Understanding Energy Consumption of Appliances

Group members

Patel charmiben kamleshkumar - Stepwise Regression
Uppala Sanjay - Ridge Regression
Vamsi Eragandhula - Lasso Regression
Murali krishna G - PCA
Krunal panchal - CCA

Executive summary:

Application and Dataset

The research aims to predict the energy consumption of household appliances based on various temperature and humidity levels within a residence and surrounding weather data. The dataset spans 4.5 months, with measurements taken at 10-minute intervals. It includes temperature and humidity readings from different rooms, collected via a wireless sensor network, as well as energy consumption data from smart meters and weather data from a nearby airport.

Background

With rising energy consumption and its environmental impact, there is a growing need to understand and manage energy usage in residential settings. Appliances are a significant contributor to household energy consumption, making it essential to identify factors influencing their energy demand. By accurately predicting appliance energy consumption, we can not only help consumers save money but also facilitate the potential for generating revenue by returning excess energy to the grid.

Methods

To tackle the complexity of the data and address multicollinearity, we employed several statistical techniques:

Stepwise regression: Iteratively adding and removing predictors based on their significance, providing insights into the impact of multicollinearity.
Ridge regression: Introducing a penalty term to reduce the influence of correlated predictors and improve model performance.
Lasso regression: Promoting sparsity and shrinking less influential coefficients to zero, allowing for feature selection and a more efficient predictive model.
Principal Component Analysis (PCA): Reducing the dimensionality of the data and understanding the variance explained by different components.
Canonical Correlation Analysis (CCA): Examining the relationship between sets of temperature and humidity variables and their contribution to appliance energy consumption.

Results

The stepwise regression revealed the presence of multicollinearity in the data, with some predictors having a positive relationship (e.g., lights, temperature in various rooms, humidity in kitchen) and others having a negative relationship (e.g., humidity in various rooms, temperature in living areas) with appliance energy consumption.

Ridge regression effectively addressed multicollinearity by shrinking coefficient estimates, providing more stable and reliable results. The cross-validated mean squared error and residual plots demonstrated the model's effectiveness in handling correlated predictors.

Lasso regression identified 15 key influencing variables, including temperature and humidity levels in specific rooms, outside temperature, humidity, and wind speed. The model equation quantifies the impact of each variable on appliance energy consumption.

PCA helped reduce the dimensionality of the data, with four principal components (temperature, room humidity, outside environment, and utilities) accounting for 72.3% of the total variance.

Limitations and Future Work

While the research provides valuable insights, there are some limitations to consider:

The dataset covers a relatively short period (4.5 months), which may not account for seasonal variations or long-term trends.

The analysis focuses on temperature and humidity factors, but there may be other influential variables (e.g., occupancy patterns, appliance usage habits) that could further improve prediction accuracy.

The research is specific to a single residential setting, and the results may vary for different household configurations or geographic locations.

Future work could involve collecting data over a longer period to capture seasonal effects, incorporating additional variables related to occupant behavior and appliance usage, and validating the models across multiple residential settings to enhance generalizability.

Conclusions:

This research has successfully employed various statistical techniques to predict appliance energy consumption based on temperature, humidity, and weather data. The findings highlight the significant impact of specific variables, such as temperature and humidity levels in different rooms, on energy consumption. By addressing multicollinearity and reducing dimensionality, the analysis provides a robust predictive model that can aid in energy management and potential cost savings for residential consumers.

Journal article

Abstract

The research project delves into forecasting energy consumption in residential buildings by analyzing temperature, humidity, and weather data. The study employs statistical methods like Principal Component Analysis (PCA) to reduce dimensionality and Canonical Correlation Analysis (CCA) to explore the relationship between temperature and humidity variables. The findings reveal a strong correlation between temperature and humidity, indicating their significant impact on energy consumption prediction.

Furthermore, multilinear regression techniques such as Stepwise, Ridge, and Lasso regressions are utilized to predict appliance energy consumption. Stepwise regression identifies multicollinearity issues, while Ridge regression effectively addresses this problem by shrinking coefficient estimates. Lasso regression enhances the model by promoting sparsity and identifying relevant features for improved predictive accuracy.

