GRAU EN INTELIGENCIA ARTIFICIAL

IAA

PRÀCTICA DE ANTISPOOFING

Sergi Flores

28 DE DESEMBRE 2024

Introducció

Aquesta pràctica consisteix en fer un model que, donat un conjunt de característiques acústiques d'àudios reals i modificats digitalment amb intel·ligència artificial, sigui capaç de predir quins són els àudios autèntics i quins no.

La base de dades original té un total de 80.816 files d'àudios en 2 fitxers: un amb features numèriques (processades amb smile), i un altre amb features categòriques. Ambdós datasets seràn particionats mitjançant un conjunt de train i de test que trobem en 2 altres csv. Aquesta partició es pot realitzar mitjançant la vàriable UniqueID, de manera que per cada cas podem relacionar les 3 taules entre elles.

1. Preprocessament de les dades

Un cop ajuntats els datasets, hem de descartar les variables que no ens proporcionen informació útil pel modelatge, així com identificadors o variables amb només missings.

1.1. Anàlisi de les variables categòriques

Del conjunt de variables categòriques, les uniques que aporten informació rellevant són:

- Sex
- Country
- Utteracy

Les quals hem format a partir de varies variables (Sex i Target_Sex; Country i Target_Country; Utteracy, Source_Utteracy i Target_Utteracy).

Això es deu a que els àudios alterats amb intel·ligència artificial no disposaven d'una categoria d'aquestes de base, sinó que tenien una procedència i un destí, de manera que s'ha considerat la característica final com a la base.

Tot i això, la variable Utteracy és força problemàtica, ja que representa un identificador de la transcripció acústica del text de l'àudio. Això indica possibles problemes en relació amb Data Leakage i Overfitting, degut a que quan es vulgui aplicar el model amb frases no vistes en el conjunt d'entrenament, aquest tendirà a fer males prediccions, tot i que dintre del el nostre conjunt d'entrenament afegir aquesta variable pugui semblar beneficiós.

Així doncs, finalment les variables categòriques escollides al preprocessament són:

- Sex: indica si la persona que parla a l'àudio és un home o una dona
- Country: indica si la persona que parla a l'àudio procedeix d'un dels següents paisos de llatinoamèrica: Argentina, Chile, Venezuela, Perú o Colòmbia

1.2. Anàlisi de les variables numèriques

Les variables númèriques restants després de desfer-nos dels ids i variables repetides són les següents, les quals s'han renombrat per tal de facilitar la llegibilitat:

Variable Original	**Transcripció**
F0semitoneFrom27.5Hz_sma3nz_stddevRisingSlope	Semitone_Rising
F0semitoneFrom27.5Hz_sma3nz_meanFallingSlope	Semitone_Mean
F0semitoneFrom27.5Hz_sma3nz_stddevFallingSlope	Semitone_Falling
loudness_sma3_amean	Loudness_Mean
spectralFlux_sma3_stddevNorm	SpectralFlux
mfcc1_sma3_amean	mfcc1_mean
mfcc1_sma3_stddevNorm	mfcc1_stddev
mfcc2_sma3_amean	mfcc2_mean
mfcc2_sma3_stddevNorm	mfcc2_stddev
mfcc3_sma3_amean	mfcc3_mean
mfcc3_sma3_stddevNorm	mfcc3_stddev
jitterLocal_sma3nz_amean	Jitter_Mean
slopeUV500-1500_sma3nz_amean	Slope_Mean

Semitone_Rising (F0semitoneFrom27.5Hz_sma3nz_stddevRisingSlope)

Aquest variable representa la desviació estàndard de la pendent creixent de les semitones de F0 (freqüència fonamental) calculades a partir de 27,5 Hz.

Aquesta mètrica analitza la variabilitat en l'increment de la freqüència fonamental.

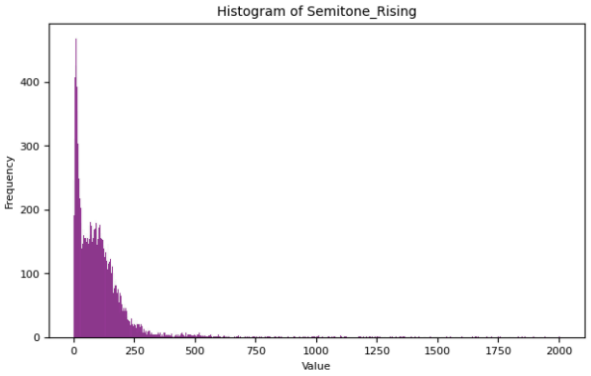


fig1.distribució semitone_rising

Aquesta distribució és un cas singular, ja que presenta dues distribucions en una mateixa variable, o el que és el mateix, és una distribució bimodal. Aquest tipus de distribució es deu a la presència de dos grups diferenciats dins del conjunt de dades. És probable que aquests grups estiguin associats al sexe, ja que els homes acostumen a tenir freqüències fonamentals més baixes que les dones a causa de les seves característiques biològiques.

Semitone_Mean (F0SemitoneFrom27.5Hz_sma3nz_meanFallingSlope)

Representa la mitjana de la pendent decreixent de les semitones de F0 calculades a partir de 27,5 Hz. Indica com de suau o pronunciat és el descens de la freqüència fonamental.

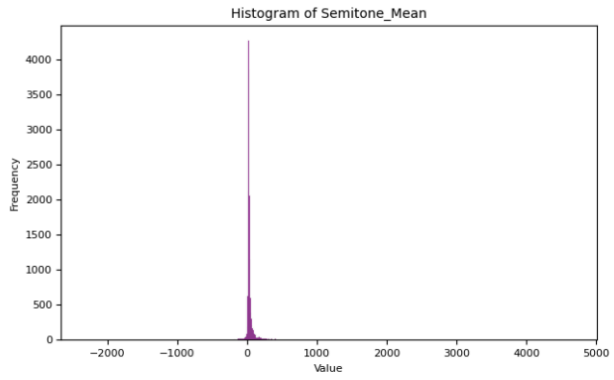


fig2.distribució semitone_mean

D'aquesta distibució es pot comentar la presència de grans outliers que se separen molt de la mitjana, cosa que ens indica la necessitat d'imputar outliers o fer alguna transformació.

