



ITU AI/ML in 5G Challenge - 2021

ITU-ML5G-PS-012: Radio Link Failure Prediction

Team: MLGrads

Deena Mukundan

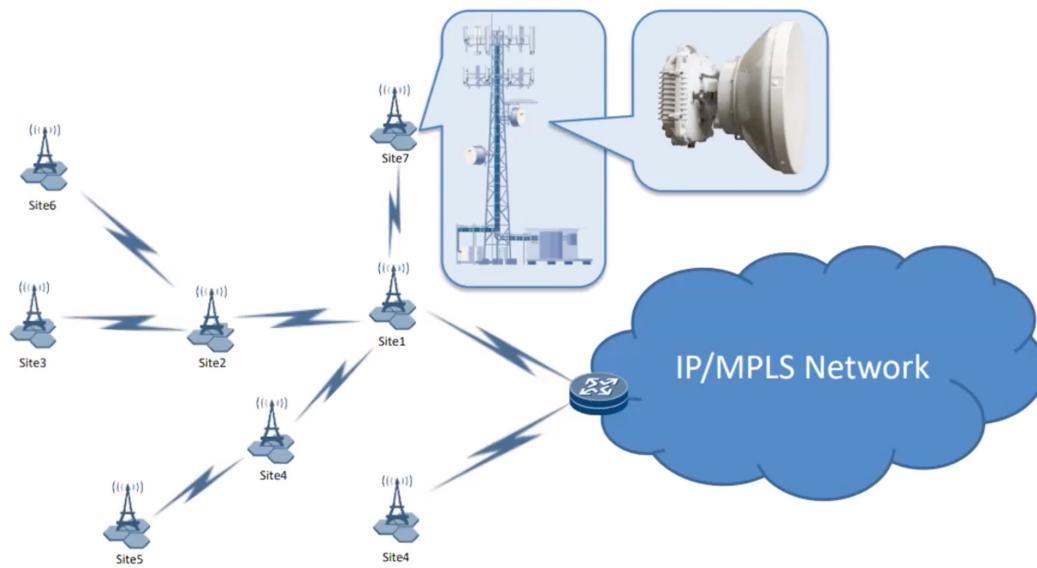
Vijay Venugopal

Agenda

- 1 Problem Statement
- 2 Training Data Observation
- 3 Model Architecture
- 4 Data Imputation Pipeline
- 5 Solution Workflow
- 6 Data Analysis
- 7 Model Training & Validation Results
- 8 Train & Test Data Drift

Problem Statement

- Cloud, rain, snow, and other weather-related phenomena affect the performance of radio links. This is especially applicable to backhaul links operating at GHz frequencies. A generic regional weather forecast data is available which lists expected conditions and coarse temperatures along with actual -precise- realizations.
- Adding to the complexity are the spatial nature of the data (regions of weather data and RLF needs to be aligned) as well as the time sync needed to correlate various occurrences. Over a period of time, we have compiled and anonymized region-wise data which corresponds to weather forecasts, radio link performances and radio link failures derived from our networks.
- Given the region-wise, historical data sets derived from our networks and weather forecasts from the meteorology stations predict the occurrence of radio link failures in:
 - i) in the next day
 - ii) in the following 5 days

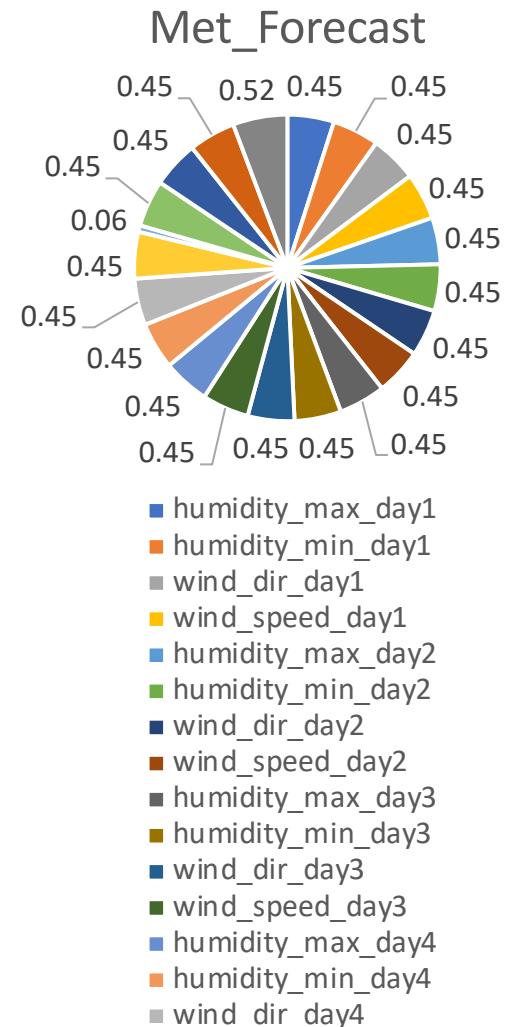
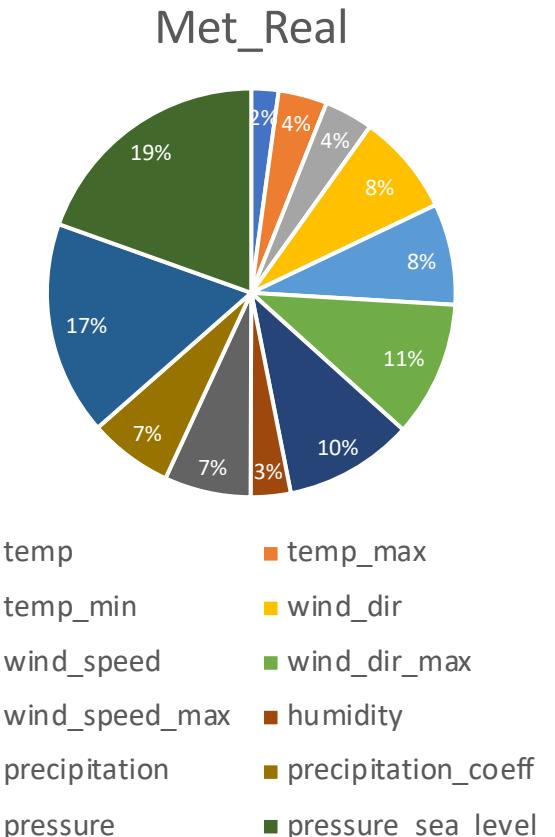


Training Data Observation

Dataset provided are from various sources like radio link (RL) KPIs, RL site details, meteorological real time and forecast data for 5 days. In addition, there is **distance** matrix that stores pairwise distances between weather stations and RL Sites.

