

# IMPORTING LIBRARIES

```
In [1]: import numpy as np
import pandas as pd
```

## DATA PRE - PROCESSING

```
In [36]: data = pd.read_csv("C:\\Users\\ASHISH TIWARI\\Downloads\\Fake_News_Data.csv")
data
```

```
Out[36]:
```

	URLs	Headline	Body	Label
0	http://www.bbc.com/news/world-us-canada-414191...	Four ways Bob Corker skewered Donald Trump	Image copyright Getty Images\nOn Sunday mornin...	1
1	https://www.reuters.com/article/us-filmfestiva...	Linklater's war veteran comedy speaks to moder...	LONDON (Reuters) - "Last Flag Flying", a comed...	1
2	https://www.nytimes.com/2017/10/09/us/politics...	Trump's Fight With Corker Jeopardizes His Legi...	The feud broke into public view last week when...	1
3	https://www.reuters.com/article/us-mexico-oil-...	Egypt's Cheiron wins tie-up with Pemex for Mex...	MEXICO CITY (Reuters) - Egypt's Cheiron Holdin...	1
4	http://www.cnn.com/videos/cnnmoney/2017/10/08/...	Jason Aldean opens 'SNL' with Vegas tribute	Country singer Jason Aldean, who was performin...	1
...	...	...	...	...
4004	http://beforeitsnews.com/sports/2017/09/trends...	Trends to Watch	Trends to Watch\n% of readers think this story...	0
4005	http://beforeitsnews.com/u-s-politics/2017/10/...	Trump Jr. Is Soon To Give A 30-Minute Speech F...	Trump Jr. Is Soon To Give A 30-Minute Speech F...	0
4006	https://www.activistpost.com/2017/09/ron-paul-...	Ron Paul on Trump, Anarchism & the AltRight	NaN	0
4007	https://www.reuters.com/article/us-china-pharm...	China to accept overseas trial data in bid to ...	SHANGHAI (Reuters) - China said it plans to ac...	1
4008	http://beforeitsnews.com/u-s-politics/2017/10/...	Vice President Mike Pence Leaves NFL Game Beca...	Vice President Mike Pence Leaves NFL Game Beca...	0

4009 rows × 4 columns

```
In [37]: data.shape
```

```
Out[37]: (4009, 4)
```

```
In [38]: data.isnull().sum()
```

```
Out[38]: URLs      0
Headline    0
Body        21
Label       0
dtype: int64
```

```
In [39]: data.dropna(inplace=True)
data.isnull().sum()
```

```
Out[39]: URLs      0
Headline    0
Body        0
Label       0
dtype: int64
```

```
In [29]: data.shape
```

```
Out[29]: (3988, 4)
```

```
In [40]: data.head(10)
```

	URLs	Headline	Body	Label
0	http://www.bbc.com/news/world-us-canada-414191...	Four ways Bob Corker skewered Donald Trump	Image copyright Getty Images\nOn Sunday mornin...	1
1	https://www.reuters.com/article/us-filmfestiva...	Linklater's war veteran comedy speaks to moder...	LONDON (Reuters) - "Last Flag Flying", a comed...	1
2	https://www.nytimes.com/2017/10/09/us/politics...	Trump's Fight With Corker Jeopardizes His Legi...	The feud broke into public view last week when...	1
3	https://www.reuters.com/article/us-mexico-oil-...	Egypt's Cheiron wins tie-up with Pemex for Mex...	MEXICO CITY (Reuters) - Egypt's Cheiron Holdin...	1
4	http://www.cnn.com/videos/cnnmoney/2017/10/08/...	Jason Aldean opens 'SNL' with Vegas tribute	Country singer Jason Aldean, who was performin...	1
5	http://beforeitsnews.com/sports/2017/09/jetnat...	JetNation FanDuel League; Week 4	JetNation FanDuel League; Week 4\n% of readers...	0
6	https://www.nytimes.com/2017/10/10/us/politics...	Kansas Tried a Tax Plan Similar to Trump's. It...	In 2012, Kansas lawmakers, led by Gov. Sam Bro...	1
7	https://www.reuters.com/article/us-india-cenba...	India RBI chief: growth important, but not at ...	The Reserve Bank of India (RBI) Governor Urjit...	1
8	https://www.reuters.com/article/us-climatechan...	EPA chief to sign rule on Clean Power Plan exi...	Scott Pruitt, Administrator of the U.S. Enviro...	1
9	https://www.reuters.com/article/us-air-berlin-...	Talks on sale of Air Berlin planes to easyJet ...	FILE PHOTO - An Air Berlin sign is seen at an ...	1

```
In [41]: data['news'] = data['Headline'] + data['Body']
data
```