1.3. Balanceig de classes

Per tal d'evitar que els models tendeixin a predir una categoria de la variable objectiu més que l'altra hem de comprovar que la classe estigui balancejada. Així doncs observem a la distribució de *Realornot*, especialment al conjunt de train.

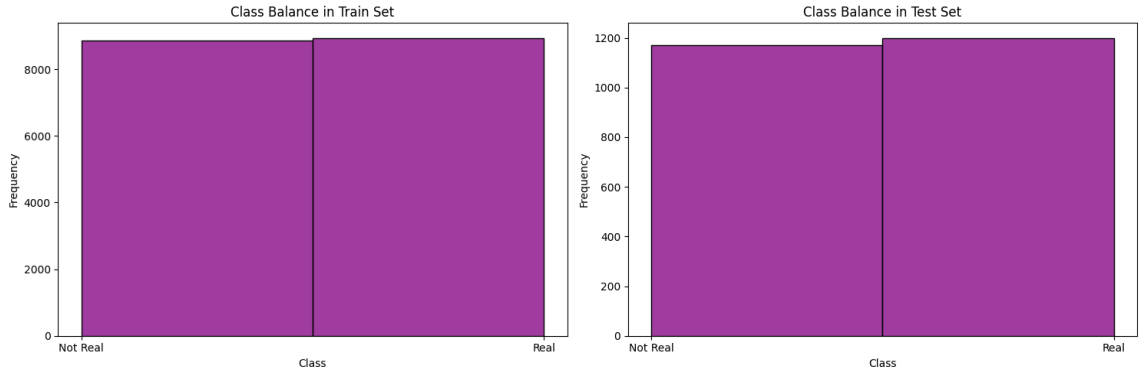


fig3.balanceig de classes

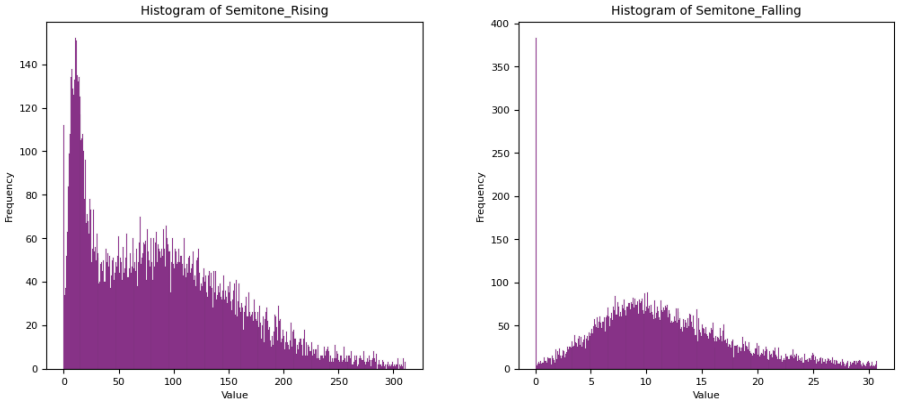
Com es pot comprovar, tant al train com al test la variable objectiu està balancejada, de manera que no ens hem de preocupar per realitzar mètodes de balanceig de les dades ja que teòricament no s'afegirà aquest esbiaix a les prediccions.

1.4. Tractament de Missings

A l'hora d'identificar missings veiem a les dades que no apareixen valors nuls, tot i això, volem identificar si els 0 poden ser en algun cas considerats com a missing. Així doncs, volem visualitzar les dues variables amb un percentatge de zeros més alt:

```
Unset
Semitone_Rising     148
Semitone_Mean         1
Semitone_Falling    455
```

Per fer-ho, com que apareixen valors molt alts a les distribucions inicials, eliminarem temporalment els outliers per tal d'identificar si aquests zeros s'haurien d'identificar com a missings.



Com es pot veure als histogrames, els valors iguals a zero suposen un desequilibri notable en les distribucions, cosa que ens podria indicar que aquestes dues variables sí contenen missings.

Així doncs, un cop identificats hem de decidir si eliminar-los o imputar-los. Així doncs, ja que tampoc tenim masses missings decidim imputar-los amb knn, utilitzant aquest mètode per ser més fiel a la realitat de les dades que si utilitzessim la mediana o mitjana directament.

Conjunt de test

Cal destacar que per tal d'evitar *Data Leakage* el particionat de les dades ha estat realitzat des de l'inici, de manera que el *knn imputer* només ha estat entrenat amb els conjunt de train i posteriorment aplicat al test.

