|  |  |
| --- | --- |
| نام و نام خانوادگی | **امیرحسین پورداود – امیرغرقابی – محمد ابوذری – مهدی سلیمانی زادگان** |
| شماره دانشجویی | **810101120 – 810102217 – 810101087 - 810102176** |
| تاریخ ارسال گزارش | **۱۴۰2.11.۰4** |

|  |  |  |
| --- | --- | --- |
|  | **به نام خدا**  **دانشگاه تهران**  **دانشکده‌ مهندسی برق و کامپیوتر** |  |
| **درس یادگیری ماشین**  **گزارش پروژه پایانی** | | |

**فهرست**

[**قسمت 1**. **پیش پردازش داده ها** 4](#_Toc157512031)

[1.1. تمیزکردن داده ها 4](#_Toc157512032)

[2.1. استخراج ویژگی داده ها 4](#_Toc157512033)

[**قسمت ۲**. **طبقه‌بندی داده ها** 5](#_Toc157512034)

[1.2. مدل های پیشنهادی 5](#_Toc157512035)

[2.2. روش های بهبود نتیجه 5](#_Toc157512036)

[3.2. نتیجه‌گیری و مقایسه 13](#_Toc157512037)

[**قسمت ۳** **– خوشه بندی داده ها** 14](#_Toc157512038)

[1.3. مدل های پیشنهادی 14](#_Toc157512039)

[2.3. روش های بهبود نتیجه 15](#_Toc157512040)

[3.3. نتیجه گیری و مقایسه 15](#_Toc157512041)

[مدل **(ASR) Automatic Speech Recognition** 15](#_Toc157512042)

**شکل‌ها**

شکل 1. تعداد نمونه های مرد و زن در داده‌های آموزشی و تست 6

شکل 2. توصیف داده‌های آموزشی 6

شکل 3. ابعاد داده‌ی آموزشی قبل از بسط دادن 7

شکل 4. ابعاد داده‌ی آموزشی پس از بسط دادن 7

شکل 5. نمودار مقادیر ویژه بر حسب componentها 7

شکل 6. ابعاد داده ها پس از اعمال الگوریتم PCA 7

شکل 7. ابعاد داده ها پس از اعمال الگوریتم LDA 8

شکل 9. دقت آموزش مدل SVM توسط 3 کرنل مختلف 8

شکل 10. نتایج کرنل rbf 8

شکل 11. ماتریس آشفتگی کرنل rbf 9

شکل 12. نمودار ROC 9

شکل 13. دقت آموزش مدل SVM توسط 3 کرنل مختلف (LDA) 10

شکل 14. نتایج کرنل rbf (LDA) 10

شکل 15. ماتریس آشفتگی کرنل rbf (LDA) 10

شکل 16. نمودار ROC (LDA) 11

شکل 17. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 92.25% 11

شکل 18. ماتریس آشفتگی الگوریتم RF 12

شکل 19. نمودار ROC الگوریتم RF 12

شکل 20. نتایج طبقه بندی با الگوریتم Logistic regression با دقت نهایی 88.88% 13

شکل 21. ماتریس آشفتگی الگوریتم Logistic Regression 13

شکل 22. مقایسه سه الگوریتم استفاده شده برای طبقه بندی صدای مرد و زن 14

**جدول‌ها**

[جدول 1. عنوان جدول نمونه **Error! Bookmark not defined.**](#_Toc115453324)

# **قسمت 1**. **پیش پردازش داده ها**

* توضیح مختصر درباره نحوه کارکرد و علت استفاده از روشهایی که برای قسمت پیش پردازش

## 1.1. تمیزکردن داده ها

* + کم کردن نویز داده های صوتی

## 2.1. استخراج ویژگی داده ها

* + استخراج ویژگی های زمانی و فرکانسی صوت ها و دلیل انتخاب این ویژگی ها

# **قسمت ۲**. **طبقه‌بندی داده ها**

در بخش قبل مراحل تمیز کردن دیتا (Data cleaning) و همچنین استخراج ویژگی از روی داده های صوتی بیان شد. در این بخش قصد داریم تا از ویژگی هایی که در بخش قبل استخراج شد، برای طبقه بندی داده داده های صوتی در دو طبقه "Male" و "Female" استفاده کنیم.

