oxagile

Deep learning approach to classify inappropriate review

with text and additional information

+1 855 466 9244 | contact@oxagile.com | www.oxagile.com

oxagile

Vlad Artemyev

has been working in the field of Deep Learning for more than 4 years, solving various problems related to Computer Vision, OCR, NLP.

oxagile
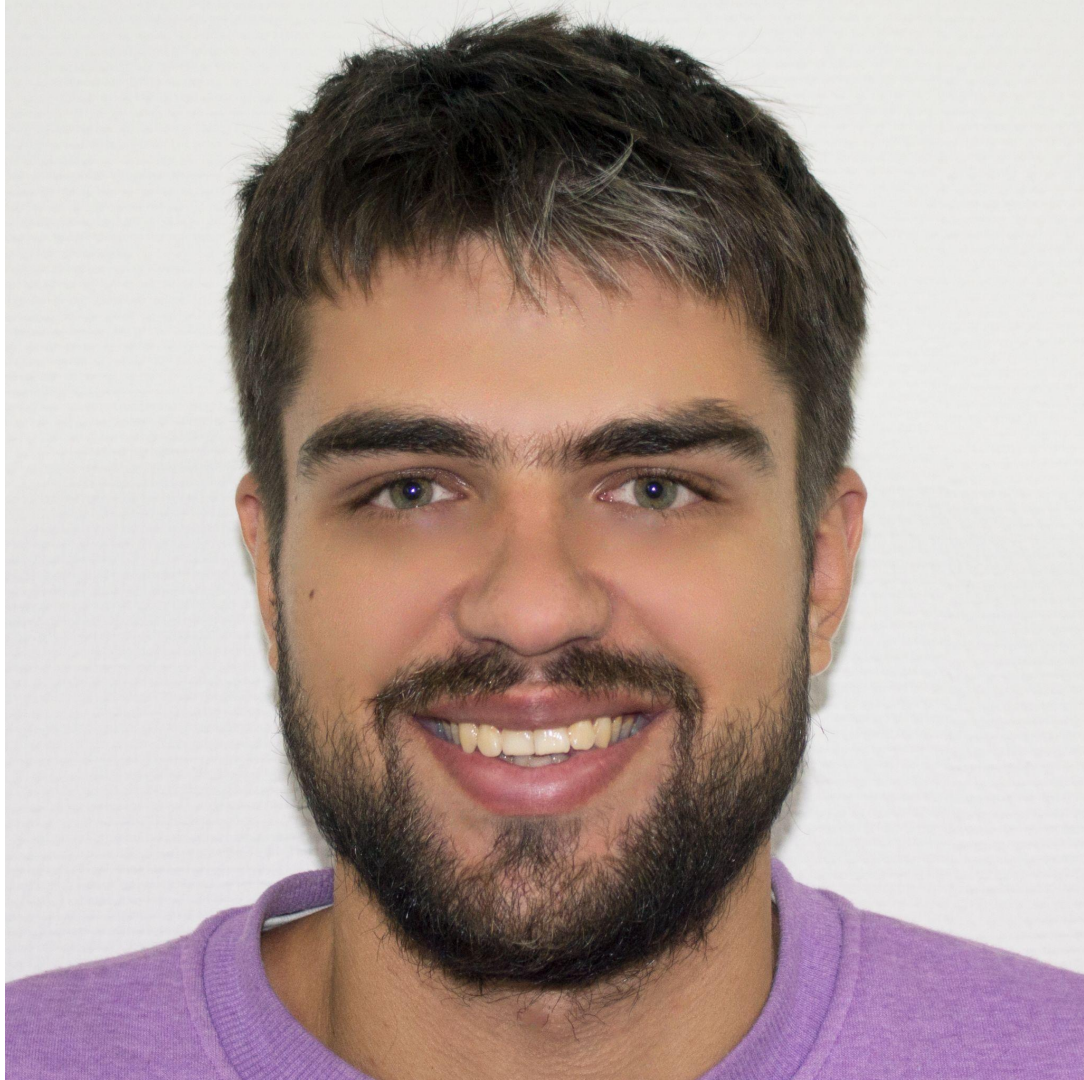
Speech Plan

- Formulation of the problem
- Data Analysis
- Model selection
- Experiments

Formulation of the problem

Formulation of the problem

Reviews are an important part of every business. Based on business reviews, you can analyze the experience of business applications by other people and make a decision for yourself whether to use this business.

