



دانشگاه تهران
دانشکده مهندسی
برق و کامپیوتر



درس یادگیری ماشین
گزارش کار پروژه پایانی

نام و نام خانوادگی	میلاد محمدی، عرفان باقری، میلاد محدودی سوق
شماره دانشجویی	810100455، 810100298، 810100462
تاریخ ارسال گزارش	1401/11/21

فهرست مطالب

3	1- مقدمه
4	2- پیش پردازش
4	2-1 تمیزسازی داده‌ها
4	2-2 استخراج ویژگی
5	3- طبقه‌بندی
5	3-1-1 روش KNN
8	3-1-2 روش SVM
10	3-1-3 روش Logistic regression
12	3-1-4 روش MLP
14	3-1-5 روش Naïve Bayes
16	3-1-6 روش Ensemble learning
17	3-2 شبکه‌های عصبی عمیق (LSTM)
19	4- خوشه‌بندی
19	4-1-1 K-Means(k=2)
22	4-1-2 K-Means(k=7)
24	4-2 DBSCAN

*** آقای میلاد محدودی سوق با هماهنگی استاد ابوالقاسمی و آقای سلطانی جدیدا به گروه اضافه شده‌اند.

1- مقدمه

موسیقی همواره از قدیم تا امروزه بخش مهمی از زندگی بشری بوده است که در لحظات شادی و غم از آن استفاده شده است. برای دسترسی آسان به موسیقی دلخواه، شناسایی و طبقه‌بندی انواع مختلف موسیقی حائز اهمیت فراوان است. با توجه به وجود ژانرهای مختلف موسیقی و سرعت تولید محتوا در جهان امروزی، عملاً دسته‌بندی انواع مختلف موسیقی توسط انسان کار بسیار سختی است. بر همین اساس یک رویکرد کارآمد برای دسته‌بندی موسیقی‌ها می‌تواند استفاده از مدل‌های یادگیری عمیق و یادگیری ماشین باشد که امروزه برای حل بسیاری از مسائل از آن‌ها استفاده می‌شود.

امروزه تشخیص ژانرهای مختلف موسیقی از سوی شرکت‌های بزرگ موسیقی مانند Spotify و SoundCloud مورد توجه قرار گرفته است. این شرکت‌ها در واقع موسیقی‌هایی را که یک فرد دانلود و یا جستجو کرده است با استفاده از سیستم‌های توصیه‌گر موسیقی‌های جدیدی را که شبیه به علاقه‌مندی‌های فرد است، پیدا کرده و برای فرد پیشنهاد می‌دهند. از کاربردهای دیگر تشخیص ژانر موسیقی، دسته‌بندی این موسیقی‌ها در گروه‌های مختلف است که می‌توانیم با این کار تعداد آهنگ‌های هر دسته و نیز محبوبیت هر ژانر از موسیقی را کشف کنیم.

در نگاه علمی، دستگاه‌های موسیقی به مجموعه‌ای از چند نغمه (گوشه) اطلاق می‌شود که با هم در گام، کوک و فواصل نت هم‌آهنگی دارند (ویکی‌پدیا). دستگاه‌های موسیقی ایرانی شامل هفت دستگاه (شور، ماهور، همایون، سه‌گاه، چهارگاه، راست‌پنج‌گاه و نوا) و هر یک از این دستگاه‌ها شامل پنج آلت موسیقی (تار، سه‌تار، کمانچه، سنتور و نی) هستند. موسیقی از عناصر مختلفی همچون ملودی، ریتم، هارمونی و... تشکیل شده که ترکیب هنری این عناصر باعث تنوع گسترده‌ای در انواع موسیقی می‌شود. درک تفاوت‌ها میان فرم‌های مختلف موسیقی نیاز به تخصص دارد. به همین دلیل اغلب سامانه‌های طبقه‌بندی که طراحی می‌شوند با دیدگاهی ساده با استفاده از برچسب‌هایی که در فراداده‌ی فایل‌ها وجود دارد طبقه‌بندی را انجام می‌دهند. این برچسب‌ها توصیف‌گرهای متنی مثل نام خواننده، نام ترانه، متن ترانه و... هستند. پیاده‌سازی چنین روشی ساده بوده ولی نیازمند این است که برای یک فایل تمامی مقادیر فراداده آن از قبل توسط انسان تعریف شده باشد. از طرف دیگر در این روش از محتوای آهنگ (سیگنال صوتی) استفاده نمی‌شود. در حالی که محتوای صوتی آهنگ‌ها حاوی اطلاعات با ارزشی همچون گام صدا، ملودی و... است که می‌تواند در طبقه‌بندی بسیار موثر باشد. یکی از مسائلی که در این پژوهش به دنبال حل آن هستیم استخراج ویژگی مناسب از فایل‌های ترانه به منظور افزایش دقت سامانه است.

در این پژوهش، هدف ما تشخیص دستگاه‌های سنتی ایرانی با استفاده از موسیقی‌هایی است که با این دستگاه‌ها تولید شده‌اند. قصد داریم تا با استفاده از مدل‌های مختلف یادگیری عمیق و یادگیری ماشین، مدلی را پیدا کنیم که بهترین دقت را در میان تمام مدل‌های استفاده شده داشته باشد. برای این کار نکته مهمی که باید به آن توجه داشته باشیم،

داده‌های ما برای انجام این پژوهش است. مدل‌های یادگیری عمیق و یادگیری ماشین همواره با داده‌های تمیز، بهترین عملکرد را دارند.

برای جمع‌آوری دادگان برای هر دستگاه ایرانی، پنج ترانه را از سایت‌های ایرانی جمع‌آوری کرده‌ایم. این داده‌ها دارای فرمت mp3 و کیفیت‌های مختلف هستند ولی در مراحل بعدی پژوهش لازم است بر روی آن‌ها پیش پردازش انجام شود. در مرحله پیش پردازش ویژگی‌های هر ترانه استخراج خواهد شد، همبستگی بین ویژگی‌ها سنجیده خواهد شد، در صورت وجود مقادیر پرت این مقادیر اصلاح خواهد شد و در صورت نیاز بر روی آن‌ها نرمال‌سازی انجام خواهد شد تا در مرحله‌ی طبقه‌بندی داده‌هایی تمیز داشته باشیم. یکی از لازمه‌های استفاده از الگوریتم‌های ماشین لرنینگ برای حل یک مسئله، وجود داده‌های لازم برای آن مسئله است. چالش اولی که در رابطه با داده‌ها وجود دارد این است که آیا تمامی افراد داده‌های خود را به طور صحیح جمع‌آوری و برچسب‌گذاری کرده‌اند؟ چالش دوم تشابه موسیقی‌های سنتی در دستگاه‌ها به همدیگر است که در جمع‌آوری داده‌ها در بسیاری از صفحه‌های وب نام آن‌ها در کنار یکدیگر آورده شده بود (رفتار و حالت دستگاه ماهر و راست‌پنج‌گاه شبیه همدیگر است. رفتار و حالت دستگاه شور و نوا شبیه همدیگر است) و کار تشخیص را کمی سخت می‌کند. همچنین چالش سوم در رابطه با داده‌ها، تمیز کردن داده می‌باشد. با توجه به داده‌هایی که جمع‌آوری کرده‌ایم، در ابتدای فایل موسیقی، گوینده نام موسیقی و دستگاه را ذکر می‌کند که برای تشخیص توسط ماشین، این صداها اضافی هستند. علاوه بر این طول موسیقی‌ها متفاوت است و برای استفاده از الگوریتم‌های یادگیری ماشین باید طول یکسان داشته باشند. البته مدل‌های یادگیری عمیق مانند LSTM که برای داده‌های sequential مناسب هستند، با طول موسیقی متفاوت نیز می‌توان از آن‌ها استفاده کرد.

