

MANISH VANGARA

(572) 205-6042 • manish.vangara07@gmail.com • github.com/ManishVangara

Data Scientist with 5+ years of experience turning complex data into actionable insights that drive measurable business impact. Skilled in clustering, regression, decision trees, and building agentic AI systems that automate decision-making and boost efficiency. Delivered \$2M in cost savings and 20% operational gains through data-driven strategies and clear communication across teams.

SKILLS

Programming & Data Manipulation: Python (NumPy, Pandas, SciPy, Statsmodels), R (dplyr, ggplot2, caret), SQL (PostgreSQL, MySQL), PySpark

Data Mining & Statistical Methods: Clustering (K-Means, Hierarchical, DBSCAN), Regression Analysis (Linear, Logistic, Polynomial), Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Collaborative Filtering, Market Basket Analysis, Matrix Factorization, Dimensionality Reduction (PCA, t-SNE), Time Series Analysis (ARIMA, Prophet)

Machine Learning & Advanced Analytics: Scikit-learn, PyTorch, TensorFlow, XGBoost, Random Forest, Gradient Boosting, Ensemble Methods, Statistical Testing, Hypothesis Testing, Cross-Validation, Model Selection, Feature Engineering, Hyperparameter Tuning, Reinforcement Learning

Agentic AI & LLM Systems: LangChain, LangGraph, OpenAI API, Autonomous Agents, Multi-Agent Systems, Tool Calling, ReAct Frameworks, Chain-of-Thought Prompting, Agent Memory Systems, Workflow Automation

Natural Language Processing: Sentiment Analysis, Text Classification, Entity Recognition, Word Embeddings (Word2Vec, GloVe), Transformers

Cloud & Data Platforms: AWS (SageMaker, S3, Glue, Redshift), Databricks, Snowflake, ETL pipelines, Spark Streaming

Data Visualization & Communication: Power BI, Tableau, Matplotlib, Seaborn, Plotly, D3.js, Data Storytelling, Technical Report Writing

PROFESSIONAL EXPERIENCE

HALLIBURTON

Houston, TX

Data Scientist

August 2024 – Present

- Extracted complex drilling sensor data using advanced clustering techniques (K-Means, DBSCAN) and decision tree algorithms, uncovering patterns that reduced non-productive time by 15% and delivered \$2M in annual cost savings through predictive maintenance insights.
- Applied collaborative filtering and matrix factorization methods to analyze equipment performance across 50+ drilling sites, identifying optimal maintenance schedules and transforming raw data into actionable plans that improved operational efficiency by 20%.
- Pioneered agentic AI system using LangChain and autonomous agents to automate drilling parameter optimization workflows, enabling real-time decision-making that reduced manual intervention by 40% and improved operational response times.
- Architected statistical models using regression analysis and support vector machines (SVM) to process 10K+ sensor readings/minute, serving real-time predictions through RESTful APIs to 200+ field engineers with 99.7% system uptime.
- Implemented k-nearest neighbors (KNN) algorithms for real-time anomaly detection in drilling operations, processing streaming data with sub-second latency and automated alert generation to prevent equipment failures.

UNIVERSITY OF CENTRAL OKLAHOMA

Edmond, OK

Graduate Data Scientist

Jan 2024 – August 2024

- Analyzed 15,000+ student records using clustering and regression analysis techniques, uncovering enrollment patterns and performance trends that informed strategic interventions and cut reporting delays by 25%.
- Developed RESTful API endpoints for ML model predictions with authentication, rate limiting, and comprehensive logging, serving real-time student risk assessments to academic advisors.
- Leveraged decision trees and statistical modeling to process complex datasets, collaborating with academic teams to transform research information into actionable insights for student success initiatives with 88% prediction accuracy.
- Automated ETL pipelines using Python, SQL, and AWS Glue for data ingestion, transformation, and model feature engineering.
- Designed responsive Power BI dashboards and ad hoc & canned reports, providing stakeholders with real-time KPI monitoring.

VERTOCITY

Hyderabad, India

Data Scientist

Jan 2020 – May 2023

- Identified at-risk student populations using decision trees, random forests, and support vector machines (SVM) to mine complex behavioral data, enabling targeted interventions that boosted completion rates by 8%.
- Validated model reliability through rigorous statistical analysis using regression techniques and cross-validation methods, collaborating with product teams to transform findings into actionable business strategies.

VERTOCITY

Hyderabad, India

Data Analyst (Applied Data Science)

May 2019 – Aug 2019

- Developed data pipelines using Python and SQL to process and analyze student performance data across multiple educational platforms.
- Built initial prototypes of predictive models for student engagement tracking, contributing to 5% improvement in early intervention rates.
- Created data visualizations and dashboards using Tableau to present insights to product and engineering teams.

EDUCATION

UNIVERSITY OF CENTRAL OKLAHOMA

Edmond, OK

Master of Science in Data Science

2023 – 2025

- Awards: **Presidential Honor Award**

- Coursework: Advanced Statistics, Machine Learning, Applied Regression, Artificial Intelligence, Data Mining, Software Engineering, Data Structures

PROJECTS

LLM-Powered Student Support Chatbot | Python (LangChain, OpenAI), React.js, Docker, AWS

- Built full-stack conversational AI using GPT-4 and LangChain with React.js frontend, reducing support response time by 60%.

Retail Customer Segmentation & Recommendation Engine | Python (Scikit-learn), Market Basket Analysis, Collaborative Filtering

- Used collaborative filtering, KNN, and market basket analysis to reveal customer segments and drive personalized recommendations.

CERTIFICATIONS

Salesforce Certified Agentforce Specialist

(In Progress)

AWS Certified Cloud Practitioner