Fuzzy Systems and Soft Computing
ISSN : 1819-4362

EVALUATING LINEAR FACTOR MODEL FOR ENERGY CONSUMPTION FORECASTING: INSIGHTS & LIMITATIONS

Dr.Minal Atul Sarode, Assistant Professor, Dept of Information Technology Sies(Nerul)College of arts, Science & Commerce(Autonomous) :minals@sies.edu.in

ABSTRACT:
The escalating global energy demand necessitates advanced management systems to optimize consumption and enhance efficiency. Information Technology (IT)-driven Energy Management Systems (EMS) have emerged as pivotalsolutions, leveraging computational models to monitor, control,andpredictenergyusage.ThispaperdelvesintotheapplicationoftheLinearFactorModel (LFM) withinIT-drivenEMS, aiming to elucidate its efficacyin forecasting energyconsumption patterns and facilitating informed decision-making. A comprehensive literature review underscores the evolution of EMS, highlighting the integration of linear models and IT frameworks. Acasestudyispresented, whereintheLFM is implementedinaresidentialbuilding equipped with smart meters and IoT devices. The methodology encompasses data collection, modelformulation, and validationprocesses. Findings indicatethat the LFM effectivelycaptures the relationship between various factors influencing energy consumption, such as temperature, occupancy, and appliance usage. The results demonstrate a significant improvement in energy forecasting accuracy, leading to optimized energy distribution and cost savings. The paper concludes byaffirming the potentialofLFMs inenhancing IT-drivenEMS and suggests avenues for future research, including the integration of machine learning techniques to further refine predictive capabilities.
Keywords:LinearFactorModel,energyconsumption,MSA,MSE,Featureselection

INTRODUCTION:

The surge in global energy consumption has prompted the development of sophisticated Energy Management Systems (EMS) that harness Information Technology (IT) to monitor, control, and optimize energy usage. These systems are instrumental in achieving energy efficiency, reducing operational costs, and minimizing environmental impacts. A critical component of EMS is the ability to accurately forecast energy demand, which facilitates effective planning and resource allocation. A Linear Factor Model (LFM) is a mathematical framework used to describe the relationship between a dependent variable and multiple independent factors. It assumes that the dependent variable can be expressed asa linear combinationofseveralfactors,eachweighted by a corresponding coefficient. The general form of an LFM is:$Y=\beta 0+\beta 1X1+\beta 2X2+...+\beta nXn+\epsilon$
...................(1)
Where: Y= Dependent variable (outcome) X1,X2,
.... ,Xn=Independentfactors(predictors)
$\beta 0$=Intercept(constantterm)
$\beta 1,\beta 2,,....,\beta n$=Factorcoefficients(weights)
ϵ=Errorterm(unexplainedvariation)
The major reason why LFMs have gained importance is due to its features of being simple & interpretable: LFMs provide a straightforward way to quantify the impact of each factor on the outcome.Second feature being its predictive capacity in forecasting, risk analysis, and decision-makingacrossvariousfields.AssumptionofLinearityisyetanotherwhichassumesalinear

relationship, which may not always hold in complex systems. LFMs are widely used in finance (asset pricing), economics, machine learning, and social sciences.
ThusmakinguseofLFMtopredict energyconsumptionbyanalyzingvariousinfluencing factors suchastemperature,occupancy, applianceusage, andelectricitypricing istheaimofthestudy.In this context the LFM model is framed as
$E_t=\beta_0+\beta_1 T_t+\beta_2 O_t+\beta_3 A_t+\epsilon_t$ ....(2)
Where:

- E_t=Energyconsumptionat timet
- T_t=Temperature
- O_t= Occupancy levels
- A_t=Appliance usage
- $\beta_0,\beta_1,\beta_2,\beta_3$=Modelcoefficients
- ϵ_t=Errorterm

LITERATUREREVIEW:

The integrationofInformationTechnology(IT)inEnergyManagement Systems(EMS)hasbeen extensively studied, leading to the development of various models aimed at improving energy forecasting and optimization. Zhou et al. (2016) explored data-drivensmart energymanagement, emphasizing the role of big data analytics in refining energy consumption strategies. Similarly, Diamantoulakis et al. (2015) examined big data techniques for dynamic energy management in smart grids, highlighting their potential to enhance operational efficiency.

