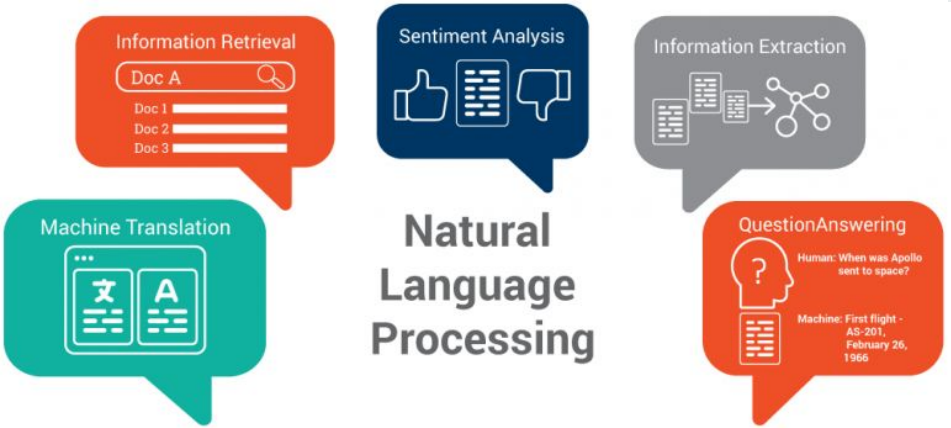
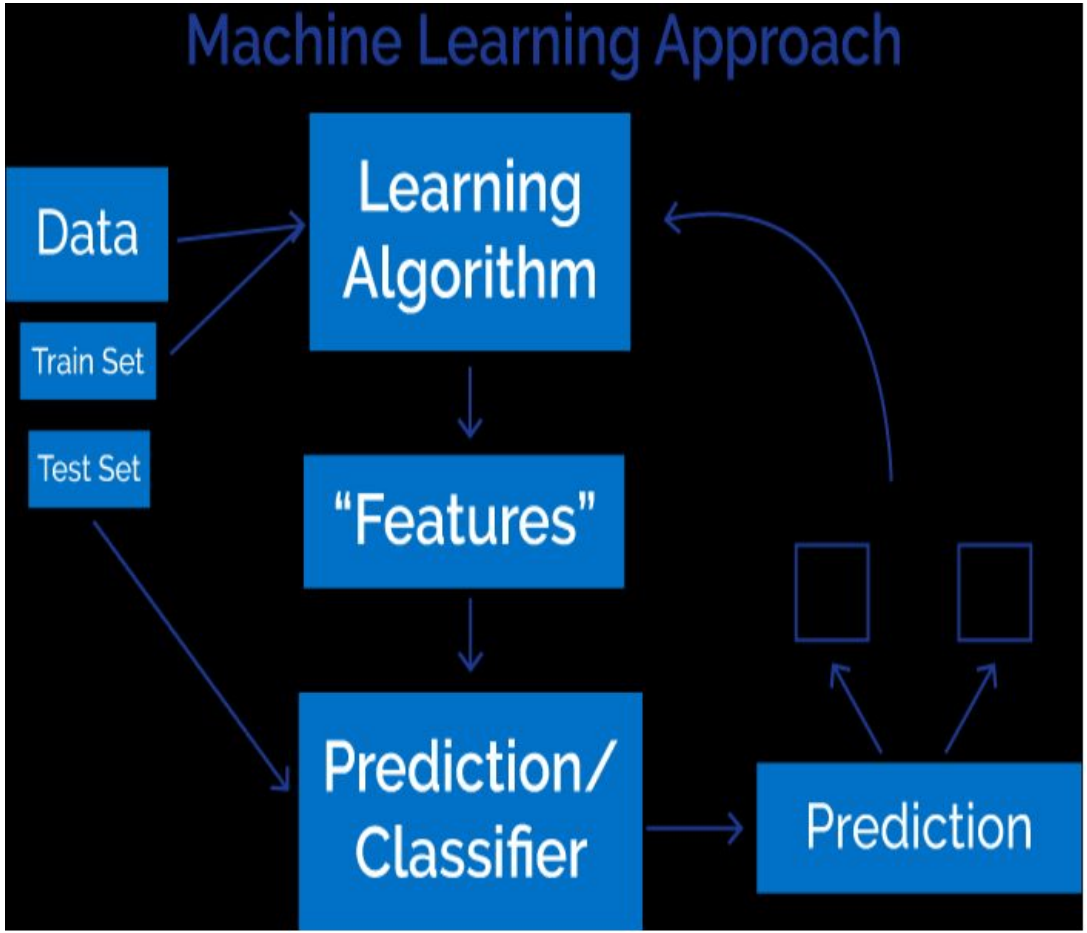
NLP analysis of Restaurant Reviews