Findings:

- Met Real
 - There are 20 unique station_no
 - Datetime have missing dates
 - Missing values found (as shown)
 - Shape - **479241, 17**
- Met Forecast data
 - There are 11 unique station_no
 - Datetime have morning & evening data
 - Missing values found (as shown)
 - Shape - **20323, 39**

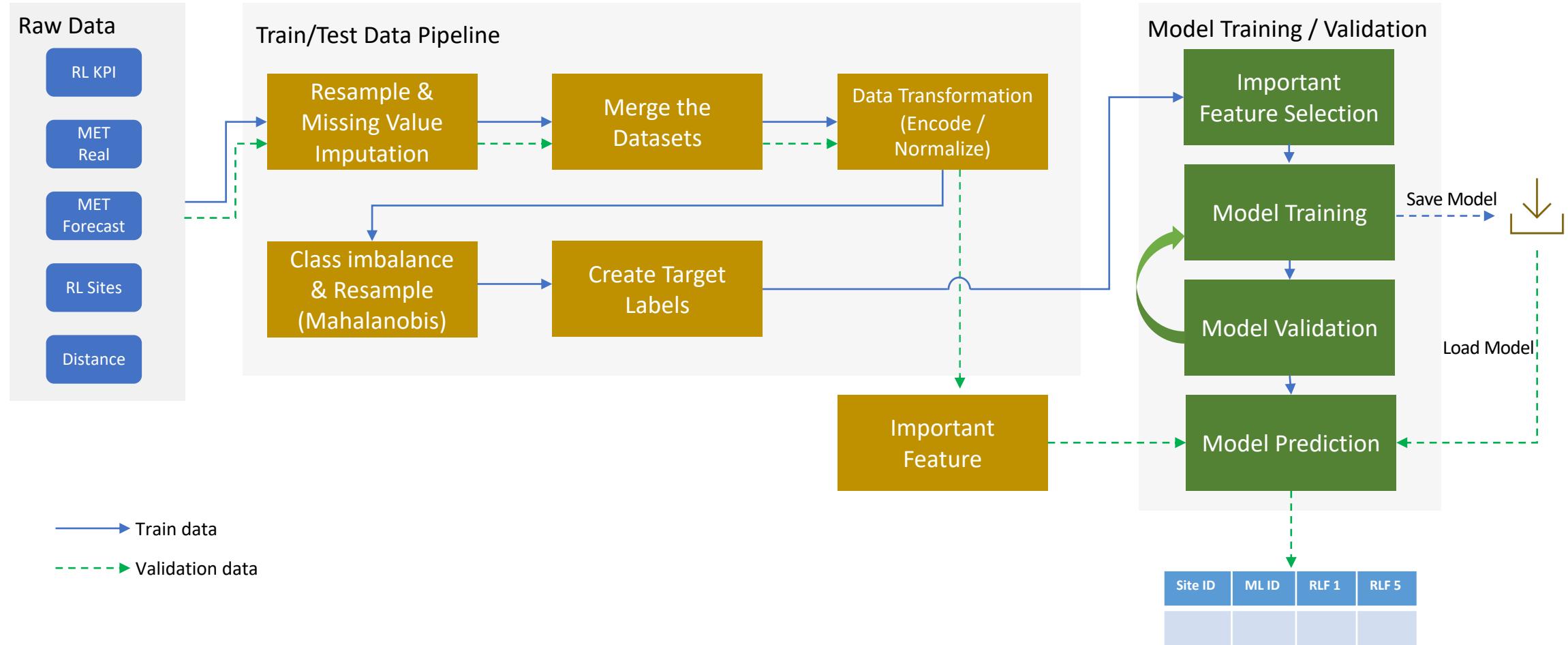


Training Data Observation (contd.)

Findings:

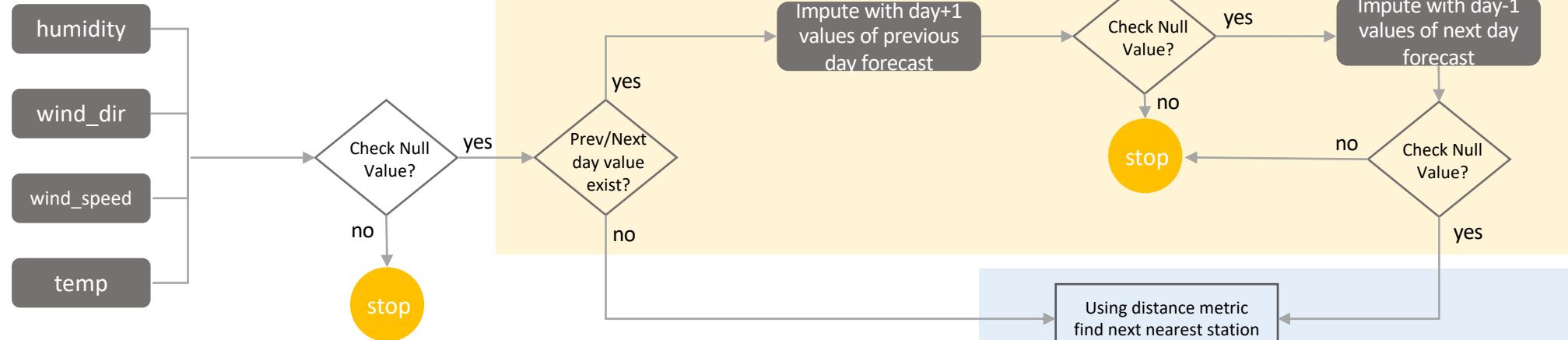
- RL_KPI data:
 - There are 1674 unique site_id
 - There are no stationid in RL KPI
 - Datetime have missing dates
 - Missing values found – freq_band 1.17% & capacity – 0.0082%
 - Shape - **1992986, 19**
- RL_Sites data:
 - There are 1674 unique site_id
 - height & clutter_classes are mapped to the RL sites
 - Shape - **1674, 4**
- Met-Stations data:
 - There are
 - Height & clutter classes are mapped to weather stations
- Distance Metrics
 - Equiviladean distance matrix is provided between all weather stations & RL sites

