**Fine-Grained**

**Sentiment Analysis on Financial Microblogs**

**and News**

**Task 1: Sentiment Analysis of Microblogs:**

* **Dataset :**

Training Set – 1204

Test Set – 335

Validation Set -134

**Approach 1 Using Gensim WordVectors:**

**Pre-Processing:**

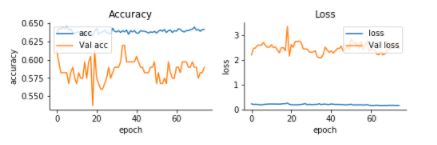
* Removal of special characters, punctuations and numbers.
* Removal of URLs, user names mentioned in a tweet message.
* Removal of words with length less than three in order to reduce the dimensionality of feature space.
* Conversion of tweet text into lower case.
* Concatenation of spans to form a unified string. For the empty spans field, we considered the whole preprocessed message text for feature extraction.

**Features:**

* Word2Vec genism model built using [Finance model (all\_fin\_model\_lower)](https://github.com/apmoore1/semeval/blob/master/models/word2vec_models)  which had collection of 189,206 financial articles.
* Average Weighted vectors built by combination of above word2vec model and tf-idf vector calculated from the dataset.
* The span column is the only text data fed as input.

**Model 1 MLP with unaltered sentiment values:**

* Sequential layer followed by dense layers and dropout of 0.2 and 0.5 respectively.
* Tanh activation function is used .
* RmsProp optimizer is used, with binary\_cross\_entropy as loss function.
* epochs=75, batch\_size=120

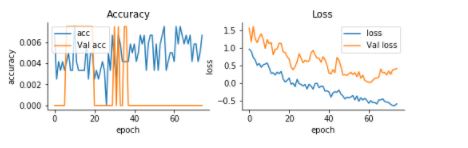




* Accuracy : Less than 1%
* **Cosine\_Similarity: 0.46058743**

**Model 1 MLP with Sentiment scores normalized to 0’s and 1’s:**

* 



* Accuracy: 55.2238812198
* **Cosine\_Similarity:** 0.55062247

**Approach 2**