International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I1P211

Volume 71 Issue 1, 124-133, January 2023

Original Article

Model of Neural Networks: Probabilistic Prediction of Floods in Banana Agricultural Field

Holiver Trujillo Moreno[1], Renzon Javier Gómez Márquez[2], Miguel Angel Cano Lengua[3], Laberiano Andrade-Arenas[4]

[1,2,3,4] *Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú*
[3]*Universidad Nacional Mayor de San Marcos*

[3]*Corresponding Author : mcanol@unmsm.edu.pe*

Received: 17 July 2022 Revised: 07 January 2022 Accepted: 17 January 2023 Published: 24 January 2023

Abstract *- During the latest events caused by climate change and the current of the child, Peru has been affected by these natural disasters, such as the flood, which directly affect the Peruvian economy and especially the department of Piura. To prevent and mitigate the problems that affect the department of Piura with respect to flooding, the development of a probabilistic system has been proposed with the use of machine learning that will allow us to prevent possible climatic changes and avoid material damage to the area based on predictions. Likewise, the data found in the repository of the free data web page provided by SENAMHI will be extracted to be reused internally and can contribute to the development of the application through neural networks that will facilitate the use of the data. Given this, it has been decided to use the data scientific method, which consists of 10 phases that allow us to identify the main points that contribute to the model of the proposal. This allows us to carry out the necessary validations to make the proposed system feasible. To obtain, as a result, a model that can predict and give warning about the threat of flooding based on the weather behavior of the area. In addition, it is concluded that the prediction models with the help of artificial intelligence tools have better efficiency in terms of forecasts.*

Keywords *- Climate change, Scientific data, Flood, Machine learning, Predictions, Neural networks.*

1. Introduction

It has been shown in the world, according to studies in recent years, that there is a link between precipitation, climatic signals, and large discharges from hydrographic basins [1]. Countries like Mexico are affected by major droughts, a product of the El Niño phenomenon, which comes with great power in regions such as Veracruz, Tabasco, and Yucatan. On the other hand, in dry regions of the world, the climate often exhibits nonlinear behavior. Flood forecasting is extremely challenging, such as in Iran, which is in Western Asia and shares a border with the Caspian Sea. The Persian and Oman Gulfs use sea level pressure (SLP) and indicators of sea surface temperature (SST), which will serve as variables to consider achieving the hydro-meteorological prediction [2]. One of the hardest years for Peru was during 2017, when the root cause was the El Niño phenomenon, with devastating power that affected many departments on the north coast of Peru [3]; one of them was the department of Piura, which obtained around 1.7 million people affected, in view of this a large amount of humanitarian aid has been granted to the affected population and in general to the department of Piura. The damage suffered by the department of Piura was of great magnitude since it was the most affected by the weather phenomenon; the rains that occurred flooded the northern capital causing losses [4], which caused shortages in agriculture, greatly affecting Measured at the national and export levels, the decrease in agricultural production caused a high price of necessities and a decrease in the quality of life of the inhabitants. In recent years, multiple ways of predicting the behavior of the climate have been proposed. Several of these studies were designed and applied through hydrological modelling [5], which is based on the study of the basins near the vulnerable place; these will allow knowing a warning system for possible floods to prevent any natural disaster [6]

The objective is to develop a model of neural networks that simulate a probabilistic prediction of flood warnings in banana agricultural fields in the Piura region located on the north coast of Peru.

The structure of the paper is as follows: Section II explains the review of the literature on the positions of different authors, Section III explains the methodology that has been carried out, Section IV explains the results through the findings, Section V the discussions regarding the review of the literature, the Section VI the conclusions and future work, as suggestions.

2. Literature Review

In this section are going to detail, as a first point, the basic theory to have a better understanding of the issues to be able to move on to the next point, which is the related work that will talk about the studies carried out by other researchers in relation to weather forecasts with the use of machine learning.

2.1. Theoretical Basis

To achieve a deep understanding of the research topics, one must carry out a bibliographic review of the general concepts that must support the research on the probabilistic flood alarm system. The concepts that will use: Machine learning and Neural Networks in connection with the weather, achieving a good fit [7].

For the identification and extraction of data or values that are used for the development of this application, the following points have been taken into account: Extraction: During this stage, the data scientist model consists of extracting the data from various sources in the case from the SENAHMI page, and being able to analyze the data and be able to complement it with other sources that allow us to obtain more information for the system. These data can be classified according to their size as Small or Big Data.

2.1.1. Cleaning

During this stage, the data that is not aligned with the need of the system will not be considered, and the amounts of information that is required are also prepared so that they can go through the necessary process to obtain quality data

2.1.2. Processing

During this stage, various statistical methods are used, for which the data scientist is responsible for processing the volumes of data and extracting the necessary information. Through mathematical analysis, predictive methods, or clustering (which consists of the grouping of elements that have similar characteristics) [8], during this process, the possible data will be interpreted and thus give forecasts about the future or seek a relationship between some of the variables.

2.1.3. Visualization

Finally, the results that were obtained will be presented so that the information provided is understandable and that anyone, whether they are experts in data science. For this, they will be accompanied by graphs that allow us a better interpretation and power to visualize better the values obtained.

