



**Click 'Rate Session'
to rate session
and ask questions.**



follow us @gotochgo



Localization with NLP: Global Empire-Building for Fun & Profit

Michelle Casbon

whoami

- Where I work: Qordoba, Director of Data Science
- Where I used to work: **iDIBON**
- What I love
 - Natural language processing
 - Distributed systems
 - Emoji One

@texasmichelle



Data Science Engineer



What my friends think I do



What my boss thinks I do



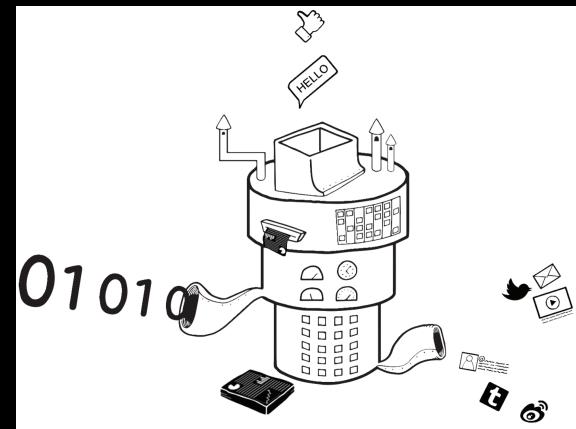
What my parents think I do



What I think I do

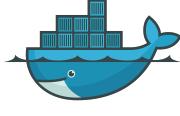


What society thinks I do



What I actually do

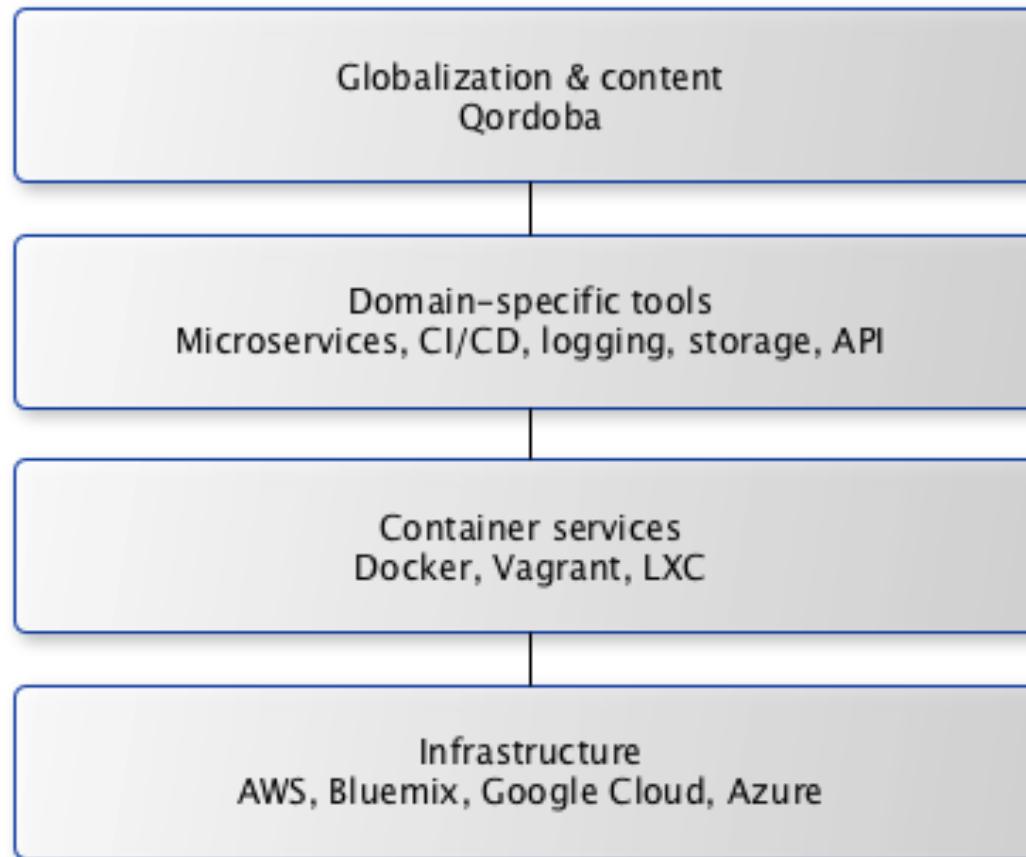
TL;DR

- Mission
 - To enable products that feel native to every user
 - Default: a product is in 100+ markets
- There's a better way to do localization
 - Hint: it involves open source    
- How to operationalize & scale NLP models

@texasmichelle



The of content

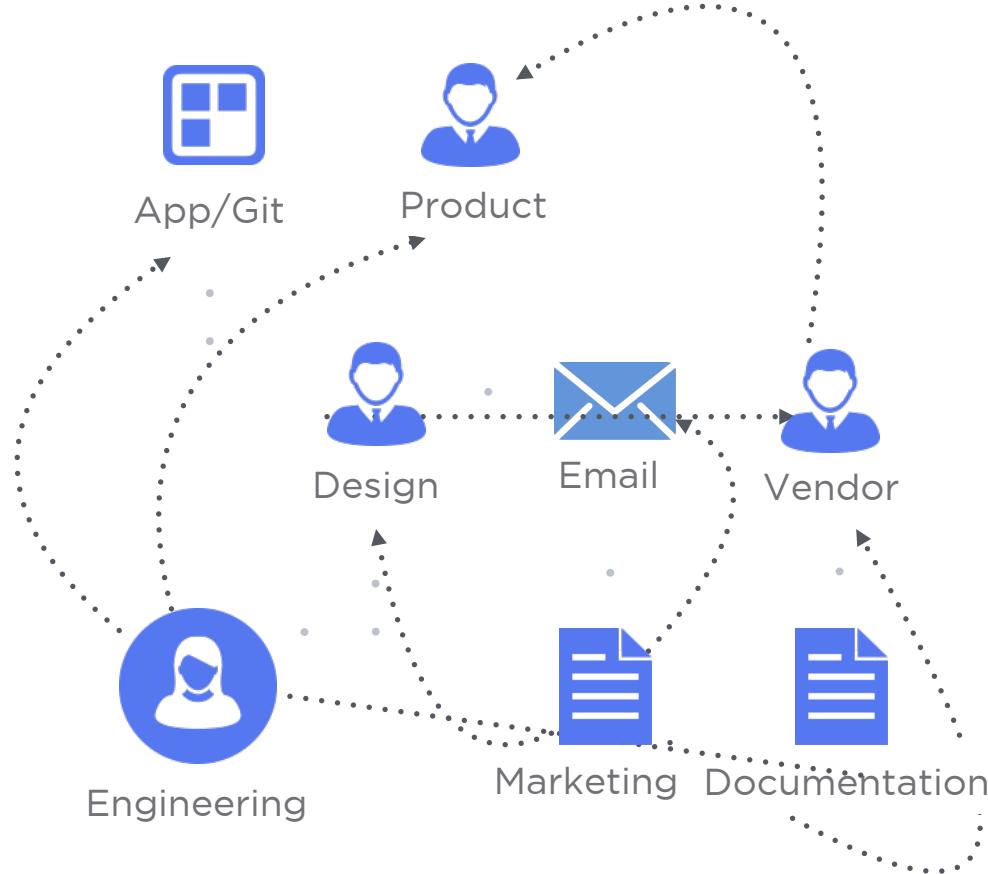


@texasmichelle



Localization is hard

- People
 - Product managers
 - Marketers
 - Designers
 - Linguists
 - Engineers
- Things
 - Copies of copy
 - String files
 - Emails
 - Pull requests
- Results
 - Got milk?
 - Hotline bling
- Wash, rinse, repeat



@texasmichelle





Ich sehe ein, wann diese
Telefondienst auffällt, das kann
nur eine Sache bedeuten...

I realize that when this
telephone service flashes, it
can only mean one thing...

I KNOW WHEN THAT HOTLINE BLING,
THAT CAN ONLY MEAN ONE THING...

NICOLEMILLER.COM
MOBILE IS LIVE!

THE FIRST 20 CUSTOMERS WILL GET A COMPLIMENTARY
IPHONE 6 CASE GIFT WITH MOBILE PURCHASE.

SHOP NOW

Localization is hard

- People
 - Product managers
 - Marketers
 - Designers
 - Linguists
 - Engineers
- Things
 - Copies of copy
 - String files
 - Emails
 - Pull requests
- Results
 - Got milk?
- Wash, rinse, repeat

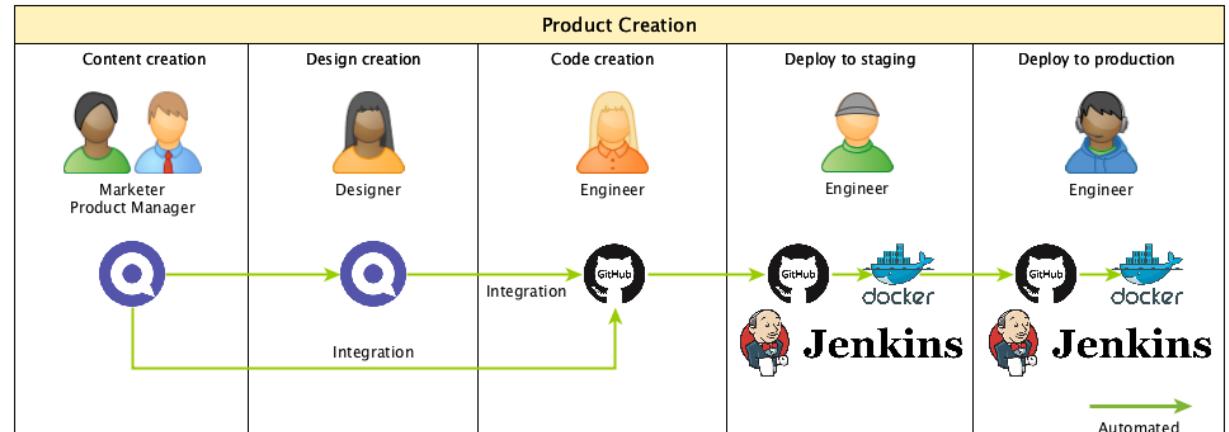


@texasmichelle



Platform

- Code points to a dynamic link 
- Other entities update the 
 - Marketers
 - Linguists
 - ML models
- Live changes
- Consistency among mediums
- Real-time translations



Platform

- Context
 - Linguists
 - Designers
- Github repo receives PRs



The image shows a GitHub pull request interface with two status cards side-by-side. The left card is red and says "All checks have failed" with one failing check from "Qordoba". The right card is green and says "All checks have passed" with two successful checks from "jenkins-ci" and "Qordoba/I10n". Both cards include a "Merge pull request" button at the bottom.

Status	Check Details
All checks have failed	1 failing check from Qordoba
All checks have passed	2 successful checks from jenkins-ci and Qordoba/I10n

@texasmichelle





BUY
ALL THE THINGS





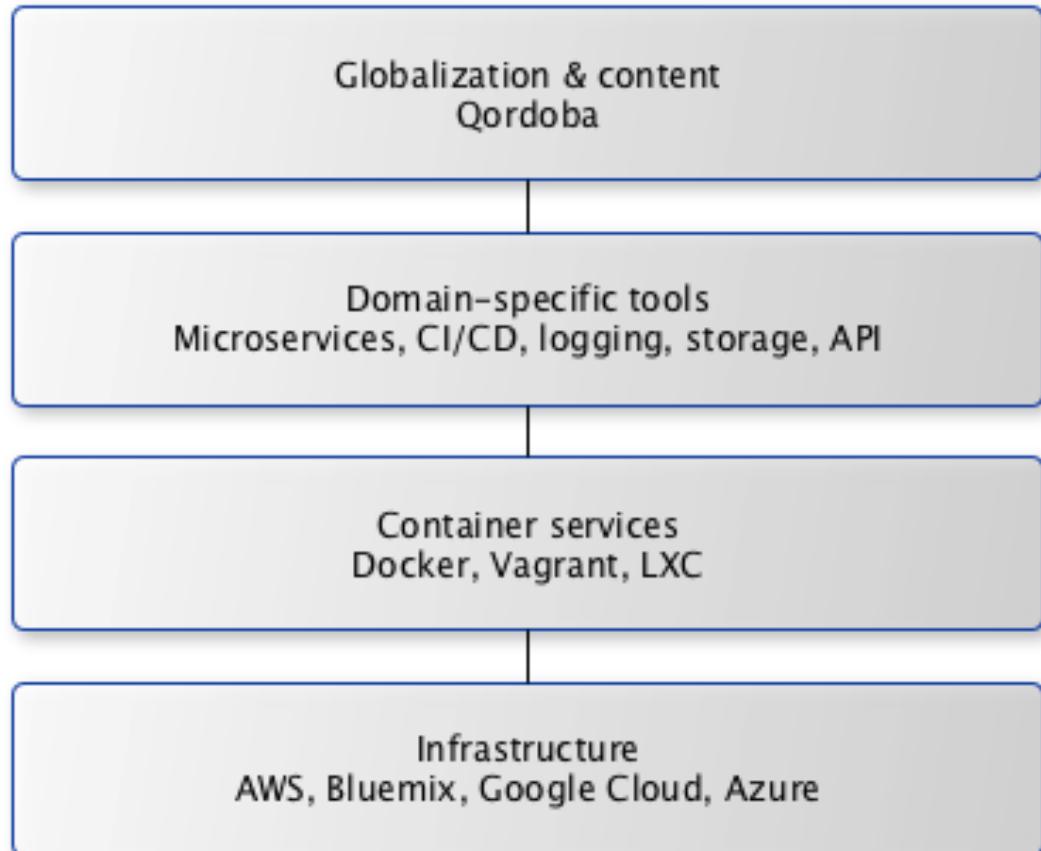
I Followed My Stolen iPhone Across The World, Became A Celebrity In China, And Found A Friend For Life ?

