

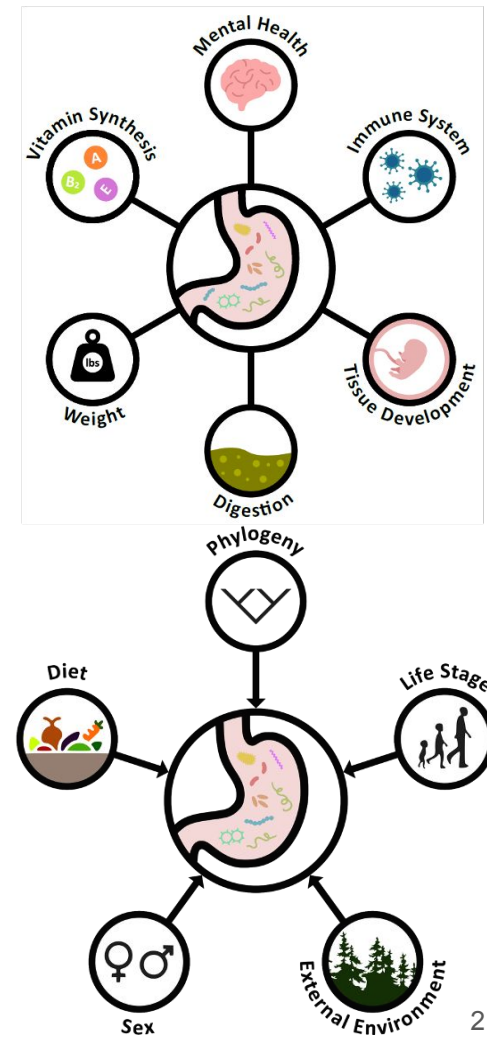
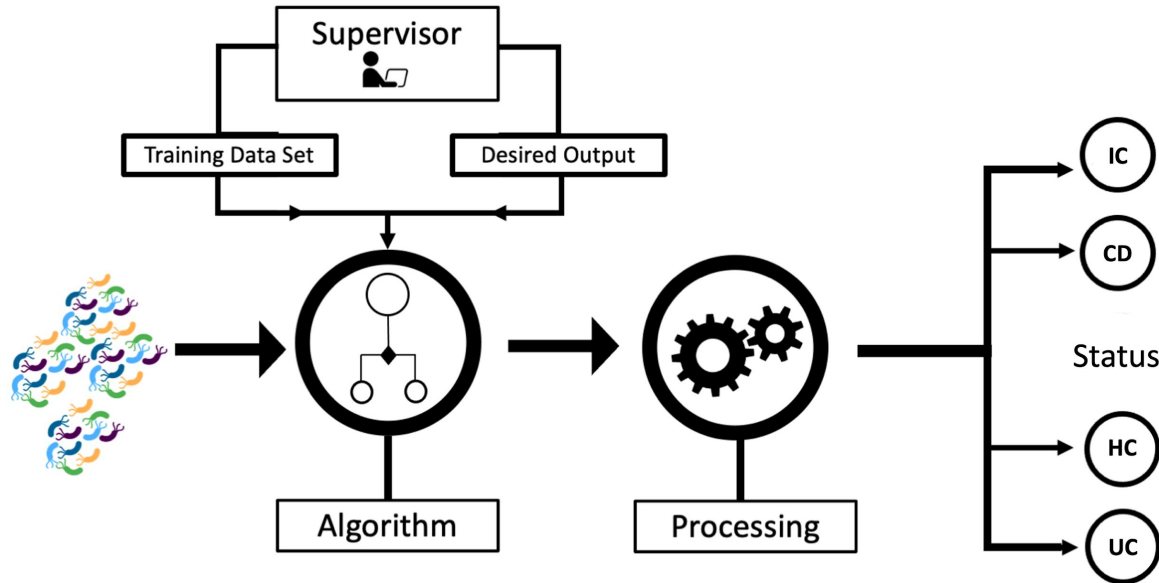
Supervised Machine Learning for Microbiome Data

Binnan Yu, Lillian Tatka, Kristina Herman, David Lee, Sierra Gillman

Introduction


The problem being addressed

Increasing accessibility to machine learning for microbiome research



The Data

4 tabundance ables of Operational Taxonomic Units (OTUs) and their metadata files