Formulation of the problem

Reviews are an important part of every business. Based on business reviews, you can analyze the experience of business applications by other people and make a decision for yourself whether to use this business.

From our side, it is useful for the user to see the most relevant reviews, and, for example, off-topic or toxic reviews should be removed.

oxagile

Formulation of the problem

Given review text and some additional information about review identify whether we should delete the review or denied removal.

In other words, we simulate the behavior of moderators using a neural network.

Achieve 80% accuracy on test data.

oxagile.com

Data Analysis

Data Analysis

Train data features available

- Account: Account Name
- Account: Industry
- Dispute Reason
- Comments

- When was the review posted?
- When was this negative removed?
- Status
- Star Rating
- Review Text

Data Analysis

Test data features available

- Account: Account Name
- Account: Industry
- When was the review posted?
- Star Rating
- Review Text

Data

analysis

Describe train data

oXagile

	Account: Account Name	Account: Industry	Dispute Reason	Comments	When was the review posted?	When was this negative removed?	Status	Star Rating	Review Text
0	College Hunks Hauling Junk Kansas City	General	Negative/Inappropriate Photo	Photo of a desk flagged for spam content.	11/16/17	3/23/21	Removed	1 Star	Photo of a desk flagged for spam content.
1	Los Angeles School of Gymnastics	General	NaN	Please remove this review. It is no longer dis...	NaN	12/9/13	Removed	NaN	NaN
2	Trails at Dominion Park	General	Off topic	Katy is an ex employee leaving this review too...	9/25/20	NaN	Disputed	1 Star	The office employees are rude and the apartmen...
3	Lakeside Family Dental	Dentist	NaN	this place is pretty good if you like mediocre...	NaN	NaN	Denied Removal	NaN	NaN
4	Utopia Management � Orange County	General	NaN	NaN	NaN	NaN	Removed	NaN	NaN
...	...	...	...	...	...	...	...	...	...
157651	Off Road Unlimited	Auto	NaN	This reviewer states, "I was looking forward t...	12/11/15	NaN	Denied Removal	1 Star	I was looking forward to having my truck worke...
157652	Petland - Kennesaw	General	NaN	--- This is not an experience with the busines...	5/31/16	6/10/16	Removed	1 Star	Julie Fleiser in the last week Buys dogs from ...
157653	Legend Dental - Austin PLLC	Dentist	NaN	legenddentalaustin@gmail.com 2525Wanderson 08/...	5/21/16	NaN	Disputed	1 Star	Scheduled a cleaning about 2-3 months in advan...
157654	Tint World Boca Raton, FL	General	Inappropriate Content	This user's name is offensive, as well as usin...	6/28/18	3/18/21	Removed	3 Stars	UPDATE & Resolution in process!!! ALTHOUGH TH...
157655	Tint World Boca Raton, FL	General	Off topic	This review is off-topic. It is not describing...	1/1/18	NaN	Denied Removal	1 Star	Horrible service!

157656 rows × 9 columns

oxagile.com

Review Example

Account Name

- Porcelain Tub Restorations

Account: Industry

- General

Dispute Reason

- Off topic

Comments

- Off topic rant that does not describe any valid content with this review user is ranting about no sense please remove this review.

Posted at?

- 9/24/20

Removed At?

- 10/15/20

Status

- Removed

Star Rating

- 1 Star

Review Text

- …

Review Text Example

Why on earth woukd anyone do business with a person who writes things like this online in public "try to get therapy for your autism" I said you were sad and angry, and you really are. If your going to insult someone dont stoop to making light of something as serious as autism. Like I said people stay clear of Mrs Angry.

