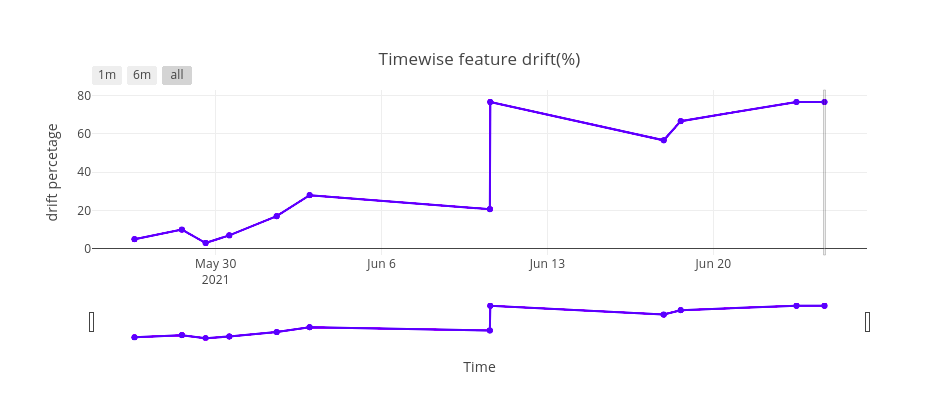
**ML Monitoring**

**1.Data drift detection:-**

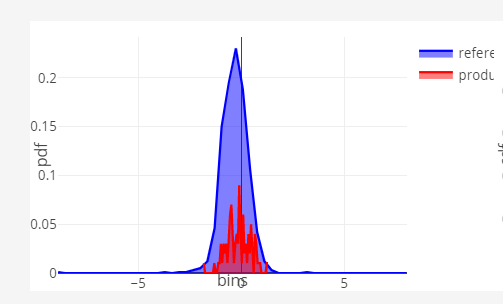
**Time-based line plot:**



* This plot basically represents the timewise percentage of features having drifted in the production dataset so when user clicks on any point then it will display feature-wise plots and information for the particular timestamp which user has clicked

**Feature-wise Plots :**

**PDF plot:**

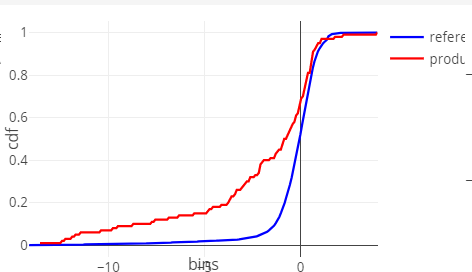


* Using this plot user would be able to understand explanation behind the data drift of production data
* In this plot there will two PDF plots

1. Reference PDF
2. 2) Production PDF

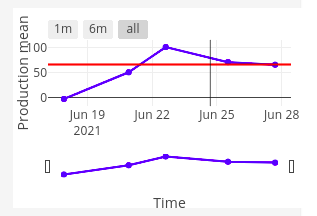
* Using these two plots, it is easy to compare the distribution of reference and production data for that particular feature, and it could bolster the probability of having data drift in production feature data .

**CDF plot:**



* Again cdf plot is having two cdf lines of reference and production
* It gives clear picture of the data-drift result
* If the shape and placement of reference cdf and production cdf are same means there is not distance between these two lines, then there will be no data-drift in production data and if shape and placement are different of reference and production cdf, then there will be a data-drift in production data

