Pragati Khekale

Dallas, TX, USA | +1 (214) 494-1191 | khekalepragati01@gmail.com
LinkedIn | GitHub |Tableau|Porfolio

SKILLS

Languages & Tools: Python, R, SQL, C++, Git, Jupyter, Unix/Linux, ArcGIS Pro.

Machine Learning: Regression, Classification, CNN (Deep Learning), XGBoost, SVM, KNN, Transfer Learning, Open Set Recognition.

ML/Statistical Libraries &Tools: Tensorflow, Keras, PyTorch, Scikit-Learn, GeoPandas, Statistical Analysis, A/B Testing, Hypothesis Testing

Analytical & Data Eng: Data Cleaning, Feature Engineering, ETL, GeoPandas, Numpy, pandas, etc

Data Visualization: Tableau, Power BI, Matplotlib, Seaborn

Cloud &Database Infrastructure: Snowflake, BigQuery, AWS

Soft Skills: Stakeholder Collaboration, Cross-Functional/ verbal and written Communication skillls, Project Management, Data interpretation, Technical Training

EXPERIENCE

Atmos Energy | Data Analyst | *Dallas, TX | Aug 2023 – Present*

- *Regulatory Data Stewardship*: Collaborated with Engineering and Compliance teams to define long-term data workflows, ensuring data integrity and alignment between analytical initiatives and federal regulatory requirements.
- *Experimental Research & Design*: Developed novel predictive models (XGBoost Regression , Deep Learning) using PyTorch, ArcGIS Pro as an R&D initiative to identify complex patterns in infrastructure data, reducing operational inefficiencies and $100k+ cost savings.
- *Data Pipeline Management*: Designed and automated end-to-end ETL pipelines in Python to integrate disparate datasets across database & business units, resolving data discrepancies and maintaining high data quality standards.
- *Technical Reporting & Publication*: Analyzed and visualized comprehensive PHMSA annual reports, translating complex technical findings for both internal executive leadership and external government stakeholders.

CCC Intelligent Solutions | Data Scientist Intern | Chicago, IL | May 2022 – Aug 2022

- Optimized a privacy-preserving ML update pipeline using SISA machine unlearning, enabling selective retraining with ~lower retrain cost versus full retraining while maintaining ~95.6% model accuracy under high deletion workloads.(Python)

Visteon Corporation | Machine Learning Intern | *Pune, India | Jul 2019 – Sep 2019*

- Integrated PoseNet into ADAS systems for real-time object detection, improving openvino model efficiency (C++).
- Contributed to the full ML model lifecycle, from design to deployment, supporting cross-functional collaboration in engineering and product teams.

PROJECTS

(Enterprisewide) Key Performance Indicators Dashboards Enterprisewide| Atmos Energy | *2024–Present*

- Led a cross-functional analytics initiative to define and monitor six key KPIs (risk assessment, asset performance, compliance, etc.) for Pipeline Integrity.
- Managing complete data lifecycle from data mining, ETL processes, resolving data discrepancies, Data Cleaning to Data Validation to maintain Data Integrity.
- Built multiple executive dashboards in Tableau to support decision-making and operational planning.
- Enhanced data transparency and enabled predictive insights for proactive risk management.

OpenMax Deep Learning Optimization for Classification | Illinois Institute of Technology | *2022*

- Implemented transfer learning-based computer vision model with OpenMax activation for open-set recognition in Tensorflow.
- Trained base model on CIFAR-10 and transferred weights to MNIST using a 4-layer convolutional architecture within a transfer learning model.
- Reduced false positives through open-set recognition, improving real-world model reliability.

Machine Unlearning - Data Privacy |CCC Intelligent Solutions | *2022*

- Built a privacy-aware ML system using SISA machine unlearning, enabling targeted retraining instead of full model retrains and reducing retraining cost by ~70–90% under high deletion volumes while maintaining ~95.6% accuracy.
- Analyzed accuracy–latency tradeoffs across shard counts and unlearning request rates (1–5% of data), uncovering non-linear (U-shaped) performance behavior and informing scalable, production-grade model update strategies under strict privacy constraints.

Netflix Data Analysis and Recomendation Using R| Illinois Institute of Technology | *2021*

- Executed an end-to-end analytics system in **R** to analyze 6K+ titles, applying NLP techniques (corpus construction, tf-idf) to extract features for content similarity.
- Performed data cleaning and normalization across countries, genre, categories, cast, and directors, and built visual insights on catalog growth, geography, duration, and rating mix for movies vs TV shows.
- Note: This project demonstrates the combination of text/NLP and structured data analysis, relevant to Multimodal concepts.

Oral Cancer Detection using CNN | Visteon Corporation | 2020

- Developed a high-performance computer vision pipeline utilizing VGG-16 and U-Net architectures for automated melanoma detection, achieving 94% classification accuracy through specialized transfer learning and image segmentation techniques.
- Engineered the system using Flask and Python, integrating advanced image augmentation and real-time preprocessing (RGB-to-grayscale, thresholding) to optimize model generalization across diverse medical datasets.

EDUCATION

Illinois Institute of Technology | Master's in Data Science | Jan 2021 – Dec 2022