

Sreecharan Vanam Data Scientist

vanamsreecharan05@gmail.com | (940)-843-3120 | Dallas, TX | LinkedIn | GitHub

Summary

Data Scientist with 4+ years in machine learning, data engineering, and analytics. Proficient in Python, SQL, and R, with expertise in big data tools like Apache Spark, Delta Lake, and Snowflake. Experienced in deploying predictive models, optimizing data pipelines, and utilizing cloud platforms (AWS, Azure). Skilled in advanced AI, data visualization (Tableau, Power BI), and delivering actionable insights to drive growth and operational efficiency.

Technical Skills

- **Programming Languages:** Python, R, SQL, NoSQL, HTML, CSS, C++, JAVA
- **Frameworks & Tools:** Flask, FlaskAPI, Git, GitHub, Eclipse, Google Colab, A/B Testing, Splunk, MS SQL, MySQL, PostgreSQL, MongoDB
- **Cloud Platform & Streaming:** AWS, Azure, Kubernetes, Docker
- **Data Processing & Streaming:** PySpark, Apache Kafka, Pandas, NumPy, Apache Airflow, Snowflake, ETL, MS Excel, Matplotlib, Seaborn, Plotly, scikit-learn, statsmodels, Data Flow, Data Bricks
- **Machine Learning Algorithms:** Random Forest, Linear Regression, PyTorch, SVM, Decision Tree, ARIMA, XGBoost, TensorFlow, Deep Learning Algorithms, LSTM, SciPy, NLTK, FFNN, RNN, CNN, Siamese Network, GAN, LLM, BERT, TFLOPS, NLP, Word2Vec
- **Statistical Analysis and Modeling:** ARIMA, PCA, ANOVA, descriptive statistics, probability distributions, Regression models, LDA, TF-IDF, Mixed-effects models and Longitudinal data analysis, hypothesis testing and correlation analysis
- **Visualization:** Power BI, Tableau, Matplotlib, Seaborn, Plotly

Professional Experience

Data Scientist, NCR Atleos

08/2024 – Present | Frisco, TX

- Engineered predictive models for ATM cash flow forecasting using Python, ARIMA, and XGBoost, analyzing historical transaction data to optimize cash regime strategies and reduce cash shortages across deployed ATMs.
- Utilized SQL and AWS Glue to preprocess and analyze 15 million transaction records, enforcing ETL processes that improved data quality and accessibility, thereby enhancing the accuracy of cash flow predictions
- Developed interactive dashboards in Power BI to visualize cash flow trends and operational metrics, empowering stakeholders to make informed decisions rapidly and improving response times for effective cash management.
- Collaborated with cross-functional teams using Agile methodologies to integrate predictive insights into existing workflows, facilitating seamless model deployment and enhancing communication between data science, finance, and operations for strategic decision-making.
- Monitored model performance using metrics like precision and recall, using Python and Scikit-learn for steady gain, for predictive accuracy and reliability in cash flow forecasting for a responsive cash management approach.
- Conducted wide analyses of cash flow patterns, using statistical techniques and data visualization tools to present findings to stakeholders, improved overall financial planning and resource allocation.

Research Assistant, University of North Texas

09/2022 – 11/2023 | Denton, TX

- Enhanced the lab's responsible AI initiatives by improving ETL processes and model validation techniques, leading to more robust and reliable AI systems.
- Boosted the lab's research output by leading scientific writing and data management efforts, crucial for developing accurate and effective predictive models.
- Spearheaded the integration of Large Language Models (LLMs) into generative AI frameworks, significantly enhancing algorithm performance and accuracy, achieving a 20% improvement in model prediction accuracy.

Data Scientist, IPExcel Services Pvt. Ltd.

