Machine Learning Methods for Clustering Architectural Precedents

Classifying the relationship between building and ground

Abdulrahman Alymani[1], Wassim Jabi[2], Padraig Corcoran[3]
[1,2,3]Cardiff University
[1,2,3]{Alymaniaa|JabiW|CorcoranP}@cardiff.ac.uk

Every time an object is built, it creates a relationship with the ground. Architects have a full responsibility to design the building by taking the ground into consideration. In the field of architecture, using data mining to identify any unusual patterns or emergent architectural trends is a nascent area that has yet to be fully explored. Clustering techniques are an essential tool in this process for organising large datasets. In this paper, we propose a novel proof-of-concept workflow that enables a machine learning computer system to cluster aspects of an architect's building design style with respect to how the buildings in question relate to the ground. The experimental workflow in this paper consists of two stages. In the first stage, we use a database system to collect, organise and store several significant architectural precedents. The second stage examines the most well-known unsupervised learning algorithm clustering techniques which are: K-Means, K-Modes and Gaussian Mixture Models. Our experiments demonstrated that the K-means clustering algorithm method achieves a level of accuracy that is higher than other clustering methods. This research points to the potential of AI in helping designers identify the typological and topological characteristics of architectural solutions and place them within the most relevant architectural canons

Keywords: *Machine Learning, Building and Ground Relationship, Clustering Algorithms, K-means cluster Algorithms*

INTRODUCTION

There are different clear approaches to the way architects connect their buildings to the ground. For example, in 1942, Rudolph Schindler established his Harris House on an existing rock which he used as a foundation. However, King Road House (1922), by the same architect, rests on the earth without adapting to the ground beneath it (Berlanda 2014). In La Tourette church 1956, Le Corbusier implements a strong Integrative relationship with the ground. The church rests on the ground in one part and its slope on the other. By lifting the church off the ground, Le

Corbusier allows the basement to intermediate the different levels (Arnold and Cling 2002). Luigi Snozzi, in most of his hillside houses, generates one solid volume embedded in the ground, while other Sonzzi work looks to rest upon a carpet (Snozzi et al. 1995).

Clustering an architect's style through their approach to the relationship between building and ground allows the categorisation of an extensive database into semantic groups belonging to specific historical periods, building types, and regions. These database groups enable an effective and fast retrieval of this data. Moreover, grouping the architecture and the architect's style has received much attention over the past decade. The first study of clustering and classifying architectural styles attempted to build a machine vision system for classifying windows according to architectural styles (Shalunts, Gayane, Yll Haxhimusa, 2011). A year later, (Shalunts, Haxhimusa and Sablatnig, 2012) and (Shalunts, 2012) offered a similar digital approach based on clustering and utilising local features to classify building facades according to Gothic and Baroque styles. According to (Xu et al., 2014) most of the current architectural style classification algorithms focus on the effective extraction of distinctive local patches or patterns (Berg, Grabler and Malik, 2007) and, (Philbin et al., 2007). The last five years have seen a new development in architectural style classification using deep convolutional neural networks (DCNN). (Obeso et al., 2016) aimed to use DCNN to classify Mexican historical buildings, while (Yoshimura et al., 2019) proposed to apply it during the classification of architectural design.

Previous work has been limited to using an image database to classify and cluster architectural styles. Although these approaches are interesting, they do not allow the machine to cluster the data without supervision or tagging of images. To the best knowledge of the authors, no work has been carried out using unlabelled data (unsupervised) to cluster architectural styles.

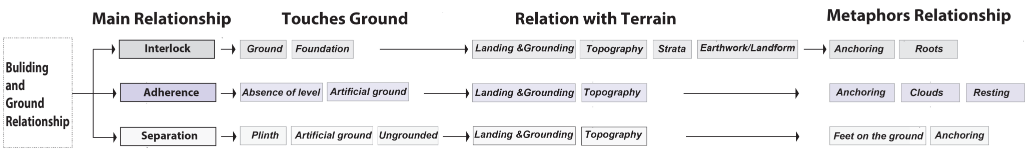
The primary purpose of this paper is to establish an architectural description framework of the figure-ground relationship within various architectural styles. The main concern is to clarify the different architects' building styles and their relationship with the ground. The significance of this research is to help both academic and practising architects understand and identify the typological and topological characteristics of their architectural figure-ground syntax and place it within the most relevant architectural canons. This, in turn, will supply a large amount of knowledge of the various building-ground relationships. This study attempts to use a computational machine learning algorithm to develop a prediction model which helps analyse and understand the building-ground relationship. Cluster machine learning techniques will be employed to create a taxonomy based on discovered similarities between building and ground relationships.

