

## Chapter 06. 프로모션 효율 예측 (Random Forest)

# 분석의 목적

- Random Forest 를 이용하여,  
프로모션에 반응할 고객을 예측
- 고객 데이터와 거래 데이터를 통합 활용



		예측값	
		0	1
실제값	0	400 TN	100 FP Type 1
	1	200 FN Type 2	300 TP

		예측값	
		0	1
실제값	0	400	100
	1	200	300

$$\text{Accuracy} = (400 + 300) / \text{All}$$

		예측값	
		0	1
실제값	0	400	100
	1	200	300

$$\text{Precision} = 300 / (100 + 300)$$

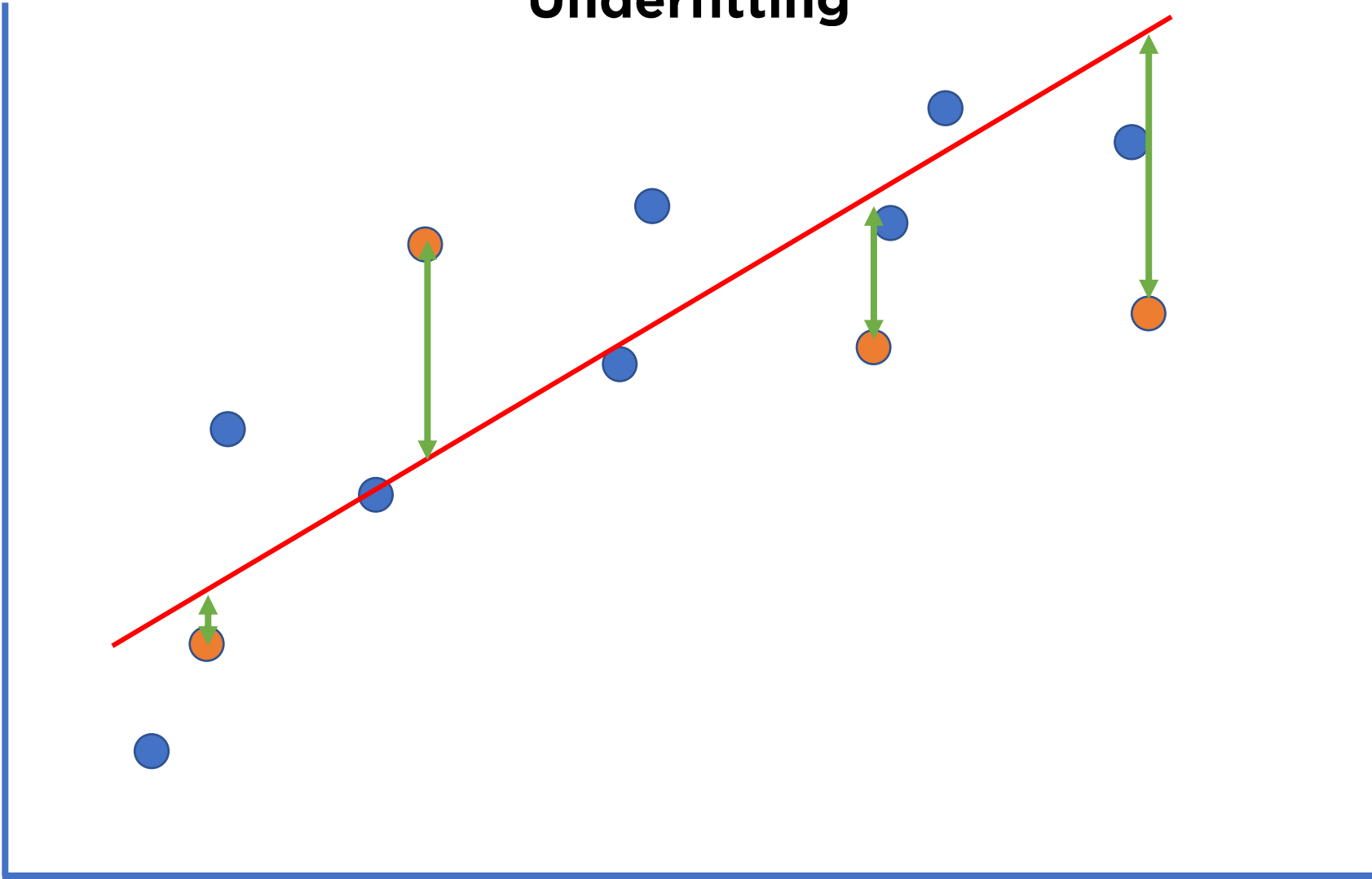
		예측값	
		0	1
실제값	0	400	100
	1	200	300

$$\text{Recall} = 300 / (200 + 300)$$

		예측값	
		0	1
실제값	0	400	100
	1	200	300

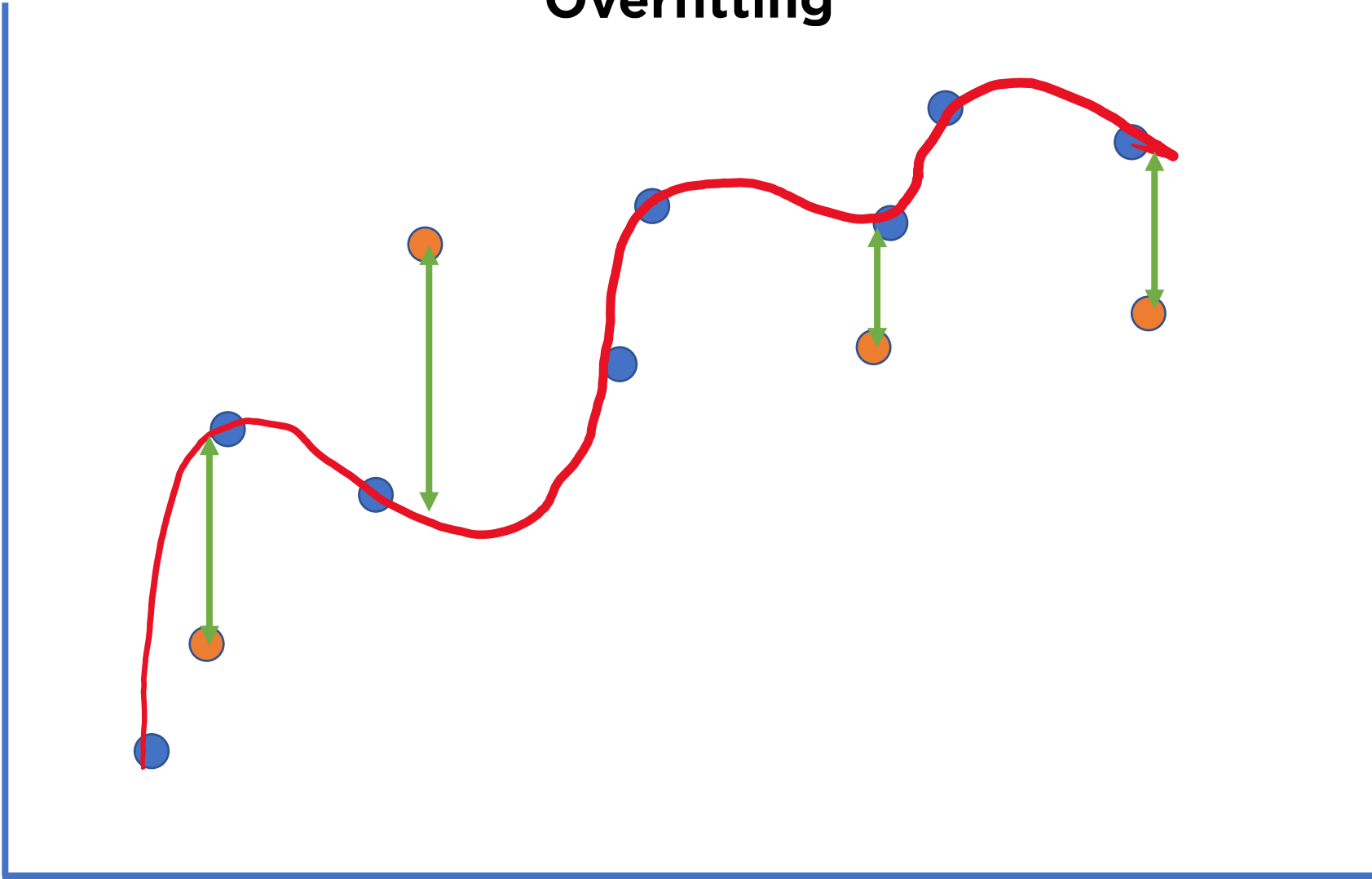
$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

