Exploratory aspect-based sentiment analysis approach gauging novel performance from online reviews

Robin Jindal
Indian Institute of Technology Delhi, New Delhi, India

Nicolle Clements
Saint Joseph's University, Philadelphia, USA

Vipul Gupta
Saint Joseph's University, Philadelphia, USA

Abstract

This study examines the dependence of novel sales based on Amazon online reviews by an aspect-based approach. The analysis was conducted on the weighted sentiment score for four different aspects from consumer reviews of popular novels published in 2015. This approach is useful to other novelists who can search for better performing novels and also for publishers who aim at monitoring public sentiment for their novels. This study reveals a feature that is correlated with Amazon's novels ranking system. Results from this study show a dependence of sales on the novel's character aspect.

Keywords: Online Novel reviews, Sales prediction, Rank of Novel, Sentiment analysis, aspect-based approach, regression model, Online Buzz.

INTRODUCTION

The book publishing industry in the United States started with a single printing press in the early 17th century and is now boasting to more than 2,600 publishing houses and generating nearly $25 billion per year revenue in 2000 (Encyclopedia, 2016). According to the reports of StatShot Annual, the U.S. book and journal publishing industry generated $27.98 billion in net revenue for the year 2014, representing 2.7 billion in units. Also, there is around 4% annual growth rate in revenue ("U.S. Publishing industry's annual survey reveals $28 Billion in revenue in 2014," 2015). With being a large industry, the prediction of the success of new novels has made its publishers, professional book reviewers, and writers very curious. There are many factors that influence the success of a novel, some of which may concern the intrinsic content and quality of the book, such as character likeness, action, plot, genre, and good storyline. There are also external factors that impact novel success such as publicity, time of release, competition, and even partly good fortune. Perhaps due to so many dynamic factors involved in the success of a novel, there is room for improvement in the previous work that has been done by the researchers in this area.

The last decade of the 20th century witnessed a boom of the internet which revolutionized the manner of living and business. One of the most important changes brought upon by the internet is the role that the consumers play in our information-driven society. Novels and books are now sold online and are also discussed and critiqued by book reviewers, publishers, and the general audience. In particular, Amazon first began letting customers post reviews of products in 1995, providing a platform to discuss their products. According to a survey conducted by global marketing firm Mintel in 2015, nearly 70% of consumers in America rely on online reviews before making purchase (Kieler, 2015). This revolution generated a lot of electronic word of mouth (eWOM) regarding these books. Recently, the eWOM has become very important in the field of social media mining and is a rich source of business intelligence. In fact, word-of-mouth (WOM) has been acknowledged as one of the most influential means of conveying information in modern history (Godes & Mayzlin, 2004; Maxham & Netemeyer, 2002; Reynolds & Beatty, 1999). With the explosion of internet usage, the electronic word of mouth has become even more influential. With this recent proliferation in the growth of user-generated content on the internet, many organizations are carrying out sentiment analysis and opinion mining of online review postings (Thet, Na & Khoo, 2010).

Analyzing opinions expressed on online platforms has become increasingly important for effective organizational decision making. In this study, the authors have attempted to analyze the eWOM of the consumer reviews posted on Amazon.com via aspect-based sentiment analysis (ABSA) to predict their dependence of the sales and rank of 17 popular novels published in 2015 to gauge their performance. For the purpose of this study, ABSA and Buzz methodology are used. ABSA involves classifying the sentiment (positive or negative) with respect to some aspects of the text whereas Buzz methodology disregards the sentiment and focuses solely on frequency of aspect mentions. Both of these methodologies are discussed in detail in a later section.

This remainder of the paper is structured as follows. First, a review of the literature is presented, followed by fundamental concepts of aspect-based sentiment analysis and buzz methodology. Next, the methods and data used for the purpose of this study are presented, of which results are discussed in following section. The paper wraps up with a discussion, conclusion, and limitations of the study.

LITERATURE REVIEW

According to Medhat, Hassan & Korashy (2014) "sentiment analysis or opinion mining is the computational study of people's opinions, attitudes, and emotions towards an entity. The entity can represent individuals, events or topics that are most likely to be covered by reviews". Sentiment analysis is a particular form of social media mining. It involves the use of a range of technologies to determine personal sentiments expressed regarding particular topics on social media platforms in order to measure the ambient, or general sentiment (Andrejevic, 2011; Arvidsson, 2011; Grimm et al., 2014). Xu, Liao, Li & Song (2011) used a graphical and two-level Conditional Random Field (CRF) model to extract and visualize comparative relations between products from Amazon customer reviews. This tool is effective to support enterprise risk management and decision making. Gunter, Koteyko & Atanasova (2014) focused on examining Emergent Research relevance to applied opinion and behavior measurement and represented a valuable resource to guide political and business decisions. Preethi, Uma & Kumar (2015) developed a generalized prediction model based on temporal sentiment analysis from Twitter to identify the causal relation between events. Ashok, Feng & Choi (2013) examined the quantitative connection between writing style and successful literature. They tried to identify characteristic stylistic elements that were more prominent in successful writings. Khan, Bashir & Qamar (2014) used three-way classification algorithm to achieve an average accuracy of 85.7% with 85.3% precision and 82.2% recall for Twitter feeds classification.

Additionally in the literature, Khoo, Nourbakhsh & Na (2012) evaluated a particular linguistic framework for adoption in the manual and automatic sentiment analysis of news text by using Appraisal theory and automated sentiment analysis. Thelwall, & Buckley (2012) performed mood setting and Lexicon Extension for analyzing topic-based sentiment analysis for the social web. Researched focused on unstructured customer reviews on social networking sites was done by Sam & Chatwin (2013) using Generalizable System Model, Sentiment Analysis Model, and Ontology-based model. Demographic analysis by Gaski (2008) proved that marketing is only marginally related to the consumer sentiment construct. Some shocking results like 'the impact of social media being stronger than conventional media' were revealed as a result of advanced sentiment analysis technique used by Yu, Duan, & Cao (2013). Balahur, Mihalcea & Montoyo (2014) tried to handle different domain of natural language processing tasks by using computational methods. Medhat, Hassan & Korashy (2014) used Ensemble approach to reducing noise sensitivity related to language ambiguity and provided a more accurate prediction for polarity. Silva, Hruschka & Hruschka Jr. (2014) tried to automatically classify the sentiment of Twitter's tweets.

