SAMRUDHI BHOSALE

Pune, India | +91 9689891872| bhosalesamruddhi3535@gmail.com |Linkedin | Portfolio | Github

EDUCATION

Queen Mary University of London, MSc Big Data Science **Sep 2023- Sep 2024**
Grade Predicted: Distinction.
Core Modules: Data Mining, Machine Learning (ML), Neural Networks & Deep Learning (DL), Big Data Processing, Cloud Computing, Natural Language Processing (NLP), Applied Statistics.

Pune University, Computer Science **Jun 2015- Jun 2019**
Grade Obtained: Distinction.
Core Modules: Data Structures and Algorithms, Object Oriented Programming (OOPS), Computer Architures, Database Management Systems, Computer Networks, Discrete Mathematics.

WORK EXPERIENCE

Locomotion, London, UK **May 2024 - Current**
Junior Data Scientist, *Remote*

- Developing an AR/VR app that helps dementia patients talk and interact with digital avatars. The app uses natural language processing (NLP) and Conversational AI to enable meaningful conversations and help patients express their thoughts and feelings.
- Applied LLMs and LangChain to manage conversations. Added a guard rail system to filter out negative or harmful prompts, making sure interactions remain positive and supportive.
- Used Azure Cosmos Database to store and manage patient conversations, allowing for efficient data organisation and retrieval.
- Added conversation summary memory feature to keep track of previous chats, ensuring smooth and coherent interactions with the avatar.

Queen Mary University of London, London, UK **Feb 2024 - May 2024**
Demonstrator, *Part Time*

- Delivered lessons on social media analysis strategies and explained intricate machine learning topics in simple terms, using everyday examples to ensure clarity and better understanding.
- Led immersive workshops on advanced social media tools and analytics, which led to a 40% increase in student engagement and a 25% boost in skill proficiency.
- Supervised and guided the students for a project analysing the Twitch platform, focusing on identifying popular streamers, subscriber patterns, and relationships between influencers. Assisted students in conducting research and presenting their findings.

ZS Associates, Pune, India **Sep 2021-Jan 2023**
Software Engineer, *Full Time*

- Upgraded from legacy systems to a modern platform on AWS, significantly enhancing the efficiency of patient data analysis and supporting key healthcare decisions.
- Utilised Python libraries, such as scikit-learn, to develop models that improved data insights and decision-making processes, reducing data processing time by 40%.
- Employed PowerBI to design interactive dashboards that provided clear, actionable insights into patient data, improving operational efficiency by 25%.
- Handled extensive healthcare datasets and employed cloud technologies (AWS S3, EC2) to ensure effective data management and governance.

Accenture, Bangalore, India **Aug 2019 – Sep 2021**
Application Development Analyst, *Full time*

- Created machine learning models and used natural language processing (NLP) to analyse customer data, gaining insights into customer behaviour using scikit learn , nltk library.
- Applied big data analytics like Spark , Hadoop to personalise marketing campaigns, improving targeting and engagement.
- Built detailed customer profiles using data analysis to predict future behaviours and refine product offerings.
- Integrated insights into marketing strategies, boosting sales and customer loyalty.

PROJECTS

- **Deepfake Detection Using XceptionNet Model :QMUL :**Result: **Distinction (Achieved)** **(May 2024 - Aug 2024)**
  - Developed a deepfake detection system using deep learning model Xception Net and PyTorch, achieving an 82% accuracy rate on the CelebDF dataset, the highest reported in recent research for high-quality celebrity deepfake videos.
  - Evaluated multiple CNN architectures, including VGG16 and ResNet, but Xception Net outperformed, demonstrating superior results in detecting subtle facial manipulations.
  - Employed DLib for robust data preprocessing, extracting video frames and aligning faces with corrective rotation based on eye positioning; utilised Grad-CAM to visualise model attention during training, enhancing interpretability of the results.
- **Blog Generation using LLaMA 2 model :** **(Apr 2024 – May 2024)**
  - Integrated streamlit to develop an interactive web application, allowing users to input prompts and receive instant, high-quality blog posts generated by the Large Language model (LLM) LLaMA 2 model, demonstrating the practical application of AI in content creation.
  - Refined the generation process to optimise the coherence and relevance of the blog content, ensuring outputs were not only accurate to the input prompts but also engaging for readers, which enhanced user interaction and satisfaction.

