APPOINTMENT SCHEDULING WITH ML

Pranjal Bahore

15/12/21

According to surveys, on an average, nearly 30% of people miss their medical appointments.

Abstract

No-show appointments significantly impact the functioning of healthcare institutions, and much research has been performed to uncover and analyse the factors that influence no-show behaviour. Patients who schedule clinic appointments and fail to keep them have a negative impact on the workflow of a clinic in many ways. In this report we have try to build a model which can predict early, whether a person is going to attend a scheduled appointment or not. It can help the medical organization to provide more services and deal with the problem of overbooking unnecessary slots. We find that the no-show rate and patients' heterogeneity have a significant impact on the optimal schedule and should be taken under consideration.

1. Introduction

No-show appointments (also commonly referred to as broken or missed appointments) are a burden to essentially all healthcare systems, significantly impacting revenue, cost and use of resources. It is a well-known fact that no-show decreases the provider's productivity and efficiency, increases healthcare costs, and limits the health clinic's effective capacity. Negative effects are also felt by patients who keep their appointments, such as dissatisfaction with high waiting time and perception of overall decrease in service quality. In addition to creating financial costs for providers, non-attendance generates social costs related with unused staff time, ineffective use of equipment and possible misuse of patients' time. There is a markedly growing interest from the healthcare community in uncovering and understanding the issues involved in no-show behaviour. However, given the variability in context and specificities of health care delivery and systems, it is unlikely that a general agreement may be reached regarding the variables that statistically influence no-show behaviour.

We study an overbooking model for scheduling arrivals at a medical facility, with patients having different no-show probabilities and different weights. The different weights correspond to different customer classes, and the no-show probability assigned to each patient reflects her history in attending scheduled appointments. An optimal appointment schedule balances the trade-offs between the benefits of efficient resource utilization and the costs of patients' waiting time and physician's overtime. When patients have different characteristics, the sequencing of the arrivals is also of interest, that is, the order at which patients are scheduled to arrive at the medical facility.

2. Problem Statement

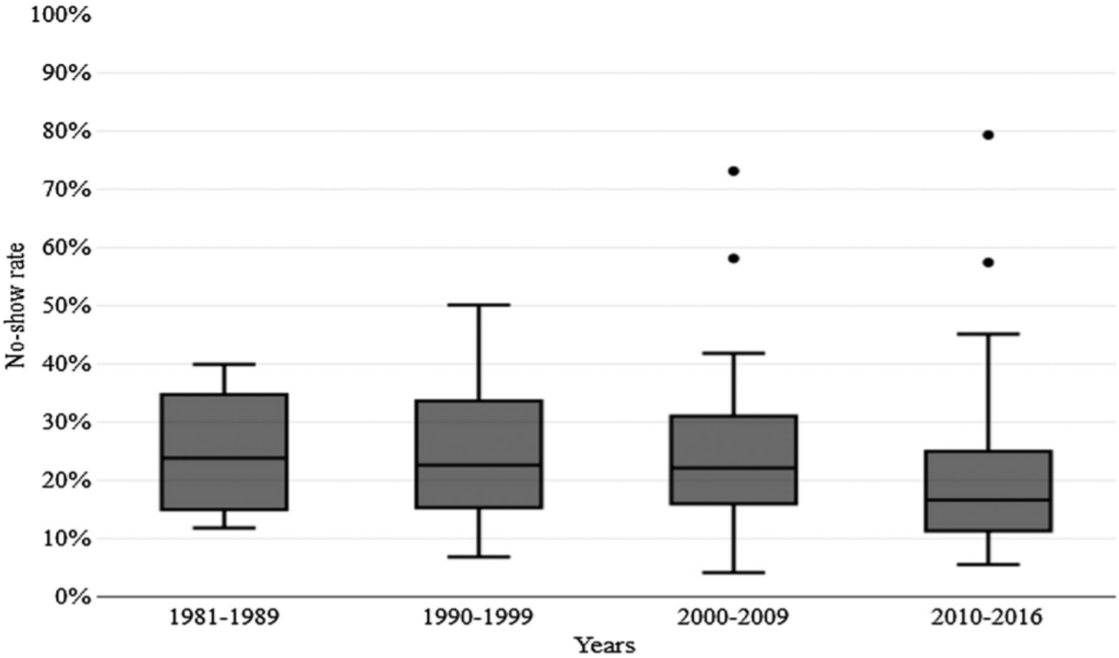
In our respective study we will take a deep dive into factors which are responsible for the no-show to the scheduled appointments and subsequently take help of these factors as the features of our Machine learning model.

