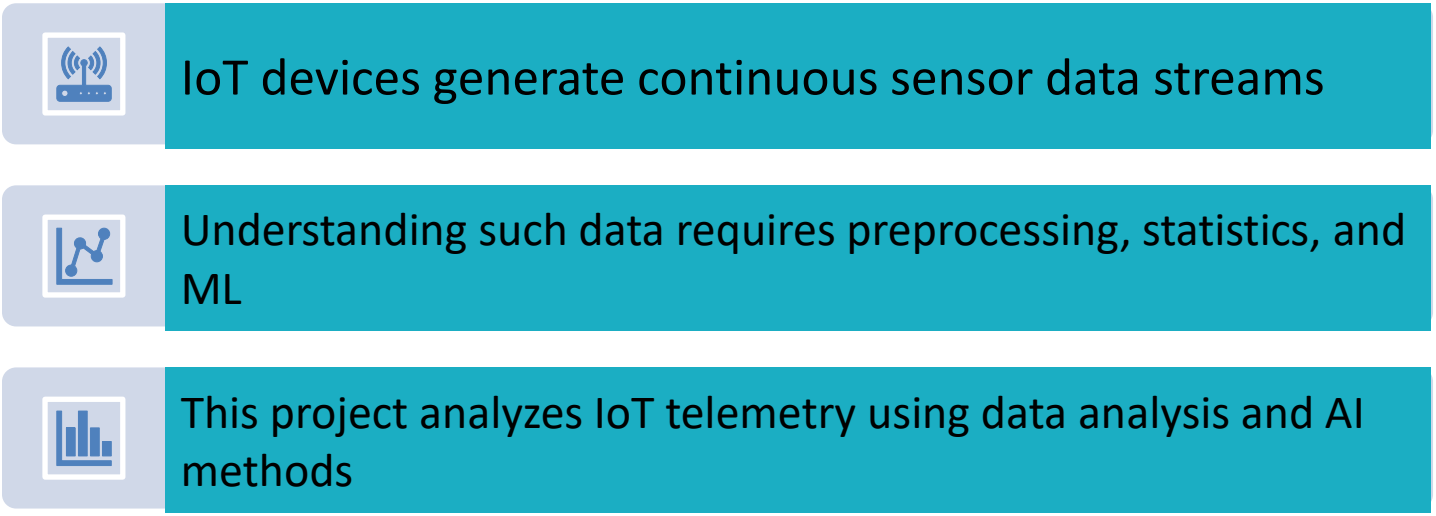
FH | JOANNEUM
Electronic Engineering

IoT Telemetry Data Analysis with Statistics & Machine Learning

Team: Fantastic Three
Group Members:
Md Ehtashamul Huque
Md Shamsur Rahman Shishir
Sheikh Md. Nayeem

Introduction



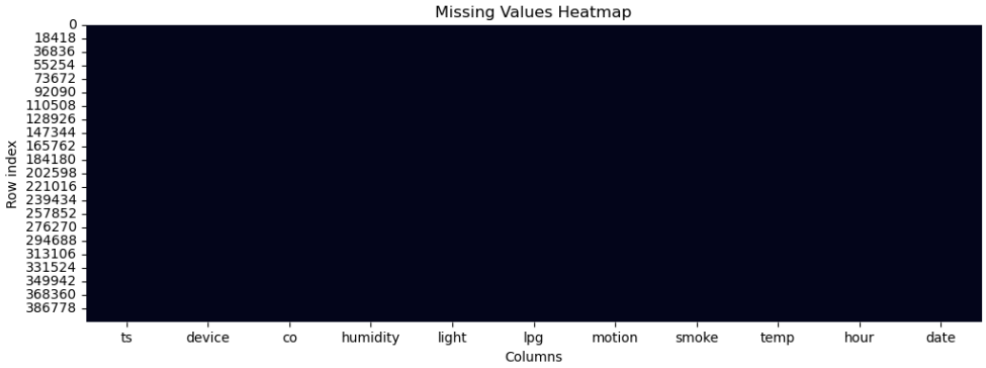
Dataset Description

- ❑ Time-stamped IoT telemetry data from multiple devices
- ❑ Sensors: temperature, humidity, CO, smoke, LPG
- ❑ Events: motion detection and light state