**Time-based Mean line plot:**

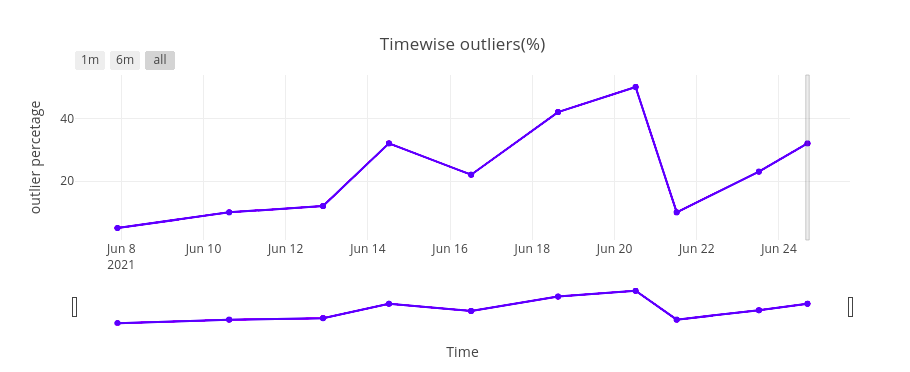


* + Time-based mean line plot represents the timewise mean value of the particular feature of the production data as a blue line in above plot and red line indicates the mean value of the reference data feature. So one can understand that how mean value is changing of particular feature of production data with time.

Other than plots in drift detection we are showing feature level information, feat-impact; (feature Importance) , p\_value and drift detection status.

**2.Outlier detection:-**

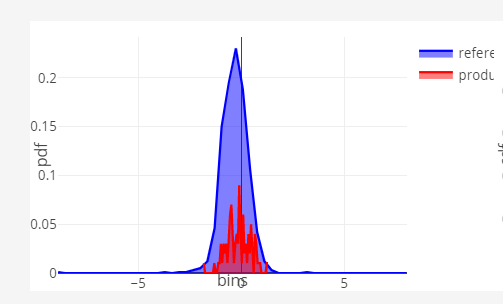
**Time-based line Plot:-**



* This plot basically represents the timewise percentage of outliers in the production dataset so when user clicks on any point and then it will display feature-wise plots and information for the particular timestamp which user has clicked

**Feature-wise:**

**PDF plot:**



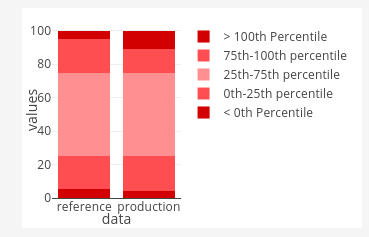
* Like data drift detection here in outlier detection this pdf plot can be very helpful
* In this plot there will two PDF plots

1) Reference PDF

2) Production PDF

* In pdf plot, tale of the distribution can provide the stronger reasons for having a outliers in either production or reference feature’s data

**Stacked bar plot:**



* In this plot there are two stacked bar plots

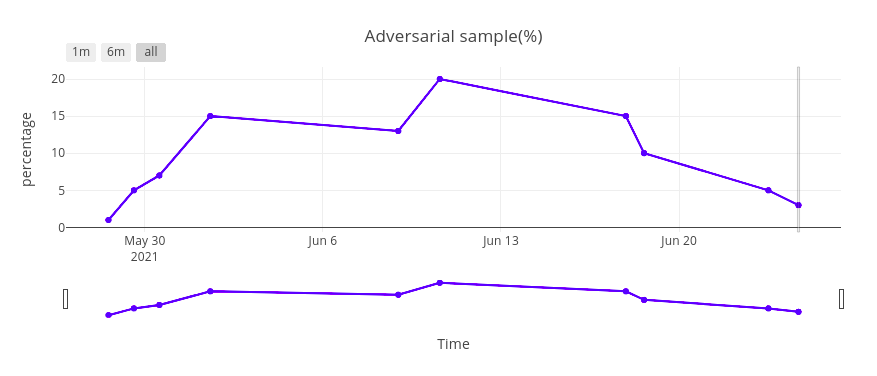
1. Reference data stacked bar plot

2) Production data stacked bar plot

* These two stacked bar plots are being plotted using IQR calculations, which is useful in outlier detection, but here in our dashboard, we are using IQR ranges to plot stacked bar-plot. By referring to these plots the user can get the idea about how many data points are considered as an outlier in both production and reference dataset for that specific feature.
* By using stacked bar p plot user can get an idea about how many percentage of points are between <0th percentile, 0th-25th percentile, 25th -75th percentile,75th- 100th percentile and > 100th percentile for the particular feature of the reference data and production data
* We can get an idea if there are any outliers in a given feature. Outlying points are those points which are <0th percentile and >100th percentile.
* In above stacked bar plot x axis indicates reference data and production data and y axis indicates particular feature values
* We can also get answers like, is the data symmetrical?, how tightly is the data grouped, if the data is skewed and if so, in what direction etc.

