

Importation du modèle

```
In [5]: from transformers import AutoTokenizer, AutoModelForSequenceClassification
tokenizer = AutoTokenizer.from_pretrained("nlptown/bert-base-multilingual-uncased-sentiment")
model = AutoModelForSequenceClassification.from_pretrained("nlptown/bert-base-multilingual-uncased-sentiment")
C:\Users\Noura\anaconda3\envs\SDSProject\lib\site-packages\tqdm\auto.py:22: TqdmWarning: IProgress not found. Please update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html
from .autonotebook import tqdm as notebook_tqdm
```

```
In [3]: import pandas as pd
```

Importation des données

```
In [4]: raw_tweets = pd.read_csv("test_clean.csv")
raw_tweets
```

	date	fav_count	favorited	data.id	is_quote_status	language	possibly_sensitive	retweet_count	retweeted	text	user.id	user.name	user.screen_name	user.location
0	Tue Nov 29 15:31:21 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	I take 0 action without strategy	2.904332e+09	Massoud	mxssoud	Ottawa, Ontario
1	Tue Nov 29 15:31:20 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@GuidoDisalle gm Guido!	1.370000e+18	aaronrferguson.eth (he/him) ua Слава Україні!	aaronrferguson	Ottawa, ON Canada
2	Tue Nov 29 15:31:18 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@SleuthieGoosie Absolutely agree with you. Not...	2.993299e+08	Middy T	middyt	Ottawa, Ontario, Canada
3	Tue Nov 29 15:31:16 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@ThisIsKyleR @elonmusk Yes, life must be sooooo...	2.798719e+07	GeeGee23	GeeGee23	Ottawa
4	Tue Nov 29 15:31:13 +0000 2022	0	False	1.600000e+18	False	en	False	0	False	@funstorpaleo So looking forward to seeing the...	7.250000e+17	Michelle Campbell Mekarski PhD	MichelleCbell	Ottawa, Ontario
...
522	Tue Nov 29 23:44:16 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@SenSanders It's spoken like a true mooooch.	1.060000e+18	Rock	1Objectivist	Ottawa, Ontario
523	Tue Nov 29 23:44:15 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@birgitomo I just saw this news. How AWESOME f...	7.570000e+17	Toivi	marktoivanen	Ottawa, Ontario
524	Tue Nov 29 23:44:14 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@the_habitant @HabsOwnSens Austin Snow	1.580000e+18	Cole's 50 goals season	caufieldmania	Ottawa, Canada
525	Tue Nov 29 23:44:14 +0000 2022	0	False	1.600000e+18	False	en	False	0	False	Story #7 is "The Liminal Men" by @ThomasHaWrit...	1.220000e+18	Fusion Fragment	FusionFragment	Ottawa, ON
526	Tue Nov 29 23:44:10 +0000 2022	0	False	1.600000e+18	False	en	NaN	0	False	@SeamusORegan @R_Boissonnault @PascaleStOnge_ ...	1.490000e+18	Samuel Johnson	LastAtTheCentre	Ottawa

527 rows × 14 columns

Colonnes de la base de données

```
In [6]: raw_tweets.columns
```

```
Out[6]: Index(['date', 'fav_count', 'favorited', 'data.id', 'is_quote_status', 'language', 'possibly_sensitive', 'retweet_count', 'retweeted', 'text', 'user.id', 'user.name', 'user.screen_name', 'user.location'], dtype='object')
```

Filtrage des données

```
In [8]: # Cleaning/ filtering tweets
```

```
#Those are tweets from 2022-11-29
# Will only look at data.id, user.id, language, text, user.location, possibly_sensitive, fav_count
tweets = raw_tweets[['data.id', 'user.id', 'language', 'text', 'user.location', 'possibly_sensitive', 'fav_count']].copy()
```

```
tweets
```

Out[8]:	data.id	user.id	language	text	user.location	possibly_sensitive	fav_count
0	1.600000e+18	2.904332e+09	en	I take 0 action without strategy	Ottawa, Ontario	NaN	0
1	1.600000e+18	1.370000e+18	en	@GuidoDisalle gm Guido!	Ottawa, ON Canada	NaN	0
2	1.600000e+18	2.993299e+08	en	@SleuthieGoosie Absolutely agree with you. Not...	Ottawa, Ontario, Canada	NaN	0
3	1.600000e+18	2.798719e+07	en	@ThisIsKyleR @elonmusk Yes, life must be soooo...	Ottawa	NaN	0
4	1.600000e+18	7.250000e+17	en	@funstonpaleo So looking forward to seeing the...	Ottawa, Ontario	False	0
...
522	1.600000e+18	1.060000e+18	en	@SenSanders It's spoken like a true mooth.	Ottawa, Ontario	NaN	0
523	1.600000e+18	7.570000e+17	en	@birgitomo I just saw this news. How AWESOME f...	Ottawa, Ontario	NaN	0
524	1.600000e+18	1.580000e+18	en	@the_habitannt @HabsOwnSens Auston Snow	Ottawa, Canada	NaN	0
525	1.600000e+18	1.220000e+18	en	Story #7 is "The Liminal Men" by @ThomasHaWrit...	Ottawa, ON	False	0
526	1.600000e+18	1.490000e+18	en	@SeamusORegan @R_Boissonnault @PascaleStOnge_...	Ottawa	NaN	0

527 rows × 7 columns

Test de l'algorithme sur un échantillon de la base de donnée

- Échantillon: Les 50 premiers tweets de la base de donnée

Préparation du modèle

```
In [14]: from transformers import pipeline
classifier = pipeline("text-classification", model = "nlptown/bert-base-multilingual-uncased-sentiment")
```

