

# **/r/houseplants /r/gardening**

An exploration in  
classification modeling

Nolan Smurro

# Table of Contents

---

01.

Overview

02.

Data Collection

03.

Data Cleaning  
and Exploration

04.

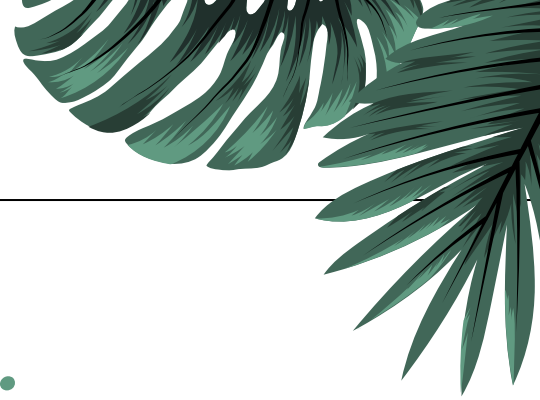
Model 1

05.

Model 2

06.

Model  
Comparison





01.

# Overview



# PlantMa

We empower houseplant hobbyists and gardening enthusiasts to keep their thumbs green!



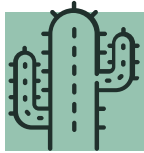
# Project Mission Statement

The aim of this project is to classify discussions and engagement on reddit to gain insights on PlantMa's customer base and their interests.



# Subreddit Classification

---



[/r/houseplants](#)

A space to discuss  
and admire plants in  
your house!



[/r/gardening](#)

A place for the best guides,  
pictures, and discussions of  
all things related to plants  
and their care.



# Data Collection

02.


---

# PRAW Collection



---

The PRAW (Python Reddit API Wrapper) Library was utilized to collect a total of 9,271 unique posts from /r/houseplants and /r/gardening utilizing the following filters:

- new
  - hot
  - rising
  - top (all, year, month, week)
- 





03.

# Data Cleaning and Exploration



# After Initial Count Vectorization...



**/r/houseplants  
dataframe size**

(4736, 33702)

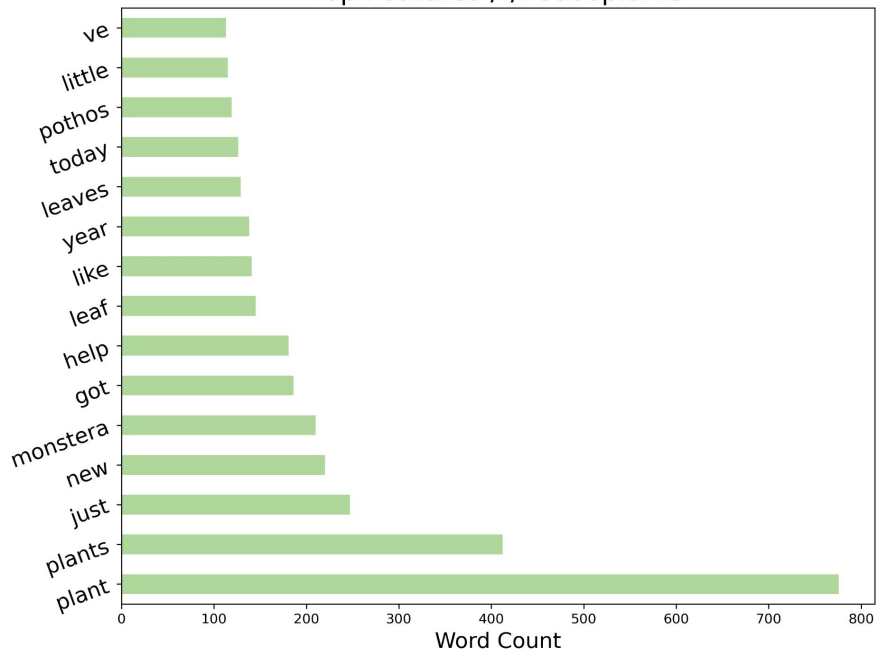
**/r/gardening  
dataframe size**

(4535, 33702)

