HEMANT RATTEY

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

hr5259@rit.edu | linkedin.com/in/hemantrattey/ | hemantrattey.github.io/

OBJECTIVES

Data Science enthusiast with over 2 years of experience in analytics and data-driven decision making, adept in Machine Learning, Deep Learning, and Data Science methodologies. Currently pursuing a Master's in Data Science at RIT, seeking full-time opportunities starting May 2024 to drive impactful insights and solutions in a dynamic environment.

EDUCATION

Rochester Institute of Technology — Rochester, NY
Master of Science in Data Science (**GPA: 4.0/4.0**) — Exp May '24

Indraprastha Institute of Information Technology — New Delhi, IN
Bachelor of Technology in Computer Science and Engineering — Dec '19

TECHNICAL SKILLS

Language: Java, Python, R, HTML, CSS, XML
Frameworks: Scikit-learn, TensorFlow, PyTorch, Keras, Pandas, Streamlit, RASA
Databases: MySQL, MongoDB, PostgreSQL, Neo4j, Pinecone

WORK EXPERIENCE

Rochester Institute of Technology — Rochester, NY
Graduate Teaching Assistant — Aug '23 – Present
- 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.

GCCIS Tutoring Center, Rochester Institute of Technology — Rochester, NY
Tutor — Jan '23 – Present
- Mentored and tutored over 400 students, delivering one-on-one assistance in Python and Java programming concepts.

iTech Mission Pvt. Ltd. — New Delhi, IN
Data Analyst and Visualization Intern — 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 — Mumbai, IN
Data Analyst — 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

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.

Predicting Readmission of Hospital Patients — Aug '22 – Dec '22
Aim: To implement a classification model that predicts the readmission of a patient
- Engineered a robust data pipeline to perform data preprocessing on a dataset spanning 10 years of clinical care at 130 US hospitals, with data on 100,000+ patients having 50+ features.
- Implemented Logistic Regression, Random Forest, and XGBoost algorithms to predict readmission of hospital patients; achieved 72% F1-score, 81% Precision, and 67% Recall through hyperparameter tuning.