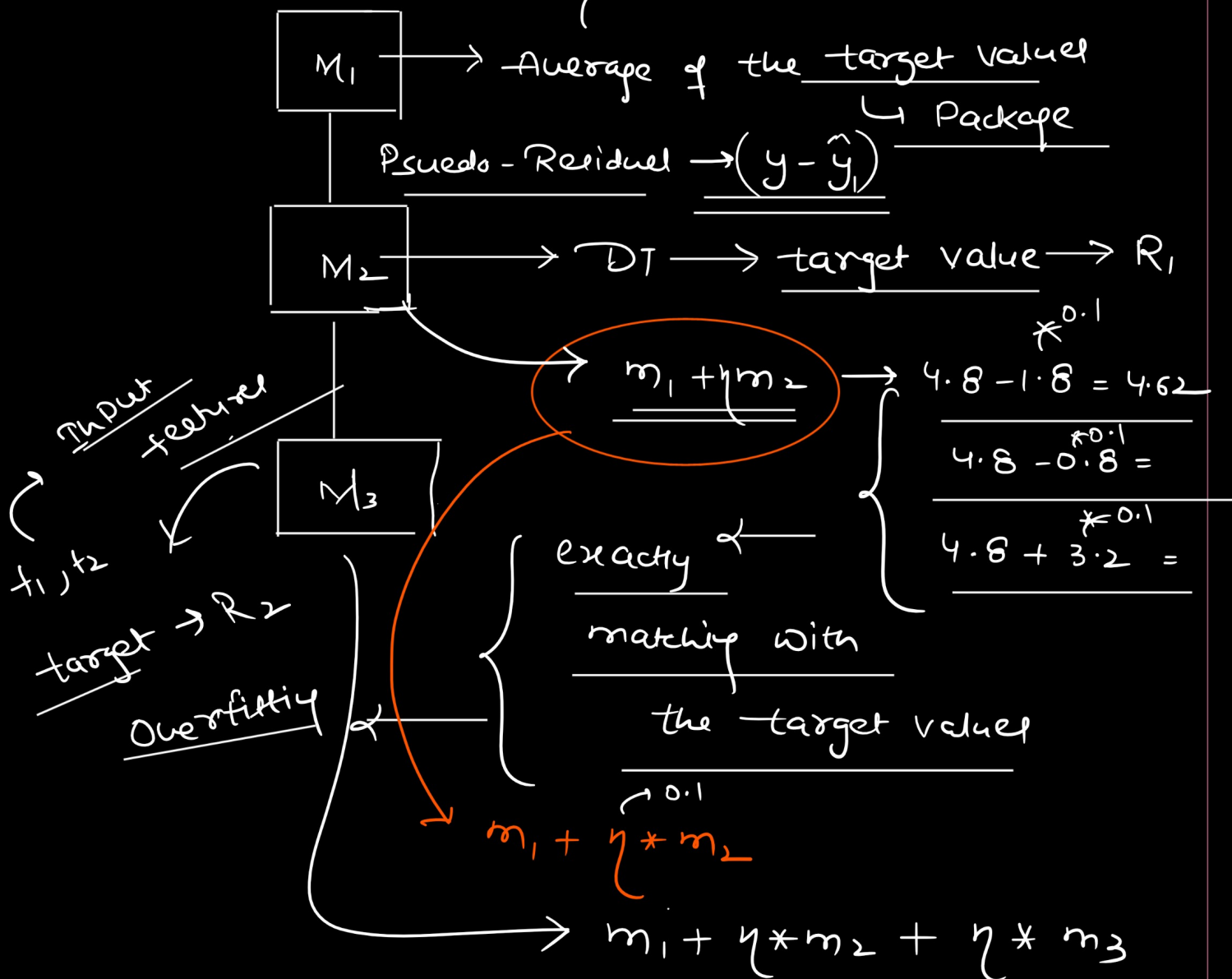


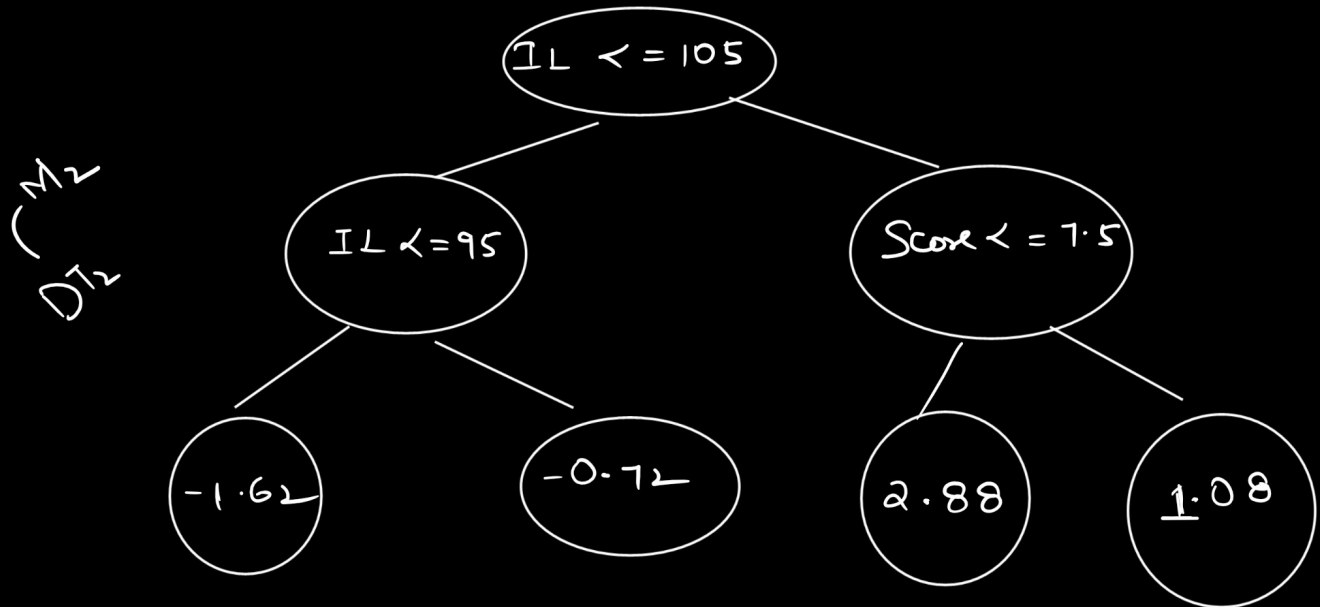
