SOEN 6611 - SOFTWARE MEASUREMENT: THEORY AND PRACTICE

Project Report on Task 1 and Task 2

WINTER 2022

Course Instructor: Dr. Olga Ormandjieva

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

TEAM 5	
Student #	**Name**
40163582	Mohammod Suhel Firdus
40157109	Vivekananda Reddy Gottam
40184906	Saswati Chowdhury
40156971	Milesh Kotadia

Table Of Contents

PROJECT STEP 1

SOEN6611/W22 Project Step 1 (3 points, due on March 19th): Identify SMART (Specific, Measurable, Achievable, Realistic, and Timely) measurement goals and derive the corresponding questions

1. Organisational Goals Related to Measurement:

Step 1 summary:

Clarify the business goals. Write a summary of the expected benefits from the use of the measurement results, which should be a summary of the reasons for the measurement-related efforts. Define the measurement goals.

1. **Clarify the business goals that are relevant to measurement (Hint: Big Data quality)**

Solution:

Business Goal: To Select a big data dataset that has the required quality which can improve the decision-making process of the organisation. The quality of the big data set can be achieved by achieving individual quality at each level of six Big data V's as described below:

Business sub-goal Label:	Goal Summary	Description
BSG-01	Increasing the **Volum**e of big data sets	Select a volume of the dataset which is easily sourced, and can be used for an ML algorithm.
BSG-02	Enhancing **Variety** in Big Data	To make better decisions, keeping a track of the multitude of data that comes over time
BSG-03	Accelerate the Big Data set **Velocity**	The availability of new and latest generated big data dataset and the frequency at which the new dataset

		is available. Also, to optimise the speed of newly available datasets by the ML Model.
BSG-04	Enhancing **Veracity** of Big Data set	Filtering the growing data in order to remove the unnecessary data and process the important ones to improve the quality of big data dataset.
BSG-05	Validate Big Date set **Validity**	To select a Big data dataset that has valid data for analysing with comparison to other datasets
BSG-06	Continuously monitoring Big Data set **Vincularity**	To improve the connectivity or relation between datasets by carefully examining and selecting which can improve the Vincularity of the Big Data Set which ultimately contributes to the overall throughput of ML algorithm.

2. Measurement stakeholders for the organisation:

2. Stakeholder & measurement needs

2.1 Identify the measurement stakeholders for the organisation.

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.

Solution:

2.1 Organisational Stakeholders

Stakeholder	Description
Product Owner/Project Manager	Person responsible for the entire ML algorithm software/product/code

Developers/Data Scientist	Person who develops the machine learning algorithm/Trains/evaluates the algorithm
Testers/QA	Person who validates the functionality and ensures the product meets the requirements
Sales and Marketing Team	People who are responsible to see how to sell the product and to market the product to attract customers
End Users	People who use the algorithm or the product post-release

2.2 Stakeholders' measurement needs

StakeHolder	**Measurement Need**
Product Owner/Project Manager	Product Owners are responsible for ensuring their companies' data projects deliver a value-adding outcome. Resource Allocation: No of people required for the completion of the project, hardware and other tools requirement
Developers/Data Scientist	Plan individual work effort. Selecting appropriate Programming language, algorithm Strategy for training, re-training, and valuation
Testers/QA	To find the errors/bugs in ML code. To check the behaviour of the algorithm on different sizes of the testing sets.
Sales and Marketing Team	Their job is to design advertising campaigns as well as plan to promote the product.
End Users	To get an enhanced experience of using the software/application

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

3. Define the measurement goals

3. Define the measurement goals (Hint : the 6V's)

Solution:

Measurement Goal Label:	Description	Corresponding business goal (label)
MG1	Volume refers to the amount of data that exists for processing the dataset. To select a dataset with a sufficient volume of records such that a decisive analytical application can be built using the same.	**Volume**
MG2	Despite the fact that the dataset changes constantly, the goal is to quickly acquire that dataset and process it without much difficulty using the existing infrastructure the organisation already has.	**Velocity**
MG3	Selecting the dataset which will contain the data specific to the requirement will be useful in improving the validity of the dataset instead of selecting a dataset that has more invalid or unrelated data.	**Validity**
MG4	Veracity increases the level of trust and authenticity for the collected data by removing the incomplete ones such that a suitable customer experience can be derived.	**Veracity**
MG5	Big data sets that have records that are of similar nature and can be used for comparison can be used for analytics. The better-connected data gives better input for the next process than data records that have no similarity and cannot be compared with each other.	**Vincularity**

MG6	To select the dataset with different types (but same genre) of records helping to categorise and segregate data. Datasets may be of different formats, but as for our interest, we should differentiate the same and select the format which yields easier procurement and maintains the quality.	**Variety**

4. Questions

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 such a way that they can be answered with measures or indicators. Above all, you should avoid closed questions (i.e. yes or no answers).