2- پیش‌پردازش

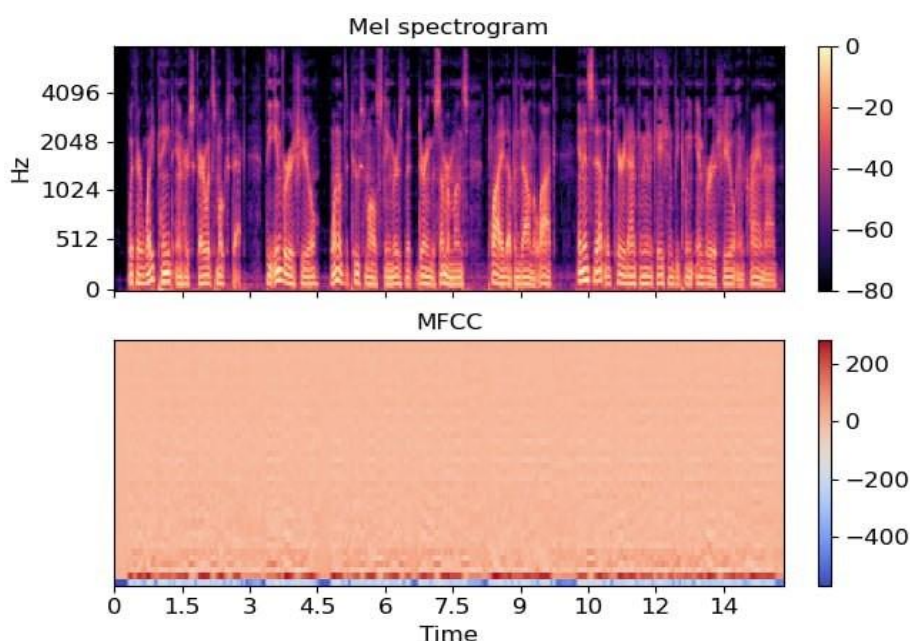
2-1 تمیزسازی داده‌ها

در بین داده‌های جمع‌آوری شده برای انجام پروژه فرمت‌های مختلفی دیده می‌شد. به عنوان مثال چندین آهنگ دارای فرمت mp4 بودند و همه آن‌ها را به فرمت یکسان mp3 تبدیل کردیم. یکی از محدودیت‌هایی که در جمع‌آوری آهنگ‌ها وجود داشت، حداقل طول آهنگ به میزان 20 ثانیه بود ولی با این حال در بین داده‌ها فایل‌هایی با طول کمتر از 20 ثانیه وجود داشت و این آهنگ‌ها را حذف کردیم. در کل 906 آهنگ جمع‌آوری شد که شامل 124 آهنگ برای دستگاه شور، 124 آهنگ برای دستگاه سه‌گاه، 140 آهنگ برای دستگاه ماهر، 145 آهنگ برای دستگاه همایون، 116 آهنگ برای دستگاه راست پنج‌گاه، 138 آهنگ برای دستگاه نوا و 142 آهنگ برای دستگاه چهارگاه می‌باشد.

2-2 استخراج ویژگی

برای انجام feature extraction از کتابخانه Librosa استفاده شده است. برای استخراج ویژگی‌ها از دو روش مختلف MFCC, Mel-spectrogram که frequency based هستند، استفاده شده است. به منظور حذف بعد زمان،

میانگین قدرت هر فرکانس در طی بعد زمان گرفته شد و برای از دست ندادن اطلاعات موجود در واریانس، انحراف معیار نیز در نظر گرفته شد و به عنوان فیچر به میانگین concat شد. این کار سبب شد که هر فایل صوتی در Mel-spectrogram و MFCC به یک بردار 256 بعدی تبدیل شود. با اجرای pca برای انتخاب 99.5 درصد واریانس کل، تعداد ابعاد بر روی داده‌های Mel-spectrogram به 50 و بر روی داده‌های MFCC به 110 کاهش یافت.



شکل 1. استخراج ویژگی با استفاده از *MFCC* , *Mel-spectrogram*

3- طبقه‌بندی

برای هر کدام از طبقه‌بندهای استفاده شده نمودار ROC و Classification report مربوط به دو روش استخراج ویژگی در ادامه آورده شده است.

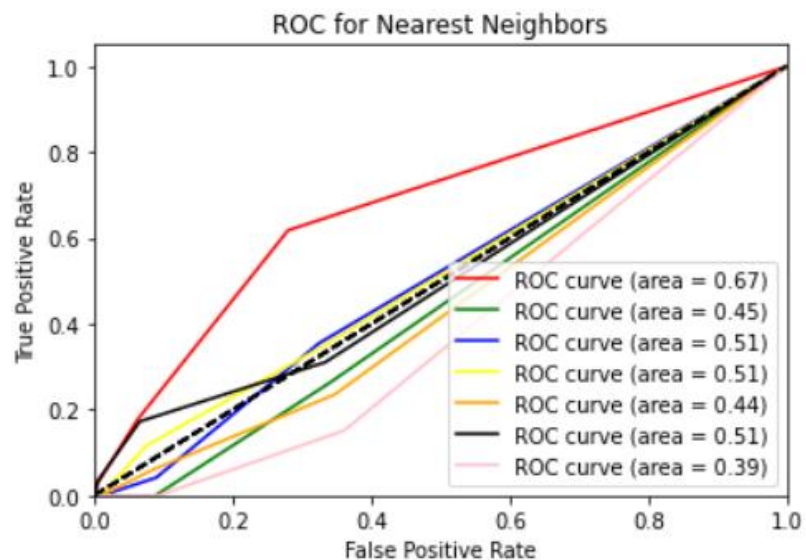
داده‌ها برای طبقه‌بندی به صورت زیر جدا شدند:

داده‌های آموزشی: 75 درصد داده‌ها

داده‌های تست: 25 درصد داده‌ها

3-1-1 روش KNN

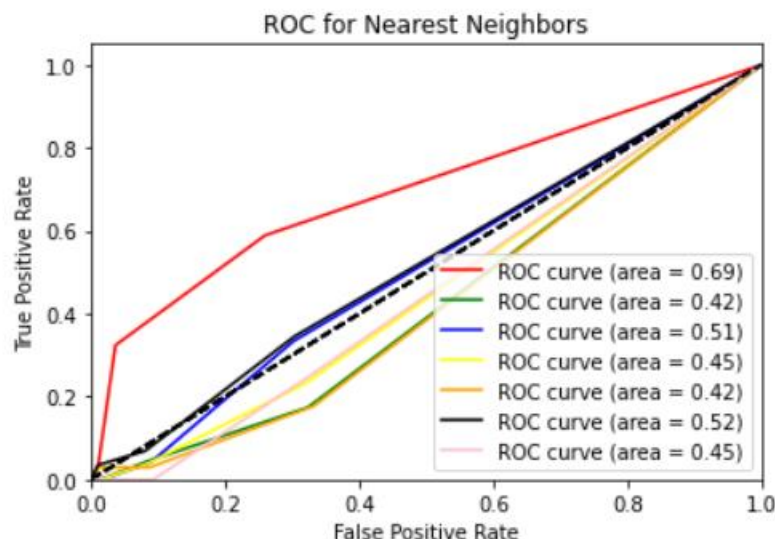
از روش KNN با $k=3$ برای طبقه‌بندی دستگاه‌ها استفاده شد. نتایج استفاده از این طبقه‌بند به صورت زیر است:



شکل 2. نمودار ROC روش *KNN* با استفاده از *Mel-spectrogram*

Classification with: Nearest Neighbors				
	precision	recall	f1-score	support
D_0	0.29	0.53	0.38	34
D_1	0.25	0.48	0.33	23
D_2	0.29	0.21	0.24	48
D_3	0.30	0.27	0.29	26
D_4	0.22	0.12	0.15	34
D_5	0.32	0.28	0.30	29
D_6	0.52	0.33	0.41	33
accuracy			0.30	227
macro avg	0.31	0.32	0.30	227
weighted avg	0.32	0.30	0.29	227

شکل 3. Classification report روش *KNN* با استفاده از *Mel-spectrogram*



شکل 4. نمودار ROC روش *KNN* با استفاده از *MFCC*

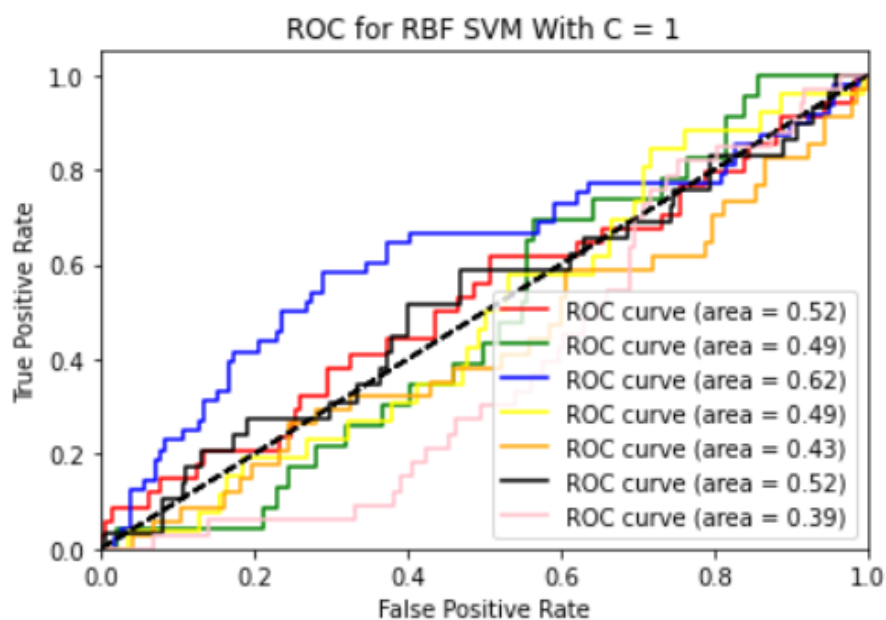
Classification with: Nearest Neighbors				
	precision	recall	f1-score	support
D_0	0.32	0.53	0.40	34
D_1	0.42	0.65	0.51	23
D_2	0.40	0.33	0.36	48
D_3	0.53	0.35	0.42	26
D_4	0.44	0.47	0.46	34
D_5	0.43	0.45	0.44	29
D_6	0.83	0.30	0.44	33
accuracy			0.43	227
macro avg	0.48	0.44	0.43	227
weighted avg	0.48	0.43	0.43	227

شکل 5. Classification report روش *KNN* با استفاده از *MFCC*

همان‌طور که از نتایج به دست آمده مشاهده می‌شود، مساحت زیر نمودار در روش *MFCC* در بعضی از دستگاه‌ها بیشتر از روش *Mel-spectrogram* و در بعضی کمتر است ولی در کل دقت طبقه‌بند با استفاده از روش *MFCC* 13 درصد بیشتر از روش *Mel-spectrogram* است. این اختلاف می‌تواند به این دلیل باشد که ویژگی‌های روش *MFCC* مستقل از هم هستند و می‌توانند تمایز خوبی ایجاد کنند. همچنین زمان محاسبات برای *MFCC* کمتر از *Mel-spectrogram* است که می‌تواند محاسباتی فشرده باشد. استفاده از *MFCC* می‌تواند محاسبات الگوریتم *KNN* را کاهش دهد و عملکرد آن را بهبود بخشد. علاوه بر این، *MFCC* به گونه‌ای طراحی شده‌اند که نسبت به تغییر در سیگنال، مانند تغییرات در زیر و بم، سرعت صحبت و صدا مقاوم باشند که می‌تواند تاثیر نویز و

تغییرپذیری در داده‌ها را کاهش دهد. این می‌تواند تشخیص دقیق الگوهای موجود در داده‌ها و پیش‌بینی را برای KNN آسان‌تر کند.

