Project 3

Filip Severin von der Lippe

December 11, 2022

GitHub repository containing code and further instructions on how to reproduce the results of this report: https://github.com/Fslippe/FYS-STK4155/tree/main/project3 https://openarchive.usn.no/usn-xmlui/handle/11250/2581934 https://www.kaggle.com/datasets/jsphyg/weather-dataset-r

Abstract

Contents

1 Introduction **3**

2 Method **3**

2.1 Structure of dataset 3

2.2 Initializing of dataset 3

A Dataset **6**

1 Introduction

2 Method

2.1 Structure of dataset

https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package The dataset includes about 10 years of weather observations from numerous Australian weather stations. The observations include features such as temperature, wind, rainfall, and sunshine hours. A full overview of the dataset and its features can be found in table 2 in appendix A. The main purpose of the dataset is to make a prediction based on today's weather if it is going to rain or not tomorrow. This prediction is either Yes or No, and we are therefore looking at a classification problem.

2.2 Initializing of dataset

The initializing and importing of the datasets require a few steps before we can train our models. These include importing the dataset to a pandas dataframe and removing all rows (days) which has at least one measurement missing. This allows for an easy way to load a training set to our models by excluding Not a Number (NaN) values and their possible negative influence on our models. This leaves us with less data to work with, which may not have an influence on the performance of our models because of the already large amount of data.

- initializing a pandas dataframe,
- remove all rows with at least one measurement missing,
- Look at sized of data from the different stations - spanning from 534 to 3062 for 22 features
- will need package tensorflowdecisionforests
- Bootstrap to reduce overfitting and improve generalization ability
- perform bootstrap on different weight and bias initializations
- Current small scale weather stations - lots of data
- Compairison between areas - assume that masked values not happen at specific times and weather conditions

The dataset comes with wind directions given in a 16-wind compass rose as seen in figure 2. This may cause some problems for neural networks or other methods not built to handle letters or words as input. To solve this problem we translate them to labels from 0 to 15 going clockwise from North (0) to North North-West (15) as seen in table 1. The location of the weather station is given in names which may cause problems for a model using all the available weather data from all the weather stations. This is solved by restricting each train and test sample to a specific weather station. We choose the weather station Cobar with 534 days of data.

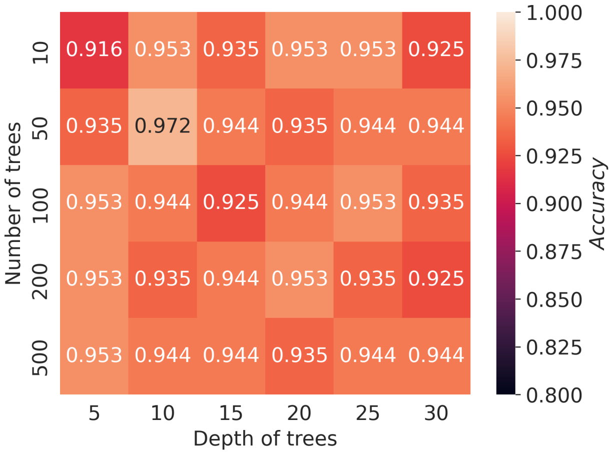

Figure 1: 16-wind compass rose used to describe wind directions. Labels are defined as N (North), S (South), W (West) and E (East).

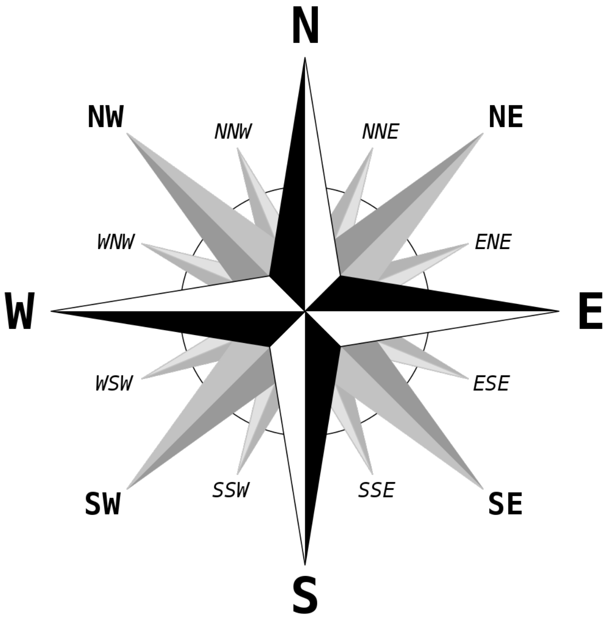

Figure 2: 16-wind compass rose used to describe wind directions. Labels are defined as N (North), S (South), W (West) and E (East).

Table 1

N	NNE	NE	ENE	E	ESE	SE	SSE	S	SSW	SW	WSW	W	WNW	NW	NNW
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

A Dataset

Table 2: Description of dataset features to predict the last feature ”RainTomorrow” if it is going to rain tomorrow or not

Feature	**desctiption**	**Unit**
Location	Common name of the weather station	name
MinTemp	Minimum temperature	degrees Celcius
MaxTemp	Maximum temperature	degrees Celcius
Rainfall	Amount of rainfall recorded in the day	mm
Evaporation	The ”Class A” pan evaporation in the 24 hours	mm
Sunshine	Number of hours with bright sunshine in the day	hours
WindGustDir	direction of the strongest wind gust in the 24 hours	16-wind compass rose
WindGustSpeed	Speed of the strongest wind gust in the 24 hours	km/h
WindDir9am	wind direction at 9am	16-wind compass rose
WindDir3pm	wind direction at 3pm	16-wind compass rose
WindSpeed9am	Wind speed at 9am	km/h
WindSpeed3pm	Wind speed at 3pm	km/h
Humidity9am	Relative humidity at 9am	percent
Humidity3pm	Relative humidity at 3pm	percent
Pressure9am	Pressure reduced to mean sea level at 9am	hPa
Pressure3pm	Pressure reduced to mean sea level at 3pm	hPa
Cloud9am	Fraction of sky covered by clouds at 9am	oktas (units of eights)
Cloud3pm	Fraction of sky covered by clouds at 3pm	oktas (units of eights)
Temp9am	Temperature at 9am	degrees Celcius
Temp3pm	Temperature at 3pm	degrees Celcius
RainToday	Rain exceeding 1mm over 24 hours today	Yes or No
RainTomorrow	Rain exceeding 1mm over 24 hours tomorrow	Yes or No