The results underscore the importance of these statistical methods in refining energy consumption predictions in residential settings. By leveraging PCA to understand variance and CCA to assess variable relationships, the study enhances the accuracy and interpretability of energy consumption forecasts. The implementation of multilinear regressions further refines the predictive models, emphasizing the significance of temperature, humidity, and weather data in forecasting energy usage in residential buildings.

Introduction

The globe is starting to take notice of the rising trend in energy consumption because it is causing an increase in greenhouse gas emissions and carbon emissions each year. The industrial sector uses the majority of the electricity generated, but the residential sector also uses a significant amount of it. Given that home appliances use the most energy in a residence, it is critical to research how energy-consuming behavior occurs in this sector of the economy. The goal of this project is to forecast how much energy household appliances will use depending on temperature and humidity levels of various rooms in the facility as well as surrounding weather data. The data set spans 4.5 months, with 10 minute intervals. The temperature and humidity levels in the home are monitored via a ZigBee wireless sensor network. Each wireless node transmitted temperature and humidity data for approximately 3.3 minutes. The wireless data was then averaged across 10-minute intervals. Every 10 minutes, m-bus energy meters recorded energy data. The experimental data sets were combined with the weather data from the nearest airport weather station (Chievres Airport, Belgium) that was extracted from a public data set from Reliable Prognosis. (rp5.ru). The data set includes two random variables for testing regression models and excluding non-predictive characteristics (parameters). In the age of smart homes, the ability to predict energy consumption can not only save money for the end user, but also help the user generate money by returning excess energy to the grid. In this case, regression analysis will be used to predict appliance energy consumption based on data collected from various sensors.

Literature Review

There have been many previous studies made related to energy consumption and based on the data available, data driven decision was conducted which shows the relationships between appliances energy consumption and different predictors. It highlighted several factors due to which various features are affected. The energy consumption of appliances represents a significant portion of the aggregated electricity demand of the residential sector due to the increased number of appliances owned, making it important to identify which ones are the main contributors to the energy consumption. Apart from that, the previous study also suggests that weather parameters have been proven relevant to predict the electricity energy consumption in buildings. This consumption typically includes the HVAC contributions, which may have different trends with respect to temperature for example. Thus, it would be desirable to know if exterior weather parameters can also help in the prediction of appliances energy use. A prediction system for the problem of individual appliance prediction was presented in [2]. The system used information such as past consumption, hour, day, season and month. The system is capable of learning from past data. One of the main conclusions was that the last 24 h are the most relevant for prediction. This study showed different models which predict the appliances. All those studies suggest some or the other way but there is no mention of statistical analysis using methods like PCA or CCA and selecting relevant features to make the model predict appliances more accurately. So the aim is to get the relevant features which effects the appliances and increase the model prediction power

Methods:

The aim of this study is to find the energy consumption in residence. The data has various temperatures and humidity levels in each room in a residence for a timeline of 4.5months and with a span of 10 mins. The first step we did is cleaning the data, exploratory data analysis, removing outliers and making sure the dependent variable(appliance) is normally distributed. We used correlation plots and VIF values to find if the data has multicollinearity and the results have shown data has a good amount of multicollinearity. So, to tackle this and do our research on this data we used three multilinear regressions. The first one is stepwise. Which iteratively adds and removes predictors based on their significance level, providing insights into the impact of multicollinearity on the model's performance. Second is ridge regressions and third lasso regressions, they add a penalty term which helps to reduce multicollinearity and improve model's performance. The Lasso regression adds a penalty term to the linear regression cost function, promoting sparsity in the model and has the ability in shrinking the less influential coefficients to zero. This ability of lasso regression can identify relevant features and build a model around them with a more efficient predictive model. And PCA(Principal Component Analysis) was used to understand how much variance in the data is explained by different components. By reducing the dimensionality of the data, PCA can reduce the multicollinearity issues. Then we tried applying CCA(Canonical Correlation Analysis) to check the relation between sets of temperature and humidity and how they contribute towards the dependent variable. The features are divided into two sets - temperature which contains all the features related to temperature, and humidity which contains features of humidity. Each canonical correlation coefficients the strength of the relationship between the sets of variables, but for a different pair of linear combinations of the variables.

Discussion and Results:

Stepwise method:

After looking at the correlation plots we came to know there is multicollinearity. So, to get a clear picture of it we implemented, stepwise method to understand the impact of it on the model. We trained null and full model and applied stepwise method got R^2 value of 0.2782 with p-value < 2.2e-16 and F-statistic: 285.6 on 21 and 15561 DF, even though the R^2 value is less and says the model is not performing well the p-value suggest that there is at least one predictor significant for predicting appliance energy of a residence.

Table 1. VIF values for stepwise model

```
       lights humid_station  temp_laundry   temp_office    humid_teen humid_kitchen  humid_living     temp_teen  temp_parents
     1.245243     47.832501      9.850873      8.406901      6.499803     16.246974     20.583099      7.175488     26.031634
 temp_outside   temp_living    humid_bath     Windspeed  temp_station humid_laundry humid_parents     temp_iron humid_outside
    32.811600     26.871766      1.380440      1.480328    142.470522     10.260136      6.077605     14.460303      9.557382
    Tdewpoint  temp_kitchen     temp_bath
    82.232027     20.084923     10.217069
```

As shown in table1 there is a lot of multicollinearity in the data. We have compared the null model , full model and stepwise model and the results show us both full model and stepwise model are similar even though we have continued our research with the stepwise model. As for the coefficients there are some predictors with positive relation(lights,temp_laundry ,humid_kitchen,temp_teen,temp_outside,humid_bath,Windspeed,humid_laundry,humid_outside, Tdewpoint,temp_kitchen,temp_bath) with appliance which means if there is any increase value of this predictors there will be increase in usage of energy consumed by appliance and some with predictors has negative relation (humid_station , temp_laundry ,humid_teen, humid_living, temp_parents, temp_living,temp_station ,humid_parents ,temp_iron) with appliances which means if there is any decrease in value of this predictors there will be decease in consumption of energy.

Then finally we checked all the assumptions of multilinear regression and the residuals have a normal distribution. There is multicollinearity in this data so, failing this assumption. There is a clear pattern in homoscedasticity, failing this assumption too. By looking at all of these results, especially the weak R^2 value and violating some assumptions, this stepwise model doesn't have good performance to predict energy consumption of appliances in residence.

Ridge Regression :

Initially, an examination of multicollinearity was conducted by calculating theVIFs. The analysis revealed that several VIFs exceeded the threshold of 10, indicating high multicollinearity among the predictor variables. Multicollinearity can lead to inflated standard errors and unreliable coefficient estimates, which can compromise the model's interpretability and predictive accuracy.

To reduce the multicollinearity we have implemented ridge regression which adds a penalty term to the sum of squared residuals, shrinking the coefficient estimates toward zero and reducing the influence of correlated predictors. The ridge regression model was fitted using the `cv.glmnet` function from the `glmnet` package in R, which performs cross-validation to determine the optimal value of the tuning parameter (lambda) that controls the strength of the penalty term.

