Effect of Commonsense on Data Analysis Output, Statistical and Machine Learning Views

Olebara Comfort Chinaza
Department of Computer Science
Faculty of Science
Imo State University,
Owerri, Imo State, Nigeria.
E-mail: chiiprime@gmail.com
ORCID ID: 0000-0002-5891-4206

ABSTRACT

The never-ending need to collect, manipulate, search for insights, and interpret data samples hasgreatly advanced research in many fields, such as computing, medicine, agriculture,manufacturing, and every aspect of human development. Data is the food on which Research andDevelopment thrive, and R and D is the live-wire of many successful world economies. The insatiable quest for data has led to the emergence of data repositories, where researchers dump data that has been utilized, and other researchers come requesting to reuse the data. Data can be generated or collected from the available repositories. The major challenge however is tailoring the available data to answer research questions that range from simple to critical and even life- dependent. This work aims to unravel the position of common sense, also known as perception, orsound judgment, in the process of data analysis. The role of commonsense at various levels of dataprocessing is highlighted through a presentation of various stages in the data processing life cycle,and the task definition of each stage. The results are: data analysis checklist that helps data analysts gain insights into the human-machine roles in a data analysis life cycle; and theconcretization of the need for a deeper commitment to the planning stage, where research questions and hypotheses require human intelligence in identifying solution variables, especially in cases where those solution variables were not predetermined before data collection. The researcher therefore concludes that: common-sense feeds data, and data feeds research.

Keywords: Effect, Commonsense, Data Analysis, Output, Statistical and Machine Learning, Perception.
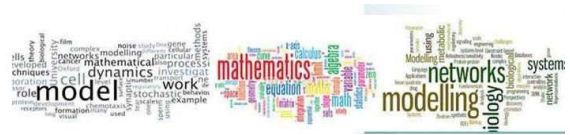
1. INTRODUCTION

This study tries to establish the role of data analysts in their analysis, in the era of automated data analysis environments such as SPSS, Tanagra, Machine Learning IDEs such as WEKA etc. It is common knowledge that statistical methods and algorithms are formula-based and models can be manually computed; hence the advent of automated data capturing, processing, analysis, and interpretation offer speed, higher accuracy, and less human effort. This does not, however, eliminate the vital role humans play in deciding what problems to solve, data analysis planning, and finding variables in the data whose interaction would yield an optimum result. This human participation requires thinking, intelligence and discretion. These faculties put together are the make-up of common sense in the context of this paper.

2. LITERATURE REVIEW

According to [1], data analysis and interpretation is the process of assigning meaning to information collected as well as determining the significance of the information, and its implications. "Assigning meaning" phrase here agrees with the opinion [2], which defines data analysis as having the capacity to view the world from the position of others and having critical thinking skills needed for interrogation and interpretation. What these two views have in common is their recognition of the role humans play in the data interrogation, analysis, insight gaining as well as interpretation of the result. In this study, the researcher aims to establish the role of human input in the automated fields of data analysis. According to [3] common sense is defined as sound and prudent judgment based on a simple perception. According to [4], Aristotle's account of perception states that it is part of human activities.

[4] in his study of Aristotle on common sense tried to establish a relationship between common sense, thinking, and intelligence. It could be stated, therefore, that the act of thinking, planning, data collection, as well as gaining deep insights, and making informed decisions with a carefully analyzed dataset requires perception, thinking, and intelligence. Common sense by [5] cites Aristotle's description of common sense, as the higher-order perception that humans have adding that the judging principle of humans coincides with his perception. [6] suggests that mining common sense from available data and using it in building intelligent systems will be the next trend in the computing world. [7] worked on developing common-sense services that construct computational models that mimic the cognitive capabilities of children.

Common sense and Logic
Common sense which has been defined as sound judgment that is based on the perception of a situation has to do with wisdom and is assumptive. Logic, on the other hand, can be defined as "science that deals with the principles and criteria used for validating inferences and is based on the calculation of collected facts [8]. In data analysis, however, these two characteristics are required at one point or another throughout the life cycle of the data analysis.

