

Lichen Classifier

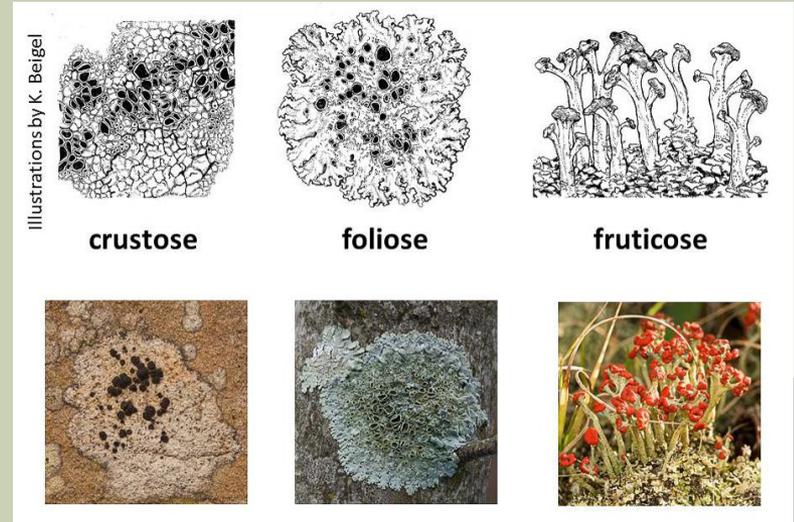
Elise Hachfeld



Overview

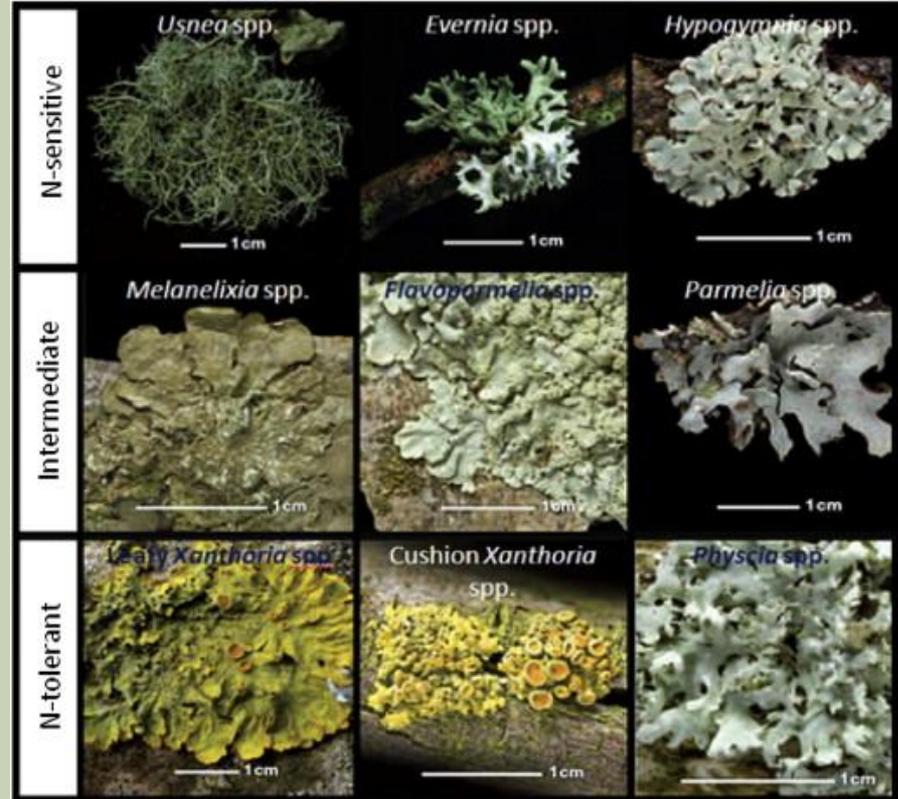
What is lichen?

- Lichens are symbiotic organisms composed of fungi and algae.
- Lichens usually grow on wood or rock and often live near moss.
- Some lichen live in the most difficult conditions on earth, like deserts and the arctic tundra.
- They produce different chemicals which produce their coloration and can even wear rocks (crustose lichen)



Purpose

- Lichens convert carbon dioxide into oxygen through photosynthesis like plants.
- They are bioindicators because some species are very sensitive to air pollution, especially with nitrogen and sulfur dioxide.
- Fruticose and foliose lichen appear more in areas with clean air.
- Monitoring of lichen species provides a cost effective way to estimate pollution levels.

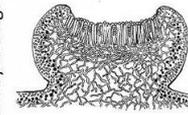


Classification Pitfalls

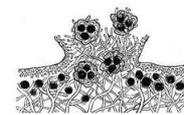
- Some species are notoriously difficult to identify.
- Identification is often based on lobe shape, branching, and the presence of reproductive structures called soredia, isidia, and apothecia.
- Color can be a useful indicator in some species, but rain/moisture causes a color change in most lichen.
- Several species of lichen often coexist, and lichen often grows next to moss.
- iNaturalist images have varying quality.



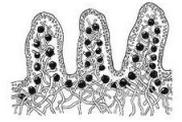
Illustrations by K. Beigel



apothecium



soredia



isidia



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Data Collection:

- Since lichens do not form a monophyletic group, observations were filtered on the genus level using Wikipedia's list of [lichen genera](#).
- Data was scraped from the citizen science website [iNaturalist.com](#).

Filters:

- Research grade ($\frac{2}{3}$ consensus) on species level ID
- Most identifiers agreed
- Open geoprivacy (location not obscured)
- Located in the United States

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

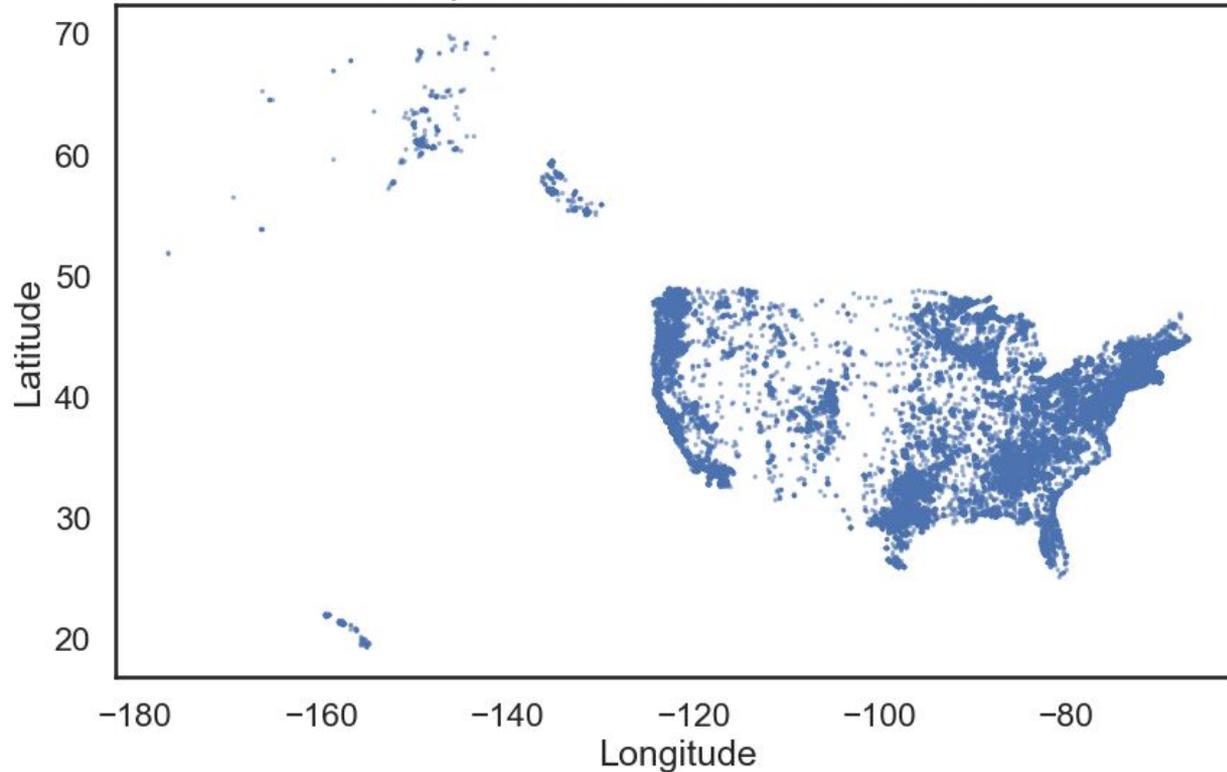
Model training

fitted models

Model evaluation

model insights

Geospatial Distribution of Observations



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

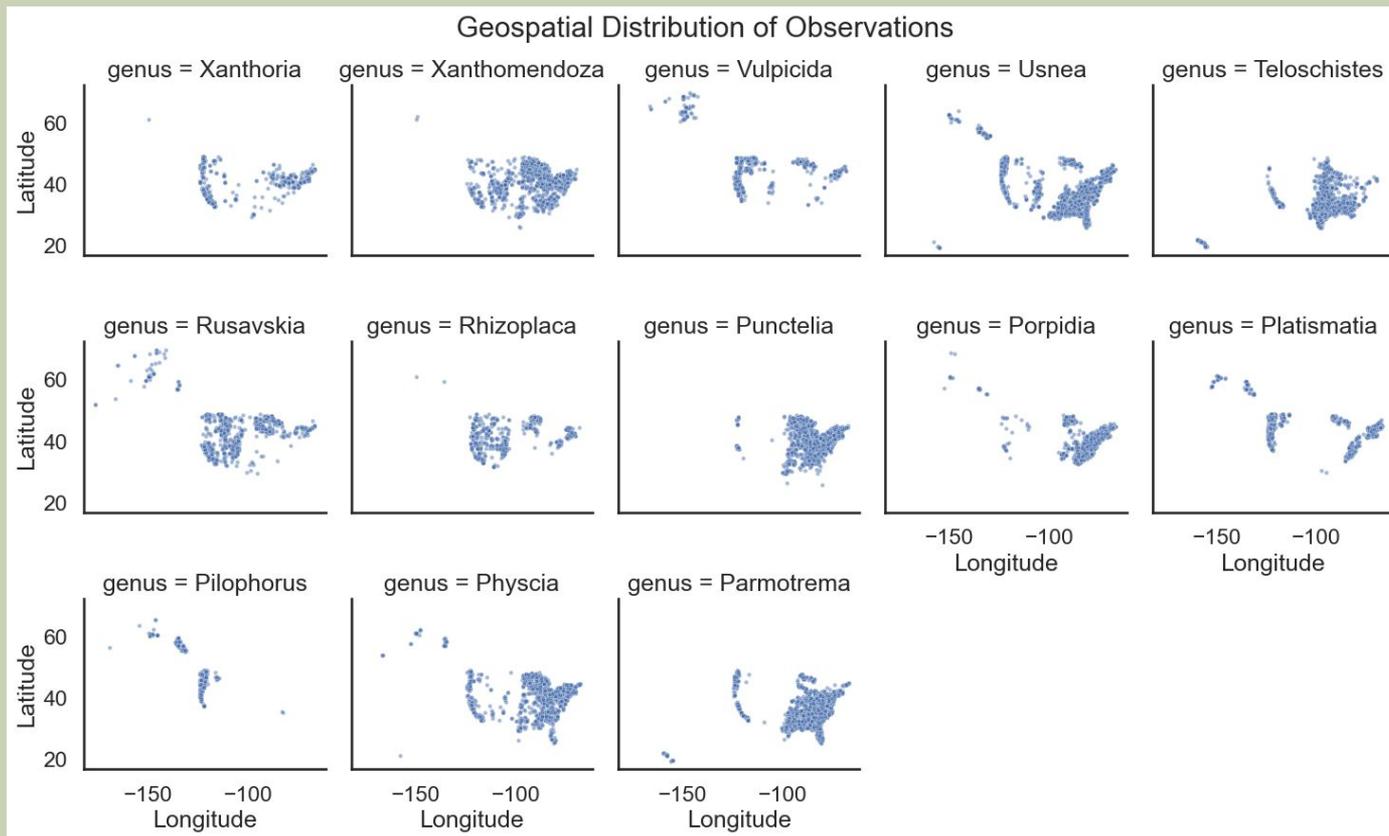
selected algorithms

Model training

fitted models

