

Data Engineer

PROFILE

I'm a Software engineer with profound interest and specializations in the domain of Data Sciences, Data Engineering and Warehousing. My experience lies in teaching machines and making data centric applications. Current role reprises in programming backend for a couple of compliance reliant products which consists of creating data pipelines, ETLs, and calculating risk scores in T-SQL and SSIS whilst clustering and classifying the results using Python based algorithms. A pythoneer, having command on numerous Python frameworks (Business Intelligence, Machine Learning, and Web Dev). I'm a passionate coder, always breaking and repairing my Linux, and trying to automate any task I can. A good team player with excellent communication skills altogether.

WORK EXPRIENCE

Qordata

Data Engineer (October 2020 – Present)



- Aggregate Spend (Generate and Monitor Compliance spend reports)
- Compliance Monitoring (Risk based assessment).
- Data Pipelines and Scripting using T-SQL.
- ETL packages using SSIS and C# scripts.
- DB/DWH development.
- Risk Score classification implemented using Fisher Jenks.
- Data mappings with fuzzywuzzy.
- Jira interactive dashboard with Plotly/Dash.
- Compliance Monitoring admin panel with Vue.js and Flask.

ProLeadSoft

Software Engineer (June 2020 – October 2020)

PROLEADSOFT

- Full Stack Developer.
- Particularly focused on the MERN and MEVN Stack.
- Developed backend routes in express.js and consumed them via the front-end.

NOTABLE PROJECTS WORKED ON

Compliance Monitoring

Risk Based Assessment

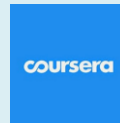
I'm currently assigned as one of the core developer for CM. A product that analyzes speaker and event spendings for Pharma companies and generates risk score based on the set of logics defined in the Risk Engine which works on the underlying principles of multiple Stats models including classifying the severity. CM follows an OLAP architecture with high dynamicity in mind, adjustable weights and algorithms. Current features being worked on include data mappings with NLP and OCR for invoice scanning.

Aggregate Spend

Compliance Insights

I was part of the team for Aggregate Spend at Qordata. Agg. Spend is one of the products that reprises its role in being an all in one solution for Pharmaceutical companies to monitor and generate reports for their compliance spendings. Created and implemented Data Cleaning, Data Pipelining in T-SQL along with Python scripts and SSIS packages. I was a vital part of scheduling, executing, developing, deploying, configuring and maintaining SQL Server Agent Jobs.

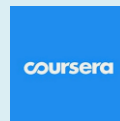
CERTIFICATIONS



Julia Scientific Programming (with Honors)



Deep Neural Networks with **PyTorch** (with Honors)



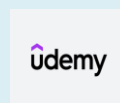
Neural Networks and Deep Learning



Mastering SQL Server 2016 Integration Services (**SSIS**)-Part 1



Interactive Python Dashboards with **Plotly and Dash**



Python REST APIs with **Flask**, Docker, MongoDB, and AWS DevOps



Time Series Analysis in SQL Server



Supervised Learning with **scikit-learn**



Data Processing in **Shell**



Introduction to Data Visualization with **Seaborn**

ATM Cash Prediction Using Time Series Approach

Final Year Project

This was our FYP which later proved to be a part of couple of research journals. Detailed work on Time Series analysis which included Amazon SageMaker, Deep Learning and Stats Models. Python notebooks implemented on real time ATM transactions to predict future withdrawals. Website developed in Django and Vue.js with Plotly Dash Dashboards. Deployment on AWS using EC2 instance connected to a S3 bucket.

BlockV: Peer to Peer Ride Sharing Service

Blockchain and Cryptocurrency

A blockchain enabled peer to peer ride sharing service running on the ethereum network with live transaction mining and ether payment facility, metamask wallet and escrow enabled. Smart contracts written in Solidity with a React.js frontend.

Satellite Image Detection

Deep Learning for Perception

Multi feature detection of a large set of satellite imagery using Unet and Resnet50 implemented in Python with the power of google collab.

Sarcasm Detection Application

Artificial Intelligence

This project involved semantic analysis on English sentences using Support Vector Machines and Natural Language Processing. Fully coded in Python with the front-end achieved through wx-Python.

Skincare Recommendation Android Application

Recommendation System

This app uses dataset from Kaggle and scrapped data from cosmetics websites to work a Tf-IDF vectorizer for content based filtering, and KNN and Decision trees for collaborative based filtering. The notebook also contains other approaches for POC including SVD. Backend APIs are based on Flask, Android application is made using Java with Android Studio whereas Firebase acts as the database and the middleware for relaying login information as well to serve the data to the application. Deployed onto Heroku with Firebase as the serverless backend.

SKILLS & INTERESTS

Dedicated Volunteer

Looking for an opportunity to volunteer. Never to say 'no' and always give back.

Captivated & Productive

Among a huge list of skills FAST has taught me is to be always doing something and be able to say "Yea! I did that".

Eager to Learn more

I'm always looking for an adventure, opportunities to learn something new and utilize them for my personal and professional career.

EDUCATION

BS Computer Science

National University of Computer & Emerging Sciences, Karachi, Pakistan.
2016 - 2020

COMPUTER SCIENCE SKILLS

Scientific Programming

PyTorch, Julia

Scripting

Vue.js, Web3.js, Bash, Solidity, Node.js, LaTeX, T-SQL

Programming

Python, C#, Docker

Python Frameworks

Flask, Django, Plotly/Dash, wx-Python, Flask & Django RESTful API

Database and Querying

T-SQL, Oracle, MySQL, MongoDB, Firebase, SSIS, DWH

Infrastructure

AWS: IAM, EC2, S3, SageMaker, SNS; Heroku, GitHub, GitLab, Bit Bucket, Jira

PUBLICATIONS

Towards optimal ATM cash replenishment using time series analysis

Journal of Intelligent & Fuzzy Systems - Oct 4, 2021

ATM Cash Prediction Using Time Series Approach

IEEE (iCoMET) - Apr 23, 2020

RELEVANT COURSEWORK

- Information Retrieval
- Data Science
- Probabilistic Programming
- Deep Learning for Perception
- Blockchain & Cryptocurrency
- Research Methods