University Canada West

Walter Andrés Paz Callizo 2239884

BUSI 651 Machine Learning Tools and Techniques

CAMPUS-FALL23-66: July 8th, 2024 - Sept. 20th, 2024

Vancouver House: Room: W-401: On Campus

Professor Sarah Gholibeigian

Due before 11:59 PM (PT) on Sunday, August 11th, 2024.

Contents

[**Introduction** 4](#_Toc174371945)

[**I.** **Answers** 5](#_Toc174371946)

[**Conclusion** 14](#_Toc174371947)

[**References** 15](#_Toc174371948)

[**Appendix** 16](#_Toc174371949)

[**Figure N** - Number of Appliances and Insulation Thickness (Inches) Relation 5](#_Toc174102376)

[**Figure T** - Insulation Thickness (Inches) and Number of Appliances Relation 6](#_Toc174102377)

[**Figure BB** - Number of Appliances and Energy Consumption Relation 6](#_Toc174102378)

[**Figure DD**- Insulation Thickness (Inches) and Energy Consumption Relation 6](#_Toc174102379)

[**Figure FF** - Training of Model and Linear Regression Instructions for Points 1, 2,3 and Default 7](#_Toc174102380)

[**Figure GG** - Accuracy of 98.33% in the case of Default Model 7](#_Toc174102381)

[**Figure HH** - 98.11% of Accuracy in Point 1 7](#_Toc174102382)

[**Figure JJ** - Residual Plot Point 1 Predicted Values vs Residuals. 7](#_Toc174102383)

[**Figure KK** - Actual vs Predicted Plot Point 1 8](#_Toc174102384)

[**Figure LL** - 98.11% Accuracy in Point 2 8](#_Toc174102385)

[**Figure OO** - Predicted vs Actual Values in Point 2 8](#_Toc174102386)

[**Figure PP** - 98.11% Accuracy in Point 3 8](#_Toc174102387)

[**Figure SS** - Predicted vs Actual Values Plot 3 9](#_Toc174102388)

[**Figure TT** - Mean Squared Error (MAE) of selected values. 9](#_Toc174102389)

[**Figure VV** - KNN for Question II with K1 9](#_Toc174102390)

[**Figure YY** - KNN with K3 for Question II 10](#_Toc174102391)

[**Figure AAA** - Points used for the Classification. 10](#_Toc174102392)

[**Figure BBB** - Index Generation for KNN 10](#_Toc174102393)

[**Figure CCC** - KNN- One Neighbor for Data Set 15](#_Toc174102394)

[**Figure DDD** - KNN Diagram with One Neighbor 16](#_Toc174102395)

[**Figure EEE** - KNN with Three Neighbors 17](#_Toc174102396)

[**Figure FFF** - KNN Predictions for Three Neighbors 17](#_Toc174102397)

[**Figure E** - Preparation of Coding Jupiter Notebook 22](#_Toc174102398)

[**Figure F** - First step is to build a model in Linear Regression (LR) with an initial testing model. 22](#_Toc174102399)

[**Figure G** - Room Area and Outside Temperature Relation 23](#_Toc174102400)

[**Figure H**-Room Area and Insulation Thickness (inches) Relation 23](#_Toc174102401)

[**Figure I** - Room Area and Average Temperature in last 24 hours (C) Relation 24](#_Toc174102402)

[**Figure J** - Room Area and Number of Appliances (NOA) Relation 24](#_Toc174102403)

[**Figure K** - Room Area and Number of Appliances Relation 25](#_Toc174102404)

[**Figure L** - Number of Appliances and Room Area Relation 25](#_Toc174102405)

[**Figure M** - Number of Appliances and Outside Temperature Relation 26](#_Toc174102406)

[**Figure N** - Number of Appliances and Insulation Thickness (Inches) Relation 26](#_Toc174102407)

[**Figure O** - Outside Temperature (C) and Room Area Relation 27](#_Toc174102408)

[**Figure P** - Outside Temperature and Number of Appliances Relation 27](#_Toc174102409)

[**Figure Q** - Outside Temperature (C) and Insulation Thickness (inches) Relation 28](#_Toc174102410)

[**Figure R** - Outside Temperature (C) and Average Temperature in the last 24 hours (C) Relation 28](#_Toc174102411)

