Probability And Random Variables

Hackathon

Input 1 - The flight was amazing! The seats were comfortable, the staff were very helpful, and the food was delicious. I would fly with them again.

output 1 -

```
Predicted Overall Rating: 10
Predicted Seat Comfort Rating: 5
Predicted Cabin Staff Rating: 5
Predicted Food Beverages Rating: 5
Predicted Inflight Entertainment Rating: 5
Predicted Value Money Rating: 5
```

Input 2- The flight was terrible. The seats were uncomfortable, the food was cold, and the staff was rude. The entertainment system didn't work, and it was not worth the money.

output 2-

```
Predicted Overall Rating: 1
Predicted Seat Comfort Rating: 1
Predicted Cabin Staff Rating: 1
Predicted Food Beverages Rating: 1
Predicted Inflight Entertainment Rating: 1
Predicted Value Money Rating: 1
```

Input 3- The seats were somewhat cramped, but the cabin staff was friendly, and the food was decent. Entertainment options were limited, but overall, the value for money was good.

output 3 -

```
Predicted Overall Rating: 8
Predicted Seat Comfort Rating: 4
Predicted Cabin Staff Rating: 5
Predicted Food Beverages Rating: 4
Predicted Inflight Entertainment Rating: 4
Predicted Value Money Rating: 4
```

Probability of each word for a given rating (0-10)- p(w1/r)
eg.- ratings are seat_comfort_rating.

```
word_probability_dicts['seat_comfort_rating']
```

	0	1	2	3	4	5	6	7	8	9	10
Outbound	0.004348	0.106522	0.110870	0.306522	0.321739	0.150000	0.0	0.0	0.0	0.0	0.0
flight	0.029352	0.211035	0.137868	0.240716	0.253863	0.127167	0.0	0.0	0.0	0.0	0.0
FRA/PRN	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.0	0.0	0.0	0.0	0.0
A319.	0.022472	0.112360	0.056180	0.168539	0.415730	0.224719	0.0	0.0	0.0	0.0	0.0
2	0.026476	0.231127	0.152057	0.257245	0.232379	0.100716	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...
'box'.	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
hours"	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
congestion"	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
ruse	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
Warsaw-Modlin	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0

I have calculated p(w1/r) for all the different types of ratings and stored them in dictionary, and for all the filtered words (content after removing stop words).