EAM2025

XI Conference

23RD - 25TH JULY 2025

Spain Tenerife
Canary Islands

European Association of Methodology

Decision Tree-Based Adaptive Testing in Psychodiagnostic Screening

Daiana Colledani & Pasquale Anselmi

Universidad de La Laguna
Cabildo de Tenerife
Tagoro
cajasiete
Gobierno de Canarias
Consejería de Universidades, Ciencia e Innovación y Cultura
Agencia Canaria de Investigación, Innovación y Sociedad de la Información
Instituto Canario de Igualdad
tea
hogrefe

Background & Motivation

- Traditional mental health assessments may be lengthy often including redundant items.

- This can lead to respondent fatigue, which reduces engagement and the quality of collected data, ultimately compromising diagnostic accuracy.

- Furthermore, this may prevent clinicians from screening all relevant constructs within the available time or resources.

- There is a growing need for more efficient and scalable screening solutions. Machine learning-based computerized adaptive testing (ML-CAT) offers a promising alternative by reducing assessment time while maintaining diagnostic accuracy.

Limitations of Traditional Methods

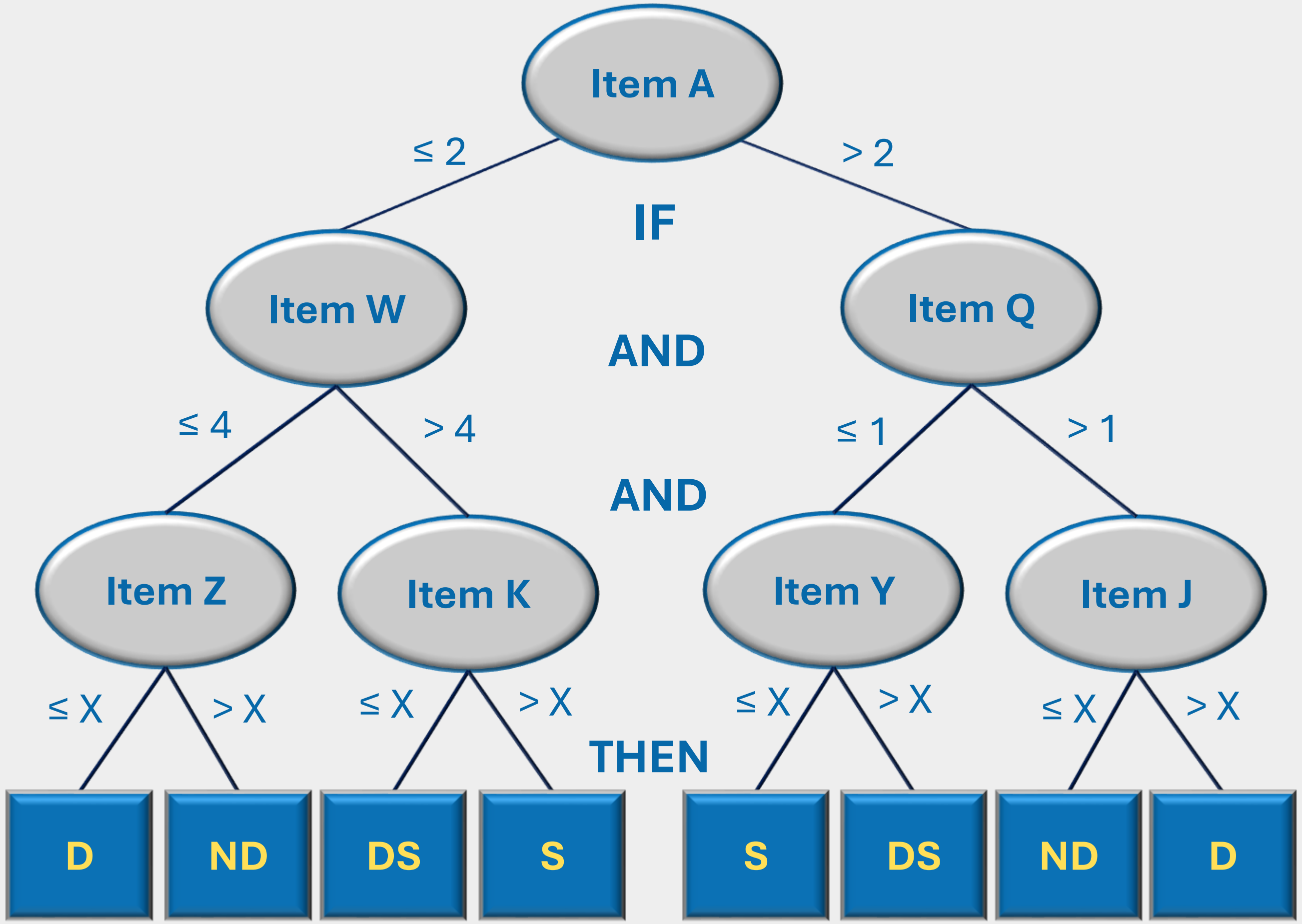
- Full-length questionnaires place a high burden on both patients and clinicians.
- Short scales are not always available and may lack sufficient construct validity.
- IRT-based adaptive tests are a valuable solution, but they require complex item calibration and rely on assumptions that are often violated in clinical data — such as unidimensionality and normal distribution. This limits their flexibility and increases development costs.

ML-Based CAT as an Alternative

- ML offers a valid alternative to traditional psychometric methods.
- In particular, Classification and Regression Trees (CART) algorithms have shown strong accuracy and efficiency in psychodiagnostic assessment.
- They can handle diverse types of data — including multidimensional and non-normal data — predict both scores and categorical diagnoses, and are suitable for tracking longitudinal changes in respondents' trait levels.
- Their decision tree (DT) structure enables interpretable and accurate classifications, making them particularly well-suited for CAT.

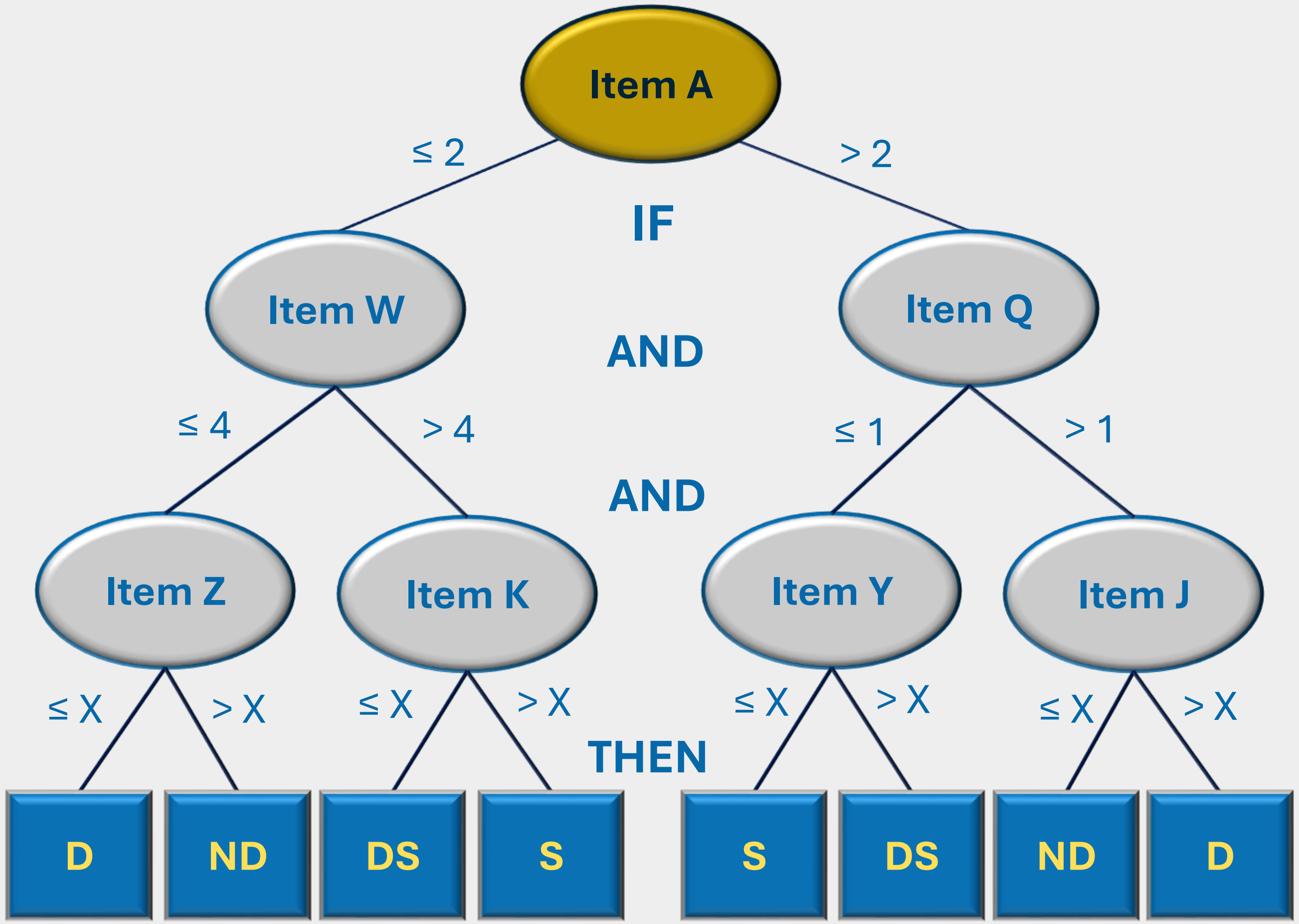
Decision Trees: Basic Structure

- DTs look like a flowchart, defining a set of “if-then” rules
- They use a set of continuous input variables (e.g., item responses) to predict a discrete outcome (e.g., diagnosis)
- They include a root-node (starting point), a series of internal nodes (items), branches (sequence of items connected by specific scoring rules), and leaves (classification)



