From Zero to Quarter Bil

How I bootstrapped $250 MM of customer value using Cloud AutoML.

Code Examples: https://goo.gl/iALoDJ

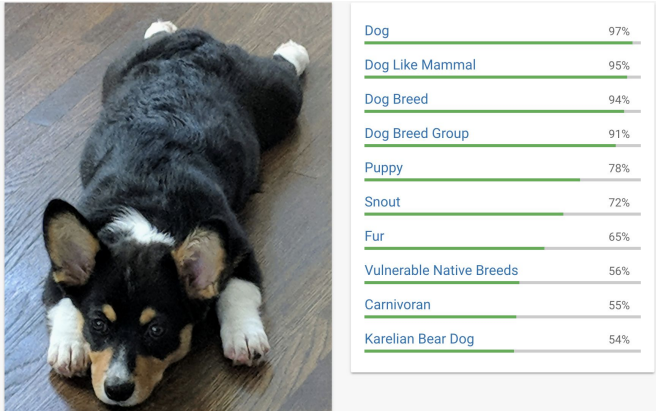
Leigh Johnson
Founder, Grant
@grepLeigh

What is machine learning?

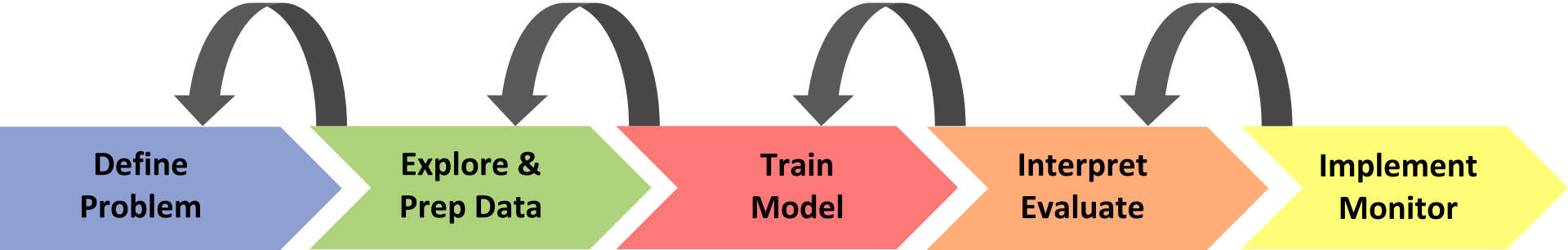
Automated pattern recognition

What is Automated Machine Learning?

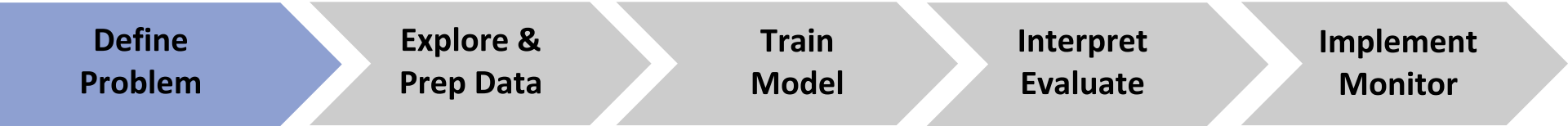
Pre-trained Model APIs	Automated ML	Custom Models
Vision, Speech, NLP, Translation	Cloud AutoML	ML Engine