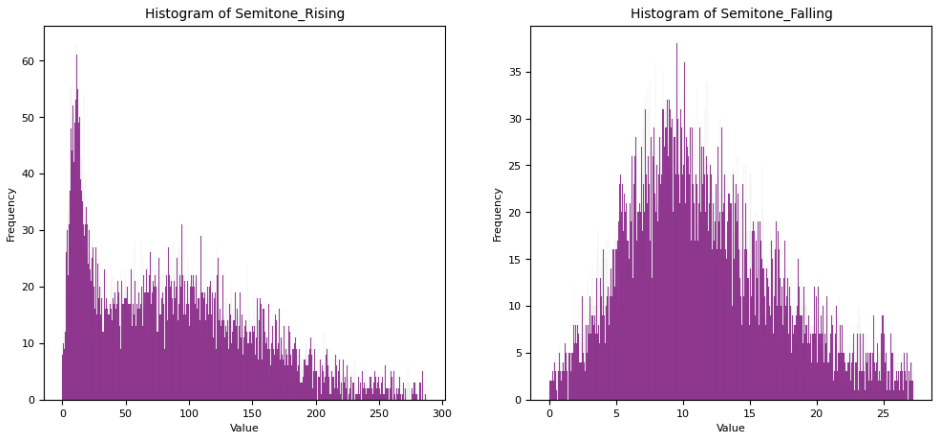


fig5.distribucions després d'imputar missings

1.5. Tractament d'outliers

Pel tractament d'outliers ens trobem en un cas delicat, ja que com podem veure en les distribucions de les variables hi ha vàries que contenen forces outliers. Tot i això, en el context de la pràctica, on estem tractant amb dades sonores, hem de tenir en compte que aquests outliers es poden deure a silencis naturals, alteracions dels models de ia, entre d'altres.

Per aquest motiu, al no disposar del coneixement tècnic necessari per avaluar la situació, s'ha decidit realitzar diverses proves amb els models, tant amb el tractament d'outliers com sense, de manera que podem determinar la seva rellevancia en la variable objectiu.

Tractament:

En les proves on sí s'han tractat outliers, s'ha decidit detectar outliers mitjançant els quartils exteriors de la distribució i imputar-los amb knn.

També es va considerar utilitzar Local Outlier Factor per tal de detectar els outliers, tot i que en ser un mètode basat en densitats tampoc permetia imputar-ne masses i donava força problemes a la hora d'imputar.

Conjunt de Test

Per evitar filtracions en el conjunt de test, a l'hora de tractar amb outliers utilitzem els límits dels quartils del train i a l'hora d'imputar amb knn només utilitzem un transform.

2. Preparació de variables

2.1. Anàlisis de variables categòriques

A l'hora de preparar les variables hem de tenir en compte que les categòriques estiguin equilibrades, ja que podrien afegir un esbiaix al model.

Així doncs, comprovem la seva proporció al conjunt d'entrenament:

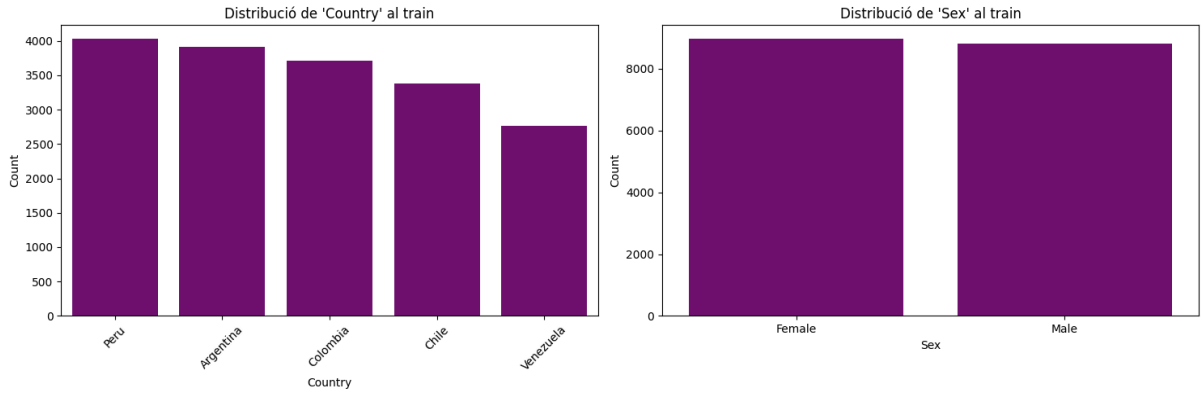


fig6.balanceig de categòriques

Com es pot veure a les gràfiques anteriors, les variables estan força equilibrades, tot i que en el cas de Country la diferència és més notable. Tot i això, abans d'arribar a cap conclusió primer hem d'analitzar com es troben distribuïdes en relació a la variable objectiu.

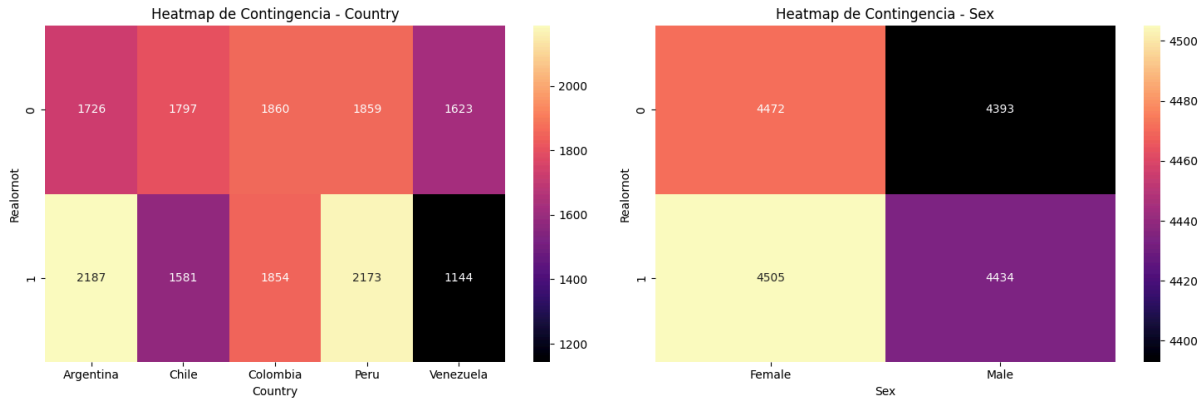


fig7.heatmap de contingència de les categòriques

Analitzant els heatmap veiem que en el cas de la variable Sex les classes es troben balancejades. Per la part del Country, veiem que hi ha petites diferències entre classes però no arriben a ser massa significatives respecte a la variable Realornot, tot i que en el cas de Venezuela es podria arribar a balancejar la classe.

