

Music Recommender and Genre Classification System

(CSCI 5502)

Team Overview

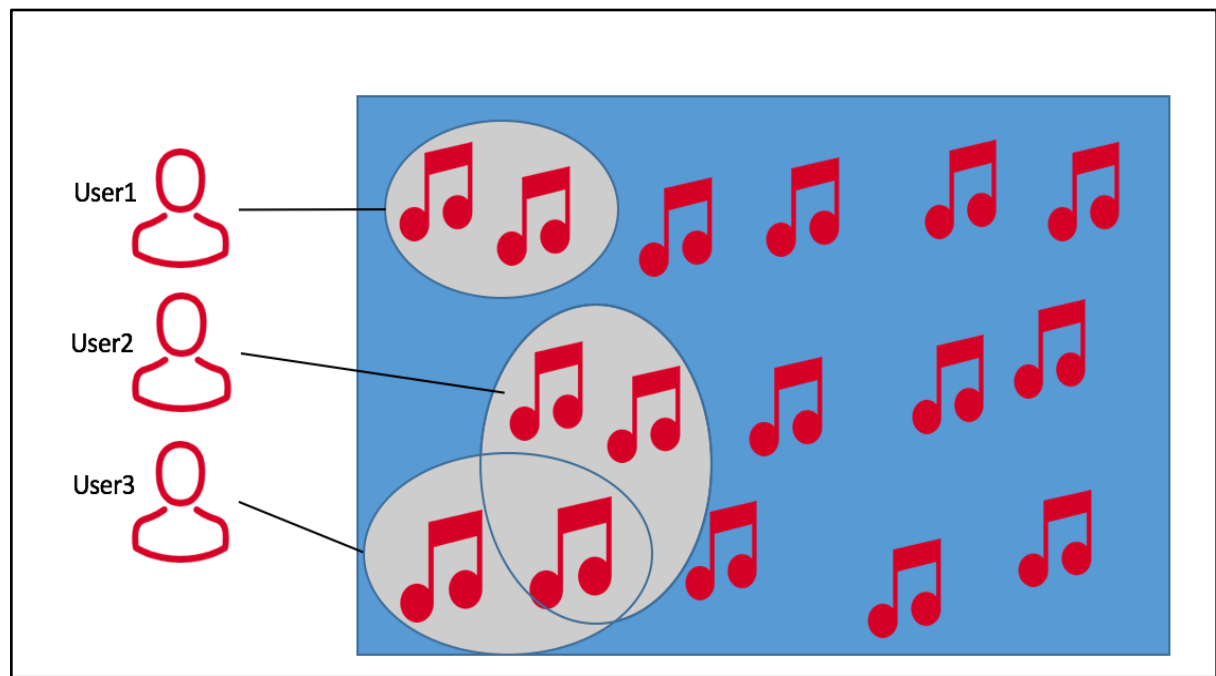
- Suresh Kumar (suku3476)
- Sunil Baliganahalli Narayana Murthy (suba5417)
- Praveen Kumar Devaraj (prde1873)

Problem

How to recommend music to Users ?

Given

- Set of Songs
- Set Users
- User profile



Previous Works

- Recommendation Systems produce recommendation in two ways - through collaborative or content-based filtering.

Collaborative filtering

- Depends on Community of Users
- Based on User Similarity
- “Cold Start” Problem

Content Based filtering

- Based on Song Similarity
- Depends on Song Features

		Songs				
user		1	2	3	4	5
1		✓	✗	✓	?	?
2		✓	✓	?	?	✓
3		✗	?	✓	✗	✗
4		?	✓	?	✗	?
5		✗	?	✓	✓	?

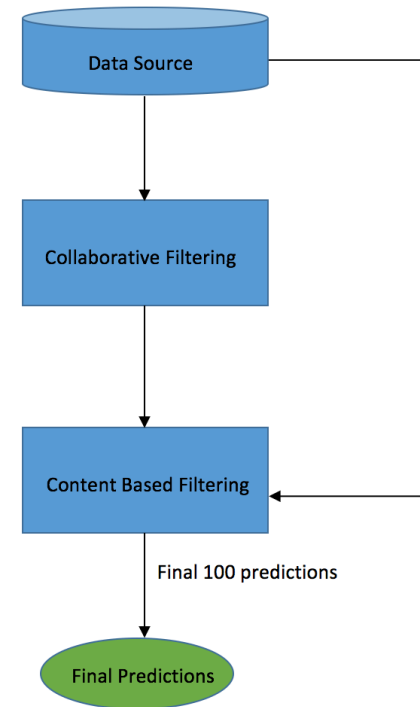
Collaborative

		Songs				
user		1	2	3	4	5
1		✓	✗	✓	?	?
2		✓	✓	✗	?	?
3		✗	✗	✓	?	?
4		✗	✓	✓	?	?
5		✗	✓	✓	?	?

