

Multi-label classification

Procedure

1. Selective picking of HTML tags' content to include in the dataset.
2. Removal of accented and non-ascii characters, stopwords, languages other than English, and punctuations.
3. Lowercasing and tokenization.
4. Perform TF-IDF on the dataset.
5. Pick tokens having highest TF-IDF weights in the document.

Selection criteria:

1. If more than 20 tokens, pick 60% of tokens, TF-IDF weight sorted.
2. If between 6 and 20 tokens, pick 70% of tokens, TF-IDF weight sorted.
3. If less than 6 tokens, include all tokens.
6. Using FastText, generate embeddings of the picked tokens from the document.
7. Calculate similarity of every picked token with every generalized token from the list – agriculture, banking, blockchain, business, crime, cryptocurrency, economy, education, entertainment, environment, fashion, finance, food, governance, health, network, politics, security, software, sport, technology, web.
8. Assign the generalized token with the highest similarity score to the token.

Assignment criteria: only assigned if similarity is more than 60%; manual assignment of certain labels such as {onion, tor} to network.

9. Create a set of assigned generalized tokens per document – thereby used as the labels in the classification.
10. Split the dataset into train and test sets.

Split criteria: 25% testing set and 75% training set.

11. Fit the classifier neural network model.
12. Evaluate the model using the test set.

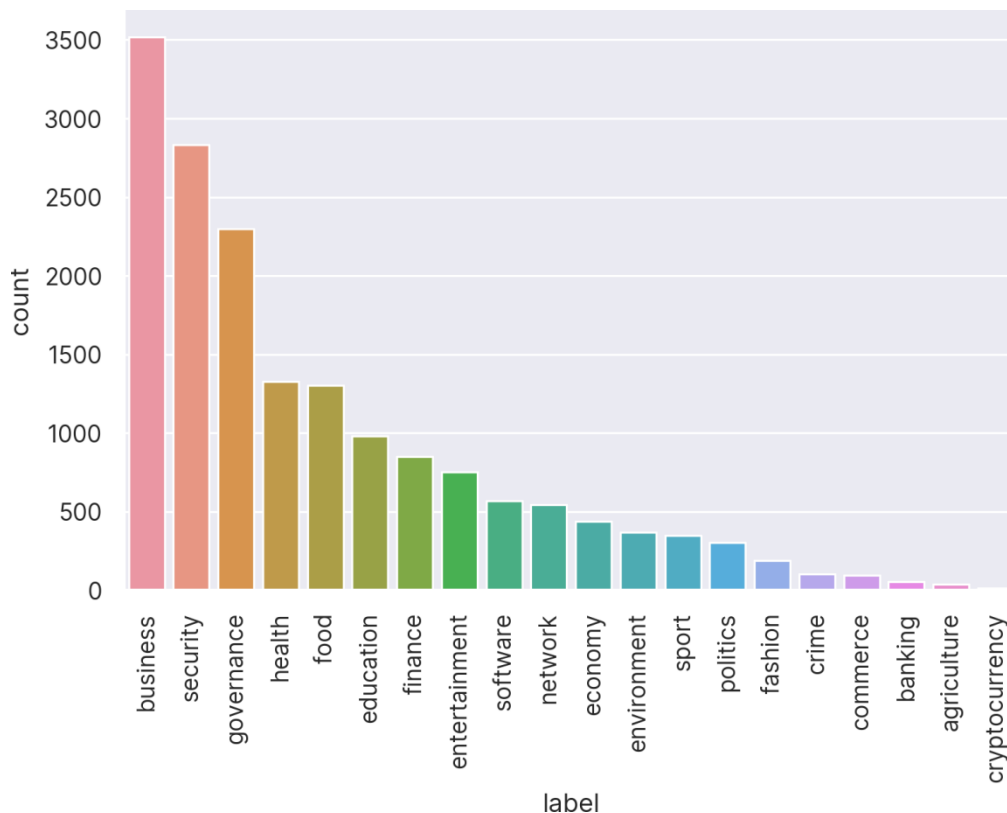
Results and Discussion

Generic Hidden Data Wiki

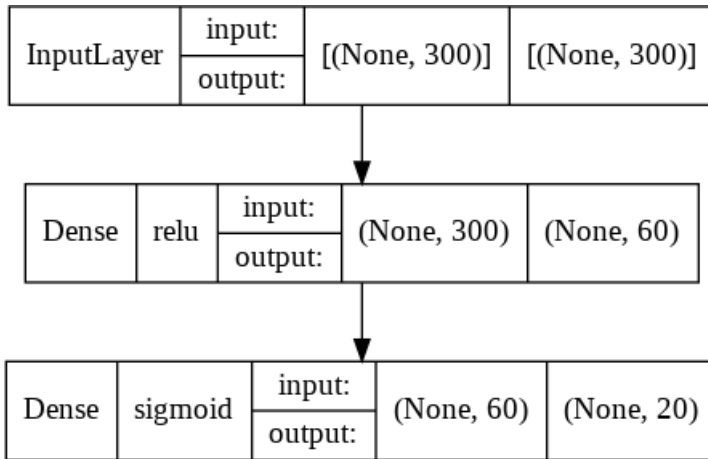
Given: Raw HTML pages

```
<!DOCTYPE html>
<html lang="zh-Hans" id="facebook" class="no_js">
<head><meta charset="utf-8" /><meta name="referrer" content="default" id="meta_referrer" /><script nonce="Yuhwhab6">window._cstart=new
Date();</script><script nonce="Yuhwhab6">function envFlush(a){function b(b){for(var c in a)b[c]=a[c]}window.requireLazy?
window.requireLazy(["Env"],b):(window.Env=window.Env||
{}).b(window.Env)}envFlush({"ajaxpipe_token":"AXjSfGrugGy21VT2ux4","timeslice_heartbeat_config":
{"pollIntervalMs":33,"idleGapThresholdMs":60,"ignoredTimesliceNames":{"requestAnimationFrame":true,"Event listenHandler
mousemove":true,"Event listenHandler mouseover":true,"Event listenHandler mouseout":true,"Event listenHandler
scroll":true},"isHeartbeatEnabled":true,"isArtilleryOn":false},"shouldLogCounters":true,"timeslice_categories":
{"react_render":true,"reflow":true,"sample_continuation_stacktraces":true,"dom_mutation_flag":true,"stack_trace_limit":30,"timesliceBuffer
Size":5000,"show_invariant_decoder":false,"compat_iframe_token":"AQ7PSIPXu_CkJVGO2xQ","isQuick":false});</script><style nonce="Yuhwhab6">
</style><script nonce="Yuhwhab6">_DEV_=0;CavalryLogger=window.CavalryLogger||function(a)
{this.lid=a,this.transition=!1,this.metric_collected=!1,this.is_detailed_profiler=!1,this.instrumentation_started=!1,this.pagelet_metrics=
