



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
UNIVERSITY OF PIRAEUS

# Ανίχνευση Συναισθημάτων σε Ηχητικά Δεδομένα

Μηχανική Μάθηση 2025 – 2026

---

Γκούβρα Ελπίδα, Πλέσσιας Γιώργος, Τσαγκέτας Ορέστης

# Εισαγωγή

## Π Αντικείμενο

Ανίχνευση συναισθημάτων σε πραγματικά ηχητικά σήματα, εστιάζοντας στην εξαγωγή χαρακτηριστικών από την ομιλία.

## ⚠ Πρόκληση

Δημιουργία συστήματος που κατανοεί τις διαφορές μεταξύ των ανθρώπων στην έκφραση των συναισθημάτων.

## ◎ Στόχος

Ταξινόμηση 4 κλάσσεων (Angry, Happy, Neutral, Sad) με κλασικά μοντέλα ML.

# Δεδομένα & Μεθοδολογία

## Δεδομένα

### IEMOCAP Dataset

Interactive Emotional Dyadic Motion Capture

5 Sessions 5,531 Samples

Το κύριο dataset εκπαιδευσης. Ηχητικά δεδομένα από διαλόγους ηθοποιών.

Angry Happy+Exc Neutral Sad



### CREMA-D Dataset

Crowd-sourced Emotional Multimodal Actors

91 Actors 4,900 Samples

Χρήση αποκλειστικά για Zero-Shot Evaluation.



## Μεθοδολογία

### Προεπεξεργασία Ήχου

Resampling στα 16kHz, Normalization (max abs value), αφαίρεση σιωπής (Trim silence top\_db=20) και Framing.

### Εξαγωγή Χαρακτηριστικών

Χρήση pyAudioAnalysis. Παράθυρο 50ms / Βήμα 25ms.

272 Features: Time-domain, Spectral (MFCCs, Chroma), Deltas

### Μοντέλα Μηχανικής Μάθησης

Εκπαίδευση 7 κλασικών ταξινομητών:

XGBoost

SVM

Random Forest

Logistic Reg

KNN

Naive Bayes

Decision Tree

# Σενάρια Εργασίας



## 1. Σενάριο 1: IEMOCAP 80/20 split

Preprocess > Split 80/20 > StandardScaler > Train/Tune



## 2. Σενάριο 2: IEMOCAP LOSO

Preprocess > Split > 5-Fold Session Split > Fold Training > Train/Tune



## 3. Σενάριο 3: IEMOCAP 80/20 + PCA

Preprocess > Split > PCA > Train/Tune



## Zero-Shot Evaluation

CREMA-D Dataset

Όλα τα μοντέλα από κάθε workflow δοκιμάζονται αυτόμata στo CREMA-D χωρίς καμία επανεκπαίδευση.

# Αρχιτεκτονική

## Steps για κατασκευή των pipelines

- **Προεπεξεργασία:** Καθαρισμός, κανονικοποίηση και προετοιμασία χαρακτηριστικών.
- **Στρατηγικές Split:** Διαφορετικά splits (Hold-out, Cross-validation, LOSO).
- **Εκπαίδευση:** Πολλαπλά μοντέλα (SVM, RF, XGB, κ.α.).
- **Βελτιστοποίηση:** Fine-tuning υπερπαραμέτρων για καλύτερη απόδοση.
- **Αξιολόγηση:** Μετρικές, πίνακες σύγχυσης και learning curves.
- **Ανάλυση PCA:** Μείωση διαστάσεων για πιο ελαφρύ μοντέλο.
- **Συγκριτική Αξιολόγηση:** Τελική σύγκριση όλων των pipelines με AUC.

## Pipelines: Σενάρια

### Pipeline

- Κάθε pipeline συνθέτει τα βήματα σε ένα πλήρες, αυτοματοποιημένο σενάριο.
- Ενιαία ροή από preprocessing μέχρι αξιολόγηση, χωρίς χειροκίνητα κενά.

### Απομόνωση

- Κάθε πείραμα έχει τη δική του δομή (features → models → results) για καθαρό διαχωρισμό.
- Εύκολη διαχείριση και καθάρισμα ανά σενάριο.

