In [ ]:
```
print(dataset_array.shape)
print(labels_array.shape)
```

In [ ]:
```
X_train, X_test, y_train, y_test = train_test_split(dataset_array, labels_array,
```

In [ ]:
```
scaler = StandardScaler()
X_trained_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

In [ ]:
```
# The ratio tells us how much information (or variance) from the original
# dataset is captured by each principal component (PC)
pca = PCA(n_components=0.65)
X_train_pca = pca.fit_transform(X_trained_scaled)
X_test_pca = pca.transform(X_test_scaled)
print("Original shape:", dataset_array.shape)
print("PCA Transformed shape:", X_train_pca.shape)
explained_variance = pca.explained_variance_ratio_.sum()
print(explained_variance)
```

In [ ]:
```
print(sum(pca.explained_variance_ratio_[:300]))
```

In [ ]:
```
cumsum = np.cumsum(pca.explained_variance_ratio_)
print(cumsum)
```

In [ ]:
```
# Explained variance as a function of the number of dimensions
plt.figure(figsize=(7,5))
plt.plot(cumsum, marker='o')
plt.xlabel("Dimensions")
plt.ylabel("Explained Variance")
plt.grid()
plt.show()
```