

# Predictive Modeling and Explainable AI for Autism Spectrum Disorder Diagnosis

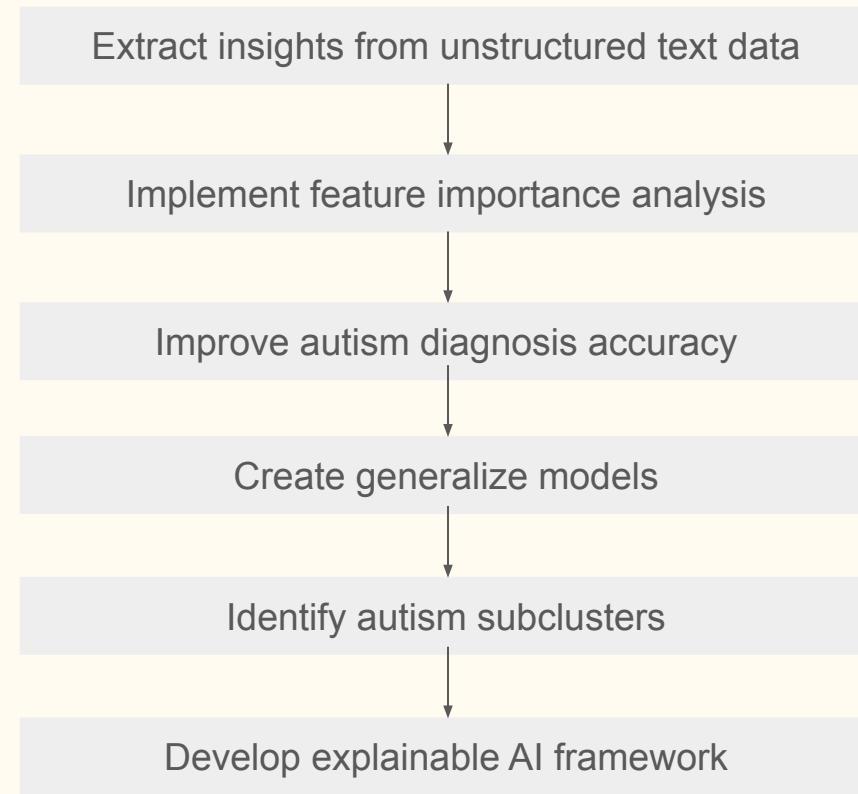
---

Advisor: Dr. Amir Jafari and Dr. Gabriela Rosenblau  
Students: Anirudh Rao, Ramana Bhaskar Kosuru, Wilona Nguyen

# Problem Statement

How can interpretable ML systems that jointly model behavioral performance and linguistic expression capture the multidimensional heterogeneity of ASD, enabling more precise diagnosis, stable subtype discovery, and improved generalizability across diverse populations?

# Objective



# PIPELINE AND DATA

# Overall Project Pipeline

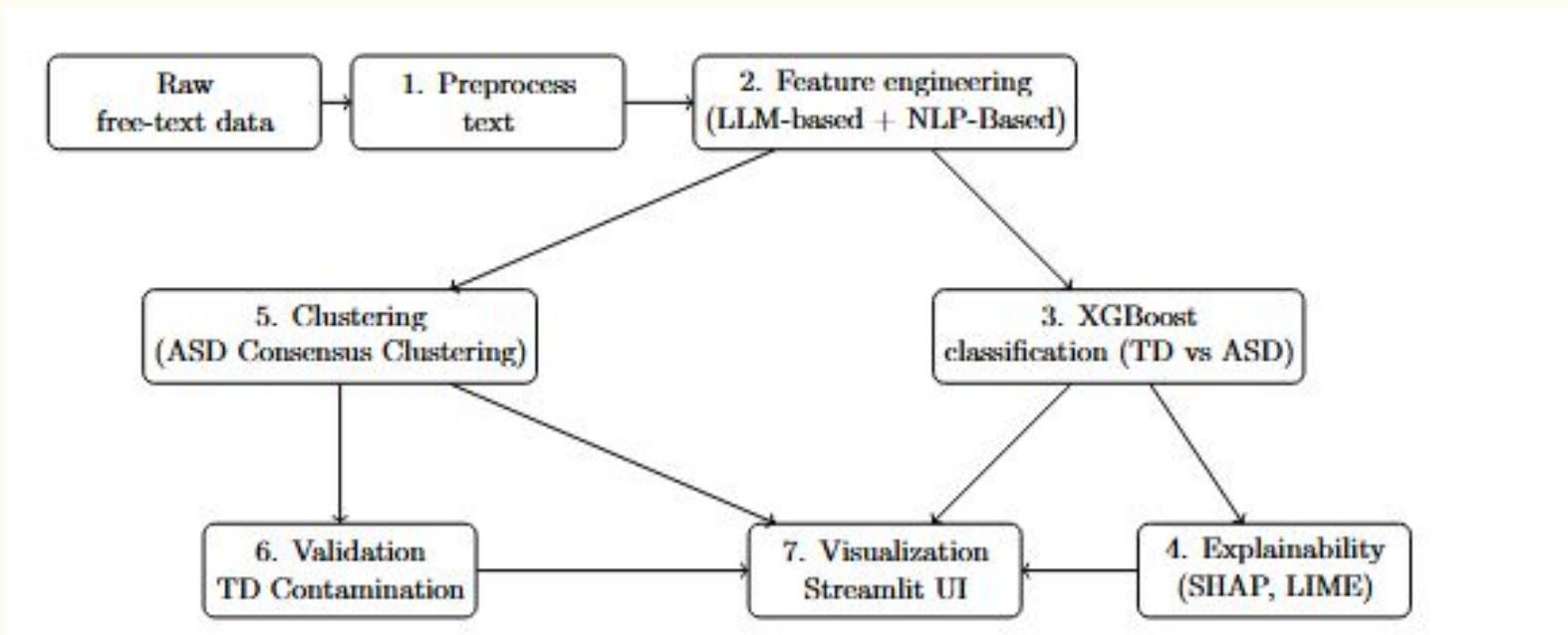


Table 1: Data Dictionary for All Datasets

Data Version	Column Name	Description
V1, V2, V3	sub	Participant ID
V1, V2	subject	Participant ID followed by a random code for the current profile
V1, V2, V3	td_or_asd	Diagnostic group (0 = TD, 1 = ASD)
V1	asd_diagnosis_text	Which ASD diagnosis does the participant have?
V1	trial	The trial number (1-60) within the current profile.
V1, V2	profile	Peer profile (TDprof_norm, ASDprof_norm, ASDprof_unif)
V1	image	The image shown on the current trial. Order is randomized within profile for each participant.
V1	cat	Image category (1 = activities, 3 = foods)
V1	subcat	Image subcategory (1 = arts & crafts, 2 = music & instruments, 3 = sports, 4 = games & gadgets, 9 = fast food, 10 = fruits & vegetables, 11 = healthy & savory, 12 = desserts)
V1	concept	Descriptive image concept label
V1	cnum	Image concept number
V1	selfpref	How much does the participant like this image?
V1	asd_meanpref	On average, how much do ASD participants like this image?
V1	td_meanpref	On average, how much do TD participants like this image?
V1	slider_rating	The participant's rating of how much they think the peer in question like the image shown.
V1	profile_rating	The correct rating for the given image and profile (this is the feedback given to participants)
V1	PE	The absolute difference between a participant's rating of how much they think the peer likes an item and the peer's actual liking of the item (actual liking = feedback).
V1, V2	avg_PE or profile_avg_PE	Average prediction error on each peer profile
V1	free_response	Participant's free response - what they think about the peer they just learned about.
V2, V3	SRS.Raw	Raw SRS-2 social responsiveness score (higher = more difficulties)
V2, V3	FSR	Flexibility Scale-Revised score (higher = less flexibility)
V2, V3	BIS	Behavioral Inflexibility Scale score (higher = less flexibility)
V1, V2	free_response	Participant's free-text description of the peer
V2	LPA_Profile_grand_mean	Latent profile (grand-mean centered)
V2	LPA_Profile_ASD_only	Latent profile (ASD-only analysis)
V3	TDNorm_avg_PE	Across all trials, average PE for TDNorm profile
V3	overall_avg_PE	Across all trials, average PE for all three profiles
V3	TDNorm_concept_learning	Across all trials, slope of PE for TDNorm profile
V3	overall_concept_learning	Across all trials, slope of PE across all three profiles
V3	free_response_TDprof_norm	Free response text from TDprof_norm profile
V3	free_response_ASDprof_norm	Free response text from ASDprof_norm profile
V3	free_response_ASDprof_unif	Free response text from ASDprof_unif profile

# Data Dictionary

V1 - Trial Level Information (187,187)

V2 - Subject/Profile Level (2,648)

V3 - Subject Level with Concept Learning (1,119)

**TARGET : *td\_or\_asd***

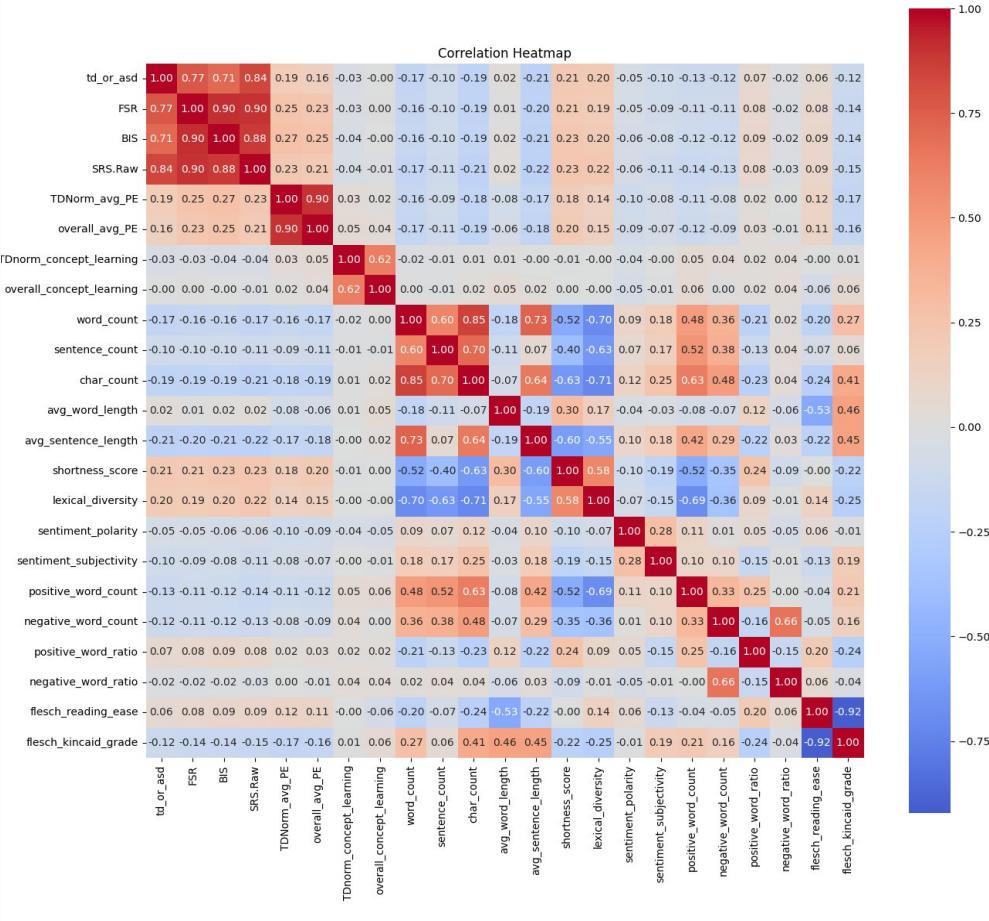
TD : Typically Developing

ASD: Autism Spectrum Disorder

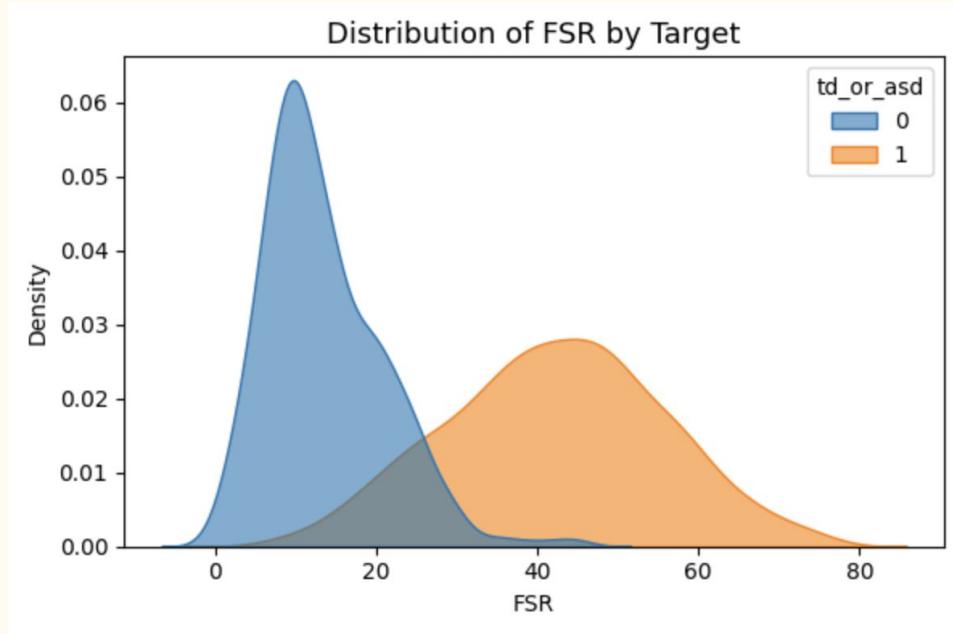
# EDA

# Correlation Analysis

SRS.Raw is considered to be a Target Leakage variable due to its extremely high corr with Target



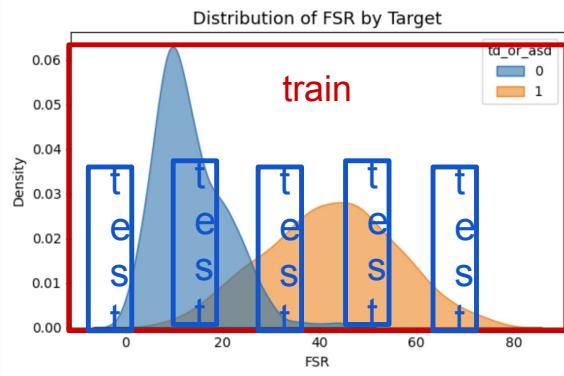
# KDE Distribution of FSR



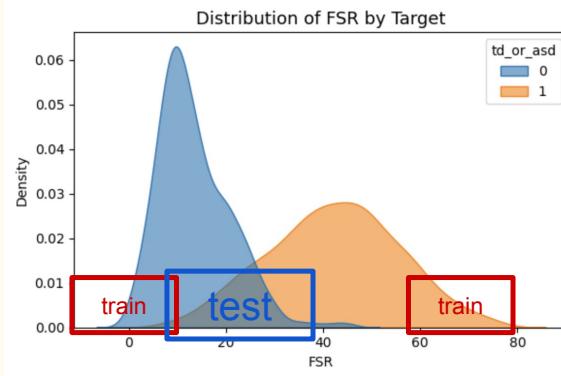
## FSR: Flexibility Scale Revised

- Higher scores indicate lower flexibility/ greater rigidity

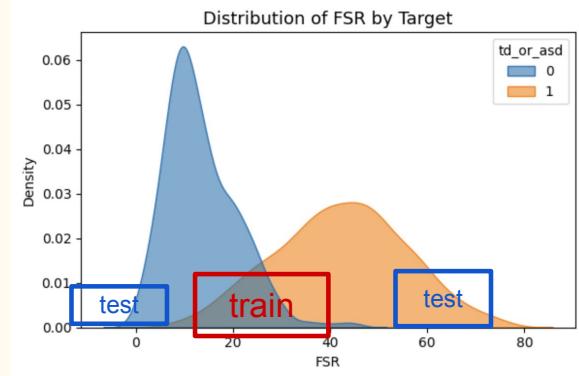
# Train - Test Split Formulation



(1) Stratified 80/20 train-test



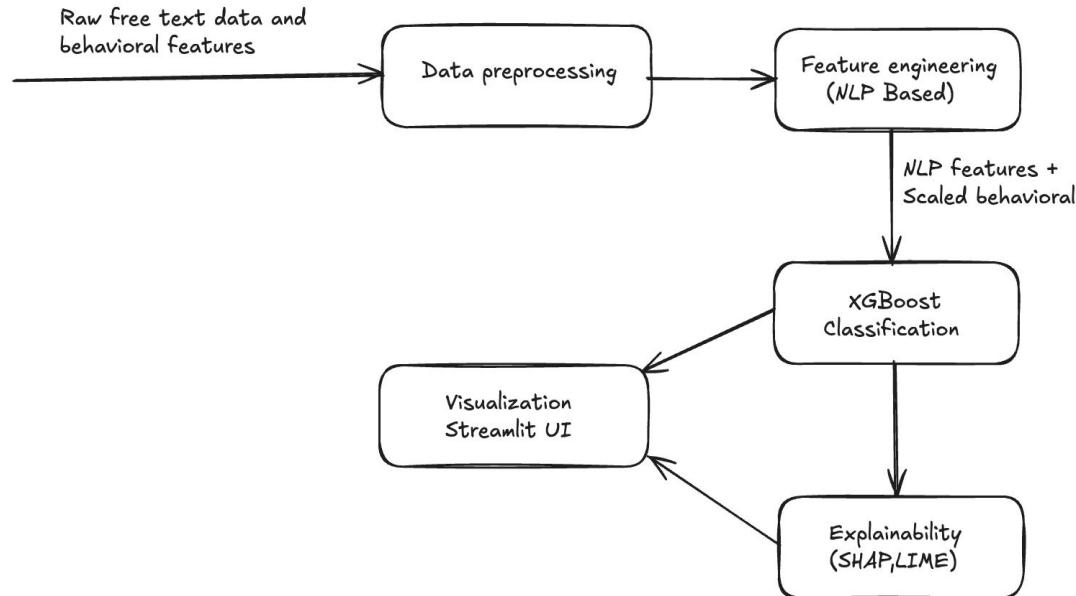
(2) test on overlap region



(3) train on overlap region

# Classification Model

# Pipeline



# Hyperparameters for XGBoost

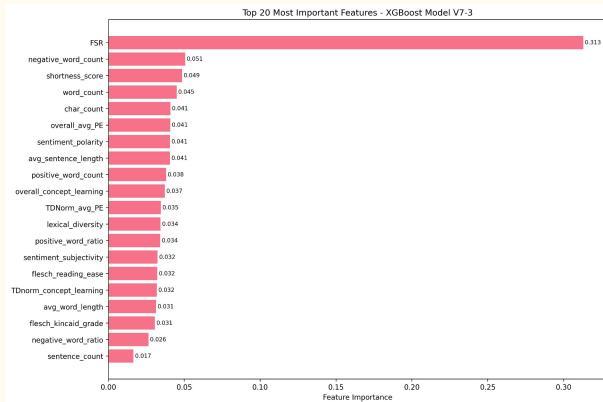
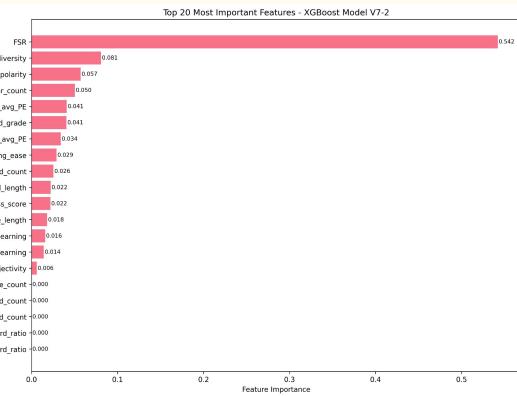
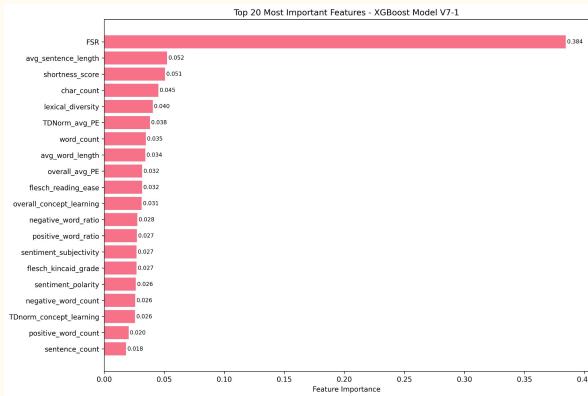
Hyperparameter	Value
<code>n_estimators</code>	100
<code>max_depth</code>	4
<code>learning_rate</code>	0.05
<code>subsample</code>	0.7
<code>colsample_bytree</code>	0.7
<code>min_child_weight</code>	3
<code>gamma</code>	0.1
<code>reg_alpha</code>	0.1
<code>reg_lambda</code>	1.0
<code>eval_metric</code>	<code>logloss</code>

# Model Versions and Performance

Model	Data File	LLM Agent	Features	Train Acc.	Test Acc.
V1	LLM Data	Sonnet	Sub, profile, subject, SRS.RAW, FSR, BIS, avg_PE, free_response, LPA_Profile_grand_mean, LPA_Profile_ASD_only, NLP features	0.916	0.900
V2	LLM Data	Sonnet	FSR, avg_PE, free_response, NLP features	0.914	0.880
V3	LLM Data	Qwen	FSR, avg_PE, free_response, NLP features	0.914	0.869
V4	LLM Data	Llama	FSR, avg_PE, free_response, NLP features	0.913	0.902
V5	LLM Data	None	FSR, avg_PE, free_response, NLP features	0.905	0.890
V6 (Optuna)	LLM Data	None	FSR, TDNorm_avg_PE, overall_avg_PE, TD-Norm_concept_learning, overall_concept_learning, NLP features	0.892	0.906
V7-1	LLM Data Aggregated	None	FSR, avg_PE (TD and Overall), concept learning (TD and Overall), NLP features, 80-20 stratified split	0.895	0.893
V7-2	LLM Data Aggregated	None	V7-1 features; Test on FSR-overlap region	1.000	0.473
V7-3	LLM Data Aggregated	None	V7-1 features; Train on FSR-overlap region	0.830	1.000
V8-1	LLM Data Aggregated	None	V7-1 features minus FSR; TDNorm_avg_PE, overall_avg_PE, TDNorm_concept_learning, overall_concept_learning; stratified split	0.639	0.628
V8-2	LLM Data Aggregated	None	V7-1 features minus FSR; Test on FSR-overlap region	0.806	0.532
V8-3	LLM Data Aggregated	None	V7-1 features minus FSR; Train on FSR-overlap region	0.545	0.663

# Model V7

**Features:** *FSR, avg\_PE(TD\_norm and overall), concept\_learning(TD\_norm and overall), NLP features*



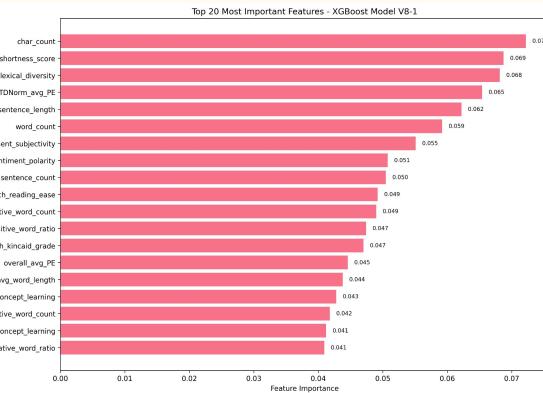
(1) Stratified 80/20 train-test

(2) test on overlap region

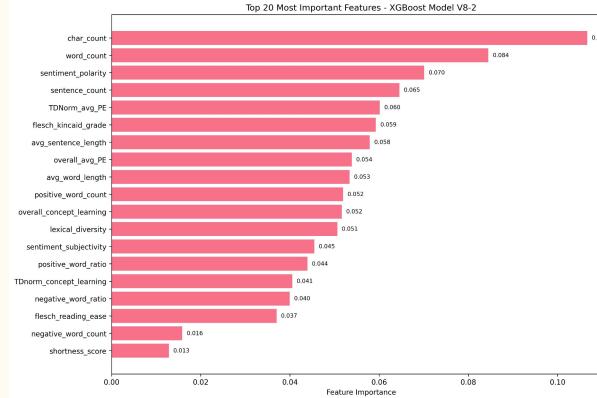
(3) train on overlap region

# Model V8 (without FSR)

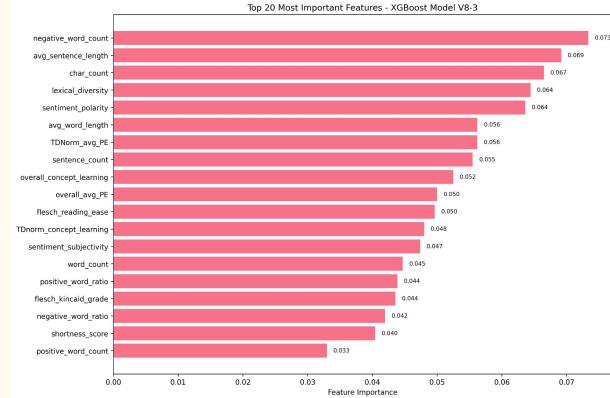
**Features:**  $avg\_PE(TD\_norm \text{ and } overall)$ ,  $concept\_learning(TD\_norm \text{ and } overall)$ , NLP features



(1) Stratified 80/20 train-test



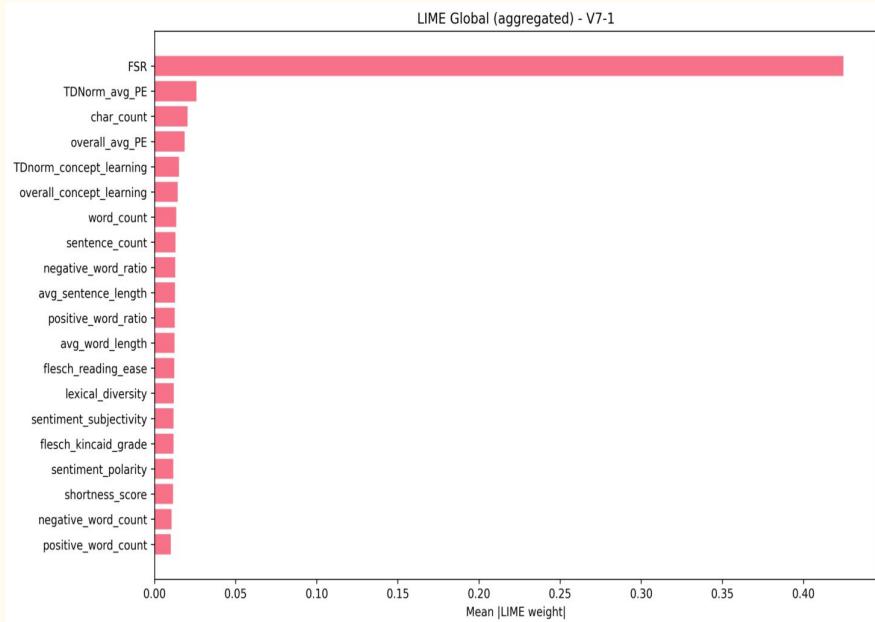
(2) test on overlap region



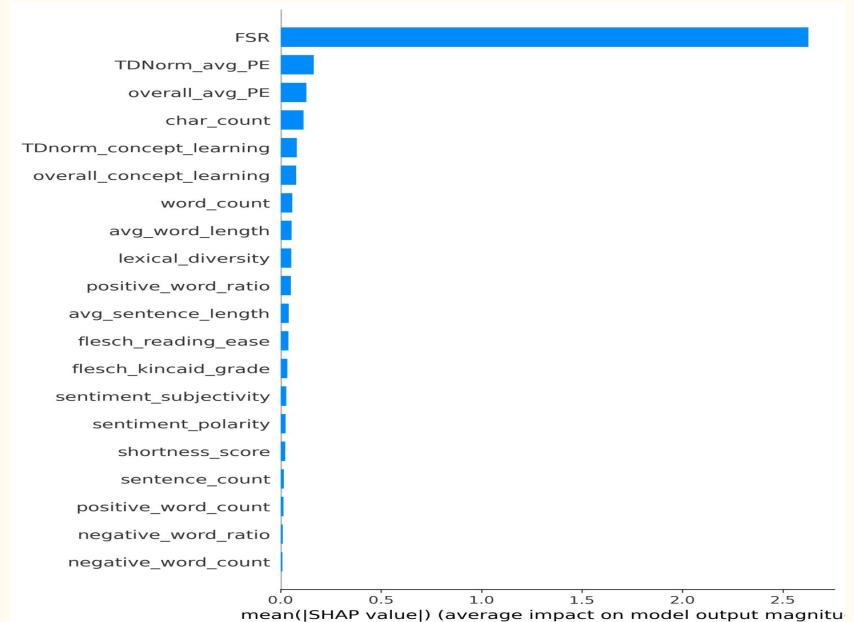
(3) train on overlap region

# xAI - Explainability

# Global Explainability

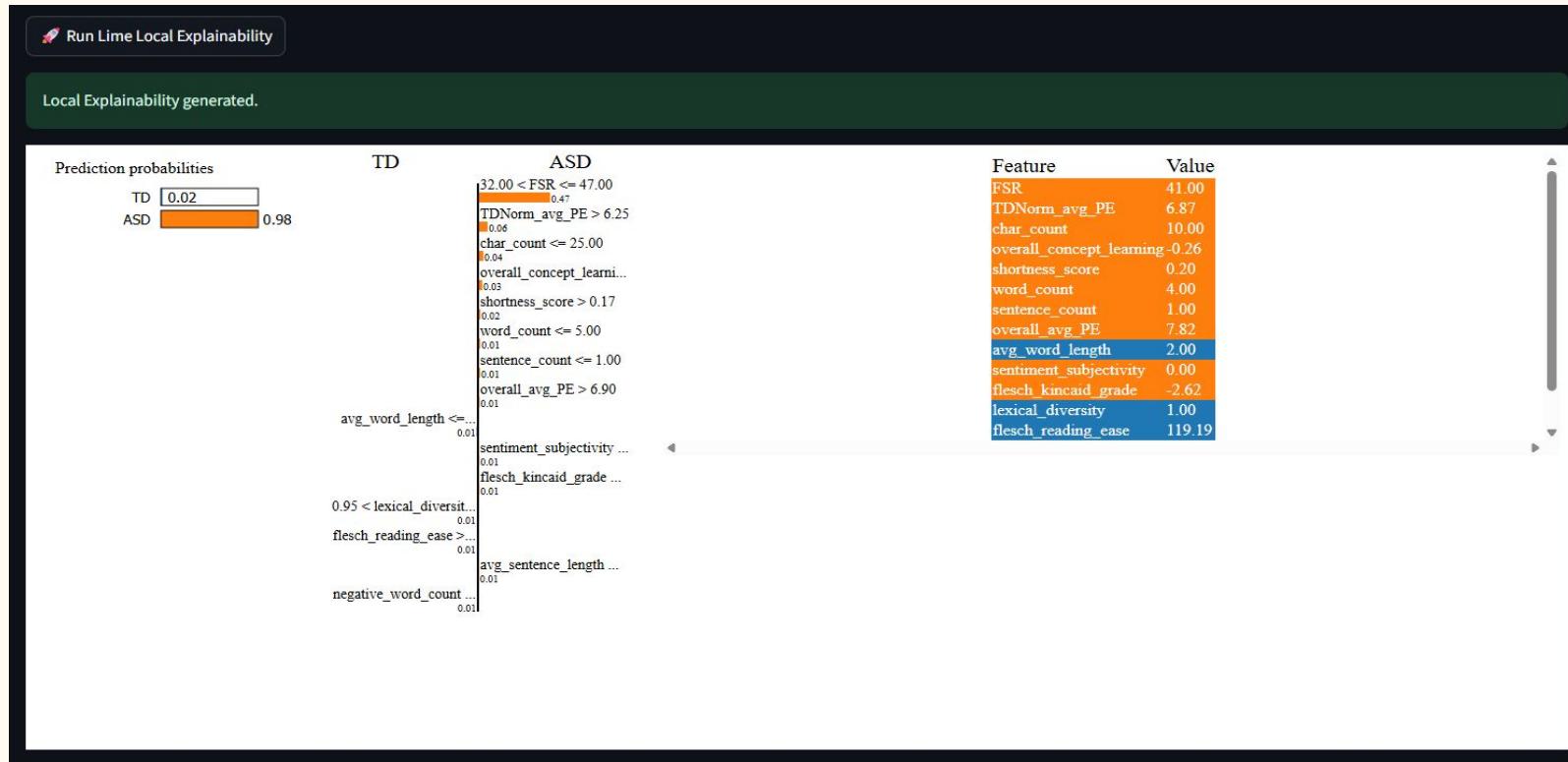


LIME

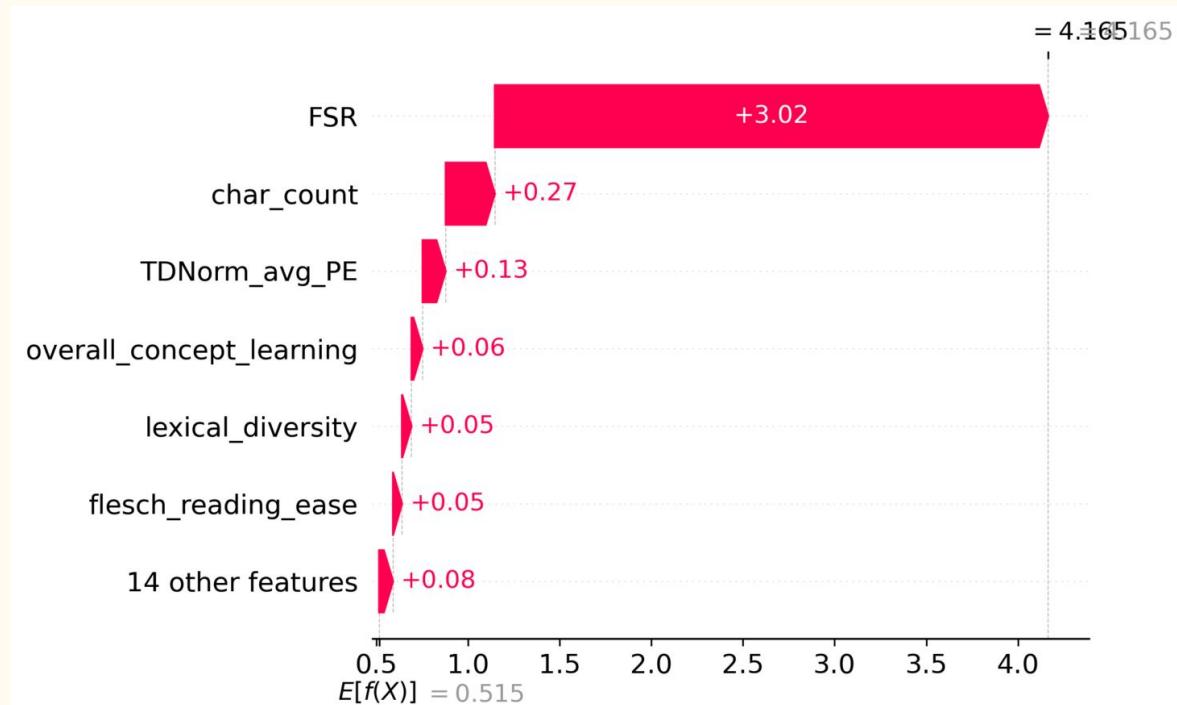


SHAP

# Local Explainability - LIME



# Local Explainability - SHAP



$$p = \frac{1}{1 + e^{-4.165}} \approx 0.985$$

Model predicts 98.5% probability of ASD

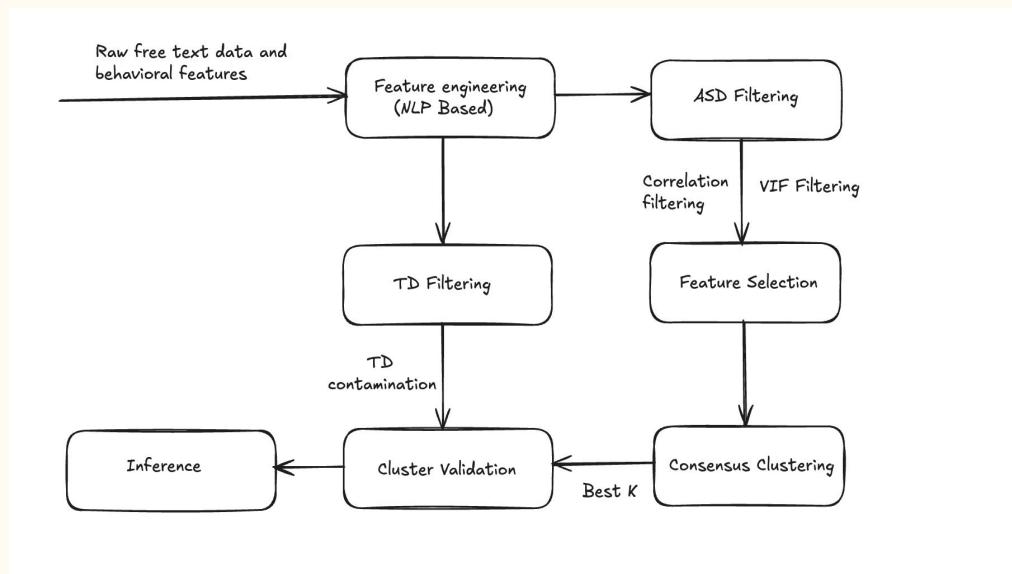
# Clustering

# Pipeline

**S Monti et al 2003 :** Consensus Clustering

**Yasin Şenbabaoğlu 2014 :**  
PAC for best K selection and  
Cluster Wise Stability  
Computation

Statistical Testing : One way  
ANOVA, KRUSKAL, DUNN  
PostHoc



# Feature Selection and PCA

Metadata Parameters

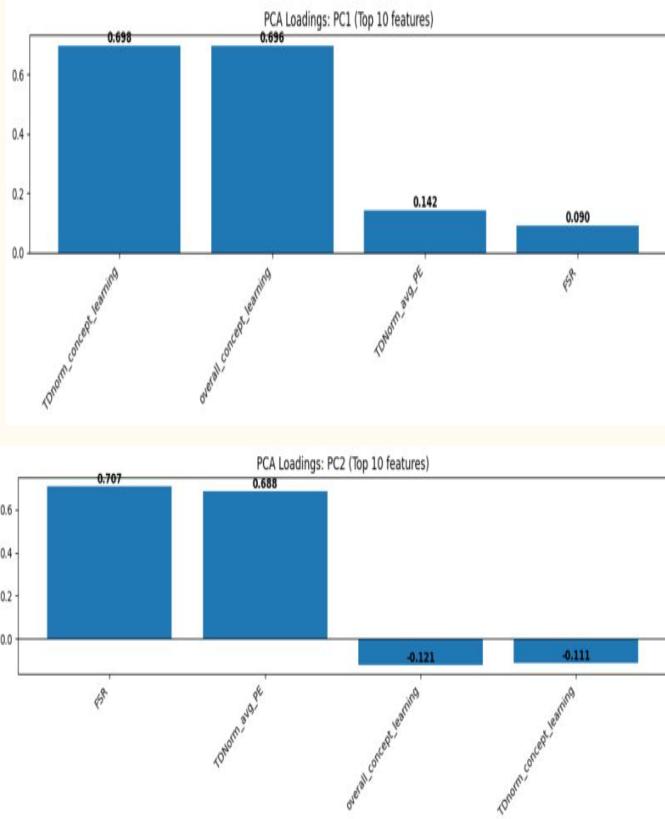
	Parameter	Value
0	KMIN	2
1	KMAX	12
2	ITERATIONS	1000
3	ASD_ONLY	True

Clustering

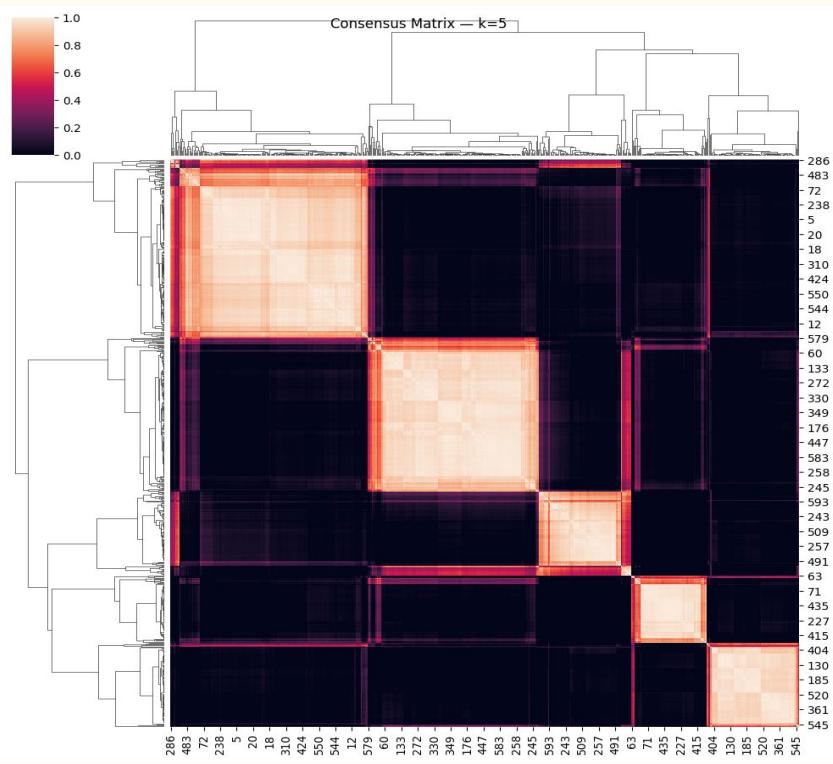
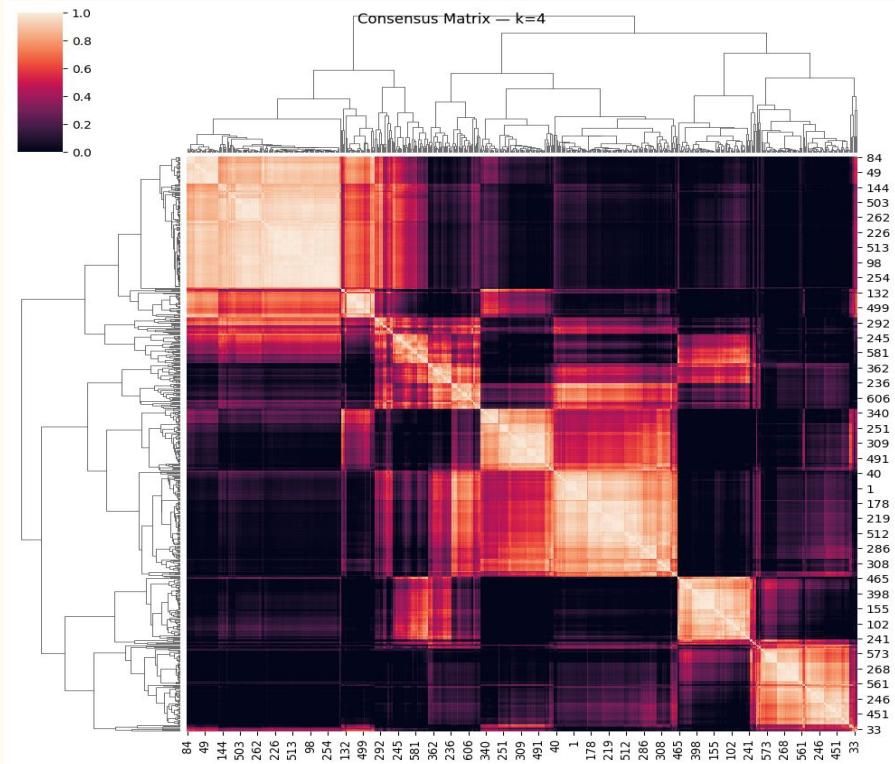
	Parameter	Value
0	KMIN	2
1	KMAX	12
2	ITERATIONS	1000
3	ASD_ONLY	True

Preprocessing

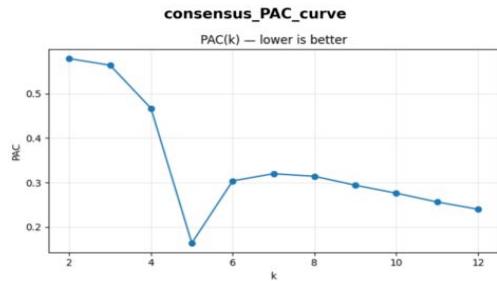
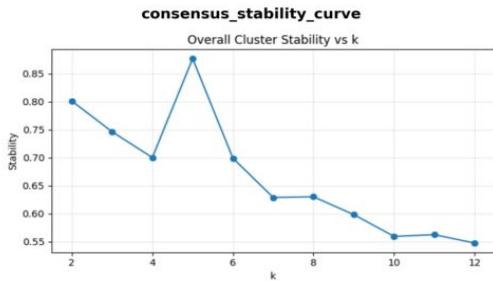
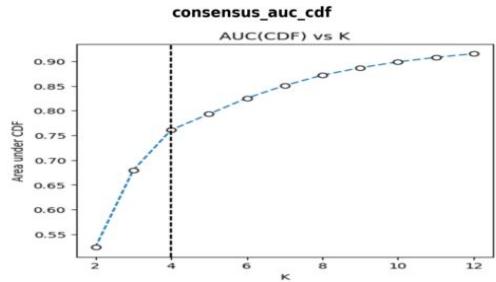
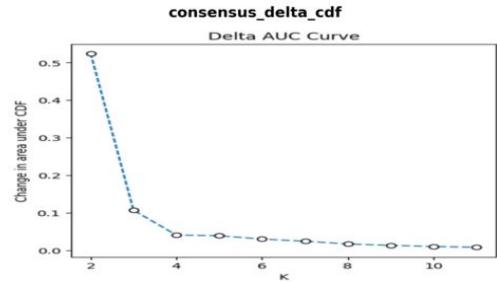
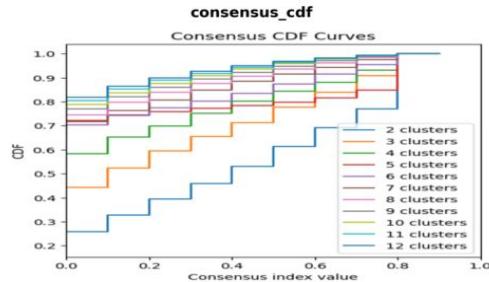
	Parameter	Value
0	CORR_THRESH	0.7
1	USE_NLP	False
2	HIGH_CORR_FEATS	overall_avg_PE,BIS,SRS.Raw
3	SELECTED_FEATURES	FSR,TDNorm_avg_PE,TDnorm_concept_learning,overall_concept_learning



# Best K Selection



# Best K Selection



## Best-K Summary

- PAC Minimum: 5
- PAC Knee: 12
- Max Stability: 2
- Stability Knee: 2
- AUC Knee: 4
- AUC Knee: 2

# Cluster Validation

## Significance Test Results

### One Way ANOVA and Kruskal Test Results for Consensus Clusters

feature	anova_F	anova_p	kruskal_H	kruskal_p
FSR	142.747	0.0	328.161	0.0
TDNorm_avg_PE	196.72	0.0	249.754	0.0
TDnorm_concept_learning	208.447	0.0	306.228	0.0
overall_concept_learning	163.215	0.0	287.5	0.0

significance.png

### Dunn Posthoc Test for Feature: FSR

	0	1	2	3	4
0	1.0	0.0	0.001	0.0	0.0
1	0.0	1.0	0.0	0.095	0.0
2	0.001	0.0	1.0	0.972	0.0
3	0.0	0.095	0.972	1.0	0.0
4	0.0	0.0	0.0	0.0	1.0

### Dunn Posthoc Test for Feature: overall\_concept\_learning

	0	1	2	3	4
0	1.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	1.0	0.008
2	0.0	0.0	1.0	0.0	0.0
3	0.0	1.0	0.0	1.0	0.362
4	0.0	0.008	0.0	0.0	1.0

### Dunn Posthoc Test for Feature: TDNorm\_avg\_PE

	0	1	2	3	4
0	1.0	0.828	0.264	0.0	0.003
1	0.828	1.0	0.0	0.0	0.174
2	0.264	0.0	1.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0
4	0.003	0.174	0.0	0.0	1.0

### Dunn Posthoc Test for Feature: TDnorm\_concept\_learning

	0	1	2	3	4
0	1.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	0.057	0.01
2	0.0	0.0	1.0	0.0	0.0
3	0.0	0.057	0.0	1.0	1.0
4	0.0	0.01	0.0	1.0	1.0

# TD Contamination and Inference

Cluster Profile for Feature: FSR

	count	mean	std	median
0	90.0	37.189	10.46	38.0
1	191.0	52.309	7.576	51.0
2	73.0	45.068	11.355	45.0
3	90.0	48.967	13.441	49.0
4	167.0	29.574	7.881	31.0

Cluster Profile for Feature: overall\_concept\_learning

	count	mean	std	median
0	90.0	-0.201	0.133	-0.175
1	191.0	-0.034	0.098	-0.033
2	73.0	0.239	0.139	0.213
3	90.0	-0.029	0.126	-0.029
4	167.0	0.01	0.086	0.009

Cluster Profile for Feature: TDNorm\_avg\_PE

	count	mean	std	median
0	90.0	5.646	1.057	5.417
1	191.0	5.32	0.723	5.3
2	73.0	6.088	1.236	5.85
3	90.0	8.205	0.915	8.233
4	167.0	5.092	0.858	4.967

Cluster Profile for Feature: TDnorm\_concept\_learning

	count	mean	std	median
0	90.0	-0.332	0.196	-0.325
1	191.0	-0.026	0.131	-0.025
2	73.0	0.37	0.203	0.357
3	90.0	0.032	0.169	0.035
4	167.0	0.033	0.125	0.032

## TD Contamination Results:

cluster	ASD_count	TD_count	total	td_fraction
0	0.0	90.0	54.0	0.375
1	1.0	191.0	2.0	0.01036
2	2.0	73.0	31.0	0.29808
3	3.0	90.0	3.0	0.03226
4	4.0	167.0	278.0	0.62472

# TD Contamination and Inference

## Cluster 1 – “High FSR ASD subtype”

- Almost pure ASD.
- Strongly elevated FSR vs TD, with mildly depressed concept metrics.
- Statistically distinct from TD cluster on all major features.

## Cluster 3 – “High PE ASD subtype”

- Almost pure ASD.
- Extremely high TDNorm\_avg\_PE and elevated FSR, but near-norm concept-learning.
- Statistically distinct from all other clusters on PE.

## Cluster 0 – “Low concept-learning, mixed ASD/TD group”

- Clear, statistically unique deficit in concept metrics,
- But substantial TD contamination

## Cluster 2 – “High-performing mixed group”

- High concept and TDnorm\_concept scores.
- Mixed ASD and TD

## Cluster 4 – “TD-dominant baseline cluster”

- Lowest FSR and PE, near-norm concept scores.
- Majority TD

# Dashboard

# Future Work

1. Test Modeling and Clustering with external datasets for additional validation and confirmation of Generalized Detection Model
2. Perform Clustering with Overall Participants to understand cluster formation
3. Integrate xAI with Cluster Assignment