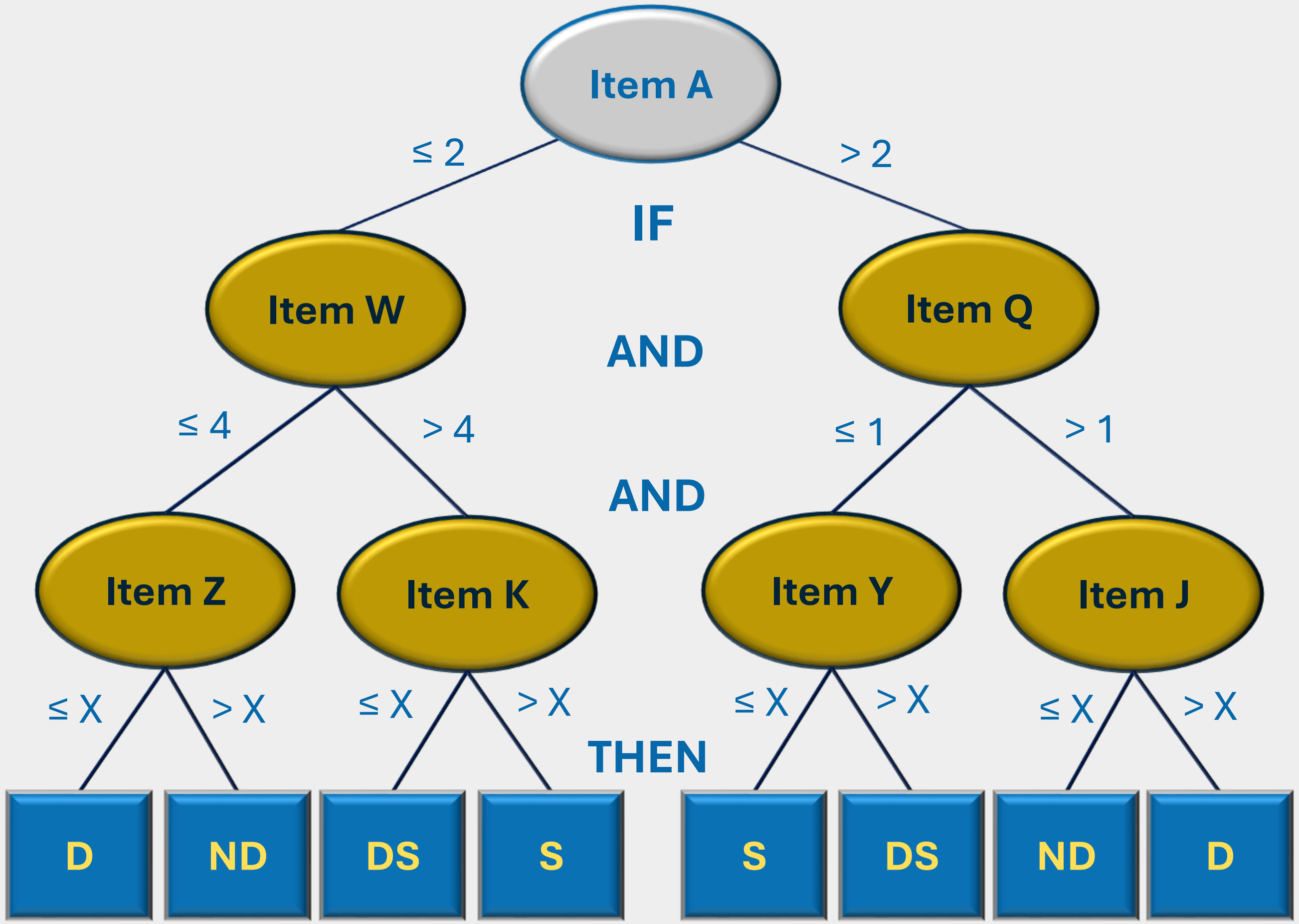
Decision Trees: Basic Structure

- DTs look like a flowchart, defining a set of "if-then" rules
- They use a set of continuous input variables (e.g., item responses) to predict a discrete outcome (e.g., diagnosis)
- They include a root-node (starting point), a series of internal nodes (items), branches (sequence of items connected by specific scoring rules), and leaves (classification)



Decision Trees: Basic Structure

- DTs look like a flowchart, defining a set of "if-then" rules
- They use a set of continuous input variables (e.g., item responses) to predict a discrete outcome (e.g., diagnosis)
- They include a root-node (starting point), a series of internal nodes (items), branches (sequence of items connected by specific scoring rules), and leaves (classification)

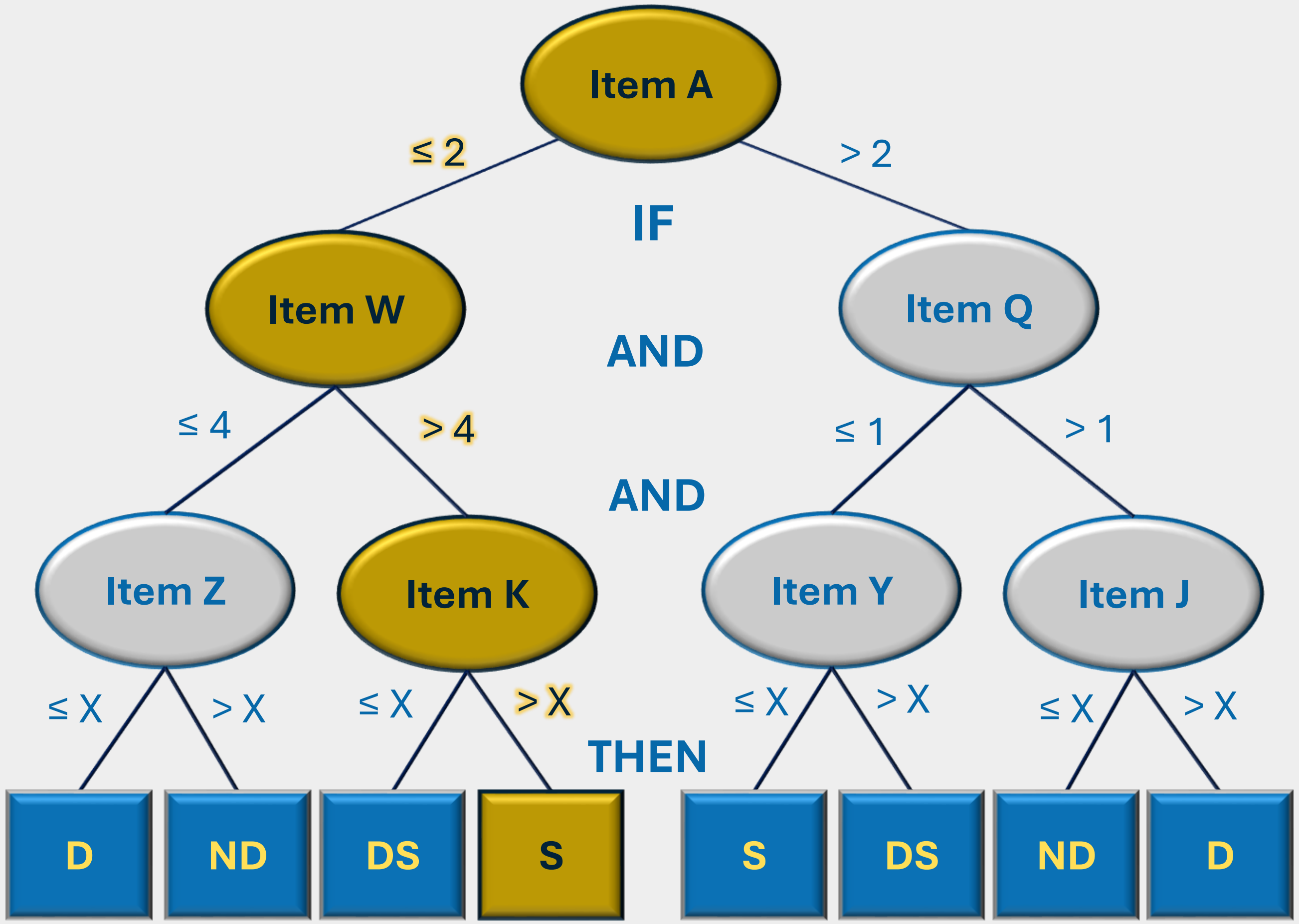


Decision Trees: Basic Structure

- DTs look like a flowchart, defining a set of "if-then" rules

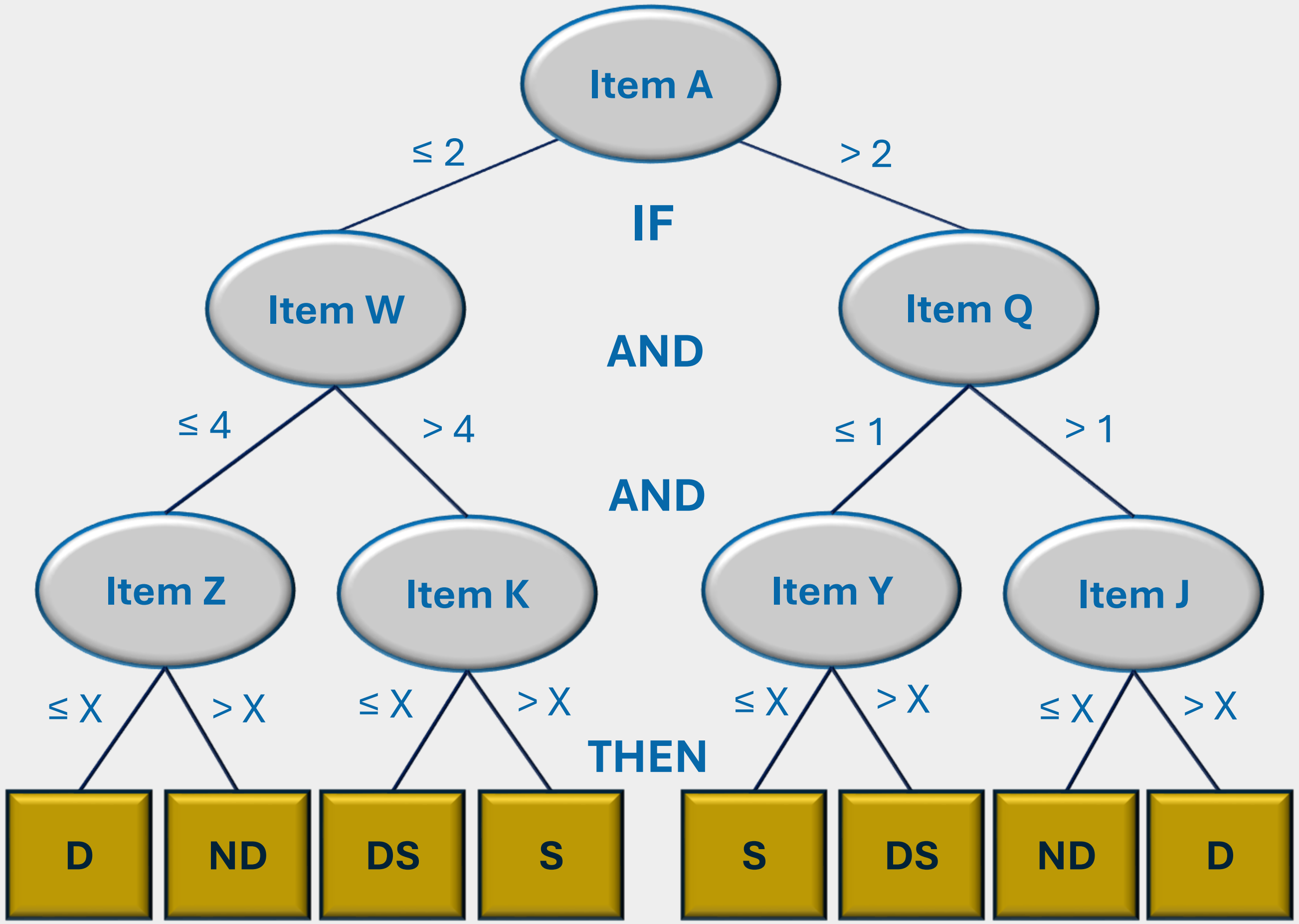
- They use a set of continuous input variables (e.g., item responses) to predict a discrete outcome (e.g., diagnosis)

