**Sentiment Analysis of Yelp Reviews**

**Project Overview:**

Sentiment Analysis is a process of analyzing the text in order to determine whether the attitude of the writer is positive, negative or neutral. It is also called Opinion Analysis or Opinion Mining. It is one of the techniques of Natural Language Processing, a field of Artificial Intelligence.

This process of analyzing the text to determine the opinion/attitude of the writer has been of great interest and value in social networks and blogs. Semantic Analysis is a great tool for businesses to bring changes based on the customer reviews. The dataset from Yelp, a mobile application which publishes crowd-sourced reviews about businesses is being used in this project to perform the Semantic Analysis.

**Proposed Implementation, Programming Language:**

The Sentiment Analysis will be carried out on Yelp Dataset which consists of 2.2 million reviews represented in JSON format. The idea is to capture the review data (text, stars, other information) and analyze the text to guess the rating(stars). The preferred programming language is Python.

**Papers to read:**

1. Michael Gamon, Sentiment classiﬁcation on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis.
2. Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques.
3. Lionel Martin, Pearl Pu, Prediction of Helpful Reviews Using Emotions Extraction.
4. Alexander Pak, Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining.

**AI Category and Algorithms to be implemented:**

The project needs sufficient amount of training(learning) to be done where Machine Learning(ML) comes into play and Sentiment Analysis is essentially Natural Language Processing(NLP). Hence ML, NLP are the main areas of application of AI knowledge gained. The following are the algorithms that will be implemented:

1. Naïve Bayes Classification – Supervised Learning
2. Support Vector Machines – Supervised Learning

**Job Description:**

The following tasks are involved in this project:

1. Capture Yelp review data
2. Preprocess the data into required format
3. Train the Classifiers (Naïve Bayes, Support Vector Machine) with the train data
4. Test the Classifier against test data
5. Review results to predict accuracy