The Machine Learning Canvas (v0.4)

Designed for: Diabetes Prediction | Designed by: Boya Sun | Date: 04/23/2020 | Iteration: 1

Decisions

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

Hospitals:
- Filter out 25% randomly (hold-out set).
- Select patients who have the highest risk for diabetes and suggest them take a diabetes test.

ML task

Input, output to predict, type of problem.

< Does this patient have diabetes? >
- Classification task

Value Propositions

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

Build a system to predict the onset of diabetes based on diagnostic measures:

Hospitals can use this system to quickly evaluate patients' risk for diabetes by entering patients' diagnostic measures and suggest those with higher risk should take a diabetes test.

The end goal is to find out those who have highest risk for diabetes and have them tested as soon as possible.

Data Sources

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

Open Data from:
- The National Institute of Diabetes and Digestive and Kidney Diseases
- Hospitals

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Every month request data on:
- Diagnostic measures
- Diabetes test results

Making Predictions

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

Every month we make predictions for new patients for the risk for diabetes (using all diagnostic measures except for diabetes outcome).

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Test on labeled data, and compute:
- F1 score
- precision
- recall
- ROC AUC

Features

Input representations extracted from raw data sources.

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- Blood Pressure
- Skin Thickness
- Insulin
- BMI: Body mass index
- Diabetes Pedigree Function
- Age

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?

Update model every quarter (with new data available).

Live Evaluation and Monitoring

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

- Compare diabetes outcome to prediction on hold-out set.
- Decrease late diabetes diagnosis rate.

Licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Any feedback or suggestions? Email me at louis@louisdorard.com