Task A Data Preprocessing

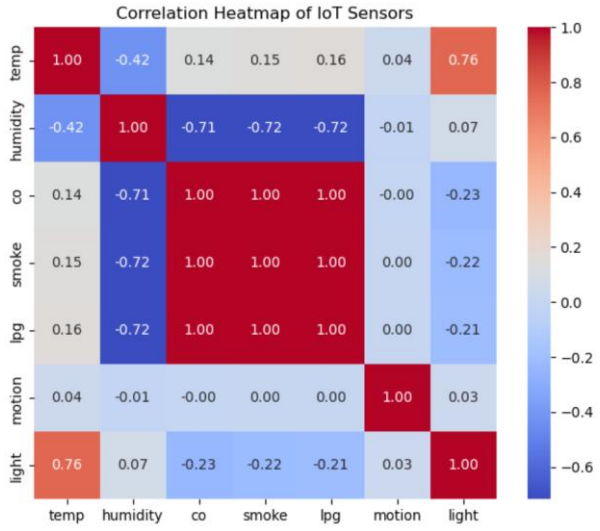
- Converted timestamps and sorted data chronologically
- Handled missing values using forward/backward filling
- Outliers reduced using quantile-based clipping
- Encoded devices and standardized numeric features

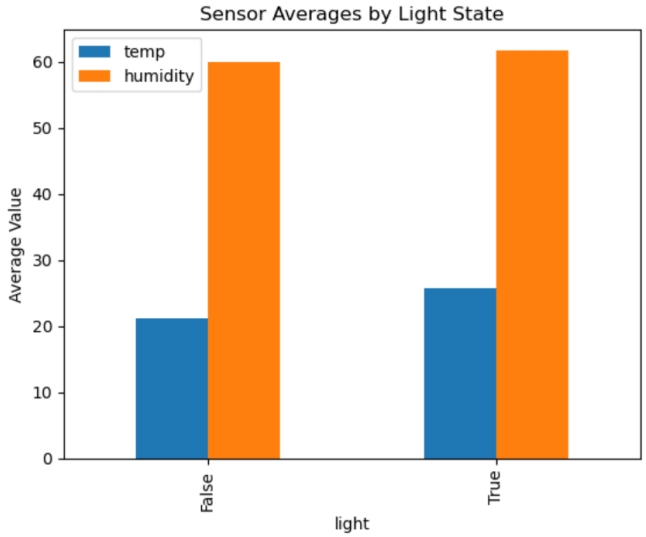
	missing_count	missing_percent
ts	0	0.0
device	0	0.0
co	0	0.0
humidity	0	0.0
light	0	0.0
lpg	0	0.0
motion	0	0.0
smoke	0	0.0
temp	0	0.0



