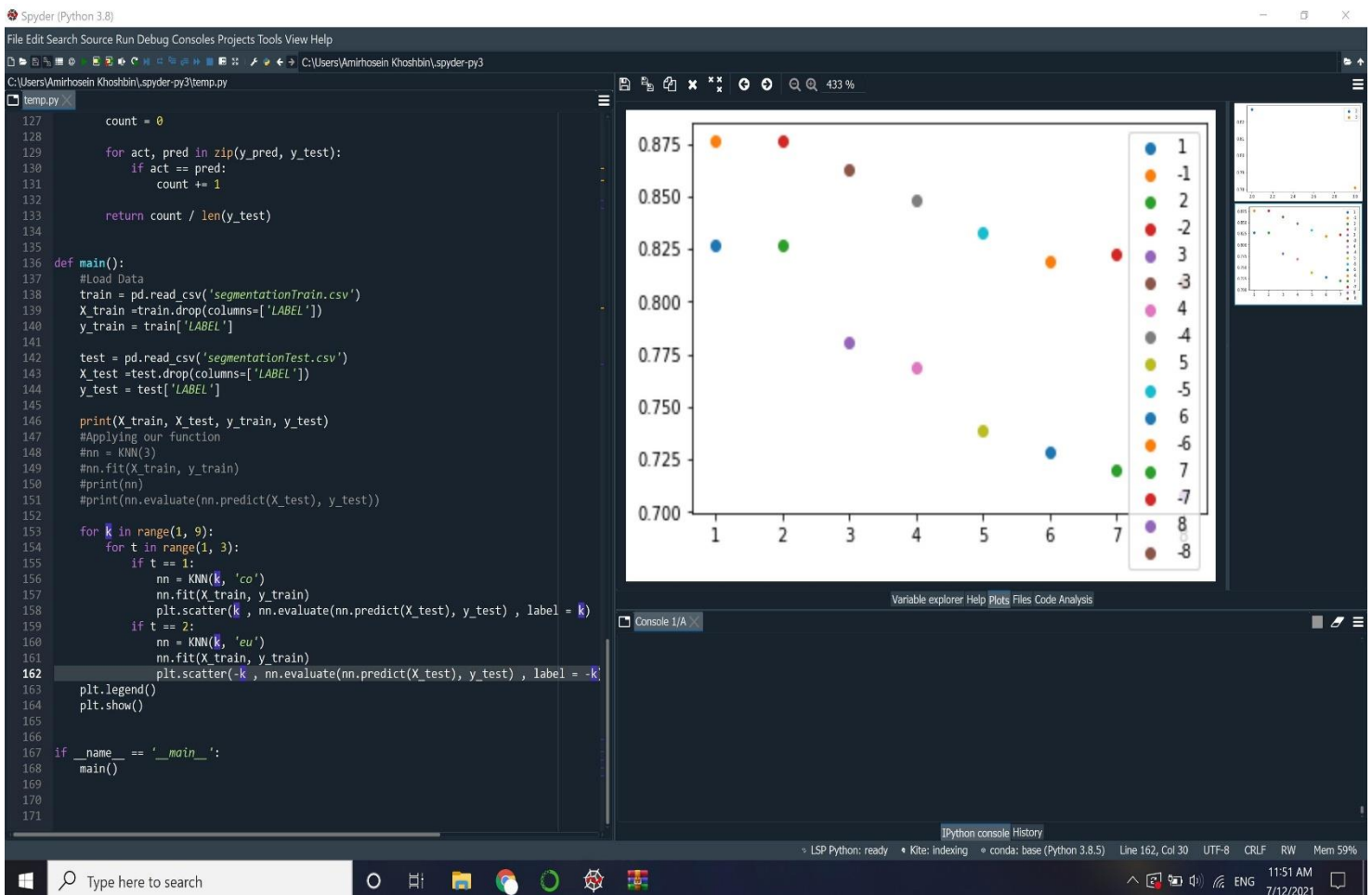


امیرحسین خوش بین ۹۶۳۱۰۲۴

تمام کد ها ضمیمه شدا است.

