AIRLINE REVIEWS SENTIMENT ANALYSIS

**Text Mining**

Winter Semester 20/21

**–** Term Paper **–**

Submitted to

Prof. Dr.

Chair for Data Science and Digitization

at

Faculty of Economics and Business Studies

Submission date:

Submitted by:

Student ID:

E-Mail:

# **Abstract**

This text mining project aims to perform airline customers’ reviews sentiment analysis and text classification using dictionary generation and machine learning classification algorithms using R as a statistical analysis tool. Airport reviews dataset containing 2 (two) variables of interest (Text and Rating) with 2027 reviews (entries) from airline customers was used for the text mining analysis to uncover the reviews’ social and market norms, as well as the word polarity to identify the sentiments. This research uses 6 (six) sentiments dictionaries (Bing-Dictionary, NRC-Dictionary, GI-Dictionary, HE-Dictionary, LM-Dictionary and QDAP-Dictionary) for sentiment analysis and lounge services were found to be of more interest. It also uses both machine learning algorithms (Decision Tree, KNN and Naïve Bayes) and dictionary generating methods (LESSO Regularization, tf weighting, Pike and Slab regression and OLS Regression) for text classification where the dictionary generation method was found to be more accurate (approximately 100% accurate) than machine learning algorithms.

# **TABLE OF CONTENT**

Contents

[**Abstract** ii](#_Toc96968710)

[**TABLE OF CONTENT** iii](#_Toc96968711)

[INTRODUCTION 4](#_Toc96968712)

[**2. Dictionary-Based Sentiment Analysis** 5](#_Toc96968713)

[**2.1 Introduction** 5](#_Toc96968714)

[2.2 Sentiment Lexicons Dictionaries 5](#_Toc96968715)

[2.2.1 BING Dictionary 5](#_Toc96968716)

[2.2.2 NRC-Dictionary 6](#_Toc96968717)

[**2.2.3 GI-Dictionary** 7](#_Toc96968718)

[**2.2.4 HE-Dictionary** 8](#_Toc96968719)

[**2.2.5 LM-Dictionary** 9](#_Toc96968720)

[**2.2.6 QDAP-Dictionary** 10](#_Toc96968721)

[**2.3 Word correlation** 11](#_Toc96968722)

[**3. Dictionary Generation** 14](#_Toc96968723)

[**3.1 Introduction** 14](#_Toc96968724)

[**3.2 LASSO Regularization** 14](#_Toc96968725)

[**3.3 pike-and-slab regression** 14](#_Toc96968726)

[**3.4 TF Weighting** 15](#_Toc96968727)

[**3.5 OLS Regression** 15](#_Toc96968728)

[**4. Machine Learning Algorithms** 16](#_Toc96968729)

[**4.1 Introduction** 16](#_Toc96968730)

[**4.2 Naïve Bayes** 17](#_Toc96968731)

[**4.3 Decision Tree Algorithm** 17](#_Toc96968732)

[**4.4 K – Nearest Neighbors (KNN)** 18](#_Toc96968733)

[**5. COMPARISON** 18](#_Toc96968734)

[**6. CONCLUSION** 19](#_Toc96968735)

[**REFERENCES** 20](#_Toc96968736)

INTRODUCTION

The global airline industry is facing increased competition between airlines and regions due to the expansion of the Treaty on Open Skies, private participation in airline and airport operations, and inter aircraft partnerships and mergers. To survive this competition, airlines continue to make efforts to improve service quality as survival strategies. Just as the product quality revolution in manufacturing determines a company's competitiveness, in the service sector, service quality innovation is a factor that determines a company's win or loss. In addition, the development of service quality is perceived as a means of securing a competitive advantage with loyal customers. However, the airline industry is not aware of the customer's needs, and the provision of quality of service is compromised. As a result, customer needs have consistently attracted the attention of scholars as a fundamental variable in customer service delivery and should be more important to airlines to identify customer needs and provide the right quality of service. To be able to know where improvement is needed, sentiment analysis is required to be performed.