3. Market/Customer Need Assessment

There is been a huge surge of AI in the sector of healthcare & many applications such as disease prediction are already present, here we will try to predict the onset of medical appointments. For the required model, we have taken a dataset from Kaggle https://www.kaggle.com/joniarroba/noshowappointments

	Features
0	Gender
1	Age
2	Scholarship
3	Hipertension
4	Diabetes
5	Alcoholism
6	Handcap
7	SMS_received
8	No-show

The above figure shows the respective used as factors while predicting the result of the machine learning model.



The above boxplots is showing that how there has been a declined in no-show rate of appointments.

- **Data Collection**

  To collect the required data for the above purpose we can have some solutions as listed below:

  1] OFFLINE APPOINTMENTS
  We can provide a form to the patient for filling it up and apart from general details we can make add on positions for our above features and collect the data of our patients

  2] ONLINE APPOINTMENTS
  We can provide a google form to fill the suffice details mentioned above or if telephonically we can ask them to put it out.

4. Target Specifications

The above supposed system will carry benefits manifolds:

- It will help the healthcare organizations to deal with the issue of overbooking appointments.
- It will also help the patients to pre-book the required slots on particular day.
- Predicting no-show appointments will lead to advance vacancy of slots for that day or prepare for the next further day itself.

5. Business opportunity

If we already have the data of the patients and their probability of no-show to scheduled date, we can leverage that & make revenue out of it as:

- We can create a **Waiting portal** which can be activated and provide advance notification to the patients in the waiting list along with their chance of converting the appointments and will charge them for providing early access.

- We can also charge hospitals and doctors for not making their resources and time go of no use, rather we will provide them with extra customers for the same slots which are going to be no-show slots of that particular date.

6. Benchmarking

There are many emerging companies in the healthcare sector using AI/ML in their respective domain to help businesses grow to larger area and extend their reach & productivity:

- **LiveHealth**

The vision of the company is to diagnose data at the center of the healthcare ecosystem, where healthcare providers and customers are connected in a single environment.

- **ChironX**

ChironX has achieved its name in one of the top 10 healthcare AI companies in India, by developing best software for healthcare industry. ChironX builds intelligent software to detect diseases that impact large populations from medical images.

- **Artelus**

Artelus based in Karnataka ties up with technology firms Health care organizations to access medical data in order to provide timely, doctor assisted solutions utilizing technologies such as portable devices, cloud computing, and deep learning.

- **Qure.ai**

  Qure.ai's core team combines deep learning expertise with clinical, scientific and regulatory knowledge. The company's advisory panel consists of radiologists, other doctors and public health experts. Qure.ai works with specialists to define clinically relevant problems and design real-world solutions.

- **Lybrate**

  Based in New Delhi, Lybrate is also one of the top 10 healthcare AI companies in India. Lybrate provides medical healthcare communication and it also known as the best delivery platform. Lybrate has facilities for both, doctors as well as patients. It believes in providing world class facilities.

6.1 Applicable Patents

- **Patent1**

  The average no-show rate across all studies was found to be 23.0%, and further analysis revealed that this rate was highest in the African continent (43.0%) and lowest in Oceania (13.2%). We also verified that psychiatry and primary care were the most investigated specialties, and that various statistical methods were used in the reviewed papers to identify significant predictors of no-show, among which the most common were chi-squared tests, t-tests and multiple logistic regression models.

- **Patent 2**

  This study considers static as well as sequential appointment scheduling and provides guidelines for the use of overbooking to compensate for no-shows. We demonstrate that the no-show rate and patients' heterogeneity have a significant impact on the optimal schedule and should be taken under consideration. Structural properties as well as a new sequencing rule are presented for the optimal offline schedules.

- **Patent 3**

  Three interconnected themes emerged from the interviews as barriers to appointment attendance: emotional barriers, perceived disrespect of the patient's beliefs and time by the health care system, and distrust and lack of understanding of the scheduling system. Transportation and child care were logistical barriers, but respondents noted they could be overcome.

6.2 Applicable constraint

- Data Collection and privacy.
- Lack of Understanding of the Scheduling System.
- Emotional Barriers to Keeping Appointments.
- Waitlisting Platforms.
- Communication to the patients.

7. Concept Generation

This product will require the information of patients to be used as a data for the machine learning model we are going to be build, here as above mentioned I have used a dataset from the Kaggle here.

