**Summary**

***Their Classification***

The purpose of the paper is combining network science with machine learning to classify the ENSO events.

They used the global daily near-surface (1000 hPa) Air Temperature data (SAT), from the National Oceanic and Atmospheric Administration (NOAA) and the Reanalysis I project of the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR). They used this dataset instead of the SST (Sea Surface Temperature), because it also takes temperatures of land areas. This dataset covers the period from 1948 to 2018. For each day, they subtracted its temperature by the average temperature of the same day over the past years to obtain SAT anomalies (SATa).

They firstly applied network science to SATa dataset and extracted nine topological measures for each month from 1948 to 2018 which will be the features in classification:

1. the transitivity
2. the number of links
3. average degree
4. average shortest-path length
5. modularity
6. global average link distance
7. the averages of Coreness
8. Eccentricity
9. Eigenvector centralities.

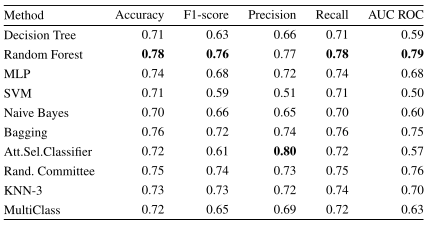
Additionally, each month from 1948 to 2018 has its ONI value.

After constructing a climate network, they applied machine learning to classify El nino and La nina events. Used ten classifier algorithms:

1. **k-Nearest Neighbors** - a new data point is classified based on similarity in the specific group of neighboring data points.
2. **Naive Bayes** - is a probabilistic classifier based on Bayes theorem.
3. **Decision Tree** - build tree in which each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents the target variable.
4. **MultilayerPerceptron** - is a feedforward artificial neural network.
5. **Support-Vector Machine** - constructs a set of hyperplanes in a high-dimensional space that maximizes the distance from each example to the nearest data point on each side.
6. **Bootstrap Aggregation** (Bagging) - a statistical estimation technique which performs some statistical measure like mean from multiple random samples with replacement from the dataset.
7. **Random Forest** - contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
8. **Random Committee** - generates several random base classifiers by using a different random number seed.
9. **AttributeSelectedClassifier** - reduces the dimensionality of training and test data by attribute selection before passing the data to a classifier.
10. **MultiClass** - is a meta classifier for handling multi-class datasets with 2-class classifiers.

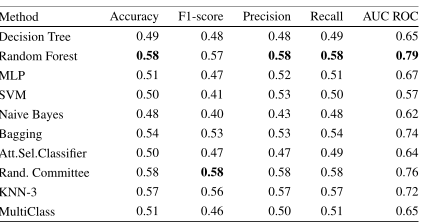
They set up 3 experimental dataset and classified them using ML:

1. Dataset1 - 2 classes with El Nino (EN) and anything else (AE). (EN 234, AE 588)



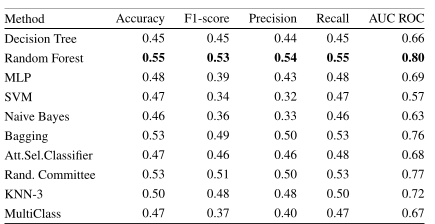
This dataset represents a binary classification (2 classes). Accuracy is around 70% and Random Forest gives the best result.

1. Dataset2 - 3 classes with El Niño (EN), La Niña (LN) events and regular years (RY) that means neither El Niño nor La Niña occurred. (EN 234, LN 227, RY 361)



This dataset is a bit more difficult than Dataset1 because it represents a multiclass classification (3 classes). Accuracy is decreased to 48-58% and Random Forest gives the best result again.

1. Dataset3 - 5 classes with strong El Niño (SEN), weak El Niño (WEN), strong La Niña (SLN), weak La Niña (WLN), and regular year (RY). (SEN 137, SLN 54, RY 361, WEN 97, WLN 143)

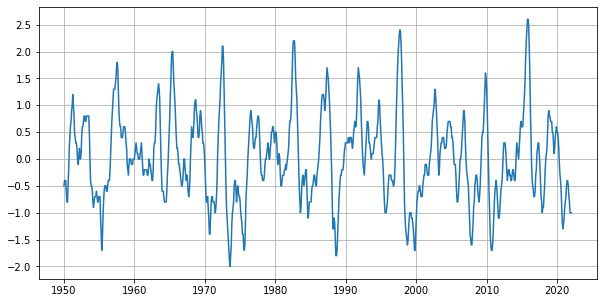


This dataset is more difficult than Dataset1 and Dataset2 because it represents a multiclass classification (5 classes). Accuracy is decreased to 46-55% and Random Forest gives the best result again.

In conclusion, Random Forest classifier gives the best result for 3 datasets and so it is the best model to classify ENSO events.

