

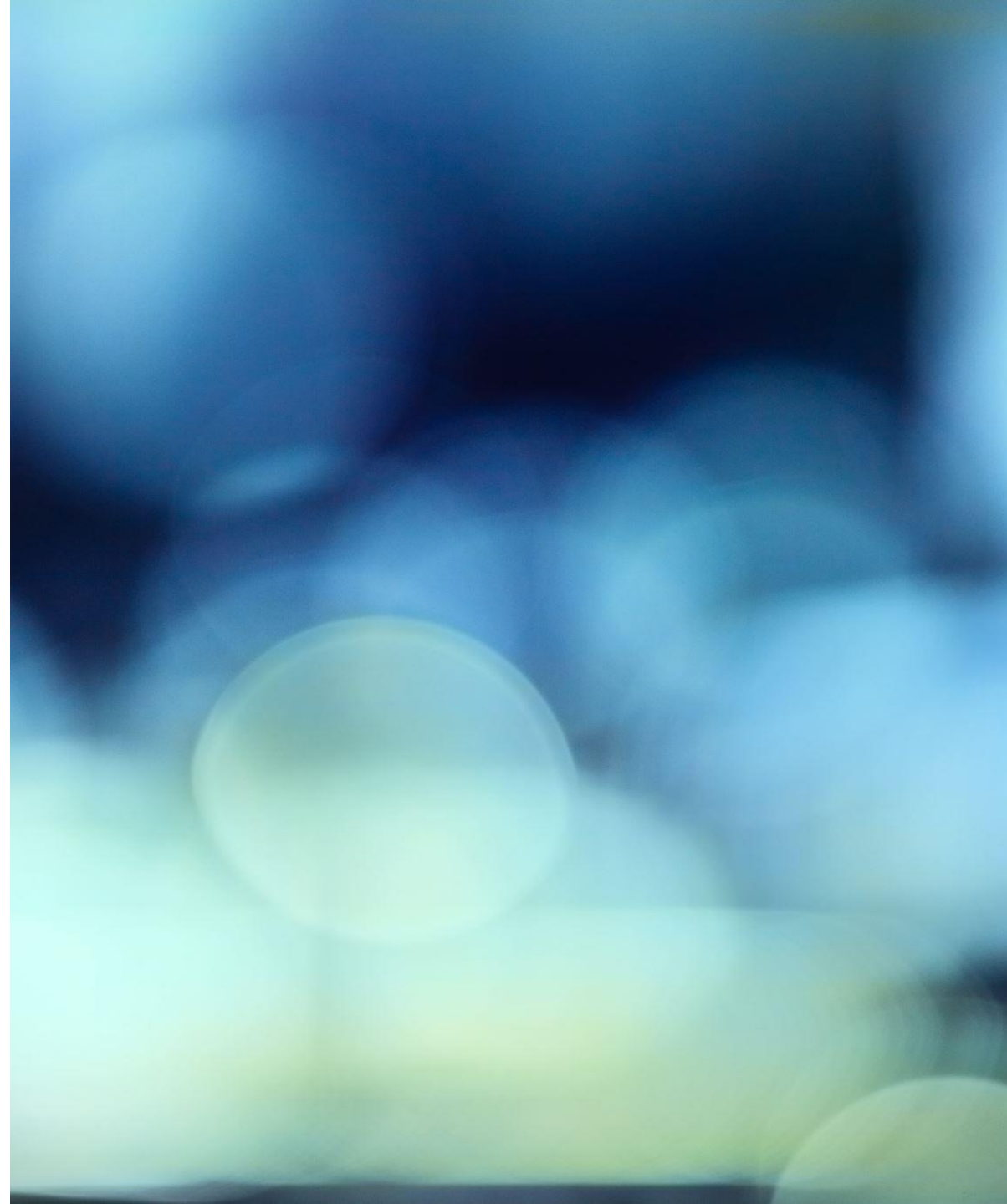
---

# CAS KAGGLE: SPOTIFY MUSIC DATABASE

Guillem Centelles Pavon

[https://github.com/Megaguille11/CasKaggle\\_Spotify](https://github.com/Megaguille11/CasKaggle_Spotify)