Model evaluation

model insights



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

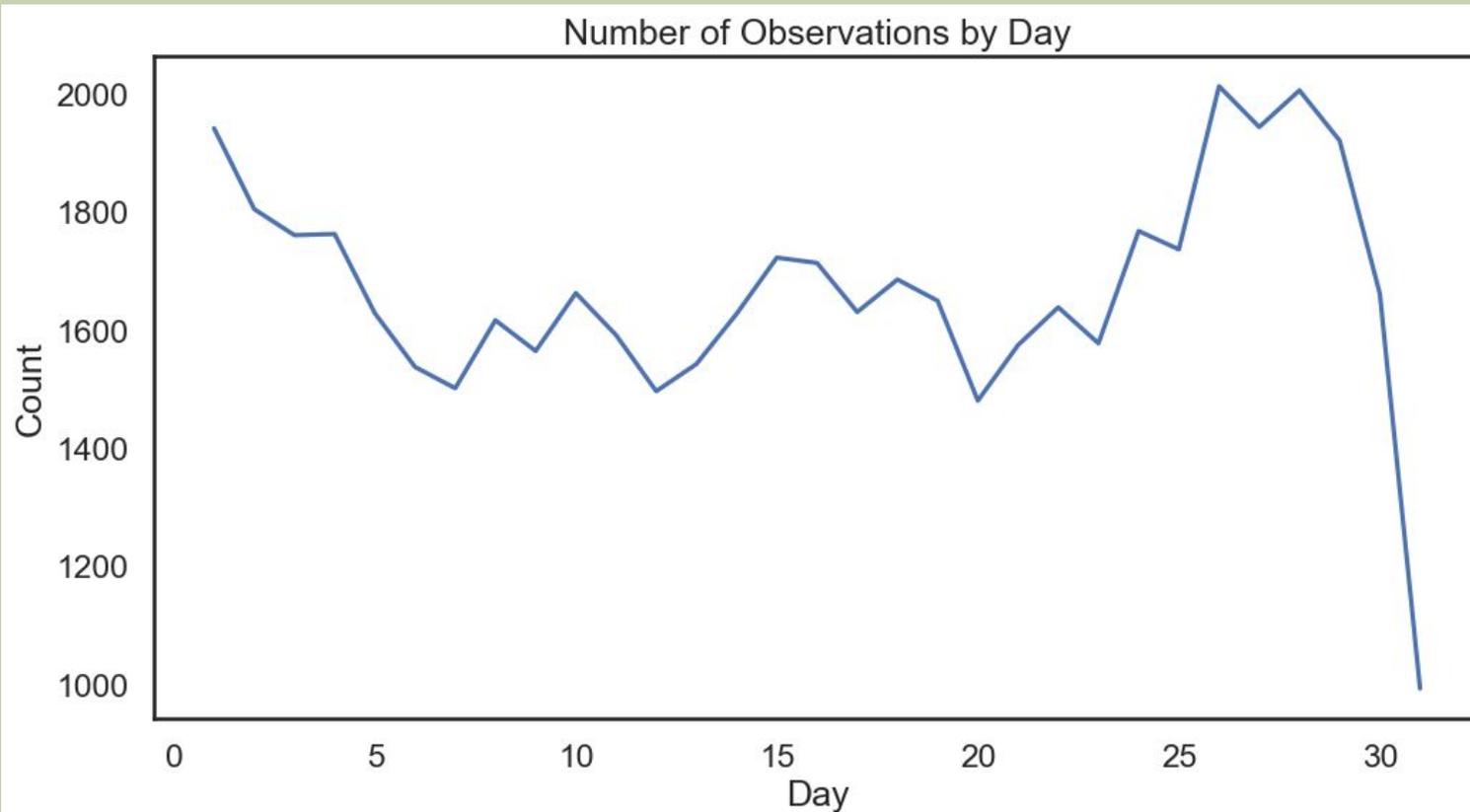
selected algorithms

Model training

fitted models

Model evaluation

model insights



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

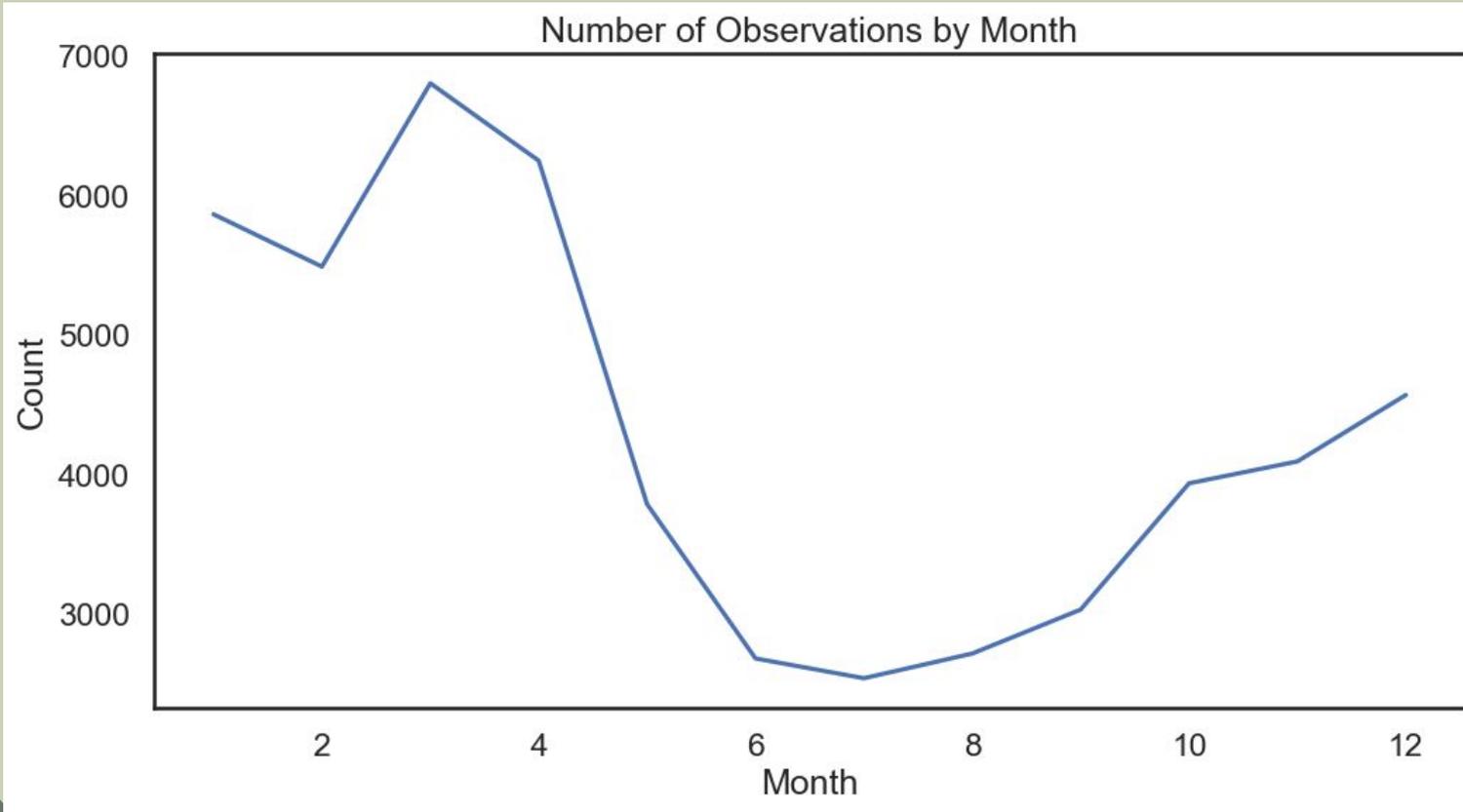
selected algorithms

Model training

fitted models

Model evaluation

model insights



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights



Typical Shield Lichens
(Subfamily Parmelioideae)

