Machine Learning Canvas

	PREDICTIONS	OBJECTIVES	DATA
GRO	**End-user** Who will use the predictive system / who will be affected by it? Sellers, marketplace and customers	**Value proposition** What are we trying to do for the system's users? (e.g. spend less time on X, increase Y...) Increase the accuracy of filling in product information. Spend less time filling in information.	**Data sources** Where do/can we get data from? (internal database, 3rd party API, etc.) Public dataset
E SF	**Problem** Question to predict answers to (on behalf of user) Sellers set up an incorrect product category. Input (i.e. question "parameter") information about product Possible outputs (i.e. "answers") category Type of problem (e.g. classification, regression, recommendation...) classification Baseline: simple, alternative way of making predictions (e.g. manual rules) logistic regression, decision trees, boosting	**Performance evaluation** Domain-specific / bottom-line metrics for monitoring performance in production accuracy weighted, precision, recall, f1-score, confusion matrix Prediction accuracy metrics (e.g. MSE if regression; % accuracy, #FP for classification) - % accuracy - F-measure - #FP - F1-score Offline performance evaluation method (e.g. cross-validation or simple training/test split) cross-validation	**Data preparation** How do we get training data (inputs, and outputs if supervised learning)? How many data points? Kaggle dataset. Encode categorical data. Split data to label and Xs data. Label is category. So we predict some label by some Xs. 464433 data points. Input features (extracted from data sources). If too many, list types of features and mention key ones. itemid, shopid, item_name, item_description, item_variation, price, stock, category, cb_option, is_preferred, sold_count, item_creation_date
GRAT	**Using predictions** When do we make predictions and how many? While filling out product information. One category. What is the time constraint for making those predictions? During the course of filling out the product card. How do we use predictions and confidence values? Tell the user which category to use when filling out a product.	**Learning models** When do we create/update models? With which data / how much? Updating the model after integration can take place manually when a certain volume of new items is accumulated. Usually each 3 months. What is the time constraint for creating a model? 6 hours Criteria for deploying model (e.g. minimum performance value — absolute, relative to baseline or to previous model) Baseline on all metrics by at least	

Reset Form

Machine Learning Canvas v0.1