

Αναγνώριση Συναισθήματος από Ήχο

Κυριαζόπουλος Χρήστος
ΑΜ:2207

Μάθημα: Μηχανική Μάθηση
14/02/2023



Περιεχόμενα

1	Εισαγωγή	4
2	Σετ Δεδομένων	4
3	Ανάλυση Διαδικασιών και Αποτελέσματα	5
3.1	Acted Emotional Speech Dynamic Database - Αποτελέσματα	6
3.1.1	Συμπεράσματα Acted Emotional Speech Dynamic Database	12
3.2	EmoVDB - Αποτελέσματα	13
3.2.1	Συμπεράσματα EmoVDB	20
3.3	Acted Emotional Speech Dynamic Database + EmovDB - Αποτελέσματα	21
3.3.1	Συμπεράσματα EmoVDB+AESDD	28
4	Τελικά Σχόλια - Επίλογος	28

Λίστα Εικόνων- Γραφημάτων

1	Περισσότερο σημαντικά χαρακτηριστικά(Acted Emotional Speech Dynamic Database)	6
2	Λιγότερο σημαντικά χαρακτηριστικά(Acted Emotional Speech Dynamic Database) . .	6
3	Ακρίβεια στα validation δεδομένα(Acted Emotional Speech Dynamic Database) . . .	7
4	Ακρίβεια στα δεδομένα εκπαίδευσης(Acted Emotional Speech Dynamic Database) . .	7
5	Learning Curve Logistic Regression(Acted Emotional Speech Dynamic Database) .	8
6	Learning Curve Decision Tree(Acted Emotional Speech Dynamic Database)	8
7	Learning Curve SVM (Acted Emotional Speech Dynamic Database)	9
8	Learning Curve KNN(Acted Emotional Speech Dynamic Database)	9
9	Learning Curve GNB (Acted Emotional Speech Dynamic Database)	9
10	Metrics Logistic Regression(Acted Emotional Speech Dynamic Database)	10
11	Metrics Decision Tree(Acted Emotional Speech Dynamic Database)	10
12	Metrics SVM (Acted Emotional Speech Dynamic Database)	11
13	Metrics KNN(Acted Emotional Speech Dynamic Database)	11
14	Metrics GNB (Acted Emotional Speech Dynamic Database)	11
15	Metrics Log.Reg. και SVM (Acted Emotional Speech Dynamic Database)	12
16	Περισσότερο σημαντικά χαρακτηριστικά(EmoVDB)	13
17	Λιγότερο σημαντικά χαρακτηριστικά(EmoVDB)	13
18	Ακρίβεια στα validation δεδομένα(EmoVDB Database)	14
19	Ακρίβεια στα δεδομένα εκπαίδευσης(EmoVDB)	15
20	Metrics Logistic Regression(EmoVDB)	16
21	Metrics Decision Tree(EmoVDB)	16
22	Metrics SVM (EmoVDB)	17
23	Metrics KNN(EmoVDB)	17
24	Metrics GNB (EmoVDB)	17
25	Metrics Logistic Regression(EmoVDB)	18
26	Metrics Decision Tree(EmoVDB)	18
27	Metrics SVM (EmoVDB)	19
28	Metrics KNN(EmoVDB)	19
29	Metrics GNB (EmoVDB)	19
30	Περισσότερο σημαντικά χαρακτηριστικά(EmoVDB+AESDD)	21
31	Λιγότερο σημαντικά χαρακτηριστικά(EmoVDB+AESDD)	21
32	Ακρίβεια στα validation δεδομένα(EmoVDB Database)	22
33	Ακρίβεια στα δεδομένα εκπαίδευσης(EmoVDB +AESDD)	23
34	Metrics Logistic Regression(EmoVDB+AESDD)	24
35	Metrics Decision Tree(EmoVDB+AESDD)	24
36	Metrics SVM (EmoVDB + AESDD)	25
37	Metrics KNN(EmoVDB+AESDD)	25
38	Metrics GNB (EmoVDB + AESDD)	25
39	Metrics Logistic Regression(EmoVDB +AESDD)	26
40	Metrics Decision Tree(EmoVDB + AESDD)	26
41	Metrics SVM (EmoVDB+AESDD)	27
42	Metrics KNN(EmoVDB +AESDD)	27
43	Metrics GNB (EmoVDB+AESDD)	27

Λίστα Πινάκων

1	Cross Validation Stats(AESDD)	7
2	Cross Validation Stats(EmoVDB)	14
3	Cross Validation Stats(EmoVDB+AESDD)	22

1 Εισαγωγή

Σκοπός της παρούσας εργασίας αποτελεί η χρήση τεχνικών μηχανικής μάθησης για την αναγνώριση συναισθήματος από ήχο. Παρακάτω παρουσιάζεται μια αναλυτική αναφορά της διαδικασίας που ακολουθήθηκε και των τεχνικών που χρησιμοποιήθηκαν. Επιπλέον δίνονται τα αποτελέσματα της ανάλυσης που προσέφερε κάθε μοντέλο που εκπαιδεύτηκε και γίνεται σύγκριση των μοντέλων ως προς τις επιδόσεις τους βάσει κάποιων μετρικών που αφορούν τόσο το *performance* στο σύνολο αλλά και άλλων που έχουν να κάνουν με επιμέρους κλάσεις (συναισθημάτα) κάθε φορά .

2 Σετ Δεδομένων

Τα σετ των δεδομένων που χρησιμοποιήθηκαν για την ανάλυση που θα ακολουθήσει είναι τα:

- Acted Emotional Speech Dynamic Database
(<https://mega.nz/folder/0ShVXY7C#-73kVoK050jTPEA95UUvMw>)
- EmoVDB
(<https://openslr.org/115/>)

Το κάθε σετ δεδομένων χρησιμοποιήθηκε για εκπαίδευση μοντέλων ξεχωριστά και στην συνέχεια συνενώθηκαν μέρη από το καθένα και εκπαιδεύτηκαν εκ νέου μοντέλα και αξιολογήθηκαν ενώ επίσης έγινε σύγκριση μεταξύ των αποτελεσμάτων.

1 Το πρώτο σετ δεδομένων αποτελείται από ηχητικά αρχεία (τύπου .wav) ηχογραφήσεων ηθοποιών που υποδύονται συναισθήματα στην ελληνική γλώσσα. Οι κλάσεις συναισθημάτων που έχουμε σε αυτό το σετ είναι :

- Χαρά (*happiness*)
- Λύπη (*sadness*)
- Απέχθεια (*disgust*)
- Φόβος (*fear*)
- Θυμός (*anger*)

Περιέχει 120 εγγραφές για κάθε κλάση (συνεπώς 600 εγγραφές συνολικά) και χρησιμοποιήθηκε ολόκληρο στο πρώτο κομμάτι της εργασίας. Οι ομολητές ήταν αρκετοί σε αριθμό (δεδομένου του μικρού μεγέθους του σετ δεδομένων) και επομένως δεν έγινε διαχωρισμός τους στην συνέχεια όσον αφορά την διάσπαση του σετ σε training set και test set (δηλαδή κάποιον να μπουν στο σετ εκπαίδευσης ενώ άλλοι να μπουν στο τεστ σετ). Οφείλουμε να αναφέρουμε εκ των προτέρων ότι το μικρό αυτό μέγεθος σε συνδυασμό με την πολυπλοκότητα του classification προβλήματος αναμένεται να μας οδηγήσουν σε ζητήματα που θα παρουσιάζονται στην συνέχεια της εργασίας.

