COVID-19

Primary Contributing Factors for Contraction
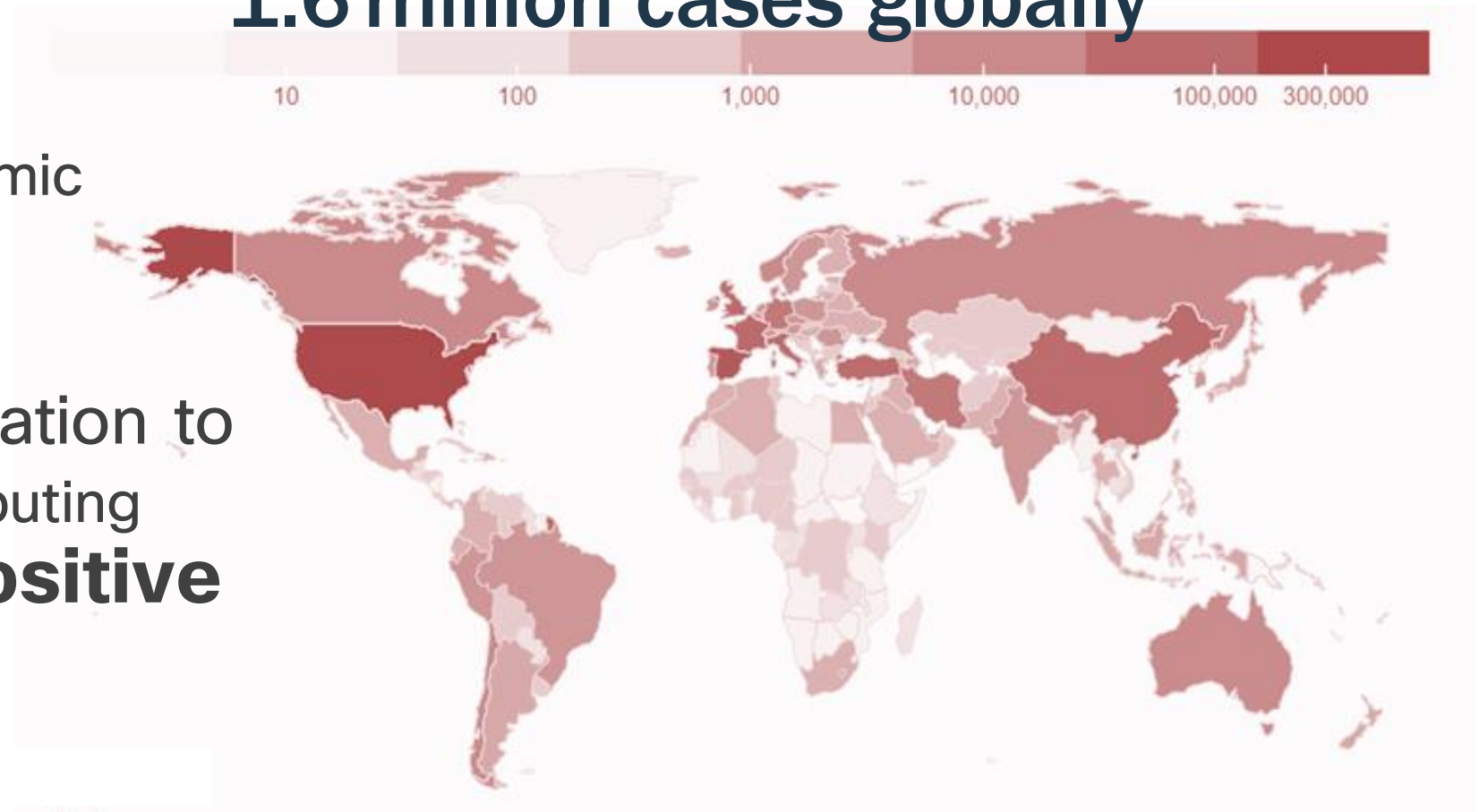
Presented by: Andi Osika

Background:

April 10, 2020:
1.6 million cases globally

- COVID-19: Pandemic

- Goal: Use classification to Identify primary contributing factors for **testing positive** for the disease.

