Group Members	**Tehreem Fatima (Bscs201926), Sania Bibi (BSCS201967), Seemi Shaheen (BSCS201965)**
Project	**Machine Learning**
Submitted to	**DR. Malik Jahan**
Date	**18 May 2022**

Table of Contents

- **Comprehension:**

  According to the NASA study, the worst drought happens in the eastern Mediterranean that begin in 1998.In this research the drought prediction is resolved using the SPI (Standardize Precipitation index), that using the four-machine learning algorithm. Although there are many machines learning algorithm, but in this article the author uses only for algorithm. That are RF, RT, RSS, and BG. The author uses these algorithms for the prediction of the drought. With the prediction, these algorithms are also compared based on the Correlation, RMSE. After comparing we came to know that the "Bagging Algorithm" is the best among all of four. Because BG algorithm has high Corelation and low RMSE.

- **Overview:**

  Natural phenomenon that effects all ecosystem with the lack of water supply called drought. In this modern era, we can easily make predictions on drought events and then control it. So, for this purpose, we can use some machine learning algorithms. The comparison of these four algorithms is done through RMSE and Correlation Coefficient. The algorithm with low RMSE and high Correlation Coefficient are more prior than others.

- **SPI method for Drought Prediction:**

So Why SPI method is used? because it is dimensionless index, where the negative value represents the drought and the positive value represent the wet condition.

Table 3
SPI categories for drought classification.

Drought group	SPI values
Normal	0 <
Mild drought	0 to −0.5
Moderate drought	−0.5 to −0.84
Severe drought	−0.84 to −1.28
Extreme drought	−1.28 to −1.65
Very extreme drought	< −1.65

- **Menn Kendal Test:**

In this research the Mann Kendall test is used for the drought prediction. This test is used to determine that whether the time series is trending lower or higher monotonically.

- **Machine Learning Methods:**

In the drought prediction, all the modelling was performed using the WEKA Software. This program includes the GUI. We, use this interface because it does not need much timing in the implementation. The parameter used especially classifiers such as M5P, Gaussian process regression (GPR), SMO-Support Vector Machine (SMO-SVM), REPTree and Additive regression (AR) are implemented. These above parameters are considered the best parameter for the modelling process. REPTree is considered is best among all others.

- **Analysis:**

Here, the main point was that we have to analyse the different ML algorithms and methods for drought prediction. We have studied these algorithms and find out some points according to the performance of these algorithms and methods.

- **Four ML Algorithms:**

Bagging Algorithm is used to deal with bias-variance trade-offs and reduces the variance of a prediction model. **Random Subspace** attempts to reduce the correlation between estimators in an ensemble by training them on random samples of features instead of the entire feature set. **Random Forest** is a classifier that contains a few decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. **Random Tree** was established using a conventional method and then refined using decision trees on an arbitrary column subdivision.

- **Findings and Claims:**

In this research, we have been observed that they have used SPI method for drought prediction. But this method already had a few disadvantages. This was deal with rainfall data but not with some external factors like temperature, radiations etc. Another method we have found for drought prediction is SPEI which consider all these necessary factors.

- **Model Performance:**

The performance of ML algorithm is measure through SPI by comparing it's estimated and observed values. Statistically, these four algorithms are implemented using the indicators which are MAE, RMSE, RAE, RRSE and Correlation Coefficient. The algorithm with the best performance after measuring it through indicators can be implemented in a different station

for validation. After calculating the above indicators that discuss above, we predict that the highest drought trend increases in the JB, that has the range between the 0.013 per decade (SPI-3) and 0.032 per decade (SPI-24). In training and testing stage, we came to know that the JB was the highest climate station (950 mm) with the average rainfall of 1561.48 mm (1960-1990). So, after calculating the drought indicators in different time scale the SPI-13 was chosen for the agricultural drought, SPI-12 was chosen for the hydrological drought. So, these algorithms will categorize as RT > RF > BG> RSS for all the stations. Here, 70% of the dataset is used for the training, and 30% is used for testing.

- **Synthesis:**
- **Weak Points:**
  - In this research, SPI only uses the rainy data and did not look upon the other climatic factors like temperature, radiation etc. It was not able to predict the start and end time of drought.
  - Other weak point that I have noticed was that, although BG algorithm had a good performance than others, but they have only discussed it. According to my opinion, all other three algorithms should have been discussed.
- **Strong Points:**
  - It was easy for us to understand the article because the methods such as correlation and RMSE are used for drought prediction.
  - Due to simplicity of the SPI method, help us easily to understand the data that is used in this article.
- **References:**

https://www.simplilearn.com/tutorials/machine-learning-tutorial/bagging-in-machinelearning#:~:text=Bagging%2C%20also%20known%20as%20Bootstrap,variance%20of%20a%20prediction%20model.

https://en.wikipedia.org/wiki/Random_subspace_method

https://www.javatpoint.com/machine-learning-random-forest-algorithm