

Feature Generation

Feature Generation

Feature generation is creating new features from the existing data

Feature Generation: Binning

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Binning

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Binning

id	gender	age	stroke
30468	Male	58	1
16523	Female	8	0
56543	Female	70	0
46136	Male	14	0
32257	Female	47	0
52800	Female	52	0
41413	Female	75	0
15266	Female	32	0
28674	Female	74	1
10460	Female	79	0
64908	Male	79	0
63884	Female	37	0
37893	Female	37	1
67855	Female	40	0
25774	Male	35	1

Feature Generation: Binning

id	gender	age	stroke
30468	Male	58	1
16523	Female	8	0
56543	Female	70	0
46136	Male	14	0
32257	Female	47	0
52800	Female	52	0
41413	Female	75	0
15266	Female	32	0
28674	Female	74	1
10460	Female	79	0
64908	Male	79	0
63884	Female	37	0
37893	Female	37	1
67855	Female	40	0
25774	Male	35	1

Method
1

BIN 1
Age > 35

Bin 2
Age ≤ 35

Feature Generation: Binning

id	gender	age	stroke	age_35
30468	Male	58	1	1
16523	Female	8	0	0
56543	Female	70	0	1
46136	Male	14	0	0
32257	Female	47	0	1
52800	Female	52	0	1
41413	Female	75	0	1
15266	Female	32	0	0
28674	Female	74	1	1
10460	Female	79	0	1
64908	Male	79	0	1
63884	Female	37	0	1
37893	Female	37	1	1
67855	Female	40	0	1
25774	Male	35	1	0

Method
1

BIN 1
Age > 35

Bin 2
Age ≤ 35

Feature Generation: Binning

id	gender	age	stroke
30468	Male	58	1
16523	Female	8	0
56543	Female	70	0
46136	Male	14	0
32257	Female	47	0
52800	Female	52	0
41413	Female	75	0
15266	Female	32	0
28674	Female	74	1
10460	Female	79	0
64908	Male	79	0
63884	Female	37	0
37893	Female	37	1
67855	Female	40	0
25774	Male	35	1

Method
2

BIN 1
 $0 < \text{Age} < 12$

BIN 3
 $20 \leq \text{Age} < 60$

Bin 2
 $12 \leq \text{Age} < 20$

Bin 4
 $60 \leq \text{Age} < 100$

Feature Generation: Binning

id	gender	age	stroke	age_gourp
30468	Male	58	1	adult
16523	Female	8	0	child
56543	Female	70	0	senior
46136	Male	14	0	teenager
32257	Female	47	0	adult
52800	Female	52	0	adult
41413	Female	75	0	senior
15266	Female	32	0	adult
28674	Female	74	1	senior
10460	Female	79	0	senior
64908	Male	79	0	senior
63884	Female	37	0	adult
37893	Female	37	1	adult
67855	Female	40	0	adult
25774	Male	35	1	adult