Tot i això, degut a que tampoc és una diferència tant notable com per afegir un esbiaix significatiu al model respecte la poca influència de la variable, s'ha decidit no realitzar una tècnica de balanceig a la categoria.

2.2. Normalització de variables

Variables Categòriques

Per tal de poder tractar amb les variables categòriques als models, necessitem poder-les tractar com a distàncies.

Així doncs amb un label encoder s'han transformat aquestes variables a números d'un rang concret (en el cas de Country), que després hem reescalat mitjançant un MinMax-Scaler per tal de restringir el seu rang entre 0 i 1 i evitar possibles esbiaixos amb les distàncies.

En el cas de la variable Sex simplement s'ha binaritzat.

Variables Numèriques

Per tal de tractar amb les variables numèriques, volem que les nostres variables segueixin distribucions normals per evitar incompatibilitats amb els models, així que hem d'aplicar transformacions logarítmiques a les variables que segueixin una distribució log-normal.

Hem aplicat la transformació a les següents variables: Semitone_Rising, Semitone_Falling, mfcc2_stddev, mfcc3_stddev, Loudness_Mean] (a l'hora de tractar sense outliers s'ha aplicat la transformació a més variables, ja que tractant els outliers aplanàvem les distribucions).

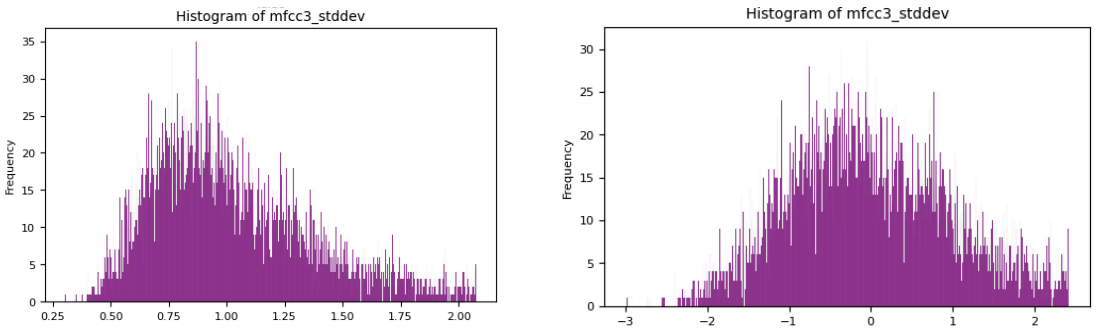


fig8.exemple d'abans i després d'aplicar una transformació logarítmica

Un cop normalitzades les dades (tant al train com al test), les hem d'escalar, de manera que s'utilitza un StandardScaler per tal de tenir mitjana 0 i desviació estàndard 1 a totes les distribucions.

Això ho fem tant en les dades amb outliers tractats com sense tractar (en el cas de tractament d'outliers podriem haver utilitzat un MinMax Scaler, tot i que s'ha vist experimentalment que no donava tants bons resultats). Tot i això sense tractament d'outliers s'ha prioritzat l'StandardScaler ja que el MinMax no tracta tant bé amb outliers i s'havia de considerar la possibilitat que hi fossin presents a les dades.

2.3. Eliminació de variables redundants

Variables Categòriques

Per tal d'eliminar variables categòriques redundants, s'ha decidit fer un chi-square per veure quines variables estàn relacionades amb la variable objectiu.

El chi-quadrat mesura la diferència entre les freqüències observades i les esperades, de manera que un valor baix indica poca diferència. Aquests són els resultats del test:

```
Unset
Sex: chi2=0.0024642795791689, p-value=0.9604080666805683
Country: chi2=175.2022615307704, p-value=7.993768458933562e-37
```

Sex: com que el valor p és molt alt (0.9604), no hi ha evidència suficient per rebutjar la hipòtesi nul·la. De manera que no hi ha una associació significativa entre la variable "Sex" i la variable "Realornot".

Country: com que el valor p és extremadament baix (7.993768458933562e-37), hi ha evidència suficient per rebutjar la hipòtesi nul·la. Cosa que significa que hi ha una associació significativa entre la variable "Country" i la variable "Realornot".

D'aquesta manera comprovem que la variable Country està relacionada amb la variable objectiu, de manera que l'afegirem al model.

Variables Numèriques

Per tal de veure les variables numèriques més relacionades s'ha realitzat una correlation matrix i un test anova, incloent la variable objectiu.

D'aquesta manera podem veure les variables més correlacionades amb la variable objectiu i així reduir el soroll.

```
Python
Selected features using ANOVA F-test: Index(['Loudness_Mean',
'SpectralFlux', 'mfcc1_mean', 'mfcc2_mean',
       'mfcc2_stddev', 'mfcc3_mean', 'mfcc3_stddev']
```