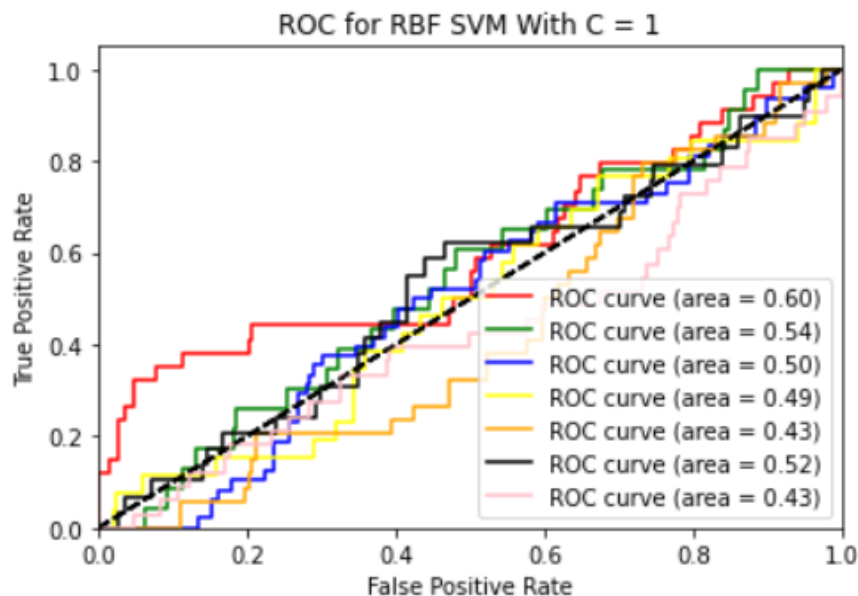
3-1-2 روش SVM(RBF, C=1)



شکل 6. نمودار ROC روش SVM با استفاده از Mel-spectrogram

Classification with: RBF SVM With C = 1				
	precision	recall	f1-score	support
D_0	0.00	0.00	0.00	34
D_1	0.50	0.43	0.47	23
D_2	0.50	0.04	0.08	48
D_3	0.15	0.88	0.26	26
D_4	0.00	0.00	0.00	34
D_5	0.22	0.34	0.27	29
D_6	0.56	0.15	0.24	33
accuracy			0.22	227
macro avg	0.28	0.27	0.19	227
weighted avg	0.28	0.22	0.16	227

شکل 7. Classification report روش SVM با استفاده از Mel-spectrogram



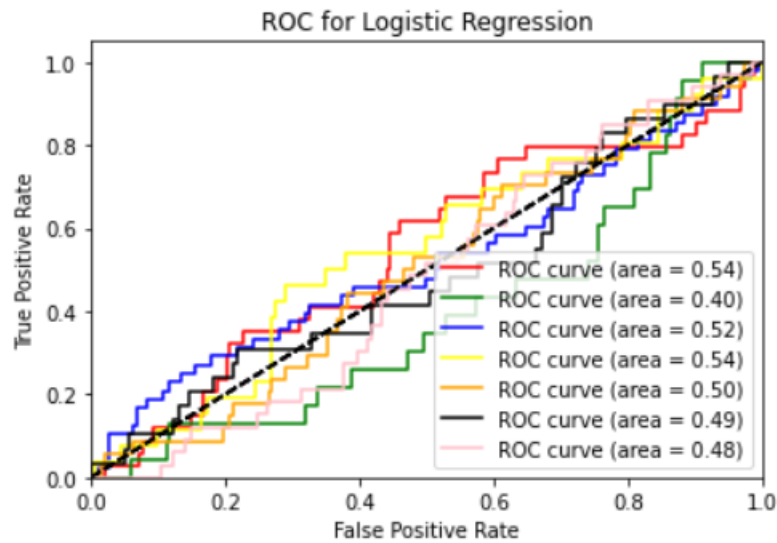
شکل 8. نمودار ROC روش SVM با استفاده از MFCC

Classification with: RBF SVM With C = 1				
	precision	recall	f1-score	support
D_0	0.80	0.12	0.21	34
D_1	0.73	0.35	0.47	23
D_2	0.71	0.10	0.18	48
D_3	0.16	0.85	0.27	26
D_4	0.60	0.09	0.15	34
D_5	0.43	0.76	0.55	29
D_6	0.80	0.24	0.37	33
accuracy			0.32	227
macro avg	0.60	0.36	0.31	227
weighted avg	0.62	0.32	0.29	227

شکل 9. Classification report روش SVM با استفاده از MFCC

با توجه به نتایج به دست آمده مشاهده می‌شود که عملکرد SVM با استفاده از روش MFCC بهتر از Mel-spectrogram است. ویژگی‌های طیفی سیگنال‌های صوتی را به‌طور مؤثرتری نسبت به Mel-spectrogram استخراج می‌کنند و به گونه‌ای طراحی شده‌اند که در برابر تغییرپذیری در سیگنال مقاوم باشند. این می‌تواند منجر به نمایش تبعیض آمیزتر داده‌ها شود و عملکرد SVM را بهبود بخشد.

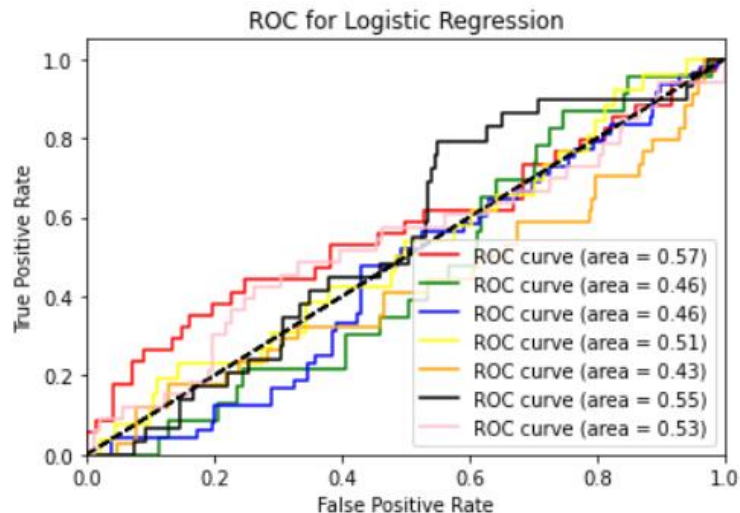
3-1-3 روش Logistic regression



شکل 10. نمودار ROC روش Logistic regression با استفاده از Mel-spectrogram

Classification with: Logistic Regression				
	precision	recall	f1-score	support
D_0	0.06	0.03	0.04	34
D_1	0.15	0.35	0.21	23
D_2	0.46	0.27	0.34	48
D_3	0.20	0.35	0.25	26
D_4	0.13	0.06	0.08	34
D_5	0.24	0.31	0.27	29
D_6	0.25	0.24	0.25	33
accuracy			0.22	227
macro avg	0.21	0.23	0.21	227
weighted avg	0.23	0.22	0.21	227

شکل 11. Classification report روش Logistic regression با استفاده از Mel-spectrogram



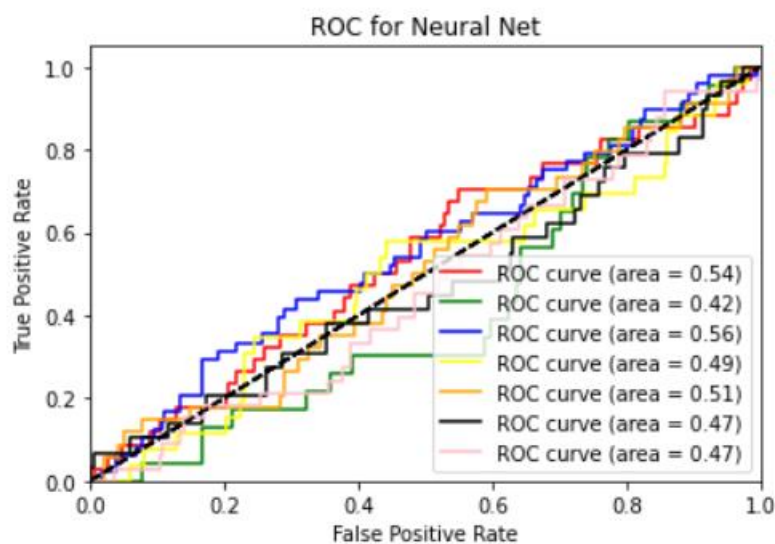
شکل 12. نمودار ROC روش *Logistic regression* با استفاده از MFCC

