Experimental Analysis and Interpretation of Electrical Energy Consumption

CMPT318 - Group 12 - Spring 2024
Brian Le - 301437410
Arthur Li - 301456006
Trevor Favel - 301420619
Aditya Kulkarni - 301437152

Outline

- Introduction
- Experiments
- Results
- Lessons Learned

Introduction

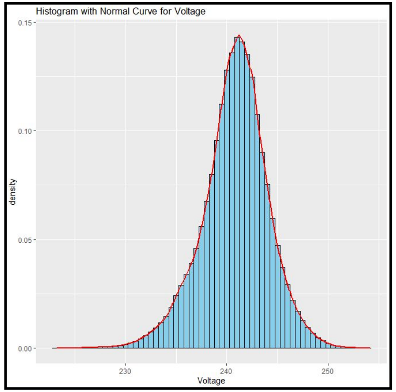
- The team was assigned to build an automated anomaly detection system using the data generated by an electronic power system.
- Increasing dependence on supervisory control systems in critical infrastructure.
- Escalating risks of cyber-attacks due to expanded automation.
- Need for anomaly detection-based intrusion detection techniques for enhancing cybersecurity.

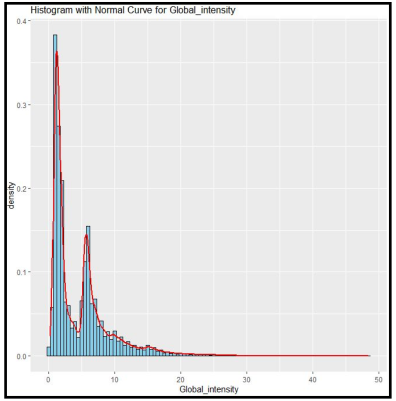
Experiments and Methodologies: Data Cleaning

- Data cleaning: interpolation or elimination
  - Linear Interpolated every NA value
  - Linear Interpolated time(date, time), elimination on numeric columns
  - Elimination on every NA value

Experiments and Methodologies: Feature Scaling

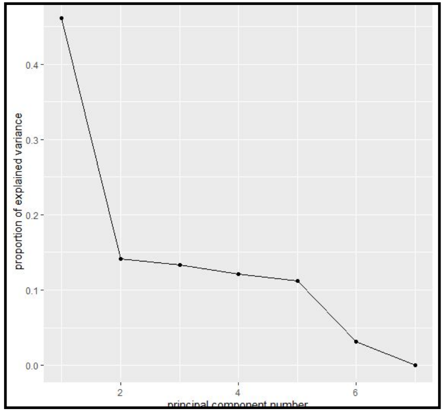
- It is crucial preprocessing step in machine learning, ensuring that each feature contributes equally to model training regardless of its scale or units of measure.
- Normalization or Standardization
- Observation on histogram





Experiments and Methodologies: Feature Engineering

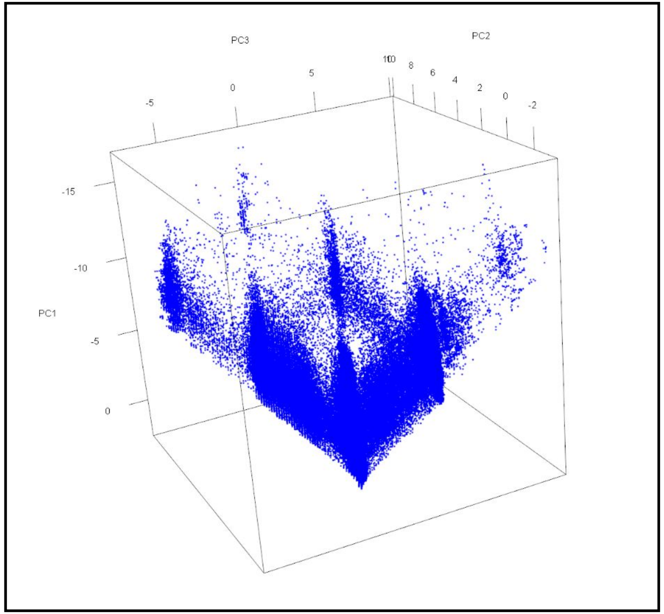
- PCA and response variable
- Coverage and proportion: threshold