In [102]:
```
data.head(10)
```

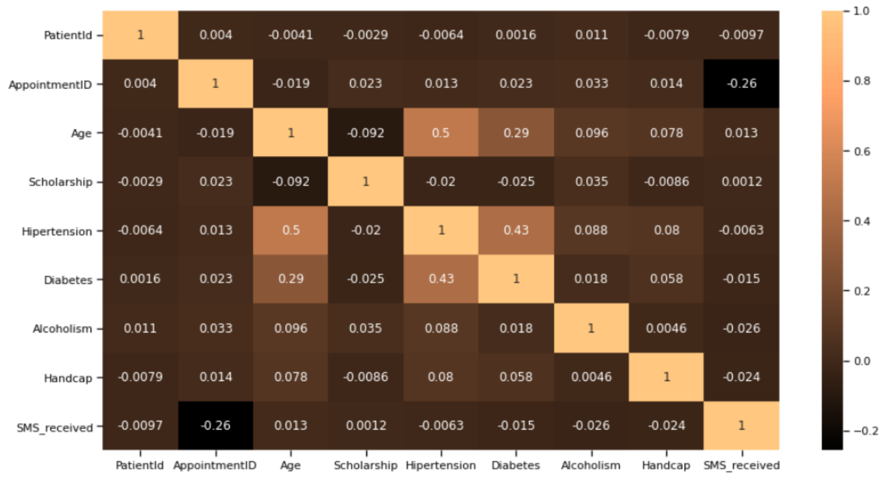
Out[102]:

AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap	SMS_received	No-show
5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62	JARDIM DA PENHA	0	1	0	0	0	0	No
5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA	0	0	0	0	0	0	No
5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z	62	MATA DA PRAIA	0	0	0	0	0	0	No
5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z	8	PONTAL DE CAMBURI	0	0	0	0	0	0	No
5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA	0	1	1	0	0	0	No
5626772	F	2016-04-27T08:36:51Z	2016-04-29T00:00:00Z	76	REPÚBLICA	0	1	0	0	0	0	No
5630279	F	2016-04-27T15:05:12Z	2016-04-29T00:00:00Z	23	GOIABEIRAS	0	0	0	0	0	0	Yes
5630575	F	2016-04-27T15:39:58Z	2016-04-29T00:00:00Z	39	GOIABEIRAS	0	0	0	0	0	0	Yes
5638447	F	2016-04-29T08:02:16Z	2016-04-29T00:00:00Z	21	ANDORINHAS	0	0	0	0	0	0	No
5629123	F	2016-04-27T12:48:25Z	2016-04-29T00:00:00Z	19	CONQUISTA	0	0	0	0	0	0	No

➢ Figure above tells us the descriptive stats of the dataset.

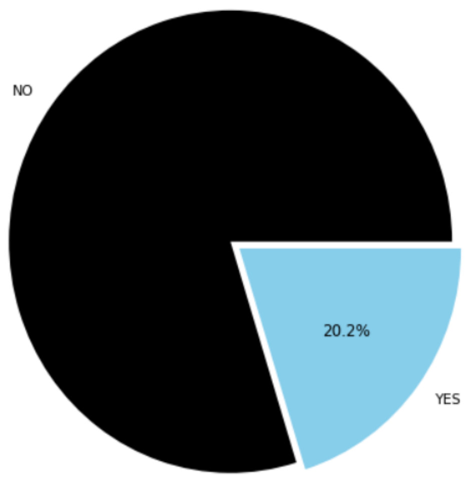
We have tried to figure out the features which are in correlation with the output and used pearson correlation matrix for that:

In [106]:
```
plt.figure(figsize=(15,8))
sns.heatmap(data.corr(),annot=True,cmap='copper')
sns.set_context('notebook',1)
```

