Vikas Reddy

Data Scientist | Machine Learning Engineer |AI Engineer

+1 (510)-820-3632 | vikasmle25@gmail.com | www.linkedin.com/in/vksrdy/ | Hayward, CA

SUMMARY

- Data Scientist with 3.5 years of experience specializing in data warehousing, extraction, and modeling, leveraging tools like Python, SQL, AWS, Machine Learning, AI and Power BI for efficient data processing and deriving actionable insights.
- Expertise in developing and optimizing machine learning algorithms, including regression, classification, and time series models, boosting prediction accuracy and improving business outcomes, enabling more informed decision-making and strategic growth.
- Strong background in statistical modeling and analysis, Responsible AI, utilizing Python, R, and advanced techniques to extract actionable insights from complex datasets, driving data-driven decision-making and enhancing problem-solving capabilities across various domains while maintaining data integrity.
- Proven track record in collaborating with cross-functional teams to implement data-driven solutions, enhancing decision-making, optimizing operational efficiencies, and delivering measurable, impactful results across a wide range of industries, business sectors.

EXPERIENCE

Johnson & Johnson, NJ | Data Scientist (ML Engineer) **Feb 2024 - Current**

- Developed a cloud-native MLOps platform on AWS for scalable AI/ML deployment and management using Python, TensorFlow, and Kubernetes, ensuring 99.9% availability while handling petabytes of medical imaging data.
- Utilized TensorFlow and PyTorch for deep learning analysis of medical images, achieving 90% accuracy on 100,000 Xray/MRI scans. Conducted A/B testing to compare model performance for data-driven decision-making.
- Integrated HL7 standards and Mirth for seamless interoperability between Electronic Health Record (EHR) systems and ML pipelines, automating 80% of manual data entry tasks.
- Implemented secure Flask/FastAPIs with JWT authentication and data encryption for HIPAA compliance. Also designed automated anomaly detection to identify unusual medical data patterns for proactive intervention.
- Employed advanced optimization techniques such as quantization and pruning (TensorFlow-Model-Optimization) to reduce ML model size by 60% without sacrificing accuracy.
- Integrated wearable device data from sources like Apple Watch and Fitbit using RESTful APIs and web sockets for continuous monitoring and personalized healthcare insights.
- Established a scalable data ingestion pipeline on AWS (S3, Glue, Athena) for processing and storing terabytes of medical data from diverse sources, ensuring reliability and scalability.
- Utilized pre-built machine learning algorithms (scikit-learn, XGBoost) for predictive analytics on medical data, resulting in an 18% improvement in diagnosis accuracy.
- Designed, built, and trained ML models such as Linear Regression, Decision Tree, Random Forest Regressor, Boosted Decision Trees regressor, CatBoost, LightGBM, XGBoost Regressor, Time series forecasting -ARIMA, ARIMAX
- Build and train Machine Learning models to process videos, image, text and evaluate Deep Learning models performance (ANN, CNN), NLP (Word2vec, Glove, Tf-IDF matrix, TensorFlow, BERT).
- Collaborated with cross-functional teams to present findings, communicated technical concepts to both technical and non-technical stakeholders, demonstrating exceptional communication and problem-solving skills while driving project success.

TCS, India | Data Engineer **Feb 2020 - Dec 2022**

- Built and optimized data pipelines with Azure DF and Synapse to process and analyze large volumes of satellite radio streaming data for audience insights.
- Deployed machine learning models on Azure VMs to personalize content recommendations, enhancing listener engagement based on real-time preferences and behavior.
- Leveraged Azure Synapse for advanced analytics to provide data-driven insights on user behavior, ad performance and content popularity, optimizing SiriusXM's programming strategy.
- Experimented with GATO for audio enhancements, including sound classification and adaptive streaming.
- Implemented Backend Microservices with Java, Spring Boot, and rest API for secure patient data processing adhering HIPAA rules.
- Optimized distributed systems using Multi Threading, confluent Kafka, Qlik, improving system throughput by 20%.
- Streamlined CI/CD pipelines automation with Docker, Jenkins increasing reliability in deployment cycles.
- Collaborated with other teams in agile sprints to help troubleshoot, code reviews and enhance performance across production environments.
- Deployed and managed cloud-based infrastructure on AWS, including EC2 instances, S3 buckets, RDS databases, and Lambda functions, ensuring system scalability, high availability, and enhanced performance across the environment.

Academic Projects

AI-Driven Job Discovery and Recommendation Engine

- Developed and deployed an AI-powered job search engine using RAG pipelines, BAAI-BGE-Large-EN, and ChatGPT, improving candidate-job relevance and cutting skill-to-job matching time by 50%.
- Built predictive models and recommendation systems on academic/research datasets using supervised & unsupervised learning, enhancing personalization and discovery.
- Piloted large-scale job relevancy engine with 50+ students, handling 1000+daily queries, validating scalability in real-world conditions.
- Managed AWS cloud infrastructure (EC2, S3, scaling policies) for high availability and cost efficiency, while mentoring 50+ students on Python, ML, NLP, and databases.

Study to predict Diabetes in Women

- Developed Statistical Modeling techniques to predict the susceptibility of diabetes in women from the Pima Indian Tribe, utilizing Logistic Regression, XG Boost, and Random Forests.
- Attained an 88% accuracy with the Random Forests model, while XG Boost and Logistic Regression demonstrated accuracies of 82% and 78%, respectively.
- Employed advanced feature engineering methods such as feature selection, scaling, and dimensionality reduction to enhance the performance of the models

TECHNICAL SKILLS

Languages: Python, Java, SQL, R
Machine Learning & AI: Regression, Classification, Decision Trees, Random Forests, Neural Networks (ANN, CNN), SVM, KNN, K-Means, Bayesian Methods, Time Series Forecasting, A/B Testing, Hypothesis Testing, ANOVA, NLP, Sentiment Analysis, LLMs, Generative AI, RAG, Deep Learning, MLOps.
Frameworks & Packages: TensorFlow, PyTorch, Scikit-learn, Keras, NumPy, Pandas, SciPy, Matplotlib, Seaborn, Spark, LangChain
Cloud & Data Engineering: AWS (SageMaker, Bedrock, Glue, Athena, Lambda, Step Functions, EventBridge, S3, EC2), Azure ML, GCP, Apache Kafka, Snowflake, dbt, Docker, Kubernetes, Jenkins, GitHub Actions, CI/CD
Databases & Visualization: MySQL, PostgreSQL, SQL Server, Oracle, MongoDB, Cassandra, Vector DB, BigQuery, Power BI, Tableau, Excel
Development Tools & IDEs: VS Code, PyCharm, Jupyter Notebook, Google Colab, Databricks, Figma
Other Skills: Data Wrangling, EDA, Data Modeling, Predictive Analytics, Data Visualization, Pattern Recognition, Statistical Analysis, Agile, SDLC, Jira.
Operating Systems: Windows, Linux, Mac OS.

CERTIFICATION

AWS certified Machine Learning Engineer

AWS Certified AI Practitioner

Snowflake certified SnowPro Associate (Platform)

EDUCATION

Master of Science – Statistics-Data Science

California State University- East Bay, Hayward, CA

Bachelor of Technology in Mechanical Engineering

Jawaharlal Nehru Technological University Hyderabad, India.