The Treatment-Naive Microbiome in New-Onset Crohn's Disease

Dirk Gevers,¹ Shihua Kugathasan,^{2,3,4} Lee A. Denson,^{5,6} Yoshiki Vázquez-Baeza,⁷ Will Van Treuren,¹ Boqin Ren,⁸ Emma Schwager,⁹ Dan Knights,^{1,10} De Jin Kong,¹ Marco Yassou,¹ Kaceli C. Morgan,¹ Aleksandar D. Kostic,¹ Chongren Luo,¹ Andrew Gonzalez,¹ Daniel McDonald,¹ Yael Hershman,¹ Thomas Walters,¹ Susan Baker,¹ Joel Rhee,¹¹ Michael Stephens,¹² Malen Heyman,¹³ James Markowitz,¹⁴ Robert Baldassano,¹⁵ Anne Griffiths,¹⁶ Francisco Snyvester,¹⁷ David Mack,¹⁸ Bianca Kim,¹⁹ Wallace Crowell,²⁰ Jeffrey Hyman,²¹ Curtis Huttenhower,¹ Rob Knight,^{1,2,3} and Harnik J. Xavier^{1,2,3*}

Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases

Jason Lloyd-Paet,¹ Cesar Aron,² Ashwin N. Ananthakrishnan,³ Melina Schirmer,⁴ Julian Anhe-Pacheco,⁵ Tiffany W. Peat,⁶ Elizabeth Andrews,⁷ Nadine I. Ajami,⁸ Kevin S. Busham,⁹ Colin J. Brislawn,¹⁰ David Gower,¹¹ Holly Gourtrey,¹² Antonio Gonzalez,¹³ Thomas C. Crowder,¹⁴ A. Bradley Karp,¹⁵ Katherine Lake,¹⁶ Camille I. Leclercq,¹⁷ James M. Macpherson,¹⁸ Danyal K. Pritchard,¹⁹ Mahadev Prasad,²⁰ Ghodratollah Rahnavard,²¹ Jonny Saiki,²² Dmitry Shugart,²³ Yoshiki Vázquez-Baeza,²⁴ Richard A. White III,²⁵ William K. Hargrett-Nelson,²⁶ Jonathan Bramer,²⁷ Lee A. Denson,²⁸ James K. Janssen,²⁹ Rob Knight,^{30,31} Shihua Kugathasan,³² Derron D. B. McGovern,³³ Joseph F. Petrosino,³⁴ Thaddeus S. Stappenbeck,³⁵ Harland S. Winter,³⁶ Gary B. Chisholm,³⁷ Eric A. Franzosa,³⁸ Hana Vlamacki,³⁹ Harnik J. Xavier,^{3,32,33} & Curtis Huttenhower^{3,32,33}

Cell Host & Microbe
Resource

BRIEF COMMUNICATION

<https://doi.org/10.1038/s41592-018-0141-9>

nature | methods

Qiita: rapid, web-enabled microbiome meta-analysis

Antonio Gonzalez^{1,12}, Jose A. Navas-Molina^{1,2,10,12}, Tomasz Kosciolk¹, Daniel McDonald¹, Yoshiki Vázquez-Baeza¹, Gail Ackermann¹, Jeff DeReus¹, Stefan Janssen¹, Austin D. Swafford³, Stephanie B. Orchanian³, Jon G. Sanders¹, Joshua Shorenstein^{1,11}, Hannes Holste^{1,12}, Semar Petrus⁴, Adam Robbins-Pianka⁵, Colin J. Brislawn⁶, Mingxun Wang⁷, Jai Ram Rideout⁸, Evan Bolyen⁸, Matthew Dillon⁸, J. Gregory Caporaso^{8,9}, Pieter C. Dorrestein^{1,3,7} and Rob Knight^{1,2,3*}

Qiita

nature microbiology

ARTICLES

PUBLISHED: 13 FEBRUARY 2017 | VOLUME: 2 | ARTICLE NUMBER: 17004

Dynamics of the human gut microbiome in inflammatory bowel disease

Jonas Halfvarson¹, Colin J. Brislawn², Regina Lamendella³, Yoshiki Vázquez-Baeza⁴, William A. Walters⁵, Lisa M. Bramer⁶, Mauro D'Amato^{7,8}, Ferdinando Bonfiglio⁹, Daniel McDonald¹⁰, Antonio Gonzalez¹¹, Erin E. McClure¹², Mitchell F. Dunkleberger³, Rob Knight^{4,10,11} and Janet K. Jansson^{2,*}

The Users



The Ecologist: Loves R

Is interested in seeing if habitat quality can be predicted based on fecal microbiome community composition in American marten (*Martes americana*). They want to determine if they can classify individuals from primary or disturbed habitat as this could be a powerful tool for conservation and management!