2 Το δεύτερο σετ δεδομένων αποτελείται από ηχητικά αρχεία (τύπου .wav) ηχογραφήσεων 4 ατόμων που υποδύονται συναισθήματα στην αγγλική γλώσσα. Οι κλάσεις συναισθημάτων που έχουμε σε αυτό το σετ είναι :

- Ενθουσιασμός (*amusement*)
- Θυμός (*anger*)
- Απέχθεια (*disgust*)
- Ουδέτερη Κατάσταση (*Neutral*)
- Νύστα (*sleepy*)

Περιέχει περίπου 1200 εγγραφές για κάθε κλάση (περί τις 6500 εγγραφές συνολικά) και χρησιμοποιήθηκε μέρος του σετ αυτού. Πιο συγκεκριμένα για το δεύτερο μέρος της εργασίας που αφορά το classification συναισθήματος αυτού μόνο του σετ περιορίστηκαμε στις πρώτες 4 κλάσεις και συνεπώς είχαμε περίπου στις 5000 εγγραφές, ενώ για το δεύτερο μέρος της εργασίας που αφορά την συννένωση των δύο σετ χρησιμοποιήθηκαν οι κλάσεις ' Θυμός απέχθεια, χαρά ' από το πρώτο και ' Θυμός απέχθεια, ενθουσιασμός ' από το δεύτερο σετ και έγινε η θεώρηση ότι οι κλάσεις ενθουσιασμός και χαρά είναι ισοδύναμες.

3 Ανάλυση Διαδικασιών και Αποτελέσματα

Οι αλγόριθμοι που χρησιμοποιήθηκαν για εκπαίδευση είναι οι εξής:

- Logistic Regression
- Decision Tree
- Support Vector Machine
- K Nearest Neighbors
- Gaussian Naive Bayes

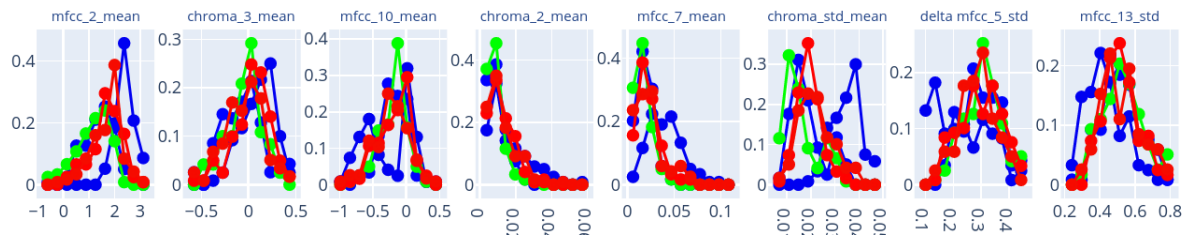
Τα βήματα που ακολουθήθηκαν είναι τα εξής:

1. Εξαγωγή Χαρακτηριστικών και Κανονικοποίηση (Feature Extraction) με pyAudioAnalysis .
2. Επιλογή Χαρακτηριστικών (Feature Selection) με την βοήθεια ενός meta-transformer (logistic regression ή support vector machine) .
3. Διαχωρισμός του σετ δεδομένων σε train set και test set με την χρήση της train _ test _ split της scikit-learn σε ποσοστά 80% και 20%. Να σημειωθεί ότι ο διαχωρισμός έγινε με τέτοιο τρόπο ώστε κάθε κλάση να εκπροσωπείται κατά το ίδιο ποσοστό στα δύο σετ.
4. Hyper-Parameter Tuning (μέσω grid search).

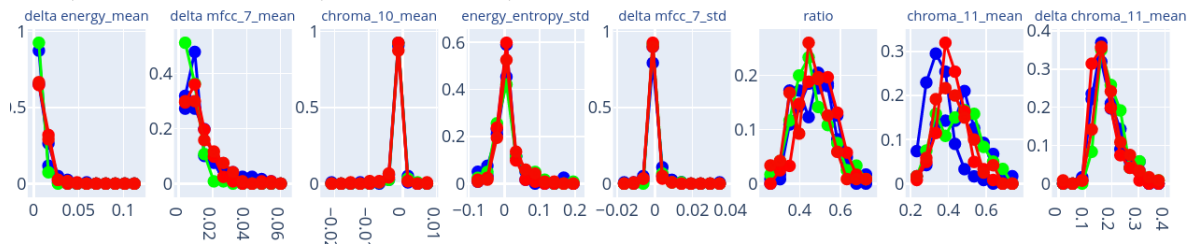
5. Cross-Validation χρησιμοποιώντας την Stratifiedkfold της scikit-learn με 10 splits .
6. Αξιολόγηση της απόδοσης του κάθε μοντέλου πάνω στα τεστ δεδομένα βάση μετρικών(accuracy, f1 score, confusion matrix ect.).

3.1 Acted Emotional Speech Dynamic Database - Αποτελέσματα

Σε πρώτη φάση χρησιμοποιήθηκε ολόκληρο το σετ δεδομένων Acted Emotional Speech Dynamic Database και ύστερα από επιλογή χαρακτηριστικών καταλήξαμε με τα 58 πιο σημαντικά features. Στην συνέχεια για να έχουμε μια αίσθηση του κατά πόσο ικανά είναι τα χαρακτηριστικά αυτά να διαχωρίσουν τα δεδομένα μας επιλέξαμε τα 8 πιο σημαντικά και 8 λιγότερο σημαντικά και δημιουργήθηκαν γραφήματα της κατανομής των τιμών τους για κάθε μία από τις κλάσεις μας.



Σχήμα 1: Περισσότερο σημαντικά χαρακτηριστικά(Acted Emotional Speech Dynamic Database)



Σχήμα 2: Λιγότερο σημαντικά χαρακτηριστικά(Acted Emotional Speech Dynamic Database)

Από τα παραπάνω γραφήματα (ιδίως των περισσότερο σημαντικών χαρακτηριστικών) μπορούμε ήδη να παρατηρήσουμε ότι υπάρχει μία κλάση η 'θλίψη', η οποία διαφοροποιείται αρκετά από τις υπόλοιπες ειδικά ως προς το χαρακτηριστικό mfcc2mean και chroma.std.mean . Αυτό το αποτέλεσμα όπως θα διαπιστώσουμε παρατηρείται και κατά την διαδικασία της τελικής αξιολόγησης.

Αποτελέσματα Hyper parameter tuning

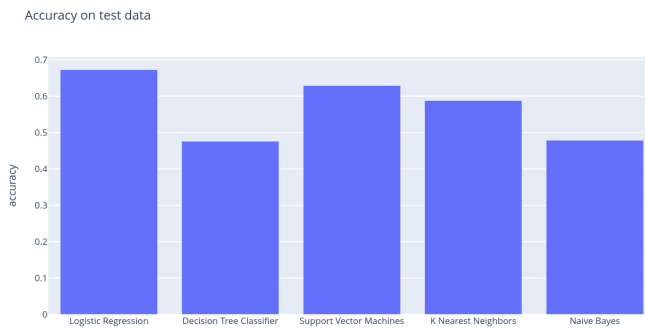
- | | |
|---------------------------|--|
| 1) Logistic Regression | — > Best parameters— > $\{C = 1, 'penalty' = 'l2'\}$ |
| 2) Support Vector machine | — > Best parameters— > $\{C = 0.5, kernel = 'linear', gamma = 0.001\}$ |
| 3) Decision Tree | — > Best parameters— > $\{'max_depth' : 20, 'max_features' : 'sqrt'\}$ |
| 4) Knn | — > Best parameters— > $\{'n_neighbors' : 7\}$ |
| 5) Naive Bayes | — > Best parameters— > $\{'var_smoothing' : 1e - 15\}$ |

Αποτελέσματα Cross Validation

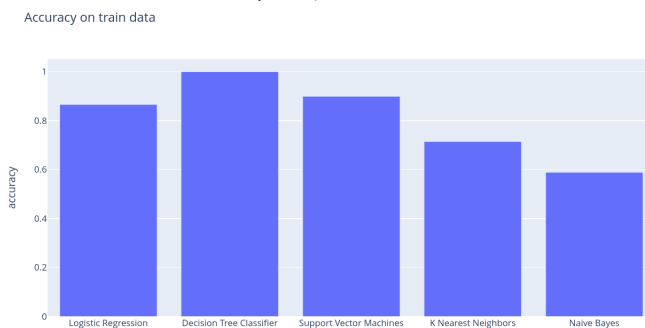
Table 1: Cross Validation Stats(AESDD)

