



01

SDAIA ACADEMY

T5 Boot Camp

LOGISTIC REGRESSION PRESENTATION

Supervised by: Mr. Mohamed El-Deeb

Logistic Regression module

AGENDA

- Goal
- Data Collection
- EDA
- Data Cleaning
- Pre-Processing
- Modeling
- Models Evaluation

GOAL

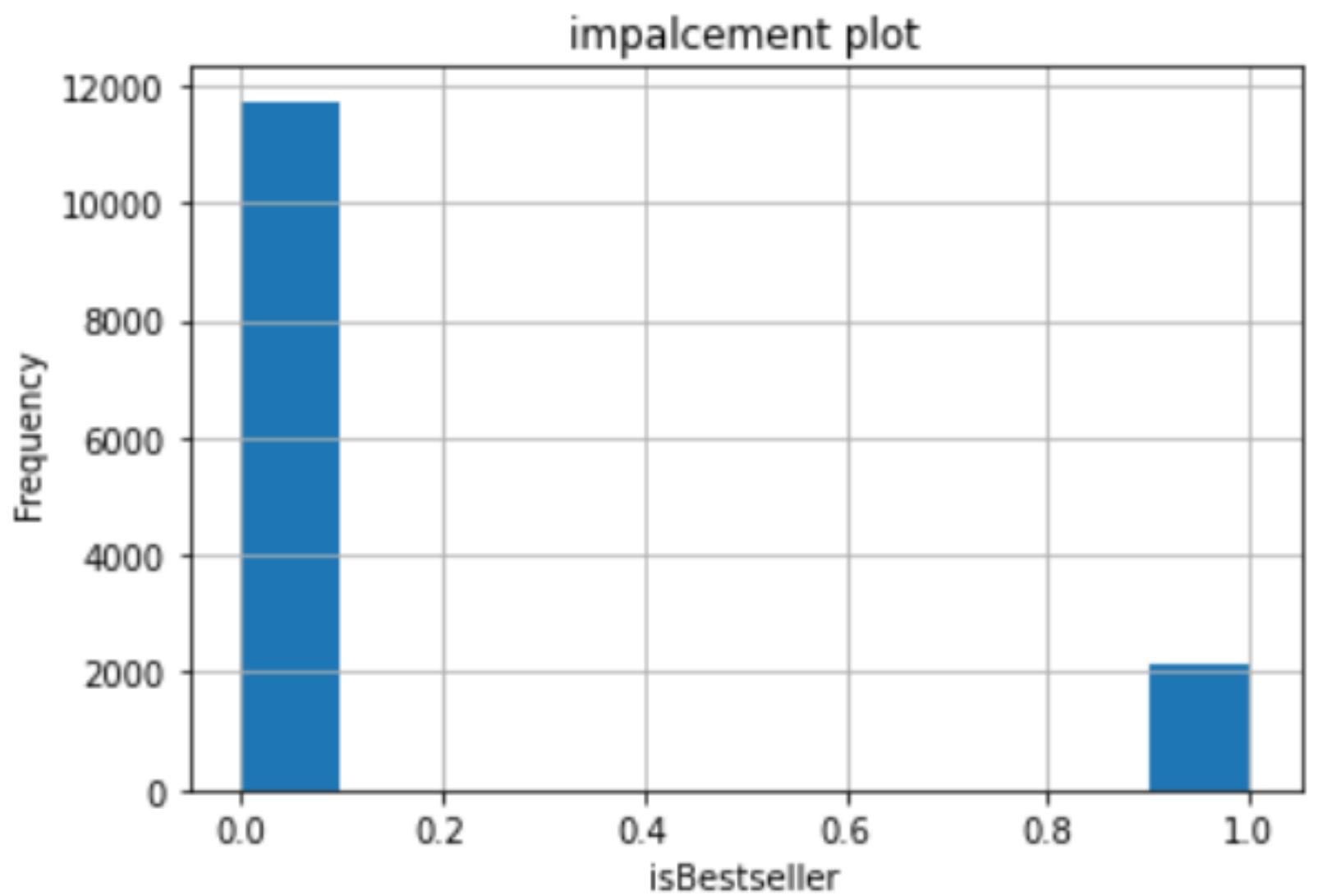
- Predicting the isBeaseller.

