Yash Bhardwaj

+91-9024363101 | yashbhardwaj920@gmail.com | linkedin.com/in/yash | github.com/yash

Education

Birla Institute of Technology and Science — Pilani, India
Bachelor of Engineering in Computer Science — *Aug. 2017 – Aug. 2021*

Research Experience

Georgia Institute of Technology — May'24 - Present
Research Intern | *Financial Services Innovation Lab*

- **VideoConviction: A Multimodal Benchmark for Human Conviction and Stock Market Recommendations** | **KDD 2025**
  - Built a first-of-its-kind dataset of stock recommendations from YouTube videos, comprising 6,000+ expert-annotated gold labels, capturing multimodal conviction at a granular level (Facial expressions, Tone, Body language).
  - Utilized OpenAPI's Whisper large-v2 model to generate video transcripts.
  - Benchmarked several multi-modal models LLaMa-3.1, LLaVa-Next-Video, Gemini, and ChatGPT to predict stock recommendation from transcript and video, respectively.
  - Using the gathered recommendations, built multiple stock portfolios and compared them with S&P 500.
  - Published our work in the Conference on Knowledge Discovery and Data Mining (KDD) 2025.

Indraprastha Institute of Information Technology Delhi — Jan'21 - June'21
Research Intern | *LINK* — *New Delhi, India*

- **Author profiling - Extreme Multi-Label Classification** | *KNN, GNNs, Google Colab, GPU acceleration*
  - Worked on identifying authorship patterns in scientific publications based on writing styles and citations.
  - I used a graph-based GNN-XML model and a KNN-based AnnexML model on a graph with 400K nodes (Processed 2 million articles from the S2ORC dataset) to predict authorship and achieved a P@1 of 23 and P@3 of 15.
- **Citation Generation - Sequence to Sequence Learning** | *PyTorch, Transformers, Encoder-Decoder Design*
  - Given the source and target research papers, we aimed to generate the citation text for the target paper.
  - Leveraging the BART transformer, I designed a model to generate citation text using the abstract and conclusion of research papers as contextual inputs.
  - The model's effectiveness was verified through a group of 30 individuals, with 20% of the citations voted suitable to cite target articles.
- **Keyphrase Generation** | *BERT, BART, T5, Longformer model, Zero shot and few-shot learning, fine-tuning*
  - We examined the performance of pre-trained language models for the task of key-phrase generation for the KP20k dataset.
  - Leveraging the hugging face library, trained and fine-tined encoder-only transformer BERT2BERT and encoder-decoder transformers BART, T5, and Longformer on PyTorch.
  - We achieved P@5 of 28.3 for BERT2BERT and P@5 of 29.5, 30.3, and 20.7 for BART, T5, and Longformer.

Work Experience

Urban Company(E-Commerce) — July 2023 – March 2025
Software Development Engineer 2 — *Bangalore, India*

- **Product Catalog Cache** | *TypeScript, NodeJS, Kafka, Redis and MongoDB*
  - To cater to our rising product catalog (40% YoY), I designed and built a cached datastore with precomputed prices for 27000 products in our catalog across five countries and currencies.
  - Employed concepts of database design, caching, asynchronous operations, and events.
  - The new data store helped build a recommendation system for our products using statistical techniques and reduced pricing system latency by 40%.
- **Dynamic Pricing construct and a new Payment construct**| *System design, time-series analysis and predictive-modelling*
  - Built a dynamic pricing system that could adjust charges and discounts as per market conditions, demographics, and demand seasonality.
  - Used time-series forecasting (SARIMA) and predictive modeling (XGBoost) to support the creation of dynamic pricing configurations.

    - The tested control group showed a 4% revenue increment.
    - Built a multi-dimensional config to control how and when to take money from the user, has collected $80 million in revenue
- **Master Database Restructuring** | *AWS, parallel processing and performance analysis*
    - Restructured the core *orders* database with over 10 million entries - while maintaining concurrent reads, writes, and updates.
    - Added a script to deploy AWS EC2 compute units based on the database's current load due to day traffic.
    - Using batching and asynchronous operations, I achieved the restructuring in 6 days (initial estimate 46 days)

Software Development Engineer 1 *Jun 2021 - July 2023*

- **Careers Website LINK** | *React, MongoDB, AGILE, Scrum, AWS and Software lifecycle*
    - Built the company's career website from scratch on the MERN stack to create a seamless experience for the applicants and the HR Department.
    - I used React-native for the front end and MongoDB to store the job and applicant data.
    - The website gets 20K interactions per week and is our company's primary hiring mode.

Bridge your Network (Networking Platform) Jan 2021 - June 2021

Software Engineering Intern *Dublin, Ireland*

- **Bridge App** | *Javascript, React, Ruby on Rails*
    - I worked on the professional networking platform's web and iOS app that helps expand business networks and raise funding (Over 40M people).
    - Built the user profile screen on React for their iOS app and designed multiple APIs on Ruby for their backend.

Steel Authority of India, Government of India May 2019 - July 2019

Web Development Intern *New Delhi, India*

- **Srujani Web App** | HTML/CSS, Bootstrap, and MySQL
    - Tasked with building a web app for employee suggestion scheme for 15000 employees of Indian Government
    - I built the front-end on HTML/CSS and Bootstrap and stored the data in MySQL.
    - The scheme has saved over $1 million in costs.

Selected Projects (Independent Study Projects)

Bias Detection and Mitigation Framework | LINK | *scikit-learn, IBM AIF360* Oct 2024 – Dec 2024

- Built a framework to detect and mitigate bias in ML models using IBM AIF360 and scikit-learn.
- Analyzed fairness metrics (Disparate Impact, Statistical Parity Difference) and applied mitigation techniques (Reweighting, Disparate Impact Remover), improving fairness by 40%.
- Evaluated logistic regression models and conducted attribute-swapping to identify and address gender bias.

Biomedical Named Entity Recognition | LINK | *RNNs, LSTM, Keras, Tensorflow* Aug 2020 – Dec 2020

- Trained a bi-directional LSTM model, using Keras with Tensorflow at its backend to identify entities in literature.
- Incorporated a one-dimensional spatial dropout layer to reduce overfitting and used the Adam optimizer to achieve an accuracy of 98%.
- Using Transfer Learning, fine-tuned the model on the DECA dataset to identify biomedical named entities and matched the state-of-the-art results with an F1-score of 76.03.

Netflix Recommendation System | LINK | *Recommendations, Ranking, NumPy, Pandas* Jan 2020 – May 2020

- I built a hybrid Netflix-like recommendation system in Python based on content-based and collaborative filtering.
- The algorithm used TF-IDF and Single Vector Decomposition to learn the relationship between movies and users.

Technical Skills

Teaching Experience: Teaching Assistant for Compiler Construction course at BITS Pilani | LINK
Competitive Programming: Multiple top-10 ranks in Codechef programming competitions with over 10K participants | LINK
Languages: C, C++, Java, Python, Javascript, Typescript, SQL, R
Data Science Frameworks And Libraries: PyTorch, Keras, TensorFlow, NLTK, spacy, NumPy, pandas, sklearn

Extra Curricular Activities

College Sports Fest Organization: Organized events for inter-college sports fest with a footfall of over 10,000
Volunteer Experience: We carried out relief and welfare initiatives for the underprivileged during the pandemic with the UJAS organization.