

# CLASSIFICATORE DI GENERI MUSICALI

---



**Machine Learning**

**A.S. 2020-2021**

**Renzo Arturo Alva Principe**

Matricola: 746799

## Abstract

Le espressioni artistiche e in particolare la musica, sono un linguaggio frutto della creatività espressa basandosi su accorgimenti tecnici allo scopo di trasmettere emozioni o messaggi.

Sebbene nel corso della storia una disciplina artistica si sviluppi in diverse direzioni le motivazioni di questi cambiamenti rimangono per di più legate alla natura umana. Per questo motivo l'arte nonostante poggi su delle basi tecniche e stilistiche è in continua evoluzione ed è difficile da descrivere in termini assoluti: usiamo l'arte per esprimere ciò che non possiamo esprimere con le parole allora come si può spiegare l'arte in termini assoluti?

La musica è forse la forma d'arte che nell'era attuale è consumata in scala maggiore. In particolare i servizi di streaming forniscono accesso immediato a milioni di brani musicali. Uno dei problemi principali dei servizi di streaming è la raccomandazione di brani agli utenti che siano potenzialmente graditi. Il noto servizio di streaming Spotify conta con più di 5000 generi musicali diversi per cui riuscire ad assegnare ad ogni brano un genere musical specifico può aiutare nel determinare una similarità tra brani dello stesso genere facilitandone la raccomandazione.

Questo progetto ha come obiettivo lo sviluppo di un modello di classificazione di brani per macro-genere musicale usando come uniche features quelle fornite da Spotify e limitate a fattori strettamente legati alla multimedialità del brano e non a quelle legate al concetto di consumo, per cui features derivanti dall'interazione con utenti come la popolarità non sono state prese in considerazione allo scopo di rendere la classificazione indipendente dalla piattaforma di streaming.

## Dominio applicativo

Che cos'è un genere musicale?

*“Un genere musicale è una categoria convenzionale che identifica e classifica i brani e le composizioni in base a criteri di affinità. Le musiche possono essere raggruppate in base alle loro convenzioni formali e stilistiche, alla tradizione in cui si inseriscono, allo spirito dei loro temi, alla loro destinazione o, se presente, al loro testo.”* [[Wikipedia](#)]

Come Wikipedia afferma la musica è un'espressione artistica che può essere classificata seguendo diversi parametri di tipo tecnico-musicale, culturale e circa le tematiche affrontate. La caratterizzazione di un brano può in alcuni casi prendere in considerazione aspetti non propri del brano ma dell'artista stesso, un esempio potrebbe essere i vestuari indossati in concerto o nei video musicali.

La musica inoltre è un fenomeno culturale e dinamico che non ha delle vere e proprie regole. È vero infatti che la teoria musicale vuole formalizzare, spiegare e fornire una matrice comune alle composizioni al fine di poter facilitarne lo studio e la rappresentazione in un linguaggio comune, tuttavia gli artisti per definizione sono liberi e possono non seguire le regole. Ad esempio nel rinascimento era tendenza che nelle composizioni sia la durata delle note che la distanza tra esse rispettassero delle precise proporzioni matematiche dando poco valore invece a composizioni più libere. Nei giorni nostri possiamo vedere come le “regole” della musica vengano oltrepassate come nel caso del rapper Tedua il quale nei suoi brani segue un tempo diverso da quello della base. Inoltre la popolarità di uno stile può determinare una tendenza nel corso della storia che a posteriori può essere classificato come un genere musicale.

Possiamo riassumere la complessità di assegnare un genere specifico a un brano prendendo in considerazioni i seguenti aspetti:

- **evoluzione:** un genere musicale dalla sua nascita alla sua morte cambia. Se il genere cambia troppo oppure segue linee diverse può dare origine ad altri generi.
- **generi morti/revival:** i generi musicali possono morire a causa di un calo di popolarità o perchè il fattore sociale che li ha dati alla luce cessa di esistere. Tuttavia esistono dei generi musicali chiamati “revival” che riprendono vecchie sonorità di un genere non più in voga ma in chiave moderna. È il caso del punk e il punk-revival.
- **contaminazione:** la contaminazione è la presenza di sonorità di altri generi in modo più incisivo rispetto all'evoluzione. L'evoluzione muta in modo graduale e naturale. Negli anni 60' i gruppi rock-psichedelico Cream e Santana sebbene entrambi avessero un

repertorio rock-psichedelico è possibile notare in Santana chiare contaminazioni di musica latina e jazz. Inoltre, nel tempo una contaminazione può affermarsi e dare luogo a un nuovo genere.

- **generi e sottogeneri:** il sottogenere di un genere ne eredita le caratteristiche. Il sottogenere ha però peculiarità che il supergenere non ha, ma non è mai chiaro in che misura. E' il caso del thrash metal e il technical thrash metal. Il secondo è caratterizzato da ritmiche più complesse, tempo più veloce e assoli più articolati. Tuttavia non c'è una quantificazione esatta di queste caratteristiche.
- **testi e tematiche:** due canzoni stilisticamente molto simili possono appartenere a generi diversi solo per il fatto di toccare tematiche diverse nei testi come è il caso della salsa, la salsa romantica e salsa intelectuàl oppure l'horror-punk che rispetto al punk più classico racconta nei suoi testi di storie horror o ancora l'epic metal che racconta di storie fantasy.
- **filosofie:** e' il caso della musica indie in contrapposizione alla musica pop. L'indie non è un vero e proprio genere ma una filosofia di produzione musicale. Gli artisti indie storicamente venivano supportati da case discografiche minori discostandosi da quelle maggiori per via dei vincoli stilistici che queste ultime imponevano allo scopo di abbracciare un pubblico più ampio. Nel tempo però l'indie ha perso le sue ragioni filosofiche avvicinandosi di più al concetto di genere musicale intenso come sonorità. Come la musica pop, l'indie rimane un genere sterminato con moltissimi sottogeneri molto eterogenei tra di loro.
- **provenienza:** la provenienza di un artista può influenzarne lo stile, è il caso di gruppi come i francesi Daft Punk che rispetto ai contemporanei statunitensi The Chemical Brothers avevano sonorità più morbide e insieme a molti gruppi francesi degli anni '90 che hanno adottato questo stile hanno dato luogo alla french-house.
- **strumenti e tecniche:** due brani apparentemente simili possono appartenere a generi diversi solo per il fatto di usare diversi strumenti (tradizionali o non convenzionali), effetti, oppure il modo stesso di suonare uno strumento predileggendo determinate tecniche.

Per le ragioni sopraelencate, i generi musicali sono liquidi e vaghi per cui è molto difficile definire un modo efficace e sistematico per classificare i brani musicali

## Dataset

Il dataset preso in considerazione è fornito da [Kaggle](#) e appartiene al dominio musicale, più precisamente contiene dati su 131.562 brani musicali ottenuti interrogando le API di Spotify e selezionando 20 features di cui quelle rilevanti per questo lavoro sono 14.

Features numeriche:

- Danceability**: quanto un brano è ballabile basato sul tempo, satabilità del ritmo, e forza dei beat.
- Energy**: [0,1] intensità percepita. Brani energici sono veloci e rumorosi
- Loudness**: rumorosità media di un brano espressa in decibel
- Speechness**: [0,1] presenza di parole in un brano
- Acousticness**: [0,1] si avvicina a 1 se il brano è acustico altrimenti si avvicina a 0
- Instrumentalness**: [0,1] si avvicina a 1 se un brano non contiene parti vocali
- Liveness**: [0,1] presenza di un pubblico nella registrazione di un brano
- Valence**: [0,1] positività di un brano. Brani con valenza alta suonano più allegri ed eufoici mentre brani con bassa valence suonano tristi, arrabbiati e negativi
- Tempo**: tempo medio stimato per un brano espresso in beats per minuto (BPM)
- Duration**: durata del brano in ms.

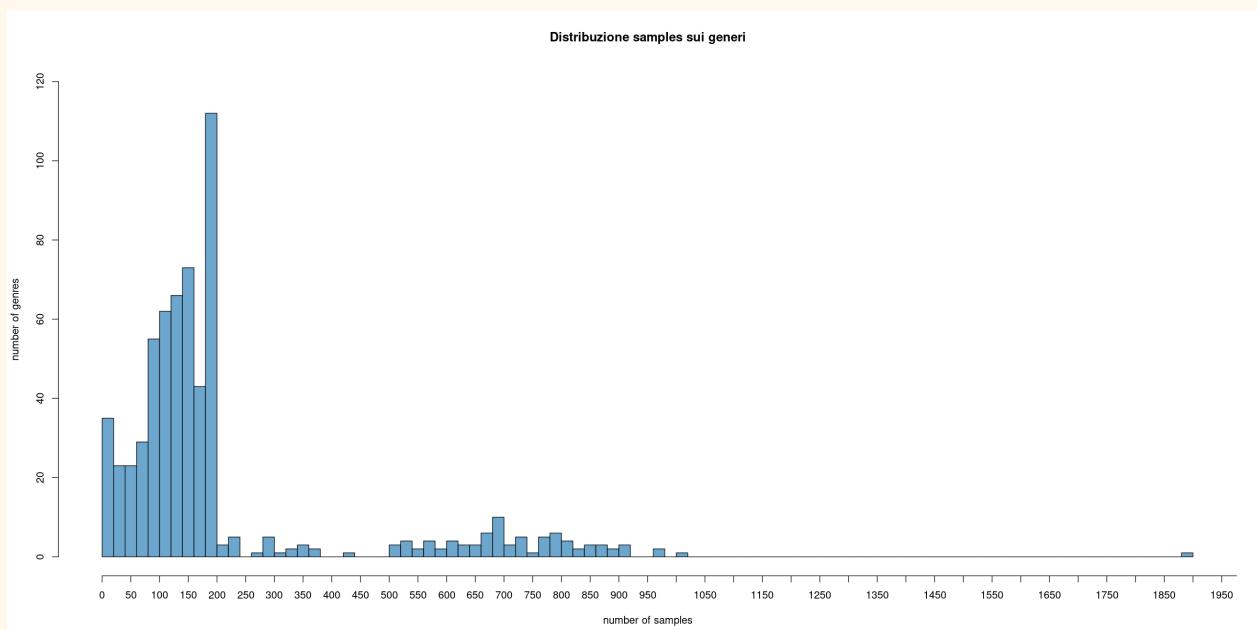
Features categoriche:

- Key**: tonalità del brano
- Mode**: modalità (maggiore o minore) del brano.
- Time\_signature**: indicazione di tempo del brano (4/4, 3/4, ecc)
- Genre**: (target) genere musicale del brano

E' interessante che molte delle features numeriche presentino valori relativi all'analisi sonora del brano ma anche ad analisi più complesse fornendo dei valori più intuitivi come Valence e Energy. Un'altro aspetto toccato dalle features riguarda la teoria musicale con Key, Mode e Time\_signature. A giudicare dalla documentazione di Spotify le features numeriche in questo dataset sono aggregate e ad alto livello.

Quanti sono i generi musicali presenti nel dataset?

La totalità di generi nel dataset ammonta a 626. Inoltre la distribuzione dei samples sui generi non è per nulla uniforme. Nella figura in basso è rappresentato la distribuzione dei generi sul numero di samples. Intuitivamente si nota che nei valori tra 100 e 200 ci sono molti generi, mentre su altri valori abbiamo molti meno generi. In altre parole gran parte dei generi ha tra 100 e 200 samples, gli altri invece spazzano da quelli meno di 50 samples a quelli con più di 1800 samples. C'è da considerare inoltre che per fare questo grafico sono stati create dei bin da 25 e all'interno la situazione potrebbe essere molto più variegata.



## Task

Il task che si desidera affrontare è quello di sviluppare un classificatore in grado di assegnare ad ogni brano un solo genere musicale tra 18 super-generi creati ad-hoc. Il motivo per cui i generi sono stati raggruppati in 18 super generi è dovuto in parte alla natura del dominio applicativo e alle sopraelencate caratteristiche “liquide” che rendono la classificazione a grana fine difficoltosa. Il fatto che il dataset tratti 626 generi con numero di samples insufficiente per molti di questi aggrava il problema.

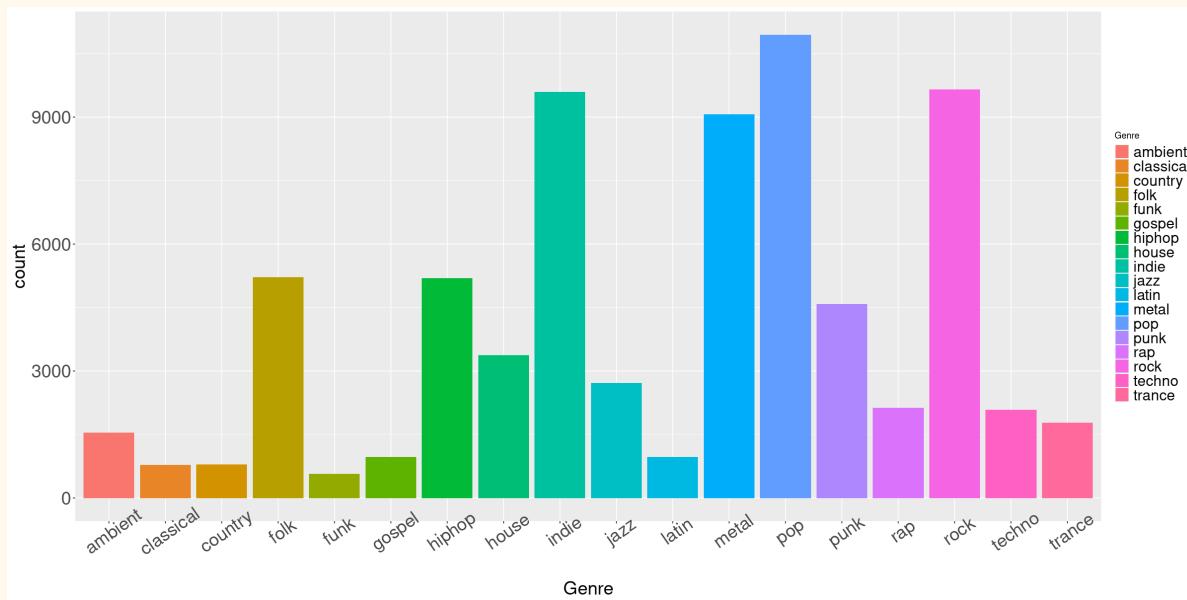
## Preprocessing dei dati:

Per ridurre la numerosità dei generi è stato fatto un raggruppamento in gran parte manuale:

- 1) sono stati tenuti solamente i generi con almeno 100 samples poichè generi con pochi brani potrebbero essere difficili da riconoscere
- 2) i generi musicali rimasti sono stati raggrupparli inizialmente sulla base del nome poichè molti nomi di generi hanno una radice. Ad esempio “symphonic power metal”, “instrumental progressive metal” vengono assegnati al gruppo “metal”
- 3) i gruppi ottenuti sono stati sottoposti a un controllo manuale onde evitare posizionamenti errati dal punto di vista stilistico-musicale
- 4) i generi senza radice comune sono stati manualmente posizionati dentro i gruppi idonei
- 5) per cercare di migliorare la bontà del raggruppamento, per ogni genere sconosciuto è stata fatta una ricerca su spotify al fine da valutare “ad orecchio” se fosse il caso di mantenere il gruppo di appartenza oppure cambiarlo.

I 18 super generi così ottenuti sono i seguenti:

ambient, classical, country, folk, funk, gospel, hiphop, house, indie, jazz, latin, metal, pop, punk, rap, rock, techno e trance.

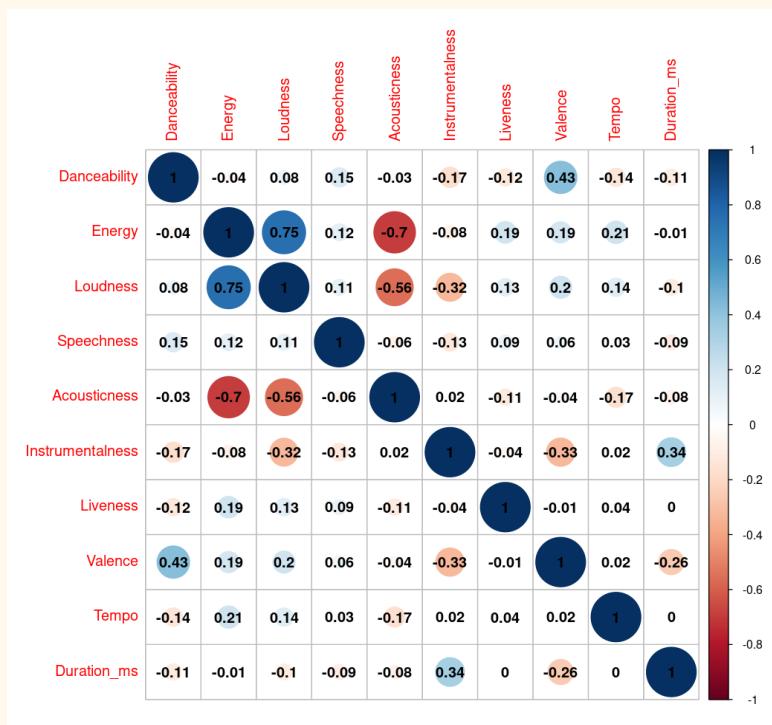


Le features numeriche del dataset sono state successivamente normalizzate.

## Analisi delle features

Allo scopo di analizzare le features numeriche è stata calcolata la matrice di correlazione di queste. La matrice di correlazione ci dice per ogni coppia di features quanto sono correlate e in che modo. La correlazione prende valori nell'intervallo [-1,1]. Due features positivamente correlate (correlazione vicina a 1) indicano che al crescere di una, la seconda cresce. Se la correlazione risulta negativa (correlazione vicina a -1) significa che al crescere di una l'altra decresce. Due features risultano scorrelate invece se non è osservabile una relazione tra queste e si manifesta con una correlazione pari a 0.

L'utilità di questa analisi sta nel fatto che se esiste un certo grado di correlazione tra le variabili, allora parte dell'informazione veicolata da uno è già contenuta nell'altra per cui questo risulterebbe vantaggioso in caso di riduzione della dimensionalità.



La matrice di correlazione per il dataset ci mostra che le variabili correlate sono poche e comunque non particolarmente correlate.

