# Investigating Hit Song Prediction

CARL DATH, CARL NYSTRÖMER, JAKOB VYTH & JOEL EKELÖF

Supervised by Bob Sturm

- We are replicating the work presented in Herremans et al. "Dance Hit Song Prediction", J. New Music Research 2014.
- Herremans et al. use machine learning to predict dance hit songs using features extracted from music recordings.
- They claim that their models "*can successfully predict if a dance song is going to be a top 10 hit versus a lower positioned dance song*".
- We want to know what their models are doing and how they can be fooled.
- We have replicated their experiments.
- We are going to try to break their model by manipulating noise and get the model to classify it as a hit.
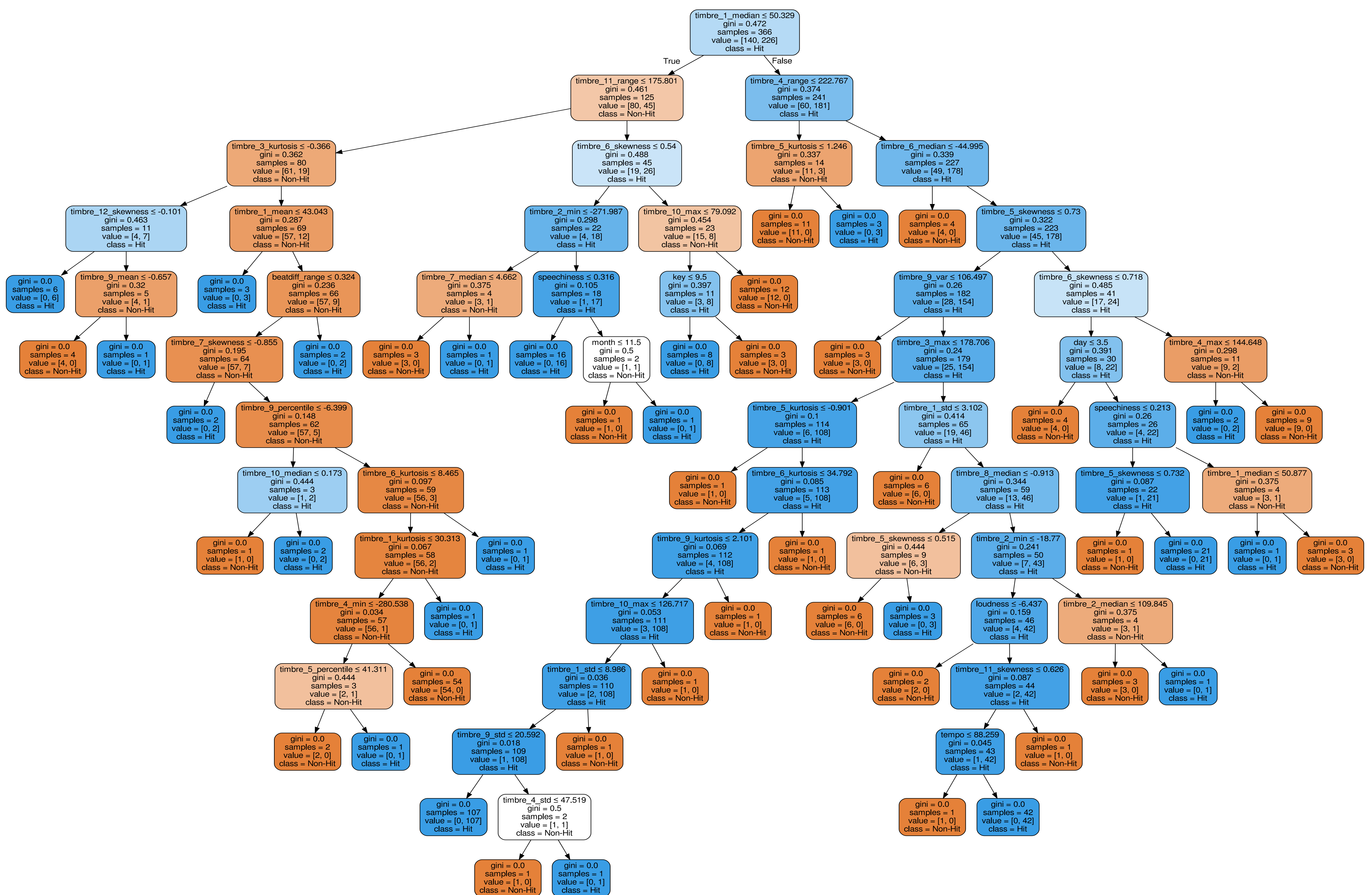
Decision tree trained on all features

Table 1: Results with 10-fold validation (accuracy) from Herremans et al.

Accuracy (%)	D1		D2		D3	
	-	FS	-	FS	-	FS
C4.5	*57.05*	*58.25*	*54.95*	*54.67*	*54.58*	*54.74*
RIPPER	*60.95*	*62.43*	*56.69*	*56.42*	*57.18*	*56.41*
Naive Bayes	**65**	**65**	*60.22*	*58.78*	59.57	59.18
Logistic regression	**64.65**	64	**62.64**	**60.6**	**60.12**	*59.75*
SVM (Polynomial)	**64.97**	**64.7**	**61.55**	**61.6**	**61.04**	**61.07**
SVM (RBF)	**64.7**	**64.63**	*59.8*	*59.89*	**60.8**	60.76

FS = feature selection, $p < 0.01$: italic, $p > 0.05$: bold, best: **bold**.