

# Aspect-Oriented Analytics of Big Data

No'aman M. Ali<sup>1,2</sup>[0000-0002-3922-7136]

<sup>1</sup> Saint Petersburg State University,  
Faculty of Mathematics & Mechanics, St. Petersburg, Russia

<sup>2</sup> Port Said University,  
Faculty of Management Technology and Information Systems, Port Said, Egypt  
no3man\_mohamed@himc.psu.edu.eg

**Abstract.** Social media platforms are one of the most significant contributors to big data; it enables consumers to provide their views or opinions about products and services. These abundant reviews contain substantial and valuable knowledge and have become a significant resource for both consumers and firms. Therefore, enterprises seek real-time insights and relevant information on how the market responds to products and services. The proposed framework employs the sentiment analysis and aspect-based sentiment analysis in parallel to customer reviews to support decision-makers regarding Marketing and Manufacturing domains. Our proposal presents a multilayer classifier for consumers' reviews. The first layer is used to categorize reviews into the aspect and non-aspect classes. The second layer is used to break every review involved in the aspect-based category into opinion units based on the product aspects. Next, we plan to measure the polarity of the reviews and opinion units. Finally, we plan to visualize the results in the form of domain-oriented reports. Also, we present a description of the testing and evaluation criteria.

**Keywords:** Big Data Analytics · Sentiment Analysis · Aspect-based Sentiment Analysis · Decision Making.

## 1 Introduction

Traditionally, organizations recognized that the analytics of owned data could broadly improve their business performance through the means of Business Intelligence (BI) [1]. Several decisions making and forecasting domains depend on big data such areas involve business analysis, product development, loyalty, health care, tourism marketing, transportation, etc. Big Data can help organizations to employ smart and effective business decisions by choosing the most appropriate informative strategic direction, increasing operational efficiency, providing better customer service, etc. [2].

Recently, there is a steady increase in customers' desire to express their views or opinions about products and services. These abundant reviews that contain substantial and valuable knowledge become a significant resource for

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

both consumers and firms. Therefore, enterprises seek real-time insights and relevant information on how the market responds to products and services [3–5].

The proposed framework employs the Sentiment Analysis (SA) and Aspect-Based Sentiment Analysis (ABSA) in parallel on customer reviews to support decision-makers regarding the Marketing and Manufacturing domains. Our proposal introduces two classes of reports; Market-oriented reports, and Product-oriented reports as depicted in section 3. Initially, we limit our analysis plan to includes electrical products; in the future, the analysis may expand to include other types of products or services as well.

In this proposal, we present our ongoing research on the developing ABSA model that comprises a multilayer classifier. Also, we explore some of the related works in the section of state of the art. We have designed an implementation and evaluation plan for our research to follow within the next period of the PhD. We desire to develop a solution comparable to other models by following the mentioned plan.

The rest of our proposal has the following structure: in Section 2, we briefly introduce some of the related works and directions regarding SA and ABSA. Section 3 comprises the proposed framework and considerations. In Sections 4 and 5, respectively, we state our research process in the form of clarified steps with identified objectives, as well as describe the desired testing and evaluation scheme.

## 2 State of the Art

The comprehensive development of e-commerce promotes the growing expansion of using online markets via electronic platforms like Amazon, eBay, Walmart, Best Buy, Wish, etc. Besides, the evolution in using social media platforms plays a significant role in encouraging enterprises to give a high priority to analyze users' activities through such platforms [6].

Big data analytics offers various solutions to get insights in real-time and provides valuable information about how the market is responding to products and campaigns [7–12]. Several works introduced for analyzing consumers' reviews to get insights such as sentiment analysis. This technique comprises the automated process of analyzing textual data and classifying opinions, as well as the extraction of properties of reviews like Polarity, subject, and opinion holder [13–15]. On the other hand, aspect-based sentiment analysis examines each review to recognize distinct aspects and identify the corresponding sentiment for each one [16–18]. Unlike sentiment analysis, it enables the association of specific sentiments with various aspects of a product or service [19].

Generally, the use of these techniques enables enterprises to realize how the public feels about something at a particular moment by analyzing their emotions, attitudes, or opinions toward various products or issues. Also, it enables enterprises to track how consumers' opinions change over time. There exist several approaches that are either based on linguistic resources or machine learning [20–23].

