Nistha Kumar

+1-(831)-295-7710 • nkumar20@ucsc.edu • linkedin.com/in/nisthak • Github • Portfolio

EDUCATION

Master of Science (MS), Computer Science *Graduating in Dec 2024*
University of California, Santa Cruz, CA, US GPA: 4.00/4.00
Coursework: Machine Learning, Deep Learning, Statistical Data Analysis, Computer Graphics, Analysis of Algorithms, System Design

PG Diploma, Data Science *Mar 2020 – Apr 2021*
International Institute of Information Technology, Bengaluru, India GPA: 4.00/4.00
Coursework: Machine Learning, Deep Learning, Computer Vision

Bachelor of Engineering, Computer Science and Engineering *Sept 2012 – Jun 2016*
Visvesvaraya Technological University, Belagavi, India
Coursework: Algorithms, Data Structures, DBMS, Computer Network, Operating System

SKILLS

Programming Languages: Python, R, SQL, NoSQL, C, Java, JavaScript, Bash
Tools & Frameworks: PyTorch, TensorFlow, NumPy, Pandas, Power BI, Tableau, Grafana, Redshift, S3, EC2, Docker, Kubernetes, Jupyter, Azure, AWS, Springboot, Kafka, REST, Git
Industry Knowledge: Machine Learning, Computer Vision, Deep Learning, Generative AI, LLM, Statistics, Data Warehousing, ETL, Data Analysis, CI/CD, System Design, Data Structures and Algorithms

PROFESSIONAL EXPERIENCE

Graduate Student Researcher, AIEA Lab - UCSC, California, US *Jan 2024 - Present*
- Led framework development that uses **few-shot prompting** to produce captions for **~10K matches** using **LLMs**. (Python)
- Leveraged traditional **Computer Vision** to create a spatial representation of Pacman's GIF frames for **prompt engineering**.

Data Scientist, Gravity iLabs, Bengaluru, India *Feb 2021 – Jul 2022*
- Led a team of **4 junior data scientists** and trained an **ANN** to predict project risks with an **F1-score of 0.73.** (Python)
- Employed **Active Learning** strategy to retrain the classifier upon adding a new risk class or project data points.
- Achieved **92% accuracy** in project completion status prediction by developing a **KNN classifier**. (Python)
- Designed and developed a scalable **ETL** data warehouse with a throughput of **2000 RPS** achieved via **AWS S3** and **Redshift**.
- Collaborated with **Data Engineers** to build a centralized dataset for **50+ dashboards,** boosting **operational efficiency** by **70%**.
- Deployed **stored procedures** on **Azure Data Factory,** reducing processing time within the **data lake** by **40%.** (SQL)
- Led a team of **BI developers** to build a **multi-tenancy model** for dashboards, reducing **TAT** by **80%** for new clients. (Power BI)

Data Scientist, Nokia, Bengaluru, India *Jun 2016 – Feb 2021*
- Collaborated with the analytics team to build an **LSTM** model for **VM rightsizing**, reducing costs by 17%.
- Led **fault prediction** for network elements, employing **LSTM** and **ARIMA** models to enhance reliability and reduce costs.
- Designed and developed a **Power BI** dashboard, which utilized **REST** and **JIRA APIs**, resulting in efficient decision-making.
- Simulated failures for **High Availability Testing** to enhance **robustness** & **fault tolerance** that reduced downtime by 40%.
- Accomplished a **90% reduction in manual effort** and saved 15+ hrs/week by automating dataset preparation using **Python, Postgres, and ETL** processes with scheduled cron jobs.

ACADEMIC PROJECTS

AI-Powered Pac-Man Commentary, UC Santa Cruz *Jan 2024 – Present*
- Engineered a novel approach to analyze Pac-Man gameplay using computer vision to generate spatial representation of game.
- Leveraged advanced LLMs and prompt engineering to generate informative captions for every 5 seconds of gameplay.
- Fine-tuned the model to produce high-quality, contextually relevant commentary, enhancing the overall viewer experience.

An analysis of PUBG gameplay statistics, UC Santa Cruz *Nov 2023 – Dec 2023*
- Built **Multiple Linear Regression** model, performed **Model selection** using **AIC & BIC** criterion to predict Winning percentile.
- Conducted **pair t-tests** to assess group significance and utilized **p-values** to determine the **significance** of predictors.

Image Classification using Transfer Learning, UC Santa Cruz *Nov 2023 – Dec 2023*
- Trained a **Swin Large Vision Transformers** on a custom dataset to classify images into 100 classes.
- Utilized **Hugging Face's PyTorch** image models library to build and train the transformer and achieved an **accuracy of 82%**.

Eye for the Blind, IIIT Bangalore *Jan 2021 – Mar 2021*
- Designed **Encoder-Decoder** neural network using **self-attention and GRU** to generate descriptive image captions, empowering visually impaired individuals to comprehend images; achieved **accessibility and inclusiveness** for 1000+ users.
- Deployed the model on **AWS** using **FLASK** with a TAT of 2 seconds.