Sentiment Analysis Service in a DevOps environment

Oindrilla Chatterjee

Data Scientist

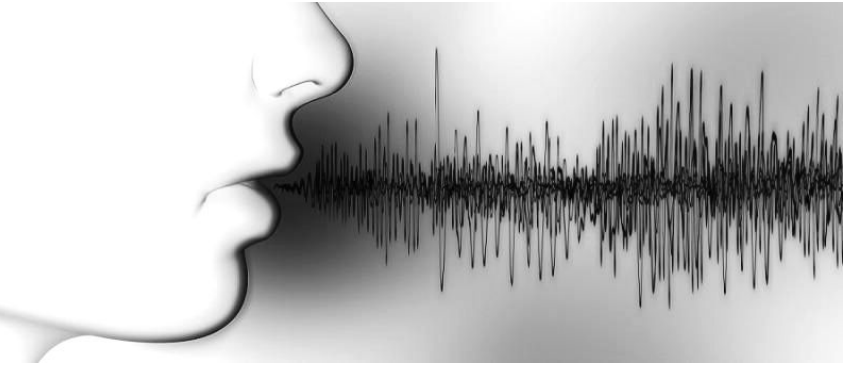
AI Center of Excellence

Red Hat

Sentiment Analysis

NLP?

How does it work?



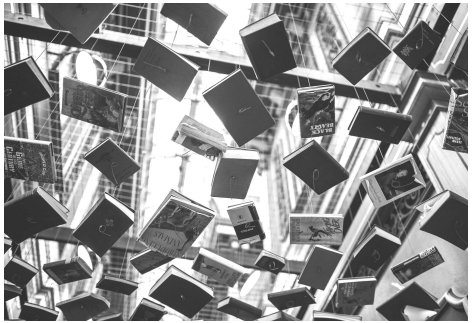
Natural language processing helps computers understand, interpret and manipulate human language.

Red Hat

What will we learn today?

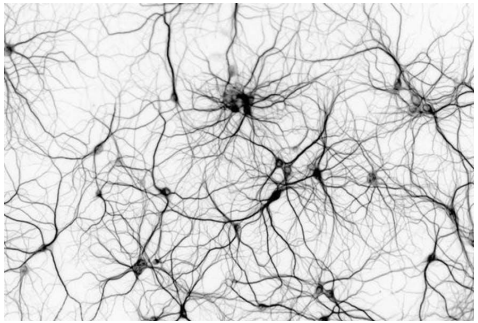
- How we narrowed down to a single service
- How we improved the service
- Lessons learnt going through the process
- Usage of the framework
- How to implement such a service

Ways of performing sentiment analysis



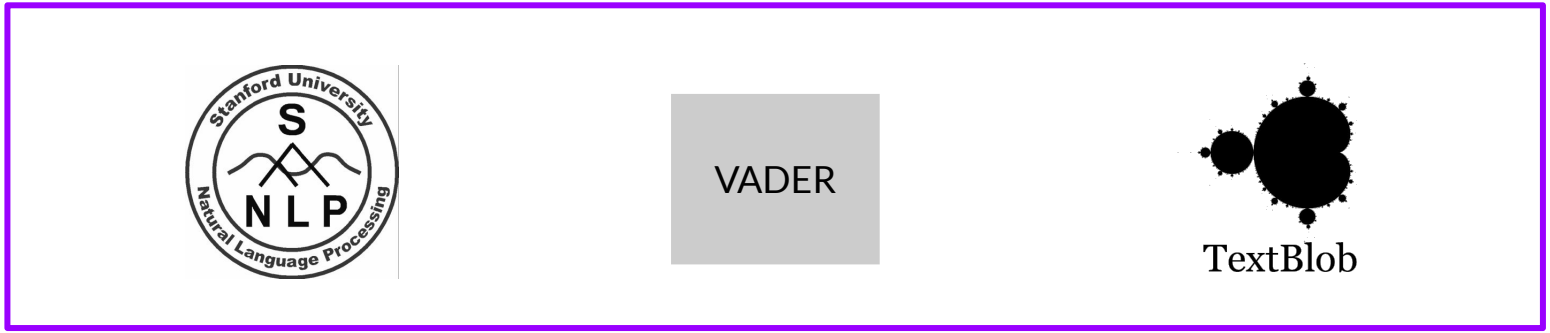
Lexicon based approach

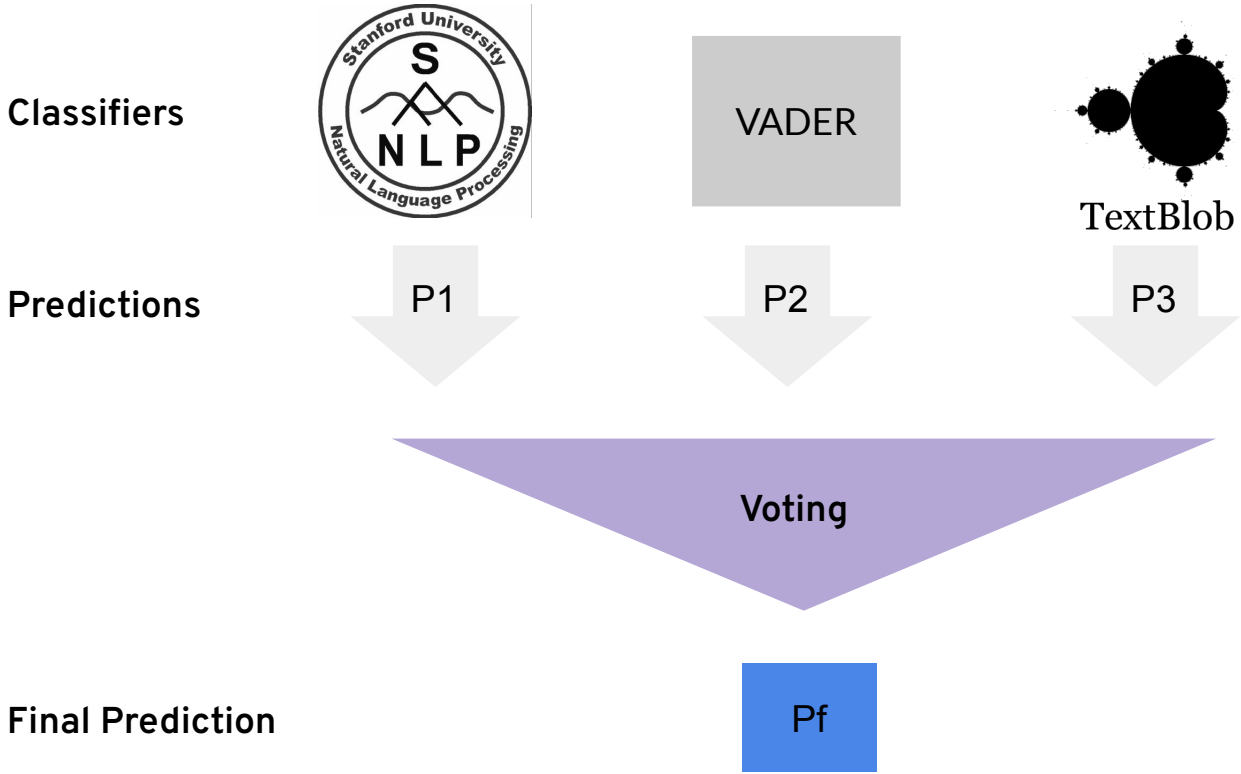
Machine Learning Classifier



Deep Learning Model

Composition of our initial service




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

Accuracy of sentence classification

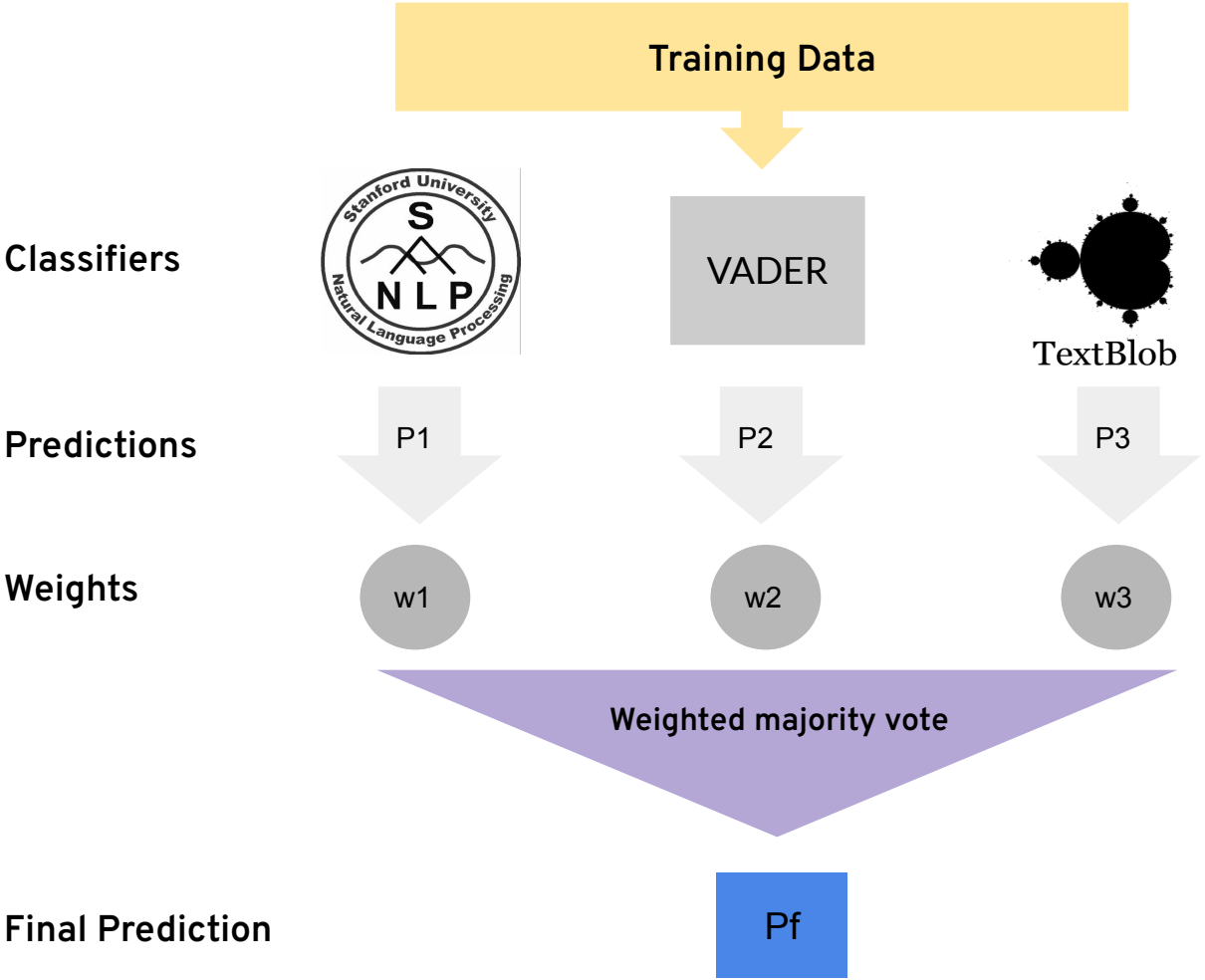
Eg. Total Negative Sentences in Dataset : 159 | **Misclassified** : 136

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.

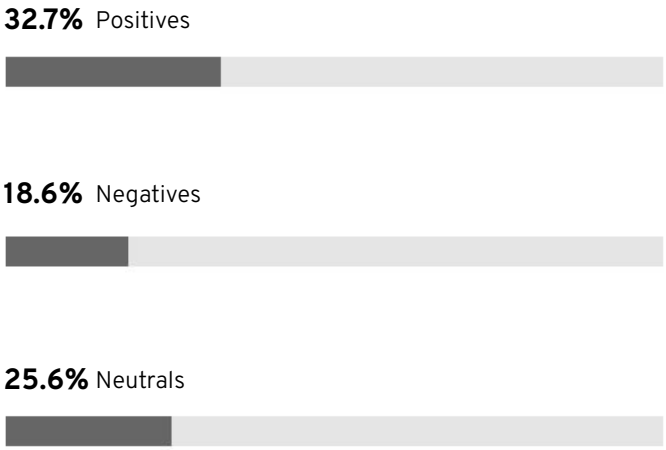

Training Data
Classifiers
Stanford University
S
N L P
Natural Language Processing
VADER
TextBlob
Predictions
P1
P2
P3
Weights
w1
w2
w3
Weighted majority vote
Final Prediction
Pf

Red Hat

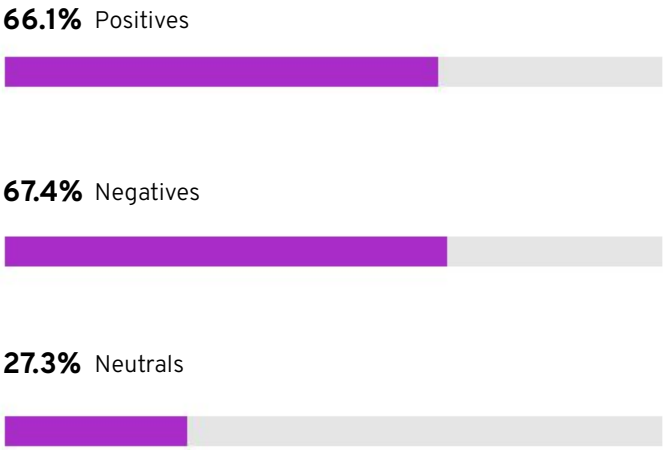
VADER

Improvements in performance

01 F1 scores of voting ensemble service



02 F1 scores of wtd average service



Red Hat

What is Entity Detection?

Entities in the context of NLP are terms/tokens which signify important information. They are mainly proper nouns in the sentence pertaining to places, people, products, organizations, events etc

Red Hat

Unable to identify context-specific terms

Eg. Ansible, RHEL7, RHEL 8, OpenShift, Red Hat Cloudforms etc - terms were not recognized by the NER system

Non-entities classified as entities

Quick updates can be put in via the support portal .
[{'name': 'Quick', 'category': 'ORG'}]

Mis-classify entities

Proactive support from the RHEL engineer :
[{'name': 'RHEL', 'category': 'ORG'}]

Steps Taken

Created Balanced training data to re-train SpaCy's NER model

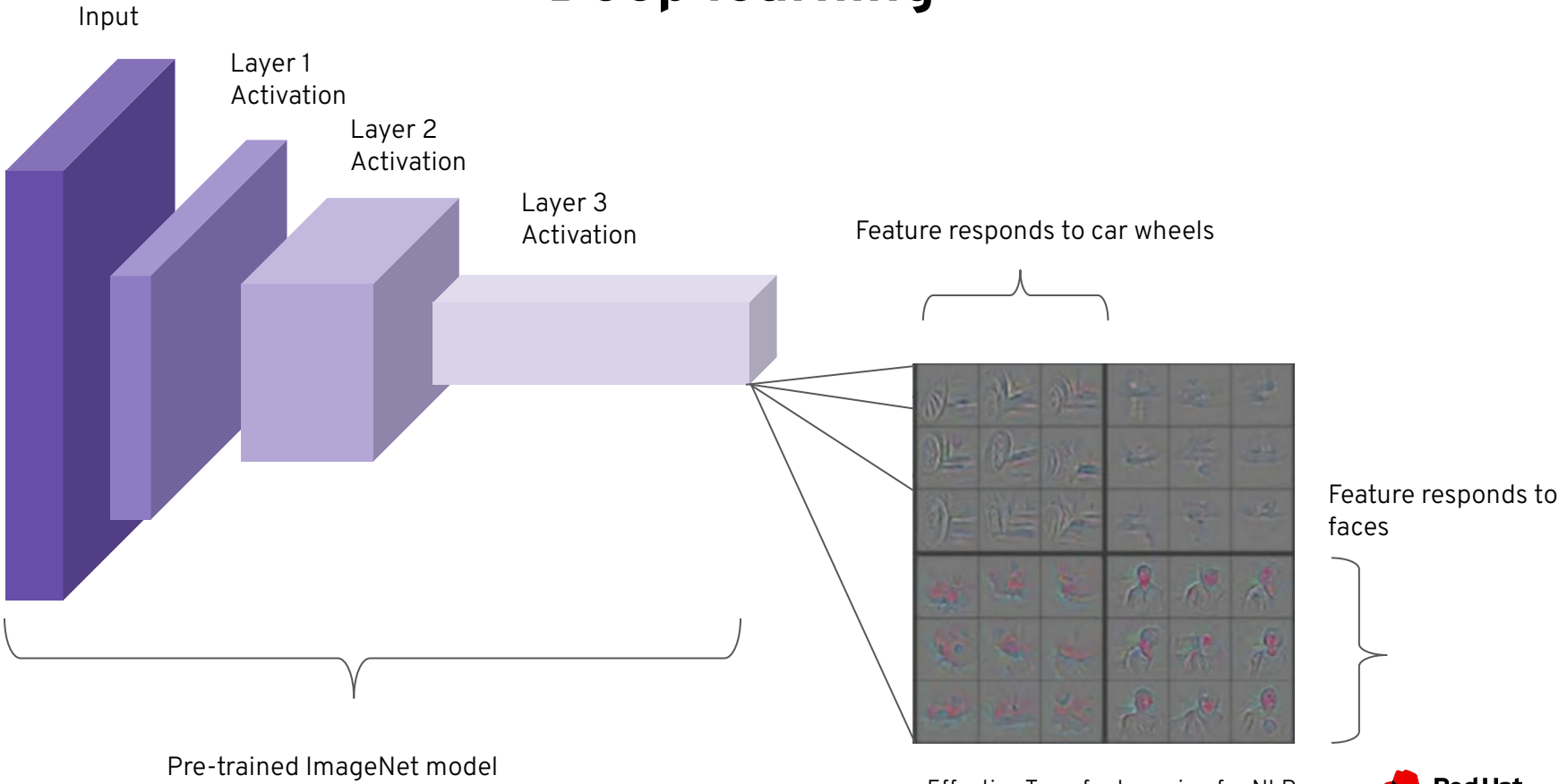
Examples of terms introduced:

Entities	True Classification
Red Hat	ORG
Redhat	ORG
OpenShift	PRODUCT
rhel	PRODUCT
RH	ORG
Ansible	PRODUCT
RHEL	PRODUCT
Red Hat Cloudforms	PRODUCT
RHEL 7	PRODUCT
RHEL7	PRODUCT
RHEL 6	PRODUCT
RHEL8	PRODUCT
RHEL 8	PRODUCT
RHEL6	PRODUCT

Entity results on a sentence with previous NER model vs updated NER model

- Higher number of context specific entities recognized
- Better classification of entities
- Lower non-entity detection

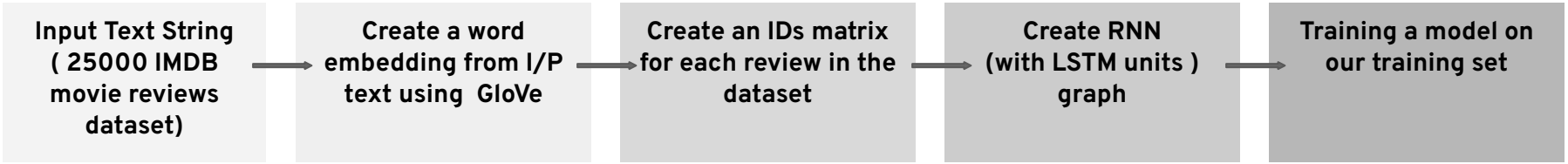
Deep learning



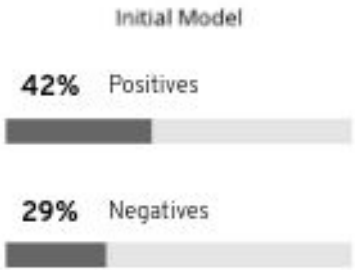
Effective Transfer Learning for NLP
- Madison May, 2018

Red Hat

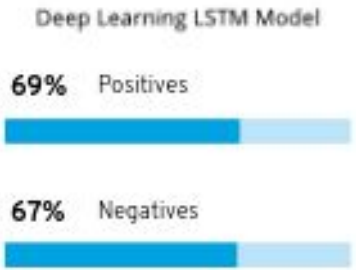
Recurrent Neural Networks(RNNs) with Long Short Term Memory(LSTM) Units



01 F1 scores of Initial Service

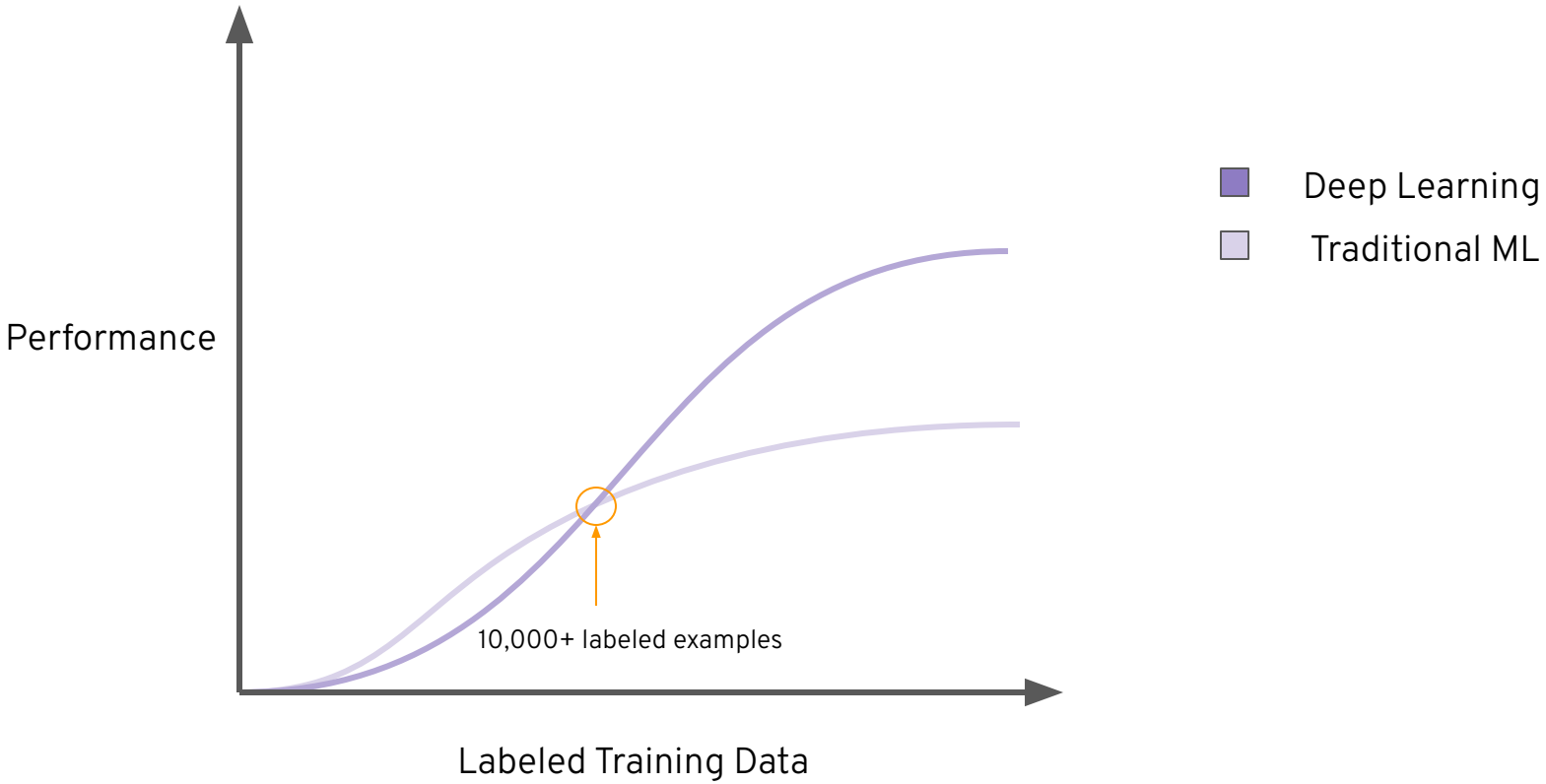


02 F1 scores of LSTM based model



Red Hat

Training data requirements



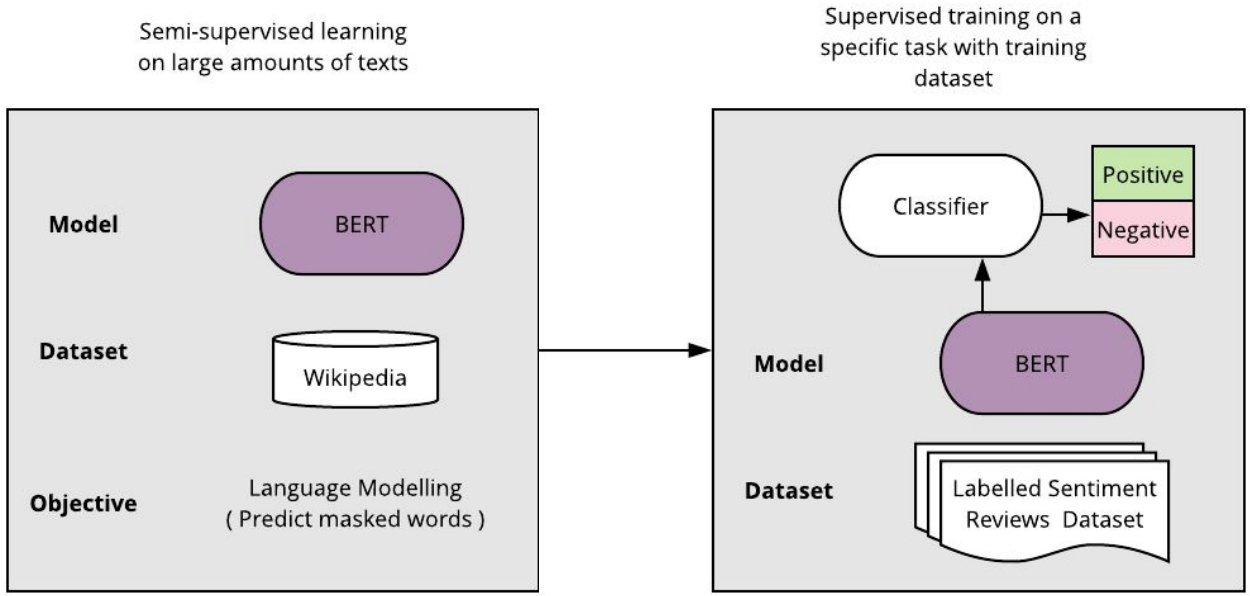
Effective Transfer Learning for NLP
- Madison May, 2018

Red Hat

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

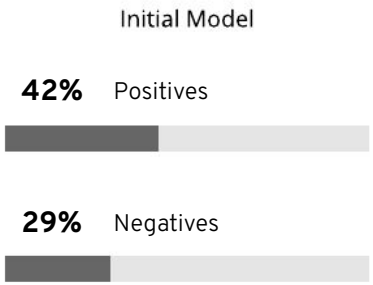
Is the second sentence the actual next sentence that comes after the first or just a random sentence?

Transformer Architecture

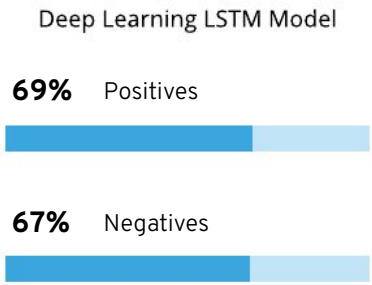
Compared to RNNs which is very sequential, Transformer architecture is able to take full advantage of GPUs and TPUs and make training must faster

Comparative Results

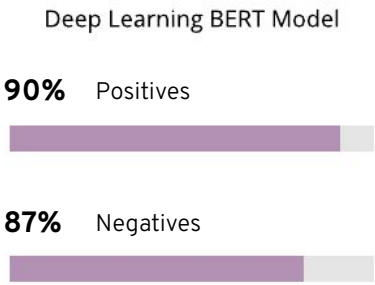
01 F1 scores of Initial Service



02 F1 scores of LSTM based model



03 F1 scores of BERT based model



Red Hat

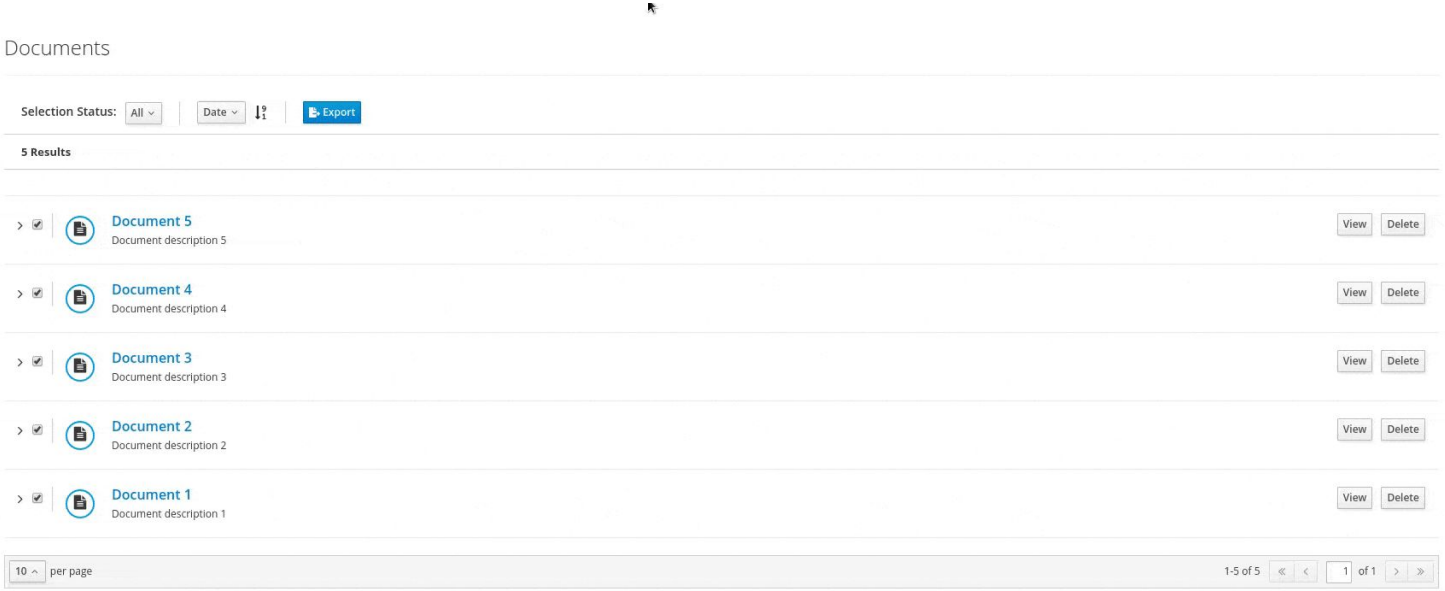
Why is this important?

1. **Better** classification or sentiment annotation

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

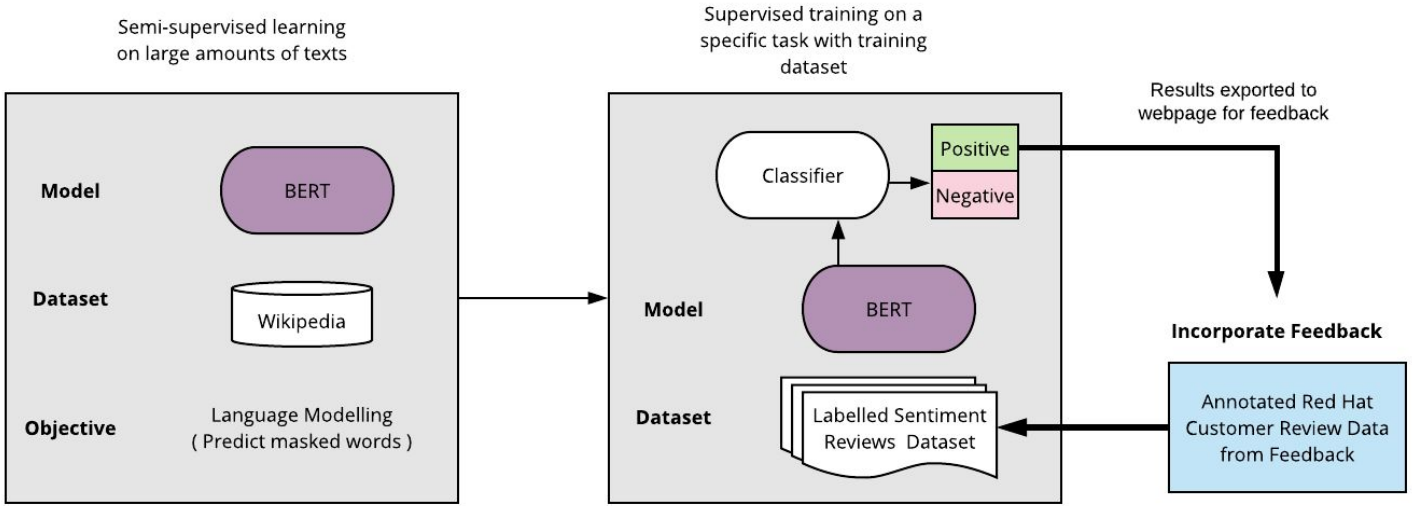
3. Models take much lesser time to train using **GPUs**

Continual improvement of the system



Red Hat

Introducing Feedback

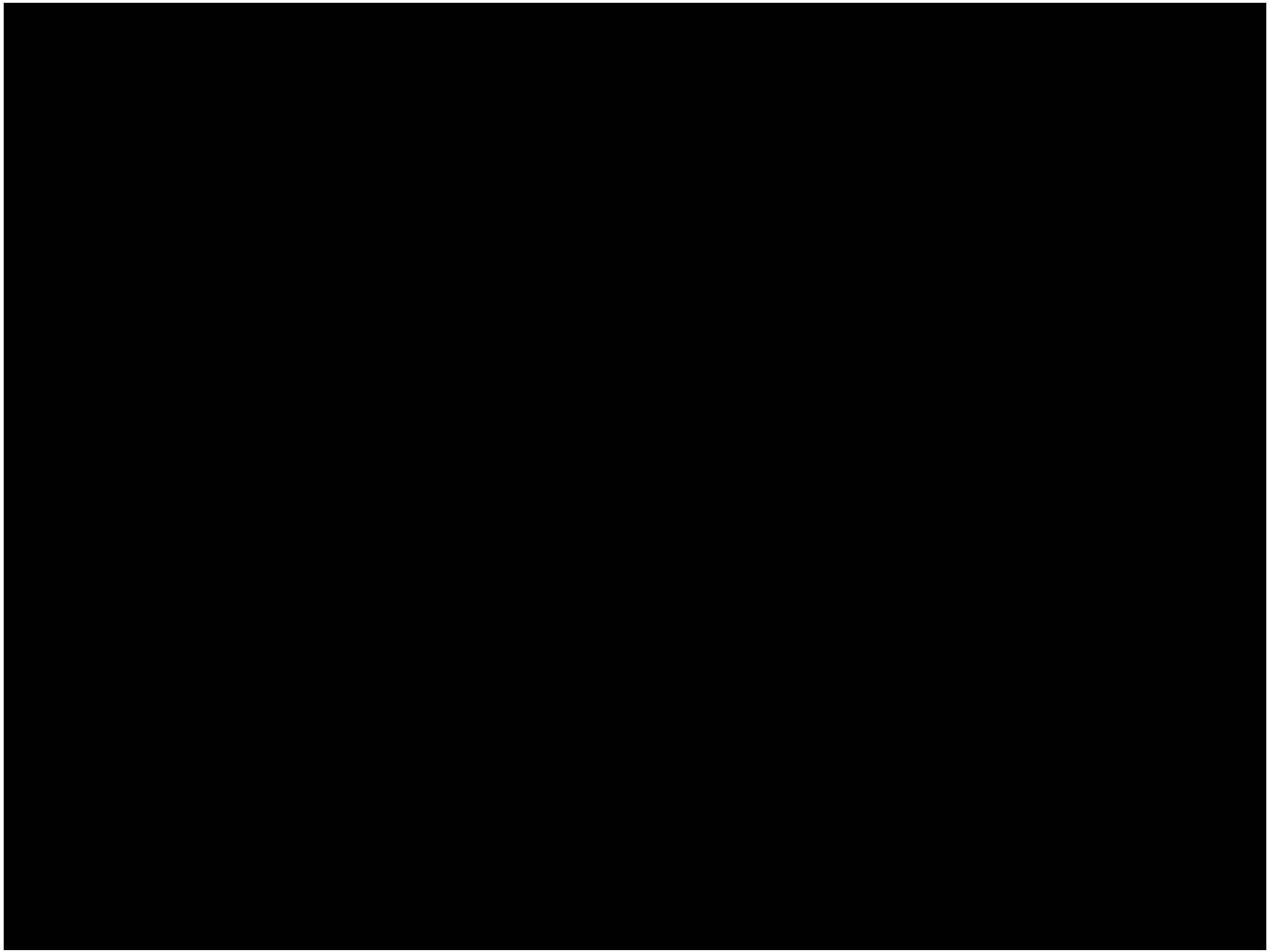


Red Hat

Lessons Learnt

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

Demo



Red Hat

How do you build such a system?

- Open Data Hub is an open-source data and AI platform for the hybrid cloud: https://opendatahub.io
- Meta project to integrate open source projects into a practical service
- Red Hat's internal data science and AI platform

OPEN DATA HUB

AI Platform powered by Open Source

To learn more:

Watch recording of Pete MacKinnon and Juana Nakfour's talk on ML pipelines with Kubeflow, Argo and Open Data Hub

Red Hat

Thank you

ochatter@redhat.com

Red Hat