(۱)

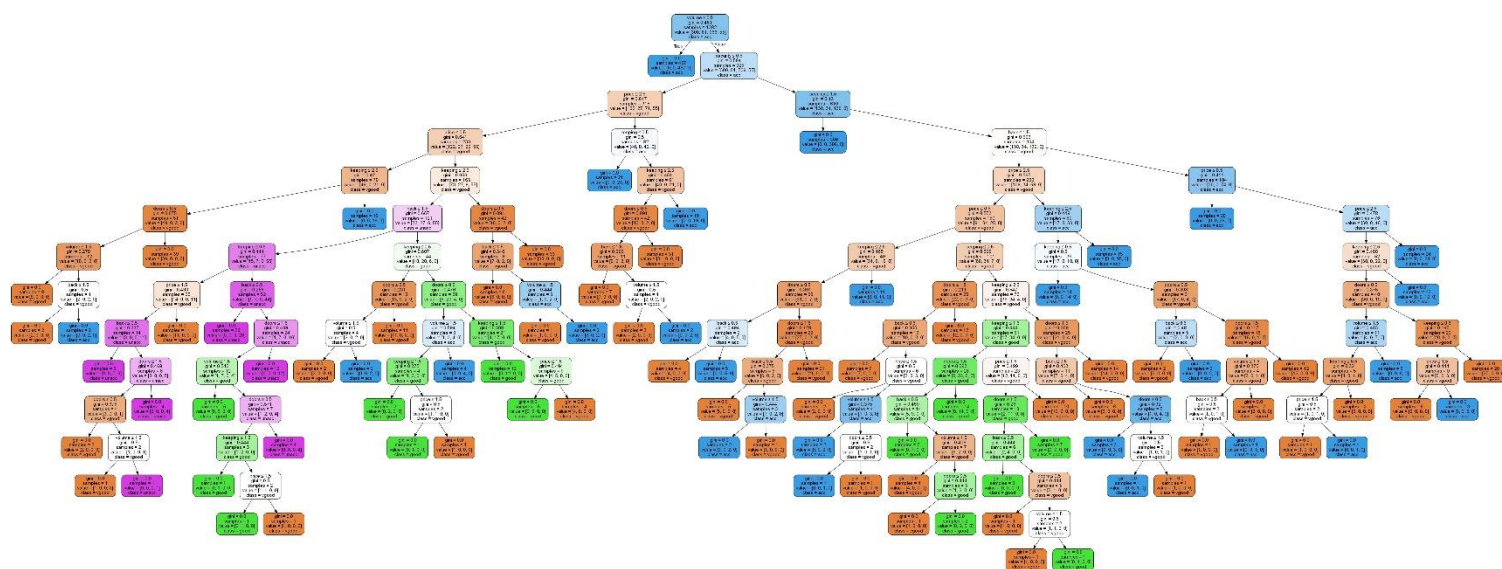
یک کلاس knn وجود دارد که euclidean\_dist و cosin\_dist دو نوع معیار فاصله را حساب میکنند. K\_neighbors به تعداد k نقاط همسایه یک نقطه را مشخص میکند. Get\_nn تمام نقاط را برای تابع k\_neighbors حساب میکند و تابع predict لیبل هر نقطه را با استفاده از vote\_count حساب میکند و در نهایت evaluate تابعی است که لیبل پیشبینی شده را با لیبل تعیین شده مقایسه میکند و با تقسیم بر تعداد لیبل ها دقت را خروجی میدهد.



