

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:

2019-07-15

Deposited version:

Post-print

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Guerreiro, J. & Rita, P. (2019). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*. N/A

Further information on publisher's website:

10.1016/j.jhtm.2019.07.001

Publisher's copyright statement:

This is the peer reviewed version of the following article: Guerreiro, J. & Rita, P. (2019). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*. N/A, which has been published in final form at <https://dx.doi.org/10.1016/j.jhtm.2019.07.001>. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

How to Predict Explicit Recommendations in Online Reviews Using Text Mining and Sentiment Analysis

Abstract

Opinions shared by peer travelers help tourists decrease the risks of making a poor decision. However, the increasing number of reviews per experience makes it difficult to read all reviews for an informed decision. Therefore, reviewers who make a personal and explicit recommendation of the services by using expressions such as “I highly recommend” or “don’t recommend” may help consumers in their decision-making process. Such reviews suggest that the reviewer was satisfied to a point that (s)he would advise others to try or was unsatisfied and will for sure avoid coming back. The current research note explores what may drive reviewers to make direct endorsements in text. A text mining method was applied to online reviews to identify drivers of explicit recommendations. Lack of competences from the provider and negative attitudes are triggers of negative direct recommendations, whereas positive feelings predict a positive recommendation in the body of the review.

Keywords

eWOM; Online Recommendations; Text Mining; CART

1. INTRODUCTION

Travelers increasingly rely on online opinions, rather than promotional ads or guides, to choose experiences more aligned with their own interests and to know what to expect when they reach their destination (Litvin *et al.*, 2005). If not guided by previous experiences and consumer ratings, choosing a restaurant in a new city, for example, is typically a challenging task. Although previous research has examined the role of direct

recommendations on purchase intention (Ordenes *et al.*, 2017; Packard & Berger, 2017), and Xiang *et al.* (2015) suggested that direct recommendations may be important markers to explain satisfaction ratings, no studies have explored what may drive such explicit recommendations. When positive, such direct language has a strong persuasion effect on consumers by showing that endorsers not only enjoy the experience but have also high domain expertise (Packard & Berger, 2017). However, a negative recommendation usually leads to a considerable decrease in purchases (Chevalier & Mayzlin, 2006).

The current research note uses a text mining technique based on lexicon-based approach to guide the search for factors within text that may explain explicit recommendations as dependent variable (Medhat, Hassan, & Korashy, 2014). As a theoretical contribution, the current paper shows what drives a consumer to write a review encouraging the reader to try a given experience. The managerial contributions obtained allow marketers to focus their efforts on those reviewers with the highest probability to recommend.

2. LITERATURE REVIEW

Online recommendations are one of the most important factors in consumers' decision-making processes, as consumers believe that the opinions of their peers are more reliable than those posted by the service provider (Filieri *et al.*, 2015). A positive review often helps consumers avoid choosing a restaurant or hotel that might not meet their expectations (Bronner & de Hoog, 2011). Reviewers who gather the largest number of fans are typically those who alert their followers either to things that should be avoided or to the most pleasurable items to pursue (Guerreiro & Moro, 2017).

Online reviews are generally free from rules dictating mandatory topics, and reviewers may choose to write in a way that best represents their experience (Wattanacharoensil *et al.*, 2017). Reviewers are free to take a negative or positive tone when they discuss service attributes, such as their feelings about the experience, the attitude of the staff and how much time they waited to be served. However, few reviews explicitly recommend a service to readers, though sentences such as “I highly recommend this restaurant” or “I will certainly come back” are sometimes included in reviews to show peers that the service is a worthy experience. Such endorsements were recently studied in the literature as drivers of self-reported star-ratings (Ordenes *et al.*, 2017) and purchase intention (Packard & Berger, 2017). Although there is no clear agreement as to whether a direct recommendation is an implicit or an explicit expression, there is an agreement that it is a directive act that strongly affects consumer decisions. The current research note grounds the definition of an explicit recommendation on the “speaker’s declaration that the object is appropriate for others” (Packard & Berger, 2017: 573). However, consumer knowledge and awareness are known to affect explicit knowledge, in a way that such knowledge is reflected on how reviewers rate their experiences. Attitudes are important markers of consumers’ satisfaction, particularly the attitudes of consumers towards the disposition of the establishment’s staff. For example, words such as “attentive” and “courteous” are more often used to describe hotels with a five-star rating than lower ranked hotels (Stringam & Gerdes, 2010). Previous research has also found that feelings play an important role in making consumers satisfied (Briggs, *et al.* 2007), and competence and functioning may also affect reviewer ratings, with the former being more associated with the competence of the staff and the latter to facilities and other functional attributes (Pantelidis, 2010). Price (budget) is more often mentioned in comments with higher rankings (Stringam & Gerdes, 2010). The examples given above suggest that there

may be markers in the text that not only predict satisfaction but also convey into an explicit recommendation, especially if reviewers have a high degree of expertise.

