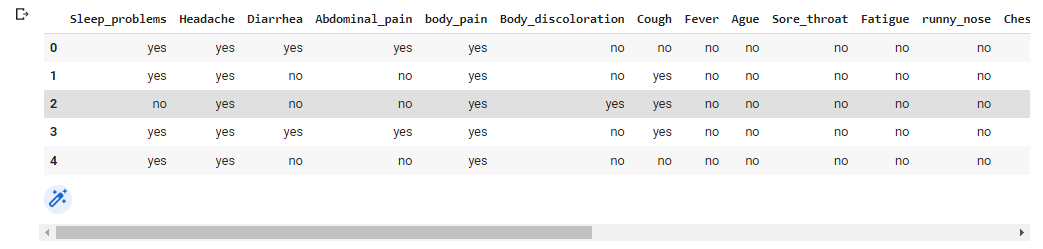
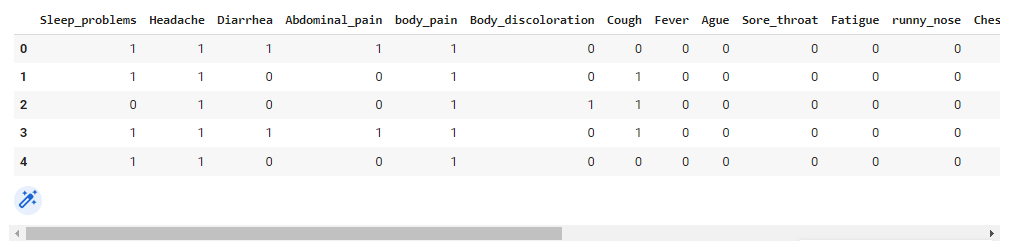
ابتدا کتابخانه های مورد نیاز یعنی numpy وpandas را import میکنیم.

دیتا ست covid.xlsx را از گوگل درایو آپلود میکنیم.

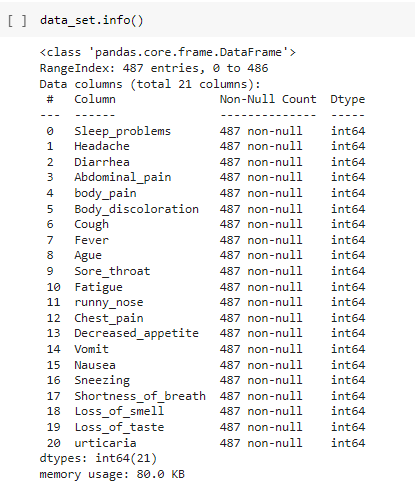


سپس ستون های # و age را به دلیل داشتن سطر های null و عدم نیاز به آنها از دیتاست drop میکنیم.

سپس تابع encode را برای کد گزاری کردن yes های موجود در سطر ها به 1 و no ها به 0 پیاده سازی میکنیم(این کار را در بلوک بعدی با یک حلقه ی for و متود apply انجام میدهیم).



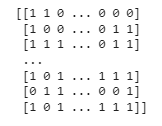
پس از اعمال کار های فوق دیتاست به صورت زیر در می آید.



حال با دستور data\_set.drop\_duplicates(inplace=True) سطر های تکراری را از دیتاست حذف میکنیم.ابعاد دیتاست پس از این بلوک :

[283 rows x 22 columns]

سپس با استفاده از این بلوک ماتریسی به ابعاد 500 در 21 از اعداد رندوم 0و1 میسازیم

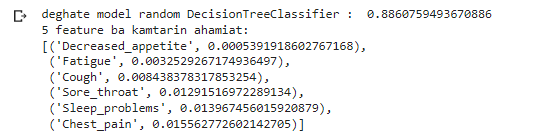


سپس این دیتافریم را با همون ستون ها به اضافه ستون covid که البته مقدار ان این دفعه 0 است در نظر میگیریم تا بتوانیم آن را با دیتا ست خود concat کرده و دوباره با دستور drop\_duplicates تک تک سطر هارا باهم مقایسه کرده (البته بدون ستون covid) چون میخواهیم اگر سطری در این سطور رندوم ما مشابه سطری از دیتاست ما بود ان را حذف کنیم و اگر نبود آن را نگه داشته و ستون covid را برای آن 0 مقدار دهی کنیم.

پس از این ما یک دیتاست کامل با label 0و1 داریم.

حالا بخش train و test را جدا میکنیم. و به ترتیب الگوریتم های خواسته شده را روی دیتاست اعمال کرده و دقت را بدست می آوریم.

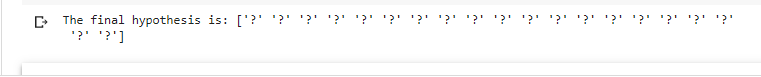
Dessision tree



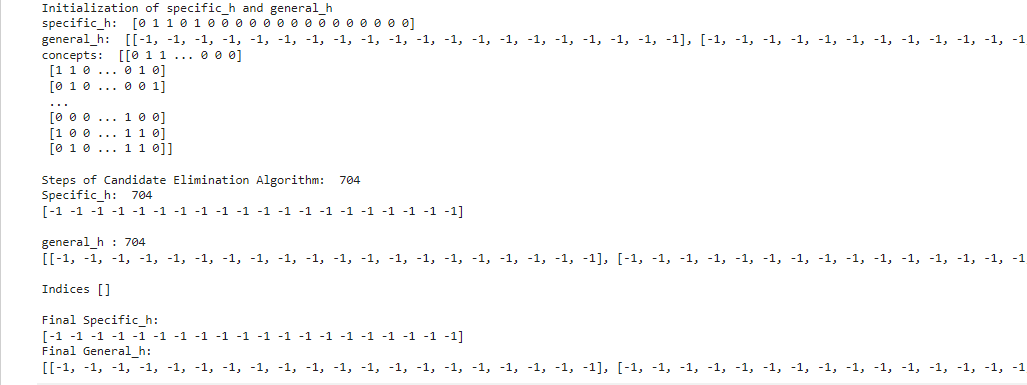
ID3



Finds



Ce



Bayse



KneighborsClassifier



k means

