

# Comparing Clustering algorithms on FIFA Dataset

Yang Song (ys638), Zhengxing Xue (zx252), Yue Liu (yl992)

- Goal: we want to compare the performance of different clustering algorithms on FIFA player dataset
- What do you plan to do?
  - We have run K Means on a subset of FIFA players (50 features) and classified players into 4 categories (Goalkeeper, Defender, Midfield, Forward). And we used TSNE to visualize the results (on right).
  - We will implement different clustering algorithms in Python
- What're some recent (or not so recent...) advances in this problem? What papers have you read or do you plan to read?
  - K-means, EM, hierarchical clustering, CURE, optimizing for global (not local) minima
  - Paper 1: <http://infolab.stanford.edu/~ullman/mmds/ch7.pdf>
  - Paper 2: <http://www.mmds.org/mmds/v2.1/ch07-clustering.pdf>
- What challenges do you anticipate? Anything you're not certain about?
  - Measuring performance could be difficult, need to clearly define this
  - FIFA Dataset is high-dimensional, need dimensionality reduction for visualization