Préparation des données

```
In [62]: text = tweets['text']
```

```
In [63]: test = text.head(50).tolist()
test
```

```
Out[63]: ['I take 0 action without strategy',
 '@GuidoDisalle gm Guido!',
 '@SleuthieGoosie Absolutely agree with you. Not normal.',
 '@ThisIsKyleR @elonmusk Yes, life must be sooooo hard for you. Being able to murder whoever you want isn't enough,... https://t.co/Mzp3Li38s0',
 '@funstonpaleo So looking forward to seeing the results that come out of this!! https://t.co/o9Esvh5PXw',
 '@HugoFaz gm Hugo!',
 "★ Donner en GRAND à votre communauté aujourd'hui ★\nNous lançons Bateaux pour Tous - une campagne pour s'assurer que... https://t.co/gkZVQKP99X",
 "★ Give back to your community in a BIG way today! ★\nWe're launching Boats for All - a campaign to ensure our races... https://t.co/Sdr7UnXfc",
 'y'all wouldn't get it https://t.co/9rzxQRwd06',
 '@WashedSummon OH MY GOODNESS WHAT A BUILD IM ABOUT TO BE FRICKING T'D',
 '@BeFootball Il va rien faire du tout',
 "The problem is that #BC isn't planning to generate enough electricity to meet its #2030ClimateTargets," before eve... https://t.co/Ri7S1j7xfW",
 'The Ottawa Police Services Board abruptly ended its meeting Monday as public delegates held a protest demanding act... https://t.co/WtF4D8qxiD',
 '@alexandrepratte Ton per diem ne te fournira pas 2000 calories par jour 😊',
 "@GraemeNichols I've been of the opinion that the team must know at least something but trading an asset under these... https://t.co/fB2PuvLi2J",
 '@benui thanks, all done in a custom game engine',
 "@118Thumper118 @ElseSlayer What's that got to do with the price of butter? 😊\nBut seriously, yes, yes and yes, except for Scrappy Doo!",
 '@NatashaZimmers As a Canadian student and teacher I know what it is. My daughter denies having heard it called that... https://t.co/oUxMmYpv3t',
 'Pain is starting to subside. https://t.co/B85iuW36lQ',
 '@JenStewartOtt will you help me with a RT? \n\nhttps://t.co/rpv5ikYyWa https://t.co/iy7e0tnm7d',
 '@TheJFreakinC Traumatic head injuries are mandatory to be elected to the Republican Party.',
 'Seeds Canada Members are about to hear all about what we have been busy working on from our senior management team... https://t.co/dAy15557ws',
 'Being part of a community that shares your views on important ... More for Aries https://t.co/bhzrBI9zzh',
 "@DPatel_PharmD And yet his common sense trumps your opponents so-called compassion, why don't you argue the issues... https://t.co/xhg1Bgw4Jd",
 "CORRECTION! Watson est là. Je ne l'avais pas vu.",
 'Pratique des Sénateurs ce matin, Anton Forsberg est présent.',
 'New collaboration between CAGBC & @UofT to research solutions for embodied carbon reductions from building material... https://t.co/LkMR0dkNiv',
 '@jbru11 @MikeDrucker @GeraldLeroy6 He left after the fall of apartheid. So yes, to avoid military service, but not... https://t.co/9UwFgNirNF',
 'Insurance can get complicated. Our latest blog post explains the basics. If you have questions, we can help... https://t.co/0kyHcdT1Wj',
 'Vote for the #LionsCelly for #BudLightCelly of the Week!',
 'Balancing recent family and career commitments could be a chal... More for Virgo https://t.co/GCmNxAND1W',
 '@aureliusraines2 May tomorrow be so boring it almost brings you to tears',
 '@PaulChampLaw You know I agree with this. They could probably hold an entirely separate inquiry on this point thoug... https://t.co/0dhJUT8hAr',
 'Worth reading. https://t.co/Akkunkbfr via @nationalpost',
 "Hi, my name is David Frey. I don't usually endorse other people's products, but in this case I'm making an excepti... https://t.co/r2Ggqidh2v",
 'The SSI journey begins.\n#ottawa @BsidesOttawa #bsidesottawa #identity #SelfSovereignIdentity https://t.co/Je2MYOWQvb',
 '@Lordsweetpotato @StephenPunwasi Lot braver than me looking that far out. Climate is playing a bigger role every ye... https://t.co/FLmZt2ComQ',
 '@shortstack_dan @jamiezjohnson @QuantumNFT gm Dan! Best wishes to your sick computer &lt;3',
 "@tersthebear So so happy this post came to you at the right time! Hope you're feeling a bit better after reading it",
 '@RoughChopOttawa Gotta love the memorial graphic that Sgt Maria Keen (@sgtonpatrol) shared recently. In addition to... https://t.co/5EARkj3J4kB',
 'Do you want to win a NYC Holiday Getaway to see @mariahcarey at Madison Square Garden and $10,000!? \n\nTune in to MO... https://t.co/0Mj9JRhG5V',
 '@Partyof3blog You too, my friend. ❤️',
 '@Sethalos @TorontoStar About what?',
 '@smwgilbert will you help me with a RT? \n\nhttps://t.co/rpv5ikYyWa https://t.co/iy7e0tnm7d',
 'The #K4DM22 Marketplace continues this week with policy-relevant research updates on #ClimateChange, gender equalit... https://t.co/rsvWTvZZSW',
 'Le marché #K4DM22 continue avec des mises à jour de recherches pertinentes pour les politiques sur les changements... https://t.co/KghRg2Syzd',
 '@StephenRwade3 @TrueNorthCentre They fought each other. That convoy should have been way more effective but too ma... https://t.co/ScB9LtDfgD',
 'PODCAST: Wake Up with Rob Snow - November 29, 2022 https://t.co/1sRZR2hSBn',
 '@mariststiles If your budget is dependent on one time development fees to provide service, then your tax rate is too... https://t.co/7abBZj56he',
 "You're probably eager to share a good time with friends today.... More for Aquarius https://t.co/xN5uNaZsre"]
```