DATA COLLECTION

- Web Sribing "Udemy"
- 12 Features --> 8 Features
- Almost 20K Record --> 13K "en"

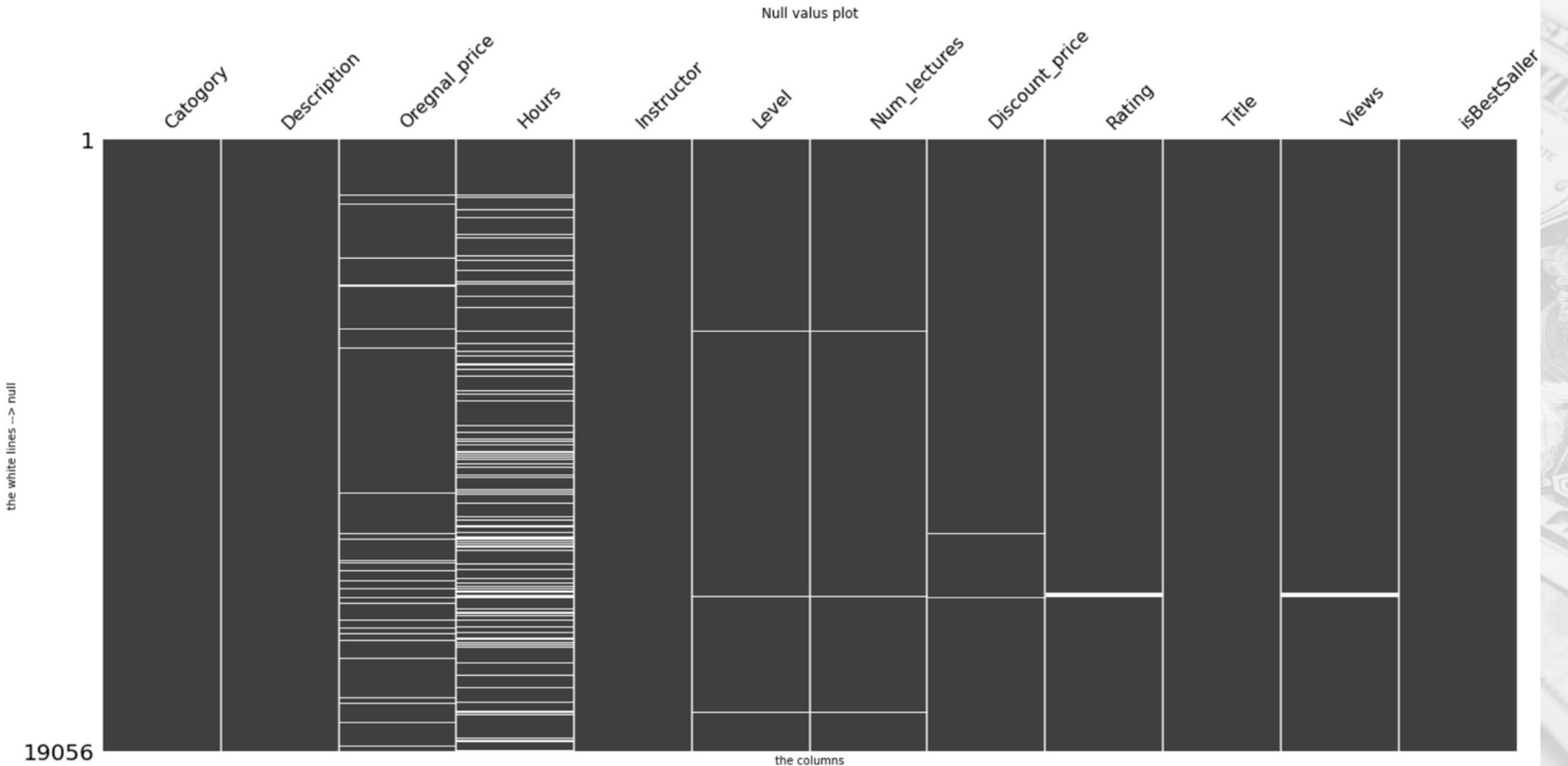
EDA

- impalcement checking



EDA

- Null Values

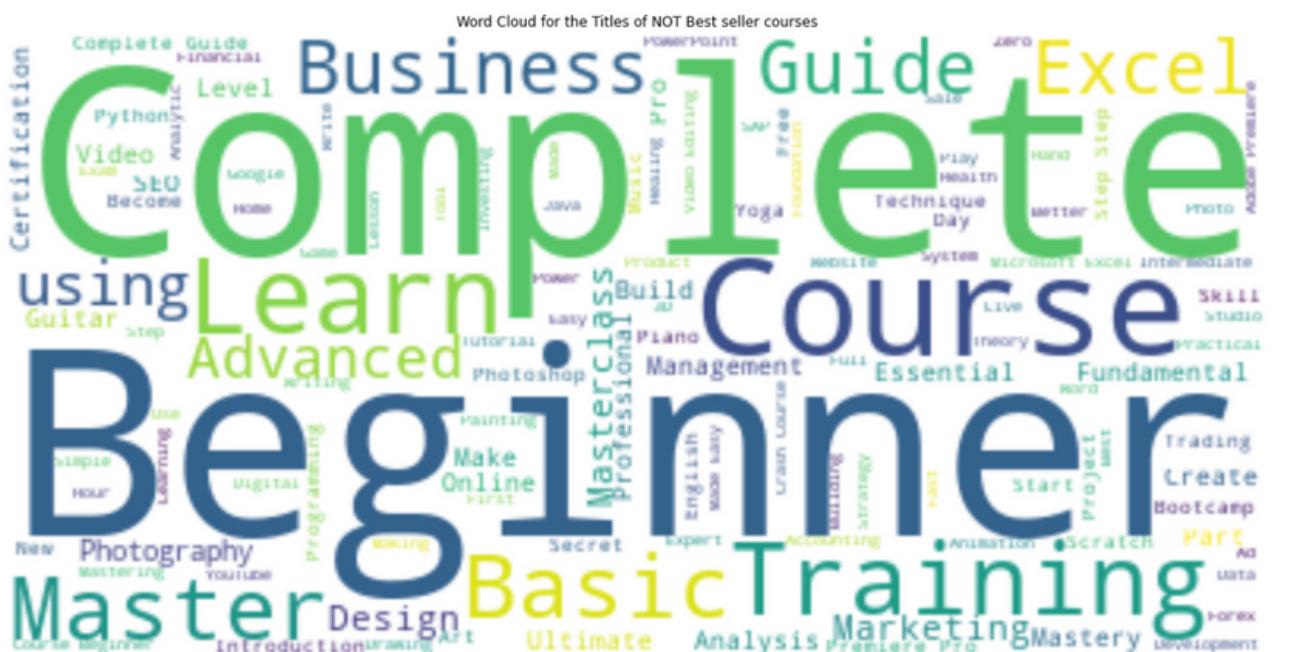