Out[41]:

	URLs	Headline	Body	Label	news
0	http://www.bbc.com/news/world-us-canada-414191...	Four ways Bob Corker skewered Donald Trump	Image copyright Getty Images\nOn Sunday mornin...	1	Four ways Bob Corker skewered Donald Trump\nImagin...
1	https://www.reuters.com/article/us-filmfestiva...	Linklater's war veteran comedy speaks to moder...	LONDON (Reuters) - "Last Flag Flying", a comed...	1	Linklater's war veteran comedy speaks to moder...
2	https://www.nytimes.com/2017/10/09/us/politics...	Trump's Fight With Corker Jeopardizes His Legi...	The feud broke into public view last week when...	1	Trump's Fight With Corker Jeopardizes His Legi...
3	https://www.reuters.com/article/us-mexico-oil-...	Egypt's Cheiron wins tie-up with Pemex for Mex...	MEXICO CITY (Reuters) - Egypt's Cheiron Holdin...	1	Egypt's Cheiron wins tie-up with Pemex for Mex...
4	http://www.cnn.com/videos/cnnmoney/2017/10/08/...	Jason Aldean opens 'SNL' with Vegas tribute	Country singer Jason Aldean, who was performin...	1	Jason Aldean opens 'SNL' with Vegas tributeCou...
...	...	...	...	...	...
4003	http://beforeitsnews.com/u-s-politics/2017/10/...	CNN and Globalist Exposed - Steve Quayle and A...	Vietnam Is in Great Danger, You Must Publish a...	0	CNN and Globalist Exposed - Steve Quayle and A...
4004	http://beforeitsnews.com/sports/2017/09/trends...	Trends to Watch	Trends to Watch\n% of readers think this story...	0	Trends to WatchTrends to Watch\n% of readers t...
4005	http://beforeitsnews.com/u-s-politics/2017/10/...	Trump Jr. Is Soon To Give A 30-Minute Speech F...	Trump Jr. Is Soon To Give A 30-Minute Speech F...	0	Trump Jr. Is Soon To Give A 30-Minute Speech F...
4007	https://www.reuters.com/article/us-china-pharm...	China to accept overseas trial data in bid to ...	SHANGHAI (Reuters) - China said it plans to ac...	1	China to accept overseas trial data in bid to ...
4008	http://beforeitsnews.com/u-s-politics/2017/10/...	Vice President Mike Pence Leaves NFL Game Beca...	Vice President Mike Pence Leaves NFL Game Beca...	0	Vice President Mike Pence Leaves NFL Game Beca...

3988 rows × 5 columns

In [42]: `data.drop(['URLs', 'Headline', 'Body'], axis=1, inplace=True)`

In [43]: `data`

Out[43]:

	Label	news
0	1	Four ways Bob Corker skewered Donald TrumpImag...
1	1	Linklater's war veteran comedy speaks to moder...
2	1	Trump's Fight With Corker Jeopardizes His Legi...
3	1	Egypt's Cheiron wins tie-up with Pemex for Mex...
4	1	Jason Aldean opens 'SNL' with Vegas tributeCou...
...	...	...
4003	0	CNN and Globalist Exposed - Steve Quayle and A...
4004	0	Trends to WatchTrends to Watch\n% of readers t...
4005	0	Trump Jr. Is Soon To Give A 30-Minute Speech F...
4007	1	China to accept overseas trial data in bid to ...
4008	0	Vice President Mike Pence Leaves NFL Game Beca...

