Chapter V

WEB SENTIMENT ANALYSIS: COMPARISON OF SENTIMENTS WITH STOCK PRICES USING CORRELATION AND REGRESSION
5.1 Introduction

A survey conducted by marketing firm Comscore in January 2013 has revealed that almost 81% of Internet users have done online research on a product at least once and, among them, 87% reported that reviews had a significant influence on their purchase. This was clearly stated that user-generated opinions are used to make decisions.

Opinion mining is a type of Natural language processing for tracking the attitudes, feelings or appraisals of the public about a particular topic, product or services. Sentiment analysis can be useful in several ways. Such opinion oriented studies include among others, emotion and mood recognition, ranking, relevance computations, perspectives in the text, identification of the source of text and opinion oriented summarization. The new algorithm proposed in the last chapter will help to obtain the significant features used for analyzing the overall sentiment for each object by computing the weighted average for all the sentiments in the textual data.

This chapter deals with the technique that is used for predicting possible relations between sentiment analysis scores and stock prices using correlation and regression.
5.2 Back Ground

Siamak Faridani, [Sia, 2011], introduced an aspect to generalize the concept of multiple dimensions to estimate user ratings along multiple axes such as service, price and value. They use Canonical correlation analysis and derive a mathematical model that can be used as a multivariate regression tool. This model has a number of valuable properties: it can be trained on offline and used efficiently on live stream of texts like blogs and tweets. It can be used for visualization and data clustering and labeling, and finally it can potentially be incorporated into natural language product search algorithms.

Marcela Charfuelan et al. [Mar, 2013], investigated possible correlations between sentiment analysis scores obtained for sentences of Mark Twain’s novel. The Adventures of Tom Sawyer and acoustic features extracted from the same sentences in the corresponding audio book. They have found that scores derived from movie reviews or categorization of emotional stories seem to be more close to the acoustics in the narrative, in particular more correlated with average energy and mean fundamental frequency. They have designed an experiment intended to predict the levels of acoustic expressivity in arbitrary text using sentiment analysis scores and the number of words in the text.

Chuan-Ju Wang et al. [Chu, 2013], this paper attempted to identify the importance of sentiment words in financial reports on financial risk. By using a finance specific sentiment lexicon, they applied regression and ranking techniques to analyze the relations between sentiment words and financial risk. The experimental results show that, based on the bag-of-words model, models trained on
sentiment words only result in comparable performance to those on original texts, which confirmed the importance of financial sentiment words on risk prediction.

Dean P. Foster et al. [2013], Modern data streams routinely combine text with numerical data used in regression analysis. For modeling such data, they described several methods that convert such text into numerical features suitable for regression analysis. The proposed featuring techniques create repressor directly from text, requiring minimal user input. The techniques range naive to subtle. One can simply use raw counts of words, obtain principal components from these counts, or build repressor from counts of adjacent words.

5.3 Proposed System

From the review of literature, it is known that a number of approaches were used to improve the accuracy of prediction of the web sentiment analysis. The aim of study approach is to evolve a technique that is used for predicting possible relations between sentiment analysis scores and stock prices using correlation and regression.

5.3.1 Steps for Comparison of Sentiments with Stock Prices

In opinion mining there are various levels of sentiment analysis such as word level, feature-level, entity-level, sentence-level and document-level. Data sets are collected from different accounts by web crawling. Extracting data from SMN (Twitter): use Twitter API to extract data from twitter domain. There are two APIs used to extract data. REST APIs are having the following resources: Time lines, tweets search, streaming, direct message, friends and followers, users, suggested
users, favorites, lists, saved searchers, place and Geo, trends, spam reports, OAuth, help. These types of APIs use the pull strategy for data retrieval. To collect information a user must explicitly request it. Streaming API provides a continuous stream of public information from Twitter. These types of APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user.

