ANALISA OBROLAN PELANGGAN MENGGUNAKAN METODE TEXT MINING UNTUK MANAJEMEN HUBUNGAN PELANGGAN (STUDI KASUS: PAPERLUST.CO)

ANALYSIS OF CUSTOMER CHAT USING TEXT MINING FOR CUSTOMER RELATIONSHIP MANAGEMENT (CASE STUDY: PAPERLUST.CO)

Muhammad Apriandito Arya Saputra [1], Andry Alamsyah [2]
Undergraduate of International ICT Business Major, School of Economics and Business, Telkom University
[1]m.apriandito@gmail.com, [2]andrya@telkomuniversity.ac.id

Abstrak

Di era Industry 4.0, persaingan di dunia bisnis semakin menigkat dan mendorong setiap perusahaan untuk selalu berkembang dan beradaptasi dengan perubahan untuk memenangkan persaingan. Pesatnya perkembangan teknologi informasi telah mengubah banyak kegiatan bisnis, salah satunya adalah bisnis berbasis internet. Aplikasi praktis dari teknologi internet dalam dunia bisnis kini semakin berkembang, yaitu perdagangan di dunia maya dengan memanfaatkan internet yang sering disebut dengan e-commerce (perdagangan elektronik)..Penelitian ini bertujuan untuk menganalisis obrolan pelanggan dari Paprlust.Co menggunakan Metode Text Mining. Dalam penelitian ini metode yang digunakan adalah analisis sentimen yang digunakan untuk mengetahui persepsi pelanggan / emosi dan metode pemodelan topik untuk mengekstraksi topik atau masalah yang dibahas dari setiap kelas analisis sentimen berdasarkan data obrolan pelanggan di platform live chat Paperlust. Semoga penelitian ini bermanfaat untuk informasi dalam industri e-commerce khususnya Paperlust.co sebagai wawasan untuk mengevaluasi dan meningkatkan bisnis mereka untuk memberikan nilai lebih dan memperoleh lebih banyak pendapatan.

Kata kunci — Analisia Sentimen; Pemodelan Topik; Manajemen Hubungan Pelanggan

Abstract

In the era of Industry 4.0, the competition in the business world is increasing and encouraging each company to always develop and adapt with the change to win the competition. The rapid development of information technology has changed a lot of business activities, on of it is the internet-based business. The practical application of internet technology in the business world is now growing, namely trading in the virtual world by utilizing the internet which is often referred to as e-commerce (electronic commerce). E-commerce business has changed many things in the business world, things that become the focus not only on goods but on services, information and intelligence focus on customers. This research purposed to analysis customer chat of Paperlust.Co using Text Mining Method. In this study the method used is sentiment analysis used to find out customers perception/emotion and topic modelling methods to extract topics or issue discussed from each class of sentiment analysis based on customer chat data in Paperlust's live chat platform. This research hopefully useful for information in e-commerce industry especially Paperlust.co as insight to evaluate and improve their business to deliver more value and gained more revenue.

Keywords—Text Mining; Sentiment Analysis; Topic Modelling; Costumer Relationship Management

1. Introduction

In the era of Industry 4.0 , the competition in the business world is increasing and encouraging each company to always develop and adapt with the change to win the competition. The rapid development of information technology has changed a lot of business activities, on of it is the internet-based business referred as e-commerce (electronic commerce). E-commerce business is a means of commercial transactions between and between organizations and individuals digitally with the use of the internet, web and online business applications [1].

Customer Relationship Management (CRM) is a strategy, which focuses on building and maintaining customer relationships that enable customer loyalty.

