False Alarm Detection

Problem Statement-This project was made for a chemical industry which had sensors installed in various parts of the factory to detect H2S gas which is hazardous to health. Every time one or multiple sensors detected the H2S leak, an emergency alarm rings to alert the workers. For every alarm, the industry calls a team which sanitizes the place and checks for the leak and this was a big cost to the company.

A few of the alarms that ring are not even hazardous. The company gave us the data for each alarm with a final column stating the alarm was dangerous or not.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ambient Temperature** | **Calibration(days)** | **Unwanted substance deposition (0/1)** | **Humidity (%)** | **H2S Content(ppm)** | **Dangerous (0/1)** |

The data was first pre-processed and analysis libraries like Numpy and Pandas were used to make it ready to be utilized by a machine learning algorithm.

Problems like standard scaling, categorial data and missing values were handled with appropriate techniques.

Then, we used Naïve Bayes model to make a classifier with first five column as independent columns and dangerous column as dependent/target column.

Now whenever, there is a leakage and the alarm rings, the data is sent to us and we predict if it is dangerous or not. If found dangerous the only the team is called to sanitize the place and fix the leak. This saved a lot of money for the company.

Form extraction service

We were given structured forms from a company and they had various attributes to be extracted like insurance company name, insured’s name, Total amount etc.

We used Open CV to first crop out various parts of the form. One major challenge was that an entity title was written in different ways like insurance company name was written in one form as Company name, in the other form as Insurance provider etc. Hence, every attribute title has various names in various forms.

We trained a machine learning classification model to get a field, applied NLP techniques to do stemming and remove stop words to identify what does the title say and then extract the value of the corresponding attribute.

For each form we made a data frame of title and values and exported that as an excel sheet.