Kim, Jha & Kang (2015) performed sensitivity analysis of general algorithm parameters for the highway alignment optimization case study. They found that 78% accuracy could be achieved in aspect classification and polarity identification of product reviews using natural language text classification (Vasudevan, Joshi, Shekokar, Bhadane, Dalal & Doshi, 2015). Khan, Bashir & Qamar (2014) used three-way classification algorithm to achieve an average accuracy of 85.7% with 85.3% precision and 82.2% recall for Twitter feeds classification. An interesting concept called SentiCircles was used to assign context-specific sentiment orientation to words (Saif, He, Fernandez & Alani, 2016). Maeyer (2012) did a substantive literature review on the relationship between online consumer reviews and sales with a broad overview on firm actions such as pricing. Ghose & Ipeirotis (2007) designed a review ranking system by judging the usefulness and impact of the online reviews. Ashok, Feng & Choi (2014) have predicted the success of literary works by their writing styles.

Taking into account the work done in previous literature, in this paper the authors attempt a new direction by giving a different methodological path to find dependence between the success of literary works and the sentiment scores of four aspects found in their online reviews.

Aspect Based Sentiment Analysis (ABSA) and Buzz Methodology

According to Bagheri, Saraee, & Jong (2014), aspects are defined as attributes that describe the object. For example in the sentences 'The voice quality of my phone is not good' and 'The mileage on my bike is high' the aspects of the phone and the bike are denoted by 'voice quality' and 'mileage' respectively. According to Medhat, Hassan & Korashy (2014), aspect level sentiment analysis aims to classify the sentiment with respect to the specific aspects of entities. The first step in an aspect-based sentiment analysis (ABSA) is to identify the entities and their aspects. Thet, Na & Khoo (2010) have proposed many methods for automatic ABSA but their accuracies are mostly in the range of 80%. Therefore, hoping for more accurate results, the manual ABSA technique under a strict framework has been adopted. The independent variables that have been adopted for the purpose of this methodology are:

a. Weighted aggregated sentiment score for each of the four aspects.
b. Weighted rating.

Weighted aggregated sentiment score for each aspect is calculated by the weighted average of the individual sentiment scores. The weight for a review is equal to the fraction of people who found that review useful out of the total reviews. Similarly, the weighted rating is the proportional average of the ratings for individual reviews. More details on the calculations of these variables are given in the following section.

Asur & Huberman (2010) demonstrated how social media content can be used to predict real-world outcomes. Oh, Hu & Yang (2016) explored the relationship between social media information diffusion and economic outcomes. The social media content generates an online "buzz", a measure of content frequency. The dependence of the Novel's sales based on this online chatter is studied by buzz methodology. For buzz methodology, the number of total sentiments (NTS) has been adopted as a metric for the buzz frequency. NTS specifies the number of occurrences of a particular aspect in a review dataset. For example in a review dataset of N reviews, NTS score for a particular aspect is defined as the number of reviews expressing some sentiment (except neutral) for that particular aspect.

Similarly, the number of positive sentiment (NPS) and number of negative sentiment (NNS) for a particular aspect are defined as the number of reviews depicting an overall positive and an overall negative sentiment score for that particular aspect respectively. The NTS, NPS, and NNS scores are used as a metric for online buzz. This study attempts to use these scores to find any dependence of the sales and rank of a novel.

Average rank, all time sales, and Amazon customer rating are treated as outcome variables for the purpose of this study. Average rank refers to the ranking of a book (lower is better) on Amazon.com and is updated on an hourly basis. All time sales refer to the number of hardcover copies of a novel sold via all the domains of Amazon. Amazon customer rating is the tangible star-rating given on Amazon.com for a particular novel on a scale of 1-5.

METHODOLOGY

Research Framework

Figure 1 (Appendix A) shows a cyclic flow chart showing the research framework. The online reviews given by readers influences the future sales of the novel by influencing 'the decision of buying the novel or not' for the prospective novel buyers. When prospective novelists purchase the books and post their reviews online, they influence more buyers by affecting their decisions and thus creating a chain in which, ultimately, the online reviews affects the sales of a novel. This chain of online reviews creates a social ‘buzz’ in the market, in turn affecting the sales of a novel. The sales data directly influences the rank associated with that novel. The better the sales of a novel, the better will be its rank (lower is better) compared to other novels in the market. Also, the online reviews given by these novel buyers generates public feedback about the novel which itself is a factor that can be used to study the sales of a novel. Improved qualitative and quantitative study for the sales and rank of novels can be done by considering the rating of these reviews which will give a better public opinion on a novel.

Dataset Construction

For the purpose of this study, a sample of 17 popular novels published in 2015 was taken (see Appendix A for the complete list of novels selected). The user-reviews of these novels were collected from the website Amazon.com. This site is an American electronic commerce and cloud computing company and currently is the largest Internet-based retailer in the United States. Amazon allows its users to mark a particular review as ‘helpful’ in case the user found that particular review to be beneficial in making a purchasing decision. For the purpose of this study, the 30 most helpful reviews from Amazon.com for each novel were taken so that aggregation of the review scores would satisfy normality assumptions due to Central Limit Theorem. The sales data and rank of the novel were taken from an online resource called NovelRank (www.novelrank.com), which gives the average all-time sales of a particular novel sold via all domains of Amazon and the ranking of the book on Amazon.com (see Appendix A for a table showing the sales and rank data for the novels). By using NovelRank, the Amazon sales data across the world can be tracked and the sales for particular novels can be exposed. These 30 reviews of 17 novels from Amazon.com, novel sales, and rank data from NovelRank were collected in order to find dependence between the electronic word of mouth spread via online user reviews and the corresponding sales and rank data to gauge their performance.

Methodology of Aspect Based Sentiment Analysis (ABSA)

Each of the online user reviews was processed manually by aspect-based sentiment analysis (ABSA) technique to get the sentiment score of review based on different aspects. The four most frequently occurring aspects in novel reviews were: Story, Character, Narration, and General. The general aspect carries the public opinions on the novel and its comparison with other novels. These aspects were identified in the reviews by implementing predefined rules which are discussed in later details. The sentiment score for a particular aspect was also obtained by a well-defined structure using an online API Indico.io whose framework is also discussed in a later section. The process that is followed to obtain the sentiment score mainly consists of two steps: 1. Aspect Identification and 2. Sentiment Score Allotment for each identified aspect. In other words, to obtain a sentiment score with respect to a particular aspect, the first step is to perform aspect identification on each sentence in the review followed assigning a sentiment score to the aspect identified. The details of these steps are discussed next.