E' comunque possibile dare una spiegazione ad alcune delle coppie:

- **Acousticness-Energy:** (-0.7): pezzi acustici sono generalmente meno energici dei pezzi con strumenti alimentati ad energia elettrica

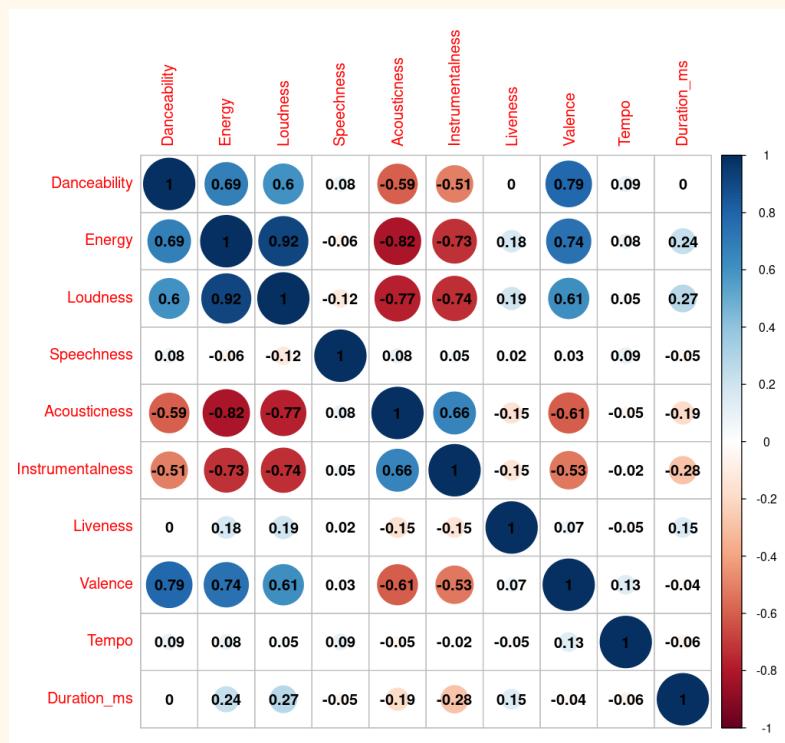
- **Acousticness-Loudness:** (-0.56) i brani acustici tendenzialmente non hanno un impatto violento
- **Loudness-Energy:** (0.75) i brani sono per scelta rumorosi allo scopo di conferire vigore
- **Valence-Danceability:** (0,43): i brani ballabili hanno per di più un mood positivo, ma non sempre, i lenti infatti hanno un mood negativo.
- **Duration- Instrumentalness:** (0,34) chi fa un brano strumentale spesso non lo dirige al grande pubblico il quale è molto esigente in termini di durata (max 3.5 min). Tutta via questa è una libertà e non un vincolo per cui l'artista può decidere di far durare quanto vuole il proprio brano causandone in alcuni casi l'allungarsi della durata.

I dati non evidenziano correlazioni forti per cui non ci sono sono features ridondanti e perciò una riduzione della dimensionalità non sembra essere necessaria. La spieagazione può essere dovuta al fatto che le features sono valori molto aggregati. Ritengo che dati più grezzi avrebbero evidenziato più pattern tra le variabili.

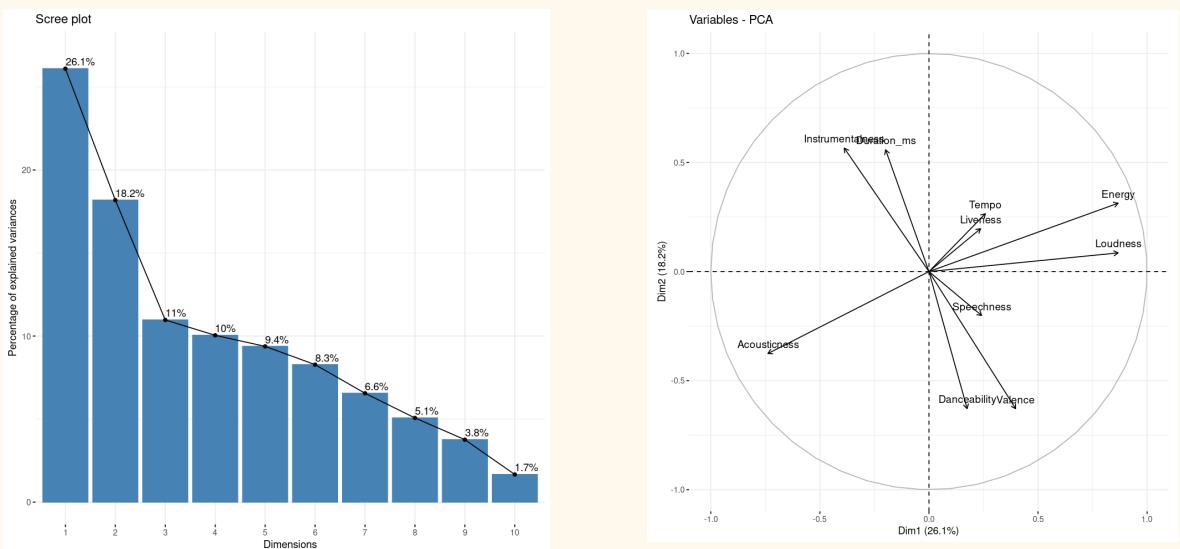
Sebbene non possiamo cambiare le features possiamo prendere due generi molto diversi e analizzare la correlazione per vedere se in questi casi estremi anche le features aggregate possono dare luogo a ridondanze.

Consideriamo quidni i generi classical e latin. Le correlazioni risultano molto più marcate.

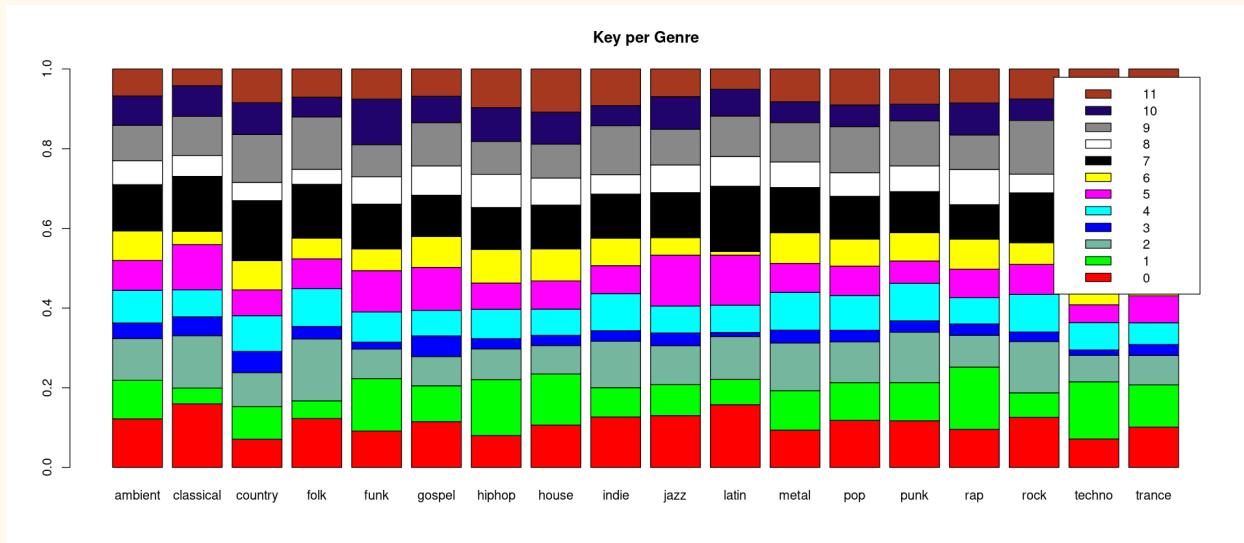
Eclatanti risultano le coppie Valence-Danceability (0,79) e Acousticness-Energy (-0.82) è infatti presumibile che ci sia una certa ridondanza: considerando la musica classi e la musica latina come la salsa, la Valence (mood) e la Danceability sono fattori "gemelli" infatti la salsa è tipicamente allegra, estremamente ballabile , mentre la musica classica non è frequentemente ballabile e può capitare che abbia un mood negativo.



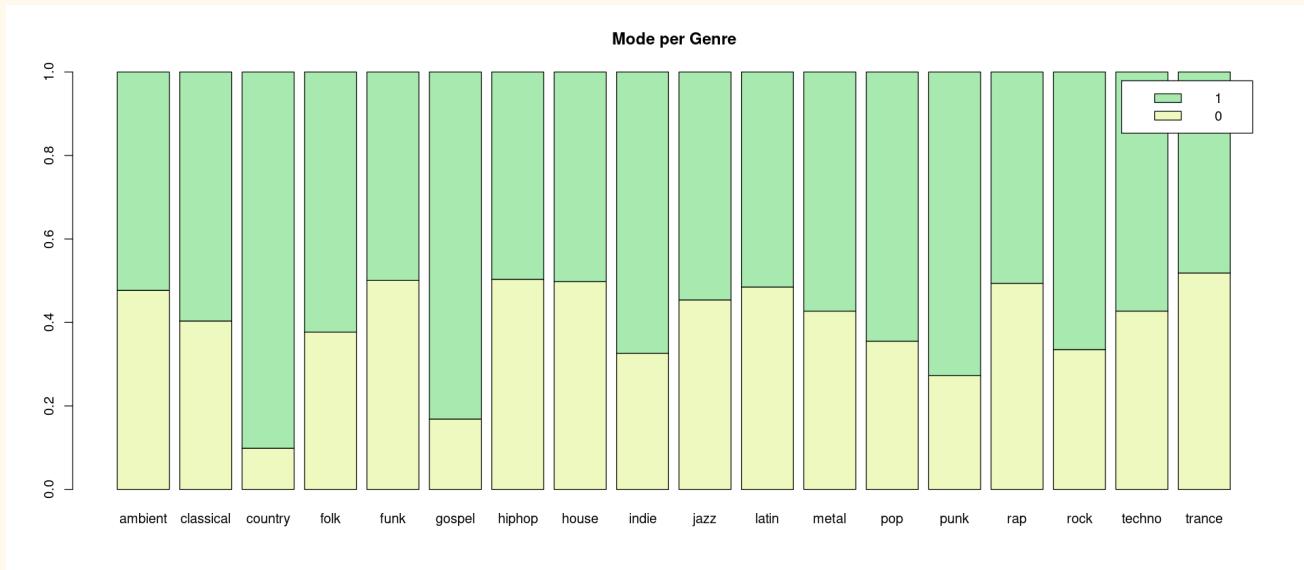
Un ulteriore conferma arriva dalla PCA che ci mostra come la varianza è spalmata praticamente su tutte le componenti rendendo tuttavia possibile ridurre lo spazio di due dimensioni ottenendo comunque un'explainingability del 94.5%. Il grafico a cerchio invece ci mostra il contributo delle features numeriche sulla PC1 (26.1%) e PC2 (18.2%) e conferma le correlazioni tra Energy-Loudness, Danceability-Valence, Energy-Acousticness, Energy-Loudness, ecc. Ci indica però che Energy-Instrumentalness sono scorrelate ma può essere dovuto al fatto che PC1 e PC2 spiegano insieme solo il 44% della varianza.



