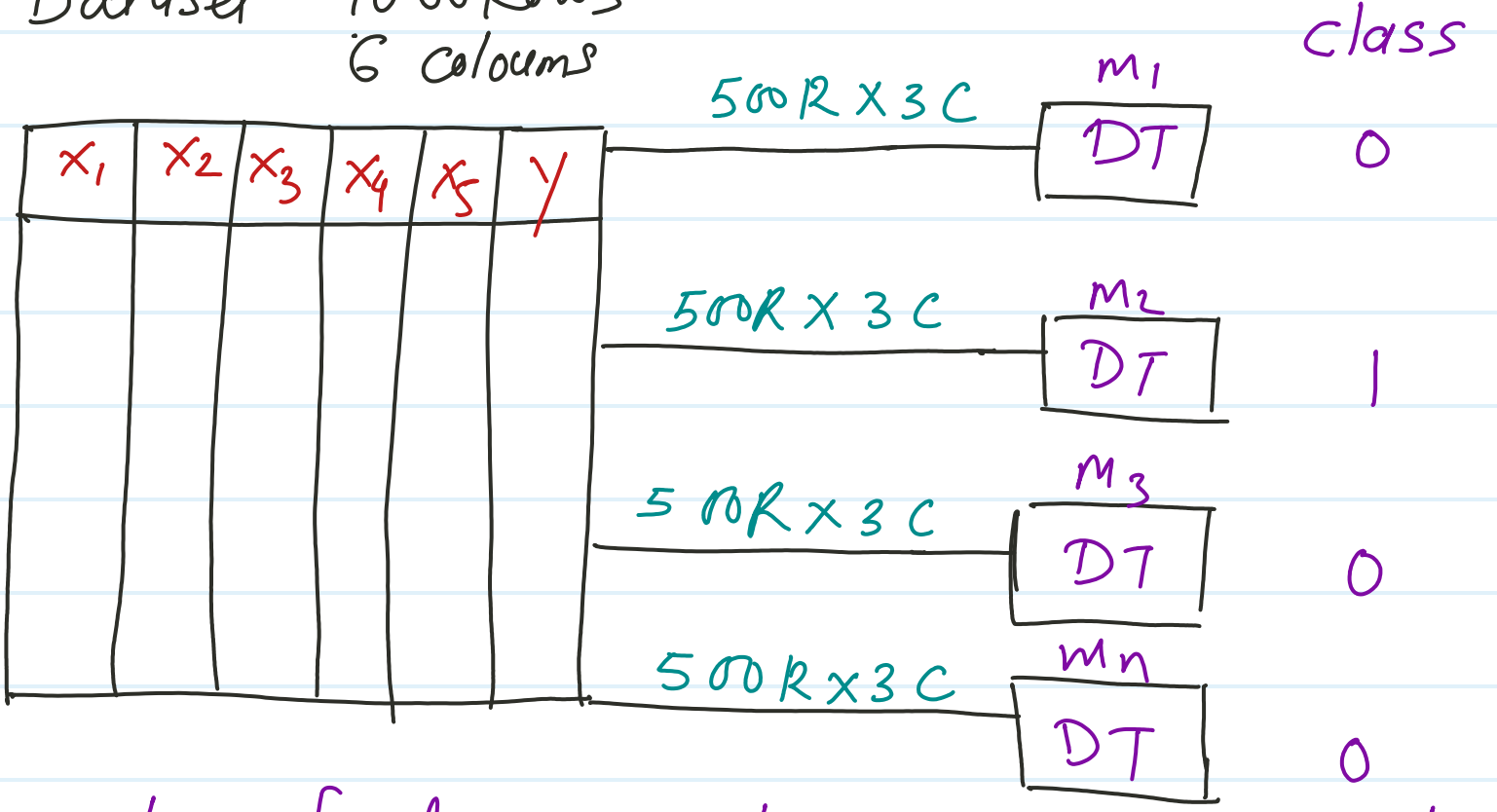
Bagging Method

* "Random Forest" classifier
* "Random forest" Regressor

Dataset - 1000 Rows
6 Colums



classification works on voting based output

entire Dataset - D

subset Dataset - d'

$D > d'$

since we are using ensemble technique. Trade of for Biased and variance would be

DT { low Biased — overfitting
high variance

RF { low Bias — Best fit
high Variance

Pros -

① Random Forest used to make robust model over a decisicy tree overfitting issue

② Random forest widely used in Regression problem due to its capabilities of handle non-linear dataset.

③ outlier / noise not affected it.

④ Can handle high dimensional data.

⑤ RF does not required feature scaling.

Equation of RF

T = Num. of DT in forest

$\hat{Y}_t(x)$ = Prediction of the tth DT for input x.

$\hat{Y}_t$ = final Random forest prediction for input x.

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t(x)$$

we calculate average of all DT. and it will be our final output.

For hyperparameter tuning, mostly Randomsearch CV prefer due to computational constraint.

- Grid search CV = $4 \times 4 \times 4 = 64$
- Randomsearch CV = [1,2,3,4] [1,2,3,4] [1,2,3,4] (69)

64 = 10

10mm 1

Random forest 100 DT / 200 DT