Model Architecture

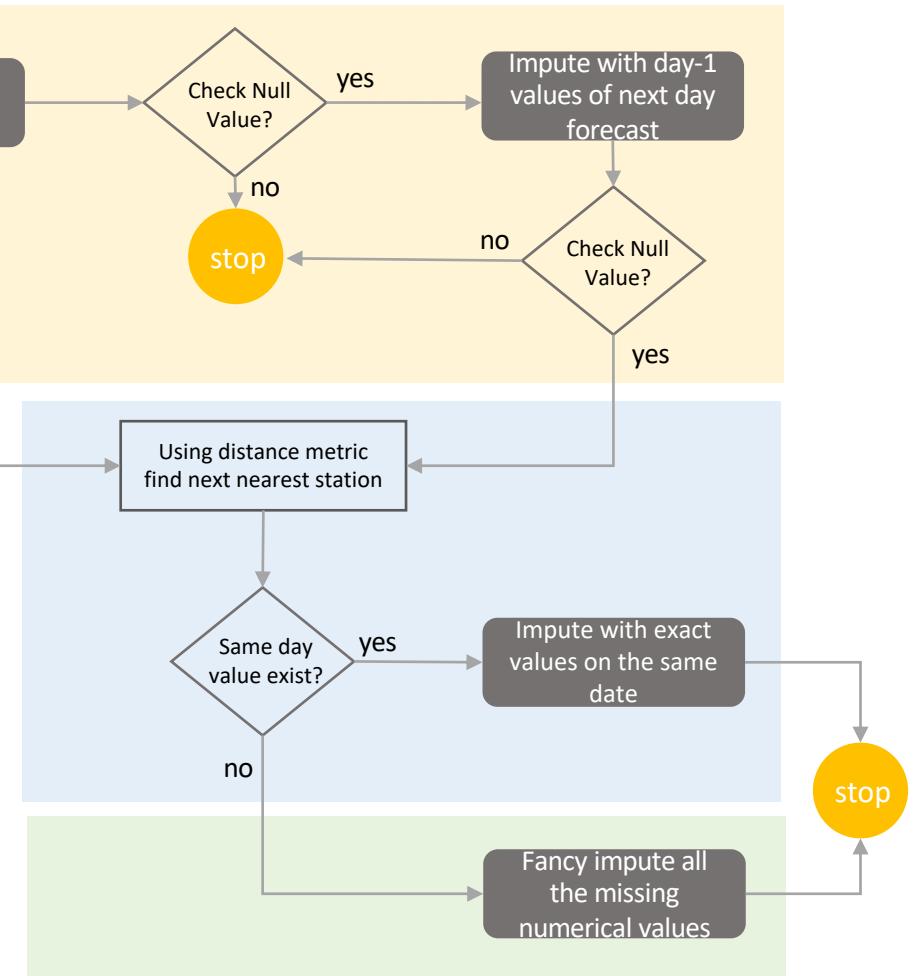
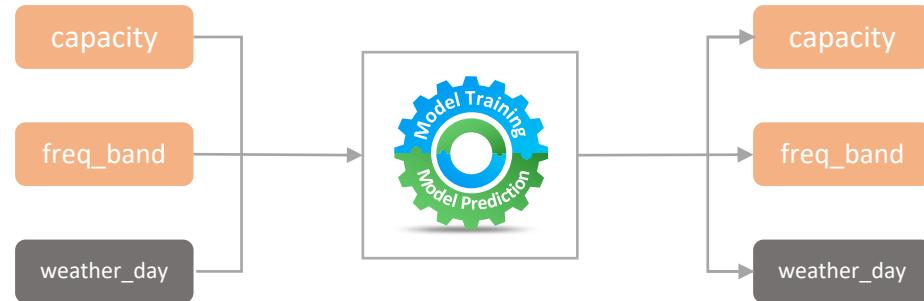


Data Imputation – Met Forecast & RL KPI

Technique 1 – Using Logical Analysis



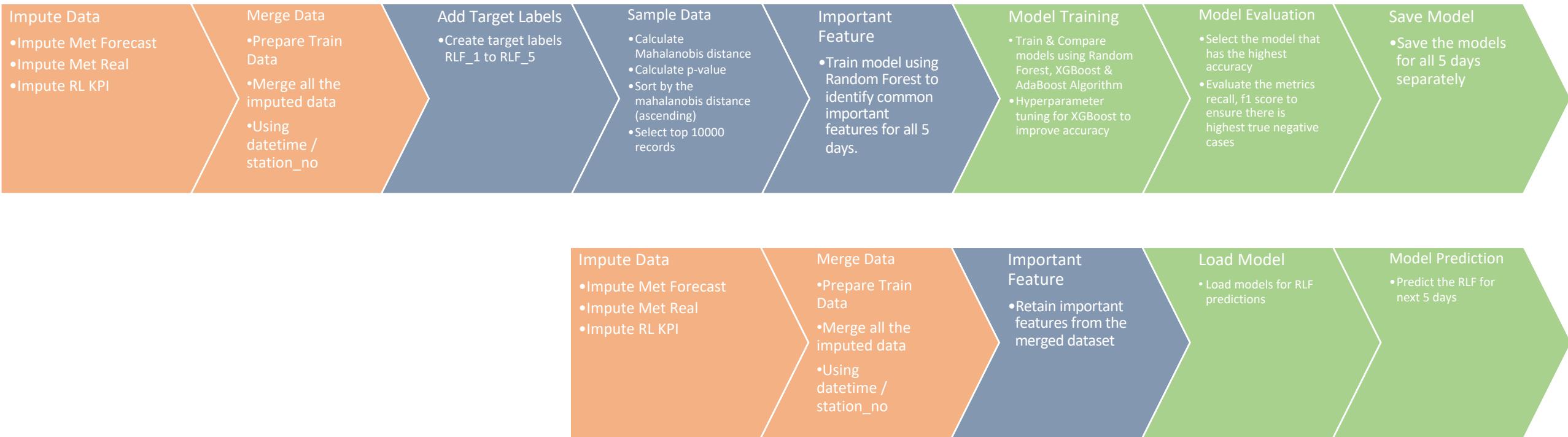
Technique 2 – Using Trained Models



Technique 3 – Using Fancy Imputer (KNN)

Note: Distance metrics is used to find the nearest station when the weather forecast is not available.

