

## Chapter 59

# Discovery of Sustainable Transport Modes Underlying TripAdvisor Reviews With Sentiment Analysis: Transport Domain Adaptation of Sentiment Labelled Data Set

**Ainhoa Serna**

 <https://orcid.org/0000-0003-4750-3222>

*University of the Basque Country, Spain*

**Jon Kepa Gerrikagoitia**

*BRTA Basque Research and Technology Alliance, Spain*

### **ABSTRACT**

*In recent years, digital technology and research methods have developed natural language processing for better understanding consumers and what they share in social media. There are hardly any studies in transportation analysis with TripAdvisor, and moreover, there is not a complete analysis from the point of view of sentiment analysis. The aim of study is to investigate and discover the presence of sustainable transport modes underlying in non-categorized TripAdvisor texts, such as walking mobility in order to impact positively in public services and businesses. The methodology follows a quantitative and qualitative approach based on knowledge discovery techniques. Thus, data gathering, normalization, classification, polarity analysis, and labelling tasks have been carried out to obtain sentiment labelled training data set in the transport domain as a valuable contribution for predictive analytics. This research has allowed the authors to discover sustainable transport modes underlying the texts, focused on walking mobility but extensible to other means of transport and social media sources.*

DOI: 10.4018/978-1-6684-6303-1.ch059

## INTRODUCTION

Mobility with motor vehicles has a negative environmental impact. Over time, means of transportation have emerged as an alternative to the use of motor vehicles. Examples of sustainable transport are bicycle, public transport (subway, tram, bus), electric car, and so on. Conforming to the EU Transport Council in 2001, “*a sustainable transport system is one that allows individuals and societies to meet their needs for access to areas of activity with total safety, in a manner consistent with human and ecosystem health, and that is also balanced equally between different generations*” (European Commission - Mobility and Transport, 2018).

Additionally, tourist activity generates wealth in the receiving place and is an excellent great source of employment. However, as a counterpart, it can also be a destructive activity. It is estimated that tourism activity produces up to 8% of global greenhouse gas emissions from 2009 to 2013 (Lenzen, Sun, Faturay et al., 2018). Even if we take into account the energy used in hotels, transport or hygiene products, it represents up to 12.5% (Sánchez, 2018). Moreover, cities across Europe have adopted or strengthened Low Emission Zones (LEZ) in response to the growing air pollution crisis. These measures have been taken by more than 250 EU cities. A study shows that 67% of interviewees favour the adoption of LEZ either strongly or slightly. LEZ should move forward to zero-emission mobility zones (ZEZ), that will eventually be turned into policies to promote transitioning to healthier alternatives like walking, cycling jointly with the electrification of all forms of transport like taxis, public transport and private vehicles (Müller and Le Petit, 2019).

Furthermore, tourism produces large quantities of the content generated by users (User Generated Content) that is rapidly growing. There is a wide variety of subjects in this type of content, and one of them is mobility. On the other hand, the different languages and contexts are relevant to react when consumers around the world are speaking various languages and as digital platforms increase the range of users on these platforms, such as Social Media data of TripAdvisor platform. Being platforms worldwide that include users from different countries, the variety and richness of the data that can be extracted and the knowledge that can be created with them can be very relevant for different companies, both public and private.

In recent years, in particular, digital technology and research methods have developed the concept of Natural Language Processing that has become a preferred means for better understanding consumers and what they share in. Regarding the economic and business relevance of NLP, forecasts that Global NLP Market is projected to rise to \$26.4 billion by 2024 and the CAGR (Compound Annual Growth Rate) of 21% from 2019 (MarketsandMarkets, 2019) will continue to increase. Given the current importance of this area and future forecasts, this research will focus on the application of NLP in the field of transport, since the contribution of this research can be relevant both at the level of global and local business. For this reason, this investigation analyses the different transport modes, focus on sustainable transports. In this research, natural language processing techniques are applied to Social Media data (UGC), to evaluate the impressions of visitors regarding success factors that can be used as planning aid tools. The study has been developed according to transport mode used and languages.

Regarding the novelty of this research, it should be noted that there are numerous TripAdvisor articles but mainly focused on tourism, such as monuments, hotels, restaurants, attractions...etc. There are hardly any studies in transportation analysis with TripAdvisor, and moreover, there is no a complete analysis of sentiment analysis. This article proposes TripAdvisor as a data source for the study of modes of transport, user ratings and automated sentiment-detection algorithms.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/discovery-of-sustainable-transport-modes-underlying-tripadvisor-reviews-with-sentiment-analysis/308537](http://www.igi-global.com/chapter/discovery-of-sustainable-transport-modes-underlying-tripadvisor-reviews-with-sentiment-analysis/308537)

## Related Content

---

### Semantics-Aware Advanced OLAP Visualization of Multidimensional Data Cubes

Alfredo Cuzzocrea, Domenico Sacca and Paolo Serafino (2007). *International Journal of Data Warehousing and Mining* (pp. 1-30).

[www.irma-international.org/article/semantics-aware-advanced-olap-visualization/1791](http://www.irma-international.org/article/semantics-aware-advanced-olap-visualization/1791)

### Sentiment Analysis and Sarcasm Detection (Using Emoticons)

Vibhu Dagar, Amber Verma and Govardhan K. (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1600-1610).

[www.irma-international.org/chapter/sentiment-analysis-and-sarcasm-detection-using-emoticons/308565](http://www.irma-international.org/chapter/sentiment-analysis-and-sarcasm-detection-using-emoticons/308565)

### Extended Adaptive Join Operator with Bind-Bloom Join for Federated SPARQL Queries

Damla Oguz, Shaoyi Yin, Belgin Ergenç, Abdelkader Hameurlain and Oguz Dikenelli (2017). *International Journal of Data Warehousing and Mining* (pp. 47-72).

[www.irma-international.org/article/extended-adaptive-join-operator-with-bind-bloom-join-for-federated-sparql-queries/185658](http://www.irma-international.org/article/extended-adaptive-join-operator-with-bind-bloom-join-for-federated-sparql-queries/185658)

### Statistical Sampling to Instantiate Materialized View Selection Problems in Data Warehouses

Mesbah U. Ahmed, Vikas Agrawal, Udayan Nandkeolyar and P. S. Sundararaghavan (2007). *International Journal of Data Warehousing and Mining* (pp. 1-28).

[www.irma-international.org/article/statistical-sampling-instantiate-materialized-view/1776](http://www.irma-international.org/article/statistical-sampling-instantiate-materialized-view/1776)

### Frequent Pattern Discovery and Association Rule Mining of XML Data

Qin Ding and Gnanasekaran Sundarraj (2012). *XML Data Mining: Models, Methods, and Applications* (pp. 243-263).

[www.irma-international.org/chapter/frequent-pattern-discovery-association-rule/60912](http://www.irma-international.org/chapter/frequent-pattern-discovery-association-rule/60912)