Tejas Khanna | Lead Data Scientist

shirleyhouse20@gmail.com • (91) 95127-67443 • https://www.linkedin.com/in/tejaskhanna/ • https://github.com/insomniac-tk

EDUCATION

Gandhinagar Institute of Technology | Ahmedabad, Gujarat

B.E. in Computer Engineering | CGPA: 8.35/10 **April 2019**

Relevant Courses: Artificial Intelligence, Data Structures, Analysis and Design of Algorithms, Big Data Analysis

TECHNICAL SKILLS

Programming Languages: Python (Advanced), Java (Intermediate), C (Beginner), C++ (Beginner)

Operating Systems: Windows XP/Vista/7/8/10, Ubuntu, MacOS

PROFESSIONAL EXPERIENCE

Lead Data Scientist | F(x) Data Labs **Oct '23 - Present**

- Served as the technical guide and mentor for **3 machine learning project teams in the problem domains of entity resolution, QA chatbots and stable diffusion based solutions.**
- Assam Government Guiness World Record Attempt
    - Served on the machine learning team of the official auditors of the GWR attempt by Assam govt. to create the largest online photo album.
    - Resnet50 model was used here to get the embeddings of the images.
    - My role here was **to create a pipeline for deduplication and invalidation of the candidate images which were numbered around 10 million**(**1.0 Cr**) and then present findings for the same. After all the checks, there were about 55,29,904 valid images. This was an official GWR.
    - **Technologies used**: python, celery, scikit-learn
- Pareit
    - Worked on improving and cutting the cost of a machine learning pipeline for medical document summarization. Worked on introducing LLMs into the existing system to supplant expensive machine learning services in AWS.
    - **Technologies used:** python, langchain, docker-compose, docker, aws s3, aws lambda, aws ec2, aws sts, aws rds, aws sqs, git

Senior Software Engineer | F(x) Data Labs **Dec '21 - Sep'23**

- Bitcoin.com, Games department
    - Developed REST-ful backends for game engines for one of the largest online crypto casinos.
    - Designed and deployed a suspicious player detection service with a production cluster setup.
    - Led a project related to an RTP service for BI reporting from ideation to deployment.
    - Spearheaded the investigations of any suspicious activities from players of the casino. Looked for suspicious patterns using data analytics libraries.
    - **Technologies used:** python, mysql, sqlalchemy, docker,docker-compose, kubernetes, kOps, FastAPI, scikit-learn, Pandas, Numpy, matplotlib
- Brightedge
    - Worked closely on customer reported issues related to the latency of the Brightedge API and worked on query optimization for the same.
    - Owned the new search engine additions feature for the platform and made significant changes in the process to make it more efficient and reduce lag between customer requests and delivery for new search engines to be added on the platform.

- **Technologies used**: python, mysql, beautifulsoup, grafana, fastapi, pyro, google reverse geo-coding apis

Software Engineer | F(x) Data Labs **Jan '20 - Nov '21**

- Sanskrit OCR
  - Built the ML backend using FastAPI for doing OCR of sanskrit manuscripts using a line segmentation model followed by an OCR model.
  - Major challenge was labeling for getting line segmentation masks, which was accomplished using Labelbox to manually curate the dataset.
  - **Technologies used:** python, ocroseg model(nvidia), calamari ocr, docker, docker-compose, FastAPI

- GMR, Market Clearing Price Prediction
  - Worked in a team of ML developers for building an ensemble of three predictive models for predicting the block wise electricity prices for the Day-Ahead-Market of the Indian Energy Exchange(IEX) for GMR group.
  - Developed robust Extract-Transform-Load(ETL), re-training and prediction pipelines using celery.
  - Successfully delivered the model which predicted the market clearing price with a 80% accuracy within 10% tolerance for the IEX market from 2020-2022.
  - **Technologies used**: python, celery, mysql, django, pytorch, tensorflow, scikit-learn, pandas, boto3, docker, docker-compose , aws ec2
- Riskfintech
  - Built a POC for an ETL pipeline for stress testing for the Bank Of England.
  - The entire project was hosted on Jupyter notebooks.
  - Stress testing is a process conducted by banks to assess how resilient they are to adverse economic conditions and financial shocks. It helps them evaluate their ability to withstand and manage potential risks.
  - Pandas was used extensively to model and simulate all the possible scenarios here.
  - **Technologies used**: Jupyter notebooks, pandas

- U.P. Government, Guinness World Record Attempt
  - Built the machine learning backend powered by celery to process, verify and perform deduplication for over 3 lakh videos of people saluting looking into the camera as they recite "vande mataram"
  - Successfully performed deduplication on more than a lakh videos and delivered value to the client by conclusively proving that the final deduplicated videos were all unique.
  - **Technologies used:** python, celery , opencv, docker, docker-compose

- Symphony, Fraud Detection Model POC
  - Successfully developed and implemented a fraud detection model for Symphony, addressing the challenge of service providers falsely claiming reimbursement for parts not changed.
  - Established a baseline model with 80% accuracy using manual labeling for fraudulent records, employing a Random Forest model.
  - **Technologies used**: python, pandas, scikit-learn

NOTABLE ACHIEVEMENTS

- As part of my Final Year Major project, I'd worked on an IOT based Carbon Monoxide Sensor using ESP8266 Microcontroller along with MQ-7 Gas Sensor. Used an Arduino IDE for uploading the code on the chip. Awarded the second best final year project in the computer engineering department.
- Hosted a webinar on "Machine Learning Basics" with 200+ student attendees.

CERTIFICATIONS

- Neural Networks and Deep Learning , Coursera
- Machine Learning, Stanford Online, Coursera
- Machine Learning With Python, IBM, Coursera
- Complete Python Bootcamp, Udemy
- Introduction to Deep Learning & Neural Networks with Keras, IBM, Coursera