***My Classification***

I used ONI dataset from Climate Prediction Center. It is the running 3-month mean SST anomaly for the Nino 3.4 region (the tropical Pacific). This dataset covers the period from 1950 to present. The index is defined as the average of monthly SST anomalies. The monthly anomalies are calculated by subtracting the average temperature in the same month over the past 30 years. Then the running three-months average is calculated.



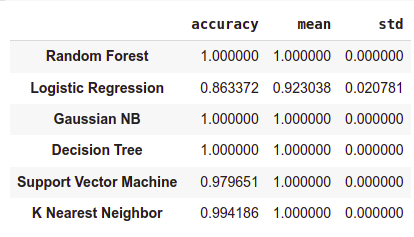
**Classification of WE, ME, SE, VSE and AE**

Firstly, I took 3 months average ONI values as feature data and built target data based on feature data where if oni value is

* greater or equal than 0.5 and less than 1 - 1 (weak el nino interval) (139 values)
* greater or equal than 1 and less than 1.5 - 2 (moderate el nino interval) (55 values)
* greater or equal than 1.5 and less than 2 - 3 (strong el nino interval) (29 values)
* greater or equal than 2 - 4 (very strong el nino interval) (19 values)
* otherwise - 0 (other climate events) (617 values)

Then, split these datasets to create train (60% of all data) and test sets (40% of all data). And applied 6 machine learning classification models:

* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified values 2-2.5 (very strong el nino) as strong, values 1-1.3 (moderate el nino) as weak and value 0.5 (weak el nino) as regular climate event.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified value 2 (very strong el nino) as strong and value 1 (moderate el nino) as weak el nino.
* **K Nearest Neighbor** - classified value 2 (very strong el nino) as strong el nino.

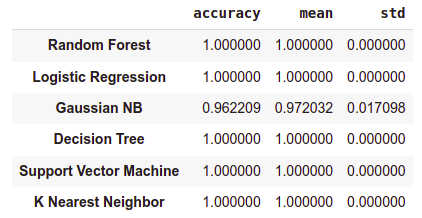


In conclusion, Random Forest, Gaussian NB and Decision Tree classified all ONI values correctly due to their intervals.

**Classification of EN and AE**

if oni value is

* greater or equal than 0.5 - 1 (el nino) (242 values)
* otherwise - 0 (another climate events) (617 values)

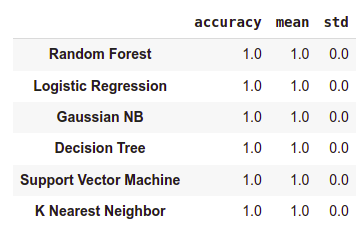


* **Gaussian Naive Bayes** - classified value 0.5 as regular year.
* **Others** - classified all intervals correctly.

**Classification of EN, LN and RY**

if oni value is

* greater or equal than 0.5 - 1 (el nino) (242 values)
* less or equal than -0.5 - 2 (la nina) (248 values)
* otherwise - 0 (another climate events) (369 values)

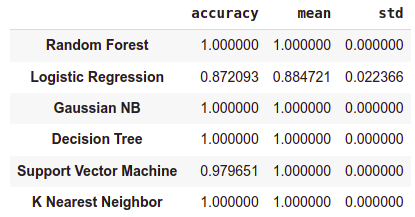


* **ALL** - classified all intervals correctly.

**Classification of SEN, SLN, RY, WEN and WLN**

if oni value is

* less or equal than -1 - 0 (strong la nina) (91 values)
* greater than -1 and less or equal than -0.5 - 1 (weak la nina) (157 values)
* greater than -0.5 and less than 0.5 - 2 (regular year) (369 values)
* greater or equal than 0.5 and less than 1 - 3 (weak el nino) (139 values)
* greater or equal than 1 - 4 (strong el nino) (103 values)

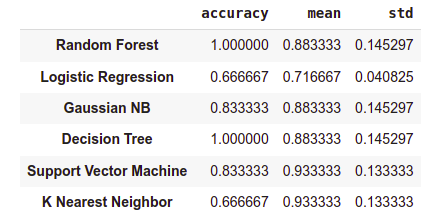


* **Logistic Regression** - classified values -0.5 and 0.5 as regular year, -1 as weak la nina and 1 as weak el nino.
* **SVM** - classified value 1 as weak el nino.
* **Others** - classified all intervals correctly.

