Udacity Machine Learning Nano Degree Capstone Project
Ian Denness

DEEPAR TIMESERIES SNOW DEPTH PREDICTION MODEL

Domain Background

Predicting the level of snow within a ski resort during winter would allow pundits to monitor the season based on current levels to better plan their trip. Currently there are only short term snow fall predictions and weather predictions put out by the weather bureau.
As such, its proposed for this assignment is to explore natural snow depths within the Australian Alpine region. Ultimately, the intent would be to provide an interface that that will use the current seasons results to predict the natural snowfall levels for Falls Creek, Victoria will be used.

Problem Statement

Can we use Time Series modelling to predict a ski resort natural snow depth. This would take historical information (historical data) and try and predict the future output based on current season information. This would rapidly update as the season unfolded. The advantage of hindcasting to compare against previous years is that snow seasons year on year can be radically different.

Its proposed to use AWS SAGEMAKER DEEPAR for the process. Its set up for learning from many different time series, in this case different seasons.   DEEPAR is noted to be good at learning seasonal dependencies, so its proposed that its quite well suited for predicting the dataset. Its noted that this is a univariate process, and at the moment the prediction will only be based on historical information.

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 - 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*

Solution Statement

The process will consist of the following process.
1. Download the dataset
2. Check data for consistency and missing values, treat accordingly
3. Determine the seasonal range of the dataset
4. Determine baseline model – Average Daily Values across all years
5. Create Test and Train Datasets
6. Convert datasets to Json Format as required by DEEPAR
7. Upload to AWS S3
8. Train the DEEPAR Estimator
9. Test the Estimator, and compare against baseline model.
10. Deploy the model.

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.

Project Design

The process will consist of the following process.
1. Download the dataset
2. Check data for consistency and missing values, treat accordingly
3. Determine the seasonal range of the dataset
4. Determine baseline model – Average Daily Values across all years
5. Create Test and Train Datasets
6. Convert datasets to Json Format as required by DEEPAR
7. Upload to AWS S3
8. Train the DEEPAR Estimator
9. Test the Estimator, and compare against baseline model.
10. Deploy the model.