

# MAYANK GOEL

Hyderabad, India · mayank.goel@research.iiit.ac.in · +91 7890545454 · <https://mayankgoel28.github.io/>

## EDUCATION

**International Institute of Information Technology**  
BTech + MS Computational Linguistics Dual *GPA: 8.18*

Hyderabad  
July 2019 -

## EXPERIENCE

### Proxzar.ai

*Research Intern*

Hyderabad, India  
Dec 2021 - Present

- Used Apache Nutch/Selenium to crawl multiple websites, along with BS4 and related tools to extract relevant content to create NLP Datasets. Solved problems such as login barriers and utilized parallelization and other techniques to make the data extraction process efficient.
- Built a Spacy3 model utilizing Transformers with an iterative approach to maximize accuracy and efficiency for a production-grade custom NER model for a construction e-commerce company, as part of an end-to-end system, including a pipeline for data augmentation.

### Lotusdew Wealth and Investment Advisors

*SDE Intern*

Hyderabad, India  
May 2022 - June 2022

- Built an Neo4j dashboard to analyze semantic similarity between companies. Finetuned BERT on finance corpora for calculating semantic similarity.
- Built a wrapper over Zerodha API using Phoenix Web Framework

### Quriate

*SDE Intern*

Remote  
Aug 2021 - Aug 2021

- Built multiple data analytics components by extensively using pandas on user data for an examination portal, with an emphasis on fast computation and vectorized code to work on a centralized MongoDB database. Was also involved in ideation for useful metrics and visualizations for end-users (students and teachers).

### Trivedi Centre for Political Data, Ashoka University

*Research Associate*

Remote  
May 2021 - June 2021

- Made data visualizations using Plotly, displaying them on a dashboard made using Streamlit, with interactive features and parameters to view the data, for an internal diagnostics tool. Used numpy and pandas to develop an extensive set of features.

## SKILLS

Languages: Python, C++, Javascript, Bash  
Frameworks, Libraries and other tools: Spacy3, PyTorch, Apache Nutch, MERN

## PROJECTS

### Echodex Python

<https://github.com/chaudhary1337/Echodex>

Wrote a generalized tool to fetch data from a news API, Twitter and Reddit. Worked on fetching and cleaning the data to a frontend made using streamlit, after using NLP tools for analysis such as sentiment analysis and document classification.

### Job Application Portal MERN

<https://github.com/MayankGoel28/JobApplicationPortal>

A full-stack web app with functionality to support use cases of recruiters as well as job seekers. Includes a JWT based user authentication system.

### Joke Generator Python

<https://github.com/MayankGoel28/JokeGenerator>

Made a novel model to generate jokes, using a statistical technique based on core theories of humor to extract templates and utilize BERT's MLM architecture for infilling. A presentation based around this project was accepted to the 11th Humor Research Conference.

## ADDITIONAL EXPERIENCE AND REWARDS

**Best Project (Theme track)**

Natural Language API hackathon, expert.ai

**1st Prize, Equity Track**

HackUMBC, University of Maryland

**Best Web Application**

HackAtHome, Brown University

**3rd Prize, Walmart Track**

TAMU Datathon, Texas A&M University

**Teaching Assistant**

Introduction to NLP  
Spring'22