Task B Exploratory Data Analysis

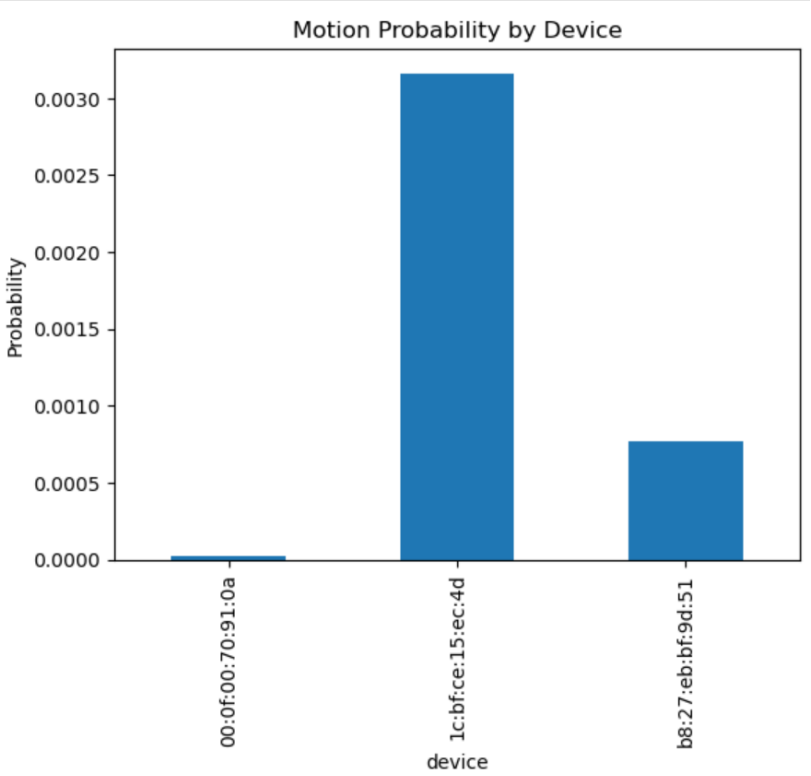
- Time-series plots revealed trends and daily cycles
- Correlation analysis showed strong relationships among gas sensors
- Event-based analysis linked motion and light to sensor behavior

