It Burns When I Decision Tree

Kevin Fagan

Introduction

- **Business problem:**
  - Finding fires in an petrochemical plant while they are in the incipient stage, before they become a bigger problem.
- **Goal**
  - Determine if fire can be detected with existing infrastructure for Monroe Energy's Trainer Refinery.

Data

- **A dataset of over 62M smoke detector sensor readings was obtained from Kaggle.**
- **The dataset was generated by Stefan Blattmann for his IOT project Real-time Smoke Detection with AI-based Sensor Fusion.**
- **Many different environments and fire sources have to be sampled to ensure a good dataset for training. A short list of different scenarios which are captured:**
  - Normal indoor
  - Normal outdoor
  - Indoor wood fire, firefighter training area
  - Indoor gas fire, firefighter training area
  - Outdoor wood, coal, and gas grill
  - Outdoor high humidity
  - etc.

Data

- **The dataset features in detail:**
  - Air Temperature
  - Air Humidity
  - TVOC: Total Volatile Organic Compounds; measured in parts per billion (Source)
  - eCO2: co2 equivalent concentration; calculated from different values like TVCO
  - Raw H2: raw molecular hydrogen; not compensated (Bias, temperature, etc.)
  - Raw Ethanol: raw ethanol gas (Source)
  - Air Pressure
  - PM 1.0 and PM 2.5: particulate matter size < 1.0 µm (PM1.0). 1.0 µm < 2.5 µm (PM2.5)
  - Fire Alarm: ground truth is "1" if a fire is there
  - CNT: Sample counter
  - UTC: Timestamp UTC seconds
  - NC0.5/NC1.0 and NC2.5: Number concentration of particulate matter. This differs from PM because NC gives the actual number of particles in the air. The raw NC is also classified by the particle size: < 0.5 µm (NC0.5); 0.5 µm < 1.0 µm (NC1.0); 1.0 µm < 2.5 µm (NC2.5);

Data

- **The dataset features in detail:**
  - Air Temperature
  - Air Humidity
  - TVOC: Total Volatile Organic Compounds; measured in parts per billion (Source)
  - eCO2: co2 equivalent concentration; calculated from different values like TVCO
  - Raw H2: raw molecular hydrogen; not compensated (Bias, temperature, etc.)
  - Raw Ethanol: raw ethanol gas (Source)
  - Air Pressure
  - PM 1.0 and PM 2.5: particulate matter size < 1.0 µm (PM1.0). 1.0 µm < 2.5 µm (PM2.5)
  - Fire Alarm: ground truth is "1" if a fire is there Target
  - CNT: Sample counter Dropped, unnecessary
  - UTC: Timestamp UTC seconds Dropped, unnecessary
  - NC0.5/NC1.0 and NC2.5: Number concentration of particulate matter. This differs from PM because NC gives the actual number of particles in the air. The raw NC is also classified by the particle size: < 0.5 µm (NC0.5); 0.5 µm < 1.0 µm (NC1.0); 1.0 µm < 2.5 µm (NC2.5);

Method

- The data was used to create classification models (declare fire based on sensor data) to predict if there is fire present
- Models were then analyzed by Recall score, as correctly classifying actual fire is more important than false alarms
- The best performing model was selected and run using only the data that an oil refinery would have available to it
- Business recommendations will be made based on that

Models

- Logistic Regression gave us an accuracy of 89% and a recall of 94.6% after tuning.

True Negatives: 2726	False Positives: 849
False Negatives: 482	True Positives: 8469

