

PRAGATI KHEKALE

Dallas, TX | +1 (214) 494-1191 | khekalepragati01@gmail.com

[LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Scientist with 2.5+ years building production machine learning (ML) models, predictive analytics systems, and automated data pipelines at Fortune 500 companies. Delivered \$100K+ cost savings through predictive modeling (XGBoost, deep learning) and statistical analysis. Expertise in Python, SQL, Machine Learning, Natural Language Processing (NLP), Deep Learning, and data visualization (Tableau). MS in Data Science with strong foundation in statistics, experimentation design, and cross-functional collaboration. Passionate about leveraging data-driven insights and AI to solve complex business problems.

TECHNICAL SKILLS

Machine Learning & AI: XGBoost, Random Forest, Regression, Classification, Clustering, Deep Learning (CNNs, Transfer Learning), NLP, LLMs, LangChain, Computer Vision (OpenCV, PoseNet), Feature Engineering, Model Evaluation, A/B Testing
Programming & Data Engineering: Python (Pandas, NumPy, SciPy, Scikit-learn, Matplotlib, Seaborn), SQL, R, ETL Pipelines, Data Cleaning, Data Mining, Data Wrangling, Big Data, Apache Spark
Deep Learning Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, Flask, FastAPI, Docker, MLflow
Statistics & Analytics: Hypothesis Testing, Statistical Modeling, Predictive Analytics, Causal Inference, Bayesian Analysis, Regression Analysis, Time Series, Experimentation Design
Tools & Platforms: Tableau, Power BI, Snowflake, BigQuery, MySQL, AWS (S3, EC2, SageMaker), Git, Jupyter, ArcGIS Pro, Agile/Scrum

PROFESSIONAL EXPERIENCE

Atmos Energy (Fortune 500) Dallas, TX

Data Scientist August 2023 – Present

- **Predictive Modeling:** Built and deployed end-to-end machine learning models (XGBoost, deep learning, causal inference) in Python/PyTorch for predictive maintenance, improving forecasting accuracy by 40% & delivering \$100K+ in cost savings
- **Data Pipeline Engineering:** Designed automated ETL pipelines integrating 5+ disparate data sources using Python, SQL, and Snowflake, achieving 95%+ data quality and reducing manual data processing by 60%
- **Statistical Analysis & KPIs:** Led cross-functional collaboration to define 50+ business KPIs and experimentation frameworks, driving data-informed decision-making across Pipeline Integrity division. Applied causal inference to measure the impact of KPI's
- **Geospatial Data Science:** Developed Python-based consequence layer algorithm deployed in ArcGIS Pro for spatial risk modeling and predictive analysis of pipeline infrastructure
- **Data Visualization:** Created executive-level Tableau dashboards and data visualizations for PHMSA compliance, translating complex analytical findings into actionable business insights

CCC Intelligent Solutions Chicago, IL

Data Scientist Intern May 2022 – August 2022

- **ML Pipeline:** Engineered machine learning pipeline implementing SISA machine unlearning algorithm in Python/Scikit-learn, reducing model retraining costs by 70-90% while maintaining 95.6% accuracy
- Conducted A/B experimentation across shard configurations to optimize production deployment; built scalable data science pipeline with model monitoring and performance optimization
- Collaborated cross-functionally with engineering and data privacy teams, establishing production-ready ML frameworks and increasing customer engagement by 20%

Visteon Corporation Pune, India

Machine Learning Engineer Intern July 2019 – September 2019

- Integrated PoseNet computer vision model into ADAS for real-time object detection, achieving 30% inference efficiency improvement through model optimization
- Optimized ML model deployment pipeline (conversion, quantization, profiling) in C++ and Python for safety-critical automotive applications

KEY PROJECTS

Oral Cancer Detection – Deep Learning Classification | Python, TensorFlow, VGG-16 (CNN), Flask, Docker

- Developed deep learning classification model achieving 94% accuracy on 6,500+ medical images using transfer learning (VGG-16); deployed as Flask REST API for real-time predictive inference, accelerating diagnostic processes by 30%

Personalized Job Search AI Agent | Python, LangChain, LLM, NLP, MLflow, Git

- Engineered autonomous AI agent using LangChain and Large Language Models (LLMs) with MLflow experiment tracking, automating data collection and NLP-driven analysis, reducing manual tasks by 60%

Netflix Content Recommendation Engine | R, NLP (tf-idf), Statistical Modeling, Data Mining, Clustering

- Built NLP-powered recommendation system analyzing 6,000+ titles using tf-idf vectorization, content similarity, and clustering, delivering personalized suggestions through statistical modeling

EDUCATION

Master of Science in Data Science January 2021 – December 2022

Illinois Institute of Technology, Chicago, IL

Coursework: Machine Learning, Deep Learning, Statistical Learning, Computer Vision, Data Mining, Big Data Technologies, Database Systems