



Coffee Rust Prediction



Group 6: Rebecca Crust, Michael Garcia, Jessalene Ea

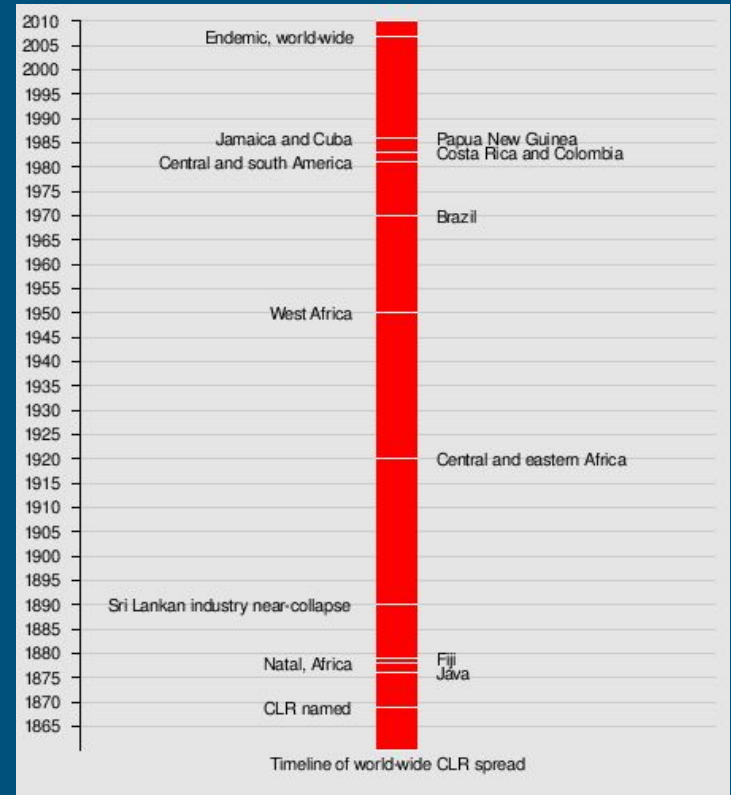


Introduction

- Problem:
 - Coffee Rust is a disease that has major economic impacts
 - Preventative measures are expensive, but necessary if environmental conditions are ideal for disease
- Our approach to a solution:
 - Create an AI that predicts severe outbreaks of Coffee rust based on weather conditions

Data Preparation

- Find locations and times of coffee rust epidemics
 - APSnet.org, various articles talking about epidemic
 - Once information about location and time is obtained, then use wunderground.com to collect data from that time period
 - Guatemala, Ecuador, Colombia, Nicaragua
 - Rainy season = conditions are ideal for coffee rust outbreak



Graph from wikipedia.org

Data Preparation

- Get data from that information
 - Data from weather underground website
 - <https://www.wunderground.com/history/airport/MGGT/2012/9/1/WeeklyHistory.html?&reqdb.zip=&reqdb.magic=&reqdb.wmo=>
 - Data straight from article (Guatemala below)
 - Required a lot of conversions in order for our data to be the same format

Year	Monthly rainfall accumulation (mm)												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	All
2012	12.6	20.9	24.5	89.8	236.9	227.2	167.8	325.4	199.5	221.4	24.1	18.9	1569
Climatology	16.0	15.4	23.1	54.6	160.6	294.5	240.9	244.4	296.6	206.3	61.8	26.4	1641
Bias	-3.4	5.6	1.4	35.2	76.3	-67.2	-73.0	81.0	-97.1	15.2	-37.7	-7.4	-71.3
%	79.0	136.3	105.9	164.4	147.5	77.2	69.7	133.1	67.3	107.4	39.0	71.8	95.7

Data Spreadsheet

Naive Bayes Round 1

- Discrete Naive Bayes Algorithm
 - Classes created were based in training data set
 - Classes the model had not seen before → 0.5 weight
 - Variables (classes) were continuous
- Results yielded about 68% accuracy
 - Unexpectedly high for our first trial
- Realized we needed to change our algorithm to accommodate the continuous classes

Naive Bayes Round 2

- Gaussian Naive Bayes Algorithm
 - Calculate mean and standard deviation of each class
 - Probabilities calculated by

$$p(x = v \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

- Testing was easier
- Predictions were the same as the discrete NB program

Changing the training/Testing data

- We compiled a second training/testing set using 80% training and 20% testing
- Wanted to see if small sample was an issue
- Results were worse
 - 61.9% accuracy