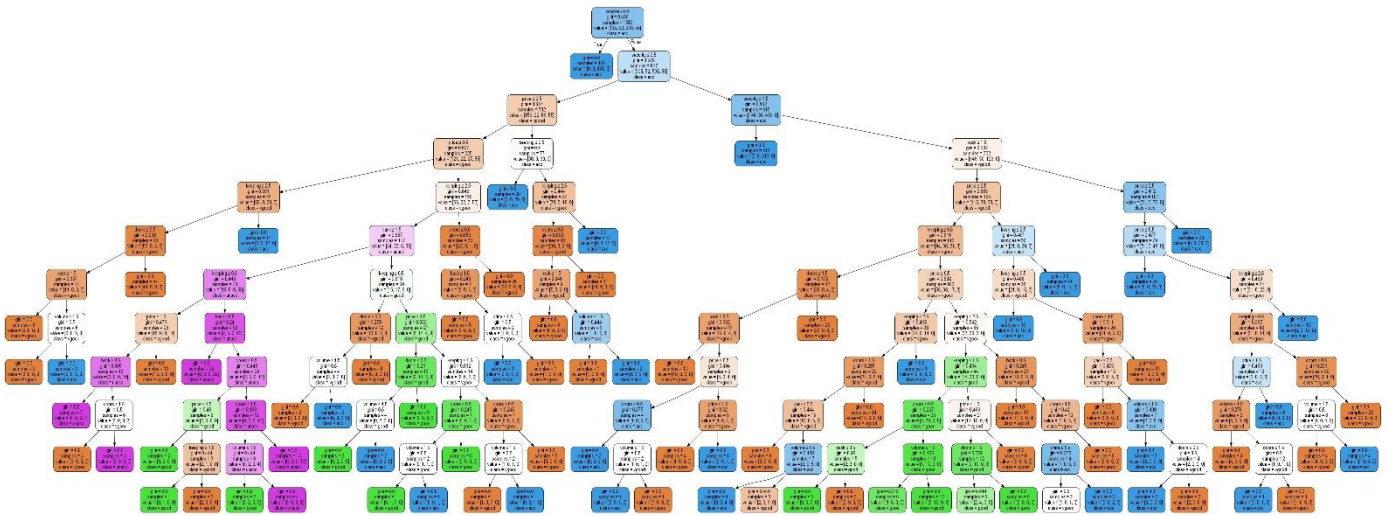
نمودار بالایی که دقت بالاتری دارد بر اساس معیار اقلیدسی است و  
بالاترین دقت برای  $k=1,2$  است.

(۲)