The data will be obtained from the free data platform provided by SENAMHI, ready for download. This data will be transferred to an Excel format to be compatible, and it's easier to put it through an ETL (data extraction, transformation, and cleansing) process [9]. Of that way will get the data ready for use in a script format. After the ETL process, the data is prepared to connect with the [10] programming IDE, in this case, PYTHON. First, the connection is made based on data by adding the necessary libraries and linking them with the variables in charge of containing them. Python offers us the ease of implementing a library that takes care of the process of machine learning, specifically neural networks, which begins to analyze the data, and, in the execution, it throws us possible scenarios, which come to be the predictions, with its respective statistical analysis used. To make the software look more attractive to the user, a web interface will be created with the ability to display this data dynamically and intuitively. The goal is for the system to be free and free for the target audience. Because it is free, the system will not have user access. User, which means it will directly send us to the main menu, where you can view predictions directly.

Machine Learning (ML): Machine learning has been called a subfield of computer science and, at the same time, would be a branch of the IoT, with the purpose of developing learning techniques for computers [11]. That means that, with the use of data, variables, experiences, behavior patterns, models and hypotheses, computers can obtain the ability to solve problems, giving machines the ability to learn based on a trend or behavior.

Neural Networks (ANN): It is an analytical and mathematical technique inspired by the neurons of the nervous system of animals; it comes from the motivation of artificial intelligence and neuroscience, starting in the 40s [12]. Neural network models can accept a larger amount of input data in a weighted manner, usually creating nonlinear algebraic functions or decision trees to reach the result that would be the output data [13].

In the same way as biological neurons, this is one of the fascinations for neural networks is their ability to learn and adjust to the conditions of their input layers and their output layer; I mean that, through an algorithm of learning, the neural networks adjust to the architecture and its parameters in a way that they can minimize the percentage of the error that varies the degree of adjustment of the data [14].

It means that as more data is obtained for the model, the margin of error is smaller, from which more accurate predictions are obtained, and, at the same time, the complexity of the network grows.

The algorithm has 3 fundamental phases, starting with the input phase, which must describe all possible variables influencing the result. The second point is learning, which consists of multiple simulations of scenarios that can occur with the input data. As the last phase has the output phase, which basically consists of the most repeated and coincident

tendency or behavior in the learning phase. This result is the possible scenario to happen based on the variables, also called prediction.

2.2. Related Work

The study carried out allowed knowing the importance of machine learning [15] when using it as a key tool to make weather forecasts. The articles studied have various intelligent methodologies that help us understand the logic of the problem based on empirical data that demonstrate the learning that the system can capture with respect to each simulated scenario.

A group of Canadian researchers affirm that the temperature variable greatly influences climate prediction and that an increasing trend can be seen with respect to reducing the margin of error. With the application of deep learning, they achieved 96% success in predicting floods with the help of the Canadian disaster database, which was linked to data from climate change indices in a region of the same country, such as part of a [16] training and testing phase. What is shown in this research is the application of IoT technologies helps to reduce the gap between results and events; this means that an optimal and reliable result is achieved every time.

The Iranians also have good results based on percentages on the prediction and use their method of neural networks based on wavelets, taking as an example the data of the institute of geophysics and the Polish academy of sciences. Likewise, various linear regression methods are used with the help of Deep Learning and Machine Learning to reduce the margin of error that occurs due to anomalies and allow early simulation of climate change [17].

The study of neural networks has been applied in multiple fields, facilitating their better understanding and assimilation to them. Thanks to processing information through mathematical functions, these perform multiple possibilities that can predict behavior through the previously identified parameters. In this way, it is possible to emulate multiple cases and associate them by identifying which data they have in common. This was assimilated to what is known as an experience where the data that is already available serves as a basis for analysing future cases that have something in common [6].

The use of neural networks to make predictions has a wide field, one of these great contributions is evidenced in the prediction of frosts [18]; for this, it is necessary to study the field to be treated and identify which model or method is appropriate For the capture of parameters and their use, in the case of frost, a structure was used which was made up of nodes that had the capacity to receive data and convert it into information that can be delivered. This is divided into phases, which begin with the training phase, where the interaction of the data begins to determine the connection between inputs; during the second phase, the data that was designed to provide results were validated to obtain results that validate the entered information.

According to the investigation carried out by Guerrero Ciprian affirms, during this period, the flow of the rivers had their maximum arrival with a flow above the historical records, appearing as an overflow affecting the valleys that are close to the basins and rivers, causing floods that directly affect agriculture in the northern region of Peru. The data obtained is stored in software called ArcGIS to avoid these disasters, for which the data obtained from the National Water Authority (ANA) were used; by means of the geographical coordinates of the department of Piura, the necessary information can be obtained to detect which are the places where the basins and rivers present the greatest flow [19]. On the other hand, the Ministry of Agriculture and Irrigation sent a report indicating the UTM coordinates for the georeferenced location of the places.

During this event, the surrounding populations suffered a great material loss since most of them were dedicated to agriculture. To identify the great impact that this natural phenomenon had, many organizations carried out reviews in all the affected places within which they were established. Recommendations based on experiences in similar phenomena and providing good practices to remain calm before the arrival of this natural phenomenon [20]. To do this, the damage that affected the region was identified, and a record of the affected families and the amount of land that was flooded; this information obtained serves to have a record of how great the magnitude of this phenomenon was, the information obtained will serve to implement a system based on neural networks to predict the possible dates that would affect the region.

