**Can Artificial Neural Networks Improve the Scope and Accuracy of Existing Short-Term Weather Predictions**

**STUDENT NAME:** Gearoid Lacey

**STUDENT NUMBER:** C00183380

**COURSE NAME:** Masters in Data Science

**DEPARTMENT:** Department of Computing and Networking

**COURSE CODE:** CWS07

**DATE OF SUBMISSION:** 16-02-2018

**CONTENTS**

introduction 3

Weather pREDICTION Literary analysis 4

pROPosed research METHODOLOGY 6

Sourcing and Collection of Data 6

Data Description 6

Data Management 7

Timeframe 7

Proposed Technologies 8

System Validation 8

Strengths and Weaknesses 8

references 10

# introduction

Predicting the weather has the potential to be incredibly useful across many sectors. Existing studies like (Nikam and Meshram, 2013) look specifically at predicting rainfall. The aim of this study is to use an Artificial Neural Network (ANN) to attempt prediction of multiple weather phenomenon at once. Agriculture alone would benefit greatly in the accurate prediction of weather (Shoba and Shobha, 2014). Accurate weather prediction would allow farmers worldwide to plan in advance their weather critical tasks (Gumaste and Kadam, 2016). Increased accuracy of the weather prediction could also eliminate the risk of crop damage due to extreme weather conditions meaning farmers income is less affected.

The ability to predict the weather not only affects Agriculture but could also be useful as a safety mechanism. The potential to forecast the weather in advance could be utilised to predict and reduce road traffic incidents (Mais et. al, 2016). Furthermore, the ability to predict the weather could be used to estimate the total hours of sunshine per day. As a result, this information could be used in the solar energy industry for more efficient and cost-effective energy generation. As such, the impact of an accurate weather prediction model could have a range of diverse uses.

As such the author aims to analyse ANN machine learning algorithms to provide short term weather predictions up to twenty-four hours in advance. This analysis will detail the optimum type of ANN for weather predictions based on multivariate time series data. Once the optimum ANN is determined, it will be implemented to generate the weather predictions. Generating weather predictions twenty-four hours in advance seems like the optimal solution for this study, as it provides a reasonable warning period for undertaking weather critical tasks or for preparation for extreme weather conditions. The weather predictions will include predicted rainfall, wind speed, temperature and humidity. If time permits, the latest weather predictions will then be made available through a location based mobile Application.

# Weather pREDICTION Literary analysis

The purpose of this study is to analyse different types of ANN algorithms to determine which is most suited to forecasting based on multivariate time series data. Once the optimal algorithm is determined it will be used to determine weather forecasts of up to twenty-four hours in advance. Therefore, the research question will be defined as; Can Artificial Neural Networks Improve the Accuracy of Existing Short-Term Weather Predictions.

With the noticeable changes in climate around the world, weather reliant industries such as Agriculture are becoming more difficult to plan and prepare for (Pandey et. al, 2017) . In Agriculture specifically short-term weather prediction could be used to help management decisions such as when is an adequate time to plant, fertilize or harvest crops (Saha and Chauhan, 2017). There are many examples of weather prediction using machine learning many of which focus on one weather element. (Nikam and Meshram, 2013) predict rainfall using Bayesian Classification with reported accuracies ranging from 81.66% to 96.15%. Noticeably as the size of the sample and training set increased so did the accuracy of predictions. (Nikam and Meshram, 2013) also outline that this approach to rainfall prediction is more focused on data analysis rather than computational power.

Bayesian Classification is fundamentally based on the principle of certain attributes being associated with certain classes (Nikam and Meshram, 2013). (O’Hagan, 2008) provides an interesting description for Bayesian Classification; “*You use the data x to learn about the parameters θ*”. In essence, the classification is based on the parameters that have been input being placed in a class or being categorised. In the case of (Nikam and Meshram, 2013) the classification is will it, or will it not rain.

(Sreenivasa et. al, 2014) propose a system that uses Artificial Neural Nets (ANN’s) to predict wind forecasts. Theoretically ANN’s are intended to simulate the human brain where they learn through the occurrence of certain events (Sreenivasa, Agarwal and Kumar, 2014). There are multiple implementations of ANN’s some of which are feed forward ANN’s and ANN’s using back propagation. The authors system uses Adaptive Neuro Fuzzy Inference Systems (ANFIS’s) which is a form of ANN that uses back propagation when its being trained. The main difference between feed forward and back propagation is back propagation adjusts the level of error and retrains the Neural Net based on the new error levels. In comparison feed forward ANN’s run to completion straight away.

(Prasad et. al, 2009) use logistic regression to predict monthly rainfall. Interestingly the model was able to predict the occurrence of extreme weather events a month in advance. Although too much rainfall can be problematic for sectors such as Agriculture, it is the extreme weather events that can be the most damaging. Therefore a system like (Prasad, Dash and Mohanty, 2009) could be hugely beneficial.

