Department of Master of Computer Applications

25-06-2024

IV Semester
MCA491P – Major Project

Enhancement of Data Quality and Development of Assessment Application

Data Quality is a major concern for organizations across various domains. Reliable data is essential for informed decision-making and efficient operations. 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 will introduce a Data Quality Assessment Application, a user-friendly tool designed to address these limitations. The application will provide a guided GUI for users to assess the quality of their datasets. It will simplify the process of Data Quality Assessment (DQA) by automating several key steps involved in a data quality framework to a large extent.

The application will utilize a combination of modular and object-oriented programming techniques in Python at the backend. Streamlit, a Python framework for building data driven applications will be utilized for the development of an interactive and intuitive user interface. A tool called Soda-Core will help in generation of Data Contracts with Data Quality Checks along with capturing the results of the checks. The application will utilize Snowflake cloud database to store all the information that will be generated in the Data Quality Assessment process with the consent of the data owner. The application will consist of various modules to achieve Data Ingestion, Data Profiling, generation of Data Quality Rule, storing the results of Data Quality Assessment and Data Grading.

The expected outcomes 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**