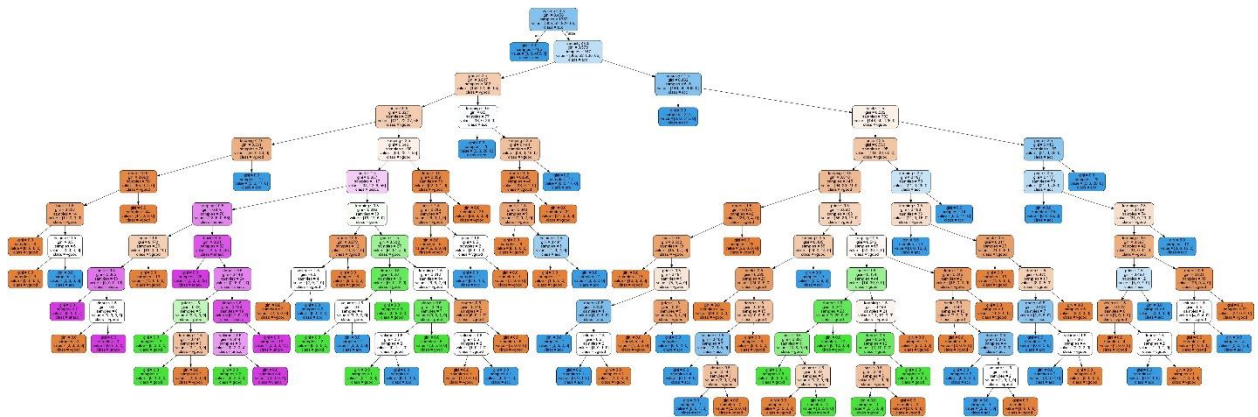
بدون استفاده از max\_depth



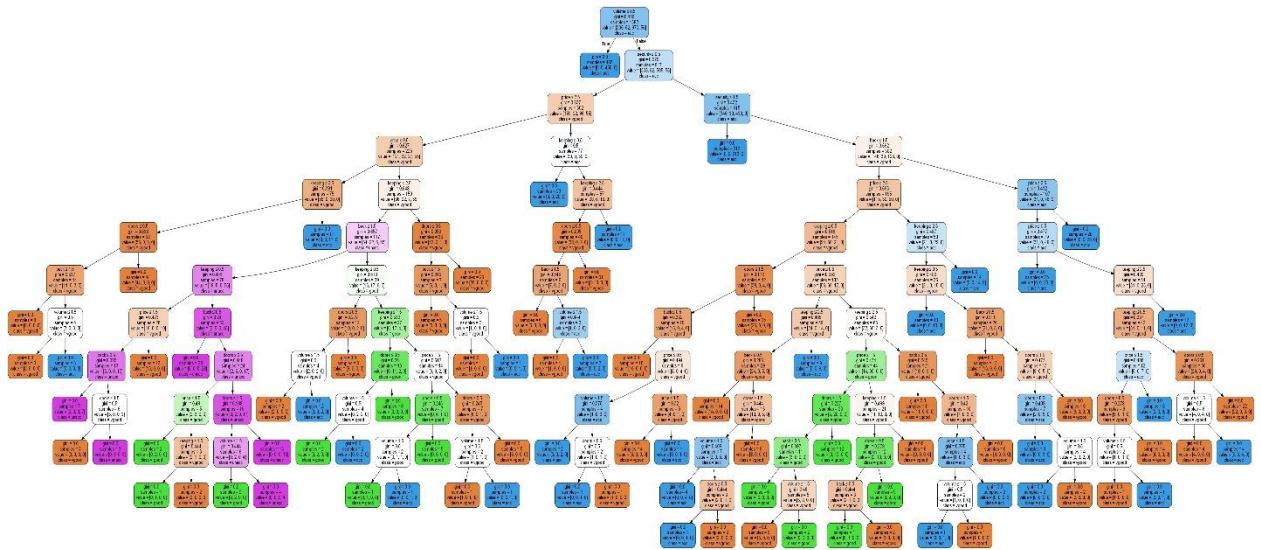
Max\_depth=11



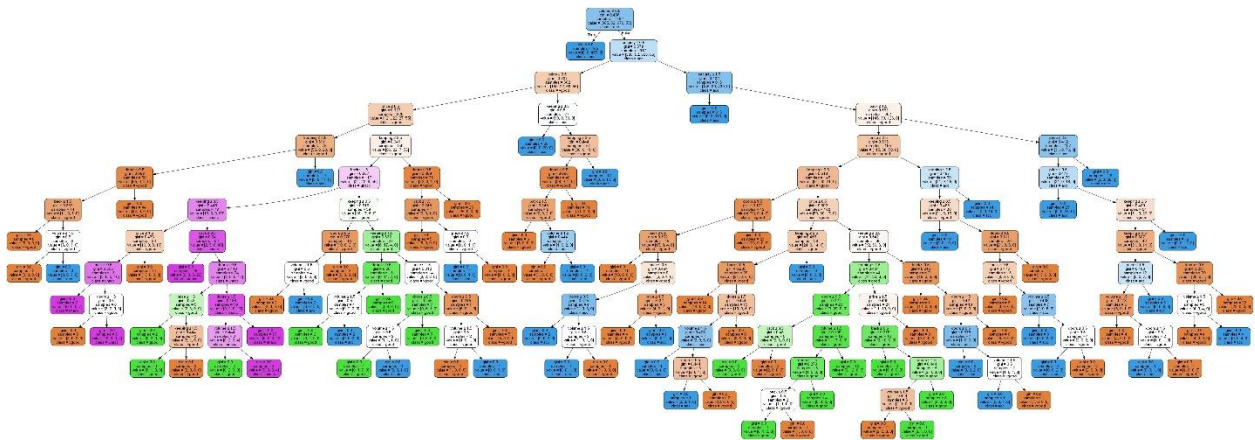
Max\_depth=12



Max depth=13

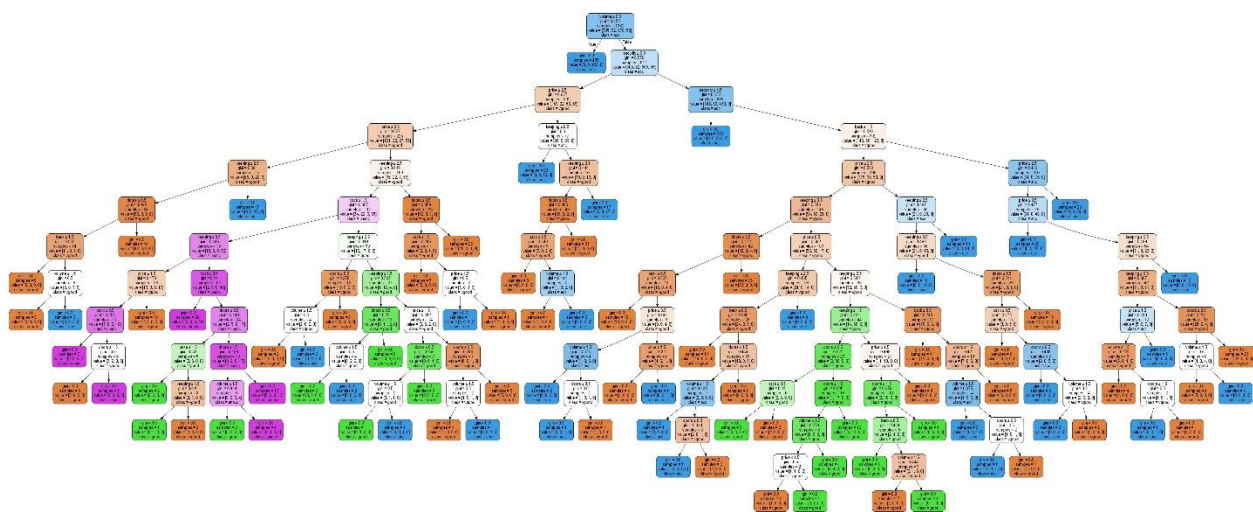


Max depth=14

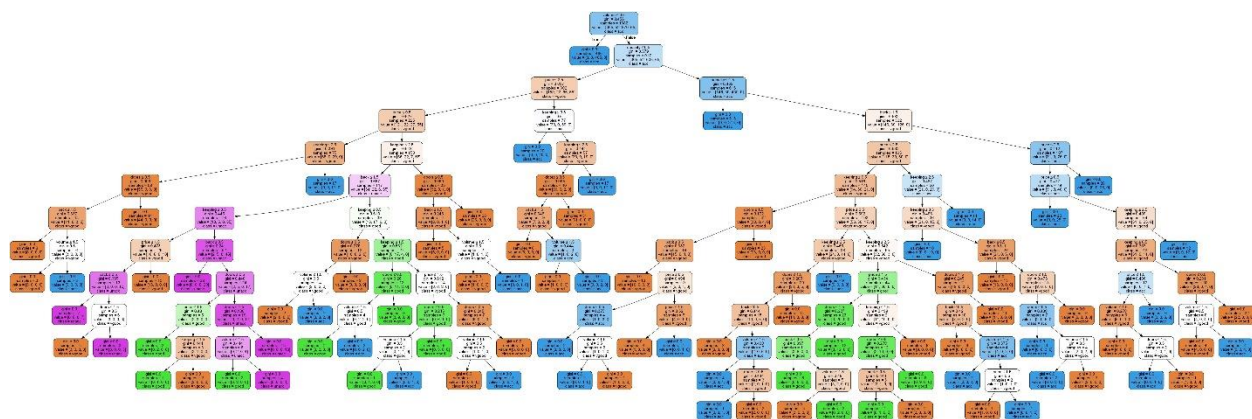


Max depth=15





Max depth=16



بهترین با عمق ۱۴ بود دقت ۰.۹۸۸ و confusion matrix به شکل زیر است.

```
In [8]: runfile('C:/Users/Amirhosein Khoshbin/.spyder-py3/untitled0.py', wdir='C:/Users/Amirhosein Khoshbin/.spyder-py3')
```

Accuracy: 0.9884393063583815 score: 0.9884393063583815

