



دانشگاه صنعتی شریف

دانشکده مهندسی برق

بیواينسترومنت

پروژه‌ی امتیازی

دانشجو

سید ابوالفضل مرتضوی

۴۰۲۲۰۰۱۹۱

بهمن ۱۴۰۲

فهرست

بخش اول (دادگان شخصی).....	۴
۱-۱- توضیحات دادگان	۴
۲-۱- پیاده‌سازی	۵
۳-۱- ۵۹.۸۸	۵
۴-۱- ۶۴.۳۰	۶
بخش دوم	۷
۱-۲- توضیحات دادگان	۷
۲-۲- پیاده‌سازی	۷
۱-۲-۲- SVM(kernel=Poly)	۸
۲-۲-۲- KNN	۹
۳-۲-۲- Decision tree	۹
۴-۲-۲- Random Forest	۱۰
منابع	۱۲

توضیحات

با توجه به کم بودن نمره‌ی امتیازی پروژه، اعضای گروه علاقه‌ای به انجام آن نداشتند، از این رو این پروژه به تنهایی انجام شده است. در این مورد با آقای دکتر صحبت کردم و ایشان فرمودند در صورت انجام پروژه به صورت فردی، امکان دادن نمره‌ی کامل به ازای انجام بخشی از پروژه بررسی می‌شود. با این حال من تمام تلاشم را برای انجام هرچه کامل‌تر این پروژه انجام دادم. در این خصوص می‌توانید این مورد را از دکتر استعلام بفرمایید.

نکته‌ی دیگر این که این پروژه بر روی دو دیتاست مختلف که یکی از آن‌ها دیتاست شخصی است که توسط یکی از دوستان بنده جمع‌آوری شده است و دیگری دیتاستی از سایتی با لینک دسترسی زیر است. به علت بالا بودن حجم دادگان شخصی، این دادگان در فایل پیوست قرار داده نشدند.

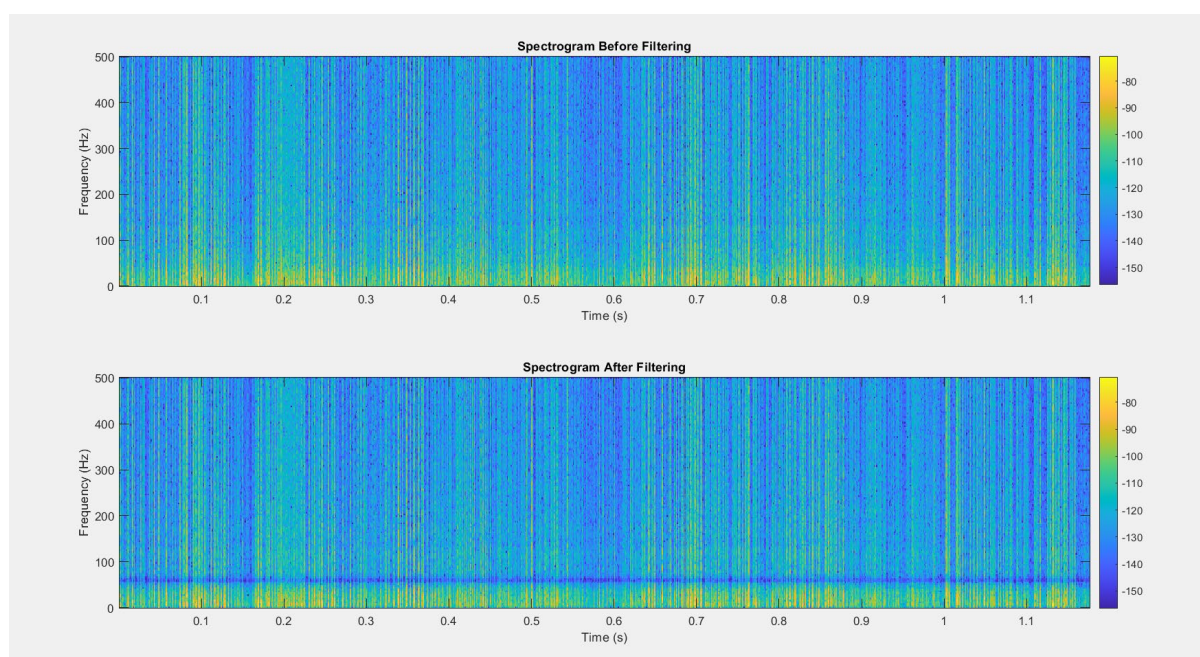
<https://archive.ics.uci.edu/dataset/481/emg+data+for+gestures>

