

Deep Learning for NLP with PyTorch

Dec 12-13, 2020

Topics

- Sequence to Sequence Model using RNN
- Text Classifiers
 - TFIDF and Logistic Regression
 - Pretrained Models, LSTM
- Text Summarization
- Question and Answering models (Chatbots)
- Tools: PyTorch, Gensim, NLTK

Bio

- **Ravi Ilango**
- Lead Data Scientist, Stealth Startup, SantaClara
- 10+ Years at Apple, Sr Data Scientist at FogHorn Systems, Sr DataScientist at StatesTitle
- Education:
 - BE Mech (Madras University, India)
 - Masters Program in Aero and Production (IIT Madras, India)
 - MBA (Santa Clara University)
 - Graduate Certificate in Data Mining and Machine Learning (Stanford)
- email: ravi.ilango@gmail.com

Trust is the new currency in business.

RELYANCE.AI

Trust as a Service
Data Protection Compliance & Legal Ops
Completely Re-Imagined

FUNDING & TEAM

Well-Funded



Unusual
Ventures

Top Notch Talent



ORACLE®

Carnegie
Mellon
University



Massachusetts
Institute of
Technology



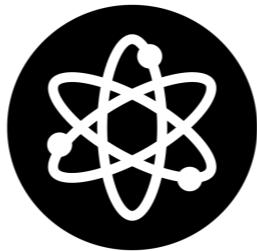
Berkeley
UNIVERSITY OF CALIFORNIA

Google

workday®

APPDYNAMICS

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE



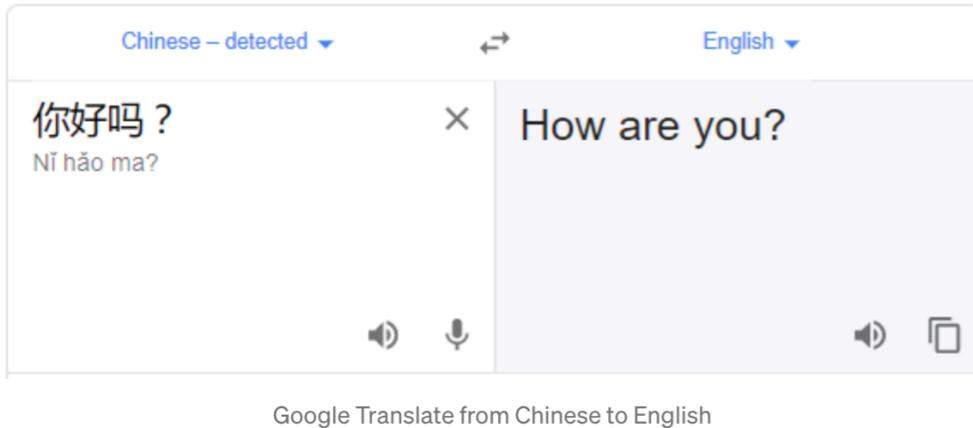
We're hiring:

<https://relyance.ai/careers>

NLP Everyday

Google Translate

used by 500 million people every day



Google Translate from Chinese to English

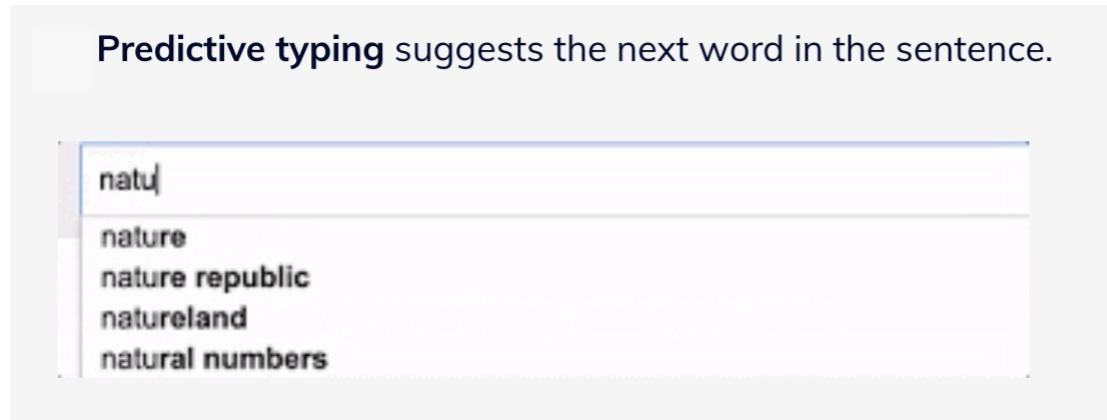
Chatbot



Our Financial Consultants use the askPRU chatbot to provide a better service experience to customers.

Email assistant

Predictive typing suggests the next word in the sentence.



Google



3.5B searches/daily

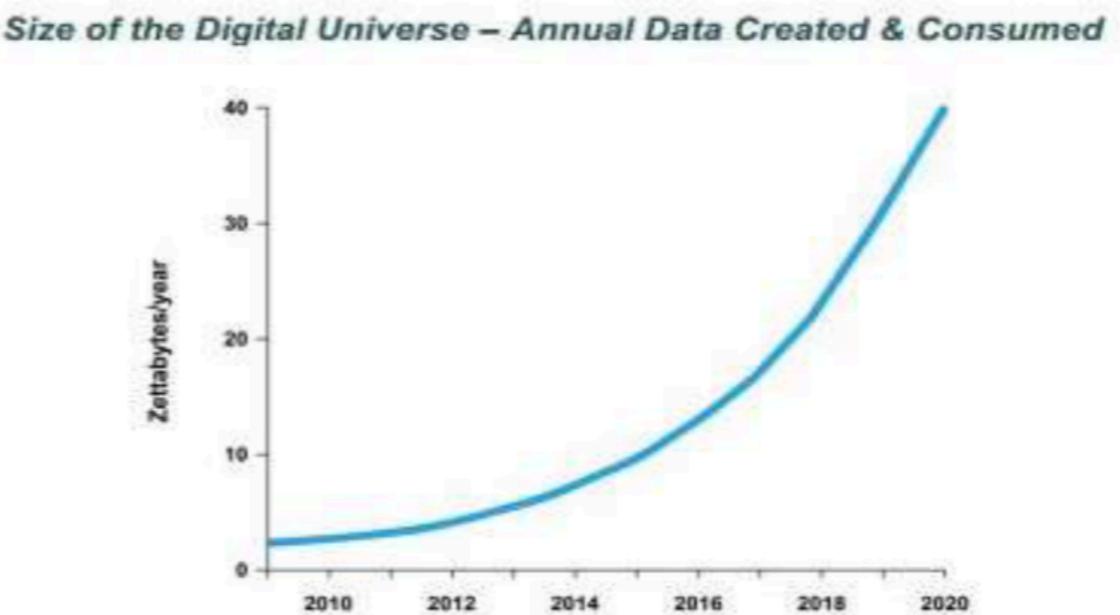
Extract and Summarization

NLP in business operation

- Achieve efficiencies using machine processing of unstructured text data

Text data

- About 80% of data in organizations are in text format
- Harder to analyse than structured data
- Huge amount of textual documents
 - Only in biomedicine 2200 scientific papers are published every day
- Growing exponentially



NLP Use Cases in Enterprises

- **Text Classification-** It is the process of categorizing or tagging text based on its content
 - organizing customer support tickets, customer reviews, emails
- **Text Extraction-** It is the process of extracting specific information from documents
 - extract key info from business documents (invoices, agreements), emails

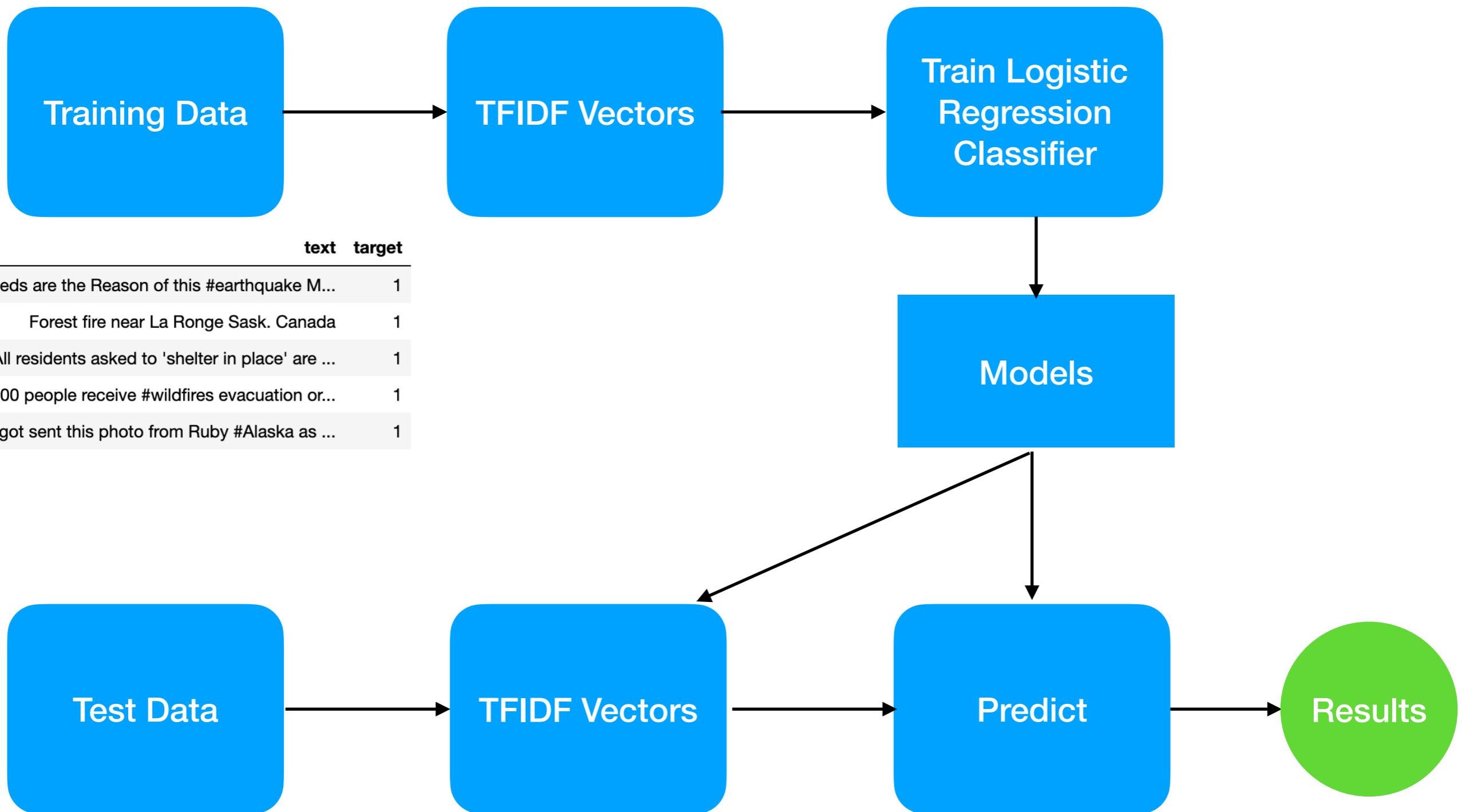
NLP Data Scientist - (some) tools



NLP Data Scientist - (some) concepts & techniques

- Word vectors, tokenization, document vectorization
- POS (parts of speech), NER
- Document parsing techniques, Regex
- Supervised ML, Classifiers, Deep Learning and how to use word embeddings
- Using Pretrained models

Lab: Disaster Detection



Lab - Disaster Detection using TFIDF and Logistic Regression

Click this link to begin

[https://colab.research.google.com/github/ravi-ilango/acm-dec-2020-nlp/blob/main/lab2/disaster detection tfidf.ipynb](https://colab.research.google.com/github/ravi-ilango/acm-dec-2020-nlp/blob/main/lab2/disaster%20detection%20tfidf.ipynb)