

Aastha Tandon

✉ atandon@horizon.csueastbay.edu 📞 408-416-8709 🌐 <https://aasthatandon.github.io/>

Education

California State University East Bay

M.S. in Business Analytics (Machine Learning & Data Science) | **GPA: 3.97/4.0**

August 2022 – May 2024

Relevant Coursework: Text Mining, Deep Learning, Big Data, Data Mining, DBMS, BI, Data Analytics

University of Delhi, India

M.Sc in Operational Research

August 2015 – May 2017

B.Sc in Computer Science

August 2011 – May 2014

Technical Skills

Programming Languages: Python, R, SAS, SQL, Java, C++, LaTeX

Learning Algorithms: CNN, LSTM, Transformers, Large Language Models (LLM), Diffusion Models

ML Frameworks/Others: PyTorch, Tensorflow, Keras, Scikit-learn, NLTK, Azure, GCP, AWS, MongoDB, MapReduce, Hive, Pig, Hadoop, ApacheSpark

Experience

California State University East Bay

Graduate Student Researcher

June 2023 – Present

- **Graduate curriculum design with GPT:** Designed an end to end system that scrapes data from web for various job titles and utilize GPT model to structure data and perform topic modeling. Recommended 10% more skills compared to traditional topic modelling approaches like LDA and BERT.
- **Dialogue summarization employing LLM:** Applied the parameter efficient fine-tuning method LORA to refine the Flan T5 base LLM model for the dialogue summarization task using the DialogSum dataset. Employed ROUGE metrics for comparison against a fully fine-tuned model, demonstrating comparable results.
- **Mitigate toxicity in an LLM:** Reduced toxicity in output of a Flan-T5 base model by incorporating Meta AI's hate speech reward model. Employed Proximal Policy Optimization (PPO) for the fine-tuning.
- Coached 100+ students in Python and Excel, enhancing programming, data analysis & management skills in 6 classes in courses ITM 300 & ITM 330.

UnitedHealth Group

Associate Data Analyst

April 2019 – June 2021

- Constructed ML models to prevent customer churn in UHG healthcare plans, increasing customer retention by 15%.
- Formulated ensemble technique for early detection of recurrent breast cancer, improving accuracy by 20%.
- Designed multi-headed Neural Network to estimate joint occurrence of Type 2 Diabetes and Cardiovascular diseases based on patient's past medical history, resulting in a 19% increase in accuracy.

Complete HEOR Solutions (CHEORS)

Senior Data Analyst

June 2017 – March 2019

- Analyzed treatment patterns data, revealing opportunities for new drug launches and shaping strategic decision making
- Created data-driven recommendation engines to promote medication resulting in projected sales growth of 10%.
- Published poster "**Costs and Healthcare Resource Utilization for Cervical Conization in Mid-Adult Women in the United States**" in EUROGIN, 2018.

Academic Projects

Breast Cancer Detection in Histopathology Images

- Built efficient CNN models with MobileNet and Regularization as backbone to detect cancer in histopathology images.
- Gathered data from Kaggle and proposed solution achieved high accuracy and AUC of 88%.

Topic Modeling on Disneyland Visitor Reviews

- Implemented Latent Dirichlet allocation to construct topics contributing to both positive and negative feedback.
- Concluded topics across different locations and extracted actionable insights to improve customer experiences.

Wikipedia Toxic Comment Classification

- Developed ML models: Binary Relevance, Chain Classifier, LSTM to estimate different types of toxicity in comment.
- Experimented and compared results of using different feature embeddings like TFIDF, Glove, BERT.

Awards and Certifications

- Received several **Bravos** for Excellent Performance, Innovation, and Cross functional teamwork at UnitedHealth.
- **Excellence Award** for teamwork and ownership at CHEORS.