The Algorithm Applications in the Sharing Economy:

Building a regression model of Airbnb prices using python

KHELFAOUI Mounia[1,*] , SEDKAOUI Soraya[2],

[12] Djilali Bounaama University, Khemis-Miliana, (Algeria)

تطبيقات الخوارزمية في الاقتصاد التشاركي:

بناء نموذج انحدار لأسعار Airbnb باستخدام برنامج python

خلفاوي منية[1*] ، صدقاوي صورية[2]

2.1 جامعة الجيلالي بونعامة، خميس مليانة – الجزائر

Received: 04/02/2020; **Accepted:** 05/01/2021; **published:** 06/06/2021

Summary:

The purpose of this study is to explore the relationship between algorithmic applications and the sharing economy. To explore this relationship we applied a regression algorithm using Python for predicting Airbnb listing price. The results of the study show that Airbnb apartment prices in the city of Paris depend widely on the amenities and location. In fact, variables leading to price increase are related to these two factors. In order to see whether these factors constitute the main predictors in Airbnb apartments' prices at Paris, we can use other algorithms, such as random forest", and decision tree.

Keywords: Algorithms, Sharing Economy, Modeling, Linear Regression, Airbnb

Jel Classification Codes: C81 ; Q01

ملخص:

تهدف هذه الدراسة الى استكشاف العلاقة بين التطبيقات الخوارزمية والاقتصاد التشاركي. لاستكشاف هذه العلاقة ، طبقنا نموذج انحدار باستخدام Python للتنبؤ بأسعار Airbnb. تظهر نتائج الدراسة أن أسعار Airbnb في مدينة باريس مرتبطة بشكل كبير بوسائل الراحة والموقع. في الواقع ، ترتبط المتغيرات التي تؤدي إلى زيادة الأسعار بهذين العاملين. من أجل معرفة ما إذا كانت هذه العوامل تشكل عوامل التنبؤ الرئيسية في أسعار Airbnb في باريس ، يمكننا استخدام خوارزميات أخرى ، مثل خوارزمية الغابة العشوائية ، وشجرة القرار.

الكلمات المفتاحية: ال خوارزميات، الاقتصاد التشاركي، النمذجة ، الانحدار الخطي، Airbnb

الترميز الاقتصادي(JEL): C81 ; Q01

* Corresponding author, e-mail: mouniakhelfaoui@yahoo.fr

I- Introduction :

For over 20 years, society's daily consumption behavior has changed. With a view to economic logic, people agree to coexist, to travel in the same car, work in the same sphere, and, without being aware of each other or part of the same family and be implicated in joint initiatives. They choose to work together to produce, distribute, and in the areas of production, consumption, and distribution. People thus bringing in a new economic strategy involving partnering and a sharing feeling: usually called: "collaborative economy" or "sharing economy". The expansion of this economy has already been guided by a range of determinants, like developing devices and diversity of technologies, global economic development, global economic crises, and ecological awareness of environmental damage. The efficiency of the resources used has deteriorated. The sharing economy refers to an innovative, commodity-based socio-economic mechanism between individuals (cars, houses, tools, etc.) and services (coaching, etc.). This could include profits-based transactions, that is to say, monetary exchanges, or giving, trading, or volunteering.

What really is the perception of this economic activity? This new business model, which is also known as "platform economy", is famous for its technical tool (the digital platform), and offers all these opportunities for communication. It is defined as trade between "peers with platforms who act as brokers among them" (Nicot 2017). Currently, more and more companies are dependent on intensified digital utilization that, besides the necessary trust, enables an easy link between demand and supply. Moreover, the algorithms are the technical basis of the platforms. Therefore, there are many algorithms in both categories that are using for Big Data analysis. The main algorithms are regression and classifications that are extremely important to data processing. These algorithms seem to be the Machine Learning algorithms that have mostly been used. Regression is also used in prediction. There are not just these two algorithms in the area of Big Data Analytic applications. Cluster method has been also provided; an algorithm can be used for platform data analysis. The above review of an application for Machine Learning algorithms (supervised and unsupervised) is being perceived mostly as an indication of what will happen as it is present across the sharing economy.

Thus, the main question of this study is to demonstrate the leading role of algorithms across digital platforms in the expansion of the sharing economy.

Many scientists note that traditional methods unable to analyze Big Data. To develop such a phenomenon, it is necessary to develop a number of newly integrated IT tools that allow the use of data through different via various real applications. So the big data capacity has been established on various algorithms in the programming process. Studies also showed that each algorithm has a particular function (Regression, Classification, Clusters). These algorithms can also be combined for other functions (Classification / Regression).

1. The sharing economy: A Literature review

The common actions are natural for us to respond to tangible comfort (products and services exchange), and thereby our self-satisfaction, to implement our humanity in the connections between each other. People realize from a very early age to share, and over time, a feeling of possession leads to a sense of sharing. Besides this, religious doctrine excites the desire to escape ego from most of us. Religious traditions demand support for fundamental, bases, and practices to be shared.

The principle of sharing in society has been, therefore, made clear through various ways: cooperatives, mutual associations, voluntary work, and more. In recent years, the spirit of sharing has been part of the development of digital technology. The latest technological devices are transforming the usage of Transport, accommodation, products, and other facilities in cities. The sharing economy and certain platform-based markets linking clients, companies, and society have disrupted almost all industries (Hodkinson, 2017, Sedkaoui and Khelfaoui, 2020).

