MANYA RAMPURIA

+1 (470) 452-4894 | mrampuria1@student.gsu.edu | https://www.linkedin.com/in/manya-r/ | https://github.com/mrampuria1

Data science professional with intensive experience in ML and NLP to extract insights from high quality data, and adept in implementing machine learning models to solve real-world problems. Published an app on the Google Play Store with over 600+ downloads using Flutter. Seeking a role as Data Scientist to leverage my expertise to drive impactful results.

EDUCATION

Georgia State University, Atlanta, GA *Dec 2023*
M.S - Computer Science, *GPA : 4/4*

IIIT Naya Raipur, India *July 2022*
B.Tech - Computer Science, *GPA : 8.93/10*

Coursework : Artificial Intelligence, Predictive Modelling, Machine Learning, Deep Reinforcement Learning, DBMS, NLP

TECHNICAL SKILLS

Programming & DB : Python, Java, C, MATLAB, SQL
Frameworks : Visual Studio, Pandas, Sklearn, Keras, Tensorflow, Pytorch,Tableau, Git, AWS, A/B Testing, Flutter, Flask
ML Algorithms : Random Forest, LSTM, RNN, CNN, LLMs, BERT, GPT, Ensemble Learning, Decision Trees, Naive Bayes
Web Technologies : JavaScript, HTML, CSS

EXPERIENCE

Graduate Research Assistant, Georgia State University **- Adobe Research** *Aug.22' - Pst.*
[Marketing, SNScraper, MongoDB, JSON, statistical analysis, exploratory data analysis, topic modeling, OpenCV, GPU]
- Led a team of 2 developers to analyze **marketing strategies** by **predicting user engagement on Twitter**
- Utilized **Git for version control** and collaborative source code management to create repositories and commit changes
- Trained an XGBoost Regressor using GridSearchCV with 0.73 R2-score on the **multimodal data** (text, image, time-series) to forecast user engagement, with a **21% improvement** due to the inclusion of engineered features

Data Science Intern, Cognistx, Pittsburgh *Dec.21' - Jan.22'*
[urllib**,** BeautifulSoup, transformers, coherence score, ROC-AUC, precision-recall curve, F1 score, xpdf, PdfMiner]
- Designed a **web crawler** that boosted relevant **data collection efficiency by 30%**
- Improved classification accuracy by 17% through threshold analysis on **zero-shot classification** confidence scores
- Integrated **BertTopic** with domain-specific legal embeddings that resulted in the expansion from 2 to 4 well-defined clusters

Deep Learning Intern, Teksands.ai, Bangalore *May.21' - July.21'*
[Knowledge Graphs, Cosine Similarity, Bert]
- Implemented a **siamese architecture** to identify contextual **similarities between two documents** by incorporating transformer-based legal bert embeddings with ConceptNet entities

Machine Learning Intern, Clintx, Kolkata *Mar.21' - May.21'*
[text-preprocessing, AWS]
- Analyzed healthcare tweets and extracted biomedical entities using **ScispaCy and UMLS**
- Utilized **AWS database** to store data and accomplished the task by deploying the solution on the **AWS lambda platform**

PROJECTS

Leveraging Social Media for Crisis Management *Feb.23'*
- Implemented an efficient **event extraction system** on disaster tweets
- Improving aid distribution through the integration of NER, hierarchical clustering, and topic modeling

DEFEND- Detection of fake news during pandemic using Deep Learning *May.21'*
- Deployed the **BERT** model with **96%** accuracy in a web app using **Flask and Heroku** for practical use

Reddit flair classification using NLP *Apr.20'*
- Successfully web scraped 30k Reddit posts with subreddit "r/india" from scratch using **Pushshift and PRAW**
- Performed data preprocessing and feature extraction techniques (word2vec and GloVe embeddings) to train a MLP classifier

CERTIFICATIONS

1. Introduction to Generative AI (Google)
2. NN, DL and Improving Deep Neural Networks (deeplearning.ai)
3. DNC, Sort & Search (Stanford Online)
4. Data Structures & Algorithms (Udacity)