Chong, A.Y.L., et al. [24] proposed a combination of sentiment analysis and a neural network to examine the importance of every predictors' variables for online retailers' sales predictions. They used datasets that contain predictor variables like online reviews, consumer sentiments, and online promotional strategies. They observed that retailers could increase sales by specifying "how" and "where" to display online reviews carefully and increasing their social interactions with consumers.

Salehan, M., and Kim, D.J. [25] introduced an approach to discuss the predictors of readership and helpfulness of online consumer reviews (OCR) using sentiment analysis for big data analytics. The presented approach could be adopted by online vendors to develop scalable, automated systems for sorting and classifying of big OCR data that will be useful to both vendors and consumers.

Wallaart, O., and Frasinca, F. [26] proposed a two-stage sentiment analysis algorithm based on ABSA. The introduced algorithm employed a lexicalized domain ontology beside neural networks with a rotatory attention mechanism to work on sentence-level. They applied their model to SemEval-2015, and SemEval-2016 datasets, which include restaurant reviews. They found that machine learning methods can effectively find words that carry sentiment, with different performance and accuracy regarding the given aspect.

Similarly, industrial enterprises seek to analyze user reviews to determine the suitability of the product to their requirements. Besides, to monitor the product life cycle in the market to support sustainable smart manufacturing [27]. Moreover, to develop future strategies for the design of new products in addition to the possibility of offering other versions of existing products after their redesign and to ensure that the problems in the current versions are addressed successfully [28].

### 3 The Proposed Framework

The proposed work strives to support enterprises through exploiting the existence of tremendous amounts of consumer reviews available over social media platforms, electronic markets, etc., by providing decision-makers with oriented feedbacks. The processing of massive amounts of data represents a challenging task due to the diversity of data types and structures that impose difficulties in data integration and storage. Here we plan to make implementation using Apache Hadoop and MapReduce as an open-source framework for distributed storage and processing of data.

The categorization of reviews into aspect-based and non-aspect classes is still a bit tricky task since the identification of entities represents a challenge. Performing a binary classification becomes more appropriate for this task. We plan to apply SA, and ABSA, on the first and second classes, respectively, to assist decision-makers concerning two primary Fields: Marketing and Manufacturing.

Regarding the ABSA, the main task is to extract and identify the entity and attribute pairs. It involves the extraction of opinion units corresponding to the

target entity. Additionally, the recognition of sentiment words and classifying into predefined sets are vital. Here we plan to investigate a new scheme for measure polarity based on fuzzy sets so that each review has a scored polarity. That scheme will be used as well in the second class to assign a membership degree with suitable items from the set.

Finally, we plan to support decision-makers in these fields by providing them with up to date valuable insights in the form of Domain-oriented reports. This task involves the visualization and summarization of data using Python to create clear charts. Also, a comparison will be made between extracted and real attributes to state the missing and required set.

### 3.1 Case Study: An Electric Clothes Iron

*Product Aspects.* The common parts of an electric clothes iron may include Sole plate, pressure plate, heating element, the cover plate, handle, pilot lamp, etc. The main features of choosing electric irons may include Iron surface (e. g. Stainless steel coating, Teflon coating, Ceramic coating), availability of steam, electric power of iron, weight, etc. Product parts represent the number of pieces that come with the original products (e.g., Portfolio, base, extra cable).

