**Department of Master of Computer Applications**

18-06-2024

**IV Semester**

**MCA491P – Major Project**

**Data Quality Assessment Application**

Data Quality is a major concern for organizations across various domains. Reliable data is essential for informed decision-making and efficient operations. However, ensuring data quality often involves manual processes and complex frameworks, making it challenging for users with limited technical expertise to follow along the process. This project introduces the Data Quality Assessment Application, a user-friendly tool designed to address these limitations. The application provides a guided GUI for users to assess the quality of their datasets. It simplifies the process by automating several key steps involved in a data quality framework to a large extent.

The application utilizes a combination of modular and object-oriented programming techniques in Python at the backend. Streamlit, a Python framework for building data driven applications facilitates the development of an interactive and intuitive user interface. A tool called Soda-Core helps to generate data quality checks and run these on the datasets. It also captures the count of records passing or failing the checks. The application utilizes Snowflake cloud database to store all the information that is generated in the Data Quality assessment process with the consent of the data owner. The application involves various modules to achieve Data Ingestion, Data Profiling, Data Quality Rule creation, Data Quality Assessment Result Storage and Data Grading.

The expected output of the project will be that it will enable people with limited technical knowledge to define Data Quality rules. The application will minimise the manual effort put in the Data Quality assessment process via automated data profiling and rule validation. It will enable in faster identification and diagnosis of data quality issues and promote data reliability. Overall, the application will lead to improved data quality and henceforth lead to better data driven insights.

**Shrivatsa A Shetty Dr.Deepika K**

**1RV22MC089 Assistant Professor**