TRIPADVISOR PREDICTION

Predicció del ”rating” a partir de comentaris

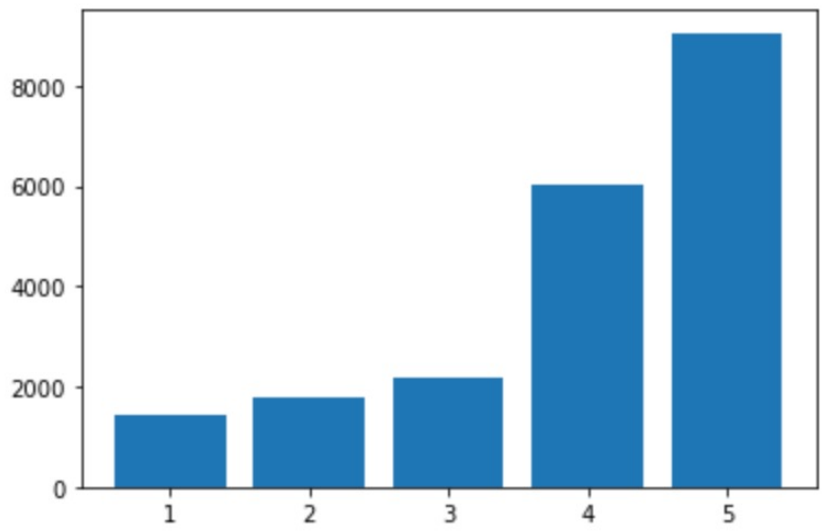
https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviews

Gerard Burgués Llavall

Introducció

	Review	Rating
0	nice hotel expensive parking got good deal sta...	4
1	ok nothing special charge diamond member hilto...	2
2	nice rooms not 4* experience hotel monaco seat...	3
3	unique, great stay, wonderful time hotel monac...	5
4	great stay great stay, went seahawk game aweso...	5
...	...	...
20486	best kept secret 3rd time staying charm, not 5...	5
20487	great location price view hotel great quick pl...	4
20488	ok just looks nice modern outside, desk staff ...	2
20489	hotel theft ruined vacation hotel opened sept ...	1
20490	people talking, ca n't believe excellent ratin...	2

[20491 rows x 2 columns]


8000
6000
4000
2000
0
1
2
3
4
5

room
hotel
day
time
nice
resort
night
got
restaurant
lot
people
beach
place
good
great location
looking
going
week
small
better
want
wonderful
way
need
bar
bed
hour
excellent
staff
friendly
great
said
overall
area
wanted
guest
loved
went
little
really
breakfast
lobby
problem
left
make
best
check
food
stay
drink
trip
city
thing
fine
morning
look
arrived
water
pool
price
street
think
told
great
couple
experience
bit
used
say
ok
come
building
hotel
punta cana
service
door
n't

Que volem resoldre?

Preprocessing

- Lower Case
- Stop Words
- Lemmanize
- Stemmer

Resultats I Model

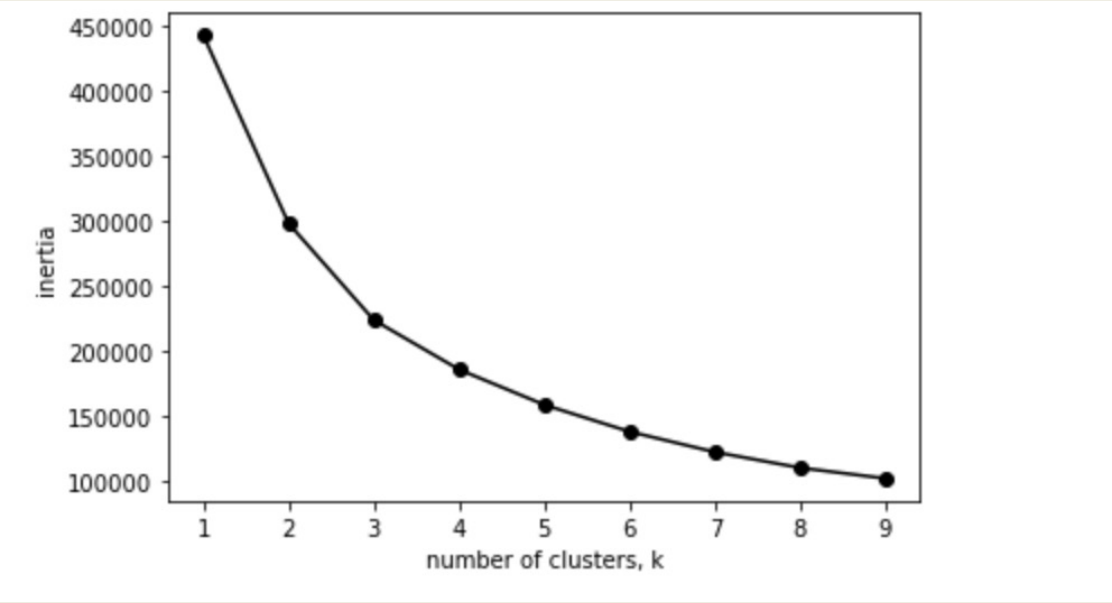
Gaussian Naïve Bayes vs Bernoulli Naive Bayes

