Divyesh Pratap Singh

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

Profile

- Machine Learning Engineer with 6 years of full-time experience in automotive and credit card industry.
- Awarded Best Graduate Engineering Trainee in Suzuki Motors, and Best Performing Decision Scientist in EXL Banking.

Experience

SUNY Research Foundation: ***ML Research Project Assistant*** July 2024 – May 2025

AVATAR: Audio Visual Active Tracking Annotation (Advisor: Dr. Venu Govindaraju) *NY, USA*

- Achieved sub-0.75 second latency in near real-time active speaker localization and transcription system, by architecting a multi-threaded producer-consumer framework. Reduced annotation process manhours by 10 times.
- Improved lip-sync classification performance to 0.89 F1 score on noisy multi-speaker videos by training a 125M parameter speaker separation model using contrastive learning with multi-similarity loss on over 2 million videos from the Oxford VoxCeleb and Google AVSpeech datasets, trained with PyTorch Distributed Data Parallel (DDP) on 4×NVIDIA A6000.
- Delivered face tracking in live camera feeds by integrating InsightFace/ SORT for face recognition & Nvidia NeMo for audio biometrics in producer thread of the AV pipeline, transitioning from previously offline-only ASD models to real time setup.

NVson: Noun Verb Semantic Observation Network (Advisor: Dr. Ifeoma Nwogu)

- Enabled early language disorder screening by designing a hypothesis driven ML pipeline to classify toddlers as Late Talkers or Typically Developing using CHILDES parent-child audio conversations.
- Achieved 91% accuracy for Manner/ Result verb classification (vs 79% human baseline) by automating large scale linguistic annotation of MASC, VerbNet, and Intercorp datasets using LLM (ReAct agentic framework), and finetuned a RoBERTa clasifier on this synthetic dataset. Submitted a paper in EMNLP 2025 (A* conference).
- Identified Manner-Result verb ratio in child speech as statistically significant linguistic predictors through A/B hypothesis testing on outputs from the 2-stage NLP system integrating audio transcription and POS annotation with generative AI.

EXL Analytics (Client: American Express): ***Lead Decision Scientist*** Nov. 2021 – June 2023

CDIT: Continuous Data Integrity Tool *Gurugram, India*

- Delivered end-to-end platform to ensure data integrity across 270+ datasets with 81% mean anomaly alert accuracy by implementing a scalable anomaly detection framework (ADF) that flagged outliers based on dynamic upper and lower bounds using statistical trend analysis (Holt-Winters, IQR, MA) to monitor high-volume credit & fraud data.
- Reduced manual data verification efforts by 40% by collaborating with business stakeholders and data stewards to identify data governance issues and integrating ADF into existing ETL workflows.
- Enabled automated asynchronous batch processing for 13 million average daily customer transactions by leveraging Hadoop for distributed storage, Hive/PySpark for efficient SQL querying, and CI/CD pipeline deployment on OpenShift (K8s).

Maruti Suzuki India Limited: ***Machine Learning Engineer*** Aug 2017 – Nov 2021

Vehicle Alarm and Warning Audio Event Detector *Gurugram, India*

- Reduced false positives by 63% (from 0.19% to 0.07%) across 1.5M+ vehicles/year by building an ML audio event classification system that replaced FFT analyzers and manual inspection in production lines.
- Improved precision recall AUC from 0.948 to 0.983 by curating a 14k training dataset, and fine-tuning Google YAMNet.
- Facilitated real-time OK/NG decisions by deploying an INT8 quantised (TFLite + ONNX) model to an edge compute module (PLC) with 1.5 seconds end-to-end latency with Apache Kafka for audio signal transfer.

Projects

Automobile Inspector, University at Buffalo April 2024 – June 2024

- Automated vehicle visual damage assessment steps to under 5 clicks, by deploying a Flask-based web app that integrates computer vision for vehicle inspection and a RAG-based conversational agent for repair assistance, and order retrieval.
- Achieved high-accuracy localization across 6 damage categories with an mAP of 0.91, by engineering a segmentation pipeline using Mask R-CNN (ResNet-101) and Deformable Convolutional Networks (DCN).
- Incorporated response relevance in repair queries by building a context-aware bot (CarBot) using Ollama, FAISS vectorDB with iterative re-ranking retrieval, and LLaMA/ gemini for personalized history-driven answers.

Education

University at Buffalo, New York 3.8/ 4 GPA

Masters in Artificial Intelligence *Aug. 2023 – May 2025*

Thapar University, India 3.5/ 4 GPA

Bachelors in Electrical Engineering *Aug. 2013 – June 2017*

Technical Skills

Programming: Python, PyTorch, TensorFlow, HIVE, PySpark, SQL, Shell scripting, HTML, CUDA, JAVA, Latex, PowerBI.
Machine Learning: Transformers, Classification, Predictive Analytic, Clustering, Statistics, sklearn, SpaCy, pandas, VLLM.
Developer Tools: Agile, JIRA, REST API, MLFlow, Docker, Hadoop, GCP, AWS EC2, sagemaker, ECR, lambda, FastAPI.
Artificial Intelligence: Langchain, Agentic AI, Prompt Engineering, LoRA, Hugging Face, Retrieval Augmented Generation.