Method
2

BIN 1
 $0 < \text{Age} < 12$

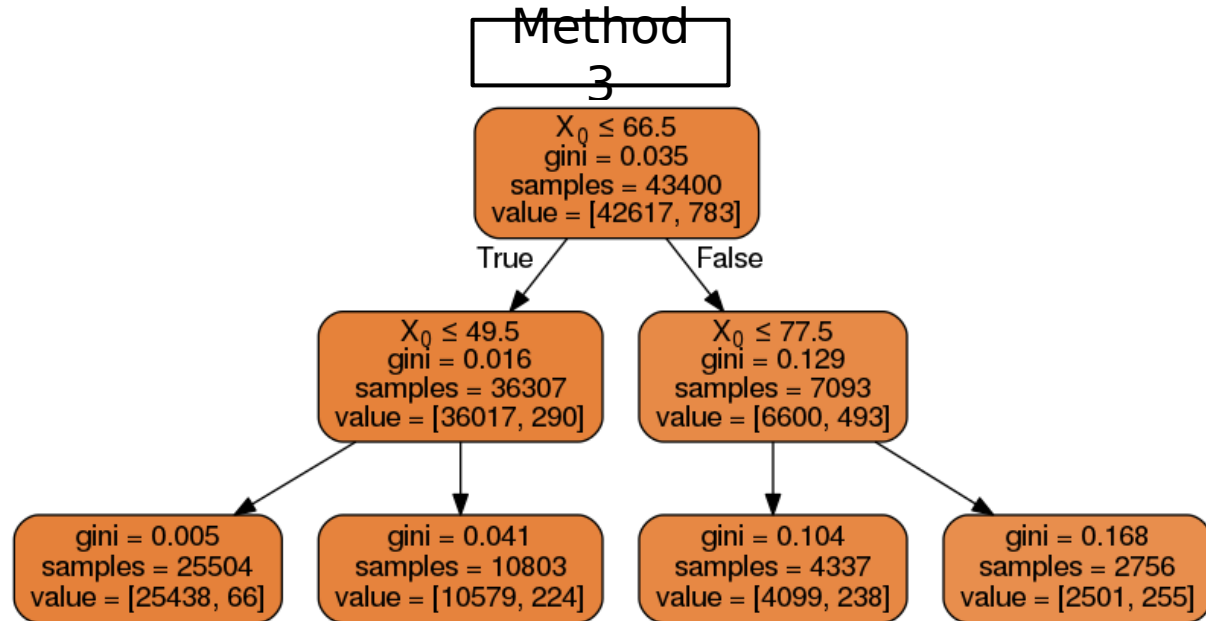
BIN 3
 $20 \leq \text{Age} < 60$

Bin 2
 $12 \leq \text{Age} < 20$

Bin 4
 $60 \leq \text{Age} < 100$

Feature Generation: Binning

id	gender	age	stroke
30468	Male	58	1
16523	Female	8	0
56543	Female	70	0
46136	Male	14	0
32257	Female	47	0
52800	Female	52	0
41413	Female	75	0
15266	Female	32	0
28674	Female	74	1
10460	Female	79	0
64908	Male	79	0
63884	Female	37	0
37893	Female	37	1
67855	Female	40	0
25774	Male	35	1



Feature Generation: Binning

id	gender	age	stroke
30468	Male	58	1
16523	Female	8	0
56543	Female	70	0
46136	Male	14	0
32257	Female	47	0
52800	Female	52	0
41413	Female	75	0
15266	Female	32	0
28674	Female	74	1
10460	Female	79	0
64908	Male	79	0
63884	Female	37	0
37893	Female	37	1
67855	Female	40	0
25774	Male	35	1

Method
3

BIN 1
 $0 < \text{Age} < 49.5$

BIN 3
 $66.5 \leq \text{Age} < 77.5$

Bin 2
 $49.5 \leq \text{Age} < 66.5$

Bin 4
 $77.5 \leq \text{Age} < 100$

Feature Generation: Binning

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Binning

Feature Generation

Feature Generation

Loan_ID	Gender	Married	Education	ApplicantIncome	CoapplicantIncome	LoanAmount	Property_Area	Loan_Status
LP001002	Male	No	Graduate	5849	0		Urban	Y
LP001003	Male	Yes	Graduate	4583	1508	128	Rural	N
LP001005	Male	Yes	Graduate	3000	0	66	Urban	Y
LP001006	Male	Yes	Not Graduate	2583	2358	120	Urban	Y
LP001008	Male	No	Graduate	6000	0	141	Urban	Y
LP001011	Male	Yes	Graduate	5417	4196	267	Urban	Y
LP001013	Male	Yes	Not Graduate	2333	1516	95	Urban	Y
LP001014	Male	Yes	Graduate	3036	2504	158	Semiurban	N
LP001018	Male	Yes	Graduate	4006	1526	168	Urban	Y
LP001020	Male	Yes	Graduate	12841	10968	349	Semiurban	N
LP001024	Male	Yes	Graduate	3200	700	70	Urban	Y
LP001027	Male	Yes	Graduate	2500	1840	109	Urban	Y
LP001028	Male	Yes	Graduate	3073	8106	200	Urban	Y
LP001029	Male	No	Graduate	1853	2840	114	Rural	N
LP001030	Male	Yes	Graduate	1299	1086	17	Urban	Y

