PREDICTING DROUGHT USING GEOSPATIAL DATA

Project Done By,

Srija M

ABSTRACT

- Drought is a long period of time when a region experiences less than normal precipitation.
- The datasets are utilised from 3 platforms : (NASA Langley Research Centre (LaRC), National Drought Mitigation Centre, Harmonised World Soil Database).
- The dataset is consist of Meteorological indices and Geospatial Data.
- Used 25 Machine Learning Algorithms and Keras and Tensor-flow in Deep Learning.
- Using GIS tools (QGIS) one can analyse , edit, predict and visualise the drought events.

PROBLEM DEFINITION

- Meteorological drought indices (**USDM United States Drought Monitor**) and soil datasets are used to predict the drought.

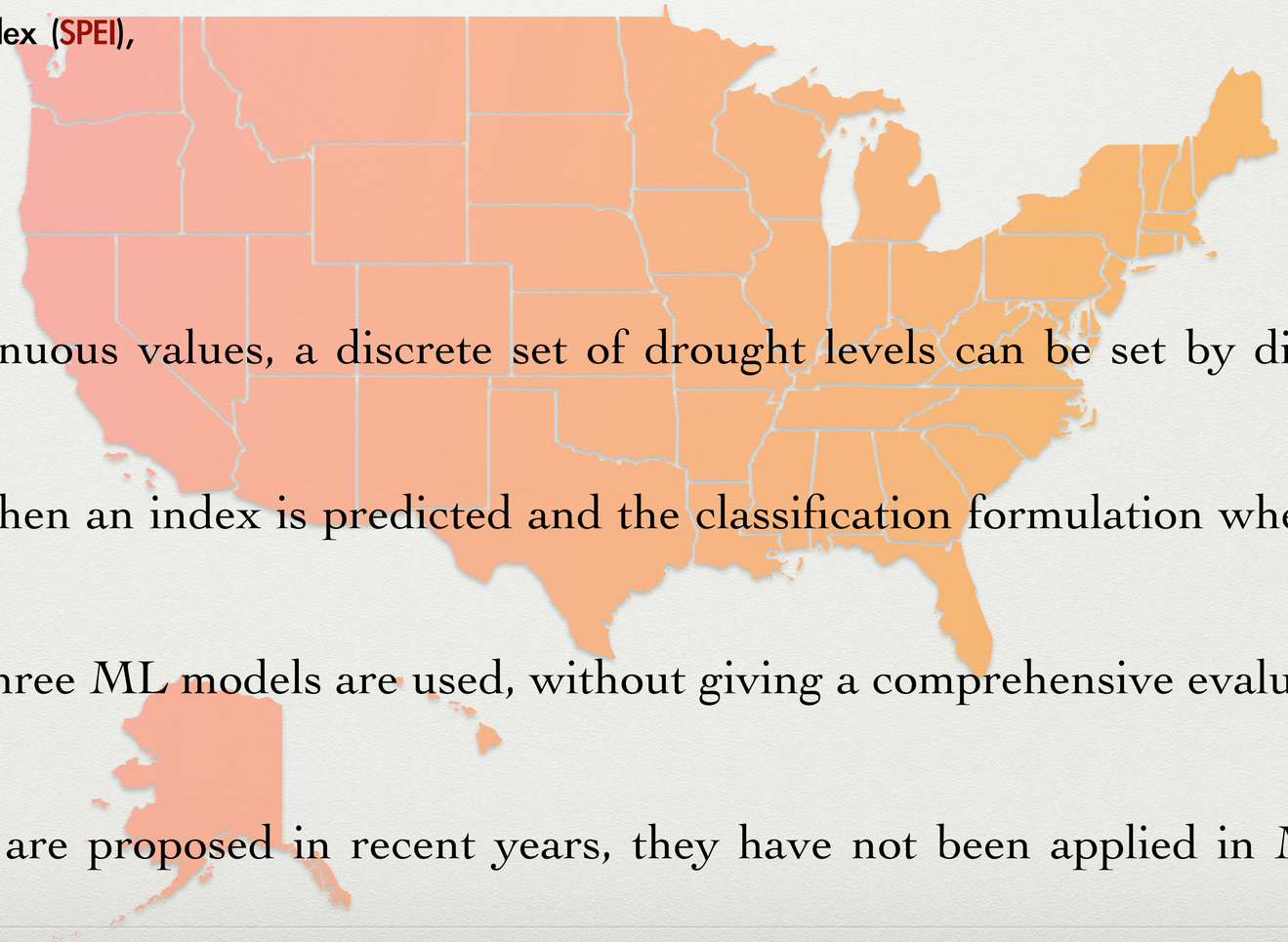
- Droughts can be forecasted using prediction models by anticipating these indices.

- Allowing society to take preventative actions in the case of a drought.

EXISTING SYSTEM

- Various drought indices are used in previous studies includes

  - Standard Precipitation Evaporation Index (SPEI),
  - Standardized Precipitation Index (SPI)
  - Groundwater Resource Index (GRI)
  - Effective Drought Index (EDI).



- While these indices are continuous values, a discrete set of drought levels can be set by dividing different value ranges.

- The regression formulation when an index is predicted and the classification formulation when the drought level is predicted.

- In most studies, only two or three ML models are used, without giving a comprehensive evaluation or consideration for other possible alternatives.

- Many Deep learning models are proposed in recent years, they have not been applied in Meteorological studies yet.

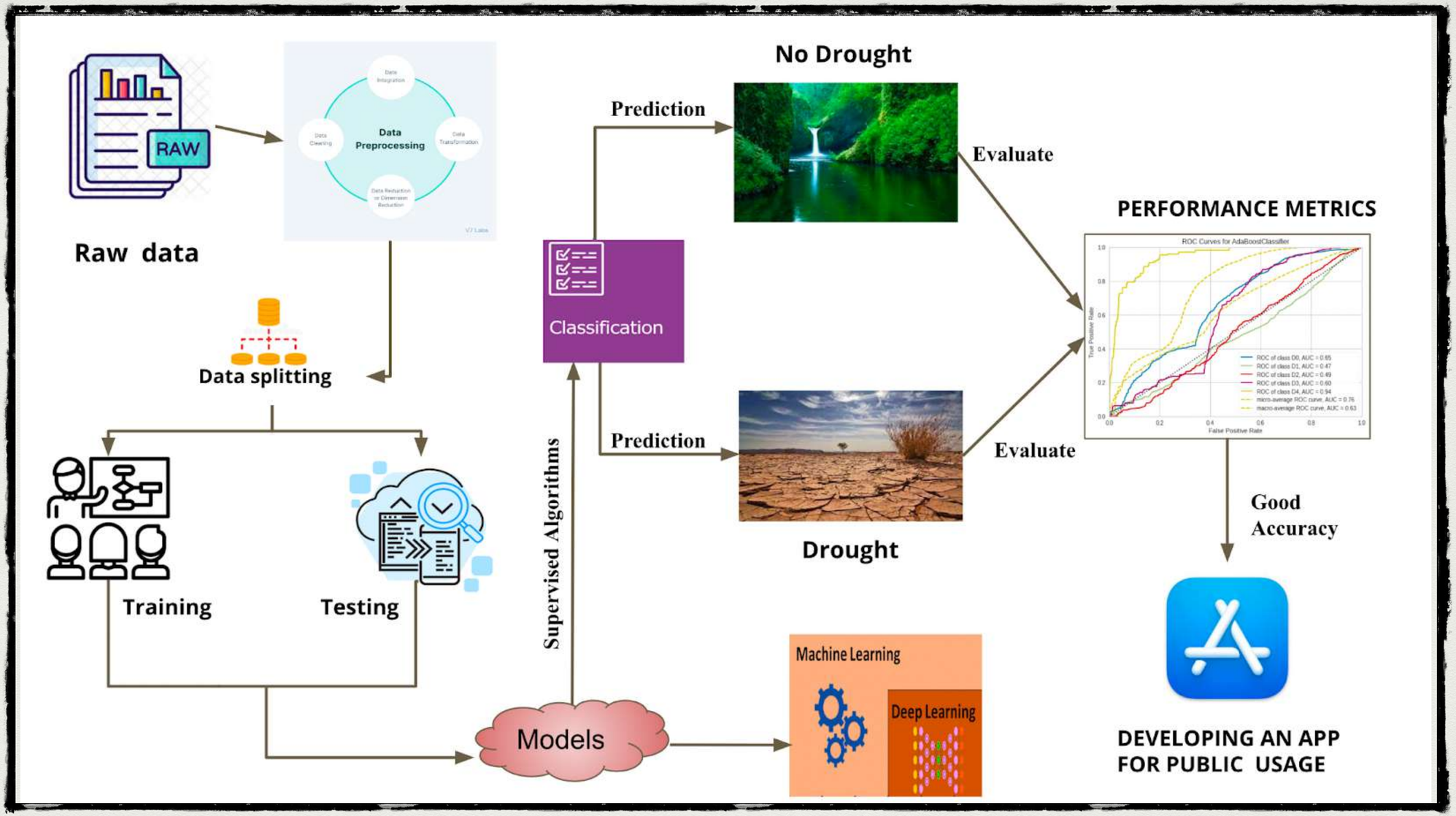
PROPOSED SYSTEM

- In this study, DSI (Drought Severity Index)is used.
- The main aim is to give a comprehensive evaluation that covers most of the popular machine learning and deep learning models for drought prediction.
- GIS tool is used to visualise the drought affected area which is easy to understand by the user.

LITERATURE REVIEW

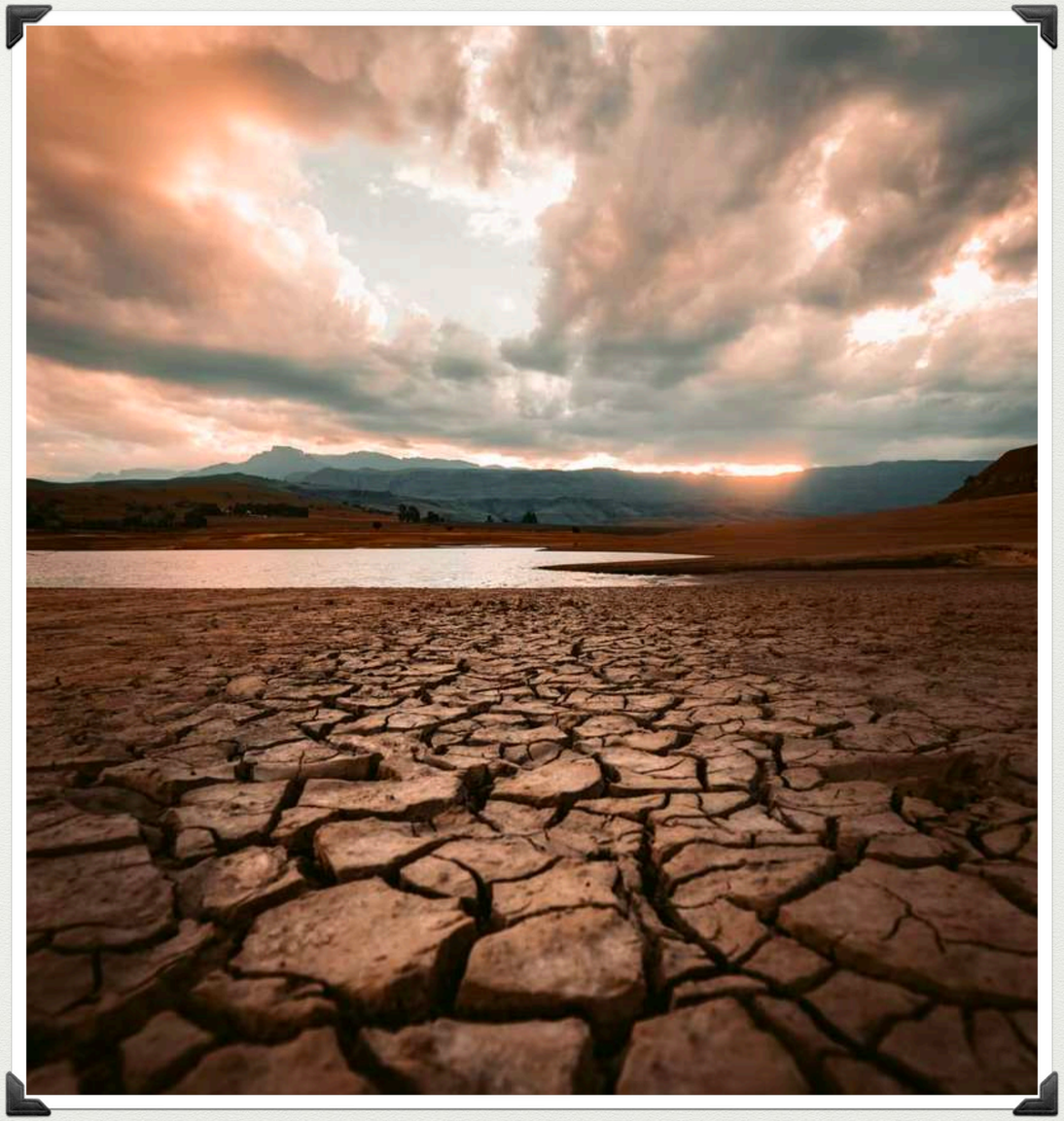
Authors	Methodology Used	Their Findings
Sheffield et al.	• downscaling technique	• decades of weather data are required
Guttman(1998)	• comparison done between PDSI and SPI indices	• SPI are more successful than PDSI
kulkarni(2015)	• atmospheric electricity	• easy to calculate and does not involve the use of probability
Otkin et al. (2015) and Park et al. (2018)	• short-term forecasting	• Useless for the sudden changes
Hao et al.	• Designed a system to predict soil moisture & severity of drought focus on seasonal drought	• global size - difficult to use and to develop regional drought mitigation plans