To understand how the principle of “sharing” is moved towards sharing economy, let’s get a dive into their history, bases, and practices. The Latin parts agree etymologically speaking, “sharing” means “to make equal parts” and “to push”, and “to activate” Consequently, since ancient times, the concept exists, combining two significances: separating a part from a particular reason. It could be a sharing of the heritage for all hereditary as well as abilities shared to assign specific tasks.

Each word has kept its particularity as for the concept of “sharing economy”. Sharing always involves leaving one or more people with some of what we have, irrespective of how this action is performed. The two words combined make a difference.

The sharing economy, through IT advancements, has become a leading location of Solidarity Economics and experienced impressive growth over recent years. Software or, in more detail, platforms increase exponentially business exchanges and discussions among service and their suppliers.

The Oxford dictionary added the term “sharing economy” in 2015. Likewise, the concept contains relatively new scientific literature. Researchers have used different names to examine its various aspects.

Amongst the used term, we could indeed talk about the peer-to-peer economy, collaborative economy, collaborative production, and collaborative consumption, the access economy, commonality-based peer production, and mesh size, on-demand economy, platform economy, and more. Sharing economy concept in the literature, however, is the most common (Ranjbari and al., 2018).

For breaking from the industrial age with consumption behaviors, the sharing economy is compelled to consume too much. This really proceeded after Second World War, reaching its height in the so-called Glorious Thirty years. Consumer culture arose when Taylor’s theory was born and assembly lines were launched. This process has led to a wealth of products and lower prices that are now turned available for all. To this must be added the captivating power of advertising which handles the consumer at producers’ whim.

It is certain that the consumer society allowed people to live in inevitable comfort and to meet all wishes for goods and services. Moving cars, water, and brought sophisticated entertainment objects and many other devices into the home. It has nonetheless had adverse consequences, denounced in the 1960s especially by consumer society terminology’s inventor George Katona, who initiated consumer psychology.

The preference to this terminology was an implausible, collaborative society that has for several years been concerned with hyper-consumption. This growing concept is at the heart of “collaborative consumption”, interpreted as “the set of resource circulation systems that allow consumers to use and provide valuable resources temporarily or permanently, through direct interaction with another consumer or through a mediator” (Decrop, 2017).

Felson and Spaeth initiated in 1978 the conception of collaborative consumption long before the advent of mobile technology and social networks. The aim was multiple from “the concept that has now been attributed” (Ertz 2017).

In collegial consumption, the perceptual opinion including the value of the product and services, popularly known in developments of economic theory, was improved. The concept of “access/use, not purchase/ ownership, of goods, capital or services” bases collaborative consumption through combining their assets (Decrop, 2017). In an attempt to address challenges in social structure, collaborative consumption is associated with the economic model using old principles like gift and selflessness.

The technical tools, the digital platform that provides all of these opportunities and renders this new business model known. The economy of sharing is termed the “platform economy”. It is interpreted as trade between “peers with platforms, acting between them as brokers” (Nico, 2017). The sharing economy is a pluralistic principle, endorsed by socio-economic relationships, entrepreneurs, and environmental groups. There is still no conceptual consensus basis of the term. Advocates of such an economy, however, have deployed an assessment of theory to classify common practices: P2P and the free market system, the donations economies, and the service economy (Borel and al. 2015).

Platforms ensure the sharing economy's success as they have multiple processes. The platforms first furnish an algorithm to efficiently match labor and users. Secondly, technology lowers the amount of cost transaction since microtransactions can be facilitated by platforms. Thirdly, platforms offer services to reduce or manage market transaction risk, thus addressing market failures, i.e. incomplete or misleading information (Drahokoupil and Fabo 2016).

2. Big data and Machine Learning Algorithms to assist the sharing economy

The sharing company modified the way people who participated and clients connect and relate all across the sharing process. However, a number of serious cases measures that apply in various procedures lie behind this process.

Due to its changing nature, sharing practices are very hard to comprehend, because sharing, as we today know it, totally varies from some few years earlier, and tomorrow will keep changing anywhere without a doubt.

We’ve seen in recent years the application of highly advanced IT tools and methods that only large companies had previously available for use. This has made many ways of creating innovative ideas easier.

Many succeed because the value chains are disrupted and established players are seriously upset. Naturally, this is just the start, because it is a trend that is growing: Uber, BlaBlaCar, Airbnb, and more. Companies in 2019 have been organized in such a way that they did not work in 1999, but they did in turn.

This paradigm involves the development of new ways in which actors can interact and reassess and learn the fundamental principles of the traditional channels. The digital revolution, therefore, brought about a revolution in data that enabled companies to gather large sets of data. Therefore, Big Data is the defining feature of this trend (Sedkaoui & Khelfaoui 2019).

The data analyzes are used to develop practices and determine who is targeting all the economic sharing platforms. These platforms also analyze these data in order to create customized products and services. These data are more used today and we can mention several factors (Sedkaoui 2018a):

- Data storage costs are constantly decreasing;
- Computing power increase;
- The explosion of large quantities of data mostly unstructured in nature requiring various operational methodologies than the conventional methods of analysis.

The main areas that have characterized the digital ecosystem are big data and the sharing economy. Large data volume' analysis has made it possible for user applications to start sharing companies. The use of these users' platforms leads to increased data volumes.

