|  |  |
| --- | --- |
| 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** 5](#_Toc157103364)

[1.1. تمیزکردن داده ها 5](#_Toc157103365)

[1.1.1 :Resample 5](#_Toc157103366)

[2.1.1. Trim 5](#_Toc157103367)

[3.1.1. Normalization 5](#_Toc157103368)

[4.1.1. PreemPhasis 6](#_Toc157103369)

[2.1. استخراج ویژگی داده ها 7](#_Toc157103370)

[1.2.1. MCFF 7](#_Toc157103371)

[2.2.1. Chorma\_stft 7](#_Toc157103372)

[3.2.1. Spectral\_Contrast 8](#_Toc157103373)

[4.2.1. Spectral\_Centrolid 8](#_Toc157103374)

[5.2.1. Zero\_Crossing\_rate 9](#_Toc157103375)

[6.2.1. Piptrack 9](#_Toc157103376)

[7.2.1. Harmonic 9](#_Toc157103377)

[8.2.1. Energy 10](#_Toc157103378)

[9.2.1. Beat\_track 10](#_Toc157103379)

[**قسمت ۲**. **طبقه‌بندی داده ها** 11](#_Toc157103380)

[1.2. مدل های پیشنهادی 11](#_Toc157103381)

[2.2. روش های بهبود نتیجه 11](#_Toc157103382)

[3.2. نتیجه‌گیری و مقایسه 18](#_Toc157103383)

[**قسمت ۳** **– خوشه بندی داده ها** 20](#_Toc157103384)

[1.3. مدل های پیشنهادی 20](#_Toc157103385)

[2.3. روش های بهبود نتیجه 20](#_Toc157103386)

[3.3. نتیجه گیری و مقایسه 20](#_Toc157103387)

[مدل **(ASR) Automatic Speech Recognition** 21](#_Toc157103388)

**شکل‌ها**

شکل 1. تعداد نمونه های مرد و زن در داده‌های آموزشی و تست 12

شکل 2. توصیف داده‌های آموزشی 12

شکل 3. ابعاد داده‌ی آموزشی پس از بسط دادن 13

شکل 4. دقت آموزش مدل SVM توسط 3 کرنل مختلف 13

شکل 5. نتایج کرنل rbf 13

شکل 6. ماتریس آشفتگی کرنل rbf 14

شکل 7. نمودار ROC 14

شکل 8. دقت آموزش مدل SVM توسط 3 کرنل مختلف (LDA) 15

شکل 9. نتایج کرنل rbf (LDA) 15

شکل 10. ماتریس آشفتگی کرنل rbf (LDA) 15

شکل 11. نمودار ROC (LDA) 16

شکل 12. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 96.09% 16

شکل 13. ماتریس آشفتگی الگوریتم RF 17

شکل 14. نمودار ROC الگوریتم RF 17

شکل 15. نتایج طبقه بندی با الگوریتم Logistic regression با دقت نهایی 93.51% 18

شکل 16. ماتریس آشفتگی الگوریتم Logistic Regression 18

شکل 17. مقایسه سه الگوریتم استفاده شده برای طبقه بندی صدای مرد و زن 19

**جدول‌ها**

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

# **قسمت 1**. **پیش پردازش داده ها**

## 1.1. تمیزکردن داده ها

پیش پردازش داده‌ها در داده‌های صوتی تأثیر قابل توجهی دارد. این فرآیند شامل مجموعه‌ای از تکنیک‌ها و الگوریتم‌ها است که به منظور بهبود کیفیت و قابلیت استفاده از داده‌های صوتی انجام می‌شود. تاثیرات اصلی پیش پردازش داده‌های صوتی عبارتند از:

### 1.1.1 :Resample

ibrosa.resample یک تابع در کتابخانه‌ی librosa است که برای بازنمونده کردن یک سیگنال صوتی به یک نرخ نمونه‌برداری مختلف استفاده می‌شود. این تابع با گرفتن یک سیگنال صوتی و نرخ نمونه‌برداری مطلوب به عنوان ورودی، سیگنال صوتی بازنمونه‌برداری شده را برمی‌گرداند. استفاده از این تابع در وظایف پردازش صوتی متداول است که نرخ نمونه‌برداری برای تحلیل یا مقایسه‌ی بعدی نیاز به تغییر دارد.