[**Figure S** - Insulation Thickness (Inches) and Room Area Relation 29](#_Toc174102412)

[**Figure T** - Insulation Thickness (Inches) and Number of Appliances Relation 29](#_Toc174102413)

[**Figure U** - Insulation Thickness (Inches) and Outside Temperature (C) Relation 30](#_Toc174102414)

[**Figure V** - Insulation Thickness (Inches) and Average Temperature in last 24 hours (C) Relation 30](#_Toc174102415)

[**Figure W** - Average Temperature in the last 24 hours (C) and Room Area Relation 31](#_Toc174102416)

[**Figure X** - Average Temperature in last 24 hours (C) and Insulation Thickness (Inches) Relation 31](#_Toc174102417)

[**Figure Y** - Average Temperature in last 24 hours (C) and Number of Appliances Relation 32](#_Toc174102418)

[**Figure Z** - Average Temperature in las 24 hours (C) and Outside Temperature (C) Relation 32](#_Toc174102419)

[**Figure AA** - Energy Consumption and Room Area Relation 33](#_Toc174102420)

[**Figure BB** - Number of Appliances and Energy Consumption Relation 33](#_Toc174102421)

[**Figure CC** - Outside Temperature (C) and Energy Consumption Relation 34](#_Toc174102422)

[**Figure DD**- Insulation Thickness (Inches) and Energy Consumption Relation 34](#_Toc174102423)

[**Figure EE** - Average Temperature in last 24 hours (C) and Energy Consumption Relation 35](#_Toc174102424)

[**Figure FF** - Training of Model and Linear Regression Instructions for Points 1, 2,3 and Default 35](#_Toc174102425)

[**Figure GG** - Accuracy of 98.33% in the case of Default Model 35](#_Toc174102426)

[**Figure HH** - 98.11% of Accuracy in Point 1 36](#_Toc174102427)

[**Figure II** - Feature Importance Plot Point 1 36](#_Toc174102428)

[**Figure JJ** - Residual Plot Point 1 Predicted Values vs Residuals. 37](#_Toc174102429)

[**Figure KK** - Actual vs Predicted Plot Point 1 37](#_Toc174102430)

[**Figure LL** - 98.11% Accuracy in Point 2 37](#_Toc174102431)

[**Figure MM** - Feature Importance Plot Point 2 38](#_Toc174102432)

[**Figure NN** - Residual Plot of Predicted vs Residuals in Point 2 38](#_Toc174102433)

[**Figure OO** - Predicted vs Actual Values in Point 2 39](#_Toc174102434)

[**Figure PP** - 98.11% Accuracy in Point 3 39](#_Toc174102435)

[**Figure QQ** - Feature Importance Plot Point 3 39](#_Toc174102436)

[**Figure RR** - Residuals vs Predicted Values Point 3 40](#_Toc174102437)

[**Figure SS** - Predicted vs Actual Values Plot 3 40](#_Toc174102438)

[**Figure TT** - Mean Squared Error (MAE) of selected values. 41](#_Toc174102439)

[**Figure UU** – Procedure. 41](#_Toc174102440)

[**Figure VV** - KNN for Question II with K1 42](#_Toc174102441)

[**Figure WW** – Procedure. 42](#_Toc174102442)

[**Figure XX** - Procedure. 43](#_Toc174102443)

[**Figure YY** - KNN with K3 for Question II 43](#_Toc174102444)

[**Figure ZZ** – Procedure. 44](#_Toc174102445)

[**Figure AAA** - Points used for the Classification. 44](#_Toc174102446)

[**Figure BBB** - Index Generation for KNN 45](#_Toc174102447)

# **Introduction**

Taking care of supervised and unsupervised data with the process of orientating a Machine Learning Algorithm for the superlative construction of a business decision through the usage of Machine Learning and the interpretation for the future outcomes such as new features mixing up the most influential variables among the correct correlation based on the proximity towards the line of Regression and the Variance encountered inside of the model while experiencing with less energy consumption in construction. This is the main reason a contractor needs to check on various features for the construction and installation of electrical equipment to a customer’s building or residential homes in the databases of the construction company.