Solution:

For GQM analysis, we have selected 2 stakeholders viz - The product owner and the Developer. The Questions with respect to each stakeholder are listed below:

4.1 Product Owner

Question Label	Description	Corresponding measurement goal (label)
PO-1	Is the Volume of the big data set sufficient enough for developing a model? How many records are available in the data set?	**MG1**
PO-2	Is the dataset frequently changing? How easy is it to source a newly changed dataset?	**MG2**
PO-3	Are we selecting the right data source for our business requirement?	**MG3**

PO-4	What is current customer satisfaction? Is this Dataset going to impact customer satisfaction?	**MG4**
PO-5	Does the selected dataset have connected/related data required for developing a product?	**MG5**
PO-6	Does the dataset contain any incomplete information? What will be the level of complexity to remove that data?	**MG6**

4.2 Developer

Question Label	Description	Corresponding measurement goal (label)
D-1	Do we have pre-processed data in a recognizable format? Is there sufficient data for processing? Is it too huge to process and current programming techniques can handle the same?	**MG1**
D-2	Does the algorithm consider the new stream of data? Is it synchronised or not? Do we need to adjust our code for a changed dataset?	**MG2**
D-3	To what extent the dataset is valid? Is it meaningful data which can be used as input for ML?	**MG3**
D-4	Do we have all the correct data information? Does the collected dataset lead us to get insightful information? Which method will be effective for removing outliers?	**MG4**
D-5	In what way can results be evaluated when the processed data is compared with the other data?	**MG5**
D-6	Is the dataset organised? How to categorise the different forms of data?	**MG6**

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

PROJECT STEP2

SOEN6611/W22 Project Step 2 (3 points, due on March 19th): Operationalize Goals, Derive Success Criteria, and Indicators

1. Step 2 - Part 1: Operationalized Goals

The objective of the first part of this step is to express your measurement goal derived in Step 1, in a structured statement that identifies the object, purpose, quality focus & perspective, environment, and constraints. This information is typically needed to gain required insight and/or to enhance decision-making. The objective of the second part is to develop success criteria and success indicators that will allow you to answer the measurement questions (from Step 1) quantitatively and then communicate the results to others.

Step 2 Part 1 will be evaluated as the following:

- **The purpose is clear and specific**
- **The purpose is consistent with the perspective**
- **The perspective is the perspective of the "group/role"**
- **The purpose and quality focus applies realistically**
- **The object of interest represents some underlying model and attributes**

Operationalized Goals

For each measurement goal you documented in Step 1:
• Operationalize it as a structured statement that identifies the object, purpose, quality focus & perspective, and environment and constraints.
• Document your results using the template below

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

Solution:

Part 1: Operationalized Goals

Operationalized Goal : Label and description	**MG1** Select the dataset with high-quality Volume.
Corresponding Measurement Goal label	**Volume**
Object of interest	Big dataset
Purpose	Analyse the high volume of dataset in order to compare the result at every stage over time to fulfil the organisational goals..
Quality Focus, Perspective	Examine the **size of the dataset** and select the one which is suitable for the ML algorithm from the point of view of **Project Manager.**
Environment and Constraints	We may need extra environmental support (like hardware/expertise of Data scientist/ tools) based on the Volume of the dataset. There are multiple sources of datasets and selecting one with the right volume which contains quality data can be challenging. **Factors and parameters** ● **Application factors :** Capacity of the Software processing the data ● **Customer factors :** To produce better customer experience ● **People factors :** Data Scientists, Developers using the data for analysis ● **Methods :** Comparing size of Big data set ● **Resource factors :** Availability of various datasets ● **Tools :** Size calculation tool ● **Process factors:** Quality assessment Process ● **Constraints** : Availability of datasets of various volumes