Skytrax (originally known as Inflight Research Services) is a [United Kingdom](https://en.wikipedia.org/wiki/United_Kingdom)-based consultancy that runs an [airline](https://en.wikipedia.org/wiki/Airline) and [airport](https://en.wikipedia.org/wiki/Airport) review and ranking site. Skytrax conducts research for commercial airlines, as well as taking [surveys](https://en.wikipedia.org/wiki/Statistical_survey) from international travelers to rate cabin staff, airports, airlines, airline lounges, [in-flight](https://en.wikipedia.org/wiki/In-flight_entertainment) entertainment, on-board catering, and several other elements of air travel. Apart from these evaluations, Skytrax has an airline forum where passengers give potential passengers insights and opinions about an airline. It also hosts flight reviews, flight checks, and satisfaction surveys.

The aim of this project is to perform sentiment analysis on airline customer reviews. The dataset used was obtained from Skytrax with four variables; Author, Text, Rating, and normalized rating

# **2. Dictionary-Based Sentiment Analysis**

## **2.1 Introduction**

Sentiment analysis is a research branch located at the heart of natural language processing (NLP), computational linguistics and text mining. Dictionary-based sentiment analysis is a computational approach to measuring the feeling that a text conveys to the reader. In the simplest case, sentiment has a binary classification: positive or negative; but it can be extended to multiple dimensions such as good, bad, anger, joy, etc. This method relies heavily on a predefined list (or dictionary) of sentiment-laden words. There are a variety of methods and dictionaries that exist for evaluating the opinion or emotion in text. The tidytext package provides access to several sentiment lexicons.

2.2 Sentiment Lexicons Dictionaries

2.2.1 BING Dictionary

This lexicon is based on unigrams, i.e., single words. It contains many English words and the words are assigned scores for positive/negative sentiment, and also possibly emotions like joy, anger, sadness, and so forth. The Bing lexicon categorizes words in a binary fashion into positive and negative categories.

From sentiment analysis performed on the airport customers’ reviews data, the top 10 most positive words are good, available, nice, comfortable, well, free, clean, hot, great and better. The top 10 most negative words are poor, limited, crowded, cold, dirty, worst, bad, disappointing, slow, tired and difficult. The word cloud below shows the word contribution to the sentiment where a word with large font implies a higher contribution and with which words with green font color are positive while those with red font color are negative.



Figure 1: Bing word cloud.

2.2.2 NRC-Dictionary

NRC lexicon dictionary is also based on single word (unigram). Like Bing lexicon, NRC lexicon contain many English words and the words are assigned scores for positive/negative sentiment, and also possibly emotions like joy, anger, sadness, and so forth. The NRC lexicon categorizes words in a binary fashion (“yes”/“no”) into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The word cloud below shows word contribution to sentiment where the larger the font size the greater contribution to sentiment.

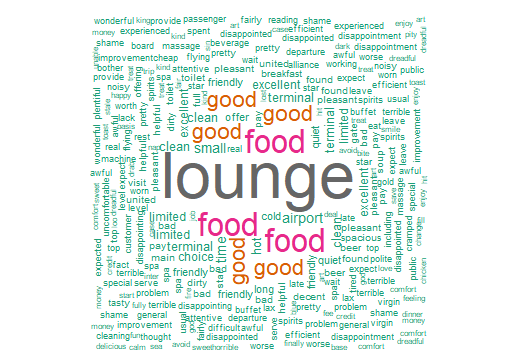


Figure 2: NRC word cloud.

### **2.2.3 GI-Dictionary**

**General Inquirer (GI)** is a Harvard-IV dictionary. It is a dictionary with a list of positive and negative words according to the psychological Harvard-IV dictionary as used in the General Inquirer software. This is a general-purpose dictionary developed by the Harvard University. It is one of many dictionaries available in the SentimentAnalysis package. The word cloud below shows the word contribution to the sentiment where a word with large font implies a higher contribution and with which words with green font color are positive while those with red font color are negative.

