PROJECT REPORT

METRICSTICS (DELIVERABLE - 1)

Prepared By

Vishal Karmakar(40220935)
Abhishek Kanuganti(40224734)
Madhava Sai Kumar Karnati(40227757)
Dharamjeet Kaur(40227330)
Simranjeet Kaur(40232877)

Under the Guidance of
Prof. Pankaj Kamthan

Submitted at
CONCORDIA UNIVERSITY

Github:
`https://github.com/karmakarvishal/SOEN6611_TEAM_F`
By Team F

Contents

Chapter 1

About The Project

In the present era, where data plays a central role, the skill of effectively interpreting data is absolutely essential for making well-informed decisions. The project aims to create a cohesive set of interrelated artifacts that enable the measurement and analysis of various statistical metrics. Our foremost focus is on enhancing user engagement, ensuring satisfaction, and optimizing time efficiency as core principles in our system's development The combination of precision and efficiency in METRICSTICS will empower users to conduct data analysis with confidence, promoting data-driven decision-making.

Chapter 2

Problem 1: GQM

2.1 Goal-Question-Metric (GQM)

SMART Goal

Goal : "To develop a METRICSTICS system that can accurately compute descriptive statistics (minimum, maximum, mode, median, arithmetic mean, mean absolute deviation, and standard deviation) of a random dataset within 5% margin of error, with a processing time not exceeding 10 seconds".

SMART Principles

- **Specific :** The goal clearly specifies what needs to be achieved - developing a METRICSTICS system that computes specific descriptive statistics. It also mentions the acceptable margin of error (5%) and processing time (10 seconds).
- **Measurable :** The goal includes quantifiable measures of success. It can be measured by comparing the system's computed statistics to known or benchmark values, ensuring accuracy within a 5% margin of error, and by monitoring processing time.
- **Attainable :** While the goal is ambitious, it is realistic and achievable within the specified timeframe, assuming adequate resources, expertise, and technology are available.
- **Relevant (Realistic):** The goal is relevant to the project's purpose: creating a METRICSTICS system for computing descriptive statistics. It is aligned with the project's overall objectives and contributes to its success.
- **Timely:** The goal sets a clear timeframe for completion. This ensures the project has a sense of urgency and a well-defined endpoint.

2.2 Questions and Metrics

Question 1:

How can we ensure METRICSTICS accurately computes descriptive statistics with a margin of error within 5% for a wide range of random datasets?

Metrics:

1. Percentage of datasets for which descriptive statistics calculated by METRICSTICS fall within the specified 5% margin of error.
2. Frequency of software updates required to maintain or improve accuracy across diverse datasets.
3. User feedback on the accuracy of METRICSTICS-generated statistics for their specific datasets.

Mechanism:

1. Implement Robust Statistical Algorithms.
2. Continuous Testing and Validation.

Question 2:

What measures can be taken to optimize the processing time of METRICSTICS, ensuring it does not exceed 10 seconds for computing descriptive statistics?

Metrics:

1. Average processing time of METRICSTICS for datasets of varying sizes and complexities.
2. Comparison of processing times between METRICSTICS and similar descriptive statistics tools.
3. Analysis of processing time reduction achieved through algorithmic improvements and parallel processing techniques.

Mechanism:

1. Parallel Processing Implementation
2. Algorithm Optimization and Caching

Question 3:

How can we make METRICSTICS easy to use for people with minimal statistical knowledge, such as high school or college students?

Metrics:

1. User-Friendliness

Mechanism:

1. Make user interface improvements and provide clear tooltips or help features to enhance user-friendliness.

Question 4:

How can we ensure that METRICSTICS remains reliable despite varying data sizes and complexities?

Metrics:

1. Scalability and Robustness

Mechanism:

1. Test METRICSTICS with datasets of different sizes and complexities. Measure its performance and accuracy with both small and large datasets

Question 5:

What is the impact of data quality on the accuracy of computed statistics?

Metrics:

1. The frequency at which data cleaning or preprocessing operations are applied to the input data.

Mechanism:

1. Identify the specific data cleaning and preprocessing operations performed on the input data. These operations can include tasks such as removing duplicates, handling missing values, outlier detection, data format validation, and more.

2. Integrate monitoring or logging mechanisms within your METRICSTICS system to record when data cleaning or preprocessing operations are initiated and completed. This can be done through custom code or by leveraging existing data processing tools and frameworks.

Question 6:

How often does the system encounter errors or anomalies in the input data?

Metrics:

1. The percentage of input datasets or data points that trigger errors or anomalies during computation.

Mechanism:

1. Implement thorough input data validation checks to ensure that the data conforms to expected formats and ranges. This includes checking for missing values, outliers, or any data points that could lead to errors during computation.

2. Establish a comprehensive testing framework that includes various scenarios and edge cases for input data. Conduct regular testing to simulate different data conditions and identify potential issues before releasing updates. This can include unit testing, integration testing, and stress testing to assess system behavior under various conditions.

Question 7:

How often does the system require updates or maintenance to maintain accuracy and performance?

Metrics:

1. Maintenance and Update Frequency

Mechanism:

1. Maintain a version control system for your METRICSTICS software. Document all changes, updates, and enhancements in a change log. This log should capture details about what was modified, added, or fixed in each version.

2. Maintain detailed documentation of each maintenance or update event, including the reasons, actions taken, and outcomes. Communicate changes and updates to relevant stakeholders, both internally and externally.

Question 8:

How efficiently does the system utilize available resources?

Metrics:

1. Resource Utilization Efficiency

Mechanism:

1. Thresholds and Alerts: Define resource utilization thresholds that trigger alerts when exceeded. For example, if CPU utilization consistently exceeds 90%, an alert is triggered. These thresholds should be set based on system performance requirements.

2. Reporting and Review: Generate regular reports on resource utilization efficiency and review them with the system administrators and stakeholders. Discuss strategies for ongoing resource optimization.

Question 9:

Is there a continuous reduction in the real-time margin of error for computed descriptive statistics as the system processes data over time?

Metrics:

1. Real-Time Margin of Error Reduction

Mechanism:

1. Assess and quantify the gradual decrease in the margin of error as time elapses, allowing for a comprehensive understanding of how the degree of inaccuracy or deviation from the expected value diminishes over an extended period. This measurement provides valuable insights into the system's continuous improvement or decline in accuracy over the course of time.

2. Evaluate the alterations in the margin of error that occur over a specifically designated and defined period of time. This involves closely observing and recording the shifts or modifications in the degree of inaccuracy or deviation from the expected value within the predetermined time frame. This measurement method allows for a focused analysis of how the margin of error evolves and behaves over the chosen time interval, contributing to a more detailed understanding of temporal trends and fluctuations.

Question 10:

Is the system continuously benchmarking its performance against known or historical data in real-time?"

Metrics:

1. Real-Time Benchmarking Frequency

Mechanism:

1. This metric assesses how frequently the METRICSTICS system engages in real-time benchmarking against known or historical data.

2. Monitor and record the frequency with which the METRICSTICS system actively engages in benchmarking activities in real-time. This metric involves keeping a meticulous tally of how many instances the system initiates benchmarking processes to evaluate its present performance against established standards or historical data. A higher count signifies a proactive and vigilant approach to continually assessing and enhancing its accuracy and processing time, as benchmarking provides a valuable yardstick for gauging progress and identifying areas for improvement in real-time.

Chapter 3

Problem 2: Use Case Model

3.1 Use Case Diagram

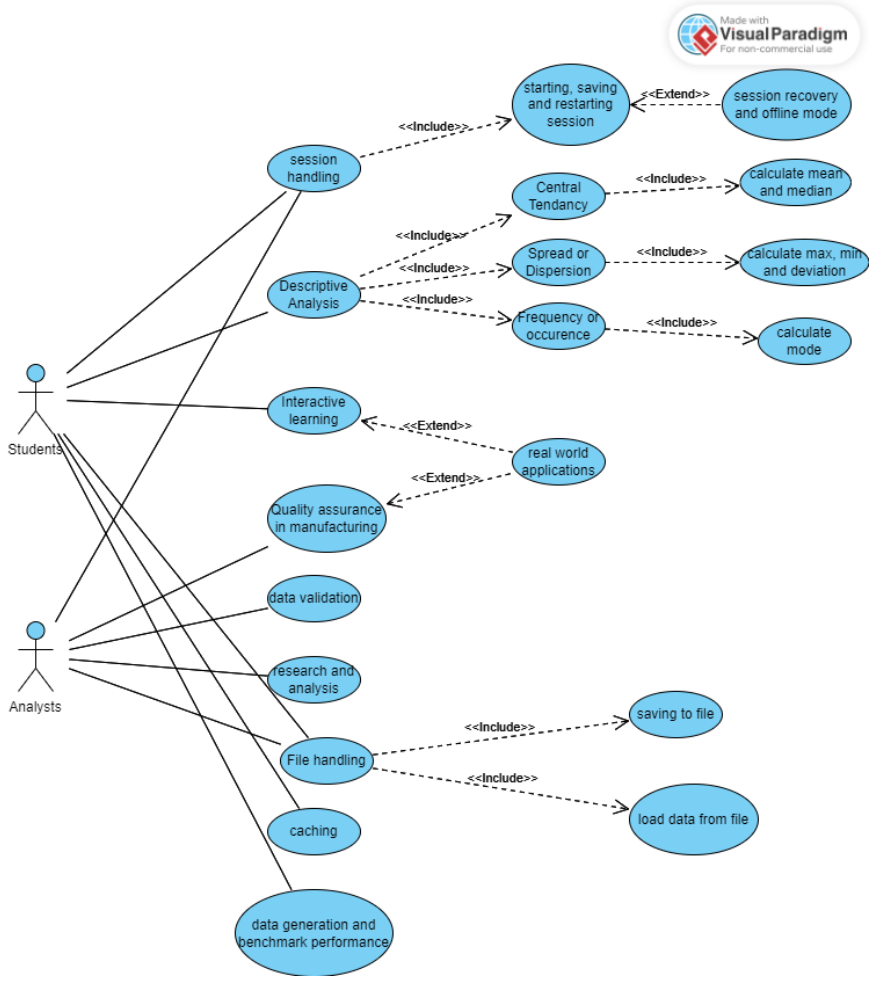

Figure 3.1: Use Case Model

3.2 Use Cases

3.2.1 Use Case-1

Use Case ID	UC-1
Use Case Name	Session handling
Primary Actors	Student
Priority	High
Description	The Session Handling use case is responsible for managing the user sessions within the Metristics system. It ensures that students can securely and efficiently interact with the system while performing statistical calculations on random datasets.
Pre-conditions	The Metristics system is operational and accessible.The Student (User) has successfully logged into the system
Post-conditions	The Student (User) can interact with the Metristics system.The session remains active until the Student (User) logs out or becomes inactive for an extended period.
Normal Flow	1 The Student accesses the Metristics system and provides valid login credentials. 2 The system validates the login credentials and authenticates the Student and logins to system. 3 If the Student wishes to log out, system terminates the session, clears session-related data, and returns the Student to the login screen.

3.2.2 Use Case-2

Use Case ID	UC-2
Use Case Name	Descriptive analysis
Primary Actors	Student
Priority	High
Description	The Descriptive Analysis use case is responsible for performing statistical analysis on a random dataset within the Metristics system. It calculates essential statistical measures.
Pre-conditions	The Student has successfully logged into the system. A random dataset is available for analysis
Post-conditions	The calculated statistical measures are displayed to the Student. The Student can use the generated statistics for further analysis or decision-making.
Normal Flow	1 The Student logs into the Metristics system and navigates to the Descriptive Analysis feature. 2 The system performs the following statistical calculations. 3 The system ensures that all calculations are accurate.

3.2.3 Use Case-3

Use Case ID	UC-3
Use Case Name	Interactive Learning
Primary Actors	Student
Priority	High
Description	The Interactive Exploration use case within the Metristics system provides Student Analysts with an immersive and interactive environment. This use case allows students to interact with practical examples, datasets, and simulations, fostering a deeper understanding of how statistical concepts are applied in real-world scenarios.
Pre-conditions	The Metristics system is operational and accessible. The Student has successfully logged into the system.
Post-conditions	Student Analysts have actively engaged in interactive exploration.
Normal Flow	1 The Student logs into the Metristics system and navigates to the Descriptive Analysis feature. 2 he system provides a rich set of interactive resources and tools.

3.2.4 Use Case-4

Use Case ID	UC-4
Use Case Name	Quality assurance in manufacturing
Primary Actors	Analysts
Priority	High
Description	The Quality Assurance use case within the Metristics system enables Data Analysts to ensure the accuracy, reliability, and quality of statistical analyses and reports generated within the system. It involves a systematic process of reviewing, validating, and verifying analysis results, data integrity, and adherence to best practices and standards.
Pre-conditions	The Data Analyst has successfully logged into the system. Statistical analyses or reports have been generated.
Post-conditions	Data Analysts have performed quality assurance checks and ensured the accuracy and reliability of analysis results. QA findings and documentation are available for review and auditing.
Normal Flow	1 The Data Analyst logs into the Metristics system. 2 The Data Analyst selects the statistical analysis or report that requires quality assurance. 3 The Data Analyst documents their findings, including any issues or concerns identified during the quality assurance process.

3.2.5 Use Case-5

Use Case ID	UC-5
Use Case Name	Data validation
Primary Actors	Analysts
Priority	High
Description	This use case empowers Data Analysts to assess data integrity and suitability for analysis, thereby enhancing the accuracy of statistical results and insights derived from the data.
Pre-conditions	Data Analysts have successfully logged into the system A dataset is available for validation
Post-conditions	Data Analysts have verified and validated the dataset for quality and accuracy. Identified data issues or discrepancies are documented and reported. The validated dataset is available for further statistical analysis
Normal Flow	1 A Data Analyst logs into the Metristics system using their credentials. 2 The system presents the Data Analyst with an option to access the Data Validation module. 3 The Data Analyst selects the dataset they want to validate from a list of available dataset

3.2.6 Use Case-6

Use Case ID	UC-6
Use Case Name	Research and analysis
Primary Actors	Analysts
Priority	High
Description	This use case provides the necessary tools and functionalities for data exploration, hypothesis testing, and statistical modeling, enabling analysts to discover trends, patterns, and correlations within the data.
Pre-conditions	Data Analysts have successfully logged into the system. A dataset is available for analysis.
Post-conditions	Data Analysts have conducted research and statistical analysis on the dataset. Insights and findings from the analysis are documented and available for reporting. Statistical results are accurate and reliable for decision-making.
Normal Flow	1 A Data Analyst logs into the Metristics system using their credentials. 2 The Data Analyst selects the dataset they want to analyze from a list of available datasets, the Data Analyst will initiate the validation . 3 Once the analysis is complete, the Data Analyst will generate reports.

3.2.7 Use Case-7

Use Case ID	UC-7
Use Case Name	File handling
Primary Actors	Analysts, Students
Priority	Medium
Description	The File Handling use case within the Metristics system allows Data Analysts to save data, reports, and analysis artifacts to files for storage, sharing, or future reference.
Pre-conditions	Data Analysts and students have successfully logged into the system. Data or analysis artifacts are available for export and file handling
Post-conditions	Data Analysts and students have successfully saved data or analysis artifacts to files. Saved files are stored in a designated location and are accessible for further use. Data Analysts may share saved files with team members or stakeholders as needed.
Normal Flow	1 A Data Analyst or student logs into the Metristics system using their credentials. 2 The Data Analyst selects the data or analysis artifacts they want to save to a file. This can include datasets, analysis reports, visualizations, or any other relevant information. 3 The Data Analyst and students confirms the file-saving operation.

3.2.8 Use Case-8

Use Case ID	UC-8
Use Case Name	Caching
Primary Actors	• Analysts
Priority	Medium
Description	The Caching use case in the Metristics system allows Data Analysts to improve the performance and efficiency of data retrieval and analysis by temporarily storing frequently accessed data or computation results in a cache.
Pre-conditions	• Data Analysts have successfully logged into the system. Data or computation results are available for caching.
Post-conditions	• Data Analysts have effectively cached data or computation results. Cached data is available for quick retrieval during subsequent analysis or operations.
Normal Flow	1 A Data Analyst logs into the Metristics system using their credentials. 2 The system presents the Data Analyst with an option to access the Caching module. 3 The Data Analyst selects the data or computation results they want to cache. This can include frequently used datasets, intermediate results of calculations, or other relevant data. 4 The Data Analyst specifies the cache duration. 5 The Data Analyst confirms the caching operation.

3.2.9 Use Case-9

Use Case ID	UC-9
Use Case Name	Benchmark Performance
Primary Actors	• Analysts
Priority	High
Description	The Benchmark Performance use case in the Metristics system enables Data Analysts to assess and compare the performance of different statistical models or algorithms when applied to the same dataset. This use case helps analysts identify the most efficient and accurate approach for their specific analysis needs, contributing to data-driven decision-making.
Pre-conditions	• Data Analysts have successfully logged into the system. Multiple statistical models or algorithms are available for benchmarking. A dataset is prepared and ready for benchmarking.
Post-conditions	• Data Analysts have benchmarked the performance of different statistical models or algorithms. Comparative results are available for analysis and model selection.
Normal Flow	1 A Data Analyst logs into the Metristics system using their credentials. 2 The Data Analyst configures benchmarking parameters, such as evaluation metrics, validation methods (e.g., cross-validation), and performance criteria (e.g., accuracy, precision, recall). 3 The system validates the benchmarking settings to ensure they comply with system guidelines and requirements.

References

1.GQM Approach
https://www.cs.umd.edu/users/mvz/handouts/gqm.pdf.

2. Use Case Diagram.,
https://www.visual-paradigm.com/guide/uml-unified-modeling-language/what is use case diagram/

3. Goal-Question-Metric (GQM) Approach for Software Quality Measurement[1]

4. Applying the Goal-Question-Metric (GQM) Framework to Improve Project Management[4]

5. Goal-Question-Metric (GQM): A Practical Introduction[2]

6. A Use Case Study for E-commerce Checkout[5]

7. Enhancing User Experience: A Mobile Banking Use Case[3]

8. Use Cases in Software Engineering[6]