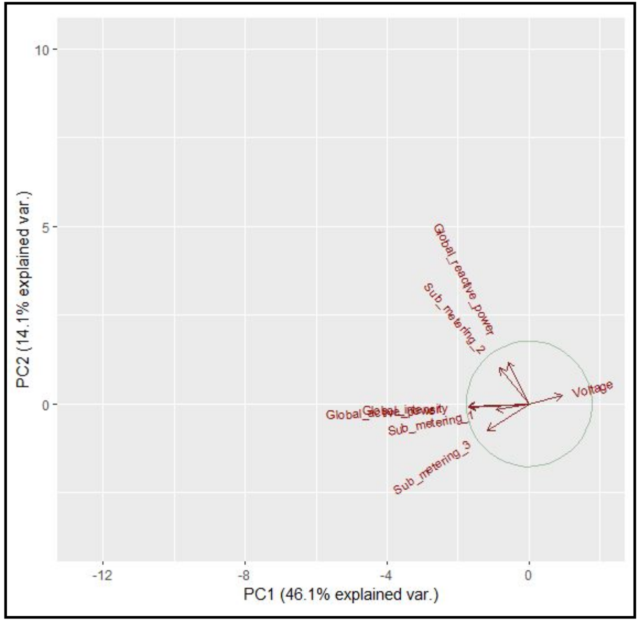



```
> summary(pc)
Importance of components:
                          PC1    PC2    PC3    PC4    PC5     PC6     PC7
Standard deviation     1.7957 0.9947 0.9657 0.9211 0.8868 0.46672 0.02714
Proportion of Variance 0.4607 0.1413 0.1332 0.1212 0.1123 0.03112 0.00011
Cumulative Proportion  0.4607 0.6020 0.7352 0.8564 0.9688 0.99989 1.00000
```

```
Rotation (n x k) = (7 x 7):
                           PC1         PC2         PC3         PC4         PC5         PC6          PC7
Global_active_power   -0.5381097 -0.05112019  0.02978743 -0.04476800  0.15868069  0.42622395  0.705752668
Global_reactive_power -0.1867954  0.67037848 -0.34034177  0.62679327 -0.02649741 -0.07819877  0.013441617
Voltage                0.2961116  0.14239228 -0.05952547 -0.04939298  0.93956025  0.05646484 -0.009883380
Global_intensity      -0.5398184 -0.03167064  0.01911187 -0.04194898  0.14069402  0.42915594 -0.708165563
Sub_metering_1        -0.2953051 -0.10046871 -0.74088814 -0.43078174  0.06308568 -0.40525145  0.002067051
Sub_metering_2        -0.2660841  0.57326777  0.50049983 -0.46511374  0.02618255 -0.36474801  0.002411597
Sub_metering_3        -0.3720936 -0.43355614  0.28276549  0.44611987  0.25862083 -0.57235055 -0.011236558
```

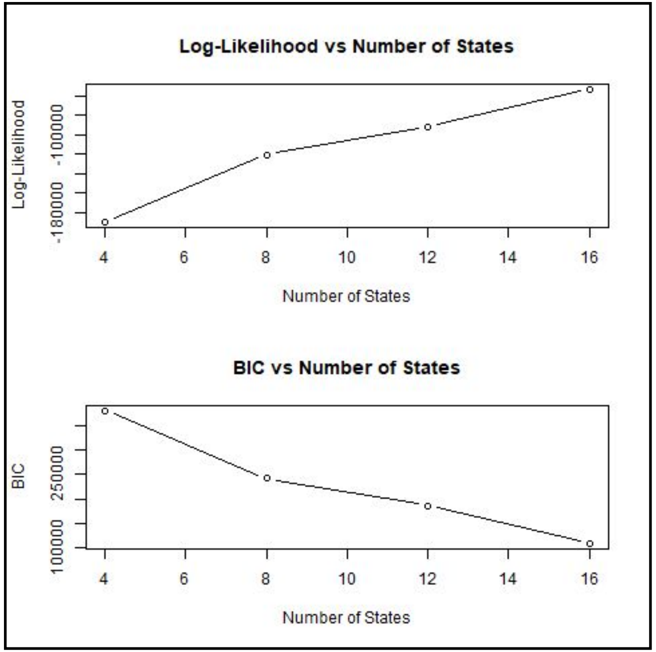
Experiments and Methodologies: Feature Engineering





Experiments and Methodologies: HMM Training and Testing

- Response variables chosen from PCA
- Data partition: time window selection
- Log-likelihood and BIC
  - Experimented with states 4-20
  - Optimal model: 16 states



Experiments and Methodologies: Anomaly Detection

- Log likelihood threshold for normal behavior
    - 10 subsets of consecutive test data weeks

```
Log likelihood threshold:  0.6288294
Log likelihood threshold:  0.745193
Log likelihood threshold:  0.03925129
Log likelihood threshold:  0.2353644
Log likelihood threshold:  0.4143268
Log likelihood threshold:  0.03202562
Log likelihood threshold:  0.8842873
Log likelihood threshold:  0.324022
Log likelihood threshold:  0.4648408
Log likelihood threshold:  0.5080238
> cat('Log likelihood threshold: ', max(subsetLogsDev))
Log likelihood threshold:  0.8842873
> 
```

Lessons Learned

- Observation and interpretation through statistical analysis such as featured scaling and engineering, alongside data cleaning—are equally important
- Careful observation of the impact of various methods for cleaning and scaling raw data
- What makes an HMM effective for anomaly detection

The End
Q&A session