Analyse

```
In [64]: raw_list = classifier(test)
```

Traitement des résultats de l'analyse

```
In [65]: raw_result = pd.DataFrame.from_dict(raw_list)
raw_result
```

```
Out[65]:
```

	label	score
0	1 star	0.924041
1	5 stars	0.498549
2	1 star	0.283423
3	1 star	0.315153
4	5 stars	0.428461
5	5 stars	0.478129
6	5 stars	0.445332
7	5 stars	0.677167
8	1 star	0.607453
9	1 star	0.256044
10	5 stars	0.354743
11	1 star	0.324010
12	1 star	0.613772
13	1 star	0.480750
14	4 stars	0.272111
15	5 stars	0.569570
16	3 stars	0.264587
17	4 stars	0.288386
18	2 stars	0.359314
19	1 star	0.242647
20	1 star	0.459253
21	5 stars	0.364008
22	4 stars	0.366059
23	5 stars	0.403067
24	3 stars	0.328802
25	4 stars	0.474071
26	5 stars	0.450861
27	1 star	0.518113
28	4 stars	0.368586
29	5 stars	0.520244
30	3 stars	0.367253
31	1 star	0.583530
32	3 stars	0.287752
33	4 stars	0.462365
34	3 stars	0.298967
35	1 star	0.225022
36	5 stars	0.424655
37	5 stars	0.409506
38	5 stars	0.705695
39	5 stars	0.555300
40	1 star	0.414112
41	5 stars	0.643906
42	1 star	0.399027
43	1 star	0.253801
44	4 stars	0.442195
45	4 stars	0.441337
46	3 stars	0.314041
47	1 star	0.283709
48	1 star	0.330353
49	3 stars	0.363568

```
In [66]: label = {"1 star": 1,  
                 "2 stars": 2,  
                 "3 stars": 3,  
                 "4 stars": 4,  
                 "5 stars": 5}
```

Résultats de l'analyse

```
In [67]: raw_result["label"] = raw_result["label"].map(label)  
raw_result
```

	label	score
0	1	0.924041
1	5	0.498549
2	1	0.283423
3	1	0.315153
4	5	0.428461
5	5	0.478129
6	5	0.445332
7	5	0.677167
8	1	0.607453
9	1	0.256044
10	5	0.354743
11	1	0.324010
12	1	0.613772
13	1	0.480750
14	4	0.272111
15	5	0.569570
16	3	0.264587
17	4	0.288386
18	2	0.359314
19	1	0.242647
20	1	0.459253
21	5	0.364008
22	4	0.366059
23	5	0.403067
24	3	0.328802
25	4	0.474071
26	5	0.450861
27	1	0.518113
28	4	0.368586
29	5	0.520244
30	3	0.367253
31	1	0.583530
32	3	0.287752
33	4	0.462365
34	3	0.298967
35	1	0.225022
36	5	0.424655
37	5	0.409506
38	5	0.705695
39	5	0.555300
40	1	0.414112
41	5	0.643906
42	1	0.399027
43	1	0.253801
44	4	0.442195
45	4	0.441337
46	3	0.314041
47	1	0.283709
48	1	0.330353
49	3	0.363568

Moyenne générale de sentiment

```
In [68]: raw_result['label'].mean()
Out[68]: 3.06
```

Overview

```
In [131...]: raw_df = tweets.head(50).copy()
raw_df = pd.concat([raw_df, raw_result], axis=1)
raw_df
```