بخش اول (دادگان شخصی)

۱-۱- توضیحات دادگان

** کدها و فایل‌های مربوط به این بخش در پوشه‌ی EMG1 قرار دارد.

داده‌های مورد استفاده در این بخش از ۱۰ نفر با استفاده از ۸ حسگر بر روی ساعد جمع‌آوری شده است. ۸ حرکت برای هر فرد انجام شده که هر حرکت ۶ بار تکرار شده است. حرکات به طول ۶ ثانیه انجام شده‌اند و بین هر دو حرکت ۳ ثانیه زمان استراحت بوده است. تصاویر حرکات در فایل پیوست در پوشه‌ی Pictures قرار دارد. داده‌های جمع‌آوری شده به دو دسته‌ی آموزش و آزمون تقسیم شدند، بدین صورت که تکرار دوم و پنجم برای هر فرد برای داده‌ی آزمون در نظر گرفته شد و بقیه، داده‌ی آموزش بودند. جمع‌آوری داده با ۸ بیت و فرکانس نمونه‌برداری ۲۰۰ هرتز انجام شده است. دادگان فیلتر شده بودند و نویز برق شهر و فرکانس پایین حذف شده بود. در شکل زیر اسپکتروگرام قبل و بعد از فیلتر شدن داده‌ها نشان داده شده است. همانطور که در شکل زیر مشهود است تفاوت چندانی در شکل قبل و بعد از فیلتر شدن وجود ندارد که نشان از فیلتر شدن دادگان از قبل دارد.



۱-۲- پیاده‌سازی

دادگان ابتدا به صورت فایل mat. بودند و هیچ پردازش اولیه‌ای به جز فیلترینگ بر روی آن‌ها انجام نشده بود و به همین دلیل حرکات مختلف، طول‌های مختلفی داشتند، برای مثال ۱ حرکت دارای ۱۳۰۰ نمونه و دیگری ۱۵۰۰ نمونه بود. به همین منظور در اولین اقدام دادگان پیش‌پردازش شدند. در این مرحله تمامی تکرارها برای تمامی حرکات، به طول ۱۲۰۰ نمونه (۶ ثانیه) جدا شدند. دادگان جدا شده در ماتریس‌هایی با اسم هر شخص برای آزمون و آموزش به طور جداگانه ذخیره شدند. کدهای مربوط به این بخش در فایل پیوست در بخش Preprocess قرار دارد.

بعد از جداسازی دادگان، این دادگان برای استخراج ویژگی‌ها استفاده شدند. برای استخراج ویژگی از ویژگی‌های زیر استفاده شد.

Skewness, mean, std, MAV, Kurtosis, zero crossing, variance, Willison amplitude, RMS and Average energy

توابع مربوط به این ویژگی‌ها یا در متلب موجود بودند یا نوشته شده و در فایل پیوست در پوشه‌ی features قرار دارند.

برای استخراج ویژگی‌ها از چندین مدل پنجره با طول‌ها و هم‌پوشانی‌های مختلف و همچنین با ویژگی‌های مختلف استفاده شد. در نهایت دو مدل که بهترین عملکرد را داشتند به ترتیب در پوشه‌های ۵۹.۸۸ و ۶۴.۳۰ قرار داده شدند. نام پوشه‌ها از درصد صحت خروجی مدل‌ها برگرفته شده است. در ادامه به معرفی این مدل‌ها پرداخته می‌شود.