### 2.1.1. Trim

تابع trim در پردازش داده‌های صوتی به معنی برش (کوتاه کردن) قسمت های ابتدایی و/یا انتهایی یک سیگنال صوتی است. این عملیات معمولاً برای حذف بخش‌های بی‌فایده یا ناکارآمد از سیگنال صوتی استفاده می‌شود و می‌تواند بهبودی در کارایی الگوریتم‌های پردازش صوتی و موارد دیگر ایجاد کند. عموماً می‌توان از تفاوت زمان شروع و پایانی یک سیگنال صوتی با توجه به معیارهای مشخصی (مانند آستانه بالا/پایین، توان، انرژی، آمارهای زمانی و فرکانسی و غیره) استفاده کرده و قسمت‌های غیرمربوط را حذف نمود. این عملیات برای حذف سکوت‌ها، نویزها و قسمت‌هایی از سیگنال صوتی که در تحلیل یا استفاده‌ی بعدی مورد نیاز نیستند، مفید می‌باشد.

### 3.1.1. Normalization

نرمال‌سازی یک تکنیک متداول در پیش پردازش داده‌های صوتی است که به هدف تغییر مقیاس و مقادیر سیگنال صوتی ورودی به یک محدوده استاندارد استفاده می‌شود، معمولاً بین 0 تا 1 یا -1 تا 1. این فرآیند به تطبیق مقیاس و دامنه مقادیر ویژگی‌ها یا مقادیر ورودی کمک می‌کند، که برای الگوریتم‌های یادگیری ماشین و وظایف پردازش سیگنال مفید است.

در این مسئله سعی شده است داده های صوتی با استفاده از فرمول (max-min) /X-min مقایس بندی شوند.

### 4.1.1. PreemPhasis

پری‌امفاسیس (Preemphasis) یک تکنیک است که در پیش پردازش داده‌های صوتی به کار می‌رود. این تکنیک برای بهبود بخش‌های خاص فرکانسی صدا و افزایش قابلیت درک و کیفیت کلی سیگنال استفاده می‌شود. هدف اصلی از پری‌امفاسیس، تاکید بر بخش‌های فرکانسی بالا و کاهش بخش‌های فرکانسی پایین می‌باشد.

در داده‌های صوتی، انرژی سیگنال معمولاً در محدوده‌ی فرکانسی پایین قرار دارد. با اعمال پری‌امفاسیس، ما از طریق تقویت بزرگی مولفه‌های فرکانسی بالا نسبت به مؤلفه‌های فرکانسی پایین، کیفیت سیگنال را بهبود می‌بخشیم. برای این کار، با اعمال فیلتر پاس‌بالا به سیگنال، فرکانس‌های بالاتر تقویت و فرکانس‌های کمتر ضعیف می‌شوند.

با اعمال پری‌امفاسیس ، می‌توانیم نسبت سیگنال به نویز را افزایش دهیم، تعادل طیفی را افزایش دهیم و عملکرد الگوریتم‌های پردازش گفتار بعدی مانند تشخیص گفتار یا شناسایی گوینده را بهبود بخشیم. اجزای فرکانس بالا تقویت شده می تواند به حفظ جزئیات مهم و کاهش تأثیر نویز یا اعوجاج در طول تجزیه و تحلیل گفتار یا انتقال کمک کند.

## 2.1. استخراج ویژگی داده ها

استخراج ویژگی‌ها در داده‌های صوتی یکی از مهمترین مراحل در پردازش سیگنال صوتی و برای دستیابی به نتایج دقیق در وظایف پردازش صوتی است. در واقع، استخراج ویژگی‌ها به معنای تبدیل سیگنال صوتی از فرمت زمانی به فرمت فضایی است، که داده‌های صوتی را قابل تحلیل و استفاده‌ی بیشتر می‌کند.

استخراج ویژگی‌ها به منظور بهبود دقت تشخیص گفتار و تشخیص صدا و همچنین بهبود عملکرد مدل‌های یادگیری ماشین مورد استفاده قرار می‌گیرد. بسته به وظایف مورد نظر، ویژگی‌های مختلفی مانند: طول پالس، نرخ تکرار و کوتاهی‌های متوسط سیگنال، انرژی و توان سیگنال، باندهای طیفی، شتاب، سرعت تغییر و ... استخراج می‌شوند. هر یک از این ویژگی‌ها، نشان دهنده‌ی خصوصیات خاصی از سیگنال صوتی هستند و می‌توانند به نتایج بهتر در وظایف تشخیص و تفکیک صدا، ترجمه صوتی و دیگر وظایف پردازش را عامل شود.

