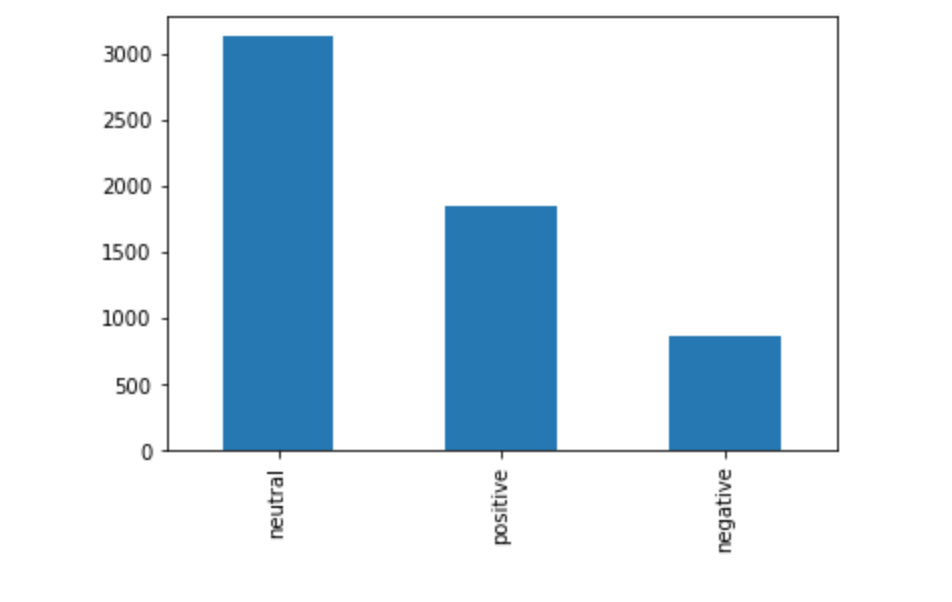
**FINANCIAL STATEMENT SENTIMENT ANALYSIS**

**OBJECTIVE:** To classify financial statements based on sentiment analysis as positive, negative or neutral

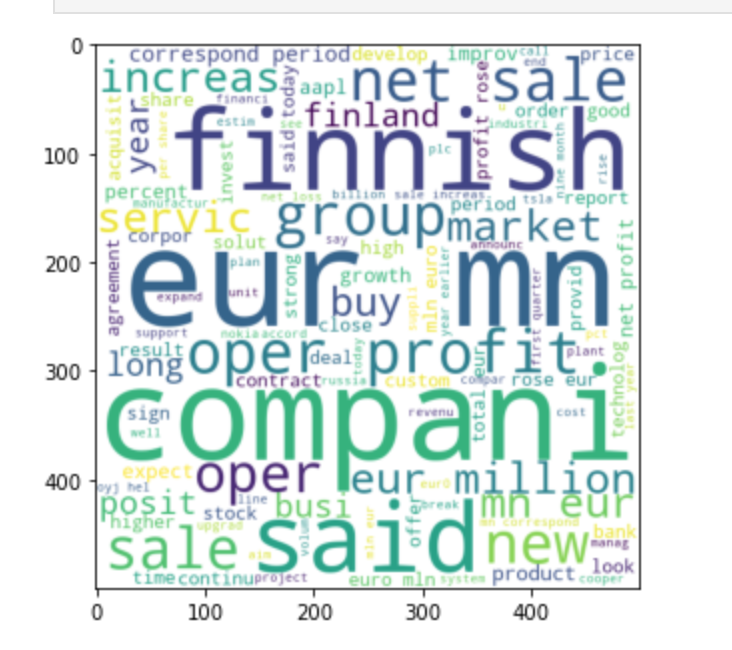
**DATA**: Dataset consisted of 5842 financial statements and their sentiments. The split of sentiment count is shown below:



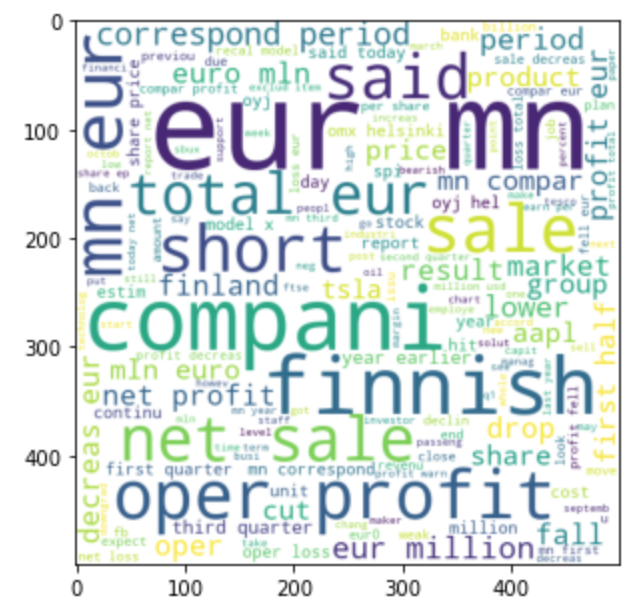
**PREPROCESSING**: Performed lower case transformation, word tokenization, removing special characters, removing stop words and performing stemming on the text data.

**WORD CLOUD**: The frequency distribution of words based on each class can be seen below:

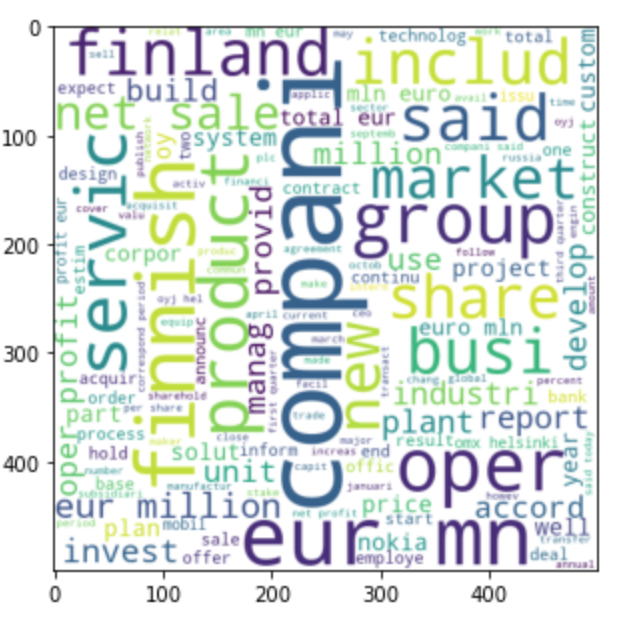
For Positive sentiment texts:



For negative sentiment texts:



For neutral sentiment texts:



**MODEL CREATION**: After using count vectorizer and tfidf vectorizer, multinomial naïve bayes algorithm is used to predict the sentiment of financial statement. The accuracy yielded from both the vectorizer techniques were close to ~67% on test data.

After adding word embedding, LSTM and Dense layers on a sequential neural network, and keeping epochs at 15 with a batch size of 32, the RNN model yielded an accuracy of ~67% on test data