Understanding the motivations behind explicit recommendations may help managers selectively improve their offerings to better meet consumers' expectations. Nevertheless, due to the unstructured nature of online reviews, it is difficult to construct a model that uses text to predict the antecedents of explicit recommendations without using formal methods to structure such information. Text mining has been used successfully in recent studies (Calheiros *et al.*, 2017) to overcome such challenges, and this method was used in the current research.

3. METHOD

Reviews were analyzed using a text mining approach based on natural language processing (NLP), thus, words in the lexicon were labeled (positive or negative) according to their semantic context (Collobert *et al.*, 2011). Unlike traditional qualitative research methods, text mining automatically structures text into groups of words based on contextual and semantic information (Aggarwal, 2012). IBM SPSS Modeler Text Analytics was used as a sentiment analysis tool (IBM, 2019a). The extraction and transformation of the text was done using the SPSS Modeler Text Analytics extraction engine. Data was cleaned by removing non-linguistic entities such as phone numbers, social security numbers, percentages and http addresses. Punctuation errors were also removed from the text. Then, the extraction engine used Natural Language Processing (NLP) to identify *uniterms* and *multiterms*, removing all irrelevant *stopwords* and grouping terms together using stemming and lemmatization procedures. The tool has a general dictionary with a part-of-speech code for each term (e.g. noun, verb, adjective, adverb). Such base dictionary is then complemented with synonyms to come up with a larger dictionary that may be applied to classify and label the text. A set of lexicons were

used to group uni-terms or multi-terms into 17 different categories: seven categories with positive connotations, seven with negative connotations and three additional markers classifying wait-time and customer support (see Appendix 1). The current paper used the base dictionaries of IBM SPSS Modeler that grouped terms in categories such as Attitudes, Feelings, Competences, Functioning, Budget, and others. This paper uses only the categories that were found to be relevant to explain recommendation according to the above literature review. The part-of-speech codes for each term were used to identify uni-terms or multi-terms. For example adjectives that were followed by a noun were grouped together as a multi-term (e.g. “friendly staff”). After identifying the terms, the extraction engine labelled them as positive or negative according to their semantic information (IBM, 2019b).

The sample was selected from the Academic Yelp Dataset that contained reviews from users of multiple types of sector industries. After an initial screening, a random sample of 1,112,708 reviews from 47,263 different restaurants in 661 different cities written between 2004 and 2017, was extracted from the Yelp dataset (Yelp, 2017).

Each review was classified according to the sentiment categories in Appendix 1, table 3. Reviews were classified as 1 or 0, depending on the presence or absence, respectively, of n-grams that related to positive or negative attitudes, feelings, competences, functioning, budget, wait time and customer support. For example, all reviews in which reviewers complained about the waiting time in the restaurant were classified as such using terms like “wait in line” and “queue,” while positive and negative terms related to budget (e.g. cheap, affordable) and competence (e.g., able to resolve, efficient) were also binarily classified according to those different markers. Sentiment markers were used to classify explicit recommendations—for example, the n-grams “come back” and “give them a try”—indicating positive recommendations (classified as “recommended”) and

terms like “can’t recommend” and “do not plan to return,” indicating negative recommendations (classified as “not recommended”).

From the total sample, 96,257 (71.44%) reviews contained positive recommendations, 38,489 (28.56%) contained negative recommendations. Such sample follows a distribution skewed toward positive sentiments, which is aligned with past empirical findings (Xiang, *et al.*, 2017).

The data was divided into training (70% random sample) and testing sets (30% random sample) due to its effectiveness in building the best classification models (Sarkar, 2016). The training dataset was balanced between positive and negative recommendations to evaluate the model regarding its ability to predict tourist recommendations. The final training dataset had a total of 26,948 reviews with a positive recommendation (49.99%) and a total of 26,959 reviews with a negative recommendation (50.01%). An imbalanced dataset in the training set is known to affect modeling performance, resulting in worse classifications (Chawla *et al.*, 2004). Thus, balancing data constitutes a valid approach to address this issue (He & Garcia, 2008). The models’ accuracy was then tested against an unbalanced testing dataset with 28,887 positive recommendations (71.47%) and 11,530 negative recommendations (28.53%).

