Remote Machine Diagnostic

PURDUE UNIVERSITY | The Data Mine

JOHN DEERE

Cai Chen, Richard Chen, Pradyumna Danigond, Khan Md Monsurul Islam, Neal Lunsavat, Arnab Mitra, Nikhil Saxena, Varun Sundaram, Aniruddh Suresh, Sahithi Tummala, Saimonish Tunguturu, Yihua Zou

Problem Statement Overview

Problem

Imagine you are driving your car to your house 20 miles away and 10 miles in you get a warning that notifies you that the cruise control function in your car was broken. How would you react compared to if you saw an engine failure warning 10 miles away from your house. Through this project we are looking to uncover customer intent and analyze machine uptime results to better categorize the overall customer impact of DTCs (Diagnostic Trouble Codes).

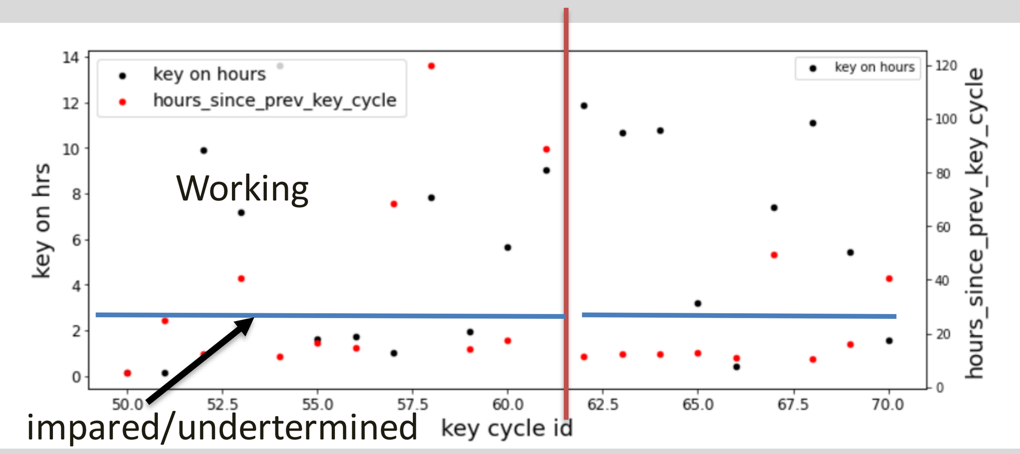
Goal

The goal of our project was to analyze machine uptime results and operational flags using Python by manipulating and labelling data sets of DTCs (Diagnostic Trouble Codes) and creating a training data set to leverage AI and advanced modelling techniques to derive customer intent and to better categorize the overall customer impact of DTCs.

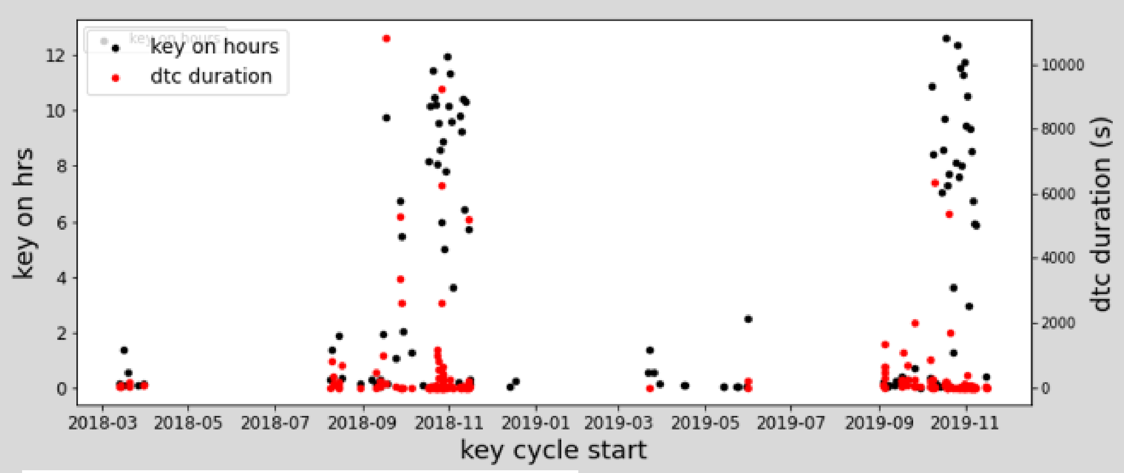
Objectives

- Interactive Visualization of Data using shiny App and Python Dashboard
- Data Cleaning and Merging
- Data Labeling

