melanie.wu [10:55 AM]

joined dat9 along with 9 others.

melanie.wu [3:51 PM]

set up a reminder “Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!” in this channel at 4:50PM every Saturday, Singapore Standard Time.

Ashu Goel [7:20 PM]

joined dat9 along with Harriet.

Yeo Wee Kiang [11:49 AM]

Hello everybody :wave::skin-tone-2:, I will be your Instructor for this course. Looking forward to seeing you really soon. :slightly\_smiling\_face:

Please complete all the \*Python\*-related topics in your \_pre-work\_ \*before\* this Saturday. :computer:

KMAN52 [1:31 PM]

joined dat9 along with 3 others.

Yeo Wee Kiang [10:18 AM]

this is a test message.

test

print("Hello")

Yeo Wee Kiang [3:21 PM]

https://s3.amazonaws.com/answer-board-image/25c3a3f5-9b48-4881-ac37-b7c4f0cc3f1f.jpeg

https://s3.amazonaws.com/answer-board-image/25c3a3f5-9b48-4881-ac37-b7c4f0cc3f1f.jpeg

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

RyanPeralta [5:31 PM]

joined dat9.

Yeo Wee Kiang [7:21 PM]

Course materials for this Saturday's class have been pushed. Please go through them and start attempting Project 1. (edited)

Yeo Wee Kiang [3:47 AM]

Introducing the Facebook Field Guide to Machine Learning video series – Facebook Research

https://research.fb.com/the-facebook-field-guide-to-machine-learning-video-series

Facebook Research

Introducing the Facebook Field Guide to Machine Learning video series

The Facebook Field Guide to Machine Learning is a six-part video series developed by the Facebook ads machine learning team. The series shares best real-world practices and provides practical tips about how to apply machine-learning capabilities to real-world problems. Machine learning and artificial intelligence are in the headlines everywhere today, and there are many resources…

https://research.fb.com/wp-content/uploads/2018/05/ml-academy-hero-graphic.png

Yeo Wee Kiang [9:28 AM]

Python Functions: print versus return

def function\_that\_prints():

print("I printed")

​

def function\_that\_returns():

return "I returned"

shanthi [10:18 AM]

joined dat9.

Yeo Wee Kiang [4:07 PM]

https://archive.ics.uci.edu/ml/index.php

https://www.kaggle.com/datasets

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Yeo Wee Kiang [2:20 PM]

I have pushed course materials for \_Experiments & Hypothesis Testing\_ lesson. Please pull.

Project 1 is due on \*1st September 2018 (Saturday), 6am\* : Submit here https://goo.gl/forms/QYqsonQ2eWQ3KFrd2

Google Docs

DAT9 Project Submission Form

1. Create a repository for this project in your GitHub.com account. 2. Push your files onto your GitHub.com repository. 3. Submit the URL (web link) to that particular project repository here.

https://lh4.googleusercontent.com/VLLdL2Q2ZiVhsujX5KvRABx31YtwKG2oZpoh5INuHvu6FFeTLZVN\_4B8nUuomgy9CaM

Harriet [9:54 PM]

left and rejoined dat9.

Yeo Wee Kiang [6:59 AM]

Another AI winter could usher in a dark period for artificial intelligence | Popular Science

https://www.popsci.com/ai-winter-artificial-intelligence

Popular Science

Another AI winter could usher in a dark period for artificial intelligence

Artificial intelligence researchers have already weathered several "AI winters" of decreased funding and public skepticism. They may face another one soon.

https://www.popsci.com/sites/popsci.com/files/styles/opengraph\_1\_91x1/public/images/2018/08/number-love-heart-red-broken-toy-926052-pxhere.com\_.jpg?itok=uzE99kTG

Yeo Wee Kiang [2:27 PM]

https://www.youtube.com/watch?v=dSjvK-Z3o3U

YouTubePyData

Brian Granger, Chris Colbert & Ian Rose - JupyterLab+Real Time Collaboration

Yeo Wee Kiang [3:24 PM]

https://www.youtube.com/watch?v=gZB\_ENJD34E

YouTubesentdex

Args and Kwargs - Intermediate Python Programming p.25

Yeo Wee Kiang [11:30 AM]

Gentle reminder: Please submit your Project 1 if you have not done so, https://goo.gl/forms/QYqsonQ2eWQ3KFrd2 For those who have submitted, received with thanks.

Google Docs

DAT9 Project Submission Form

1. Create a repository for this project in your GitHub.com account. 2. Push your files onto your GitHub.com repository. 3. Submit the URL (web link) to that particular project repository here.

https://lh4.googleusercontent.com/VLLdL2Q2ZiVhsujX5KvRABx31YtwKG2oZpoh5INuHvu6FFeTLZVN\_4B8nUuomgy9CaM

joe [12:37 PM]

hi. where is that jupyter page during day 1 morning session? the one with project schedule deadlines and time lines? i can't see them on the "classes" anymore

joe [12:44 PM]

we need to submit project 1 and project 4 before 6am tomorrow, right? what is project 4 ? i kinda need dteails...

nelson [12:49 PM]

https://git.generalassemb.ly/dat-sg-09/classes/blob/master/README.md

Sheena [12:54 PM]

Hi @joe project 4 proposal and dataset is for next week along with project 2 :slightly\_smiling\_face: for this week it’s just the python technical code challenges!

joe [1:02 PM]

ok. project 1 it is

Sheena [4:01 PM]

Hi all, can anyone let me know who has github id jotmy? Thank you!

and KMAN2208

Sheena [4:45 PM]

got it @Joyce and @KMAN52 :slightly\_smiling\_face:

Yeo Wee Kiang [9:28 AM]

Spurious Correlations :laughing: http://www.tylervigen.com/spurious-correlations

tylervigen.com

15 Insane Things That Correlate With Each Other

Why do these things correlate? These 15 correlations will blow your mind. (Is this headline sensationalist enough for you to click on it yet?)

http://tylervigen.com/images/spurious-correlations-share.png

Yeo Wee Kiang [9:36 AM]

A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population mean (μ).

confidence\_interval.png

Yeo Wee Kiang [2:23 PM]

http://contrib.scikit-learn.org/imbalanced-learn/stable/auto\_examples/combine/plot\_smote\_tomek.html

Yeo Wee Kiang [3:05 PM]

http://contrib.scikit-learn.org/imbalanced-learn/stable/under\_sampling.html#tomek-links

Yeo Wee Kiang [4:26 PM]

https://hangouts.google.com/call/voUD8Ce49Qx89bYZWWoCAAEN

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

joe [9:10 PM]

Statistics - The Art and Science of Learning from Data 3e - Agresti, Franklin. This is 3rd edition though. the one we saw today is the 4th edition.

PDF

Statistics - The Art and Science of Learning from Data 3e - Agresti, Franklin.pdf

42 MB PDF

Yeo Wee Kiang [7:55 AM]

http://cleverowl.uk/2016/02/06/curse-of-dimensionality-explained/

Clever Owl

Curse of Dimensionality Explained | Clever Owl

The Curse of Dimensionality, a term initially introduced by Richard Bellman[1], is a phenomena that arises when applying machine learning algorithms to highly-dimensional data. Let's take a simple example as an illustration of the issue. Say we have a data set of observations where , and is a scalar (). We also have a set of target classes , where is a class variable expressing association with one of two possible classes - either

Feb 7th, 2016

Yeo Wee Kiang [4:08 PM]

http://www.nada.kth.se/kurser/su/DA2001/sudata16/examination/python-quick-ref.pdf

Yeo Wee Kiang [6:18 AM]

The 50 Best Free Datasets for Machine Learning - Gengo AI

https://gengo.ai/datasets/the-50-best-free-datasets-for-machine-learning/

Gengo AI

The 50 Best Free Datasets for Machine Learning - Gengo AI

What are some open datasets for machine learning? We at Gengo decided to create the ultimate cheat sheet for high quality datasets.

Jun 13th

https://gengo.ai/wp-content/uploads/2018/06/heroplacehold\_50datasets-1.jpg

Ashu Goel [8:27 AM]

http://www.nada.kth.se/kurser/su/DA2001/sudata16/examination/python-quick-ref.pdf

Yeo Wee Kiang [10:36 AM]

https://www.searchenginejournal.com/google-introduces-new-search-engine-for-finding-datasets/268337/

Search Engine Journal

Google Introduces New Search Engine for Finding Datasets - Search Engine Journal

Google has launched a new type of search engine designed specifically around helping people find data.