Data analysis

Describe train data

oXagile

	Account: Account Name	Account: Industry	Dispute Reason	Comments	When was the review posted?	When was this negative removed?	Status	Star Rating	Review Text
0	College Hunks Hauling Junk Kansas City	General	Negative/Inappropriate Photo	Photo of a desk flagged for spam content.	11/16/17	3/23/21	Removed	1 Star	Photo of a desk flagged for spam content.
1	Los Angeles School of Gymnastics	General	NaN	Please remove this review. It is no longer dis...	NaN	12/9/13	Removed	NaN	NaN
2	Trails at Dominion Park	General	Off topic	Katy is an ex employee leaving this review too...	9/25/20	NaN	Disputed	1 Star	The office employees are rude and the apartmen...
3	Lakeside Family Dental	Dentist	NaN	this place is pretty good if you like mediocre...	NaN	NaN	Denied Removal	NaN	NaN
4	Utopia Management ◆ Orange County	General	NaN	NaN	NaN	NaN	Removed	NaN	NaN
...	...	...	...	...	...	...	...	...	...
157651	Off Road Unlimited	Auto	NaN	This reviewer states, "I was looking forward t...	12/11/15	NaN	Denied Removal	1 Star	I was looking forward to having my truck worke...
157652	Petland - Kennesaw	General	NaN	--- This is not an experience with the busines...	5/31/16	6/10/16	Removed	1 Star	Julie Fleiser in the last week Buys dogs from ...
157653	Legend Dental - Austin PLLC	Dentist	NaN	legenddentalaustin@gmail.com 2525Wanderson 08/...	5/21/16	NaN	Disputed	1 Star	Scheduled a cleaning about 2-3 months in advan...
157654	Tint World Boca Raton, FL	General	Inappropriate Content	This user's name is offensive, as well as usin...	6/28/18	3/18/21	Removed	3 Stars	UPDATE & Resolution in process!!! ALTHOUGH TH...
157655	Tint World Boca Raton, FL	General	Off topic	This review is off-topic. It is not describing...	1/1/18	NaN	Denied Removal	1 Star	Horrible service!

157656 rows × 9 columns

oXagile.com

Clean Data

- Remove empty reviews

oxagile

oxagile.com

oxagile

Data
Analysis

We have

117K

Not empty reviews

Clean Data

- Remove empty reviews
- Remove useless statuses

Remove Useless Statuses

Unique Statuses

- Removed
- Disputed
- Denied Removal
- Flagged
- Re-Disputed
- Responded as Owner

- nan
- Filtered
- Reallocated
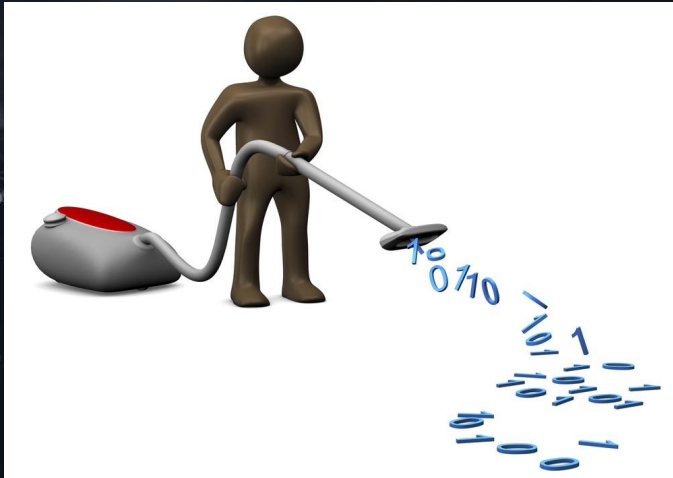- Escalated
- Please Dispute Next

oxagile

Data Analysis

We have

89K

Reviews after deleting useless statuses

Clean Data

- Remove empty reviews
- Remove useless statuses
- Remove duplicated reviews?

oxagile

oxagile.com

Duplicated Reviews

We have 58928 unique reviews. But for 3641 reviews we have duplicated ones with different status.

