

Master of Technology

Unit 2/6: Computational Intelligence I

Workshop (5): Titanic Survivor Case Study with Python & Orange3

GU Zhan (Sam)
Institute of Systems Science
National University of Singapore

© 2018 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of ISS, NUS other than for the purpose for which it has been supplied.

Objectives

On completion of this workshop, students will

- » **have practical understanding on machine learning powered data analysis with Python & Orange3 data mining workflow tools, for Titanic survived passenger case study**
- » **conduct exploratory data analysis, machine learning, and evaluations for an end to end data analysis mini project**
- » **use and train neural net and support vector machine models from Python scikit-learn library**
- » **familiarize and participate in *Kaggle* data science competition to harness computational intelligence technical skills**

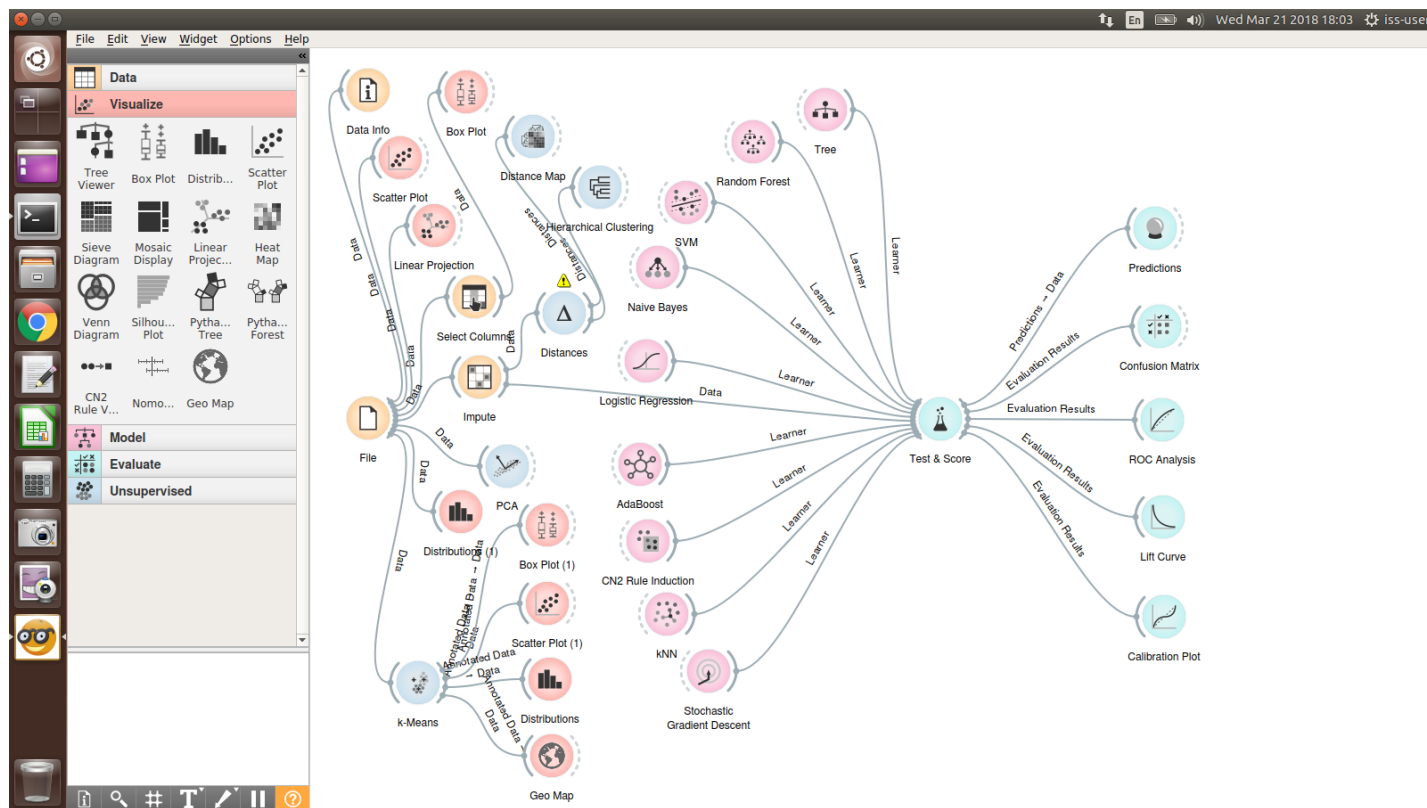
Workshop (5)