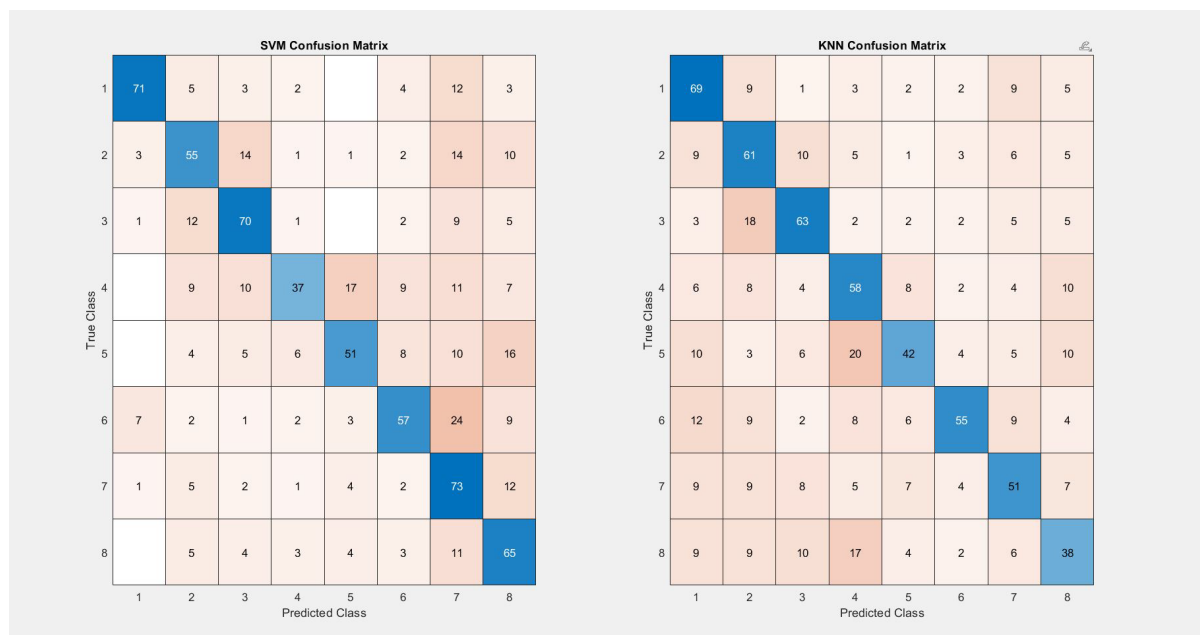
۱-۳- ۵۹.۸۸

در این مدل برای استخراج ویژگی‌ها از پنجره‌ای به طول ۲۰ بدون هم‌پوشانی استفاده شد. در هر پنجره برای هر کانال به طور جداگانه ویژگی‌های mean, RMS و std محاسبه می‌شدند و در آرایه‌هایی با نام‌های x_train و X_test ذخیره می‌شدند. همچنین برچسب‌ها نیز در آرایه‌های Y_train و Y_test برای دادگان ذخیره می‌شدند. در نهایت پس از استخراج ویژگی‌ها طبقه‌بندهای SVM(Kernel=poly) و KNN بر روی دادگان پیاده شدند و صحت خروجی‌آنها به ترتیب برابر با ۵۹.۸۸ و ۵۴.۶۳ بود که مقدار قابل قبولی نبود. در شکل زیر مقدار صحت هر مدل نشان داده شده است.

accuracy	0.5988
accuracy_KNN	0.5463

قبل از پیاده‌سازی طبقه‌بندها، ویژگی‌ها نرمال شدند یعنی میانگین صفر و واریانس ۱.

همچنین confusion matrix مدل‌ها به صورت زیر است. در این ماتریس محور عمودی کلاس‌های واقعی است و محور افقی کلاس‌های پیش‌بینی شده. این ماتریس نشان می‌دهد که برای مثال چند داده که متعلق به کلاس ۱ بوده‌اند به درستی پیش‌بینی شده‌اند. پیش‌بینی‌های صحیح بر روی قطر اصلی ماتریس قرار دارند. با استفاده از این ماتریس معیارهای مختلفی مانند F1score, precision و... محاسبه می‌شود که برای سنجش مدل استفاده می‌شوند.

