

APPLYING ARTIFICIAL INTELLIGENCE METHODS TO BUSINESS PROCESS AUTOMATION



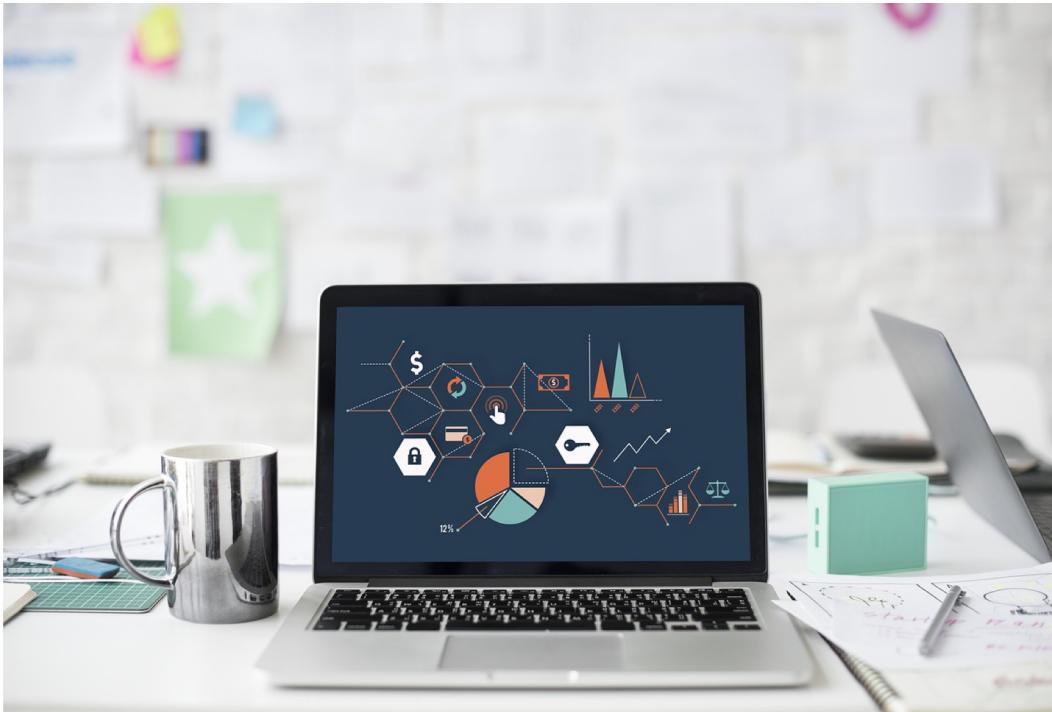
Senior Data Scientist Adam Karwan, PhD
16th November 2018, Gliwice

AGENDA

1. Why Data Science?
2. Personal Experience
 - a. Evolution
 - b. Machine Learning Challenges
3. Business Process Automation
 - a. RPA
 - b. Chatbots
 - c. Data Analytics
 - d. Data Science
4. Demo
 - a. Multiple Text Classification
 - b. Deep Learning vs Standard NLP approach
5. Summary

WHY DATA SCIENCE ?