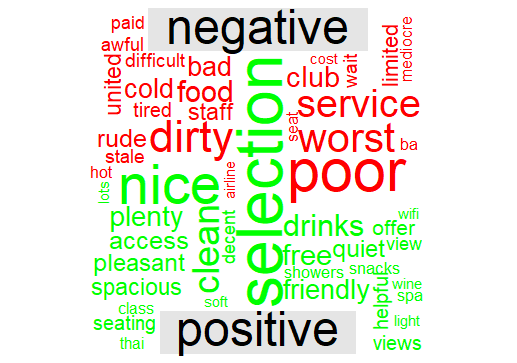


Figure 3: GI dictionary word cloud.

### **2.2.4 HE-Dictionary**

HE is a dictionary with a list of positive and negative words according to the Henry’s finance-specific dictionary. This dictionary was first presented in the Journal of Business Communication among one of the early adopters of text analysis in the finance discipline.

The word cloud below shows word contribution to sentiment where the larger the font size the greater contribution to sentiment.



Figure 4: HE Dictionary word cloud.

From the word cloud above, it is seen that “lounge” is often used in commenting by the airport customers. Lounges are places where passengers can enter under special cases. Lounges are mostly seen within the accommodation, entertainment and aviation sectors and they are areas much more comfortable than the waiting areas. The word “lounge” is often used in commenting since most customers have experience lounge services.

### **2.2.5 LM-Dictionary**

LM-Dictionary is a dictionary with a list of positive, negative and uncertain words according to the Loughran-McDonald finance-specific dictionary. This dictionary was first presented in the Journal of Finance and has been widely used in the finance domain ever since.

The word cloud below shows word contribution to sentiment where the larger the font size the greater contribution to sentiment.

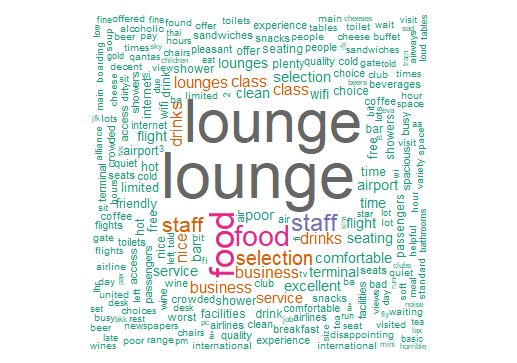


Figure 5: LM Dictionary word cloud.

It is also evident that the word “lounge” is often used in airport customers’ comments. The reason is just the same as what we have discussed on the previous part (section 2.2.4).

### **2.2.6 QDAP-Dictionary**

The word cloud below shows the word contribution to the sentiment where a word with large font implies a higher contribution and with which words with green font color are positive while those with red font color are negative.



Figure 6: QDAP Dictionary word cloud.

## **2.3 Word correlation**

In word analysis, we look at the association of words using the consecutive words. Another way is to investigate how often any two words appear in a given comment. For example, if a comment mentioned that beer is plenty, we might be interested in knowing what other words were mentioned in the same context. The function pairwise\_count in the R package [widyr](https://cran.r-project.org/package=widyr) was used to count how many times a pair of words appear together in a comment. Word association analysis was performed and the following word network graph showing associated words was obtained.

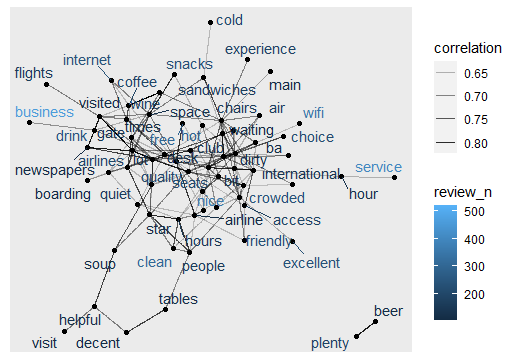


Figure 7: Word network graph

Here we can see clusters of word networks most commonly used together.

Correlation coefficient of greater than 0.6 shows strong correlation. In the graph above, the bold the connecting line is, the stronger the correlation.