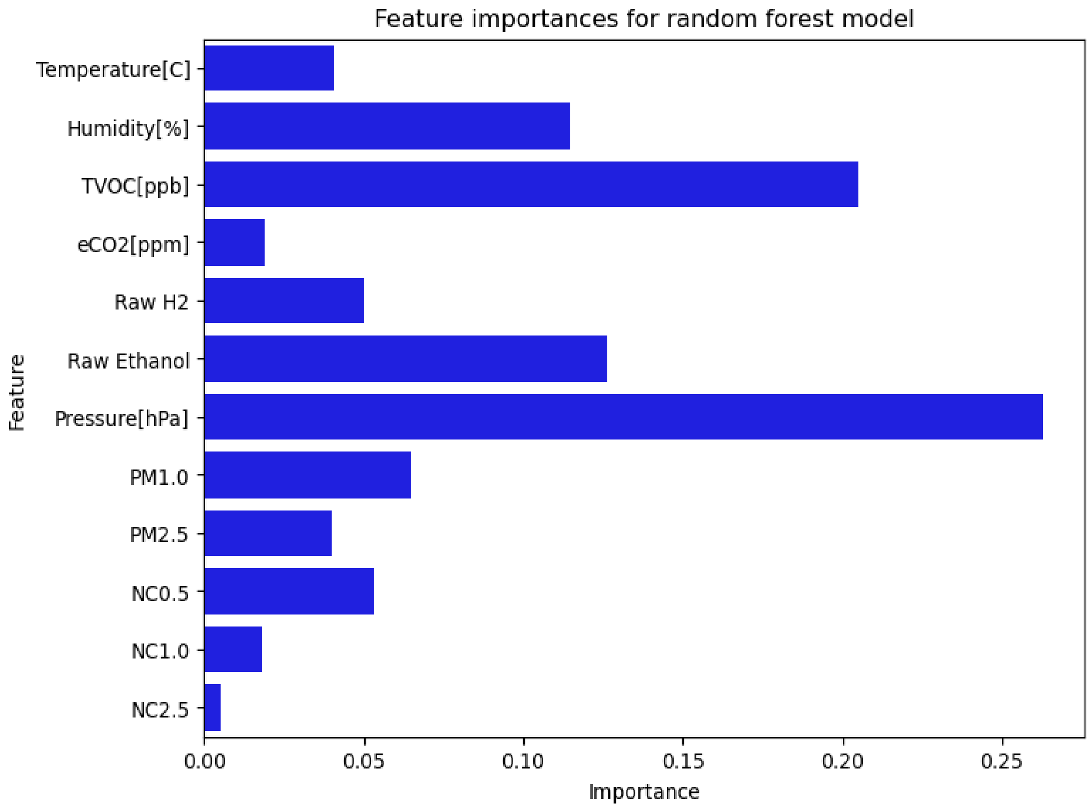
- A Decision Tree gave us 100% accuracy and recall.

True Negatives: 3575	False Positives: 0
False Negatives: 0	True Positives: 8951

Random Forest

- A Random Forest model also gave us 100% accuracy and recall.

True Negatives: 3575	False Positives: 0
False Negatives: 0	True Positives: 8951



Select features

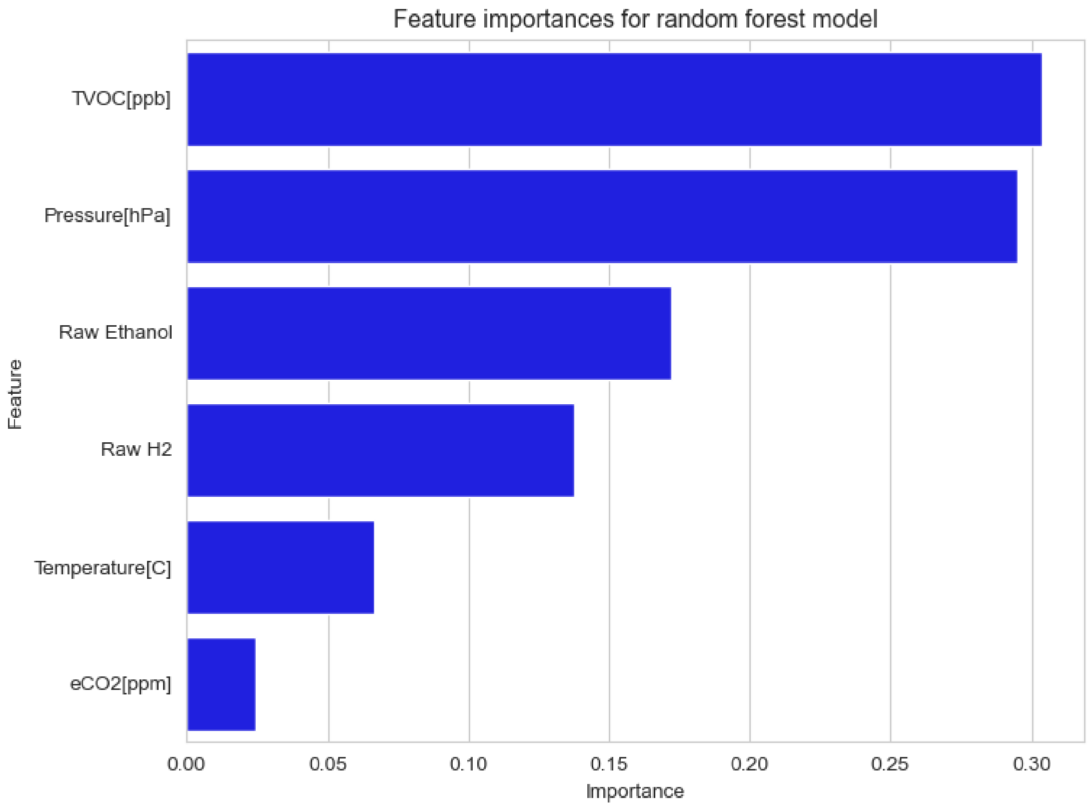
- Now we know smoke detectors, which our refinery does not have, can detect fire. Lets try this again with instrumentation we do have.