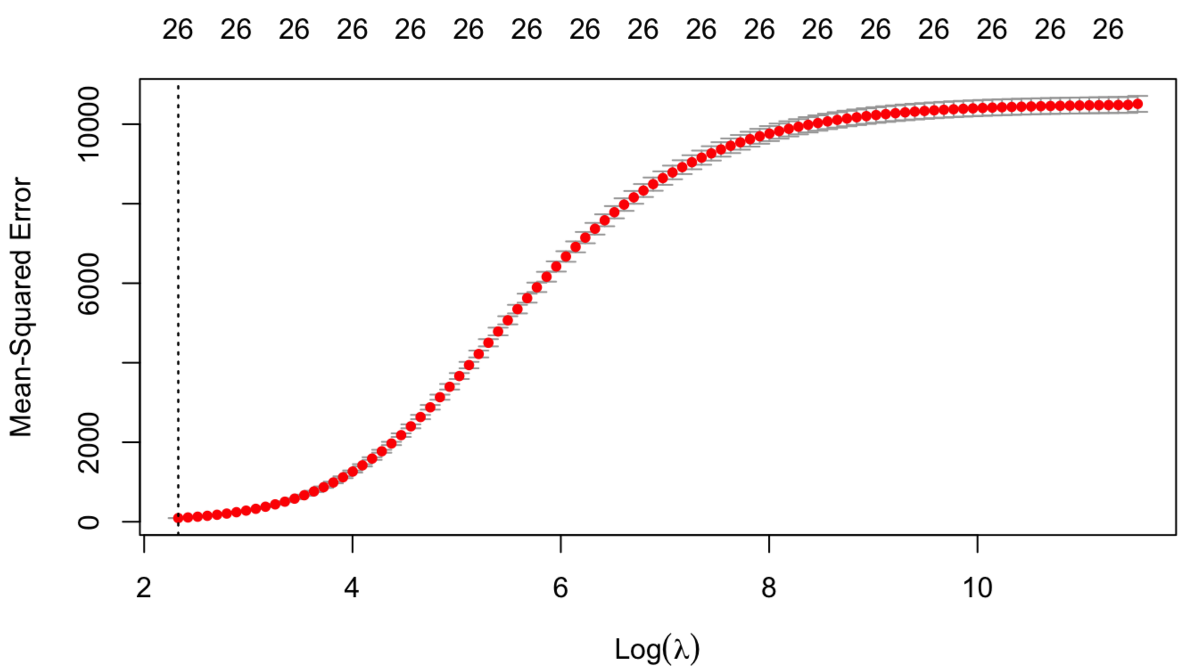

Figure 1:Cross-Validated Mean Squared Error vs. Log(Lambda)

The ridge regression coefficients were obtained using the `coef` function, and predictions were made using the `predict` function. To evaluate the model's performance, the cross-validated mean squared error was plotted against lambda. This plot (Figure 1) visualized the trade-off between model complexity and predictive accuracy, aiding in the selection of an appropriate lambda value.

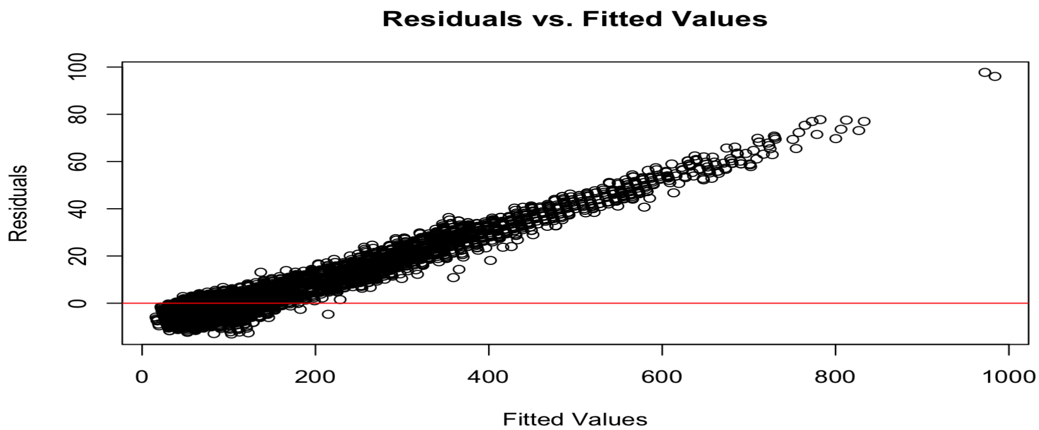

Figure 2: Residual plots vs Fitted Values

Additionally, residual plots were generated to assess the model's assumptions and identify potential outliers or patterns in the residuals. The residual plot for the ridge regression model (Figure 2) showed a relatively even distribution of residuals around the zero line, indicating that the model's assumptions were reasonably met.
Overall, the implementation of ridge regression effectively addressed the multicollinearity issue in the dataset, providing more stable and reliable coefficient estimates compared to ordinary least squares regression. This regularization technique allowed for the interpretation of individual predictor variables' impact on energy consumption while mitigating the adverse effects of multicollinearity.

