

Defining an Integrated and Computed Methodology Approach for Sentiment and Psychographic Analysis in Tourism Research

Federica Izzo

Department of Education, psychology and communication, University "Suor Orsola Benincasa", Naples, Italy

Quirino Picone

Department of Education, psychology and communication, University "Suor Orsola Benincasa", Naples, Italy

Received: 14 May 2022. **Revision received:** 18 July 2022. **Accepted:** 1 December 2022

Abstract

High-performance computational resources and artificial intelligence-based tools can enhance tourism research and marketing. However, a formal methodological approach using digital technologies in this field is still missing. This research work presents the preliminary results of defining an integrated computational methodology in tourism research and marketing. In addition, the paper aims to provide guidelines for a methodological approach leveraging technological resources and Big Data. The proposed research method is based on online User-Generated Content (UGC) analysis through a psychographic approach based on the Big Five Model, Sentiment Analysis, and Machine Learning techniques. The study is supported by high-performance computing resources, artificial intelligence-based tools, and open-source Python-based software for data collection, text analysis, and psychographic attribution. Results show a remarkable performance of the BFF prediction model and confirm the role of personality in the tourists' decision-making and appreciation of a site. Future developments of this project involve using the acquired structured dataset labeled with sentiment and psychographic attribution to create a further prediction model on tourist segments and appreciation as part of a marketing strategy in tourism management. Future research should push forward the development of further integrated and performing computer-based methodology in tourism research and marketing, leveraging the massive amount of data and the potential of high-performance computing techniques. The main contribution of this research effort is twofold: the definition of a general-purpose BFF/Sentiment Analysis methodology and the development of a prediction model from online UGC based on the Big Five personality traits in the tourism research scenario.

Key Words: Big Five Factor model, Sentiment Analysis, Machine Learning.

JEL Classification: C38, D91, Z30

Reference: Izzo, F., Picone, Q. (2022). Defining an Integrated and Computed Methodology Approach for Sentiment and Psychographic Analysis in Tourism Research. *Journal of Tourism and Services*, 25(13), 1-21. doi: 10.29036/jots.v13i25.393

1. Introduction

Customer profiling and market segmentation based on psychographics may be the foundation of a marketing and communication strategy in tourism. Moreover, digital technologies and procedures can enhance this type of analysis, such as high-performance computational resources and artificial intelligence-based tools. However, a formal methodological approach to using artificial intelligence-based data analysis and knowledge extraction in this field of application is still missing.

The Big Five Factor model is successfully implemented in consumer behavior studies. It is also a suitable marketing analysis, segmentation, and prediction method through Big Data and User-Generated Content (UGC) provided by the Internet and social media. The model, developed by Costa and McCrae (1985), gets back to the idea of personality traits as an explanation of behavior: *“Personality traits, in our model, account (in part) for the motives, habits, and attitudes that directly affect behavior, and thus provide instead an indirect explanation of behavior”* (McCrae, 1995, p. 242). According to the theory, traits can characterize individual thoughts, actions, and feelings stably and can be quantitatively detected (Costa, 1999). Inspecting the English language to find terms to describe personality traits, Costa and McCrae defined a taxonomy of characteristics based on five primary factors, known as the Big Five or the OCEAN model:

1. Openness to Experience (curiosity, need for variety, imagination, and a wide range of interests);
2. Conscientiousness (self-discipline, organization, persistence, and a need for achievement);
3. Extraversion (sociability, warmth, assertiveness, and a need for excitement and positive experiences);
4. Agreeableness (sympathy, cooperation, altruism, and harmony-seeking);
5. Neuroticism (anxiety, hostility, self-consciousness, impulsiveness, and a general predisposition to experience negative affects).

Albeit with some delay compared to consumer behavior studies, this model has recently become popular (Tsiakali, 2018) and well-accepted (Hu, 2018) in tourism research and marketing. Leri (2020) reports that BFFs are prevailing on other scales in tourism and hospitality research, and Bogicevic (2021) highlights the growing attention to psychographic segmentation leveraging the attribution of customers' personality traits. Nevertheless, a methodological gap exists in data acquisition, text analysis, and psychographic attribution. The main reason is that the research approach is not along with the current availability of technological resources and Big Data. Digital technologies offer new opportunities in tourism research. Moreover, digital technologies create new innovative tools for marketing purposes (Civelek et al., 2020; Ključnikov et al., 2021). However, their potential is still unexplored due to the lack of community-accepted practices. This scenario calls for a redefinition of research methods, technologies, and procedures, moving the traditional research approach forward.

Here we present the preliminary steps to define an integrated and computed methodology approach in tourism research and marketing. The proposed paradigm is supported by the extensive use of high-performance computing resources and artificial intelligence-based tools (needed for data acquisition, preparation, processing, and interpretation). Leveraging data science techniques for Big Data automatic learning, the proposed approach consists of two separate phases: training and predicting.

We can resume the training procedure in the following steps: 1) To perform the training phase, leveraging a literature-supported method, first, we associated the psychographic analysis based on the Big Five Factor (BFF) Model with a vector-represented sentence dataset. 2) We collected online User-Generated Contents (UGCs), performing the same vectorization done in the previous step. 3) We used Sentiment Analysis and Machine Learning techniques to train a predictive model able to characterize by BFF components a given text.

In the predicting phase, a generic UGC is positioned in a 5-dimension space where each coordinate axis is one of the BFFs. The main contribution of this research effort is twofold: the definition of a general-purpose BFF/Sentiment Analysis methodology and the development of a prediction model from online UGC based on the Big Five personality traits in the tourism research scenario. From the practical implementation point of view, the research design is based on open-source Python-based software for data collection, text analysis, and psychographic attribution. Data collection is conducted by web scraping UGC (users' reviews and ratings) from the TripAdvisor pages of the Archaeological Park of Pompeii, Herculaneum, and Paestum to build an original and domain-consistent dataset. Text analysis is operated with SEANCE - Sentiment Analysis and Cognition Engine, a Natural Language Processing tool. The reference measure for BFFs is adapted from a 300-item inventory based on Costa and McCrae's NEO-PI-R Facets available in the International Personality Item Pool (IPIP). Finally, BFF assessment is

managed through fuzzy sentiment analysis using Deep Neural Networks. The final result is a structured dataset of processed data from UGC labeled with sentiment and psychographic attribution. Future developments of this project involve using the acquired structured data for a prediction model on tourist segments and appreciation as part of a marketing strategy designed to increase the number of visits to a cultural site or to reallocate tourists from an overcrowded site.

This work is structured as follows: Section 2 presents a literature review on the use of the BFF model in tourism research, knowledge gaps, and the theoretical framework supporting the proposed approach; the tools and methodologies applied in this work for the data collection process, the datasets organization, and, finally, the development of supervised Machine Learning model is described in Section 3; Section 4 introduces the preliminary results of Machine Learning prediction model for BFF attribution; a discussion of results and model performance is provided in Section 5; theoretical contributions, managerial implications, limitations, and further works, are concluded in Section 6.

2. Literature review

Many business studies refer to the BFF model to highlight the role of personality in entrepreneurship (Qin, 2021; Mohammadali, 2019; Petrikova, 2016), leadership (Paulienė, 2012), employee behavior, and motivation (Yousaf, 2019; Ullah, 2020). Among the BFF, Conscientiousness is the trait that recruiters most appreciate in job candidates (Wroblowska, 2015), while creativity (under Openness to experience) is positively associated with entrepreneurship and job satisfaction (Ahmed, 2013). Moreover, Agreeableness, Extraversion, and Openness to experience can enhance organizational attractiveness (Rozsa, 2019). However, despite the diffusion of the BFF model, the taxonomy of personality traits remains an open research area (Wroblowska, 2015; 2019).

This section aims to analyze how the Big Five model of personality is applied in tourism research and marketing from both a theoretical and methodological perspective in the most recent research products from 2015 to 2021. The main focus topics are the use of the BFF model within a methodological research approach, data collection techniques, and the suitability of the instruments used to assess personality traits. Findings mean identifying any knowledge gap and theoretical support for developing an integrated and computational research methodology to leverage the current availability of technological resources and Big Data.

