enem-variance

January 21, 2024

1 Results - ENEM dataset

Exam codes:

1.1 CH - Humanities

1.2 CN - Natural Sciences

1.3 MT - Math

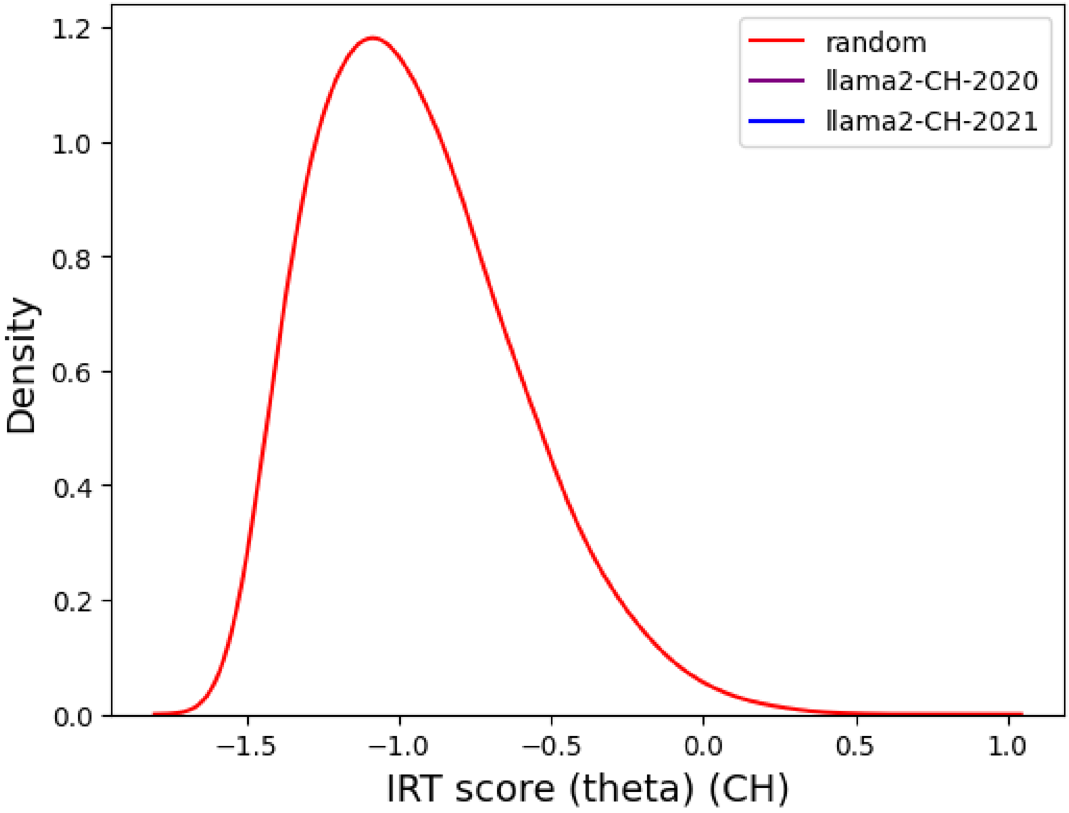
1.4 LC - Languages

```
Loading… CH 2022 mistral simple-zero-shot
Loading… CH 2020 mistral paper-nunes-2023-zero-shot
Loading… CH 2021 mistral paper-nunes-2023-zero-shot
Loading… CH 2022 mistral paper-nunes-2023-zero-shot
Loading… MT 2020 mistral simple-zero-shot
Loading… MT 2021 mistral simple-zero-shot
Loading… MT 2022 mistral simple-zero-shot
Loading… CN 2021 mistral simple-zero-shot
Loading… CN 2022 mistral simple-zero-shot
Loading… LC 2021 mistral simple-zero-shot
Loading… LC 2022 mistral simple-zero-shot
Loading… CH 2020 llama2 simple-zero-shot
Loading… CH 2021 llama2 simple-zero-shot
```

```
mistral CH simple-zero-shot
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
Cell In[6], line 11
      9     print(llm, exam, year)
     10     # curves['random'] = dic_random_scores[llm][exam][year]['CTT_SCORE']
  ↪/ 45
---> 11     curves[f'{year}'] = dic_scores[llm][exam][year]['CTT_SCORE'] / 45
     13 g = sns.kdeplot(
     14     curves,
     15     bw_adjust=2,
     16     common_norm=False,  # Normalize each distribution independently
     17     palette=palette,  # Use palette for multiple colors
)
plt.title(f"{dic_exam_to_paper_name[exam]}")

KeyError: 'CTT_SCORE'
```

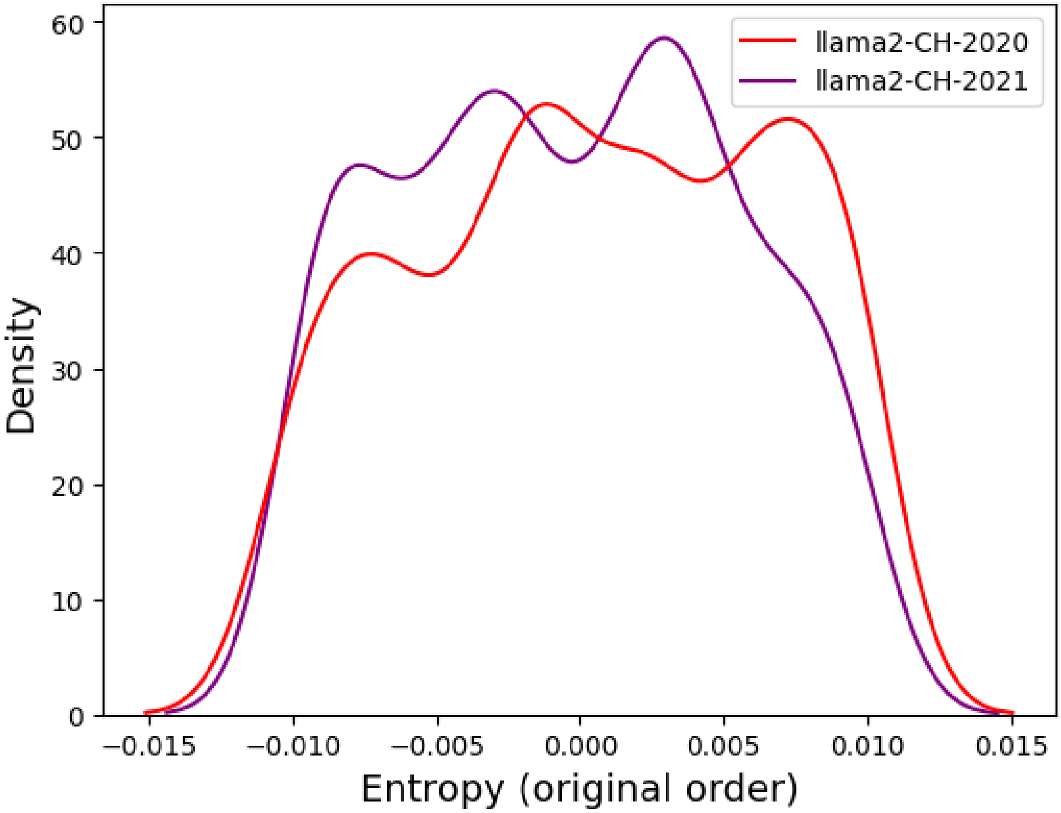
```
C:\Users\pedro\AppData\Local\Temp\ipykernel_560\371188463.py:12: UserWarning:
The palette list has more values (4) than needed (3), which may not be intended.
  g = sns.kdeplot(
C:\Users\pedro\AppData\Local\Temp\ipykernel_560\371188463.py:12: UserWarning:
Dataset has 0 variance; skipping density estimate. Pass `warn_singular=False` to
disable this warning.
  g = sns.kdeplot(
```



```
<Figure size 640x480 with 0 Axes>
```

```
llama2 CH
```