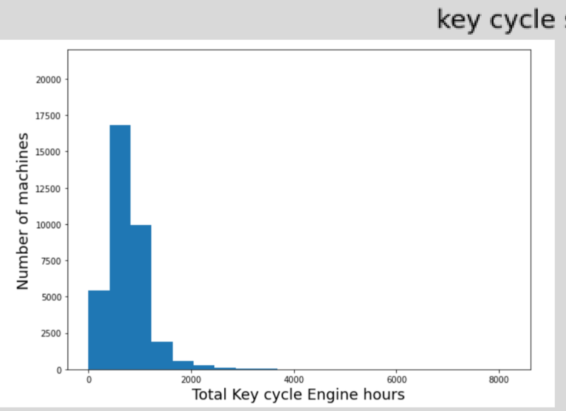
Data Labeling

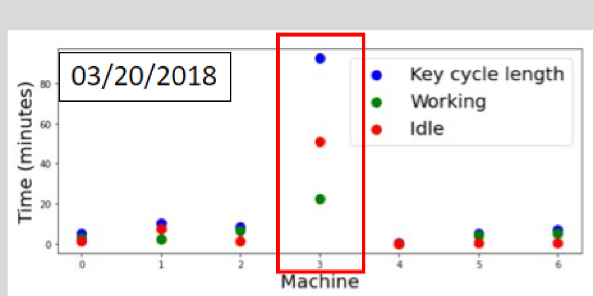


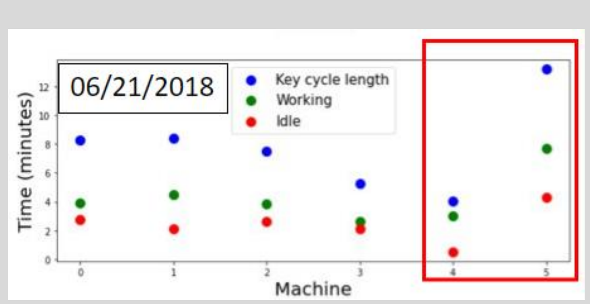
- Measurement data provide information if the key cycle is idle/transport/working.
- Key cycle along with dtc location and measurement data is used to label the key cycle that will be fed into the model



- Filter key cycles by location and date
- Find local abnormal key cycles
- Abnormal key cycles have different duration and composition







Data cleaning and merging

Data Given

Key cycle table

- Total 35000 machine data is provided with two years data.
- Each machine has key cycles , key on hours and key cycle engine hours
- For machine id, long anonymized id is provided

DTC table

- The capture time of dtc, dtc duration is provided
- The dtc is provided with long anonymized id for fmi, spn and tla.

Data cleaning

- Long anonymized id has been converted to shorted id such as tla_37, spn_6
- A function has been generated to move back to the original has id.

```
coded_id_list = {'tla_37','spn_359','fmi_19'}
```

Encrypted dtc → Function to transform any encrypted id to original hash id → Original hash id

Example

```
get_label_encoded_value(dicts,coded_id_list)
```

```
value assigned for spn_359 is 845d5a1e8cf7e84a8e5c9fc1bd4a6490005393b86b44dd2bbf9e8da39b7bd529
value assigned for tla_37 is e8185ca78486dfe6a3683cfe6f1323525db3cd50eab3bd93b7119eb01344f94c
value assigned for fmi_19 is a56ba88e62a90d0c10a27fc28e35c79fbb22d244d24d9471381ff5655c6e2e94
```

Data Given

Measurement table

- Include the state of the machine if it is idle, working or in transport.
- 4 machine ids were given as sample to write a function that can generate data.
- Function has been delivered to team to run for all the key cycle that needs less storage capacity.

Merging

- Key cyle and dtc table has been merged.
- Key cycle and measurement table has been merged

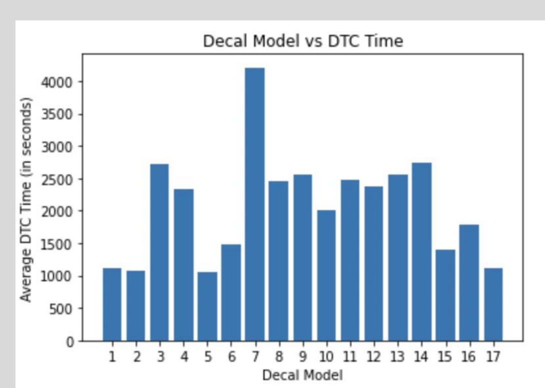
Key cycle table + Dtc table = Merged table ⇒ Dtc capture time is between key cycle start and key cycle end

