

MANO BALAJI CHEEPURUPALLI

660-864-4221 | manochEEPurupalli@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Software Engineer with 3 years of experience in data engineering and analytics, specializing in automating reporting pipelines and ETL processes using Python, Flask, and SQL. Skilled in cleaning, joining, and validating multi-source datasets to ensure data accuracy and integrity. Experienced in preparing slide-ready charts and dashboards for leadership audiences and partnering with stakeholders to translate business questions into data solutions.

TECHNICAL SKILLS

- **Programming:** Python, SQL
- **Machine Learning:** Scikit-learn, Random Forest, Logistic Regression, SVM, TF-IDF, Model Evaluation (F1, ROC-AUC)
- **Data Science:** EDA, Feature Engineering, Data Cleaning, Data Quality Assurance (DQA), Statistical Analysis
- **Databases:** Relational Databases, Query Optimization (PostgreSQL, MySQL), Redis
- **APIs & Frameworks:** Flask, FastAPI, REST APIs
- **Cloud:** AWS Sage maker, Azure, GCP
- **Visualization:** Power BI, Matplotlib, Streamlit
- **Tools:** Git, GitHub, CI/CD, Version Control Systems, AI Tools For Development
- **Software Engineering:** Data Structures, Algorithms, Software Development Lifecycle, Debugging, Troubleshooting
- **Professional Skills:** Data Analysis, Automation, Stakeholder Partnership

WORK EXPERIENCE

LV Prasad Eye Institute | *Software Engineer (Data Engineering & Analytics)* Nov 2022 - Oct 2023

- Built and maintained Python/Flask backend systems and SQL-based data pipelines to ingest, clean, and structure clinical data, partnering with clinical stakeholders to inform the ETL design and support patient workflows, analytics, and hospital operations, while leveraging Git and GitHub.
- Designed scalable databases, automated data validation, and delivered real-time Power BI dashboards and slide-ready reports, improving data accuracy, visibility, and decision-making while ensuring HL7/FHIR compliance and following the full software development lifecycle.

GEICO | *Software Engineer (Data Science & Analytics)* Oct 2021 - Oct 2022

- Conducted in-depth data analysis of millions of daily claims and policy transactions using Python and SQL, building scalable ETL pipelines and optimized databases in a distributed computing environment by applying efficient data structures and algorithms.
- Streamlined reporting and data workflows to improve operational efficiency, reduce manual processing, troubleshoot data issues, and enable faster, data-driven decision-making across business and engineering teams.

Value Momentum | *Software Engineer Intern (Data Engineering & APIs)* Jun 2021 - Oct 2021

- Developed Python and SQL-based ETL pipelines and backend services to process large-scale clinical and enterprise data, performing data analysis to improve data availability and system reliability by incorporating AI tools for development.
- Integrated internal tools and external platforms via REST APIs, supported testing and debugging within Agile sprints, and contributed to enterprise software best practices.

PROJECTS

Customer Churn Prediction – End-to-End ML System | [Link](#)

- Built an end-to-end churn prediction pipeline using Python and Scikit-learn, engineering automated feature preprocessing (one-hot encoding, scaling) for high-cardinality data.
- Deployed a production-grade REST API with Flask and an interactive Streamlit dashboard, achieving <200ms inference latency for real-time risk analysis.

Email & SMS Spam Classification System – NLP | [Link](#)

- Engineered an NLP pipeline using TF-IDF and supervised classifiers (SVM, Naive Bayes), achieving 99% F1-score via optimized n-gram ranges and stratified cross-validation.
- Productized the model as a high-performance REST API using FastAPI with Joblib artifacts for low-latency, real-time message filtering.

ArticleGen - AI-Powered SaaS Platform | [Link](#)

- Architected and deployed ArticleGen, a production-grade AI SaaS platform built with Next.js 15 (App Router) and Supabase, integrating Google Gemini Pro for real-time streamed SEO content generation; implemented Server Actions for secure backend execution, TextDecoder-based streaming for low-latency UX, Stripe-powered credit monetization with webhook validation, PostgreSQL triggers for transactional integrity, and role-based authentication via Supabase Auth.
- Designed a scalable, cloud-native infrastructure by containerizing the full-stack application with Docker, publishing versioned images to GitHub Container Registry, and deploying to AWS EC2; established CI/CD workflows, ensured secure environment configuration, implemented real-time credit tracking and usage persistence, and delivered a polished, responsive UI using Tailwind CSS and Framer Motion with dark-mode support.

CERTIFICATIONS

- Microsoft Certified: Power BI Data Analyst Associate: Earned: Sep 2025; Expires: Sep 2026
- IBM Data Analyst Professional Certificate (Coursera): Completed: Feb 2023
- Python: Master Programming and Development with 15 Projects (Udemy): Completed: Jan 2023

EDUCATION

University of Central Missouri

Jan 2024 - Dec 2025

Master of Science, Data Science (GPA: 3.5)

- **Coursework:** Deep Learning, Machine Learning, Natural Language Processing, Big Data Analytics, NoSQL & Database Systems, Cloud Computing, Statistical Learning