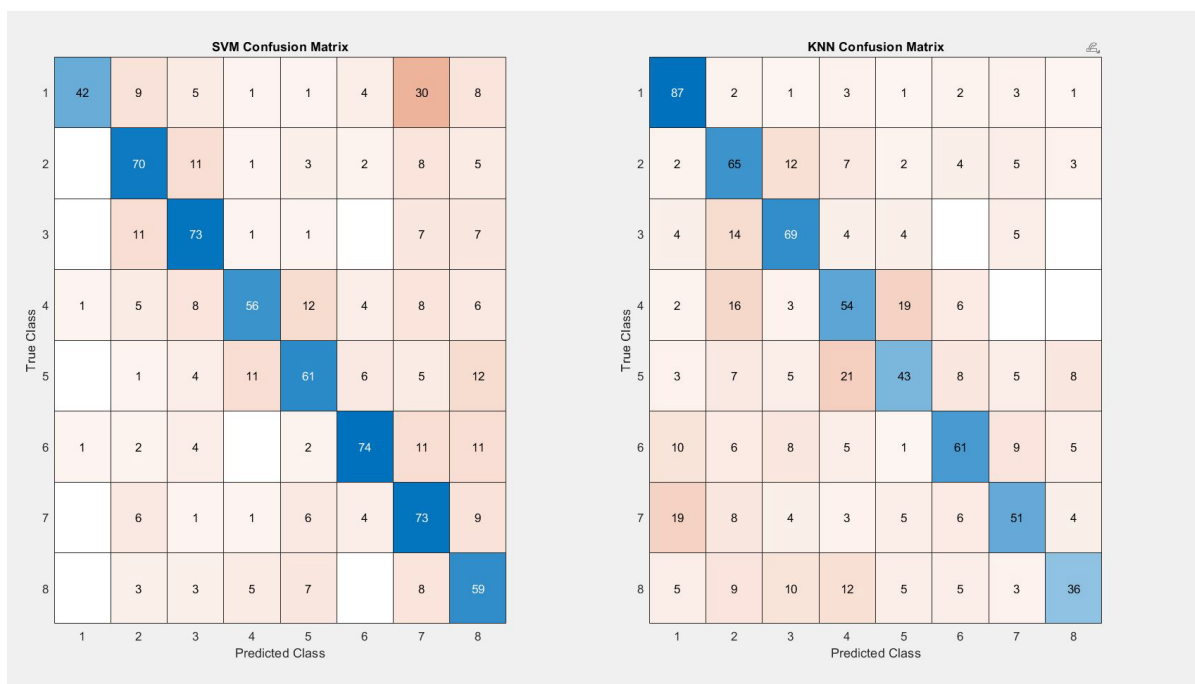


۴-۱- ۶۴.۳۰

در این مدل نیز از پنجره‌ای به طول ۴۰ و هم‌پوشانی ۲۰ استفاده شد. ویژگی‌ها نیز مانند قبل و برای هر کانال محاسبه شدند. در نهایت خروجی صحت برای طبقه‌بندهای SVM و KNN به ترتیب برابر با ۶۴.۳۰ و ۵۹.۸۸ به دست آمد. در شکل زیر این صحت‌ها نشان داده شده‌اند.

accuracy	0.6430
accuracy_KNN	0.5899

همچنین confusion matrix برای این بخش به صورت زیر است.



*** در هر دو پوشه، هر دو طبقه‌بندهای با فرمت mat. ذخیره شده‌اند تا در صورت نیاز به بررسی، از انجام آموزش دوباره خودداری شود.

بخش دوم

*** کدها و فایل‌های مربوط به این بخش در پوشه‌ی EMG2 قرار دارد.

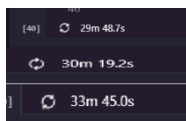
۲-۱- توضیحات دادگان

این دادگان توسط یک تیم از ۳۶ نفر جمع‌آوری شده بود. هر شخص ۷ حرکت را دو بار تکرار کرده بود. دادگان از ۸ کانال بر روی ساعد برای هر فرد ثبت شده بودند. هر حرکت ۳ ثانیه انجام شده بود و در بین هر دو حرکت ۳ ثانیه زمان استراحت وجود داشت. دادگان این بخش نیز فیلتر شده بودند. در مورد فرکانس نمونه‌برداری و... اطلاعاتی در دسترس نبود.

۲-۲- پیاده‌سازی

در این بخش برای پیاده‌سازی مدل‌ها و پردازش دادگان از پایتون استفاده شد. ابتدا تمامی بخش‌هایی که داده‌های از دست رفته یا null داشتند با میانگین ستون مربوطه جایگزین شدند. در این بخش ابتدا سعی شد

که با استفاده از پنجره‌گذاری ویژگی‌های مختلف برای دادگان به دست بیاید اما با توجه به حجم محاسبات بالا این امر با سیستم شخصی محقق نشد. در ادامه از سایت colab برای تسریع عملکرد استفاده شد اما با توجه به مشکلات اینترنت چهار بار در میانه‌های استخراج ویژگی‌ها، کولب متوقف شد و کار نیمه تمام ماند. در تصویر زیر مدت زمان‌هایی برای استخراج ویژگی‌ها صرف شد نشان داده شده است.



(48)	29m 48.7s
30m 19.2s	
33m 45.0s	

در نهایت با توجه به مشکلات گفته شده از پنجره‌گذاری دستی صرف نظر شد و با جست و جو در اینترنت به دنبال بهینه‌ترین روش برای استخراج ویژگی‌ها پرداخته شد. نتیجه‌ی جست و جو استفاده از تابعی از کتابخانه‌ی pandas بود که دستور این تابع در زیر نشان داده شده است.

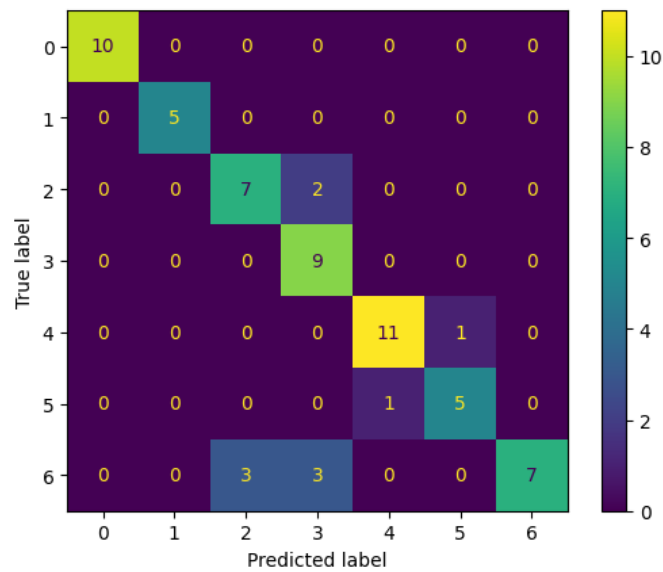
```
1 #feature extraction
2 dataset=dataset.agg(['min','max',rms,SSI,abs_diffs_signal,np.ptp])
```

در این تابع، توابعی که برای استخراج ویژگی نیاز هستند برای دادگان مورد استفاده، فراخوانده می‌شوند و در نهایت خروجی به دست می‌آید.

با استخراج ویژگی‌ها در این بخش ۴ طبقه‌بند مختلف بر روی دادگان پیاده شد.

