

Visualization Term Project

20141087 Ryeongyang Kim 20161206 Jaewoong Lee

November 24, 2019

Contents

1	Introduction	3
2	Materials	3
2.1	Building Layout	3
2.1.1	Basic	3
3	Methods	3
3.1	Scikit-learn: Machine Learning in Python	3
3.2	Matplotlib	4
3.3	Pandas	4
3.4	SciPy	4
3.5	TSNE	4
4	Results	4
4.1	Question 1	4
4.2	Question 2	4
4.3	Question 3	5
4.4	Question 4	5
5	Discussion	5

List of Tables

List of Figures

1	Main Layout of the building	3
2	Energy Zone of the Building	3
3	Prox zone of the Building	3
4	TSNE Plot for General Building Data	4
5	Abnormality Founded by Algorithms	5

1 Introduction

2 Materials

2.1 Building Layout

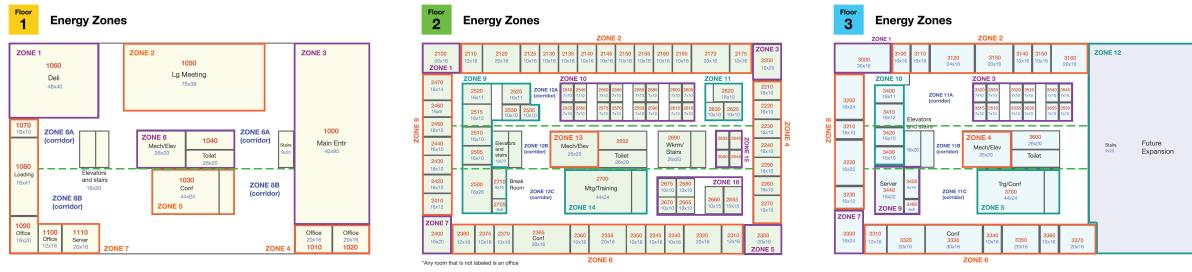
2.1.1 Basic



(a) First Floor (b) Second Floor (c) Third Floor

Figure 1: Main Layout of the building

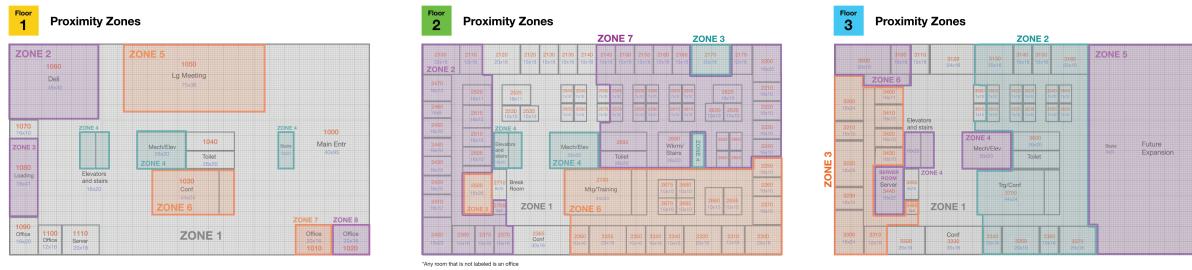
The main layout of this building is as figure 1.



(a) First Floor (b) Second Floor (c) Third Floor

Figure 2: Energy Zone of the Building

The energy zone of this building is as figure 2.



(a) First Floor (b) Second Floor (c) Third Floor

Figure 3: Prox zone of the Building

The prox zone of this building is as figure 3.

3 Methods

3.1 Scikit-learn: Machine Learning in Python

Scikit-learn is a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems [1].

3.2 Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms [2].

3.3 Pandas

Pandas is a Python library of rich data structures and tools for working with structured data sets common to statistic, finance, social sciences, and many other fields [3].

3.4 SciPy

SciPy is a Python-based ecosystem of open-source software for mathematics, science, and engineering [4].

3.5 TSNE

T-distributed Stochastic Neighbor Embedding (TSNE) is a machine learning algorithm for visualization high-dimensional data in a low-dimensional space [5].

4 Results

- 4.1 What are the typical patterns in the prox card data? What does a typical day look like for GAStech employees?
- 4.2 Describe up to five of the most interesting patterns that appear in the building data. Describe what is notable about the pattern and explain its possible significance.

First of all, we drew the distribution of the building data using TSNE technique. The TSNE plot of general building data is shows as figure 4.

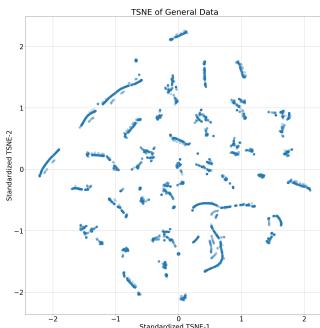


Figure 4: TSNE Plot for General Building Data

To find patterns which appear in the building data, we should find that normality/abnormality in the building data. However, there are over 400 columns in the general building data; therefore, it is almost impossible to find abnormality column-by-column by human. Hence, we used these four algorithms which are included in scikit-learn: *EllipticEnvelope* [6], *OneClassSVM*, *IsolationForest* [7, 8], and *LocalOutlierFactor* [9].

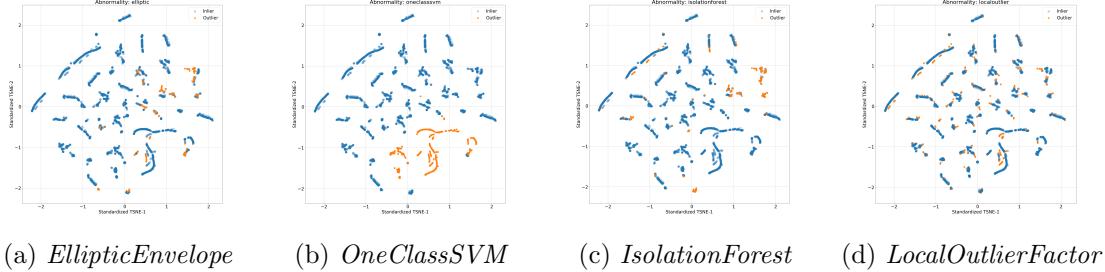


Figure 5: Abnormality Founded by Algorithms

- 4.3 Describe up to five notable anomalies or unusual events you see in the data. Prioritize those issue that are most likely to represent a danger or a serious issue for building operations.
- 4.4 Describe up to three observed relationships between the proximity card data and building data elements. If you find a causal relationship, describe your discovered cause and effect, the evidence you found the support it, and your level of confidence in your assessment of the relationship.

5 Discussion

References

- [1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in python,” *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [2] J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in science & engineering*, vol. 9, no. 3, p. 90, 2007.
- [3] W. McKinney, “pandas: a foundational python library for data analysis and statistics,” *Python for High Performance and Scientific Computing*, vol. 14, 2011.
- [4] E. Jones, T. Oliphant, P. Peterson, *et al.*, “Scipy: Open source scientific tools for python,” 2001.
- [5] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [6] P. J. Rousseeuw and K. V. Driessens, “A fast algorithm for the minimum covariance determinant estimator,” *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
- [7] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, IEEE, 2008.
- [8] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, p. 3, 2012.
- [9] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “Lof: identifying density-based local outliers,” in *ACM sigmod record*, vol. 29, pp. 93–104, ACM, 2000.