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
Graduate Teaching Assistant (GTA) for CS 6220 Big Data Systems and Analytics for Fall 22

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

Experience

Indian Institute of Technology (IIT), Gandhinagar, Research Intern May 2021 - Jan 2022
- Won **Best Presentation Award** for work done at LINGO lab on probing and robustness evaluation of automated review generators to adversarial perturbations.
- Formulated desiderata for an ideal generator system and provided a public leaderboard along with a framework for unified & comprehensive measurement of model performance.

Unicode Research, Research student & TA Aug 2020 - Jan 2022
- **Projects**: Estimating the causal effect of mentorship on student career outcomes; and small-world simulation to model opinion polarization of online communities using Pyro.
- **TA: Google Research** funded 9-week Machine Learning Course UMLSC 2021 with 100+ students.
- Participated in Shalizi-Stats reading group for Bayesian Statistics and Cosma Shalizi's book.

Dwarkadas J. Sanghvi College of Engineering, Undergraduate Research Assistant Jan 2021 - June 2021
- Implemented 4 efficient aggregation stratergies for federated learning on non-iid medical data, using ResNet & U-Net
- Trained UMLFiT & AWD-LSTM models for detection of Spear Phishing on a corpus of ∼73k emails.
- Published 2 chapters with *Dr. Ramchandra Mangrulkar* in Chapman and Hall/CRC in the domain of FL & NLP.

Feople Org., Machine Learning Intern July 2020 - Dec 2020
- Devised a dynamic pricing strategy for a restaurant which helped in improving the sales by 28%.
- Developed recommender system pipeline using PyToch and rendered visualizations with MERN stack.

Margosatree Technologies, Software Engineer Intern Jan 2020 - June 2020
- Used Spark-Scala to gain useful insights from large-scale real-time data generated by machines.
- Worked on a diverse array of client and internal projects like report generation software using Selenium & Pandas.

Projects

Hate Speech Detection using Federated Learning *Guide: Zeerak Talat*
- Standardized 9 hate-speech datasets and experimented with AWD-LSTM, BERT, FNet, DistilBERT, RoBERTa, etc. in federated & centralized settings. (Under review at **EMNLP 2022**)

A Federated Approach to Predict Emojis in Hindi Tweets *Guide: Zeerak Talat*
- Conducted cost sensitive learning and SMOTE for imbalanced emoji data using FedProx for training
- Curated a dataset of ∼1M tweets in low resource Hindi language & conducted emoji prediction using bi-LSTM, mBERT, IndicBERT, Hindi-Electra, XLM-R, etc. (Under review at **ACL ARR**)

Automotive Component Failure Prediction *Guide: Dr. Kriti Srivastava*
- Collaborated with a Big4 Consultancy firm to predict tyre life in vehicles using models such as MLP, XGB, etc.
- Designed a case study for the firm regarding tyre life uncertainty after extensive analysis of presented data.

Publications

[1] **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, .

[2] **Jash Mehta**, Deep Gandhi, Govind Thakur, and Pratik Kanani, "Music genre classification using transfer learning on log-based mel spectrogram," in *2021 5th ICCMC*, pp. 1101–1107, IEEE, 2021 .

[3] Deep Gandhi*, **Jash Mehta***, and Pranit Bari, "Ablation analysis of seq2seq models and vanilla transformers for spanish to english translation," in *Proceedings of the 3rd ICADCML*, Springer Nature, .

[4] Deep Gandhi, **Jash Mehta**, Nemil Shah, and Ramchandra Mangrulkar, "Federated learning for brain tumor segmentation on the cloud," pp. 261–278, Chapman and Hall/CRC, 2021 .

Skills

Languages:	Python, R, Javascript, C, C++
Libraries/Frameworks:	PyTorch, Huggingface, Pyro, Node.js, React.js, GraphQL, Pandas, scikit-learn, NumPy
Databases:	SQL, MongoDB, Redis, PostgreSQL
Tools:	Git, Jupyter, Docker, Bash, Heroku, AWS, Azure, LaTeX