World Academy of Science, Engineering and Technology
International Journal of Pharmacological and Pharmaceutical Sciences
Vol:14, No:05, 2020

Predicting Success and Failure in Drug Development Using Text Analysis

Authors : Zhi Hao Chow, Cian Mulligan, Jack Walsh, Antonio Garzon Vico, Dimitar Krastev

Abstract : Drug development is resource-intensive, time-consuming, and increasingly expensive with each developmental stage. The success rates of drug development are also relatively low, and the resources committed are wasted with each failed candidate. As such, a reliable method of predicting the success of drug development is in demand. The hypothesis was that some examples of failed drug candidates are pushed through developmental pipelines based on false confidence and may possess common linguistic features identifiable through sentiment analysis. Here, the concept of using text analysis to discover such features in research publications and investor reports as predictors of success was explored. R studios were used to perform text mining and lexicon-based sentiment analysis to identify affective phrases and determine their frequency in each document, then using SPSS to determine the relationship between our defined variables and the accuracy of predicting outcomes. A total of 161 publications were collected and categorised into 4 groups: (i) Cancer treatment, (ii) Neurodegenerative disease treatment, (iii) Vaccines, and (iv) Others (containing all other drugs that do not fit into the 3 categories). Text analysis was then performed on each document using 2 separate datasets (BING and AFINN) in R within the category of drugs to determine the frequency of positive or negative phrases in each document. A relative positivity and negativity value were then calculated by dividing the frequency of phrases with the word count of each document. Regression analysis was then performed with SPSS statistical software on each dataset (values from using BING or AFINN dataset during text analysis) using a random selection of 61 documents to construct a model. The remaining documents were then used to determine the predictive power of the models. Model constructed from BING predicts the outcome of drug performance in clinical trials with an overall percentage of 65.3%. AFINN model had a lower accuracy at predicting outcomes compared to the BING model at 62.5% but was not effective at predicting the failure of drugs in clinical trials. Overall, the study did not show significant efficacy of the model at predicting outcomes of drugs in development. Many improvements may need to be made to later iterations of the model to sufficiently increase the accuracy.

Keywords : data analysis, drug development, sentiment analysis, text-mining
Conference Title : ICDDDD 2020 : International Conference on Drug Development and Drug Design
Conference Location : Berlin, Germany
Conference Dates : May 21-22, 2020

Open Science Index, Pharmacological and Pharmaceutical Sciences Vol:14, No:05, 2020 publications.waset.org/abstracts/121298.pdf