https://www.buzzfeed.com/mjs538/i-followed-my-stolen-iphone-across-the-world-became-a-celebr?utm_term=.ut02j4dGN4#.txnrxkyjGk



Containers

- Powered by   
- All the way down
 - Legal 
 - Informal 
 - Somber 



@texasmichelle



Affect Detection

- What is it?
- Why do we want it?
 - Hands-off translations 🙌
 - Workflow transitions 🔄



@texasmichelle



Affect Detection

- I had dinner with my wife  fear
- I had dinner with my girlfriend  joy
- Help Wanted  sadness
- Busca empleo  anger
- This apartment is in a killer location  joy
- Diese Wohnung befindet sich an einem mörderischen Standort  fear



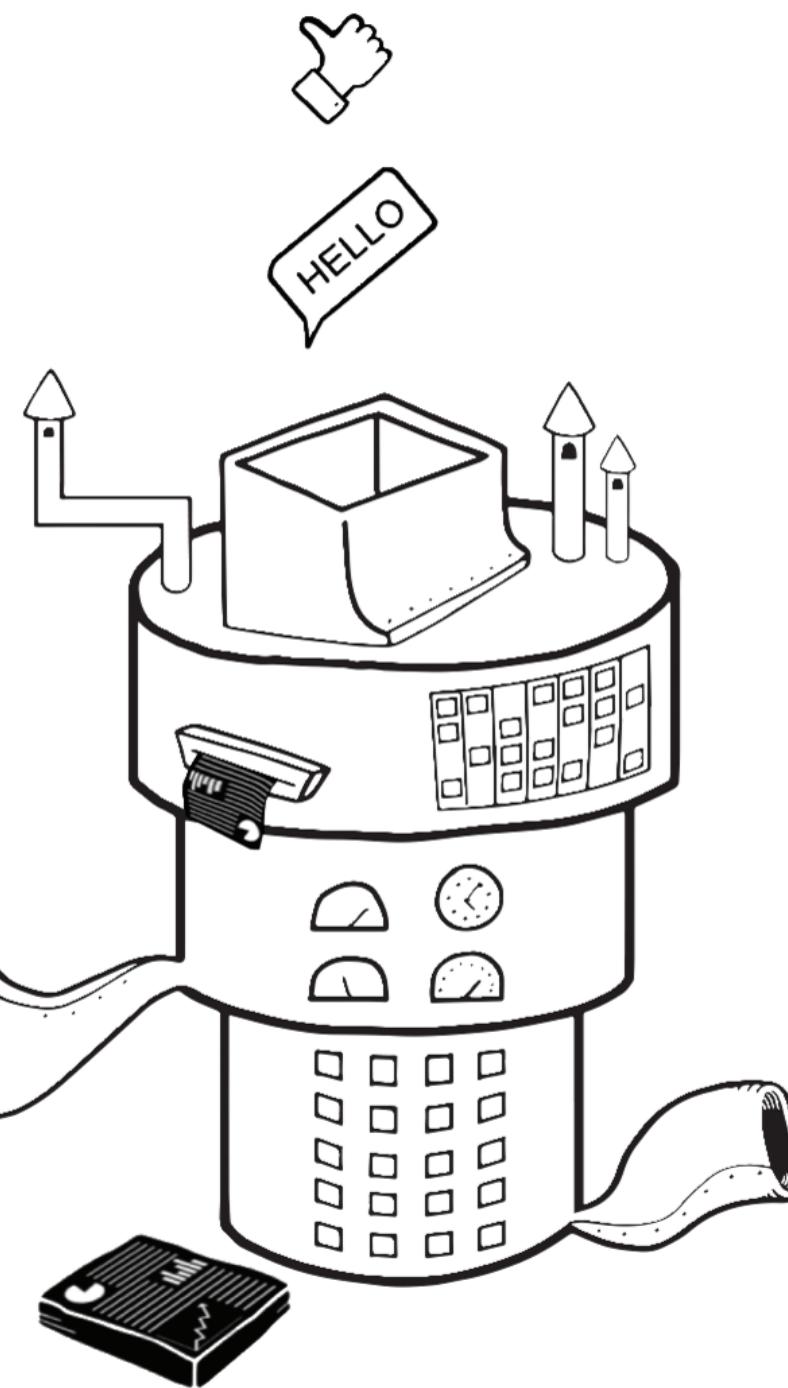
Requirements

- REST interface 
 - Response time
- Scalability   
 - Deployment
 - Models
 - Languages
- Accuracy 
 - Accuracy
- Open source 
 - Open source

@texasmichelle



01010



Affect Detection

-  came early
 - PredictionIO entered ASF Incubator
 - Existing Apache community 
 - REST interface
 - Response time
 - Scalability
 - Clean objects 
 - Engine
 - Evaluator
 - Algorithm
 - ...

@texasmichelle



Not like this....



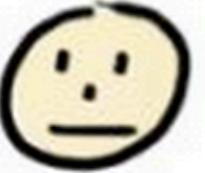
1

2

3

4

Like this!



1



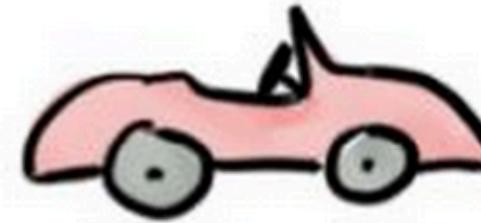
2



3



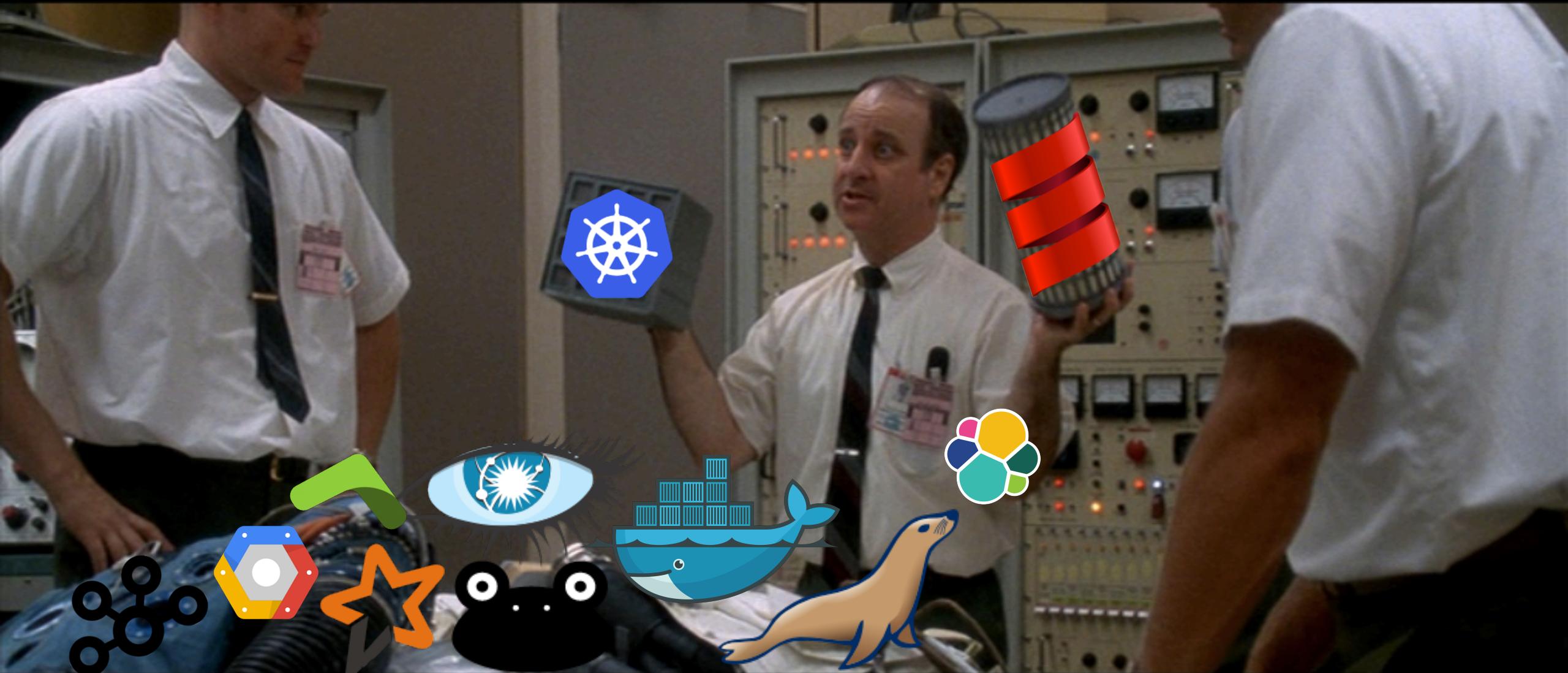
4



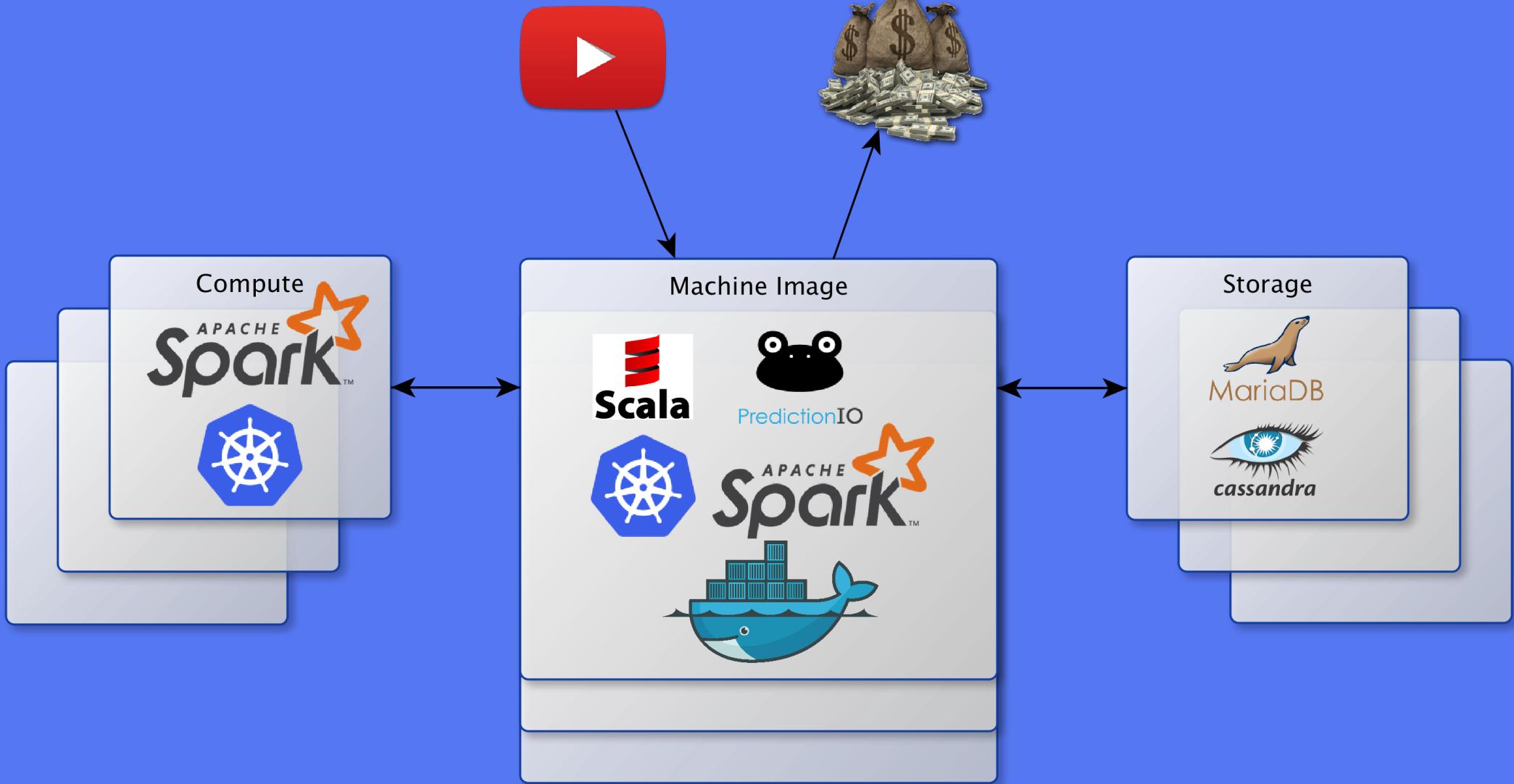
5

What does the  look like?