# Underfitting

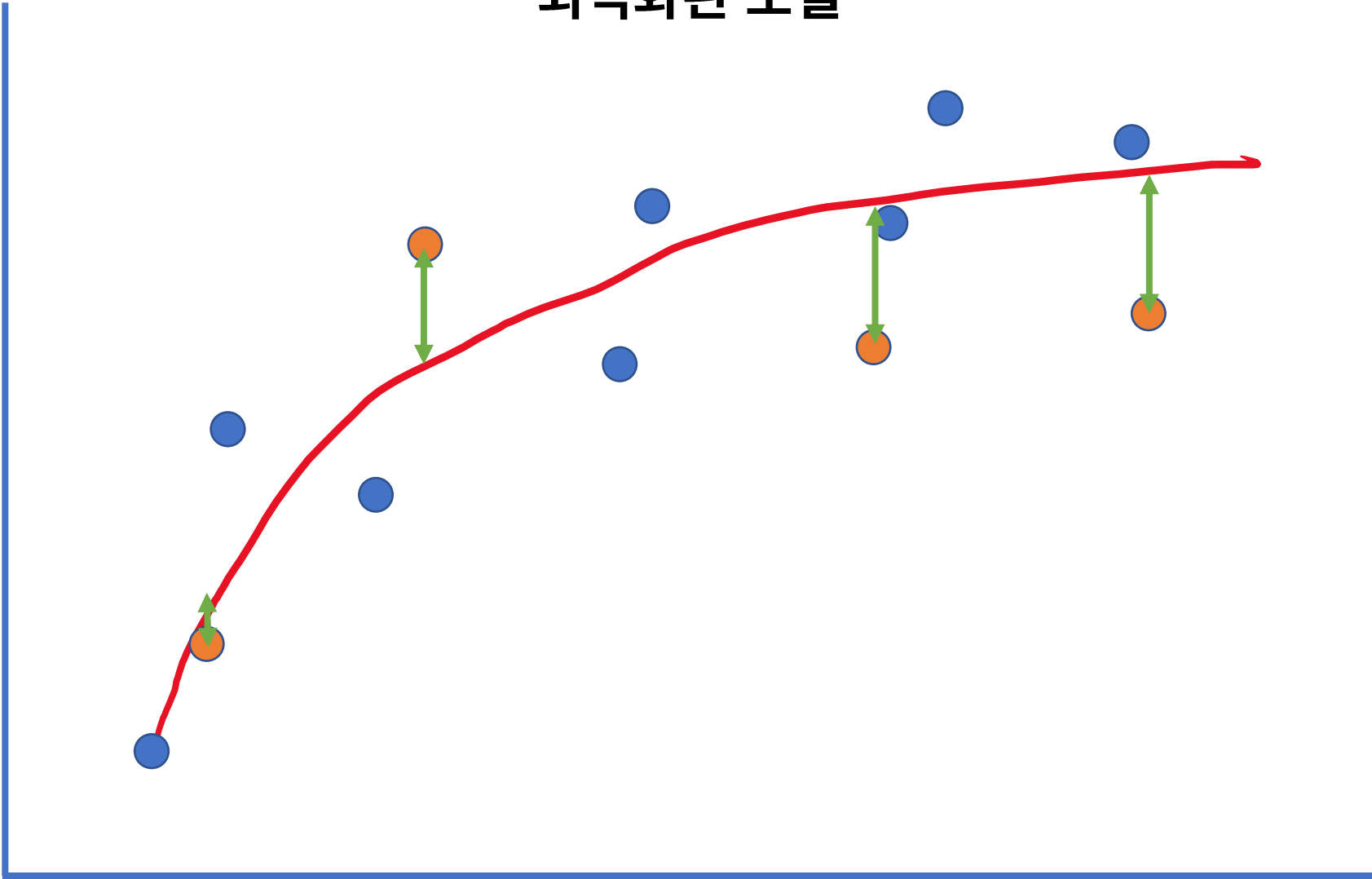




# Overfitting



# 최적화된 모델



# Bagging: 관측치를 복원 추출로 뽑아 다수의 Subset을 생성

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
937349	5	Surburban
278861	2	Urban
612950	4	Rural
413548	3	Urban
432575	5	Surburban
211216	9	Rural
798032	11	Surburban
705393	5	Surburban



id	recency	zip_code
906145	10	Surburban
184478	6	Rural
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
413548	3	Urban
432575	5	Surburban
211216	9	Rural