	Avg Accurscy	Avg F1 score
Logistic Regression	67.28	66.78
Decision Tree	47.62	46.72
SVM	62.93	62.18
KNN	58.80	57.45
GNB	47.85	45.79

Και τα αντίστοιχα γραφήματα που δείχνουν το μέσο accuracy σε δεδομένα του σετ εκπαίδευσης και του validation set παρουσιάζονται παρακάτω:



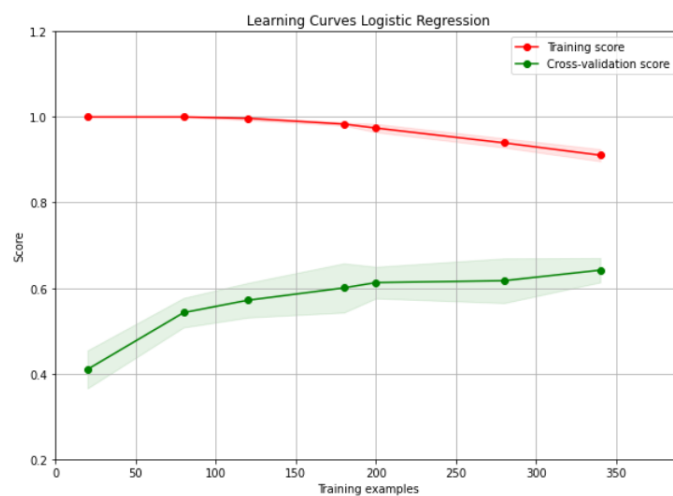
Σχήμα 3: Ακρίβεια στα validation δεδομένα(Acted Emotional Speech Dynamic Database)



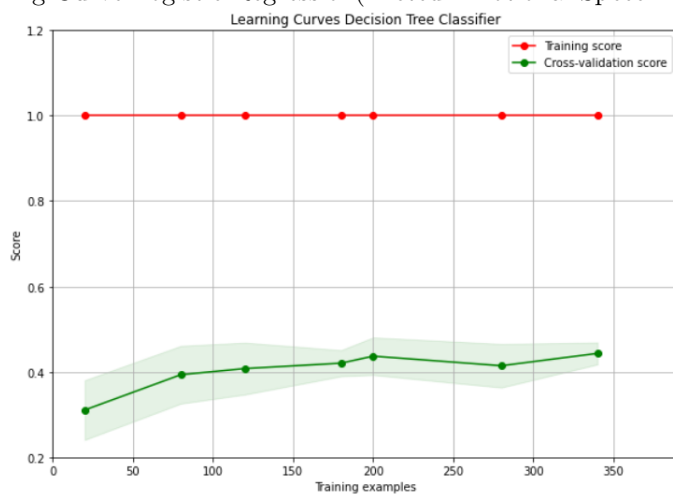
Σχήμα 4: Ακρίβεια στα δεδομένα εκπαίδευσης(Acted Emotional Speech Dynamic Database)

Το μέσο accuracy και f1 όπως παρουσιάζεται στα validation δεδομένα είναι αρκετά άσχημο. Παρόλα αυτά την καλύτερη απόδοση φαίνεται να έχουν το Logistic Regression και ο SVM ενώ το Decision Tree φαίνεται πως είναι overfitted καθώς η ακριβειά του στο σετ εκπαίδευσης είναι άριστη ενώ δεν μπορεί να κάνει γενίκευση και να έχει αντίστοιχα καλή ακρίβεια στα validation δεδομένα. Για λόγους πληρότητας και ενδεικτικά με το ποια μοντέλα είναι overfitted παρουσιάζονται και τα learning curves για τον

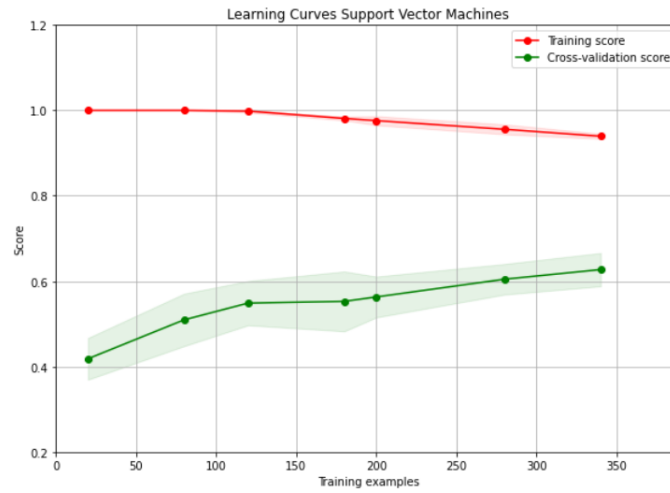
καθένα.



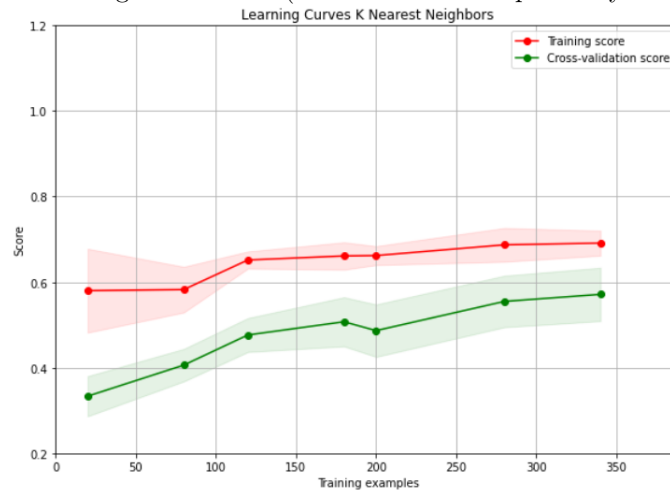
Σχήμα 5: Learning Curve Logistic Regression(Acted Emotional Speech Dynamic Database)



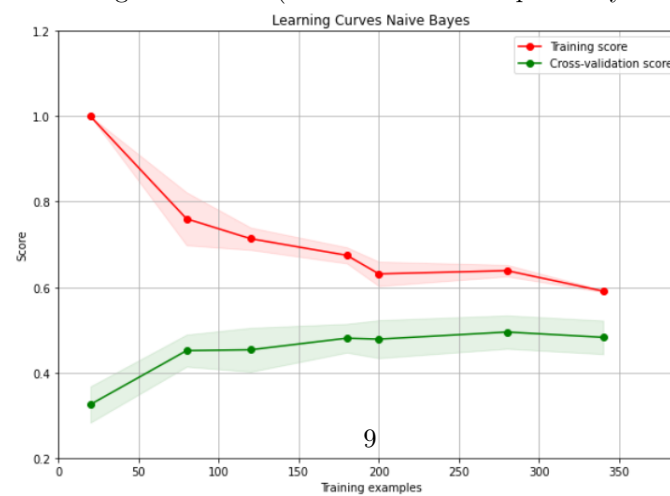
Σχήμα 6: Learning Curve Decision Tree(Acted Emotional Speech Dynamic Database)



Σχήμα 7: Learning Curve SVM (Acted Emotional Speech Dynamic Database)



Σχήμα 8: Learning Curve KNN(Acted Emotional Speech Dynamic Database)



Σχήμα 9: Learning Curve GNB (Acted Emotional Speech Dynamic Database)

Αποτελέσματα Evaluation στα τεστ δεδομένα

Logistic Regression Test Accuracy: 64.46%

Classification Report:

	precision	recall	f1-score	support
0.0	0.81	0.92	0.86	24
1.0	0.46	0.50	0.48	24
2.0	0.70	0.67	0.68	24
3.0	0.77	0.68	0.72	25
4.0	0.48	0.46	0.47	24
accuracy			0.64	121
macro avg	0.64	0.64	0.64	121
weighted avg	0.65	0.64	0.64	121

