In [2]:
```python
import pandas as pd
import numpy as np
import seaborn as sn
import matplotlib.pyplot as plt
```

In [4]:
```python
path = "Fake.csv"
path2 = "True.csv"
```

In [6]:
```python
true = pd.read_csv(path)
fals = pd.read_csv(path2)
```

In [8]:
```python
true.head(3)
```

Out[8]:

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017

In [10]:
```python
true = true.drop(["date" , "title" , "subject"] , axis = 1)
fals = fals.drop(["date" ,"title" , "subject"] , axis = 1)
```

In [12]:
```python
true.head(3)
```

Out[12]:

	text
0	Donald Trump just couldn t wish all Americans ...
1	House Intelligence Committee Chairman Devin Nu...
2	On Friday, it was revealed that former Milwauk...

In [14]:
```python
true['label'] = 1
fals['label'] = 0
```

In [16]:
```python
true.head(3)
```

Out[16]:

	text	label
0	Donald Trump just couldn t wish all Americans ...	1
1	House Intelligence Committee Chairman Devin Nu...	1
2	On Friday, it was revealed that former Milwauk...	1

In [18]:
```python
fals.head(3)
```

Out[18]:

	text	label
0	WASHINGTON (Reuters) - The head of a conservat...	0
1	WASHINGTON (Reuters) - Transgender people will...	0
2	WASHINGTON (Reuters) - The special counsel inv...	0

In [20]:
```python
data = pd.concat([true, fals])
data = data.sample(frac=1).reset_index(drop=True)
```

In [22]:
```python
data.head()
```

Out[22]:

	text	label
0	WASHINGTON (Reuters) - U.S. officials are disc...	0
1	LONDON (Reuters) - It was meant to silence her...	0
2	The function of socialism is to raise sufferin...	1
3	When far left publications like Democracy Now ...	1
4	DAR ES SALAAM (Reuters) - Tanzanian opposition...	0

In [24]:
```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44898 entries, 0 to 44897
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   text    44898 non-null  object
 1   label   44898 non-null  int64
dtypes: int64(1), object(1)
memory usage: 701.7+ KB
```

In [26]:
```python
data.isnull().sum()
```

Out[26]:
```
text     0
label    0
dtype: int64
```

In [28]:
```python
data.duplicated().sum()
```

Out[28]: 6251

In [30]:
```python
data.shape
```

Out[30]: (44898, 2)

In [32]:
```python
data = data.drop_duplicates()
```

In [34]:
```python
data.duplicated().sum()
```

Out[34]: 0

In [36]:
```python
import string
def remove_punctuation(text):
    return text.translate(str.maketrans('', '', string.punctuation))
```

In [38]:
```python
data['text'] = data['text'].apply(remove_punctuation)
```

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

Out[40]:

	text	label
0	WASHINGTON Reuters US officials are discussin...	0
1	LONDON Reuters It was meant to silence her cr...	0
2	The function of socialism is to raise sufferin...	1
3	When far left publications like Democracy Now ...	1
4	DAR ES SALAAM Reuters Tanzanian opposition la...	0

In [42]:
```python
from wordcloud import WordCloud
```

In [44]:
```python
fake_news = data[data["label"] == 0]
true_news = data[data["label"] == 1]
```

In [46]:
```python
print(fake_news.shape)
print(true_news.shape)
```

```
(21192, 2)
(17455, 2)
```