Linear regression has been widely applied in predictive modeling to estimate energy demand. Manojpraphakar andSoundarrajan(2020) demonstratedtheeffectivenessofregressionmodels in capturing real-time energy consumption trends. Additionally, mixed integer linear programming (MILP) has been utilized in home energy management systems. Putra et al. (2019) proposed an MILP-based model incorporating dynamic pricing strategies to optimize electricitycosts, further demonstrating the applicabilityoflinear models in EMS.Zervaset al. (2019)investigated hybrid power systems integrating photovoltaics, electrolyzers, and fuel cells, comparing them to standalone photovoltaic setups. Meanwhile, El-Shater et al. (2019) focused on energy flow management in hybrid fuel systems, and Jurado and Saenz (2018) implemented a neuro-fuzzy controllerforwind-dieselhybridenergysystems.Onaretal.(2017)designedcontrollersforhybrid power systems featuring wind turbines, photovoltaics, fuel cells, and supercapacitors.

Kabasi(2018)analyzedahybridwind-photovoltaicsystemwithproportional-integral(PI)control, utilizing maximum power point tracking for optimized photovoltaic energy conversion. Sinha et al. (2019) conducted a comprehensive review of hybrid renewable energy systems, covering design,sizing,andanalyticaltools.Sharifietal.(2020)appliedmulti-objectiveoptimizationusing acumulativeparticlealgorithmtoahybridsystemincorporatingwindturbines,photovoltaics,and hydrogen storage. Izadyar et al. (2020) assessed renewable energy resource capacities by considering geographic and technical factors for microgrid power generation. Sanajaoba et al. (2021) utilized the cuckoo algorithm to optimize microgrid components, factoring in cost and reliability metrics.

FurtherstudiesbyRajannaetal.(2022)employedamulti-objectivegeneticalgorithmtooptimize powerdistributionacrossagridwithfouroperationalareas,thoughwithoutaddressingtotalpower balance. Upadhyay et al. (2021) introduced a bird-inspired optimization algorithm for power allocation in microgrids, integrating diesel generators and battery banks into the control framework.Maet al.(2022)exploredpowermanagement inenergystoragesystems,proposinga circuit that integratesbatteriesand supercapacitors for efficient energydistribution. Collectively, these studies underscore the expanding role of IT-driven models in energy management, reinforcing the significance of Linear Factor Models (LFMs) in optimizing energy consumption and ensuring grid stability.

CASESTUDY:

The study involves use of two different datasets: a synthetic dataset: which is a dummy organizations energy consumption details produced synthetically anda kaggle dataset :This dataset provides a comprehensive overview of electricity-related metrics, environmental conditions, and additional influencing factors, spanning from 2016 to 2022.

1. **DummyDataset:** Sincewedonothavereal-world energyconsumptiondata,wegenerate a synthetic dataset using numpy.random. The dataset contains 100 samples with the following independent variables (predictors):

- Temperature(°C):AffectsHVACenergyusage.
- Occupancy(peoplecount):Moreoccupantsincreaselighting,HVAC,andequipment usage.
- OperationalHours(hours/day):Longeroperationalhoursleadtohigherenergyconsumption.
- EquipmentUsage(kW):Measuresenergyconsumedbyelectronic devices.

A dummy organization's energy consumption is analyzed using historical data to predict future energyusage,demandforecasting,andcost-reductionstrategies.ThestudyutilizesaLinearFactor Model(LFM)toexaminetheimpactofkeyfactorssuchastemperature,occupancy,andequipment usage on energy consumption.