Opinion Retrieval involves retrieving desired information from bag-of-words or Twitter textual data to measure ad hoc information. This system needs a test collection consisting of three things. First one is a document collection. Second one is test suite of information which can be expressed as queries or tags. Third one is a set of relevant judgments, through standard a binary assessment as either relevant or non-relevant for each query-document pair. Sentiment Extraction: Finding or discovering of target entity. It uses various methods to extract the sentiment from the document using unsupervised learning, supervised learning and lexicon based approach. Sentiment Classification: Positive or Negative -Score Analysis: To find weather a piece of text is opinionated or not and to find the polarity of the text. This classification may be binary or multiclass classification. Up to this step we have already completed and calculated sentiment score. Now in this chapter we are going to compare these score values with stock prices of the companies. The company stock prices are collected from the official website of NSE for a period of 50 days from 1\textsuperscript{st} November, 2013 to 20\textsuperscript{th} December 2013. The twitter data for these companies were also collected for the same period of 50 days. Sentiment score analysis has been carried out on this date and sentiment scores or values were arrived at using this analysis.
5.3.2 **Web Sentiments with Correlation analysis**

Pearson correlation coefficient indicates correlation relationship, with two-tailed test of significance. The Pearson correlation coefficient is a measure of the linear correlation between two variables, returning a value between +1 and -1. The correlation coefficient is calculated by:
\[ r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \]

An equivalent expression gives the correlation coefficient as the mean of the products of the standard scores. Based on a sample of paired data \((X_i, Y_i)\), the sample Pearson correlation coefficient is

\[ r = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right) \]

where

\[ \frac{X_i - \bar{X}}{s_X}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \text{and} \quad s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} \]

5.4 Experiment Result:
Comparison of sentiments with stock using Correlation

A correlation is a single number that describes the degree of relationship between two variables. Dependence is a statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice.

This study analyzes the Pearson correlation between overall sentiment and stock prices of a company (i.e.) both Opening and closing prices. The research used Pearson correlation coefficient, which is a linear measure between the variables. It returns a value between +1 and -1. The Pearson correlation coefficient, \( r \), can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association.
that is, as the value of one variable increases, so does the value of the other variable. A value less than 0 indicates a negative association that is, as the value of one variable increases, the value of the other variable decreases. In this research the sample size for the correlation analysis is 35, which represents the working days of stock market during the selected period of analysis.

5.4.1 Analysis with Bharathi Airtel Data set

The following table 5.1 shows the calculated values of correlation between Bharathi Airtel sentiment scores and the company’s share prices together with the ‘P’ values.

<table>
<thead>
<tr>
<th></th>
<th>Airtel Close Price</th>
<th>Airtel Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airtel Open Price</td>
<td>.913**</td>
<td>.684**</td>
</tr>
<tr>
<td>Airtel Close Price</td>
<td>1</td>
<td>.803**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

The ‘p’ value is 0.5, which says that any value of correlation is considered as statistically significant. So, the Airtel overall sentiment is correlated with both the opening price and closing price of Airtel. It is observed from Table 5.1 that the value above 0.6 represents a very strong relationship. Hence it may be concluded that there is strong relationship between Airtel overall sentiment and stock prices and it is statistically significant.
Fig. 5.2: Comparison of Company Sentiments and Stock Prices - Airtel

The Figure 5.2, a line graph was drawn to show the relationship between the overall sentiment and Stock Prices of Airtel.

5.4.2 Analysis with Titan Data set

The following table 5.2 shows the calculated values of correlation between Titan sentiment scores and the company’s share prices together with the ‘P’ values.

Table 5.2
Correlation of Sentiments and Stock Prices – Titan

<table>
<thead>
<tr>
<th></th>
<th>Titan Close Price</th>
<th>Titan Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titan Open Price</td>
<td>.944**</td>
<td>.840**</td>
</tr>
<tr>
<td>Titan Close Price</td>
<td>1</td>
<td>.843**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
The ‘p’ value is 0.5, which says that any value of correlation is considered as statistically significant. The Titan Industries overall sentiment is correlated with both the opening price and closing price of Titan Industries. It is observed from Table 5.2 that the value is more than 0.8 which represents a very strong relationship. Hence it may be concluded that there is a strong relationship between Titan Industries overall sentiment and stock prices and it is statistically significant.