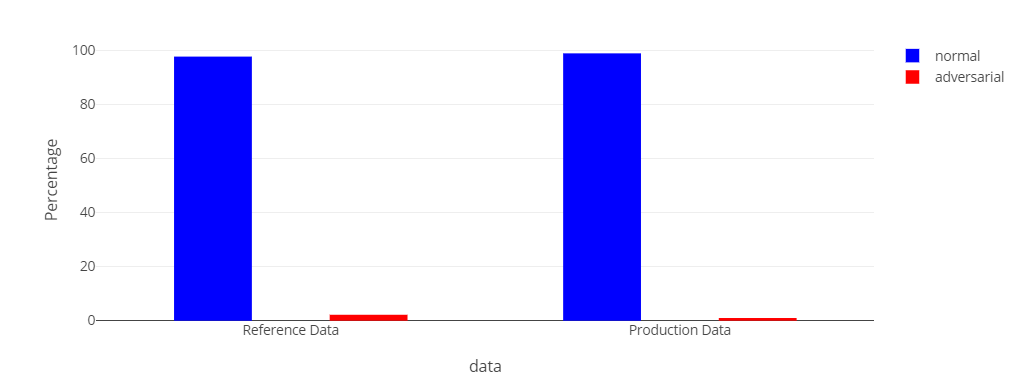
**3. Adversarial detection:**

**Time-based line plot :**



* This plot basically represents the timewise percentage of adversarial sample in the production dataset so when user clicks on any point and then it will display the reference and production data plots with adversarial percentage for the particular timestamp which user has clicked on.

**Bar plot:**

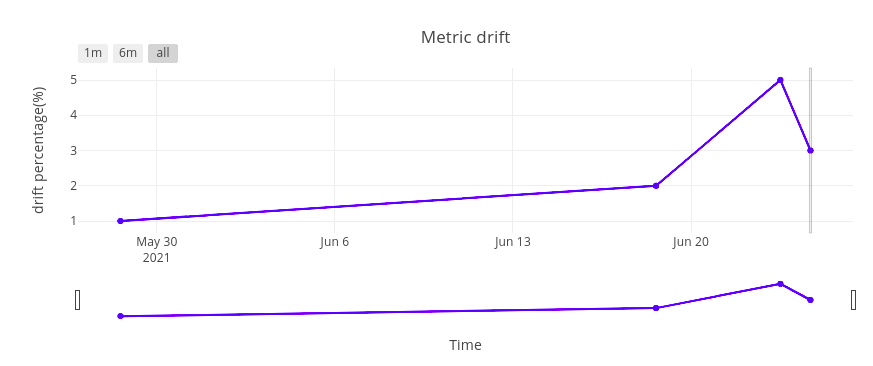


* Here there are total 4 bars:

1. Reference not adversarial: which represents the percentage of points that are not detected as an adversarial datapoint in the reference dataset.
2. Reference adversarial: which represents the percentage of points that are detected as an adversarial datapoint in reference dataset.
3. Production not adversarial: which represents the percentage of points that are not detected as an adversarial datapoint in the production dataset.
4. Production adversarial: which represents the percentage of points that are detected as an adversarial datapoint in the production dataset.