Solution Workflow

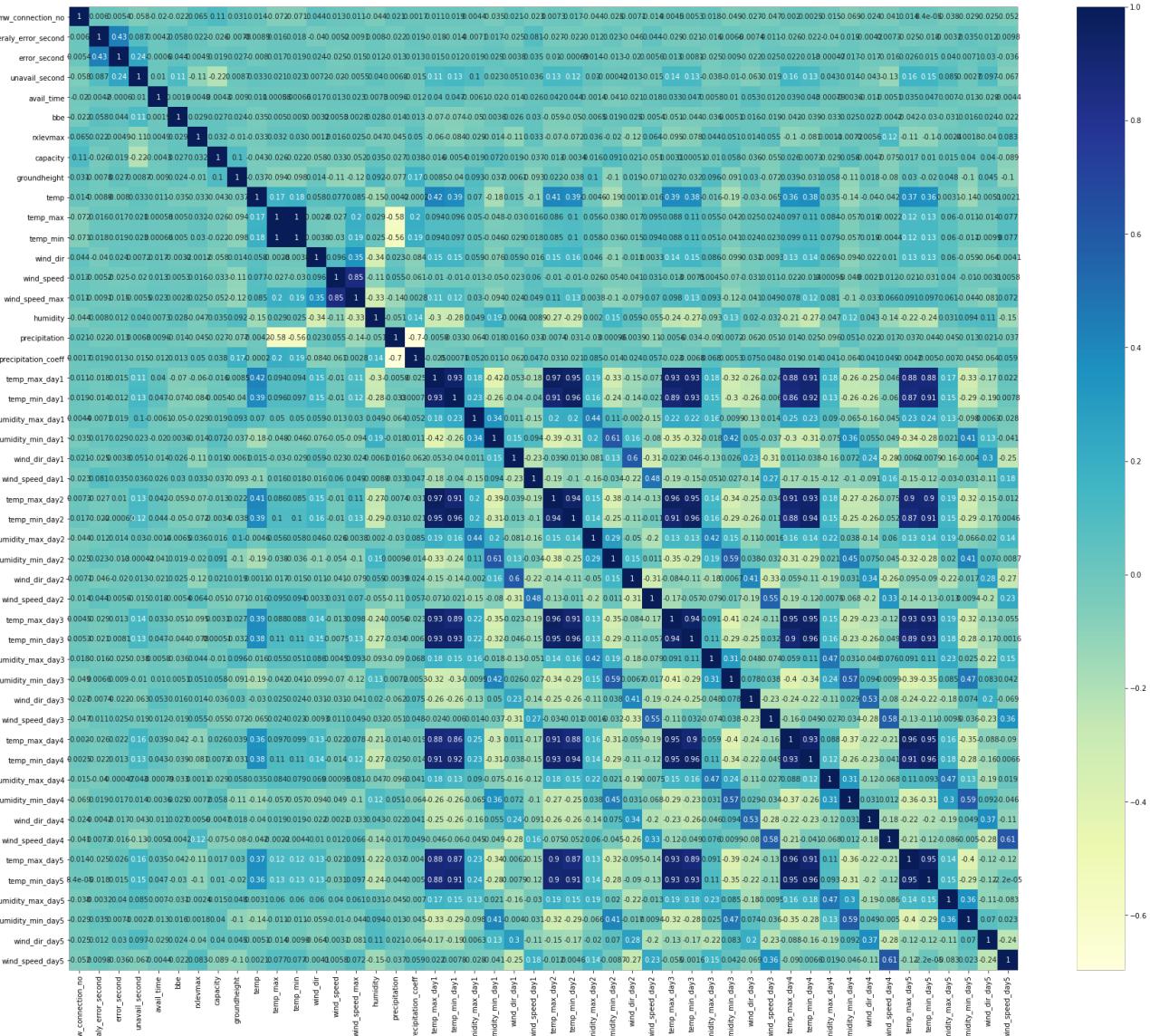


GitHub Repository: <https://github.com/deenammd/MLgrads>

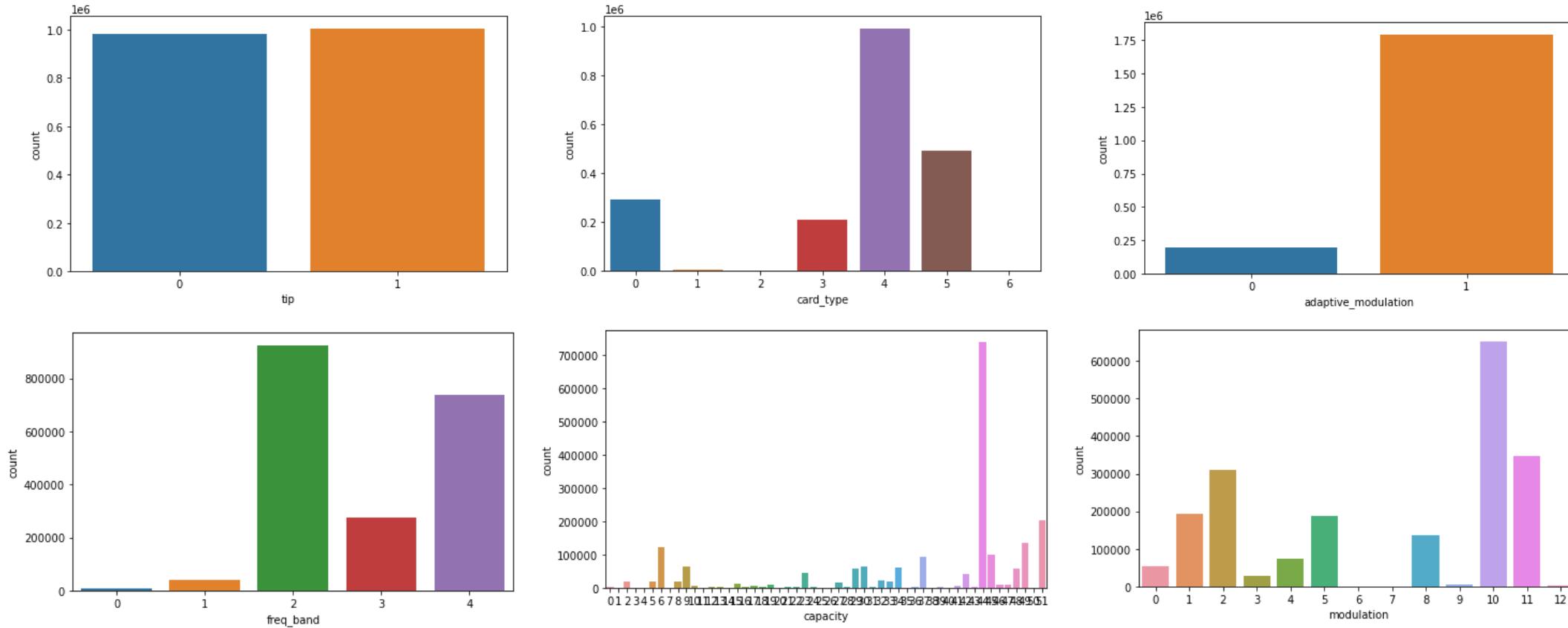
Data Analysis

Plot 1:

From the correlation plot of all numerical variables, it is observed that temp_max and temp_min values are highly correlated.