در زیر فیچر های مطرح شده در این پروژه را مطرح مینماییم

### 1.2.1. MCFF

MFCC یا مجموعه ویژگی‌های همینوسیکل فرکانسی، یک روش استخراج ویژگی از سیگنال صوتی است که معمولاً برای تشخیص و تمایز دادن بین صداها استفاده می‌شود. این روش بر اساس تجزیه سیگنال صوتی به بخش‌های مختلف با استفاده از نمونه‌های زمانی کوتاه تقسیم می‌شود. پس از تقسیم سیگنال، معیارهای آماری مانند انرژی، میانگین و واریانس برای هر بخش محاسبه شده و سپس با استفاده از تبدیل فوریه معیارهای زمانی به معیارهای فرکانسی تبدیل می‌شوند. در نهایت، ضرایب MFCC حاصل به عنوان بردار ویژگی برای تشخیص و تمایز دادن بین صداها استفاده می‌شوند.

### 2.2.1. Chorma\_stft

Chroma\_STFT یک روش استخراج ویژگی از سیگنال صوتی است که برای تشخیص و تمایز دادن بین آکوردهای موسیقی استفاده می‌شود. این روش بر اساس تبدیل فوریه کوتاه مدت زمانی (STFT) سیگنال صوتی استخراج می‌شود. در این روش، ابتدا سیگنال صوتی تقسیم می‌شود و برای هر بخش کوتاه زمانی STFT محاسبه می‌شود. سپس با استفاده از STFT، انرژی هر باند فرکانسی در هر بخش محاسبه می‌شود. در نهایت، از انرژی هر باند فرکانسی برای محاسبه ماتریس کروما استفاده می‌شود.

ماتریس کروما نشان می‌دهد که در هر بخش زمانی، آکوردهای مختلف موسیقی در چه میزان وجود دارند. این ماتریس به طور معمول از 12 ستون تشکیل شده است که هر ستون نمایانگر یک آکورد موسیقی است. ارزش هر سلول در ماتریس نشان دهنده حضور یا عدم حضور آکورد مربوطه در بخش زمانی موردنظر است. در کل، Chroma\_STFT یک روش مفید است که برای تحلیل و تمایز آکوردهای موسیقی در سیگنال صوتی استفاده می‌شود.

### 3.2.1. Spectral\_Contrast

Spectral\_Contrast یک روش استخراج ویژگی از سیگنال صوتی است که برای تشخیص تفاوت‌های طیفی در آن استفاده می‌شود. این روش بر اساس تبدیل فوریه کوتاه مدت زمانی (STFT) سیگنال صوتی استخراج می‌شود.

در این روش، ابتدا سیگنال صوتی تقسیم می‌شود و برای هر بخش کوتاه زمانی STFT محاسبه می‌شود. سپس با استفاده از STFT، طیف فرکانسی برای هر بخش محاسبه می‌شود. در Spectral\_Contrast، برای هر باند فرکانسی معیارهای مختلفی مانند میانگین و واریانس طیف در ناحیه آن باند محاسبه می‌شوند. سپس با استفاده از این معیارها، کنتراست طیفی بین باندهای مختلف محاسبه می‌شود.

مقادیر کنتراست طیفی نشان می‌دهد که در هر بخش زمانی، تفاوت‌های طیفی در باندهای مختلف برجسته است یا خیر. این ویژگی می‌تواند برای تشخیص و تمایز دادن بین سیگنال‌های صوتی با مشخصات طیفی متفاوت مانند صداهای موسیقی ، گفتار و غیره استفاده شود..

در کل، Spectral\_Contrast یک روش مفید است که برای تشخیص ویژگی‌های طیفی در سیگنال صوتی و تمایز دادن بین آنها استفاده می‌شود.