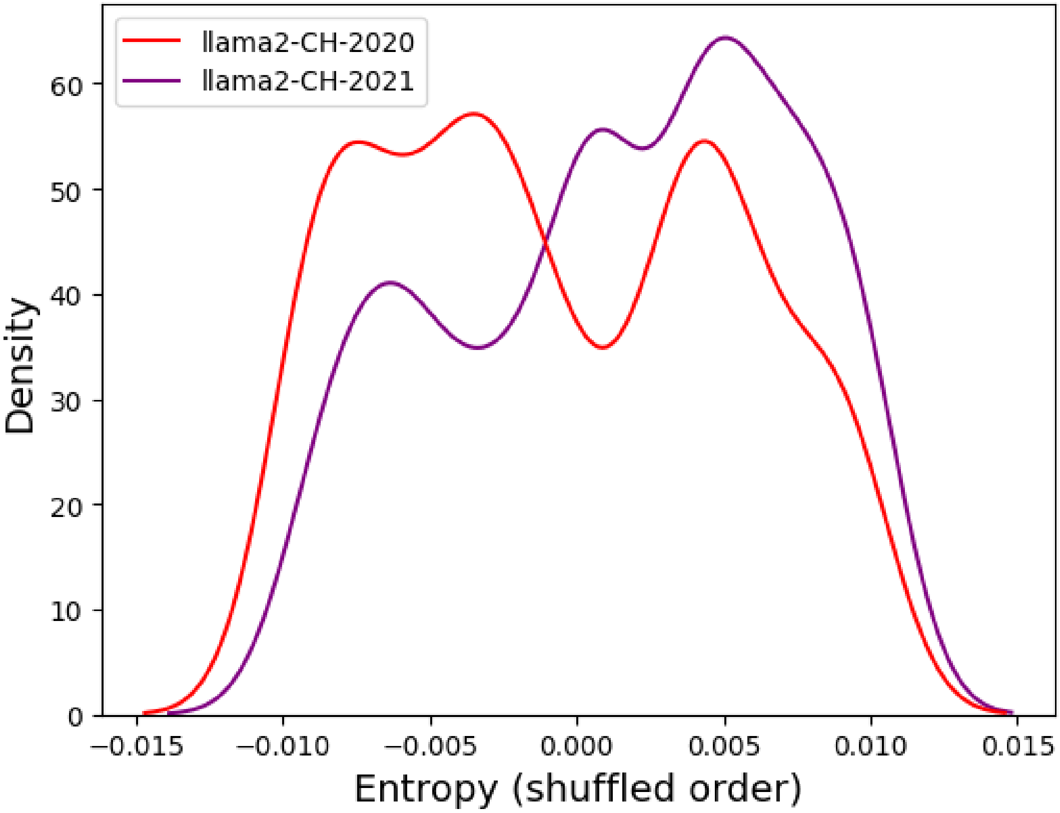
```
C:\Users\pedro\AppData\Local\Temp\ipykernel_560\3526799247.py:13: UserWarning:
The palette list has more values (4) than needed (2), which may not be intended.
  g = sns.kdeplot(
```



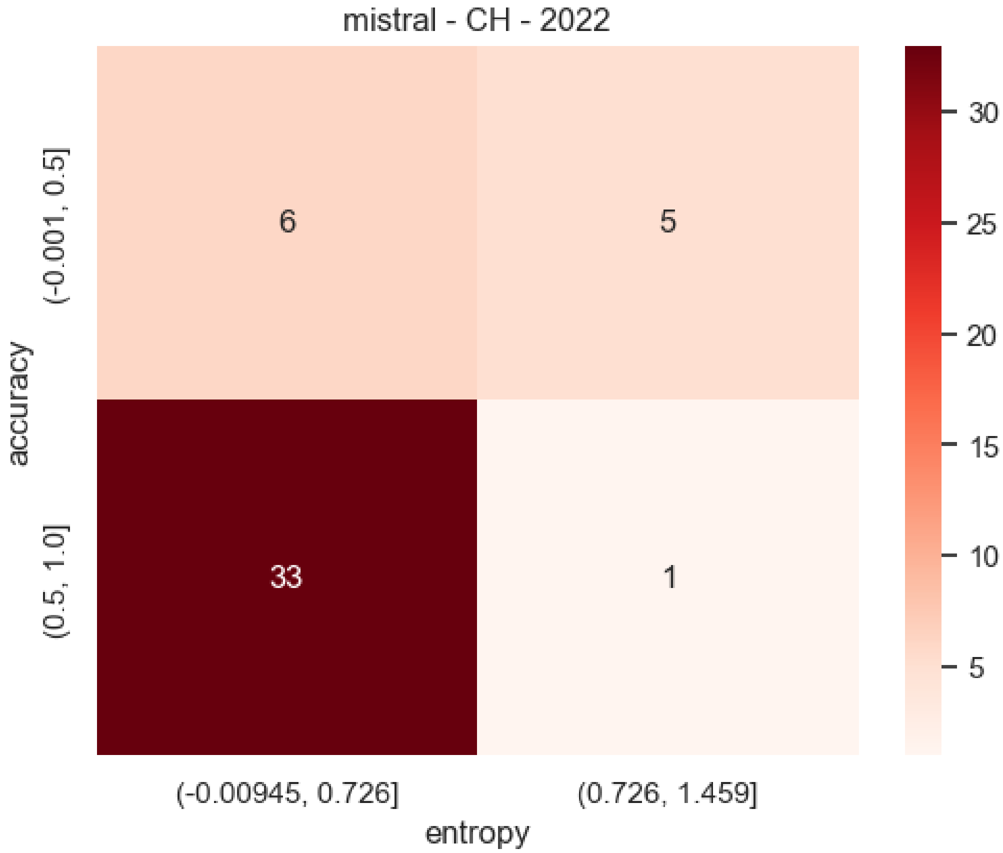
```
<Figure size 640x480 with 0 Axes>
```