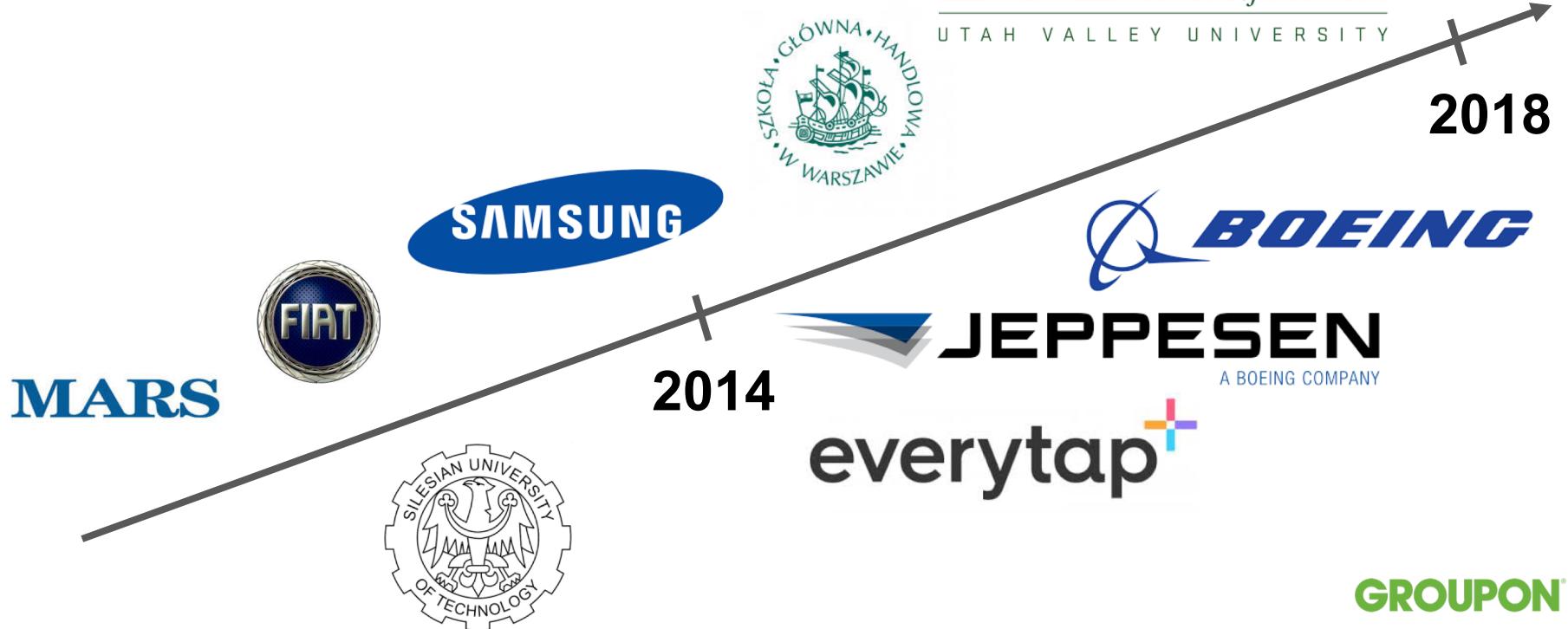
*Data Science will be future and I can work
on the interesting projects for the next years*



GROUPON®

PERSONAL EVOLUTION

Personal Evolution



Kaggle - Data Science as a Competition

- ★ Data Sets
- ★ Contests
- ★ Codes



- ★ Self-learning
- ★ Recruitment
- ★ Algorithms for Companies

BUSINESS PROCESS AUTOMATION

RPA (Robotics Process Automation)

- ★ Automation of the business processes within organization
 - minimize costs, time and human errors
 - existing toolkits or open source to implement scripts



Steps to implement RPA

- I. **Find** and describe business process
- II. **Estimate** complexity and cost saving
- III. **Implementation, deployment and monitoring**



robonomika.pl

www.economist.com/business/2018/09/15/ai-may-not-be-bad-news-for-workers

GROUPON

RPA - use cases, a little brainstorm



Possible areas

- Customer Service
- Data Management
- Financial Services
- Human Resources
- Operations
- Procurement
- Reporting
- Sales

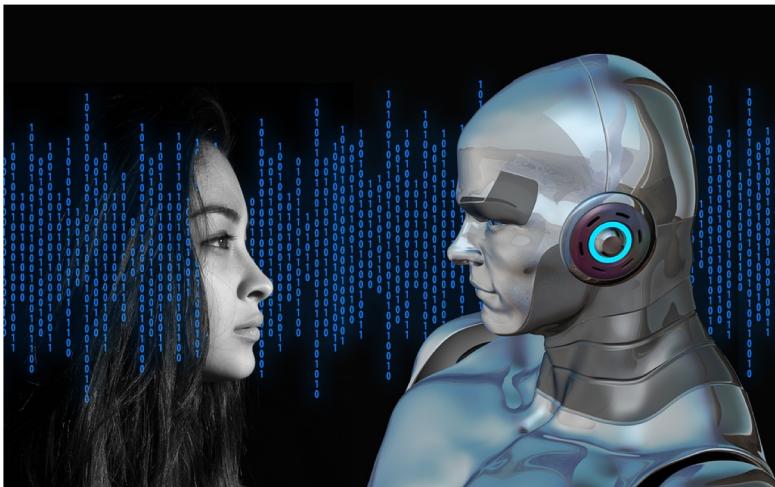
blog.appliedai.com/robotic-process-automation-use-cases

