CSCI-6401-01 **BLACK SHARKS**

WEATHER PREDICTION

Team name: Black Sharks

Team members:

1. Sudheer Kumar Reddy Reddymalla
   sredd1@unh.newhaven.edu
   00719922
2. Durga Guna Sekhar Reddy Savanam
   dsava2@unh.newhaven.edu
   00755099
3. Vamsi Krishna Regalla
   vrega1@unh.newhaven.edu
   00730353

Research Question:

To predict the weather based on the patterns previously collected using data mining techniques. Here, in this case we use Seattle's weather data.

Data Set:

The data set we taken is from Kaggle website. Here is the link for our data set: https://www.kaggle.com/datasets/ananthr1/weather-prediction It has 6 column attributes.

Data Mining Techniques:

- Naïve Bayes Classification
- K-Star algorithm
- SMO model
- Decision Tree and
- Random forest

Parameters and hyperparameters:

Precipitation: All forms in which water falls on the land surface and open water bodies as rain, sleet, snow, hail, or drizzle.

temp_max : shows the results which has maximum temperature it may range from -1.6 to 35.6 .

temp_min : shows the results which has minimum temperature it may range from -7 to 18.3 .

wind: shows the wind speed on those specific dates it may range from 0.4 to 9.5.

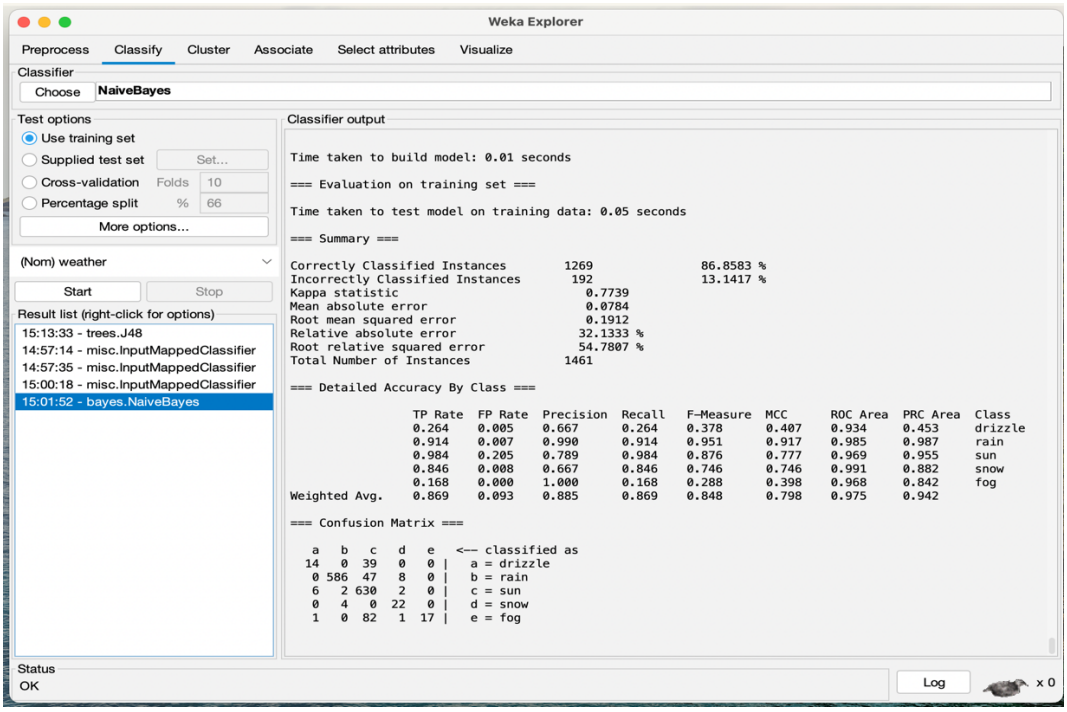
other parameters such as drizzle, rain, sun, snow, fog.

