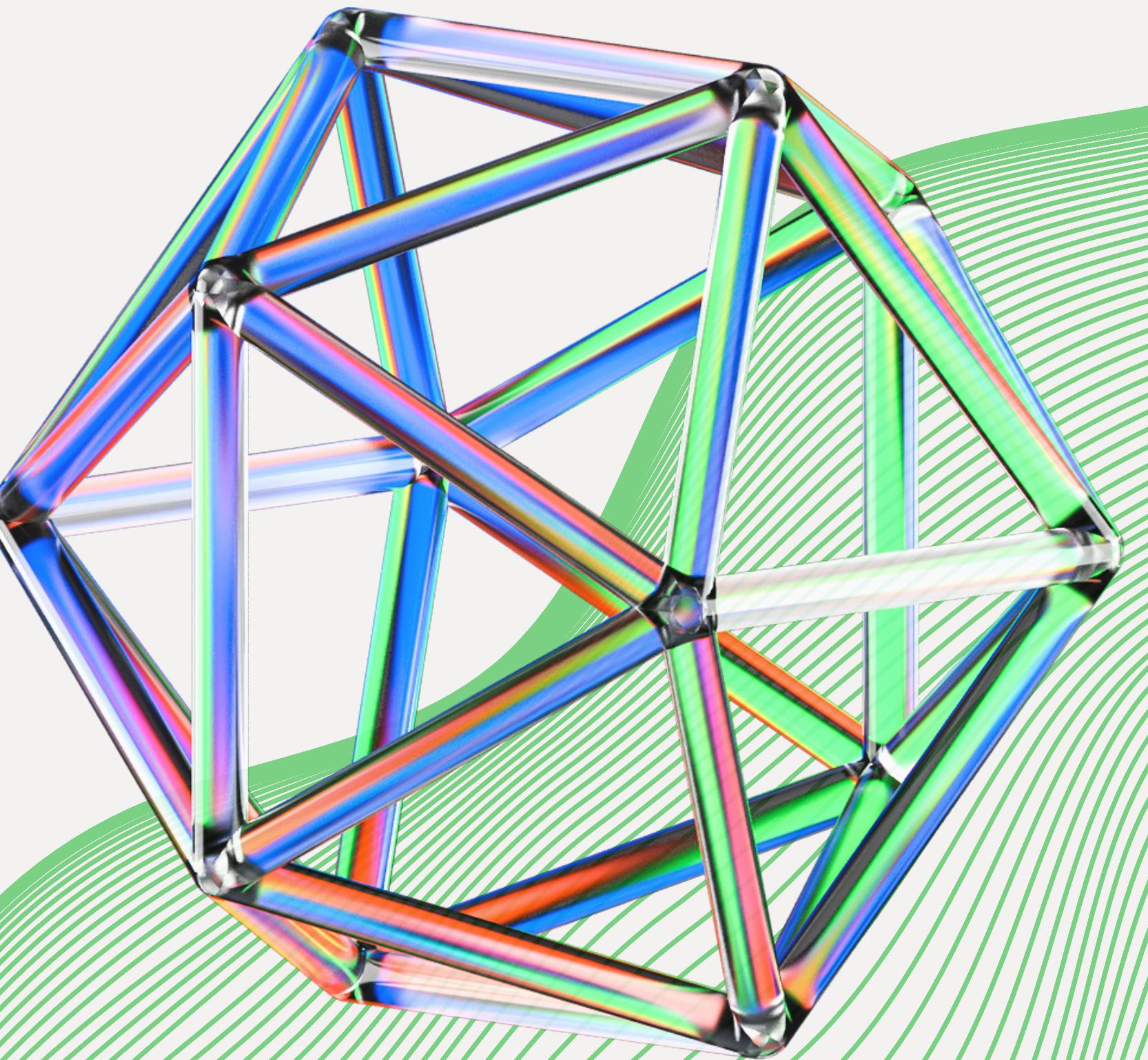
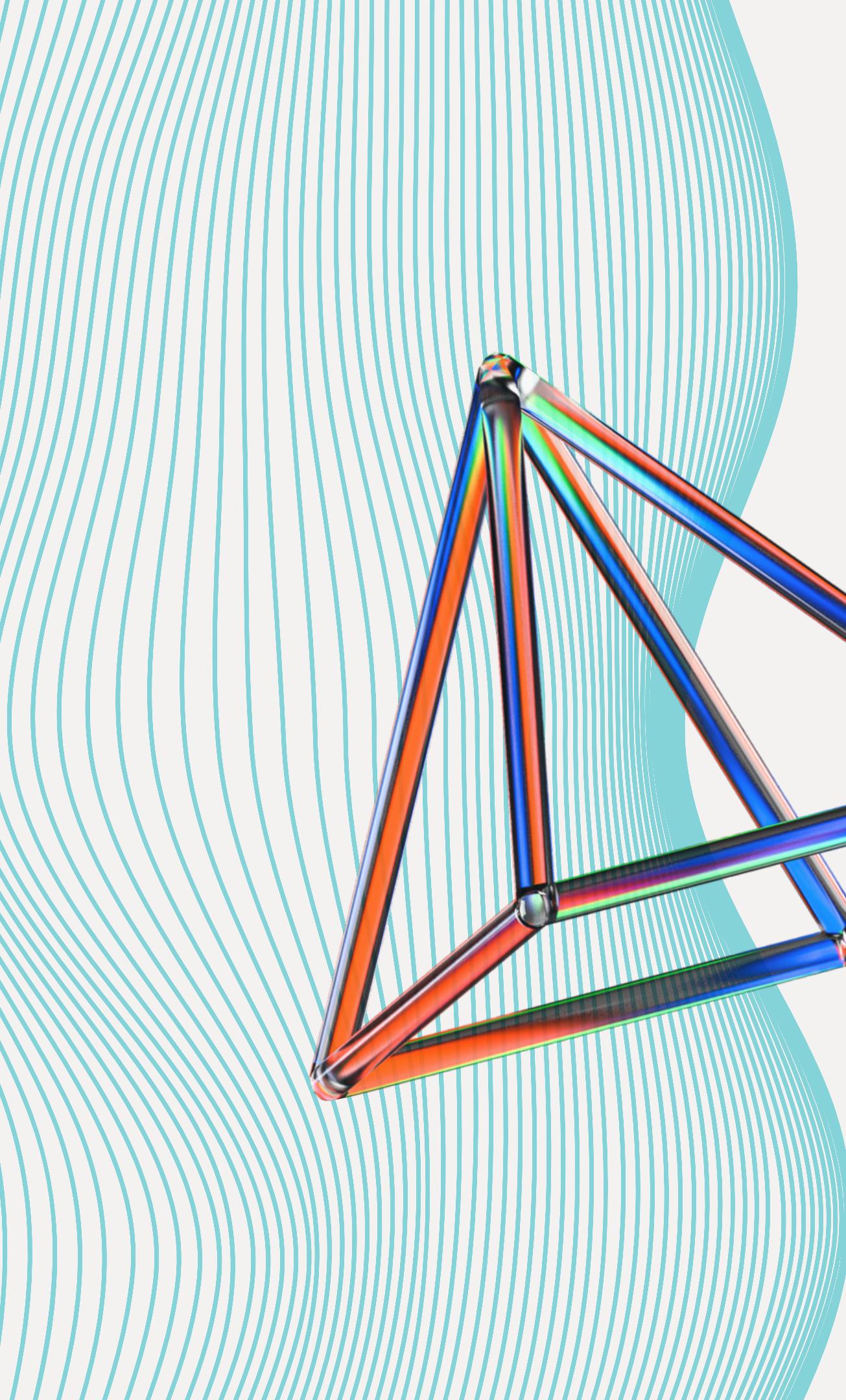


