Activity based

Project Report on

Artificial Intelligence Machine Learning

Phase - II

Submitted to Vishwakarma University, Pune

Under the Initiative of

Contemporary Curriculum, Pedagogy, and Practice (C2P2)

By

Shravan Sudhir Meshram

SRN No : 202101425

Roll No : 31

Div : E

Third Year Engineering

Department of Computer Engineering

Faculty of Science and Technology

Academic Year

2023-2024

Artificial Intelligence Machine Learning: Project Phase II
Project Name: Classify weather conditions based on features like temperature and humidity using decision trees, random forest, and ensemble models.

- ## Introduction

Classifying weather conditions using predictive modeling is a fascinating and crucial task that involves analyzing patterns in meteorological data to predict future weather events. This process not only aids in understanding the environment better but also plays a vital role in planning and decision-making across various sectors such as agriculture, transportation, and emergency management. Among the plethora of machine learning techniques available, decision trees, random forests, and ensemble models stand out for their ability to handle the complexity and variability of weather data. This introduction will briefly explore these models, setting the stage for a deeper dive into their application in classifying weather conditions based on features like temperature and humidity.

Decision Trees

Decision trees are a type of supervised learning algorithm that models decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It's a straightforward and intuitive method that splits the data into branches to make predictions. In the context of weather classification, a decision tree would analyze the dataset, breaking it down into smaller subsets based on the features (e.g., temperature, humidity) and making decisions at each node based on these features. The end result is a tree-like model of decisions and their possible consequences, which can be used to classify weather conditions.

Random Forests

Random Forest is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forests correct for decision trees' habit of overfitting to their training set, providing a more robust and accurate model. For weather classification, a random forest would use multiple decision trees to make its predictions, taking into account the variability and interdependence of weather features to improve accuracy.

Ensemble Models

Ensemble models combine multiple machine learning techniques to improve predictive performance compared to models that make predictions individually. This approach can blend different types of models or multiple instances of the same model to harness their collective intelligence. In weather classification, ensemble models might combine decision trees, random forests, and other algorithms, leveraging their individual strengths and compensating for their weaknesses. This method is particularly effective in handling the unpredictable and multifaceted nature of weather data.

- **Glimpse of Dataset**

Overview

Weather datasets compile extensive records of atmospheric conditions over time, gathered from a wide array of sources including ground-based weather stations, satellites, radar systems, and weather balloons. These datasets can range from historical records spanning several decades to near-real-time data feeds. The granularity of the data can also vary, from minute-by-minute observations to daily, monthly, or annual summaries.

