# CS3103 – Machine Learning
# Mid-Term Project Report

Group ID: 25GK26

## 1 Title and Group Details

**Project Title:** Renewable Energy Production Prediction using Hybrid Machine Learning Models

**Group ID:** 25GK26

| Name | Roll Number | Role and Responsibilities |
|---|---|---|
| Juhi Sahni | 2301CS88 | Define overall roadmap and ensure integration. Implement XGBoost and oversee version control. |
| Saniya Prakash | 2301CS49 | Data preprocessing: handle missing values, outlier detection, normalization, and feature engineering. |
| Mihika | 2301CS31 | Build predictive models: implement Linear Regression, Random Forest, and hyperparameter tuning. |
| Shefali Bishnoi | 2301CS87 | Implement SVR with multiple kernels. Prepare research documentation. |

Table 1: Group Members and Their Responsibilities

## 2 Introduction

Due to the rapid depletion and very slow production of fossil fuels, switching to renewable energy sources like wind and solar is imperative. However, because of shifting environmental conditions, their production is unpredictable. For energy management, accurate forecasting is essential. By identifying both linear and nonlinear patterns in environmental data, this project creates a hybrid machine learning model to forecast the production of renewable energy.

## 3 Related Work

A thorough analysis of LSTM models is given in [2], which emphasizes how well they manage exploding/vanishing gradient issues, which makes them perfect for time-series forecasting applications such as renewable energy prediction.

[3] evaluated XGBoost and Linear Regression for solar energy forecasting, showing that XGBoost was better with an MAE of 38.08 compared to 80.23 for Linear Regression. In a similar vein, [4] used Random Forest to predict solar energy and achieved 99.82% accuracy in classification tasks.

Although these studies offer insightful information, their primary focus is on predicting a single energy source. By creating a thorough framework for multi-output regression of simultaneous PV and wind production using hybrid CNN-LSTM architectures, our research fills this gap.

# 4 Methodology

## 4.1 Dataset Description

The dataset [1] contains 38,880 samples collected at 5-minute intervals, featuring meteorological parameters (solar irradiance DHI/DNI/GHI, wind speed, humidity, temperature) and energy production data (PV and wind generation). The dataset includes temporal features (season, day of week) and covers a complete annual cycle, enabling comprehensive modeling of renewable energy patterns.

## 4.2 Block Diagram / Workflow

The comprehensive workflow for the renewable energy prediction system is shown in Figure 1 (see Appendix).

## 4.3 Proposed CNN-LSTM Hybrid Architecture

We suggest a model that uses Convolutional Neural Networks (CNN) to find spatial features and Long Short-Term Memory (LSTM) networks to learn temporal sequences. We chose this hybrid method because CNNs are great at finding spatial patterns and correlations between weather features, and LSTMs are made to deal with temporal dependencies and long-term patterns in time-series data. This combination is especially good for predicting renewable energy because weather features have spatial relationships and energy production follows temporal patterns.

After extracting significant spatial features from meteorological data using convolutional layers, the hybrid architecture models the time-dependent nature of energy generation using LSTM layers. This allows the model to learn intricate relationships between energy output and environmental conditions over time, producing precise simultaneous predictions for wind and PV energy production from multi-dimensional weather data.

## 4.4 Mathematical Formulation

The CNN-LSTM model can be represented as a composite function:

$$\hat{\mathbf{Y}} = f_{LSTM}(f_{CNN}(\mathbf{X})) \tag{1}$$

where $\mathbf{X}$ represents the input feature matrix and $\hat{\mathbf{Y}}$ contains the predicted PV and wind production values.

The convolutional operation extracts spatial features:

$$\mathbf{H}^{conv} = \sigma(\mathbf{W}_{conv} * \mathbf{X} + \mathbf{b}_{conv}) \tag{2}$$

The LSTM component processes temporal sequences through its gating mechanism:

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \tag{3}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{C}_t) \tag{4}$$

The model is trained to minimize the prediction error through backpropagation.