![Graph showing comparison of company sentiments and stock prices](image)

**Fig. 5.3: Comparison of Company Sentiments and Stock Prices - Titan**

In figure 5.3 a line graph was drawn between the overall sentiment and Stock Prices of Titan Industries. The graph clearly shows a very close relationship between the closing prices and opening prices with the overall sentiment of Titan Industries.
5.4.3 Analysis with Bosch Data set

The following table 5.3 shows the calculated values of correlation between Bosch sentiment scores and the company’s share prices together with the ‘P’ values.

**Table 5.3**
Correlation of Sentiments and Stock Prices – Bosch

<table>
<thead>
<tr>
<th></th>
<th>Bosch Close Price</th>
<th>Bosch Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bosch Close Price</td>
<td>.746**</td>
<td>.618**</td>
</tr>
<tr>
<td>Bosch Open Price</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosch Close Price</td>
<td>1</td>
<td>.586**</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).

The ‘p’ value is 0.5, which says that any value of correlation is considered as statistically significant. So, the Bosch overall sentiment is correlated with both the opening price and closing price of Bosch.

**Fig. 5.4: Comparison of Company Sentiments and Stock Prices - Bosch**
The Figure 5.4, a line graph was drawn between the overall sentiment and Stock Prices of Bosch. The graph clearly shows relationship of closing prices and opening prices with the overall sentiment of Bosch. Even though the sentiment of Bosch is correlated with the stock prices, the degree of correlation is not very strong compared to the other companies. Hence it may be concluded that the sentiment can’t be used to predict the stock prices of this company.

5.4.4 Analysis with TCS Data set

The following table 5.4 shows the calculated values of correlation between TCS sentiment scores and the company’s share prices together with the ‘P’ values.

<table>
<thead>
<tr>
<th>TCS Open Price</th>
<th>TCS Close Price</th>
<th>TCS Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCS Close Price</td>
<td>.767**</td>
<td>.774**</td>
</tr>
<tr>
<td>TCS Close Price</td>
<td>1</td>
<td>.801**</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).

The ‘p’ value is 0.5, which says that any value of correlation is considered statistically significant. The TCS overall sentiment is correlated with both the opening price and closing price of TCS. It is observed from Table 5.4 that the value above 0.8 represents a very strong relationship. Hence it may be concluded that there is a strong relationship between TCS overall sentiment and stock prices and it is statistically significant.
The Figure 5.5, a line graph was drawn between the overall sentiment and Stock Prices of TCS. The graph clearly shows relationship of closing prices and opening prices with the overall sentiment of TCS.

5.4.5 Analysis with Colgate Data set

The following table 5.5 shows the calculated values of correlation between TCS sentiment scores and the company’s share prices together with the ‘P’ values.
Table 5.5
Correlation of Sentiments and Stock Prices – Colgate

<table>
<thead>
<tr>
<th></th>
<th>Colgate Close Price</th>
<th>Colgate Overall Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colgate Open Price</td>
<td>.789**</td>
<td>.731**</td>
</tr>
<tr>
<td>Colgate Close Price</td>
<td>1</td>
<td>.761**</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).

The ‘p’ value is 0.5, which says that any value of correlation is considered statistically significant. So, the Colgate overall sentiment is correlated with both the opening price and closing price of Colgate. It is observed from Table 5.5 that the value above 0.7 represents a strong relationship. Hence it may be concluded that there is a strong relationship between Colgate overall sentiment and stock prices and it is statistically significant.