(Kuligowski and Barros, 1998) compared an ANN using back propagation to linear regression for rainfall prediction. The authors note that linear regression proved marginally more accurate overall although the ANN was a better predictor for extreme weather conditions. Noticeably the authors used the root mean squared error (RMSE) as a measure of accuracy.

(Hall et. al, 1999) use an ANN to predict twenty-four hour average rainfall. Following this they then determine the probability of this amount of rainfall happening. According to the authors, this method produces impressive results. The authors state that, of the 436 days with a rainfall probability of less than five percent, 435 days had no rainfall. They also state that they had an overall accuracy of 95% for precipitation forecasts.

(Baboo and Shereef, 2010) propose a system that implements an ANN using back propagation to predict the temperature. Noticeably the authors carried out one experiment where they only use two hundred rows of training data on their ANN. Despite the small amount of training data used, the system still predicted accurate results. Also, the attributes in the dataset used by the authors are quite similar to the dataset proposed in this experiment.

# pROPosed research METHODOLOGY

Based on existing studies on similar data, the proposed machine learning algorithm to provide weather predictions is an ANN. Although systems have already been developed to predict the weather using ANN’s, most systems predict one weather attribute only. The proposed system differs from these as it aims to predict four weather attributes, .

## Sourcing and Collection of Data

The proposed source of the data for this experiment is (Met Éireann, n.d.). Met Éireann provides historical weather data for twenty-five weather stations around Ireland. Some of these weather stations are automated and the rest use manual observations. Each weather station provides updated weather reports hourly from 2003. It is worth noting the hourly records for each month are not uploaded until the beginning of the next month. As there is no direct link to download each dataset for each weather station, it is proposed that a python script will be used to scrape the webpage inputting the relevant values in the form and retrieving each dataset automatically.

Met Éireann uses the HARMONIE - AROME (HIRLAM ALADIN Regional Mesoscale Operational NWP in Europe) (Application of Research to Operations at Mesoscale) model for short range weather prediction.(Éireann, n.d.). This model is developed by a consortium of European countries. The model provides Met Éireann with weather predictions up to 54 hours in advance four times a day (Whelan *et al.*, 2016).

## Data Description

The data provided in each of the datasets is described in Figure 1. With the exception of the date column, each column is populated with numerical data.



**Figure 1.** Column Name, Description and Measurement Units

## Data Management

Having downloaded the dataset for Oak park in County Carlow which is an automated weather station, it is evident that some data pre-processing will be required. Initial inspection of the data showed rows of data that were empty for small periods of time. This could be due to maintenance of the weather station. Having cleaned the data by removing empty rows and any obvious outliers, the data will then be stored in a MySQL database.

In the dataset for Oak park in County Carlow alone, there are approximately 126,000 rows of data, some of which contain all null values. The amount of data per station will fluctuate. This is apparent in the datasets for weather stations located in Ballyhaise County Cavan and Shannon Airport County Clare where they have 96,000 and 263,000 rows of data respectively. Having cleaned the data, and assuming the other weather stations report a reasonable amount of data, it is expected that there will be over one million rows of data available for the experiment.

## Timeframe

The following table provides an approximate outline for project completion. The least amount of time has been allocated to the sourcing and collection of data as programmatically retrieving the datasets should be relatively straightforward. As the author has no prior experience of implementing ANN algorithms it is expected that a large proportion of time will be spent both learning and implementing the algorithms.

**Table 1**. Expected Timeframe.

|  |  |  |
| --- | --- | --- |
| Task | Task Description | Expected Timeframe |
| Sourcing and Collection of Data | Create a Python script to scrape data from Met Éireann using Beautiful Soup | 1 - 2 weeks |
| Data Management | Data curation and storage | 1 - 2 weeks |
| Data Analysis | Learning and Implementing ANN algorithms | 12 weeks |
| Application Development | Incorporate Task 3 into a location based weather forecasting application | 4 weeks |

## Proposed Technologies

The proposed programming language for use in this experiment is Python. There are numerous reasons to use the python programming language in an experiment like this, the main reasons being the large quantity of machine learning libraries and vast quantities of documentation to supplement these libraries.

If time permits and development can commence on the mobile application, it is suggested that the application be developed as a native Android application. It would be possible to develop a hybrid application using technologies like (*Adobe PhoneGap Build*, n.d.) which would make the application available for both Android and iOS. Although developing hybrid applications can reduce development time, they have reduced performance in comparison to native applications (Canavesi, 2016). Therefore a native application is the preferred option.

## System Validation

Validation on time series data is slightly more difficult than traditional validation in machine learning. The reason for this is each row of data in a time series dataset is not independent of the events that happened before it (Brownlee, 2016). In meteorology the validation is referred to as hindcasting. There are several methods for validating predictions based on time series data. The most intuitive of these is Walk Forward Validation (WFV). Using WFV you pick the minimum number of rows to begin training on but there is only one row in the test set. Following the first iteration you add the new row to your training set and update your test set with a new row of data. Therefore WFV will create numerous prediction models which can be computationally expensive but the accuracy is greater.