طبقه بندی نوعی از Binary Classification است که راه های مختلفی برای جداسازی داده های مرد و زن می توان استفاده کرد. از بهترین آنها می توان به روش های زیر اشاره کرد:

* SVM
* Random Forest
* KNN
* Decision Tree

## 1.2. مدل های پیشنهادی

در دانشگاه VIT پژوهش انجام شد برای طبقه بندی داده های صوتی زن و مرد. در این مقاله از چندین روش برای طبقه بندی استفاده شد که دو تا از بهترین روش هایی که گزارش داده است، SVM و Random Forest است. در واقع SVM بیشتر دقت و RF بیشترین مقاومت (Robustness) را در میان سایر روش ها داشته است[1].

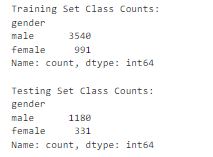
## 2.2. روش های بهبود نتیجه

در درس، روش هایی برای بهبود نتایج طبقه بندی تدریس شد مانند کاهش بعد (Dimensionality reduction) که یک ابزار ارزشمند برای بهبود عملکرد مدل است که از ویژگی های به نسبت کم ارزش چشم‌پوشی می کند و همچنین نرمالایز کردن(Normalization) یا اسکیل کردن فیچرها که یکی از مراحل مهم پیش پردازش است و برای مدل هایی که به اندازه فیچرهایی ورودی حساس هستند حیاتی است. در زیر توضیح مختصری در مورد نحوه عملکرد روش های بهبود نتیجه‌ی ذکر شده داده می شود.

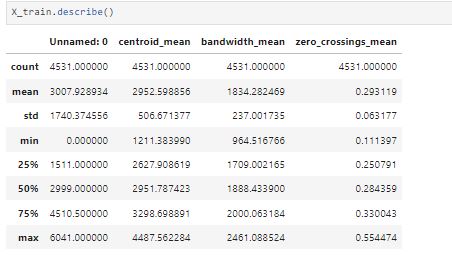
* + کاهش بعد: از PCA برای کاهش بعد استفاده شده است. بدین ترتیب که تعداد فیچر ها به 20 فیچر کاهش یافته است.
  + کاهش بعد: علاوه بر روش قبلی، می توان از تکنیک LDA (Linear Discremenant analysis) نیز استفاده کرد.
  + نرمالایز کردن: با استفاده از کتابخانه StandardScaler، مقادیر فیچر ها نرمالایز شده اند تا در ادامه بتوان از آن ها برای یادگیری مدل SVM و RF استفاده کرد.

فایل transcripts\_features.csv توسط کتابخانه pandas خوانده می شود. 5 ردیف اول و همچنین توصیفات این فایل در زیر قابل مشاهده است.

داده ها به دو دسته train و test و با نسبت 75%-25% تقسیم می شوند طوری که هیچکدام از این دو دسته نسبت به داده های male و female بایاس نباشند. در واقع چون تعداد داده های male بیشتر است امکان دارد این اتفاق بیافتد.

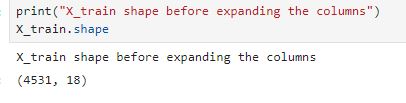


شکل 1. تعداد نمونه های مرد و زن در داده‌های آموزشی و تست

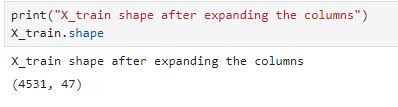


شکل 2. توصیف داده‌های آموزشی

در ادامه، فیچرهایی که شامل چند مقدار بوده اند بسط داده شده اند. به طور مثال برای ستون mfcc\_mean که شامل یک لیست با 13 مقدار است، 13 ستون مجزا در نظر گرفته شده است تا بتوان در ادامه آن ها را اسکیل کرد.