Hardware:

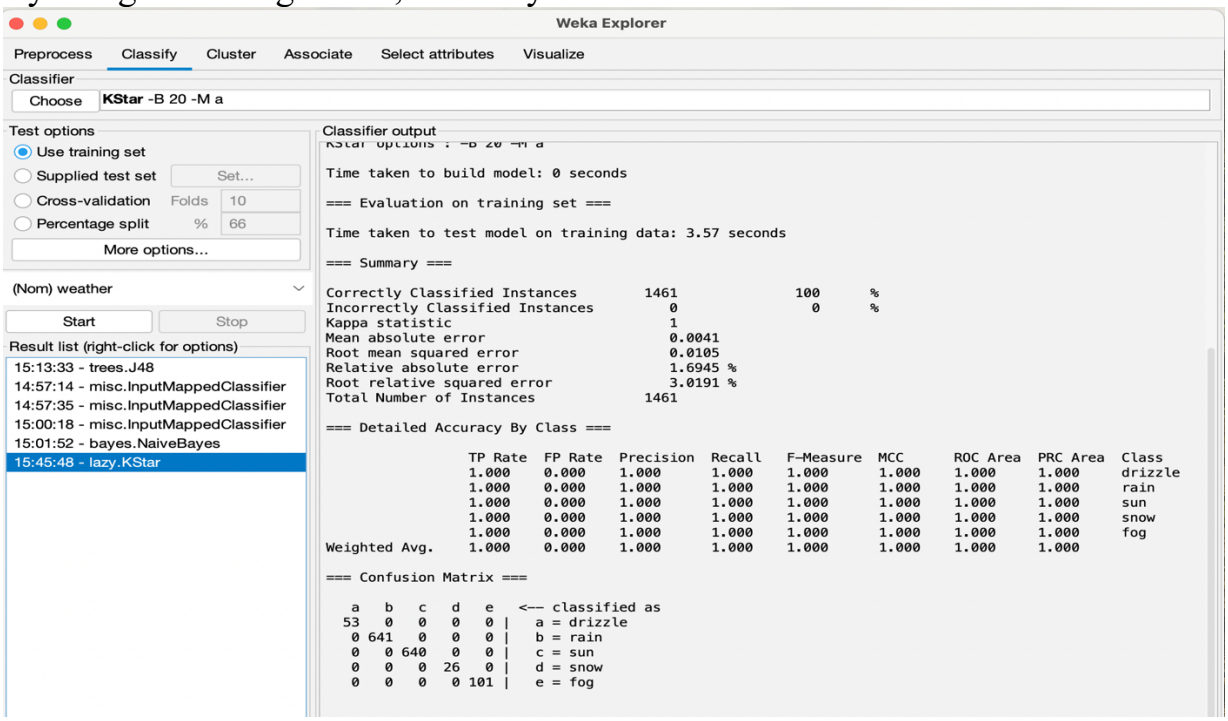
Processor: M1 chip
Ram: 8GB

Outcomes and Visualization Techniques:

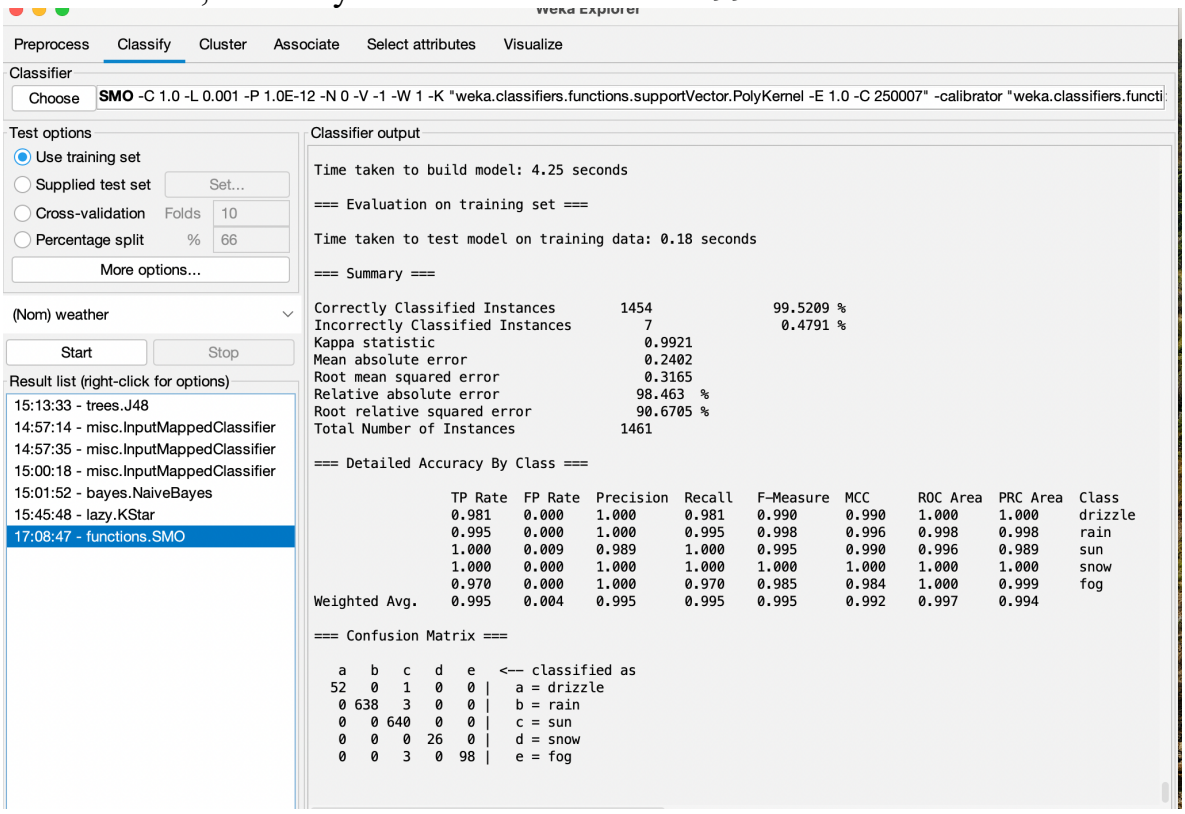
Naïve Bayes Classification: A probabilistic classifier is the Naive Bayes algorithm for classification. It is based on probability models that make substantial assumptions about independence. The independence presumptions frequently do not affect reality. They are therefore viewed as being naive.
By using Naive bayes model, correctly classified instances are 1269 which is 86.85%.



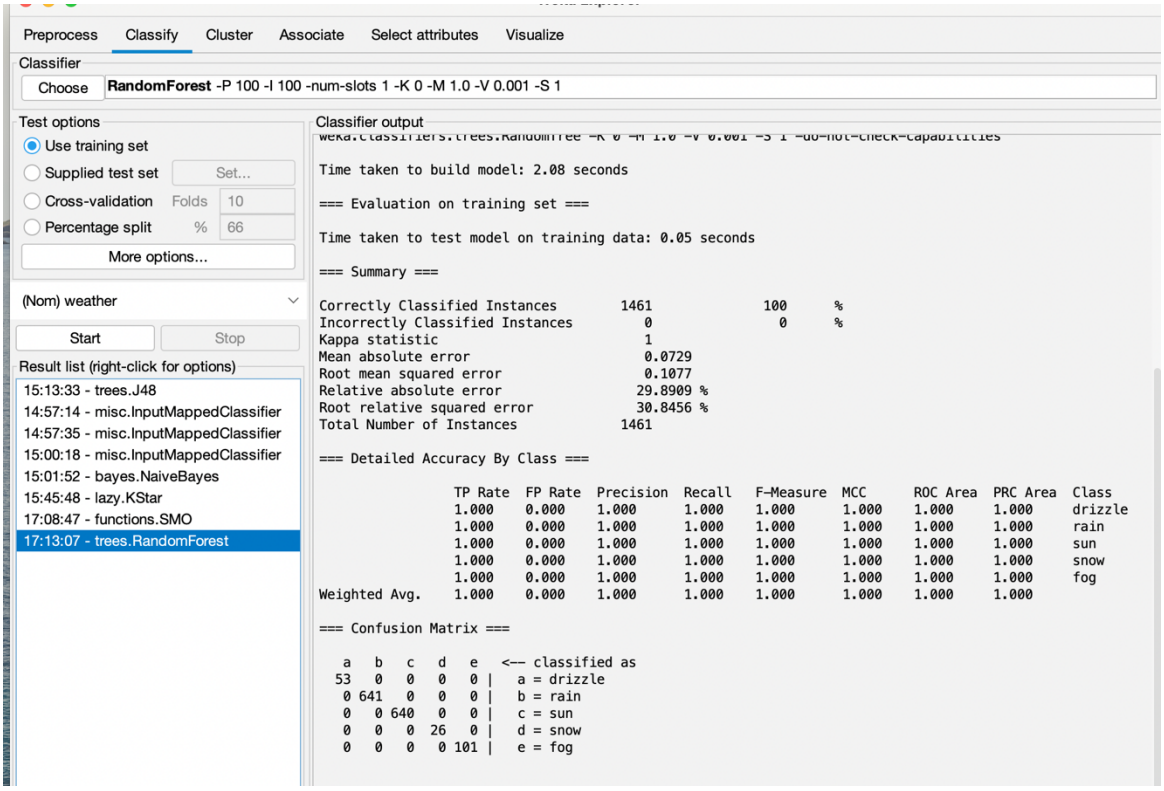
K-Star algorithm: K-star is an instance-based classifier, meaning that it bases the classification of a test instance on the classification of training instances that are like it as specified by some similarity function. Use of a modelling and simulation distance function sets it apart from other instance-based learners.
By using K-star algorithm, correctly classified instances are 100%.



