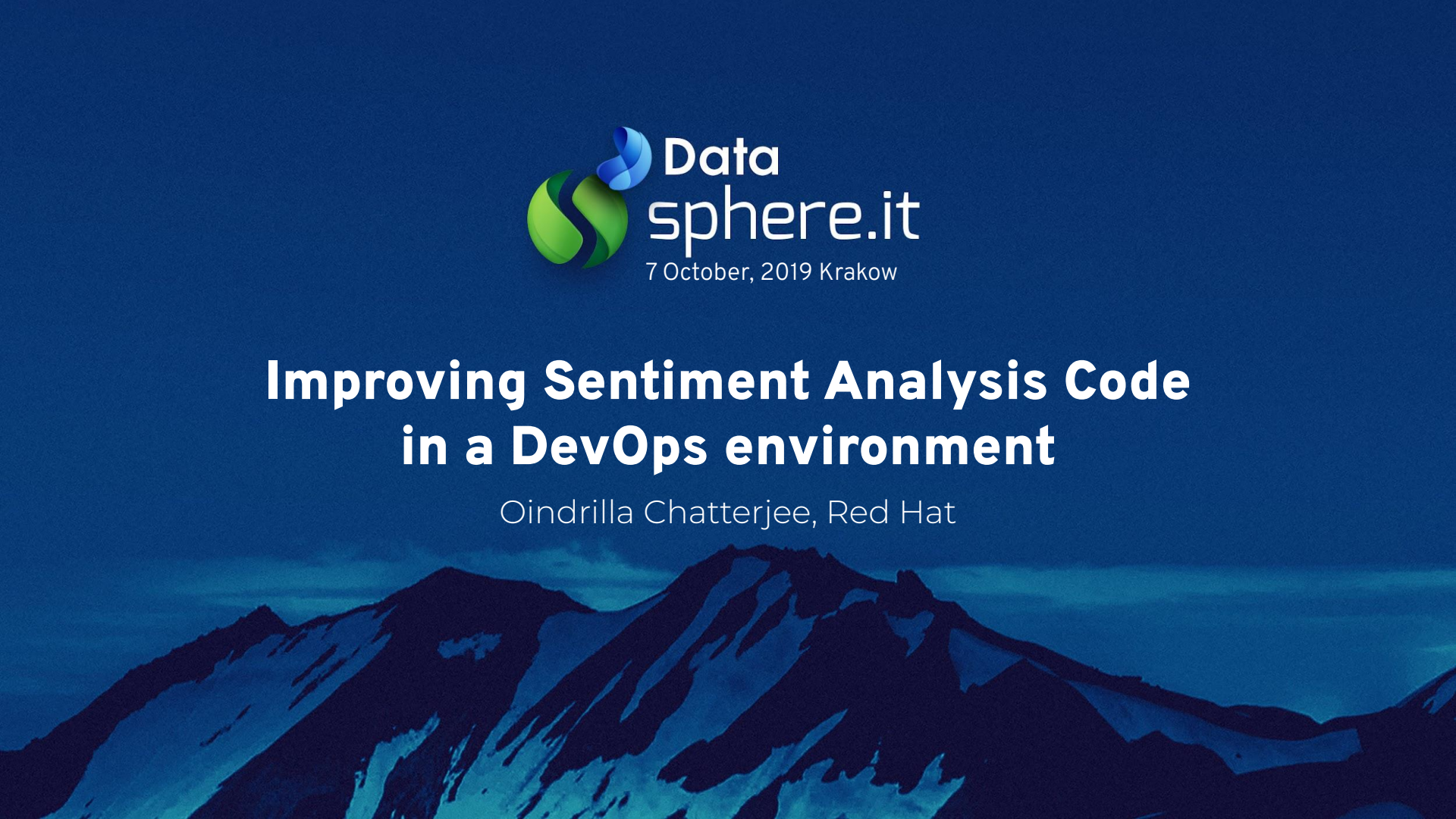
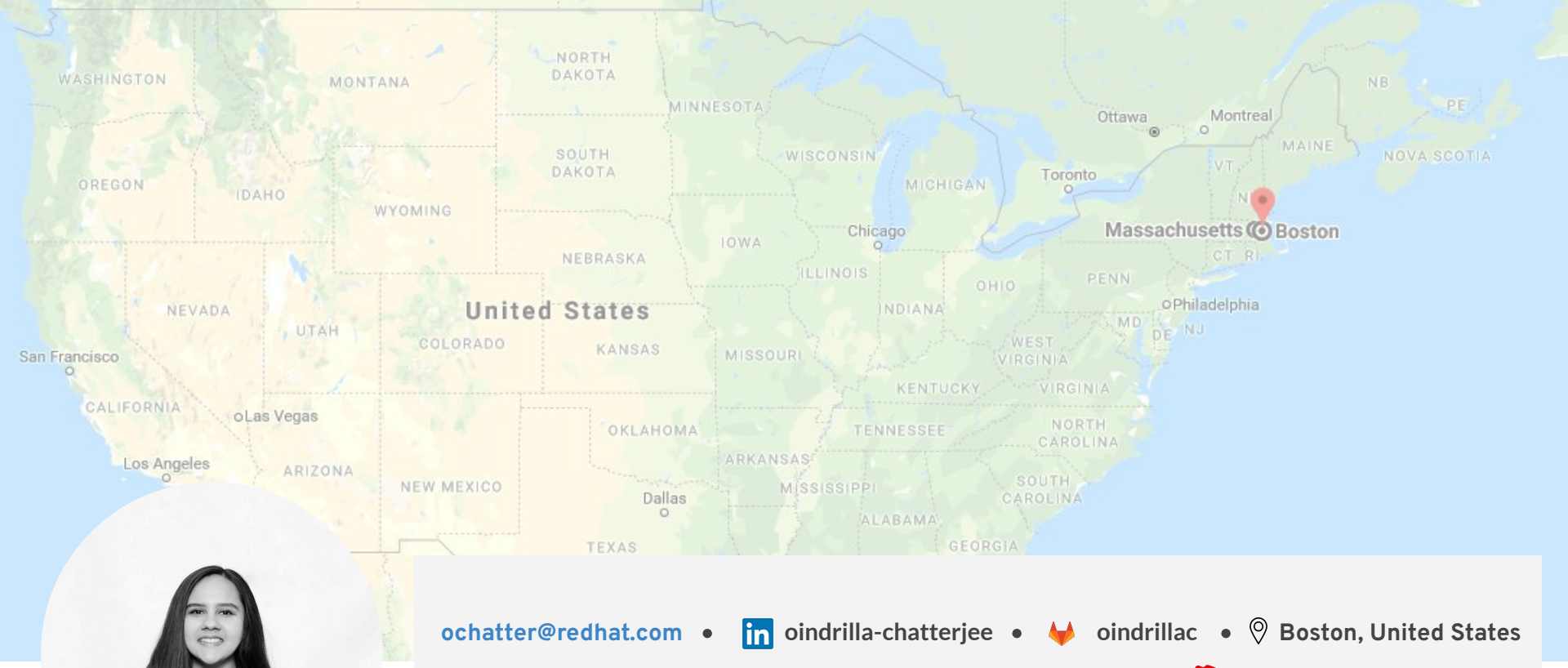


Improving Sentiment Analysis Code in a DevOps environment

Oindrilla Chatterjee, Red Hat



ochatter@redhat.com • oindrilla-chatterjee • oindrillac • Boston, United States

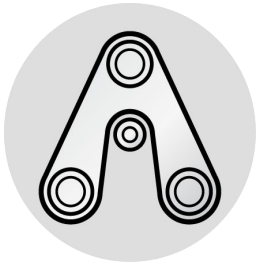
Data Scientist at **Office of the CTO | AI Center of Excellence |** Red Hat

What will we learn today?

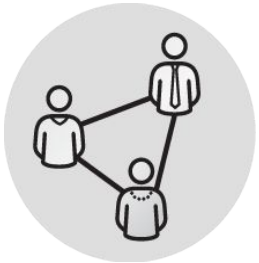
- How we see AI at Red Hat
- How we narrowed down to a single service
- Lessons learnt going through the process
- Usage of the framework

How Red Hat Sees AI

Represents a workload requirement for our **platforms** across the hybrid cloud.



Applicable to Red Hat's existing core business in order to increase **Open Source** development and production **efficiency**.



Valuable to our customers as specific services and product capabilities, providing an **Intelligent Platform** experience.

Enable customers to build **Intelligent Apps** using Red Hat products as well as our broader partner ecosystem.

thoth-station.ninja

AI Ops https://bit.ly/2n5vFps

opendatahub.io

Data as the Foundation

Red Hat

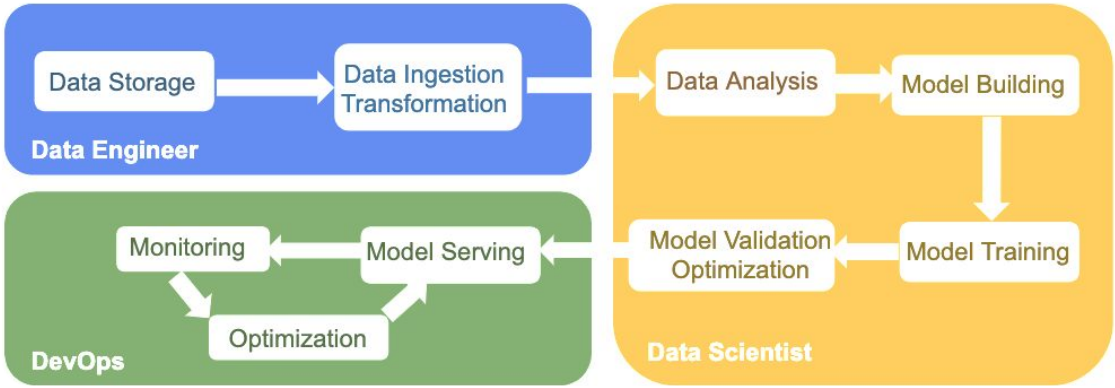
The Open Data Hub Project

Collaborate on a Data & AI platform for the Hybrid Cloud
https://opendatahub.io/

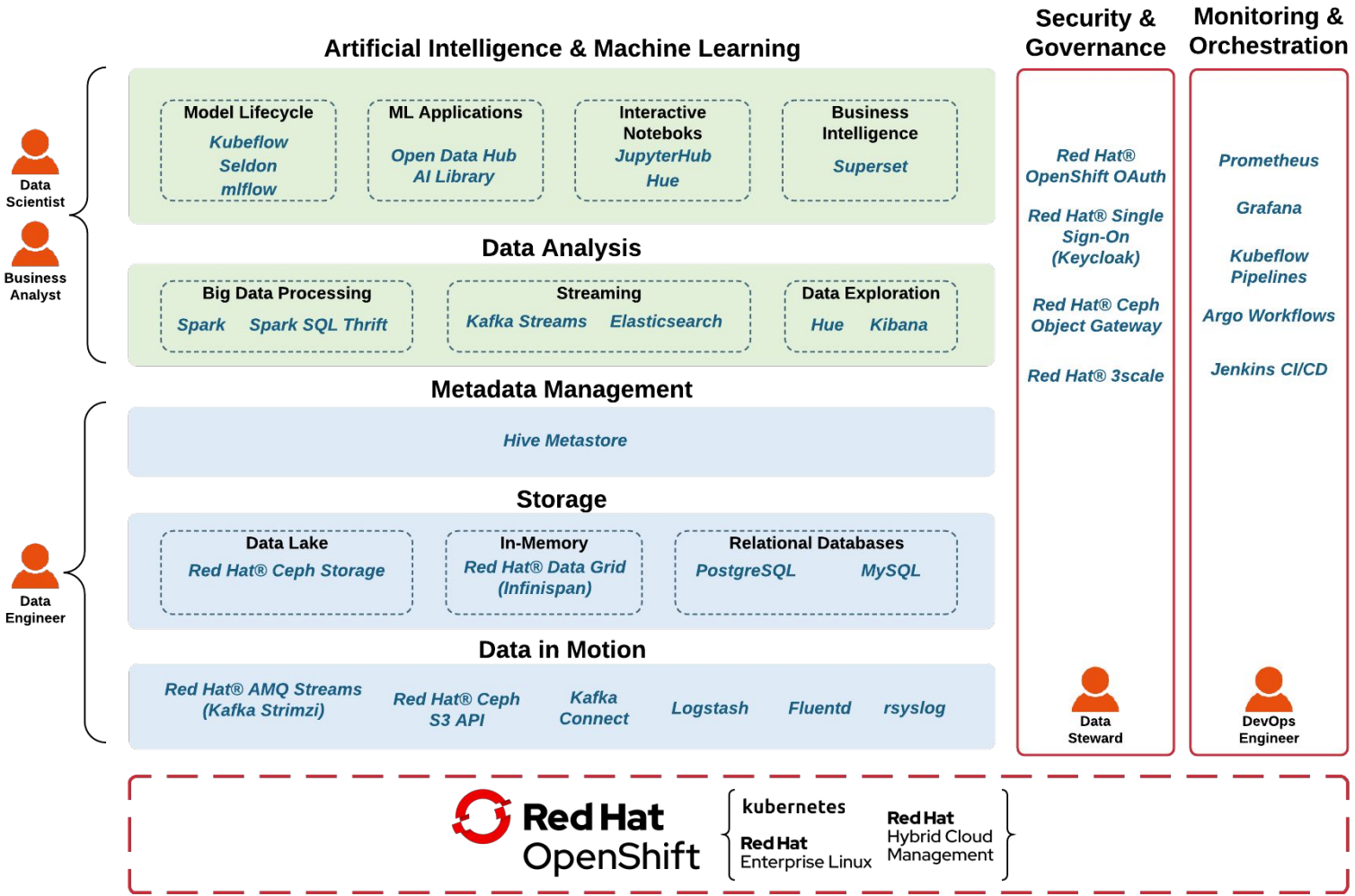
- Meta-operator that integrates open source AI/ML/Data projects
- Blueprint architecture for AI/ML on OpenShift
- Red Hat's internal Data Science and AI platform
- Open Data Hub Architecture: https://opendatahub.io/docs/architecture.html

Red Hat

OPEN DATA HUB
AI Platform powered by Open Source



Red Hat


Artificial Intelligence & Machine Learning
Model Lifecycle
Kubeflow
Seldon
mlflow
ML Applications
Open Data Hub
AI Library
Interactive Noteboks
JupyterHub
Hue
Business Intelligence
Superset
Data Scientist
Business Analyst
Data Analysis
Big Data Processing
Spark
Spark SQL Thrift
Streaming
Kafka Streams
Elasticsearch
Data Exploration
Hue
Kibana
Metadata Management
Hive Metastore
Storage
Data Lake
Red Hat® Ceph Storage
In-Memory
Red Hat® Data Grid (Infinispan)
Relational Databases
PostgreSQL
MySQL
Data Engineer
Data in Motion
Red Hat® AMQ Streams (Kafka Strimzi)
Red Hat® Ceph S3 API
Kafka Connect
Logstash
Fluentd
rsyslog
Security & Governance
Red Hat® OpenShift OAuth
Red Hat® Single Sign-On (Keycloak)
Red Hat® Ceph Object Gateway
Red Hat® 3scale
Data Steward
Monitoring & Orchestration
Prometheus
Grafana
Kubeflow Pipelines
Argo Workflows
Jenkins CI/CD
DevOps Engineer
Red Hat OpenShift
kubernetes
Red Hat Enterprise Linux
Red Hat Hybrid Cloud Management
Red Hat

AI Library

https://gitlab.com/opendatahub/ai-library

- An Open source collection of AI Components
  - Machine learning algorithms
  - Machine learning solutions to common use cases
- Allows rapid prototyping of ideas

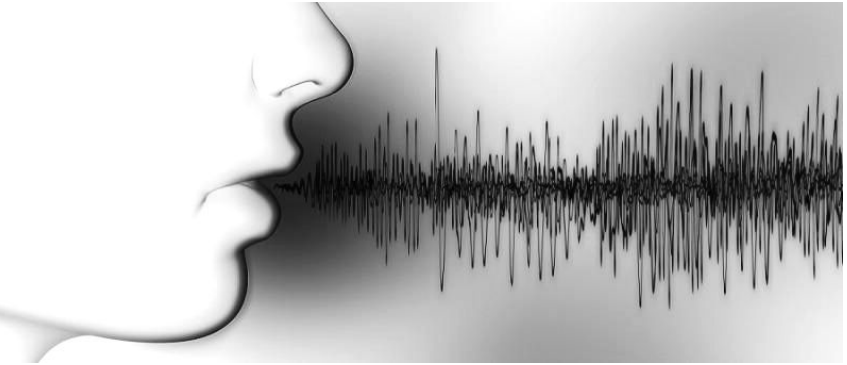
Red Hat

Sentiment Analysis

Red Hat

NLP?

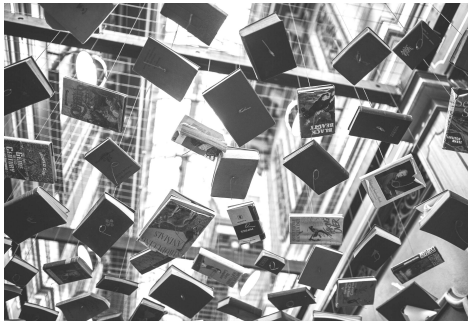
How does it work?



Natural language processing helps computers understand, interpret and manipulate human language.

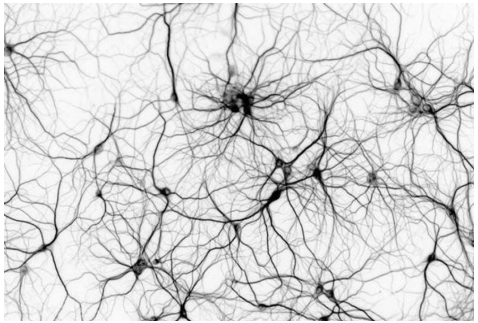
Red Hat

Ways of performing sentiment analysis



Lexicon based approach

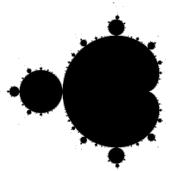
Machine Learning Classifier



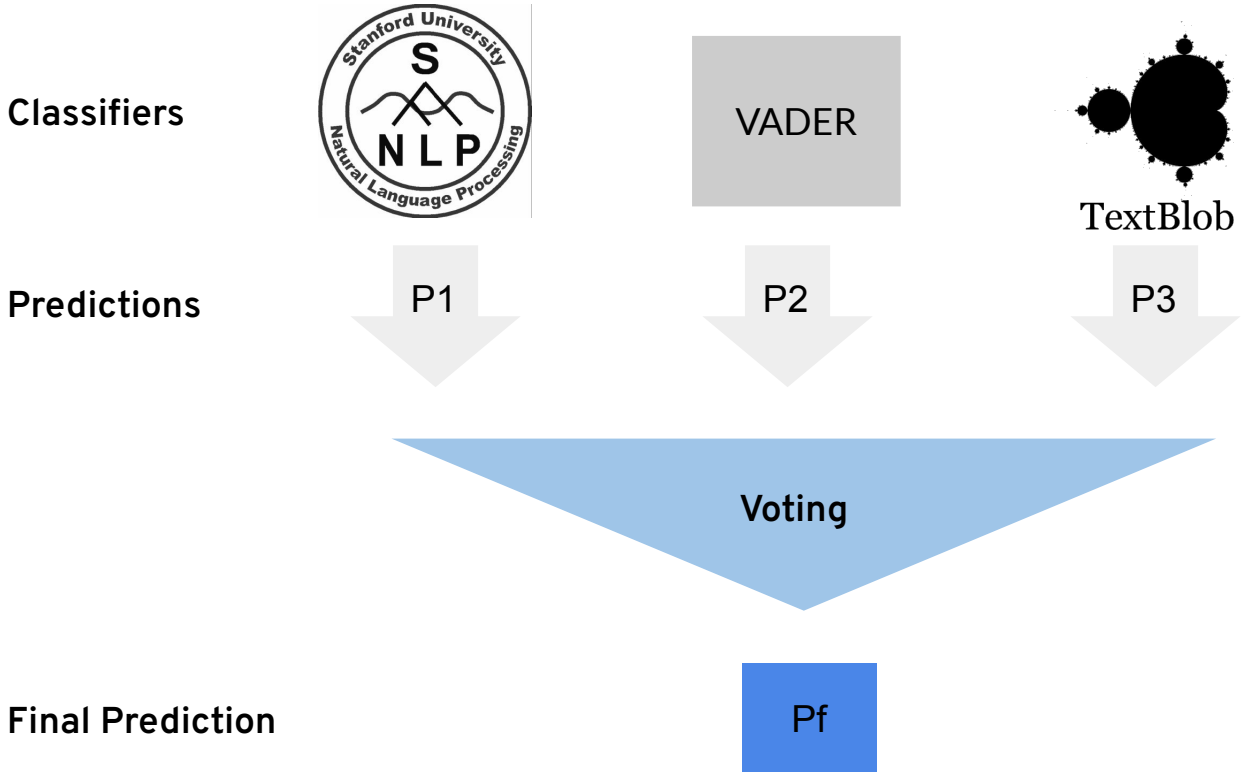
Deep Learning Model

Composition of our initial service

VADER



TextBlob


Classifiers
Stanford University
S
N L P
Natural Language Processing
VADER
TextBlob
Predictions
P1
P2
P3
Voting
Final Prediction
Pf

Red Hat

Accuracy of sentence classification

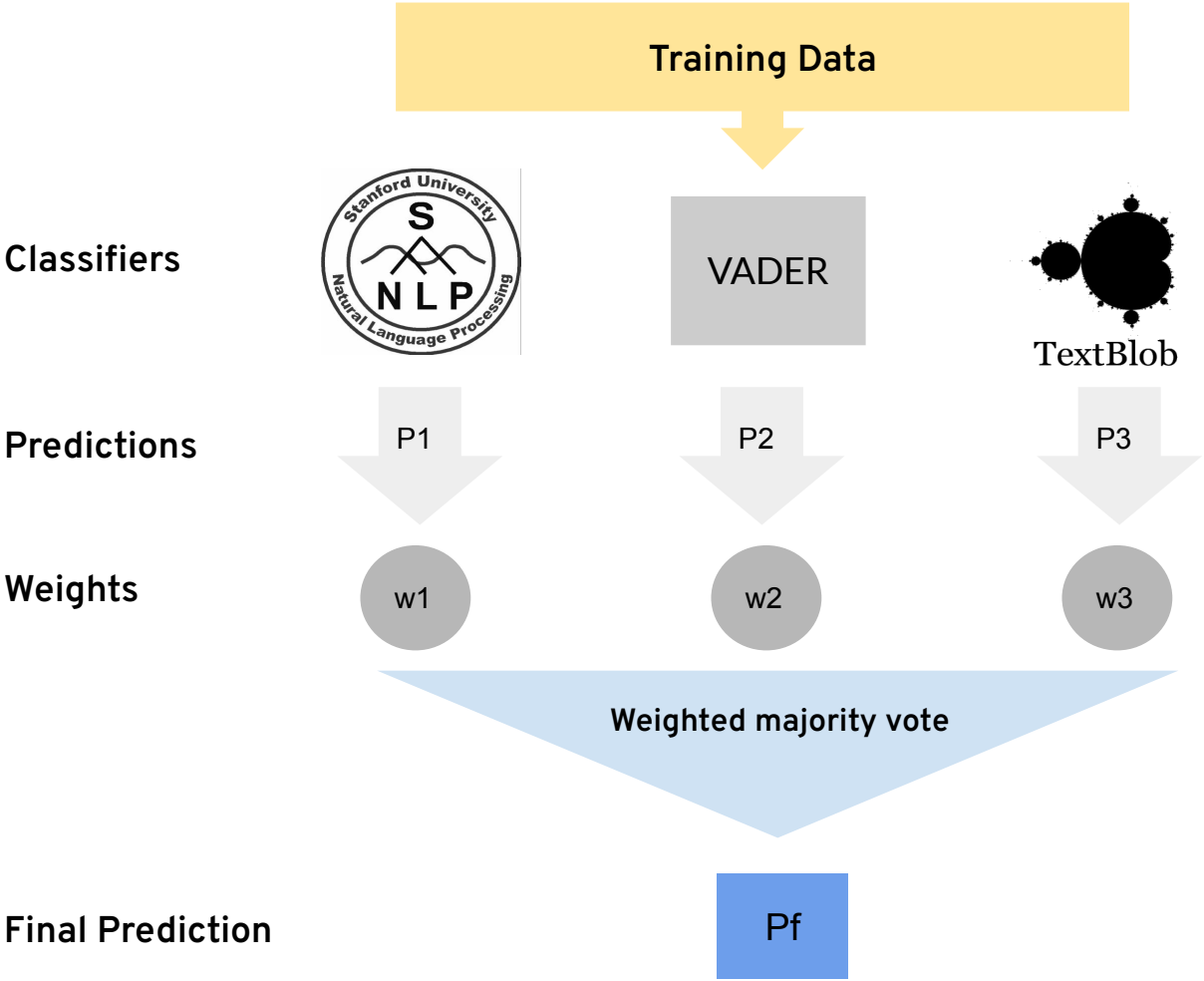
Eg. Total Negative Sentences in Dataset : 159 | **Misclassified** : 136

Foreign languages were being misinterpreted as English

Eg. En realidad no me ayudaron , el SLA de 1hora nunca se cumplió y lo solucioné por mi cuenta en este caso - tagged as **Positive**
Means Negative

Domain specific sentences mis-classified

Quick responses from the analyst assigned to the case. - marked as **Negative**.
Should be **Positive** in context of customer reviews.
Words like consistent, immediate, in-depth, skilled, polite etc indicate positive sentiment in the context of customer service.

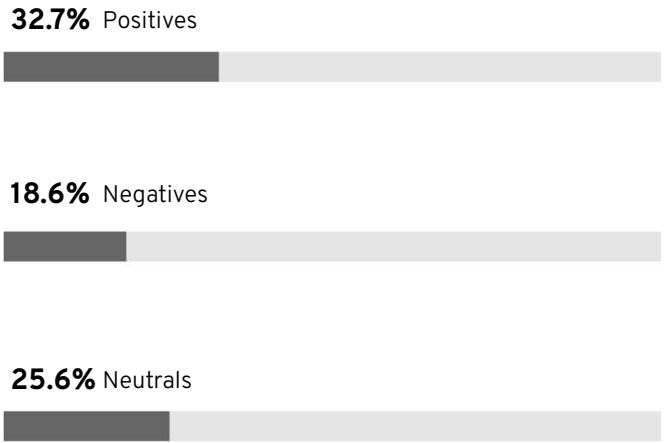

Training Data
Classifiers
Stanford University
S
N L P
Natural Language Processing
VADER
TextBlob
Predictions
P1
P2
P3
Weights
w1
w2
w3
Weighted majority vote
Final Prediction
Pf

Red Hat

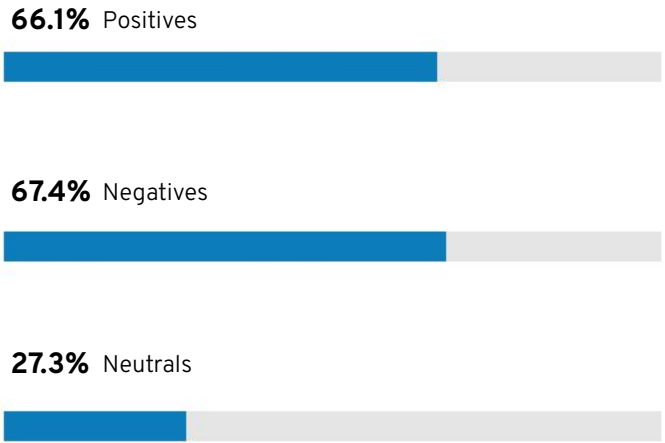
VADER

Improvements in performance

01 F1 scores of voting ensemble service

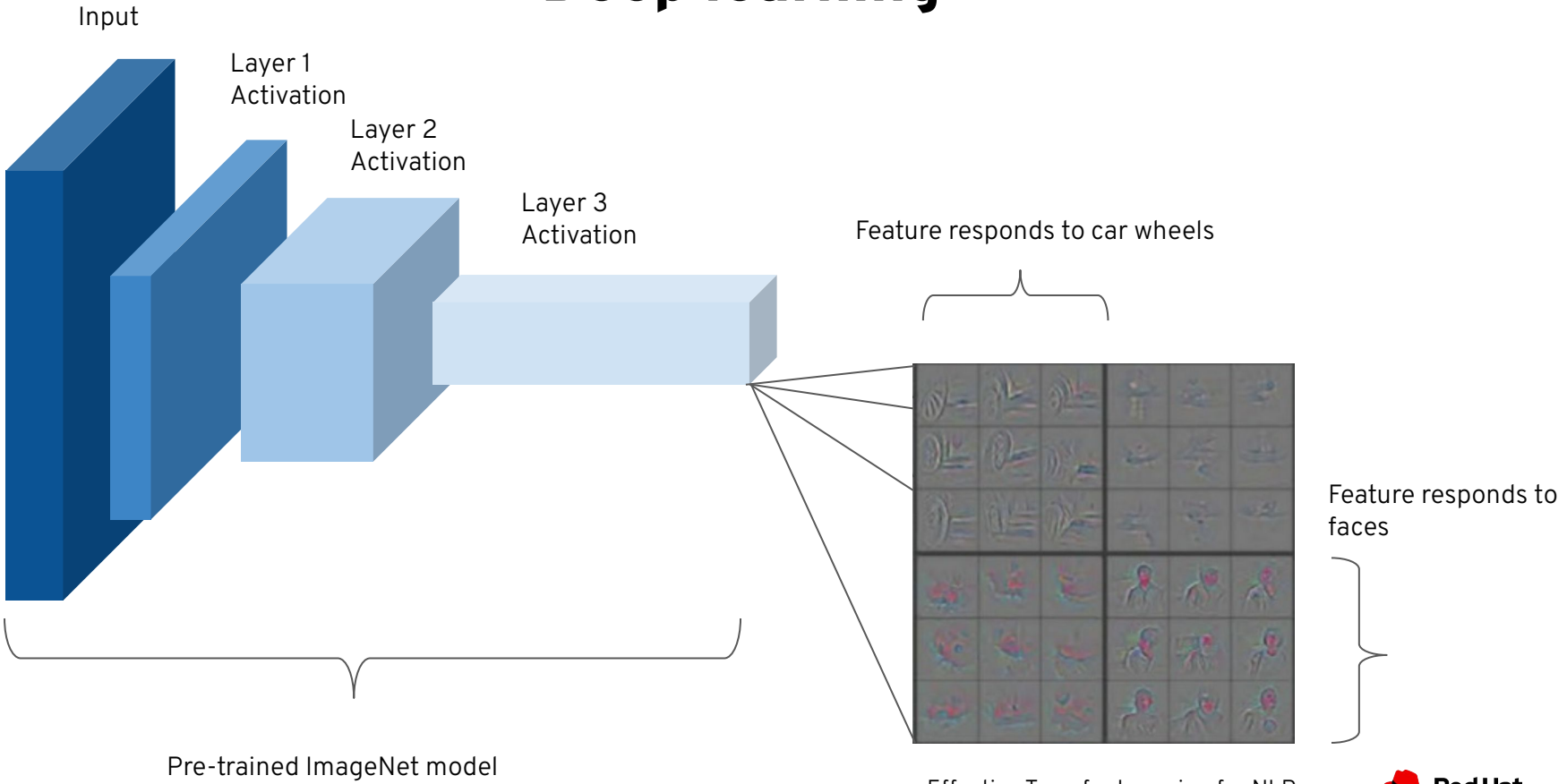


02 F1 scores of wtd average service



Red Hat

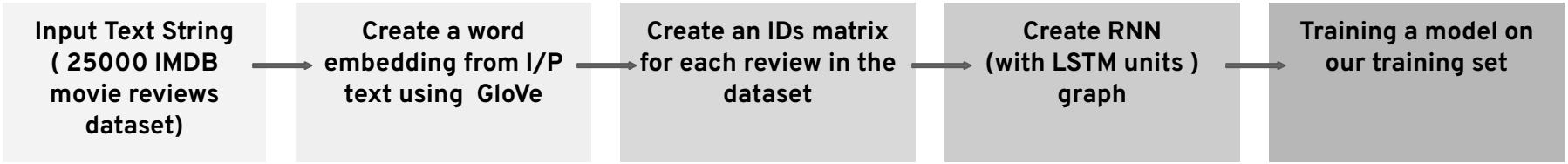
Deep learning



Effective Transfer Learning for NLP
- Madison May, 2018

Red Hat

Recurrent Neural Networks(RNNs) with Long Short Term Memory(LSTM) Units



01 F1 scores of Initial Service

Initial Model

42% Positives

29% Negatives

02 F1 scores of LSTM based model

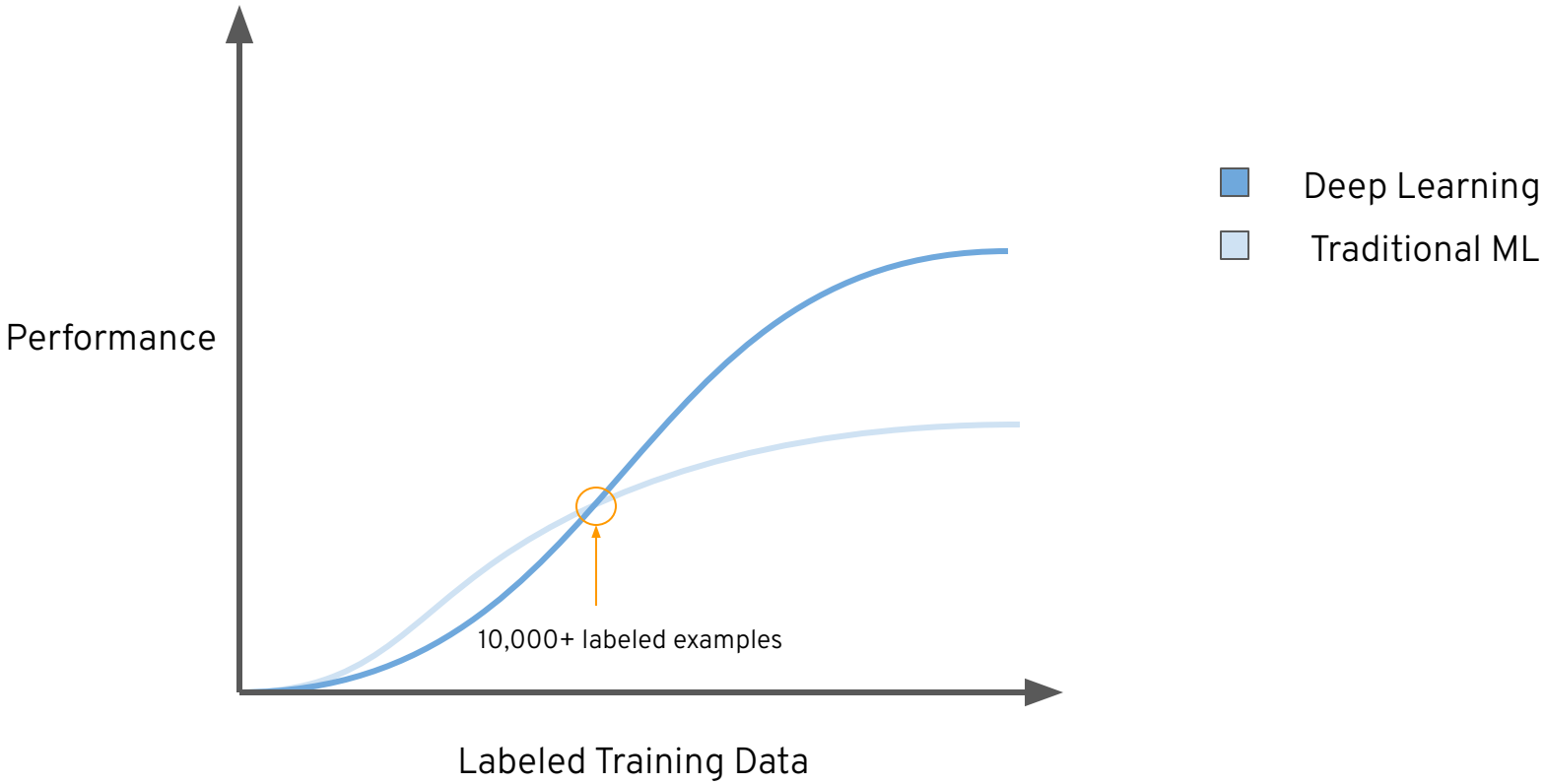
Deep Learning LSTM Model

69% Positives

67% Negatives

Red Hat

Training data requirements

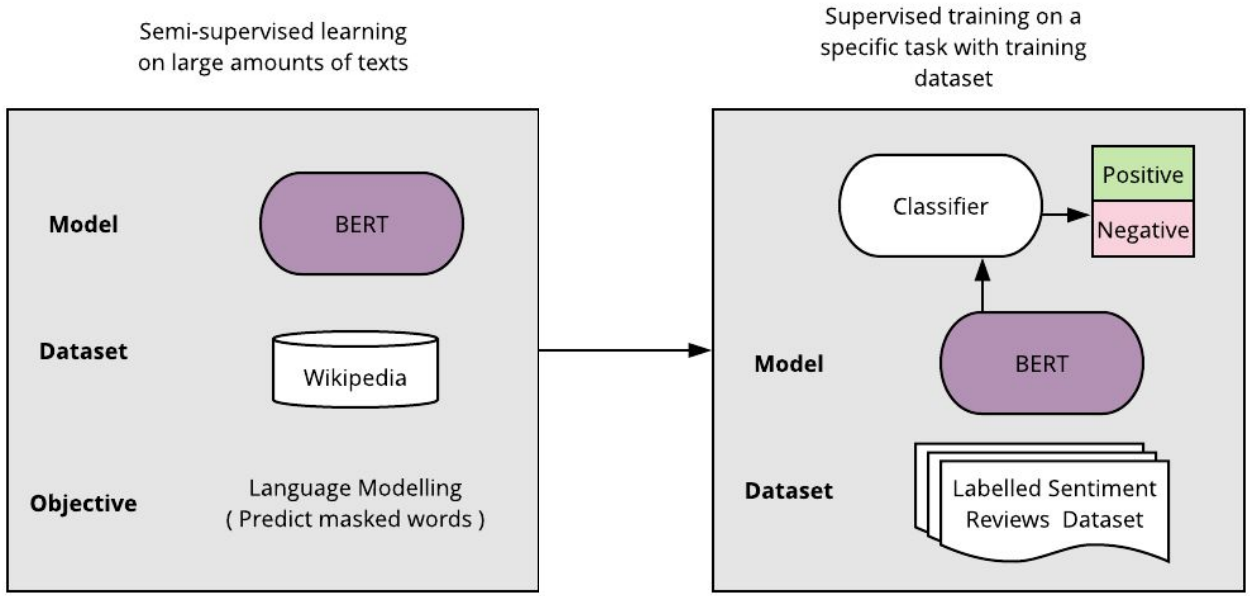


Effective Transfer Learning for NLP
- Madison May, 2018

Transfer Learning ? How does it apply in NLP?

- Pre-training allows a model to capture and learn a variety of linguistic phenomena
- Address difficult challenges in ML research, availability of data and resources

Bidirectional Encoder Representation from Transformers (BERT)



Red Hat

What makes BERT different?

Contextual representation

Contextual models instead generate a representation for each word in a sentence based on context

Deeply bidirectional unsupervised language representation

Technique of masking out some words in the input and then condition each word bidirectionally to predict the masked words

Models relationships between sentences

Is the second sentence the actual next sentence that comes after the first or just a random sentence?

Transformer Architecture

Compared to RNNs which is very sequential, Transformer architecture is able to take full advantage of GPUs and TPUs and make training must faster

Comparative Results

01 F1 scores of Initial Service

Initial Model

42% Positives

29% Negatives

02 F1 scores of LSTM based model

Deep Learning LSTM Model

69% Positives

67% Negatives

03 F1 scores of BERT based model

Deep Learning BERT Model

90% Positives

87% Negatives

Red Hat

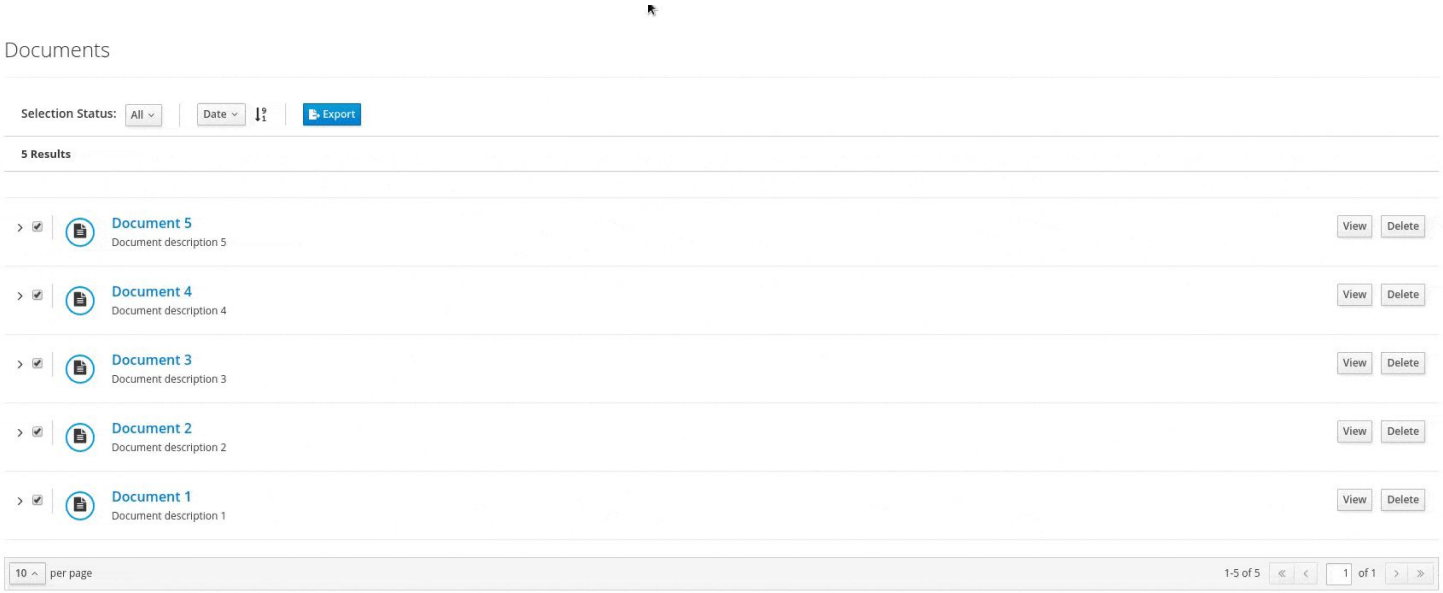
Why is this important?

1. **Better** classification or sentiment annotation

2. Annotating a single sentiment for **entire review/piece of text/paragraph** as opposed to initial service

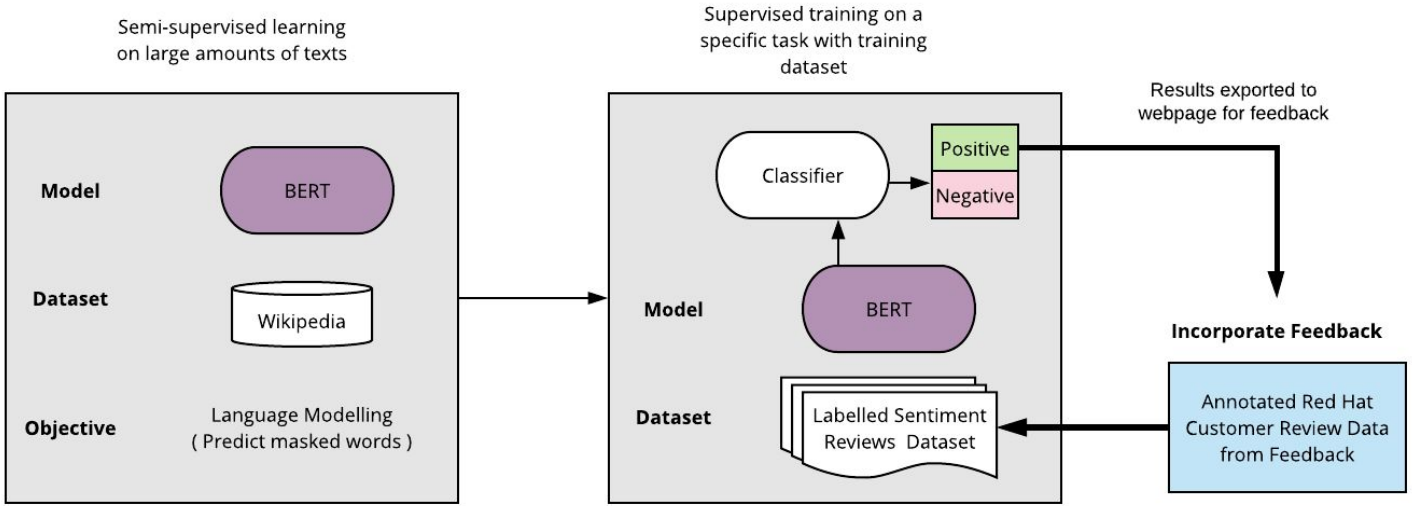
3. Models take much lesser time to train using **GPUs**

Continual improvement of the system



Red Hat

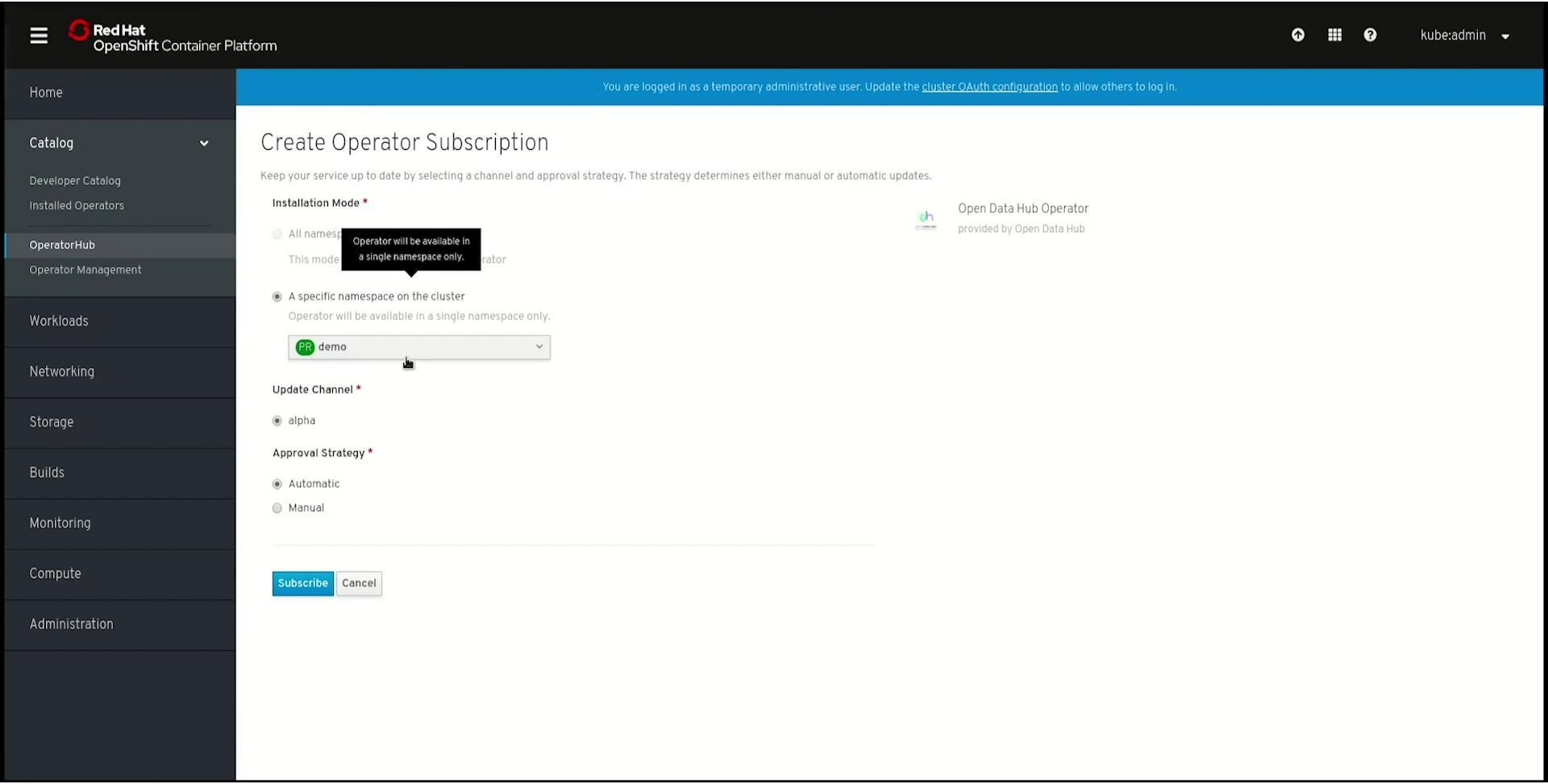
Introducing Feedback

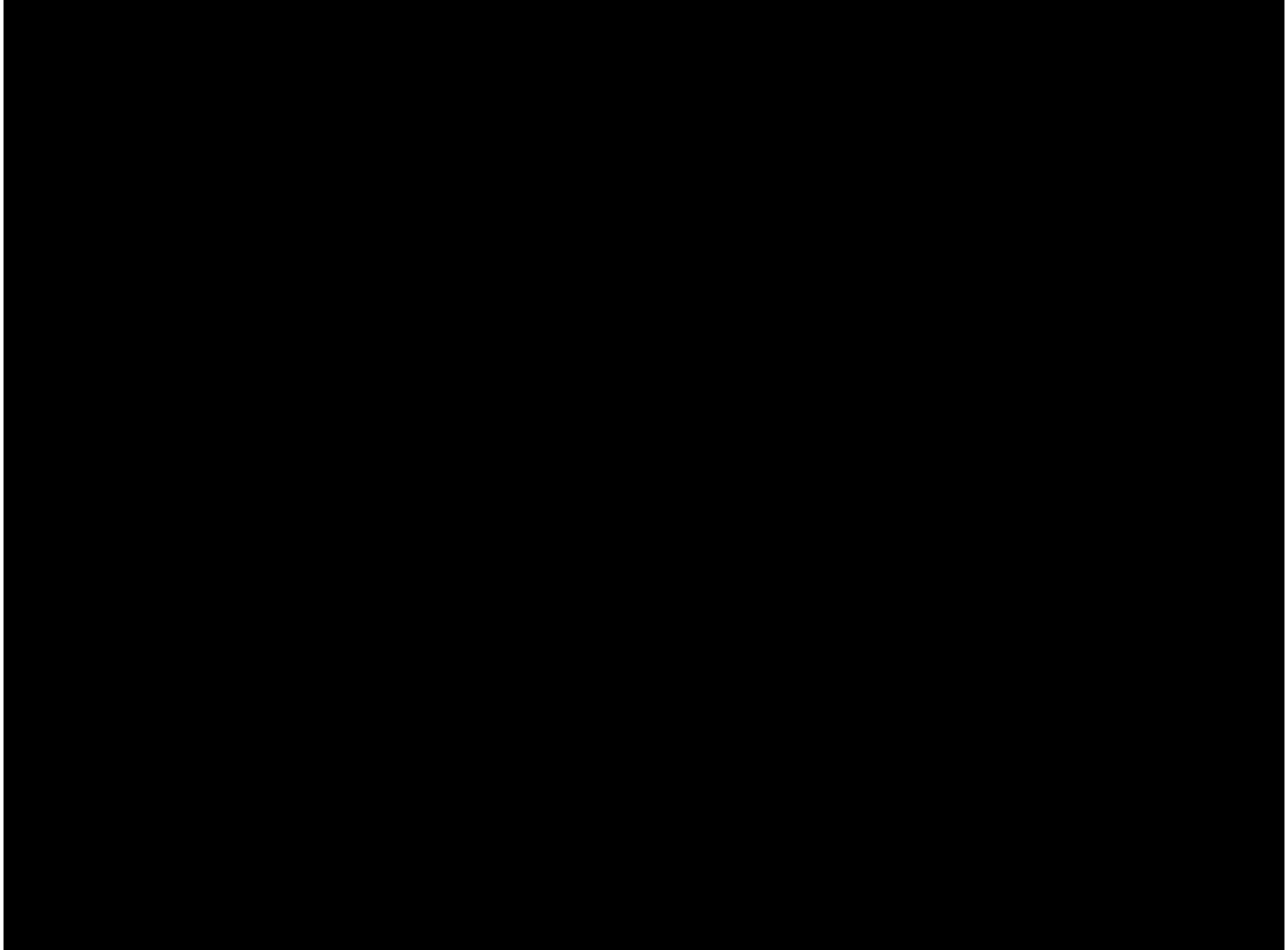


Red Hat

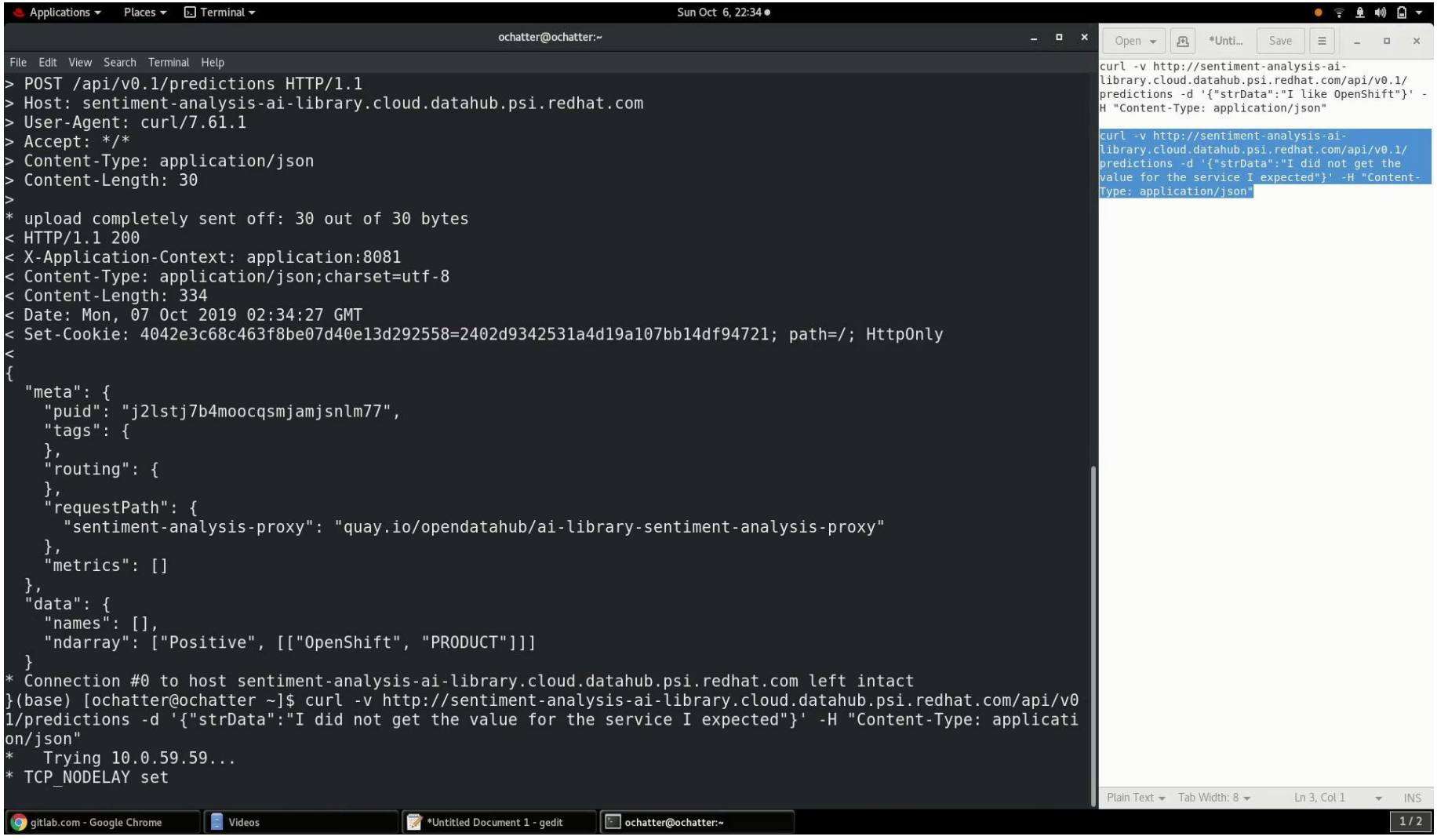
Lessons Learnt

- Evaluation of versions of the model -- we are not limited to F1 Scores
- Lack of variability when developing context-specific models


Red Hat
OpenShift Container Platform
kube:admin
Home
Catalog
Developer Catalog
Installed Operators
OperatorHub
Operator Management
Workloads
Networking
Storage
Builds
Monitoring
Compute
Administration
You are logged in as a temporary administrative user. Update the cluster OAuth configuration to allow others to log in.
Create Operator Subscription
Keep your service up to date by selecting a channel and approval strategy. The strategy determines either manual or automatic updates.
Installation Mode *
All namesp
This mode
rator
Operator will be available in a single namespace only.
A specific namespace on the cluster
Operator will be available in a single namespace only.
PR demo
Update Channel *
alpha
Approval Strategy *
Automatic
Manual
Subscribe
Cancel
Open Data Hub Operator
provided by Open Data Hub



Red Hat


Applications
Places
Terminal
Sun Oct 6, 22:34
ochatter@ochatter:~
File Edit View Search Terminal Help
> POST /api/v0.1/predictions HTTP/1.1
> Host: sentiment-analysis-ai-library.cloud.datahub.psi.redhat.com
> User-Agent: curl/7.61.1
> Accept: */*
> Content-Type: application/json
> Content-Length: 30
>
* upload completely sent off: 30 out of 30 bytes
< HTTP/1.1 200
< X-Application-Context: application:8081
< Content-Type: application/json;charset=utf-8
< Content-Length: 334
< Date: Mon, 07 Oct 2019 02:34:27 GMT
< Set-Cookie: 4042e3c68c463f8be07d40e13d292558=2402d9342531a4d19a107bb14df94721; path=/; HttpOnly
<
{
"meta": {
"puid": "j2lstj7b4moocqsmjamjsnlm77",
"tags": {
},
"routing": {
},
"requestPath": {
"sentiment-analysis-proxy": "quay.io/opendatahub/ai-library-sentiment-analysis-proxy"
},
"metrics": []
},
"data": {
"names": [],
"ndarray": ["Positive", [["OpenShift", "PRODUCT"]]]
}
* Connection #0 to host sentiment-analysis-ai-library.cloud.datahub.psi.redhat.com left intact
}(base) [ochatter@ochatter ~]$ curl -v http://sentiment-analysis-ai-library.cloud.datahub.psi.redhat.com/api/v0
1/predictions -d '{"strData":"I did not get the value for the service I expected"}' -H "Content-Type: applicati
on/json"
*   Trying 10.0.59.59...
* TCP_NODELAY set
Open
*Unti...
Save
curl -v http://sentiment-analysis-ai-library.cloud.datahub.psi.redhat.com/api/v0.1/predictions -d '{"strData":"I like OpenShift"}' -H "Content-Type: application/json"
curl -v http://sentiment-analysis-ai-library.cloud.datahub.psi.redhat.com/api/v0.1/predictions -d '{"strData":"I did not get the value for the service I expected"}' -H "Content-Type: application/json"
Plain Text
Tab Width: 8
Ln 3, Col 1
INS
gitlab.com - Google Chrome
Videos
*Untitled Document 1 - gedit
ochatter@ochatter:~
1 / 2

Thank you!

ochatter@redhat.com

OPEN DATA HUB
AI Platform powered by Open Source

Red Hat

AI Ops https://bit.ly/2n5vFps

thoth-station.ninja

opendatahub.io

https://gitlab.com/opendatahub/ai-library

Data sphere.it

Join the conversation #sphereIT #DataSphere