When ordering online, their system showed items as available, but they were not available. They were back-ordered. I had to call to find out. Also their web site did not work well using a Samsung Android phone.

Overpriced

```
['Denied Removal',
 'Removed',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal']


['Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Removed',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Removed',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal',
 'Denied Removal']
```

Clean Data

- Remove empty reviews
- Remove reviews with useless statuses
- Remove duplicated reviews?
- Remove reviews with inappropriate dispute reason

oxagile

oxagile

Data Analysis
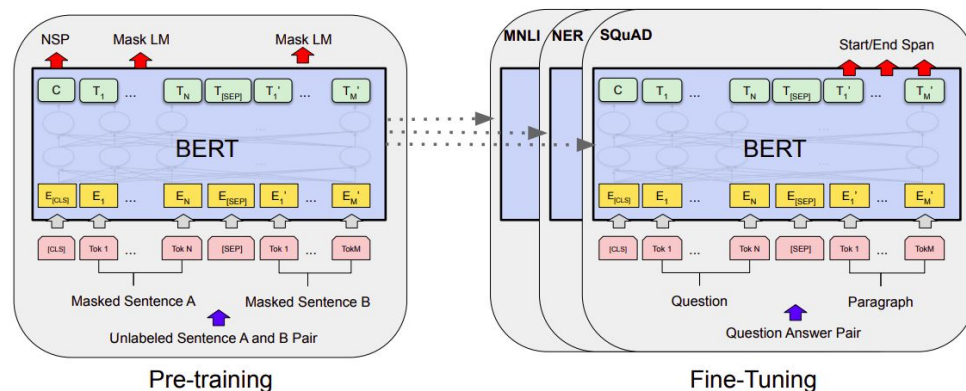
We have

56K

Reviews after deleting reviews with dispute reason equals to Negative/Inappropriate Photo
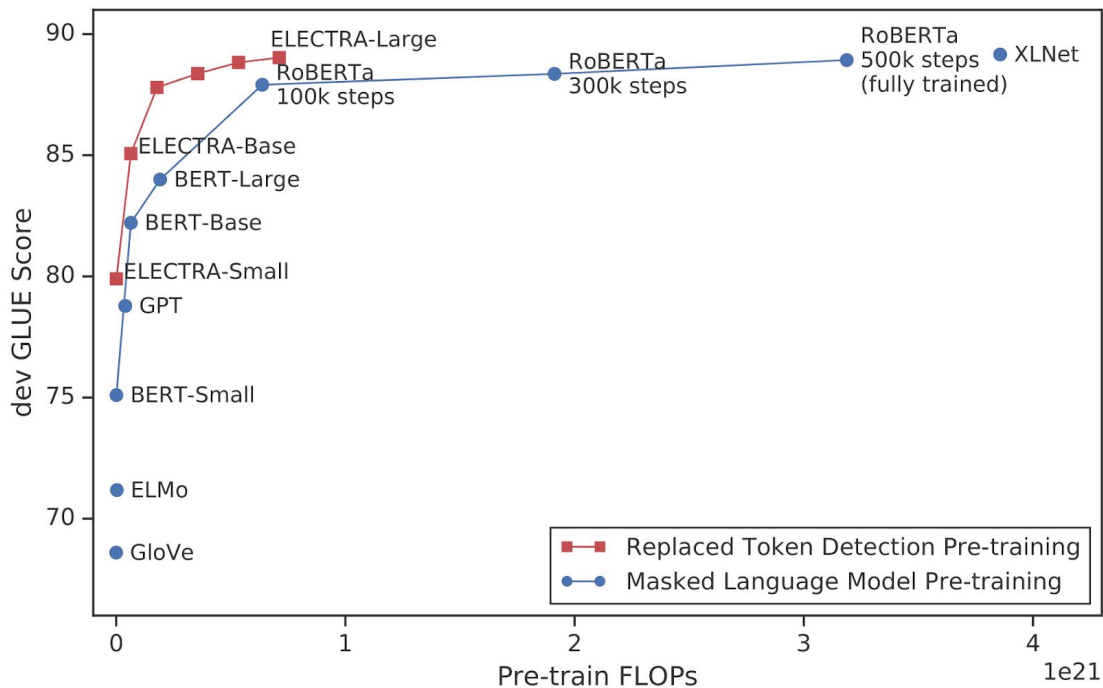
Model Selection

