

آشنایی با ساختار داده‌ای که در اختیارمان قرار داده شده است :

```
1 trainData = pd.read_csv('Dataset/train.csv')
2 print(trainData)
```

	label	content
0	...	فناوری گزارش های منتشر شده حاکی از آن است که کاربران
1	...	ورزشی سوپر استار سینما و از قهرمانان سابق ووشو — کو
2	...	حوادث مدیرعامل شرکت عمران آب کیش از فوت یک نفر در آت
3	...	فناوری یک نوجوان انگلیسی به اتهام هک حساب های کاربری
4	...	سلامت دانشمندان در جدیدترین مطالعات خود اثرات جدید و
...
8995	...	حوادث شانزدهم مهر امسال در پی تماس با مرکز فوریت های
8996	...	ورزشی اتحادیه فوتبال ترینیداد و توباگو اعلام کرده تی
8997	...	فرهنگی/هنری اسامی 18 فیلم بخش مسابقه مستند ششمین دوره جشنو
8998	...	حوادث کامیونت حامل بار تخته های چوبی عصر دیروز چهارش
8999	...	فناوری چهارمین لیگ بازی های رایانه ای ایران از روز ۶

پس همانطور که مشاهده می شود هر سطر از دیتاست شامل یک عنوان یا برچسب خبری و محتوای مربوطه است.

فاز اول : پیش پردازنده

در ابتدا تأثیر stemming و lemmatization را نشان می دهیم. به طور کلی در stemming کلمات با ریشه خود جایگزین می شوند و در lemmatization کلمات با معانی مشابه و کلمات هم خانواده در یک گروه قرار می گیرند.

مثالی از تأثیر lemmatization :

```
1 lemmatizer = Lemmatizer()
2 stemmer = Stemmer()
3
4 sampleWords = ['رفتم', 'رفت', 'رفتند', 'رفتی']
5
6 for word in sampleWords :
7     print(lemmatizer.lemmatize(word))
```

رفت#رو
رفت#رو
رفت#رو
رفت#رو

مثالی از تأثیر stemming :

```

1  lemmatizer = Lemmatizer()
2  stemmer = Stemmer()
3
4  sampleWords = ['رفتم', 'رفت', 'رفتند', 'رفتی']
5
6  for word in sampleWords :
7      print(stemmer.stem(word))

```

رف
رف
رفتند
رفت

در گام پیش پردازش برای تمیز کردن متن، ابتدا متن را normalize میکنیم (در این صورت به متن نیم فاصله افزوده خواهد شد). در ادامه هر کاراکتر غیر حروف الفبا را از متن حذف میکنیم. سپس کلمات متن را جدا کرده و آن‌ها را lemmatize میکنیم و در انتها هر کلمه‌ای که در stopwords خود کتابخانه hazm باشد را از متن حذف میکنیم:

```

1  normalizer = Normalizer()
2  stopWords = stopwords_list()
3  punctuations = ['(', ')', '"', "'", '.', ',', ':', ';', '!', '?', '»', «', '{', '}', '[', ']']
4
5  def preprocessor(text) :
6      tmpTxt = []
7      #delete numbers
8      for letter in text :
9          if not letter.isdigit() :
10             tmpTxt.append(letter)
11      text = ''.join(tmpTxt)
12      #using semi space
13      text = normalizer.normalize(text)
14      words = word_tokenize(text)
15      words = [word for word in words if word not in punctuations]
16      words = [lemmatizer.lemmatize(word).split('#', 1)[0] for word in words]
17      words = [word for word in filter(None, words) if word not in stopWords]
18      return words

```

فاز دوم : فرآیند مسأله

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)}$$

x : مجموعه‌ای از کلمات دیده شده در متن

c : عنوان هر خبر (*lable*)

$p(c|x)$: با فرض این که مجموعه کلمات x دیده شده‌اند، چقدر احتمال دارد عنوان خبری مربوط به این کلمات را عنوان خبری c پیش‌بینی کنیم.

$p(x|c)$: با فرض این که عنوان خبری، c باشد، چقدر احتمال دارد مجموعه کلمات x دیده شوند.

$p(c)$: احتمال آن که عنوان یک خبر c باشد، که این احتمال در حالت کلی برابر تعداد تکرارهای عنوان c به تعداد تکرارهای تمام عناوین خبری است.

$p(x)$: احتمال دیده شدن مجموعه کلمات x در عناوین خبری.

