A case study of applying machine learning models to Drought Prediction

ADRIANA MIHAELA COROIU* and ALINA DELIA CĂLIN*, Babeş-Bolyai University, Romania

This paper analyses the application of machine learning models (ML) as an approach to predict droughts, with the purpose of early draught detection and reducing its impact by ensuring effective planning and adequate resources allocation year by year. Regression and Long short-term memory (LSTM) are the methods used to perform the analysis, yielding promising results.

CCS Concepts: • **Information systems** → *Data analytics*; • **Theory of computation** → **Models of learning**.

Additional Key Words and Phrases: regression, lstm, prediction, drought

ACM Reference Format:
Adriana Mihaela Coroiu and Alina Delia Călin. 2018. A case study of applying machine learning models to Drought Prediction . 1, 1 (June 2018), 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION AND SHORT LITERATURE REVIEW

Droughts are a major cause of human losses, with a detrimental role in the evolution of food crises by declining agricultural and livestock production, but also with global economic impact by rising inflation or energy crises. Although there are mitigation measures to combat droughts (irrigation, cultivation of drought-resistant plant species, fertilizers, etc.), they are conditional on early detection of droughts to enable appropriate and timely action.

The drought monitor in the United States of America was used by Hao et al. [2, 3] in a statistical method of categorical drought prediction, with positive RPSS scores for the majority of the studied regions, both for one-month and three-month predictions. The model developed by Zhang et al. [4] predicts droughts using past drought index, weather measurements and climate signals from 32 stations in 1961-2016 in China's Shaanxi Province. The methods they test include: a distributed lag nonlinear model (DLNM), a neural network and an XGBoost model. Overall, their model had the highest prediction accuracy for general drought (89% −97%).

2 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 purpose 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.

*Both authors contributed equally to this research.

Authors' address: Adriana Mihaela Coroiu, adriana.coroiu@ubbcluj.ro; Alina Delia Călin, alina.calin@ubbcluj.ro, Babeş-Bolyai University, M. Kogalniceanu, 1, Cluj-Napoca, Cluj, Romania, RO-400084.

Manuscript submitted to ACM

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 it that there is a predicted value for drought every seven records (as this activity is performed once a week).

3 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. Using Adam optimiser, and the mean squared error for loss, batches of 14,000 entries are trained for 1,000 epochs. We obtained a training accuracy between 80% and 94% and a loss between 0.45 and 0.20. After the initial experiment, we excluded from the data set 90% of the data that had the corresponding class 0. After this process the data set became more balanced. For the second regression model we obtained an accuracy of 7% and a loss of 0.38 (this may be due to not choosing the right metrics, loss, or we encountered the problem of the vanishing gradient). 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 1.

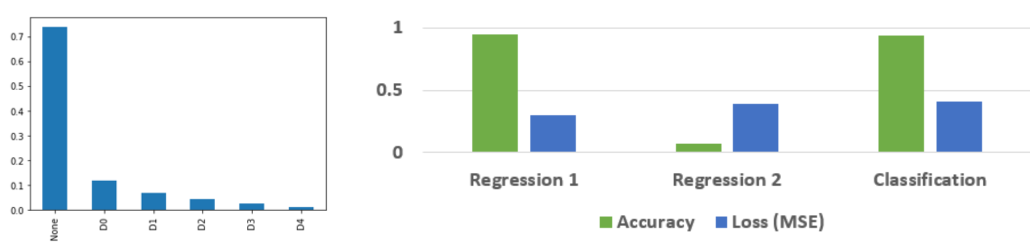

Fig. 1. Data distribution (left) and the models results (right)

4 CONCLUSIONS

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 draught by timely detection and strategic allocation of resources.

5 BIBLIOGRAPHIES

REFERENCES

[1] Tatiana Ermolieva, Günther Fischer, Eva Hizsnyik, Sylvia Prieler, Laixiang Sun, Geza Toth, HT van Velthuizen, and David Wiberg. 2012. Land Use Change and Agriculture Program: MetaData v. 1.0-Main models, databases, tools (covering period 2005-2011). (2012).

[2] Zengchao Hao, Vijay P Singh, and Youlong Xia. 2018. Seasonal drought prediction: advances, challenges, and future prospects. *Reviews of Geophysics* 56, 1 (2018), 108–141.

[3] University of Nebraska-Lincoln National Drought Mitigation Center. 2018. United States Drought Monitor. (2018). https://droughtmonitor.unl.edu/

[4] Rong Zhang, Zhao-Yue Chen, Li-Jun Xu, and Chun-Quan Ou. 2019. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. *Science of the Total Environment* 665 (2019), 338–346.