It is largely dependent on the capability of businesses to analyze available data, generate value, and increase profitability. How would you do this, nevertheless? The answer is the algorithms for machine learning.

Algorithms of machine learning can learn of high levels of current data or observations (Sedkaoui 2018a). Given that these data come from various sources, their analysis may vary in degrees of difficulty. That is why the algorithms make it possible to machine to learn without being programmed (Samuel, 1959). Algorithms of machine learning are helpful during the data processing phase; not surprisingly, a wide range of simplification technologies utilization has been developed.

The various algorithms we may use to extract value for machine learning are now being examined. The two main categories of algorithms are:

- Supervised algorithms: Get findings and discover throughout much knowledgeable data by identifying patterns, like spam filters, for example;
- Unsupervised algorithms: the input data, for this instance, have no validation set of class, and the algorithm is responsible for revealing the structure of the data. For example, a new client of Amazon checking for a particular thing, for instance. The Amazon system has no data about this client. A customer groups that are reaching the same thing is associated with the client. Therefore, the algorithm will characterize the client on the basis of information similitudes.

It is important to note that, regardless of what type of algorithm we apply to huge data analysis, not all algorithms have the same objective. Algorithms are typically divided into two different types:

–The manner of their learning process ;

–The form of the issue to be examined and analyzed.

3. Algorithms and use of data in digital platforms

There are a lot of algorithms we would use for big data analysis in the two-mentioned categories. The most popular algorithms can be found in Table 1. Regression and classification are essential in the analysis of data and are the most frequently used algorithms for machine learning. Its applications vary. For instance, when looking for words in texts (persons, locations, etc.), we choose a classification, if we try to recognize people by voice or image recording.

Regression is however used to identify the number of customers with special resources that are to be spent in exchanging products or services, to help Uber drivers too, for example, predict which part vehicle will probably fail, preventing fraud in payments, and develop a robust rating system.

Both regression and classification methods can be combined and used to designate potential clients for Airbnb apartments or to search for nearby tourist destinations of a certain location. Those are very important techniques, but these two algorithms are not the only ones in the list of big data analytics applications.

Cluster analysis is also provided, an algorithm we might use to analyze data generated by platforms. The attitudes of users in the platform for the sharing economy can be analyzed through this type of analysis; for example, by combining socio-demographic elements with other sharing variables, groups (or clusters) can be found on the basis among those factors. Such an analysis can be applied to examples in which we attempt to evaluate users' intentions relying on shared behavior, for example, for peer-to-peer platforms.

The above detail of applications (supervised and unsupervised) algorithms for machine learning can really be viewed as an indication of what's going to happen, as it is present throughout the sharing economy.

Authors think that it is more important to deploy the sentiment analysis algorithm since it allows companies of sharing economy to optimize client recommendations systems efficiently with the production of graphics to assess their satisfaction.

Companies need tooling to gather, analyze and store and collect the data they generate every day to fully exploit the potential of big data. Therefore, various kinds of equipment, as well as technology, were established to simplify analyzing of data to produce a perfect setting over the use of varied algorithms. Hadoop, Spark, Python, Matlab, R, as well as additional functionalities and tools for large data management and analysis provide diverse application algorithms.

II– Methodology:

We will underline the role of prediction algorithms in the economic sharing context. In these two concepts, a relationship is a development of the sharing economy as a new economic model. To do so, we decided to build a model using the database of the company Airbnb by selecting a linear regression algorithm. That is, we plan to offer Airbnb customers to rent an apartment. To that end, the prices of the apartment should be defined through a linear regression approach that allows us to construct a model that explains this price according to certain variables.

First, we explain the identification of regression issues. The algorithm predicts a true worth of output in a regression problem. In other words, based on previous comments (inputs), regression predicts the value (Sedkaoui 2018a). Therefore, the main issue of regression is the construction of a model that uses data to predict which new data do not explain. The objective is to forecast potential values that are based on existing values.

There should be three factors that must be taken into consideration in the design of a regression model before the start of the modeling process (Sedkaoui, 2018b):

- Description: the first key stage is to describe the phenomenon modeled on the question to be replied to;
- Prediction: model could be applied for future action prediction. Prediction: For example, it can be used to identify potentially interested customers for a service provided (accommodations, etc.).

- Decision-making: features and tools of forecasting offer additional valuable decision-making information. The model's intelligence leads to improve company results activities and actions.

The model provides answers to future behavior questions and to previously unknown features of a phenomenon to identify particular profiles.

The model is mathematically presented as follows:

As the name indicates, the model of linear regression presupposes a linear connection with both independent variables and dependent. Mathematically such a linear relation is determined as follow:

$$y = \alpha + \sum_{i=1}^{n} \beta i \, xi + \varepsilon$$

Where:

- y: the dependent variable;
- x_i : the independent variables (for i=1, 2, 3,……,n);
- α: the value of y when x_i is equal to 0;
- β: the change in y on the basis of one unit in a change in x_i;

This model is built to predict (Y) from (X). The development of a linear model of regression includes an estimation of the values or parameters used for α and β. For this purpose, we may use many techniques, but the most commonly used method of estimating the parameters is OLS.

1. Building a predictive price model with linear regression algorithms: A case of Airbnb[1] b

Having discussed the regression method principle before, let us now consider how to build a model and use data. To apply the algorithm to the real data, we will be using a sample dataset. We aim to show how important algorithmic applications are in the context of the sharing economy. Therefore, we will use databases we've compiled in this economy from specific companies' sources.

Remember that the objective is clear: Airbnb apartment price modeling in Paris. Our price suggestion model is to be developed.

In December 2018 the database used in this particular situation was assembled in CSV[2] format. It includes 59.881 findings and 95 factors, all accessible on the Airbnb platform.

This database is analyzed in three stages to develop the model that enables us to learn a lot about its features:

- Data preparation stage, aimed at cleaning up the data and selecting the most important models;
- The analysis phase of exploratory data;
- Modeling by linear regression.

However, to transform our data into knowledge Python[3] is used before we perform these three phases. So we'll import the panda library, making loading of CSV files easy. Therefore, data preparation is the first stage of our analysis. Consider the database for a while before you take this step (Figure 2).

Now in this first phase of analysis, we are going to remove and smooth certain factors. The aim is to maintain those who can supply the modeling process with information. Because 95 variables are contained in the database file, we have taken who already appear irrelevant (e.g. id, name, host_id, country, etc.) and other columns with information on the host, the neighborhood, etc. Of course, the analysis will not use this type of variable.

All inputs managed to retain and produce the linear predictor feature, measured with a linear regression model structure. That's why we have carried out the initial examination of 39 separate factors, 59,881 of which are used to map out the estimated value in euros in an apartment for a night's accommodation.

Consequently, the nature of these variables is qualitative and quantitative. For example, the quantitative variables include the number of comments, specific features such as latitude/ longitude. Neighborhood, room_type, etc., are qualitative variables.

We can now continue data preparation because we know our variables. Using the Python description function, the variables with missing values can be identified which we outlined in Figure 2.

We notice the removal of all non-numeric factors. Because of the type of regression equation, independent variables (x_i) must be continuous. Categorical variables must be transformed into continuous variables.

Use the empty percent function to figure out which variables in the model are empty and show them graphically (Figure 3). This figure indicates that the variable square feet have no value (almost 100%).

Likewise, license, security_deposit, and cleaning_fee variables have missing values over 30%. Therefore, these variables will be removed. Other variables containing missing values, importantly: neighborhood, zipcode, as well as jurisdiction_names, since the proportion is small, the models can be incorporated.

For the rest of the factors, we remove the street variable like the calendar_updatedand calendar_last_scraped variables are already in the zipcode.

Thus, one will see that there are certain factors include "NA", which we will not use as they do not have the modeling value in the next part of the analysis. This stage offers descriptive statistics of our models for the dependent variable as shown in Tables 2, such as the mean and standard deviations as well as the min and max values.

These statistics are, of course, very useful to understand the data used, in addition to the graphics. For example, the average Paris apartment price is 110,78 euros per night in all observations.

One other important point is we've got a very strong limit of 25.000 compared with the average. Therefore, our task is not completed. The thesort_values function is applied to retrieve price variable data (Figure 4). Figure 4 shows that the price is not inconsistent but large only.

III - Results and discussion :

1. Exploratory analysis

It is now important to consider if each of these supplies provides fully or partially overlapping information following the determination of the explanatory variables used in the model. has been classified as "multicollinearity" in the regression process, which exists if some separate variables are strongly correlated.

To recognize the significance of such a regression problem, the correlations need to be analyzed. To fully comprehend this data and determine important features via the construction of the model will explore variables that can correlate with one another.

Here it is important to note that we will primarily use panda libraries to navigate the workflow of exploratory data and to control the relevant data, together with seaborn and matplotlib, to draw the required graphics.

Then, the corr function is used and figure 5 shows the results. When examined by comparison pairs, results outline the Pearson correlation coefficient.

Through this, we would like to illustrate how a visual aspect is generated (data visualization) to make the values clearer. The correlations among the multiple factors are better known in two colors (red and blue).

The color red refers to a negative correlation, and the scaling intensity depends on its precise value. In the meantime, a positive relationship is a color blue.

An analysis of the correlations we have found relatively effective relationships (positive or negative) between the various characteristics. The findings of the analysis of correlation, varying between the positive and negative correlations, show that a positive correlation exists:

–Between availability_60 and availability_90 (0.94);

–Between host_total_listings_countand calculated_host_listings_count (0.88);

– Between accommodates and beds (0.86).

However, the remaining variables were just interrelated moderately or weakly. We removed some and keep other variables like availability_90 since the correlation is weak. We will only maintain the variable host_total_listings_count, etc.

We’re also going to remove variable(s) with NA results (figure empty lines or columns 5 that include: requires_license, has_ availability, and is_business_travel_ready that won’t help our model.

The (independent) explanatory variables have now been selected. Finally, we have data with 20 and 58.740 variables (more than 97% of all observations). A total of 75 different factors have been excluded. Also, we need to address the properties of the dependent variable, “apartment price”. We are going to import the shuffle function, which mixes the data set to produce a number of good patterns.

The distribution of the dependent variable is essential to consider, which is provided by Figure 6. We see how highly asymmetric the prices of this data pack are (2.39). Figure 6 also shows that most data points are less than 250 (A). We used a subset of the database with prices from 50 to 250 to remove high and low values. The target variable will also be transformed decreasing asymmetry (B).

Then, we will evaluate the influence on prices of various factors. Figure 7 shows box plots for certain variables about the Airbnb price. It shows variables like room_type, bed_type, and property_type which might influence the dependent variable.

