

**Comments Mining With TF-IDF: The Inherent Bias and Its Removal**

**Abstract:**

Text mining have gained great momentum in recent years, with user-generated content becoming widely available. One key use is comment mining, with much attention being given to sentiment analysis and opinion mining. An essential step in the process of comment mining is text pre-processing; a step in which each linguistic term is assigned with a weight that commonly increases with its appearance in the studied text, yet is offset by the frequency of the term in the domain of interest. A common practice is to use the well-known tf-idf formula to compute these weights.