Machine Learning II

PREDICTING EMPLOYEE ATTRITION

S2 Group 7





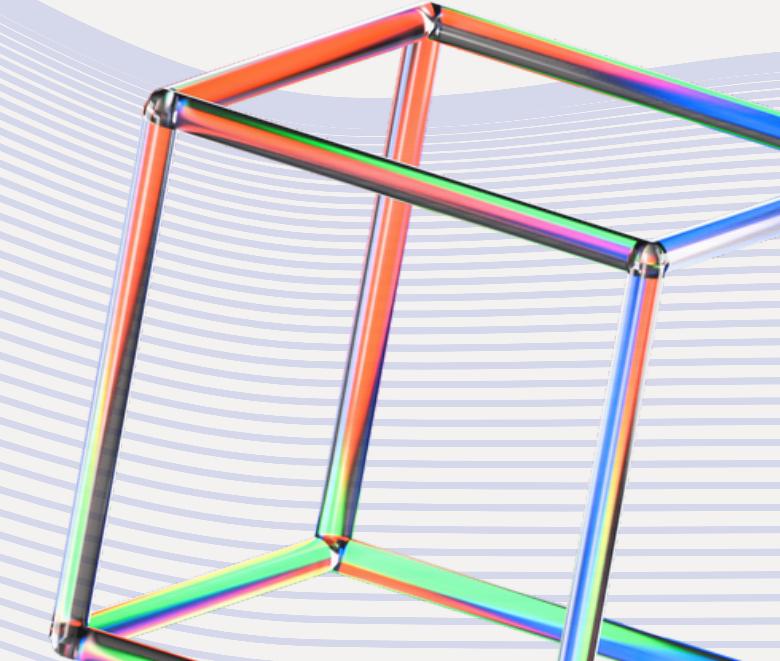
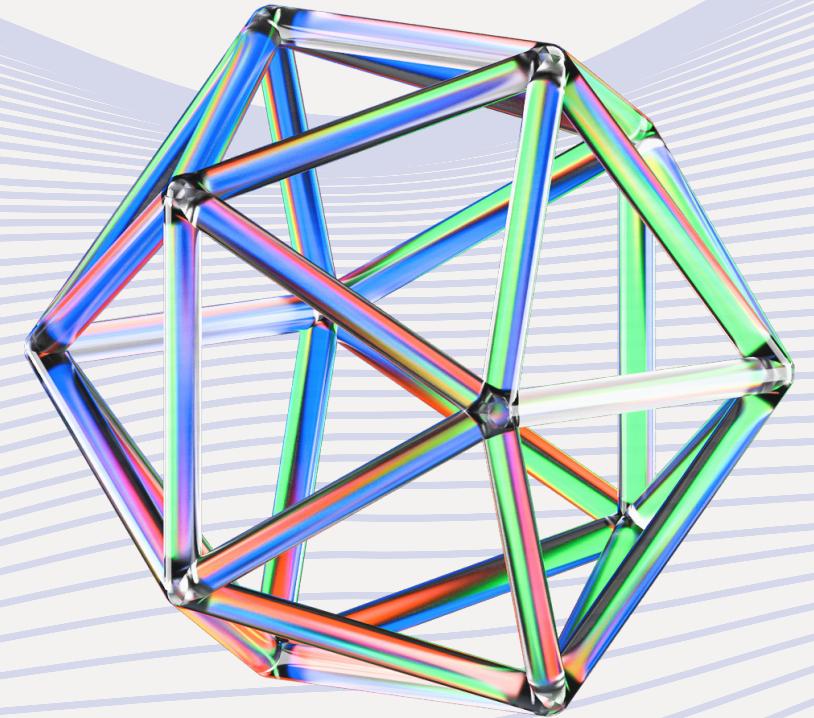
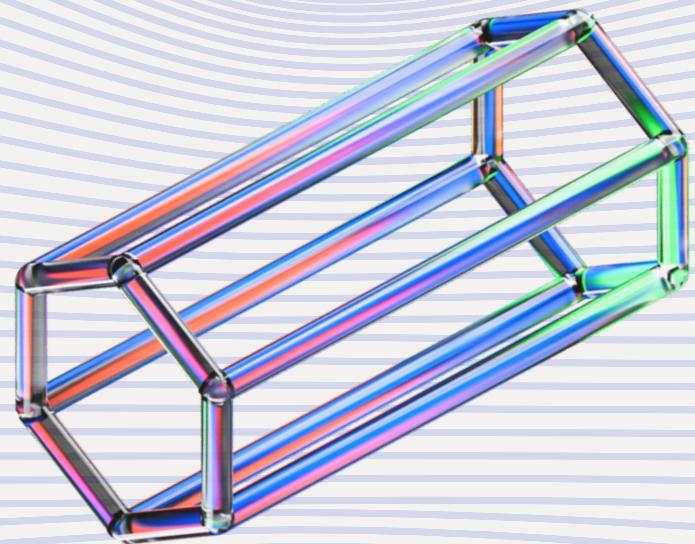
INDEX

- 01 Dataset
- 02 Feature eng.
- 03 Algorithm
- 04 Results
- 05 Conclusions
- 06 Business Case

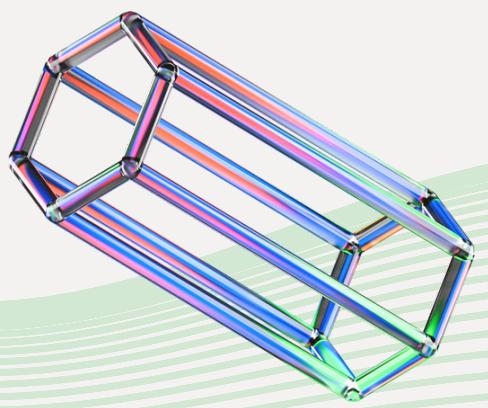
OBJECTIVES

Predict the likelihood of employees leaving the company.

