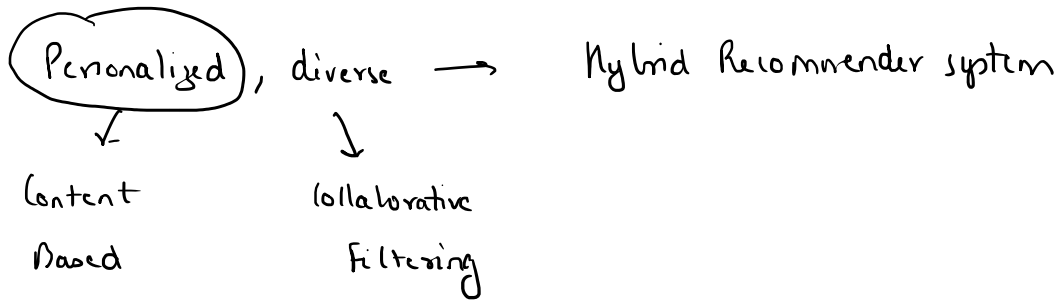


Hybrid Recommender system

User engagement, User retention ↑

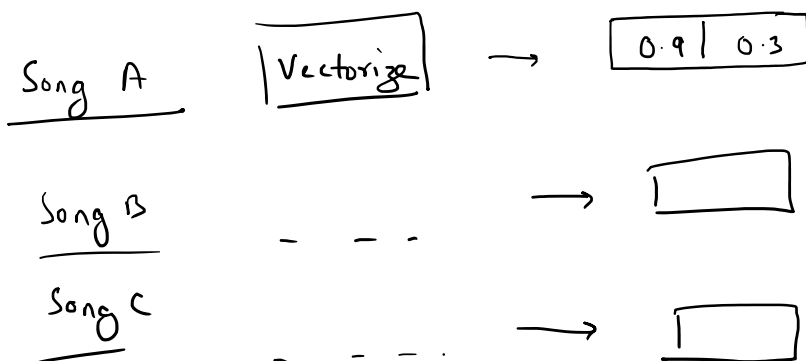
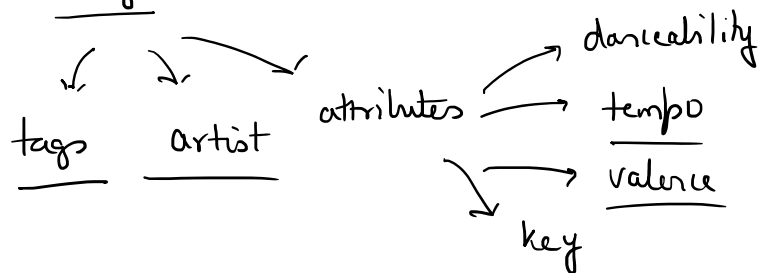


Phase I → Only Content Based

Phase II → Only Collaborative

Phase III → Weighted approach = $\frac{\alpha_1 CB + \alpha_2 CF}{\text{Hybrid Recommender system}}$

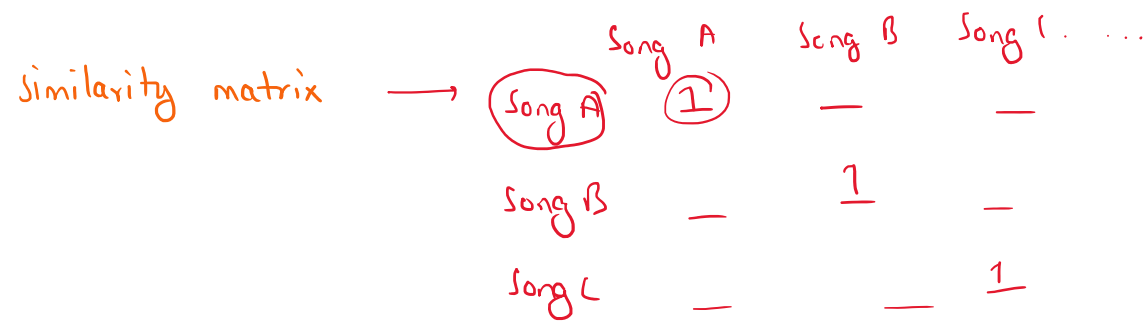
By calculating similarity b/w songs → metadata.



Cosine similarity

$$k(x, y) = \frac{xy^T}{\|x\| \|y\|}$$

x, y -1 to +1



Songs dataset → 50k

genre → 56% values missing → Drop
tags → Kill

Data cleaning → duplicate → 9
missing vals → genre drop
tags → No tags

name, artist, tags → lower case

Content Based → Similar attributes

↖ Name, Id → drop.


Transformations → Vectorize → Matrix
Row → Song
cols →

Similarities → cosine similarity

↓

Recommend → Top similarity scores.

dataset (transformed) , input vector ^T
 (50674×8431) (1×8431)
 (8431×1)



(50674×1) Song A Song A
 Song B —
 Song C —