## Strengths and Weaknesses

It is anticipated that the main weakness of this study will involve comprehending in detail, how existing weather forecasting models work. Although there is a large quantity of information on these modes, meteorology is primarily based on the study of physics. As the author has no background in physics, comprehending this information becomes quite difficult. Despite this, it is expected that a basic knowledge of the workings of these systems can be attained.

One of the possible strengths of this experiment would be the reduced hardware requirements for short term weather predictions. Current medium range weather models, which predict up to ten days ahead (Sahin, 2014), require supercomputers as the task at hand is so computationally expensive. Another possible strength of this experiment would be increasing the accuracy of weather forecasting for the basic set of weather elements.

# references

*Adobe PhoneGap Build* (n.d.). Available at: https://build.phonegap.com/ [Accessed: 8 February 2018].

Brownlee, J. (2016) ‘How To Backtest Machine Learning Models for Time Series Forecasting’, *Machine Learning Mastery*, 19 December. Available at: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/ [Accessed: 8 February 2018].

Canavesi, B. (2016) *BENEFITS AND DISADVANTAGES OF HYBRID MOBILE APPLICATIONS*. Available at: https://www.linkedin.com/pulse/benefits-disadvantages-hybrid-mobile-applications-brooks-canavesi [Accessed: 8 February 2018].

Éireann, M. (n.d.) *Met Éireann - The Irish Weather Service(HIRLAM)*. Available at: https://www.met.ie/news/display.asp?ID=277 [Accessed: 6 February 2018].

Gumaste, S. S. and Kadam, A. J. (2016) ‘Future weather prediction using genetic algorithm and FFT for smart farming’, in *Computing Communication Control and automation (ICCUBEA), 2016 International Conference on*. IEEE, pp. 1–6.

Hall, T., Brooks, H. E. and Doswell, C. A. (1999) ‘Precipitation Forecasting Using a Neural Network’, *Weather and Forecasting*, 14(3), pp. 338–345. doi: 10.1175/1520-0434(1999)014<0338:PFUANN>2.0.CO;2.

Kuligowski, R. J. and Barros, A. P. (1998) ‘Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks’, *Weather and Forecasting*, 13(4), pp. 1194–1204. doi: 10.1175/1520-0434(1998)013<1194:LPFFAN>2.0.CO;2.

Mais, D., Lloyd, D. and Davies, J. (2016) ‘Modelling weather effects on road casualty statistics’, *Significance*, 13(1), pp. 28–31. doi: 10.1111/j.1740-9713.2016.00880.x.

Met Éireann (n.d.) *Met Éireann - The Irish Weather Service*, *Display and Download Historical Data*. Available at: http://www.met.ie/climate-request/ (Accessed: 11 January 2018).

Nikam, V. B. and Meshram, B. B. (2013) ‘Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach’, in. IEEE, pp. 132–136. doi: 10.1109/CIMSim.2013.29.

O’Hagan, A. (2008) ‘The Bayesian approach to statistics’, *Rudas, T. Handbook of Probability: Theory and Applications. Thousand Oaks, CA: SAGE Publications, Inc*.

Pandey, A. K., Agrawal, C. P. and Agrawal, M. (2017) ‘A hadoop based weather prediction model for classification of weather data’, in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–5. doi: 10.1109/ICECCT.2017.8117862.

Prasad, K., Dash, S. K. and Mohanty, U. C. (2009) ‘A logistic regression approach for monthly rainfall forecasts in meteorological subdivisions of India based on DEMETER retrospective forecasts’, *International Journal of Climatology*, p. n/a-n/a. doi: 10.1002/joc.2019.

Saha, G. and Chauhan, N. (2017) ‘Numerical weather prediction using nonlinear auto regressive network for the Manaus region, Brazil’, in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*. *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pp. 1–4. doi: 10.1109/IPACT.2017.8245061.

Sahin, C. (2014) *Medium range forecast graphical products (charts)*, *ECMWF*. Available at: https://www.ecmwf.int/en/forecasts/documentation-and-support/medium-range/medium-range-forecast-charts [Accessed: 9 February 2018].

Shoba, G. and Shobha, G. (2014) ‘Rainfall prediction using Data Mining techniques: A Survey’, *Journal of Engineering and Computer Science*, 3(5), pp. 6206–6211.

Sreenivasa, S. C., Agarwal, S. K. and Kumar, R. (2014) ‘Short term wind forecasting using logistic regression driven hypothesis in artificial neural network’, in *2014 6th IEEE Power India International Conference (PIICON)*. *2014 6th IEEE Power India International Conference (PIICON)*, pp. 1–6. doi: 10.1109/POWERI.2014.7117710.

Whelan, E. *et al.* (2016) ‘Met Éireann NWP Highlights 2015’, *ALADIN-HIRLAM Newsletter*, 6, pp. 76–78.

Dr. S. Santhosh Baboo and I. Kadar Shereef, "An Efficient Weather Forecasting System using Artificial Neural Network," International Journal of Environmental Science and Development vol. 1, no. 4, pp. 321-326, 2010.