

Netflix Recommendation System

Nisarg Vinchhi - nvinchh

Rohit Mandge - rnmandge

Vaibhav Rajadhyaksha - vrajadh

Claims:

- For the project, we will be using the Netflix dataset which has approximately 200 million records. We will be using collaborative filtering to recommend to the users, top X movies which they haven't watched yet.
- The time taken to process the given dataset for test purposes was inordinately long on a standalone machine. The configuration of the machine was 8GB RAM, Core i7 (64-bit) processor.
- The data structure that we plan to use is nested dictionaries in Python. Since the size of the dataset is large, we plan to use Apache Spark for processing the data in a parallel and distributed manner.
- We plan to run spark on top of yarn for efficient management of clusters.