Machine Learning Canvas

PREDICTIONS	OBJECTIVES	DATA

GRO

End-user

Who will use the predictive system / who will be affected by it?

Prediction system will be used by Music Label to find potentially popular songs. Authors of these songs are affected by this prediction system.

Value proposition

What are we trying to do for the system's users? (e.g. spend less time on X, increase Y...)

Signing the contract with songs author's at the earlies stages of their career results in higher future returns for Music Label and faster promotion for the authors.

Data sources

Where do/can we get data from? (internal database, 3rd party API, etc.)

internal database, spotify api

E SF

Problem

Question to predict answers to (on behalf of user)

How popular is this song going to be?

Input (i.e. question "parameter")

Information about song

Possible outputs (i.e. "answers")

estimated popularity in range from 0 to 100

Type of problem (e.g. classification, regression, recommendation...)

regression

Baseline: simple, alternative way of making predictions (e.g. manual rules)

manual rules

Performance evaluation

Domain-specific / bottom-line metrics for monitoring performance in production

Hit rate - ratio of close predictions and overperforming authors over total predictions

Prediction accuracy metrics (e.g. MSE if regression; % accuracy, #FP for classification)

MSE
R2

Offline performance evaluation method (e.g. cross-validation or simple training/test split)

Cross-validation is the main evaluation method.

Data preparation

How do we get training data (inputs, and outputs if supervised learning)? How many data points?

All data is gathered from spotify api

Input features (extracted from data sources). If too many, list types of features and mention key ones.

song genres: provided as a list of labels
available markets: provided as a list of labels
information about artist: number of followers
information about album: release date, number of tracks, album_name, ...
information about the track: loudness, danceability, energy, valence, ...

GRAT

Using predictions

When do we make predictions and how many?

What is the time constraint for making those predictions?

How do we use predictions and confidence values?

Learning models

When do we create/update models? With which data / how much?

What is the time constraint for creating a model?

Criteria for deploying model (e.g. minimum performance value — absolute, relative to baseline or to previous model)

Reset Form

Machine Learning Canvas v0.1