



به نام خدا  
دانشگاه تهران  
دانشکده مهندسی برق و کامپیوتر



## درس یادگیری ماشین گزارش پروژه پایانی

نام و نام خانوادگی	امیرحسین پورداد - امیرغرقابی - محمد ابوذری - مهدی سلیمانی زادگان
شماره دانشجویی	۸۱۰۱۰۱۱۲۰ - ۸۱۰۱۰۲۲۱۷ - ۸۱۰۱۰۱۰۸۷ - ۸۱۰۱۰۲۱۷۶
تاریخ ارسال گزارش	۱۴۰۲، ۱۱، ۰۴

## فهرست

قسمت ۱. پیش پردازش داده ها.....	۴
۱,۱. تمیزکردن داده ها.....	۴
۲,۱. استخراج ویژگی داده ها.....	۴
قسمت ۲. طبقه‌بندی داده ها.....	۵
۱,۲. مدل های پیشنهادی.....	۵
۲,۲. روش های بهبود نتیجه.....	۵
۳,۲. نتیجه‌گیری و مقایسه.....	۱۳
15..... قسمت ۳ - خوشه بندی داده ها	
۱,۳. مدل های پیشنهادی.....	۱۵
۲,۳. روش های بهبود نتیجه.....	۱۵
۳,۳. نتیجه‌گیری و مقایسه.....	۱۵
مدل (ASR) Automatic Speech Recognition.....	۱۶

## شکل‌ها

- شکل ۱. تعداد نمونه های مرد و زن در داده‌های آموزشی و تست..... ۶
- شکل ۲. توصیف داده‌های آموزشی..... ۶
- شکل ۳. ابعاد داده‌ی آموزشی قبل از بسط دادن..... ۷
- شکل ۴. ابعاد داده‌ی آموزشی پس از بسط دادن..... ۷
- شکل ۵. نمودار مقادیر ویژه بر حسب component ها..... ۷
- شکل ۶. ابعاد داده ها پس از اعمال الگوریتم PCA..... ۷
- شکل ۷. ابعاد داده ها پس از اعمال الگوریتم LDA..... ۸
- شکل ۹. دقت آموزش مدل SVM توسط ۳ کرنل مختلف..... ۸
- شکل ۱۰. نتایج کرنل rbf..... ۸
- شکل ۱۱. ماتریس آشفتگی کرنل rbf..... ۹
- شکل ۱۲. نمودار ROC..... ۹
- شکل ۱۳. دقت آموزش مدل SVM توسط ۳ کرنل مختلف (LDA)..... ۱۰
- شکل ۱۴. نتایج کرنل rbf (LDA)..... ۱۰
- شکل ۱۵. ماتریس آشفتگی کرنل rbf (LDA)..... ۱۰
- شکل ۱۶. نمودار ROC (LDA)..... ۱۱
- شکل ۱۷. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 92.25%..... ۱۱
- شکل ۱۸. ماتریس آشفتگی الگوریتم RF..... ۱۲
- شکل ۱۹. نمودار ROC الگوریتم RF..... ۱۲
- شکل ۲۰. نتایج طبقه بندی با الگوریتم Logistic regression با دقت نهایی 88.88%..... ۱۳
- شکل ۲۱. ماتریس آشفتگی الگوریتم Logistic Regression..... ۱۳
- شکل ۲۲. مقایسه سه الگوریتم استفاده شده برای طبقه بندی صدای مرد و زن..... ۱۴

## جدول‌ها

جدول ۱. عنوان جدول نمونه.....**Error! Bookmark not defined.**

## قسمت ۱. پیش پردازش داده ها

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

### ۱.۱. تمیز کردن داده ها

○ کم کردن نویز داده های صوتی

### ۲.۱. 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 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

### ۱.۲. 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].

### ۲.۲. 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 چشم پوشی 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 نتیجه ی ذکر شده داده می شود.

- کاهش بعد: از PCA برای کاهش بعد استفاده شده است. بدین ترتیب که تعداد فیچر ها به ۲۰ 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. ۵ ردیف اول و همچنین توصیفات این فایل در زیر قابل مشاهده است.