For these variables, this figure shows a comparative box plot. We keep in mind that generally there is a huge gap between the condition of the room and the location. Cleanliness, the form of the room, and the category of the bed are therefore very complex and expensive options to choose from in Airbnb.

Regarding other features available, like bedrooms and bathrooms, the results show that physical characteristics and prices are in a positive connection. This is reasonable because the apartments that are best equipped are without a doubt the costliest (Gibbs et al.2018).

The influence of the number of bedrooms and bathrooms on Airbnb's price per night in Paris is shown in Figure 8. The correlation coefficient between the two factors validates that these factors are absolutely vital in price definition (0.62).

However, bedrooms and bathrooms are not the only factors that influence the dependent variable; there is also another relevant factor that can affect prices. These include facilities, specifically the comforts of an apartment.

There are more than 65 services in the database, and the top 20 features are shown in Figure 9. Airbnb apartments offer customers the most common facilities, heating, basic equipment, kitchen, and television.

The apartment position is also an important factor, which often significantly affects prices. We showed the price as part of the neighborhood in Figure 10 in order to see how important it is. This figure shows that the apartment's location is important and influences strongly the price.

We have selected numerical variables and explored them together. In this way, these variables can be translated into a form to improve the accuracy of the results through machine learning algorithms.

2. **Modeling**

The analysis showed that the apartment's overall condition affected the price. We should also consider cleanliness, room and bed type, square footage, and so on. Certain other information is extremely useful for setting apartment prices, such as shopping access and public transport.

We selected 20 independent variables based on the exploratory analysis we conducted (see Table 3). In the present stage, the data contains the thousands of separators (","), as well as the symbol "$", to handle certain aspects of the dependent variable. We did the following operation for this purpose:

```
prices_new['price'] = prices_new['price'].map(lambda x:
                                              str(x)
                                              .lstrip('$')
                                              .replace(',',''))
prices_new['price'] = prices_new['price'].map(lambda x:
                                              int(float(x)))
```

Then we went on to the model including model assessments (among the tested models), which produces the most efficient and permits the generalization of unused data. Since this case study is designed to create an Airbnb price model in Paris, linear regression is the used model type. We, therefore, chose a linear regression that takes the various dependencies found in the data into account.

We even used the regression of Lasso and Ridge to solve several separate variables' multicollinearity problems. Data must be split up into test sets before using the model to create an unchanged data set to estimate the model performance.

```
X_train, X_test, y_train, y_test = train_test_split(updated_data, prices_new, test_size=0.3)
```

In this respect :

–X_train: all predictors in the test data-set ;
–Y_train: the target variable in the test ;
–X_test: all predictors for the entire test ;
–Y_test: the target variable in the test set; in this situation, it's the price.

In this regard, various combinations of factors have been pre-processed and designed in the earlier previous stage, using a number of variables to perform these operations. The parameters for each model have been adapted and then chosen the best model possible accuracy.

The last step before reporting results is this optimization process. It allows a robust and powerful parameter or set to be identified by using a particular problem algorithm.

We have tested various parameters α for Ridge and Lasso regression models (0.001, 0.01, 0.1, 1, 10, 100). Table 4 describes the hyperparameters and validation scores for each model.

Findings show that linear regression does not predict the estimated price because many linked data points are included in our set of variables. The Airbnb Apartments were over-adjusted to produce dreadful scores of RMSE.

It can be concluded that the prices of Airbnb depend heavily on the equipment and location available in Paris.

It must be noted that we can also use other algorithms, for example, random forests and decision tree, to determine whether such features are the main factors of the Paris price for Airbnb apartments.

IV- Conclusion:

From this study, we learned what Algorithmic Applications are, why in business contexts this phenomenon has become so important, how we can draw value from data and how companies could enhance sharing economy potential from data. These corporations are encouraged to implement an approach driven by data in many manners. The application provides a description of how enterprises could build the innovative pattern in terms of the sharing economy through data analysis technologies to gather huge quantities of information and create value through the use of algorithm application, explore new methods for correlating between, and develop techniques and tools that handle a wide arsenal of data.

We have explored in this study the role of algorithms in the expansion of sharing economy across digital platforms that using different types of algorithms. For that, we have used the Linear Regression algorithm approach. Therefore, this study illustrates the importance of algorithms and their applications for sharing economy companies. In addition, the expansion of this innovative economic model is dependent on algorithmic applications technology

References:

Borel, S., Demailly, D., Massé, D. (2015). Les fondements théoriques de l'économie collaborative. Dans Le Rapport moral sur l'argent dans le monde. AEF [Online] .Available at: https://www.researchgate.net/publication/287935542/download.

Decrop, A. (2017). *La consommation collaborative : Enjeux et défis de la nouvelle société de partage*. De Boeck supérieur, Louvain-la-Neuve

Drahokoupil, J., Fabo, B. (2016). The platform economy and the disruption of the employment relationship. *ETUI Policy Brief*, 5, 1–6.

Ertz, M. (2017). Quatre essais sur la consommation collaborative et les pratiques de multiples vies des objets. Administration PhD Thesis, Université du Québec, Montreal [Online] .Available at: https://archipel.uqam.ca/11099/1/D 3350.pdf.

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., Goodwill, A. (2018). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1).

Hodkinson, P. (2017). *Media, Culture and Society: In Introduction*. SAGE Publications, Londres

Khelfaoui, M & Sedkaoui, S. (2019). *L'Economie de Partage et le Big Data Analytics*. ISTE editions, Londres

Nicot, A.-M. (2017). Le modèle économique des plateformes: économie collaborative ou réorganisation des chaînes de valeur?. *La Revue des Conditions de Travail*, 6, 1–10.

Ranjbari. M., Morales-Alonso.G & Carrasco-Gallego. R. (2018). Conceptualizing the Sharing Economy through Presenting a Comprehensive Framework. *Sustainability 2018*, 10, 2336 ; doi : 10.3390/su 10073336.

Sedkaoui, S. (2018a). Data analytics and big data. ISTE Ltd, London, and Wiley, New York.

Sedkaoui, S. (2018b). Big Data Analytics for Entrepreneurial Success: Emerging Research and Opportunities. *IGI Global*, New York.

[1] Airbnb is the world's most popular accommodation or rental facility in cities. Airbnb has become a multinational empire around the world. This platform shows the best price ranges for each person (Khelfaoui and Sedkaoui, p.14).

[2] The CSV is an open text format representing tabular information in the form of comma-separated values, known by the acronym common-separated values.

[3] Python is an interpreted programming language for multi-paradigm and multi-platforms. It promotes structured and functional programming which is essential. It has a strong dynamic type, automatic waste collection memory management, and a system for exceptional management.

- Appendices:

Table (1): The different applications of Machine Learning algorithms

Analysis	Algorithms	Learning mode	Problem to process
Simple	Simple Regression	Supervised	Regression
	Multiple regression	Supervised	
	Naïve Bayes	Supervised	Classification
	Logistic regression	Supervised	
Complex	Hierarchical classification	Unsupervised	Cluster analysis
	K-means	Unsupervised	
	Decision tree	Supervised	Classification / Regression
	Random Forest	Supervised	
	Bootstrapping	Supervised	
	Support Vector Machine (SVM)	Supervised	
	Neural Networks	Supervised	
	kNN	Supervised	

The source: Sedkaoui, (2018), p.70

Table (2): Descriptive statistics

Parameter	Value
Mean	110.78
Standard deviation	230.73
Min	0
Max	25.000

The source: Sedkaoui and Khelfaoui, (2019), p. 180

Table (3): List of explanatory variables

Dependent variable	Independent variable	
Price	Neighborhood	host_total_listings_count
	property_type	Zipcode
	room_type	Latitude
	Accommodates	Longitude
	Bedrooms	Bathrooms
	bed_type	Beds
	guests_included	Amenities
	availability_365	extra_people
	minimum_nights	maximum_nights
	calculated_host_listings_count	instant_bookable

Table (4): RMSE for different regression models

Model	Hyperparameter	RMSE
Linear regression	By default	2368.59
Ridge	Alpha = 1.00	14.832
Lasso	Alpha = 0.1	15.214

Figure (1): Overview of the data-set

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
0	2577	Loft for 4 by Canal Saint Martin	2827	Karine	NaN	Entrepôt	48.869933	2.362511	Entire home/apt	125	3	0
1	3109	zen and calm	3631	Anne	NaN	Observatoire	48.833494	2.318518	Entire home/apt	75	3	7
2	5396	Explore the heart of old Paris	7903	Borzou	NaN	Hôtel-de-Ville	48.851001	2.358690	Entire home/apt	115	2	148
3	7397	MARAIS - 2ROOMS APT - 2/4 PEOPLE	2626	Franck	NaN	Hôtel-de-Ville	48.857576	2.352751	Entire home/apt	115	10	231
4	7964	Large & sunny flat with balcony !	22155	Anaïs	NaN	Opéra	48.874642	2.343411	Entire home/apt	99	3	6

Figure (2): Variables with missing values

	host_total_listings_count	neighbourhood_group_cleansed	latitude	longitude	accommodates	bathrooms
count	59873.000000	0.0	59881.000000	59881.000000	59881.000000	59805.000000
mean	8.456984	NaN	48.864166	2.345693	3.041148	1.108603
std	66.884320	NaN	0.018452	0.033477	1.531710	0.642671
min	0.000000	NaN	48.813083	2.220731	1.000000	0.000000
25%	1.000000	NaN	48.850827	2.324459	2.000000	1.000000
50%	1.000000	NaN	48.865229	2.348563	2.000000	1.000000
75%	2.000000	NaN	48.879374	2.371054	4.000000	1.000000
max	1305.000000	NaN	48.905774	2.473815	17.000000	50.000000

Figure (3): The emptiness of the function for all variables

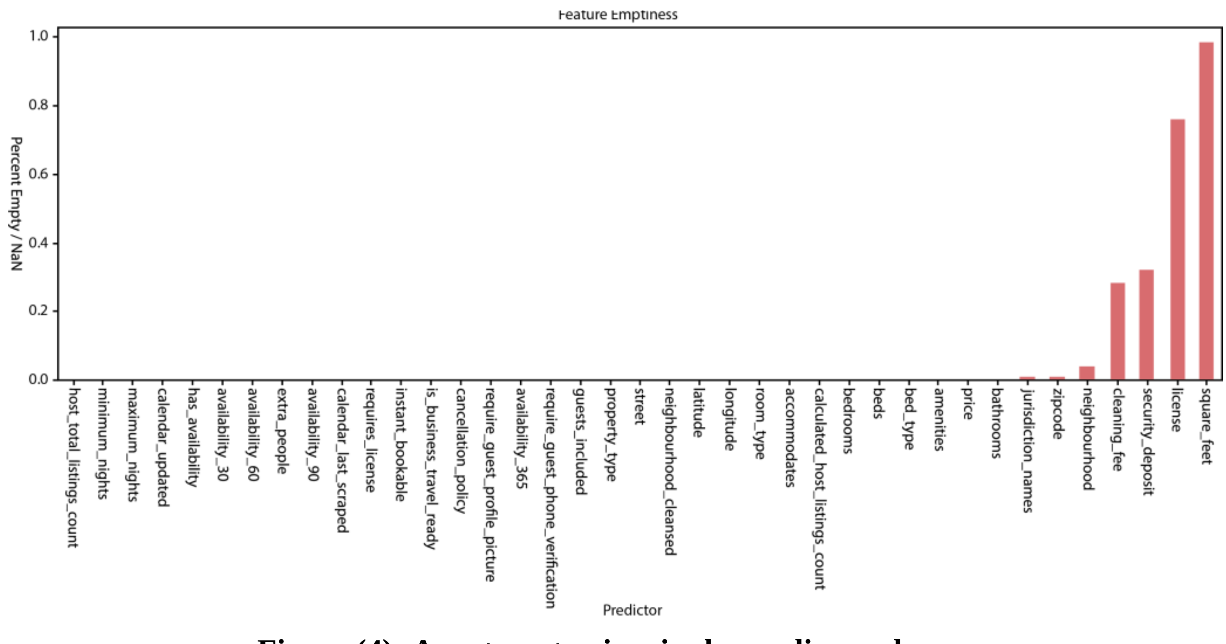


Figure (4): Apartment prices in descending order

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price
47118	24441452	Audacity & Chic close to Montmartre	184585722	Sandra	NaN	Élysée	48.881951	2.326527	Private room	25000
47117	24441436	Audacity & Chic in Paris 9th - Chambre Uno	184585722	Sandra	NaN	Élysée	48.881508	2.326565	Private room	25000
47120	24441474	Audacity & Chic close to the Sacré Cœur	184585722	Sandra	NaN	Opéra	48.881448	2.328070	Private room	25000
14852	7225849	Artistic apartment, Montmartre	34063120	Martin	NaN	Buttes-Montmartre	48.890107	2.346739	Entire home/apt	9379
53787	27608896	Petit studio parisien	170409506	Salma	NaN	Observatoire	48.832456	2.323958	Private room	8803
49042	25448670	cosytiti Paris 15	75744897	Anne	NaN	Vaugirard	48.838110	2.297415	Entire home/apt	8576
11597	6088687	Lovely Duplex 30m2 - Le Marais	2071795	Mathilde	NaN	Temple	48.865385	2.353980	Entire home/apt	8500
19064	9442661	Lively studette in 19th district	2865311	Léo	NaN	Buttes-Chaumont	48.877121	2.386713	Entire home/apt	5000
4847	2344562	1 room Apartment	1882417	Laurent	NaN	Gobelins	48.834035	2.361325	Entire home/apt	5000

Figure (5): Correlation matrix

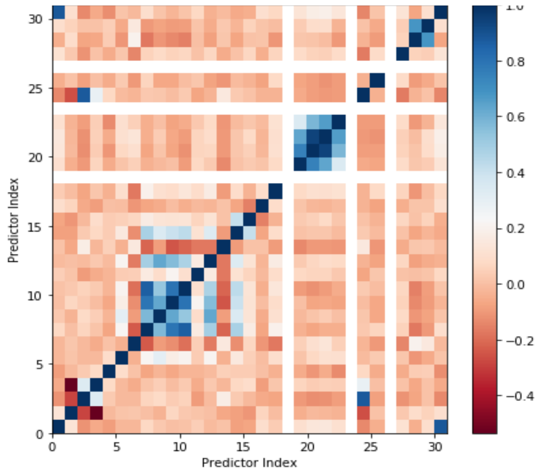


Figure (6): Price distribution

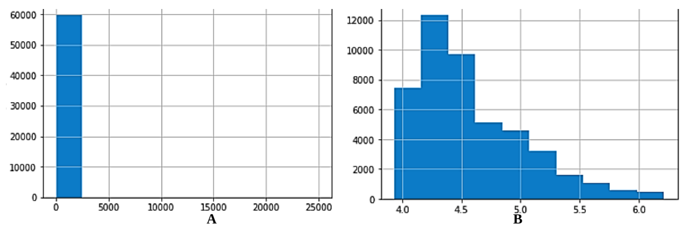


Figure 7: Prices of apartments depending on certain features

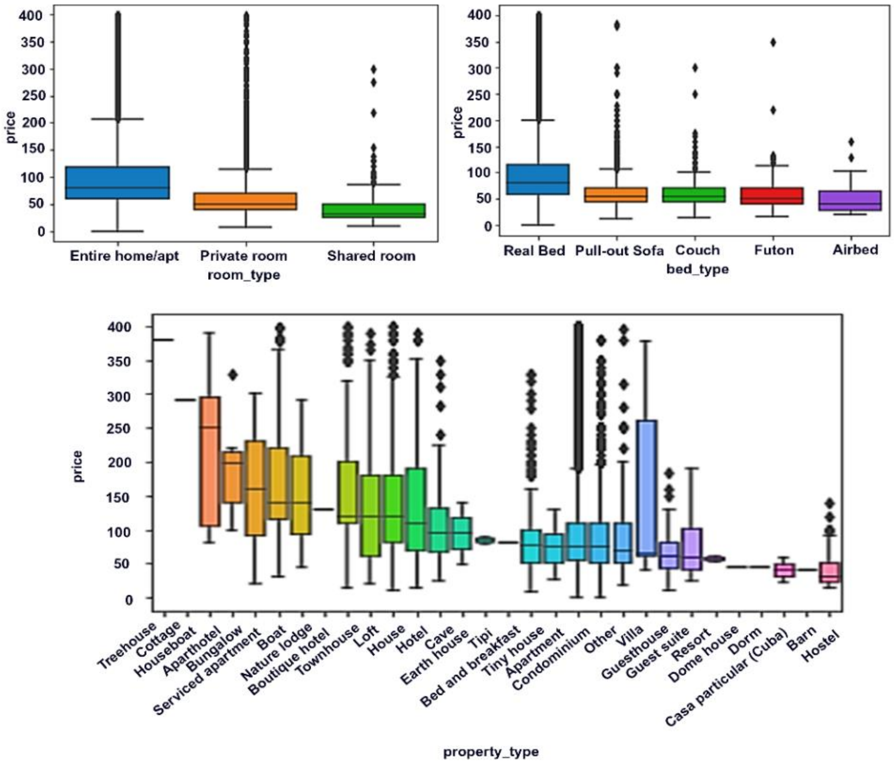


Figure 8: Price per room and bathroom number

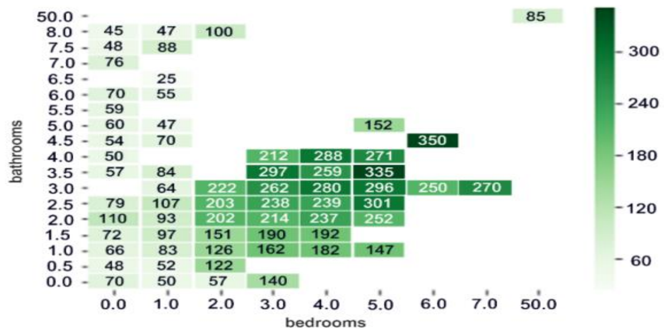


Figure 9: Pricing distribution based on amenities

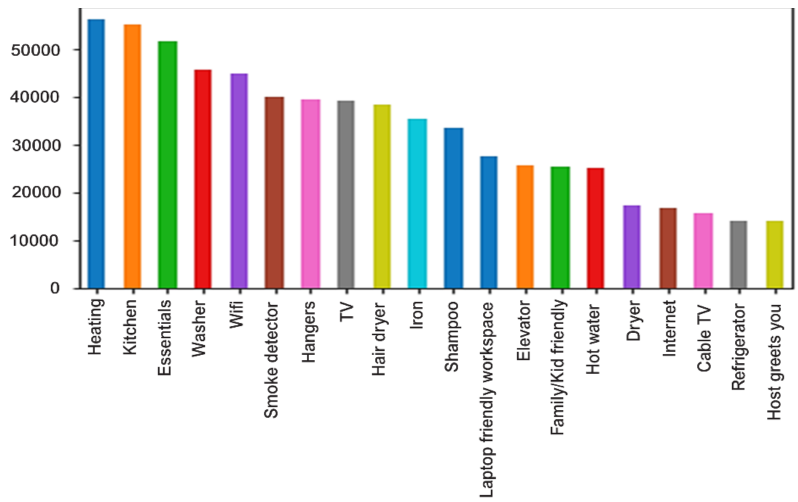
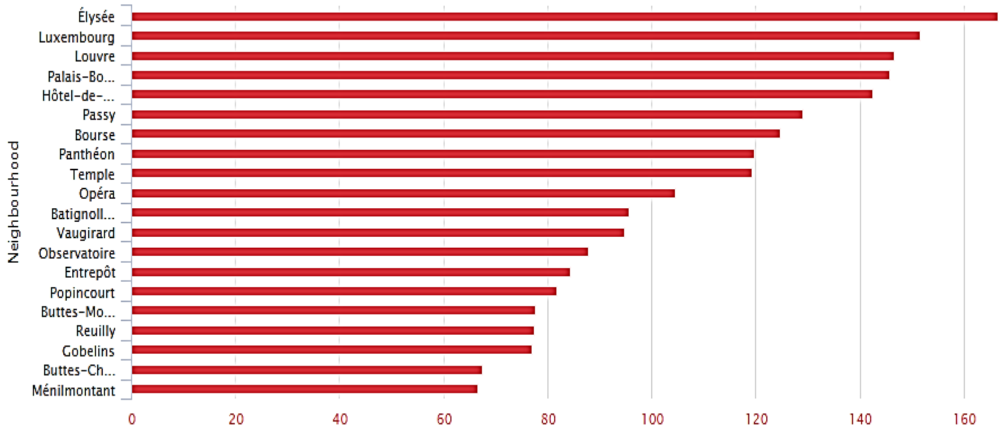


Figure 10: The price/neighborhood relationship



How to cite this article by the APA method:

Khelfaoui Mounia, Sedkaoui Soraya. (2021), Algorithmic Applications in the Sharing Economy. Predictive Price Model with Linear Regression Algorithms. Case of Airbnb Company, Roa Iktissadia Review, Volume 11 (Number 01), Algeria: Hamma Lakhther University Eloued, pp. 729-745.