

Model Name: Decision Tree	
Pre-Processing	Data standardization
Experiments	Decision Tree <ul style="list-style-type: none"> • Base model • Gridsearch with base model • Oversampled • Gridsearch with base model Audio features analysis <ul style="list-style-type: none"> • Significance of audio features with bar chart • Recommended range for popular music
Prediction Generation	Overall accuracy by balanced accuracy
File Name	Group2_DecisionTree_Natapong_Sornprom.ipynb Group2_DecisionTree_Natapong_Sornprom.html
Environment details	Jupyter Notebook
Data File	train_and_validation.csv

Model Names: KMeans & AdaBoost	
Pre-Processing	Data standardization, bucketing, PCA
Experiments	K Means <ul style="list-style-type: none"> • K = 4 clustering and visualization • K = 2 clustering and visualization AdaBoost <ul style="list-style-type: none"> • Base model • Gridsearch with base model • Upsample with gridsearch • Undersample with gridsearch
Prediction Generation	Overall accuracy
File Name	Group2_AdaBoost_Tevfik_Gurkan.ipynb Group2_AdaBoost_Tevfik_Gurkan.html
Environment details	Jupyter Notebook
Data File	train_and_validation.csv

Model Name: Logistic Regression	
Pre-Processing	Data standardization
Experiments	<ul style="list-style-type: none"> • Logistic regression <ul style="list-style-type: none"> ○ Base model with cut off point popularity > 50 and < 50 ○ Base model with cut off point popularity > 70 and < 70 ○ Base model (popularity>70) with PCA ○ Base model (popularity >70) with SMOTE & Gridsearch ○ Smote & gridsearch model with PCA • Random Forest <ul style="list-style-type: none"> ○ Base model ○ Base model sample with smote ○ Base model with smote & gridsearch
Prediction Generation	Overall accuracy R2 AUC Recall, precision, f1 score
File Name	Group2_Logistic&RF_Zhenyu_Zhou.ipynb Group2_Logistic&RF_Zhenyu_Zhou.html
Environment details	Jupyter Notebook
Data File	data .csv (logistic regression) train_and_validation.csv (RF)

Model Name: K-nearest Neighbors	
Pre-Processing	Data standardization, Log-transformation
Experiments	K-nearest neighbors <ul style="list-style-type: none"> • K = [1,3,5,7,9,11,13,15,17,19,21,23,25,27,29] • Gridsearch with different K • Data standardization • Log-transformation
Prediction Generation	Overall accuracy Recall, precision, f1 score
File Name	Group2_K-nn_Bowen_Qiu.ipynb Group2_K-nn_Bowen_Qiu.html
Environment details	Jupyter Notebook
Data File	data .csv

Model Name: Support Vector Classification	
Pre-Processing	Data standardization, transform target var to multi classification
Experiments	SVC modeling <ul style="list-style-type: none"> • With and without stratified data • Manual Hyperparameter (stratified data) tuning <ul style="list-style-type: none"> ◦ Class Weight: None vs Balanced ◦ C tuning ◦ Gamma tuning • RBF vs Linear kernels(with stratified data) • Compare data split (non-stratify vs stratify)
Prediction Generation	Overall accuracy Recall, precision, f1 score
File Name	Group2_SVC_DevindraSawh.ipynb Group2_SVC_DevindraSawh.html
Environment details	Mainly Jupyter Notebook , some iterations via Google CoLab
Data File	data.csv