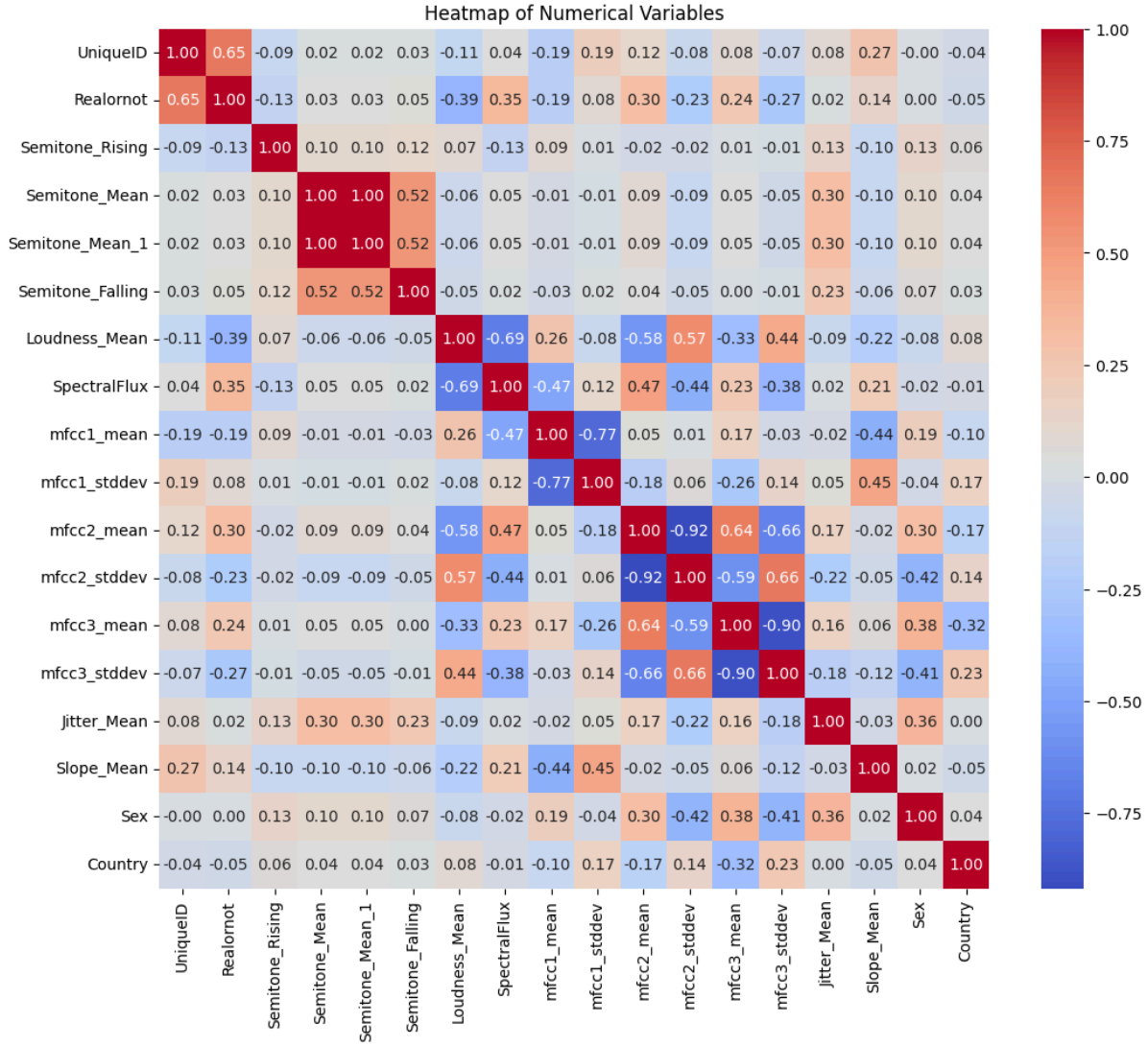


fig9.correlation matrix

Les variables seleccionades són Loudness_Mean, SpectralFlux, mfcc1_mean, mfcc2_mean, mfcc2_stddev, mfcc3_mean, mfcc3_stddev.

2.4. Reducció de dimensionalitat

Un cop ja tenim seleccionades les variables que utilitzarem, tenim l'opció d'aplicar-lis una tècnica de reducció de dimensionalitat abans de fer l'entrenament del model.

Tot i això, com que només hem escollit 6 variables numèriques, no s'ha considerat necessari l'ús d'un PCA o cap altra tècnica de reducció de dimensionalitat, ja que aquestes són beneficioses principalment en casos on hi ha una alta dimensionalitat i per tant les distàncies es veuen afectades a l'hora de ser considerades per un model.

Així doncs, en un cas on tampoc disposem de tantes dimensions, el fet de fer un PCA no ens aporta cap benefici a part de la pèrdua d'informació i explicabilitat.

