

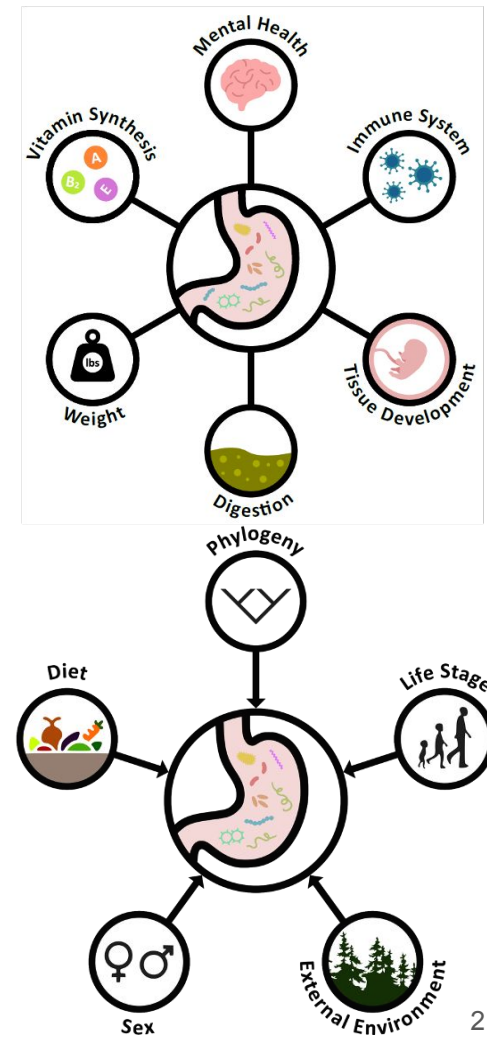
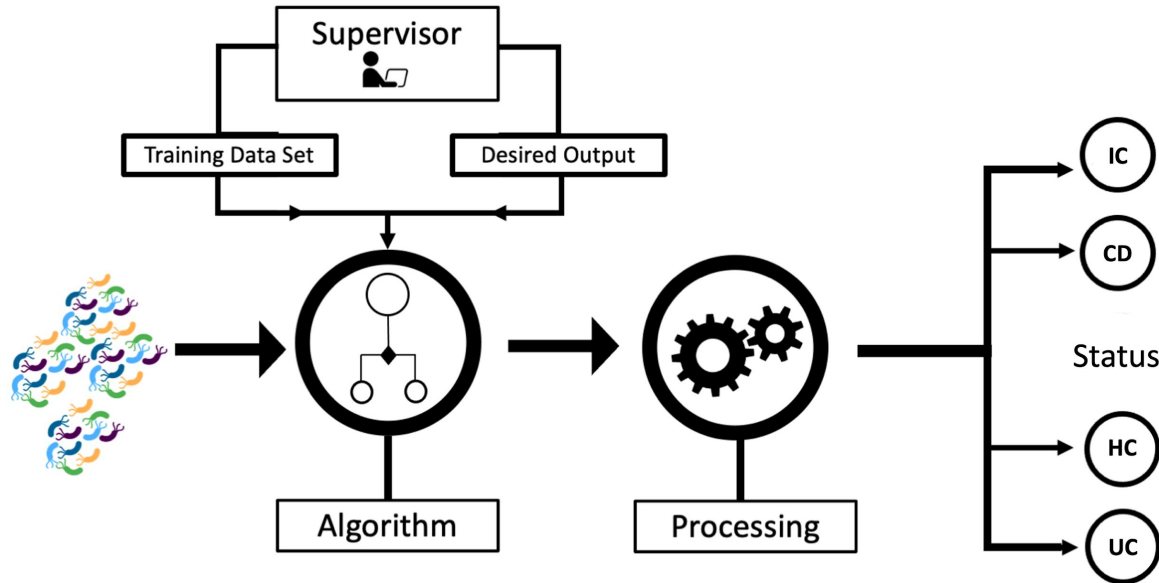
Supervised Machine Learning for Microbiome Data

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 Yassani,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 Sylvestre,¹⁷ David Mack,¹⁸ Bianca Kim,¹⁹ Wallace Clements,²⁰ Jeffrey Hyman,²¹ Curtis Huttenhower,¹ Rob Knight,^{1,2,3} and Harnik J. Xavier^{1,2,3*}