The Medical Clinician: Can google

Wants to be able to determine if suspected patients have inflammatory bowel disease so they can begin effective treatment before their health deteriorate, and will use BioME to classify people!



The Microbiologist: bioinformatic buff

Is here for the preliminary results/confirmation. The lab intern might have mislabeled some (~100) of the samples. Does say that a *Turdus turdis* or just *Turtle's turds*? They don't want to have to throw out all those samples..



The Users



The Ecologist: Loves R

The ecologist collects fecal samples from wild martens and puts the data into BioME where he sees that support vector classification is able to sort out the samples with high accuracy.



The Medical Clinician: Can google

Wants to be able to determine if suspected patients have inflammatory bowel disease so they can begin effective treatment before their health deteriorate, and will use BioME to classify people!



The Microbiologist: bioinformatic buff

Is here for the preliminary results/confirmation. The lab intern might have mislabeled some (~100) of the samples. Does say that a *Turdus turdis* or just *Turtle's turds*? They don't want to have to throw out all those samples..



The Users



The Ecologist: Loves R

The ecologist collects fecal samples from wild martens and puts the data into BioME where he sees that support vector classification is able to sort out the samples with high accuracy.



The Medical Clinician: Can google

BioME is used to implement a bowel disease classification model. A clinician enters the patient's data into BioME to classify their sample as healthy or diseased.



The Microbiologist: bioinformatic buff

Is here for the preliminary results/confirmation. The lab intern might have mislabeled some (~100) of the samples. Does say that a *Turdus turdis* or just *Turtle's turds*? They don't want to have to throw out all those samples..



The Users



The Ecologist: Loves R

The ecologist collects fecal samples from wild martens and puts the data into BioME where he sees that support vector classification is able to sort out the samples with high accuracy.



The Medical Clinician: Can google

The clinician gives BioME data from healthy patients and patients with inflammatory bowel disease. She then inputs the patient's data for BioME to classify as healthy or diseased.

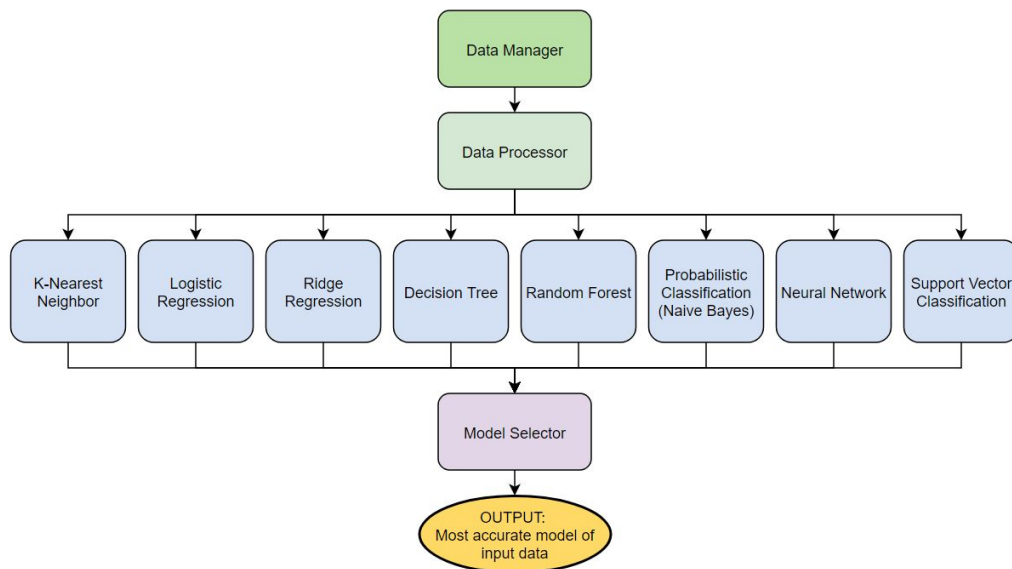


The Microbiologist: bioinformatic buff

The microbiologist gives sample data to BioME to verify the identity of mislabeled samples. Later, he gives his verified data to BioME to compare models.