Common sense and data analysis life cycle
[9] outlined steps in the planning of a data analysis task. The researcher reiterates these steps and also separates tasks for common sense(human), and those for logic (machine automated). The authors in [9] suggest dividing pre-analysis tasks into two: Planning and Management. Data analysis planning stage has tasks such as:

Data Planning tasks

Generating research question(s) and/or hypothesis: The researcher after finding a problem that he wishes to research, asks a question whose answer would either bring solution to the discovered problem or provide insight towards proffering a solution. The hypothesis proposes or suggests a testable outcome to the research question using an alternative or null hypothesis, wherethe null(H0) is the negation of the alternative hypothesis(H1). Decision variables are developed here. Independent variables are variables that cause a change in the system while dependent variables are those affected by independent variables.

Role of common sense:

This stage is purely human effort. The researcher decides on a problem to solve and generates theappropriate question/hypothesis

Select/access/build a dataset

This involves deciding the source of data to be analyzed. The researcher may take advantage of numerous free datasets online from sites like Kaggle, Google dataset search data, Data.Gov, UCImachine learning repository, CERN Open data portal, and many more. Another option would be to design a questionnaire and collect responses from a target group.

Role of common sense:

If a researcher decides to use dataset from online source, s/he would only have to search for a dataset that meets research criteria even if not completely. The other option involves designing thequestionnaire with the required questions.

List inclusion/exclusion

This stage involves removing contents of the acquired dataset that are not required in research andreworking the dataset if necessary.

Role of common sense:

The researcher's action at this level to a great extent determines successful analysis of the data. This is because choosing the right variables and determining what insights to search for will helpthe researcher present appropriate questions or choose the right criteria from acquired data.

Choosing the right variable for analysis

The decision variables (Independent and dependent variables) are selected from the dataset and anappropriate statistical method is chosen and applied.

Role of common sense

It is human activity to pick out independent and dependent variables, and this must be carried outwith utmost care, as failure to select the right decision variables leads to obtaining the wrong outcome, thereby making the wrong decisions. Choosing the right statistical method depends on the type of data collected: ordinal, nominal, interval, etc. However, the robustness of statistical methods could be leveraged upon to gain the required insight. According to [10]
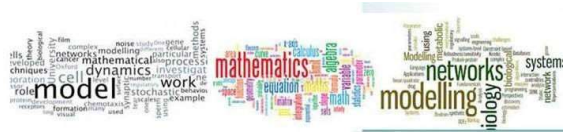
it is better to generate multi-item questions for determining a single variable. Nested or follow-upquestions also provide deeper understanding and insight. Likert scale is ordinal type data. However, when a variable has multi-item Likert questions, a composite of the scores is calculatedand the data changes from ordinal form to interval form. According to [11] and [12], two schoolsof thought exist on the issue of Likert scale data should be analyzed. While the first believe that itshould remain ordinal with nonparametric analysis methods, the second agree to leverage on the robustness of statistical methods, allowing the researcher to compute composite, categorizing using the mean, all depending on the insight the researcher wishes to gain. Computing the data inthis way allows parametric analysis methods to be performed on data as they are changed from ordinal to interval cited in [13].

Data Coding
Data coding is a vital part of data preparation. Scaled data are either positively or negatively worded with the same scale format given. It involves changing textual data into numeric equivalents that represent its presence, order, or category.

Using common sense on coding and reverse-coding of data.
It is important to code and reverse code data items that aim to achieve a common goal. For instance:"I spend more time on social media daily than on reading" with scale: 1: Strongly disagree, 2: Disagree, 3: Neutral, 4: Agree 5: Strongly Agree. In the stated case, the question is negatively worded, hence people who "Totally Disagree" should score higher than those who agree. To further buttress this, if a variable has multi-item questions, some positively worded and others negativelyworded, it is important to reverse code them, so each score will reflect the word pattern used [14].

Data Management
This involves such tasks as:

Creating Dataset to work with
Coding may be carried out in a spreadsheet such as excel, but the dataset should be imported intoa data analysis software such as SPSS or loaded into a machine learning environment.

Clean the data
Data cleaning involves the removal of missing numbers and unrequired data.

### Create an analysis file/Analyze
This involves appropriate variable naming, computation of composites, value labeling, dummy variables encoding, and all preps before data analysis proper. After these preparations, the actual analysis is carried out.

Data Interpretation
The interpretation of analyzed data follows the outcome of the analysis. These could be in percentages or numbers whose level is used to gain insight into the performance of the decision variables.