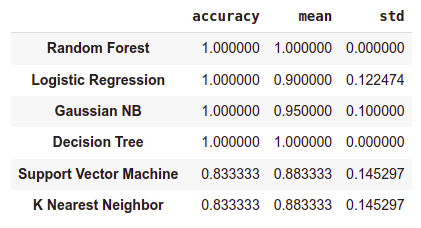
**Classification of 3 VSE events for 1.5 - 4 years**

There have been three very strong El Nino (VSE) events since 1950 : 1982-1983, 1997-1998, 2015-2016. I also classified these VSE events for 1.5, 2, 2.5, 3, 3.5 and 4 years with the same method as above. Here target data takes 2 labels: when oni value is greater or equal than 2, label will be 1, otherwise 0.

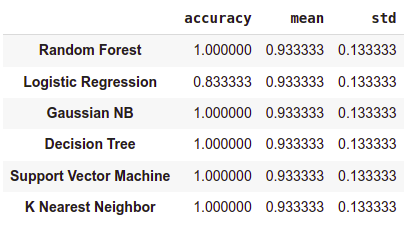
* 1982-1983 VSE for 1.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified values 2 and 2.2 as strong.
* **Gaussian Naive Bayes** - classified value 2 as strong.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified value 1.9 as very strong.
* **K Nearest Neighbor** - classified values 2 and 2.2 as strong.



* 1997-1998 VSE for 1.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified all intervals correctly.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified value 1.9 as very strong.
* **K Nearest Neighbor** - classified value 1.9 as very strong.

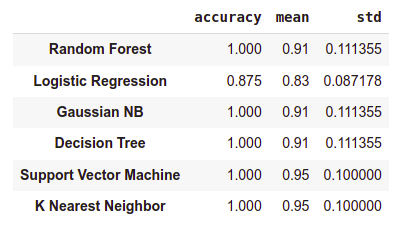


* 2015-2016 VSE for 1.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified value 2.1 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.

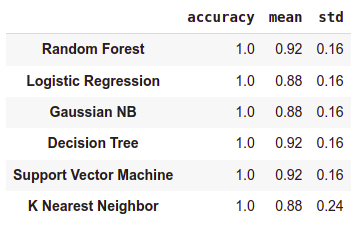


In conclusion, Random Forest and Decision Tree classified all ONI values correctly due to their intervals for each dataset.

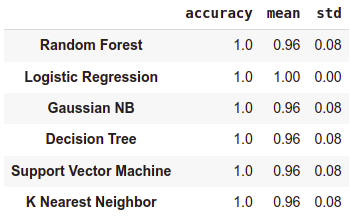
* 1982-1983 VSE for 2 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified value 2.2 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.



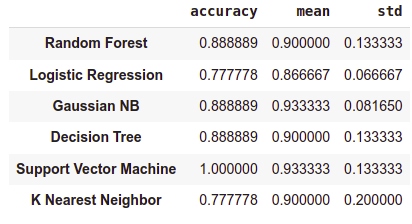
* 1997-1998 VSE for 2 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified all intervals correctly.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.



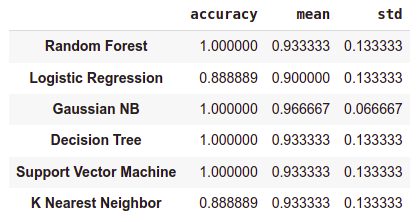
* 2015-2016 VSE for 2 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified all intervals correctly.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.



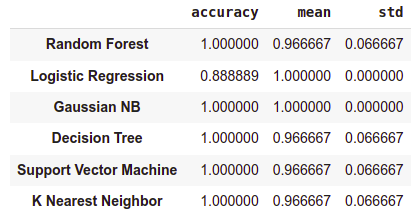
* 1982-1983 VSE for 2.5 year
* **Random Forest** - classified value 1.9 as very strong.
* **Logistic Regression** - classified value 2.2 as strong.
* **Gaussian Naive Bayes** - classified value 1.9 as very strong.
* **Decision Tree** - classified value 1.9 as very strong.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified value 2.2 as strong.



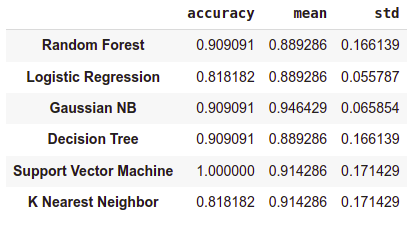
* 1997-1998 VSE for 2.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified value 2.2 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified value 1.9 as very strong.



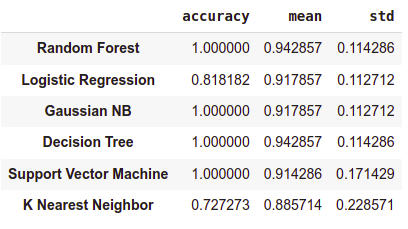
* 2015-2016 VSE for 2.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified value 2.1 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.