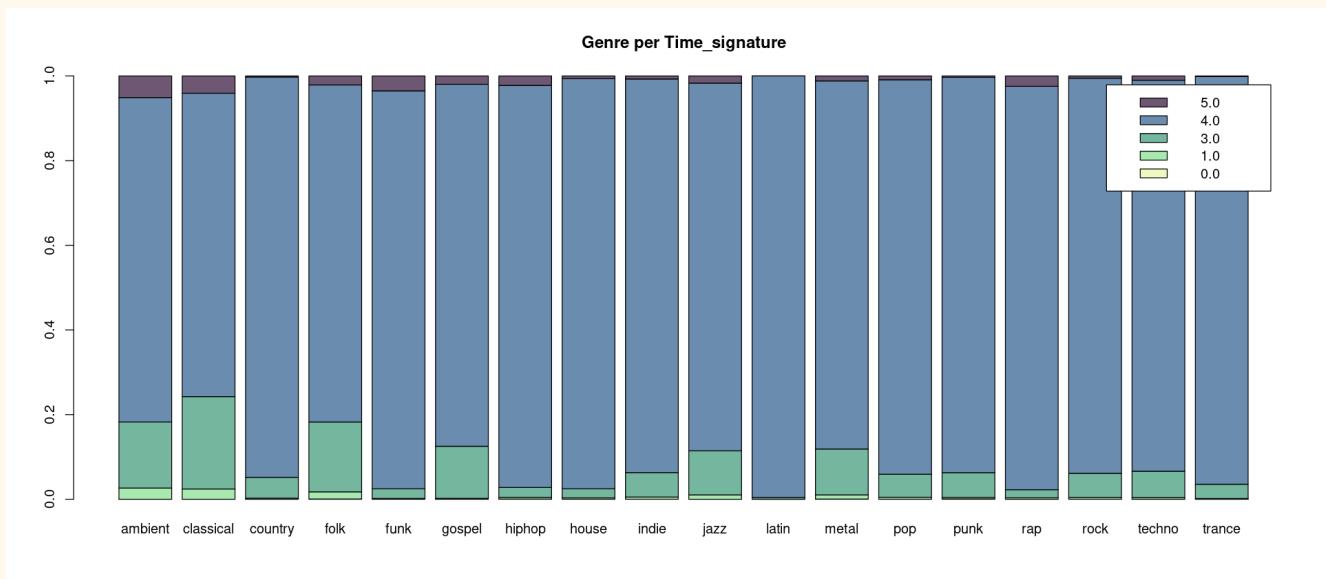
Plotando invece la separabilità dei generi considerando le variabili numeriche a coppie non sono presenti cluster evidenti di samples appartenenti allo stesso genere.



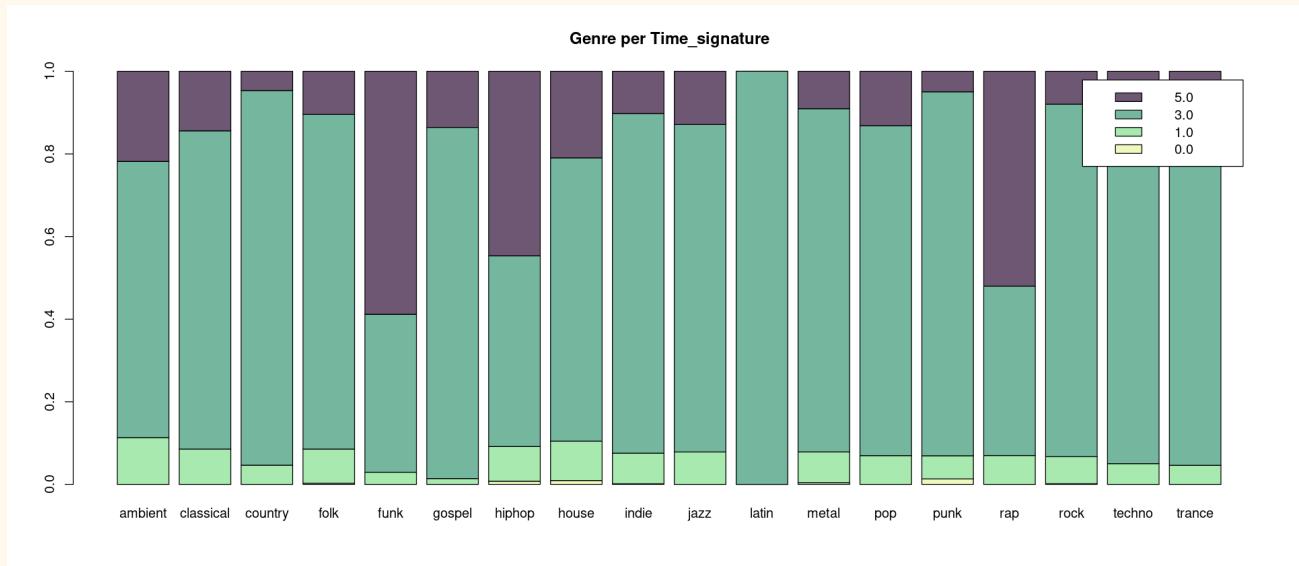
L'analisi delle tre features categoriche (Key, Mode e Time\_signature) risultano poco caratterizzanti dei generi musicali. Prendendo Key (tonalità) e analizzando la sua presenza all'interno dei singoli generi possiamo vedere nella figura in alto che la distribuzione delle tonalità è molto simile in tutti i generi. Più diversificata è invece la Mode sui vari generi dove possiamo vedere come il country, il gospel e il punk siano caratterizzati da una modalità prevalentemente maggiore (1)



E' impressionante invece come la stragrande maggioranza dei brani abbia Time\_signature pari a 4 (presumibilmente 4/4) per cui sapere che un brano è in 4/4 non ci è utile in termini di predizione del genere.



Tuttavia se ignoriamo i samples con Time\_signature pari a 4 e analizziamo i restanti vediamo casi interessanti come il funk e il rap che hanno Time\_signature 5 per il 50% dei casi mentre negli altri generi tale percentuale è molto minore. Determinante invece sembra essere per il genere latin che in questo plot assume interamente il valore 3. Questo significa che se un brano non ha Time\_signature 4 e nemmeno 3 allora non può essere di genere latin.



## Modellazione e training

Per il problema di multiclass classification sono stati usati i seguenti modelli:

### **Naive Bayes:**

Questo algoritmo è indicato per la classificazione sia binaria che multiclasse. E' chiamato Naive perchè la peculiarità principale dell'algoritmo, oltre ad avvalersi del teorema di Bayes, è che si basa sul fatto che tutte le caratteristiche non siano correlate l'una all'altra. La presenza o l'assenza di una caratteristica non influenzano la presenza o l'assenza di altre. Il calcolo ne risulta quindi agevolato.

### **Random Forest**

Un random forest consiste in un grande numero di alberi di decisione individuali che operano come un ensemble, ovvero ogni albero di decisione predice una classe e la classe con più "voti" diventa la predizione del modello. Il motivo per cui il random forest è un algoritmo in genere molto performante è per via del numero elevato di alberi decisionali che votano insieme per decidere la classe vincente, in questo modo sebbene ci possano essere degli alberi che sbagliano, la "saggezza delle folle" aiuta a ridurre questo problema.

### **Deep Neural Network:**

Le reti neurali profonde sono particolari algoritmi organizzati a strati di neuroni. La peculiarità delle reti neurali profonde è che il modello internamente trasforma le features aumentando il grado di astrazione man mano si prosegue con gli strati. Tutta via le reti neurali sono sensibili all'overfitting per cui è importante applicare tecniche di regolarizzazione che mitighino il problema. Tra le tecniche di regolarizzazione esiste il dropout dove a ogni iterazione esclude dei neuroni con una certa probabilità.

Nel caso specifico del progetto, le features categoriche sono state codificate usando l'one-hot encoding raggiungendo le 27 features (numeriche + categoriche codificate) ed è stata usata una rete con la seguente architettura (la cui configurazione è stata decisa sperimentalmente usando la CV, vedi dopo):

- input layer 27 unità
- layer denso 32 unità (dropout 40%)
- layer denso 16 unità
- layer denso 8 unità
- layer denso 4 unità
- output layer 18 unità (softmax)

La divisione tra training set e test set è stata fatta seguendo le proporzioni 80%-20% usando un sampling stratificato.

