Chapter 06. 프로모션 효율 예측 (Random Forest)

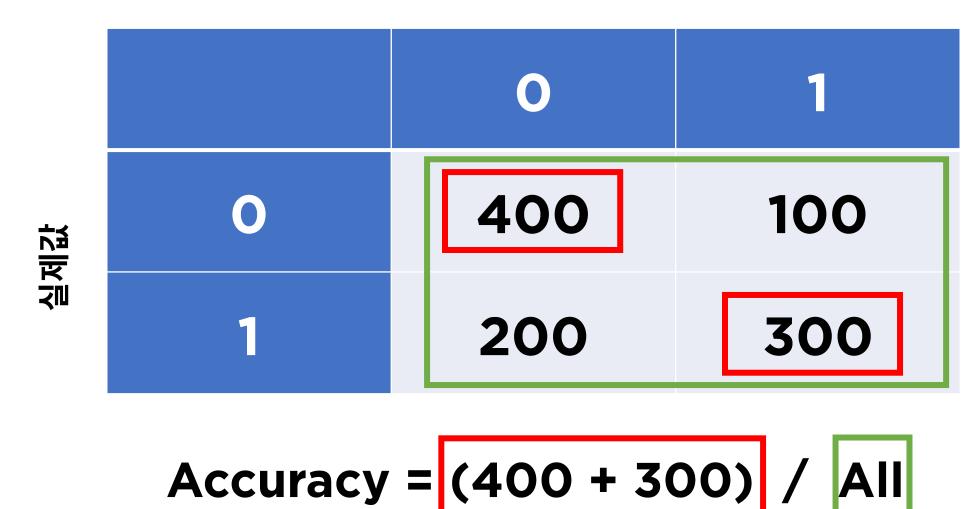
분석의 목적

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





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





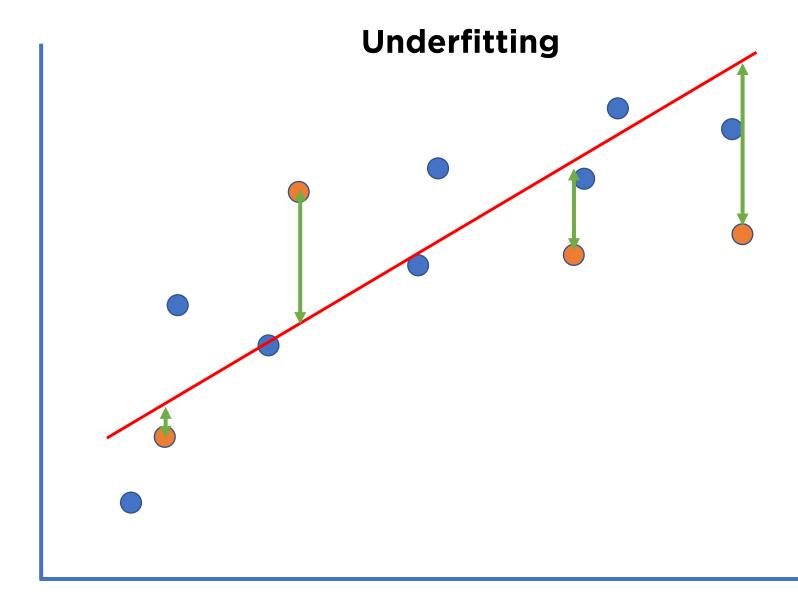
		0	1
실제값	0	400	100
ŹŪ	1	200	300

		0	1
실제값	0	400	100
	1	200	300
	Recall	= 300 / (20	00+300)

