Introduction:

Detecting misinformation on social media is an extremely important but also a technically challenging problem. Fake news on social media has experienced a resurgence of interest due to the recent political climate and the growing concern around its negative etc. Not only does it provide a source of spam in our lives, but fake news also has the potential to manipulate public perception and awareness in a major way. For example, in January 2017, a spokesman for the German government stated that they “are dealing with a phenomenon of a dimension that they have not seen before”, referring to the proliferation of fake news[1]. As a response, numerous articles and blogs have been worn to raise public awareness and provide tips on disorientating truth from falsehood. Here we are using Tf-Idf for extracting features in NLP and Tf-Ldf will give more accurate results in case of fake news detection.

Literature Survey:

The task of detecting fake news has undergone a variety of labels, from misinformation, to rumour, to spam. Just as each individual may have their own intuitive definition of such related concepts, each paper adopts its own definition of these words which conflicts or overlaps both with other terms and other papers. For this reason, we specify that the target of our study is detecting news content that is fabricated, that is fake. There has been a large body of work surrounding text analysis of fake news and similar topics such as rumours or spam[2].

Ideas :

In this work, we propose a method for detecting fake news that has the potential to influence public opinion. Here for extracting features in NLP doc2vec is being used in many papers and researches. We are using Tf-Idf instead of doc2vec because it will give us a more accurate result in case of fake news detection. In this case we can say that Tf-Idf works better than doc2vec which is a generalising version of word2vec. Here if we do not do data pre-processing then the accuracy will increase a lot more in case of detecting fake news.

Potential challenges:

In this paper the potential challenges that we are facing is that we have a very insufficient amount of datasets. The TF-IDF method relies on Google search results. So it is heavily keyword driven.

Conclusion:

Fake news detection is a challenging task to take on. Over the years, we have seen much research in this sector. However, the problem seems to be ever growing at this age. Here, we propose our ideas to face this challenging task. To sum it up, our idea for implementing Tf-Idf for fake news detection could be the solution to this problem.

Bibliography:

[1] 2017. Germany investigating unprecedented spread of fake news online. (January 2017). www.theguardian.com/world/2017/jan/09/ germany-investigating-spread-fake-news-online-russia-election 1

[2] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).