Sounak Mukherjee

✉ m.sounak2000@gmail.com | in mSounak | ☎ +91-6290480560 | ○ github.com/mSounak | k kaggle/redhunter20

EXPERIENCE

Pratilipi ⬀ May 2022 - present

Data Scientist Bangalore

- Performed multiple experiments on **Collaborative Filtering** models and **Content Based** models to create user personalized recommendations which improve total reads by **50%** and also **2x** the subscription purchases.
- Identified issues with the legacy ML workflow deployment pipeline, and built a proof-of-concept to orchestrate the ML workflows using **Metaflow** after evaluating multiple workflow orchestration tools.
- Built and maintained the internal Machine Learning platform end to end in the organization using **Metaflow**, **AWS Batch**, **AWS EventBridge**, **AWS StepFunctions** and **Jenkins**.
- Migrated the ML workflows from legacy deployment pipeline to **Metaflow**, which led to a more reliable system and also helped reduce cost by **50%**.
- Instrumented ML workflows across the organization to generate metrics for **InfluxDB**, created dashboards and alerts in **Grafana** to improve observability and reliability.
- Written data ingestion scripts for AWS Personalise, and deployed new campaigns, which helped to achieve daily reads of **150k**.
- Performed hyper-parameter tuning and data cleaning for algorithms that power our recommendation systems improving the total number of reads by **20%** across the user base.
- Built a Bi-LSTM model from scratch to predict the gender of a person from their name and achieve an F1-score of **93%**.
- Implemented Infrastructure as Code, to automate AWS ECS services deployment pipeline for the organization, using **Terraform**.
- Built a system to collect data, and create embeddings for semantic search using **Sentence Transformers** and **TF-IDF**. Also created an API using **FastAPI** to serve it to the end user.

Bobble.ai ⬀ February 2022 - May 2022

AI Intern Remote

- Wrote data cleaning pipelines to process terabytes of conversational data.
- Performed **intent recognition** on **hindi macronics** corpus.
- Developed a model to find the sentiment of brands in the conversational texts.
- Trained **FastText** on large datasets to create sentence embedding and word embedding.
- Performed clustering on large conversational data.

PROJECTS

Targeted-sentiment-analysis ⬀ December 2021

Implemented Transformation Net for targeted sentiment analysis

- Built a custom model to get the polarity of the target in the sentence. Engineered a powerful backend API with FastAPI, containerized the solution using Docker, and seamlessly deployed it on a Kubernetes cluster for optimal performance and scalability.

Machine-Translation ⬀ November 2021

Trained a Transformer from scratch and used for machine translation

- Encoder-Decoder architecture to produce an input sequence representation with applied positional embedding to increase token position awareness in sentence. Decoder connects it all together and generates required translation. It translates English sentences to German have plans to expand its support for Indian language. Created the front-end in streamlit and deployed the application in EC2 instance.

TECHNICAL SKILLS

- **Languages:** Python, C/C++
- **Framework:** TensorFlow, Keras, PyTorch, Scikit-Learn, Matplotlib, Pandas, Numpy, Seaborn, Spacy
- **Others:** Linux, Git, SQL, AWS, Streamlit, Docker, FastAPI, Terraform, MetaFlow, MLFlow, Weight&Biases

EDUCATION

Kalinga Institute of Industrial Technology Bhubaneswar

B.Tech Computer Science and Engineering, **CGPA: 9.15** 2019–2023