Paperlust.Co is one of the e-commerce that focused on online printing. Data that are collected from Paperlust's Live Chat Platform is unstructured data, and to extract the information it necessary use Text Mining Methodology. Text Data Mining or Knowledge-Discovery in Text (KDT), refers generally to the process of extracting non-trivial information and knowledge from unstructured text [2] Sentiment Analysis and Topic Model are model of Text Mining Methodology. Sentiment analysis, is used to analyze feelings , evaluation, acceptance, attitudes, and emotions of individuals about the discussion of phenomena such as products, services, organizations or general issues that are being circulated. Focused sentiment analysis expresses public opinion or implies positive and negative impressions of an opinion [3]. This method is also able to measure the quality of service of the company and specifically can also measure public sentiment on each performance feature of the existing company [4]. Topic modeling using the LDA alogrithm can extract important topics in a large document text. Topic modelling methods are generally used for automatically organizing, understanding, searching, and summarizing large electronic archives. Topic models discover the hidden themes through out the collection and annotate the documents according to those themes. [5]

2. Literature Review

2.1 Customer Relationship Management (CRM)

Customer Relationship Management is a comprehensive approach for creating, maintaining and expanding customer relationships. Customer Relationship Management based on e-commerce systems is used to increase customer satisfaction to increase sales.

2.2 E-Commerce

E-commerce is defined as the use of the internet and the web for business transactions. Formally, it allows digital commercial transactions between organizations and individuals.

2.3 Data Mining

Data mining is a term used to discover hidden knowledge in a database. Data mining is a semi-automatic process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify potentially useful information that stored in large databases. [6]

2.4 Text Mining

Text mining is an interdisciplinary field that draws on information retrieval, data mining, machine learning, statistics, and computational linguistics. A substantial portion of information.

2.5 Machine Learning

Machine learning investigates how computer scan learn (or improve their performance) based on data. A main research area is for computer programs to automatically learn to recognize complex patterns and make intelligent decisions based on data. For example, a typical machine learning problem is to program a computer so that it can automatically recognize handwritten postal codes on mail after learning from a set of examples

2.6 Sentiment Analysis

Sentiment analysis is also called opinion mining studies that examine the opinions, sentiments, evaluations, acceptance, attitudes, and emotions of people on an entity such as products, services, organizations, individuals, issues, events, topics and other attributes that represent a considerable problem. Analysis sentiment and opinion mining focus on opinions that elicit positive or negative sentiment responses. Sentiment Analysis is often referred to as Opinion Mining, since the results of the analysis come from the opinions or attitudes of the person.

2.7 Topic Modelling

Topic modelling is an important research direction in machine learning, natural language processing and information retrieval. Topic modelling is the process of identifying the underlying semantic structure of a document with the use of a hierarchical Bayesian analysis on the collection of documents.

2.8 Research Framework

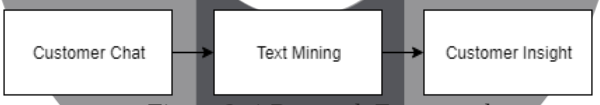

Figure 2. 1 Research Framework
Source: Author

Data received by the Paperlust's.Co Live Chat Platform will be extracted using the text mining method. This result is expected to be emotional data from each customer and also expected by using topic modeling method of important points from each of the sentiment levels can be extracted. These insight results are expected to become information for Paperlust.Co to Evaluate and Improve the quality services in accordance with the Customer Relationship Management (CRM) strategy.