# Επισκόπηση Σεναρίων

## 1. Σενάριο 1

80/20 STRATIFIED

### Baseline αξιολόγηση

- Step 1: Preprocess IEMOCAP.
- Step 2: Stratified 80/20 split + normalize.
- Step 3: Train models.
- Step 4: Evaluate models.
- Step 5: Hyperparameter tuning.
- Step 6: Evaluate tuned models.
- Step 7: Zero-Shot Evaluation (CREMA-D).

## 2. Σενάριο 2

LOSO STRATEGY

### Γενίκευση Προβλέψεων

- Step 1: Preprocess IEMOCAP.
- Step 2: LOSO split (folds).
- Step 3: Train/tune per fold.
- Step 4: Evaluate per fold.
- Step 5: Zero-Shot Evaluation per fold.
- Step 6: Aggregate LOSO results.

## 3. Σενάριο 3

PCA OPTIMIZATION

### Μείωση Διαστάσεων

- Step 1: Preprocess IEMOCAP.
- Step 2: 80/20 split (raw, no normalize).
- Step 3: PCA στο train (~99.03% variance).
- Step 4: Train models (no-scale).
- Step 5: Evaluate models.
- Step 6: Hyperparameter tuning.
- Step 7: Evaluate tuned models.
- Step 8: Zero-Shot Evaluation (CREMA-D).

# Σενάριο 1

1

## Προεπεξεργασία Δεδομένων

Εξαγωγή χαρακτηριστικών από τα ηχητικά αρχεία IEMOCAP με pyAudioAnalysis.

2

## Διαχωρισμός Train/Test

Stratified split 80/20 (Train: 4.424, Test: 1.107) με κανονικοποίηση.

3

## Εκπαίδευση Μοντέλων

Εκπαίδευση 7 ταξινομητών (XGBoost, SVM, RF, Logistic Reg, KNN, NB, DTR).

4

## Αρχική Αξιολόγηση

Αξιολόγηση με confusion matrices, learning curves και metrics.

## 1 Βασικές Πληροφορίες

### Dataset

IEMOCAP (5.531 samples)

Train/Test Split	Train Samples	Test Samples
80% / 20% (stratified)	4.424	1.107

### Features

272 (68 × 4 stats)

### Models

XGBoost SVM RF Logistic Reg KNN  
NB DTR

# Σενάριο 1

5

## Hyperparameter Tuning

RandomizedSearchCV με weighted F1-score.

6

## Τελική Αξιολόγηση

Αξιολόγηση των βελτιστοποιημένων μοντέλων με τα ίδια metrics για σύγκριση.

7

## Zero-Shot Evaluation

Έλεγχος γενίκευσης σε CREMA-D dataset. Τα μοντέλα αξιολογούνται χωρίς επανεκπαίδευση.



### Zero-Shot Ranking (CREMA-D) - Accuracy

1

Naive Bayes

54.29%

2

XGBoost

52.22%

3

SVM

50.82%

4

KNN

48.82%

5

Random Forest

47.04%

6

Logistic Reg

44.37%

7

Decision Tree

35.24%

# Σενάριο 2

1

## Preprocessing & Feature Extraction

Καθαρισμός δεδομένων IEMOCAP και εξαγωγή χαρακτηριστικών.

2

## Iterative Splitting

Επανάληψη διαδικασίας για κάθε ομιλητή. Test: 1 Speaker | Train: Remaining Speakers.

3

## Normalization & Training

Εφαρμογή Scaler μόνο στο Train set. Εκπαίδευση μοντέλων όπως SVM, RF, XGB ή βελτιστοποίηση υπερπαραμέτρων.

4

## Evaluation

Αξιολόγηση στο Test set του τρέχοντος fold. Καταγραφή μετρικών απόδοσης.

## Χαρακτηριστικά & Στόχοι

### Speaker Independence

Διασφαλίζει ότι το μοντέλο μαθαίνει συναίσθηματικά χαρακτηριστικά και όχι την ταυτότητα του ομιλητή. Αποφεύγει το overfitting σε συγκεκριμένες φωνές.

### Ανθεκτικότητα

Προσομοιώνει ρεαλιστικές συνθήκες όπου το σύστημα καλείται να αναγνωρίσει το συναίσθημα ενός νέου, άγνωστου χρήστη.

### Data Efficiency

Χρησιμοποιεί το μεγαλύτερο μέρος των δεδομένων για εκπαίδευση σε κάθε fold (~80-90%), διατηρώντας παράλληλα αυστηρό διαχωρισμό για την αξιολόγηση.