The implementation of technology in daily life is a very relevant fact since, through it, daily activities can be facilitated; the use of it in different areas of the lives considerably improves production and helps to have a better lifestyle. Therefore, the implementation of sensors on the banks of the rivers will allow the detection of certain climatic alterations and thus avoid meteorological risks; according to [21], it was proposed to implement them in the streams and rivers of the department of Piura in this way it is obtained data in real-time and determine if this would affect the population, generating an early warning and carrying out an early evacuation. However, this is not the only way to anticipate these risks. A predictive system was also proposed using neural networks; these are carried out using data obtained during the natural phenomenon and thus simulate the possible behavior that rainfall may have in the department; it has come to have a large percentage of effectiveness in this type of data collection having a 98.4% success rate during its testing stage, while during the first

semester of 2016, a 100% effectiveness was obtained. In this way, the implementation of sensors to obtain data and the implementation of neural networks would allow the prediction of the different natural phenomena that the El Niño [22] current may bring.

The study carried out in Asia indicates the development of models based on the data collected to forecast the main drainages in the Nile to avoid floods. The data of these models are based on (MLR, ANN and WT-ANN). This study can have a great impact on the phenomenon that occurred on the coast of Peru by using the records obtained over time can measure the degree of risk in stages and thus be able to establish a similarity that allows identifying the most important and have better information for the prediction of the records [23], for this must consider the highlighted information, where the most relevant is taken, obtaining a better performance at the time of taking the data. These can be classified according to their characteristics or by the criteria granted; the classification will depend on the degree of information you can provide. By granting a higher degree of qualification, predictions can be made more effectively. In such a way, when developing a model that predicts a natural phenomenon's behaviour, it will depend on how much true information can be gathered and the degree of importance and impact that [32] can have.

In conclusion, it is evident that artificial intelligence tools help to achieve a prediction with a high level of precision. However, it was also noted that the lack of data variables makes the prediction's perfection decrease. It is stated very clearly by the Canadian researchers; also the Iranians who used data from other external sources, the lack of technology, such as support for satellite heat sensors, wind sensors or other technologies with the ability to capture metrics, is what prevents us from compliance of the project objective.

3. Materials and Methods

Considering all the parameters that the development of the research asks of us, a methodology is needed that can meet each of the proposed objectives, so based on research recommendations on the same topic related to data management and artificial intelligence, it was concluded that the scientific data method was the most suitable for the research Fig.1.

3.1. Stages of the Methodology

3.1.1. Business Understanding

To understand the project, you must first be very clear about the objective of the research, the characteristics, and the proposed scope of the work. At this stage, a solid structure must be defined to guarantee the expected success of the project.

3.1.2. Analytic Approach

Once the problem is structured in the most optimal way, the scientific data will be able to determine an analytical approach, allowing us to solve the problem. For this, during the course implies the expression of the problem in the environment of the development of statistical techniques and automatic learning. Within the organization, the appropriate way to obtain the desired result can be determined. Given the following example, to determine if the target of a prediction is expected to have an answer that would have the values" yes" or "no", there is an analytical way by which one can define the foundation, test and run for classification modeling [25].

3.1.3. Data Requirements

The data must be selected carefully to obtain a very rich information format. The data must be aligned with the project requirement, which means that any impurity or data not aligned with what is needed in the project would lead to error and failure to implement the probabilistic predictive system [26].

3.1.4. Data Collection

In the first stage of data collection, data is identified and collated from available data sources (structured, unstructured, and semi-structured) of relevant weeks. They often must choose to make additional investments in less accessible pieces of information. It is better to delay funding decisions to have more information about possible data and models. There is the possibility of having gaps at the time of obtaining the data, and this would be possible in view of this; a review of the characteristics and requirements established for the data for its collection should be carried out.

Although the data obtained and the segmentation of it is of great importance, today's high-performance platforms and database analysis capabilities allow working with much larger data sets, if not all available data. By having the data input, the prediction of the models can give an adequate account of unforeseen events, such as the appearance of diseases or failures in the system [27].

The sampling, cleaning, transformation, cleaning and ordering of data is a very important part of any system that works with a relational or non-relational database. However, according to new research and the advancement of new methodologies in the field of data, it concluded the greater the variety of data, the more accurate response can be obtained regarding the predictions because all the data that has influenced the useful life of a process is being included, even if it seems little importance can also influence the result of the system. In this way, events or situations can be represented with incidence frequency categorized as rare, for example, some diseases, also some system failures or natural disasters.

