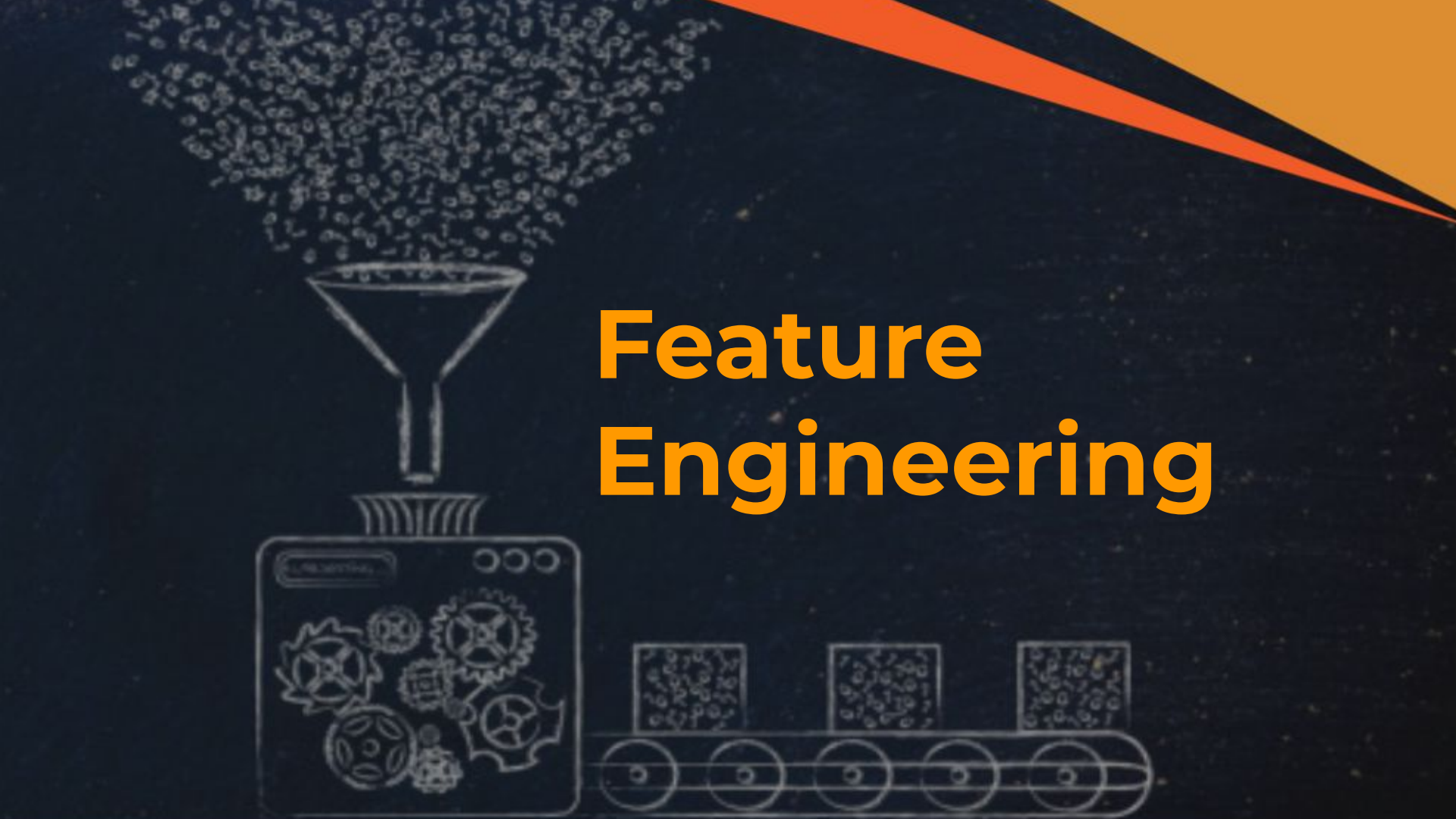
Titanic Survival Predictions

Problem Statement

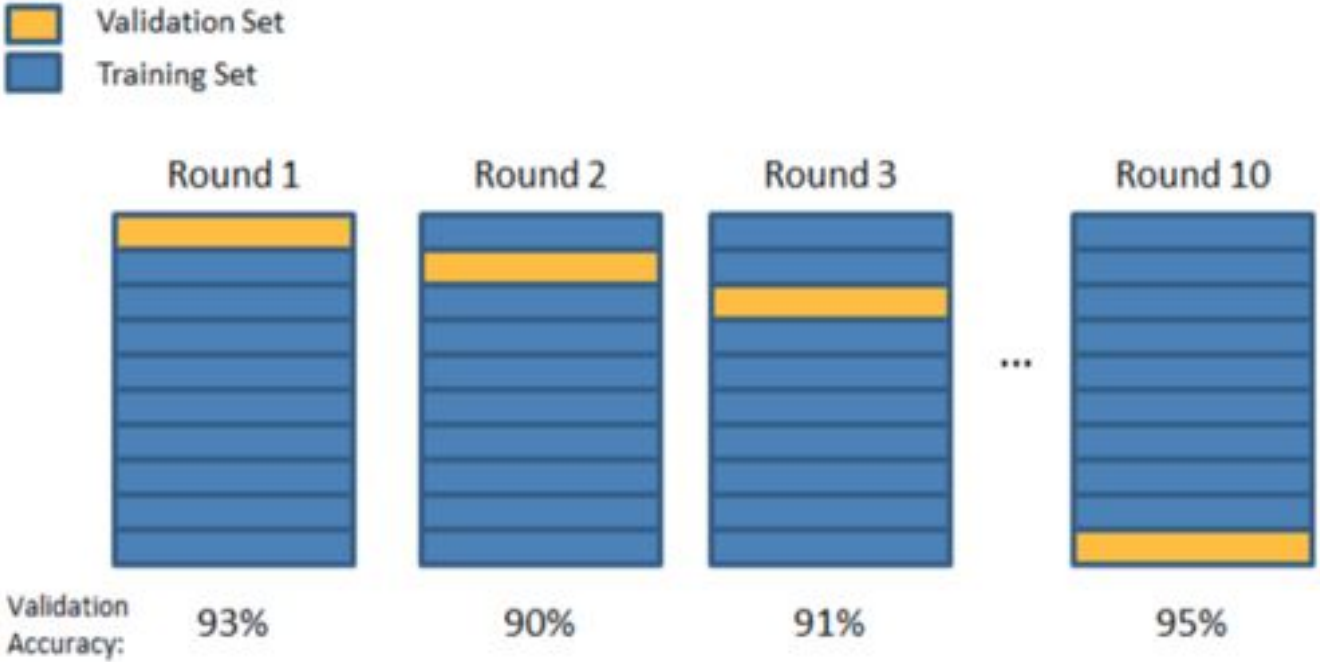
Data Collection, Loading And Exploration


Feature
Engineering

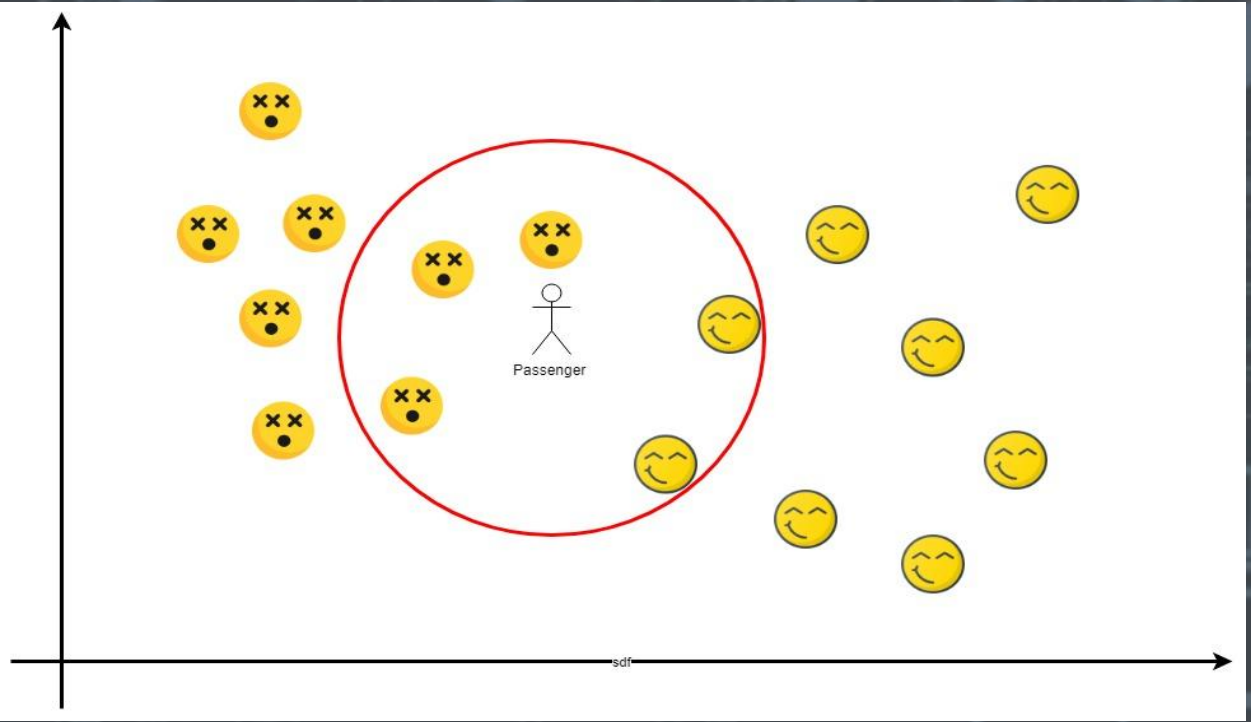
Implementation Of Feature Engineering

- Fill in the missing fields
- Binning
- Vector Mapping

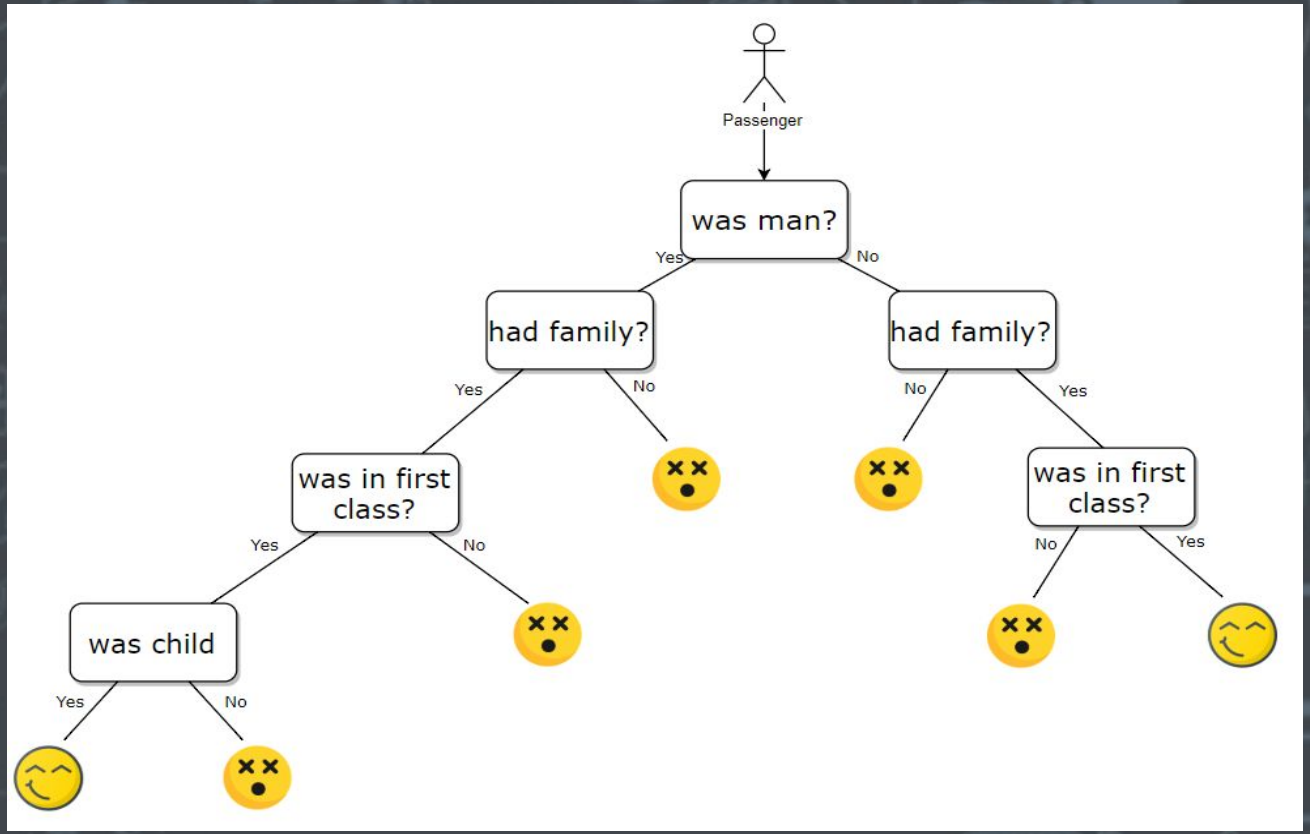
Modelling and Cross Validation


Validation Set
Training Set
Round 1
Round 2
Round 3
Round 10
...
Validation Accuracy:
93%
90%
91%
95%
Final Accuracy = Average(Round 1, Round 2, ...)

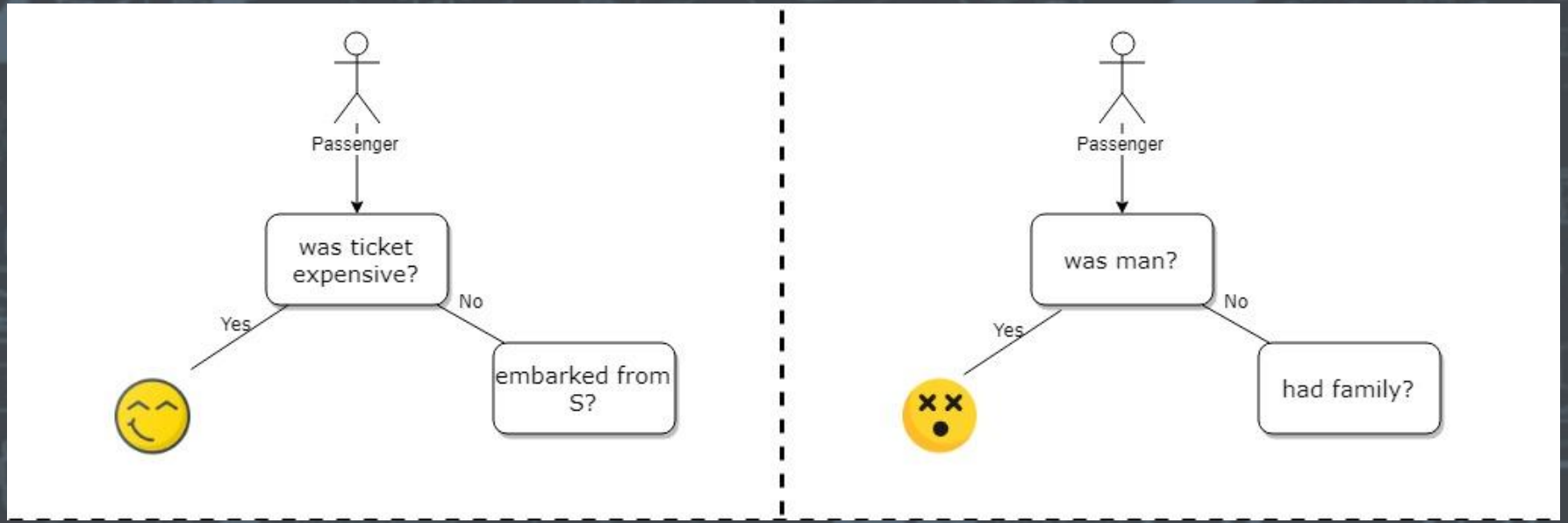
K-nearest neighbors



Decision Tree



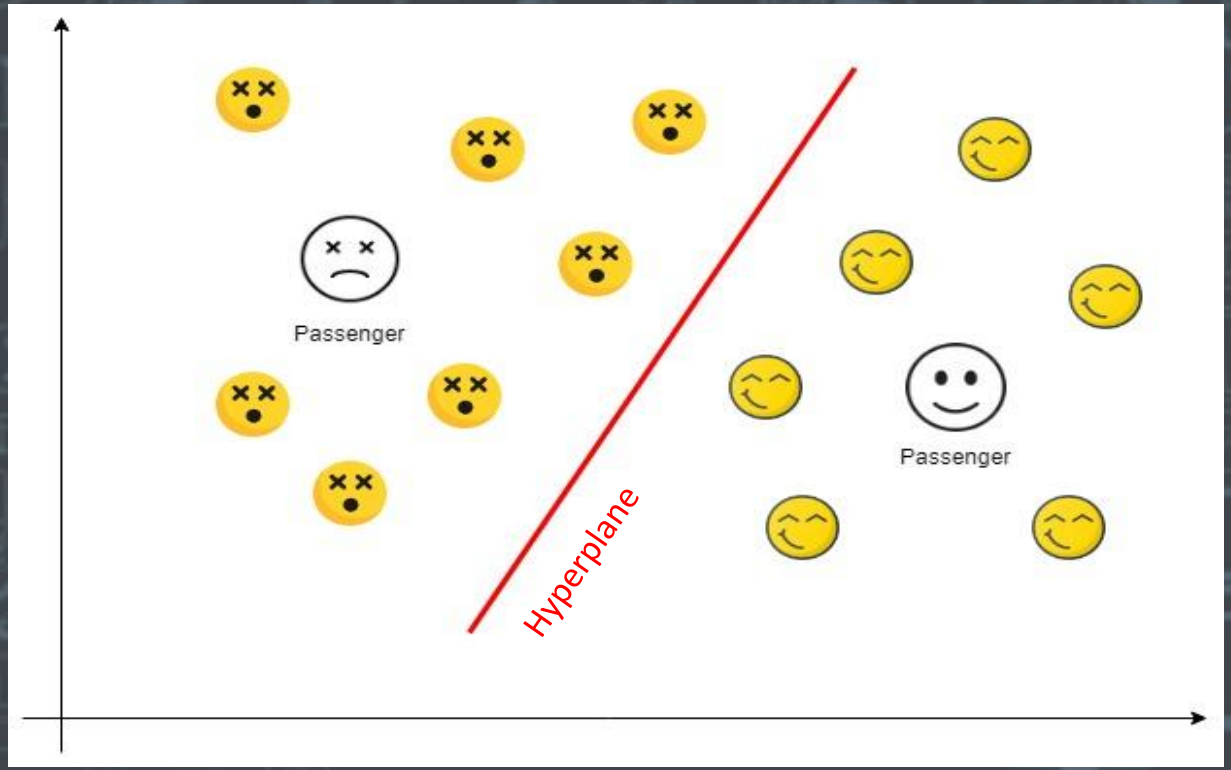
Random Forest



Gaussian Naïve Baye's

$$P(Survival|f_1, .., f_n) = \frac{P(Survival) * P(f_1, .., f_n|Survival)}{P(f_1, .., f_n)}$$

Support Vector Machine



Result

Classifier	Accuracy after k-fold cross validation	Accuracy on test data
K Nearest Neighbors	79.19%	81.25%
Decision Tree	79.64%	70.40%
Random Forest	79.64%	73.99%
Gaussian Naïve Baye's	77.71%	80.71%
Support Vector Machine	82.93%	84.30%