Confusion Matrix:

```
[[22  0  0  2  0]
 [ 1 12  3  0  8]
 [ 1  5 16  0  2]
 [ 0  4  2 17  2]
 [ 3  5  2  3 11]]
```

Σχήμα 10: Metrics Logistic Regression(Acted Emotional Speech Dynamic Database)

Decision Tree Classifier Test Accuracy: 49.59%

Classification Report:

	precision	recall	f1-score	support
0.0	0.76	0.79	0.78	24
1.0	0.35	0.38	0.36	24
2.0	0.54	0.54	0.54	24
3.0	0.41	0.44	0.42	25
4.0	0.42	0.33	0.37	24
accuracy			0.50	121
macro avg	0.50	0.50	0.49	121
weighted avg	0.49	0.50	0.49	121

Confusion Matrix:

```
[[19  2  0  1  2]
 [ 1  9  6  5  3]
 [ 0  8 13  3  0]
 [ 3  3  2 11  6]
 [ 2  4  3  7  8]]
```

Σχήμα 11: Metrics Decision Tree(Acted Emotional Speech Dynamic Database)

Support Vector Machines Test Accuracy: 59.50%

Classification Report:

	precision	recall	f1-score	support
0.0	0.81	0.88	0.84	24
1.0	0.41	0.46	0.43	24
2.0	0.70	0.67	0.68	24
3.0	0.63	0.48	0.55	25
4.0	0.46	0.50	0.48	24
accuracy			0.60	121
macro avg	0.60	0.60	0.60	121
weighted avg	0.60	0.60	0.60	121

Confusion Matrix:

```
[[21  0  0  2  1]
 [ 1 11  4  2  6]
 [ 0  6 16  0  2]
 [ 2  5  1 12  5]
 [ 2  5  2  3 12]]
```

Σχήμα 12: Metrics SVM (Acted Emotional Speech Dynamic Database)

Naive Bayes Test Accuracy: 48.76%

Classification Report:

	precision	recall	f1-score	support
0.0	0.57	0.83	0.68	24
1.0	0.48	0.50	0.49	24
2.0	0.78	0.29	0.42	24
3.0	0.43	0.48	0.45	25
4.0	0.33	0.33	0.33	24
accuracy			0.49	121
macro avg	0.52	0.49	0.48	121
weighted avg	0.52	0.49	0.48	121

Confusion Matrix:

```
[[20  0  0  3  1]
 [ 5 12  0  3  4]
 [ 2 11  7  2  2]
 [ 3  0  1 12  9]
 [ 5  2  1  8  8]]
```

Σχήμα 13: Metrics KNN(Acted Emotional Speech Dynamic Database)

Naive Bayes Test Accuracy: 48.76%

Classification Report:

	precision	recall	f1-score	support
0.0	0.57	0.83	0.68	24
1.0	0.48	0.50	0.49	24
2.0	0.78	0.29	0.42	24
3.0	0.43	0.48	0.45	25
4.0	0.33	0.33	0.33	24
accuracy			0.49	121
macro avg	0.52	0.49	0.48	121
weighted avg	0.52	0.49	0.48	121

Confusion Matrix:

```
[[20  0  0  3  1]
 [ 5 12  0  3  4]
 [ 2 11  7  2  2]
 [ 3  0  1 12  9]
 [ 5  2  1  8  8]]
```

Σχήμα 14: Metrics GNB (Acted Emotional Speech Dynamic Database)

3.1.1 Συμπεράσματα Acted Emotional Speech Dynamic Database

Βάση των αποτελεσμάτων που παρουσιάστηκαν παραπάνω μπορούμε να εξάγουμε τα εξής συμπεράσματα όσον αφορά την καταλληλότητα του σετ δεδομένων μας, τους αλγόριθμους μάθησης που χρησιμοποιήθηκαν καθώς και την επιτυχία ή όχι του στόχου κατηγοριοποίησης των συναισθημάτων:

1. Το σετ δεδομένων μας αποδείχθηκε πολύ μικρό σε σχέση με τα απαραίτητα χαρακτηριστικά που επιλέχθηκαν (58 από τα 138). Έτσι ως αποτέλεσμα παρουσιάζεται το overfitting που συνέβη ιδίως στην περίπτωση του δέντρου αποφάσεων γεγονός που αποτυπώνεται και στο αντίστοιχο learning curve .
2. Οι αλγόριθμοι με την μεγαλύτερη επιτυχία όσον αφορά το accuracy φαίνεται να είναι το logistic regression και ο SVM .
3. Το πιο σημαντικό ίσως αποτέλεσμα που αποτυπώνεται στο confusion matrix και το recall όλων των αλγορίθμων είναι το ότι η κλάση 0 που αντιστοιχεί στη λύπη εντοπίζεται με επιτυχία σε πολύ μεγάλο ποσοστό γεγονός που επιβεβαιώνει την διαίσθησή που αναπτύξαμε κατά την οπτικοποίηση των σημαντικότερων χαρακτηριστικών (όπως αναφέρθηκε παραπάνω). Το αποτέλεσμα αυτό εξακριβώνεται και μέσω ενός one vs the rest classifier όπου εκπαιδεύτηκαν οι δύο πιο αποδοτικοί αλγόριθμοι (logistic regression, SVM) και τα αποτελέσματα φαίνονται παρακάτω.
4. Στα learning curves ειδικά του SVM και του logistic regression φαίνεται πως με την προσθήκη δεδομένων υπάρχει η τάση αύξησης της ακρίβειας στο validation .

```
Logistic Regression Test Accuracy: 96.69%
Classification Report:
      precision    recall  f1-score   support

    0.0         0.92     0.92     0.92         24
    1.0         0.98     0.98     0.98         97

 accuracy         0.97         121
  macro avg       0.95     0.95     0.95         121
 weighted avg     0.97     0.97     0.97         121

Confusion Matrix:
[[22  2]
 [ 2 95]]
Support Vector Machines Test Accuracy: 96.69%
Classification Report:
      precision    recall  f1-score   support

    0.0         0.92     0.92     0.92         24
    1.0         0.98     0.98     0.98         97

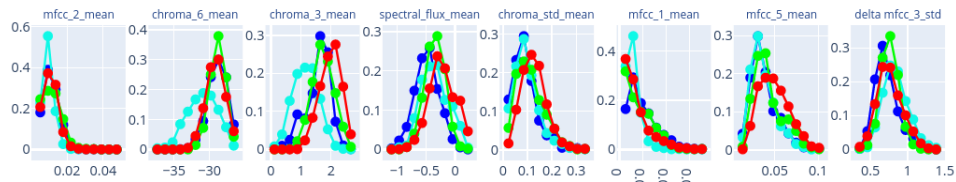
 accuracy         0.97         121
  macro avg       0.95     0.95     0.95         121
 weighted avg     0.97     0.97     0.97         121

Confusion Matrix:
[[22  2]
 [ 2 95]]
```

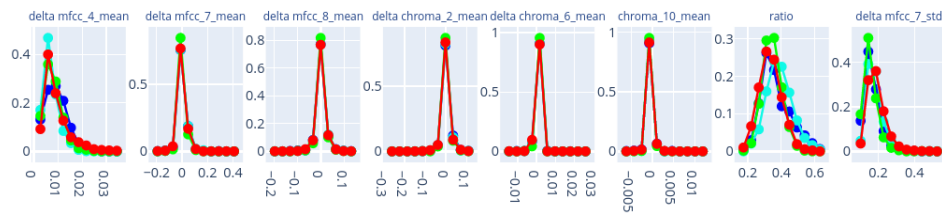
Σχήμα 15: Metrics Log.Reg. και SVM (Acted Emotional Speech Dynamic Database)

3.2 EmoVDB - Αποτελέσματα

Στην συνέχεια χρησιμοποιήθηκε μέρος του σετ δεδομένων EmoVDB και συγκεκριμένα οι κλάσεις amusement, anger και disgust . Η διαδικασία που ακολουθήθηκε είναι εντελώς παρόμοια με πριν. Επομένως έγινε επιλογή χαρακτηριστικών και πιο συγκεκριμένα διαλέχθηκαν τα 52 πιο σημαντικά. Ενδεικτικά παρουσιάζονται ξανά οπτικοποιημένες οι κατανομές για κάθε κλάση των 8 πιο σημαντικών και 8 λιγότερο σημαντικών.



