

Anuja Jyothi Vijayakumar

anujajyothiv@gmail.com | +44 7404 619499 | London, United Kingdom, E1 4EP | <https://www.linkedin.com/in/anujajyothi> |

<https://github.com/AnujaJyothiVK>

SUMMARY

Graduated with distinction in MSc Big Data Science with the desire to shift towards the world of Data Science. Looking forward to joining an organization where my growth can be aligned with the organizational growth. Possess 6+ years of experience in Software industry in the software division of an international Bank. Have mentored and lead teams across different geographical regions including India, China, and UK.

SKILLS

Programming Languages	: Python, Familiarity with HTML, CSS, SQL, JSON
Big Data Frameworks	: Familiarity with Apache Spark, Hadoop
Deep Learning Frameworks	: Familiarity with Keras, PyTorch
Machine Learning Frameworks	: Scikit-learn
Visualization Libraries	: Familiarity with Matplotlib, Seaborn
Data Science Libraries	: Numpy, Pandas, Familiarity with NLTK
Data Modelling/Management	: SQL
Web Scraping	: Familiarity with Beautiful Soup
Development Tools	: Visual Studio Code, Jupyter Lab

EDUCATION

MSc, Big Data Science

September 2020 – September 2021

Queen Mary University of London

United Kingdom

Modules: Applied Statistics|Big Data Processing|Data Mining|Deep Learning and Computer Vision|Digital Media and Social Networks|Natural Language Processing|Neural Networks and NLP|Machine Learning

B.Tech, Information Technology

July 2003 – April 2007

Anna University

India

Relevant Modules: Object Oriented Programming|Probability and Statistics|Data Structures and Algorithms|Database Management System|Software Engineering|Program Design and Development|Object Oriented Analysis and Design|Software Quality Management|Total Quality Management

ACADEMIC PROJECTS

1. NLP project - Text Classification for Deception Detection in Amazon Reviews. – Pre-processed using Regex, trained using K fold cross validation and Linear SVC classifier, determined the precision recall accuracy and f-score. A comparison was done by adding additional features and different stemming and lemmatization process. Finally, did an exploratory analysis on the dataset using seaborn and Matplotlib to explore some of the linguistic and stylistic traits of the reviews and compared the two classes.
2. NLP project - Sequence Labelling for Movie Queries. - CRF Tagging of Movie. Tagging on the movie queries dataset first done using BIO tags and then improved by adding more features and using POS tags.
3. NLP project - Identifying the genders of the characters in the Eastenders British Soap Opera. The following are performed to experiment and optimize the classification: preprocessing, tokenization, different ngram models, line and sentence level features, cross validation and more.
4. NN & NLP project - Word Representation and Text Classification with Neural Networks. – Sentiment Analysis of Twitter and IMDB Large Movie Review Own word embedding is obtained by training a skip gram neural network model using negative sampling. Glove Embeddings was also used as pretrained embedding. Various algorithms like LSTM and CNN were explored. Finally, all these are applied real world text classification tasks like the sentiment analysis in twitter and the IMDB Movie reviews.
5. NN & NLP Project - Neural Machine Translation and Neural Dialogue Systems - Implemented a version of the seq2seq NMT model, and enriched that with attention, using a pre-trained BERT as a classifier model, implemented a series of dialogue act taggers and created an end-to-end dialogue system.
6. Deep learning Course-works: Critical analysis and study performed on Generative Adversarial Nets and on super-resolution using Convolutional Neural Networks. These two were used upon the MNIST dataset and the PyTorch framework is explored.

7. **Data Mining Course-works:** Data Preprocessing, Data exploration and visualisation (exploring sklearn, seaborn, pandas, numpy, and matplotlib libraries), Data Warehousing and On-line Analytical Processing (exploring cubes library), Classification, clustering, and Association Analysis (exploring mlxtend library), Outlier Detection, Web Mining (exploring BeautifulSoup library)
8. **Big Data Processing project:** Ethereum Analysis involving Time Analysis, Scam analysis, and comparative evaluation of the usage of Spark and MrJob. Thus, the following are performed: Number of transactions per month, average transactions per month, top ten most popular stories, top ten most active miners, Lucrative scams, Scam over time and correlation, Gas used over time.

WORK EXPERIENCE

Freelancer, Quest Global Technology (India) Pvt. Ltd

Dec 2017 – Jul 2018

Delhi, India

- Worked as freelancer developing visualforce pages and to certain extend Apex codes in salesforce
- Mentored as a Salesforce trainer.

Software Engineer, HSBC Software Development (India) Pvt. Ltd.

Jan 2008 – Aug 2014

Pune, India & Sheffield, United Kingdom

- Hands on experience on HP Exstream Dialogue (a CCM solution), Mainframe, JCL & CICS testing environment.
- Involved in the different phases of the project from creation of Requirements Gathering, Engagement Level Documents, Statement of Work, Implementation Tasks, and Coordination with the Different business personnel and Post Implementation Support.
- Involved in creating specifications from scratch for projects involving migration from one technology to another.
- As a Module Lead role, involved in allocation of tasks to subordinates and handling queries.
- Trained and mentored over 15 junior programmers and developers in India and China.
- Took additional responsibility of Internal Quality Auditor (IQA) role.
- Worked in UK at HSBC Griffin House (Sheffield, UK) as an Onsite Coordinator
- Have been recognized and awarded for individual and team achievements

CERTIFICATIONS

Udemy, Certificate of Completion: 2022 Complete Python Bootcamp from Zero to Hero in Python	2022
Kaggle, Certificate of Completion: Python	2021
Kaggle, Certificate of Completion: Intro to SQL	2021
Kaggle, Certificate of Completion: Advanced SQL	2021
Kaggle, Certificate of Completion: Pandas	2021
Kaggle, Certificate of Completion: Intro to Machine Learning	2020
Kaggle, Certificate of Completion: Intermediate Machine Learning	2020
Kaggle, Certificate of Completion: Data Visualization	2020
CDAC: A Certified Mainframe Professional	2008

REFERENCES

Syeda Mehwish Naseem
Data Engineer, SCSK Europe, London
Mobile: 07922945724
Email: smehwish.naseem@gmail.com

Rasika Kale
Scrum Master/ Agile Project Manager, Newbury
Mobile : 07555757240
Email : rasikakale111@gmail.com

Poornima Sampath Kumar
Data Engineer, AXA Insurance, London
Mobile: 07404918766
Email: poornimasampathkum@gmail.com