**Salman valiyakath**

Data scientist | Generative AI | NLP | AI Engineer

Al Hamriya, Dubai | +971 543136468 | valiyakathsalman@gmail.com | https://www.linkedin.com/in/salman-valiyakath

**PROFESSIONAL SUMMARY**

Data scientist with strong engineering background and 7 years of experience using Predictive modelling, Data analysis, Generative AI. Capable of Creating, Developing, Testing, and Deploying highly adaptive diverse services using NLP, Generative AI, Computer vision. Collaborating with cross functional teams and ensure the accuracy and integrity around the data.

**LANGUAGE & FRAMEWORKS:** Python | Docker | SQL | AWS Sagemaker | MLflow | Postgres | Linux | Kubernetes | TensorFlow | Git

**SKILLS:** NLP | Deep Learning | Agile | Azure ML | Generative AI | Vector Database| Fast API | AWS EKS | AWS ECR | AWS S3 | REST API

**TCHNOLOGIES AND TOOLS:** Scikit-learn, spaCy, BERT, NumPy, Pandas, GitHub, LLM, Langhain, Retrieval Augmented Generation | MLOps

**WORK EXPERIENCE**

**Capgemini Engineering, Bangalore**

***Senior Data scientist* May 2022 –** **Dec 2024**

* Engineered and deployed XGBoost machine learning models for predicting medical imaging protocols in healthcare, utilizing AWS SageMaker for model training and inference, Docker for containerization, and AWS EKS and AWS ECR for cloud deployment.
* Developed and optimized multi-model endpoints with multi-tenant capabilities in AWS SageMaker to enhance model deployment efficiency. Collaborated on the implementation of Generative AI DevStudio, automating product backlog generation, code development, and test creation to streamline the software development lifecycle using Google Generative AI LLMs.
* Facilitated communication with Product and Business stakeholders to decode business rules, logic, and operational requirements, driving data-driven decision-making processes.
* Developed a Generative AI-based Chatbot to support technical support engineers with troubleshooting, employing retrieval-augmented generation (RAG) techniques to enhance query resolution.
* Built and trained multi-class classification models to categorize Jira tickets, aiding technical engineers in streamlining the issue resolution process.
* Optimized data processing workflows, achieving a 50% reduction in time to detect and identify similar Jira tickets.
* Designed and implemented a Proof of Concept (PoC) for Log Explainer, generating meaningful insights and recovery plans from complex server error logs.
* Mentored junior members in best practices for data visualization, analysis techniques, and data quality assessment methodologies.
* Provided guidance to new joiners on contemporary practices and advanced techniques in machine learning and Generative AI.
* Conducted PoC for building, training, evaluating, and deploying machine learning models using AzureML, gaining hands-on experience with AzureML, Azure Storage, and Azure MLOps for end-to-end machine learning lifecycle management.
* Researched and established connections between Azure Storage and customer private datastore.
* Deployed and invoked machine learning models for predictions or API services using Azure managed endpoints, Azure Container Instances, or Azure AKS Cluster.
* Researched and trained models using Federated Learning frameworks such as Flower and Openfl.
* Collaborated on the design and implementation of the KYC Validator project, focusing on extracting key information from KYC documents and detecting document forgery.
* Trained and deployed a YOLO-based object detection model to detect bobbins and analyse rolled fiber cable alignment.

**Blackbuck Insight P Ltd, Bangalore**

***Data scientist* Sep 2020 –May 2022**

* Built an AI Quality Check Agent in python (NLP) for HealthCare Domain to identify and tag unauthorized documents.
* Conceptualized and built PoC for Identifying and Extracting required sections from the Patients Medical Record to fulfil the minimal ROI (Request of information) for an ROI Request.
* Created the NER Model using SpaCy v3 with Roberta Model, to extract Patients Medical and Demographic Entities from EHR.
* Build advanced algorithms to solve complex business problems of in a computationally and statistically effective manner
* Mastered skills in Python, SQL, Data analysis. Knowledge in Hypothesis testing, Machine learning, NLP and Deep Learning
* Developed ETL pipeline in python, for collecting data from multiple sources and transforming them into aggregated table in postgress, for powering Kibana Dashboard.
* Train the Classification model (fasttext) to Identify and tag No-medical Pages in Patient’s Electronic Health Record with 91% Accuracy.
* Analyze the Data and Identify trends or patterns in complex and large Ddataset’s to extract key insights, Drive optimization and improvement of product and communicate the results with the stakeholders for the Business purpose.
* Application of various machine learning algorithms and statistical Modeling like decision trees, Text Analytics, natural language processing (NLP), supervised and unsupervised, regression models, social network analysis, neural networks, deep learning, clustering to identify Volume using Scikit-learn package in python from aws Sagemaker Jupiter notebook.
* Extensive Data analysis - Data Cleansing, features engineering, Data imputation and Ccommunicate the insights from the Data with the Customer through the Confusion metrics and Charts to visualize the insights.

**Huawei Technologies, Bangalore**

***Automation Engineer* July 2018 – Apr 2020**

* Contributed for automating Repetitive and proactive task of Huawei MS Department & Pioneered Process automation under
* the domain specific areas in Telecom; supported in the product development.
* Developed and successfully implemented process automation for Saudi ZAIN MS Project at Huawei.
* IP - Vlan Data analyzing tool (Tools: - VBA, Ant robot, Batch script, Macros): - Designed and developed tool to support the Analysis team to significantly reduce the workload.
* Controller Utilization Processing tool (Tool: - VBA, Ant robot, Batch script): - Associated with Data Analysis team to prepare the 2G/3G Controller Utilization report tool.

**Pathfinder Enterprise Solutions P Ltd, Bangalore**

***Jr Technical Associate* Oct 2017 – May 2018**

* Responsible for Deploying and Configuring Retail Software and Helping the back-end team to make the audit report, troubleshooting of the retail software, Procurement of Retailer POS report in first of every month, Proficient with Microsoft.

**PROJECTS**

**Imaging Protocol recommendation system**

Developed and deployed a recommendation system to assist Radiologists in selecting optimal imaging protocols for MR, CT, and X-ray machines in healthcare settings. Utilized XGBoost Machine Learning algorithms and AWS ML Stack to design and implement the system, improving accuracy and efficiency in imaging diagnostics.

**Generative AI Studio**

Designed a system that leverages Google Generative AI Large Language Models (LLMs) to assist developers, managers, and product owners in generating product backlogs, code, and tests. This solution significantly streamlined the software development lifecycle, enhancing efficiency and reducing development time.

**KYC Validator**

Engineered a solution to automate the extraction of demographic data from customer identity cards, simplifying and accelerating the Know Your Customer (KYC) validation process. Ensured accurate identity verification while reducing manual labor and error rates.

**ACADEMIC DETAILS**

**Bachelor of Technology (B. Tech) in ECE (2017)**

Cochin University of Science and Technology, Kerala, India

**RELEVENT COURSES & CERTIFICATES**

* Computer Vision - Object Detection with OpenCV **Cousera**
* Machine Learning A to Z. Hands-on python in DS - **Udemy**
* Introduction to python – **Analytics Vidhya**
* Generative AI with Large Language Models – **Cousera**
* Introduction to Generative AI - **Cousera**

**PERSONAL INFORMATION**

Passport No: R6081661 | Nationality: Indian