Venkata Phani Krishna Puppala

venkatapuppala31@gmail.com | +1(341)- 356-2422| United States

SUMMARY

- Over 4 years of experience in machine learning engineering and data engineering across diverse domains.
- Expertise in building end-to-end ML workflows including data ingestion, preprocessing, feature engineering, model training, tuning, and deployment.
- Expert in applying advanced machine learning algorithms such as Random Forest, SVM, LSTM, CNN, Transformer models, and cutting-edge generative AI techniques to solve complex problems.
- Extensive experience architecting scalable data pipelines and cloud solutions using AWS (S3, Redshift) and Snowflake, ensuring robust data ingestion and processing at scale.
- Designed and deployed fully automated, optimized ETL pipelines with data encryption to support high-performance AI/ML model training and real-time analytics.
- Engineered scalable RESTful APIs in Python to seamlessly integrate machine learning models into production environments, enabling data-driven decision-making.
- Proficient in performing in-depth exploratory data analysis (EDA), statistical modeling, and advanced visualization using Python, R, Tableau, Power BI, and SharePoint.
- Collaborated closely with cross-functional stakeholders to translate business goals into actionable AI/ML solutions, driving measurable impact and insights.
- Strong command over multiple programming languages including Python, R, SQL, Scala, Java, C++, and shell scripting, adaptable across diverse OS platforms.
- Leveraged TensorFlow and Keras frameworks to rapidly prototype, train, and deploy state-of-the-art deep learning models for various predictive analytics tasks.
- Demonstrated expertise in managing large-scale datasets with rigorous data governance, ensuring data integrity, security, and compliance with industry standards.

EDUCATION

University of the Pacific | Stockton, CA | GPA: 3.47+ **May 2025**

Master of computer science and engineering

Jawaharlal Nehru Technological University | Hyderabad | CGPA: 6.2 **Jan 2023**

Bachelor of computer science and engineering

WORK EXPERIENCE

Gen AI Engineer (Data analytics) | Sakshi Ltd, Hyderabad **June 2022 – July 2023**

- Built end-to-end ML workflows: data ingestion (AWS, Snowflake), preprocessing, modeling, and deployment.
- Conducted EDA with Pandas, NumPy, Seaborn, Stats models, and Pandas Profiling.
- Trained Random Forest models on customer web activity for conversion prediction.
- Applied PCA, feature engineering, and hyperparameter tuning to optimize models.
- Worked with ML algorithms: Linear/Logistic Regression, Decision Trees, SVM, K-Means, Boosting.
- Built LSTM models for time series and predictive analytics.
- Solid foundation in Generative AI, deep learning, and synthetic data techniques.
- Designed and implemented data pipelines for large-scale model training and evaluation.
- Integrated ML models into production environments with APIs and cloud services.
- Utilized TensorFlow/Keras for model prototyping and rapid experimentation.
- Collaborated with cross-functional teams to align AI models with business objectives.
- Explored Generative AI techniques including transformer models and synthetic data generation for advanced analytics.

Data Engineer | Raghava Constructions (India) Private Limited, India **Feb 2020 – May 2022**

- Built AWS Data Pipelines for automating data loads from S3 to Redshift and other destinations.
- Performed ETL using Redshift and Pyspark, transforming data from sources like Excel, CSV, Oracle, and flat files.
- Developed T-SQL scripts, stored procedures, views, and triggers for business reporting.
- Designed logical/physical data models and metadata repositories using ERwin and MB MDR.
- Automated ETL processes via shell scripts and DataStage job orchestration.
- Created Pyspark scripts for data encryption using hashing algorithms on sensitive fields.
- Built Python-based REST APIs for revenue tracking and analytics.
- Supported AI/ML teams with clean, structured datasets for model training and evaluation.
- Analyzed departmental data and presented KPIs to leadership via SharePoint-based dashboards.
- Assisted in sourcing content for ML model reference and data versioning.
- Optimized query performance in Redshift using distribution keys, sort keys, and vacuum strategies.
- Implemented version control and CI/CD practices for ETL code using Git and deployment pipelines.

Junior Data Analyst | Yogicareers, India **Aug 2019 – Jan 2020**

- Analyzed and documented business requirements to support recruitment data needs.
- Collaborated with Business Analysts to ensure accurate candidate and client data reporting.
- Executed ad-hoc SQL queries for recruitment analysis and trend monitoring.
- Created automated monthly/quarterly reports using Teradata SQL and Unix BTEQ scripts.
- Validated data accuracy and consistency in staffing reports.
- Queried and analyzed Hadoop data using Hive to identify data quality issues.
- Performed source-to-target data validation with JSON and aggregation functions.

KEY PROJECTS

Customer Conversion Prediction

- Built end-to-end ML pipeline using AWS and Snowflake to predict customer conversions from web activity.
- Conducted EDA, feature engineering, and applied PCA for dimensionality reduction.
- Trained and optimized Random Forest models with hyperparameter tuning.
- Deployed models via APIs for real-time prediction, aligning results with business goals.
- Leveraged TensorFlow/Keras to prototype and deploy LSTM and CNN models for time series forecasting and classification tasks.
- Collaborated with cross-functional teams to ensure AI models address business needs and drive actionable insights.

Automated ETL Pipeline for AI/ML Support

- Developed automated ETL pipelines in AWS (S3 to Redshift) using Pyspark and SQL.
- Optimized data workflows and query performance; implemented data encryption for sensitive fields.
- Built Python REST APIs for revenue analytics; ensured version control and CI/CD for ETL code.
- Delivered clean datasets to AI/ML teams, improving model training efficiency.
- Designed logical and physical data models using ERwin to support scalable data architecture.
- Automated ETL orchestration with shell scripts and DataStage jobs, enhancing pipeline reliability.
- Monitored and analyzed KPIs via SharePoint dashboards, enabling data-driven leadership decisions.

TECHNICAL SKILLS

Machine Learning & AI: Linear & Logistic Regression, Decision Trees, Random Forest, SVM, K-Means, KNN, Naive Bayes, Gradient Boosting, AdaBoost, PCA, LDA, NLP, Deep Learning (ANN, CNN, RNN), Transformer Models, TensorFlow, Keras, Generative AI, Synthetic Data

Data Engineering & Big Data: AWS (S3, Redshift), Pyspark, Hadoop Ecosystem (HDFS, Hive, Pig, Sqoop, Yarn, Spark, Spark SQL, Kafka), Hortonworks, Cloudera, Data Pipeline Automation, ETL, Shell Scripting, Data Modeling (ERwin), CI/CD, Git.

Data Analytics & Visualization: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), R (ggplot2), SQL (Teradata, Oracle, MySQL, SQL Server), Tableau, Power BI, QlikView, D3.js, SharePoint Dashboards

Programming Languages: Python, R, SQL, Scala, Java, C, C++, C#, MATLAB, UNIX Shell Scripting, COBOL

Databases & Tools: Teradata, Oracle 9i/10g, DB2, SQL Server, MySQL, Teradata SQL Assistant, PyCharm, Autosys

Operating Systems: Linux, UNIX, Windows, Mac OS, z/OS