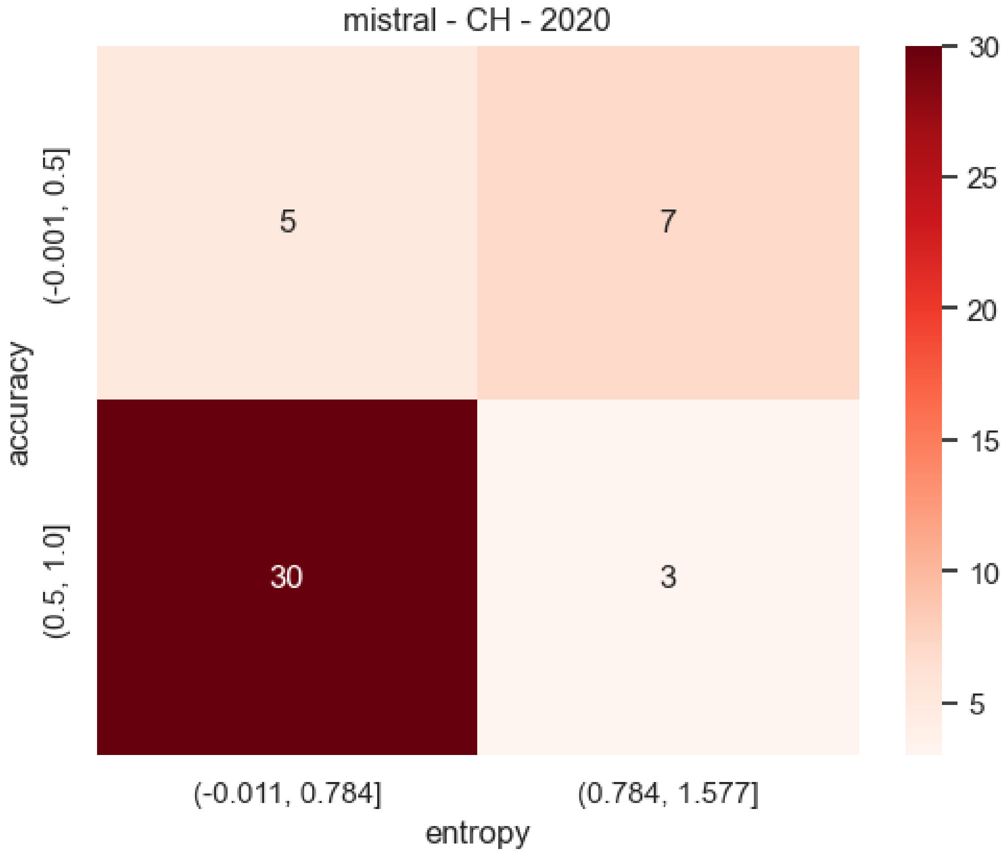
```
llama2 CH
```

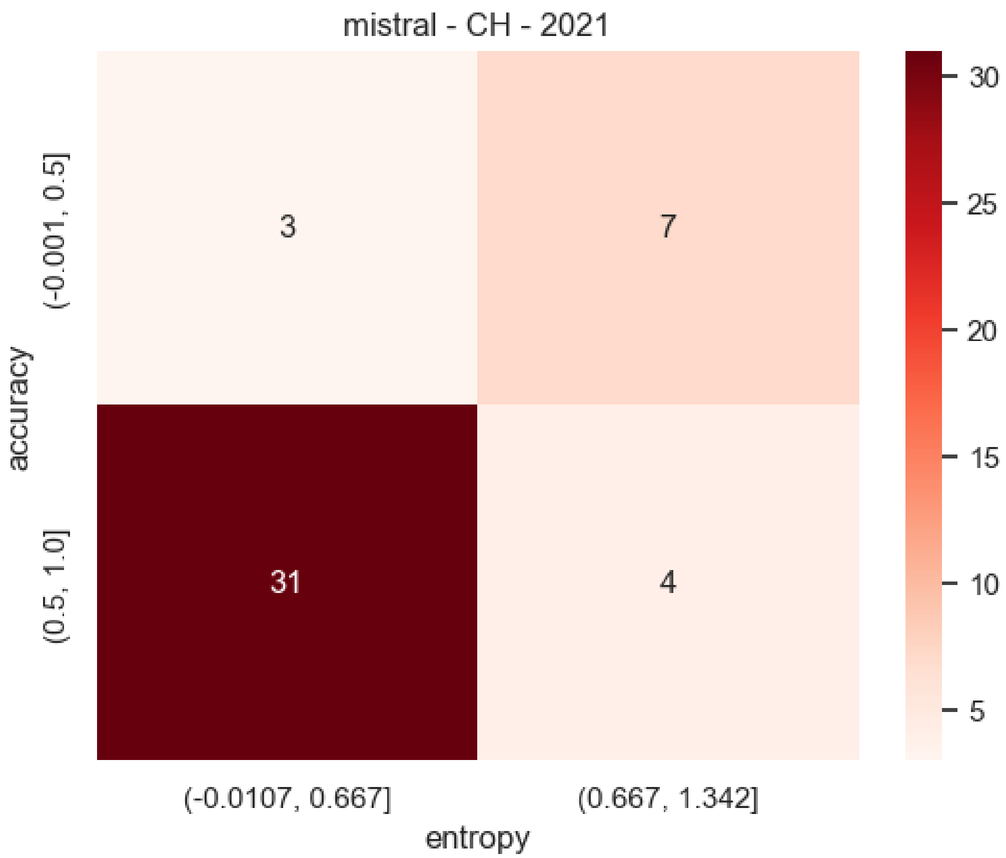
```
C:\Users\pedro\AppData\Local\Temp\ipykernel_560\4259430781.py:13: UserWarning:
The palette list has more values (4) than needed (2), which may not be intended.
  g = sns.kdeplot(
```