ML Project Lifecycle

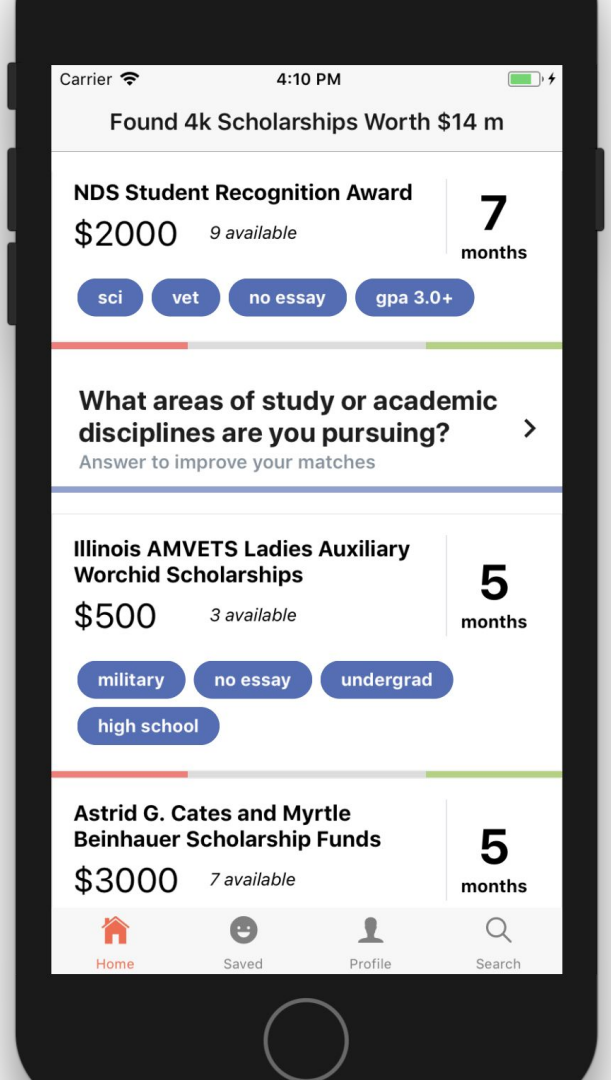


Define Business Problem



$2.3B scholarships went unclaimed in 2017

swim



Annotate scholarships to improve search and recommendations

Financial need
School
Transcript
Disability
Heritage
Military Service
Interest/Hobby

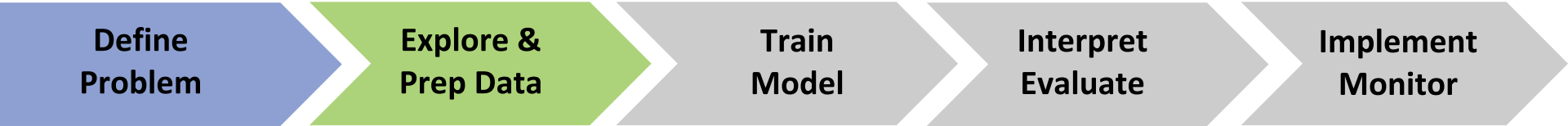
GPA
No Essay
Area of Study
Grade level
Region
Interview
Athletic ability

Amelia Earhart Fellowships are awarded annually to women pursuing graduate Ph.D/doctoral degrees in aerospace-related sciences or aerospace-related engineering ...

Female

Aviation

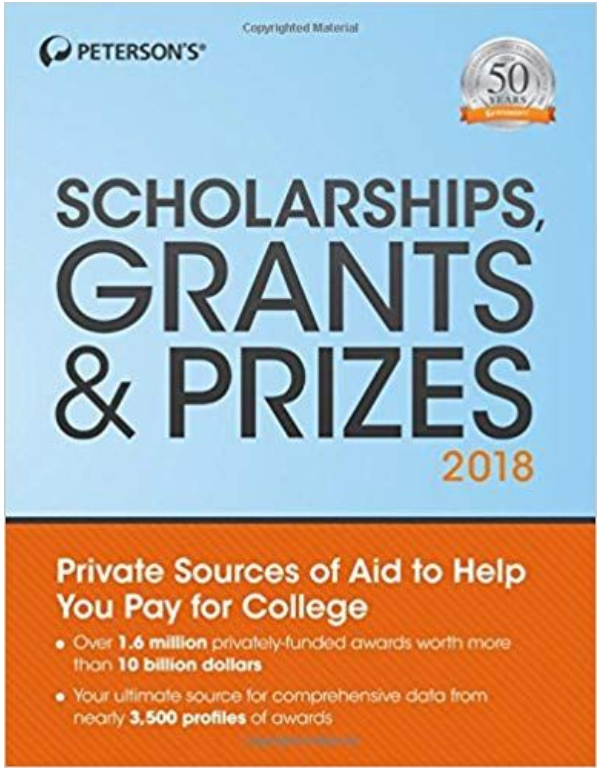
Explore & Prep Data



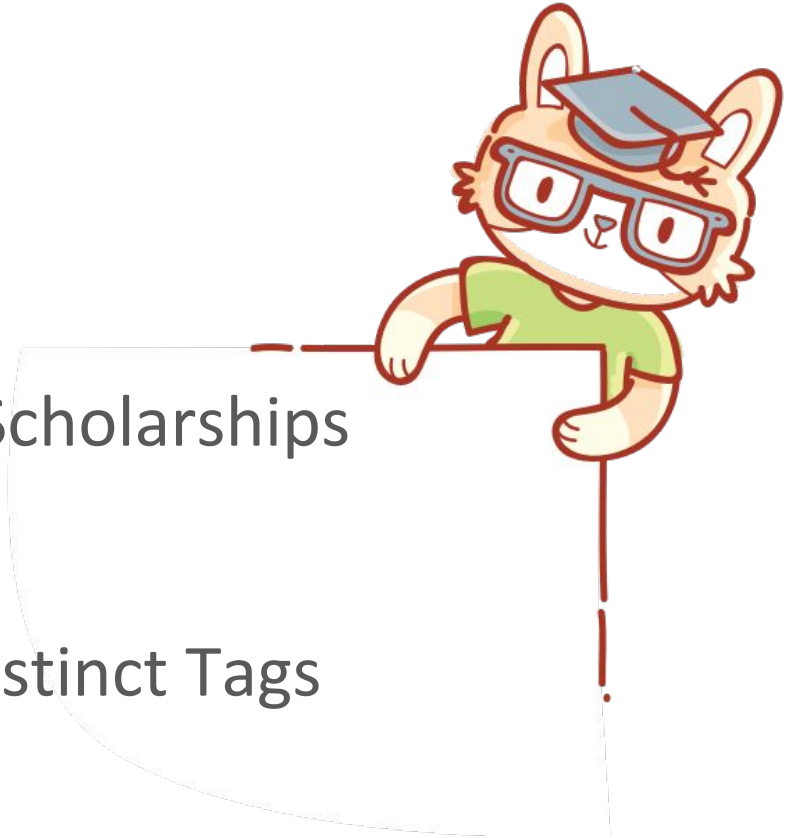
Looking for a dataset?

