Ahmad Nayyar Hassan . Jarvis Consulting

Around two years of working experience in the field of data analytics and software development. My expertise lies towards Java programming and big data. I have sound analytical and data engineering capabilities and hands on experience with database management using SQL and various machine learning algorithms using R and python. I have demonstrated strong communication and leadership skills through successful extracurricular initiatives and team lead role within a few months of professional experience.

Skills

Proficient: Java, Bash, SQL, Python, Spark, Docker, Keras, Machine Learning, Data Analysis, Statistics, Software Development

Competent: Maven, C++, MATLAB, Docker, Jenkins, Hadoop, Tableau, Mockito

Familiar: AWS, MongoDB, Spring Framework,ETL

Development Projects

Project source code:https://github.com/jarviscanada/jarvis_data_eng_ahmad/tree/develop

- **Linux & SQL::** Developed a cluster management solution using *Bash* to record hardware specifications and periodically monitor resource usages of individual nodes in a cluster of servers.Used *Docker* instantiated RDBMS *PostgreSQL* database for data persistence. Gathered and registered real time data in the database instance using *crontab* jobs.
- **Java Grep:** Implemented Unix grep command in Java using *Stream API* and *Lambda Expressions*.
- **Java JDBC:** Developed an application using *JDBC* along with *PostgreSQL* to provide abstraction using different design patterns.Used both *DAO(Data Access Object)* and *Repository pattern* to provide abstraction on database/data files incorporating complex joins and queries.
- **Twitter CLI App:** Used *MVC* architecture to implement a twitter application which allows user to Post, Delete and show Tweets through command line.Performed unit testing using *Mockito* framework and handled database interactions through *JDBC*.Managed dependencies of different layers within the application using ***Spring Boot*** framework.

Professional Experiences

Data Engineer,Jarvis, Toronto(2020-Present): Develop Linux and Java-based applications with ***maven*** automation tool. These applications handling databases through JDBC and interacting with ***REST*** APIs such as of Twitter. Unit testing is performed using Mockito framework and auto-configuration of dependencies files is done using ***Spring Boot*** framework.

Data Scientist, LUMS, Lahore (2017-2018): Performed exploratory data analysis using ***Tableau***, data pre-processing using Python and data base management using MySQL.Carried out telecom coverage analysis using Machine Learning Algorithms including SVR (Support Vector Regression), ANN (Artificial Neural Network) with ***Keras***, ***Tensor Flow*** and Outlier Detection through cluster-based approaches including K-Means clustering and DBSCAN. Managed and guided a team of full and part-time researchers.

Education & Academic Projects

University of Waterloo (2018-2019), Master of Engineering, Computer Engineering

- **Sale Price Prediction:** Designed an end to end Machine Learning *Pipeline*(*object-oriented*) to predict the price of item listings on online shopping platform.Developed front end ***REST API*** using ***flask*** framework, used Jenkins for designing Continuous Integration/Continuous Deployment (***CI/CD***) pipeline and deployed the model on the ***AWS-EC2*** platform using ***Docker*** for containerization. The proposed architecture achieved a RMSE (Root Mean Squared Error) of 0.241
- **Data Mining on Yelp Dataset:** Analyzed Yelp Data to find patterns and perform customer analytics. Set-up and stored data in MySQL server database. Wrote complex and nested queries to perform data cleaning and extract meaningful customer information. Performed ***CRUD*** Operations, imposed sanity and consistency conditions. Added

indexes for efficient data querying and created ER .Implemented both the ***Decision Tree Classifier*** and ***Naive Bayes*** Predictor as stored procedures in ***MySQL*** through an interactive GUI in Python.

- **Page Rank in PySpark:** Performed graph analysis using a graph of Gnutella server network.Implemented single source personalized page rank and handled dead end and spider trap problems using ***Spark***.
- **Web Traffic Time Series Forecasting**:Built a prediction model using LSTM neural network to estimate future web traffic on distinct Wikipedia pages.Performed in depth exploratory data analysis (***EDA***) to detect trends and seasonality in the data using pandas, seaborn and matplotlib. Performed differencing to achieve stationarity and used Augmented Dicky fuller test to identify it.Developed a **LSTM** model with baseline ***ARIMA*** model, the MAPE score improved significantly from 8.2 to 7.27 percent.

Lahore University of Management Science (2013-2017), Bachelor of Applied Science, Electrical Engineering

- **Optimized installation of security cameras at traffic intersections:** Developed Python and C++ computer-based application for the local police to optimize the number of street traffic cameras by solving the vertex cover.Integrated python scripts with C++ using inter process communication via piping, multi threading and solved vertex cover problem using SAT solvers.

Certificates & Awards & Activities

- Dean's Honor List ,2017
- **DataCamp**: Big Data Fundamentals via PySpark, Supervised Learning with scikit-learn, Unsupervised learning in Python, Deep Learning in Python, Statistical Thinking in Python, Python Data Science Toolbox (1 and 2)
- **Udemy:** The Ultimate Hands on Hadoop
- Director Debates Society, 2017