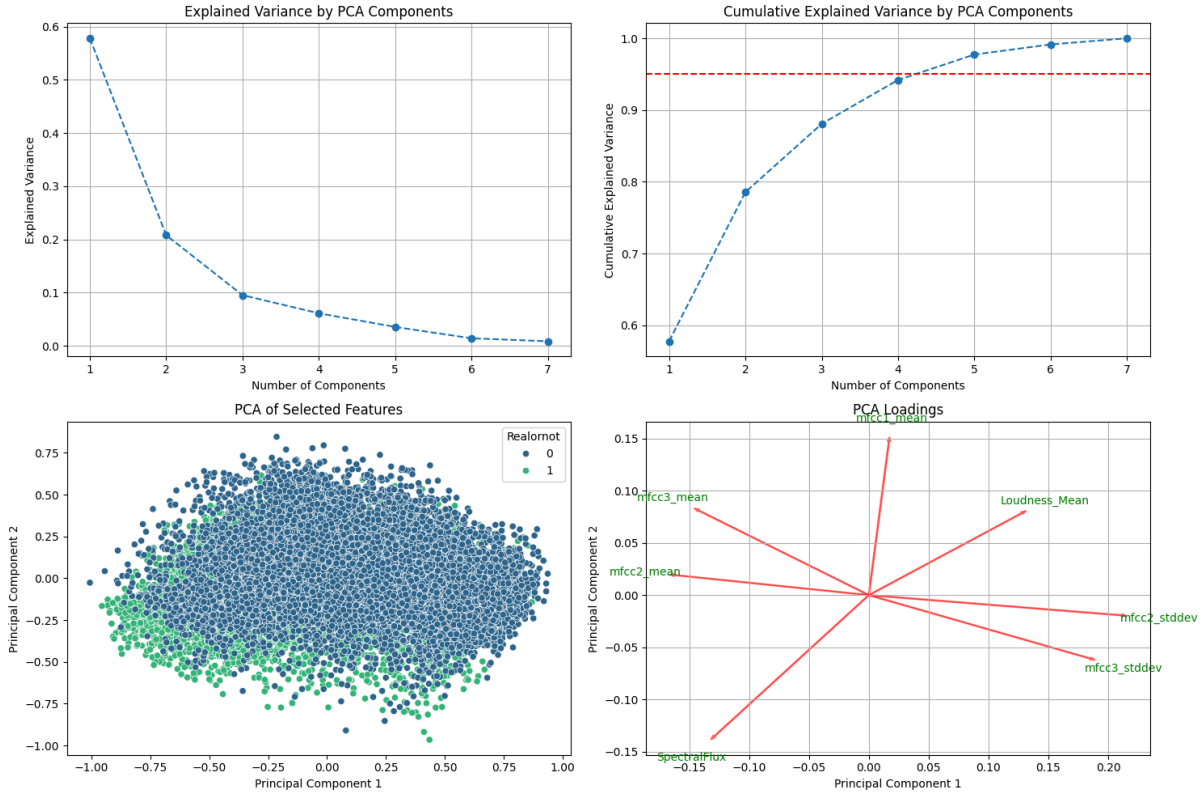


fig10. Gràfiques del PCA

Com podem veure als gràfics anteriors, podriem arribar a reduir la dimensionalitat en certa mesura dintre el conjunt de train. El número idoni maximitzant la variança seria de 5 components, cosa que es relaciona amb les similituds dels mfcc_mean i mfcc_sttdev.

Tot i això, tenint en compte que estem parlant de només 7 dimensions, com ja s'ha mencionat aquseta pèrdua d'informació no és compensativa pels guanys que podria arribar a aportar al model.

3. Definició de models

3.1. Definició de mètriques

Com que en teoria tenim les classes força balancejades, les mètriques que utilitzarem seràn F1-score i Accuracy.

F1-score:

Mesura la mitjana armònica entre el precision i el recall, de manera que podem observar si hi ha un equilibri entre aquestes dues mètriques, tot i que no reflexa la capacitat del model per manipular múltiples classes

Accuracy:

Mesura la proporció de prediccions correctes sobre el total de prediccions, és intuïtiva i interpretable tot i que pot ser enganyosa si hi ha classes desbalanceadas.

3.2. Motivació del primer model triat

El model que s'ha considerat prioritari ha sigut l'SVM, ja que la naturalesa del problema requereix un equilibri entre la gestió de dades (amb probablement relacions no lineals) i la capacitat de generalització.

Per aquest motiu l'SVM despunta sobre els altres, al ser KNN un model molt simple i un arbre de decisió més susceptible a overfitting. A més, l'SVM permet gestionar dades amb marges separadors que maximitzen la generalització, fent-lo especialment robust en situacions amb dades que presenten certa ambigüitat o que poden no estar clarament separades en l'espai de característiques, com és el cas.

Interpretabilitat

Respecte la interpretabilitat, l'SVM amb kernels no lineals és menys interpretable, però força més potent en dades amb separacions complexes com és el cas.

Gestió d'hiperparàmetres

Un altre aspecte fonamental és la capacitat de personalitzar el rendiment del model mitjançant l'ajust dels hiperparàmetres, permetent optimitzar el comportament del model segons les necessitats específiques del problema.

- **Valor de C:** Gestiona l'equilibri entre ampliar el marge de separació de l'hiperplà i reduir els errors de classificació en el conjunt de dades d'entrenament.
- **Kernel:** defineix la funció de transformació que l'SVM aplica per mapejar dades que no són linealment separables a un espai de dimensió superior, on aquestes dades es poden separar de manera més efectiva.
- **Gamma:** Determina l'abast de la influència d'un punt d'entrenament específic. Es tracta d'un paràmetre del kernel que regula la importància relativa dels punts propers en l'espai de característiques. Valors elevats de gamma poden portar a un overfitting del model, ja que es dóna massa pes als punts més propers.

Volum de dades i eficiència

El volum de dades juga un paper crucial. L'SVM és força bo per a conjunts de dades petits o moderats, ja que pot maximitzar la separabilitat amb un nombre limitat de mostres, oferint bons resultats amb menys dades disponibles.

3.3. Discussió dels hiperparàmetres

Valor de C	Kernel	Gamma
0.1	linear	0.01
1	rbf	0.1
100		1

C: aquest hiperparàmetre determina la intensitat de la regularització. Un valor alt de C redueix la regularització, fent que el model sigui més sensible als punts de dades específics, mentre que un valor baix afavoreix un model més senzill i amb major capacitat de generalització.

Kernel: aquest hiperparàmetre determina la funció utilitzada per transformar les dades d'entrada al nou espai de característiques.

- linear: Utilitza una frontera de decisió lineal.
- rbf: Utilitza una funció de base radial per capturar la complexitat no lineal.

Gamma: aquest hiperparàmetre controla la influència d'un punt de dades concret en la funció de decisió.

- Valors alts de gamma: cada punt de dades té una influència molt local, cosa que pot conduir a un model altament complex que tendeix a l'overfitting
- Valors baixos de gamma: els punts de dades tenen una influència més àmplia, resultant en una funció de decisió més suau i general

3.4 Anàlisis dels resultats

Els resultats s'han obtingut de fer un crossvalidation de 3 folds.