3988 rows × 2 columns

```
In [47]: X = pd.DataFrame(data['news'], columns=['news'])
Y = pd.DataFrame(data['Label'], columns=['Label'])
```

```
In [48]: X.head()
```

Out[48]:

	news
0	Four ways Bob Corker skewered Donald TrumpImag...
1	Linklater's war veteran comedy speaks to moder...
2	Trump's Fight With Corker Jeopardizes His Legi...
3	Egypt's Cheiron wins tie-up with Pemex for Mex...
4	Jason Aldean opens 'SNL' with Vegas tributeCou...

```
In [49]: Y.shape
```

Out[49]: (3988, 1)

```
In [82]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X.news, Y.Label, test_size=0.21, random_state=5)
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer()
train_data = vectorizer.fit_transform(X_train)
test_data = vectorizer.transform(X_test)
```

## MODEL TRAINING (RandomForestClassifier, LINEAR SVM , LogisticRegression)

```
In [71]: from sklearn.ensemble import RandomForestClassifier

rfc = RandomForestClassifier()
rfc.fit(train_data, y_train)
rfc_predict = rfc.predict(test_data)
```

```
In [72]: from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix, precision_score, recall_score, f1_score
import matplotlib.pyplot as plt
import seaborn as sns

rfc_acc = accuracy_score(y_test, rfc_predict) * 100
rfc_prec = precision_score(y_test, rfc_predict) * 100
rfc_rec = recall_score(y_test, rfc_predict) * 100
rfc_f1 = f1_score(y_test, rfc_predict) * 100
```

```

rfc_cm = confusion_matrix(y_test, rfc_predict)
labels = ['True Neg', 'False Pos', 'False Neg', 'True Pos']
labels = np.asarray(labels).reshape(2,2)

print('RANDOM FOREST CLASSIFIER\n')
print('Accuracy: {:.2f}%\n'.format(rfc_acc))
print('Precision: {:.2f}%\n'.format(rfc_prec))
print('Recall: {:.2f}%\n'.format(rfc_rec))
print('F1 Score: {:.2f}%\n'.format(rfc_f1))
print('Confusion Matrix:')
sns.heatmap(rfc_cm, annot=labels, fmt='')
plt.show()

```

RANDOM FOREST CLASSIFIER

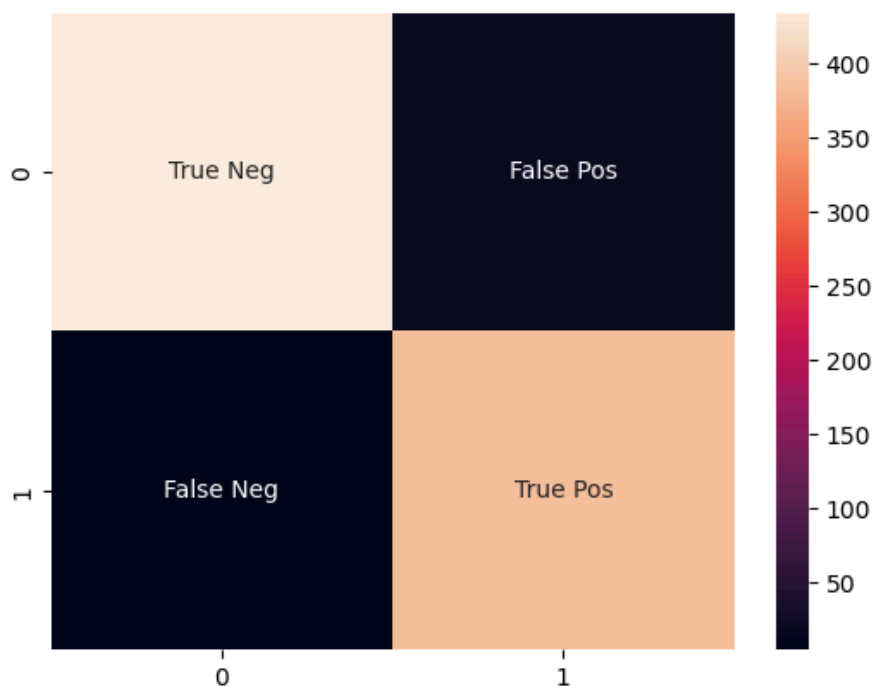
Accuracy: 97.37%

Precision: 95.50%

Recall: 98.96%

F1 Score: 97.20%

Confusion Matrix:



In [73]: `from sklearn.svm import LinearSVC`

```

svm = LinearSVC()
svm.fit(train_data, y_train)
svm_predict = svm.predict(test_data)

```

C:\ANACONDA\Lib\site-packages\sklearn\svm\\_classes.py:32: FutureWarning: The default value of `dual` will change from `True` to `auto` in 1.5. Set the value of `dual` explicitly to suppress the warning.  
 warnings.warn(

