# Manual for use of the system

The aim of this document is to explain how to use the system. It is a big data system designed to work with big data related to rainfalls and temperature. The information is downloaded by the system from an open repository. The information is downloaded on a year basis and then the user can create databases from which obtaining minable data. After that, the system lets the user visualize data and apply data mining techniques to predict.

The steps to use the system are:

1. After running the system, we will be requested to choose the language (Spanish or English). We choose "English" or "Spanish" and press on ACEPTAR button.
2. In the field "Year" (up-left) we introduce the year (e.g. "2018") for which we want to download data from all over the world. After that, we press "Download" button and, after indicating the path in our hard disk drive, the download of a zipped file will start. There is a progress bar. This step duration will depend on the Internet connection, which must be active.
3. Then we press "Unzip" and select the file previously downloaded in step 2 from our file directory. We indicate the path where the file will be unzipped.
4. We press "Export CSV to Mongo" and select the file previously unzipped in step 3 from our file directory. This step duration is long since a database will be created on MongoDB for the selected year and the whole world.
5. We repeat steps 2-4 for each year we want to create a database for.
6. After that, we need to create a database (right part of the screen) from which we will take data for analysis. We can create databases at 3 levels: region (Asia, Western Europe, …), country (Jordan, Spain, …) or particular station inside a country. If we select a region, its countries will be listed. If we select a country, its stations will be listed. We can select a station and press the button "See the location on the map", which will open a navigator window to show a Google map with the location of that station. Imagine, for instance, that we want to create a database for the country "Jordan"; we just select the country "Jordan" and press on "Create a database for the selected country". This will create an empty database called "Jordan". We can create as many databases as we want.
7. To include data in the database created in step 6, we just navigate to the left-central part of the application and select the database created; we then select the year (e.g. "2018") to incorporate to the database (e.g. "Jordan") and press on "Update MongoDB". This will load the selected year data onto the selected database. This step is long too.

8. We repeat step 7 for every year we intend to load in the created database. We will do steps 7-8 for any database we want to load with data.
9. After that, we navigate to the down part of the screen where we find 4 functionalities:

CLEAR DATA

10. We can mark "Clear data" and press on "Send" to clear all the data displayed in the central table at any time we wish.

SHOW ON DIAGRAM

10. We mark the option "Show on diagram".
11. We select the temporal range From-to (all the selected year must be included in the created database).
12. We select the variable we intend to visualize.
13. We choose the database we want to use.
14. We press "Send" and data will be displayed. The system shows average data for each month. If our database represents a country or a region, the data shown are the average values from all its stations.

CREATE .ARFF FILE FOR TIME SERIES

10. The purpose of this functionality is to create a file with data prepared to time series data mining. We mark the option "Create file .arff for timeseries DM".
11. We select the temporal range From-to of years to be included in our analysis (all the selected year must be included in the created database).
12. We select the database (The rest of fields and buttons are irrelevant here).
13. Press "Send".
14. We select the location in our folder directory for the data file to be stored and provide a file name (e.g., jordantraining.arff) (otherwise, a default name will be given).
15. We then move to the down-right part of the screen, and select "univariate analysis" and press "Submit".
16. We select the variable to analyse.
17. We select the algorithm to use.
18. We press Submit.
19. We select the file created in step 14 (or any other created for analysis) and data will be displayed. The displayed data are a prediction for the 12 months of the year immediately subsequent to the years loaded in the database. For instance, if we used the database "Jordan" with years "2016" and "2017" and run this

functionality, it will show the prediction for the 12 months of 2018. (Note that railfall in Jordan are very little and prediction is quite poor; temperature data are more accurate).

20. We can also ran "multivariate analysis" repeating steps 15-19, selecting the option "multivariate analysis" in step 15. When we want to predict values for a variable (e.g. rainfall), univariate analysis uses only rainfall previous data for prediction, while multivariate analysis uses all the variables for prediction (min-temperature, max-temperature and rainfall itself).

## CREATE .ARFF FILE FOR CLASSIFICATION

10. In some countries or stations, there are very few data available and predictions are poor. In order to deal with this issue, we have included this functionality to predict in a few-data station with models trained with data from other stations (called many-data stations) nearby with more data availability. We must have a database created for each of those nearby stations. We mark the option "Create file .arff for classification".

11. We select the temporal range From-to of years to be included in our analysis (all the selected year must be included in the many-data stations database to select).

12. We select a variable (we focus on a variable here).

13. In Database, we select one many-data station database whose data will be included for training the model (it only works with stations, no with countries or areas).

14. Press the button "Add for analysis(classification)" and the station data will be included in the model (although no confirmation is displayed).

15. We repeat steps 11-14 for each many-data station database to train the model.

16. Press "Send".

17. We select the location in our folder directory for the data file to be stored and provide a file name (e.g., jordantrainingclassif.arff) (otherwise, a default name will be given).

18. We then move to the down-right part of the screen, and select "classification" and press "Submit".

19. Below the button "Submit" we have to select the few-data station for which we want to predict. If we navigate on Areas/Countries/Stations on the central-right part of the screen, it will be easy for us to choose the few-data station. Once we have selected the correct few-data station, we press on "Submit" (the one which is completely below).

20. We select the file created in step 17 (or any other created for this kind of analysis) and data will be displayed.