PROPOSED ARCHITECTURE & MODULE DESCRIPTIONS



SYSTEM ARCHITECTURE USING ML AND DL ALGORITHM

DATA DESCRIPTION



Dataset 1 (Binary Classification) year 2000-2020

No. of columns : 21
No. of rows : 23,841,468
Target value : Score
0 - No drought
1 - Drought

Split	Year Range	Data Sample Number	Data Percentage (appx)
Training	2000-2016	19300680	47%
Validation	2017-2018	2268840	10%
Testing	2019-2020	2271948	43%

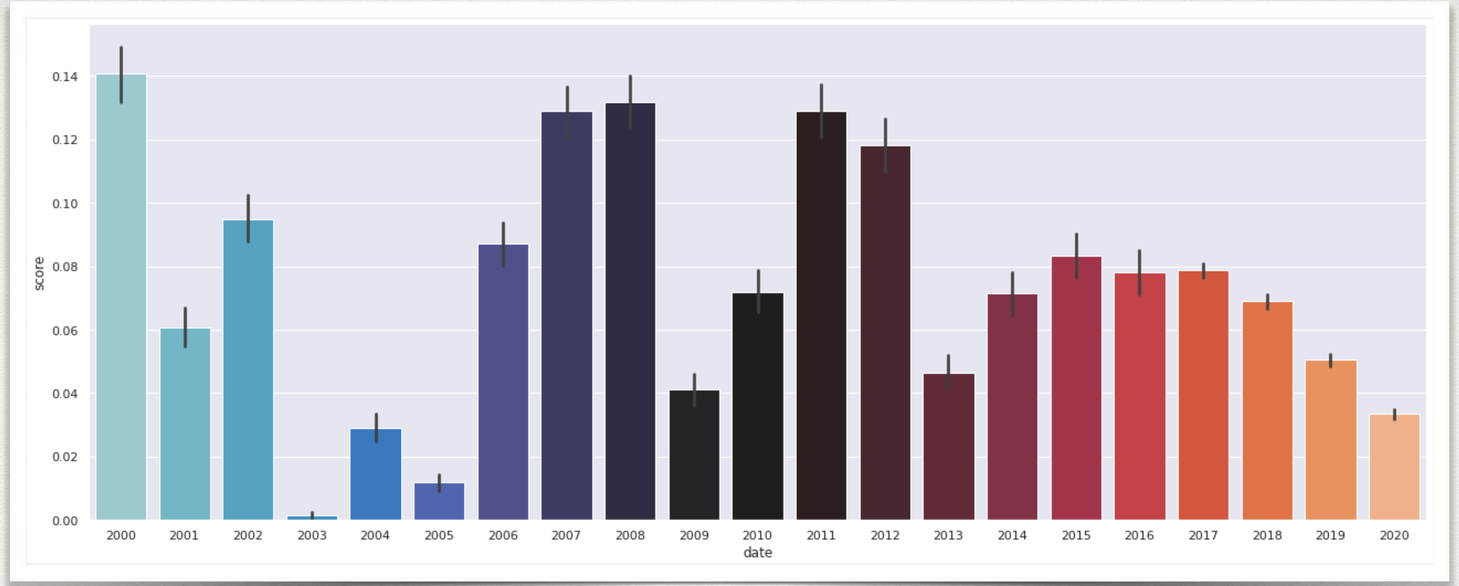
Dataset 2 (Multi-class Classification) year 2000-2021

No. of columns : 26
No. Of rows. : 56792
Target value : LEVEL

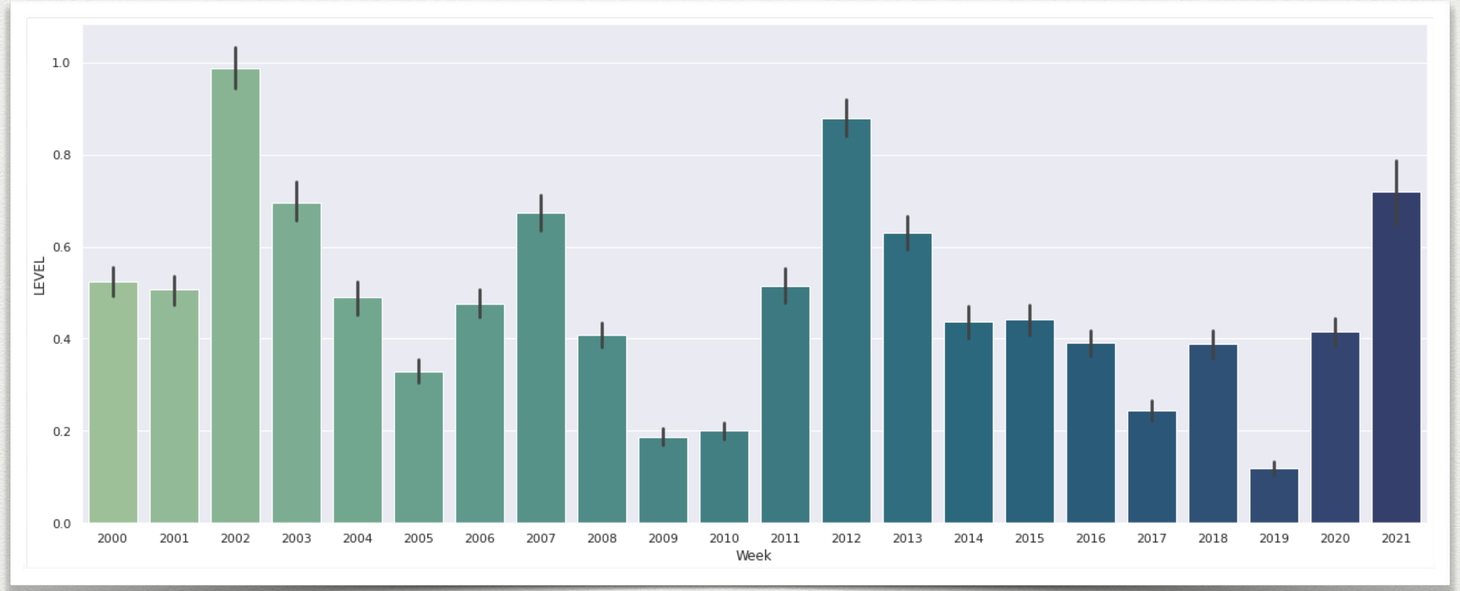
Category of drought types	Threshold value	Description
D0	Less than 0	Abnormally dry
D1	0 to 1	Moderate drought
D2	1 to 2	Severe drought
D3	2 to 3	Extreme drought
D4	3 to 4 or more	Exceptional drought

S.NO	INDICATOR	DESCRIPTION
1.	Week	certain dates
2.	State	Name of the US State
3.	Postal Code	US postal code
4.	Fips	Country code of USA
5.	Longitude	Longitude of Conus
6.	Latitude	Latitude of Conus
7.	WS10M_MIN	Minimum Wind Speed at 10 Meters (m/s)
8.	QV2M	Specific Humidity at 2 Meters (g/kg)
9.	T2M_RANGE	Temperature Range at 2 Meters (C)
10.	WS10M	Wind Speed at 10 Meters (m/s)
11.	T2M	Temperature at 2 Meters (C)
12.	WS50M_MIN	Minimum Wind Speed at 50 Meters (m/s)
13.	T2M_MAX	Maximum Temperature at 2 Meters (C)
14.	WS50M	Wind Speed at 50 Meters (m/s)
15	TS	Earth Skin Temperature (C)
16.	WS50M_RANGE	Wind Speed Range at 50 Meters (m/s)
17.	WS50M_MAX	Maximum Wind Speed at 50 Meters (m/s)
18.	WS10M_MAX	Maximum Wind Speed at 10 Meters (m/s)
19.	WS10M_RANGE	Wind Speed Range at 10 Meters (m/s)
20.	PS	Surface Pressure (kPa)
21.	T2MDEW	Dew/Frost Point at 2 Meters (C)
22.	T2M_MIN	Minimum Temperature at 2 Meters (C)
23.	T2MWET	Wet Bulb Temperature at 2 Meters (C)
24.	PRECTOT	Precipitation (mm day-1)
25..	Aland_SQMI	Land Soil Quality
26.	LEVEL	Drought level from D0,D1,D2,D3,D4

S.NO	INDICATOR	DESCRIPTION
1.	WS10M_MIN	Minimum Wind Speed at 10 Meters (m/s)
2.	QV2M	Specific Humidity at 2 Meters (g/kg)
3.	T2M_RANGE	Temperature Range at 2 Meters (C)
4.	WS10M	Wind Speed at 10 Meters (m/s)
5.	T2M	Temperature at 2 Meters (C)
6.	WS50M_MIN	Minimum Wind Speed at 50 Meters (m/s)
7.	T2M_MAX	Maximum Temperature at 2 Meters (C)
8.	WS50M	Wind Speed at 50 Meters (m/s)
9.	TS	Earth Skin Temperature (C)
10.	WS50M_RANGE	Wind Speed Range at 50 Meters (m/s)
11.	WS50M_MAX	Maximum Wind Speed at 50 Meters (m/s)
12.	WS10M_MAX	Maximum Wind Speed at 10 Meters (m/s)
13.	WS10M_RANGE	Wind Speed Range at 10 Meters (m/s)
14.	PS	Surface Pressure (kPa)
15	T2MDEW	Dew/Frost Point at 2 Meters (C)
16.	T2M_MIN	Minimum Temperature at 2 Meters (C)
17.	T2MWET	Wet Bulb Temperature at 2 Meters (C)
18.	PRECTOT	Precipitation (mm day-1)
19.	Score	Target (0- No drought , 1- Drought)



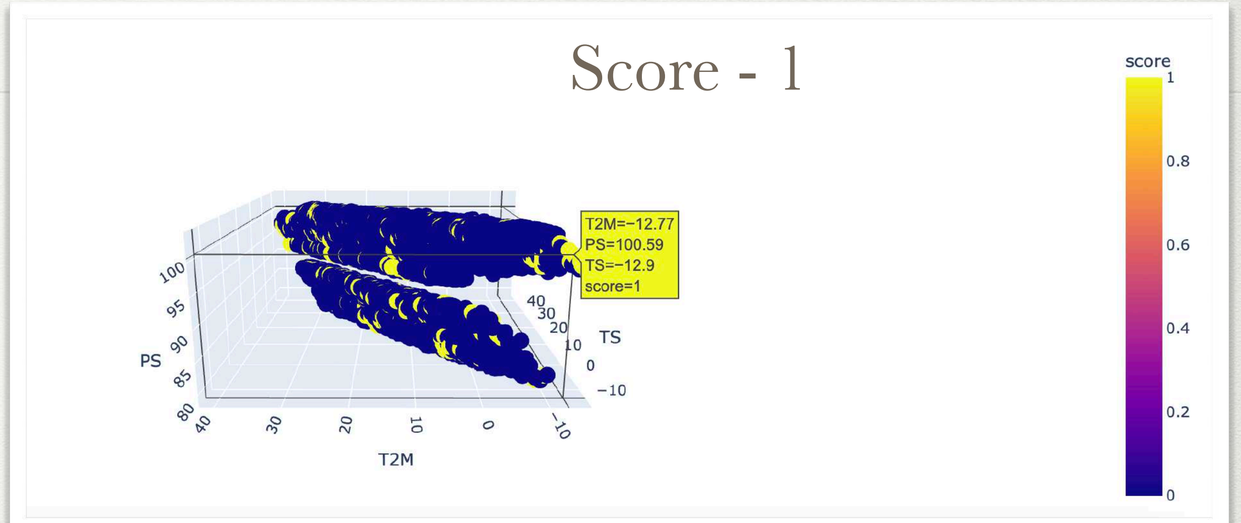
Drought history by year (2000-2020) Binary class