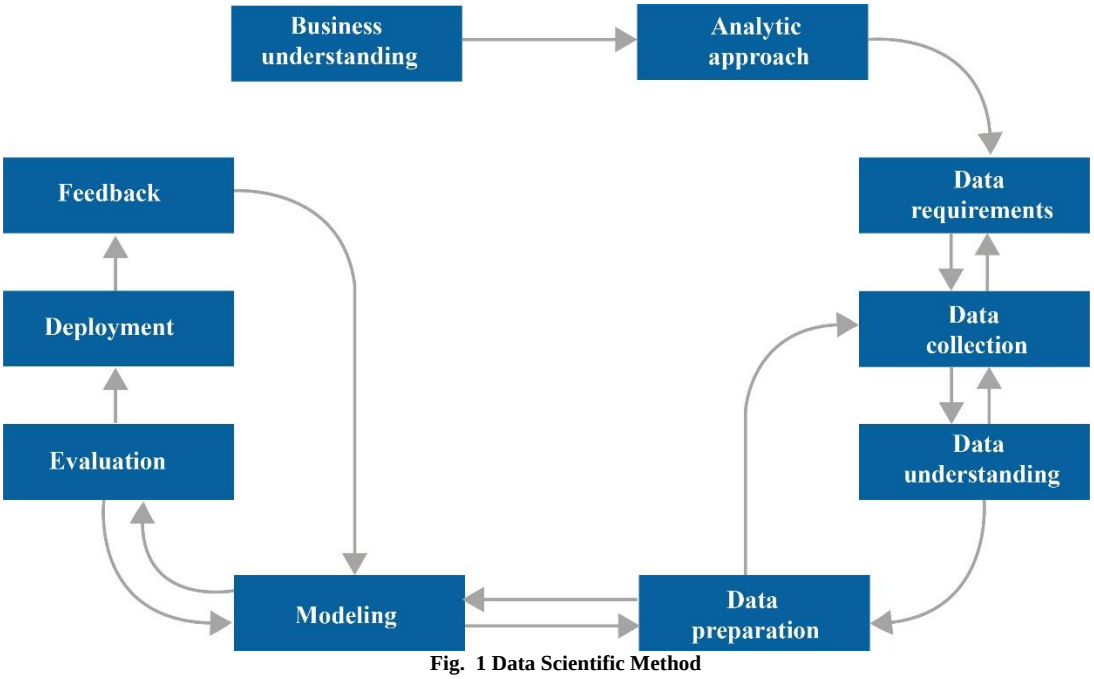

Fig. 1 Data Scientific Method

3.1.5. Data Understanding

Then you must perform a series of statistical tests, dynamic graphs and various strategies to understand the nature of the data and ensure the quality of the information [28].

This means that from a step before the software development, the data must have a correct trend or behavior based on indicators; in this case, as a great example, it can be the trend of precipitation with respect to the last 10 or 50 years.

3.1.6. Data Preparation

This step includes all the dataset construction operations that will be used in the next modeling step. The operations for an ETL process consist of data processing to filter empty, missing, duplicate or invalid data, matching the expected format. Data from other reliable sources can also be combined to align with the project's requirements.

According to the investigations of data scientists, they use a model that they call feature engineering, which consists of the generation of explanatory variables, which are called metrics.

These are structured variables. When text data, such as SENAMHI web records or site observations, are available in different formats, it is sought to balance and standardize the data structure so that it is compatible and helps us enrich the system's reliability based on accurate results.

All data scientists agree that the ETL process can be the most time-consuming step depending on the size and provenance of the data in building projects. Getting to synthesize the entire volume of data can even occupy more than 3/4 of the time of the entire development life of the project. For this reason, the automation of some of the data preparation steps is a proposal that can speed up the process by minimizing the time of custom preparation based on anticipation or preparation in advance. With high-performance batch parallelism systems and analytics capabilities where data is currently stored, DataMart's can help us prepare data in such a simple, fast, and functional way that they can be joined into large Datawarehouse units [29].

3.1.7. Modeling

In this process, the first tests can be carried out with the algorithm to adjust the ideal preparation with the data. Multiple simulations can be performed with the algorithm to find the most suitable model [30].

Being simple modeling, pilot tests can be carried out with few data; in this way, the reliability of the system can be guaranteed with respect to the prediction with real data; it is worth mentioning that the data must be the most important or relevant in the influence of the system.

3.1.8. Evaluation

During model development and prior to project implementation, the data will be evaluated against the model so that you can ensure that it solves your stated problem accurately with an acceptable margin of error. When evaluating the implicit models, several measures will be calculated for the ideal diagnosis and with their results to evaluate coincidences and differences, such as tables and

graphs, which allow analyzing the data. From these tools, the quality of the model and its effectiveness in solving the problem are explained. For most predictive models, the data uses a test set independent of the training set, but it must obey the same type of probability distribution and outcome knowledge. The test suite is used for evaluation.

Sometimes the final form will be applied to a validation set to run last. In addition, the data can identify models, a statistical significance test or go through analysis based on indicators and graphic models as an additional test for its quality; this test can be very important to justify the implementation of the data model to be used [31].

3.1.9. Deployment

Already meeting all the requirements needed for quality data and having a solid idea about the necessary development of the system, it begins to be carefully implemented little by little following good practices and considering the necessary backups to advance without interrupting development without the need to rollback deployment.

You can start with the prediction based on 10 scenarios, then 15 or 20; everything depends on the equipment, also the time with respect to the learning speed; it should be emphasized that the more time you give learning, the more accurate data can be obtained.

3.1.10. Feedback

When obtaining the implemented model's results, feedback on the model's performance and its repercussion for the environment in which it was implemented will be obtained. Therefore, to perform the feedback, it can be applied in percentages in relation to the responses that can be obtained with respect to the amount of humidity in the environment and, in this way, identify the degree of severity that can influence the population.

The extracted data can be analyzed within the feedback to determine which model highlights utility and accuracy. These will determine if they can be less or fully automated during the evaluation of the models and their collection for feedback, to determine their adjustment and the re-adjustment of the model that will allow the acceleration of the processes that will allow the update for the model of so that can get better results.

3.2. Development of the Methodology

3.2.1. Business Understanding

Through machine learning, this system will allow the synthesis of a methodology, giving rise to the ability to improve processes by learning from their own experience and from the data entered to create a probabilistic model. In this way, the process was perfected and facilitated without having been specifically programmed to do so. This way, through climatology data, where the necessary parameters that influence was identified, will help us predict flood risks in the department of Piura, one of the places most affected by weather phenomena such as the El Niño phenomenon (FEN).

