Principal Component Analysis

* Dimensionality Reduction Algorithm
* Suppose 200 variables can come down to 30 Principal Components
* Model can be trained on 30 PCs
* Reduces the model complexity
* Con:
  + We lose all the information regarding the original variables
  + Model Explainability becomes less
* Steps for finding PCs:

# Calculation of Covariance Matrix

A picture containing text, person

Description automatically generated

* Helps in finding the variance present inside the data

# Understanding Principal Components

A picture containing map

Description automatically generated

# Principal Components Generation

1. Find the covariance matrix
2. Find Eigen Values and Eigen Vectors from the Covariance Matrix
3. Sort the values in Descending order
4. Each Eigen Vector will represent a Principal Component
5. Redraw the scatter plot having Principal Components only

Text

Description automatically generated with medium confidence

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Redrawing Principal Components**

Diagram, text, whiteboard

Description automatically generated