HEMANT RATTEY

380 John Street, Rochester, NY 14623
(585) 719-4929 | hr5259@rit.edu | www.linkedin.com/in/hemantrattey

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, ggplot, plotnine
Databases: MySQL, MongoDB, PostgreSQL, Neo4j
Tools: Jupyter Notebooks, Tableau, LaTeX, Git, MS Excel

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.
- Demonstrated ability to foster a deeper understanding of the material, instill confidence, and achieve substantial grade improvements among students.

iTech Mission Pvt. Ltd.   New Delhi, IN
Data Analyst and Visualization Intern   May '21 – Jan '22
- Executed data analysis and visualization tasks to provide insights into complex data sets, utilizing tools such as Excel, R, Python and Tableau to create reports and dashboards for clients.
- Collaborated with a team to create a chatbot for the UN India Intranet Data Portal, utilizing open-source RASA and Python.
- Worked closely with cross-functional teams to understand client requirements and ensure that data analyses and visualizations were aligned with business objectives.
- Played a key role in developing and maintaining data management systems, employing Python scripts for data cleansing, transformation, modeling, while also providing quality control and testing support to ensure iTech Mission's software products achieved exceptional accuracy and usability standards.
- Successfully managed and analyzed a large-scale dataset of over 100GB with more than 3 million rows using MySQL for reporting and implemented automation using crontab during the Poshan Maah event.

Johnson & Johnson   Mumbai, IN
Data Analyst   Apr '20 – Jan '21
- Cooperated with the sales team to develop and implement data-driven strategies to drive sales growth and utilized statistical techniques to identify patterns, trends, and insights from large datasets for E-commerce and Key Accounts.
- Utilized Python for data analysis and script writing, including data cleaning, data manipulation, and data modeling.
- Created SQL queries to extract, manipulate, and analyze data from various sources, including internal databases & external APIs.
- Worked with cross-functional teams to identify opportunities for data-driven insights and to communicate findings to key stakeholders. Conducted ad-hoc analysis to support business decisions and address sales team's inquiries.
- Leveraged Tableau to create interactive dashboards and reports for the sales team and senior management. Supported the team to develop the Decision Cockpit View dashboard for analyzing consumer sales data in APAC region.

PROJECTS

Web Search Application for Google Local Data   Aug '22 – Dec '22
Aim: To build a web search application for seamless retrieval of Google Local Data
- Conducted 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 hundreds of business images from Google Maps using Selenium and loaded the cleaned data and images into a MongoDB database, leveraging GridFS for efficient storage and retrieval.
- Implemented geocoding and geospatial query functionalities using geopy and OpenCage API.
- Utilized Flask to develop the web search application to search for review information and business metadata.

Predicting Readmission of Hospital Patients   Aug '22 – Dec '22
Aim: To build a classification model that predicts the readmission of a patient
- Designed pipeline to perform data cleaning, preprocessing and feature engineering along with EDA.
- Leveraged supervised learning classification algorithms namely Logistic Regression, Random Forest and XGBoost for predicting readmission of hospital patients.
- Evaluated the models with hyperparameter tuning to perform better with 72% F1-score, 81% Precision and 67% Recall.

Understanding Link Farming on Twitter (FriendsFarmer)   Aug '17 – Dec '17
Aim: To study the behavior of accounts on Twitter and classify them as bot or not
- Gathered a dataset of more than 1,000 twitter accounts using the Twitter API and Tweepy library.

- Developed a robust data processing pipeline to extract relevant features such as account metadata, tweet content and engagement metrics.
- Implemented various supervised machine learning classification algorithms such as Random Forest, Decision Trees and Logistic Regression to classify the accounts as bot or not based on the extracted features.
- Achieved an accuracy of 77% on Random Forest classifier in classifying Twitter accounts, demonstrating the effectiveness of the developed classification model.
- Conducted in-depth analysis on characteristics of bot accounts involved in link farming, identifying traits such as high tweet frequency, high following-follower ratio, similar posting patterns and low engagement with genuine users.