Tejas Shinde

Los Angeles, CA | 978-274-8822 | tejasashinde15@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Senior Data Scientist with over 5 years of experience building and deploying end-to-end data science and machine learning solutions across Energy, Finance, Health and HR domains. Proven expertise in integrating statistical modelling, developing CI/CD pipelines, deploying advanced NLP frameworks, with business intelligence production ready dashboards. Awarded for delivering measurable business outcomes, including significant cost savings and operational efficiency. Passion to mentor teams and collaborating cross-functionally to create a safe space for innovation and playing with numbers together as a team.

SKILLS

Technical Skills	:	Python, SQL, R, Rust, ReactJS, MongoDB, SQLite3.
Algorithms & Statistics	:	Regression, Classification, Clustering, LLMs, Neural Networks.
Data Visualization:	:	Tableau, PowerBI, Streamlit, D3.js.
Cloud:	:	Git, Databricks, Azure, AWS, Docker, Terraform.
Frameworks:	:	TensorFlow, Keras, HuggingFace, A/B Testing (Kameleoon), REST, SOAP.
GenAI APIs:	:	OpenAI, Grok, Perplexity, Llama.
Tools:	:	Advanced Excel, PowerPoint, Word, VBA, Jira, Oracle BI.
Soft Skills:	:	Decision Making, Clear Communication, Leadership Skills, Agile Framework.

WORK EXPERIENCE

Accenture (Client: San Diego Gas & Electric) — **Los Angeles, CA**
Senior Data Scientist — May 2024 – Present

- Designed and implemented robust, end-to-end CI/CD machine learning pipelines, integrating multiple data sources and processing over 8 million records. Delivered actionable insights that identified key drivers for Diverse Business Enterprise, significantly enhancing organizational strategy saving ~$120000.
- Applied advanced statistical techniques to balance covariates and implemented K-Means clustering on cost probabilities derived from XGBoost and Random Forest propensity models deployed using Databricks. Improved forecast accuracy and saved up to 2 FTE's effort, equating to an annual savings of ~ $80,000
- Developed and deployed 18 production-ready Power BI and Tableau dashboards enriched with NLP capabilities to analyze open-text data, leveraging LLMs (RoBERTa, BERT) layered with RAG pipelines deployed using Azure DevOps and Docker for seamless wall time, enabling real-time insights and decision-making.
- Led the development of ensemble models combining Random Forest, XGBoost, and LightGBM to predict cost drivers and optimize resource allocation. Improved model accuracy by 12% compared to baseline models.

HSBC — **Bengaluru, India**
Data Science Manager — Jun 2022 - May 2024

- Built a comprehensive Employee Lifecycle analytics platform that consolidated employee skills, mood, churn, and pay indicators, resulting in a 20% improvement in talent retention insights and a 15% reduction in churn forecasting error and hosted the solution on a Linux server on Azure.
- Integrated Selenium-based REST API GitHub calls for scraping to extract over 100,000 unique job skills, leveraged advanced NLP techniques (zero-shot classification and custom embeddings) to analyze 200,000+ open-text survey responses, and deployed predictive models (linear regression, logistic regression, optimization, time series forecasting, and SQL) to deliver actionable insights to HR and leadership teams.
- Improved Employee survey submissions from 60% to 75% by collaborating with senior management across customer success and finance teams, identifying and addressing customer pain points to drive engagement.
- Mentored a team of 28 on advanced Python techniques, including feature engineering, model validation, and explainable AI techniques, accelerating the adoption of best-in-class data science practices.
- Earned 1 Regional Award, 1 Global Award, and 1 promotion in recognition of impactful contributions.

HSBC — **Bengaluru, India**
Senior Data Scientist — Jul 2020 - Jun 2022

- Led the initiative of refactoring existing reporting techniques and helped automate 835 PowerPoint cuts for each business using Python PPTX, APIs, MongoDB and Databricks for faster querying which cut down time from 2 months of reporting to 22 hours of ready reports in the inbox of employees ranging from Peoples Manager to Business Heads.
- Created Statistical and Machine Learning Models to forecast the FTE & Cost at Employee Level which served as an organization wide benchmark base to track churn in form of monthly executive reports presented to the board. The board would take actions based off of the numbers which were A/B Tested and Audit screened.
- Worked cross functionally with Data Science Leads at Microsoft, Glass Door and TechWolf to enhance the inhouse capability of Employee Engagement and see how the actual data is doing against the Market Data to predict Employee Mood scale.
- Documentation of coding standards for the wider team & setting up default GitHub repositories for code ready environments in Python & R.
- Secured 3 Global awards and 2 promotions.

EDUCATION

Symbiosis International University — **India**
Master of Science – Computer Applications (CGPA: 8.62/10.00) — Apr 2020

- Ranked #2 in 3 out of 4 semesters, Appointed Data Science exchange representative with Ritsumeikan University.
- Relevant Coursework: Data Science Major, Data Structures & Algorithms, Machine Learning, End to End pipelines, Natural Language Processing.

Symbiosis International University — **India**
Bachelor of Computer Applications (CGPA: 7.86/10.00) — Apr 2018

- Ranked #1 & #2 in 2 & 1 out of 6 semesters respectively, Secured 2nd place amongst top universities in India.
- Relevant Coursework: Exploratory Data Analysis in Python & R, JavaScript, Java, DS Android Application.

PROJECTS & RECENT CONTRIBUTIONS (Academy Project Portfolio)

Kaggle: Detect AI Generated Texts using LLMs (repo)
Kaggle: Diabetes Diagnosis Prediction (repo)
Neural Network & Deep Learning (repo)
Machine Learning (repo)
OpenCV Projects (repo)
Auto Encoders (repo)
Reinforcement Learning (repo)
Full Stack Data Science Webapps (repo)