



Microbiome

Geolocation

Data

User-selected
Preprocessing
Method

User-selected
Accuracy (A)

Random-Forest based ML
training with partial data,
followed by testing with
hidden data to provide
accuracy A' for instance N'

$A' \geq A$

Yes

No

Report Bioindicator species
based on average weight
associated for each microbe

User-selected
number of
instance N

$N' < N$

Yes

Select tree and the
weight associated to
each microbe