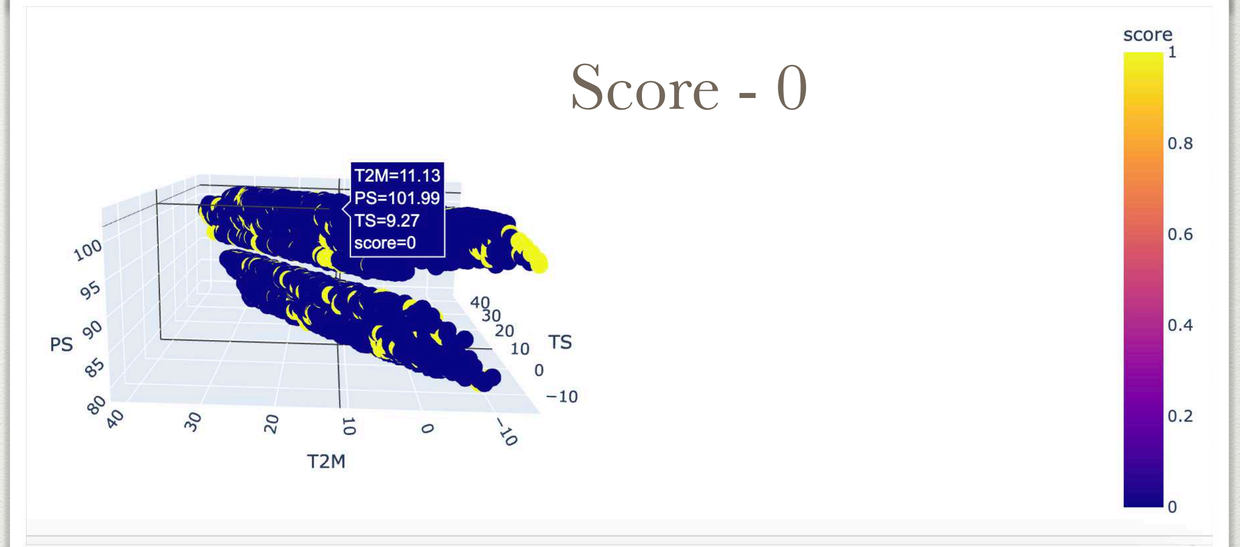


Drought history using soil quality by year (2000-2021) Multi-class class

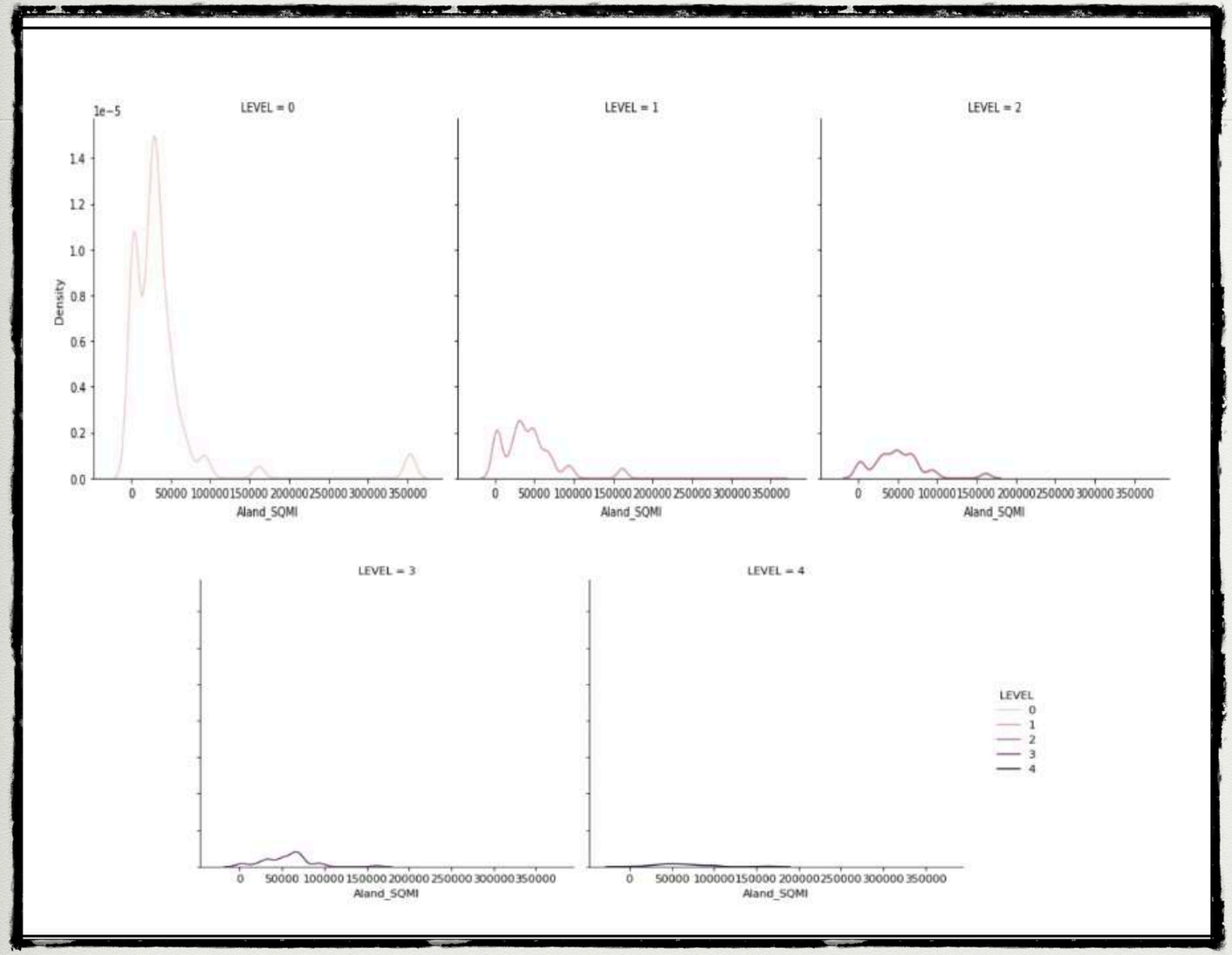
DATASET -1

S.NO	INDICATOR	DESCRIPTION
1.	WS10M_MIN	Minimum Wind Speed at 10 Meters (m/s)
2.	QV2M	Specific Humidity at 2 Meters (g/kg)
3.	T2M_RANGE	Temperature Range at 2 Meters (C)
4.	WS10M	Wind Speed at 10 Meters (m/s)
5.	T2M	Temperature at 2 Meters (C)
6.	WS50M_MIN	Minimum Wind Speed at 50 Meters (m/s)
7.	T2M_MAX	Maximum Temperature at 2 Meters (C)
8.	WS50M	Wind Speed at 50 Meters (m/s)
9.	TS	Earth Skin Temperature (C)
10.	WS50M_RANGE	Wind Speed Range at 50 Meters (m/s)
11.	WS50M_MAX	Maximum Wind Speed at 50 Meters (m/s)
12.	WS10M_MAX	Maximum Wind Speed at 10 Meters (m/s)
13.	WS10M_RANGE	Wind Speed Range at 10 Meters (m/s)
14.	PS	Surface Pressure (kPa)
15	T2MDEW	Dew/Frost Point at 2 Meters (C)
16.	T2M_MIN	Minimum Temperature at 2 Meters (C)
17.	T2MWET	Wet Bulb Temperature at 2 Meters (C)
18.	PRECTOT	Precipitation (mm day-1)
19.	Score	Target (0- No drought , 1- Drought)





DATASET 2



S.NO	INDICATOR	DESCRIPTION
1.	Week	certain dates
2.	State	Name of the US State
3.	Postal Code	US postal code
4.	Fips	Country code of USA
5.	Longitude	Longitude of Conus
6.	Latitude	Latitude of Conus
7.	WS10M_MIN	Minimum Wind Speed at 10 Meters (m/s)
8.	QV2M	Specific Humidity at 2 Meters (g/kg)
9.	T2M_RANGE	Temperature Range at 2 Meters (C)
10.	WS10M	Wind Speed at 10 Meters (m/s)
11.	T2M	Temperature at 2 Meters (C)
12.	WS50M_MIN	Minimum Wind Speed at 50 Meters (m/s)
13.	T2M_MAX	Maximum Temperature at 2 Meters (C)
14.	WS50M	Wind Speed at 50 Meters (m/s)
15	TS	Earth Skin Temperature (C)
16.	WS50M_RANGE	Wind Speed Range at 50 Meters (m/s)
17.	WS50M_MAX	Maximum Wind Speed at 50 Meters (m/s)
18.	WS10M_MAX	Maximum Wind Speed at 10 Meters (m/s)
19.	WS10M_RANGE	Wind Speed Range at 10 Meters (m/s)
20.	PS	Surface Pressure (kPa)
21.	T2MDEW	Dew/Frost Point at 2 Meters (C)
22.	T2M_MIN	Minimum Temperature at 2 Meters (C)
23.	T2MWET	Wet Bulb Temperature at 2 Meters (C)
24.	PRECTOT	Precipitation (mm day-1)
25..	Aland_SQMI	Land Soil Quality
26.	LEVEL	Drought level from D0,D1,D2,D3,D4

METHODOLOGY

DOMAIN

- Machine Learning

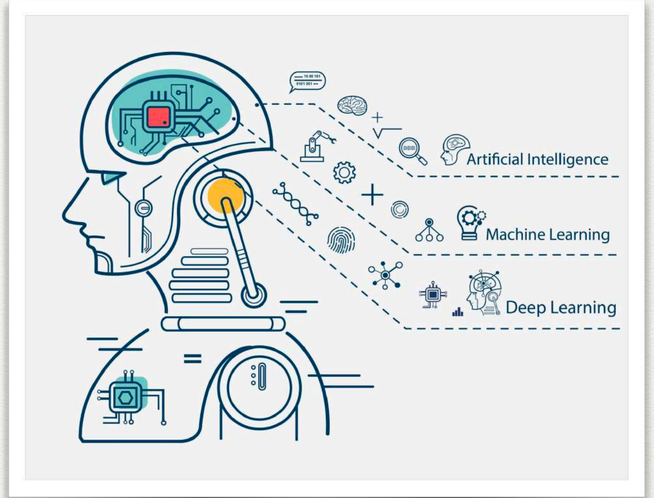
Machine learning is one of the tree of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

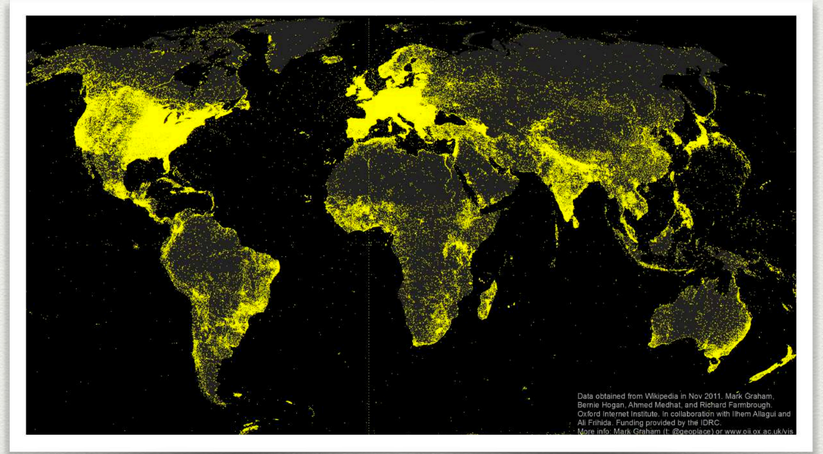
- Deep Learning

Deep Learning is a computer software that mimics the network of neurons in the brain.

- GeoSpatial Data

Geospatial data (also known as “spatial data”) is used to describe data that represents features or objects on the Earth’s surface.





ALGORITHM

- **Binary Classification**

Classification problems with two class labels are referred to as binary classification. "Normal" "Abnormal.