	precision	recall	f1-score	support
acc	0.97	0.99	0.98	92
good	0.93	0.88	0.90	16
unacc	1.00	1.00	1.00	225
vgood	1.00	1.00	1.00	13
accuracy			0.99	346
macro avg	0.98	0.96	0.97	346
weighted avg	0.99	0.99	0.99	346

(۳)

بعد از انجام preprocessing ها به ترتیب اول حذف punctaution و بعد token کردن lowercase میکنیم و stopword ها رو حذف میکنیم و stem میکنیم. بعد ماتریس tf و idf را تشکیل میدهم و بعد با ترکیب آنها، خروجی را به عنوان ورودی به naïve bayes میدهم.

```
190
191 gnb = GaussianNB()
192
193 #Train the model using the training sets
194 gnb.fit(X_train, y_train)
195
196 y_pred = gnb.predict(X_test)
197
198 # Model Accuracy, how often is the classifier correct?
199 print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
200
```

```
accuracy      0.97      346
macro avg     0.94      0.95      0.94      346
weighted avg  0.98      0.97      0.97      346
```

```
In [28]: runfile('C:/Users/Amirhosein Khoshbin/.spyder-py3/untitled1.py', wdir='C:/Users/Amirhosein Khoshbin/.spyder-py3')
1000
<class 'dict'>
Accuracy: 0.5
```

Python console History

ISP Python: ready • Kite: indexing • conda: base (Python 3.8.5) | Line 199, Col 58 | UTF-8 | CRLF | RW | Mem 6

to search

دقت ۵۰ درصد است.

(۴)