BRIEF COMMUNICATION


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





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

Jason Lloyd-Paet,¹ Cesar Aron,² Ashwin N. Ananthakrishnan,³ Melinda Schrimmer,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. Groves,¹⁴ A. Bradley Lippell,¹⁵ Karlijn Lake,¹⁶ Camille I. Landow,¹⁷ James M. Leach,¹⁸ Danyal K. Plicht,¹⁹ Mahadev Prasad,²⁰ Ghodanah Ramanan,²¹ Jonny Saik,²² Dmitry Shugart,²³ Yoshiki Vázquez-Baeza,²⁴ Richard A. White III,²⁵ William K. Wernke,²⁶ Jonathan Wernke,²⁷ Lee A. Denson,²⁸ James K. Janssen,²⁹ Rob Knight,^{30,31} Shihua Kugathasan,³² Denson 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*}



Dynamics of the human gut microbiome in inflammatory bowel disease

Jonas Halfvarson,¹ Colin J. Brislawn,² Regina Lamendella,³ Yoshiki Vázquez-Baeza,⁴ William A. Walters,⁵ Lisa M. Bramer,⁶ Mauro D'Amato,^{7,8} Ferdinando Bonfiglio,⁹ Daniel McDonald,¹⁰ Antonio Gonzalez,¹¹ Erin E. McClure,¹² Mitchell F. Dunkleberger,¹³ Rob Knight,^{14,15,16} 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

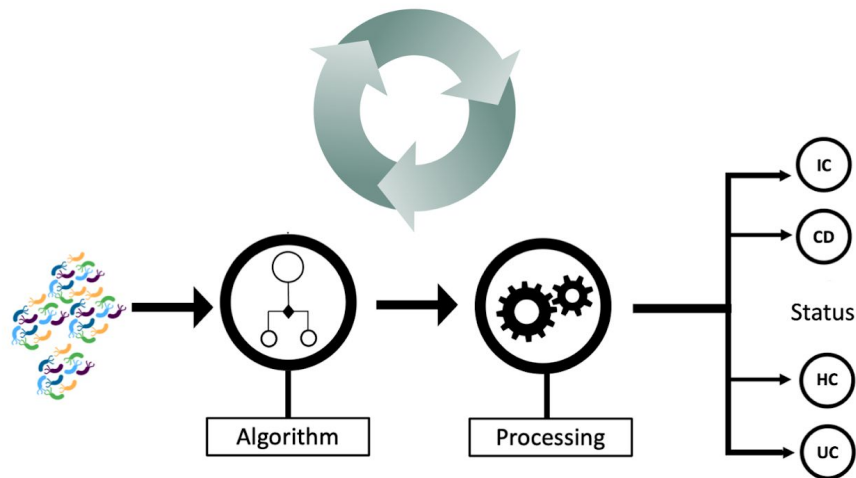
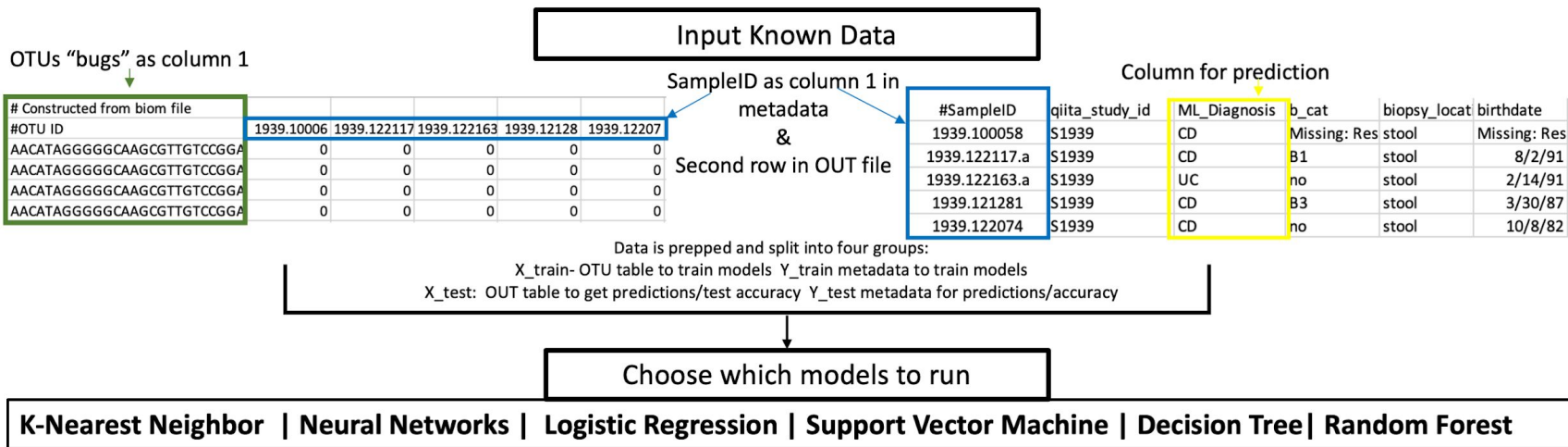
Wants to be able to determine if suspected patients have inflammatory bowel disease so they can begin effective treatment before their health deteriorate, and will use BioME to classify people!