The mathematical concept used in calculating each pair correlation coefficient (phi coefficient) is given by the following formula;

Where:

n11 is the total number of comments with both word x and word y in it.

n10 is the total number of comments with word x but without word y.

n01 is the total number of comments with without word x but with word y.

n00 is the total number of comments with neither word x nor word y.

n1. is the total number of comments with word x.

n0. is the total number of comments without word x.

* n.1is the total number of comments with word y.
* n.0 is the total number of comments without word y.
* n..n.. is the total number of comments.

# **3. Dictionary Generation**

## **3.1 Introduction**

In SentimentAnalysis package, “generateDictionary” function is used to get the list of positive and negative words associated with rating.

Routine applies method for dictionary generation (LASSO, ridge regularization, elastic net, ordinary least squares, generalized linear model or spike-and-slab regression) to the document-term matrix in order to extract decisive terms that have a statistically significant impact on the response variable.

## **3.2 LASSO Regularization**

LASSO regularization process generated weighted (words with individuals scores) dictionary type with 38 total entries where 19 (50%) are positive entries and 19 (50%) are negative entries. The average word score was 0.003286664 with a Standard deviation of 0.05843928 and Skewness of 0.6384296.

This model generated the following predictive results; word correlation = 0.5316, correlation t – statistic = 28.2412, lm t. value = 28.2412, r-squared = 0.2826, RSME = 1.0999 and accuracy score = 1.

The in-sample performance lasso generated the following results:

Cor = 0.539259, cor.t.statistic = 28.8154740, lm.t.value = 28.815474,

r.squared = 0.29080, RMSE = 1.09088885, Accuracy = 1 and

avg.sentiment.pos.response = 3.30439.

## **3.3 pike-and-slab regression**

The Pike and Slab regression generated weighted (words with individuals scores) dictionary type with 50 total entries where 24 (48%) are positive entries and 26(52%) are negative entries. The average word score was 0.0008706261 with a Standard deviation of 0.06074601 and Skewness of 0.2913962.

This model generated the following predictive results; word correlation = 0.5521, correlation t – statistic = 29.7978, lm t. value = 29.7978, r-squared = 0.3048, RSME = 3.3888, accuracy score = 1 and avg.sentiment.pos.response = 0.09089.

## **3.4 TF Weighting**

The tf weighting generated weighted (words with individuals scores) dictionary type with 34 total entries where 16 (47.06%) are positive entries and 18 (52.94%) are negative entries. The average word score was 0.007646263 with a Standard deviation of 0.1787446 and Skewness of 0.6849168.

This model generated the following predictive results; word correlation = 0.5331, correlation t – statistic = 28.3523, lm t. value = 28.3523, r-squared = 0.2842, RSME = 1.4510, accuracy score = 0.983719 and avg.sentiment.pos.response = 3.390367.

## **3.5 OLS Regression**

The OLS Regressiongenerated weighted (words with individuals scores) dictionary type with 65 total entries where 32 (49.23%) are positive entries and 33 (50.77%) are negative entries. The average word score was -0.0002953292 with a Standard deviation of 0.063848 and Skewness of -0.3195831 (left skewed).

This model generated the following predictive results; word correlation = 0.5562, correlation t – statistic = 30.115, lm t. value = 30.1150, r-squared = 0.3093,

RSME = 1.0708, accuracy score = 1 and avg.sentiment.pos.response = 3.30439.

# **4. Machine Learning Algorithms**

## **4.1 Introduction**

You can perform text classification in two ways: manual or automatic.

Manual text classification involves a human annotator, who interprets the content of text and categorizes it accordingly. This method can deliver good results but it’s time-consuming and expensive. Automatic text classification applies machine learning, [natural language processing (NLP), and other AI-guided techniques](https://monkeylearn.com/blog/nlp-ai/) to automatically classify text in a faster, more cost-effective, and more accurate manner. There are many approaches to automatic text classification, but they all fall under three types of systems: Rule-based systems, Machine learning-based systems and Hybrid systems.

