Samrudhi Bhosale

+91 9689891872
bhosalesamruddhi3535@gmail.com
Github
LinkedIn
Portfolio

Results-driven Data Science professional specializing in Machine Learning, Large Language Models, and Generative AI, eager to apply analytical expertise and innovative technologies to deliver impactful solutions for leading organizations.

PROFILE SUMMARY

- Over 4+ years of experience in Data Science and Machine Learning across finance, healthcare, and automotive domains.
- Skilled in building and evaluating machine learning models, automating data pipelines, and delivering production-ready solutions.
- Currently working in the automotive domain as a Software Engineer, focused on CI/CD pipelines and model evaluation using large language models.
- Experience with function-call-based and chat-based evaluations for assessing model performance before deployment.
- Strong background in data cleaning, annotation review, and automation of data quality checks using Python scripts and tools.
- Previously developed AR/VR-based applications to support dementia patients, combining computer vision and NLP for social impact.
- Proficient in Python, Git, Docker, and cloud platforms including AWS and Azure.
- Comfortable working across data science, MLOPS, and software engineering tasks, with a focus on building scalable, efficient solutions.
- Proven ability to lead and mentor peers, contributing to knowledge sharing and collaborative problem-solving.

EDUCATION

2024 — **MS in Big Data Science**
Queen Mary University of London

2019 — **BE in Computer Engineering**
Savitribai Phule Pune University

SKILLS

- **Programming Languages:** Python (Matplotlib, Seaborn, NumPy, Pandas), Java (Spring Boot, REST API)
- **ML Frameworks:** TensorFlow, PyTorch , Keras, Scikit-learn,
- **Big Data Technologies:** Hadoop, Spark
- **Database:** MySQL, PostgreSQL, NoSQL, MongoDB, Azure CosmosDB
- **Cloud Platforms:** AWS, Microsoft Azure, DataBricks
- **Deployment & Containerization:** Docker, TeamCity, Bitbucket, CI/CD, Jenkins, Gitlab
- **Visualization & Analysis:** Power BI, JIRA, Excel
- **MLOps:** Git, Jenkins, Airflow, MLFlow, Opik
- **APIs:** Flask, Google Dialog Flow

CERTIFICATIONS

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

CORE COMPETENCIES

- CI/CD Pipelines
- Cloud Computing
- Deep Learning
- Natural Language Processing
- Big Data Technologies
- Large Language Models
- Conversational AI
- Generative AI
- Image processing

WORK EXPERIENCE

Cerence AI, Pune, India as Software Engineer
March 2025 - Current

- Designed and implemented CI/CD pipelines to automate machine learning model evaluation workflows before deployment.
- Integrated LLMs for automated function-call and chat-based model evaluations, improving evaluation accuracy by 20%.
- Worked on Agentic AI systems; evaluated in-car agents for tasks like car control, navigation, and media handling.
- Checked agent selection based on query intent, ensuring correct routing and validating verbalizer outputs.
- Developed a pipeline to evaluate Level 2 (L2) agents by verifying specific function/tool calls for each domain.
- Performed evaluations for cross-domain queries to ensure multi-functional requests were correctly handled.
- Automated 61K records dataset review and cleaning using Python scripts, reducing data preprocessing time by 40% and improving annotation quality scores by 10%.
- Evaluated multiple large language models (GPT-3.5-Turbo, LLaMA, GPT-OSS, Gemma-3N) to generate inferences and benchmark their outputs against ground truth data for model selection.
- Utilized Python, Git, Jenkins, Docker, and Azure cloud services for development and deployment.

ZS Associates, Pune, India as Software Engineer
September 2021 - January 2023

- Designed and implemented machine learning models aimed at predicting heat stroke outcomes, enhancing predictive accuracy and contributing to public health initiatives.
- Streamlined data pipelines for healthcare applications by utilizing Amazon Redshift, effectively managing both SQL and NoSQL datasets to optimize data ingestion processes.
- Transitioned legacy systems to a modern AWS-based architecture, improving scalability and performance while ensuring seamless integration with existing workflow.
- Monitored model version control and monitoring in production environments using Bitbucket and TeamCity, adhering to best practices in DevOps to ensure efficient CI/CD processes.
- Collaborated with cross-functional teams to identify business requirements and translate them into technical specifications for data-driven solutions.
- Conducted thorough testing and validation of machine learning models to ensure reliability and accuracy in real-world applications.