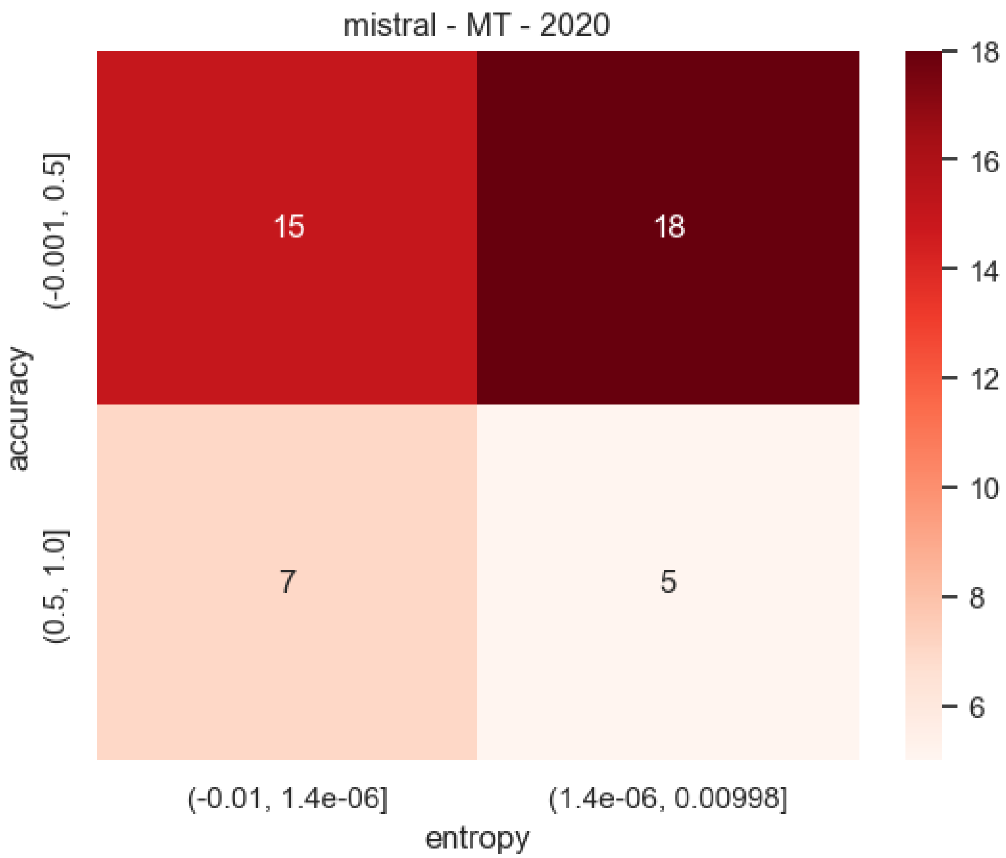

llama2-CH-2020
llama2-CH-2021
Density
0
10
20
30
40
50
60
−0.015
−0.010
−0.005
0.000
0.005
0.010
0.015
Entropy (shuffled order)

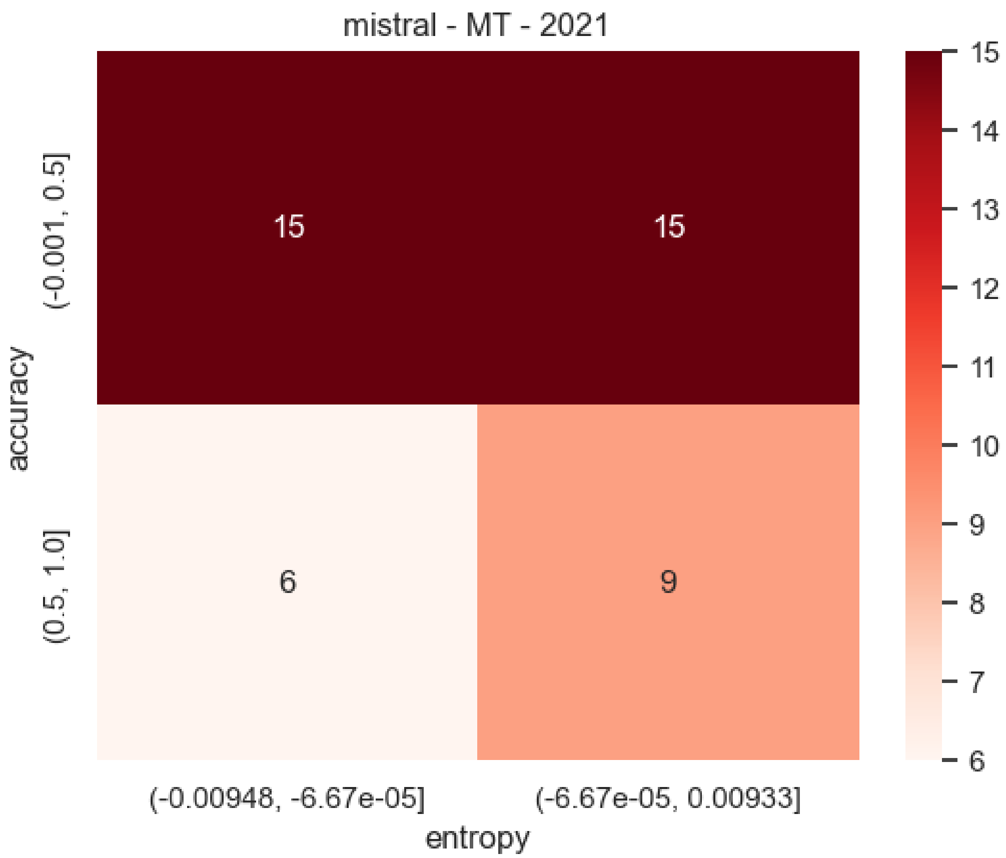
```
<Figure size 640x480 with 0 Axes>
```

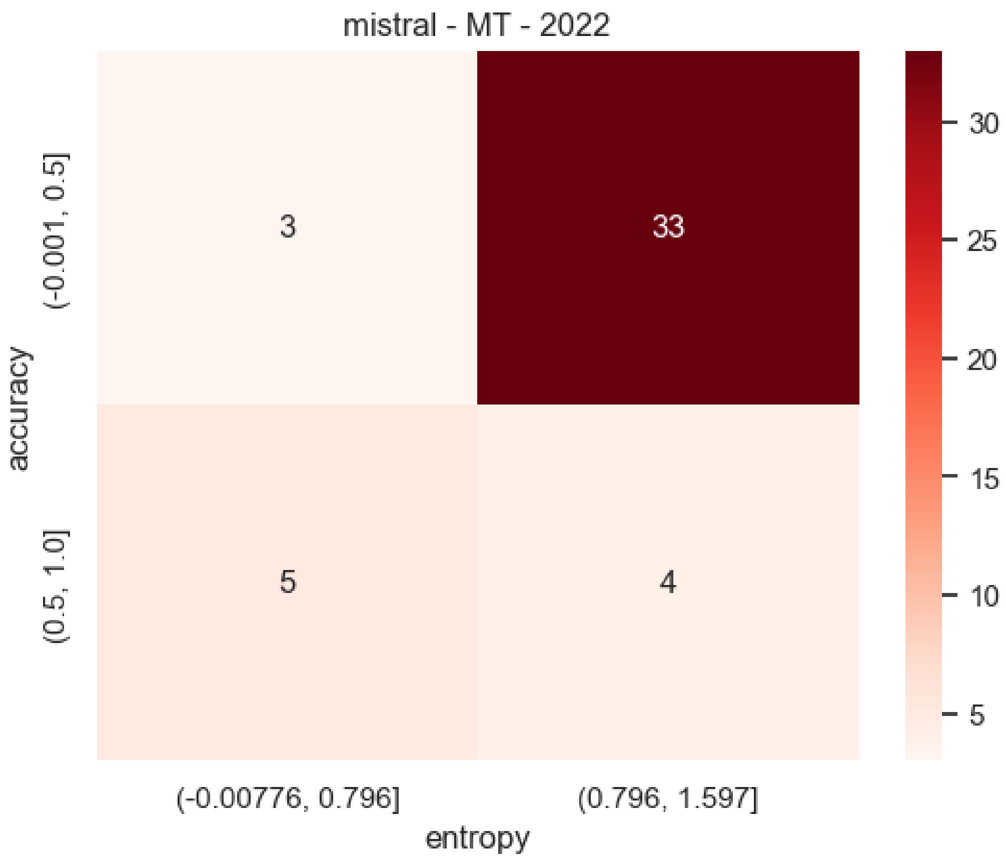

mistral - CH - 2022
accuracy
(-0.001, 0.5]
(0.5, 1.0]
6
5
33
1
(-0.00945, 0.726]
(0.726, 1.459]
entropy
30
25
20
15
10
5