In [48]:
```python
type(fake_news)
type(true_news)
```

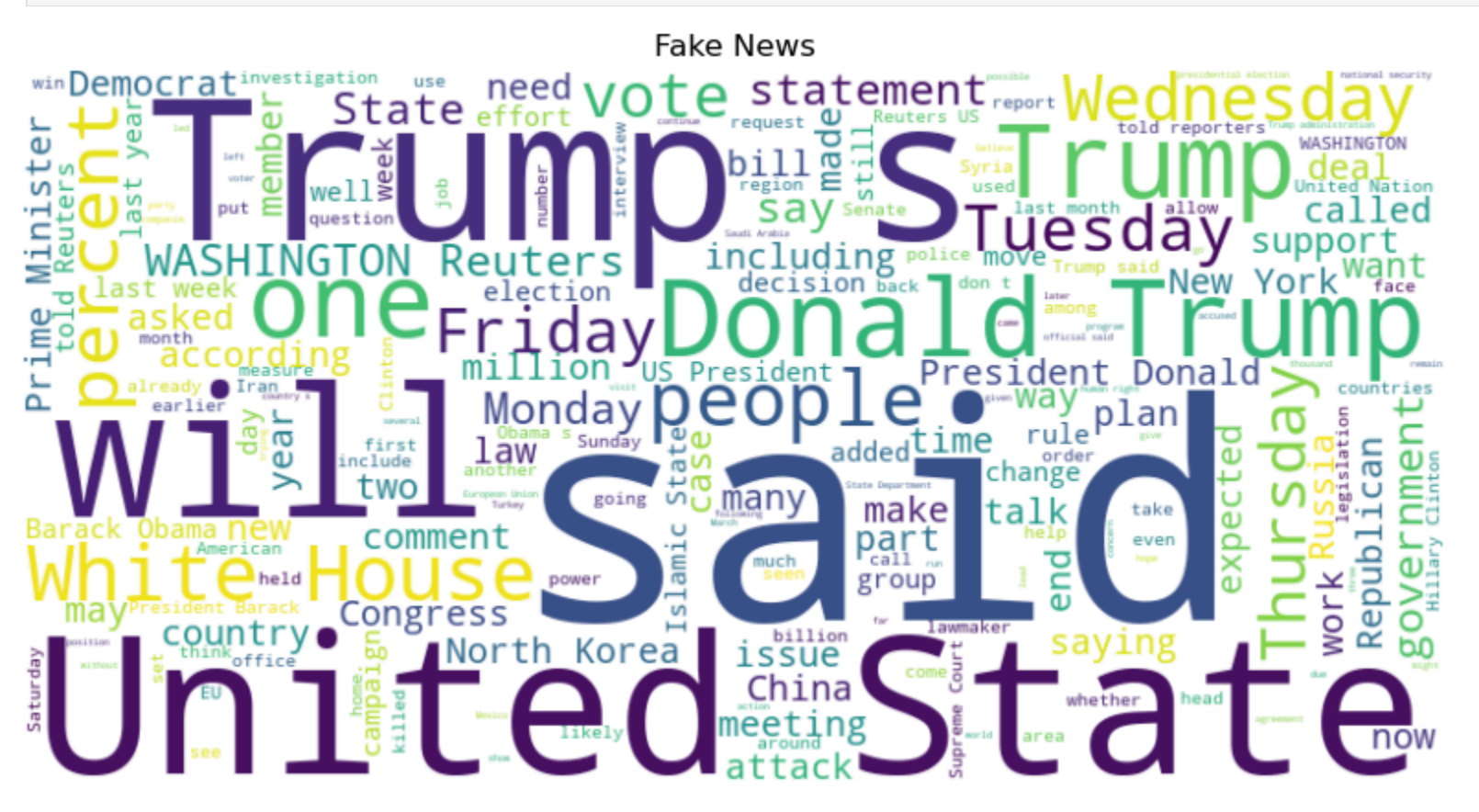
Out[48]: pandas.core.frame.DataFrame

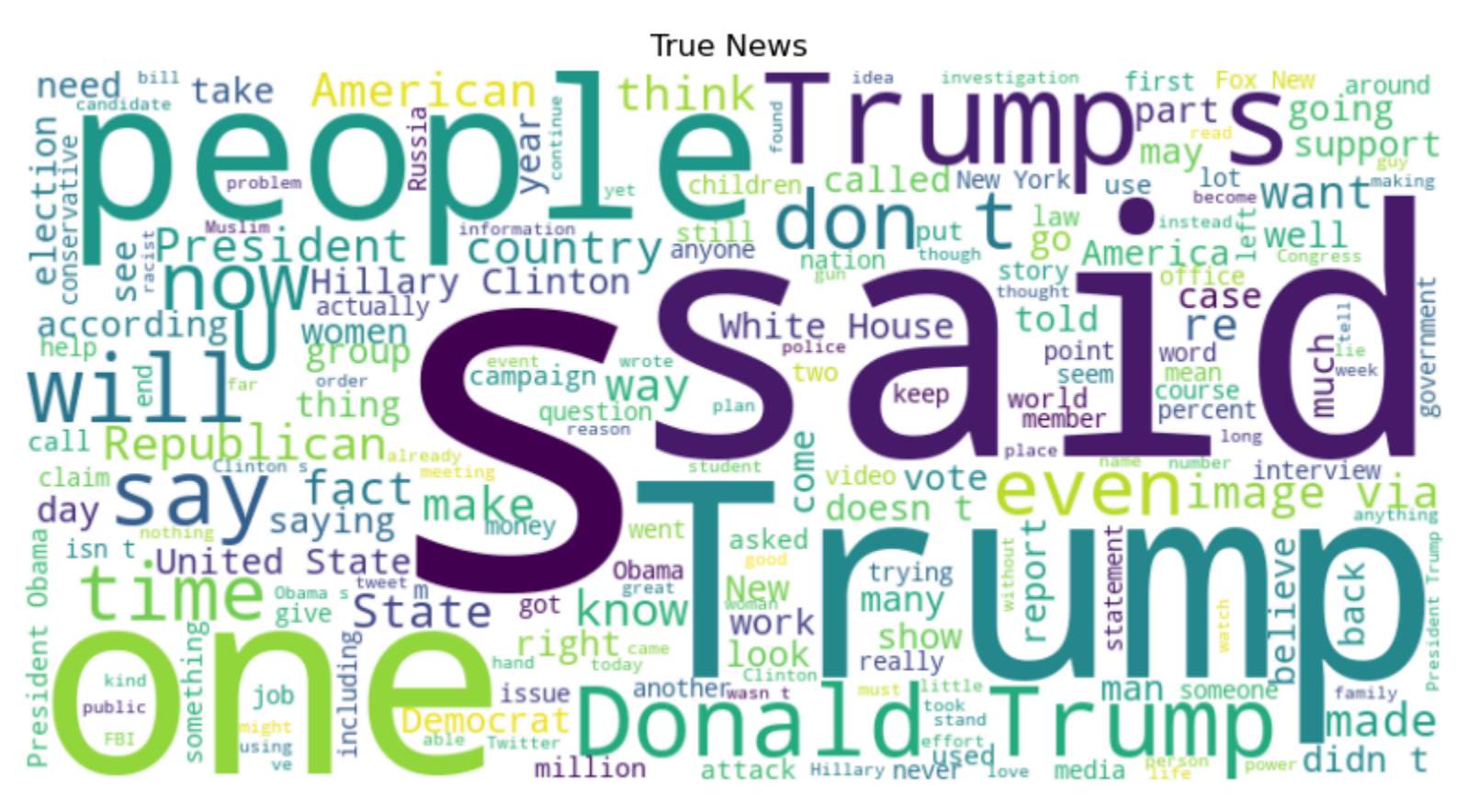
In [50]:
```python
fake_news.columns
true_news.columns
```

Out[50]: Index(['text', 'label'], dtype='object')

In [52]:
```python
fake_news_text = ' '.join(fake_news["text"].astype(str))
true_news_text = ' '.join(true_news["text"].astype(str))
```

In [54]:
```python
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(fake_news_text)
plt.figure(figsize=(10, 10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("Fake News")
plt.show()

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(true_news_text)
plt.figure(figsize=(10, 10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title("True News ")
plt.show()
```





In [56]:
```python
data.head()
```

Out[56]:

	text	label
0	WASHINGTON Reuters US officials are discussin...	0
1	LONDON Reuters It was meant to silence her cr...	0
2	The function of socialism is to raise sufferin...	1
3	When far left publications like Democracy Now ...	1
4	DAR ES SALAAM Reuters Tanzanian opposition la...	0

In [58]:
```python
from sklearn.feature_extraction.text import TfidfVectorizer
tdf = TfidfVectorizer()
```

In [60]:
```python
x = data[["text"]]
y = data["label"]
```

In [75]:
```python
x
```

Out[75]:
```
<38647x225975 sparse matrix of type '<class 'numpy.float64'>'
	with 8068580 stored elements in Compressed Sparse Row format>
```