- **Chatbot for Currency Conversion :** **(Mar 2024 - Apr 2024)**
  - Implemented a currency converter chatbot leveraging Google's Dialogflow for agent and intent setup and training.
  - The chatbot seamlessly communicates with backend code to relay currency parameters, with a Flask application processing the conversion and delivering the result. To make the application accessible, it was hosted using Ngrok.
- **Data Analysis and Visualization for NewYork Taxi Trip Analysis Using PySpark :** **(Feb 2024 - Mar 2024)**
  - Designed and executed PySpark scripts to analyse large-scale data from Uber and Lyft, merging rideshare and taxi datasets to identify the busiest pickup locations, most profitable routes, and urban mobility patterns. Presented findings using detailed visualisations that highlighted trends and patterns in daily taxi operations.
  - Gained practical skills in managing and optimising Spark environments, focusing on efficient data processing using Spark clusters, executors, and Resilient Distributed Datasets (RDDs) to handle complex queries and large data volumes.
- **Empowering Webcams with Real-Time Object Detection :** **(Jan 2024 - Feb 2024)**
  - Utilised YOLO models for swift object recognition, enabling real-time detection of items in video feeds from a webcam.
  - Integrated cvlib, a tool that facilitates the use of pre-trained models, streamlining project development and enhancing webcam functionality for immediate practical applications.
- **Machine Learning Classification Project, QMUL :**Result: **Distinction (Achieved)** **(Nov 2023 – Dec 2023)**
  - Designed two Machine Learning pipelines with MLEnd Yummy Data set, One using classification model - RandomForest, achieving 89.42% and 66.07% training & testing Accuracies for 'Rice' or 'Chips' classification.
  - Another classifying ‘Vegetarian’ or ‘Non-Vegetarian’ dish achieved 93.44% and 79.45% training & testing accuracies using Text analysis and NLP technique TF-IDF algorithm.
- **Kitchen Wizard- A Python Flask Application, QMUL :**Result: **Distinction (Achieved)** **(Oct 2023 – Nov 2023)**
  - Built a user-friendly Python app using Flask that instantly suggests recipes based on the fridge ingredients, like a virtual "magic pantry".Deployed the app on GCP for seamless user experience around the world.
  - Streamlined robust Hash authentication to keep the data safe and private, because delicious recipes shouldn't come at the cost of security.
- **Content Based - Movie Recommender System :** **(Oct 2023 – Nov 2023)**
  - Constructed a content-based movie recommender system using the TMDB 5000 dataset, implementing data cleaning, dimensionality reduction, and transformation of complex data structures for optimal data quality and analysis.
  - Applied the NLP technique bag of words and cosine similarity to analyse movie relationships, enhancing recommendation accuracy by focusing on genres, overviews, keywords, actors, and directors.

CERTIFICATIONS

- AWS Certified Cloud Practitioner
- Azure AI Fundamentals (AI900)
- Generative AI Fundamentals Databricks Certified

TECHNICAL SKILLS AND LANGUAGES

- Database: MYSQL, PostgreSQL, NoSQL, MongoDB.
- Programming Languages: Python - PyTorch, Keras, Numpy, Pandas, TensorFlow ; Java.
- Big Data Technologies: Hadoop, Spark, Hive, HBase, HPC, MapReduce.
- Data Warehousing: Sql Server, Redshift, Big Query, and Snowflake.
- Cloud Platforms: Google Cloud Platform(GCP), Amazon Web Services(AWS), Microsoft Azure for scalable data storage and processing, DataBricks.
- Data Analysis and Visualisation Tools : Excel, PowerBI, Tableau, JIRA.
- Deployment and Maintenance tools: Git, TeamCity.
- MLOPS : Github Actions, Airflow, Bitbucket, Teamcity.
- Languages: **English** (native proficiency), **Hindi** (native proficiency), and **Marathi** (native proficiency)