Sep 6th

Yeo Wee Kiang [9:52 AM]

https://plot.ly/python/ipython-notebook-tutorial/

plot.ly

Jupyter Notebook Tutorial

Jupyter notebook tutorial on how to install, run, and use Jupyter for interactive matplotlib plotting, data analysis, and publishing code

Yeo Wee Kiang [10:12 AM]

http://dorienherremans.com

Yeo Wee Kiang [1:26 PM]

https://www.ibm.com/developerworks/rational/library/content/RationalEdge/sep04/bell/index.html

ibm.com

The class diagram

from The Rational Edge: As the most important example of the new structure diagram type in UML 2, the class diagram can be used by analysts, business modelers, developers, and testers throughout the software development lifecycle. This article offers a comprehensive introduction.

https://www.edx.org/course/introduction-to-computer-science-and-programming-using-python

edX

Introduction to Computer Science and Programming Using Python

An introduction to computer science as a tool to solve real-world analytical problems using Python 3.5.

Jul 6th

https://prod-discovery.edx-cdn.org/media/course/image/956319ec-8665-4039-8bc6-32c9a9aea5e9-7080122292de.small.jpg

Yeo Wee Kiang [2:07 PM]

https://www.youtube.com/watch?v=apACNr7DC\_s

YouTubeSocratica

Python Classes and Objects || Python Tutorial || Learn Python Programming

Yeo Wee Kiang [2:24 PM]

https://www.analyticsvidhya.com/blog/2017/08/introduction-to-multi-label-classification/

Analytics Vidhya

Solving Multi-Label Classification problems (Case studies included)

This article introduces multi label classification problems. It covers approaches to solve these problem using multi-learn (scikit) library in python.

Aug 26th, 2017

Yeo Wee Kiang [2:32 PM]

https://venturebeat.com/2017/04/07/how-olay-skin-advisor-built-their-deep-learning-algorithms/

VentureBeat

How Olay uses deep learning to personalize skin care

The beauty and skincare world is oversaturated, especially if you include all the affordable convenience store brands. If you’re a shopper with a budget, you are likely mixing different produ…

Apr 8th, 2017

https://venturebeat.com/wp-content/uploads/2017/04/venturebeat\_olay\_1200px.jpg?fit=1200%2C642&strip=all

https://olayskinadvisor.ca/

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Yeo Wee Kiang [5:28 AM]

Forecasting at Uber: An Introduction

https://eng.uber.com/forecasting-introduction/

Uber Engineering Blog

Forecasting at Uber: An Introduction

In this article, we provide a general overview of how our teams leverage forecasting to build better products and maintain the health of the Uber marketplace.

Sep 7th

http://eng.uber.com/wp-content/uploads/2018/09/twitter.png

Yeo Wee Kiang [7:18 AM]

https://www.youtube.com/watch?v=m2DfpM6MyB8&feature=share

YouTubeGoogle Cloud Platform

How to Make a Data Science Project with Kaggle

Yeo Wee Kiang [6:27 AM]

https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333

Towards Data Science

Web Scraping TripAdvisor, Text Mining and Sentiment Analysis for Hotel Reviews

Study after study has shown that TripAdvisor is becoming terrifyingly important in a traveler’s decision making process. However…

Reading time

10 min read

Aug 6th

https://cdn-images-1.medium.com/max/1200/1\*4M\_Cxv6wIrI1o16MDp3c3A.jpeg

Yeo Wee Kiang [10:57 AM]

http://queirozf.com/entries/one-hot-encoding-a-feature-on-a-pandas-dataframe-an-example

queirozf.com

One-Hot Encoding a Feature on a Pandas Dataframe: an Example

One-hot encoding is a simple way to transform categorical features into vectors that are easy to deal with. Learn how to do this on a Pandas DataFrame.

Nov 27th, 2015

http://queirozf.com/images/landscape/tiger.png

http://www.mien.in/2017/10/14/handling-categorical-features-in-python/

The Mien Blog

Handling Categorical Features With Python

What are Categorical variables?Ordinal/Ordered categorical variableNominal categorical variableSimple approach to encode categorical featuresLabel EncodingOne Hot Encoding/One of K schemeImplementing

Yeo Wee Kiang [5:37 PM]

http://community.asdlib.org/imageandvideoexchangeforum/files/2013/07/Figure5.10.jpg

http://community.asdlib.org/imageandvideoexchangeforum/files/2013/07/Figure5.10.jpg

chiadaozhe [10:00 PM]

hi wee kiang, can i check how much details we need to write for the project 4 proposal?

Yeo Wee Kiang [10:01 PM]

Enough for us to assess if the project is viable. Don't need to write a thesis

chiadaozhe [10:01 PM]

ok thanks!

Yeo Wee Kiang [7:15 AM]

https://medium.com/nanonets/topic-modeling-with-lsa-psla-lda-and-lda2vec-555ff65b0b05

Medium

Topic Modeling with LSA, PSLA, LDA & lda2Vec – NanoNets – Medium

From matrix factorization to deep learning

Reading time

12 min read

May 25th

https://cdn-images-1.medium.com/max/1200/1\*\_ZMgTsJGmR743ngZ7UxN9w.png

Yeo Wee Kiang [9:32 AM]

https://www.youtube.com/watch?v=U3B5z2HQlaQ

YouTubeTrevor Payne

Let's Learn Python #13 - UML - Unified Modeling Language

Yeo Wee Kiang [8:49 AM]

overfit vs underfit|melvin

image\_d954ce84-a637-4dc9-869a-664966af446720180516\_212200.jpg

Yeo Wee Kiang

Guide to Numpy: http://web.mit.edu/dvp/Public/numpybook.pdf

Posted in #dat8May 17th

Yeo Wee Kiang

For those of you who are more finance-inclined, you can refer to https://www.quantopian.com/lectures

Posted in #dat8May 20th

Yeo Wee Kiang

Deep and Shallow Copies in Python

https://github.com/UWSEDS/LectureNotes/blob/master/Deep-and-Shallow-Copies.ipynb

Thread in #dat8Jun 3rd

Medium

Three techniques to improve machine learning model performance with imbalanced datasets

This project was part of one my recent job interview skill test for a “Machine learning engineer” position. I had to complete the project…

Reading time

5 min read

Apr 15th

https://cdn-images-1.medium.com/max/1200/1\*r6VqMHOut9tDB4qMjPjcrQ.jpeg

Towards Data Science

Topic Modeling and Latent Dirichlet Allocation (LDA) in Python

Topic modeling is a type of statistical modeling for discovering the abstract “topics” that occur in a collection of documents. Latent…

Reading time

5 min read

May 31st

https://cdn-images-1.medium.com/max/1200/1\*LeJ3vMxW4ZvcS4ZsDyJDYA.jpeg

nbviewer.jupyter.org

Notebook on nbviewer

Check out this Jupyter notebook!

Yeo Wee Kiang

Here’s another article on handling imbalanced datasets: https://blog.dominodatalab.com/imbalanced-datasets/

Posted in #dat8Jun 18th

r2d3.us

A visual introduction to machine learning, Part II

Learn about bias and variance in our second animated data visualization.

http://www.r2d3.us/static/pages/decision-trees-part2-v2/part2-preview-en.png

Yeo Wee Kiang [9:34 AM]

https://www.statsmodels.org/stable/examples/index.html

https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/

Analytics Vidhya

A Complete Tutorial on Ridge and Lasso Regression in Python

Here is a complete tutorial on the regularization techniques of ridge and lasso regression to prevent overfitting in prediction in python

Jan 28th, 2016

Yeo Wee Kiang [3:43 PM]

https://hangouts.google.com/call/NWoTX9DwyTF2Xc1auZsEAAEM

Yeo Wee Kiang [3:48 PM]

http://setosa.io/ev/ordinary-least-squares-regression/

Explained Visually

Ordinary Least Squares Regression explained visually

http://setosa.io/ev/ordinary-least-squares-regression/fb-thumb.png

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

RyanPeralta [4:54 PM]

@nelson can we borrow your Project 2? Amazing stuff! (edited)

Yeo Wee Kiang [11:19 AM]

https://www.youtube.com/watch?v=leTyvBPhYzw&feature=share

YouTubeDatabricks

Art of Feature Engineering for Data Science - Nabeel Sarwar

Yeo Wee Kiang [11:34 AM]

https://www.youtube.com/playlist?list=PLOU2XLYxmsIIuiBfYad6rFYQU\_jL2ryal

Yeo Wee Kiang [6:39 PM]

A Feature Selection Tool for Machine Learning in Python

https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0

Towards Data Science

A Feature Selection Tool for Machine Learning in Python

Using the FeatureSelector for efficient machine learning workflows

Reading time

10 min read

Jun 22nd

https://cdn-images-1.medium.com/max/1200/1\*rqeNwjiakRS-SqsW9wCgrA.jpeg

Yeo Wee Kiang [8:04 AM]

https://twimlai.com

This Week in Machine Learning & AI

This Week in Machine Learning and AI Podcast

Keep up with the most interesting and important stories from the world of machine learning, deep learning, artificial intelligence and bots.