In [ ]:
```python
x = tdf.fit_transform(x["text"])
```

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

In [ ]:
```python
x.shape
```

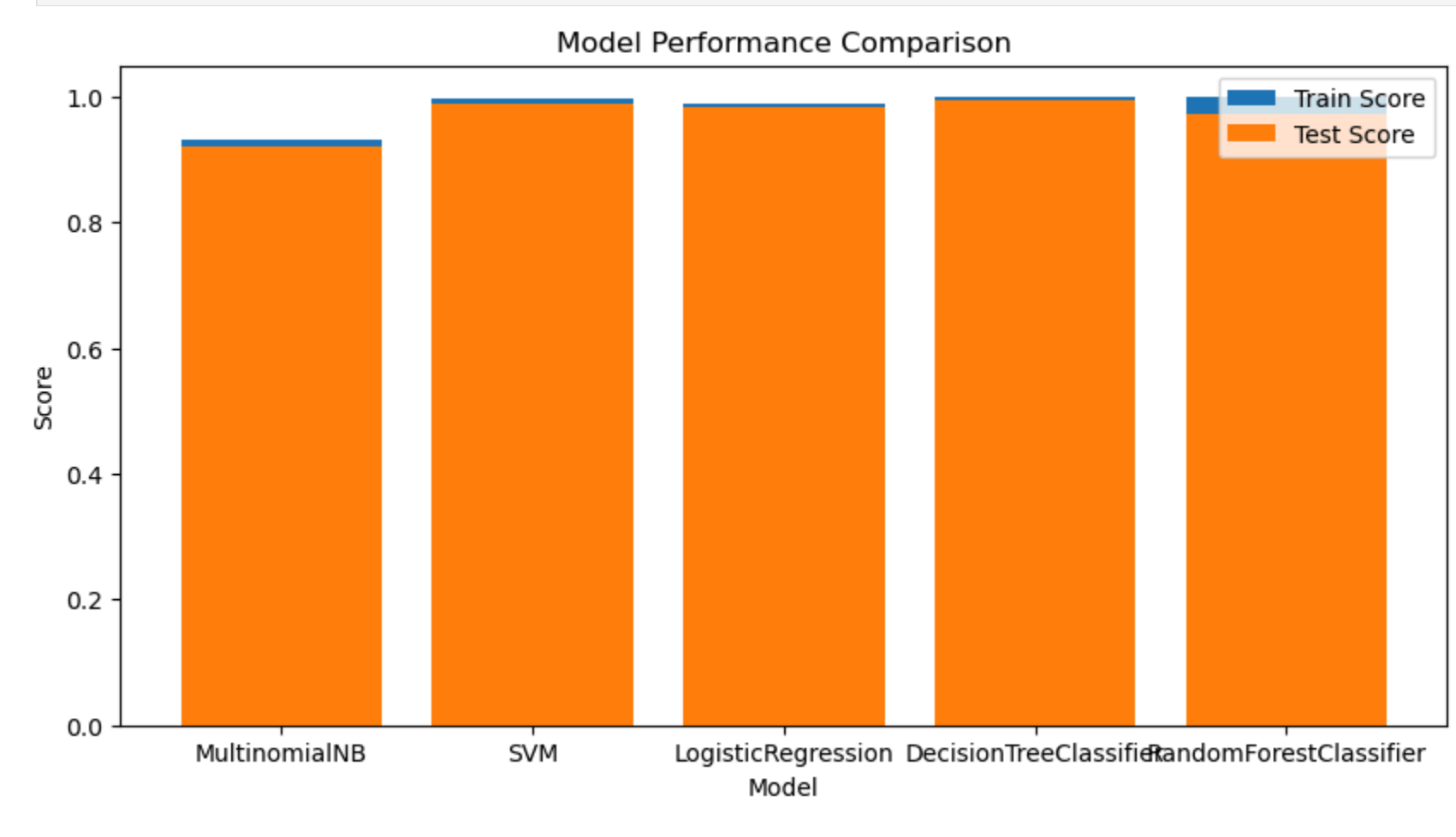
In [70]:
```python
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.30, random_state=41)
```

In [72]:
```python
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
mb = MultinomialNB()
svm = SVC()
lr  = LogisticRegression()
dct = DecisionTreeClassifier()
rfc = RandomForestClassifier()
```

In [74]:
```python
mb.fit(x_train , y_train)
svm.fit(x_train , y_train)
lr.fit(x_train , y_train)
dct.fit(x_train , y_train)
rfc.fit(x_train , y_train)
```