True Negatives: 3575	False Positives: 0
False Negatives: 0	True Positives: 8951

- Good news. This includes ambient temperature, VOCs, CO2, H2, Ethanol, and ambient pressure, for which there are standing monitors around the refinery.

Instruments and analyzers

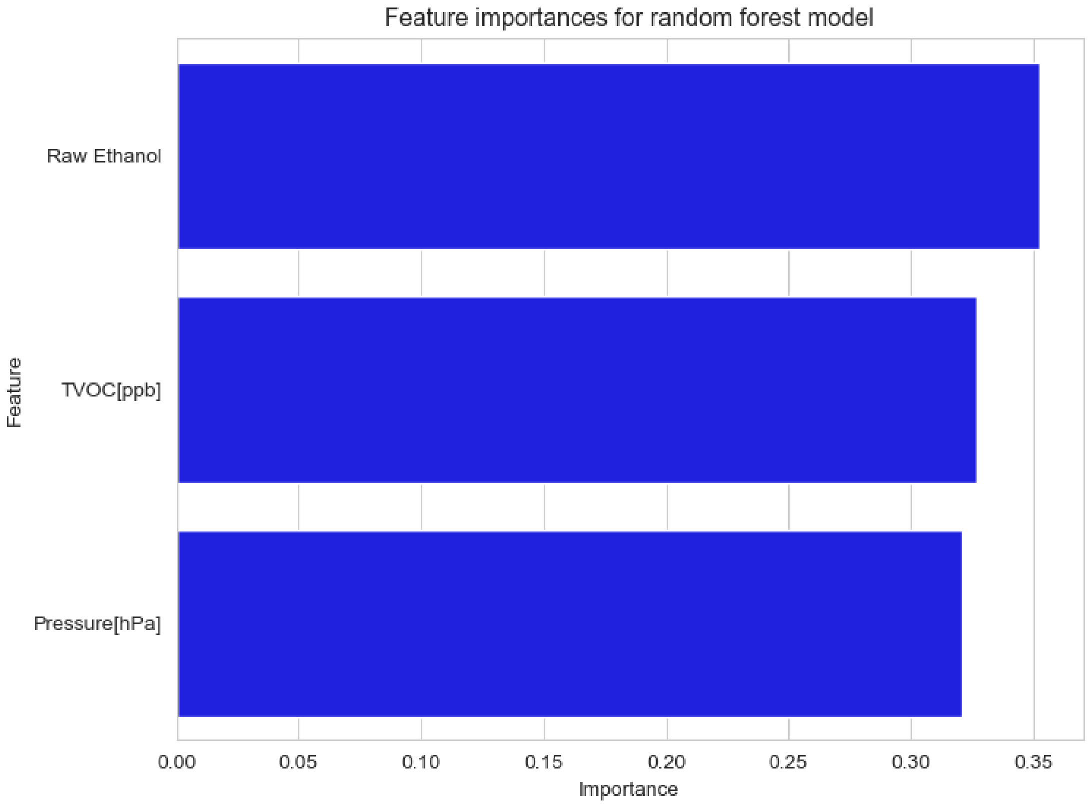
- These exist in varying frequencies throughout the refinery.
- Ambient pressure can also be inferred from process monitoring instruments by detecting changes in groups of physically close points.