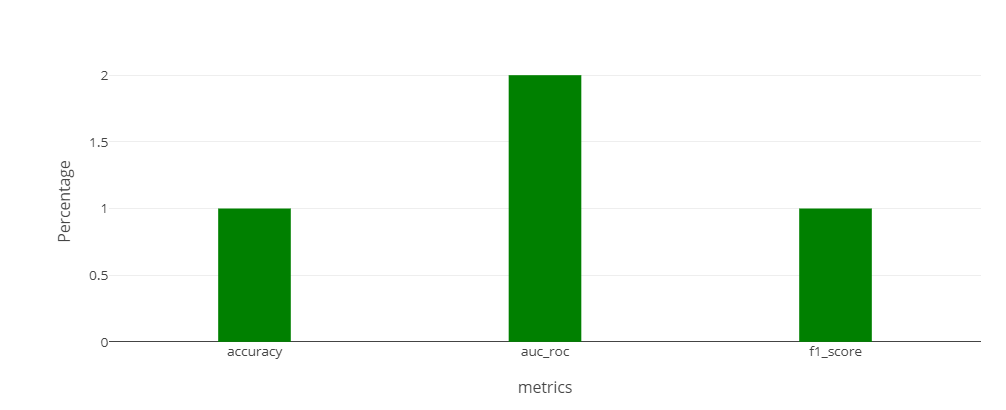
**3.Concept drift :**

**Time-based line plot :**



* This plot basically represents the timewise percentage of accuracy drift in the production dataset so when user clicks on any point and then it will display the accuracy, auc roc and f1 score drift percentage for the particular timestamp which user has clicked on.

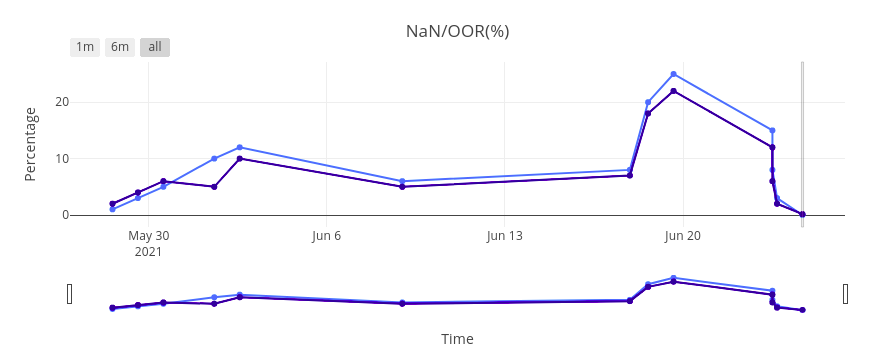
**Bar plot:**



* In the sigmared monitoring platform, we are providing a performance metric bar plot to get drift information of the metrics like accuracy, auc\_roc and f1 score. Bar plot indicates percentage of drift occured in particular metric. By using this bar plot one can understand that particular metric is drifting so the model needs to be retrained.

**4. Data Integrity:**

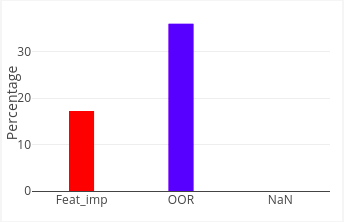
**Time-based line plot :**



* This plot basically represents the timewise percentage of NaN(Null) and OOR(Out of range) in the production dataset so when user clicks on any point then it will display feature-wise plots and information for the particular timestamp which user has clicked

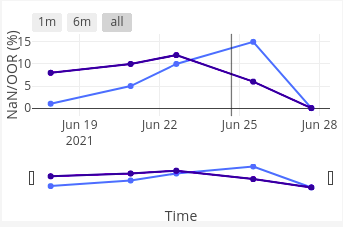
**Feature-wise plots**

**Bar plot:**



* In the sigmared monitoring platform, we are providing a data integrity module which helps users to identify the number of NaN values and OOR(out of ranges) values.
* Basically, this plot represents the total percentage of Nan(null) values and oor(out of range) values in production data.
* Feature-wise bar plot represents the feature level analysis of data integrity module by presenting the percentage of Nan(null) and OOR(Out of range) values for that particular feature alongside with feature-importance of that given feature.
* So that using this plot user can understand how important/crucial this feature is when it comes to data integrity.

