Machine Learning Presentation

A Data Science Bootcamp Capstone Project

by Russell Kua (28 Apr 2023)

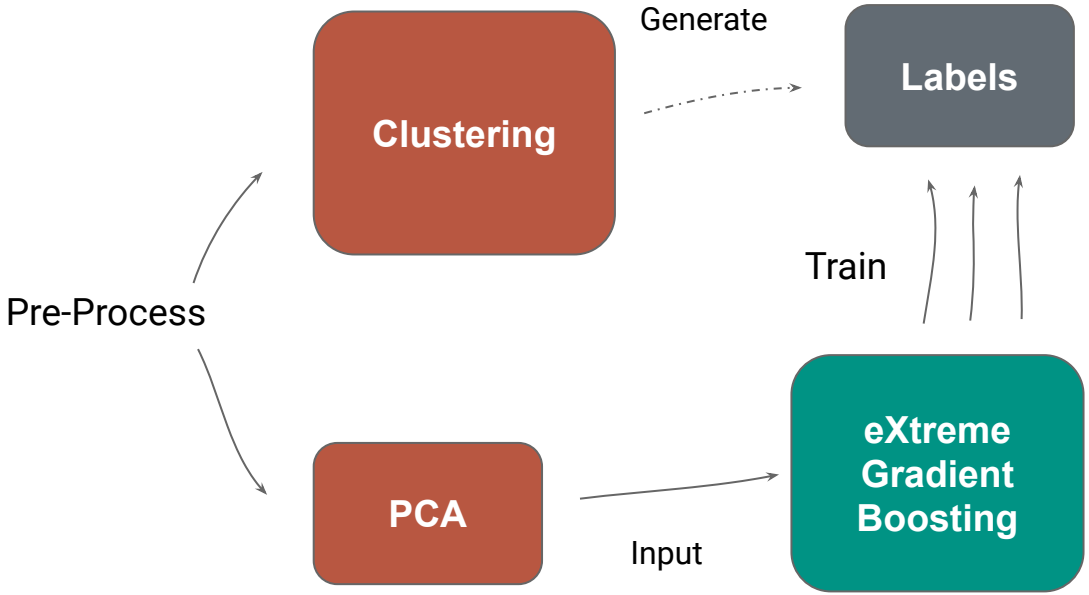
The Task

To predict if caregivers would use caregiver services

Building the Predictive Model

Model

- *Unsupervised Learning*
  - Clustering: Generate labels
  - PCA: Derive principal components as input variables
- *Supervised Learning*
  - XGB: Train and tune the predictive model



Results and Analysis

	NA	*U*
Predicted NA	729 (65%)	76 (33%)
Predicted U	397 (35%)	152 (67%)

- 67% correct on respondents who used services
- 65% correct on respondents who did not use services

This means:

- 33% of respondents who would use the services are missed
- 35% of respondents who would not use the services are wrongly classified

Implication?

Challenges & Future Directions

Challenges

- Matching data with specific use cases
- Treating the data right

Future Directions

- Explore new ideas for capturing data
  - Semi-supervised learning methods
- Combine insights from other research methods
  - Eg., NLP models for FGDs

Thank You