Personalized metabolomics pathway analysis

USER MANUAL

Web: http://lilikoi.garmiregroup.org

Github: https://github.com/lanagarmire/lilikoi

Technical support

Email: falakwaa@hawaii.edu

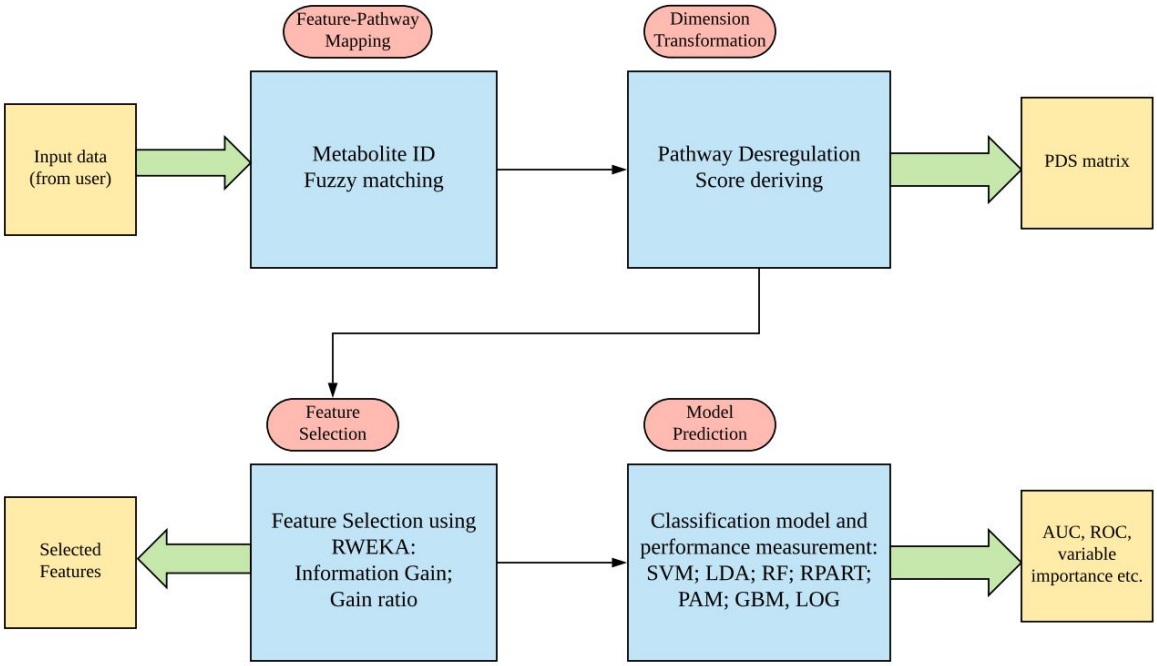
Contents

Preface

This document guides users to successfully use the Shiny version of the Lilikoi R package. Lilikoi is a novel tool for personalized pathway analysis of metabolomics data. Lilikoi is funded by K01ES025434 awarded by NIEHS, through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (www.bd2k.nih.gov), P20 COBRE GM103457 awarded by NIH/NIGMS, R01 LM012373 awarded by NLM, and R01 HD084633 awarded by NICHD to L.X. Garmire.

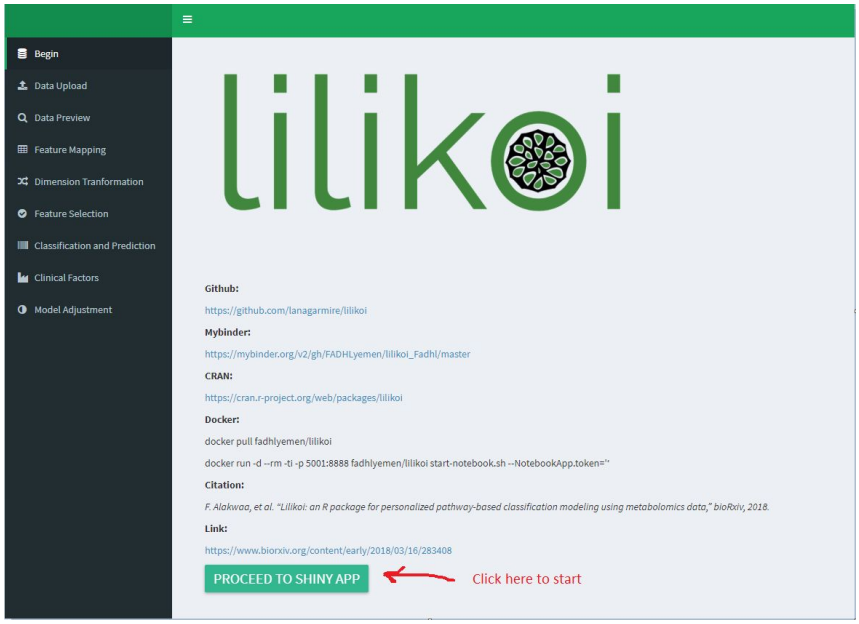
Introduction to Lilikoi:

Lilikoi (Hawaiian word for passion fruit) is a new and comprehensive R package for personalized pathway based classification modelling using metabolomics data. Four basic modules are presented as the backbone of the package: 1) Feature mapping module, which standardizes the metabolite names provided by users, and map them to pathways. 2) Dimension transformation module, which transforms the metabolomic profiles to personalized pathway-based profiles using pathway deregulation scores (PDS). 3) Feature selection module which helps to select the significant pathway features related to the disease phenotypes, and 4) Classification and prediction module which offers various machine-learning classification algorithms.

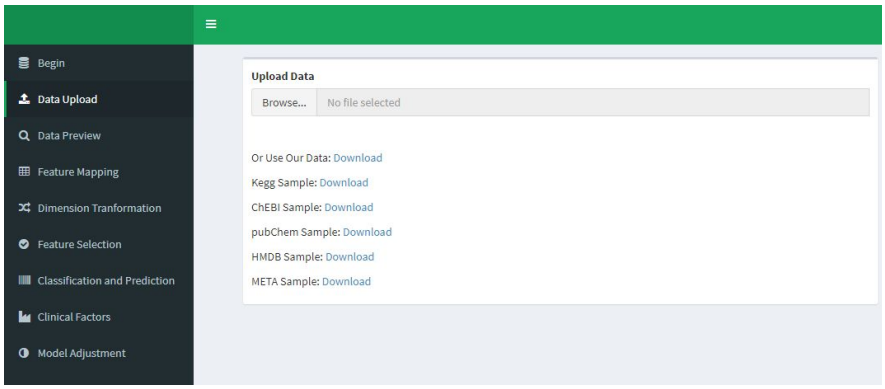


Users can access the Lilikoi Shiny version from the link below:

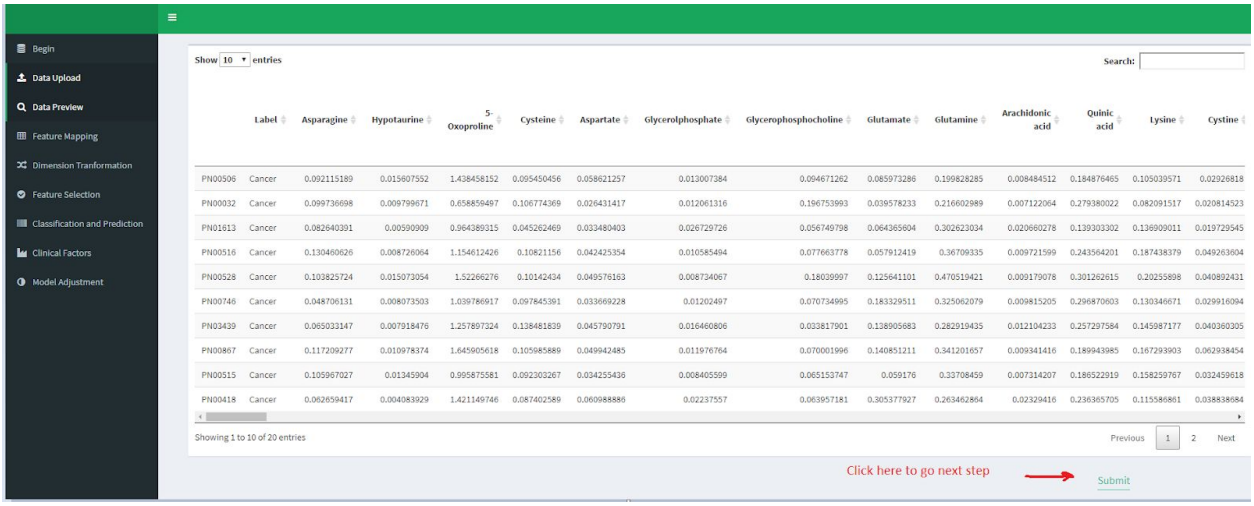
http://lilikoi.garmiregroup.org



You can upload your own data or try Lilikoi with our data samples. Information on how to format your data is below.

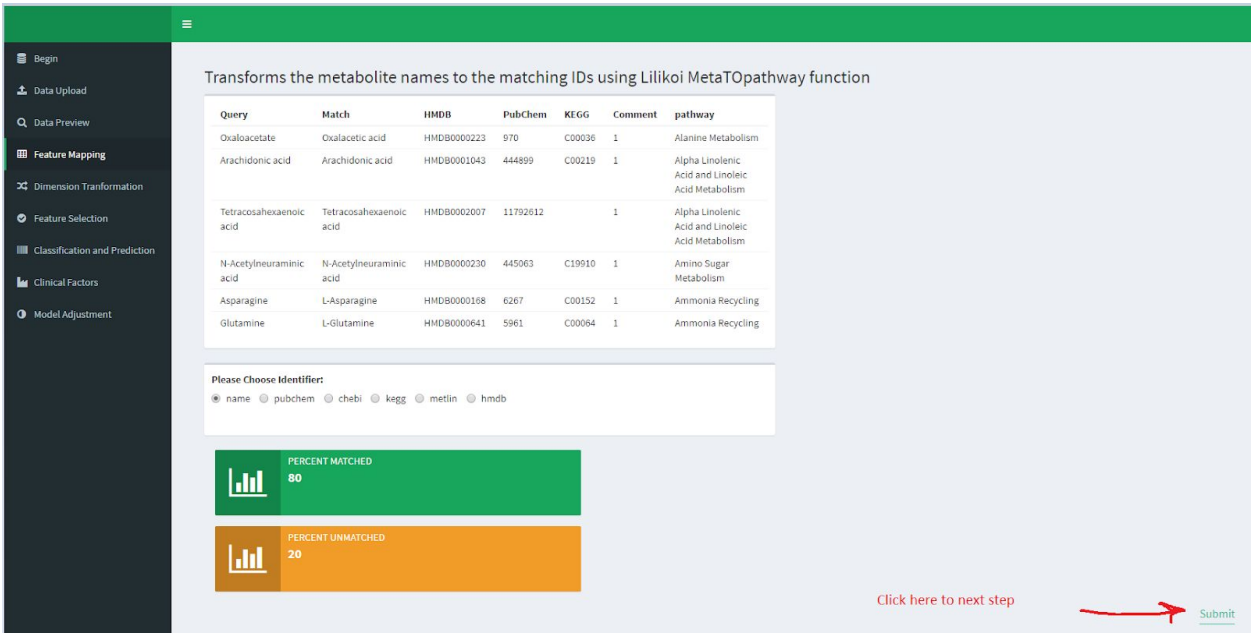


Your data should be in the below format which Lilikoi can use. Your samples should be in rows and your metabolites should be in the columns. The actual values should be integers or floats with the metabolite measurements. Both rows and columns should have names. The second column should be the label of the samples (case vs. control) from which the classifiers will be trained to predict.


Begin
Data Upload
Data Preview
Feature Mapping
Dimension Tranformation
Feature Selection
Classification and Prediction
Clinical Factors
Model Adjustment
Show 10 entries
Search:
Label
Asparagine
Hypotaurine
5-Oxoproline
Cysteine
Aspartate
Glycerolphosphate
Glycerophosphocholine
Glutamate
Glutamine
Arachidonic acid
Quinic acid
Lysine
Cystine
PN00506 Cancer 0.092115189 0.015607552 1.438458152 0.095450456 0.058621257 0.013007384 0.094671262 0.085973286 0.199828285 0.008484512 0.184876465 0.105039571 0.02926818
PN00032 Cancer 0.099736698 0.009799671 0.658856497 0.106774369 0.026431417 0.012061316 0.196753993 0.039578233 0.216602989 0.007122064 0.279380022 0.082091517 0.020814523
PN01613 Cancer 0.082640391 0.00590909 0.964389315 0.045262469 0.033480403 0.026729726 0.056749798 0.064365604 0.302623034 0.020660278 0.139303302 0.136909011 0.019729545
PN00516 Cancer 0.130460626 0.008726064 1.154612426 0.10821156 0.042425354 0.010585494 0.077663778 0.05791241 9 0.36709335 0.009721599 0.243564201 0.187438379 0.049263604
PN00528 Cancer 0.103825724 0.015073054 1.52266276 0.10142434 0.049576163 0.008734067 0.18039997 0.125641101 0.470519421 0.009179078 0.301262615 0.20255898 0.040892431
PN00746 Cancer 0.048706131 0.008073503 1.039786917 0.097845391 0.033669228 0.01202497 0.070734995 0.183329511 0.325062079 0.009815205 0.296870603 0.130346671 0.029916094
PN03439 Cancer 0.065033147 0.007918476 1.257897324 0.138481839 0.045790791 0.016460806 0.033817901 0.138905683 0.282919435 0.012104233 0.257297584 0.145987177 0.040360305
PN00867 Cancer 0.117209277 0.010978374 1.645905618 0.105985889 0.049942485 0.011976764 0.070001996 0.140851211 0.341201657 0.009341416 0.189943985 0.167293903 0.062938454
PN00515 Cancer 0.105967027 0.01345904 0.995875581 0.092303267 0.034255436 0.008405599 0.065153747 0.059176 0.33708459 0.007314207 0.186522919 0.158259767 0.032459618
PN00418 Cancer 0.062659417 0.004083929 1.421149746 0.087402589 0.060988886 0.02237557 0.063957181 0.305377927 0.263462864 0.02329416 0.236365705 0.115586861 0.038838684
Showing 1 to 10 of 20 entries
Previous 1 2 Next
Click here to go next step
Submit