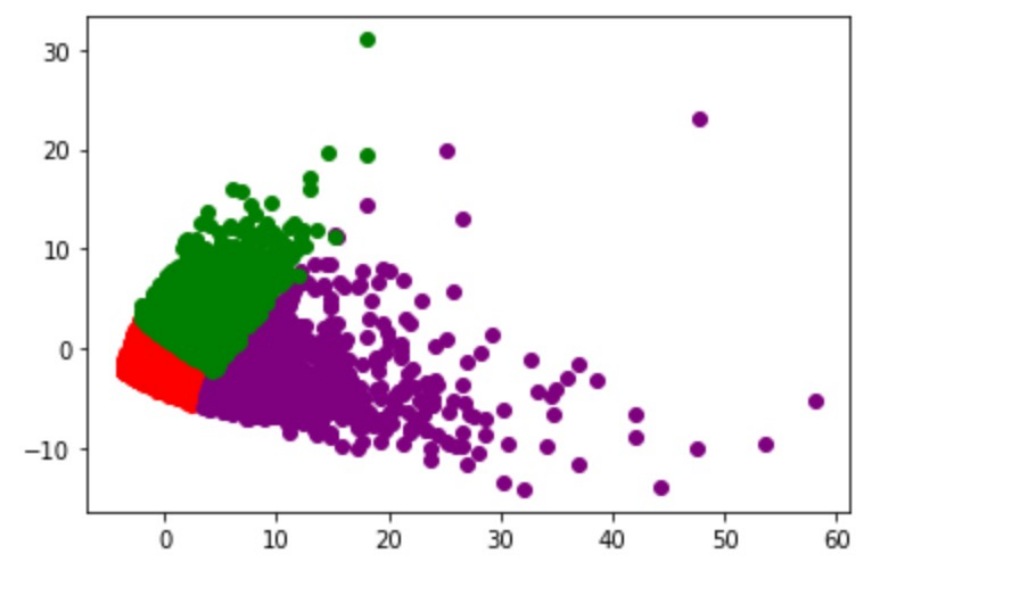
```
Accuracy: 0.4991461332032203
Precision: 0.4991461332032203
Confusion Matrix:
 [[ 199   53   13    4   23]
 [ 141   83   47   22   40]
 [  78   61  100   67  126]
 [  54   64  148  246  740]
 [  35   42   64  231 1418]]
**********************************************************
Accuracy_NB2: 0.5174432788484996
Precision_NB2: 0.5174432788484996
ConfusionMatrix 2:
  [[ 192   63    7    2   28]
 [  91  112   59   21   50]
 [  31   82  103   90  126]
 [  15  121   86  373  657]
 [  16  101   32  300 1341]]
```