{}},this.events={},this.ongoing_watch={},this.values={t_cstart:window._cstart},this.piggy_values={},this.bootloader_metrics=
{}},this.resource_to_pagelet_mapping=
{}},this.initializeInstrumentation&&this.initializeInstrumentation(),CavalryLogger.prototype.setIsDetailedProfiler=function(a)
{this.is_detailed_profiler=a;return this},CavalryLogger.prototype.setTTIEvent=function(a){this.tti_event=a;return
this},CavalryLogger.prototype.setValue=function(a,b,c,d){d=d?this.piggy_values:this.values;(typeof d[a]===undefined||c)&&(d[a]=b);return
this},CavalryLogger.prototype.getLastTtiValue=function(){return
this.lastTtiValue},CavalryLogger.prototype.setTimestamp=CavalryLogger.prototype.setTimestamp||function(a,b,c,d){this.mark(a);var
e=this.values.t_cstart||this.values.t_start;e=d?e+d:CavalryLogger.now();this.setValue(a,e,b,c);this.tti_event&&a=this.tti_event&&
(this.lastTtiValue=e,this.setTimestamp("t_tti",b));return this},CavalryLogger.prototype.mark=typeof console===object&&console.timestamp?
function(a){console.timestamp(a):function(){}},CavalryLogger.prototype.addPiggyback=function(a,b){this.piggy_values[a]=b;return
this},CavalryLogger.instances={},CavalryLogger.id=0,CavalryLogger.prototype.getInstance=function(a){typeof a===undefined&&
(a=CavalryLogger.id);CavalryLogger.instances[a]||(CavalryLogger.instances[a]=new CavalryLogger(a));return
CavalryLogger.instances[a]},CavalryLogger.setPageID=function(a){if(CavalryLogger.id===0){var
b=CavalryLogger.getInstance();CavalryLogger.instances[a]=b;CavalryLogger.instances[a].lid=a;delete
CavalryLogger.instances[0]}CavalryLogger.id=a,CavalryLogger.now=function(){return
window.performance&&performance.timing&&performance.timing.navigationStart&&performance.now?
performance.now():performance.timing.navigationStart:new Date().getTime()};CavalryLogger.prototype.measureResources=function()
{CavalryLogger.prototype.profileEarlyResources=function(){}},CavalryLogger.getBootloaderMetricsFromAllLoggers=function()
{CavalryLogger.start_js=function(){}},CavalryLogger.done_js=function()
{}},CavalryLogger.prototype.setTTIEvent("t_domcontent");CavalryLogger.prototype.measureResources=function(a,b)
{if(!this.log_resources)return;var c="bootstrap/"+a.name;if(this.bootloader_metrics[c]!==void 0||this.ongoing_watch[c]!==void 0)return;var
d=CavalryLogger.now();this.ongoing_watch[c]=d;"start_"+c in this.bootloader_metrics||(this.bootloader_metrics["start_"+c]=d);b&&("tag_"+c
in this.bootloader_metrics&&(this.bootloader_metrics["tag_"+c]=b);if(a.type==="js")
{c="js_exec/"+a.name;this.ongoing_watch[c]=d},CavalryLogger.prototype.stopWatch=function(a){if(this.ongoing_watch[a]){var
b=CavalryLogger.now(),c=b-this.ongoing_watch[a];this.bootloader_metrics[a]=c;var d=this.piggy_values;a.indexOf("bootstrap")===0&&
(d.t_resource_download||(d.t_resource_download=0),d.resources_downloaded|
```

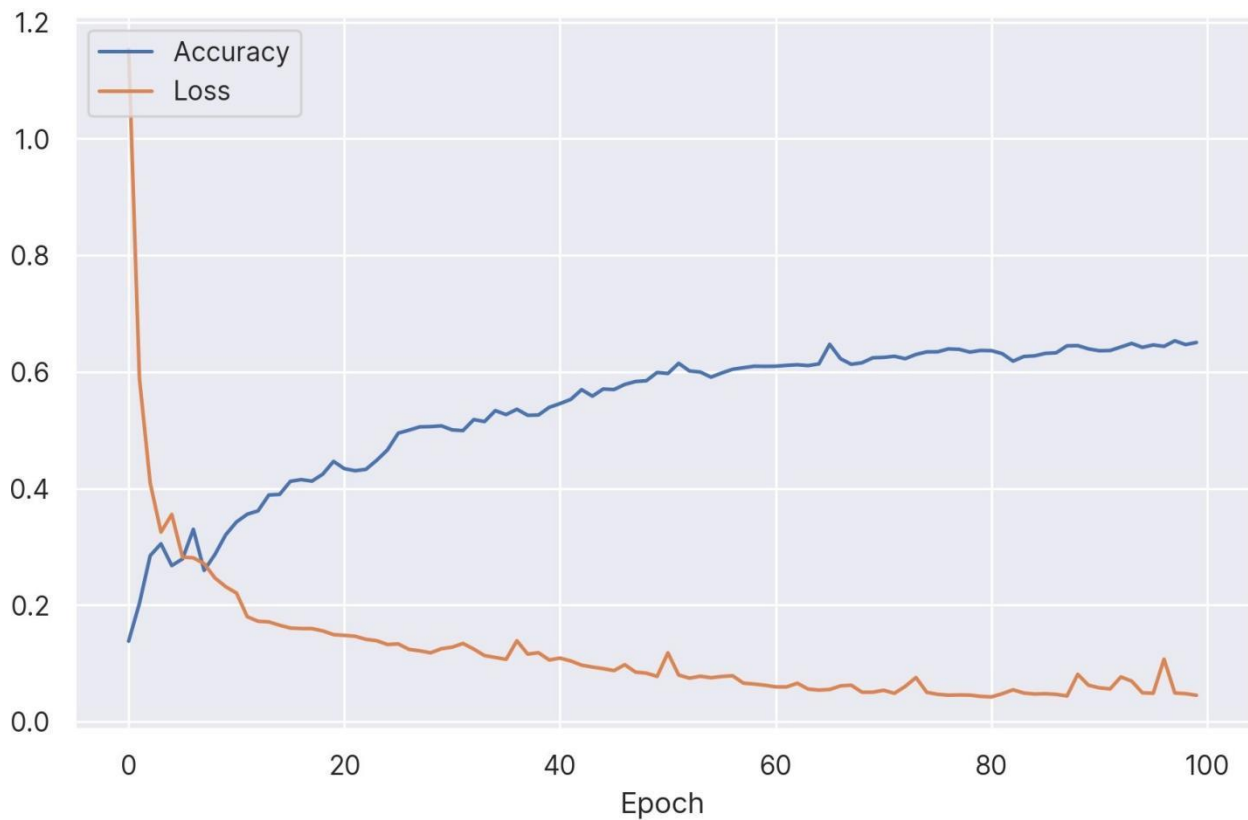
Assigned generalized labels' distribution



Trained TensorFlow Neural Network model architecture



Model fitting plots



Model evaluation report

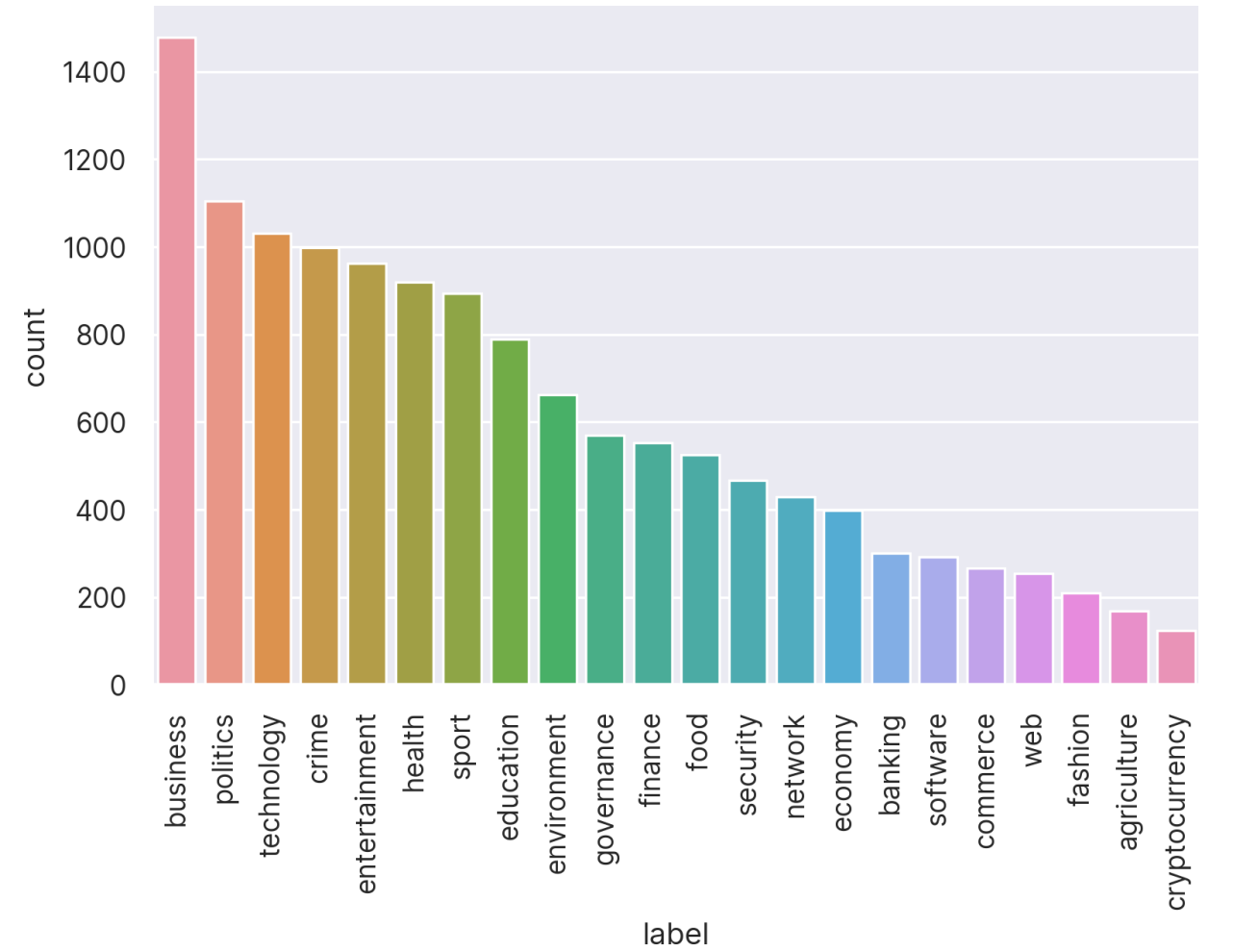
	precision	recall	f1-score	support
agriculture	0.75	0.75	0.75	8
banking	0.28	0.70	0.40	10
business	0.95	0.92	0.94	878
commerce	0.71	0.63	0.67	19
crime	0.44	0.36	0.40	22
cryptocurrency	0.00	0.00	0.00	2
economy	0.68	0.83	0.75	110
education	0.93	0.73	0.82	247
entertainment	0.94	0.65	0.77	201
environment	0.72	0.54	0.61	80
fashion	0.65	0.51	0.57	47
finance	0.76	0.84	0.80	219
food	0.92	0.92	0.92	330
governance	0.95	0.94	0.95	548
health	0.82	0.84	0.83	320
network	0.79	0.64	0.71	131
politics	0.88	0.63	0.74	73
security	0.97	0.94	0.95	677
software	0.66	0.82	0.73	142
sport	0.92	0.87	0.90	94
micro avg	0.88	0.85	0.87	4158
macro avg	0.74	0.70	0.71	4158
weighted avg	0.89	0.85	0.87	4158
samples avg	0.92	0.91	0.91	4158