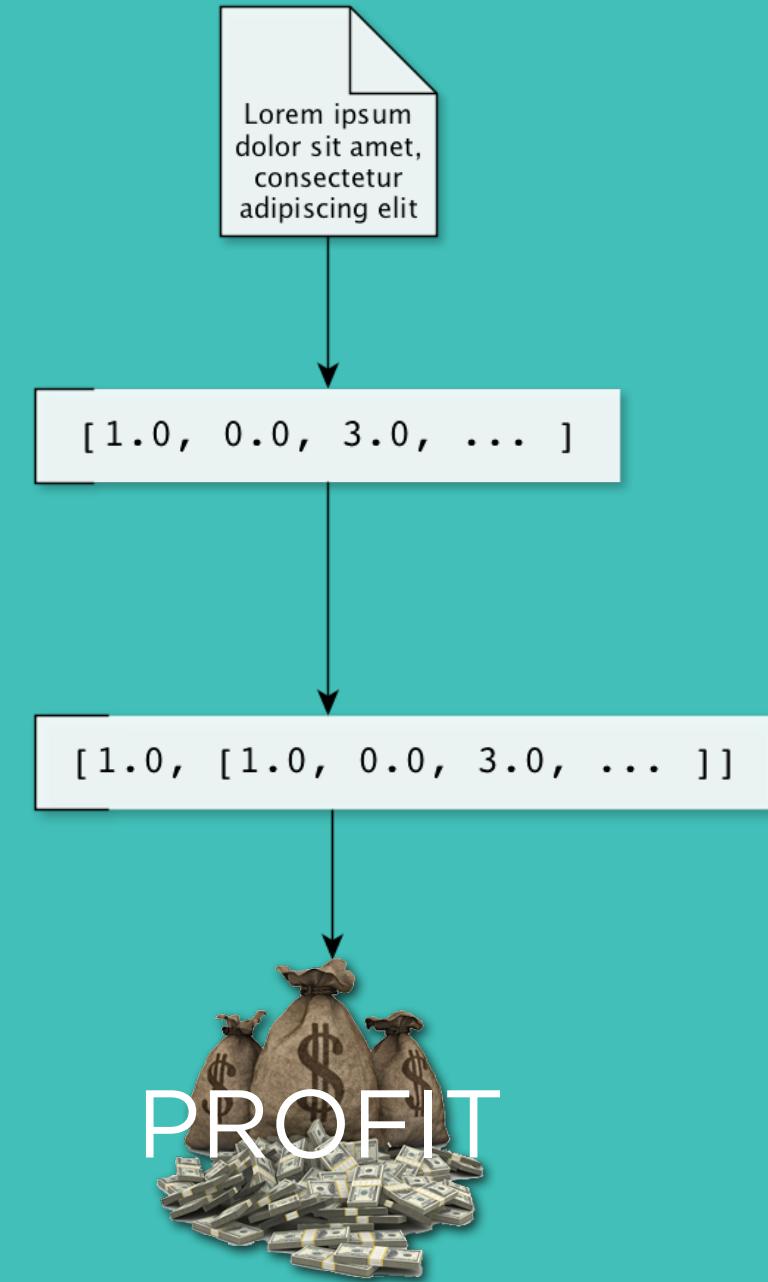
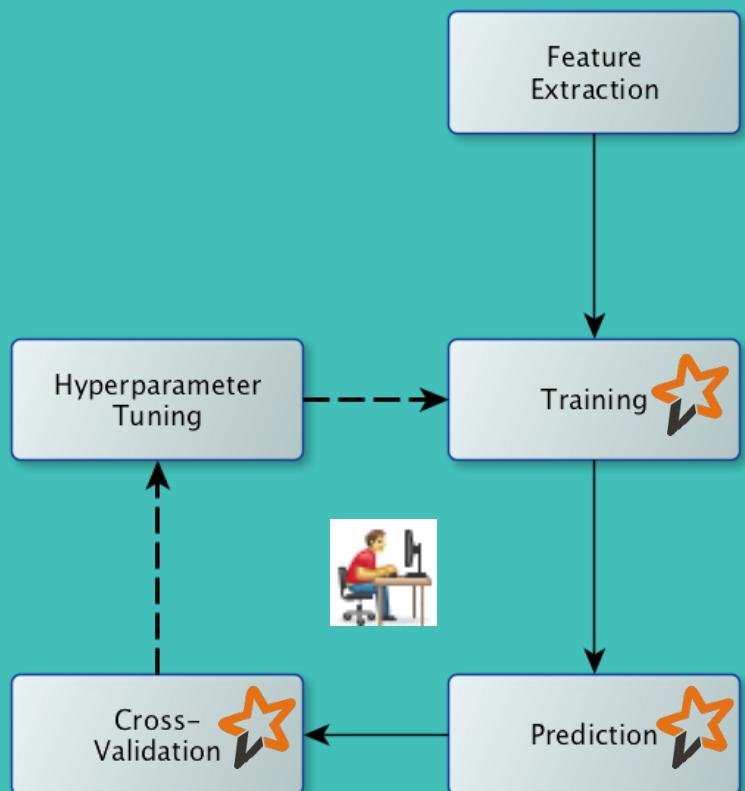