Content Based

Proposed Work

- Combine the advantages of collaborative filter and content-based filter to form a hybrid recommendation system
- Genre Classification to categorize music
- Recommendation based on time of the day and User mood



Hybrid Model for Recommendation

Libraries / Frameworks Used

- Python
- Apache Spark (MLlib + MapReduce)
- Sklearn
- Theano
- Tensorflow google
- Numpy

Completed Tasks

- **Collaborative Filtering**
 - Apache Spark model-based collaborative filter using Alternating Least Squares (ALS) algorithm.
- **Content Based Filtering**
 - Logistic Regression
 - SVM
- **Hybrid Model**
- **Genre classifier:**
 - Constructed feature vector using attributes like “Loudness”, “Tempo”, “Time-Signature”, “Mode”, “Key”, “Average Timbre”, “Duration” and “Variable Timbre”.
 - Logistic Regression
 - Naïve Bayes
 - Random Forest
 - Decision Tree
- **Evaluation Metrics**

Datasets

- Million Song dataset (<http://labrosa.ee.columbia.edu/millionsong/>) - Open Source
- Available Datasets
 - Million Songs dataset by EchoNest
 - SecondHandSongs dataset -> cover songs
 - musiXmatch dataset -> lyrics
 - Last.fm dataset -> song-level tags and similarity
 - **Taste Profile subset -> user data**
 - thisismyjam-to-MSD mapping -> more user data
 - **tagtraum genre annotations -> genre labels**
 - Top MAGD dataset -> more genre labels
- The raw data consists of listening history of a million users in HD5 format.
- Only a subset (1 million entries) of the data is used for testing.
- Full dataset (46 million entries) – AWS (~4hr) Collaborative Filtering
- K-Fold Cross-Validation (4:1)

Performance Evaluation

- Mean Squared Error and Root Mean Squared Error.
- Precision and recall or DCG to assess the quality of the recommendation system.
- K Fold Cross Validation in the ratio of 4:1 (Split the input data into training set and test set)
- RMSE formula :

$$RMSE = \sqrt{\frac{1}{n} \sum_{p,q} |P_{u,v} - R_{u,v}|^2}$$

- Evaluation of Genre Classifier – Predicted genre for songs of known genre and compare the predicted genre with the actual genre of the song.

Genre Classifier

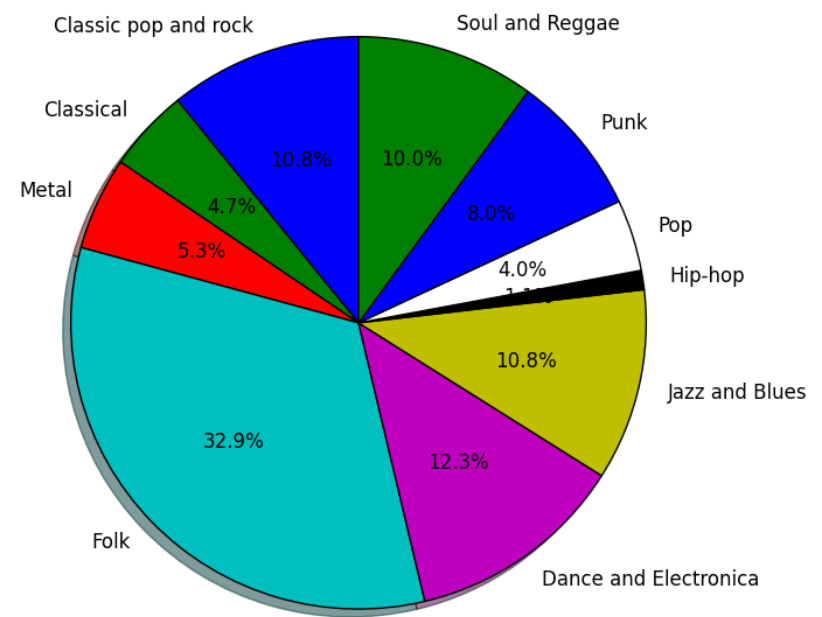
- Automatic Genre Classification – Categorize music based on genre
- Data mining techniques and Machine learning algorithms on existing data.
- Feature vector include
 - Loudness,
 - Tempo
 - Danceability
 - Key
 - Energy
 - Duration
 - Artist Name etc
- Model is constructed using above feature vector to predict genre.
- Modelling techniques –
 - Support Vector Machines (SVM), Nearest-Neighbor (NN) classifiers
 - Gaussian Mixture Models, Linear Discriminant Analysis (LDA), Random Forest

Key Results

- Genre Classifier

Classifier	Accuracy
Logistic Regression	53.76
Decision Tree	47.58
Naïve Bayes	24.52
Random Forest	58.37
SVM	41.73
SVM + Boosting	49.64