# **Answers**

**Question 1:** After the procedure of multiple **Linear Regression’s (LR’s)** inside of the **Collaboratory Notebook for Jupiter**.

|  |
| --- |
| a) Primary feature: Insulation Thickness Secondary feature: Number of Appliances.  Reason: In both the feature of Number of Appliances and Insulation Thickness not only share better relationship with the Energy Consumption after managing the data with Multiple -Linear Regression but it also conducts a better prediction for a more precise analysis through the Regression Model prepared on the data. |
| b) Feature not contributing: Type of Building and the HVAC System (HVACS).  Mitigation strategy: As an analyst is better to either take it out of the possible variable analysis as these variables or features are merely categorical and will not influence the Linear Regression (LR) as values can only be interpreted in binary code or Boolean Values (1 and 0’s) which is better only to try with Numerical Values and not the Categorical Values. Categorical Values are only used for True and False Tables.    If no such feature exists, justify your claim: -----. |
| c) Apply multiple linear regression:    Energy consumption for point 1: 408.127. Energy consumption for point 2: 443.678. Energy consumption for point 3: 361.018.  The energy consumption for the Points 1, 2 and 3 are the following: For Point 1 the energy consumption of 408.127, in Point 2 the energy consumption predicted is of  443.678 and an energy consumption of 361.018. |
| d) Mean Squared Error (MSE) regression loss: The Mean Squared (MSE) Regression Loss is of 40.359 within Point 1, Point 2 and Point 3 out of the process of complex crossing Multiple-Linear Regression (MLR) with new points for analysis where MSE goes up to 40.359 for the three points alone but it increases as one when the three are used due to some features added or modified but stay similar not only due to similar values but remain in relation to one close with a 0.98 R-Squared where some decimals are the change upon the three new points. |
| e) Recommendation to include the new feature (with reason): A new feature should include certain considerations from the Insulation Thickness (IT) and the Number of Electrical Appliances due to a stable alignment in Linear Regression (LR) delivering a closer linear fitting inside of the diagrams for the three-dimensional comparison of X, Y and Z features. Additionally, the residuals are closer to the Regression Line (RL) signaling a further co-relation of variables through variance and the variation in data explained by the R-Squared furthermore inside of the code with values close to one or one hundred percentage of relation. However, an essential feature to consider inside arguments due to a better control in any electrical and construction projects can be the need of an Investment Risk Assessment as financials have a huge impact in decision-making from the business and customer’s needs, identification for risks, performance, planning and time management for the building built. |

**Question 2:** After KNN modeling with one and three Nearest Neighbors

|  |
| --- |
| **Decision Boundary for K=1:** In the **Decision Boundary for K=1** the new area covered by the K-Nearest Neighbors which is the following:      **Figure CCC** - KNN- One Neighbor for Data Set    **Figure DDD** - KNN Diagram with One Neighbor  As it is understandable the value for (8,4) remains right in the border of the highlighted area versus the point (8,6) inside of the selected area where (8,4) stands for Blue and (8,6) for yellow figures inside of the studied KNN depending on if switching from 1 to 3 KNN selected in grouped area with predominant values.  **Decision Boundary for K=3:** In the **Decision Boundary for K=3** the new area covered by the K-Nearest Neighbors which is the following:      **Figure EEE** - KNN with Three Neighbors    **Figure FFF** - KNN Predictions for Three Neighbors |
| **Classify:**   |  |  | | --- | --- | |  | **Prediction** | | **point (8, 6)** | **Blue Square Yellow Circle** | | **point (8, 4)** | **Blue Square Yellow Circle** | |

Question 3: Secondary K-Nearest Neighbor Analysis done through modeling.

**Question 3:**

|  |
| --- |
| **Predicted labels (+ or -):**  Predictions for new points: ['+' '-']  Point [7.81 5.33] is classified as +  Point [9.43 5.29] is classified as -  **Point 1:** [7.81,5.33] +  **Point 2:** [9.43,5.29] - |
| **Index of closest neighbors: Indices of the closest neighbors for each point: [[13 11 14]**  **[11 13 19]]**  **For Point 1:** [13,11,14]  **For Point 2:** [11,13,19] |

# **Conclusion**

Overall, the importance of keeping any of the subsets and datasets covered with Linear Regression (LR) and the concept of K-Nearest Neighbor is for the variable studies while discovering a novelty into the possibilities regarding various scenarios before dealing with issues and any deductions coming from the current data. Investing in the future for more features to be analyzed like investment can save time and efforts for the business to obtain closer results to the deletion of human error with proper measurements taken into consideration when making a choice or consideration for any commercial or residential client.