**Time-wise line plot**

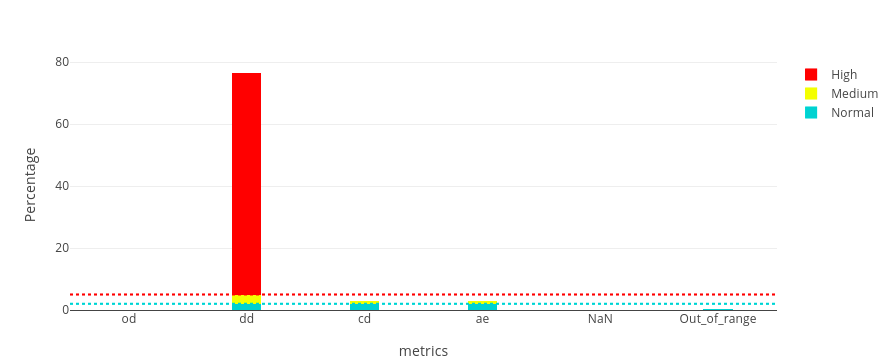


* Time-based line plot represents the timewise percentage of NaN(null) and OOR(Out of range) value for the particular feature of the production data. Dark blue line and light blue line in above plot indicates the OOR and NaN value of the reference data feature. So one can understand that how NaN and OOR values are changing over time for the particular feature of production data.

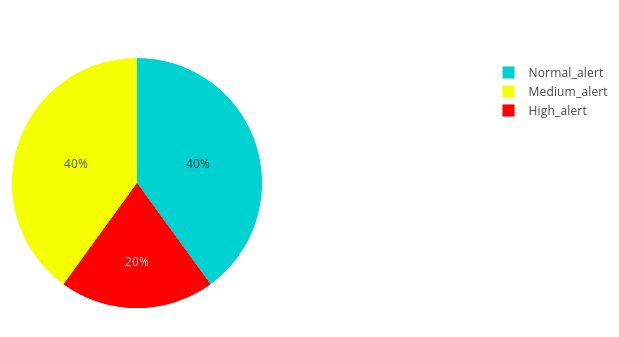
**5. ML Alert**

1. **Monitoring Alert**

* **Stacked bar plot**



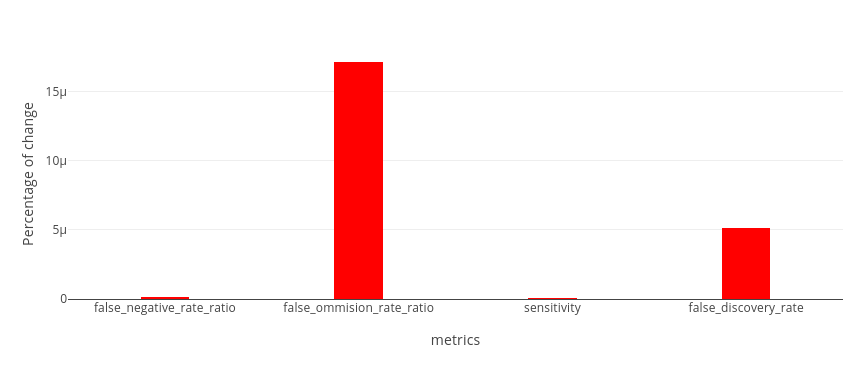
* Monitoring alerts indicate the percentage of outliers, drifted features, adversarial samples, Nan(null) values and OOR(out of range) values in the production data that go beyond the predetermined threshold value and is highlighted through a different colour. If the metric percentage is less than 2% then it is considered as a normal alert (blue), if it is between 2% to 5% then it will appear as medium alert (yellow) and if the metric percentage is beyond 5% then it will appear as a high alert (red).
* **Pie chart**



* The pie chart indicates the Percentage of normal, medium and high alerts.

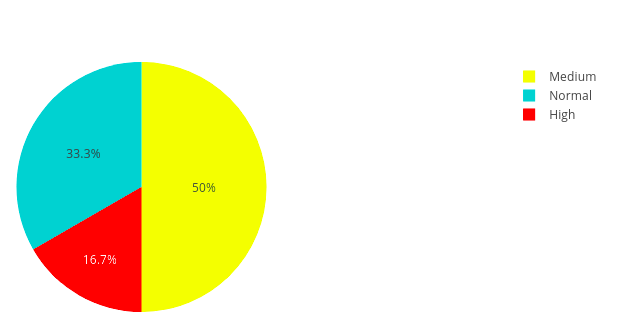
1. **Bias Alert**

* **Bar plot**



* The bias alert bar chart only display metrics that have changed in production. The y-axis indicates the percentage of change that occurred in these metrics.