oxagile

BERT

is a transformer-based machine learning technique for natural language processing (NLP) pre-training developed by Google.

BERT

BERT is better than previous SOTA models.

oxagile

oxagile.com

Google Search

On October 25, 2019, Google Search announced that they had started applying BERT models for English language search queries within the US.

On December 9, 2019, it was reported that BERT had been adopted by Google Search for over 70 languages.

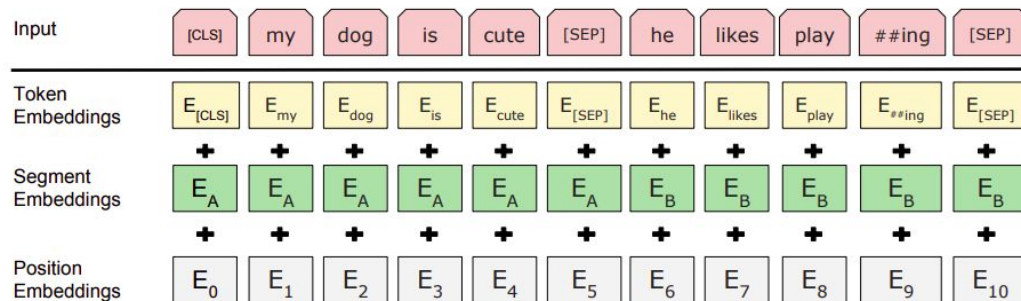In October 2020, almost every single English-based query was processed by BERT.

BERT

- Not recurrent
- Context-based model
- Bidirectional

BERT Embeddings

The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{\#\#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}