Try toolbox.google.com/datasetsearch

Acquire Labeled Data



Preprocess Data

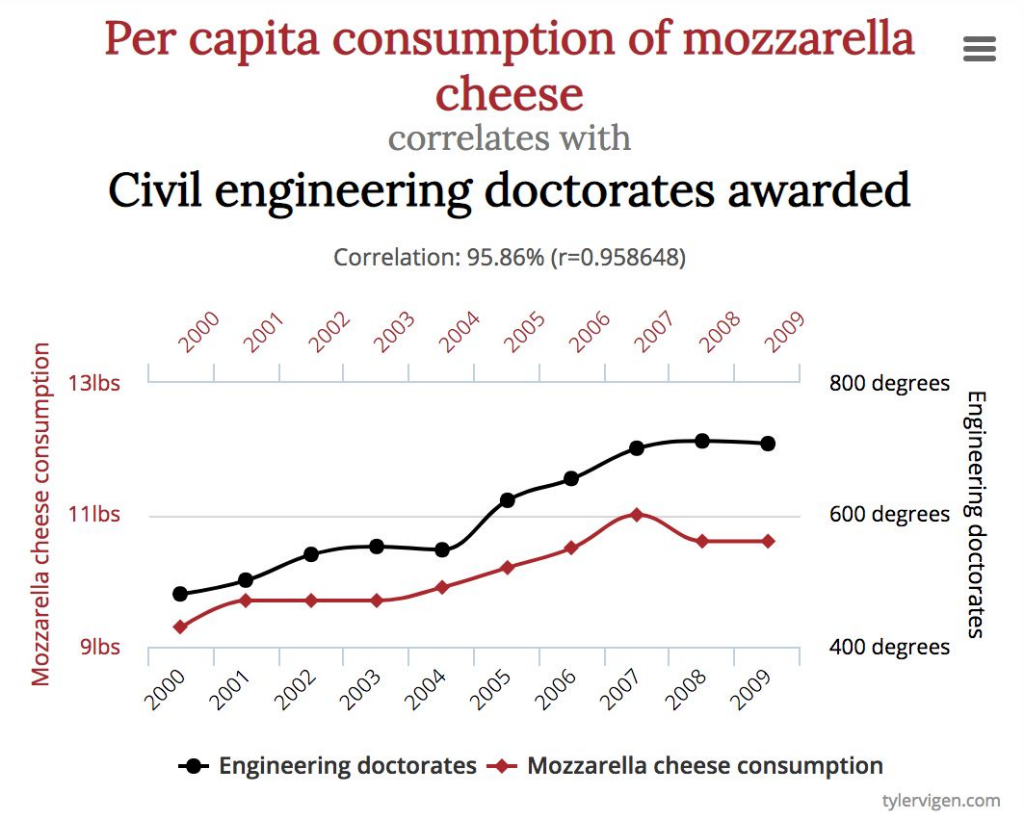


4,267 Scholarships

343 distinct Tags

Code samples:
https://goo.gl/iALoDJ#Training-Dataset-Exploration

Explore Data, Subdivide Problems, Reduce Collinearity



Instead of one generic model...

GPA, essay required, area of study, grade level, region, interview, financial need, school type, transcript, disability, heritage, state residence, athletic ability, military service

Train multiple, specific models

GPA, academic level, school type

Area of study

Disability

Heritage

State residence / use

Subdividing Another ML problem

Generic Model Predicts:

(Limited to 1,000 labels)

White
Wedding Dress
Cocktail Dress

Image Credit: modcloth.com

Color Model:
Ivory

Article type Model:
Dress

Neckline Model:
V-Neck

Length Model:
Sleeveless
Midi

Material Model:
Satin

Train Model

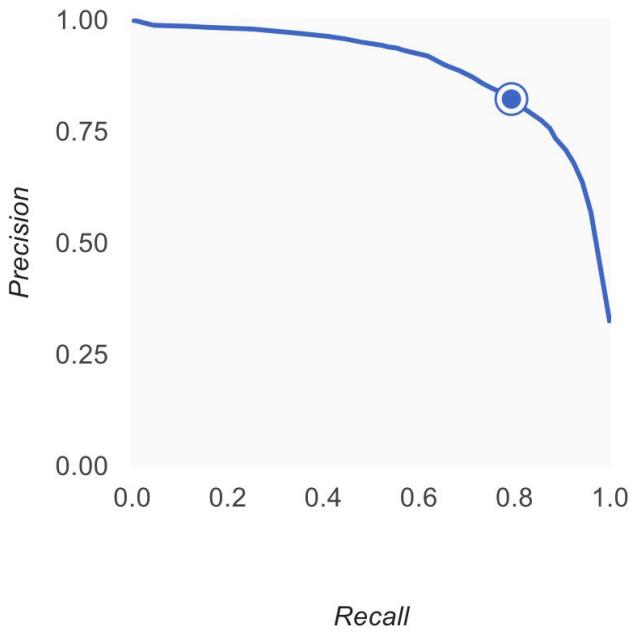
Demo

Evaluate Model

What is a Confusion Matrix?

