Machine Learning in Tourism: A Brief Overview

Generation of Knowledge from Experience

Roman Egger

Learning Objectives
- Illustrate the intuition behind machine learning
- Explain the different ML paradigms
- Provide an overview of ML algorithms
- Discuss areas of application in the field of tourism
- Understand the limitations and challenges of ML

1 Introduction and Theoretical Foundations

Machine Learning (ML) is undoubtedly one of the most significant and far-reaching technological developments to currently shape our times (Jamal et al., 2018). These technologies can be found in all areas of our lives, providing us with information and knowledge derived from data, albeit in an inconspicuous way. They serve as the backbone of voice assistants, such as Siri, Cortana, Bixby, or Alexa, are the cornerstones of chatbots, support personalized marketing, and predict customer behavior. They optimize processes of all kinds, filter and classify spam from our emails, form the basis of fraud prevention, and also provide the foundation for plagiarism checks. Additionally, ML technologies are constantly being employed in the use of social media platforms without users even noticing or being aware of it. For example, ML is used when photos are uploaded to Facebook and faces are automatically recognized and suggested for tagging or when sentiments behind

R. Egger (✉)
Innovation and Management in Tourism, Salzburg University of Applied Sciences, Salzburg, Austria
e-mail: roman.egger@fh-salzburg.ac.at

R. Egger (ed.), *Applied Data Science in Tourism*, Tourism on the Verge,
https://doi.org/10.1007/978-3-030-88389-8_6

emojis are recognized on Instagram and auto-hashtags are suggested. ML provides the central infrastructure for Artificial Intelligence (AI) and often lays the foundation for data science projects, linking computer science and statistics in order to develop algorithms and statistical model-based theories (Althbiti & Ma, 2020) with the objective of high predictive performance and generalizability (Jordan & Mitchell, 2015).

Although the beginnings of machine learning date back to the 1950s, its potential has only recently become apparent as the amount of data available today has reached enormous proportions. ML is often mentioned in the context of "Big Data" (Kelm et al., 2020), where it is true that its influence increases in parallel with the amount of data available. At the same time, however, it must be mentioned—and this is particularly relevant for scientific research projects—that one does not always have to be in possession of several gigabytes of data in order to be able to successfully apply ML. As will be exemplified later on, even smaller datasets containing several hundreds or thousands of instances might be sufficient enough to use ML approaches to automatically identify patterns in complex data. The knowledge of these patterns can then help to predict future events and make complex decisions with confidence.

There have been numerous attempts to define the term *machine learning*, like the one from Arthur Samuel, who first coined the term "Machine Learning" in 1959, describing it as "a field of study that gives computers the ability to learn without being explicitly programmed" (the quote is often cited, but cannot be found in his papers. It can therefore be seen as a gist of Arthur Samuel's 1959 paper). Similarly, Akerkar (2019b) defines ML as "computational methods using the experience to improve the performance or to make accurate predictions. [...] It is the study of algorithms that learn from examples and experience instead of hardcoded rules" (p. 19). The term "experience" is used here to refer to existing databases and their properties (the training data), which are used to learn and train a model (Mohri et al., 2018). The aim is to identify patterns in the data that allow to either better describe the data, increase performance, or perform the most accurate possible prediction (Jamal et al., 2018). Let us first take a closer look at some general terms before delving into the different parts of ML.

Datasets are a set of examples that contain *features* to solve a problem. If we think of data in a tabular form, each row is an *instance,* and each column is a feature. Features are measurable pieces of data that are fed into a machine learning algorithm and help to understand the problem. The result is a model, which is to be understood as the trained representation of what the algorithm has learned. For example, a random forest algorithm can be trained with training data, and the output is a random forest model. New, unknown data can now be fed into the model in order to obtain predictions, classify the data, and much more, depending on the algorithm used. Thus, a predictive algorithm creates a predictive model that, when fed with new data, produces a prediction based on the data it was trained on (Kelm et al., 2020).

Artificial intelligence, big data, and machine learning are frequently mentioned in connection with the currently popular and (semi-)overused terms "smart tourism" and "smart destinations." According to Gretzel, Sigala, Xiang, and Koo (2015), what supposedly makes destinations, cities, and tourism "smart," generally speaking, includes the various information and communication technologies (ICTs) integrated into the physical infrastructure, smart experiences that attempt to optimize travel experiences through personalization, contextualization, and real-time analysis (Buhalis & Amaranggana, 2015) as well as the business ecosystem geared toward smartness. Smartness thus requires the processing of big data, available as transactional data, user-generated content, data provided by integrated devices and measured by sensors, etc. (Gretzel et al., 2015; Koo et al., 2016). All data types, i.e., image, audio, video, text, and metadata such as date and time values, geospatial data, tags, and more, are of great relevance. For detailed information on analytics in smart tourism design, it is recommended to read Xiang and Fesenmaier (2017).

In order to process data accordingly, recognize structures and patterns within a dataset, extract new informative features, perform far-reaching analyses and forecasts, and personalize recommendations, among other tasks, various machine learning approaches can be implemented. An analysis of multiple publications containing the search query "Tourism" AND "Machine Learning" in the article title, abstract, or keywords revealed 390 papers in Scopus and 216 papers in Web of Science. As Fig. 1 shows, increased ML use in tourism research can be observed, especially from

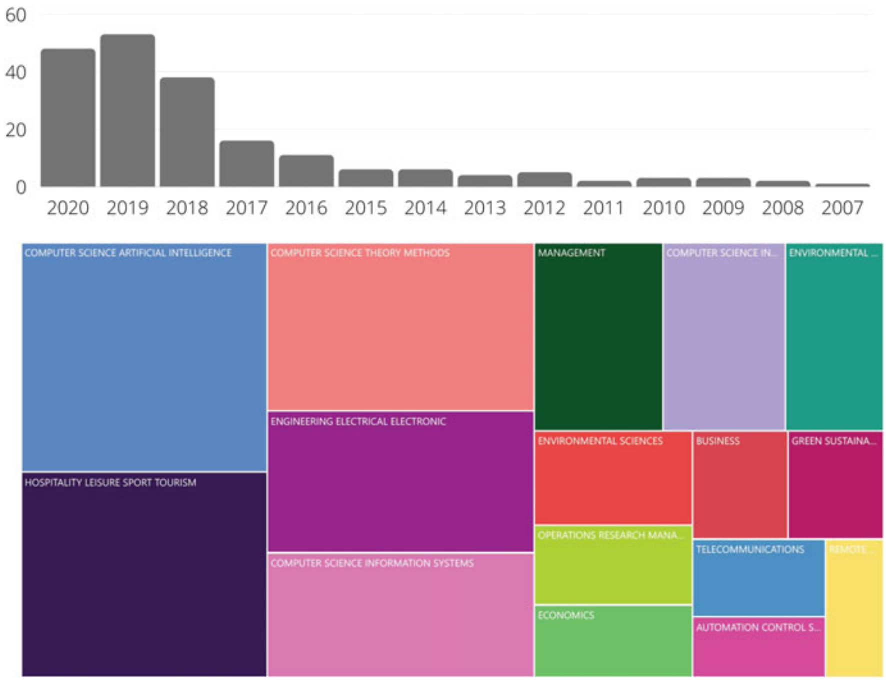

Fig. 1 ML-methods in tourism literature. Source: Author's depiction based on Scopus and Web of Science

2018 onwards. Moreover, the treemap depicts the subject areas in which papers have been published, with the field of "Computer Science Artificial Intelligence" still containing the most published articles, followed by the categories "Hospitality Leisure Sport Tourism" and "Computer Science Theory Methods."