oxagile

BERT Classification
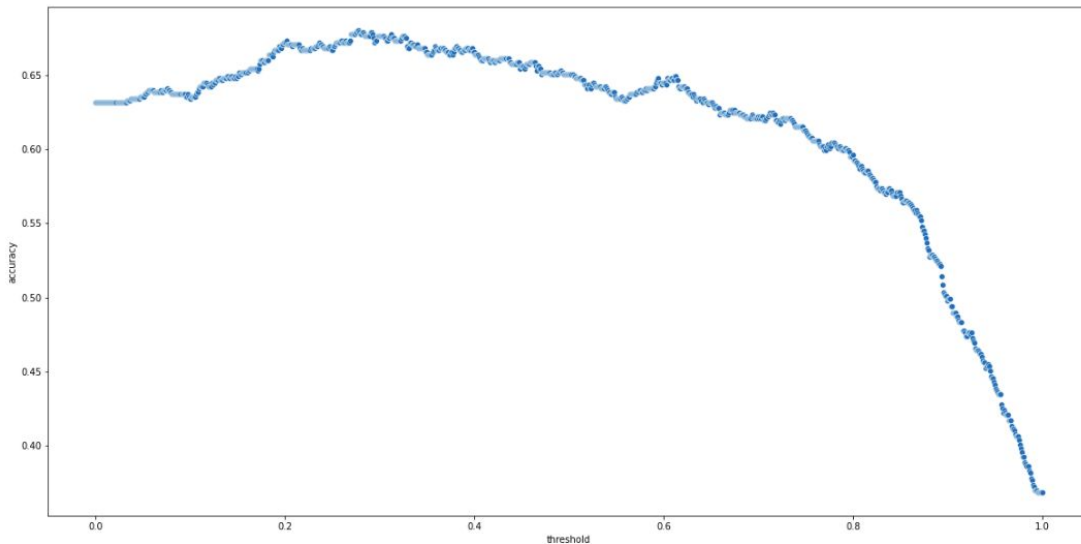
Finetuning pretrained model for classification task.
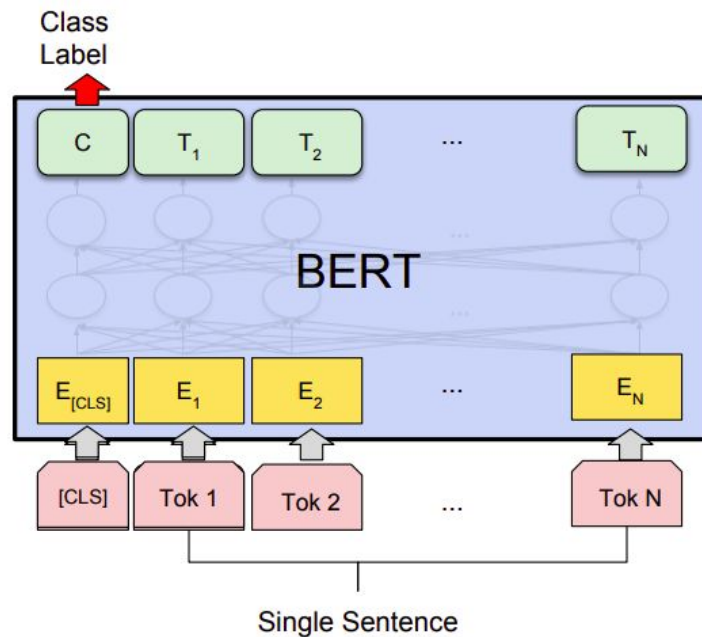
(b) Single Sentence Classification Tasks: SST-2, CoLA

Experiments

Train only with text

68% accuracy on test data is very slight improvement. What is the reason?

Dispute Reason

The most popular dispute reason is offtopic.

```
Ввод [70]:  df['Dispute Reason'].describe()

Out[70]:  count          79589
          unique            27
          top        Off topic
          freq           58550
          Name: Dispute Reason, dtype: object
```

157637	Cal Spas of Minnesota	General	Off topic	This review has nothing to do with the service...
157641	Petland Round Lake Beach	General	Off topic	Tells people to go to petsmart and is only ran...
157655	Tint World Boca Raton, FL	General	Off topic	This review is off-topic. It is not describing...

58550 rows × 10 columns

Data Analysis

Test data features available

- Account: Account Name
- Account: Industry
- When was the review posted?

- Star Rating
- Review Text

Train with concatenated fields

76% accuracy on test data is good but it's still less than 80%.
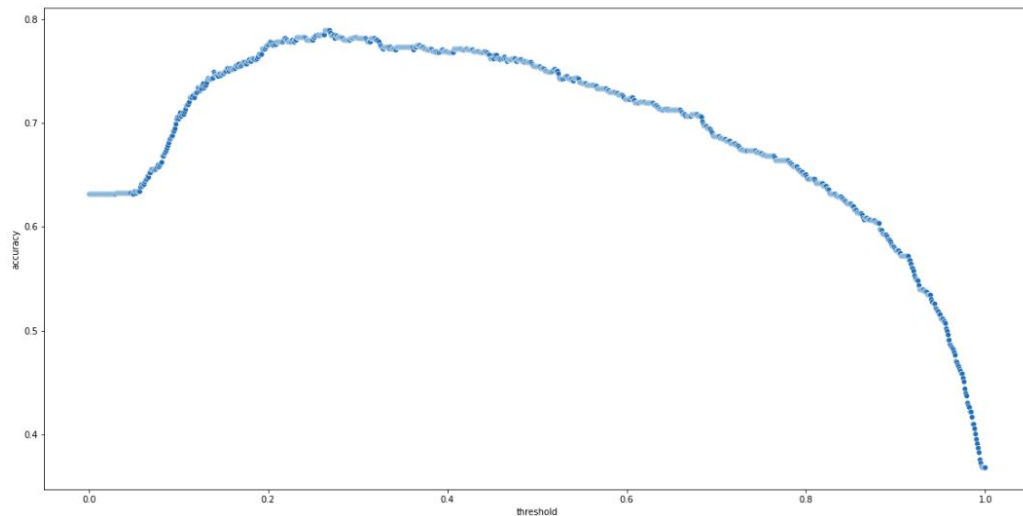
How to improve accuracy?