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
612950	4	Rural
413548	3	Urban
432575	5	Surburban
211216	9	Rural
798032	11	Surburban
705393	5	Surburban

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban

id	recency	zip_code
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
937349	5	Surburban
278861	2	Urban
612950	4	Rural
413548	3	Urban
798032	11	Surburban
705393	5	Surburban

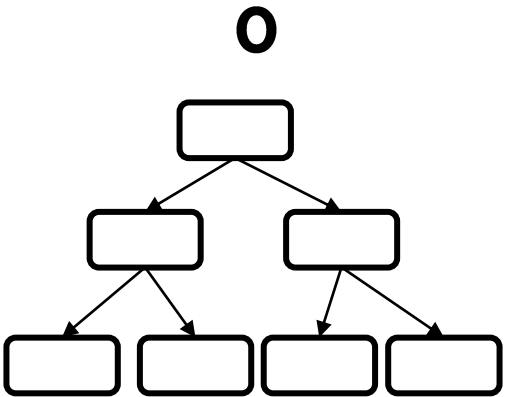
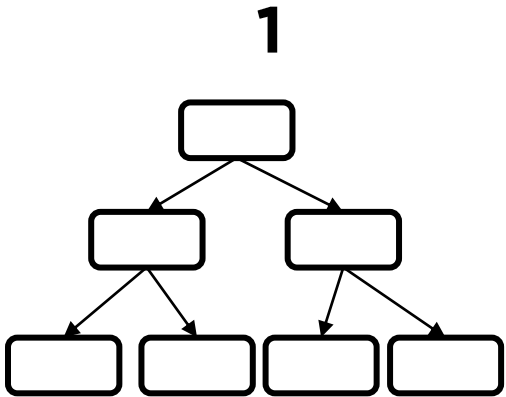
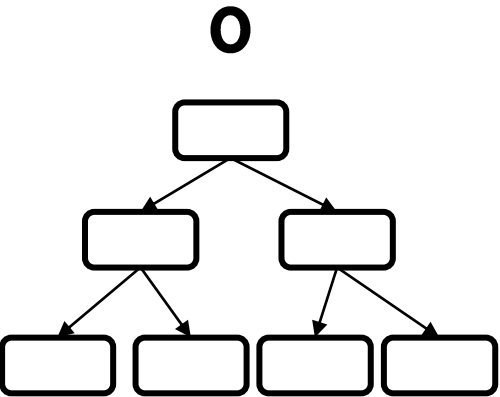
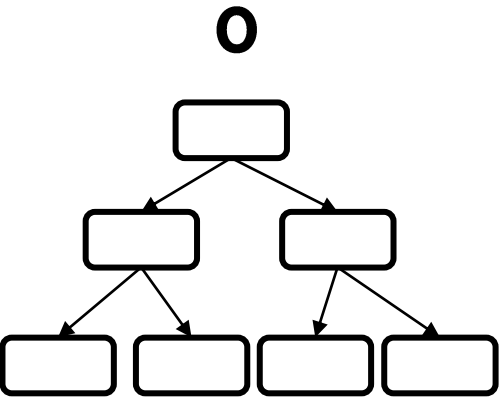
# Bagging: 각 Subset별로 별도의 독립된 트리를 만들고, 각 트리별 예측값들을 종합하여 최종예측

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
413548	3	Urban
432575	5	Surburban
211216	9	Rural

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
612950	4	Rural
413548	3	Urban
432575	5	Surburban
211216	9	Rural
798032	11	Surburban
705393	5	Surburban

id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban

id	recency	zip_code
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
937349	5	Surburban
278861	2	Urban
612950	4	Rural
413548	3	Urban
798032	11	Surburban
705393	5	Surburban