شکل 3. ابعاد داده‌ی آموزشی قبل از بسط دادن



شکل 4. ابعاد داده‌ی آموزشی پس از بسط دادن

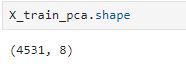
همانطور که مشاهده می شود، ابعاد هر نمونه آموزشی زیاد است. بنابراین به کمک الگوریتم های کاهش بعد یعنی PCA و LDA اقدام به کاهش شده و در ادامه نتایج هر الگوریتم طبقه بندی برای داده‌های کاهش بعد یافته گزارش می شود.

برای یافتن بهترین تعداد component در الگوریتم PCA، به دنبال بیشینه کردن واریانس هستیم. برای بیشینه کردن واریانس، بردارهایی مفید هستند که مقدار ویژه متناظر با آن ها بیشینه باشند. نمودار مقادیر ویژه را به ازای component ها رسم می شود.



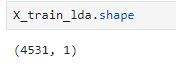
شکل 5. نمودار مقادیر ویژه بر حسب componentها

نمودار در n=8 شکستگی دارد بنابراین از همین عدد به عنوان آرگومان ورودی الگوریتم PCA استفاده می کنیم.



شکل 6. ابعاد داده ها پس از اعمال الگوریتم PCA

همچنین از LDA نیز برای کاهش بعد داده ها استفاده شده است.

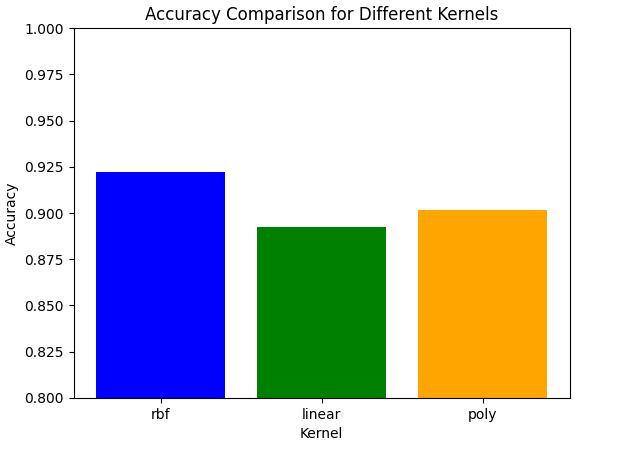


شکل 7. ابعاد داده ها پس از اعمال الگوریتم LDA

1. **طبقه بندی با الگوریتم SVM**

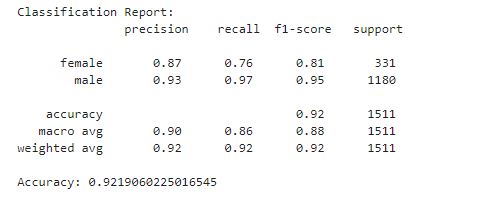
مدل اول برای طبقه بندی صداهای مرد و زن توسط الگوریتم SVM(Support Vector Machine) آموزش داده شد. نتیجه دقت آموزش این مدل برای سه کرنل مختلف یعنی RBF، خطی و چند جمله ای در زیر قابل مشاهده است.

در این بخش ابعاد ویژگی از 40 تا با کمک PCA به 8 فیچر کاهش داده شده است.



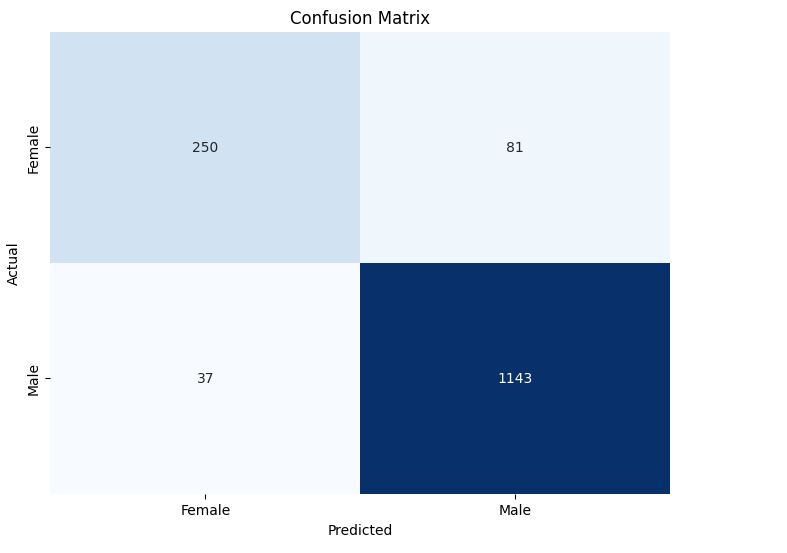
شکل 9. دقت آموزش مدل SVM توسط 3 کرنل مختلف