Supervised algorithms designed for large datasets were used to predict explicit recommendations in the text as they have successfully predicted variables of large datasets in the past (Wu *et al.*, 2008). Five algorithms were tested on the data. A Probit algorithm, a binomial logistic algorithm, and three decision tree algorithms (CHAID, C&RT and Random Forest algorithm). Table 1 shows the accuracy results from the four models.

-- PLACE TABLE 1 AROUND HERE --

All five algorithms present a similar accuracy and thus, binary logistic regression and CHAID were selected to show the results. The Binomial logistic regression has the advantage over the Probit model of being less complex in terms of interpretation, being computationally efficient on large datasets and in this case having the same accuracy of the Probit model. Although random forests fit the data marginally better its complexity reduces interpretability. Therefore, the CHAID model was selected due to its interpretability and the use of a Bonferonni adjustment for the p-value as a splitting criteria (Pawar & Gawande, 2012).

4. RESULTS

The binary logistic regression, using a stepwise method of variable selection, explained 21% of the variance in recommendations (Nagelkerke R^2) and correctly classified 66.86% of reviews in the training sample and 63.2% in the test sample.

The results (Table 2) show that positive feelings have the largest positive effect on recommendations ($\text{Exp}(B)=2.949$, $p<.05$), and negative competences have the largest opposite effect ($\text{Exp}(B)=-.353$, $p<.05$).

-- PLACE TABLE 2 AROUND HERE --

The CHAID decision tree (Figure 1) correctly predicted reviews in 66.05% and 61.28% of cases in the training and testing data, respectively. The decision tree found that feelings and attitudes were the most important predictors of explicit recommendations in a review. When reviewers write comments that indicate a negative attitude (e.g. arrogant, bad-tempered), 79.3% of them write a review that contains an explicit negative recommendation. A review that contains no negative feelings (e.g. bitter, dirty) or negative attitudes leads to a positive recommendation 74.9% of the time. Although the total sample has more positive than negative explicit recommendations, which may be

explained by a higher number of positive than negative experiences overall (Xiang, *et al.*, 2017), the current study shows that negative experiences have a stronger impact on explicit recommendations.

-- PLACE FIGURE 1 AROUND HERE --

5. DISCUSSION AND CONCLUSIONS

This research contributes to the literature by showing the most important predictors of an explicit recommendation in the review versus a non-recommendation advice.

A lack of competence and negative attitudes are important factors that predict the presence of a negative recommendation within a review. For example, if reviewers feel that the staff had a bad attitude, they will usually recommend others to avoid the service. The same applies when reviewers felt that the provider was not able to efficiently resolve any problems that occurred. These findings are aligned with the study by Stringam & Gerdes (2010), who found that words suggesting that a provider lacked competence (e.g., “apology”, “refund”) were negatively associated with consumer ratings. The current research note shows that such an effect may also lead to a negative explicit recommendation in the review.

Cleanliness and the quality of decor have been positively associated with the popularity of restaurants (Zhang *et al.*, 2010). We are aligned with such findings, demonstrating that positive feelings usually predict a positive recommendation in the body of the review, particularly when the review contains no negative feelings. If consumers are completely satisfied with a service, not only in functional terms, but especially if the service triggers positive feelings (e.g., attractive decoration, clean environment), they usually recommend it to their peers.

Despite the contributions of our study and its large sample size, we were limited by our use of a single information source (Yelp) and a single type of business (restaurants). Future research may confirm our results using other information sources and focusing on other services in hospitality and tourism. Future research may also investigate whether specific experiences (e.g., wait time) influence specific sentiment categories (e.g. competence), which would help marketers manage the antecedents of such predictive behaviour.