2.1 BFFs considered in tourism research

The most recent literature confirms that the Big Five-Factor model of personality is currently widely applied in tourism research. Some researchers report that previous works only focus on one or two factors of personality (Passafaro, 2015; Kokeny, 2018; Alves, 2020). Nevertheless, from 2015 to the present, most papers consider the whole set of personality dimensions from the Big Five/OCEAN model. Many scholars agree on the moderating role of the Big Five Factors on tourists' motivation (Liao, 2018; Hu, 2018), satisfaction (Delic, 2017; Bogicevic, 2021; Moghavvemi, 2021; Han, 2021), preferences (Delic, 2016; Alves, 2020), intention to revisit a location (Khoi, 2021), affective and cognitive response (Jani, 2015; Hahn, 2018).

The BFF model is often combined with other factors to adapt psychographic segmentation to different research purposes. Additional factors may be grouped into main categories related to: word of mouth (Hu, 2018; Liao, 2018; Tapanainen, 2021; Bogicevic, 2021), tourists' behavior (Tsiakali, 2018; Medina, 2019; Polanco, 2021; Hahn, 2018; Delic, 2016), travel type (Delic, 2017, 2018; Akhrani, 2020; Han, 2021), community behavior (Moghavvemi, 2021; Khoi, 2020; Jordan, 2015), destination features (Jani, 2015; Kokeny, 2018; Leri, 2020; Mesquita, 2018; Wang, 2021). Some scholars also combine the Big Five Factors with other structured models. For example, Tran (2015) applied the five dimensions in

tourist behavior characteristics to explain how personality traits affect customers' choice and satisfaction with different types of recreation. Boodnah (2016) applied the 5E's model for tourists classification to determine how personality traits, coupled with other factors, influence individual appreciation at Port-Louis's 'Porlwi-By-Light' festival (Mauritius). Tsiakali (2018) used the EBM model to find that BFFs support tourists' trust in User-Generated Content (UGC) more than in Marketing-Generated Content (MGC) in the travel decision-making stages. Moghavvemi (2021) and Khoi (2020) considered community factors to examine moderating effects of the BFFs on residents' satisfaction with tourism.

The evidence found in the literature proves that most psychographic studies in tourism research aim to describe tourists' characteristics and behavior. Some scholars discussed the role of personality factors in better understanding tourists' preferences and predicting their travel choices (Akhrani, 2020). Results from the psychographic analysis using the BFF model are considered of utmost importance for marketing (Tapanainen, 2021) and management (Akhrani, 2020) purposes for future development. In agreement with Alves (2020), the descriptive nature of psychographic studies is a limitation that should be overcome. The current availability of data and technologies can answer the need for prediction in the tourism industry. On this trend, some works move forward the analysis step and apply the information retrieved to develop a prediction model on tourist preferences. Within TheRoute project, Mesquita (2018) used the BFF model to define a visitor's profile and allow an intelligent route generation system to suggest personalized routes based on tourists' preferences for the categories of points of interest and personality traits. Alves (2020) related the Big Five personality dimensions with eleven categories of tourist attractions to build a model for a Recommender System that automatically predicts tourist preferences and suggests activities accordingly. Because of the COVID-19 pandemic spread, Polanco (2021) developed a recommendation system to manage crowd levels in interior places or point-of-interests (POI). The recommendation algorithm leverages the BFF model to evaluate a user's personality and predict his/her ratings to the POI proposed, enhancing the decision-making process and trust evolution.

From the above discussion, it is possible to observe that the prediction potential of psychographic analysis with the BFF model remains unexplored. Conversely, most recent works support the concept that the Big Five Factors, possibly combined with other variables, can be used for prediction models.

2.2 Data collection and acquisition methodologies

There is growing interest and discussions on Big Data and the possibilities of current computing technology. Nevertheless, the potential information from unstructured data and User-Generated Content is not effectively exploited in the most used methodologies to acquire and collect information for psychographic analysis in tourism marketing. Questionnaires are the chosen way to gather information for psychographic analysis, even in the most recent works. Questionnaire methodology, with slight differences in the administration method (self-administered, interview, online/email, APP registration), is used in most research in the field. A few studies consider web browsing techniques to collect data from web portals, online platforms, or social media. Rojas (2017) associated the BFFs with ten South American countries to identify their destination personality. Text content was scraped from all web pages (4814 total) of their official English tourism websites and then processed as a single record for each page for computer-aided qualitative and descriptive analysis. Wang (2021) analyzed brand-generated content to assess the personality factors of hotels and peer-to-peer (P2P) accommodation platforms according to the BFF model. Twint, a Twitter scraping tool in Python, was used to obtain a dataset of 33,157 tweets from fourteen P2P and hotel pages. The personality dimensions of each brand were computed through the IBM Watson™ Personality Insights software. Han (2021) examined how the reviewer's personality affected the satisfaction with hotels on TripAdvisor.com. A programmed and customized web crawler was used to collect 43,816 reviews on 751 hotels. Text analysis was performed with a computerized text analysis tool called Linguistic Inquiry and Word Count (LIWC), and the output was submitted to BFF attribution.

Despite being the most used, some scholars agree on the limitations of the questionnaire survey approach. First, a questionnaire's answers can be affected by a social desirability bias. Even when confidentiality of the data and anonymity are assured (Moghavvemi, 2021), individuals may try, deliberately or subconsciously, to present a better image of themselves by reporting false responses (Van de Mortel, 2008). Social desirability response bias is common with questions related to personality and might affect the validity of a test. This risk is taken into consideration in Jani (2015), Verma (2017), Leri (2020), Alves (2020), and Bogicevic (2021). Secondly, questionnaires prepared by English-speaking research teams for non-English-speaking respondents may fall into translation, linguistic, and cultural biases. Back-and-forth translations are done in the questionnaire creation and pre-test phases to avoid or limit linguistic bias, as described in Hahn (2018), Leri (2020), Khoi (2021), Moghavvemi (2021), and Tapanainen (2021). Indeed, cultural and linguistic bias may affect the research results (Jordan, 2015).

A further limitation is the number of samples provided by questionnaire surveys. The size of datasets effectively collected in studies using questionnaires as a data source, after incomplete and invalid responses are excluded, is highly variable, e.g., it may range from 104 samples in Tran (2015) to 997 in Delic (2016). The difference between administered questionnaires and usable responses is substantial; Boodnah (2016) notes that a 50% response rate should be considered a desirable threshold. Hahn (2018) claims that response bias is common when a questionnaire is perceived as too long, and researchers should investigate different methods to overcome this limitation. Sample size does not represent a specific constraint in psychographic analysis based on statistics, as in most of the literature considered in the present review. However, in the case of a prediction model leveraging Machine Learning, a more extensive dataset can lead to more accurate predictions and stable results (Ajiboye, 2015). As reported above, the current availability of appropriate software makes obtaining more data accessible. Moreover, online and social media content language naturally reflects the Big Five personality traits. For these reasons, recent literature is growingly investigating UGC as a more effective and efficient approach for psychological trait assessment (Wang, 2021).

The latest literature highlights the need to overcome the traditional survey approach due to method bias and the need for a massive amount of data. Thus, Web browsing techniques can provide a more extensive and valid psychographic analysis dataset using the BFF model.

2.3 BFF inventories used in tourism research

According to the literature, different scales are used for BFF assessment in research works. A BFF test is usually a variable list of short statements rated in a Likert-scale format according to respondents' agreement with each item. BFF attribution differs according to data collection methods: questionnaire surveys include a section with BFF items, and the acquired dataset already contains information on the BFFs; web browsing techniques obtain a dataset with no BFF label, and psychological factors are attributed at a later stage. The use of validated inventories confirms the scientific nature of psychographic research. For this reason, the scales applied in tourism research papers are tested and approved in previous studies.

Many measures for the BFF model were developed from the NEO Personality Inventory (NEO-PI) by Costa (2008) and subsequent versions and are currently used in tourism research (Delic, 2017). Goldberg (1990) developed and tested a set of Big-Five markers composed of 100 synonym adjective clusters to define a more economical BFF measure. Later, this and other measurement tools were collected in the International Personality Item Pool (IPIP), an inventory of more than 3000 items and 250 scales for personality assessment collected by the Oregon Research Institute (Goldberg, 1999, 2006). IPIP is a free reference source for psychographics and consumer behavior researchers, and the IPIP website enlists 877 scientific papers using the inventories published between 1999 and 2019. Inventories from Goldberg and the IPIP project are used by Hu (2018), Hahn (2018), and Alves (2020). In addition, smaller inventories have been developed, such as the 44-item BFI (John, 1999) used in Mesquita (2018),

Khoi (2020), and Moghavvemi (2021). The 25-item BFF scale (Yoo, 2011) was employed to prove the influence of personality on travelers' motivation to use and create consumer-generated media. It was also applied in Jani (2015), Tsiakali (2018), Wu (2019), and Leri (2020). One of the most used BFF measurement models is the 20-item scale (Mini-IPIP) (Donnellan, 2006), specifically designed for a questionnaire survey data collection approach. This inventory is used in Jordan (2015), Kvasova (2015), Delic (2017, 2018), Verma (2017), and Tapanainen (2021).

