The Machine Learning Canvas (v0.4)

Designed for: **Abnormal Detection Task** Designed by: Jia-Ying Lin Date: 2017/09/04 Iteration: 1 .

ML task

Input, output to predict, type of problem.

Type of problem: Abnormal Detection

Input為紅酒成份資料

Output為normal或abnotmal

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

預測結果品質不在正常範圍的紅酒,交由品酒師做進一步確認

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

由紅酒成份資料預測品質是否不在正常範圍

Data Sources

Which raw data sources can we use (internal and external)?

UC Irvine Machine Learning Repository: Wine Quality Data Set

Collecting Data

How do we get new data to learn from ?

先將quality介於4-8之間的資料設為正常,其餘的資料設為異常

從正常資料中取出80%作為訓練檔,剩下的20%與異常資料合併成測試檔

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

每釀出一桶紅酒即做預測

目標：快速檢驗出品質不在正常範圍的紅酒

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

將預測為正常的紅酒交由品酒師驗證, 錯誤率必須低於1%

Features

Input representations extracted from raw data sources.

1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH
10 - sulphates
11 - alcohol
12 - quality (score between 0 and 10)

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

利用品質正常的紅酒成份資料作為訓練檔

用One-Class SVM演算法建立預測模型

Live Evaluation and Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

每天抽查5%預測為正常的紅酒交由品酒師驗證, 錯誤率必須低於1%