Pragati Khekale

Dallas, TX, USA | +1 (312) 536-8480 | khekalepragati01@gmail.com

LinkedIn | GitHub |Tableau

SUMMARY

Data Scientist with expertise in building predictive models, deploying analytical solutions, and optimizing ML pipelines across cloud and enterprise environments. Skilled in Python, ML frameworks, and geospatial analytics with a strong ability to align technical solutions with business strategy. Experienced in developing deep learning models, enhancing model performance, and supporting cross-functional engineering and product teams.

SKILLS

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

Machine Learning: Regression, Classification, CNN, XGBoost, SVM, KNN, Transfer Learning, Open Set Recognition

Libraries &Tools: Tensorflow, Keras, PyTorch, Scikit-Learn, GeoPandas,

: Data Cleaning, Feature Engineering, ETL, GeoPandas.

Visualization: Tableau, Power BI, Matplotlib, Seaborn

Cloud & Infrastructure: Snowflake, BigQuery

Soft Skills: Stakeholder Collaboration, Cross-Functional Communication, Project Management, Strategic Thinking

EXPERIENCE

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

- Developed predictive models for unlocatable data points using ArcGIS Pro, XG Boost, KNN Regression, and deep learning, Saving more than $100k (estimate) cost and reducing operational inefficiencies.
- Designed and implemented KPI for Pipeline Integrity projects and created and automated various ETL pipelines using Python and power automate, integrating datasets across business units to support planning and strategy initiatives.
- Partnered with engineering, compliance and operations teams to align data workflows with long-term business goals.
- Presented executive annual reports on capital and operating data, influencing roadmap planning and corporate strategy across enterprise.

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

• Optimized ML models across six use cases analyzing 250K+ data points, achieving 96% model accuracy for operational forecasts.

• Collaborated with engineering and data privacy teams to develop analytical frameworks supporting business operations and customer-facing solutions.

• Produced performance metrics and reports to aid senior leadership in product strategy and process improvement.

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

- Integrated PoseNet into ADAS systems for real-time object detection, improving efficiency by 40%.
  • Contributed to the full ML model lifecycle, from design to deployment, supporting cross-functional collaboration in engineering and product teams.

PROJECTS

KPI Dashboards | 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.

• Built 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 open-set recognition using OpenMax in TensorFlow, reducing false positives by 12%.

• Conducted comparative analyses of classification models to optimize accuracy and business reliability.

Machine Unlearning|CCC Intelligent Solutions *| 2022*

- Implemented machine unlearning within SISA architecture for privacy- preserving model behaviour.
- Achieved 96% accuracy on model evaluations, conducted comparative analysis of shared-level performance to assess robustness.

Netflix Data Analysis and Recomendation Using R| Visteon Corporation *| 2020*

- Executed an end-to-end Netflix analytics and recommendation system in R to analyze 6K+ titles enriched with IMDb ratings and genre.
- 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.
- Applied NLP techniques such as corpus construction, tokenization, and tf-idf on descriptions and genres to extract high-signal featuresfor content similarity and recommendation use cases
- **Oral Cancer Detection using CNN | Visteon Corporation** | 2020
- Developed and trained a VGG-16 based model on 6,500+ X-ray images achieving 91% accuracy.
- Applied transfer learning to improve training efficiency by 40%, demonstrating scalable model deployment.

EDUCATION

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

CERTIFICATIONS

- AWS Machine Learning Specialty (In Progress)
- Google Data Analytics Professional Certificate