داده ها به دو دسته train و test و با نسبت ۷۵٪-۲۵٪ 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.

```
Training Set Class Counts:
gender
male      3540
female    991
Name: count, dtype: int64

Testing Set Class Counts:
gender
male      1180
female    331
Name: count, dtype: int64
```

شکل ۱. تعداد نمونه های مرد و زن در داده های آموزشی و تست

X_train.describe()				
	Unnamed: 0	centroid_mean	bandwidth_mean	zero_crossings_mean
count	4531.000000	4531.000000	4531.000000	4531.000000
mean	3007.928934	2952.598856	1834.282469	0.293119
std	1740.374556	506.671377	237.001735	0.063177
min	0.000000	1211.383990	964.516766	0.111397
25%	1511.000000	2627.908619	1709.002165	0.250791
50%	2999.000000	2951.787423	1888.433900	0.284359
75%	4510.500000	3298.698891	2000.063184	0.330043
max	6041.000000	4487.562284	2461.088524	0.554474

شکل ۲. توصیف داده های 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 ۱۳ مقدار است، ۱۳ ستون مجزا در نظر گرفته شده است تا بتوان در ادامه آن ها را اسکیل کرد.

$(4531, 18)$ 

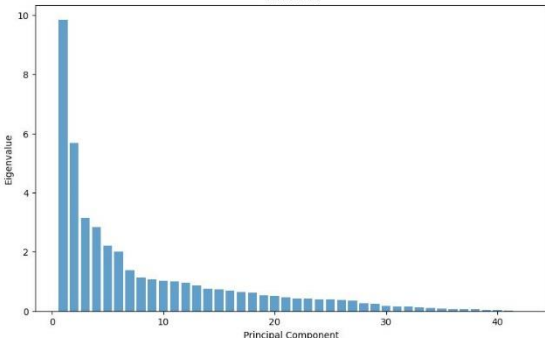
شکل ۳. ابعاد داده‌ی آموزشی قبل از بسط دادن

 $(4531, 47)$ 

شکل ۴. 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 داده های 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 ۵. نمودار مقادیر ویژه بر حسب componentها

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

 $(4531, 8)$ 

شکل ۶. ابعاد داده ها پس از اعمال الگوریتم PCA

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



```
X_train_lda.shape
```

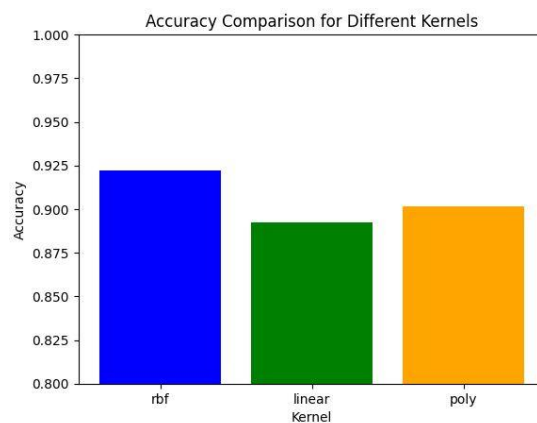
```
(4531, 1)
```

شکل ۷. 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 ۴۰ تا با کمک PCA به ۸ فیچر کاهش داده شده است.



شکل ۸. دقت آموزش مدل SVM توسط ۳ کرنل مختلف

مشاهده می شود که کرنل **rbf** بیشترین دقت یعنی ۹۲٪ را داشته است که مقادیر **recall**، **f1-score**، و صحت آن در شکل زیر نیز قابل مشاهده است.

```
Classification Report:
      precision    recall  f1-score   support

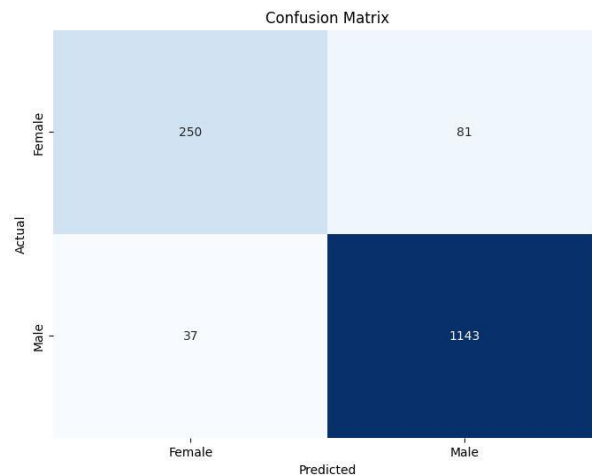
 female      0.87      0.76      0.81       331
    male      0.93      0.97      0.95      1180

 accuracy      0.92      0.92      0.92      1511
 macro avg      0.90      0.86      0.88      1511
weighted avg      0.92      0.92      0.92      1511

Accuracy: 0.9219060225016545
```

