
LOVELY PROFESSIONAL UNIVERSITY

**Name –** Yuvraj SIngh

**Regd_no – 11804228**

**Section – KM007**

**Dat-set –**

**Annealing**

Lovely Professional University
Phagwara, India

# Introduction:-

In metallurgy and materials science, annealing is a heat treatment that alters the physical and sometimes chemical properties of a material to increase its ductility and reduce its hardness, making it more workable.

**Problem statement**:

I have been provided with an Excel dataset . My task is to analyse the dataset and predictthe class of the Coil

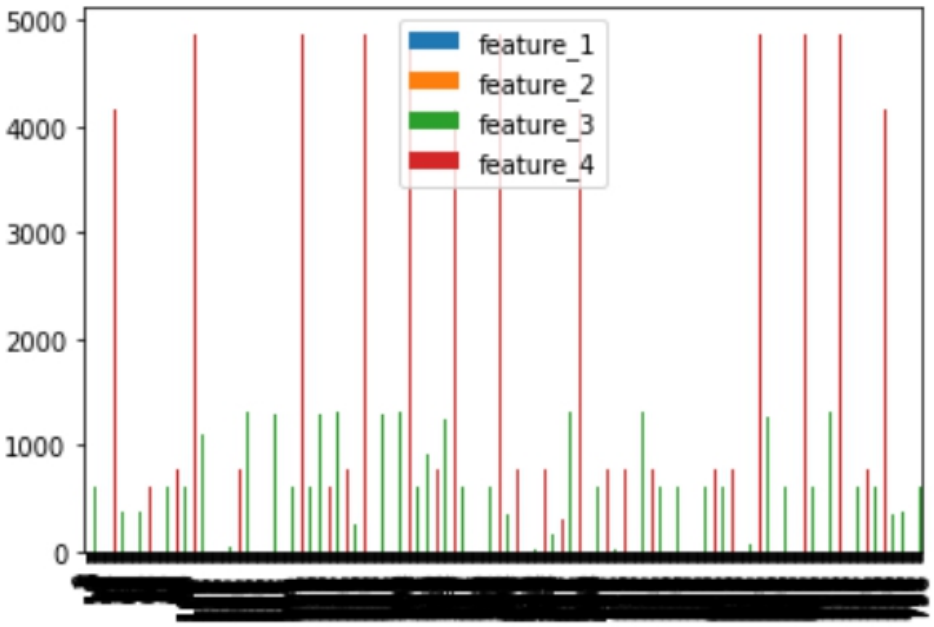
# Cleaning and feature selection of Data-set

Firstly I started with renaming the columns with proper names after that I checked for null values in the data-set. Since there is no null values, I checked for special characters and I found that in some of the columns. After that I assigned NaN to all the special characters and later on I dropped them from the datset.

# Data Visualization

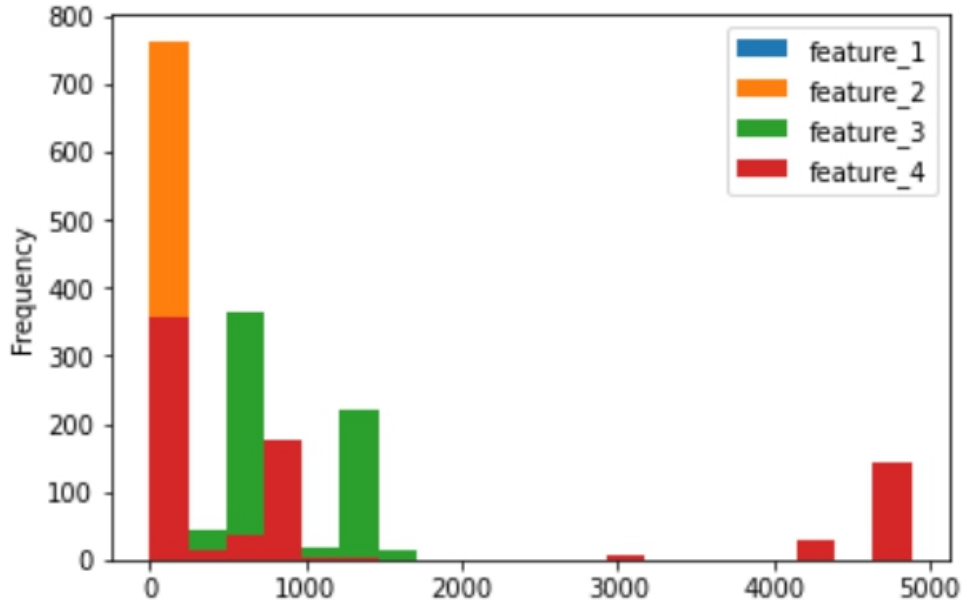
```
df.plot.bar()
```

```
<AxesSubplot:>
```