The objective of the project is to identify the variables that influence when identifying the behavior of the climate; in this way, the risk of flooding can be prevented, which helps to mitigate the impact that it would have on the region where its economy is based on In national and export trade agriculture, one of the functions that would be had is to predict the behavior of the climate through these already identified variables and, thanks to machine learning, translate this information and process it for the common use of the inhabitants so that they take into account the possible months in which a natural catastrophe may occur and thus take preventive measures so that it does not greatly affect its economy and agriculture, specifically the banana plantation in the department of Piura between the districts of Sullana and Sechura.

3.2.2. Analytic Approach

As a second phase, define the research topic under an analytical approach that allows statistical graphs to be made with the data to ensure the project's feasibility. The analytical approach is used in these cases because the data that will be used for the project requires a statistical analysis with empirical results.

3.2.3. Data Requirements

As a third phase define that the data will be extracted from the free platform offered by the National Service of Meteorology and Hydrology of Peru (SENAMHI) through external data tools, can obtain and import any web content in a table format, creating a set of data belonging to the web page. Which allows us to obtain the information in real-time and to be able to work it in the Excel spreadsheet. These data will be established with the variables aligned to the project.

3.2.4. Data Collection

The fourth phase starts with the data purification obtained through the excel spreadsheet, which will identify and select only the precise information, in this case, will focus on the data of the Department of Piura, specifically in the city of Sechura, which shows us the precise data that will be used at the time of the creation of the probabilistic system.

3.2.5. Data Understanding

In the fifth phase began to carry out a data extraction, treatment, and cleaning process (ETL) to guarantee quality data for the project structure. However, as the first tests, pilots were carried out with only 2 variables to seek the viability and reliability of the project. This process, in general, is unnecessary because there is a greater amount of data input, so a more optimal result can be obtained. However, it did it in this case because it had to accommodate the formats and failures in the repository export.

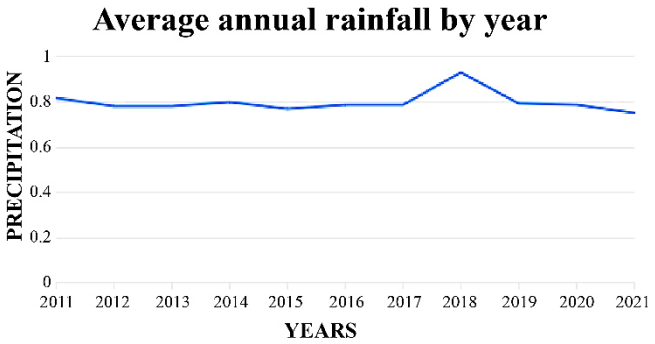

Fig. 2 Precipitation per year

3.2.6. Data Preparation

After cleaning and organizing the data, the sixth phase begins to verify its behavior with a statistical graph indicating the precipitation during the last years. Where it is shown that the year 2018 had higher rainfall Fig.2.

This test helps us identify the reliability of the data in the real scenario because it is recalled that between the last months of 2017 and the beginning of 2018, a natural disaster of flood occurred in the department of Piura in Peru, caused by the phenomenon of the child, as shown by the behavior of precipitation in the graph. In this way, it attests to the quality of the data which are dealing.

4. Results and Discussion

4.1. Modeling

As a seventh phase, it is proposed to declare the TensorFlow and Numpy libraries, respectively, that are being used in the Python language Fig.3.

```
'''Google artificial intelligence library'''
import tensorflow as tf
'''Library to work with arrangements'''
import numpy as np
```

Fig. 3 Added Libraries

Tensor Flow is an exclusive open-source platform for machine learning, which can be installed on a system by importing its web service.

On the other hand, Numpy is a library to work with two-dimensional arrays in python. This way, our variables will be able to store the data together and orderly.

Any IDE that supports such a language can be used for the python language. It is also possible to work with the necessary libraries for machine learning, as the developer likes.

For the pilot test, use the Google platform as a programming tab, then proceed to declare only two variables for the tests, starting with the minimum temperature and the precipitation that are the most influential for the pilot test Fig.4.

```
temperaturemin=np.array([9,12,13,6,7,10],dtype=float)
precipitation=np.array([2,1.4,1.2,2.3,2.1,1.7],dtype=float)
```

Fig. 4 Definition of variables

On the other hand, it could be a test with precipitation and years; in this way, it was possible to predict the average temperature of a given year, obtain a consistent result under verification, and check the system's reliability. Then define the number of layers and inputs that are going to use in the algorithm, in this case, only 1 input and, in the same way, 1 single layer Fig.5.

```
cap=tf.keras.layers.Dense(units=1,input_shape=[1])
model=tf.keras.Sequential([cap])
```

Fig. 5 Learning Speed Function

The relationship between layers and input variables can be related. Finally, when defining the algorithm, the learning speed has been considered a variable, which has been established at 0.1; this will allow us to obtain a better response based on the data entered in Fig.6.

```
model.compile(
  optimizer=tf.keras.optimizers.Adam(0.1),
  loss='mean_squared_error')
```

Fig. 6 Definition of the Inputs

During the development, it has been identified that some variables meet certain characteristics, such as while the speed time is greater, the learning will be more precise; in this way, the analysis of the data sent has a greater time to perform the calculations, this grants a higher percentage of reliability because it does not rush when processing them. The internal processes that they have would have greater precision.