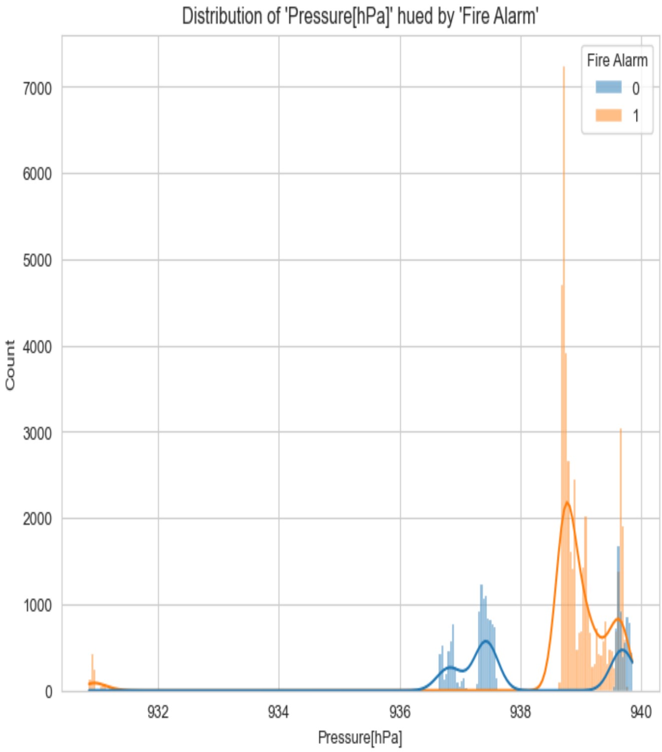
Even less equipment

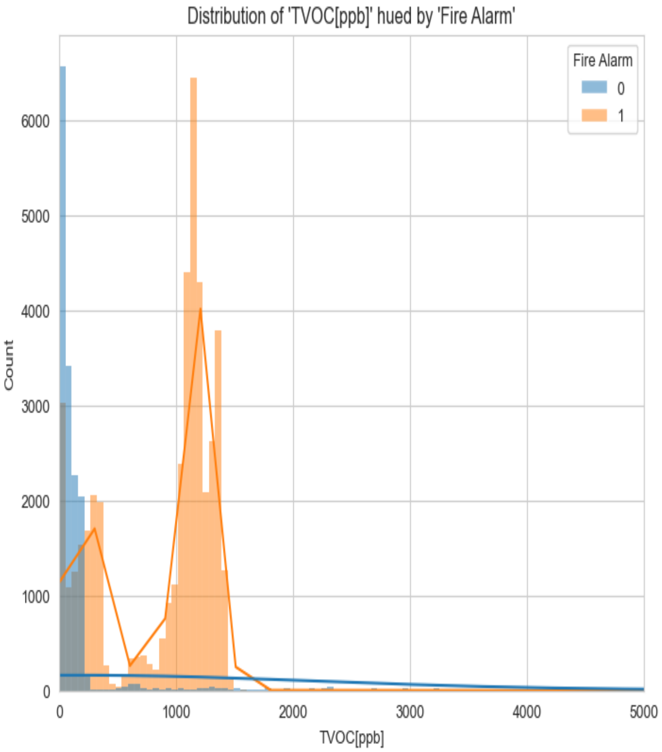
- All possible combinations of available existing equipment were used to predict fire, and ranked by recall score.
- We get acceptable results with only three sensors. Recall of 99.99%

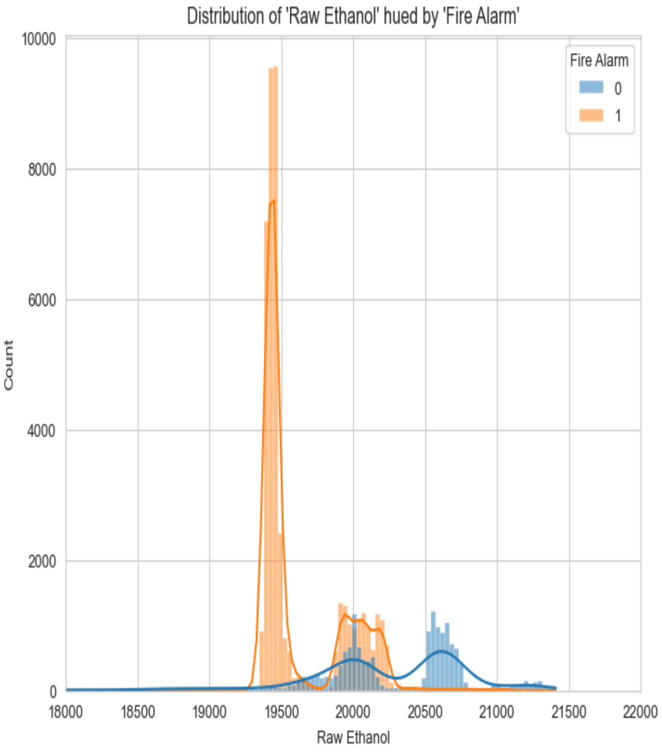
True Negatives: 3572	False Positives: 3
False Negatives: 1	True Positives: 8950



Distributions







Recommendations and Next Steps

- It is clear that we can leverage existing equipment to improve the refineries ability to detect fire at the incipient stage.
- We should evaluate the physical location of the available instruments and analyzers to find gaps in protection.
- Then, more analyzers can be added at a lower cost than installing smoke detectors (which would include the cost of supporting and training personnel on new equipment). It is difficult to estimate the cost of a dedicated fire protection system, but as an example, a small tank farm (24 tanks) in Bavaria estimated that it would cost them approximately $1.3MM.

Recommendations and Next Steps

- **Next steps:**
  - Ethanol detection was used in this case because in the real world testing that created this dataset, ethanol was used as the Class B fuel. LEL detection would be more versatile and should swap right in, but real world testing should be conducted to confirm.
  - If possible, find data regarding the physical distance limits of these readings, to plan and optimize the density of our instruments.

Questions?

Kevin Fagan

https://www.linkedin.com/in/kevin-m-fagan/
kevinmfagan@gmail.com