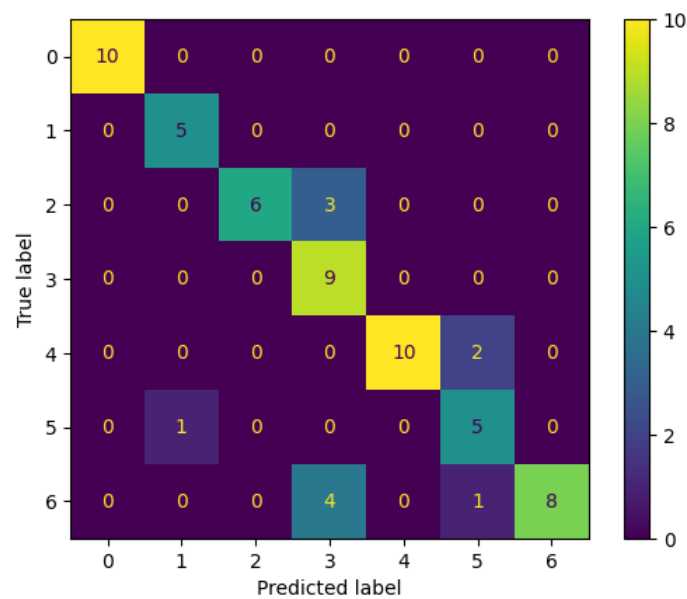
۲-۲-۱ - SVM(kernel=Poly)

این طبقه‌بند با استفاده از کتابخانه‌ی scikit learn پیاده‌سازی شد. برای پیاده‌سازی داده‌ها به صورت ۷۵ درصد آموزش و ۲۵ درصد آزمون پیاده شدند. پیش از پیاده‌سازی مدل نیز، داده‌ها مانند بخش قبل نرمال شدند. صحت خروجی این طبقه‌بند برابر با ۸۴ درصد به دست آمد. در شکل زیر نیز confusion matrix این طبقه‌بند نشان داده شده است. همانطور که قابل ملاحظه است تعداد کمی داده بر روی قطر اصلی نیستند که نشان از پیش‌بینی غلط آن‌ها دارد.



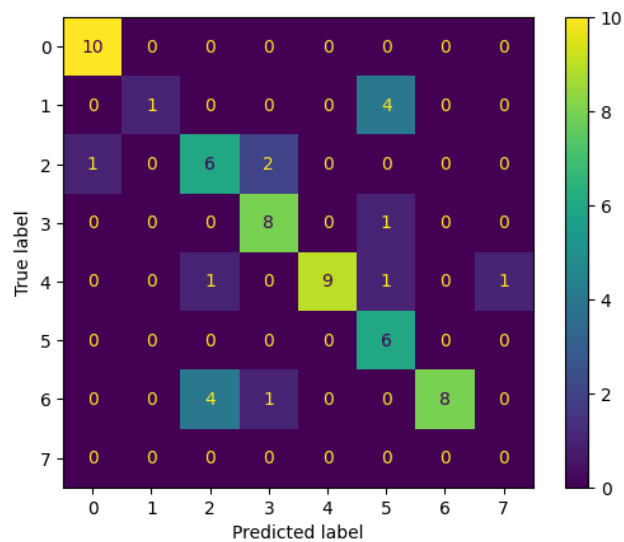
KNN - ۲-۲-۲

این طبقه‌بند نیز مشابه طبقه‌بند قبلی پیاده شد. صحت این طبقه‌بند برابر با ۸۲ درصد است. confusion matrix نیز به صورت زیر است.



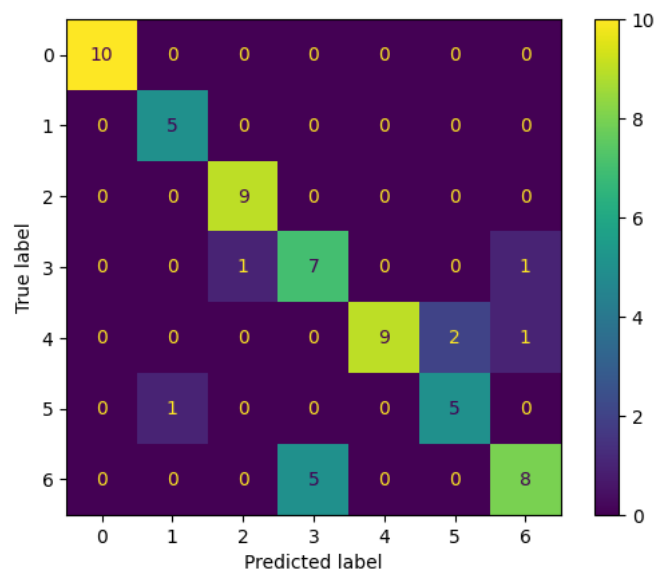
Decision tree - ۳-۲-۲

در این طبقه‌بند نیز صحت به دست‌آمده برابر با ۷۶ درصد است. در شکل زیر نیز confusion matrix مربوط به این طبقه‌بند نشان داده شده است.