Σχήμα 16: Περισσότερο σημαντικά χαρακτηριστικά(EmoVDB)



Σχήμα 17: Λιγότερο σημαντικά χαρακτηριστικά(EmoVDB)

Από τα παραπάνω γραφήματα (ιδίως των περισσότερο σημαντικών χαρακτηριστικών) μπορούμε ήδη να παρατηρήσουμε ότι υπάρχει μία κλάση ο 'θυμός'(γαλάζιο χρώμα), η οποία διαφοροποιείται αρκετά από τις υπόλοιπες ειδικά ως προς το χαρακτηριστικό mfcc2mean και chroma.3.mean . Αυτό το αποτέλεσμα όπως θα διαπιστώσουμε παρατηρείται και κατά την διαδικασία της τελικής αξιολόγησης.

Αποτελέσματα Hyper parameter tuning

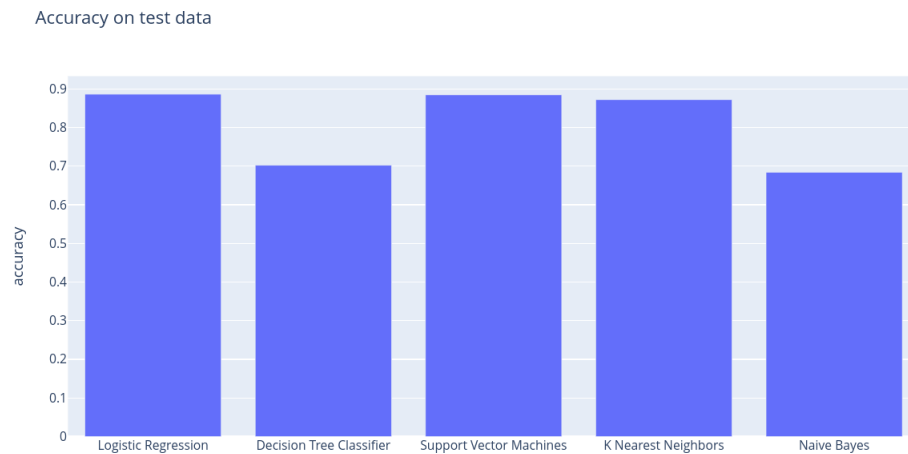
- 1) Logistic Regression — > Best parameters — > $\{C = 1, 'penalty' = 'l2'\}$
- 2) Support Vector machine — > Best parameters — > $\{C = 0.1, kernel = 'linear', gamma = 0.001\}$
- 3) Decision Tree — > Best parameters — > $\{'max_depth' : 10\}$
- 4) Knn — > Best parameters — > $\{'n_neighbors' : 7\}$
- 5) Naive Bayes — > Best parameters — > $\{'var_smoothing' : 0.00075\}$

Αποτελέσματα Cross Validation

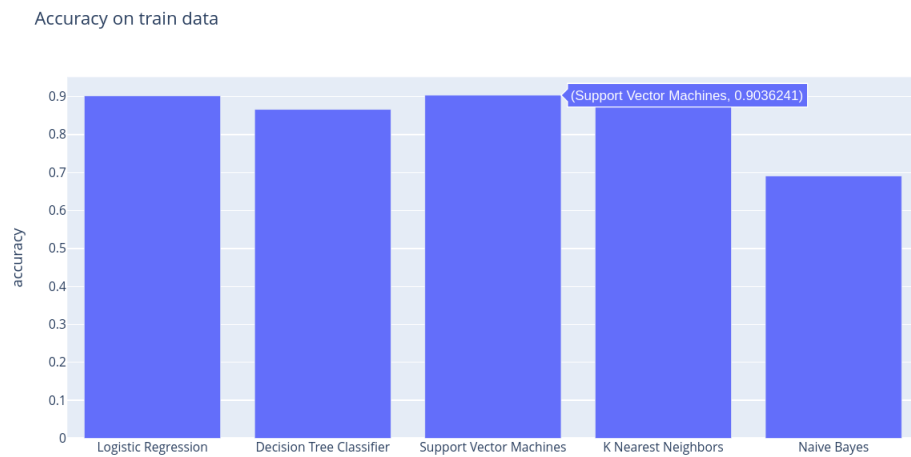
Table 2: Cross Validation Stats(EmoVDB)

	Avg Accurscy	Avg F1 score
Logistic Regression	88.63	88
Decision Tree	70.22	70.2
SVM	88.46	88.46
KNN	87.21	87.24
GNB	68.43	67.46

Και τα αντίστοιχα γραφήματα που δείχνουν το μέσο accuracy σε δεδομένα του σετ εκπαίδευσης και του validation set παρουσιάζονται παρακάτω:

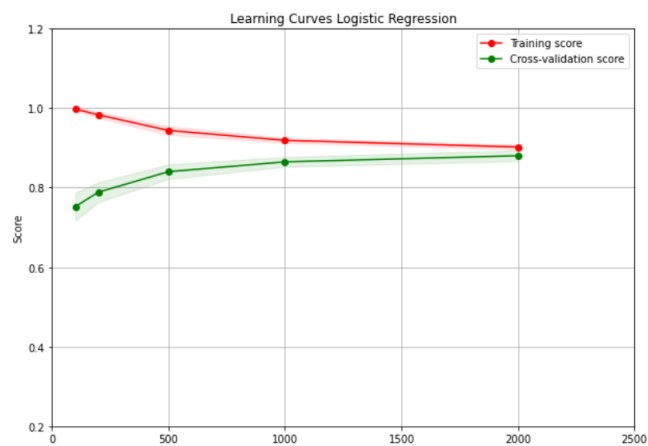


Σχήμα 18: Ακρίβεια στα validation δεδομένα(EmoVDB Database)

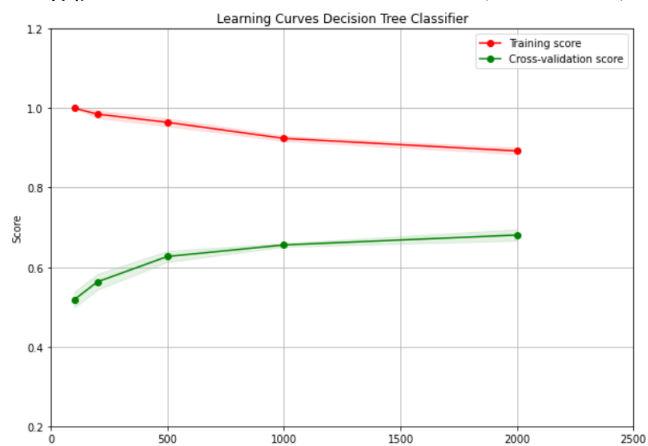


Σχήμα 19: Ακρίβεια στα δεδομένα εκπαίδευσης(EmoVDB)

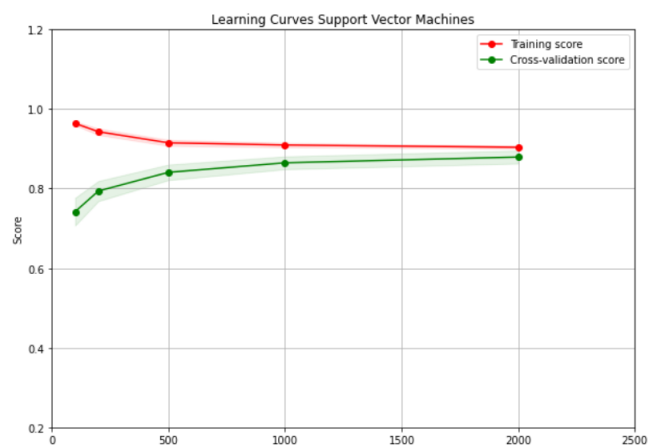
Τα αντίστοιχα learning curves παρουσιάζονται στις επόμενες σελίδες και όπως φαίνεται δεν υπάρχει το πρόβλημα που υπήρξε στην περίπτωση του προηγούμενου σετ δεδομένων (όχι overfitted μοντέλα).



