Vivek Singh

LinkedIn +91 7355695707 viveksingh.com vivekkgp97@gmail.com

Profile

Data Scientist with 5 years of experience in Data Science, Machine Learning, Neural Network, MLOps, System Identification, Computer Vision, Smart Manufacturing & IIoT, LLM. Currently leading a Data Science Team in transforming business requirements into actionable data driven solutions. Worked closely with MIT's device realization lab team.

Skills

- **Programming:** Python, Matlab, SQL, C++
- **Machine Learning/Robotics:** Pytorch, Tensorflow, (KMeans, DBSCAN, GMM), Supervised (Decision Tree, SVM, Autoregresor), Neural Network, NLP, Reinforcement Learning, System Identification, Matlab & Simulink, Computer Vision (openCV), NLP, Gym
- **Gen AI:** LLM, Langchain, Huggingface, LLM-finetuning, RAG, LLAMA 2, Mistral 7B, Code T5 Plus,LLM-finetuning, RAG
- **Data Science & Miscellaneous Technologies:** Flask, HTML, CSS, Dash,Plotly, Seaborn, Matplotlib, OOP, Bagging, Boosting, Kernel, Clustering, Dimensionality Reduction (PCA, t-SNE, UMAP)
- **MLOps**: Git, Dockerization, Vertex AI, Artifact Registry, Auto ML, CI/CD Pipelines, Cloud Function, RESTful API, deploying models to production using GCP

Work Experience

Data Scientist (Deputy Manager) **STL Technology Solutions** Jul'19 – Present

MIT's Device Realization Lab, Prof. Brian Anthony (banthony@mit.edu) *Pune, Maharashtra*

Data Driven Modeling of system dynamics and control design

- Led the development and implementation of LSTM-based deep learning models for adaptive control systems.
- Played a key role in designing, training, and validating RNN and system identification models tailored for Digital Twins. This design tool resulted in annual savings of approximately $ 2 million.
- Collaborated with MIT to develop a proof of concept (POC) for real-time process correction in fiber drawing processes using Deep RL.
- The POC demonstrated superior performance compared to traditional PI and QDMC control methods when tested on a custom tabletop fiber drawing system, achieving state-of-the-art results.

Automated customer enquiries using LLM

- Created an automated question-answering system utilizing Retrieval Augmentation Generation (RAG) with Langchain and the Mistral 7B model.
- Utilized historical product catalog, manuals, and specification documents for training and refinement. Realized a substantial decrease in manual effort, estimated at 75% to 80%.
- Key Technologies used: Langchain, Mistral 7B, ChromaDB.

Automated quick reading of PLC logs using LLM

- Implemented Parameter Efficient Fine-Tuning (PEFT) techniques to fine-tune a Large Language Model (LLM) for anomaly detection in controller & hardware from PLC logs.
- It resulted in a substantial reduction in downtime with an accuracy score exceeding 80%.
- Utilized the Code T5 Plus framework for implementation, achieving superior performance compared to LLAMA 2 and FLAN T5 XL.

Data Driven Adaptive Control of Fibre Coating Process

- Implemented and integrated a system identification model into PLC as a lookup table and transfer function, resulting in a 25% reduction in coating scrap, equating to annual savings of approximately $1.5 million.
- Processed high volumes of IoT data, used noise reduction and tapering techniques that resulted in significant performance boost.

Defect Detection System using Computer Vision

- Designed and implemented a Computer Vision platform (Edge Detection) in Google Cloud to improve the Mfg. process quality and productivity.
- It involved deploying a python job on Google App Engine that can fetch the images from GCS Bucket and apply edge detection algorithms and save geometrical and spatial information of edges in the BigQuery.
- Saved approx 100 Man hours/month.

Deployed ML prediction models

- Deployed end to end Bend Insensitive Fiber Scrap Reduction ML model, starting from data collection from different data sources to deployment to Google Cloud using App Engine and managing data pipelines with CronJob. Impact was an annual savings of $1M.
- Deployed prediction models to GCP that reduces raw materials waste and the testing time of batch by 6 man hours/day

Optimization of glass making process

- Led the team to develop and deploy ML model as a REST API using Flask, Dockerization, and Google Cloud services that takes the machine generated file (IoT data) from cloud storage and predicts glass optical quality and saves the record to Bigquery, optimizing business processes.
- This tremendously reduced the manual testing effort,raw material wastes, saving 200 Man hours/month

Research Internships/ Projects

NeoAero, Controller Designer, *VTOL autonomous drones for last mile logistics* Apr'19 – Jun'19
Simulation and position control of VTOL drone *NeoAero, London, UK*

- Designed and implemented position control algorithms for a food delivery quadcopter using MATLAB, Simulink, and Simscape.

Taiwan Tech ,Research Scholar, *Distinguished Prof. Chyi-Yeu Jerry Lin* Apr'18 – Aug'18
Control and synchronization of motor driven 7 bar mechanism *Taiwan Tech, Taiwan*

- Implemented a control algorithm for a motor-driven mechanism mimicking a jogger's face on a treadmill.
- Established serial communication b/w camera sensor and faulhaber dc motor This involved taking feedback from camera sensor and motor to real-time synchronization through serial communication

SKALA: *Stair climbing mobile robot* Nov'16 – Apr'17
Vision based navigation *IIT Kharagpur*

- Developed real-time object tracking for an autonomous vision-based control system.

Roles & Responsibilities

- Key person from company to manage the collaboration with MIT, USA
- Led cross-functional collaboration to define project scope, business benefits, and expectations.
- Shared findings to management and stakeholders for model improvement and identifying of additional data/ measurement solutions
- Led, mentored, and monitored team performance to ensure process efficiency.

Education

Indian Institute of Technology Kharagpur, *West Bengal* 2015 - 2019
B.Tech, *CGPA 8.01/10, Mechanical Engineering*

Massachusetts Institute of Technology 2023 - present
Micromaster*, Statistics and Data Science*

Relevant Coursework – Machine Learning with Python, Statistical Modeling and Computation, Probability, Fundamental of Statistics, Deep Learning Specialization by Andrew NG, Introduction to Artificial Intelligence, Reinforcement Learning by David Silver, Statistical Method, Data Analysis and Algorithm, Advanced Image Processing, Digital Signal Processing, MLOps Fundamental, Docker Mastery, Optimal control, Robotics, Soft Computing, System and Control.