Recommend actions that the company can take to alleviate attrition.

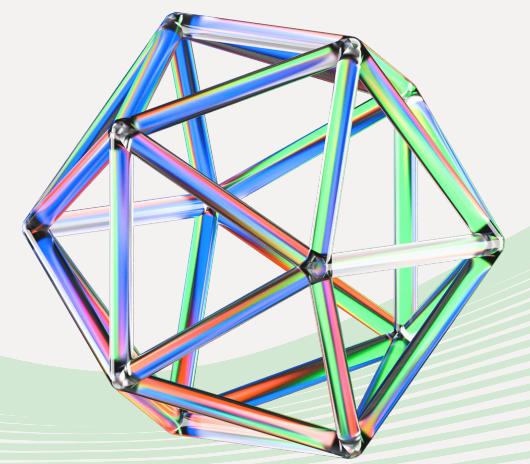


OVERALL VIEW



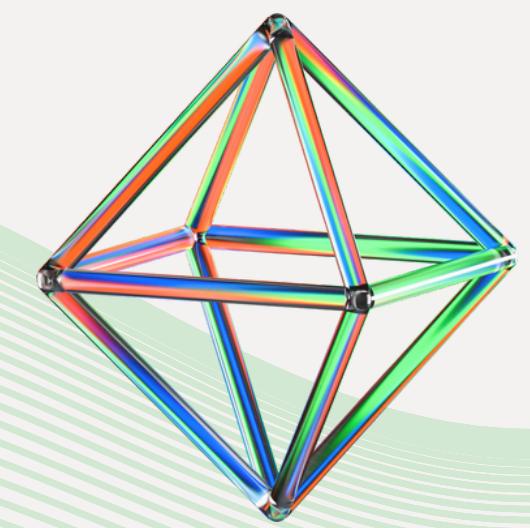
Feature Engineering

Transforming raw data into clean informative features



Algorithm

Choosing the **best** model to obtain the **best** performance



Results

Analysing the results & establish **recommendations**

DATASET

EDA

- Box Plot
- Histograms
- Profile Report
- Manual Checks

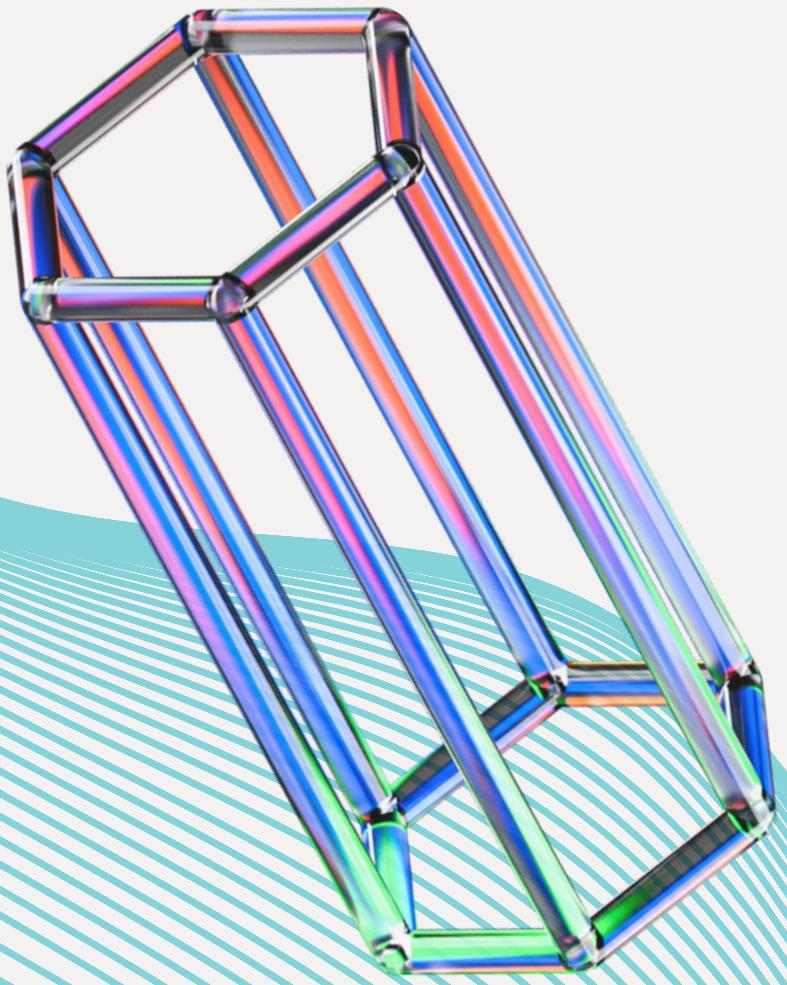
EXTRACT & SELECT

- Extracting 6 columns
- Arbitrary columns
- Exclude Over18, ID, StandardHours, EmployeeCount, PerformanceRating

CLEAN

- Dropping arbitrary columns
- Imputes mismatched entries
- Fill missing values

FEATURE ENGINEERING



FEATURE CREATION

1 Outlier-specific columns

Age and Distance From Home

2 "Total Satisfaction"

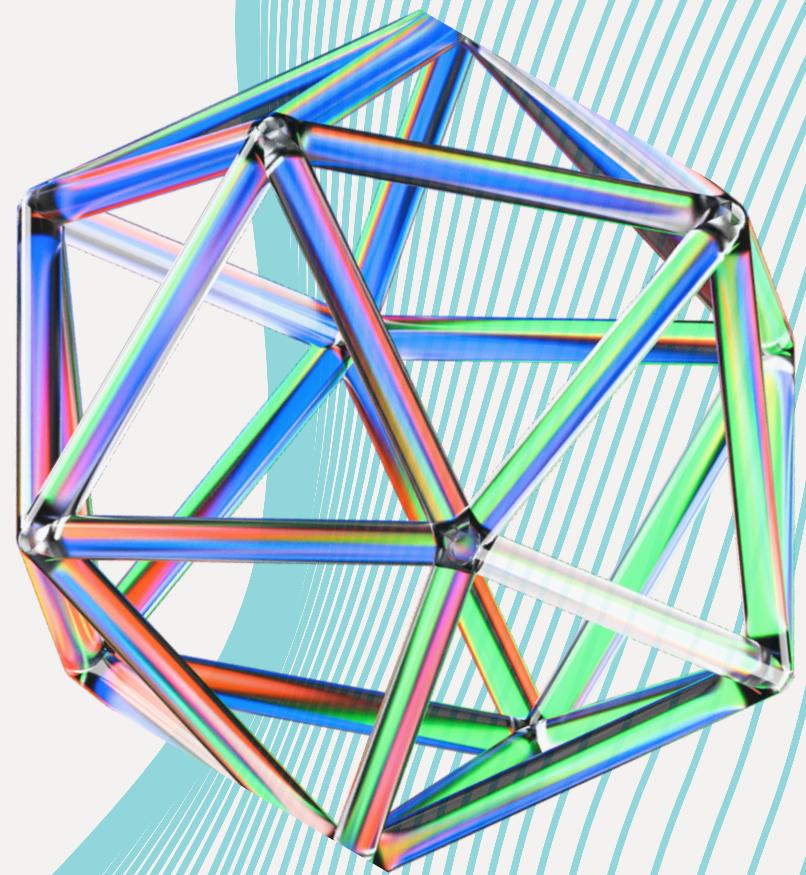
a compound score of Job Involvement, Environment Satisfaction, Job Satisfaction, and Relationship Satisfaction

