Prashant Gaurav

prashant.grv07@gmail.com | +1 (623) 212-6588 | www.linkedin.com/in/prashant-gaurav-01

Education

Arizona State University — **August 2023 - May 2025**
Masters of Science in Computer Science — *3.61/4.0 Tempe, Arizona*
Coursework: Data Mining, Data Visualization, Statistical Machine Learning, Data Processing at Scale.

University of Delhi, Netaji Subhas Institute of Technology (NSIT) — **August 2017 - July 2021**
Bachelor of Engineering in Computer Engineering — *3.21/4.0 New Delhi, India*
Coursework: Machine Learning, Data Structures, Computer Networks, Advanced Algorithms

Work Experience

Kaaj.ai — *San Francisco, CA*
AI/ML Intern, June 2024 - August 2024

- Developed an AI-powered financial document processing system using OCR and prompt engineering, reducing analysis time from several hours to 120 seconds per document while maintaining high accuracy and achieving a cost-efficiency of $0.15-$0.30 per document for balance sheets, invoices, tax documents, and financial statements.
- Conducted experiments with state-of-the-art LLMs including Llama 3.1 and GPT-4-Omni, optimizing performance through meticulous benchmarking and prompt refinement.
- Architected and deployed a scalable document processing service using Flask and Celery, creating specialized endpoints for various document types. Implemented a PostgreSQL database for efficient storage and retrieval of processed financial data, ensuring seamless integration with existing systems.

Yellow.ai — *Bangalore, India*
Software Development Engineer, Data Science and NLP, July 2021 - July 2023

- Achieved a ~2 million sentence embedding corpus through research and development, optimizing microservices to streamline chatbot development and enhance training outcomes using embedding vector search algorithms and cosine similarity techniques.
- Developed multiple data pipelines responsible for gathering, cleaning, and preprocessing organizational data, facilitating analytics and training for a proprietary LLM using Pandas, Databricks and Pyspark.
- Improved internal utterance clustering algorithms by experimenting with K-means, density-based, and hierarchical unsupervised learning algorithms, leading to enhanced performance in multiple microservices.
- Investigated the performance of the OpenAI Whisper model on multiple Indian languages using Word Error Rate, resulting in significant cost savings by eliminating third-party Speech-to-Text services.
- Developed an efficient OpenAPI with rate limiting capabilities to deploy the team's internal zero-shot model, resulting in its utilization by Microsoft for testing DynamicNLP performance.

Samsung R&D Institute India — *Bangalore, India*
Software Trainee, Deep Learning, May 2020 - July 2020

- Developed a parental control model using deep learning and data analysis techniques, achieving 84% accuracy in classification. Built a collaborative filtering-based recommendation system for personalized content recommendations.

Projects

Training Data Suggestion

- Explored unsupervised learning methods like density and hierarchical-based clustering to enhance the suggestion model, resulting in a 21% increase in silhouette score.
- Empowered users to create bots, contributing 80% to training data and significantly improving bot performance due to enhanced data quality.

Technical Skills

Languages: Python, C, C++, Java, JavaScript
Tools and Frameworks: Pytorch, TensorFlow, Flask, Pytest(Unit Testing), Streamlit, Scikit-learn, Numpy, Pandas
Others : Machine Learning, Artificial Intelligence, Natural Language Processing, Computer Vision, Generative AI, SaaS, Jupyter, DevOps, Git, Web Development, Databricks, CI/CD, ElasticSearch, MongoDb, Redis, MySQL, Jenkins, OOPs, Linux, Jira, Docker, Kubernetes, Excel, PyTest, RESTful, Kafka, Matplotlib, Matlab, Azure