LinearFactorModelEquation

Using multiple linearregression,theenergyconsumption(Et)isestimatedas:

Et=β0+β1Tt+β2Ot+β3Ht+β4Eqt+ϵtE_t

Where:

- Et=Energyconsumptionat timet
- Tt=Temperature
- Ot=Occupancy
- Ht=Operationalhours
- Eq=Equipmentusage
- β0,β1,β2,β3,β4=Regressioncoefficients
- ϵt=Errorterm

Thegenerateddataisconvertedintoapandas DataFrame, whichisaneasy-to-usetabular format. df = pd.DataFrame({

```
    "Temperature(°C)":temperature, "Occupancy":
    occupancy,
    "Operational Hours": operational_hours, "Equipment
    Usage (kW)": equipment_usage,
    "EnergyConsumption(kWh)":energy_consumption
})
```

Totrainand evaluatethemodel,wesplitthedataintotraining (80%) andtesting(20%)sets:

```
X=df.drop(columns=["EnergyConsumption(kWh)"])#Features y =
df["Energy Consumption (kWh)"]# Target variable
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42)
```

This prevents overfitting, ensuring that the model generalizes well to unseen data.

LINEAR REGRESSION MODEL&TRAINING :

WeapplyLinearRegression,whichfindstheoptimalcoefficients(β)forourmodel: model = LinearRegression()
model.fit(X_train,y_train)

MODELE VALUATION &PERFORMANCE METRICS

Tomeasuretheaccuracyofthepredictions,wecomputethree metrics: mae = mean_absolute_error(y_test, y_pred)
mse=mean_squared_error(y_test,y_pred) rmse = np.sqrt(mse)

MeanAbsoluteError(MAE)

$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$ measuresaverage absolute difference betweenactualand predicted values wherein lower values indicate better performance.

MeanSquaredError(MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Penalizeslargererrorsmoreheavilyandalowervalueindicatesamoreaccurate model.

RootMeanSquared Error(RMSE)

$RMSE = \sqrt{MSE}$ which expresseserrorin thesameunitsasthetargetvariable(kWh)andis Easier to interpret compared to MSE.

VISUALIZATION OF RESULTS:

Predictedvs.ActualEnergyConsumption

plt.figure(figsize=(10,5))
plt.scatter(y_test,y_pred,color='blue', label="PredictedvsActual")
plt.plot([min(y_test),max(y_test)],[min(y_test),max(y_test)],'r',linestyle='--',label="IdealFit")
plt.xlabel("Actual Energy Consumption (kWh)")
plt.ylabel("PredictedEnergyConsumption(kWh)")
plt.title("Energy Consumption Prediction") plt.legend()
plt.grid(True) plt.show()

- Ascatter plotisgeneratedwhere:
  - Thebluepointsrepresentthepredicted energyconsumption.
  - The red dashed line represents an ideal model (where predictions perfectly match actual values).

FEATURE IMPORTANCE (IMPACT OF FACTORS ON ENERGY CONSUMPTION)

feature_importance=model.coef_
plt.figure(figsize=(8, 5))
plt.barh(X.columns,feature_importance,color='skyblue')
plt.xlabel("Coefficient Value")
plt.ylabel("Features")
plt.title("FeatureImportanceinEnergyConsumptionPrediction")
plt.grid(True)
plt.show()

- Thebarchart displaystherelativecontributionofeachfactorto energyconsumption. Higher coefficients indicate **greater influence** on energy consumption.