In this paper, we’re going to focus on a Machine learning-based systems text classification.

Machine learning text classification learns to make classifications based on past observations. By using pre-labeled examples as training data, machine learning algorithms can learn the different associations between pieces of text, and that a particular output (i.e., tags) is expected for a particular input (i.e., text). The first step towards training a machine learning NLP classifier is feature extraction: a method is used to [transform each text into a numerical representation](https://monkeylearn.com/blog/beginners-guide-text-vectorization/) in the form of a vector. One of the most frequently used approaches is [bag of words](https://machinelearningmastery.com/gentle-introduction-bag-words-model/), where a vector represents the frequency of a word in a predefined dictionary of words.

Then, the machine learning algorithm is fed with training data that consists of pairs of feature sets (vectors for each text example) and tags (e.g. sports, politics) to produce a classification model. Once it’s trained with enough training samples, the machine learning model can begin to make accurate predictions. The same feature extractor is used to transform unseen text to feature sets, which can be fed into the classification model to get predictions on tags.

In this work we will look at three text classification algorithms: Naïve Bayes, K-NN and Decision tree algorithms.

## **4.2 Naïve Bayes**

Naive Bayes is based on Bayes’s Theorem, which helps us compute the conditional probabilities of the occurrence of two events, based on the probabilities of the occurrence of each individual event. So we’re calculating the probability of each word to be positive or negative.

The probability of A, if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true.

The algorithm was implemented in R using the airport reviews dataset and an accuracy score of 0.3185 was obtained with a 95% confidence interval of (0.2835, 0.3551).

## **4.3 Decision Tree Algorithm**

The decision tree Algorithm belongs to the family of supervised machine learning algorithms. It can be used for both a classification problem as well as for regression problem. The goal of this algorithm is to create a model that classify text according to its polarity trained using rating values, for which the decision tree uses the tree representation to solve the problem in which the leaf node corresponds to a class label and attributes are represented on the internal node of the tree.

The algorithm was trained with the airport review data set and obtained its prediction accuracy of 0.3704 with 95% confidence interval of (0.3338, 0.408).

## **4.4 K – Nearest Neighbors (KNN)**

KNN algorithm is used to classify by finding the K nearest matches in training data and then using the label of closest matches to predict. Generally, neighbors share similar characteristics and behavior that's why they can be treated as they belong to the same group. We consider the K-Nearest Neighbors of the unknown data we want to classify and assign it the group appearing majorly in those K neighbors. For K=1, the unknown/unlabeled data will be assigned the class of its closest neighbor.

The algorithm was implemented using R to classify words used in airport reviews with polarity and an accuracy for predicting new dataset was obtained to be 0.3348 with 95% confidence interval of (0.2943, 0.3718)

From the three text classification algorithms used in this study, the algorithm with the highest accuracy score was Decision Tree algorithm.

# **5. COMPARISON**

There are two predictive approaches used in this paper namely; dictionary generation method and machine learning approach which generated different predictive results. Four dictionary generation methods are used. LASSO regularization with a predictive accuracy score of 1, pike and slab regression with an accuracy score of 0.5328, tf weighting with an accuracy score of 0.9837 and ordinary least squares (OLS) regression with an accuracy score of 1. Three machine learning algorithms are used; Decision Tree, Naïve Bayes and KNN with predictive accuracy scores 0.3704, 0.3185 and 0.3348 respectively.

Generally, dictionary generation methods have high predictive accuracy scores than machine learning algorithms.

# **6. CONCLUSION**

Based on the result of this study, lounge services are of more concern in the airline industry since “lounge” is frequently used by airline customers in writing comments and reviews on airline products and services. Internet technology and sentiment analysis techniques are of importance in identifying which products or services need to be improved in the airline industry in order to ensure quality services and the best customer experience. This is necessary to be able to survive in the competitive environment of airline markets. The best approach to build sentiment predictive models is implementing the dictionary generation methods since they produce more accurate results than machine learning algorithms. Among the dictionary generation methods, the best techniques are LASSO Regression and OLS Regression.