Feature Generation

Loan_ID	Gender	Married	Education	ApplicantIncome	CoapplicantIncome	LoanAmount	Property_Area	Loan_Status
LP001002	Male	No	Graduate	5849	0		Urban	Y
LP001003	Male	Yes	Graduate	4583	1508	128	Rural	N
LP001005	Male	Yes	Graduate	3000	0	66	Urban	Y
LP001006	Male	Yes	Not Graduate	2583	2358	120	Urban	Y
LP001008	Male	No	Graduate	6000	0	141	Urban	Y
LP001011	Male	Yes	Graduate	5417	4196	267	Urban	Y
LP001013	Male	Yes	Not Graduate	2333	1516	95	Urban	Y
LP001014	Male	Yes	Graduate	3036	2504	158	Semiurban	N
LP001018	Male	Yes	Graduate	4006	1526	168	Urban	Y
LP001020	Male	Yes	Graduate	12841	10968	349	Semiurban	N
LP001024	Male	Yes	Graduate	3200	700	70	Urban	Y
LP001027	Male	Yes	Graduate	2500	1840	109	Urban	Y
LP001028	Male	Yes	Graduate	3073	8106	200	Urban	Y
LP001029	Male	No	Graduate	1853	2840	114	Rural	N
LP001030	Male	Yes	Graduate	1299	1086	17	Urban	Y

Feature Generation

Loan_ID	Gender	Married	Education	ApplicantIncome	CoapplicantIncome	LoanAmount	Property_Area	Loan_Status
LP001002	Male	No	Graduate	5849	0		Urban	Y
LP001003	Male	Yes	Graduate	4583	1508	128	Rural	N
LP001005	Male	Yes	Graduate	3000	0	66	Urban	Y
LP001006	Male	Yes	Not Graduate	2583	2358	120	Urban	Y
LP001008	Male	No	Graduate	6000	0	141	Urban	Y
LP001011	Male	Yes	Graduate	5417	4196	267	Urban	Y
LP001013	Male	Yes	Not Graduate	2333	1516	95	Urban	Y
LP001014	Male	Yes	Graduate	3036	2504	158	Semiurban	N
LP001018	Male	Yes	Graduate	4006	1526	168	Urban	Y
LP001020	Male	Yes	Graduate	12841	10968	349	Semiurban	N
LP001024	Male	Yes	Graduate	3200	700	70	Urban	Y
LP001027	Male	Yes	Graduate	2500	1840	109	Urban	Y
LP001028	Male	Yes	Graduate	3073	8106	200	Urban	Y
LP001029	Male	No	Graduate	1853	2840	114	Rural	N
LP001030	Male	Yes	Graduate	1299	1086	17	Urban	Y

Addition

Feature Generation: Feature Interaction

Loan_ID	Gender	Married	Education	ApplicantIncome	CoapplicantIncome	LoanAmount	Property_Area	Loan_Status
LP001002	Male	No	Graduate	5849	0		Urban	Y
LP001003	Male	Yes	Graduate	4583	1508	128	Rural	N
LP001005	Male	Yes	Graduate	3000	0	66	Urban	Y
LP001006	Male	Yes	Not Graduate	2583	2358	120	Urban	Y
LP001008	Male	No	Graduate	6000	0	141	Urban	Y
LP001011	Male	Yes	Graduate	5417	4196	267	Urban	Y
LP001013	Male	Yes	Not Graduate	2333	1516	95	Urban	Y
LP001014	Male	Yes	Graduate	3036	2504	158	Semiurban	N
LP001018	Male	Yes	Graduate	4006	1526	168	Urban	Y
LP001020	Male	Yes	Graduate	12841	10968	349	Semiurban	N
LP001024	Male	Yes	Graduate	3200	700	70	Urban	Y
LP001027	Male	Yes	Graduate	2500	1840	109	Urban	Y
LP001028	Male	Yes	Graduate	3073	8106	200	Urban	Y
LP001029	Male	No	Graduate	1853	2840	114	Rural	N
LP001030	Male	Yes	Graduate	1299	1086	17	Urban	Y

