

(12) PATENT APPLICATION PUBLICATION

(19) INDIA

(22) Date of filing of Application : 26/10/2021

(21) Application No. 202111048903 A

(43) Publication Date : 05/11/2021

(54) Title of the invention : A SYSTEM OF META-HEURISTIC MODEL OF FEATURE SELECTION FOR SENTIMENT ANALYSIS

(51) International classification : G06N0020000000, G06K0009620000, G06Q0030020000, G06F0040300000, G06N0005000000
(86) International Application No : NA
Filing Date : NA
(87) International Publication No : NA
(61) Patent of Addition to Application Number : NA
Filing Date : NA
(62) Divisional to Application Number : NA
Filing Date : NA

(71) Name of Applicant :

1) ARPITA

Address of Applicant : PH.D. SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS, KURUKSHETRA UNIVERSITY, KURUKSHETRA -----

2) Dr PARDEEP KUMAR
3) Dr KANWAL GARG

Name of Applicant : NA

Address of Applicant : NA

(72) Name of Inventor :

1) ARPITA

Address of Applicant : PH.D. SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS, KURUKSHETRA UNIVERSITY, KURUKSHETRA -----

2) Dr PARDEEP KUMAR

Address of Applicant : ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS, KURUKSHETRA UNIVERSITY, KURUKSHETRA -----

3) Dr KANWAL GARG

Address of Applicant : ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND APPLICATIONS, KURUKSHETRA UNIVERSITY, KURUKSHETRA -----

(57) Abstract :
A huge proportion of electronic data is produced with elevated accessibility to computerized information. Accumulation of this huge amount of data although arouses the problem of scalability, yet persuade researchers to critically analyze the data with aim of extracting utmost benefits in view of wiser decisiveness. This research focuses on superintending the problem of scalability for extrication of accurate sentiment interpretations from massive content over Twitter using supervised machine learning algorithm. Towards this end, first requirement is curtailing of text to a better structured format by pre-processing of data collected through Twitter Streaming API. For this research, raw data at step of pre-processing is filtered with fine sieve of two processes i.e. cleaning and transformation. Further, it was observed that feature extraction, dimensionality reduction and feature selection were three major phases of producing reduced set of attributes. But, all three had some limitations in tackling enormous set of features. Therefore, a hybrid meta-heuristic model collaborating extraction, reduction and selection is suggested as second phase of sentiment analysis in this paper. Thereafter, for third phase of experimentation, five supervised machine learning classifiers named NB, Random Forest, SVM, Decision Tree and (LR) Logistic Regression models were applied over three secondary training datasets. First being Mixed dataset of movie reviews and news, second being Airline dataset and the final one was dataset of Amazon product reviews. Results demonstrated 52.07%, 45.63% and 50.3% reduction in feature subset for Mixed, Airline and Amazon dataset respectively without compromising the accuracy. Ultimately, Support Vector Machine which is observed to be outperforming other four classifiers for all three datasets provides a scrutiny of sentiments over tweets related to Taliban Government.

No. of Pages : 28 No. of Claims : 8

Chairman
Deptt. of Computer Science
and Applications
K.U. Kurukshetra