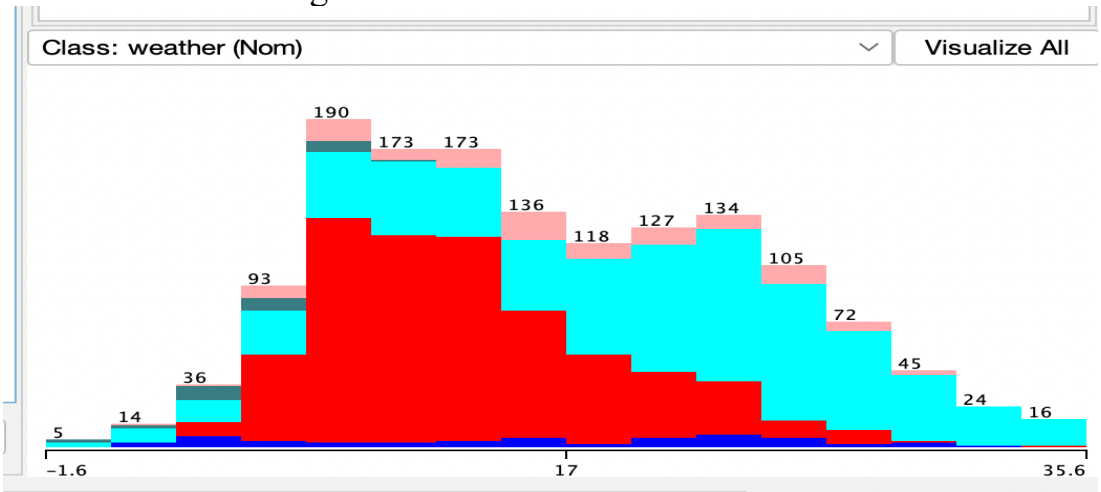
SMO model: is a method for resolving the quadratic programming (QP) problem that appears during training support vector machines (SVM). Support vector machines are frequently trained using SMO, which is implemented by the well-liked LIBSVM tool. As a result, that earlier SVM training techniques were far more complicated and needed pricey third-party QP solvers.
In this model, correctly classified instances are 99%.



Random Forest: Supervised machine learning algorithms like random forest are frequently employed in classification and regression issues. It creates decision trees from several samples, using the majority vote for classification and the average in the case of regression.
In this random forest model, correctly classified instances are 100%.



Visualization: This gives overall information about the weather.



Conclusion:

By comparing all these modelling techniques, we can say that Random Forest and k-star have most instances with 100%. Whereas, Naive Bayes have 86% and SMO having 99 % correctly classified instances.

By using these modelling techniques in Weka tool, we can predict weather.

GitHub:

https://github.com/sudheerredde/BLACKSHARKS.git