Addition

Ratio

Feature Generation: Feature Interaction

Loan_ID	Gender	Married	Education	ApplicantIncome	CoapplicantIncome	LoanAmount	Property_Area	Loan_Status
LP001002	Male	No	Graduate	5849	0		Urban	Y
LP001003	Male	Yes	Graduate	4583	1508	128	Rural	N
LP001005	Male	Yes	Graduate	3000	0	66	Urban	Y
LP001006	Male	Yes	Not Graduate	2583	2358	120	Urban	Y
LP001008	Male	No	Graduate	6000	0	141	Urban	Y
LP001011	Male	Yes	Graduate	5417	4196	267	Urban	Y
LP001013	Male	Yes	Not Graduate	2333	1516	95	Urban	Y
LP001014	Male	Yes	Graduate	3036	2504	158	Semiurban	N
LP001018	Male	Yes	Graduate	4006	1526	168	Urban	Y
LP001020	Male	Yes	Graduate	12841	10968	349	Semiurban	N
LP001024	Male	Yes	Graduate	3200	700	70	Urban	Y
LP001027	Male	Yes	Graduate	2500	1840	109	Urban	Y
LP001028	Male	Yes	Graduate	3073	8106	200	Urban	Y
LP001029	Male	No	Graduate	1853	2840	114	Rural	N
LP001030	Male	Yes	Graduate	1299	1086	17	Urban	Y

Addition

Ratio

Difference

e

Feature Generation: Missing Values

Feature Generation: Missing Value

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Missing Value

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Missing Value

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	smoking_status_nan	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	0	1
16523	Female	8	0	0	No	Private	Urban	17.6		1	0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		1	0
32257	Female	47	0	0	Yes	Private	Urban	50.1		1	0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	0	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		1	0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	0	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	0	1

Feature Generation: Frequency Encoding

Feature Generation: Frequency Encoding

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

Feature Generation: Frequency Encoding

id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	bmi	smoking_status	stroke
30468	Male	58	1	0	Yes	Private	Urban	39.2	never smoked	1
16523	Female	8	0	0	No	Private	Urban	17.6		0
56543	Female	70	0	0	Yes	Private	Rural	35.9	formerly smoked	0
46136	Male	14	0	0	No	Never_worked	Rural	19.1		0
32257	Female	47	0	0	Yes	Private	Urban	50.1		0
52800	Female	52	0	0	Yes	Private	Urban	17.7	formerly smoked	0
41413	Female	75	0	1	Yes	Self-employed	Rural	27	never smoked	0
15266	Female	32	0	0	Yes	Private	Rural	32.3	smokes	0
28674	Female	74	1	0	Yes	Self-employed	Urban	54.6	never smoked	1
10460	Female	79	0	0	Yes	Govt_job	Urban	35		0
64908	Male	79	0	1	Yes	Private	Urban	22	formerly smoked	0
63884	Female	37	0	0	Yes	Private	Rural	39.4	never smoked	0
37893	Female	37	0	0	Yes	Private	Rural	26.1	formerly smoked	1
67855	Female	40	0	0	Yes	Private	Rural	42.4	never smoked	0
25774	Male	35	0	0	No	Private	Rural	33	never smoked	1

work_type	work_type_freq
Private	0.6
Private	0.6
Private	0.6
Never_worked	0.1
Private	0.6
Private	0.6
Self-employed	0.2
Private	0.6
Self-employed	0.2
Govt_job	0.1
Private	0.6
Private	0.6
Private	0.6
Private	0.6
Private	0.6