Classification with: Logistic Regression				
	precision	recall	f1-score	support
D_0	0.38	0.24	0.29	34
D_1	0.35	0.57	0.43	23
D_2	0.49	0.35	0.41	48
D_3	0.22	0.27	0.24	26
D_4	0.22	0.21	0.21	34
D_5	0.44	0.62	0.51	29
D_6	0.45	0.39	0.42	33
accuracy			0.37	227
macro avg	0.36	0.38	0.36	227
weighted avg	0.37	0.37	0.36	227

شکل 13. Classification report روش *Logistic regression* با استفاده از MFCC

دقت طبقه‌بند در روش MFCC بیشتر از Mel-spectrogram است که می‌تواند به دلایلی که قبلاً در طبقه‌بند KNN بیان شد، این اتفاق بیفتد. همچنین دقت طبقه‌بند *Logistic regression* بیشتر از SVM شده است. *Logistic regression* از نظر محاسباتی کارآمد است و آن را به انتخاب خوبی برای مجموعه داده‌های بزرگ تبدیل می‌کند. از طرفی، SVM زمانی که داده‌ها ابعاد بالایی دارند، از نظر محاسباتی به خوبی عمل نمی‌کند.

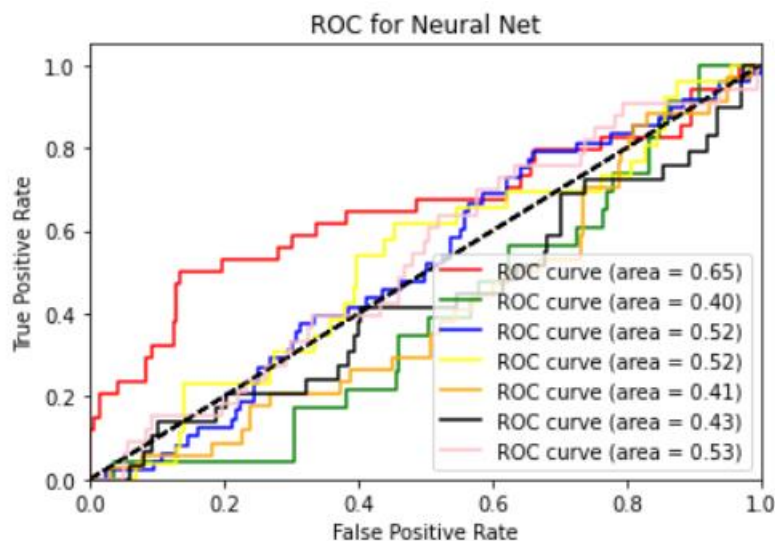
3-1-4 روش Neural Net (MLP)



شکل 14. نمودار ROC روش *MLP* با استفاده از *Mel-spectrogram*

Classification with: Neural Net				
	precision	recall	f1-score	support
D_0	0.67	0.06	0.11	34
D_1	0.19	0.13	0.15	23
D_2	0.70	0.15	0.24	48
D_3	0.14	0.65	0.23	26
D_4	0.00	0.00	0.00	34
D_5	0.35	0.24	0.29	29
D_6	0.23	0.39	0.29	33
accuracy			0.22	227
macro avg	0.32	0.23	0.19	227
weighted avg	0.36	0.22	0.19	227

شکل 15. Classification report روش *MLP* با استفاده از *Mel-spectrogram*



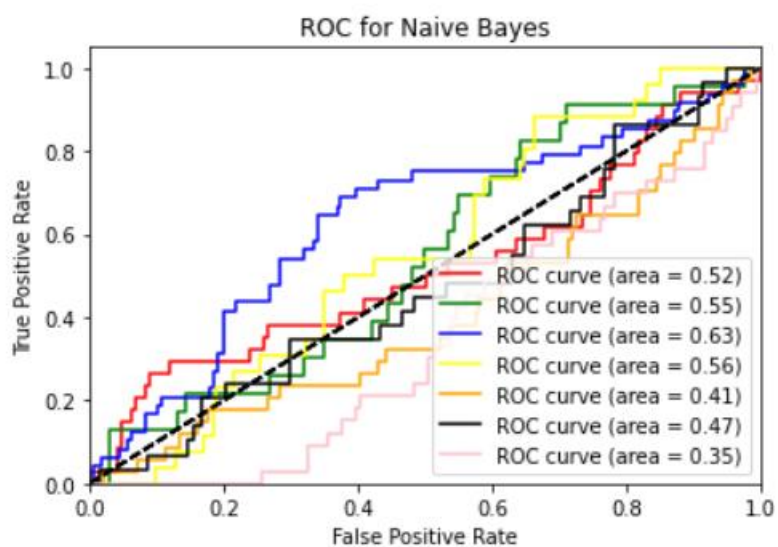
شکل 16. نمودار ROC روش *MLP* با استفاده از *MFCC*

Classification with: Neural Net				
	precision	recall	f1-score	support
D_0	0.38	0.29	0.33	34
D_1	0.38	0.61	0.47	23
D_2	0.42	0.17	0.24	48
D_3	0.25	0.50	0.33	26
D_4	0.60	0.26	0.37	34
D_5	0.35	0.59	0.44	29
D_6	0.45	0.39	0.42	33
accuracy			0.37	227
macro avg	0.40	0.40	0.37	227
weighted avg	0.41	0.37	0.36	227

شکل 17. Classification report روش *MLP* با استفاده از *MFCC*

نتایج به دست آمده نشان می‌دهد که همچنان عملکرد طبقه‌بندها با استفاده از روش *MFCC* بهتر از روش *Mel-spectrogram* است. دقت طبقه‌بند *MLP* و *Logistic regression* برابر و بیشتر از *SVM* است. *MLP* یک مدل قدرتمندتر و انعطاف پذیرتر از *SVM* است، زیرا می‌تواند روابط غیر خطی پیچیده بین ورودی‌ها و خروجی‌ها را بیاموزد. این باعث می‌شود که برای کارهایی که داده‌ها به صورت غیر خطی قابل تفکیک هستند مناسب‌تر باشد. علاوه بر این ویژگی‌هایی که طبقه‌بندها بر اساس آن عمل طبقه‌بندی را انجام می‌دهند، مهم است و تا به اینجا روش *MFCC* بهتر از روش *Mel-spectrogram* بوده است.

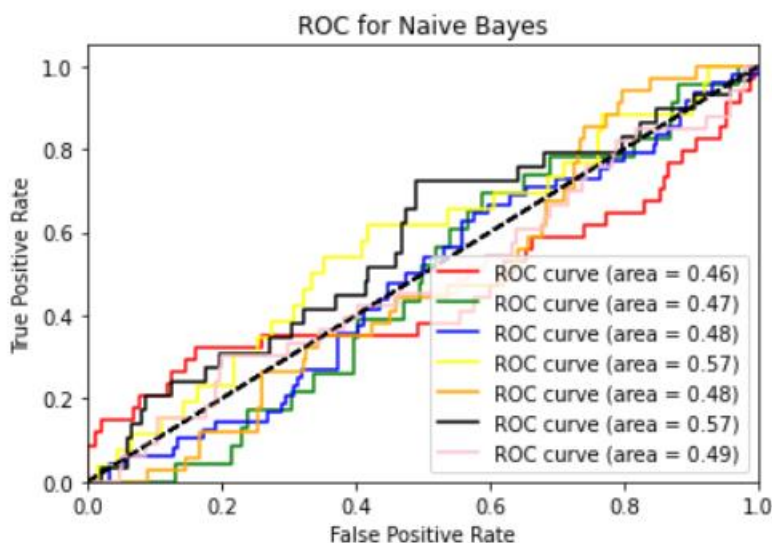
3-1-5 روش Naïve Bayes



شکل 18. نمودار ROC روش Naïve Bayes با استفاده از Mel-spectrogram

Classification with: Naive Bayes				
	precision	recall	f1-score	support
D_0	0.33	0.15	0.20	34
D_1	0.12	0.74	0.20	23
D_2	0.33	0.06	0.11	48
D_3	0.17	0.15	0.16	26
D_4	0.20	0.03	0.05	34
D_5	0.40	0.21	0.27	29
D_6	0.08	0.03	0.04	33
accuracy			0.16	227
macro avg	0.23	0.20	0.15	227
weighted avg	0.24	0.16	0.14	227

شکل 19. Classification report روش Naïve Bayes با استفاده از Mel-spectrogram



شکل 20. نمودار ROC روش *Naïve Bayes* با استفاده از MFCC

Classification with: Naive Bayes				
	precision	recall	f1-score	support
D_0	0.26	0.24	0.25	34
D_1	0.26	0.43	0.32	23
D_2	0.50	0.25	0.33	48
D_3	0.28	0.35	0.31	26
D_4	0.26	0.35	0.30	34
D_5	0.29	0.28	0.28	29
D_6	0.37	0.30	0.33	33
accuracy			0.30	227
macro avg	0.32	0.31	0.30	227
weighted avg	0.33	0.30	0.30	227

شکل 21. Classification report روش *Naïve Bayes* با استفاده از MFCC

تعداد دستگاه‌ها 7 تاست و برای طبقه‌بندی 7 کلاس داریم. در حالت رندوم دقت طبقه‌بند $0.15 \approx \frac{1}{7}$ است. با توجه به شکل 19 مشاهده می‌شود که طبقه‌بند *Naïve Bayes* با روش Mel-spectrogram شبیه به حالت base line

عمل می‌کند در حالی که با استفاده از روش MFCC دقت طبقه‌بند 30 درصد است. این نشان‌گر این است که ویژگی‌های روش MFCC مستقل از هم هستند.

