Principal component analysis report

Machine learning and Data mining II

Nguyễn Trọng Minh – 22BI13304

**TABLE OF CONTENT**

1. **Study the dataset**
2. Choose the dataset
3. Features of the dataset
4. Labels
5. Mathematical properties
6. Missing data
7. **PCA**
8. Apply PCA
9. Principle components
10. Visualize
11. **Study the dataset**
12. Choose the datasets

For this lab work, we chose 2 datasets from Kaggle. One is the dataset of 80 Valorant players in pro league, contains data present their performance on the game. And the second dataset is data about weather (mostly rainy and snowy) recorded in some days from 2006 to 2016.



* 1. Valorant players dataset



* 1. Weather dataset

1. Features of the dataset

* For the valorant players dataset, the “Player” and “Team” columns indicate names and a team, these features are qualitative and discrete because they present properties of the data and can only be some fixed value. The next column – “Rounds played” indicates the number of rounds that player has played in 2023. This feature is discrete and quantitative because it can only be integer. The last 2 columns are “KD” and “Rating”, these stats can be any number in their range, so they are quantitative and continuous.



2.1 Features classification for player dataset

* For weather dataset, only the first 2 columns, “Formatted Date” and “Summary” are qualitative and discrete data because they present the time and weather state at that time. The rest of the data set are all quantitative and continuous since they present a value measured at a time.  
  
  1. Features classification for weather data

1. Labels