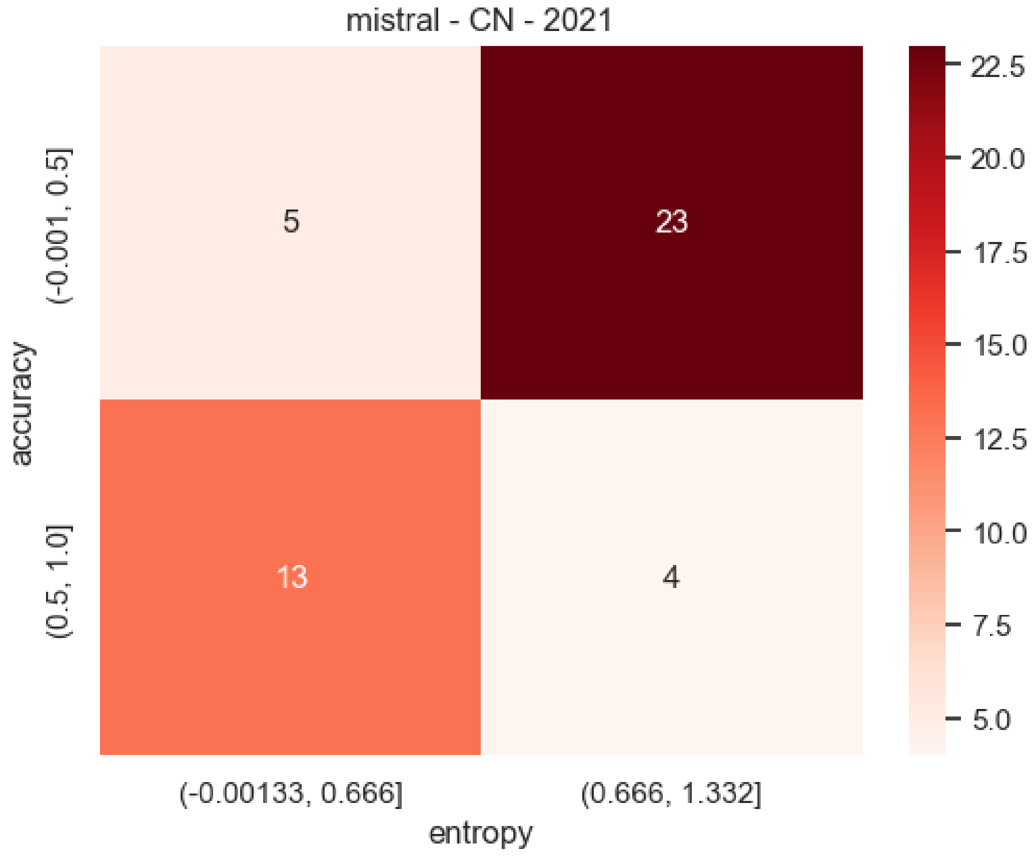

mistral - CH - 2020
accuracy
(-0.001, 0.5]
(0.5, 1.0]
5
7
30
3
(-0.011, 0.784]
(0.784, 1.577]
entropy
30
25
20
15
10
5

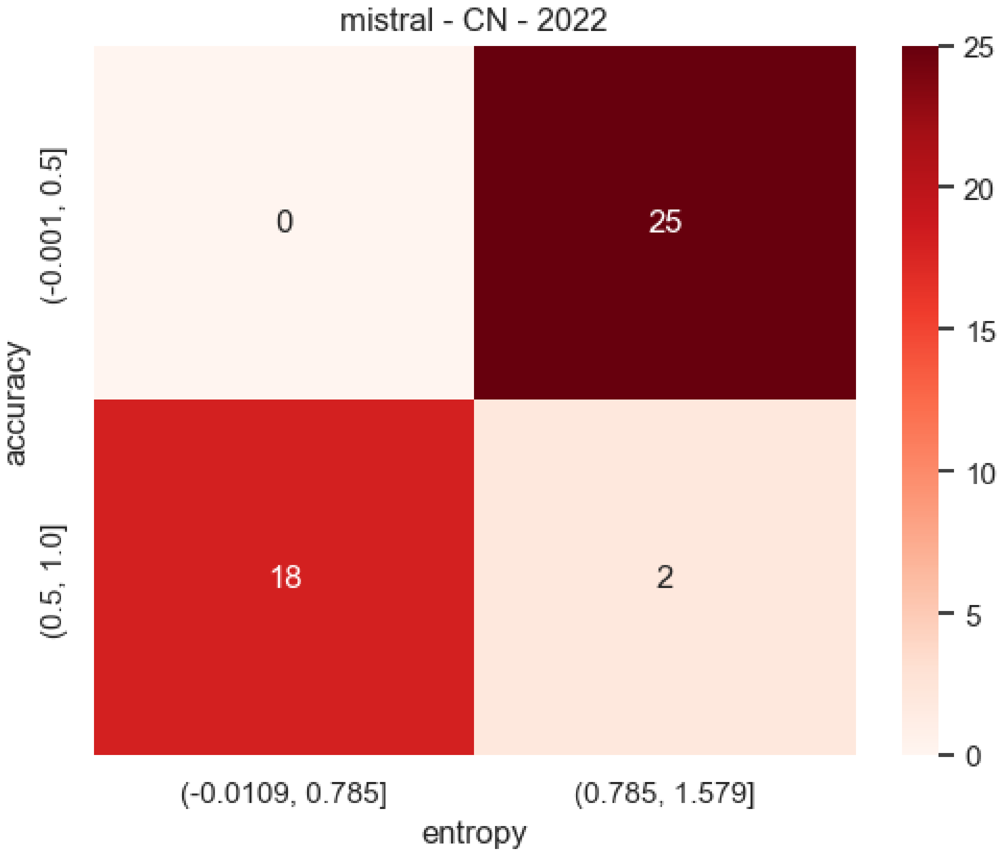

mistral - CH - 2021
accuracy
(-0.001, 0.5]
(0.5, 1.0]
entropy
(-0.0107, 0.667]
(0.667, 1.342]
3
7
31
4
30
25
20
15
10
5