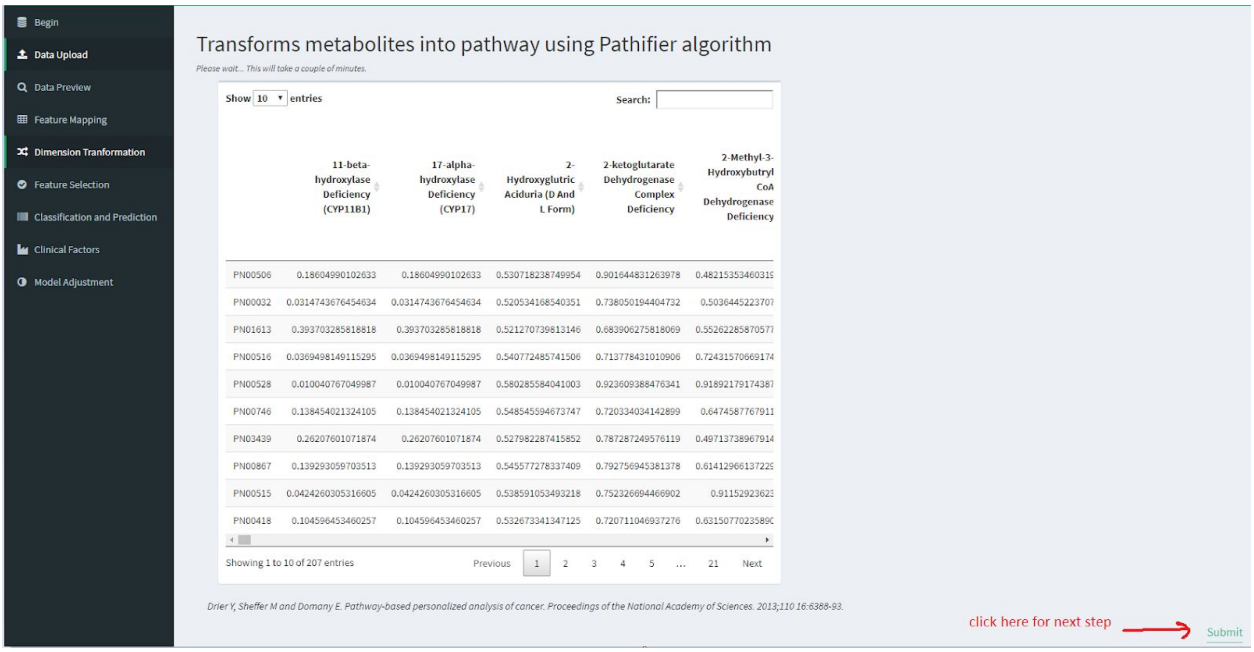
Part 1: Feature mapping module

When you press “submit” in the last figure you will go directly to the “feature mapping module”. In this module, all the metabolites are mapped to our curated database. Users have to select their data identifiers such as metabolites names, KEEG IDs, PubChem IDs, and HMDB IDs. The final result (see below figure) shows also the percentage of matched and unmatched metabolites. Unmatched metabolites will be filtered out in the next step.



