data from Inspirehep API.

Social And Volunteer Experiences

Administrator of a Social Media Channel for Particle Physics 2017 - Present

In September 2017, I created a Telegram (Social Media app) channel to post important arXiv papers, news, and other related topics to particle physics such as machine learning. A few months later, my friend, Majid Ekhterachian from the University of Maryland, joined the channel as an administrator. Our channel has over 500 subscribers which are mainly graduate students and faculty in physics. The contents are in Persian and English languages and our mission is to spread the news about particle physics in our country. The link for this media is https://t.me/IPMParticlePhysics . Since 2020, Reza Ebadi from the University of Maryland and Mohammad Hassan Hassanshahi from Imperial College London are joined to channel as administrators.

Education

PhD in Particle Physics, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran 2015 – 2020

MSc in Black Hole Physics, Sharif University of Technology, Tehran, Iran 2013 – 2015

BSc in Physics, Amirkabir University of Technology, Tehran, Iran 2008 – 2013

Certificates

Introduction to Relational Databases (RDBMS), by Lin Joyner, Rav Ahuja, Rose Malcolm and Sandip Saha 2022
Joy, IBM on Coursera. Certificate earned at 2022-09-08 (96.0%) [Link]

Introduction to Data Engineering, by Rav Ahuja and Priya Kapoor, IBM on Coursera. Certificate earned at 2022-08-07 (95.2%) [Link] 2022

Hands-on Introduction to Linux Commands and Shell Scripting, by Rav Ahuja, Ramesh Sannareddy and Sam Prokopchuk, IBM on Coursera. Certificate earned at 2022-07-24 (98.0%) [Link] 2022

Mathematical Biostatistics Boot Camp 1, by Johns Hopkins University on Coursera. Certificate earned at 2022-07-06 (90.36%) [Link] 2022

Graph Data Modeling Fundamentals, by Neo4j Academy on Neo4j. Certificate earned at 2022-05-08 (100.0%) [Link] 2022

Cypher Fundamentals, by Neo4j Academy on Neo4j. Certificate earned at 2022-04-27 (100.0%) [Link] 2022

Neo4j Fundamentals, by Neo4j Academy on Neo4j. Certificate earned at 2022-04-27 (100.0%) [Link] 2022

Blockchain Basics, by Bina Ramamurthy, The State University of New York on Coursera. Certificate earned at 2022-02-04 (95.0%) [Link] 2022

Natural Language Processing with Classification and Vector Spaces, by Younes Bensouda Mourri and Lukasz Kaiser, deeplearning.ai on Coursera. Certificate earned at 2021-01-07 (100.0%) [Link] 2021

Business English: Management and Leadership, by Wanda Huber, Andrea Haraway and Jenny Young, Arizona State University on Coursera. Certificate earned at 2020-04-21 (88.0%) [Link] 2020

Databases and SQL for Data Science, by Rav Ahuja, IBM on Coursera. Certificate earned at 2020-02-04 (100.0%) [Link] 2020

Data Science Methodology, by Alex Aklson and Polong Lin, IBM on Coursera. Certificate earned at 2019-12-30 (100.0%) [Link] 2019

Data Analysis with Python, by Joseph Santarcangelo, IBM on Coursera. Certificate earned at 2019-12-29 (100.0%) [Link] 2019

Open Source tools for Data Science, by Polong Lin, IBM on Coursera. Certificate earned at 2019-12-07 (100.0%) [Link] 2019

Python for Data Science and AI, by Joseph Santarcangelo, IBM on Coursera. Certificate earned at 2019-11-24 (100.0%) [Link] 2019

What is Data Science?, by IBM on Coursera. Certificate earned at 2019-11-22 (99.0%) [Link] 2019

Neural Networks and Deep Learning, by Andrew Ng, Stanford University on Coursera. Certificate earned at 2019-10-21 (100.0%) [Link] 2019

Machine Learning, by Andrew Ng, Stanford University on Coursera. Certificate earned at 2019-09-26 (98.4%) [Link] 2019

The Unix Workbench, by Johns Hopkins University on Coursera. Certificate earned at 2019-09-11 (91.8%) [Link] 2019