**Spec Answers**

**Introduction**

In our day-to-day lives, decisions are often made using the recommendations of others and how people ‘perceive’ various aspects of products, predictions etc. Hence, “what people are thinking” is an important factor of the choices made by people and organizations.

This has become even more prominent due to the rise of the World Wide Web. For example, 81% of Internet users have done online research on a product and 20% of Americans do this daily.

(reference <http://www.cse.iitb.ac.in/~pb/cs626-449-2009/prev-years-other-things-nlp/sentiment-analysis-opinion-mining-pang-lee-omsa-published.pdf>).

With growing number of opinion rich resources such as Twitter, online review sites, blogs etc. there is opportunity of using automated tools to calculate the market sentiment useful in making consumer decisions. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the classification of opinion & sentiment in text, has thus occurred response to the interest in new systems that deal directly with opinions as a ﬁrst-class object.

This project deals with classifying sentiments from Twitter and News Corpuses. This decision stems from the ‘live’ nature of Twitter and News which gives opportunity for decisions to be made on an intraday basis. This feature strongly correlates with requirements from agents used in the Finance Industry such as automated trading tools.

Sentiment Analysis, currently showing various degrees of performance, is a topic of high interest for researchers due to the scope of various applications such as market analysis, risk management and as mentioned, automated trading strategies.

Although this research domain shows huge potential, researchers currently face various challenges. Firstly, there is the case of lack of training data required for such a task. One of the motivations behind this project is to investigate how effective Corpuses can be collected and used to accurately predict sentiment.

Actual sentiment analysis also poses interesting challenges. One of which is Multi-Label Classification. To make a successful sentiment analyser, the ability to be flexible in the classification label is important. This project has the interesting challenge of being able to associate the input with multiple labels simultaneously.

However, the main challenge this project aims to solve is of Domain Adaptation. As of today, sentiment analysers perform well on certain type of documents but not on others due to the data they have been trained on. For example, an analyser classifying movie reviews will work well on new instances of movie reviews but not on Tweets regarding company stocks. Hence, it is exciting to research various data labeling techniques which combine the training data from different domains to build a classifier which works across domains.

This powerful result can be further extended to build a real life application in form of a trading agent. The challenges give opportunity to apply this research field within real life problems such as algorithmic trading that processes ‘domain-general’ data to predict ‘domain-specific’ data.

**Group Organisation and Working Culture**

The management structure of this group will consist of a flat structure with each of the three members taking ownership of equally challenging tasks. The group will adhere to certain working practices:

* Group will be proactive with regards to their own tasks as well as improving the standard of other’s responsibilities.
* Development will be using agile methodologies with continuous integration regarding the code.
* Every member will be proactive in their work and will constantly look to add improvements.
* Each group member will be open about the status of tasks and will always be open for collaboration.

Furthermore, within the Group the overall responsibilities will be based on the following role titles:

**Supervisor – Maria Liakata**

**Customer – Rob Procter**

**Project Manager – Ojas Kulkarni.**

The project manager will ensure that the project is being completed with accordance to the working practices defined above and the timetable shown in this specification. Furthermore, it will be project manager’s responsibility to check each deliverable and sign of on its completion if it meets the quality standards decided by the team & supervisor.

**Team Coordinator– Richard Townsend.**

It is the Team Coordinator’s responsibility to set the short-term goals with respect to the objective being completed. The timekeeper will ensure the correct strategy to be used in order to achieve the target defined. The project manager will then ensure that the tasks get performed correctly and within the schedule.

**Lead Developer – Aaron Kalair**

It is the Lead Developer’s responsibility to make sure that all individual components are integrated once submitted by team members. Furthermore, the lead developer will make decisions regarding system architecture & code standards. The lead developer also enforces the development methodology.