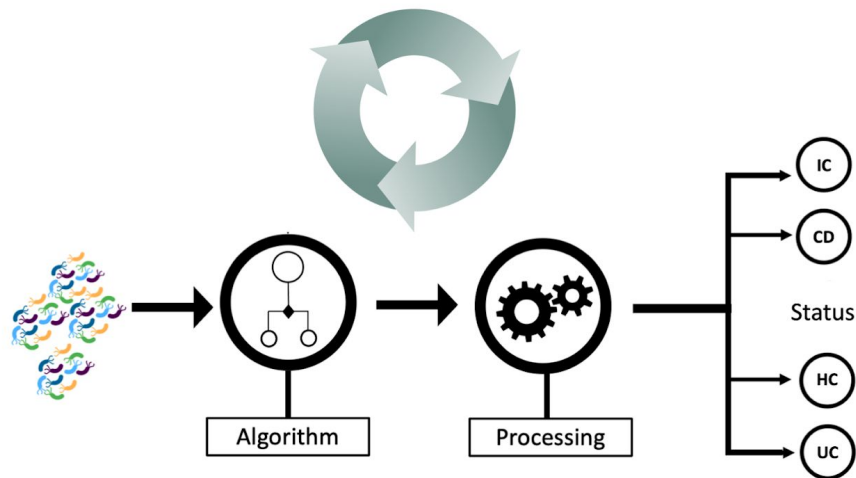
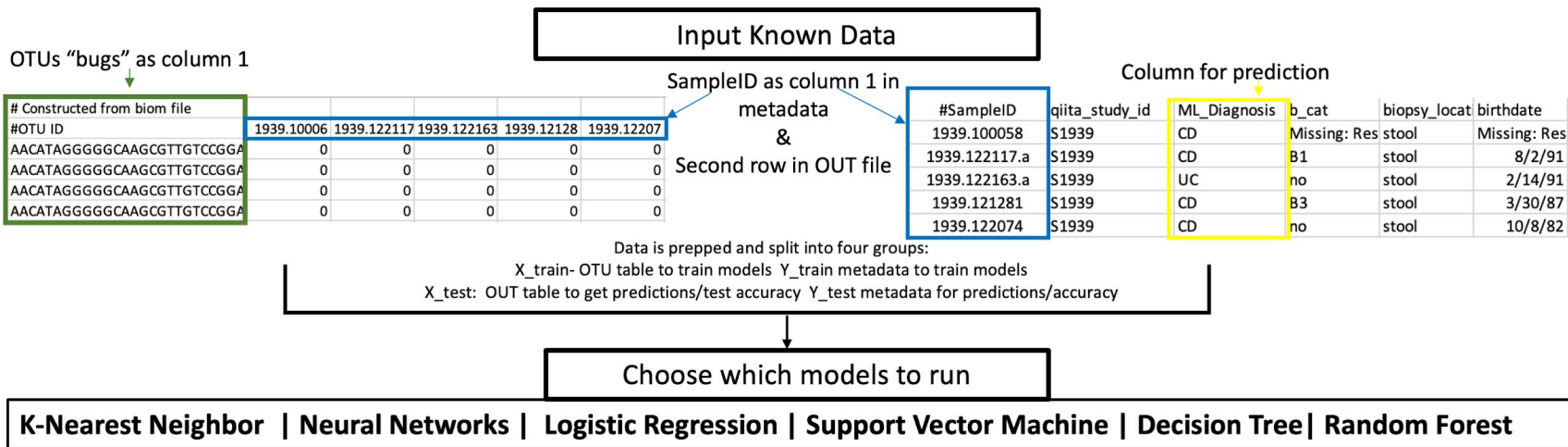


Functional Specifications



BioME/

- | - README.md
- | - biome/
 - | - __init__.py
 - | - prep_split_data.py
 - | - select_model.py
 - | - train_mlp.py
 - | - knn.py
 - | - dtree.py
 - | - logistic.py
 - | - ridge.py
 - | - random.py
 - | - SVC.py
 - | - naive_bayes.py
 - | - scripts/
 - | - biome_run.py
 - | - tests/
 - | - ...
 - | - Data/
 - | - bug_OTU_rel.tsv
 - | - bug_OTU_raw.tsv
 - | - FecesMeta.txt
 - | - query_point.tsv
- | - docs/
 - | - FunctionalSpec.md
 - | - ComponentSpec.md
 - | - Technology Review Presentation
 - | - images/
 - | - ...
- | - setup.py
- | - LICENSE
- | - BioME_environment.yml



