Ramakrishna Reddy Vennam
Data Scientist
ramakrishna.r@careernb.com | +1 (656) 208-4865 | FL, USA. | **LinkedIn** | **GitHub**

Professional Summary

- Results-driven Data Scientist with almost **4 years** of experience in extracting, transforming, and analyzing complex datasets to deliver actionable business insights.
- Adept at designing and implementing scalable data models, advanced statistical methods, and customized reporting frameworks to support strategic decision-making.
- Experienced in applying a range of statistical techniques such as regression, classification, clustering, sentiment analysis, and ensemble methods across structured and unstructured data sources.
- Proven track record of enhancing process efficiency and supporting cross-functional teams through data-driven solutions and clear visual reporting.

Education

Masters in Computer Science, University of South Florida, FL**,** USA**.**   **May 2025**

Work Experience

Charles Schwab, FL.   **Jan 2025   Current**
Data Scientist.

- Applied advanced **Natural Language Processing** techniques and anomaly detection methods to analyze transaction narratives and customer behaviour, significantly improving the accuracy of fraud and risk segmentation.
- Employed **SQL, Python, and Pandas** to perform large-scale data cleansing, transformation, and validation across datasets exceeding **3 million records**, leveraging Hadoop and parallel processing to ensure high data quality and performance.
- Engineered and optimized predictive features, implementing machine learning models such as **K-Nearest Neighbours (KNN)** and **Gradient Boosting**, which improved credit risk classification accuracy by **25%.**
- Conducted comprehensive **Exploratory Data Analysis (EDA)** using **Matplotlib and Microsoft Excel** to uncover patterns and correlations in financial datasets, supporting data-driven investment decisions.
- Deployed machine learning models via **Flask-based REST APIs,** utilizing Flask SQLAlchemy for seamless database interaction and integrating AWS services including IAM and RDS for secure and scalable deployment.
- Designed interactive financial dashboards in **Power BI**, delivering real-time performance tracking and executive reporting that enhanced operational efficiency by **30%**.
- Integrated **LLaMA-based large language models** into financial analysis pipelines to enable advanced inference, contextual analysis, and intelligent retrieval of domain-specific insights.

Accenture, India.   **May 2020   Jul 2023**
Data Scientist.

- Engineered custom **PyTorch Data Loaders** to handle complex data formats, enhancing ingestion workflows and reducing pre-processing time by **40%**.
- Built, trained, and evaluated machine learning models using **Scikit-learn, accelerating experimentation cycles** and shortening development timelines.
- Developed and deployed advanced classification models such as **Decision Trees, Random Forests, and SVM** to support targeted marketing efforts, resulting in a **10% increase in average customer lifetime value**.
- Processed and structured large-scale unstructured text data using key NLP techniques, including tokenization, lemmatization, and stop-word filtering.
- Optimized **SQL queries** and resource allocation strategies to improve data processing efficiency, achieving reduction in overall execution time.
- Designed and delivered interactive Tableau dashboards to visualize critical health metrics in real time, supporting leadership with actionable insights and improving strategic decision-making by **25%.**
- Implemented **scalable MLOps pipelines** leveraging FastAPI, Docker, and CI/CD integrations to ensure reliable deployment of machine learning models into production environments.
- Developed resilient data pipelines in **Azure using Data Factory and Data Lake,** enabling robust batch and real-time processing with distributed frameworks like Hadoop and Apache Spark.

Skills

- **Programming Languages:** SQL, Python, R, Scala, C++
- **Database:** MySQL, Mongo DB, PostgreSQL, NoSQL
- **Tools:** R Studio, Jupyter Notebook, Hadoop, Advanced Excel-Analytic Solver Platform, Docker, R-shiny.
- **Python Libraries & Data Analysis:** Pandas, Numpy, Scikit-learn, Py-Torch, Tensor-flow, Keras, Matplotlib, Seaborn, Django, Flask, Flask SQLAlchemy.
- **Machine Learning Skills:** Deep Learning, Decision Trees (CART), Ridge Lasso Regression, Res Net, Naïve Bayes, KNN, SVM, Boosting Random Forest.
- **Cloud Platform:** Amazon Web Services (IAM, EC2, RDS), GCP (Big Query), MS Azure (Azure data factory, Azure Data Lake)
- **Methodologies:** Waterfall, SDLC, SCRUM, Agile.
- **Software Packages:** Microsoft Office (Excel, Word, PowerPoint), SharePoint, Google Suite.
- **Data Skills:** Data Manipulation, Data Visualization, Data Analysis
- **BI Tools:** Tableau, Power BI
- **Soft Skills:** Analytical Thinking, Curiosity & Continuous Learning, Problem-Solving Ability, Adaptability, Time Management

Projects

Social AI Development Project | SuperINTRO   **Nov 2023**
PRESENT Designed and implemented scalable APIs using Python, Django, Flask, and Node.js to support AI-powered agent functionalities. Integrated LangChain, LLMs, and Retrieval-Augmented Generation (RAG) for intelligent decision-making and natural language understanding. Optimized backend systems with MongoDB, Firebase, and Express.js for enhanced performance and scalability. Deployed AI solutions using Docker and Google Cloud Platform (GCP) to ensure high availability and efficient monitoring. Followed CI/CD best practices, automated testing, and code reviews to maintain high-quality standards. Collaborated with cross-functional teams to align project goals and drive innovation in AI-driven networking solutions.

Generative AI Project   **SEP 2022**
NOV 2023 Built an AI-powered search system integrating Azure AI Search, OpenAI GPT-4, and machine learning models. Optimized data retrieval and relevance by embedding AI-powered search capabilities with supervised learning algorithms. Deployed a fully functional system using Azure Functions, ensuring scalability and automation in cloud environments.

Deep Learning & OCR   Number Plate Detection   **SEP 2022**
NOV 2023
Developed a computer vision system using custom YOLO models for real-time number plate detection. Implemented Optical Character Recognition (OCR) models to extract alphanumeric data with 95% accuracy. Deployed the solution as a scalable API, enabling integration with security and automation systems.

Machine Learning   Targeted Article Recommendation System   **MAR 2022**
SEP 2022
Designed a recommendation engine using ML algorithms to personalize content based on user preferences. Achieved 98% accuracy through hyper parameter tuning and model optimization. Successfully deployed on a cloud platform, integrating CI/CD, DevOps, and automated testing for continuous delivery.