- They include a root-node (starting point), a series of internal nodes (items), branches (sequence of items connected by specific scoring rules), and leaves (classification)



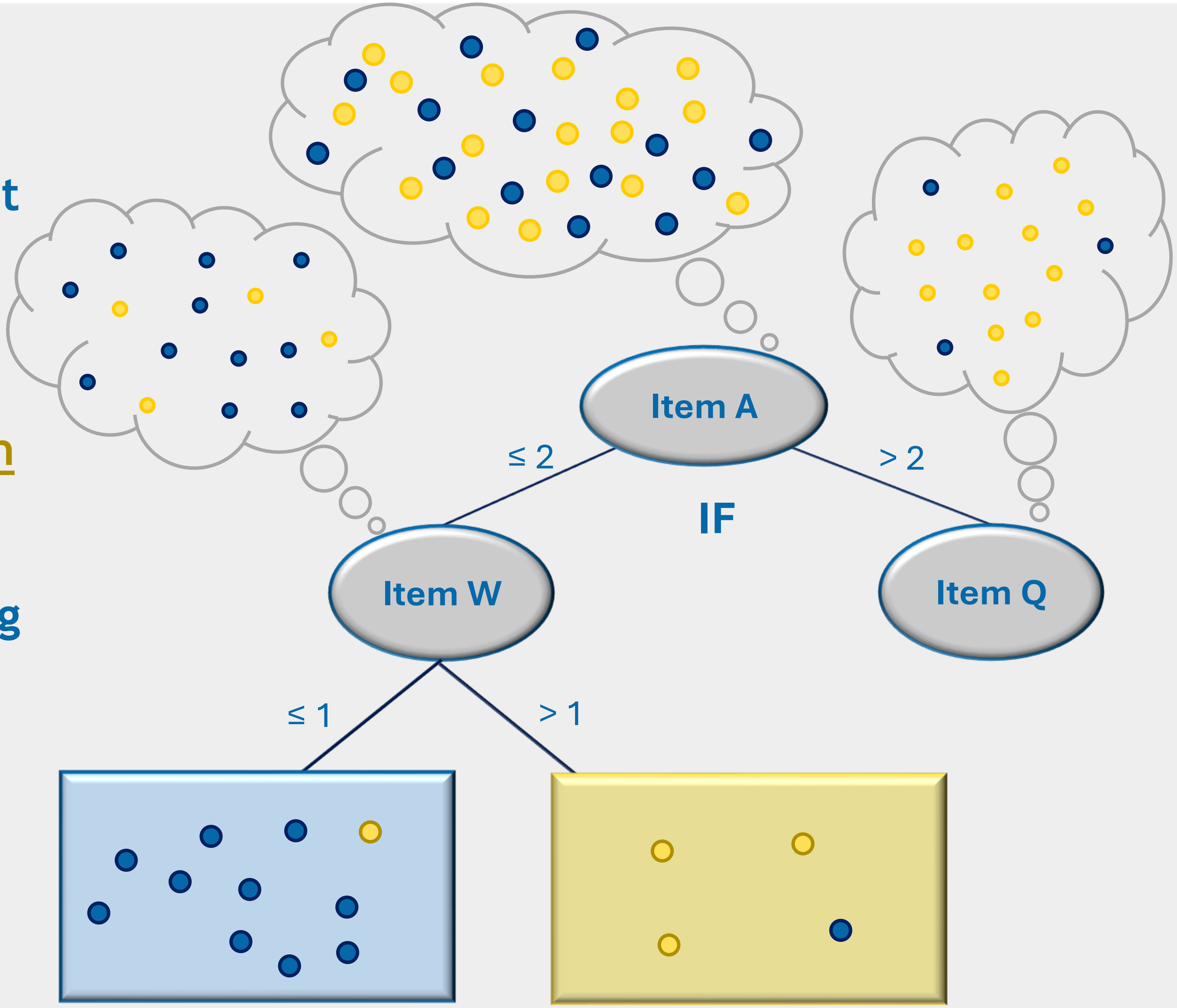
Decision Trees: Basic Structure

- DTs look like a flowchart, defining a set of “if-then” rules
- They use a set of continuous input variables (e.g., item responses) to predict a discrete outcome (e.g., diagnosis)
- They include a root-node (starting point), a series of internal nodes (items), branches (sequence of items connected by specific scoring rules), and leaves (classification)



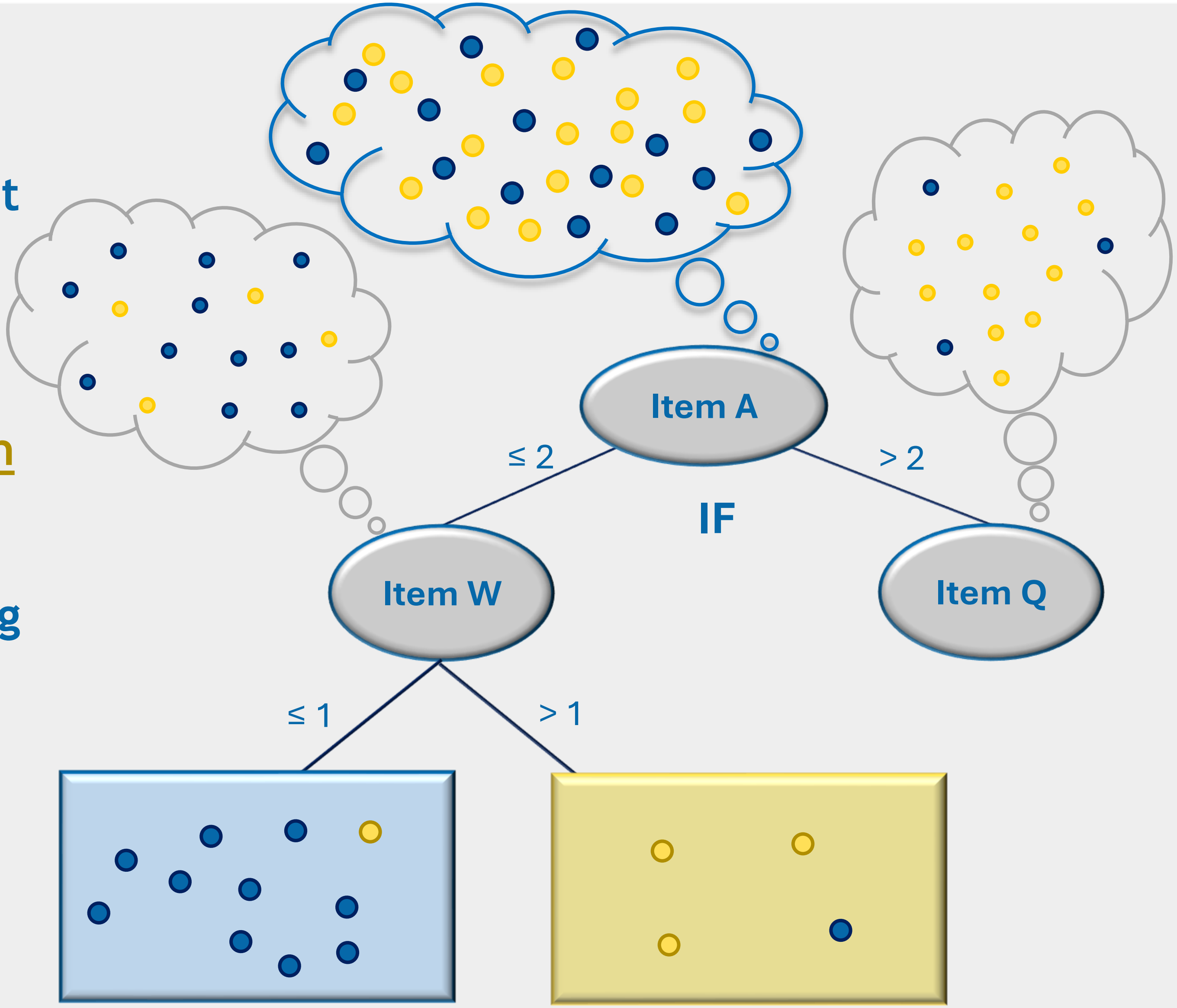
Decision Trees: Basic Structure

- The algorithm begins by selecting the root node, containing the full dataset.
- At each step, it chooses the attribute (item) and the splitting rules (scoring rule) that provide the highest information gain, that is, the variable that most effectively separates individuals by outcome, increasing entropy and reducing uncertainty
- The goal is to create the purest nodes: subgroups where cases are as homogeneous as possible with respect to the target classification (e.g., diagnosis).