		Actual Class	
		Dog	Non Dog
Predicted Class	Dog	True Positive	False Positive
	Non Dog	False Negative	True Negative

Beyond Accuracy: **Precision** and **Recall**

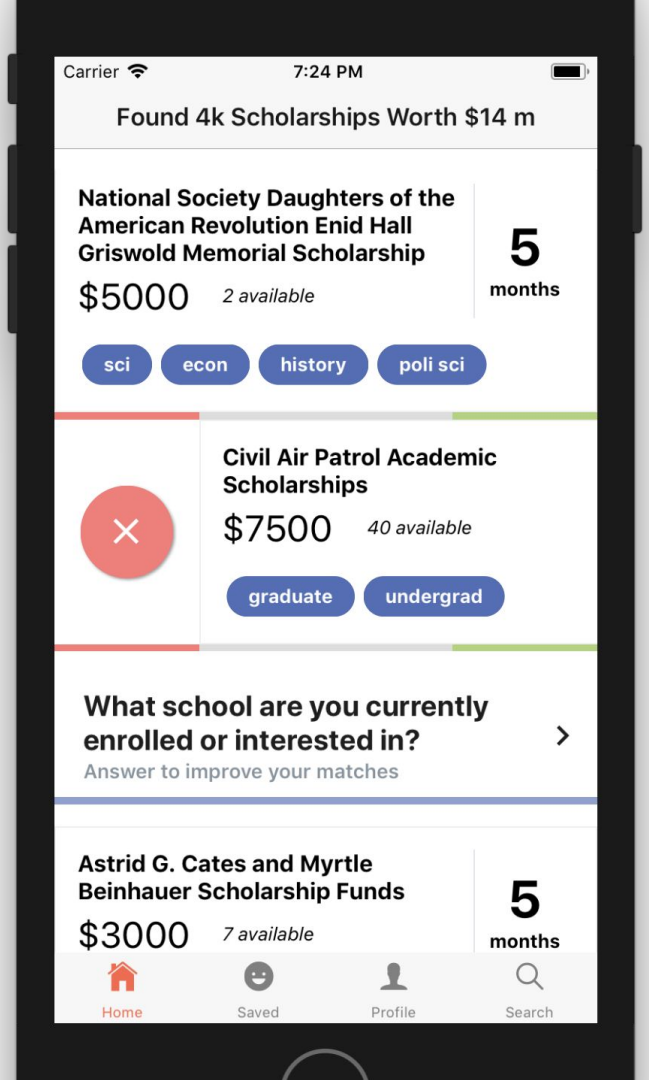


High Precision, Low Recall =>
More False Negatives

High Recall, Low Precision =>
More False Positives

Optimizing for **Recall**

High Recall, Low Precision =>
More False Positives



Implement Model REST API or Python

```
import sys

from google.cloud import automl_v1beta1
from google.cloud.automl_v1beta1.proto import service_pb2


def get_prediction(content, project_id, model_id):
  prediction_client = automl_v1beta1.PredictionServiceClient()

  name = 'projects/{}/locations/us-central1/models/{}'.format(project_id, model_id)
  payload = {'text_snippet': {'content': content, 'mime_type': 'text/plain' }}
  params = {}
  request = prediction_client.predict(name, payload, params)
  return request  # waits till request is returned

if __name__ == '__main__':
  content = sys.argv[1]
  project_id = sys.argv[2]
  model_id = sys.argv[3]

  print get_prediction(content, project_id,  model_id)
```

```
$ python predict.py "SCHOLARSHIP TEXT HERE"
<PROJECT_ID> <MODEL_ID>
```

Scrape Unlabeled Data

Code Example:

https://goo.gl/iALoDJ#Scrape-unlabeled-scholarship-text

Thanks for listening!

Try out Cloud AutoML! cloud.google.com/automl

Inspired by Sarah Robinson's "From Zero to ML"

Tell students & teachers about Grant! trygrant.com