

# Vasu Kondapalli

Email: [vasudeva.rao@quadranttechnologies.com](mailto:vasudeva.rao@quadranttechnologies.com) Phone: 8249692289 LinkedIn: [linkedin.com/in/skswain172](https://www.linkedin.com/in/skswain172)

---

## Data Scientist | AI Engineer | LLM Solutions Developer | Gen AI

With over 3 years of expertise in **Data Science**, focusing on **Machine Learning (ML)**, **Natural Language Processing (NLP)**, and **Deep Learning**, to create innovative and impactful data-driven solutions. With strong proficiency in **Python** and ML tools such as **Scikit-Learn**, **TensorFlow**, and **PyTorch**, I aim to contribute to the development of efficient **machine learning pipelines** and scalable **data engineering systems**. My goal is to harness my skills in **cloud technologies**, **database management**, and **data visualization** to support insightful decision-making and deliver measurable value to the organization.

## SKILLS

---

- **Machine Learning:** Supervised and Unsupervised Algorithms
- **Natural Language Processing:** RNN, LSTM, Transformers, BERT, GPT, Encoder-Decoder
- **Generative AI:** OpenAI, ChatGPT, Llama, Gemini, Claude Hugging Face, RAG, Fine Tuning
- **Deep Learning:** ANN, CNN
- **Computer Vision:** OpenCV
- **Programming:** Python
- **Databases:** MySQL, MongoDB, Vector DB
- **Visualization:** Matplotlib, Seaborn, Power BI, Tableau
- **Frameworks:** TensorFlow, Keras, Scikit-Learn, Streamlit, Lang Chain, Flask API
- **Cloud Platforms:** AWS, Azure (including Azure Cognitive Services)
- **Tools:** Docker, Git, GitHub Actions, CI/CD, MLflow, Cline

## EXPERIENCE

### Data Scientist

**ZAVJET INFOTEC PRIVATE LIMITED** Jul 2022 – Apr 2025

#### LLM-Driven Role Extraction Framework for Business Documentation

**Technologies:** Python, Gemini Flash, Claude Sonnet 3.5 & 4, Prompt Engineering, YAML, Markdown, Regex

- Built a scalable system to extract and structure role definitions from unstructured business documents using **Gemini Flash** and **Claude Sonnet models**.
- Designed a **template-agnostic prompt pipeline** with smart fallback, enabling accurate role, entity, and objective mapping across diverse formats.
- Automated structured output generation in **YAML and Markdown**, ensuring schema compliance and filename safety.
- Achieved over **95% completeness** and consistency across document types while reducing manual extraction effort significantly.
- Delivered a lightweight, production-ready solution for document intelligence in enterprise settings.

#### AI-Powered HR Document Intelligence System using OCR + LangChain + Gemini

**Technologies Used:** Python, Flask, Tesseract OCR, OpenCV, LangChain, Google GenAI Embeddings, ChromaDB, Gemini, Azure Blob Storage

- Developed a scalable enterprise-grade RAG system for intelligent HR document understanding and real-time Q&A.
- Extracted text from scanned PDFs (e.g., appraisal letters, policy manuals) using Tesseract OCR and OpenCV.

- Converted raw OCR output into semantic chunks and embedded them using **Google GenAI Embeddings**, storing vectors in ChromaDB with custom metadata (e.g., document type, department, date).
- Powered the question-answering pipeline using Gemini Pro via LangChain's RetrievalQA, integrating hallucination control and fallback logic.
- Delivered a secure Flask-based interface for uploading, embedding, and querying documents with audit logging support.
- Reduced manual HR lookup times by **70%**, enhancing employee self-service and compliance responsiveness.

### Customer Churn Classification Using BERT

- Developed a robust customer churn classification model using fine-tuned BERT on telephonic conversation transcripts.
- Pre-processed data through advanced text cleaning, tokenization, padding, and label encoding to ensure compatibility with BERT.
- Implemented learning rate scheduling, early stopping, and regularization techniques to optimize model performance during fine-tuning.
- Achieved 88% accuracy with high precision, recall, F1-score, and an AUC-ROC of 0.91, evaluating the model using confusion matrices and classification reports.
- Integrated MLflow for experiment tracking and version control of models, enhancing reproducibility.
- Automated CI/CD pipeline setup using Azure DevOps for seamless deployment.
- Configured scalable deployment with continuous monitoring and retraining workflows.

## Python Developer

**ZAVJET INFOTEC PRIVATE LIMITED** Jan 2021- June 2022

### Project: Delivery Duration for E-commerce Company

- Built a pipeline to analyse and predict delivery times from historical data.
- Extracted and engineered features like delivery duration from timestamp data.
- Cleaned and validated data for accurate model training. Visualized key insights using Pandas, NumPy, and Seaborn.

### Project: Job Portal Analysis

- Analysed job listings to identify trends in salaries and ratings.
- Parsed and normalized salary estimates using regular expressions.
- Generated insights through visualizations with Matplotlib and Seaborn.
- Structured unorganized data for actionable insights.

## EDUCATION

### Bachelor of Science in Physics

*Utkal University, Bhubaneswar, Odisha*

*Graduated 2016*

## CERTIFICATION

Generative AI: Introduction and Applications

Link: <https://coursera.org/verify/AFNLH80PMIXO>

Project: Generative AI Applications with RAG and Lang Chain

Link: <https://coursera.org/verify/PHVT84T9BWEL>

Machine Learning with Python

Link: <https://courses.cognitiveclass.ai/certificates/8a108a53db8b4844ad3a69f9850f7c34>