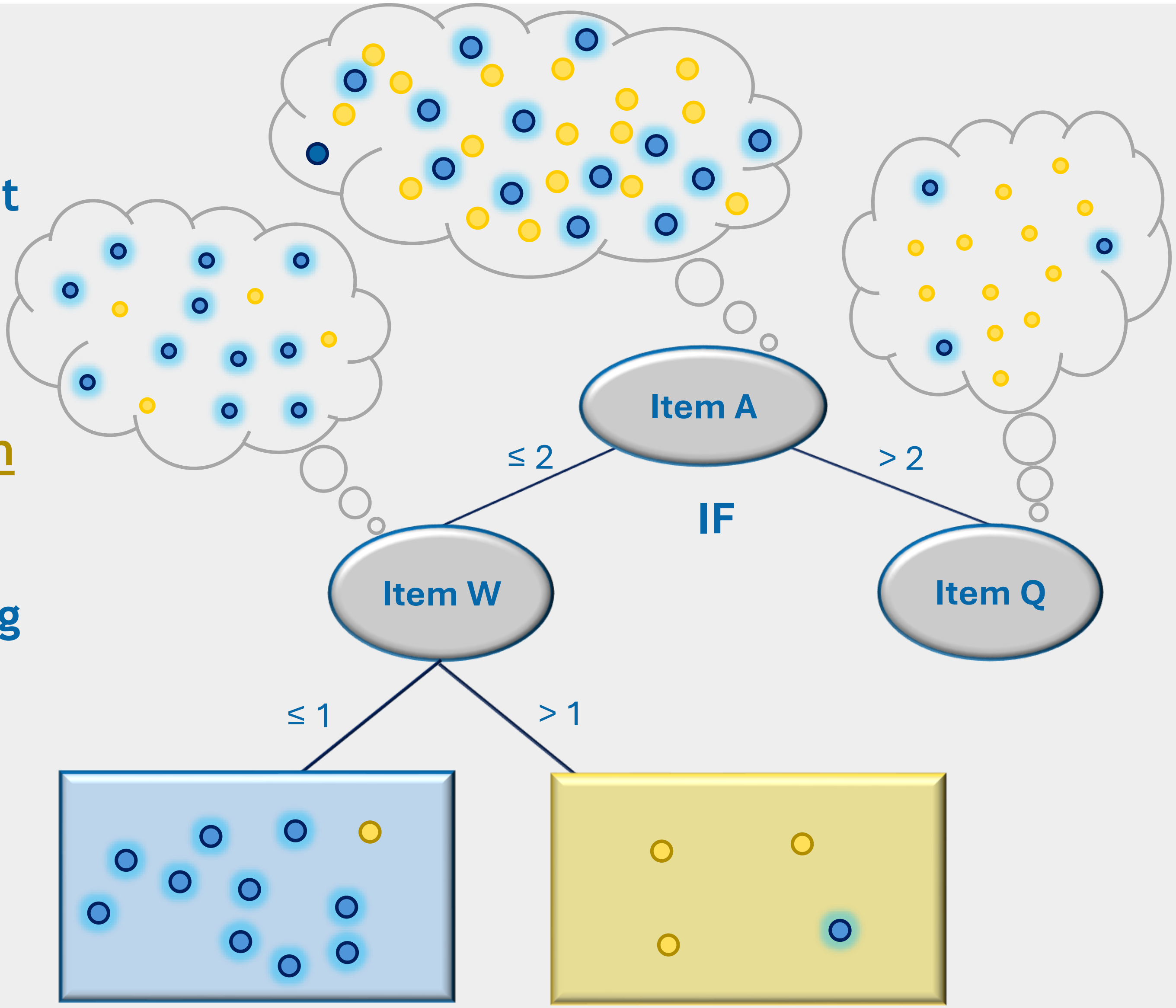
Decision Trees: Basic Structure

- The algorithm begins by selecting the root node, containing the full dataset.
- At each step, it chooses the attribute (item) and the splitting rules (scoring rule) that provide the highest information gain, that is, the variable that most effectively separates individuals by outcome, increasing entropy and reducing uncertainty
- The goal is to create the purest nodes: subgroups where cases are as homogeneous as possible with respect to the target classification (e.g., diagnosis).



Decision Trees: Basic Structure

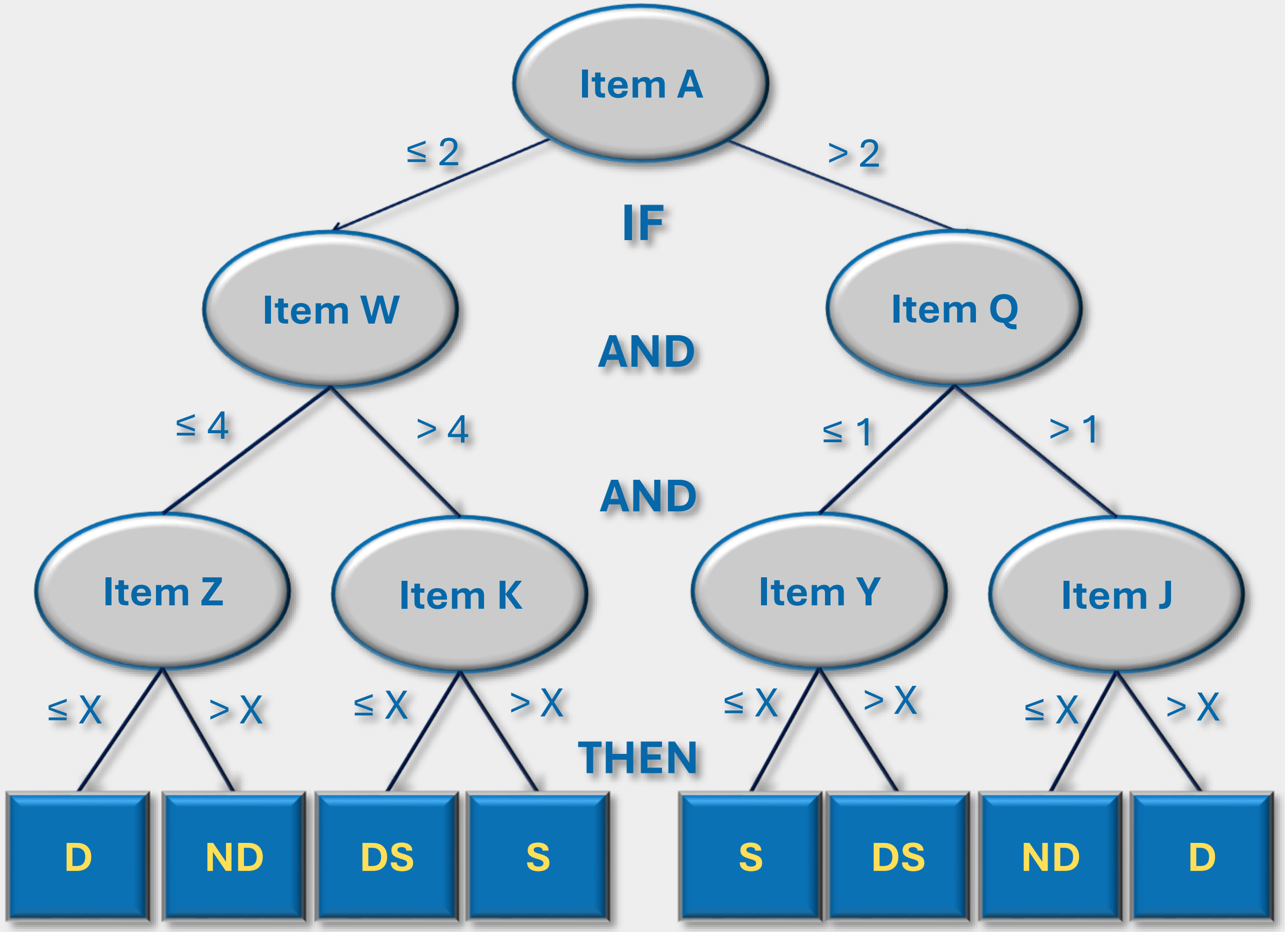
- The algorithm begins by selecting the root node, containing the full dataset.
- At each step, it chooses the attribute (item) and the splitting rules (scoring rule) that provide the highest information gain, that is, the variable that most effectively separates individuals by outcome, increasing entropy and reducing uncertainty
- The goal is to create the purest nodes: subgroups where cases are as homogeneous as possible with respect to the target classification (e.g., diagnosis).



Building a Decision Tree

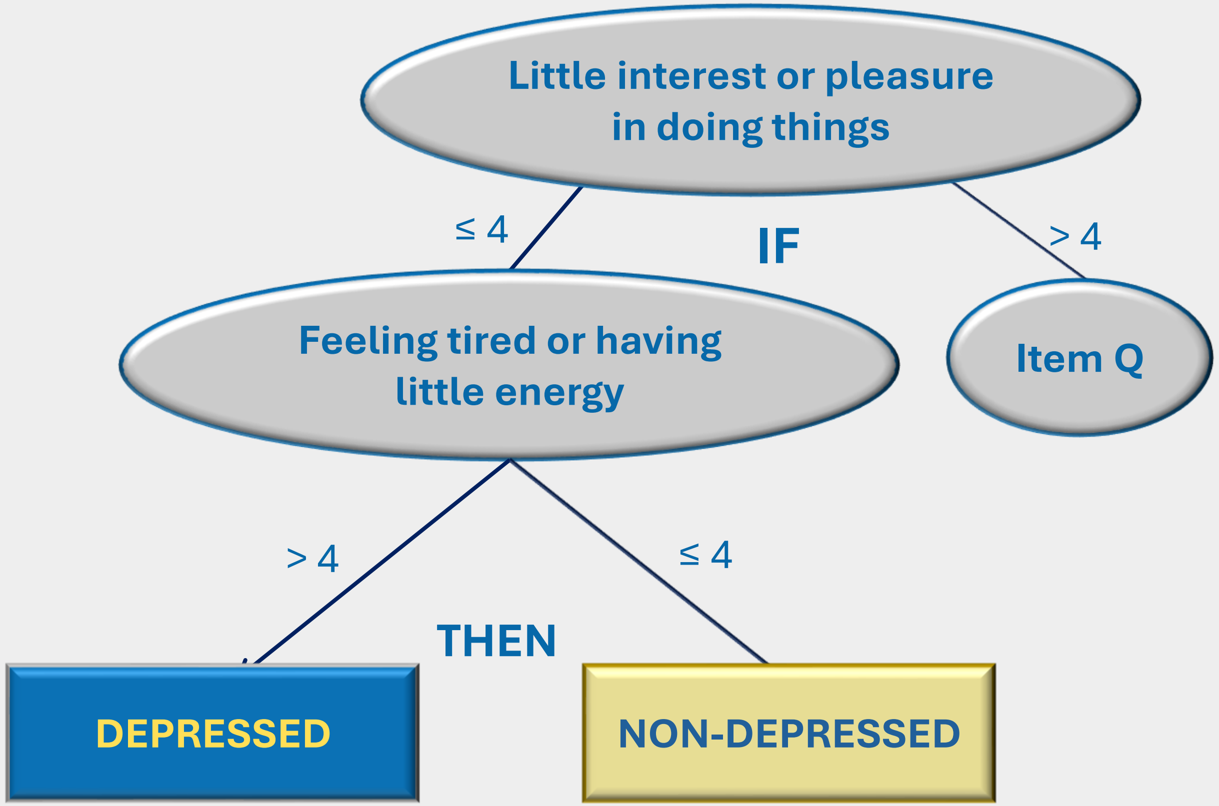
- DTs are built on a training dataset using a recursive "divide and conquer" process that creates branches until stopping criteria are met (e.g., max depth or no gain).

