KATHOLISCHE UNIVERSITÄT EICHSTÄTT-INGOLSTADT

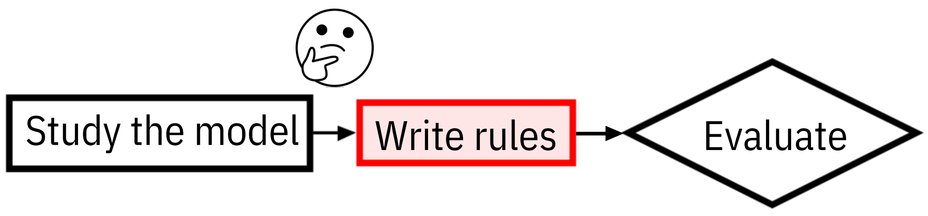
MIDS

Machine Learning Project: Rain in Australia

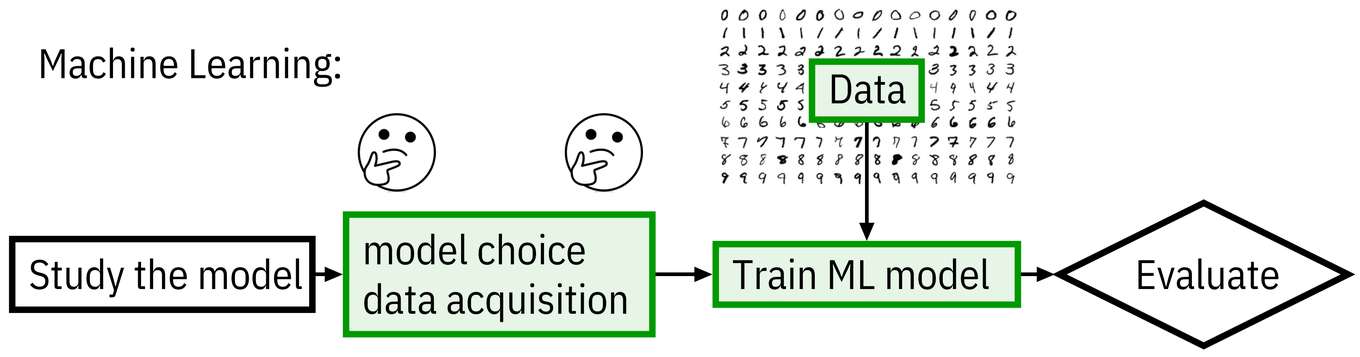
Hands on Machine Learning 2024 Project, worked by: Ali Guliyev & Denis Hoti

What is Machine Learning (ML)?

Classical Programming:



Machine Learning:

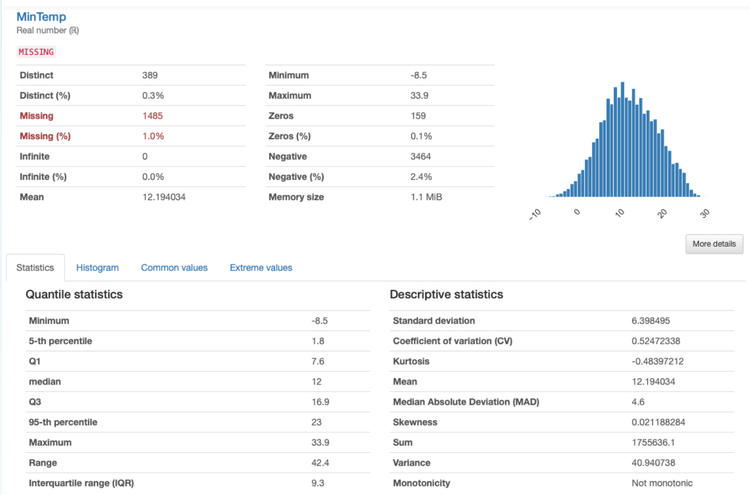


The problem

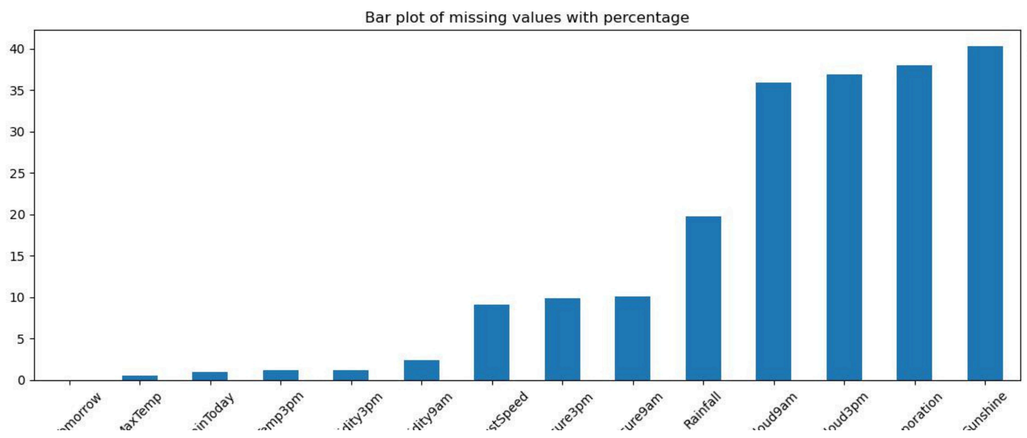
Question: **Will it rain tomorrow or not? Should you carry an umbrella tomorrow or not?**

About the Dataset

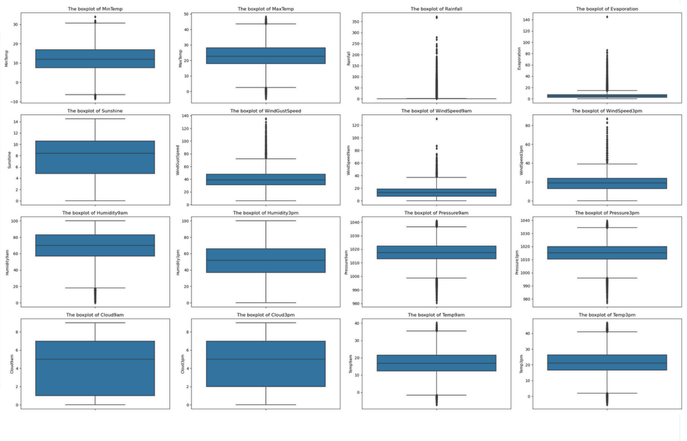
The “Rain in Australia” dataset is provided by Australian Goverment, consist of 10 years of daily weather observations, from multiple cities.