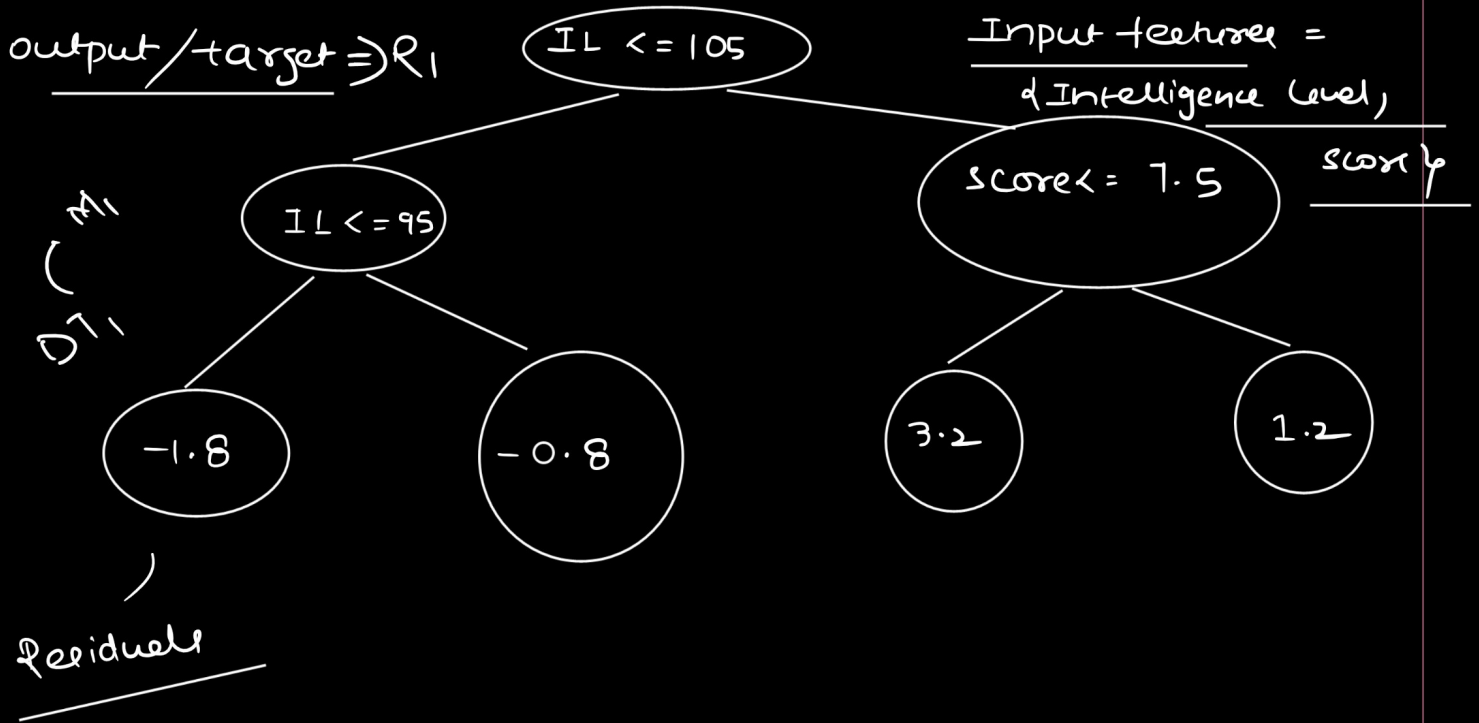
Gradient Boosting

(f1) ✓ Intelligence Level	(f2) ✓ Score	y (target) Package	Regression \hat{y}_1	R_1	\hat{y}_2	Pred_2	R_2	\hat{y}_3	Pred_3
① 90	8	3	4.8	-1.8	-1.8	4.62	-1.62	-1.62	4.458
② 100	7	4	4.8	-0.8	-0.8	4.72	-0.72	-0.72	4.648
③ 110	6	8	4.8	3.2	3.2	5.12	2.88	2.88	5.408
④ 120	9	6	4.8	1.2	1.2	4.92	1.08	1.08	5.028
⑤ 80	5	3	4.8	-1.8	-1.8	4.62	-1.62	-1.62	4.458
⑥ 115	7								

test data

$$(3+4+8+6+3)/5 \approx 4.8$$





$$M_3 \rightarrow m_1 + \eta * m_2 + \eta * m_3$$

$$\rightarrow 4.8 - 0.18 - 0.162$$

1st Record $\rightarrow 4.458$

2nd Record

$$4.8 - 0.08 - 0.072$$

$$\rightarrow \underline{4.648}$$

$$\underline{\text{3rd Record}} \rightarrow 4.8 + 0.1 \times 3.2 +$$

$$0.1 \times 2.88$$

$$\rightarrow 4.8 + 0.32 +$$

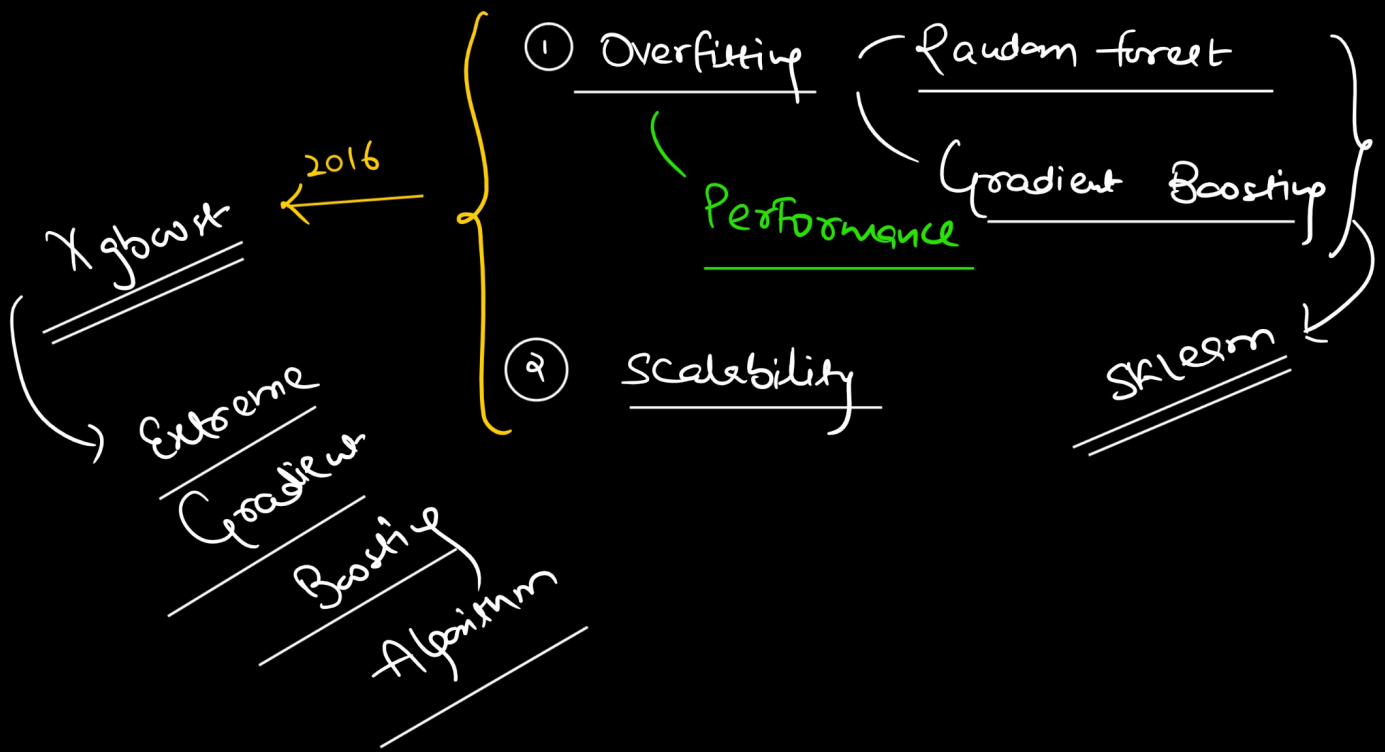
$$0.288$$

\rightarrow

$$\underline{\text{4th Record}} \rightarrow 4.8 + 0.12 + 0.108$$

$$\rightarrow 5.028$$

$$\underline{\underline{\hspace{1.5cm}}}$$



Sparse Data												
