Vaibhav Kumar

Webpage: web.cs.ucla.edu/~vaibhavk
Github: github.com/TimeTraveller-San
Email: vaibhavk@cs.ucla.edu
Mobile: +1 (424) 402-7223

Education

- **University of California, Los Angeles** — Sep 2021 – Mar 2023
  Master of Science in Computer Science — *Los Angeles, California*
- **Delhi Technological University** — 2016 – 2020
  Bachelor of Technology in Computer Science; GPA 8.25/10. — *Delhi, India*

Skills Summary

- **Programming Languages**: Python, C++, JavaScript, Java, Bash and Solidity
- **Frameworks**: PyTorch, Numpy, Pandas, OpenAI Gym, Scikit-Learn, Jax, Flask, NodeJS, Geth, LightGBM.
- **Development**: Docker, Flask, SQLite, jQuery, HTML, CSS, Web3.

Experience

- **UnitedHealth Group (Optum)** — Bangalore, India (Remote)
  Associate Data Scientist — *Aug 2020 – Jul 2021*
  - **Machine Learning**: Designed unsupervised models for automated intent classification and integrated them with the UHG customer care chat support. Achieved a **weighted F1 score of 0.88** – evaluated over an internally classified dataset carefully labelled by a diverse group of customer representatives.
  - **Impact**: The pipeline reduced the average case-assignment time from **10 seconds (manual) to** < 1 **second**.
  - **Development**: Wrote cron job schedules using Bash, Apache Spark and Python to fetch, aggregate, modify and update the internal customer archive databases with interactive visualization.
- **Laboratory for Computational Social Systems, IIIT Delhi** — Delhi, India
  Research Intern — *Mar 2019 – Dec 2020*
  - **NLP and Fairness**: Created novel multi-objective optimization-based methods in the word vector space to mitigate gender bias in non-contextual word emebeddings using PyTorch autograd. Also applied similar methods to the hyperbolic space using the geoopt library to debias the Poincaré GloVe embeddings.
  - **Impact**: Achieved state-of-the-art results in both downstream evaluation task **(co-reference resolution – 97.4%)** and benchmark datasets **(SemBias – 21.4%)**. Formulated the theory and presented the results in the form of two first-authored papers in the **TACL Journal (MIT Press)** and **NLP-OSS at EMNLP 2021**.
  - **Development**: Led the iterative code-development process for both Euclidean and Hyperbolic geometry in PyTorch. Open-sourced all the code and the associated datasets with proper reproducibility guidelines.

Projects

- **BSC Mainnet MEV Implementations**: Implemented several Miner Extractable Value (MEV) techniques like arbitrage and competitive front/back running over the BSC mainnet with custom profitable strategies devised using **gas-optimized smart contracts (Solidity)** and **Web3 (NodeJS and Python)**.
- **Fair Embedding Engine (FEE)**: FEE provides a framework for quantifying, mitigating and visualizing gender-bias in non-contextual word embeddings using a repertoire of state-of-the-art algorithms and benchmarks.
- **yTermPlayer**: A minimal, terminal based YouTube playlist streaming program written in python. Published on PyPI with more than **16,000 total downloads** (*pip install ytermplayer*).
- **Epidemic Spread Mapper**: A real-time epidemic mapping system build to use crowd-sourced data for both web and android. Utilized the past **temporal and spatial data** to predict future epidemic spreads using **LightGBM**.
- **Crypto Portfolio Alexa Skill**: An Alexa Skill for getting crypto price change alerts, portfolio management and transaction details. Backend API created from scratch using Python and Flask for multiple cryptocurrencies.

Publications

- **Nurse is Closer to Woman than Surgeon? Mitigating Gender-Biased Proximities in Word Embeddings**: **Vaibhav Kumar**, Tenzin Singhay Bhotia, Vaibhav Kumar, Tanmoy Chakraborty - Published in the Transactions of the Association for Computational Linguistics (**TACL**), Volume 8, 2020 p.486-503.
- **Fair Embedding Engine: A Library for Analyzing and Mitigating Gender Bias in Word Embeddings**: **Vaibhav Kumar**, Tenzin Singhay Bhotia, Vaibhav Kumar - Accepted at the 2^{nd} Workshop for NLP Open Source Software at Empirical Methods in Natural Language Processing (**NLP-OSS, EMNLP**), 2020.

Achievements

- **Generative Dog Images:** Won my second Kaggle silver medal and levelled up to a Kaggle competitions expert.
- **Santander Customer Transaction Prediction**: Bagged a silver Kaggle medal with an AUC-ROC score of 0.91