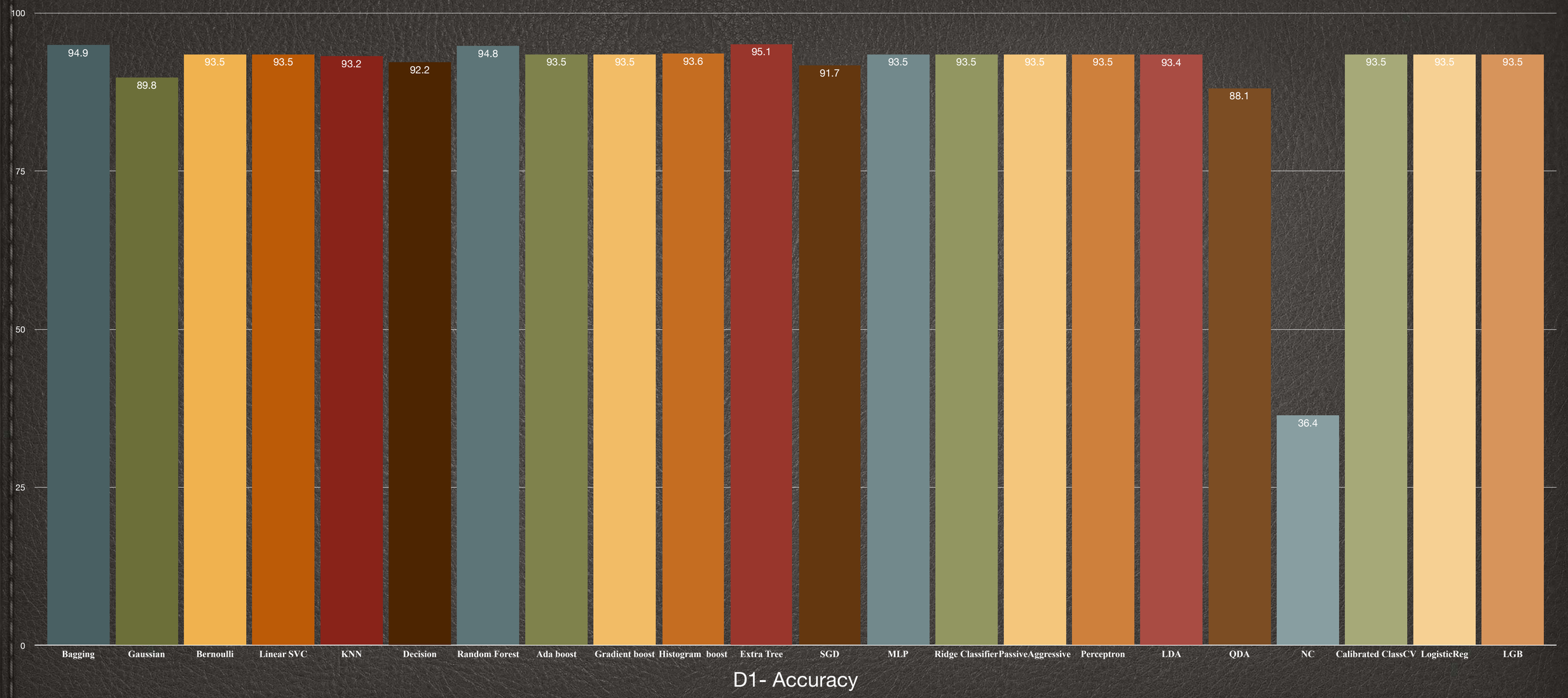
"No Drought" is the normal state and "**Drought"** is the abnormal state.

- **Multiclass Classification**

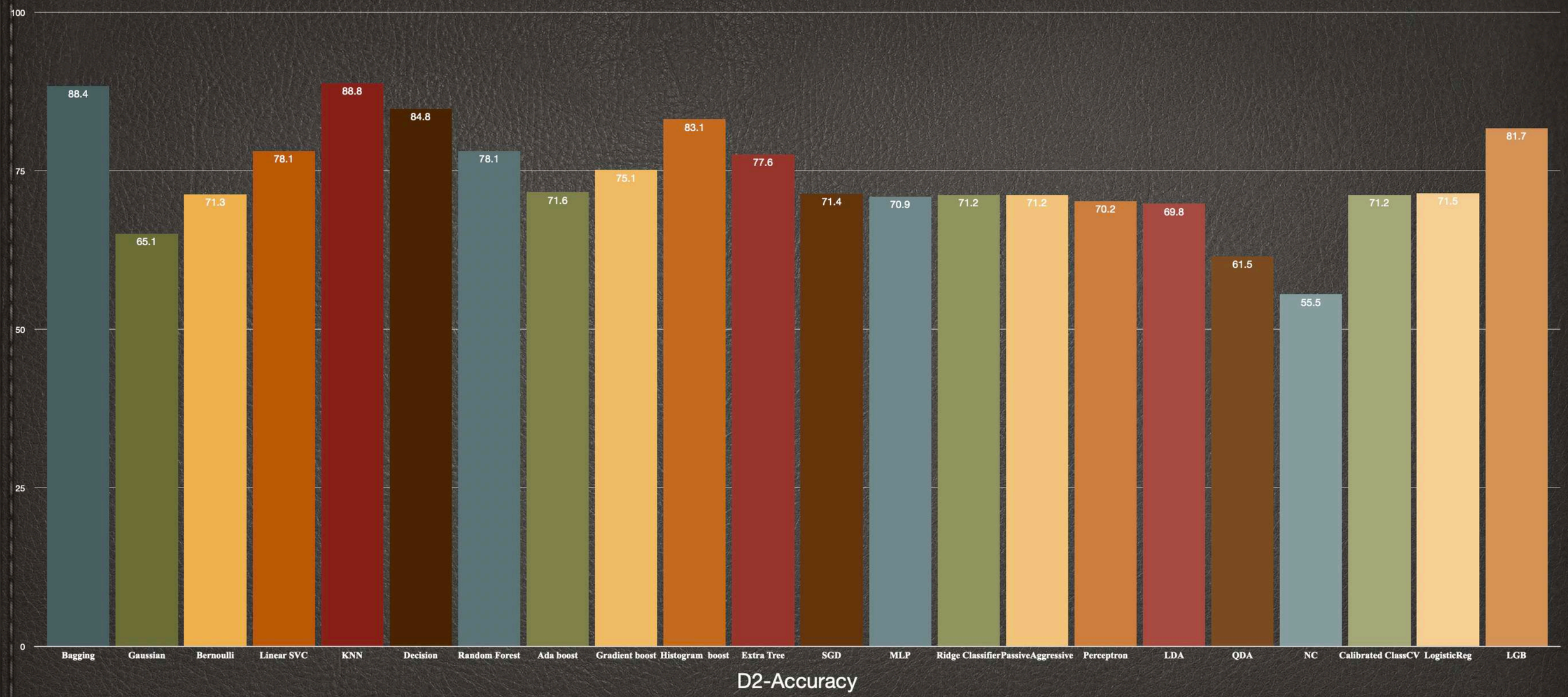
Classification tasks with more than two class labels are referred to as Multi-class Classification tasks.

More than two labels : **"D0", "D1", "D2", "D3", "D4"**

ACCURACY Comparison of ML models



ACCURACY Comparison of ML models



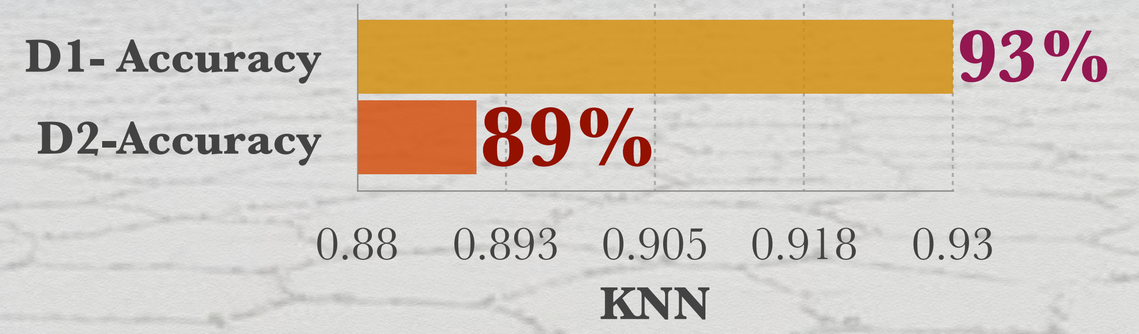
D2-Accuracy

Machine Learning

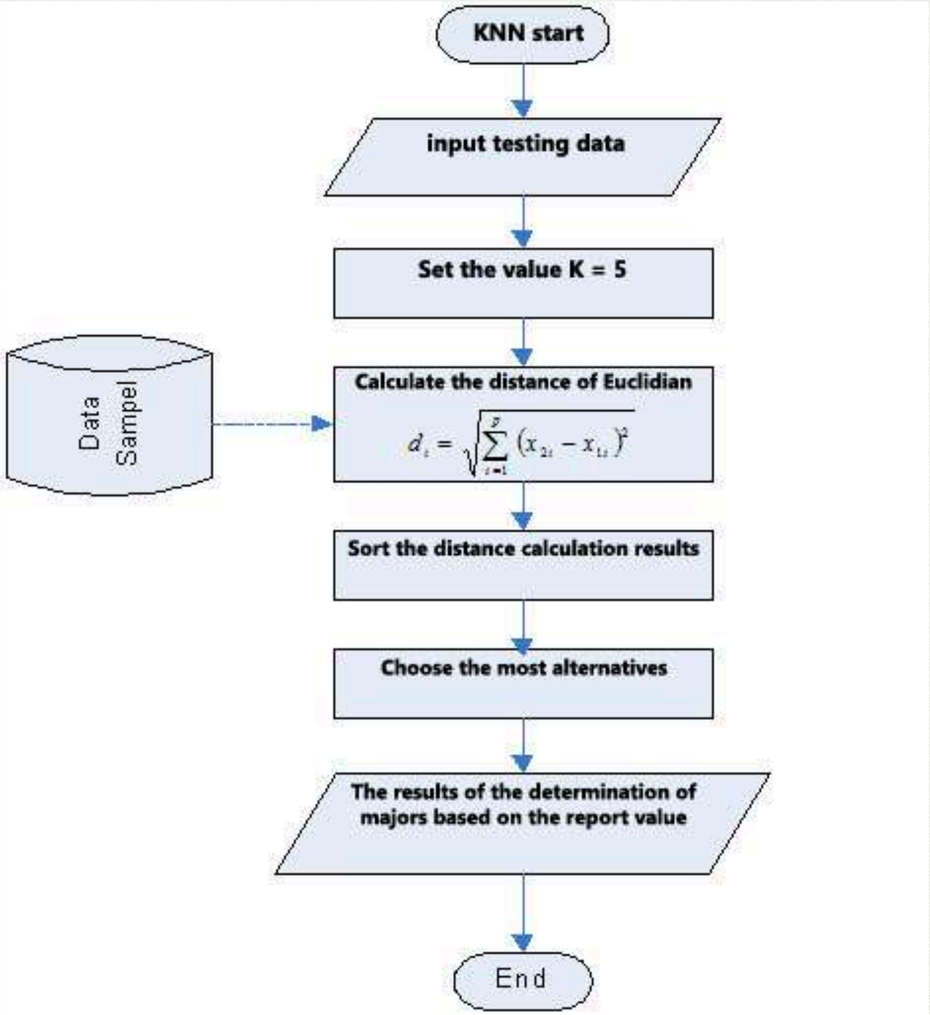
BEST WORKING

- **K Nearest Neighbour**

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on **Supervised Learning technique**.
- K-NN algorithm can be used for **Regression** as well as for **Classification**.
- K-NN is a **non-parametric algorithm.**
- It is also called a **lazy learner algorithm.**
- Feature Similarity.