	C	gamma	kernel	f1_svm	accuracy_svm	f1_svm_no_outliers	accuracy_svm_no_outliers
8	100.0	1.00	rbf	0.706026	0.759275	0.412703	0.715008
5	1.0	1.00	rbf	0.701116	0.745363	0.461738	0.759275
7	100.0	0.10	rbf	0.703912	0.739039	0.384853	0.770236
2	0.1	1.00	rbf	0.684524	0.695616	0.552665	0.755059
6	100.0	0.01	rbf	0.679158	0.688027	0.523235	0.762226
4	1.0	0.10	rbf	0.680156	0.684654	0.545636	0.764755
3	1.0	0.01	rbf	0.660994	0.658094	0.509737	0.695616
1	0.1	0.10	rbf	0.666554	0.657673	0.568018	0.739039
0	0.1	0.01	rbf	0.666874	0.632378	0.593117	0.667369

Influència del kernel

El kernel RBF sembla oferir consistentment millors resultats d'accuracy en comparació amb el lineal. Això indica que la naturalesa del problema podria aprofitar les capacitats del kernel RBF per capturar relacions no lineals complexes de manera més efectiva.

Impacte del paràmetre C

Per al kernel RBF, valors intermedis de C ofereixen un bon equilibri entre accuracy i robustesa als outliers. Quan C és molt alta, l'f1 disminueix lleugerament en la majoria de casos, cosa que pot indicar un possible overfitting.

Impacte de gamma

Es pot observar que principalment valors de gamma baixos tendeixen a donar més bons resultats.

3.5 Resultat final

Per evaluar la millor combinació d'hiperparàmetres, hem de tornar a fer una predicció, però sobre el conjunt de test (Model amb tractament d'outliers).

```
Unset
Millors hiperparàmetres: {'C': 100, 'gamma': 1, 'kernel': 'rbf'}

Accuracy Test: 0.7592748735244519

F1 Test: 0.7670338637290902

Classification Test Report:
               precision    recall  f1-score   support

           0       0.77      0.73      0.75      1172
           1       0.75      0.78      0.77      1200

    accuracy                           0.76      2372
   macro avg       0.76      0.76      0.76      2372
weighted avg       0.76      0.76      0.76      2372
```

Per al kernel RBF, els millors hiperparàmetres obtinguts són: C = 100, gamma = 1, amb una accuracy en el test de 0.76 i un F1-score de 0.77. Això indica un rendiment acceptable del model però es poden destacar alguns aspectes

- **Classe 0:** Amb un F1-score de 0.75 i un recall de 0.70, el model té dificultats per identificar correctament els exemples d'aquesta classe. Això es deu a un nombre elevat de falsos positius, tot i tenir una precisió alta (0.80).
- **Classe 1:** El model té millor rendiment amb un F1-score de 0.78 i un recall elevat (0.83), capturant la majoria dels exemples correctament. No obstant, la precisió és més baixa (0.74), amb més falsos positius en aquesta classe.

Conclusió:

El model és consistent i separa bé les classes, però caldria millorar la identificació de la classe 0 reduint els falsos positius. Això es podria deure a un desbalanceig en la naturalesa de les dades o a la decisió de no balancejar la variable Country tot i que caldria fer més proves al respecte.

Cal destacar que s'ha seleccionat el model amb tractament d'outliers ja que tot i obtenir millors accuracy sense tractar-los, obtenim un f1 score molt més baix que indica un mal equilibri entre separació de classes.

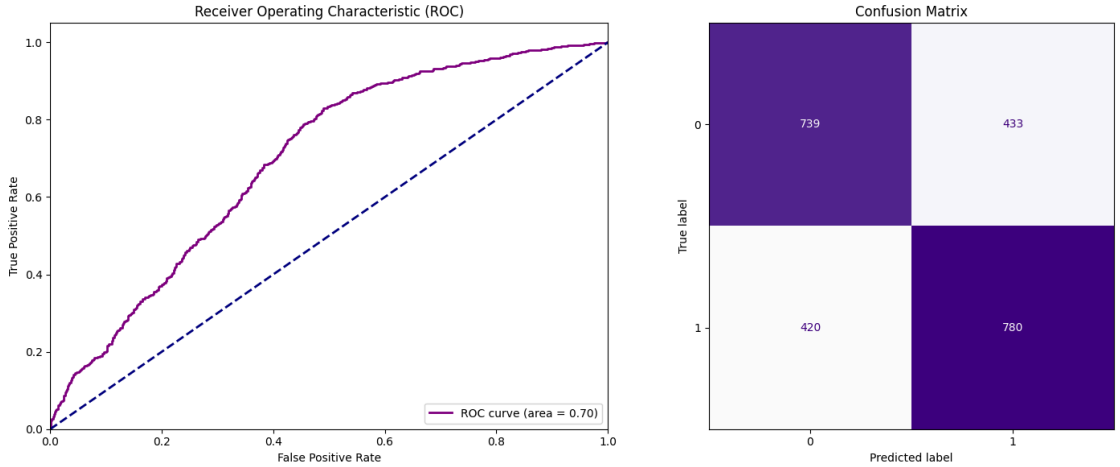


fig11.curva ROC i confussion matrix

3.6 Motivació del segon model triat

El model K-Nearest-Neighbors (KNN) pot ser una bona segona opció, especialment des d'una perspectiva d'explicabilitat.

KNN classifica les dades tenint en compte la proximitat dels veïns en l'espai de característiques. Aquest procés és fàcil de comprendre, ja que simplement busca els individus més propers i pren la decisió en funció d'ells. A més, no depèn de funcions matemàtiques complexes, a diferència de l'SVM, que utilitza kernels i conceptes com els marges de separació.

Quan el model classifica un individu en una classe concreta, es pot verificar que els veïns han influït en aquesta decisió, proporcionant així una explicació clara de per què es va prendre aquesta elecció.

L'hiperparàmetre 'k' té també una interpretació directa i permet ajustar el model. Valors baixos de 'k' fan que el model sigui més sensible a les dades locals, centrant-se en casos

individuals, mentre que valors més alts de 'k' suavitzen la classificació, tenint en compte més veïns per fer la predicció.