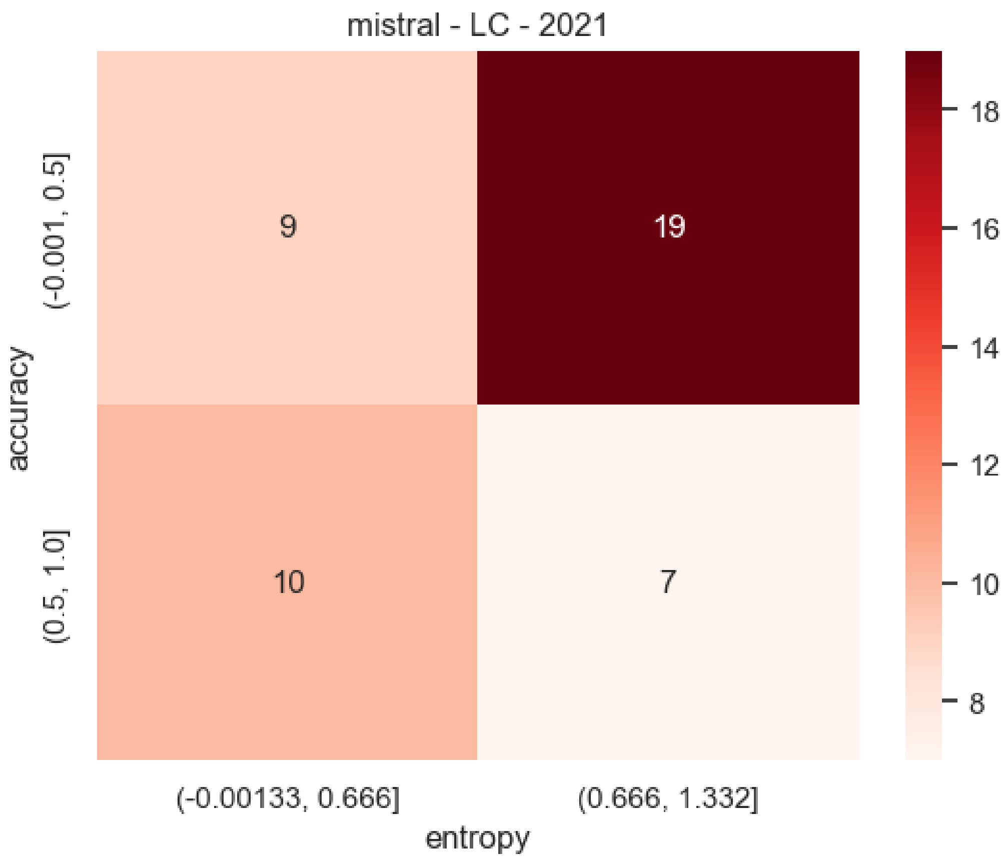

mistral - MT - 2020
accuracy
(-0.001, 0.5]
(0.5, 1.0]
15
18
7
5
(-0.01, 1.4e-06]
(1.4e-06, 0.00998]
entropy
18
16
14
12
10
8
6

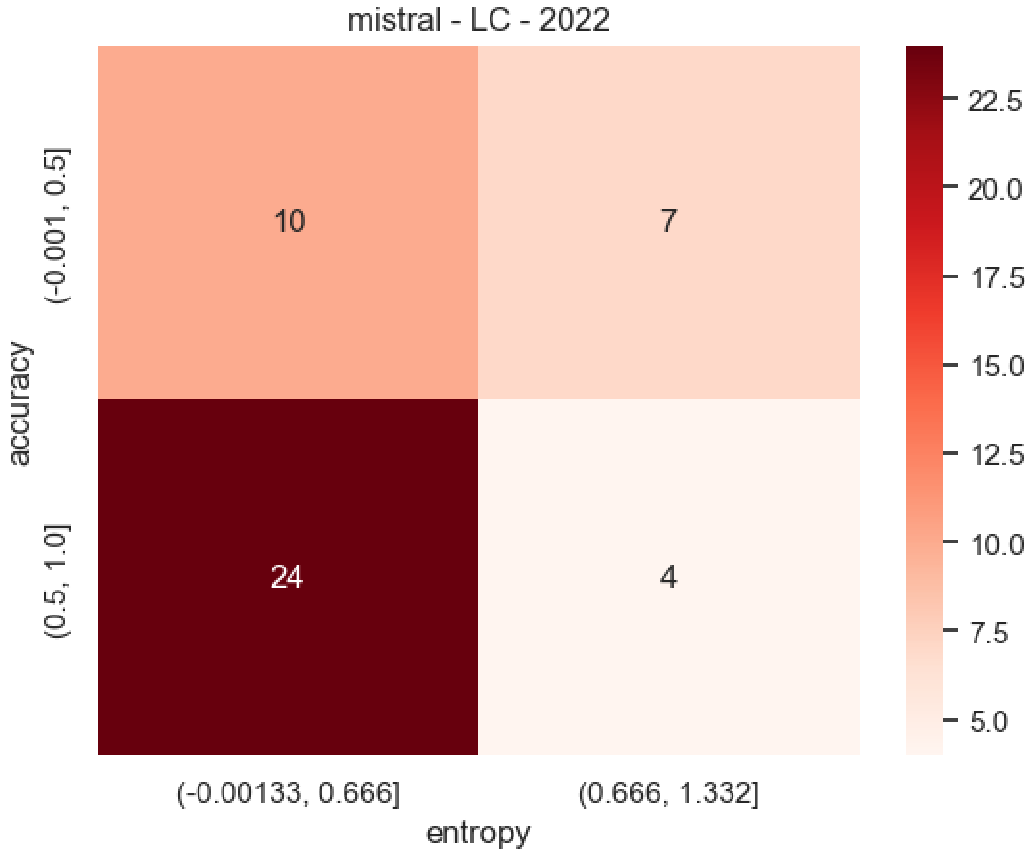

mistral - MT - 2021
accuracy
(-0.001, 0.5]
(0.5, 1.0]
entropy
(-0.00948, -6.67e-05]
(-6.67e-05, 0.00933]
15
15
6
9
15
14
13
12
11
10
9
8
7
6

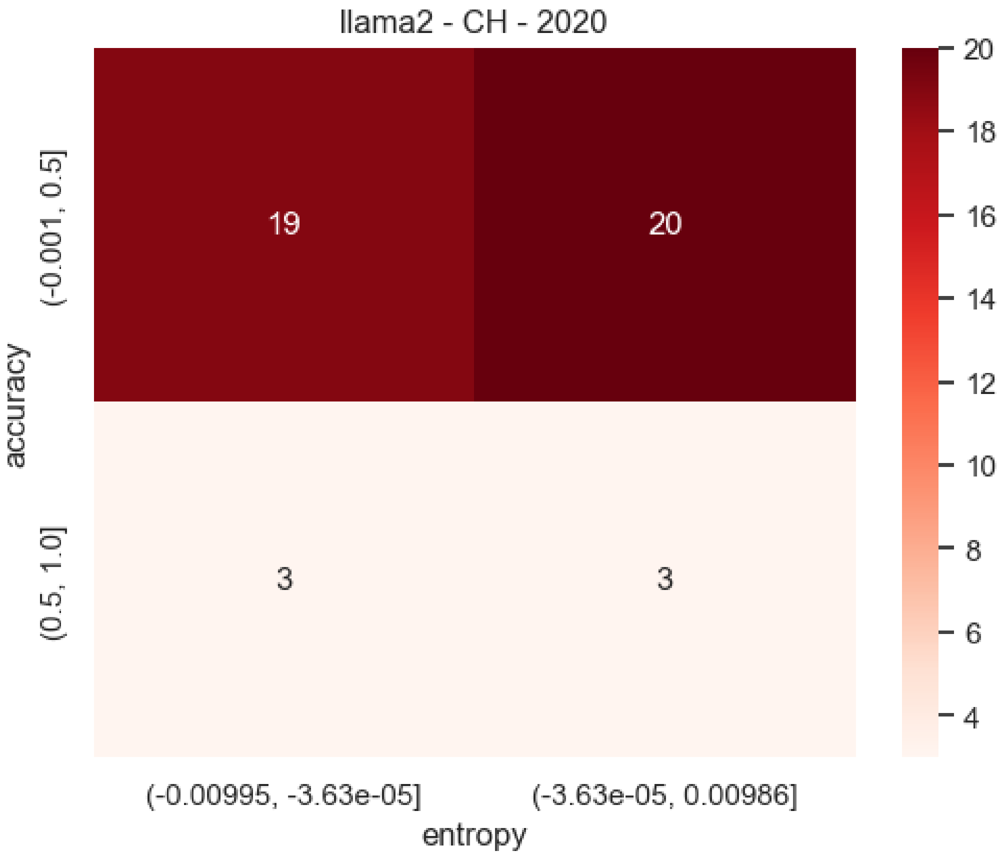

mistral - MT - 2022
accuracy
(-0.001, 0.5]
(0.5, 1.0]
3
33
5
4
(-0.00776, 0.796]
(0.796, 1.597]
entropy
30
25
20
15
10
5