![Fig. 5.6: Comparison of Company Sentiments and Stock Prices - Colgate](image-url)
The Figure 5.5, a line graph was drawn between the overall sentiment and Stock Prices of Colgate. The graph clearly shows relationship of closing prices and opening prices with the overall sentiment of Colgate.

5.5 Findings and Interpretations

In the correlation analysis results were shown for all four companies’ sentiment score and their stock price having strong relationship except one. The Bosch data set analysis shows the sentiment significance value is below the optimum value. In mathematical optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function. It is observed from Bosch data set analysis, it may be concluded that the sentiment can’t be used to predict the stock prices. For further analysis of this research we neglect the company Bosch but the research needed to get accurate and perfect reason for neglecting this for further analysis hence this research undertakes regression analysis to confirm the earlier prediction.

5.6 Web Sentiments with Regression analysis

In general the purpose of multiple regression is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable. Multiple regression is a flexible method of data analysis that may be appropriate whenever a quantitative variable is to be examined in relationship to any other factors. Relationships may be nonlinear, independent variables may be quantitative or qualitative, and one can examine the effects of a single variable or multiple variables with or without the effects of other variables taken into account.
Many practical questions are involved in the relationship between a dependent or criterion variable of interest ($Y$) and a set of $K$ independent variables or potential predictor variables ($X_1, X_2, X_3, ..., X_k$), where the scores on all variables are measured for $N$ cases.

A multiple regression equation for predicting $Y$ can be expressed as follows:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 ... + \beta_k X_k$$

where,

$Y$ is the value of the Dependent variable ($Y$), what is being predicted or explained

$\alpha$ is the Constant or intercept

$\beta_1$ is the Slope (Beta coefficient) for $X_1$

$X_1$ is the First independent variable that is explaining the variance in $Y$

$\beta_2$ is the Slope (Beta coefficient) for $X_2$

$X_2$ is the Second independent variable that explains the variance in $Y$

$\beta_3$ is the Slope (Beta coefficient) for $X_3$

$X_3$ is the Third independent variable that explains the variance in $Y$

To apply the equation, each $X_j$ score for an individual case is multiplied by the corresponding $\beta_j$ value, the products are added together, and the constant $\alpha$ is added to the sum. The result is $Y'$, the predicted $Y$ value for the case.

For a given set of data, the values for $\alpha$ and the $\beta_j$ are determined mathematically to minimize the sum of squared deviations between predicted $Y'$ and the actual $Y$ scores. Calculations are quite complex, and best performed with the help of a computer, although simple cases with only one or two predictors can be solved by hand with special formulas.
The correlation between $Y'$ and the actual $Y$ value is also called the multiple correlation coefficient, $R_{y.12...k}$, or simply $R$. Thus, $R$ provides a measure of how well $Y$ can be predicted from the set of $X$ scores. The following formula can be used to test the null hypothesis that in the population there is no linear relationship between $Y$ and prediction based on the set of $k_X$ variables from $N$ cases:

$$F = \frac{R^2_{y.12...k} / k}{(1 - R^2_{y.12...k}) / (N - k - 1)}, \quad df = k, N - k - 1.$$  

For the statistical test to be accurate, a set of assumptions must be satisfied. The key assumptions are that cases are sampled randomly and independently from the population, and that the deviations of $Y$ values from the predicted $Y$ values are normally distributed with equal variance for all predicted values of $Y$.

Alternatively, the independent variables can be expressed in terms of standardized scores where $Z_1$ is the $Z$ score of variable $X_1$, etc. The regression equation then simplifies to:

$$Z_Y = \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3$$

The value of the multiple correlation $R$ and the test for statistical significance of $R$ are the same for standardized and raw score formulations.