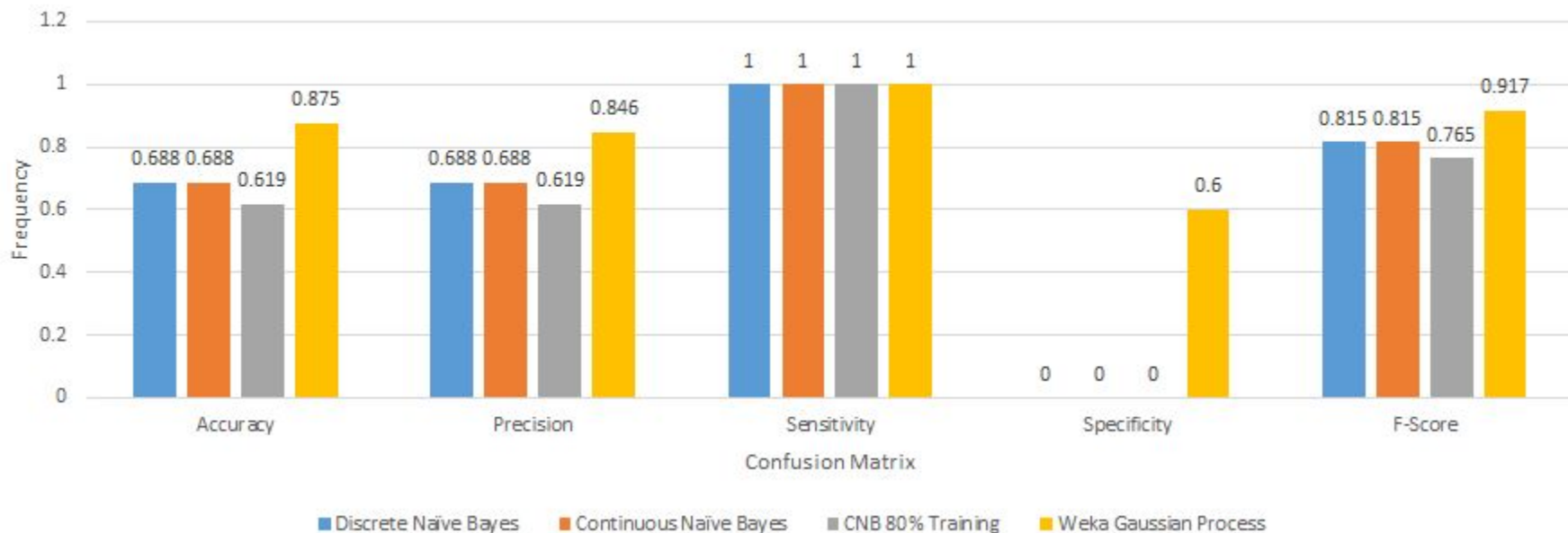
Demo

SVM

- Have yet to implement because of setbacks in our Naive Bayes program
- Ideal because it can handle multiple classes
- Possible Con:
 - Our small sample size might make the algorithm less accurate
- Likely Kernel: Linear
- Using open ML library scikit-Learn

Results

Compared Results



Conclusion and Discussion

- What went wrong?
 - Collecting our own data for training and testing programs.
 - Output models that aren't efficient
- What went right?
 - Our programs all run correctly
 - Model is correct more than half of the time
- What is not solved?
 - How to create a database for our programs to utilize
 - Looking into other models to improve results

Conclusion and Discussion

- Proposed timeline?
 - Expected to present in finals week, had to rush certain tasks
- What is important for our ML algorithm? → Needs more training data; a larger database

TIMELINE

- ★ WEEK OF OCTOBER 16, 2017
 - WHAT PROJECT TO WORK ON
 - WHICH SPECIFIC PATHOGEN TO WORK
 - WHICH VARIABLES TO USE
- ★ WEEK OF OCTOBER 23, 2017
 - GATHERING DATA
 - FINDING RESOURCES
 - PARTITION WORK FOR PROPOSAL
 - SETTING UP/SHARING SLIDES, EXCEL, GITHUB
 - INDIVIDUAL SLIDES AND INFORMATION
- ★ WEEK OF OCTOBER 30, 2017
 - REVIEW SLIDES TOGETHER
- ★ MEETING UP DATES
- ★ EARLY NOVEMBER: PUTTING DATA TOGETHER
 - EXCEL
 - GRAPH
 - MAKE MODEL
- ★ LATE NOVEMBER: PROGRAM WRITING
- ★ DECEMBER: ASSESSING ACCURACY, REVISING METHODS
- ★ DEAD/FINALS WEEK: FINALIZE PROJECT

References

www.wunderground.com

https://en.wikipedia.org/wiki/Hemileia_vastatrix

<http://www.apsnet.org/edcenter/intropp/lessons/fungi/Basidiomycetes/Pages/CoffeeRust.aspx>

<https://www.conservation.org/publications/Documents/Coffee-rust-crisis-in-Central-America.pdf>

[Guatemala 2012 Epidemic](#)

<http://scikit-learn.org/stable/modules/svm.html#multi-class-classification>

<http://observer.com/2015/09/central-americas-slow-recovery-from-coffee-ravaging-epidemic/>

Questions?



© ELYAN PARKER

