## **Sri Charan**

**Email ID:** **ccharan1805@gmail.com**

**LinkedIn:** [**Sri Charan Chowdary**](https://www.linkedin.com/in/sri-charan-chowdary-15043b349/)

**Phone no: +1 567 294 3334**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Professional Summary**

* AI/ML Engineer with over 3+ years of hands-on experience in building machine learning models and deploying intelligent systems in production environments.
* Specialized in Natural Language Processing (NLP) with experience in text classification, sentiment analysis, NER, and text summarization using tools like spaCy, NLTK, and Hugging Face Transformers.
* Proficient in Python, with strong knowledge of Pandas, NumPy, Scikit-learn, and Matplotlib for data manipulation, analysis, and visualization.
* Proven expertise in .NET Core, C#, and Microsoft’s AI tools including Semantic Kernel, Azure AI Services, and Agentic AI architectures.
* Experienced in training and fine-tuning models using TensorFlow, PyTorch, and Hugging Face, with a focus on real-world performance and interpretability.
* Developed and deployed RESTful APIs for ML models using Flask and FastAPI, integrating them into enterprise applications and client-facing platforms.
* Experience integrating Retrieval-Augmented Generation (RAG) pipelines with AWS Bedrock and foundation models (e.g., Claude, Titan) to build scalable, LLM-powered chatbots and knowledge retrieval systems for enterprise use cases.
* Built and managed data pipelines involving ETL processes, data validation, and feature engineering for structured and unstructured data sources.
* Deployed models and applications on cloud platforms such as AWS EC2, S3, and SageMaker, and used Docker to containerize ML services.
* Applied hyperparameter tuning using GridSearchCV and RandomizedSearchCV to optimize model performance and reduce overfitting.
* Skilled in building RAG (Retrieval-Augmented Generation) pipelines and integrating LLMs (Large Language Models) into production environments.
* Implemented MLOps practices, including Git-based version control, basic CI/CD automation, model logging, and monitoring.
* Proficient in prompt engineering and evaluation workflows with Langfuse, ensuring continuous quality improvements in LLM-based systems.
* Built dashboards and visual reports using Power BI, Seaborn, and Matplotlib to present insights and model performance to stakeholders.
* Enthusiastic about Generative AI, LLM applications, LangChain, and leveraging AI to solve practical business problems.
* Strong communicator with the ability to explain technical concepts to non-technical audiences and drive actionable insights from data.
* Proficient in end-to-end ML lifecycle including data preprocessing, model selection, hyperparameter tuning, deployment, and performance monitoring.
* Skilled in designing scalable and maintainable Python codebases for both automation scripts and AI-powered solutions.
* Experience integrating ML models into production using tools like Docker, Flask/FastAPI, and cloud platforms such as AWS and GCP.
* Strong background in building and scaling generative AI applications with real-time insights via Langfuse and OpenAI APIs.
* Collaborative team player with proven ability to work across cross-functional teams in Agile/Scrum environments to deliver impactful results.
* Versed in building ML APIs with Flask/FastAPI and deploying services on AWS and Dockerized environments.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | Python, SQL (PostgreSQL, MySQL, SQL Server). |
| **Operating Systems** | Windows, Linux |
| **ML Frameworks & Tools** | TensorFlow, PyTorch, LangChain, Scikit-learn, Hugging Face Transformers, Apache Spark. |
| **Generative AI Technologies** | OpenAI, Google Gemini Pro, Large Language Models (LLMs) like GPT & BERT, GANs, Retrieval-Augmented Generation (RAG), AWS Bedrock, Claude, Titan, Amazon Knowledge Bases |
| **NLP & LLMs** | Text Classification, Sentiment Analysis, NER, Transformers (Hugging Face) |
| **Data Handling** | Pandas, NumPy, Regex, BeautifulSoup |
| **Deployment & APIs** | Flask, Docker (basic), FastAPI (basic), REST APIs |
| **Model Tuning & Evaluation** | GridSearchCV, RandomizedSearchCV, Precision, Recall, ROC-AUC |
| **Data Science & Visualization** | Exploratory Data Analysis (EDA), Statistical Modeling, Feature Engineering, Data Preprocessing, Data Pipelines, Time Series Analysis, Data Mining, Power BI, Tableau, Matplotlib, Seaborn. |