Yeo Wee Kiang [6:02 AM]

https://etav.github.io/python/vif\_factor\_python.html

etav.github.io

Variance Inflation Factor (VIF) Explained - Python

Applied Data Science, progamming and machine learning projects

Yeo Wee Kiang [9:27 AM]

https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

Analytics Vidhya

Introduction to KNN, K-Nearest Neighbors : Simplified

This article explains k nearest neighbor (KNN),one of the popular machine learning algorithms, working of kNN algorithm and how to choose factor k in simple terms.

Mar 26th

https://www.fabienplisson.com/finding-the-right-k-in-k-nearest-neighbors/

Yeo Wee Kiang [10:25 AM]

https://chrisalbon.com/machine\_learning/linear\_regression/selecting\_best\_alpha\_value\_in\_ridge\_regression/

chrisalbon.com

Selecting The Best Alpha Value In Ridge Regression

How to select the best alpha value when conduct in ridge regression in scikit-learn for machine learning in Python.

Dec 21st, 2017

https://codingstartups.com/practical-machine-learning-ridge-regression-vs-lasso/

Coding Startups

Practical machine learning: Ridge regression vs. Lasso - Coding Startups

For many years, programmers have tried to solve extremely complex computer science problems using traditional algorithms which are based on the most basic condition statement: if this then that. For example, if the email contains the word “free!” it should be classified as spam. In recent years, with the rise of exceptional cloud computing technologies, … Read More Read More

Aug 3rd, 2017

Yeo Wee Kiang [2:31 PM]

http://scikit-learn.org/stable/modules/generated/sklearn.model\_selection.GridSearchCV.html#sklearn.model\_selection.GridSearchCV

Yeo Wee Kiang [4:07 PM]

https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/

https://www.unm.edu/~schrader/biostat/bio2/Spr06/lec11.pdf

Yeo Wee Kiang [4:18 PM]

https://gist.githubusercontent.com/curran/a08a1080b88344b0c8a7/raw/d546eaee765268bf2f487608c537c05e22e4b221/iris.csv

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Yeo Wee Kiang [4:33 PM]

https://towardsdatascience.com/stock-prediction-in-python-b66555171a2

Towards Data Science

Stock Prediction in Python – Towards Data Science

Make (and lose) fake fortunes while learning real Python

Reading time

12 min read

Jan 20th

https://cdn-images-1.medium.com/max/1200/1\*dSbEKnRVhfCJMgLfQDtrCw.jpeg

Yeo Wee Kiang [7:02 PM]

https://medium.com/@elye.project/applying-tensorflow-in-android-in-4-steps-to-recognize-superhero-f224597eb055

Medium

Applying TensorFlow in Android in 4 steps – Elye – Medium

In this blog, I’ll show in 4 steps on how to use a set of pictures (superhero), from training them in the TensorFlow to recognize them…

Reading time

7 min read

Mar 25th

https://cdn-images-1.medium.com/max/1200/1\*6ZlDpvHywohbieIHipPa3g.png

Paul [9:13 PM]

Hey guys, if any of you want a nice and easy to understand / read intro to Logisitc regression, I though this would help:

https://medium.com/data-science-group-iitr/logistic-regression-simplified-9b4efe801389

Medium

Logistic Regression. Simplified. – Data Science Group, IITR – Medium

After the basics of Regression, it’s time for basics of Classification. And, what can be easier than Logistic Regression!

Reading time

5 min read

Apr 1st, 2017

https://cdn-images-1.medium.com/max/1200/1\*D70wG-VO7GFq85VFejundA.jpeg

Yeo Wee Kiang [7:07 PM]

https://medium.com/@mjspeck/presenting-code-using-jupyter-notebook-slides-a8a3c3b59d67

Medium

Presenting Code Using Jupyter Notebook Slides – Matthew Speck – Medium

When you make a slide presentation, there are a few programs you likely think of: Microsoft PowerPoint, Google Slides, Prezi (just…

Reading time

3 min read

May 16th, 2017

https://cdn-images-1.medium.com/max/1200/1\*kt\_bEri7QAArBhP1qPJgMg.png

Yeo Wee Kiang [10:47 PM]

https://rawgraphs.io/learning/introduction-to-rawgraphs/

Yeo Wee Kiang [9:37 AM]

https://www.hackster.io/baqwas/visualize-your-zero-dbff3f

Hackster.io

Visualize Your Zero

A portable advanced visual recognition solution can help you overcome many situations where a picture is worth billions of calculations. By Matha Goram.

https://hackster.imgix.net/uploads/attachments/581343/wp\_20170515\_12\_26\_22\_pro\_B6tAV565Dp.jpg?auto=compress%2Cformat&w=600&h=450&fit=min

https://www.raspberrypi.org/magpi/tensorflow-ai-raspberry-pi/

The MagPi Magazine

How to install TensorFlow on Raspberry Pi - The MagPi Magazine

Google TensorFlow 1.9 officially supports the Raspberry Pi, making it possible to quickly install TensorFlow and start learning AI techniques with a Raspberry Pi. Back in The MagPi issue 71 we noted that it was getting easier to install TensorFlow on a Raspberry Pi. This latest news makes installing TensorFlow 1.9 as simple as using …

Aug 6th

Yeo Wee Kiang [1:03 PM]

Good or not?

I built a machine learning model to solve a binary classification problem. The accuracy of myclassifier is 99%, and the AUC is 0.51. Is my model doing well? Why or why not?

martin [1:18 PM]

maybe ..."The model is not doing well because since AUC is about 0.51... classifier is not able to differentiate between the labels... random classifier. Higher the AUC, the better the model and vice versa. However, the AUC is computed with the prediction scores specified as a parameter in the AUC method." #dat9

Yeo Wee Kiang [2:58 PM]

https://imbalanced-learn.readthedocs.io/en/stable/auto\_examples/combine/plot\_smote\_tomek.html

Yeo Wee Kiang [4:23 PM]

https://hangouts.google.com/call/2DKVmXbvGYp7z1aUmkDFAAEM

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Yeo Wee Kiang [5:17 AM]

Interpretable Machine Learning

https://christophm.github.io/interpretable-ml-book/

christophm.github.io

Interpretable Machine Learning

Machine learning algorithms usually operate as black boxes and it is unclear how they derived a certain decision. This book is a guide for practitioners on how to make machine learning decisions more interpretable.

Yeo Wee Kiang [9:56 AM]

https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0

Towards Data Science

A Feature Selection Tool for Machine Learning in Python

Using the FeatureSelector for efficient machine learning workflows

Reading time

10 min read

Jun 22nd

https://cdn-images-1.medium.com/max/1200/1\*rqeNwjiakRS-SqsW9wCgrA.jpeg

joe [1:56 PM]

Sir, i got a question: this is my confusion matrix...

Pasted image at 2018-10-02, 1:56 PM

and this is my ROC.

Pasted image at 2018-10-02, 1:57 PM

teh confusion matrix itself looks wrong. then the ROC looks ok?

Yeo Wee Kiang [9:54 PM]

http://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion\_matrix.html

y\_pred should be assigned with the ground truth labels of the target variable.

Yeo Wee Kiang [2:33 PM]

http://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html

t-redactyl.io

Using VADER to handle sentiment analysis with social media text

Using VADER to handle sentiment analysis with social media text written April 08, 2017 in python,programming tips,text mining

https://github.com/cjhutto/vaderSentiment

GitHub

cjhutto/vaderSentiment

VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social ...

