


THE MACHINE LEARNING CANVAS

Iteration: .

Designed for:

Designed by:

Date:

PREDICTION TASK  Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation? Identificação e classificação da qualidade de vinhos tintos e brancos.	DECISIONS  How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that. Decidir se um vinho é considerado "bom" ou "ruim" com base em suas características físico-químicas.	VALUE PROPOSITION  Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces. Fornecer aos produtores de vinho uma ferramenta automatizada para avaliar a qualidade de seus produtos com base em características objetivas.	DATA COLLECTION  Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes. Coleta de dados sobre características físico-químicas de vinhos tintos e brancos.	DATA SOURCES  Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc. <i>Origem dos dados: "Modeling wine preferences by data mining from physicochemical properties" de P. Cortez et al. (2009).</i>
IMPACT SIMULATION  Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct decisions? <u>Fairness constraint</u> ? Simular o impacto do modelo de classificação na avaliação da qualidade dos vinhos e fornecer informações sobre o desempenho do modelo.	MAKING PREDICTIONS  When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target? Utilização de técnicas de Machine Learning, como Random Forest, XGBoost e Regressão Logística, para fazer previsões sobre a qualidade dos vinhos.		BUILDING MODELS  How many prod models are needed? When would we update? Time available for this (including featurization and analysis)? Treinamento de modelos de classificação usando os algoritmos de Random Forest, XGBoost e Regressão Logística.	FEATURES  Input representations available at prediction time, extracted from raw data sources. Variáveis preditoras: <ul style="list-style-type: none">Acidez fixa: medida da concentração de ácidos fixos no vinho.Acidez volátil: medida da concentração de ácidos voláteis no vinho.Ácido cítrico: medida da concentração de ácido cítrico no vinho.Açúcar residual: quantidade de açúcar residual no vinho.Cloretos: quantidade de cloretos presentes no vinho.Dióxido de enxofre livre: quantidade de dióxido de enxofre livre no vinho.Dióxido de enxofre total: quantidade total de dióxido de enxofre no vinho.Densidade: medida da densidade do vinho.pH: medida do nível de acidez ou alcalinidade do vinho.Sulfatos: quantidade de sulfatos presentes no vinho.Álcool: teor alcoólico do vinho. Variável alvo: <ul style="list-style-type: none">Qualidade: avaliação da qualidade do vinho, representada por uma escala numérica de 0 a 10.
	MONITORING  Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?	Monitoramento contínuo do desempenho do modelo de machine learning e atualização dos dados de treinamento para manter a precisão das previsões.		



ONLINE COURSE

Master the Machine Learning Canvas

Learn a step-by-step process to get to a complete and detailed Machine Learning Canvas. This will help you...

- Validate the feasibility of your ML use case ideas.
- Boost collaboration within your team.
- Anticipate issues that would otherwise come up during implementation or in production.

More details at ownml.co/plan

Miro

Brainstorming MLC

Share

CHURN

Prediction Task	Decisions	Value Proposition	Data Collection	Data Sources
ENTRYPY: customer WHAT FOR: subscription to and use > 1 mo. OUTCOMES: <ul style="list-style-type: none">• Churn (Predict churn) > 90%• Renewal (Repeat) > 90%	<ul style="list-style-type: none">• Filter out customers due to revenue > 1 mo. wait for 12 mo. or in limited set• Filter out customers whose churn proba > 90%• Predict if customer can be retained (repeat system)• Filter out customers whose retention proba < 90%• Sort by desired churn proba * yearly_revenue * retention_prob	OBJECTIVES: reduce churn rate among OnMML subscribers; improve repeat rate of retention efforts by understanding why customers may churn. WORKFLOW: Every month, get list of subscribers likely to churn (i.e. > 90% > 90%). Target top 20 cluster according to prod optimization, and retention response in 70% discount to each cluster. QUESTIONS to answer: <ul style="list-style-type: none">• Who are the most important and most fragile customers?• Who are we losing, among them? (to be addressed with a separate system)• Which activities are shared among churners? What drives churn problems? Are there patterns that could help us better retention response?	INITIAL: roughstart taken at time T-1. 10% of customers due to revenue > 1 mo. later - not used First 10, observe relevant: 1.5K / 10 x 1 mo. > 90% customers. CONTINUOUS: Hidden test at time T-10% of all customers due to revenue at T-1, selected at random. > 90% customers: Hidden devices: don't send any mail. Observed customers will be added to test set.	<ul style="list-style-type: none">• Drops for payment info (basic churn/renewal)• Highfidelity for OnMML platform analytics• Connect to the mail response• Contacts for professional info, industry, risk, company size• Available for custom info (average, notes, ...)
Impact Simulation	Making Predictions	Building Models	Features	Live Monitoring
TEST SET: at time T, we observed customer since T-1; these are all customers due to revenue at time T-1 and not targeted (< 90%) GAIN vs COSTS: during targeted customers: <ul style="list-style-type: none">• TP provides gain of subscription_value * 90% * retention_prob• FP incurs cost of subscription_value * 10% DEVELOPMENT: if total gain > 0, when using number of TP and FP in test set and applying selfless select	<ul style="list-style-type: none">• Every month, for all customers not filtered out by application, and customers in limited set• Make predictions for all of them (i.e. 2000-1000 = 1000)• Get explanations for targeted customers (20)• All inner rules and over-ought retention_prob• Directly stored in database	FREQUENCY: 1 prediction per month, for all customers in limited set		

machinelearningcanvas.com by Louis Dorard, Ph.D.

28%