## Sul training set:

allo scopo di fare tuning per migliorare le performance sono state valutate attraverso la 10-fold cross validation diverse alternative che considerassero i seguenti parametri:

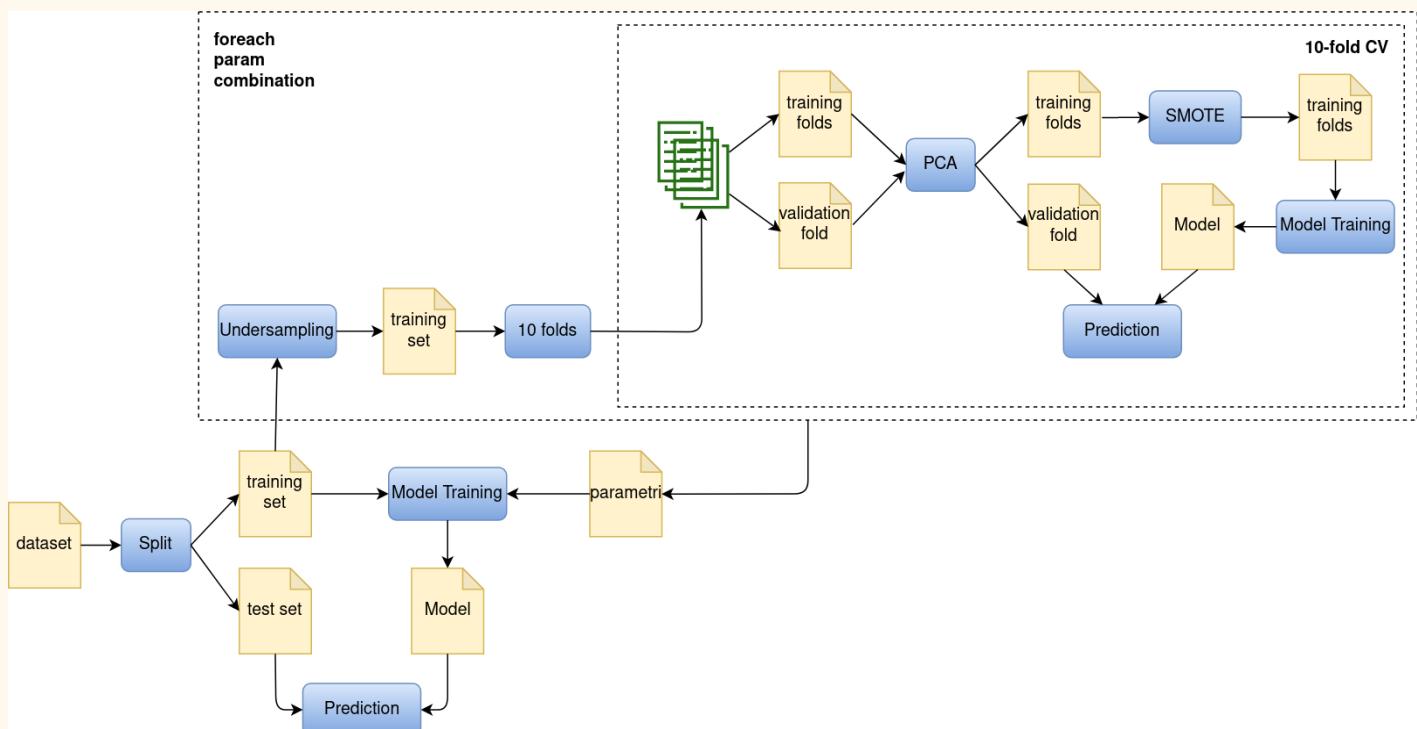
- **undersampling**: nessun undersampling o in caso contrario quanto applicarne
  - **oversampling (SMOTE)**: oversampling tramite generazione di dati sintetici. Se si, quanto?
  - **PCA**: riduzione della dimensionalità

In aggiunta, grazie alla cross-validation è stato possibile fare tuning sui parametri della rete neurale valutandone profondità e numero di neuroni per strato.

In fine i modelli vengono allenati su tutto il training set con la configurazione migliore ottenuta precedentemente.

**Sul test set:**

### misure di performance



## Tuning

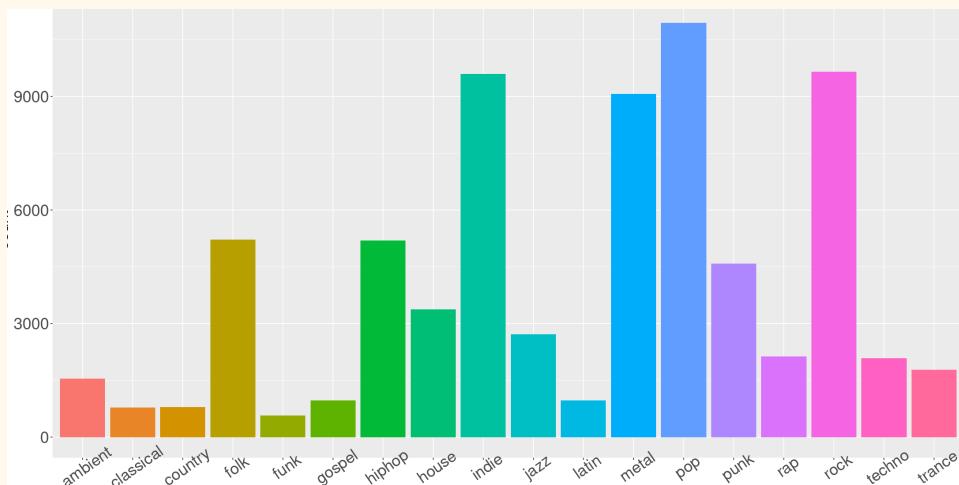
Come si vede chiaramente dalla distribuzione dei generi, viene affrontato un problema di multiclass classification su classi fortemente sbilanciate dove però tutte le classi hanno uguale importanza.

La cross-validation ci aiuterà quindi a capire se l'uso di undersampling, oversampling e PCA ha un effetto positivo o negativo sui 3 modelli scelti.

In particolare l'oversampling e l'undersampling sono stati scelti attentamente allo scopo di evitare che vengano generati troppi sample sintetici. In particolare se ricordiamo la distribuzione dei samples sui generi ci rendiamo conto come evitando l'undersampling avremmo delle classi altamente sbilanciate che potremmo pareggiare con l'oversampling ma a costo di generare troppi dati sintetici, per cui nella zona rossa della tabella sottostante abbiamo scelto di generare dati sintetici per le classi minori fino al 60% rispetto alla classe maggiore. Per cui dopo l'oversampling i samples per le classi minori saranno il 60% di quella maggiore.

In modo analogo accade per undersampling=3000 e SMOTE=0.7. Riducendo le classi fino a 3000 samples possiamo permetterci di fare oversampling fino al 70% della classe top. Con undersampling=2000 possiamo puntare a pareggiare la classe top. Con undersampling=1000 e nessun oversampling si avrebbe una distribuzione quasi uniforme. In fine con undersampling=600 si avrebbe una distribuzione uniforme. Naturalmente in questi ultimi due casi non abbiamo bisogno dell'oversampling in quanto la distribuzione ce lo permette.

In tutti questi è stato valutato anche l'uso della riduzione dello spazio per le features numeriche a 8 dimensioni tramite PCA.



Possiamo vedere in tabella come in tutti casi l'uso di SMOTE per oversampling peggiori le performance.

L'undersampling invece risulta avere un impatto positivo e più si pareggiano le classi tramite undersampling e maggiore è l'accuracy dei modelli. La riduzione delle dimensionalità con PCA migliora le performance solo nel caso di NB mentre per RF e NN la PCA ha un'influenza negativa.

In conclusione questa valutazione ci dice che senz'altro le performance peggiori si ottengono non usando undersampling e pareggiando le classi solo con SMOTE, mentre le performance migliori sono quelle dove si applica undersampling fino a 1000 senza PCA e senza SMOTE riuscendo a raggiungere un'accuracy del 50,6 % con RF.