EDA

- Title word cloud



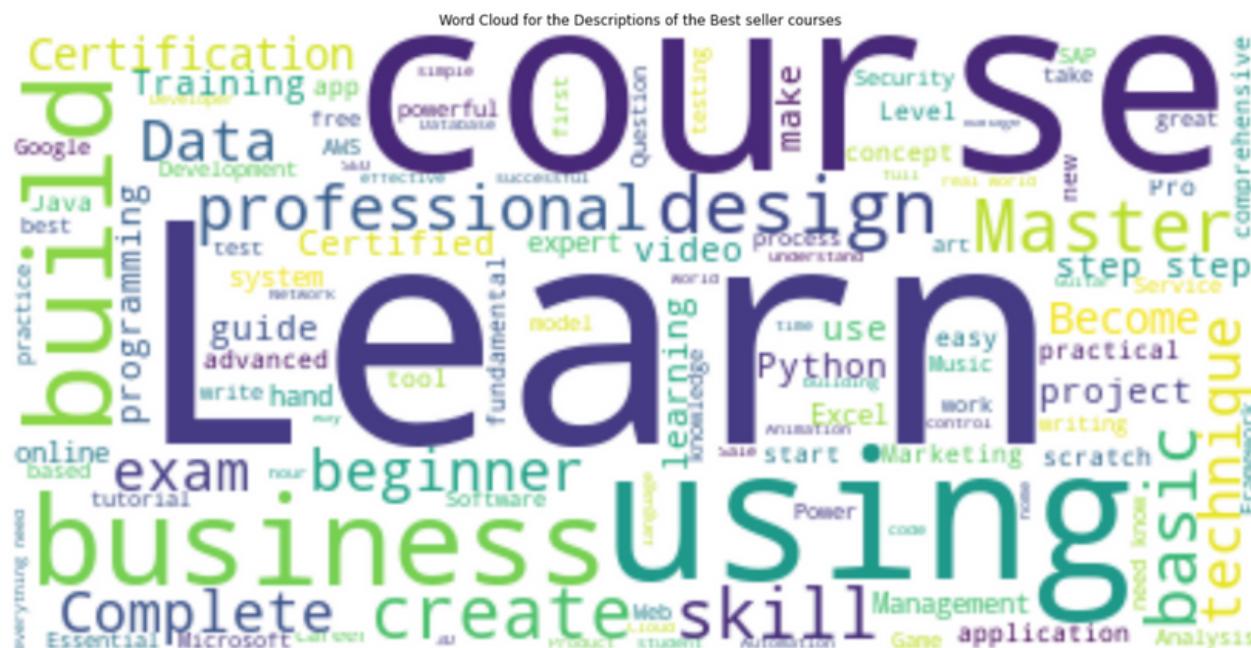
Bestseller



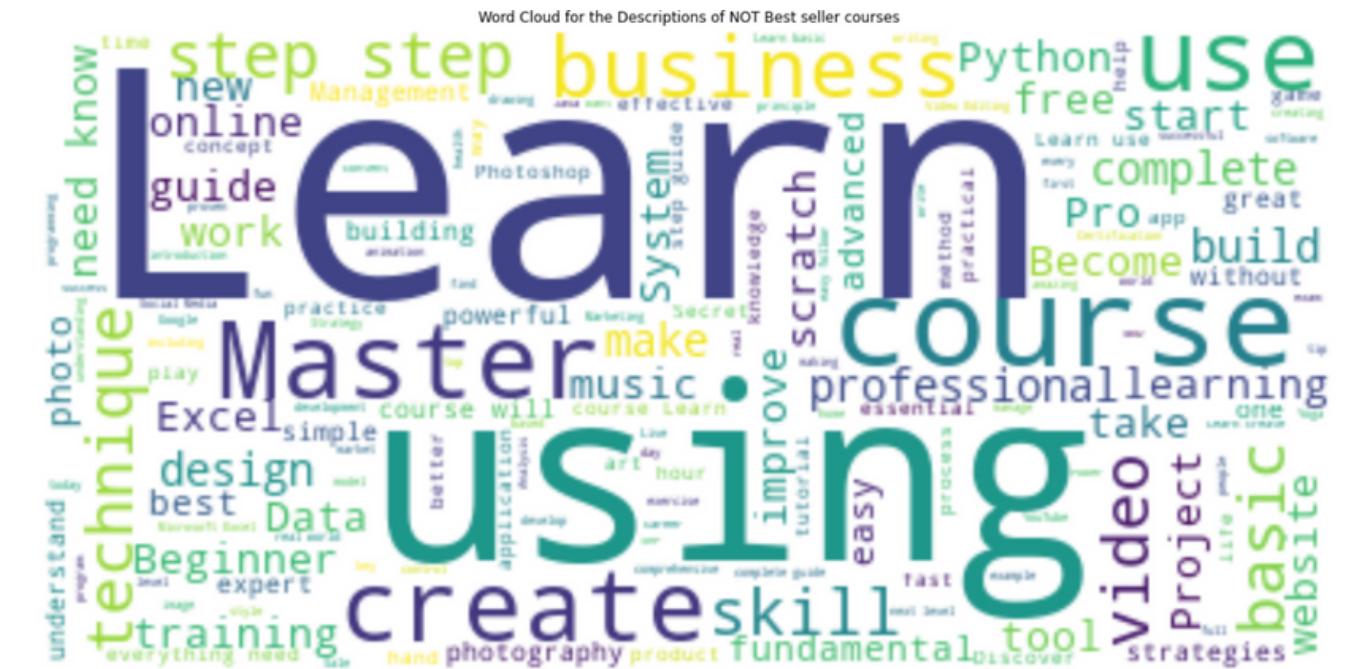
NOT Bestseller

EDA

- Description word cloud

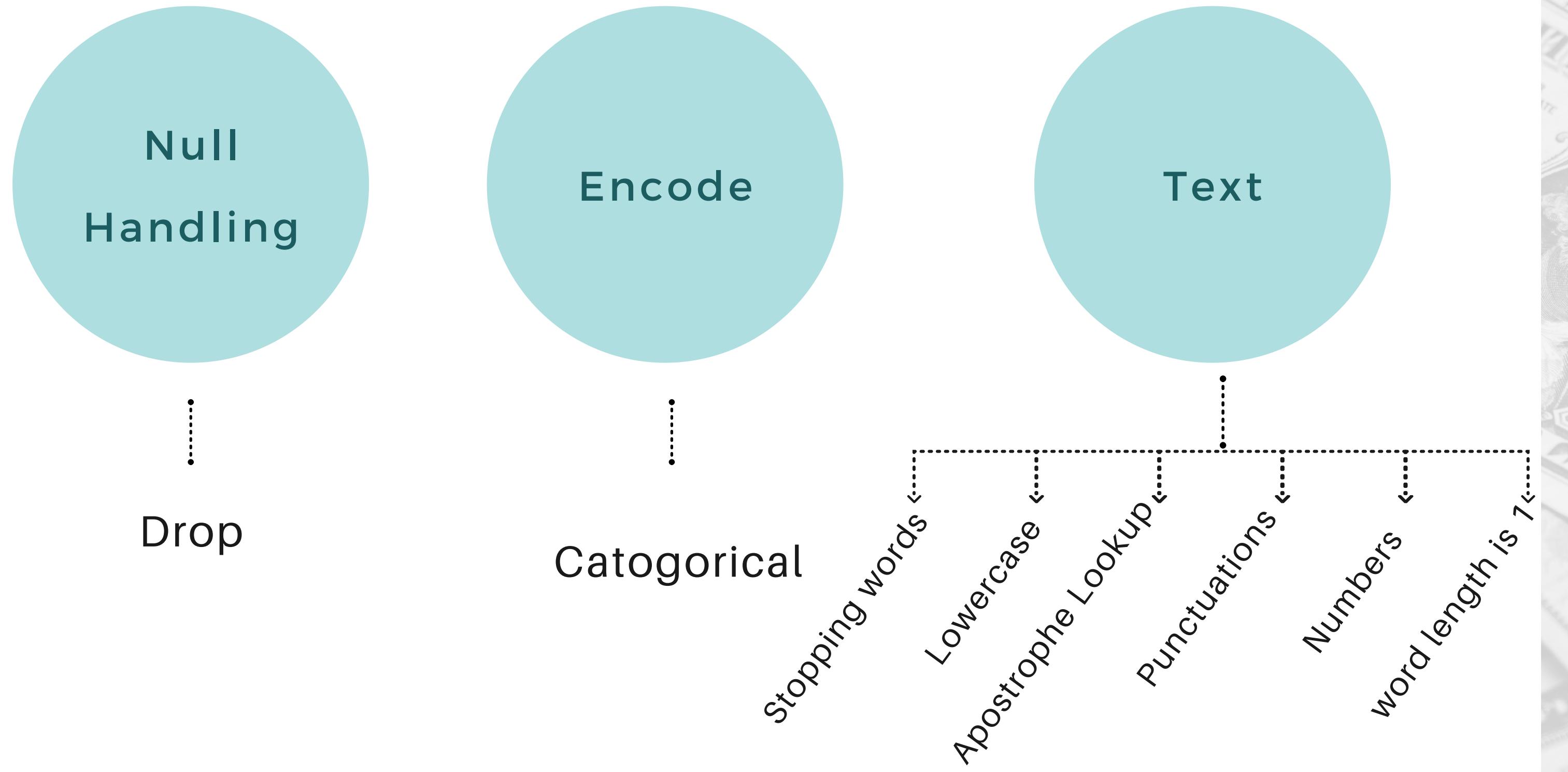


