# Sentiment Analysis for Stock Market Prediction Proposal

**Project Overview**

The aim of the project is to create a 2- layered machine learning system in which the first layer performs the retrieval and extraction of sentiment from news articles. The sentiment is then pooled with other relevant attributes as features into a classifier to result in a predicted stock market price. The system will be designed so that over time, the performance of the system becomes better as the classifier learns to predict the stock market value of the target company.

In addition, after the initial focus on the prediction of the stock market performance of a handful entity (or company) and while the classifier is improving its performance, extensions will the made to the system to provide it with a web interface. With this web interface, more entities will be monitored – with the understanding that for these new entities, new data will have to be collected and learning will have to take place – and methods of visualisation will be added to make the complete system interactive.

**Project Background**

It is well understood that stock market price is determined by the value that the market places on the stock and in turn, this is determined by what the perceived value of the company’s assets are. Hence, predicting the stock market can be broken down further into “predicting the perceived value of assets.” In the age of blogs and online news websites that publish news articles as soon as they are known, determine the perceived value becomes a lot easier. The main outcome of the system will be to determine whether or not the performance of <classifier> based stock market prediction can be aided by the inclusion of sentiment as one of the features.

**Objectives**

1. Build a sentiment analyser that will take as input news from online news blogs and will produce as an output the sentiment polarity and strength.
2. Build a classifier which will take in as input the output from the sentiment analyser as well as other suitable features in order to make a prediction of the stock market value
3. Build a web system which will act as an interface to the classifier and sentiment analyser and will provide suitable visualisations in order that the effect of the market sentiment can easily be measured against the performance of the stock market
4. Determine the performance of the classifier against that of traditional classifiers that do not take into account the sentiment.
5. Determine the performance of classifier against that of newer stock market price classifiers that do not take into factor other features such as momentum, relative strength index, commodity channel index and other features as identified.

**Programming Languages and Libraries**

Python will be used to program all parts of the system. This is because the Python programming language make the manipulation of raw data very easy. In addition, web frameworks such as Django, Flask make the transmission from Python objects into web page graphs very easy. For web development, technologies such as HTML, CSS, D3 along will be used.

**Project Workflow**

Develop Client

Develop Server

Literature Search

Perform classification of (training and test) data

Literature Review

Determine data properties

Collect, Generate Data, preprocess data

Train SVM on Data

Pre-process data: (Convert to unigram, remove stopwords, perform feature selection).

Evaluate results.

Write Complete Report

Credit Scoring Application