Yeo Wee Kiang [4:32 PM]

http://blog.aylien.com/a-review-of-the-recent-history-of-natural-language-processing/

AYLIEN

A Review of the Neural History of Natural Language Processing - AYLIEN

This is the first blog post in a two-part series. The series expands on the Frontiers of Natural Language Processing session organized by Herman Kamper and me at the Deep …

Oct 1st

Yeo Wee Kiang [5:21 PM]

https://www.bogotobogo.com/python/python\_if\_\_name\_\_equals\_\_main\_\_.php

bogotobogo.com

Python Tutorial: \_\_name\_\_ == \_\_main\_\_ - 2018

Python Tutorial: \_\_name\_\_ == \_\_main\_\_

Yeo Wee Kiang [8:36 AM]

For those who wants a math-heavy machine learning textbook, you can refer to http://www.math.chalmers.se/Stat/Grundutb/CTH/mve187/1819/Bishop.pdf

Yeo Wee Kiang [8:41 AM]

https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets

Yeo Wee Kiang [8:55 AM]

https://www.youtube.com/watch?v=Gzun8PpyBCo

YouTubeEnthought

Getting Started with JupyterLab (Beginner Level) | SciPy 2018 Tutorial | Jason Grout

Yeo Wee Kiang [9:22 AM]

What is an API?

https://www.youtube.com/watch?v=s7wmiS2mSXY

YouTubeMuleSoft Videos

What is an API?

https://github.com/toddmotto/public-apis

GitHub

toddmotto/public-apis

A collective list of public JSON APIs for use in web development. - toddmotto/public-apis

Yeo Wee Kiang [10:34 AM]

https://www.quantopian.com/lectures

quantopian.com

Online Lectures in Statistical and Financial Topics

Hands-on Lectures for Quantitative Finance: Teaching Scientific and Statistical Rigor Through Real Data Examples

https://media.quantopian.com/public/og-image-default.png

Yeo Wee Kiang [10:44 AM]

https://www.mytransport.sg/content/mytransport/home/dataMall.html

Yeo Wee Kiang [10:56 AM]

https://doc.scrapy.org/en/latest/intro/tutorial.html

Yeo Wee Kiang [1:07 PM]

https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60

Towards Data Science

PCA using Python (scikit-learn) – Towards Data Science

My last tutorial went over Logistic Regression using Python. One of the things learned was that you can speed up the fitting of a machine…

Reading time

8 min read

Dec 5th, 2017

https://cdn-images-1.medium.com/max/1200/1\*Gob8ZbScyM7hHUHjvrMJYg.png

Yeo Wee Kiang [1:20 PM]

https://k-drama-finder.herokuapp.com/index/

Yeo Wee Kiang [1:32 PM]

http://www.automationtestinghub.com/download-chrome-driver/

AutomationTestingHub

How to download Chrome Driver for Selenium - AutomationTestingHub

This article provides detailed steps on how to download chrome driver. It also provides some basic information about chrome driver and why it is needed

Jun 7th

Yeo Wee Kiang [3:31 PM]

http://scikit-learn.org/stable/auto\_examples/ensemble/plot\_forest\_importances.html

RhysZH [4:00 PM]

https://www.theguardian.com/books/booksblog/2017/dec/13/harry-potter-botnik-jk-rowling

the Guardian

'He began to eat Hermione's family': bot tries to write Harry Potter book – and fails in magic ways

After being fed all seven Potter tales, a predictive keyboard has produced a tale that veers from almost genuine to gloriously bonkers

Dec 14th, 2017

https://i.guim.co.uk/img/media/eb055678f11104d510fe0c4ca6f7c8134552bf53/0\_150\_2745\_1647/master/2745.jpg?width=1200&height=630&quality=85&auto=format&fit=crop&crop=faces%2Centropy&bm=normal&ba=bottom%2Cleft&blend64=aHR0cHM6Ly9hc3NldHMuZ3VpbS5jby51ay9pbWFnZXMvb3ZlcmxheXMvZDM1ODZhNWVmNTc4MTc1NmQyMWEzYjYzNWU1MTcxNDEvdGctZGVmYXVsdC5wbmc&s=bd945933f787016df97a837d4c9e394c

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Yeo Wee Kiang [8:39 AM]

https://www.youtube.com/watch?v=uvCS\_lfxFWI&feature=share

YouTubeKaggle

Kaggle Live-Coding: How to turn text into structured data for NLP

Yeo Wee Kiang [8:54 AM]

https://towardsdatascience.com/machine-learning-word-embedding-sentiment-classification-using-keras-b83c28087456

Towards Data Science

Machine Learning — Word Embedding & Sentiment Classification using Keras

In the previous post, we discussed various steps of text processing involved in Nature Language Processing (NLP) and also implemented a…

Reading time

9 min read

Oct 4th

https://cdn-images-1.medium.com/max/1200/1\*jvJUYSgLHSHw\_cdkNjF7Yg.jpeg

Yeo Wee Kiang [6:42 AM]

https://medium.freecodecamp.org/the-hitchhikers-guide-to-machine-learning-algorithms-in-python-bfad66adb378

freeCodeCamp.org

The Hitchhiker’s Guide to Machine Learning in Python

Featuring implementation code, instructional videos, and more

Reading time

8 min read

Aug 2nd, 2017

https://cdn-images-1.medium.com/max/1200/1\*D4v4JceAnfoAfj8DwkwA8w.png

Yeo Wee Kiang [11:16 AM]

https://medium.com/@aneesha/recursive-feature-elimination-with-scikit-learn-3a2cbdf23fb7

Medium

Recursive Feature Elimination with Scikit Learn – Aneesha Bakharia – Medium

Datasets used to train classification and regression algorithms are high dimensional in nature — this means that they contain many…

Reading time

2 min read

Feb 27th, 2016

Yeo Wee Kiang [4:17 PM]

https://medium.com/@yanhann10/a-brief-view-of-machine-learning-pipeline-in-python-5f50b941fca8

Medium

A brief view of machine learning pipeline in python

As I step out of R’s comfort zone and venture into Python land, I find pipeline in scikit-learn useful to understand before moving on to…

Reading time

1 min read

Jun 7th, 2017

Yeo Wee Kiang [4:31 PM]

https://www.kaggle.com/baghern/a-deep-dive-into-sklearn-pipelines

Yeo Wee Kiang [10:02 AM]

https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d

Medium

Your Easy Guide to Latent Dirichlet Allocation – Lettier – Medium

Before we get started, I made a tool (here’s the source) that runs LDA right inside your browser (it’s pretty neat). Be sure to have that…

Reading time

10 min read

Feb 24th

https://cdn-images-1.medium.com/max/1200/1\*eqArgnH7PNT76mhB9pQjIg.jpeg

Yeo Wee Kiang [10:10 AM]

https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/

Machine Learning Mastery

Linear Discriminant Analysis for Machine Learning

Logistic regression is a classification algorithm traditionally limited to only two-class classification problems. If you have more than two classes then Linear Discriminant Analysis is the preferred linear classification technique. In this post you will discover the Linear Discriminant Analysis (LDA) algorithm for classification predictive modeling problems. After reading this post you will know: The …

Apr 6th, 2016

https://3qeqpr26caki16dnhd19sv6by6v-wpengine.netdna-ssl.com/wp-content/uploads/2016/03/Linear-Discriminant-Analysis-for-Machine-Learning.jpg

https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d

Yeo Wee Kiang [10:55 AM]

http://nbviewer.jupyter.org/github/bmabey/hacker\_news\_topic\_modelling/blob/master/HN%20Topic%20Model%20Talk.ipynb

nbviewer.jupyter.org

Notebook on nbviewer

Check out this Jupyter notebook!

Yeo Wee Kiang [12:06 PM]

https://github.com/facebook/prophet

GitHub

facebook/prophet

Tool for producing high quality forecasts for time series data that has multiple seasonality with linear or non-linear growth. - facebook/prophet

Yeo Wee Kiang [2:40 PM]

https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/

Yeo Wee Kiang [2:53 PM]

https://plot.ly/scikit-learn/plot-kmeans-silhouette-analysis/

plot.ly

Selecting the number of Clusters with Silhouette Analysis on KMeans Clustering

Plotly is the easiest way to graph and share your data.

Yeo Wee Kiang [3:49 PM]

https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent

Google Developers

Reducing Loss: Stochastic Gradient Descent | Machine Learning Crash Course | Google Developers

Yeo Wee Kiang [4:12 PM]

https://keras.io/

slackbot [4:50 PM]

Reminder: Please fill out the exit ticket for today’s lesson here: https://goo.gl/8iZfWG ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

Google Docs

Welcome to Today's Exit Ticket

This survey is designed to help your instructor as well as the GA team to understand how you felt about the lesson and how effective the instruction was today. Your honest feedback will help you improve your learning experience in real time.

https://lh4.googleusercontent.com/TFqiSlQJquGdsWLLzJQhGbr6UqHdivv9ZgQiombU2GnpJ4VoTgEzBoLLKRSFxvyRpx0

Jonathan [7:05 PM]

joined dat9.

joe [11:55 PM]

Ok, I have the winning model and pickled it for further use. How do I deploy?

