


Divyesh Pratap Singh

+1 (716) 573 1947 | dsingh27@buffalo.edu | github.com/Divyeshpratap | linkedin.com/in/divyesh-pratap-singh
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 Analyst** in EXL Banking and Finance 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 **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 casual modelling 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**.

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 automated extract & transform (**ETL**) pipeline to safeguard against **data discrepancies** at an early stage.
- Engineered via a scalable anomaly detection framework (ADF) that **ensembles four ML algorithms**. Optimized for high-volume data, big data pipeline processes over million 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%**.
- **Tech stack: Time Series Forecasting**, EWMA, Holt-Winter double exponential smoothing, ACF, PACF, Augmented Dickey Fuller test, **CI/ CD**.
- **Risk Strategy and Remediation Tool** (Product Management, Risk Mitigation & Compliance).
- Developed a data-driven strategy to adjust credit lines based on real-time transactional behaviors. Directed a team of three towards development of backend systems on JAVA, and Business Logic Model on Python.
- I oversaw the optimization of database meta data and interactions, utilizing **Hadoop** & **HBase** for optimal data management & accessibility, powered by **Hive** over **Tez** engine. Engineered model consistently outperformed **FICO scores** in accuracy, hitting significant positive lift and **GINI** coefficient improvements. Strategic advancement resulted in a **5% reduction in delinquent losses**.
- **Tech stack: PySpark**, Gradient Boosting, GINI, reject inferencing, Jenkins, Kubernetes, Spring MVC, JAVA.

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 | Advance Algorithm Analysis & Data Structures | Numerical Math | Data Intensive Computing | Deep Learning | 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.