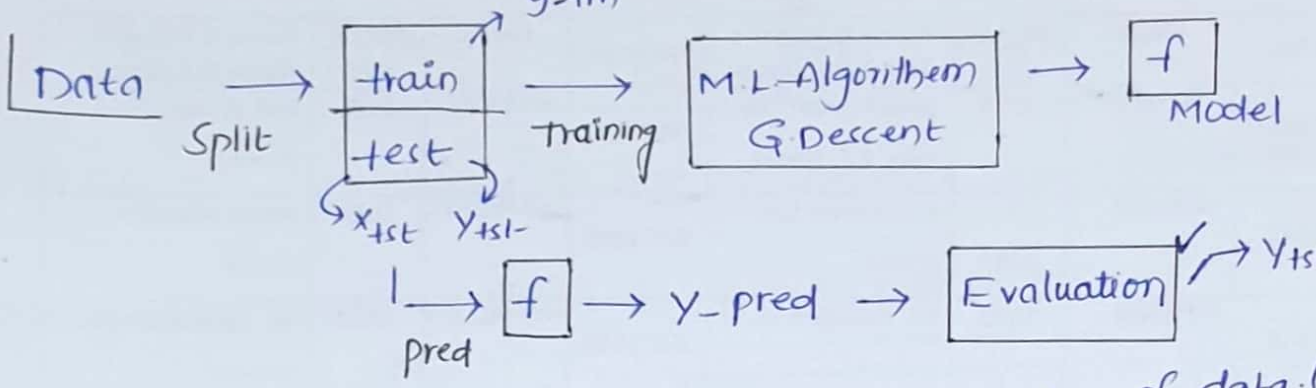
Evaluation Metrics-1 :-

Pipe line of Evaluation :- y-trn → Discrete → Classification (or) ∈ R → Regression



There are two types of Evaluations based on type of data (y-trn)

if it a 1) classification task,

Perform diff matrices — ① Accuracy

like — ② Confusion Matrics
③ Precision & Recall
④ F1 score
⑤ ROC - AUC
⑥ Log - loss

2) Regression task we have to perform Evaluatrn Matrices like -

① MAE - Mean Absolute Error
② MSE - Mean Square " "
③ RMSE - Root Mean " "
④ R^2_score → R2 score
⑤ MAD → Median Absolute deviation

Regression Evaluation Matrics :-

MAE :- is nothing but measure of errors between paired Observations like (y_{act}, y_{pred}) ie mean difference b/w $y_{act} - y_{pred}$

$$M.A.E. = \frac{\sum_{i=1}^{n} |y_{act_i} - y_{pred_i}|}{n}$$

In Other terms we can Say - the Sum of the absolute difference between actual and predicted values.

M.S.E :- it is defined as mean or Avg of the Square of the diffrence b/w actual and predicted value

$$M.S.E = \frac{1}{n}\sum_{i=1}^{n}(y_{act_i} - y_{pred_i})^2$$

R.M.S.E :- is defined as root of mean of Square of the diff b/w actual and predicted value (or) Simply root of M.S.E.

$$R.M.S.E = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{act_i} - y_{pred_i})^2} \text{ or } \sqrt{M.S.E.}$$

M.AD :- is defined as median of Sum of absolute differences b/w yactual to ypredict (or) Median absolute deviation
↳ Error

$$M.A.D = \text{Median}\,|y_{act_i} - y_{pred_i}|$$

R^2 - score :- is a Coefficent of determination it is the total variance explained by the model based on correlation b/w actual & predicted value

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad 0 \leq R^2 \leq 1$$

where —

$SS_{res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ → residual Sum of Square

$SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ → Total Sum of Square

$\bar{y}_i$ → mean

Note : In SS_{tot} we get $\bar{y}_{i(pred)}$ from Simple mean model

In SS_{res} we get $\hat{y}_{i(pred)}$ from linear regression model

Case - I :- If we make a best model ie where there is no errors. in that $SS_{res} = 0$

$$R^2 = 1 - \frac{0}{SS_{tot}} = 1$$

Case - II :- if a model (regression) behave like Simple mean model in that $SS_{res} = SS_{total}$

$$R^2 = 1 - \frac{S_{tot}}{S_{tot}} = 0$$

Case - III :- if our model contain more errors ie the model behaves like worrest than Simple mean model $SS_{res} > SS_{tot}$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = -ve$$

