Applied Mathematical Sciences, Vol. 18, 2024, no. 9, 419 - 441
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2024.919155

Shapley Value in Machine Learning Modeling: Optimizing Decision-Making in Coworking Spaces

Tiziana Ciano

University of Aosta Valley
Department of Economic and Political Sciences
Strada Cappuccini, 2A, Aosta, Italy

Massimiliano Ferrara

University Mediterranea of Reggio Calabria
Department of Law, Economics and Human Sciences
Via dell'Universit, 25, Reggio Calabria, Italy

Abstract

Game Theory is a mathematical approach to interactive decision-making situations, focusing on players and strategies. The Shapley Value is a fundamental concept in cooperative Game Theory, as it provides a fair method for distributing gains or costs among players. This study calculates the Shapley Value within machine learning models to determine the marginal contribution to the success of collaborative projects, helping to identify investments to maximize the success of coworking spaces in mountain areas. Machine learning models offer valuable insights to predict investments and strategic decisions in a mountain coworking space, ensuring and maximizing its success. The Gradient Boosting model excels at identifying key features such as internet connectivity and accessibility in mountain environments, allowing decision makers to invest in high-quality network infrastructure and accessibility improvements for coworking spaces.

Mathematics Subject Classification: 91A06, 91A12, 68T05, 68T20

Keywords: Shapley Value, predictive models, coworking space, mountain area.

1 Introduction

Over the last decade, a relatively new phenomenon has emerged known as coworking spaces (see [1, 2, 3, 4]) or subscription-based workspaces in which individuals and teams from different companies work in one space shared and common [5]. The modern workspace has evolved due to technological advancements, changing work models and the global pandemic [6]. Local coworking spaces offer a dynamic and economical alternative to traditional office structures [7]. Furthermore, they provide a flexible, diverse and dynamic working environment for various professionals at reduced costs [8, 9]. Coworking spaces offer a desk or workspace for rent and provide access to a community of like-minded people, thus creating a node of professional and private life, called a social hub [3, 4, 10, 11, 12, 13]. The birth and development of coworking spaces have proven to be a predominantly urban phenomenon. In recent years, however, peripheral and rural areas are becoming very attractive for this type of new workplace, even if the literature on this topic is limited [14]. Numerous studies have shown that coworking spaces are mainly concentrated in urban centers, where knowledge workers and urban services are concentrated. These services can be both productive and non-productive, such as good access to restaurants, cafes, shops, cultural and entertainment services, and a good level of environment [15]. This explains why the literature on new spaces of work mainly deals with large areas urban and metropolitan regions (see [14]). In the wake of the Covid-19 pandemic, attention has shifted to more suburban and rural workplaces that can accommodate employees working remotely. In fact, the Covid-19 pandemic has influenced the types of work, work environments and geography of work, making rural and peripheral areas more attractive than before [14]. Many coworking spaces are located in large metropolitan areas in close proximity to their customers, usually highly qualified ICT professionals, freelancers, employees etc [16]. However, non-urban coworking spaces have received much less attention than their metropolitan counterparts, especially in systematizing explanations of location factors [17]. In [18] explore the determinants of the location of coworking spaces in Italy, focusing on 549 spaces in 2018. The results show that coworking is mainly an urban phenomenon, with major cities favored for their innovation and business environment [18]. The study also explores whether coworking spaces can foster development in peripheral and internal areas, especially during the Covid-19 pandemic, where smart working is more widespread [18]. Capdevila's [19] study explores the spread of coworking in rural areas of Catalonia, highlighting the shift from urban centers to remote working due to technological advances and digital advances. Coworking is seen by policy makers as an opportunity to promote socioeconomic growth and urban regeneration [20]. In [21] examined the aspect of urban planning for Toronto (Canada) and considers coworking as an essen-

tial factor to achieve the sustainability goal, as this style of working promotes a stronger collaboration culture, reduces traffic and moves workers to more remote regions, creating opportunities for urban infrastructure planning [22]. The objective of this work is the application of predictive models combined with the Shapley Value to identify which features most influence the success of a collaborative project within a coworking space in a mountain area.
The rest of the work is divided as follows: in section 2 the literature review concerning the Shapley Value and machine learning is described, while in sections 3, 4 and 5 an overview of the modeling relating to the Shapley Value with the corresponding topological properties. In section 6 we present an application of the Shapley Value and predictive models in coworking contexts and present the corresponding results. Finally, in section 7 an analysis of the critical dependencies for the success of coworking in mountain areas is developed.

2 Literature review: Shapley Value and machine learning

Game Theory is a mathematical approach to interactive decision-making situations, in which agents make choices based on their preferences. It uses board game terminology to describe these situations, the players and their strategies, as introduced by John von Neumann and Oskar Morgenstern [23]. The Shapley Value is a fundamental concept in cooperative game theory, proposed by Lloyd Shapley in 1953. It provides a fair method for distributing the gains (or costs) resulting from cooperation among players in a game, based on each player's individual contributions participant. This value is particularly useful in contexts where it is difficult to quantify the specific contribution of each player. The Shapley Value has become the basis for several methods that attribute a machine learning model's prediction on an input to its basic features. The use of the Shapley Value is justified by citing the uniqueness result of Shapley [24], which shows that it is the only method that satisfies some good properties (axioms) [25]. Predictive models are increasingly utilized in managerial and operational decision-making due to the use of complex machine learning algorithms, growing computing power, and increased data acquisitions [26]. Borgonovo et al.[26] study how to improve the connection between Shapley Values and sensitivity analysis for managerial modeling and insights by linking value functions from local to global scales, introducing finite-change Shapley Values and implementing a glocal approach. In [27, 28] use the Shapley Value to attribute the goodness of fit (R^2) of a linear regression model to its features by retraining the model on different subsets of features. In [29, 30] apply the Shapley Value to study the importance of a feature for a given function, using it to identify the variance explained by the feature. Lundberg &

