

Multiclass music genre classification

Daniel Scalena 844608
Alessandro Albi 817769

Università degli studi di Milano-Bicocca

Overview dei dati

- Dataset di classificazione di 10 generi musicali
 - Dati prevalentemente numerici
-

Exploratory Data Analysis

- Struttura del dataset
 - Analisi esplorativa
 - Preprocessing del dataset
-

Struttura del dataset

Il dataset è composto da 50005 istanze e 18 attributi.

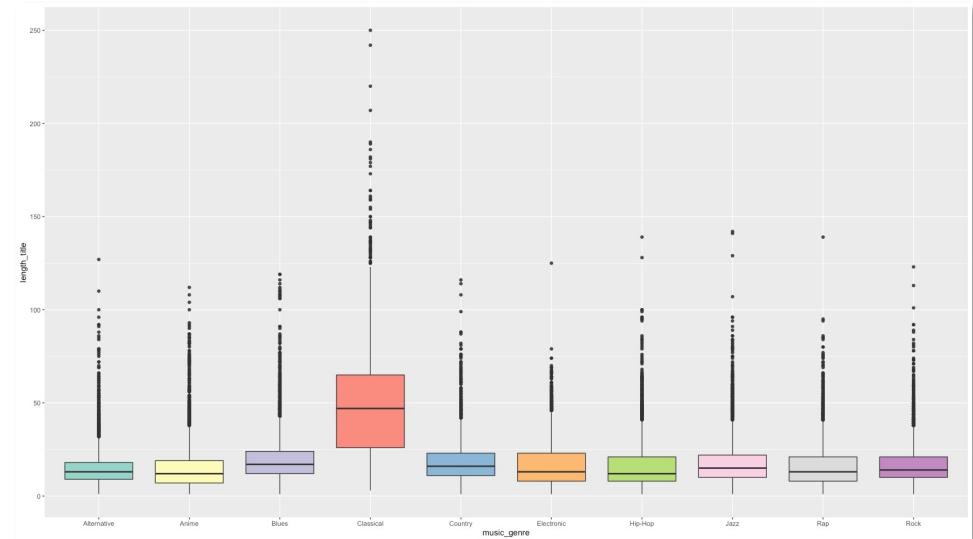
Gli attributi sono per la maggior parte di tipo numerico, ove ciò non sia vero verrà riportato.

- **Instance_id:** identificativo di una traccia
- **Artist_name:** nome dell'artista del brano
- **track_name(str):** il nome della traccia
- **popularity:** popolarità di una specifica traccia
- **Acousticness:** descrive l'acustica di una determinata traccia
- **Danceability:** descrive quanto una traccia sia propensa alla danza o al ballo
- **Duration_ms:** descrive la durata di una traccia in millisecondi
- **Energy:** descrive l'energia di una traccia
- **Instrumentalness:** descrive la strumentalità di un brano
- **Key(str):** stima effettuata con tool automatici
- **Liveness:** descrive la vitalità della traccia
- **Loudness:** misurazione media in decibel riguardo alla registrazione della traccia
- **Speechiness:** misura quanto una traccia sia parlata/cantata
- **Tempo:** usa la nozione di tempo musicale per identificare le tracce
- **Obtained_date(str):** data dell'acquisizione del brano
- **Valence:** determina la valenza di una traccia
- **Music_genre(str):** indica il genere del brano.
10 classi, ognuna il 10% del dataset.

Preprocessing e analisi esplorativa

Len_title:

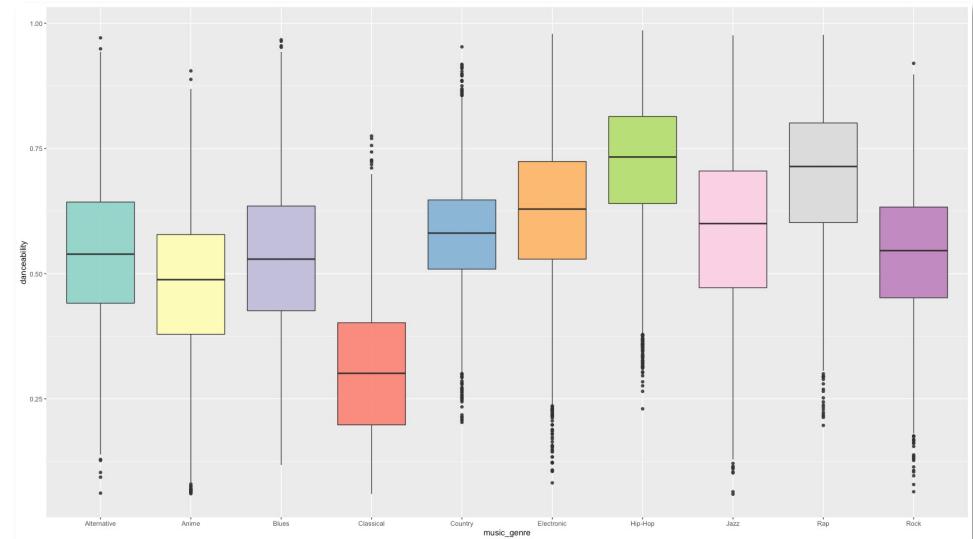
- Aggiunta feature sulla **lunghezza dei titoli** delle tracce



Preprocessing e analisi esplorativa

Danceability:

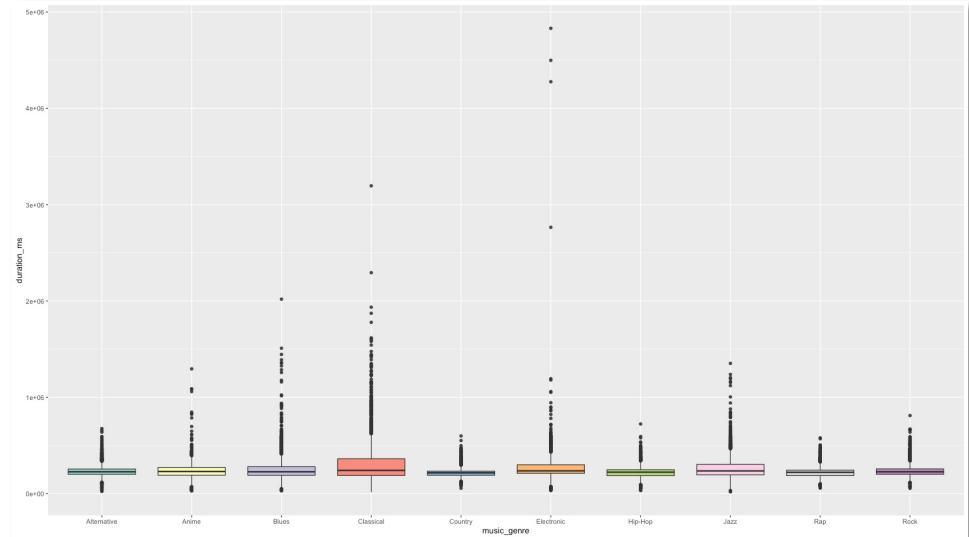
- La classe della **musica classica** riesce a separarsi da tutte le altre classi



Preprocessing e analisi esplorativa

Duration_ms:

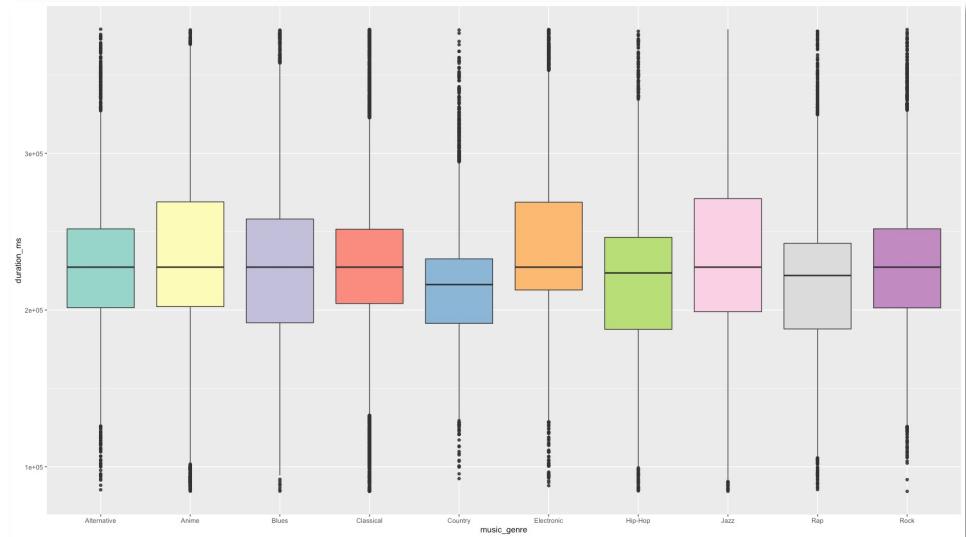
- Forte presenza di outliers
- Rimozione degli outliers con la mediana



Preprocessing e analisi esplorativa

Duration_ms:

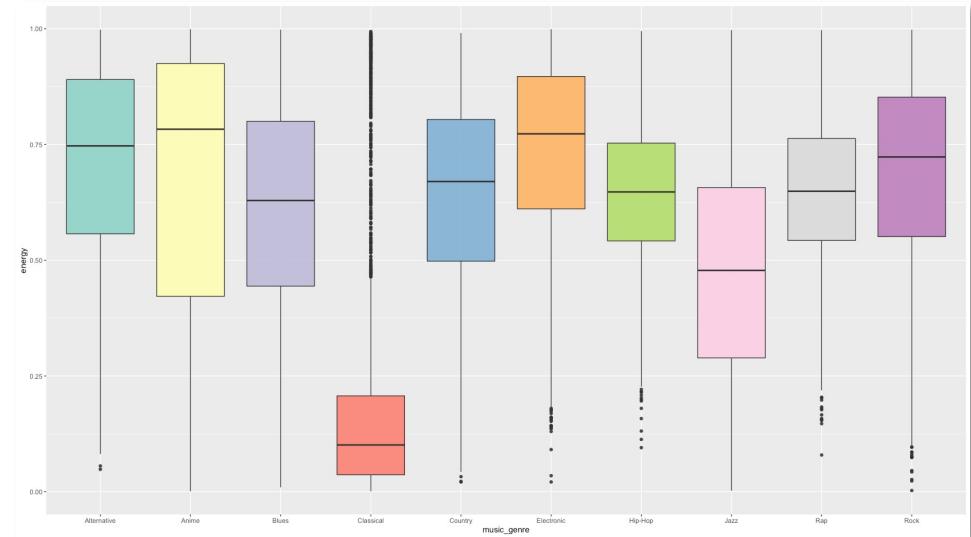
- Forte presenza di outliers
- Rimozione degli outliers con la mediana



