RRL Matrix

**COLINA**, Rita Isabel P.

Topic: **Examining Tourist Sentiments in the Philippines Through NLP Before and After COVID-19**

Adviser: **Dr. Maria Art Antonette D. Clariño**

| **Literature** | **Major findings** | **Theories and Methodologies Used** | **Relation to the study** | **Research gap** |
| --- | --- | --- | --- | --- |
| Naciye Güliz Uğur, Adem Akbıyık,  Impacts of COVID-19 on global tourism industry: A cross-regional comparison,  Tourism Management Perspectives,  Volume 36,  2020,  100744,  ISSN 2211-9736,  https://doi.org/10.1016/j.tmp.2020.100744.  (https://www.sciencedirect.com/science/article/pii/S2211973620301112) | The tourism sector is very sensitive and easily affected by global crises  Cancellation of trips were mention in 40.81% of all cases  It takes extra time for travelers to return to their old mobility even after the crisis has ended  Travel insurance scored as the most frequently repeated phrase after coronavirus  Magnitude of C19 consequences is not comparable to previous crises | Text mining in TripAdvisor, documents, web pages, social media, consumer comments  Word cloud  Word/phrase frequency count | This study applies sentiment analysis using statistical techniques | This is more of a statistical approach, no Natural Language Processing / AI techniques were applied  Technique used was word frequency and text analysis. Polarity of text may be applied to gain more insight in the sentiments of travelers  Dataset was from Tripadvisor |
| Fun in The Philippines: Automatic Identification and Sentiment Analysis of Tourism-related Tweets  Julia Camille L. Menchavez  Kurt Junshean P. Espinosa | * The system was able to automatically classify tweets as tourism / non-tourism related * Svm best accuracy and f-score * Most classifiers were unable to deal with short tweets * Naive bayes with n-grams yielded great results * Classifier was unable to correctly classify short and misspelled tweets * Sentiment analysis struggled with neutral sentiments | Naive bayes  Mapbox   1. Identification of tourism/nontourism tweets    1. Feature extraction       1. Bag of words       2. Top words       3. Top words w weights    2. Classification       1. Logistic regression       2. Naive bayes       3. Support vector machines 2. Sentiment Analysis | Study uses AI to automatically classify tweets  Study used AI to analyze sentiments | Does not compare pre-covid lockdown vs post-covid lockdown sentiments  \*recommendations:  Separate words in hashtags  Normalize intentionally misspelled words  Handle neutral sentiment tweets  Tweets with both positive and negative sentiments should be handled as two separate tweets  May be better to focus on a specific city/province rather than the entire country |
| Twitter Sentiment Analysis of Bangkok Tourism During COVID-19 Pandemic Using Support Vector Machine Algorithm  https://doi.org/10.20965/jdr.2021.p0024 | Linear kernel type had the highest accuracy  Most used words were “Food”, “City”, “Temple”  Sentiment of people who wanted to visit Bangkok during lockdown was shown  People piqued interest in Bangkok metropolitan area because of the food and temples while the negative class related to transportation  Strengths and weaknesses of Bangkok tourism were highlighted | Twitter API to collect tweets  Manual annotation of Positive, Negative, and Neutral tweets  Lemmatized using NLTK  SVM  C and gamma shuffle-split cross-validation | This research is very similar to the one that I want to do but in a different locale and lacks the comparison between covid before and after sentiments | Research locale is in Bangkok  They performed comparative studies on different NLP Sentiment analysis tools  There is no comparison between Pre/Post COVID-19 sentiments |
| Natural language processing applied to tourism research: A systematic review and future research directions  Journal of King Saud University – Computer and Information Sciences 34 (2022) 10125–10144 | Most important NLP process: preprocessing, representation methods, machine learning algorithms, and performance metrics  5 major categories of NLP: sentiment analysis, destination branding, question-answering, NLP for assisting in tourism, misc  Preprocessing: tokenization, removing stop words, steaming  Representation methods: Bags of n-grams  Algorithms: Decision trees and SVMs for data classification  Quality of nature of data: noisy data, missing values, class unbalance, and other defects corrected during preprocessing  Hotel reviews and data from tripadvisor were the preferred source  Deep learning techniques are most successful but:  The amount of data  Many languages are involved  Many speak the same language but are not from the same place | 227 studies reviewed  Objective is to know how NLP is used in the tourism industry and to comprehend the current status of NLP research in hospitality and tourism  PRISMA was used to analyze the studies | Provides insight on NLP techniques to use for tourism research | Analyzes studies that use NLP for tourism research |
|  |  |  |  |  |

1. **Introduction**
   1. Tourism in the Philippines
      1. It’s more fun
      2. GDP
   2. COVID-19 Outbreak
      1. Lockdown
      2. Cancelled events (sinulog, etc)
   3. Social networking services
      1. Twitter
   4. Objectives
      1. Identify tourist sentiment toward PH tourism before and after the COVID-19 lockdown
      2. Identify the key factors that affect PH tourism after the COVID-19 lockdown
         1. Safety hazards
         2. Travel protocols
         3. Changes in tourism infrastructure
      3. Identify the impact of COVID-19 on tourism sentiments
2. **Review of Related Literature**
3. **Methodology**