Improving luxury hotel service and resource management: A sentiment analysis approach.

BUSN9200 Research Proposal

Name: Charlie Martin

ID – CAM205

Word Count – 1097

Contents

Introduction

In their study of hospitality and tourism online reviews, Liu and Law (2015) found that online reviews play a critical role in sales for the industry. Similarly, Kaplan and Haenlein (2010) concluded that online reviews have become the word of mouth of the digital age due to being up to date and reliable. As Duan, Gu, and Whinston (2008) confirmed, word of mouth has long been recognised as the most influential resource of transmission in society. Evidently, online reviews are a crucial resource for hotels and extracting maximum information from them can provide a competitive edge, but extracting meaningful insights from a catalogue of unorganised personal experiences is no mean feat. This project will provide a framework for extracting insights from online reviews with a sentiment analysis model. Various models will be discussed and analysed for accuracy and suitability. Additionally, a new perspective will be suggested for applying said insights specifically to the hotel industry, where both hotel and reviewer attributes are considered. This proposal will have a short literature review, followed by the outlined research questions and proposed methods, limitations, and timeline.

Relevant Literature

Sentiment analysis is one of the fastest growing research areas in computer science. As Mäntylä, Graziotin, and Kuutila (2018) found, the availability of subjective texts online has caused an outbreak in computer-based sentiment analysis, with 99% of related papers being published after 2004. This is further supported by Pang, Lee and Vaithyanathan (2002) who found that a crucial characteristic of the rapidly growing online discussion sites is the overall sentiment. This study also found that sentiment classification can be beneficial specifically in business intelligence applications, with machine learning techniques out-performing human baselines. Clearly, sentiment analysis is a promising new field and machine learning techniques deployed have promising results for extracting insights. One of challenges of sentiment analysis is the feature extraction itself.

Asgher, M. et al (2014) found in a review of the literature, feature selection is usually categorized within 4 types, with NLP (natural language processing) being significantly more frequent. This study also concluded that proper feature selection techniques play a significant role in identifying relevant attributes and increasing accuracy. This is taken further by Ahuja, R. et al (2019), whose extensive study and testing of feature extraction methods and machine learning algorithms found that TF-IDF (a NLP based extraction method) performed on average 3-4% better. This study also concluded that logistic regression algorithms were best for sentiment analysis, which has also been concluded in studies such as that of Prabhat and Khullar (2017). However, Hasanli and Rustamov (2019) found that logistic regression only performed better with TF-IDF feature selection, and bag of words approach resulted in a Naïve Bayes algorithm having higher accuracy, which suggests further study may be needed. It is well established in the literature that sentiment analysis has great potential for business intelligence applications, and there are variations of sentiment analysis models that are favourable for increased accuracy. Next, practical applications to the hotel industry.

As mentioned, it is accepted that online reviews are of increasing importance, particularly to the hotel and tourism industry, and sentiment analysis may have the ability to unlock the potential of a growing resource. Regarding sentiment analysis for the hotel industry, several studies have been

undertaken. These studies generally outline key features for hotel managers to focus on to improve service. For example, Tran, Ba and Huynh (2019) found that location and restaurants were the 2 most important aspects for good reviews, while Sodanil (2016) found that the room was significantly more important than a restaurant. Most notable is a study by Chang et al (2020) on luxury hotel reviews. As mentioned in this study, customers for luxury hotels tend to focus more on details as upscale amenities are expected, making these reviews the ideal dataset for such analysis. It is mentioned in this study that future research may benefit from further investigation of customer profiles, which is a gap in the literature of this field in general. Hence, the goal of this project to provide some insight.

This is important to know particularly for the hotel industry as they are in the unique position of seeing customers profiles before they arrive, allowing for opportunity to manage resources accordingly, namely room allocations. Room allocation is noted as being particularly important and various complex models exist, even being referred to as a statistics game by Song, Parla and Yuan (2010). Identifying trends in reviews from customer categories could provide hotel managers with a quick and practical framework for allocating rooms to maximise good reviews, thereby maximising revenue.

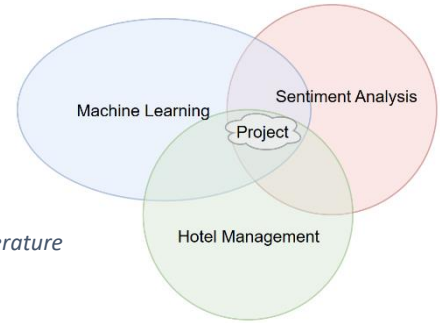


Figure 1 - Placement of study in literature

Research Question and Hypotheses

This project has a variety of research questions and associated hypotheses. The primary hypotheses are:

- H0 – Reviewer attributes have no impact on review sentiment.
- H1 – Reviewer attributes impact review sentiment.

Key Research Questions

- Most accurate feature extraction method and machine learning algorithm combination?
- Key attributes in positive and negative reviews, how can hotels manage this?
- How much do some common guest attributes impact overall sentiment?

Other Hypotheses

- Location and room are most important for positive reviews.
- Customer service is the biggest indicator for negative reviews.
- Bookings being ‘leisure’ or ‘business’ will significantly impact review sentiment.