Ensemble learning روش 3-1-6

	precision	recall	f1-score	support
D_0	0.33	0.26	0.30	34
D_1	0.21	0.70	0.33	23
D_2	0.53	0.17	0.25	48
D_3	0.20	0.54	0.29	26
D_4	0.00	0.00	0.00	34
D_5	0.38	0.41	0.39	29
D_6	0.50	0.12	0.20	33
accuracy			0.28	227
macro avg	0.31	0.31	0.25	227
weighted avg	0.33	0.28	0.24	227

شکل 22. Classification report روش Ensemble learning با استفاده از Mel-spectrogram

	precision	recall	f1-score	support
D_0	0.34	0.35	0.35	34
D_1	0.41	0.65	0.50	23
D_2	0.53	0.33	0.41	48
D_3	0.30	0.54	0.39	26
D_4	0.53	0.29	0.38	34
D_5	0.44	0.59	0.50	29
D_6	0.52	0.33	0.41	33
accuracy			0.42	227
macro avg	0.44	0.44	0.42	227
weighted avg	0.45	0.42	0.41	227

شکل 23. Classification report روش Ensemble learning با استفاده از MFCC

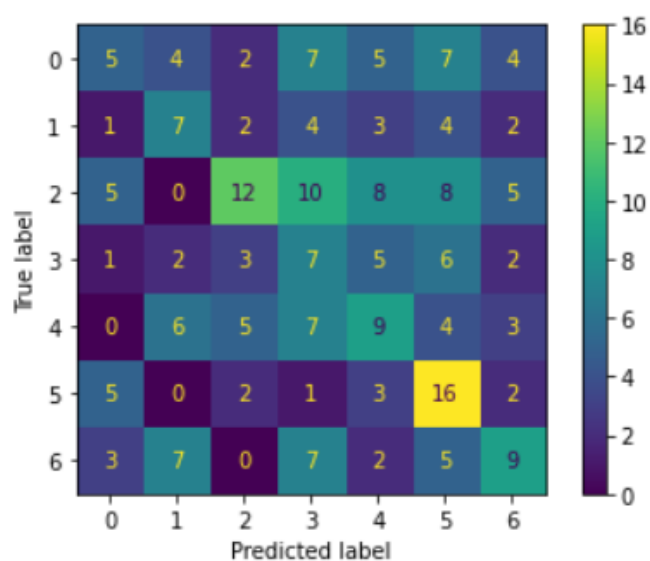
دقت طبقه‌بندی در Ensemble learning با استفاده از روش MFCC، 42 درصد و در روش Mel-spectrogram 22 درصد شده است. دلیل کم بودن دقت در روش Mel-spectrogram می‌تواند انتخاب ضعیف ویژگی باشد. اگر ویژگی‌های مورد استفاده برای کار طبقه‌بندی مرتبط یا آموزنده نباشد، طبقه‌بند کننده‌ها ممکن است نتوانند پیش‌بینی‌های دقیقی انجام دهند.

3-2 شبکه‌های عصبی عمیق (LSTM)

```
accuracy_score(Y_test_new, y_pred, normalize=False)
```

```
8/8 [=====] - 1s 10ms/step  
65
```

شکل 24. دقت *LSTM* با استفاده از *Mel-spectrogram*

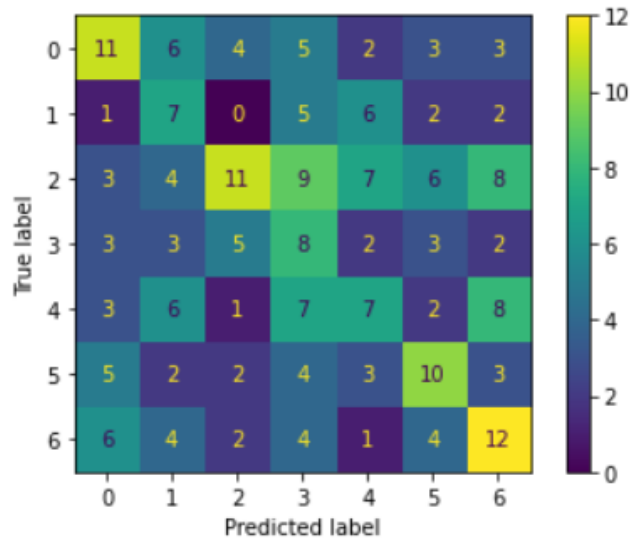


شکل 25. ماتریس آشفتگی *LSTM* با استفاده از *Mel-spectrogram*

```
accuracy_score(Y_test_new, y_pred, normalize=False)
```

```
8/8 [=====] - 1s 15ms/step  
66
```

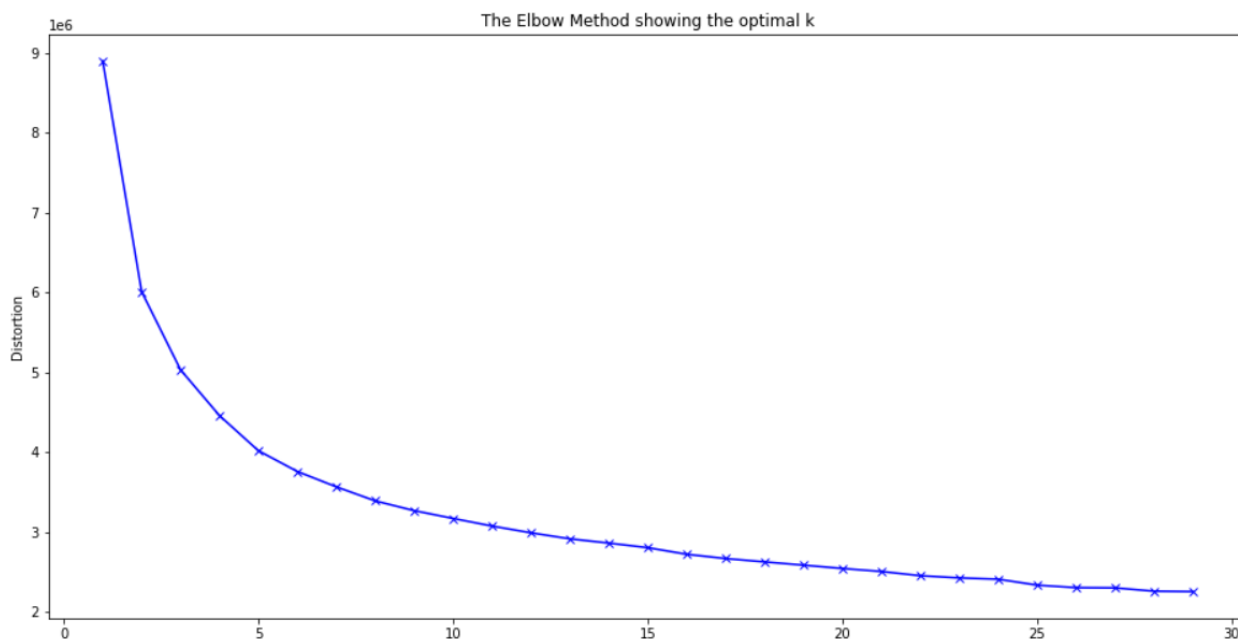
شکل 26. دقت *LSTM* با استفاده از *MFCC*



شکل 27. ماتریس آشفتنگی *LSTM* با استفاده از *MFCC*

دقت طبقه‌بندی با استفاده از *LSTM* در مقایسه با سایر طبقه‌بند بسیار بالا بوده است. به دلیل پیچیدگی و منحصربه‌فرد بودن الگوهای موسیقی، این الگوها توسط طبقه‌بندی‌کننده‌های ساده مانند *KNN*، *SVM*، *logistic regression* و *Naive Bayes* به خوبی درک نشود. در مقابل، *LSTM*، به عنوان یک نوع شبکه عصبی بازگشتی، قادر به مدیریت داده‌های متوالی هستند و آنها را برای تشخیص الگوها در سیگنال‌های موسیقی مناسب می‌کند. از جمله چالش‌هایی که در این پروژه وجود دارد، تعداد آهنگ‌های کم در مقایسه با 7 کلاس می‌باشد. همچنین شباهت برخی دستگاه‌ها به همدیگر (رفتار و حالت دستگاه ماهر و راست‌پنج‌گاه شبیه همدیگر است. رفتار و حالت دستگاه شور و نوا شبیه همدیگر است) یکی دیگر از عواملی است که باعث شده تا طبقه‌بند دقت کافی را نداشته باشند.