## Svm بهتر است

بزرگترین تفاوت بین مدل‌هایی که از نظر "ویژگیها" می‌سازیم این است که Naive Bayes با آنها به عنوان مستقل رفتار می‌کند ، در حالی که SVM تعاملات بین آنها را تا حدی بررسی می‌کند .

```
196 print("Accuracy naive bayes:", metrics.accuracy_score(y_test, y_pred))
197
198 #Create a svm Classifier
199 clf = svm.SVC() # Linear Kernel
200
201 #Train the model using the training sets
202 clf.fit(X_train, y_train)
203
204 #Predict the response for test dataset
205 y_pred = clf.predict(X_test)
206
207 print("Accuracy svm:", metrics.accuracy_score(y_test, y_pred))
```

```
In [44]: runfile('C:/Users/Amirhosein Khoshbin/.spyder-py3/untitled2.py', wdir='C:/Users/Amirhosein Khoshbin/.spyder-py3')
1000
<class 'dict'>
Accuracy naive bayes: 0.49
Accuracy svm: 0.53

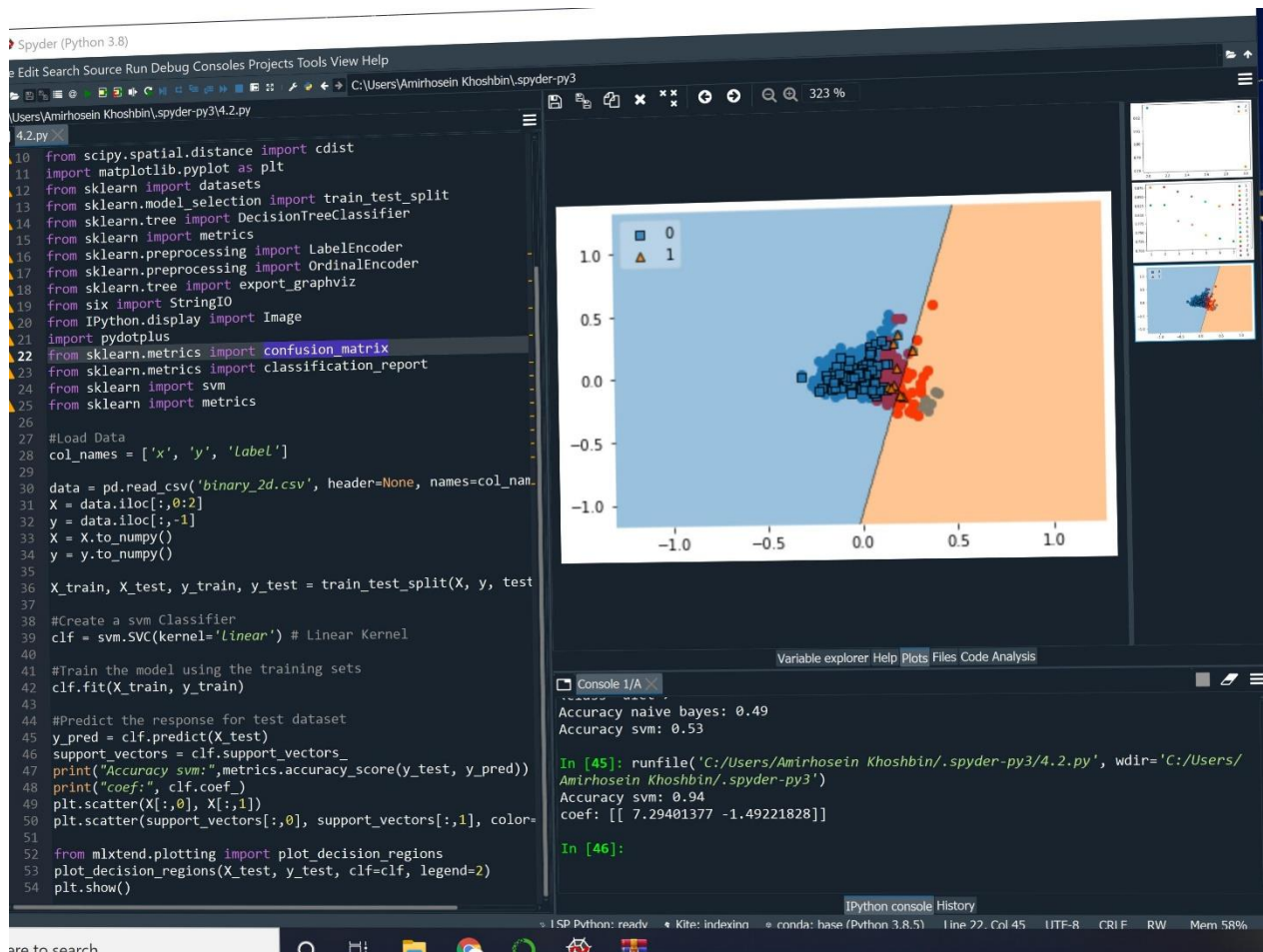
In [45]:
```

Python console History

SP Python: ready Kite: indexing conda: base (Python 3.8.5) Line 190, Col 41 UTF-8 CRI F RW Mem 60%

ere to search

قسمت دوم:



94 درصد