Literature shows a general tendency to consider a reduced number of inventory items. The questionnaire survey data collection approach induces this trend. Even the studies that apply the NEO-Five-Factor model created by Costa and McCrae (Costa, 2008) use only a limited selection from the original items list, as in Kokeny (2018) and Khoi (2021). The research based on a computed data collection approach diverges from this trend. In Rojas (2017), the complete NEO-Five-Factor set is exploited for destination personality assessment. Han (2021) retrieved his psychographic attribution approach from Yarkoni (2010) to find correlations between semantic categories of textual data processed with LIWC and the Big Five personality traits.

Research in tourism reports the need to contain the number of items in a questionnaire survey as a primary reason for short BFF assessment scales. Khoi (2021) observes that using long scales in a questionnaire can cause boredom, lack of attention and care in the testing sample, and, as a result, affect the reliability of data and the study's validity. Although short scales proved to be a good and reliable instrument for BFF attribution (Gosling, 2003), they do not provide an assessment of lower-order personality facets (Kvasova, 2015). On the contrary, generating higher-level BFF domains from lower-level facets makes comprehensive inventories more useful in some research contexts (Goldberg, 1999). In tourism marketing, lower personality facets appreciable in user-generated and social media content can be considered a driver of consumer choices (Tsiakali, 2018).

Much research has argued that psychographic measurements tailored for a tourism context should replace poorly performing tools (Jordan, 2015). For this purpose, currently available high-performance computational resources can carry out a more advanced text processing (Rojas, 2017). The availability of more significant amounts of data and the need for deeper insights into tourists' behavior for marketing and management applications call for more specific and rigorous measures (Kvasova, 2015), possibly with psychometrically complex statements (Kokeny, 2018).

Using computational data recollection methods overcomes the need to apply a short BFF inventory. It offers the possibility to exploit the potential of a more vast scale for a more precise assessment of personality domain facets. More extensive scales can provide specific measures for the Big Five Factors and deeper insight into personality traits attribution. Based on the above conversation, we can hypothesize that using a more exhaustive BFF inventory for personality assessment is more effective in a computed methodology approach and can provide a deeper understanding of tourists' behavior.

Findings in the literature analyzed show a methodological gap in data acquisition, text analysis, and psychographic attribution. Furthermore, current studies do not refer to a shared research path leveraging Big Data availability and technological, computed-aided instruments. Therefore, the methodology proposed in the following sections provides a foundation for further development in tourist behavior studies based on data mining, sentiment analysis, and Machine Learning software.

3. Methods

This section describes the software tools, data sources, data collection techniques, and the dataset creation procedure. The paper aims to provide guidelines for a new methodological approach to tourism research and marketing leveraging technological resources and Big Data. To this aim, we provide a practical implementation of the methodology using Python-based open-source software, publicly available data, and Machine Learning.

3.1 Tools

Python is a high-level, multi-paradigm (functional, procedural, and object-oriented) programming language. It is widely used in data science and Machine Learning applications. The Python-based software used to accomplish the primary purpose of this work are:

1. for data gathering and acquisition: Selenium WebDriver;
2. for sentiment analysis: Sentiment Analysis and Cognition Engine (SEANCE);
3. for neural networks and Machine Learning: TensorFlow, Keras.

These resources are freely available, open-source, and downloadable online.

Selenium WebDriver is a project embedding a collection of libraries and tools to browse the web automatically. It was developed at the global software consultancy Thoughtworks in Chicago in 2004 by J. Huggins for automatic application testing. It evolved into a web testing framework able to emulate user interaction with the major browsers (Firefox, Safari, Edge, Chrome, Internet Explorer). Due to the possibility of extending the core software with a range of plugins and drivers, this tool supports a whole ecosystem of open-source projects.

SEANCE - Sentiment Analysis and Cognition Engine is a Natural Language Processing tool developed by professors S. A. Crossley, K. Kyle (Georgia State University), and D. S. McNamara (Arizona State University) in 2016 (Crossley, 2017). It is easy to use, customizable, and freely downloadable for Mac and Windows. A Python implementation is available open-source on Github. SEANCE turns text into vectors using eight freely available, domain-independent source databases (*General Inquirer*, *Lasswell*, *Geneva Affect Label Coder*, *Affective Norms for English Words*, *EmoLex*, *SenticNet*, *Valence Aware Dictionary for Sentiment Reasoning*, *Hu-Liu polarity*). In addition, 2725 linguistic micro-features, 20 aggregated macro-features on sentiment, cognition, and social order, a negation control feature, and part of speech (POS) tags are also available. The software processes texts as plain-text (.txt) files in a single input folder. The output is a comma-separated values (.csv) spreadsheet containing all the texts processed, one for each row. The default analysis includes 20 component scores, but it is possible to customize the indices and the individual databases to include, as well as the POS and negation control features. Component score values are Z-score normalized. Complete documentation on SEANCE installation, features, index overview, and a list of all papers citing or using this software are available on the SEANCE web portal.

TensorFlow is a programming model for Machine Learning (ML) and Deep Neural Networks implementations. It was developed for internal use as an evolution of a Google Brain project and open-sourced in 2015. It manages the data flow of multidimensional arrays (tensors), setting the communication between different nodes, and represents both the algorithm computation and state in a unified graph (Abadi, 2016). It is currently used in many Google products, such as *Search*, *Maps*, and *YouTube*, and ML applications by different industries and companies. *Keras* is a TensorFlow-based framework providing APIs and layers to make deep learning models faster and more user-friendly.

3.2 Data acquisition and organization

We collected two different datasets from publicly available sources. The reviews dataset is composed of UGC from TripAdvisor reviews and ratings. The BFF dataset comprises text labeled with psychographic attribution and was adapted from the International Personality Item Pool (IPIP). Details on data collection techniques, data processing, and the composition of the output datasets are described below.

3.2.1 The reviews dataset

Investigations on Big Data from social media, e-commerce, and online review sites have generated many online corpora based on UGC, such as tweets, posts, and reviews. Some are also freely

available, such as the Stanford Large Network Dataset Collection (Leskovec, 2016). In the online tourism scene, TripAdvisor can be considered one of the largest travel platforms in the world, with over 859 million user reviews and opinions, as stated on the web portal. Annotated datasets specifically crawled from TripAdvisor were used in Natural Language Processing (NLP) tests on sentiment prediction. Wang (2010) collected a 235,793 hotel reviews dataset to test their model to solve the Latent Aspect Rating Analysis (LARA) problem. From Wang’s publicly available dataset, Marcheggiani (2014) randomly extracted and labeled a set of 442 reviews to evaluate aspect-specific opinions at the sentence level. An extension of this English-only dataset, with the addition of Italian and Spanish reviews, was built by Jimenez-Zafra (2015) to provide a multilingual corpus for aspect-oriented opinion mining.

Some studies used data recollected from TripAdvisor for text and sentiment analysis on an Italian basis. For example, Clarizia (2018) used online reviews to analyze sentiment from users’ experiences in the cultural heritage sites of Pompeii, Paestum, and the National Park of Vesuvius. In addition, Cortese (2019) performed a quali-quantitative preliminary analysis on TripAdvisor data to boost cultural and religious tourism in Pompeii. Nevertheless, to the best of our knowledge, no dataset of TripAdvisor reviews on the three most important archeological areas in the Campania region (Pompeii, Herculaneum, Paestum) is currently available.

