In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
pure = pd.read_csv("../reports/knn_0.3.csv")
pca = pd.read_csv("../reports/pca_then_knn_0.3.csv")
```

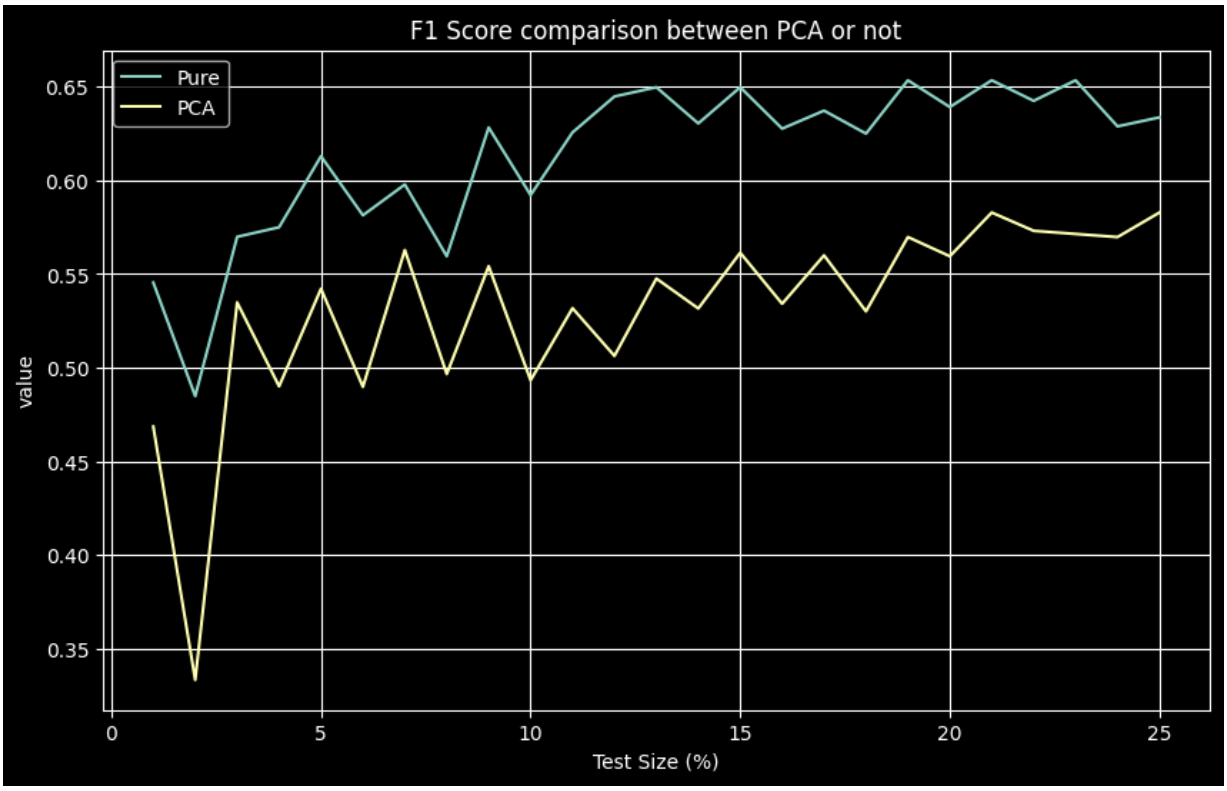
In [3]:
```
pure.head()
pca.head()
```

Out[3]:

	k	F1 Score	Accuracy	Precision	Recall
0	1	0.468571	0.523077	0.455556	0.482353
1	2	0.333333	0.548718	0.468085	0.258824
2	3	0.534884	0.589744	0.528736	0.541176
3	4	0.490066	0.605128	0.560606	0.435294
4	5	0.542169	0.610256	0.555556	0.529412

In [4]:
```
plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure['F1 Score'], label='Pure')
plt.plot(pca['k'], pca['F1 Score'], label='PCA')

