Q What is F_1 score

Ans:

It's a statistical method. It's a measure of a test's accuracy. It considers both Precision & recall.
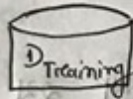
$$F_1 \, score = \frac{2 \times Precision \times recall}{Precision + recall}$$

Q What is k-fold Cross Validation.

Ans

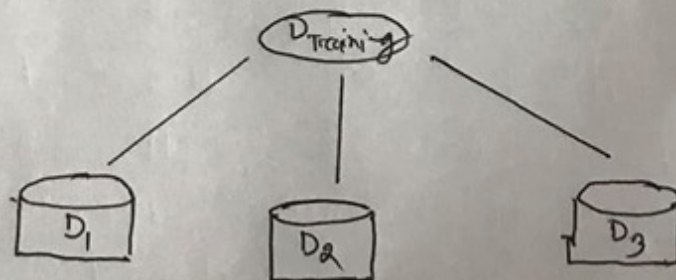k-fold Cross Validation is a Procedure used to estimate the skill of the model on new data.

k fold Cross validation =>

আবার কাছে big data আছে, যেখানে বলি $D_{Training}$. Initially কোন test data নাই,

এর $D_{Training}$ কে k সংখ্যক ভাগে ভাগ করবো,

ধরি k=3, তাহলে তিন ভাগে ভাগ করি-

আমরা এই ক্ষেত্রে Big data কে k সংখ্যক ভাগে ভাগ করাটাকে বলা হয় fold করা।

ধরি k এর মান তিন, তাহলে তিনবার Model ও train এবং testing হবে।
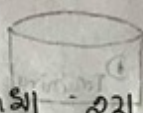
Training	Testing
(i) $T_1 + T_2$	T_3
(ii) $T_2 + T_3$	T_1
(iii) $T_3 + T_1$	T_2

যখন Testing হবে, test data গুলো unseen

এই Testing এর সময় দেখা হয় কয়গুলো k সংখ্যক iteration এর এই কয়গুলো instance সঠিকভাবে classified হয়েছে। যেখান থেকে Precision / Error এর average consider করা হয়।

285/259
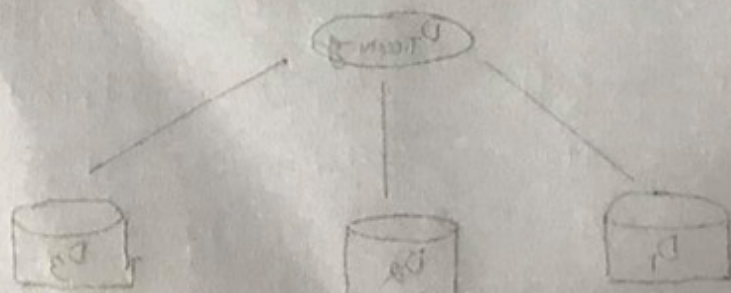
Q. Write decision-tree induction algorithm

<u>Ans:</u>　　Input: $D = \{x_1, \ldots, x_i, \ldots, x_n\}$

Output: A decision tree, DT.

<u>Method:</u>

$DT = \phi$

find the root node be with best spliting, $A_j \in D.$

$DT = $ create the root node;

$DT = $ add arc to root node for each split Predicate & label;

for each arc do

D_j. created by applying splitting Predicate to D;

if stopping Point reached for this Path then

$DT' = $ Create a leaf node & label it with C_i;

else

$DT' = DT$ Build (D_j);

end if

$DT = $ add DT' to arc;

end for