Figure 1. Data collection from TripAdvisor pages

ARCHAEOLOGICAL PARK OF POMPEII	ARCHAEOLOGICAL PARK OF HERCULANEUM	ARCHAEOLOGICAL PARK OF PAESTUM
Overall evaluation: 4.5/5	Overall evaluation: 4.5/5	Overall evaluation: 4.5/5
Total reviews: 22389	Total reviews: 7867	Total reviews: 4946
Language: English (11259)	Language: English (4387)	Language: English (1110)
Rating: 1-5	Rating: 1-5	Rating: 1-5
Dates: 08/2005 - 10/2020	Dates: 08/2011 - 01/2020	Dates: 10/2010 - 11/2020
		

Source: authors’ own contribution. Pictures from TripAdvisor.

Due to the need to gather data from online UGC, an original and domain-consistent corpus was built. Data were web scraped from the TripAdvisor pages of the Archaeological Park of Pompeii, the Archaeological Park of Herculaneum, and the Archaeological Park of Paestum, last consulted on 10 Feb. 2021. Data recollection for each archeological site considered reviews written in English and with all rating values (1 to 5, according to the TripAdvisor scale), as shown in Fig. 1.

Han (2021) pointed out that most available scientific literature considered either review rating or review text. On the contrary, a joint analysis of both variables can offer a more accurate insight into tourists' experience and satisfaction. The variance in the period considered for each site is due to the attempt to get equal reviews for all ratings, despite the total number of reviews difference.

Data was collected by a Selenium WebDriver-based scraper sorting the main elements of the review, extending and improving an already available software component as open-source (Gambino, 2019). Modifications were needed to fit the last upgrades to the TripAdvisor page structure. Additions to the original software allowed it to manage rating and language selection, previous and next page

navigation, and avoid automatic reviews translation from English to Italian. The final web scraper developed for this research is available on GitHub (<https://github.com/Fede83/Scraping-TripAdvisor-with-Python-2020>). Data was collected on the posting date, rating, title, body of the review, and the "read more" button, as shown in Fig. 2. No personal information about the reviewer's profile was collected.

Figure 2. TripAdvisor page structure and data selection



Source: authors' own contribution. Picture from TripAdvisor.

Figure 3. The total number of reviews acquired from TripAdvisor

POMPEII			HERCULANEUM			PAESTUM		
Rating	Pages	Reviews	Rating	Pages	Reviews	Rating	Pages	Reviews
5	25	125	5	25	125	5	25	125
4	25	125	4	25	125	4	25	125
3	118	591	3	28	139	3	4	20
2	37	182	2	8	38	2	2	7
1	17	82	1	3	13	1	1	2

Source: authors' own contribution.

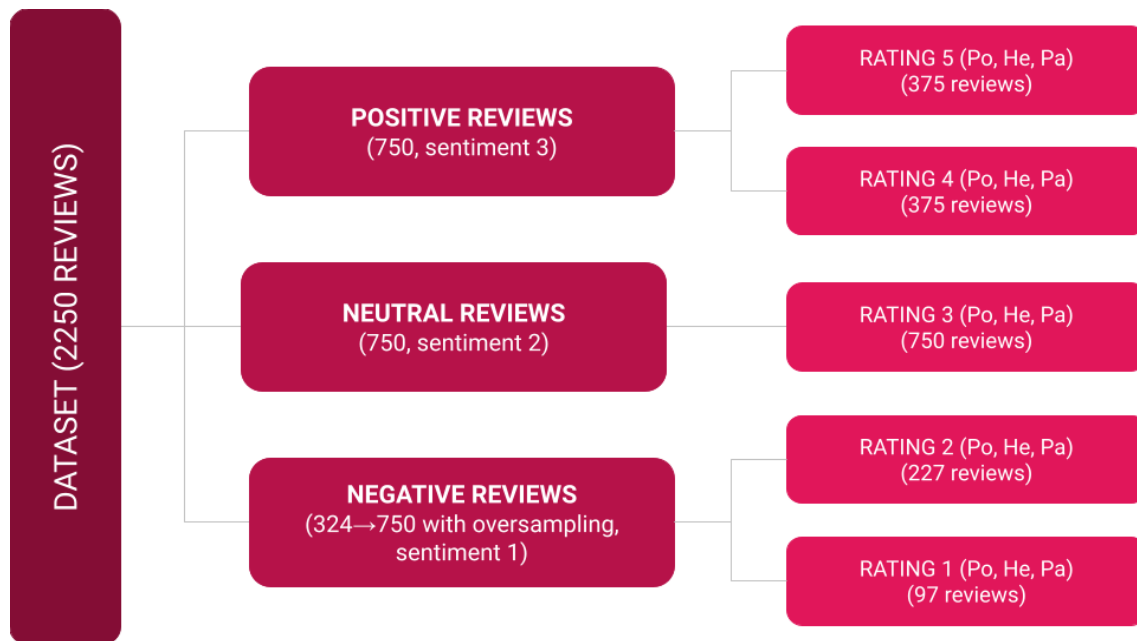
Fig. 3 shows the total number of reviews acquired from TripAdvisor. The considerable disparity among the number of browsed pages and downloaded reviews for each rating and site is functional to build a balanced dataset.

The dataset was created by grouping all reviews from Pompeii (Po), Herculaneum (He), and Paestum (Pa), according to their TripAdvisor ratings, into three sentiment classes to facilitate the neural network analysis and labeling:

1. sentiment 3 or positive includes ratings 4 and 5 (750 reviews);
2. sentiment 2 or neutral corresponds with rating 3 (750 reviews);
3. sentiment 1 or negative includes ratings 1 and 2 (324 reviews).

Within this data structure, negative reviews are distinctly underrepresented. The reviews labeled with sentiment 1 or negative were randomly oversampled to the number of 750 to match classes 2 and 3 and avoid performance errors in the prediction model accuracy caused by an unbalanced dataset, as pointed out by Chawla (2002), t. Although not functional in the current paper, this step is necessary to make the dataset suitable for further research developments. The composition of the resulting dataset, with overall 2250 reviews, is shown in Fig. 4.

Figure 4. **Composition of the reviews dataset**



Source: authors' own contribution.

3.2.2 The BFF dataset

A second dataset, labeled with the Big Five Factors, was built as a training set for BFF attribution. Data were retrieved from the International Personality Item Pool (IPIP). A 300-item inventory based on Costa and McCrae's NEO-PI-R Facets (Costa, 2008), with positive and negative keys, was selected among all available resources. Before the SEANCE analysis, the dataset was pre-processed as follows:

1. all items were reworded in complete sentences, adding the first singular and plural personal pronouns, generally the most used pronouns in reviews;
2. negative-key items were converted into positive-key sentences adding negative verb forms or adjectives;
3. the 30 facet labels were grouped in the corresponding Big Five Factors of personality.

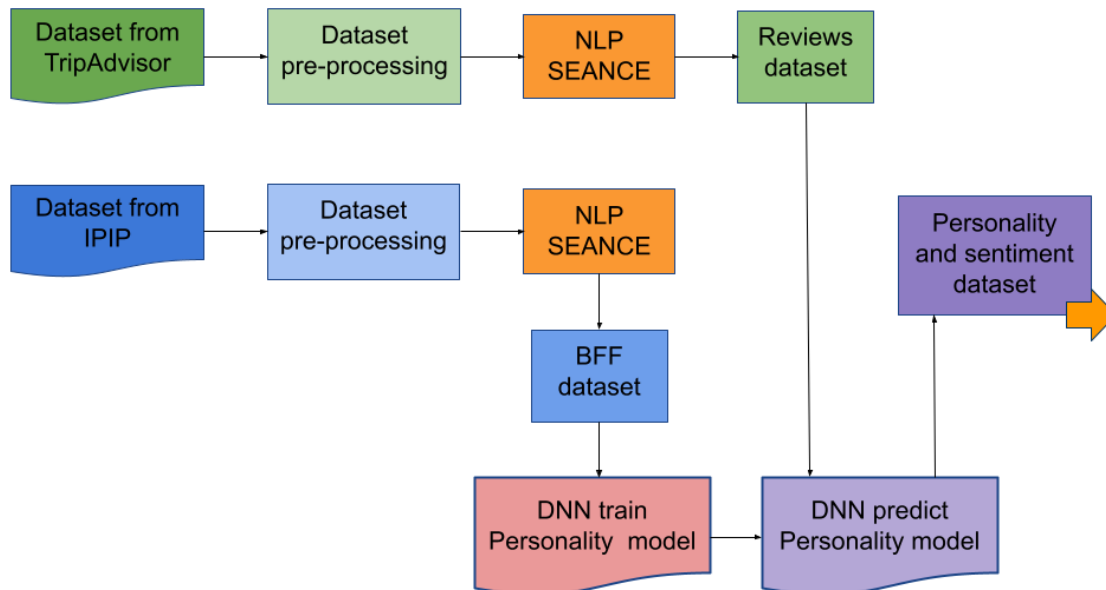
The resulting dataset comprises 300 sentences labeled according to the psychographic attribution (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) stated within the inventory.

3.3 Machine Learning models for BFF attribution

In order to perform the two separate phases of the proposed Machine Learning approach for psychographic attribution, training, and predicting, both the BFF dataset and the reviews dataset were vectorized using SEANCE. In this analysis, no one of the individual databases included in SEANCE was selected, all words were analyzed, and the POS and negation control features were taken into account.

The final result consists of two different datasets: a reviews dataset with one column for the number of words, 20 columns for SEANCE component scores, and a label column with the corresponding sentiment; a BFF dataset with one column for the number of words, 20 columns for SEANCE component scores, and a label column with the corresponding BFF. For both datasets, the “number of words” column was intentionally dropped, as this feature is considered not relevant in the present stage of the research. The data flow diagram is shown in Fig. 5.

Figure 5. Data flow diagram



Source: authors' own contribution.

3.3.1 Training phase: BFF classification model

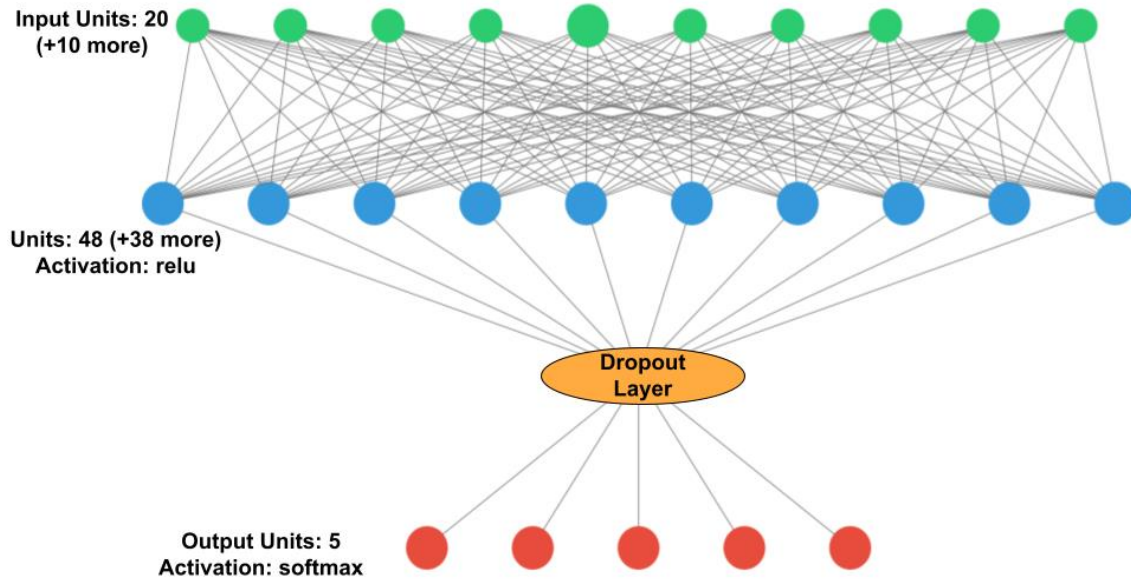
The BFF dataset was further processed by changing BFF labels to numeric data type and categories and split into an 80% train set and 20% test set. Then, it was used for supervised learning in a multi-class Deep Neural Network (DNN) classification model for BFF attribution. The model, leveraging the Tensorflow/Keras framework, is built as follows:

1. the input layer contains 20 neurons as the 20 SEANCE component scores selected as feature parameters;
2. the first dense layer contains 48 neurons, as a result of a fine adjustment process, and ReLu activation to manage the input weights;
3. the regularization dropout layer prevents model overfitting;
4. the output layer contains five neurons as the different BFF labels and Softmax activation to give a fuzzy output, assigning a decimal probability to each class.

The model architecture is shown in Fig. 6.

The “Categorical Cross Entropy” function computes the loss between the labels and predictions. This class is specifically for a model with a one-hot representation of two or more label classes and floating-point values per feature. The number of epochs is regularized by the EarlyStopping callback, which halts training when the model stops improving. In our model, reiteration stops when validation loss reaches its minimum.

Figure 6. The BFF classification model architecture

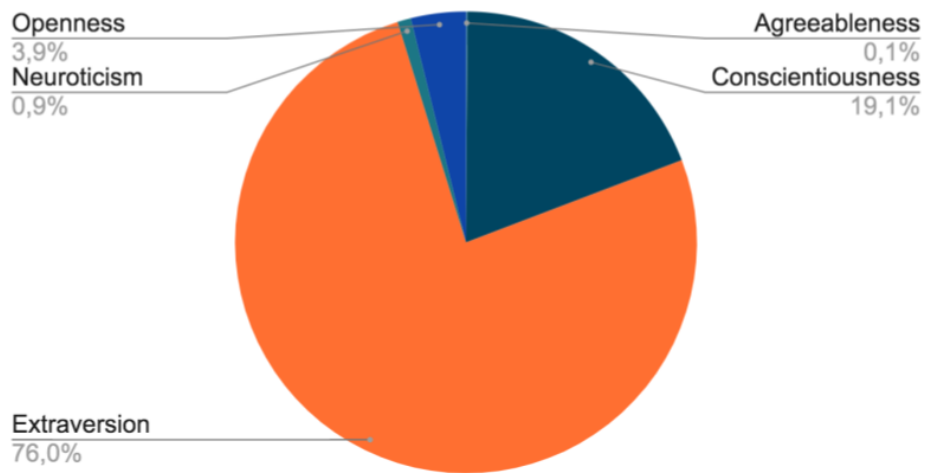


Source: authors' own contribution.

Figure 7. Example of an item from the output structured dataset: a review with sentiment and psychographic attribution

Review #14, sentiment 3, BFF: Extraversion

"Massive- spend the whole day walking If you get there by public transport do not listen to the guides by the train station who will tell you to go upstairs to get tickets. Keep walking and you will get there in 2 mins and buy your tickets directly from there. The site is massive, do expect to spend whole day walking around."



Source: the authors' own contribution.

After the fine tuning of the variables, the model hyperparameters are:

1. learning_rate = 0.001;
2. epochs = 4096 (a random deliberately high number that will never be reached due to EarlyStopping);
3. batch_size = 16;

4. `validation_split = 0.1`.

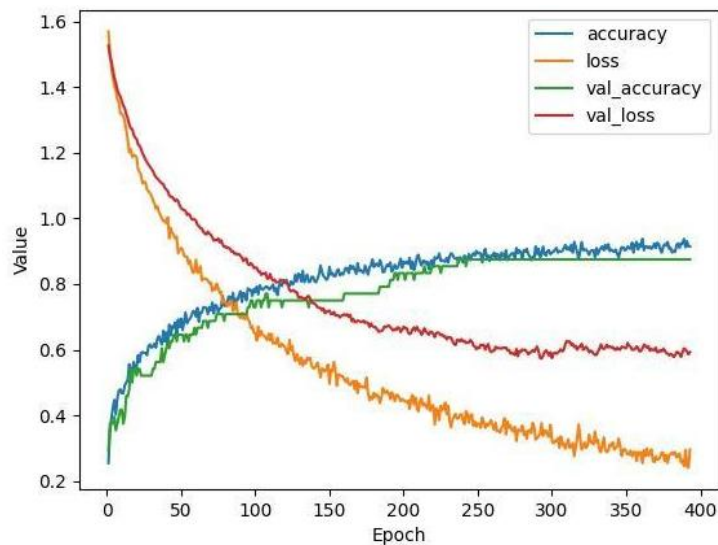
The model output predicts a decimal probability of the BFF attribution for each item.

The BFF classification model was saved and applied to perform psychographic predictions on the reviews dataset via the Keras *predict* function. The output predictions generated are a decimal probability of the BFF attribution for each review. Only the BFF label corresponding to the highest prediction value was chosen for each review for data readability reasons. Therefore, this study does not consider the label column with the corresponding sentiment, but it is preserved for future analysis. The final output of this model is a structured dataset of processed data from UGC labeled with sentiment and psychographic attribution. An output example is shown in Fig. 7. The BFF prediction model and the resulting structured dataset are available on GitHub (<https://github.com/Fede83/bff>).

4. Results

The preliminary results of the BFF classification model showed that a larger train dataset was needed to perform satisfactorily. For this reason, the BFF dataset was augmented, oversampling to 120 items for each class (against the original 60 items per class, 300 total) before the train-test dataset split. Such alteration resulted in a lower test loss and a higher test accuracy. For example, Fig. 8. shows about 80% accuracy on the test set, although we should expect slightly different values each time we run the model.

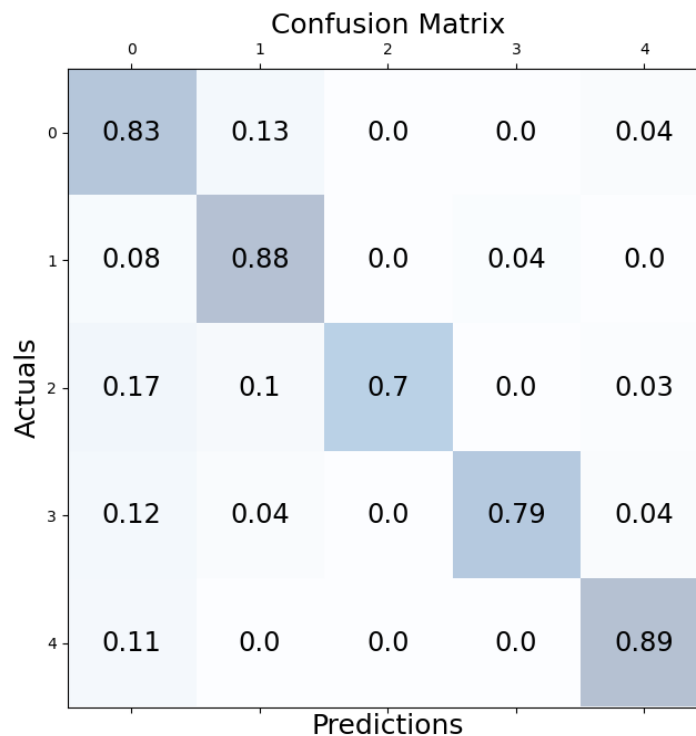
Figure 8. **BFF classification model performance after dataset augmentation**



Source: the authors' own contribution.

A normalized confusion matrix was built to summarize and evaluate the BFF classification model prediction performance. Such a technique is helpful to get a deep insight into the prediction accuracy and errors for each class in the test set. This confusion matrix comprises the five BFF classes, where the x axis represents the classes predicted by the model. In contrast, the y axis represents the actual classes labeled in the dataset. Values are a relative percentage of the items for each class in the BFF test set. The results in Fig. 9 show a remarkable percentage of correct predictions for each class: 83% for Agreeableness, 88% for Conscientiousness, 70% for Extraversion, 79% for Neuroticism, and 89% for Openness. Further performance evaluation, e. g., using statistical methods or robust indices, is behind the scope of this work.

Figure 9. Confusion matrix on BFF classification model prediction performance (0=Agreeableness, 1=Conscientiousness, 3=Extraversion, 4=Neuroticism, 5=Openness)



Source: the authors' own contribution.

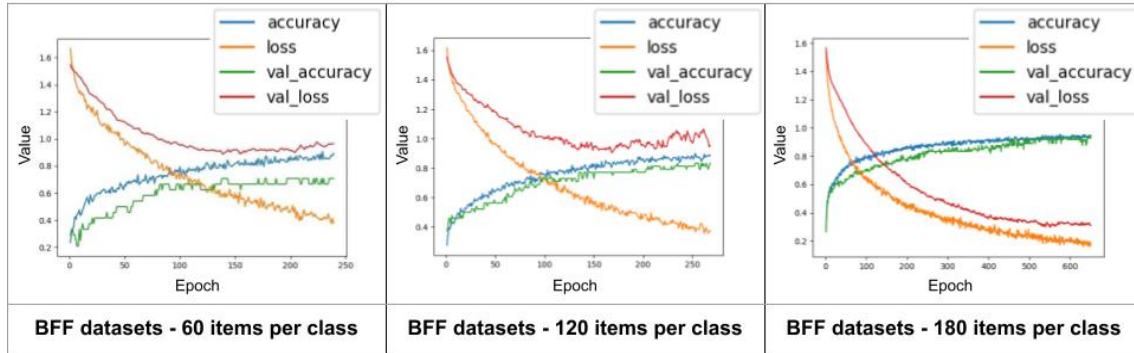
5. Discussion

Although this paper takes an essential step towards defining a methodological paradigm in tourism research and marketing, there is a significant limitation due to the number of items in the BFF inventory. Therefore, the BFF classification model dataset was oversampled from the original 60 to 120 items for each class to achieve a better model performance. Further tests showed better results for loss and accuracy values for a higher data oversampling index, as shown in Fig. 10.

The downside of the oversampling technique is that the model reiterates the same items, raising the risk of overfitting. This result suggests that using a more extensive inventory as the full NEO-PI-R (Costa, 2008) and subsequent versions may improve the Machine Learning classification model. Moreover, leveraging the availability of UGC on online platforms, a more extensive reviews dataset could also avoid the need to oversample an imbalanced dataset and provide a deeper insight into tourists' segments and preferences.

Observations of the BFF prediction model results confirm the distinctive characteristics attributed by Costa and McCrae (1985) to each BFF and prove a connection between personality and reviewer satisfaction. For example, the Openness trait is identified by curiosity, need for variety, imagination, and a wide range of interests. Furthermore, the reviews labeled with Openness are the most numerous in the final dataset: this shows the willingness of their authors to travel abroad and discover different cultures. Indeed, although reviewers' nationality is not a type of data collected in this study, it is highly improbable that an Italian would write a review on a touristic site in Italy in a language other than Italian.

Figure 10. BFF classification model performance with 60, 120, and 180 items per class



Source: the authors' own contribution.

Moreover, most Openness-labeled reviews are associated with positive sentiment (sentiment 3): this shows an aptitude to overcome the unfamiliarity of a new place for the sake of experience and knowledge. The high number of visitors is consistent with Alves's (2020) analysis which demonstrates that individuals characterized by Openness to Experience prefer visiting museums and enjoying historical artifacts and are satisfied with aesthetic experiences. The high sentiment rank is in line with findings on the positive influence of Openness in interest-type travel curiosity (Jani, 2014). Results comply with Han's (2021) demonstration of the positive impact of this personality trait on overall travel satisfaction and the content and rating of online reviews.

Table 1. BFF prediction model results

BFF	Sentiment 1	Sentiment 2	Sentiment 3	Total
Openness	201	254	388	843
Conscientiousness	139	118	153	410
Extraversion	96	147	122	365
Agreeableness	59	75	59	193
Neuroticism	255	156	28	439

Source: the authors' own contribution.

The second most represented BFF in the result dataset is Neuroticism, which is associated with the highest number of negative reviews (sentiment 1). This result confirms the general predisposition of people identifying with this psychological trait to experience negative feelings and anxiety, probably caused by impulsiveness in the trip planning or a lack of predisposition to overcome inconveniences in an unfamiliar place easily. The findings are consistent with previous studies. The high number of visitors rated high on Neuroticism confirms that these individuals prefer to choose well-known and tested destinations more than try something for the first time (Kahle, 2005). Nevertheless, the low sentiment rate suggests that these individuals may enjoy more predictable and controlled vacation activities, like going to the beach, swimming, and sunbathing (Delic, 2016). This result is also in line with Khoi's (2021) confirmation of the moderating role of Neuroticism in the travel experience and the negative influence of this trait on travel inspiration and emotional response.

A summary of findings is shown in Table 1.

Personality plays a role in the tourism decision-making process, in the disposition to share one's own experience online, and in appreciating the site visited. For this reason, the definition of users according to the Big Five model may provide a deep insight into travelers' profiling and segmentation.

Despite being considered demanding in terms of dataset preparation, model development, and testing, ML classification seems suitable for content and sentiment analysis in a skyrocketing unstructured data availability scenario (Kirilenko, 2017). Our preliminary results show an overall remarkable performance in classification, and prediction shows coherence with manual psychographic attribution made in previous studies using traditional BFF assessment inventories. The novelty of the proposed approach is to put together well-grounded psychological theory, high-performance computing techniques, and open source tools in a one-of-a-kind methodology to contribute to the advancement in tourism and marketing research.

6. Conclusion

This study answers a methodological gap in tourism research and marketing, proposing the application of the Big Five Factor model, data collection techniques, and personality assessment tools within a methodological research approach. The availability of online UGC and high-performance computational resources supports the proposed paradigm.

The methodology described in this paper and the related results from the reported use case imply a twofold novel contribution to the advance of knowledge about the use of artificial intelligence in tourism and services: theoretical contributions and managerial contributions.

Regarding theoretical contributions, first, a general-purpose BFF/Sentiment Analysis methodology was defined to develop a psychographic assessment model leveraging only open-source, freely available instruments, such as SEANCE and IPIP. To the best of our knowledge, no correspondence existed between SEANCE component scores and the BFF model, unlike other paid sentiment analysis software, e.g., the BFF attribution to the Linguistic Inquiry and Word Count (LIWC) output by Yarkoni (2010). Second, the new opportunities currently offered by technological resources and Big Data are still unexplored in tourism research and marketing, different from consumer behavior studies. Therefore, this scenario calls for a redefinition of research methods, technologies, and procedures, moving forward with the traditional approach. Based on the examined literature, this study is one of its first kind to apply computed data recollection from online UGC to build a prediction model based on the Big Five personality traits and user preferences in the context of an Italian tourism scenario. Moreover, when publicly released, the produced datasets represent a valuable data source that can be reemployed in future research studies.

Market segmentation and customer profiling proved crucial as demand, offer, and communication currently acts at a microtargeting level. However, UGC analysis and psychographic attribution can offer a deeper insight into tourism targets and boost data-driven, actionable strategies, from tourism marketing and communication to territorial management. Regarding managerial contributions, the development of this work can involve different fields, from planning an advertisement campaign to improving strategic management of a single cultural heritage site, a cluster of similar sites, or at a regional level. For example, it may be the foundation of a marketing strategy designed to increase the number of visits to a cultural site or to reallocate tourists from an overcrowded site, promoting cooperation among different territorial stakeholders and providing a more sustainable, safe, and satisfactory travel experience. Moreover, the methodological paradigm proposed, unlike customer analysis and segmentation operated by private organizations specialized in buyer personas profiling or by big business platforms (Facebook/Meta for Business, Google Ads and Analytics), allows one to keep complete control of data and manage data analysis according to the individual needs and plans of action.

Future research should push forward the development of further integrated and performing computer-based methodology in tourism research and marketing, leveraging the massive amount of data currently available and the potential of high-performance techniques.

In the present research stage, a structured dataset of UGC labeled with sentiment and psychographic attribution has been produced. Although the classification and prediction performance gave remarkable preliminary results, further confrontation with human, statistical, and other automated classification and rating systems is needed to assess the model's accuracy and efficiency. Future developments of this work involve using such an acquired structured dataset for a prediction model on tourist segments and appreciation as part of a marketing strategy designed to increase the number of visits to a cultural site or reallocate tourists from an overcrowded site.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: A System for {Large-Scale} Machine Learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265-283).
2. Ahmed, S. N., Rehman, A., & Amjad, M. (2013). Job satisfaction and intrapreneurship: The moderating effect of personality trait. *Journal of International Studies*, 6(1).
3. Ajiboye, A. R., Abdullah-Arshah, R., & Hongwu, Q. (2015). Evaluating the effect of dataset size on predictive model using supervised learning technique.
4. Akhrani, L. A., & Najib, A. (2020). Tell me who you are and i will tell you where to go: Examining the role of big five personality in soft-adventure traveling type. *Geo Journal of Tourism and Geosites*, 32(4), 1299-1303.
5. Alves, P., Saraiva, P., Carneiro, J., Campos, P., Martins, H., Novais, P., & Marreiros, G. (2020, July). Modeling Tourists' Personality in Recommender Systems: How Does Personality Influence Preferences for Tourist Attractions?. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization* (pp. 4-13).
6. Bogicevic, V., & Bujisic, M. (2021). Comparison of composites, dichotomous, and latent factor measurement operationalizations in hospitality research on moderating effects. *Cornell Hospitality Quarterly*, 62(1), 121-138.
7. Boodnah, K. D., Armoogum, V., Jaunky, V. C., & Armoogum, S. (2016, August). Towards smart tourism: an individual appreciation of porlwi-by-light festival. In *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)* (pp. 323-328). IEEE.
8. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
9. Civelek, M., Gajdka, K., Světlík, J., & Vavrečka, V. (2020). Differences in the usage of online marketing and social media tools: evidence from Czech, Slovakian and Hungarian SMEs. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(3), 537-563. doi: 10.24136/eq.2020.024
10. Clarizia, F., Colace, F., De Santo, M., Lombardi, M., & Pascale, F. (2018, October). A sentiment analysis approach for evaluation of events in field of cultural heritage. In *2018 Fifth international Conference on social Networks analysis, management and Security (SNAMS)* (pp. 120-127). IEEE.
11. Cortese, F., D'Ambrosio, I., & Petracca, M. (2019). A possible synergy between culture and religion for the sustainability of tourism of Pompeii. *Sustainability*, 11(8), 2231.
12. Costa Jr, P. T., & McCrae, R. R. (2008). The Revised Neo Personality Inventory (neo-pi-r).
13. Costa, P. T., & McCRAE, R. R. (1999). A five-factor theory of personality. *The Five-Factor Model of Personality: Theoretical Perspectives*, 2, 51-87.

14. Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior research methods*, 49(3), 803-821.
15. Delic, A., Neidhardt, J., Nguyen, T. N., & Ricci, F. (2018). An observational user study for group recommender systems in the tourism domain. *Information Technology & Tourism*, 19(1), 87-116.
16. Delic, A., Neidhardt, J., Rook, L., Werthner, H., & Zanker, M. (2017). Researching individual satisfaction with group decisions in tourism: experimental evidence. *Information and communication technologies in tourism 2017*, 73-85.
17. Delić, A., Neidhardt, J., & Werthner, H. (2016). Are sun lovers nervous. *E-Rev. Tourism Res.(eRTR)*, ENTER.
18. Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, 18(2), 192.
19. Durán-Polanco, L., & Siller, M. (2021). Crowd management COVID-19. *Annual reviews in control*, 52, 465-478.
20. Gambino, G., & Pirrone, R. (2019, November). Investigating Embeddings for Sentiment Analysis in Italian. In *NLAAI@ AI* LA*.
21. Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6), 1216.
22. Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7(1), 7-28.
23. Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1), 84-96.
24. Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6), 504-528.
25. Hahn, S. S., Preis, M. W., Lee, C. K., & Kim, M. J. (2018). Effects of affective and cognitive responses on loyalty of mega-event visitors and the moderating role of big five personality traits. *관광연구저널*, 32(10), 87-101.
26. Han, M. (2021). Examining the effect of reviewer expertise and personality on reviewer satisfaction: an empirical study of TripAdvisor. *Computers in Human Behavior*, 114, 106567.
27. Hu, Y., & Kim, H. J. (2018). Positive and negative eWOM motivations and hotel customers' eWOM behavior: Does personality matter?. *International Journal of Hospitality Management*, 75, 27-37.
28. Jani, D. (2014). Relating travel personality to Big Five Factors of personality. *Tourism: An International Interdisciplinary Journal*, 62(4), 347-359.
29. Jani, D., & Han, H. (2015). Influence of environmental stimuli on hotel customer emotional loyalty response: Testing the moderating effect of the big five personality factors. *International Journal of Hospitality Management*, 44, 48-57.
30. Jiménez-Zafra, S. M., Berardi, G., Esuli, A., Marcheggiani, D., Martín-Valdivia, M. T., & Fernández, A. M. (2015, September). A multi-lingual annotated dataset for aspect-oriented opinion mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2533-2538).
31. John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999), 102-138.
32. Jordan, E. J., Vogt, C. A., & DeShon, R. P. (2015). A stress and coping framework for understanding resident responses to tourism development. *Tourism Management*, 48, 500-512.
33. Kahle, L. R., Matsuura, Y., & Stinson, J. (2005). Personality and personal values in travel destination preference. *ACR Asia-Pacific Advances*.

34. Khoi, N. H., Le, A. N. H., & Tran, M. D. (2021). Tourist inspiration and its consequences: The moderating role of neuroticism. *International Journal of Tourism Research*, 23(5), 901-913.
35. Khoi, N. H., Phong, N. D., & Le, A. N. H. (2020). Customer inspiration in a tourism context: an investigation of driving and moderating factors. *Current Issues in Tourism*, 23(21), 2699-2715.
36. Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 57(8), 1012-1025.
37. Ključnikov, A., Civelek, M., Vavrečka, V. and Nėtek, V. (2021). The Differences in the Usage of Social Media between SMEs operating in the Iron and Mining Industries. *Acta Montanistica Slovaca*. 26 (2) 185-194 DOI: <https://doi.org/10.46544/AMS.v26i2.01>
38. Kokeny, L., & Kiss, K. (2018). The relationship between personality types and consumer behavior in tourism. *Ekonomické Rozhľady/Economic Review*, 47(2).
39. Kvasova, O. (2015). The Big Five personality traits as antecedents of eco-friendly tourist behavior. *Personality and individual differences*, 83, 111-116.
40. Leri, I., & Theodoridis, P. (2020). How do personality traits affect visitor's experience, emotional stimulation and behaviour? The case of wine tourism. *Tourism Review*.
41. Leskovec, J., & Krevl, A. (2016). SNAP Datasets: Stanford large network dataset collection; 2014. URL <http://snap.stanford.edu/data>, 49.
42. Liao, Y. K., Nguyen, P. T., & Thi, K. N. N. (2018). Examining the role of online customer reviews on tourist motivation: destination and revisit intention. *The Journal of Social Sciences Research*, 4(12), 413-422.
43. Marcheggiani, D., Täckström, O., Esuli, A., & Sebastiani, F. (2014, April). Hierarchical multi-label conditional random fields for aspect-oriented opinion mining. In *European conference on information retrieval* (pp. 273-285). Springer, Cham.
44. McCrae, R. R., & Costa Jr, P. T. (1995). Trait explanations in personality psychology. *European Journal of Personality*, 9(4), 231-252.
45. McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, 49(3), 710.
46. Medina-Ojeda, N., Martín-Santana, J. D., & Medina-Muñoz, D. R. (2019). The personality of the tourist as a segmentation criteria of sun and beach destinations: an application to the destination Gran Canaria. *Cuadernos de Turismo*, 44, 565-568.
47. Mesquita, R., & Lima, L. S. (2018, November). A Context-Awareness Approach to Tourism and Heritage Routes Generation. In *Ambient Intelligence—Software and Applications—, 9th International Symposium on Ambient Intelligence* (Vol. 806, p. 10). Springer.
48. Moghavvemi, S., Woosnam, K. M., Hamzah, A., & Hassani, A. (2021). Considering residents' personality and community factors in explaining satisfaction with tourism and support for tourism development. *Tourism Planning & Development*, 18(3), 267-293.
49. Mohammadali, Z. M., & Abdulkhaliq, S. S. (2019). Prospects and Challenges of Entrepreneurship Development in the Kurdistan Region of Iraq: An Overview. *International Journal of Entrepreneurial Knowledge*, 7 (2), 4-16. doi: 10.12345-0006.
50. Paulienė, R. (2012). Transforming leadership styles and knowledge sharing in a multicultural context. *Business, Management and Economics Engineering*, 10(1), 91-109.
51. Passafaro, P., Cini, F., Boi, L., D'Angelo, M., Heering, M. S., Luchetti, L., ... & Triolo, M. (2015). The "sustainable tourist": Values, attitudes, and personality traits. *Tourism and hospitality research*, 15(4), 225-239.
52. Petrikova, D., & Sorokova, T. (2016). Managerial and entrepreneurial skills as determinants of business. *Polish Journal of Management Studies*, 14.

53. Qin, S. (2021). Perceptual Enablers Of Global Nascent Entrepreneurs: A Quantitative Investigation Based On Global Entrepreneurship Monitor Data. *International Journal of Entrepreneurial Knowledge*, 9(1), 112-126.
54. Rojas-Mendez, J. I., & Hine, M. J. (2017). Countries' positioning on personality traits: Analysis of 10 South American national tourism websites. *Journal of Vacation Marketing*, 23(3), 233-247.
55. Rozsa, Z., Formánek, I., & Maňák, R. (2019). Determining the factors of the employees' intention to stay or leave in the slovak's SMEs. *International Journal of Entrepreneurial Knowledge*, 7 (2), 63-72.
56. Tapanainen, T., Dao, T. K., & Nguyen, T. T. H. (2021). Impacts of online word-of-mouth and personalities on intention to choose a destination. *Computers in Human Behavior*, 116, 106656.
57. Tran, X., Nguyen, B. L., & Nguyen, M. C. (2015). Effects of the big five personality traits on recreation Types-The case of Vietnam tourism.
58. Tsiakali, K. (2018). User-generated-content versus marketing-generated-content: Personality and content influence on traveler's behavior. *Journal of Hospitality Marketing & Management*, 27(8), 946-972.
59. Ullah, I., Elahi, N. S., Abid, G., & Butt, M. U. (2020). The impact of perceived organizational support and proactive personality on affective commitment: Mediating role of prosocial motivation. *Business, Management and Economics Engineering*, 18(2), 183-205.
60. Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, 25(4), 40-48.
61. Verma, V. K., Kumar, S., & Chandra, B. (2017). Big Five personality traits and tourist's intention to visit green hotels. *Indian Journal of Scientific Research*, 15(2), 79-87.
62. Wang, H., Lu, Y., & Zhai, C. (2010, July). Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 783-792).
63. Wang, X., Cheng, M., Wong, I. A., Teah, M., & Lee, S. (2021). Big-five personality traits in P2P accommodation platforms: similar or different to hotel brands?. *Current Issues in Tourism*, 24(23), 3407-3419.
64. Wroblowska, Z. (2019). Personality requirements for Ukrainian and Czech managers creating brands as a competitive advantage. *Journal of Competitiveness*, 11(1), 152-167.
65. Wroblowska, Z., & Ruda, T. (2015). The identification and comparison of the requirements placed on product managers during the recruitment process. *Journal of Competitiveness*, 7(3).
66. Wu, C., & Mursid, A. (2019). The relationship between personality, customer participation, customer value and customer satisfaction in tourism service. *European Journal of Tourism Research*, 23, 156-171.
67. Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3), 363-373.
68. Yoo, K. H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in human behavior*, 27(2), 609-621.
69. Yousaf, K., Abid, G., Butt, T. H., Ilyas, S., & Ahmed, S. (2019). Impact of ethical leadership and thriving at work on psychological well-being of employees: Mediating role of voice behaviour. *Business, Management and Economics Engineering*, 17(2), 194-217.

Brief description of Authors:

Federica Izzo

ORCID ID: <https://orcid.org/0000-0002-5136-1696>

Affiliation: Department of Education, psychology and communication, University "Suor Orsola Benincasa", Naples, Italy.

<https://www.unisob.na.it/universita/facolta/formazione/index.htm?vr=1&lg=en>

Email: federica.izzo@studenti.unisob.na.it

Mrs. Izzo is a Ph.D. scholar in the "Humanities and Technologies: an integrated research path" program at University "Suor Orsola Benincasa", Naples, Italy. She is a Subject Expert and Teaching Assistant in Web Marketing. Her research reflects marketing, communication, consumer behavior, business and management studies, data analysis, machine learning, and new technologies applied to qualitative research.

Quirino Picone, Ph.D.

ORCID ID: <https://orcid.org/0000-0002-4224-1836>

Affiliation: Department of Education, psychology and communication, University "Suor Orsola Benincasa", Naples, Italy.

<https://www.unisob.na.it/universita/facolta/formazione/index.htm?vr=1&lg=en>

Email: quirino.picone@docenti.unisob.na.it

Dr. Picone has a Ph.D. in "Humanities and Technologies: an integrated research path". He is an adjunct professor in Web Marketing, Web 2.0, Internet and New Media. He is also the academic coordinator of Master's courses on multimedia communication in the food and wine sector and e-commerce management. He manages all social media, communication, and web marketing activities for the University "Suor Orsola Benincasa". He is a speaker at Italian and international conferences on marketing and management.