

Comprehensive Evaluation of Indian Goods and Services Tax Rollout using Deep Learning and Machine Learning of Social Media Impressions and Information System Success Model

Pankaj Dixit

Abstract

The GST tax regime was rolled out in June 2017. Whenever such a new tax law is introduced, certain section of the users has satisfaction whereas others have grievances. They express their feelings on social media like Twitter and Facebook. This thesis aims at providing detailed analysis of the newly introduced GST system using machine learning and deep learning techniques in order to improve the system. The thesis also aims to evaluate the success of the GST system. Tweets by the taxpayers with respect to the GST regime and the system collected from the year June 2017 to May 2020 formed the database for this analysis. The entire analysis is in five different modules. In the first module sentiment analysis has been carried out using deep learning techniques viz. CNN, LSTM and bi-directional LSTM with different context-based embeddings. In the second module attention based deep learning model is used to show how the attention weights of the key words appearing in the tweets (related to GST) vary over the quarters in the three-year period. In the third module it has been attempted to find the importance of pairs of terms appearing in the tweets using a hybrid approach of co-occurrence graph and attention based deep learning technique. Co-occurrence graphs have been constructed pertaining to different categories in GST. The frequency of edges (in the co-occurrence graphs) between the key words and special terms given by GST was used to find the important pairs of terms. Attention weights for these important pairs have been evaluated using bidirectional LSTM model. Finally, a new measure called 'attention factor' has been coined, which gives the highly important pair of terms occurring with the keywords in the tweets. A helpdesk was established by GSTN to answer the queries posed by the taxpayers. In the fourth module automatic QA system has been developed by using clustering of embeddings of questions with context-based embeddings like BERT and ROBERTA and three choices of answers have been provided for each question. In the fifth module, the success of the GST e-governance information system has been evaluated using a modified DeLone and McLean information system success model. The reliability analysis along with convergent and divergent validity is carried out followed by structured equation modelling. Regression and path analysis is also performed to evaluate the success of the GST e-governance system.