Role of common sense in data management
Tasks within the data management and data analysis are out-of-the-book tasks and do not requirethe researcher's perceptive. Once an appropriate data analysis method has been selected, and assumptions confirmed, the analysis process in SPSS, machine learning, or any software is performed using statistical algorithms that encapsulate the details and offer results that are used togain desired insight.
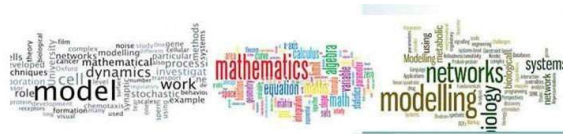
3. SUMMARY

Successfully carrying out any data analysis task begins with a researcher conceiving an idea, encountering a challenge, or seeking to fill a research gap. This is followed by developing research question(s) and hypothesis, planning which includes tasks such as dataset selection or generation, data cleaning, analysis variable selection, method/algorithm selection, actual data analysis, and finally the interpretation.

4. CONCLUSION AND RECOMMENDATION

After a research idea has been conceived, human common sense drives the initial phases of the solution-seeking until all inputs have been provisioned, then machine receives the inputs, perform chosen algorithm with human assistance, and produces outputs with which humans can gain insights, and make informed decisions. The researcher, therefore, recommends that analysts pay more attention to the tasks outlined as requiring common sense, as their depth and quality influences the accuracy of the final outcome.

Endnote/Copyright Note
A copy of this paper was published in the Journal of Technological and Academic Research Vol .2, 141

REFERENCES

[1] "Analyzing and Interpreting Data - Office of Institutional Research – Syracuse University." https://institutionalresearch.syr.edu/assessment/asesspp/analyzing-and-interpreting-data/ (accessed Mar. 04, 2019).

[2] P. Bazeley, "Interpreting Data," Integr. Anal. Mix. Methods Res., pp. 54–68, Dec. 2018, doi: 10.4135/9781526417190.N3.

[3] "Common sense Definition & Meaning - Merriam-Webster." https://www.merriam-webster.com/dictionary/common sense (accessed Mar. 04, 2019).

[4] P. Gregoric, "Aristotle on the Common Sense," Aristotle on the Common Sense, pp. 1–272, 2007, doi: 10.1093/acprof:oso/9780199277377.001.0001.

[5] "Ccommon sense | The Chicago School of Media Theory." https://lucian.uchicago.edu/blogs/mediatheory/keywords/common-sense/ (accessed Mar. 04, 2019).

[6] N. Tandon, A. S. Varde, and G. De Melo, "Commonsense Knowledge in Machine Intelligence."

[7] D. Gunning, "Machine Common Sense Concept Paper," 2018.

[8] "Differences Between Logic and Common Sense (Spoiler Alert: We Need Both)." https://msmarketinstudio.wordpress.com/2019/10/17/differences-between-logic-and- common-sense-spoiler-alert-we-need-both/ (accessed Mar. 05, 2019).

[9] "Creating an AnalysisPlan." https://www.cdc.gov/globalhealth/healthprotection/fetp/training_modules/9/creating-analysis-plan_pw_final_09242013.pdf Accessed: 2019-03-04

[10] A Technology Acceptance Model for Empirically Testing New End-User Information Systems."https://www.researchgate.net/publication/35465050_A_Technology_Acceptance_Model_for_Empirically_Testing_New_End-User_Information_Systems (accessed Mar. 11, 2019).

[11]

[12]   R. Dwivedi and N. N. Pandey, "Analysis of Evaluation Data Collected on Likert Type Items:Humanities-Courses," J. Data Anal. Inf. Process., vol. 09, no. 02, pp. 102–121, 2019, doi:10.4236/JDAIP.2019.92007.

[13]   J. Carifio and R. J. Perla, "Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban

[14]   Legends about Likert Scales and Likert Response Formats and their
Antidotes," J.Soc.   Sci.,   Vol.   3,   No. 3,   Pp.   106–116,Mar.   2007,
doi:10.3844/JSSP.2007.106.116.

[15]   A. Joshi, S. Kale, S. Chandel, and D. K. Pal, "Article no.BJAST.2015.157 Opinion ArticleJoshi et al," vol. BJAST, no. 4, p. 157, 2015, doi: 10.9734/BJAST/2015/14975.

[16]   "An   Easy   Way   to   Reverse   Code   Scale   items   -   The   AnalysisFactor."
https://www.theanalysisfactor.com/easy-reverse-code/ (accessed Mar. 11, 2019).