* **Pie chart**



* The pie chart indicates the number of Normal, Medium and high alerts. The alert message button contains the information of the changed metrics along with their percentage value

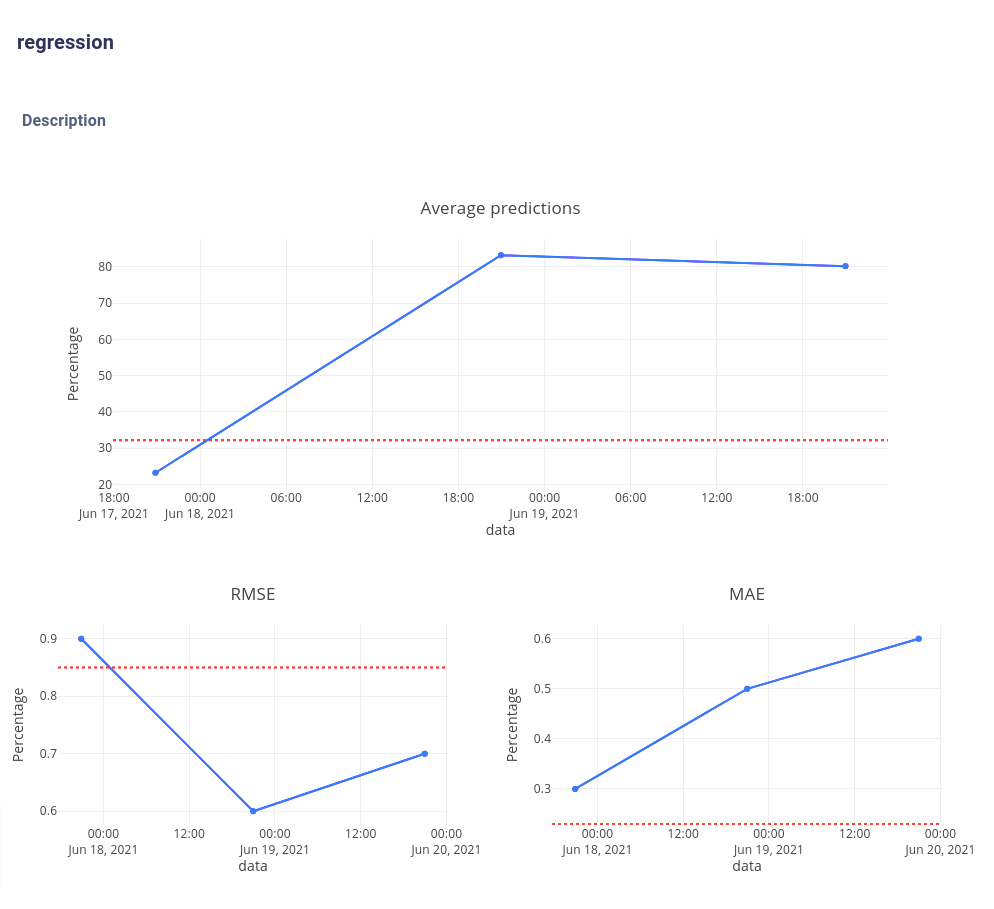
**Model Performance**

1. **Classification**

**Time-based line plot and gauge meter**



* When there is classification model given for the data then model performance tab automatically displays line plot and empty gauge meters.
* This plot basically represents the timewise accuracy percentage of the classification model. When user clicks on any of the point then it will fill the accuracy,f1 score, precision and recall gauge meter with the particular timestamp metric measures.

1. **Regression** 

* When there is a regression model given for the data then model performance tab automatically displays the time-based line plot of the average value of the prediction, RMSE(Root Mean Squared Error) and MAE(Mean Absolute Error) of the regression model. Red line represents average value of the prediction, RMSE(Root Mean Squared Error) and MAE(Mean Absolute Error) of the reference data.