Department of Mechanical Engineering
Indian Institute of Technology (BHU) Varanasi

Prediction of wind speed in coastal areas using Machine learning models

Manish Kumar

Supervisor : Arnab Sarkar

Mentor : Mohan Gupta

PROJECT OVERVIEW AND MOTIVATION

- The project aimed to develop an accurate, interpretable, and automated machine learning model for wind speed prediction in the coastal region of Jamnagar, Gujarat . This area is characterized by dynamic coastal winds influenced by monsoons, sea breezes, and complex interactions between land and sea temperatures.

- Disaster Management: Early warnings for storms, cyclones, and coastal flooding to protect lives and property.

- Renewable Energy: Reliable forecasts enable efficient planning and operation of coastal and offshore wind farms—a key to India's green energy transition.

PROBLEM STATEMENT AND OBJECTIVES

Problem Statement :

- Current methods for predicting wind speed don't work well for coastal areas complicated weather. We need better, easy-to-understand machine learning models and automated data tools to make accurate wind forecasts.

Objectives:

- Automate ERA5 data acquisition for Jamnagar.
- Preparing data and feature engineering capturing seasonal/diurnal effects.
- Develop and compare RF, XGBoost and ANN ,
- Apply SHAP for model explainability.

PROJECTS UNDERTAKEN – MODEL DEVELOPMENT

- Developed, tuned, and compared Random Forest, XGBoost, and Artificial Neural Network models for wind speed prediction in Jamnagar.
- Selected and retained the most important meteorological features using model-based importance scores and SHAP explainability methods.
- Applied SHAP for model interpretability, making predictions transparent and explaining key feature contributions.
- Systematically analyzed model results using metrics like R2 score , MSE, and MAE.

Data Acquisition and Automation

- Used ERA5 data from Climate Data Store for Jamnagar (22.5°N–22.3°N, 69.9°E–70.2°E) for year 2021, on hourly basis at 950 hPa.
- Manual downloads were time-consuming and error-prone due to large data volume.
- Automated downloads using Python CDS API with batching by months and day ranges.
- Processed complex NetCDF climate datasets and converted them into structured CSV format.

Data prepration and Feature engineering

- Cleaning: Imputed missing values with the mean value of corresponding column.
- Features: Constructed from raw meteorological fields—wind components, temperature, humidity, etc.
  - We discarded some meteorogogical variables like crwc which had not significant contribution to the model.
- Split data: 80% train, 20% test for robust evaluation.

MACHINE LEARNING MODELS

1.Random Forest Regressor summary

- It is an ensemble ML algorithm that predicts continuous values by averaging the results of multiple decision trees, providing high accuracy and robustness to overfitting.
- n_estimators=100 and feature importance of >= 0.02 are only used.

- Time Taken : 29.35 seconds
- R^2 Score : 0.7565699
- MSE : 2.387225

Random Forest

where
z - Geopotential Height
t- Temperature
w- Vertical velocity
q- specific humidity
r - relative humidity
vo - vorticity
d - divergence
crwc - cloud liquid water content

XGBOOST REGRESSOR SUMMARY

- XGBoost Regressor is a gradient boosting machine learning algorithm that builds an ensemble of weak prediction trees sequentially, optimizing model accuracy by minimizing a differentiable loss function with regularization to prevent overfitting.