Classifier Vs Accuracy

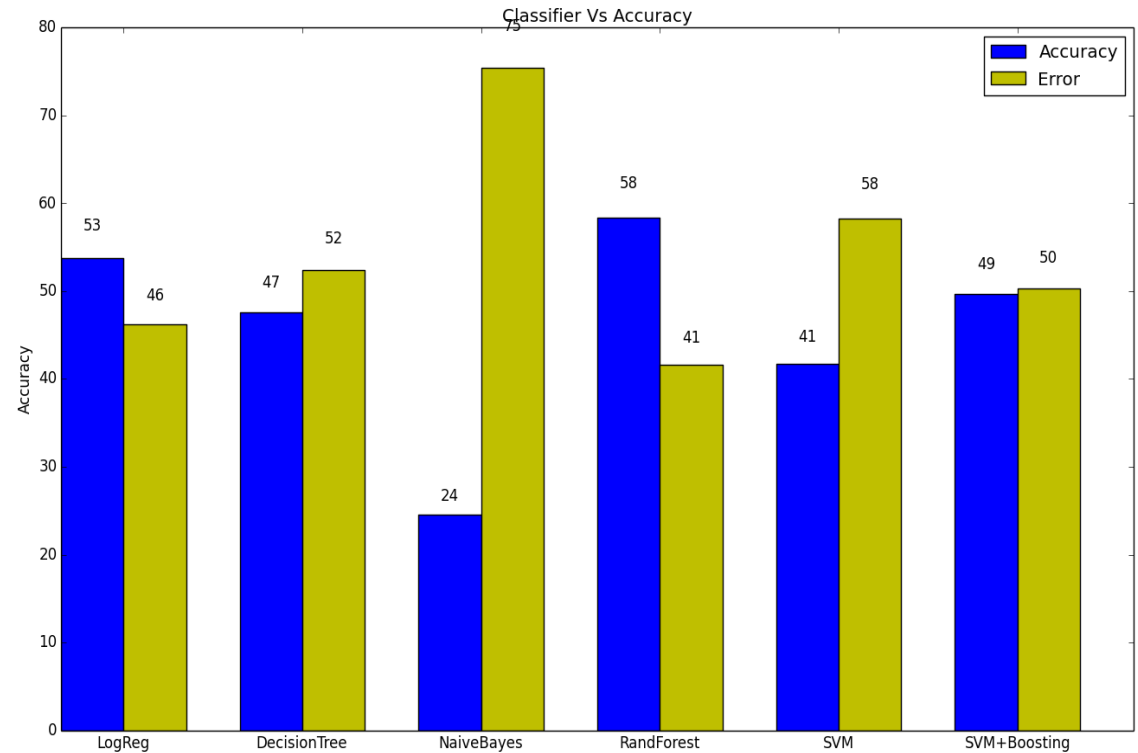


Distribution of Genre

Key Results

Genre Classifier

- Random Forest – Highest Accuracy – Ensemble method

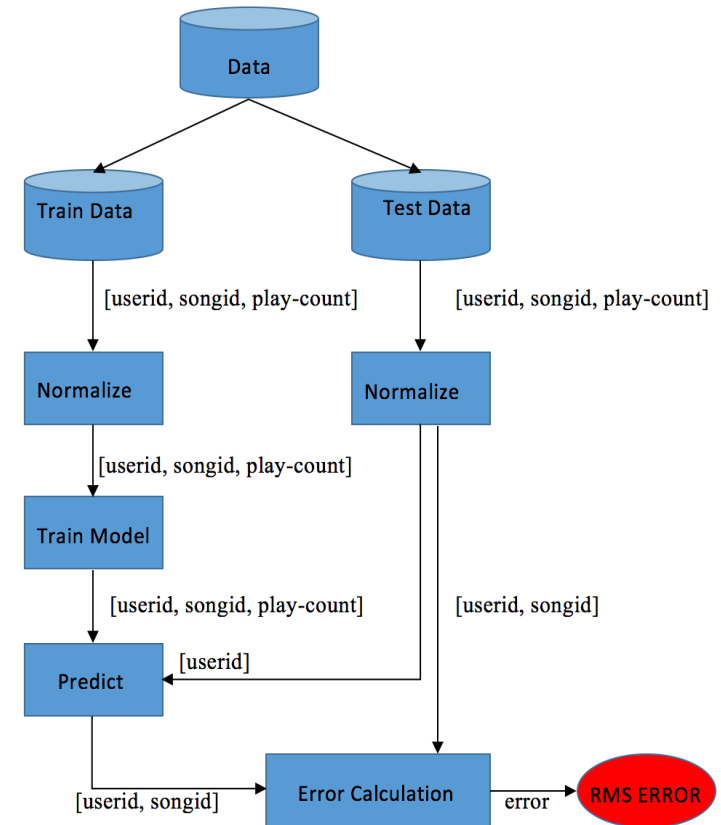


Showing Top 20 rows in Genre Classification using Random Forest

genre	track_id	features	indexedLabel	rawPrediction	probability	prediction	predictedLabel
classic pop and rock	TRAAAGR128F425B14B	[-7.322,123.989,4...	0.0	[0.30337078651685...	[0.15168539325842...	2.0	dance and electro...
classic pop and rock	TRAACPE128F421C1B9	[-11.939,110.189,...	0.0	[0.87855449075979...	[0.43927724537989...	0.0	classic pop and rock
classic pop and rock	TRAACQW128F428854F	[-7.032,47.271,4...	0.0	[0.98946869070208...	[0.49473434535104...	0.0	classic pop and rock
classic pop and rock	TRAAGJV128F1464090	[-9.091,129.122,1...	0.0	[0.90697350069735...	[0.45348675034867...	0.0	classic pop and rock
classic pop and rock	TRAAGNL128F4299BF1	[-11.335,137.29,4...	0.0	[1.09313479623824...	[0.54656739811912...	0.0	classic pop and rock
classic pop and rock	TRAAIAN12903CFF16D	[-11.479,151.911,...	0.0	[0.96922310797627...	[0.48461155398813...	0.0	classic pop and rock
classic pop and rock	TRAAMUY128F4283222	[-7.947,108.94,4...	0.0	[0.57220342053330...	[0.28610171026665...	1.0	folk
classic pop and rock	TRAAOAU12903D0154B	[-15.146,77.27,4...	0.0	[0.88663745892661...	[0.44331872946330...	0.0	classic pop and rock
classic pop and rock	TRAAUSW128F426646E	[-10.173,102.055,...	0.0	[0.84805175267037...	[0.42402587633518...	0.0	classic pop and rock
classic pop and rock	TRAAXRS128F932F05D	[-14.26,121.93,4...	0.0	[1.38208708930594...	[0.69104354465297...	0.0	classic pop and rock
classic pop and rock	TRABEKP128E078C123	[-17.772,116.267,...	0.0	[0.64463445530220...	[0.32231722765110...	1.0	folk
classic pop and rock	TRABJYG128F92EB8DC	[-8.482,120.809,4...	0.0	[1.28955547736335...	[0.64477773868167...	0.0	classic pop and rock
classic pop and rock	TRABLS0128F14A4707	[-13.134,152.434,...	0.0	[0.31194968553459...	[0.15597484276729...	1.0	folk
classic pop and rock	TRABMTM12903D083D2	[-4.942,92.014,1...	0.0	[1.15726233421274...	[0.57863116710637...	0.0	classic pop and rock