Factor Datatypes Collected:

Biological	Behavioral	Environmental
Age	Drug(prescription/non) Use	Number of Contacts
Sex	Alcohol Use	Co-Habitation Numbers
BMI	Hand Washing, Social Distancing	Travel Habits
Preexisting Health Conditions	Mask Wearing Outside of Home	

Data Background:

- **Nexoid**

- **618K+ Self-Reported Samples**
  - Top Countries: US, Canada and UK

2.3%

5%

87%

COVID-19 Rates

Percent

.13% Tested Positive

Negative ::: Positive
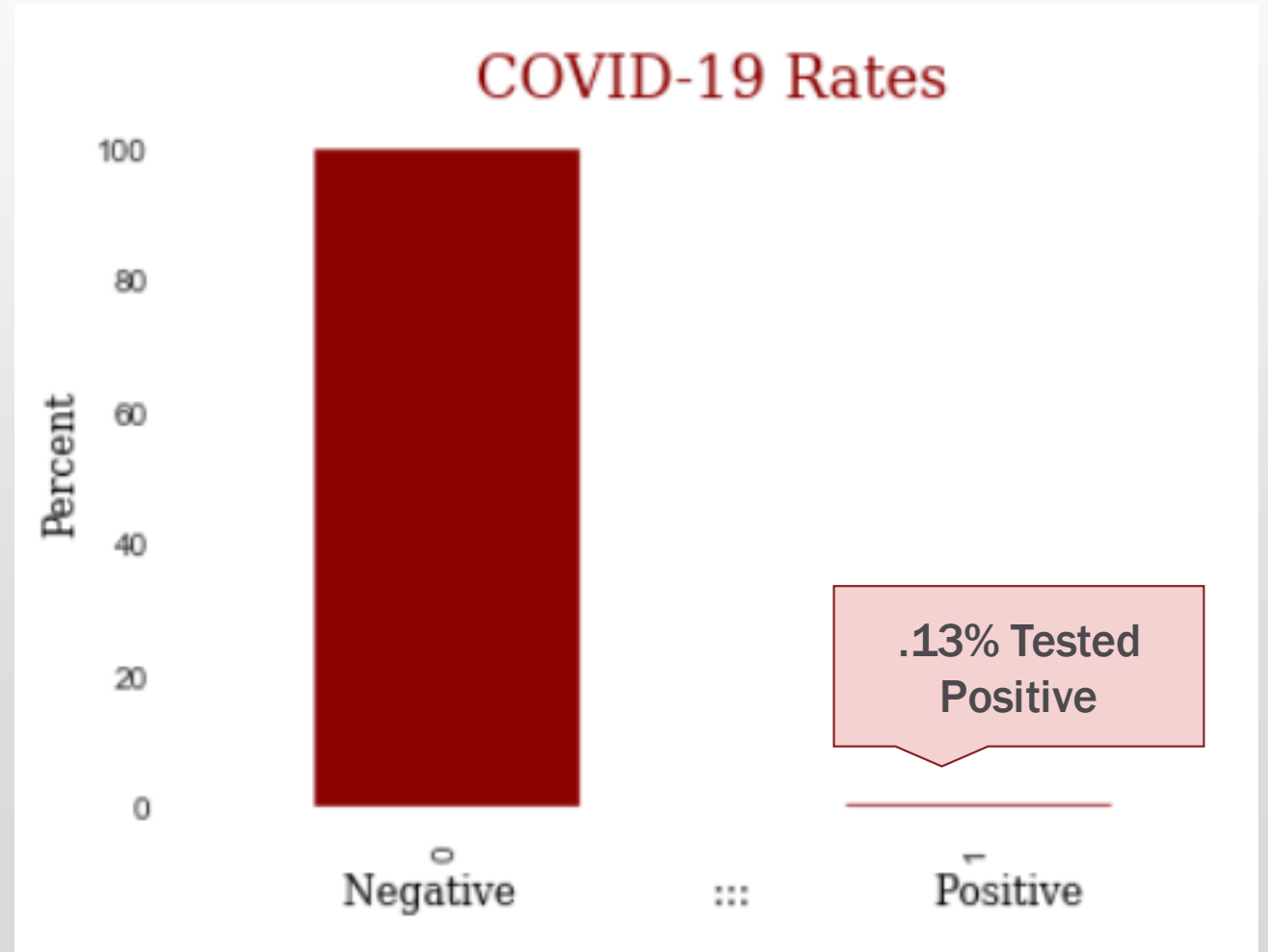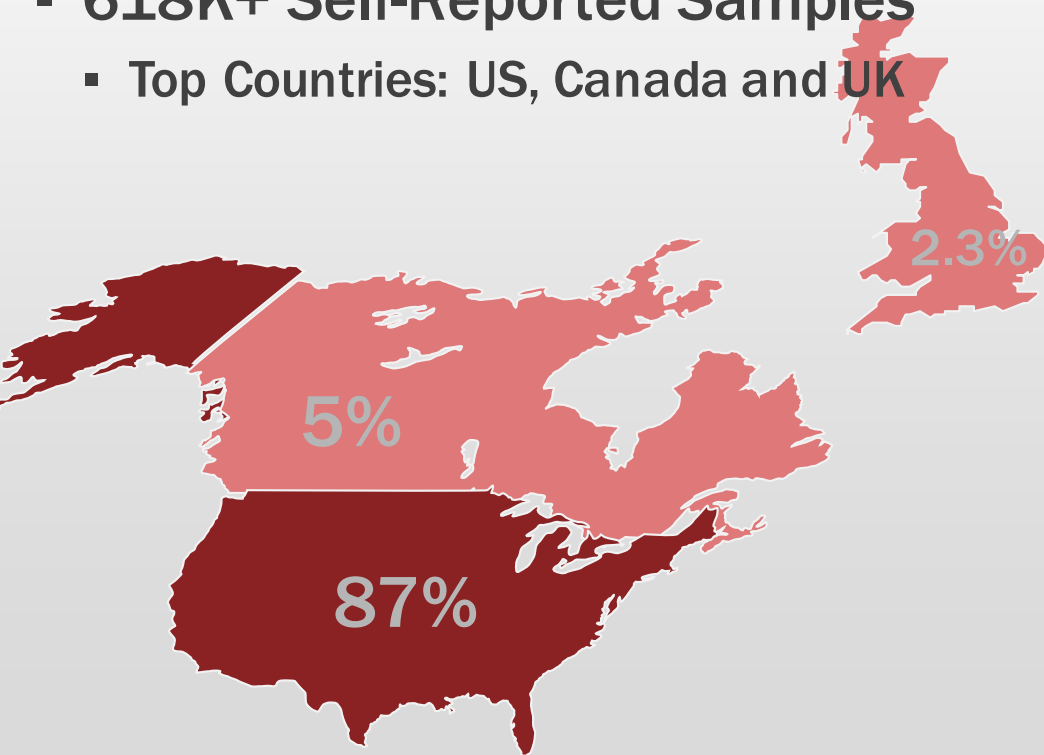
Methodology:

Classification Modeling: Random Forest

Random Forest Classification:

Decision Tree:

Biological, Behavioral, Environmental

Data

(COVID-19 +)
Practicing Social Distancing & Hygiene
(low score)