در ادامه توسط دو تابع زیر به محاسبه این احتمالات می‌پردازیم :

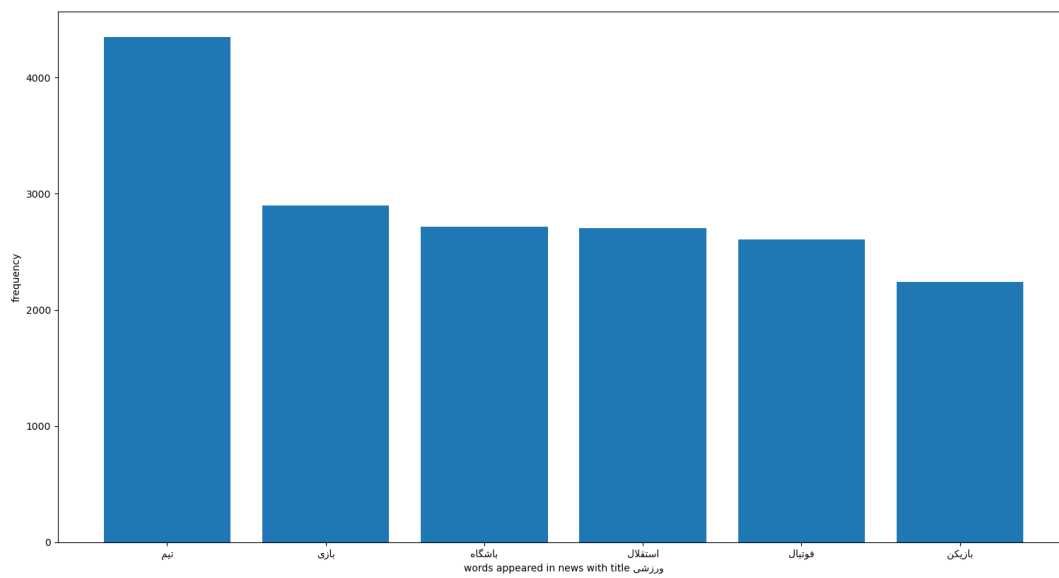
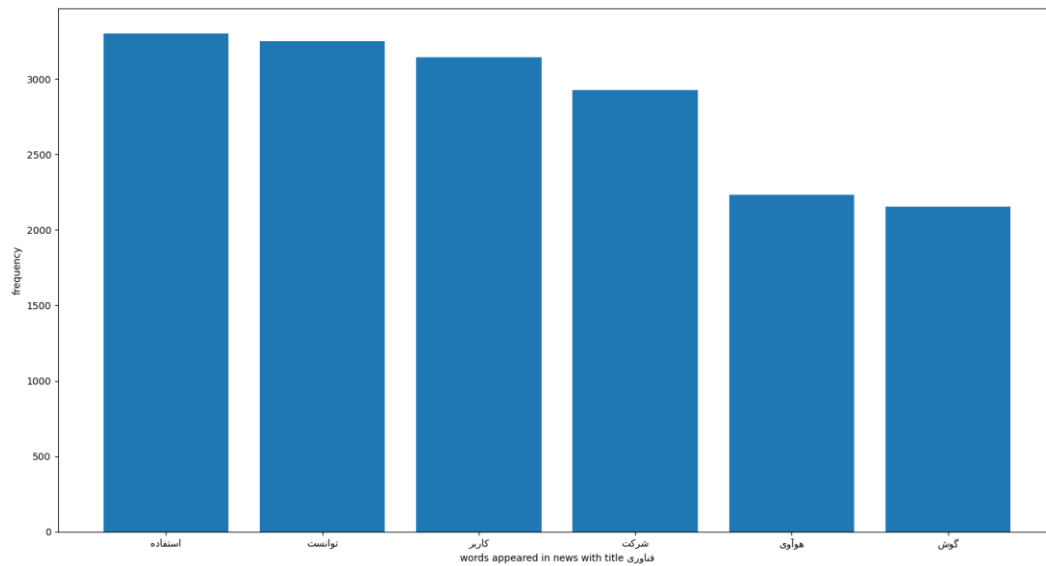
```
1 titles, wordsCount = {}, {}
2 #titles is a dictionary that its keys are titles
3 #and its values are count of each title
4 # wordsCount shows count of each word in all data
5 #wordsCountInTitles shows count of each word for each title
6 totalWordsCount, distinctWordsCount = 0, 0
7 wordsCountInTitles, totalWordsOfTitles, seenWords = {}, {}, {}
```