In [74]:

```

svm_acc = accuracy_score(y_test, svm_predict) * 100
svm_prec = precision_score(y_test, svm_predict) * 100
svm_rec = recall_score(y_test, svm_predict) * 100
svm_f1 = f1_score(y_test, svm_predict) * 100
svm_cm = confusion_matrix(y_test, svm_predict)
labels = ['True Neg', 'False Pos', 'False Neg', 'True Pos']
labels = np.asarray(labels).reshape(2,2)

print('LINEAR SVM KERNEL\n')
print('Accuracy: {:.2f}%\n'.format(svm_acc))
print('Precision: {:.2f}%\n'.format(svm_prec))
print('Recall: {:.2f}%\n'.format(svm_rec))
print('F1 Score: {:.2f}%\n'.format(svm_f1))

```

```
print('Confusion Matrix:')
sns.heatmap(svm_cm, annot=labels, fmt='')
plt.show()
```

LINEAR SVM KERNEL

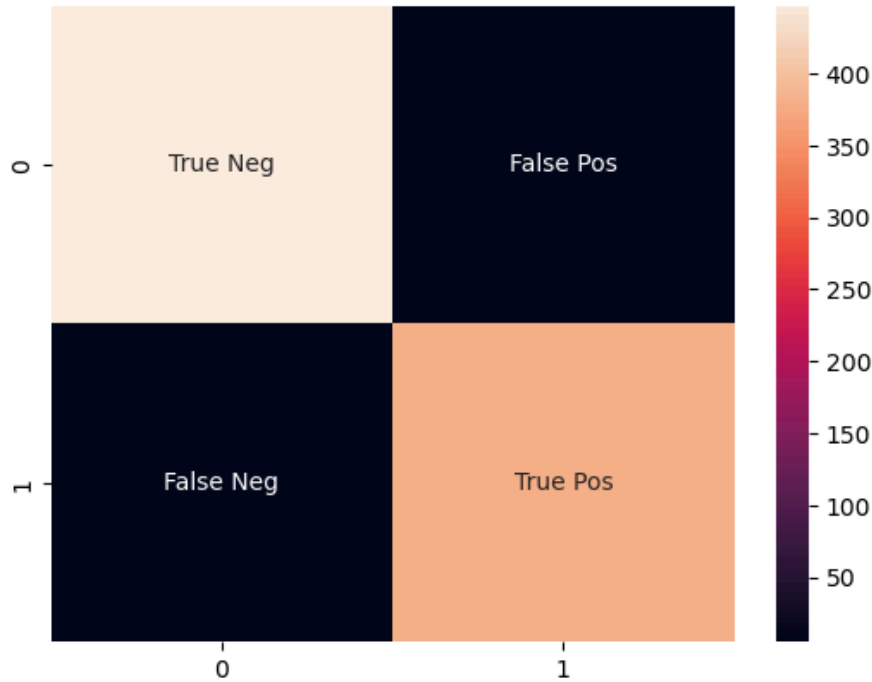
Accuracy: 98.45%

Precision: 98.69%

Recall: 97.93%

F1 Score: 98.31%

Confusion Matrix:



```
In [83]: from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
log.fit(train_data, y_train)
```

```
Out[83]: ▼ LogisticRegression
LogisticRegression()
```

```
In [84]: lr_predict = log.predict(test_data)
```

```
In [85]: lr_acc = accuracy_score(y_test, lr_predict) * 100
lr_prec = precision_score(y_test, lr_predict) * 100
lr_rec = recall_score(y_test, lr_predict) * 100
lr_f1 = f1_score(y_test, lr_predict) * 100
lr_cm = confusion_matrix(y_test, lr_predict)
labels = ['True Neg', 'False Pos', 'False Neg', 'True Pos']
labels = np.asarray(labels).reshape(2,2)

print('LogisticRegression\n')
print('Accuracy: {:.2f}%\n'.format(lr_acc))
print('Precision: {:.2f}%\n'.format(lr_prec))
print('Recall: {:.2f}%\n'.format(lr_rec))
print('F1 Score: {:.2f}%\n'.format(lr_f1))
print('Confusion Matrix:')
sns.heatmap(svm_cm, annot=labels, fmt='')
plt.show()
```

LogisticRegression

Accuracy: 95.58%

Precision: 94.63%

Recall: 95.85%

F1 Score: 95.24%

Confusion Matrix:

