Sushovan Saha

+91 7001505741 sushovansaha95@gmail.com LinkedIn Github Kaggle

Education

Indian Institute of Technology, Guwahati **2021 - 2023**
M.Tech in Data Science *8.66/10 CGPA*

Indian Institute of Engineering Science and Technology (IIEST), Shibpur **2015 - 2019**
B.Tech in Computer Science & Technology *8.01/10 CGPA*

Ramakrishna Mission Vidyapith, Purulia **2013 - 2015**
Board - WBCHSE *90 Percent*

Work Experience

TVS Motor Company **July 2023 – Current**
Software Engineer I *Bengaluru*

- Developing an image classification ML model for low-performance microcontrollers using Keras and PyTorch, applying quantization to reduce model size by **70%** with minimal impact on accuracy.

Projects

1. MTP : News Article based Question-Answering System

Thesis : *Dr. Ashish Anand, Dept. of CSE and Dr. Prithwijit Guha, Dept. of EEE, IIT Guwahati* *GitHub*

- Built a QA model by fine-tuning **FlanT5** on **NewsQA Dataset** generated from popular Indian news media sources.
- Made model to work with incremental knowledge base with the help of efficient context retrieval using **BM25, KNN, and SentenceTransformer.**
- Scores on different NLP metrics for generated answers on the test set : **METEOR 96.67%, BLEU 64.80% and RoGUE 96.58%**.

2. Multi Approach Similarity Based Article Bias Prediction using LSTM and BERT

Course Project : *Dr. Sanasam Ranbir Singh (Dept. of CSE, IIT Guwahati)* *GitHub*

- Determined the political bias of news articles using **vanilla LSTM, Co-Matching, Entailment, Multi-Head Attention, and BERT**.
- Did comparative study of their performance **CoMatching 43.38%, Entailment 74%, MHA 79%, BERT 72.8%**.

3. Brand Logo Recognition using Transfer Learning and App Deployment in GCP with MLOps

Self Project *GitHub*

- Classification of **27** brands' logos on **Flickr27 Dataset** using **Transfer Learning** with several models **Xception, MobileNetV2, EfficientNet, InceptionV3** using **Keras & Pytorch-Lightning**
- Made a Streamlit WebApp and Successfully Deployed in **GCP** following **MLOps** and CI/CD Pipeline

4. Finetune Gemma2B Instruct for QA Task using RAG

Self Project *GitHub*

- Finetuned **Gemma2B** LLM for Question Answering task
- Results are shown for No Finetune with prompt, with Finetuned, and using **RAG Framework**

5. Rider Driven Cancellation Prediction

Self Project *GitHub*

- Built ML models to predict rider-driven cancellations using real-world order data, with appropriate preprocessing and feature engineering
- Handled highly class imbalance with **SMOTE** and **Stratified KFold CrossValidation**.
- Improved the accuracy from **52%** to **63.5%** using **Stacking of Classifiers**.

Technical Skills & Courses

Languages:C++, Python, SQL
ML Frameworks:NumPy, Pandas, Scikit-learn, PyTorch, PyTorch-Lightning, Transformers, LLM, MLOps, MLflow, Github Actions
Developer Tools:VS Code, GitHub
Courses:Statistics, Machine Learning, Deep Learning, Data Structures and Algorithms, DBMS

Achievements

- **Kaggle Expert** : Kaggle Notebook Expert
- **TVS PGET Hackathon** : Secured 2nd Position among 5 teams

Extracurricular

- **Teaching Assistant** : Machine Learning Lab and Python Laboratory (CS-594), IIT Guwahati