Jonathan [10:33 AM]

left dat9.

melanie.wu [10:33 AM]

made this channel \*private\*. Now, it can only be viewed or joined by invitation.

Yeo Wee Kiang [5:12 PM]

https://stats.stackexchange.com/questions/57640/why-is-there-a-sharp-elbow-in-my-roc-curves

Cross Validated

Why is there a sharp elbow in my ROC curves?

I have some EEG data sets that I am testing against two classes. I can get a decent error rate from LDA (the class-conditional distributions aren't Gaussian, but have similar tails and good enough

Yeo Wee Kiang [9:06 AM]

https://dataskeptic.com/

http://lineardigressions.com/

http://datastori.es/

http://nssdeviations.com/

https://www.oreilly.com/topics/oreilly-data-show-podcast

http://www.learningmachines101.com/

http://www.thetalkingmachines.com/

https://twimlai.com/

http://partiallyderivative.com/

https://www.datacamp.com/community/podcast

Yeo Wee Kiang [10:36 AM]

https://developers.google.com/web/progressive-web-apps/

Google Developers

Progressive Web Apps | Web | Google Developers

https://developers.google.com/web/images/social-webfu-16x9.png

Yeo Wee Kiang [11:01 AM]

Untitled

Dao Zhe Chia

Albert Nelson Antonydoss

Rhys Morgan

Kai En Gwee

Ryan Peralta

https://hangouts.google.com/call/wVoGtEAFYEzneT5qhWaXAAEM

Ashu Goel [4:36 PM]

Slack Chat

dat9.md

## #dat9

\_\_\_\_

​

### Tue, October 2nd 2018

9:56 AM \*\*wee.kiang\*\*: <https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0>

​

1:56 PM \*\*jonatslim\*\*: Sir, i got a question: this is my confusion matrix...

​

1:57 PM \*\*jonatslim\*\*: and this is my ROC.

​

1:58 PM \*\*jonatslim\*\*: teh confusion matrix itself looks wrong. then the ROC looks ok?

​

9:54 PM \*\*wee.kiang\*\*: <http://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion\_matrix.html>

​

9:55 PM \*\*wee.kiang\*\*: y\_pred should be assigned with the ground truth labels of the target variable.

​

​

​

### Fri, October 5th 2018

2:33 PM \*\*wee.kiang\*\*: <http://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html>

​

2:34 PM \*\*wee.kiang\*\*: <https://github.com/cjhutto/vaderSentiment>

​

4:32 PM \*\*wee.kiang\*\*: <http://blog.aylien.com/a-review-of-the-recent-history-of-natural-language-processing/>

​

5:21 PM \*\*wee.kiang\*\*: <https://www.bogotobogo.com/python/python\_if\_\_name\_\_equals\_\_main\_\_.php>

​

​

​

### Sat, October 6th 2018

8:36 AM \*\*wee.kiang\*\*: For those who wants a math-heavy machine learning textbook, you can refer to <http://www.math.chalmers.se/Stat/Grundutb/CTH/mve187/1819/Bishop.pdf>

​

8:41 AM \*\*wee.kiang\*\*: <https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets>

​

8:55 AM \*\*wee.kiang\*\*: <https://www.youtube.com/watch?v=Gzun8PpyBCo>

​

9:22 AM \*\*wee.kiang\*\*: What is an API?

<https://www.youtube.com/watch?v=s7wmiS2mSXY>

​

9:26 AM \*\*wee.kiang\*\*: <https://github.com/toddmotto/public-apis>

​

10:34 AM \*\*wee.kiang\*\*: <https://www.quantopian.com/lectures>

​

10:44 AM \*\*wee.kiang\*\*: <https://www.mytransport.sg/content/mytransport/home/dataMall.html>

​

10:56 AM \*\*wee.kiang\*\*: <https://doc.scrapy.org/en/latest/intro/tutorial.html>

​

1:07 PM \*\*wee.kiang\*\*: <https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60>

​

1:20 PM \*\*wee.kiang\*\*: <https://k-drama-finder.herokuapp.com/index/>

​

1:32 PM \*\*wee.kiang\*\*: <http://www.automationtestinghub.com/download-chrome-driver/>

​

3:31 PM \*\*wee.kiang\*\*: <http://scikit-learn.org/stable/auto\_examples/ensemble/plot\_forest\_importances.html>

​

4:00 PM \*\*rhyszh.morgan\*\*: <https://www.theguardian.com/books/booksblog/2017/dec/13/harry-potter-botnik-jk-rowling>

​

4:50 PM \*\*slackbot\*\*: Reminder: Please fill out the exit ticket for today’s lesson here: <https://goo.gl/8iZfWG> ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

​

​

​

### Sun, October 7th 2018

8:39 AM \*\*wee.kiang\*\*: <https://www.youtube.com/watch?v=uvCS\_lfxFWI&feature=share>

​

​

​

### Mon, October 8th 2018

8:54 AM \*\*wee.kiang\*\*: <https://towardsdatascience.com/machine-learning-word-embedding-sentiment-classification-using-keras-b83c28087456>

​

​

​

### Tue, October 9th 2018

6:42 AM \*\*wee.kiang\*\*: <https://medium.freecodecamp.org/the-hitchhikers-guide-to-machine-learning-algorithms-in-python-bfad66adb378>

​

11:16 AM \*\*wee.kiang\*\*: <https://medium.com/@aneesha/recursive-feature-elimination-with-scikit-learn-3a2cbdf23fb7>

​

4:17 PM \*\*wee.kiang\*\*: <https://medium.com/@yanhann10/a-brief-view-of-machine-learning-pipeline-in-python-5f50b941fca8>

​

4:31 PM \*\*wee.kiang\*\*: <https://www.kaggle.com/baghern/a-deep-dive-into-sklearn-pipelines>

​

​

​

### Sat, October 13th 2018

10:02 AM \*\*wee.kiang\*\*: <https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d>

​

10:10 AM \*\*wee.kiang\*\*: <https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/>

​

10:10 AM \*\*wee.kiang\*\*: <https://medium.com/@lettier/how-does-lda-work-ill-explain-using-emoji-108abf40fa7d>

​

10:55 AM \*\*wee.kiang\*\*: <http://nbviewer.jupyter.org/github/bmabey/hacker\_news\_topic\_modelling/blob/master/HN%20Topic%20Model%20Talk.ipynb>

​

12:06 PM \*\*wee.kiang\*\*: <https://github.com/facebook/prophet>

​

2:40 PM \*\*wee.kiang\*\*: <https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>

​

2:53 PM \*\*wee.kiang\*\*: <https://plot.ly/scikit-learn/plot-kmeans-silhouette-analysis/>

​

3:49 PM \*\*wee.kiang\*\*: <https://developers.google.com/machine-learning/crash-course/reducing-loss/stochastic-gradient-descent>

​

4:12 PM \*\*wee.kiang\*\*: <https://keras.io/>

​

4:50 PM \*\*slackbot\*\*: Reminder: Please fill out the exit ticket for today’s lesson here: <https://goo.gl/8iZfWG> ! This will help Wee Kiang and team address any concerns or questions you have! Thank you!

​

​

​

### Mon, October 15th 2018

7:05 PM \*\*jonathanfromsg\*\*: jonathanfromsg has joined the channel

​

​

​

### Wed, October 17th 2018

11:55 PM \*\*jonatslim\*\*: Ok, I have the winning model and pickled it for further use. How do I deploy?

​

​

​

### Thu, October 18th 2018

10:33 AM \*\*jonathanfromsg\*\*: jonathanfromsg has left the channel

​

10:33 AM \*\*melanie.wu\*\*: made this channel \*private\*. Now, it can only be viewed or joined by invitation.

