HEMANT RATTEY

380 John Street, Rochester, NY 14623
(585) 719-4929 | hr5259@rit.edu | https://hemantrattey.github.io/

EDUCATION

Rochester Institute of Technology   Rochester, NY
Master of Science in Data Science **GPA: 4.0/4.0**   Exp Aug '24

Indraprastha Institute of Information Technology Delhi   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, Pandas, Keras, PyTorch, Numpy, Matplotlib, RASA, Seaborn
Databases: MySQL, MongoDB, PostgreSQL, Neo4j

WORK EXPERIENCE

GCCIS Tutoring Center, Rochester Institute of Technology   Rochester, NY
Tutor   Jan '23 – May '23
- Mentored and tutored over 400 students enrolled in GCIS 120, GCIS 123, and GCIS 124 courses, delivering personalized, 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% and 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.
- Conducted ad-hoc reporting for Key Account and Ecommerce business managers, providing timely and customized insights to support strategic data driven decision-making.
- 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

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
- Perfomed data cleaning and preprocessing using Pandas on 700 million reviews for 5 million businesses, aggregating data for a few selected states; structured the data into JSON files, facilitating seamless processing and analysis in subsequent stages.
- Scraped over 700 images from Google Maps using Selenium; loaded the cleaned data and images into a MongoDB database, 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 build a classification model that predicts the readmission of a patient
- Investigated the problem of hospital readmission by analyzing electronic health records and worked to develop a predictive model that would assist hospital administrators in improving patient outcomes.
- Engineered a robust data pipeline to perform data cleaning and 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.

Understanding Link Farming on Twitter (FriendsFarmer)   Aug '17 – Dec '17
Aim: To study the behavior of accounts on Twitter and classify as bot or not
- Extracted over 1,000 Twitter accounts based on keywords indicating spam behavior using Tweepy and the Twitter API.
- Developed a robust data processing pipeline to extract essential features, including account metadata, tweet content, and engagement metrics, ensuring high-quality data for analysis.
- Achieved 77% accuracy in classifying Twitter accounts using Random Forest classifier; optimized classification model to identify bot accounts involved in link farming by analyzing characteristics, including tweet frequency and engagement metrics.