Bestseller

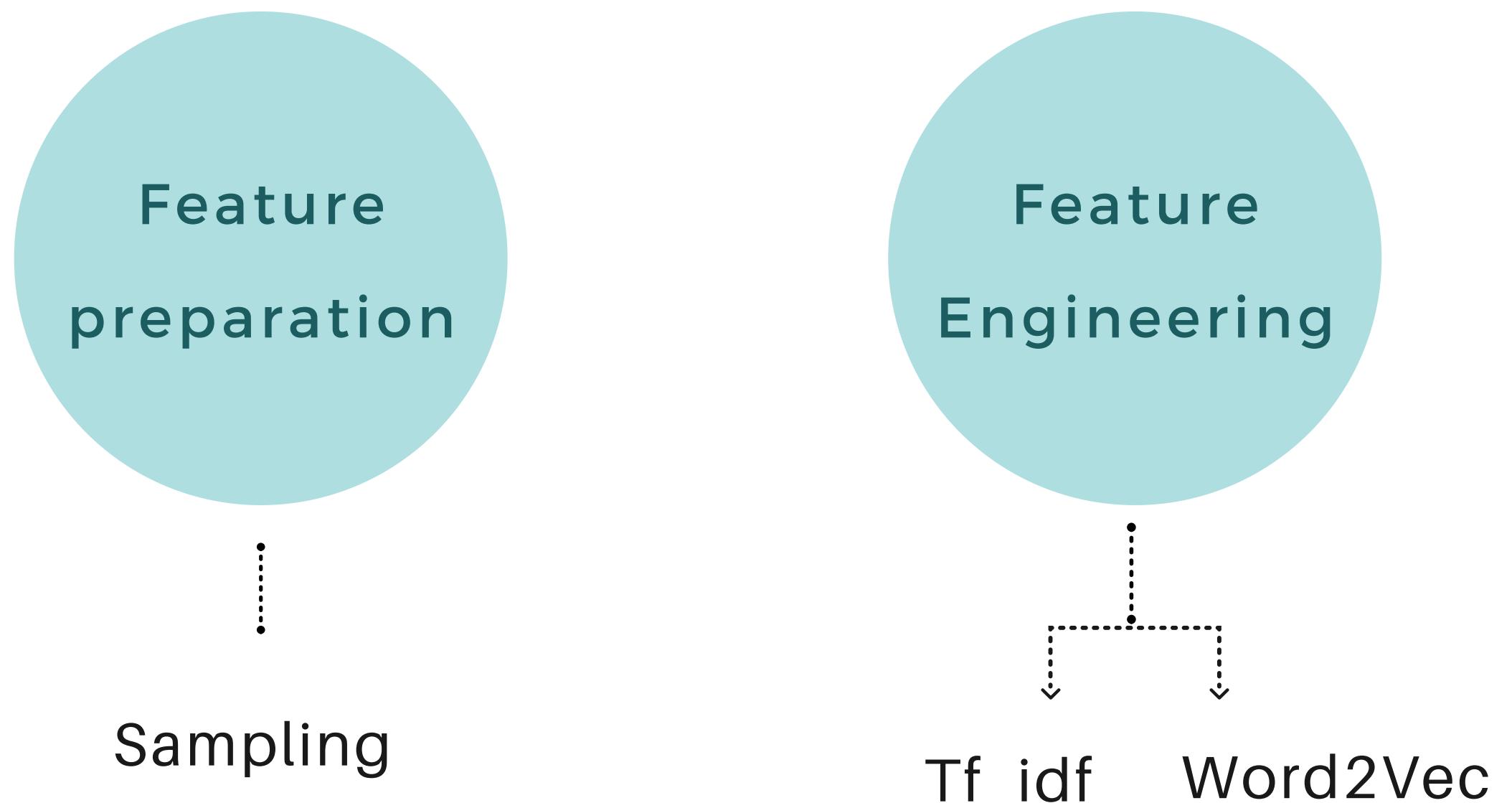


NOT Bestseller

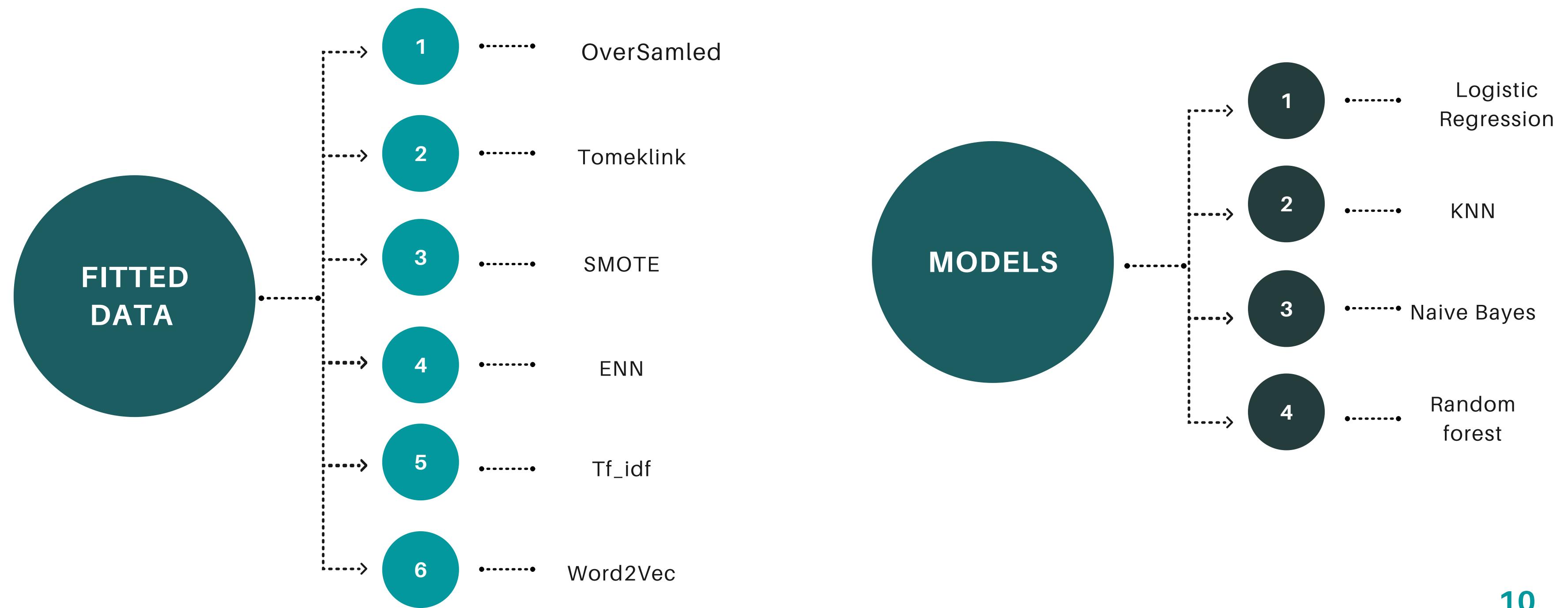
DATA CLEANING



PRE-PROCESSING



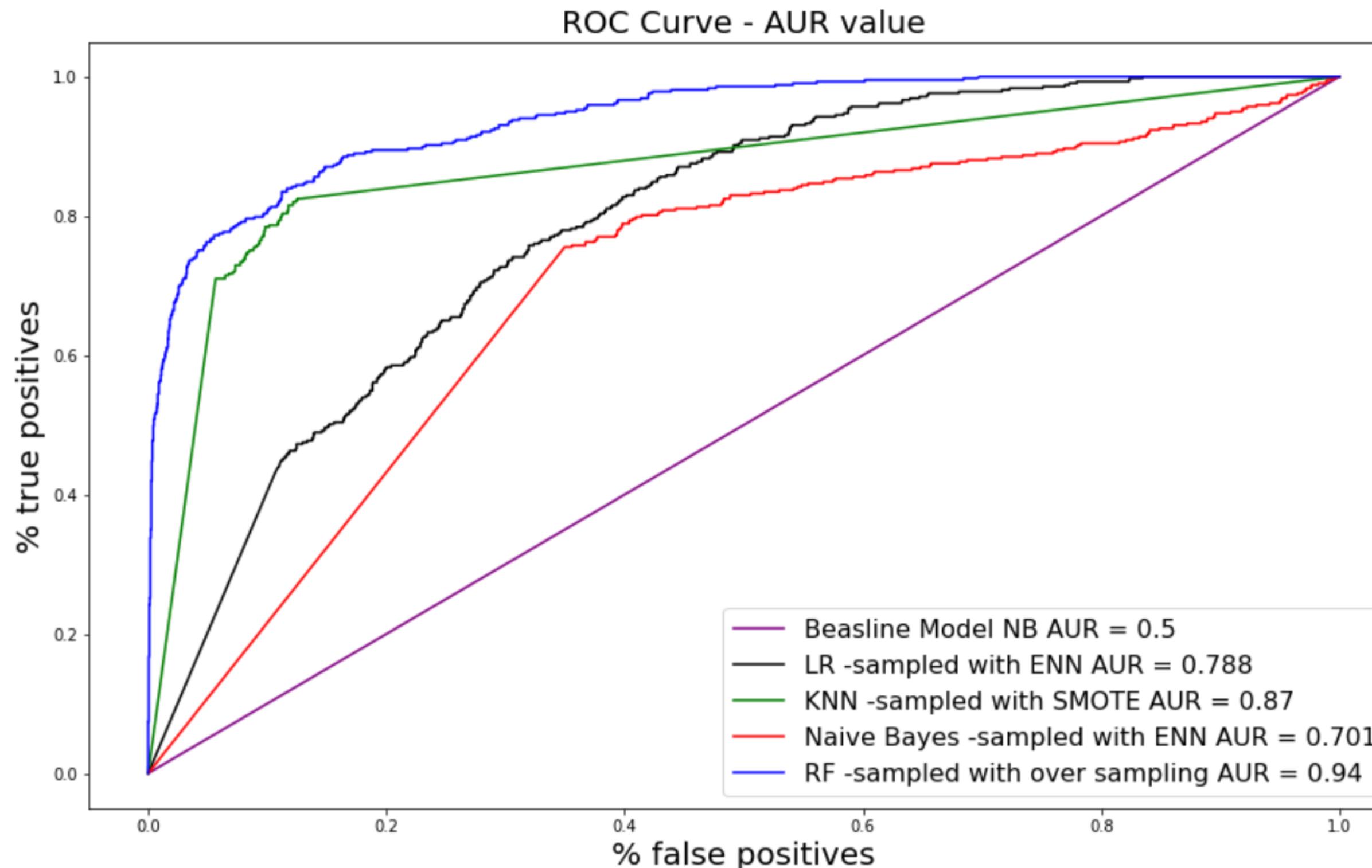
MODELING



MODELS EVALUATION