By visualizing the cross-validated mean squared error and residual plots, the analysis demonstrated the effectiveness of ridge regression in handling correlated predictor variables and obtaining a robust model for predicting energy consumption in buildings.

Lasso Regression:

We used Lasso regression as one of our techniques. Lasso regression is used when we have a limited set of variables as our dataset is very much suited for this. Optimization technique is used in the lasso regression where this optimization technique introduces a penalty term to the linear regression cost function, promoting sparsity in the model and has the ability in shrinking the less influential coefficients to zero. Its capacity to effectively shrink the coefficients of insignificant variables to zero makes it valuable in scenarios with a small number of predictors. This ability of lasso regression can identify relevant features and build a model around them with a more efficient predictive model.

The energy dataset has some unwanted data like Date, and non-dimensional data that are removed for the lasso analysis by exploring the correlation between them and the target variable

Appliances. Removed outliers in dataset. We used the “glmnet” package in R to implement the lasso regression. Then two models minimum lambda model (Lasso_model_min) and 1 standard error lambda (Lasso_model_1se) are built with different lambda values. The minimum lambda value model built by the smallest lambda value possible that can minimize the cross-validation error. This will build a complex model possible with good fit to training data. 1 standard error rule provides a model with less features , less prone to overfitting which is built using lambda value which is 1 standard error away that minimizes the cross validation error.

Table 2: Lasso model output

```
Call:  cv.glmnet(x = as.matrix(standardized_predictors), y = response_variable,      alpha = 1)

Measure: Mean-Squared Error

    Lambda Index Measure    SE Nonzero
min 0.0453    67    8459 190.4      22
1se 1.1750    32    8634 209.9      15
```

The above Table 2 shows the outputs of the Lasso model and the interpretation done below.

Let's analyze the provided output for both the minimum lambda (min) and 1 standard error lambda (1se):

Minimum Lambda Model (Min model):

- MSE: 8459
- Lambda : 0.0453
- Nonzero Coefficients: 22

1 Standard Error Lambda Model (1se model):

- MSE: 8634
- Lambda : 1.1750
- Nonzero Coefficients: 15

The better model is basically chosen by the less mean squared error , the lower the MSE values the better accurate model is built. Anyways the choice between the two models depends on the specific goals and considerations weighing the trade-off between model accuracy and simplicity. As shown above The Minimum Lambda model has 22 Non zero variables which is more complex compared to 15 Non zero variables in 1se Lambda Model. But the Min model is more accurate compared to the 1se model because of its lower MSE value. This 1se model with 15 Non zero variables reduces the overfit of data,potentially improves generalization to new data and there is almost or slight difference with MSE of both models. I chose the 1se model as the better choice for my analysis.

Table 3: %Deviance of Min lasso model

```
Call:  glmnet(x = as.matrix(standardized_predictors), y = response_variable,      alpha = 1, lambda = 
0.0453) 

  Df  %Dev Lambda
1 23 16.88 0.0453
```

Table 4: %Deviance of 1se Lasso model

```
Call:  glmnet(x = as.matrix(standardized_predictors), y = response_variable,      alpha = 1, lambda = 
1.175) 

  Df  %Dev Lambda
1 15 15.07  1.175
```

The above tables Table 3 and Table 4 show the percentage of deviance of the two models and the percentage of deviance is a common term in model evaluation. It is calculated as the percentage reduction in deviance brought about by the model compared to a null model (a model with no predictors). The lower deviance indicates a better fit to the data. So the 1se model with 15 Non zero variables with 15.07 percent deviance is a better choice for the model compared to the Min model. So I chose 1 se model as my Final model for analysis.