​

​

​

### Fri, October 19th 2018

5:12 PM \*\*wee.kiang\*\*: <https://stats.stackexchange.com/questions/57640/why-is-there-a-sharp-elbow-in-my-roc-curves>

​

​

​

### Sat, October 20th 2018

9:06 AM \*\*wee.kiang\*\*: <https://dataskeptic.com/>

<http://lineardigressions.com/>

<http://datastori.es/>

<http://nssdeviations.com/>

<https://www.oreilly.com/topics/oreilly-data-show-podcast>

<http://www.learningmachines101.com/>

<http://www.thetalkingmachines.com/>

<https://twimlai.com/>

<http://partiallyderivative.com/>

<https://www.datacamp.com/community/podcast>

​

10:36 AM \*\*wee.kiang\*\*: <https://developers.google.com/web/progressive-web-apps/>

​

11:02 AM \*\*wee.kiang\*\*: <https://hangouts.google.com/call/wVoGtEAFYEzneT5qhWaXAAEM>

Collapse

Yeo Wee Kiang [4:42 PM]

https://www.linkedin.com/in/weekiang/

Ashu Goel [4:43 PM]

Chat

dat9 (1).md

## #dat9

<https://etav.github.io/python/vif_factor_python.html>

<https://twimlai.com> – podcast

A Feature Selection Tool for Machine Learning in Python

<https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0>

<https://www.youtube.com/watch?v=leTyvBPhYzw&feature=share> – art of Feature Eng

<http://setosa.io/ev/ordinary-least-squares-regression/>

<https://www.statsmodels.org/stable/examples/index.html>

<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/>

<https://blog.dominodatalab.com/imbalanced-datasets/>

A visual introduction to machine learning, Part II

Topic Modeling and Latent Dirichlet Allocation (LDA) in Python

<http://www.mien.in/2017/10/14/handling-categorical-features-in-python/>

<http://queirozf.com/entries/one-hot-encoding-a-feature-on-a-pandas-dataframe-an-example>

<https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333>

<https://www.youtube.com/watch?v=m2DfpM6MyB8&feature=share>

<https://www.analyticsvidhya.com/blog/2017/08/introduction-to-multi-label-classification/>

<http://www.r2d3.us/visual-intro-to-machine-learning-part-2/>

<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/>

<https://stats.stackexchange.com/questions/257182/why-does-my-classical-logistic-regression-model-perform-better-than-its-elastic>

<https://www.quantopian.com/lectures>

<https://doc.scrapy.org/en/latest/intro/tutorial.html>

<http://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html>

[**https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0**](https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0)

[**https://stackoverflow.com/questions/39163354/evaluating-logistic-regression-with-cross-validation**](https://stackoverflow.com/questions/39163354/evaluating-logistic-regression-with-cross-validation)

[**https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/**](https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/)

[**http://nbviewer.jupyter.org/github/bmabey/hacker\_news\_topic\_modelling/blob/master/HN%20Topic%20Model%20Talk.ipynb**](http://nbviewer.jupyter.org/github/bmabey/hacker_news_topic_modelling/blob/master/HN%20Topic%20Model%20Talk.ipynb)

[**https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6**](https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6)

[**https://dataskeptic.com/**](https://dataskeptic.com/)

[**http://lineardigressions.com/**](http://lineardigressions.com/)

[**http://datastori.es/**](http://datastori.es/)

[**http://nssdeviations.com/**](http://nssdeviations.com/)

[**https://www.oreilly.com/topics/oreilly-data-show-podcast**](https://www.oreilly.com/topics/oreilly-data-show-podcast)

[**http://www.learningmachines101.com/**](http://www.learningmachines101.com/)

[**http://www.thetalkingmachines.com/**](http://www.thetalkingmachines.com/)

[**https://twimlai.com/**](https://twimlai.com/)

[**http://partiallyderivative.com/**](http://partiallyderivative.com/)

[**https://www.datacamp.com/community/podcast**](https://www.datacamp.com/community/podcast)

**DEPLOYMENT**

[**https://xcitech.github.io/tutorials/heroku\_tutorial/**](https://xcitech.github.io/tutorials/heroku_tutorial/)

**html template responsive**

**https://**[**www.w3schools.com/**](http://www.w3schools.com/)

Heroku app

Django

Google – Progressive Web Apps – mimics smartphone apps, but app is not native

[**https://developers.google.com/web/progressive-web-apps/**](https://developers.google.com/web/progressive-web-apps/)

Yeo Wee Kiang [6:27 AM]

https://github.com/modin-project/modin/blob/master/README.md

GitHub

modin-project/modin

Modin: Speed up your Pandas workflows by changing a single line of code - modin-project/modin

Yeo Wee Kiang [12:02 PM]

https://chrome.google.com/webstore/detail/open-in-colab/iogfkhleblhcpcekbiedikdehleodpjo

chrome.google.com

Open in Colab

Open a Github-hosted notebook in Google Colab

martin [11:39 PM]

https://cloud.google.com/blog/products/ai-machine-learning/introducing-ai-hub-and-kubeflow-pipelines-making-ai-simpler-faster-and-more-useful-for-businesses

Google Cloud Blog

Introducing AI Hub and Kubeflow Pipelines: Making AI simpler, faster, and more useful for businesses | Google Cloud Blog

Introducing the AI Hub and Kubeflow Pipelines to help make AI and machine learning simpler and more useful for more businesses.

Sheena [2:01 PM]

Hi @kaien :slightly\_smiling\_face: awesome job in getting on top of project 1! I’m gonna send back the jupyter file with my comments in blue after every answer.

Please let me know if you have any questions!

Binary

proj\_1.ipynb

Binary

kaien [7:41 AM]

Thank you Sheena

Sheena [2:35 PM]

PDF

project2.pdf

508 kB PDF

kaien [3:08 PM]

Thank you Sheena for your comments :slightly\_smiling\_face:

Sheena [3:08 PM]

np! please do let me know if you have any question regarding the projects/topics/whatever :slightly\_smiling\_face:

kaien [3:15 PM]

ok, thanks will do. I like the topics and trying to pick up the knowledge as fast as I can but… in a busy work day that slow me down..

Sheena [4:19 PM]

that's great to hear :slightly\_smiling\_face: yeahh i know it can be tough, but I think the following lessons are a bit better to grasp now that you guys have a lot of the basics

or foundation topics

kaien [8:23 AM]

Good morning, Sheena

What functions do the two lines below have?

plt.stylr.use(‘fivethirtyeight’)

%config InlineBackend.figure\_format= ‘retina’

Sheena [12:23 PM]

Hi kai en. Sorry what do you mean?

For the first one, fivethirtyeight is a type of plot style which mimics fivethirtyeight.com

The second one i believe is the format type the figure would be

Hope that answers your question? :) if not can you explain what you’re trying to ask a little bit more?

kaien [8:59 PM]

Hi Sheena, thanks for reply.

I was referring to the command lines used in

classes/linear-regression/practice/solution-code/kobe-shots-made-regularization-lab-solutions.ipynb

kaien [9:09 PM]

I have no idea how could a website (fivethirtyeight.com) help in the Kobe data model.

And What can a ‘retina’ format do?

Sheena [11:18 PM]

It’s just a style. And that’s where it was derived from. :joy:

As for retina- means that you are using a format that supports retina display

I believe there is just an improvement in color display and all that - compatible for retina

kaien [11:37 PM]

thanks:)

Sheena [5:04 PM]

PDF

project3.pdf

319 kB PDF

kaien [1:00 PM]

Thank you , Sheena

Regarding smote Tomek, I don’t know how to apply to our model with reference to the example in https://imbalanced-learn.readthedocs.io/en/stable/auto\_examples/combine/plot\_smote\_tomek.html

The example uses make\_classification which already by itself created a model

Sheena [1:21 PM]

It created a classification dataset (X,y) with two classes. Not a classification model! So u can use it the same way as when u have X=df[[features]] and y=df[target]

:)

kaien [1:48 PM]

So when we use X =df[[features]] and y=df[target], we can replace make\_classification line, right ?

Sheena [2:26 PM]

Yup. :) cause make\_classification line basically gives the user a dummy classification data

So X and y values

kaien [4:02 PM]

Thanks

Will try tonight

Sheena [4:02 PM]

np! :slightly\_smiling\_face: if anything just slack me

if you are still having troubles we can work through it together

kaien [9:49 PM]

Hi, Sheena. I am testing one of the feature column called LIMIT\_BAL, but still can’t get the smote Tomek

Image from iOS

kaien [10:18 PM]

would you be able to advise?

Sheena [10:20 PM]

Hi kai en

For X, it has to be double brackets

So X= Credit[[‘LIMIT\_BAL’]]

Y is fine as it is

kaien [10:30 PM]

am I right?