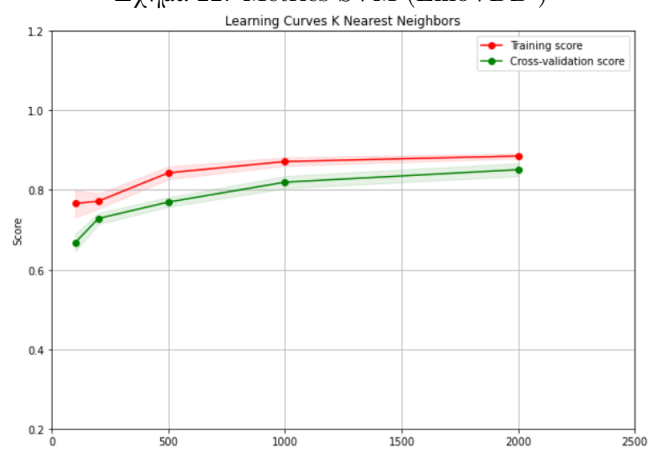
Σχήμα 20: Metrics Logistic Regression(EmoVDB)



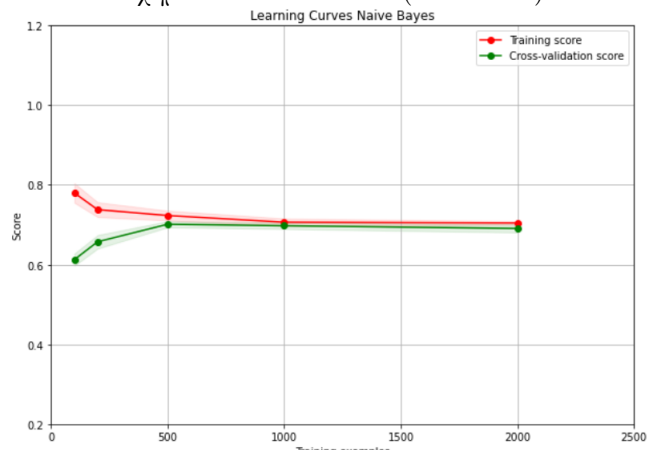
Σχήμα 21: Metrics Decision Tree(EmoVDB)



Σχήμα 22: Metrics SVM (EmoVDB)



Σχήμα 23: Metrics KNN(EmoVDB)



Σχήμα 24: Metrics GNB (EmoVDB)

Αποτελέσματα Evaluation στα τεστ δεδομένα

Logistic Regression Test Accuracy: 88.61%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.82	0.87	0.84	230
1.0	0.94	0.93	0.94	254
2.0	0.83	0.82	0.82	204
3.0	0.93	0.91	0.92	313
accuracy			0.89	1001
macro avg	0.88	0.88	0.88	1001
weighted avg	0.89	0.89	0.89	1001

Confusion Matrix:

```
[[199 10 18 3]
 [ 14 237 2 1]
 [ 16 4 167 17]
 [ 13 1 15 284]]
```

Σχήμα 25: Metrics Logistic Regression(EmoVDB)

Decision Tree Classifier Test Accuracy: 71.23%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.66	0.63	0.64	230
1.0	0.79	0.78	0.78	254
2.0	0.59	0.65	0.62	204
3.0	0.77	0.77	0.77	313
accuracy			0.71	1001
macro avg	0.70	0.70	0.70	1001
weighted avg	0.71	0.71	0.71	1001

Confusion Matrix:

```
[[144 27 40 19]
 [ 34 197 11 12]
 [ 20 11 132 41]
 [ 20 13 40 240]]
```

Σχήμα 26: Metrics Decision Tree(EmoVDB)

Support Vector Machines Test Accuracy: 88.91%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.85	0.86	0.85	230
1.0	0.93	0.96	0.95	254
2.0	0.83	0.80	0.82	204
3.0	0.92	0.91	0.92	313
accuracy			0.89	1001
macro avg	0.88	0.88	0.88	1001
weighted avg	0.89	0.89	0.89	1001

Confusion Matrix:

```
[[197 13 15 5]
 [ 8 243 2 1]
 [ 17 3 164 20]
 [ 10 1 16 286]]
```

Σχήμα 27: Metrics SVM (EmoVDB)

K Nearest Neighbors Test Accuracy: 85.61%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.82	0.86	0.84	230
1.0	0.93	0.83	0.88	254
2.0	0.76	0.90	0.83	204
3.0	0.91	0.85	0.88	313
accuracy			0.86	1001
macro avg	0.85	0.86	0.85	1001
weighted avg	0.86	0.86	0.86	1001

Confusion Matrix:

```
[[197 11 18 4]
 [ 30 211 5 8]
 [ 3 2 184 15]
 [ 11 2 35 265]]
```

Σχήμα 28: Metrics KNN(EmoVDB)

Naive Bayes Test Accuracy: 69.83%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.66	0.47	0.54	230
1.0	0.75	0.83	0.79	254
2.0	0.57	0.77	0.66	204
3.0	0.79	0.71	0.75	313
accuracy			0.70	1001
macro avg	0.69	0.70	0.69	1001
weighted avg	0.71	0.70	0.69	1001

Confusion Matrix:

```
[[107 44 54 25]
 [ 28 212 6 8]
 [ 11 10 158 25]
 [ 17 17 57 222]]
```

Σχήμα 29: Metrics GNB (EmoVDB)

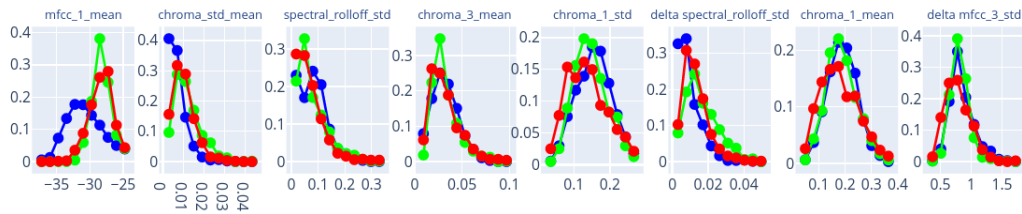
3.2.1 Συμπεράσματα EmoVDB

Βάση των αποτελεσμάτων που παρουσιάστηκαν παραπάνω μπορούμε να εξάγουμε τα εξής συμπεράσματα όσον αφορά την καταλληλότητα του σετ δεδομένων μας, τους αλγορίθμους μάθησης που χρησιμοποιήθηκαν καθώς και την επιτυχία ή όχι του στόχου κατηγοριοποίησης των συναισθημάτων:

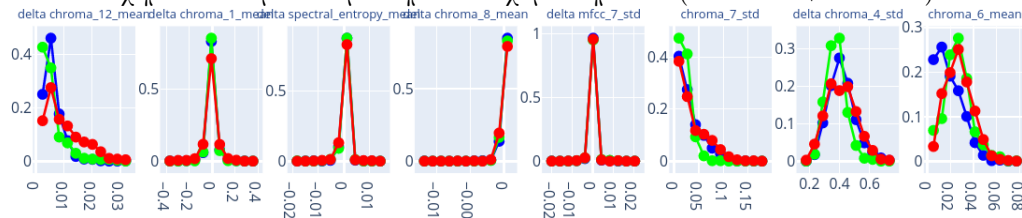
1. Η απόδοση όλων των αλγορίθμων είναι αρκετά ικανοποιητική με καλύτερη εκείνη του SVM και του logistic regression γεγονός που αποτυπώνεται σε όλα τα metrics που παρουσιάστηκαν.
2. Ο όγκος του σετ δεδομένων που είχαμε στην διαθεσή μας είναι αρκετά μεγαλύτερος και επομένως επαρκής για να αποφευχθεί το overfitting όπως φαίνεται από τα learning curves . Ακόμη και στην περίπτωση του δέντρου αποφάσεων που έχει την τάση να κάνει overfit στην περιπτωσή μας δεν έχει συμβεί αυτό αλλά όπως φαίνεται υπάρχει η τάση εάν προστίθονταν επιπλέον δεδομένα για αύξηση της ακρίβειας στο validation σετ.
3. Το πιο σημαντικό ίσως αποτέλεσμα που αποτυπώνεται στο confusion matrix και το recall όλων των αλγορίθμων είναι το ότι η κλάση 1 που αντιστοιχεί στον θυμό και η κλάση 3 που αντιστοιχεί στην ουδέτερη κατάσταση εντοπίζονται με επιτυχία σε πολύ μεγάλο ποσοστό γεγονός που επιβεβαιώνει την διαισθησή που αναπτύξαμε κατά την οπτικοποίηση των σημαντικότερων χαρακτηριστικών (όπως αναφέρθηκε παραπάνω στην παρουσίαση των αποτελεσμάτων).

3.3 Acted Emotional Speech Dynamic Database + EmovDB - Αποτελέσματα

Ως τελευταίος πειραματισμός χρησιμοποιήθηκαν μέρη και των δύο σετ δεδομένων. Χρησιμοποιήθηκαν οι κλάσεις Θυμός απέχθεια, χαρά ' από το πρώτο σετ από το δεύτερο και ' Θυμός απέχθεια, ενθουσιασμός ' όπου έγινε η θεώρηση ότι οι κλάσεις χαρά και ενθουσιασμός είναι ταυτόσημες. Ακολουθώντας τα ίδια βήματα με προηγουμένως έχουμε αρχικά την οπτικοποίηση της κατανομής των κατανομών για κάθε κλάση των 8 πιο σημαντικών και 8 λιγότερο σημαντικών χαρακτηριστικών από τα συνολικά 47 που επιλέχθηκαν.



Σχήμα 30: Περισσότερο σημαντικά χαρακτηριστικά(EmoVDB+AESDD)



Σχήμα 31: Λιγότερο σημαντικά χαρακτηριστικά(EmoVDB+AESDD)

Αποτελέσματα Hyper parameter tuning

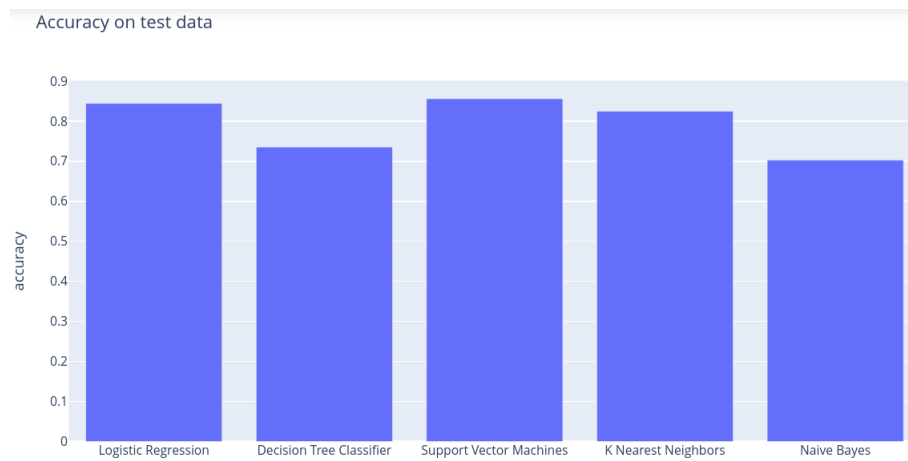
- 1) Logistic Regression —> Best parameters —> $\{C = 1, 'penalty' = 'l2'\}$
- 2) Support Vector machine —> Best parameters —> $\{C = 0.01, kernel = 'linear', gamma = 0.001\}$
- 3) Decision Tree —> Best parameters —> $\{'max_depth' : 8\}$
- 4) Knn —> Best parameters —> $\{'n_neighbors' : 13\}$
- 5) Naive Bayes —> Best parameters —> $\{'var_smoothing' : 1e - 15\}$

Αποτελέσματα Cross Validation

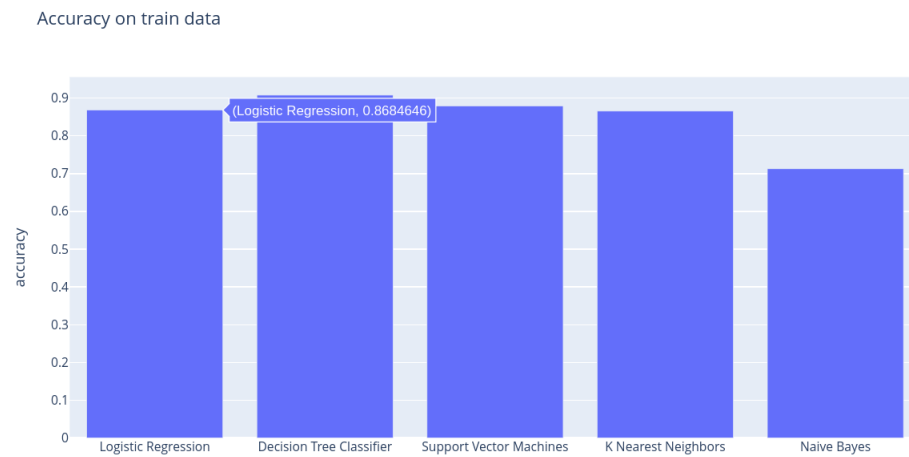
Table 3: Cross Validation Stats(EmoVDB+AESDD)

	Avg Accurscy	Avg F1 score
Logistic Regression	84.45	84.44
Decision Tree	73.57	73.61
SVM	85.63	85.60
KNN	82.52	82.45
GNB	70.31	69.74

Και τα αντίστοιχα γραφήματα που δείχνουν το μέσο accuracy σε δεδομένα του σετ εκπαίδευσης και του validation set παρουσιάζονται παρακάτω:

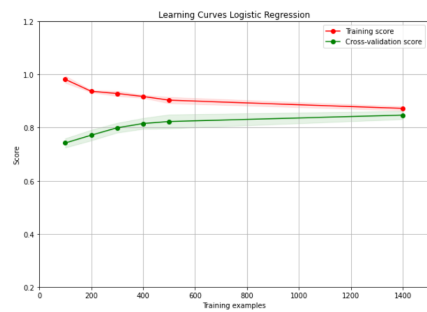


Σχήμα 32: Ακρίβεια στα validation δεδομένα(EmoVDB Database)

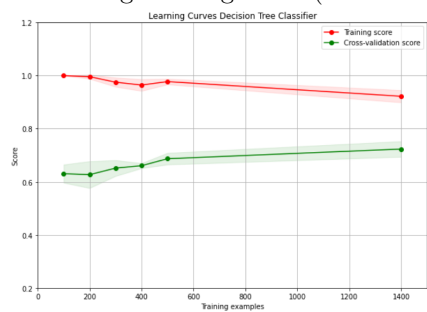


Σχήμα 33: Ακρίβεια στα δεδομένα εκπαίδευσης(EmoVDB + AESDD)

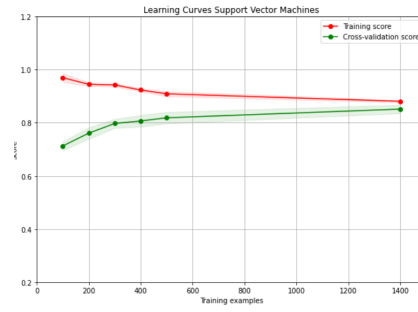
Τα αντίστοιχα learning curves παρουσιάζονται στις επόμενες σελίδες και όπως φαίνεται δεν υπάρχει το πρόβλημα που υπήρξε στην περίπτωση του πρώτου σετ δεδομένων (όχι overfitted μοντέλα).