مشاهده می شود که کرنل rbf بیشترین دقت یعنی 92% را داشته است که مقادیر f1-score، recall، و صحت آن در شکل زیر نیز قابل مشاهده است.



شکل 10. نتایج کرنل rbf

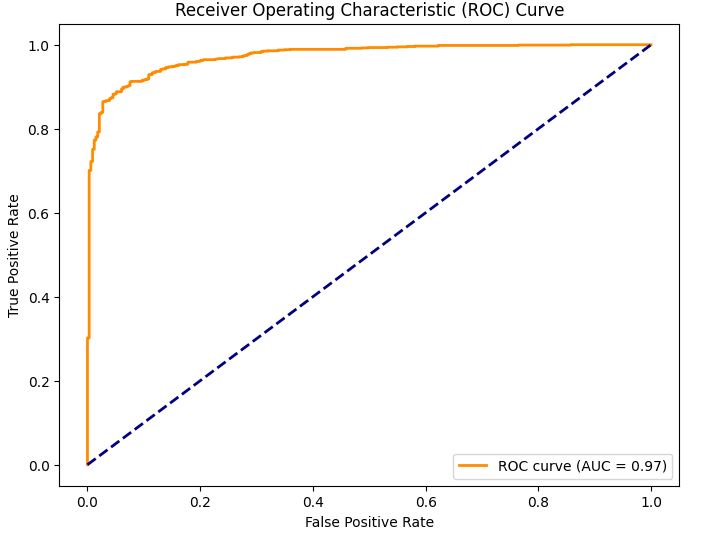
ماتریس آشفتگی مدل SVM با کرنل rbf در شکل زیر قابل مشاهده است.



شکل 11. ماتریس آشفتگی کرنل rbf

ماتریس آشفتگی دقت بسیار بالای طبقه بندی صدای مرد را گزارش می دهد که از 1180 نمونه تنها 37 عدد از آنها را در دسته‌ی صدای زن قرار داده است. اما در طبقه بندی صدای زن، در 331 نمونه، 81 مورد را صدای مرد تشخیص داده است.

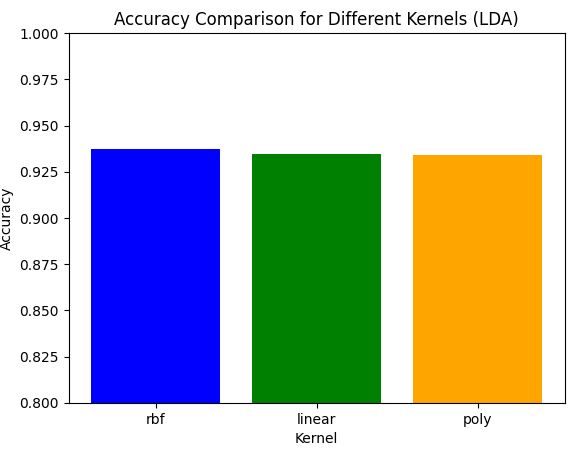
نمودار ROC در تصویر زیر قابل مشاهده است.



شکل 12. نمودار ROC

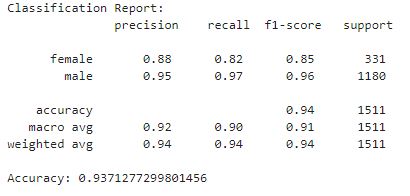
سطح زیر نمودار ROC نزدیک حدود 0.92 است که نشان از عملکرد نسبتا خوب مدل دارد.

در ادامه، ابعاد فیچرها به کمک تکنیک LDA کاهش داده شده و سپس نتایج ارائه می شوند. تعداد component ها در تکنیک LDA، باید از min(features, classes-1) کمتر باشد بنابراین در این بخش آرگومان n\_components=None در نظر گرفته شده است.