Case - IV :- if Simple mean model Contain more-Errors as Compare residual

ie $SS_{tot} > SS_{res}$

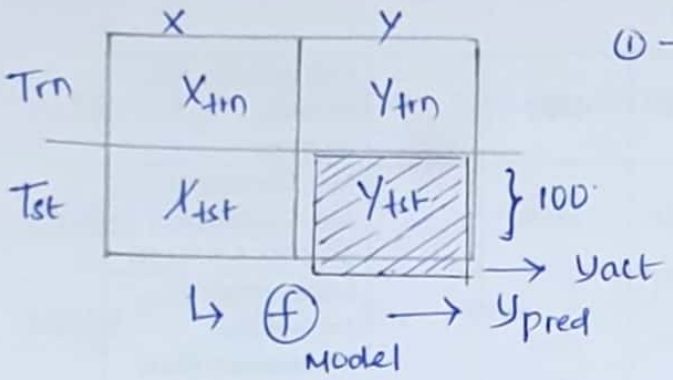
$$R^2 = \frac{SS_{res}}{SS_{tot}} = \text{lies blw } [0,1]$$

Classification Evaluation Matrics :-

① Accuracy :- it is the most intuitive performance measure and it is simply a ratio of correctly classified observation to total observations.

if we have high Accuracy then the model is best Model.

Accuracy Calculations :-



(i) - After Split if Suppose test data Contain 100 rows in that 60 are +ve classified & remaining 40 -ve classified

(ii) After model making predicted values Shows as 53 +ve positively classified data. Consist 53(+) Correctly classified and 7(-) are mis classified, Similarly in Negitively classified data Correctly classified 35(-), miss classified (+v)5 Out of 40 Shown below.

Acutal values	predicted values	
	$X_{tst} \Rightarrow Y_{tst\_pred}$	
Total Data points 100 → 60 +ve	M →	Correct: 53 +ve, 7 -ve; Miss classifier
→ 40 -ve	M →	35 -ve, 5 +ve
	Correctly → 88	Mis → 12

$$\text{Accuracy} = \frac{\text{No of Correctly Classified}}{\text{Total No of data points}} = \frac{88}{100}$$

$$= 88\% \text{ or } 0.88$$

Imbalance data set :-

if we have a descrete type of data set if consist two type classifications like +ve & -ve for Suppose we have 1000 data points On that +ve type of data is 950 and -ve type of data is 50 there is huge mismatch of data. this type of data called as Imbalanced Data.

Case II – for Suppose. +ve class Consist 500 data points & -ve class Contain 500 data points then we called it as balanced Data set

Note: So if we have Imbalanced data set then we have a big problem to predictions before going to any machine learning model we have to handle this type of data. (ie balance the data set).

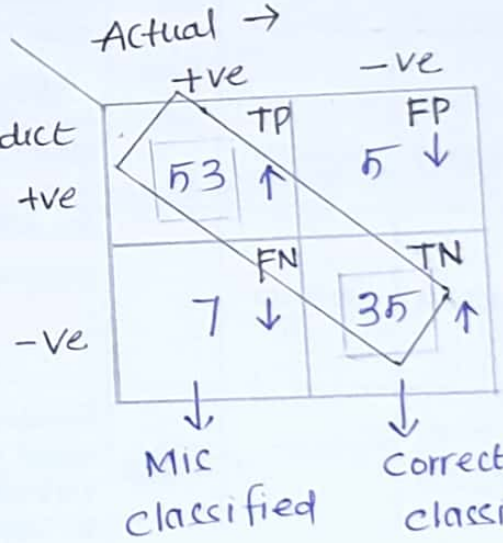
Ex:- Generally we got this type of Imbalanced data sets mostly in Health Cares Organizations, BSFI, Banking Domains, Credit Card fraud if we have a balanced data set then Accuracy will good option, In other factors like Imbalanced data then Accuracy matrics is not good option So that's why we go with other matrices like –

Confusion Matrix :- it is a table that is often used to describe the performance of a model (classification) on set of test data for which the true values are known.

if we take the same example of 100 data points set there are

correctly classified as – 88

Mis classified as – 12



if we put this data into Confusion matrix terminology.

