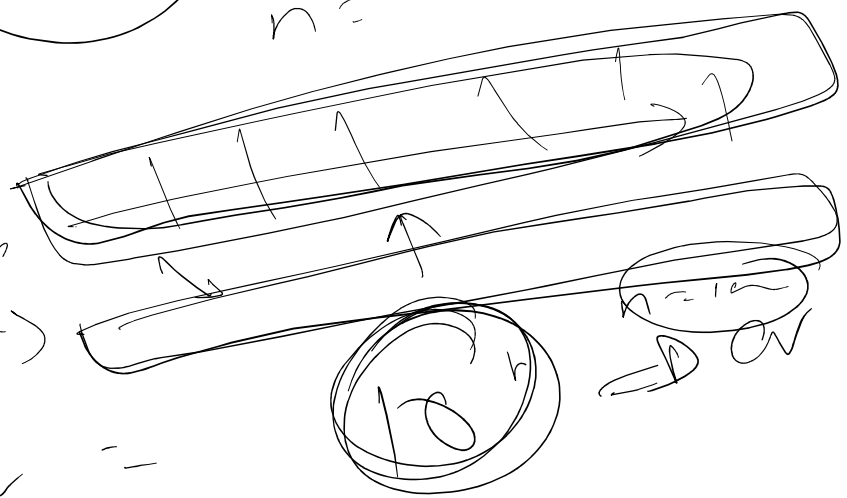
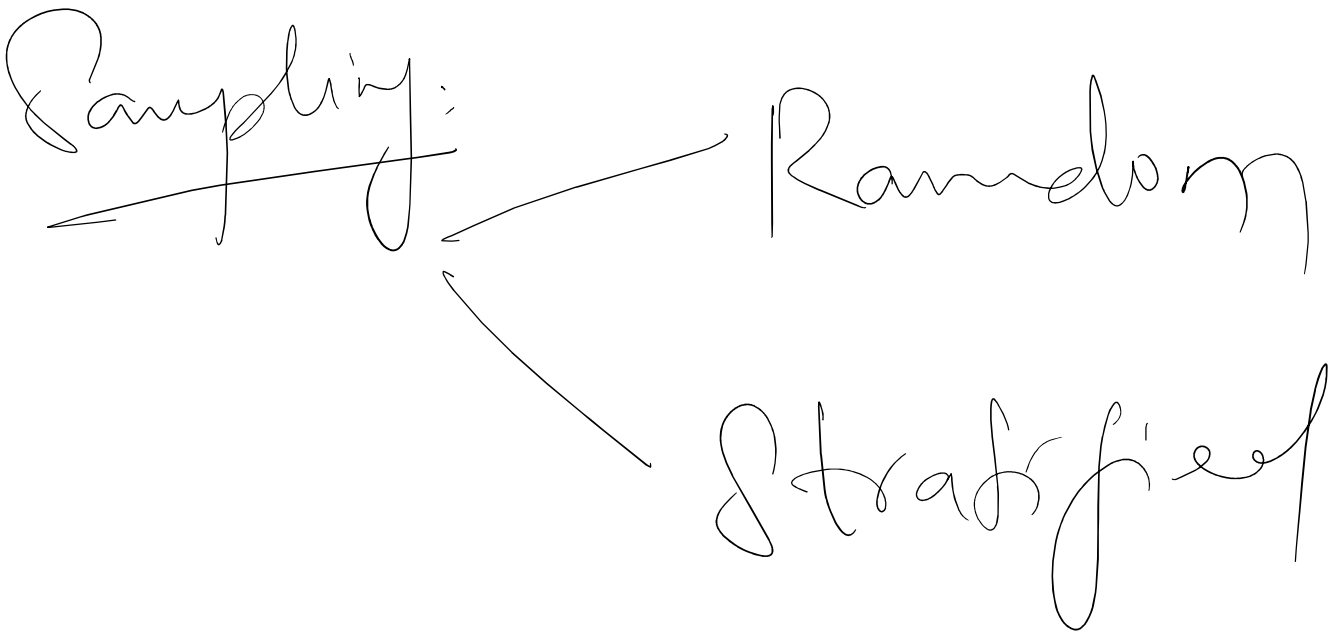
$$\frac{G}{M} * 1\omega = CV$$

