Vivekanand Education Society's
Institute of Technology
Department of Artificial Intelligence and Data Science

WILD FIRE PREVENTION USING AI/ML

Team Members:
Atharva Sardal
Ashish Patil
Aum Kulkarni
Sanika Dhuri

Guide Name:
Mrs. Sangeeta Oswal

CONTENT TABLE

INTRODUCTION

NEED & RELEVANCE

They pose a threat not only to the forest wealth but also to the entire regime to fauna and flora seriously disturbing the ecology and environment of a region.

When there is no rain for months, the forests become littered with dry senescent leaves and twinges, which could burst into flames ignited by the slightest sparkt

Objectives

- To study and evaluate various machine learning models

- Using machine learning model to predict information about forest fires

Problem statement

Every year forest fires destroy a huge area of forest cover, leaving large-scale destruction of flora and fauna in it's wake. Devise an effective solution to minimize their effects using AI/ML

About the Data set

- Obtained on UCI-repository
- Country Specific data set based in Portugal
- Time of Data set creation-2008

Data Set height- 517 rows

Total Data Points : (517*13)=6721

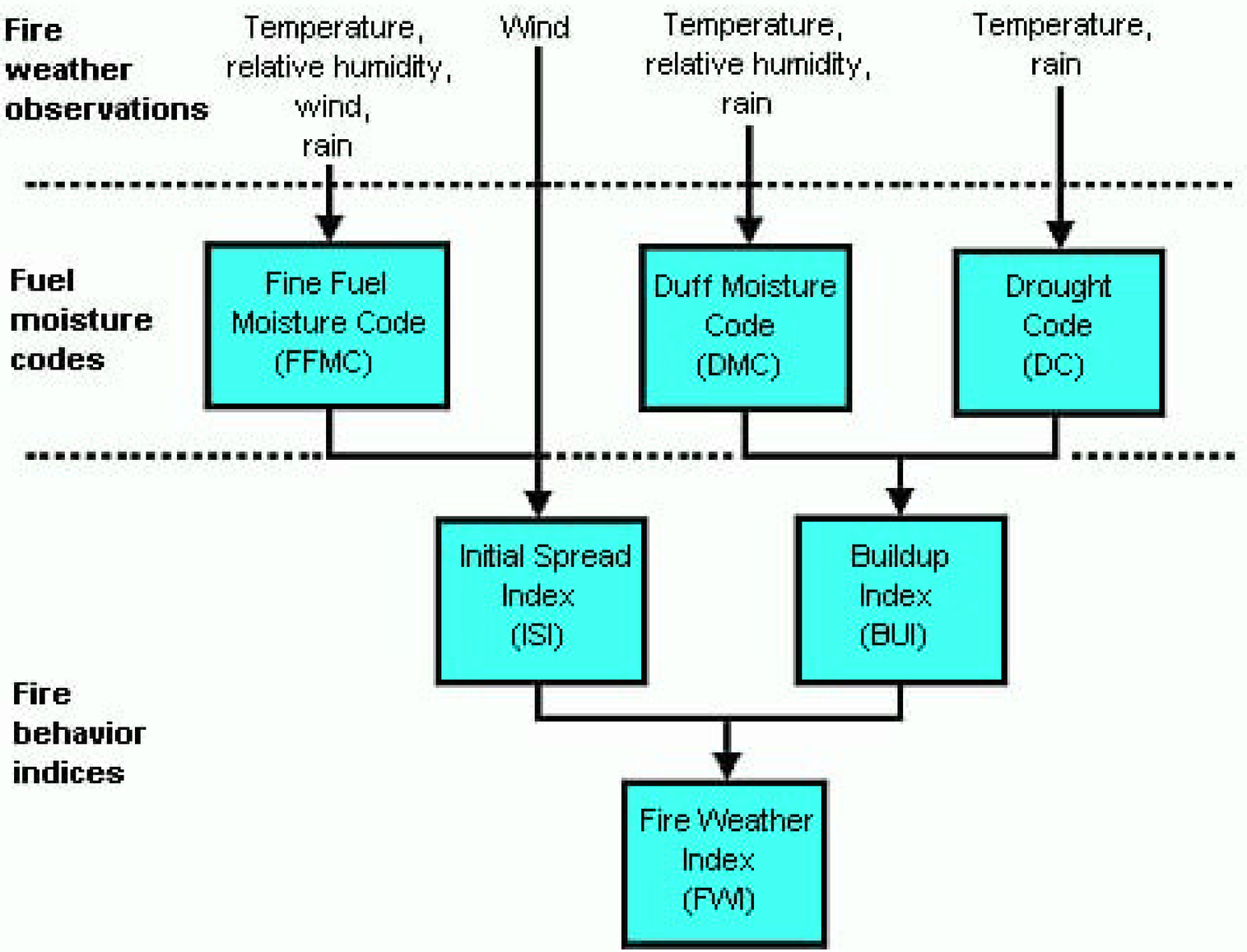
NO NULL VALUES

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMC - FFMC index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84