# **References**

Chat Gpt (2024). Chat Gpt. <https://chatgpt.com/>

Gholibeigian, S. (2024). Classification - KNN.ipynb. <https://colab.research.google.com/drive/1eOOh4Kc09rFg9D7vjzPLwyGuHKFBE9gM#scrollTo=cK9x9SfLBfYp>

Paz Callizo, W.A. (2024). Construction Company Assignment I Individual ML UCW.ipynb. <https://colab.research.google.com/drive/1pilXuoNFniC-XrCb5ZQLCbTxZavr5CnV#scrollTo=zug_npRJoyHi>

# **Appendix**

**Procedure**

Question 1:

A graph of energy consumption and hvac systems

Description automatically generated

**Figure N** - Number of Appliances and Insulation Thickness (Inches) Relation

A graph of energy consumption

Description automatically generated

**Figure T** - Insulation Thickness (Inches) and Number of Appliances Relation

A graph with blue dots

Description automatically generated

**Figure BB** - Number of Appliances and Energy Consumption Relation

A graph with blue dots

Description automatically generated

**Figure DD**- Insulation Thickness (Inches) and Energy Consumption Relation

A screenshot of a computer

Description automatically generated

**Figure FF** - Training of Model and Linear Regression Instructions for Points 1, 2,3 and Default

A white background with black numbers and symbols

Description automatically generated

**Figure GG** - Accuracy of 98.33% in the case of Default Model

A screenshot of a computer

Description automatically generated

**Figure HH** - 98.11% of Accuracy in Point 1

**Figure JJ** - Residual Plot Point 1 Predicted Values vs Residuals.

A graph with blue dots and red lines

Description automatically generated

**Figure KK** - Actual vs Predicted Plot Point 1

A screenshot of a computer error

Description automatically generated

**Figure LL** - 98.11% Accuracy in Point 2

A graph with blue dots and red lines

Description automatically generated

**Figure OO** - Predicted vs Actual Values in Point 2

A computer code with numbers and symbols

Description automatically generated

**Figure PP** - 98.11% Accuracy in Point 3

A graph with blue dots and red lines

Description automatically generated

**Figure SS** - Predicted vs Actual Values Plot 3

**A screenshot of a computer

Description automatically generated**

**Figure TT** - Mean Squared Error (MAE) of selected values.

Question 2:

A screen shot of a graph

Description automatically generated

**Figure VV** - KNN for Question II with K1

A screen shot of a graph

Description automatically generated

**Figure YY** - KNN with K3 for Question II

Question 3:

A screenshot of a computer code

Description automatically generated

**Figure AAA** - Points used for the Classification.

A screenshot of a computer program

Description automatically generated

**Figure BBB** - Index Generation for KNN

Question 1:

A screenshot of a computer

Description automatically generated

**Figure E** - Preparation of Coding Jupiter Notebook

A graph of a graph of temperature

Description automatically generated with medium confidence

**Figure F** - First step is to build a model in Linear Regression (LR) with an initial testing model.

A graph of energy consumption

Description automatically generated

**Figure G** - Room Area and Outside Temperature Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure H**-Room Area and Insulation Thickness (inches) Relation

A graph of energy consumption

Description automatically generated

**Figure I** - Room Area and Average Temperature in last 24 hours (C) Relation

A graph of energy consumption

Description automatically generated

**Figure J** - Room Area and Number of Appliances (NOA) Relation

A graph of a diagram

Description automatically generated

**Figure K** - Room Area and Number of Appliances Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure L** - Number of Appliances and Room Area Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure M** - Number of Appliances and Outside Temperature Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure N** - Number of Appliances and Insulation Thickness (Inches) Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure O** - Outside Temperature (C) and Room Area Relation

A graph of a temperature measurement

Description automatically generated with medium confidence

**Figure P** - Outside Temperature and Number of Appliances Relation

A graph of a graph showing the temperature of a temperature

Description automatically generated with medium confidence

**Figure Q** - Outside Temperature (C) and Insulation Thickness (inches) Relation

A graph of energy consumption and heating

Description automatically generated

**Figure R** - Outside Temperature (C) and Average Temperature in the last 24 hours (C) Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure S** - Insulation Thickness (Inches) and Room Area Relation

