

# Análisis de Datos con PCA en el Conjunto de Datos del Vino

Angel Espinosa, Gonzalo Alipio, Julieth Mahecha  
Universidad INFOTEP

April 2, 2025

## 1 Introducción

Desarrollamos un análisis sobre el conjunto de datos del vino mediante PCA.

## 2 Importar librerías y cargar datos

```
# Importar librerías
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Cargar datos
df = pd.read_csv("wine.csv")
```

Se importan las librerías correspondientes para cargar los datos del archivo con la información de los vinos.

## 3 Preparando datos para PCA

```
# Normalizar los datos
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df.iloc[:, 1:])
```

## 4 Aplicando PCA

```
# Aplicar PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)
```

## 5 Visualizando resultados

```
# Crear gráfico de dispersión
plt.scatter(X_pca[:, 0], X_pca[:, 1])
plt.xlabel('Componente Principal-1')
plt.ylabel('Componente Principal-2')
plt.title('PCA en el Conjunto de Datos del Vino')
plt.show()
```

## 6 Conclusión

Se ha aplicado PCA al conjunto de datos del vino, reduciendo su dimensionalidad y facilitando su visualización.