An especially useful application of multiple regression analysis is to determine whether a set of variables (Set B) contributes to the prediction of $Y$ beyond the contribution of a prior set (Set A). The statistics of interest here, $R$ squared added, is the difference between the $R$ squared for both sets of variables ($R^2_{Y.AB}$) and the $R$ squared for only the first set ($R^2_{Y.A}$). Let $k_A$ be the number of variables in the first set and $k_B$ be
the number in the second set, a formula to test the statistical significance of R squared added by Set B is:

\[ F = \frac{\left( R^2_{Y,AB} - R^2_{Y,A} \right) / k_B}{\left( 1 - R^2_{Y,AB} \right) / (N - k_A - k_B - 1)} \], \quad df = k_B, N - k_A - k_B - 1.

Each set may have any number of variables.

5.7 Experiment Result II: Comparison of Sentiments with Stock Prices

This research work involves techniques used for predicting possible relations between sentiment analysis scores and stock prices. The study finds out the regression between the sentiment and stock prices of the selected companies. It will show how sentiment can be related with company stock price movement. In this study we are interested in predicting the closing price of stock (Y) using information on opening price (X₁), and overall sentiment of previous day (X₂) using regression analysis. The whole idea is to determine, whether the sentiment of the previous day and today’s Opening price have an effect or together are they good indicator of today’s closing price.

5.7.1 Analysis with Bharathi Airtel Data set

In this analysis the dependent variable is company Closing price while the independent variables are company opening price and overall Sentiment of a company.
Table 5.6
Multiple Regression of Sentiments and Stock Prices – Airtel

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.954a</td>
<td>.910</td>
<td>.904</td>
<td>3.86681</td>
</tr>
</tbody>
</table>

a - Predictors (Constant), Overall Sentiment, Opening price

The above Table 5.6 presents the model summary of the regression carried out to predict the closing share price movements with the help of investor sentiment derived from twitter and the opening prices. The R square value of 0.910 indicates that the regression model is fit for predicting the closing price with the selected predictors namely sentiments and opening price.

Table 5.7
Significance Test of Predictors – Airtel

<table>
<thead>
<tr>
<th>Predictors</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening Price</td>
<td>.679</td>
<td>.000</td>
</tr>
<tr>
<td>Overall Sentiment</td>
<td>.339</td>
<td>.000</td>
</tr>
</tbody>
</table>

Dependent Variable : Close price

The Table 5.6 represents the predictors taken into account to predict the closing price of the shares of their Airtel Beta values and the ‘P’ values. ‘P’ value less than 0.05(p<0.05) is considered to be statistically significant (i.e.) the predictor can be used. In this case, the ‘p’ value of both prediction are below 0.05 which represents that the company’s Opening Price and overall sentiments are good predictors of the company’s closing price.
Fig. 5.7: Scatter Plot Matrix – Regression Line - Airtel

The scatter plot generally has the predictor variables in the horizontal axis and outcome variable in the vertical axis. The straight line represents the linear regression line. When the line is linear and increasing, it is said that both the variables are linearly correlated (i.e.) if predictor variable increases, the outcome variables also increase. In figure 5.7 the scatter plot continues the Goodness of the model. There were no patterns visible that deviate from the straight line fitted using this regression model.

5.7.2 Analysis with Titan Data set

In this analysis the dependent variable is company closing price while the independent variables are company opening price and company Overall Sentiment.
Table 5.8
Multiple Regression of Sentiments and Stock Prices – Titan

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.958&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.918</td>
<td>.913</td>
<td>3.06727</td>
</tr>
</tbody>
</table>

<sup>a</sup> - Predictors (Constant), Overall Sentiment, Opening price

The above Table 5.8 presents the model summary of the regression carried out to predict the closing share price movements with the help of investor sentiment derived from twitter and the opening prices. The R square value of 0.918 indicates that the regression model is good for predicting the closing price with the selected predictors namely sentiments and opening price.