References

- Aggarwal, C. C. (2012). Mining Text Data. In C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 1–10). New York, NY: Springer.
- Briggs, S., Sutherland, J., & Drummond, S. (2007). Are hotels serving quality? An exploratory study of service quality in the Scottish hotel sector. *Tourism Management*, 28(4), 1006–1019. <https://doi.org/10.1016/j.tourman.2006.08.015>.
- Bronner, F. & de Hoog, R. (2011). Vacationers and eWOM: Who posts, and why, where, and what? *Journal of Travel Research*, 50(1), 15-26. <https://doi.org/10.1177/0047287509355324>.
- Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 1-19. <https://doi.org/10.1080/19368623.2017.1310075>.
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Special issue on learning from imbalanced data sets. *ACM Sigkdd Explorations Newsletter*, 6(1), 1-6. <https://doi.org/10.1145/1007730.1007733>.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online

- book reviews. *Journal of Marketing Research*, 43(3), 345-354.
<https://doi.org/10.1509/jmkr.43.3.345>.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, 2493-2537.
- Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travellers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174-185.
<https://doi.org/10.1016/j.tourman.2015.05.007>.
- Guerreiro, J., & Moro, S. (2017). Are Yelp's tips helpful in building influential consumers? *Tourism Management Perspectives*. 24, 151-154.
<https://doi.org/10.1016/j.tmp.2017.08.006>.
- He, H., and Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, 21(9), 1263-1284.
<https://doi.org/10.1109/TKDE.2008.239>.
- IBM. (2019a). <https://www.ibm.com/products/spss-modeler> Accessed 01 July 2019.
- IBM. (2019b).
https://www.ibm.com/support/knowledgecenter/bg/SS3RA7_15.0.0/com.ibm.spss.ta.help/intro_extract_defined.htm Accessed 01 July 2019.
- Litvin, S. W., Bloise, J. E., & Laird, S. T. (2005). Tourists' use of restaurant webpages: Is the internet a critical marketing tool? *Journal of Vacation Marketing*, 11(2), 155–161. <https://doi.org/10.1177/1356766705052572>.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and

- applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
<https://doi.org/10.1016/j.asej.2014.04.011>.
- Ordenes, F. V., Ludwig, S., Ruyter, K., & Wetzels, M. (2017). Unveiling what is written in the stars: Analysing explicit, implicit and discourse patterns of sentiment in social media. *Journal of Consumer Research*, 43, 875-894.
<https://doi.org/10.1093/jcr/ucw070>.
- Packard, G & Berger, J. (2017). How language shapes word of mouths' impact. *Journal of Marketing Research*. 54, 572-588. <https://doi.org/10.1509/jmr.15.0248>.
- Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, 51(4), 483–491.
<https://doi.org/10.1177/1938965510378574>.
- Pawar, P. Y., & Gawande, S. H. (2012). A comparative study on different types of approaches to text categorization. *International Journal of Machine Learning and Computing*, 2(4), 423-426. <https://doi.org/10.7763/IJMLC.2012.V2.158>.
- Sarkar, B. K. (2016). A case study on partitioning data for classification. *International Journal of Information and Decision Sciences*, 8(1), 73-91.
<https://doi.org/10.1504/IJIDS.2016.075788>.
- Stringam, B. B., & Gerdes, J. (2010). An analysis of word-of-mouth ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management*, 19(7), 773–796. <https://doi.org/10.1080/19368623.2010.508009>.
- Wattanacharoensil, W., Schuckert, M., Graham, A., & Dean, A. (2017). An analysis of the airport experience from an air traveler perspective. *Journal of Hospitality and Tourism Management*, 32, 124-135. <https://doi.org/10.1016/j.jhtm.2017.06.003>.

- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37. <https://doi.org/10.1007/s10115-007-0114-2>.
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130. <https://doi.org/10.1016/j.ijhm.2014.10.013>.
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65. <https://doi.org/10.1016/j.tourman.2016.10.001>.
- Yelp. Yelp Challenge Dataset 2017. (2017). https://www.yelp.com/dataset_challenge Accessed 01 September 2017.
- Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694-700. <https://doi.org/10.1016/j.ijhm.2010.02.002>.

	Probit	Binary Logistic	CHAID	C&RT	Random Forest
Accuracy	66.86%	66.86%	66.05%	66.05%	66.83%

Table 1. Predictive Models Accuracy

	B	Wald	Sig.	Exp(B)
Intercept	-.052	10.733	.001	
Customer Support	-.221	113.482	.000	.802
Negative Attitude	-.893	804.229	.000	.409
Negative Budget	-.661	488.685	.000	.516
Negative Competence	-1.042	93.615	.000	.353
Negative Feeling	-.847	1169.048	.000	.429
Negative Functioning	-.431	311.019	.000	.650
Positive Attitude	.808	1079.210	.000	2.243
Positive Budget	.105	9.325	.002	1.111
Positive Competence	.228	41.602	.000	1.256
Positive Feeling	1.082	2940.617	.000	2.949
Positive Functioning	.450	48.335	.000	1.569
WaitTime	-.241	83.000	.000	.786

