Clustering technique: In order to cluster user reviews, we passed the vector text files containing information about the user's movie preferences, and in another situation information if a user has a seen a movie, into a script that first converts the string into a 2-D numpy array. From there, we compute pair-wise Euclidean distances. Now that we have our distance matrix, we can now move onto clustering the data.

  The algorithm begins with a forest of clusters that have yet to be used in the hierarchy being formed. When two clusters *s* and *t* from this forest are combined into a single cluster *u*, *s* and *t* are removed from the forest, and *u* is added to the forest. When only one cluster remains in the forest, the algorithm stops, and this cluster becomes the root. At each iteration, the algorithm must update the distance matrix to reflect the distance of the newly formed cluster *u* with the remaining clusters in the forest.[1]

---

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html#scipy.cluster.hierarchy.linkage