About FFMC,DMC ,ISI,BUI,DC

- The **Canadian Forest Fire Weather Index (FWI) System** consists of six components that account for the effects of fuel moisture and weather conditions on fire behavior.

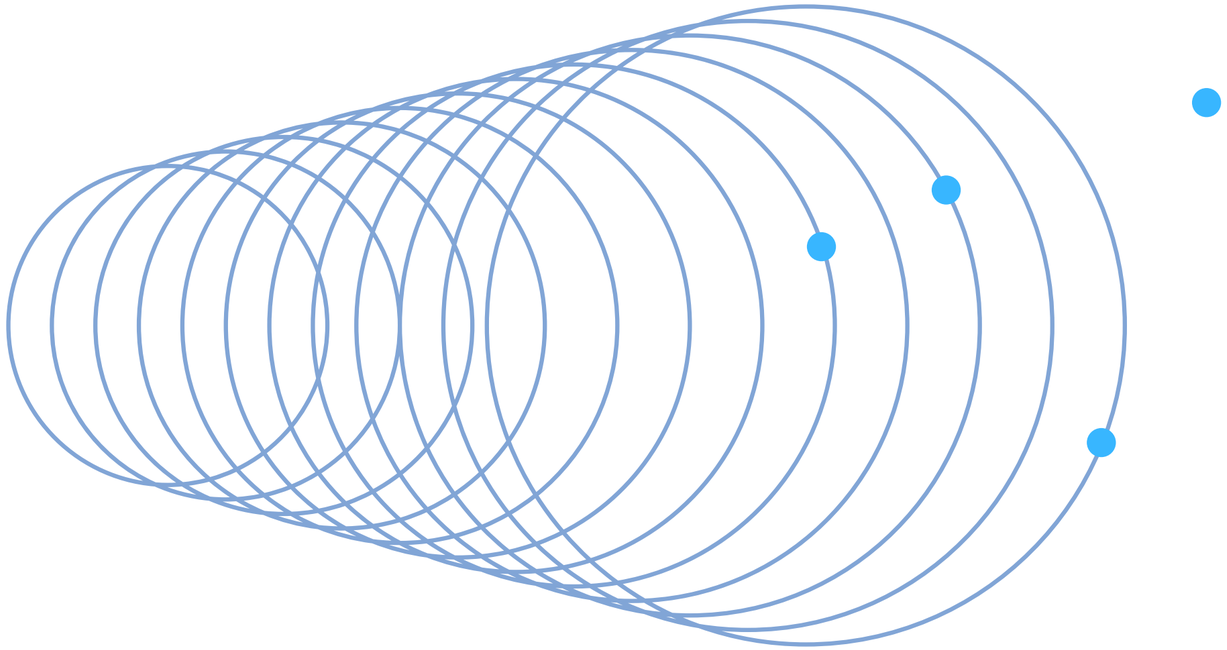
- They are- FFMC,DMC,DC,ISI,BUI


Fire weather observations
Temperature, relative humidity, wind, rain
Wind
Temperature, relative humidity, rain
Temperature, rain
Fuel moisture codes
Fine Fuel Moisture Code (FFMC)
Duff Moisture Code (DMC)
Drought Code (DC)
Initial Spread Index (ISI)
Buildup Index (BUI)
Fire behavior indices
Fire Weather Index (FWI)

