

Advanced Detection of Critical Failure Events

Rashmi Ananth, Narmadha Balraj, James Hubbard, Sandra Lee, Justin Mathew, Natalie McGuckin, Elias Winters

Project Description:

- Caterpillar provides customers with condition monitoring services
- Provide actionable service recommendations to the customers
 - Uses machine data: Diagnostic Trouble Codes (DTC).
- Figure 1: DTC Classification Levels

Figure 1

Severity Level 1	Severity Level 2	Severity Level 3
<ul style="list-style-type: none"> Cautions Warnings 	<ul style="list-style-type: none"> Continued machine abuse Harmful for the machine 	<ul style="list-style-type: none"> Critical event Machine shut down

- Goal: Predict critical DTCs ≥ 2 weeks in advance, leveraging all related DTCs with all severities

Figure 2



Our Data is From The CAT 777G Off-Highway Truck

About the Data

- A set of 'Severity Level 3' DTCs are selected for a particular Caterpillar quarry truck sales model and many more DTCs are used as inputs.
- Our group used data of DTCs history for a group of quarry trucks as seen in the visual below.

Figure 3

candidate	date	hours	event	occur_count	svrty_level	
0	7TG10100	17251	22.0	Machine_SensorJ_exceededLimit	1	1
1	7TG10100	17252	43.0	Comp06B_discreteSensorH_Up	1	2
2	7TG10100	17253	65.0	Comp06B_discreteSensorH_Up	1	2
3	7TG10100	17253	65.0	Machine_SensorJ_exceededLimit	2	1
4	7TG10100	17253	65.0	Machine_SensorJ_exceededLimit	1	1

Figure 3 (above): This data shows different severity levels and time stamps for each incident

Exploratory Data Analysis:

- There are 4 critical event tags in the data that we used to solve our problem.
- The information we focused on included the date, number of occurrences, severity level, and service meter hours.
- Figure 4: Visualize all components and look for correlations

Figure 4

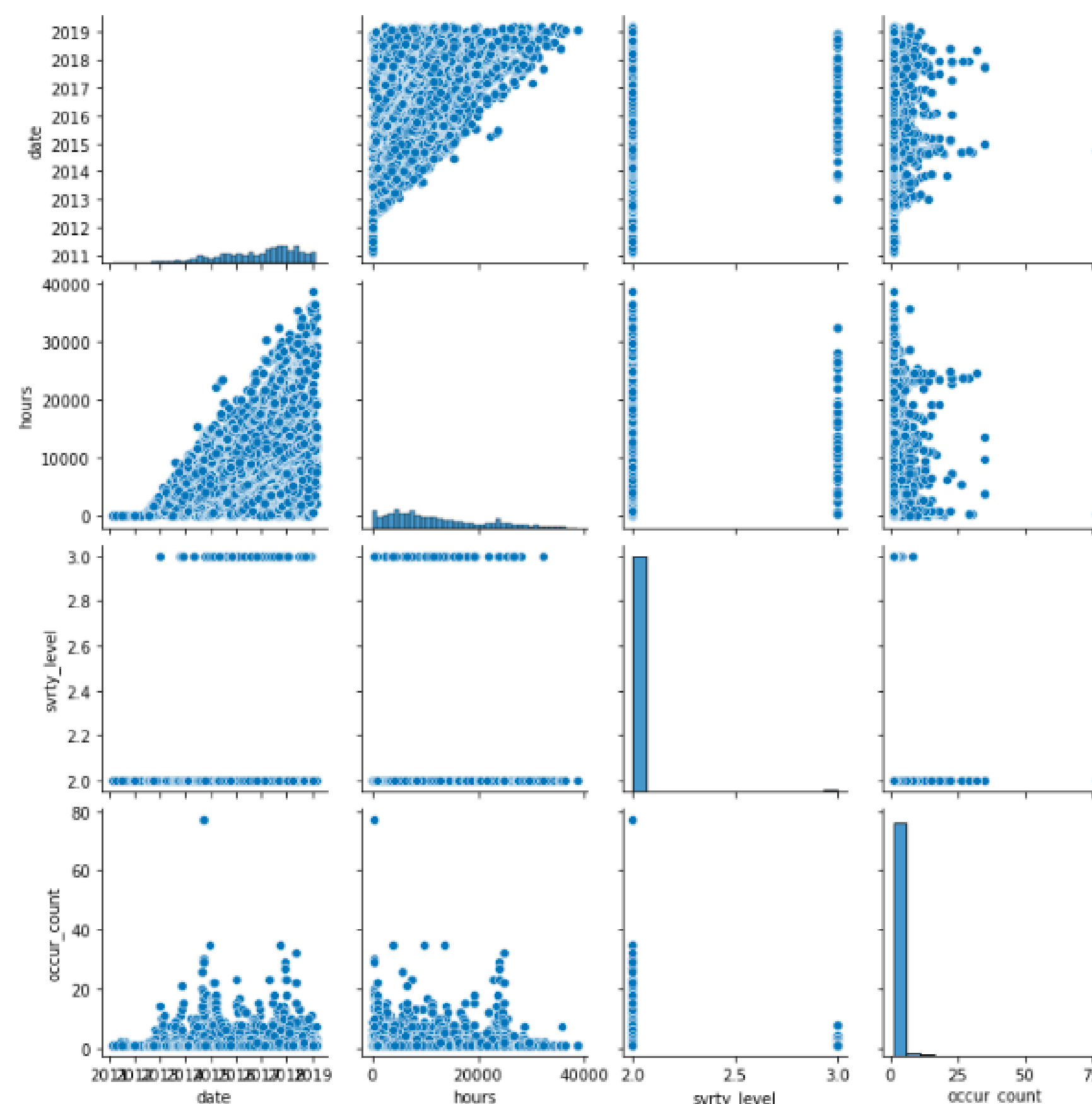
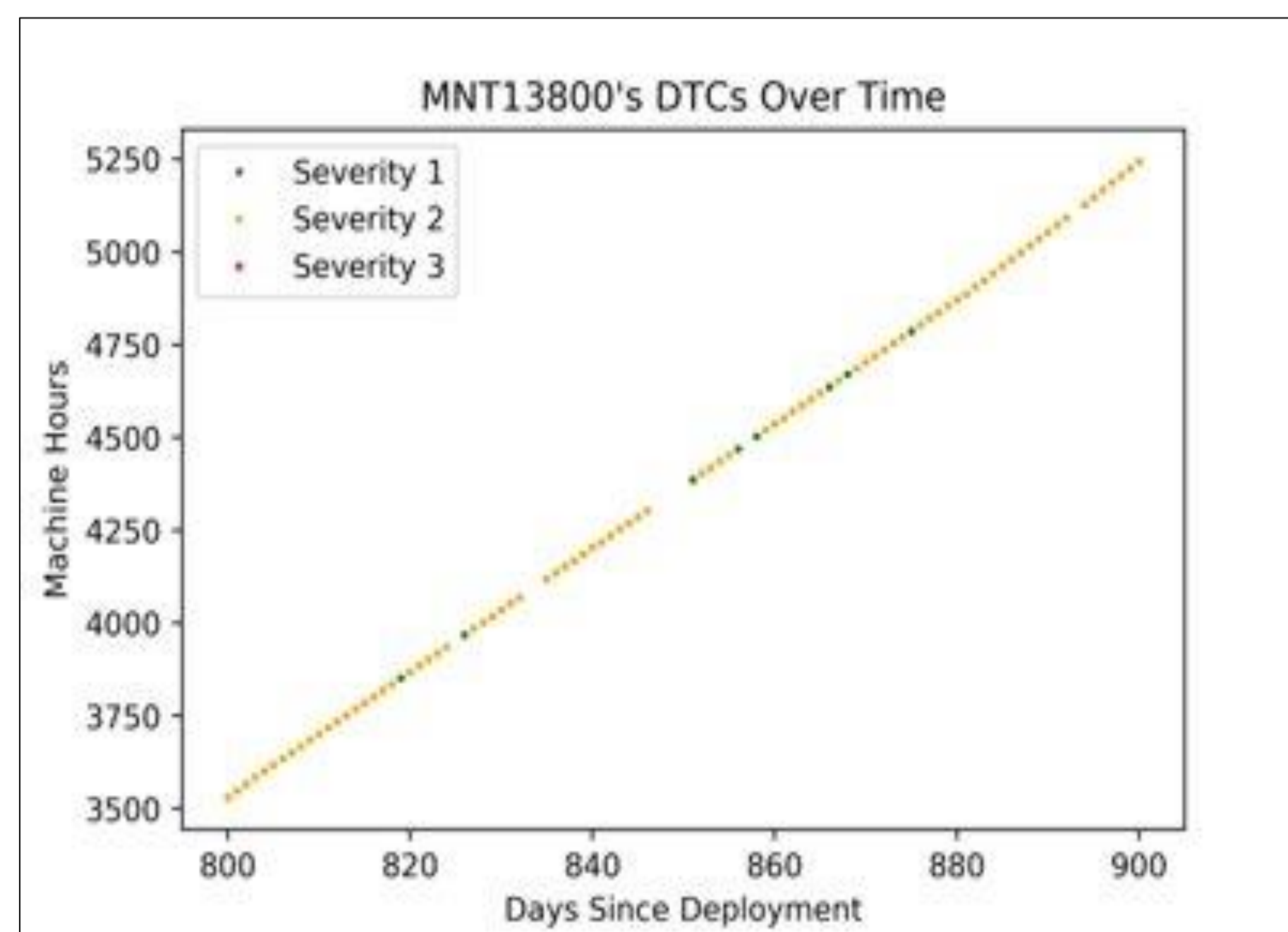


Figure 5



- Chains of severity level 1 and 2 events often precede severity level 3 events, followed by long gaps.
- Severity level and occurrence count are less relevant than the actual types of events occurring and their relationships with one another in time.

Potential Model

- In order to pick out significant events related to the critical failures we have been asked to predict, our team plans to make use of some form the TF*IDF metric.
- TF*IDF is a method for weighting the items in a dataset based on their relative rarity.
- Our goal is to compare the behavior of a single candidate to the behavior of candidates in general.
- It's possible that several of these weights could be calculated across different scopes to improve specificity and noise-resistance.
- The figures below show TF*IDF formulas.

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

Above: The numerator is the raw frequency of an event in a given candidate's event history, and the denominator is the sum of the frequencies of all other events in that candidate's history.

Below: The numerator is the total number of event histories, and the denominator is the number of event histories in which a given event occurs

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

Future Goals/Conclusion

- In the future, we want to create a system that can perform similar analysis to other datasets.
- The next steps to reach this goal include:
 - Choosing a model
 - Developing and train the model
 - Test and evaluate the model's performance on testing datasets
 - Integrate the finalized model into a system which can easily be applied to other data sets.
- We hope to eventually be able to predict up to 1 month before a level 3 Diagnostic Trouble Code appears.

Acknowledgements

- Thank you to our mentors Chris Ha, Sridhar Ramaswamy, Vijay Yalamanchili and Harsheel Shah from CAT Digital
- Thank you to the Data Mine Faculty and Staff