TRUE

FALSE

(Covid-19 +)
Other Feature

(Covid-19 -)
Other Feature
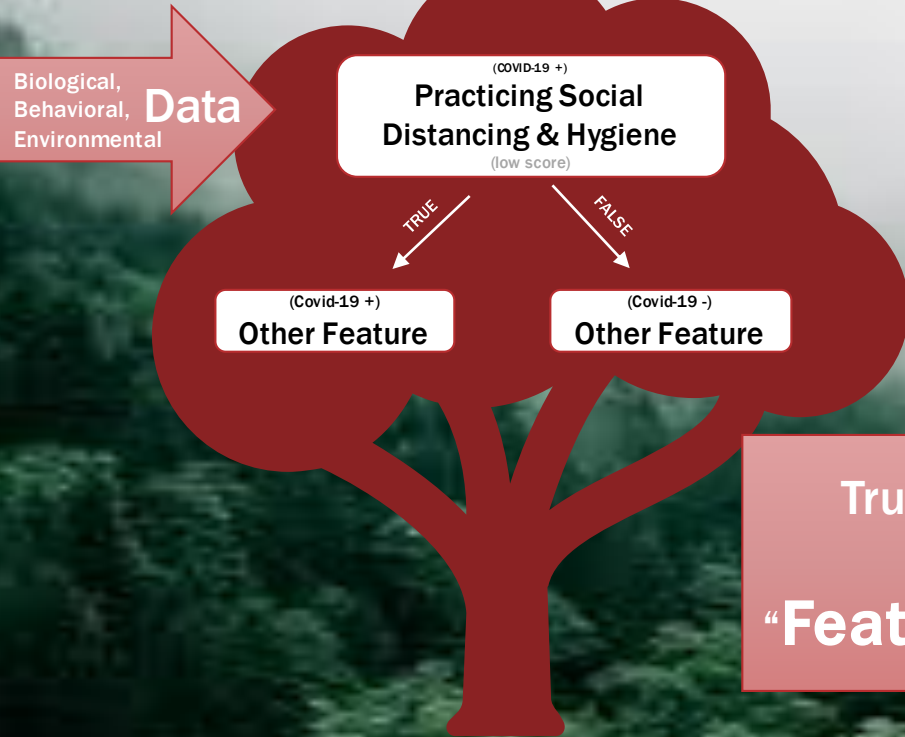
True/False Decisions Develop
"Feature Importance"

Many Decision Trees

FINDINGS:

TOP FACTORS THAT CONTRIBUTE TO CONTRACTING COVID19

Modeling Results from Random Forest:
95.8% Accuracy

Feature Importance in Predicting the Testing Positive for COVID-19
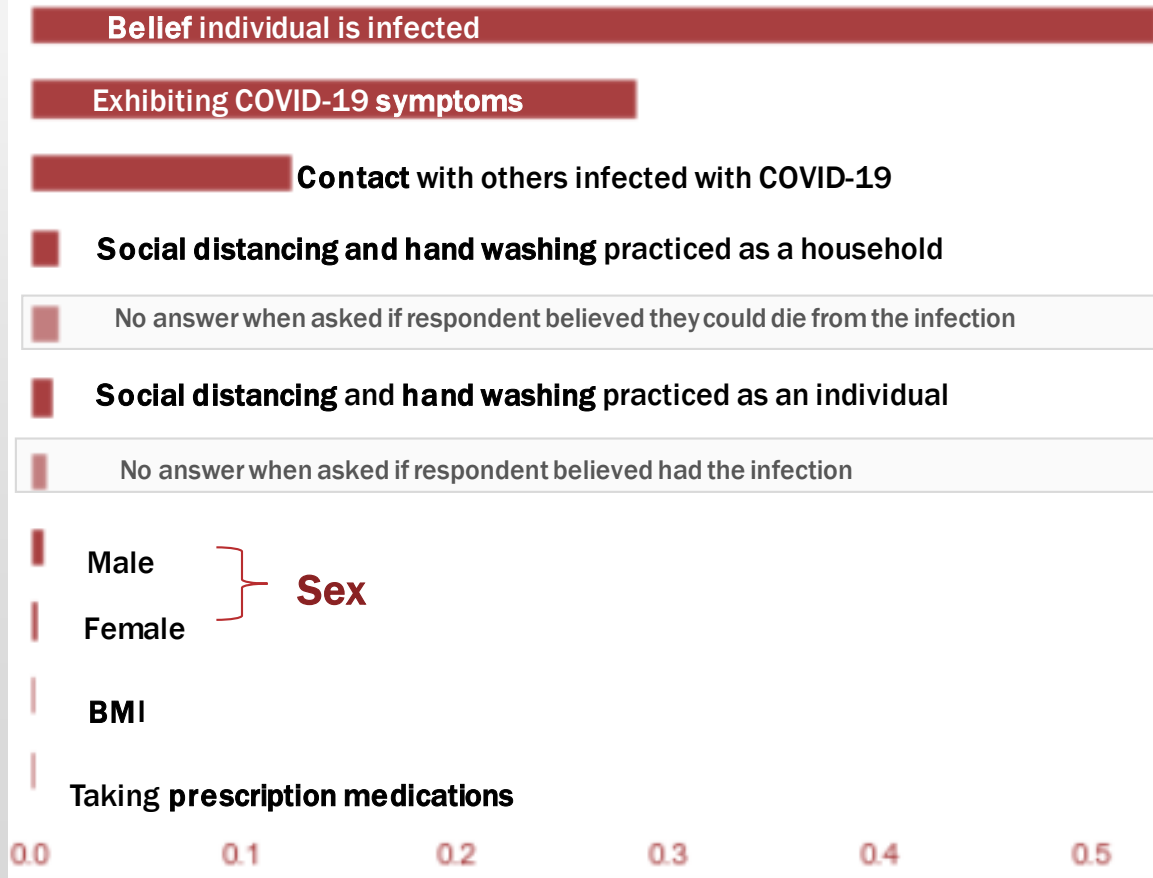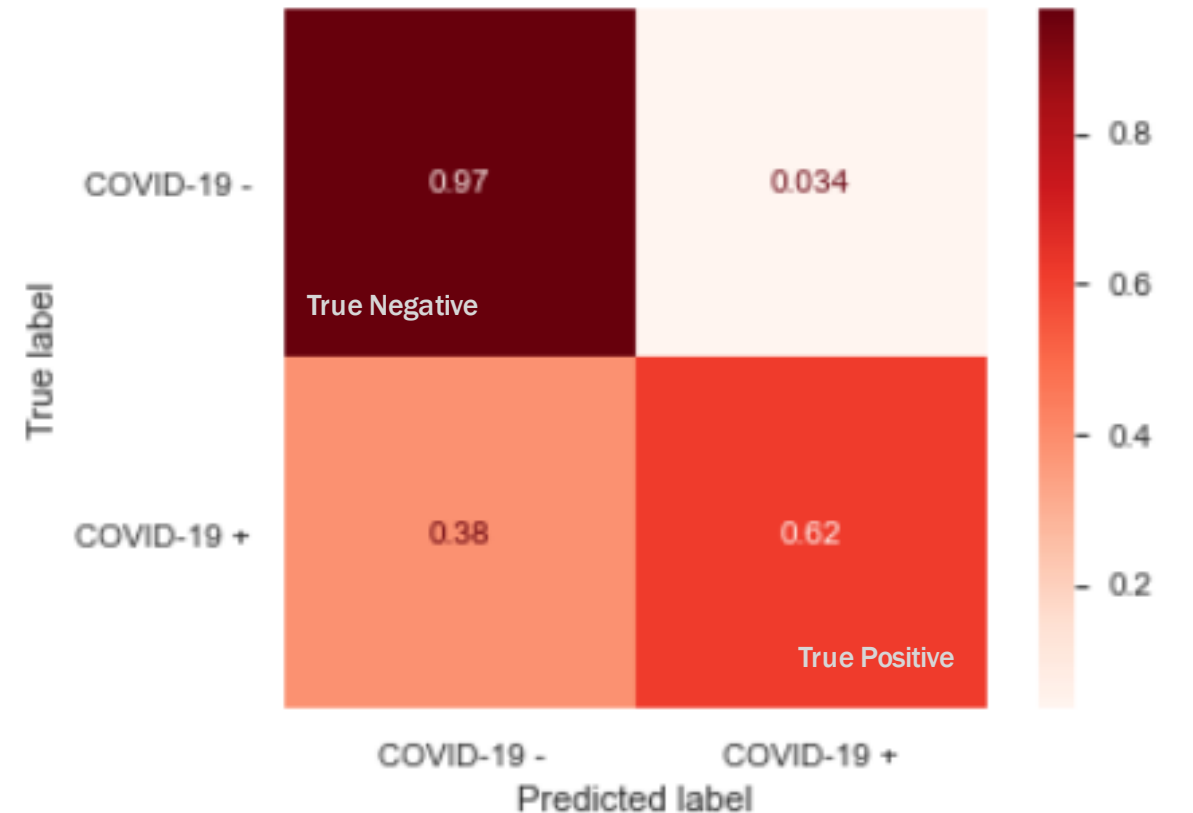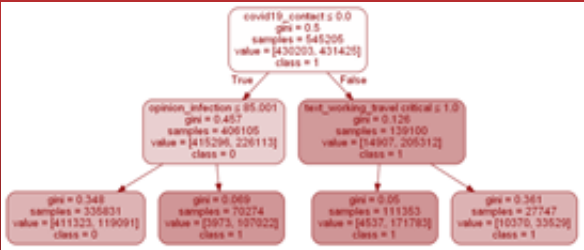
- **Belief** individual is infected
- Exhibiting COVID-19 **symptoms**
- **Contact** with others infected with COVID-19
- **Social distancing and hand washing** practiced as a household
- No answer when asked if respondent believed they could die from the infection
- **Social distancing** and **hand washing** practiced as an individual
- No answer when asked if respondent believed had the infection
- Male
- Female — **Sex**
- **BMI**
- Taking **prescription medications**