This paper is organised as follows: Firstly, there will be an overview of the building-ground relationships and the taxonomy of these links. Secondly, the related work carried out using the clustering algorithm will be highlighted. Thirdly, there will be an overview of the algorithm that uses and describes the methodology of the K-Means, K-Modes and Gaussian Mixture algorithms. Fourthly, there will be an analysis of the results experiment. Fifthly, the results and limitations will be discussed. The final section will conclude the research and make suggestions for future study.

BUILDING AND GROUND RELATIONSHIP

The building and ground problem has long been discussed in the architecture discipline. Historically, architects have observed the ground as reliable physical and conceptual support for their work. However, recent technological, philosophical and geopolitical changes have improved the notion of connections with the ground (Porter, 2018). Modern architects respond to these conditions by inventing formal topologies, while some modern architects desired steel frameworks. Reinforced concrete is consistently used in providing a physical disengagement with the building form and the ground. For instance, the Convent of La Tourette by Le Corbusier facilitated

Figure 1
Building and Ground Relationship Taxonomy



a surprisingly diverse range of approaches relative to the ground (Samuel, 2013). Other architects deconstructed the architectural objects by connecting and permeating the interior space into surrounding landscapes (Leatherbarrow, 2004). Contemporary architects used similar methods with the ground; some adopted the idea of disregarding it through the use of autonomous objects while others made an effort to the case of the division between landscape and building with megastructures, field conditions, landform building and landscape urbanism (Porter, 2018). However, building and ground problems have long been discussed in our discipline, yet, there is an apparent lack of assessment criteria of how a building touches the ground.

Building and Ground Taxonomy

According to the (Toma Berlanda 2014) lexicon, the current building -ground relationship taxonomy is divided into three main categories: separation; adherence; and interlock. Moreover, the building touches the ground via different categories; grounded; ungrounded; foundation; plinth; artificial ground; and absence of level. Additionally, the building's relationship with the terrain is defined through, topography; landing and grounding; strata and earthwork-landform. Finally, the metaphorical relationship involves feet on the ground, anchoring, roots and clouds. (See Figure1).

RELATED WORK OF MACHINE LEARNING CLUSTERING ALGORITHM METHODS

Clustering plays an essential role in data analysis. Clustering algorithms have been implemented in many problem domains and have continued to develop in a variety of areas for different types of application. This is because not all of these are suitable for all types of applications (Dharmarajan and Velmurugan, 2016). This section describes the related clustering algorithm work carried out by the researcher in a different discipline. However, focusing on this section will be in the clustering algorithm applied to the field of architecture.

(Glaser and Peng, 2003) examined the LiQuID tool, which was used to cluster lighting simulation data. They aimed to reduce large complex sets of photographs by classifying them into representative prototypes. LiQuID and Light Sketch tools help architects create a quick design and decide on the building light quality. However, these methods suffer from several pitfalls. Firstly, the illuminance data is not considered in the classification. Secondly, the visualisation of clusters needs to develop further. Finally, the similarities between larger temporal units need to be addressed.

Experiments examining the resulting population of alternate designs and providing insight into the relationship between architectural features and design performance were conducted by (Chen, Janssen and Schlueter, 2015). The experiments show that it is possible to gain general knowledge by linking architectural features to design performance. There is still considerable ambiguity about this information because it is not easy to rely on the specific groups being compared. Moreover, it does not seem to have any architectural features, which in turn makes it complicated, concluding anything in terms of performance.

(Lee and Lee, 2016) investigated the colour pattern difference between Eastern and Western cultures using a case study of Disneyland Paris and Tokyo Disneyland. The result indicated that the former was using green and bluish colours while the latter featured more red and yellowish colours based on CIELAB colour space values. This paper had some difficulty in obtaining building images due to trees or visitors in the park. Therefore, the result will be improved with higher resolution images.