Step -I - We have to devided table as two classes namely Actual and predicted in that also devide classes as binary format ie two classes only (like - Yes/No, True/false Male/Female etc.) Confusion Matrix Only Supports to binary format.

Step -II - According to that classes we devided the classifier as shown in figure as (+ve, -ve) lables ie total positive points 60 Correctly classified ie In Actual/prediction give true value in this case 53 (+ve) put in first cell, in below cell we have put misclassified data point ie 7 (-ve) Similarly in actual data set and predicted data (40) correctly classified (-ve) class put beside cell 35 (-ve) finally above the cell put misclassified 5 (+ve)

Step -III - Now if we observe correctly classified data points and misclassified data points set in diagonally ie the diagnal frame of data gives us high values (maximum) always.

Step -IV - Now giving the names to each cell according to we put the names above (TP, FN, TN, FP)

Note :- In case of Imbalanced data set Confusion Matrix are good option to caliculate Accuracy.

$$\text{Accuracy} = \frac{T.P + TN}{\text{Total } (TP+FN+TN+FP)} = \frac{35+53}{53+7+35+5}$$

$$= 0.88$$

Terminologies in Confusion Matrix :-

$y \in \{+ve, -ve\}$

Dataset → D_{Test} values → 1000 → 900 (-ve) → M → 850 -ve, 50 +ve

1000 → 100 (+ve) → M → 94 +ve, 6 -ve

P ↑ / A →	+ve	-ve
+	TP: 94 ↑	FP: 50 ↓
-	FN: 6 ↓	TN: 850 ↑

Total - 100 (+ve, P) ; 900 (-ve, N)

$\Rightarrow$ T.P.R = True positive Rate = $\frac{TP}{P} = \frac{94}{100} = 94\%$

T.N.R = True Negative Rate = $\frac{TN}{N} = \frac{850}{900} = 94\%$

F.N.R = False Negative Rate = $\frac{FN}{P} = \frac{6}{100} = 6\%$

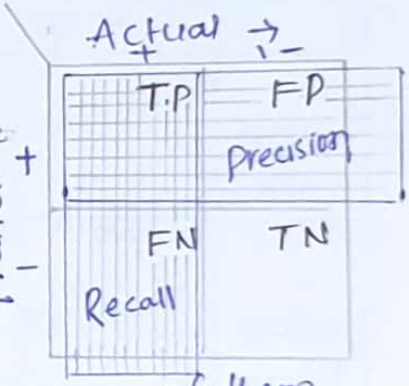
F.P.R = False positive Rate = $\frac{F.P}{N.} = \frac{50}{900} = 50\%$

Note :- TPR is also called as Sensitivity & TNR is called as Specificity

Precision & Recall :- This matrices are derived from Confusion Matrix

① **Precision :-** out of all the points (+ve) the model predicts to be positive

$$\text{Precision} = \frac{TP}{TP+FP}$$



② **Recall :-** Also called as Sensitivity (T.P.R)

Out of all the actual positive points, what percentage of them have you predicted to be positive

$$\text{Recall} = T.P.R = \frac{TP}{P} \simeq \frac{TP}{TP+FN}$$

Both precision & Recall lies in the range of [0, 1]

Note :- if any body asking you precision & Recall of your model both of them high and maximum ie close to 1,

F1 Score :- F_1 Score also called as F Score it is measure of model's Accuracy on dataset., F_1 Score is way of combaining the Precision & recall of the model and it is defined as hormonic mean of models precision & Recall.