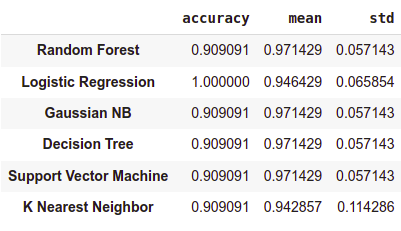
* 1982-1983 VSE for 3 year
* **Random Forest** - classified value 2 as strong.
* **Logistic Regression** - classified values 2 and 2.2 as strong.
* **Gaussian Naive Bayes** - classified value 2 as strong.
* **Decision Tree** - classified value 2 as strong.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified values 2 and 2.2 as strong.



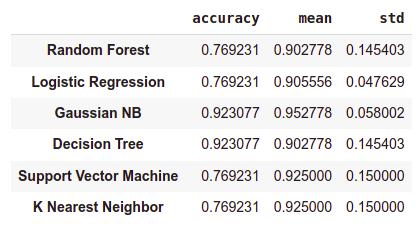
* 1997-1998 VSE for 3 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified values 2.1 and 2.3 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified values 2.1, 2.3 and 2.4 as strong.



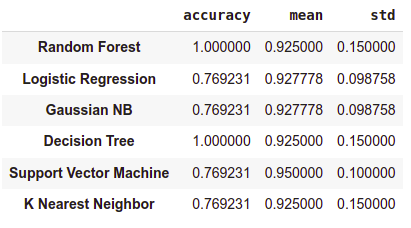
* 2015-2016 VSE for 3 year
* **Random Forest** - classified value 1.9 as very strong.
* **Logistic Regression** - classified all intervals correctly.
* **Gaussian Naive Bayes** - classified value 1.9 as very strong.
* **Decision Tree** - classified value 1.9 as very strong.
* **Support Vector Machine** - classified value 1.9 as very strong.
* **K Nearest Neighbor** - classified value 1.9 as very strong.



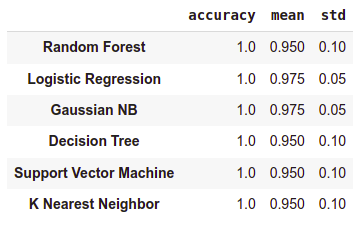
* 1982-1983 VSE for 3.5 year
* **Random Forest** - classified values 2 and 2.2 as strong.
* **Logistic Regression** - classified values 2 and 2.2 as strong.
* **Gaussian Naive Bayes** - classified value 2 as strong.
* **Decision Tree** - classified value 2 as strong.
* **Support Vector Machine** - classified values 2 and 2.2 as strong.
* **K Nearest Neighbor** - classified values 2 and 2.2 as strong.



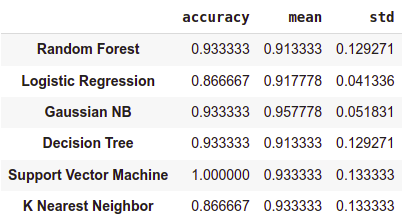
* 1997-1998 VSE for 3.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified values 2.3 and 2.4 as strong.
* **Gaussian Naive Bayes** - classified values 2.3 and 2.4 as strong.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified values 2.3 and 2.4 as strong.
* **K Nearest Neighbor** - classified values 2.3 and 2.4 as strong.



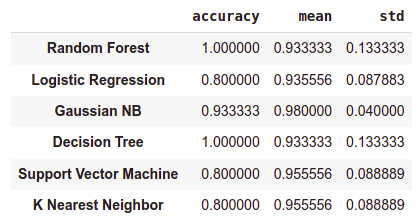
* 2015-2016 VSE for 3.5 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified all intervals correctly.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.



* 1982-1983 VSE for 4 year
* **Random Forest** - classified value 2 as strong.
* **Logistic Regression** - classified values 2 and 2.2 as strong.
* **Gaussian Naive Bayes** - classified value 2 as strong.
* **Decision Tree** - classified value 2 as strong.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified values 2 and 2.2 as strong.



* 1997-1998 VSE for 4 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified values 2.1, 2.3 and 2.4 as strong.
* **Gaussian Naive Bayes** - classified value 2.1 as strong.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified values 2.1, 2.3 and 2.4 as strong.
* **K Nearest Neighbor** - classified values 2.1, 2.3 and 2.4 as strong.



* 2015-2016 VSE for 4 year
* **Random Forest** - classified all intervals correctly.
* **Logistic Regression** - classified value 2.2 as strong.
* **Gaussian Naive Bayes** - classified all intervals correctly.
* **Decision Tree** - classified all intervals correctly.
* **Support Vector Machine** - classified all intervals correctly.
* **K Nearest Neighbor** - classified all intervals correctly.