Lee [31] also investigate the Shapley Value with conditional expectations; they construct various approximations that make assumptions about the function or distribution and apply them compositionally on modules of a deep network. In [32, 33] use the Shapley Value to solve the attribution problem, i.e. the importance of features for a specific prediction. Strumbelj et al.[32]; they apply the Shapley Value by retraining the model on every possible subset of the features. Strumbelj & Kononenko [33] apply the Shapley Value to the conditional expectation of a specific model (without retraining). Lundberg et al.[34] computes the Shapley Value with conditional expectations efficiently for trees; However, it is not very clear about his assumptions about the distribution of features. In [35] apply the Shapley Value to the conditional expectations of the model function with an artificial distribution that is the product of the marginal distribution of the underlying features. In this work we identify five features that most influence the success of a collaborative project within a coworking space in a mountain area. By collaborative project we mean any collaborative initiative or task that a team in a coworking environment undertakes to accomplish. These projects can vary widely in nature and purpose, depending on the type of coworking and the goals of the individuals involved. There have been several studies addressed in the literature on the Shapley Value and predictive models but little literature, at least as far as we know, is found on the applications of these models to coworking spaces. Pan, et al.[36], conducted a study in London and used machine learning to analyze occupancy levels and user behavior in a flexible coworking space. The results showed that shared areas near windows are preferred for communication and work, while semi-enclosed spaces are preferred for focused work. The study provides insights for future human-centered space planning, particularly in hybrid work setups and coworking systems. In [37] coworking spaces act as a Schelling point, providing a focal point for finding people, ideas and resources when coordination is lacking. The model tests predictions about successful organizational and institutional forms of coworking spaces. Therefore, in this work, we calculate the Shapley Value, within the machine learning models, for each feature to determine its marginal contribution to the success of the collaborative project complexity helping to identify investments that maximize the success of the coworking space in a mountain area.

3 The Shapley Value: a topological approach

In a cooperative approach of dynamical competitive situations, the Shapley Value represents a fair sharing of the total utility available for the grand coalitions N. Unlike the core and stable sets, the Shapley Value uniquely determine an utility vector, so that it can be interpreted as a point-solution for a game.

3.1 Formalizations

Let us consider a *TU* (as Transferable Utility) game (N, v). Unless we will assume the contrary, v will be an arbitrary real-valued function defined on 2^N, with $v(\tau) = 0$ (not necessarily superadditive).

Definition 3.1. *The Shapley Value of the game (N, v) is the vector $\tau(v)$, where, for each $i \in N$*

$$\tau(v) = \sum_{\sigma \in \sigma_n} \frac{1}{n!} \left[v\left(P_i^\sigma \cup \{i\}\right) - v\left(P_i^\sigma\right) \right] \tag{1}$$

and

$$P_i^\sigma = \{ j \in N \mid \sigma(j) < \sigma(i) \} \quad \forall \sigma \in \sigma^n$$

(Here, σ_n is the group of all permutations on N).
If, for every $\sigma \in \sigma_n$, one defines the n-vector $a^\sigma(v)$ of components $a^\sigma(v)_i = v(P_i^\sigma \cup \{i\}) - v(P_i^\sigma)$, then the Shapley Value is the aritmetical mean of the vector $a^\sigma(v)$, $\sigma \in \sigma_n$, i.e $\phi(v) = \sum_{\sigma \in \sigma_n} \frac{1}{n!} a^\sigma(v)$. The vector $a^\sigma(v)$ are called "marginal worth vector" characteristic function v. An equivalent definition of the Shapley Value follows from the next propositions.

Proposition 1. *For every $i \in N$, one has:*

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1!}{n!} \left[v(S \cup \{i\}) - v(S) \right] \tag{2}$$

Proof: For any $S \subseteq N \setminus \{i\}$, denote by $\sigma_n(i, S) = \{\sigma \in \sigma_n | P_i^\sigma = S\}$. Obviously $\cup_{S \subseteq N \setminus \{i\}} \sigma_n(i, S) = \sigma_n$. On the other hand, $|\sigma_n(i, S)| = |S|!(n - |S| - 1)!$. Then, grouping the terms in 1, directly obtain (2).

Remark 1. *Denote by $\gamma_s = \frac{|S|!(n-|s|-1)!}{n!}$.*
Then

$$\sum_{S \subseteq N \setminus \{i\}} \gamma_S = \sum_{S=0}^{n-1} \frac{(n-1)}{S} \; \frac{S!(n - S - 1)!}{n!} = 1$$

Remark 2. *The summation in (2) can be extended up to all subsets of N. Indeed, if $i \in S$, then $S \cup \{i\} = S$. So that the corresponding term in the sum is* 0.

4 Topological properties of the Shapley Value and their formalization

In the sequel we will consider τ as a mapping which associate to each n-person *TU* cooperative game an n-vector $\tau(v)$, in our vision, this vector can be the Shapley Value.

The set of all n-person games is naturally identified with the set of all characteristic functions $\mathbb{N} = \{v : 2^N \longrightarrow \mathbb{R} \, | v(\phi) = 0\}$, which in turn can be identified with the $2^n - 1$ dimensional euclidean space (the value for the empty set is dropped). In this way, $\mathbb{N}$ is organized as a real vector space, so that the sum of the two games (characteristic functions) and the product with scalars are well defined. For any $\sigma \in \sigma_n$ define the game $\sigma v(S) = v(\sigma(S))$.

Proposition 2. *Let be $\tau : \mathbb{N} \longrightarrow \mathbb{R}^n$ such that $\tau(v)$ is the Shapley Value of the game (N, v) for every $v \in \mathbb{N}$.Then τ satisfies the following properties:*

1. *Linearity:* $\tau(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 \tau(v_1) + \alpha_2 \tau(v_2), \ \forall \alpha_1, \alpha_2 \in \mathbb{R}, v_1, v_2 \in \mathbb{N}$
2. *Anonymity:* $\tau(\sigma(v)_i) = \tau(v)_{\sigma(i)}, \forall \alpha_1, \alpha_2 \in \mathbb{N}$
3. *Symmetry:* $v(S \cup \{i\}) = v(S \cup \{i\}), \forall S \subseteq N$ *for some* $i, j \in N, \ \tau(v)_i = \tau(v)_j$
4. *"Dummy player" axiom:* $v(S \cup \{i\}) = v(S), \forall S \subseteq N$ *for some* $i \in N \longrightarrow \tau(v)_i = 0$
5. *Efficiency:* $\sum_i \in N \tau(v)_i = v(N)$
6. *Monotonicity with respect to i:* $v(S \cup \{i\}) - v(S) \geq v'(S \cup \{i\}), \forall S \subseteq N$ *for some* $v, v' \in \mathbb{N}$, *e* $i \in N \rightarrow \tau(v)_i \geq \tau(v')_i$
7. *Individual rationality:* $v(S) = v(S \cap T), \ \forall S \subseteq N$ *for some* $T \subseteq N \rightarrow \sum_i \in T \tau(v)_i = v(T)$
8. $\tau(v)_i \geq v(\{i\})$, *whenever* v *is superadditive.*