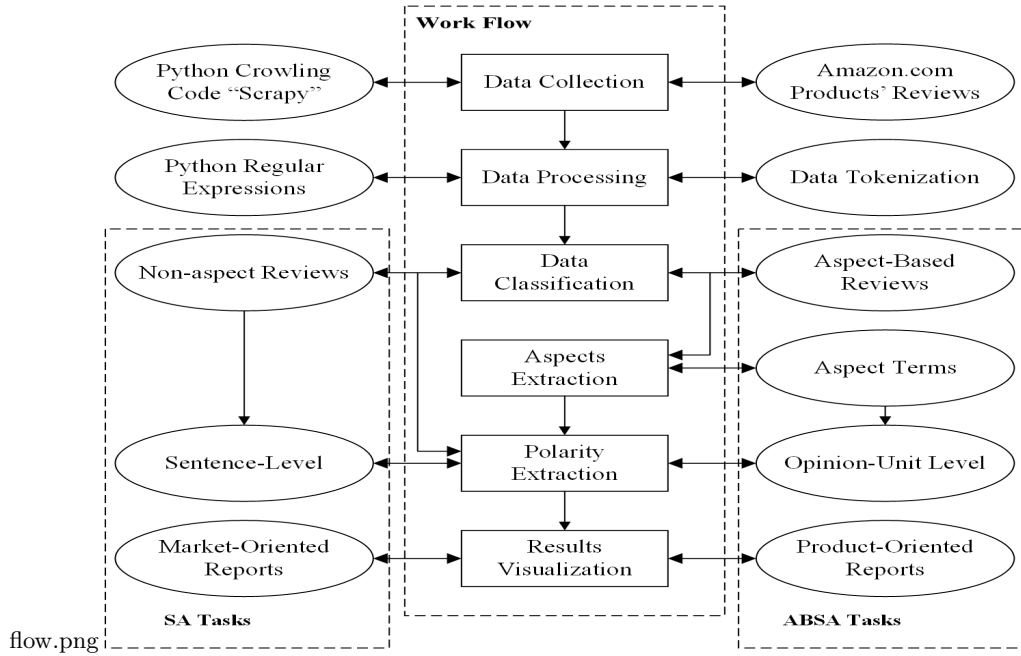
*Sentiment Analysis.* We plan to perform the analysis to measure the overall performance and consumer satisfaction concerning the product by performing SA in conjunction with users' ratings and merge results with the results realized from the next section.

*Aspect-based Sentiment Analysis.* Concerning computing the performance and quality of the distinct components and parts of the product, we plan to extract a list of product aspects. For every list will associate with only one product, which has a unique identifier as well. On the other hand, each item will assign a unique identifier to eliminate repetition, which enables the sharing of one part over several lists of products that have one or more identical parts or features. During the analysis process, we will use the mentioned lists to measure the aspects polarity separately. Next, we investigate the collection of aspects sentiments to conclude the results.

*Output.* Results involve the evaluation of consumers' satisfaction based on their reviews. Reports will state the degree of suitability of a particular component like the handle; indicators may differ among users like comfortable, regular, or hard. Another important part, is the anticipation of the needed parts or features by consumers, in addition to identifying the main competing products for the current release. This information allows the redesign of the product to eliminate disadvantages and the design of new products that meet consumers' needs.

## 4 Research Method

We have identified clear steps that we plan to follow in our proposed work to achieve work objectives, as depicted in Fig. 1:



**Fig. 1.** System design and work flow for the proposed framework.

**Step 1.** Gathering data from various sources like social media platforms, online reviews (e.g., Amazon.com), surveys, etc. We plan to develop a web scraper to collect data from the internet using Scrapy as it is an open-source framework written in Python.

**Step 2.** Preprocessing Data to be suitable for analysis purposes, besides noises cleaning. That involves the process of breaking a stream of text up into words "Tokens," using Python regular expressions.

**Step 3.** Classification of collected reviews into two classes: Aspect-Based reviews and Non-aspect reviews. Such a process concerning the type of analysis to be applied.

**Step 4.** Extracting aspects from reviews, this step concerning the aspect-based analysis only. That involves the identification of every entity attribute pairs (opinion units).

**Step 5.** Extracting and identifying the polarity of sentences by detecting sentiment words. This task will run over two levels: Opinion-Unit level and Sentence level according to the type of analysis.

**Step 6.** Visualization of the results and deliver this feedback to decision-makers. This includes the creation of an easy-to-understand visual report with simultaneous interpretation in a way that everyone in the company can understand. Visualization can represent either a combination of results or a separation for each data source.

## 5 Testing and Evaluation Criteria

Given the proposed framework that states the idea for solving the research problem and objectives to be achieved, in addition to the working mechanism of our research method, we can compose the following plan for testing and evaluation through various steps of work.

**Testing.** Hereabouts we plan to monitor the ongoing processes through the implementation phases. We plan to input a random set of data for every disjoint task to ensure the quality and efficiency of outputs and make a comparison with human-made processing (e.g., Tokenization, classification, aspect extraction, etc.).

**Evaluation.** We plan to get published sales data regarding a particular product during a specific period that suffixes to the period in which collected reviews belong, and compare our results and recommendation with this data to assert the consistency of real data with achieved results. Similarly, concerning the needed features, we plan to survey similar products that have already added these features. On the other hand, to comprehensively evaluate the performance of the proposed work, we desire to experiment with a widely used ABSA dataset; the Laptops and Restaurant datasets of SemEval-16 Track 2 Task 5.