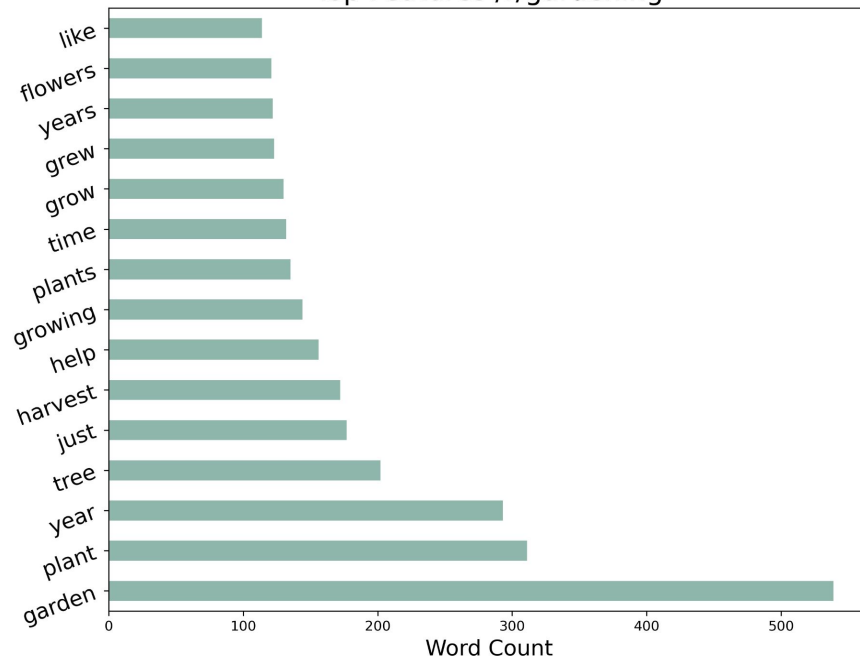


# Top Features

Top Features /r/houseplants



Top Features /r/gardening



The slide features a decorative border of green leaves in the corners. The top-right corner has a cluster of leaves, including a large monstera leaf. The bottom-right corner also has a cluster of leaves, including a large monstera leaf. The bottom-left corner has a large monstera leaf. The top-left corner has a large monstera leaf.


# Model 1

Count Vectorization and  
Logistic Regression

04.

---

# Best Parameters, Model 1

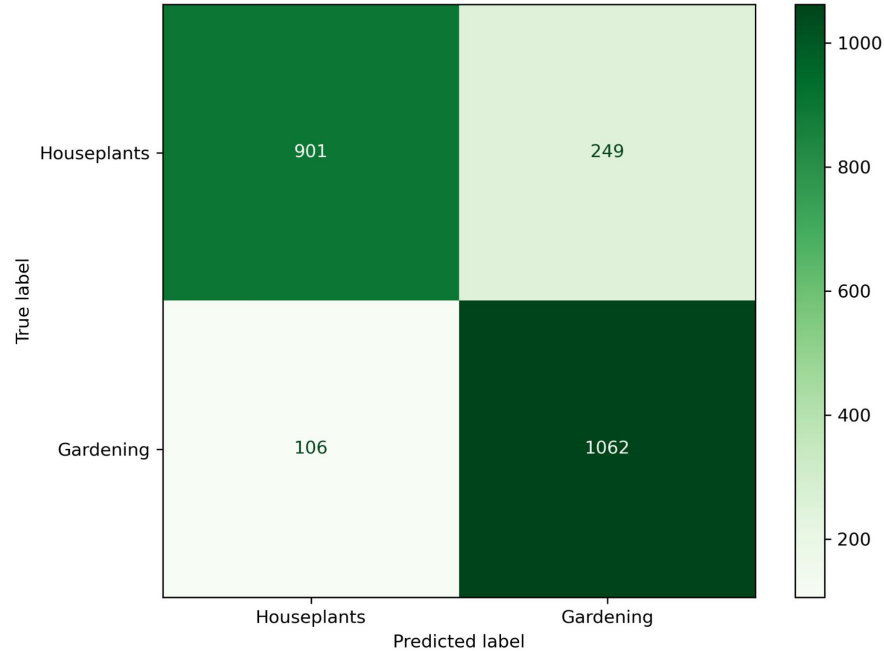
	cvec max df	cvec min df	cvec max features	cvec n gram range	cvec stop words
Model 1	0.9	1	None	(1,2)	English + Custom

Custom stop words: 've', 'help', 'little'

# Scores, Model 1

	Baseline	Benchmark Training	Benchmark Testing	Tuned Training	Tuned Testing
Model 1, cvec/ log_reg	.489	.937	.837	.977	.846

