











<h3>Decisions</h3>  <p>How are predictions used to make decisions that provide the proposed value to the end-user?</p> <ul style="list-style-type: none"><li>- Establish early warning for doctors to take actions for preventive care of diabetes disease prior to diagnosis.</li><li>-Alert patients to improve the health metrics that is related to the diabetes such as BMI etc.</li></ul>	<h3>ML task</h3>  <p>Input, output to predict, type of problem.</p> <p>Input: Numerical Patient Data</p> <p>Output: Diabetes Diagnostic Result (1- Positive, 0- Negative)</p> <p>Problem type: Classification</p>	<h3>Value Propositions</h3>  <p>What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?</p> <p>What? Accurately predict diabetes diagnostics based on patient's health metrics data to enable proactive interventions and improve recover.</p> <p>Why?</p> <p>Enhance diabetes diagnostic accuracy rate with prediction.</p>	<h3>Data Sources</h3>  <p>Which raw data sources can we use (internal and external)?</p> <ul style="list-style-type: none"><li>-Internal: Current patient health metrics data collected by hospital</li><li>-External: Existing diabetes diagnostic dataset from other research institute</li></ul>	<h3>Collecting Data</h3>  <p>How do we get new data to learn from (inputs and outputs)?</p> <p>Inputs: Patient health information and diabetes diagnostic provided by National Institute of Diabetes and Digestive and Kidney Diseases.</p> <p>Outputs: Labeled predicted diabetes diagnostic data.</p>
<h3>Making Predictions</h3>  <p>When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?</p> <p>When the model accuracy is not as desired on new predictions, we can revisit the model and incorporate new dataset to train the model or include new features that might contribute to reducing noise in the model.</p> <p>Monthly evaluation of the model accuracy is necessary and appropriate.</p>	<h3>Offline Evaluation</h3>  <p>Methods and metrics to evaluate the system before deployment.</p> <ul style="list-style-type: none"><li>-Validation on new incoming patient's diagnostic data if available.</li><li>-Test model on the desired test proportion of the entire dataset.</li></ul>	<p>Identify key health indicators that are likely associated with diabetes diagnostic.</p> <p>Who?</p> <ul style="list-style-type: none"><li>-Patients</li><li>-Doctors / Healthcare Organizations</li></ul>	<h3>Features</h3>  <p>Input representations extracted from raw data sources.</p> <ol style="list-style-type: none"><li>1. Pregnancies: Number of times pregnant</li><li>2. Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test</li><li>3. Blood Pressure: Diastolic blood pressure (mm Hg)</li><li>4. Skin Thickness: Triceps skin fold thickness (mm)</li><li>5. Insulin: 2-Hour serum insulin (mu U/ml)</li><li>6. BMI: Body mass index (weight in kg/ (height in m) ^2)</li><li>7. Diabetes Pedigree Function: Diabetes pedigree function</li><li>8. Age: Age (years)</li></ol>	<h3>Building Models</h3>  <p>When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?</p> <p>Model can be trained based on new data collected from patients on a monthly basis to improve the model.</p>
	<h3>Live Evaluation and Monitoring</h3> <p>Methods and metrics to evaluate the system after deployment, and to quantify value creation.</p>	<ul style="list-style-type: none"><li>-Compare model prediction result with diabetes diagnostic test result to see the accuracy.</li><li>-Conduct prescriptive analysis in the long term to monitor if the proactive medical treatment is affecting the diabetes diagnostic.</li></ul> 		



**Any feedback or suggestions? Email me at [louis@louisdorard.com](mailto:louis@louisdorard.com)**