## 6 Summary and Future Work

Producing large amounts of reviews by consumers via expressing their views or opinions about products and services represents a significant resource for both consumers and firms; it contains substantial and valuable knowledge. There are growing interests to analyze such behavior. SA and ABSA are promising approaches to analyze these reviews. Researchers introduced various models to realize this task that comprise a combination with other techniques like neural networks and machine learning.

In this paper, we have outlined our plan of study that aims to develop a consumers' review analysis model for electrical products. We have explored some of the existing works and briefly discussed them. Then, we presented a description of our approach with possible directions and objectives. In the future, we plan to continue our studies by executing the steps outlined in Section 4 with the obligation of evaluation criteria.

**Acknowledgments.** My thanks to my supervisor Prof. Boris Novikov, for his guidance, encouragement, and advice he has provided throughout the previous period of my doctoral studies and is still ongoing. He provided me valuable comments and feedback at various stages of this research.

## References

1. Gandomi, A., Haider, M.: Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management* **35**(2), 137–144 (2015).
2. Wang, H., Xu, Z., Fujita, H., Liu, S.: Towards Felicitous Decision Making: An Overview on Challenges and Trends of Big Data. *Information Sciences* **367-368**, 747–765 (2016). <https://doi.org/10.1016/j.ins.2016.07.007>
3. Hammou, B.A., Lahcen, A.A., Mouline, S.: Towards A Real-Time Processing Framework Based on Improved Distributed Recurrent Neural Network Variants with Fast-Text for Social Big Data Analytics. *Information Processing & Management* **57**(1), 102122 (2020). <https://doi.org/10.1016/j.ipm.2019.102122>
4. Zhao, Z., Wang, J., Sun, H., Liu, Y., Fan, Z., Xuan, F.: What Factors Influence Online Product Sales? Online Reviews, Review System Curation, Online Promotional Marketing and Seller Guarantees Analysis. *IEEE Access* **8**, 3920–3931 (2020).
5. Duan, Y., Edwards, J.S., Dwivedi, Y.K.: Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda. *International Journal of Information Management* **48**, 63–71 (2019).
6. Akter, S., Wamba, S.F.: Big Data Analytics in E-Commerce: A Systematic Review and Agenda for Future Research. *Electronic Markets* **26**(2), 173–194 (2016).
7. Jabbar, A., Akhtar, P., Dani, S.: Real-time Big Data Processing for Instantaneous Marketing Decisions: A Problematization Approach. *Industrial Marketing Management*, (2019). <https://doi.org/10.1016/j.indmarman.2019.09.001>
8. Malhotra, D., Rishi, O.P.: An Intelligent Approach to Design of E-Commerce Metasearch and Ranking System Using Next-Generation Big Data Analytics. *Journal of King Saud University - Computer and Information Sciences*, (2018).
9. Zhaoa, Y., Xu, X., Wang, M.: Predicting Overall Customer Satisfaction: Big Data Evidence From Hotel Online Textual Reviews. *International Journal of Hospitality Management* **76**, 111–121 (2019).
10. Liu, X., Shin, H., Burns, A.C.: Examining the Impact of Luxury Brand’s Social Media Marketing on Customer Engagement: Using Big Data Analytics and Natural Language Processing. *Journal of Business Research*, (2019).
11. Kumar, A., Shankar, R., Aljohani, N.R.: A Big Data Driven Framework for Demand-driven Forecasting with Effects of Marketing-mix Variables. *Industrial Marketing Management*, (2019). <https://doi.org/10.1016/j.indmarman.2019.05.003>
12. Zheng, K., Zhang, Z., Song, B.: E-Commerce Logistics Distribution Mode in Big-Data Context: A Case Analysis of JD.COM. *Industrial Marketing Management*, (2019). <https://doi.org/10.1016/j.indmarman.2019.10.009>
13. Taylor, E.M., O., C.R., Velásquez, J.D., Ghosh, G., Banerjee, S.: Web Opinion Mining and Sentimental Analysis. In: Velásquez, J.D., Palade, V., Jain, L.C. (eds.) *Advanced Techniques in Web Intelligence-2: Web User Browsing Behaviour and Preference Analysis*, pp. 105–126. Springer, Berlin, Heidelberg (2013).
14. Ramanujam, R.S., Nancyamala, R., Nivedha, J., Kokila, J.: Sentiment Analysis Using Big Data. In: *International Conference on Computation of Power, Energy, Information and Commuincation (ICCPEIC)*, pp. 480–484. IEEE, Chennai, India (2015). <https://doi.org/10.1109/ICCPEIC.2015.7259528>