# Confusion Matrix, Model 1





# 05.

## Model 2

Count Vectorization/ Extra  
Trees Classifier



# Best Parameters, Model 2

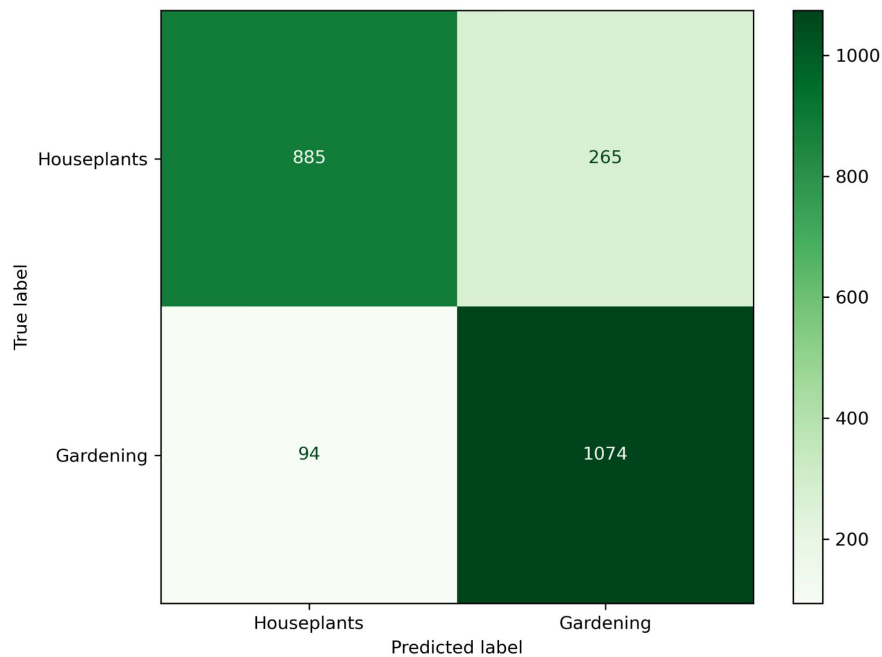


	cvec max df	cvec min df	cvec max features	cvec n gram range	cvec stop words	et max depth	et max features	et min samples leaf	et n estimators
Model 2	0.9	1	None	(1,2)	None	None	sqrt	1	500