A graph of energy consumption

Description automatically generated

**Figure T** - Insulation Thickness (Inches) and Number of Appliances Relation

A graph of energy consumption and heating

Description automatically generated

**Figure U** - Insulation Thickness (Inches) and Outside Temperature (C) Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure V** - Insulation Thickness (Inches) and Average Temperature in last 24 hours (C) Relation

A graph of energy consumption

Description automatically generated

**Figure W** - Average Temperature in the last 24 hours (C) and Room Area Relation

**A graph of energy consumption and heating

Description automatically generated**

**Figure X** - Average Temperature in last 24 hours (C) and Insulation Thickness (Inches) Relation

A graph of energy consumption and hvac systems

Description automatically generated

**Figure Y** - Average Temperature in last 24 hours (C) and Number of Appliances Relation

A graph of energy consumption and heating

Description automatically generated

**Figure Z** - Average Temperature in las 24 hours (C) and Outside Temperature (C) Relation

A graph with blue dots

Description automatically generated

**Figure AA** - Energy Consumption and Room Area Relation

A graph with blue dots

Description automatically generated

**Figure BB** - Number of Appliances and Energy Consumption Relation

A graph with blue dots

Description automatically generated

**Figure CC** - Outside Temperature (C) and Energy Consumption Relation

A graph with blue dots

Description automatically generated

**Figure DD**- Insulation Thickness (Inches) and Energy Consumption Relation

A graph with blue dots

Description automatically generated

**Figure EE** - Average Temperature in last 24 hours (C) and Energy Consumption Relation

A screenshot of a computer

Description automatically generated

**Figure FF** - Training of Model and Linear Regression Instructions for Points 1, 2,3 and Default

A white background with black numbers and symbols

Description automatically generated

**Figure GG** - Accuracy of 98.33% in the case of Default Model

A screenshot of a computer

Description automatically generated

**Figure HH** - 98.11% of Accuracy in Point 1

A graph with blue bars

Description automatically generated with medium confidence

**Figure II** - Feature Importance Plot Point 1

A diagram of a plot

Description automatically generated

**Figure JJ** - Residual Plot Point 1 Predicted Values vs Residuals.

A graph with blue dots and red lines

Description automatically generated

**Figure KK** - Actual vs Predicted Plot Point 1

A screenshot of a computer error

Description automatically generated

**Figure LL** - 98.11% Accuracy in Point 2

A graph with blue bars

Description automatically generated with medium confidence

**Figure MM** - Feature Importance Plot Point 2

A graph with blue dots and red line

Description automatically generated

**Figure NN** - Residual Plot of Predicted vs Residuals in Point 2

A graph with blue dots and red lines

Description automatically generated

**Figure OO** - Predicted vs Actual Values in Point 2

A computer code with numbers and symbols

Description automatically generated

**Figure PP** - 98.11% Accuracy in Point 3

A graph with blue and white bars

Description automatically generated

**Figure QQ** - Feature Importance Plot Point 3

A diagram of a plot

Description automatically generated

**Figure RR** - Residuals vs Predicted Values Point 3

A graph with blue dots and red lines

Description automatically generated

**Figure SS** - Predicted vs Actual Values Plot 3

**A screenshot of a computer

Description automatically generated**

**Figure TT** - Mean Squared Error (MAE) of selected values.

Question 2:

A screenshot of a computer program

Description automatically generated

**Figure UU** – Procedure.

A screen shot of a graph

Description automatically generated

**Figure VV** - KNN for Question II with K1

A screenshot of a computer program

Description automatically generated

**Figure WW** – Procedure.

A screenshot of a computer program

Description automatically generated

**Figure XX** - Procedure.

A screen shot of a graph

Description automatically generated

**Figure YY** - KNN with K3 for Question II

Question 3:

A screenshot of a computer program

Description automatically generated

**Figure ZZ** – Procedure.

A screenshot of a computer code

Description automatically generated

**Figure AAA** - Points used for the Classification.

A screenshot of a computer program

Description automatically generated

**Figure BBB** - Index Generation for KNN