

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

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

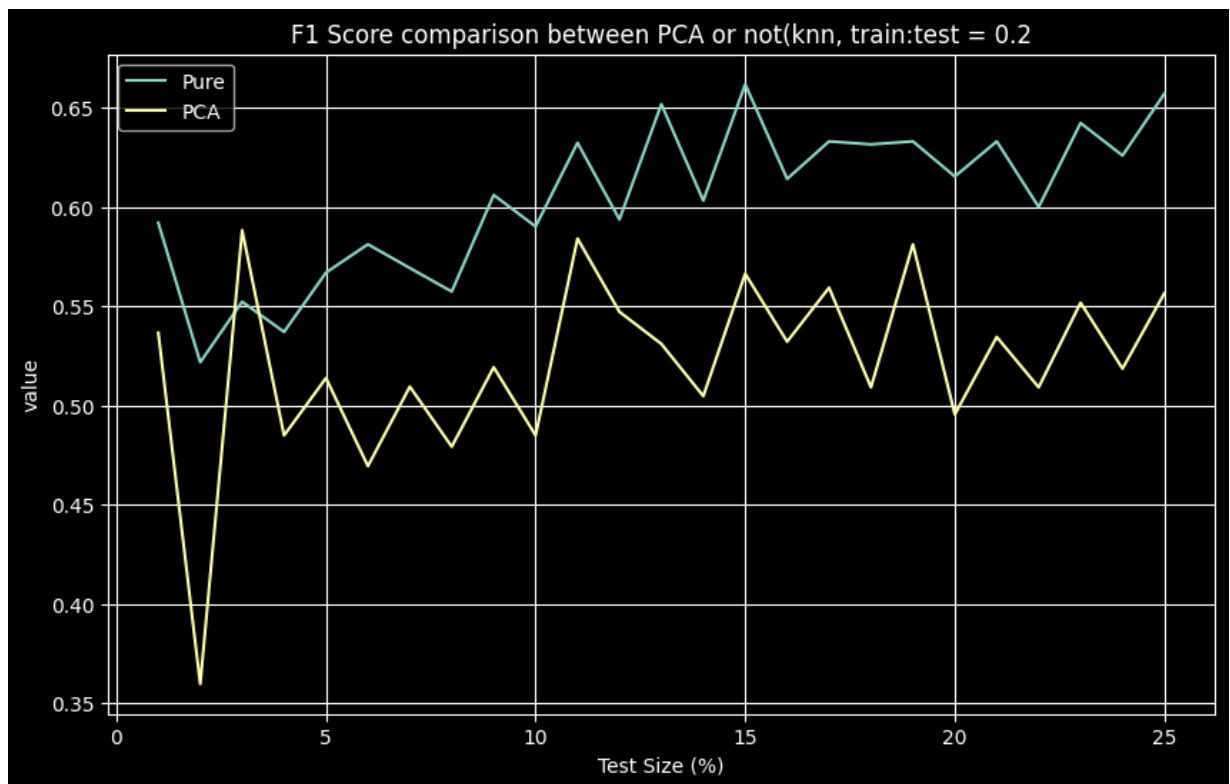
```
In [3]: pure.head()
pca.head()
```

```
Out[3]:
```

	k	F1 Score	Accuracy	Precision	Recall
0	1	0.536585	0.561538	0.523810	0.550000
1	2	0.359551	0.561538	0.551724	0.266667
2	3	0.588235	0.623077	0.593220	0.583333
3	4	0.484848	0.607692	0.615385	0.400000
4	5	0.513761	0.592308	0.571429	0.466667

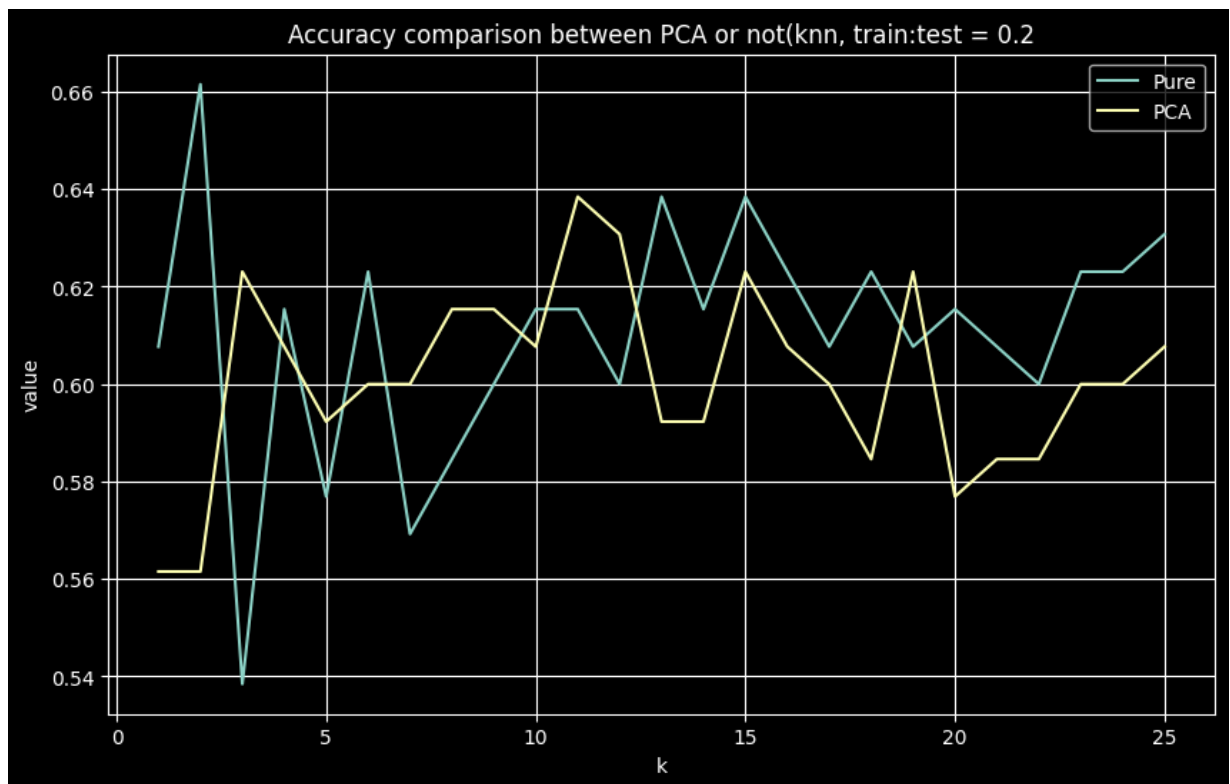
```
In [4]: plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure['F1 Score'], label='Pure')
plt.plot(pca['k'], pca['F1 Score'], label='PCA')

plt.xlabel('Test Size (%)')
plt.ylabel('value')
plt.title('F1 Score comparison between PCA or not(knn, train:test = 0.2)')
plt.legend()
plt.grid(True)
plt.show()
```