[illegible]

Please enter the relative path to the categorical data: biome/Data/FecesMeta.txt

What models would you like to test?
See README.md for abbreviations. Type all if all models should be tested: mlp1,mlp3,dtree,knn,gnb

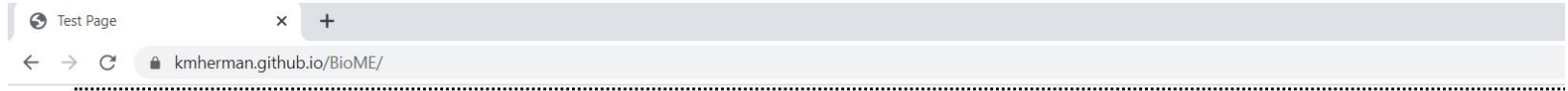
The best performing model is: MLP (single hidden layer)

Please enter the path to the data that you would like to make a prediction for: biome/Data/query point.tsv

Would you like to make another prediction? no

Future Work

Create a webapp that allows user to select a model they want to run

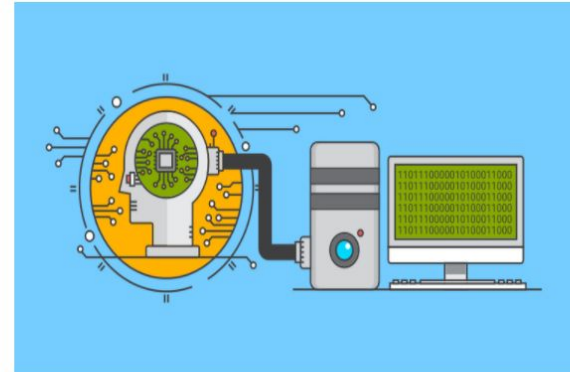
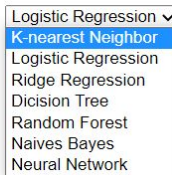


Why Our Solution is Trustworthy?

1. Algorithm

Our product is developed based on 7 different algorithms, each of it have different characteristics so that our product can handle different dataset.

Click the dropdown list to see algorithms option we provide



Supervised Machine Learning for Microbiome Data

Kristina Herman¹ Lillian Tatka² David Lee² Sierra Gillman³ Binnan Yu^{4,5}

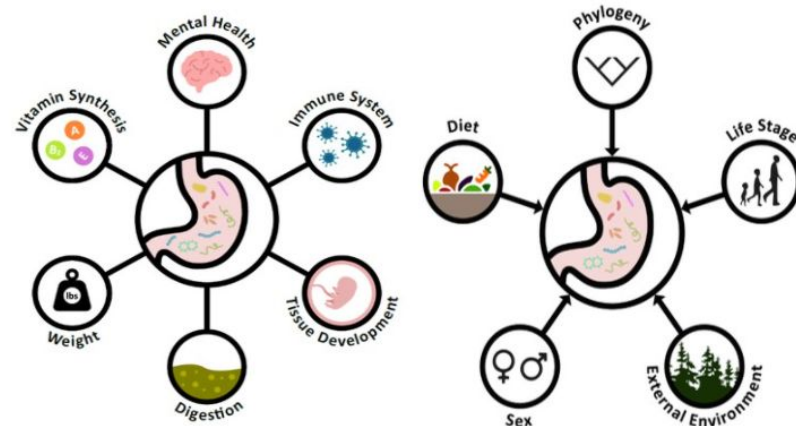
¹Department of Chemistry; ²Department of Bioengineering; ³School of Environmental and Forest Sciences;

⁴Department of Rehabilitation Medicine; ⁵Department of Applied Mathematics, University of Washington, Seattle, WA, 98105

Statement of The Problem

Once thought only to be pathogenic, the microorganisms living on and within an animal host are now recognized as playing critical roles in host health. Factors that shape gut microbial communities are multifaceted and include the host's diet and life-stage.

While microbial shifts have been implicated in numerous human ailments (e.g., obesity, anxiety, inflammatory bowel disease), research has thus far been limited to differentiating microbial communities between groups and less for predictive uses. As a result of an ever expanding microbiome data availability, microbiome research lends itself to advancement with machine learning and data science. However, the implementation of machine learning in microbiome research might feel daunting and time consuming to those outside of the realm of data science.



Thank You!