### 4.2.1. Spectral\_Centrolid

در مورد استخراج ویژگی های داده های صوتی، مرکز طیفی (Spectral Centroid) یک ویژگی است که به طور کلی استفاده می شود. این ویژگی نقطه مرکز ثقل طیف سیگنال صدا را نشان می دهد. برای محاسبه آن، از میانگین وزن داده های فرکانسی موجود در سیگنال استفاده می شود، که وزن ها توسط بردارهای میدان طیفی مربوطه تعیین می شود. مرکز طیفی اطلاعاتی درباره توزیع محتوای فرکانسی یک سیگنال صدا ارائه می دهد. مقدار بالاتر مرکز طیفی نشان دهنده این است که بیشتر از انرژی سیگنال در فرکانس های بالاتر تمرکز شده است، در حالی که مقدار پایین تر نشان دهنده تمرکز بر فرکانس های پایین تر است. مرکز طیفی می تواند به عنوان یک ویژگی در برنامه های مختلف مانند تشخیص گفتار، طبقه بندی سبک موسیقی و طبقه بندی رویدادهای صوتی مورد استفاده قرار گیرد. این ویژگی می تواند برخی از ویژگی های طیفی یک صدا را نشان دهد و به تمایز بین انواع مختلف صداها بر اساس محتوای فرکانسی آنها کمک کند.

سه فیچر مختلف را میتوان از این داده استخراج نمود که میانگین وزن های فرکانسی ، ماکزیمم آن ها و همچنین مقدار میانه این داده بعنوان سه فیچر مجزا برای این تحلیل به کار گرفته شده است.

### 5.2.1. Zero\_Crossing\_rate

Zero Crossing Rate (نرخ گذار صفر) یکی از ویژگی‌های استخراج شده از داده‌های صوتی است. این ویژگی نشان می‌دهد که چقدر سیگنال صوتی از صفر عبور می‌کند. وقتی که سیگنال صوتی از مثبت به منفی یا از منفی به مثبت تغییر می‌کند، یک گذار صفر رخ می‌دهد. نرخ گذار صفر می‌تواند اطلاعاتی درباره ریتم و تناوب سیگنال صوتی ارائه دهد. در سیگنال‌های با ریتم بالا، نرخ گذار صفر بیشتر می‌شود به این معنی که سیگنال صوتی بین مثبت و منفی بیشتر عوض می‌شود. در سیگنال‌هایی با ریتم کمتر، نرخ گذار صفر کمتر می‌شود زیرا سیگنال صوتی بین مثبت و منفی کمتر تغییر می‌کند. استفاده از نرخ گذار صفر در برنامه‌های پردازش صوتی می‌تواند در شناسایی الگوها، تشخیص گفتار، تشخیص خواننده یا سخنران و حتی تحلیل سیگنال‌های موسیقی مفید باشد.

### 6.2.1. Piptrack

Piptrack یک ویژگی در پردازش و تحلیل گفتار است که برای تخمین فرکانس بنیادی (یا پیچیدگی یا تن طبیعی) یک سیگنال صوتی استفاده می‌شود. فرکانس بنیادی نشان دهنده نت مورد شنیدار و تن طبیعی یک صدا است. Piptrack بر پایه مفهوم تحلیل هارمونیک استوار است، جایی که اجزای هارمونیکی در طیف فرکانسی یک سیگنال صوتی شناسایی می‌شوند. این روش، قله‌های طیفی یا پیک‌های موجود در دامنه فرکانس را تحلیل کرده و حرکت آنها در طول زمان را پیگیری می‌کند. با پیگیری این قله‌ها، Piptrack فرکانس بنیادی را تخمین می‌زند و اطلاعاتی درباره تغییرات تن طبیعی در سیگنال صوتی ارائه می‌دهد. این ویژگی رایج در وظایفی مانند تشخیص گفتار، سنتز گفتار و تحلیل موسیقی استفاده می‌شود. این ویژگی به استخراج اطلاعات مربوط به فرکانس بنیادی، شناسایی الگوهای ملودیک و تمایز صداها یا سازهای مختلف بر اساس فرکانس بنیادی آنها کمک می‌کند.

### 7.2.1. Harmonic

هارمونیک‌ها در تشخیص ویژگی‌های صدا در تحلیل گفتار نقش بسیار مهمی دارند. در گفتار، هارمونیک‌ها ضرب‌های صحیحی از فرکانس بنیادی (کمترین جزء فرکانسی) سیگنال صوتی هستند.