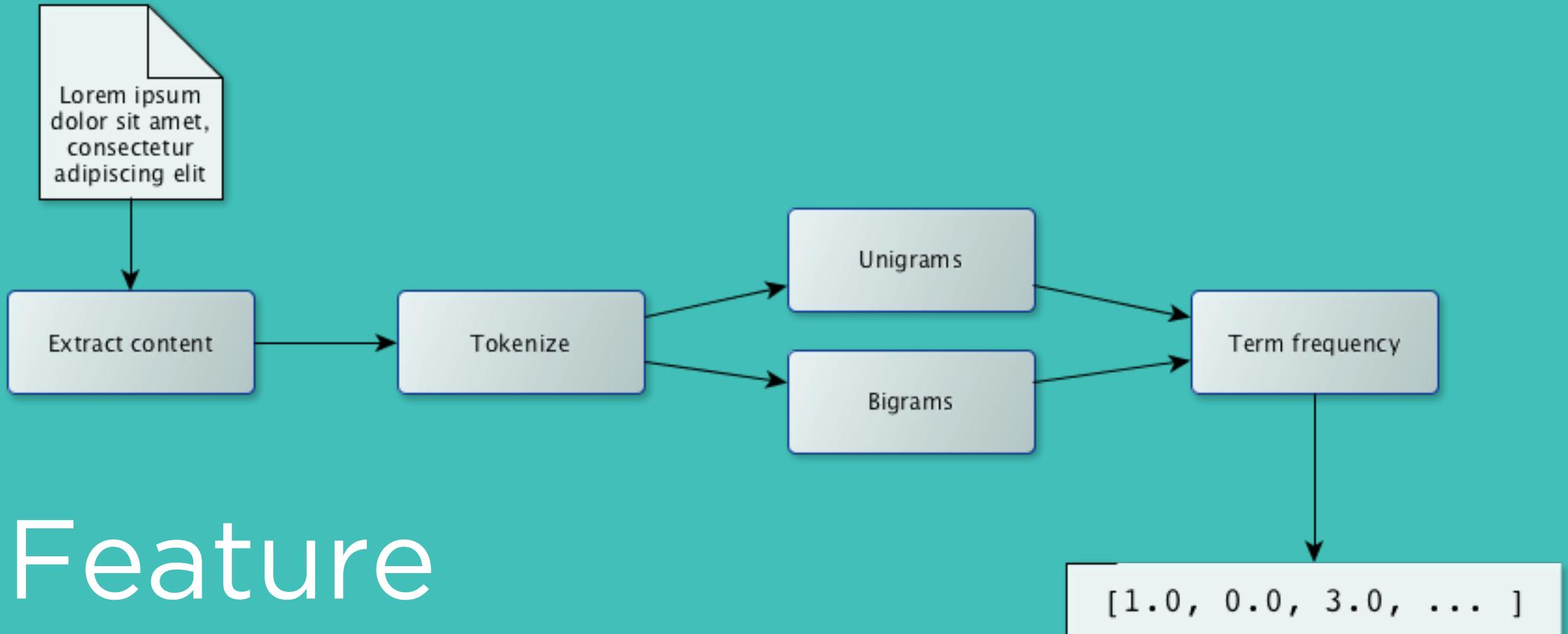


PROFIT

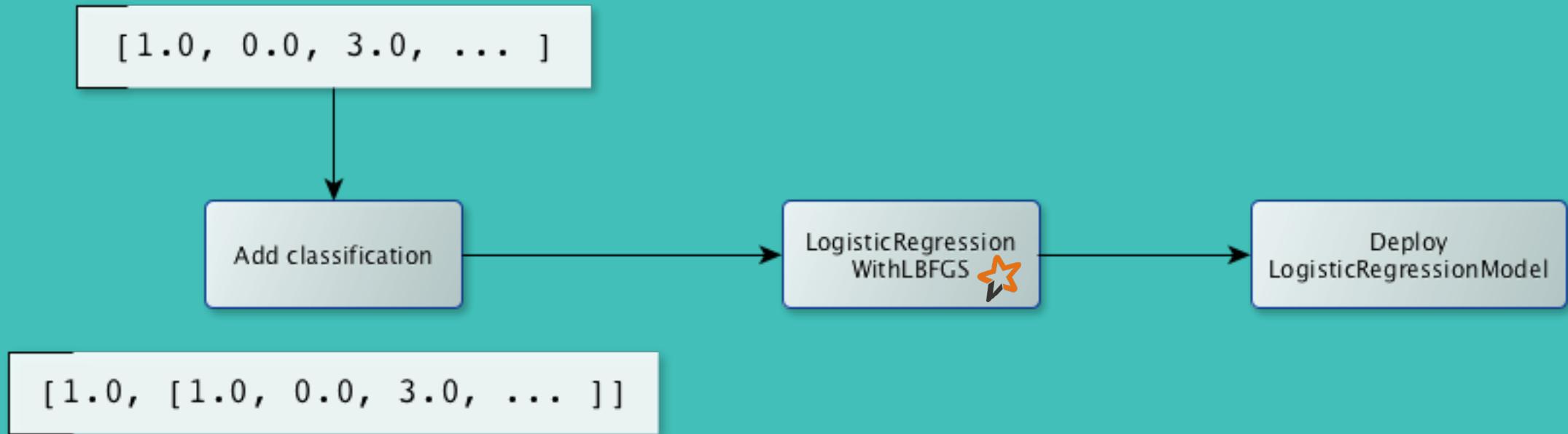


NLP

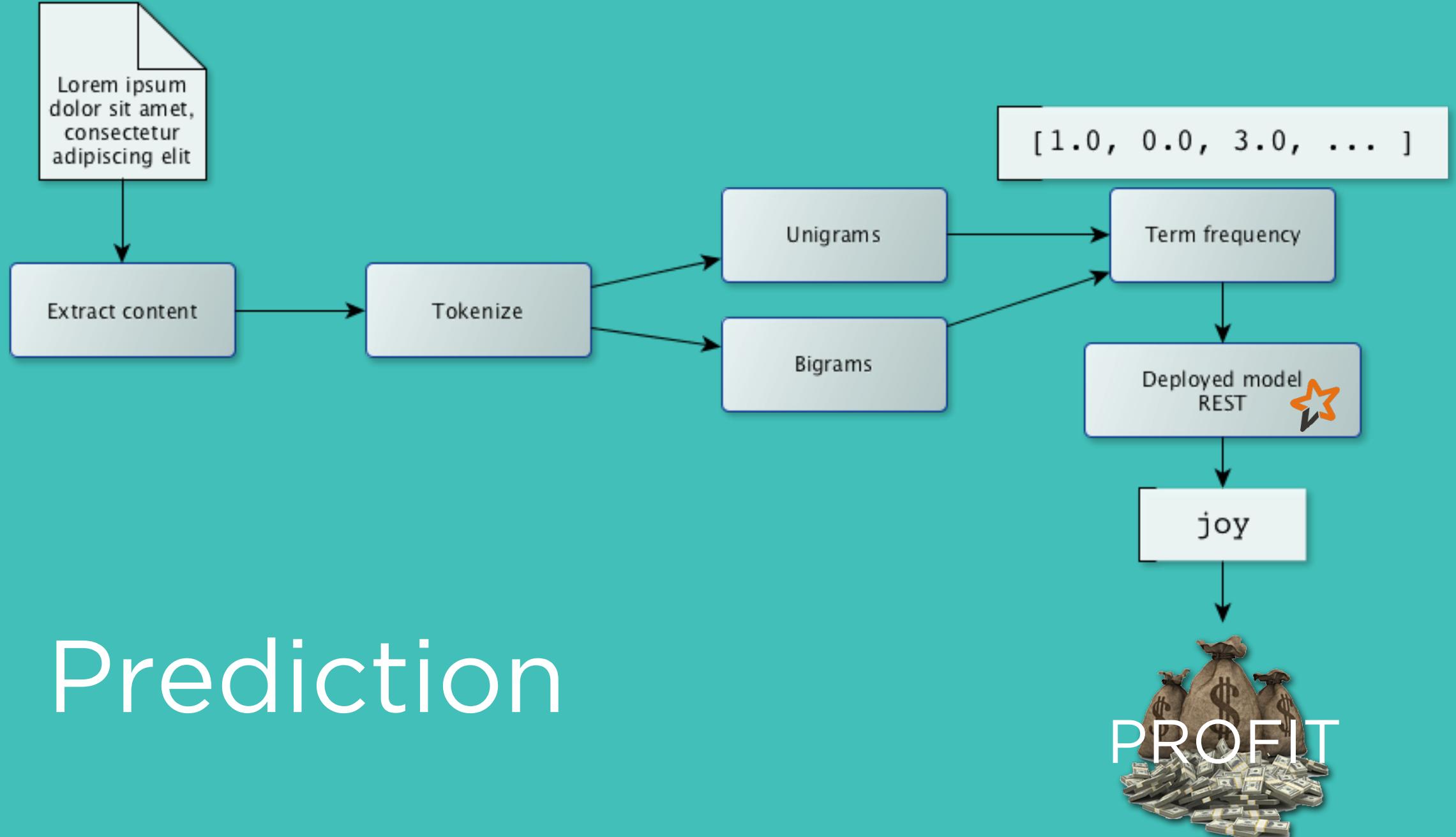




Feature Extraction

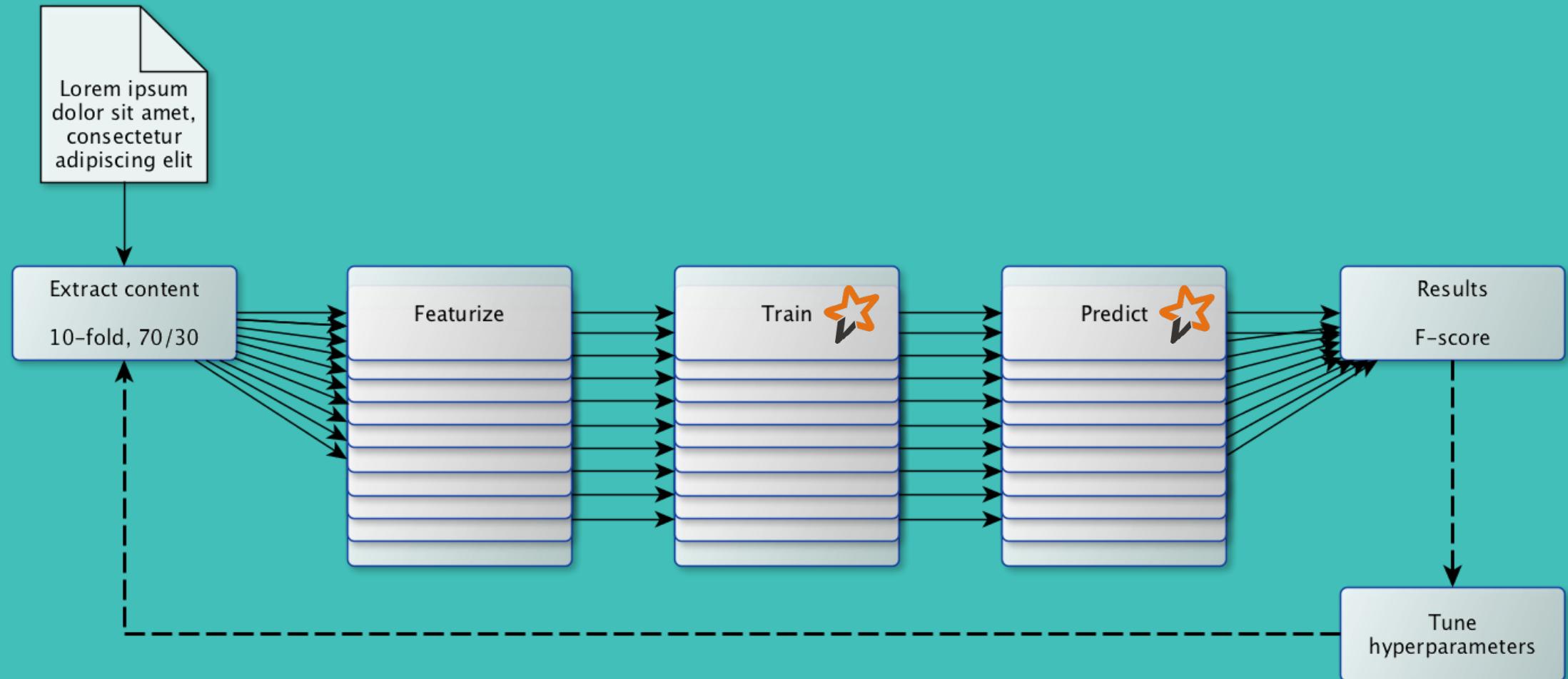


Training



Prediction

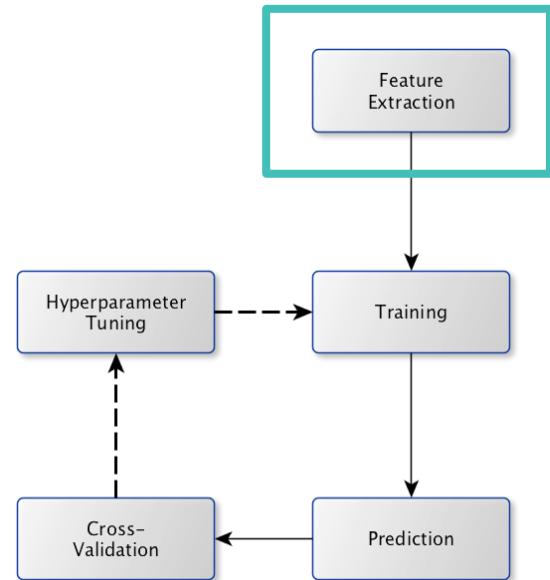
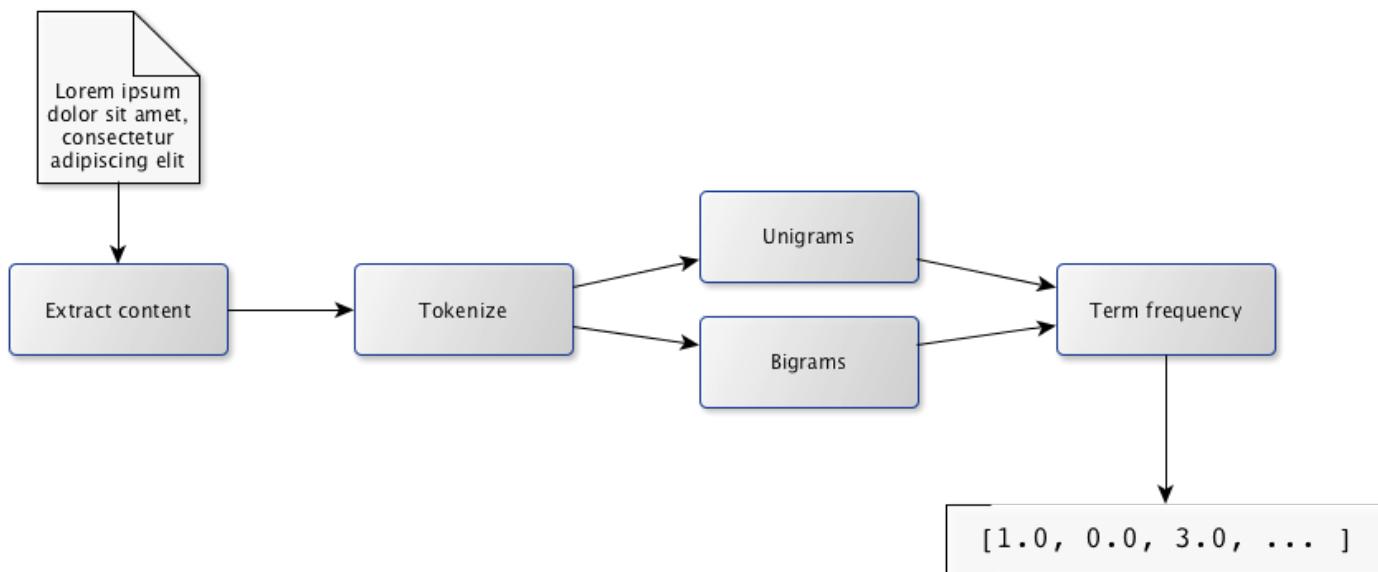




Cross-validation

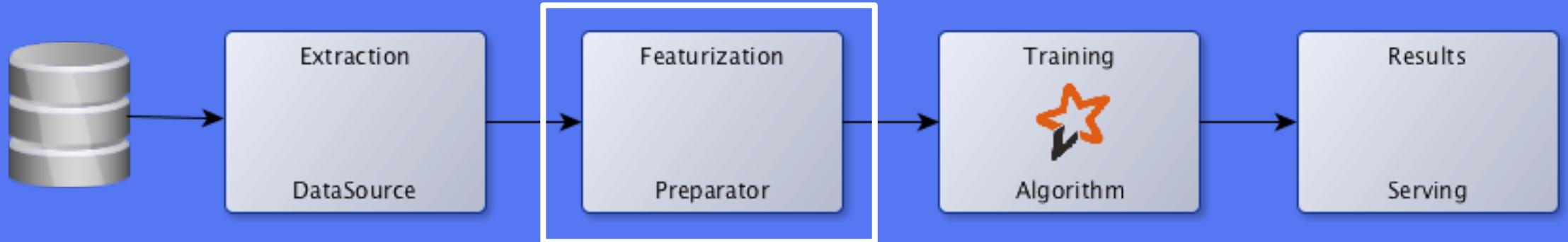
Featurization

- UTF8 support
- Simple
 - Term frequency, descending
 - Ngrams



@texasmichelle

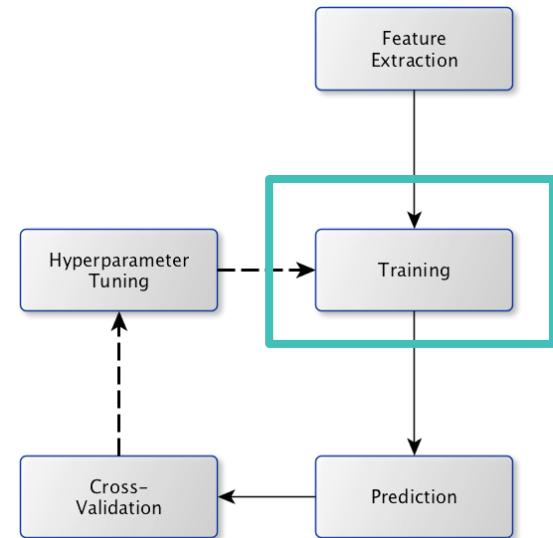
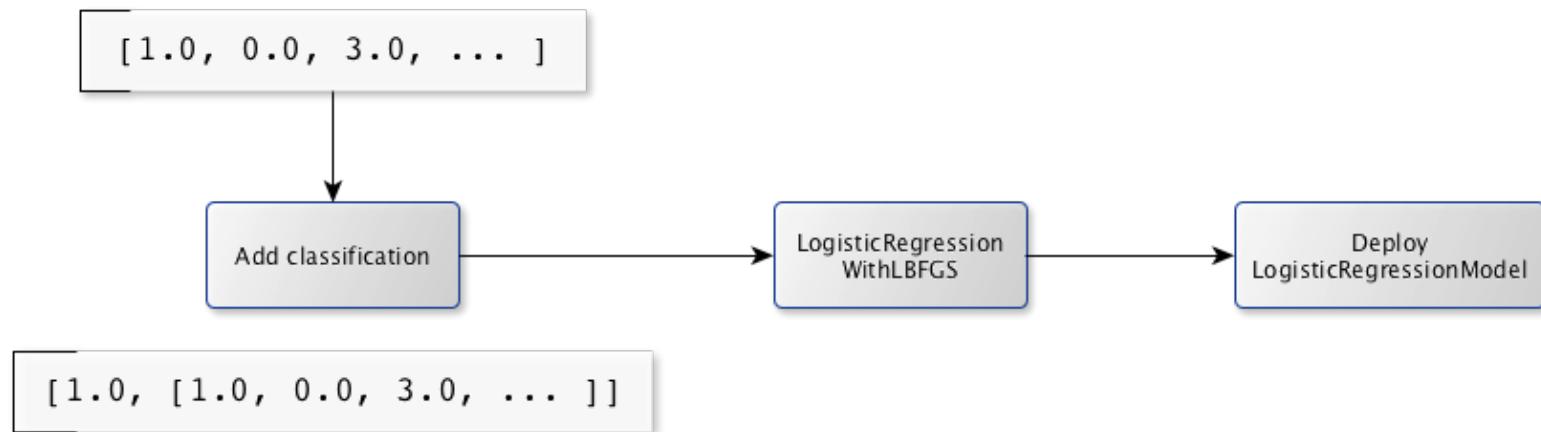