شکل ۹. نتایج کرنل **rbf**

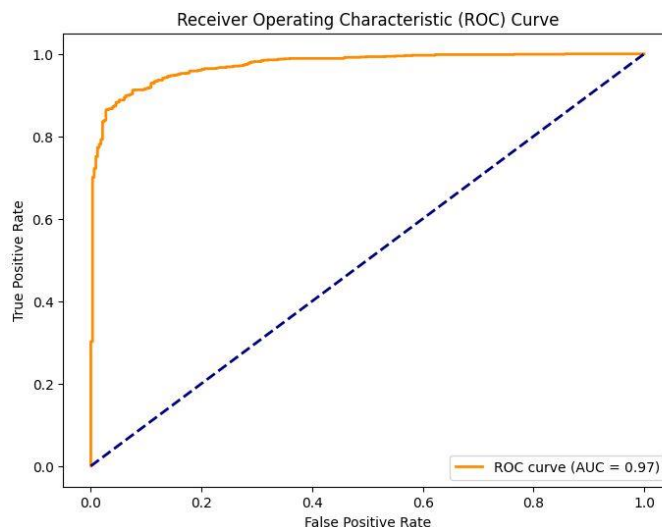
ماتریس آشفتگی مدل SVM با کرنل rbf در شکل زیر قابل مشاهده است.



شکل ۱۰. ماتریس آشفتگی کرنل **rbf**

ماتریس آشفتگی دقت بسیار بالای طبقه بندی صدای مرد را گزارش می دهد که از ۱۱۸۰ نمونه تنها ۳۷ عدد از آنها را در دسته ی صدای زن قرار داده است. اما در طبقه بندی صدای زن، در ۳۳۱ نمونه، ۸۱ مورد را صدای مرد تشخیص داده است.

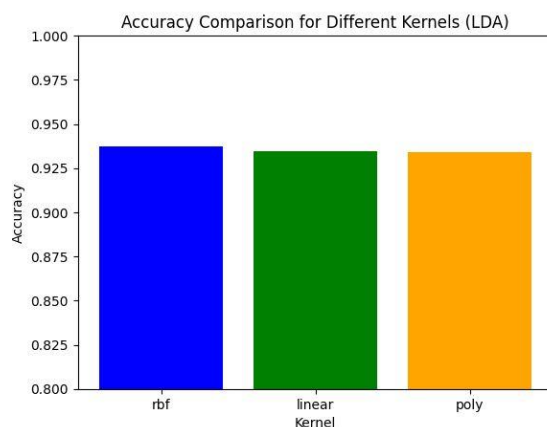
نمودار ROC در تصویر زیر قابل مشاهده است.



شکل ۱۱. نمودار ROC

سطح زیر نمودار ROC نزدیک حدود ۰٫۹۲ است که نشان از عملکرد نسبتاً 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  $\min(\text{features}, \text{classes}-1)$  کمتر باشد بنابراین در این بخش آرگومان  $n\_components=None$  در نظر گرفته شده است.



شکل ۱۲. دقت آموزش مدل SVM توسط ۳ کرنل مختلف (LDA)

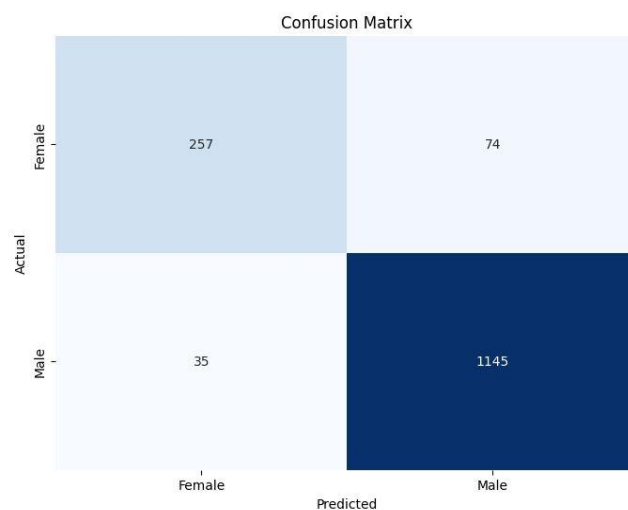
در اینجا نیز مانند حالت قبل، کرنل rbf بهترین عملکرد را داشته است که دقت و ماتریس آشفته‌گی آن در شکل زیر قابل مشاهده است.