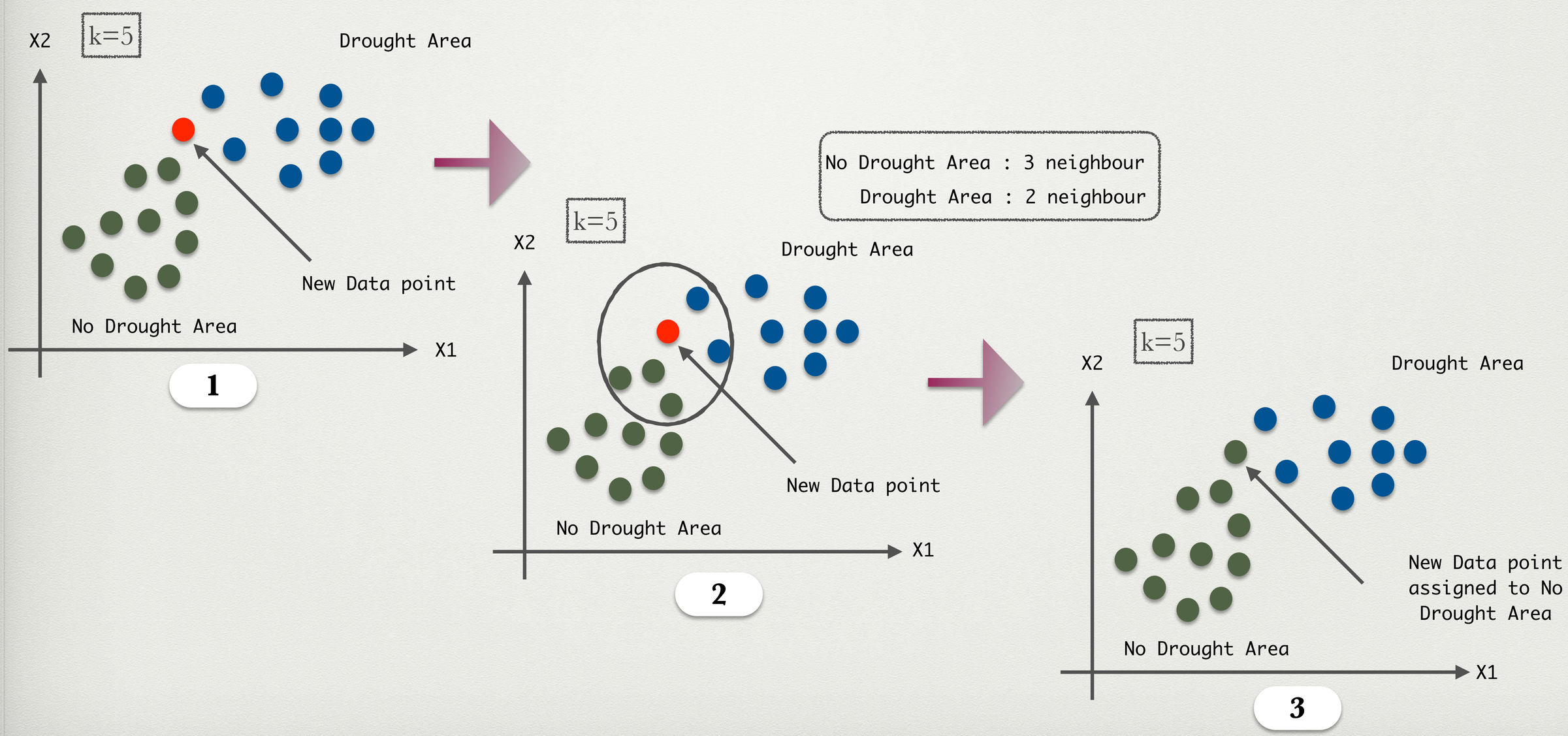


WORKING OF KNN CLASSIFIER

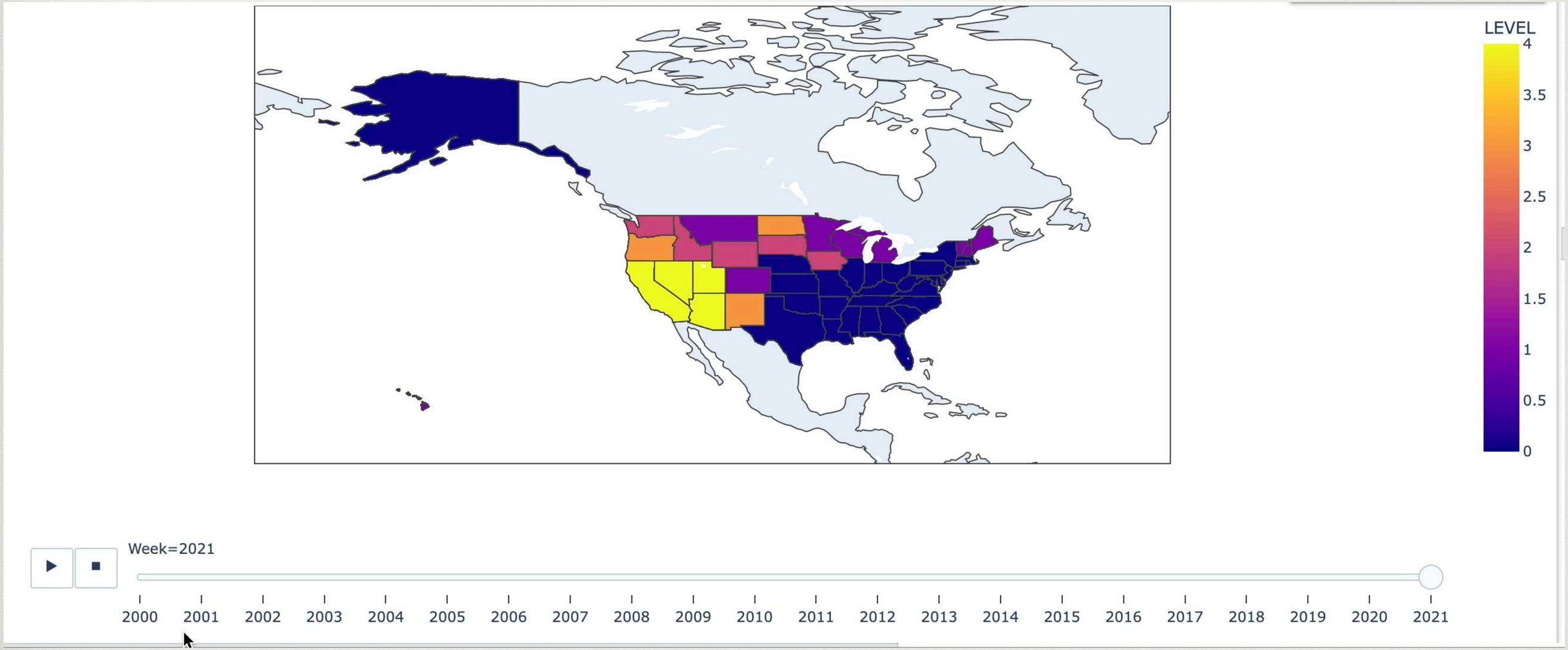


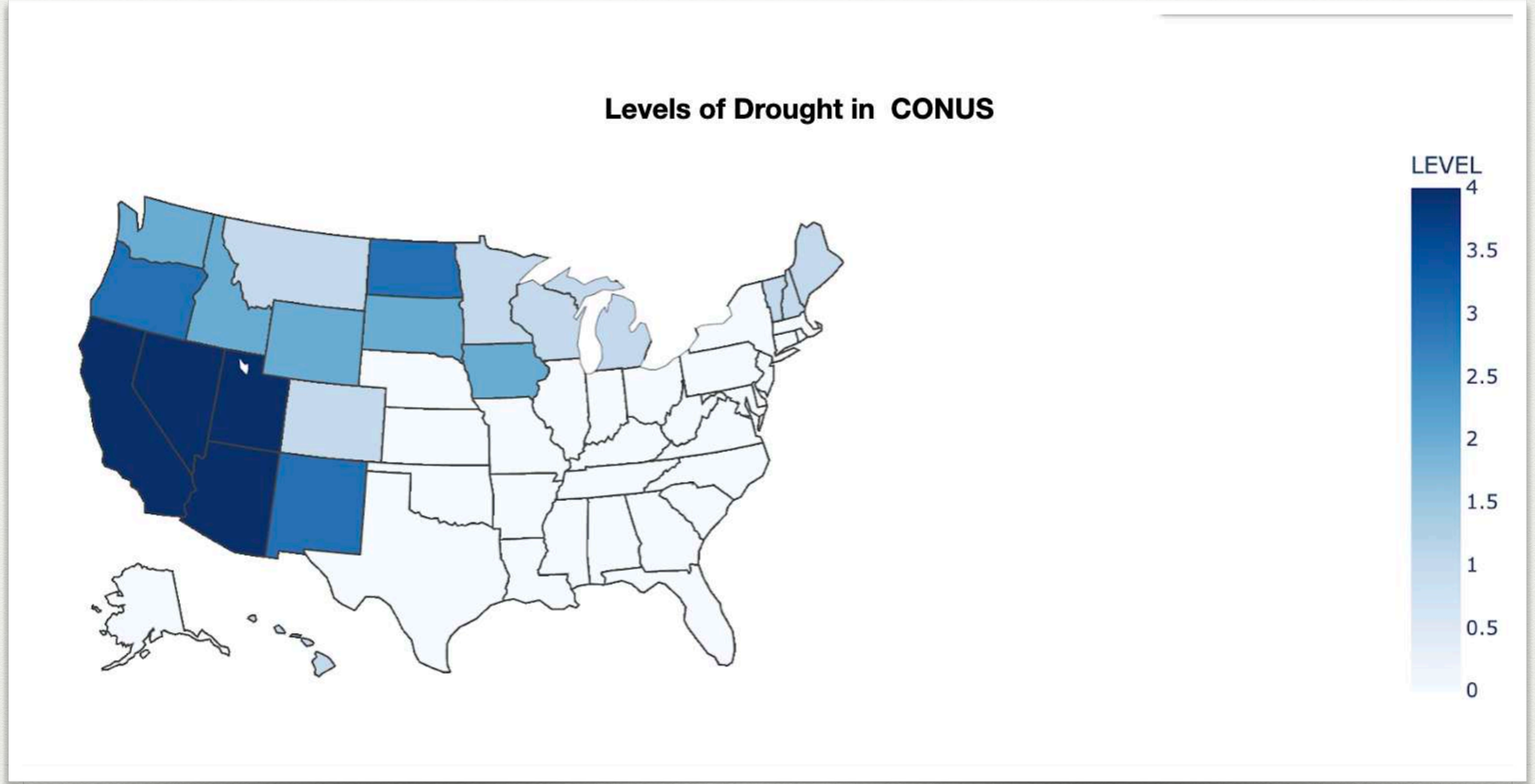
Flow Chart Representation

WORKING OF KNN CLASSIFIER



Animation of Drought Events using Plotly





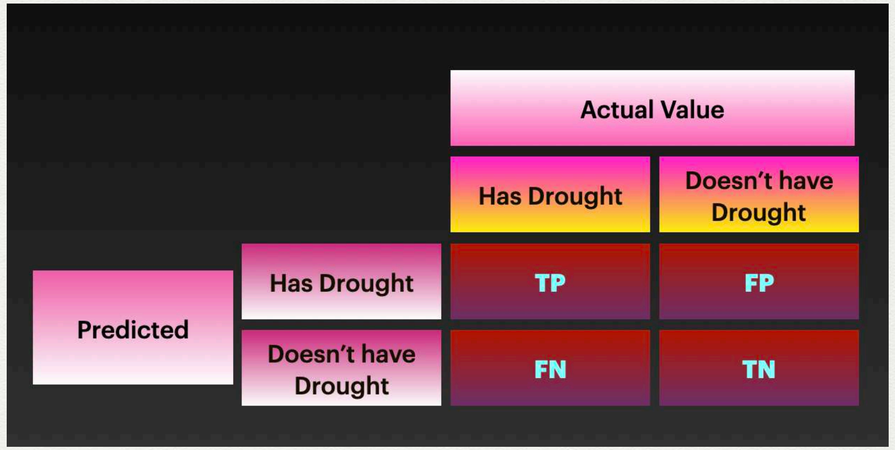
Drought events from 2000-2021

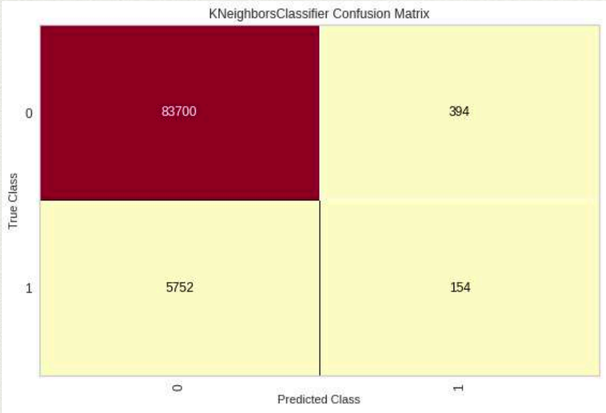
PEFORMANCE METRICS

- Here 13 evaluation metrics are used to quantify the prediction performance of each model namely : **Accuracy, Precision, Recall, F1 score , Confusion Matrix, Classification report, Log Loss, Roc_Auc_score (Area under the curve)(Receiver Operating Characteristics), ROC AUC curve, Average Precision Score, Discrimination Threshold, Precision recall curve, ROC curve.**