# Random Forest: Bagging과는 다르게, Variable을 샘플링하여 Subset 생성

id	recency	zip_code	is_referral	channel	Num_Item
906145	10	Surburban	0	Phone	1
184478	6	Rural	1	Web	3
394235	7	Surburban	1	Web	4
130152	9	Rural	1	Web	5
940352	2	Urban	0	Web	4
612635	6	Surburban	0	Phone	1
872298	9	Surburban	1	Phone	3
103308	9	Urban	0	Phone	4
283516	9	Rural	1	Phone	5
101354	10	Urban	1	Web	6
176547	7	Urban	1	Phone	1
663001	1	Urban	1	Phone	3
937349	5	Surburban	1	Multichannel	2
278861	2	Urban	0	Web	1
612950	4	Rural	1	Multichannel	3
413548	3	Urban	1	Web	1
432575	5	Surburban	0	Phone	4
211216	9	Rural	0	Web	5
798032	11	Surburban	0	Phone	6
705393	5	Surburban	1	Multichannel	1



id	recency	is_referral	Num_Item
906145	10	0	3
184478	6	1	4
394235	7	1	1
130152	9	1	6
940352	2	0	2
612635	6	0	1
872298	9	1	1
103308	9	0	2
283516	9	1	3



id	recency	zip_code	is_referral
906145	10	Surburban	0
184478	6	Rural	1
394235	7	Surburban	1
130152	9	Rural	1
940352	2	Urban	0
612635	6	Surburban	0
872298	9	Surburban	1
103308	9	Urban	0
283516	9	Rural	1



id	is_referral	channel	Num_Item
906145	0	Phone	3
184478	1	Web	4
394235	1	Web	1
130152	1	Web	6
940352	0	Web	7
612635	0	Phone	1
872298	1	Phone	3
103308	0	Ph	
283516	1	Ph	



id	recency	zip_code	channel
906145	10	Surburban	Phone
184478	6	Rural	Web
394235	7	Surburban	Web
130152	9	Rural	Web
940352	2	Urban	Web
612635	6	Surburban	Phone
872298	9	Surburban	Phone
103308	9	Urban	Phone
283516	9	Rural	Phone

# Random Forest: 마찬가지로 각 Subset별로 트리 생성 후, 각 트리별 예측값들을 종합하여 최종예측

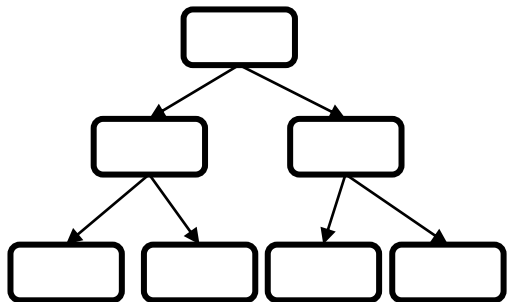
id	recency	zip_code	is_referral
906145	10	Surburban	0
184478	6	Rural	1
394235	7	Surburban	1
130152	9	Rural	1
940352	2	Urban	0
612635	6	Surburban	0
872298	9	Surburban	1
103308	9	Urban	0
283516	9	Rural	1

id	is_referral	channel	conversion
906145	0	Phone	0
184478	1	Web	0
394235	1	Web	0
130152	1	Web	0
940352	0	Web	0
612635	0	Phone	1
872298	1	Phone	0
103308	0	Phone	0
283516	1	Phone	0

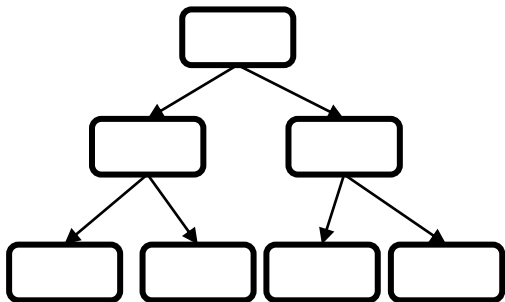
id	recency	is_referral	conversion
906145	10	0	0
184478	6	1	0
394235	7	1	0
130152	9	1	0
940352	2	0	0
612635	6	0	1
872298	9	1	0
103308	9	0	0
283516	9	1	0

id	recency	zip_code	channel
906145	10	Surburban	Phone
184478	6	Rural	Web
394235	7	Surburban	Web
130152	9	Rural	Web
940352	2	Urban	Web
612635	6	Surburban	Phone
872298	9	Surburban	Phone
103308	9	Urban	Phone
283516	9	Rural	Phone