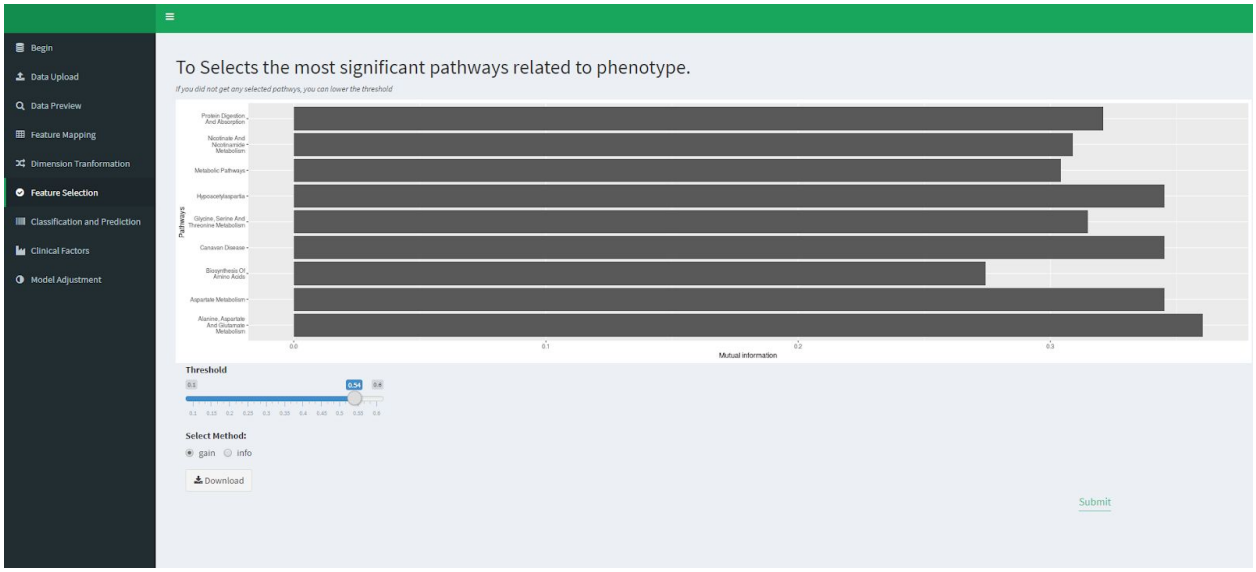
Part 2: Dimension transformation module

In this module, your metabolite samples matrix is converted to the pathways-samples matrix which is called the pathways deregulation score matrix (PDS). This step takes some time depending on the size of the input dataset. The calculation of these scores was performed using the Pathifier package. Please read our manuscript (the link in the resources section) to know more about how this conversion was performed.



Part 3: Feature selection

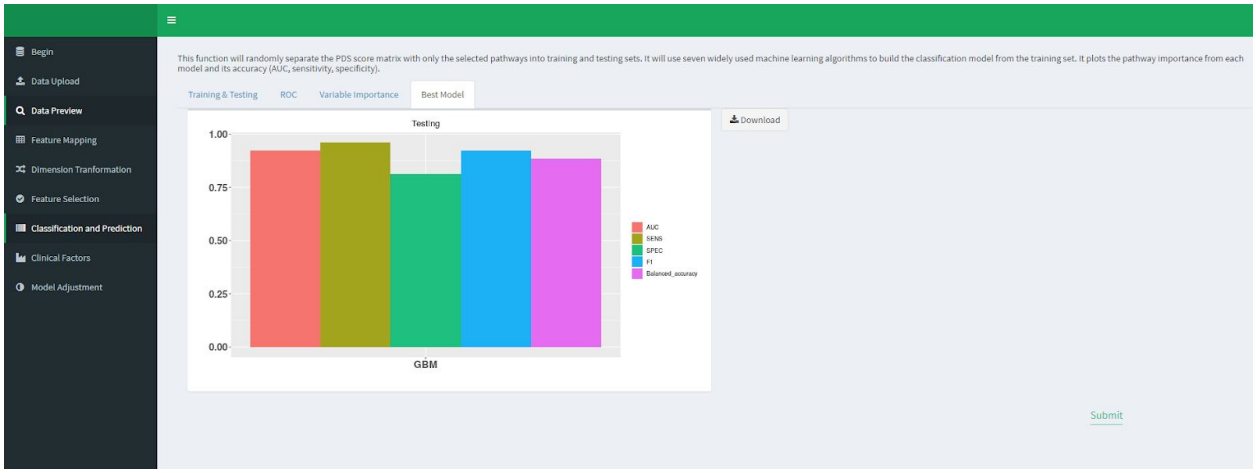
Not all the pathways in the PDS score matrix are very important. So using this module, we select only important pathways related to the phenotype. Users can adjust the threshold to limit the number of training features using one of the two available selection methods (gain or info). In the below example nine pathways are selected.



