**A Modular Ice Cream Factory Dataset on Anomalies in Sensors to Support Machine Learning Research in Manufacturing Systems**

**Alternate Title:**

An Ice Cream Factory Anomaly Detection using MIDAS (Modular Ice cream factory Dataset on Anomalies in Sensors) in ML

**Aim**:

The aim of creating a modular ice cream factory dataset on anomalies in sensors to support machine learning research in manufacturing systems is to provide a valuable resource for the development and evaluation of machine learning models and anomaly detection algorithms in the context of manufacturing.

**Abstract:**

A small deviation in manufacturing systems can cause huge economic losses, and all components and sensors in the system must be continuously monitored to provide an immediate response. The usual industrial practice is rather simplistic based on brute force checking of limited set of parameters often with pessimistic pre-defined bounds. The usage of appropriate machine learning techniques can be very valuable in this context to narrow down the set of parameters to monitor, define more refined bounds, and forecast impending issues. One of the factors hampering progress in this field is the lack of datasets that can realistically mimic the behaviors of manufacturing systems. In this paper, we propose a new dataset called MIDAS (Modular Ice cream factory Dataset on Anomalies in Sensors) to support machine learning research in analog sensor data. MIDAS is created using a modular manufacturing simulation environment that simulates the ice cream-making process. Using MIDAS, we evaluated three different supervised machine learning algorithms (Logistic Regression, Decision Tree and Random Forest) for two different problems: anomaly detection and anomaly classification. The results showed that Decision Tree is the most suitable algorithm with respect to model accuracy and execution time. We collect the Modular Ice Cream Factory Dataset from the publicly available, to enable interested researchers to enhance the state of the art by conducting further studies

**Introduction:**

 In a world where most of the commodities are mass produced by companies using automated manufacturing systems, the quality of those systems is of vital importance. Even a small deviation in parts of the system could potentially result in bad or malfunctioning products leading to customer dissatisfaction, environmental impacts or huge economic losses to the industry. This is the main reason why all components and sensors in the system have to be continuously monitored to identify anomalies, and prompt remedial actions should be provided if something goes wrong.

In a generic sense, an ‘anomaly’ is a deviation from expected behavior, and can occur for different reasons, including faults in the system or its configuration, or due to unanticipated external interference. Such interference, or even some system or configuration faults, belong to the realm of cyber security threats if the root cause is an act of bad intention. Regardless of the cause, the consequences of anomalies must be kept at acceptable levels

**Existing System:**

This research aims to create an effective prediction model using different types of ML methods to detect anomalies in Ice cream factory. First of all the datasets are collected, and then the preprocessing is accomplished via the missing values imputation. The Mean Value Imputation (MVI) method is used to impute the missing values of the dataset. Then, the categorical feature values are converted to their equivalent numerical values using the One Hot Encoding (OHE) technique. Shows that all datasets used in this work have a object and numerical features as converted into numerical features is used to alleviate this issue. After completing the initial preprocessing, the datasets feature values are scaled using different Feature Scaling techniques. The result showed that Multilayer Perceptron is the most suitable with respect to accuracy. But the accuracy is less. Then now we create a new system for better anomaly prediction. So now we move on to the proposed system.

**Proposed System**:

This research aims to create an effective prediction model using different types of ML methods to detect anomalies in Ice cream factory. First of all, the datasets are collected, and then the preprocessing is accomplished via the missing values imputation and Create an instance of RFE with the classifier and the desired number of features to select Logistic Regression classification of modeling and performance evaluation. Then we are using Decision Tree, Logistic regression and Random Forest Algorithm for prediction accuracy. Decision Tree gives a best results with respect to high accuracy. Compare to existing system our new system gives results are more Accuracy.

**Module Description:**

* Data Pre-Processing
* Algorithm Implementation
* Prediction

**Data Pre-Processing:**

We collect the Modular Ice cream factory Dataset on Anomalies in Sensors from the publicly available repositories: Kaggle and Github. Normal, Freeze, Ramp and Step these are the data sets that classify the three different anomalies from the sensors. The preprocessing is accomplished via the missing values imputation, Then, the categorical feature values are converted to their equivalent numerical values using the One Hot Encoding (OHE) technique. In the dataset there is unwanted columns are presented, these columns are removed and get a neat and clean dataset to predict more accuracy.

**Algorithm Implementation:**

 The Classification Algorithms to produce the best results. We are using Decision Tree, Logistic regression and Random Forest Algorithm to predict anomalies by using MIDAS dataset using ML. On an analysis conducted within various algorithms, the Decision Tree was found to provide highest efficiency. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

* Decision Tree
* Logistic Regression
* Random Forest

**Selected Algorithm:** Decision Tree

 Decision Trees (DTs) are a non-parametric supervised learning method used for [classification](https://scikit-learn.org/stable/modules/tree.html#tree-classification). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

For instance, in the example below, decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model

**Prediction:**

The anomalies detections are classified in to four types that’s are Normal, Freeze, Ramp and Step. Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model. Preprocessed data are trained and input given by the user goes to the trained dataset and predict the results that the accurate classification.

**Software Requirements:**

* Operating System : Windows 10 (64 bit)
* Software : Python 3.7
* Tools : Anaconda (Jupyter Note Book IDE)

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Working Diagram:**

****

**Architecture Diagram:**

 Freeze

(Ageing Cooling Temperature)

Architecture Diagram:

Dataset

Preprocessing

User Input

Trained model

Machine Learning

Anomaly

YES

 NO

Normal

 Step

 (Hardening/Temperature)

 Ramp

 (Ageing Cooling Level)

 Freeze

(Ageing Cooling Temperature)