Preprocessing e analisi esplorativa

Energy:

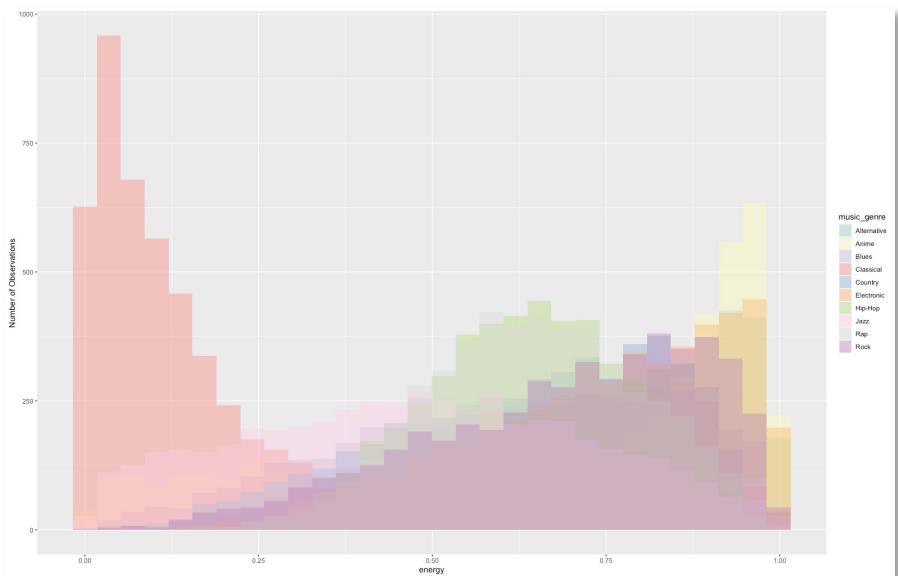
- Nuovamente la classe relativa al genere di musica classica risulta essere quella meglio separata
- Anche nell'istogramma



Preprocessing e analisi esplorativa

Energy:

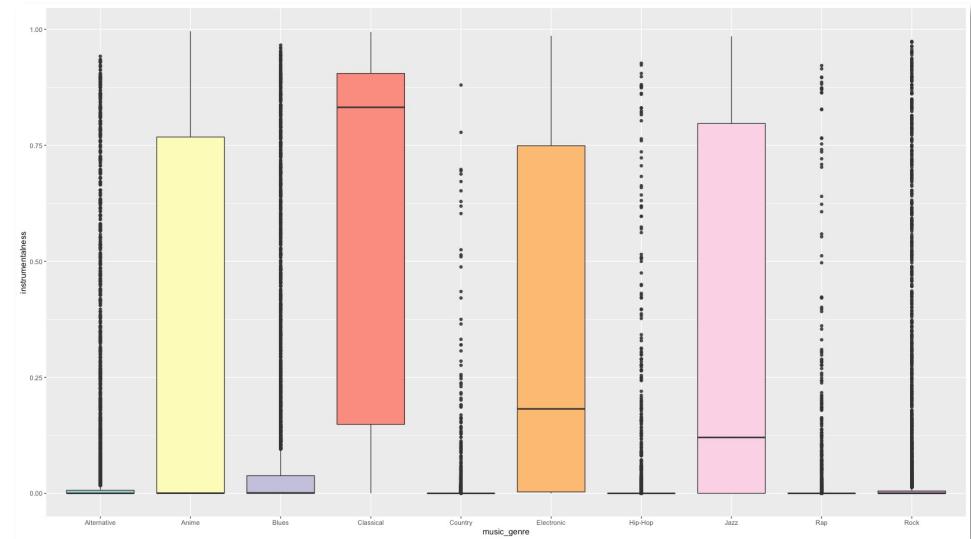
- Nuovamente la classe relativa al genere di musica classica risulta essere quella meglio separata
- Anche nell'istogramma



Preprocessing e analisi esplorativa

Instrumentalness:

- Più del 30% dei valori sono pari a 0
- Non viene considerata ai fini di classificazione



Preprocessing e analisi esplorativa

Pulizia dei dati:

- Rimozione di 5 **valori nulli**
- I **valori nulli sotto il 10%** sono stati sostituiti dalla **mediana** (e.g. tempo)
- I **valori nulli per più del 10%** delle istanze hanno causato la **perdita** totale dell'attributo (e.g. instrumentalness)
- **Rimozione** degli **outliers**