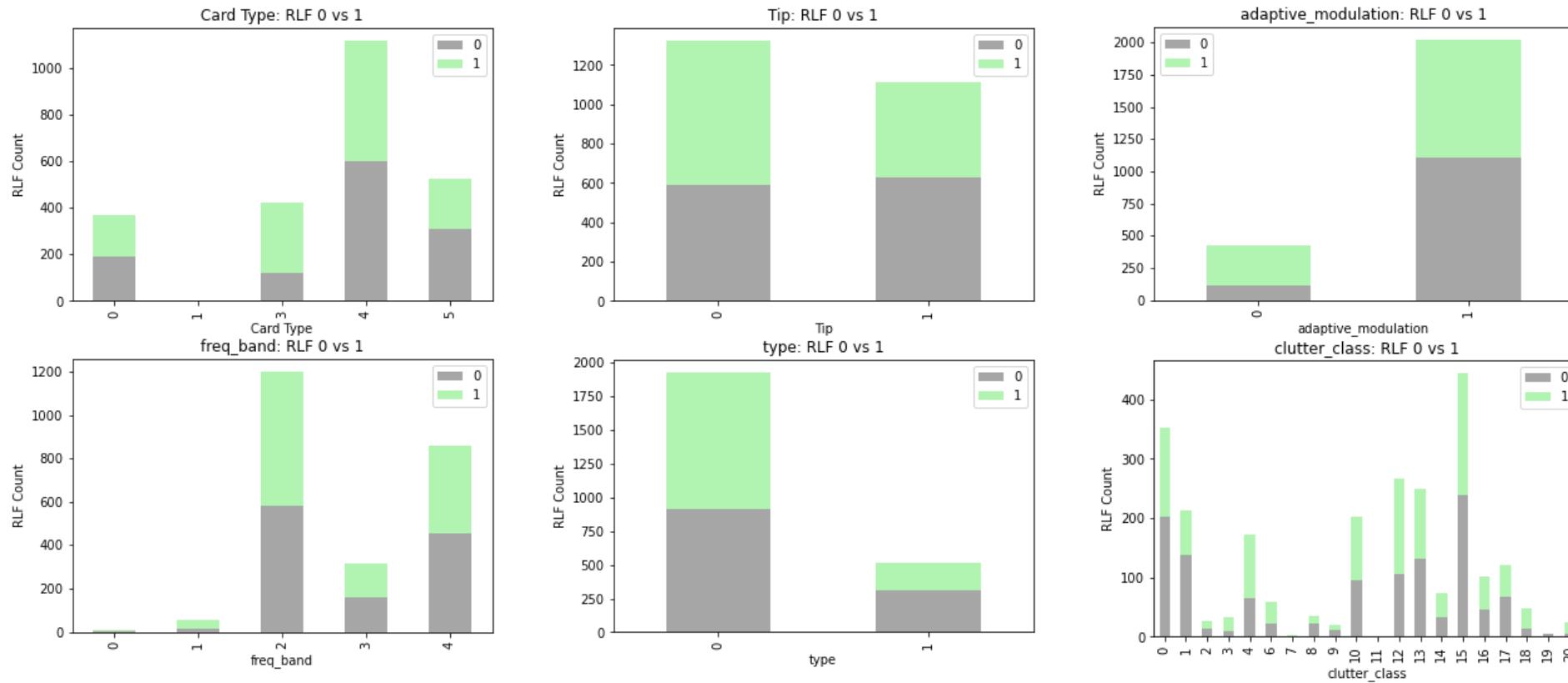
Data Analysis (Contd)



Plot 2: From the categorical variables, we observe the following

- card_type 4 has more data instances
- adaptive_modulation type 1 has more data
- freq_band 2 & 4 have more data instances
- capacity type 44 has significant data instances

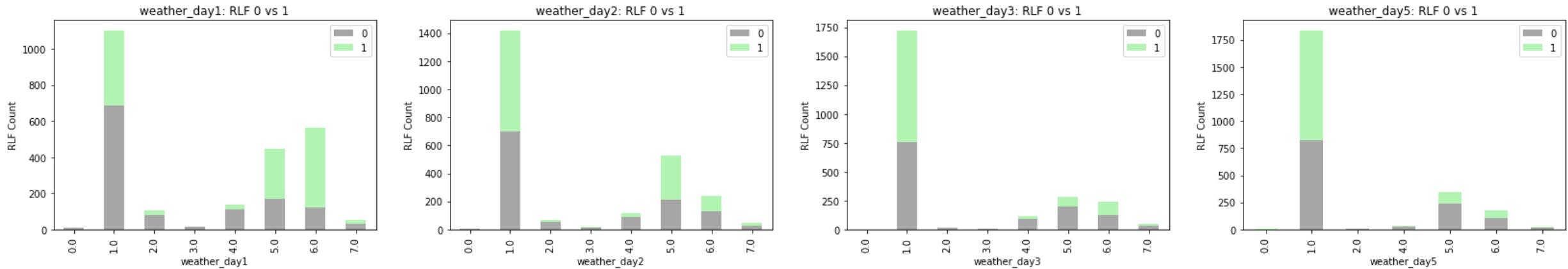
Data Analysis (Contd)



Plot 3: From the categorical variables, we observe the following

- card_type 4 and 3 have more RL failures compared with other card types.
- Tip 0 (FAR) have more RL failures compared with 1 (NEAR).
- Adaptive_modulation 1 have more RL failures compared with 0.
- Freq_band 2 and 4 have more RL failures compared with other freq bands.
- type 0 (ENK) have more RL failures compared with 1 (NEC).
- clutter_class 15 have highest number of RL failures.

