Installation

- Language requirement:
  - python 2.7
  - numpy>=1.10.0
  - matplotlib>=1.5.0
  - pandas>=0.17.0
  - scikit-learn>=0.17 (different syntaxes for v0.17 and v0.18)

- Model is that participants develop on their own machine, then upload their code to the RAMP platform where it is evaluated

- ➔ you need to be able to run with numpy, sci-kit learn, matplot lib etc...

- ➔ we recommend to install Anaconda on your laptop
  https://www.continuum.io/downloads (with python 2.7)
    - then run app Navigator...navigated to notebook .ipynb

- Otherwise, you can use the SFT ML installation on any CVMFS cluster (e.g. lxplus)

  source /cvmfs/sft.cern.ch/lcg/views/LCG_85swan3/x86_64-slc6-gcc49-opt/setup.(c)sh

- Quick test:

  Python

  >>> import matplotlib.pyplot as plt
  >>> from sklearn import linear_model
  >>> import numpy as np
Scikit-learn 17 vs 18

- Update to scikit-learn 0.18: in terminal, conda update scikit-learn (takes a while), then test:
  ```python
  >>> import sklearn
  >>> print sklearn.__version__
  0.18.1
  ```
- OR
- user_test_submission.py requires sklearn 0.18 (0.17 complains about ShuffleSplit)
- following changes user_test_submission.py
- at the beginning
  ```
  # skl 0.18
  # from sklearn.model_selection import ShuffleSplit
  # => replace with skl 0.17
  from sklearn.cross_validation import ShuffleSplit
  ```
- and later in def get_cv:
  ```
  # v0.18
  # event_cv = ShuffleSplit(
  #     n_splits=1, test_size=0.5, random_state=57)
  # for train_event_is, test_event_is in event_cv.split(unique_event_ids):
  # => replace with v0.17
  # nsamples=unique_event_ids.shape[0]
  # event_cv = ShuffleSplit(nsamples,1, test_size=0.5, random_state=57)
  # for train_event_is, test_event_is in event_cv:
  ```
- and things run fine

$ python user_test_submission.py
Reading file ...
Training ...
Testing ...
score = 0.846205321725