Aspect Identification

The guidelines for identifying a particular aspect in a novel review are summarized in Table 1 (Appendix A). The first column in the table mentions the name of the four aspects under consideration. The second column in the table represents the list of marker words which indicate the presence of a particular aspect of the review. For example, in a particular review “The writing skills of the author are phenomenal”, in this review the term ‘writing skills’ clearly, refers to the aspect ‘Narration’ of the novel by the predefined rules mentioned in Table 1 (Appendix A). All synonym words to the listed markers were also used as markers.

The third column in the table refers to the novel specific marker words for a particular aspect. For example in the review specific to the novel ‘Red Queen’ written by ‘Victoria Aveyard’ the sentence in a particular review “I really loved the world Aveyard created”, the presence of the name ‘Aveyard’ indicates to the narration aspect of the novel. A list of the novel specific words for an aspect was created so as to ease the process of identifying the aspect in a user review. The fourth column in the table defines the issues handled/questions answered by a particular aspect, creating a list of things that were consider in an aspect. For example in a particular review “The novel was a very comfortable read”, therefore this refers to the aspect ‘Narration’ as it answers the question of how comfortable was the reading of the novel in the review.

Sentiment Score Assignment

After identifying the aspect in the novel review, the sentiment score of the aspect is assigned with the help of Indico.io. This sentiment analysis API takes a sample of text (tweet, article, email, forum post, report, etc.) and returns the percentage positivity. This is similar to the notion of the text of a 5-star review should be positive, while the text of a 1-star review is negative. This API has set an accuracy standard for sentiment analysis at 93.8 percent using its sentiment HQ tool. Results were measured against the IMDB dataset, an academic gold standard. This accuracy is largely the result of using a Recurrent Neural Network (RNN) capable of understanding complex language, such as negation and sarcasm, with higher efficacy than traditional methods. The API predicts a score for each piece of text submitted and returns a value between 0 and 1. When the score is low (close to 0.0), it indicates negative sentiment. When the score is high (close to 1.0), it indicates positive sentiment. The thresholds used in this study are depicted in Figure 2 (Appendix A).

For the purpose of this study, API is used to find the sentiment score corresponding to a particular aspect of each review text. Figure 2 (Appendix A) shows the framework that has been used to assign the sentiment score for different aspects for the reviews.

The visualization below shows an illustration of the procedure that was used i.e. first identify and separate the lines in a review on the basis of aspect and then calculate the sentiment score according to predefined framework.

Buzz Methodology

Through recent research, it has been noticed that the amount of online ‘buzz’ prevalent in the social media is a major factor in predicting the success of a novel, measured by its sales data and its ranking. The number of positive sentiments (NPS) is defined as the count of reviews that have a positive sentiment for a particular aspect of a novel. Therefore all the reviews having a positive sentiment score are counted in NPS. Similarly, the number of negative sentiments (NNS) is defined as the count of reviews that carry a negative sentiment within them with respect to a particular aspect. Also, the number of total sentiments (NTS) is the count of the reviews that carry either positive or negative sentiment with respect to a particular aspect. All reviews having a sentiment score except 0 (Neutral) are counted in NTS. NTS indicates that the aspects are at least ‘talked about’ in the social media no matter whether they carry a positive or a negative sentiment. Due to the virtue of the definition of these three terms, defined below is a mathematical formula depicted in equation (1).

$$NTS = NNS + NPS$$

The NTS, NNS, and NPS scores were aggregated in a similar fashion as the sentiment analysis score for each aspect of a novel and were also used in the regression modeling for the sales and rank of the novel.

Regression Analysis

The aggregated sentiment score for each of the four aspects, weighted rating, and the buzz methodology parameters (NTS, NPS, and NNS) were used as independent variables in several exploratory multi-linear regression models (MLR). These independent variables were used to explore outcome variables including the average rank, average all-time sales, and the Amazon customer rating of the novel. All the analysis was carried out in the Data Analysis ToolPak of Excel 2013.

RESULTS

Preliminary descriptive statistical analysis done on the outcome variables is summarized in Table 2 (Appendix A). The descriptive analysis details about the measures of central tendency, spread and standard errors associated with the outcome variables. Additionally, the outcome variables were found to be roughly normally distributed.

Aspect Based Sentiment Analysis (ABSA) Model Results

The results found using the weighted sentiment scores by the ABSA model demonstrated significant predictions for several of the outcome variables and are listed in Table 3 (Appendix A). Column 1 and 2 in the table lists the independent and outcome variables between which dependence was found out to be statistically significant. Column 3 and 4 give details about the Pearson's correlation coefficient and the p-value of the dependence. Column 5, 6, and 7 provide further details of the dependence.

As listed in Table 3 (Appendix A), weighted rating and average rank were found to be negatively correlated, meaning better the weighted rating of the most helpful reviews of a novel, the better will be its Amazon-given rank. Also, the correlation between the character aspect and all-time sales was found to be negative. Meaning, the more there are negative sentiments about the character aspect in the online user reviews, the higher likelihood of sales of that novel! The scatter plots showing the dependence between these pairs of variables are visualized in the Figures 4-6 (Appendix A).

Buzz Methodology Results

The MLR results from NTS and NNS scores did not identify any significant predictions on the outcome variables. However, the NPS score from character aspect was found to have a statistically significant dependence on the all-time sales. They were found to have a weak to moderately strong negative relationship in between them (Pearson's correlation coefficient = -0.3346 and p-value = 0.029). Surprisingly, the dependence signifies that more the number of positive reviews about the 'Character' the less will be the sales of the novel! The dependence can be visualized by the scatter plot given in Figure 7 (Appendix A).

DISCUSSION

The results show that there is little significance of the sentiment lying in the online user reviews on overall performance of novels. Therefore it suggests that a prospective buyer must not make a decision solely on the basis of the sentiment of the reviews that are posted online. Hence, a prospective buyer must not:

a. make a choice of buying the novel based on the positive sentiments prevalent in the online posted reviews.

b. reject the idea of buying a novel after seeing negative sentiments in online reviews.