Classification Report:

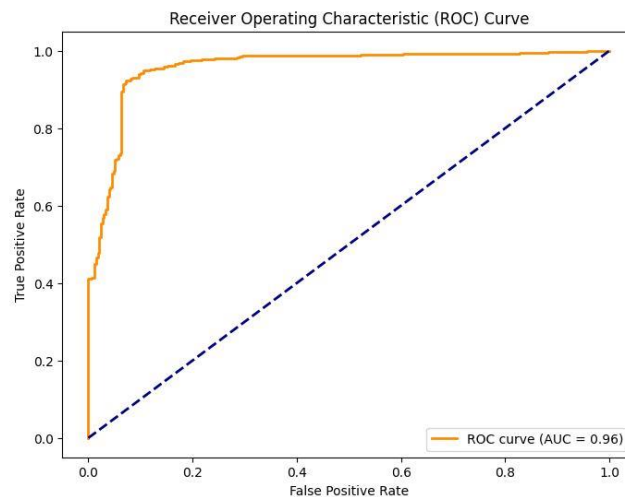
	precision	recall	f1-score	support
female	0.88	0.82	0.85	331
male	0.95	0.97	0.96	1180
accuracy			0.94	1511
macro avg	0.92	0.90	0.91	1511
weighted avg	0.94	0.94	0.94	1511

Accuracy: 0.9371277299801456

شکل ۱۳. نتایج کرنل rbf (LDA)



شکل ۱۴. ماتریس آشفته‌گی کرنل rbf (LDA)



شکل ۱۵. نمودار ROC (LDA)

## 2. 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.

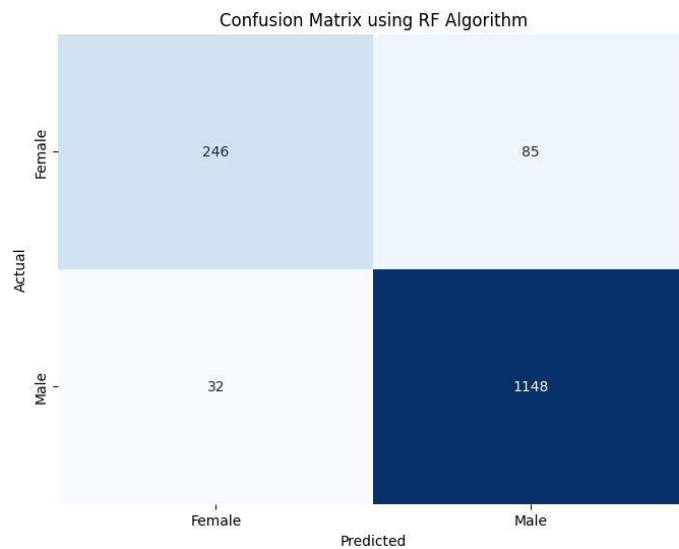
```
Classification Report:
      precision    recall  f1-score   support

 female      0.88      0.74      0.81       331
  male      0.93      0.97      0.95      1180

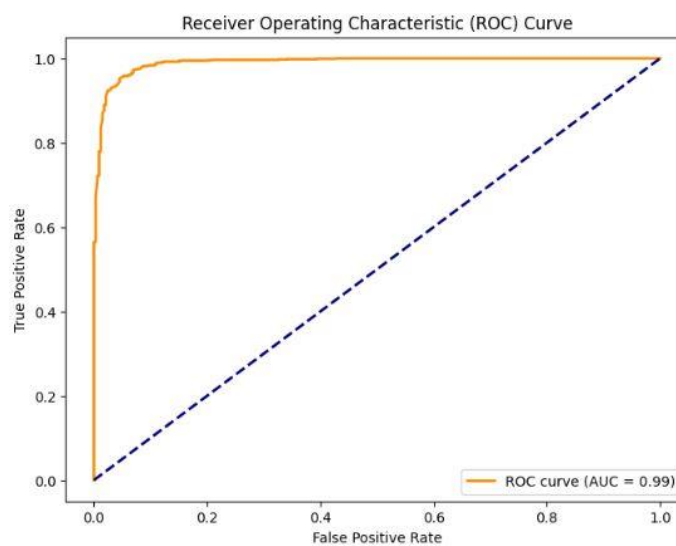
 accuracy      0.92      0.92      0.92      1511
 macro avg      0.91      0.86      0.88      1511
weighted avg      0.92      0.92      0.92      1511

Accuracy: 0.9225678358702846
```