Seminar-IT290

NLP



Natural language processing (NLP) is branch of computer science concerned with the interactions between computers and human (natural) languages.
It is about analyzing any text and handling predictive analysis.


Machine Learning Approach
Data
Train Set
Test Set
Learning Algorithm
"Features"
Prediction/ Classifier
Prediction

Classification:

Classification is the process of dividing dataset into different categories or groups by adding Labels.

Analysis of Restaurant review is a case of binary classification .The Reviews can either Positive or Negative.

Cleaning the Dataset by Removing Stopwords:

Stopwords are words which are filtered out before processing of natural language data.

Stop Words

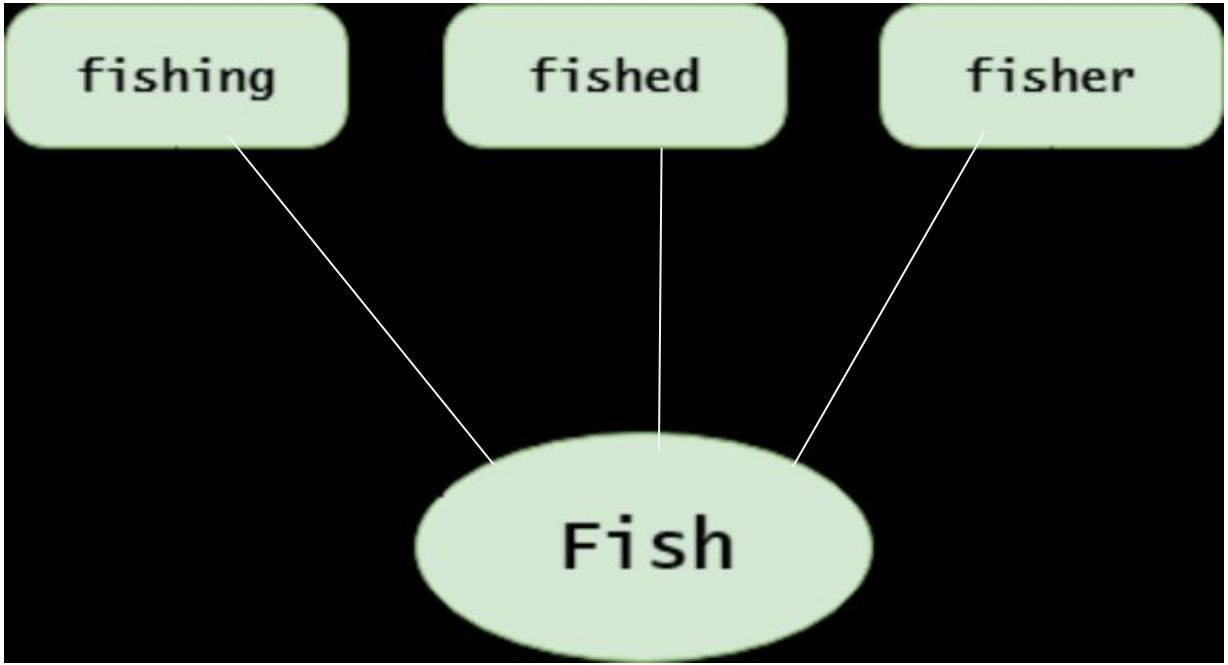
These words include:

- a
- I
- the
- in
- of
- for
- at
- to
- on
- with
- from

With Stop Words	Without Stop Words
/growing-up-with-hearing-loss/	/growing-hearing-loss/
/coming-to-terms-with-hearing-loss/	/coming-terms-hearing-loss/
/the-world-of-being-hearing-impaired/	/world-being-hearing-impaired/
/echo-in-ears-from-talking-people/	/echo-ears-from-talking-people/
/listening-is-exhausting/	/listening-exhausting/
/living-with-hearing-loss/	/living-hearing-loss/
/what-is-hearing-loss/	/what-hearing-loss/

Stemming:

Reduce words to its root word.



Dataset:

```
Review  Liked
Food arrived quickly!   1
It was not good.    0
On the up side, their cafe serves really good food. 1
Our server was fantastic and when he found out the wife loves roasted garlic and bone marrow, he added extra to
The only good thing was our waiter, he was very helpful and kept the bloddy mary's coming.  1
Best Buffet in town, for the price you cannot beat it.  1
I LOVED their mussels cooked in this wine reduction, the duck was tender, and their potato dishes were deliciou
This is one of the better buffets that I have been to.  1
So we went to Tigerlilly and had a fantastic afternoon! 1
The food was delicious, our bartender was attentive and personable AND we got a great deal! 1
The ambience is wonderful and there is music playing.   1
Will go back next trip out. 1
Sooooo good!!   1
REAL sushi lovers, let's be honest - Yama is not that good. 0
At least 40min passed in between us ordering and the food arriving, and it wasn't that busy.    0
This is a really fantastic Thai restaurant which is definitely worth a visit.   1
Nice, spicy and tender. 1
Good prices.    1
Check it out.   1
It was pretty gross!    0
I've had better atmosphere. 0
Kind of hard to mess up a steak but they did.   0
Although I very much liked the look and sound of this place, the actual experience was a bit disappointing. 0
I just don't know how this place managed to served the blandest food I have ever eaten when they are preparing
Worst service to boot, but that is the least of their worries.  0
Service was fine and the waitress was friendly. 1
The guys all had steaks, and our steak loving son who has had steak at the best and worst places said it was th
We thought you'd have to venture further away to get good sushi, but this place really hit the spot that night.
Host staff were, for lack of a better word, BITCHES!    0
Bland... Not a liking this place for a number of reasons and I don't want to waste time on bad reviewing.. I'll
Phenomenal food, service and ambiance.  1
I wouldn't return.  0
Definitely worth venturing off the strip for the pork belly, will return next time I'm in Vegas.    1
This place is way too overpriced for mediocre food. 0
Penne vodka excellent!  1
They have a good selection of food including a massive meatloaf sandwich, a crispy chicken wrap, a delish tuna
The management is rude. 0
Delicious NYC bagels, good selections of cream cheese, real Lox with capers even.   1
Great Subway, in fact it's so good when you come here every other Subway will not meet your expectations.   1
I had a seriously solid breakfast here. 1
This is one of the best bars with food in Vegas.    1
```

Name	Size	Type	Date Modified
Natural Language Processing.py	1 KB	py File	10-04-2019 22:29
Restaurant_Reviews.tsv	59 KB	tsv File	10-04-2019 22:29

Help | Variable explorer | File explorer

IPython console

Console 1/A

```
[nltk_data]   Package stopwords is already up-to-date!

In [6]: runfile('C:/Users/mSaurabh/Downloads/Machine-Learning-The-future-master/Natural
Language Processing/Natural Language Processing.py', wdir='C:/Users/mSaurabh/Downloads/
Machine-Learning-The-future-master/Natural Language Processing')
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\mSaurabh\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!

In [7]: runfile('C:/Users/mSaurabh/Downloads/Machine-Learning-The-future-master/Natural
Language Processing/Natural Language Processing.py', wdir='C:/Users/mSaurabh/Downloads/
Machine-Learning-The-future-master/Natural Language Processing')
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\mSaurabh\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!

In [8]: runfile('C:/Users/mSaurabh/Downloads/Machine-Learning-The-future-master/Natural
Language Processing/Natural Language Processing.py', wdir='C:/Users/mSaurabh/Downloads/
Machine-Learning-The-future-master/Natural Language Processing')
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\mSaurabh\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!

In [9]:
```