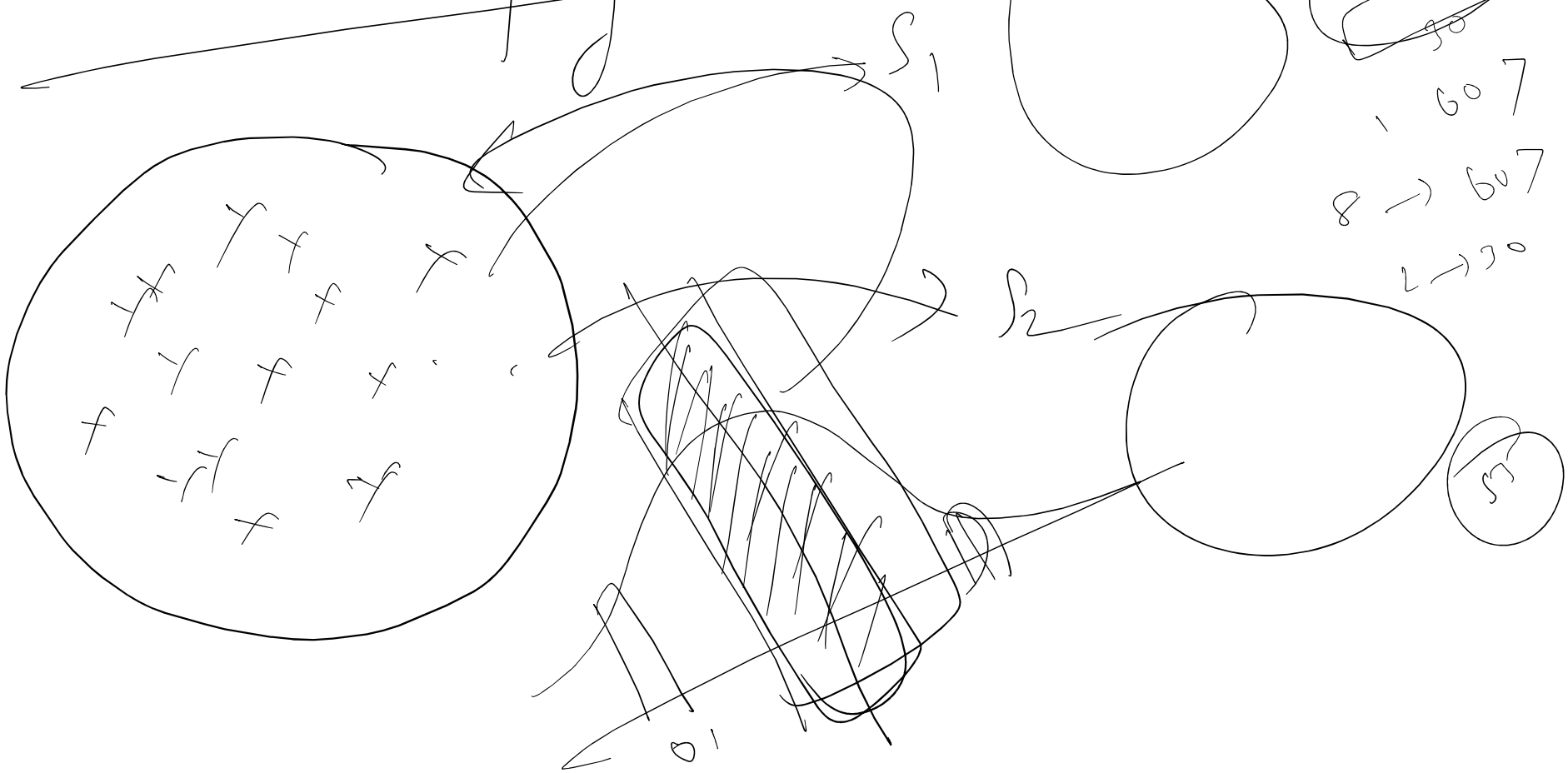
$$= \frac{0.5}{5} * 1\omega$$

$$= 0.1 * 1\omega =$$

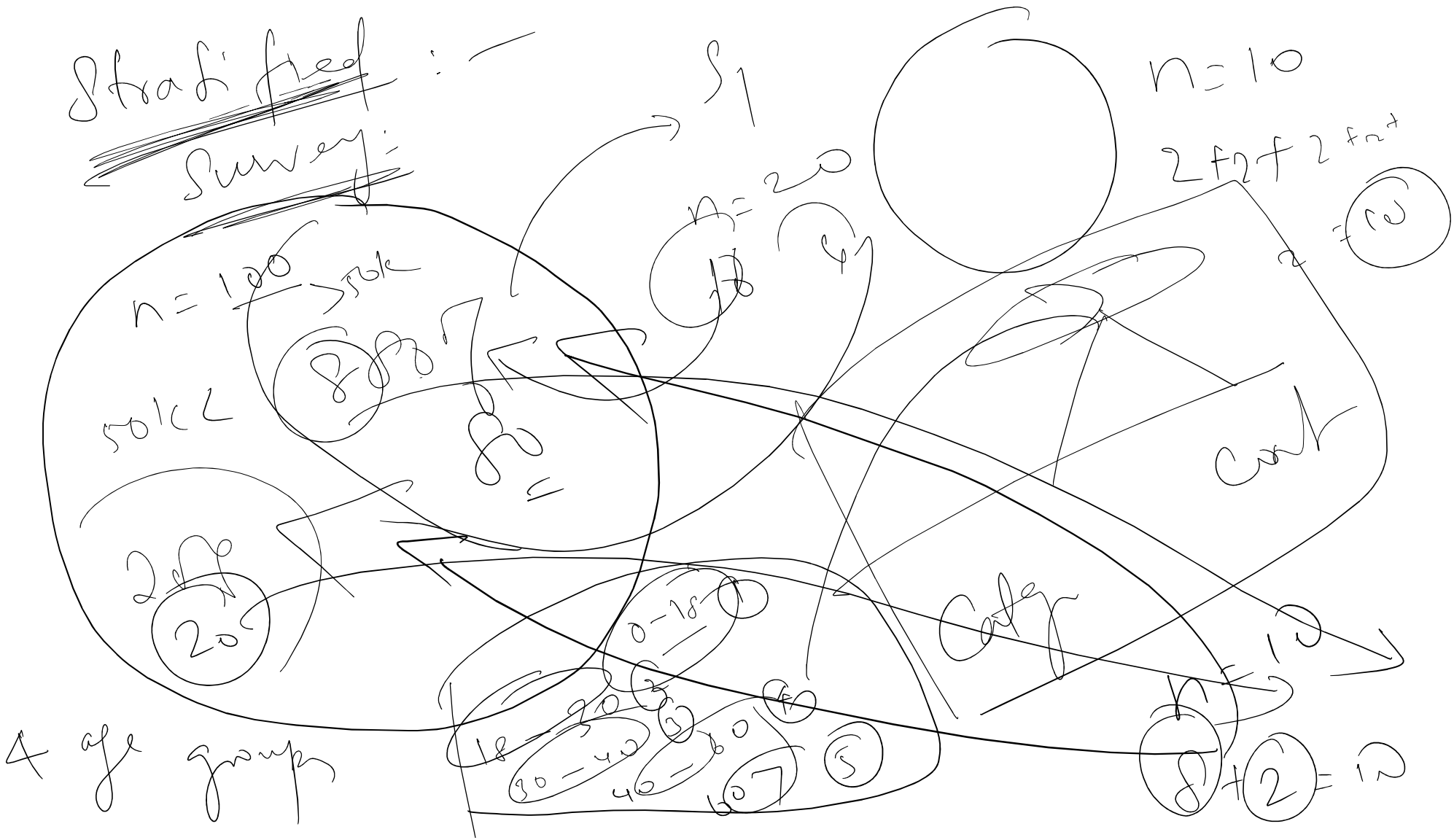



Sampling:
Random
Stratified

Random Sampling:

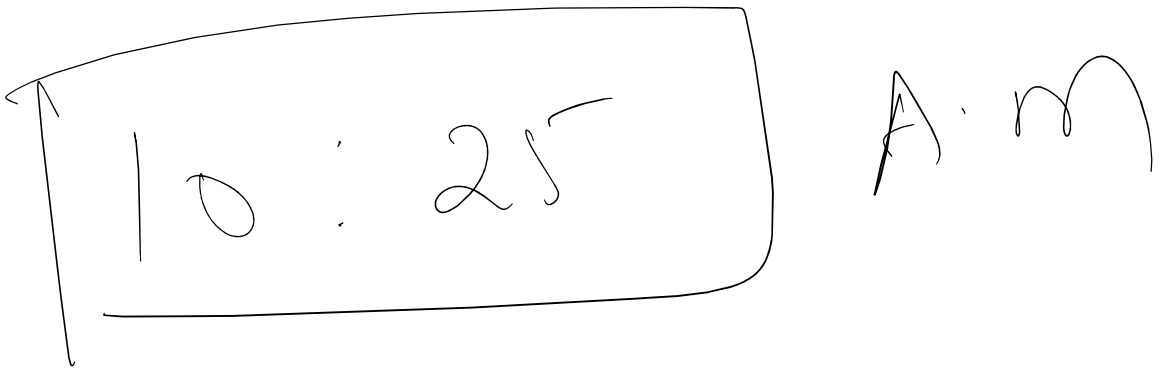


Stratified Survey:

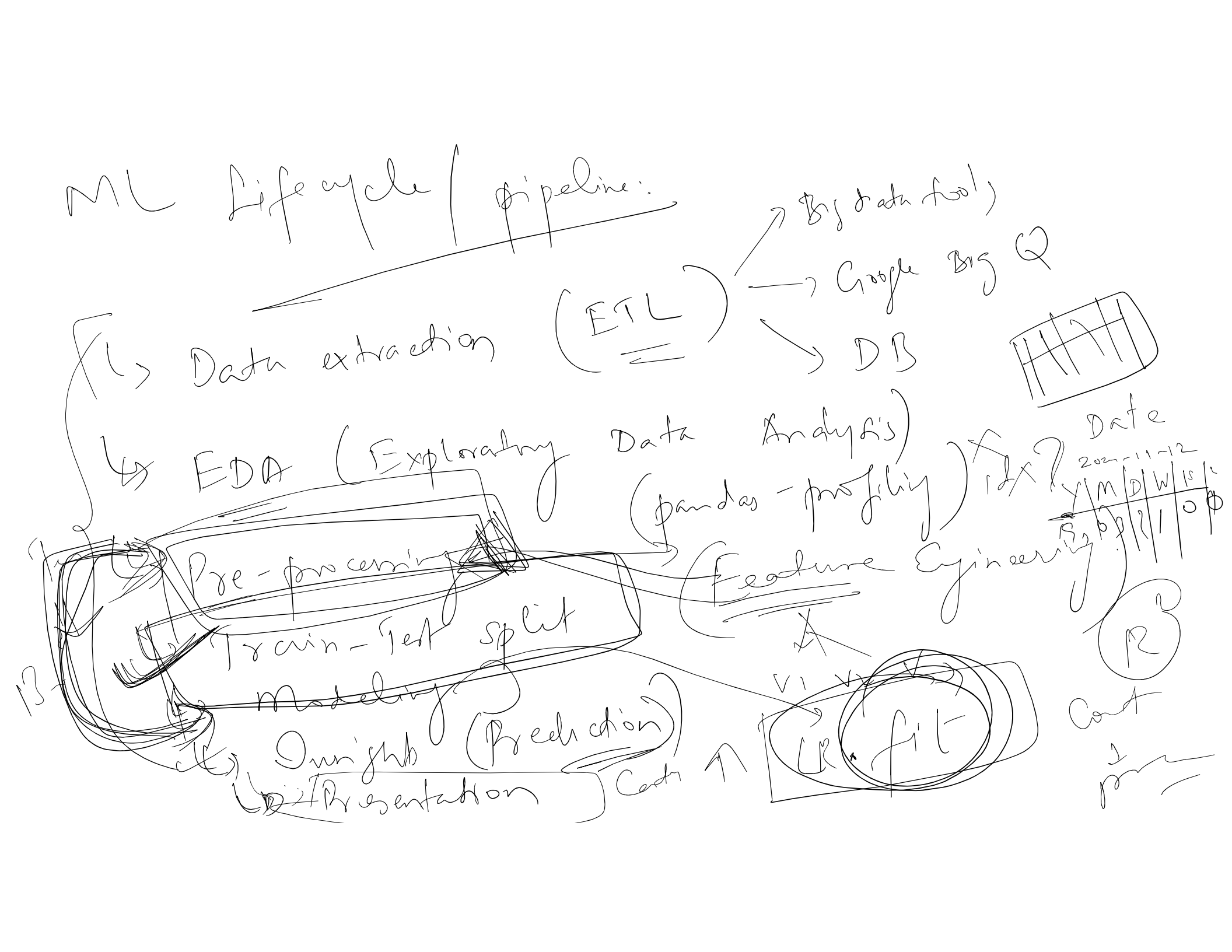


4 age groups

Be back at



A.m



ML Lifecycle / pipeline:

- Data extraction (ETL)
  - → Big data tools
  - → Google Big Q
  - → DB
- EDA (Exploratory Data Analysis) (pandas-profiling)
- Pre-processing
- Train-Test split
- Modeling
- Insights (Prediction)
- Presentation

Feature Engineering

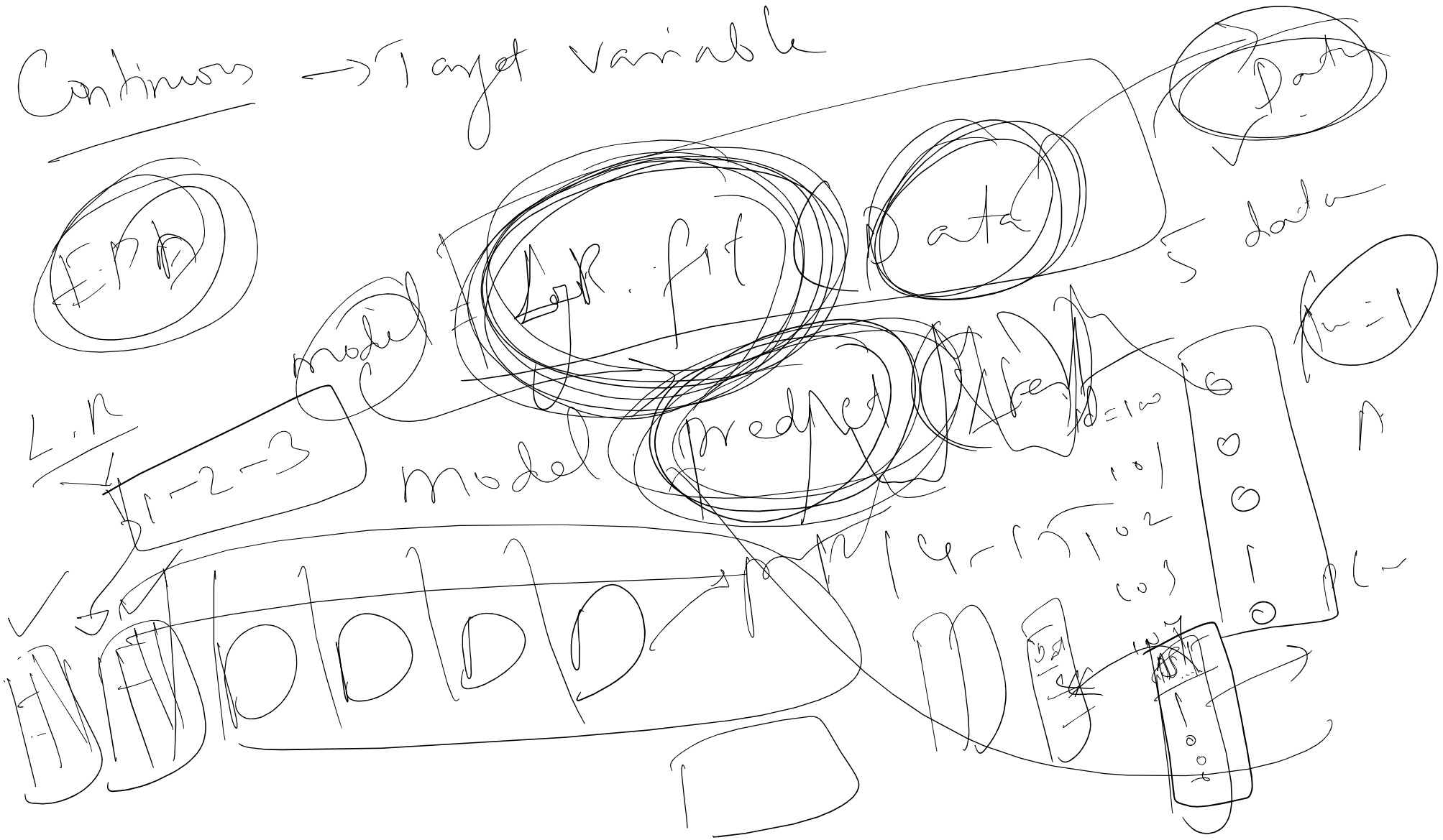
Date 2021-11-12

Y	M	D	W	IS	
5	0	3	2	1	0

R

Cont

LR fit



Continuous → Target variable

EPD

model = L.R. fit Data

Data

5 data

L.R

1-2-3

model predict

$\mu = 1$

$\mu = 100$

101

102

103

104

0

0

0

1

0

$P(\mu$

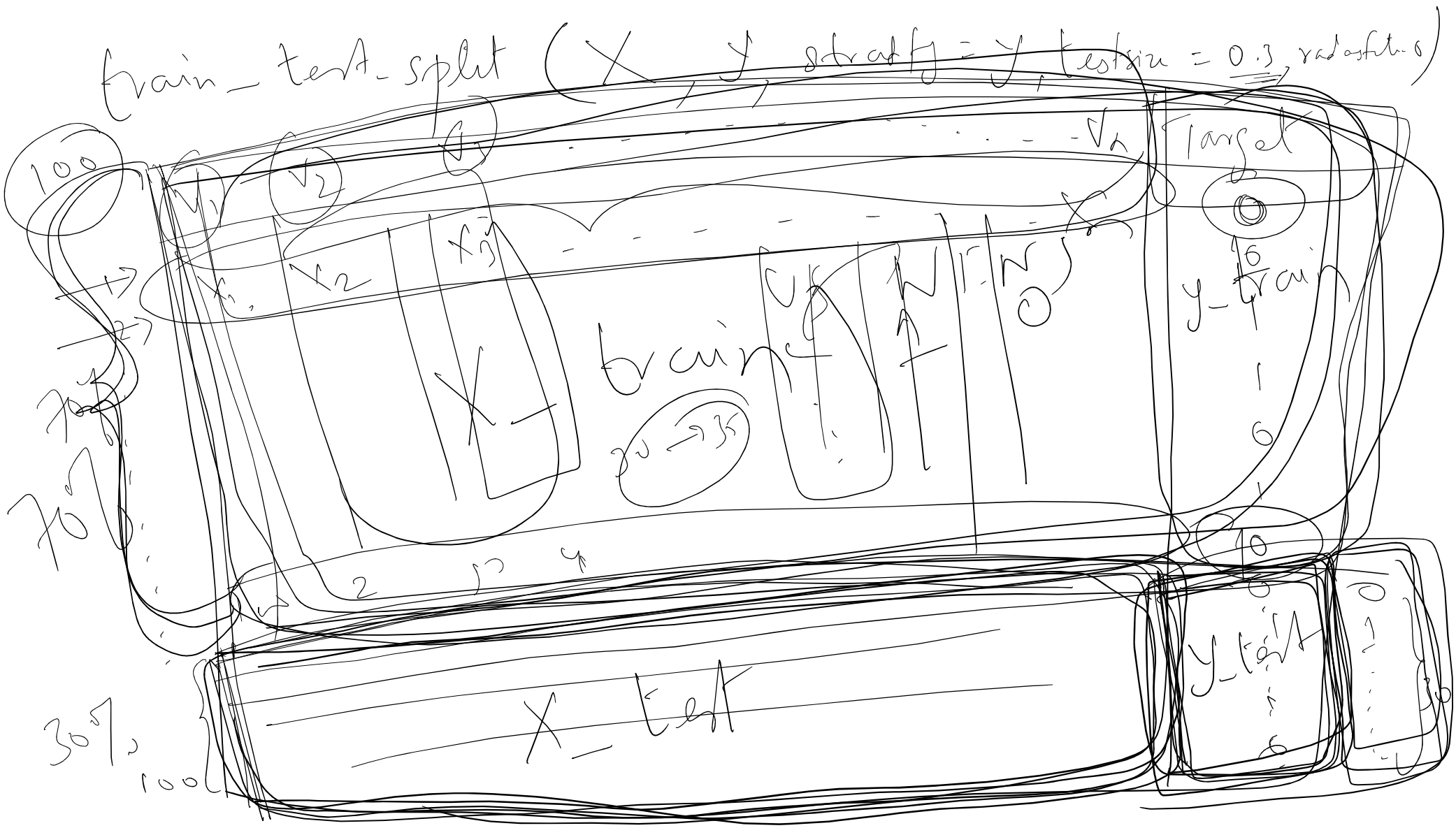
1

0

0

0

train_test_split (X, y, stratify = y, testsize = 0.3, radomstate = 0)