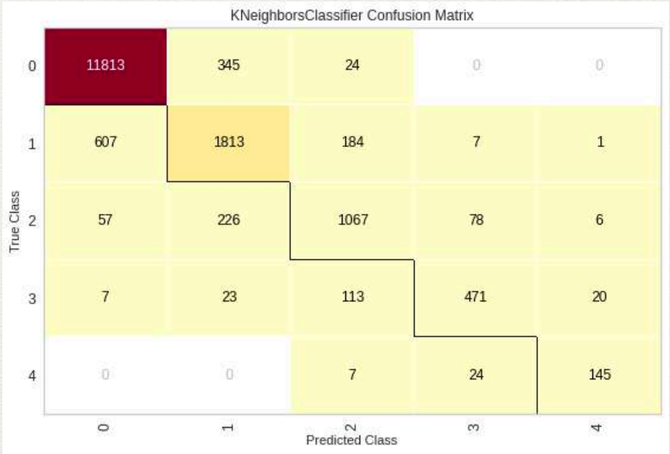
METRICS	FORMULAS
Accuracy	**Accuracy = (TP + TN) / (TP + FP + FN +TN)**
Precision	**Precision = TP / (TP + FP)**
Recall	**Recall = TP / (TP + FN)**
F1 score	**F1 score = (Precision x Recall) / (Precision + Recall)**

Confusion matrix





Dataset - 1
Binary classification



Dataset - 2
Multi-class classification

Deep Learning

- ***Keras***

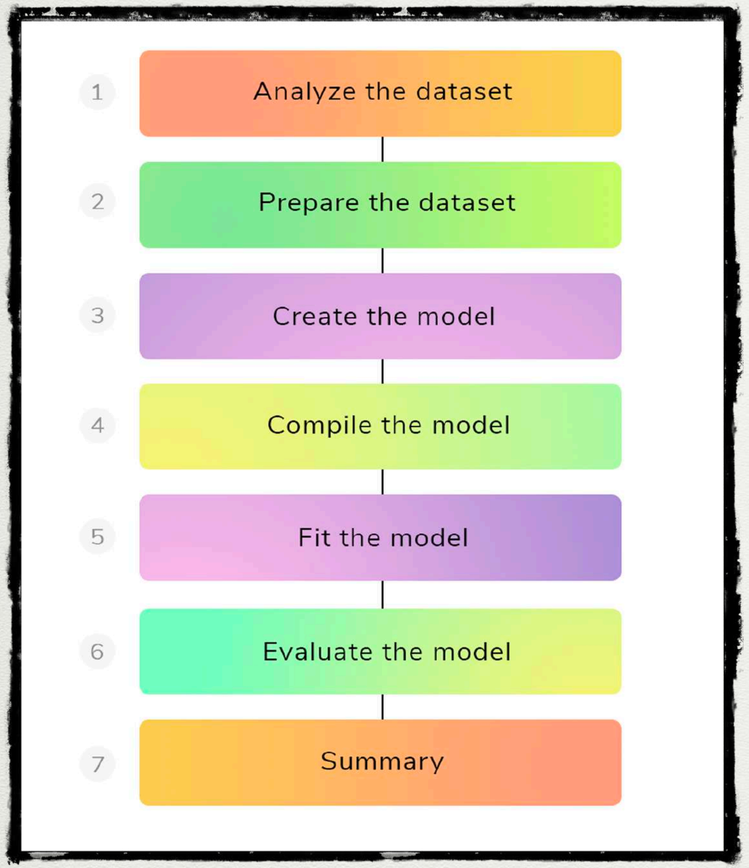
❋ **It is an Open Source Neural Network library that runs on top of Theano or Tensorflow.**

❋ **It is designed to be fast and easy for the user to use.**

❋ **It is a useful library to construct any deep learning algorithm of whatever choice we want.**

- ***Tensor Flow***

❋ **TensorFlow is an open-source platform for machine learning and a symbolic math library that is used for machine learning applications.**



Keras Flow Chart for Deep Learning

```
..
# define the keras model
model = Sequential()
model.add(Dense(12, input_dim=20, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
...
```

```
# compile the keras model
model.compile(loss='binary_crossentropy',      optimizer='adam',
metrics=['accuracy'])
...
```

```
# fit the keras model on the dataset
hist=model.fit(x,y,epochs=10,batch_size=10,validation_data=(data_test,
target_test))
```

```
# evaluate the keras model
_,accuracy = model.evaluate(x,y)
print('Accuracy: %.2f' % (accuracy*100))
```

Epoch is the term used to denote the number of iterations involved during the training process of a neural network.

```
Epoch 1/10
30000/30000 [==============================] - 117s 2ms/step - loss:
0.9027 - accuracy: 0.8767 - val_loss: 1.0055 - val_accuracy: 0.9365
Epoch 2/10
30000/30000 [==============================] - 59s 2ms/step - loss:
0.3424 - accuracy: 0.9310 - val_loss: 0.3780 - val_accuracy: 0.9365
Epoch 3/10
30000/30000 [==============================] - 59s 2ms/step - loss:
0.3222 - accuracy: 0.9353 - val_loss: 0.3869 - val_accuracy: 0.9365
Epoch 4/10
30000/30000 [==============================] - 58s 2ms/step - loss:
0.3199 - accuracy: 0.9353 - val_loss: 0.4047 - val_accuracy: 0.9365
Epoch 5/10
30000/30000 [==============================] - 59s 2ms/step - loss:
0.3185 - accuracy: 0.9353 - val_loss: 0.4506 - val_accuracy: 0.9365
Epoch 6/10
30000/30000 [==============================] - 58s 2ms/step - loss:
0.3180 - accuracy: 0.9353 - val_loss: 0.5622 - val_accuracy: 0.9365
Epoch 7/10
30000/30000 [==============================] - 60s 2ms/step - loss:
0.3173 - accuracy: 0.9353 - val_loss: 0.4856 - val_accuracy: 0.9365
Epoch 8/10
30000/30000 [==============================] - 60s 2ms/step - loss:
0.3166 - accuracy: 0.9353 - val_loss: 0.5989 - val_accuracy: 0.9365
Epoch 9/10
30000/30000 [==============================] - 59s 2ms/step - loss:
0.3162 - accuracy: 0.9353 - val_loss: 0.4720 - val_accuracy: 0.9365
Epoch 10/10
30000/30000 [==============================] - 60s 2ms/step - loss:
0.3160 - accuracy: 0.9353 - val_loss: 0.4482 - val_accuracy: 0.9365
Ellipsis
```

Dataset- 1

```
Epoch 1/5
9087/9087 [==============================] - 25s 3ms/step - loss:
29.9779 - accuracy: 0.5468 - val_loss: 0.6948 - val_accuracy: 0.7104
Epoch 2/5
9087/9087 [==============================] - 28s 3ms/step - loss:
0.7140 - accuracy: 0.6624 - val_loss: 0.6942 - val_accuracy: 0.7102
Epoch 3/5
9087/9087 [==============================] - 28s 3ms/step - loss:
0.6935 - accuracy: 0.6710 - val_loss: 0.6986 - val_accuracy: 0.7104
Epoch 4/5
9087/9087 [==============================] - 28s 3ms/step - loss:
0.6927 - accuracy: 0.7006 - val_loss: 0.6951 - val_accuracy: 0.7104
Epoch 5/5
9087/9087 [==============================] - 27s 3ms/step - loss:
0.6922 - accuracy: 0.7100 - val_loss: 0.6945 - val_accuracy: 0.7104
```

Dataset- 2

Predicted Model

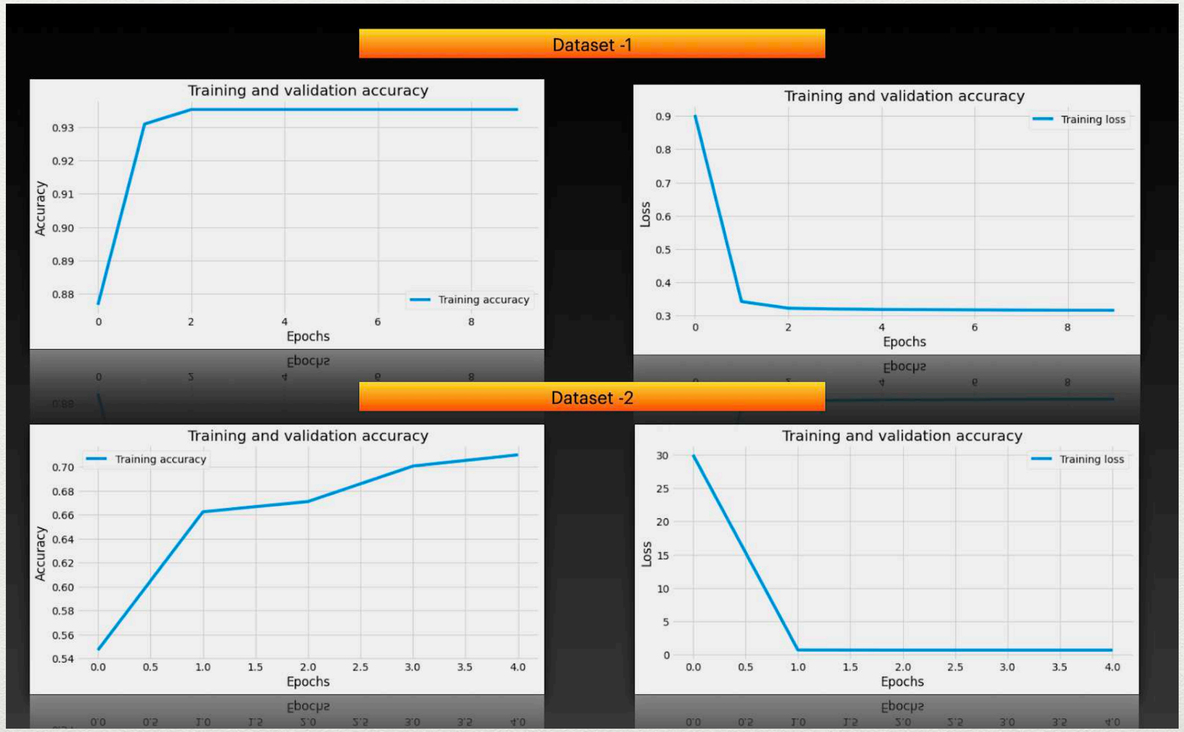
```
9375/9375 [==============================] - 35s 4ms/step - loss: 0.3148 -
accuracy: 0.9353
Accuracy: 93.53
```

Dataset- 1

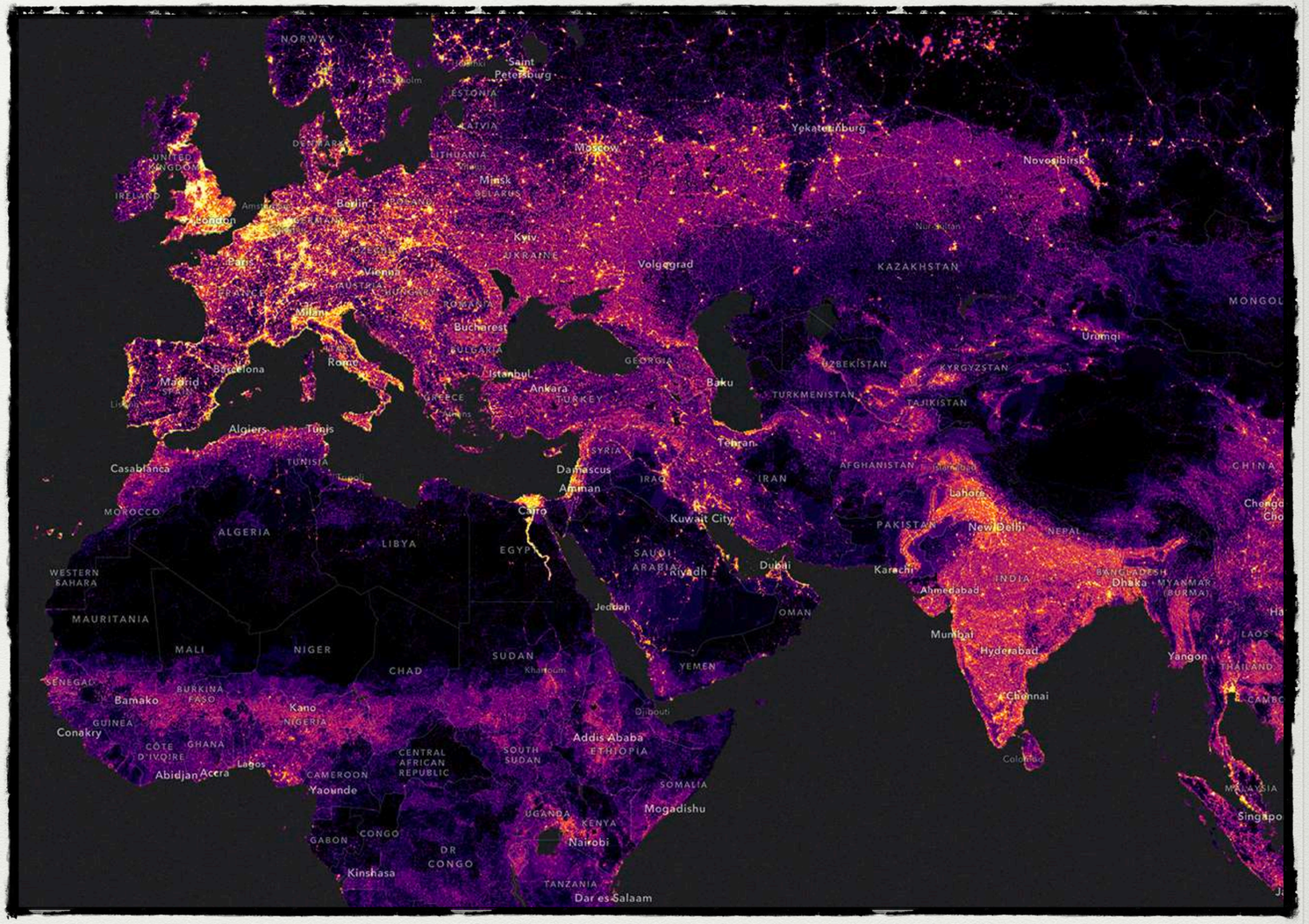
```
1420/1420 [==============================] - 2s 2ms/step - loss: 0.6924 -
accuracy: 0.7127
accuracy: 71.27
```

