Machine Failure Prediction

# Project Overview:

The given dataset contains sensor data collected from various machines. The data includes a variety of sensor readings as well as the recorded machine failures. The aim of the project is to predict machine failures in advance.

# Objectives:

* Exploratory Data Analysis
* Compare various models to find the one with highest accuracy
* Predict random user input

# Requirements/Task(s):

Python v3.10+

Sklearn

Catboost

Numpy

Pandas

Matplotlib

Seaborn

# Steps taken

* Check data for outliers and null values and handle them
* Perform Exploratory data analysis
* Select necessary features
* Train different models and the models
* Find the model with highest accuracy
* Predict custom user input

# Results:

|  |  |  |
| --- | --- | --- |
| Model | Train data accuracy | Test data accuracy |
| Logistic Regression | 0.9125827814569536 | 0.9206349206349206 |
| Decision Tree Classifier | 0.9880794701986755 | 0.8994708994708994 |
| Random Forest Classifier | 0.9880794701986755 | 0.8994708994708994 |
| KNeighbours Classifier | 0.919205298013245 | 0.91005291005291 |
| Cat Boost Classifier | 0.9549668874172186 | 0.8994708994708994 |
| Gaussian Naive Bayes | 0.9125827814569536 | 0.9153439153439153 |
| Support Vector Classifier | 0.9125827814569536 | 0.9259259259259259 |

Support Vector Classfier was used to predict user input data

# 

# Summary/ Outcome

* Exploratory data analysis was performed
* Models were trained, tested and compared
* Support vector classifier was used to predict the user input data since it gave the highest accuracy.
* User input was predicted

# Link to Project

<https://github.com/KaustubhaRam/machine_failure_prediction.git>