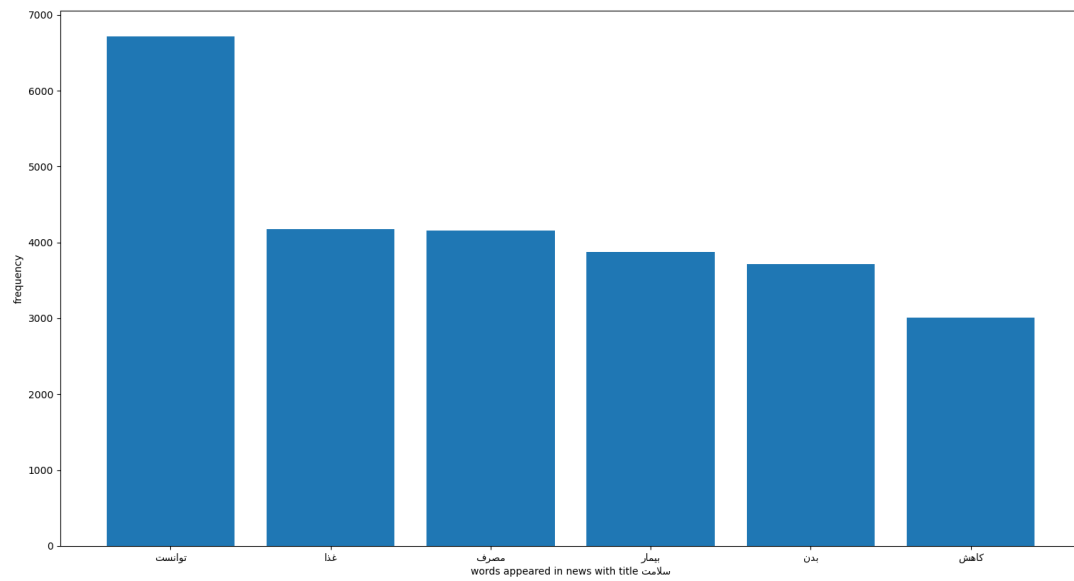
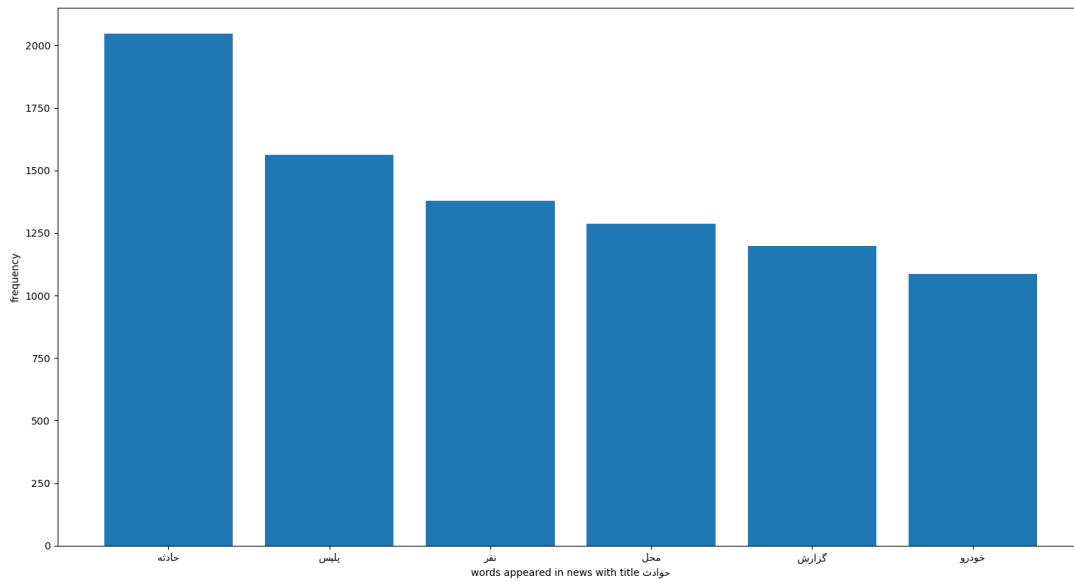
```
1 def calc_titles_frequency() :
2     for _, row in trainData.iterrows() :
3         if row.isnull().values.any() :
4             continue
5         title = row['label']
6         if title in titles :
7             titles[title] += 1
8         else :
9             titles[title] = 1
```