# Σενάριο 2

4

## Evaluation

Αξιολόγηση στο Test set του τρέχοντος fold. Καταγραφή μετρικών απόδοσης.

5

## Zero-Shot Evaluation

Πρόβλεψη σε εξωτερικό dataset CREMA-D για έλεγχο γενίκευσης σε εντελώς άγνωστα δεδομένα.

6

## Aggregation

Συγκέντρωση αποτελεσμάτων από όλα τα folds. Υπολογισμός μέσου όρου και τυπικής απόκλισης.



### Zero-Shot Ranking (CREMA-D) - Accuracy

1

Naive Bayes

53.71%

2

SVM

51.84%

3

XGBoost

51.13%

4

Random Forest

47.17%

5

KNN

47.09%

6

Logistic Reg

38.66%

7

Decision Tree

36.87%

## Σενάριο 2

Zero-Shot avá fold

MODEL	FOLD 0	FOLD 1	FOLD 2	FOLD 3	FOLD 4	AVG
Naive Bayes	53.10%	53.94%	54.08%	54.00%	53.43%	<b>53.71%</b>
SVM	54.27%	51.80%	49.27%	50.94%	52.96%	<b>51.84%</b>
XGBoost	52.04%	50.33%	52.00%	50.57%	50.73%	<b>51.13%</b>
Random Forest	50.76%	44.49%	44.47%	46.65%	49.49%	<b>47.17%</b>
KNN	46.41%	46.49%	48.88%	46.31%	47.37%	<b>47.09%</b>
Logistic Reg	36.67%	38.94%	36.02%	39.94%	41.71%	<b>38.66%</b>
Decision Tree	38.20%	39.06%	37.45%	35.73%	33.92%	<b>36.87%</b>

# Σενάριο 3

1

## Προεπεξεργασία

Καθαρισμός δεδομένων και εξαγωγή χαρακτηριστικών.

2

## Διαχωρισμός Train/Test

Διαχωρισμός 80/20 χωρίς κανονικοποίηση.

3

## PCA

Εφαρμογή PCA στο train (variance 0,99)

4

## Εκπαίδευση

Εκπαίδευση μοντέλων.

## Κύρια Σημεία

### PCA

272 διαστάσεις → 205 components (99.03% variance).

### Linear

Logistic Reg 59.08% · SVM 61.16%

### Trees

RF/XGB ασθενούν με PCA.

### KNN/NB/DTR

Χαμηλή απόδοση στο PCA setup.

# Σενάριο 3

5

## Αξιολόγηση

Αξιολόγηση μοντέλων.

6

## Tuning

RandomizedSearchCV και επαναξιολόγηση.

7

## Zero-Shot Evaluation

Zero-Shot Evaluation στο CREMA-D με ίδιο PCA/scaler.



### Zero-Shot Ranking (CREMA-D) - Accuracy

1

SVM

51.16%

2

KNN

49.06%

3

XGBoost

47.04%

4

Random Forest

45.22%

5

Logistic Reg

44.53%

6

Decision Tree

41.84%

7

Naive Bayes

30.76%

# Αποτελέσματα: Σενάριο 1

Σύγκριση Test (IEMOCAP) vs Zero-Shot (CREMA-D)

MODEL	ACC		F1-W		F1-M		UAR		PREC-W		REC-W	
	TEST	ZERO										
XGBoost	62.24%	52.22%	62.21%	52.48%	62.59%	52.26%	62.25%	52.50%	62.80%	55.20%	62.24%	52.22%
SVM	61.25%	50.82%	61.25%	50.12%	61.83%	50.08%	61.58%	51.39%	61.55%	56.00%	61.25%	50.82%
Logistic Reg	59.89%	44.37%	59.83%	44.39%	60.39%	44.14%	60.50%	44.14%	59.86%	50.92%	59.89%	44.37%
Random Forest	58.81%	47.04%	58.78%	44.45%	59.17%	44.62%	58.60%	48.71%	59.49%	61.79%	58.81%	47.04%
KNN	55.65%	48.82%	55.19%	47.60%	55.80%	47.36%	56.52%	48.92%	55.88%	49.01%	55.65%	48.82%
Naive Bayes	49.23%	54.29%	47.80%	51.35%	49.45%	51.21%	52.86%	54.38%	49.47%	53.12%	49.23%	54.29%
Decision Tree	46.88%	35.24%	46.80%	36.39%	47.25%	36.16%	47.09%	35.20%	46.94%	46.74%	46.88%	35.24%

