Vikas Ravikumar Karjigi

Chicago, IL | (312) 273-8331 | vkarjigi@hawk.illinoistech.edu | LinkedIn | GitHub

SUMMARY

Data Scientist and AI/ML Engineer with 3+ years of experience delivering ML-driven solutions and real-time data systems. Skilled in Python, SQL, TensorFlow, and cloud tools, with a proven ability to transform complex data into insights and deploy scalable ML models.

EDUCATION

Illinois Institute of Technology, Chicago, IL Aug 2024 – Present
Master of Applied Science in Data Science GPA: 4.0/4.0
Core Courses: Big Data Technologies, Data Preparation and Analysis, Regression, Machine Learning, Introduction to Time Series, Monte Carlo Methods in Finance, Project Management

RV College of Engineering, Bengaluru, India Aug 2018 – Aug 2022
Bachelor of Engineering in Electronics and Communication

TECHNICAL SKILLS

- **Programming & Databases:** Python, SQL, R, Bash, Pandas, NumPy, PySpark, Git, PostgreSQL, MySQL
- **Machine Learning & AI:** Deep Learning, CNNs, RNNs, NLP, Time Series Forecasting, Reinforcement Learning, Model Evaluation & Tuning, Scikit-learn, TensorFlow, PyTorch, OpenCV, Hugging Face, LangChain
- **Cloud & MLOps:** AWS (S3, Glue, Redshift), Azure (ADF, Databricks), Airflow, Docker, Kafka, CI/CD, ETL, Stream & Batch Processing
- **Data Visualization & Analytics:** Tableau, Power BI, Matplotlib, Seaborn, Streamlit, KPI Dashboards, A/B Testing, Statistical Modeling

EXPERIENCE

Illinois Institute of Technology Jun 2025 – Present
AI/ML Research Assistant Chicago, USA
- Developed a Reinforcement Learning pipeline using **PPO** in **OpenAI Gymnasium (Ant-v4)** with **MuJoCo** physics simulation, applying real-time training and policy optimization with **Stable-Baselines3**.
- Integrated trained agents into a custom **Unity 3D** environment with **Meta/Oculus VR** using **socket programming**, enabling immersive and interactive visualization of intelligent RL behaviors.

Illinois Institute of Technology Jun 2025 – Present
Teaching Assistant - Data Preparation and Analysis Chicago, USA
- Guided graduate students in data wrangling, feature engineering, dimensionality reduction, and model evaluation using **Python, R, Pandas, and SQL**, while explaining concepts and reinforcing best practices across the data science workflow.

Open Avenues Foundation Sep 2024 – May 2025
Data Scientist - Build Student Consultant Chicago, USA
- Developed a computer vision pipeline to detect and classify highlight-worthy moments from **120+ minutes** of volleyball footage using **OpenCV**, **XGBoost**, and motion-based features, achieving an **88% F1-score** on test clips.
- Automated generation of **1-2 minute** highlight reels, reducing manual editing time by **80%** and delivering final outputs in **under 3 minutes**.

Boeing India Private Limited Aug 2022 – Jul 2024
Data Analyst Bengaluru, India
- Improved booking efficiency by **30%** by building SQL pipelines across **MySQL** and **MSSQL** using complex joins and aggregations for real-time **KPI** dashboards — earned cross-functional recognition for business impact.
- Achieved **$60K+ in annual savings** by reducing reporting latency by **98%** (from 20 hours to 20 minutes) through stored procedure tuning, CTEs, indexing strategies, and parallel execution in MSSQL**.**
- Standardized metrics and improved **data access** by creating SQL templates across teams, earning recognition for cross-functional leadership.
- Resolved **500+** data security issues by automating validation checks with **Python**, ensuring compliance and data integrity.

Exposys Data Labs Jul 2019 – Sep 2019
Data Analyst Intern Remote, India
- Built a Python-based ETL pipeline using Pandas and MySQL to clean, normalize, and process **25K+** operational records, reducing manual prep by **60%** and enabling faster feature extraction and exploratory analysis for ML experimentation.

PROJECTS

GenAI Chatbot for Document Intelligence | LangChain, Hugging Face, FAISS, RAG, AWS, Docker
- Built a **RAG-based chatbot** using LangChain and Hugging Face Transformers to query **10K+ PDF** documents via FAISS semantic retrieval, reducing manual research time by **~70%** for analysts and domain experts.
- Deployed a Streamlit interface on **AWS (S3, Lambda)**, containerized with **Docker**, and designed for scalable serverless deployment, **CI/CD** readiness, and real-time LLM-powered Q&A on user-uploaded documents.

Real-Time Fraud Detection System for Financial Transactions | Kafka, PySpark, LightGBM, Azure, Streaming Data, Power BI
- Engineered a real-time fraud detection pipeline using Kafka, PySpark, and **LightGBM** on **Azure Databricks** to stream and score **1M+ daily transactions** with temporal risk pattern detection.
- Reduced false positives by **40%** and enabled automated risk monitoring via **Power BI** dashboard, accelerating fraud detection efficiency.

Multimodal Deep Learning for Disaster Damage Assessment | PyTorch, CNN, BERT, Hugging Face, FastAPI, AWS, Docker, Tableau
- Designed a multimodal deep learning pipeline using **CNN** and **BERT** to assess disaster damage from satellite images and field reports, achieving **88% accuracy** in classifying damage severity.
- Deployed the model via **FastAPI** on **AWS**, containerized with **Docker**, and visualized predictions in Tableau, boosting field decisions by **~60%**.