Drug dataset

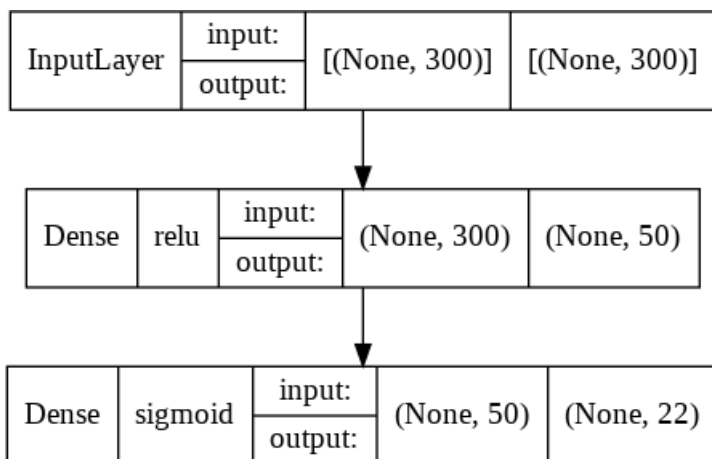
Given: unclean dataset

	body	description	title	url
0	No cookies, no javascript, no trace. We protect your privacy! In Cart In Contact In Escrow In Home Question/Answer In Category	Drugs Store - buy Ecstasy, Cocaine, LSD, Meth.	Cocaine Drugs	http://fc3rhytftqwyroq7p6537fj4knfrmd3pbdwm4sz3mye#4wfz7qd.onion/
1	No cookies, no javascript, no trace. We protect your privacy! In Cart In Contact In Escrow In Home Question/Answer In Category	Drugs Store - buy Ecstasy, Cocaine, LSD, Meth.	Cannabis	http://r3q8nxfmfdx2z3oumaseyfr5qujg4cayzv2ygy6rfsk24bbhugnyd.onion/
2	No cookies, no javascript, no trace. We protect your privacy! In Cart In Contact In Escrow In Home Question/Answer In Category	Drugs Store - buy Ecstasy, Cocaine, LSD, Meth.	Buy Drugs - Ecstasy, LSD, Meth, Cocaine, Cannabis, Seeds Cannabis	http://7rcoowrqedgbyaywnxstypkd48f22akwk73etfu2wemra3ytdhoyd.onion/
3	No cookies, no javascript, no trace. We protect your privacy! In Cart In Contact In Escrow In Home Question/Answer In Category	Drugs Shop - buy cannabis best marijuana hemp.	Cannabis - Seeds Cannabis - Drugs Shop hash	http://hpcphv5q3hiebh6c765isalthe4tr6u5dnd638eo2ic3ytr7c72qd.onion/
4	No cookies, no javascript, no trace. We protect your privacy! In Cart In Contact In Escrow In Home Question/Answer In Category	Drugs Store - buy Ecstasy, Cocaine, LSD, Meth.	DRUGS	http://ren6isw6tkbmphx7eispf7nw6mmbec7wjuyhr77ostyctf5dyd.onion/
...
20245	On the Internet, Forever-Mutating Debates In Related Articles In Outlook '98 an annual survey of the economy, markets and industry In News Highlights of 1997 From CyberTimes Forum In Predictions for the Year Ahead In By AMY HARMON In skin...	NaN	On the Internet, Forever-Mutating Debates	https://archive.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/www.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/library/financial/010598net.html
20246	On the Internet, Forever-Mutating Debates In Related Articles In Outlook '98 an annual survey of the economy, markets and industry In News Highlights of 1997 From CyberTimes Forum In Predictions for the Year Ahead In By AMY HARMON In skin...	NaN	1997 Technology Highlights	https://archive.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/www.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/library/financial/010598techni.html
20247	On the Internet, Forever-Mutating Debates In Related Articles In Outlook '98 an annual survey of the economy, markets and industry In News Highlights of 1997 From CyberTimes Forum In Predictions for the Year Ahead In By AMY HARMON In skin...	NaN	Stock Volatility	https://archive.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/www.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/library/financial/010597stox1.GIF.html
20248	On the Internet, Forever-Mutating Debates In Related Articles In Outlook '98 an annual survey of the economy, markets and industry In News Highlights of 1997 From CyberTimes Forum In Predictions for the Year Ahead In By AMY HARMON In skin...	NaN	Wall Street	https://archive.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/www.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/library/financial/010598stox1.jpg.html
20249	On the Internet, Forever-Mutating Debates In Related Articles In Outlook '98 an annual survey of the economy, markets and industry In News Highlights of 1997 From CyberTimes Forum In Predictions for the Year Ahead In By AMY HARMON In skin...	NaN	Oil Output	https://archive.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/www.nytimesn7cgmftshazwhfgzm37qyb44r64yrb2d3+62d2jsciylid.onion/library/financial/010597ok1.GIF.html
20250 rows x 7 columns				

