Consider a home price prediction model

Encoders --> Aggregators --> Heads

name: kitchen photo 1
type: Image

name: bedroom photo 1
type: Image

name: bedroom photo 2
type: Image

name: zipcode
type: Categorical

name: living area
type: Numerical

name: lot size
type: Numerical

name: description
type: Text

normalizations,
scalings,
transforms,
augmentations

Image Encoder
trunk: ResNet50
pool: avg
additional dense: [256, 128]

Aggregator
additional dense: [16]
name: size_related_features

NLP Encoder
trunk: ALBERT

kitchen_photo_1_encoding
shape=(None, 128)

bedroom_photo_1_encoding
shape=(None, 128)

bedroom_photo_2_encoding
shape=(None, 128)

size_feature_encoding
shape=(None, 16)

description_encoding
shape=(None, 128)

Aggregator
additional dense: [512]
name:all_features

Regressor
Losses: MSE
Metrics: MSE, MAPE

Classifier
Losses: CROSSENTROPY
Metrics: AUC, PRAUC

Regressor
Losses: L1

Classifier
Losses: CROSSENTROPY
Metrics: mAP

target transforms

price
(main target)
type: Numerical

city
(auxiliary target)
Type: Categorical

swimming pools,
granite counter top
(auxiliary target)
Type: BoundingBox