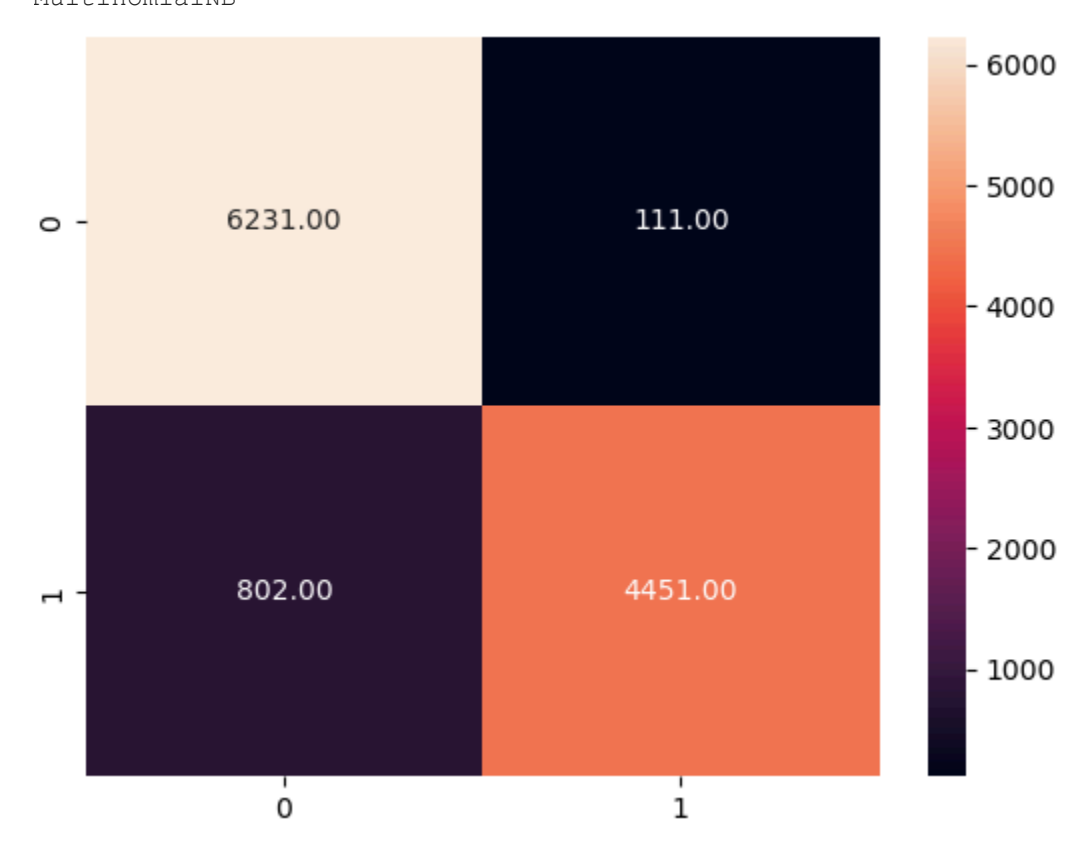
Out[74]:
```
RandomForestClassifier
RandomForestClassifier()
```

In [81]:
```python
plt.figure(figsize=(10,5))

train_scores = [mb.score(x_train, y_train), svm.score(x_train, y_train), lr.score(x_train, y_train), dct.score(x_train, y_train), rfc.score(x_train, y_train)]
test_scores = [mb.score(x_test, y_test),svm.score(x_test, y_test), lr.score(x_test, y_test), dct.score(x_test, y_test), rfc.score(x_test, y_test)]

model_names = ['MultinomialNB', 'SVM', 'LogisticRegression', 'DecisionTreeClassifier', 'RandomForestClassifier']

plt.bar(model_names, train_scores, label="Train Score")
plt.bar(model_names, test_scores, label="Test Score")

plt.xlabel('Model')
plt.ylabel('Score')
plt.title('Model Performance Comparison')
plt.legend()
plt.show()
```

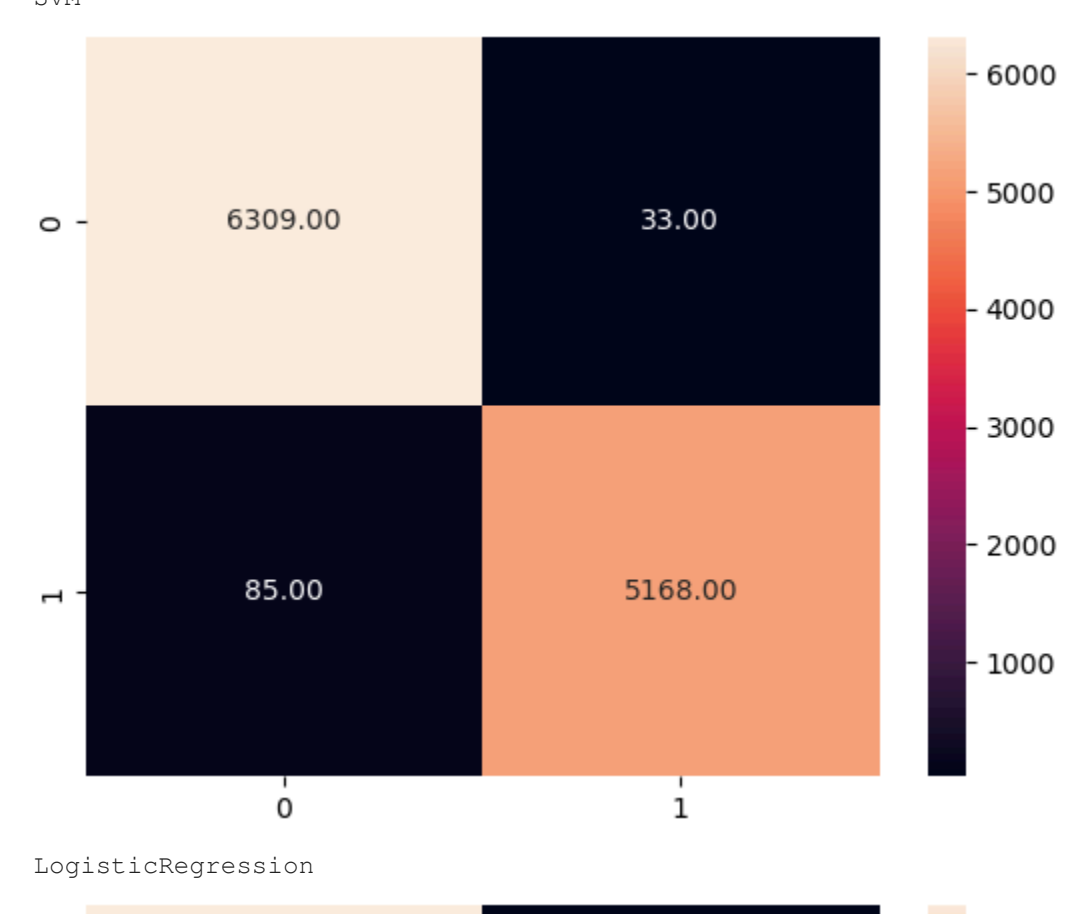


In [83]:
```python
from sklearn.metrics import confusion_matrix ,classification_report
import seaborn as sns
mod = [mb, svm, lr,dct ,rfc]
model_name = ['MultinomialNB', 'SVM', 'LogisticRegression', 'DecisionTreeClassifier', 'RandomForestClassifier']
for model, model_name in zip(mod, model_name):
    corr = confusion_matrix(y_test, model.predict(x_test))
    print(model_name)
    sns.heatmap(corr, annot=True, fmt="0.2f")
    plt.show()
```

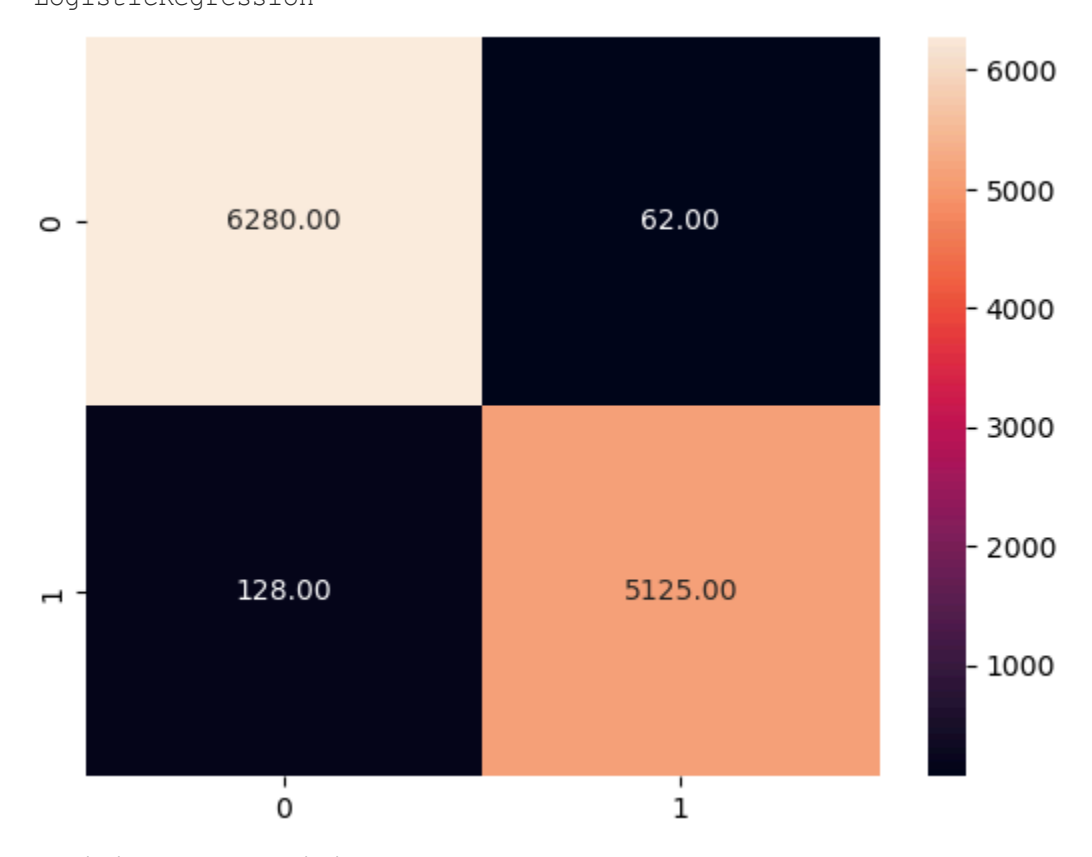
MultinomialNB



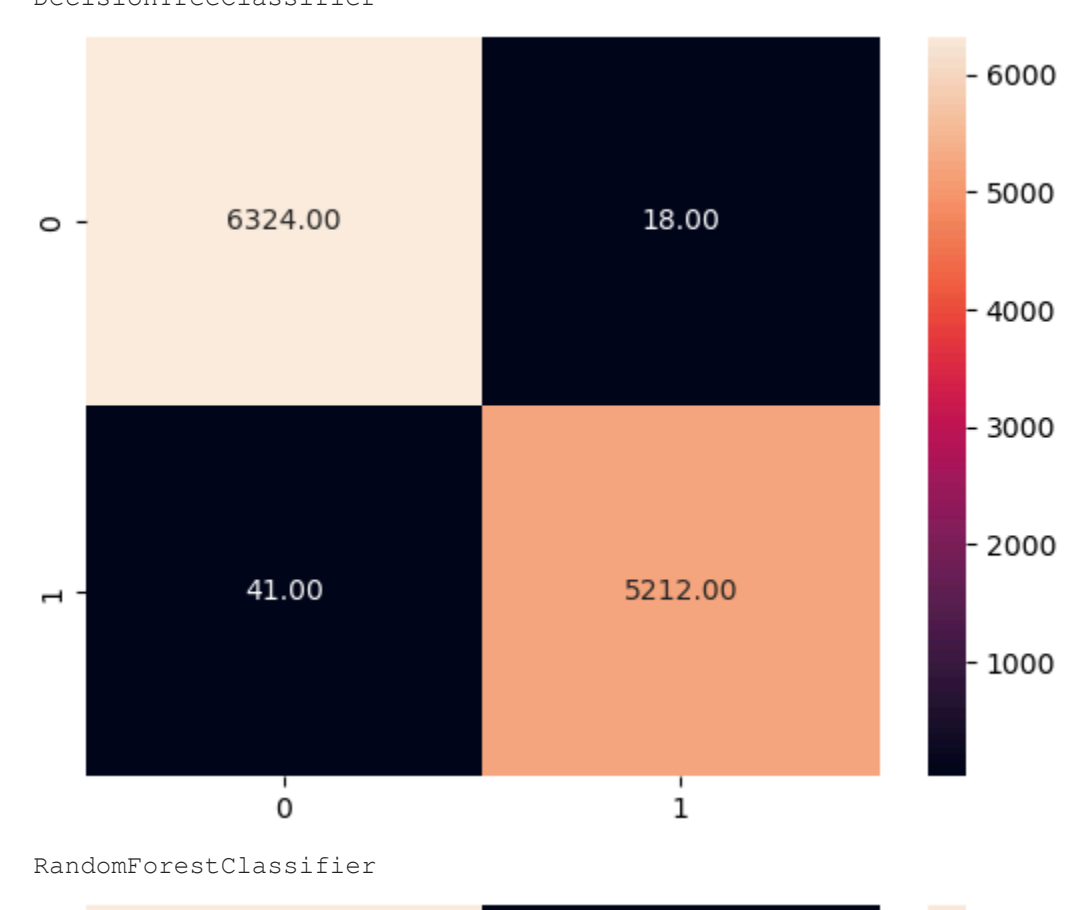
SVM



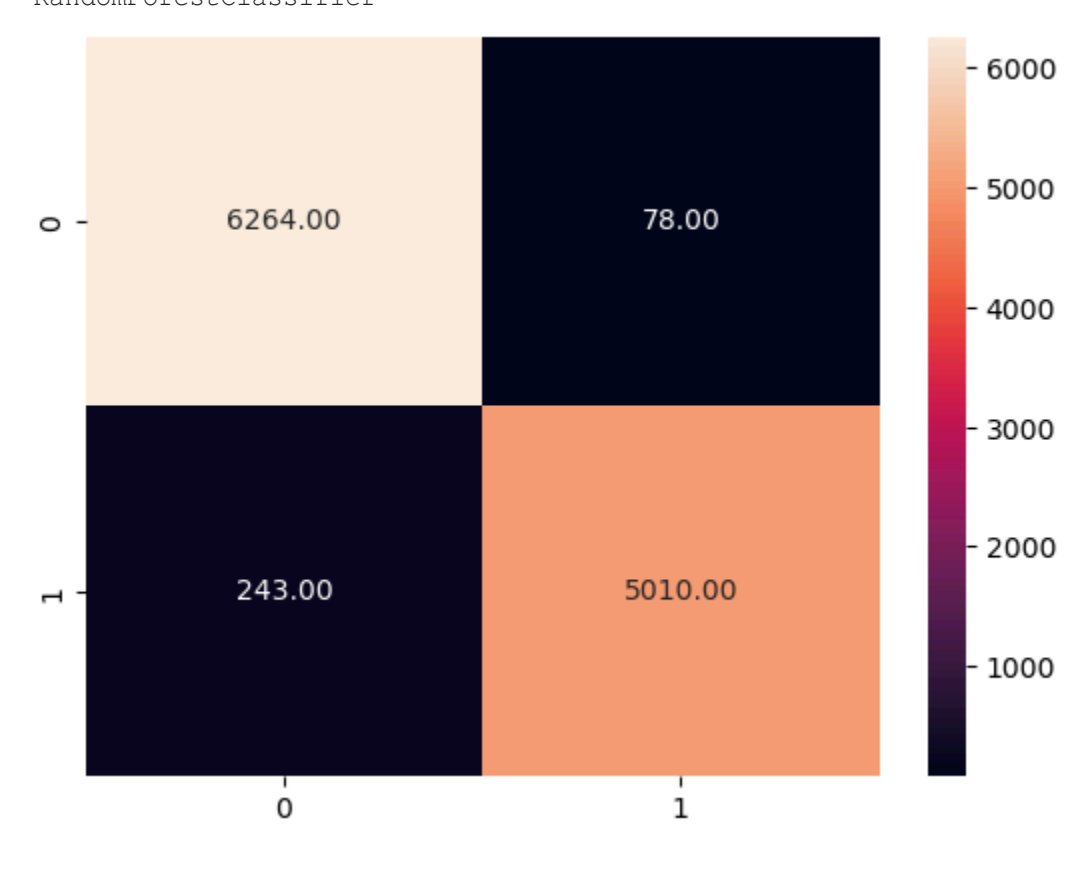
LogisticRegression



DecisionTreeClassifier



RandomForestClassifier



In [84]:
```python
for model, model_name in zip(mod, model_names):

    print(model_name)
    print(classification_report(y_test,model.predict(x_test)))
```

```
MultinomialNB
              precision    recall  f1-score   support

           0       0.89      0.98      0.93      6342
           1       0.98      0.85      0.91      5253

    accuracy                           0.92     11595
   macro avg       0.93      0.91      0.92     11595
weighted avg       0.93      0.92      0.92     11595

SVM
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      6342
           1       0.99      0.98      0.99      5253

    accuracy                           0.99     11595
   macro avg       0.99      0.99      0.99     11595
weighted avg       0.99      0.99      0.99     11595

LogisticRegression
              precision    recall  f1-score   support

           0       0.98      0.99      0.99      6342
           1       0.99      0.98      0.98      5253

    accuracy                           0.98     11595
   macro avg       0.98      0.98      0.98     11595
weighted avg       0.98      0.98      0.98     11595

DecisionTreeClassifier
              precision    recall  f1-score   support

           0       0.99      1.00      1.00      6342
           1       1.00      0.99      0.99      5253

    accuracy                           0.99     11595
   macro avg       1.00      0.99      0.99     11595
weighted avg       0.99      0.99      0.99     11595

RandomForestClassifier
              precision    recall  f1-score   support

           0       0.96      0.99      0.98      6342
           1       0.98      0.95      0.97      5253

    accuracy                           0.97     11595
   macro avg       0.97      0.97      0.97     11595
```

```
weighted avg       0.97      0.97      0.97     11546

```