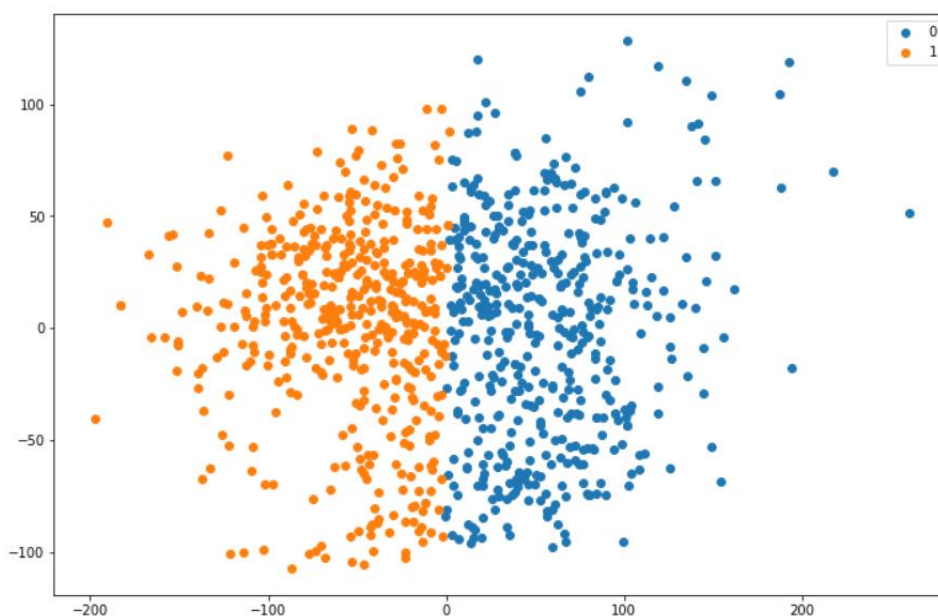
4- خوشه‌بندی



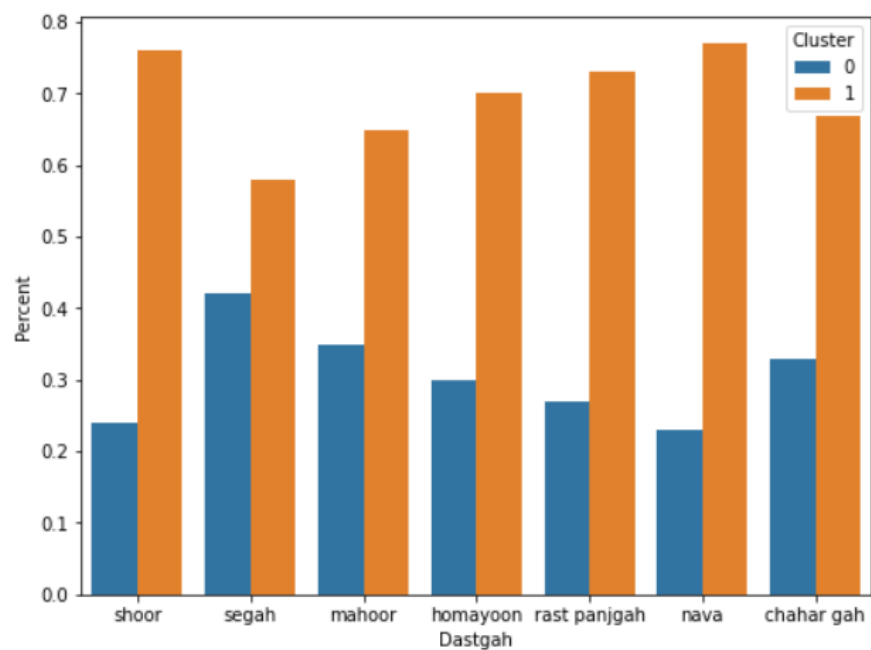
شکل 28. Elbow method

نمودار بالا نشان‌دهنده‌ی Elbow method است که برای یافتن مقدار بهینه تعداد کلاسترها استفاده می‌شود. با توجه نمودار مقدار بهینه برابر 4 یا 5 می‌باشد ولی با توجه به صورت پروژه با مقادیر تعیین شده خوشه‌بندی را انجام دادیم.

4-1-1 K-Means(k=2)



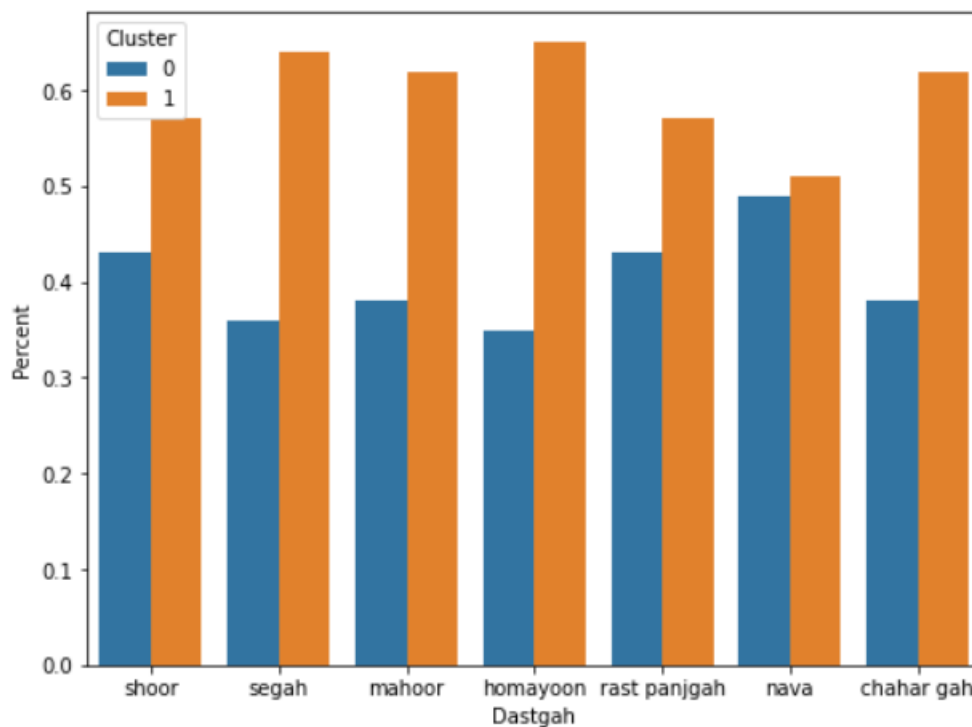
شکل 29. اختصاص دستگاه‌ها به دو خوشه با استفاده از Mel-spectrogram



شکل 30. درصد اختصاص هر دستگاه به خوشه‌ها با استفاده از *Mel-spectrogram*



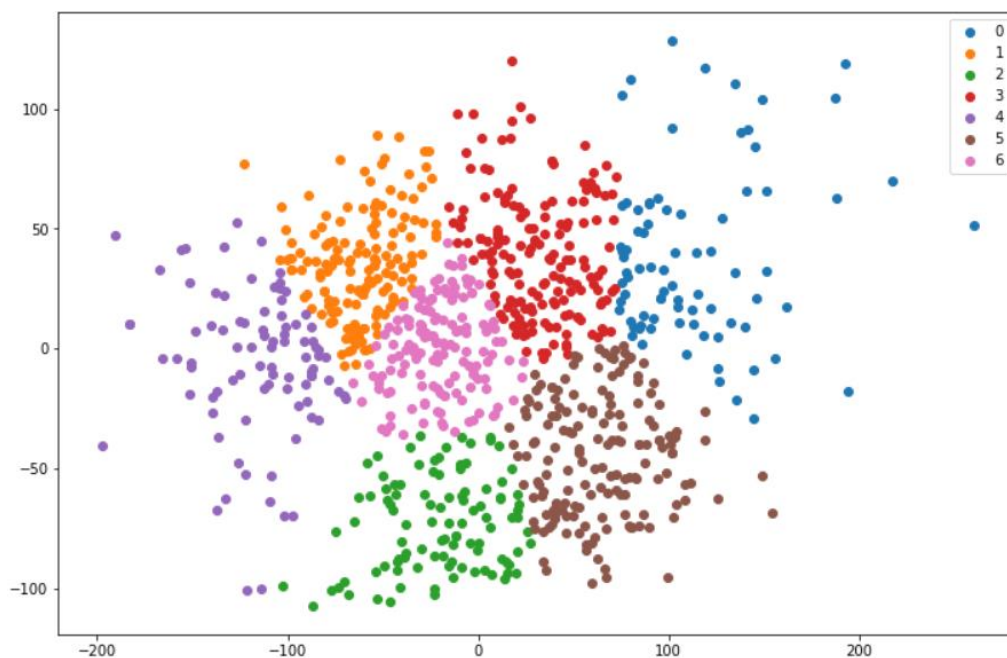
شکل 31. اختصاص دستگاه‌ها به دو خوشه با استفاده از *MFCC*



