Udacity Machine Learning Nano Degree Capstone Project
Ian Denness

DEEPAR TIMESERIES SNOW DEPTH PREDICTION MODEL

Domain Background

The domain proposed for this assignment is to explore natural snow depths within the Australian Alpine region. The intent is to provide an interface that that will use the current (i.e. todays) weather to predict the natural snowfall. To simplify the roll out, Falls Creek, Victoria will be used.

Problem Statement

The problem is providing an alternate method for predicting natural resort snow depth, based on previous time series information. Typically, forecasting is done utilising weather prediction methods.

Datasets and Inputs

The datasets utilised are the output, predicted data which is sourced from data.gov.au, and additionally historical information will be used from the Bureau of Meteorology. The information from data.gov.au is a compilation of the target variable (snow depth) and the information from the Bureau of Meteorology is considered to be the independent variables. There are several that can be chosen, but for the purposes of this exercise its proposed to use Maximum Temperature and Rainfall.

Location 1 - Bureau of Meterology

http://www.bom.gov.au/climate/data/index.shtml

This will be accessed individually and CSV files downloaded for each year for each of the parameters, maximum temperature and rainfall.

Location 2 - Victorian Alpine Resorts - Daily Snow Depth Records Falls Creek

https://arcc.vic.gov.au/wp-content/uploads/2021/06/Data-2020-Daily-Snow-Depth-Records_Falls-Creek.csv

This will be accessed directly through webscraping.

They will be joined by combined date to create a dataset with only a several independent variables. Looking at the preliminary data, there will be some manipulation to combine these into a suitable output.

Solution Statement

The model will be able to predict a future time series data points based on previous data. Its applicable to the domain as typical weather prediction uses empirical weather based predictions. This moves into the hind casting realm of using the past to predict the future. As per the DEEPAR exercise the outcome will be able to be tested against current data by splitting into test and train sets.

Benchmark Model

The benchmark model, will be to use the median snow depth across all years as a baseline compared to using a machine learning model.

Evaluation Metrics

Its proposed that because this is a Regression exercise, that RMSE will be used to compare the model vs historical outputs.

Project Design

Student summarizes a theoretical workflow for approaching a solution given the problem. Discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

The workflow, will consist of downloading the data, cleaning, transforming and rendering it into a format that is suitable for use. Its proposed not to use Time Series models such as through SAGEMAKER DEEPAR which is a recurrent neural network.