classic pop and rock	TRABTFI128F14905F6	[-18.264,151.477,...	0.0	[1.00660429279031...	[0.50330214639515...	0.0	classic pop and rock
classic pop and rock	TRABTYR128F9304934	[-16.172,127.207,...	0.0	[0.93363338820870...	[0.46681669410435...	0.0	classic pop and rock
classic pop and rock	TRABWPT128F42B06CC	[-10.548,89.966,4...	0.0	[0.57273482959268...	[0.28636741479634...	1.0	folk
classic pop and rock	TRABXHU128F147EDE9	[-14.369,157.219,...	0.0	[0.93995765275257...	[0.46997882637628...	0.0	classic pop and rock
classic pop and rock	TRACRBQ128F4263964	[-14.293,111.492,...	0.0	[0.96624803767660...	[0.48312401883830...	0.0	classic pop and rock
classic pop and rock	TRACTIQ128F4288A7C	[-8.45,130.031,4...	0.0	[1.18783945428682...	[0.59391972714341...	0.0	classic pop and rock

Collaborative Filter(CF)

- Memory based CF:
 - Based on user rating to compute the similarity between the users
 - Performance decreases as the data gets sparse
- Model Based CF:
 - Based on machine learning algorithms to find patterns on training data
 - The users and items are a small set of latent factors used to predict missing entries
 - Performance is better in case of sparse data and is scalable
 - Used Alternating Least square algorithm to learn the latent factors
 - Play count of a song is used as rating

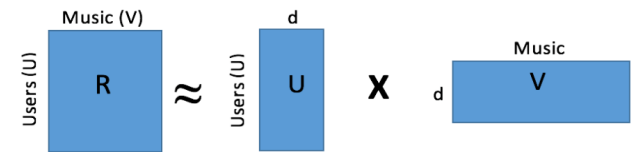


Collaborative Filter(CF)

- Used Apache Spark's Mlib - Model based collaborative – Uses Alternating Least square Algorithm.

Algorithm 1 ALS Algorithm

- 1: Set up vectors y_{Users} , y_{Items} , x_q , and x_p
 - 2: Initialize μ , a_i , b_u , y_{ui}
 - 3: **while** MSE not converged and iteration \neq max limit **do**
 - 4: Create q_i vectors
 - 5: Solve for p_u vectors
 - 6: Create p_u vectors
 - 7: Solve for q_i vectors
 - 8: Rescale each p_u and q_i vector
 - 9: Update μ , a_i , b_u , y_{ui}
 - 10: Find MSE
 - 11: **end while**
-

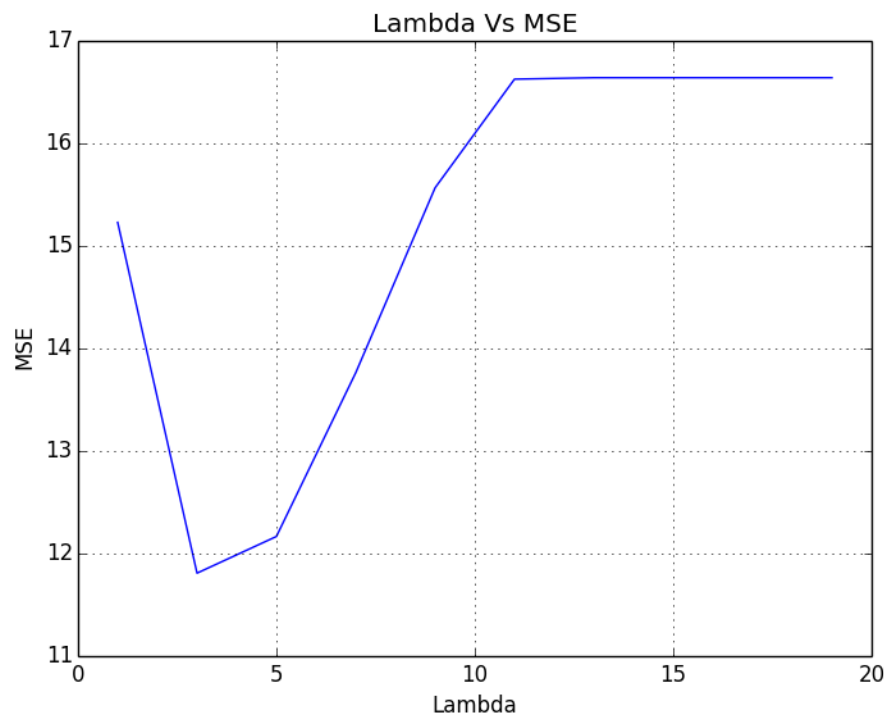


Low Rank Matrix Factorization

Key Results

- Collaborative Filtering Results

Rank = 8 Number of Iterations = 10



Lambda – Regularization Parameter	Mean Squared Error
1	15.23
3	11.80
5	12.16
7	13.76
9	15.56
11	16.62
13	16.64
:	:
20	16.64