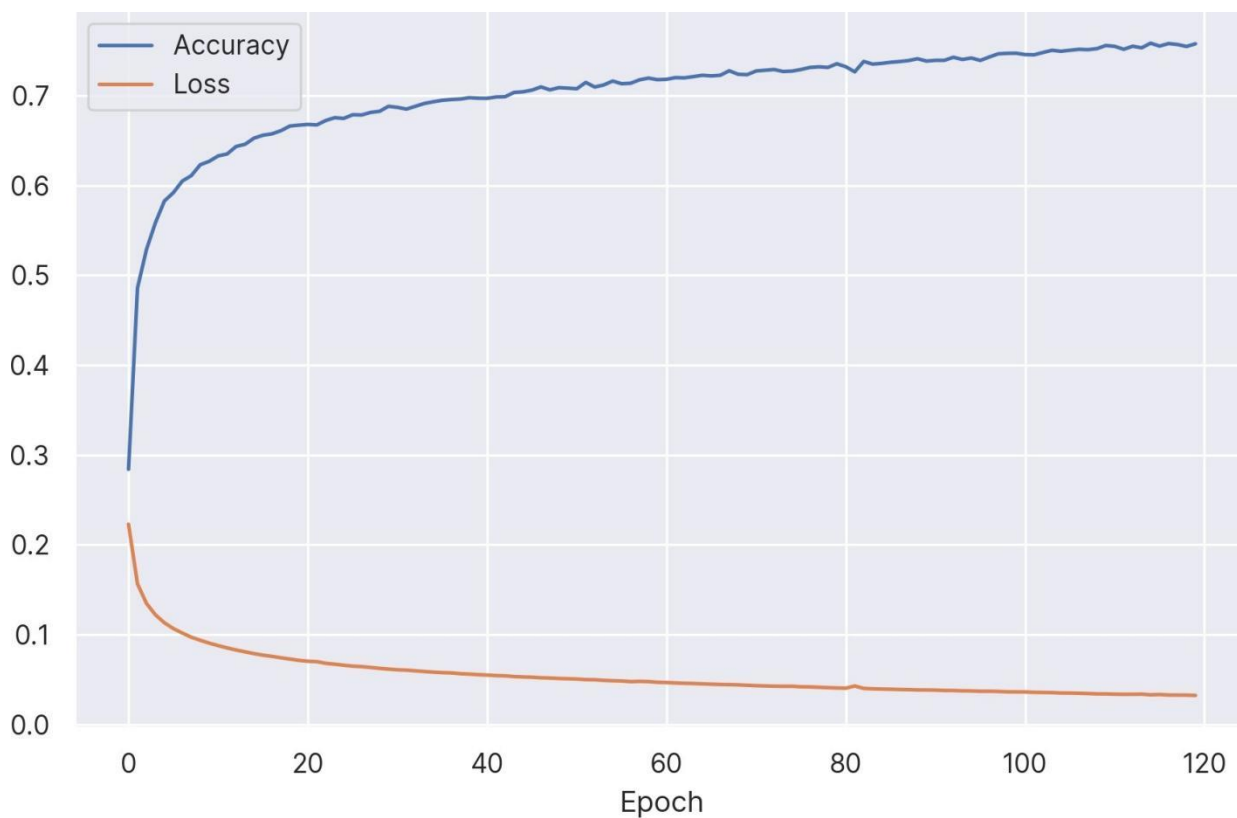
Assigned generalized labels' distribution