Among tourism-specific journals, the most noteworthy (as of June 2021) is *Tourism Management* with 81 ML publications, followed by *Annals of Tourism Research* with 30 articles and *Tourism Management Perspectives* with 26 papers.

1.1 The Machine Learning Process

Although the step-by-step procedure for ML projects can vary depending on the chosen algorithm, the ideal-typical process looks more or less like the stages presented in Fig. 2, starting with the collection of training data and concluding with the interpretation of the data.

Since many people associate big data with machine learning, there is often an attempt to use all the available data. However, Awad and Khanna (2015) point out that this is sometimes counterproductive and unnecessary as it seems to be more efficient to select only a subset of the data (features) that is useful for solving the problem. This data can be either in structured or unstructured form and must be prepared accordingly before being processed via an algorithm (Egger et al., 2022). For example, the data must be presented in a uniform format, and incorrect and missing data must be removed. In addition, it may be necessary to normalize, discretize, average, smooth, etc. the data so as to be able to process it further (Awad & Khanna, 2015).

Feature engineering and feature selection appear next and are of particular importance in ML since the statement "garbage in—garbage out" holds especially true; in other words, good features are the backbone of any machine learning model. It is understandable that a model can only be as good as the data with which it was trained on, and that fatal errors can occur if a model trained on bad data has been used (Sanchez, 2003). Optimally, only those features that have an influence on the

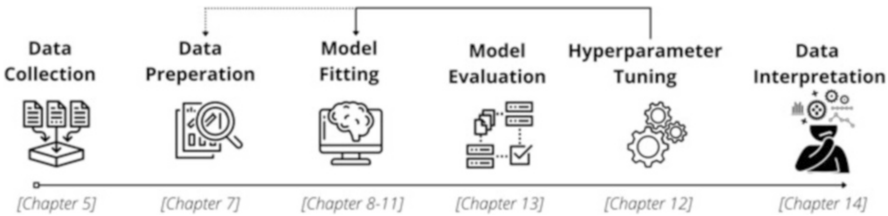

Fig. 2 ML process

quality of the model should be selected (Wennker, 2020). (See also chapter "Feature Engineering"). The subsequent step is to train the algorithm. For this purpose, the data is split into training and testing data, where the training data is used to train the algorithm, and the testing data is used to measure its performance. In contrary to supervised learning, unsupervised learning does not require the data to be split into training and testing and, thus, does not call for cross-validation either.

Once the model has been trained, it must be evaluated. As we will show below, this is not possible for unsupervised tasks in terms of calculated key metrics because there is no ground truth label. In supervised tasks, the effectiveness and performance of the algorithm can be evaluated, and hints for optimizing the data processing as well as changeable hyperparameters are obtained. Each ML system contains hyperparameters with settings that can be changed and, in turn, affect the algorithm's performance (Feurer & Hutter, 2019). Hyperparameter tuning is a sensitive process and requires comprehensive knowledge of the effects of such a change. Thus, to identify the settings that produce the best results, an iterative process between data preparation, model fitting, hyperparameter tuning, and model evaluation takes place. As a final step, the validated model should be applied to an actual task, for example, performing a prediction, and the results then need to be interpreted and put into its subject-specific context.

The "no free lunch theorem" states that, averaged across all optimization problems, each algorithm performs equally well when no resampling is performed (Adam et al., 2019). In other words, no algorithm works optimally for all tasks; each task has its own peculiarities and requires the correct choice of the appropriate algorithm (Egger, 2022). Therefore, numerous approaches have been developed to cater to specific tasks, and new types and forms of ML and their algorithms are constantly being developed and improved (Edwards, 2018).

There are three main types of ML algorithms, with the first, unsupervised learning algorithms, being covered in detail in chapters "Clustering" and "Dimensionality Reduction", and the second, supervised algorithms, being discussed in chapters "Classification" and "Regression". The third, reinforcement learning, has not been added as a separate chapter due to its comparatively low relevance for tourism cases (Jamal et al., 2018) (Fig. 3). Natural Language Processing (NLP), with algorithms for text classification, topic modeling, or sentiment analysis, is an additional, special ML case (Egger & Gokce, 2022) and will be discussed in detail in chapters "Natural Language Processing (NLP): An Introduction" to "Knowledge Graphs"; therefore, it will not be covered at this point in time.

ML approaches can be distinguished according to data type and availability of the dependent variable's label (Edwards, 2018). Thus, either continuous or discrete dependent variables are given, which, if they contain a label, can be processed with supervised algorithms, or, if no label is given, unsupervised methods are applied (Table 1).

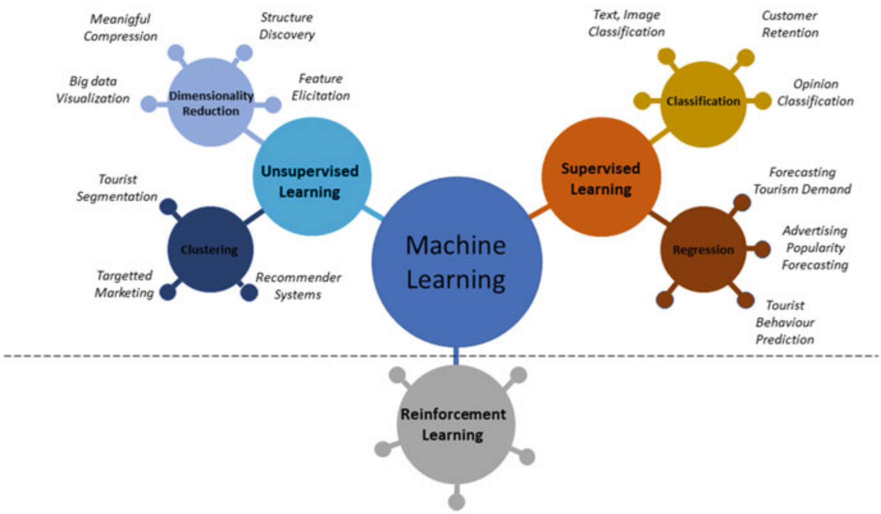

Fig. 3 Machine learning paradigms and application examples

Table 1 Learning problems and algorithms

	Supervised	Unsupervised
Discrete	Classification/categorization	Clustering
Continuous	Regression	Dimensionality reduction

Source: Skilton and Hovsepian (2018a)

In recent years, numerous ML paradigms have evolved (Fig. 4), and in addition to the supervised-, unsupervised-, and reinforcement-learning types highlighted in Fig. 3, model-based, memory-based, and deep-learning are also worth mentioning (Skilton & Hovsepian, 2018b). In particular, deep learning with neural networks can be viewed as a driver of the current ML renaissance, representing an entirely independent subfield (Choo et al., 2020). Deep learning approaches can scale the amount of data better and, subsequently, often deliver better results (Papp et al., 2019). However, it is a misconception to assume that neural networks are always superior to classical machine learning approaches (Choo et al., 2020). For a more detailed discussion on neural networks and deep learning, refer to further literature, e.g., Aggarwal (2018b) or Ekman (2021).

Fig. 4 Machine learning paradigms and algorithms. (An interactive version of this figure is available at: http://www.datascience-in-tourism.com/?p=132). Source: Dobilas (2021)

1.2 Unsupervised Learning

Unsupervised learning algorithms attempt to identify common elements (Mich, 2020) and recognize useful structures and patterns from input data without requiring the data to be labeled. Jordan and Mitchell (2015) mention two main problems for which unsupervised methods can be helpful: (1) when missing data leads to *data sparsity,* affecting the model's performance and accuracy, and (2) when data is presented in high-dimensional spaces, resulting in the *curse of dimensionality* phenomenon (Bernstein & Kuleshov, 2014). Overall, unsupervised algorithms take a set of predictors and analyze the relationships between them (Ozdemir, 2016). On the one hand, this can lead to identifying groups of observations that behave similarly due to their features, which corresponds to *clustering*, or, on the other hand, to grouping features together in order to achieve *dimensionality reduction*.