هارمونیک‌ها برای نشان دادن پوشش طیفی سیگنال صوتی استفاده می‌شوند و برای شناسایی تن (پیچیدگی) صدا بسیار حائز اهمیت هستند. حضور هارمونیک‌ها در سیگنال نشان می‌دهد که صدا دوره‌ای است، و عدم حضور هارمونیک‌ها نشان می‌دهد که صدا بی‌دوره یا نویزی است.

تشخیص هارمونیک‌ها در برنامه‌های مختلفی مانند تشخیص گفتار، تشخیص گوینده و تحلیل صدای خواننده استفاده می‌شود. این ویژگی در شناسایی فرکانس بنیادی، تعیین کیفیت صدا، و تشخیص حضور برباده‌های صدایی یا ناهنجاری‌های صدایی کمک می‌کند.

به طور خلاصه، تشخیص هارمونیک‌ها در تشخیص ویژگی‌های صدا بسیار حیاتی است زیرا اطلاعاتی درباره تن و پوشش طیفی سیگنال صوتی ارائه می‌دهند.

### 8.2.1. Energy

اندازه گیری میزان انرژی یک سیگنال دارای اطلاعات زیادی میباشد از جمله تشخیص تن صدا برای بیان یک جمله ، همچنین قدرت همچنین صدا های با انرژی بیشتر قابلیت این را دارند که بتوانند راحتتر تشخیص داده شوند و تاثیر نویز در آن ها کمتر بوده و بعنوان یک فیچر میتوان از آن ها استفاده نمود.

### 9.2.1. Beat\_track

Beat tracking یا تشخیص ضرب در موسیقی یک ویژگی از تحلیل صدا است که به کمک آن می‌توان فاصله زمانی بین ضربات را در موسیقی تشخیص داد. این ویژگی معمولاً در تحلیل پردازشی موسیقی بکار می‌رود. در ویژگی های صدا، تشخیص ضرب در موسیقی با استفاده از الگوریتم های مختلفی انجام می‌شود که توسط تحلیل موجک، تبدیل فوریه، تحلیل انرژی و غیره انجام می‌شوند.

تشخیص ضرب در موسیقی در ویژگی های صدا، همچنین می‌تواند در تحلیل ویژگی های صدای انسانی نیز مفید باشد. به عنوان مثال، با تشخیص ضرب در موسیقی با استفاده از روش های مختلف، می توان به بررسی تغییرات فاصله‌های ضربات در زمان در صداهای انسانی پرداخت.

از دیگر کاربردهای تشخیص ضرب در موسیقی به عنوان یک قابلیت صدای هوشمند در تلفن همراه و دستگاه های پخش موسیقی می توان اشاره کرد. با افزودن کاربرد تشخیص ضرب در موسیقی به این دستگاه ها، می توان به طور خودکار تنظیم های مختلف را برای پخش موسیقی انجام داد. به طور مثال، با تشخیص ضرب در موسیقی، دستگاه پخش موسیقی می تواند به طور خودکار حالت های مختلفی را اعم از پخش در حالت شلوغی یا آرامش و... به طور خودکار انتخاب کند.

# **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].

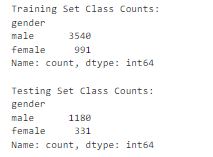
## 2.2. روش های بهبود نتیجه

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.

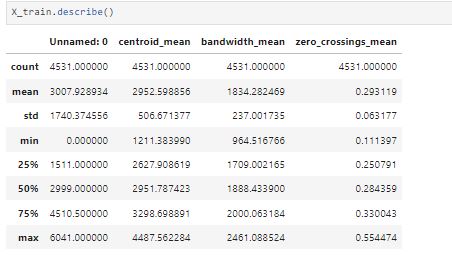
* + کاهش بعد: از PCA برای کاهش بعد استفاده شده است. بدین ترتیب که تعداد فیچر ها به 20 فیچر کاهش یافته است.
  + کاهش بعد: علاوه بر روش قبلی، می توان از تکنیک LDA (Linear Discremenant analysis) نیز استفاده کرد.
  + نرمالایز کردن: با استفاده از کتابخانه StandardScaler، مقادیر فیچر ها نرمالایز شده اند تا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

