Regression Analysis on Dataset 'Renewable Energy and Weather Conditions'

Hasin Md Daiyan
dept. of CSE
AUST
Dhaka, Bd.
hasin.cse.20200204055@aust.edu

Ahmed Al Nahian
dept. of CSE
AUST
Dhaka, Bd.
ahmed.cse.20200204063@aust.edu

Amit Karmakar
dept. of CSE
AUST
Dhaka, Bd
amit.cse.20200204056@aust.edu

The relationship between renewable energy generation and weather conditions is critical for optimizing energy production and management. This study explores the use of regression analysis techniques, including Long Short-Term Memory (LSTM) networks and Hidden Markov Models (HMM), to investigate the impact of weather variables on renewable energy output over time..

***Index Terms*—HMM, LSTM**

I. Introduction

Regression analysis is a statistical method used to examine the relationship between one dependent variable and one or more independent variables. Conventional linear regression models have limitations in capturing complex relationships between weather variables due to their linearity assumption[1]. LSTM and HMM are employed in regression analysis of renewable energy and weather conditions to leverage their abilities to handle sequential and probabilistic data respectively, thereby enhancing the accuracy and predictive power of models.

II. Motivation

Advancements in renewable energy technologies increasingly rely on accurate predictive models that integrate intricate weather dynamics to optimize energy generation and management. Recent studies, such as those employing LSTM for capturing long-term dependencies in time-series data and utilizing HMM to model probabilistic weather regimes impacting renewable energy output underscore the critical role of these methodologies. LSTM's ability to handle sequential data and HMM's capacity to identify hidden states provide robust frameworks for understanding and predicting the complex relationships between weather variables and energy production. Integrating these advanced techniques in regression analysis promises to enhance the precision of energy forecasts, inform strategic decision-making, and drive innovations towards sustainable energy solutions.

III. Literature Review

Various regression algorithms analysed are evaluated by the evaluation criteria such as root mean square error, mean absolute error, relative absolute error and relative root square error[2]. In this paper MAE of Linear Reg on Temperature was 0.4113 [2]. In yet another study , it was revealed that among Temperature and humidity has relation as parameters[3].

IV. Methodology

A. Data Collection and Preprocessing

For this study, a dataset named Renewable Energy and Weather Conditions collected . The text data underwent several preprocessing steps to clean and prepare it for analysis.The dataset was loaded using pandas and its structure was examined. Missing values were identified and visualized. Categorical variables, specifically weather-type, were encoded using OneHotEncoder to prepare them for modeling. Numerical features were normalized using StandardScaler to ensure uniform scale across variables.

B. Model Development

Two models were developed to analyze the relationship between weather conditions and renewable energy generation. An LSTM model was constructed to capture temporal dependencies in the data, reshaping it into a 3D format and trained using Keras/TensorFlow. Simultaneously, a Gaussian Hidden Markov Model (HMM) was implemented to predict latent weather states affecting energy output.

- LSTM (Long Short-Term Memory) networks are a type of recurrent neural network (RNN) designed to handle sequence data by effectively capturing long-term dependencies. They are particularly suited for time-series analysis in scenarios where past information significantly influences future outcomes, such as predicting renewable energy generation based on historical weather data
- HMM (Hidden Markov Models) are probabilistic models that assume the system being modeled is a Markov process with hidden states influencing observed outcomes. In

the context of renewable energy and weather conditions, HMMs help uncover latent weather patterns or regimes (e.g., clear skies, cloudy days) that impact energy generation

C. Model Training and Evaluation

The LSTM model was trained on the prepared dataset, with performance metrics such as Mean Absolute Error (MAE) and R² Score used for evaluation on a test set. Concurrently, a Linear Regression model was trained using HMM-predicted states as features to assess their impact on predicting energy deltas, with Mean Squared Error (MSE) and R² Score used to gauge predictive accuracy and model performance.

V. DATASET

The 'Renewable Energy and Weather Conditions' dataset offers detailed hourly insights into energy consumption and essential meteorological variables critical for renewable energy generation analysis. It includes measurements such as Global Horizontal Irradiance (GHI), which quantifies solar radiation received on a horizontal surface, along with temperature, pressure, humidity, wind speed, and precipitation data. The "Energy delta[Wh]" column tracks changes in energy consumption over specified time periods, providing key indicators of energy demand fluctuations. Additional features like isSun denote sunlight presence, while dayLength and sunlightTime detail daylight duration and available sunlight hours. The weather-type column categorizes conditions into clear, cloudy, or rainy, offering crucial context for examining how weather impacts renewable energy production trends. Organized by hour and month, this dataset facilitates comprehensive exploration of the dynamic relationship between weather dynamics and renewable energy efficiency.

VI. RESULT AND ANALYSIS

The results of the LSTM and HMM models for predicting renewable energy generation based on weather conditions are summarized as follows:

A. LSTM Model Performance:

The LSTM model achieved a Test Loss of 139166.0 and a Mean Absolute Error (MAE) of 170.196. The R² Score, indicating the model's goodness of fit, was 0.87.

B. HMM Model Performance:

The HMM-based approach resulted in a Mean Squared Error (MSE) of 138000.34 and an R-squared (R²) value of 0.0.87.
These findings suggest that while both models effectively predict energy generation based on weather variables, both models demonstrates superior performance in capturing the underlying probabilistic structure of weather conditions affecting renewable energy output compared to linear regression. Integrating these models provides comprehensive insights into optimizing energy management strategies under varying weather conditions.

TABLE I
RESULT

Linear Regression	**LSTM**	**HMM**
0.85	**(MAE)0.87**	**(MSE)0.87**

CONCLUSION

The exploration of renewable energy generation in relation to weather conditions underscores the critical role of meteorological factors in shaping energy production dynamics. Through comprehensive analysis using advanced techniques like LSTM and HMM models, this study has illuminated significant insights into how weather variables such as solar radiation, temperature, and wind speed influence energy output over time. The LSTM model demonstrated robust performance in capturing temporal dependencies, while the HMM model effectively identified hidden weather patterns impacting energy generation patterns. These findings highlight the importance of leveraging data-driven approaches to optimize renewable energy systems amidst varying weather conditions, paving the way for more sustainable and resilient energy solutions.

REFERENCES

[1] Gurwinder Singh, Anil Sharma, Pramod Vishwakarma, Manoj Gaur-Revealing non-linear trends and patterns in weather data through regression analysis approach

[2] Yeturu Jahnavi-Analysis of weather data using various regression algorithms

[3] L. Sunitha, M. Balraju,J. Sasikiran,B. Anil Kumar-FINDING RELATION BETWEEN PARAMETERS OF WEATHER DATA USING LINEAR REGRESSION METHOD