Day	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Sensor 6	Sensor 7	Sensor 8	Sensor 9	Sensor 10	Sensor 11	Sensor 12
1-Jan	0	0.89	0	0	0	0	0	0	0	0	0	0.911
2-Jan	0	0	0	0	0	0	0	0	0	0	0	0.931
3-Jan	0	0.951	0	0	0	0	0	0	0	0	0	0.951
4-Jan	0.954	0.911	0	0	0	0	0	0	0	0	0	0.899
5-Jan	0	0	0	0	0	0	0	0	0	0	0	0.897
6-Jan	0	0.899	0	0	0	0	0	0	0	0	0	0.968
7-Jan	0.895	0.911	0	0	0	0	0	0	0	0	0	0.991
8-Jan	0.911	0.962	0	0	0	0	0	0	0	0	0	0.951
9-Jan	0	0.954	0	0	0	0	0	0	0	0	0	0.898
10-Jan	0.898	0.934	0	0	0	0	0	0	0	0	0	0.962

most of the elements $\rightarrow 0$

1 Distributed computing \rightarrow Large dataset

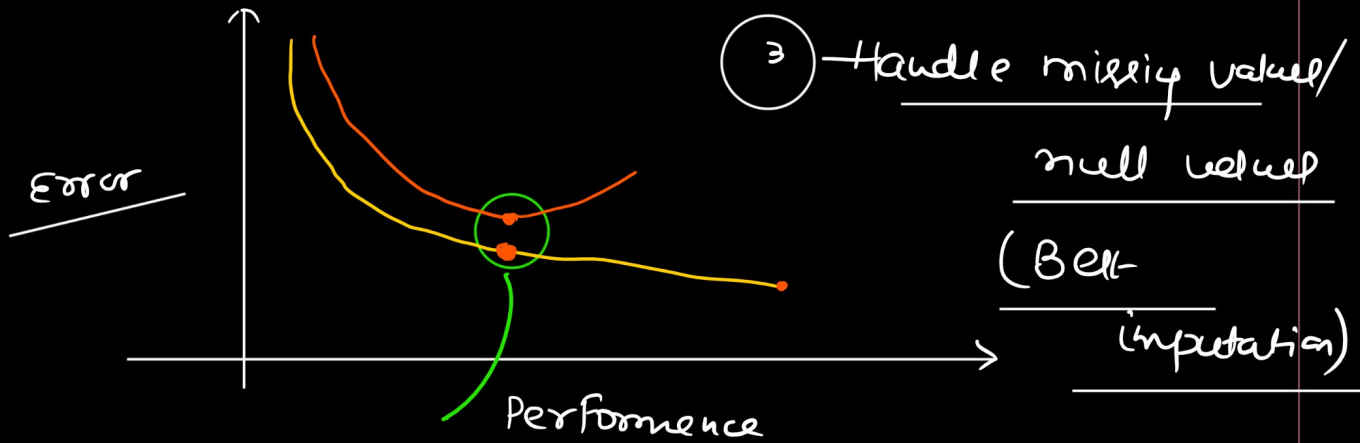
Parallel
Computation

2 Early stopping \rightarrow to prevent

overfitting

train error \rightarrow \downarrow

Validation error \rightarrow { test error \rightarrow \downarrow



Stop the training to avoid the overfitting

Kaggle \rightarrow leaderboard $\uparrow\uparrow$

6) Compatibility
 \hookrightarrow Python
Java, C++