

Shruthin Reddy Sainapuram

+1-469-350-7477 | Louisville, KY | shruthinreddysainapuram@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Summary

Data Science enthusiast and practitioner with hands-on experience in machine learning, NLP, deep learning, and business analytics. Skilled in Python, SQL, Power BI, and cloud platforms like Azure and Databricks, with a proven ability to build, optimize, and deploy scalable ML models. Adept at translating complex data into actionable insights through explainable AI and interactive dashboards. Passionate about using data to solve real-world problems and support cross-functional decision-making.

Skills

- **Programming Languages:** Python, C++, C, HTML/CSS, JavaScript, SQL
- **Databases:** PostgreSQL, Azure SQL Database
- **Machine Learning & Deep Learning:** Supervised & Unsupervised Learning, Neural Networks, CNNs, Feature Engineering
- **Libraries & Frameworks:** NumPy, pandas, scikit-learn, Matplotlib, Seaborn, PySpark, Flask, Hugging Face, ChromaDB
- **Natural Language Processing & Generative AI:** Transformers, LLMs, Vector Databases, Prompt Engineering, RAG
- **Cloud & Data Engineering:** Azure, AWS, Databricks, Azure Data Factory, ETL /ELT Pipelines, Data Integration
- **Data Visualization & Business Intelligence:** Power BI, Interactive Dashboards, Data Storytelling
- **Tools & Development Practices:** Git/GitHub, VS Code, Microsoft Excel, Google Colab, Agile Collaboration
- **Mathematics & Statistics:** Probability, Statistical Analysis, A/B Testing, Hypothesis Testing

Work Experience

Grading Assistant **September 2023 – Present**
University of Louisville *Louisville, USA*

- Reviewed **80+ weekly C/C++ programming assignments**, providing detailed code feedback and optimization suggestions, improving student performance by **15%**
- Maintained **100%** accuracy in grading records, ensuring fairness and compliance with **FERPA privacy guidelines**, while helping students identify **coding errors** and **improve problem-solving skills**.

Data Science Intern **August 2022 – May 2023**
VE Projects Private Limited *Hyderabad, India*

- Pulled and cleaned 500K+ customer records using **SQL and Python**, preparing structured datasets for churn analysis under the mentorship of a senior data scientist
- Built machine learning models (**Random Forest, XGBoost**), improving predictive accuracy from 80% to 85% by applying feature selection and tuning techniques
- Partnered with business teams to identify key churn indicators and built **Power BI** dashboards to present actionable insights for reducing attrition.

Projects

Predicting 30-Day Readmission Risk in Diabetic Patients Using Machine Learning

GitHub Link: <https://github.com/shruthin4/diabetic-readmission-prediction>

- Developed a machine learning pipeline using **Python (pandas, scikit-learn, LightGBM)** on 100,000+ diabetic patient records to predict 30-day hospital readmissions, resulting in **62% recall** and **64% accuracy** with balanced performance metrics.
- **Increased model sensitivity by over 60x** (from 0.01 to 0.62) by applying class balancing techniques and optimizing probability thresholds, improving identification of high-risk patients.
- Performed **explainable AI analysis** using **SHAP** to identify top readmission predictors, enabling actionable clinical decisions for resource allocation and discharge planning.

Interactive Resume Chatbot

GitHub Link: <https://github.com/shruthin4/Portfolio-Chatbot>

- Developed a Retrieval-Augmented Generation (RAG) system using **Google Gemini API, ChromaDB**, and **Flask** that transforms resume content into an interactive conversational experience with a responsive **HTML/CSS/JS** interface featuring quick-access buttons for common queries
- Engineered semantic search capabilities using **NLTK, spaCy**, and **vector embeddings** to accurately retrieve and contextualize information from categorized documents, improving response relevance by implementing hybrid retrieval techniques
- Implemented robust system architecture with session management, conversation context awareness, and error handling; containerized with **Docker** and deployed on Render Cloud while optimizing for consistent performance despite free-tier constraints

Handwritten Character Recognition using CNN

- Designed and trained a custom **convolutional neural network** (AdvancedNet) with **PyTorch** using data augmentation and dropout for generalization on the EMNIST handwritten letters dataset.
- Compared training performance across **CPU and GPU** configurations, implemented error analysis by saving **2,400+** misclassified samples to identify confusion patterns between similar letters.
- Built a modular training and evaluation pipeline with cross-entropy loss and Adam optimizer, achieving efficient convergence and over **87% accuracy**, demonstrating strong generalization across 26 letter classes.

News Articles Classification using NLP and Machine Learning

GitHub Link: <https://github.com/shruthin4/News-Articles-Classification>

- Developed an end-to-end text classification pipeline by applying NLP techniques (**lemmatization, TF-IDF, cosine similarity**) and trained models (**Logistic Regression, SVM**) on BBC news articles across five categories.
- Optimized model performance through **hyperparameter tuning** and **K-fold cross-validation**, resulting in consistently high evaluation scores (**96%+** across precision, recall, and F1).
- Visualized article distribution, length, and confusion matrices to interpret model decisions and uncover category-wise performance patterns.

IPL Cricket Data Analysis using Python

GitHub Link: <https://github.com/shruthin4/IPL-Cricket-Analysis-2007-2024>

- Cleaned, merged, and structured **500K+** IPL match and ball-by-ball records using pandas, handling edge cases like Super Overs, no-results, and D/L method matches.
- Performed in-depth analysis on team, player, and venue-level performance using grouped aggregations, pivot tables, and created custom plots (heatmaps, dual bar charts, trend lines) to derive actionable insights.
- Created 15+ insightful visualizations using **matplotlib** and **seaborn** including heatmaps, trend lines, bar charts, and dual-axis plots to explore player stats, team performance, toss impact, and venue-based patterns.

Certifications

- Microsoft Power BI Data Analyst Associate - PL300
- Microsoft Azure Fundamentals – AZ900
- Databricks Data Lakehouse Fundamentals
- Hands-On Approach to AI for Real-World Applications - IIT Kharagpur & TCS iON
- Oracle Cloud Infrastructure 2024 Generative AI Certified Professional
- Oracle Cloud Infrastructure 2023 AI Certified Foundations Associate

Education

University of Louisville | Master of Science in Computer Science and Engineering | **GPA: 3.96/4.0** **August 2023 – Present**

- **Coursework:** Python and Data Analytics, Data Visualization, Introduction to Machine Learning, Web Mining for E-Commerce