Proposed Methods

This project will adopt a quantitative, positivist approach and follow a deductive reasoning pathway. The data analysis will be conducted with machine learning algorithms, which give quantitative results in numbers. Naturally, a qualitative approach would not be suitable here, although the raw data is in qualitative reviews, these will be converted to quantitative measures in the analysis for more reliable results.

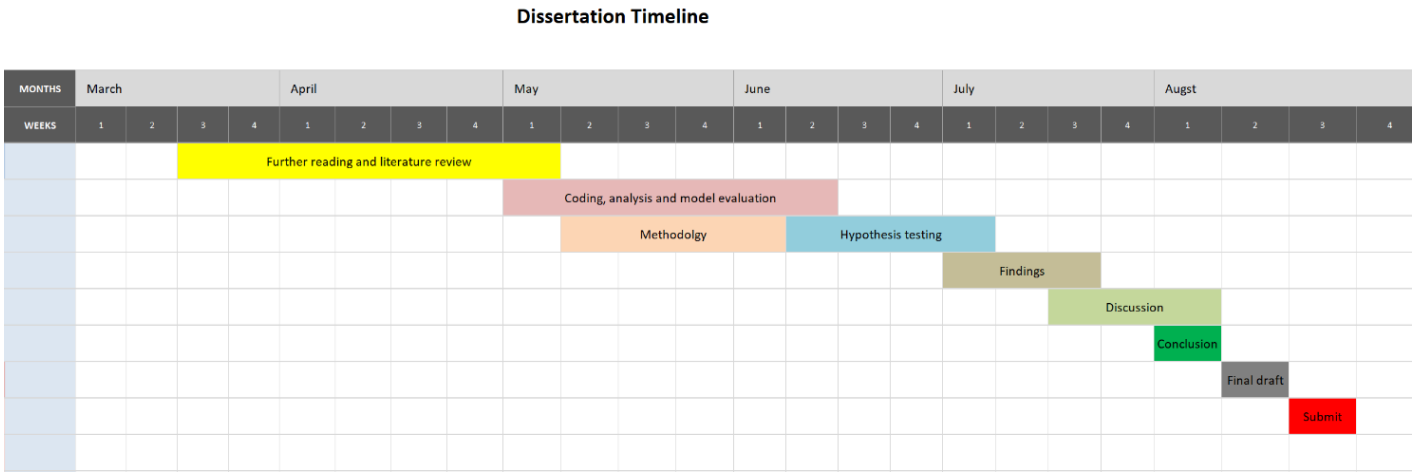
The data for this study is a public, secondary dataset of around 500,000 reviews for luxury hotels in Europe. As the dataset is public and contains no sensitive data, there is no ethical issues regarding its analysis. Potential ethical issues for future work are discussed in the potential limitations.

The text reviews from the dataset will be used for feature extraction and primary sentiment analysis. Additionally, the review 'tags' (Couple, solo travel, leisure etc) will be further analysed for novel insights.

Potential Limitations

The biggest concern with this project is that the data available for customer profiles is not detailed enough to provide a significant result or valuable insight, which may reduce the reliability and validity of any conclusions. Naturally, when customer data is in question, there are various ethical concerns regarding privacy. Due to this, the customer information available is not personal or particularly detailed. This would also be a concern for further research as naturally for better results more data would be needed of greater detail, which should not be collected without expressed permission and an emphasis on privacy and data security.

Project Timeline



References

Ahuja, R., Chug, A., Kohli, S., Gupta, S. and Ahuja, P., 2019. The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*, pp.341-348.

Asghar, M.Z., Khan, A., Ahmad, S. and Kundi, F.M., 2014. A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, *4*(3), pp.181-186.

Chang, V., Liu, L., Xu, Q., Li, T. and Hsu, C.H., 2023. An improved model for sentiment analysis on luxury hotel review. *Expert Systems*, *40*(2), p.e12580.

Duan, W., Gu, B. and Whinston, A.B., 2008. Do online reviews matter?—An empirical investigation of panel data. *Decision support systems*, *45*(4), pp.1007-1016.

Hasanli, H. and Rustamov, S., 2019, October. Sentiment analysis of Azerbaijani twits using logistic regression, Naive Bayes and SVM. In *2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-7). IEEE.

Kaplan, Andreas M., and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media." *Business horizons* 53, no. 1 (2010): 59-68.

Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, *27*, pp.16-32.

Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Prabhat, A. and Khullar, V., 2017, January. Sentiment classification on big data using Naïve Bayes and logistic regression. In *2017 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE.

Schuckert, M., Liu, X. and Law, R., 2015. Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, *32*(5), pp.608-621.

Sodanil, M., 2016. Multi-language sentiment analysis for hotel reviews. In *MATEC Web of Conferences* (Vol. 75, p. 03002). EDP Sciences.

Song, J., Parlar, M. and Yuan, Q., 2010, August. Decision Making of Hotel Room Allocation: A Statistic Game. In *2010 International Conference on Internet Technology and Applications* (pp. 1-4). IEEE.

Tran, T., Ba, H. and Huynh, V.N., 2019. Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In *Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, March 27–29, 2019, Proceedings 7* (pp. 393-405). Springer International Publishing.