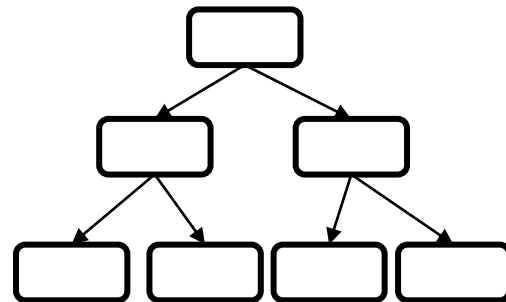
0



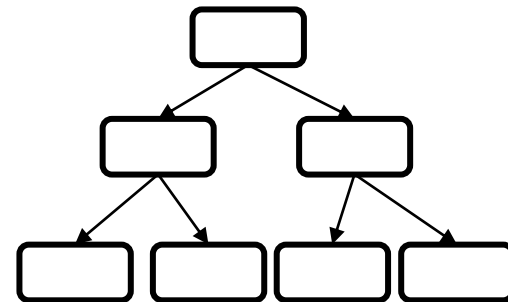
0



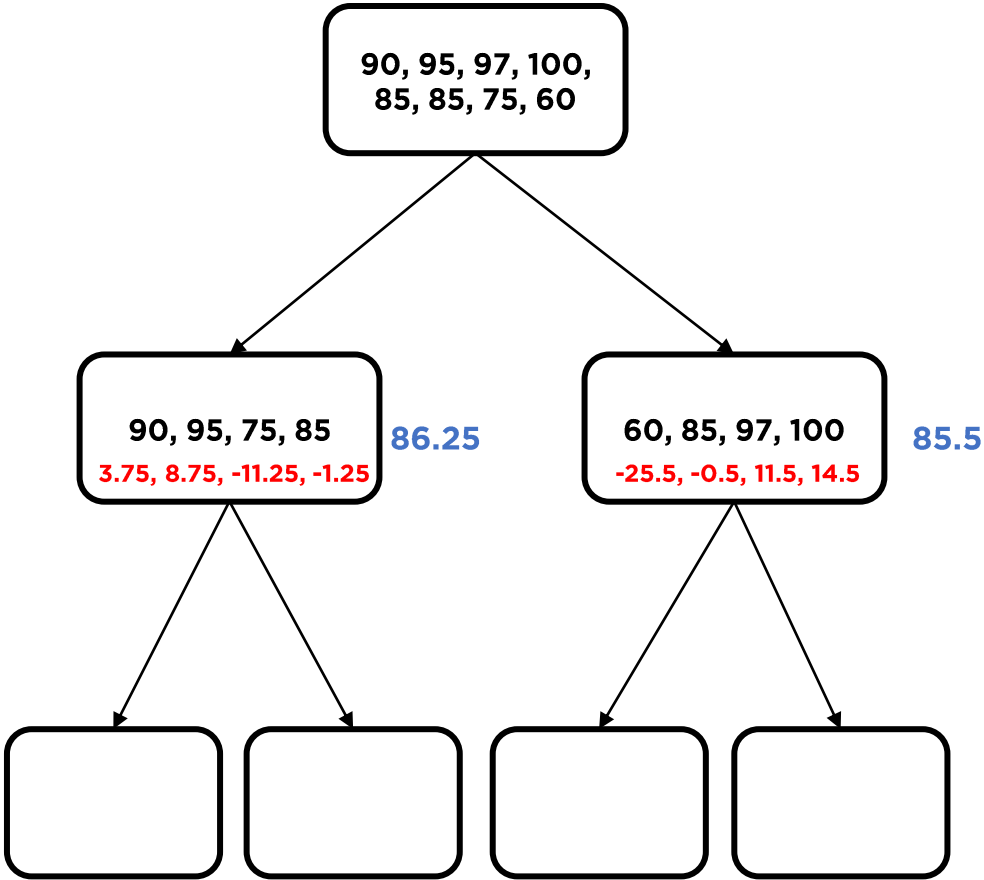
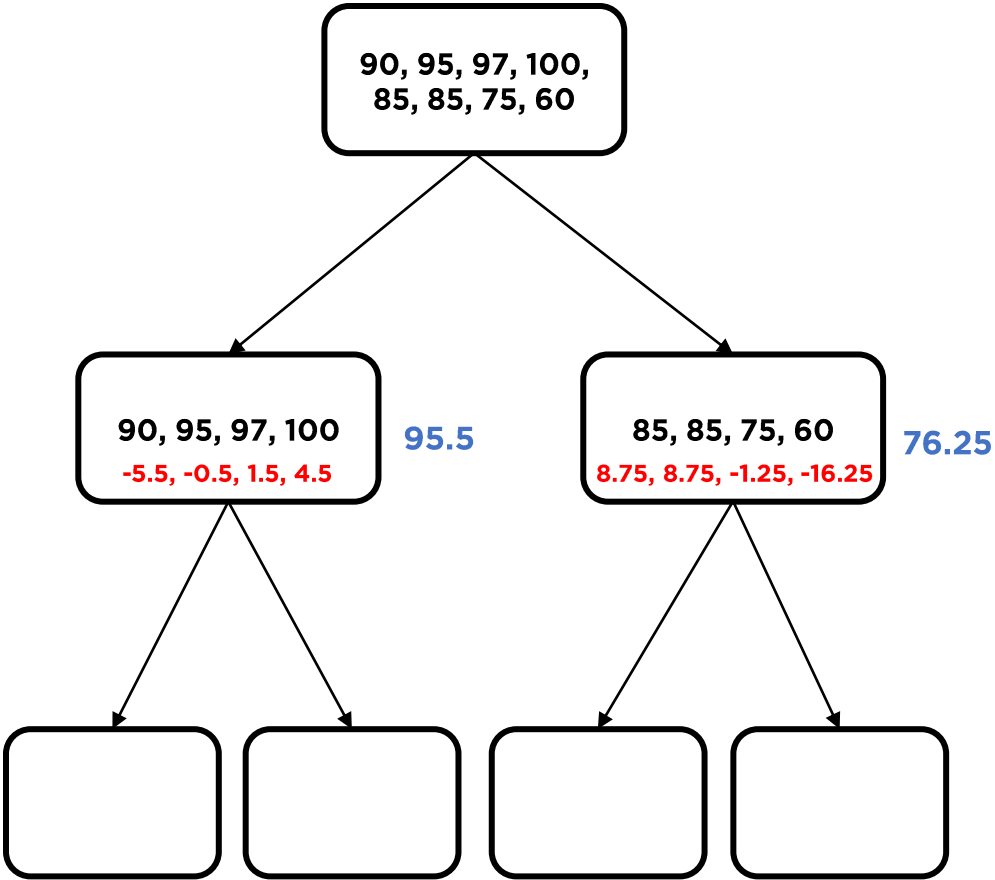
1



0



노드 내 각각의 관측치와 노드 평균과의 차이가 최소화 되는 방향으로 분리



# Regressor의 경우, 각 예측치의 평균값으로 최종 예측

id	recency	zip_code	is_referral
906145	10	Surburban	0
184478	6	Rural	1
394235	7	Surburban	1
130152	9	Rural	1
940352	2	Urban	0
612635	6	Surburban	0
872298	9	Surburban	1
103308	9	Urban	0
283516	9	Rural	1

id	is_referral	channel	conversion
906145	0	Phone	0
184478	1	Web	0
394235	1	Web	0
130152	1	Web	0
940352	0	Web	0
612635	0	Phone	1
872298	1	Phone	0
103308	0	Phone	0
283516	1	Phone	0

id	recency	is_referral	conversion
906145	10	0	0
184478	6	1	0
394235	7	1	0
130152	9	1	0
940352	2	0	0
612635	6	0	1
872298	9	1	0
103308	9	0	0
283516	9	1	0

id	recency	zip_code	channel
906145	10	Surburban	Phone
184478	6	Rural	Web
394235	7	Surburban	Web
130152	9	Rural	Web
940352	2	Urban	Web
612635	6	Surburban	Phone
872298	9	Surburban	Phone
103308	9	Urban	Phone
283516	9	Rural	Phone