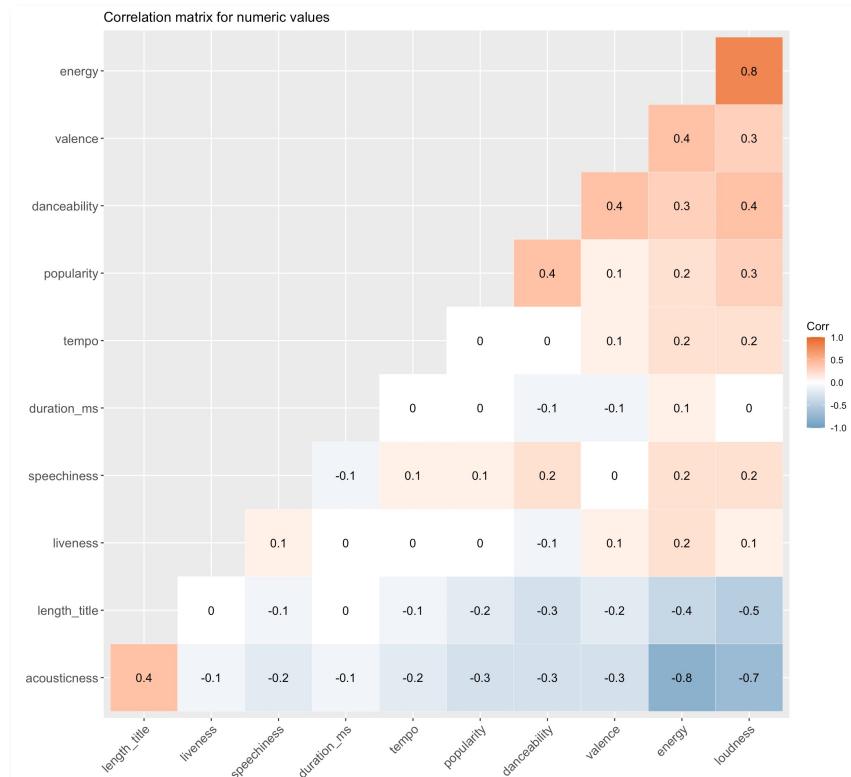
Correlazione e PCA

- Correlazione
 - Analisi bivariate
 - PCA
-

Matrice di correlazione

Dal grafico si evince una forte **correlazione diretta** tra feature *energy* e *loudness*.

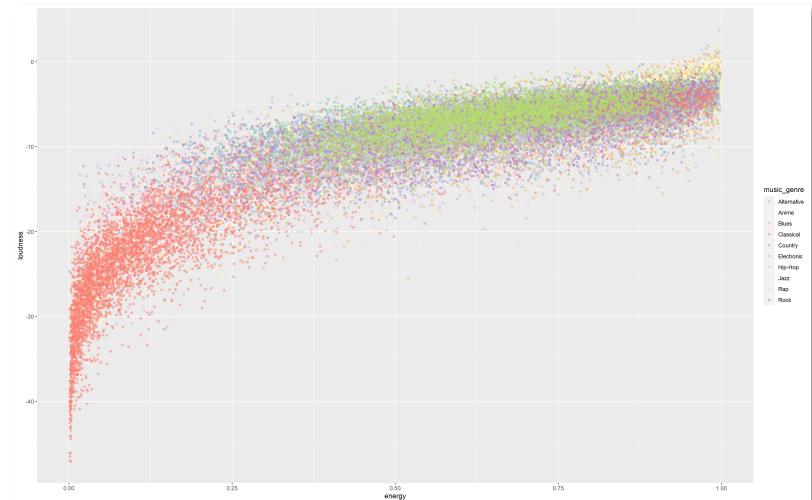
Allo stesso modo si può notare una **correlazione indiretta** tra le feature *acousticness* e *energy* e le feature *qcousticness* e *loudness*



Matrice di correlazione

Dal grafico si evince una forte **correlazione diretta** tra feature *energy* e *loudness*.

Allo stesso modo si può notare una **correlazione indiretta** tra le feature *acousticness* e *energy* e le feature *acousticness* e *loudness*



PCA

Applicando la PCA al nostro insieme di dati, otteniamo la tabella riportata.

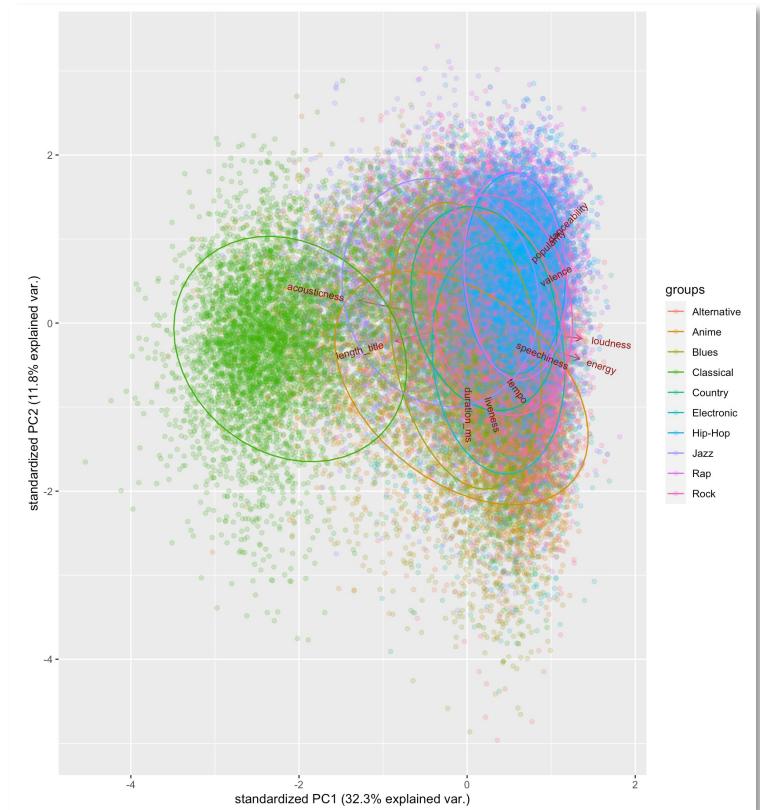
La varianza cumulata, riportata nella terza colonna, mostra come per raggiungere **almeno il 90%** della stessa siano necessarie 8 delle 11 dimensioni