Classifier Neural Network model architecture



Model fitting plot



Classifier evaluation report

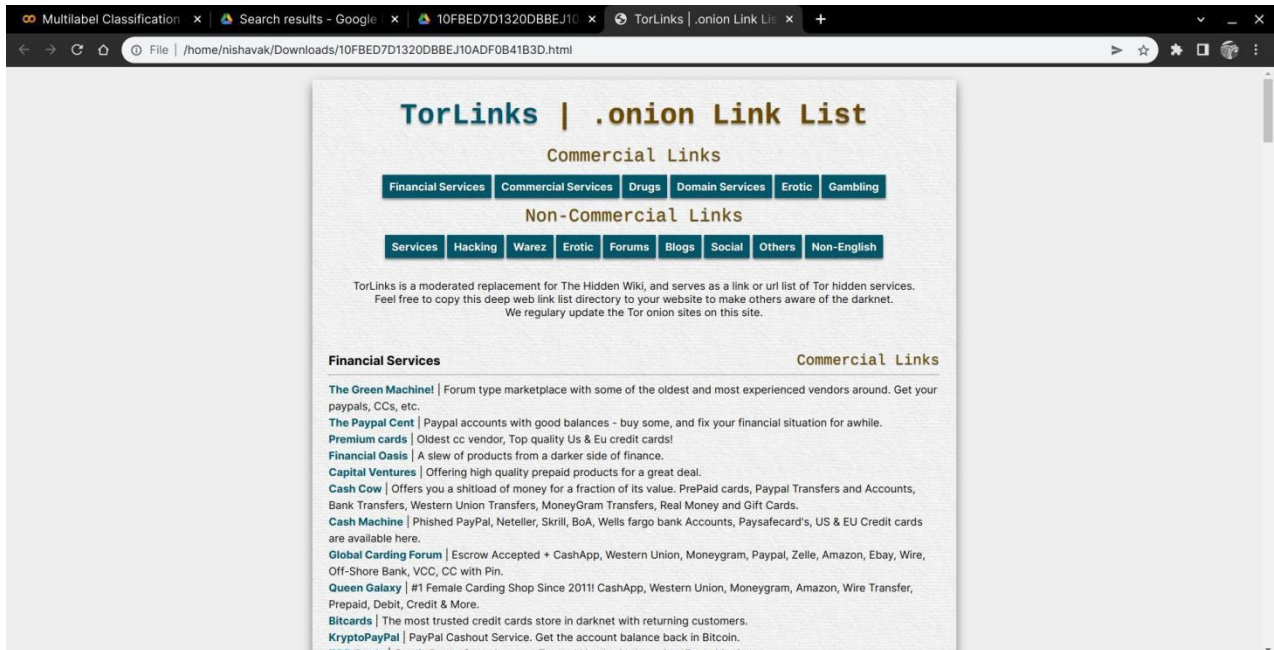
	precision	recall	f1-score	support
agriculture	0.51	0.57	0.54	53
banking	0.83	0.80	0.82	81
business	0.78	0.47	0.58	346
commerce	0.80	0.70	0.75	74
crime	0.81	0.72	0.76	234
cryptocurrency	0.88	0.82	0.85	28
economy	0.62	0.63	0.62	92
education	0.58	0.54	0.56	180
entertainment	0.85	0.64	0.73	242
environment	0.81	0.75	0.78	174
fashion	0.71	0.71	0.71	51
finance	0.74	0.52	0.61	135
food	0.73	0.72	0.73	125
governance	0.65	0.55	0.59	150
health	0.76	0.66	0.71	246
network	0.78	0.75	0.77	106
politics	0.83	0.63	0.72	311
security	0.61	0.65	0.63	104
software	0.41	0.50	0.45	56
sport	0.90	0.82	0.86	234
technology	0.82	0.42	0.55	265
web	0.94	0.88	0.91	56
micro avg	0.76	0.63	0.69	3343
macro avg	0.74	0.66	0.69	3343
weighted avg	0.77	0.63	0.68	3343
samples avg	0.65	0.64	0.63	3343

Explanation for the results

The **body** feature in the given dataset is poorly scraped and contain **ARCHIVE** sections within it, which contains additional unnecessary and irrelevant words. Presence of these words in the data is unseparable and causes the noise which reflects in the results of model evaluation as well as label generation.

Example of working model

1) Page input



Predicted labels – business, network.

2) Page input

Samsungstore

Buy original Samsung products from retailer
without VAT and taxes!

[develop free website](#)

We are back!

Important notice! We are ready to take your orders, delivery only within Europe! Smaller delays can happen during the shipment.

Only original items

We order everything directly from Samsung wholesalers, you will receive only new, original, unopened products

Fast

Fast, safe shipping currently to Europe only. Shipping takes 2-5 days. Shipping is free.

Secure

We are trusted sellers since 2013 on deepweb. We need only your name a delivery address.

Payment

We accept any kind of cryptocurrency.
Escrow is accepted.

