Music Recommendation System

MIT Capstone Project
(Applied Data Science)

Sanjay Shukla

Agenda

- Objective
- Solution Method
- Solution Recommendation
- Challenges & Approach
- Conclusion
- Appendix

Objective

Problem Statement

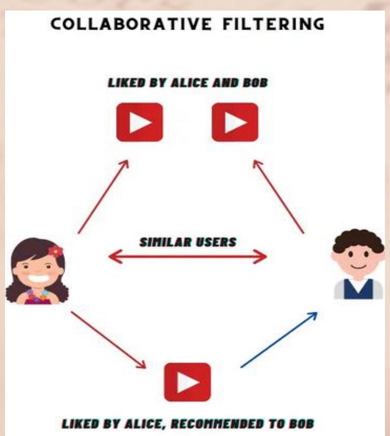
How to keep user engaged on our music platform by determining the most likely songs the users will enjoy next, resulting in revenue increase

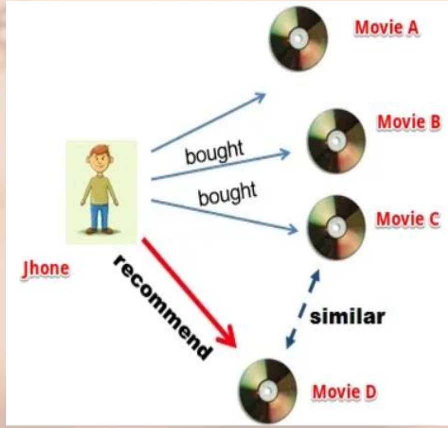
Goal

- Improve customer satisfaction and loyalty by providing personalized recommendations
- Creates a better user experience while also driving incremental revenue

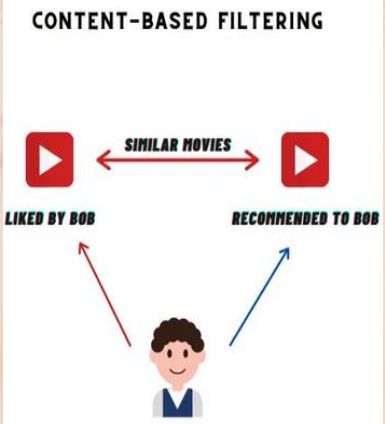
Solution Method

- Build recommendation system is to personalize content and to present the most relevant items to the user.
- For the solution following assumptions were made and other data were ignored
  - Users to have listened at least 90 songs
  - Song to be listened by at least 120 users to be considered
- Following Machine Learning recommendation model were tested, first with baseline and then performing hyperparameter optimization





	Movie1	Movie2
U1		5
U2	1	
U3	1	4
U4		
U5	3	1



Solution Recommendation

- Given the limited dataset in which no real user or song rating is available, using our solution options, we recommend model with user-based collaborative filtering that uses an estimated rating This model has performed well compared to other models with a 0.705 recall score
- This means
  - out of all the time song should have been predicted, 70.5% of the time model predicted it correctly
  - every user will be 70.3% more likely to engage with the next song they are listening to if they pick from our recommendation
- With higher engagement the business can achieve growth target
- On alternate a hybrid solution can be used including the content based model that uses text processing technic along with user-based solution

Challenge and Approach

Observed Challenge

- Selective interaction between users and songs
- The estimated rating is inadequate to capture user’s real affinity to the songs.
- By relying on one data point, the model may not recommend a relevant song to the user resulting loss of user engagement
- The data point we use as user’s rating is indirect or implicit feedback based of play count
- Recently release songs have clear preference indicating probability of younger audience using the music app

Suggested Approach

- Implement current model and perform A/B Test to predict accuracy of solution
- The current model can be improved by gathering additional dimensions like
  - genre of the song
  - user’s preference
  - User’s past behavior (includes duration the user listen to song)
- Encourage user to rate the songs
- Introduce lesser heard songs to younger audience generating interest in different genera of music