By the ABSA methodology, it was obtained that average rank and weighted rating have a negative dependence, indicating the higher the weighted rating of a novel the lesser is the average rank (recall low rank is better). This result can be interpreted to speculate that Amazon uses the weighted rating of a novel as one of its variables to give it a rank. But since the relation is only weak to moderately strong, this requires further investigation on other variables Amazon might use to assign ranks. Figure 8 (Appendix A) shows that there is a strong, positive and linear dependence between the Amazon customer rating and weighted rating. One can speculate that Amazon uses the weighted rating as one of many factors to calculate customer rating.

Also through this study, several unexpected results have been obtained. One surprising result is that the more negative sentiments about the character aspect in the online user reviews indicate higher sales of that novel. Contrasting to the popular belief that positive reviews affect the performance of a novel in a positive way, these results can be justified by recent studies which claim that negative reviews sometimes increases the sales of the product by acting as risk mitigations. Berger, Sorensen & Rasmussen (2010) argued that negative publicity can increase purchase likelihood and sales by using product awareness by sparking consumer curiosity. However, further investigation is required to see if this notion applies in this study.

CONCLUSION AND LIMITATIONS

In this paper, the dependence of novel performance from social buzz and sentiments contained in online reviews was explored by an aspect-based approach. This study takes into consideration the data from 30 most helpful reviews of 17 popular novels published in 2015. The study revealed that higher sales and better ranks of the novel are correlated with negative character reviews. The findings were not able to predict any strong dependence between a novel's performance and the sentiments prevalent in online reviews.

The study is limited to only the data of 30 most helpful reviews of 17 popular novels. Though the global sales data were used, the study only considers reviews in English language only. Therefore, future research directions include:

a. Study of a data that represents a more diverse novel population.
b. Able to analyze reviews from other languages.
c. Able to include smart features that tackle advanced things like location, how well detailed was the review etc.

REFERENCES

Andrejevic, M. (2011). Watching Back. Media and Social Justice, 177-191.

Arvidsson, A. (2011). Ethics and value in customer co-production. Marketing Theory, 11(3), 261-278.

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

Bagheri, A., Saraee, M., & Jong, F. D. (2014). ADM-LDA: An aspect detection model based on topic modeling using the structure of review sentences. Journal of Information Science, 40(5), 621-636.

Balahur, A., Mihalcea, R. & Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. Computer Speech & Language, Volume 28, Issue 1, Pages 1-6, ISSN 0885-2308.

Berger J., Sorensen, A.T. & Rasmussen, S.J (2010). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. Marketing Science, Vol 29, No. 5, pp. 815-827.

Campbell, C. (n.d.). ReviewTrackers - Online Reviews Facts & Figures Online Reviews Facts & Figures. Retrieved from http://www.reviewtrackers.com/wp-content/uploads/ReviewTrackers-Online-Reviews-Facts.pdf.

De Maeyer, P. (2012). Impact of online consumer reviews on sales and price strategies: A review and directions for future research. *Journal of Product & Brand Management*, *21*(2), 132–139.

Gaski, J.F. (2008). The index of Consumer Sentiment toward Marketing: Validation, Updated Results, and Demographic Analysis .Journal of consumer policy, Issue 2, pp 195-216.

Ghose, A., & Ipeirotis, P. G. (2007). Designing novel review ranking systems. . doi:10.1145/1282100.1282158

Godes, D. & Mayzlin, D.(2004). Using Online Conversations to Study Word-of-Mouth Communication. Marketing Science, Volume 23, pages 545-560.

Grimm, E. L., & Al., E. A. (2006). Substituted Coumarins as Potent 5-Lipoxygenase Inhibitors. ChemInform, 37(31).

Gunter, B., Koteyko, N. & Atanasova, D (2014). Sentiment analysis: A market-relevant and reliable measure of public feeling? International Journal of Market Research, Vol. 56, Issue 2.

Khan, F.H., Bashir, S. & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. Decision Support Systems, Volume 57, Pages 245-257, ISSN 0167-9236.

Khoo, C.S.G., Nourbakhsh, A., Na, J.C. (2012). Sentiment analysis of online news text: a case study of appraisal theory. Online Information Review, volume 36, pages 858-878.

Kieler, A. (2015, June 3). *Nearly 70% of consumers rely on online reviews before making A purchase*. Retrieved July 26, 2016, from https://consumerist.com/2015/06/03/nearly-70-of-consumers-rely-on-online-reviews-before-making-a-purchase/

Kim, E., Jha, M. K., & Kang, M. (2015). A Sensitivity Analysis of Critical Genetic Algorithm Parameters. International Journal of Operations Research and Information Systems, 6(1), 30-48.

Maxham, J.G. & Netemeyer, R.G. (2002). A Longitudinal Study of Complaining Customers: Evaluations of Multiple Service Failures and Recovery Efforts. Journal of Marketing, Vol. 66(pp. 57-71).

Medhat, W., Hassan, A. & Korashy, H. (2014). Sentiment analysis algorithm and applications: A survey. Ain Shams Engineering Journal, Volume 5, Issue 4, Pages 1093- 1113.

Mohanty, S., Clements, N., Gupta, V. (2015). Investigation of the relationship between the electronic word of mouth (eWOM) present in user reviews of movies and their box office success through an aspect based approach. Submitted to the International Journal of Business Analytics.

Oh, C. & Hu, H. (2016) "Social Media Information Diffusion and Economic Outcomes: Twitter Retweets and Box Office Revenue" Accepted for presentation at 20th Pacific Asia Conference on Information Systems (PACIS 2016), Chiayi, Taiwan.

Preethi, P.G., Uma, V. & Kumar, A. (2015). Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction. International Conference on Computer, Communication and Convergence, Procedia Computer Science, Volume 48, Pages 84-89, ISSN 1877-0509.

"Publishing Industry." Dictionary of American History. 2003. Retrieved July 26, 2016 from Encyclopedia.com: http://www.encyclopedia.com/doc/1G2-3401803454.html

Reynolds, K. E., & Beatty, S. E. (1999). A relationship customer typology. Journal of Retailing, 75(4), 509-523.

Saif, H., He, Y., Fernandez, M. & Alani H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, Volume 52, Issue 1, Pages 5-19, ISSN 0306-4573.

Sam, K.M. & Chatwin, C.R. (2013). Ontology-Based Sentiment Analysis Model of Customer Reviews for Electronic Products. International Journal of e-Education, e-Business, e-Management and e-Learning. Vol. 3, No. 6.

Silva, N.F.F.D., Hruschka, E.R. & Hruschka Jr., E.R (2014). Tweet sentiment analysis with classifier ensembles, Decision Support Systems, Volume 66, Pages 170-179, ISSN 0167-9236.

Stony Brook University. (2014, January 6). Some elements of writing style differentiate successful fiction. ScienceDaily. Retrieved July 22, 2016, from www.sciencedaily.com/releases/2014/01/140106094151.htm.

Thelwall, M. & Buckley, K. (2012). Topic-Based Sentiment Analysis for the Social Web: The Role of Mood and Issue-Related Words. Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK.

Thet, T.T., Na, J.N. & Khoo, C.S.G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science, Volume 36 Issue 6, Pages 823-848.

Thet, T.T., Na, J.N. & Khoo, C.S.G. (2010). Comparing sentiment expression in movie reviews from four online genres. Online Information Review, 34:2, 317-338.

U.S. Publishing industry's annual survey reveals $28 Billion in revenue in 2014. (2015). Retrieved July 26, 2016, from http://publishers.org/news/us-publishing-industry%E2%80%99s-annual-survey-reveals-28-billion-revenue-2014

U.S. Publishing Industry's Annual Survey Reveals $28 Billion in Revenue in 2014. (n.d.). Retrieved July 22, 2016, from http://publishers.org/news/us-publishing-industry's-annual-survey-reveals-28-billion-revenue-2014.

Vasudevan, H., Joshi, A.R., Shekokar, N.M., Bhadane, C., Dalal, H. & Doshi, H (2015). International Conference on Advanced Computing Technologies and Applications (ICACTA).Sentiment Analysis: Measuring Opinions, Procedia Computer Science, Volume 45, Pages 808-814, ISSN 1877-0509.

Xu, K., Liao, S.S., Li, J. & Song Y (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. Decision Support Systems, Volume 50, Issue 4, Pages 743-754, ISSN 0167-9236.

Yu, Y., Duan, W. & Cao, Q (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. Decision Support Systems, Volume 55, Issue 4, Pages 919-926, ISSN 0167-9236.

Appendix A: Tables and Figures from the manuscript

Figure 1: Research Framework

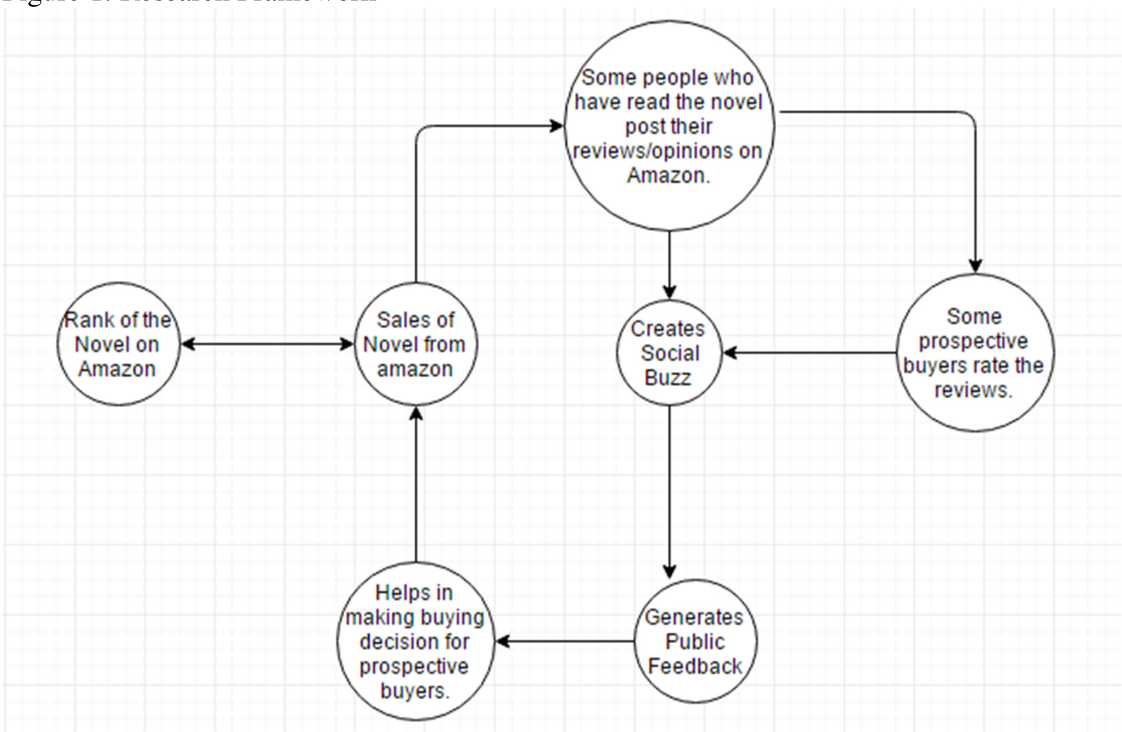


Table 1: Aspect Identification Table

Aspect	***General Wording***	***Specific Wording***	***Issues Handled by the Aspect***
Character	Hero, villain protagonist, antagonist, act, play, role, cast, performed, portrayed, character, All the words corresponding to 3rd person singular/plural pronouns.	Character Names	1. Relationship status between different characters. 2. How well did the reader bond with the character? 3. Can the reader relate himself to the character? 4. Did the character seem real? 5. Was the characterization done properly by the author?
Story	Twists, turns, plot, tale, story, timeline, climax, Thriller, psychological, mystery, premise, exploration, ending, events, contents	Topic/Main motto of the novel.	1. What is the main part of the story? 2. What is the central topic of the novel? 3. What is the literary composition used in the novel?
Narration	Writing skills, told, narrated, in one afternoon, voice, the point of view, perspective, dialogue.	Narrator's Name Author's Name	1. How good is the author? 2. Is the title appropriate for the novel? 3. How comfortable is the reading, which is judged by the pace and intensity of the novel

General	The novel, Book, NY Times best seller, devouring readers, in one go.	Novel Name	1. General comments on the novel. 2. Comparison with other novels. 3. Do the buyers think they paid the right price? 4. Packaging and a cover of the book? 5. Was it good enough for Re-Read? 6. Controversies & Publication Issues.

Figure 2: Sentiment score assignment framework.

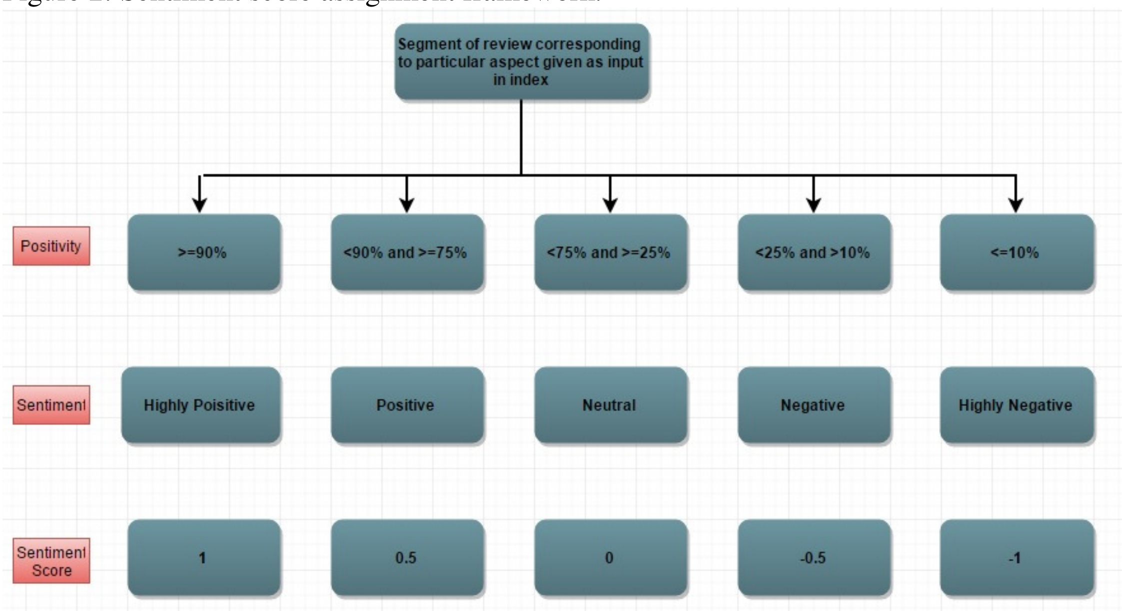
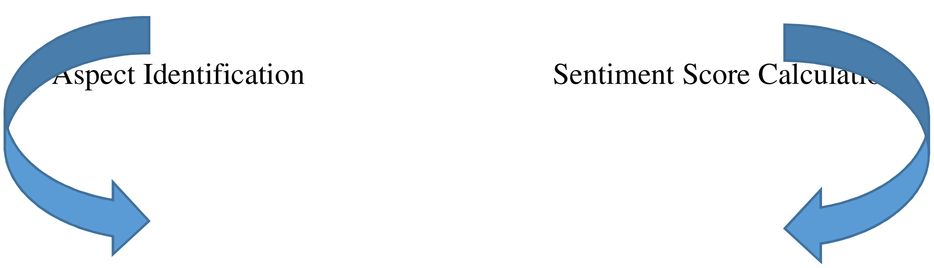


Figure 3: Illustration of methodology for aspect identification and sentiment score calculation

Sample Review for the Novel : "All the bright places"
This year instead of having New Year's resolutions I won't keep, I decided to make book goals and one of those goals was to read more YA. 25 is my goal for the year. This was the first on my list this year and it blew me away! For real, I had ALL THE FEELS while reading this. Powerful, memorable, smart, funny, sad, real… this book was just completely authentic and moving. There were so many 'quotable' things in this book. I just want to talk about the story so much but I went in blind and loved it all the more so I'll not talk much about the story. What I will say is Theodore Finch is one of the most endearing characters I've ever read about. This book made me want to wander, to get out and explore the world around me, or just the state that I live in. This book made me think about my life. This book hit important topics that people need to know more about and recognize. This book will leave you emotionally drained, but it's SO WORTH IT! All the Bright Places is a book that needs to be read. It's a brilliant and beautiful and sad and everything! If this is the way all my YA books are going to be this year, I'm pretty sure I'll exceed my 25 goals!



Aspect	Sentence	% Positivity	Sentiment Score
Story	I just want to talk about the story so much but I went in blind and loved it all the more so I'll not talk much about the story.	92	1
	This book made me think about my life. This book hit important topics that people need to know more about and recognize.	99	1
	Weighted Sentiment Score for Story		1
Character	What I will say is Theodore Finch is one of the most endearing characters I've ever read about.	99	1
	Weighted Sentiment Score for Character		1
Narration	There were so many 'quotable' things in this book	58	0
	For real, I had ALL THE FEELS while reading this.	89	0.5
	Weighted Sentiment Score for Narration		0.25
General	This was the first on my list this year and it blew me away.	99	1
	Powerful, memorable, smart, funny, sad, real… this book was just completely authentic and moving.	100	1
	This book made me want to wander, to get out and explore the world around me, or just the state that I live in.	80	0.5
	This book will leave you emotionally drained, but it's SO WORTH IT!	100	1
	All the Bright Places is a book that needs to be read.	84	0.5
	. It's a brilliant and beautiful and sad and everything.	99	1
	If this is the way all my YA books are going to be this year, I'm pretty sure I'll exceed my 25 goals!	98	1
	Weighted Sentiment Score for General		0.8571428

Table 2: Descriptive statistics of outcome variables.

	Average Rank	**All Time Sales**	**Amazon Customer Rating**
Mean	207,365.12	10014.82353	4.364705882
Median	178132	5694	4.5
Range	49506 – 499694	1403- 33333	3.4 - 4.8
Standard Deviation	132513.4306	10184.03002	0.401467895
Standard Error	32139.22771	2469.990086	0.097370267

Table 3: Details of dependence between independent and outcome variables.

Independent Variable	***Outcome Variable***	***Pearson's Correlation Coefficient***	***p-Value***	***Type of Dependence***	***Dependence Direction***	***Dependence Strength***
Weighted Rating	Average Rank	-0.5083	0.037	linear	Negative	weak to moderately strong

Character	All Time Sales	-0.5723	0.001	linear	Negative	weak to moderately strong
Weighted Rating	Amazon Customer Rating	0.6801	0.002	linear	Positive	moderately strong to strong

Figure 4: Scatter Plot of Average Rank and Weighted Rating showing a negative linear relationship.

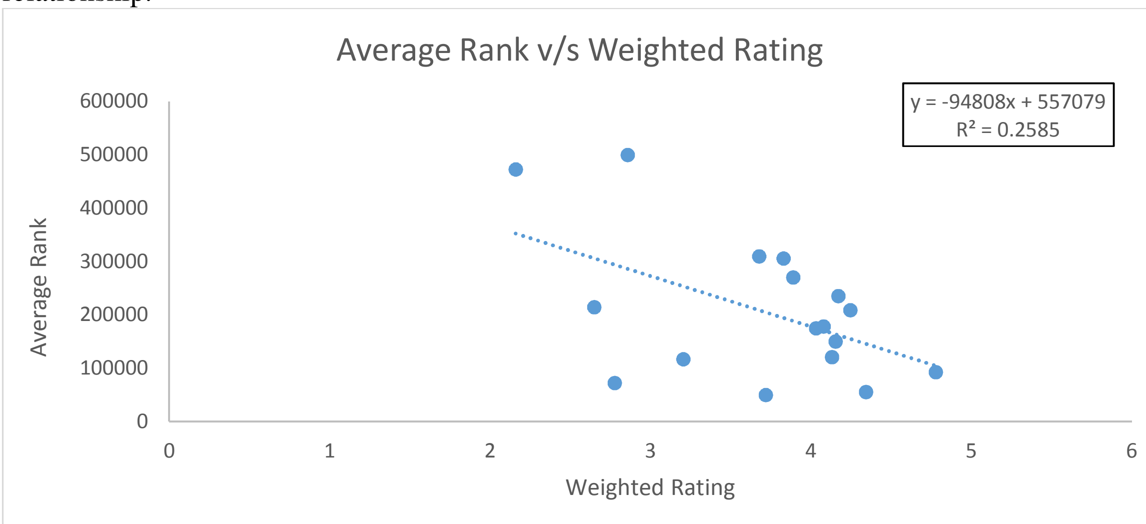


Figure 5: Scatter plot of all time sales and the weighted sentiment score of character showing a negative linear relationship.

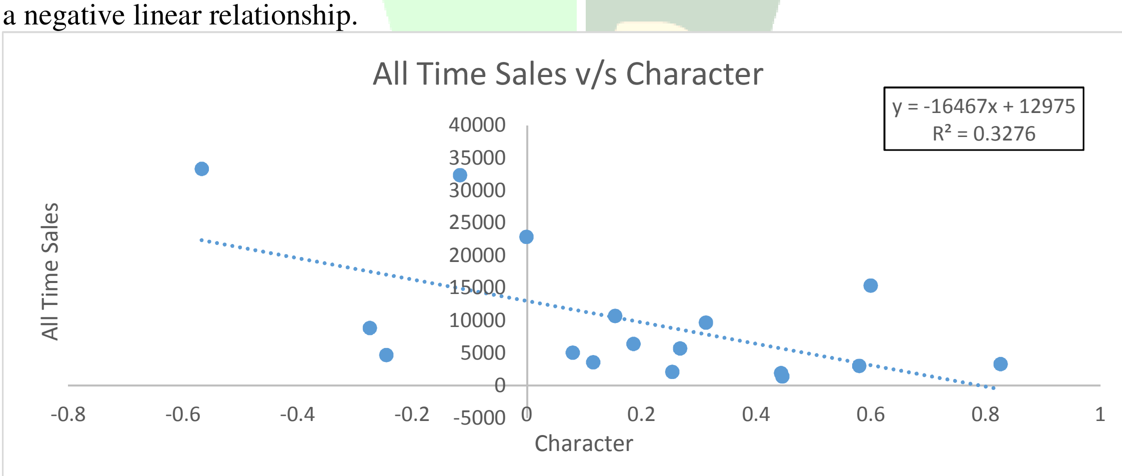


Figure 6: Scatter plot of Amazon customer rating and weighted rating showing a positive linear relationship.

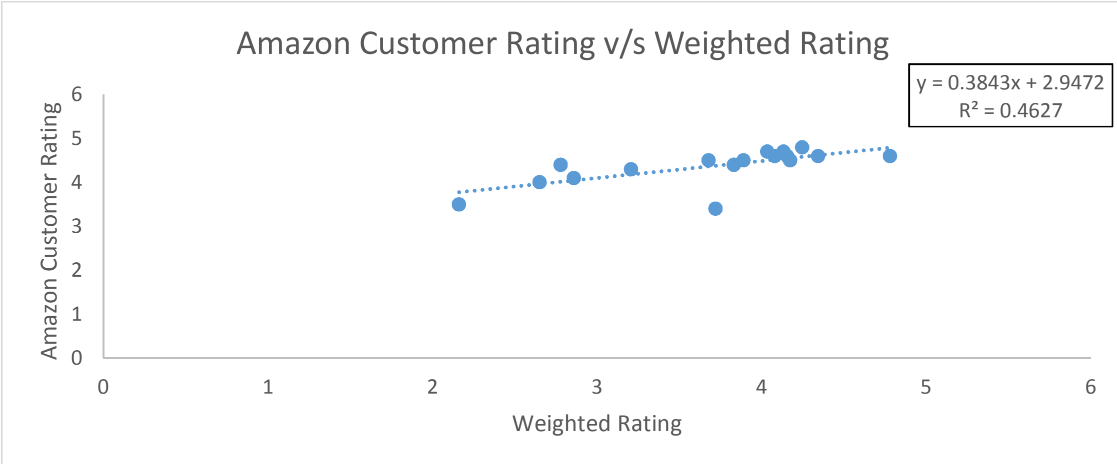
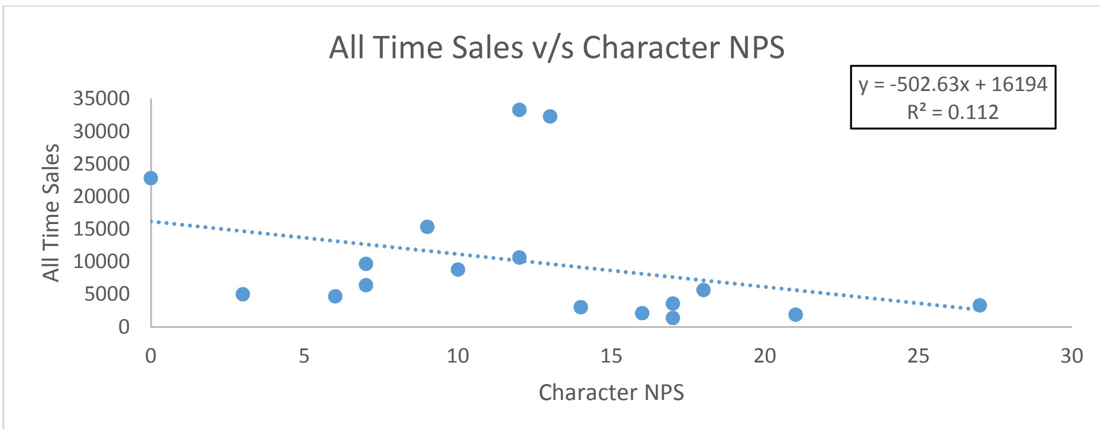


Figure 7: Scatter Plot showing a linear dependence of all time sales on the number of positive sentiments of character.



Appendix B: List of the novels and data for outcome variables.

S.No.	**Novel**	**Avg. All time sales**	**Average Rank**	**Amazon Customer Rating**
1	The Girl on the Train	32105	223,187	4
2	Red Queen (Red Queen, #1)	9462	111,537	4.3
3	Go Set a Watchman	32064	45,552	3.4
4	All the Bright Places	3317	192,766	4.6
5	A Court of Thorns and Roses (A Court of Thorns and Roses, #1)	1645	163,982	4.6
6	After You (Me Before You, #2)	7618	9021	4.1
7	Why Not Me?	6083	311,415	4.5
8	Winter (The Lunar Chronicles, #4)	2805	189,919	4.7
9	An Ember in the Ashes (An Ember in the Ashes, #1)	5526	277,513	4.5
10	Six of Crows (Six of Crows, #1)	3055	65,792	4.6
11	A Little Life	4504	98,822	4.1
12	A Darker Shade of Magic (Shades of Magic, #1)	1234	322,330	4.4

13	Queen of Shadows (Throne of Glass, #4)	1801	152,922	4.8
14	Fates and Furies	4823	395,061	3.5
15	Between the World and Me	20341	63,958	4.6
16	Dead Wake: The Last Crossing of the Lusitania	15149	226,301	4.5
17	The Sword of Summer (Magnus Chase and the Gods of Asgard, #1)	10084	139,712	4.7

Appendix C: List of the novels and the data for independent variables.

S.No.	Novel Name	Novel Rating	Aspect	Aspect Score	Buzz		
					NTS	NPS	NNS
1	The Girl on the Train	2.6	Story	-0.2613	23	2	21
			Character	-0.5675	18	7	11
			Narration	0.0082	23	7	16
			General	-0.2218	22	6	16
2	Red Queen	3.2	Story	0.3122	22	12	10
			Character	0.045	233	7	16
			Narration	-0.0179	19	7	12
			General	-0.036	22	9	13
3	Go Set a Watchman	3.71	Story	-0.116	19	8	11
			Character	0.1178	25	13	12
			Narration	0.122	25	10	15
			General	-0.024	25	8	17
4	All the Bright Places	4.079	Story	0.115	25	16	9
			Character	0.3099	21	17	4
			Narration	0.408	13	12	1
			General	0.2604	22	19	3
5	A Court of Thorns and Roses	4.15	Story	0.4432	26	24	2
			Character	0.1681	25	21	4
			Narration	0.1727	25	22	3
			General	0.5469	27	23	4
6	After You	2.858	Story	-0.274	24	13	11
			Character	-0.299	20	10	10
			Narration	-0.0485	14	8	6
			General	-0.267	23	10	13
7	Why Not Me?	3.677	Story	0.185	15	9	6
			Character	-0.012	12	7	5
			Narration	0.103	21	14	7
			General	0.1336	23	13	10
8	Winter	4.033	Story	0.579	20	16	4
			Character	0.491	19	14	5
			Narration	0.39	14	10	4
			General	0.120	21	14	7
9	An Ember in the Ashes	3.889	Story	0.2671	26	24	2
			Character	0.26	21	18	3
			Narration	0.452	25	21	4
			General	-0.004	23	17	6
10	Six of Crows	4.779	Story	0.825	24	23	1

			Character	0.680	27	27	0
			Narration	0.6637	25	24	1
			General	0.4812	23	22	1
11	A Little Life	2.777	Story	-0.2452	22	4	18
			Character	-0.392	25	6	19
			Narration	-0.19405	25	5	20
			General	-0.160	17	5	12
12	A Darker Shade of Magic	3.830	Story	0.445	24	21	3
			Character	0.3041	25	17	8
			Narration	0.5187	23	18	5
			General	0.3989	18	15	3
13	Queen of Shadows	4.245	Story	0.253	23	18	5
			Character	0.430	20	16	4
			Narration	0.427	17	13	4
			General	0.2552	18	16	2
14	Fates and Furies	2.160	Story	0.0795	22	5	17
			Character	-0.1946	21	3	18
			Narration	-0.571	22	1	21
			General	-0.527	19	5	14
15	Between the World and Me	4.343	Story	-0.0011	23	12	11
			Character	-0.0091	2	0	2
			Narration	0.5829	26	15	11
			General	0.3973	16	10	6
16	Dead Wake	4.171	Story	0.5994	18	16	2
			Character	0.218	10	9	1
			Narration	0.744	24	22	2
			General	0.002	11	8	3
17	The Sword of Summer	4.131	Story	0.1538	20	13	7
			Character	0.162	17	12	5
			Narration	0.515	20	16	4
			General	0.2210	18	14	4