Generating statistical Tables, Listings, and figures using machine learning, deep learning, and NLP techniques that provide more efficiency to the statistical programming process

Farha Feroze, and Ilango Ramanujam, Symbiance Inc;

ABSTRACT

Our in-house tool is a seamless solution to generate statistical TLG reporting in a structured and faster way with the help of AI that reduces manual effort significantly. The tool is built with an efficient framework for accelerating the TLG generation process in just three simple steps.

Generate TLG reports just by uploading TLG mock shell and ADaM datasets. AI engine will automatically annotate mock shell with ADaM variables about 70 to 80% for the first time and accuracy can be improved on subsequent study usages. Once the annotations are finalized by the user, the system will generate the reports (tables, listings, and figures) in minutes and provide the stand-alone program for each TLG.

The tool is easily configurable for any sponsor related TLG shell. The tool is trained by historical data. This paper will discuss one such algorithm of Machine Learning which was implemented as a tool.

INTRODUCTION

MACHINE LEARNING (ML)

Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data.

DEEP LEARNING (DL)

Deep learning (also known as deep structured learning) is part of machine learning methods based on artificial neural networks with representation learning.

NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing is a subfield of artificial intelligence concerned with the interactions between computers and human language. The target of NLP is to make a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them.

NEED OF ML, DL & NLP FOR TLG GENERATION

- To Extract table structure from TLG mockshell
- To Identify appropriate ADaM variable
- To identify required statistical methods for the given table
- To train the system from users while it encounters new input information every time
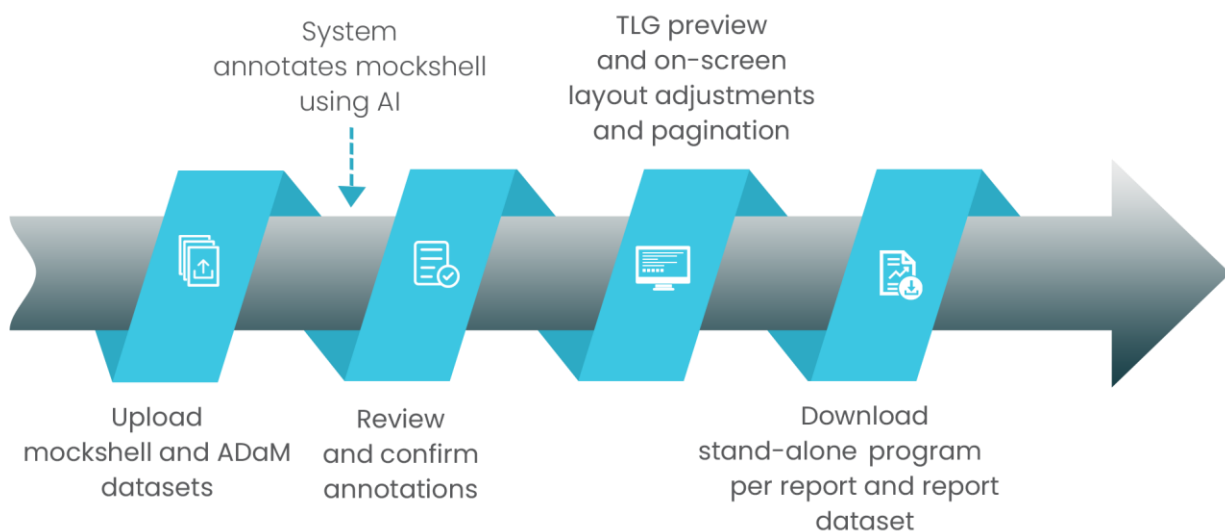- To re-use the learning and produce more accurate results

WHAT ARE WE TRYING TO SOLVE?

- Creating TLGs manually is time consuming
- It requires double programming for validation purpose
- TLG mock shell may have typos or human errors and it may require modification

BENEFITS OF USING MACHINE LEARNING, DEEP LEARNING AND NATURAL LANGUAGE PROCESSING IN TLG GENERATION?

- Significant reduction of programming efforts
- Easy adaption to sponsor's TLG mock shells
- Quick & Flexible to redo TLGs towards mock shell and ADaM data changes
- Downloadable Stand-alone programs without Macros

OVERALL PROCESS

System annotates mockshell using AI

TLG preview and on-screen layout adjustments and pagination

Upload mockshell and ADaM datasets

Review and confirm annotations

Download stand-alone program per report and report dataset

STEP 1 – UPLOAD TLG MOCK SHELL AND ADAM DATASETS

Tool provides an option to upload TLG mock shell and relevant ADaM Datasets.

STEP 2 – USER REVIEW & CONFORMANCE

System's AI engine will understand the uploaded TLG mock shell information and ADaM datasets and identify relationship between the ADaM variables and the information available in the mock shell. It automatically predicts ADaM variables and annotates the TLG mock shell based on the historical data. Around, 70 to 80% of the standard ADaM variables will be automatically annotated by the system for the first time. If there are any new variables to be annotated, then, the system has an option for the user to add the ADaM variable(s) annotation manually for the first time in the uploaded TLG mock shell tables. The AI engine will learn from the user, and it re-uses that information automatically from any other subsequent studies or tables. So, ADaM variable annotations previously done will be automatically annotated by the system in any subsequent studies or tables.

STEP 3 – TLG PREVIEW

Once all the annotations are completed, the system runs the TLG generation process and TLG preview can be seen by the user in minutes. The tool handles the following five different types of tables and mostly in any TLG mock shell, the following type constitutes 90 to 95% of the tables.

- Categorical
- Continuous
- Mixed (both continuous and categorical in a table)
- Change from baseline
- Shift table

Column Re-sizing, layout and pagination

The tool enables an option to resize the report column width on-screen in real time. Layout and pagination are adjusted while resizing column width. Those changes will be reflected in downloadable program.

DOWNLOAD

The tool provides the following two down-loadable files as an outcome:

- Stand-alone program per report without macros
- Report Datasets (before final output)

CONCLUSION

The AI techniques (ML, DL and NLP) can be utilized to read the unstructured data (document) such as TLG mock shell and identify correlation between ADaM variables and information in the mock shell. This can be utilized well with any sponsor or type of TLG mock shell. Once annotations are finalized, the AI engine produces stand-alone programs and actual TLGs.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Farha Feroze
Product Manager
Symbiance Inc
fferoze@symbiance.com

ilango Ramanujam
President and Chief Executive Officer
Symbiance Inc
iramanujam@symbiance.com