Key Results

- User Listening History

song_id	title	play_count	artist_name	duration	year	track_id
SOCHXIG12A58A779A0	Must Be The Ganja	8	Eminem	242.88608	2009	TRFKSWC128F9300987
SOJTDUS12A6D4FBF0E	None Shall Pass (...)	6	Aesop Rock	245.65506	0	TRRJDEP128F422DE06
SOGDDKR12A6701E8FA	My Dad's Gone Crazy	5	Eminem / Hailie Jade	268.59058	2006	TRRELZC128E078ED67
SOAGGLN12AB018AC9B	Empire Ants (Feat...)	5	Gorillaz	283.24527	2010	TRLUMNG12903CB652B
SOHEMBB12A6701E907	Superman	5	Eminem / Dina Rae	350.11874	2003	TROTSBD128E078ED75
SOMHSGR12AB018EAA6	Come Gangsta	4	Tech N9ne	335.35956	2006	TROALNF12903CC5DE7
SOFCGSE12AF72A674F	Starshine	4	Gorillaz	212.32281	2000	TRTSFXC128E0781334
SOVXUKQ12A67021875	Kids With Guns	4	Gorillaz	228.88445	2005	TRLWTAG128E0789037
SOZGDUH12AF72A6DB2	Tomorrow Comes Today	4	Gorillaz	192.91383	2000	TRYUBWR128E07840BA
SOZBHW12AB0187D0B	Rogue Bagel	3	OTT	508.47302	2008	TRIMJUU12903CD43E5
SOEWPY12AF729F7FC	Say What You Say	3	Eminem / Dr. Dre	309.83792	2002	TRKSRLI128E078ED79
SOUVTSM12AC468F6A7	Drop The World	3	Lil Wayne / Eminem	229.58975	0	TREQNRF12903CF2405
SOVXLZU12AF729FEE3	The Rain	3	TECH N9NE feat. A...	259.83954	0	TRQLUIP12903CC5E00
SOEMQPZ12A6D4F7317	Blazing Arrow	3	Blackalicious	159.89507	2002	TRLERKR128F1456BB4
SOPCTBB12AF72A1B64	Hailie's Song	3	Eminem	320.83545	2002	TRBKFKL128E078ED76
SOSENY12AC468B565	Feel So Good (Alb...	2	Mase	203.62404	0	TRQHRXE12903D072F9
SODIJEG12AB018E49F	Somersetttler	2	OTT	451.02975	2001	TRCSVYO12903CD43AD
SORSFFE12AB018B024	The P.A.S.E.O. (T...	2	Tech N9ne	346.46158	2007	TROPYVM12903CC5D96
SOSQAYJ12A6311082E	Man Research (Cla...	2	Gorillaz	272.4567	2000	TRMFYJI128E078132D
SONCPWY12AB01810F5	Insane	2	Eminem	181.34158	2009	TRKNXNF12903CED077
SONCPWY12AB01810F5	Insane	2	Eminem	181.34158	2009	TRRSQZX12903CCAB16
SOSHFFR12A6D4F93DB	Mary Jane	2	Rick James	295.31384	1978	TROMBCU128F425FA4B
SOBHKHS12A58A7E426	Texas	2	The String Cheese...	592.4044	1997	TRSZVWK12903D02A2D
SOSVAMY12A58A7CFFF	Recoup This (LP V...	1	Coolio	80.61342	1995	TRUCPPB128F1477E50
SOQRPBK12AB017C4E9	Rainbow Serpent	1	The String Cheese...	237.66159	2005	TRCJQSQ128F930AB49
SOMNXML12A6D4F8386	One More Time (Sh...	1	Daft Punk	236.38159	2000	TRELQBX128F146DE1F
SOPGABS12AC46867D6	Jellyfish	1	The String Cheese...	850.8599	1997	TRWVWA12903CF6916
SOHYSXA12AB0186704	Mockingbird	1	Eminem	250.8273	2004	TRYNJBB128F931A7EE
SORKKEN12AF72AA5FA	I'm Me	1	Lil Wayne	294.86975	0	TRVXZVA128F425E27
SOLCHMG12AB018422E	Echo (Explicit Al...	1	Gorilla Zoe	237.89668	0	TRGBKEF12903C9E511

Key Results

- Collaborative Filtering Results

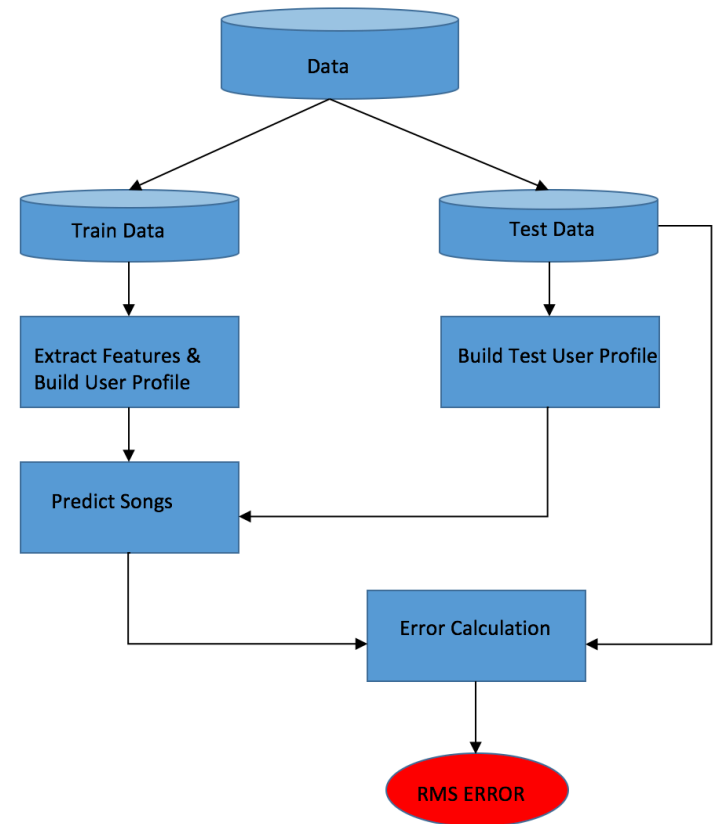
```
----- Top 20 Songs Recommended by Collaborative Filter(ALS) -----
```

track_id	title	artist_name	duration	year	predicted_score
TREQNRF12903CF2405	Drop The World	Lil Wayne / Eminem	229.58975	0	10.0
TRYNJBB128F931A7EE	Mockingbird	Eminem	250.8273	2004	9.642213618791946
TRDFZTB12903CF1CBD	Pursuit Of Happin...	Kid Cudi / MGMT / ...	295.67953	2010	9.251939227477454
TRBZGSM128E078EDB4	Without Me	Eminem	251.55873	2002	8.873327393602377
TRLGMFJ128F4217DBE	Horn Concerto No....	Barry Tuckwell/Ac...	277.15872	0	8.2877383176119
TRLBBLG128F933C0E1	Day 'N' Nite	Kid Cudi Vs Crookers	162.97751	2008	8.15348437384867
TREOIML128E078ED7A	'Till I Collapse	Eminem / Nate Dogg	297.9522	2002	8.031105058696792
TRJRDDL128F147CBE8	Master Of Puppets	Metallica	515.2125	1986	7.92378467515709
TROTSBD128E078ED75	Superman	Eminem / Dina Rae	350.11874	2003	7.915957998362179
TRRGQKP128F1469E20	Représente	Alliance Ethnik	252.21179	1999	7.3639571763352585
TRIDNKO128F92D5902	Terre Promise	O'Rosko Raricim	90.56608	0	7.36384111682264
TRRNFFH128F92D262D	Savior	Rise Against	242.25914	2008	7.1907413015408075
TRZJNDO128EF343498	The Gift	Angels and Airwaves	302.0534	2006	7.0667437813899046
TRLRGVX128E078EC1B	My Name Is	Eminem	268.5122	1999	6.982347582964827
TROVBAL12903CC4F1B	My Name Is	Eminem	269.03464	1999	6.982347582964827
TROYTES128F148C87E	Provider	Days Of The New	353.27954	1999	6.704776846881755
TRKFWPM12903CCAB5C	Forever	Drake / Kanye Wes...	357.66812	0	6.699876345816497
TRABFDT12903CADD73	Up Up & Away	Kid Cudi	227.34322	2009	6.687169853986756
TRQTLTB128F92F785B	Bleed It Out [Liv...	Linkin Park	494.99384	2008	6.288765066302588
TREPEEF128F9311356	Ghosts 'n' Stuff ...	Deadmau5	374.20364	2009	6.265890735080141

only showing top 20 rows

Content-Based Filtering

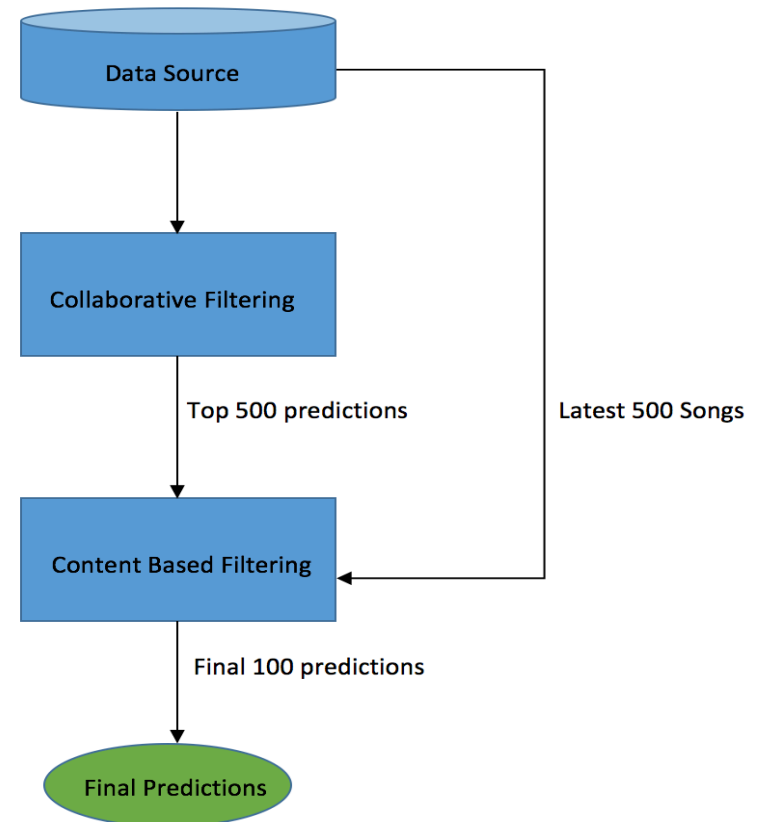
- Depends on the attributes and features of the songs
- Song attributes considered include danceability, tempo, energy, timbre, key, loudness
- User profile is based on a weighted vector of songs. These weights denote importance of each feature.
- Machine learning algorithms like Bayesian Classifiers, cluster analysis, decision trees, artificial neural networks, etc., were used