GROUPON

CHATBOTS



Automatic conversation with Customers (Virtual Agent)



Brief History

1966

NLP Computer program ELIZA, @MIT AI Lab, Weizenbaum

2018

Real Conversation via Phone Call and IOT Devices

- ★ TTS (Text to Speech)
- ★ SR (Speech Recognition Engine)

AI Assistant calls local business

youtube.com/watch?v=D5VN56jQMWM

GROUPON

CHATBOTS

Command based and structured
structured questions and answers

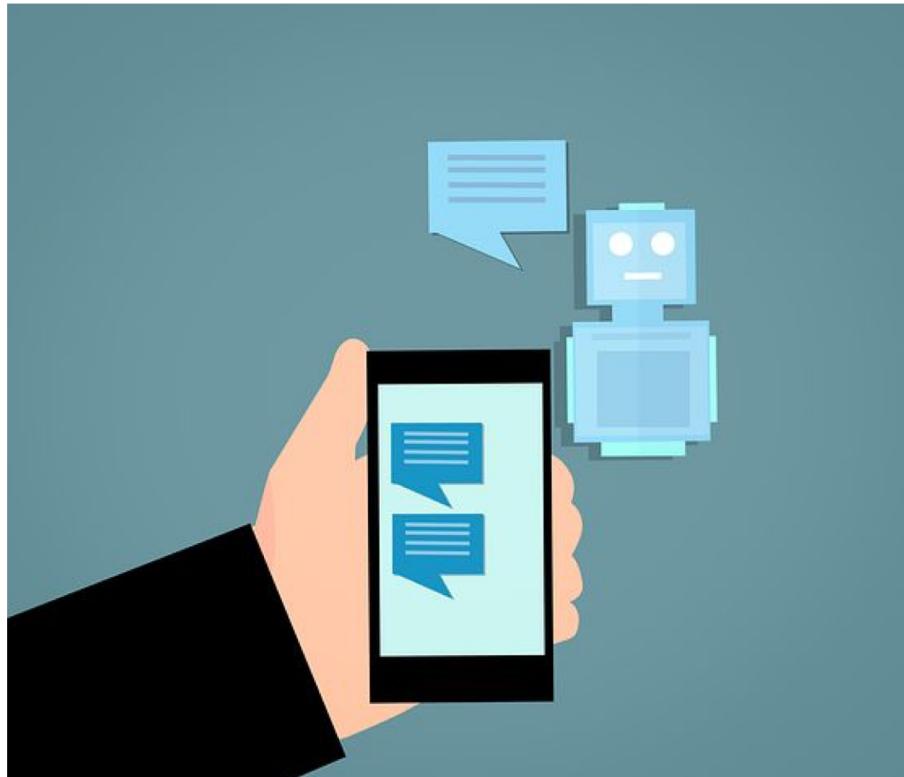
AI based
based on Machine Learning

Pros

- fast
- cheap
- 24/7
- no queues

Cons

- not every business
- limited audience
- retention
- mistakes



CHATBOTS - tips and tricks

- ★ **Plug and play solution**, market research (big players, startups, tailored)
- ★ Implement bots **from scratch** (additional cost, better control)
 - possible, tutorials, online courses, samples codes available in Internet
- ★ How to **feed chatbot with dat**
 - Historical data (Logs from Data Warehouse)
 - Clean and extract conversation via emails, historical chat session, website forms
 - Analize the most common issues (e.g. 70% of all contacts can be in TOP 30)
 - Reorganization of workflow in FAQ, quick answers to the most common issues
 - Create user friendly error messages
 - Monitor, analyze, keep learning, testing, switch on/off ML (self-learning)
- ★ Microsoft **Tay Project** AI ChatBot via Twitter, closed after 16 hours
 - Insane Chatbot, Offensive tweets (racism, holocaust, etc.)

Data Analysis



Standard KPIs

- ★ Satisfaction
- ★ Response Time
- ★ Retention
- ★ Productivity
- ★ Turn Over
- ★ Cash Flows

are the current metrics enough?

Advanced Analytics

Behavioral Analysis | Gamification | Marketing Campaigns
Segmentations | Cross and Upselling Strategies
Personalize Notifications | Predictive Modeling

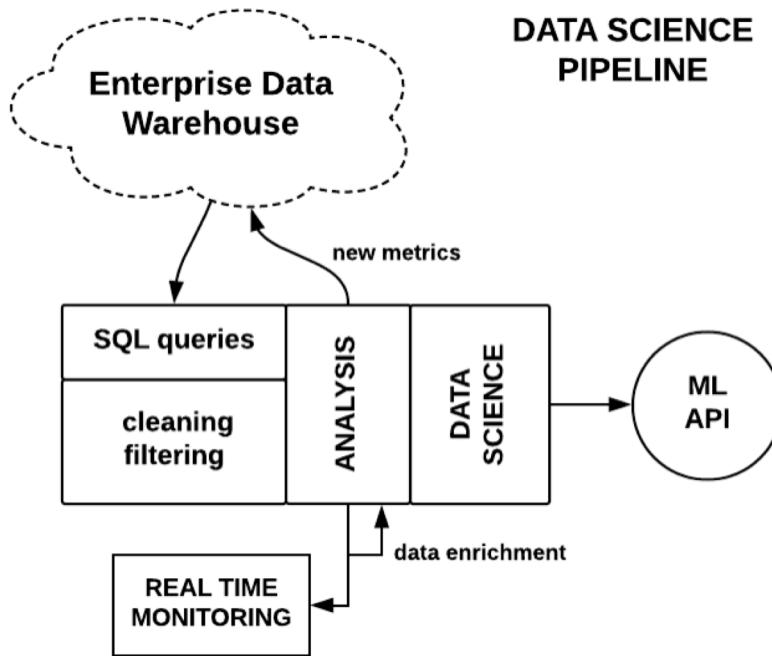
Data Analysis - GDT



Global Decision Tree with all possible policy

- ★ **easily expandable**
- ★ can differ according to geolocation (policy)
- ★ require **constant reorganization**
- ★ could be a possible **bottleneck**
 - remove “dead paths”
 - periodical simplification

Data Science Workflow



4C in DS
Creativeness
Cleverness
Curiosity
Communicativeness

- ★ Accuracy (%)
- ★ Retrain Models

Machine Learning is 90% building pipelines and 10% running algorithms

Automatic Document Categorization

Content based classification | Document organization | Knowledge discovery

- ★ Spam filtering
- ★ **Genre detection**
- ★ Sentiment analysis
- ★ Language identification

- ★ **Emails, Chats, Logs**
- ★ Documents, Scans (with OCR)
- ★ Websites



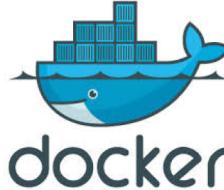
Automatic Document Categorization

- Engines, **APIs** with pretrained models or build your own
- Open Source & SOTA** (state of the art) approach
 - Pipeline (Standard NLP)
 - Deep Learning (RNN + BLSTM)
 - etc.

