

# JASH MEHTA

☎ +1 (470) 909-9476 · ✉ jmehta73@gatech.edu · 🌐 jash-mehta-3300 · 🌐 jashmehta3300.github.io · 🌐 jashmehta3300

## EDUCATION

**Georgia Institute of Technology** Aug 2022 - Present  
M.S. in Computer Science, Specialization: Machine Learning GPA: **4.0/4.0**

- Graduate Teaching Assistant (**GTA**) for CS 6220: **Big Data** Systems and Analytics for Fall 22 & 23
- Graduate Teaching Assistant (**GTA**) for CS6675/CS4675: Adv. Internet Computing Systems for Spring 23
- **M.S. Project:** Mitigating Data Poisoning attacks in **Large-scale Distributed Systems**

**University of Mumbai** Aug 2018 - May 2022  
Bachelor of Engineering in Computer Engineering GPA: **9.67/10**

- Research Assistant (**RA**) under Prof. Ramchandra Mangrulkar for NLP & distributed (federated) systems.
- **B.E. Project:** Towards a Multi-Modular Decentralized System for Dealing with Healthcare data 🎓

## EXPERIENCE

**Cisco**, Software Engineer II (Intern) May 2023 - Present

- Creating tools for K8s clusters and CI/CD pipelines to enhance software supply chain security.

**Unicode Research**, Research student Jan 2022 - Apr 2023

- Served as TA for **Google Research** funded 9-week Machine Learning Course UMLSC 2021 with 100+ students.
- Built data pipeline for mining Twitter data and managed workflows using **Airflow**, AWS EC2 & S3, and Docker.
- Worked with the SimPPL team to build better civic integrity tools that support newsrooms to better understand their audiences on social networks. (supported by **Wikimedia** Foundation, **Google**, **AWS**, NYC Media Lab).

**Research Collaboration**, Independent Researcher Jan 2022 - Apr 2023

- Implemented statistical methods: re-sampling, cost-sensitive learning, and SMOTE for data with class-imbalance.
- Performed data mining to curated ~1M tweets in low resource Hindi language & conducted emoji prediction using bi-LSTM, mBERT, XLM-R, etc. (**Published: EMNLP 2022**)
- Standardized 9 hate-speech datasets & implemented LSTM, BERT, RoBERTa, etc. (**Published: EACL 2023**)
- Developed **distributed FL architecture** to obtain **14.52%** improvement in F1-score while preserving privacy.

**People Org**, Machine Learning Intern July 2020 - Dec 2020

- Devised and implemented dynamic pricing strategy resulting in **28%** increase in restaurant sales.
- Deployed ML models in production through **AWS EC2**, Amazon **Sagemaker** endpoints, ML APIs, and **Docker**.
- Worked on predictive analytics & AI modelling (Usage Forecast & Recommender System). Conducted analysis using statistics tools (**A/B Testing**).

**Margosatree Technologies**, Software Engineer Intern Jan 2020 - June 2020

- Performed anomaly detection with **Hadoop** and **Spark** to find 129 anomalies in 2TB of real-time IoT data.
- Data analysis using **PCA** & multivariate linear modeling; rendered visualizations using **Python & SQL**.

## SELECTED PUBLICATIONS [*\* represents equal contribution*] 🎓 FULL LIST:

- [1] Deep Gandhi\*, **Jash Mehta\***, Nirali Parekh, Karan Waghela, Lynette D'Mello, and Zeerak Talat, "A federated approach to predict emojis in hindi tweets," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, 2022 🎓.
- [2] Jay Gala\*, Deep Gandhi\*, **Jash Mehta\***, and Zeerak Talat, "A federated approach for hate speech detection," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Short Papers*, Association for Computational Linguistics, 2023 🎓.
- [3] **Jash Mehta\***, Deep Gandhi\*, Naitik Rathod, and Sudhir Bagul, "Indicfed: A federated approach for sentiment analysis in indic languages," in *Proceedings of 18th ICON 2021*, ACL Anthology, 🎓.

## SKILLS

Languages: Python, Javascript, Golang, C, C++, HTML  
Libraries/Frameworks: PyTorch, Tensorflow, Huggingface, Airflow, scikit-learn, NumPy, Flask, Node.js, React.js  
Databases: SQL, MongoDB, Redis, MySQL  
Tools: Git, Azure, AWS, Docker, Jupyter, Bash, Heroku, L<sup>A</sup>T<sub>E</sub>X