# **REFERENCES**

1. Heriyanto, D.S.N.; Putro, Y.M. Challenges and Opportunities of the Establishment ASEAN Open Skies Policy. *Padjadjaran J.*

*Ilmu Huk. (J. Law)* 2019, *6*, 466–488, doi:10.22304/pjih.v6n3.a3.

1. Chen, L.; Li, Y.Q.; Liu, C.H. How airline service quality determines the quantity of repurchase intention-Mediate and moderate effects of brand quality and perceived value. *J. Air Trans. Manag.* 2019, *75*, 185–197.
2. Gupta, H. Evaluating service quality of airline industry using hybrid best worst method and VIKOR. *J. Air Transp. Manag.* 2018,*68*, 35–47, doi:10.1016/j.jairtraman.2017.06.001.
3. Han, H.; Yu, J.; Kim, W. Environmental corporate social responsibility and the strategy to boost the airline’s image and customer loyalty intentions. *J. Travel Tour. Mark.* 2019, *36*, 371–383 doi:10.1080/10548408.2018.1557580.
4. Balakrishnan, B.K.; Dahnil, M.I.; Yi, W.J. The Impact of Social Media Marketing Medium toward Purchase Intention and Brand

Loyalty among Generation Y. *Procedia-Soc. Behav. Sci.* 2014, *148*, 177–185, doi:10.1016/j.sbspro.2014.07.032.

1. Alnsour, M.; Ghannam, M.; Alzeidat, Y. Social media effect on purchase intention: Jordanian airline industry. *J. Internet Bank.*

*Commer.* 2018, *23*, 1–1.

1. Kim, S.; Kandampully, J.; Bilgihan, A. The influence of eWOM communications: An application of online social network framework.

*Comput. Human Behav*. 2018, *80*, 243–254.

1. Lyberg, L.E.; Weisberg, H.F.; Wolf, C.; Joye, D.; Smith, T.; Fu, Y.-C. Total Survey Error: A Paradigm for Survey Methodology.
2. Basias, N.; Pollalis, Y. Quantitative and qualitative research in business & technology: Justifying a suitable research methodology.

*Rev. Integr. Busi. Econo. Res.* 2018, *7*, 91–105.

1. Ban, H.-J.; Kim, H.-S. Understanding Customer Experience and Satisfaction through Airline Passengers’ Online Review. *Sustainability*

2019, *11*, 4066, doi:10.3390/su11154066.

1. Liu, B. Sentiment Analysis and Opinion Mining. *Synth. Lect. Hum. Lang. Technol.* 2012, *5*, 1–167,

doi:10.2200/s00416ed1v01y201204hlt016.

1. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* 2017, *58*, 175–191,

doi:10.1177/0047287517747753.

1. Ali, F.; Kwak, D.; Khan, P.; El-Sappagh, S.; Ali, A.; Ullah, S.; Kwak, K.S. Transportation sentiment analysis using word embedding

and ontology-based topic modeling. *Knowl. Based Syst.* 2019, *174*, 27–42.

1. Park, E.; Kang, J.; Choi, D.; Han, J. Understanding customers’ hotel revisiting behaviour: A sentiment analysis of online feedback

reviews. *Curr. Issues Tour.* 2018, *23*, 605–611, doi:10.1080/13683500.2018.1549025.

1. Tran, T.; Ba, H.; Huynh, V.-N. Measuring Hotel Review Sentiment: An Aspect-Based Sentiment Analysis Approach. In *Proceedings*

*of the Computer Vision*; Springer International Publishing: Cham, Switzerland, 2019; pp. 393–405.

1. Nakayama, M.; Wan, Y. The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Infor.*

*Manag.* 2019, *56*, 271–279.

1. Knorr, A. Big data, customer relationship and revenue management in the airline industry: What future role for frequent flyer

programs? *Rev. Integr. Bus. Econo. Res*. 2019, 8, 38–51.

**APPENDIX: R code**