Natural
Language
ToolKit



gensim



GROUPON®

Challenge description

Categorize **consumer financial complaints** into **11 classes**

US Consumer Finance Complaints

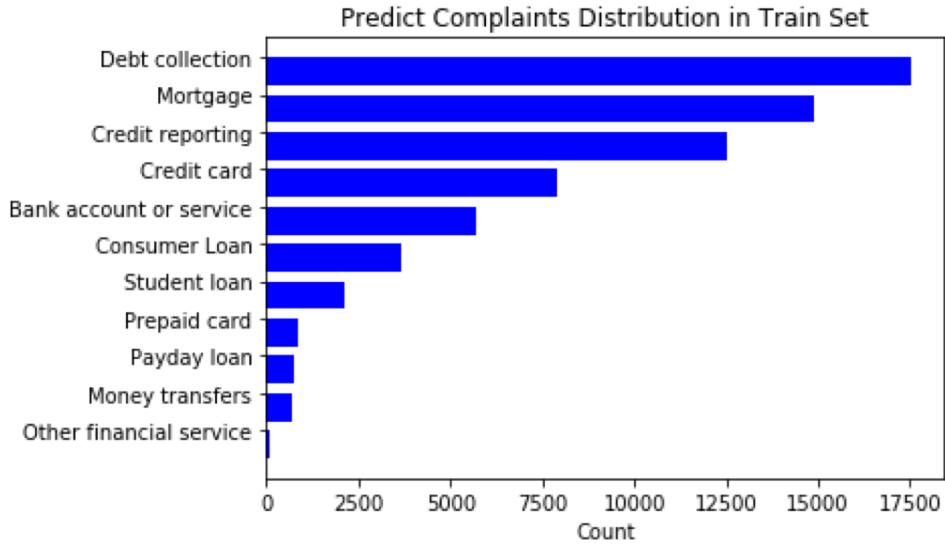
US consumer complaints on financial products and company responses

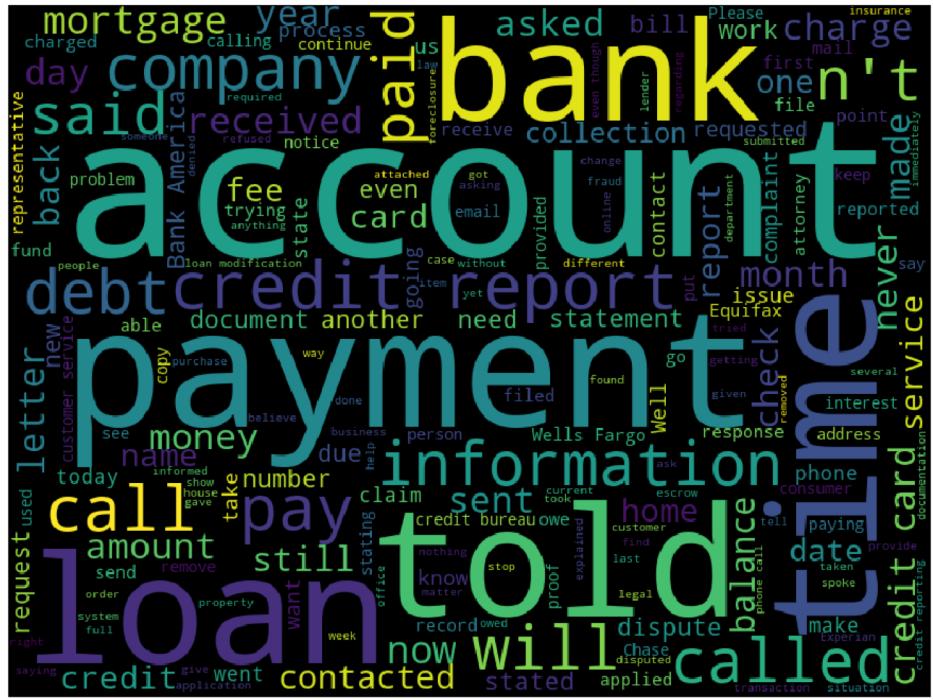
Thousands of the consumer complaints weekly
about financial products and services

Randomly ~9% Acc

Need classification before sending final respond within 15 days

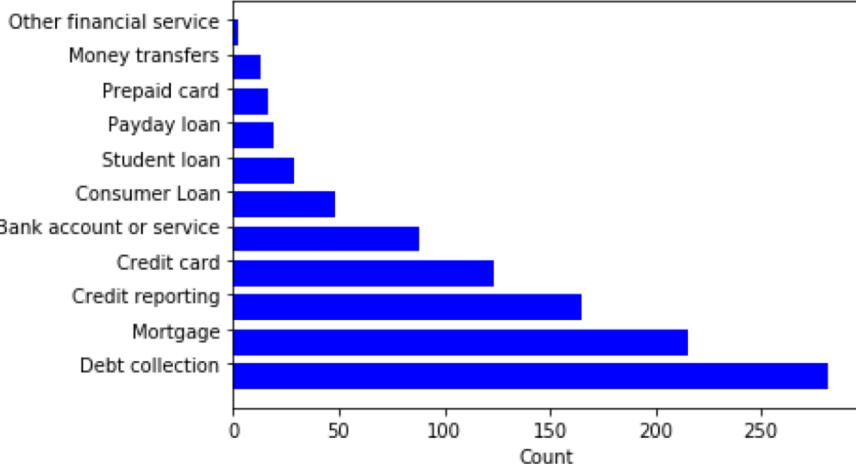
Data Train / Test 66.8k / 1k





Test 1k Data

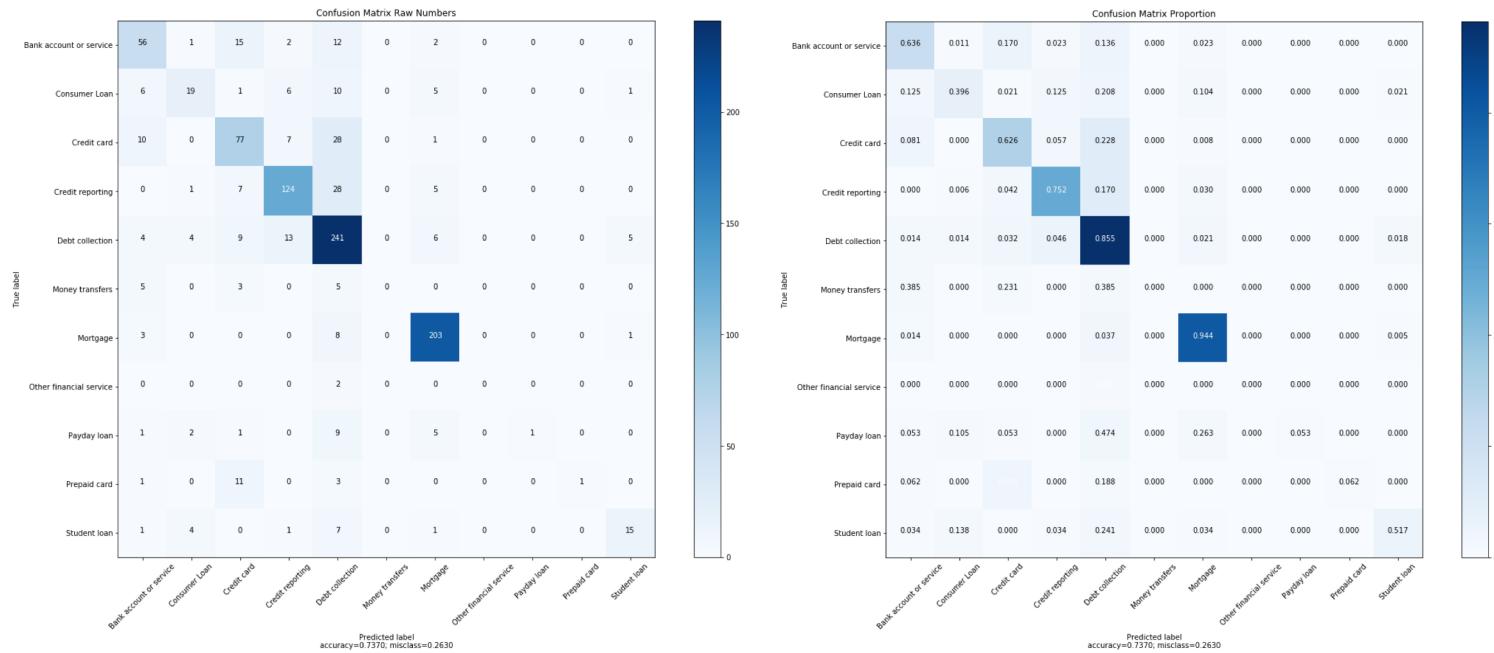
Predict Complaints Distribution



'Debt collection'