۲-۲-۴ Random Forest

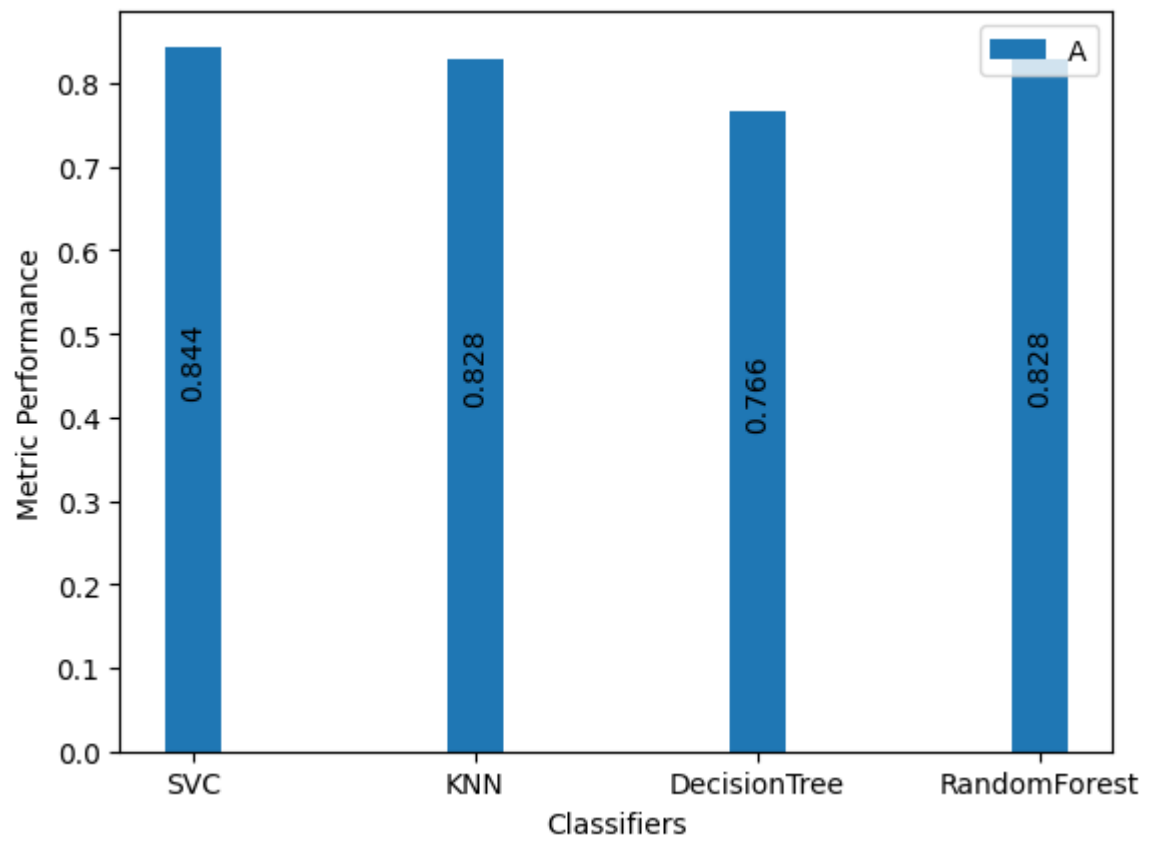
صحت به دست آمده برای این طبقه‌بند برابر با ۸۲ درصد بود. همچنین confusion matrix آن نیز به صورت زیر است.



در آخر نیز مقایسه‌ی صحت این طبقه‌بندها به صورت زیر است.

بهترین طبقه‌بند: SVM

بدترین: Decision tree



منابع

1. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.agg.html>
2. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>
3. <http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
4. <http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
5. <http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
6. <http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>