HARIKA KANAKAM

📍 San Jose, CA ✉ harikakanakam98@gmail.com 📞+1(812)778-4329 in LinkedIn 🔲 Github 🪪 Portfolio

EDUCATION

Indiana University   Bloomington, IN, USA
Masters in Data Science | GPA: 3.83/4   Aug 2022 - May 2024
Relevant Courses: Machine Learning (Supervised, Unsupervised, Neural Networks), Applied Algorithms, Cloud Computing.

Madanapalle Institute of Technology and Science   Madanapalle, India
Bachelors in Computer Science | GPA: 9.16/10   Jul 2015 – Apr 2019

WORK EXPERIENCE

Knowledge Lens: A Rockwell Automation Company   Bangalore, India
Senior Big Data Engineer   Apr 2021 – Jul 2022
- Collaborated with **cross-functional teams** and business stake holders to analyze risks and provide unique optimized strategies.
- Performed quantitative research to drive continuous improvement of product, resulting in a **10% reduction in production costs**.
- Monitored user sessions using **AWS** Cognito, CloudWatch Logs, Redis cache datastore delivering **40% spike in productivity**.
- Revamped **REST APIs** by mentoring the team through code reviews resulting in **20%** faster delivery and **35%** bug reduction.
- Deployed network restricted applications using **AWS Kubernetes** pods to enhance network security and technical solutions.

Junior Data Scientist   May 2019 - Mar 2021
- Performed hands-on data analysis and modeling with large data sets, including discovering data from sources(local, s3, azure), getting data access, cleaning up data, and making them **"model-ready"**.
- **Automated the job monitoring** process and executed data migrations seamlessly, transferring data from MySQL to AWS S3.
- Utilized the capability of ML and data mining for **predictive analysis**, to forecast deployment status with 95% accuracy.
- Designed migration flow for **microservices** platform hosted on AWS using FastAPI, conducted analytics on stakeholder feedback.
- Collaborated with DevOps team to orchestrate Linux services and **GIT CI/CD** deployments, **reducing execution time by 10%**.

Junior Data Science Intern   Jan 2019 - Apr 2019
- Optimized job success rates by applying **problem-solving** skills to existing ML models, elevating job success from **82% to 93%**.
- Proficiently engineered backend services to execute GIT operations and validate **Docker** deployment files using Python **Flask**.

Tech stack: PyCharm, AWS(EKS, Sagemaker, Athena, RDS, EMR, Secrets Manger, Cloud watch, Cognito, Load Balancer), Tensorflow,Keras, MySQL, Docker, Kubernetes, Fast API, Flask, Swagger, Redis, GIT, Jenkins, Bit bucket, ngrok, SSO, JWT, PostgreSQL, Mongo DB, Spark, Hadoop, Hive.

ACADEMIC EXPERIENCE

Indiana University, *Research Assistant*   Summer 2023
- Conducted **data extraction** and transformation by parsing XML and SAS files to convert them into datasets to train ML models.
- Evaluated and analyzed **time horizons of companies**, helping investors with valuable insights to identify potential risk factors.

ACADEMIC PROJECTS

The Vision of Machine Unlearning
- Implemented machine unlearning algorithms to remove the effect of unwanted samples from the trained ML model.
- Implemented teacher-student (**knowledge distillation**) and zero-shot architecture for class-based unlearning by defining a custom loss function as an extension for the existing **KL-Divergence loss**, thereby outperforming the base implementation by 7%.

User Interactive Image Segmentation
- Developed a user interactive platform to extract the objects with/without the background from the user uploaded images.
- Utilized **Mask R-CNN** for image classification and segmentation, achieving a 95% extraction rate for the classes known by model.

Optical Music Recognition Using Template Matching
- Developed an automated system for detecting music note symbols and analyzing their pitch in digital music sheets.
- Implemented robust pre-processing techniques, including **Gaussian blur** and staff line removal, significantly enhancing detection accuracy and making it accessible to musicians and educators for **error-free music notation** handling.

Home Credit Default Risk Prediction
- Developed **preprocessing pipelines** to handle missing values, categorical variables, and feature scaling, ensuring data quality and compatibility for modeling. Built and evaluated ML models to predict loan default likelihood with **92% accuracy**.

TECHNICAL SKILLS

Data Visualization : Excel, Tableau, PowerBI
Programming and Databases : Python, MySQL, PostgreSQL
Statistical Models : Hypothesis Test, Z-Test, T-Test, Chi-Squared Test, A/B Test
Cloud Services : Amazon Web Services(AWS)(S3, EC2, RDS, ECR, Sagemaker, Kubernetes)
Deep Learning : Neural Networks, NLP, Convolutional and Recurrent Neural Networks, DNN, Transformers, Word Embeddings
Libraries : PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face, Pandas, NumPy, OpenCV, matplotlib, seaborn, ggplot, PySpark