Pasted image at 2018-10-10, 10:30 PM

Sheena [10:40 PM]

Yup :)

Oh and take out the make classification

Cause it overrides rhe X and y that u just called

kaien [10:42 PM]

hi, Sheena. after I took out the make\_clasification, the charts show nothing

Pasted image at 2018-10-10, 10:41 PM

the full content of my smote tomek:

import matplotlib.pyplot as plt

from sklearn.datasets import make\_classification

from sklearn.decomposition import PCA

from imblearn.combine import SMOTETomek

print(\_\_doc\_\_)

X= Credit[feature\_cols]

y= Credit.default\_nextMTH

# Generate the dataset

#X, y = make\_classification(n\_classes=2, class\_sep=2, weights=[0.22, 0.78],

#n\_informative=3, n\_redundant=1, flip\_y=0,

#n\_features=18, n\_clusters\_per\_class=1,

#n\_samples=30000, random\_state=10)

# Instanciate a PCA object for the sake of easy visualisation

pca = PCA(n\_components=2)

# Fit and transform x to visualise inside a 2D feature space

X\_vis = pca.fit\_transform(X)

# Apply SMOTE + Tomek links

sm = SMOTETomek()

X\_resampled, y\_resampled = sm.fit\_sample(X, y)

X\_res\_vis = pca.transform(X\_resampled)

# Two subplots, unpack the axes array immediately

f, (ax1, ax2) = plt.subplots(1, 2)

c0 = ax1.scatter(X\_vis[y == 0, 0], X\_vis[y == 0, 1], label=“Class #0”,

alpha=0.5)

c1 = ax1.scatter(X\_vis[y == 1, 0], X\_vis[y == 1, 1], label=“Class #1",

alpha=0.5)

ax1.set\_title(‘Original set’)

ax2.scatter(X\_res\_vis[y\_resampled == 0, 0], X\_res\_vis[y\_resampled == 0, 1],

label=“Class #0”, alpha=0.5)

ax2.scatter(X\_res\_vis[y\_resampled == 1, 0], X\_res\_vis[y\_resampled == 1, 1],

label=“Class #1", alpha=0.5)

ax2.set\_title(‘SMOTE + Tomek’)

# make nice plotting

for ax in (ax1, ax2):

ax.spines[‘top’].set\_visible(False)

ax.spines[‘right’].set\_visible(False)

ax.get\_xaxis().tick\_bottom()

ax.get\_yaxis().tick\_left()

ax.spines[‘left’].set\_position((‘outward’, 10))

ax.spines[‘bottom’].set\_position((‘outward’, 10))

ax.set\_xlim([-6, 8])

ax.set\_ylim([-6, 6])

plt.figlegend((c0, c1), (‘Class #0’, ‘Class #1’), loc=‘lower center’,

ncol=2, labelspacing=0.)

plt.tight\_layout(pad=3)

plt.show()import matplotlib.pyplot as plt

from sklearn.datasets import make\_classification

from sklearn.decomposition import PCA

from imblearn.combine import SMOTETomek

print(\_\_doc\_\_)

X= Credit[feature\_cols]

y= Credit.default\_nextMTH

# Generate the dataset

#X, y = make\_classification(n\_classes=2, class\_sep=2, weights=[0.22, 0.78],

#n\_informative=3, n\_redundant=1, flip\_y=0,

#n\_features=18, n\_clusters\_per\_class=1,

#n\_samples=30000, random\_state=10)

# Instanciate a PCA object for the sake of easy visualisation

pca = PCA(n\_components=2)

# Fit and transform x to visualise inside a 2D feature space

X\_vis = pca.fit\_transform(X)

# Apply SMOTE + Tomek links

sm = SMOTETomek()

X\_resampled, y\_resampled = sm.fit\_sample(X, y)

X\_res\_vis = pca.transform(X\_resampled)

# Two subplots, unpack the axes array immediately

f, (ax1, ax2) = plt.subplots(1, 2)

c0 = ax1.scatter(X\_vis[y == 0, 0], X\_vis[y == 0, 1], label=“Class #0”,

alpha=0.5)

c1 = ax1.scatter(X\_vis[y == 1, 0], X\_vis[y == 1, 1], label=“Class #1",

alpha=0.5)

ax1.set\_title(‘Original set’)

ax2.scatter(X\_res\_vis[y\_resampled == 0, 0], X\_res\_vis[y\_resampled == 0, 1],

label=“Class #0”, alpha=0.5)

ax2.scatter(X\_res\_vis[y\_resampled == 1, 0], X\_res\_vis[y\_resampled == 1, 1],

label=“Class #1", alpha=0.5)

ax2.set\_title(‘SMOTE + Tomek’)

# make nice plotting

for ax in (ax1, ax2):

ax.spines[‘top’].set\_visible(False)

ax.spines[‘right’].set\_visible(False)

ax.get\_xaxis().tick\_bottom()

ax.get\_yaxis().tick\_left()

ax.spines[‘left’].set\_position((‘outward’, 10))

ax.spines[‘bottom’].set\_position((‘outward’, 10))

ax.set\_xlim([-6, 8])

ax.set\_ylim([-6, 6])

plt.figlegend((c0, c1), (‘Class #0’, ‘Class #1’), loc=‘lower center’,

ncol=2, labelspacing=0.)

plt.tight\_layout(pad=3)

plt.show()

feature\_cols = [‘LIMIT\_BAL’,‘SEX’,‘EDUCATION’,‘MARRIAGE’,‘AGE’,‘PAY\_0’,‘PAY\_2’,‘PAY\_3’,‘PAY\_4’,‘PAY\_5’,‘PAY\_6’,‘BILL\_AMT1’,‘PAY\_AMT1’,‘PAY\_AMT2’,‘PAY\_AMT3’,‘PAY\_AMT4’,‘PAY\_AMT5’,‘PAY\_AMT6’]

Sheena [10:44 PM]

Hi Kai en.. again, u need to double bracket your X :)

Did you call your feature\_cols before the first time you assigned it to X?

Image from iOS

kaien [10:50 PM]

I called the feature\_cols in a cell before this one.

after using double bracket, I get the error message:

Automatically created module for IPython interactive environment

---------------------------------------------------------------------------

KeyError Traceback (most recent call last)

<ipython-input-402-c4953efc737a> in <module>()

7 print(\_\_doc\_\_)

8

----> 9 X= Credit[[feature\_cols]]

10 y= Credit.default\_nextMTH

11

~/anaconda3/lib/python3.6/site-packages/pandas/core/frame.py in \_\_getitem\_\_(self, key)

2677 if isinstance(key, (Series, np.ndarray, Index, list)):

2678 # either boolean or fancy integer index

-> 2679 return self.\_getitem\_array(key)

2680 elif isinstance(key, DataFrame):

2681 return self.\_getitem\_frame(key)

~/anaconda3/lib/python3.6/site-packages/pandas/core/frame.py in \_getitem\_array(self, key)

2721 return self.\_take(indexer, axis=0)

2722 else:

-> 2723 indexer = self.loc.\_convert\_to\_indexer(key, axis=1)

2724 return self.\_take(indexer, axis=1)

2725

~/anaconda3/lib/python3.6/site-packages/pandas/core/indexing.py in \_convert\_to\_indexer(self, obj, axis, is\_setter)

1325 if mask.any():

1326 raise KeyError(‘{mask} not in index’

-> 1327 .format(mask=objarr[mask]))

1328

1329 return com.\_values\_from\_object(indexer)

KeyError: “[(‘LIMIT\_BAL’, ‘SEX’, ‘EDUCATION’, ‘MARRIAGE’, ‘AGE’, ‘PAY\_0’, ‘PAY\_2’, ‘PAY\_3’, ‘PAY\_4’, ‘PAY\_5’, ‘PAY\_6’, ‘BILL\_AMT1’, ‘PAY\_AMT1’, ‘PAY\_AMT2’, ‘PAY\_AMT3’, ‘PAY\_AMT4’, ‘PAY\_AMT5’, ‘PAY\_AMT6’)] not in index”

https://github.com/gkaien/GA18Aug/blob/master/Project4/Project-4.ipynb

GitHub

gkaien/GA18Aug

Test repo. Contribute to gkaien/GA18Aug development by creating an account on GitHub.

Sheena [10:59 PM]

Oh sorry you dont need the double bracket if feature\_cols is a list

Can u print X\_vis and show me what it looks like

kaien [11:03 PM]

Pasted image at 2018-10-10, 11:03 PM

Sheena [11:19 PM]

hi kai en:)

got it

sooo if you see your values are in like -xxxxxx.xxxxx

so just take this part out

ax.set\_xlim([-6, 8])

ax.set\_ylim([-6, 6])

and your image will show

Pasted image at 2018-10-10, 11:20 PM

kaien [11:23 PM]

GOT IT. thanks, Sheena

Pasted image at 2018-10-10, 11:23 PM

but when I check my target column, how come is still the same imbalanced. 22%:78%

23364 vs 6636

Sheena [11:56 PM]

what was the initial though?

if im not wrong is it not around 90+ to ~10%?

kaien [11:59 PM]

Initial was already at 23364 vs 6636

Sheena [12:55 AM]

Ohhh okay. Sorry. This is because your values do not change -> the values that you are using are the in the variable X\_resampled if you’re looking at smotetomek and X\_vis if you are using the smotetomek with pca :)

Sorry i didnt get to check my phone until now

kaien [6:49 AM]

its ok, Sheena. I was worrying to disturb you so late. may I know what is the difference with pca?

kaien [6:58 AM]

hi, Sheena. how can I check the column target ‘default\_nextMTH’ having balanced data counts using value\_counts() function?