0.0   0.1   0.2   0.3   0.4   0.5

Random Forest Classifier Prediction Rates

	COVID-19 -	COVID-19 +
COVID-19 -	0.97 (True Negative)	0.034
COVID-19 +	0.38	0.62 (True Positive)

True label / Predicted label

Top Feature Correlations:

Positively Correlated

Individual believed they contracted COVID-19

Individual exhibited COVID-19 symptoms

Individual came in contact with another who was COVID-19+

Individual *did not respond* if they believed they had COVID-19

Individual *did not respond* if they believed they could die from COVID-19

Individual was male

Body Mass Index (kg/m** 2)

The individual was taking prescription medication

Negatively Correlated

Household practiced social distancing and hygiene

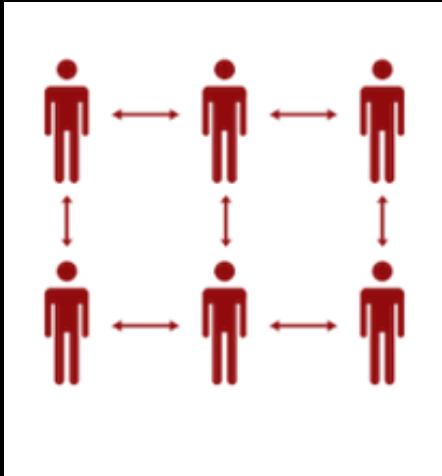Individual practiced social distancing and hygiene

Behavior Matters:

Individual was female

Recommendations:

- Practice social distancing and good hygiene.
  - Individually & Collectively

- Avoid contact with those known to be infected.
  - Especially true for males, and those with higher BMI

If taking prescription medications, discuss with physician possibility for increased risk for infection.

Future Work:

Prescription Medications:

Out of the 1200 unique medications:

- Identify what medications or combinations could be contributing factors

- Evaluate if increase in number of prescriptions are contributing factors

BMI

- In addition to this result, studies show that BMI is an important factor. This feature could be further examined to determine what a possible threshold might be.

