**Datasheet**

**Motivation**

Smoke detectors save a lot of lives. For example, the number of fire victims fell by more than 48% in France from 1982 to 2012 and 56% in the UK from 1982 to 2013. These reductions can largely be linked to increased fire safety regulations and smoke detectors ([Source](https://www.modernbuildingalliance.eu/fire-safety-statistics/)). In the U.S. 96% of all homes have smoke alarms and approximately 20% of homes with smoke alarms have non-operational smoke alarms. It is estimated that if every home had working smoke alarms, U.S. residential fire deaths could drop by 36%, with nearly 1100 lives saved per year ([Source](https://www.nist.gov/el/smoke-alarm-research#:~:text=Smoke%20alarms%20are%20now%20installed,1100%20lives%20saved%20per%20year.))). With an increasing number of smoke detectors, false alarms became a problem ([Source](https://www.bafe.org.uk/bafe-news/home-office-statistics-highlight-fire-false-alarms-remain-an-issue)). The number of false fire alarms is increasing continuously, which is a severe issue for firefighters.

The dataset was created by a DIY hacker by the name of Stefan Blattmann. Details can be found on the website https://www.hackster.io/stefanblattmann

**Composition**

The dataset (smoke\_detection\_iot.csv) was downloaded from <https://www.kaggle.com/code/dhavalrupapara/smoke-detection-using-machine-learning>

and contains data captured from the sensors onboard an Arduino Nicla Sense microcontroller.

The data captured is composed of the following features:

*Air temperature*

* *TVOC: Total Volatile Organic Compounds (ppb)*
* *eCO2: Equivalent CO2 concentration*
* *Raw H2: Raw molecular hydrogen – not compensated*
* *Raw Ethanol*
* *Air pressure*
* *PM1.0: Particulate matter size < 1 micron*
* *PM2.5: Particulate matter size < 2.5 microns*
* *CNT: Sample counter*
* *UTC: Time stamp*
* *NC0.5: Number concentration of particulate matter (actual number of particles <0.5 micron)*
* *NC1.0: Number concentration of particulate matter (actual number of particles <1 micron)*
* *NC2.5: Number concentration of particulate matter (actual number of particles <2.5 micron)*
* Fire Alarm: Ground truth is ‘1’ if a fire is present

**Collection process**

The data was collected by the dataset’s author and was made available in 2022.

60,000 samples were collected from various on-board sensors at a frequency of 1 Hz (one sample per second).

Different environment and fire scenarios were sampled to provide a good dataset for training ML models, specifically:

* 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

A time stamp was added to every sensor reading.

**Preprocessing/cleaning/labelling**

These details are not given by the author; however, it is assumed that the data in the datasets has not been cleaned or pre-processed in any way.

**Uses**

The dataset and project hardware details are provided on the author’s hacker.io website and is therefore made available for fellow hackers to use.

**Distribution**

No specific distribution details are given.

**Maintenance**

No information is provided regarding maintenance of the dataset. It appears to be a one-off dataset that was used to test the feasibility of a sensor hardware project aimed at reducing false fire alarms.