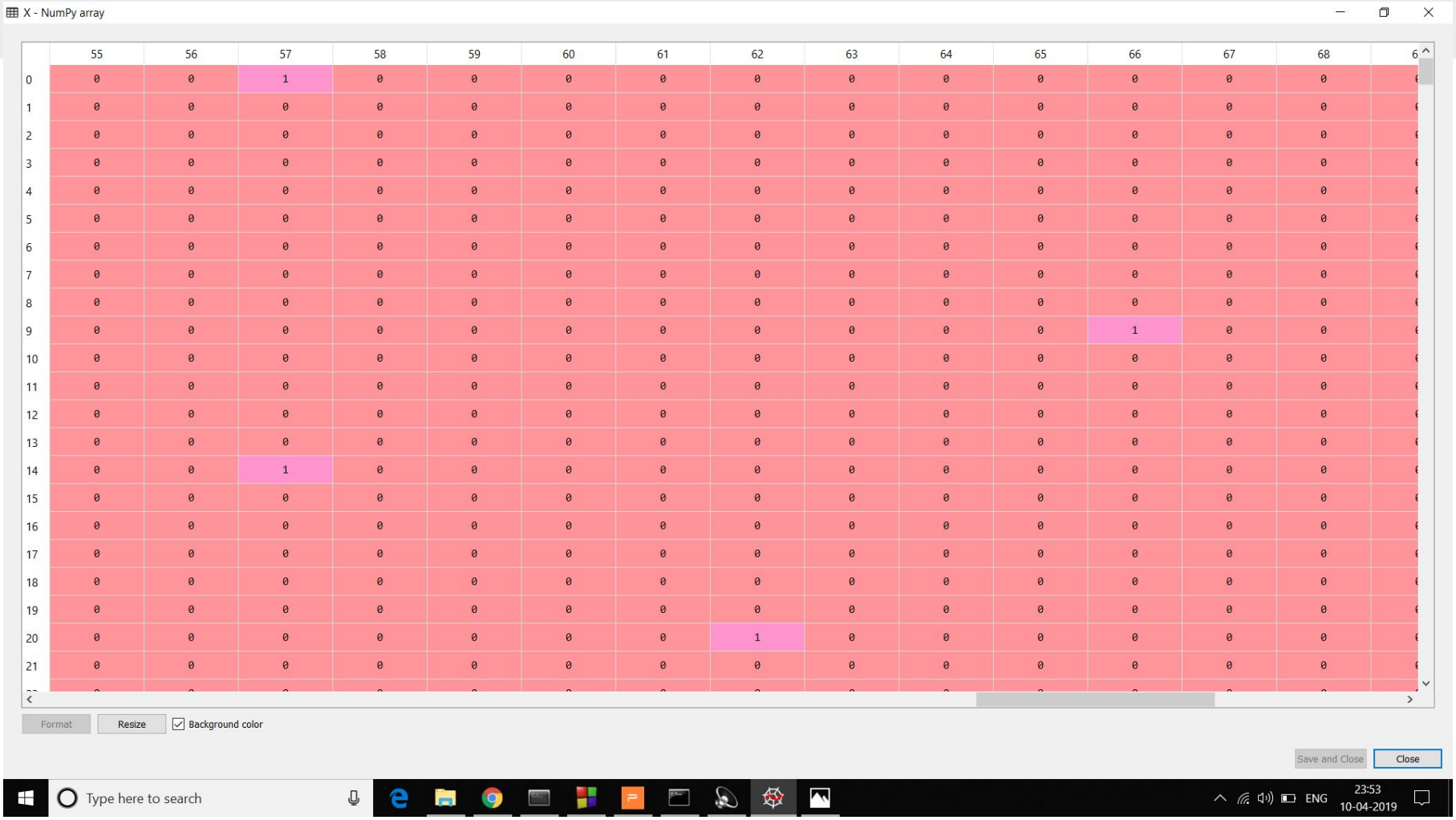
IPython console | History log

Dataset after Stemming and removing Stopwords:

Index	Type	Size	Value
0	unicode	1	food arriv quickli
1	str	1	good
2	unicode	1	side cafe serv realli good food
3	unicode	1	server fantast found wife love roast garlic bone marrow ad extra meal ...
4	unicode	1	good thing waiter help kept bloddi mari come
5	str	1	best buffet town price cannot beat
6	unicode	1	love mussel cook wine reduct duck tender potato dish delici
7	unicode	1	one better buffet
8	unicode	1	went tigerlilli fantast afternoon
9	unicode	1	food delici bartend attent person got great deal
10	unicode	1	ambienc wonder music play
11	str	1	go back next trip
12	str	1	sooooo good
13	unicode	1	real sushi lover let honest yama good
14	unicode	1	least min pass us order food arriv busi
15	unicode	1	realli fantast thai restaur definit worth visit
16	unicode	1	nice spici tender
17	unicode	1	good price
18	str	1	check
19	unicode	1	pretti gross
20	unicode	1	better atmospher
21	unicode	1	kind hard mess steak
22	unicode	1	although much like look sound place actual experi bit disappoint
23	unicode	1	know place manag serv blandest food ever eaten prepar indian cuisin

Feature Extraction:

1. The process of extracting data from the files is called feature extraction.
2. The goal of feature extraction is to obtain a set of informative and non-redundant data.


X - NumPy array
55 56 57 58 59 60 61 62 63 64 65 66 67 68
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7 0 0 0 0 0 0 0 0 0 0 0 0 0 0
8 0 0 0 0 0 0 0 0 0 0 0 0 0 0
9 0 0 0 0 0 0 0 0 0 0 0 1 0 0
10 0 0 0 0 0 0 0 0 0 0 0 0 0 0
11 0 0 0 0 0 0 0 0 0 0 0 0 0 0
12 0 0 0 0 0 0 0 0 0 0 0 0 0 0
13 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 0 0 1 0 0 0 0 0 0 0 0 0 0 0
15 0 0 0 0 0 0 0 0 0 0 0 0 0 0
16 0 0 0 0 0 0 0 0 0 0 0 0 0 0
17 0 0 0 0 0 0 0 0 0 0 0 0 0 0
18 0 0 0 0 0 0 0 0 0 0 0 0 0 0
19 0 0 0 0 0 0 0 0 0 0 0 0 0 0
20 0 0 0 0 0 0 0 1 0 0 0 0 0 0
21 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Format
Resize
Background color
Save and Close
Close
Type here to search
ENG
23:53
10-04-2019

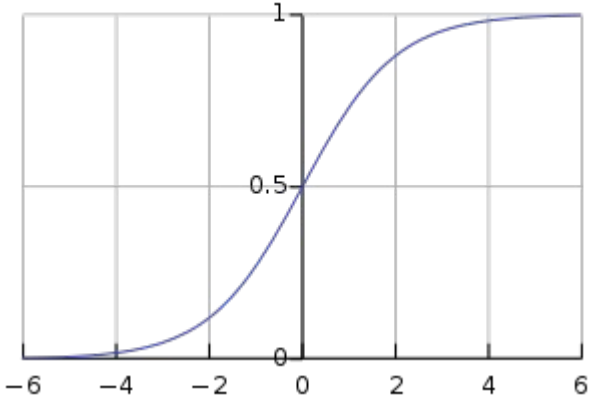
Logistic Regression:Classifier for Prediction

$$h(x_i) = g(\beta^T x_i) = \frac{1}{1 + e^{-\beta^T x_i}}$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

If h<0.5 , y=0;
If h>=0.5 , y=1;