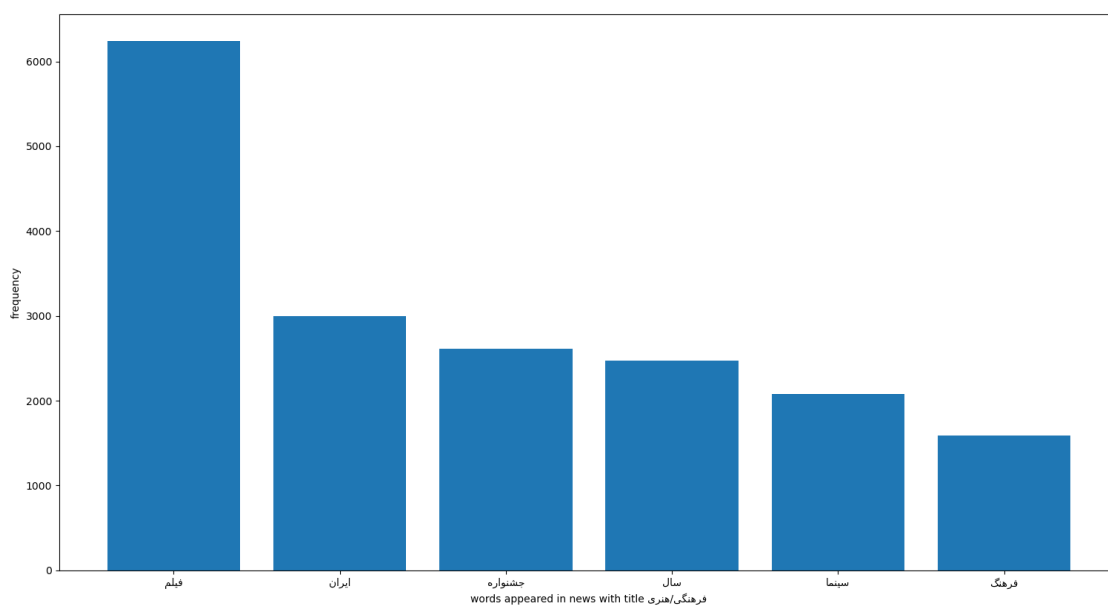
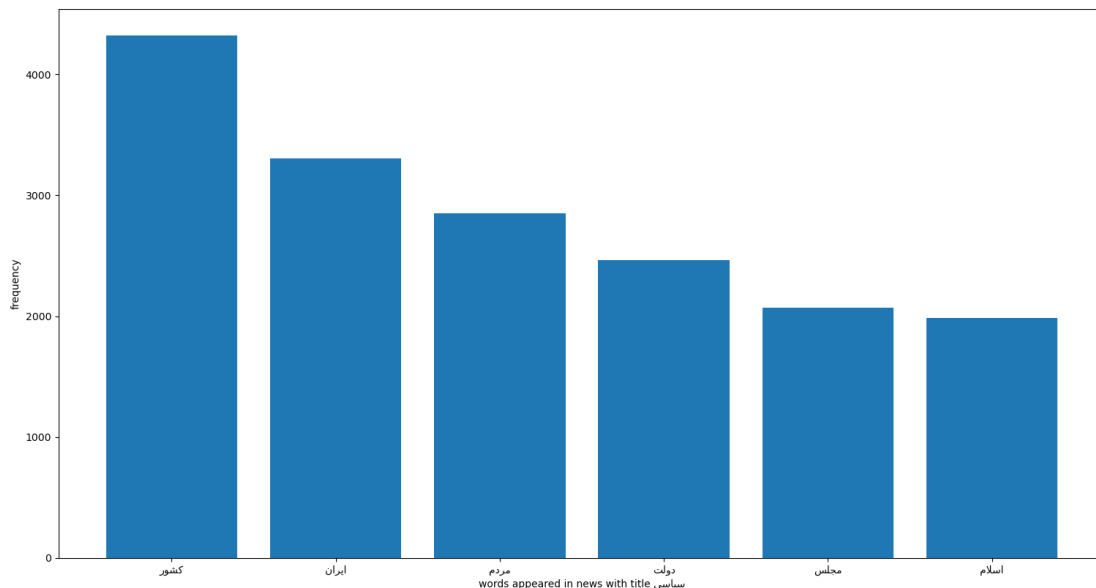
```
1 calc_titles_frequency()
2 #print(titles)
3
4 for title in titles.keys() :
5     wordsCountInTitles[title] = {}
6     totalWordsOfTitles[title] = 0
```

```
1 def calc_words_frequency() :
2     global totalWordsCount, distinctWordsCount
3     for _, row in trainData.iterrows() :
4         if row.isnull().values.any() :
5             continue
6         title = row['label']
7         content = row['content']
8         words = preprocessor(content)
9         n = len(words)
10        totalWordsOfTitles[title] += n
11        totalWordsCount += n
12        for word in words :
13            if word not in seenWords.keys() :
14                distinctWordsCount += 1
15                seenWords[word] = True
16            if word in wordsCountInTitles[title].keys() :
17                wordsCountInTitles[title][word] += 1
18            else :
19                wordsCountInTitles[title][word] = 1
20            if word in wordsCount.keys() :
21                wordsCount[word] += 1
22            else :
23                wordsCount[word] = 1
```