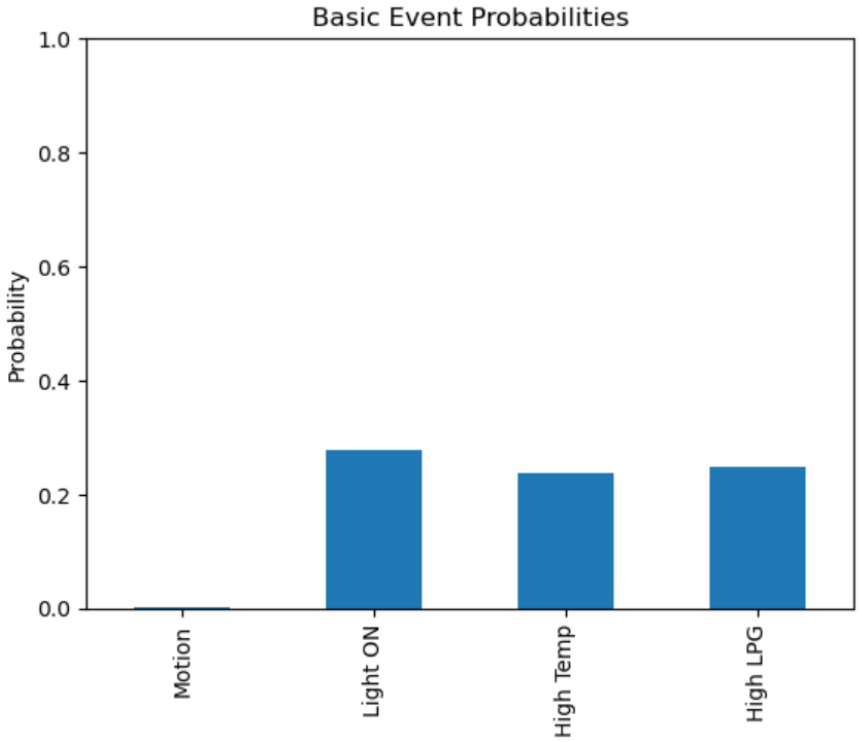




Task C Probability & Event Analysis

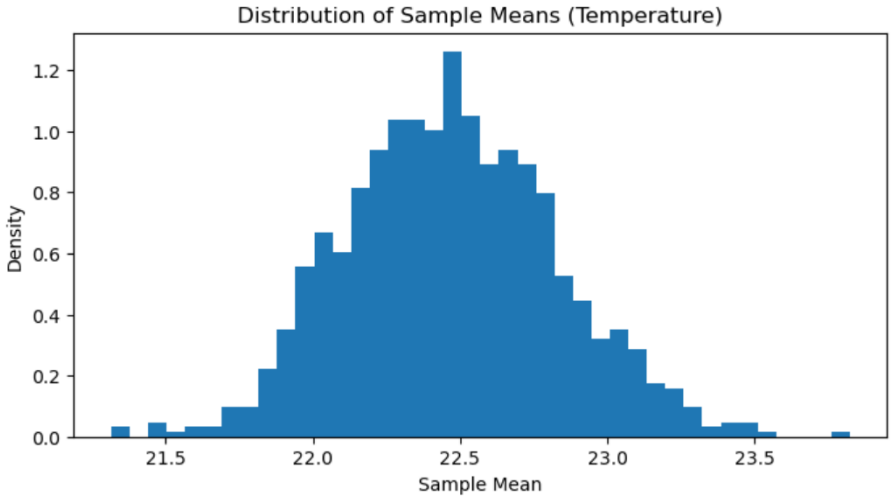
- Defined events such as motion, light, and high temperature
- Computed event and conditional probabilities
- Verified Bayes' rule using empirical data
- Tested event independence using chi-square tests

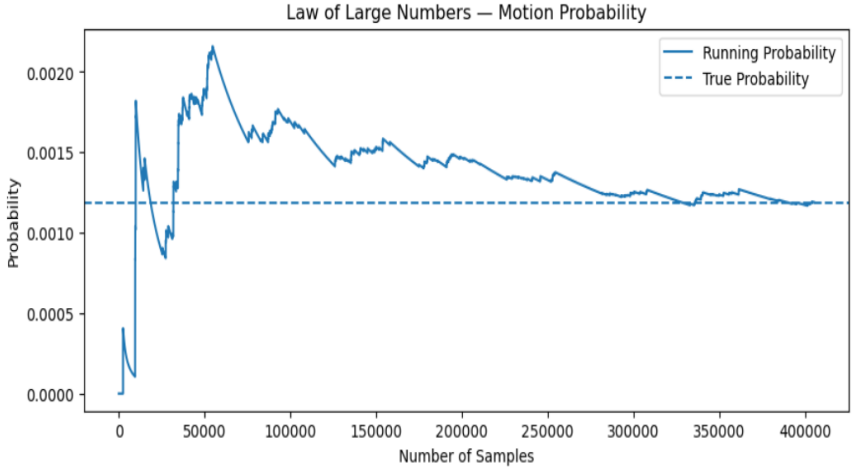




Task D Statistical Theory Validation

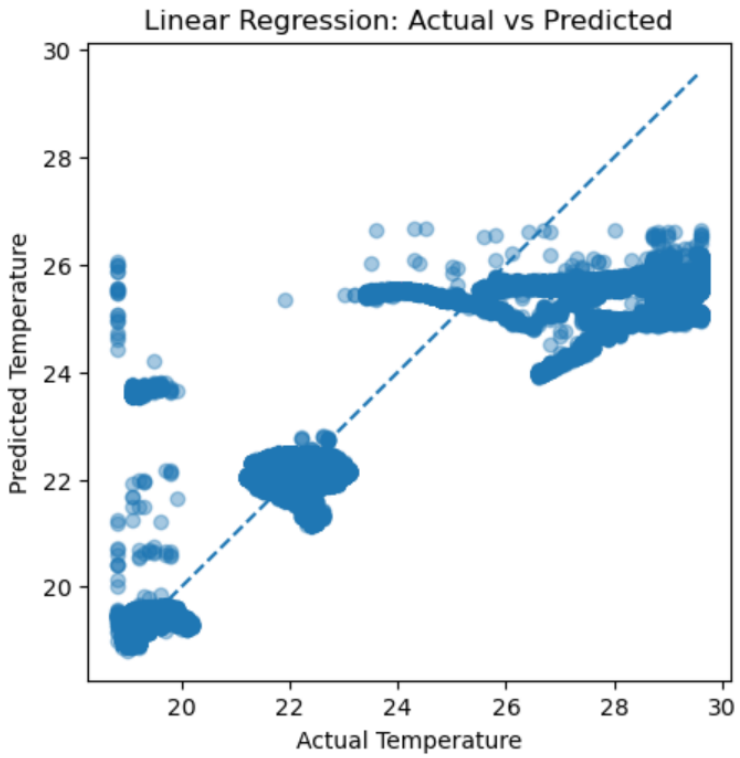
- Law of Large Numbers demonstrated using running averages
- Central Limit Theorem shown via sampling distributions
- Confidence intervals computed for key parameters

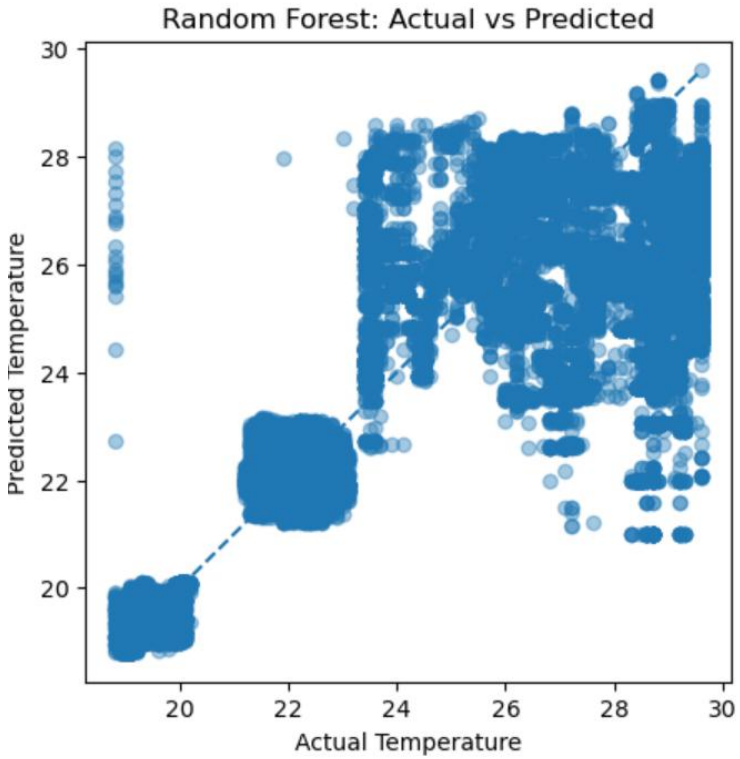




Task E Machine Learning

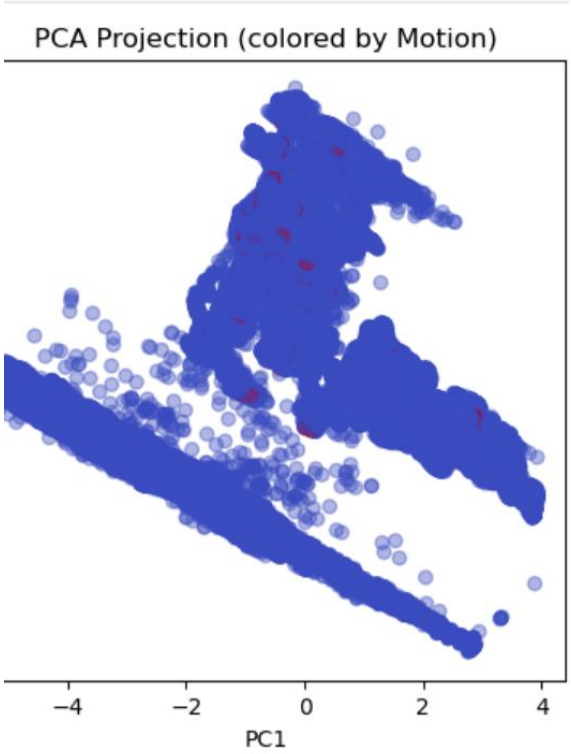
- Regression task: predict temperature from sensors and events
- Models: Linear, Ridge, Lasso, Random Forest
- Time-aware train-test split used
- Random Forest achieved best performance