شکل ۱۶. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 92.25%



شکل ۱۷. ماتریس آشفتگی الگوریتم RF



شکل ۱۸. نمودار ROC الگوریتم RF

### 3. 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.

```

Classification Report for logistic regression:
              precision    recall  f1-score   support

    female       0.88       0.74       0.81         331
    male         0.93       0.97       0.95        1180

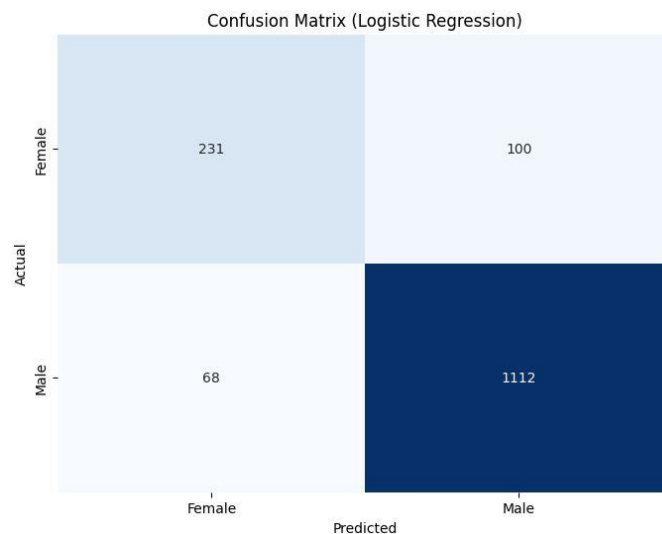
 accuracy              0.92         1511
 macro avg       0.91       0.86       0.88         1511
 weighted avg    0.92       0.92       0.92         1511

Accuracy: 0.8888153540701522

```

شکل ۱۹. 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 ۹۲٪ است.

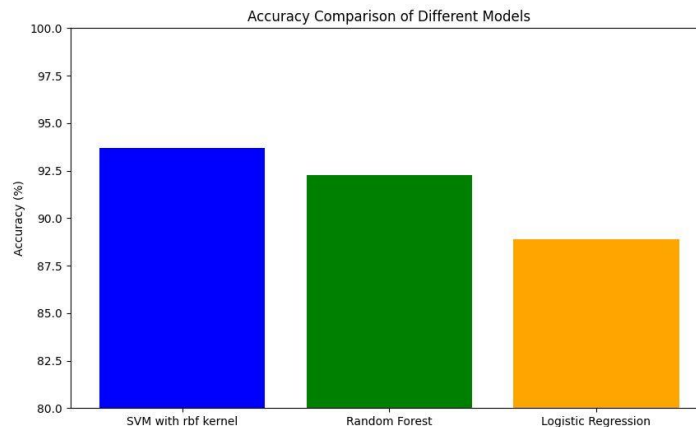


شکل ۲۰. ماتریس آشفتگی الگوریتم **Logistic Regression**

## ۳,۲. نتیجه گیری و مقایسه

در این بخش از ۳ الگوریتم مختلف برای طبقه بندی صوت استفاده شده است.

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



شکل ۲۱. 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 (۹۳٪) 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 ۳ 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

### ۱,۳. مدل های پیشنهادی

- توضیح مختصر درباره علت انتخاب مدل های خوشه بندی

### ۲,۳. روش های بهبود نتیجه

- توضیح مختصر درباره علت و نتیجه استفاده از روش های تدریس شده در درس مانند روش های کاهش بعد و normalization در 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

### ۳,۳. 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 خوشه بند و به ازای هر تعداد خوشه
- به ازای 2 مقدار 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 گرفته اند 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 WER7 را 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>