Accenture, Bangalore, India as Application Development Analyst
Aug 2019 - Sep 2021

- Integrated diverse financial and consumer data sources for Putnam Investments, developing robust solutions that enhance decision-making capabilities and drive business value through insightful data visualizations.
- Created automated scripts for data cleaning and normalization, ensuring high data integrity across extensive financial datasets and facilitating accurate analysis.
- Applied advanced statistical analysis and natural language processing techniques to gain insights into customer behavior, enabling improved segmentation and targeting financial products.
- Assisted in the containerization of applications using Docker, contributing to enhanced deployment processes and scalability of solutions.
- Collaborated with cross-functional teams to identify business needs and develop tailored solutions that align with organizational goals.

INTERNSHIP

Locomotion, London, UK as Junior Data Scientist
May 2024 – Feb 2025

- Spearheading the development of an innovative AR/VR application designed specifically for dementia patients, utilizing advanced Natural Language processing techniques to facilitate engaging conversations through digital avatars.
- Employing large language models, including OpenAI's GPT-3.5 and LLaMA, to create a sophisticated dialogue management system that ensures safe and meaningful interactions by implementing guardrail filtering mechanisms.
- Optimizing data storage and retrieval processes by leveraging Azure Cosmos Database, thereby enhancing the efficiency of patient data management and utilization.
- Integrating memory features into the application to maintain conversational context, improving user engagement and interaction quality across multiple sessions.
- Collaborating with cross-functional teams to ensure the successful deployment of the application, focusing on user feedback and iterative improvements to enhance functionality.

Queen Mary University of London, London, UK as Demonstrator
February 2024 - May 2024

- Facilitated engaging lessons on social media analysis strategies, breaking down complex machine learning concepts into easily digestible formats to enhance student comprehension and retention.
- Led hands-on workshops focused on advanced social media tools and analytics, fostering an interactive learning environment that encourages student participation and collaboration.
- Guided students through a comprehensive project analyzing the Twitch platform, assisting them in identifying key trends and relationships among popular streamers and their audiences.
- Provided mentorship and support to students during their research projects, ensuring they develop critical thinking and analytical skills necessary for success in the field.
- Evaluated student performance and providing constructive feedback to promote continuous improvement and skill development.
- Collaborated with faculty to design curriculum materials that align with industry standards and emerging trends in data science and analytics.

PROJECTS

Deepfake Detection Using XceptionNet Model : QMUL :Result: Distinction (Achieved) (May 2024 - Aug 2024)

- Developed a deep fake 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, showcasing 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; utilized Grad-CAM to visualize 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, showcasing the LLM proof of concept for content creation.
- Refined the generation process to optimize 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.

Data Analysis and Visualization for New York Taxi Trip Analysis Using PySpark :

February 2024 - March 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 visualizations that highlight trends and patterns in daily taxi operations.
- Gained practical skills in managing and optimizing Spark environments, focusing on efficient data processing using Spark clusters, executors, and Resilient Distributed Datasets (RDDs) to handle complex queries and large data volumes.

Diet Classifier, QMUL: Result: Distinction (Achieved)
November 2023 – December 2023

- Developed machine learning models for food image classification, focusing on dietary categories such as vegetarian and non-vegetarian, and distinguishing between food types like rice and chips.
- Utilized advanced feature extraction techniques, including data augmentation and food area cropping based on high-saturation regions.
- Implemented classification algorithms like Logistic Regression, Support Vector Machine, Naïve Bayes, and CNN models such as ResNet, VGG16, and GoogleNet. Applied data preprocessing and augmentation to improve model accuracy, with practical applications in dietary monitoring and personalized nutrition planning.

Content Based - Movie Recommender System:
October 2023 – November 2023

- Designed and implemented AI-powered solution, a 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 for analyzing movie relationships, enhancing recommendation accuracy by focusing on genres, overviews, keywords, actors, and directors.

PERSONAL DETAILS

Address	: Pune, India
Date of Birth	: 3rd April 1997
Languages Known	: English, Hindi, and Marathi