Data Analysis (Contd)

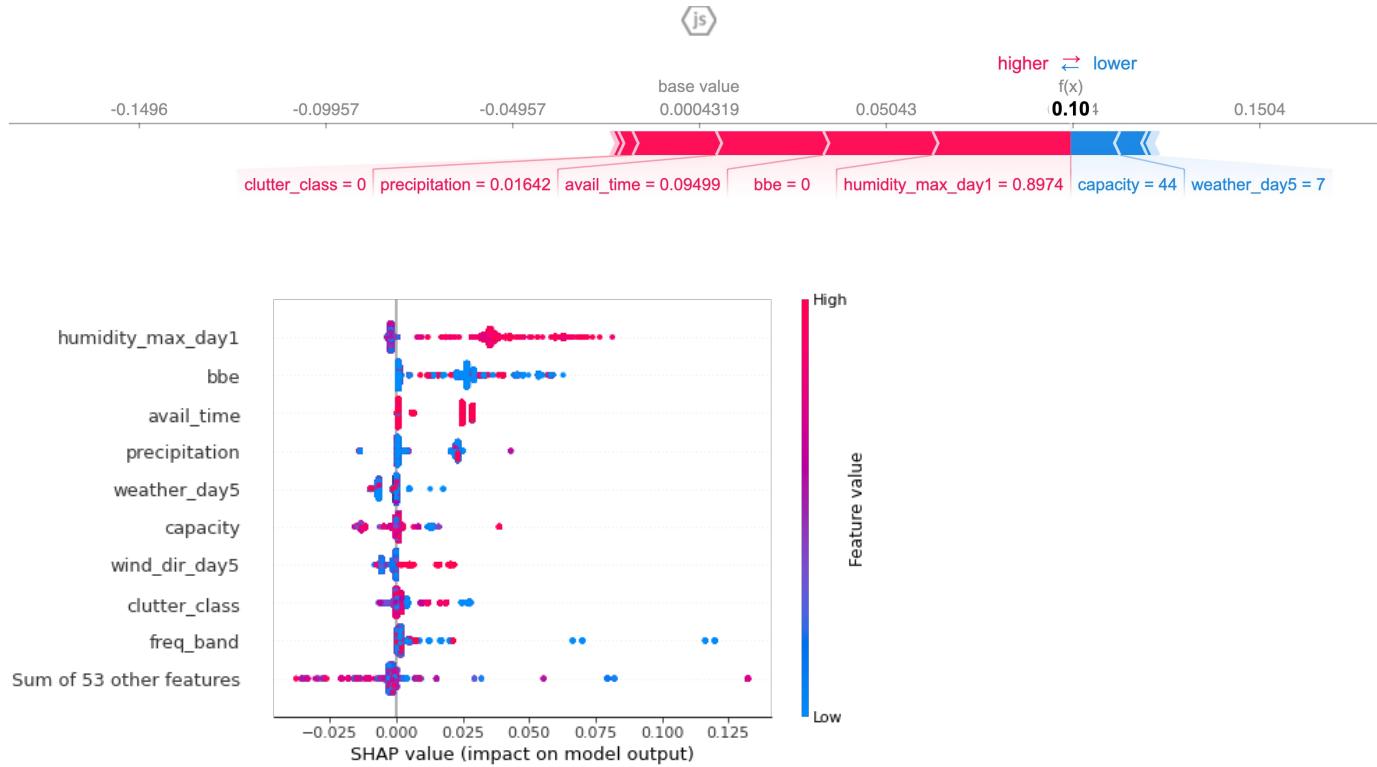
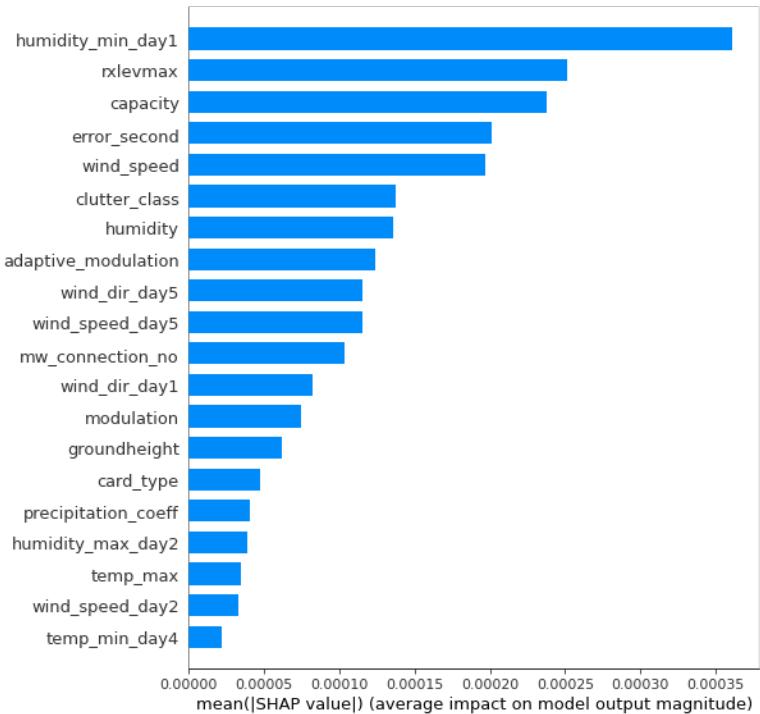


Plot 4: Weather Day correlation chart

- It is observed that weather day has high correlation with the RL failures for all the 5 days.

hot day	0
clear sky	0
scattered clouds	1
few clouds	1
overcast clouds	1
misty	2
foggy	2
windy	3
light rain	4
light rain showers	4
light intensity shower rain	4
rain	5
heavy rain	5
heavy rain showers	5
thunderstorm with heavy rain	6
heavy thunderstorm with rain showers	6
light snow	7
snow	7
sleet	7

Data Analysis (Contd)



Training Data Interpretation:

Training data is explained using SHAP library, there are highly correlated variables as shown in the figure above, humidity_min_day1, bbe, avail_time, precipitation & clutter_class are some of the visible features that contribute to the True class prediction.