👍 2 🗨️ 1

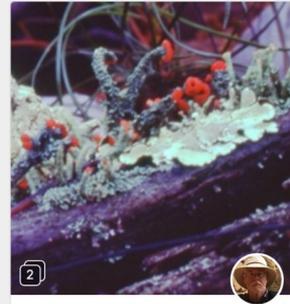
🕒 Jul '77



Common Greenshield Lichen
(*Flavoparmelia caperata*)

🏆 Research Grade 👍 2

🕒 Aug '78



Lipstick Powderhorn
(*Cladonia macilenta*)

👍 1

🕒 Aug '81



Lipstick Powderhorn
(*Cladonia macilenta*)

👍 1

🕒 Sep '81



Flaking Tarpaper Lichen
(*Collema flaccidum*)

👍 2 🗨️ 3 ⭐ 2

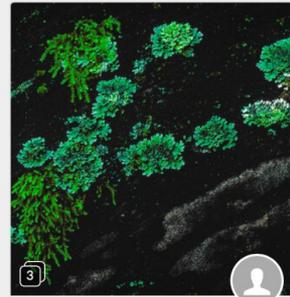
🕒 Oct '96



Genus Xanthomendoza

👍 3 🗨️ 1

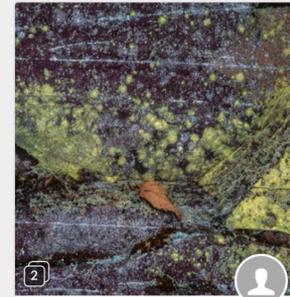
🕒 Mar '98



Frosted Lichens
(Genus *Physconia*)