**Professional Experience**

**AI/ML Engineer**

**Client: Cortracker360 | Novi, Michigan. Feb 2024 – PRESENT**

**Responsibilities:**

* Led the end-to-end development and deployment of machine learning models for predictive analytics, customer segmentation, and NLP-driven automation solutions that supported key business initiatives and client needs.
* Designed and maintained scalable ML pipelines using Python, TensorFlow, Scikit-learn, and MLflow, integrating version control and experiment tracking for model reproducibility and lifecycle management.
* Developed advanced NLP models using Hugging Face Transformers, BERT, and spaCy to perform text classification, named entity recognition (NER), and sentiment analysis on customer support data, reducing manual processing time by 40%.
* Designed and deployed scalable .NET Core microservices integrated with AI-powered modules for intelligent advisory and document analysis.
* Built real-time recommendation systems using collaborative filtering and content-based methods to personalize user experience, resulting in a 15% increase in customer engagement on platform features.
* Built and optimized Agentic AI solutions using LangChain, enabling context-aware, autonomous task execution within business workflows.
* Collaborated with cross-functional teams including data engineers, backend developers, and product managers to align model development with business goals and ensure seamless deployment.
* Developed and deployed RAG pipelines using Pinecone and Azure Cognitive Search to improve document retrieval accuracy by 38%.
* Designed and deployed lightweight Retrieval-Augmented Generation (RAG) pipelines using AWS Bedrock with foundation models like Claude for building chatbots that automate RFP responses and knowledge retrieval tasks.
* Conducted thorough exploratory data analysis (EDA), feature engineering, and dimensionality reduction using Pandas, NumPy, and PCA to enhance model performance.
* Implemented MLOps practices for model CI/CD, containerization, and monitoring using Docker, FastAPI, Git, and AWS services like S3, EC2, SageMaker, and CloudWatch.
* Trained and evaluated multiple ML algorithms (XGBoost, LightGBM, Random Forest, Logistic Regression, SVM) for classification and regression tasks, optimizing hyperparameters with Optuna and GridSearchCV to achieve high accuracy and F1-scores.
* Integrated Langfuse to track, log, and monitor large language model (LLM) interactions, improving observability and debugging capabilities.
* Designed and deployed custom anomaly detection algorithms for fraud and abnormal behavior identification, resulting in early alerts that reduced potential financial losses by 20%.
* Developed internal tools and scripts for automated data preprocessing, missing value imputation, data validation, and performance reporting, significantly reducing project turnaround time.
* Researched and implemented state-of-the-art deep learning architectures (CNNs, LSTMs, and Transformers) for computer vision and sequence modeling tasks based on client requirements.
* Integrated models into production-grade microservices with robust REST APIs, enabling scalable, low-latency inference.
* Implemented secure API integrations to AWS Bedrock for prompt orchestration and dynamic response generation using Amazon-hosted LLMs.
* Mentored two junior ML engineers, conducted code reviews, and led knowledge-sharing sessions on best practices in ML, code quality, and MLOps workflows.
* Automated Langfuse logging within LangChain/RAG pipelines to gain full-stack visibility of AI interactions and system health.
* Participated in sprint planning, daily standups, and retrospectives as part of an agile team, ensuring timely delivery of project milestones.
* Contributed to technical documentation including architecture diagrams, model cards, and API specifications, improving team collaboration and onboarding processes.
* Explored and experimented with unsupervised learning techniques (K-means, DBSCAN, autoencoders) for clustering and dimensionality reduction in exploratory projects.
* Collaborated with cross-functional teams to incorporate Langfuse insights into product decision-making and prompt iteration cycles.
* Built custom connectors between Bedrock-hosted models and internal document repositories (e.g., S3, RDS) to enable real-time conversational access to enterprise knowledge bases.
* Actively stayed updated with the latest advancements in AI/ML through research papers, webinars, and implementing experimental projects to test new ideas and methodologies.

**Environment:** Python, SQL, Bash, Scikit-learn, XGBoost, LightGBM, CatBoost, TensorFlow, Keras, PyTorch, Hugging Face Transformers, spaCy, NLTK, BERT, Matplotlib, Seaborn, Plotly, Streamlit, Dash, MLflow, DVC, Docker, FastAPI, Flask, ONNX, AWS EC2, AWS S3, AWS SageMaker, AWS Lambda, Bedrock, AWS CloudWatch, Pandas, NumPy, Dask, PySpark, PostgreSQL, MongoDB, AWS RDS, Redis, Git, GitHub, GitLab, Jenkins, GitHub Actions, Bitbucket Pipelines, Jupyter Notebooks, VS Code, PyCharm, Google Colab, Docker Compose, Conda, Virtualenv, Ubuntu Linux, Windows 11, Jira, Trello, Slack, Microsoft Teams.