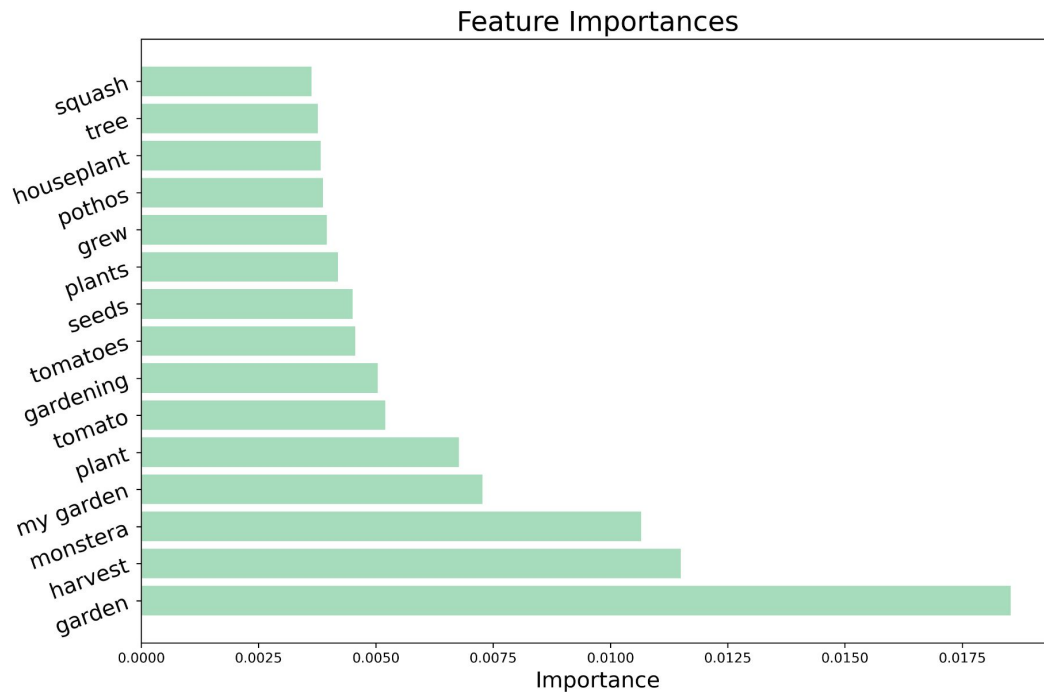
# Scores, Model 2

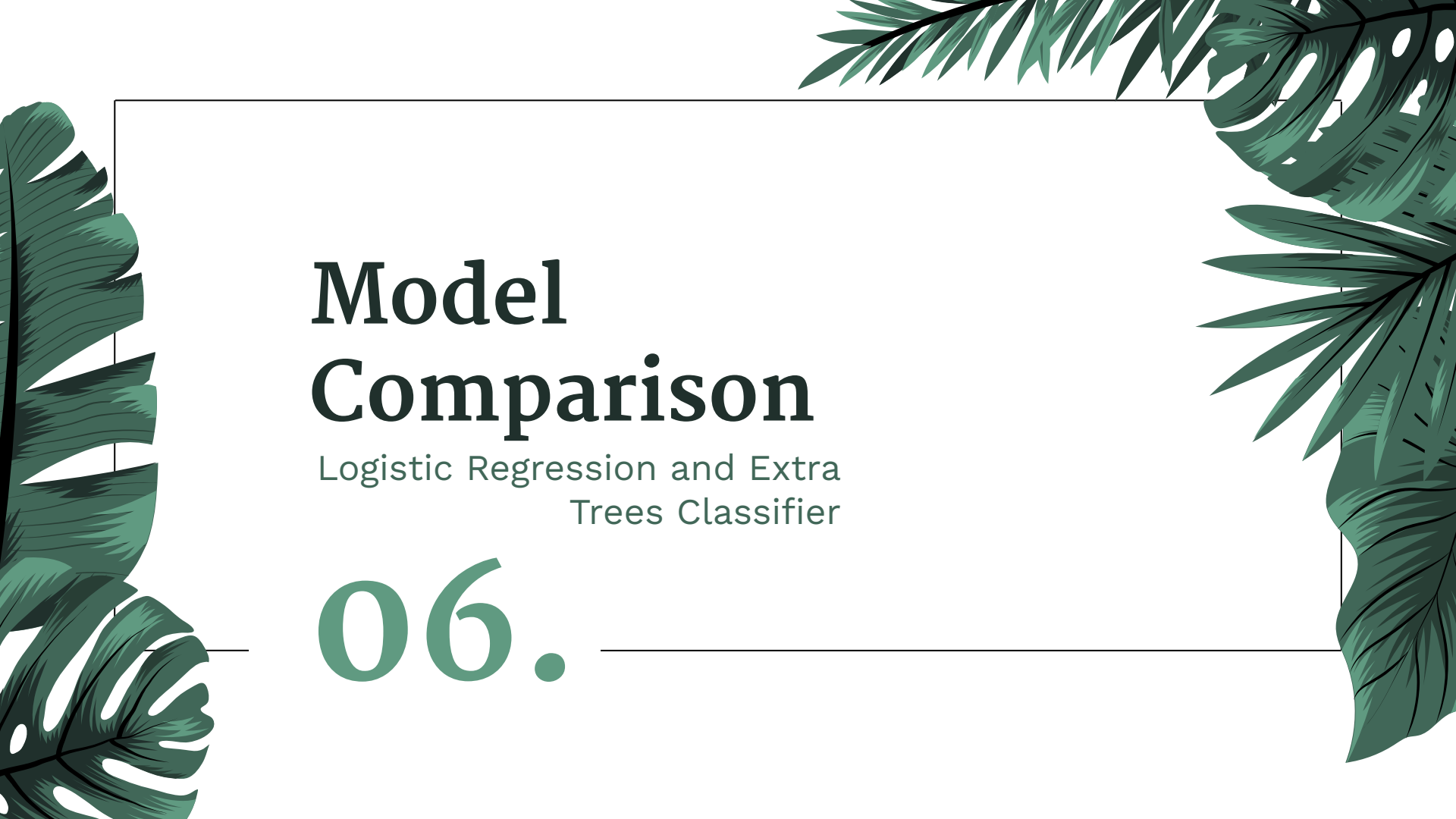
	Baseline	Benchmark Training	Benchmark Testing	Tuned Training	Tuned Testing
Model 2, cvec/ extra trees	.489	.995	.839	.995	.845

# Confusion Matrix, Model 2



# Feature Importances, Model 2



The slide features a white background with a thin black rectangular border. In the four corners, there are stylized green leaves. The top-left and bottom-right corners have large, detailed monstera leaves. The top-right and bottom-left corners have smaller, more delicate palm-like leaves. The main text is centered within the white area.

# Model Comparison

Logistic Regression and Extra  
Trees Classifier


06.

---

# Scores, Comparison



	Baseline	Benchmark Training	Benchmark Testing	Tuned Training	Tuned Testing
Model 1, cvec/ log_reg	.489	.937	.837	.977	.846
Model 2, cvec/ extra trees	.489	.995	.839	.995	.845





## /r/houseplants Misclassified

“How it started Vs how it’s growing”

“She’s doing a thing”

“Am I doing it right?”

“Why I’m single”

“How to manage this big boi.”

“Dogs and UPS delivery don’t mix 🙄”

## /r/gardening Misclassified

“ain’t stupid if it works”

“Today on ‘Reddit made me buy it”

“so proud of my babies”

“Poisonous?”

“Forgive me pruning god, for I have sinned”

“My mom got furious after discovering her cat's new warm sleeping spot”

# Thanks

Do you have any questions?

**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