	data.id	user.id	language	text	user.location	possibly_sensitive	fav_count	label	score
0	1.600000e+18	2.904332e+09	en	I take 0 action without strategy	Ottawa, Ontario	NaN	0	1	0.924041
1	1.600000e+18	1.370000e+18	en	@GuidoDisalle gm Guido!	Ottawa, ON Canada	NaN	0	5	0.498549
2	1.600000e+18	2.993299e+08	en	@SleuthieGoosie Absolutely agree with you. Not...	Ottawa, Ontario, Canada	NaN	0	1	0.283423
3	1.600000e+18	2.798719e+07	en	@ThisIsKyleR @elonmusk Yes, life must be soooo...	Ottawa	NaN	0	1	0.315153
4	1.600000e+18	7.250000e+17	en	@funstonpaleo So looking forward to seeing the...	Ottawa, Ontario	False	0	5	0.428461
5	1.600000e+18	1.370000e+18	tl	@HugoFaz gm Hugo!	Ottawa, ON Canada	NaN	0	5	0.478129
6	1.600000e+18	1.431515e+08	fr	✖️ Donner en GRAND à votre communauté aujourd'hu...	Ottawa, ON	False	0	5	0.445332
7	1.600000e+18	1.431515e+08	en	✖️ Give back to your community in a BIG way toda...	Ottawa, ON	False	0	5	0.677167
8	1.600000e+18	1.220000e+18	en	y'all wouldn't get it https://t.co/9rzxQRwdO6	Ottawa, Ontario	False	0	1	0.607453
9	1.600000e+18	1.180000e+18	en	@WashedSummon OH MY GOODNESS WHAT A BUILD IM A...	Ottawa, Ontario	NaN	0	1	0.256044
10	1.600000e+18	1.130000e+18	fr	@_BeFootball Il va rien faire du tout	Ottawa, Ontario	NaN	0	5	0.354743
11	1.600000e+18	2.587749e+09	en	"The problem is that #BC isn't planning to gen...	Ottawa	False	0	1	0.324010
12	1.600000e+18	2.121260e+07	en	The Ottawa Police Services Board abruptly ende...	Ottawa, Ontario	False	0	1	0.613772
13	1.600000e+18	1.218876e+08	fr	@alexandrepratt Ton per diem ne te fournira pa...	Gatineau, Québec	NaN	0	1	0.480750
14	1.600000e+18	1.300000e+18	en	@GraemeNichols I've been of the opinion that t...	Ottawa, Ontario	NaN	1	4	0.272111
15	1.600000e+18	3.227156e+08	en	@_benui thanks, all done in a custom game engine	Ottawa, ON	NaN	0	5	0.569570
16	1.600000e+18	2.417086e+07	en	@118Thumper118 @ElseSlayer What's that got to ...	Ottawa, Canada	NaN	0	3	0.264587
17	1.600000e+18	7.166381e+07	en	@NatashaZimmers As a Canadian student and teac...	Ottawa, Ontario	NaN	0	4	0.288386
18	1.600000e+18	3.368386e+09	en	Pain is starting to subside. https://t.co/B85i...	Ottawa, Ontario	False	0	2	0.359314
19	1.600000e+18	1.220000e+18	en	@JenStewartOtt will you help me with a RT? \n...	Ottawa, Ontario	False	0	1	0.242647
20	1.600000e+18	1.643439e+09	en	@TheJFreakinC Traumatic head injuries are mand...	Ottawa ON	NaN	0	1	0.459253
21	1.600000e+18	4.255745e+08	en	Seeds Canada Members are about to hear all abo...	Ottawa, Ontario	False	0	5	0.364008
22	1.600000e+18	2.023695e+08	en	Being part of a community that shares your vie...	Ottawa	False	0	4	0.366059
23	1.600000e+18	1.470000e+18	en	@DPatel_PharmD And yet his common sense trumps...	Ottawa, Ontario	NaN	0	5	0.403067
24	1.600000e+18	2.419062e+07	fr	CORRECTION! Watson est là. Je ne l'avais pas vu.	Ottawa, Ontario	NaN	1	3	0.328802
25	1.600000e+18	8.926511e+07	fr	Pratique des Sénateurs ce matin, Anton Forsber...	Ottawa, ON	NaN	0	4	0.474071
26	1.600000e+18	1.799580e+08	en	New collaboration between CAGBC & @UofT to...	Ottawa	False	0	5	0.450861
27	1.600000e+18	1.298700e+08	en	@jbru11 @MikeDrucker @GeraldLeroy6 He left aft...	Ottawa, Ontario	NaN	0	1	0.518113
28	1.600000e+18	4.805246e+08	en	Insurance can get complicated. Our latest blog... Ottawa, Ontario Canada	False	0	4	0.368586	
29	1.600000e+18	1.570000e+18	en	Vote for the #LionsCelly for #BudLightCelly of...	Ottawa, Ontario	NaN	0	5	0.520244
30	1.600000e+18	1.869026e+07	en	Balancing recent family and career commitments...	Ottawa, Ontario	False	0	3	0.367253
31	1.600000e+18	8.350000e+17	en	@aureliusraines2 May tomorrow be so boring it ...	Ottawa, Canada	NaN	0	1	0.583530
32	1.600000e+18	3.347822e+09	en	@PaulChampLaw You know I agree with this. They...	Ottawa, Ontario	NaN	0	3	0.287752
33	1.600000e+18	7.170000e+17	en	Worth reading. https://t.co/Akkunbxfr via @na...	Ottawa, Canada	False	0	4	0.462365
34	1.600000e+18	5.750402e+06	en	"Hi, my name is David Frey . I don't usually en...	Ottawa	False	0	3	0.298967
35	1.600000e+18	3.857182e+08	en	The SSI journey begins.\n#owasp #ottawa @Bside...	Ottawa	False	1	1	0.225022
36	1.600000e+18	1.490000e+18	en	@lordsweetpotato @StephenPunwasi Lot braver th...	Ottawa, Ontario	NaN	0	5	0.424655
37	1.600000e+18	1.370000e+18	en	@shortstack_dan @jamiezjohnson @QuantumNFT gm ... Ottawa, ON Canada	NaN	0	5	0.409506	
38	1.600000e+18	8.767406e+07	en	@tersthebear So so happy this post came to you...	Ottawa, Ontario	NaN	0	5	0.705695
39	1.600000e+18	2.076640e+08	en	@RoughChopOttawa Gotta love the memorial graph...	Ottawa, Ontario	False	0	5	0.555300
40	1.600000e+18	4.964396e+07	en	Do you want to win a NYC Holiday Getaway to se...	Ottawa	False	0	1	0.414112
41	1.600000e+18	3.710908e+07	en	@Partyof3blog You too, my friend. ❤️	Ottawa, Canada	NaN	0	5	0.643906
42	1.600000e+18	1.576301e+09	en	@Sethalos @TorontoStar About what ?	Ottawa Ontario	NaN	0	1	0.399027
43	1.600000e+18	1.220000e+18	en	@smwgilbert will you help me with a RT? \n\nht...	Ottawa, Ontario	False	0	1	0.253801
44	1.600000e+18	4.006517e+07	en	The #K4DM22 Marketplace continues this week wi...	Ottawa	False	0	4	0.442195
45	1.600000e+18	4.006517e+07	fr	Le marché #K4DM22 continue avec des mises à jo...	Ottawa	False	0	4	0.441337
46	1.600000e+18	1.160558e+08	en	@StephenRWade3 @TrueNorthCentre They fought ea...	Ottawa	NaN	0	3	0.314041
47	1.600000e+18	1.907386e+08	en	PODCAST: Wake Up with Rob Snow - November 29, ...	Ottawa, Ontario	False	0	1	0.283709
48	1.600000e+18	1.580000e+18	en	@maristiles If your budget is dependent on on...	Ottawa	NaN	0	1	0.330353
49	1.600000e+18	1.640973e+08	en	You're probably eager to share a good time wit...	Ottawa, ON	False	0	3	0.363568