plt.xlabel('Test Size (%)')
plt.ylabel('value')
plt.title('F1 Score comparison between PCA or not')
plt.legend()
plt.grid(True)
plt.show()
```

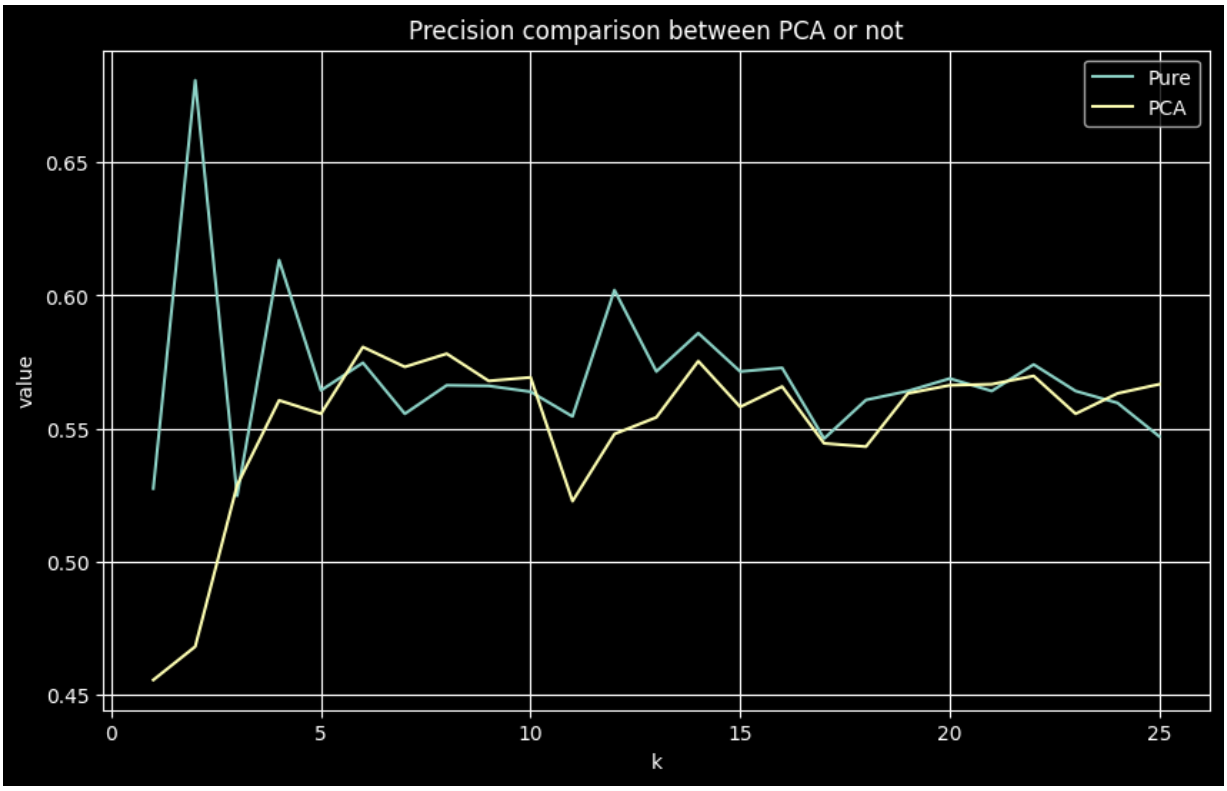


In [5]:
```
plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure["Accuracy"], label='Pure', )
plt.plot(pca['k'], pca['Accuracy'], label='PCA')

plt.xlabel('k')
plt.ylabel('value')
plt.title('Accuracy comparison between PCA or not')
plt.legend()
plt.grid(True)
plt.show()
```

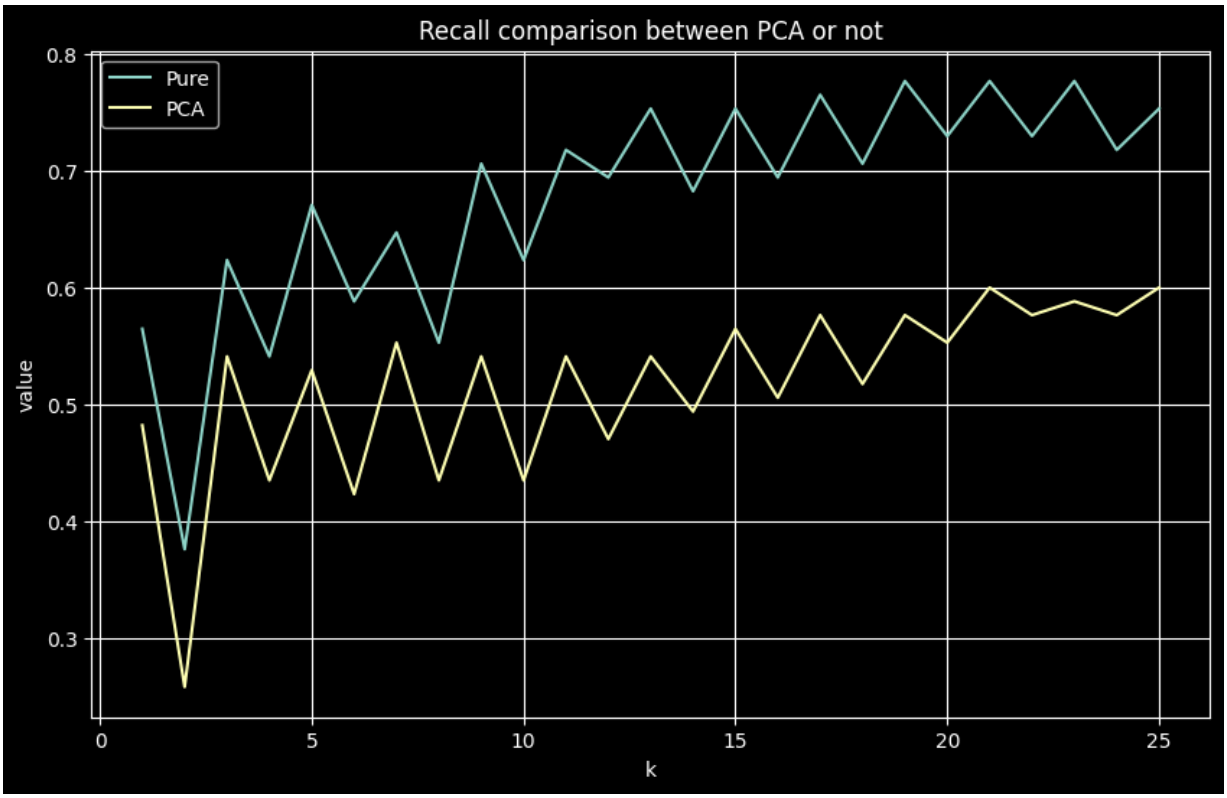
In [6]:
```
plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure["Precision"], label='Pure', )
plt.plot(pca['k'], pca['Precision'], label='PCA')

plt.xlabel('k')
plt.ylabel('value')
plt.title('Precision comparison between PCA or not(knn, train:test = 0.3')
plt.legend()
plt.grid(True)
plt.show()
```



In [7]:
```
plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure["Recall"], label='Pure', )
plt.plot(pca['k'], pca['Recall'], label='PCA')

plt.xlabel('k')
plt.ylabel('value')
plt.title('Recall comparison between PCA or not(knn, train:test = 0.3')
plt.legend()
plt.grid(True)
plt.show()
```


Recall comparison between PCA or not
Pure
PCA
value
k
0.8
0.7
0.6
0.5
0.4
0.3
0
5
10
15
20
25