**Model Card**

**Model Description**

**Input features:**

* UTC
* Temperature
* Humidity
* TVOC
* eCO2
* Raw\_H2
* Raw\_Ethanol
* Pressure
* PM1\_0
* PM2\_5
* NC0\_5
* NC1\_0
* NC2\_5
* CNT

**Output predictor:**

* Fire\_Alarm

**Model Architecture:**

The models used come from sklearn.

|  |  |
| --- | --- |
| **Model** | **Best Performance** |
| LogisticRegression() | 'C': 100, 'penalty': 'l2', 'solver': 'liblinear'. |
| DecisionTreeClassifier() | Default settings |
| RandomForestClassifier() | Default setting |
| SVC(kernel='rbf') | Default settings |
| GradientBoostingClassifier() | Default settings |

**Performance**

I have trained several models which are summarised in the table – below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Best Accuracy** | **Best Recall** | **Best Precision** |
| Logistic Regression | 0.991218266006706 | 0.994961370507222 | 0.9927382415372584 |
| Decision Tree Classifier | 0.9999201660546064 | 1.0 | 0.9998880429914913 |
| Random Forest Classifier | 0.9998403321092129 | 0.9997760609114321 | 1.0 |
| Support Vector Machine | 0.9993613284368513 | 0.9998880304557161 | 0.9992167393980083 |
| Gradient Boosting | 0.9999201660546064 | 0.9998880304557161 | 1.0 |

The models are trained using a training dataset that has been split from the testing dataset prior to any pre-processing. Three performance metrics are given, using *sklearn import metrics*. All performance metrics are derived from the test dataset, as follows:

*metrics.accuracy\_score(y\_test, y\_pred)*

*metrics.recall\_score(y\_test, y\_pred)*

*metrics.precision\_score(y\_test, y\_pred)*

**Limitations**

The model has been trained and tested on a small dataset that was provided by an Arduino hacker as part of a DIY project. It should not be used in place of an approved smoke alarm.

**Trade-offs**

The dataset is for educational purposes only