Arpit Raghav

MSc Big Data and High Performance Computing
University of Liverpool, United Kingdom

EDUCATION

- **MSc in Big Data and High-Performance Computing, University of Liverpool, United Kingdom**
  Core Modules - Big Data Analytics, Multi-Core and Multi-Processor Programming, Linear Programming, Applied Algorithmics, Data Mining and Visualization, Machine Learning, and Bioinspired Optimisation.
  Expected Grade - First Class with Distinction
- **Bachelor of Computer, SRM Institute of Science and Technology, India**
  Relevant Modules - Programming in C, Object Oriented Programming using C++ and Java, Relational Database Management System, Data Structures, Digital Logical Fundamentals, Statistical Methods, PHP & MySQL, Computer Networks, Data Mining and Warehousing.
  Grade - First Class

Technical Skills

- **Programming Languages:** ***Python***, C, C++, Java, R, Parallel Programming using MPI and Open MP
- **Programming Experience:** 2 Years in Python & 3+ Years in C & C++
- **Web Programming Technologies:** HTML, CSS, PHP
- **Data Visualization Tools:** ***Tableau, Power BI, Google Data Studio & Advanced Excel***
- **Data ETL Tools:** ***Apache Spark, Hadoop, Azure Data Factory(Currently Learning)***
- **Platforms:** Windows, Linux
- **General Tools -** ***Advanced SQL***, JIRA, MS Office, Barkla (UOL HPC resource)

Work Experience

University of Liverpool, United Kingdom **Aug 21 to Present**
Reference - Dr. Floriana Grasso - Module Coordinator (floriana@liverpool.ac.uk)
Teaching Assistant - Database Development and Design Systems

- Taking Labs and Tutorials for 2nd-year undergraduate CS students for the subject Database Development.
- Taking Tutorials for 1st year undergraduate CS students for the subject Design Systems
- Conducting Lab Sessions the labs for multiple batches of 120 students in total and grading their assignments.

Deloitte Consulting, India **Aug 20 to Aug 21**
Reference - Sidharth Singal - Senior Manager (sisingal@deloitte.com)
Associate Analyst

- Cleaning and Visualizing large ad datasets for the region of EMEA, Northern America, and Asia.
- Responsible for reviewing advertisements as part of the project for the client Verizon Media.
- Support in tracking and implementing project deliverables.
- Support with Queue monitoring and work allocation/prioritization on a daily basis and visualizing large datasets on Google Data studio.
- Intra-day real-time monitoring of service levels for all Pods. Drive real-time adherence to the expected capacity against the actual performance to achieve service levels and efficiency metric goals.
- Build relationships with clients to ensure there is no hurdle to achieving the goals.
- Assist Team Leads with capacity planning and management and Handle the team in the absence of the Team Lead as a backup lead.

Projects

Secure FTS in a healthcare network using BLOCKCHAIN (Bachelors Major Project)

- **Encryption Technique -** Asymmetric Encryption
- **Digital Signature Algorithm -** ECDSA, RSA
- **Hashing Algorithm -** SHA 256, Blowfish
- **Distributed FS and Privacy Guard -** IPFS & GNUPG

Created a secure File Transfer System over a local network, using multiple encryption and decryption and aligned them of inter Planetary File System to create a Blockchain-like security System.

Data Visualization and insights for a Foodchain (Bachelors Minor Project)

- **Tools Used - XAMPP, EXCEL & Google Data Studio**
- **Languages - PHP, MYSQL**

Worked with a small food chain to present them with business, sales, profit, and loss insights using easy-to-understand pictorial methods and graphs.

Projects (Continued) Note - The detailed description of each project is available on my github page

Eliminating Algorithmic bias in AI and ML systems (Literature Review)
Presented a literature review based on multiple research papers discussing the major and minor biases across multiple AI and ML platforms. Collated multiple proposed ways to eliminate/reduce the same and proposed ways to reduce the bias on these platforms. (The paper can be accessed by clicking on the name of the paper)

Social Network Analysis using Pandas Dataframe and Netwokx in Python
Performed Social Network analysis for a small social network of 209 nodes and 4000 edges. Visualized the network using multiple types of graphs and did analysis such as centrality, in degree, etc.

Principle Component Analysis
Performed multiple analysis operations using pandas data frame and performed principal component analysis on the open and close dataset of 500 companies for a span of 5 years

Binary Class Perceptron
Implemented Binary Class Perceptron algorithm implementation on Iris Dataset using python. Performed multiple operations as mentioned below-

- Trained and Tested the algorithm on train and test dataset
- Use the binary perceptron to train classifiers to discriminate between class 1 and class 2, class 2 and class 3, and class 1 and class 3 and reported the train and test classification accuracies for each of the three classifiers after training for 20 iterations.
- Extended the binary perceptron that you implemented in part 3 above to perform multi-class classification using the 1-vs-rest approach. Reported the train and test classification accuracies after training for 20 iterations.
- Added an l2regularisation term to your multi-class classifier implemented in part 4. Set the regularisation coefficient to 0.01, 0.1, 1.0, 10.0, 100.0 and compare the train and test classification accuracies.

Deep Learning
In the following project, I imported and trained an OpenAI gym game called Lunar Lander-V2. The following operations are performed in the following Depp Learning Project.

- Imported an OpenAI gym game names Lunar Lander V2
- Created a 4 layer Neural Network
- Connected the game to the Neural Network
- Implemented the deep reinforcement learning model

Reinforcement Learning

- Implemented Reinforcement learning while comparing a greedy vs 2 e greedy method on 10 armed testband using python
- Implemented Reinforcement learning for Comparing SARSA and Q-Learning in the cliff walking learning model using python

Natural Language Processing (Currently Doing)
In this project, I use natural language processing (NLP) techniques to automatically extract symptoms and diagnosis information from medical texts. More specifically, the steps in the project can include the following:

- Identify a suitable dataset comprising of medical abstracts for a set of pre-defined diseases
- Use named entity recognition tools to identify sentences with symptoms and diseases.
- Use an NLP method to identify the relationship between symptoms and disease.
- Steps 1-3 will result in a knowledge graph of symptoms and disease
- Use an appropriate method to store and query the knowledge graph

The dataset I have used are the synthetic abstracts of text-based data. The text-based dataset is MEDLINE's dataset which is freely available to use provided the proper references were made.

Covid 19 Tableau Dashboard
Visualized the reported spread and deaths around the globe due to Covid 19 using Tableau.

Personal & Contact Information

Email - arpitraghav13@gmail.com/arpitraghav9@gmail.com
Linkedin - www.linkedin.com/in/arpit-raghav-a45907190/
Phone - + 44(0)7824061065
Github - https://github.com/arp-raghav
Website - https://arp-raghav.github.io/#