Pred	Eigenvalue	Variance percent	Cumulative variance percent
Dim.1	3.5578039	32.343672	32.34367
Dim.2	1.3023699	11.839726	44.18340
Dim.3	1.0785425	9.804932	53.98833
Dim.4	0.9808640	8.916945	62.90528
Dim.5	0.9535021	8.668201	71.57348
Dim.6	0.8472323	7.702112	79.27559
Dim.7	0.7706874	7.006249	86.28184
Dim.8	0.6548674	5.953340	92.23518
Dim.9	0.4734266	4.303878	96.53906
Dim.10	0.2657594	2.415995	98.95505
Dim.11	0.1149445	1.044950	100.00000

Biplot PC1 e PC2

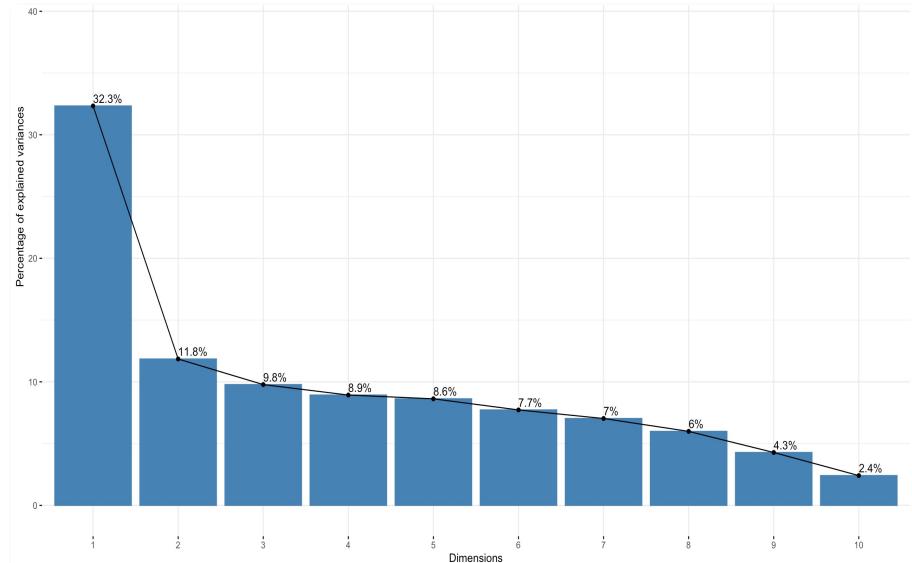
Si nota come le prime **due componenti principali** esprimono **circa il 45%** della variazione totale dei dati.

L'unica informazione utile si può ottenere notando una buona suddivisione della musica classica.



Scree plot

- Necessarie **8 componenti per spiegare il 90% di varianza nel dataset**
- Sono presenti **11 feature iniziali e non risulta conveniente** usare 3 feature in meno eliminando il 10% di varianza
- Viene usato il **dataset originale senza PCA**



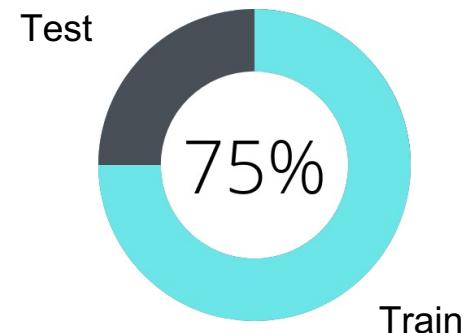
Split del dataset

Split del dataset in train e test:

- 75% Train
- 25% Test

Vengono mantenute le **proporzioni delle classi** sia in train che in test (10% per ogni classe)

I valori sono **scalati** per risultare confrontabili, aiutando l'addestramento



SVM

Neural Network

- Parametri per l'**addestramento**
 - **SVM**: C = 1, kernel = poly;
 - **NN**: Layers: (11, 10, 10, 10).
 - Risultati sul dataset di **test**
-

Risultati: Support Vector Machine

Matrice di confusione

- Accuratezza: 54,0%
- Musica classica ben classificata

pred	Alternative	Anime	Blues	Classical	Country	Electronic	Hip-Hop	Jazz	Rap	Rock
Alternative	490	40	47	20	110	77	125	62	117	220
Anime	7	934	128	56	17	73	1	30	1	4
Blues	20	86	593	35	77	73	1	30	1	4
Classical	4	57	7	1037	1	7	0	42	0	3
Country	236	33	140	7	659	96	41	125	39	152
Electronic	89	74	97	30	56	737	19	141	14	8
Hip-Hop	92	1	4	0	24	50	599	30	515	28
Jazz	80	21	192	59	123	95	12	597	7	30
Rap	33	0	2	0	11	19	382	9	412	57
Rock	200	5	41	4	173	24	71	37	143	747

Metriche sul dataset di test

Stats	Balanced Accuracy	AUROC	precision	recall	F1
Alternative	0.6595	0.61	0.3916867	0.3746177	0.3829621
Anime	0.85921	0.96	0.7466027	0.7466027	0.7466027
Blues	0.71617	0.72	0.4740208	0.5583804	0.5127540
Classical	0.91009	0.88	0.8309295	0.8955095	0.8620116
Country	0.72477	0.87	0.5267786	0.4312827	0.4742713
Electronic	0.77110	0.43	0.5891287	0.5826087	0.5858506
Hip-Hop	0.70634	0.61	0.4788169	0.4460164	0.4618350
Jazz	0.71168	0.49	0.4783654	0.4909539	0.4845779
Rap	0.64227	0.41	0.3301282	0.4454054	0.3791993
Rock	0.76754	0.62	0.5971223	0.5169550	0.5541543