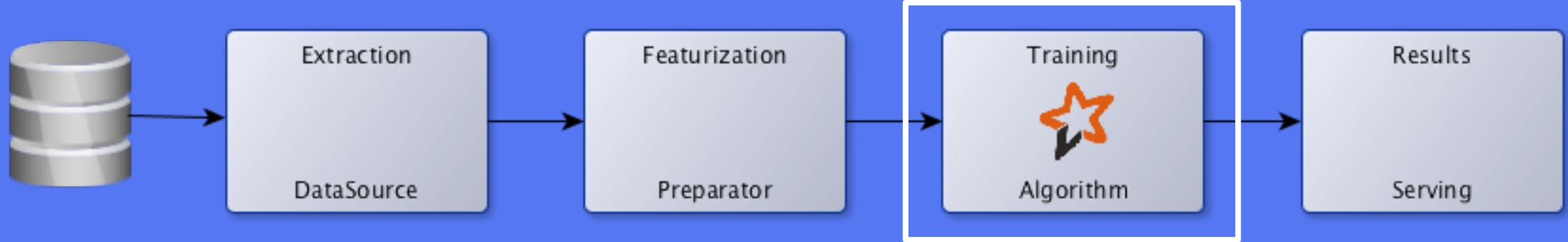


Featurization with PredictionIO

Training

- Supervised models
- Multi-class classification
 - Logistic regression

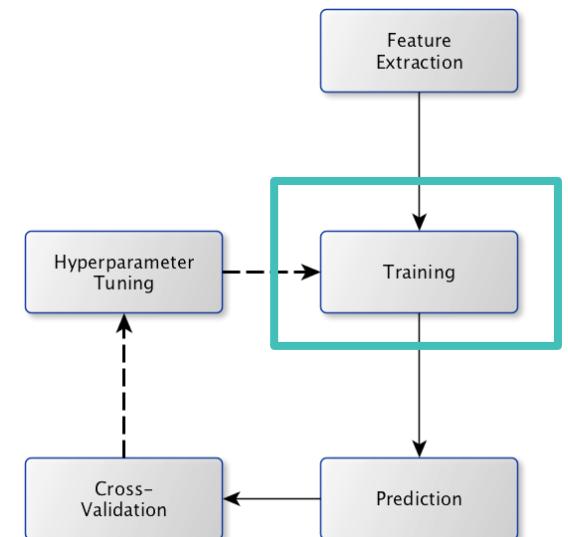
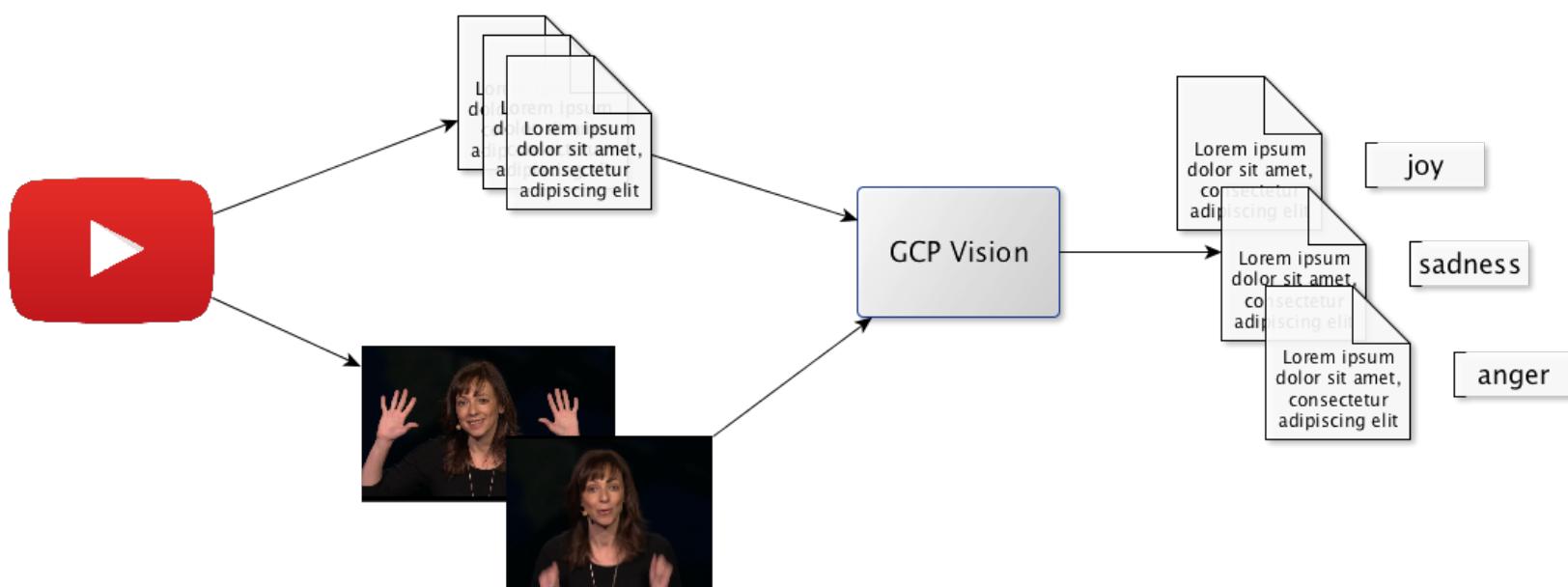




Training with PredictionIO

Training

- Auto-generated

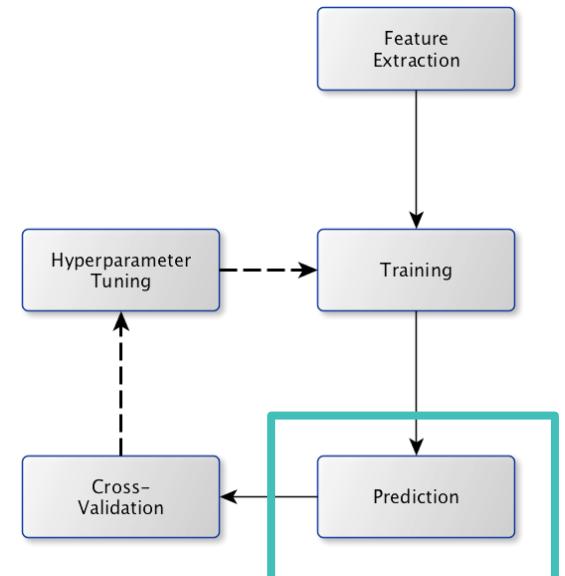
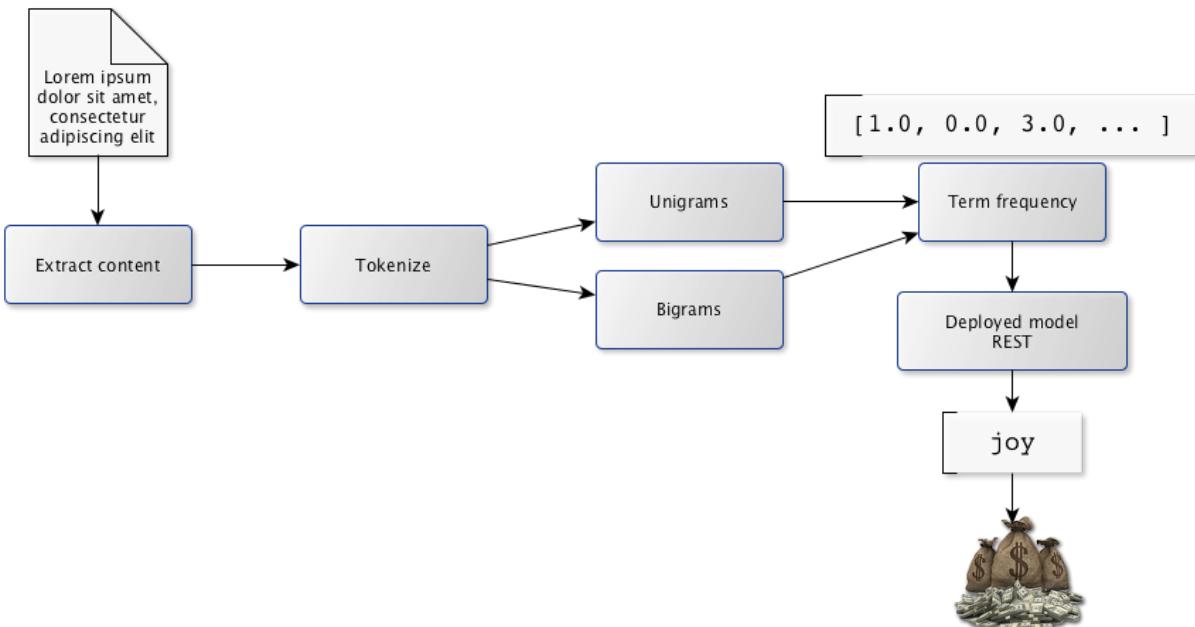


@texasmichelle



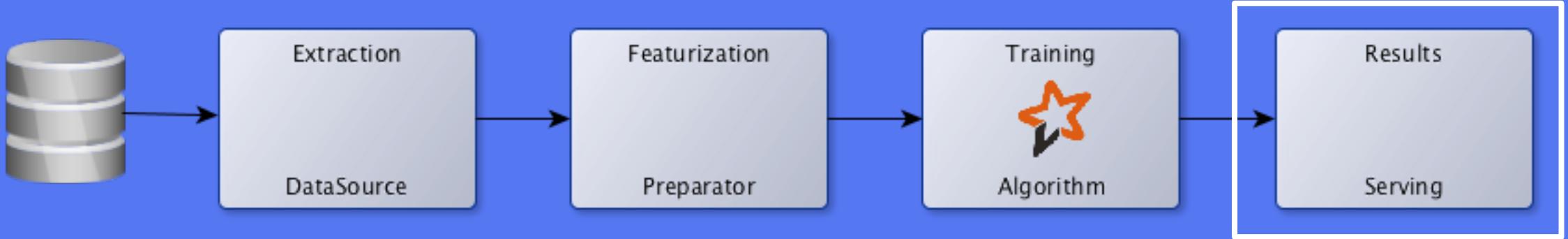
Prediction

- Featurization
- Router
 - REST
 - Pub/Sub



@texasmichelle

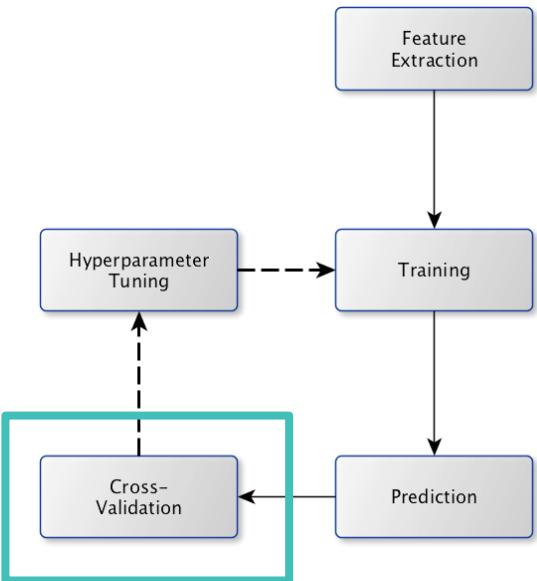
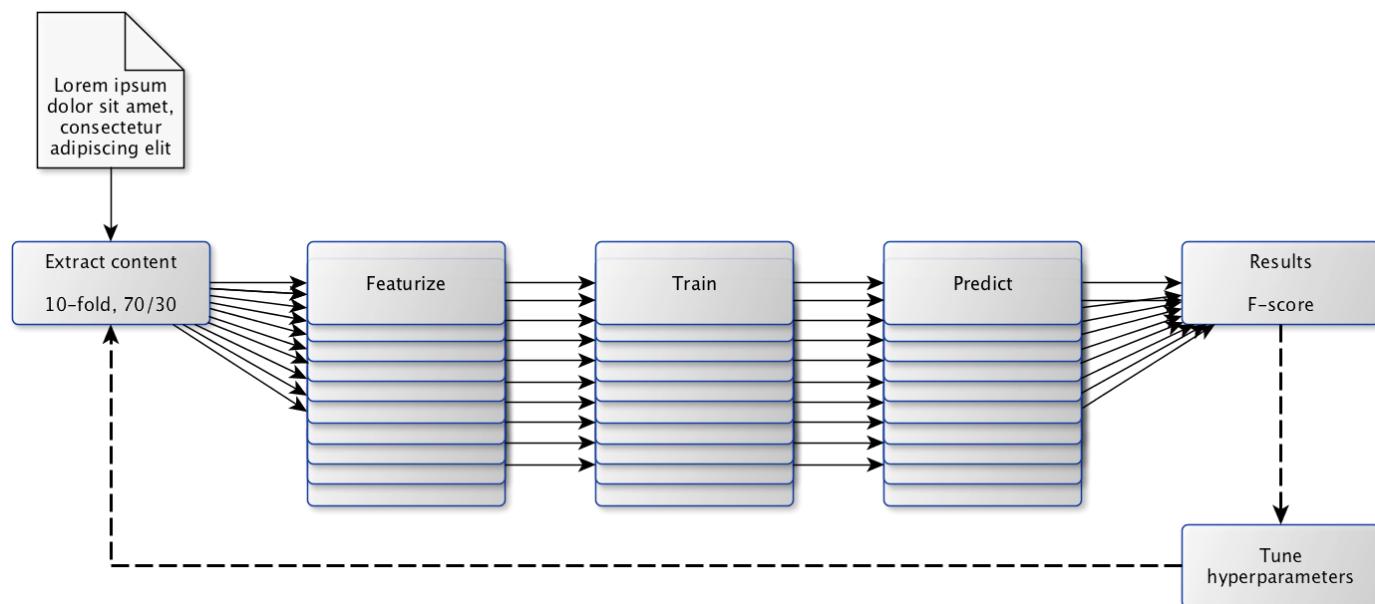