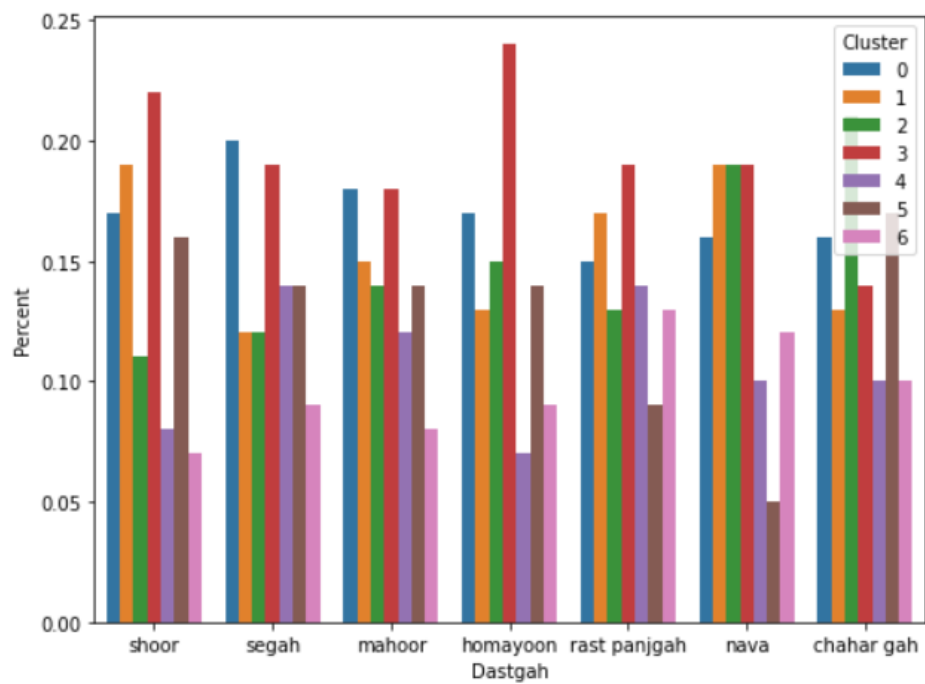
شکل 32. درصد اختصاص هر دستگاه به خوشه‌ها با استفاده از MFCC

دستگاه‌های موسیقی دارای چندین نوع آهنگ می‌باشند و هر کدام از این آهنگ‌ها ریتم‌های مختلفی دارند که در برخی دستگاه‌ها نقاط اوج و فرود شبیه به هم دیگر می‌باشد. با توجه به شکل 30 در روش Mel-spectrogram درصد بسیاری از آهنگ‌های هر دستگاه در خوشه یک قرار گرفته‌اند ولی در روش MFCC اختصاص هر دستگاه به خوشه‌ها تقریباً نرمال‌تر است و این به دلیل تفاوت ویژگی‌های این دو روش است که هر کدام ویژگی‌های خاصی از آهنگ را مورد توجه قرار داده‌اند. همچنین دستگاه‌های ماهور و راست پنج‌گاه که شبیه به هم هستند اختصاص آهنگ‌های مربوط به این دستگاه‌ها به هر کدام از خوشه‌ها شبیه به هم بوده است. به طور کلی آهنگ‌هایی که قدرت فرکانس آن‌ها شبیه به هم بوده، به همدیگر نزدیک‌تر بوده و در یک خوشه قرار گرفته‌اند.

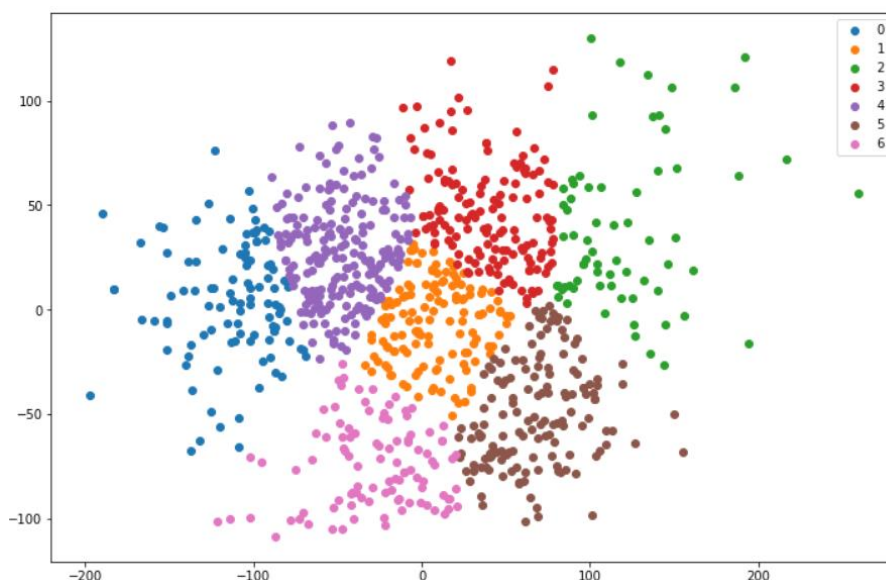
K-Means(k=7) 4-1-2



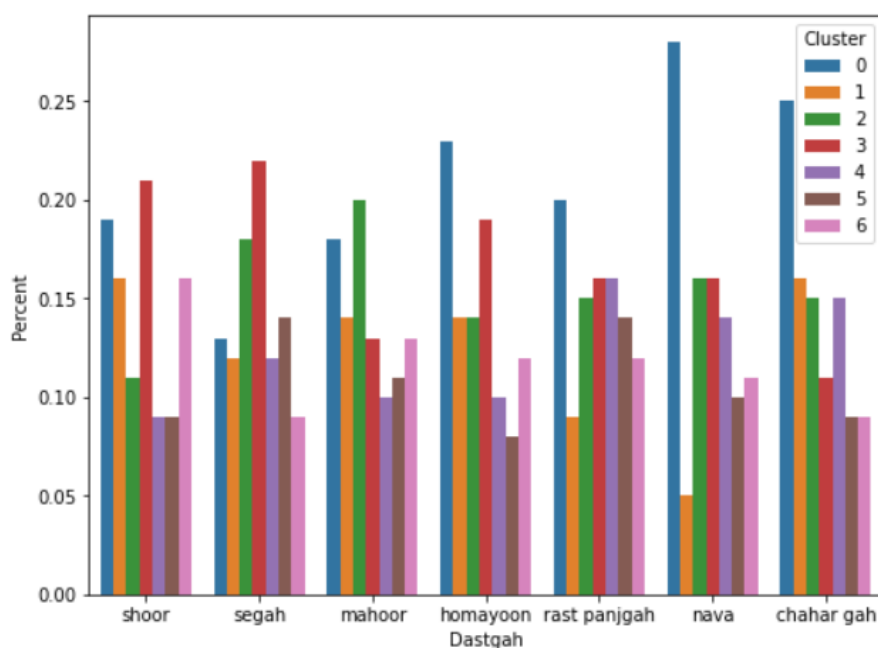
شکل 33. اختصاص دستگاه‌ها به 7 خوشه با استفاده از *Mel-spectrogram*



شکل 34. درصد اختصاص هر دستگاه به خوشه‌ها با استفاده از *Mel-spectrogram*



شکل 35. اختصاص دستگاه‌ها به 7 خوشه با استفاده از MFCC



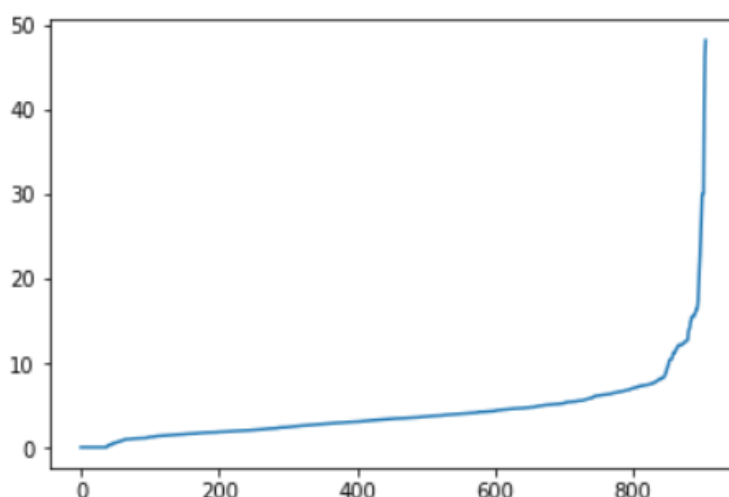
شکل 36. درصد اختصاص هر دستگاه به خوشه‌ها با استفاده از MFCC

در خوشه‌بندی با $k=7$ نیز آهنگ‌هایی که از نظر قدرت فرکانس به هم شبیه بوده‌اند، به دیگر نزدیک‌تر بوده و در یک خوشه قرار گرفته‌اند. دستگاه‌ها با هم همپوشانی دارند و به خوبی از هم جدا نیستند و این می‌تواند باعث شود که انتساب خوشه مبهم باشد و در نتیجه هر کلاس در چندین خوشه وجود داشته باشد. همچنین اگر خوشه‌ها کروی نباشند، می‌تواند باعث همپوشانی کلاس‌ها شود و در نتیجه هر کلاس در چندین خوشه وجود داشته باشد. تعداد داده‌ی کافی و انتخاب ویژگی‌های مناسب باعث می‌شود که خوشه‌بندی به طور مناسبی انجام شود. در این پروژه داده‌های ما کافی نبودند و به