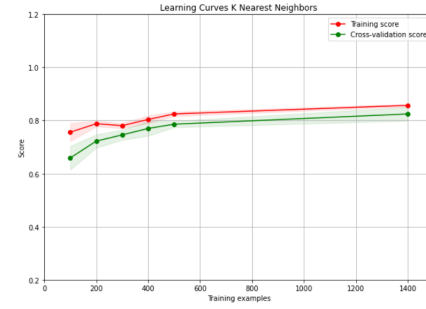
Σχήμα 34: Metrics Logistic Regression(EmoVDB+AESDD)



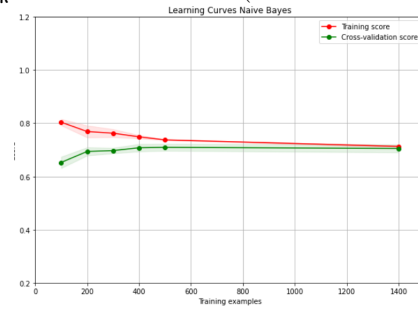
Σχήμα 35: Metrics Decision Tree(EmoVDB+AESDD)



Σχήμα 36: Metrics SVM (EmoVDB + AESDD)



Σχήμα 37: Metrics KNN(EmoVDB+AESDD)



Σχήμα 38: Metrics GNB (EmoVDB + AESDD)

Αποτελέσματα Evaluation στα τεστ δεδομένα

Logistic Regression Test Accuracy: 84.58%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.88	0.86	0.87	184
1.0	0.83	0.87	0.85	129
2.0	0.82	0.81	0.81	154
accuracy			0.85	467
macro avg	0.84	0.85	0.84	467
weighted avg	0.85	0.85	0.85	467

Confusion Matrix:

```
[[159 11 14]
 [ 4 112 13]
 [ 18 12 124]]
```

Σχήμα 39: Metrics Logistic Regression(EmoVDB +AESDD)

Decision Tree Classifier Test Accuracy: 74.52%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.78	0.74	0.76	184
1.0	0.71	0.76	0.73	129
2.0	0.74	0.74	0.74	154
accuracy			0.75	467
macro avg	0.74	0.75	0.74	467
weighted avg	0.75	0.75	0.75	467

Confusion Matrix:

```
[[136 21 27]
 [ 17 98 14]
 [ 21 19 114]]
```

Support Vector Machines Test Accuracy: 84.15%

Σχήμα 40: Metrics Decision Tree(EmoVDB + AESDD)

Support Vector Machines Test Accuracy: 84.15%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.86	0.85	0.86	184
1.0	0.81	0.91	0.86	129
2.0	0.84	0.77	0.81	154
accuracy			0.84	467
macro avg	0.84	0.84	0.84	467
weighted avg	0.84	0.84	0.84	467

Confusion Matrix:

```
[[157 12 15]
 [ 5 117  7]
 [ 20 15 119]]
```

K Nearest Neighbors Test Accuracy: 84.15%

Σχήμα 41: Metrics SVM (EmoVDB+AESDD)

K Nearest Neighbors Test Accuracy: 84.15%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.84	0.85	0.84	184
1.0	0.85	0.87	0.86	129
2.0	0.83	0.81	0.82	154
accuracy			0.84	467
macro avg	0.84	0.84	0.84	467
weighted avg	0.84	0.84	0.84	467

Confusion Matrix:

```
[[156 10 18]
 [ 10 112  7]
 [ 20  9 125]]
```

Σχήμα 42: Metrics KNN(EmoVDB +AESDD)

Naive Bayes Test Accuracy: 72.16%

Classification Metrics:

	precision	recall	f1-score	support
0.0	0.71	0.82	0.76	184
1.0	0.72	0.75	0.73	129
2.0	0.75	0.58	0.66	154
accuracy			0.72	467
macro avg	0.73	0.72	0.72	467
weighted avg	0.72	0.72	0.72	467

Confusion Matrix:

```
[[150 18 16]
 [ 18 97 14]
 [ 44 20 90]]
```

Σχήμα 43: Metrics GNB (EmoVDB+AESDD)

3.3.1 Συμπεράσματα EmoVDB+AESDD

Βάση των αποτελεσμάτων που παρουσιάστηκαν παραπάνω μπορούμε να εξάγουμε τα εξής συμπεράσματα όσον αφορά την καταλληλότητα του σετ δεδομένων μας, τους αλγορίθμους μάθησης που χρησιμοποιήθηκαν καθώς και την επιτυχία ή όχι του στόχου κατηγοριοποίησης των συναισθημάτων:

1. Η απόδοση των αλγορίθμων είναι αισθητά καλύτερη σε σχέση με το πρώτο σετ που χρησιμοποιήθηκε και ελαφρώς καλύτερη και από την περίπτωση που χρησιμοποιήσαμε το EmoVDB μόνο. Βέβαια αυτό δεν είναι αντιπροσωπευτικό επειδή σε αυτό το πείραμα οι κλάσεις που είχαμε ήταν 3 και συνεπώς ήταν πιο εύκολη η εκπαίδευση των αλγορίθμων.
2. Εδώ σε αντίθεση με τους προηγούμενους πειραματισμούς όλες οι κλάσεις διαφοροποιούνται εξίσου καλά από τις υπόλοιπες όπως αποτυπώνεται στα metrics παραπάνω.

4 Τελικά Σχόλια - Επίλογος

Το πρόβλημα αναγνώρισης συναισθήματος αποτελεί ένα από τα δυσκολότερα classification προβλήματα που μπορεί να συναντήσει κανείς. Οι δυσκολίες λόγω των οποίων τα μοντέλα στην παραπάνω ανάλυση έχουν bias είναι:

- Περιορισμένα- μικρά σετ δεδομένων.
- Τεχνητά σετ δεδομένων. Ηθοποιοί υποδύονται συναισθήματα δεν είναι πραγματικά τα δεδομένα.
- Το στοιχείο της υποκειμενικότητας όταν ένα συναίσθημα γίνεται annotate από κάποιον σε μία συγκεκριμένη κλάση.
- Στην περίπτωση του πρώτου σετ δεδομένων υπάρχουν αρκετοί ομιλητές αλλά στην περίπτωση του δεύτερου οι ομιλητές είναι μόλις 4 και έτσι υπάρχει κίνδυνος τα μοντέλα μας να έχουν αποκτήσει ευαισθησία στα χαρακτηριστικά της ομιλίας των συγκεκριμένων ομιλητών.

Βασικό λοιπόν θετικό αποτέλεσμα της ανάλυσης που κάναμε για τις 3 περιπτώσεις και αυτό με επιφύλαξη θα μπορούσε κανείς να θεωρήσει την κατασκευή των μοντέλων στην πρώτη και δεύτερη πειραματική φάση τα οποία είναι σε θέση όπως είδαμε να ξεχωρίζουν συγκεκριμένα συναισθήματα από κάποια άλλα. Σε μελλοντική δουλειά θα μπορούσε κανείς να τα συνδυάσει και να ελέγξει σε καινούρια δεδομένα την αποτελεσματικότητά τους σε σχέση με την κλάση που ξεχωρίζει καλύτερα το καθένα.