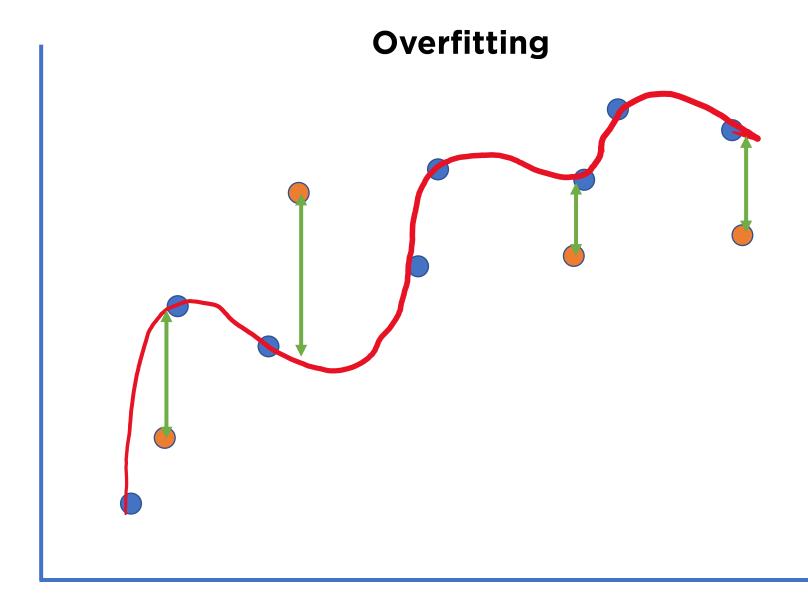


		0	1
실쪠값	0	400	100
心	1	200	300



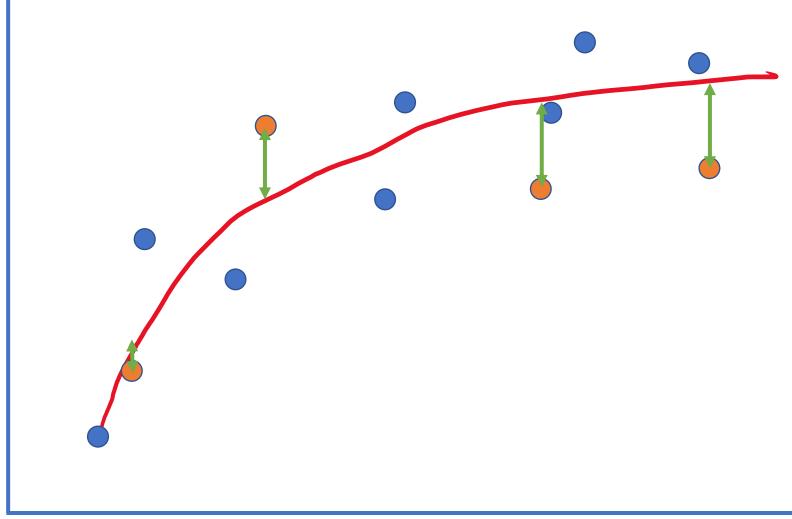














Bagging: 관측치를 복원 추출로 뽑아 다수의 Subset을 생성

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
/05535		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

	Ia	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

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



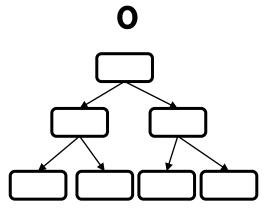
Bagging: 각 Subset별로 별도의 독립된 트리를 만들고, 각 트리별 예측값들을 종합하여 최종예측

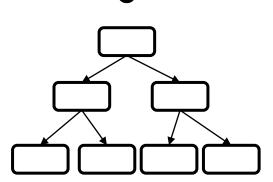
id	recency	zip_code
906145	10	Surburban
184478	6	Rural
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban
413548	3	Urban
432575	5	Surburban
211216	9	Rural

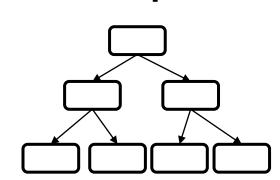
recency	zip_code
10	Surburban
6	Rural
7	Surburban
9	Rural
2	Urban
6	Surburban
4	Rural
3	Urban
5	Surburban
9	Rural
11	Surburban
5	Surburban
	10 6 7 9 2 6 4 3 5 9

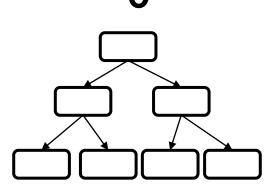
id	recency	zip_code
906145	10	Surburban
184478	6	Rural
394235	7	Surburban
130152	9	Rural
940352	2	Urban
612635	6	Surburban
872298	9	Surburban
103308	9	Urban
283516	9	Rural
101354	10	Urban
176547	7	Urban
663001	1	Urban

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

				ic	d	rece	ncy	is_referral	Num_Item
				906	145	10	0	0	3
				184	478	e	5	1	4
				394	235		7	1	1
	id	recency	zip_	code	is_re	ferral		1	6
	906145	10	Surbi	urban)	2.	0	2
	184478	6	_	ral		i		0	1
	394235	7	Surbi	urban		1		1	1
	130152	9	Ru	ral		1		0	2
	940352	2	Urk	oan)		1	3
	612635	6	Surbi	urban)			
	872298	9	Surbi	urban		1			
	103308	9	Urk	oan)			
	283516	9	Ru	ral	•	1			
I	id	ic referral	cha	nnel	Mirros	ltom			
		is_referral				_ltem_			
	906145	0	_	one		3			
	184478	1		eb		1			
	394235	1		eb		1			
	130152	1	W	eb	(5			
	940352	0	W	eb		7			
	612635	0		one		l			
7	872298	1	Pho	ne		3			
	103308	0	Pho	ic		rece		zip_code	channel
	283516	1	Pho	906		10		Surburban	Phone
-		-		184		6		Rural	Web
				394		7		Surburban	Web
				130		9		Rural	Web
				940		2		Urban	Web
				612	635	ϵ		Surburban	Phone
				872	298	9		Surburban	Phone
				103		9		Urban	Phone
				283	516	ç)	Rural	Phone



