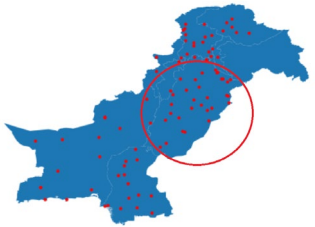
DATA3001 DATA SCIENCE AND DECISIONS

Rainfall in Pakistan

Authors: Eu Shaun Lim, Rachel Ha



Objective

- To improve the rainfall forecast.
- Our model helps to predict whether the stations around Punjab have heavy rainfall.

Overall Methodology

1. Data Cleaning
2. Exploratory Data Analysis
3. Data Modelling
4. Evaluation/Result

Data Description

- Dataset consists of average values of all variables for all stations in Punjab.
- Observations recorded at 3 hour interval
- Response: **RAIN**
- Predictors: **TEMP, DEWPOINT, RELHUMIDITY, SEALVLPRS, WINDSPEED, DIRECTION, TLTCLOUD, VISIBILITY**

Data Cleaning

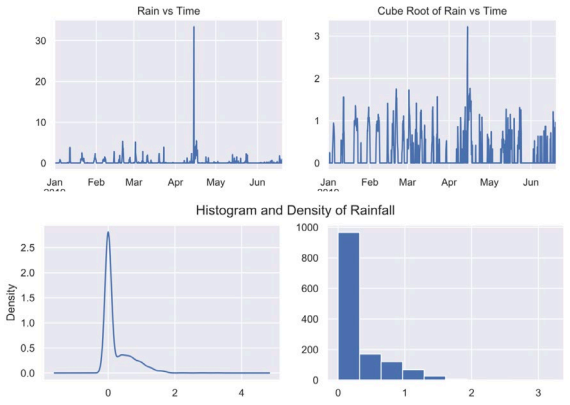
Before shifting

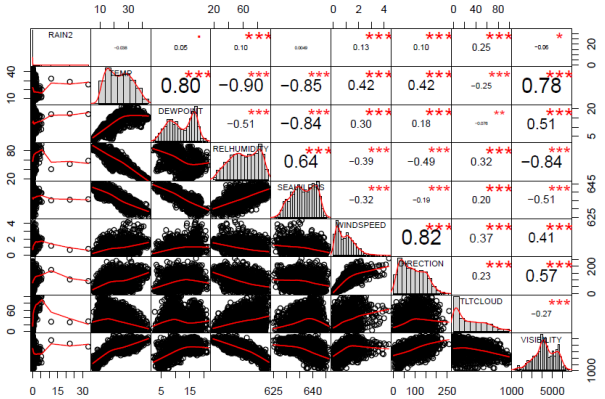
- Replace -9999 values with NaN values (i.e. null value)
- Remove missing date rows

After shifting

- Shifting: shift the rainfall observations three hours back
- Used linear interpolation to fill in the missing values

Exploratory Data Analysis

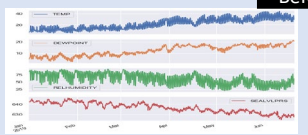




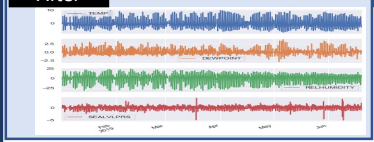
ARIMA Model

1 Differencing

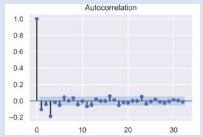
Before



After

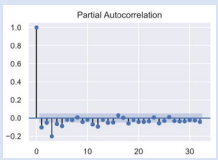


2 ACF and PACF



Provides values of auto-correlation in each series with the lagged values.

Finds the correlation of the residuals with the next lagged value.



3 Model Selection

	coef	std err	z	P>|z|
DEWPOINT	0.0363	0.010	3.622	0.000
RELHUMIDITY	0.0040	0.001	3.543	0.000
SEALVLPRS	-0.0318	0.009	-3.530	0.000
WINDSPEED	0.0828	0.025	3.340	0.001
TLTCLOUD	0.0040	0.001	5.298	0.000

- Stepwise regression → forward/backward selection → AIC = -146.422
- Grid Search → RMSE = 0.217 → p = 6, q = 2

Result

Testing period: 15 April 2019 – 21April 2019

TRUE_RAIN	PRED_RAIN
1	0
1	1
1	1
1	1
1	1
1	0
1	1

- TRUE_RAIN: heavy rain in testing set
- PRED_RAIN: predicted heavy rain by our model

Conclusion/Future Research

- Given that heavy rain has occurred, our model predicted heavy rain with a 71% accuracy.
- The "heavy rain" predicted is moderate rain, as we have little observations on heavy rain.
- Have yearly data instead of a few months data
- Walk forward validation provide more thorough validation method.
- Different models should be considered in the future i.e. Artificial Neural Networks