

Report comparison between tf-idf and CBOW

➤ The best model is TF-IDF because it has a higher accuracy than other models.

- **TF-IDF Model using SVC classification:** Accuracy Score = 84.09 %
 - First Ten weights: [0.08623831, 0.00067443, 0.00742417, 0.10901635, -0.051393, -0.0256965, -0.0256965 , 0.45350374, 0.13317529, -0.03227555])
- **CBOW Model using SVC classification:** Accuracy Score = 61.97 %
 - First Ten weights: [0.96915751, -0.61393173, 0.69102031, -1.37128383, 1.21836854, -1.6814669 , 0.99303347, -0.93448003, -0.07090225, 0.18539607]
- **CBOW Model using CNN classification:** Accuracy Score = 70.76 %
 - Shape = (10, 5): [[-0.02187942, -0.1012679 , 0.04078932, -0.02307892, 0.00393769], [-0.07589015, -0.6135594 , 0.18429352, -0.05991728, -0.19143105], [-0.17613725, 0.38137412, 0.18920319, 0.29770893, 0.0302038], [0.00400581, -0.12271068, 0.06805567, -0.92764616, -0.20548436], [0.01635927, -0.38749725, -0.21248342, -1.0566291 , -0.08099387], [0.20160654, -1.7769299 , -0.22258946, -0.48694164, 0.00922324], [0.12965584, 0.59611404, -0.10566047, 0.0368999 , -0.09732449], [0.10822973, -0.32830673, 0.23095767, -0.27286118, -0.21433295], [0.2007957 , -1.9508222 , 0.16215785, -0.90371805, -0.11279596], [-0.04627476, -0.29559678, 0.06679402, -0.24283446, -0.14110461]]

➤ When we get the most similar for 'good' => `model1.most_similar('good')` the result is:

```
[('bad', 0.8474155068397522),  
 ('nice', 0.8347944021224976),  
 ('funny', 0.807961106300354),  
 ('great', 0.7707611918449402),  
 ('quite', 0.7314134836196899),  
 ('perfect', 0.7313023805618286),  
 ('fine', 0.7258012890815735),  
 ('cool', 0.7153827548027039),  
 ('smart', 0.6937596797943115),  
 ('decent', 0.6834757924079895)]
```

And maybe that's the reason of low accuracy in the CBOW.

By:	Ahmed Osama	20160006
	Mariam Nasser	20170284
	Moaaz Hasan	20170411