

DESCRIPTION

Dataset: Pacific_train.csv and Pacific_test.csv

Dataset Description:

The NHC publishes the tropical cyclone historical database in a format known as HURDAT, short for HURricane DATabase. These databases (Atlantic HURDAT2 and NE/NC Pacific HURDAT2) contain six-hourly information on the location, maximum winds, central pressure, and (starting in 2004) size of all known tropical cyclones and subtropical cyclones.

Columns:

- ID
- Name
- Date
- Time
- Event
- Status
- Latitude
- Longitude
- Maximum Wind
- Minimum Pressure
- Low Wind NE
- Low Wind SE
- Low Wind SW
- Low Wind NW
- Moderate Wind NE
- Moderate Wind SE
- Moderate Wind SW
- Moderate Wind NW
- High Wind NE
- High Wind SE
- High Wind SW
- High Wind NW

Problem Statement

You are provided with two data sets “Pacific_train.csv” and “Pacific_test.csv” having hurricane and typhoon information.

You are required to make a multi-class classification model where the target variable is “Status” to classify hurricanes and typhoons into the correct category.

Carry out the following tasks and select the appropriate features and make classification models using the following algorithms having a **10-fold cross validation score** :

1. Decision Trees (Applying different criterion and choosing the best)
2. Random Forest
3. Naive Bayes
4. SupportVectorClassifier

HINT: Use correlation to select the most appropriate features.

Write python functions for the following and compare the performance of algorithms used above:

1. Recall
2. Precision
3. Accuracy

The Recall, Precision is to be computed for each label and algorithm pair

1. Which is the best model?

Hint: Implement all the above-mentioned models and then calculate the value of recall, precision and accuracy of each of them to finally select the best model.

NOTE: You *MUST* implement all the 4 models mentioned above in the question.

Final Output Sample:

	A
1	GaussianNB
2	0.7

NOTE: Let's say Naive Bayes Algorithm is the best algorithm for the above scenario with accuracy 0.7, then print GaussianNB(**Name of Naive Bayes function in sklearn**) and its respective accuracy score in a CSV file in the above-mentioned format.

Output Format:

- Perform the above operations and write your output to a file named **output.csv**, which should be present at the location **/code/output/output.csv**
- **output.csv** should contain the answer to **each question** on consecutive rows.

DATASETS

- [Training dataset](#)