Risultati: Neural Network

Matrice di confusione

- Accuratezza: 55,6%
- Musica classica ben classificata

pred	Alternative	Anime	Blues	Classical	Country	Electronic	Hip-Hop	Jazz	Rap	Rock
Alternative	482	8	11	0	158	56	124	114	26	269
Anime	48	895	103	64	39	62	0	31	1	8
Blues	52	109	610	12	144	57	2	198	1	66
Classical	39	47	20	1061	8	29	0	38	0	7
Country	132	20	60	3	576	49	20	115	9	266
Electronic	122	72	87	6	85	671	32	111	20	45
Hip-Hop	35	1	0	0	13	14	852	21	236	79
Jazz	63	14	156	81	128	120	39	573	7	69
Rap	42	2	2	0	6	12	728	17	274	166
Rock	109	2	1	5	29	11	48	25	56	965

Metriche sul dataset di test

Stats	Balanced Accuracy	AUROC	precision	recall	F1
Alternative	0.68075	0.75	0.4288256	0.3862179	0.4064081
Anime	0.86677	0.94	0.7649573	0.7154277	0.7393639
Blues	0.76249	0.87	0.5809524	0.4876099	0.5302043
Classical	0.92226	0.98	0.8612013	0.8494796	0.8553003
Country	0.71305	0.89	0.4856661	0.4608000	0.4729064
Electronic	0.78497	0.79	0.6207216	0.5363709	0.5754717
Hip-Hop	0.71217	0.82	0.4617886	0.6810552	0.5503876
Jazz	0.70042	0.78	0.4609815	0.4584000	0.4596871
Rap	0.67639	0.61	0.4349206	0.2193755	0.2916445
Rock	0.73517	0.74	0.4974227	0.7713829	0.6048261

Featuring engineering and second Neural Network

- **Estrazione di feature**
aggiuntive viste le **scarse performances**

Featuring engineering

- Performance **sotto le aspettative**
- Vengono riconsiderate **variabili scartate** nella prime fasi
- Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
- Validazione idea
- Creazione **dummy variables**
- Addestramento di una **nuova rete neurale** più profonda

Featuring engineering

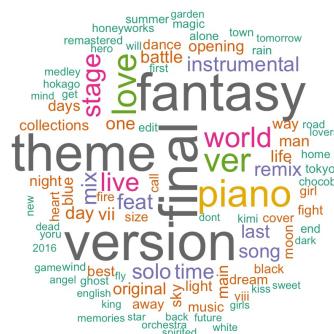
- Performance **sotto le aspettative**
- Vengono riconsiderate **variabili scartate nella prime fasi**
- Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
- Validazione idea
- Creazione **dummy variables**
- Addestramento di una **nuova rete neurale** più profonda

Featuring engineering

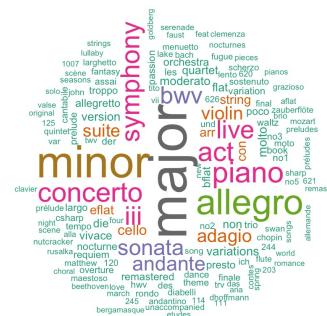
- Performance **sotto le aspettative**
- Vengono riconsiderate **variabili scartate nella prime fasi**
- Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
- Validazione idea
- Creazione **dummy variables**
- Addestramento di una **nuova rete neurale** più profonda

Featuring engineering

- Performance **sotto le aspettative**
 - Vengono riconsiderate **variabili scartate** nella prime fasi
 - Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
 - Validazione idea
 - Creazi
 - Addes



Anime



Classic



Electronic

Featuring engineering

- Performance **sotto le aspettative**
- Vengono riconsiderate **variabili scartate nella prime fasi**
- Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
- Validazione idea
- Creazione **dummy variables**
- Addestramento di una **nuova rete neurale** più profonda

Featuring engineering

- Performance **sotto le aspettative**
- Vengono riconsiderate **variabili scartate nella prime fasi**
- Uso di **Bag of Word** per la feature sul **nome delle tracce musicali**
- Validazione idea
- Creazione **dummy variables**
- Addestramento di una **nuova rete neurale** più profonda

Risultati: Neural Network + BoW

Matrice di confusione

- Accuratezza: 72,6% (+19%)
- Tutti i generi musicali sono meglio classificati

pred	Alternative	Anime	Blues	Classical	Country	Electronic	Hip-Hop	Jazz	Rap	Rock
Alternative	482	8	11	0	158	56	124	114	26	269
Anime	48	895	103	64	39	62	0	31	1	8
Blues	52	109	610	12	144	57	2	198	1	66
Classical	39	47	20	1061	8	29	0	38	0	7
Country	132	20	60	3	576	49	20	115	9	266
Electronic	122	72	87	6	85	671	32	111	20	45
Hip-Hop	35	1	0	0	13	14	852	21	236	79
Jazz	63	14	156	81	128	120	39	573	7	69
Rap	42	2	2	0	6	12	728	17	274	166
Rock	109	2	1	5	29	11	48	25	56	965

Metriche sul dataset di test

- Miglioramenti significativi:
 - Alternative, Blues, Country, Rap, Electronic, Hip Hop e Jazz (7 classi su 10)