3. Theoretical Framework

3.1 Research Stages

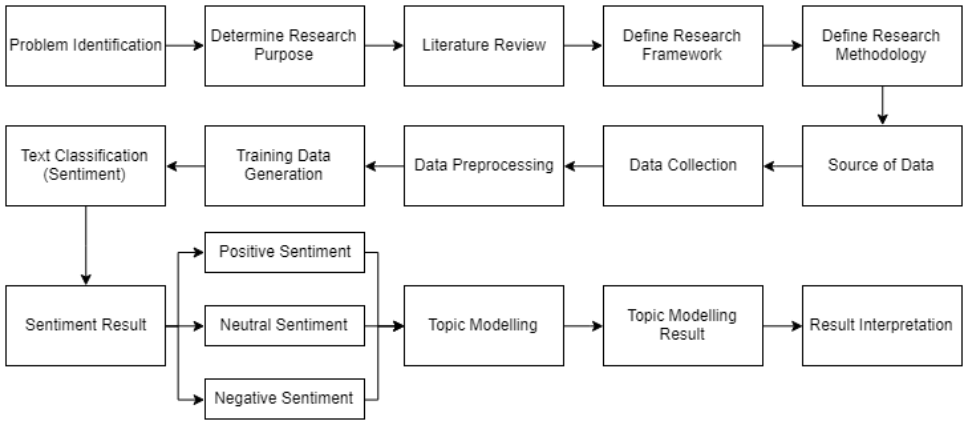

Figure 3. 1 Research Stages
Source: Author

3.2 Data Analysis Technique

After all data are collected properly from Paperlust.Co data server, there are two major stages that will be accomplished, which is Sentiment Analysis and Topic Modelling. In Sentiment Analysis the process will be, pre-processing, creation of training data [7]. Techniques to be used are Naïve Bayes Classifier. Then the text that contain sentiment will be separated to different file, and use topic modelling to uncover the topic in each sentiment dataset.

3.3 Naïve Bayes

Naive Bayes is one of the simplest and commonly used classification methods in research. This classification calculates the probability of a class based on the distribution of words in a document. This classification uses the Bayes theorem to predict the probability of features that have been labeled. Some empirical studies suggest that when Naïve Bayes compared to the decision tree and neural network has higher accuracy and speed when applied to large databases. NBC assumes that the effect of attribute values in that class is independent of the other attributes.

3.4 Text Performance Evaluation

Text Performance Evaluation is conducted in order to test the results that have been done by the machine. One of the measurements is to use confusion matrix. Predicted label is the number of labels predicted by the researchers before being processed in the machine. While the actual label is the result of the number of labels either positive or negative. Then True Positive (TP) is a class that is predicted to be positive and classified positively by the machine, True Negative (TN) is a class that is predicted to be negative and classified negatively by the machine, False Positive (FP) is a class that is predicted to be positive and classified otherwise and False Negative FP) is a class that is predicted to be negative and classified otherwise. The results obtained from the matrix will be tested in the formula

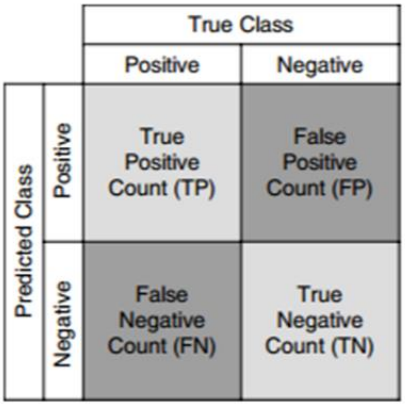

Figure 3. 2 Confusion Matrix [12]

4. Result and Disscussion

4.1 Text Performance Evaluation

After performing various processes, the training data and test data that have been made under the terms is 30:70 or if detailed, 30 percent is training data and 70 percent is the test data that has the highest accuracy, precision and recall if classified.

Table 4.1 Text Performance Evaluation

Accuracy	Relative Error	Kappa
81.18%	18.82%	0.607

4.2 Sentiment Analysis

As shown in Figure 4.1 from all the chat data that received into Paperlust's live chat platform. 67% of the chat have neutral sentiment, 8% have negative sentiment, 25% have positive sentiment. The chat was dominated by neutral sentiment chat, and the percentage of positive chat is bigger that percentage of negative chat.

Table 4.1 Sentiment Analysis Result

Positive Chat	Neutral Chat	Negative Chat
3713	10091	1212

4.3 Topic Modelling

To extract information of majority topic and point in each sentiment class, the author separate sentiment analysis result into 2 files (positive and negative). Neutral classified text is not used since it contains a lot of data noises (e.g. no sentiment words, etc.). From the result, we use saliency (term's frequency) measurement method, these quantities measure on how much information a term conveys about topics by computing the Kullback-Liebler divergence between the distribution of topics given the term and optionally weighted by the term's overall frequency. (Sievert & Shirley, 2014). For a thorough overview of the dataset, the conduct topic modelling for the whole dataset, positive and negative respectively to capture the whole topic in the dataset. 5 major topic for each sentiment class (positive and negative) will be explained in this analysis. Light blue color represent the overall term frequency, while the red one represent estimated term frequency within the selected topic.