Model Training Results

Day 1 Prediction:

[[7411 79] [0 7581]]				
	precision	recall	f1-score	support
False	1.00	0.99	0.99	7490
True	0.99	1.00	0.99	7581
accuracy			0.99	15071
macro avg	0.99	0.99	0.99	15071
weighted avg	0.99	0.99	0.99	15071

RF

5 Days Prediction:

[[6387 22] [34 6255]]				
	precision	recall	f1-score	support
False	0.99	1.00	1.00	6409
True	1.00	0.99	1.00	6289
accuracy			1.00	12698
macro avg	1.00	1.00	1.00	12698
weighted avg	1.00	1.00	1.00	12698

[[7325 165] [0 7581]]				
	precision	recall	f1-score	support
False	1.00	0.98	0.99	7490
True	0.98	1.00	0.99	7581
accuracy			0.99	15071
macro avg	0.99	0.99	0.99	15071
weighted avg	0.99	0.99	0.99	15071

XGBoost

[[6396 13] [14 6275]]				
	precision	recall	f1-score	support
False	1.00	1.00	1.00	6409
True	1.00	1.00	1.00	6289
accuracy			1.00	12698
macro avg	1.00	1.00	1.00	12698
weighted avg	1.00	1.00	1.00	12698

[[6699 791] [789 6792]]				
	precision	recall	f1-score	support
False	0.89	0.89	0.89	7490
True	0.90	0.90	0.90	7581
accuracy			0.90	15071
macro avg	0.90	0.90	0.90	15071
weighted avg	0.90	0.90	0.90	15071

AdaBoost

[[6251 158] [243 6046]]				
	precision	recall	f1-score	support
False	0.96	0.98	0.97	6409
True	0.97	0.96	0.97	6289
accuracy			0.97	12698
macro avg	0.97	0.97	0.97	12698
weighted avg	0.97	0.97	0.97	12698

[[7458 32] [0 7581]]				
	precision	recall	f1-score	support
False	1.00	1.00	1.00	7490
True	1.00	1.00	1.00	7581
accuracy			1.00	15071
macro avg	1.00	1.00	1.00	15071
weighted avg	1.00	1.00	1.00	15071

ExtraTree

[[6371 38] [18 6271]]				
	precision	recall	f1-score	support
False	1.00	0.99	1.00	6409
True	0.99	1.00	1.00	6289
accuracy			1.00	12698
macro avg	1.00	1.00	1.00	12698
weighted avg	1.00	1.00	1.00	12698

5 Days Prediction:

It is observed that XGBoost Classifier has the highest f1-score 100% for both true & false class.

Day 1 Prediction:

It is observed that ExtraTree Classifier has the highest f1-score 100% for both true & false class.

Model Evaluation Results



Model Validation Results

Day 1 Prediction: AdaBoost

```
=====Confusion Matrix=====:
```

```
[[16378 21332]
 [ 4     5]]
```

```
==== Classification Report====:
```

	precision	recall	f1-score	support
False	0.9998	0.4343	0.6056	37710
True	0.0002	0.5556	0.0005	9
accuracy		0.4343	0.4343	37719
macro avg	0.5000	0.4949	0.3030	37719
weighted avg	0.9995	0.4343	0.6054	37719

5 Days Prediction: AdaBoost

```
=====Confusion Matrix=====:
```

```
[[ 1931 35753]
 [ 4     31]]
```

```
==== Classification Report====:
```

	precision	recall	f1-score	support
False	0.9979	0.0512	0.0975	37684
True	0.0009	0.8857	0.0017	35
accuracy			0.0520	37719
macro avg	0.4994	0.4685	0.0496	37719
weighted avg	0.9970	0.0520	0.0974	37719

Validation Result Summary:

It is observed that AdaBoost Classifier is predicting True class with 55% recall (or 0.05% f1-score) for Day 1 & AdaBoost Classifier predicting True class with 88% recall (or 0.17% f1-score) for 5 Days prediction. We have investigated the problem, we found there is data drift for some of the variables which is shown in the next slide. We have employed various other ensemble methods like AdaBoost with Logistic Regression, SVC, RandomForest, without much progression with the accuracy. With the same approach, we have obtained highest accuracy on the training dataset.

Model Validation Results (Contd.)

Day 1 Prediction: XGBoost

```
[[33123 4587]
 [ 9   0]]

===== Classification Report=====:

      precision    recall   f1-score   support
  False       0.9997   0.8784   0.9351    37710
  True        0.0000   0.0000   0.0000     9

accuracy          0.8782    37719
macro avg       0.4999   0.4392   0.4676    37719
weighted avg    0.9995   0.8782   0.9349    37719
```

5 Days Prediction: XGBoost

```
[[ 81 37603]
 [ 0   35]]

===== Classification Report=====:

      precision    recall   f1-score   support
  False       1.0000   0.0021   0.0043    37684
  True        0.0009   1.0000   0.0019     35

accuracy          0.0031    37719
macro avg       0.5005   0.5011   0.0031    37719
weighted avg    0.9991   0.0031   0.0043    37719
```

Day 1 Prediction: Random Forest

```
[[37710   0]
 [ 9   0]]

===== Classification Report=====:

      precision    recall   f1-score   support
  False       0.9998   1.0000   0.9999    37710
  True        0.0000   0.0000   0.0000     9

accuracy          0.9998    37719
macro avg       0.4999   0.5000   0.4999    37719
weighted avg    0.9995   0.9998   0.9996    37719
```