Table 5: Coefficients of 1se Final model

```
26 x 1 sparse Matrix of class "dgCMatrix"
                      s0
(Intercept)   95.3971511
lights        16.1963406
temp_kitchen  -3.7172994
humid_kitchen 30.3376384
temp_living     .       
humid_living -17.8747995
temp_laundry  27.3012323
humid_laundry  5.7160743
temp_office   -7.3555021
humid_office    .       
temp_bath     -1.4070362
humid_bath      .       
temp_outside   3.3140327
humid_outside   .       
temp_iron       .       
humid_iron    -6.2411671
temp_teen      1.5979023
humid_teen   -24.6231980
temp_parents -18.0130492
humid_parents   .       
temp_station    .       
Press_mm_hg     .       
humid_station -0.1012667
Windspeed      3.3825759
Visibility      .       
Tdewpoint       .       
```

As shown in Table 5 above these are the 15 Non zero variables and the coefficients of those variables after the less influential variables are shrunk to zero.

The key influencing variables where one unit increase is associated with an increase of 16.196 lights, 30.337 humidity in kitchen , 27.301 temperature in laundry ,5.716 humidity in laundry, 3.314 increase temperature outside , 1.597 temperature in teenager room, and 3.382 increase in wind speed. And one unit increase associated with decrease of 17.874 humidity in living room, 24.623 humidity in teenager room, 6.24 humidity in ironing room, 18.01 in temperature in Parents room , 3.717 temperature in kitchen, 7.3555 temperature in office, 1.407 temperature in bathroom, 0.010 in humidity in station. By this we can say that Lasso regression can make the model simple by variable selection as said by making less influential coefficients to zero and increases the performance by reducing the overfitting of data and improving interpretability.

PCA

We conducted PCA to understand how much variance in the data is explained by different components and addressed multicollinearity issues by reducing dimensionality. This analysis is aimed to uncover the underlying patterns and relationships among variables while retaining the most informative features. The data consisted of 27 variables 2 of them with cutoff< 0.3 has been removed and we continued with only 25 variables in the data. The Kaiser-Meyer-Olkin test score was 0.86 and the results of Bartlett's test of sphericity is $X2 = 791598.192$,$df = 351$, p-value $<$ 2.22e-16, the both results suggest that we can use PCA for this data, it was performed with varimax rotation and we got 6 components, however we have selected 4 components for our main research. The first factor is temperature explained 33.5% of total variance (eigenvalue = 9.32), the second factor is room humidity explained 26.2% of total variance(eigenvalue = 7.063), the third is outside environment explained 7.8% of total variance(eigenvalue = 1.87), the fourth is utilities explained 4.8% of total variance(eigenvalue =1.27), these four components accounted for 72.3% of total variance.

Table 6 Principal component analysis

```
Loadings:
               RC1    RC2    RC3    RC4
temp_kitchen   0.928
temp_living    0.786         0.395
temp_laundry   0.934
temp_office    0.936
temp_bath      0.939
temp_outside   0.732         0.578
humid_outside -0.762  0.446
temp_iron      0.943
temp_teen      0.891
temp_parents   0.964
temp_station   0.740         0.568
humid_kitchen         0.898
humid_living          0.819 -0.324
humid_laundry         0.910
humid_office          0.953
humid_iron            0.938
humid_teen            0.918
humid_parents         0.905
Tdewpoint      0.610  0.636
humid_station -0.394  0.471 -0.684
Windspeed                    0.646
Appliances                          0.639
lights                              0.778
humid_bath            0.382         0.375
Press_mm_hg          -0.314
Visibility
random_1

                 RC1   RC2   RC3   RC4
SS loadings    9.055 7.074 2.103 1.299
Proportion Var 0.335 0.262 0.078 0.048
Cumulative Var 0.335 0.597 0.675 0.723
```

All the four components have mean 0 which says they have a balanced distribution. The 1st component temperature has a min and max value of -2.3680 and 2.7723, which are lower limit and upper limit of the temperature component. The 2nd component room humidity has a min and max value of -2.6572 and 2.8124 which are lower limit and upper limit of the room humidity component. The 3rd component outside the environment has a min and max value of -2.18197 and 2.95996 here the upper limit is little higher than the lower limit and there might be a chance of outliers. The 4th component has min and max of -1.5930 and 7.0252 here the upper limit is so high and it indicates there is a huge chance this is because of the outliers.