Étude des langues des tweets

```
In [132]: raw_language = raw_df['language']  
raw_language.unique()
```

```
Out[132]: array(['en', 'tl', 'fr'], dtype=object)
```

```
In [133]: eng = raw_df[raw_language == 'en']  
len(eng)/len(raw_df)
```

```
Out[133]: 0.86
```

```
In [134]: fr = raw_df[raw_language == 'fr']  
len(fr)/len(raw_df)
```

```
Out[134]: 0.12
```

```
In [135]: other_lang = raw_df[ (raw_language != 'fr') & (raw_language != 'en') ]  
len(other_lang)/len(raw_df)
```

```
Out[135]: 0.02
```

Langues des tweets:

- Anglais (86 %)
- Français (12 %)
- Autres (2%)

In []:

Visualization

Préparation des graphes

In [56]: `import altair as alt`

```
In [164... hist = alt.Chart(raw_df).encode(
    y= alt.Y("mean(label)", title = "Moyenne des labels sentiments"),
    color = alt.Color('language', title = "Langue", scale=alt.Scale(scheme='set2'))

).mark_bar(size=50).facet(
'language:N',
    columns = 3,
    title="Moyenne des sentiments par langage"

)
```

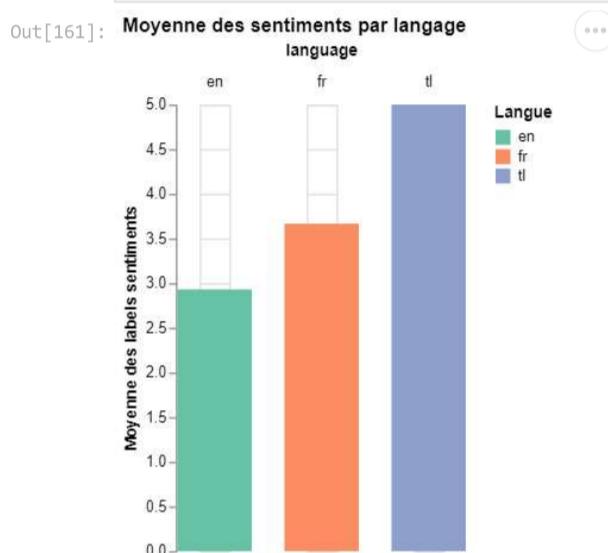
Échelle : Négatif 1 -----5 Positif

In [165... chart2 = alt.Chart(eng, title="Confidence de classification par niveau de positivité").encode(

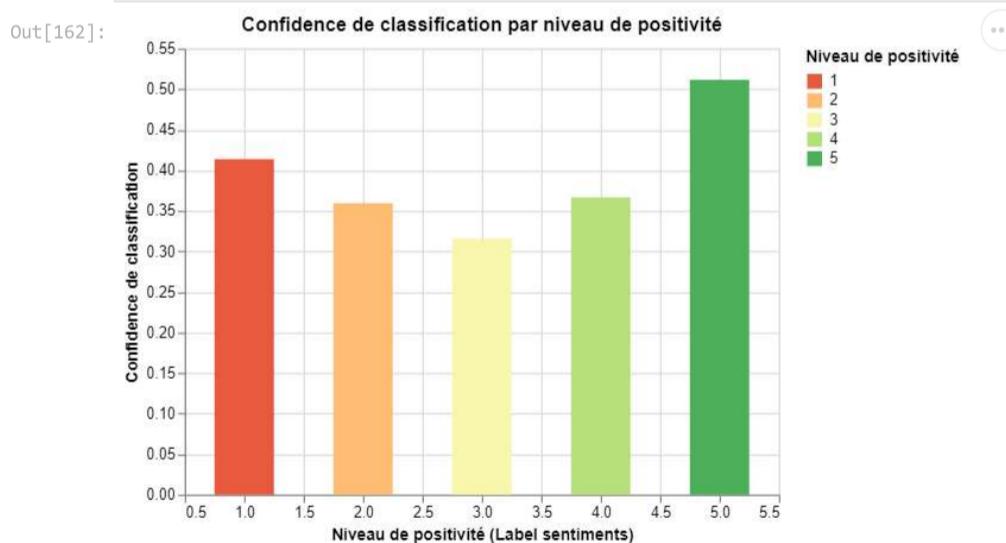
```
    y = alt.Y("mean(score)", title = "Confidence de classification"),
    x= alt.X("label", title = "Niveau de positivité (Label sentiments)"),
    color = alt.Color("label:O", title = "Niveau de positivité", scale=alt.Scale(scheme='redyellowgreen'))
).mark_bar(size=40)
```

Graphes

In [161... hist



In [162... chart2



In []: