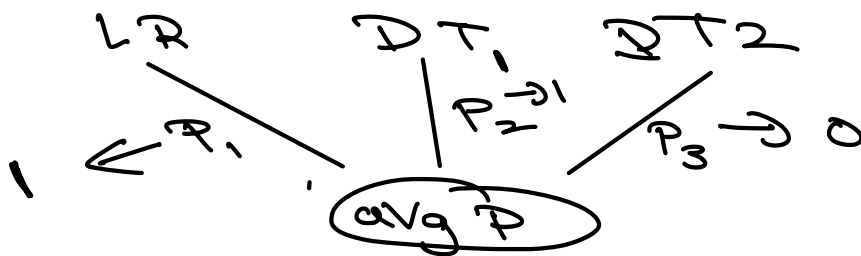


* Using DT classifier (max-depth=4)

Train-Score $\approx 85\%$

Test-Scores $\approx 78\%$

How to improve this?



Base-Learners + Combine/Aggregate Result
(as Unique as Possible)

Variants of Ensemble Technique

1) Bagging (Ex:- RF) *

2) Boosting* (Ex:- GBDT, XGBOOST, LightGBM)

3) Stacking

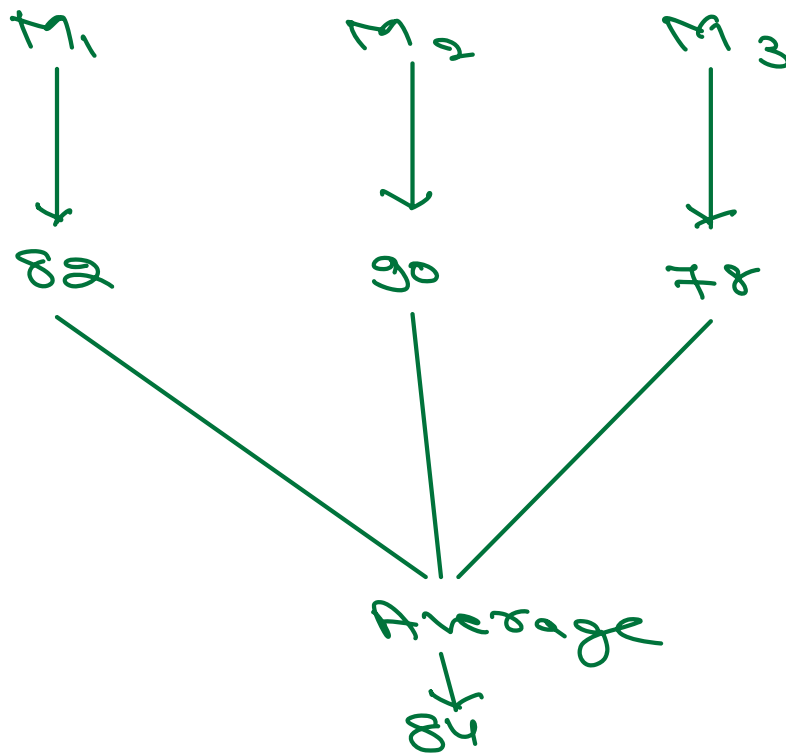
4) Cascading

} not so important from industry perspective
(Highly used ML Competitions)

Bagging

(Bootstrap Aggregation)

Bootstrap Sampling + Aggregation



Random Forest

(Bagging Model)

Each tree shall be
trained on Randomly
Sampled subset of Data

① Sample Rows (Row and Col Sampling)
② Sample Columns



CS + RS Leads to
1) Set of underfitted DT's
2) Different opinions



Row Sampling → oob Out of Bag Data points

Left out Row after Row Sampling can be Evaluate to Validate Performance

OOB Score ⇒ Variation of Cross Val Score



Pending

③ Implement RF with Data

③ Bias Variance in RA

③ HyperParam Tuning with Grid Search and Random Search