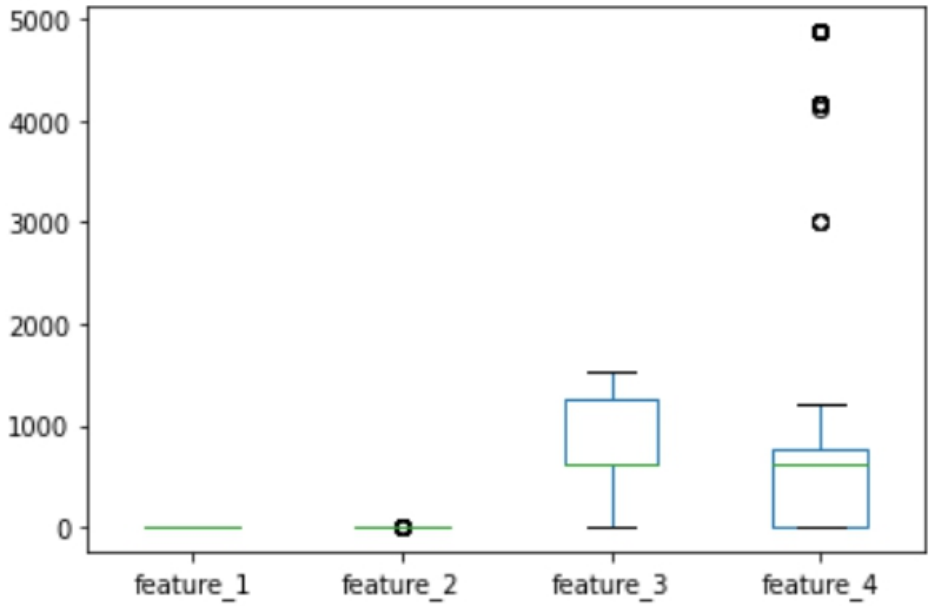
```
df.plot.hist(bins=20)
```

```
<AxesSubplot:ylabel='Frequency'>
```



```
df.plot.box()
```

```
<AxesSubplot:>
```




## Splitting Data and Model selection

After identifying feature that to be trained and tested I have splitted the data in the manner that 70% to be trained and 30% to be tested.

### Model Selection

After splitting the data I have to select the model. I have used three different model. And those are:-

1. Random Forest Classifier
2. Logistic Regression
3. SVM

### 1. Random Forest Classifier

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an esemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction

I got the accuracy above 85%. In this model.

### 2. Logistic Regression

Logistic Regression is one of the easiest and most commonly used supervised Machine learning algorithms for categorical classification. The basic fundamental concepts of Logistic Regression are easy to understand and can be used as a baseline algorithm for any binary (0 or 1) classification problem.

It is a Statistical predicting model that can predict either a 'Yes'(1) or 'No'(0).

I got the accuracy above 75%. In this model.

### 3. SVM

I got the accuracy above 75%. In this model.

### Result

After Traning and testing the data-set on three different models we got accuracy 82%, 75% and 71% respectively. Hence, our first model Random forest classifier is best for our data-set.

## Prediction on Unknown data

After Doing all the things from scratch , I have tested My best model which is Random Forest Classifier on unknown data. In this I will take data from user,

## Conclusion

In final words, firstly I fetch the data-set cleaned it and made that perfect, later on trained on three different models and got 82%, 75% and 75% accuracy respectively.