Text Mining Based Prediction Model for Incident Occurrences in Steel Plant

Conference Paper · November 2016

CITATIONS
0

READS
58

4 authors, including:

Sobhan Sarkar
Indian Institute of Technology Kharagpur
20 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Irshad Ansari
Indian Institute of Technology Kharagpur
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Session WA06 - Text Mining I

1 - Text Mining Based Prediction Model for Incident Occurrences in Steel Plant

November 16, 2016, 8:00 - 9:30 AM | 102A-MCC

Authors

Sobhan Sarkar, Vishal Lakha, Irshad Ansari, Jhareswar Maiti, IIT, kharagpur, India. Contact: sobhan.sarkar@gmail.com

Abstract

The aim of this study is to provide the predictive solution using text mining and classification algorithms. Data on accident occurrences for a period of four years from a steel industry was collected. The outputs of text mining have been fed into four binary classification algorithms (SVM, k-NN, Random Forest, Maximum Entropy) which were tested further for evaluation of the best fit model to predict the accident classes (injury or property damage). The year-wise results showed that RF outperforms other classifiers with higher accuracy, i.e., 92.7%, 91.4%, and 91.8%, and area under curve values as 0.926, 0.910, and 0.912, respectively as tested with 10-fold cross validation for years 1, 3, and 4.