'This creditor is harrassing me at work. They are calling my children repeatedly. \n'

TensorFlow Model, train in 2 minutes



Final Prediction 73.7% on 1000 samples

GROUPON®

Transformation	Stemming	TF-IDF	Keras	DL Topology	Final Model
Data Filtering	Case Folding	Embedding	BRNN & LSTM		Ready to Deploy
Removing: - tags - duplicates	Removing - Stop words - Double Letters	FastText Word2Vec	Naive Bayes Regression Linear SVM SVM (Kernel) Random Forests etc.	Hyperparameters Grid Search Ensemble methods	Retrain and Update Model
Formatting	Add New Features	Data Enrichment (non Textual) Other embeddings ELMo		Confusion Matrix	

e.g. standard NLP approach

```
model = Pipeline([('vect', CountVectorizer(ngram_range=(1,3))),
                  ('tfidf', TfidfTransformer(use_idf=False)),
                  ('clf', CalibratedClassifierCV(LinearSVC(), cv=2))])
```

- ★ Category label
- ★ Probability of the correct categorization

Deep Learning, model in Keras



```
# Topology
inp = Input(shape=(maxlen, )) # maxlen=300
embed_size = 128
x = Embedding(max_features, embed_size)(inp)
x = Bidirectional(LSTM(40, return_sequences=True, name='lstm_layer'))(x)
x = GlobalMaxPool1D()(x)
x = Dropout(0.1)(x)
x = Dense(20, activation="tanh")(x) # 50
x = Dropout(0.1)(x)
x = Dense(11, activation="softmax")(x) # original "sigmoid"

# Prepare Model
model = Model(inputs=inp, outputs=x)
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

batch_size = 32 # original 32
epochs = 2
```

Layer (type)	Output Shape	Param #
input_7 (InputLayer)	(None, 300)	0
embedding_6 (Embedding)	(None, 300, 128)	2560000
bidirectional_6 (Bidirection (None, 300, 80)		54080
global_max_pooling1d_6 (Glob (None, 80)		0
dropout_11 (Dropout)	(None, 80)	0
dense_11 (Dense)	(None, 20)	1620
dropout_12 (Dropout)	(None, 20)	0
dense_12 (Dense)	(None, 11)	231
<hr/>		
Total params:	2,615,931	
Trainable params:	2,615,931	
Non-trainable params:	0	

training ~15 minutes

Github vs Standard vs Keras



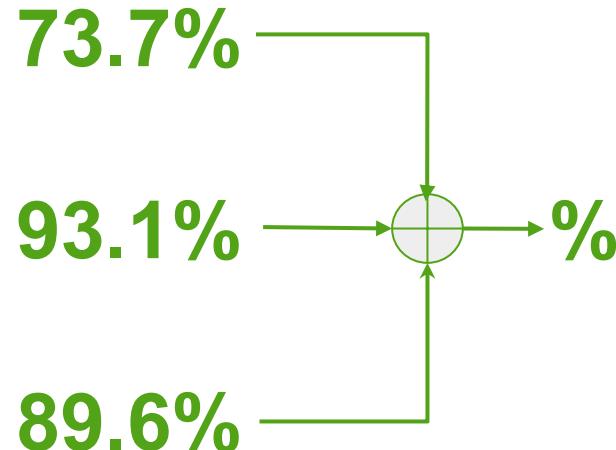
multi-class-text-classification-cnn

```
model = Pipeline([('vect', CountVectorizer(ngram_range=(1,3))),  
                  ('tfidf', TfidfTransformer(use_idf=False)),  
                  ('clf', CalibratedClassifierCV(LinearSVC(), cv=2))])
```



Keras RNN with BLSTM

- ★ Ensemble method
- ★ Imbalanced data (Imputation, e.g. SMOTE Algorithm)
- ★ Provide correct annotations



GROUPON®

ML new opportunities TPU/GPU/Clouds

- ★ **TPU/GPU** (speed up Deep Learning Models trained locally)
- ★ **Cloud Computing** (VM's with possible ML solutions)

- Big players & new solutions
- Cheap data storage per hour
- Decreasing storage and computational costs in time
(Energy price is limit)
- Leverage dataset with storage system
- Additional cost with data migration, transition
- Speed up data scientist work
- Limitations (downtime, security, privacy, flexibility)



SUMMARY

Welcome!

Learn about GPN through our guides and explore the APIs

GROUPON PARTNER NETWORK

partner-api.groupon.com

github.com/groupon



Groupon API

Work with our data warehouse to access data and perform your own data analysis



Link Guide

Use our extensive range of banners and widgets



How to get paid

Learn how and when you will get paid

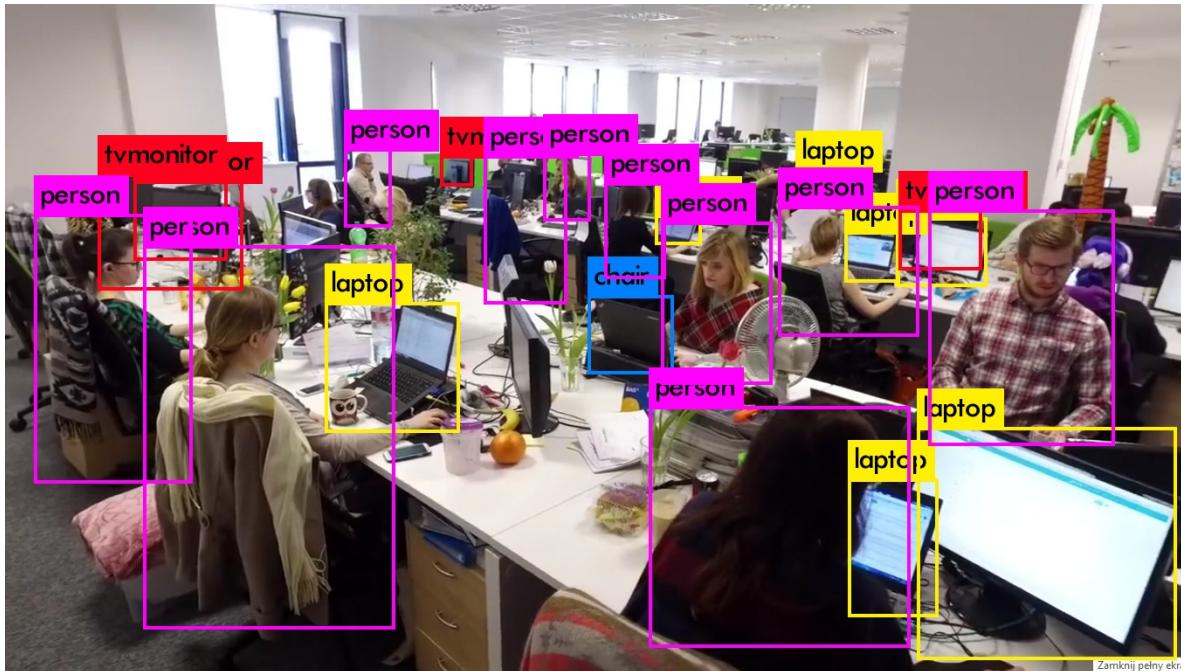
We are growing in Katowice



1 - 10 - 6k - 15 - 49M - 1.5B

GROUPON®

Convolutional Neural Networks



<https://www.youtube.com/watch?v=VOC3huqHrss> YOLO version 2

Invitation

mlgdansk.pl

on going bi-weekly Technical meetings about Data Science in Tricity

AI is the future, if I did not believe it, I would not do it



akarwan@groupon.com

GROUPON[®]