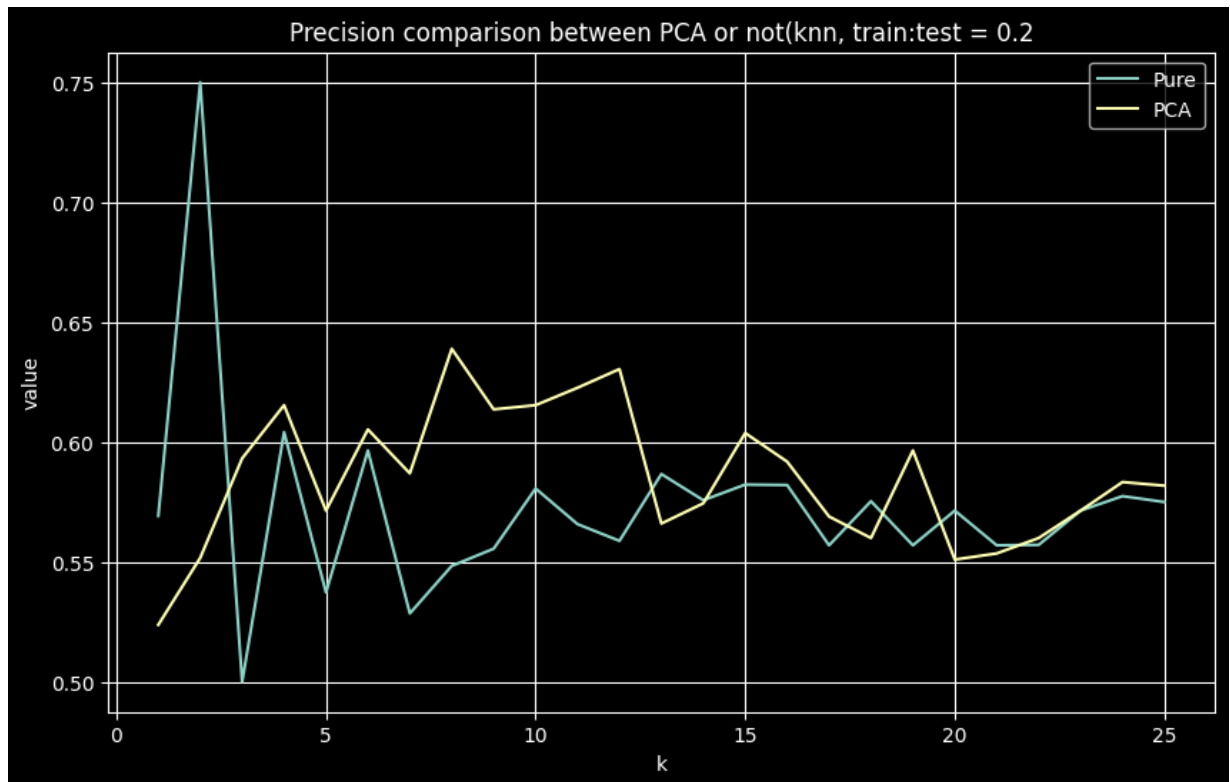
```
In [5]: plt.style.use('dark_background')
plt.figure(figsize=(10, 6))
plt.plot(pure['k'], pure["Accuracy"], label='Pure', )
plt.plot(pca['k'], pca['Accuracy'], label='PCA')

plt.xlabel('k')
plt.ylabel('value')
plt.title('Accuracy comparison between PCA or not(knn, train:test = 0.2)')
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 = 0.2)')
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 = 0.2')
plt.legend()
plt.grid(True)
plt.show()
```