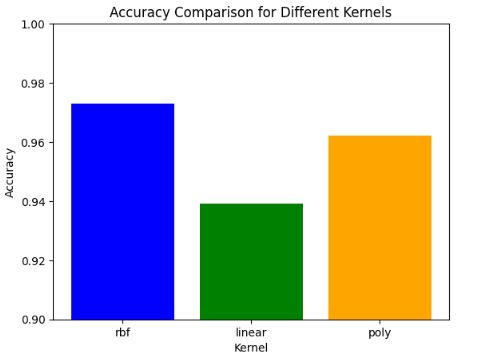


شکل 3. ابعاد داده‌ی آموزشی پس از بسط دادن

1. **طبقه بندی با الگوریتم SVM**

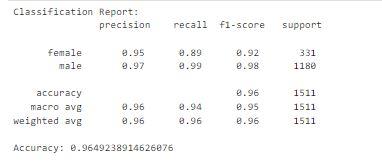
مدل اول برای طبقه بندی صداهای مرد و زن توسط الگوریتم SVM(Support Vector Machine) آموزش داده شد. نتیجه دقت آموزش این مدل برای سه کرنل مختلف یعنی RBF، خطی و چند جمله ای در زیر قابل مشاهده است.

در این بخش ابعاد ویژگی از 40 تا با کمک PCA به 20 فیچر کاهش داده شده است.



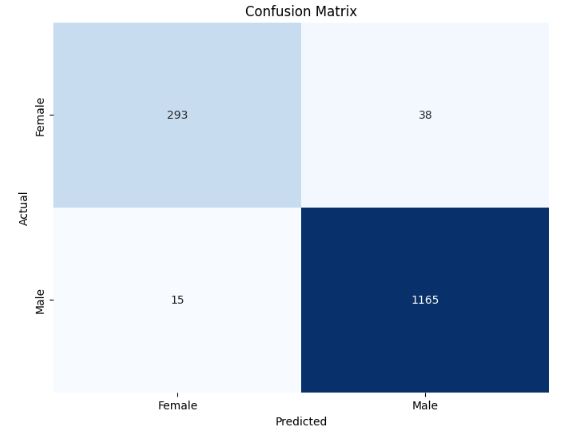
شکل 4. دقت آموزش مدل SVM توسط 3 کرنل مختلف

مشاهده می شود که کرنل rbf بیشترین دقت یعنی 96.49% را داشته است که مقادیر f1-score، recall، و صحت آن در شکل زیر نیز قابل مشاهده است.



شکل 5. نتایج کرنل rbf

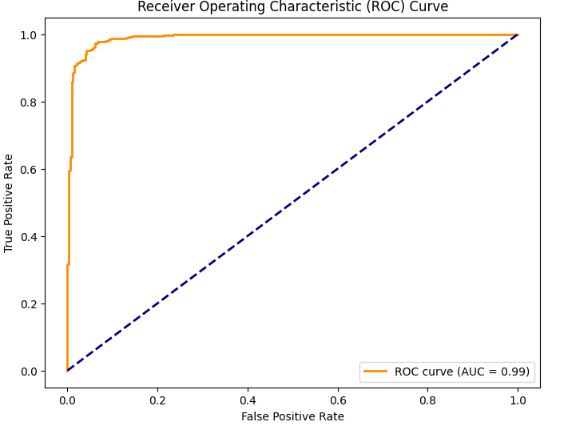
ماتریس آشفتگی مدل SVM با کرنل rbf در شکل زیر قابل مشاهده است.



شکل 6. ماتریس آشفتگی کرنل rbf

ماتریس آشفتگی دقت بسیار بالای طبقه بندی صدای مرد را گزارش می دهد که از 1180 نمونه تنها 15 عدد از آنها را در دسته‌ی صدای زن قرار داده است. اما در طبقه بندی صدای زن، در 331 نمونه، 38 مورد را صدای مرد تشخیص داده است.

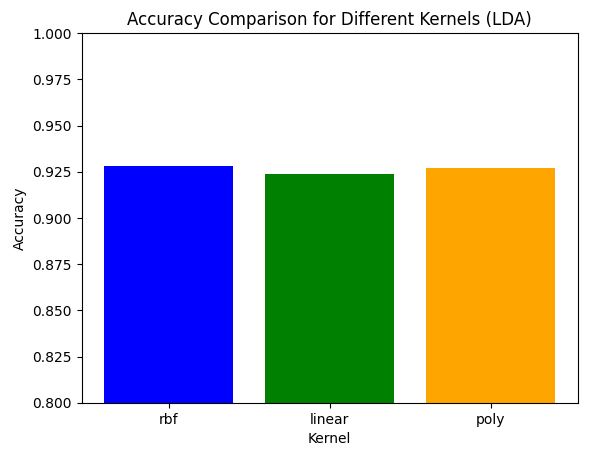
نمودار ROC در تصویر زیر قابل مشاهده است.



