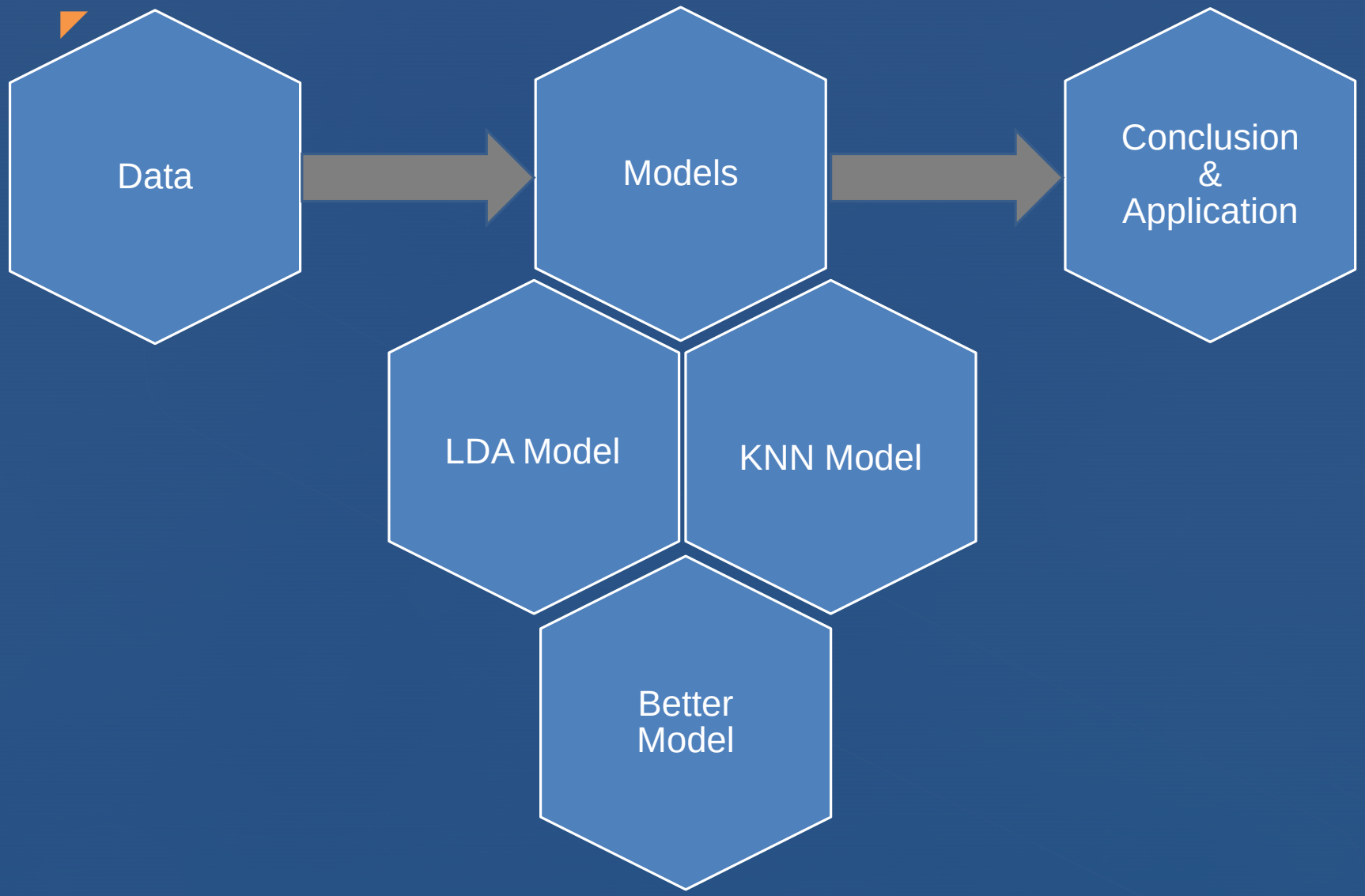
Music Recommendation System

Chris Chen, Lin Du

Purpose

- Similar songs could share common features, like genre, lyrics, and correlated tags labeled by users.
- Our goal is to find a way to classify types of target music for people with different tastes. So we would be able to recommend similar music according to songs they like.