CLUSTERING ANALYSIS
Generally, cluster analysis is grouping objects, as observations or events, based on the datum found in the data describing the objects or their relationships (Sharma, Bajpai and Litoriya, 2012). Clustering is the task of grouping a set of similar objects into the same group called a cluster. Cluster analysis is a ML task which can be performed using a ML algorithm. The clustering task can be achieved by implementing diverse algorithms that differ widely in their concept and the process of deciding the output.

UNSUPERVISED MACHINE LEARNING (CLUSTERING ALGORITHM)
Machine learning is a field of research and practice related to developing computer programmes that are configured to improve their performance in a given task through the acquisition and processing of incremental data (Mitchell, 1997). The simplest machine learning definition is "learning from data" (Geron, 2019). Using machine learning has high potential. It is excellent for simplifying code to perform better than hand training or writing an extended code of rules. Machine learning is useful for finding a solution when using a traditional approach is unsuccessful. According to the volume and type of supervision, the machine learning algorithm can be classified into different categories: supervised; unsupervised; semi-supervised; and reinforcement learning (Géron, 2017). In unsupervised machine learning algorithms, training data is unlabelled, so the algorithm learns without requiring a teacher.

Centroid Clustering Models
In the centroid model, the concept of similarity is derived from the closest point to the centre of the clusters. Many algorithms use this approach of clusterings, such as k-means, k-medoids and k-modes. These models require the user to input the desired number of clusters, which makes it essential to have prior knowledge of the dataset. Centroid clustering models need to be run iteratively to find the optimal solution. (See Figure 2)

K-Means Cluster Algorithm
K-means is a prevalent machine learning approach for clustering (Hartigan.J.A and Wong M.A., 2001). K-means is an unsupervised machine learning algorithm that is distance-based and uses an intrinsic relationship between data points (Stasiuk and Thomsen, 2014). The initialisation steps of k-means algorithm can be explained as follows: (1) selecting the first division with K clusters; (2) generating a new division by assigning each point to its closest cluster centre; and (3) calculating new cluster centres (Jain and Dubes, 1988). Step (4) reiterates steps (2) and (3) repeatedly until reaching a stable state, in which the data points no longer change between clusters, meaning centroids do not require any recalculation (Stasiuk and Thomsen, 2014).

K-Modes Cluster Algorithm
A similar alternative to k-means is the k-modes clustering algorithm which replaces the "means" of clus-

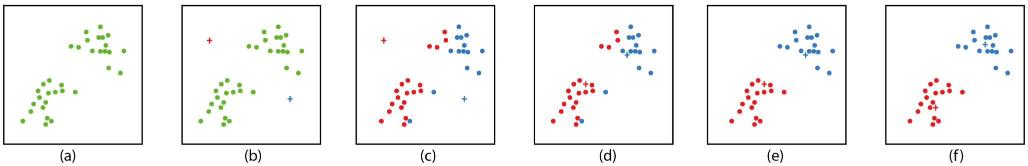


Figure 2
The centroid clustering Algorithm processes

ters with "modes". K-modes is an unsupervised learning option because there are no assumptions about the data. An important k-modes feature is explicitly optimising a "matching" metric, which corresponds to the loss function (Chaturvedi, Fooda and Green, 2001).

Gaussian mixture models (GMM)

Gaussian mixture models (GMM) are an extension of the k-means algorithm model in which clusters are modelled with Gaussian distributions. GMM is used mainly for probability density estimation, which is also known as soft clustering (He et al., 2017). Furthermore, the concept of using GMM is to find clusters that share similar properties, which means data is overlapped (Yambem and Nandakumar, 2016). A GMM consists of several Gaussians, each identified by k ⊠ {1,..., K}, where K is the number of clusters. Each Gaussian k consists of the following parameters: μ mean, which defines its centre; Σ probability, which defines its width; and π, which defines how large the Gaussian function will be (Carrasco, 2019).

METHODS

To understand the architect's approach to the building and ground relationship, several case studies need to be clustered and analysed. A novel proof-of-concept workflow was developed to enable a machine learning computer system to learn how to cluster aspects of an architect's style when designing a building and determine how these buildings relate to the ground. Several clustering algorithm models were evaluated to find the most suitable algorithm that works with this kind of problem.

The data was collected and then archived using Microsoft Access database software. All models were run in a Jupyter Notebook using the Scikit-learn library which is a free machine learning software library for the Python programming language. All experiments were executed on a regular laptop computer running macOS Catalina 10.15.2 operating system with the following configuration: a 2.7 GHz Quad-Core Intel Core i7 with 16 GB memory.

The k-means and k-modes models were chosen for three reasons: (1) a general-purpose form of the problem is required; (2) similar cluster sizes; and (3) a moderate number of clusters. The reason for using Scikit learn library to implement the clustering algorithm is because it provides an easy way to use the interface with the Python language (Pedregosa Fabian et al., 2011). Moreover, GMM was used to see if there is any uncertainty in clustering methods. K-means can provide information about which data point belongs to which cluster, while GMM offers details of the possible groups. K-means can be seen as a case of GMM with equal expectations per component.

Data Collection and Pre-Processing

A total of 139 architectural precedents were recruited for this study. The architectural examples were collected during the initial step. All of the collected building data focused on residential buildings from three specific periods in history: modernism, post-modernism and contemporary.

After the data has been collected, the clustering model was constructed. The first step after reading the excel data file is to translate the non-numerical

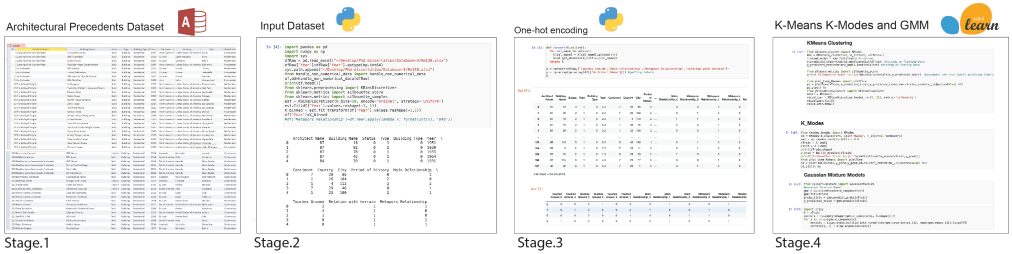


Figure 3
Methods Workflow

data to numerical data to enable the machine to read it (see Table 1). Machine learning algorithms cannot work on label data directly. They require all input and output variables to be numerical data. This means that categorical data must be converted to a numerical data form. There are two steps for converting categorical data to numerical data: integer encoding and one-hot encoding. In this experiment, a one-hot encoding array was used as a way of changing the input data. One-hot encoding is a process of converting a categorical variable into a form that enables the ML algorithm to learn more clearly. In one-hot encoding form, a0 indicates a non-existing category, while 1 means an existing category (see Table 2).

The purpose of using one-hot encoding is that it will allow the model to assume a natural ordering between categories (Brownlee, 2017).

Architect	Touches Ground	Main Relationship	Metaphors Relationship	Relation with Terrain
50	2	1	1	2
50	1	2	1	0
50	0	0	1	0
50	2	5	1	2
46	2	5	1	2
46	2	5	1	2
50	2	1	1	2

Table 1
Noun-numerical Data to Numerical Data

Touches Ground						Main Relationship			Metaphors Relationship					Relation with Terrain			
0	4	5	3	2	1	0	1	2	0	1	2	3	4	0	1	2	3
1	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	0
0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0

Table 2
One Hot Encoding array

Performance Evaluation

Generally, two types of performance evaluation can be used to evaluate the algorithm model. The first, external evaluation, utilises information about the data sets such as normalised mutual information, Rand index and F-measure. Meanwhile, the second, internal evaluation, assesses the data set itself using silhouette score, Davies-Bouldin index, Partition coefficient and others. The silhouette index score was used to measure the separation distance between the resulting clusters. The silhouette score displays how close each data point in one group is to the neighbouring groups. It has a range of [-1, 0, 1], which means the +1 rating indicates this data is far from the adjacent clusters while the score close to -1 implies that the data might have been allocated to the wrong group. In this paper, the coefficient score was used as a parameter to discover the best clustering. Furthermore, the squared Euclidean distance was used to measure the dissimilarity between objects for k-means and k-modes clustering as well as computing the silhouette score.

EXPERIMENTAL RESULTS

Three different unsupervised algorithms were adapted: k-means; k-modes; and Gaussian mixture models (GMM). An attempt was made to cluster the architectural precedent into different groups called "Architect's styles". In the experiment, the training and testing ratio was set to 70 percent training data and 30 percent of the testing data. The only changed hyperparameters were the k number and fitting time.

In this experiment, we used t-SNE methods to demonstrate the clustering results. t-Distributed Stochastic Neighbor Embedding (t-SNE) is used to visualize unsupervised high-dimensional data by projecting each datapoint to a spot in a 2D or 3D-dimensional map (Van Der Maaten and Hinton, 2008). The t-SNE algorithm computes a measure of similarity between pairs of states in high-dimensional space and in low-dimensional space. Then it tries to improve these two similarities using the cost function.

Moreover, we used Elbow methods to determine the optimal number of clusters. Elbow methods is a method that looks at the contrast ratio shown as a function of the number of clusters (Bholowalia and Kumar, 2014). We identified the "elbow criterion" through following the line chart that resembles an arm, then the point of inflection on the curve is an indication to the best model provided at that point (Yellowbrick, 2016).

K-Means and K-Modes experiment results

The k-means algorithm was run with different numbers of clusters (see Figure 4). The best silhouette score related to the appropriate time was found in

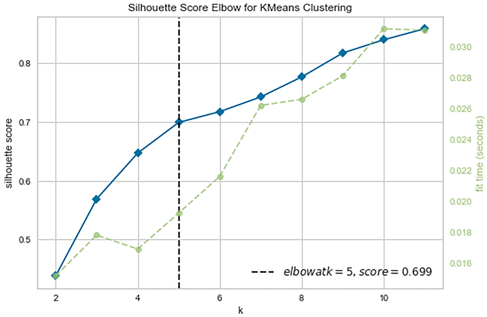


Figure 4
Silhouette score Elbow for K-Means Clustering

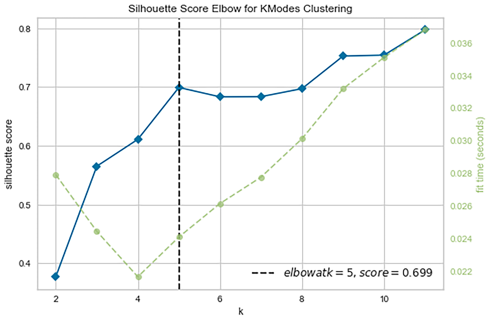


Figure 5
Silhouette score Elbow for K-Modes Clustring

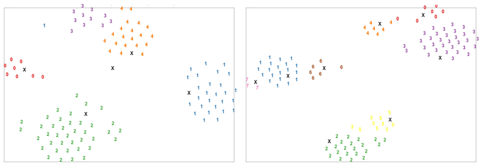


Figure 6
t-SEN K-Means plot for 5 clusters (Left) and 8 clusters (Right)

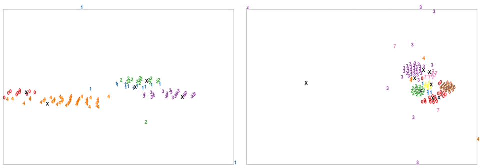


Figure 7
t-SEN K-Modes plot for 5 clusters (Left) and 8 clusters (Right)

five clusters with a rating of 0.69 silhouette score (84.5 percent). In k-means, the time was not affected by the model like the other algorithms; therefore, the best result was found at k=11, which reached 0.87 silhouette score (93.5 percent). However, the purpose of using clustering algorithms is to find a small number of clustering. Therefore, the acceptable number of clusters compared with this piratical kind of problem was found in eight clusters where the silhouette score reached 0.78 silhouette score (89 percent).

In parallel, the k-modes algorithm was run using the same number of clusters (see Figure 5). The best silhouette score was found with five clusters with a rating of 0.69 silhouette score (84.5 percent). This was similar to the k-means algorithm. Furthermore, in a case that was abandoned, the best result was found at 11k, which reached 0.80 silhouette score (90 percent). This result was worse than k-means. Moreover, in the k-modes model, there was no improvement of the score after the 5 k; an increase only appeared in cluster nine with a rating of 0.76 silhouette score (88 percent).

While both clustering algorithms revealed strong accuracy, the k-means method yielded a better overall silhouette score with less fit time.

To validate the results obtained from k-means, a plot t-SNE was used to show the clustering distribution. In the t-SNE graph below, both clusters five and eight show a clear group separation. Therefore, both are good in terms of the partitioning concept (Figure 6). Equally, in the k-modes algorithms, the five-cluster plot demonstrates clearer separation than the eight-cluster model (Figure 7). A comparison between k-means and k-modes shows both models have a clear plot t-SNE, which means both are successful in describing the partitioning idea.

To visualise the results, the testing result was split into groups of images. Architectural precedents and their designers are shown in (Figure 8). Cluster 0 presented the similar architectural styles which involved interlock approaches to the ground. Taking cluster 0 as an example, the case studies were (1) Lovell House for Richard Neutra and (2) Wolfe House for Rudolph Schindler. Both have a similar approach to the ground which is interlock. Meanwhile, in cluster one, the case study involved (1) Mirror Point Cottage for MacKay-Lyons Sweetapple Architects and (2) The 7th Room by Snohetta. Here, both show a similar approach to the ground, which is separation. Moreover, in cluster two the precedent (1) Dana Thomas

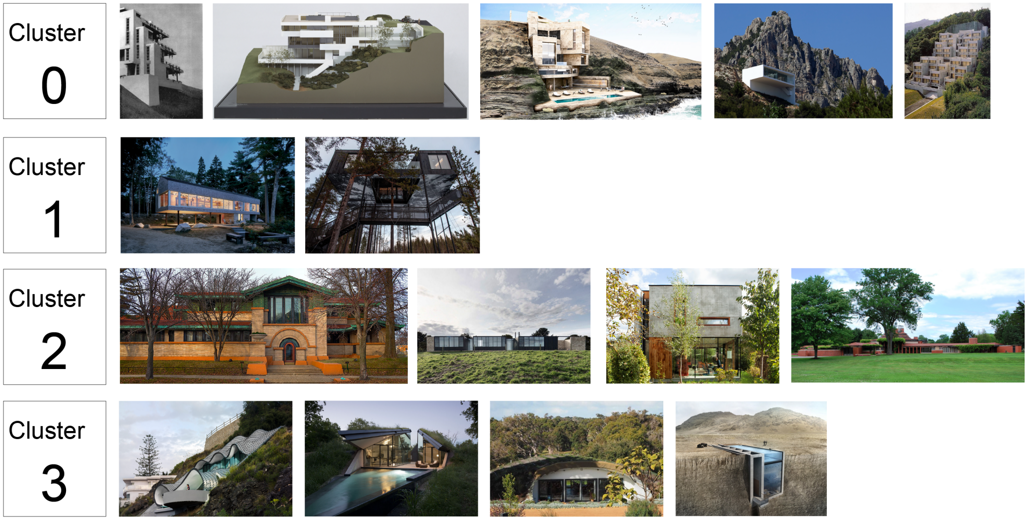


Figure 8
Clusters Architects styles in different groups of Building and Ground Relationship

House for Frank Lloyd Wright has a similar approach to the case study (2) Fish Creek House by Edition Office in designing a building with an adherence relationship to the ground. Finally, in cluster three, all the examples show a similar approach to the ground, which is interlock with more anchoring to go under the ground.

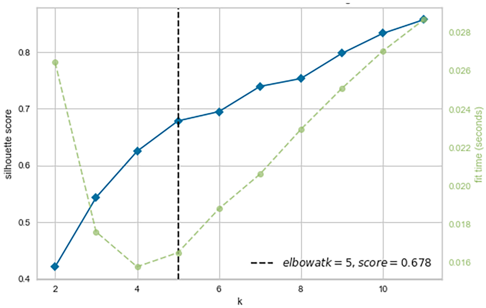


Figure 9
Silhouette score Elbow for GMM Clustering

GMM experiment results

The GMM algorithm was run in a different number of clusters (see Figure 9). The purpose of using GMM was to locate uncertainty in the groups. The t-SNE of GMM experiment shows that there was no overlap between the clusters which means that all the groups were certainly partitioning (see Figure 10). Moreover, it still obtained a good accuracy result such as 0.67 silhouette score (83.5 percent) at cluster five, approximately 0.75 silhouette score (87.5 percent) at cluster eight and approximately 0.87 silhouette score (93.5 percent) at cluster 11.

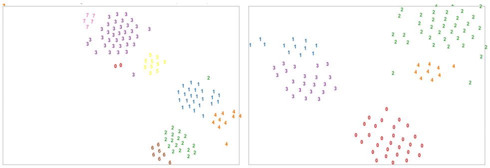


Figure 10
t-SEN GMM plot for 8 clusters (Left) and 5 clusters(Right)

DISCUSSION

Further data collection is required to understand the architects' styles through their approaches to the ground. Also, the paper results are preliminary and could have a bigger sample to test in future studies. According to the results shown in Table 1, all the models reach a high level of accuracy. To have high model performance the optimum number of clusters must be selected. In this case, the ideal k number was eight clusters. The k-means model offers the most consistent accuracy. The purpose of using GMM was to locate uncertainty in the groups. However, in this model GMM does not provide us with this probability, because the data points are well separated into groups. Although the GMM model does not cluster soft assignments, it clusters hard assignments with similarly high accuracy to k-means. Our findings seem to demonstrate that K-Means, K-Modes, and GMM achieved high accuracy on clustering architects' styles. Therefore, our approach would lend itself well for use by researchers to create a taxonomy based on a similarity between architectural precedents.

CONCLUSIONS

This paper has suggested a new proof of concept workflow that enables machine learning to automatically discover how different architects design the relationship between a building and its surrounding ground. We applied three machine learning algorithm models to a collected set of architectural precedents and found that K-means performed best for our type of data. Our investigation into this area is still ongoing and seems likely to confirm our hypothesis of using unlabelled data to cluster building and ground relationships. We have obtained accurate results proving that all ML algorithms cluster the problem with high accuracy. The highest accurate results were found consistently with the K-means algorithm, which is approximately 0.78 silhouette score (89 percent) for 8 clusters. There are many alternative approaches to clustering that can be used, such as K-Medoids, K-Medians, and fuzzy c-means. This paper is part of an ongoing PhD-research which is devoted to the exploration of the relationship between the building and ground. Future work will concentrate on using different data to generalise this approach. We are confident that our research will serve as a basis for future studies on using unsupervised ML algorithms to help the architectural discipline with similar clustering problems.

ACKNOWLEDGMENTS

The authors would like to thank Dr Wesley Aelbrecht, Cardiff University for his help with building and ground historical background and Mr Akhil Meethal, ÉTS Montreal for his help with the python programming environment.

REFERENCES

Berg, Alexander, Grabler, Floraine and Malik, Jitendra 2007 'Parsing images of architectural scenes', *Proceedings of the IEEE International Conference on Computer Vision*

Berlanda, Toma 2014, *Architectural Topographies*, Routledge

Bholowalia, Purnima and Kumar, Arvind 2014, 'EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN', *International Journal of Computer Applications*, 105, pp. 0975-8887

Chaturvedi, Anil K, Fooda, Kraft and Green, Paul 2001, 'K-modes Clustering', *Journal of Classification*, 18, pp. 36-55

Chen, Kian, Janssen, Patrick and Schlueter, Arno 2015 'Analysing Populations of Design Variants Using Clustering and Archetypal Analysis', *Proceedings of the 33rd eCAADe Conference*

Table 3
Comparison between the algorithms

Clustering method	K-Manes			K-Modes			GMM		
Number of K	Silhouette score	% accuracy	Time	Silhouette score	% accuracy	Time	Silhouette score	% accuracy	Time
5	0.69	**84.5%**	0.020sec	0.69	**84.5%**	0.024sec	0.67	**83.5%**	0.017sec
8	0.78	**89%**	0.026sec	0.69	**84.5%**	0.033sec	0.75	**87.5%**	0.023sec
11	0.87	**93.5%**	0.030sec	0.80	**90%**	0.036sec	0.87	**93.5%**	0.028 sec

Dharmarajan, A and Velmurugan, T 2016, 'Efficiency of k-Means and k-Medoids Clustering Algorithms using Lung Cancer Dataset', *International Journal of Data Mining Techniques and Applications*, 150, pp. 150-156

Geron, Aurelien 2019, *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow.*, O'Reilly Media

Glaser, Daniel and Peng, James 2003, 'On Classifying Daylight for Design', *International Journal of Architectural Computing*, 1, pp. 205-217

Géron, Aurélien 2017, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media

He, Conghui, Fu, Haohuan, Guo, Ce, Luk, Wayne and Yang, Guangwen 2017, 'A Fully-Pipelined Hardware Design for Gaussian Mixture Models', *IEEE Transactions on Computers*, 66, pp. 1837-1850

Jain, Anil K and Dubes, Richard 1988, *Algorithms for Clustering Data*, Prentice-Hll, Inc.

Leatherbarrow, David 2004, *Topographical Stories: Studies in Landscape and Architecture*, University of Pennsylvania Press

Lee, Ji Ho and Lee, ji-Hyun 2016 'Cultural Difference in Colour Usages for Building Facades focusing on Theme Park Buildings', *Living Systems and Micro-Utopias: Towards Continous Designing: Proceedings of the 21st International Conference of CAADRIA*

Van Der Maaten, Laurens and Hinton, Geoffrey 2008, 'Visualizing data using t-SNE', *Journal of Machine Learning Research*, 9, pp. 2579-2625

Mitchell, Tom 1997, *Machine Learning*, McGraw-Hill Science/Engineering/Math;

Obeso, Abraham, Benois-Pineau, Jenny, Acosta, Alejandro and Vázquez, Mireya 2016, 'Architectural style classification of Mexican historical buildings using deep convolutional neural networks and sparse features', *Journal of Electronic Imaging*, 26, p. 011016

Pedregosa, Fabian, Varoquaux, Gael, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, Vanderplas, Jake, Passos, Alexandre, Cournapeau, David, Brucher, Matthieu, Perrot, Matthieu and Duchesnay, Edouard 2011, 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825-2830

Samuel, Flora 2013, *Sacred concrete : the churches of Le Corbusier*, Basel : Birkhauser

Shalunts, Gayane 2012 'Architectural Style Classification of Building Facade Towers', *Advances in Visual Computing*, pp. 346-357

Shalunts, Gayane, Haxhimusa, Yll and Sablatnig, Robert 2011, 'Architectural Style Classification of Building Facade Windows', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9474, pp. 285-294

Shalunts, Gayane, Haxhimusa, Yll and Sablatnig, Robert 2012 'Classification of gothic and baroque architectural elements', *2012 19th International Conference on Systems, Signals and Image Processing, IWSSIP 2012*, pp. 316-319

Sharma, Narendra, Bajpai, Aman and Litoriya, Ratnesh 2012, 'Comparison the various clustering algorithms of weka tools', *International Journal of Emerging Technology and Advanced Engineering*, 2, pp. 73-80

Stasiuk, David and Thomsen, Mette 2014 'Learning to be a Vault - Implementing learning strategies for design exploration in inter-scalar systems', *Fusion, Proceedings of the 32nd International Conference on eCAADe*

Xu, Zhe, Tao, Dacheng, Zhang, Ya, Wu, Junjie and Chung Tsoi, Ah 2014, 'Architectural Style Classification using Multinomial Latent Logistic Regression', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8689 LNCS, pp. 600-615

Yambem, Nandita and Nandakumar, A N 2016, 'A Technical Insight into Clustering Algorithms & Applications', *International Research Journal of Engineering and Technology (IRJET)*, 3, pp. 529-533

Yoshimura, Yuji, Cai, Bill, Wang, Zhoutong and Ratti, Carlo 2019, 'Deep learning architect: Classification for architectural design through the eye of artificial intelligence', *Lecture Notes in Geoinformation and Cartography*, NA, pp. 249-265

Yousif, Shermeen and Yan, Wei 2018 'CLUSTERING FORMS FOR ENHANCING ARCHITECTURAL DESIGN OPTIMIZATION', *Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*

Zhao, Peipei, Miao, Qiguang, Song, Jianfeng, Qi, Yutao, Liu, Ruyi and Ge, Daohui 2018 'Architectural Style Classification Based on Feature Extraction Module', *IEEE Access*

[1] https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95

[2] https://www.zacharytateporter.com/assorted-grounds

[3] https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/

[4] https://www.scikit-yb.org/en/latest/api/cluster/elbow.html