# Αποτελέσματα: Σενάριο 2

Σύγκριση Test vs Zero-Shot

MODEL	ACC		F1-W		F1-M		UAR		PREC-W		REC-W	
	TEST	ZERO										
SVM	56.38%	51.84%	56.34%	50.29%	57.04%	50.29%	57.13%	52.72%	58.01%	57.48%	56.38%	51.84%
XGBoost	56.26%	51.13%	56.13%	50.70%	56.63%	50.49%	57.13%	51.31%	57.22%	53.28%	56.26%	51.13%
Random Forest	54.55%	47.17%	54.47%	44.45%	54.99%	44.64%	55.01%	48.78%	56.53%	60.41%	54.55%	47.17%
Logistic Reg	53.58%	38.66%	53.20%	36.68%	53.65%	36.32%	54.62%	38.02%	54.20%	45.48%	53.58%	38.66%
KNN	51.53%	47.09%	51.05%	46.06%	51.37%	45.75%	52.26%	47.10%	52.98%	48.28%	51.53%	47.09%
Naive Bayes	47.95%	53.71%	46.37%	50.50%	47.52%	50.34%	51.22%	53.74%	48.10%	52.22%	47.95%	53.71%
Decision Tree	43.67%	36.87%	43.67%	36.69%	43.65%	36.62%	43.95%	37.25%	44.57%	41.30%	43.67%	36.87%

## Παρατηρήσεις

Με βάση το Accuracy, το καλύτερο Test μοντέλο είναι SVM (56.38%). Στο Zero-Shot (CREMA-D) ξεχωρίζει το Naive Bayes (53.71%), δείχνοντας καλύτερη γενικευση.