1. Download and install virtual machine iss-vm (25 GB) if necessary

<http://bit.ly/iss-vm>

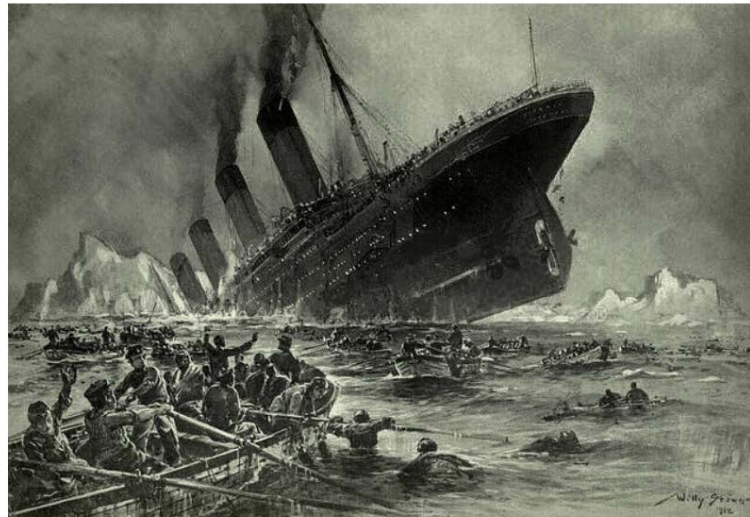
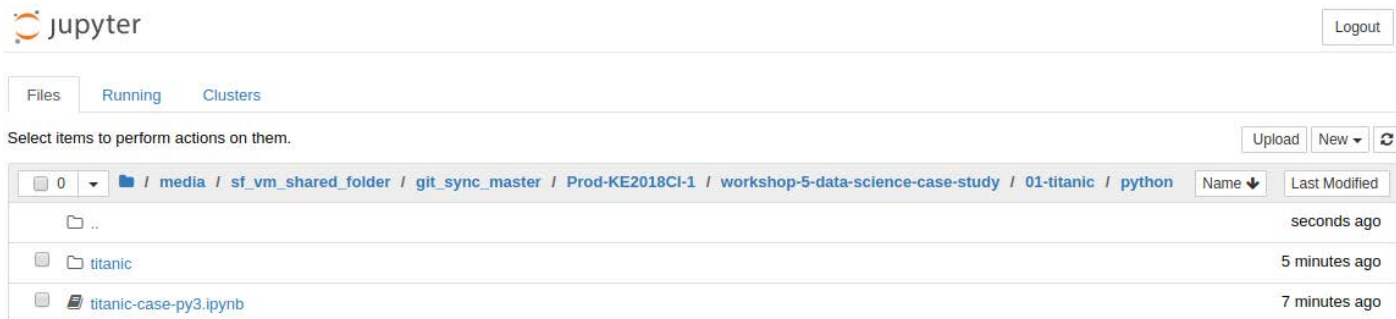
2. Click 'Tool Python3 Orange' on desktop to start Orange workflow platform: Open

`workshop-5-data-science-case-study/01-titanic/orange/orange-titanic.ows`



Workshop (5)

3. Click 'Tool Python3 Jupyter Notebook' on desktop to start Python: Open workshop-5-data-science-case-study/01-titanic/python/titanic-case.py3.ipynb



The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

<https://www.kaggle.com/c/titanic>

Workshop (5)

4. Register Kaggle.com; Participate; Submit to obtain competition Leaderboard Score:
 < Titanic: Machine Learning from Disaster > <https://www.kaggle.com/c/titanic/>

The screenshot shows the Kaggle website interface. At the top, there's a navigation bar with links for Competitions, Datasets, Kernels, Discussion, Jobs, and a Sign In button. Below this is a banner for the 'Titanic: Machine Learning from Disaster' competition, which includes the text 'Start here! Predict survival on the Titanic and get familiar with ML basics'. A red dashed line highlights the 'Kernels' tab in the navigation bar. Below the banner, there are tabs for Overview, Data, Kernels, Discussion, Leaderboard, and Rules. The 'Kernels' tab is selected. Below the tabs, there are filters for Public, Your Work, and Favorites, and a 'Sort by' dropdown set to 'Hotness'. A search bar for kernels is also present. The main content area displays a list of kernels, including 'Titanic: 2nd degree families and majority voting' and 'Introduction to Ensembling/Stacking in Python'.