Prediction with PredictionIO

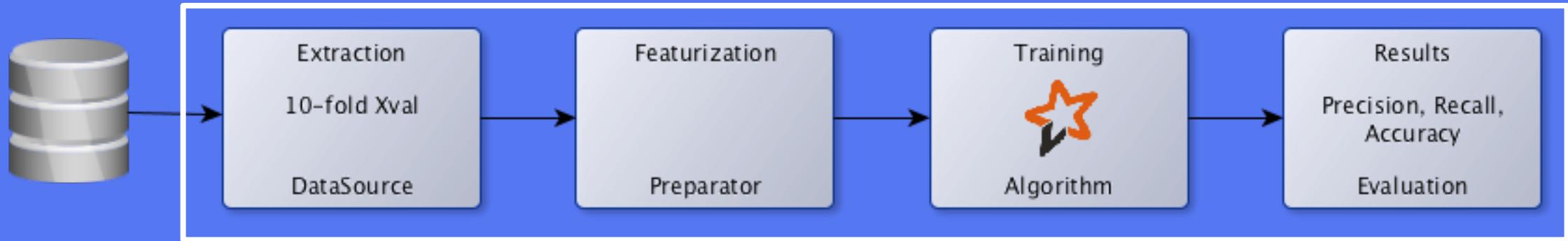
Cross-validation

- F-score
 - Precision
 - Recall
 - Accuracy
- AUC



@texasmichelle

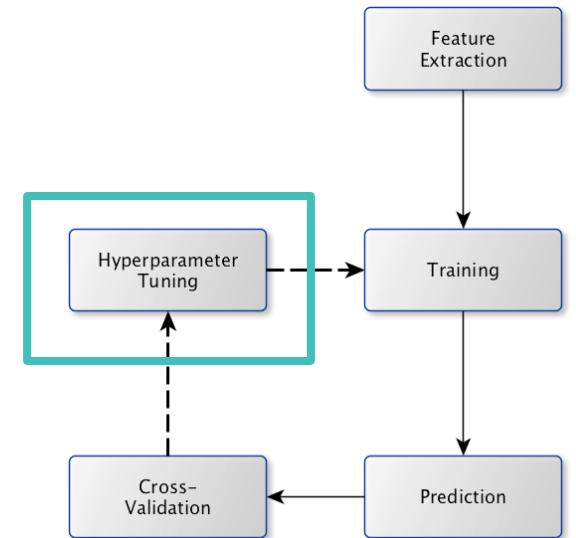
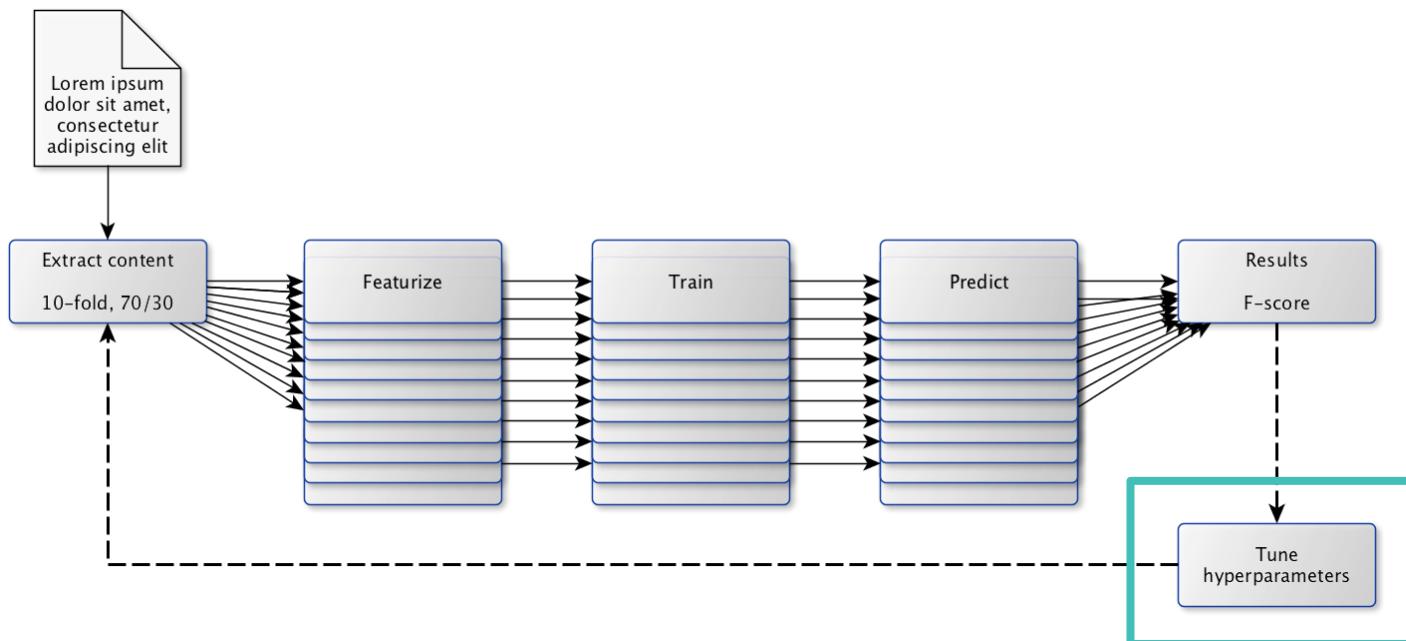




Cross-validation with PredictionIO

Hyperparameter Tuning

- Max iterations
- Regularization parameter



```
{  
    "id": "default",  
    "description": "Default settings",  
    "engineFactory": "com.qordoba.ml.classification.ClassificationEngine",  
    "datasource": {  
        "params": {  
            "appName": "LRBFGS",  
            "languageCode": "en",  
            "countryCode": "US",  
            "numRecords": 60000,  
            "distributeLabelsEvenly": false  
        }  
    },  
    "preparator": {  
        "params": {  
            "tokenizerLanguage": "en",  
            "tokenizerCountry": "US"  
        }  
    },  
    "algorithms": [  
        {  
            "name": "logisticLBFGS",  
            "params": {  
                "maxIter": 10,  
                "regParam": 0.55  
            }  
        }  
    ]  
}
```

Hyperparameter tuning with PredictionIO



logstash

Logging & Reporting



Continuous Delivery



Jenkins



Scala



akka



Microservices



GitHub



Event Bus



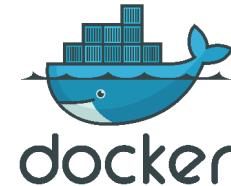
Cloud Pub/Sub



kafka



JFrog Artifactory



docker

Storage



MariaDB



cassandra

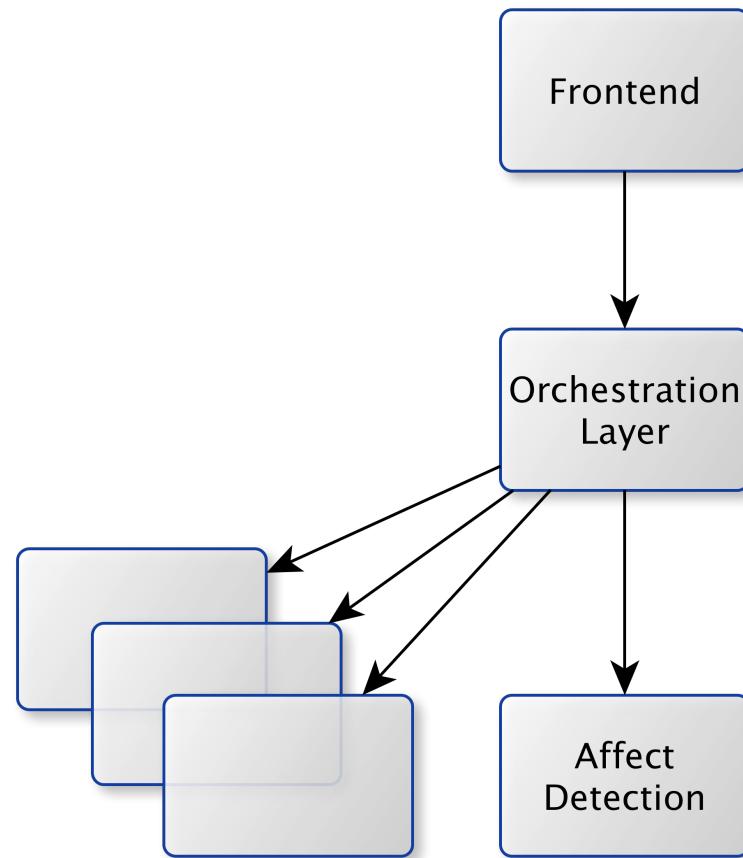


MESOS



kubernetes

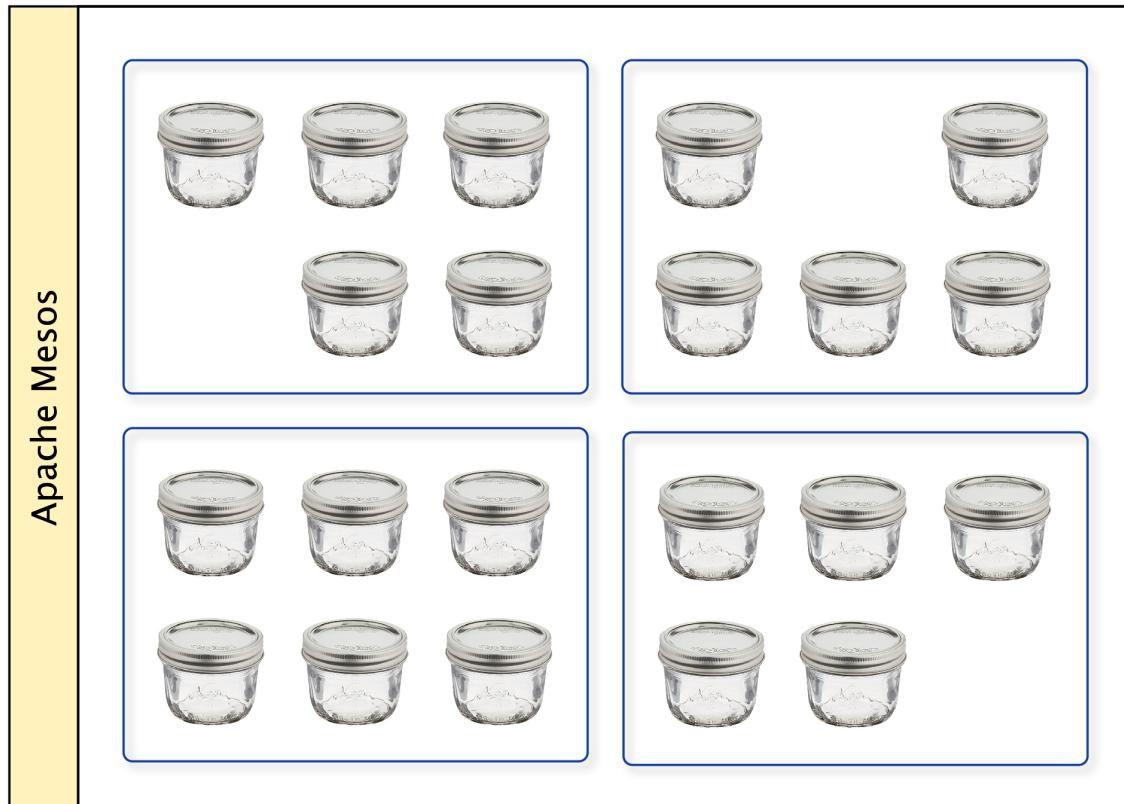
System Components



@texasmichelle



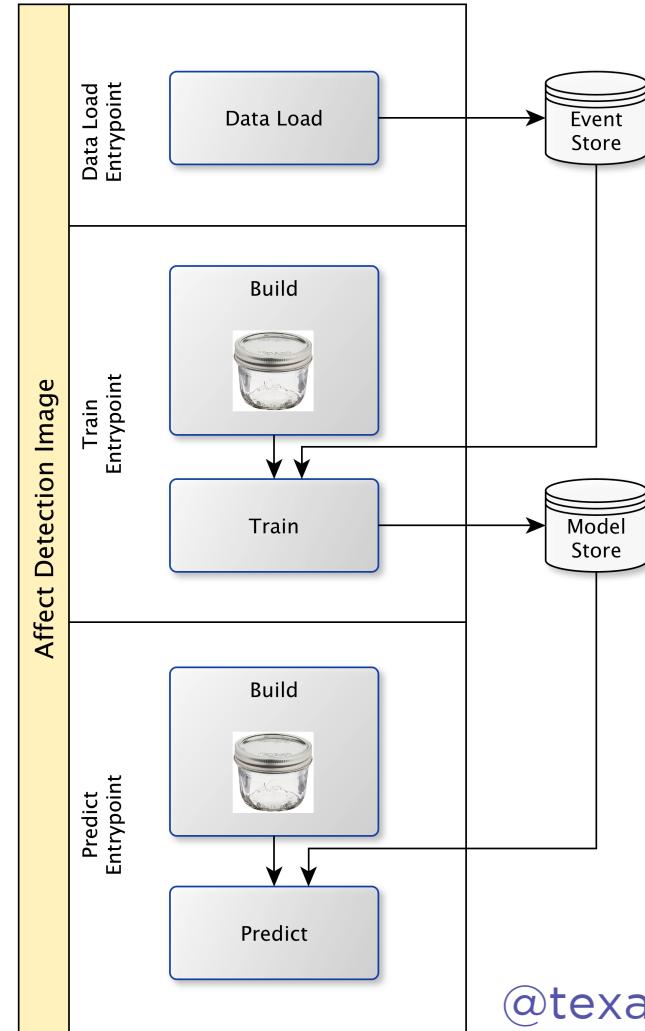
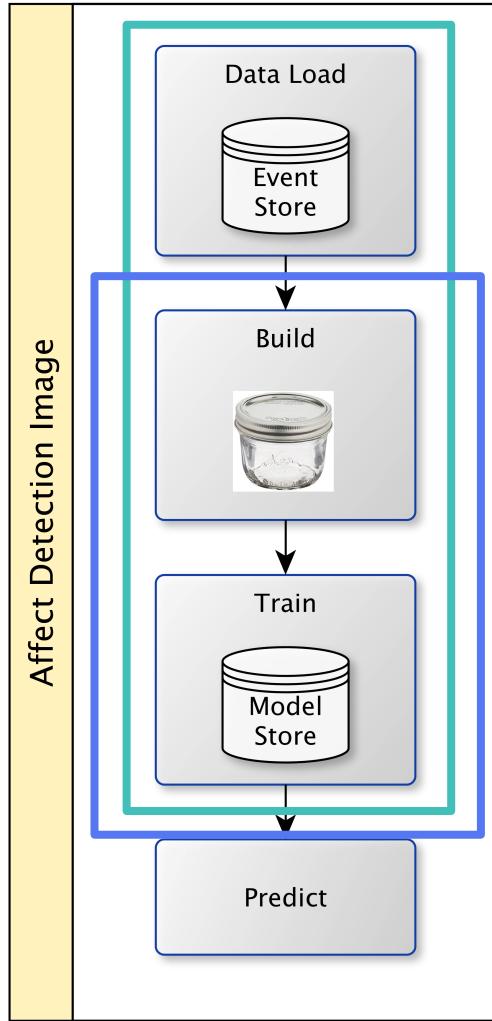
Before



@texasmichelle



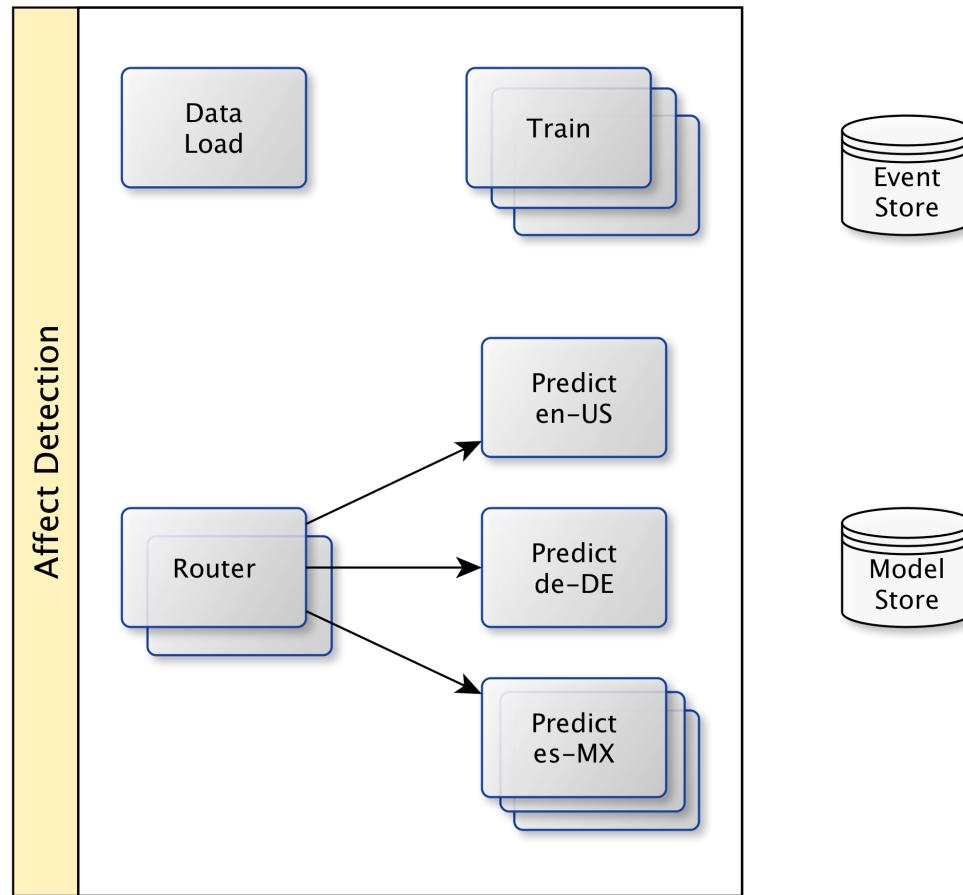
Single Model



@texasmichelle



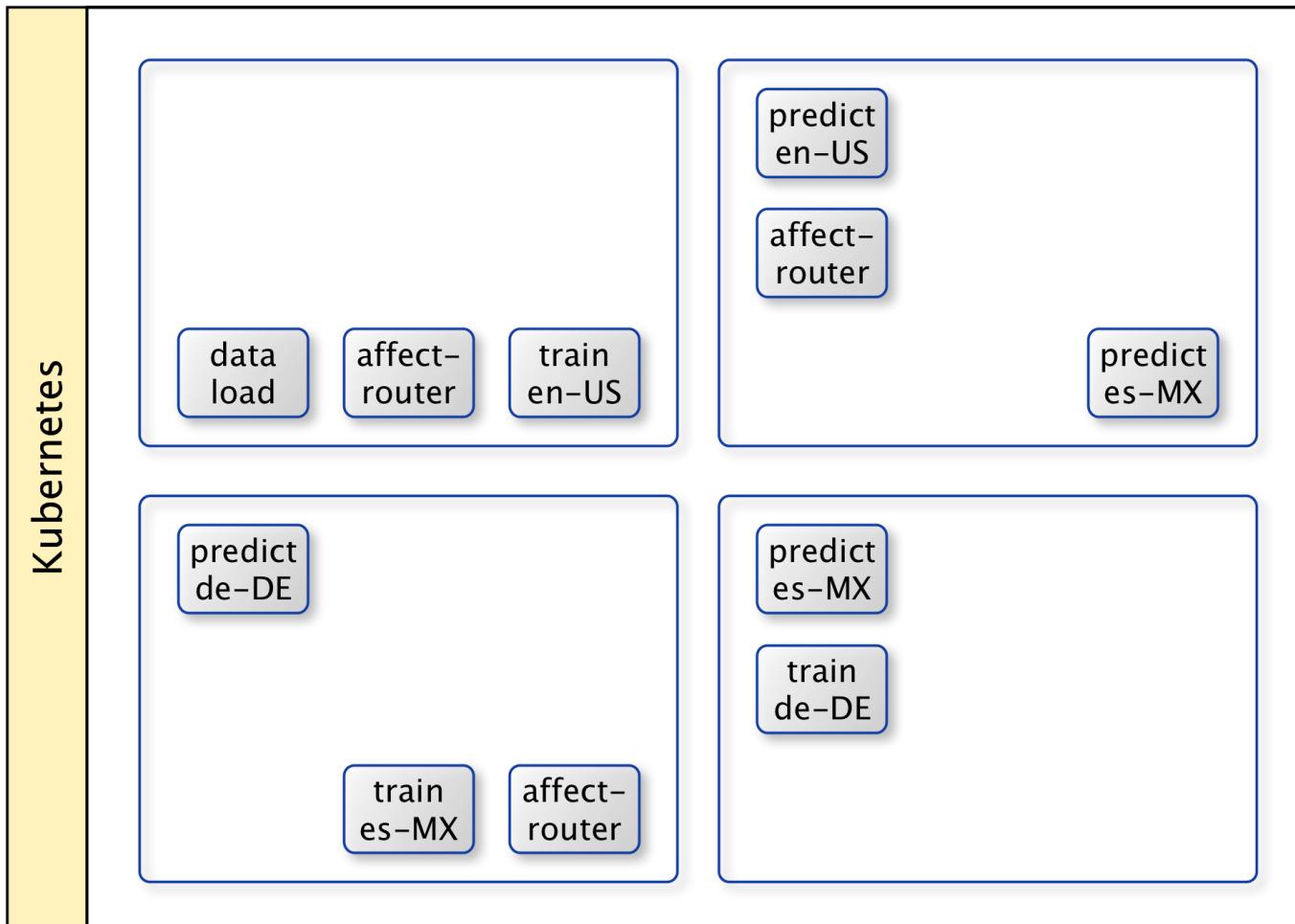
Many Models



@texasmichelle



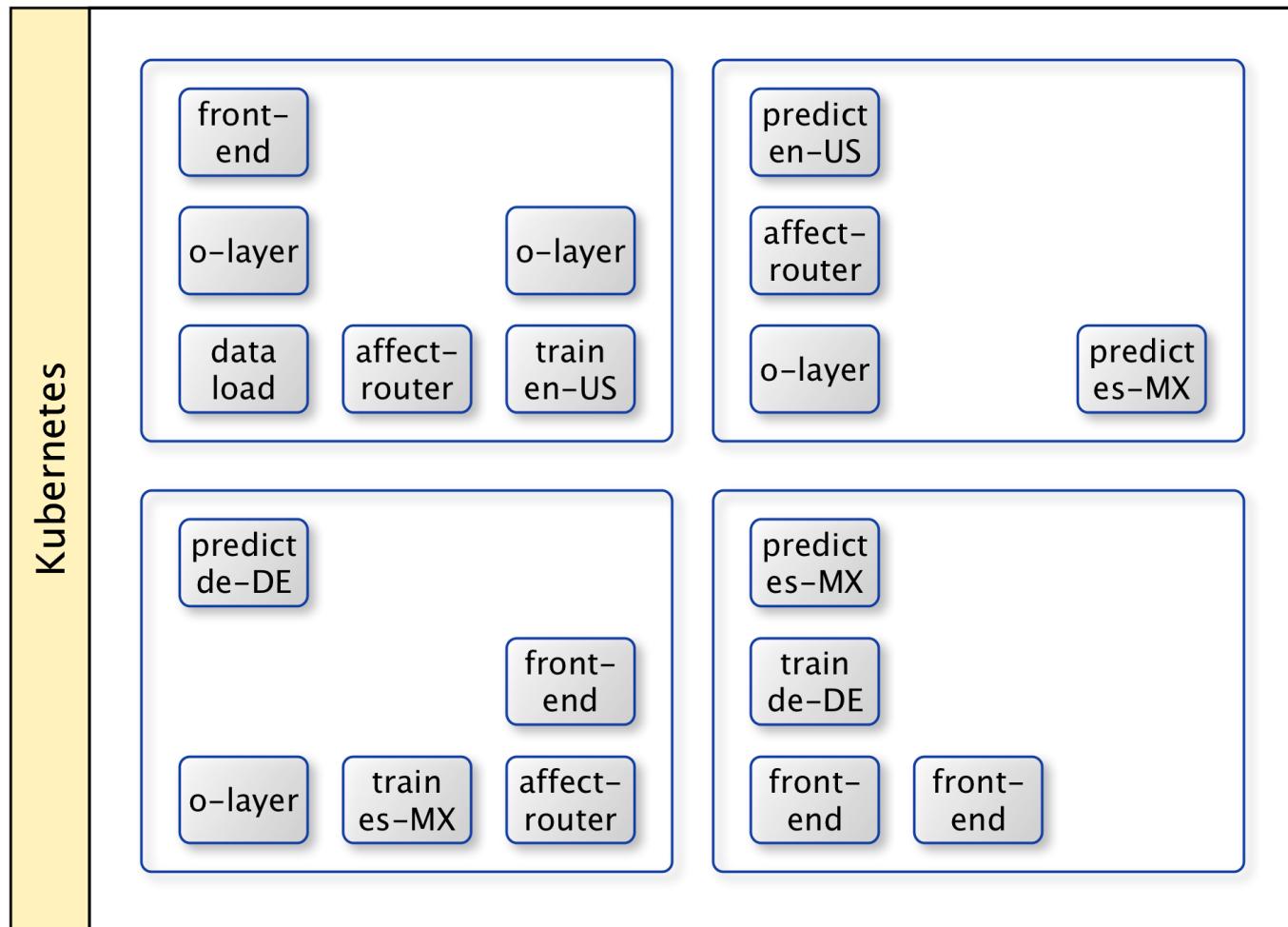
Deployment



@texasmichelle



After



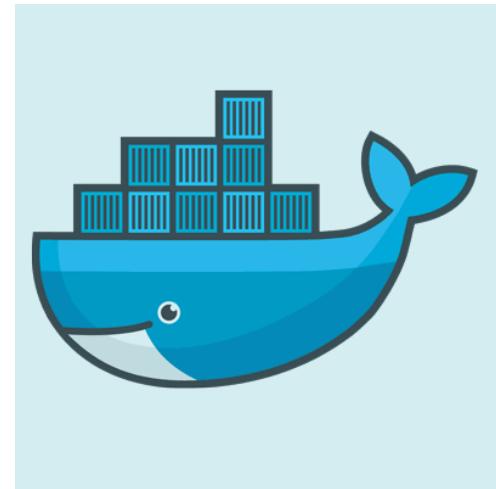
@texasmichelle



Affect Detection Image

Affect Detection
Mysql Client
Apache PredictionIO (incubating) 0.10.0
Apache Spark 1.6.3 Native BLAS
Java 8
Ubuntu 16.04

Affect Detection
Mysql Client
Apache PredictionIO (incubating) 0.10.0
Apache Spark 1.6.3 Native BLAS
Java 8
Ubuntu 16.04



