

## Automatic Lexicon Generation for Domain Specific Dataset for Mining Public Opinion on China Pakistan Economic Corridor

**Authors :** Tayyaba Azim, Bibi Amina

**Abstract :** The increase in the popularity of opinion mining with the rapid growth in the availability of social networks has attracted a lot of opportunities for research in the various domains of Sentiment Analysis and Natural Language Processing (NLP) using Artificial Intelligence approaches. The latest trend allows the public to actively use the internet for analyzing an individual's opinion and explore the effectiveness of published facts. The main theme of this research is to account the public opinion on the most crucial and extensively discussed development projects, China Pakistan Economic Corridor (CPEC), considered as a game changer due to its promise of bringing economic prosperity to the region. So far, to the best of our knowledge, the theme of CPEC has not been analyzed for sentiment determination through the ML approach. This research aims to demonstrate the use of ML approaches to spontaneously analyze the public sentiment on Twitter tweets particularly about CPEC. Support Vector Machine SVM is used for classification task classifying tweets into positive, negative and neutral classes. Word2vec and TF-IDF features are used with the SVM model, a comparison of the trained model on manually labelled tweets and automatically generated lexicon is performed. The contributions of this work are: Development of a sentiment analysis system for public tweets on CPEC subject, construction of an automatic generation of the lexicon of public tweets on CPEC, different themes are identified among tweets and sentiments are assigned to each theme. It is worth noting that the applications of web mining that empower e-democracy by improving political transparency and public participation in decision making via social media have not been explored and practised in Pakistan region on CPEC yet.

**Keywords :** machine learning, natural language processing, sentiment analysis, support vector machine, Word2vec

**Conference Title :** ICDM 2019 : International Conference on Data Mining

**Conference Location :** Dublin, Ireland

**Conference Dates :** July 29-30, 2019