Bharath Shroff

+919811982503 | bharathsshroff@gmail.com | https://www.linkedin.com/in/bharathshroff | https://github.com/bssmsi

Summary

Results-driven professional with 5+ years of experience in AI/ML and data science consistently leveraging cutting-edge technologies to drive innovation. Proven expertise in deploying scalable ML pipelines/solutions to Production and delivering actionable insights for global stakeholders. Skilled in AI/ML, Python, RAG and cloud platforms like Databricks and Azure.

Work Experience

MCSquared AI   Remote Contract
Data Scientist   Aug 2024 - Present
- Led the team to build a pipeline in Databricks feeding into a map view dashboard containing proximity hotspots of leads around business provided site locations leveraging Bing Maps API and 3rd party Real world data sources like Citeline, Health Verity, IQVIA.
- Deployed an unsupervised K-means clustering model to categorize doctors into 9 different categories based on a variety of features effectively enriching the data with classification of doctors for downstream use.
- Providing predictive solutions for data driven initiatives for clinical trial recruitments for manufacturing of new products.

Trilogy   Remote Contract
AI Innovation Specialist - Finance   Aug 2024 - Oct 2024
- Deriving Financial Insights using LLM chatbot built on React for the frontend and Express JS for the backend, which updated the RAG Vector DB upon new file uploads, reducing manual analysis time by an hour.
- Reporting with automated consolidated financial reports with AI Insights including Profit and Loss (P&L) reports and Ledgers across 10 different entities saving 5-6 hours of manual effort required every month end.
- Analyzed and identified over 100 AI-first tools and built PoCs to speed up accounting processes, reducing the need for building from scratch and improving efficiency.

MCSquared AI   Remote Full Time
Data Scientist   May 2022 - Jul 2024
- Deployed Machine Learning Survival model to production replacing the previous XGBoost model on Databricks using the medallion architecture capable of self re-training every month with new data and auto archive or promote to production based on the champion model using MLFlow for model versioning and evaluating the model performance based on C-score, a metric similar to Area Under the Curve (AUC score). Improved precision to 20% from 6% to narrow down sales leads by more than 80% making data-backed decisions and improve adoption of product by providing lower number but higher quality leads.
- Deployed pipelines and several PoCs in Databricks to automate EDA and model complex datasets from 5 different real-world data sources and third-party API sources, ensuring the data warehouse was updated with both delta load and full load functionalities to support timely business decisions feeding into Power BI / Tableau dashboards.
- Deriving insights from the data using Contours for PoC and Data Quality Management using Ontology Objects on the Palantir Foundry platform, creating dashboards crucial for global business decisions in the pharmaceutical industry.

Isha Foundation   Coimbatore, India
Full Stack Developer Volunteer   Aug 2021 - Apr 2022
- Developed a web application using the open-source Odoo Framework built on Python, streamlining processes and digitizing multiple forms required to be filled by hand by 100s of visitors saving hours of work both for the visitors and the staff.
- Developed APIs for communication across four different ecosystems, enhancing integration and data flow.

ITC Infotech   Bengaluru, India
Associate IT Consultant   Jun 2019 - Jul 2021
- Deployed end-to-end modules using Git DevOps for Continuous Integration/Continuous Deployment across the 4 stages (DEV->QA->UAT->PROD), ensuring seamless transitions and operational efficiency for MLOps.
- Migrated project codebase from Scala on HDInsight Cluster to PySpark on Databricks with optimizations.
- Automated daily tasks using Python (Selenium) and PowerShell scripts, including taking screenshots, verifying counts, and emailing recipients for Data Quality Monitoring, which improved accuracy and saved time.
- Redesigned pipelines for performance by cleaning up the Azure Data Factory, reducing 300 redundant datasets to 4 and 30 pipelines to 3, which streamlined data processing and improved efficiency.
- Created and modified Power BI dashboards to support business development by enabling data-informed decisions and providing actionable recommendations, leading to improved strategic planning.

DELL EMC   Bengaluru, India
RnD Intern   May 2018 - Jul 2018
- Developed Python scripts for automated reporting, flagging approximately 100 high-priority reports daily, enhancing efficiency in report management
- Built Web Spider using scrapy for notifying outdated open-source packages which scrapes 4 different open-source websites.
- Established a robust priority flow between four websites ensuring comprehensive package updates and improving data consistency

Computer Institute of Japan   Yokohama, Japan
RnD Intern   May 2016 - Jul 2016
- Helped in improving the accuracy of multi-class Classification of emails and Achieved 70%+ accuracy.
- Developed python GUI using Tkinter with Progress bar for the 3 different Training Testing, Validation stages.
- Used gensim, SVM, Naive Bayes and TF-IDF pre-processor for best performance.

Education

Indian Institute of Technology (IIT) Hyderabad, India
Bachelor of Technology, Mechanical Engineering   Jul 2015 - May 2019

Projects

BeingAwareMatters
CTO at beingawarematters.com   Jul 2024 - Sep 2024
- Single-handedly developed a cross-platform app using React Native. Deployed web version on Vercel.

Buzzzest
CTO at FashionTech   Dec 2023 - Jun 2024
- Spearheading the tech platform development single-handedly.
- Built a B2B dashboard using Next.JS
- Built a chatbot leveraging RAG technique on LangChain in Python.
- Building a cross-platform using React Native

ITC Infotech
Network traffic analysis   Oct 2020 - Oct 2020
- Extracting insights by transforming 6 million + Apache server logs and visualizing through plots showing traffic originates from 10 different countries.
- Link to Open-Source GitHub - https://github.com/BssMsi/PySpark-Training.

Melanoma Classification   Aug 2020
- Achieved 85% AUC score in Identifying Melanoma using Convolutional Neural Network (CNN) models.
- Link to Kaggle Notebook - https://www.kaggle.com/code/bssmsi/melanoma-classification.

Image classification of fruits   May 2020 - Jul 2020
- Multi Class Classification of Fruits using images, dataset used from Kaggle with 90380 annotated images.
- Leveraging Pretrained Machine Learning (ML) models like VGG, ResNet, AlexNet, Mobile Net for mobile deployable model.
- Link to the Notebook using Open Source Kaggle - https://www.kaggle.com/code/bssmsi/fruits-recognition.

IIT Hyderabad
IITH Main Website   Jan 2019 - Mar 2019
- Built our college website from scratch using WordPress Templating which included integrating from over 10 departments.

Certifications

Certified Azure Data Engineer Associate (DP-200, 201): Microsoft | 2021

Skills

Python, Big Data, SQL, Databricks Cloud, PySpark, Machine Learning, Data Science, Data Engineering and Manipulation, PyTorch, LangChain, LLM, RAG, Next JS, React Native, Odoo