👍 1

🕒 Sep '98



Genus Psilolechia

👍 2 🗨️ 1

🕒 Jan '99

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

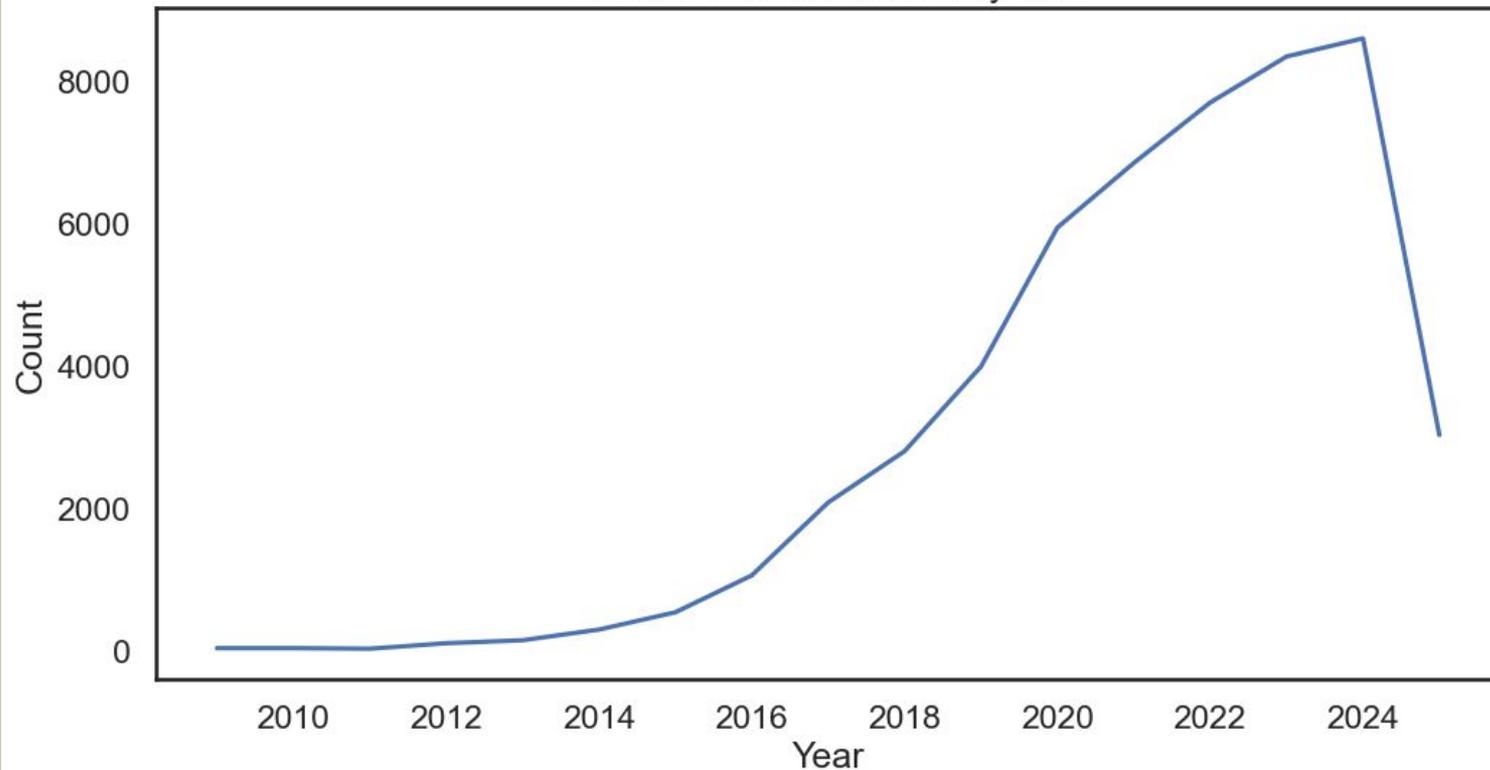
Model training

fitted models

Model evaluation

model insights

Number of Observations by Year



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

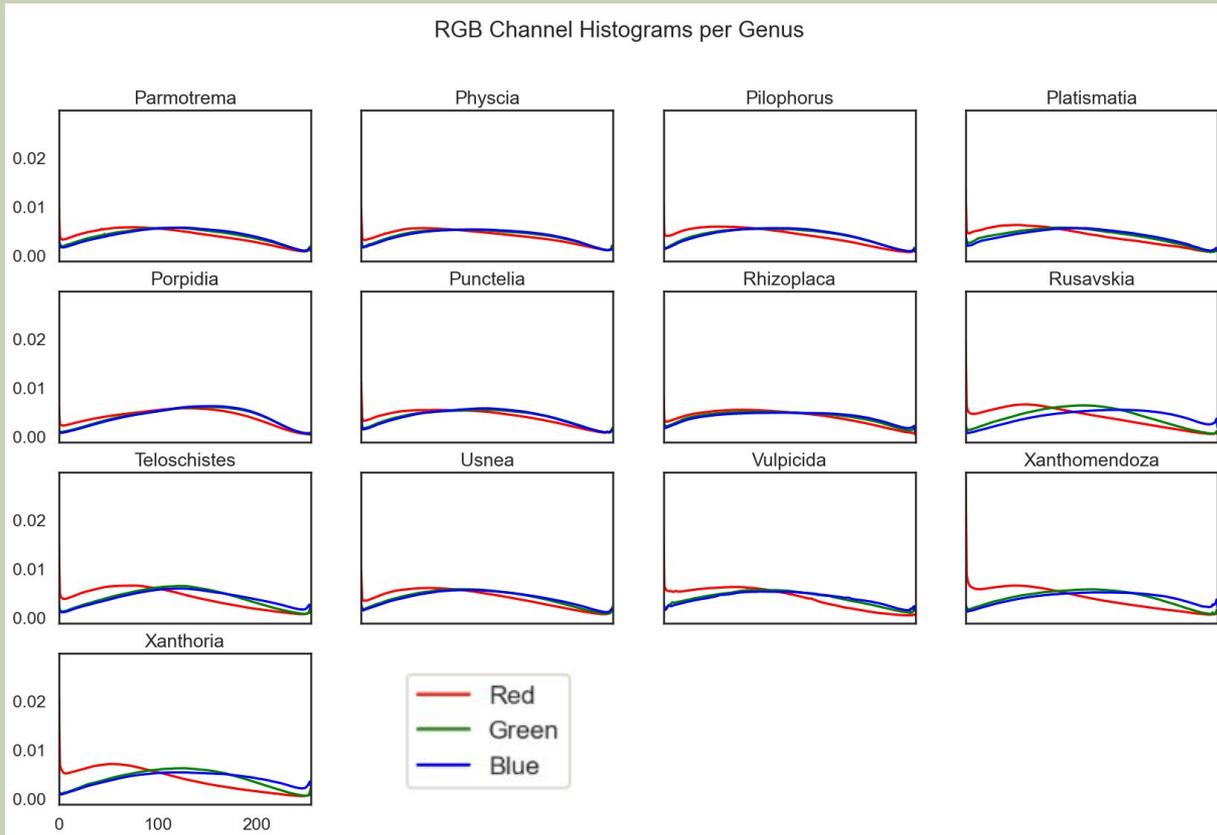
selected algorithms

Model training

fitted models

Model evaluation

model insights



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

- Removed duplicate rows based on UUID.
- Created genus column from scientific name.
- Converted dates and times to pandas datetime types.
- Scraped high resolution images from iNaturalist.
- Removed rows from dataframe for the images that failed to download.
- Split images into training (70%), validation (15%) and testing (15%) sets.
 - Stratified by species
 - Dropped species folder if it had 3 or less observations

Cleaning



Teloschistes

Xanthomendoza



Physica



Expectation





Reality

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

All subsets:

- Keras preprocessing function for base model
- Resize (256x256)
- Random crop (224 x 224)

Only training set:

- Random flip (horizontal)
- Random rotation (0.2)
 - Factor = 0.2 [-72, 72]
- Random contrast (0.2)
 - Range: [0.8, 1.2]
- Random translation (0.2, 0.2)
 - Range: [-0.2, 0.2], [-0.2, 0.2]

Note: I did NOT grayscale the images.

Augmentation

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Transfer Learning

- **Base model:** imagenet weights

Custom layers:

- **Dense:** 1024 neurons, SELU activation, lecun_normal
- **Batch Normalization**
- **Dropout :** 30%
- **Dense** 512 neurons, SELU activation, lecun_normal
- **Batch Normalization**
- **Dropout :** 20%
- **Dense:** 128 neurons, SELU activation, lecun_normal
- **Batch Normalization**
- **Dropout :** 10%
- **Output:** 13 neurons, softmax activation

Base Models Tested:

ResNet

- ResNet50
- ResNet50V2
- ResNet101
- ResNet152

EfficientNet

- EfficientNetB0
- EfficientNetV2B0

Architecture

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 224, 224, 3)	0
preprocess (Lambda)	(None, 224, 224, 3)	0
resize (Resizing)	(None, 256, 256, 3)	0
center_crop (CenterCrop)	(None, 224, 224, 3)	0
random_flip (RandomFlip)	(None, 224, 224, 3)	0
random_rotation (RandomRotation)	(None, 224, 224, 3)	0
random_contrast (RandomContrast)	(None, 224, 224, 3)	0
random_translation (RandomTranslation)	(None, 224, 224, 3)	0
base_model (Functional)	(None, 7, 7, 2048)	23,587,712
global_max_pool (GlobalMaxPooling2D)	(None, 2048)	0
dense_1024 (Dense)	(None, 1024)	2,098,176
batch_norm_1 (BatchNormalization)	(None, 1024)	4,096
dropout_1 (Dropout)	(None, 1024)	0
dense_512 (Dense)	(None, 512)	524,800
batch_norm_2 (BatchNormalization)	(None, 512)	2,048
dropout_2 (Dropout)	(None, 512)	0
dense_128 (Dense)	(None, 128)	65,664
batch_norm_3 (BatchNormalization)	(None, 128)	512
dropout_3 (Dropout)	(None, 128)	0
output_layer (Dense)	(None, 13)	1,677

Total Parameters: 26,284,685

Trainable Parameters: 2,693,645

Base Model: ResNet50

Architecture

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Batch Size

- 32 samples before updating model

Optimizer

- Adam
- Tried SGD, tended to overfit

Learning Rate Scheduling

- Cosine decay with restarts
- 10,000 decay steps
- Converged better than exponential decay

Initial Learning Rates

- Coarse: 0.00001 (1e-5)
- Fine: 0.000005 (5e-6)

Label Smoothing

- Set to 0.05
- Reduced overfitting

Parameters

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Stage 1

- Trained the model with the base model frozen
- Only the weights of the custom layers were updated
- Avoids destroying the imagenet weights

Stage 2

- Trained the model with a lower learning rate
- Unfroze the last **50** layers of the base model
- Fine tunes the model by making the weights trainable

Fine Tuning

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

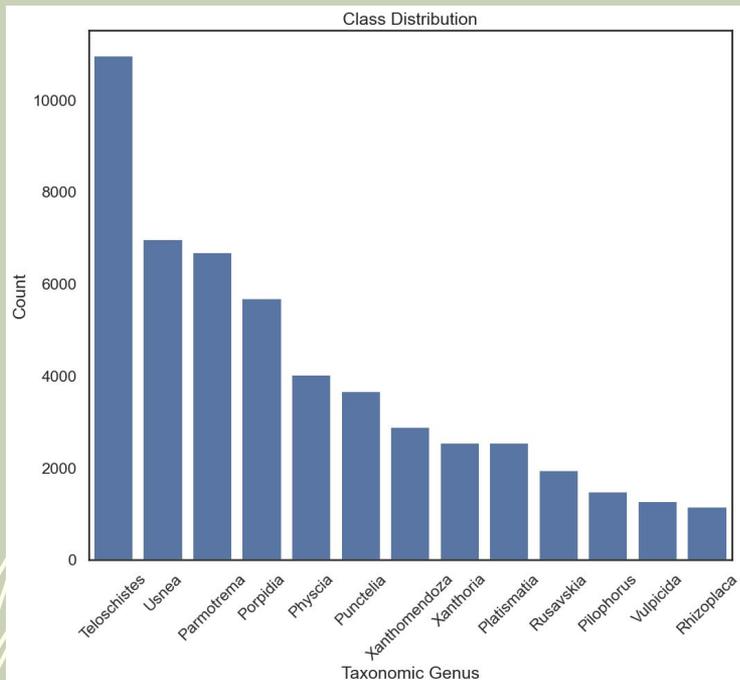
selected algorithms

Model training

fitted models

Model evaluation

model insights



Porpidia	14	48	10	11	93	54	12	21	26
Punctelia	31	63	28	29	191	113	18	19	54
Rhizoplaca	8	19	10	18	58	30	4	11	12
Rusavskia	16	38	10	19	82	75	11	11	28
Teloschistes	80	216	59	106	526	320	70	85	181
Usnea	53	149	37	67	312	236	38	53	93
Vulpicida	9	25	7	16	61	35	7	11	18
Xanthomendoza	27	47	14	27	132	89	16	25	51
Xanthoria	22	63	11	17	105	80	15	29	38

Class Imbalance

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Early Stopping + Checkpoint:

- Ended training after 10 epochs of no improvement in validation loss.
- Saved model with the best weights for both coarse and fine training.

Evaluated model performance with visual inspection of confusion matrix and classification report.

Metrics

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

	precision	recall	f1-score	support
Vulpicida	0.7502439024390240	0.865990990990991	0.8039728175640360	888.0
Pilophorus	0.6320400500625780	0.8501683501683500	0.7250538406317300	594.0
Parmotrema	0.9	0.8959276018099550	0.8979591836734690	221.0
Punctelia	0.8789237668161440	0.519893899204244	0.6533333333333330	377.0
Rhizoplaca	0.9692307692307690	0.9658018867924530	0.96751329001772	848.0
Xanthomendoza	0.781431334622824	0.73992673992674	0.7601128880526810	546.0
Porpidia	0.8044692737430170	0.8470588235294120	0.8252148997134670	170.0
Xanthoria	0.7658227848101270	0.8344827586206900	0.7986798679867990	290.0
Platismatia	0.9763888888888890	0.8557516737674990	0.91209886052546220	1643.0
Teloschistes	0.8664323374340950	0.9499036608863200	0.90625	1038.0
Physcia	0.73	0.7724867724867720	0.7506426735218510	189.0
Rusavskia	0.773109243697479	0.42990654205607500	0.5525525525525530	428.0
Usnea	0.6207627118644070	0.7710526315789470	0.687793427230047	380.0
accuracy	0.8265895953757230	0.8265895953757230	0.8265895953757230	0.8265895953757230
macro avg	0.8037580818161040	0.7921809486014190	0.7877828753486390	7612.0
weighted avg	0.8391532472211960	0.8265895953757230	0.8244525827165990	7612.0

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

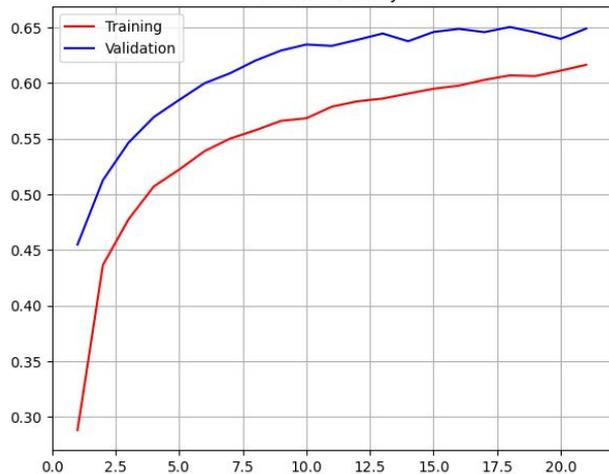
Model training

fitted models

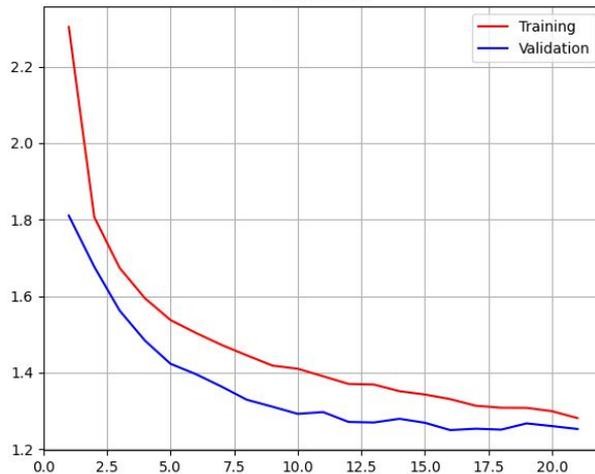
Model evaluation

model insights

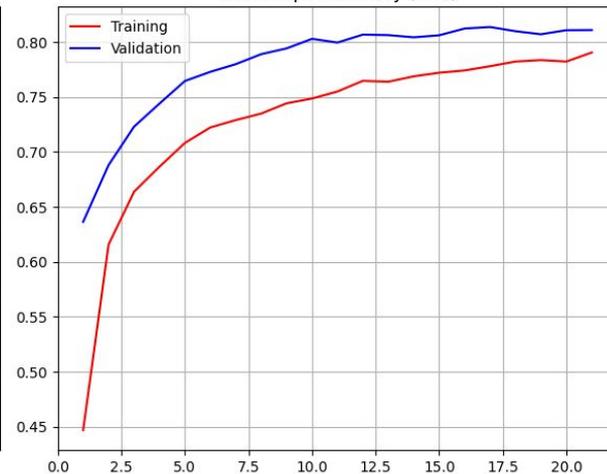
Model Accuracy



Model Loss



Model Top-k accuracy (k=2)



Exploring Data

data insights

Data cleaning

cleaned data

Model selection

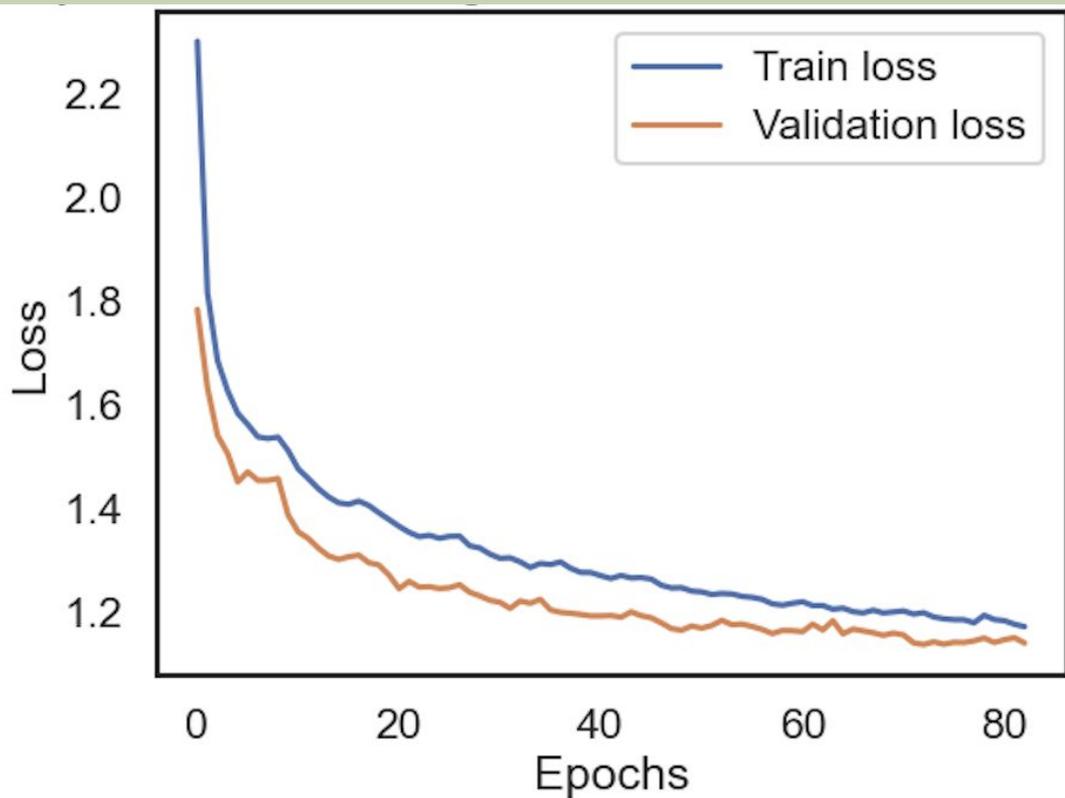
selected algorithms

Model training

fitted models

Model evaluation

model insights



Coarse

- Accuracy: 68.48%
- Top-k: 84.3%

Fine

- Accuracy: 82.66%
- Top-k: 93.52%

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Coarse

True: Vulpicida

1. Vulpicida (93.07%)
2. Teloschistes (2.82%)
3. Xanthomendoza (1.44%)

True: Rhizoplaca

1. Rhizoplaca (95.38%)
2. Porpidia (1.78%)
3. Pilophorus (1.47%)

True: Usnea

1. Usnea (71.79%)
2. Xanthoria (18.94%)
3. Physcia (2.59%)

True: Xanthomendoza

1. Xanthomendoza (43.60%)
2. Vulpicida (24.19%)
3. Pilophorus (17.49%)

True: Platismatia

1. Platismatia (99.66%)
2. Usnea (0.13%)
3. Rusavskia (0.11%)



Fine

True: Vulpicida

1. Vulpicida (96.20%)
2. Parmotrema (0.90%)
3. Rhizoplaca (0.61%)

True: Rhizoplaca

1. Rhizoplaca (93.72%)
2. Vulpicida (1.12%)
3. Physcia (1.09%)

True: Usnea

1. Xanthoria (44.40%)
2. Usnea (39.47%)
3. Rusavskia (8.64%)

True: Xanthomendoza

1. Xanthomendoza (69.20%)
2. Vulpicida (24.76%)
3. Pilophorus (3.77%)

True: Platismatia

1. Platismatia (99.58%)
2. Porpidia (0.25%)
3. Punctelia (0.05%)

Exploring Data

data insights

Data cleaning

cleaned data

Model selection

selected algorithms

Model training

fitted models

Model evaluation

model insights

Potential Improvements:

- Further tune hyperparameters using grid search.
- Add additional genus categories, especially those important for air quality.
- Use this model as a base model to train species identification.
- Create an ensemble model with location data to further refine predictions.

Next Steps

Thanks!



Do you have any questions?
youremail@freepik.com +91 620 421 838
yourcompany.com

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

Please keep this slide for attribution