Anoop Kunjumon Scariah

+1682-283-5887 | anoopkunju@gmail.com| LinkedIn | GitHub

TECHNICAL SKILLS

- **Programming languages:** Python, SQL, Scala.
- **Deep Learning Frameworks:** HuggingFace, TensorFlow, PyTorch.
- **NLP & ML Libraries:** NLTK, Scikit-learn, LangChain, LammaIndex, SpaCy, PySpark, Pandas.
- **Cloud Platforms:** AWS (SageMaker, Comprehend, Glue, EC2, S3, Athena, Step Functions, Redshift,Lambda), Azure (AI Studio, SQL, Synapse, Databricks), Tableau.
- **Big Data Technologies:** Apache Spark, Chroma, Faiss.
- **Version Control & Collaboration:** GitHub, Jenkins, Trello, Jira, Rally.
- **Containers & Orchestration:** Docker, Kubernetes.

PROFESSIONAL EXPERIENCE

Data Engineer/Data Scientist **July 2021 - Present**

UnitedHealth Group

- Implemented RAG based LLM application using HuggingFace and LangChain for analytics information retrieval assistant increasing the Analytics team efficiency by 70%.
- Engineered advanced NLP models for Topic Modeling on unstructured data for categorization to enhance the patient outcomes prediction by 60%.
- Built and optimized automated data pipelines ETL and ELT using AWS Glue, Step functions, S3 and EMR for HL7, EMR, 835 Claims data from Parquet, JSON data format to facilitate data integration for analytics, machine learning, and predictive modeling products.
- Automated client acquisition data validation process using Python and SQL resulting in 50% reduction in data errors.
- Conducted Data Analysis to identify patterns and trends in data generation, validation, and system performance post-ETL process for ensuring data quality.
- Improved performance tuning and optimization using advanced/complex SQL to reduce the runtime of raw data extraction and transformation.
- Conducted Gap Analysis, Stage Data Analysis and Data modeling to comprehend the client's HL7 file data structure and requirements for business needs.
- Executed data and code migration from Oracle architecture to Big Data and AWS cloud environment.
- Led communication with different teams in the organization to accommodate changes in data and client requirements.

Graduate Research Assistant **June 2020 - July 2021**

Department of Data Science at The University of Texas at Arlington, Arlington, TX

- Developed object detection ML model using computer vision deep learning YOLOv3 architecture on AWS SageMaker, for detecting use of firearms.
- Designed data pipelines for integrating data from disparate sources and conducted ETL and data pre-processing using AWS Glue.
- Conducted business research & SQL analysis for product scope, finding potential buyers and vendors for the system to fit the unique specification.

Assistant Manager **January 2018 - July 2019**

Kotak Securities Ltd, Mumbai, India

- Spearheaded development lifecycle implementation from strategic planning through tactical & technical execution and deployment of Stock Trading REST API product that lets users execute the order in real-time, and stream live market data which increased the profit of the company by 45%.
- Planning Software development lifecycle following agile methodology using Jira and maintaining reports after the production process.
- Conducted and monitored performance, reliability, and functional testing of RESTful APIs using tools like Postman.
- Assisted cross-departmental project managers in product roadmap creation, evangelism, and execution.
- Conducted technical/progress meetings to discuss progress and any issues regarding the project with various levels of the organization.
- Assisted senior project managers with creating dashboards for analysis, staging process, and communicating with vendors.
- Wrote SQL stored procedures and scheduled them to populate data tables increasing CRM Analysis performance by 65%.

Software Engineer **July 2017 - January 2018**

- Coordinated with cross-departmental changes in the company process and technical implementation to understand functional requirements, and problem specification to establish system KPIs, database infrastructure, and goals for the Fintech software development project.
- Designed, built, evaluated, and deployed efficient and scalable full-stack system modules for the stock trading platform which has a customer base of 1.4 million users and around 500,000 trades per day using technologies like C++, Python, SQL, HTML, and JavaScript.
- Enhanced efficiency by automating SQL and MongoDB queries using Python REST API, resulting in a 30% increase in Stock Upload per second on a distributed system architecture over the network.
- Developed an automated system for application deployment and backup system with an integrated reporting dashboard to improve the speed and the efficiency of the end-of-day (EOD) process which helped to reduce the runtime by 80%.

CERTIFICATION

- **AWS Certified Cloud Practitioner** **January 2023**
- **AWS Certified Machine Learning - Specialty** **October 2023**

PUBLICATION

- Published research paper on "**Developing a multi-model transport system by linkage of local public transport with commuter trains using software as a service (SaaS) architecture.**" International conference on **ICTIS** 2018 by **Springer** publications.
- Published conference research paper "**Smart Ticketing System for Railway in Smart Cities using Software as Service**" at **IEEE** International Conference on **I-SMAC** (IoT1` in Social, Mobile, Analytics, and Cloud) 2017.

PROJECTS

Natural Language Processing: NLTK | NLP| Scikit-Learn | Python | Excel | Text processing | Tableau | AWS SageMaker |Web Scraping |D3.js | Flask | Beautiful soup.

- Developed NLP semantics models to accomplish Text-mining of historical content data using **NMF, LSA, LDA** to conduct business analysis and research to uncover product similarity and generate KPI measure, as well as Topic Modelling utilizing **Multidimensional Scaling, Clustering, and K-means** algorithms.
- Created a **Speech Recognition** Voice Assistant for Police to interact with a call, process data & filter the calls to enhance productivity by 75% with **Google DialogFlow**.

Sales Analysis for future prediction: Python | Pandas | Keras | SQL | Plotly | Tableau.

- Developed **Tableau dashboard** with Market Basket Analysis for finding the correlation between goods purchased to sales to increase annual profit by 75% of the store.

Heart Disease prediction application: TensorFlow | Pandas | Keras | NumPy | Plotly | PySpark | Tableau | AWS SageMaker | AWS Beanstalk | AWS Lambda.

- Developed **Ensemble ML model** which can detect the chance of heart disease in a client using the health data collected from wearable technology.
- Developed ML model using **AWS SageMaker** and deployed application on **AWS Beanstalk** with auto-scaling with load balancing and fault tolerance of 90%.

Bank Loan Acceptance Prediction Analysis: TensorFlow | Pandas | Keras | NumPy | Plotly | SQL | XGBoost Algorithm| Decision Trees Classifier Algorithm.

- Analyzed and processed complex banking data using advancing querying, visualization, and analytics tools to provide insights for KPIs measure for **acceptance forecasting**.

Product rating predictor: Python | RESTful API | AWS | NLP | Data Pipelines | Machine Learning Models | Web Application | NLTK | Scikit-learn | Pandas.

- Conducted data cleaning & ETL processing **terabytes** of data i.e., 15million+ customer review content data using AWS Glue pipeline for Machine Learning model generation.
- Conducted feature extraction using **TFID Vectorizer** and developed a Natural Language processing model using **Multinomial Naïve Bayes** Classifier.

EDUCATION

- **Master's in Computer Science.** The University of Texas at Arlington, TX **May 2021**
- **Bachelor's in Information Technology.** University of Mumbai, India **May 2017**