Time	Energy de	GHI	temperat	pressure	humidity	wind_spe	rain_1h	snow_1h	clouds_al	isSun	sunlightT	dayLengtl	SunlightT	weather_	hour	month
01-01-2017 00:00	0	0	1.6	1021	100	4.9	0	0	100	0	0	450	0	4	0	1
01-01-2017 00:15	0	0	1.6	1021	100	4.9	0	0	100	0	0	450	0	4	0	1
01-01-2017 00:30	0	0	1.6	1021	100	4.9	0	0	100	0	0	450	0	4	0	1
01-01-2017 00:45	0	0	1.6	1021	100	4.9	0	0	100	0	0	450	0	4	0	1
01-01-2017 01:00	0	0	1.7	1020	100	5.2	0	0	100	0	0	450	0	4	1	1
01-01-2017 01:15	0	0	1.7	1020	100	5.2	0	0	100	0	0	450	0	4	1	1
01-01-2017 01:30	0	0	1.7	1020	100	5.2	0	0	100	0	0	450	0	4	1	1
01-01-2017 01:45	0	0	1.7	1020	100	5.2	0	0	100	0	0	450	0	4	1	1
01-01-2017 02:00	0	0	1.9	1020	100	5.5	0	0	100	0	0	450	0	4	2	1
01-01-2017 02:15	0	0	1.9	1020	100	5.5	0	0	100	0	0	450	0	4	2	1
01-01-2017 02:30	0	0	1.9	1020	100	5.5	0	0	100	0	0	450	0	4	2	1
01-01-2017 02:45	0	0	1.9	1020	100	5.5	0	0	100	0	0	450	0	4	2	1
01-01-2017 03:00	0	0	2	1019	100	5.7	0	0	100	0	0	450	0	4	3	1
01-01-2017 03:15	0	0	2	1019	100	5.7	0	0	100	0	0	450	0	4	3	1
01-01-2017 03:30	0	0	2	1019	100	5.7	0	0	100	0	0	450	0	4	3	1
01-01-2017 03:45	0	0	2	1019	100	5.7	0	0	100	0	0	450	0	4	3	1
01-01-2017 04:00	0	0	2.5	1018	100	5.6	0	0	100	0	0	450	0	4	4	1
01-01-2017 04:15	0	0	2.5	1018	100	5.6	0	0	100	0	0	450	0	4	4	1
01-01-2017 04:30	0	0	2.5	1018	100	5.6	0	0	100	0	0	450	0	4	4	1
01-01-2017 04:45	0	0	2.5	1018	100	5.6	0	0	100	0	0	450	0	4	4	1
01-01-2017 05:00	0	0	2.6	1017	100	5.7	0	0	100	0	0	450	0	4	5	1
01-01-2017 05:15	0	0	2.6	1017	100	5.7	0	0	100	0	0	450	0	4	5	1
01-01-2017 05:30	0	0	2.6	1017	100	5.7	0	0	100	0	0	450	0	4	5	1
01-01-2017 05:45	0	0	2.6	1017	100	5.7	0	0	100	0	0	450	0	4	5	1
01-01-2017 06:00	0	0	2.8	1017	100	6	0	0	100	0	0	450	0	4	6	1
01-01-2017 06:15	0	0	2.8	1017	100	6	0	0	100	0	0	450	0	4	6	1
01-01-2017 06:30	0	0	2.8	1017	100	6	0	0	100	0	0	450	0	4	6	1
01-01-2017 06:45	0	0	2.8	1017	100	6	0	0	100	0	0	450	0	4	6	1
01-01-2017 07:00	0	0	2.9	1016	100	6.1	0	0	98	0	0	450	0	4	7	1
01-01-2017 07:15	0	0.2	2.9	1016	100	6.1	0	0	98	1	15	450	0.03	4	7	1
01-01-2017 07:30	0	2.7	2.9	1016	100	6.1	0	0	98	1	30	450	0.07	4	7	1
01-01-2017 07:45	0	6.4	2.9	1016	100	6.1	0	0	98	1	45	450	0.1	4	7	1
01-01-2017 08:00	5	10.6	3.5	1016	99	6	0	0	98	1	60	450	0.13	4	8	1
01-01-2017 08:15	33	6	3.5	1016	99	6	0	0	98	1	75	450	0.17	4	8	1
01-01-2017 08:30	44	2.8	3.5	1016	99	6	0	0	98	1	90	450	0.2	4	8	1
01-01-2017 08:45	61	3.1	3.5	1016	99	6	0	0	98	1	105	450	0.23	4	8	1
01-01-2017 09:00	65	3.5	3.6	1016	97	6.2	0	0	100	1	120	450	0.27	4	9	1
01-01-2017 09:15	83	3.8	3.6	1016	97	6.2	0	0	100	1	135	450	0.3	4	9	1
01-01-2017 09:30	69	4.1	3.6	1016	97	6.2	0	0	100	1	150	450	0.33	4	9	1
01-01-2017 09:45	98	4.3	3.6	1016	97	6.2	0	0	100	1	165	450	0.37	4	9	1
01-01-2017 10:00	138	4.5	3.8	1015	93	5.6	0	0	94	1	180	450	0.4	4	10	1
01-01-2017 10:15	161	4.7	3.8	1015	93	5.6	0	0	94	1	195	450	0.43	4	10	1
01-01-2017 10:30	119	6.4	3.8	1015	93	5.6	0	0	94	1	210	450	0.47	4	10	1
01-01-2017 10:45	95	11.9	3.8	1015	93	5.6	0	0	94	1	225	450	0.5	4	10	1

- **CODE:**

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report
from IPython.display import display


data = pd.read_csv('weather_data.csv')

# Split the data into training and testing sets
```

```
X = data[['temperature', 'humidity']]  # Features
y = data['weather_condition']  # Target variable

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Scale the data
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Train decision tree
dt_classifier = DecisionTreeClassifier(random_state=42)
dt_classifier.fit(X_train, y_train)

# Evaluate decision tree
y_pred_dt = dt_classifier.predict(X_test)
accuracy_dt = accuracy_score(y_test, y_pred_dt)
report_dt = classification_report(y_test, y_pred_dt)

print("Decision Tree")
print(f"Accuracy: {accuracy_dt:.2f}")
print(f"Classification Report:\n{report_dt}\n")
display(pd.DataFrame({"Accuracy": [accuracy_dt], "Classification Report":
[report_dt]}))