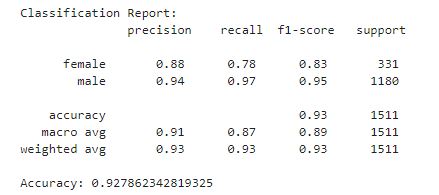
شکل 7. نمودار ROC

سطح زیر نمودار ROC نزدیک به 1 است که نشان از عملکرد خوب مدل دارد.

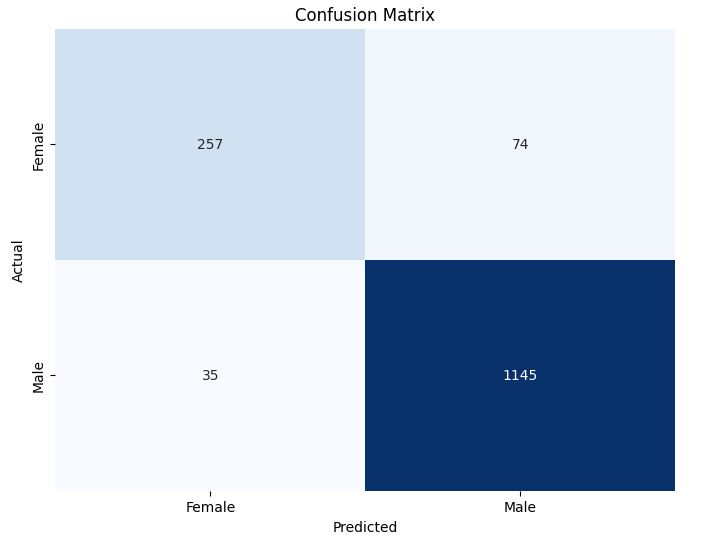
در ادامه، ابعاد فیچرها به کمک تکنیک LDA کاهش داده شده و سپس نتایج ارائه می شوند. تعداد component ها در تکنیک LDA، باید از min(features, classes-1) کمتر باشد بنابراین در این بخش آرگومان n\_components=None در نظر گرفته شده است.



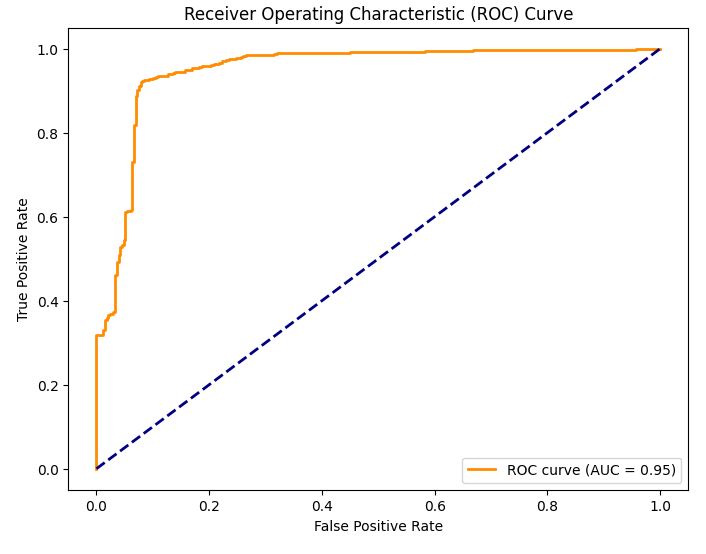
شکل 8. دقت آموزش مدل SVM توسط 3 کرنل مختلف (LDA)



شکل 9. نتایج کرنل rbf (LDA)



شکل 10. ماتریس آشفتگی کرنل rbf (LDA)

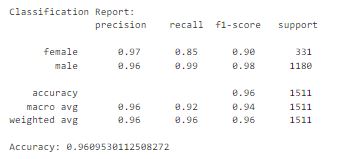


شکل 11. نمودار ROC (LDA)