Proofs of each property:

1. *Trivially follows from (2).*
2. *By using (2) we can write*

$$\tau(v)_i = \sum_{S \subseteq N \setminus \{\cdot\}} \tau_S[v(\sigma(S \cup \{i\}) - v(\sigma(S))].$$

Put $Q = \sigma(S)$ *since* $|Q| = |\sigma(S)| = |S|$ *it follows that* $\gamma_s = \gamma_Q$ *and then*

$$\tau(v)_i = \sum_{Q \leq N \setminus \{v(i)\}} \gamma_Q[v(Q \cup \{\sigma(i)\}) - v(Q)] = \tau(v)_{\sigma(\cdot)}.$$

3. *Let be* $\sigma \in \sigma_n$ *such that* $\sigma(i) = j$, $\sigma(j) = i$ *and* $\sigma(k) = k$, $\forall k \in N \setminus \{i, j\}$. *Obviously,* $\sigma v = v$ *and then, it follows by 2, that* $\tau(v) = \tau(\sigma v) = \sigma \tau(v)$. *Particularly,* $\tau(v)_i = \tau(v)_j$.

4. *Follows from property 2.*

5. *It suffices to verify that*

$$\sum_{i\in N} a^{\sigma}(v)_i = v(N), \forall \sigma \in \Sigma_n,$$

denote by $i_k = \sigma^{-1}(k)$ *for each* $k \in N$. *Then*

$$\sum_{i\in N} a^{\sigma}(v)_i = \sum k \in N a^{\sigma}(v)_i k = \sum_{k\in n} [v(P_i k^{\sigma} \cup \{ik\}) - v(P_i k^{\sigma})].$$

But

$$P_i k^{\sigma} = \{j \in N | \sigma(j) < k = \{i_1, i_2, ..., i_{n-1}\}.$$

Hence

$$\sum_{i\in N} a^{\sigma}(v)_i = v(N) - v(\{i_1, i_2, ...i_{n-1}\} + v\ \{i_1, i_2, ...i_{n-2}\}) + v\ (i_1 - v(\phi)) = v(N)$$

6. *Obvious from 2.*

7. *If* $i \notin T$ *then*

$$\tau(v)_i = \sum_{s\leq N\setminus\{i\}} \gamma_s [v(S\cup\{i\}) - v(S)] = \sum_{S\cup N\setminus\{i\}} \gamma_s [v(S\cap T) - v(S\cap t)] = 0$$

Therefore, it follows by v. *That*

$$\sum_{i\in T} \tau(v)_i = \sum_{i\in N} \tau(v)_i = V(N) = v(T)$$

8. *If* v *is superadditive then* $v(S\cup\{i\}) - v(S) \geq v\{i\}$ *for every* $i \in N$.

Remark 3. *A weaker form of linearity is the additivity:*

$$\tau(v_1 + v_2) = \tau(v_1) + \tau(v_2), \forall\ v_1, v_2 \in \mathbb{N}.$$

Remark 4. *Any* i *satisfying the assumptions of 4* $(v(S\cup\{i\} = v(S)\forall S \subseteq N)$ *is called "dummy player". He has null worth* $(v\{i\} = v(\phi))$ *and can nothing for any coalition.*

Remark 5. *Any coalition* T *satisfing the assumptions of 6 is called a "carrier of* v*".*

Remark 6. *As it follows from 5 and 7 the Shapley Value of any superadditive game is an imputation. The individual rationality (property 8) can fail if 5 is not superaadditive.*

5 The Shapley Value and its axiomatic approach: new findings

Starting from the properties have just showed, let us introduce the following:

Theorem 5.1. *The Shapley Value can be viewed as a unique value function which satisfies all properties 1, 3, 4, 5. This function is the Shapley Value.*

Proof: *As it was precedently proved, the Shapley Value (denoting by τ) satisfies the mentioned properties. Let us show that if ψ is a value functions satisfying the above four properties, then $\psi \equiv \tau$ for every $T \subseteq N$, define the characteristic function $\rho_T \in \mathbb{N}$, by:*

$$\rho_T(S) = \begin{cases} 1, & \text{if } S \supseteq T, \\ 0, & \text{otherwise.} \end{cases}$$

a) *Prove that $\phi(\alpha_{\rho_T}) = \psi(\alpha_{\rho_T})$, $\forall\ T \supseteq N$ and $\forall \alpha \in \mathbb{R}$. Note first that if $i \notin T$ then $\rho_T(S \cup \{i\}) - \rho_T(S) = 0$, so that i is a dummy player and by 4 it follows $\psi(\alpha_{\rho_T})_i = 0 = \tau(\alpha_{\rho_T})_i$. Pick now $i, j \in T$, and $S \subseteq N \setminus \{i, j\}$. Then, $\rho_T(S \cup \{i\}) = \rho_T(S \cup \{j\})$ and by 3 it follows that $\psi(\alpha_{\rho_T})_i = \psi(\alpha_{\rho_T})_j$ by v, one has*

$$\sum_{i \in N} \psi(\alpha_{\rho_T})_i = \sum_{i \in N} \psi(\alpha_{\rho_T})_i = \alpha_{\rho_T}(N) = \alpha.$$

Therefore, $\psi(\alpha_{\rho_T})_i = \frac{\alpha}{|T|}$ for every $i \in T$. One can easily verify that $\tau(\alpha_{\rho_T})_i = \frac{\alpha}{|T|}$ for every $i \in T$, so that condition (a) holds.

b) *Prove that every $v \in \mathbb{N}$ can be expresses as*

$$v = \sum_{T \in 2^N} \alpha_T \rho_T$$

where

$$\alpha_T = \sum Q \subseteq T(-1)^{|T|-|Q|} v(Q)$$

Let us verify the above equality for each value of the argument:

$$\sum_{T \in 2^N} \alpha_T \rho_T(S) = \sum_{T \subseteq S} \alpha_T = \sum_{T \supseteq S} \sum_{Q \supseteq T} (-1)^{|T|-|Q|} v(Q) = \sum_{T \subseteq S} (-1)^{|T|} \sum_{Q \subseteq T} (-1)^{|Q|} v(Q)$$

$$= \sum_{Q \subseteq S})[(-1)^{|Q|} + \binom{|S|-|Q|}{1} (-1)^{|Q|+1} + \binom{|S|-|Q|}{2} (-1)^{|Q|+2} + \ldots +$$

$$+ \binom{|S|-|Q|}{|S|-|Q|} (-1)^{|S|} v(Q)] = \sum_{Q \subseteq S} (-1)^{|Q|} (1-1)^{|S|-|Q|} v(Q) + v(S) = v(s)$$

c) By using the additivity of both ψ and τ we complete the proof:

$$\psi(v) = \sum_{T\in 2^N} \psi(\alpha_{T\rho_T}) = \sum_{T\in 2^N} \tau(\alpha_{T\rho_T}) = \tau(v)$$

Basically the above theorem claims the utility of the value function under the given assumptions. The leading idea of the proof is that the theorem are determined by their values for the simple games (characteristic functions) $\rho_T, T \in 2^N$. For such games any ψ has the same value ad τ. Therefore, it was not necessary to obtain explicitly the value function for an arbitrary characteristic function. We will show in the next that the explicit expression of the of the SV can be obtain from the axioms required by the theorem.

Proposition 3. *The unique value function τ of 5.1 associates to each game (N, v) the vector $\tau(v)$ defined by (2)*

Proof: As it was show at the step (b), $\tau(v) = \sum_{T\subseteq N} \alpha_T\ \tau(\rho_T)$ from (a) it follows that $\tau(\rho_T) = \frac{1}{|T|}e^T$, where $e^T \in \mathbb{R}^n$

$$e_i^T = \begin{cases} 1, & \text{if } i \in T \\ 0, & \text{otherwise.} \end{cases}$$

Denote By $t = |T|$ and $s = |S|$. Then

$$\tau(v)_i = \sum_{\substack{T\subseteq N\\ i\in T}} \alpha_T\, \frac{1}{t+1} = \sum_{T\subseteq N\setminus\{i\}} \alpha_{T\cup\{i\}}\, \frac{1}{t+1} = \sum_{S\subseteq N} \left(\sum_{\substack{T\subseteq N\setminus\{i\}\\ S\setminus\{i\}\subseteq T}} \frac{1}{t+1}(-1)^{t+1-s} \right) v(s)$$

$$= \sum_{S\subseteq N\setminus\{i\}} \sigma_i^s v(s) + \sum_{\substack{s\subseteq N\\ i\in s}} \delta_i^s v(s)$$

$$= \sum_{S\subseteq N\setminus\{i\}} \delta_i^S v(s) + \sum_{S'\subseteq N\setminus\{i\}} \delta^{s'\cup\{i\}} v(s'\cup\{i\})$$

where we have used the notation

$$\delta_i^s = \sum_{S\setminus\{i\}\subseteq T\subseteq N\setminus\{i\}} \frac{1}{t+1}(-1)^{t+1-s}$$

but

$$\delta^{S'\cup\{i\}} = \sum_{S'\subseteq T\subseteq N\setminus\{i\}} \frac{1}{t+1}(-1)^{t-s}$$

so that

$$\tau(v)_i = \sum_{S\subseteq N\setminus\{i\}} \left(\sum_{S\subseteq T\subseteq N\setminus\{i\}} \frac{1}{t+1}(-1)^{t-s} \right) [v(S \subseteq \{i\}) - v(S)]$$

observe finally, that

$$\sum_{S\subseteq T\subseteq N\setminus\{i\}} \frac{1}{t+1}(-1)^{t-s} = \sum_{t=s}^{n-1} \binom{n-s-1}{t-s} (-1)^{t-s} \int_0^1 x^T\,dx =$$

$$= \int_0^1 x^s \left(\sum_{t=s}^{n-1} \binom{n-s-1}{t-s} (-1)^{t-s}\, x^{t-s} \right) dx =$$

$$= \int_0^1 x^s (1-x)^{n-s-1}\,dx =$$

$$= B(s+1, n-s) = \frac{\Gamma(s+1)\Gamma(n-s)}{\Gamma(n+1)} = \frac{s!(n-s-1)!}{n!}$$

and hence 2 is established

■

In theorem 5.1 axiom 4 and 5 can be replaced by axiom 7, as it follows from the next propositions.

Proposition 4. *The value function τ verifies 7 if and only if it verifies 4 and 5.*

Proof: *Assume first ψ satisfy 7 obviously, N is a carrier of every $v \in \mathbb{N}$. Then 7 implies the efficiency. Let i by a dummy player then, $N \setminus \{i\}$ is a carrier of 5, so that, by 7 one has*

$$\sum_{j\in N\setminus\{i\}} \tau(v)_j = v(N \setminus \{i\}) = v(N)$$

Altogether the efficiency and this relations give us $\tau(v)_i = 0$ assume now that psi satisfies 4 and 5. Let T be a carrier of 5. Then, by 5, it follows

$$v(T) = v(N) = \sum_{i\in N} \tau(v)_i$$

Pick an $i \notin T$. The for every S, one has

$$v(S) = v(S \cap T) = v((S \cup \{i\}) \cap T) = v(S \cap \{i\})$$

so that i is dummy. By 4 one has $\psi(v)_i = 0$. Therefore,

$$v(T) = \sum_{i\in T} \psi(v)_i$$

6 Shapley Value and Coworking modeling: dynamics and predictive models

In coworking spaces, collaboration and interaction between different skills and resources are pivot. The Shapley Value offers a robust methodology for assigning value to individual contributions in a group context. This approach can be particularly useful for optimizing resources and teams in coworking environments. We define a set of features $X = x_1, x_2, ...x_n$ and a predictive model $f(X)$ that estimates the success of the project. The cooperative game (N, v) consists of the set N of all features and the function $v(S)$, which evaluates the model using only the features in $S \subseteq N$. In other words, $v(S)$ is the performance or accuracy of the predictive model using only the features in S. We define the Shapley Value as in (2) where $\tau(v)$ represents the importance of the feature x in determining the success of the project, that is, it determines the marginal contribution of each feature to the overall success of the project, allowing you to identify and prioritize the most influential features. The objective of this study is to optimize the distribution of human and material resources to maximize the productivity and effectiveness of coworking in a mountain area. Suppose we want to establish a coworking space in a mountain area that attracts not only the local population but also tourists and professionals looking for a quiet retreat to work on creative and technological projects. We consider as features:

1. Internet connectivity: essential to ensure that professionals can work effectively without interruptions.

2. Accessibility: Important as some mountain areas can be difficult to reach, especially in adverse weather conditions.

3. Environmental Sustainability: important to maintain harmony with the mountain environment and to attract environmentally conscious individuals.

4. Ease of Collaboration: This is crucial because coworking spaces are designed not only to provide a physical place to work, but also to promote the exchange of ideas, the building of professional relationships and the creation of new collaboration opportunities between members.

5. Support Services: such as the availability of accommodation, restaurants and other facilities that enhance the coworking space experience.

As mentioned above, these features represent the set N, while $v(s)$ is the expected success of the project when only the features in S are available. For example v(Internet Connectivity, Accessibility) could be the evaluation of the success of the coworking space when only these two features are optimized. The Shapley Value for each feature will be calculated to determine their marginal contribution to the overall success of the project. This value will help identify which investments in specific features bring the greatest benefits in terms of the success of the coworking space. Using a predictive model that includes the Shapley Value helps to better understand how different features contribute to the success of a coworking space in a mountain area.

6.1 A synthetic dataset for Modeling tests

We generate a synthetic dataset where we create artificial data that mimics the features of a real dataset. The synthetic data was generated with the aim of replicating the main features that one would expect to observe in a real dataset, while maintaining a certain degree of variability and complexity. We opted since we did not have a real dataset available, we opted to use synthetic data, which allowed us to proceed with the analysis effectively. We are going into the project we are promoting (see acknowledgments) to collect a dataset from a sample by Aosta Valley Region (Italy) and the involved stakeholders. The synthetic, artificially generated data was critical to being able to explore the dynamics of our study without the need to access sensitive or difficult-to-obtain information. In the context of a coworking space in a mountain area, we can imagine that the identified features are measured in the following way:

1. Internet Connectivity: The speed of your internet connection, measured in Mbps.
2. Accessibility: the ease of access to the location, rated on a scale of 0 to 10, where 0 indicates that it is very difficult to reach and 10 that it is very easy.
3. Environmental Sustainability: measures the environmental impact, evaluated on a scale from 0 to 10.
4. Ease of Collaboration: how much the spaces facilitate collaboration between members, rated on a scale of 0 to 10.
5. Support services: number of services (such as restaurants, accommodations, etc.) available, evaluated as an integer representing the quantity of services available (e.g. 0 to 20).

After defining the features, the next step is to generate data for each features. Therefore, we use the functions of the numpy library of Python which allows us to generate random numbers following different distributions and the pandas library to manage the dataset. Additionally, we define a function to simulate project success based on features, while also adding an element of noise to make the data more realistic. In particular, we generate a synthetic dataset with 1000 samples, each of which represents a coworking space with different features. A "success" score is calculated for each sample, based on a weighted combination of features. Each feature contributes to the total score with a specific weight: internet connectivity has a weight of 30%; accessibility has a weight of 25%, environmental sustainability has a weight of 15%; ease of collaboration has a weight of 20%; support services has a weight of 10%. In a digital age, the quality of internet connectivity is critical to attracting and retaining members, especially in remote locations such as mountain areas, where connection may be problematic. Professionals who depend on the internet for teleconferencing, cloud computing, and other work needs will strongly value this. We assign the highest percentage because we believe it is the most decisive feature for the success of a coworking space in a mountain area as it represents the backbone of daily operations for the majority of modern professionals who work in smart working. Without a reliable connection, the functionality of the coworking space would be severely compromised, limiting its ability to attract and retain members. We then normalize the features so that their values lie between 0 and 1. Finally, a small normally distributed random value (noise) with mean 0 and standard deviation 5 is added to the final score to simulate real variability. So we get a function that returns the simulated success as a numeric value. We proceed to train three predictive models, namely Random Forest Regressor, K-Nearest Neighbors and Gradient Boosting, and evaluate their performance to identify the best model.

6.2 Estimating of the Shapley value

Once the model is trained we use the *SHAP* library to calculate the Shapley Values, which will help us understand the importance of each feature in determining the success of the project. The training dataset is used as a background to compare the predicted values and evaluate the importance of the features. Shapley Values quantify the contribution of each feature to the model's prediction for a given sample. A positive value indicates that the feature increases prediction, while a negative value indicates that it decreases prediction. Shapley Values are calculated for each sample in the test set and indicate the contribution of each feature to the model's predictions. We use these values to visualize and interpret how each feature influenced the model's predictions.

6.3 Results

These values are related to each feature and provides informations on the relative importance as well as how variations in the features affect the predictions of the Random Forest model.

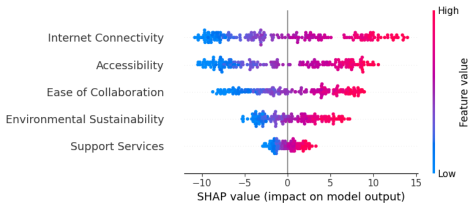

Figure 1: Distribution of Shapley Values for each feature: Random Forest model

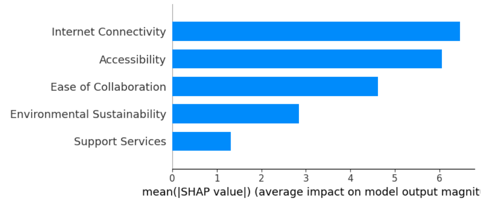

Figure 2: Absolute average features: Random Forest model

Each Shapley Value represents the marginal contribution of a specific feature to the model's prediction for a given sample. The graph shows that the features are sorted vertically by decreasing importance, i.e. the features at the top of the graph have the greatest impact on the model's predictions. This allows us to see not only which feature is most important, but also how its impact varies depending on the value of the feature itself. The color of the dots indicates whether the feature values are high or low. For example, if the red dots (high values of the feature) are mainly on the right, it means that high values of that feature tend to increase the prediction of the model and therefore increase the prediction of the success of the coworking space. We note that internet connectivity is the most important feature, with a long horizontal bar and a distribution of Shapley Values showing how different internet speeds affect the predictions. This suggests that high internet speed is crucial to the success of a coworking space in such areas, where stable and fast connection can be a decisive factor in attracting professionals. Accessibility and other features are ranked below, with shorter bars, indicating they have less impact than internet connectivity. Thanks to this representation we are able to identify the most influential features on the successful model of a coworking space in a mountain area and understand how the values of the features influence the prediction of the model (if they increase or decrease the predicted value) and finally analyze the heterogeneity of the contribution of each feature to the predictions (see Figure 2). We apply, on the same dataset, the K-Nearest Neighbors (KNN) and Gradient Boosting models. We visualize the results of the models in Figures 3-6.

We calculate the evaluation metrics for the various models and note that from the results obtained, Gradient Boosting is the best model among the three, since it has the lowest MSE (Mean Squared Error) and RMSE (Root

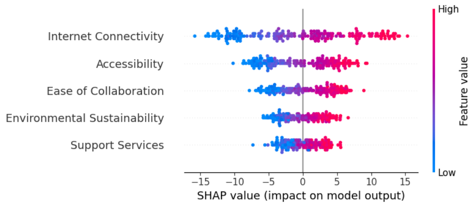

Figure 3: Distribution of Shapley Values for each feature: K-Nearest Neighbors

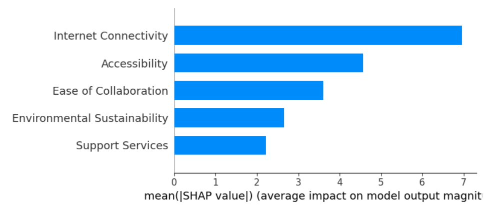

Figure 4: Absolute average features: K-Nearest Neighbors

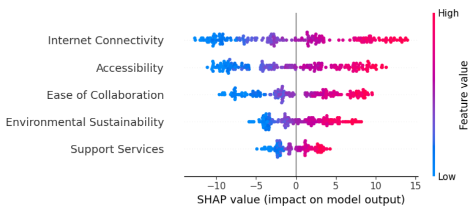

Figure 5: Distribution of Shapley Values for each feature: Gradient Boosting

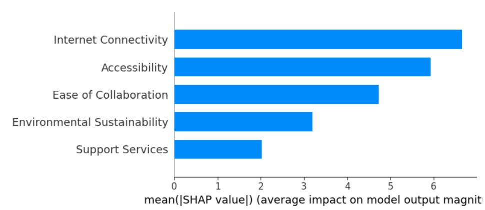

Figure 6: Absolute average features: Gradient Boosting

Mean Square Error), which means that makes fewer errors in predictions. It also has the highest value of R^2 (coefficient of determination), which indicates that it explains the variability in the data better. Finally, it has the lowest MAE (Mean Absolut Error), which means that its predictions are the most accurate in terms of mean absolute error (see table 1).

Evaluation metric	Random Forest	K-Nearest-N	Gradient Boosting
MSE	41.18	60.93	34.33
RMSE	6.42	7.81	5.86
R^2	0.80	0.70	0.83
MAE	5.02	6.02	4.61

Table 1: Evaluation metrics of the predictive models

Random Forest is the second best model, with good overall performance, but slightly worse than Gradient Boosting. While KNN has the worst performance among the three models, with larger errors and a lower ability to explain the variability in the data. Therefore, accurately predicting the success of a coworking space based on these features can help optimize resource distribution and make informed decisions. Gradient Boosting highlighted that Internet connectivity is the most influential feature, followed by accessibility and environmental sustainability. These findings are consistent with the idea that in a mountain area, a robust Internet connection and easy access are key to attracting professionals and tourists. Random Forest also showed good results, but slightly worse than Gradient Boosting. Its classification metrics are robust, but less precise. For Random Forest, Internet connectivity was also found to be the most influential feature, but the model showed a slight variation in the relative importance of other features compared to Gradient Boosting. This confirms that key features are consistent, but the model is less fine-tuned in identifying their relative importance. The performance of the KNN was significantly lower than the other two models. This is due to the nature of KNN, which is not well suited to capturing the complexities of feature relationships in heterogeneous datasets such as the one used here. KNN has shown difficulty in clearly identifying feature importance, reflecting its lower performance.

7 Planning an efficient Coworking management in Mountain Areas: a new approach based on Shapley Value

Features such as internet connectivity, accessibility, environmental sustainability, ease of collaboration and support services are all critical features in a coworking context in a mountain area. Since our goal is to analyze a model that predicts the success of a coworking space in a mountain area, we are going to calculate the contribution of each features to the predictions made by the Gradient Boosting model. Figures 7-11 show the marginal contribution of a specific features to the model prediction. This is represented by the Shapley Values along the y-axis. The x-axis represents the specific features values (for example, the level of Internet connectivity). In particular, Figure 7 represents how changes in Internet connectivity (e.g., increased speed) influence the model's prediction (prediction of coworking success), supporting us to better interpret the model's decisions.

Internet connectivity has emerged as one of the most influential features. A fast and reliable internet connection is essential to attract professionals, freelancers and tourists who depend on a stable connection to work. The graph highlighted that accessibility (the interaction between accessibility and connectivity) can further amplify the positive impact of a good Internet connection. In Figure 8, accessibility, or the ease with which you can reach the coworking space, has proven to be a crucial factor. In a mountain area, where infrastructure may be limited, improving accessibility is critical to the success of coworking. The graph highlighted a strong positive trend: coworking spaces that are easily accessible are more likely to succeed. Furthermore, this effect is enhanced when combined with good Internet connectivity. A good Internet

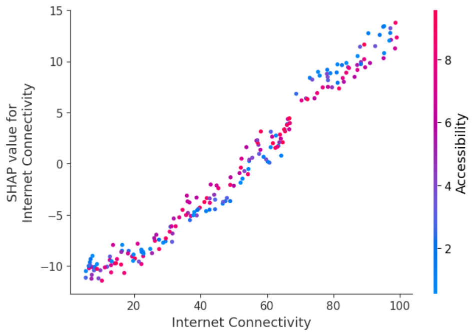

Figure 7: Dependency graph generation for internet connectivity

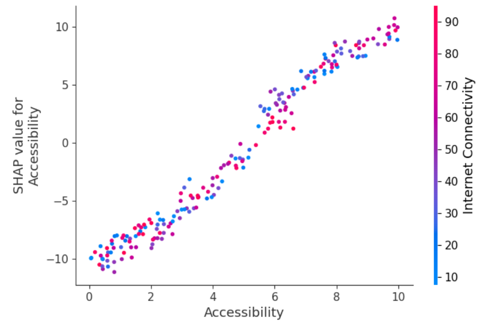

Figure 8: Dependency graph generation for accessibility

connection combined with easy access makes coworking spaces more attractive and functional, particularly in mountain areas.

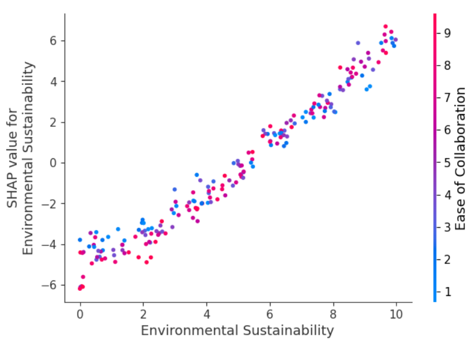

Figure 9: Dependency graph generation for environmental sustainability

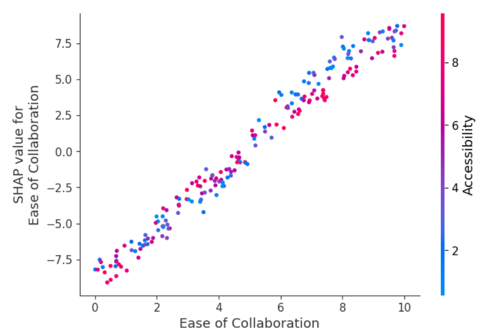

Figure 10: Dependency graph generation for ease of collaboration

In Figure 9, environmental sustainability is particularly relevant in a mountain context, where environmental conservation is a priority. The graph showed that coworking spaces that adopt sustainable practices are associated with greater success. This is further amplified if the space also facilitates collaboration between members. Investing in renewable energy, reducing ecological impact, and promoting responsible use of resources are strategies that can improve the profile of coworking and attract a more aware audience. Promoting green and sustainable practices can not only improve the image of the coworking space, but also attract a wider and more aware audience. In Figure 10, ease of collaboration is a crucial aspect for any coworking space. The graph highlighted that spaces designed to facilitate interaction and collaboration between members tend to be more successful. This effect is particularly strong when combined with high accessibility. In a mountain area, where the community may be more isolated, creating opportunities for networking and collaboration is essential to attracting and retaining members.

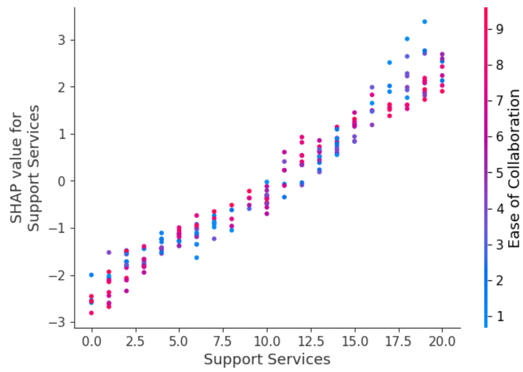

Figure 11: Dependency graph generation for Support Services

Finally in Figure 11, support services, such as restaurants, accommodation, transportation, and other amenities, emerged as a significant factor in the success of the coworking space. Spaces that offer or facilitate access to a full range of support services tend to be more attractive and more likely to succeed. This effect is enhanced if the space also promotes collaboration between members. In a mountain area, where access to amenities may be limited, it is strategic to ensure that coworking space members have access to all the amenities needed for a comfortable and productive stay. Coworking spaces that facilitate interaction between members and offer a full range of support services tend to be more successful. Combining a collaborative environment with well-organized support services is crucial to creating an attractive and productive coworking experience. Therefore, it is clear from the analysis that to maximize the success of a coworking space in a mountain area, it is essential to: Design spaces that encourage collaboration between members; Promote environmental sustainability as a key value; invest in infrastructure that ensures reliable internet connectivity and easy access and finally ensure that essential support services are available or easily accessible. These strategies, combined, can transform a coworking space in a mountain area into an attractive, functional and successful work environment.

8 Conclusions

In a coworking space in a mountain area, it is crucial to make accurate predictions to identify which investments and strategic decisions would maximize the success of the pace. Machine learning models provided valuable insights. Gradient Boosting emerges as the best model for this context, with superior performance in terms of predictive accuracy and interpretability of key features. This model precisely identified the most influential features, such as internet connectivity and accessibility, which are crucial in a mountain environment. Decision makers can use these insights to focus investments on high-quality network infrastructure and accessibility improvements to maximize the attractiveness of the coworking space. Random Forest, while slightly inferior to Gradient Boosting, offers good robustness and similar interpretability. This model also confirms that the same features are fundamental, strengthening confidence in the results. K-Nearest Neighbors (KNN) has proven to be less suitable for this type of prediction. Gradient Boosting is the best choice. This model not only provides accurate predictions but also offers a clear interpretation of the features that drive success, helping decision makers to invest more effectively. Random Forest is a good alternative for those looking for robustness and simplicity, while KNN should be avoided in this case, given its lower performance.

Acknowledgements. This publication is part of the project NODES which has received funding from the MUR M4C2 1.5 of PNRR funded by the European Union - NextGenerationEU (Grant agreement no. ECS00000036)).

References

[1] Schopfel J, Roche J, Hubert G, Co-working and innovation: new concepts for academic libraries and learning centres, *New Libr World*, **116** (2015), 6778. https://doi.org/10.1108/nlw-06-2014-0072

[2] Bouncken RB, Reuschl AJ, Coworking-spaces: how a phenomenon of the sharing economy builds a novel trend for the workplace and for entrepreneurship, *Review of Managerial Science*, **12** (2018), 317334. https://doi.org/10.1007/s11846-016-0215-y

[3] Brinks V., Netzwerke(n) und Nestwrme im coworking space-Arbeiten zwischen Digitalisierung und Re-Lokalisierung, *Geographische Zeitschrift*, **100** (2012), 129145. https://doi.org/10.25162/gz-2012-0016

[4] Garrett LE, Spreitzer GM, Bacevice PA, Co-constructing a sense of community at work: the emergence of community in coworking spaces, *Organization Studies*, **38** (2017), 821842. https://doi.org/10.1177/0170840616685354

[5] Howell. T, Coworking spaces: An overview and research agenda, *Research Policy*, **51** (2022), no. 2, 104447.
https://doi.org/10.1016/j.respol.2021.104447

[6] Pan, J., Cho, T. Y., Sun, M., Debnath, R., Lonsdale, N., Wilcox, C., & Bardhan, R., Future workspace needs flexibility and diversity: A machine learning-driven behavioural analysis of co-working space, *Plos One*, **18** (2023), no. 10, e0292370. https://doi.org/10.1371/journal.pone.0292370

[7] Berbegal-Mirabent J., What do we know about co-working spaces? Trends and challenges ahead, *Sustainability*, **13** (2021), no. 3, 1416.
https://doi.org/10.3390/su13031416

[8] Kojo, I., & Nenonen S., Typologies for co-working spaces in Finlandwhat and how?, *Facilities*, **34** (2016), no. 5/6, 302-313.
https://doi.org/10.1108/f-08-2014-0066

[9] Tan, T. H., & Lau K., Understanding users and hosts motives to co-working space: Case of Kuala Lumpur, Malaysia, *Open House International*, **46** (2021), no. 1, 81-95. https://doi.org/10.1108/ohi-07-2020-0077

[10] Capdevila I., Knowledge dynamics in localized communities: coworking spaces as microclusters, *SSRN Electronic Journal*, (2013).
https://doi.org/10.2139/ssrn.2414121

[11] Gandini A., The rise of coworking spaces: a literature review, *Ephemera*, **15** (2015), 193205.

[12] Garrett LE, Spreitzer GM, Bacevice PA, Co-constructing a sense of community at work: the emergence of community in coworking spaces, *Organization Studies*, **38** (2017), 821842.
https://doi.org/10.1177/0170840616685354

[13] Neuberg, B. (2014). Blog available at:
http://codinginparadise.org/ebooks/html/blog/ startofcoworking.html

[14] Akhavan, M., Mariotti, I., & Rossi F., The rise of coworking spaces in peripheral and rural areas in Italy, *Territorio*, **97** (2021), no. 2, supplemento, 2021, 35-4.

[15] Mariotti I., Akhavan M., Rossi F., The preferred location of coworking spaces in Italy: an empirical investigation in urban and peripheral areas, *European Planning Studies*, **31** (2021), 467-489.
https://doi.org/10.1080/09654313.2021.1895080

[16] Di Marino M, Rehunen A, Tiitu M, Lapintie K, New working spaces in the Helsinki Metropolitan area: understanding location factors and implications for planning, *European Planning Studies*, **31** (2023), 508527. https://doi.org/10.1080/09654313.2021.1945541

[17] Vogl, T., Sinitsyna, A., & Micek, G., Systematic literature review of location factors of coworking spaces in non-urban areas, *SpringerBriefs in Applied Sciences and Technology*, 2024, 83-94. https://doi.org/10.1007/978-3-031-50868-4_8

[18] Mariotti, I., Akhavan, M., & Rossi F., The preferred location of coworking spaces in Italy: an empirical investigation in urban and peripheral areas, *European Planning Studies*, **31** (2023), no. 3, 467-489. https://doi.org/10.1080/09654313.2021.1895080

[19] Capdevila I., Spatial processes of translation: how coworking diffused from urban to rural environments. The case of Coworking in Catalonia, Spain. In: Hracs B.J., Brydges T., Haisch T., Hauge A., Jansson J., Sjholm J. (eds.), *Culture, Creativity and Economy*, Collaborative practices, value creation and spaces of creativity. London: Routledge, 2021. https://doi.org/10.4324/9781003197065-8

[20] Boutillier S., Le coworking, lempreinte territoriale Essai danalyse dune agglomration industrielle en reconversion, *Revue Interventions conomiques*, **60** (2018), 10-31. https://doi.org/10.4000/interventionseconomiques.4845

[21] Petch Z., *The Urban Planners Guide to Coworking: A Case Study of Toronto, Ontario*, Ryerson University Master of Planning in Urban Development, 2015. https://doi.org/10.32920/ryerson.14649492.v1

[22] Yu R., Burke M., Raad N., Exploring impact of future flexible working model evolution on urban environment, economy and planning, *Journal of Urban Management*, **8** (2019), no. 3, 447-457. https://doi.org/10.1016/j.jum.2019.05.002

[23] Von Neumann, J., & Morgenstern, O., *Theory of Games and Economic Behavior*, 2nd rev., 1947.

[24] Shapley, L. S., *A Value of N-Person Games*, Contributions to the Theory of Games, 1953, 307317.

[25] Sundararajan, M., & Najmi, A., *The Many Shapley Values for Model Explanation*, In International Conference on Machine Learning, 2020, 9269-9278.

[26] Borgonovo, E., Plischke, E., & Rabitti, G. The many Shapley values for explainable artificial intelligence: A sensitivity analysis perspective, *European Journal of Operational Research*, **318** (2024), 911-926. https://doi.org/10.1016/j.ejor.2024.06.023

[27] Lindeman, R., Merenda, P., and Gold R., *Introduction to Bivariate and Multivariate Analysis*, Scott, Foresman, 1980.

[28] Gromping U., Estimators of relative importance in linear regression based on variance decomposition, *The American Statistician*, **61** (2007), no. 2, 139147. https://doi.org/10.1198/000313007x188252

[29] Owen, A., Sobol indices and shapley value, *SIAM/ASA Journal on Uncertainty Quantification*, **2** (2014), no. 1, 245251. https://doi.org/10.1137/130936233

[30] Owen, A. and Prieur, C., On shapley value for measuring importance of dependent inputs, *SIAM/ASA Journal on Uncertainty Quantification*, **5** (2017), no. 1, 9861002. https://doi.org/10.1137/16m1097717

[31] Lundberg, S. and Lee, S.-I. A unified approach to interpreting model predictions. In NIPS, 2017.

[32] Strumbelj, E., Kononenko, I., and Sikonja, M. R., Explaining instance classifications with interactions of subsets of feature values, *Data & Knowledge Engineering*, **68** (2009), no. 10, 886904. https://doi.org/10.1016/j.datak.2009.01.004

[33] Strumbelj, E. and Kononenko, I., Explaining prediction models and individual predictions with feature contributions, *Knowledge and Information Systems*, **41** (2014), no. 3, 647665. https://doi.org/10.1007/s10115-013-0679-x

[34] Lundberg, S. M., Erion, G. G., and Lee, S., Consistent individualized feature attribution for tree ensembles, 2018, CoRR,abs/1802.03888.

[35] Datta, A., Sen, S., and Zick, Y., Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems, *In 2016 IEEE Symposium on Security and Privacy (SP)*, (2016), 598617. https://doi.org/10.1109/sp.2016.42

[36] Pan, J., Cho, T. Y., Sun, M., Debnath, R., Lonsdale, N., Wilcox, C., & Bardhan, R., Future workspace needs flexibility and diversity: A machine learning-driven behavioural analysis of co-working space, *Plos One*, **18** (2023), no. 10, e0292370. https://doi.org/10.1371/journal.pone.0292370

[37] Waters-Lynch, J., & Potts, J., The social economy of coworking spaces: a focal point model of coordination, *Review of Social Economy*, **75** (2017), no. 4, 417-433. https://doi.org/10.1080/00346764.2016.1269938

Received: August 19, 2024; Published: October 1, 2024