A big advantage of unsupervised learning is that the data do not require labels, which, in practice, makes it much easier to find suitable data material (Provost & Fawcett, 2013b). On the other hand, however, the predictive power is lost since only the response variable contains the information necessary for a prediction. Another major disadvantage is that it is difficult to measure how well the model works or to what extent it performs successfully (Dy & Brodley, 2004). Since there is no response variable, the model cannot be evaluated with respect to its performance, and the result ends in differences and similarities that require subjective human judgment (Ozdemir, 2016).

1.2.1 Clustering

A very common method of unsupervised learning is clustering, to which the aim is to identify distinct groups in the data. The data should be homogeneous within a group and show similar characteristics, and the individual groups should differ distinctly from one another, i.e., they should be heterogeneous. Thus, the goal, in this case, is not to make a prediction but, rather, to learn something about the structure and patterns inherent in the data (Arefieva et al., 2021) (Fig. 5).

Skilton and Hovsepian (2018a) list the following as information that a clustering algorithm tries to determine when searching for subsets: What is the number of subsets, and what is their size? Since there is no labeled data in unsupervised learning to specify the number of groups, this is often a nontrivial task. While there are indeed metrics to help the researcher determine the number of clusters, it ultimately boils down to a decision that must be made individually by evaluating the data. Furthermore, the question arises as to what common characteristics and

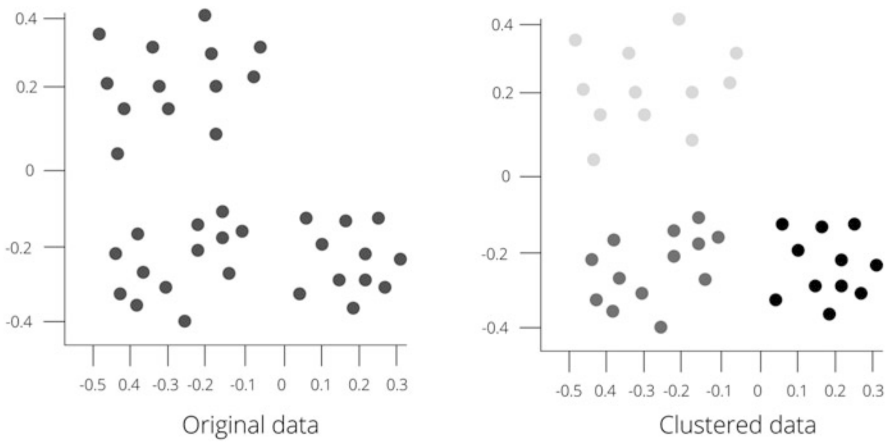

Fig. 5 Clustering

properties the members of a group have and whether these subsets themselves exhibit structures and patterns. Typical questions could be, for example, "can tourists be divided into natural groups in terms of perceived risk with regard to COVID-19? (Neuburger & Egger, 2021)" or "into which spatial units can a city be divided in terms of its tourist use?" As shown in Fig. 4, there are many different clustering algorithms available, each developed for specific problems, which should be chosen carefully based on the intended application.

A special form of clustering is known as *community detection*. These are graph-based approaches in which similarities are not identified based on the data's features but, instead, on the relationships between the data. In contrast to clustering methods that process tabular data, community detection (Ghosh et al., 2018) algorithms (such as Louvain, Leiden, or Markov clustering) use networks as a data source. They process a matrix of edges and nodes and detect commonalities based on membership features in networks (Fortunato & Hric, 2016). The question, therefore, is how to represent the relationships in a graph in a compact way. According to McAuley (2017), a community can be defined in cases where the members are connected to each other and where there are few edges between the communities, a high density inside and few corners outside, and a sense of "cliquishness."

Clustering methods have always played an essential role in the context of tourism, especially to typologize tourists and their behavior, but also to group photos, reviews, destinations and their characteristics, etc. into homogeneous groups. Up until now, hierarchical clustering and k-means methods have primarily been applied. Hierarchical clusterings have been used, for example, by Neuburger and Egger (2021) to group travelers according to their perceived COVID-19 risk, Derek, Woźniak, and Kulczyk (2019), who created a typology of outdoor tourists, Batista e Silva, Barranco, Proietti, Pigaiani, and Lavalle (2020), who developed a new systematic classification of EU regions based on the predominant location of hotels, or by Del Chiappa, Atzeni, and Ghasemi (2018) to analyze residents' perceptions and attitudes towards tourism development in Costa Smeralda, Italy. Less frequently, one may also come across studies that have used k-means clustering to investigate a tourism context. For instance, Srihadi, Hartoyo, Sukandar, and Soehadi (2016) segmented the Jakarta travel market by creating a typology of tourists based on their lifestyles, and Chua, Meng, Ryu, and Han (2021) used k-means clustering to group tourists' volunteering in terms of their life satisfaction and attitudes toward volunteering.

The use of density-based clustering algorithms is still quite rare in tourism, although these methods are particularly suitable for spatial analysis. Park, Xu, Jiang, Chen, and Huang (2020) studied spatial structures of tourism destinations by applying a trajectory data mining approach using mobile big data. The goal was to better understand the spatial structures and interactions of tourist attractions, for which they used density-based spatial clustering of applications with noise (DBSCAN). This algorithm was also applied by Ma, Kirilenko, and Stepchenkova (2020) to cluster Instagram photos of the solar eclipse. Community detection algorithms, such as Louvain or Leiden, have also hardly been used in tourism thus far. One example is Yu and Egger's (2021) study, in which they analyzed the relationship between colors and user engagement of Instagram images. The Louvain algorithm was used to categorize annotated images using Google Cloud Vision on the basis of their labels. In this case, the network-based algorithm was applied to cluster Instagram photos according to their interconnected nodes, where the labels were considered as nodes.

See chapter "Clustering" for a more explicit discussion on the topic of clustering and its different approaches.

1.2.2 Dimensionality Reduction

Oftentimes, high-dimensional data must be processed, and the so-called "curse of dimensionality," seen as an obstacle when using many methods, requires the complexity of the data to be reduced through the act of decreasing the dimensions. This can be achieved by identifying features that best represent the data, which, in turn, results in a smaller, more efficient dataset. Obtained in this way, the smaller size of the data is then easier to process and should help render significant information more recognizable (Bernstein & Kuleshov, 2014). It is important, however, that the loss of information associated with the reduction of data (in favor of gaining better insights into the data) remains within a reasonable range (Provost & Fawcett, 2013a).

The best-known way of reducing dimensions in data is principal component analysis (PCA). In recent years, however, algorithms that are better able to preserve the local and global structure of the data and are used to reduce high-dimensional data, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP), have emerged. In the context of tourism, usually the classical methods of dimension reduction are preferred; in most cases, PCA is used, while t-SNE and UMAP still rarely make their presence known. This is most likely due to the fact that most studies work with survey data, which hardly produce high-dimensional vectors unless they are generated during the process of feature engineering. These algorithms are also often used for the reduction of high-dimensional data in order to visualize them in 3D or 2D.

For example, Li, Li, Hu, Zhang, and Hu (2018) performed sentiment classification paired with a topic model approach using a bidirectional recurrent neural network. By applying lda2vec for topic identification, multidimensional word vector representations were created. In order to observe and visualize the process of sentiment classification in more detail, the data had to be reduced to a 2D projection, for which they used t-SNE as a dimension reduction algorithm. Another example is a study from Payntar, Hsiao, Covey, and Grauman (2021) in which high-dimensional features of geotagged internet photos were generated by applying ResNet50 and ImageNet, which are convolutional neural networks. In order to be able to map the photos onto a map for interpretation purposes, a visual embedding with t-SNE was performed afterward.

Chapter "Dimensionality Reduction" deals with the different methods of dimension reduction in more detail, with a special focus on "rarely used" methods such as t-SNE and UMAP.

1.3 Supervised Learning

In contrast to unsupervised learning, the target variable of a training dataset has a label, i.e., a more detailed description that can be used to train the algorithm. This makes supervised learning, especially suitable for prediction tasks. Depending on whether the target variable is continuous or discrete, it becomes either a regression or a classification problem (Fig. 6). In both cases, the attempt is made, by means of a trained prediction model, to predict an output variable y by approximating a function $f(x)$ (Hastie et al., 2009). In this sense, the labeled training data is used to synthesize the model function by trying to generalize the relationship between the input data and the output data (Awad & Khanna, 2015). The choice of features is of particular importance here because having too many features can confuse the learning algorithm (Jamal et al., 2018). The goal is to train the algorithm, with the help of a labeled training set, so well that it is able to predict the correct class labels for a new, unseen dataset as accurately as possible (Awad & Khanna, 2015). Thus, the quality of a model is highly influenced by the training data, especially when it comes to supervised learning approaches. In practice, one may often be confronted with the

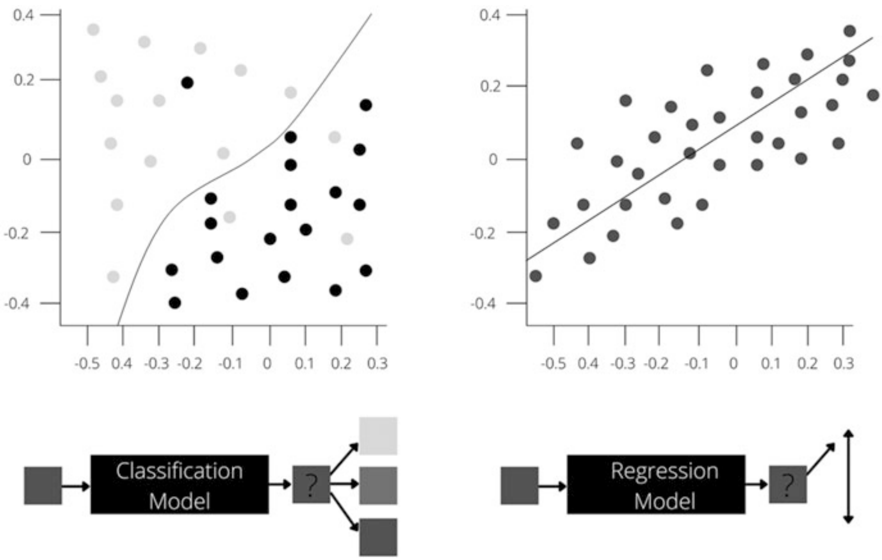

Fig. 6 Classification vs. Regression. Source: Adapted and expanded upon from Langs and Wazir (2019)

issue that such required labels are unavailable in the training dataset and have to be created first. This can be solved partially with unsupervised approaches (semi-supervised, self-supervised tasks) (Saeed et al., 2019), but if one wants to seriously ensure data quality, then one also has to rely on human labeling, which can quickly become a complex and expensive (time-consuming) task. On the other hand, one is then able to measure the quality of the results accurately, provided that the aspect of overfitting (see chapter "Model Evaluation"), where models fit too precisely to the training data at the expense of generalization, has been taken care of as well (Provost & Fawcett, 2013a).

Oftentimes in practice, it is not so straightforward or easy to decide whether a situation is a classification or a regression problem, especially when the features are taken from a typical Likert scale and, as in most cases, can be interpreted as both ordinal and interval scaled data.

1.3.1 Classification

Classification is one of the most commonly used ML applications, most likely due to the fact that a myriad of different use-cases containing classification problems exist. The overall goal of classification is to divide similar data points into different classes. The most commonly used approaches include decision trees, rule-based methods, probabilistic methods, support vector machines (SVM) methods, instance-based methods, and neural networks (Aggarwal, 2015). By using different classification algorithms, various rules, depending on which of these classifications have been applied, can be identified (Cleve & Lämmel, 2020). For example, customers can be divided into classes according to their creditworthiness. For this particular case, the basis for rule generation would involve a dataset consisting of customer data where, in addition to numerous other features, the characteristic of creditworthiness would be recognized as a label. The machine would then learn from the training data and find the rules and patterns that expel the most errorless classification of the new and previously unknown data. At first glance, classification and clustering methods seem to appear very similar as they both try to segment different groups based on characteristics. Yet, the vital difference is that, for classification, the structure of the groups is determined by the given labels, while in clustering, the segmentation is done on the basis of feature similarities, with the main objective being to reveal the hidden structures in the dataset (Arabie et al., 1996).

Apart from regression cases, classification tasks can be seen as the most widely used ML approach. In their study, Ramos-Henríquez, Gutiérrez-Taño, and Díaz-Armas (2021) aimed to operationalize the value proposition of hosts on Airbnb. From more than 250 variables, they first identified those that contribute most to being classified as a "superhost." The authors then used a SVM classifier for binary classification and naïve Bayes and Logit models as baseline models when comparing analyses. Moreover, Deng and Li (2018) presented an approach with regard to the correct selection of photos for destination image communication based on classifying them into affection categories. For this purpose, the authors trained a naïve Bayes model with numerous "content-emotion" pairs using Flickr images to predict an emotion classification for new photos based on their content. In another study, Martinez- Torres and Toral (2019) attempted to classify reviews using a text-based ML approach by examining the content of online reviews from the hospitality sector. The goal of their paper was to identify those features in order to successfully classify reviews into either deceptive or nondeceptive reviews. For this purpose, the authors determined the TF-IDF value from the words of the reviews and used them as input data for the classification process. Keeping the "no-free lunch theorem" in mind, they applied six different classifiers and compared their results. They trained k-NN, logistic regression, SVM, random forest, gradient boosting, and multilayer perceptron (MLP) classifiers for this task.

For more information, chapter "Classification" goes into detail on the topic of classification and describes, together with an example, the different approaches.

1.3.2 Regression

Regression is commonly known as a set of statistical tools that models the relationship between explanatory variables and a target variable, thus describing the average relationship between numerical attributes (Shalev-Shwartz & Ben-David, 2014). There are many different regression methods in machine learning, and, in contrast to classification, continuous numerical values, rather than discrete features in the form of classes, are predicted. If one divides the predicted numerical feature into a defined number of intervals, any regression algorithm can also be used as a classification one (Akerkar, 2019b; Althbiti & Ma, 2020). Thus, regression analyses can be used to solve both prediction problems and classification problems (Skilton & Hovsepian, 2018b). An extension of linear regression is logistic regression; here, the dependent variable has a categorical characteristic, typically with binary values of 0 and 1.

In tourism literature, regression is mostly used to explain one variable with the help of a dependent variable. In contrast to this frequently used explanatory function, regression in ML mainly involves prediction. An excellent overview of tourism forecasting research using internet data is provided by Li, Law, Xie, and Wang (2021), and chapter "Time Series Analysis" also deals with the topic of time series analysis in order to forecast tourism demand.

Prediction models based on regression approaches are used in a wide variety of areas. For example, in terms of sustainability, models for predicting the impact of tourism on nature are highly relevant but widely missing. In this sense, Jahani, Goshtasb, and Saffariha (2020) used different regression models to investigate which ecological factors are associated with vegetation density regeneration and how these can be predicted. For this purpose, they compared SVM, radial basis, and multilayer perceptron models. Similarly, Han (2020) also investigated the relationship between local tourism development and its impact on environmental resources so as to develop a quantitative SVM prediction model for turnover development in Chengdu, China.

Some additional tourism examples in which different regression approaches have been used come from, for example, Guerreiro and Rita (2020), who explored the connotation of explicit recommendations based on Yelp reviews by applying logistic regression, as well as Martin-Fuentes, Fernandez, Mateu, and Marine-Roig (2018), who strengthened the efficacy of using SVM for multiclass classification. By taking properties listed on Airbnb as the case context, the accuracy of SVM proved to be higher than the logistic model in general. Moreover, to group a large number of travel blog entries, Shibata, Shinoda, Nanba, Ishino, and Takezawa (2020) applied ensemble learning using SVM with RBF kernel.

Chapter "Regression" provides further information on various regression approaches.

1.4 Reinforcement Learning

The third approach, which will not be discussed further in this book, is reinforcement learning, a form of automated, goal-directed learning and decision making (Akerkar, 2019a). Here, the focus is on the characteristics of the learning problem and not on the learning algorithm itself. In this way, any algorithm can be selected to solve the problem as long as it is suitable for solving that particular problem (Skilton & Hovsepian, 2018a). The system is given a task, and it is supposed to learn and evolve on its own based on positive or negative feedback in an attempt to maximize a numerical reward value (Ngyuen & Zeigermann, 2021).

One of the very scarce examples of applying reinforcement learning in a tourism context comes from Lu, Meng, Timmermans, and Zhang (2021). Using a deep reinforcement learning model, random forest, and a multinomial logit model, the authors investigated the hesitation of passengers to choose a connecting airport when they have a large number of online sales options via several channels at their disposal. Their results yielded that the reinforcement learning model achieved the highest accuracy, followed by the random forest algorithm and that the multinomial logit model performed much worse than the alternative approaches.

1.5 Neural Networks

Artificial neural networks (ANNs) take on a special role in ML in part because, while other algorithms may have each been developed with a specific type of task in mind, ANNs are highly adaptable and can, theoretically, learn any mathematical function (Aggarwal, 2018a). Thus, almost all networks developed for ML are "Turing complete," meaning that they can simulate any learning algorithm, given they have enough training data (Pérez et al., 2021). Generally speaking, neural networks attempt to simulate the decision process of the neuron networks in the human brain (Graupe, 2013) so as to produce artificial intelligence. While other ML approaches use highly complex mathematical functions designed to solve a very specific problem, ANNs work with numerous, yet elementary, computational operations, possess self-organizing properties and solve complex, mathematically ambiguous, and nonlinear or stochastic problems (Graupe, 2013). Additionally, ANNs are characterized by high parallelism, a distributed memory, and adaptability (Adeli & Hung, 1994). Compared to other ML approaches, this renders ANNs particularly robust and tolerant to error and noise (Palmer et al., 2006).

There are a variety of different network architectures that differ mainly due to their net topologies and connection types (e.g., single-layer, multi-layer, feedforward, feedbackward networks). For instance, since feedforward neural networks only pass signals from input to output, there are no cycles or loops in this network. In contrast, recurrent neural networks (RNNs) have additional connections between neurons in the same layer and previous layers. RNNs can, therefore, not only learn from the input and weights but also from the previously learned so-called hidden states.

As shown in Fig. 7, the basic structure of an ANN provides a primary input layer in which neurons receive and process the weighted input signals, where the weights reflect the strength of the connection between neurons. Each neuron has an activation function whose output is computed by functions, such as hard limit, threshold logic, and sigmoid (Ebrahimpor-Komleh & Afsharizadeh, 2015). Suppose a certain output value is reached, then the neuron "fires" and forwards the output to the next neuron. The higher the forwarded output value is, the more significant its input dimension was (Raschka & Mirjalili, 2018). The output dimensions are then combined in the next layer (hidden layer) with new further dimensions, but this step is hardly comprehensible and leads to black box models. The continuation of these processes results in a complex network with multiple connections. As mentioned before, ANNs are highly adaptable, and this adaptability threshold is reached by providing the network with feedback based on its output. As such, the ANN can

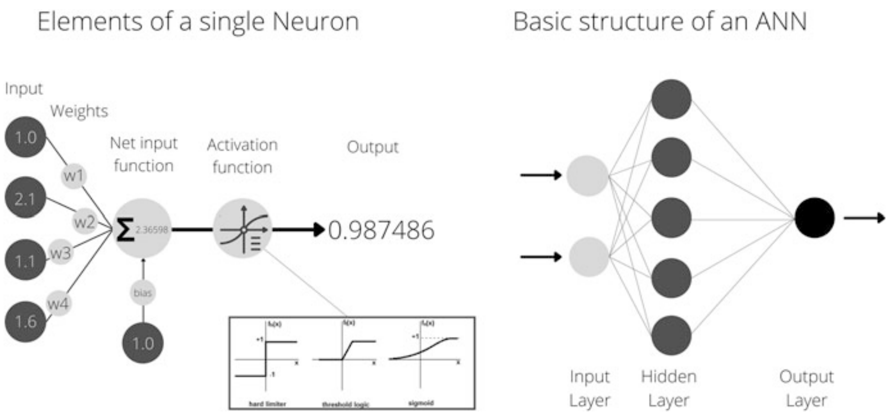

Fig. 7 Elements and architecture of a simple ANN. Source: Adapted and extended from Ebrahimpor-Komleh and Afsharizadeh (2015)

make better predictions with each next step as it updates and adjusts the weights of the connections. By repeating this “backpropagation” several times, with larger and larger amounts of data, the system learns based on the rules it creates itself (Moolayil, 2019).

More information on the role of the bias neuron, the different activation functions, the error measurement of the forward pass, and details regarding backpropagation are beyond the scope of this chapter. For the interested reader, Graupe (2013) or Aggarwal (2018b) are recommended at this stage.

As with other ML approaches, overfitting or underfitting can also occur in neural networks. In overfitting, the algorithm learns and memorizes the training data by heart, so to speak, and processes new, unknown data inadequately. To solve this issue, Wennker (2020) recommends reducing the complexity of a network by removing neurons. However, the problem here is that one cannot know in advance how many neurons need to be removed from the network. Mathematically, an L1 and L2 regularization can help. On the other hand, to counteract underfitting, new neurons and weights should be included.

Deep learning is a subfield of machine learning, and deep artificial neural network uses. The success of deep learning is based on the advancements of network architectures such as RNNs and CNNs (Aggarwal, 2018b). As the name suggests, deep learning uses multilayer neural networks in which more layers mean a larger parameter space to be used for the learning process (Ba & Caruana, 2013). In particular, deep learning is used in image recognition and natural language processing.

Neural networks are also becoming increasingly important in tourism since, as mentioned above, they can basically be applied to any task. Phillips, Zigan, Santos Silva, and Schegg (2015), for example, used neural network analysis to examine the determinants of hotel performance by investigating the relationships of user-generated online reviews, hotel characteristics, and Revpar. Additionally, Palmer et al. (2006) suggested an ANN for tourism time series forecasting, Bloom (2005) applied a neural network to segment the international tourism market in Cape Town, South Africa, and, similarly, Kim, Wei, and Ruys (2003) used an ANN to segment the West Australian senior tourist market. Furthermore, Tsaur, Chiu, and Huang (2002) analyzed the determinants of guest loyalty to international tourist hotels, and Abubakar, Namin, Harazneh, Arasli, and Tunç (2017) used a structural equation model and an ANN to examine the influence of favoritism/nepotism and supervisor incivility on employee cynicism and job disengagement as well as the moderating role of gender. These diverse examples show that neural networks offer a vast field for application and can be used extensively in tourism contexts in order to perform a wide variety of ML tasks.

1.6 Machine Learning Limitations and Challenges

Nonetheless, despite ML's advantages, its uses and applications also come with risks and challenges. Similar to classical statistics, there are numerous points that need to be taken into account throughout the duration of the ML process in order to achieve valid and satisfying results. This starts with the quantity and quality of the data. Most of the time, data is unstructured and messy, and appropriate preprocessing is required. In addition to the quality of the data, a corresponding quantity thereof is also necessary for numerous procedures. Depending on whether the process requires an unsupervised or supervised task, this can quickly become a significant challenge. In the case of unsupervised procedures, their lack of evaluation capabilities calls for high-quality data, whereas when it comes to supervised procedures, high-quality labeled data is often not available in sufficient quantities.

In most cases, features that seem particularly promising for use as input data must first be generated and then selected (see chapter "Feature Engineering"). Moreover, during the actual ML process, the appropriate ML model must be selected, and the correct hyperparameters have to be chosen (see chapter "Hyperparameter Tuning"), which requires a good understanding of the setting values and their potential impacts on the results. Another main problem is the overfitting or underfitting issue as well as the evaluation of models (see chapter "Model Evaluation"). Finally, and importantly, the results have to be interpreted correctly; since many ML algorithms have a black box characteristic, making interpretation difficult, this is yet an additional hurdle that must be overcome (see chapter "Interpretability of Machine Learning Models").

1.7 Auto-ML

To make the benefits of machine learning available to a wider range of users and shorten a human's working time during the data science process, companies such as Google, Amazon, Microsoft, and IBM are currently focusing heavily on developing auto-ML approaches. Another aim thereof is to attract larger groups of customers to their services. Although most of the steps in such a process are automated, a human being is still necessary to provide and define the required training data for the input. In most cases, only the dataset needs to be uploaded to the cloud, and the corresponding categories need to be labeled. The system then prepares the data accordingly, selects the right algorithm, and tunes the hyperparameters. Ultimately, the result is a REST endpoint for using predictions (Janakiram, 2018). All of this happens automatically "in the background," with the entirety of the decision-making being data driven and objective, and should, in turn, save the user both time and the necessary expertise (Gijsbers et al., 2019). In addition, Auto-ML approaches are mostly offered as a cloud solution, guaranteeing sufficient memory and computing power. Hutter, Kotthoff, and Vanschoren (2019) summarize this by saying, "this can be seen as a democratization of machine learning: with AutoML, customized state-of-the-art machine learning is at everyone's fingertips" (p.ix).

In recent years, numerous Auto-ML solutions have been launched on the market, with certain systems being named as leading solutions. For instance, H20 AutoML was founded as a platform for nonexperts to experiment with ML (H20, 2017), and Amazon offers a comprehensive package with AWS Sagemaker Autopilot (Das et al., 2020). Other similar solutions include Microsoft's Azure AutoML, Google Cloud AutoML, and IBM's AutoAI. Auto-WEKA (Hutter et al., 2019; Thornton et al., 2013), Auto-sklearn (Feurer et al., 2019), MLBox Auto-ML (Romblay, 2017), and TPOT (Olson & Moore, 2019) also exist as Auto-ML solutions but can only be used locally with your own hardware and software environment since they are not offered as cloud solutions.

For ML projects, Python, with its countless available modules, is the typical go-to software, yet R or Julia (Bezanson et al., 2017) can also be considered alternatives. For those who have no prior experience with scripting or programming languages, numerous visual computing solutions, such as the Konstanz Information Miner (KNIME), Rapidminer, or Orange3, are available. With these applications, analysis pipelines are assembled in the sense of workflows via drag and drop, and these solutions have a molar structure and are continuously expanded upon with new components. In KNIME, for example, almost all WEKA methods are available (Cleve & Lämmel, 2020). For more information on software solutions, see chapter "Software and Tools".

Further Readings & Other Sources

Dulhare, U. N., Ahmad, K., & Ahmad, K. A. B. (Eds.). (2020). *Machine learning and big data: Concepts, algorithms, tools and applications*. Wiley.

Google's Machine Learning Crash Course. https://developers.google.com/machine-learning/crash-course

Ng, A. (2021). *Machine Learning—Coursea Course with more than 4 million participants*. https://tinyurl.com/andrewng-ml

Nwanganga, F., & Chapple, M. (2020). *Practical machine learning in R*. Wiley.

Raschka, S., & Mirjalili, V. (2017). *Python machine learning: Machine learning and deep learning with python* (2nd ed.). Scikit-Learn, and TensorFlow.

Theobald, O. (2021). *Machine learning for absolute beginners: A plain English introduction* (3rd ed.). Scatterplot Press.

Venturi, D. (2019). *Provides an extensive list of machine learning courses provided on the Internet*. https://www.freecodecamp.org/news/every-single-machine-learning-course-on-the-internet-ranked-by-your-reviews-3c4a7b8026c0/

References

Abubakar, A. M., Namin, B. H., Harazneh, I., Arasli, H., & Tunç, T. (2017). Does gender moderates the relationship between favoritism/nepotism, supervisor incivility, cynicism and workplace withdrawal: A neural network and SEM approach. *Tourism Management Perspectives, 23*, 129–139. https://doi.org/10.1016/j.tmp.2017.06.001

Adam, S. P., Alexandropoulos, S.-A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No free lunch theorem: A review. In I. C. Demetriou & P. M. Pardalos (Eds.), *Springer optimization and its applications. Approximation and optimization* (Vol. 145, pp. 57–82). Springer International. https://doi.org/10.1007/978-3-030-12767-1_5

Adeli, H., & Hung, S.-L. (1994). *Machine learning: Neural networks, genetic algorithms, and fuzzy systems*. 97804710. Retrieved from https://scholar.nctu.edu.tw/zh/publications/machine-learning-neural-networks-genetic-algorithms-and-fuzzy-sys

Aggarwal, C. C. (2015). *Data classification: Algorithms and applications*. Chapman & Hall/CRC data mining and knowledge discovery series. CRC Press. Retrieved from http://proquest.tech.safaribooksonline.de/9781466586741

Aggarwal, C. C. (2018a). *Neural networks and deep learning*. Springer International. https://doi.org/10.1007/978-3-319-94463-0

Aggarwal, C. C. (2018b). *Neural networks and deep learning: A textbook*. Springer International. Retrieved from https://books.google.at/books?id=achqDwAAQBAJ. https://doi.org/10.1007/978-3-319-94463-0

Akerkar, R. (Ed.). (2019a). *SpringerBriefs in business. Artificial Intelligence for Business*. Springer International. https://doi.org/10.1007/978-3-319-97436-1

Akerkar, R. (2019b). Machine learning. In R. Akerkar (Ed.), *SpringerBriefs in business. Artificial intelligence for business* (pp. 19–32). Springer International. https://doi.org/10.1007/978-3-319-97436-1_2

Althbiti, A., & Ma, X. (2020). Machine learning. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of big data* (pp. 1–5). Springer International. https://doi.org/10.1007/978-3-319-32001-4_539-1

Arabie, P., Hubert, L., & de Soete, G. (1996). *Clustering and classification*. World Scientific.

Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. *Tourism Management, 85*, 104318. https://doi.org/10.1016/j.tourman.2021.104318

Awad, M., & Khanna, R. (Eds.). (2015). *Books for professionals by professionals. Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Apress Springer Science+Business Media. https://doi.org/10.1007/978-1-4302-5990-9

Ba, L. J., & Caruana, R. (2013, December 21). *Do deep nets really need to be deep?* Retrieved from http://arxiv.org/pdf/1312.6184v7

Batista e Silva, F., Barranco, R., Proietti, P., Pigaiani, C., & Lavalle, C. (2020). A new European regional tourism typology based on hotel location patterns and geographical criteria. *Annals of Tourism Research*, 103077. https://doi.org/10.1016/j.annals.2020.103077

Bernstein, A., & Kuleshov, A. (2014). Dimensionality reduction in statistical learning. In *ICMLA '14: Proceedings of the 2014 13th international conference on machine learning and applications* (pp. 330–335). IEEE Computer Society. https://doi.org/10.1109/ICMLA.2014.59

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review, 59*(1), 65–98. https://doi.org/10.1137/141000671

Bloom, J. Z. (2005). Market segmentation. *Annals of Tourism Research, 32*(1), 93–111. https://doi.org/10.1016/j.annals.2004.05.001

Buhalis, D., & Amaranggana, A. (2015). Smart tourism destinations enhancing tourism experience through personalisation of services. In *Information and communication technologies in tourism 2015* (pp. 377–389). Springer. https://doi.org/10.1007/978-3-319-14343-9_28

Choo, K., Greplova, E., Fischer, M. H., & Neupert, T. (2020). *Machine learning kompakt*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32268-7

Chua, B.-L., Meng, B., Ryu, H. B., & Han, H. (2021). Participate in volunteer tourism again? Effect of volunteering value on temporal re-participation intention. *Journal of Hospitality and Tourism Management, 46*, 193–204. https://doi.org/10.1016/j.jhtm.2020.12.003

Cleve, J., & Lämmel, U. (2020). *Data mining* (3. Aufl.). De Gruyter Studium. De Gruyter Oldenbourg. Retrieved from https://books.google.at/books?id=T0YCEAAAQBAJ

Das, P., Ivkin, N., Bansal, T., Rouesnel, L., Gautier, P., Karnin, Z., et al. (2020). Amazon SageMaker autopilot. In *Proceedings of the fourth international workshop on data management for end-to-end machine learning* (pp. 1–7). https://doi.org/10.1145/3399579.3399870

de Romblay, A. A. (2017). *MLBox's official documentation.* Retrieved from https://mlbox.readthedocs.io/en/latest/

Del Chiappa, G., Atzeni, M., & Ghasemi, V. (2018). Community-based collaborative tourism planning in islands: A cluster analysis in the context of Costa Smeralda. *Journal of Destination Marketing and Management, 8*, 41–48. https://doi.org/10.1016/j.jdmm.2016.10.005

Deng, N., & Li, X. (2018). Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. *Tourism Management, 65*, 267–278. https://doi.org/10.1016/j.tourman.2017.09.010

Derek, M., Woźniak, E., & Kulczyk, S. (2019). Clustering nature-based tourists by activity. Social, economic and spatial dimensions. *Tourism Management, 75*, 509–521. https://doi.org/10.1016/j.tourman.2019.06.014

Dobilas, S. (2021, July 2). Random Forest models: Why are they better than single decision trees? *Towards Data Science*. Retrieved from https://towardsdatascience.com/random-forest-models-why-are-they-better-than-single-decision-trees-70494c29ccd1

Dy, J., & Brodley, C. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research, 5*, 845–889. https://doi.org/10.1007/SpringerReference_302701

Ebrahimpor-Komleh, H., & Afsharizadeh, M. (2015). Improving prediction based digital image reversible watermarking by neural networks. In *2015 international congress on technology, communication and knowledge (ICTCK)* (pp. 201–208). IEEE. https://doi.org/10.1109/ICTCK.2015.7582671.

Edwards, G. (2018, November 18). Machine learning | An introduction—Towards data science. *Towards Data Science*. Retrieved from https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0#6246

Egger, R. (2022). Introduction: Data science in tourism. In R. Egger (Ed.), *Tourism on the verge. Applied data science in tourism* (pp. ix–xxiv). Springer.

Egger, R., & Gokce, E. (2022). Natural language processing (NLP): An introduction. In R. Egger (Ed.), *Tourism on the verge. Applied data science in tourism* (pp. 307–334). Springer.

Egger, R., Kroner, M., & Stöckl, A. (2022). Web scraping. In R. Egger (Ed.), *Tourism on the verge. Applied data science in tourism* (pp. 67–84). Springer.

Ekman, M. (2021). *Learning deep learning: Theory and practice of neural networks, computer vision, Nlp, and transformers using Tensorflow*. [S.l.]: ADDISON-WESLEY. Retrieved from https://books.google.at/books?id=W8dFzgEACAAJ

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning* (pp. 3–34). Springer International.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., & Hutter, F. (2019). Auto-sklearn: Efficient and robust automated machine learning. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning* (pp. 113–134). Springer International.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports, 659*, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Ghosh, S., Halappanavar, M., Tumeo, A., Kalyanaraman, A., Lu, H., Chavarria-Miranda, D., et al. (2018). Distributed Louvain algorithm for graph community detection. In *2018 IEEE International Parallel 2018* (pp. 885–895). https://doi.org/10.1109/IPDPS.2018.00098

Gijsbers, P., LeDell, E., Thomas, J., Poirier, S., Bischl, B., & Vanschoren, J. (2019, July 1). *An open source AutoML benchmark*. Retrieved from http://arxiv.org/pdf/1907.00909v1

Graupe, D. (2013). *Principles of artificial neural networks. Advanced series in Circuits & Systems: v.7*. World Scientific. Retrieved from https://ebookcentral.proquest.com/lib/subhh/detail.action?docID=1336559

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets, 25*(3), 179–188. https://doi.org/10.1007/s12525-015-0196-8

Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management, 43*, 269–272. https://doi.org/10.1016/j.jhtm.2019.07.001

H20. (2017). *AutoML: Automatic machine learning—H2O 3.32.1.3 documentation*. Retrieved from http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html

Han, J. (2020). Application of SVM model to environmental resource analysis in tourism development. *Journal of Physics: Conference Series, 1629*, 12007. https://doi.org/10.1088/1742-6596/1629/1/012007

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Springer eBook collection mathematics and statistics) (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated machine learning*. Springer International. https://doi.org/10.1007/978-3-030-05318-5

Jahani, A., Goshtasb, H., & Saffariha, M. (2020). Tourism impact assessment modeling of vegetation density for protected areas using data mining techniques. *Land Degradation and Development, 31*(12), 1502–1519. https://doi.org/10.1002/ldr.3549

Jamal, S., Goyal, S., Grover, A., & Shanker, A. (2018). Machine learning: What, why, and how? In A. Shanker (Ed.), *Bioinformatics: Sequences, structures, phylogeny* (pp. 359–374). Springer Singapore. https://doi.org/10.1007/978-981-13-1562-6_16

Janakiram, M. (2018, April 15). *Why AutoML is set to become the future of artificial intelligence*. *Forbes*. Retrieved from https://www.forbes.com/sites/janakirammsv/2018/04/15/why-automl-is-set-to-become-the-future-of-artificial-intelligence/?sh=bcfd730780ae

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science (New York, N.Y.), 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kelm, O., Gerl, K., & Meißner, F. (2020). Machine learning. In I. Borucki, K. Kleinen-von Königslöw, S. Marschall, & T. Zerback (Eds.), *Handbuch Politische Kommunikation* (pp. 1–9). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-26242-6_55-1

Kim, J., Wei, S., & Ruys, H. (2003). Segmenting the market of West Australian senior tourists using an artificial neural network. *Tourism Management, 24*(1), 25–34. https://doi.org/10.1016/S0261-5177(02)00050-X

Koo, C., Shin, S., Gretzel, U., Hunter, W. C., & Chung, N. (2016). Conceptualization of smart tourism destination competitiveness. *Asia Pacific Journal of Information Systems, 26*(4), 561–576. https://doi.org/10.14329/apjis.2016.26.4.561

Langs, G., & Wazir, R. (2019). Machine learning. In S. Papp, W. Weidinger, M. Meir-Huber, B. Ortner, G. Langs, & R. Wazir (Eds.), *Handbuch Data Science: Mit Datenanalyse und Machine Learning Wert aus Daten generieren* (pp. 178–198). Hanser.

Li, Q., Li, S., Hu, J., Zhang, S., & Hu, J. (2018). Tourism review sentiment classification using a bidirectional recurrent neural network with an attention mechanism and topic-enriched word vectors. *Sustainability, 10*(9), 3313. https://doi.org/10.3390/su10093313

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management, 83*, 104245. https://doi.org/10.1016/j.tourman.2020.104245

Lu, J., Meng, Y., Timmermans, H., & Zhang, A. (2021). Modeling hesitancy in airport choice: A comparison of discrete choice and machine learning methods. *Transportation Research Part A: Policy and Practice, 147*, 230–250. https://doi.org/10.1016/j.tra.2021.03.006

Ma, S., Kirilenko, A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse. *Tourism Management, 77*, 104021. https://doi.org/10.1016/j.tourman.2019.104021

Martinez-Torres, M. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management, 75*, 393–403. https://doi.org/10.1016/j.tourman.2019.06.003

Martin-Fuentes, E., Fernandez, C., Mateu, C., & Marine-Roig, E. (2018). Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management, 69*, 75–83. https://doi.org/10.1016/j.ijhm.2017.10.016

McAuley, J. (2017). *CSE 158: Web mining and recommender systems*. Retrieved from https://cseweb.ucsd.edu/classes/wi17/cse158-a/

Mich, L. (2020). Artificial intelligence and machine learning. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-tourism* (pp. 1–21). Springer International. https://doi.org/10.1007/978-3-030-05324-6_25-1

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. In *Adaptive computation and machine learning. Cambridge Massachusetts* (2nd ed.). The MIT Press.

Moolayil, J. (2019). *Learn Keras for deep neural networks*. Apress. https://doi.org/10.1007/978-1-4842-4240-7

Neuburger, L., & Egger, R. (2021). Travel risk perception and travel behaviour during the COVID-19 pandemic 2020: A case study of the DACH region. *Current Issues in Tourism, 24*(7), 1003–1016. https://doi.org/10.1080/13683500.2020.1803807

Ngyuen, C. N., & Zeigermann, O. (2021). *Machine learning: Kurz & gut* (2. Auflage). dpunkt. verlag, O'Reilly, Preselect.media Gmbh.

Olson, R. S., & Moore, J. H. (2019). TPOT: A tree-based pipeline optimization tool for automating machine learning. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *The springer series on challenges in machine learning. Automated machine learning* (pp. 151–160). Springer International. https://doi.org/10.1007/978-3-030-05318-5_8

Ozdemir, S. (2016). *Principles of data science: Learn the techniques and math you need to start making sense of your data*. Packt. Retrieved from http://proquest.tech.safaribooksonline.de/9781785887918

Palmer, A., José Montaño, J., & Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781–790. https://doi.org/10.1016/j.tourman.2005.05.006

Papp, S., Weidinger, W., Meir-Huber, M., Ortner, B., Langs, G., & Wazir, R. (Eds.). (2019). *Handbuch Data Science: Mit Datenanalyse und Machine Learning Wert aus Daten generieren*. Hanser. https://doi.org/10.3139/9783446459755

Park, S., Xu, Y., Jiang, L., Chen, Z., & Huang, S. (2020). Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Annals of Tourism Research, 84*, 102973. https://doi.org/10.1016/j.annals.2020.102973

Payntar, N. D., Hsiao, W.-L., Covey, R. A., & Grauman, K. (2021). Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. *Tourism Management, 82*, 104165. https://doi.org/10.1016/j.tourman.2020.104165

Pérez, J., Barceló, P., & Marinkovic, J. (2021). Attention is Turing-complete. *Journal of Machine Learning Research, 22*, 1–35. Retrieved from https://www.jmlr.org/papers/volume22/20-302/20-302.pdf

Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management, 50*, 130–141. https://doi.org/10.1016/j.tourman.2015.01.028

Provost, F., & Fawcett, T. (2013a). *Data science for business: What you need to know about data mining and data-analytic thinking* (1st ed.). O'Reilly Media. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=619895

Provost, F., & Fawcett, T. (2013b). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly.

Ramos-Henríquez, J. M., Gutiérrez-Taño, D., & Díaz-Armas, R. J. (2021). Value proposition operationalization in peer-to-peer platforms using machine learning. *Tourism Management, 84*, 104288. https://doi.org/10.1016/j.tourman.2021.104288

Raschka, S., & Mirjalili, V. (2018). *Machine Learning mit Python und Scikit-Learn und TensorFlow: Das umfassende Praxis-Handbuch für Data Science, Deep Learning und Predictive Analytics* (K. Lorenzen, Trans.) (2., aktualisierte und erweiterte Auflage). mitp.

Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(2), 1–30. https://doi.org/10.1145/3328932

Sanchez, J. S. (2003). Analysis of new techniques to obtain quality training sets. *Pattern Recognition Letters, 24*(7), 1015–1022. https://doi.org/10.1016/S0167-8655(02)00225-8

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.

Shibata, N., Shinoda, H., Nanba, H., Ishino, A., & Takezawa, T. (2020). Classification and visualization of travel blog entries based on types of tourism. In J. Neidhardt & W. Wörndl (Eds.), *Information and communication Technologies in Tourism 2020* (pp. 27–37). Springer International. https://doi.org/10.1007/978-3-030-36737-4_3

Skilton, M., & Hovsepian, F. (2018a). Machine learning. In M. Skilton & F. Hovsepian (Eds.), *The 4th industrial revolution* (pp. 121–157). Springer International. https://doi.org/10.1007/978-3-319-62479-2_5

Skilton, M., & Hovsepian, F. (Eds.). (2018b). *The 4th industrial revolution*. Springer International. https://doi.org/10.1007/978-3-319-62479-2

Srihadi, T. F., Hartoyo, Sukandar, D., & Soehadi, A. W. (2016). Segmentation of the tourism market for Jakarta: Classification of foreign visitors' lifestyle typologies. *Tourism Management Perspectives, 19*, 32–39. https://doi.org/10.1016/j.tmp.2016.03.005

Thornton, C., Hutter, F., Hoos, H., & Leyton-Brown, K. (2013). Auto-WEKA: Combined selection and Hyperparameter optimization of classification algorithms. In I. S. Dhillon & R. L. Grossman (Eds.), *Kdd 2013: The 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 11–14, 2013, Chicago, Illinois, USA ; [including] the Industry Practice Expo (IPE) [and] co-located workshops* (pp. 847–855). ACM.

Tsaur, S.-H., Chiu, Y.-C., & Huang, C.-H. (2002). Determinants of guest loyalty to international tourist hotels—A neural network approach. *Tourism Management, 23*(4), 397–405. https://doi.org/10.1016/S0261-5177(01)00097-8

Wennker, P. (2020). Machine learning. In P. Wennker (Ed.), *Künstliche Intelligenz in der Praxis* (pp. 9–37). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-30480-5_2

Xiang, Z., & Fesenmaier, D. R. (2017). *Analytics in smart tourism design*. Springer International. https://doi.org/10.1007/978-3-319-44263-1

Yu, J., & Egger, R. (2021). Color and engagement in touristic Instagram pictures: A machine learning approach. *Annals of Tourism Research*, 103204. https://doi.org/10.1016/j.annals.2021.103204