Dataset- 2

Output of DL



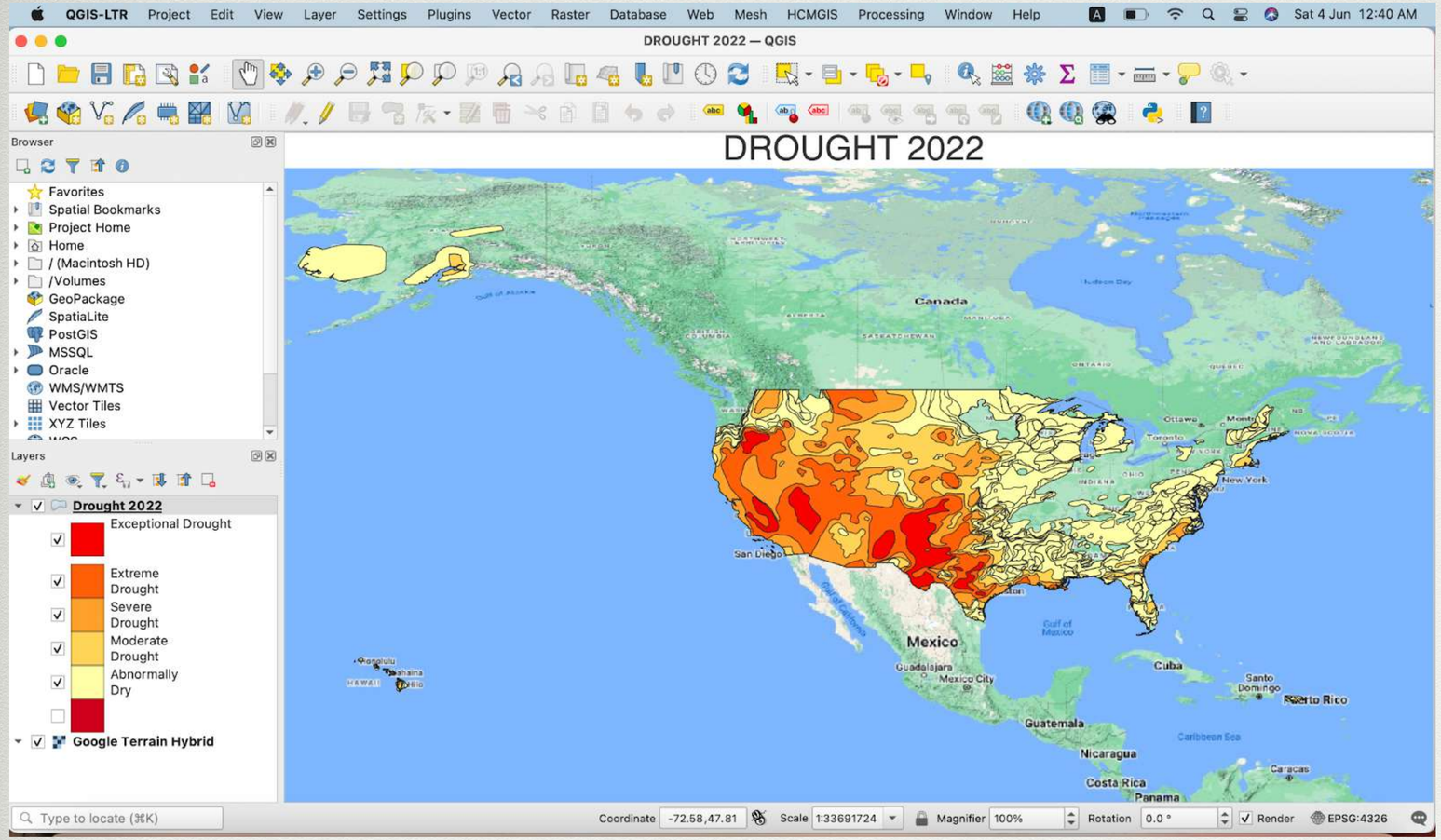
GIS visualisation



- **Geospatial data**, or spatial data is information that has a geographic aspect to it.

- **QGIS** is a free and open-source cross-platform desktop geographic information system (GIS) application that supports viewing, editing, printing, and analysis of geospatial data

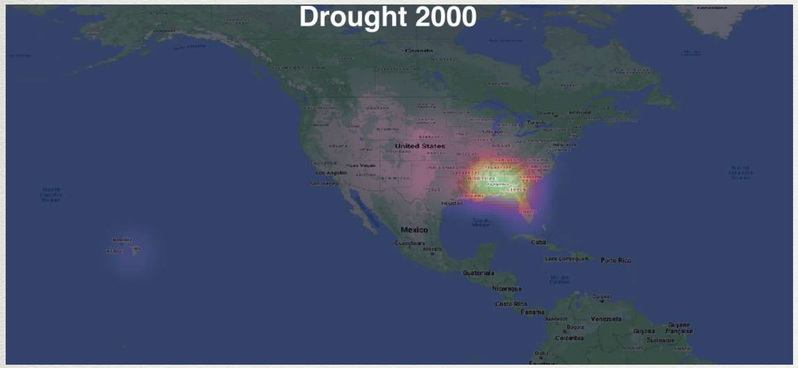
- MOLUSCE plugins (**Modules for Land use Change Evaluation**).

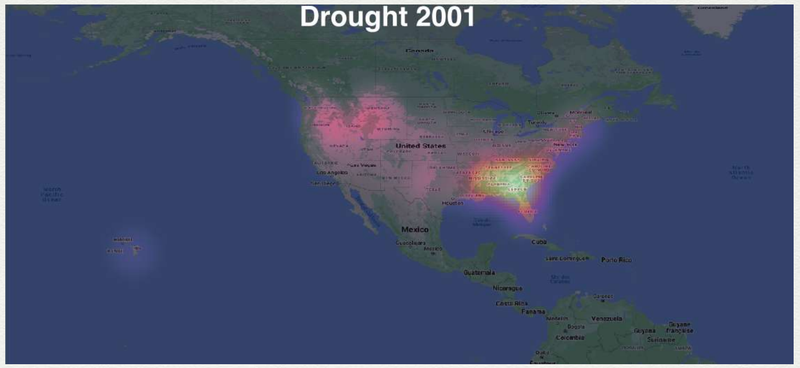


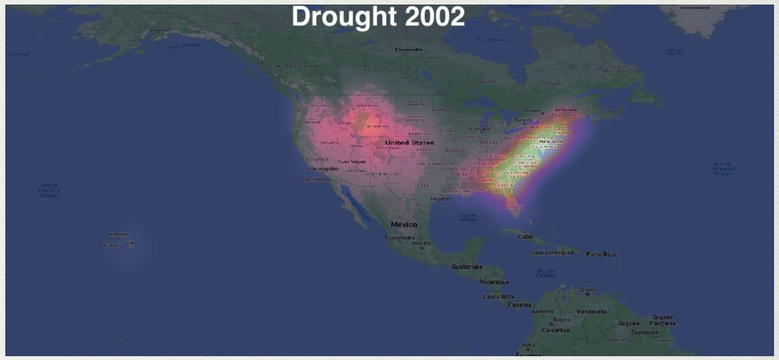
Drought in 2022

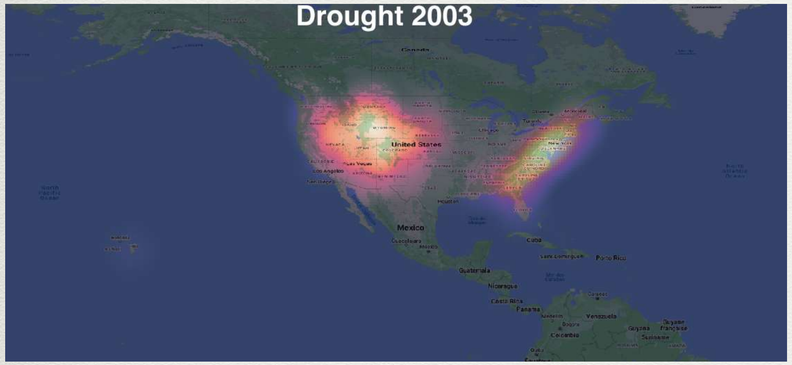
Category of drought types	Threshold value	Description
D0	Less than 0	Abnormally dry
D1	0 to 1	Moderate drought
D2	1 to 2	Severe drought
D3	2 to 3	Extreme drought
D4	3 to 4 or more	Exceptional drought

