










<b>ML task</b>  Input, output to predict, type of problem.  Type of problem : Classification  Input : iris的sepal(萼片)長寬和petal(花瓣)長寬  Output : class of iris plant.	<b>Decisions</b>  How are predictions used to make decisions that provide the proposed value to the end-user?  利用既有的Iris的sepal(萼片)長寬和petal(花瓣)長寬建立種類的分類模型, 有新的Iris資料進來可以預測出類別	<b>Value Propositions</b>  What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?  由iris的sepal(萼片)長寬和petal(花瓣)長寬來分類iris的種類	<b>Data Sources</b>  Which raw data sources can we use (internal and external)?  UC Irvine Machine Learning Repository: <a href="#">Iris Data Set</a>	<b>Collecting Data</b>  How do we get new data to learn from ?  從iris.data中各iris類別分別取80%的資料作為訓練檔  各iris類別剩下20%資料作為測試檔
<b>Making Predictions</b>  When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?  每一次新的田野調查後, 就將蒐集到的Iris資料用模型資料分類  目標: 快速辨別Iris種類, 以利研究建檔	<b>Offline Evaluation</b>  Methods and metrics to evaluate the system before deployment.  利用既有的Iris的sepal(萼片)長寬和petal(花瓣)長寬建立種類的分類模型  再用有類別標籤的Iris資料作預測, 錯誤率需小於1%		<b>Features</b>  Input representations extracted from raw data sources.  1. sepal length in cm 2. sepal width in cm 3. petal length in cm 4. petal width in cm 5. class: -- Iris Setosa -- Iris Versicolour -- Iris Virginica	<b>Building Models</b>  When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?  將iris作為訓練檔  用knn演算法建立預測模型
	<b>Live Evaluation and Monitoring</b> Methods and metrics to evaluate the system after deployment, and to quantify value creation.	每一次新的田野調查後, 就將蒐集到的Iris資料用模型資料分類並抽查5%的資料交由植物學家判斷, 錯誤率需小於1%	