Resultats I Model

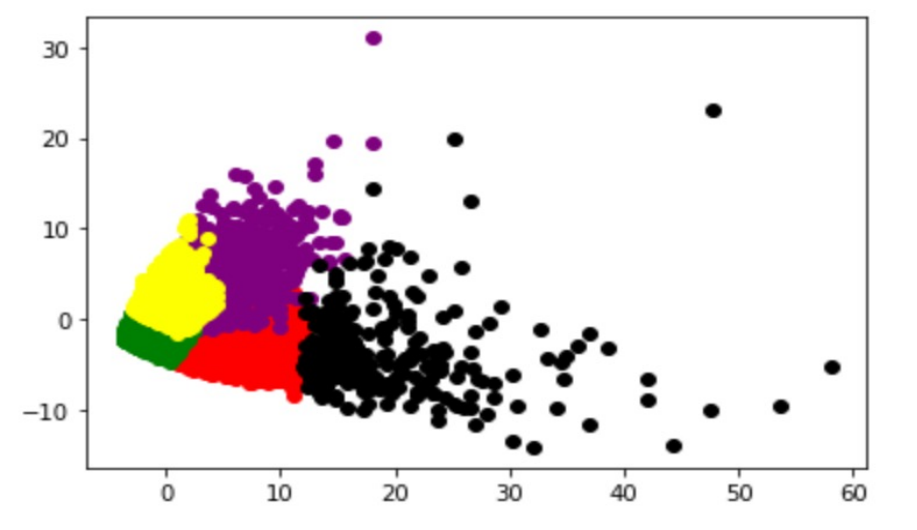
Random Forest

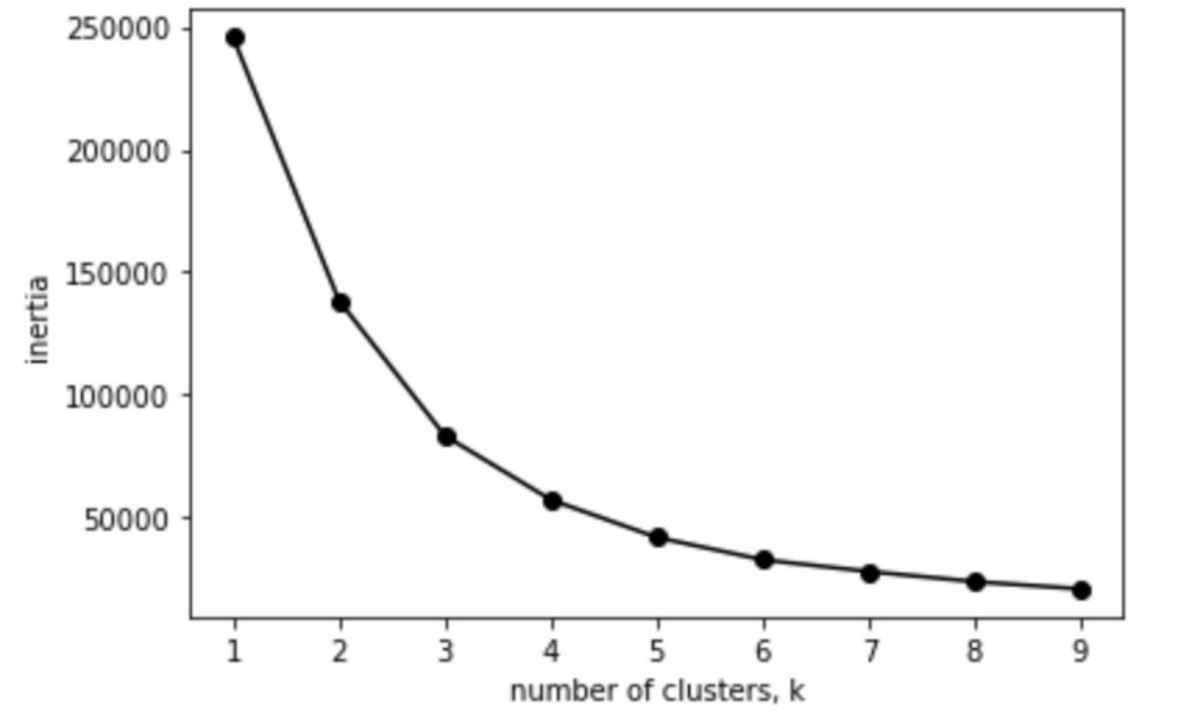
```
Accuracy: 0.4366918760673335
Precision: 0.4366918760673335
```





Kmeans Anàlisis amb PCA





KMEANS ANÀLISIS AMB T-SNE

CONCLUSIÓ