- Leaf nodes represent the final classification. For categorical variables, splits are based on discrete values; for numerical variables, the algorithm identifies optimal cut-points (e.g., Likert score ≤1 vs. >1).



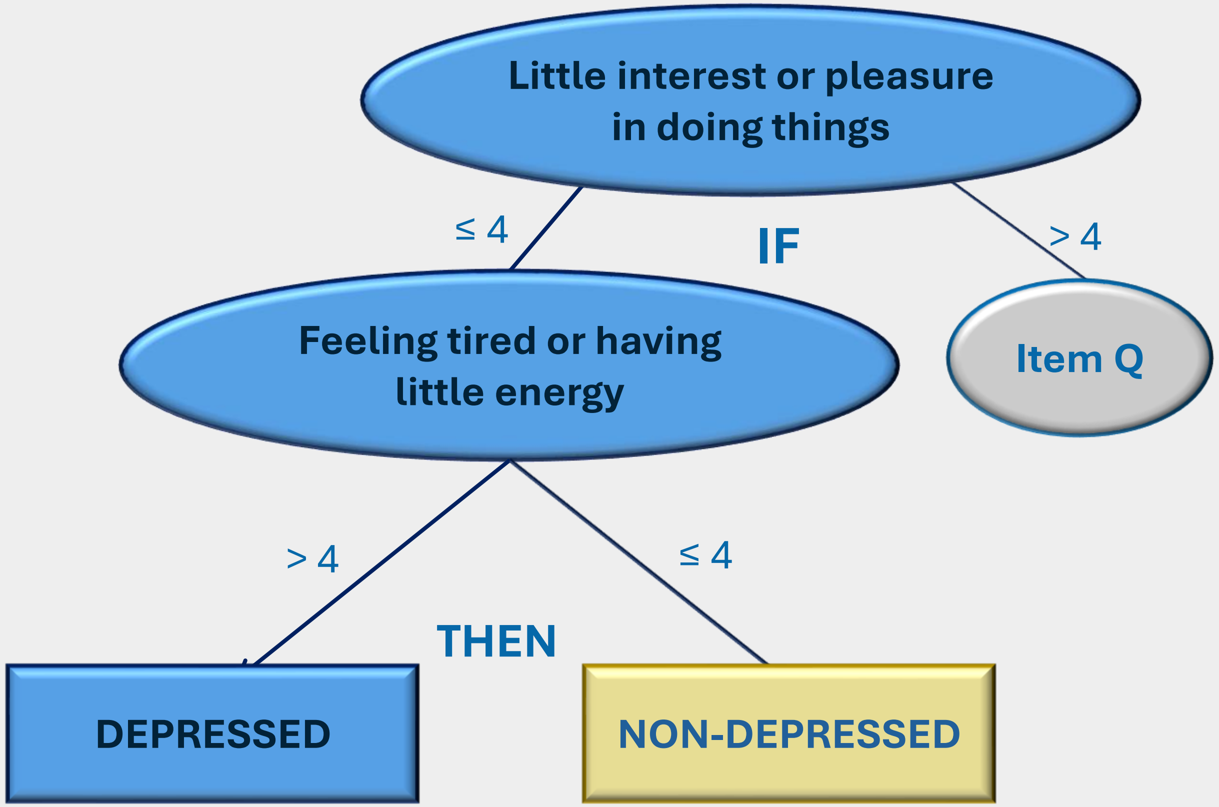
DTs in Psychodiagnostic Testing

- Once the structure of the Decision Tree is trained on a large dataset — using test items as input variables — it can be used to administer items adaptively to new respondents.
- New respondents are administered only the items along the relevant branch, rather than the full test.
- This adaptive process leads to a final diagnosis using a reduced item set, based on individual responses.
- The result: faster assessment with minimal loss in diagnostic precision.



DTs in Psychodiagnostic Testing

- Once the structure of the Decision Tree is trained on a large dataset — using test items as input variables — it can be used to administer items adaptively to new respondents.
- New respondents are administered only the items along the relevant branch, rather than the full test.
- This adaptive process leads to a final diagnosis using a reduced item set, based on individual responses.
- The result: faster assessment with minimal loss in diagnostic precision.



This work - Aim

- Usually, ML algorithms are trained to detect a single diagnostic condition
- In this context, they offer fast and accurate classifications.
- Given their success in single-condition assessment, it's plausible they could be extended to simultaneous screening of multiple disorders.
- A multi-disorder ML-CAT would

  ✓enable efficient detection of multiple diagnostic areas,
  ✓guide decisions for further clinical evaluation,
  ✓reduce the burden of multiple assessments, and
  ✓enhance the screening experience for both patients and clinicians.

This work - Method

- Sample

1,486 questionnaires investigating multiple mental health conditions on a sample of Spanish university students (68.3% female, Mage = 21.30, *SD* = 3.64)

- Measures

Social Anxiety Disorder Dimensional Scale (SAD-D) -> 10 item

Agoraphobia Dimensional Scale (AG-D) -> 10 item

Data were retrieved from a publicly available OSF repository (doi: 10.17605/OSF.IO/JZ4GE; Vidal-Arenas et al., 2023)

This work - Method

- Analysis

- A real-data simulation approach was used: The J48 algorithm was trained on part of the sample (*N*=1,040) and tested on unseen data (*N*=446).

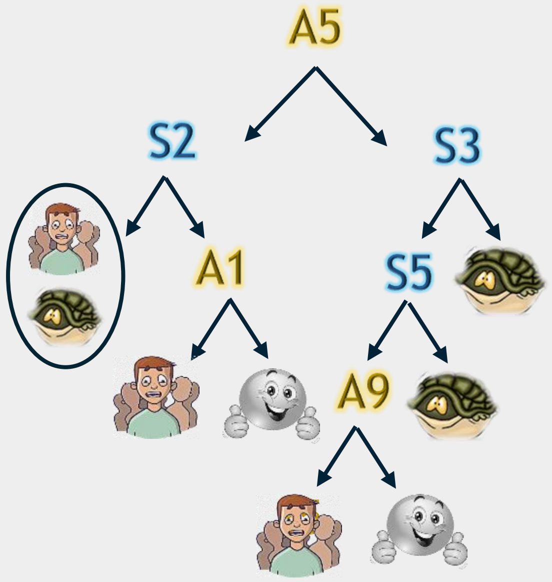
- The algorithm focused on 4 diagnostic groups:

Diagnostic Group	Training Dataset	Testing Dataset
Social anxiety (S)	55	24
Agoraphobia (A)	32	14
Social anxiety and Agoraphobia (SA)	42	18
No diagnosis (N)	911	390

This work - Method

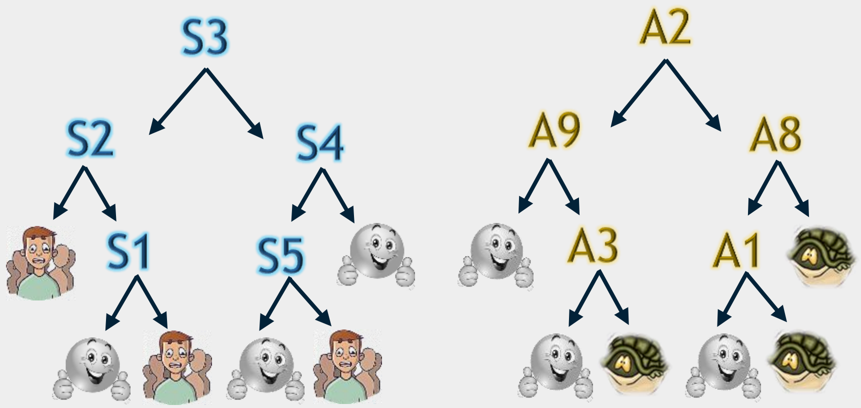
- <u>Two training conditions</u>

✓ TR4: model trained to classify all 4 diagnostic categories simultaneously (based on the 20 items of the SAD-D and AG-D)



A single DT based on 20 items, two mental health conditions (SAD-D and AG-D) and four diagnostic categories (social anxiety only, agoraphobia only, both, or neither)

✓ TR2: models trained separately for each disorder (presence/absence; based on the 10 items of the SAD-D or AG-D)



Two distinct DTs each based on a single mental health condition, 10 items, two diagnostic categories at a time (i.e., present absent)

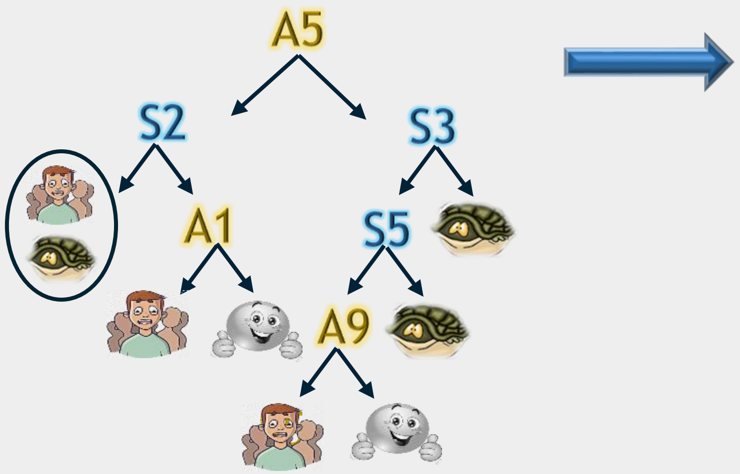
This work - Method

- Two testing conditions

✓ **TE4: test on 4 diagnostic categories**

✓ **TE2: test on 2 diagnostic categories (single disorder)**

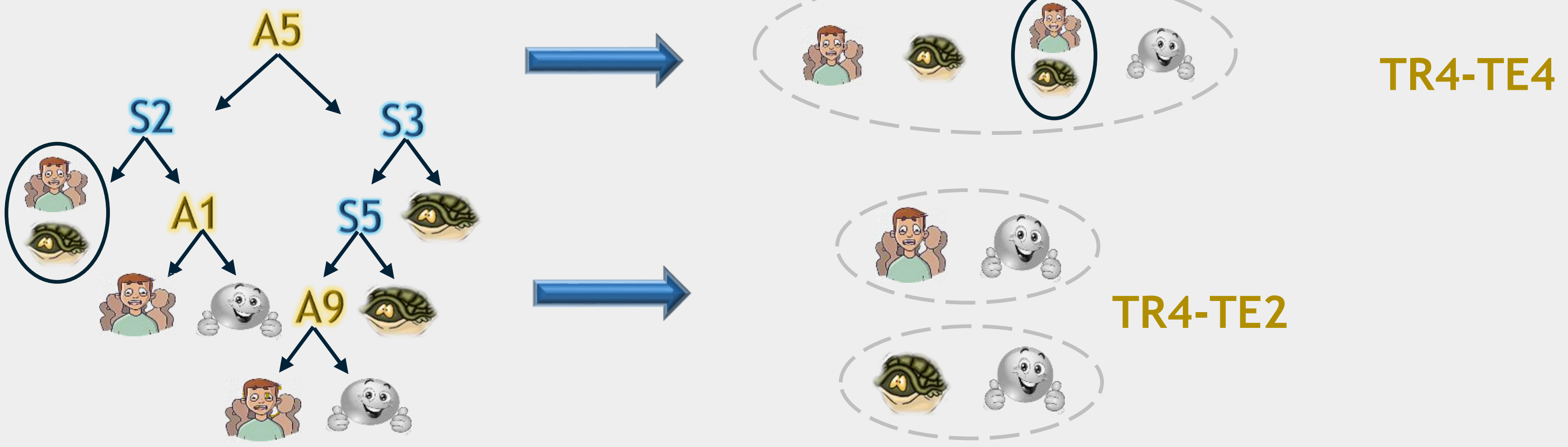
The two training conditions and two testing conditions lead to four settings:



TR4-TE4

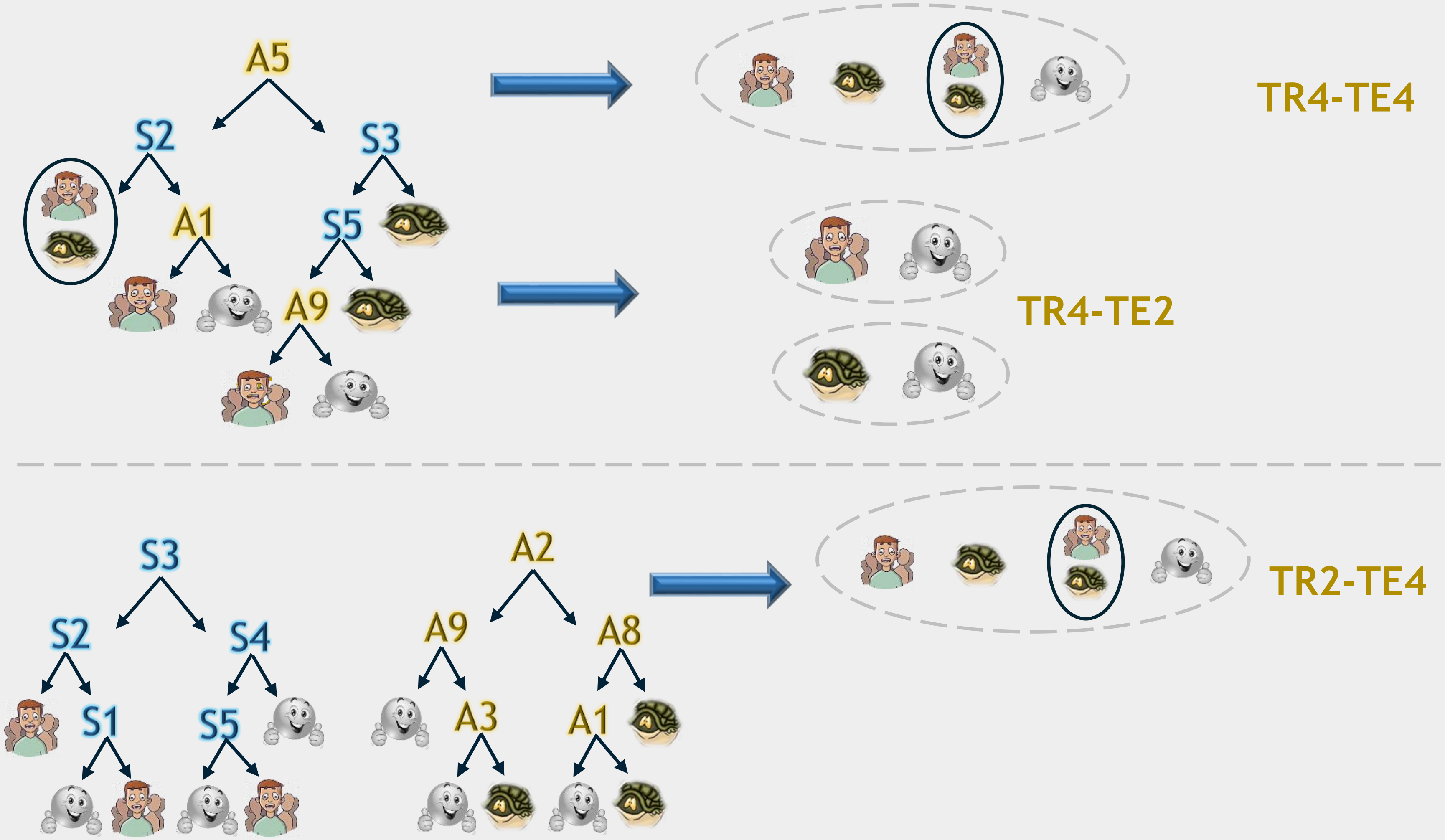
TRAINING	TESTING
TR4-TE4	
S, A, SA, N	S, A, SA, N
TR4-TE2	
S, A, SA, N	S/NON-S A/NON-A
TR2-TE4	
S/NON-S A/NON-A	S, A, SA, N
TR2-TE2	
S/NON-S A/NON-A	S/NON-S A/NON-A

The two training conditions and two testing conditions lead to four settings:



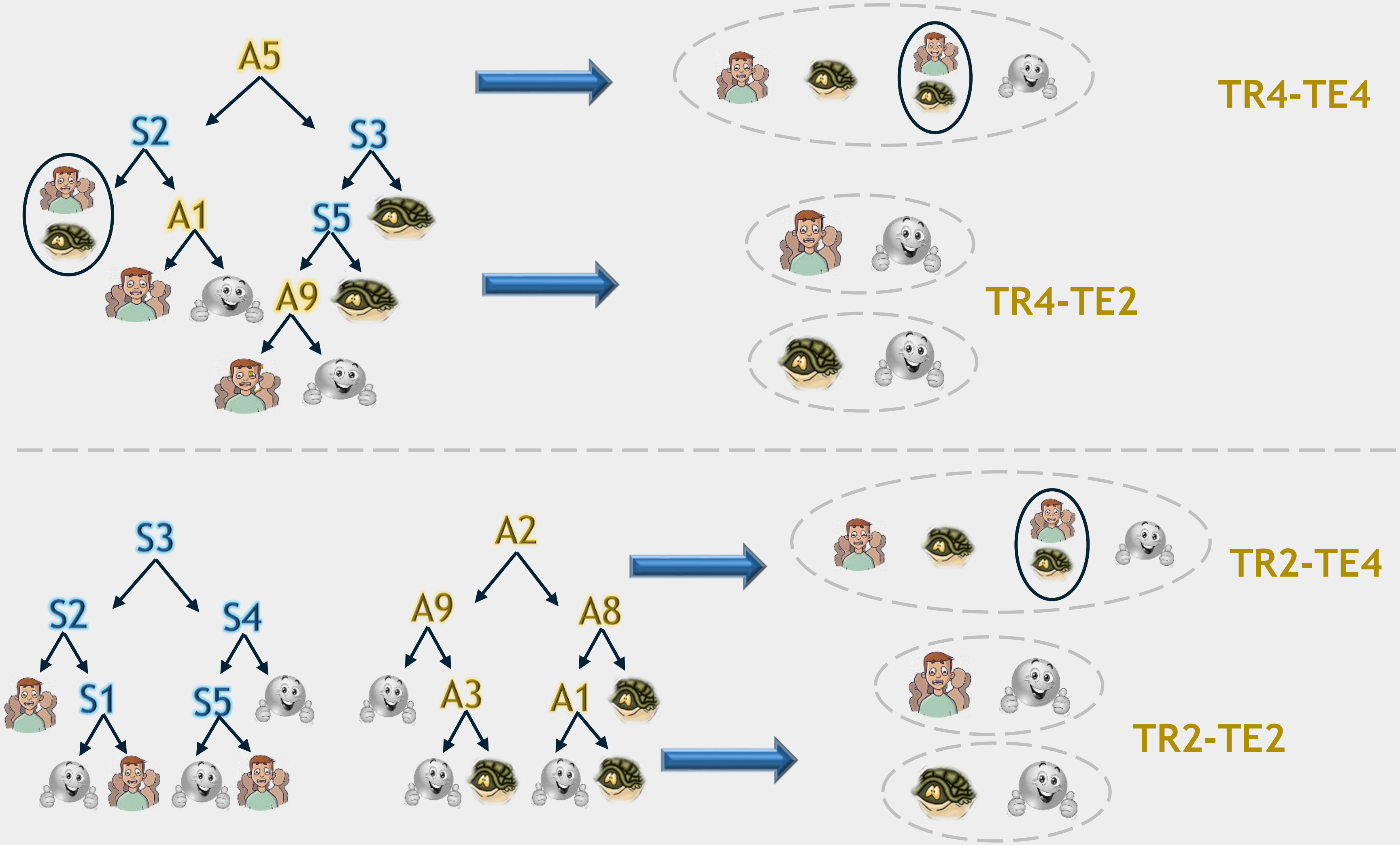
TRAINING	TESTING
TR4-TE4	
S, A, SA, N	S, A, SA, N
TR4-TE2	
S, A, SA, N	S/NON-S A/NON-A
TR2-TE4	
S/NON-S A/NON-A	S, A, SA, N
TR2-TE2	
S/NON-S A/NON-A	S/NON-S A/NON-A

The two training conditions and two testing conditions lead to four settings:



TRAINING	TESTING
TR4-TE4	
S, A, SA, N	S, A, SA, N
TR4-TE2	
S, A, SA, N	S/NON-S A/NON-A
TR2-TE4	
S/NON-S A/NON-A	S, A, SA, N
TR2-TE2	
S/NON-S A/NON-A	S/NON-S A/NON-A

The two training conditions and two testing conditions lead to <u>four settings</u>:

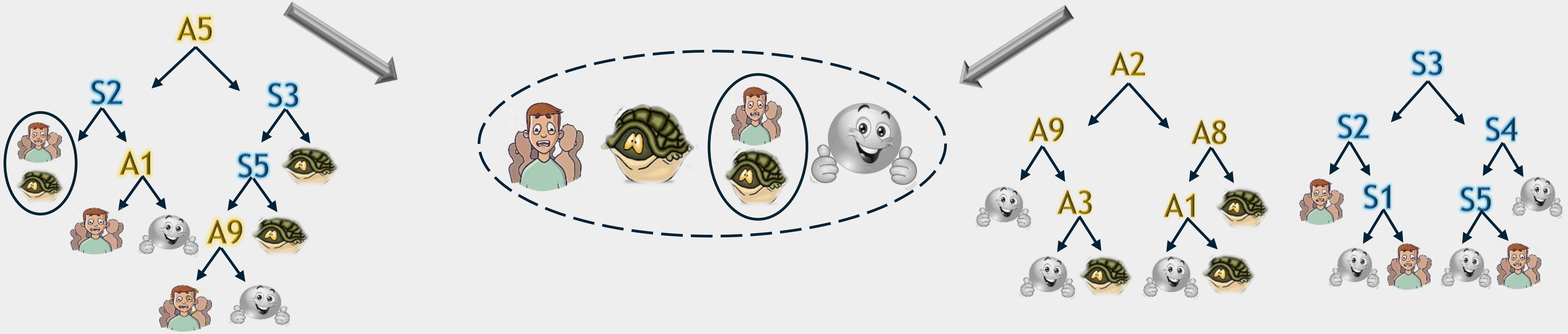


TRAINING	TESTING
TR4-TE4	
S, A, SA, N	S, A, SA, N
TR4-TE2	
S, A, SA, N	S/NON-S A/NON-A
TR2-TE4	
S/NON-S A/NON-A	S, A, SA, N
TR2-TE2	
S/NON-S A/NON-A	S/NON-S A/NON-A

Results on the testing dataset

TR4-TE4

TR2-TE4



Category	Sensitivity	Specificity
SA	.83	.99
A	.43	.99
S	.71	.97
N	.96	.80

Accuracy = .93

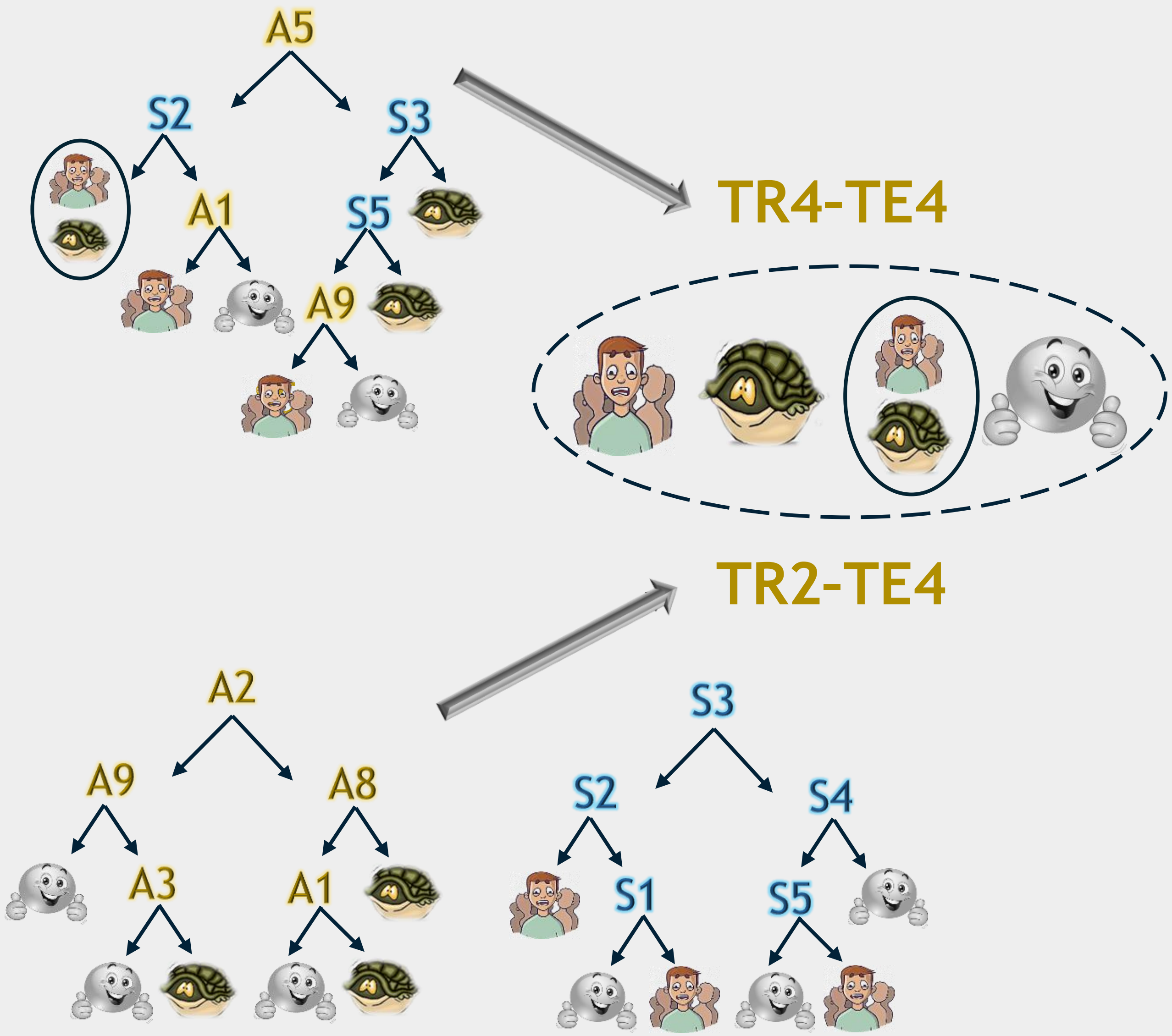
4.44 items required by the ML-CAT to complete the assessment (77.8% item reduction)

Category	Sensitivity	Specificity
SA	.78	1.00
A	.86	.96
S	.71	.99
N	.96	.88

Accuracy = .93

6.33 items required by the ML-CAT to complete the assessment (68.4% item reduction)

Comparing Conditions TR4-TE4 and TR2-TE4

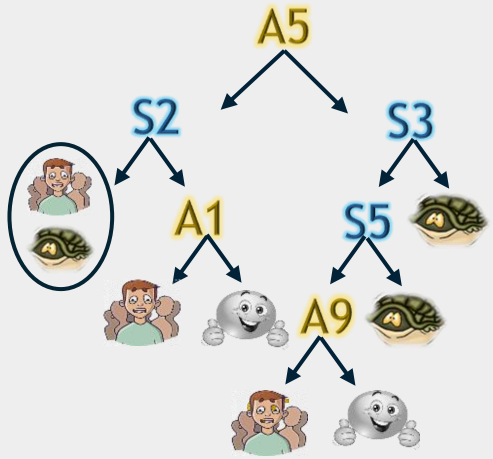


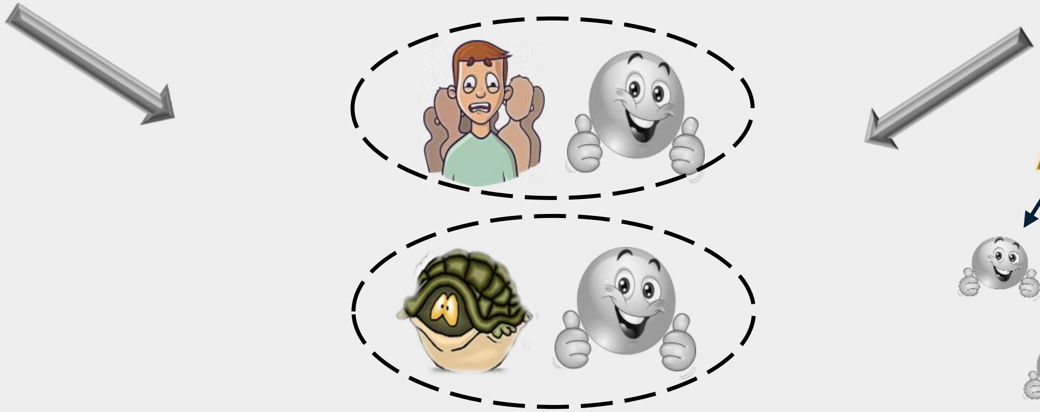
- The McNemar's test showed no significant difference in performance between conditions TR4-TE4 and TR2-TE4 ($\chi^2(1) = 0.24$, $p = .62$)

- However, the condition TR2-TE4 required administering significantly more item than the TR4-TE4 condition ($t(445) = 38.5$, $p < .001$, $d = 1.83$), yet significantly fewer than the full-length test administration ($t(445) = -387$, $p < .001$, $d = -18.3$)

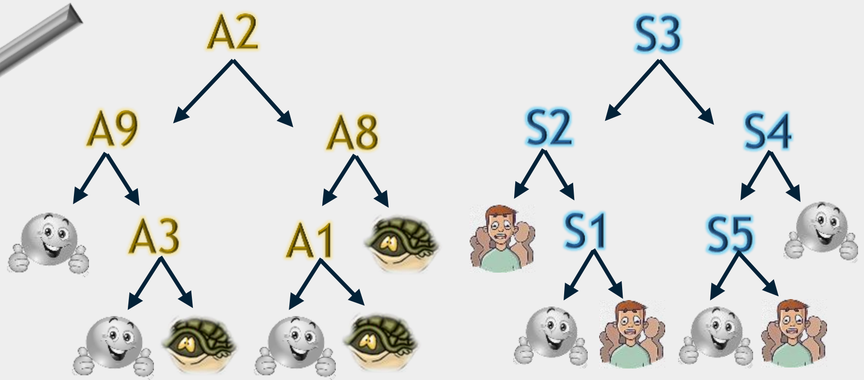
Results on the testing dataset

TR4-TE2





TR2-TE2



Category	Sensitivity	Specificity
A	.94	.97

Category	Sensitivity	Specificity
S	.76	.99

Accuracy = .96 for both agoraphobia and social anxiety.

The ML-CAT required on average 4.44 items to complete the assessment.

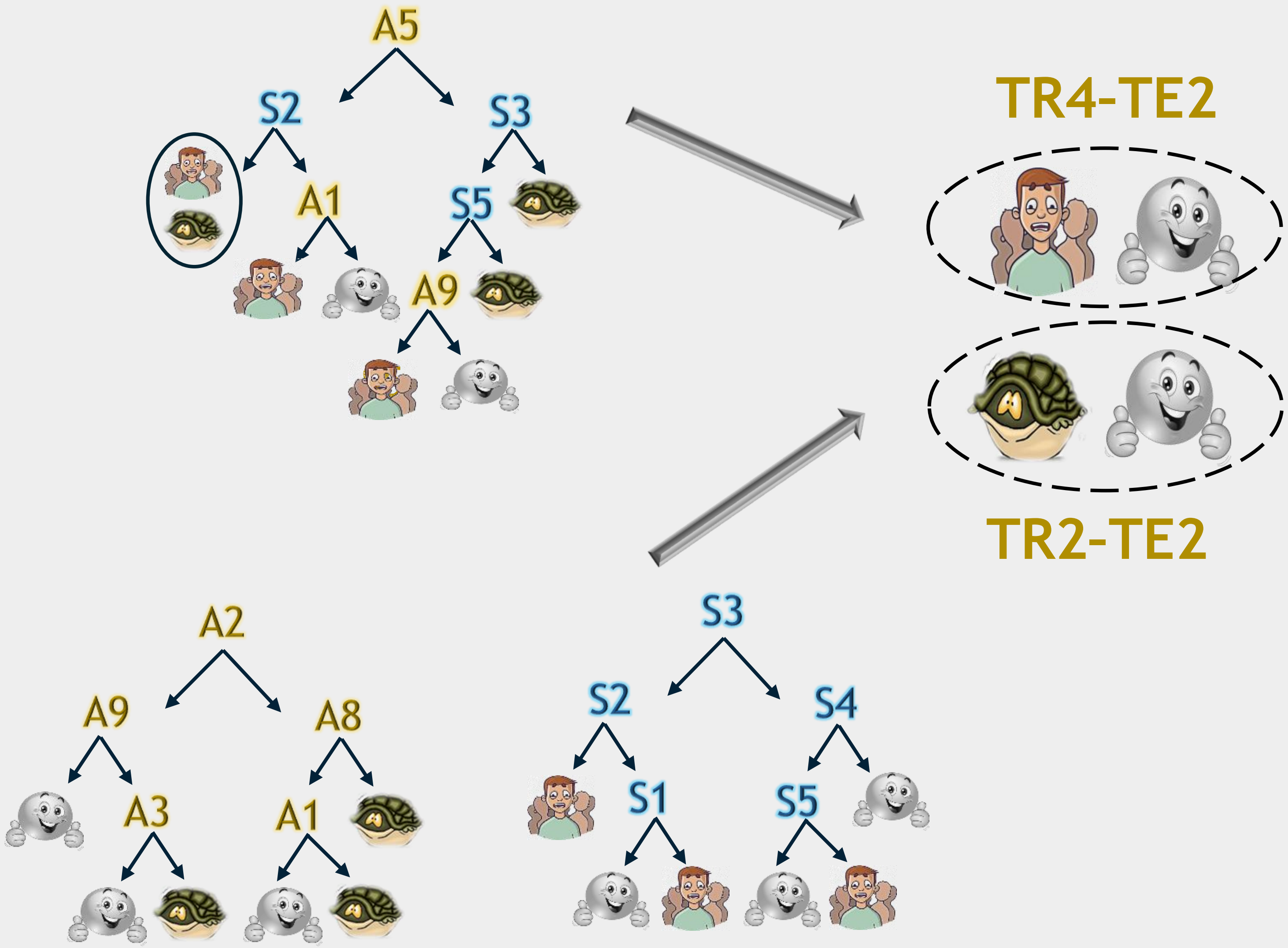
Category	Sensitivity	Specificity
A	.94	.97

Category	Sensitivity	Specificity
S	.76	.99

Accuracy = .96 and .97 for social anxiety and agoraphobia, respectively.

The ML-CAT required on average 3.17 items to complete the assessment of agoraphobia and 3.16 for social anxiety.

Comparing Conditions TR4-TE2 and TR2-TE2



For both agoraphobia and social anxiety, the McNemar tests indicated that the differences in performance between the two training methods were not statistically significant (for agoraphobia, $\chi^2(1) = 0.39$, $p = .53$; for social anxiety, $\chi^2(1) = 0.20$, $p = .66$).

Conclusion

ML-based CAT using decision trees shows strong potential for adaptive psychodiagnostic screening.

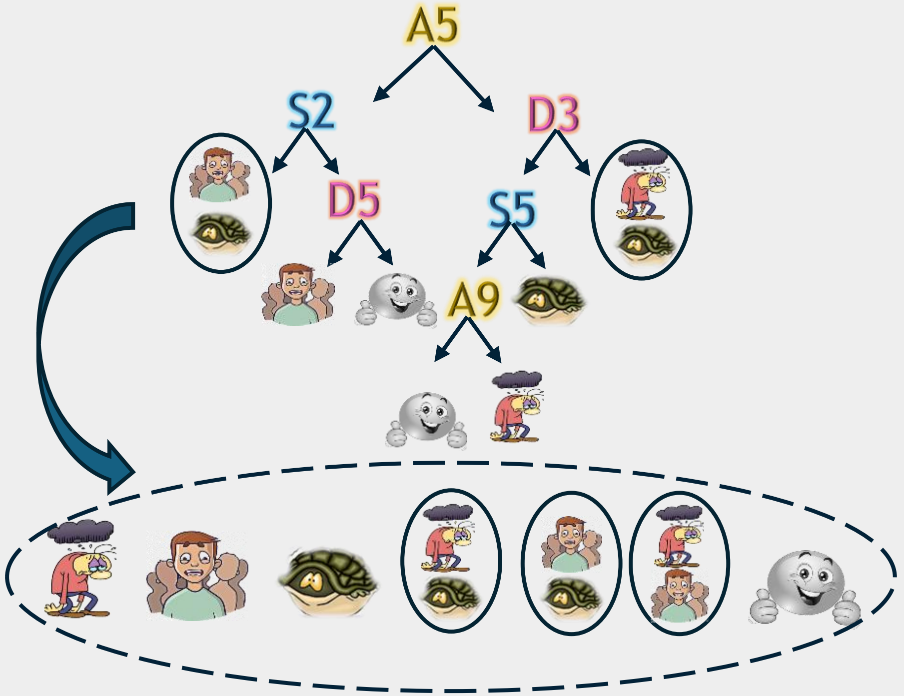
- ✓ It achieves high diagnostic accuracy with a significant reduction in items administered (up to 77.8%).
- ✓ Simultaneous classification (TR4) is more efficient than single-disorder approaches (TR2), while maintaining comparable accuracy.
- ✓ However, the best approach depends on context.

Conclusion

- TR4 (multi-disorder training)

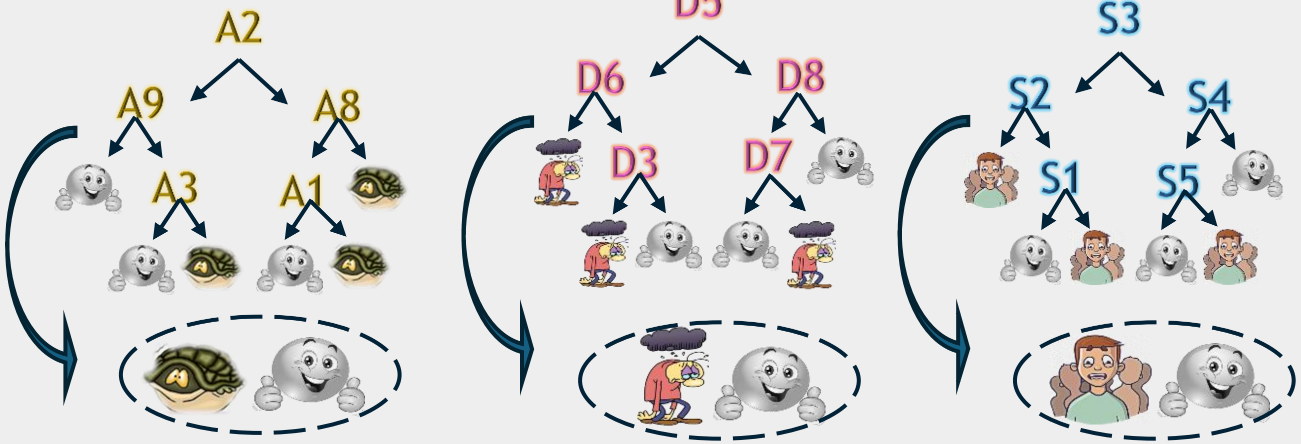
✓ Is ideal for focused screening of a limited number of disorders, offering high efficiency.

✓ However, it might becomes computationally complex as the number of disorders increases (exponential growth in diagnostic combinations).



- TR2 (single-disorder training)

✓ Is more scalable and flexible, and better suited to settings where many disorders are assessed



Thank You!