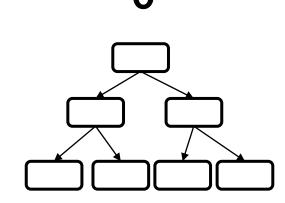
Random Forest: 마찬가지로 각 Subset별로 트리 생성 후, 각 트리별 예측값들을 종합하여 최종예측

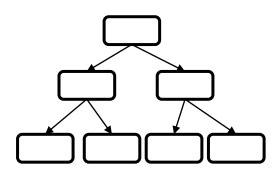
id	recency	zip_code	is_referral
906145	10	Surburban	0
184478	6	Rural	1
394235	7	Surburban	1
130152	9	Rural	1
940352	2	Urban	0
612635	6	Surburban	0
872298	9	Surburban	1
103308	9	Urban	0
283516	9	Rural	1

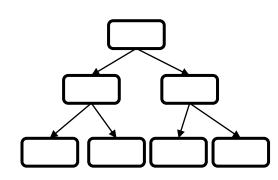
id	is referral	channel	conversion
Iu	15_referral	Chamilei	Conversion
906145	0	Phone	0
184478	1	Web	0
394235	1	Web	0
130152	1	Web	0
940352	0	Web	0
612635	0	Phone	1
872298	1	Phone	0
103308	0	Phone	0
283516	1	Phone	0

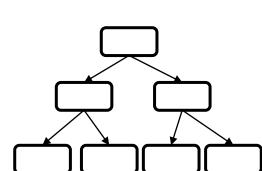
id	recency	is_referral	conversion
906145	10	0	0
184478	6	1	0
394235	7	1	0
130152	9	1	0
940352	2	0	0
612635	6	0	1
872298	9	1	0
103308	9	0	0
283516	9	1	0

recency	zip_code	channel
10	Surburban	Phone
6	Rural	Web
7	Surburban	Web
9	Rural	Web
2	Urban	Web
6	Surburban	Phone
9	Surburban	Phone
9	Urban	Phone
9	Rural	Phone
	10 6 7 9 2 6 9	10 Surburban 6 Rural 7 Surburban 9 Rural 2 Urban 6 Surburban 9 Surburban 9 Urban



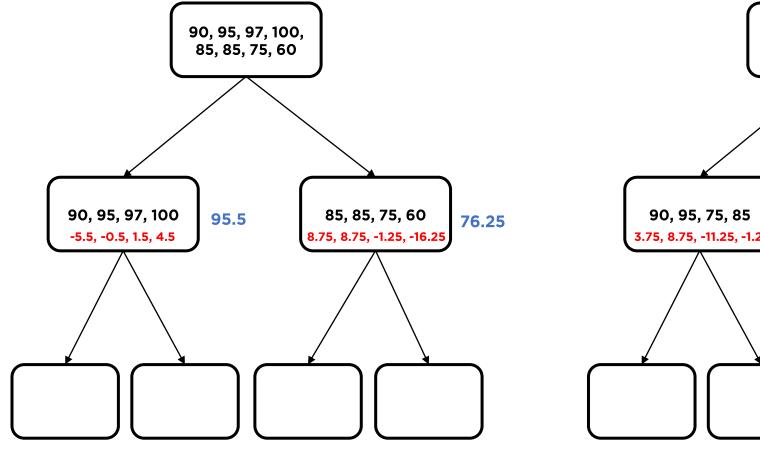


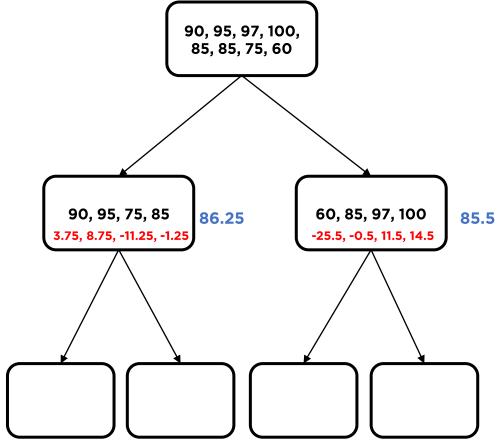






노드 내 각각의 관측치와 노드 평균과의 차이가 최소화 되는 방향으로 분리





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

