KAUSHIK KUMAR

Tucson, USA +1-520-609-4957 kaushikkumar.208@gmail.com Linkedin

SUMMARY

Data Scientist with proven expertise in developing and deploying **machine learning models** and **data pipelines**. Achievements include reducing testing time by over 70%, saving over **$72000 a month**, and building real-time monitoring systems using **PowerBI** dashboards. Proficient in **Python, SQL, Statistical Techniques, and MLOps** with hands-on experience in scalable architectures and **A/B testing**.

EDUCATION

UNIVERSITY OF ARIZONA **Aug 2024 - Present**
Master of Science, Data Science (GPA: 4.0/4.0)
- **Coursework:** Data Analytics & Visualization, Data Mining, Neural Networks, Advanced Machine Learning, SQL/NoSQL, Data Ethics

VELS INSTITUTE OF SCIENCE AND TECHNOLOGY **Jun 2019 - May 2023**
Bachelor of Engineering, Computer Science (GPA: 3.65/4.0)

EXPERIENCE

SELECTOR AI | *Software Engineering Intern* **Jan 2022 - Mar 2022**
- Designed and deployed **log mining pipelines** (ingestion, normalization, event inference, dashboards) using **Promtail, Loki**, and **rsyslog**, and brought up the production environment (**S2AP**).
- Replaced legacy Random Forest classification with **BM25S-based retrieval**, achieving **<2s indexing for 100k+ logs**; validated through **shadow runs on multiple clients**, improving both scalability and interpretability of event detection..

UNIVERSITY OF ARIZONA | *Graduate Researcher & Summer Tutor (**SALT**)* **Jan 2025 - Present**
- Co-developed a **novel safety-augmented RL** framework integrating **ShieldNN** with **LTL**, **DeepDFA**, and probabilistic shielding, achieving robust constraint adherence in **OpenAI Gym** environments.
- Enhanced **Supervised PCA** for high-dimensional tasks; outperformed **PCA, LDA, kernel** methods on biomedical and vision datasets.
- **Tutoring** undergrads for their **Math, English**, and **Computer Science** courses.

JOHNSON ELECTRIC | *Data Scientist* **Jun 2023 - Jun 2024**
- Reduced EOL testing time by 70%, saving $72K+/month by building a complete **MLOps** pipeline for a **hybrid AI** model (Kalman Filters and regression-based leakage prediction) for **Tesla's** air pumps, built **SQL pipelines, dockerized deployment,** and **Power BI** dashboards for anomaly detection and real-time analytics.
- Engineered ensemble classification models (**Gradient Boosting, Random Forest**) to detect faulty actuators with 93% accuracy on **timeseries data**, reducing cycle time by 83%, and **PyQt5 GUI** along with **Power BI** for interactive failure analysis dashboards.
- Built a real-time production monitoring system using **Power BI, Dask, and PySpark**, deploying parallelized **ETL** and **JSD/KL Divergence**-based unsupervised anomaly detection for zero false-positive predictive maintenance and **sub-step root cause analysis**.
- Applied advanced **statistical modeling** and **skew-normal curve fitting** to reduce industrial water pump (**Ford Motors**) test time by 70%, achieving F1 score of 0.93 and zero false positives, using **SVR, extrapolation**, and synthetic **time-series morphing**.
- A PLM Teamcenter chatbot using **Azure OpenAI, Chroma DB, Huggingface models, Marqo AI,** and **Solr**, integrated with SharePoint and **Graph API** to serve **domain-specific engineering data** with real-time feedback and **CRM** insights.
- An NG Engineering chatbot utilizing **LangChain, FAISS,** and **OpenAI APIs** for scalable document search and retrieval, featuring web embedding pipelines and auto-retraining triggers for robust production support. Integrated into Microsoft teams to enable access control

HEWLETT PACKARD & NATIONAL UNIVERSITY OF SINGAPORE | *Data Science Researcher* **Dec 2022 - Jan 2023**
- Designed and developed a language-invariant hate speech and gender bias detection model by combining an **n-gram** approach for language identification with a multimodal transformer (**MBERT**) for feature extraction and classification, trained on **GPU clusters**.
- Led as a **Team Leader** to deploy the model on **Azure VM** instance, implementing a **Flask-based RESTful API** backend and a **React** frontend to ensure responsiveness and efficient data processing. Finally deployed as a browser extension.

RESEARCH PROJECTS

ONLINE AUCTION MANAGEMENT SYSTEM USING ETHEREUM BLOCKCHAIN
- Developed a decentralized auction system using **Ethereum**, **Solidity**, **Truffle**, and **MetaMask**, enabling secure, transparent bidding via smart contracts using **Solidity**. Integrated **encryption** and **NFC-based data storage** for secure access
- Published "A Decentralized Online Marketplace Using Ethereum Blockchain" at NCRTCI'23 Conference

CROP SURVEILLANCE SYSTEM USING ANN AND DEEP LEARNING
- Designed and trained a deep convolutional neural network (**CNN**) using **PyTorch** and **TorchVision** to classify crop leaf diseases with over 93% accuracy, implementing key modules such as **segmentation**, feature extraction, and curative recommendation..
- Built robust image pipeline with **preprocessing**, **GLCM** features, and **SVM** classification for disease detection and curative suggestions.

TECHNICAL SKILLS

- **Data Mgt & Visualization**: PowerBI, Tableau, MongoDB, Airflow, SQL, NoSQL, Databricks, ETL, Excel, Azure Services (DevOps, Data-Studio, Data-Lake, Machine Learning, Blob, Synapse, Timeseries, VM)
- **ML Programming And Development**: Python, TensorFlow, Scikit-learn, NLTK, Spark, Pytorch, OpenCV, Keras, LLM, VLM, Deep Learning, NN Pruning, RLHF, Timeseries Forecasting, PySpark, Hyperparameter Tuning, Torch Vision, Pandas, NumPy
- **Analytical And Statistical Techniques**: Statistical Models, Data Mining, Predictive Modelling, Pattern Recognition, Kalman Filters, Marqo AI, Solr, RAG tools, Synthetic Data Generation, GAN, A/B Testing