VIT-AP

UNIVERSITY

Prof: Dr.D.Sumathi

Slot : A2+TA2

<u>Submitted by:</u>

22BCE9938 – S. Lalith Kumar

22BCE8033 – N. Manideep

22BCE8436 – K.V.V.Jitin

Smart Home Dataset With Weather Forecast Dataset

Abstract—**The integration of smart sensor data from the home with external weather information represents an innovative approach to energy consumption patterns predictive analytics, optimizing the efficiency of automation in a home environment. Within this paper, various sources of real-time data, such as indoor temperature, humidity levels, appliance energy consumption, and weather conditions outside, are combined in the context of developing machine learning models able to analyze and predict the dynamics of energy usage. By using this information, the proposed framework provides insights on the level at which outdoor environmental factors impact indoor energy demands and, therefore, high correlations between external weather patterns and residential energy consumption behavior. Preliminary findings indicate that strong relations do indeed lie between meteorological conditions like temperature fluctuations, moisture variations, and other behavioral patterns and indoor energy requirements. These results have high potential for the use of predictive modeling in smarter energy management to bring about improved energy savings with less waste and a more sustainable home energy management approach. This research sets foundations for future frontier developments in intelligent energy automation that contribute toward environmental sustainability as well as greater household energy efficiency.**

Keywords—*Smart home, energy forecasting, Deep learning, data analysis, weather impact*.

I. Introduction

The integration of external weather forecasting into smart home technologies has, in fact, emerged as a promising avenue for energy-consuming efficiency through the current drive for sustainable living and energy efficiency. This synergy ensures that decisions made in real time optimize energy management, particularly in the residential sector, since well-operational carbon emissions are the main source of climate-changing emissions.[7] Current studies show how important the role machine learning algorithms play in predicting energy usage, by adapting to data from sensors and from other external, variable conditions. For example, Lasso Regression and fuzzy decision matrices have had proven efficiency in discovering energy consumption patterns. However, one problem persists: different algorithms perform differently on different datasets [1].

1D-CNN-based classifiers stress the need to overcome challenges like class imbalance and concept drift on dynamic streams of weather data. These models boost prediction stability, especially on imbalanced datasets, along with superior performance metrics in terms of recall and F1 scores[2]. Similar innovation techniques such as Edge Computing coupled with machine learning have allowed the real-time electricity demand forecasting, reducing latencies and inefficiencies associated with these distributed cloud-based systems[4].

Leveraging this research, this study aims to develop a machine learning-based framework that integrates real-time smart home sensor data (temperature, humidity, and appliance energy usage) with external weather information. By understanding the correlations between weather patterns and energy demand, we propose to enhance energy efficiency while promoting sustainable living. Our methodology not only improves the accuracy of models for the prediction of energy but also improves key gaps, which include incorporating localized microclimatic variations and balancing energy management strategies.[3]

This study is potential for practical application in intelligent home automation, revealing ways to optimize energy usage while minimizing wastage and ensuring a more sustainable future.

II. Literature Survey

There has been research to delineate the relationship between external weather conditions and smart home energy consumption based on improved energy efficiency and sustainability. In recent times, tremendous advancements have been made with regard to understanding and forecasting the patterns of energy consumption through machine learning and data analytics[6].

Studies like Iram et al.'s, have used machine learning algorithms Lasso Regression which predict energy usage by combining weather input data with appliance-wise consumption profiles. In this study, it developed fuzzy decision algorithms to rank algorithms and had addressed the issues of having optimum predictive models based on changing weather [1].Alex et al. developed the Self-Organizing Auto-Encoder-based 1D-CNN framework to address the class imbalance and concept drift phenomena in weather data streams. In this work, deep learning techniques result in higher accuracy and recall values and thus are robust to changing data distributions[2].

Localized microclimates also revealed the discovery aspect of weather data, with Mehmood et al. working on its impact for urban energy consumption models.

Their study bridges the existing gaps between preceding researches because they studied actual real-time microclimatic changes instead of subtropical conditions at night, thereby making their predictions more accurate in terms of urban residential areas[3]. addition, Edge Computing in smart homes succeeded in solving latency and scalability issues which is most common case with cloud-based systems. Short-term electricity demand forecasting through XGBoost and LightGBM made real-time prediction along with a decrease in network congestion[8].Although these studies are very informative, the harmonization of different datasets and comprehensive weather metrics still has gaps in relating diverse datasets to one another. Existing research often focuses on isolated aspects of energy consumption or weather data, thus overlooking the interdependencies between indoor environments, appliance usage, and external weather conditions. By overcoming such challenges, future studies can leverage advanced algorithms and integrated frameworks to create smarter, more adaptive energy management systems for sustainable living.

III. DATASET OVERVIEW

This dataset creates an enriched set of data streams that articulate the interconnected dynamics of indoor climate control, the usage of appliances in energy consumption, and external conditions in weather. Climatic needs by the home are well depicted through indoor temperature and humidity data; insights can be gained into how HVAC units respond to different conditions of climate. Analysis of these factors allows one to pinpoint energy usage patterns and find chances to optimize settings for maintaining comfort while curbing energy waste.

Thus, these data add more granular detail to appliance-specific energy consumption data, promoting further insight into how some particular device is causing increases in energy demand. For instance, outside temperature and humidity can have such massive variations in HVAC system usage. In arrears of other data regarding appliances, such approaches are set to better target strategies toward improving energy efficiency-through scheduling heavy-use devices for off-peak hours, for instance, or through smart automation systems that would adjust usage according to expected future conditions.

Outdoor weather data, including air temperature, humidity, wind speed parameters, and information regarding the weather, forms the prime external data required for effective forecasting and modeling energy. Time-stamped records further elaborate on the dataset and allow the study of temporal variations in daily consumption patterns and seasonal patterns. This dataset, developed by a collaboration of researchers and practitioners, contributes toward creating predictive models that improve energy management in smart homes. The models can ensure adaptability in balancing comfort with efficiency while serving as a base for sustainable living solutions.

IV. METHODOLOGY

A. Data Preprocessing

It comprises 5,03,912 rows and 24 columns.

Data Cleaning: removal of outliers and missing value handling assures the data's credibility.

Normalization: scales features, needful to compare within the model; thus, the more accurate a model.

Feature Engineering: new features like temperature variation, humidity gradients, and usage time slots are introduced, which, in turn increases the accuracy of the model predictions.

.

B. Exploratory Data Analysis (EDA)

Correlation Analysis: this determines if outdoor temperature and humidity correlate with indoor energy usage
Trend Analysis: to understand the peak usage times in terms of daily or seasonal changes.

C. Model Selection and Training

Trained Linear Regression, Decision Trees, and finally, LSTM for sequence prediction in order to determine which model would perform well for predicting outdoor temperatures and humidity and subsequently energy usage,

Divided training sets to see how well models could predict temperatures, humidity, and energy consumption.

D. Model Evaluation

For the evaluation, Mean Absolute Error, Root Mean Square Error, and R-squared were used.

LSTM is more accurate than other models for predicting indoor conditions using weather patterns.

V. RESULTS AND DISCUSSION

Results show that outdoor weather significantly affects indoor energy demand:

Temperature Differences: Outdoor general low temperatures result in increased heating needs indoors.

Humidity Dependency: Devices such as dehumidifiers are conditionally dependent on outdoor as well as indoor humidity and hence energy consumption is affected overall.

Model Performance: LSTM outperformed and gave better results with low RMSE, thus proper to be applied in smart homes for time-dependent energy prediction.

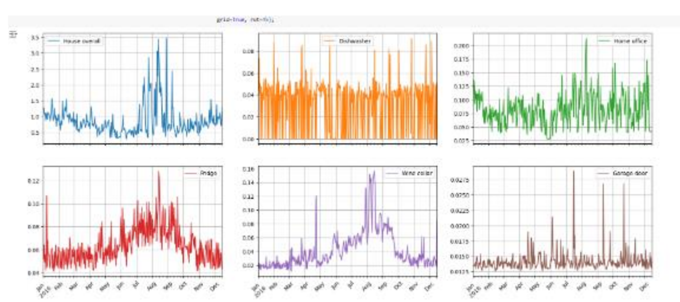
This research findings propose that the real-time weather forecasts can enhance home energy management by pre-emptive changes of appliances according to condition.

1. Data Pre-processing: Importing the data, cleaning (removing rows with invalid information), preparing the features (like converting time stamps and renaming columns).

2. Data Transformation: Summing up duplicate columns (for example, the same appliances), changing units, and looking at data types.

3. Data Visualization: Making visualizations in order to get insights into the relationships between the elements.

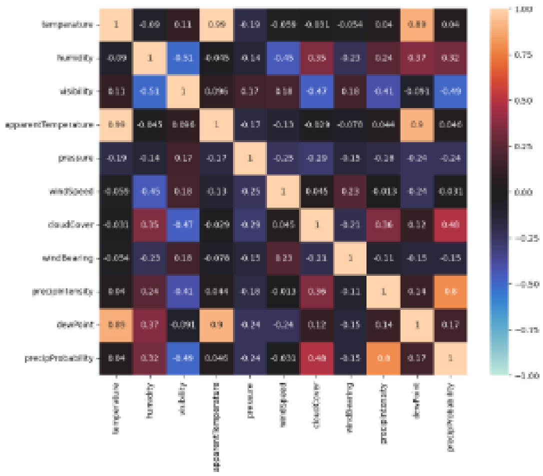
#Visualize wheather data:



.

4. Correlation Analysis: Checking relationships among other features.

Weather correlation:



5. Time-series Analysis-Patterns in time.

6. Anomaly detection-detecting unusual value or behavior.

7. Time-series forecasting-model based prediction of future data-points.

6. Acknowledgments:

This work was supported by Kaggle for DataSet.

VI. Conclusion

The integration of weather data into smart home energy prediction models can provide better capabilities to manage energy usage. In this article, we have shown that there exist substantial correlations between weather and energy demand by fitting machine learning models. For further research, real-time model application for real applications will be researched upon by extending the parameters of weather.

Model	**MSE**	**RMSE**	**MAE**	**MAPE**	**R^2**
Baseline	0.071	0.266	0.177	0.236	0.077
Arima_basic	0.259	0.509	0.463	0.722	-2.379
Arima_dynamic	0.069	0.263	0.176	0.229	0.094
Sarima	0.107	0.327	0.266	0.397	-0.399
Sarimax	0.101	0.317	0.243	0.363	-0.319
LSTM_Univar	0.068	0.261	0.173	0..307	0.106
LSTM_Multivar	0.022	0.150	0.110	0.173	0.700

Signigicance of Models:

1. Baseline Model

What It Is: A simple model used as a benchmark for evaluating the performance of more complex models.

Example: Last observed value (Naïve), Mean of previous values, or a simple moving average.

Significance:

Helps to set a performance threshold; more advanced models should outperform the baseline.

Quick and easy to implement for initial insights.

2. ARIMA (Autoregressive Integrated Moving Average)

What It Is: A statistical model combining autoregressive (AR), differencing (I), and moving average (MA) components to capture patterns in stationary time series.

Significance:

Effective for univariate time series with non-seasonal data.

Relies on assumptions of stationarity (differences in trends removed).

Widely used for its simplicity and interpretability.

3. ARIMA Basic vs. ARIMA Dynamic

ARIMA Basic:

Forecasts future values using a pre-trained ARIMA model.

Does not update itself based on new incoming data during the forecast period.

Significance:

Suitable for scenarios where historical trends are stable.

Limited adaptability to sudden changes in the data.

ARIMA Dynamic:

Updates predictions dynamically by incorporating actual observed values as they become available.

Significance:

Better adaptability for real-time forecasting.

More accurate for time series with sudden shifts or evolving patterns.

4. SARIMA (Seasonal ARIMA)

What It Is: Extends ARIMA by adding seasonal components to handle periodic patterns.

Seasonal terms: AR, MA, and differencing components specific to the seasonal cycle.

Significance:

Ideal for time series with strong seasonal trends (e.g., monthly sales, temperature data).

Captures both regular and seasonal patterns effectively.

5. SARIMAX (Seasonal ARIMA with Exogenous Variables)

What It Is: SARIMA with an added capability to include exogenous variables (independent predictors) that influence the target variable.

Significance:

Incorporates external factors like promotions, holidays, or economic indicators.

Useful when the time series is influenced by variables outside of its own historical values.

6. LSTM Univariate

What It Is: A Long Short-Term Memory (LSTM) neural network trained to predict a single variable's future values based solely on its past values.

Significance:

Handles nonlinear patterns and long-term dependencies better than ARIMA/SARIMA.

Suitable for non-stationary and complex time series data.

Requires substantial data and computational resources.

7. LSTM Multivariate

What It Is: An LSTM model trained to predict a variable's future values using multiple correlated variables as input (multivariate data).

Significance:

Captures relationships between the target variable and influencing variables.

Useful for time series with complex interactions between multiple features.

Computationally intensive and requires a good understanding of feature engineering.

VII. REFERENCES

[1]. Iram, Shamaila, et al. "An innovative machine learning technique for the prediction of weather based smart home energy consumption." ***IEEE Access*** **11 (2023): 76300-76320.**

[2]. Sharma, Navin, et al. "Predicting solar generation from weather forecasts using machine learning." ***2011 IEEE international conference on smart grid communications (SmartGridComm)*****. IEEE, 2011.**

[3]. Alex, Suja A., Uttam Ghosh, and Nazeeruddin Mohammad. "Weather prediction from imbalanced data stream using 1d-convolutional neural network." ***2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22)*****. IEEE, 2022.**

[4]. Indikawati, Fitri Indra, and Guntur Maulana Zamroni. "Household power consumption forecasting using IoT smart home data." ***Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*** **5.1 (2019): 8-15.**

[5]. Naudiyal, Anjali, et al. "A Review Analysis: Comparative Study On Various Machine Learning Techniques for Load Forecasting In Electric Power Distribution System with Multiprocessing." ***2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*****. IEEE, 2023.**

[6]. Zhou, Sha, and Lei Zhang. "Smart home electricity demand forecasting system based on edge computing." ***2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*****. IEEE, 2018.**

[7]. Jukaria, Mukta, Ashutosh Upadhyay, and Anuj Kumar. "Vector Auto-Regression-Based Predictive model for Smart Meter." ***2023 International Conference on Device Intelligence, Computing and Communication Technologies,(DICCT)*****. IEEE, 2023.**

[8]. Brahim, Ghassen Ben. "Weather Conditions Impact on Electricity Consumption in Smart Homes: Machine Learning Based Prediction Model." ***2021 8th International Conference on Electrical and Electronics Engineering (ICEEE)*****. IEEE, 2021.**

GitHub link for the Project:

https://github.com/LalithSiramdasu/Smart-Home-Dataset-With-Weather-Forecast

Submitted To:

Dr.D.Sumathi
Professor Grade-1
VIT-AP University
9942887233/8778542153
sumathi.d@vitap.ac.in