Rutvi Bheda

Data Enthusiast

585-498-0563 | LinkedIn | GitHub | rutvibheda317@gmail.com

EDUCATION

Master of Science in Computer Science   **Jan 2023 – Present**
Rochester Institute of Technology   Rochester, New York
Bachelor of Engineering in Computer Engineering   **Aug 2017 – Jul 2021**
University of Mumbai   Mumbai, India
Coursework: Algorithms, Big Data Analytics, Machine Learning, Artificial Intelligence, Information Retrieval, Visual Analytics

TECHNICAL SKILLS

Programming Languages	: Python, R, Java, C, C++, HTML, CSS, PHP, JavaScript, Unix Shell Scripting, Linux
Databases	: SQL, Spark SQL, Oracle RDBMS, MongoDB, PostgreSQL, NoSQL
Libraries	: Pandas, NumPy, NLTK, Matplotlib, TensorFlow Scikit-Learn, OpenCV, Gensim, Plotly
Tools and Frameworks	: Git, JIRA, Docker, Django, AWS, Azure, Apache Spark, Kafka, Scala, Tableau, Power BI, Looker
Technical Skills	: Data Science, Machine Learning, Data Modeling, Database Management Systems

WORK EXPERIENCE

Retail Data Engineer | *Auxiliary Services at Rochester Inst. of Technology*   **May 2023 – Present**
- Analyzing RIT's dining data systems by working closely with the team and leveraging Agilysys software focusing in point of sale by customizing and managing **ETL processes** for OnDemand items, multiple menus, kiosks, and registers
- Performed **data cleansing** and **manipulation** on a dataset of approximately **30,000 inventory items**, ensuring accuracy and consistency by handling missing values, duplicates, and standardizing data formats to improve accuracy

Data Engineer | *Accenture*   **Aug 2021 - Sep 2022**
- Optimized retail data processes for a banking client by directing **ETL workflows** using **Ab Initio** and **Autosys**, resulting in an **80% improvement** in data processing efficiency and significantly enhancing operational speed
- Deployed **advanced ETL** components like **Aggregate**, **Join**, **Reformat**, **Rollup**, and **Scan** to effectively manage and transform large data sets and executed **complex SQL queries** thereby increasing the throughput by over **50%**
- Directed monitoring of daily scheduled jobs, used **Jira** following **Agile** methodology for feature tracking, conducted detailed **run reviews**, ensuring error-free deployments and the consistent delivery of functional software sprints
- Implemented **parallelism techniques** for processing data simultaneously, reducing processing times by **40%**

PROJECTS

AWS ETL Pipeline on YouTube Data| *AWS, Python, Spark, Data Analysis* *Quicksight Dashboard*
- Designed an **end-to-end data pipeline** utilizing **AWS S3** for storage, **Lambda** for event-driven processing, and **Spark** combined with **Athena** to transform and analyze the YouTube trending videos dataset containing over **1 million rows**
- Directed the **ETL** process using **AWS Glue crawler** and built **QuickSight dashboards** to **visualize key metrics**, enabling actionable insights into video trend patterns and make informed decisions

New York City Crash Map Data Mining | *Python, Postgres, Plotly, Data Mining*
- Performed **exploratory data analysis** on a dataset of **two million** traffic accident records in the Brooklyn borough using Postgres and Python for data cleaning and mining to understand accident patterns for further analysis
- Executed the **KMeans** clustering due to its efficiency to identify accident hotspots and visualized using **Plotly** for generating **heat maps** and **bar plots**, effectively getting patterns and trends of traffic accidents over consecutive years

CLEANSE - Hate Speech Classifier and Euphemistic Text Generator | *Python, Pandas, Scikit-Learn*
- Developed a hate speech detection system using Python and achieving an accuracy of **96%** using logistic regression and used NLP methods such as **sentiment analysis**, **semantics**, and **unigram method** on a dataset of **39k tweets**
- Employed **WordNet** to generate euphemistic substitutes for the offensive language in the tweets to decrease the hateness and seamlessly integrating into an API and documenting the project's methodology in a **technical paper**

British Airways Reviews | *Tableau, Data Analysis* *Tableau Link*
- Developed a **Tableau** dashboard with **interactive visualizations** and **maps** of the British Airways reviews to monitor performance metrics/**KPIs** and enabling data-driven decisions to identify and address the areas for improvement

PUBLICATIONS AND CERTIFICATIONS

- Identification of Hate Speech using Natural Language Processing and Machine Learning, IEEE (Paper Link)
- Educational Advancements in the Field of Augmented Reality and Virtual Reality, IEEE (Paper Link)
- **AWS Certified Cloud Practitioner**, Amazon Web Services (Link)
- **Hands-On Essentials: Data Warehousing Workshop**, Snowflake (Link)

CO-CURRICULARS

- SWE 2024 In-Person Attendee, Chicago, Illinois (October 2024)
- GHC 2023 Attendee, Virtual (September 2023)
- On-Stage Head of Entrepreneurship Cell (2019-2020)