Conclusion

Implement Music Recommendation system

Observer Profitability

Enhance User engagement

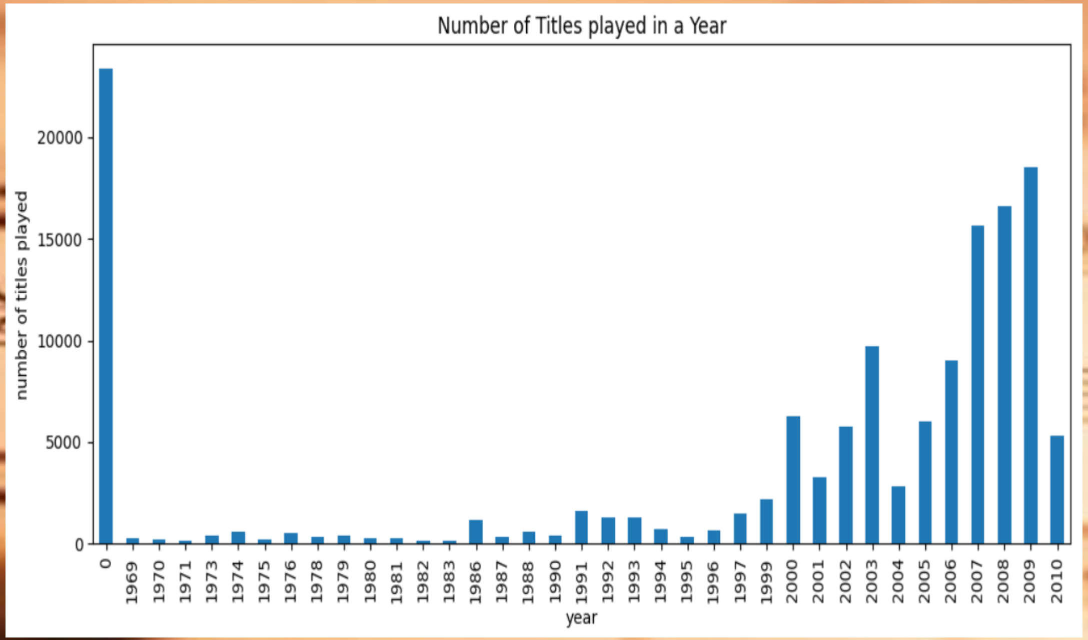
Continuously monitor and enhance model


APPENDIX

Observation : Data

- User interaction with song that they listen, but no user rating is available
- Song play count is used as indirect feedback to show users affinity to the songs indicating if the us song, user listening to, leading higher play count

count	138301.000000
mean	1.698614
std	1.088205
min	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
max	5.000000



Date :

- total of 3,337 unique users, 620 unique songs and 247 unique artists
- peak in plays from the years 2007 to 2009 indicating the chart is right-skewed distribution
- over 23000 transactions has no year information available

Observation : Solution Matrices Comparison

Model Type ->	**User- User**		**Item - Item**		**Matrix Factorization**		**Clustering**	
Matrices	*Baseline*	*Optimized*	*Baseline*	*Optimized*	*Baseline*	*Optimized*	*Baseline*	*Optimized*
RMSE	1.0817	1.0224	1.032	1.0239	1.0026	0.9848	1.0389	1.0335
Precision	0.401	0.443	0.316	0.361	0.432	0.443	0.398	0.393
Recall	0.705	0.654	0.572	0.568	0.654	0.641	0.594	0.575
F_1 score	0.511	0.528	0.407	0.441	0.52	0.524	0.477	0.467

Matrices :

- RMSE: The average deviation in predictions from the actual values. It penalizes large errors. It would be better if we bring this and is better for large values … iction.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

- *Precision* : Out of all the songs predicted, how many times model was correct?

  Precision = TP/Predicted positive = TP/(TP+FP)

- *Recall :* Out of all the times song should have been predicted, the model predicted it correctly.

  Recall for Dog= TP/Actual positive = TP/(TP+FN)

- F1 Score : Harmonic mean (weighted average) of the model’s precision and recall.

  F1 Score = 2* (Precision*Recall)/(Precision+Recall)

- The content-based recommendation model merges key song information, such as title, album, and artist, into a unique text for each song, recommending music based directly on the descriptive characteristics of the songs

Technical Solution Approach