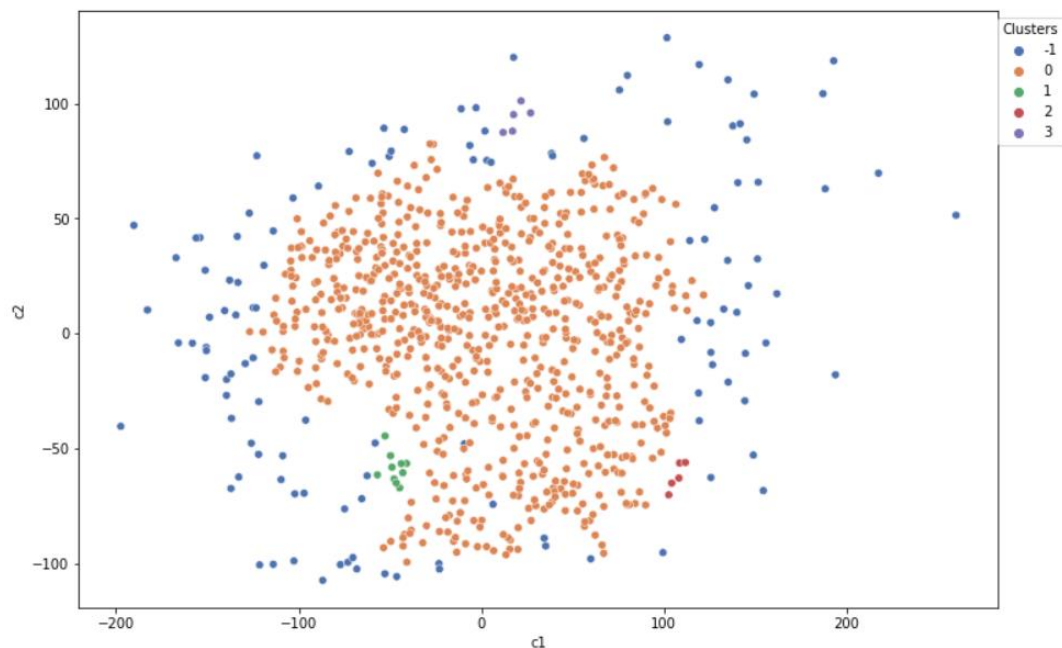
همین دلیل عمل خوشه‌بندی که مورد انتظار بود (اختصاص هر کدام از دستگاه‌ها به یک خوشه و در بدترین حالت اختصاص هر دستگاه به دو خوشه) انجام نپذیرفت.

DBSCAN 4-2

جهت استفاده از این الگوریتم باید فرآپارامترهای `epsilon` و `min_samples` را تنظیم نمود. چون این دو پارامتر به شدت به هم وابسته هستند یکی از آنها را ثابت و دیگری را متغیر در نظر گرفته و `tuning` انجام دادیم. در اینجا ما `min_samples` برابر 5 در نظر گرفته‌ایم. برای یافتن مقدار بهینه شعاع، ابتدا از KNN استفاده کردیم تا فاصله‌ای را که در آن داده‌ها به هم نزدیکی دارند پیدا کنیم. با توجه به نمودار زیر از نقطه 10 فاصله‌ها به صورت نمایی افزایش می‌یابد و بهترین `epsilon` برابر 10 است.

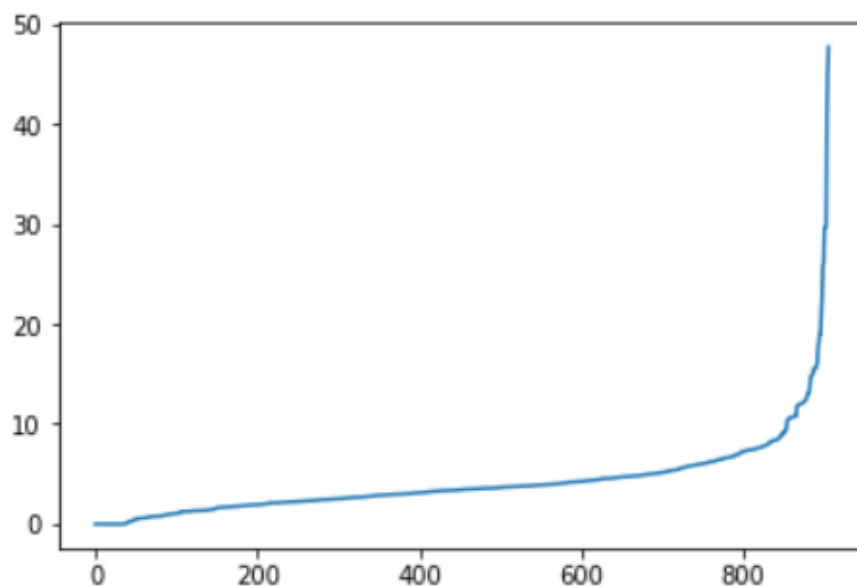


شکل 37. یافتن شعاع مناسب با استفاده از *Mel-spectrogram*

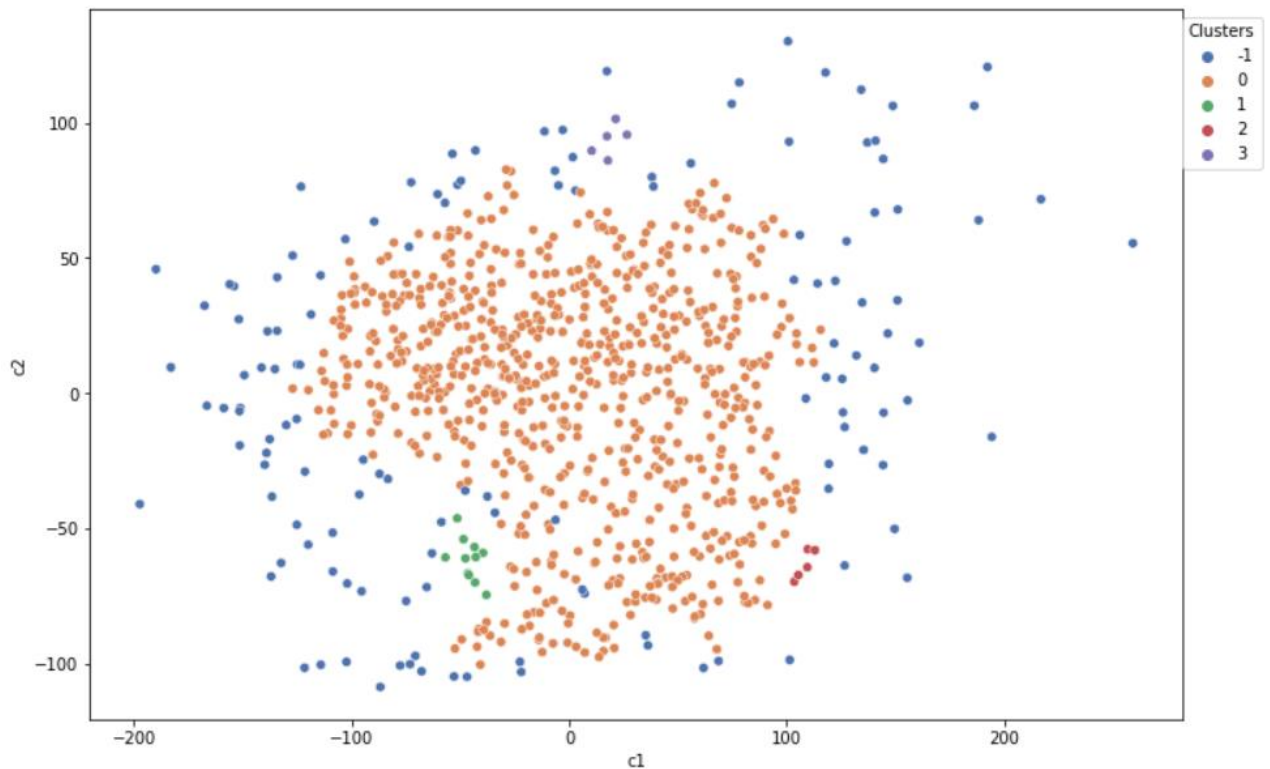


شکل 38. اختصاص دستگاه‌ها به 5 خوشه با استفاده از *Mel-spectrogram*

*** در هر دو روش *Mel-spectrogram* و MFCC به مقدار شعاع یکسانی دست یافتیم.



شکل 39. یافتن شعاع مناسب با استفاده از MFCC



شکل 40. اختصاص دستگاه‌ها به 5 خوشه با استفاده از MFCC

DBSCAN یک الگوریتم خوشه‌بندی مبتنی بر چگالی است، به این معنی که نقاطی را که نزدیک به یکدیگر در یک منطقه متراکم هستند، خوشه‌بندی می‌کند. دستگاه‌های موسیقی شور و نوا و همچنین ماهر و راست پنج‌گاه که به هم دیگر شبیه هستند در یک خوشه قرار گرفته‌اند. ابعاد داده‌ها یا تعداد ویژگی‌های مورد استفاده برای نمایش دستگاه‌های موسیقی نیز می‌تواند بر نتایج الگوریتم DBSCAN تأثیر بگذارد. داده‌ها ابعاد بالایی دارند و تشخیص دقیق خوشه‌ها دشوار است به همین دلیل برخی از آهنگ‌ها به عنوان outlier شناخته شده‌اند.