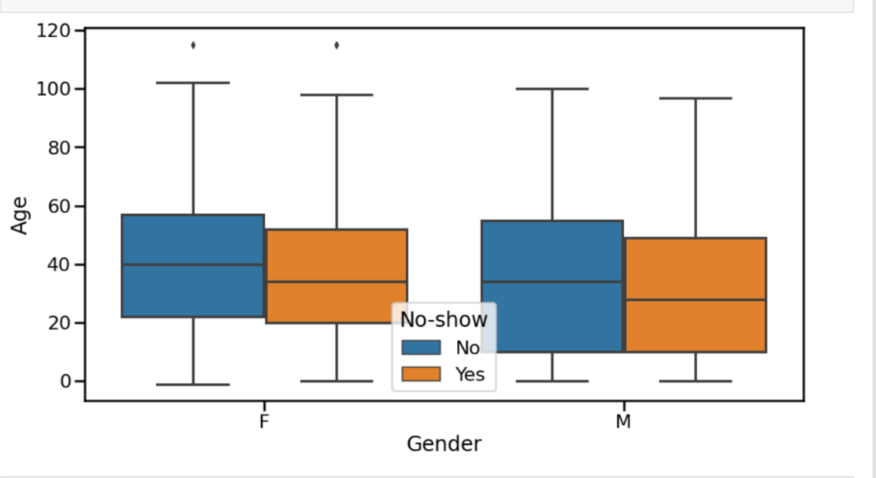


➢ Heatmap showing correlation between features.

In [84]:
```
plt.pie(data['No-show'].value_counts(),labels=['NO','YES'],radius=2,autopct='%0
plt.show()
```

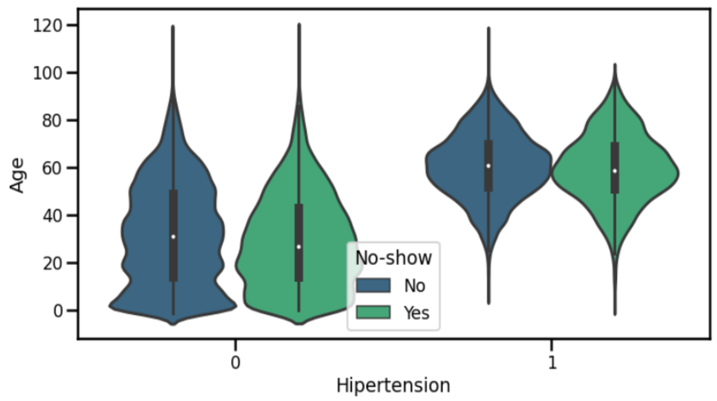


➢ Data shows around 20.2% people had missed their scheduled appointments.



➢ Female Patients tends to skip their appointments more as compared to male ones.

In [98]:
```
plt.figure(figsize=(12,6))
sns.violinplot(x='Hipertension',y='Age',data=data,hue='No-show',palette='viridis')
sns.set_context('poster',font_scale=0.8)
```



➢ Age has a direct relation with hypertension and came out to be a leading factor to miss appointments.

Machine Learning can be vital for the analysis of the whole dataset as:

- We can figure out the no. of patients having higher probability to skip appointments.
- We can deal with the issue of Overscheduling.
- We can save revenue for the hospitals and organisation.
- Managements can be efficient & robust.

8. Final Product Prototype

The final product when implemented into the business will provide us detail information regarding the following:

- **Predicting No-show Rate**
  By a creating the machine learning model we can predict the customers who are going to skip the scheduled appointments.