---



---

# INTRODUCCIÓ

- Model classificador sobre si cançó és acústica.
- Atributs:

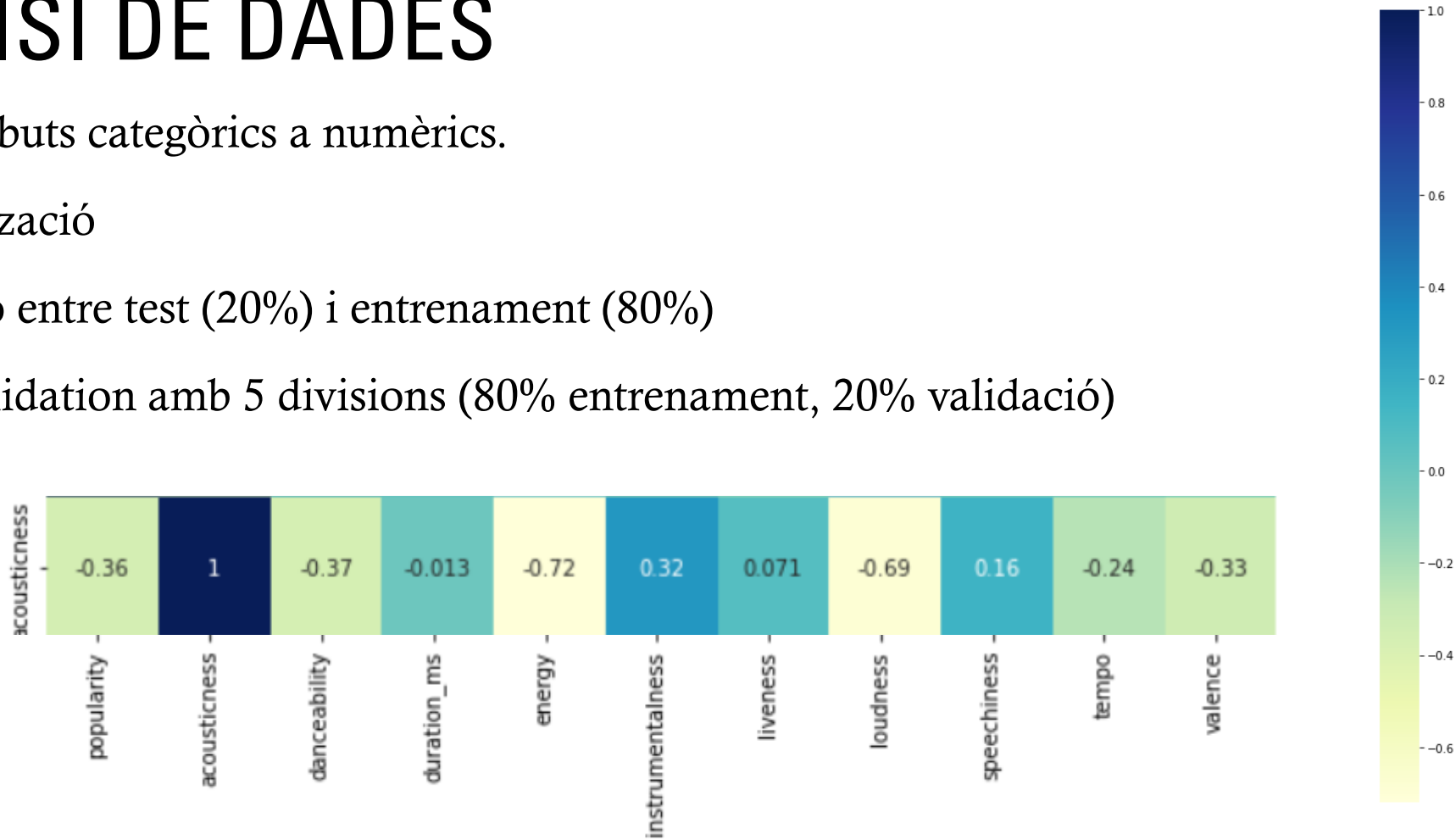
<i>genre</i>	<i>artist_name</i>	<i>track_name</i>	<i>track_id</i>
<i>popularity</i>	<b><i>acousticness</i></b>	<i>danceability</i>	<i>duration_ms</i>
<i>energy</i>	<i>instrumentalness</i>	<i>key</i>	<i>liveness</i>
<i>loudness</i>	<i>mode</i>	<i>speechiness</i>	<i>tempo</i>
<i>time_signatura</i>	<i>valence</i>		

---

---

# ANÀLISI DE DADES

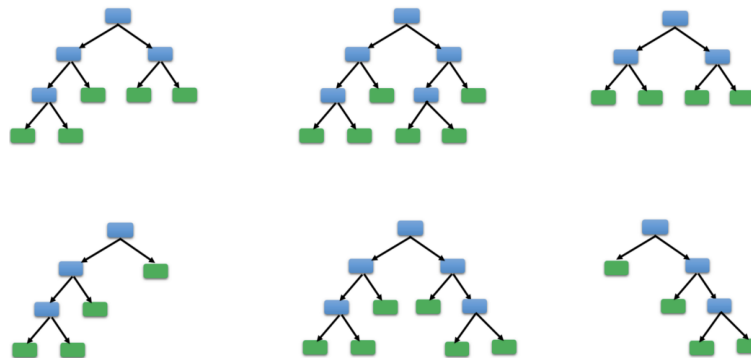
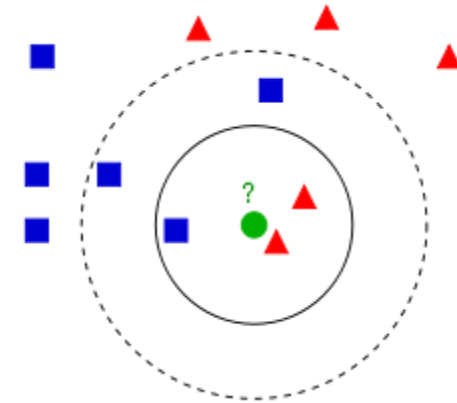
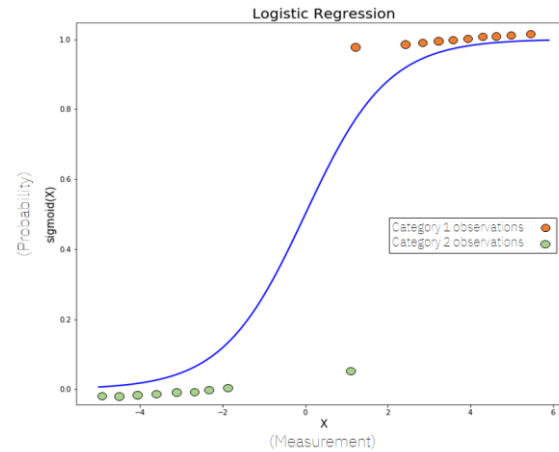
- Pas d'atributs categòrics a numèrics.
- Normalització
- Separació entre test (20%) i entrenament (80%)
- Cross-Validation amb 5 divisions (80% entrenament, 20% validació)



---

# MÈTODES D'APRENENTATGE

- Regressió Logística
- Arbre de decisió
- *Random Forest*  
(100 i 200 arbres)
- *K-Nearest Neighbors*  
( $K = 5$  i  $K = 10$ )



---

# RESULTATS

Model	Temps d'execució	Accuracy del test
Regressió Logística	0.4715 s	87.00%
Decision Tree	2.3490 s	87.77%
Random Forest (100 arbres)	36.7268 s	92.18%
Random Forest (200 arbres)	77.6756 s	92.25%
K-Nearest Neighbors (K = 5)	68.9664 s	87.74%
K-Nearest Neighbors (K = 10)	213.0671 s	87.70%

---

---

# CONCLUSIONS

- Classificadors funcionant correctament.
    - (Desviació estàndard entre 0.01 i 0.02 en *l'accuracy* de validació)
  - *Random Forest* ofereix els millors resultats.
  - K-NN molt costós pel poc rendiment.
  - Regressió logística ràpid de crear
-