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

**SOEN6611/S22 Project Step 1 (1.5 points, due before midnight on July 11th)**

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

**Step 1 tasks:**

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

BG1: To improve the performance of Machine Learning models to provide relevant insights.

BG2: To analyze the performance of the system over time for varying data loads to mitigate issues related to data handling and processing.

BG3: To improve the decision making for improving internal operations processes.

BG4: To ensure that data is accurate and collected from reliable sources in order to stay compliant with organization and international data policies.

BG5: To better understand the linkage between datasets and derive the additional insights using the linkage information

BG6: To ensure that data is available when required and is accurate and complete for making accurate decisions.

* 1. **Stakeholders & their measurement needs:**

1.2.1 Identify the stakeholders.

1. Project Manager
2. Developers
3. Data Scientist
4. Data Engineer
5. End User

1.2.2 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.

1. Volume- Increasing the size and amount of big data that an organization analyzes and manages.

2. Velocity – Improving speed at whlich data is being generated/processed or handled.

3. Variety - Increasing diversity data elements across dataset i.e. more diversity and range of different data types/formats.

4. Veracity – Analyzing and Improving the degree that which the data is precise, accurate, and trustworthy.

5. Validity – Improving the validity of big data in terms of its accuracy and correctness for usage.

6. Vincularity – Improving data set by increasing the traceability of data for use in the Machine learning algorithm.

| Stakeholder | Measurement need | Benefits |
| --- | --- | --- |
| Project Manager | measure data generated by different project teams | provide accurate decisions to stakeholders regarding the status of projects. |
| Developers | measure the quality of data for business purposes. | Based on the quality of data, developers can write the appropriate code for further processing of data. Example handling of null values or missing values. |
| Data Scientists | measure volume and quality of big data for getting useful insights from the data | * to improve accuracy the machine learning model * to support accurate decision making for management. |
| Data Engineer | measure velocity, volume and quality of big data | to improve the data pipeline for efficient data handling and processing from different data sources. |
| End User | measure the accuracy and validity of data provided by the service provider | to enable better decision making and negotiations with the service provider |

* 1. **Measurement Goals:** Derive the measurement goals from the above 6 measurement needs

*Use the table below*

| **Measurement Goal Label:** | **Description** | **Corresponding business goal (write its label)** |
| --- | --- | --- |
| MG1 | Volume refers to vast amount of data generated by the world. It is evaluated by measuring the number of information bits across all records required to specify the information content of multiple datasets | BG1 - the more the data the better the machine learning model will perform. |
| MG2 | Velocity refers to the speed of processing of data in any form of handling, recording and publishing of data. It also refers to the speed at which data is being generated. It is evaluated by the measuring the relative growth of data set over time. | BG2 - system should be able to handle data loads of varying speeds. |
| MG3 | Variety refers to the ever-increasing different forms that data can come in. This includes formats such as text, images, sound, videos, 3d models, and much more. | BG3 - different types of data are collected from multiple sources to make decisions. |
| MG4 | Validity of Big Data refers to accuracy and correctness for the purpose of usage. | BG4 - ensure that data is collected from reliable sources |
| MG5 | Vincularity refers to the connectivity or linkage of data and to determine the traceability of data attributes to its sources. | BG5 - To better understand the linkage between datasets and derive the additional insights using the linkage information |
| MG6 | Veracity refers to the degree that data is accurate, trusted and precise. It is not only the accuracy of the data itself but the trustworthiness of the data source, type, and processing of it | BG6 - ensure that data is collected from trusted source and available when required to make accurate business decision |

* 1. **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 big is the data volume? | MG1: Volume |
| **Q2** | What is the velocity of big data? | MG2: Velocity |
| **Q3** | How diverse is the big data? | MG3: Variety |
| **Q4** | What is the validity of big data? | MG4: Validity |
| **Q5** | What is the current level of traceability of big data? | MG5: Vincularity |
| **Q6** | What is the quality and accuracy of big data? | MG6: Veracity |