Hybrid Model = Collaborative Filter + Content-based Filter

Two approaches are pipelined – Output of Collaborative Filtering is fed as input for Content based Filtering

Feed latest songs for content based filtering to recommend the latest songs also – to overcome the disadvantage of Collaborative Filtering.



Key Results

- Content based Filtering Results

track_id	title	artist_name	duration	year	predicted_score
TRMCRHR128E0781333	Latin Simone (Que...	Gorillaz	217.57343	2001	10.0
TRWOGEW12903CB6538	To Binge (Feat. L...	Gorillaz	235.59792	2010	8.390461186733406
TRPAEKT128E07840C5	Slow Country	Gorillaz	215.82323	2000	8.219245435756925
TRFSRXO12903CCEA58	My Mom	Eminem	319.81668	2009	6.939293167412522
TRQZHJK128E07813B9	M1 A1	Gorillaz	230.68689	2001	6.7123689493867245
TRYGCJM128E07813B3	Double Bass	Gorillaz	284.9432	2000	6.347336901181175
TRFKTDC128F4295F8B	P.N.C. (LP Version)	Smif-n-Wessun	325.72037	0	6.336248067676789
TRYQYBV128C719635F	Sky Is Falling	Blackalicious	146.5465	2002	5.034276712040043
TRMTGGR12903CD43E4	The Queen Of All ...	OTT	472.24118	2008	5.034276712040043
TRTQOQJ128E07840BE	Punk	Gorillaz	96.67873	2000	4.935361565109643
TRYZOAQ12903CD43E6	Daisies and Rubies	OTT	618.23956	2008	4.746598003607228
TRMLBXM128F42134D1	Alphabet Aerobics	Blackalicious	133.79874	1999	4.604042389952389
TRDFSHW12903CC5F14	Red Nose	Tech N9ne Collabos	326.45178	0	4.080677640494916
TRUQKVX128E07813B2	Sound Check (Grav...	Gorillaz	279.9277	2001	3.3561844746933622
TRSVHSG128F92C2C0E	Just A Friend	Biz Markie	243.09506	1989	3.353351854996721
TRQYOMR12903CD43AE	Splitting An Atom	OTT	441.83466	2003	3.338326217103519
TRMQOOQ128E07813AB	Re-Hash	Gorillaz	218.40935	2000	3.201142176700428