**Python Developer**

**Client: Concentrix | Chennai, India Aug 2019 - Dec 2021**

**Responsibilities**

* Developed scalable backend services and RESTful APIs using Flask and FastAPI, improving system response time by 35%.
* Built automation scripts for customer support processes, reducing manual work by 50% using Python, Selenium, and Pandas.
* Designed and optimized SQL queries and stored procedures to improve data retrieval performance from MySQL and PostgreSQL databases.
* Implemented data validation and cleansing pipelines for large customer datasets, ensuring 99.8% data accuracy.
* Collaborated with DevOps teams to containerize applications using Docker and deploy on AWS EC2 instances.
* Integrated third-party APIs (e.g., Salesforce, Twilio) into internal dashboards for real-time customer insights and alerts.
* Used Pandas and Matplotlib to build ad hoc data analysis reports for operations and management teams.
* Developed internal tools and dashboards for QA and testing teams, enhancing traceability and reducing bug detection time.
* Developed automation scripts using Selenium and Python to streamline manual web-based data entry processes, reducing operational effort by 30%.
* Built robust ETL pipelines to extract and transform data from external APIs (including Salesforce and Twilio), and load into PostgreSQL for reporting.
* Designed and implemented custom RESTful APIs using FastAPI to support internal tools and cross-platform integration.
* Wrote efficient and reusable Python code modules for data parsing, cleansing, and validation, improving downstream data quality for analytics teams.
* Collaborated closely with QA to write unit and integration tests using PyTest, resulting in increased code coverage and fewer production issues.
* Participated in daily Agile ceremonies (stand-ups, sprint planning, retrospectives) and maintained task tracking via JIRA, ensuring timely delivery and team collaboration.
* Managed source control and collaboration using GitHub and GitLab; contributed to CI/CD setup using Jenkins and GitHub Actions.
* Worked in Agile environment with 2-week sprints, actively participating in sprint planning, code reviews, and retrospectives.

**Environment:** Python, Flask, FastAPI, SQL, PostgreSQL, MySQL, Pandas, NumPy, Selenium, Docker, Git, GitHub, GitLab, Jenkins, AWS EC2, REST API, Linux, JIRA, Matplotlib, JSON, XML, Postman, VS Code, Agile, CI/CD, Salesforce API, Twilio API

### **Project Portfolio – AI & ML**

### **AI-Based Text Summarizer**

### Built an NLP-based extractive and abstractive summarizer using Hugging Face Transformers and BART models.

### Integrated with Flask REST API and deployed on AWS EC2 for internal document processing.

* Implemented tokenization, attention masking, and input truncation strategies for optimal summarization using transformer-based models.
* Evaluated model performance using ROUGE and BLEU scores, iterating on results to fine-tune output quality for different document types.

### Achieved a 40%-time reduction in manual report writing tasks.

### **End-to-End Scalable ML Model Deployment using AWS, Kubernetes, and CI/CD**

### Achieved 98% accuracy in fraud detection using optimized ML models.

* Reduced deployment time by 50% using CI/CD automation.
* Designed a Docker-based microservices architecture to modularize the model pipeline and enable easy replication across environments.
* Automated deployment workflow using GitHub Actions and Jenkins, enabling seamless model updates and environment provisioning.
* Enhanced model performance monitoring by integrating real-time logging and alerting.
* Improved scalability by deploying on AWS Kubernetes (EKS) and ensuring auto-scaling.

### **Customer Churn Prediction**

* Developed classification models using Random Forest and XGBoost on customer behavior data.
* Improved retention strategy insights by identifying high-risk churn segments with 89% model accuracy.
* Performed extensive EDA and feature engineering, including customer tenure bucketing and frequency-based behavior segmentation.
* Deployed model predictions into a dashboard using Streamlit, enabling the business team to explore churn drivers interactively.
* Visualized feature importances and confusion matrix using Seaborn.

**Licenses & Certifications**

**Advance Your Skills in AI and Machine Learning**

**A blue square with white letters

AI-generated content may be incorrect. Issued by LinkedIn Learning — Apr 2025**Skills: Machine Learning and Artificial Intelligence (AI) [Credential Available on LinkedIn](https://www.linkedin.com/learning/certificates/70f726ab86d7809be1bf993f9ca265b77e740dbb838f8cabb30a8cec9df44be2)

**Career Essentials in Generative AI**

**A four squares of different colors

AI-generated content may be incorrect. Issued by Microsoft & LinkedIn Learning — Apr 2025**Skills: Generative AI, Artificial Intelligence (AI)  
[Credential Available on LinkedIn](https://www.linkedin.com/learning/certificates/28fdf5309318630bf8e3c24dc645fa4c70219ed989609bca9d7ce0336efc9832?lipi=urn%3Ali%3Apage%3Ad_flagship3_profile_view_base_certifications_details%3BohGin%2FoNQbG7Z2thJVyMqg%3D%3D)

**Education**

**Master of Science in Applied Security & Analytics**

University of Findlay, OH, United States.

**Bachelors of Technology in Computer Science & Engineering**

Panimalar Engineering College, Chennai, India.