$$F_1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

ROC - AUC :- Stands for Reciever operating charecterstic curve

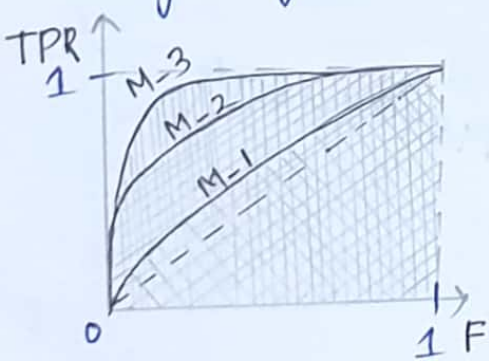
AUC Stands for Area Under the Curve.

Roc - AUC Matrix only works On Binary classification (ie if we have only two classes) if we have more than two classes this matrix is not good.

Roc :- It is an evalution Matrix for binary classification problems. it is the probability Curve that plots the TPR againest FPR at various threshold values and essentially Separates the "Signal from the noise"

AUC :- it is measure of the ability of a classifier to distinguish between classes and is used as a Summery of the roc Curve.

Note :- the higher the AUC, the better the performance of the model distinguishing b/w the positive and negative classes.



M-1 = Model 1

from graph we can say M3 is better than as compare with (M1 & M2)

from Confusion Matrix —

FPR = 1-TNR

FNR = 1-TPR

Log loss :- Can also called as logarithem loss, it is a most important classification matric based on probabilities for any given problem. alower logloss value means better predictions, logg-loss is a slight twist on something called as likelihood-function, Infact logloss is -1 * the log of the likelihood-function.

Note :- Small logloss value gives better performance.

$$\text{logloss} = -\frac{1}{n}\sum_{i=1}^{n}\left\{(y_i * \log(p_i)) + ((1-y_i) * \log(1-p_i))\right\}$$

Short Review of Performance Matrics

Regression

① M.A.E.
② M.S.E.
③ R.M.SE.
④ MAD.

} Errors made by model so it is as if possible maintain low value

⑤ R^2 → As Near to 1
if '1' then better performance

Classification

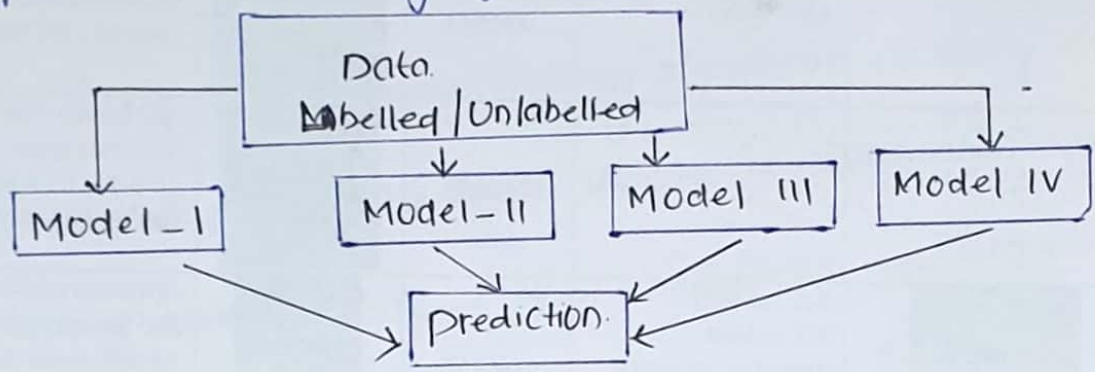
① Accuracy → big problem of Imbalance dataset
② Confusion Matrix
③ Precision & Recall
④ F1 score

} Best Matrics for Multiple classification.

⑤ ROC-AUC → only for binary classification.
⑥ log loss ↓ its Impact With Imbalance
↳ Smaller the best (Near to '0')

Note : In classification the performance matrics (C.M, P&R, F1 score) are best matrice because it is well perform even in multiple classification as well handle the Imbalanced datasets.

