Divyesh Pratap Singh

+1 (716) 573 1947 | dsingh27@buffalo.edu | github.com/Divyeshpratap | linkedin.com/in/divyesh-pratap-singh | My Personal Website

December 2024 Masters in Artificial Intelligence graduate seeking Full Time roles in Machine Learning and Data Science.

Profile

- **Machine Learning**, and Software engineer with **5 years of experience** solving business problems in automotive and credit card industry.
- Working as **Graduate Research Assistant** in UB Deep Learning Lab on multimodal Active Speaker Localization, & Child Speech modelling.
- Awarded **Best Graduate Engineering Trainee** in Suzuki Motors, and **Best Performing Decision Scientist** in EXL Banking vertical.

Projects

Automobile Inspector: AI-Powered Car Damage Detection and Chatbot Platform GitHub ***June 2024 - August 2024***

- Deployed an AI driven web application that integrates computer vision and NLP to provide comprehensive car repair solutions, from visual damage assessment to a **Retrieval-Augmented Generation (RAG)** based conversational chatbot.
- Engineered segmentation pipeline uses **Mask R-CNN** with **ResNet-101** backbone, and **Deformable Convolution Networks (DCN)** to accurately localize vehicle damages into dents, scratches, broken lamps, glass shatter, flat tire, and cracks.
- Designed **carBot**, a context aware **chatbot** leveraging **chat history** to enhance prompt accuracy using **Ollama**. Documents are stored in Facebook's **FAISS vector store** for efficient retrieval and processed via Llama for context-specific responses.
- Architected a user-centric application using **Flask**, enabling damage detection, cost estimation & repair requests in just **5 clicks**.

Experience

The Research Foundation SUNY, University at Buffalo, *Research Assistant| NY, USA)* GitHub ***April 2024 - Present***

- **ChildSpeak**: **LLM** Powered Early Speech Pattern Analysis for Child Language Development

- Designed a robust causal modeling pipeline to identify a toddler as Late Talker or Typically Developing using parent child audio conversations.
- Implemented three key components: Audio **transcription** and **diarization** using OpenAI Whisper and Pyannote, identification of **novel part-of-speech** tags, and **causal modeling** to analyze word usage representative of late talkers.
- Annotated a large dataset comprising **2.5 million words** and a **40K vocabulary** across new categories: Shape/Non-Shape nouns and Result/Manner verbs. Leveraged **GPT 4o-mini** and **LLama 3.1** 405B models to **annotate** the dataset with a carefully crafted **prompt**.
- **Finetuned Roberta** base, achieving **94% accuracy** in **classifying Shape/ Non-shape nouns** and **97%** on **Result/ Manner Verbs**.

CUBS (Center for Unified Biometrics and Sensors), UB, *Graduate Assistant| NY, USA)* ***July 2024 - Present***

- **AVATAR**: **Audio Visual Active Tracking and Annotation Rendering**

- Designed AVATAR, a near **Real-Time Multi Modal Active Speaker Localization** and Transcription Framework using a **Producer Consumer Multi-Threading** Architecture, achieving a **latency** of **0.5 seconds**.
- Producer thread performs Face detection and Tracking on live camera feed using **MTCNN** and **SORT**. **Lip Sync** scores evaluation is using either **TalkNet** or Syncnet.
- Consumer thread enables near real-time **visualization** of **active speakers**, displaying bounding boxes, track number and confidence scores.

EXL Analytics, *Data Science Lead Assistant Manager | Hybrid (Gurugram, India)* ***November 2021 - June 2023***

- **Continuous Data Integrity Tool** (Product Development, Data Streamlining, Quality Improvement at Company Wide Scale).

- Implemented robust **data quality controls** to bolster integrity of all Credit & Fraud risk decision science models in American Express. This automates extract & transform (**ETL**) pipeline to safeguard against **data discrepancies** at an early stage.
- Engineered via a scalable **anomaly** detection framework (ADF) that **ensembles four time series** algorithms, optimized for high-volume data.
- Leveraged **Hadoop** for optimal data management and **Hive/ PySpark** for querying, enabling the big data pipeline to process millions of rows across thousands of features within minutes, ensuring real-time data fidelity..
- Attained an impressive **81% mean outlier alert accuracy** on 240 plus datasets, **reducing manual verification** efforts by **30%**.

Suzuki Motors India Limited, *Analytics Deputy Manager | Onsite (Gurugram, India)* ***July 2017 - November 2021***

- **Inventory Management & SKU Demand Forecasting** (Process Optimization, Traceability, Operations & Project Management).

- Crafted a forecasting model by integrating ABC and XYZ analysis to provide **live insights on SKU availability, safety stock, & reorder points** for over 4,000 parts. Integrated **SARIMA** with **LSTM** framework to predict delivery times of critical spare parts, enhancing the Order to Delivery Process Analysis and Inventory Optimization.
- Increased **inventory turnover ratio** from **3.2 to 4.7** leading to annual cost savings of **$7 million** by reducing wastage.

Education

3.75/4.0 **MS in Artificial Intelligence**, *University at Buffalo, The State University of New York* | NY, USA ***August 2023-December 2024***

8.23/10 **B.Tech in Electrical Engineering**, *Thapar University* | India ***July 2013-June 2017***

Courses: Machine Learning, Pattern Recognition, Computer Vision & Image Processing, Fundamental of A.I., Reinforcement Learning, , Numerical Math, Advanced Algorithm Analysis & Data Structures, Data Intensive Computing, Computational Linguistics, Information Retrieval, Linear Algebra

Skills

Programming: **Python**, HIVE, PySpark, **SQL**, Java, Shell scripting , HTML, LaTeX, Apache Solr.

Machine Learning: **Transformers**, Parameter Efficient Training, LORA, **Predictive Analytic**, **Feature Engineering**, Clustering, Advanced Statistics.

Framework/ Library: **PyTorch**, TensorFlow, **Hugging Face**, **Hadoop**, **sklearn**, **pandas**, statsmodel, NumPy, OpenMMLab, CUDA, SpaCy, NLTK.

Information Tech: **Linux**, Cloud Computing, Distributed Computing, **Big Data**, OpenShift, GIT, Postman, GCP, **Agile**, **JIRA**, Kanban.