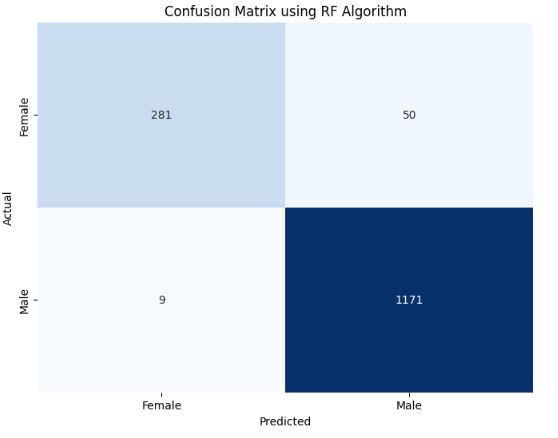
1. **طبقه بندی با الگوریتم Random Forest**

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

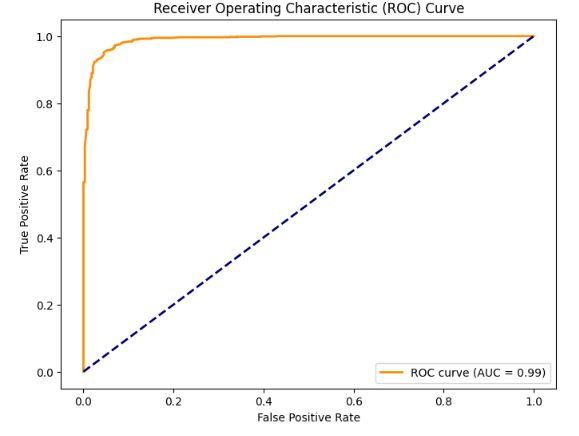
نتایج در ادامه قابل مشاهده است. در نتایج زیر، از PCA برای کاهش ابعاد فیچر ها استفاده شده است.



شکل 12. نتایج طبقه بندی با الگوریتم RF با دقت نهایی 96.09%



شکل 13. ماتریس آشفتگی الگوریتم RF

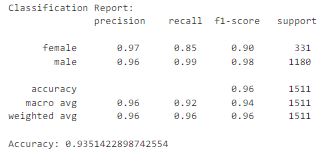


شکل 14. نمودار ROC الگوریتم RF

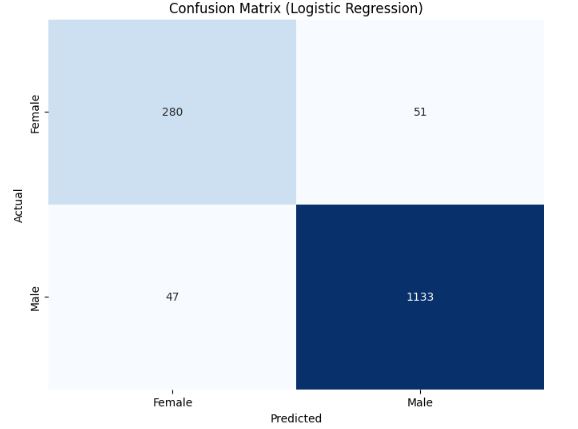
1. **طبقه بندی با الگوریتم Logistic Regression**

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

نتایج در ادامه قابل مشاهده است.



شکل 15. نتایج طبقه بندی با الگوریتم Logistic regression با دقت نهایی 93.51%

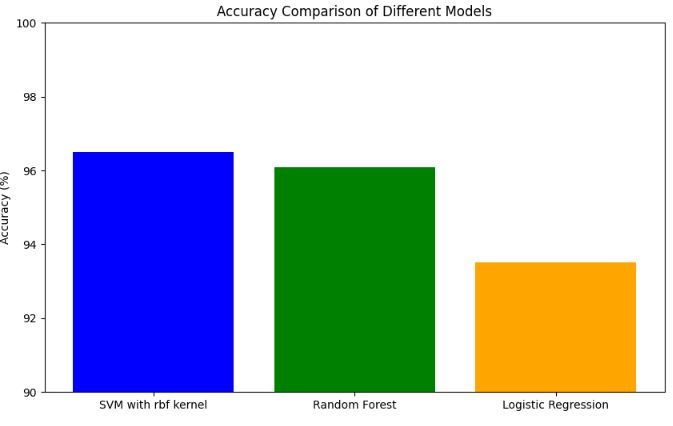


شکل 16. ماتریس آشفتگی الگوریتم Logistic Regression

## 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 17. 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(96%) 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.

# **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.