The PCA analysis successfully reduced the dimensionality of the dataset and explained variance of the data. These findings can be used to inform further analysis and decision making in understanding the variables.

CCA

For further analysis, it was found that we have two main feature sets which were contributing towards the model building. So, to get more detailed insights and how the features are related, we created two sets - temperature and humidity which contain all the features related to temperature and humidity to perform CCA.

```
## $cor
##  [1] 0.9498850 0.9089053 0.6978795 0.5973289 0.4181468 0.4094495 0.3659453
##  [8] 0.1790645 0.1328240 0.0768844
##
## $xcoef
##                       [,1]          [,2]          [,3]          [,4]
## temp_kitchen  2.460573e-03 -1.134707e-03 -4.283186e-03 -0.0006324457
## temp_living  -3.790931e-03  3.947921e-03  4.057248e-03  0.0010191018
## temp_laundry -2.851871e-04 -7.037197e-04  2.942336e-04  0.0025323507
## temp_office   3.610294e-04 -5.672221e-04  1.871756e-03 -0.0016240927
## temp_bath    -2.109130e-04 -8.112067e-05 -2.382660e-03  0.0033852006
## temp_outside  2.303094e-05 -9.236028e-04  1.386194e-03  0.0002865744
## temp_iron    -1.155919e-03 -5.844421e-04  3.668385e-05 -0.0085674902
## temp_teen     3.641705e-04  1.034327e-04  1.532812e-03  0.0011995831
## temp_parents  1.686108e-03 -1.955462e-03  2.513383e-03  0.0046908008
## temp_station -5.524238e-04  2.562840e-04 -3.254085e-03 -0.0007794603
##                       [,5]         [,6]          [,7]          [,8]
## temp_kitchen -0.0008747011 -0.010730605 -0.0045041010 -0.0077244788
## temp_living  -0.0009456172  0.001402518  0.0011134419  0.0016872388
## temp_laundry -0.0043164120  0.004415487  0.0007210351 -0.0026577193
## temp_office   0.0017119879  0.006544862  0.0003255489 -0.0020405572
## temp_bath     0.0043990571  0.001628099 -0.0067303758  0.0063338671
## temp_outside  0.0012970728 -0.002710613 -0.0005266925  0.0005648619
## temp_iron    -0.0024113594 -0.000969576 -0.0048076675  0.0015401090
## temp_teen     0.0068543767  0.001330466  0.0039357967 -0.0015730349
## temp_parents -0.0050715825 -0.004654802  0.0065818510  0.0023269926
## temp_station -0.0005447410  0.002757218  0.0013363157 -0.0005642115
##                       [,9]         [,10]
## temp_kitchen -0.0067912441 -0.0023091347
## temp_living   0.0038534814 -0.0023961335
## temp_laundry  0.0046657748  0.0059132884
## temp_office  -0.0055079343  0.0006301749
## temp_bath     0.0005572467  0.0014858991
## temp_outside -0.0025530108  0.0042367476
## temp_iron     0.0038914082  0.0034989107
## temp_teen     0.0033729677  0.0012215711
## temp_parents -0.0029009914 -0.0105969468
## temp_station  0.0021167765 -0.0040224363
##
```

```
## $ycoef
##                         [,1]          [,2]          [,3]          [,4]
## humid_kitchen -1.200403e-03  5.109044e-04  2.393829e-03  6.501760e-04
## humid_living   1.762572e-03 -1.795258e-03 -2.597615e-03 -6.732956e-04
## humid_laundry  5.226853e-05  8.928530e-04  7.044698e-04 -1.280067e-04
## humid_office  -7.649771e-04  4.595866e-04 -1.530746e-03  1.904628e-03
## humid_bath     7.738806e-06  4.423667e-05  4.236782e-05  7.375335e-05
## humid_outside  8.141849e-05  1.916713e-04 -1.710826e-04 -1.716078e-04
## humid_iron    -4.759988e-04 -3.022238e-04 -3.179654e-04 -1.748840e-03
## humid_teen     1.401539e-04  1.979666e-06 -4.192876e-04  1.290033e-03
## humid_parents -8.834184e-05 -3.503337e-04  5.308896e-04 -5.997210e-04
## humid_station  9.258651e-05  9.356934e-06  6.487118e-04  4.769393e-04
##                         [,5]          [,6]          [,7]          [,8]
## humid_kitchen -6.963090e-04  1.957914e-03  0.0019611008  0.0034289893
## humid_living   5.203062e-04  3.291440e-04 -0.0002590224 -0.0006385056
## humid_laundry -6.951376e-04  1.463611e-03 -0.0025342230 -0.0051396768
## humid_office   1.622120e-03 -4.254012e-03  0.0007581391 -0.0002724897
## humid_bath     3.977208e-04  1.210889e-04 -0.0005002296  0.0003014293
## humid_outside -1.642987e-05  4.409095e-05  0.0001428986  0.0002462038
## humid_iron    -8.757425e-04  2.336763e-04 -0.0021415140  0.0012110370
## humid_teen     1.197837e-03  1.749058e-03  0.0015524401 -0.0008661219
## humid_parents -2.215180e-03 -5.758021e-04  0.0002647064  0.0005186696
## humid_station -2.062452e-04 -2.581220e-04 -0.0002657028 -0.0001140868
##                         [,9]         [,10]
## humid_kitchen  1.978820e-05  7.719533e-04
## humid_living   4.603929e-04  5.965634e-04
## humid_laundry  8.233675e-04  2.640419e-03
## humid_office   9.324076e-05  2.489772e-04
## humid_bath     4.325625e-04 -2.099882e-04
## humid_outside -2.754602e-05 -4.148075e-06
## humid_iron    -2.102964e-03  2.371479e-04
## humid_teen    -9.355456e-04 -1.505144e-03
## humid_parents  2.536496e-03 -1.813878e-03
## humid_station -5.564835e-05  5.025773e-05
##
## $xcenter
## temp_kitchen  temp_living temp_laundry  temp_office    temp_bath
temp_outside
##    21.686571    20.341219    22.267611    20.855335    19.592106
7.910939
##    temp_iron    temp_teen temp_parents temp_station
##    20.267106    22.029107    19.485828     7.411665
##
## $ycenter
## humid_kitchen  humid_living humid_laundry  humid_office    humid_bath
##      40.25974      40.42042      39.24250      39.02690      50.94928
## humid_outside    humid_iron    humid_teen humid_parents humid_station
##      54.60908      35.38820      42.93617      41.55240      79.75042
```

Each of the canonical correlation coefficients represents the strength of the relationship between the sets of variables, but for a different pair of linear combinations of the variables.

Correlation of first 2 canonical - 0.94 0.90

There was a strong correlation. Which shows that the first linear combination from each set of variables is strongly correlated, and the second linear combination from each set of variables is also pretty strongly correlated. In other words, the temperature variables are highly related to the humidity variables.

Conclusion:
The research successfully developed predictive models for appliance energy consumption based on temperature, humidity, and weather data in a residential setting. The findings highlighted key influencing factors like room temperatures, humidity levels, outdoor conditions, and wind speed. By addressing statistical challenges, the analysis provides a robust model to aid energy management and potential cost savings for homeowners.

However, the study has limitations, including a short data collection period and exclusion of variables like occupancy patterns and appliance usage habits. Future work could involve longer data collection, incorporating additional variables, and validating models across multiple residential settings.

Overall, this research contributes to understanding residential energy usage and offers a foundation for optimizing energy consumption, reducing environmental impact, and promoting energy-efficient practices in homes.

References:

[1] Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, *140*, 81–97. https://doi.org/10.1016/j.enbuild.2017.01.083

[2] Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings*, *40*(3), 394–398. https://doi.org/10.1016/j.enbuild.2007.03.007