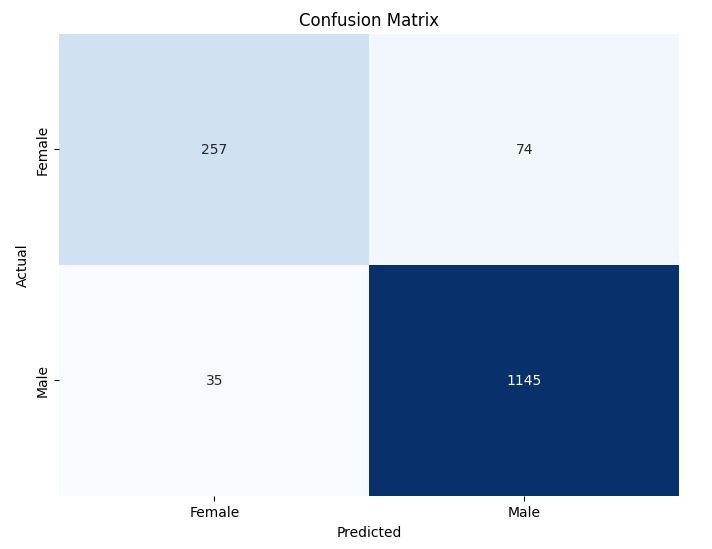


شکل 13. دقت آموزش مدل SVM توسط 3 کرنل مختلف (LDA)

در اینجا نیز مانند حالت قبل، کرنل rbf بهترین عملکرد را داشته است که دقت و ماتریس آشفتگی آن در شکل زیر قابل مشاهده است.



شکل 14. نتایج کرنل rbf (LDA)



شکل 15. ماتریس آشفتگی کرنل rbf (LDA)

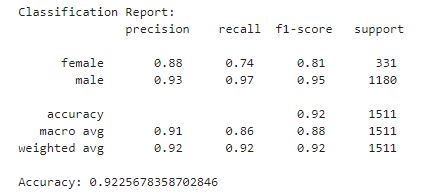


شکل 16. نمودار ROC (LDA)

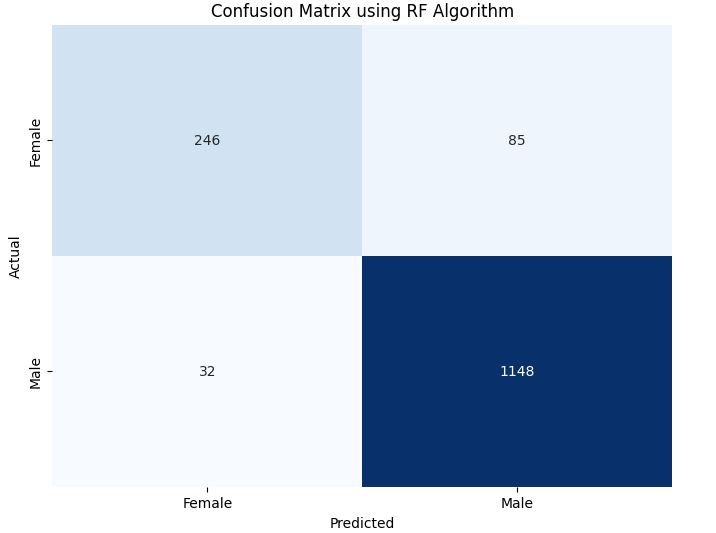
1. **طبقه بندی با الگوریتم Random Forest**

در این بخش به طبقه بندی با استفاده از الگوریتم RF می پردازیم.