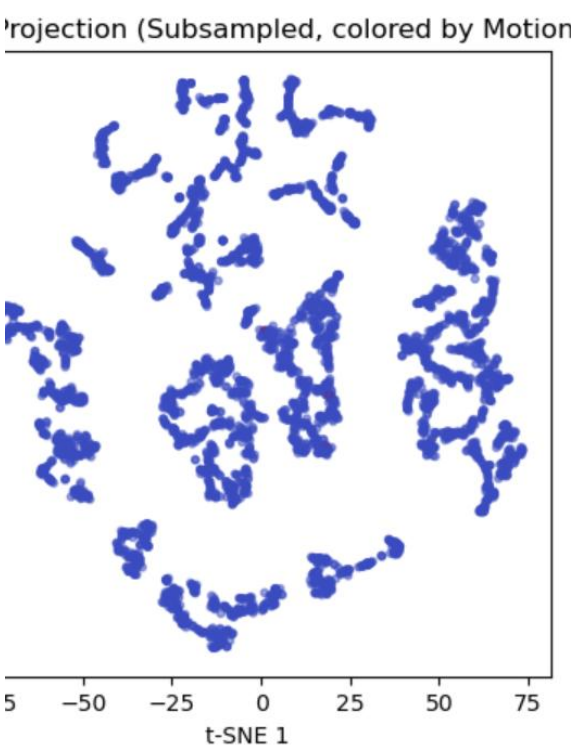


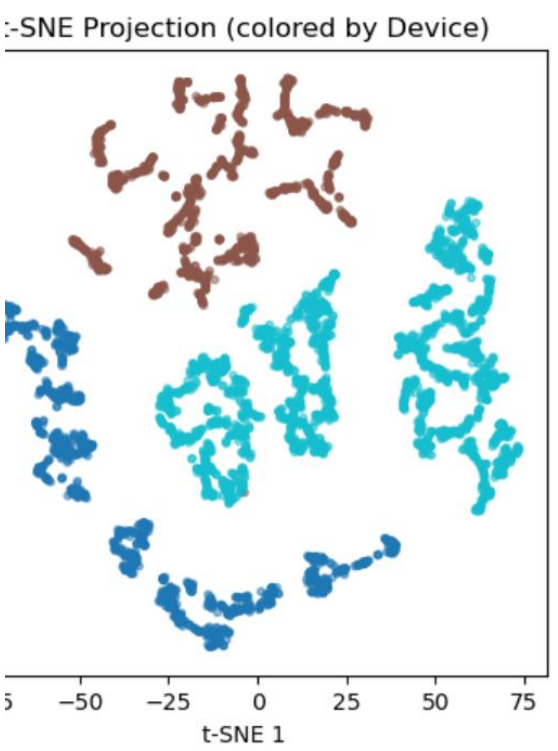


Task F Dimensionality Reduction

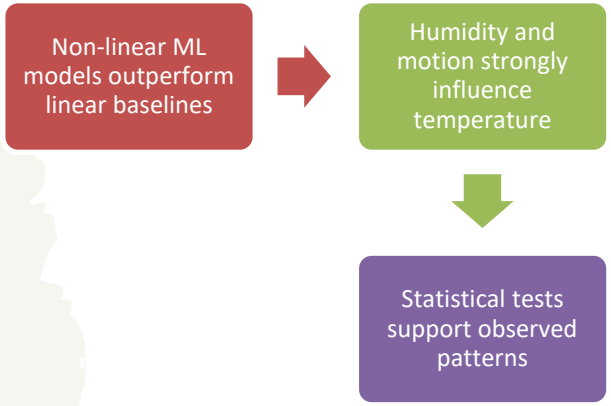
- PCA reduced dimensionality while preserving variance
- t-SNE used on a subsample to visualize non-linear structure
- Hypothesis testing validated observed relationships







Key Results



Limitations & Future Work

No external weather data included

Temporal models not explored

Future work: LSTM, anomaly detection, real-time systems

Conclusion

End-to-end IoT data analysis successfully completed

Statistics and ML combined for robust insights

Project demonstrates practical data science skills