n = 100 test size = 0.3

train = 70% ; test = 30%

RF SVM

from sklearn.linear_model import LogisticReg

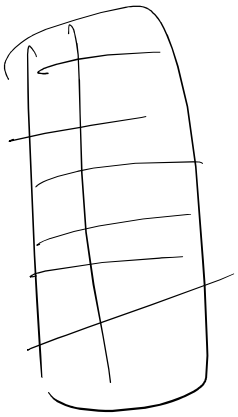
model = LogisticRegression.fit() Boost

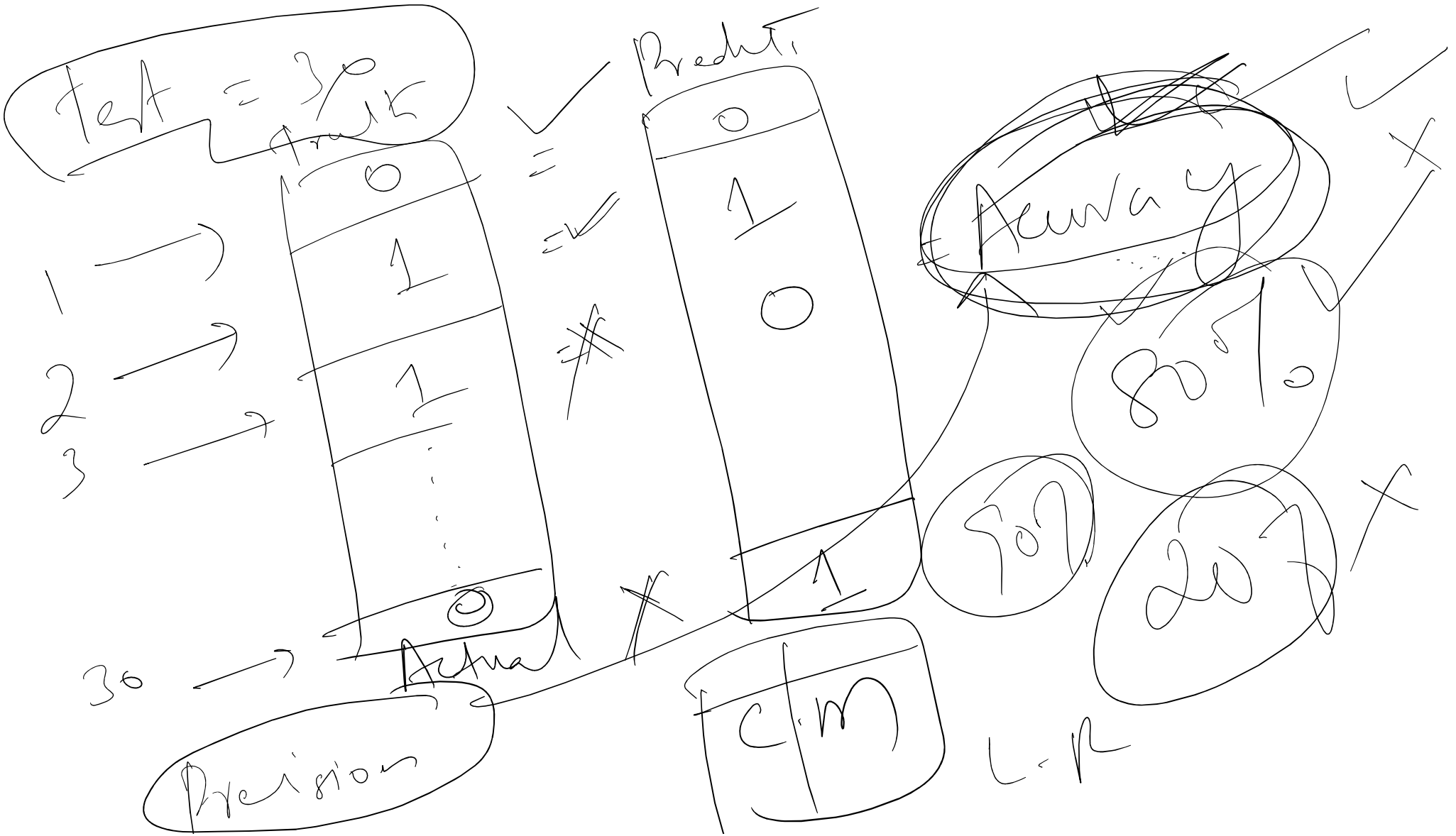
XGBoost

Catb

model.fit(Data)

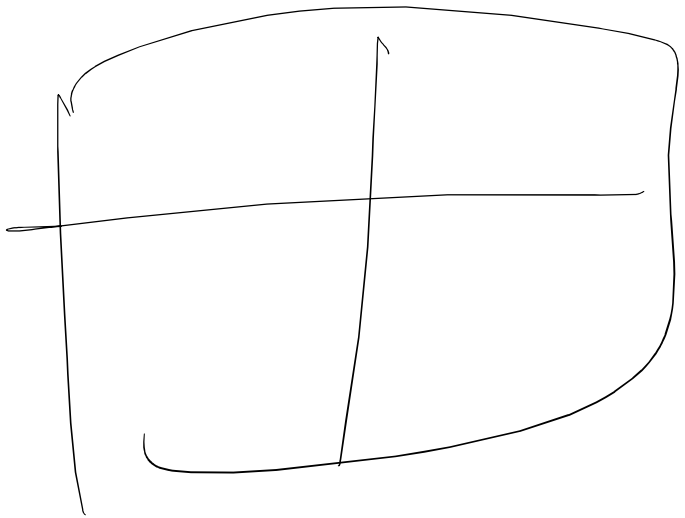
Predicting model.predict(X_test)

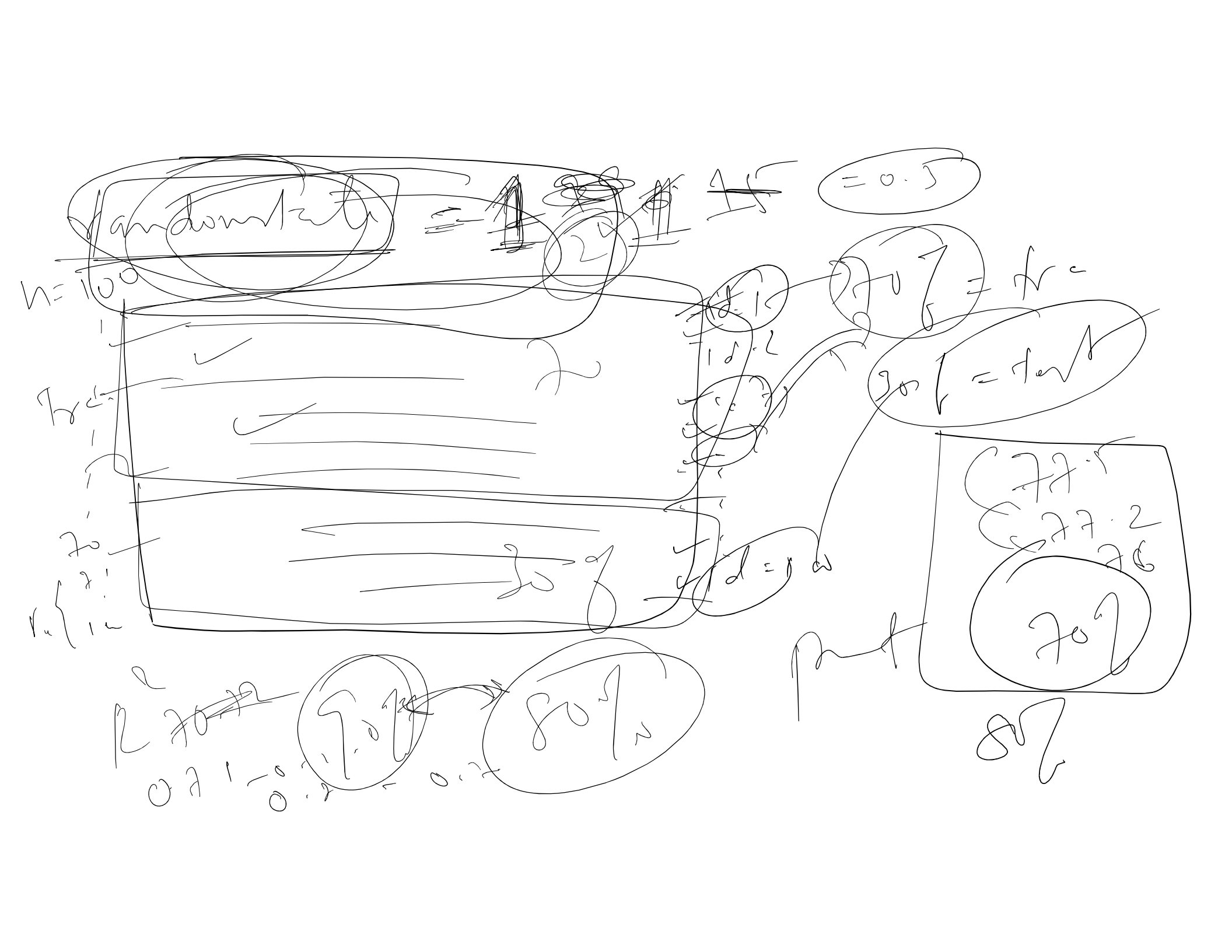


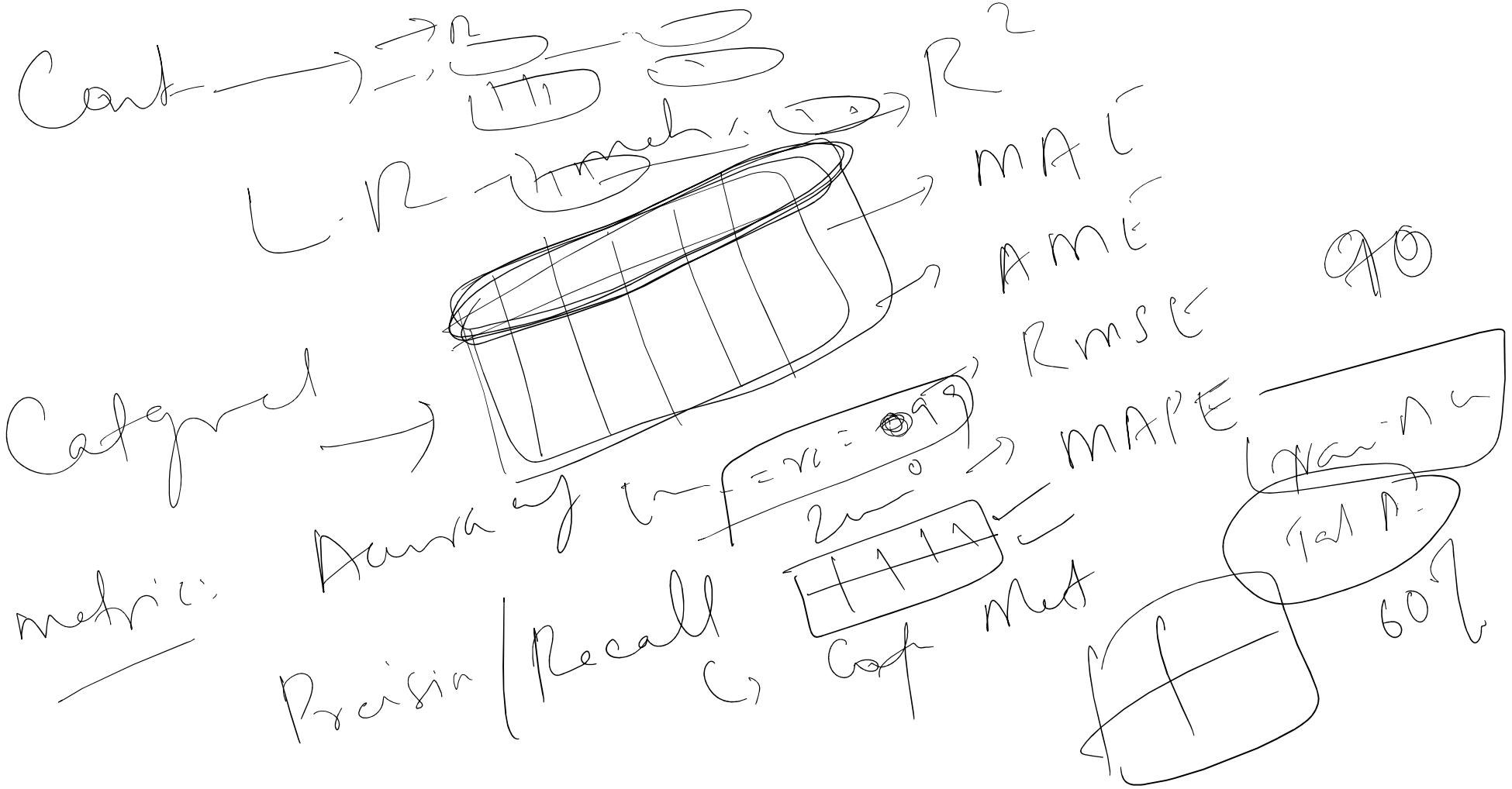

Test = 30
Actual
Predict
Accuracy
Precision
Recall
F1 Score
1
2
3
30
0
1
1
1
0
1
0
1
0

Logistic Reg

Conf Matrix







Cont → R

L.R → R^2

MAE

AME

RMSE

MAPE

Categorical →

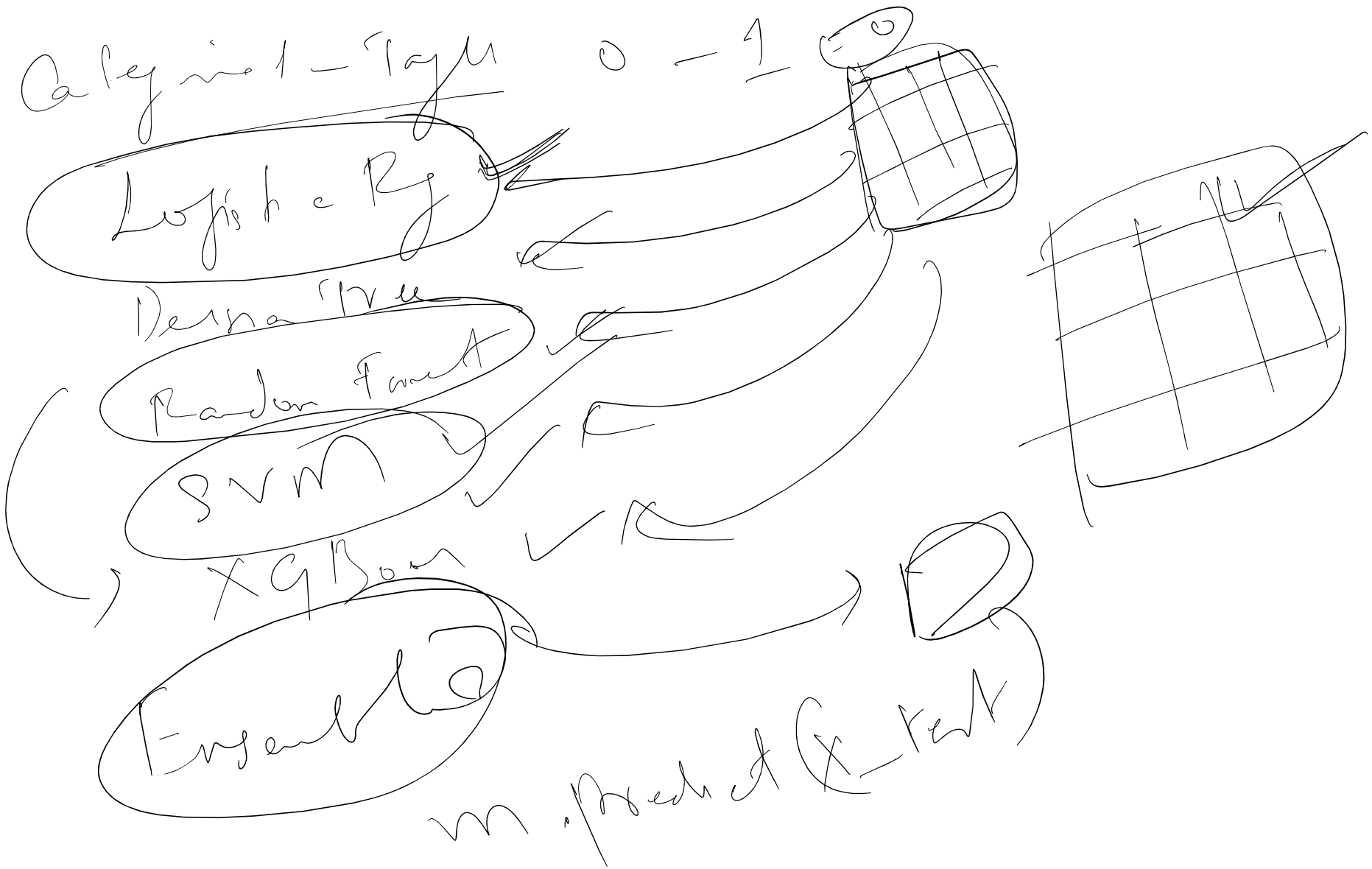
metric: Accuracy

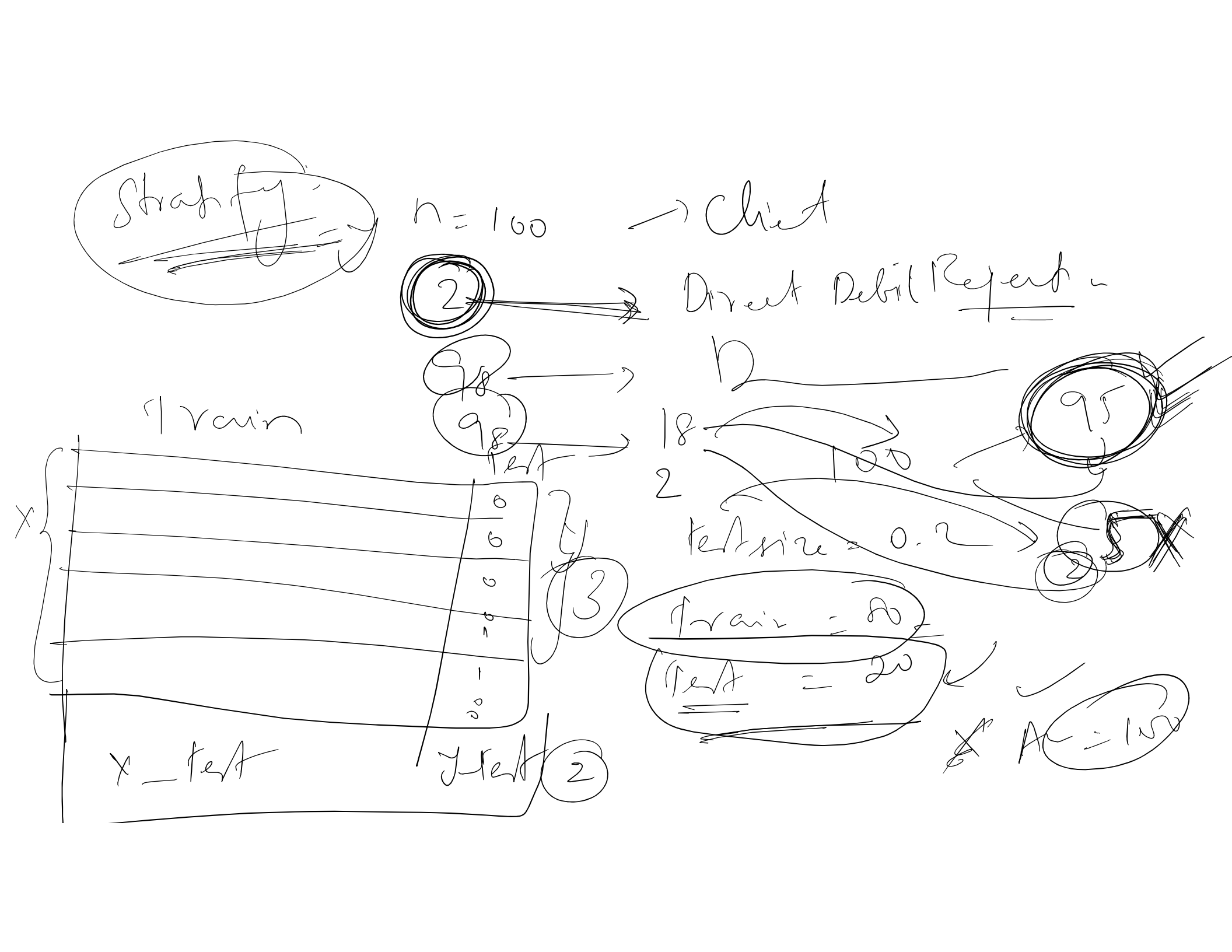
Precision / Recall

Cof Mat

90

60%


Logistic Reg
Random Forest
SVM
XGBoost
Ensemble



Stratify

n = 100 → Client

2 → Direct Debit Repeat

98 → D

98 → 18

2

100

95

Test size = 0.2

2

Train

X

Y

0
0
0
0
1
1
0
0

3

X_test

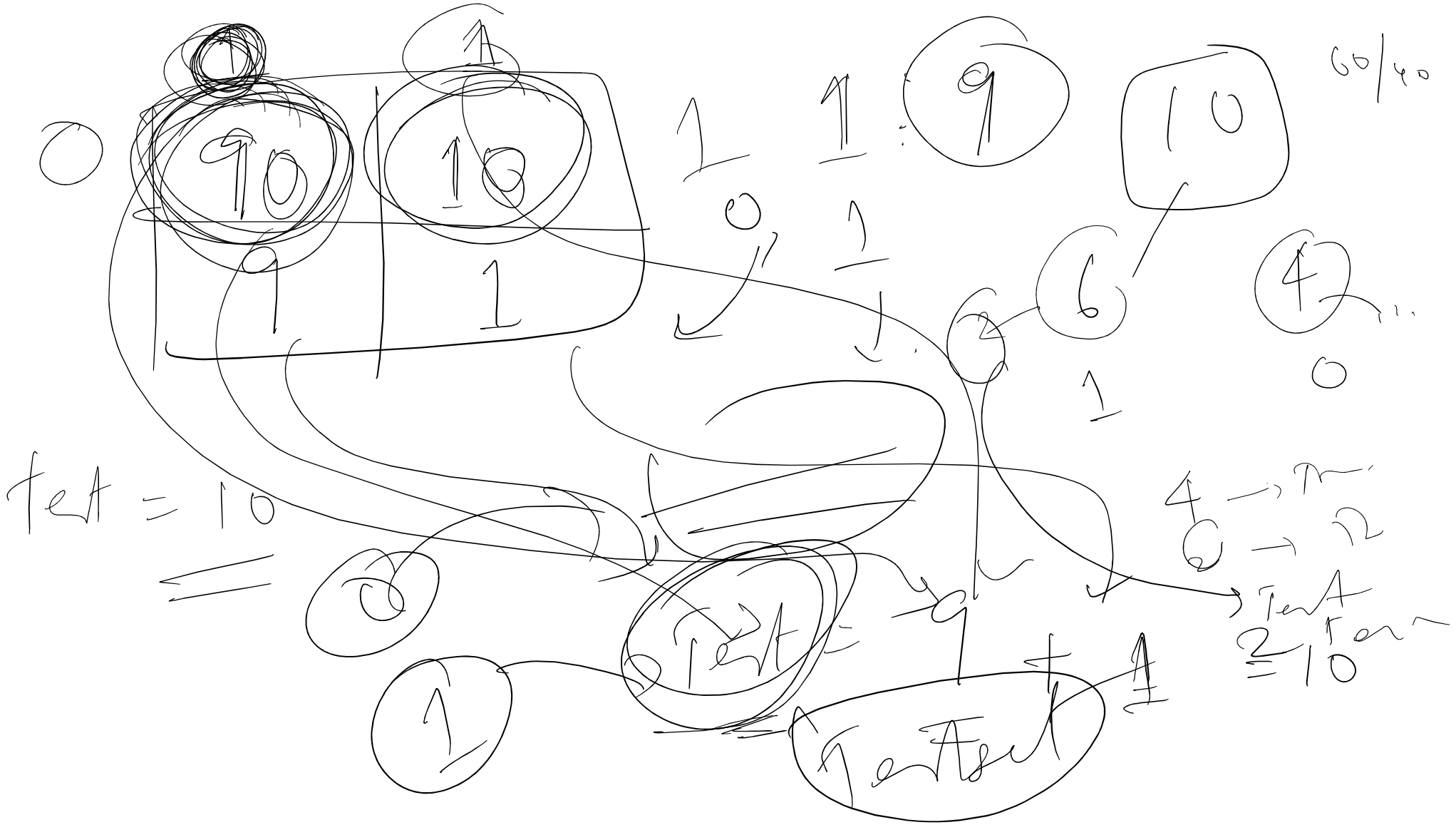
Y_test

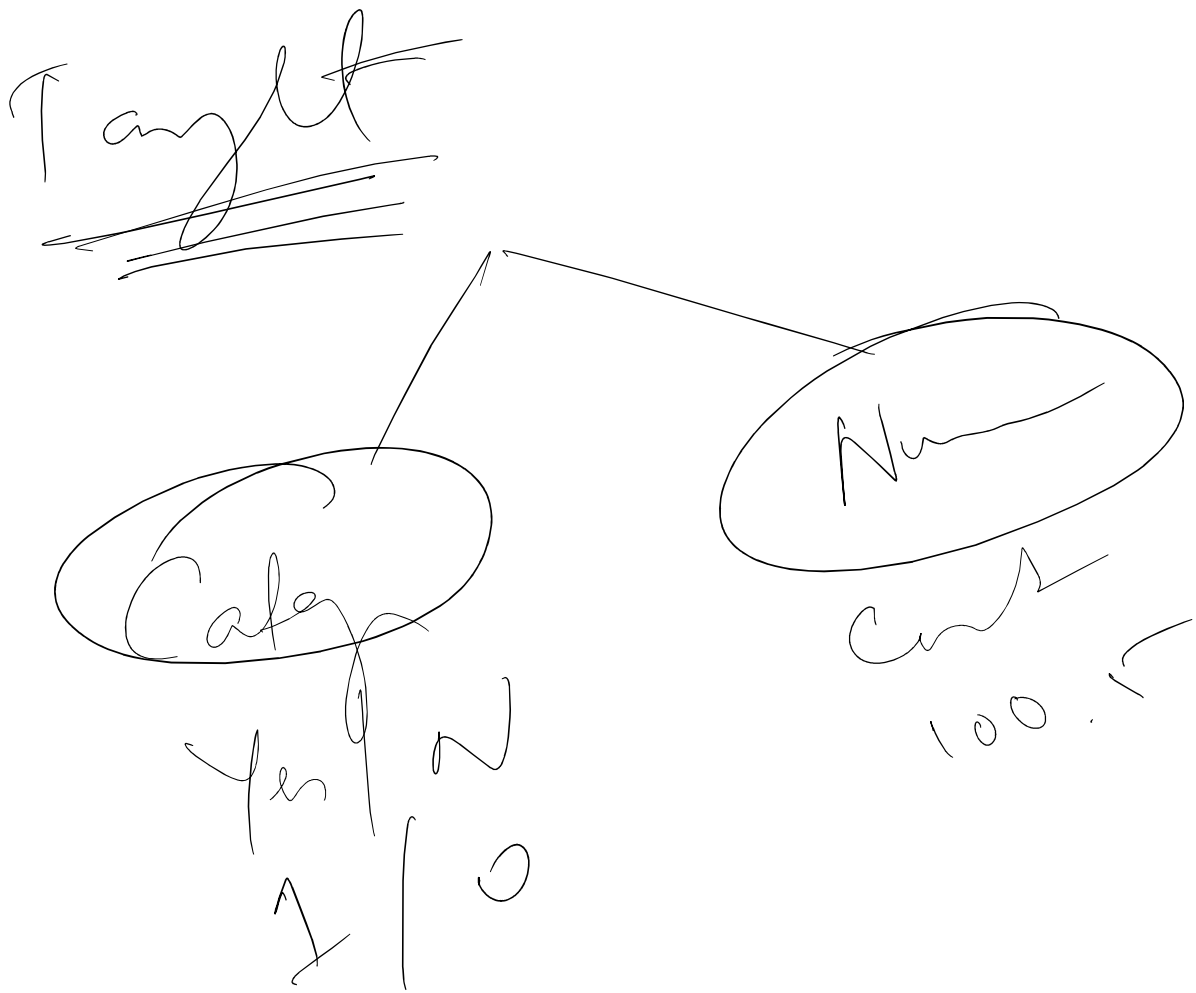
2

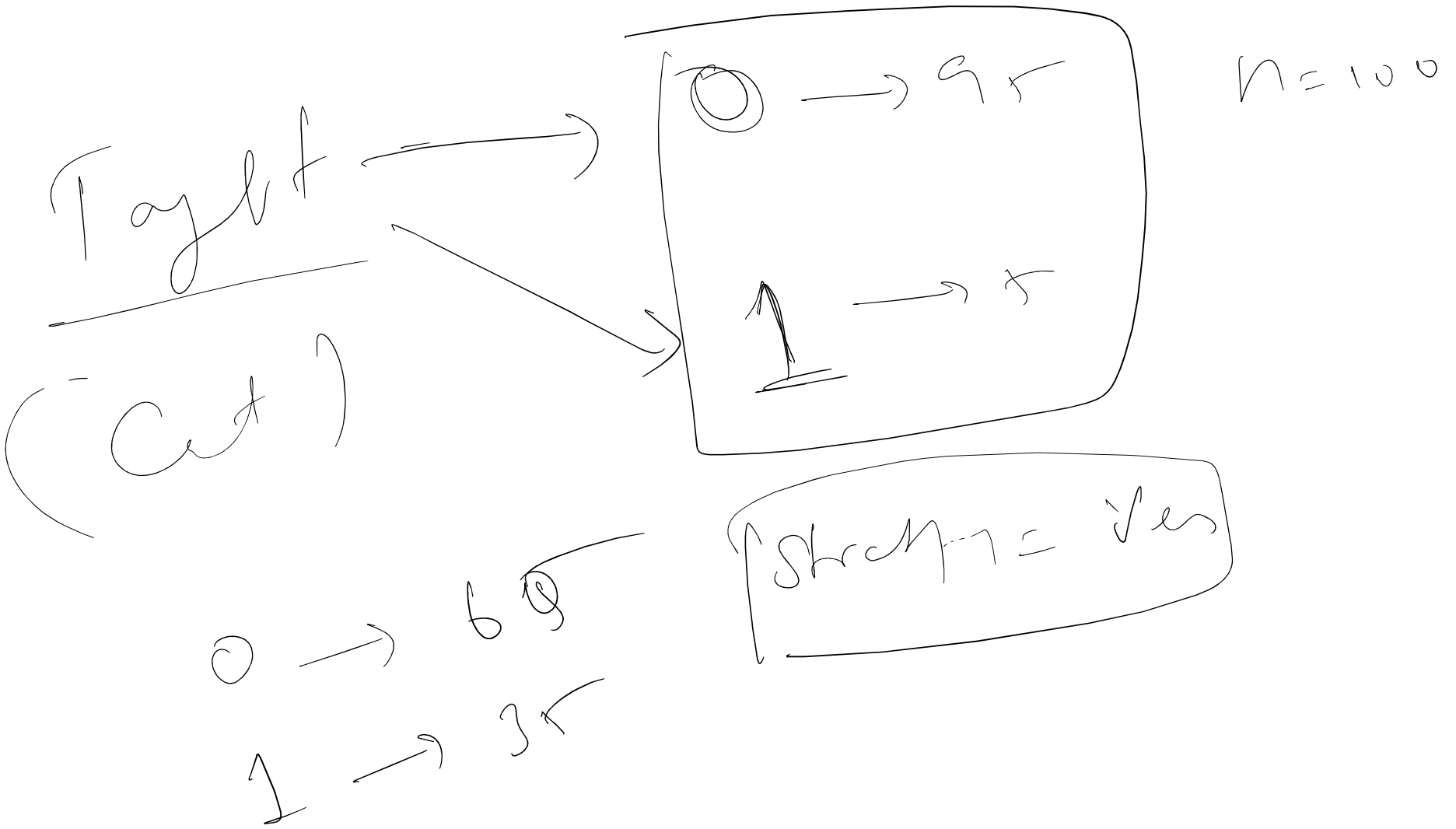
Train = 80

Test = 20

AC = 100







Taylor

(Cut)

0 → 95

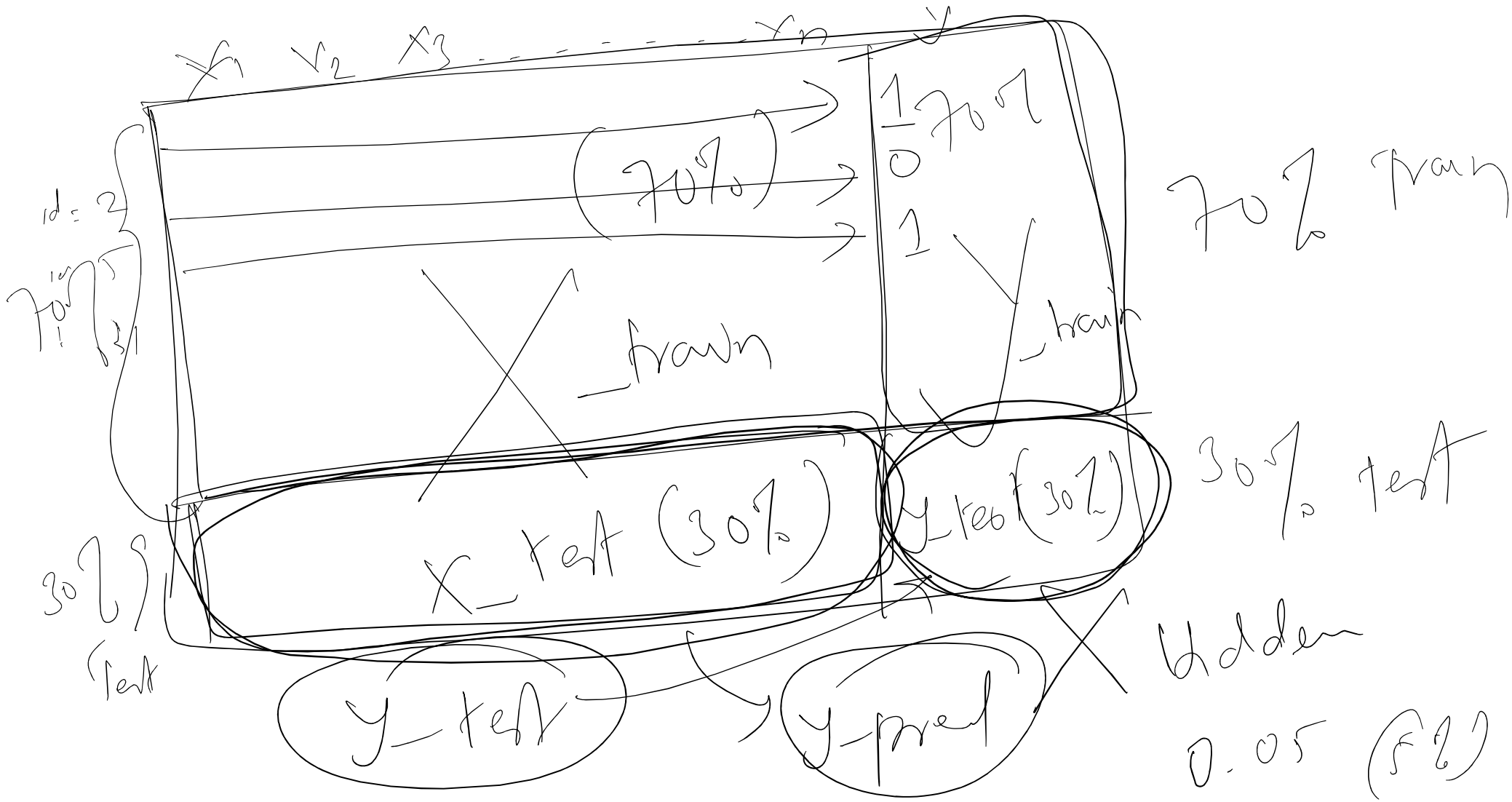
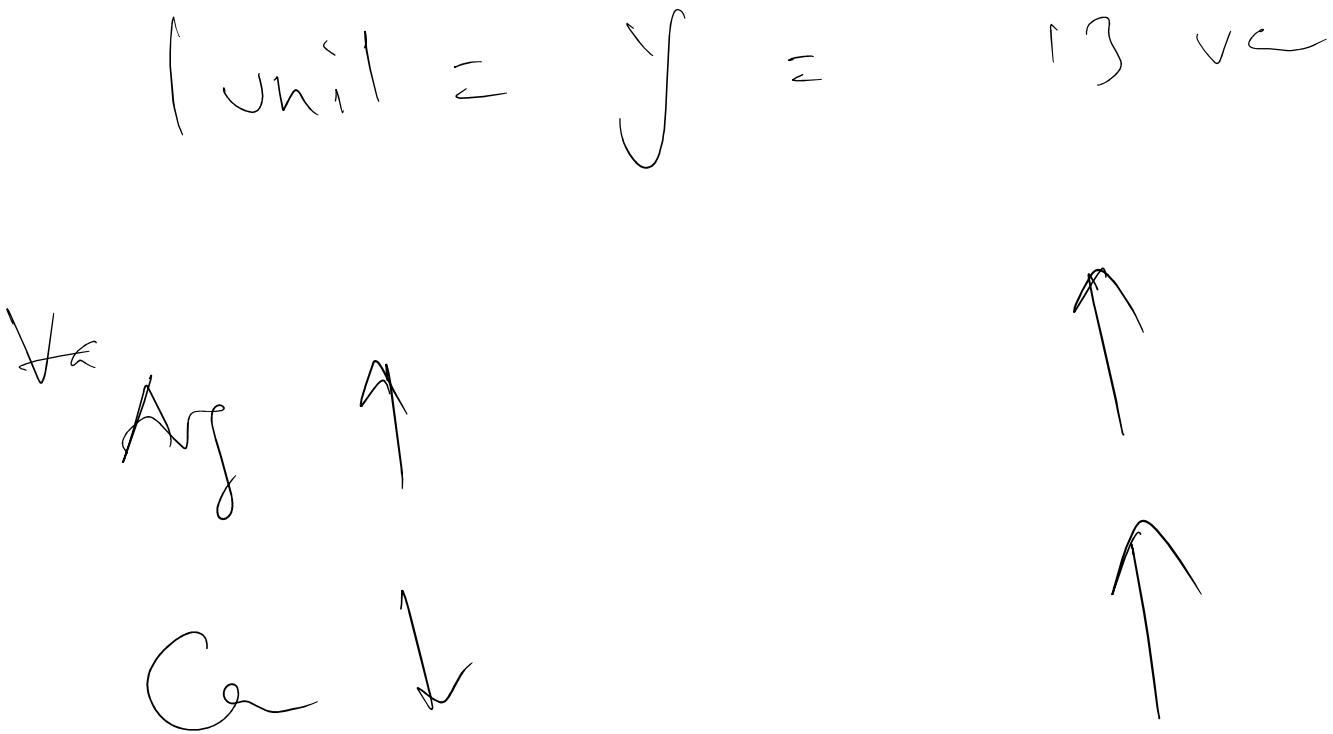
1 → 5

n = 100

Strength = Yes

0 → 65

1 → 35



x1 x2 x3 ..... xn y
1
0
1
70%
(70%)
X_train
Y_train
70% train
X_test (30%)
Y_test (30%)
30% test
30% test
y_test
y_pred
hidden
0.05 (5%)