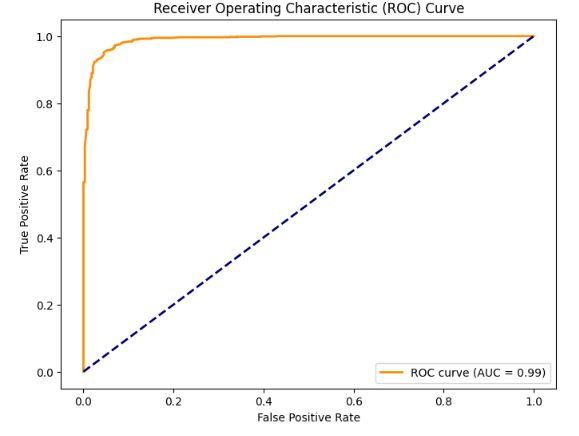
نتایج در ادامه قابل مشاهده است. در نتایج زیر، از PCA برای کاهش ابعاد فیچر ها استفاده شده است.



شکل 17. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 92.25%



شکل 18. ماتریس آشفتگی الگوریتم RF



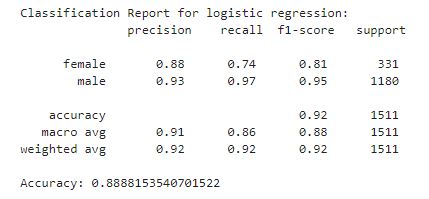
شکل 19. نمودار ROC الگوریتم RF

1. **طبقه بندی با الگوریتم Logistic Regression**

در این بخش به طبقه بندی با استفاده از الگوریتم Logistic Regression می پردازیم.

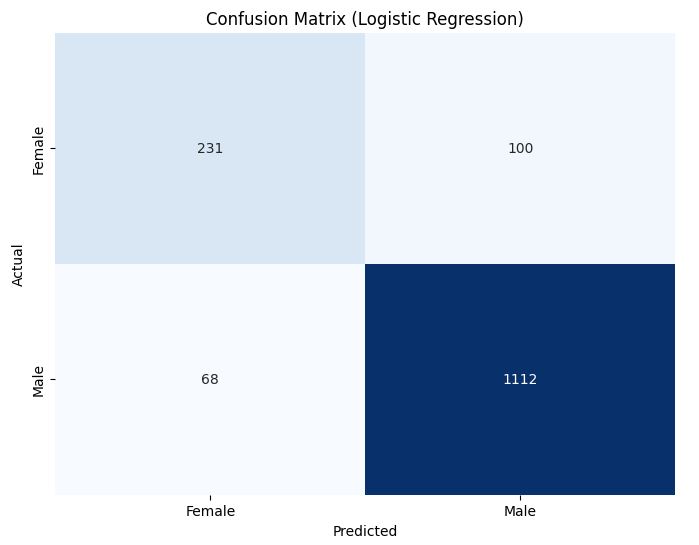
در این بخش نیز از داده هایی استفاده می شود که توسط الگوریتم PCA کاهش بعد داده شده اند.

نتایج در ادامه قابل مشاهده است.



شکل 20. نتایج طبقه بندی با الگوریتم Logistic regression با دقت نهایی 88.88%

در صورتی که از داده های اصلی (داده هایی که بعد آنها کاهش داده نشده اند) استفاده میشد، دقت حدود 92% است.

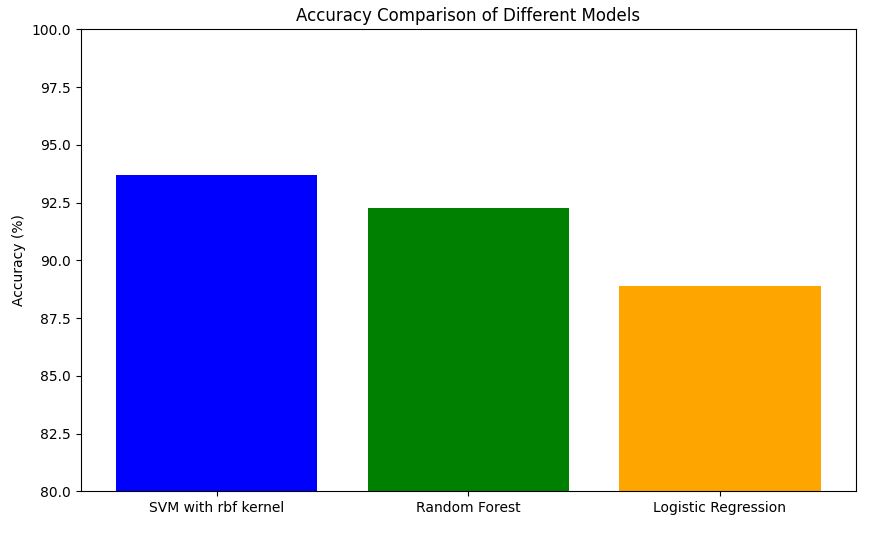


شکل 21. ماتریس آشفتگی الگوریتم Logistic Regression

## 3.2. نتیجه‌گیری و مقایسه

در این بخش از 3 الگوریتم مختلف برای طبقه بندی صوت استفاده شده است.

