+1 (514) 560 8589
Montreal, Canada
preetichouhan025@gmail.com

Preeti Chouhan

preetichouhan.github.io
github.com/preetichouhan025
linkedin.com/in/preetichouhan025

Skills

Languages:	Python, Java, C++, SQL, MATLAB
Frameworks:	Tensorflow, PyTorch, PySpark, Flask, Django, REST, Numpy, Pandas, Scikit-learn, OpenCV, NLTK, PyLucene, Matplotlib, HuggingFace, SpeechBrain, LangChain
MLOps/DevOps:	AWS Sagemaker, GCP, Heroku, Kubernetes, Docker, MLflow
Databases & Data Tools:	PostgreSQL, MySQL, Oracle, NoSQL, Redis, Cassandra, MongoDB, Dask, ElasticSearch, VectorDB (Chroma),
Machine Learning Specs:	Neural Networks, CNN, RNN, NLP, LLM, RAG, RL, sequence models, computer vision, data analysis, recommender systems, classification, regression, clustering, GANs, Hypothesis testing, PCA, Analytical and Predictive Modeling
Tools,Protocols,Others:	Git, Tableau, JIRA, Agile and Scrum, Waterfall, SDLC, Unit Testing, Linux

Professional Experience

Data Scientist (R&D) - ML | DL | NLP **Aug 2021 — Jan 2023**
Xeeva Inc. *New Delhi, India*

- Led the development and deployment of an end-to-end large-scale multi-class classification model using BERT in a custom SageMaker environment within the Spend Analysis pipeline. Achieved a 90% Macro F1 Score, reducing manual review time by 99%.
- Developed & deployed an enrichment & anomaly detection system with Naive Bayes & XGBoost for robust data quality checks. Achieved a 50% reduction in PyLucene index size & significantly improved disk I/O efficiency by optimizing with FastPFOR & Zstandard algorithms.
- Enhanced the IR pipeline, improving search accuracy by 80%, reducing processing time by 40%, and boosting semantic extraction precision by 25%. Utilized entity extraction, dependency parsing, and clustering to optimize data organization and query relevance.
- Conducted large-scale A/B testing and developed attribution models to optimize customer acquisition, enhancing data-driven decision-making and performance metrics.
- Boosted data processing by 5x and querying speed by 3x with optimized batching and cloud storage, achieving 99% system uptime.
- Continuous Integration/Deployment Pipeline Integration, pull requests, code reviews, load/stress testing, unit/integration testing.
- Mentored junior team members and conducted ML workshops, enhancing team collaboration and technical expertise.

Software Engineer - Machine Learning | Deep Learning **Jun 2018 — Jul 2021**
Larsen and Toubro Infotech Limited *Mumbai, India*

- Prototyped and deployed a Transformer-based chatbot, achieving 80% precision in response accuracy and a 70% increase in user satisfaction through fine-tuning, transfer learning, and contextual embeddings.
- Classified terabytes of documents using Random Forest and XGBoost classifiers with multiprocessing to meet latency requirements, leveraging pixel intensity and assembly features for feature extraction, achieving an F1 score of 85%.
- Productionized a multi-GPU OCR (Optical Character Recognition) pipeline using Tesseract into a live API micro-service with Docker containers, implementing high-availability and automatic failover, reducing hard copy dependency by 95%.
- Achieved >90% precision and recall in product review analysis using DistilBERT, uncovering key quality issues and customer preferences to drive strategic decisions and product improvements.
- Implemented anomaly detection for quality inspection of manufactured labels using convolutional AutoEncoders, comparing reconstruction error (RE) and reconstruction accuracy (RA) with Kernel Density Estimation (KDE) based on latent space vectors.
- Identified, assessed, and mitigated modeling risks and concerns, while ensuring the implementation of regulatory compliance throughout the entire model lifecycle.

Education

Master of Applied Computer Science, Concordia University Sep 2022 - May 2024
Accomplishments: Dean's List for Academic Excellence, Golden Key Honour Society. GPA: 4.0
Relevant Coursework: Algorithm Design, Statistics, Machine Learning, Computer Vision, Applied AI, Conversational AI, Distributed System Design, Advanced Programming Practices, Programming and Problem Solving, Geometric Modeling, Federated Learning

Bachelor of Technology in Computer Science & Engineering, University of Petroleum and Energy Studies Jul 2014 - May 2018

Academic Projects

Text-To-Speech (Neural Speech Synthesis with Transformer Network) Git
Implemented a transformer-based text-to-speech system using SpeechBrain trained on LJSpeech dataset, leveraging multi-headed self-attention and WaveNet vocoder for natural speech generation.

Novel Image Generation via text prompts using VQGAN (Vector Quantized Generative Adversarial Networks) and Transformer Git
Implemented an image generation system using VQGAN and transformer architecture to create realistic images from text prompts, enhancing 3D modeling and scene analysis. Trained on the COCO dataset for diverse and accurate outputs.