Stats	Balanced Accuracy	AUROC	precision	recall	F1
Alternative	0.68075	0.75	0.4288256	0.3862179	0.4064081
Anime	0.86677	0.94	0.7649573	0.7154277	0.7393639
Blues	0.76249	0.87	0.5809524	0.4876099	0.5302043
Classical	0.92226	0.98	0.8612013	0.8494796	0.8553003
Country	0.71305	0.89	0.4856661	0.4608000	0.4729064
Electronic	0.78497	0.79	0.6207216	0.5363709	0.5754717
Hip-Hop	0.71217	0.82	0.4617886	0.6810552	0.5503876
Jazz	0.70042	0.78	0.4609815	0.4584000	0.4596871
Rap	0.67639	0.61	0.4349206	0.2193755	0.2916445
Rock	0.73517	0.74	0.4974227	0.7713829	0.6048261

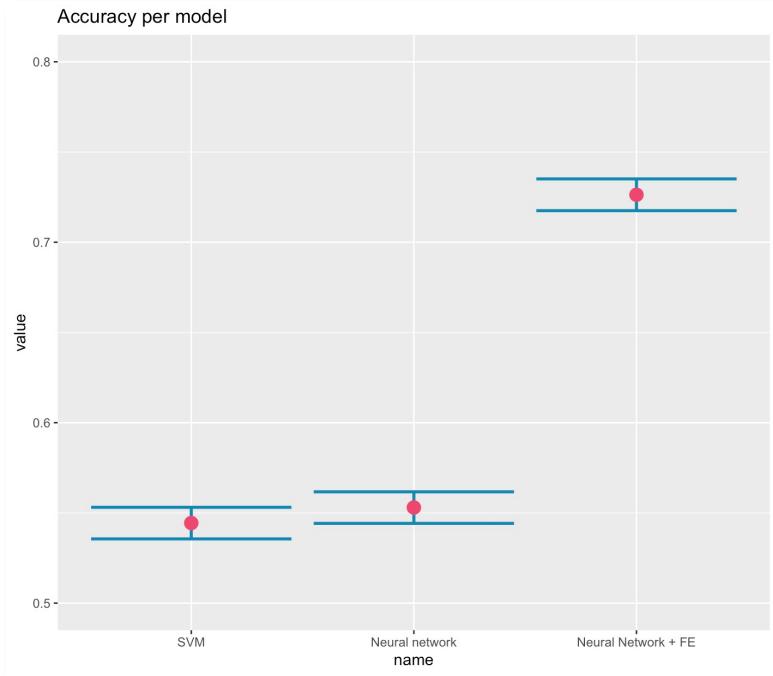
Considerazioni finali sui modelli

- Confronto tra modelli:
 - Timings
 - Performances

Risultati sperimentali

Accuratezza **complessiva**:

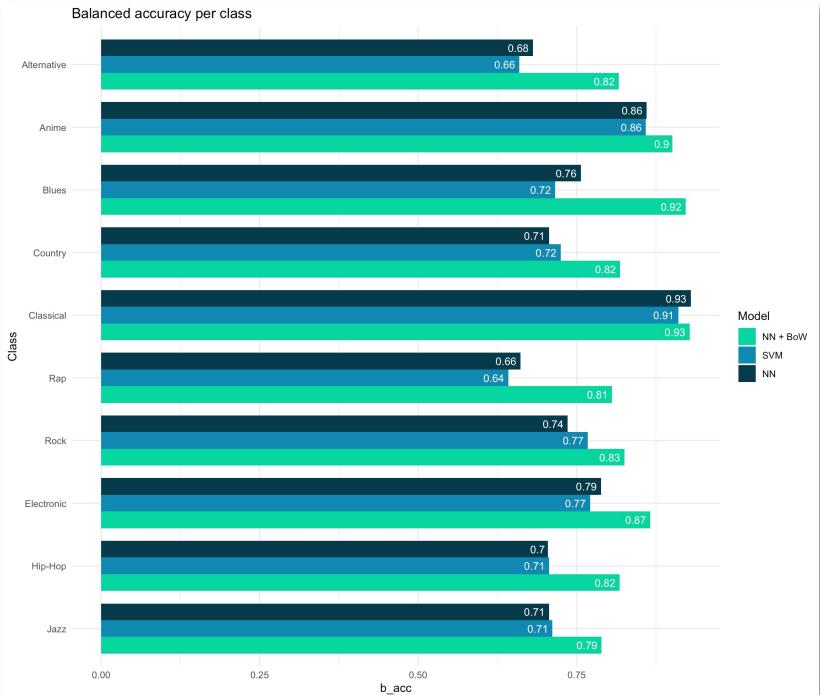
- SVM: 54,0%
- Neural network: 55,6%
- Neural network + BoW: **72,6%**



Risultati sperimentali

Balanced accuracy:

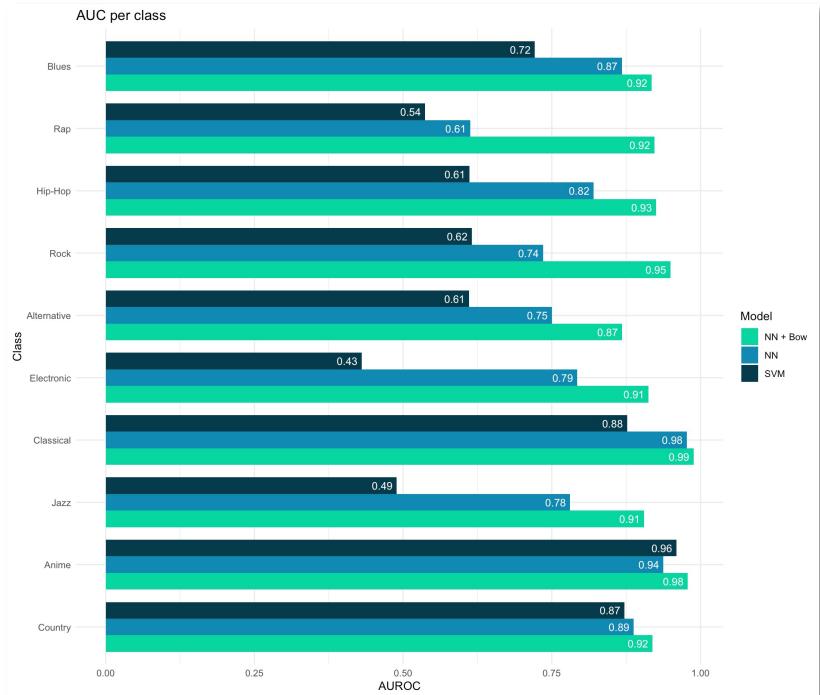
- SVM: 74,7%
- Neural network: 75,4%
- Neural network + BoW: **85,1 %**



Risultati sperimentali

Area under the ROCurve:

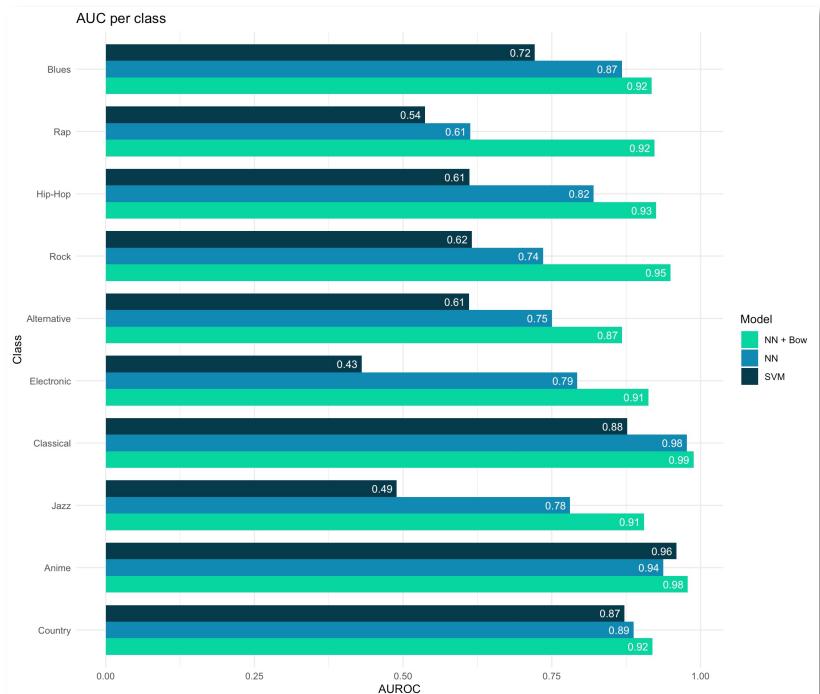
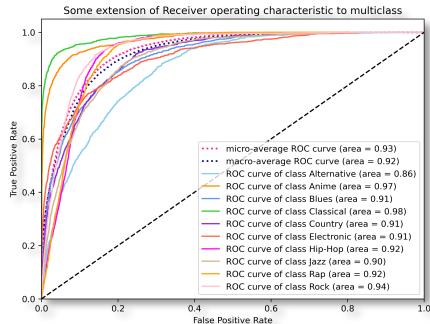
- SVM: 67,3%
- Neural network: 81,7%
- Neural network + BoW: **93,0%**



Risultati sperimentali

Area under the ROCurve:

- SVM: 67,3%
- Neural network: 81,7%
- Neural network + BoW: 93,0%



Conclusioni e sviluppi futuri

Pro NN + BoW:

- **Performance** migliori
- Miglior **separazione delle classi**

Contro NN + BoW:

- Tempo di **addestramento molto lungo** (Risolvibile con una GPU)

Conclusioni e sviluppi futuri

Pro NN + BoW:

- **Performance** migliori
- Miglior **separazione delle classi**

Contro NN + BoW:

- Tempo di **addestramento molto lungo** (Risolvibile con una GPU)

10 fold cross validation

- Tecnica di validazione, split in 10 val. set per la fase di training
-

Risultati k = 10 fold validation

Confusion matrix

- Somma di tutti i fold sul val. Set
 - Totale train set: 37 500 istanze

pred	Alternative	Anime	Blues	Classical	Country	Electronic	HipHop	Jazz	Rap	Rock
Alternative	2634	122	132	87	57	210	234	111	201	165
Anime	159	3282	129	9	66	129	42	174	30	6
Blues	99	69	2517	12	36	0	63	63	3	0
Classical	36	72	72	3567	21	27	51	291	0	3
Country	96	69	132	6	2796	201	120	186	291	303
Electronic	189	96	189	24	93	2865	6	117	63	135
HipHop	162	0	36	0	36	63	2181	147	491	84
Jazz	135	39	426	42	267	195	75	2316	60	174
Rap	69	0	21	0	63	36	752	165	2331	132
Rock	165	3	99	0	315	27	228	180	276	2751

Risultati k = 10 fold validation

Metriche

- Acc. media su CM val. set:
 - SVM: 54,3% (+/- 3,1%)
 - Neural network: 54,8% (+/- 3,6%)
 - **Neural network + BoW:** **70,1%** (+/- 2,6%)



Grazie per l'attenzione!