

Big Data Course Work 2 Proposal

Group AU: (Liam, Manual, Natalia, Johannes, Saima)

Our group has selected the topic:

"Recommender Systems using similarity analysis algorithms at a large scale"

As we understood, for this coursework we need to investigate briefly two most popular similarity metrics (Jaccard's Similarity coefficient and Cosine similarity) and choose one of it. We should implement a project which generates a list of recommendations for user. The algorithm should consider the potential enormous size of the dataset and no external libraries are allowed for implementing similarity algorithms elements.

Dataset: We are going to use recommended MovieLens dataset. The recommendations will be provided according to what user has already rated.