PAPER 1: **Time series forecasting using a hybrid ARIMA and neural network model**

PAPER 2: **Twitter as a Corpus for Sentiment Analysis and Opinion Mining**

This paper describes Sentiment analysis for Microblogging activity in twitter. They present a method to collect a corpus with positive and negative sentiments. They train sentiment classifier from collected corpus (twitter posts) which is able to have distribution of positive and negative and neutral sets of words. They used multinomial Naïve Bayes classifier which uses POS and n-grams tags as features. Although they used other classifiers like SVM, Naïve Bayes gave them best results. Further to increase accuracy they discard common n-grams as they don’t strongly imply any sentiment. They used strategies like salience and entropy to filter out the common n-grams.

We take basic elements of this paper for our work. Instead of the corpus they collected from twitter, we collect reviews of each listings provided by users in AIRBNB datasets and create classifier which classifies these reviews as positive and negative for each listings.

PAPER 3:

This paper contributes in big data by devising new KNN method that can be used for big data. Authors conducted k-means clustering to divide the dataset into various parts and then selected the nearest cluster and conducted KNN classification. They proposed new training process for KNN that blocks training dataset by clustering algorithm and find k-nearest clusters centers for each test sample and then develop new classification model depending on each cluster. They divided conventional KNN into training and testing process and describe the experiments to determine value of k in their extended KNN algorithm.

This is a very organized paper devising KNN method for big data. This paper made us understand the need for KNN classification for large scale data. This paper motivated us to use traditional k-means and KNN classification for our dataset.