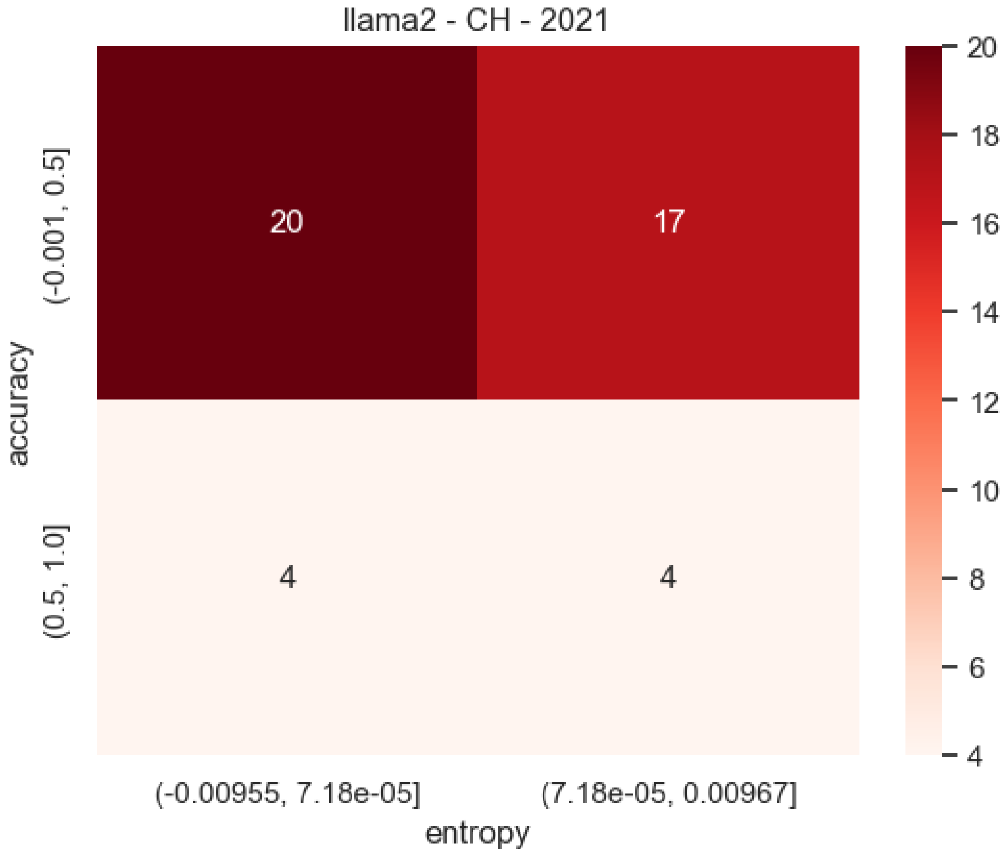

mistral - CN - 2021
accuracy
(-0.001, 0.5]
(0.5, 1.0]
entropy
(-0.00133, 0.666]
(0.666, 1.332]
5
23
13
4
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0


mistral - CN - 2022
accuracy
(-0.001, 0.5]
(0.5, 1.0]
0
25
18
2
(-0.0109, 0.785]
(0.785, 1.579]
entropy
25
20
15
10
5
0


mistral - LC - 2021
accuracy
(-0.001, 0.5]
(0.5, 1.0]
entropy
(-0.00133, 0.666]
(0.666, 1.332]
9
19
10
7
18
16
14
12
10
8


mistral - LC - 2022
accuracy
(-0.001, 0.5]
(0.5, 1.0]
10
7
24
4
(-0.00133, 0.666]
(0.666, 1.332]
entropy
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0