@texasmichelle



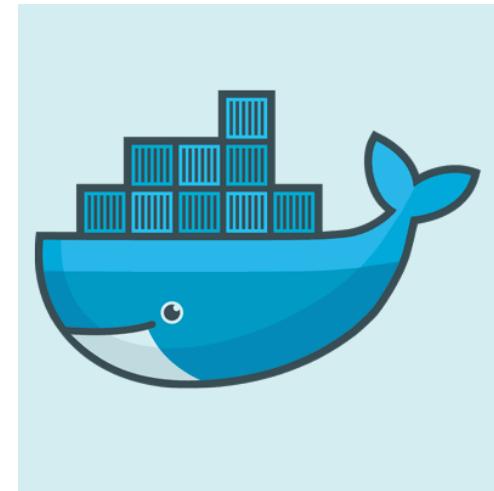
Router Image

affect-router

qordoba-builder

affect-router

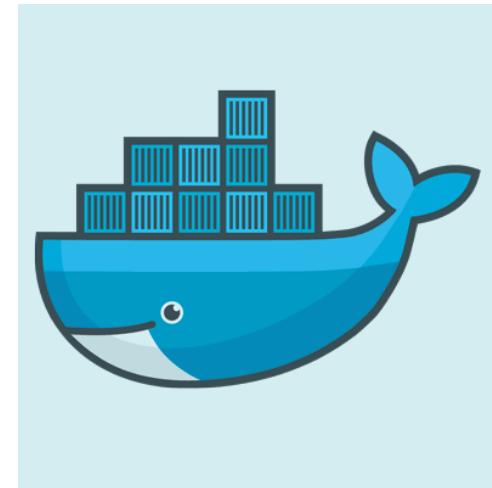
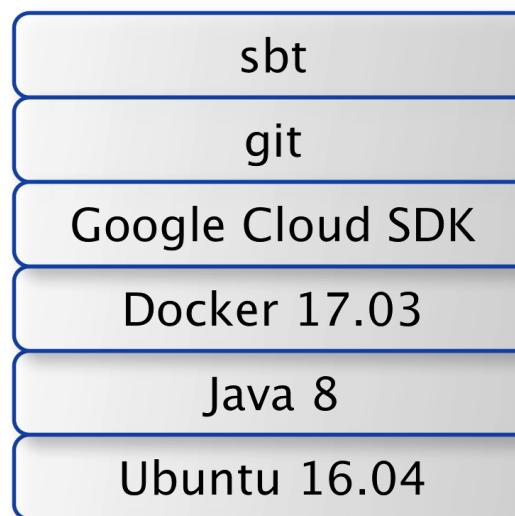
qordoba-builder



@texasmichelle



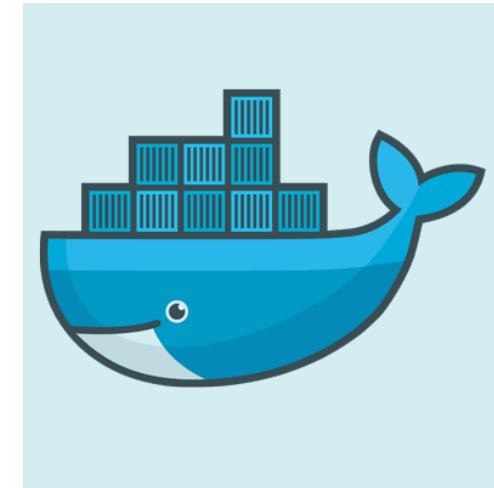
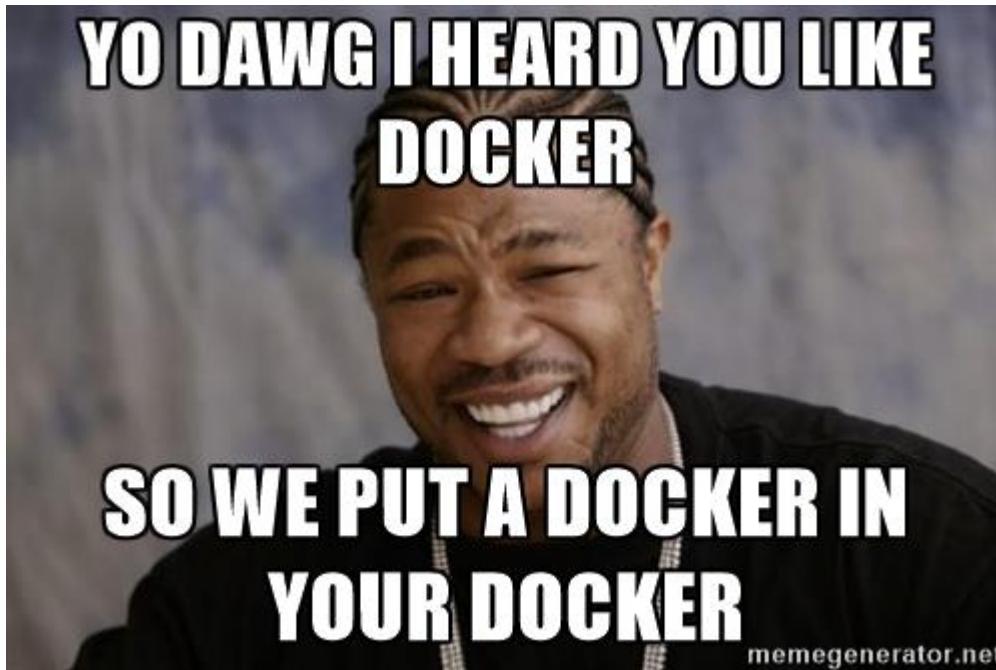
Qordoba Builder Image



@texasmichelle



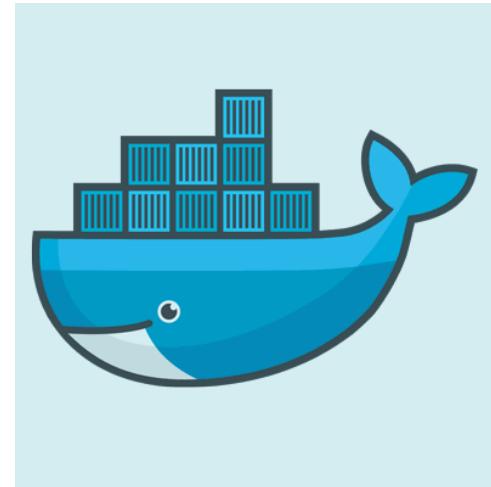
Qordoba Builder Image



@texasmichelle



Qordoba Builder Image

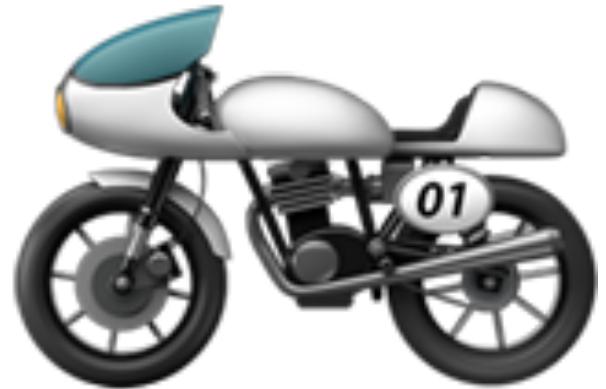


@texasmichelle



Affect Detection

- More sophisticated featurization
- Multi-label instead of mutual exclusivity
- Confidence scores
- Additional algorithms
- More evaluation metrics
- Automation
- ~~Classification~~ Graph?

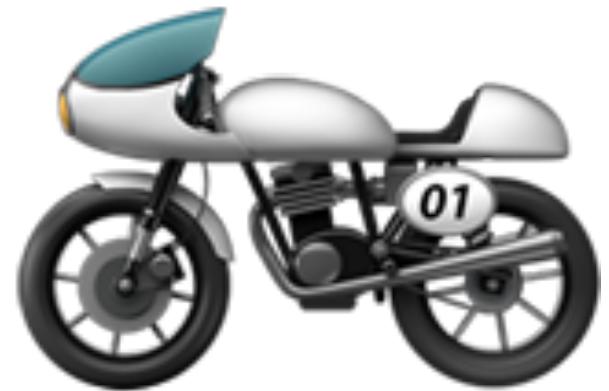


@texasmichelle



Affect Detection

- Expanded training set
 - Language
 - Source
 - Filter
 - Video Intelligence
 - Detect Faces
- Unsupervised models
- ~~All the domains~~ Generalization

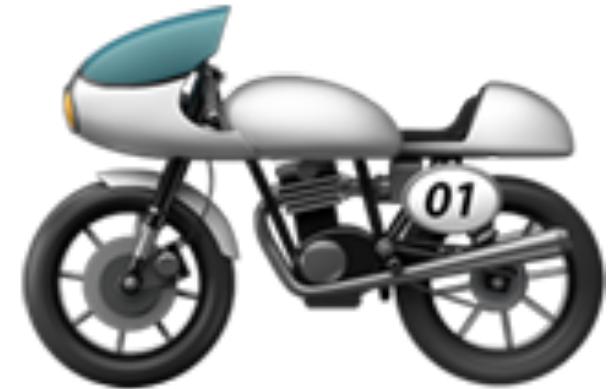


@texasmichelle



Affect Detection

- Lighter images
- Lighter deployment
- Weaveworks 



@texasmichelle



TL;DR

- There's a better way to ~~take over the world~~ do globalization
- How to operationalize & scale NLP models
- What do you want to do tonight?
 - Affect detection!

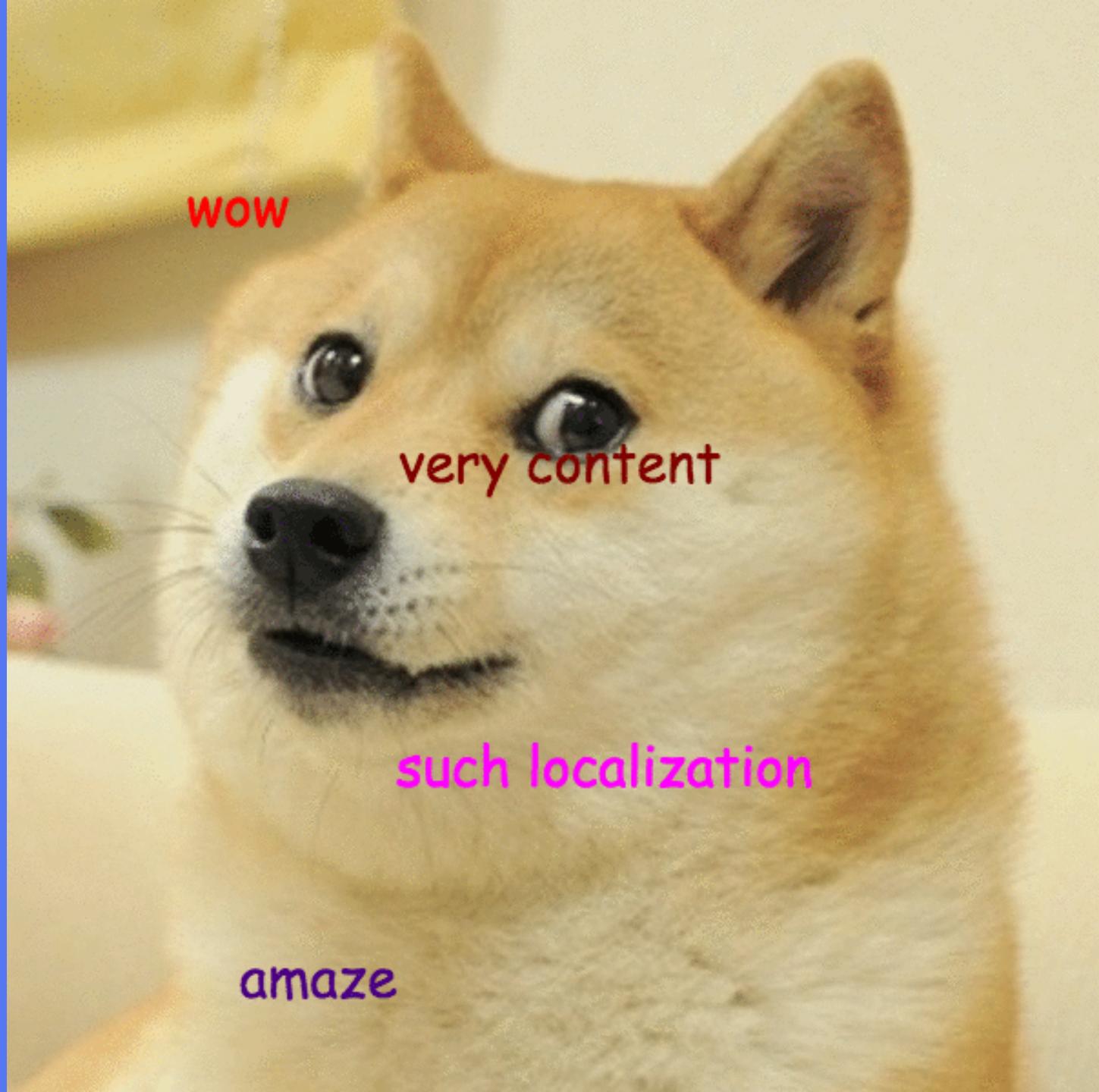


@texasmichelle



Q&A

michelle@qordoba.com
 @texasmichelle





Please

**Remember to
rate this session**

Thank you!



follow us @gotochgo