Ensemble Technique :- this techniques used in multiple learning algorithems to obtain better predective performance than could be obtain from any of the constituent learning algorithem alone

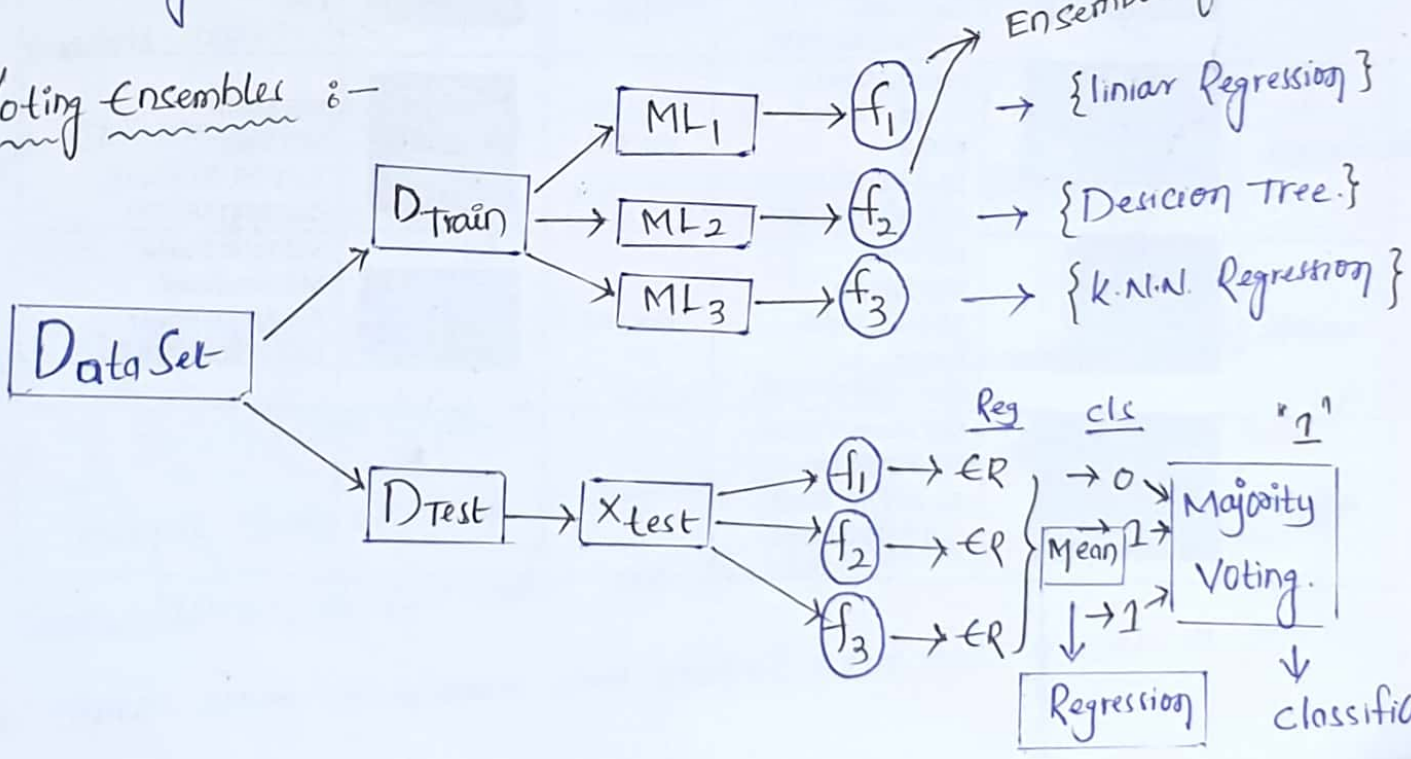


Classifications :- As name suggested ensemble means groups, previosly for certain task completions we used one algorithem only, but here in ensembles we used different type of machine learning algorithems to perform a better result. So there are diffrent techniques avialable to do this task. They are -

① Voting Ensembles ② Bagging Ensembles ③ Boosting.

④ Stacking Ensembles ⑤ Cascading Ensembles.