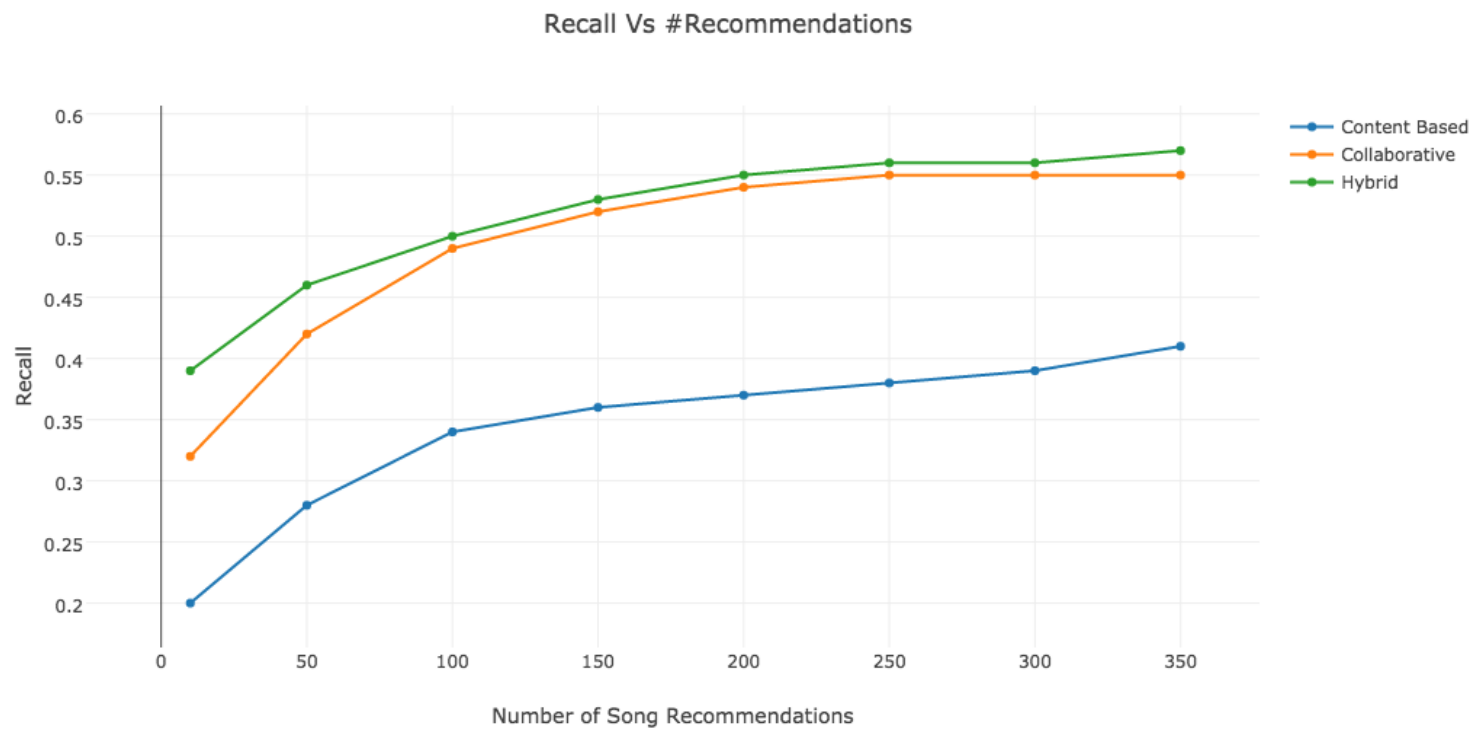
Key Results

- Hybrid Recommendation Model - Predictions

----- Your Final Recommendations are -----

track_id	title	artist_name	duration	year	predicted_score
TREQNRF12903CF2405	Drop The World	Lil Wayne / Eminem	229.58975	0	10.0
TRMCRHR128E0781333	Latin Simone (Que...	Gorillaz	217.57343	2001	10.0
TRYNJBB128F931A7EE	Mockingbird	Eminem	250.8273	2004	9.642213618791946
TRDFZTB12903CF1CBD	Pursuit Of Happin...	Kid Cudi / MGMT /...	295.67953	2010	9.251939227477454
TRBZGSM128E078EDB4	Without Me	Eminem	251.55873	2002	8.873327393602377
TRWOGEW12903CB6538	To Binge (Feat. L...	Gorillaz	235.59792	2010	8.390461186733406
TRLGMFJ128F4217DBE	Horn Concerto No....	Barry Tuckwell/Ac...	277.15872	0	8.2877383176119
TRPAEKT128E07840C5	Slow Country	Gorillaz	215.82323	2000	8.219245435756925
TRLBBLG128F933C0E1	Day 'N' Nite	Kid Cudi Vs Crookers	162.97751	2008	8.15348437384867
TREOIML128E078ED7A	'Till I Collapse	Eminem / Nate Dogg	297.9522	2002	8.031105058696792
TRJRDDL128F147CBE8	Master Of Puppets	Metallica	515.2126	1986	7.92378467515709
TROTSP128E078ED75	Superman	Eminem / Dina Rae	350.11874	2003	7.915957998362179
TRRGQKP128F1469E20	Représente	Alliance Ethnik	252.21179	1999	7.3639571763352585
TRIDNKO128F92D5902	Terre Promise	O'Rosko Raricim	90.56608	0	7.36384111682264
TRRNFFH128F92D262D	Savior	Rise Against	242.25914	2008	7.1907413015408075
TRZJNDO128EF343498	The Gift	Angels and Airwaves	302.0534	2006	7.0667437813899046
TROVBAL12903CC4F1B	My Name Is	Eminem	269.03464	1999	6.982347582964827
TRLRGVX128E078EC1B	My Name Is	Eminem	268.5122	1999	6.982347582964827
TRFSRXO12903CCEA58	My Mom	Eminem	319.81668	2009	6.939293167412522
TRQZHHJK128E07813B9	M1 A1	Gorillaz	230.68689	2001	6.7123689493867245
TROYTES128F148C87E	Provider	Days Of The New	353.27954	1999	6.704776846881755
TRKFWPM12903CCAB5C	Forever	Drake / Kanye Wes...	357.66812	0	6.699876345816497
TRABFDT12903CADD73	Up Up & Away	Kid Cudi	227.34322	2009	6.687169853986756
TRYGCJM128E07813B3	Double Bass	Gorillaz	284.9432	2000	6.347336901181175
TRFKTDC128F4295F8B	P.N.C. (LP Version)	Smif-n-Wessun	325.72037	0	6.336248067676789
TRQILTB128F92F785B	Bleed It Out [Liv...	Linkin Park	494.99384	2008	6.288765066302588
TREPEEF128F9311356	Ghosts 'n' Stuff ...	Deadmau5	374.20364	2009	6.265890735080141
TRCWHIO128F1488FR7	Harder Better Fas...	Daft. Punk	223.60771	2007	6.215374395247457

Results Comparison



Future Scope

- Use Deep learning techniques.
- Use Probabilistic Topic Modelling with Collaborative Filtering
- Recommend songs based on time of the day (Morning, Noon, Evening).
- Recommendation based on mood tags sad, happy, joyful, romantic, etc.
- Integration with Spotify/i-tunes etc