

# HEMANT RATTEY

Rochester, New York | +1 (585) 719-4929

hr5259@rit.edu | [linkedin.com/in/hemantrattey/](https://www.linkedin.com/in/hemantrattey/) | [hemantrattey.github.io/](https://github.com/hemantrattey)

## EDUCATION

### Rochester Institute of Technology

Master of Science in Data Science (GPA: 4.0/4.0)

Rochester, NY

Exp Dec '24

### Indraprastha Institute of Information Technology

Bachelor of Technology in Computer Science and Engineering

New Delhi, IN

Dec '19

## TECHNICAL SKILLS

**Language:** Java, Python, R, HTML, CSS, XML

**Frameworks:** Scikit-learn, TensorFlow, PyTorch, Keras, Pandas, Streamlit, RASA, LangChain, Tableau, Transformers

**Databases:** MySQL, MongoDB, PostgreSQL, Neo4j, Pinecone

## WORK EXPERIENCE

### Rochester Institute of Technology

*Tutor*

Rochester, NY

Jan '23 – Present

- Mentored and tutored over 400 students, delivering one-on-one assistance in Python and Java programming concepts.

### Rochester Institute of Technology

*Graduate Teaching Assistant*

Rochester, NY

Aug '23 – May '24

- Assisted professors in preparing course material and assignments catering to a variety of problem domains for ISTE230 - Introduction to Database Design and Data Modelling and ISTE 438 – Contemporary Databases; provided assessment support.

### iTech Mission Pvt. Ltd.

*Data Analyst and Visualization Intern*

New Delhi, IN

May '21 – Jan '22

- Spearheaded the automation of data analysis and visualization tasks, reducing the time spent on these tasks by 70%, enabling the team to focus on high-value work.
- Orchestrated a cross-functional team to create a chatbot for the UN India Intranet Data Portal; used open-source RASA and Python to reduce data search time by 75%.
- Led the development and implementation of Python scripts for automating data extraction and cleaning, resulting in an 80% reduction in data cleaning time and saving 10 hours per week.
- Analyzed a massive dataset of 3 million rows with over 100GB of data using MySQL and automated reporting with crontab during the Poshan Maah event; reduced data analysis time by 50% and increased event efficiency.

### Johnson & Johnson

*Data Analyst*

Mumbai, IN

Apr '20 – Jan '21

- Collaborated and executed data-driven sales strategies with the sales team, utilizing statistical analysis to identify untapped regions and increase revenue by 25% in Q3.
- Automated the sales reporting process by developing SQL queries and Python scripts that generated daily, weekly, and monthly reports, saving the team approximately 8 hours per week.
- Analyzed consumer sales data using Tableau and collaborated with the sales team to develop the Decision Cockpit View dashboard, enabling the team to gain deeper insights into customer behavior which led to a 20% increase in sales.

## PROJECTS

### From Breaking News to Breaking Civility: A Study of Toxicity Trends in Political Debates

Feb '24 – Dec '24

*Aim: To analyze toxicity trends in political debates using NLP and BERT-based models for cross-cultural insights.*

- Compiled a comprehensive dataset of prime-time political debate transcripts, downloading over 1000 Indian debate videos from YouTube and scraping 5000+ U.S. transcripts from CNN, Fox News, and MSNBC to enable cross-cultural toxicity analysis.
- Transcribed debate videos using OpenAI's Whisper for automatic speech recognition; segmenting transcripts into one-minute intervals to capture granular toxicity scores without context dilution.
- Implemented advanced NLP techniques for data preprocessing to optimize model accuracy in detecting toxicity levels in debates.
- Applied Perspective API and fine-tuned BERT-based LLM (*unitary/toxic-bert*) to perform segment-wise toxicity scoring, enabling precise analysis of hostility in Indian and U.S. debates.
- Conducted exploratory data analysis to identify patterns, revealing concentrated toxicity in the initial segments of Indian debates compared to moderate yet explicit language in U.S. debates, illustrating distinct discourse and hostility trends across regions.

### PDF Pal: Your Document Dialogue Companion

Feb '24 – Mar '24

*Aim: To develop a chatbot companion to interact with uploaded PDF files*

- Utilized PyPDF2 to seamlessly extract text from PDF files and convert it into 784 dimensional embeddings using InstructorEmbedding; leveraged Pinecone vector database for storing embeddings.
- Integrated Google Flan T5 XXL as the language model, ensuring robust conversational capabilities; employed LangChain framework and RetrievalQA chain for question-answering.
- Developed a user-friendly frontend using Streamlit, enhancing the overall user experience and accessibility.

### Web Search Application for Google Local Data

Jan '23 – May '23

*Aim: To build a web search application for seamless retrieval of Google Local Data*

- Performed data cleaning and preprocessing using Pandas on 700 million reviews for 5 million businesses; structured the data into JSON files, facilitating seamless processing and analysis in subsequent stages.
- Scraped over 700 images from Google Maps using Selenium; leveraging GridFS for efficient storage and retrieval.
- Incorporated Flask to build the application with search functionality for location, business name, and address to retrieve information about businesses and the reviews.