در تصویر زیر مقایسه دقت این سه روش قابل مشاهده است.



شکل 22. مقایسه سه الگوریتم استفاده شده برای طبقه بندی صدای مرد و زن

همانطور که در شکل بالا مشخص است، هر سه مدل دقت خوبی داشته اند اما دقت الگوریتم SVM با کرنل rbf کمی دقت بالاتری(93%) داشته است.

همچنین از ماتریس های آشفتگی مشخص است که هر سه الگوریتم در طبقه بندی صدای Female اشتباه بیشتری در طبقه بندی داشته اند. یکی از دلایل آن می تواند ناشی از بیشتر بودن تعداد داده ی صوتی Male باشد. اگرچه با تقسیم مناسب داده ها در داده های آموزشی و تست، سعی شد از بایاس شدن نتیجه جلوگیری شود. تعداد داده های اصلی male، حدود 3.5 برابر تعداد داده های female بود. با این حال هر 3 الگوریتم عملکرد مناسبی داشتند.

همچنین استفاده از الگوریتم های کاهش بعد مانند LDA و PCA منجر به کاهش دقت شد چون از تمامی ویژگی های صوتی استفاده نشده است. اما با این حال هزینه محاسباتی و همچنین زمان یادگیری مدل را کاهش داده است.

# **قسمت ۳** **– خوشه بندی داده ها**

## 1.3. مدل های پیشنهادی

* + توضیح مختصر درباره علت انتخاب مدل‌های خوشه‌بندی

## 2.3. روش های بهبود نتیجه

* + توضیح مختصر درباره علت و نتیجه استفاده از روش‌های تدریس شده در درس مانند روش‌های کاهش بعد و normalizationدر جهت بهبود کارایی مدل های خوشه بند
  + استفاده از روش هایی مثل ،silhouette score نمودار پراکندگی خوشه ها را به ازای تعداد خوشه های مختلف رسم کرده و یافتن تعداد خوشه مناسب

## 3.3. نتیجه گیری و مقایسه

* + ذکر میزان پراکندگی درون خوشه ای و میان خوشه ای و همچنین تحلیل اینکه هر خوشه نماینده چه دسته ای از داده هاست (ویژگی های مشترک داده های درون هر خوشه)، برای مدل های خوشه‌بند و به ازای هر تعداد خوشه
  + به ازای 2مقدار دلخواه دیگر برای تعداد خوشه
  + برای هر یک از تعداد خوشه های انتخاب شده، شباهت داده های درون یک خوشه و تفاوت بین خوشه ها و دلایلی که فکر می کنید برخی داده ها در یک خوشه قرار گرفته‌اند باید بررسی دقیق شوند

# مدل **(ASR) Automatic Speech Recognition**

* + میزان خطای مدل ASR و بررسی همبستگی این خطا به ویژگی ها
  + تحلیل همبستگی احتمالی میزان خطا با ویژگی های هر صوت (جنسیت، لهجه و لحن)
  + میزان خطای مدل را بر اساس آماره WER7را ذکر کرده
  + همبستگی میزان خطا را با ویژگی هایی نظیر سرعت صحبت کردن و نویز محیطی را نیز به صورت کیفی بررسی کرده

**مراجع:**

[1] Raahul, A., Sapthagiri, R., Pankaj, K., & Vijayarajan, V. (2017). Voice based gender classification using machine learning. In IOP Conference Series: Materials Science and Engineering (Vol. 263, p. 042083). IOP Publishing. https://doi.org/10.1088/1757-899x/263/4/042083