- Time Taken : 0.62
- R^2 Score : 0.64754194
- MSE : 3.4564

XGBOOST REGRESSOR SUMMARY

where

z - Geopotential Height

t- Temperature

w- Vertical velocity

q- specific humidity

r - relative humidity

vo - vorticity

d - divergence

crwc - cloud liquid water content

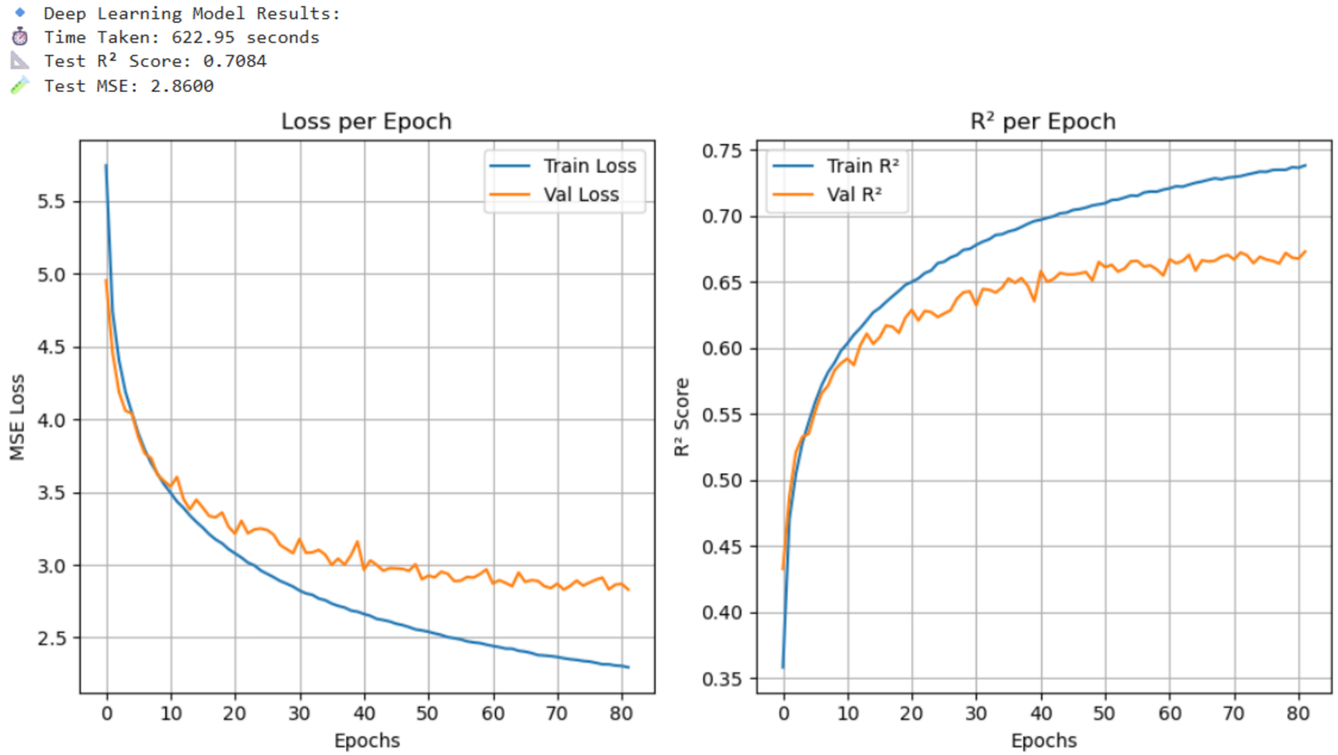
ARTIFICIAL NEURAL NETWORK (ANN) SUMMARY

- An Artificial Neural Network (ANN) Regressor is a deep learning model that consists of multiple interconnected layers of neurons which learn complex nonlinear relationships in data through iterative weight optimization, making it suitable for modeling intricate patterns such as those in meteorological wind forecasting.

- In this model we have used different number of layers having different number of neuron with different activation functions such as ReLu, RMS Prop , Adam and SGD using GridSearch and RandomSearch CV , we found best model with different combination .

ARTIFICIAL NEURAL NETWORK (ANN)

- Model consists of two dense layer each having 256 neurons.
- Model used ReLU as the activation function and was trained for 100 epochs.
- R2_Score , MSE and MAE used as accuracy matrix.



