Research title: ***Application of Machine Learning in Tracking Food Insecurity***

Abstract:

The research was on contributing to a body of knowledge in four ways in a context of applying Machine Learning to track food insecurity from; phone data, tweets, and news feed as datasets.

While phone datasets such as Call Detail Records (CDRs), airtime credit purchases, and mobile transactions have been applied to gain some information on food insecurity, limited attention has been given on the use of phone conversations and phone messages for similar purpose. A Machine Learning strategy was proposed to similarly harness these unexplored datasets. The strategy was published in a journal for feedback and validation purpose. Future work points at subjecting the strategy with relevant datasets, an opportunity missed due to complication in accessing relevant datasets from telecommunication companies.

While tweets from a community have been applied to gain insights on food insecurity, this strategy hits a snag for a low tweeting community. This study explored use of relevant tweets from both internal and external communities as an alternative for a low tweeting community. From the four classifiers (KNN, SVM, D-TREES, and N-BAYES) explored, D-TREES generated a superior performance with ROC-AUC measure of 0.84. Food insecurity trends based on D-TREES resulted into a correlation of 0.57 compared with ground truth data. This was comparable with correlation of 0.4 from previous related work. The strategy can be extended in studying other situations where local communities are not generating substantial tweets do to pattern extraction.

Limited studies have applied both similarity and polarity measures in a Machine Learning model to study food insecurity though close studies have applied polarity as an attempt. The study explored this in a context of classifying whether an article is on food insecurity or not in a news feed. Findings were promising; with N-BAYES and KNN generating ROC-AUC measures of 0.927 and 0.931 respectively better than random classifier performance of 0.91. Future work points at exploring the concept with deep learning as a sophisticated environment to enhance performance.

A strategy mimicking human cognition was modeled and blended with standard classifiers to enhance performance in classification and time complexity in a context of tracking food insecurity from news feed. KNN, SVM, N- BAYES, and D-TREES generated improvement in AUC-ROC measures ranging from 0.3% to 3% while in CNN and LSTM the measures retarded. On overall, CNN generated highest performance of AUC-ROC of 0.804 though without the strategy. Time complexities on all six classifiers improved by a factor at least 3 faster. Food insecurity trends generated with CCN were comparable with ground truth trends based on food prices at a correlation of 0.45. This was comparable with 0.4 correlation obtained with similar previous studies. Future work can consider venturing into: (1) examining any un-explored aspect(s) of the news feed, (2) alternatives options of modeling human cognition, and (3) un-explored algorithms.