1. Metrics in classification: good
2. YOLOV8 ============ yet to covered
3. Finding the prime number : good
4. CNN ========= yet
5. Opencv ===== yet
6. ML vs DL vs NLP : good
7. Read nd write : good
8. Class object : yet
9. Csv file 2 5 rows: good
10. Median imputation : good
11. Annotation images: yet
12. Train vs test: good
13. List vs tuple: good
14. Real times : good
15. ML : good
16. Precion vs recall: good

11 ======== 16

Tomorrow you have a class :

As usual

9 to 11 : ML

Friday : 10 ====== python ml

Class object packages

Saturday : 9 to 1 regression : api

Sunday : 9to1

Monday : classifaction python

Cluster with python

SVM

Bagging and boosting

Azure for students : ask friends

**Ensemble Methods**

**Ensemble : group of items**

**Ensemble Methods or algorithms: group of algorithms**

**Suppose a data has 1000 observations are there**

**Now you want develop a ML algorithms**

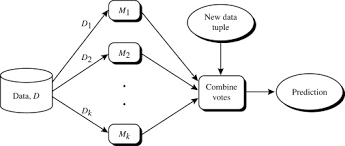
**Logistic regression on 1000 observations**

**Decision Tree on 1000 observations**

**Naïve bayes on 1000 observations**

**KNN on 1000 observations based on metrics pick the best model**

1. **Will divide 1000 observations into 4 parts**
2. **Each part has 250 observations**
3. **Part1 Part2 Part3 Part4**
4. **Logistic regression apply on Part1**
5. **DT apply on part2**
6. **KNN apply on part3**
7. **NB apply on part4**
8. **New test sample will pass through all 4 models**
9. **Every model will give one answer**
10. **Final answer : majority voting**

****

**Ensemble Methods are divided into two parts**

1. **Bagging**
2. **Boosting**

**Bagging:**

**How to divide the data?**

**Bootstrap aggregation : ?**

**First time out of 1000 observations we are picking 250 randomly**

**Qns: 250 are gone : 1000-250: 750**

**Will you use these 750 observation to pick next 250**

**0r**

**Will you pour 250 in bag again: 1000**

**Bootstrap aggregation :**

**Sample with replacement**

**Step-1: first 1000 observations are there, will pick 250 observations randomly**

**Step-2: again that 250 , you included in the bag: 1000**

**Again will pick 250 observations**

**First query: how to divide the observations**

**Answer: Bootstrap aggregation : sample with replacement**

**Second query: 250 : Srikanth 250 : Srikanth**

**250 will never match with other 250**

**63% of data will repeated, 37% will never choose**

**What is the probability to choose the observation:**

**N observations are there what is the probability of one obser = 1/N**

**Probability of not choose that observation = 1-1/N**

**Suppose im repeating n times = (1-1/N)^n ====== > 37%**

**Probability to choose the observation is 63%**

**Probability not to choose the observation is 37%**

**1, 2,3,4,5,6,7,8,9,10**

**1,2,3,4,5,6,7 8,9,10 (not able to choose)**

**The observations 37% are called OOB (out of bag data) ,**

**this data you can use as test data**

**this will avoid overfitting of the model**

**how many times we will repeat this process then ?**

**how many models you want to apply : 3 model**

**How can it avoid overfitting:**

**What is meant by overfitting**

**Low train error ------- high test error**

**5 qns ---- 5 answer :**

**Two guys will go to interview**

**Person: who is pm of Pakistan : long back butto now mushrao**

**Person2 ==== person1 :**

**Person-2: what is**

**Overfitting is happened because of less data**

**Data ==== train and test**

**Bagging what ever you are choosing : train data it is enough to train moel**

**Model overfit: less data**

**Less data =========== increase the data**

**Data augmentation**

**Test data is forming automatically : OOB data**

1. **Bagging is one of the ensemble method**
2. **Bagging means: Bootstrap aggregation**
3. **Bootstrap aggregation: sample with replacement**
4. **When do this 63% data will choose**
5. **37% data will never choose**
6. **This 37% of data is called OOB data**
7. **OOB data consider as test data**
8. **To train model we have more than 50% data i.e 63**
9. **This will avoid overfitting**
10. **63% of data consider a train data ===== train error**
11. **37% of data consider as test data ===== test error**
12. **Test error also called out of bag error**

**Best example of Bagging model : Random forest**

**Ensemble means : group of different models**

**Bagging is one type of ensemble method**

**Group of different models ========= Group of Decision tree**

**why we are applying different models in different parts?**

**Every model is unique === pattern learning**

**If you divide the data and apply different model ===== solution robust**

**You are training 67% of data**

1. **Bagging example : Random forest**
2. **Will apply DT is the only algorithm on different data**
3. **DT if you use same variable to calculate the root node**

**There might be a chance only variable come as root node**

**For all the DT**

1. **This will leads bias**
2. **To avoid this , generally in ensemble method will pick observations randomly**
3. **But in RF will pick not only observation but also variables**
4. **P = total number of predictors**
5. **Will pick**
6. **= 10 variables data**