SOEN 6611 - SOFTWARE MEASUREMENT: THEORY AND PRACTICE

Project Step 1 (1.5 points, due before midnight on July 11th)

Summer 2022

Course Instructor: Dr. Olga Ormandjieva

Source: SEI Implementing Goal-Driven Measurement course material (adapted).

Objective: Identify SMART (Specific, Measurable, Achievable, Realistic, and Timely) measurement goals and derive the corresponding questions

TEAM INFORMATION

Name	Student ID
Hasandeep Singh	40198687
Jasleen Kaur	40218417
Anushka Sharma	40159259
Kavleen Kour Sidhu	40205476

Step 1 tasks:

1.1 Business Goal(s): State the business goal(s) that are relevant to measurement. Label (each) business goal as BG# for traceability purposes

Business Goals	Goal Summary	Description
BG-01 (Volume)	Produce more than sufficient data (without being excessive).	• To specify and measure the information content of the datasets, to increase the volume of the big data • Data generated from high volume transactions

BG-02 (Velocity)	Maintain the real-time flow of information to make appropriate business decisions	• Transmitting bits of data at a near-constant rate • To stream fast-moving data into distributed storage and enhance the velocity of new data
BG-03 (Variety)	To better generalize the collective goal of the project (i.e. future prediction, classification, etc.)	• To process different kinds of data and enhance the variety of big data • Receiving end receives several types of data, including images, videos, and audio.
BG-04 (Vincularity)	Achieve overall accuracy of data for final service	• Improve connectivity and linkage of data to trace and make decisions • Improve our datasets by increasing the vincularity of the big data • Have accuracy of "things" in fusion
BG-05 (Veracity)	Attain correctness and relevancy of data to perform further procedures	• Big Data can have value only when its veracity can be established and thereby the information quality confirmed. • Estimating the quality parameter value and overall data accuracy in determining which characteristics are applicable to improve the veracity
BG-06 (Validity)	To maintain/achieve project goal and data output alignment	• Precision or reliability of the data used to derive the information • To check the authenticity of data of the datasets and improve its validity

1.2 Stakeholders & their measurement needs:

3.1 Identify the stakeholders.

3.2 Identify the stakeholders' measurement needs

Write a summary of the expected benefits from the use of the measurement results for the selected stakeholders, which should be a summary of the reasons for the measurement-related efforts.

Stakeholders	Measurement Needs	Benefits
User	Availability, credibility, and correctness	• Ensure the availability and reliability of the system • Ensure the validity of the data for the user • Ensure the correctness and relevancy of data
Strategy managers	Data correctness	• Verify the accuracy of the big data. • Evaluation of big data's improvements in volume, velocity, variety, veracity, vincularity, and validity. • Gain access to big data's added value and insights. • Evaluate data correctness and make improvements.
Big data managers	Manageability and performability	• Compare the quantity, speed, variety, truth, vinculity, and validity of big data. • Describe the growth in precision, correctness, insight gained, linking of big data, quality of big data, pace of big data accumulation, and amount of big data over various time periods of the big data life cycle. • Assess the change ramification and how to deal with it, with respect to project goals.
Data Engineers	Verifiability and functionality	• The data would be verified and processed, thus will possess the confirmation of verification,

		as someone is responsible for it. • The general or final goal is maintained and the functionality to be, is achieved.

1.3 Measurement Goals : Derive the measurement goals from the above 6 measurement needs

Measurement Goal Label:	Description	Corresponding business goal (write its label)
MG1	Compare the change in volume of big data at different time intervals. The volume of preprocessed dataset must be greater than threshold value of 60% to generate and predict results.	**BG-01**(Volume)
MG2	Despite changes to the dataset, the goal is to process the dataset without many changes in the existing infrastructure and the system must be at least scaled up to 5 to 6 replicas under increasing velocity.	**BG-02** (Velocity)
MG3	To have a dataset diversity between 10% to 12% with for unbiased dataset. Characterize the different types of big data gathered.	**BG-03** (Variety)
MG4	To determine the vincularity of the pipeline for linkage and connectivity. The system must ensure a traceability of or about 0.1 value.	**BG-04** (Vincularity)
MG5	Compare veracity and check if the data is correct and relevant to the final goal. Evaluate current ness and system must ensure data more than 10 to 15 years of age is eliminated.	**BG-05** (Veracity)
MG6	Evaluate the credibility of the dataset which can be 0 or 1, If 1 then the dataset is credible. This value can be based on factors such as rating, scores.	**BG-06** (Validity)

1.4 Questions:

For each of the above measurement goals, derive questions that the stakeholder in the selected above role might have to ask, and whose questions would be answered by the measurement results.

Important: the questions must be formulated in quantifiable way that they can be answered with indicators. Above all, you should avoid closed questions (i.e. yes or no answers).

Use the table below:

Question Label	Description	Corresponding measurement goal (label)
Q1	How many distinct records are there in the retrieved dataset which is obtained by dividing big data in that of different time frames?	**MG1** (Volume)
Q2	How rapidly is the volume of big data growing among the time frames?	**MG2** (Velocity)
Q3	What are the different classifications the data can be classified into? Do we have dataset varying and diverse by at least 10 to 12% making it unbiased?	**MG3** (Variety)
Q4	To what extent can the datasets or processed data be interrelated or traced to the reliable source?	**MG4** (Vincularity)
Q5	What is the increase in accuracy and availability of the dataset ensuring growth of veracity in different time intervals?	**MG5** (Veracity)
Q6	Is the percentage of credibility and compliant records in the dataset greater than 80%? Is our data adhering to the standards of ISO like decimal values should maintain '.' format?	**MG6** (Validity)