

ANJALI SINGH

FULL STACK DATA SCIENTIST · ELECTRONICS & COMMUNICATION MAJOR · UDACITY DEEP LEARNING GRADUATE



"I am a Full stack Data Scientist with experience in developing deep learning algorithms & deploying end-to-end cloud infrastructure for ML pipelines. Lead a team of 8 people."

SKILLS

Programming Languages : Python, C
Data Science Frameworks : PyTorch, Scikit-learn, NumPy, Pandas, Keras, Tensorflow, Matplotlib, Seaborn etc.
Cloud computing : AKS, Azure Functions, AzureML, Azure cognitive, KEDA, AAD, Amazon Sagemaker, EC2, S3 etc.
Architecture design : REST API, Microservices, Asynchronous API, Event-driven Architecture, RabbitMQ, Autoscaling
ML-ops Technologies : Kubernetes, Docker, Github Actions (CI/CD), Azure Devops Pipelines
Database : MongoDB, Elastic Search, CosmosDB, NoSQL
IDEs/Distribution : VS Code, IntelliJ IDEA, Anaconda, Jupyter Notebook
Version Control : Git
Interests / Domains : NLP, Computer Vision, Mathematics, AI, Data Visualization
Others : Unit testing, hypothesis testing, Multithreading & Multiprocessing
Languages : English (Professional Proficiency), Hindi, Punjabi

EXPERIENCE

Precily AI. Sr. DATA SCIENTIST

NEW DELHI | APRIL 2021

Precily AI. DATA SCIENTIST

NEW DELHI | JAN 2020 – MARCH 2021

PROJECT: INFRASTRUCTURE AND DEPLOYMENT ON AZURE

- Managed cross-functional (DS & Engineering Team) integrations. Administered cloud resources, IAM roles, and cost optimization decisions for deployment.
- Devised **autoscaling** based on custom metrics to optimize time complexity of information retrieval process which brought down processing time from 82 to 20 minutes (**75.6%** decrease).
- Designed **microservices** for Information retrieval, Similarity, NLQ search & Automated training platform services as a **multi-tenant** architecture which cut down maintenance costs to roughly **50%** & increased scalability.
- Planned IP structure, **vnet**, **subnet**, **NSG** for services, application gateway & Ingress controller in **Azure CNI Networking** of AKS cluster.
- Implemented sync & async APIs using **Flask** & **Celery** frameworks. Applied **jwt** tokenization for the **authentication & authorization**.
- Built automated pipelines (**CI/CD**) for deployment of ML models using **Github actions**. Orchestrated container deployments using **Kubernetes**, **Docker**, **Helm**, **Container Registry**.
- Setup **Monitoring-logging** services & **real-time error messaging**. Identified error logs using **Kusto** query language during QA and testing.

PROJECT: TST PREDICTION, DATA CATALOG, AUTOMATED LEARNING PLATFORM

- Conducted **litigation analysis** on tax data with Grid-search hyper-opt on a set of ensemble & regression models with feature embeddings, achieved an accuracy of **97%** on test set.
- Compared similarity based on **distance** b/w **embeddings** (**BERT**, **Glove**), **Elastic search**, **Annoy** in terms of accuracy & inference time. Elastic search **BM25** performed the best with the accuracy of **73%**.
- Improved **Advanced Search** & **NLQ contextual search** relevance score using elastic search boosting on index of 57k - 4 million documents, queries tuned according to the metadata relevant to the application.
- Administered 4 codebases, reviewed pull requests, git issue tracking, and version control.
- Moderated all integration aspects of data science processes including user flow diagrams, API design, **database schema design**, etc. with the stakeholders. Contributed to a **successful POC** with the client.

PROJECT: SUMMARIZATION R&D

- Generated **extractive & abstractive** summaries with NLP techniques like **transformers**, pointer generator, etc. **ROUGE** score of **66** achieved with OpenNMT on AUSTLII dataset & **39** with PG network on a legal dataset.
- Created **PG network** from scratch based on paper "**Inducing Document Structure for Aspect-based Summarization -Amazon Research**" in **Pytorch** gaining improvement over the baseline model.
- Lead a team of five data scientists for NLP research under the consultation of IIITD professor, NLP LCS2 dept.
- Distributed computing of two deep learning models using **Horovod** and **DistributedDataParallel**.

Intel Indexer LLC. ML SCIENTIST INTERN

REMOTE | DEC 2018 – MARCH 2019

- Built **four asset prediction dashboards** for stock indexes, stock companies, and cryptocurrency including data harvesting, risk analysis in international finance.
- Improved time-series modeling approaches using MLP, RNN, and attention for stock price forecasting. Created data harvesters and scrapers for different websites and databases like www.sec.gov
- Applied a research paper on '**Uncertainty in banking**' by **Paul e. Soto** to make a **lexicon** of finance using NLP in PyTorch to build a sentiment model for financial reports.
- Collaborated with a team of four to meet deadlines on time which gained next term funding for the company.

PERSONAL PROJECTS

Sentiment analysis using SageMaker

MARCH 2019

Developed a web app which interacts with **RNN** performing sentiment analysis on movie reviews with an accuracy of **83%**. Deployed & hosted the web app using **Amazon SageMaker, Lambda & S3**.

Dog breed classifier using CNN

JAN 2019

Created a **CNN** with **Conv2d, batch-norm** layers which can identify an estimate of dog's breed. Compared the results with **OpenCV** filter & transfer learning **VGG16** with an accuracy of **73%** on 133 classes.

Autoencoders, GANs

JAN 2019

Trained a **DCGAN** discriminator & generator on CelebA dataset of 2 Million to generate human images.

ACHIEVEMENTS

Gained an early promotion to a full-time role. Rewarded with an annual bonus for exemplary performance.

Effectively communicated ideas and strategies in weekly product meetings with the management.

Contributed in team building by taking technical interviews for Data Scientists and MLOps position at my current company.

EDUCATION

UDACITY – Deep Learning NanoDegree

DECEMBER 2018 – APRIL 2019

APJ Abdul Kalam University – B. Tech Electronics & Communication Engineering

LUCKNOW | 2012–2016

CERTIFICATION

Coursera Introduction to MongoDB by MongoDB Inc.