Opinion Infection:

- Meaning/Description

- Data collection surrounding this no longer exist.

THANK YOU

STAY HEALTHY!

Appendix:

Receiver Operator Curve (ROC) Results for Random Forest Classification Model:

Blue dotted line represents the 50:50 chance in predicting if someone tests positive for COVID-19.

The red line represents the diagnostic ability or Area Under the Curve (AUC) of the binomial classifier to be 87.1% reliable or approximately 37% better than randomly guessing.

Receiver operating characteristic (ROC) Curve

AUC=0.871

True Positive Rate

False Positive Rate
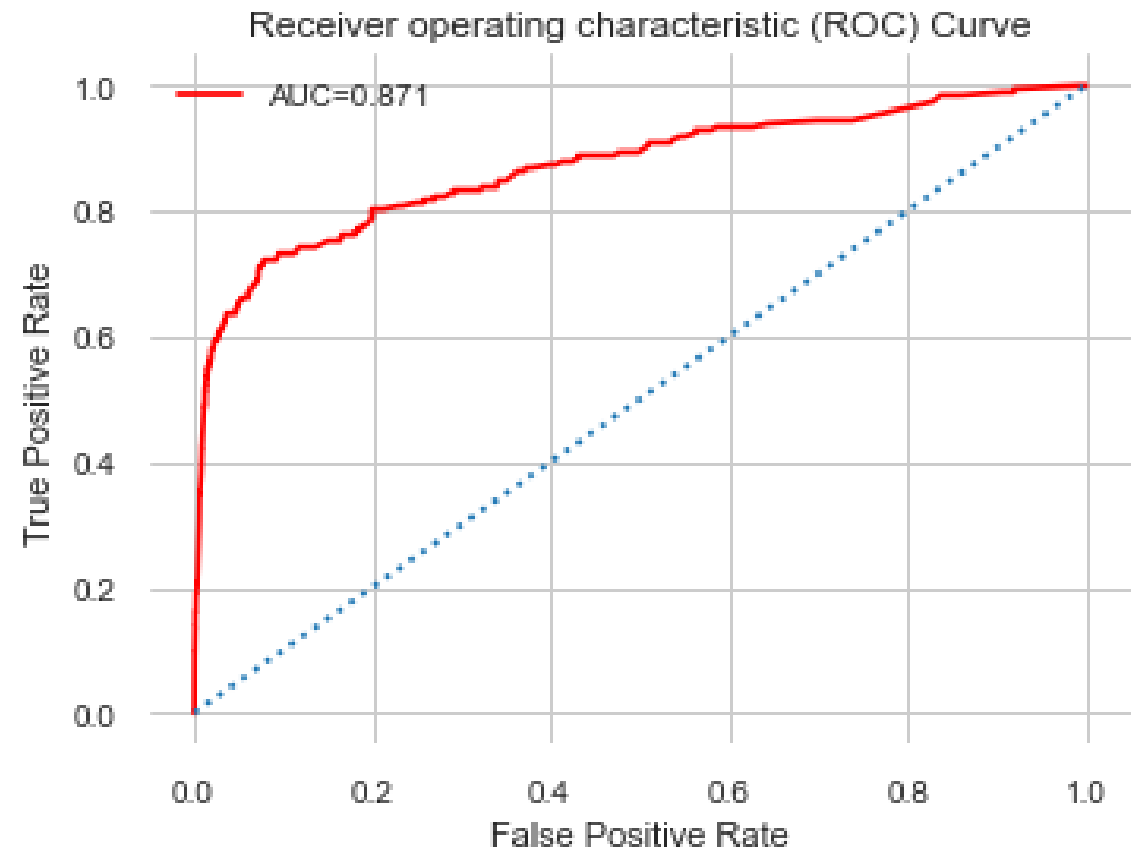
One sample tree from the random forest: (a closer look)