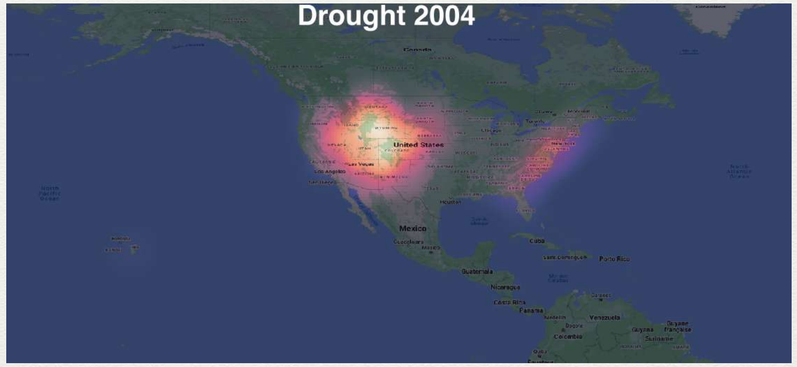
QGIS VISUALISATION

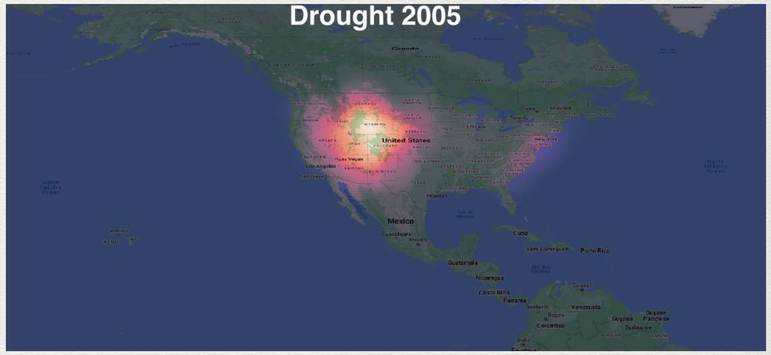


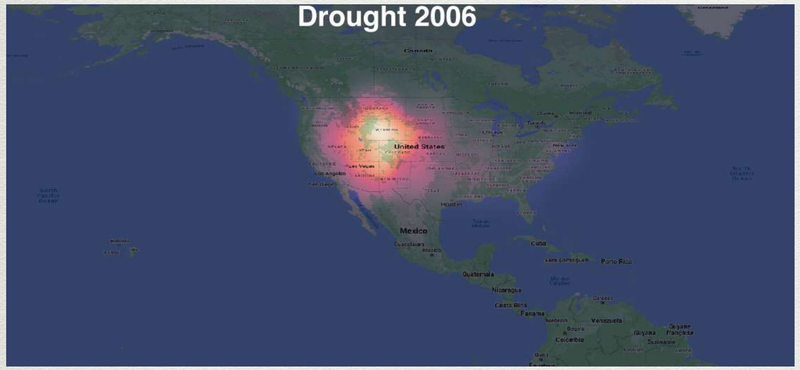


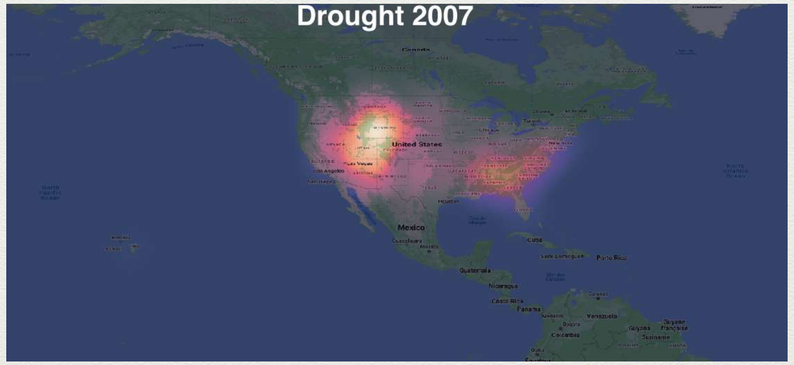


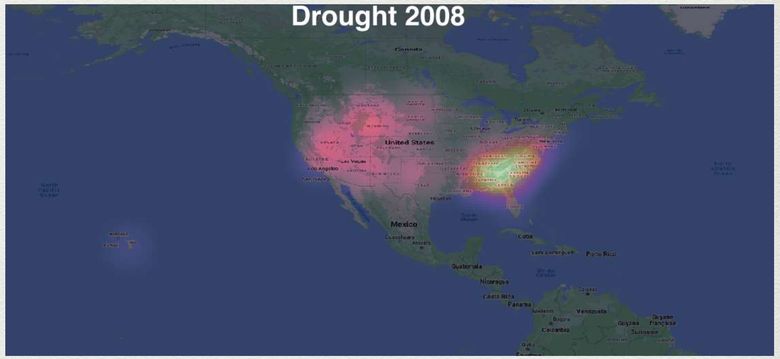




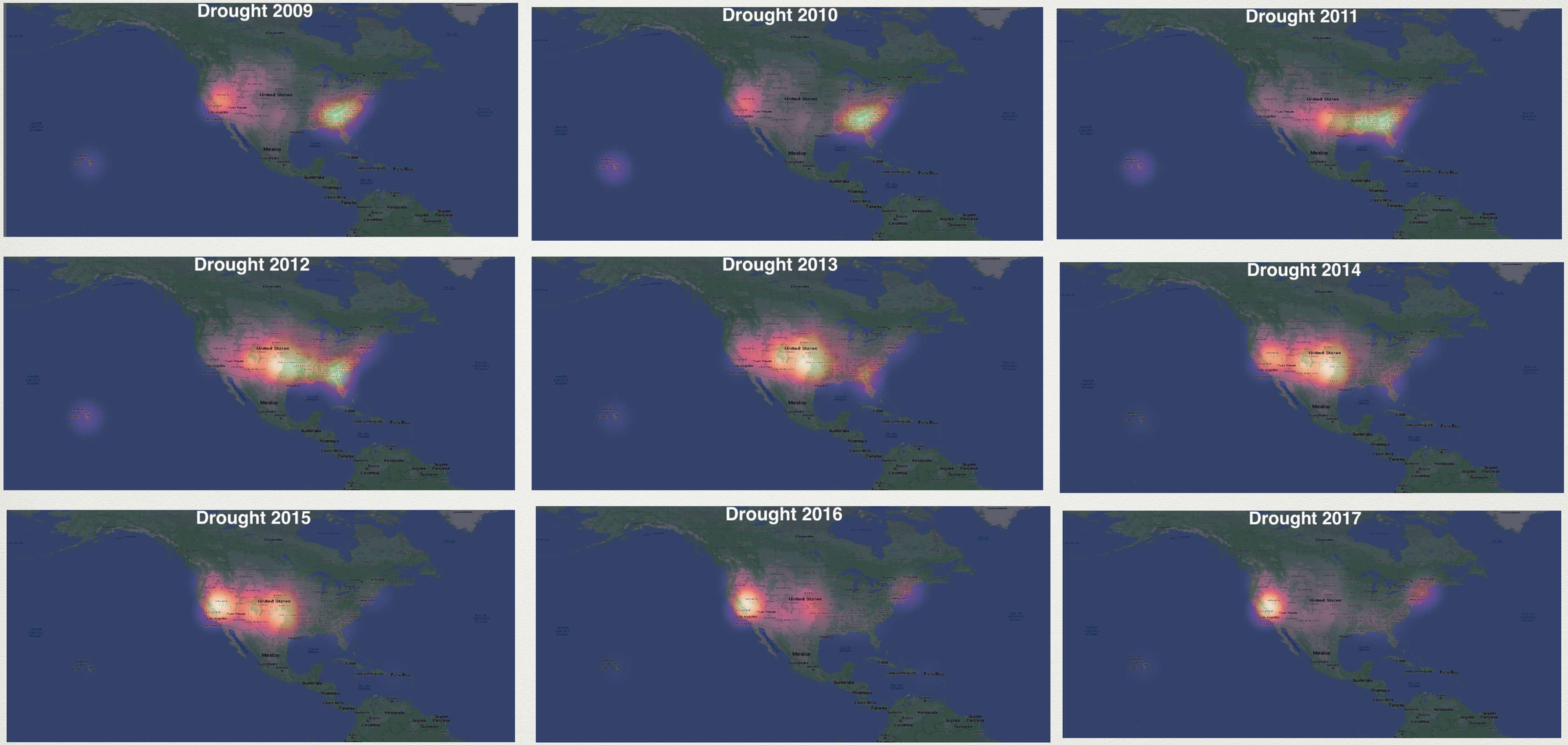




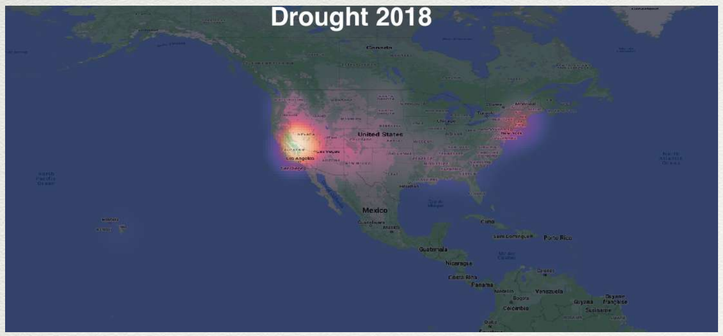


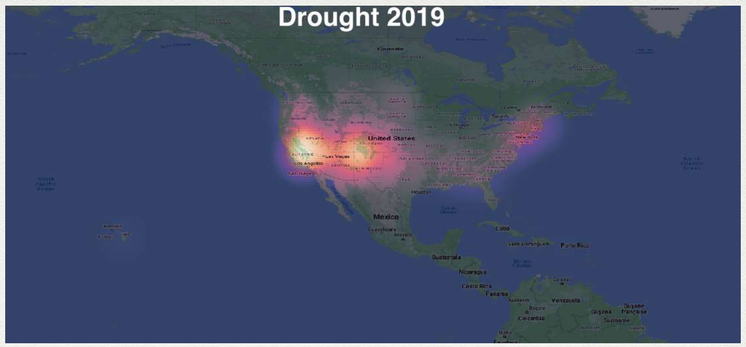


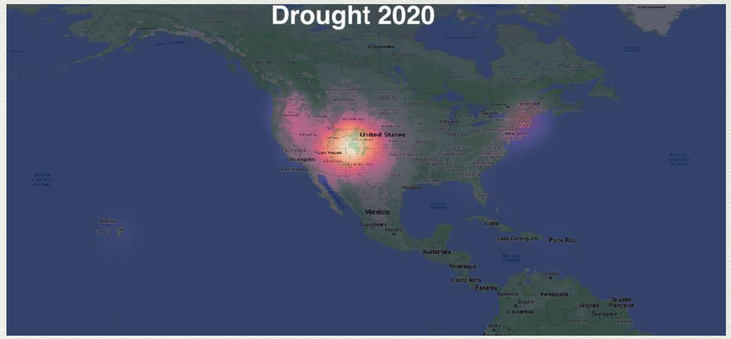
Drought 2000- 2008

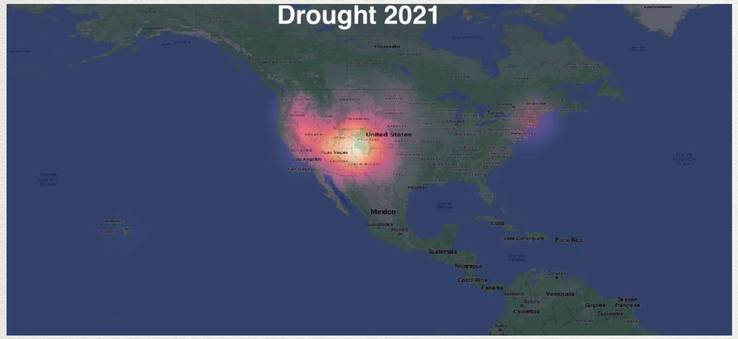


Drought 2009 - 2017









Drought 2018 - 2025

WEB APPLICATION

STREAMLIT

- Stream-lit is an open-source app framework for Machine Learning.
- It helps us create web apps for data science and machine learning in a short time.

Streamlit

Drought Prediction Using Geo-Spatial data

US Drought Prediction using Metrological indices ML App

fips	Country code of US
1001	
date	Certain dates
0	
PRECTOT	Precipitation (mm day 1)
0.22	
PS	Surface pressure (kpa)
100.51	
QV2M	Specific humidity at 2 meters (g/kg)
9.65	
T2M	Temperature at 2 meters (C)
14.74	
T2MDEW	Dew/ Frost point at 2 meters (C)
13.51	
T2MWET	Wet bulb Temperature at 2 meters (C)
20.96	
T2MMAX	Maximum Temperature at 2 meters (C)
11.46	
T2MMIN	Minimum Temperature at 2 meters (C)
12.23	

T2MRANGE	Temperature range at 2 meters (C)
21.21	
TS	Earth Skin temperature (C)
14.65	
WS10M	Wind speed at 10 meters (m/s)
2.20	
WS10MMAX	Maximum wind speed at 10 meters (m/s)
2.94	
WS10MMIN	Minimum wind speed at 10 meters (m/s)
1.49	
WS10MRANGE	Wind speed Range at 10 meters (m/s)
1.46	
WS50M	Maximum wind speed at 50 meters (m/s)
4.85	
WS50MMAX	Maximum wind speed at 50 meters (m/s)
6.04	
WS50MMIN	Minimum wind speed at 50 meters (m/s)
3.23	
WS50MRANGE	Wind speed Range at 50 meters (m/s)
2.81	

Predict

The output is [0]

CONCLUSION

- In this study the real world dataset of US spanning from 2000- 2021 is collected.
- The results show that the Machine Learning model can achieve the best performance than the Deep Learning model used in this study simultaneously.
- Among Machine Learning models, XGBoost and NB are relatively better and among deep learning models, Neural Networks are relatively better.

FUTURE ENHANCEMENT

- The *first direction* is the combination of effective **Feature Selection Algorithms** with machine learning models.

- *Another direction* is the use of an oversampling approach that creates synthetic minority class samples and helps to solve the imbalanced problem, e.g., **Synthetic Minority Oversampling Technique (SMOTE)**.

God has cared for these trees, saved them from drought disease, avalanches and thousand tempests and floods. But he cannot save them from fools.

REFERENCE

[1] Y. Tian, Y.-P. Xu, G. Wang, Agricultural drought prediction using climate indices based on support vector regression in xiangjiang river basin, Science of the Total Environment 622 (2018) 710–720.

[2] A. Dikshit, B. Pradhan, A. Huete, An improved spei drought forecasting approach using the long short-term memory neural network, Journal of environmental management 283 (2021) 111979.

[3] A. D. Mehr, Drought classification using gradient boosting decision tree, Acta Geophysica (2021) 1–10.

[4] A. Mokhtar, M. Jalali, H. He, N. Al-Ansari, A. Elbeltagi, K. Alsafadi, H. G. Abdo, S. S. Sammen, Y. Gyasi-Agyei, J. Rodrigo-Comino, Estimation of spei meteorological drought using machine learning algorithms, IEEE Access 9 (2021) 65503–65523.

[5] A. Dikshit, B. Pradhan, A. M. Alamri, Long lead time drought forecast- ing using lagged climate variables and a stacked long short-term memory model, Science of The Total Environment 755 (2021) 142638.

THANK YOU