Rahul Singh

singhrahulsb@gmail.com | Erie, PA | +1 (551)-280-6143 | Hacker Rank
LinkedIn/in/rahuls0 | tableau/Rahul Singh | github.com/singhrahulbrijesh

RELEVANT EXPERIENCE

Business Data Analyst   **PCG, Harrisburg PA**   **Aug 2023 – Present**
- Collaborated with stakeholders to capture critical data specifications and process improvements for **Medicaid data**.
- Creating reports and dashboards to present findings to stakeholders, including healthcare providers, and vendors.
- Analyze business requirements and document them appropriately in the form of user stories in **Team Foundation Server and Azure DevOps**.
- Collaborated with project and testing teams to update project artifacts in response to changing requirements.
- Conducted in-depth data analysis using tools such as Excel, **SQL**, Tableau, and **IBM Cognos Analytics** to derive actionable insights and support decision-making processes.

Data Science Research Assistant   **Gannon University, Erie PA**   **Aug 2022– May 2023**
- Trained Analyzed and tested the null hypothesis for user-purchase behavior derived from user demographics and purchase feedback data.
- Deployed a sparse adversarial model SRLGAN for hybrid User Cold-Start Recommendation using the proposed Sparse Regularizer achieving a precision of **0.53**, Normalized Discounted Cumulative Gain score of **0.52**, and Mean Reciprocal Rank score of **0.7**.
- Implemented algorithms in GANs like LSRGAN, and WGAN for Zero-Shot Classification using PyTorch accomplishing a top-1 accuracy score of **0.64** on popular datasets like AWA with a training data size of approx. **40K** seen and **10K** unseen class.

Data Assistant   **Gannon University, Erie PA**   **Oct 2021–May 2023**
- Utilized ImageNow software to scan and file confidential international student documents, ensuring proper organization.
- Proficient in Microsoft Excel (**VLOOKUP, SUMIF, INDEX-MATCH, Pivot chart**) to create and manage spreadsheets, perform data analysis, and generate reports, contributing to **30%** of the role.

Python Developer   **Make My Clinic Pvt Ltd**   **July 2019 – May 2021**
- Designed e-commerce platforms for clinic products and services using Python and related frameworks, achieving a functionality and performance rate of **88%**.
- Developed RESTful APIs in coordination with cross-functional teams, including Developer and testing teams, enabling seamless integration with third-party applications, and achieving a **93%** API success rate.

EDUCATION

M.S. Computer Information Science- Data Science	**Gannon University, PA**	*GPA*: 3.875	**Aug2021 - May 2023**
Bachelors - Information Technology	**Mumbai University, India**	*GPA*: 3.25	**June 2016 - July 2019**

SKILLS

Programming Language: Python, R- Language, C/C++, SQL, JavaScript, Database management, ReactJS.
Cloud & Technologies: Flask, Django, MongoDB, Oracle, Snowflake, POSTGRE SQL, AWS, GCP, Azure.
Quality & Project Management: MS Test Manager (MTM), Team Foundation Server (TFS), MS Project, Jira, Azure DevOps.
Big Data & Analytics Tools: Hadoop, Spark, Tableau, Power BI, Stata, SSRS, SPSS, Qlik Sense, SAS, IBM Cognos.
Libraries: - TensorFlow, Pandas, NumPy, Spark, Pytorch, OpenCV, Keras, Matplotlib, Scikit-learn, Beautiful Soup, Selenium.

RESEARCH

Automating Patch set generation from code review comments using LLM - LINK   **Nov 2023 – ongoing**
- Evaluated the effectiveness of pre-trained LLMs (**GPT-4, Google's BARD, GitHub Copilot**, etc.) by comparing their generated patch sets with historical human-generated data from the same repositories, providing a detailed assessment of their capability to replicate human tasks in code review.
- Applied research to Apache projects (**Kafka, Spark, Airflow**), emphasizing real-world pull requests. Enhancing relevance by automating outcome assessment, integrating matched code changes with **80%+** similarity, and directing lower matches for manual review by developers by interviewing them.

Categorizing the common defects in modern Web browsers by KE to LLM - LINK   **March 2022 - ongoing**
- Conducted web scraping with Selenium to gather large datasets (**6M** for Firefox, **8M** for Chrome) and used advanced NLP and GPT-4 to analyze defects, improving insights into bug patterns and browser performance.
- Developed and deployed a GPT-4 model, achieving a **30%** boost in defect categorization accuracy and cutting bug analysis time by **40%**, enhancing overall efficiency.

CERTIFICATION

- KPMG Data Analyst Virtual Internship – Developed valuable insights into the target audience and key consumer characteristics such as age, gender, and income. Translated data insights into presentations, facilitating informed decision-making for stakeholders.