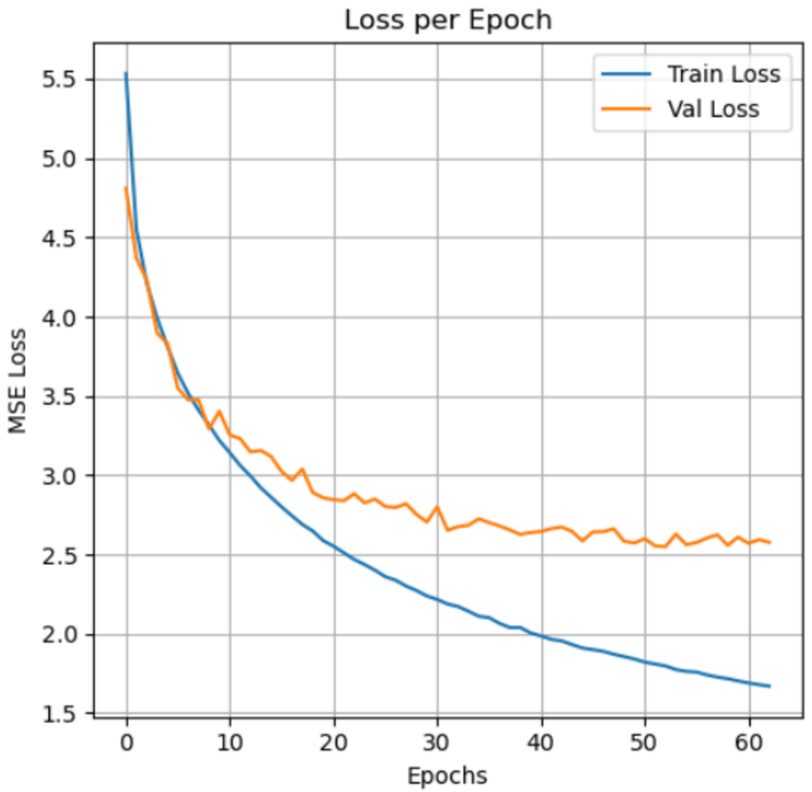
Artificial neural network (ANN)

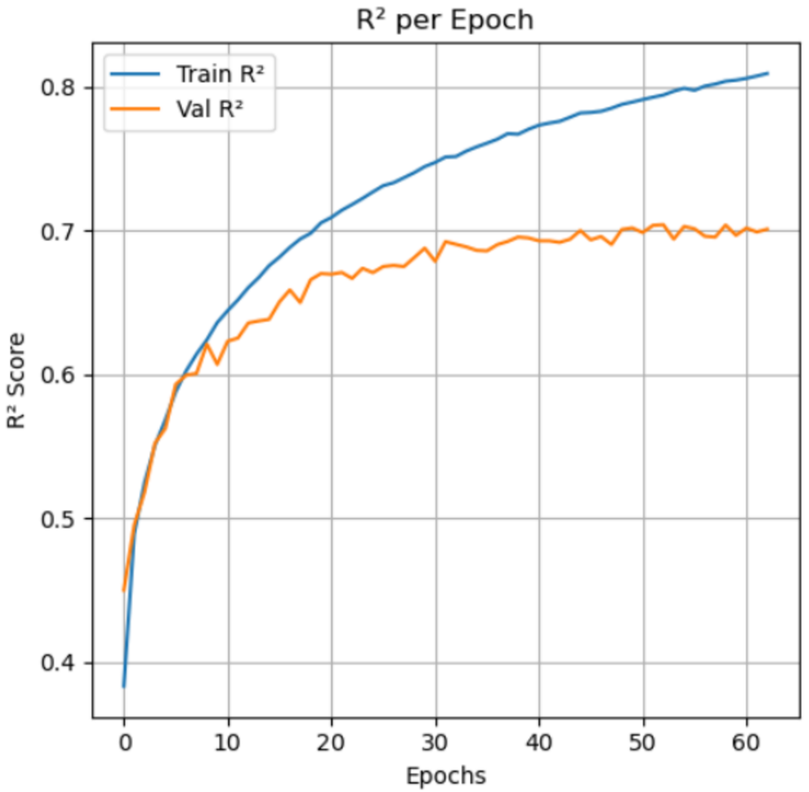
Model consists of 4 dense layer having 256, 256, 128 and 64 neurons.

Each layer has ReLu as activation function ans was trained for 100 epchos.

R2_Score and MSE used as accuracy matrix.