Data pre-processing

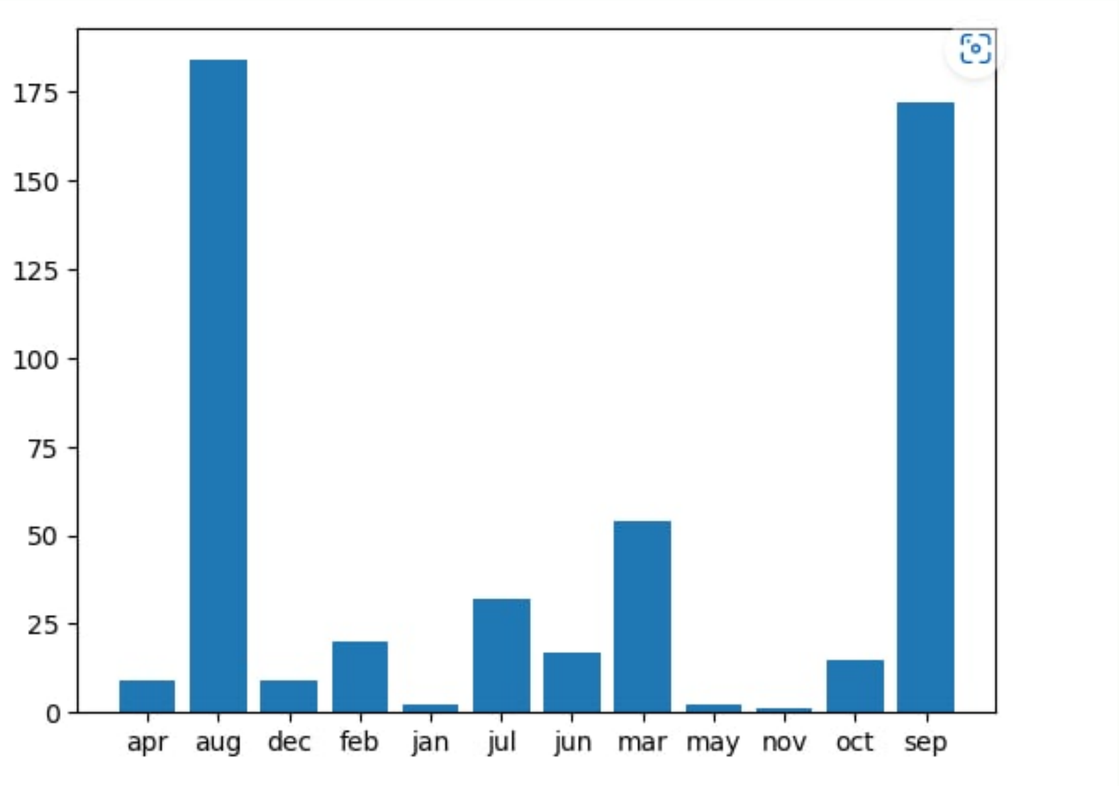
Three steps have been performed in data preprocessing:

- Up sampling
- Reducing Skewness
- One Hot encoding of categorical columns



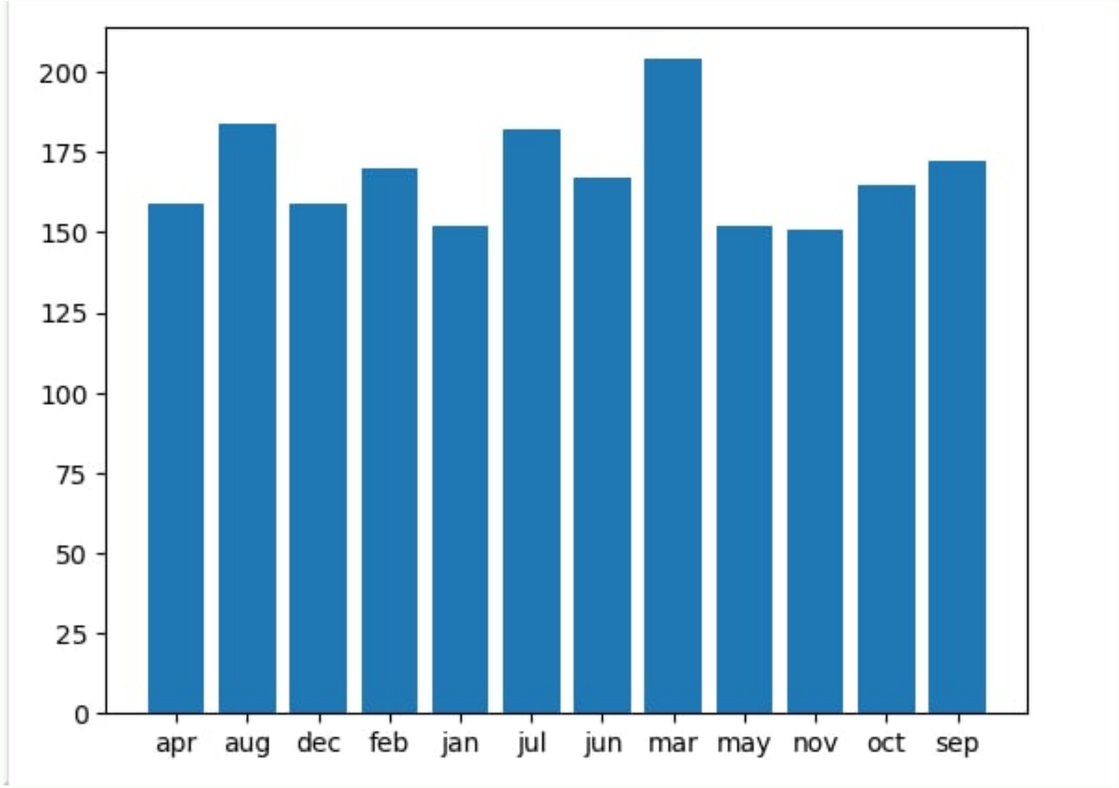
UPSAMPLING:

Upsampling is a procedure where synthetically generated data points (corresponding to minority class) are injected into the dataset. After this process, the counts of all the labels are almost the same. This equalization procedure prevents the model from inclining towards the majority class



DATA SET BEFORE UPSAMPLING

517

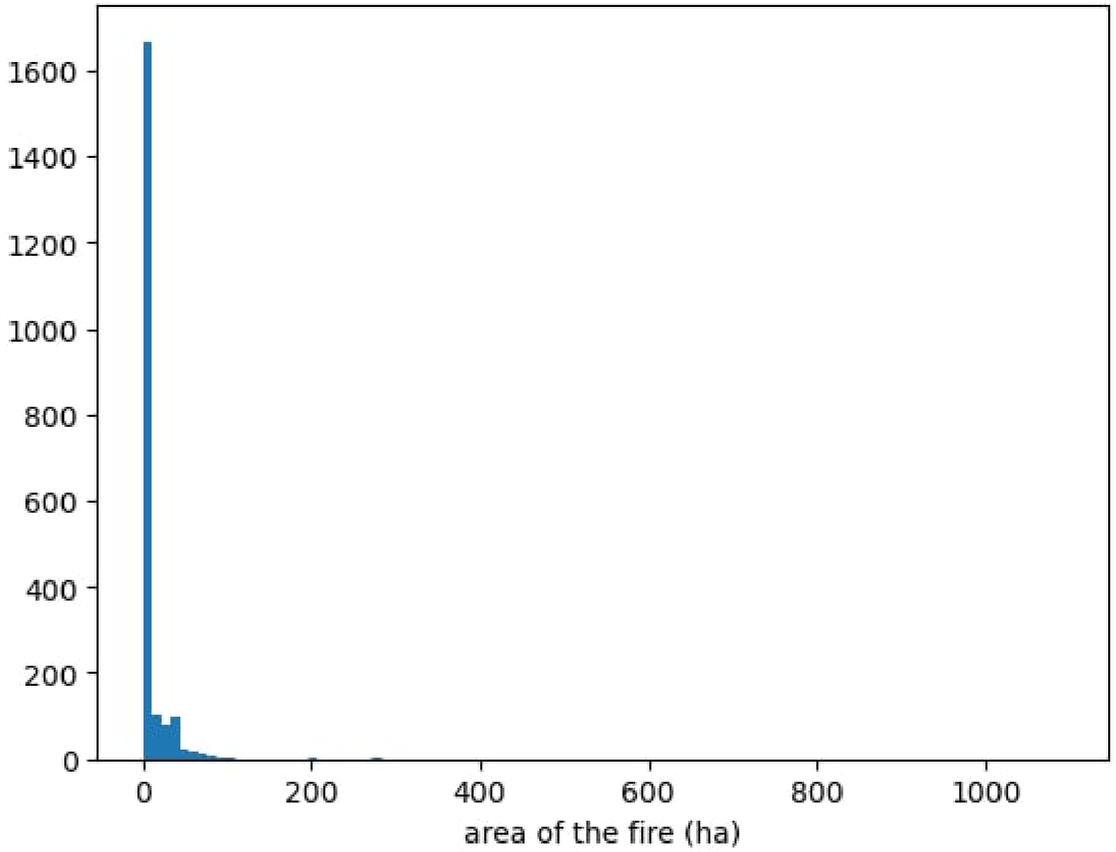


DATA SET AFTER UPSAMPLING

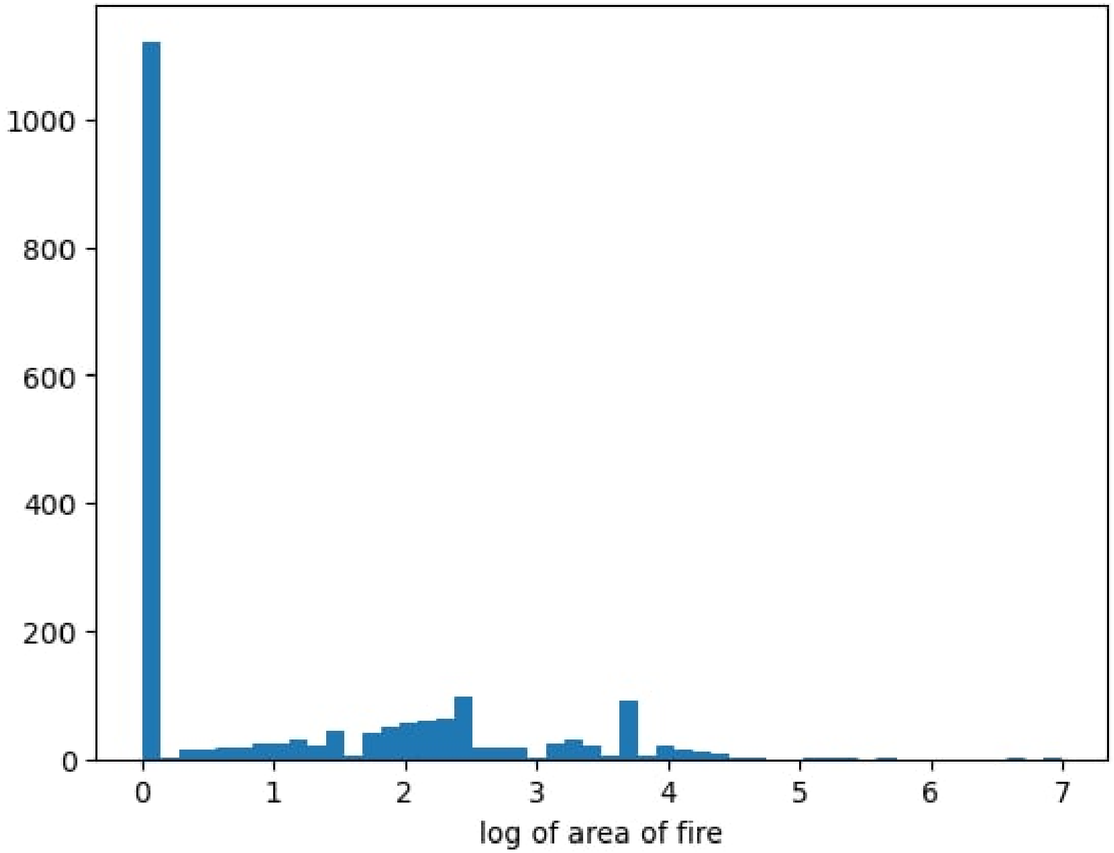
*The data points for the months have been equalized 2017

Reducing Skewness:

As the area column of the data set is very skewed to towards zero, we take the log of the area to reduce it skewness



Before taking Log



After taking Log

• One-Hot encoding

The categorical columns of months and the days are one hot encoded to create separate columns for each month and day

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

After One-hot encoding

	month_apr	month_aug	month_dec	month_feb	month_jan	month_jul	month_jun	month_mar	month_may	month_nov	month_oct	month_sep
0	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	0	0

MODEL-TRAINING

Methods used for training the model are:-

Linear Regression: Linear Regression is a supervised learning algorithm that aims to draw the best fit line that has the least mean squared error.

Support Vector Regression: Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. It is based on the concept of Support Vector Machines (SVM). The aim of this model is to find the best fit hyperplane that includes the maximum amount of points.

MODEL-TRAINING

Hyperparameters:-

kernel: the kernel type to be used. The most common kernels are rbf (this is the default value), poly or sigmoid, but you can also create your own kernel. RBF has been used

C (Regularization parameter): ells the SVM optimization how much you want to avoid miss classifying each training example. A C value of 1 is taken

gamma: The gamma parameter defines how far the influence of a single training example reaches. Lower the value of gamma, more number of data points a greater distance will be considered. A gamma value of 1 is taken.

*All the features have been scaled to using Standard Scaler of scikit-learn library before training the model

MODEL-TRAINING

Evaluation Metrics:

MSE(mean square error):

The Mean Squared Error measures how close a regression line is to a set of data points.

