Naga Pratyusha Duvvuri

Data Scientist

prtyushaduvvuri98@gmail.com | +353892624336 | Cork, Ireland | https://www.linkedin.com/in/pratyushaduvvuri98/

github.com/pratyusha1213

SUMMARY

Data Scientist and Python Developer with 4+ years of experience in software engineering, AI/ML development, and middleware integration. Proven expertise in machine learning, data preprocessing, model evaluation, and MLOps workflows, with hands-on experience across Microsoft Azure and Oracle Cloud environments. Strong background in large language models (LLMs), API integration, and end-to-end data science project execution gained through roles at Capgemini and G2i Inc. Recognized for improving AI model performance, optimizing enterprise systems, and delivering scalable, production-ready ML solutions. Certified in Microsoft Azure, Oracle Cloud, and Generative AI, with a track record of enhancing data-driven decision-making and driving automation.

PROFESSIONAL EXPERIENCE

G2i Inc, *Data Scientist - Freelancer* 10/2024 – Present | Remote

- Improved LLM output quality by **30%** through systematic evaluation, structured scoring frameworks, and targeted fine-tuning recommendations.
- Reviewed and optimized AI-generated code across **Python, SQL, Java, C**, and networking scripts, increasing accuracy and maintainability for enterprise applications.
- Identified defects and performance bottlenecks in generative AI workflows, reducing model response inconsistencies by **25%**.
- Collaborate with engineering teams using **GitHub, Python, SQL, REST APIs**, and LLM evaluation tools to enhance model alignment and code generation reliability.

Capgemini, *Associate Consultant* 09/2020 – 04/2024 | Hyderabad, India

SCUBI Maintenance Support — Cloud & Batch Processing

- Developed and maintained Python-based batch processes deployed on **Oracle Cloud (OCI)**, improving system stability and supporting high-volume employment benefits data processing.
- Automated recurring operational tasks and optimized batch scheduling, resulting in a **25% reduction in job runtime** and a **40% improvement in pipeline reliability**.
- Monitored and validated cloud-based batch jobs, resolving failures and enhancing data flow consistency across environments.
- Collaborated with cross-functional teams using **Azure, OCI, Python, SQL, Git, and Automation tools** to streamline production workflows.

Middleware Engineering — Atradius Credit Insurance

- Designed and maintained enterprise integrations using **Oracle SOA Suite, OSB, BPM, and JMS**, enabling secure and efficient communication across insurance systems.
- Built RESTful and SOAP web services that supported critical business workflows and improved integration performance by **35%**.
- Created BPMN workflows connected with SOA services and backend databases, automating decision-making and reducing manual processing time.
- Conducted performance tuning using Oracle Enterprise Manager (OEM), increasing throughput and reducing latency in middleware operations.

EDUCATION

Data Science & AI Program, *Turing College* 06/2024 – 11/2025 | Remote

Big Data Processing and Applications - Data Engineering, *University of Oulu* 03/2025 – 05/2025 | Finland

Bachelor of Technology - Information Technology, *Pragati Engineering College* 06/2016 – 09/2020 | Surampalem, India

SKILLS

Machine Learning: Supervised & Unsupervised Learning, Predictive Modeling, Forecasting, Classification, Clustering, Feature Engineering, Statistical Inference, Model Tuning (GridSearchCV), Model Evaluation (Accuracy, ROC-AUC, Precision/Recall)

Programming & Tools: Python, SQL, Django, FastAPI, GitHub, Jira, RESTful APIs, Postman, SoapUI, Power BI (Basic), Excel | Cloud Platforms: Oracle Cloud (OCI), Microsoft Azure | Version Control & Workflow Tools: Git, Jupyter Notebook

Exploratory Data Analysis (EDA): Data Wrangling, Data Cleaning, Descriptive & Inferential Statistics, Correlation Analysis, Outlier Detection, Feature Selection, Visualization (Matplotlib, Seaborn), Cloud-based Data Processing

AI & Deep Learning: Large Language Models (LLMs), Generative AI

PROJECTS

Travel Insurance Purchase Prediction, *Machine Learning Project*

- Built a predictive ML model to estimate the likelihood of customers purchasing travel insurance using demographic and travel-related features.
- Performed comprehensive **EDA**, handled missing values, created visualizations, and engineered features to improve signal quality.
- Trained multiple models — **Logistic Regression, Decision Tree, Random Forest, XGBoost** — and applied **GridSearchCV** for hyperparameter tuning.
- Developed an ensemble classifier using **VotingClassifier**, achieving improved classification stability and higher accuracy over individual models.
- Identified key drivers of insurance purchase through **feature importance analysis**, enabling actionable business interpretability.
- Technologies: **Python, Pandas, Scikit-learn, XGBoost, Matplotlib, Seaborn, Jupyter Notebook**

Home Credit Default Risk Prediction, *ML Project*

- Built a full end-to-end machine learning pipeline to predict the likelihood of clients defaulting on loans using real-world credit bureau and financial datasets.

- Performed extensive **data preprocessing**: handled missing values, encoded categorical features, managed class imbalance, engineered new features, and resolved outliers.
- Conducted **in-depth EDA** including distribution analysis, correlation heatmaps, feature visualization, and insights into key socioeconomic predictors.
- Trained and optimized multiple models — **XGBoost, LightGBM, Gradient Boosting** — with evaluation using ROC-AUC, precision/recall, and feature importance analysis.
- Achieved **0.77 ROC-AUC** using LightGBM, successfully identifying high-risk clients and improving credit decision-making accuracy.
- Deployed the final model to **Google Cloud Platform (GCP)** with an accessible **HTTP API endpoint**, enabling real-time prediction capability.
- Technologies: **Python, Pandas, NumPy, Scikit-learn, XGBoost, LightGBM, Seaborn, Matplotlib, GCP, Jupyter Notebook**

AI Interview Simulator Web App, *LLM-Powered Application*
- Developed an interactive **Streamlit web application** that simulates job interviews using AI-generated, role-specific questions and personalized feedback.
- Integrated **OpenAI/Gemini LLM APIs** to generate 10 tailored interview questions based on job title, job description, and uploaded resume (PDF/DOCX/TXT).
- Implemented multiple **prompt engineering techniques** including Zero-Shot, Few-Shot, Chain-of-Thought, Role-Based, and Self-Critique to generate high-quality questions and feedback.
- Built automated candidate assessment with **AI-generated feedback**, including strengths, weaknesses, model answers, and performance scoring.
- Added advanced customization options such as difficulty level, AI creativity (temperature), skip question feature, and raw LLM output display for debugging.
- Implemented document parsing using **PyPDF2** and **python-docx** to extract resume content for personalized question generation.
- Designed modular structure with separate utilities for prompt templates, API calls, and input validation, ensuring maintainability and scalability.
- Technologies: **Python, Streamlit, OpenAI/Gemini API, PyPDF2, python-docx, Regex, Virtual Environments**

CERTIFICATES

- Oracle Cloud Infrastructure 2024 Generative AI Certified Professional
- Microsoft Certified: DevOps Engineer Expert
- Microsoft Certified: Azure Fundamentals

PUBLICATIONS

RESEARCH ON APPLICATION OF ARTIFICIAL INTELLIGENCE IN MEDICAL EDUCATION, 05/05/2020
The International journal of analytical and experimental modal analysis
Artificial Intelligence (AI) is revolutionizing medical education by enhancing distance learning, virtual inquiry systems, and teaching efficiency. This research explores how AI-driven solutions, machine learning, and intelligent tutor systems improve medical training and personalize learning. Our findings highlight AI's potential to bridge traditional education with digital innovation in healthcare.