HW6: D(St)reams of Anomalies

Report

Pushkar Singh Negi

Ku ID: 2946319

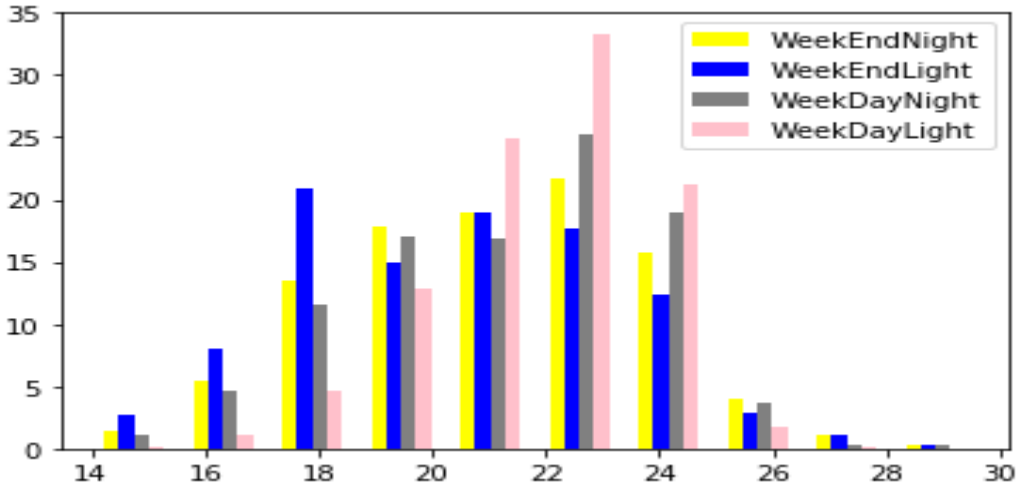
EECS 731: Introduction to Data Science

Notebook: LabProject#6_Anomalies.ipynb Purpose: Deduced Additional Information and isolation forest and One class SVM for anomalies.

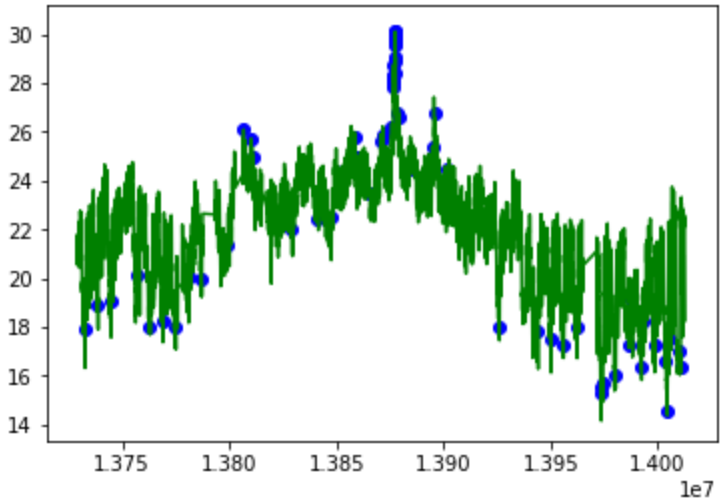
Deduced Additional Information, visualization and anomalies: Notebook: LabProject#6_Anomalies.ipynb

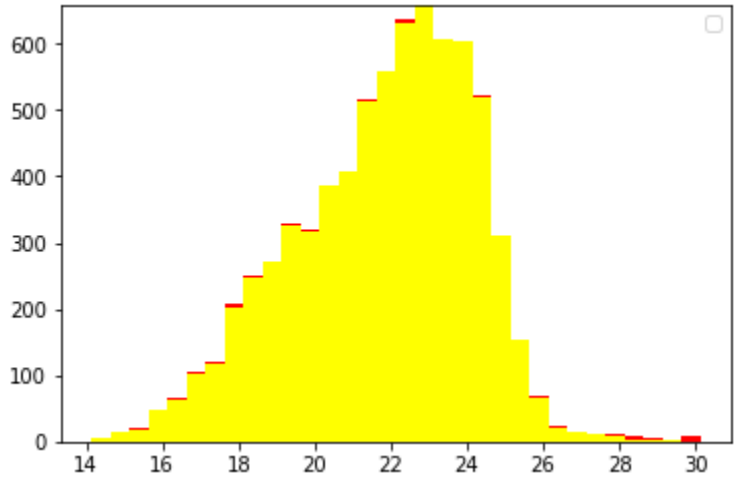
Input raw dataset **:** ambient_temperature_system_failure.csv
Jupyter Notebook File name **:** LabProject#6_Anomalies.ipynb

1. Loaded the ambient_temperature_system_failure.csv in a dataframes.
2. Additional Information #1: The hours and what time of the day is, i.e. it's night or day time.
3. Additional Information #2: To find the day of the week i.e. whether it's a week day or a weekend and assigned numerical (Monday 0 and Sunday as 6)
4. Additional Information #3: Created four categories, i.e. Weekday: day time, Weekday: Night time, Weekend:Daytime and Weekend: Nightime.
5. Visualization: Created a multiple barchart depicting the four categories genetraed in the Additional Information #3



6. Isolation Forest: Anomalies
    - Took useful features such as 'value', 'hours', 'daylight', 'DayOfTheWeek', 'WeekDay' and standardized them.
    - Next, with the help of sklearn.ensemble import IsolationForest trained the isolation forest.
    - Next, added the data in the original dataframe i.e. df.
    - Next, plotted the anomaly against time with the help of visualization (plot)
    - And, also plotted the anomaly against temperature with the help of visualization (plot)





7. One Class SVM: Anomalies
    - We started in the same way as we did for the isolation forest, i.e. took useful features such as 'value', 'hours', 'daylight', 'DayOfTheWeek', 'WeekDay' and standardized them.
    - Trained one class SVM (#nu=0.95 * outliers_fraction + 0.05)
    - Next, added the data to the original dataframe.
    - Next, plotted the anomaly against time with the help of visualization (plot)
    - And, also plotted the anomaly against temperature with the help of visualization (plot)

Reference: I searched the various approaches available on google.com and kaggle.com and took help in coding from the various kernels available on kaggle.com