The Machine Learning Canvas (v0.4)

Decisions

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

- Predict whether a patient will develop diabetes in the future
- The model can be dynamic that chances every time the patient gets physical exams and updates the records, so we can see the trend of the chances.
- The model output can be provided to doctors to make decision on whether to give some early preventative care.

ML task

Input, output to predict, type of problem.

- Problem: Classification
- Input: multiple features with discrete and continuous numeric data
- Output: the probability to develop heart disease/ or binary classification

Value Propositions

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

- Build a predictive model to predict whether a patient is prone to developing diabetes. This model can be used as a supporting tool for doctors to take proactive interventions.

Data Sources

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

- Kaggle: https://www.kaggle.com/uciml/pima-indians-diabetes-database
- The data source can be from internal medical records of healthcare institutions.

Collecting Data

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

- Continuous inflow of new patient data or updated data from old patients.

Making Predictions

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

- After we have a trained model with the precision and recall that we think is risk-controlled and cost-efficient, we can implement the model and use it to predict on new data.
- Log transformation on Insulin and DiabetesPedigreeFunction
- Scaling the features as their magnitude are quite different.

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

- Confusion matrix
- ROC/AUC
- False negative rate here is more important as we want to prevent such serious diseases, but we would like to discuss with business stakeholders to confirm, because we want to take proactive interventions for patients who have high chance to develop the disease

Features

Input representations extracted from raw data sources.

- Age
- Medical metrics such as Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction

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?

- The model can be updated when new technologies or medical intervention methods are invented
- When new findings in diabetes appear, we can potentially add new features in the model
- If more demographic information can be collected, we can add these in as well.

Live Evaluation and Monitoring

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

- Confusion matrix
- ROC/AUC

Download the free guide to the Machine Learning Canvas!