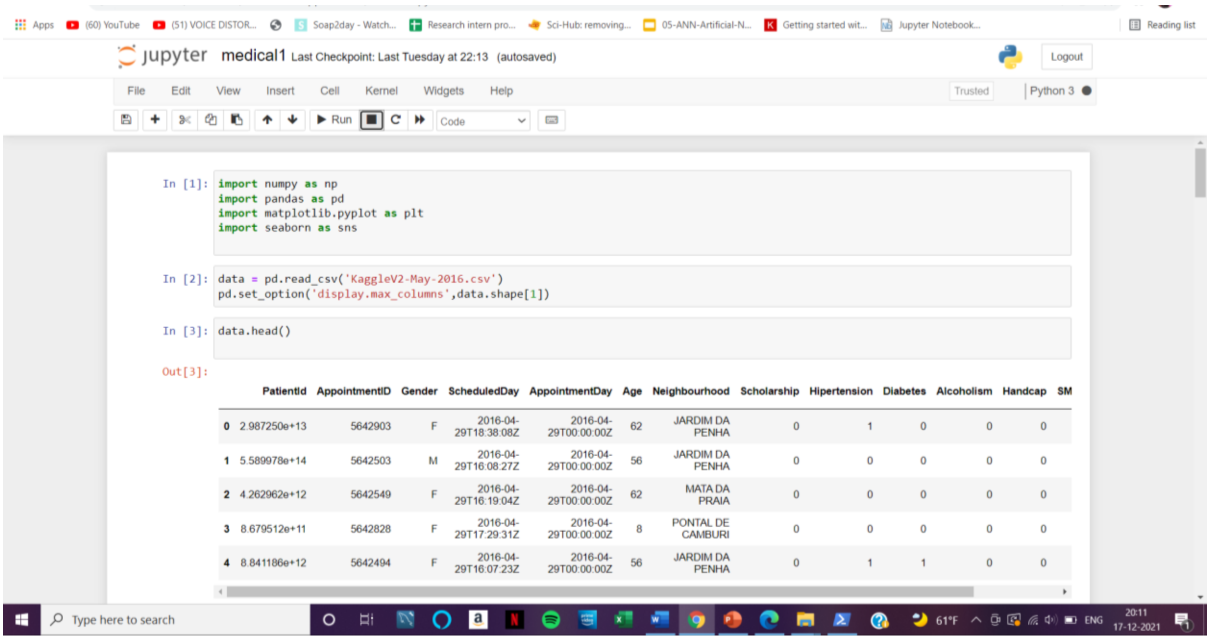
- **Dealing with Overscheduling**

  After having the details of skipped appointments we can help organisations to fill those seats with new ones which will helpful for both patients as well as hospitals.

- **Revenue Generation**
  We can offer organisation to save their time and resources which are going to be invested in false appointments.

- **Waitlisting Platform**
  We can create a platform in which waitlist patients can get early access to the appointments and allotted seats for the skipped appointments seats and thus we can charge them for that.

9. Implementation on small scale



jupyter medical1 Last Checkpoint: Last Tuesday at 22:13 (autosaved)

In [5]:
```
data['No-show'].value_counts(True)
```
Out[5]:
```
No     0.798067
Yes    0.201933
Name: No-show, dtype: float64
```

In [6]:
```
data.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110527 entries, 0 to 110526
Data columns (total 14 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   PatientId       110527 non-null  float64
 1   AppointmentID   110527 non-null  int64
 2   Gender          110527 non-null  object
 3   ScheduledDay    110527 non-null  object
 4   AppointmentDay  110527 non-null  object
 5   Age             110527 non-null  int64
 6   Neighbourhood   110527 non-null  object
 7   Scholarship     110527 non-null  int64
 8   Hipertension    110527 non-null  int64
 9   Diabetes        110527 non-null  int64
 10  Alcoholism      110527 non-null  int64
 11  Handcap         110527 non-null  int64
 12  SMS_received    110527 non-null  int64
 13  No-show         110527 non-null  object
dtypes: float64(1), int64(8), object(5)
memory usage: 11.8+ MB
```

In [7]:
```
data.nunique()
```
Out[7]:
```
PatientId          62299
AppointmentID     110527
Gender                 2
ScheduledDay      103549
AppointmentDay        27
Age                  104
Neighbourhood         81
Scholarship            2
Hipertension           2
Diabetes               2
Alcoholism             2
Handcap                5
```

Type here to search

61°F ENG 20:16 17-12-2021

jupyter medical1 Last Checkpoint: Last Tuesday at 22:13 (autosaved)

In [10]:
```
data.drop(['PatientId','AppointmentID','ScheduledDay','AppointmentDay','Neighbourhood'],axis=1,inplace=True)
```

In [11]:
```
data.head(10)
```
Out[11]:

	Gender	Age	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap	SMS_received	No-show
0	F	62	0	1	0	0	0	0	No
1	M	56	0	0	0	0	0	0	No
2	F	62	0	0	0	0	0	0	No
3	F	8	0	0	0	0	0	0	No
4	F	56	0	1	1	0	0	0	No
5	F	76	0	1	0	0	0	0	No
6	F	23	0	0	0	0	0	0	Yes
7	F	39	0	0	0	0	0	0	Yes
8	F	21	0	0	0	0	0	0	No
9	F	19	0	0	0	0	0	0	No

In [12]:
```
data.isnull().sum()
```
Out[12]:
```
Gender          0
Age             0
Scholarship     0
Hipertension    0
Diabetes        0
Alcoholism      0
Handcap         0
SMS_received    0
No-show         0
dtype: int64
```

In [13]:
```
data['Scholarship'].value_counts(True)
```
Out[13]:
```
0    0.901734
1    0.098266
Name: Scholarship, dtype: float64
```

In [14]:
```
data['No-show']=data['No-show'].replace({'Yes':0,'No':1})
```

Type here to search

61°F ENG 20:16 17-12-2021

jupyter medical1 Last Checkpoint: Last Tuesday at 22:13 (autosaved)