MobiRise

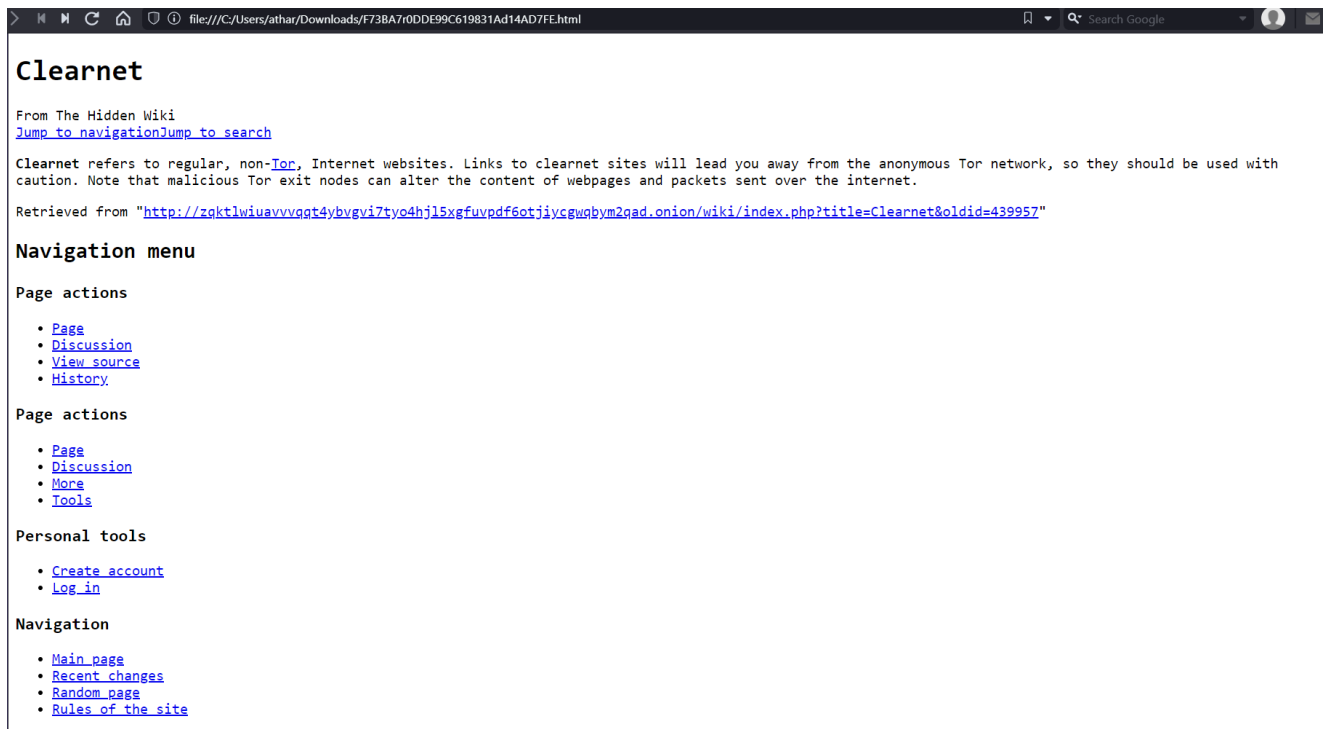
Best selling items*

*Only examples, we can order any other Samsung items

Product	Price (in EUR)
Galaxy S20 4G 128Gb	539 -
Galaxy S20 5G 128Gb	599 -

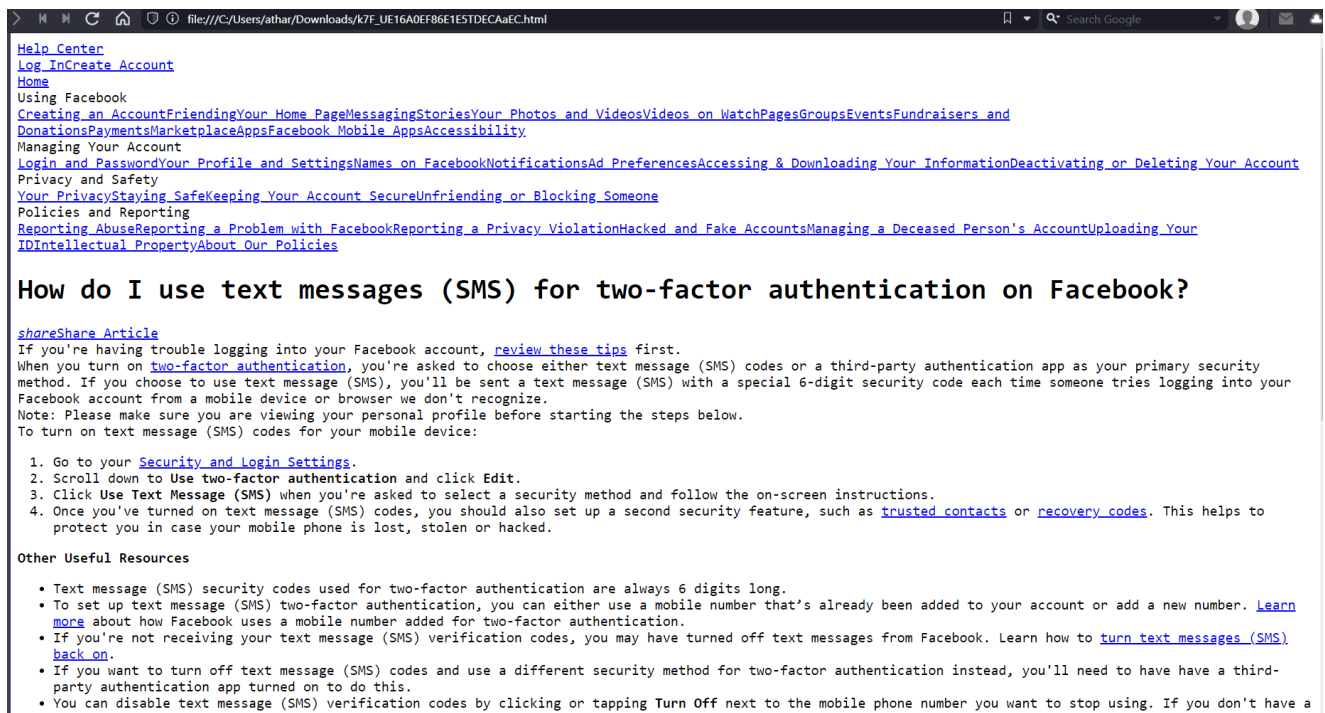
Predicted labels: business

3) Page input



Predicted labels: network

4) Page input



Predicted labels: security

5) Page input

How do I add music to my story on Facebook?

[shareShare Article](#)

To add music to [your story](#), log into the Facebook app for iOS or Android. Adding music to your story isn't available for everyone at this time. When you share a photo or video to your story on Facebook, you may be able to add a clip of a song that will play during your story. You may also see the option to add music lyrics for select songs.

To add music to your story:

1. Open the Facebook app for iOS or Android.
2. Tap + Add to Story at the top of your News Feed.
3. At the top of the screen, scroll to the right and tap Music.
4. Tap to select a song. If you want to display music lyrics, select a song with the LYRICS label next to it.
5. Drag the slider left or right to select a clip from the song, then tap Done.
6. Tap to select a display style (example: Aa) then tap Done. You can choose music lyrics if they're available for your selected song, or different sticker formats with the song name and artist.
7. The song will appear on your story. You can tap to change the style, tap and hold to move it, or use two fingers to rotate it or change the size. To remove the song, drag it to at the bottom of the screen.
8. Tap the circle at the top of the screen (example:) to change the background or add a photo or video from your camera roll.
9. Tap Share to Story.

Note: At this time, music can't be added to text stories.

To choose music before you record a video:

1. Open the Facebook app for iOS or Android.
2. Tap + Add to Story at the top of your News Feed.
3. Tap then tap or MUSIC in the bottom right.
4. Tap to select a song.
5. Drag the slider left or right to select a clip from the song.
6. Tap or Start to record your video.
7. Tap to end your recording. The song clip will appear on your story.
8. Tap Share to Story.

Learn more about [stories on Facebook](#).

Was this information helpful?

☐ Yes

☐ No

Related Articles

[What is subscription linking on Facebook?](#)

Predicted Labels: entertainment

6) Page input

[Help Center](#)

[Log In](#)[Create Account](#)

[Home](#)

Using Facebook

[Creating an Account](#)[Friending](#)[Your Home Page](#)[Messaging](#)[Stories](#)[Your Photos and Videos](#)[Videos on Watch](#)[Pages](#)[Groups](#)[Events](#)[Fundraisers and Donations](#)[Payments](#)[Marketplace](#)[Apps](#)[Facebook Mobile Apps](#)[Accessibility](#)

Managing Your Account

[Login and Password](#)[Your Profile and Settings](#)[Names on Facebook](#)[Notifications](#)[Ad Preferences](#)[Accessing & Downloading Your Information](#)[Deactivating or Deleting Your Account](#)

Privacy and Safety

[Your Privacy](#)[Staying Safe](#)[Keeping Your Account Secure](#)[Unfriending or Blocking Someone](#)

Policies and Reporting

[Reporting Abuse](#)[Reporting a Problem with Facebook](#)[Reporting a Privacy Violation](#)[Hacked and Fake Accounts](#)[Managing a Deceased Person's Account](#)[Uploading Your ID](#)[Intellectual Property](#)[About Our Policies](#)

Donating to a Fundraiser

Learn more about how you can donate to nonprofits and personal fundraisers on Facebook.

About Donations

[Who can create and donate to nonprofit fundraisers?](#)

Personal Fundraiser Review Delay

We have fewer people available to review personal fundraisers because of the coronavirus (COVID-19) outbreak. Because we're also trying hard to prioritize the review of personal fundraisers related to COVID-19, we may be unable to review your fundraiser or it may take longer than usual. Nonprofit fundraisers will not be impacted.

Nonprofit fundraisers are those that directly benefit a charitable organization. You can check whether the feature to create nonprofit fundraisers is available in your country or region below. If it is, you can create fundraisers to benefit charitable organizations. Find out how charities can [qualify for fundraising on Facebook](#).

Receiving the money raised

Money raised for nonprofit organizations is [paid out to the organization](#), not a specific person. If the nonprofit chooses to accept donations on Facebook, we cover fees so that 100% of donations made are paid to the organization.

Donating to fundraisers

To donate, check that:

- You live in a country or region where fundraising on Facebook is available.
- You currently have payments enabled.

Learn how to [start a fundraiser on Facebook](#).

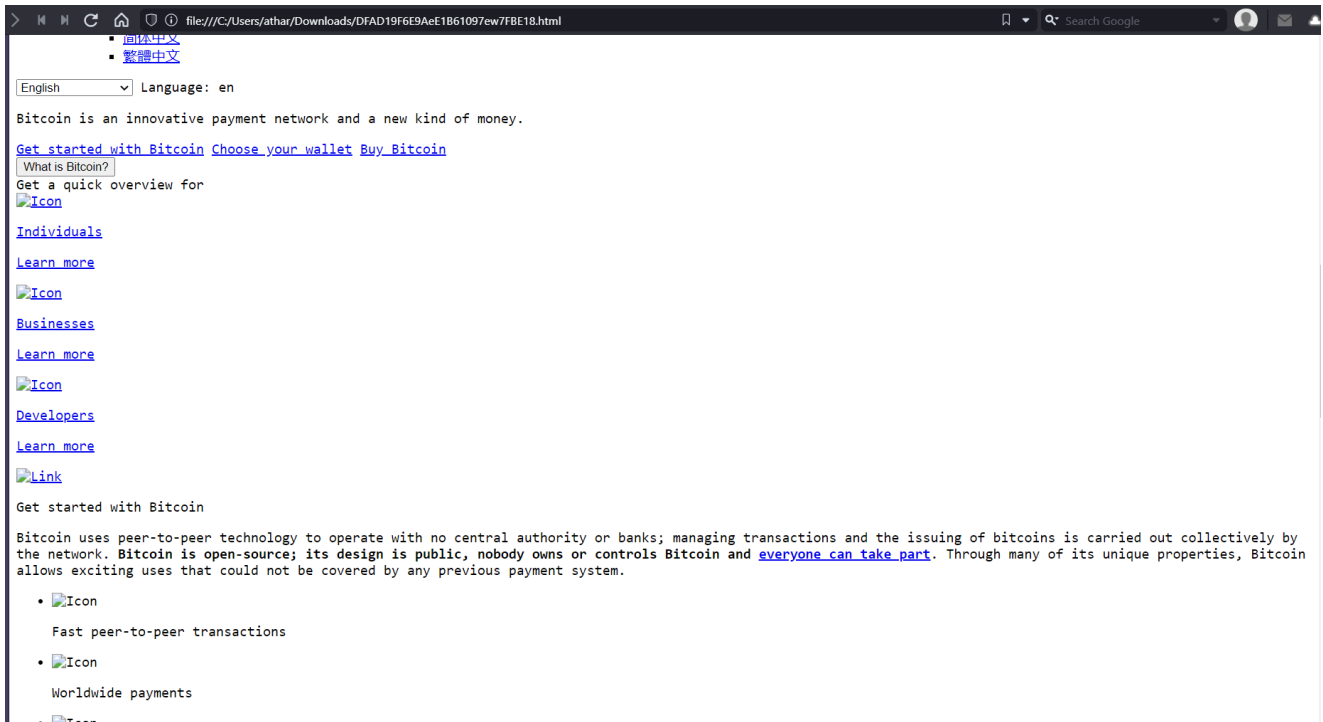
Countries/regions where you can create and make donations to nonprofit fundraisers on Facebook

Country/region	Create nonprofit fundraisers	Donate to nonprofit fundraisers
----------------	------------------------------	---------------------------------

Predicted Labels: health, business

7)

Page input:



Predicted labels: network, finance