**Algorithm 1** CNN-LSTM Hybrid Training for Energy Prediction

**Require:** Feature matrix $\mathbf{X} \in \mathbb{R}^{N \times T \times F}$, target matrix $\mathbf{Y} \in \mathbb{R}^{N \times 2}$
**Ensure:** Optimized CNN-LSTM model with parameters $\theta^*$

```
1:  Normalize features using StandardScaler and handle missing values
2:  Reshape X to spatiotemporal format (samples, timesteps, features)
3:  Split data into training, validation, and test sets
4:  Initialize CNN layers for spatial feature extraction
5:  Initialize LSTM layers for temporal sequence modeling
6:  Add Dense output layer for multi-output regression (PV and Wind)
7:  for each epoch do
8:      Forward pass: H^conv = CNN(X)
9:      h_t = LSTM(H^conv), Ŷ = Dense(h_t)
10:     Compute loss: L = (1/N) Σ_{i=1}^{N} ||Y_i − Ŷ_i||^2
11:     Update parameters via backpropagation: θ ← θ − α∇_θ L
12:     Evaluate validation loss and apply early stopping if needed
13: end for
14: Evaluate model on test set and compute performance metrics
15: return Trained model with optimized parameters θ*
```

## 5 References


[1] S. Rojas Ortega, P. Castro-Correa, S. Sepúlveda-Mora, and J. Castro-Correa, "Renewable energy and electricity demand time series dataset with exogenous variables at 5-minute interval," Mendeley Data, vol. 1, 2023.

[2] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 1-44, 2020.

[3] W. Hastomo, A. Digdoyo, A. Satyo, and D. Arif, "Enhancing solar energy efficiency: Predictive modeling with XGBoost and linear regression," *JIKA (Jurnal Informatika)*, vol. 9, no. 1, pp. 66-74, 2025.

[4] N. P, M. Sari, and K. Kusnanto, "Predicting the potential of renewable solar energy based on weather data in Indonesia using the random forest method," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 4, pp. 1771-1781, 2025.

# Appendix: Workflow Diagram

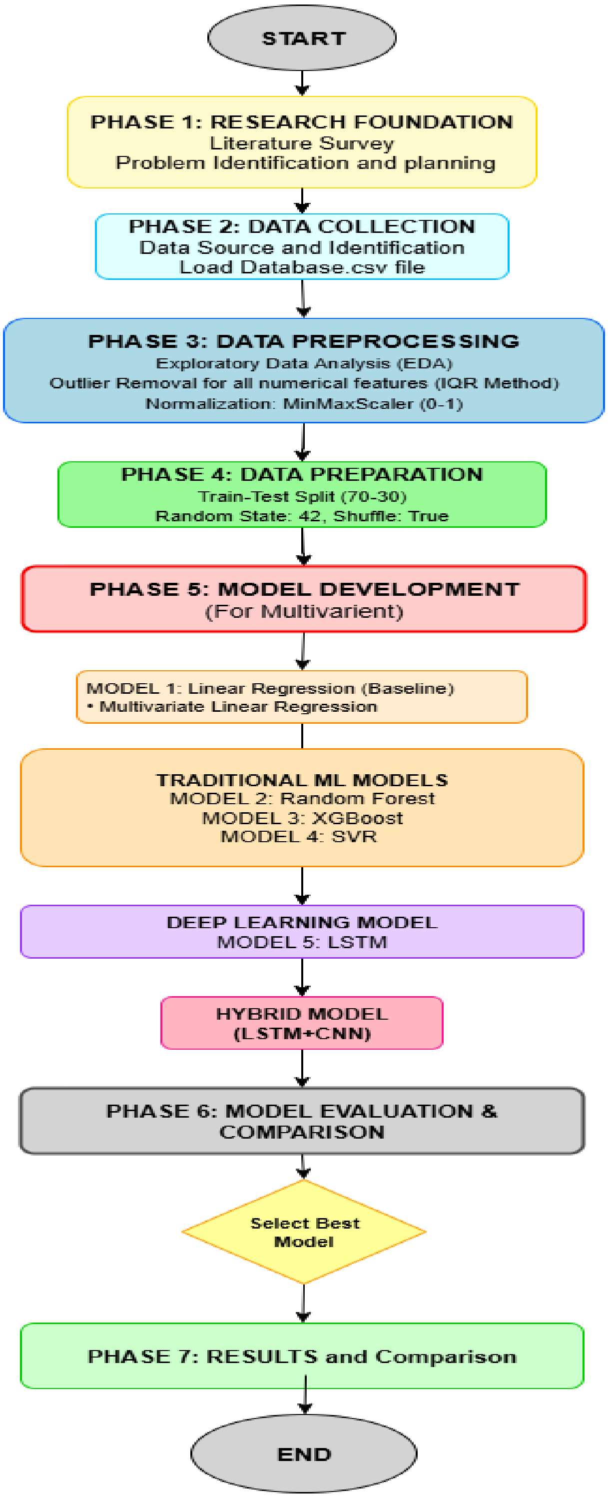


Figure 1: Comprehensive workflow for renewable energy prediction system