1. The article is headlined “Challenges of Feature Selection for Big Data Analytics”.
2. The article is written by students from Arizona State University. It was published the journal “Intelligent systems” in 2016.
3. The article is about some challenges of feature selection where part of them usually are ignored but should be considered. The authors provided open-source repository with algorithms that solve those challenges.
4. The authors write that because of high dimensional feature space there is a problem named as the curse of dimensionality. It is popular and long-stand problem that means it is impossible to see or understand some correlations or relationships between features. Therefore, it would be nice to have some solution that allows to us reduce the dimensionality. The article describes some ways to solve this kind of problem. According to the text, authors advice to categorize data on structured data, linked data, multi-source/multi-view data and streaming data. These categories help to find more right solution for better feature selection. Otherwise we intend to select irrelevant or redundant features that may be a cause of bad model prediction that use these ones. Further the authors provide open-source repository that help developers to solve the problem of feature selection for categories mentioned before. In conclusion, they also provide benchmark that gives an opportunity to make experiments and comparisons between different feature selection algorithms.
5. In my opinion the article is worth reading because it has useful solutions for the problem that is growing nowadays because of big informational world which already has a lot of data that cannot be analyzed by easy way.