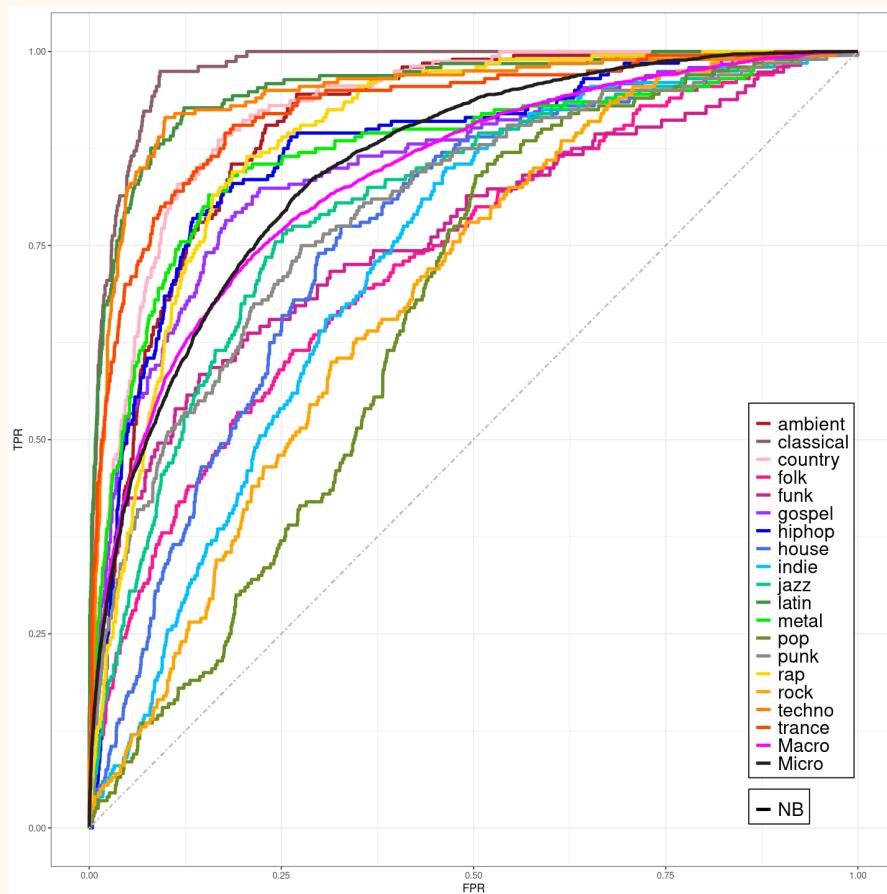
<b>Undersampling</b>	<b>PCA</b>	<b>SMOTE</b>	<b>NB Accuracy</b>	<b>RF Accuracy</b>	<b>NN Accuracy</b>
-	8	-	0,328	0,454	0,346
-	8	0,6	0,248	0,436	0,291
-	-	-	0,261	0,482	0,347
-	-	0,6	0,216	0,466	0,303
3000	8	-	0,319	0,459	0,356
3000	8	0,7	0,296	0,454	0,332
3000	-	-	0,291	0,493	0,366
3000	-	0,7	0,279	0,488	0,35
2000	8	-	0,336	0,464	0,362
2000	8	1	0,321	0,457	0,35
2000	-	-	0,327	<b>0,506</b>	<b>0,371</b>
2000	-	1	0,317	<b>0,5</b>	0,364
1000	8	-	<b>0,36</b>	0,466	<b>0,371</b>
1000	-	-	<b>0,362</b>	<b>0,506</b>	<b>0,381</b>
600	8	-	0,357	0,443	0,36
600	-	-	<b>0,361</b>	0,494	<b>0,377</b>

## Testing e analisi dei risultati

La cross-validation precedentemente descritta ha evidenziato che i tre modelli performano meglio pareggiando le classi tramite undersampling fino a raggiungere i 1000 samples per classe e senza oversampling e in fine nessuna riduzione della dimensionalità con PCA.

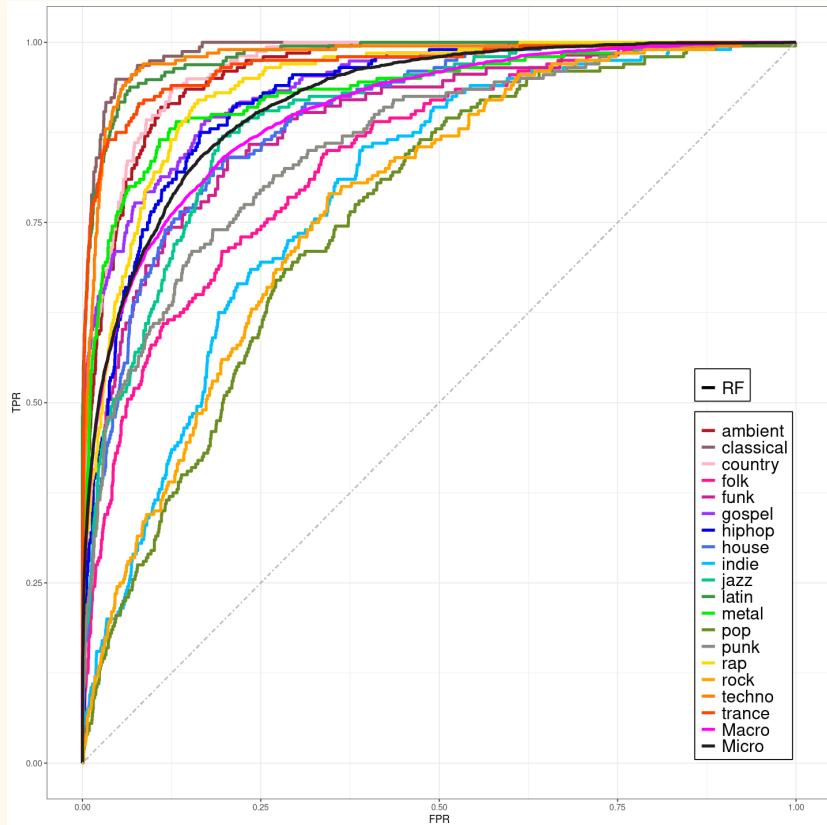
I modelli vengono quindi allenati su tutto il training set processato come sopra descritto e le performance valutate sul test set.

Per quanto riguarda le curve ROC in seguito verranno mostrate le multi ROC per tutti i modelli. La curva multi ROC ci permette di vedere in un plot unico tutte le curve ROC associate alle diverse classi. La ROC ci dà informazioni circa la separabilità di una classe rispetto alle altre. Le curve che sono molto vicine alla diagonale indicano che il modello non è in grado di distinguere la classe.



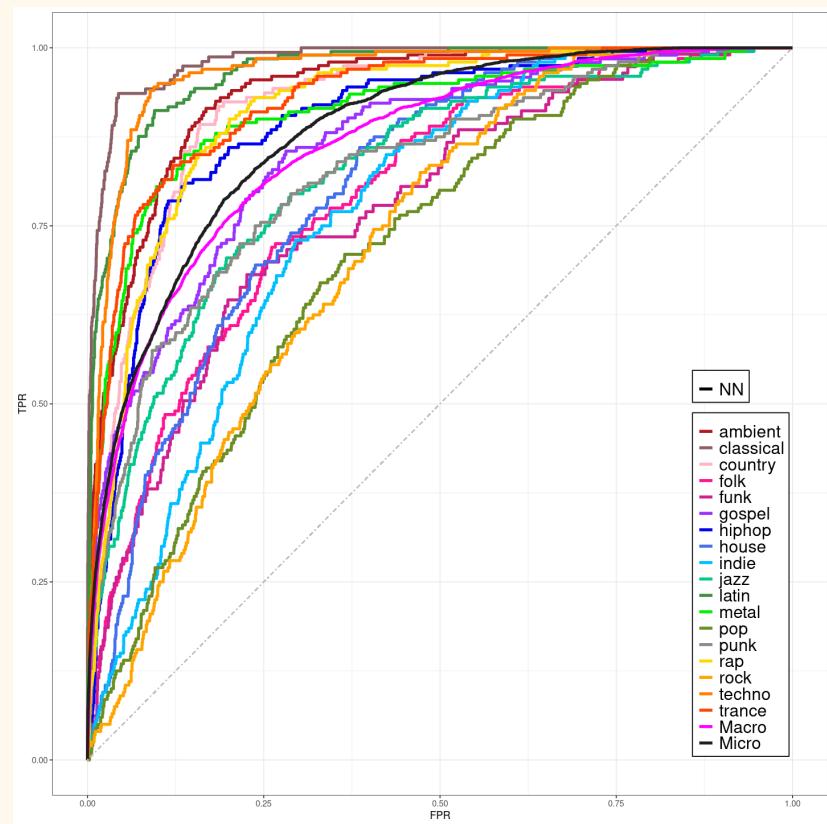
Multi ROC per NB:

peggiori: pop, rock e indie,  
migliori: classical, trance e latin



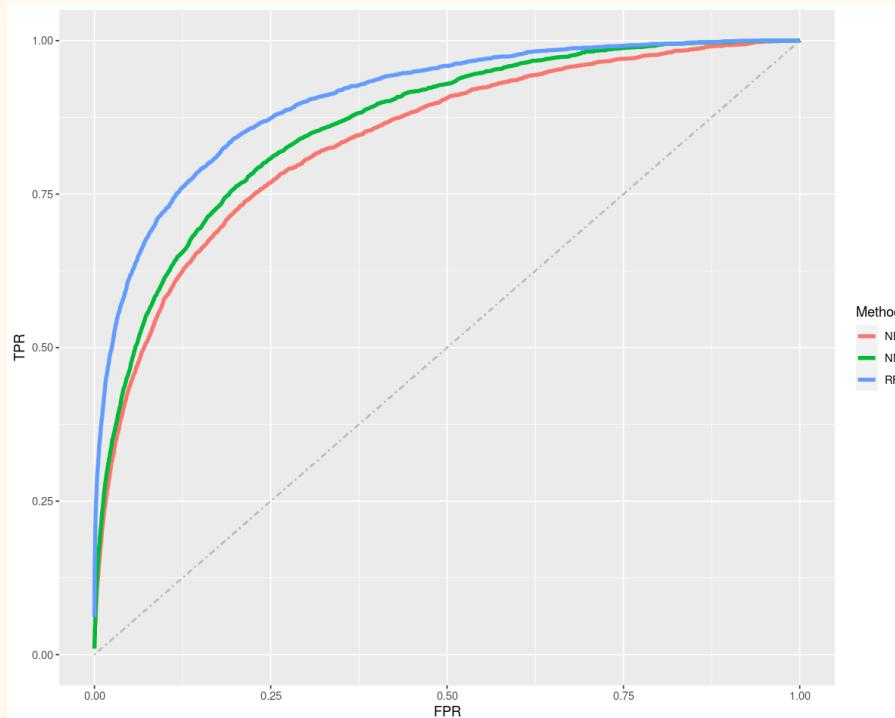
Multi ROC per RF:

peggiori: pop, rock, indie,  
migliori: classical, trance, latin



Multi ROC per NN:

peggiori: pop, rock, indie  
migliori: classical, techno, latin



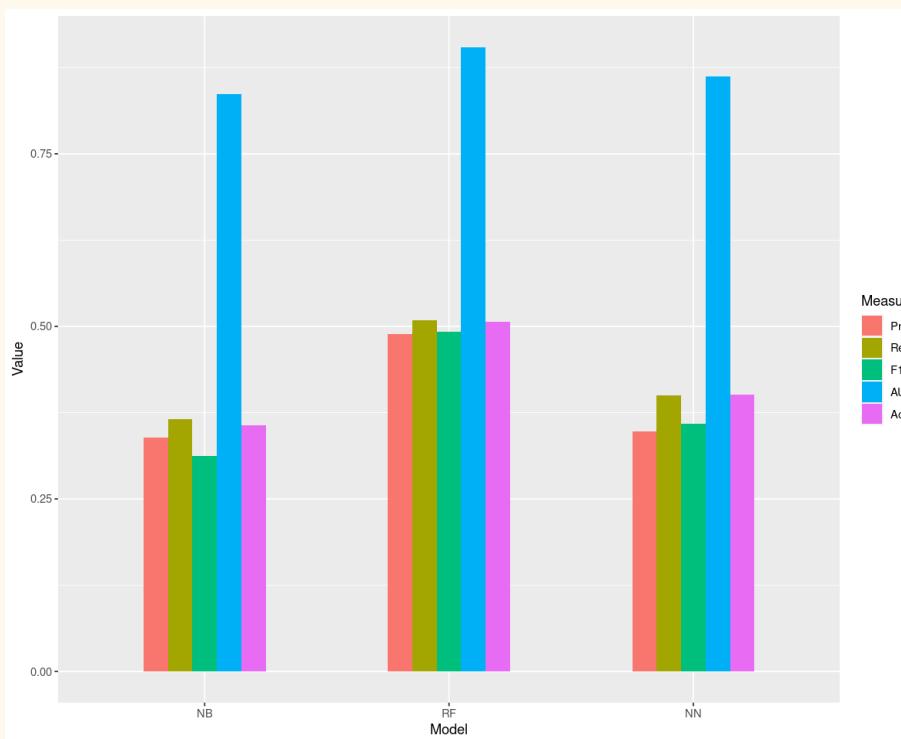
ROC globale per ogni modello

L'area sotto la ROC è descritta da AUC e in questo caso è un valore medio tra le AUC per tutte le cassi. L'AUC in particolare ci dice quanto un modello è in grado di discriminare tra le classi.

AUC NB = 0,836

AUC RF = 0,904

AUC NN = 0,861



In questo grafico è stata una misura macro, per cui una media non pesata poichè ogni classe ha uguale importanza.

Acc NB = 0.35

Acc RF = 0.50

Acc NN = 0.40

## Conclusioni

Dalle curve ROC macro e dalla accuracy il modello vincitore è il Random Forest.

Dalle curve ROC singole invece vediamo come la classifica delle top-3 migliori e peggiori curve concidano. In particolare le peggiori risultano pop, rock e indie, mentre le migliori classical, trance e latin.

Non è un caso che la capacità dei modelli nel discriminare tra le classi sia molto bassa per alcuni mentre per altri maggiore. Il motivo è che la purezza del raggruppamento non è omogeneo, ci sono gruppi come rock, pop e indie, che sono estremamente ampi, contaminati ed eterogenei mentre gruppi come techno, latin e classical più circoscritti.

La difficoltà che hanno i classificatori in un task di questo tipo è analoga alla difficoltà che hanno le persone con una vasta cultura musicale quando pongono la domanda “che musica ascolti?” e l’interlocutore risponde con “ascolto pop”. In questo caso l’unica cosa rimane ipotizzare che si intenda la musica trasmessa in radio oggi poiché anche la musica da radio negli anni è cambiata molto: si è passati negli anni da Elvis Presley ai Beatles a Madonna a Lady Gaga. In modo simile accade con il rock.

In conclusione il task prefissato era ambizioso: da una parte si è cercato di discriminare tra generi approssimativamente raggruppati, dall’altro i risultati ottenuti trovano in riscontro nella realtà delle cose.

In fine le ragioni dell’accuratezza relativamente bassa sono da attribuire a:

- features numeriche sono valori troppo aggregati
- raggruppamento manuale è approssimativo e disomogeneo

Tuttavia osserviamo che:

- l’esistenza di tanti generi è giustificabile dall’enorme ricchezza e dimensionalità della musica
- features più grezze aiuterebbero a rappresentare meglio tutta questa varianza

## Appendice

Naive Bayes confusion matrix

		Reference									
Prediction		ambient	classical	country	folk	funk	gospel	hiphop	house	indie	jazz
ambient		62	14	0	18	1	3	2	6	6	11
classical		67	130	3	36	2	6	1	0	7	27
country		1	2	135	42	9	20	21	52	75	12
folk		2	4	3	24	0	7	2	2	5	8
funk		0	0	0	1	24	0	3	2	3	17
gospel		2	1	2	6	7	72	6	2	4	13
hiphop		0	0	0	6	4	3	24	5	3	5
house		6	0	0	1	9	9	5	14	5	11
indie		1	2	2	11	0	9	1	5	8	1
jazz		5	3	1	11	12	9	7	7	6	36
latin		0	0	7	22	13	25	7	6	14	19
metal		8	0	0	9	8	7	0	3	11	7
pop		0	0	0	0	1	0	2	2	2	2
punk		0	0	2	4	2	5	0	3	26	1
rap		1	0	3	6	6	13	117	34	17	11
rock		0	0	0	2	0	3	0	0	2	4
techno		36	0	0	1	6	0	1	48	1	9
trance		9	0	0	0	9	2	1	9	5	6
		Reference									
Prediction		latin	metal	pop	punk	rap	rock	techno	trance		
ambient		0	11	0	4	1	4	10	4		
classical		0	2	7	4	0	5	0	0		
country		13	24	89	63	35	75	1	17		
folk		0	3	4	2	1	5	0	1		
funk		1	3	5	3	0	2	4	2		
gospel		11	0	8	4	6	14	0	1		
hiphop		0	0	0	0	25	4	1	1		
house		1	2	4	2	4	6	6	6		
indie		0	0	6	4	3	4	0	0		
jazz		4	1	7	2	1	7	2	1		
latin		156	4	15	8	14	24	1	2		
metal		0	94	8	29	3	11	1	7		
pop		0	1	3	1	1	0	0	0		
punk		1	31	11	55	2	16	2	5		
rap		6	6	24	11	104	9	1	4		
rock		0	1	1	1	0	5	0	1		
techno		0	1	3	0	0	1	149	27		
trance		0	16	5	7	0	8	22	121		

Overall Statistics

Accuracy : 0.3563

## Random Forest confusion matrix

		Reference									
Prediction		ambient	classical	country	folk	funk	gospel	hiphop	house	indie	jazz
ambient	126	20	0	16	3	0	2	10	5	13	
classical	16	119	0	8	0	7	1	1	1	1	3
country	0	0	106	8	0	0	3	9	20	2	
folk	3	6	5	67	2	10	6	0	14	16	
funk	1	0	0	3	41	2	3	2	5	8	
gospel	0	0	6	10	1	120	2	1	11	4	
hiphop	1	0	1	5	4	7	106	10	8	14	
house	12	0	3	3	7	9	15	77	10	7	
indie	1	3	12	9	2	5	2	7	39	2	
jazz	14	8	1	20	23	7	5	6	13	96	
latin	0	0	5	11	11	15	4	2	8	13	
metal	3	0	0	12	1	1	0	6	11	2	
pop	1	0	2	3	1	2	7	6	10	3	
punk	0	0	8	11	1	1	1	1	23	1	
rap	1	0	4	6	4	3	42	15	3	2	
rock	0	0	5	8	6	3	0	4	18	4	
techno	18	0	0	0	6	1	1	41	0	7	
trance	3	0	0	0	0	0	0	2	1	3	
		Reference									
Prediction		latin	metal	pop	punk	rap	rock	techno	trance		
ambient	0	6	2	4	1	4	12	4			
classical	0	0	2	1	0	2	0	0			
country	3	2	23	20	7	15	0	2			
folk	4	6	16	13	1	14	0	1			
funk	0	1	2	2	1	2	2	1			
gospel	9	0	7	1	3	12	0	1			
hiphop	0	2	12	7	60	8	0	1			
house	2	2	22	4	10	7	15	2			
indie	0	5	17	11	3	25	1	4			
jazz	6	3	8	5	0	10	1	1			
latin	160	1	9	7	4	15	0	1			
metal	0	138	13	26	1	13	3	6			
pop	3	2	16	4	4	9	1	2			
punk	1	19	12	79	5	22	0	4			
rap	4	4	18	3	100	7	1	0			
rock	1	3	15	8	0	32	0	1			
techno	0	0	0	1	0	0	153	14			
trance	0	6	6	4	0	3	11	155			

## Overall Statistics

Accuracy : 0.5069

Neural Network confusion matrix:

		Reference									
Prediction		ambient	classical	country	folk	funk	gospel	hiphop	house	indie	jazz
ambient	104	8	0	11	5	0	1	8	4	11	
classical	29	128	0	20	1	7	1	0	2	17	
country	0	2	100	19	7	10	4	16	33	9	
folk	10	11	8	59	3	19	3	1	22	25	
funk	1	0	0	1	3	4	4	1	0	1	
gospel	0	2	6	20	3	80	1	1	10	10	
hiphop	0	0	0	9	9	12	102	17	8	11	
house	3	0	1	2	14	5	11	40	7	9	
indie	1	0	1	0	0	0	0	0	3	0	
jazz	4	5	0	16	16	10	12	8	11	63	
latin	0	0	5	13	11	24	6	2	8	15	
metal	4	0	1	10	3	1	0	8	20	5	
pop	1	0	8	5	1	3	6	15	12	8	
punk	0	0	16	10	2	7	3	5	43	0	
rap	2	0	11	3	1	4	45	27	13	4	
rock	0	0	1	0	3	6	0	1	2	2	
techno	34	0	0	1	11	1	1	45	0	7	
trance	7	0	0	1	20	0	0	5	2	3	
		Reference									
Prediction		latin	metal	pop	punk	rap	rock	techno	trance		
ambient	0	5	1	4	0	3	7	6			
classical	0	2	4	3	0	2	0	0			
country	9	5	40	27	11	39	0	3			
folk	1	5	17	9	0	16	0	2			
funk	0	1	0	0	0	1	0	1			
gospel	14	3	11	11	1	19	0	0			
hiphop	10	1	14	7	77	17	1	2			
house	0	1	12	4	13	8	9	7			
indie	0	1	4	3	0	5	0	0			
jazz	8	0	10	3	3	10	1	1			
latin	148	0	8	5	7	16	0	0			
metal	0	111	13	30	1	16	4	19			
pop	1	1	10	4	6	8	0	2			
punk	0	40	26	80	7	29	0	7			
rap	1	6	21	5	72	7	1	4			
rock	1	1	0	1	1	0	0	0			
techno	0	0	2	0	1	0	149	31			
trance	0	17	7	4	0	4	28	115			

#### Overall Statistics

Accuracy : 0.4005

## Naive Bayes precision, recall e F-measure

	Class: ambient	Class: classical	Class: country	Class: folk
Precision	0.39490	0.43771	0.19679	0.328767
Recall	0.31000	0.83333	0.85443	0.120000
F1	0.34734	0.57395	0.31991	0.175824
Prevalence	0.05860	0.04571	0.04629	0.058599
Detection Rate	0.01817	0.03809	0.03955	0.007032
Detection Prevalence	0.04600	0.08702	0.20100	0.021389
Balanced Accuracy	0.64022	0.89103	0.84258	0.552375
	Class: funk	Class: gospel	Class: hiphop	Class: house
Precision	0.342857	0.45283	0.296296	0.153846
Recall	0.212389	0.37306	0.120000	0.070000
F1	0.262295	0.40909	0.170819	0.096220
Prevalence	0.033109	0.05655	0.058599	0.058599
Detection Rate	0.007032	0.02110	0.007032	0.004102
Detection Prevalence	0.020510	0.04659	0.023733	0.026663
Balanced Accuracy	0.599225	0.67302	0.551130	0.523017
	Class: indie	Class: jazz	Class: latin	Class: metal
Precision	0.140351	0.29508	0.46291	0.45631
Recall	0.040000	0.18000	0.80829	0.47000
F1	0.062257	0.22360	0.58868	0.46305
Prevalence	0.058599	0.05860	0.05655	0.05860
Detection Rate	0.002344	0.01055	0.04571	0.02754
Detection Prevalence	0.016701	0.03575	0.09874	0.06036
Balanced Accuracy	0.512375	0.57662	0.87604	0.71757
	Class: pop	Class: punk	Class: rap	Class: rock
Precision	0.200000	0.33133	0.27882	0.250000
Recall	0.015000	0.27500	0.52000	0.025000
F1	0.027907	0.30055	0.36300	0.045455
Prevalence	0.058599	0.05860	0.05860	0.058599
Detection Rate	0.000879	0.01611	0.03047	0.001465
Detection Prevalence	0.004395	0.04864	0.10929	0.005860
Balanced Accuracy	0.505633	0.62023	0.71814	0.510166
	Class: techno	Class: trance		
Precision	0.52650	0.55000		
Recall	0.74500	0.60500		
F1	0.61698	0.57619		
Prevalence	0.05860	0.05860		
Detection Rate	0.04366	0.03545		
Detection Prevalence	0.08292	0.06446		
Balanced Accuracy	0.85165	0.78709		

## Random Forest precision, recall e F-measure

	Class: ambient	Class: classical	Class: country	Class: folk
Precision	0.55263	0.73913	0.48182	0.36413
Recall	0.63000	0.76282	0.67089	0.33500
F1	0.58879	0.75079	0.56085	0.34896
Prevalence	0.05860	0.04571	0.04629	0.05860
Detection Rate	0.03692	0.03487	0.03106	0.01963
Detection Prevalence	0.06680	0.04717	0.06446	0.05391
Balanced Accuracy	0.79913	0.87496	0.81793	0.64929
	Class: funk	Class: gospel	Class: hiphop	Class: house
Precision	0.53947	0.63830	0.43089	0.37198
Recall	0.36283	0.62176	0.53000	0.38500
F1	0.43386	0.62992	0.47534	0.37838
Prevalence	0.03311	0.05655	0.05860	0.05860
Detection Rate	0.01201	0.03516	0.03106	0.02256
Detection Prevalence	0.02227	0.05508	0.07208	0.06065
Balanced Accuracy	0.67611	0.80032	0.74321	0.67227
	Class: indie	Class: jazz	Class: latin	Class: metal
Precision	0.26351	0.42291	0.60150	0.58475
Recall	0.19500	0.48000	0.82902	0.69000
F1	0.22414	0.44965	0.69717	0.63303
Prevalence	0.05860	0.05860	0.05655	0.05860
Detection Rate	0.01143	0.02813	0.04688	0.04043
Detection Prevalence	0.04336	0.06651	0.07794	0.06915
Balanced Accuracy	0.58054	0.71961	0.89805	0.82975
	Class: pop	Class: punk	Class: rap	Class: rock
Precision	0.210526	0.41799	0.46083	0.296296
Recall	0.080000	0.39500	0.50000	0.160000
F1	0.115942	0.40617	0.47962	0.207792
Prevalence	0.058599	0.05860	0.05860	0.058599
Detection Rate	0.004688	0.02315	0.02930	0.009376
Detection Prevalence	0.022268	0.05538	0.06358	0.031644
Balanced Accuracy	0.530663	0.68038	0.73179	0.568173
	Class: techno	Class: trance		
Precision	0.63223	0.79897		
Recall	0.76500	0.77500		
F1	0.69231	0.78680		
Prevalence	0.05860	0.05860		
Detection Rate	0.04483	0.04541		
Detection Prevalence	0.07091	0.05684		
Balanced Accuracy	0.86865	0.88143		

Neural Network precision, recall e F-measure

	Class: ambient	Class: classical	Class: country	Class: folk
Precision	0.58427	0.59259	0.29940	0.27962
Recall	0.52000	0.82051	0.63291	0.29500
F1	0.55026	0.68817	0.40650	0.28710
Prevalence	0.05860	0.04571	0.04629	0.05860
Detection Rate	0.03047	0.03750	0.02930	0.01729
Detection Prevalence	0.05215	0.06329	0.09786	0.06182
Balanced Accuracy	0.74848	0.89675	0.78051	0.62385
	Class: funk	Class: gospel	Class: hiphop	Class: house
Precision	0.166667	0.41667	0.34343	0.27397
Recall	0.026549	0.41451	0.51000	0.20000
F1	0.045802	0.41558	0.41046	0.23121
Prevalence	0.033109	0.05655	0.05860	0.05860
Detection Rate	0.000879	0.02344	0.02989	0.01172
Detection Prevalence	0.005274	0.05626	0.08702	0.04278
Balanced Accuracy	0.511002	0.68986	0.72465	0.58350
	Class: indie	Class: jazz	Class: latin	Class: metal
Precision	0.166667	0.34807	0.55224	0.45122
Recall	0.015000	0.31500	0.76684	0.55500
F1	0.027523	0.33071	0.64208	0.49776
Prevalence	0.058599	0.05860	0.05655	0.05860
Detection Rate	0.000879	0.01846	0.04336	0.03252
Detection Prevalence	0.005274	0.05303	0.07852	0.07208
Balanced Accuracy	0.505166	0.63914	0.86479	0.75649
	Class: pop	Class: punk	Class: rap	Class: rock
Precision	0.10989	0.29091	0.31718	0.000000
Recall	0.05000	0.40000	0.36000	0.000000
F1	0.06873	0.33684	0.33724	NaN
Prevalence	0.05860	0.05860	0.05860	0.058599
Detection Rate	0.00293	0.02344	0.02110	0.000000
Detection Prevalence	0.02666	0.08057	0.06651	0.005567
Balanced Accuracy	0.51239	0.66965	0.65588	0.497043
	Class: techno	Class: trance		
Precision	0.52650	0.53991		
Recall	0.74500	0.57500		
F1	0.61698	0.55690		
Prevalence	0.05860	0.05860		
Detection Rate	0.04366	0.03369		
Detection Prevalence	0.08292	0.06241		
Balanced Accuracy	0.85165	0.77225		