15. Yadav, K., Rautaray, S.S., Pandey, M.: A Prototype for Sentiment Analysis Using Big Data Tools. In: Mandal, J.K., Dutta, P., Mukhopadhyay, S. (eds.) First International Conference on Computational Intelligence, Communications, and Business Analytics (CICBA 2017), vol. 775, pp. 103–117. Springer, Singapore (2017).
16. Ma, Y., Peng, H., Cambria, E.: Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. In: The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), pp. 5876–5883. Association for the Advancement of Artificial Intelligence, New Orleans, Louisiana, USA (2018).
17. Liu, N., Shen, B., Zhang, Z., Zhang, Z., Mi, K.: Attention-based Sentiment Reasoner for Aspect-based Sentiment Analysis. *Human-centric Computing and Information Sciences* **9**(35), 17 (2019). <https://doi.org/10.1186/s13673-019-0196-3>
18. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In: Burstein, J., Doran, C., Solorio, T. (eds.) 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), vol. 1, pp. 380–385. Association for Computational Linguistics, MN, USA (2019).
19. Bandari, S., Bulusu, V.V.: Survey on Ontology-Based Sentiment Analysis of Customer Reviews for Products and Services. *Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19*, pp. 91–101. Springer (2020).
20. Yang, H., Zeng, B., Yang, J., Song, Y., Xu, R.: A Multi-task Learning Model for Chinese-oriented Aspect Polarity Classification and Aspect Term Extraction. *arXiv e-prints*, (2019).
21. See-To, E.W.K., Ngai, E.W.T.: Customer Reviews for Demand Distribution and Sales Nowcasting: A Big Data Approach. *Annals of Operations Research* **270**(1-2), 415–431 (2018). <https://doi.org/10.1007/s10479-016-2296-z>
22. Schouten, K., Frasincar, F.: Ontology-Driven Sentiment Analysis of Product and Service Aspects. *The Semantic Web: 15th International Conference, ESWC 2018*, pp. 608–623. Springer International Publishing, Cham (2018).
23. Ghosh, M., Sanyal, G.: An Ensemble Approach to Stabilize the Features for Multi-Domain Sentiment Analysis Using Supervised Machine Learning. *Journal of Big Data* **5**(1), 44 (2018). <https://doi.org/10.1186/s40537-018-0152-5>
24. Chong, A.Y.L., Li, B., Ngai, E.W.T., Ch'ng, E., Lee, F.: Predicting Online Product Sales Via Online Reviews, Sentiments, and Promotion Strategies: A Big Data Architecture and Neural Network Approach. *International Journal of Operations & Production Management* **36**(4), 358–383 (2016).
25. Salehan, M., Kim, D.J.: Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytics. *Decision Support Systems* **81**, 30–40 (2016). <https://doi.org/10.1016/j.dss.2015.10.006>
26. Wallaart, O., Frasincar, F.: A Hybrid Approach for Aspect-Based Sentiment Analysis Using a Lexicalized Domain Ontology and Attentional Neural Models. In: Hitzler, P.D.P., Fernández, M., Janowicz, K., Zaveri, A., Gray, A.J.G., Lopez, V., Haller, A., Hammar, K. (eds.) *The Semantic Web: 16th International Conference, ESWC 2019*, pp. 363–378. Springer International Publishing (2019).
27. Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., Almeida, C.M.V.B.: A Comprehensive Review of Big Data Analytics Throughout Product Lifecycle to Support Sustainable Smart Manufacturing: A Framework, Challenges and Future Research Directions. *Journal of Cleaner Production* **210**, 1343–1365 (2019).
28. Rehman, M.H.u., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P.P., Perera, C.: The Role of Big Data Analytics in Industrial Internet of Things. *Future Generation Computer Systems* **99**, 247–259 (2019). <https://doi.org/10.1016/j.future.2019.04.020>