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..

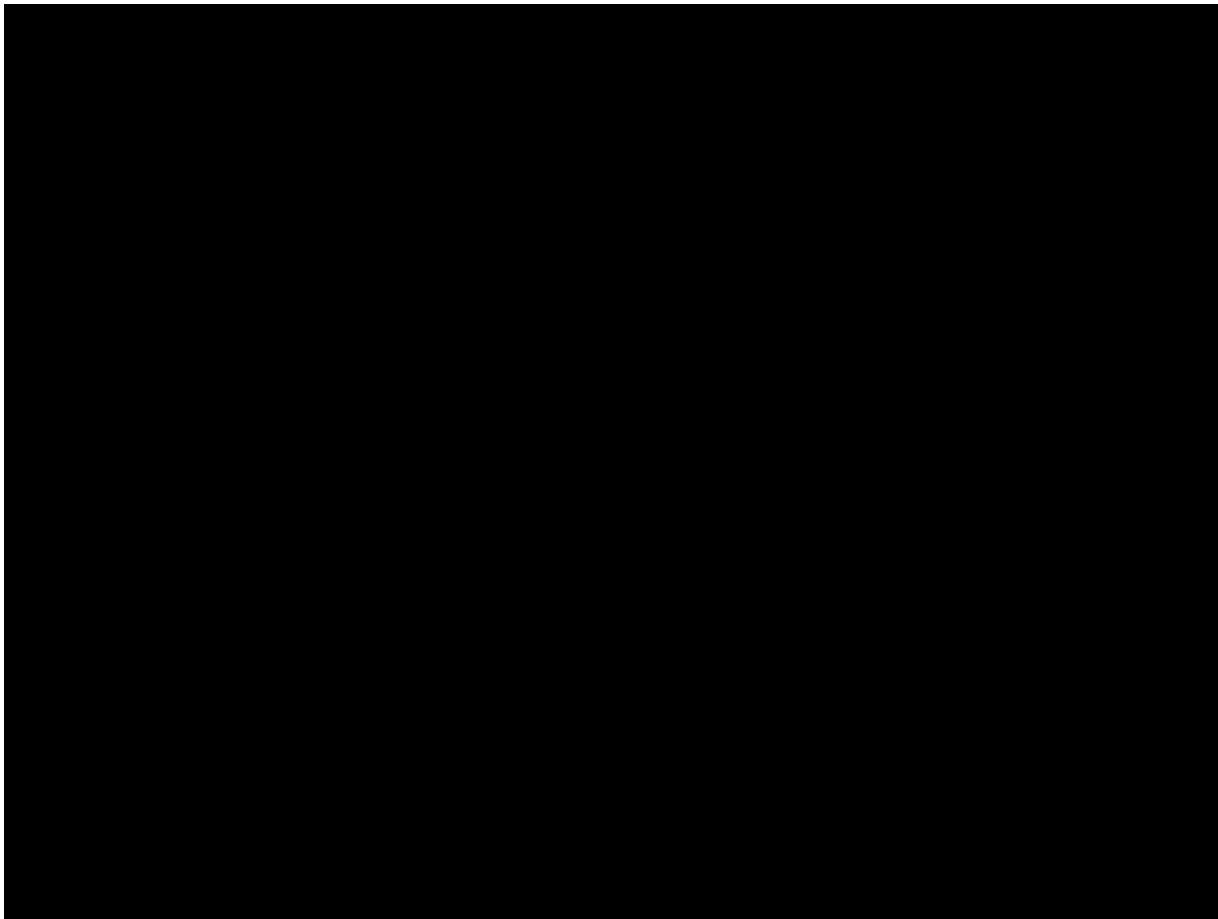




BioME Demo

BioME/

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



Lessons learned

- If we were to do this project again, we would...
 - Get Travis CI up and running earlier
 - Keep up with writing unittests as the code is being written
 - Keep up with updating documentation
 - Finish code earlier (because there is a lot to finish once the actual code is functional!)