Rahul Singh

[singhrahulsb@gmail.com](mailto:singhrahulsb@gmail.com) | Erie, PA | +1 (551)-280-6143 | Hacker Rank

[LinkedIn/in/rahuls0](https://www.linkedin.com/in/rahuls0/) | [tableau/Rahul Singh](https://public.tableau.com/app/profile/rahul.singh5386) | [[github.com/singhrahulbrijesh](https://github.com/singhrahulbrijesh)](https://d.docs.live.net/0a7dbf7c88842a6f/Documents/Resume/%20%20%20%20%20%20%20%20GitHub:%20git)

**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**.
* Developed comprehensive business documentation, including procedures, work instructions, and process flow diagrams.
* 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, 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 Aug 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- DataScience** Gannon University, PA***GPA:*** *3.875* **Aug2021** - **May 2023**

**Bachelors - Information Technology**   **Mumbai University, India** **G*PA****: 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**](https://dl.acm.org/doi/10.1145/3644815.3644981) **Nov 2023 – ongoing**

* Evaluated the Effectiveness of Pre-trained LLMs, including **GPT-4, Google’s BARD, GitHub Copilot**, and others, in replicating human tasks related to code review.
* Meticulously compared Large Language Models' patch sets with historical human-generated data in the same repositories.
* 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.

**Identification of common defects in modern web browsers | Dr. TAJMILUR RAHMAN March 2022 - ongoing**

* Performed web scraping using Selenium to collect and extract bug data from repositories and databases of Firefox and Chrome, resulting in datasets of **6 million** and **8 million** Large Datasets respectively.
* Implemented K-means clustering and **Bert model** to refine the dataset by grouping similar synonyms and removing less important words occurring below **0.3%**, resulting in a more accurate and focused dataset.

**CERTIFICATION**

* [KPMG Data Analyst Virtual Internship](https://forage-uploads-prod.s3.amazonaws.com/completion-certificates/KPMG%20AU/m7W4GMqeT3bh9Nb2c_KPMG%20AU_s4y3BdfNMzjfinoWN_1689465569067_completion_certificate.pdf) – 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.