🔹 Deep Learning Model Results:
⏱️ Time Taken: 573.62 seconds
📐 Test R^2 Score: 0.7414
📏 Test MSE: 2.5360





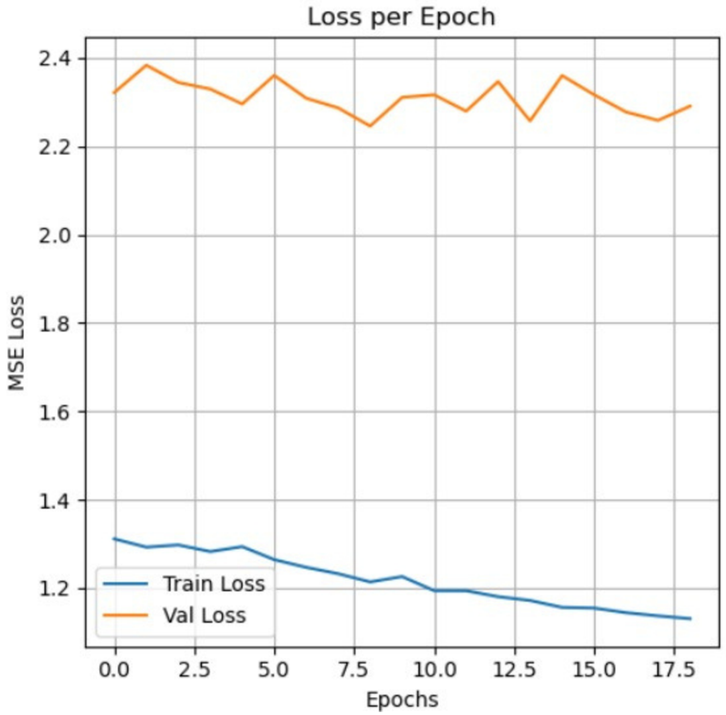
Artifical Neural Network(ANN)

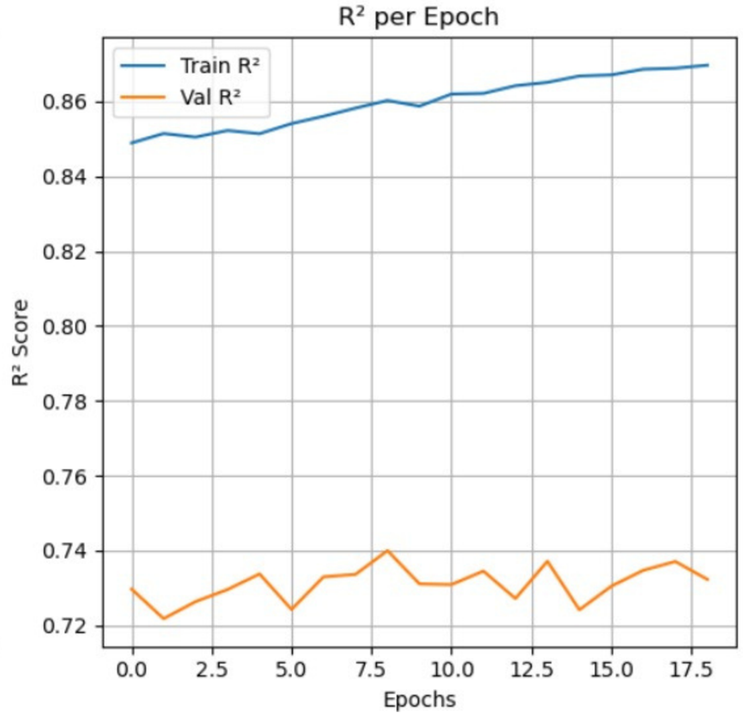
Model consists of dense layer ranging from 1 to 4 having different no. of neurons in each layer.

Different Activation function such as Adam, RMS Prop, SGD are used in each layer and trained for 100 epoches.

R2_score and MSE were used as accuracy matrix.

🔹 Tuned Deep Learning Model Results:
⏱️ Time Taken: 5704.76 seconds
📐 Test R^2 Score: 0.7710
📉 Test MSE: 2.2456





Thankyou