Data
Models
Conclusion
&
Application
LDA Model
KNN Model
Better
Model

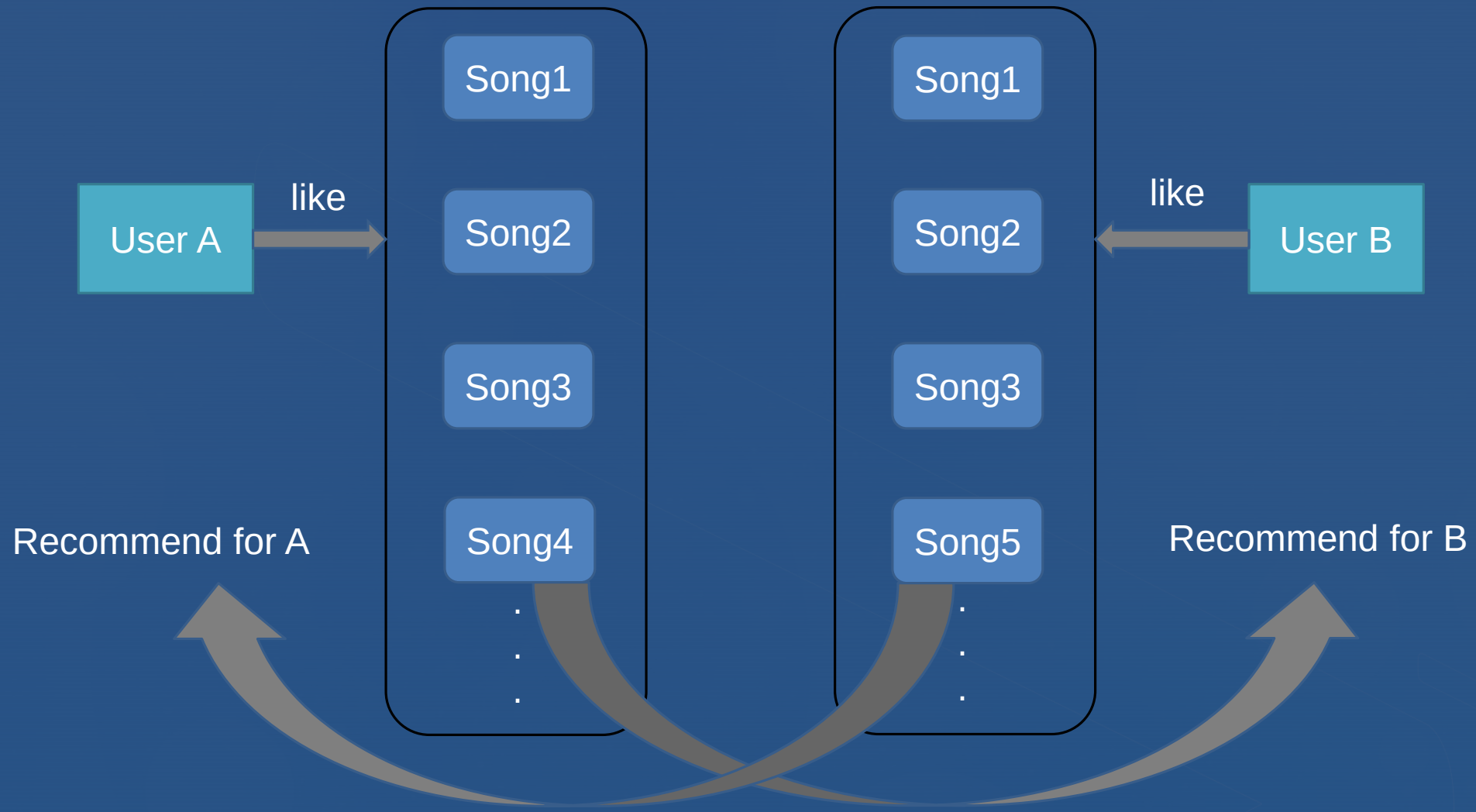
Dataset

- **Last.fm dataset**, the official song tag and song similarity dataset of the Million Song (Datasethttps://labrosa.ee.columbia.edu/millionsong/lastfm)
- List of features:
- Song Info, e.g., singer, title, publisher, year
- Lyrics
- Tags

Model: Latent Diriclet Allocation (LDA)

- Let $X = \{x_1, \dots, x_D\}$ be the set of all the lyrics. x_d is a vector of size N_d where $x_{d,i}$ is the ith word in the dth lyric.
- Let V denotes the size of unique words from all the lyrics.
- Suppose the lyrics are generated from K topics
- Let η_k be a V-dimensional probability distribution on the V words for topic k
- Let θ_d be a K-dimensional probability distribution on the K topics for lyric d
- Let $c_{d,i}$ be the latent variable that picks out the topic that the $x_{d,i}$ belongs to.
- Then the model is defined as the following

$$c_{d,i} \sim Discrete(\theta_d), x_{d,i} \sim Discrete(\eta_{c_{d,i}})$$


Song1
Song2
Song3
Song4
Song1
Song2
Song3
Song5
User A
like
like
User B
Recommend for A
Recommend for B

Model: K-Nearest Neighbors (KNN)

- Use of historical user taste information to calculate the distance between users
- Use the target user's "nearest neighbors (k-Nearest)" Evaluation of weighted evaluation values to predict the target commodity users preferences for specific commodities extent
- The system then can make recommendations based on the preferences of the target user extent.

Model Selection

- Split data into training set and testing set.
- Use cross-validation method to calculate the error rate for each model.
- Choose the model with smaller error rate and apply it to our web app, where our conclusion and application land.

Reference

- Blei, D., Ng, A. & Jordan, M. (2003) Latent Diriclet Allocation. *Journal of Machine Learning Research* 3, 993-1022
- Paisley, J., Wang, C. & Blei, D. (2012) The Discrete Infinite Logistic Normal Distribution. *Bayesian Analysis*, 7(4): 997-1034