Table 5.9
Significance Test of Predictors – Titan

<table>
<thead>
<tr>
<th>Predictors</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening Price</td>
<td>.768</td>
<td>.000</td>
</tr>
<tr>
<td>Overall Sentiment</td>
<td>.212</td>
<td>.047</td>
</tr>
</tbody>
</table>

Dependent Variable : Close price

The Table 5.9 represents the predictors taken into account to predict the closing price of the shares of their Titan Beta values and the ‘P’ values. ’P’ value less than 0.05 (p<0.05) is considered to be statistically significant (i.e.) the predictor can be used. In this case, the ‘p’ value of both prediction are below 0.05 which represents that the company’s Opening Price and overall sentiments are good predictors of the company’s closing price.
The scatter plot generally has the predictor variables in the horizontal axis and outcome variable in the vertical axis. The straight line represents the linear regression line. When the line is linear and increasing, it is said that both the variables are linearly correlated \(i.e.,\) if predictor variable increases, the outcome variables also increases. In the figure 5.8 that the scatter plot continue the Goodness of fit of the model there were no pattern visible that deviation from the straight line fitted using this regression model.

### 5.7.3 Analysis with Bosch Data set

In this analysis the dependent variable is closing price while the independent variables are opening price and overall Sentiment.
Table 5.10  
Multiple Regression of Sentiments and Stock Prices – Bosch

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.763&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.581</td>
<td>.555</td>
<td>107.26070</td>
</tr>
</tbody>
</table>

<sup>a</sup> - Predictors (Constant), Overall Sentiment, Opening price

The model summary revealed that the regression line fitted for Bosch Company’s Opening prices has a lower R square value compared with the other companies. It shows that the regression model in this case is not as robust as that of other companies. The above table 5.10 presents the model summary of the regression carried out to predict the closing share price movements with the help of investor sentiment derived from twitter and the opening prices. The R square value of 0.581 indicates that the regression model is good for predicting the closing price with the selected predictors namely sentiments and opening price.

Table 5.11  
Significance Test of Predictors – Bosch

<table>
<thead>
<tr>
<th>Predictors</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening Price</td>
<td>.620</td>
<td>.000</td>
</tr>
<tr>
<td>Overall Sentiment</td>
<td>.203</td>
<td>.171</td>
</tr>
</tbody>
</table>

Dependent Variable: Close price

The above table represents the predictors taken into account to predict the closing price of the company, its Beta value and ‘p’ is the significance value. Anything below 0.05(p<0.0.5) is considered to be statistically significant (i.e.) the predictor can be used. In this case, the ‘p’ value is below 0.05 for company opening
price, which says that it can be used as a predictor separately, while company overall sentiment is not statistically significant, since the value of ‘p’, is above 0.05. So, the overall sentiment of company cannot be a good predictor of its closing price.

In optimization the decision variables in an optimization problem are those variables whose values can vary over the feasible set of alternatives in order to either increase or decrease the value of the objective function. In an optimization problem where the objective function is to be maximized the optimal value is the least upper bound of the objective function values over the entire feasible region. Hence in this research the company was neglected for further analysis.

5.7.4 Analysis with TCS Data set

In this analysis the dependent variable is company Closing price while the independent variables are company opening price and overall Sentiment of a company.

Table 5.12
Multiple Regressions of Sentiments and Stock Prices – TCS

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.844a</td>
<td>.713</td>
<td>.695</td>
<td>23.20510</td>
</tr>
</tbody>
</table>

a - Predictors (Constant), Overall Sentiment, Opening price

The above Table 5.12 presents the model summary of the regression carried out to predict the closing share price movements with the help of investor sentiment derived from twitter and the opening prices. The R square value of 0.844 indicates
that the regression model is good for predicting the closing price with the selected
predictors namely sentiments and opening price.

Table 5.13
Significance Test of Predictors – TCS

<table>
<thead>
<tr>
<th>Predictors</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening Price</td>
<td>.378</td>
<td>.019</td>
</tr>
<tr>
<td>Overall Sentiment</td>
<td>.514</td>
<td>.002</td>
</tr>
</tbody>
</table>

Dependent Variable: Close price

The Table 5.13 represents the predictors taken into account to predict the
closing price of the shares of their TCS Beta values and the ‘P’ values. ’P’ values
less than 0.05(p<0.05) are considered to be statistically significant (i.e.) the
predictor can be used. In this case, the ‘p’ value of both predictions is below 0.05
which represents that the company’s Opening Price and overall sentiments are good
predictors of the company’s closing price.

Fig. 5.9: Scatter Plot Matrix – Regression Line - TCS
The scatter plot generally has the predictor variables in the horizontal axis and outcome variable in the vertical axis. The straight line represents the linear regression line. When the line is linear and increasing, it is said that both the variables are linearly correlated \((i.e.)\) if predictor variable increases, the outcome variables also increase. In the figure 5.9 that the scatter plot continues the Goodness of the model there were no patterns visible that deviate from the straight line fitted using this regression model.

### 5.7.5 Analysis with Colgate Data set

In this analysis the dependent variable is company Closing price of the company while the independent variables are company opening price of the company and company overall Sentiment.

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.831(^a)</td>
<td>.690</td>
<td>.671</td>
<td>10.66996</td>
</tr>
</tbody>
</table>

\(^a\) Predictors (Constant), Overall Sentiment, Opening price

The ‘P’ values of the Beta coefficient of the independent variable show that both of them are equally good prediction of the closing price of Colgate. The above table 5.14 presents the model summary of the regression carried out to predict the closing share price movements with the help of investor sentiment derived from twitter and the opening prices. The R square value of 0.690 indicates that the regression model is good for predicting the closing price with the selected predictors namely sentiments and opening price.
### Table 5.15
Significance Test of Predictors – Colgate

<table>
<thead>
<tr>
<th>Predictors</th>
<th>$\beta$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening Price</td>
<td>.478</td>
<td>.002</td>
</tr>
<tr>
<td>Overall Sentiment</td>
<td>.413</td>
<td>.007</td>
</tr>
</tbody>
</table>

Dependent Variable: Close price

The Table 5.15 represents the predictors taken into account to predict the closing price of the shares of their Colgate Beta values and the ‘$P$’ values. ‘$P$’ values less than 0.05 ($p<0.05$) are considered to be statistically significant ($i.e.$) the predictor can be used. In this case, the ‘$p$’ value of both predictions are below 0.05 which represents that the company’s Opening Price and overall sentiments are good predictors of the company’s closing price.

![Fig. 5.10: Scatter Plot Matrix – Regression Line - Colgate](image)
The scatter plot generally has the predictor variables in the horizontal axis and outcome variable in the vertical axis. The straight line represents the linear regression line. When the line is linear and increasing, it is said that both the variables are linearly correlated \((i.e.\) if predictor variable increases, the outcome variables also increase. In the figure 5.10 the scatter plot continue the Goodness of the model there were no pattern visible that deviates from the straight line fitted using this regression model.

### 5.8 Findings and Interpretations

The results of regression analysis show that four of the five companies’ sentiment scores and their stock prices have strong relationship. Excepting Bosch Ltd the R square values of the model and the ‘P’ values of the independent variable show that the sentiments of the investor are good predictions of the company stock prices. In the case of Bosch Ltd both the R square value of the model and the ‘P’ value of the independent variable indicate that investor sentiment cannot be used to predict the closing prices. Hence investor sentiments of Bosch Ltd are not taken for further analysis.

### 5.9 Chapter Summary

This study shows that company sentiment scores and their stock prices have strong relationship among them. The overall prediction is significant positive correlation between overall sentiment of a firm and its respective stock prices and when research predicts the closing prices of a firm by using a combination of opening prices and overall sentiment, the information criterion is low, thus it leads
to higher accuracy in prediction. For the best prediction take the predictors and find the best way to combine them to predict values on a single scaled outcome variable. In the following chapter the accuracy of prediction of the independent variable is optimized with the help of ALM by finding out the best ways to combine the predictions to predict values on a single scaled outcome variable.