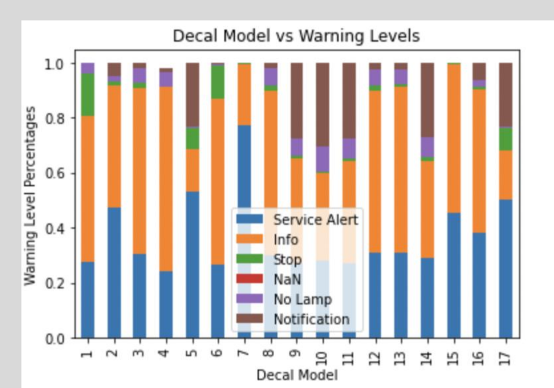
Calculate the time difference between the key cyle end and captured dtc time

Key cycle 1 – key cycle start → Dtc in dashboard → Dtc in dashboard → Key cycle end

Statistical Analysis

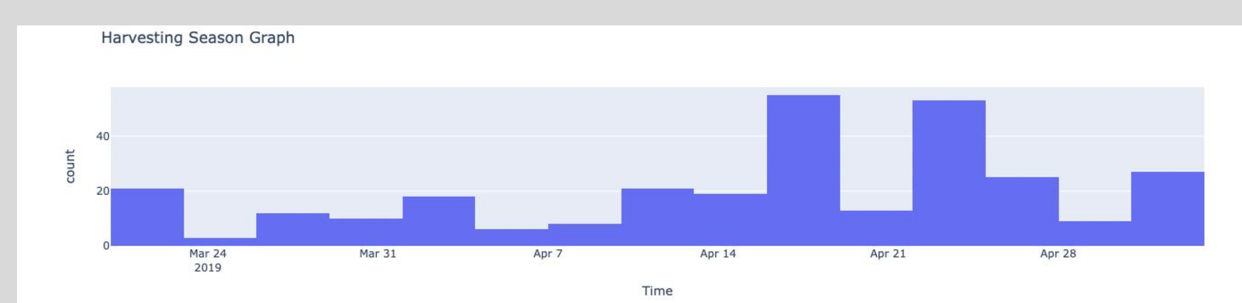
Decal Model Analysis

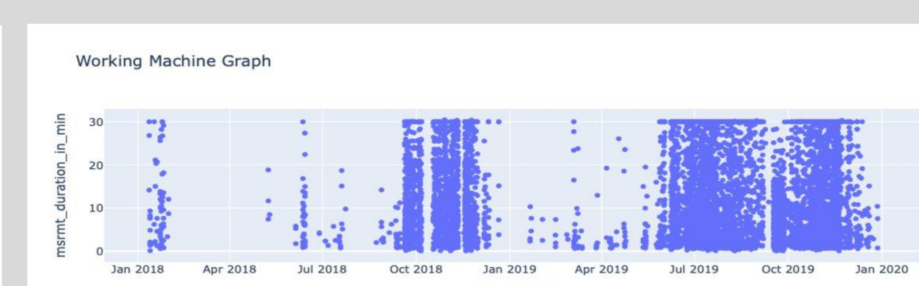




- Does the decal model have an effect on the amount of DTCs and Warning Level alerts a machine will get?
- Decal Models vs DTC Duration (left)
  - calculated the cumulative DTC duration
- Decal Models vs Warning Levels (right)
  - calculated percentages of each warning level
- Decal Model #7 is the one with the highest cumulative DTC duration and the highest number of service alerts.

Harvesting Season Graph





Purdue University | The Data Mine

Remote Machine Diagnostic

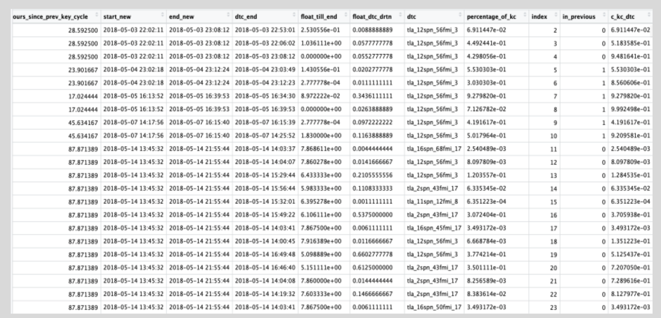
Cai Chen, Richard Chen, Pradyumna Danigond, Khan Md Monsurul Islam, Neal Lunsavat, Arnab Mitra, Nikhil Saxena, Varun Sundaram, Aniruddh Suresh, Sahithi Tummala, Saimonish Tunguturu, Yihua Zou

John Deere

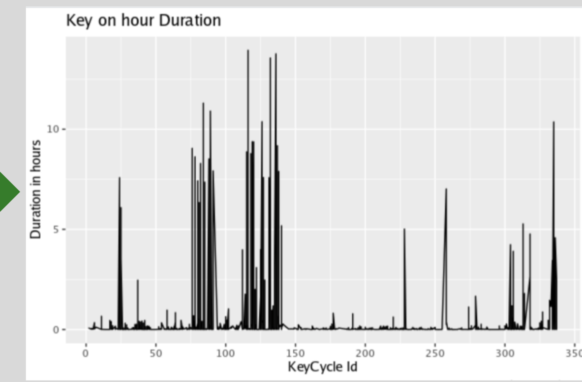
R Shiny App

Raw Data

Cleaned/Sorted Data



Visualization generated

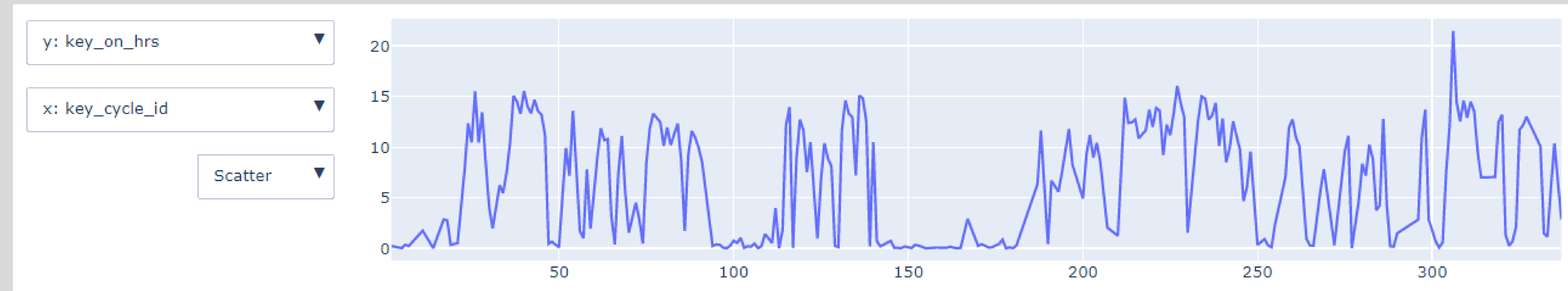


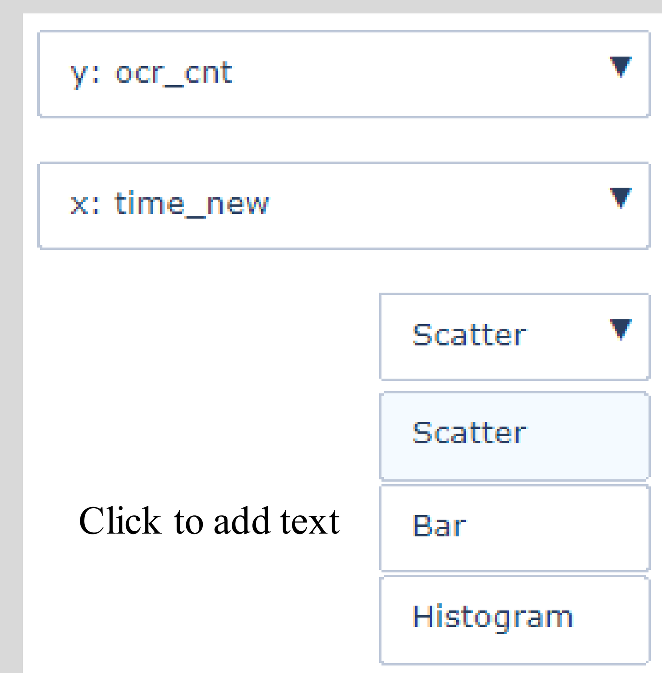
Interactive Shiny App

Purpose

Make Data more interactive/accessible and easier to work with by allowing users to generate graphs and statistics on specific criteria available in the shiny app. In the shiny app to the left, our feature allows you to select a time range and machine number, and it will display DTC duration across cycles with an interactive graph that could be zoomed in for further analysis.

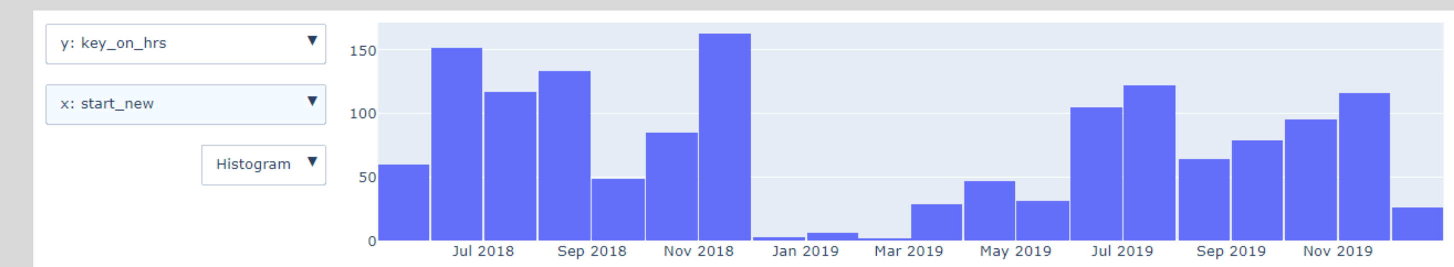
Python Dashboard

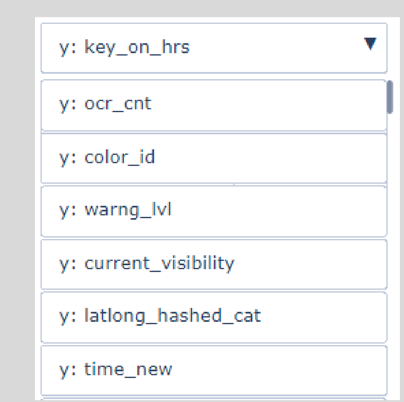




Dashboard Functionalities

The improved-python dashboard enables users and guests to interact with it without having to wait for a significant amount of time. This is achieved by using a sample data (merg.csv) of our original DTC table. On the dashboard, dropdown options are available for both axes, in which each option represents a column within any csv file we input. Graph types are also a usable feature now.





Dashboard Outcome & Application

As shown on the left side, this particular dashboard uses and offers up to 20 variables - his is extremely crucial for making comparisons between our data. The above example selects time as x variable and key-on-hours to demonstrate a pattern throughout a year, where we can see, clearly, when the harvesting seasons may potentially be.

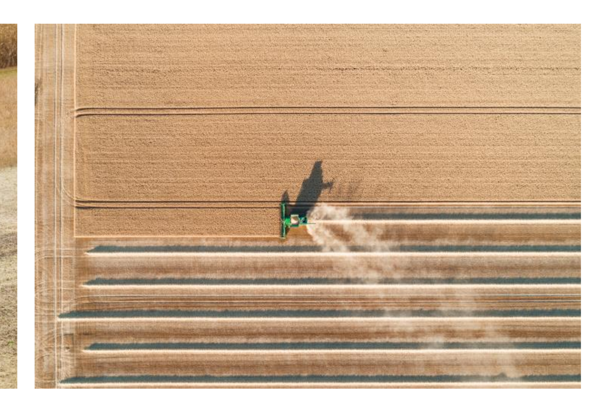
Conclusion

Learning Outcome

During this project, we were able to merge and filter the data. We then did some statistical analysis to find some irregularities within our given dataset. We then created visuals using our key cycles data, which we then used to label the data. We also worked on creating r shiny/python dashboards to make labeling data more efficient. We explored multiple methods of labeling data including using AI to label data and hand labeling data.

Barriers

Due to the complexity of combine key cycles data, it was difficult to definitively determine whether a combine was impaired based on key_cycle data.



Future Goals

1) Improve the existing UI for the dashboard
2) Hand labeling the key-cycles using measurement table and geo-location data
3) Exploring the possibility of using boosting classifier for future key-cycle labeling
4) Analyzing more into the dataset given to us and try to figure out some patterns
5) Further the labeling and exploring the specific decal model vs seasonality
6) Improve dashboard by adding further customization such as color and graph types (e.g. 3D plot).

Reference and Acknowledgments

- We would like to thank::
  - Corporate Partner Mentors: John Andruch, Kylie O'Connor, and Devin Becktell
  - Student Mentor: Cai Chen
  - And the entire John Deere team for giving us this opportunity, letting us access their internal data and systems, and guiding us as we created and refined this project.
- We would also like to thank: Dr. Mark Daniel Ward, David Glass, and Maggie Ann Betz for supporting us during the Academic Year and providing us with the resources for the successful completion of our project.

DASK