4.2. Evaluation

In the eighth phase, we already want to evaluate the test; during this phase, it is important to carry out all possible simulations to have a data store; these will allow us to establish behavior to identify them later. The proposed simulations are carried out with real data, which are extracted by the measurements made in the place; that is why to start with the training and the necessary number of simulations so that the algorithm can learn the possible behaviors of the climate in that region Fig .7.

```
print("We star training...")
record=model.fit(temperaturemin,precipitation,epochs=10,verbose=False)
print("Trained model")
```

Fig. 7 Training and Simulation

The case carried out only 10 simulations, as it was a pilot with few data and did not need to rewind many cases or scenarios. To continue with the development, we imported a library of graphic indicators, in this way to be able to obtain the results in Fig.8 graphically.

```
import matplotlib.pyplot as plt
plt.xlabel("#Epoch")
plt.ylabel("Flood probability")
plt.plot(record.history["loss"])
```

Fig. 8 Library for graphics. Finally

We can see a possible scenario that shows the probability of floods in the next 10 years Fig.9.

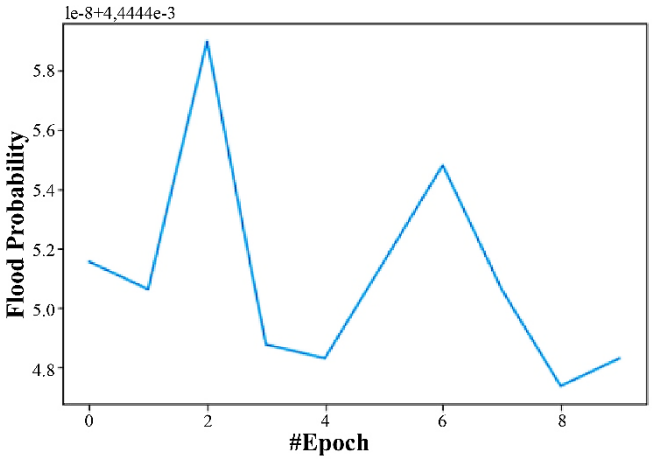

Fig. 9 Probability of flooding in 10 years.

The idea is not unreasonable because know that climate change and the winds of the El Niño phenomenon have increased today (2022-2023).

4.3. Deployment

Finally, tell our algorithm to forecast the precipitation for tomorrow Fig.10.

```
'''FIRST PREDICTION'''
print("Predict tomorrow's weather")
result=model.predict([5.0])
print("The result is"+str(result)+"precipitation")
```

```
Predict tomorrow's weather
The result is [[2.4731889]] precipitation
```

Fig. 10 Visualization of the predictions

Modify the system so that it can forecast the precipitation for the day after the pilot run and see that it returns 2,473, which is an unusual amount for the month of June during its season of the year.

Table 1. Category

VALUE	CATEGORY
$2 \leq X$	Extremely Rainy
$1.5 \leq X \leq 2$	Rainy
$1 \leq X \leq 1.5$	Moderately Rainy
$0.5 \leq X \leq 1$	Slightly Rainy
$0 \leq X \leq 0.5$	Normal
$-0.5 \leq X \leq 0$	Normal
$-1 \leq X \leq -0.5$	Slightly Dry
$-1.5 \leq X \leq -1$	Moderate Dry
$-2 \leq X \leq -1.5$	Dry
$X \leq -2$	Extremely Dry

4.4. Feedback

With two variables, logical answers can be obtained; according to the new data and scientific concepts, they indicate that the greater the amount of real data, the more precise the behavior pattern and, consequently, the projection will have less margin of error.

This answer supports the goal of creating a probabilistic flood warning system; the next challenge is implementing it in a real scenario and adding a method to calculate precipitation and flooding.

But the system not only ends with the prediction of precipitation but must categorize on what scale precipitation can become dangerous. And thus propose a mathematical model that helps us obtain the correct answer and, in the same way, convert it into an algorithm in python so that the system can identify it, display it, and notify.

The value of precipitation is subject to a standardized and categorized scale from extremely dry and ending in extremely rainy; according to this scale can determine the danger of the prediction and be able to translate it into the system; on the other hand, there are limits to which each belongs to the category from -2 to +2 that means, the water accumulation in an environment Table 1.

Consequently, to be able to send radio notifications or synchronize them to another system that sends the alert in advance to mitigate the threat of vital and material losses due to the flood.

About forecasts and predictions, certain coincidences could be seen in the opinions of the different researchers. They all say that the decision-making methodology is the most optimal to reach the answer based on linear algebra methods applied to computer science. Artificial intelligence technologies such as machine learning or deep learning allow the identification of behavior and trends based on an algorithm synthesized according to the research needs.

Regarding the research of the Canadians who used deep learning as a tool for the prediction of climate change, they

obtained a 96% success in their probability [16]; it can be concluded that the greater the capacity of data collection techniques, the better results can be obtained, just like the Iranians regarding climate change anomalies with 97% accuracy [17]. By using sensors and satellite images, the quality and reliability of the data can be guaranteed.

The application of Neural Networks and Machine learning has the potential to be applied in multiple areas; this is due to the concurrence of similar data. It means that while data with similar values are used, they can perform interactions in such a way that they allow us to obtain results that are stored and thus learn the possible results that can be obtained by having this data, can analyze it and apply it to the field to work and display the results.