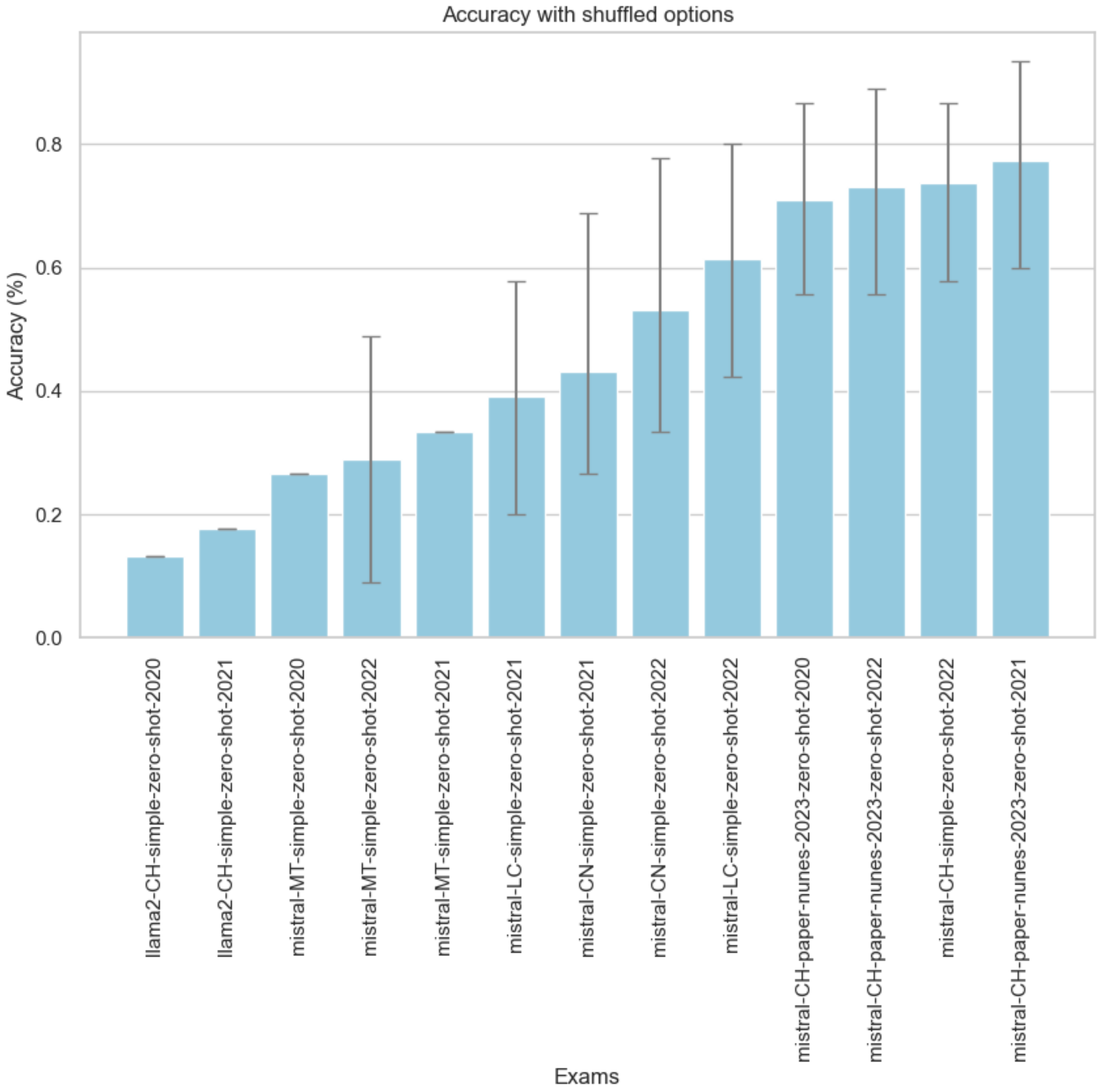

llama2 - CH - 2020
accuracy
(-0.001, 0.5]
(0.5, 1.0]
19
20
3
3
(-0.00995, -3.63e-05]
(-3.63e-05, 0.00986]
entropy
20
18
16
14
12
10
8
6
4



```
1.6094379124341005

1.6094379124341005

1.6094379124341005

1.6094379124341005
```

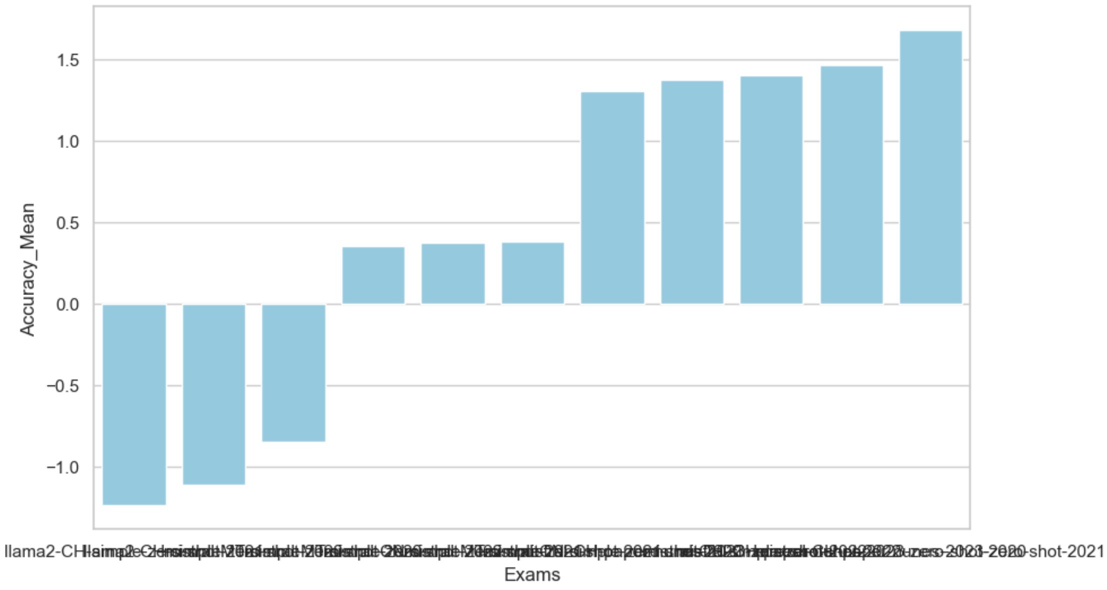
```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
Cell In[12], line 12
     10 for exam in ['CH', 'MT']:
     11     for year in [2020, 2021, 2022]:
---> 12         sns.
 ↪scatterplot(x=dic_itens['mistral'][exam][year]['human_accuracy_75_100'],␣
 ↪y=dic_itens['mistral'][exam][year]['accuracy'])
     13         plt.title(f"human accuracy vs llm accuracy - Mistral {exam}␣
 ↪{year}")
     14         plt.show()
```

```
KeyError: 2020
```



```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[5], line 39
     37 # Create the bar plot with error bars
     38 sns.barplot(x='Exams', y='Accuracy_Mean', data=data, errorbar=None,␣
  ↪color='skyblue')
---> 39 plt.errorbar(x=data['Exams'], y=data['Accuracy_Mean'], 
     40␣
  ↪              yerr=[data['Accuracy_Mean'] - data['Accuracy_Min'], data['Accuracy_Max'] - dat
     41              fmt='none', ecolor='gray', capsize=5)
     43 plt.title('Accuracy with shuffled options')
     44 plt.ylabel('Accuracy (%)')

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\matplotlib\pyplot.py:3030, in errorbar(x, y, yerr, xerr, fmt, ecolor, elinewidth, capsize, barsabove, lolims, uplims, xlolims, xuplims, errorevery, capthick, data, **kwargs)
   3009 @_copy_docstring_and_deprecators(Axes.errorbar)
   3010 def errorbar(
   3011     x: float | ArrayLike,
   (…)
   3028     **kwargs,
   3029 ) -> ErrorbarContainer:
-> 3030     return gca().errorbar(
   3031         x,
   3032         y,
   3033         yerr=yerr,
   3034         xerr=xerr,
   3035         fmt=fmt,
   3036         ecolor=ecolor,
   3037         elinewidth=elinewidth,
   3038         capsize=capsize,
   3039         barsabove=barsabove,
   3040         lolims=lolims,
   3041         uplims=uplims,
   3042         xlolims=xlolims,
   3043         xuplims=xuplims,
   3044         errorevery=errorevery,
   3045         capthick=capthick,
   3046         **({"data": data} if data is not None else {}),
   3047         **kwargs,
   3048     )

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\matplotlib\__init__.py:1478, in _preprocess_data.<locals>.inner(ax, data, *args, **kwargs)
   1475 @functools.wraps(func)
   1476 def inner(ax, *args, data=None, **kwargs):
   1477     if data is None:
-> 1478         return func(ax, *map(sanitize_sequence, args), **kwargs)
   1480     bound = new_sig.bind(ax, *args, **kwargs)
   1481     auto_label = (bound.arguments.get(label_namer)
   1482                   or bound.kwargs.get(label_namer))

File ~\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\matplotlib\axes\_axes.py:3674, in Axes.errorbar(self, x, y, yerr, xerr, fmt, ecolor, elinewidth, capsize, barsabove, lolims, uplims, xlolims, xuplims, errorevery, capthick, **kwargs)
```

```
   3671 res = np.zeros(err.shape, dtype=bool)  # Default in case of nan
   3672 if np.any(np.less(err, -err, out=res, where=(err == err))):
   3673     # like err<0, but also works for timedelta and nan.
-> 3674     raise ValueError(
   3675         f"'{dep_axis}err' must not contain negative values")
   3676 # This is like
   3677 #     elow, ehigh = np.broadcast_to(…)
   3678 #     return dep - elow * ~lolims, dep + ehigh * ~uplims
   3679 # except that broadcast_to would strip units.
   3680 low, high = dep + np.vstack([-(1 - lolims), 1 - uplims]) * err

ValueError: 'yerr' must not contain negative values
```


Accuracy_Mean
llama2-CH
shot-2021
Exams