02/2020 – 08/2022 | Bengaluru, India

- Increased data research efficiency by 30% by leveraging SQL (JOINS) for complex data extraction and transformation tasks using Python (Pandas, NumPy) for data manipulation and enforced ETL processes with AWS Glue and Redshift.
- Utilized AWS S3 for scalable storage and AWS Lambda for serverless data processing, ensuring efficient handling and real-time analytics of large datasets. Automated workflows with these tools for data operations.
- Applied statistical modeling techniques to assess patent trends and technological landscapes. Used Statsmodels for hypothesis testing and data inference, providing actionable insights for enhancing IP strategies.
- Developed and deployed predictive models for enhancing Intellectual Property portfolios using Scikit-Learn for classification and regression tasks and TensorFlow for deep learning, driving strategic decision-making for leading multi-national companies.
- Employed advanced algorithms such as Random Forest and Gradient Boosting Machines to improve predictive performance. Utilized Python libraries like XGBoost and LightGBM for model fine-tuning and accuracy.
- Improved client decision-making by 20% via data visualization with Tableau and Matplotlib, creating interactive dashboards and detailed reports to present complex insights, enhancing strategic and operational decisions.

Data Scientist, Treflo

05/2018 – 12/2019 | Bengaluru, India

- Directed the development of a recommendation system by preprocessing datasets from user interactions and 1,000 product catalogs using Python (Pandas and NumPy) for EDA to identify 20+ key features for model training.
- Enforced hybrid recommendation system combining filtering with matrix factorization techniques (SVD, ALS) and content-based filtering using Scikit-learn for traditional methods and TensorFlow for deep learning.
- Integrated the recommendation engine into the e-commerce platform via RESTful APIs. Deployed on AWS, utilizing EC2 for computing and RDS for managed databases, handling up to 100,000 real-time recommendations daily.
- Evaluated model performance using precision, recall, F1-score, and MAP, conducting A/B testing with 10,000 users to compare the new system against previous versions, analyzing results with Google Analytics.
- Developed interactive Tableau dashboards to visualize system performance and user metrics, preparing 15 detailed reports and aiding strategic decision-making.

Education

Master of Science, University of North Texas
Computer Science

08/2022 – 05/2024 | Denton, TX

Bachelor of Science, Vellore Institute of Technology
Computer Science and Engineering

06/2016 – 05/2020 | Tamil Nadu, India

Research

- Investigating Code Generation Performance of Chat-GPT with Crowdsourcing Social Data
- Collaborated with Argonne National Laboratory to develop and analyze Stable Diffusion models
- Awarded the Best Track Paper and Youngest Co-Author at IEEE COMPSAC 2023

IEEE COMPSAC 2023

March 2023

May 2023

Academic Projects

DistilBERT-based Question and Answering System Adapted for UNT FAQs | Web Scraping, Python

Sept 2023

- Improved response time by 40% for UNT FAQs by enhancing the DistilBERT-based QA system, increasing accessibility for prospective students.
- Streamlined information retrieval processes by conducting web scraping and data organization of UNT website FAQs, resulting in a 25% increase in user satisfaction.
- Implemented optimized data preprocessing and model training techniques, ensuring 95% accuracy in predictions. Empowered prospective students with swift access to accurate answers from departmental webpages, revolutionizing UNT information retrieval.

Sentiment Analysis and Evaluating Code Quality Generated by ChatGPT | Python, NLP, NLU, LLM

May 2023

- Conducted an in-depth analysis of code quality generated by ChatGPT, providing valuable insights into code assessment and public sentiment.
- Utilized a combination of sentiment analysis techniques to evaluate generated code, and incorporated computer vision methods to analyze code structure and presentation, improving overall evaluation accuracy by 30%.
- Enhanced understanding of ChatGPT’s code generation capabilities, contributing to improvements in AI code generation.

Generating Realistic and Diverse Textual Movie Reviews with GPT-2 | Python, Deep Learning

May 2024

- Engineered a deep learning model using GPT-2 to generate realistic and diverse human-like movie reviews, leveraging the IMDb dataset, resulting in a 25% increase in review authenticity.
- Optimized the GPT-2 model through meticulous preprocessing and fine-tuning, enhancing text generation accuracy by 30%, and producing high-quality, varied content.
- Implemented advanced sampling and evaluation techniques, leading to a 20% improvement in the naturalness and coherence of AI-generated text.

Fine-Tuned BERT for Domain-Specific Question Answering | Python, NLP

April 2024

- Developed and fine-tuned a BERT model for a domain-specific question-answering system using the SQuAD dataset, increasing response accuracy by 35%.
- Enhanced the model’s precision through advanced NLP preprocessing techniques, achieving a 40% improvement in F1 scores and Exact Match metrics.
- Designed a user-friendly interface for the QA system, enabling swift and accurate access to information, resulting in a 20% boost in user satisfaction.