# Train random forest
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42)
rf_classifier.fit(X_train, y_train)

# Evaluate random forest
y_pred_rf = rf_classifier.predict(X_test)
accuracy_rf = accuracy_score(y_test, y_pred_rf)
report_rf = classification_report(y_test, y_pred_rf)

print("Random Forest")
print(f"Accuracy: {accuracy_rf:.2f}")
print(f"Classification Report:\n{report_rf}\n")
display(pd.DataFrame({"Accuracy": [accuracy_rf], "Classification Report":
[report_rf]}))

# Train ensemble model
ensemble_clf = VotingClassifier(estimators=[
    ('dt', dt_classifier),
    ('rf', rf_classifier)], voting='soft')

ensemble_clf.fit(X_train_scaled, y_train)  # Use scaled data

y_pred_ensemble = ensemble_clf.predict(X_test_scaled)
```

```
accuracy_ensemble = accuracy_score(y_test, y_pred_ensemble)
report_ensemble = classification_report(y_test, y_pred_ensemble)

print("Ensemble Model")
print(f"Accuracy: {accuracy_ensemble:.2f}")
print(f"Classification Report:\n{report_ensemble}\n")
display(pd.DataFrame({"Accuracy": [accuracy_ensemble], "Classification Report":
[report_ensemble]}))
```

- **OUTPUT**

Decision Tree

```
Decision Tree
Accuracy: 0.89
Classification Report:
              precision    recall  f1-score   support

           1       0.87      0.96      0.92        28
           2       0.97      0.85      0.91        46
           3       0.81      0.86      0.83        63
           4       0.94      0.91      0.92       214
           5       0.76      0.86      0.81        49

    accuracy                           0.89       400
   macro avg       0.87      0.89      0.88       400
weighted avg       0.89      0.89      0.89       400


   Accuracy                             Classification Report
0      0.89                precision    recall  f1-score    ...
```

Random Forest

```
Random Forest
Accuracy: 0.90
Classification Report:
              precision    recall  f1-score   support

           1       0.81      0.89      0.85        28
           2       1.00      0.83      0.90        46
           3       0.87      0.84      0.85        63
           4       0.93      0.93      0.93       214
           5       0.79      0.92      0.85        49

    accuracy                           0.90       400
   macro avg       0.88      0.88      0.88       400
weighted avg       0.90      0.90      0.90       400


   Accuracy                              Classification Report
0       0.9                precision    recall  f1-score   ...
```

Ensemble Model

```
Ensemble Model
Accuracy: 0.90
Classification Report:
              precision    recall  f1-score   support

           1       0.81      0.89      0.85        28
           2       1.00      0.83      0.90        46
           3       0.87      0.84      0.85        63
           4       0.93      0.93      0.93       214
           5       0.79      0.92      0.85        49

    accuracy                           0.90       400
   macro avg       0.88      0.88      0.88       400
weighted avg       0.90      0.90      0.90       400


   Accuracy                              Classification Report
0       0.9                precision    recall  f1-score   ...
```

- **Conclusion**

The endeavour to classify weather conditions through the use of advanced machine learning techniques such as decision trees, random forests, and ensemble models demonstrates a powerful approach to interpreting complex environmental data. These methods offer distinct advantages in handling the intricate and often nonlinear relationships inherent in weather data, which includes variables like temperature and humidity. The application of decision trees provides an intuitive and straightforward mechanism for breaking down data into manageable parts, making it easier to understand the decision-making process. Meanwhile, random forests enhance this approach by integrating multiple decision trees to reduce overfitting and improve predictive accuracy, thus offering a more robust model for weather classification. Ensemble models take this a step further by combining the strengths of various predictive models to achieve superior performance, effectively leveraging their collective power to tackle the unpredictability and multifaceted nature of weather phenomena.

The exploration of weather datasets, with their rich compilation of atmospheric measurements, lays the groundwork for these advanced analytical techniques. By tapping into detailed observations from ground-based stations, satellites, radar systems, and weather balloons, we can feed our models with high-quality, diverse data sources, thereby enriching the predictive modeling process. The significance of this work extends beyond the academic and scientific realms, touching upon practical applications in agriculture, disaster preparedness, urban planning, and climate research, among others. Accurate weather classification and forecasting can lead to better decision-making and planning, ultimately contributing to the safety, efficiency, and resilience of societies worldwide.