Data Analysis for **MinTemp** attribute

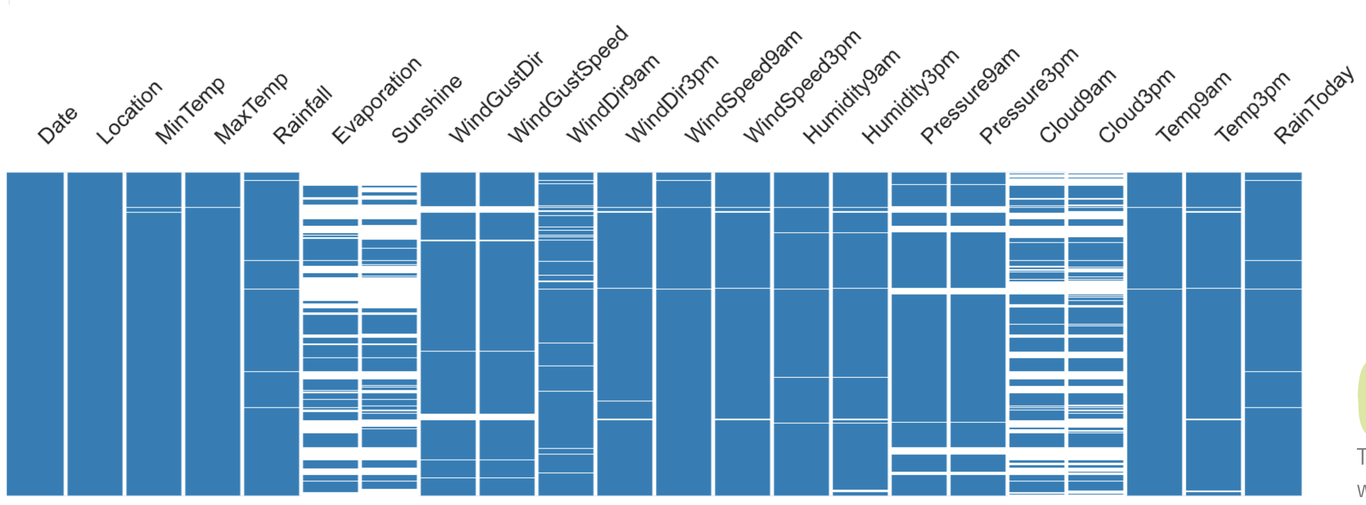


Barplot to show the missing values for each attribute



Graphs which shows the outliers in our dataset

Inputation of Missing values with Gradient Boosting - LightGBM



Columns contain missing values: **all columns** except **Date** and **Location**

How to fill missing values?

Dataset with missing values

Identify Columns with missing values

The columns metioned on the graph: **Sunshine, etc...**

Train Inputation Models

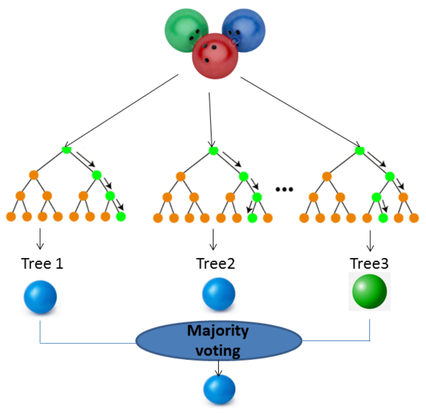
Train seperate LightGBM Models for each column with missing values

Impute Missing Values

Use the trained models to predict and fill in the missing values

Models Evaluation

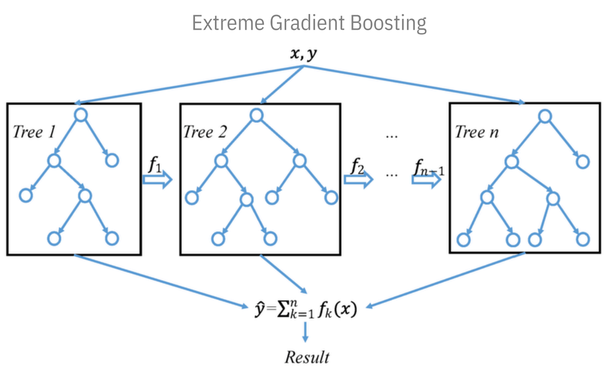
Random Forest



F1 Score Train: **80%**
F1 Score Validation: **69%**

Note: Random forest uses multiple trees to make a decision (prediction).

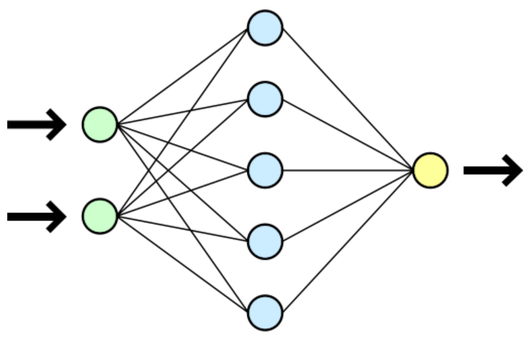
XGBoost



F1 Score Train: **81%**
F1 Score Validation: **72%**

Note: This model gives the best results, but it also takes the longest time to run.

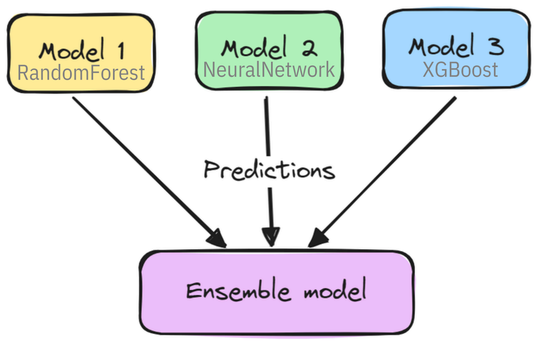
MLP - Neural Network



F1 Score Train: **71%**
F1 Score Validation: **69%**

Note: Neural Networks take a long time to train because of complexity.

Voting - Ensemble Method



F1 Score Train: **79%**
F1 Score Validation: **71%**

Note: This model is nothing but a collaboration of other models.

17.07.2024