MAE(mean Absolute error):

Its the measure of errors between two set of paired inputs expected to give the same outcome

MODEL–TRAINING

Evaluation Metrics:

R2 Score:

The R2 coefficient of determination measures how well the regression predictions approximate the real data points. An R2 of 1 indicates that the regression predictions perfectly fit the data.

Model Evaluation for SVR Model

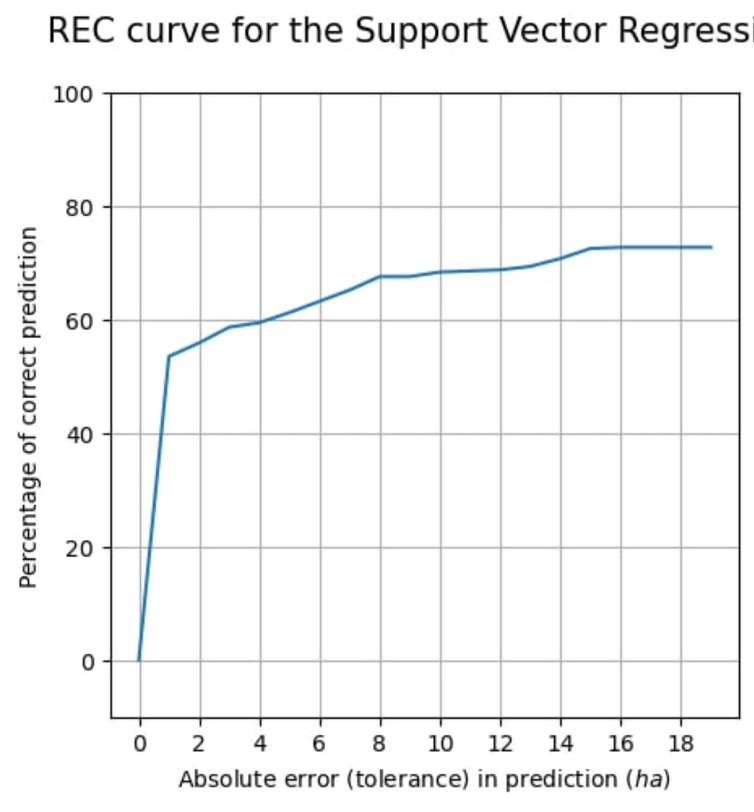
Mean Squared Error: 0.4493

Mean Absolute Error: 0.3127

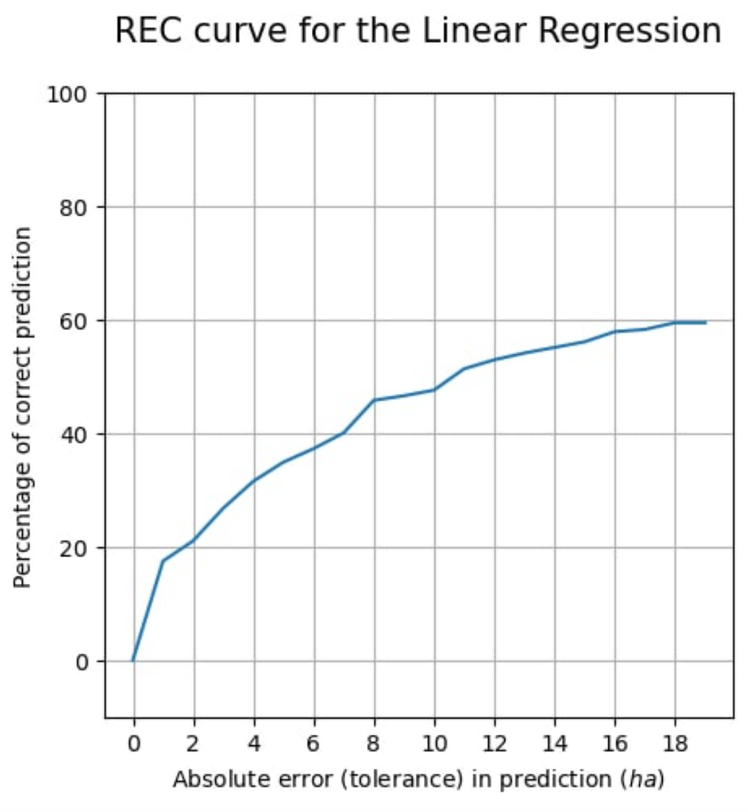
R2 Score: 0.7585

REC curves analysis

*regression characteristic Curves



The SVR predicts more than 50% of the points within a tolerance of 2%



Linear Regression which predicts only about 20% within a tolerance of 2%: