

Decomposition Analysis of Index of Multiple Deprivation (IMD) Based on Shapley Value

Candidate: Xiaohan Feng

Submission Date: 30/08/2020

Module Name: MSc Dissertation

Module ID: CASA0012

Supervisor: Dr. Huanfa Chen

Word Count: 10359

This dissertation is submitted in part requirement for the MSc in the Centre for Advanced Spatial Analysis, Bartlett Faculty of the Build Environment, UCL.

Abstract

Many government programs consider IMD a reliable technique to quantify the geographical variation of deprivation and use it as guidance to allocate resources. However, the subjectivity in the settings of the calculation process and the intrinsic flaw of the data may lower the reliability of IMD. Additionally, IMD is derived from the composition of seven domain scores. It allocates the fixed weight from the seven domains for each LSOA area and this may induce us to think that the importance of the seven fields is in the same pattern among all these areas. However, our study found that the contributions of seven deprivation domain scores to the IMDs in England vary in LSOAs, which indicates each domain has different deprivation levels among England. It is clearer after we divide the areas into 4 clusters by performing the K-Means clustering on the Shapley Value of all the LSOAs. The four clusters are distinctively disparate. Cluster 2 contains the least deprived areas with the greatest number of LSOAs (10295), while cluster 3 is the cluster of the most deprived areas in England and it covers the least land areas and has 9258 LSOAs. In contrast with cluster 2 whose contribution in all domains is negative, all the domains have a positive Shapley value in cluster 3 and after comparing the proportion of each domain's contribution from table 5 with the official weight (the benchmark), we know that the proportion of Income, Employment, Edu and Health are all larger than the benchmark, and their contribution accounts for more than 95% of the overall deprivation. Cluster 1 is the most conspicuous one on the map since it has the widest land area, bizarrely, it has the fewest number of LSOA areas. It's deprived severely in the housing and service domain. The proportion of this domain's contribution is 0.202, which is larger than its official weight 0.093 and is even larger than the contribution for Income (For other clusters, Income and Employment are the top two contributors to IMD). To cope with it, it is urgent to build more affordable housing with a suitable living environment. For cluster 0, the percentage of the Crime domain is 0.2. It is more than twice as much as the official weight of 0.093. Comparing with the clusters get by performing K-Means on domain score, although they have a similar data structure in four clusters, clusters for SHAP have a comparatively even composition, which means, each cluster corresponds to a similar number of IMD deciles. On the contrary, the clustering result for domain scores is mainly composed of two clusters, cluster 2 and 3.

Keywords: IMD, geographical variation, Shapley value

Declaration of Authorship

I, Xiaohan Feng, hereby declare that this dissertation is entirely my original work and that all sources have been acknowledged. Besides, this dissertation is 10,359 words in length.

Signed: Xiaohan Feng

Date: 30/8/2021

Contents

1.	Introduction	1
2.	Research question	1
3.	Literature Review	2
3.1.	IMD	2
3.1.1.	The definition of IMD	2
3.1.2.	The function of IMD	2
3.1.3.	The limitation of IMD	3
3.2.	Shapley Value	4
3.2.1.	Shapley Value in geography / GIS / urban research	5
4.	Methodology	6
4.1.	Shapley Value	6
4.2.	SHAP (SHapley Additive exPlanations) and KernelSHAP	7
4.2.1.	KernelSHAP	8
4.3.	K-Mean Clustering	10
4.3.1.	K-Means Algorithm: Expectation–Maximization	11
4.3.2.	Selection for K	11
5.	Experiment and Result	12
5.1.	Data presentation	12
5.2.	Data preprocessing	12
5.3.	Result	13
5.3.1.	Shapley value result	13
5.3.2.	Clustering analysis	18
6.	Discussion	25
7.	Conclusion and Limitation	27
8.	Appendix	28
9.	Reference	29

List of Figures

Figure 1 Sample 1 of force plot.....	14
Figure 2 Sample 2 of force plot.....	14
Figure 3 Scatter plot of Income domain.....	15
Figure 4 Scatter plot of Employment domain	15
Figure 5 Scatter plot of Edu domain	15
Figure 6 Scatter plot of Live domain	15
Figure 7 Scatter plot of House domain	15
Figure 8 Scatter plot of Health domain.....	16
Figure 9 Scatter plot of Crime domain.....	16
Figure 10 summary plot of the global importance	16
Figure 11 Summary plot contains more information	18
Figure 12 SSE elbow plot for SHAP.....	18
Figure 13 Line plots of Silhouette Coefficient for SHAP	18
Figure 14 Distribution of clusters from SHAP.....	19
Figure 15 Heat map of Shapley value.....	20
Figure 16 SSE elbow plot for score	21
Figure 17 Silhouette Coefficient line plots for score	21
Figure 18 Distribution of clusters from score	22
Figure 19 Heat map of score.....	23
Figure 20 Radar plot of the clusters for SHAP	24
Figure 21 Radar plot of the clusters for Decile	24
Figure 22 Radar plot of the clusters for score	24
Figure 23 Distribution of IMD's 10 deciles.....	25

List of Tables

Table 1 variable description	9
Table 2 data description	12
Table 3 average importance of each feature.....	17
Table 4 Shapley value of cluster center.....	20
Table 5 The proportion of each domain in 4 clusters	21

List of acronyms and abbreviations

IMD: Index of Multiple Deprivation

LSOAs: Lower-layer Super Output Areas

Income: Income Deprivation

Employment: Employment Deprivation

Edu: Education, Skills and Training Deprivation

Health: Health Deprivation and Disability

Crime: Crime

House: Barriers to Housing and Services

Live: Living Environment Deprivation Domain

Acknowledgments

It is not an easy time in the months of dissertation writing, not only because of the worldwide COVID 19 and the natural disasters that happen in China but also the health problem on my own. However, I received a lot of support and assistance from following people, and I cannot get through it without them.

Firstly, I would express sincere gratitude to my supervisor, Prof. huanfa Chen, for his valuable guidance and feedback throughout this research. Besides, his patience and gentleness encouraged me every time when I met the problem and let me know I am not alone.

Secondly, I want to thank my father Zhengwei Feng and mother Xueyan Wu. They give me meticulous care and education throughout my life. As they are also my teachers in the middle school, I hope they will be proud of me when they see the article.

Lastly, I need to thank my boyfriend, Junjie Li. I was inspired many times from the discussion with him and the academic articles he recommends are worthful. In addition, he warms and cheers me up during my hard time.

1. Introduction

Many government programs consider IMD a reliable technique to quantify the geographical variation of deprivation and use it as a guide to allocate resources. The Oxford University team pointed out that about 1% of all government spending refers to the IMD (OCSI, 2011). Thus, it's necessary to have a cautious and precise inspection of the mechanism of IMD. Additionally, IMD is derived from the composition of seven domain scores. It allocates the fixed weight from the seven domains for each LSOA area and this may induce us to think that the importance of the seven fields is in the same pattern among all these areas (like the income deprivation would always play 22.5% of the importance of the overall deprivation in all small areas). However, just like the government needs different strategies to tackle the same set of problems according to each district's local condition. The main contribution to deprivation in seven domains is also likely to vary in different areas. Moreover, knowing this variation is significant, because realizing how deprived each area is only can give the policymaker an idea about fund distribution among areas. Nevertheless, which domains are worth funding, how to better improve the overall situation of a certain region would need the information from each domain.

But how can we get the exact contribution for each deprivation domain score to the final IMD score, given that the calculation is complex (standardized the domain scores by ranking and then transforming to a specific exponential distribution based on their ranks)? Shapley value as an explanation method can help to handle it.

To sum up, this article aims to evaluate the contributions of the seven domains to the final deprivation of each area using Shapley value and compare the difference over areas. After ascribing the effects to the specific seven domains, we can propose the targeted policies to different areas according to the deprivation of the specific domains.

2. Research question

Are there variations of the contributions of seven deprivation domain scores to the Index of Multiple Deprivation (IMD) across LSOAs (Lower-layer Super Output Areas) in England?

Is there any clustering pattern of LSOAs by the deprivation domain scores or domain contributions?

3. Literature Review

3.1. IMD

3.1.1. The definition of IMD

It is an interesting area to quantify the spatial variation of social and economic circumstances of different areas. One of the most famous attempts is the English government's IMD, which measures deprivation in England locally. It dissects deprivation to evaluate social welfare and is considered as a combination of seven social and economic aspects by the English government. Because of that, seven domain measurements of deprivation are developed, including Income Deprivation; Employment Deprivation; Education, Skills and Training Deprivation; Health Deprivation and Disability; Crime; Barriers to Housing and Services; Living Environment Deprivation, and each of them has a bunch of sub-domains to help integrate the concept (Ministry of Housing Communities and Local Government, 2019). Nowadays the main domains of IMD are similar across the world. For example, the IMD of New Zealand is also composed of seven domains that are similar to the composition of IMD in England (Exeter *et al.*, 2017). The IMD of Germany has five domains similar to that of England. The additional special two are municipal revenue deprivation and social capital deprivation (Maier, 2017). In fact, the theoretical basis of determining deprivation is taken from Townsend's work in the 1980s, in which he defined the deprivation as lacking "diet, clothing, housing, household facilities, and fuel and environmental, educational, working and social conditions, activities and facilities" (Noble *et al.*, (2006) quoting Townsend, p. 172). To sum up, IMD is the index that uses the census and other government data to describe the situation of deprivation geographically.

3.1.2. The function of IMD

Since the data to calculate IMD are all authoritative data, it is widely accepted and used in many fields. Government can use IMD to help target resources to "priority areas". In other words, it is a useful tool to guide resource distribution around the country. For example, it was ever used to help the central government of England to determine the eligible amount of Neighborhood Renewal Fund monies to the local authorities (Cabinet Office, 2001). Some government frameworks for regeneration and funds allocation documents also made explicit references to IMD (DCLG, 2008; DCLG, 2009). Additionally, IMD is broadly used as a key indicator to identify the need or deprivation of local people within academic circles. Experts used it to propose more targeted policy suggestions or develop more realistic theorems (Bull and Jones, 2006; Kintrea, 2007; Macintyre, Macdonald and Ellaway, 2008).

3.1.3. The limitation of IMD

Given the fact that IMD is more and more widely employed in many fields, it is necessary to understand the limitation of it.

a) Limitation about multiple indexes

IMD is the combination of multiple domains, thus, it shows an overview of the local deprivation. However, on the other hand, it may also be difficult to eliminate the measurement error and aggregation irrationality of so many domains of data. As an aggregated index, it suffers that "deficits in some sectors, which actually threaten the health of the whole system" (Bossel, 1999). No matter which indexes of domains or sub-domains are wrong, they will all affect the final effectiveness of the final IMD. The relationship of the sub-index inside IMD may also cause the problem of estimate the deprivation. For example, poor health may affect employment opportunities and income, which further affect the Barriers to Housing and Services Domain (Briggs, Abellan and Fecht, 2008). In other words, some domains cannot eliminate the influence of others. As a result, the combination of the domains, IMD, may pay more attention to certain aspects of deprivation, which is wrong from the initial weight settings by the experts.

b) Limitation about subjectivity

What's more, due to lots of subjective intervene, IMD always raises concerns about paternalism. For example, the allocation of weights to different domains, the choice of indicators in each domain, and many other mechanisms are all determined by the experts' judgments (Watson *et al.*, 2019); the exponential transformation on the domains is done without explanation and implication (Deas *et al.*, 2003); the process to construct the multiple deprivations lacks the general theory support (Clelland and Hill, 2019). We cannot guarantee all these settings proposed by the experts are all suitable to current society, and it is always dilemmatic to judge the effectiveness of the subjective judgments of experts. Therefore, the validity of the results and the utility in the policy decisions may face challenges.

c) Limitation about data

Some of the individual indicators used in the computation of the local IMD were collected at a district or even national scale. Without using the exact ward data, the estimated values may distort the accuracy of the local IMD (Deas *et al.*, 2003). In addition, some indicators are double-counted across more than one domain. Such as Severe Disablement Allowance contributes to not only the Employment Deprivation Domain but also the Health Deprivation and Disability Domain. Though some experts argued that it is legitimate to count the indicators as two deprivations (University of Oxford, 2000), it still raises concerns about the veracity especially when we do the overall factor analysis of all the indicators of IMD (Deas *et al.*, 2003). Besides, some indicators which are used to calculate the domain scores are based on census data

years ago. It means they are not very recent data for the area and as a result, they cannot reflect the current circumstances and may cause measurement error.

d) Limitation about comparison

The IMD scores are only utilized to rank the deprivation among areas, but it is not a numeric measurement of the deprivation (OCSI, 2011). It cannot tell people how much one area is more deprived than the other since the score has no quantitative meaning. In addition, we cannot compare the deprivation situation of one area over time. The IMD score of one area will change relative to other areas, thus it cannot tell us how much one area develops these years. The decrease of the score of one area may be due to the degeneration of its social and economic environment, but also maybe because of the fast development of other areas. Thus, we can only consider deprivation as an order of the deprivation level among all the areas.

3.2. Shapley Value

Shapley Value aims to figure out the reasonably expected payoff for each player in a cooperative game. One of the most famous problems solved by Shapley Value is the cost-sharing problem (Gul, 1989; Pérez-Castrillo and Wettstein, 2001; CHUN, HU and YEH, 2017). For example, Siano, Gallo and Glielmo (2015) calculated the distribution of costs of the shared travel for each driver and passenger, based on the application of the Shapley value. He presented a demand and supply managing algorithm for the shared transportation system, and thus the participants of the system are in a cooperative game, sharing the benefit from the transferable utility and superadditivity of the system. The Shapley value is used to calculate the marginal cost that each participant adds to the coalition system. Also in the transportation field, Lu *et al.* (2010) designed an urban railway ticket pricing mechanism by using Shapley Value. In his model, the government, Metro Corporation, and passengers are the three participants of the cooperative game. By calculating the Shapley Value, the model can produce price adjustment advice in different service stages based on the expected profit distribution of the participants, which can ease the conflict of the participants' different interest requirements.

Along with the development of Machine Learning models, Shapley Value is popular to be applied to helping understand the intrinsic influence factors of Machine Learning models (Lundberg and Lee, 2017; Lundberg, Erion and Lee, 2018), since Machine Learning models are always the "black box" which conceals the importance of the factors for us. Smith and Alvarez (2021) studied the mortality of patients with COVID19. He used Shapley Value to interpret the results of a series of machine learning models, such as Naive Bayes, Logistic Regression, LightGBM, XGBoost, and so on. By this method, he found that ages, days in the hospital and some other factors are the robust predictors in the models. Besides, the Shapley Value can provide the marginal impact of each mortality factor on a case-by-case patient level, which is very helpful to detect anomalous patterns when treating patients.

Contributing back to the cooperative game theorems, Shapley Value is also widely used to discuss the time-consistent cases. Petrosjan and Zaccour (2003) researched the time-consistent Shapley Value to allocate the cost of pollution reduction for countries. It was proved that using the outcomes of Shapley Value, each country can receive a fair time-consistent cost and the total cost would be lower than the sum of the cost of each country assuming playing a noncooperative game. Reddy, Shevkoplyas and Zaccour (2013) also discussed time-consistent Shapley Value. Their theory decomposes the Shapley Value for dynamic stochastic discrete-time games over time. The Shapley Value results showed that their dynamic allocation method is time consistent and can be widely applied in many fields to help build a long-term cooperation relationship among players.

3.2.1. Shapley Value in geography / GIS / urban research

a) The regional policy suggestion of the CO2 emissions by Shapley value

Many experts in China using the Shapley Value decomposition method to analyze the key factors of CO2 emissions in different areas (Yu, Wei and Wang, 2014; Liang *et al.*, 2018). By this method, we can see more detailly about influenced factors of the CO2 emissions in geographic level. Li et al. (2016a) used the aggregate data from the World Input-Output Database to study the main drivers behind energy-related CO2 emission for agricultural sectors across eighteen European countries. By Index Decomposition Analysis (IDA) facilitated by the Shapley Index, he found decreasing energy intensity is the main way to decline the CO2 emission. Yan et al. (2018) used the same method as Li and studied the provincial CO2 emissions from China's thermal electricity generation. The Shapley Value results showed that economic activity is the main factor to push up the CO2 emissions in about 30 provincial power grids in China. Zhang, Wang and Dac (2014) compared the entropy and Shapley Value methods in allocation the carbon quota to different regions in China. He found the result of the entropy method has a positive effect on the task of the Shapley value method. From the final analysis results of Shapley Value, he came up with some targeted policy suggestions for the CO2 emissions reduction in China regionally. Wen and Hao (2020) combined the Shapley Value with Spectral Clustering algorithm to decompose the factors of CO2 emissions at the provincial level in China. He found carbon intensity played a significant role in most provinces, but there are other factors that affect the CO2 emissions differently in different provinces. Based on the PSO-FCM clustering method and Shapley Value, Yu, Wei and Wang (2014) clustered the 30 provinces of China into four classes according to 13 macro factors which may influence CO2 emissions. He proposed three-parts CO2 emissions reduction strategies to be suitably used at the provincial level. By Shapley Value, he also found the main approaches to reduce the CO2 emissions for each class, respectively. The Shapley Value decomposition method can give customized policy suggestions that vary from region to region in the light of local conditions.

b) Other fields using Shapley Value related to urban research

The Shapley value method is also used in many other fields to provide the intrinsic view about the contributing factors to certain problems. Chen and Li (2014) used Shapley Value to decompose the income inequality into five main aspects, including education disparity, household registration, geographic location, type of job, and gender. Geographic location plays the third important role in the analysis. The study paid extra attention to the geographic analysis, since the regional difference may affect other influence factors of income inequality. Aristondo and Onaindia (2020) applied the Shapley Value decomposition method to explore overall poverty change in terms of three poverty components' changes. He found that the incidence, intensity, and inequality should all be taken into account as the poverty measures according to Shapley Value. He applied the theory to 28 European countries and got the regional analysis results for each country. Dong *et al.*, (2020) employed the regression-based Shapley decomposition method to aid the analysis of haze pollution by quantile regression. The population density of regional inequality is found by Shapley Value to be the most important factor of regional differences in haze pollution.

4. Methodology

4.1. Shapley Value

Shapley value is created by Shapley (1953) from cooperative game theory and aims to assign player's payout according to their contribution. It is similar to other explanation methods in some cases. However, these methods are based on the certain assumption (take LIME for example, it assumes that the target black-box model can be locally approximated by some interpretable model like decision trees and linear models) and do not have the theory to support it; while Shapley value has the solid theory as it is the only method that satisfies the properties of a fair payout: Efficiency, Symmetry, Dummy, and Additivity (Molnar, 2021). Following is a detailed explanation of these properties.

a) Background

We assume there are total n people in collaboration and they create a profit of $v(N)$. What need to do is fairly distribute this profit.

$N = \{x_1, x_2, \dots, x_n\}$ and x_i is the vector of feature values of the instance (player i).

S is a subset of the N . $v(S)$ represents the value generated by the cooperation of the elements in the S . The final Shapley value for player i is $\psi_i(N, v)$.

b) Properties

1. **Efficiency:** the full yield of the game is distributed to the players, which means the sum of the Shapley value among players is equal to the total value. $\sum_{i \in N} \psi_i(N, v) = v(N)$

2. **Symmetry:** i and j are interchangeable relative to v if the contributions of them are equal to all possible coalitions. For all S that contains neither i nor j , if $v(S \cup \{x_i\}) = v(S \cup \{x_j\})$, then $\psi_i(N, v) = \psi_j(N, v)$. That is to say, interchangeable (equivalent) players receive the same amount of payments.
3. **Dummy:** player i is dummy if his contribution to any coalition is zero. That would mean, $v(S \cup \{x_i\}) = v(S)$. For all S , if i is dummy, he should have 0 Shapley value (receive nothing), $\psi_i(N, v) = 0$.
4. **Additivity:** if a game can be separated into two parts $v = v_1 + v_2$, then the distributed gains from value function v should correspond to the gains derived from v_1 and the gains derived from v_2 . For every coalition S and player i , $\psi_i(N, v_1 + v_2) = \psi_i(N, v_1) + \psi_i(N, v_2)$, where the game $(N, v_1 + v_2)$ is defined by $(v_1 + v_2)(S) = v_1(S) + v_2(S)$.

This is a prime requisite if a researcher intends to design an evaluation scheme that would be applied to the "systems of interdependent games" (Kuhn, 1997).

The Shapley value is proved by Shapley (1953) to be the only map from the set of all games to payoff vectors that satisfies all four properties.

$$\psi_i(N, v) = \sum_{S \subseteq N \setminus \{x_i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} (v(S \cup \{x_i\}) - v(S)) \quad (1)$$

From the formula above, we can get that the Shapley value of a feature value is its average marginal contribution over all possible coalitions to the payout

4.2. SHAP (SHapley Additive exPlanations) and KernelSHAP

If the model is simple, the best explanation model is itself; while if the original model is hard to understand, an explanation model, which always pertains to the class of "additive feature attribution methods" will be used to interpret it (Lundberg and Lee, 2017).

We assume the original model is f and the explanation model is g . Then, the model g which belongs to the "additive feature attribution methods" will satisfy the following formula:

$$g(z') = \phi_0 + \sum_{l=1}^M \phi_l z'_l \quad (2)$$

Where $z' \in \{0, 1\}^M$, $\phi_l \in R$ and M is the number of features.

SHAP was first proposed by Lundberg and Lee (2017) and its innovation is that it can represent Shapley value explanation as to the above formula. In SHAP, z' would be considered as the coalition vector where an entry of 0 means the feature value it corresponds to is "absent" and an entry of 1 would be "present". Thus, for x (the instance of interest), all the features are "playing" which means it is a vector of 1's. In this case,

the formula can be simplified as:

$$g(x') = \phi_0 + \sum_{i=1}^M \phi_i \quad (3)$$

Since SHAP computes Shapley values, it also satisfies the four properties mentioned above. Besides, SHAP has three properties desired to get the single unique solution just like other models that belong to the class of additive feature attribution methods.

Properties

1. **Local accuracy:** the output of the function f we are seeking to explain is equal to the sum of all feature attributions. It seems to be very similar to the “efficiency” property of Shapley value and actually can be derived from it.

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^M \phi_i x'_i \quad (4)$$

2. **Missingness:** a missing feature (such that $z'_i = 0$) attributed no importance. In theory, features that are already missing could have any arbitrary Shapley value and would not hurt other properties as they multiples by $z'_i = 0$.

$$z'_i = 0 \Rightarrow \phi_i = 0$$

3. **Consistency:** changing a model so a feature has a larger or the same impact on the model, its Shapley value will also increase or keep constant. What’s more, this property is an indication of properties for Shapley: Linearity, Symmetry and Dummy.

For any two models f and f' that satisfy:

$$f'_x(z') - f'_x(z'_i = 0) \geq f_x(z') - f_x(z'_i = 0)$$

For all $z' \in \{0, 1\}^M$ and $f_x(z') = f(h_x(z'))$, then:

$$\phi_i(f', x) \geq \phi_i(f, x)$$

4.2.1. KernelSHAP

For an instance, x , KernelSHAP estimates each feature value’s contribution to the prediction. There are 5 steps in this process:

1. Create K random sampled coalitions $z'_k \in \{0, 1\}^M$, $k \in \{1, \dots, K\}$. If the instance has four features ($M=4$), a random coalition might be like this (1, 1, 0, 0). It means we have a coalition of the first and second features as 1 means the feature present in coalition and 0 means feature absent.

2. Get predictions for each z'_k by firstly using the function $h_x(z') = z$ ($h_x: \{0,1\}^M \rightarrow \mathbb{R}^p$) to map z'_k to the original feature space (convert 1's to the corresponding value of instance x and 0's to the mean value of the corresponding feature in research) and then applying model f : $f(h_x(z'))$

Model f: The input for our f is seven domain scores and the output is the rank¹ of IMD. For each score, we need to find its rank in its domain and then transformed these ranks to a specified exponential distribution. Then, the transformed domain scores are combined using the following domain weights in table 1 to get the IMD scores and after rank the IMD scores in ascending order, we get our output. The reason we choose IMD rank instead of the IMD score as our output is that the official website where we get the data advises the user to use ranks instead of the scores.

The transformed domain score X is given by:

$$X = -23 \ln \left(1 - R(1 - \exp^{-100/23}) \right)$$

Where $R = \frac{r}{N}$, N is the number of LSOAs in England (32844), r is the rank for a certain domain score (if a score is the largest one in its domain, it's also the most deprived one, then $r=N$ and $R=1$).

The contents of the following table come from Ministry of Housing Communities and Local Government (2019). The first column is the short names of our input variables, the second one is the domain that our variables belong to. Next to it is the description of corresponding domains. The last column lists their official weight.

Table 1 variable description

Input variable	Deprivation Domain	Description	Weight
income_score	Income Deprivation Domain	Domain measures the proportion of the population experiencing deprivation relating to low income.	22.5%
employment_score	Employment Deprivation Domain	Domain measures the proportion of the working-age population in an area involuntarily excluded from the labor market.	22.5%
edu_score	Education, Skills and Training Deprivation Domain	Domain measures the lack of attainment and skills in the local population.	13.5%
health_score	Health Deprivation and	Domain measures the risk of premature death and the	13.5%

¹ The rank in this article refers to the rank that is get by ranking the scores in an ascending order. It is different with the rank of the official IMD rank, 1 is least deprived here.

	Disability Domain	impairment of quality of life through poor physical or mental health.	
crime_score	Crime Domain	measures the risk of personal and material victimization at the local level.	9.3%
house_score	Barriers to Housing and Services Domain	Domain measures the physical and financial accessibility of housing and local services.	9.3%
live_score	Living Environment Deprivation Domain	Domain measures the quality of the local environment.	9.3%

3. Compute the weight for each z'_k with the SHAP kernel. The mechanism is that for coalitions that contain few 1's or many 1's, they would get the larger weights. For example, the weight for coalitions (1, 1, 1, 0) is the same as (0, 0, 0, 1) and they have larger weight than (0, 0, 1, 1). It is easy to understand since we can learn more about individual features if we can observe their effects in isolation. However, we can learn little of the contribution of an individual feature if a coalition consists of almost half of the features, because there would be many possible coalitions. Based on it, Lundberg and Lee (2017) propose the SHAP kernel:

$$\pi_x(z') = \frac{(M-1)}{\binom{M}{|z'|} |z'| (M - |z'|)} \quad (5)$$

Where M is the maximum coalition size and $|z'|$ is the number of present features in instance z' and M is the maximum coalition size.

4. Fit the linear regression model in (2) with the kernel weight by optimizing the loss function as follows and get the coefficients of the linear model ϕ_k , which is the Shapley value we required.

$$L(f, g, \pi_x) = \sum_{z' \in Z} [f(h_x(z')) - g(z')]^2 \pi_x(z') \quad (6)$$

Where Z is the training data

4.3. K-Mean Clustering

K-Means is one of the most popular "clustering" algorithms. It searches within a multidimensional dataset that is unlabeled for a pre-determined number of clusters. A point will be classified to a particular cluster if it is closer to that cluster's center than other clusters' centers.

4.3.1. K-Means Algorithm: Expectation–Maximization

Given a training set $x^{(1)}, x^{(2)}, \dots, x^{(m)}$, the expectation-maximization approach here consists of the following procedure:

1. Initialize cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in R^n$ randomly
2. Repeat until converged
 - a) E-Step: assign each point $x^{(i)}$ to the nearest cluster center μ_j

$$c^{(i)} := \arg \min_j ||x^{(i)} - \mu_j||^2 \quad (7)$$

- b) M-Step: for each cluster $c^{(j)}$, set the cluster centers to the arithmetic mean of all the points in the cluster.

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}} \quad (8)$$

4.3.2. Selection for K

a) Elbow method

The Elbow method is commonly utilized in the K-means algorithm to find the optimal number of clusters through fitting the model with a range of values for k. It requires us to plot a line chart between SSE and clusters' numbers. Based on the plot, we need to find the "elbow point" that after which the SSE starts to decrease linearly.

SSE is the sum of squares of distance errors between the points in the cluster and the central point, The smaller the SSE, the better the clustering result.

$$SSE = \sum_{i=1}^k \sum_{j \in C_k} (x^{(i)} - \mu_j)^2 \quad (9)$$

b) Silhouette Coefficient

It is used to assess the quality of K-Means clusters fit on the data. This measure has a range of $[-1, 1]$. A value that is close to 1 is what we want, it indicates that the sample is far away from the neighboring clusters. 0 means that the sample is on the decision boundary between two neighboring clusters and a negative value indicates that the sample might have been assigned to the wrong cluster. Its calculation needs two values:

1. $a(i)$: The mean distance between the observation $x^{(i)}$ and all other data points in the same cluster $c^{(i)}$. This distance can also be called a mean intra-cluster distance.
2. $b(i)$: The mean nearest-cluster distance. It is the Mean distance between the observation and all other data points of the next nearest cluster.

The Silhouette Coefficient then can be calculated by:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (10)$$

5. Experiment and Result

5.1. Data presentation

The data we need is from “English indices of deprivation 2019”² by the Ministry of Housing, Communities & Local Government. It contains both the seven domain scores and their corresponding rank. As we want to get the Shapley value for seven domain scores, these variables are what we will focus on. We calculated their minimum/ maximum and mean, and listed them in table 2. We can see from the table that the income_score and employment_score are on the same scale (between 0 and 1) and their mean is around 0.1; health_score and crime_score have approximately 0 mean and range from -4 to 4; edu_score’s, house_score’s, and live_score’s mean are about 21.7 and their range is (0, 100).

Table 2 data description

	income_score	employment_score	edu_score	health_score	crime_score	house_score	live_score
mean	0.128	0.100	21.691	0.000	0.000	21.691	21.691
min	0.003	0.002	0.013	-3.215	-3.459	0.483	0.126
max	0.609	0.534	99.446	3.547	3.350	70.456	91.602

5.2. Data preprocessing

To get each domain’s contribution to the IMD scores, we need to implement the **model f** shown in the methodology part. However, the official data only provide seven domain scores with 3 decimals, which means there could be more than two hundred LSOAs that have the same domain scores. This would bring huge errors when we transform the scores to the ranks. Thus, we need to preprocess the scores with the help of the domain rank from the official dataset (domain rank from the official dataset is different from the rank we use. It considers the most deprived one as 1, which is the one with the highest score. Therefore, we would use $N - \text{domain rank from the official dataset} + 1$ as the rank r we use).

² The dataset could be downloaded in the official website in the following link:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/845345/File_7_-_All_IoD2019_Scores_Ranks_Deciles_and_Population_Denominators_3.csv

The method here is simple, if there are many LSOAs have the same score, denote a as the lowest rank for these scores and r is the rank of the score we would modify. The modified score would be the original score plus $(r - a) \times 10^{-6}$. The reason we choose 10^{-6} is because the largest number of the same score is within 300.

5.3. Result

5.3.1. Shapley value result

The SHAP values explain the margin output of the model, which is the change in the rank of the IMD. As there are more than 30 thousand LSOAs, it's not possible to visualize every small area's Shapley value in seven domains. Thus, we will look into it three ways: Firstly, we will choose two areas to see their force plot, so that we can know the possible difference between LSOAs in the structure of the Shapley value for seven domains; Secondly, we would use a scatter plot to show the effect a single feature has on the predictions made by the model; Finally, SHAP summary plot will help the interpretation of the prediction of all samples and make the comparison between seven domains.

a) Illustrating SHAP values using two LSOA samples

In the force plot, the color blue indicates a negative contribution and red means a positive contribution. Besides, the number beneath the color bar is the original value of the domain score. For the domain that is closer to the position of $f(x)$, its length is longer and at the same time, its contribution to IMD is also larger. In figure 1, the sample used is called "Barking and Dagenham 014A" whose LSOA code is E01000103, we can see that the primary factor for overall multiple deprivations is employment. The next most powerful indicator is income_score. Moreover, house_score and crime_score contribute least, and except these two domains that have a positive effect on the output, other domains contribute negatively to the overall IMD rank in this area. Thus, the sum of their Shapley value is just about 5879, which is much less than the base value of 18150. Let's turn to figure 2, It shows the contribution of 7 domains in an LSOA (with code E01006095) of Trafford. Most of the domains' color is red, which means nearly all of them can push the prediction higher (to the right) and the summation of their Shapley value is around 26138. However, the structure of this observation's SHAP value is not like the first one whose structure is similar to the official weight distributed to seven domains. For example, the most influential indicator is house_score, but it is one of the domains with the least weight: 9.3% (see table 1). These two figures indicate the existence of geographical variation for the influence and contribution of seven domain scores to the IMD rank.

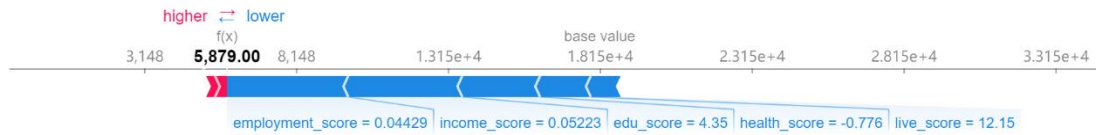


Figure 1 Sample 1 of force plot



Figure 2 Sample 2 of force plot

b) Illustrating the relationship between domain scores and their corresponding SHAP values

SHAP scatter plots represent how the model output varies by feature value and whether the relationship between the target and the features is linear, monotonic, or more complex. Each dot is a single prediction (LSOA) from the dataset. The x-axis is the value of domain scores and the y-axis is the SHAP value for that feature.

The following scatter plots from figure 3 to figure 9 give us a clearer relationship between individual domain scores and their corresponding Shapley values. The linear line in each plot describes the scenario in which all domain scores are linearly combined. In general, these scatter plots have a similar tendency: it's super-linear at the beginning, which means IMD rank increases significantly when domain scores are small. Then it becomes linear and finally tends to be sublinear. In addition, for points whose domain scores are lower than their average value, it corresponds to a negative Shapely value. It means the marginal effect this variable has on the predicted IMD rank is negative. While a higher-than-average domain score pushes the prediction higher. In addition, the vertical dispersion in the plot is only small when the domain score is around the average value, for other places, the same value for domain score can have a different impact on the model's output for different LSOAs. Then, we would inspect each domain separately to see their difference.

For Figures 3 and 4, only the very beginning part (just above 0) has super-linearity. When income_score is larger than 0.3 and employment_score is larger than 0.2, it becomes sublinear and tends to be flat.

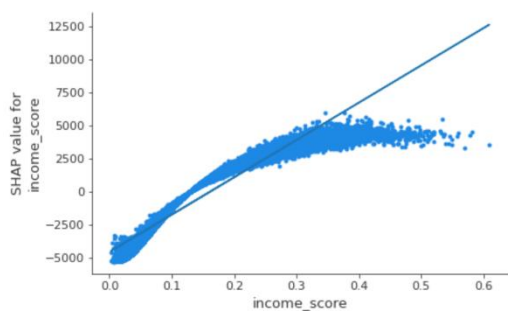


Figure 3 Scatter plot of Income domain

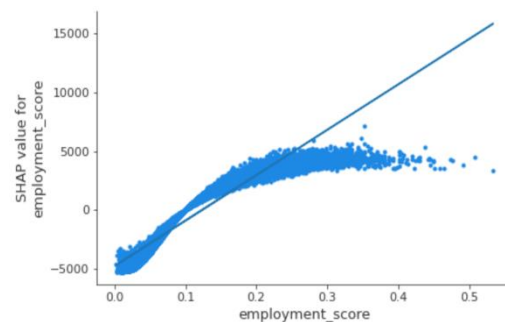


Figure 4 Scatter plot of Employment domain

As for edu_score, house_score, and live_score, their sublinear and superlinear parts are not obvious and are more like a linear line, which means if they increase a certain value, their importance will grow proportionally. However, their vertical dispersion is large, so their prediction would not be stable and easier to be affected by other indicators.

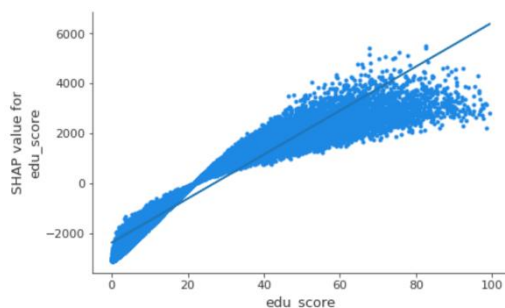


Figure 5 Scatter plot of Edu domain

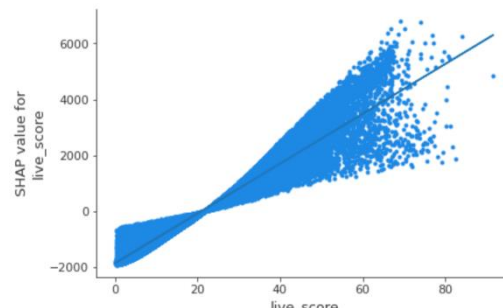


Figure 6 Scatter plot of Live domain

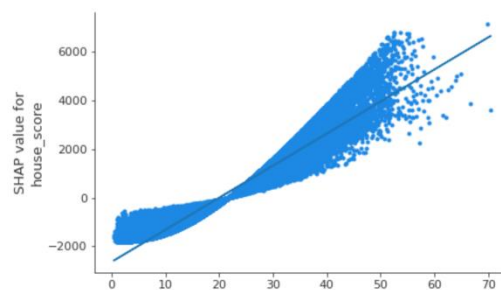


Figure 7 Scatter plot of House domain

The super-linear part of health_score (in figure 8) and crime_score (in figure 9) is larger than other plots, which is more than half of the total length. Additionally, points in this part are more concentrated. When the value of scores is after one, they turn to be more discrete and linear.

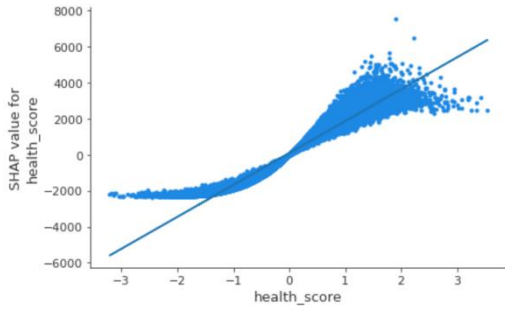


Figure 8 Scatter plot of Health domain

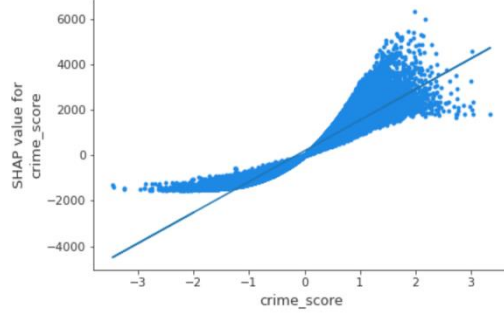


Figure 9 Scatter plot of Crime domain

c) Illustrating the overall contribution for seven domains

There are two kinds of summary plots. One is the typical bar chart of feature importance, which is obtained by taking the average absolute value of Shapley values of each feature and is actually the global importance I .

$$I_j = \sum_{i=1}^n |\phi_j| \quad (11)$$

From figure 10, we can roughly see that the importance of Employment Score, Income Score are the highest and they are almost the same; the Education, Skills, and Training Score's and the Health Deprivation and Disability Score's importance are the third and fourth one. The next two are the house_score and live_score, their importances are very similar, and the last one is the crime_score.

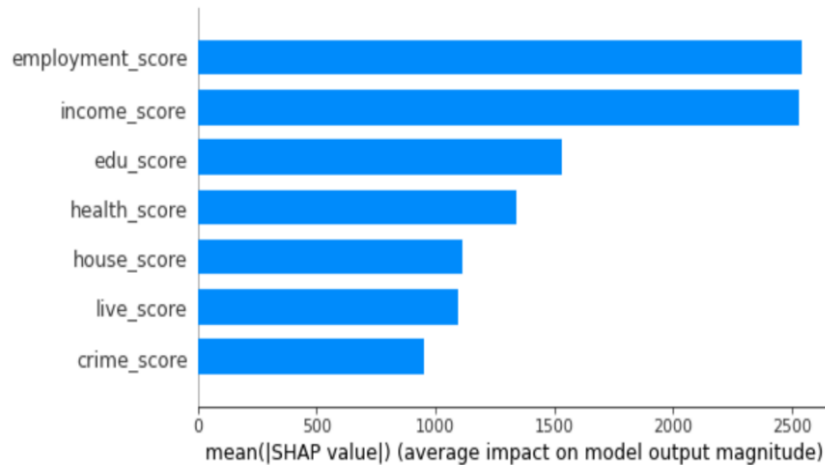


Figure 10 summary plot of the global importance

To get the specific importance of each feature, we use formula 11 to get the result and calculated their proportion (listed in table 3). Recall that the IMD combines information from the seven domains using the weights we have mentioned to produce an overall

relative measure of deprivation. We also list these weights on the rightmost column to make a comparison with the proportion of each domain's importance as follows. Although there are some differences between the calculated proportion and the weight mentioned above: for the weight derived from importance, health_score is nearly 1.5% lower than the official weight; the live_score and house_score domains have about 0.7% difference with the official one; Others are slightly higher than official weight and the difference is within 0.5%. On the whole, the overall structures between the proportion we calculated and the official weight are similar, which means the average contribution of each domain conforms to the official weight.

Table 3 average importance of each feature

feature	importance	proportion	official weight
employment_score	2542.109	0.228	22.5%
income_score	2532.332	0.229	22.5%
edu_score	1532.545	0.138	13.5%
health_score	1342.020	0.121	13.5%
live_score	1110.486	0.086	9.3%
house_score	1093.447	0.100	9.3%
crime_score	954.190	0.098	9.3%

For the second kind of summary plot, which is figure 11, the row of it results from projecting the points of a SHAP scatter plot onto the y-axis, then recoloring it by the feature value. This plot has loaded information and the biggest difference between this plot and the regular variable importance plot is that the color attribute allows us to match how changes in the value of a feature affect the change in IMD rank and it also indicates the relationships of the predictors with the target variable.

As the variables are ranked by the feature importance in descending order, their sequence is the same as in figure 10. Besides, it's interesting to get from figure 11 that the range of effects over the dataset in all domains behaves similarly. The effect of a high feature value is associated with a higher and positive prediction and when the feature values are low, they tend to contribute a negative effect to the final IMD score. Additionally, when feature values are low, they are more gathered especially for the last three domains, which also indicates that the change of the value for these features will not obviously influence the Shapley value. This can also be verified by the above scatter plot: when domain scores are below their average value, points are more gathered and the slope is relatively small. A similar phenomenon could be found for the first two domains when feature values are high.

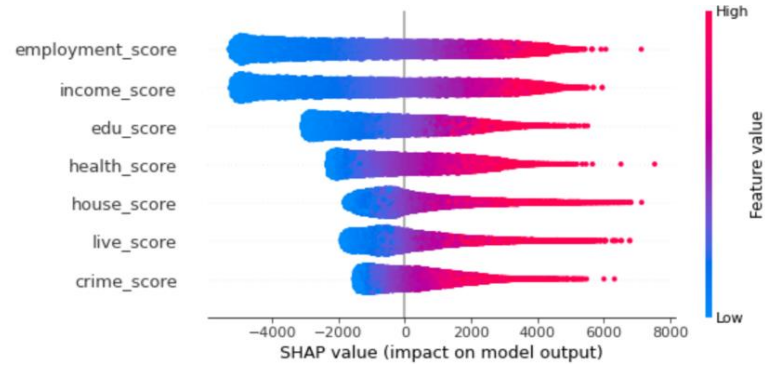


Figure 11 Summary plot contains more information

5.3.2. Clustering analysis

a) Clustering analysis for Shapley value

We have realized the existence of the variation for Shapley value in seven domains by the example of certain areas shown in force plot and non-linear pattern in scatter plot. However, it's still unclear how varied between 32844 areas. A good way is to divide the whole LSOAs into several parts by clustering and use the average Shapley value of each domain to represent each part's circumstance. K-Means Algorithm is used here to cluster the Shapley value for seven domains of LSOAs. Firstly, we need to find a suitable K. Elbow Criterion and Silhouette Coefficient is commonly used to evaluate K-Means clustering and choose K. These two indicators are plotted in figures 12 and 13 where the x-axis is k and y-axis are the value of two indicators. For the left plot, which is figure 12, it seems that after K = 4, the line lean to be linear. For figure 13, K = 4 is also acceptable although K = 2 is the best one. Combining these two figures, we decide that the K for Shapley value clustering is 4 as we might put more weight on the result of SSE.

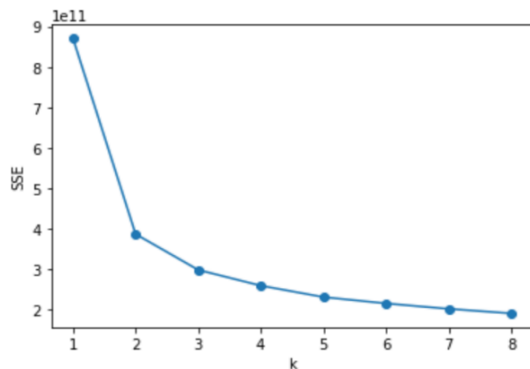


Figure 12 SSE elbow plot for SHAP

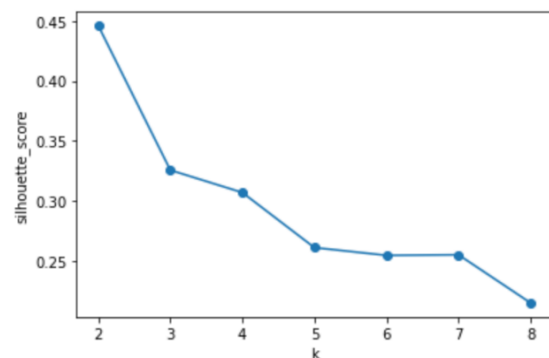


Figure 13 Line plots of Silhouette Coefficient for SHAP

Then, we project our clustering result to the map in figure 14. It seems that most of the areas in England are purple which is cluster 1. Cluster 2 takes second and they seem to be radially distributed about London. The concentrations of the areas in cluster 0 are

located in the midland and the northeast. For LSOAs in cluster 3, it mainly disperses between the yellow areas. Next, let's observe London alone. Surprisingly, the proportion of areas in cluster 1 becomes the lowest, and cluster 0 turns to be the highest. Cluster 2 areas are prone to spread near the London border. What's more, areas for cluster 3 are still randomly scattered in yellow areas in London.

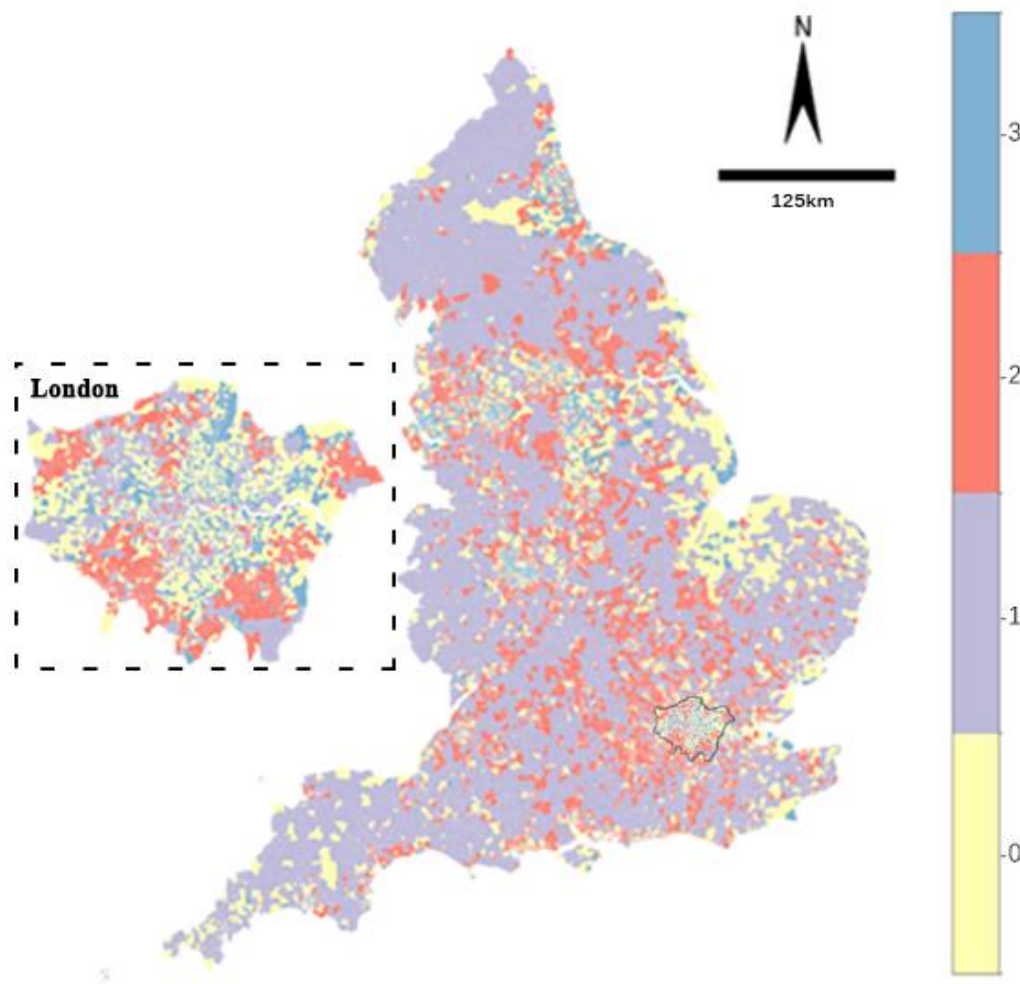


Figure 14 Distribution of clusters from SHAP

After knowing how each cluster distributed, we will see what is the composition of every cluster. Table 4 contains the cluster centers for seven domains in four clusters. If a domain's Shapley value is positive, it would have a positive contribution to the rank of IMD, which means the overall deprivation level of these areas will also be higher. In cluster 2, all the Shapley value of domains' center is negative, while for cluster 3, all of them are positive. Furthermore, except for the Shapley value for House Domain and live Domain, other domains are negative in cluster 1, and the domains in the first three columns are also negative in cluster 0. When adding up all the Shapley values in seven domains, we can get the overall influence (if we add the total Shapley value with the base value, we can get the average rank of these areas). Thus, from the total Shapley value in table 4, areas in cluster 3 are the most deprived, and areas in cluster 2 are the least deprived among the four clusters.

Table 4 Shapley value of cluster center

cluster	income_ SHAP	employment_ SHAP	edu_ SHAP	health_ SHAP	crime_ SHAP	house_ SHAP	live_ SHAP	total_ SHAP
0	-467	-457	-190	368	438	46	228	-34
1	-2519	-2926	-1559	-1079	-189	2531	1731	-4009
2	-3890	-3804	-2019	-1278	-621	-371	-842	-12826
3	2593	2623	1462	1820	1015	101	217	9831

For better comparison, a heat plot in figure 15 is created from the data in the above table, so next, let's compare each domain's contribution in the same cluster. Although the contribution in all domains is the lowest in cluster two, the house domain still needs to be focused on as it's the highest one in this cluster two. It's also the highest domain for cluster 1, besides, the live domain's Shapley value is also large. Thus, these two domains are the main reason for the deprivation of cluster 1. The color in cluster 0 did not vary a lot. It means that the contribution for the 7 domains is relatively even than other clusters and maybe health and crime domains need to be paid more attention in this domain. The last cluster is the most deprived one, furthermore, the color of income and employment domain is dull-red. Therefore, to solve the overall deprivation of cluster 3, settle the problem from income and employment is crucial.

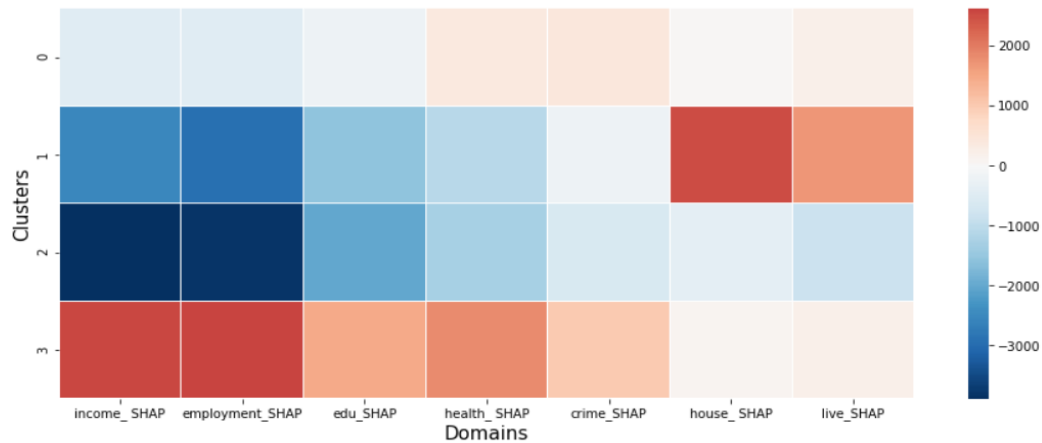


Figure 15 Heat map of Shapley value

We also calculated the proportion of each domain (using each domain's absolute Shapley value divided by the sum of the absolute value in the cluster) based on the data in table 4 to make a comparison to the official weight (listed in table 1). The proportion of cluster 0 and cluster 1 in Income is slightly smaller than the official weight (0.225), while that for clusters 2 and 3 is obviously larger than the official weight. The pattern in Employment is similar with income, except that cluster 1 is a little larger than the benchmark (the official weight for Employment is also 0.225). The proportion for clusters 1, 2 and 3 is relatively

close to 0.135 in Edu domain, while cluster 0 is significantly smaller than the official weight. For Health domain, the disparity between their proportion and the official weight is not small, cluster 1 and 2 is 0.3 smaller than the official weight (0.135), on the contrary, the percentage of cluster 0 and 3 is 0.3 more than 0.135. The official weight for the last three domains: Crime, House, and Live is 0.093. However, the difference between Crime and House is large. Their smallest value can be nearly 0.01 (0.015 for cluster 1's Crime domain and 0.01 for cluster 3's House domain) and the largest one can reach approximately 0.2 (0.2 for cluster 0's Crime domain and 0.202 for cluster 1's House domain). For the Live domain, the proportion for cluster 1 is relatively large and for cluster 3 is obviously small.

Table 5 The proportion of each domain in 4 clusters

cluster	income_ SHAP	employment_ SHAP	edu_ SHAP	health_ SHAP	crime_ SHAP	house_ SHAP	live_ SHAP
0	0.213	0.208	0.087	0.168	0.200	0.021	0.104
1	0.201	0.233	0.124	0.086	0.015	0.202	0.138
2	0.303	0.297	0.157	0.100	0.048	0.029	0.066
3	0.264	0.267	0.149	0.185	0.103	0.010	0.022

b) Clustering analysis for Score

Instead of using the K-mean clustering method on the seven domain scores directly, we standardized the data (so that the mean is their center and component-wise scale to unit variance) to eliminate the influence on the judgment of distance caused by the difference in dimension and order of magnitude. For the selection of K, the "elbow point" also happens when k equals 4. The silhouette coefficient is the largest when k = 2, but is higher in k = 4 than in k = 3. We choose K=4 as our final answer.

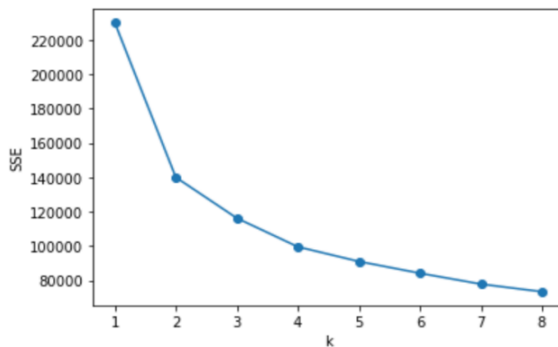


Figure 16 SSE elbow plot for score

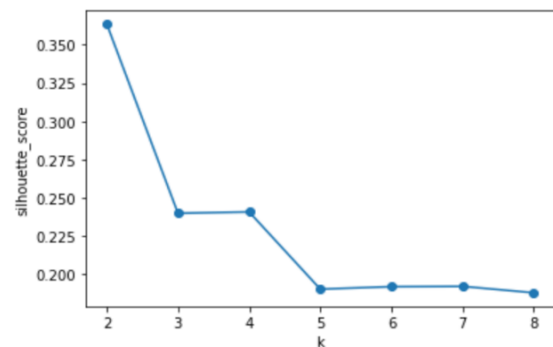


Figure 17 Silhouette Coefficient line plots for score

In figure 18, the major land of the map is red (cluster 2), which pervades all around England. Cluster 3 and cluster 1 are the second and the third largest cluster in acreage on this map, and areas in yellow which represents cluster 0 are the smallest. Take a look at

London alone. The proportional relationship of the four clusters is similar to the whole of England. Cluster 1 and 3 gather around the boundary of London. Other areas are almost all belong to Cluster 2, as areas for cluster 0 are rare and scatter randomly in London.

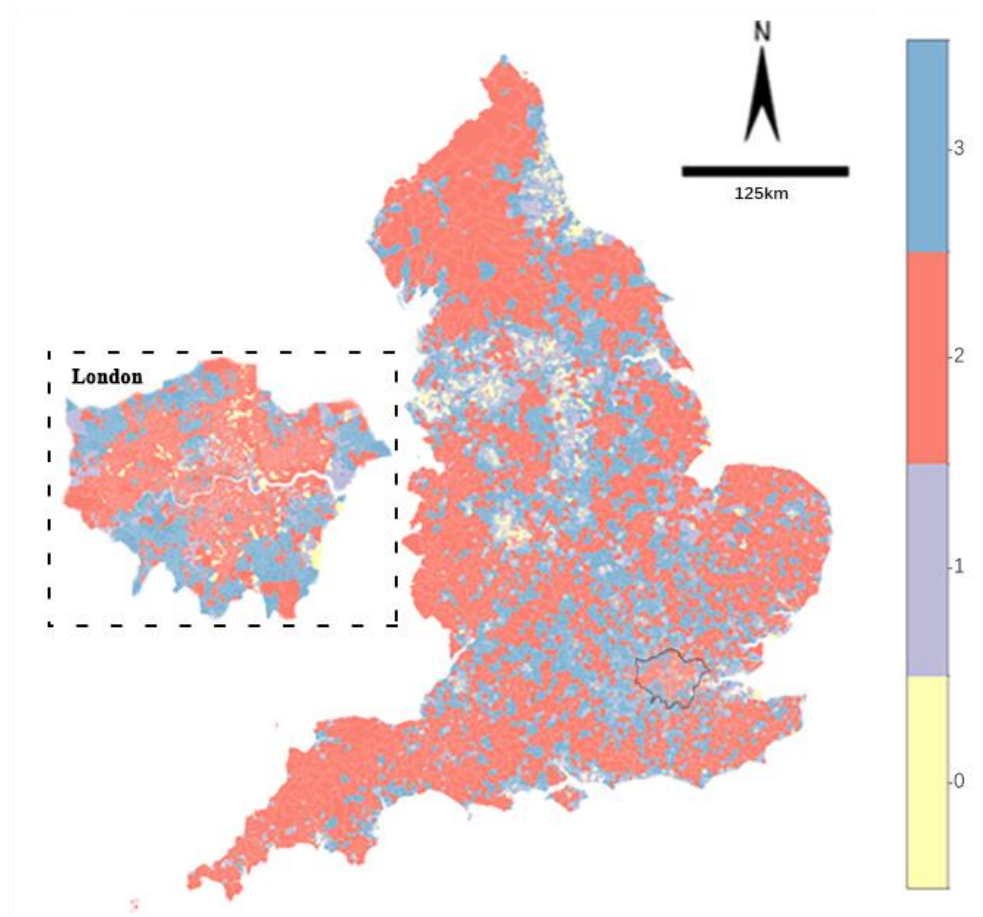


Figure 18 Distribution of clusters from score

The cluster centers of scores are transformed to follow a standard distribution. Thus, it's not necessary to list their number in a table and we just use these data to draw the heat plot in figure 19 directly. In cluster 1, six domains' centers are positive, the only negative one is House Domain. It's the same for cluster 0 but all the scores are higher. Moreover, the color of Income, Employment, and Edu are all dark red while in its corresponding cluster for SHAP, only the first two are of a similar color. It's because the weight for income and employment is higher than edu. In contrast with cluster 0, cluster 3 has the smallest values for all the domains except the House Domain and all of them are negative. Cluster 2 has the highest value in the House domain and other domains are all to the third-highest deprivation scores.

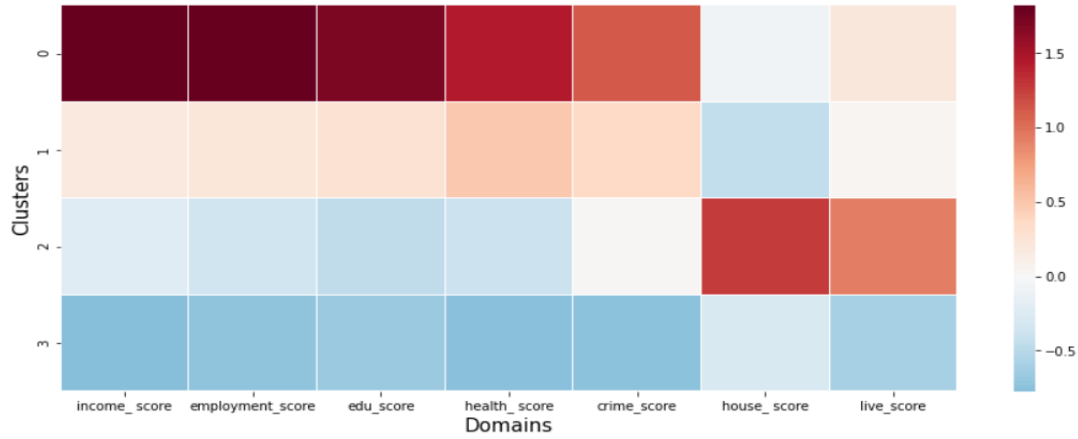


Figure 19 Heat map of score

c) Comparison

To better compare the clustering result between two datasets, we scaled their clustering centers to range 0 and 1 and visualize the result using the radar plot. Moreover, the IMD Decile (where 1 is the most deprived 10% of LSOAs) is also plotted. Figure 20 is the radar plot for Shapley value, Figure 21 is for scores and Figure 22 is for decile. It seems that except for Live and House domains, other domains' value for one cluster has the same rank. For example, cluster 2 in figure 20, cluster 3 in figure 21, and cluster 10 in figure 22 have the smallest value in all 5 domains. As for the deprived level of these clusters, we already know from table 4 that the most deprived to the least deprived SHAP cluster are clusters 3, 0, 1, and 2 based on their contribution to overall IMD. Thus, we infer that the highest to the lowest score clusters are: 0,1,2 and 3 with respect to the deprivation level. Additionally, we find that the overall most deprived area according to IMD has a relatively low score in the house domain.

Compared with the horizontal comparison (based on clusters) in Table 4 and figure 14, we focus on the vertical comparison (based on domains) here as it's more suitable to compare the relative value of the domain when all the domain's Shapley value is scaled to the same range. When observing the relative value of different clusters in the same domain in figure 20, although the House domain is of the highest Shapley value, the Health domain's contribution is close to that of cluster one. Thus, in cluster two, the Health domain might still need to be focused on. On the contrary, the Health domain is not an issue for cluster one in contrast with cluster two. The deprivation of cluster one is mainly from the Live and House domain. Moreover, their relative value for the live domain is more than double the second largest one, and for the house domain, it's five times as much as its second-largest one. The relative value is almost the same for cluster 0 and cluster 3 in the Live domain and the House domain, and in other domains, the relative value for cluster 3 is twice as much as cluster 0.

In figure 21, as the domain scores are also scaled individually, therefore we can eliminate

the effect of official weight and compare the relative value in the same domain for two figures. Cluster 3 is no longer the smallest in all domains, its house domain is higher than cluster 1. For other domains in cluster 1, their relative values are farther away from cluster 0 than the relative distance between cluster 0 and cluster 3 for SHAP. In addition, the health domain in cluster 2 is not the closest to cluster 3. In this case, although we can establish a one-to-one correspondence between SHAP clustering result and score one, their difference within clusters still exists.

Overall, from the three radar plots in Figures 20,21 and 22, we know that the contribution of seven domains has different patterns in four clusters and they also prove that there are variations of the contributions of seven deprivation domain scores to the Index of Multiple Deprivation (IMD) across LSOAs.

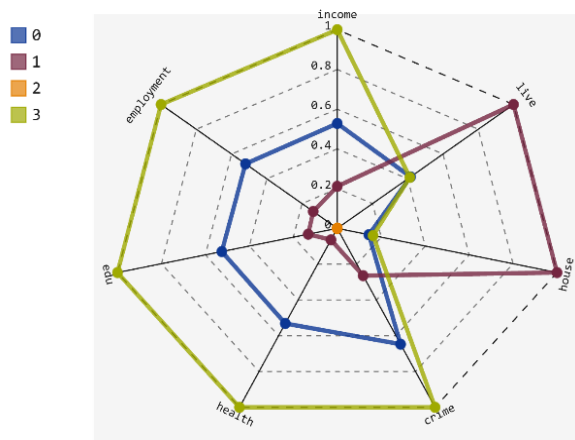


Figure 20 Radar plot of the clusters for SHAP

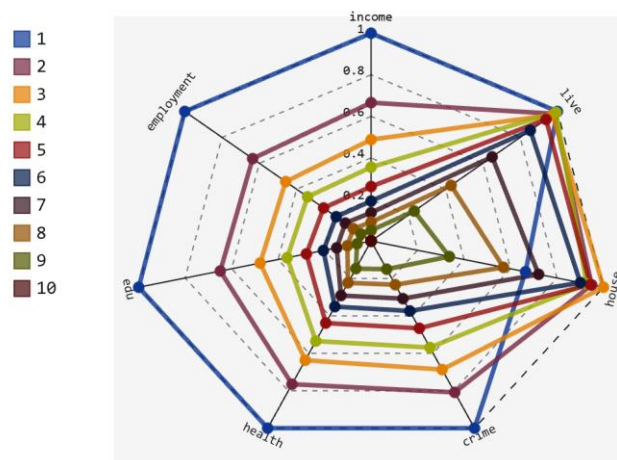


Figure 21 Radar plot of the clusters for Decile

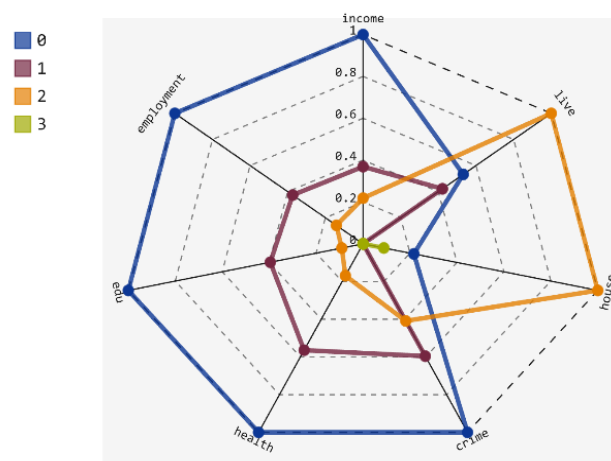


Figure 22 Radar plot of the clusters for score

We also compare figures 14 and 18 with figure 23 which depicts the decile of LSOAs where 1 is most deprived and 10 is least deprived. From the perspective of IMD, the relatively most deprived areas which are colored dark red are concentrated in the middle eastern

and northeast seaboard, and also scattered in the region of north-central and northwest-central in England. The proportion of these areas are higher in London and they disperse mainly in the middle (especially central north and central east) of London. Recall the map get by performing clustering for Shapley values in figure 14, it seems the above areas correspond to the areas for cluster 3. Similarly, the navy-blue areas in the decile map correspond to cluster 2. On the whole, the map for SHAP clustering has a comparatively even composition, which means, each cluster corresponds to a similar number of deciles. On the contrary, the clustering result for domain scores is mainly composed of two colors, red and blue, especially for London.

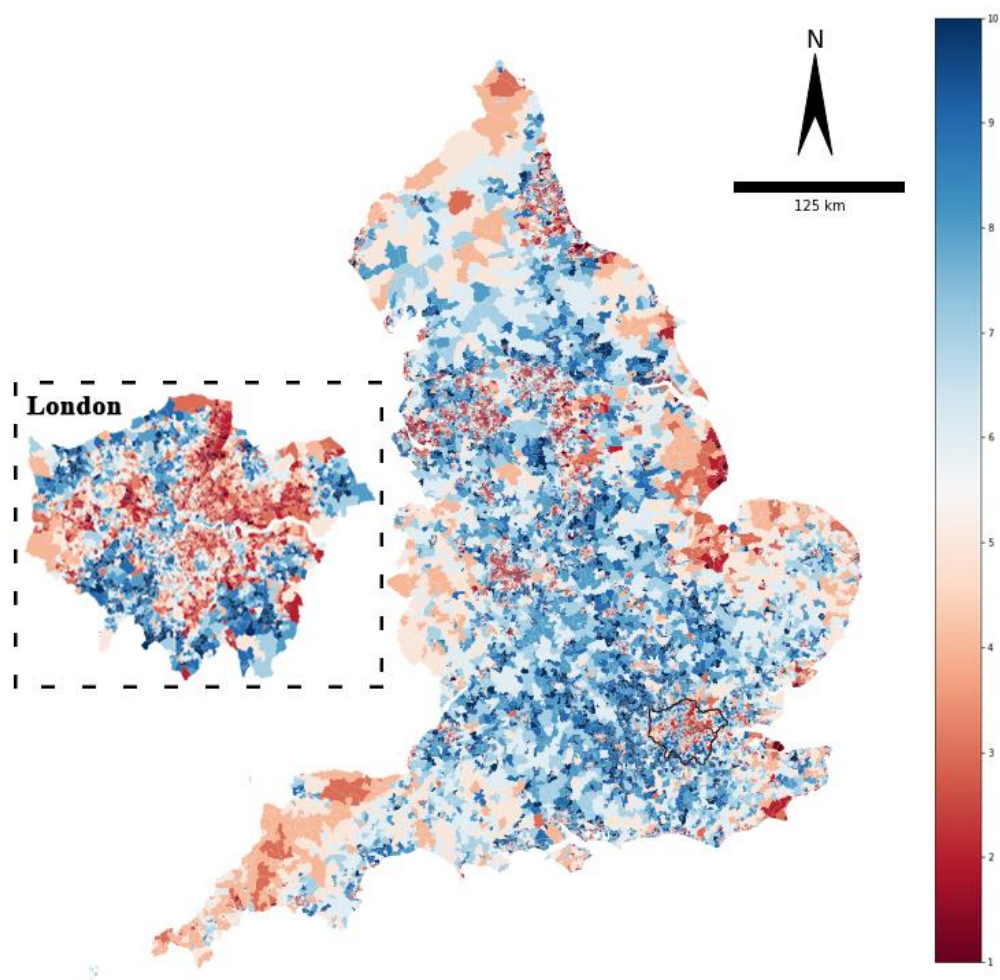


Figure 23 Distribution of IMD's 10 deciles

6. Discussion

From the two examples of the force plot, we find that the contribution of 7 domains varies a lot in these areas. They can have positive or negative effects in different areas. Their absolute contribution can also be the largest in one area but be the smallest in another area. If the contribution in LSOAs is of the same weight for seven domains, the scatter

plot should be like a linear line. However, most of them behave from the superlinear at the beginning to the sublinear at last, which means when the domain scores are below the average, the increase of score's value will speed up the growth of Shapley value, while when the score is large, the increment of the score will not significantly improve its contribution to IMD. Moreover, for lower-than-average points, they are more concentrated but when the scores are high, the vertical dispersion of the scatter plot is large. This might result from non-linear interaction effects in the model between different domains. Just as Briggs et al. (2008) claim that it is hard to eliminate the influence from other domains because these domains are not independent of each other.

From the above exploration, we have confirmed the bias of the seven domains' contribution to the overall IMD in various areas. Then, we do clustering analysis to roughly divide them to see if there is any clustering pattern of LSOAs and how disparate they are. In addition, targeted policy suggestions for each cluster could be given. We can explicitly see from table 4 and figure 14 that the contributions of seven domains vary a lot not only in the magnitude of the Shapley value but the structure of the data for 4 clusters.

Cluster 1 is the most conspicuous cluster in figure 14 since it has the widest land area, bizarrely, it has the fewest number of LSOAs (only contains 3809 LSOAs). Thus, it indicates that the average areas of each LSOAs are large and it might be the reason for the severe deprivation of House domain in cluster 1 as one of the subdomains for Housing domain is barriers to services, which reflects the geographical barrier to some peripheral support facilities like schools, post offices, supermarkets, and GP surgery. The most vulnerable group to this issue is the old tenants who are over 70 years old and at the same time have chronic health problems and decrements in cognitive and physical functioning (Golant, 2003). Lawton (1969) early realizes the problem of unmet supportive service needs for elderly residents and presses for the solution of it. For these frail residents, the financial firewall between the service delivery and the "bricks and mortar" issues should be broken down. Housing programs occupied exclusively by those seniors that integrate the housing and supportive service would be an appropriate alternative. Housing affordability is also the main problem with respect to obstacles to housing, although the subsidy for it comprised more than 3% of UK government spending and become the second-largest expenditure in the aspect of welfare (only less than the state pension) in 2011/2012 (Braakmann and McDonald 2020). For homeless people, a safe and smaller haven that could give them more privacy is an ideal transitional place before they get temporary housing. However, just as Wusinich *et al.*, (2019) claim, creating more affordable housing is the only viable way to end the problem brought by obstacles to housing.

Cluster 2 contains the least deprived areas with the greatest number of LSOAs (10295), while cluster 3 is the cluster of the most deprived areas in England and it covers the least land areas and has 9258 LSOAs. It is likely that areas there are rural areas, as it's deprived in all domains especially in the Income and Employment domain but its deprivation level in the Housing domain is relatively small (this domain only contributes 1% to the overall deprivation maybe because of the lower rental and house prices in rural areas). All the domains have a positive Shapley value in cluster 3 and after

comparing the proportion of each domain's contribution from table 5 with the official weight (the benchmark), we know that the proportion of Income, Employment, Edu and Health are all larger than the benchmark, and their contribution accounts for more than 95% of the overall deprivation. (Rafique and Khawaja, 2020) studied the association between education profile and income in Pakistan. They discovered that the failure to raise 50% to at least primary education level from the uneducated population brings a potential income loss of PKR 251 billion. For the solution to education deprivation, the research of Köthemann (2020) declared that the promotion of preschool participants can buffer the unfavorable social background from the educational deprivation. In addition, skill training for adults is also a vital part of education. Hyde and Stapleton (2017) advise the public support such as technical training and financial support could also go to the older workers who do not reach the retirement age but feel strenuous in keeping up the work (because they have health problems and have difficulty in adapting to new skills) to keep them in the labor force. This measure can also prevent work disincentives and thus reduce unemployment to some extent. After having the educated and skilled human resources, necessary physical capital and infrastructure should be provided to consume them. Otherwise, they would not become the incentives, but a burden for the government (Rafique and Khawaja, 2020). As Income is highly correlated with the Edu and Employment, the average income in cluster 3 would automatically increase if the Edu domain and Employment domain are no longer deprived.

For cluster 0, the percentage of the Crime domain is 0.2. It is more than twice as much as the official weight of 0.093. Besides, the percentage of the Health domain is 0.168, which is also larger than its official weight of 0.135. These indicate that cluster 0 faces relatively severe crime deprivation and at the same time deprived in the health domain. For the crime, Vollaard and van Ours (2011) proved that a large scale of government intervention in executing built-in security regulation can better decrease the crime (especially can lower the burglary risk) than trying to improve the precautionary consciousness through subsidies on security or public campaign. Security pacts are also effective in lowering predatory crimes like robberies and thefts (Calaresu and Triventi, 2019). Improving the Crime problem can also help the improvement of the Health domain. Harrison, Gemmell and Heller (2007) said that feeling unsafe is the potential largest barrier to physical activity on population levels. Thus, reducing crime can improve the perceived feeling of safety and therefore promote physical activity and increase the overall health status there.

7. Conclusion and Limitation

IMD is frequently used to quantify geographic deprivation; however, it has many limitations. The most important one is that they are the combination of multiple sub-indexes, which may face the risks we described previously. The subjectivity of the settings of the calculation process and the intrinsic flaw of the data may also cause bias to the final IMD. People subconsciously assume that each domain contributes the

same percentage in different places because the weight of each domain to form the IMD is fixed. However, our study found that the contributions of seven deprivation domain scores to the IMDs in England vary in LSOAs. It is clearer after we do the clustering and divide the similar small areas into 4 clusters. For clusters derived from SHAP, cluster 0 has the widest land area and it's deprived severely in the housing and service domain and live environment domain. To cope with it, it is urgent to build more affordable housing with a suitable living environment. In addition, the integration of the housing and services programs is a good way for elderly people who are in bad health. As for cluster 3, which is suspected to be rural areas, almost all domains, except for the House domain and Live domain, are very serious. These deprived domains are also correlated, so without intervention, the areas in cluster 3 will always be stuck in a vicious circle. Thus, we need to improve the deprived domains simultaneously using the suggestions we provide, especially the Education Domain and the Unemployment Domain.

There are also many limitations in the article, Firstly, when constructing the model, we use the average values as the references instead of the randomly selected scores in each domain. Though this setting can speed up the computation, it may lower the accuracy. Secondly, we classify the results into 4 clusters, which are subjective and empirical. In the future, we can try other clustering methods to auto-select suitable cluster numbers. Lastly, we only concentrate on the seven domains and do not explore more on the sub-domains or even their indicators. In the future, the subdomains, and the indicators that formed them will be studied in each cluster to see the main contributors to the domain deprivation and in this case, more detailed and suitable suggestions can be given to the policymakers.

8. Appendix

The codes and data used in this project are all in the following GitHub link:

https://github.com/Dora-fxh/CASA_Dissertation_XiaohanFeng

9. Reference

- A *New Commitment to Neighbourhood Renewal*. (2001). London. Available at: [http://www.bris.ac.uk/poverty/downloads/keyofficialdocuments/Neighbourhood Renewal National Strategy Report.pdf](http://www.bris.ac.uk/poverty/downloads/keyofficialdocuments/Neighbourhood%20Renewal%20National%20Strategy%20Report.pdf).
- Aristondo, O. and Onaindia, E. (2020). 'On measuring the sources of changes in poverty using the Shapley method. An application to Europe'. *Fuzzy Sets and Systems*. Elsevier B.V., 383, pp. 80–91. doi: 10.1016/j.fss.2018.12.011.
- Bossel, H. (1999). *Indicators for sustainable development: theory, method, applications*. International Institute for Sustainable Development Winnipeg.
- Braakmann, N. and McDonald, S. (2020). 'Housing subsidies and property prices: Evidence from England'. *Regional Science and Urban Economics*, 80. doi: 10.1016/j.regsciurbeco.2018.06.002.
- Briggs, D., Abellan, J. J. and Fecht, D. (2008). 'Environmental inequity in England: Small area associations between socio-economic status and environmental pollution'. *Social Science and Medicine*. Pergamon, 67 (10), pp. 1612–1629. doi: 10.1016/j.socscimed.2008.06.040.
- Bull, A. C. and Jones, B. (2006). 'Governance and social capital in urban regeneration: A comparison between Bristol and Naples'. *Urban Studies*, 43 (4), pp. 767–786. doi: 10.1080/00420980600597558.
- Calaresu, M. and Triventi, M. (2019). 'Governing by contract as a way to reduce crime? An impact evaluation of the large-scale policy of security pacts', 52, pp. 255–279. doi: 10.1007/s11077-018-9337-1.
- Chen, C. and Li, S. (2014). 'Market transition and income inequality in urban China: Evidence from shapley value decomposition'. *Frontiers of Economics in China*, 9 (2), pp. 309–337. doi: 10.3868/s060-003-014-0016-9.
- CHUN, Y., HU, C.-C. and YEH, C.-H. (2017). 'A Strategic Implementation of the Shapley Value for the Nested Cost-Sharing Problem'. *Journal of Public Economic Theory*, 19 (1), pp. 219–233. doi: 10.1111/jpet.12190.
- Clelland, D. and Hill, C. (2019). 'Deprivation, policy and rurality: The limitations and applications of area-based deprivation indices in Scotland'. *Local Economy: The Journal of the Local Economy Policy Unit*, 34 (1), pp. 33–50. doi: 10.1177/0269094219827893.
- DCLG. (2008). *Transforming places; changing lives: A framework for regeneration*. London.
- Deas, I., Robson, B., Wong, C. and Bradford, M. (2003). 'Measuring neighbourhood deprivation: A critique of the Index of Multiple Deprivation'. *Environment and Planning C: Government and Policy*, 21 (6), pp. 883–903. doi: 10.1068/c0240.
- Dong, F., Yu, B., Pan, Y. and Hua, Y. (2020). 'What contributes to the regional inequality of haze pollution in China? Evidence from quantile regression and Shapley value decomposition'. *Environmental Science and Pollution Research*. Environmental Science and Pollution Research, 27 (14), pp. 17093–17108. doi: 10.1007/s11356-020-07929-8.

- Exeter, D. J., Zhao, J., Crengle, S., Lee, A. and Browne, M. (2017). 'The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand'. *Exeter, D. J., Zhao, J., Crengle, S., Lee, A. and Browne, M. (2017). 'The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand'. PLoS ONE, 12 (8), pp. 1–19. doi: 10.1371/journal.pone.0181260.*
- Golant, S. M. (2003). 'Political and organizational barriers to satisfying low-income U.S. Seniors' need for affordable rental housing with supportive services'. *Journal of Aging and Social Policy*, 15 (4), pp. 21–48. doi: 10.1300/J031v15n04_02.
- Gul, F. (1989). 'Bargaining foundations of Shapley value'. *Econometrica*: JSTOR, pp. 81–95.
- Harrison, R. A., Gemmell, I. and Heller, R. F. (2007). 'The population effect of crime and neighbourhood on physical activity: an analysis of 15 461 adults'. *Journal of Epidemiology & Community Health*. BMJ Publishing Group Ltd, 61 (1), pp. 34–39. doi: 10.1136/JECH.2006.048389.
- Hyde, J. S. and Stapleton, D. C. (2017). 'Using the Health and Retirement Study for Disability Policy Research: A Review'. *Forum for Health Economics and Policy*, 20 (2), pp. 1–12. doi: 10.1515/fhep-2017-0002.
- Kintrea, K. (2007). 'Policies and Programmes for Disadvantaged Neighbourhoods: Recent English Experience'. *Housing Studies*, 22 (2), pp. 261–282. doi: 10.1080/02673030601132920.
- Köthemann, D. (2020). 'How educational systems shape the relationship between social origin and educational : Deprivation cross-national analyses in the first years after finishing secondary school'. *Comparative Sociology*, 19 (2), pp. 279–303. doi: 10.1163/15691330-BJA10013.
- Kuhn, H. W. (1997). *Classics in Game Theory*. Princeton University Press (Frontiers of Economic Research). Available at: <https://books.google.co.uk/books?id=HyTpw6H5syUC>.
- Lawton, M. P. (1969). 'Supportive Services in the Context of the Housing Environment'. *The Gerontologist*, 9 (1), pp. 15–19. doi: 10.1093/geront/9.1.15.
- Li, T., Baležentis, T., Makutėnienė, D., Streimikiene, D. and Kriščiukaitienė, I. (2016). 'Energy-related CO2 emission in European Union agriculture: Driving forces and possibilities for reduction'. *Applied Energy*, 180, pp. 682–694. doi: 10.1016/j.apenergy.2016.08.031.
- Liang, Y., Niu, D., Zhou, W. and Fan, Y. (2018). 'Decomposition Analysis of Carbon Emissions from Energy Consumption in Beijing-Tianjin-Hebei, China: A Weighted-Combination Model Based on Logarithmic Mean Divisia Index and Shapley Value'. *Sustainability*, 10 (7), p. 2535. doi: 10.3390/su10072535.
- Lu, W., Chen, Y., Zhang, N. and Shi, Y. (2010). 'Ticket pricing of urban rail transit based on Shapley value method'. *Proceedings - 2010 2nd IEEE International Conference on Information and Financial Engineering, ICIFE 2010*. IEEE, pp. 127–130. doi: 10.1109/ICIFE.2010.5609267.
- Lundberg, S. M., Erion, G. G. and Lee, S.-I. (2018). 'Consistent individualized feature attribution for tree ensembles'. *arXiv preprint arXiv:1802.03888*.

- Lundberg, S. M. and Lee, S. I. (2017). 'A unified approach to interpreting model predictions'. *Advances in Neural Information Processing Systems*, 2017-Decem (Section 2), pp. 4766–4775.
- Macintyre, S., Macdonald, L. and Ellaway, A. (2008). 'Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland'. *Social Science & Medicine*, 67 (6), pp. 900–914. doi: 10.1016/j.socscimed.2008.05.029.
- Maier, W. (2017). 'Indizes Multipler Deprivation zur Analyse regionaler Gesundheitsunterschiede in Deutschland'. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 60 (12), pp. 1403–1412. doi: 10.1007/s00103-017-2646-2.
- Ministry of Housing Communities and Local Government. (2019). 'The English Indices of Deprivation 2019: Research report', (September), pp. 1–86. Available at: <https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report><https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report><https://assets.publishing.service.gov.uk/government/uploads>.
- Molnar, C. (2021). *5.9 Shapley Values / Interpretable Machine Learning*. Available at: <https://christophm.github.io/interpretable-ml-book/shapley.html#general-idea> (Accessed: 15 June 2021).
- Noble, M., Wright, G., Smith, G. and Dibben, C. (2006). 'Measuring Multiple Deprivation at the Small-Area Level'. *Environment and Planning A: Economy and Space*, 38 (1), pp. 169–185. doi: 10.1068/a37168.
- OCSI. (2011). *Why the Indices of Deprivation are still important in the open data era. Oxford Consultants for Social Inclusion*. Available at: <https://ocsi.uk/2011/03/24/why-the-imd-is-still-important-in-the-open-data-age/> (Accessed: 12 July 2021).
- Pérez-Castrillo, D. and Wettstein, D. (2001). 'Bidding for the surplus: a non-cooperative approach to the Shapley value'. *Journal of economic theory*. Elsevier, 100 (2), pp. 274–294.
- Petrosjan, L. and Zaccour, G. (2003). 'Time-consistent Shapley value allocation of pollution cost reduction'. *Journal of Economic Dynamics and Control*, 27 (3), pp. 381–398. doi: 10.1016/S0165-1889(01)00053-7.
- Rafique, N. and Khawaja, I. (2020). 'Cost of Educational Deprivation: a Case of Pakistan'. *International Journal of Economics and Financial Issues*, 10 (5), pp. 187–202. doi: 10.32479/ijefi.10335.
- Reddy, P. V., Shevkoplyas, E. and Zaccour, G. (2013). 'Time-consistent Shapley value for games played over event trees'. *Automatica*, 49 (6), pp. 1521–1527. doi: 10.1016/j.automatica.2013.02.029.
- Shapley, L. S. (1953). 'A Value for n person Games. Contributions to the Theory of Games'. *Annals of Mathematics Studies*, 28 (2), pp. 307–317.
- Siano, G., Gallo, M. and Glielmo, L. (2015). 'A method for managing transportation requests and subdivision costs in shared mobility systems'. *VEHITS 2015 - Proceedings of the 1st International Conference on Vehicle Technology and Intelligent Transport Systems*, pp. 152–158. doi:

10.5220/0005459501520158.

Smith, M. and Alvarez, F. (2021). 'Identifying mortality factors from Machine Learning using Shapley values – a case of COVID19'. *Expert Systems with Applications*, 176, p. 114832. doi: 10.1016/j.eswa.2021.114832.

Transforming Places; Changing Lives: Taking Forward the Regeneration Framework. (2009). London.

University of Oxford. (2000). *Response to the Formal Consultations on the Indices of Deprivation 2000*.

Vollaard, B. and van Ours, J. C. (2011). 'Does Regulation of Built-in Security Reduce Crime? Evidence from a Natural Experiment'. *Economic Journal*, 121 (552), pp. 485–504. doi: 10.1111/j.1468-0297.2011.02429.x.

Watson, V., Dibben, C., Cox, M., Atherton, I., Sutton, M. and Ryan, M. (2019). 'Testing the Expert Based Weights Used in the UK's Index of Multiple Deprivation (IMD) Against Three Preference-Based Methods'. *Social Indicators Research*. Springer Netherlands, 144 (3), pp. 1055–1074. doi: 10.1007/s11205-018-02054-z.

Wen, L. and Hao, Y. (2020). 'Factor decomposition and clustering analysis of CO2 emissions from China's power industry based on Shapley value'. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*. Taylor & Francis, 00 (00), pp. 1–17. doi: 10.1080/15567036.2020.1776795.

Wusinich, C., Bond, L., Nathanson, A. and Padgett, D. K. (2019). "If you're gonna help me, help me": Barriers to housing among unsheltered homeless adults'. *Evaluation and Program Planning*. Pergamon, 76, p. 101673. doi: 10.1016/J.EVALPROGPLAN.2019.101673.

Yan, Q., Wang, Y., Baležentis, T., Sun, Y. and Streimikiene, D. (2018). 'Energy-Related CO2 Emission in China's Provincial Thermal Electricity Generation: Driving Factors and Possibilities for Abatement'. *Energies*, 11 (5), p. 1096. doi: 10.3390/en11051096.

Yu, S., Wei, Y. M. and Wang, K. (2014). 'Provincial allocation of carbon emission reduction targets in China: An approach based on improved fuzzy cluster and Shapley value decomposition'. *Energy Policy*. Elsevier, 66 (2014), pp. 630–644. doi: 10.1016/j.enpol.2013.11.025.

Zhang, Y. J., Wang, A. D. and Da, Y. Bin. (2014). 'Regional allocation of carbon emission quotas in China: Evidence from the Shapley value method'. *Energy Policy*. Elsevier, 74 (C), pp. 454–464. doi: 10.1016/j.enpol.2014.08.006.