Voting Ensembles :-



Procedure for Voting Ensemble :-

Step-I :- first of all we have to Split a dataset in-to two parts as $[D_{Train}, D_{rest}]$ train and test data sets

Step-II :- Apply Some machine learning algorithems as shown in diagram. So here we desided wether Our target variable is classification type or Regression type of variable case-I if it classification type then do some classification algorithem apply like (logisitic Reg, KNN classification, DT classification etc) to train data set, Case-II if it is regression type of algorithem apply like (linear Regression, KNN Regression, Descision tree regression etc)

Step-III :- if it is a case-I {ML-1, ML-2, ML-3} ensembles of Machine learning models as (f_1, f_2, f_3), in case-II also weget it.

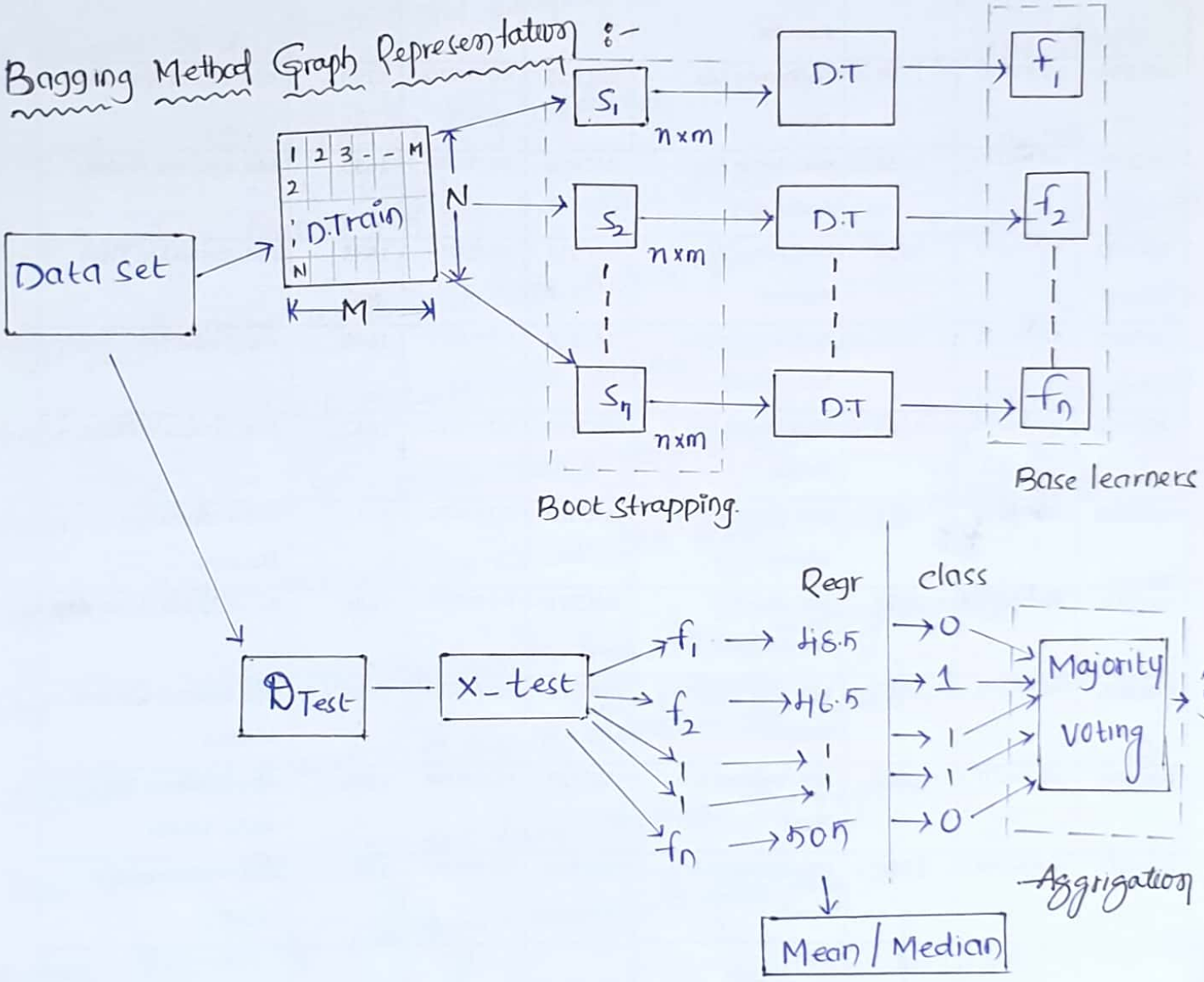
Step-IV :- NOW Consider a Dtest Data appling Some X_{tst} values in models which we applying deffrent type of algorithems ashown in graph. as per it we get (f_1, f_2, f_3) models of {linear Reg, DTReg, KNN Reg} in case-II as well we get (f_1, f_2, f_3) models of (logistic Reg, DT classification, KNN classification) in case-I.

Step-V :- if Our (f_1, f_2, f_3) values ase classifiers and predicted as (0,1) output then we take a majority of voting and decided the result as either (1 or 0), if our values are Real Numbers $\in R$ then we Consider this threes Mean (or) Median and predict which type of class or prediction.

2) Bagging :- it is also called as Bootstrap Aggrigation, is a ML Ensemble, Metal-Algorithem designed to Improve the Stability and accuracy of machine learning Algorithems used in statistical classification & Regression it also reduces Varience and help to avoid Overfitting

Random forest :- it is a Supervised Machine Learning algorithm, The "forest" it builds is an essembles of decision trees usually trained with the "bagging" method. the general idea of the bagging method is that a Combination of learning models increases the Overall model.

Bagging Method Graph Representation :-



Boot strapping :- it is Nothing but Sampling of row & column.

Procedure for Bagging :-

Step-I :- first of all we have to consider a data set and split that data set into two parts as train & test Data set. now consider a train data of a size (N,M) represent no of columns, rows as respect.

Step-II :- Now devided this train data in to diffrent 'n' no of samples as shown in figure of a dimentionality of (n x m) represent no of columns, rows respectively. after deviding the samples do Sampling of row & columns of Samples of 'n' numbers this Sampling is called as "Boot Strapping"

Step-III :- Now applying only decision Tree algorithm to each Sample of 'n' numbers because this algorithm make decision yields only as per this bagging technique we only consider decision tree algorithm, Similarly in Random forest also we used decision tree algorithm only because as per name random means Select Sample randomly from a trained data. and forest means trees.

Note :- In this procedure we discuss both "bagging method aswell Random forest"

Step-IV :- After applying Decision tree Machine learning algorithm we creat diffrent kind of models of size 'n' numbers. this models are called as base learners also in random forest algorithm. this algorithm do two types of tasks they are of our target variable is ($\in R$) any real number we do regression tasks, if it is catogorical type of data we do classification task.

So Random forest algorithm do regression, classificaton tasks.

Step-v :- Now Consider test data (D_{test}) from X_{test} Data we applying to our models ($f_1, f_2, \ldots f_n$) if it is $\in R$ any real number do regression, if it is categorical do classification task. after this we calculate mean or median to give predicted data (ie prediction), if cat then do majority voting So in this majority voting considered as that type of Data. and that give prediction ($\hat{y}$) this represent category of data.

Note :- After Sampling whenever we apply decision tree algorithm from that step to final prediction (ie outcome) internally doing "Randomforest" process

Termonology description of Random forest :-

The models (f) is called as Base learner (Internally we applying "DT" Algorithm) what ever models we used here ($f_1, f_2 \ldots f_n$) their depths (or) depths should be high as possible (ie the base learners should be Over-fitted models).

Over-fitting :- in bias termindogy Overfitting is nothing but high Varience and low bias, (ie the train data should be well fitted in the boundary.)

Bootstrapping — (Row Sampling, Column Sampling)

Aggregation — classification — Majority voting ; Regression — Mean (or) Median.

Graph for Random forest :-