5. Conclusion

The degree of incidence of weather currents can be determined with the neural network system, calculating the anomalies that occur during their stay in the weather. According to the variation of the degree with which the currents of El Niño and La Niña are presented, the threat and probability of flooding vary. While the degree of incidence is greater, the climatological results will be more devastating. Besides, the agricultural areas most likely to suffer from flood risk are those located on the coast, specifically in Sechura. Banana cultivation may be the most affected by the La Niña phenomenon. The months of January, February, March, and April have a high percentage chance of being very harsh for flood-prone agricultural areas. The synthesis of neural network models is always a challenge for researchers. However, have everything needed to meet the specific goal. By focusing on the banana farming area specifically, it simplifies data upload to the system, facilitating processing times. Pilot tests were carried out with unrealistic data on the banana agricultural area, with good results displayed in the system. The benefit obtained by developing a predictive flood system is very important to avoid agricultural and economic losses. It is also very important so that the farmer can anticipate the disaster and can develop some contingency plan, such as diversion channels, to avoid losses. Finally, this system is very important because it opens paths to other types of research, such as predictions of earthquakes, hurricanes, and all kinds of natural disasters. The importance of probabilistic systems can be used to prevent disasters and mitigate losses of all kinds, whether natural or caused by the hand of the human being. Its use can even have a soccer or banking scope; it can even replace BI indicators and predict future scenarios at the pace that a company is heading; other types of disasters, such as seismic or climatic, can be predicted in the future.

One of the limitations is that during the first stages when applying Deep learning and Machine learning, there are still error percentages, which can be reduced as long as the largest amount of data volume is made for model training. It requires a high capacity of computational resources to process these data.

As future work, a similar model can be proposed, with greater scope, focusing not only on a specific city but on an entire region as well as on a country with a better source of data, such as sensors or satellites, with which more results can be obtained. Accurate.

Acknowledgments

I would like to thank the faculty members of the Faculty of Engineering for their suggestions for the research project, as well as everyone who helped with the research in some way.

References

[1] Michela Biasutti et al., “Projected Changes in the Physical Climate of the Gulf Coast and Caribbean,” *Climatic Change*, vol. 112, pp. 819–845, 2011. *Crossref*, https://doi.org/10.1007%2Fs10584-011-0254-y

[2] Ehsan Meidani, and Shahab Araghinejad, “Long-lead Streamflow Forecasting in the Southwest of Iran by Sea Surface Temperature of the Mediterranean Sea,” *Journal of Hydrologic Engineering*, vol. 19, no. 8, p. 05014005, 2014. *Crossref*, https://doi.org/10.1061%2F%28asce%29he.1943-5584.0000965

[3] Alex Victor Luyo Luque, Fiorella Andrea Guerrero Ciprian, and Jonatán Edward Rojas Polo, “Management of Humanitarian Aid in the Event of Floods Caused by the El Niño Phenomenon, through the Optimal Location of Sensors in Streams And Rivers - Case of the Overflow of the Piura River,” *Engineering, Integration, and Alliances for a Sustainable Development. Hemispheric Cooperation for Competitiveness and Prosperity on a Knowledge-Based Economy: In Proceedings of the 18th LACCEI International Multi-Conference for Engineering, Education and Technology*, pp. 1-11, 2020. *Crossref*, https://doi.org/10.18687%2Flaccei2020.1.1.125

[4] Cesar Adolfo Alvarado Ancieta, and Bernd Ettmer, “Aggradation and Degradation of the Piura River-Realignment of the Lower Main Channel, El Niño Phenomenon, Peru,” *Ingenieria Hidraulica en Mexico*, vol. 23, no. 3, pp. 5–28, 2008.

[5] Pablo Ramirez, and Karin Katty, “*Flood Early Warning System Applying a Hydrological Modeling with HEC-HMS in the Culebras River*,” Thesis to Obtain the Professional Title of Civil Engineer, pp. 1-86, 2019.

[6] Damian Jorge Matich, “Neural Networks: Basic Concepts and Applications,” *National Technological University, Mexico*, vol. 41, pp. 12–16, 2001.

[7] David Quintero Plaza, “Machine Learning and Meteorology: A Perfect Fit,” *Meteorology Statal Agency*, pp. 291–299, 2022. [Online]. Available: http://hdl.handle.net/20.500.11765/13563

[8] Jorge Nanclares, and Ulises M A. Rapallini, “Teoria Armonica Y Machine Learning.”
[9] Hernández Leal, and Emilcy Juliana, “Model for the Process of Extraction, Transformation and Loading in Data Warehouses, An Application with Environmental Data,” *Science and Engineering New Granada*, vol. 26, no. 2, 2016.
[10] Alvaro Corujeira Belghouch,“Migration from Relational Data Model in MYSQL to Vertical in Vertica Through ETL Process,” *Information Systems*, 2021. [Online]. Available: https://oa.upm.es/68229/
[11] Sergio Ramirez-Gallego et al., “An Information Theory-Based Feature Selection Framework for Big Data Under Apache Spark,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 9, pp. 1441–1453, 2017. *Crossref*, https://doi.org/10.1109/TSMC.2017.2670926
[12] Hugo Vega Huerta et al., “Pattern Recognition using Artificial Neural Networks,” *Research Journal of Systems and Computing*, vol. 6, no. 2, pp. 17–26, 2009.
[13] Cemil Kuzey, Ali Uyar, and Dursun Delen, “The Impact of Multinationality on Firm Value: A Comparative Analysis of Machine Learning Techniques,” *Decision Support Systems*, vol. 59, pp. 127–142, 2014. *Crossref*, https://doi.org/10.1016/j.dss.2013.11.001
[14] Yosvani Orlando Lao-Leon et al., “Procedure for Forecasting Demand by Using Artificial Neural Networks,” *Holguin Sciences*, vol. 23, no. 1, pp. 43–59, 2017.
[15] Giuseppe Bonaccorso, *Machine Learning Algorithms*, Packt Publishing Ltd, 2017.
[16] May Haggag et al., “A Deep Learning Model for Predicting Climate-Induced Disasters,” *Natural Hazards*, vol. 107, pp. 1009–1034, 2021. *Crossref*, https://doi.org/10.1007%2Fs11069-021-04620-0
[17] Nguyen Thi Thuy Linh et al., “Flood Prediction Based on Climatic Signals using Wavelet Neural Network,” *Acta Geophysica*, vol. 69, no. 4, pp. 1413–1426, 2021. *Crossref*, https://doi.org/10.1007%2Fs11600-021-00620-7
[18] Gustavo Ovando, Monica Bocco, and Silvina Sayago, “Neural Networks to Model Frost Prediction,” *Technical Agriculture*, vol. 65, no. 1, pp. 65–73, 2005. *Crossref*, https://doi.org/10.4067/S0365-28072005000100007
[19] Ana Cecilia Bermeo et al., “Watershed Management to Face Climate Change and the El Niño Phenomenon: Proposal for Technological Adaptation to Climate Change and the Fen in Piura,” *Soluciones Practicas*, 2008. [Online]. Available: https://hdl.handle.net/20.500.12543/4620
[20] Villanueva Benites Stephanie, “*Logistics management of donations from Piura in Action: Niño Costero Phenomenon 2017*,” University of Piura, 2018. [Online]. Available: https://hdl.handle.net/11042/3463
[21] D. Davila, “Early Warning Systems for Floods in Latin America,” 2016.
[22] Miguel Jimenez-Carrion, Flabio GutIerrez-Segura, and Jorge Celi- Pinzon, “Modeling and Prediction of the El Niño Phenomenon in Piura, Peru through Artificial Neural Networks using Matlab,” *Technological Information*, vol. 29, no. 4, pp. 303–316, 2018. *Crossref*, https://doi.org/10.40672Fs0718-07642018000400303
[23] Bahaa Mohamed Khalil et al., “Application of Artificial Neural Networks for the Prediction of Water Quality Variables in the Nile Delta,” *Journal of Water Resource and Protection*, vol. 4, no. 6, pp. 388–394, 2012. *Crossref*, https://doi.org/10.4236%2Fjwarp.2012.46044
[24] Ismail Olaniyi. Muraina, Bernice Olufunmilola. Lesi, and Moses Adeolu. Agoi, "Review of Connection between ICT Practices and Climate Change in a Developing Nation," *SSRG International Journal of Computer Science and Engineering* , vol. 8, no. 9, pp. 1-4, 2021. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V8I9P101
[25] Nathan Walter, and Sheila T. Murphy, “How to Unring the Bell: A Meta-Analytic Approach to Correction of Misinformation,” *Communication Monographs*, vol. 85, no. 3, pp. 423–441, 2018. *Crossref*, https://doi.org/10.1080/03637751.2018.1467564
[26] Abadie Alberto, “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects,” *Journal of Economic Literature*, vol. 59, no. 2, pp. 391–425, 2021. *Crossref*, https://doi.org/10.1257/jel.20191450
[27] Donghao Zhou et al., “A Survey on Network Data Collection,” *Journal of Network and Computer Applications*, vol. 116, pp. 9–23, 2018. *Crossref*, https://doi.org/10.1016/j.jnca.2018.05.004
[28] Francisco Javier Nieto, Unai Aguilera, and Diego Lopez-de Ipina, “Analyzing Particularities of Sensor Datasets for Supporting Data Understanding and Preparation,” *Sensors*, vol. 21, no. 18, 2021. *Crossref*, https://doi.org/10.3390/s21186063
[29] Stefan Stieglitz et al., “Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation,” *International Journal of Information Management*, vol. 39, pp. 156–168, 2018. *Crossref*, https://doi.org/10.1016/j.ijinfomgt.2017.12.002
[30] Iqbql H. Sarker et al., “Mobile Data Science and Intelligent Apps: Concepts, Ai-Based Modeling and Research Directions,” *Mobile Networks and Applications*, vol. 26, pp. 285-303, 2021. *Crossref*, https://doi.org/10.1007/s11036-020-01650-z
[31] Michael Quinn Patton, “Evaluation Science,” *American Journal of Evaluation*, vol. 39, no. 2, pp. 183–200, 2018. *Crossref*, https://doi.org/10.1177/109821401876312
[32] Touria Hamim, Faouzia Benabbou, and Nawal Sael, “Student Profile Modeling using Boosting Algorithms,” *International Journal of Web-Based Learning and Teaching Technologies*, vol. 17, no. 5, pp. 1–13, 2022. *Crossref*, https://doi.org/10.4018%2Fijwltt.20220901.oa4