Clean Data

- Remove empty reviews
- Remove reviews with useless statuses
- Remove duplicated reviews?
- Remove reviews with inappropriate dispute reason

What is we do not delete duplicated review?

Train with not deleted duplicated reviews

Unexpectedly we got boost of accuracy. It is about 78.9%. Good, but still less than 80%.

Magic Augmentation

Each epoch for each additional field (except for the review text) we use random to determine whether we use this field in concatenation or not.

Example of random field sampling

Each epoch input from each review is different.

```
__ __ __ __ __ __... Translation: The meat procession that never ends...'

'Pho Hoa and Jazen Tea. __ __ __ __ __ __... Translation: The meat procession that never ends...'

'General. __ __ __ __ __ __... Translation: The meat procession that never ends...'

'Pho Hoa and Jazen Tea. General. __ __ __ __ __ __... Translation: The meat procession that never ends...'
```
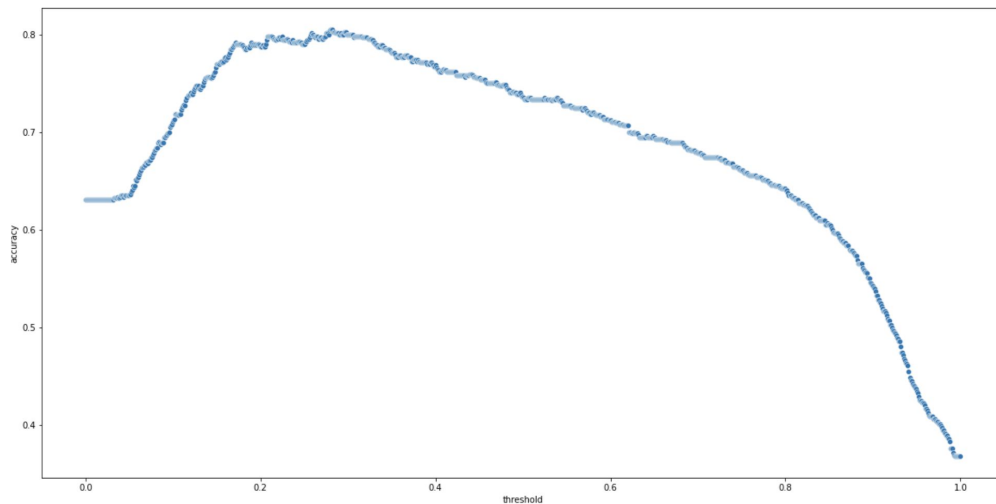
Train with random field sampling

We got 80.5% accuracy on test test which is good enough and we achieve the goal of 80%.

oxagile

Let's talk solutions

We're here to help you with your digital transformation journey. Contact us and an expert will get back to you to discuss your business needs in detail.

HEADQUARTERS

+1 855 466 9244
usa@oxagile.com
165 Broadway, 23rd Floor, New York, NY 10006, USA

REGIONAL OFFICE

+44 203 514 2366
uk@oxagile.com
Kemp House, 152-160 City Road, London, EC1V 2NX, UK

DEVELOPMENT CENTER

+375 29 619 34 09
contact@oxagile.com
57 praspiekt Dziaržynskaha, Minsk, 220089, Belarus

oxagile.com