5 Days Prediction: Random Forest

```
[[ 9 37675]
 [ 0   35]]

===== Classification Report=====:

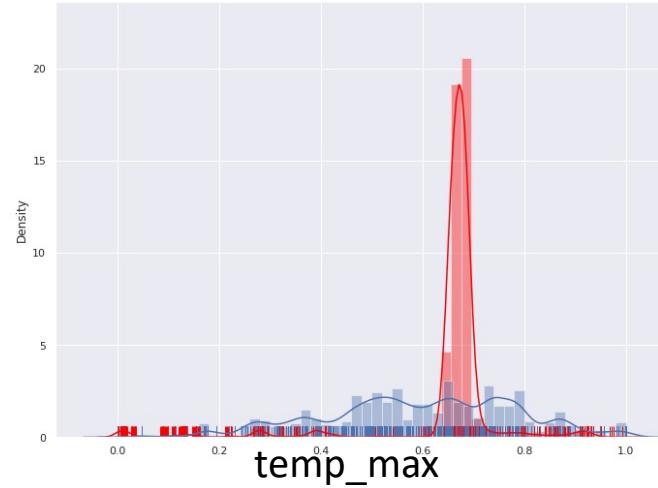
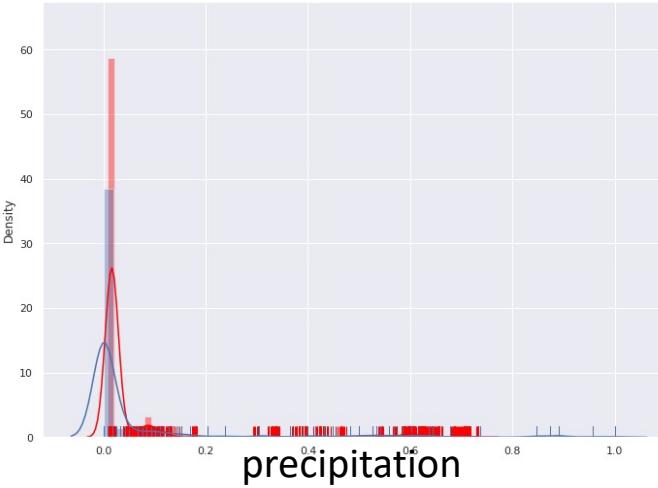
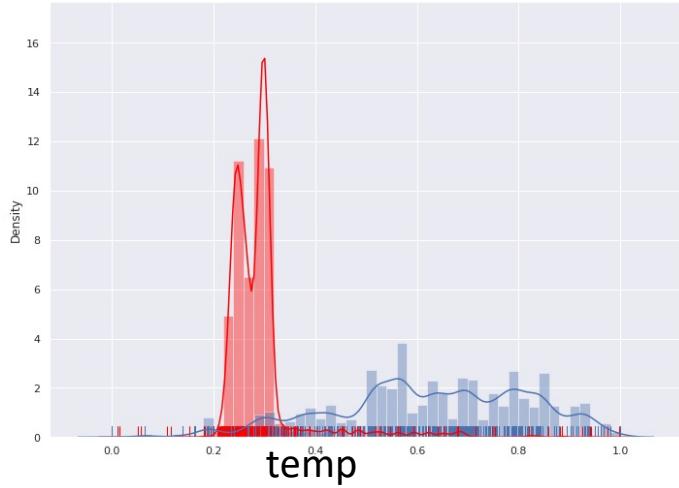
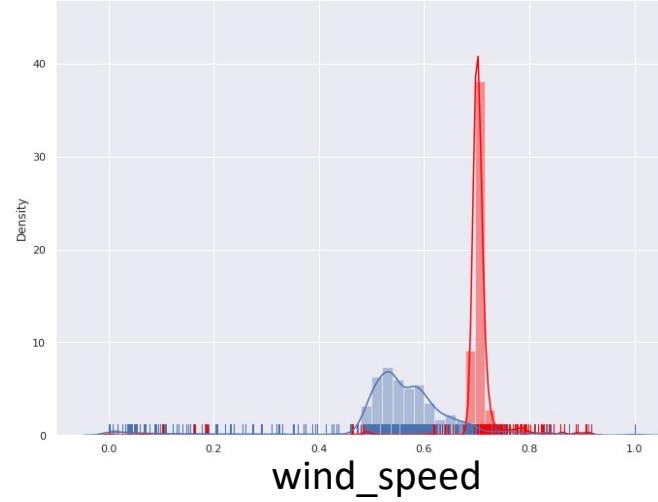
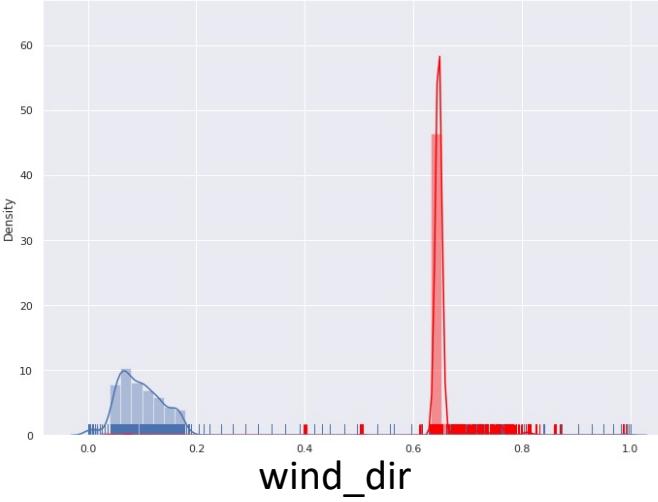
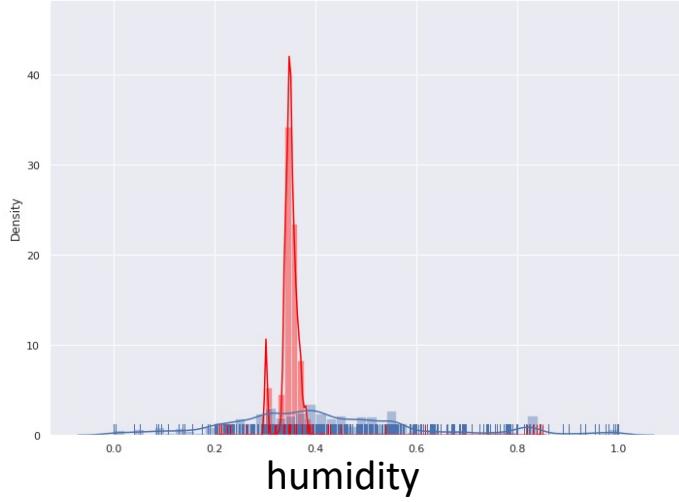
      precision    recall   f1-score   support
  False       1.0000   0.0002   0.0005    37684
  True        0.0009   1.0000   0.0019     35

accuracy          0.0012    37719
macro avg       0.5005   0.5001   0.0012    37719
weighted avg    0.9991   0.0012   0.0005    37719
```

Day 1 Predict: It is observed that XGBoost & Random Forest Classifier is predicting False class with 93% and 99% f1-score respectively & True class is NA.

5 Day Prediction: It is observed that XGBoost & Random Forest Classifier is predicting True class with 0.19% f1-score & Flase class with 0.043% and 0.005% respectively.

Train & Test Data Drifts:



We observe data distribution variations on all the above variables from validation met_real dataset.
Possible solution is to remove the drifting features and retrain the model & predict the results.
Red - Training distribution, Blue - Validation distribution