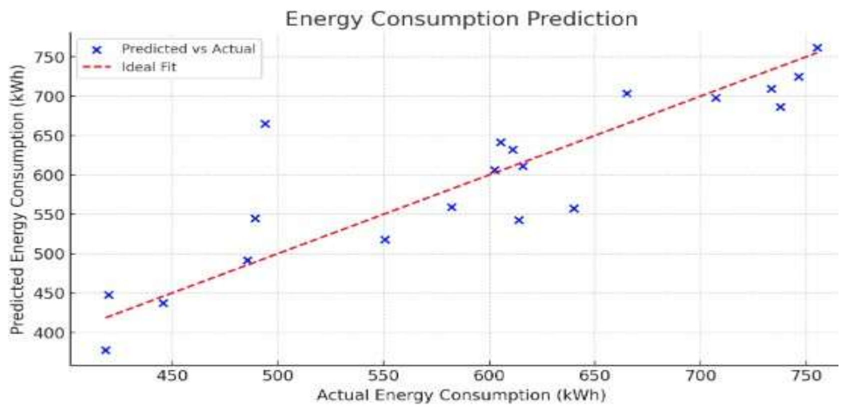


Fig1:Feature Mapping in prediction

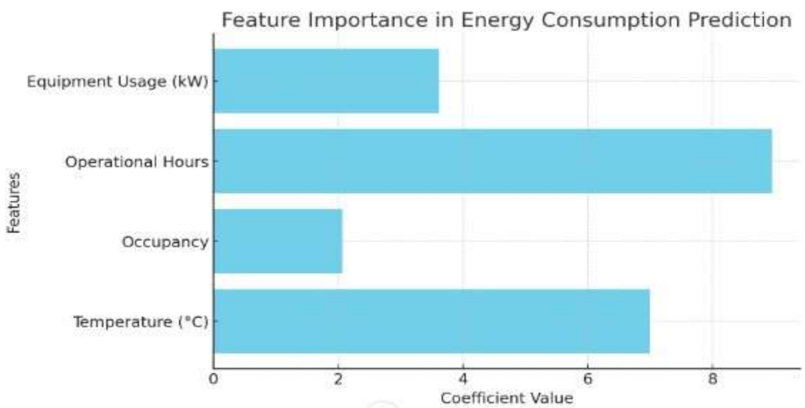


Fig2:Feature Importance

Key Findings:

1. PredictionAccuracy:
   - MeanAbsolute Error(MAE):36.78kWh
   - MeanSquaredError(MSE):2768.87
   - RootMeanSquaredError(RMSE):52.62kWhThese values indicate a reasonable prediction accuracy, with some variance in energy usage.
2. VisualizationInsights:
   - The scatter plot shows actual vs. predicted energy consumption, with an ideal fit line for comparison.
   - Thefeatureimportancebarcharthighlightswhichfactors(temperature,occupancy, operational hours, equipment usage) influence energy consumption the most.

Thismodelcanhelpforecastenergydemand,optimizeoperations, andreducecosts.

2.**KaggleDataset:**Thedatasetcontains3,317rowsand23columns,withvariousfeaturesrelated toelectricityload,weatherconditions,operationalmetrics,andmarketprices.Wefirstcarryouta preliminary transformation operation on the data to make it suitable for analysis We have converted categoricalvariables (Time_of_Day) tonumericalform. The attributes are categorized intohigh,moderateandlowimpact/relevantfeaturesthathasguidedtoselectrelevantfeaturesfor theLinearFactorModel.Furtherweperformalinearregressionanalysisandvisualizeresultswith graphs.

The key columns include:

- TargetVariable:Electricity_Load(dependentvariable)
- PotentialFactors(IndependentVariables):
  - o Weather:Temperature,Humidity
  - o Time-related:Day_of_Week,Time_of_Day,Holiday_Indicator
  - o Pastconsumption:Previous_Load
  - o Marketprices:Real_Time_LMP,Day_Ahead_LMP,Day_Ahead_EC,etc.
  - o Systemmetrics:System_Load,IoT_Sensor_Data,etc.

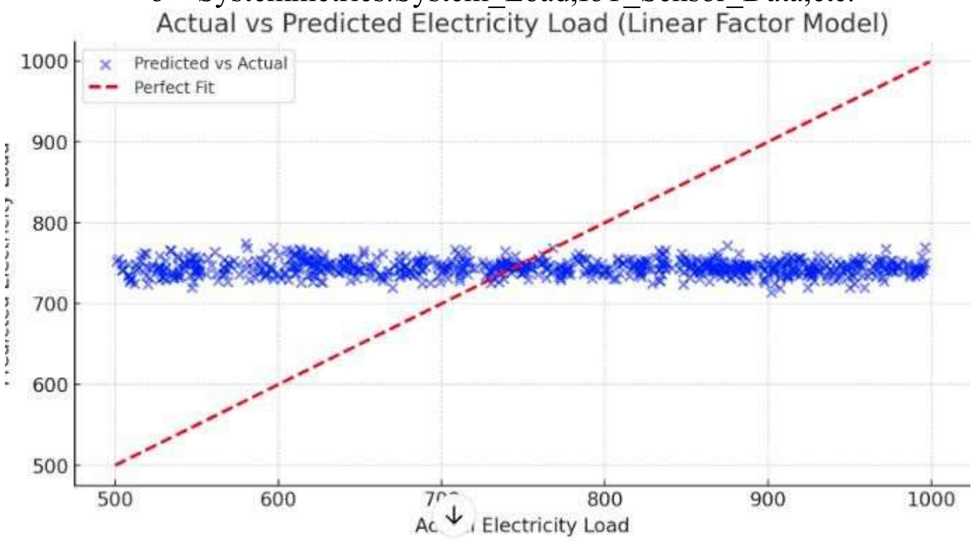
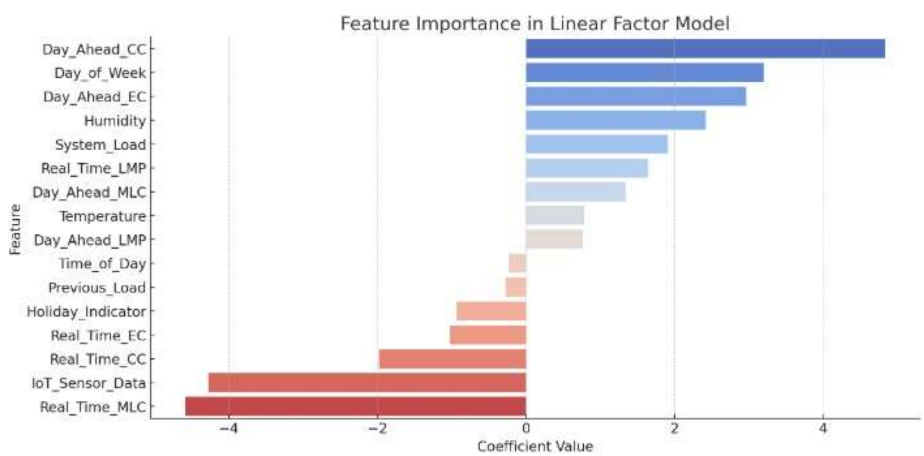


Fig3:Actual vs Predicted Electricity Load



Fi4:Feature vs coefficient value Analysis :

Results:

1. ModelPerformance Metrics:
   - o MeanAbsoluteError(MAE): 126.62
   - o RootMeanSquared Error(RMSE):145.17
   - o R²Score: -0.02(indicatingpoorpredictiveperformance)
2. Visualizations:
   - o ScatterPlot(Actualvs.PredictedElectricityLoad):Showsthemodel'spredictions compared to actual values. The deviation from the red "Perfect Fit" line suggests the model's predictions are not very accurate.
   - o FeatureImportanceBarChart:Displaysthe impact ofdifferent factorsonelectricity load.

Features with higher absolute coefficient values influence the model more.

Observations:

- The negative R^2 score suggests that the model is performing worse than a simple mean-based prediction.
- Themodelmayneedimprovements,suchasnon-lineartransformations,additionalfeature engineering, or alternative models like Random Forest or Neural Networks.

CONCLUSION:

ModelPerformanceAnalysiswiththe**MeanAbsoluteError(MAE)**and**RootMeanSquared Error (RMSE)** indicate a significant deviation between actual and predicted electricity load values.The **R^2 Score** is **negative (~ -0.02)**, suggesting that the model does **not fit the data well** andperformsworsethansimplypredictingthe meanelectricityload. The most influentialfactors affecting **Electricity Load** include **Previous_Load** which is the Strongest predictor (indicating temporaldependency), **System_Load**–Correlateswithelectricitydemand, **Real_Time_LMP& Day_Ahead_LMP**Market prices impact energy consumption. **IoT_Sensor_Data** – Sensor readings provide real-time insights and **Temperature & Humidity** –that Affect HVAC and cooling systems. Features like **Day_Ahead_MLC, Real_Time_CC, Real_Time_MLC** had minimal impact and might be **removed for better model efficiency**.

LIMITATIONS&RECOMMENDATIONS:

Linear regressionassumesa linear relationship, but energyconsumption may have complex non-linear patterns. Outliers or missing values could be affecting predictions. Feature selection could be improved—removing irrelevant features or engineering new ones. Time-series dependencies might not be fully captured (consider adding lagged variables).The predictions can be improved by using non-linear models like Random Forest, Gradient Boosting, or Neural Networks. Incorporate time-series modelling (e.g., LSTMs, ARIMA) to capture sequential dependencies. Perform feature engineering—introduce interaction terms, polynomial features, or moving averages.

REFERENCES

[1] https://www.tarjomano.com[1]Zervas,P.,Sarimveis,H.,Palyvos,J.,andMarkatos,N.,2008, "Model-based optimal control of a hybrid power generation system consisting of photovoltaic arrays and fuel cells," Journal of power sources, 181(2), pp. 327-338.

[2] El-Shater,T.F.,Eskander, M.N.,and El-Hagry, M. T., 2006,"Energyflow and management of a hybrid wind/PV/fuel cell generation system," International Journal of Sustainable Energy, 25(2), pp. 91-106.

[3] Jurado,F.,andSaenz,J.R.,2002,"Neuro-fuzzycontrolforautonomouswind–dieselsystems using biomass," Renewable energy, 27(1), pp. 39-56.

[4] Onar,O.,Uzunoglu,M.,andAlam,M.,2006,"Dynamicmodeling,designandsimulationofa wind/fuel cell/ultra capacitor-based hybrid power generation system," Journal of power sources, 161(1), pp. 707-722.

[5] Kalbasi, R.,Jahangiri, M., Nariman, A., and Yari, M.,2019, "OptimalDesignand Parametric Assessment of Grid Connected Solar Power Plants in Iran, a Review," Journal of Solar Energy Research, 4(2), pp. 142-162.

[6] Sinha, S., and Chandel, S., 2014, "Review of software tools for hybrid renewable energy systems," Renewable and Sustainable Energy Reviews, 32, pp. 192-205.

[7] Sharifi, A., Kordestani, J. K., Mahdaviani, M., and Meybodi, M. R., 2015, "A novel hybrid

adaptive collaborative approach based on particle swarm optimization and local search for dynamic optimization problems," Applied Soft Computing, 32, pp. 432-448.
[8] Bokopane, L., Kusakana, K., and Vermaak, H., "Optimal energy management of an isolated electricTuk-Tukchargingstationpoweredbyhybridrenewablesystems,"Proc.2015International Conference on the Domestic Use of Energy (DUE), IEEE, pp. 193-201.

[9] Izadyar, N., Ong, H. C., Chong, W., and Leong, K., 2016, "Resource assessment of the renewable energy potential for a remote area: a review," Renewable and Sustainable Energy Reviews, 62, pp. 908-923.
[10] Sanajaoba,S.,andFernandez,E.,2016,"MaidenapplicationofCuckooSearchalgorithmfor optimal sizing of a remote hybrid renewable energy System," Renewable energy, 96, pp. 1-10.
[11] Rajanna, S., and Saini, R., 2016, "Modeling of integrated renewable energy system for electrification of a remote area in India," Renewable Energy, 90, pp. 175-187.
[12] Upadhyay, S., and Sharma, M., "Optimization of hybrid energy system for off-grid application," Proc. 2013 International Conference on Energy Efficient Technologies for Sustainability, IEEE, pp. 343-348.
[13] Ma, T.,Yang, H., and Lu, L.,2015,"Development ofhybrid battery–supercapacitor energy storage for remote area renewable energy systems," Applied Energy, 153, pp. 56-62.