3 New ratios:

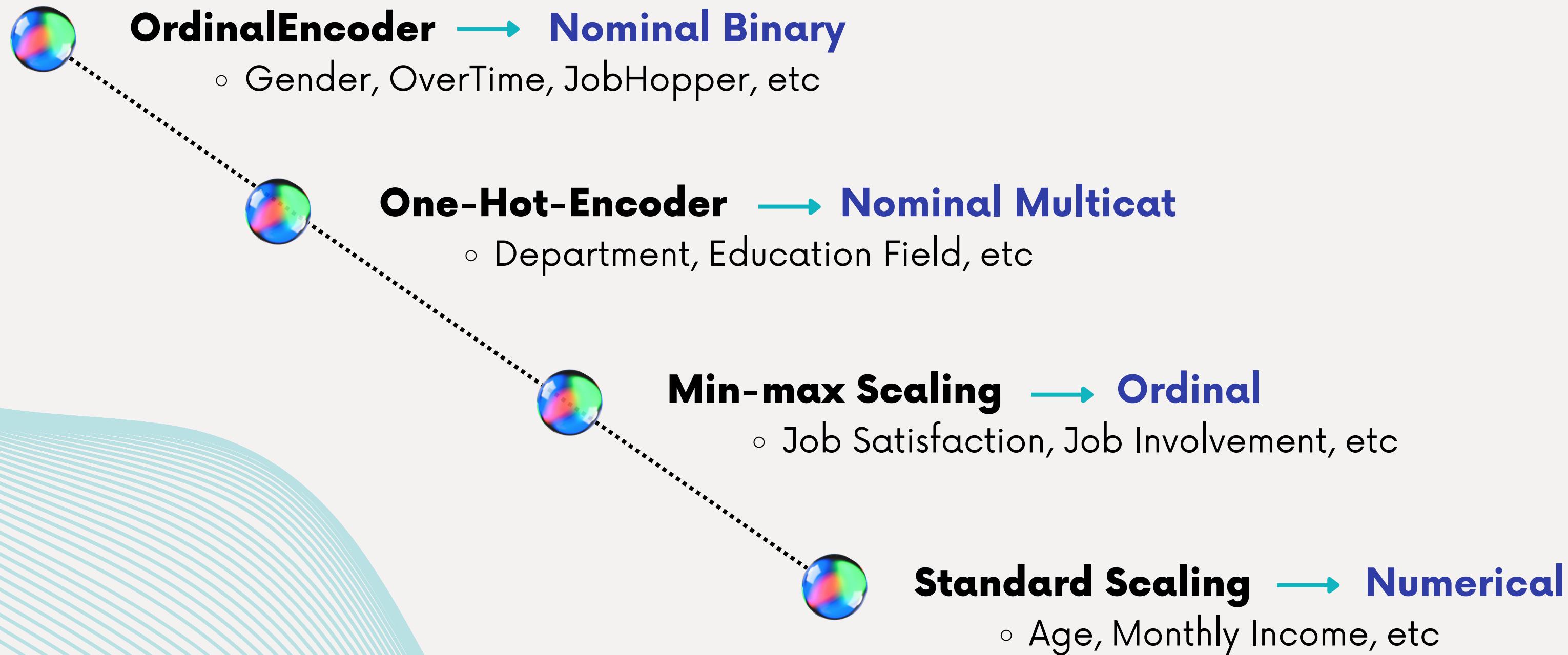
- Years With Current Manager / Total Working Years
- Total Working Years / Number of Companies Worked

4 "Job Hoppers"

Employees having worked in >4 companies, with less than 2.14 years experience on average in each company



FEATURE TRANSFORMATION & PIPELINE



ALGORITHM

AUC score on test set

XGBOOST
0.9083



RANDOM FOREST

0.8423

LOGISTIC REGRESSION

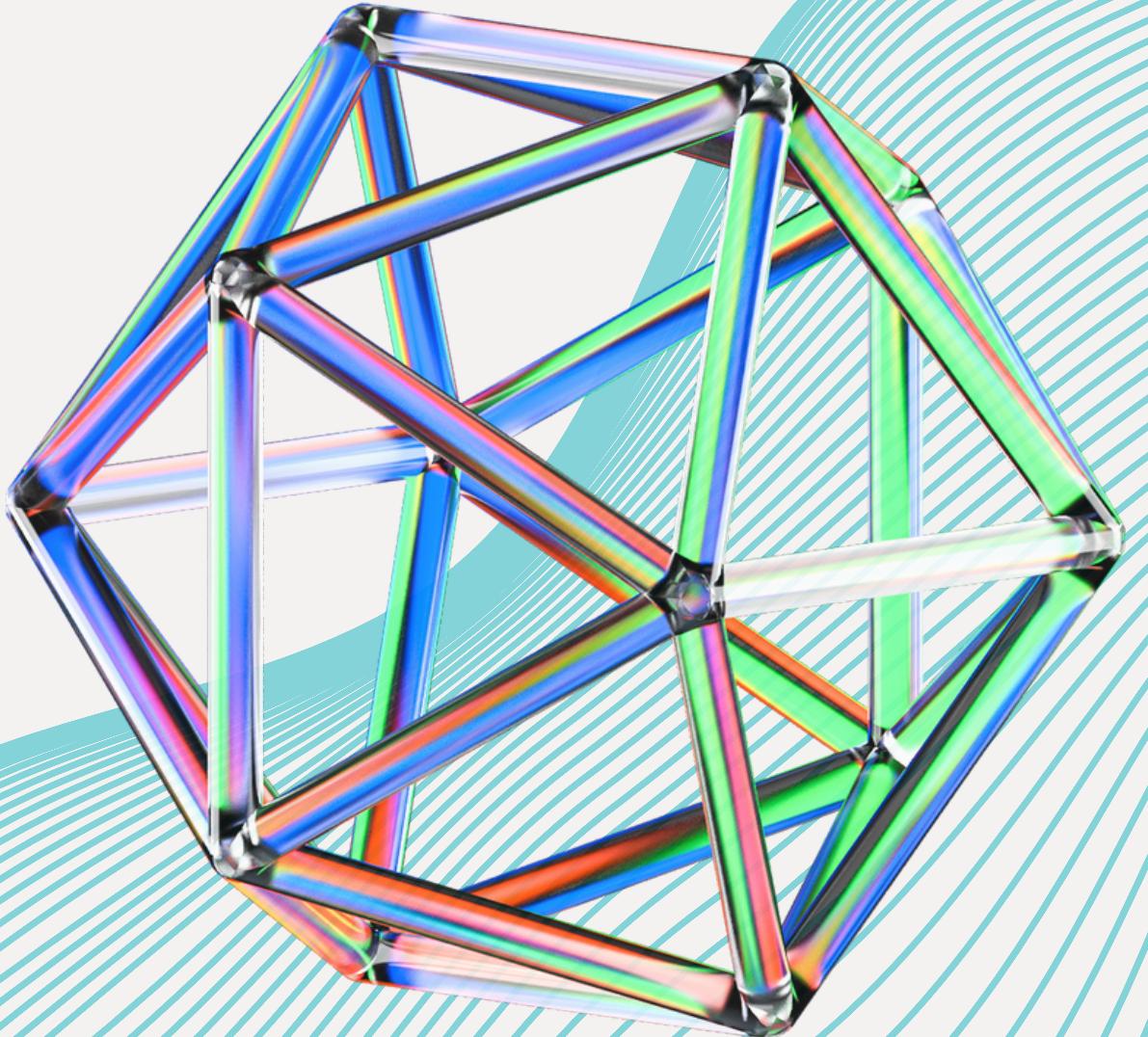
0.8270

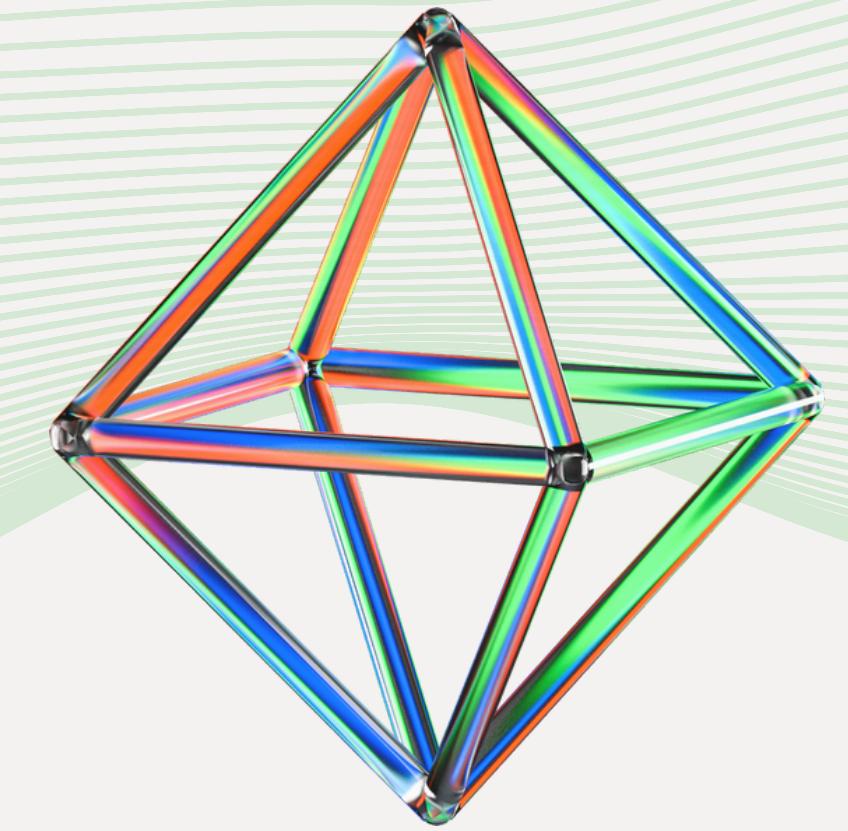
SUPPORT VECTOR MACHINES

0.8113

DECISION TREE

0.7144

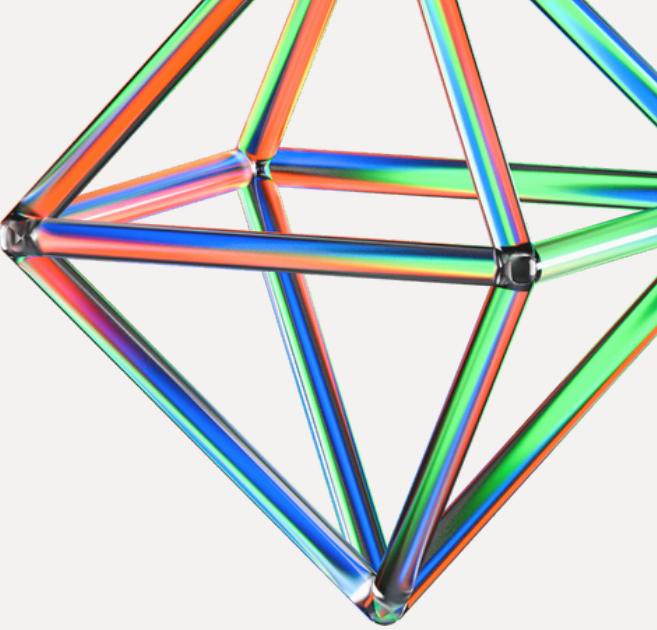




RESULTS

Identify the employees that are more likely to **leave the company**

XGBOOST RESULTS



0.903

ROC AUC

.22

Optimal threshold
Minimize the FN and FP

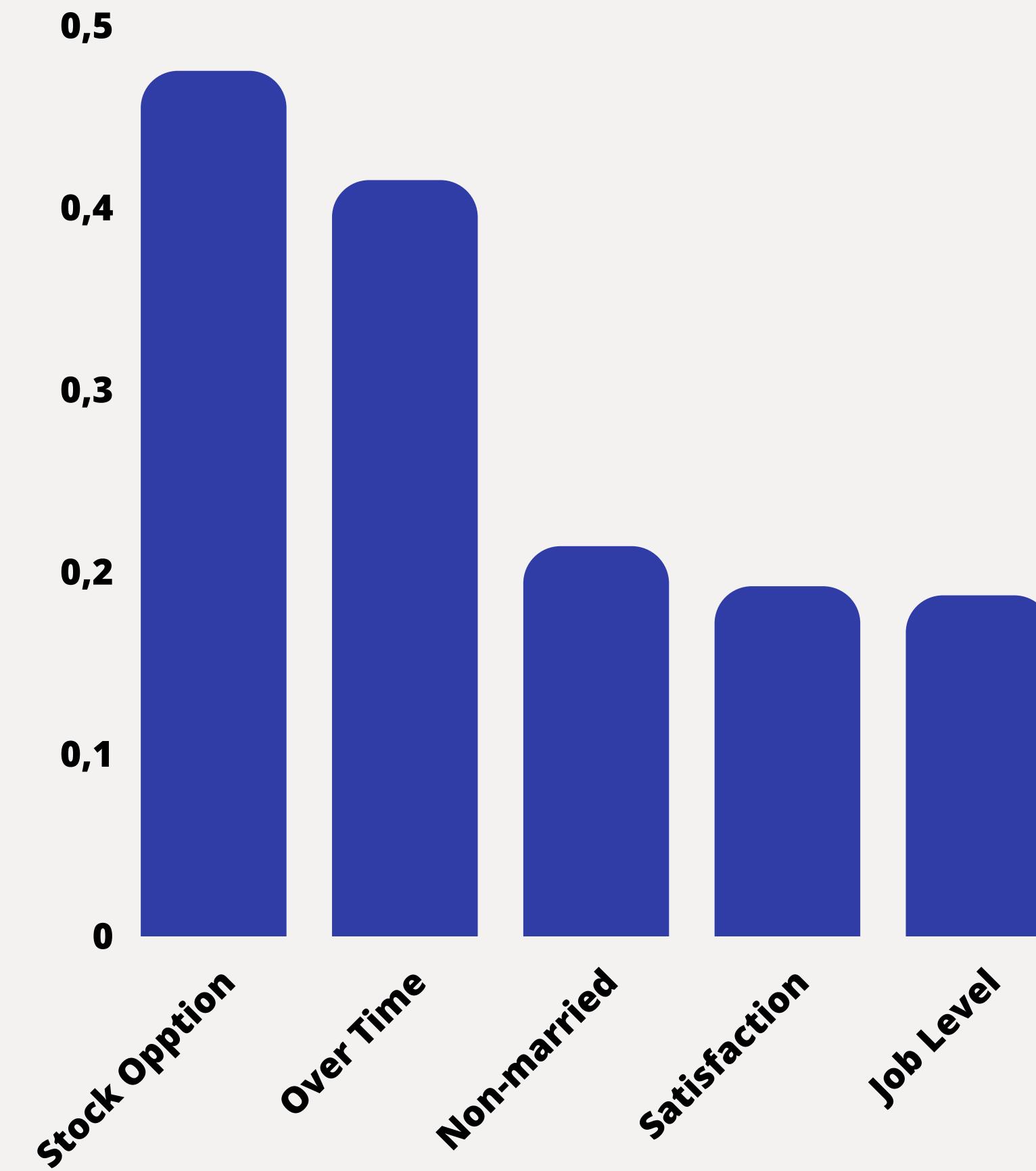
Most Relevant Features

JobInvolvement, WorkLifeBalance, TotalSatisfaction, Age, HourlyRate,
PercentSalaryHike, DistanceFromHome, StockOptionLevel, OverTime

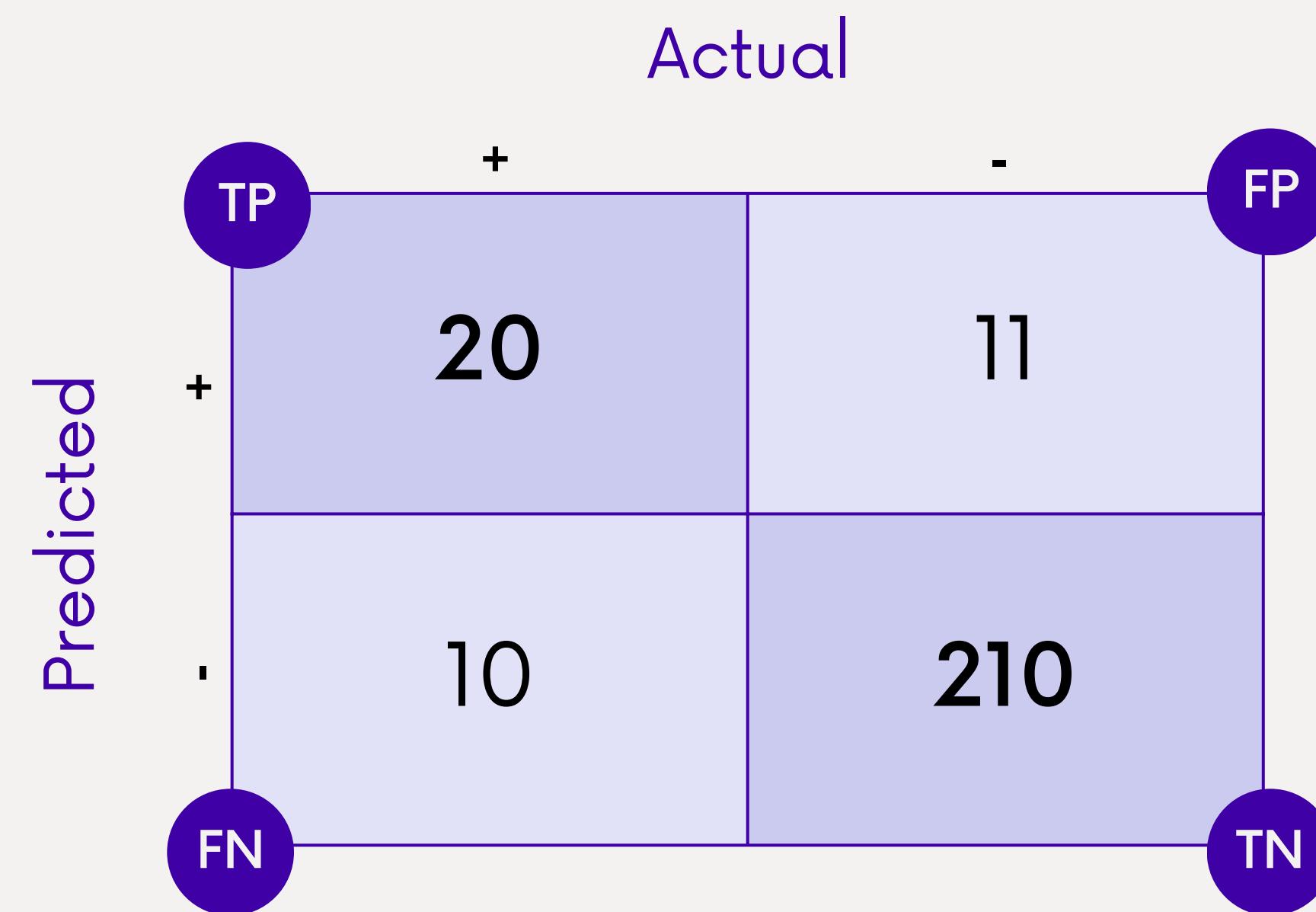
PREVENT OVERFITTING

Parameters used:

- max_depth, n_estimators, learning_rate, subsample, reg_alpha, reg_lambda, min_child_weight
- Cross Validation (Kfold = 6)
- Grid Search CV - Zoom-in
- SHAP Values to explain model predictions and identify important features. Guiding decision making.



CONFUSION MATRIX



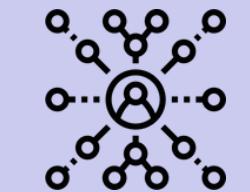
RECOMMENDATIONS



Stop
unnecessary
Over Time Work



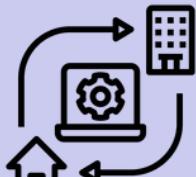
Improve the
**Work Life
Balance**



Expand
**Equity
Program**

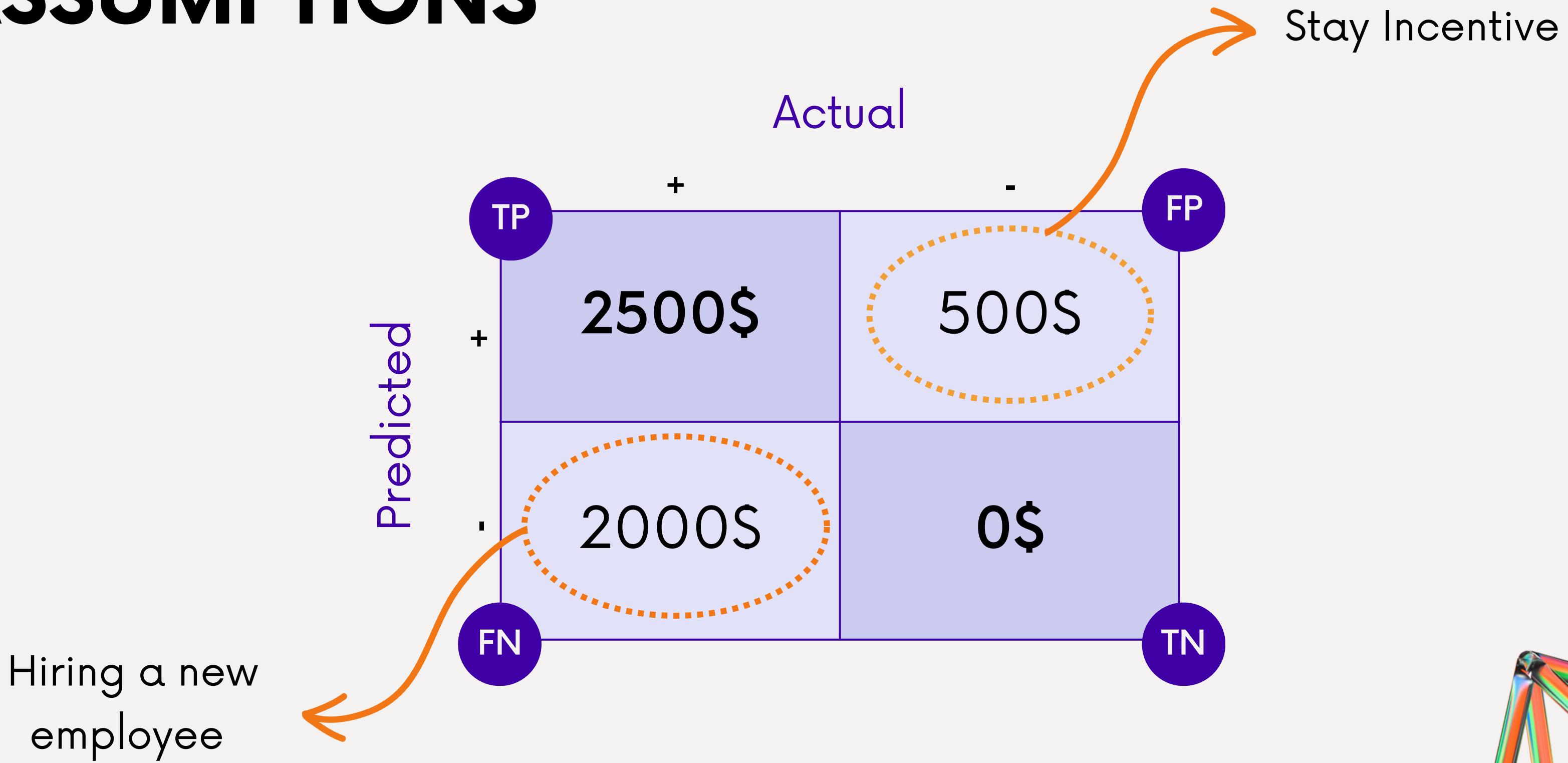


**Financial
Incentives**
based on
performance



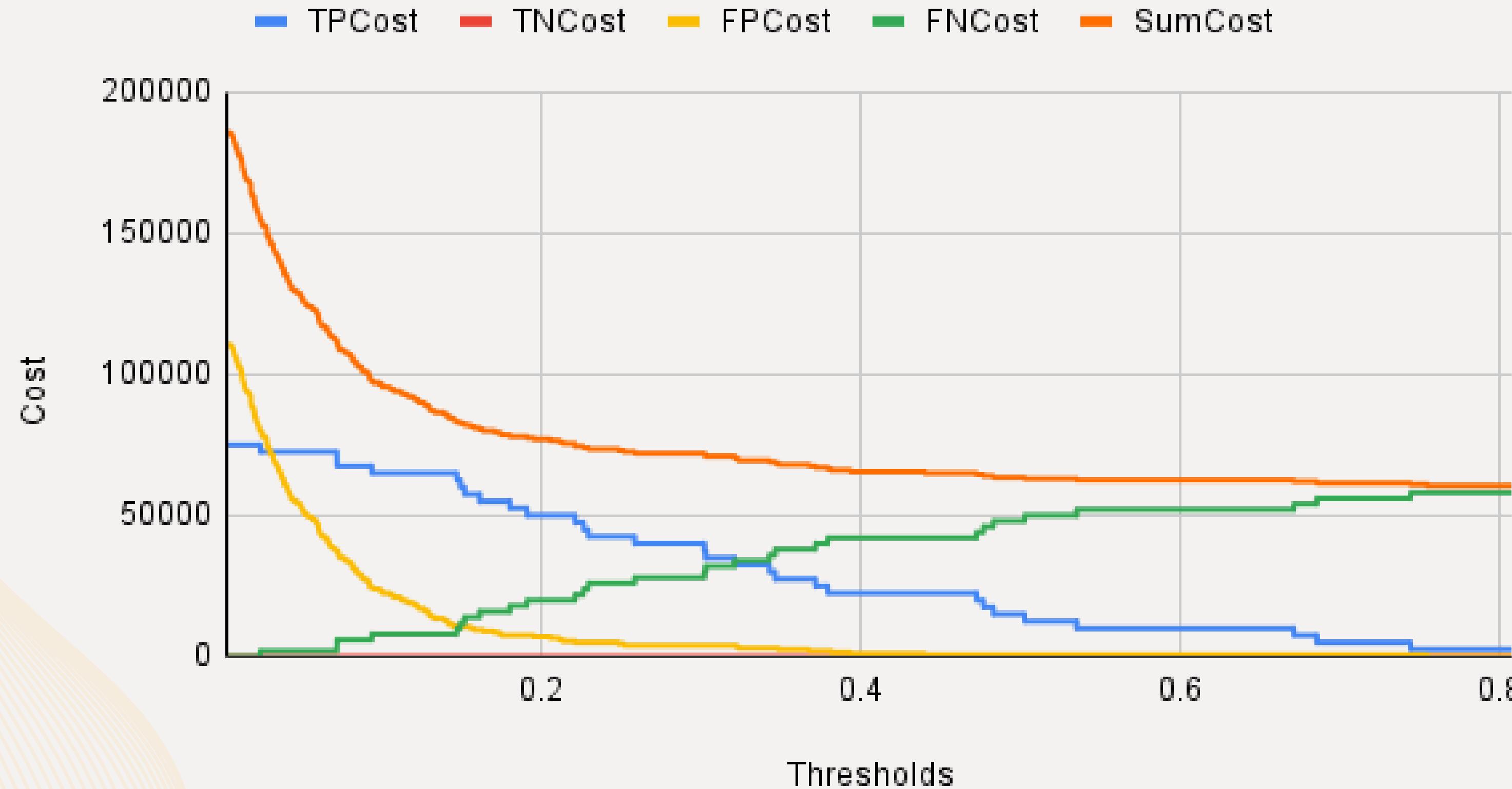
Set up a
**Hybrid
Model**

ASSUMPTIONS



Assuming that the company wants to keep all employees

IMPLICATIONS - COST BREAKDOWN



30%
252

Are leaving...

SCENARIO

THRESHOLD 0.33

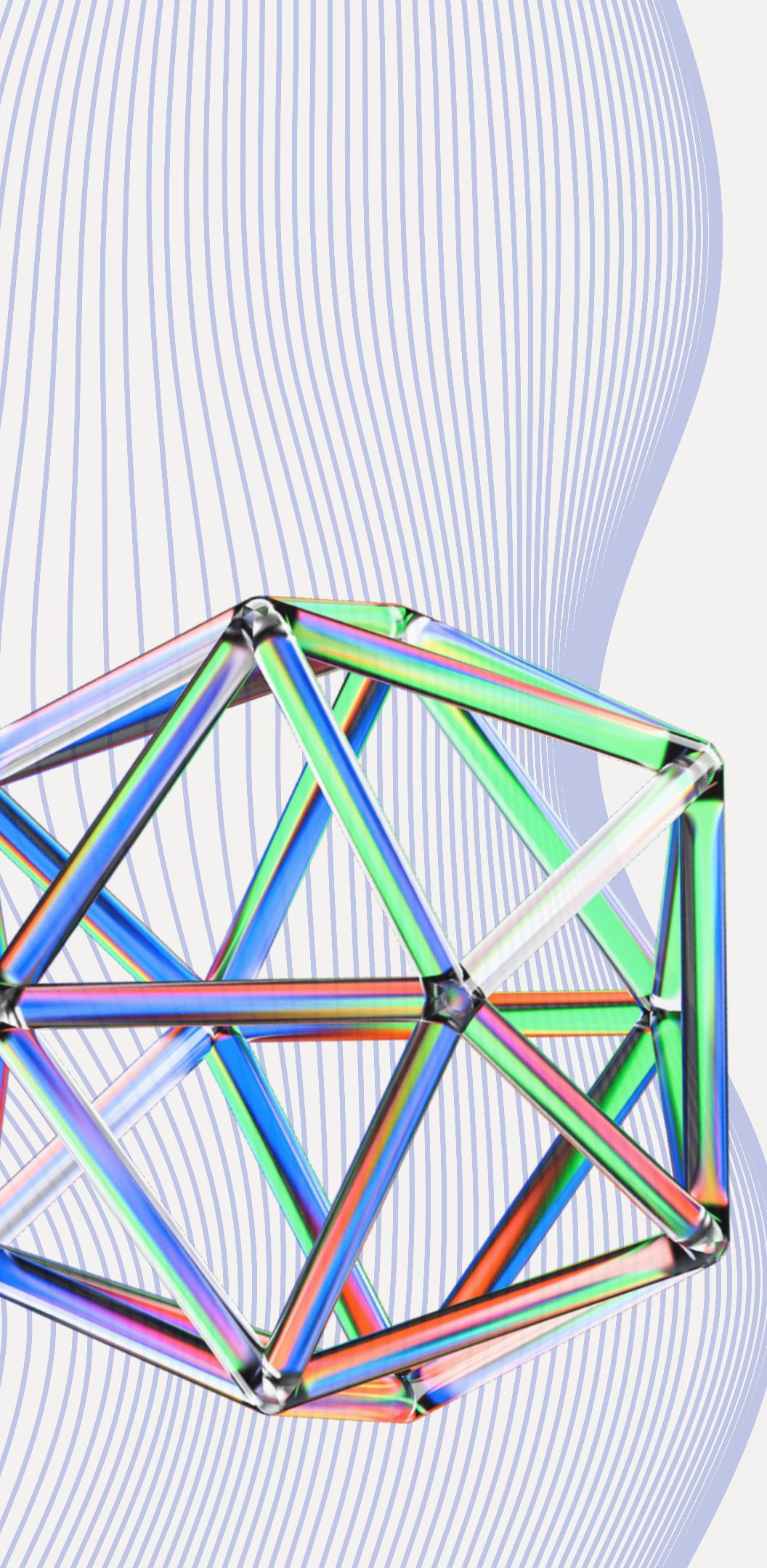
13 incentivised
7 did not know will leave
6 people were going to stay

\$69.5K

THRESHOLD 0.45

9 incentivised
21 did not know will leave
1 incentivized and stayed anyways

\$65K



CONCLUSION

- Incentivize employees in accordance with important features.
- Use the model to decide how you would like to spend your HR budget
- Monitor whether incentives are working on employees
- More data – Retrain

THANK YOU