Data	Models	Score			
		F1	Precision	Recall	Accuracy
OverSamlped	Logistic Regression	0.340	0.221	0.737	0.577
	KNN	0.292	0.286	0.298	0.786
	Naive Bayes	0.347	0.345	0.349	0.806
	Random Forest	0.428	0.387	0.478	0.811
Tomeklink	Logistic Regression	0.150	0.603	0.085	0.856
	KNN	0.848	0.470	0.193	0.273
	Naive Bayes	0.349	0.334	0.366	0.798
	Random Forest	0.322	0.485	0.241	0.850
SMOTE	Logistic Regression	0.361	0.248	0.661	0.654
	KNN	0.323	0.252	0.449	0.721
	Naive Bayes	0.350	0.347	0.354	0.806
	Random Forest	0.388	0.477	0.327	0.847
ENN	Logistic Regression	0.366	0.228	0.922	0.528
	KNN	0.323	0.195	0.927	0.424
	Naive Bayes	0.361	0.239	0.737	0.613
	Random Forest	0.361	0.225	0.920	0.519
Tf_idf	Logistic Regression	0.00	0.00	0.00	0.852
	Naive Bayes	0.293	0.212	0.476	0.661
	Random Forest	0.000	0.000	0.000	0.852
Word2Vec	Logistic Regression	0.002	0.250	0.001	0.847

MODELS EVALUATION





udemy

THANK YOU FOR
LISTENING