Table 2. Binary Logistic Regression for predicting explicit recommendations

Appendix 1

Valence	Topic	Examples of Included Terms
Positive	Positive	good; like; love
	Positive Attitude	adorable; amicable; compassionate
	Positive Feeling	attractive; beautiful; calm
	Positive Competence	accurately answered; answered all my questions; competent

	Positive Functioning	all went well; faultless; keeps improving
	Positive Budget	affordable; cheap; cost effective
	Positive Recommendation	come back; confidently recommend; definitely book again
	Negative	bad; annoying; busy
	Negative Attitude	arrogant; autocratic; careless
	Negative Feeling	bitter; bored; chalky
Negative	Negative Competence	badly prepared; contradictory; didn't explain
	Negative Functioning	barely worked; can't get to work; collapsed
	Negative Budget	a bit pricy; a little high; cost a lot of money
	Negative Recommendation	not recommend; consider leaving; do not plan to return
	Wait Time	wait in line; queue; waited
	Customer Support	customer service; call center; staff

Table 3. Sentiment categories

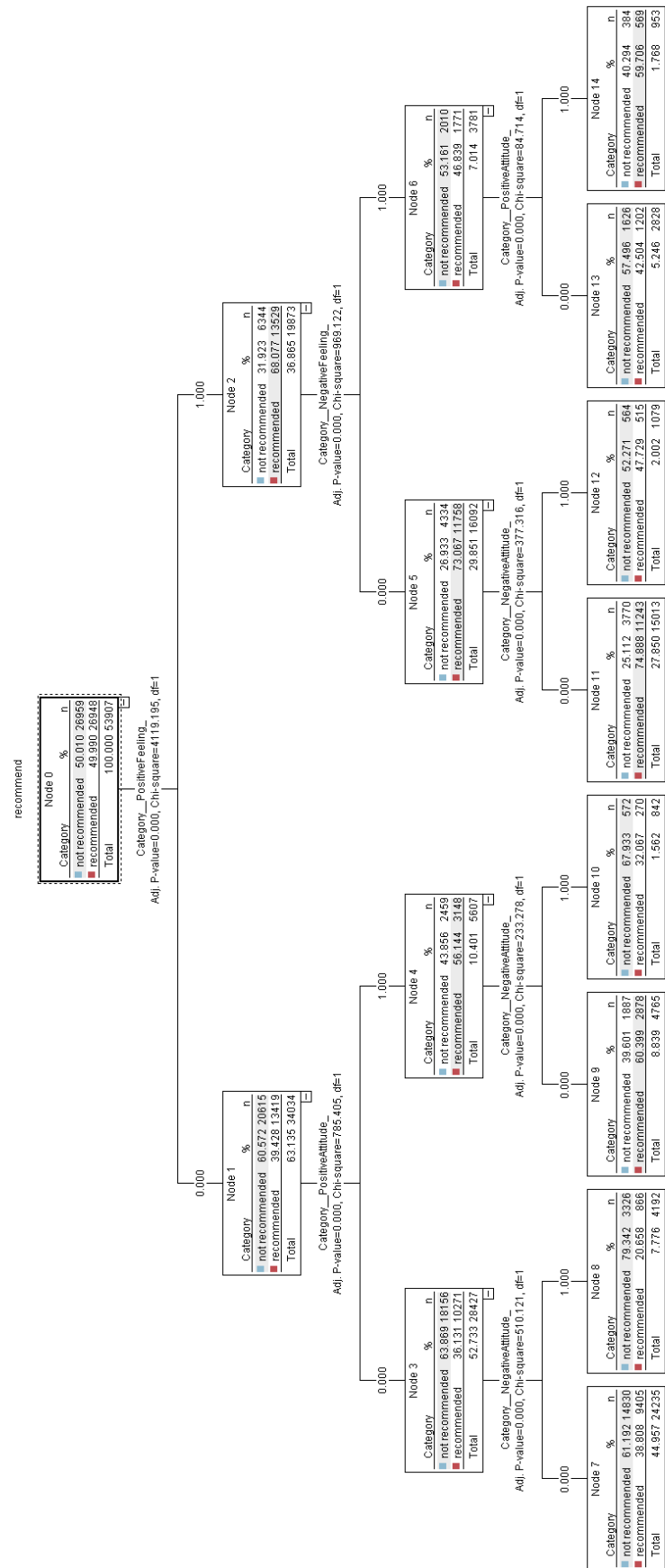


Figure 1. Decision Tree for predicting explicit recommendations