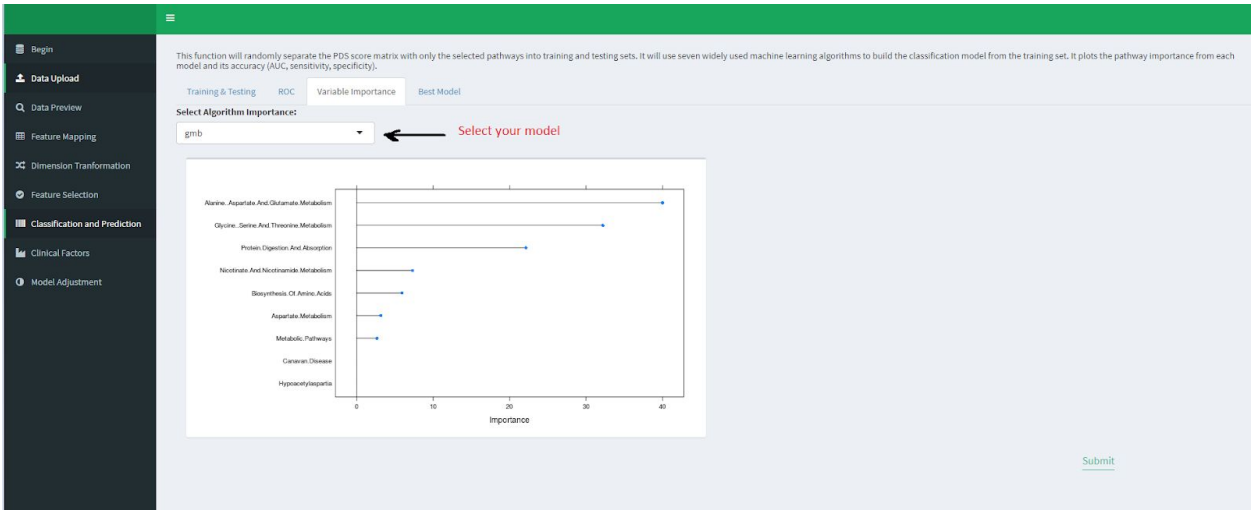
Part 4: Classification and prediction module

In this step, Lilikoi will build a model using the reduced PDS matrix. Lilikoi shows the model accuracy based on training (80%) and testing (20%) datasets. This module uses seven different machine learning algorithms and presents the results for each.

In this example, Lilikoi suggests that “GBM” is the best algorithm based on the accuracy result.



Below is the importance ranking of the nine selected pathways based on GBM model.


Begin
Data Upload
Data Preview
Feature Mapping
Dimension Tranformation
Feature Selection
Classification and Prediction
Clinical Factors
Model Adjustment
This function will randomly separate the PDS score matrix with only the selected pathways into training and testing sets. It will use seven widely used machine learning algorithms to build the classification model from the training set. It plots the pathway importance from each model and its accuracy (AUC, sensitivity, specificity).
Training & Testing
ROC
Variable Importance
Best Model
Select Algorithm Importance:
gmb
Select your model
Alanine..Aspartate.And.Glutamate.Metabolism
Glycine..Serine.And.Threonine.Metabolism
Protein.Digestion.And.Absorption
Nicotinate.And.Nicotinamide.Metabolism
Biosynthesis.Of.Amino.Acids
Aspartate.Metabolism
Metabolic.Pathways
Canavan.Disease
Hypoacetylaspartia
0
10
20
30
40
Importance
Submit

Resources:

Github: https://github.com/lanagarmire/Lilikoi

Mybinder: https://mybinder.org/v2/gh/FADHLyemen/Lilikoi_Fadhl/master

CRAN: https://cran.r-project.org/web/packages/Lilikoi

Docker:

```
docker pull fadhlyemen/Lilikoi
docker run -d --rm -ti -p 5001:8888 fadhlyemen/Lilikoi start-notebook.sh
--NotebookApp.token=''
```

Citation:

F. Alakwaa, et al. “Lilikoi: an R package for personalized pathway-based classification modeling using metabolomics data,” bioRxiv, 2018.

https://www.biorxiv.org/content/early/2018/03/16/283408