A case study of applying machine learning models to Drought Prediction
ADRIANA MIHAELA COROIU
ALINA DELIA CĂLIN
Babeș-Bolyai University, Romania

Abstract
This paper analyses the application of machine learning models (ML) as an approach to predict droughts, with the purpose of early drought detection and reducing its impact by ensuring effective planning and adequate resource allocation year by year.

Regression and Long short-term memory (LSTM) are the methods used to perform the analysis, yielding promising results.

STRUCTURE AND PREPROCESSING OF THE DATA SET
For the experiments we used a data set based on the United States Drought Monitor (USDM) [3], enriched with information from the Harmonized World Soil Database [1].

The objective was to investigate if only meteorological data is sufficient for droughts prediction, considering five levels of drought the sixth category being no drought, annotated as "None".

Each data set entry contains: a FIPS code (ID), the observations date, a value of the level of drought for a county, and meteorological indicators for the last 90 days.

There is an unequal distribution of data, with the drought-free category being considerably more prevalent, as the histogram in Figure 1 shows.

Initially, the data contained 52 columns, then it was pre-processed by eliminating columns with extreme low or high correlation value taken two by two (below 10% or over 85%), remaining 9 columns.

A particularity of the data set is that there is a predicted value for drought every seven records (as this activity is performed once a week).

EXPERIMENTS AND RESULTS
Three different models were trained. Two of them are based on regression and the third on classification using the LSTM neural network.

For the classification model, the predictions were rounded at intervals of 0.25, obtaining 21 classes from the interval [0,5].

For the regression model we chose an architecture of 3 layers with 30 neurons, with 0.2 Dropout after each second layer, ending with a fully-connected, dense layer.

For the classification model we obtained an accuracy of 93% and a loss of 0.41.

We noticed that this model gets a loss other than 0 and when it gets an accuracy of 100% on a batch, this is favorable for the model learning process.

The results are in Figure 2.

CONCLUSIONS and Future work directions
Drought forecasting starting from existing meteorological data, using an LSTM type neural network, with approaches based on both regression and classification, is feasible and providing satisfactory results, helping in combating the harmful effects of drought by timely detection and strategic allocation of resources.

New ideas ...
- more data
- real data555
- new approach using metaheuristic to determine the best values of parameters

REFERENCES


We appreciate the awesome idea of this conference!