



Evaluation metrics in dataframe



loan approval dataset 😊

☐ original data

1.1 original dataframe

	loan_id	no_of_dependents	education	self_employed	income_annum	loan_amount	loan
0	1	2	Graduate	No	9600000	29900000	
1	2	0	Not Graduate	Yes	4100000	12200000	
2	3	3	Graduate	No	9100000	29700000	
3	4	3	Graduate	No	8200000	30700000	
4	5	5	Not Graduate	Yes	9800000	24200000	

☐ label encoding

ML models with label encoding

	_id	sampler	model_type	Evaluation	accuracy
0	6650ef4eeb166451c4fb31f4	no sampler	LogisticRegression	train	0.637000
1	6650ef4eeb166451c4fb31f5	no sampler	LogisticRegression	test	0.639000
2	6650ef4eeb166451c4fb31f6	no sampler	KNeighborsClassifier	train	0.721000
3	6650ef4eeb166451c4fb31f7	no sampler	KNeighborsClassifier	test	0.552000
4	6650ef4feb166451c4fb31f8	no sampler	SVC	train	0.625000
5	6650ef4feb166451c4fb31f9	no sampler	SVC	test	0.615000
6	6650ef4feb166451c4fb31fa	no sampler	DecisionTreeClassifier	train	1.000000
7	6650ef4feb166451c4fb31fb	no sampler	DecisionTreeClassifier	test	0.981000
8	6650ef4feb166451c4fb31fc	no sampler	RandomForestClassifier	train	1.000000
9	6650ef4feb166451c4fb31fd	no sampler	RandomForestClassifier	test	0.981000

☐ dummies

ML models with dummies

	_id	sampler	model_type	Evaluation	accuracy
7	6650ef57eb166451c4fb3226	no sampler	DecisionTreeClassifier	test	0.982000
8	6650ef57eb166451c4fb3227	no sampler	RandomForestClassifier	train	1.000000
9	6650ef57eb166451c4fb3228	no sampler	RandomForestClassifier	test	0.989000
10	6650ef57eb166451c4fb3229	no sampler	AdaBoostClassifier	train	0.980000
11	6650ef57eb166451c4fb322a	no sampler	AdaBoostClassifier	test	0.977000
12	6650ef58eb166451c4fb322b	no sampler	GradientBoostingClassifier	train	0.996000
13	6650ef58eb166451c4fb322c	no sampler	GradientBoostingClassifier	test	0.988000
14	6650ef58eb166451c4fb322d	RandomUnderSampler	LogisticRegression	train	0.750000
15	6650ef58eb166451c4fb322e	RandomUnderSampler	LogisticRegression	test	0.748000
16	6650ef58eb166451c4fb322f	RandomUnderSampler	KNeighborsClassifier	train	0.713000