In [16]:
```
data['No-show'].value_counts(True)
```
Out[16]:
```
1    0.798067
0    0.201933
Name: No-show, dtype: float64
```

In [18]:
```
gender=pd.get_dummies(data['Gender'],drop_first=True)
data = pd.concat([data.drop('Gender',axis=1),gender],axis=1)
```

In [20]:
```
data.head()
```
Out[20]:

	Age	Scholarship	Hipertension	Diabetes	Alcoholism	Handcap	SMS_received	No-show	M
0	62	0	1	0	0	0	0	1	0
1	56	0	0	0	0	0	0	1	1
2	62	0	0	0	0	0	0	1	0
3	8	0	0	0	0	0	0	1	0
4	56	0	1	1	0	0	0	1	0

In [25]:
```
y = data['No-show']
```

In [26]:
```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
x=scaler.fit_transform(data.drop(['No-show'],axis=1))
```

In [27]:
```
from sklearn.model_selection import train_test_split
```

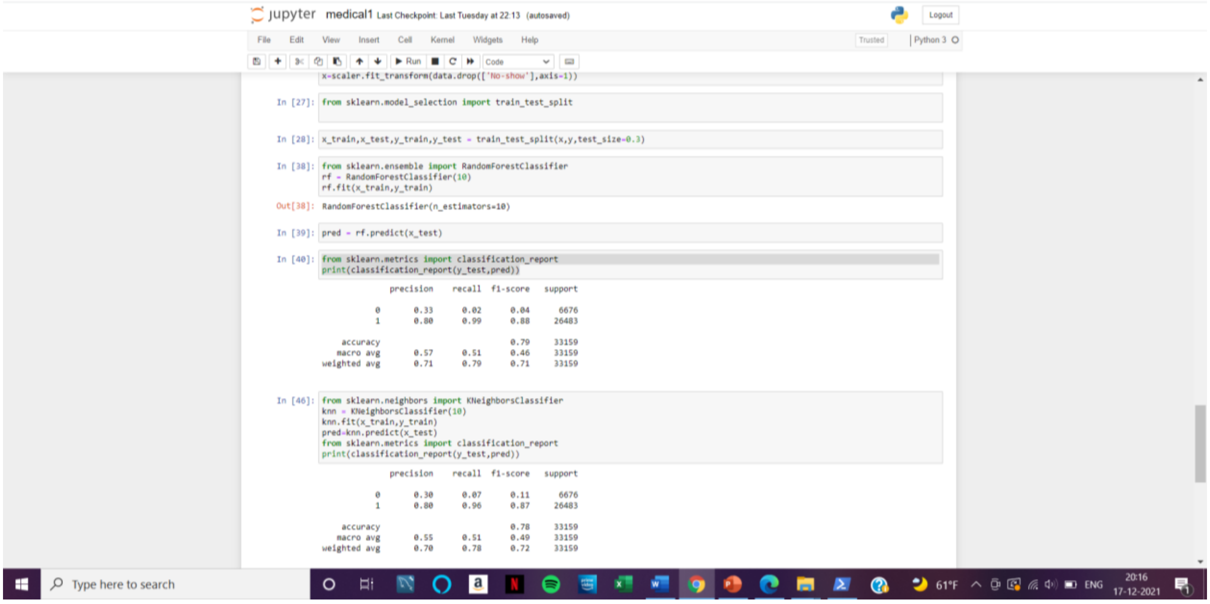
In [28]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3)
```

In [38]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(10)
rf.fit(x_train,y_train)
```
Out[38]:
```
RandomForestClassifier(n_estimators=10)
```

In [39]:
```
pred = rf.predict(x_test)
```

Type here to search

61°F ENG 20:16 17-12-2021



10. Conclusion

This study considers static as well as sequential appointment scheduling and provides guidelines for the use of overbooking to compensate for no-shows. We demonstrate that the no-show rate and patients' heterogeneity have a significant impact on the optimal schedule and should be taken under consideration. Structural properties as well as a new sequencing rule are presented for the optimal offline schedules. A heuristic solution is presented for the online scheduling problem, where requests for appointments come in gradually over time and the scheduler has to fit each patient into a growing schedule for a given day.

11. References

- file:///C:/Users/HP/Downloads/b6846d449ce40dd78ea1e4b4488b92b4.pdf
- file:///C:/Users/HP/Downloads/zacharias2013.pdf
- file:///C:/Users/HP/Downloads/lacy2004.pdf