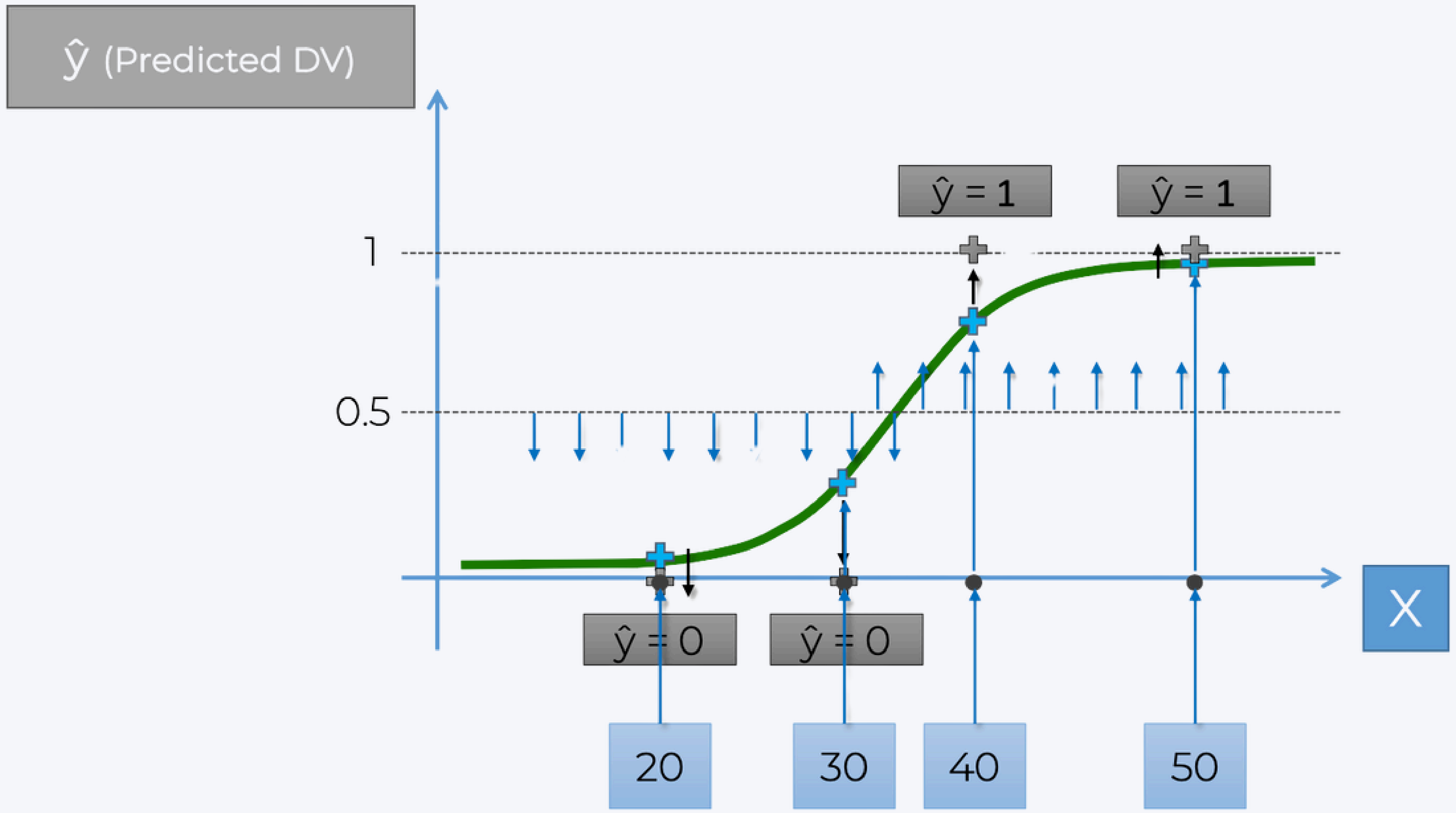
أكاديمية طويق
Tuwaiq Academy

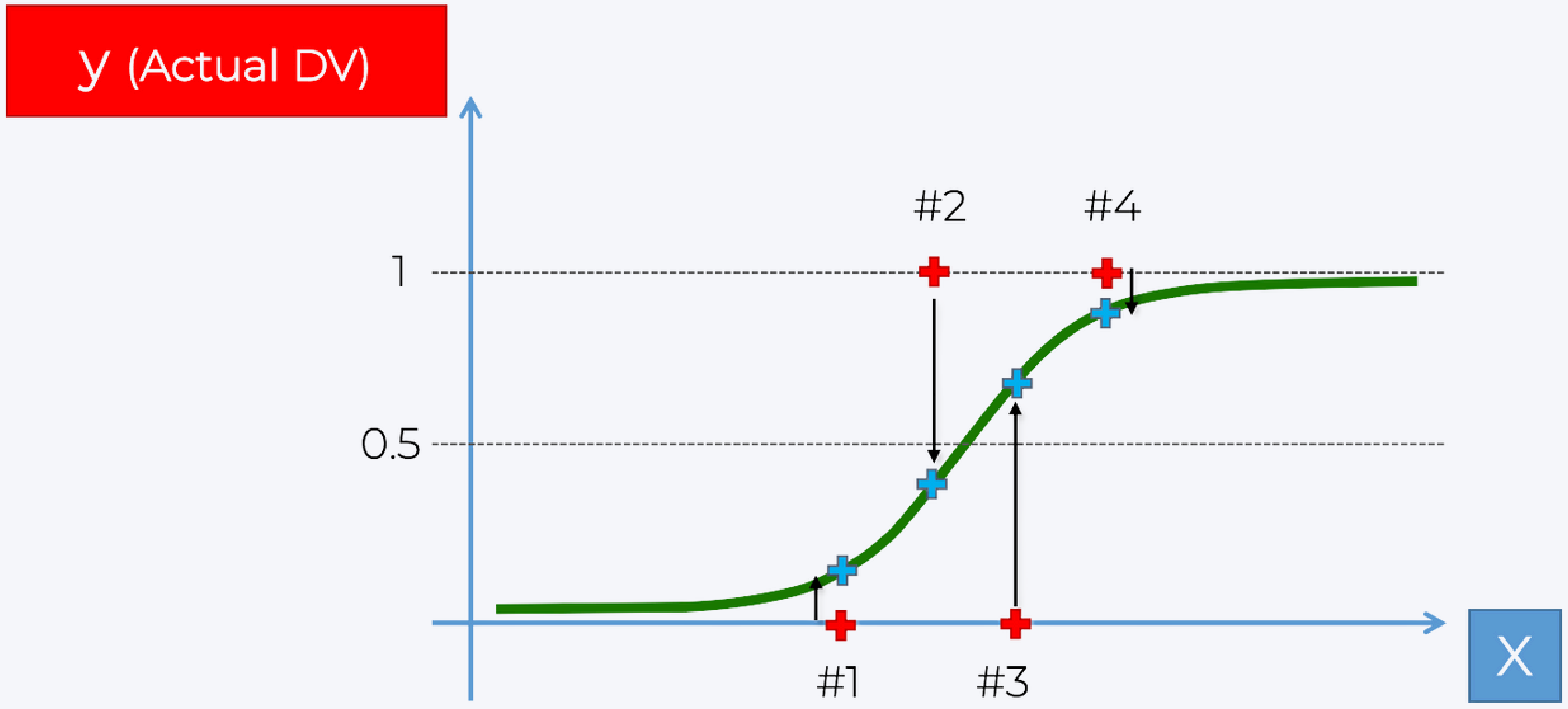
Evaluating Classification Model

Evaluating Classification Model

False Positives & False Negatives

When evaluating a classification model, we often deal with prediction errors.





https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

False Positives & False Negatives

False Positives (Type I Error)

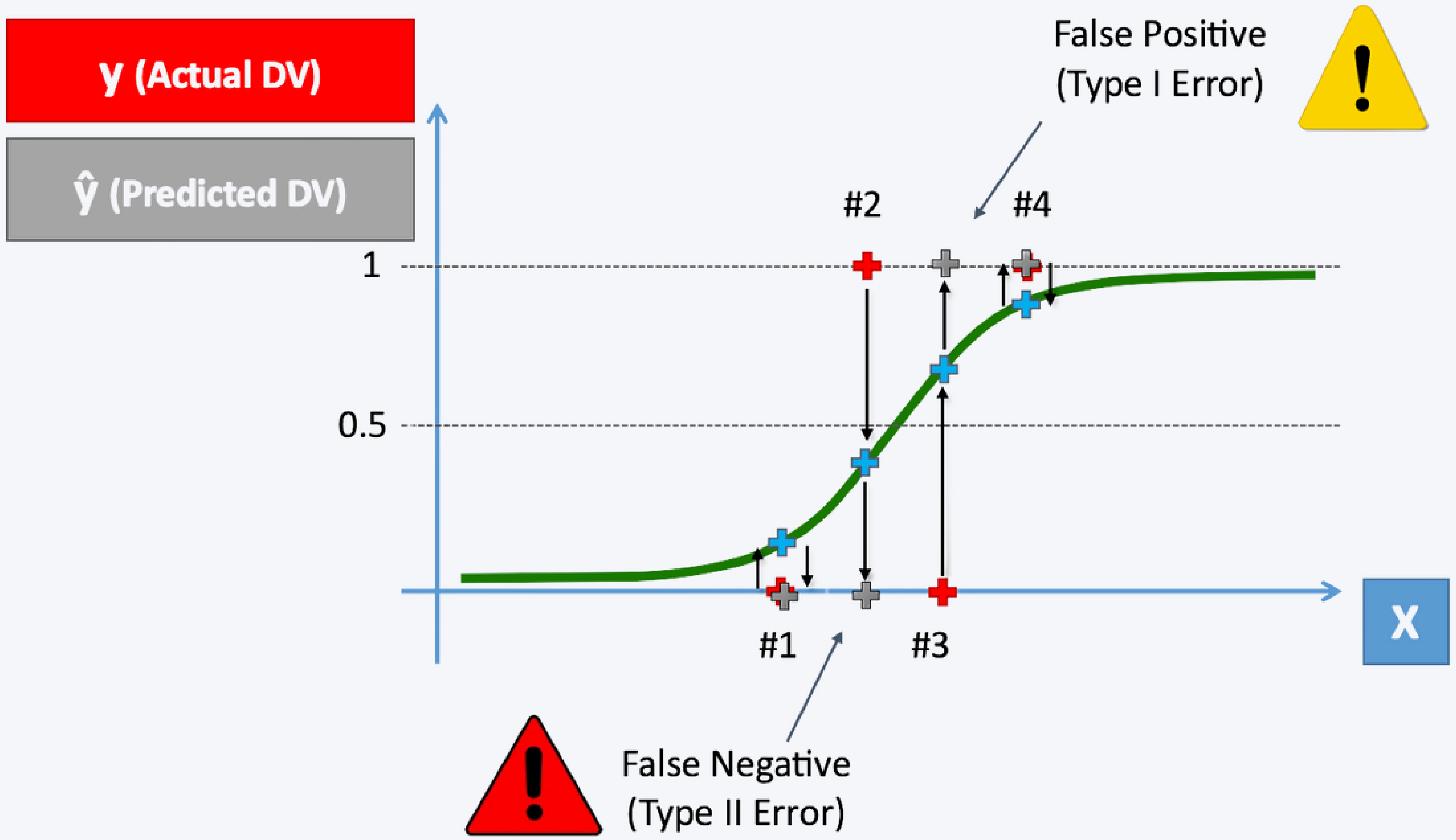
- **The model incorrectly predicts a positive outcome when the actual outcome is negative.**

- **Example**: A medical test predicts a patient has a disease when they do not.

False Negatives (Type II Error)

- **The model incorrectly predicts a negative outcome when the actual outcome is positive.**

- **Example**: A medical test predicts a patient does not have a disease when they actually do.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

How This Relates to Model Evaluation

- These errors are crucial for measuring model performance.
- A good classification model should minimize both types of errors while maintaining high accuracy.
- Depending on the application, one type of error may be more costly than the other (e.g., in fraud detection, false negatives can be more dangerous).

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

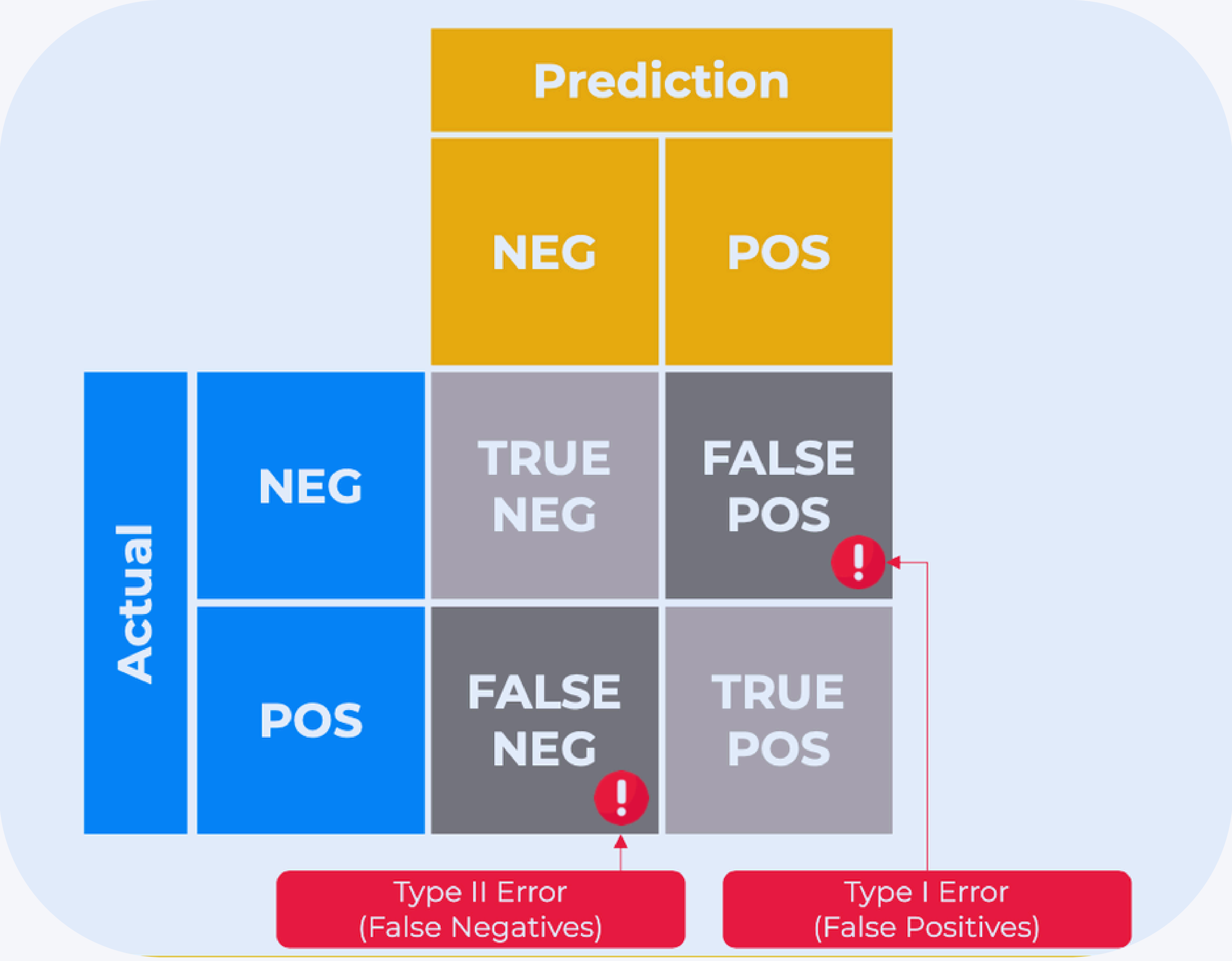
Confusion Matrix & Accuracy

The confusion matrix is a fundamental tool for evaluating classification models.

It provides a summary of prediction results by comparing actual and predicted values.

The main components are:

- **True Positive (TP):** The model correctly predicted the positive class.
- **True Negative (TN):** The model correctly predicted the negative class.
- **False Positive (FP) (Type I Error):** The model incorrectly predicted positive when the actual class was negative.
- **False Negative (FN) (Type II Error):** The model incorrectly predicted negative when the actual class was positive.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

Confusion Matrix & Accuracy

		Prediction	
		NEG	POS
Actual	NEG	43	12
	POS	4	41

Type II Error
(False Negatives)

Type I Error
(False Positives)

Accuracy Rate and Error Rate:

$$AR = \frac{Correct}{Total} = \frac{TN + TP}{Total} = \frac{84}{100} = 84\%$$

$$ER = \frac{Incorrect}{Total} = \frac{FP + FN}{Total} = \frac{16}{100} = 16\%$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

Precision, Recall, and F1-Score Explanation

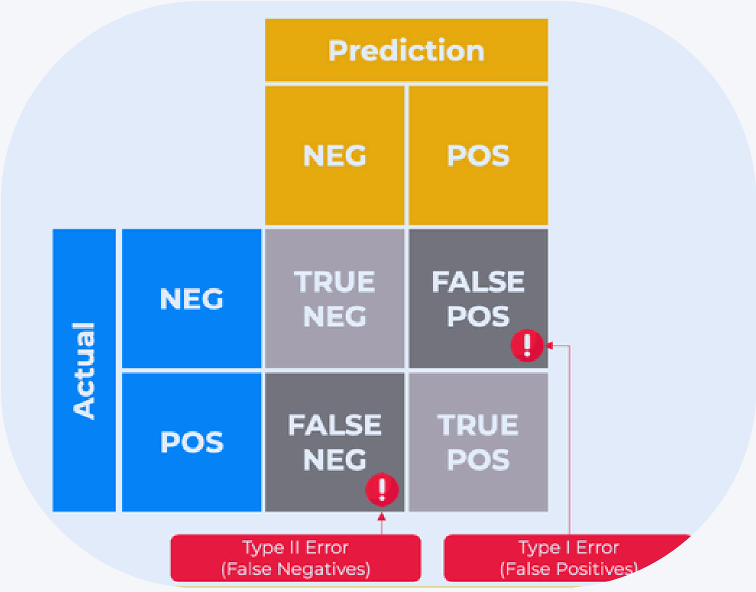
Precision (Positive Predictive Value):

Precision measures the accuracy of positive predictions made by the model.

It answers:
"Of all the instances predicted as positive, how many were actually positive?"

High Precision: Few false positives, meaning when the model predicts positive, it is often correct.

Low Precision: Many false positives, meaning the model is often incorrect in predicting positives.



$$\text{Precision} = \frac{TP}{TP + FP}$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

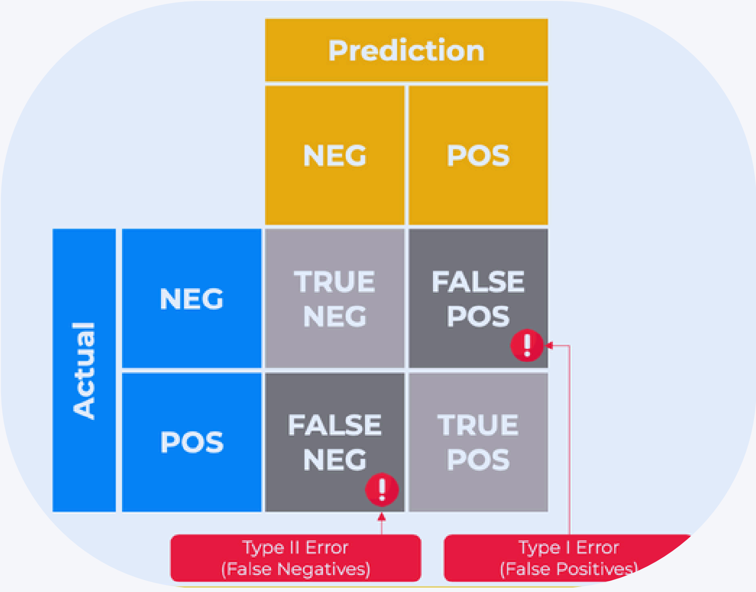
Precision, Recall, and F1-Score Explanation

Recall (Sensitivity or True Positive Rate):

Recall measures how well the model identifies all actual positive cases.

It answers:
"Of all the actual positive cases, how many did the model correctly identify?"

- **High Recall:** Few false negatives, meaning most actual positives are correctly classified.
- **Low Recall:** Many false negatives, meaning the model misses many actual positives.



$$\text{Recall} = \frac{TP}{TP + FN}$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

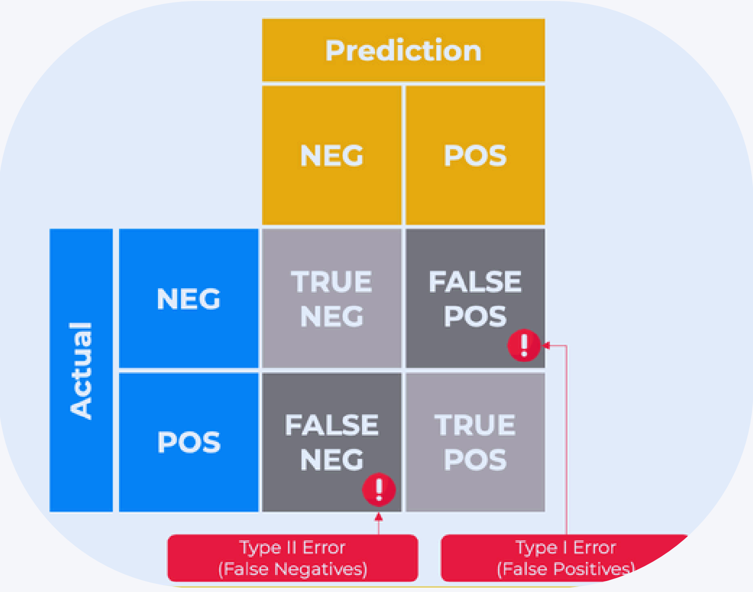
Evaluating Classification Model

Precision, Recall, and F1-Score Explanation

F1-Score (Harmonic Mean of Precision & Recall)

The F1-score balances Precision and Recall.

It is useful when we want a single metric that considers both false positives and false negatives.



$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

Precision, Recall, and F1-Score Explanation

Key Insights:

✅ Precision is important when false positives are costly (e.g., spam detection, where incorrectly flagging an important email is bad).

✅ Recall is important when false negatives are costly (e.g., medical diagnoses, where missing a disease is dangerous).

✅ F1-Score is a balanced metric, especially useful when the dataset is imbalanced.

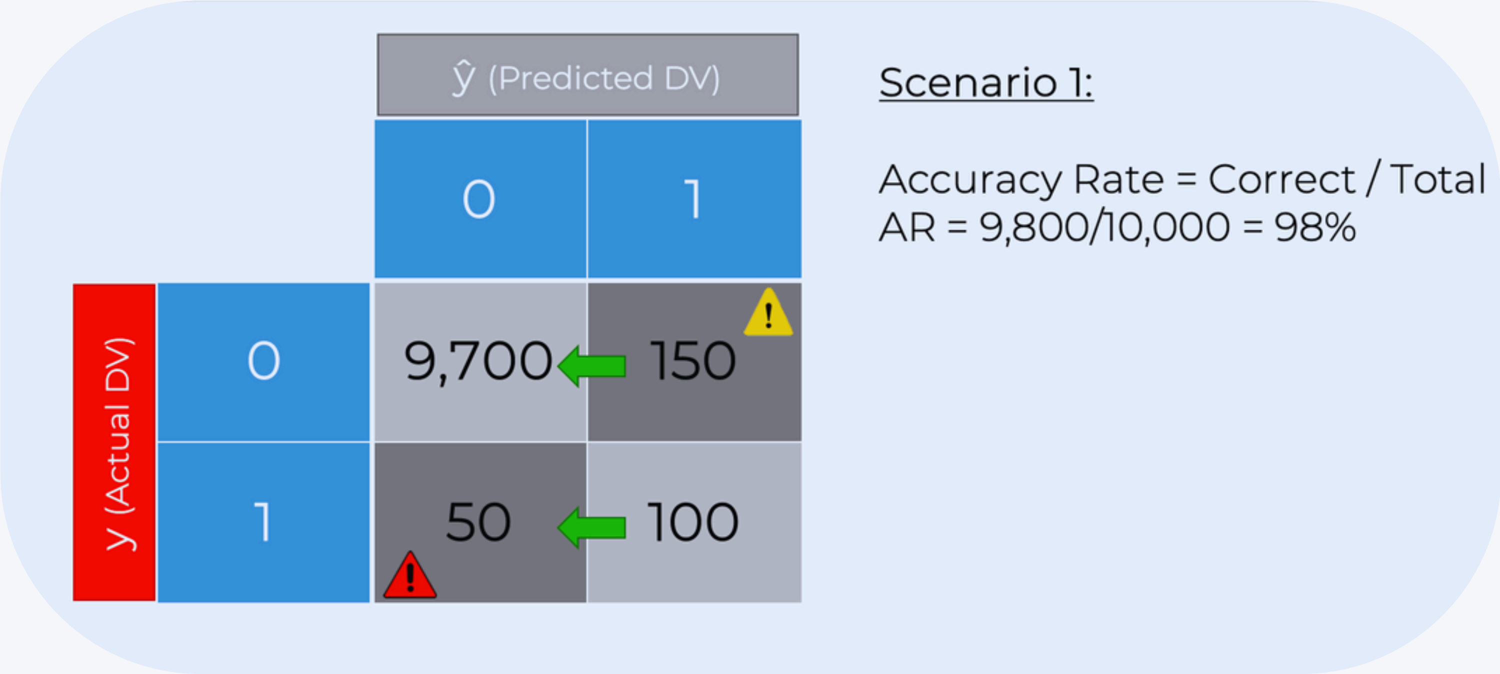
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

Accuracy Paradox - Scenario 1

- Here isa confusion matrix with two classes (0 and 1).

- The model achieves 98% accuracy, but it highlights a critical issue: **imbalanced data.**



- While most negative cases (0s) are correctly predicted (9,700), the model fails to correctly classify a significant number of positive cases (50 false negatives).

Accuracy alone does not tell the full story of model performance, especially in imbalanced datasets.

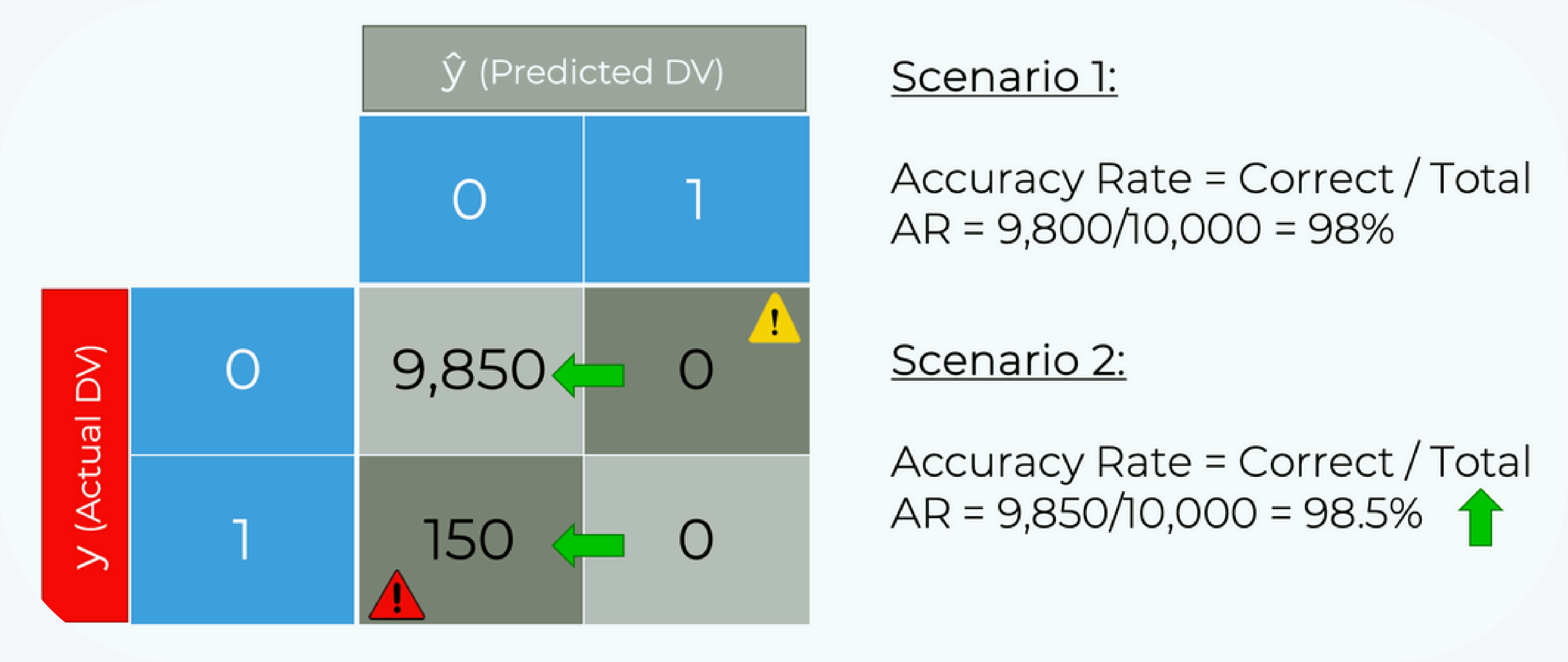
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

أكاديمية طويق
Tuwaiq Academy

Evaluating Classification Model

Accuracy Paradox - Scenario #2

- The model slightly improves, increasing accuracy to 98.5%, but it completely ignores class 1 (no correct positive predictions).
- This highlights the accuracy paradox, where accuracy increases while performance for an important class worsens.



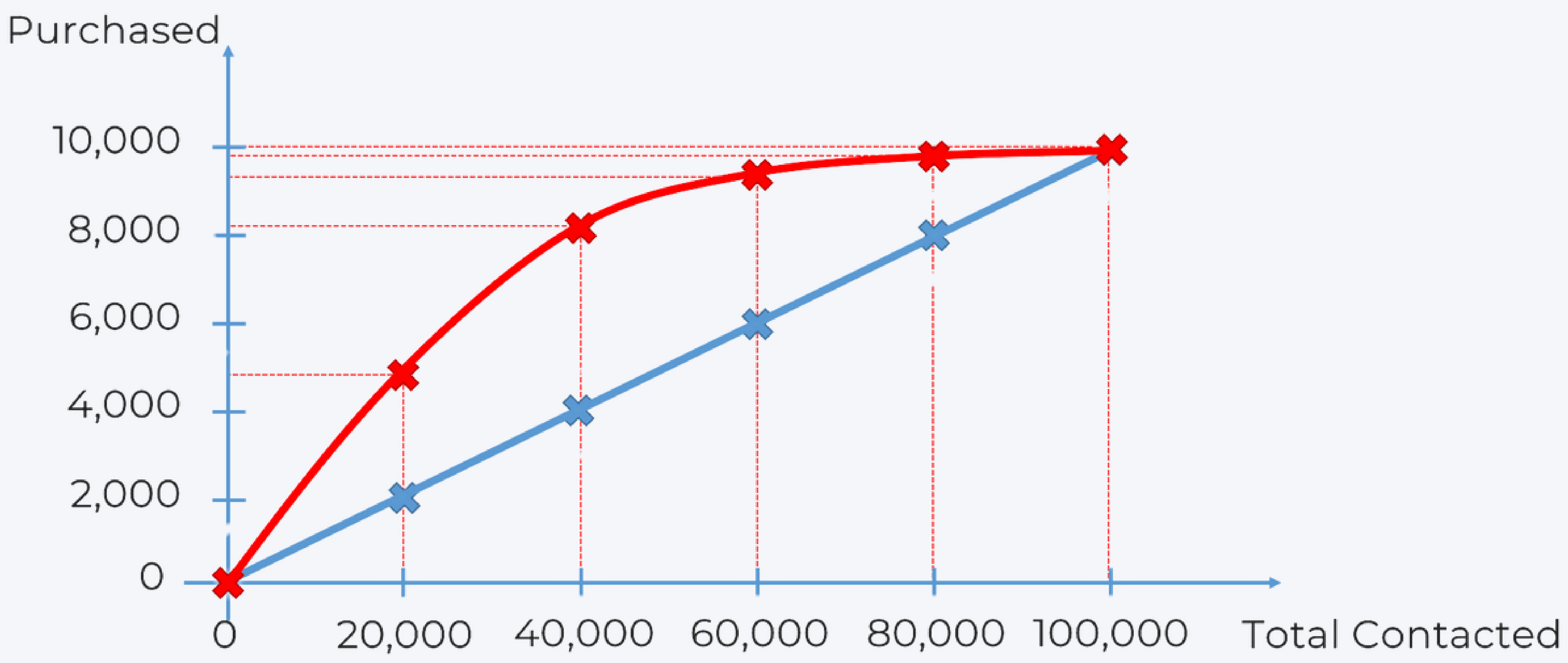
A high accuracy does not necessarily mean the model is good. Other metrics like precision and recall must be considered.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

CAP (Cumulative Accuracy Profile) - Introduction

- **CAP** is a visualization technique to evaluate classification models.



- The x-axis represents total contacted (e.g., customers, patients, etc.), and the y-axis represents the positive outcomes (e.g., purchases, disease detection).
- The red curve shows the model's performance, while the blue diagonal represents random selection.
- **A steeper curve indicates a better model.**

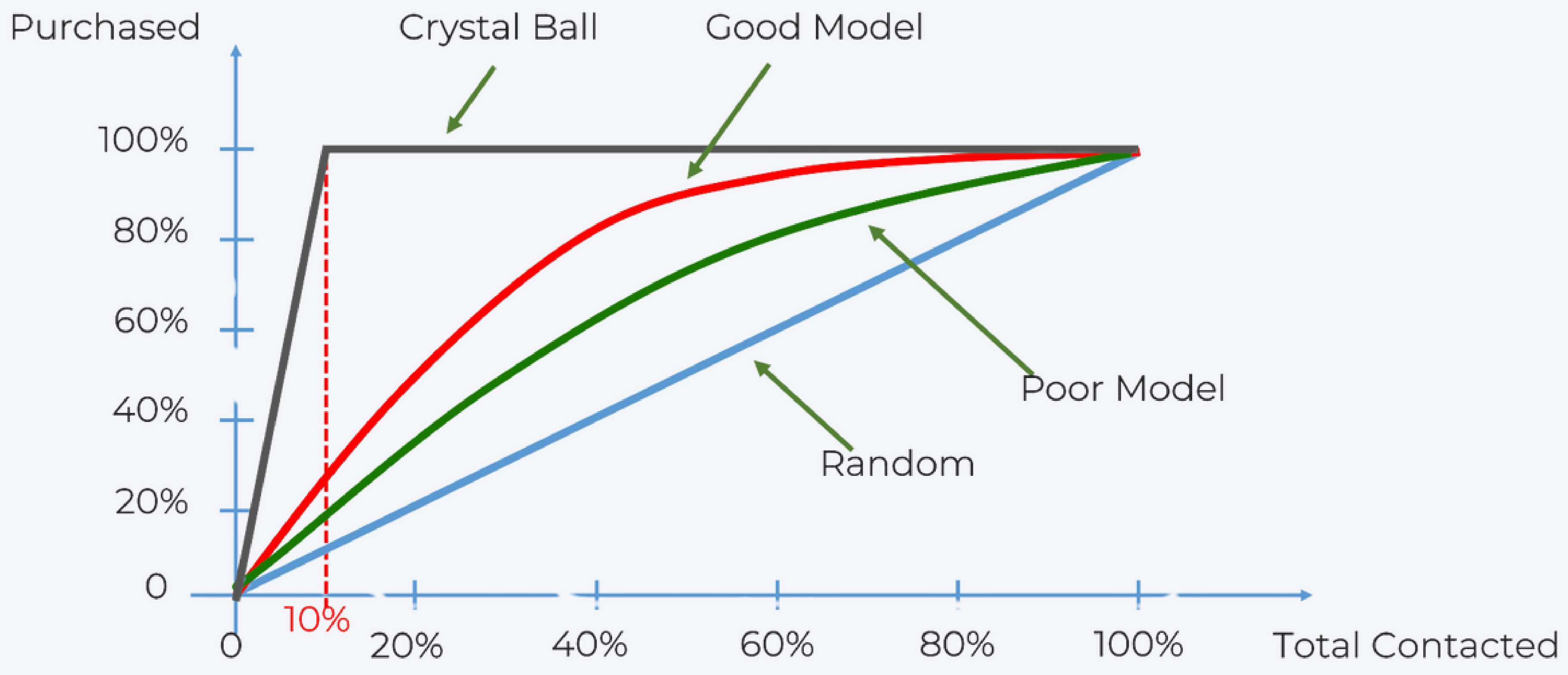
https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

CAP Curves - Performance Comparison

<u>Compares different models:</u>

- **Crystal Ball:** A perfect model (ideal case).
- **Good Model:** A practical, effective model.
- **Random Model:** A model that performs no better than random guessing.



The closer the curve is to the perfect model, the better the classifier.

https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

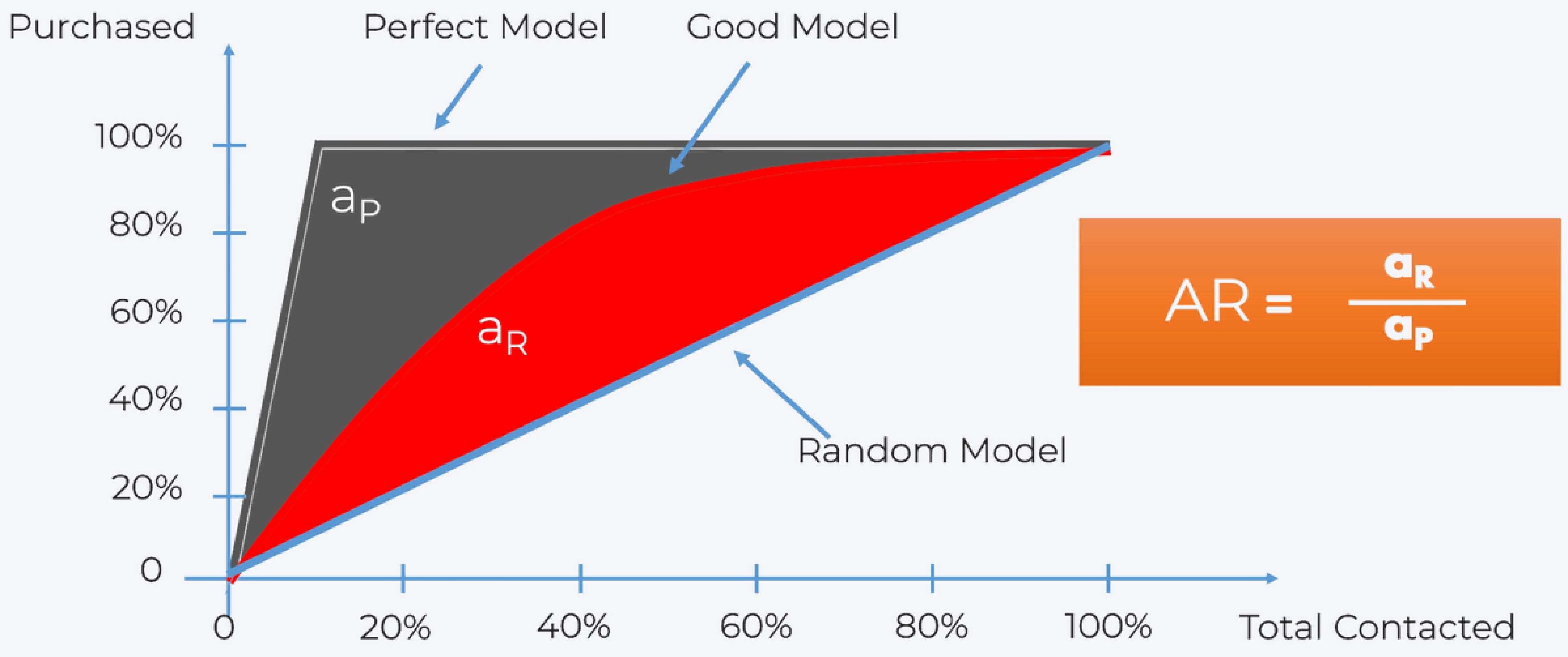
Tuwaiq Academy أكاديمية طويق

CAP Analysis - Model Evaluation

- **AR (Accuracy Ratio) Calculation:**
  - AR = aR / aP, where:
    - aR is the area under the model's curve.
    - aP is the area under the perfect model's curve.

A higher AR value means a better model.

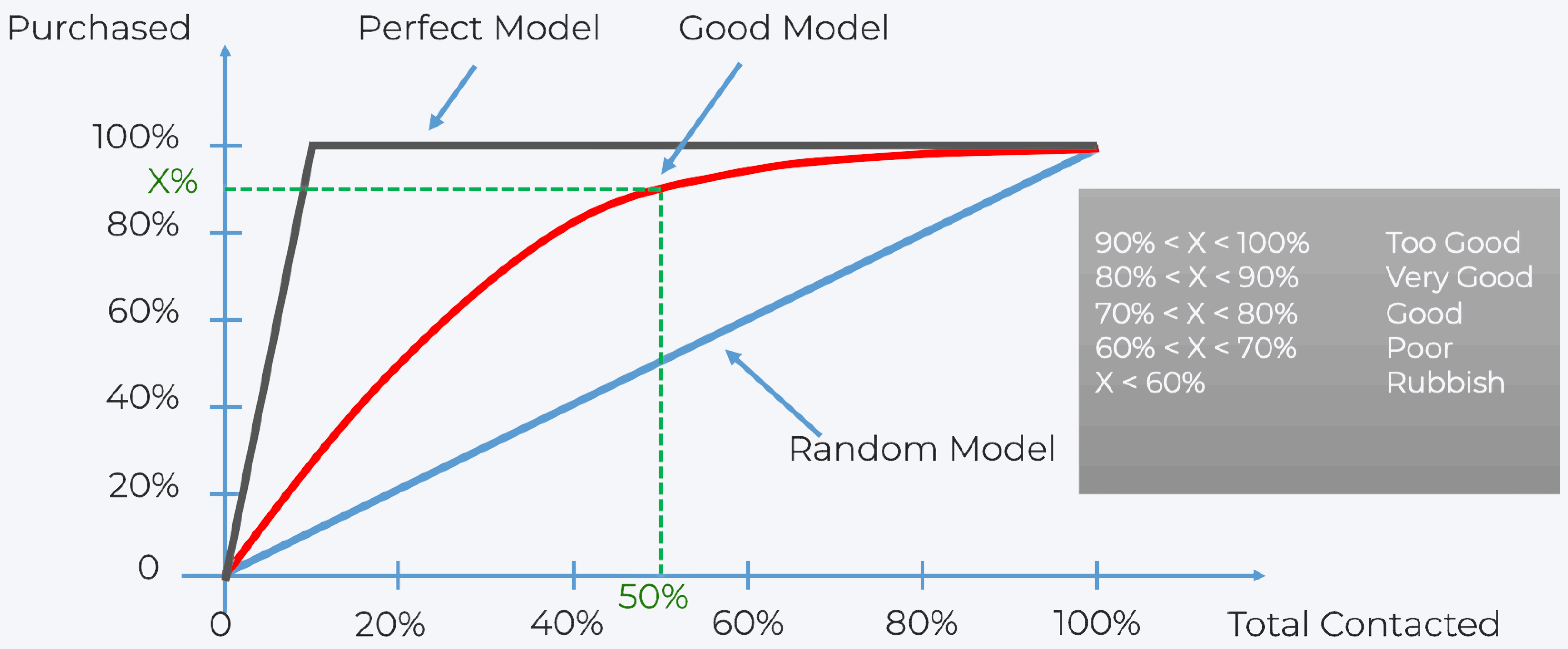
AP analysis provides a quantitative measure of model quality.



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Evaluating Classification Model

CAP Analysis - Model Evaluation



https://www.researchgate.net/publication/344389401_DS-Regression-04-SVR

Hands-On Code

Evaluating Classification Model