En resum, el model KNN seria una excel·lent segona opció, especialment si es busca un model fàcil d'interpretar i que permeti justificar les decisions preses basant-se directament en les dades. És una opció ideal per presentar els resultats a un públic menys tècnic.

```
Unset
Millors hiperparàmetres: {'metric': 'manhattan', 'n_neighbors': 11,
'weights': 'uniform'}

Accuracy Test: 0.7099494097807757

F1 Test: 0.7333333333333333

Classification Test Report:
              precision    recall  f1-score   support

           0       0.74      0.63      0.68      1172
           1       0.69      0.79      0.73      1200

    accuracy                           0.71      2372
   macro avg       0.71      0.71      0.71      2372
weighted avg       0.71      0.71      0.71      2372
```

	metric	n_neighbors	weights	f1_knn	accuracy_knn	f1_knn_no_outliers	accuracy_knn_no_outliers
8	euclidean	11	uniform	0.676316	0.714587	0.587333	0.730607
28	minkowski	11	uniform	0.676316	0.714587	0.587333	0.730607
29	minkowski	11	distance	0.676586	0.714165	0.584520	0.729764
9	euclidean	11	distance	0.676586	0.714165	0.584520	0.729764
19	manhattan	11	distance	0.678186	0.712057	0.559417	0.723862
24	minkowski	7	uniform	0.667929	0.710371	0.574400	0.721754
4	euclidean	7	uniform	0.667929	0.710371	0.574400	0.721754
16	manhattan	9	uniform	0.675500	0.710371	0.559040	0.731450
18	manhattan	11	uniform	0.678367	0.709949	0.561344	0.724283
26	minkowski	9	uniform	0.670552	0.709528	0.581016	0.722597
6	euclidean	9	uniform	0.670552	0.709528	0.581016	0.722597
25	minkowski	7	distance	0.667826	0.709106	0.572330	0.718803
27	minkowski	9	distance	0.670017	0.709106	0.579259	0.719646

Com es pot comprovar amb aquests resultats, el model no aporta tant bons resultats com en el cas del SVM, tot i que s'hi acosta bastant.

4. Model Escollit (SVM)

Com s'ha mencionat anteriorment, per la naturalesa del problema el model més adequat és l'SVM, ja que a part d'haver-ho comprovat experimentalment ofereix un equilibri entre la gestió de dades i la capacitat de generalització.

A més, havent realitzat la prova amb dades sense imputar i imputant, hem vist una lleugera millora quan no imputem les dades, de manera que probablement el problema requereix d'aquests outliers per a donar bons resultats.

Tot i això, en el cas del f1 score si no imputavem outliers aquest disminuia dràsticament, cosa que pot ser indicatiu d'overfitting i mal balanceig de classes, de manera que finalment s'ha escollit el SVM amb tractament d'outliers.

Limitacions

L'SVM no és eficient des del punt de vista computacional per a conjunts de dades grans, ja que el temps d'entrenament creix de manera exponencial a mesura que augmenta el nombre de mostres (degut a la complexitat combinatòria).

A més, els seus hiperparàmetres són abstractes, ja que en problemes no lineals o d'alta dimensionalitat la separació de classes és difícil d'interpretar.

Capacitats

En espais d'alta dimensionalitat en relació amb la mida de les dades, els SVM tenen un bon rendiment, ja que es concentren exclusivament en els vectors de suport en lloc de considerar tota la mostra.

A més l'SVM té com a objectiu maximitzar el marge entre les classes, la qual cosa contribueix a millorar la capacitat de generalització del model en noves dades. En el nostre cas, un marge òptim permet evitar classificacions errònies influenciades pel soroll present en les dades.

També hem de tenir en compte la seva variabilitat a l'hora d'escollir els hiperparàmetres i que té un bon rendiment amb mides de dades moderades.

5. Model Card

Model Details

Overview

Aquest model prediu si un àudio s'ha generat de forma artificial o ha estat enregistrat per una persona real. Està entrenat amb l'algoritme de Support Vector Machine (SVM) utilitzant un kernel rbf. El model s'ha optimitzat amb un grid search i els millors hiperparàmetres seleccionats són els següents:

- **C** = 0.1
- **gamma** = 0.01
- **kernel** = 'rbf'

El model està dissenyat exclusivament en un context educatiu i no es preté fer servir el model per a cap aplicació empresarial o comercial.

Version

- **Name:** Practica_antispoofing
- **Date:** 2024-12-28

Owners

Sergi Flores, sergi.flores@estudiantat.upc.edu

References

- Dataset original
- Article base de dades original

Considerations

Intended Users:

- Estudiants de Introducció a l'Aprenentatge Automàtic (IAA).
- Professors d'IAA.

Use Cases:

El propòsit del model és treballar amb algoritmes de classificació per predir si un àudio ha estat generat amb intel·ligència artificial o no en un context educatiu i no està dissenyat per a ser aplicat en la presa de decisions reals.

Limitations:

- **Escalabilitat:**

  L'SVM pot ser computacionalment costós amb grans conjunts de dades així que no es recomana utilitzar-lo amb grans conjunts de dades.

- **Explicabilitat limitada:**

  Els resultats del model poden ser difícils d'entendre per a usuaris no tècnics, ja que l'SVM empra tècniques complexes per establir marges i hiperplans en espais de característiques amb múltiples dimensions.

- **Sensibilitat als hiperparàmetres:**

  Tot i que els hiperparàmetres s'han optimitzat, valors incorrectes podrien causar overfitting o underfitting.

- **Dades no representatives:**

  Si les dades d'entrenament no són equilibrades o han estat mal preprocessades el model podria tenir un mal rendiment.

Validation Classification Report:

Class	Precision	Recall	F1-score	Support
0	0.77	0.73	0.75	1172
1	0.75	0.78	0.77	1200
Accuracy			0.76	2372
Macro avg	0.76	0.76	0.76	2372
Weighted avg	0.76	0.76	0.76	2372