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

Operationalized Goal : Label and description	**MG2** Improve the decision by frequently gathering data and processing it faster.
Corresponding Measurement Goal label	**Velocity**
Object of interest	Big dataset / source of the dataset
Purpose	Analysing the flow of Bigdata in order to improve the speed at which new insights are gathered over the time.
Quality Focus, Perspective	Examine the expansion and **changes in the dataset** from the perspective of the **Developer**.
Environment and Constraints	The newly gathered data needs to be monitored constantly. The new incoming data must be consistent with respect to the existing data. **Factors and parameters** ● **Application factors :** Software processing the dataset should be flexible for processing fast changing data ● **Customer factors :** Quality and efficiency ● **People factors :** Data Scientists, Developers using the data for analysis ● **Methods :** agility of the project management process ● **Resource factors :** Mindset of people to accept the change ● **Tools :** Data processing tool ● **Process factors:** Time to gather data ● **Constraints** : Managing versions of different dataset for rollback capability

Operationalized Goal: Label and description	**MG3** Improving the quality by checking the correctness of big dataset.
Corresponding Measurement Goal label	**Validity**

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

Object of interest	Big dataset
Purpose	Evaluating the correctness of the dataset in order to predict whether it is reliable for the system or not.
Quality Focus, Perspective	Examine the **quality and behaviour** of the dataset from the point of a **Data scientist.**
Environment, and Constraints	The dataset should contain only that information which fulfils the system requirements. The truthfulness of the data can be examined by tools, expertise of data scientists. **Factors and parameters** ● **Application factors :** Data types of the records in the dataset ● **Customer factors :** Information correctness ● **People factors :** Data scientist/Developers ● **Methods :** Algorithm to check data correctness ● **Resource factors :** Expertise/ perspective of the person who is validating the data ● **Tools :** Validation tools/ Data validation algorithms ● **Process factors:** Quality standard benchmark ● **Constraints** : Separate process to examine the validity of the dataset is required

Operationalized Goal: Label and description	**MG4** Improving the veracity by analysing and processing only that data which is specific to need and removing unwanted data.
Corresponding Measurement Goal label	**Veracity**
Object of interest	Big dataset
Purpose	Evaluating the dataset by cleaning it in order to make it more efficient and improving the quality.

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

Quality Focus, Perspective	Examine the **defects in the dataset** to improve the genuineness of it from the point of view of **Tester**.
Environment and Constraints	The incorrect data should be distinguishable from the correct one. **Factors and parameters** ● **Application factors :** Remove the outliers from the dataset ● **Customer factors :** Dataset cleaning ● **People factors :** Data Scientists, Developers using the data for analysis ● **Methods :** Analysing with statistical technique ● **Resource factors :** Person who has expertise in analysing quality of data ● **Tools :** Data cleaning tools ● **Process factors:** Turn around time ● **Constraints :** Correctly identifying the right valid data

Operationalized Goal : Label and description	**MG5** To improve the traceability and the relationship between data.
Corresponding Measurement Goal label	**Vincularity**
Object of interest	DataSet
Purpose	To improve the connectivity or relation between datasets by carefully examining and selecting which can improve the Vincularity of the Big Data Set which ultimately contributes to the overall throughput of ML algorithm.
Quality Focus, Perspective	The **flexibility to remove irrelevant data** (modifiability) to fine-tune the dataset from the point of view of **process improvement**.
Environment and Constraints	The data should be comparable to establish relationship **Factors and parameters** ● **Application factors :** Linking different datasets to make strong relations ● **Customer factors :** Data integration ● **People factors :** Data Scientists, Developers using the data for analysis

---

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

	● **Methods :** Standardization of data, degree of relationship between two dataset ● **Resource factors :** Existing data in dataset ● **Tools :** Data integration tools ● **Constraints** : Quality assessment of datasets

Operationalized Goal: Label and description	**MG6** The purpose of Variety is to classify and separate data through categorization and segmentation.
Corresponding Measurement Goal label	**Variety**
Object of interest	Dataset
Purpose	To select the dataset with different types (but same genre) of records helping to categorise and segregate data
Quality Focus, Perspective	Examine the **behaviour of the dataset** and select the distinct compatible data from the point of view of a **Data scientist**.
Environment and Constraints	Needs to be careful about the standardisation and distribution of the dataset. **Factors and parameters** ● **Application factors :** Capacity of processing different kind of dataset ● **Customer factors :** Variety in data ● **People factors :** Data Scientists, Developers using the data for analysis ● **Methods :** Methods to classify the data ● **Resource factors :** Person who operates the classification tool ● **Tools :** Data classification tools ● **Process factors:** Semantics of each type of data ● **Constraints** : Need adequate and consistent data

An extreme difference in the velocity of the dataset during different time frames indicates an outdated dataset.

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, "*Implementing Goal-Driven Measurement*".

Trudel & Ormandjieva (2020), adapted from the *Software Engineering Institute*, “*Implementing Goal-Driven Measurement*”.