Pasted image at 2018-10-11, 6:58 AM

so sorry to disturb you…

Sheena [11:14 AM]

So basically the one with pca makes your features only two columns because it compresses only the important features that you have and you indicated you want n\_components = 2.

kaien [11:27 AM]

Will smote Tomek overwrite my original target column such that it now become more balanced?

Sheena [11:28 AM]

Yup it should. When u called smote tek to fit\_transform your X,y it means that you are assigning them to new variables X\_resampled and y\_resampled

kaien [11:30 AM]

So based on the command: Credit.y\_resampled.value\_counts() will show me a balanced target column right?

Sheena [11:31 AM]

Dont think that’s how u would call it. Because y\_resampled would be an array so it would not have the value\_counts() column

kaien [12:46 PM]

Since the n\_components=2. It will be a 4x4 array right?

Sheena [6:55 PM]

No it's actually going to be 2 columns x the number of rows your initial data has

so basically what pca does is that it compresses your features into 2 components

as in two features

so the number of rows will be the same but your number of columns are compressed

kaien [7:38 PM]

Ok thanks . I will continue tonight

Sheena [8:58 PM]

Np!

kaien [10:42 PM]

Hi Sheena, before going through smote Tomek, my training and test scores are about 78% which is close to my imbalanced dataset.

After I’m done with smote Tomek, my training and test scores are lower at about 65%. Is this expected?

kaien [7:35 AM]

Hi, Sheena. after I dataframe the smote tomek X and y resampled. the columns look very weird. is this expected?

Pasted image at 2018-10-12, 7:34 AM

kaien [7:45 AM]

my latest progress: https://github.com/gkaien/GA18Aug/blob/master/Project4/Project-4.ipynb

GitHub

gkaien/GA18Aug

Test repo. Contribute to gkaien/GA18Aug development by creating an account on GitHub.

kaien [3:59 PM]

Hi Sheena, would be able to advise me? Sorry..., I’m really a noob when come to programming.

Sheena [5:02 PM]

Hi hi sorry

Ive neen sick since yesterday so i was pretty much knocked out today :(

Let me take a look :)

So for the question about scores - it’s normal if the scores are lower if your baseline becomes lower. What i mean is - if the score you are using is accuracy, your baseline at the start is 78%.. which means having a score of 78% is not great right? Because if u guessed 78% of the time, you’ll guess correctly that it’s not fraud. So after smote tomek, check what your baseline is. And basically how u do this is to check how your classes are now split. Is it now 50-50 as opposed to 78-22

kaien [5:17 PM]

Oh dear, hope you can get well soon. Sorry to take up your resting hours

Sheena [5:17 PM]

no worries

kaien [5:19 PM]

Sheena, I have checked the Resampled is 50-50 now. My rerun logistic regression score is now 61.9% for testing model, 61.5% for training model

Is the score of so much below 100% is good ?

Sheena [5:19 PM]

actually score that is close to 100% is good.. haha so it's not a very good model

but it is better than your first one before you did the smote tomek

kaien [5:20 PM]

:sob:

Sheena [5:20 PM]

in any case, think about your business problem

and by this i mean, for this case, what would the business want more? what do they want to maximize or "catch" in that sense

obviously most are not fraud right, so what they might want to look at is the ones that are

as for the dataframe for the smote tomek

you can do something like this if you havent already done so :slightly\_smiling\_face:

Pasted image at 2018-10-12, 5:23 PM

so try to explore your logistic regression - see what parameters you can work with and see if it helps your model.. That or you can also check if it makes sense to not do pca, and do feature importances instead! :slightly\_smiling\_face:

kaien [5:34 PM]

Hi, Sheena. I have created a dataframe for X and y resampled. the dataset now do not show the name of the headers. If I want to retrieve back headers naming, can I use rename function? or any other more efficient method?

Pasted image at 2018-10-12, 5:34 PM

Sheena [6:39 PM]

You can get the columns of your original dataset but make sure your y is the end

So u can do

Initialcol = df\_initial.columns

Then new\_df.columns = initialcol

Somethig like that :)

kaien [7:02 PM]

got it! thanks Sheena!

can I change the pca components number to get better score in theory?

Sheena [7:08 PM]

You can try that as well for sure. :)

kaien [7:41 PM]

Hmmm... PCA doesn’t help. I will try to drop somemore Low correlated to target features

kaien [9:06 PM]

Hi Sheena, do you know why the total counts in the confusion matrix is much lower than my dataset cases which is more than 40,000?

Pasted image at 2018-10-12, 9:06 PM

Sheena [9:12 PM]

This is because u split your data into train test

Which means that this is only x% of your dataset (depends how much testing percentage u put in the split)

kaien [10:00 PM]

thank you

kaien [10:29 AM]

Hi Sheena, do you have some time later? I have done get dummies, Smote tomek, train test split but the accuracy is still very poor. Hope to hear your advice.

https://github.com/gkaien/GA18Aug/blob/master/Project4/Project-4.ipynb

GitHub

gkaien/GA18Aug

Test repo. Contribute to gkaien/GA18Aug development by creating an account on GitHub.

Sheena [10:30 AM]

I think if we end relatively on time I can sit with you a bit. I have something to go to after class but should be ok!

let's look at it later :slightly\_smiling\_face:

kaien [10:31 AM]

ok thanks alot, Sheena. talk to you later

Sheena [10:31 AM]

np :slightly\_smiling\_face:

kaien [8:17 AM]

Good morning Sheena, my highest test model score is 87%.

What else can I do if in commercial work we can only allow default rate of 1%. That means test model score need to be around E.g. 97%

Sheena [9:17 AM]

What metric did u use!

kaien [9:18 AM]

You mean classifier ? I use Random Forest

Sheena [9:19 AM]

No. As in what scorig metric

kaien [9:19 AM]

Confusion matrix

Sheena [9:39 AM]

Confusion matrix doesnt have a percentage score..

The model score u mentioned - what score is that?

kaien [9:52 AM]

Sorry are you referring to the classification report matrix?

Sheena [9:52 AM]

Haha nope. U said your highest model test score is 87%..

kaien [10:00 AM]

How does the metric you are referring to look like?

Sheena [10:00 AM]

Your highest model test score - what scoring method did u use

Sheena [10:30 AM]

Is the score from rf.score?

kaien [11:01 AM]

Yes

Sheena [11:17 AM]

Ohh okay. Did you use smote tomek data?

For that score?

kaien [11:38 AM]

Yes

Sheena [12:07 PM]

Okay. Cool. Then the statment doesnt make sense that commercial work can only allow a default rate of 1%

And this is because based on the actual data without manipulation - 1% might be the threshold

To have 99% accuracy for a model is not entirely great as well. So this should not be the threshold unless you have 90-10 split class to begin with

Im assuming this 1% is based on the fact that your data is 78-22 at the start - and so if ou have 87% then it is definitely not up to the standard. However, since you are assuming 50-50 split in classes, an 87% accuracy is high enough

kaien [12:47 PM]

Great advice

Thanks a lot Sheena

Without your responsive guidance I wouldn’t be able completing the final project

Cheers!

Sheena [12:47 PM]

Np!! Here to help. :)

All the best! Im sure you guys will do great on your presentation this coming saturday.

Last stretch!!! Jiayou :)

kaien [12:48 PM]

Thanks !!! :+1::sweat\_smile:

kaien [1:57 PM]

Good afternoon, Sheena. Though my model can achieve 87% scoring in the test data set. Is there a way I can export the accounts in each of the classifications?

The reason I ask is because I’m commercial world we need to identity who are prone to default

Sheena [2:00 PM]

Definitely. So your y\_pred will give u all your results that you model predicts as true or false

kaien [2:01 PM]

Oh yes! ;)

Thank you so much :blush:

Sheena [2:05 PM]

If you want to take out the accounts

Sheena [2:50 PM]

The only thing u have to do is combine the account column of your test data with the y\_pred - and then you can filter it out to show u the accounts that were predicted to have defaulted

kaien [3:15 PM]

So when the system is able to auto feed new data set and simultaneously self train to improve model scoring from time to time, can we then defined this as machine learning ?

Sheena [4:42 PM]

Yup :)