# Αποτελέσματα: Σενάριο 3

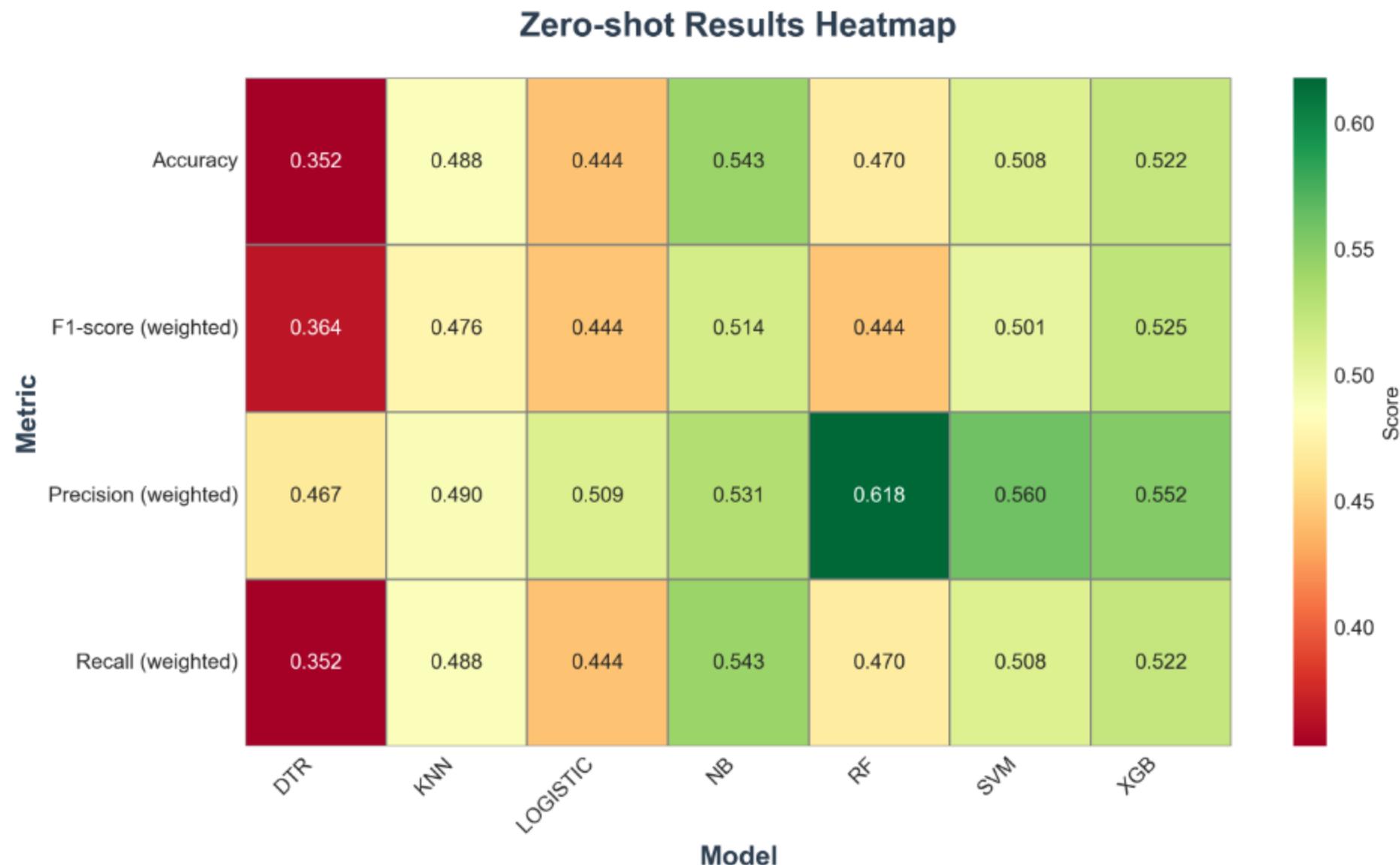
Σύγκριση Test vs Zero-Shot

MODEL	ACC		F1-W		F1-M		UAR		PREC-W		REC-W	
	TEST	ZERO										
SVM	61.16%	51.16%	61.15%	50.55%	61.66%	50.49%	61.42%	51.71%	61.42%	56.12%	61.16%	51.16%
Logistic Reg	59.08%	44.53%	59.00%	44.54%	59.61%	44.28%	59.79%	44.33%	59.05%	51.21%	59.08%	44.53%
XGBoost	56.19%	47.04%	56.18%	47.37%	56.67%	47.16%	56.36%	47.19%	56.41%	49.91%	56.19%	47.04%
KNN	55.83%	49.06%	55.41%	47.96%	56.07%	47.69%	56.72%	49.12%	56.03%	49.28%	55.83%	49.06%
Random Forest	55.19%	45.22%	54.69%	44.74%	54.25%	44.83%	53.01%	46.30%	58.30%	55.82%	55.19%	45.22%
Decision Tree	47.34%	41.84%	46.94%	42.19%	46.89%	42.12%	46.47%	42.49%	47.61%	47.35%	47.34%	41.84%
Naive Bayes	39.30%	30.76%	38.14%	22.91%	38.87%	22.69%	42.54%	30.01%	44.65%	25.06%	39.30%	30.76%

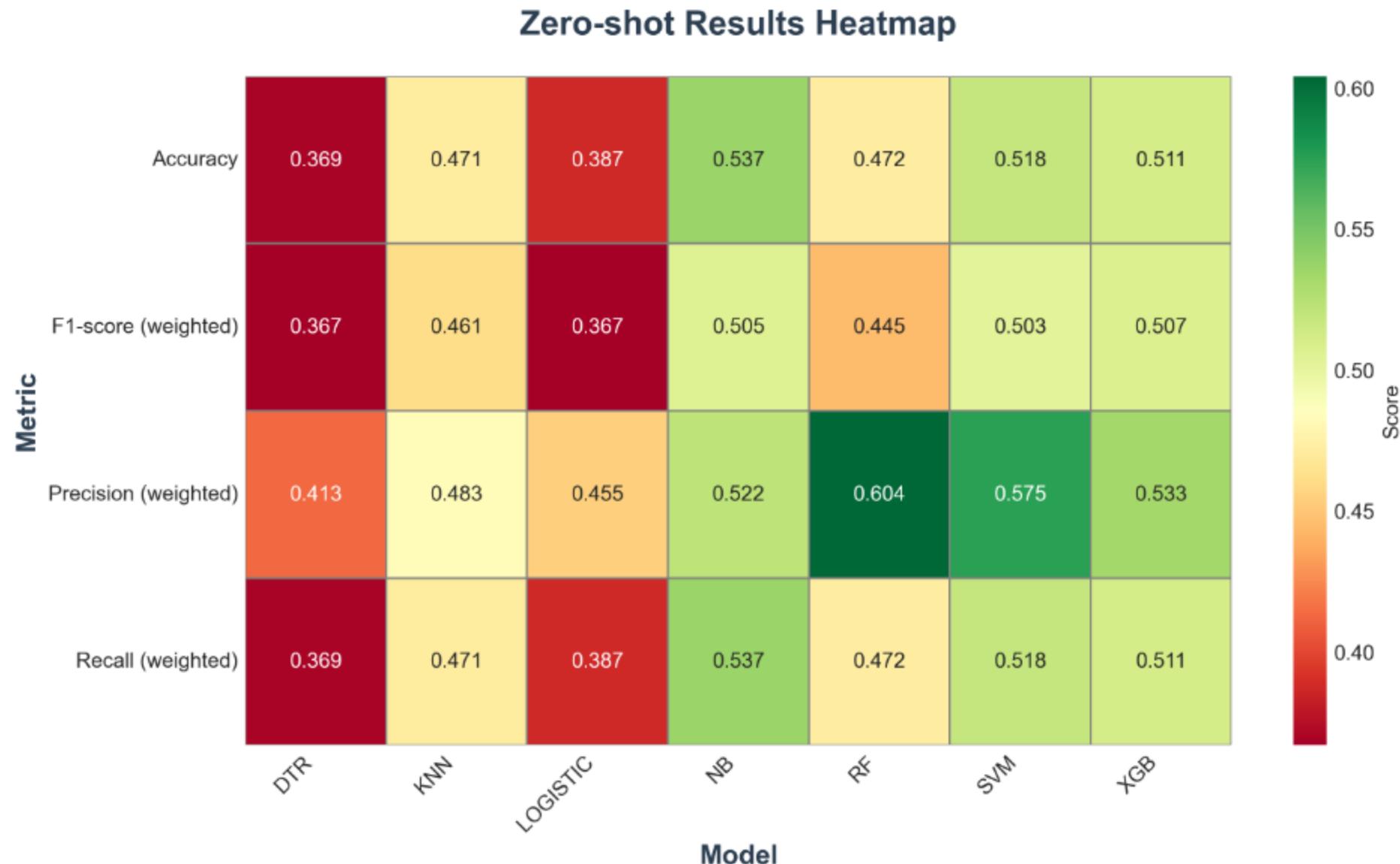
## Παρατηρήσεις

Με βάση το Accuracy, το καλύτερο Test μοντέλο είναι SVM (61.16%). Στο Zero-Shot (CREMA-D) ξεχωρίζει το SVM (51.16%), δείχνοντας καλύτερη γενίκευση.

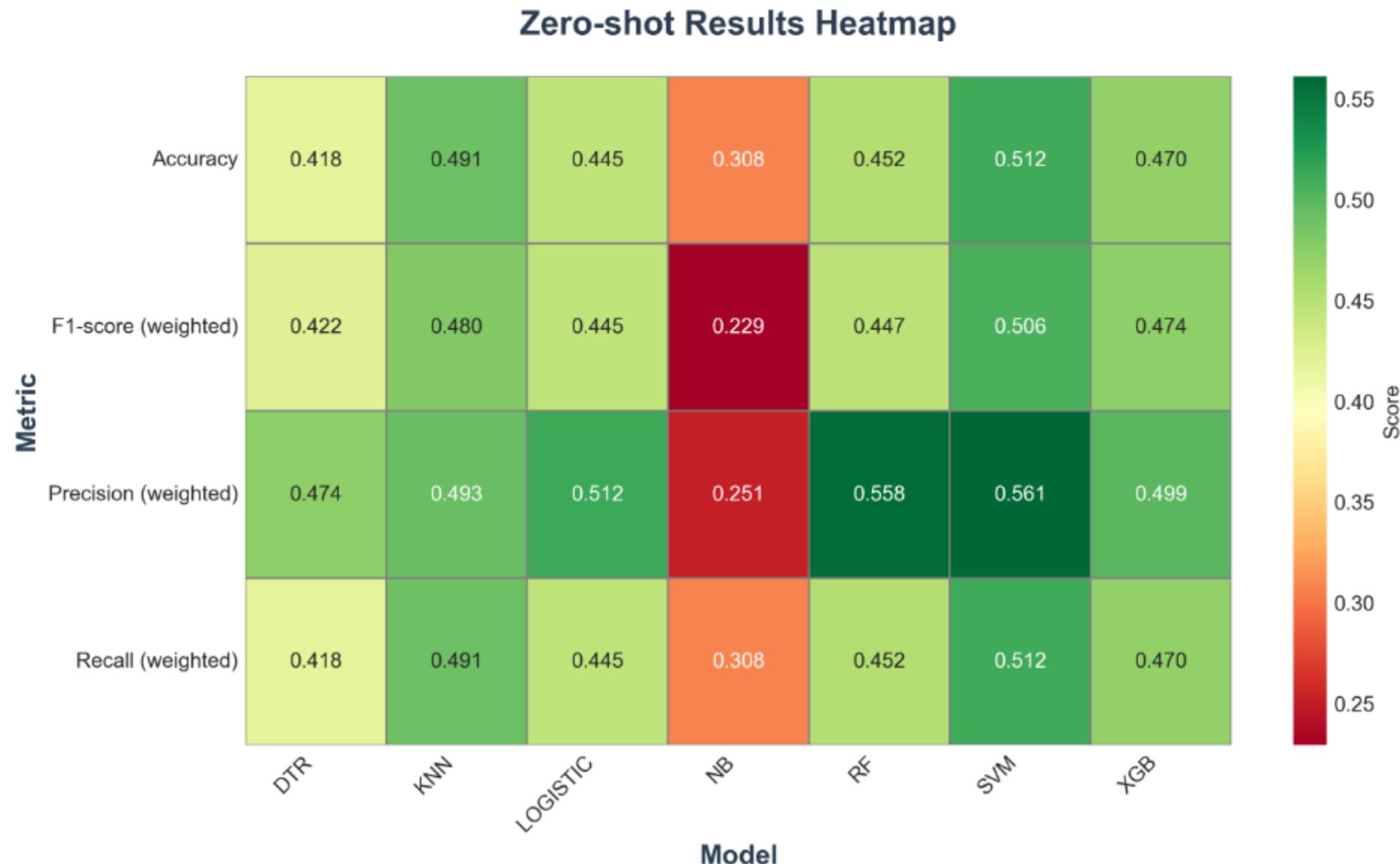
# Zero-shot Results Heatmap – Σενάριο 1



# Zero-shot Results Heatmap – Σενάριο 2



# Zero-shot Results Heatmap – Σενάριο 3

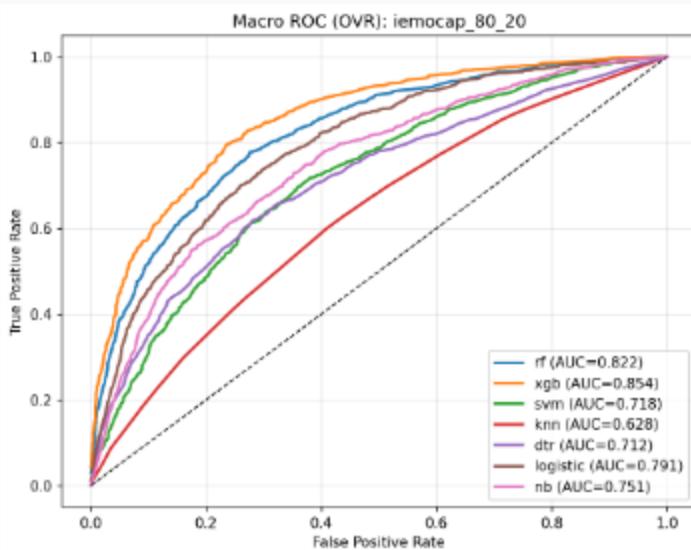


# Καμπύλες ROC ανά σενάριο

Macro AUC (one-vs-rest) ανά σενάριο

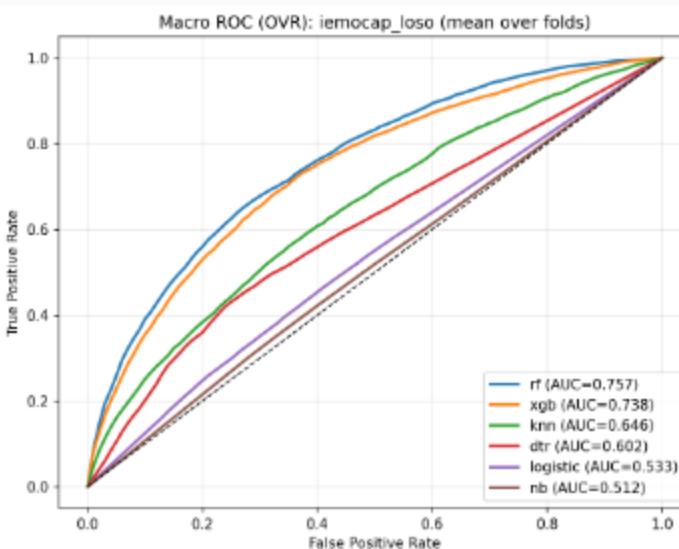
## Σενάριο 1: 80/20 Split (IEMOCAP)

Ένα train/test split · καλύτερο: XGBoost (AUC 0.8542)



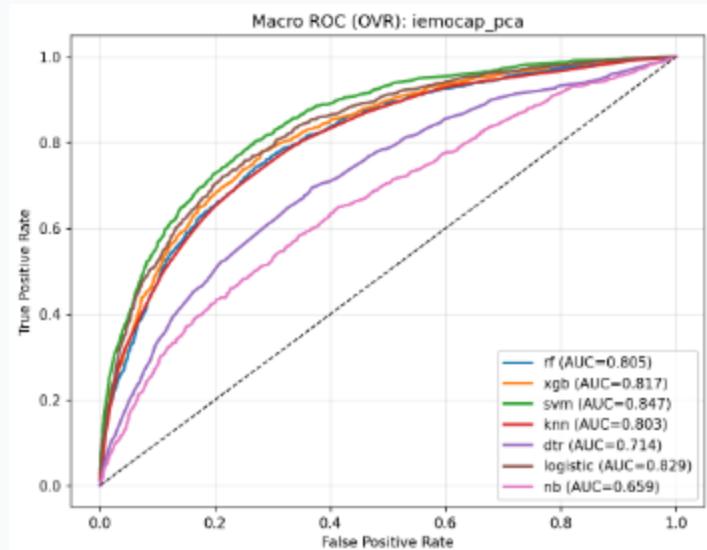
## Σενάριο 2: Leave-One-Speaker-Out (LOSO)

Μέσος AUC ανά fold · καλύτερο: Random Forest (AUC 0.7572)



## Σενάριο 3: 80/20 + PCA

Μειωμένη διαστατικότητα · καλύτερο: SVM (AUC 0.8466)



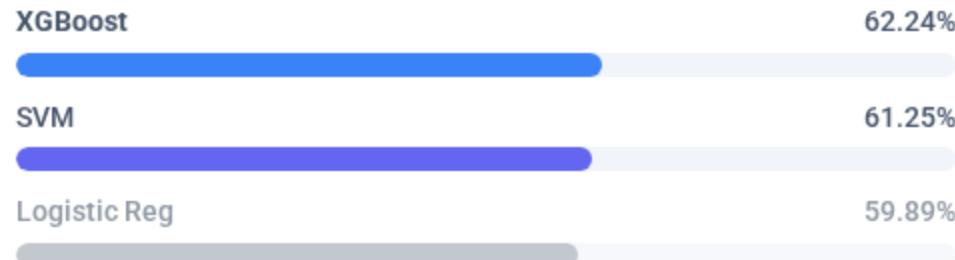
# Συμπεράσματα

Σύνοψη αποτελεσμάτων σεναρίων

## Ξ Σύγκριση Ακρίβειας (Accuracy)

### Σενάριο 1: IEMOCAP 80/20 (Baseline)

Best Case



### Σενάριο 2: IEMOCAP LOSO (Speaker-Independent)



## Σενάριο 3: Η Επίδραση του PCA

Η μείωση διαστάσεων (272 → 205, 99.03% variance) επηρέασε διαφορετικά τις οικογένειεις μοντέλων:

### LINEAR MODELS

↑ Stable

LogReg & SVM διατηρούν ~60%

### TREE MODELS

↓ 4-6%

XGBoost & RF μειώνονται



## Zero-Shot Surprise (CREMA-D)

Έλεγχος γενίκευσης

54.29%

Naive Bayes

**Παράδοξο:** Το Naive Bayes έχει σχετικά χαμηλή απόδοση στο IEMOCAP (49.23%), αλλά την καλύτερη γενίκευση στο CREMA-D.

**ⓘ** Ένδειξη overfitting των πολύπλοκων μοντέλων (XGB/SVM) στα ιδιαίτερα χαρακτηριστικά του IEMOCAP.