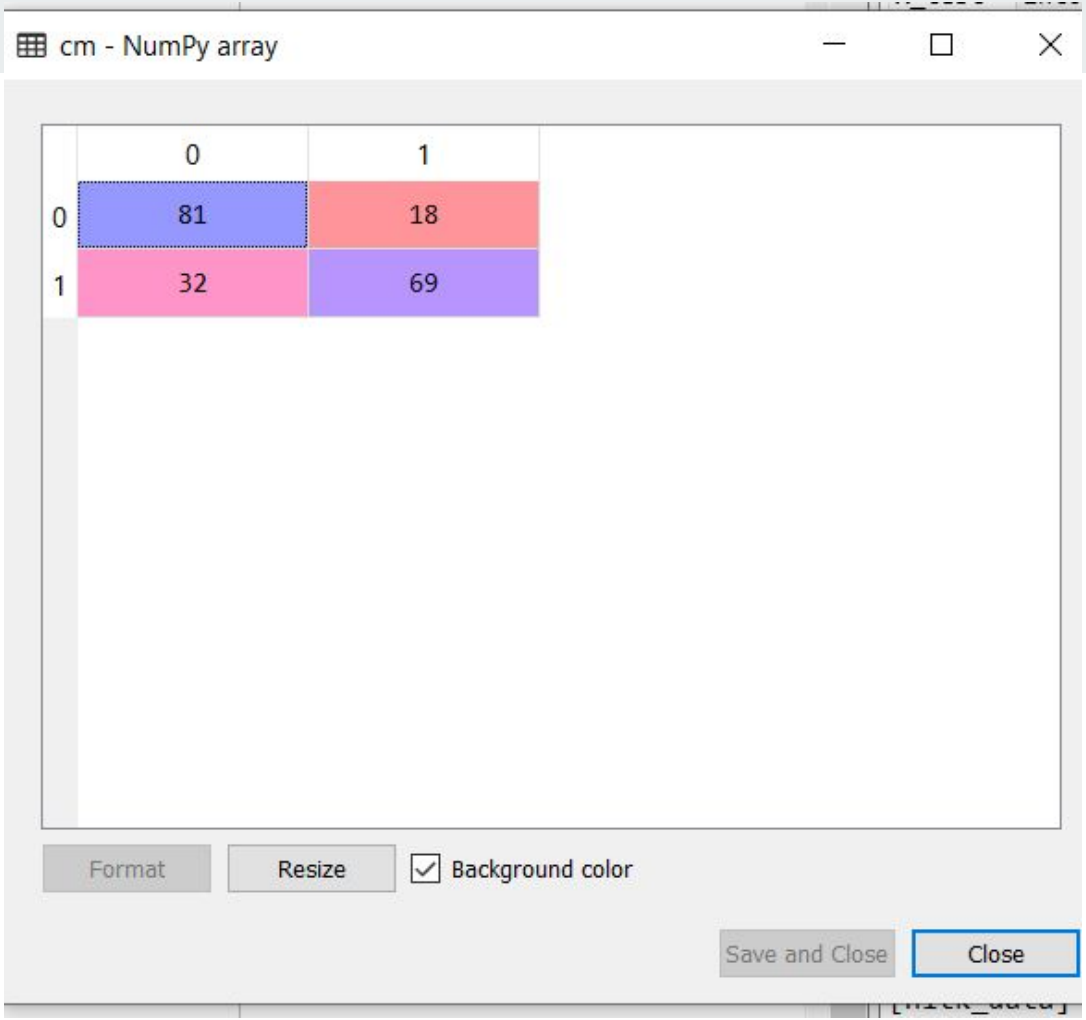
Sigmoid Function[g(z)]



Confusion Matrix:Accuracy using confusion matrix

	Class 1 Predicted	*Class 2 Predicted*
Class 1 Actual	**TP**	**FN**
Class 2 Actual	**FP**	**TN**

Accuracy = (TP + TN)/(TP+TN+FN+FP)



Accuracy=(81+69)/(81+18+32+69)
=0.75