

YASHASVI MANTHA

Data Scientist | Engineer | NLP Researcher

Education

12th Standard Electronics as Major

S.R.P Jr College - Mumbai

2016 - 2017

MHSC 75%

Bachelor of Computer Engineering

Gandhi Institute of Technology - Vizag

2017 - 2021

8.66 CGPA

Location

In order of Preference

Bombay / Hyderabad / Bangalore / Visakhapatnam

Social



+91-9121837667



yashasvimantha@gmail.com



linkedin.com/in/yashasvimantha/



github.com/YashasviMantha



yashasvimantha.com

Work Experience

ISS - ESG (Institutional Shareholder Services - ESG), Mumbai

2022 Jan - Present

Data Scientist - Innovations Lab

Keywords: Python, Kubernetes, Gitlab CI/CD, PySpark, Elastic Search, Spacy, Vector Database, Neo4J, NiFi, SKLearn, Private Cloud

- Migrated entire platform to private cloud including 40+ Micro-services and 32+ ML Models.
- Implemented a Hybrid Search based ranking with Vespa (Vector Database) for classifying ESG Specific News Articles.
- Built a Text Summarization application with Streamlit using various Transformer Based Neural Networks.
- Developed an Entity Linking pipeline using a mini - inhouse Knowledge Graph in parallel with DBPedia.
- Re-evaluated models using task specific evaluation metrics for classification and clustering.
- Designed and developed Data Annotation platforms for building domain specific datasets for a clustering task.
- Developed pipelines for various NLP tasks including NER, Entity Linking, Unsupervised Clustering, QA, etc.

Deloitte India (Office of the US), Hyderabad

**2021 Jan - May (Intern)
July - Dec (FTE)**

Intern (4 months) + Full Stack Engineer

Keywords: SQL, Stored Procedures, Agile, Backend Development

- Responsible for developing and maintaining a web-based Audit Quality Monitoring Platform that uses .NET Core and Angular along with Microsoft SQL.
- Fixed numerous bugs, LSIs, and defects that disturbed the workflow of end-users.
- Made a sample MVC application under the repository pattern under the initial training program.
- Wrote various seed scripts and stored procedures that automated the process of clearing transaction data and populating seed data in tables.
- Also wrote segments of documentation of the WebAPI that was actively being developed.
- Understood the Agile and DevOps workflow during the project timeline

Centre for Indian Language Technologies, IIT Bombay

2019 May - Aug

NLP Research Intern under Prof. Pushpak Bhattacharyya

Keywords: Wordnet, Glove, BERT, Fasttext, Pandas, Python, Phylogenetics

- Worked on 14 different Indian languages from the IndoWordNet and created Cross-Lingual word embeddings using Muse and FastText.
- Surveyed various methods to calculate a 14x14 Distance matrix for these languages including 3 low resource languages.
- Utilized edit distance algorithms like Normalized Levenshtein Distance and Angular Cosine to calculate a 14x14 Distance matrix.
- Was responsible for cleaning, script conversion and offset conversion to devanagari script for various Indian languages. Also had to handle large wikipedia dumps for FastText, MUSE model training.
- Implemented various tree construction algorithms like UPGMA, WPGMA and Neighbour-joining.
- Calculated cross-lingual word embeddings between 14 language pairs to perform phylogeny.
- Used the SkipGram Model to effectively get embeddings.
- Significantly improved Distance scored (by at least 70%) from the baseline approach to the novel approach.

Research Publications and Presentations

ACM IKDD CODS And COMAD (India):

- First author of ACM's CoDS COMAD YRS-Track 2020 titled "Harnessing Deep Cross-lingual Word Embeddings to Infer Accurate Diachronic Phylogenetic Trees". Presented at ISB (Indian School of business), Hyderabad about how the Indian Languages evolved through a tree.
- Poster Presentation under the CoDS COMAD, Young Researcher's Symposium 2020 at 7th ACM IKDD CoDS 25th COMAD about Inferring Accurate Diachronic Phylogenetic Trees.