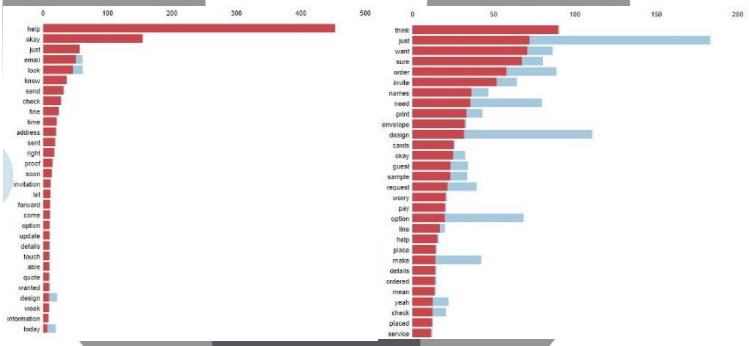

Figure 4.1 Largest Topic In Positive and Negative Sentiment

As shown in Figure 4.15, the largest dominant topic in the positive sentiment document is about the customer ask Customer Support team for help to update the information by email and check everything are fine. The largest dominant topic in the negative sentiment document is about the customer was not understand and wonder the look of colour gold, white, and blue foil in invitation card.

5. Result and Disscussion

5.1 Conclusion

Author has successfully implemented to properly classify, analyze, and summarize thousands of chats in Paperlust.Co Live Chat Platform. In this study customer chat analysis, Naïve Bayes method is suitable for text analysis and Latent Dirichlet Allocation is a proper method that has proved its capability in extracting in-depth insight about the topics discussed in the large-scale dataset. Based on the author opinion, the method that the author use in this research are better in term of a near real-time processing capability compared to the traditional way.

5.2 Suggestion

5.2.1 Theoritical Aspect

For the future research, the author wishes to add more data to enrich the training data, so the result will be more accurate. Then the tools to do data mining and data analyzing is using a full open source edition. The author also hopes that this research is expanded to multi-label classification since one customer contain more than one sentiment or dimension

5.2.2 Practical Aspect for Industry

Based on the testing result and discussion which is done in this research, the author suggests for Paperlust.Co to understand the issue occur in positive and negative sentiment. Paperlust.Co can create a better strategy for their CRM strategy, adapting to a new condition about what their customers want, delivering a better values/services, and at last gained more profit. From the conclusion above, we conclude that Paperlust.co need to improve on website usability where error or bugs might occur and interfere with user experience.

REFERENCES

[1] Laudon, K. C., & Traver, C. G. (2014). *E-Commerce 2014, 10th Edition.* Pearson.

[2] Moreno, A., & Redondo, T. (2015). Text Analytics: the convergence of Big Data InternatIonal Journal of InteractIve MultIMedIa and artIfIcIal IntellIgence, 57-63.

[3] Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publisher.

[4] Alamsyah, A., Rahmah, W., & Irawan, H. (2015). Sentiment Analysis Based On Appraisal Theory For Marketing Intelligence In Indonesia's Mobile Phone Market. *Journal of Theoretical and Applied Information Technology*.

[5] Tong, Z., & Zang, H. (2016). A Text Mining Research Based on LDA Topic Modelling. *The Sixth International Conference on Computer Science, Engineering and Information Technology.* Canada.

[6] Turban, Efraim, Strauss, Judy, Lai, & Linda. (2016). *Social Commerce.* Springer.

[7] Thakkar, H., & Patel, D. (2015). Approaches for Sentiment Analysis on Twitter: A State-of-Art study.