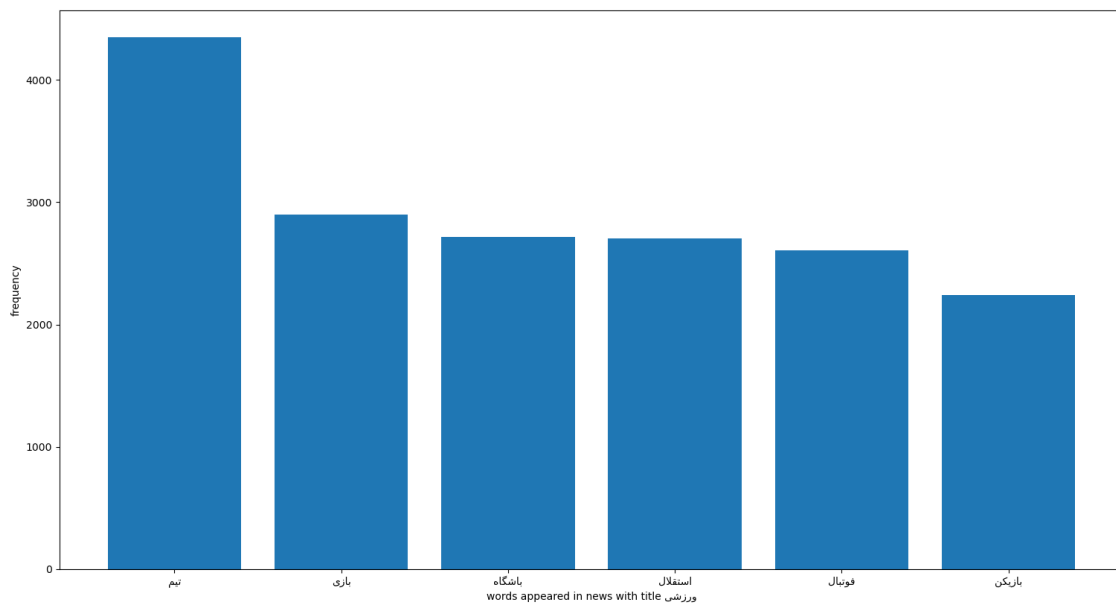
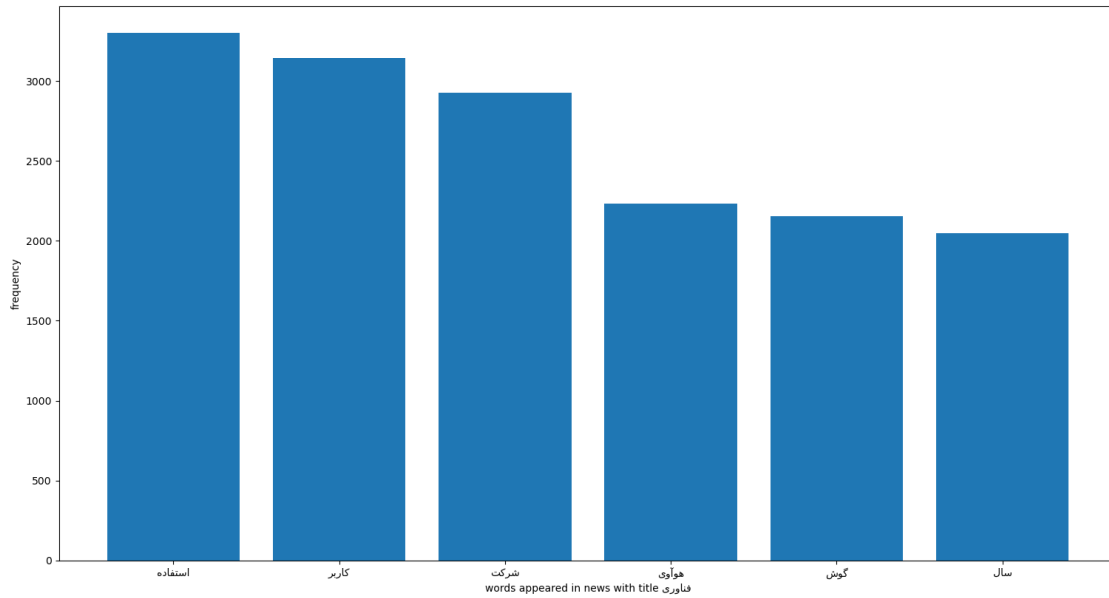
بررسی صحت:

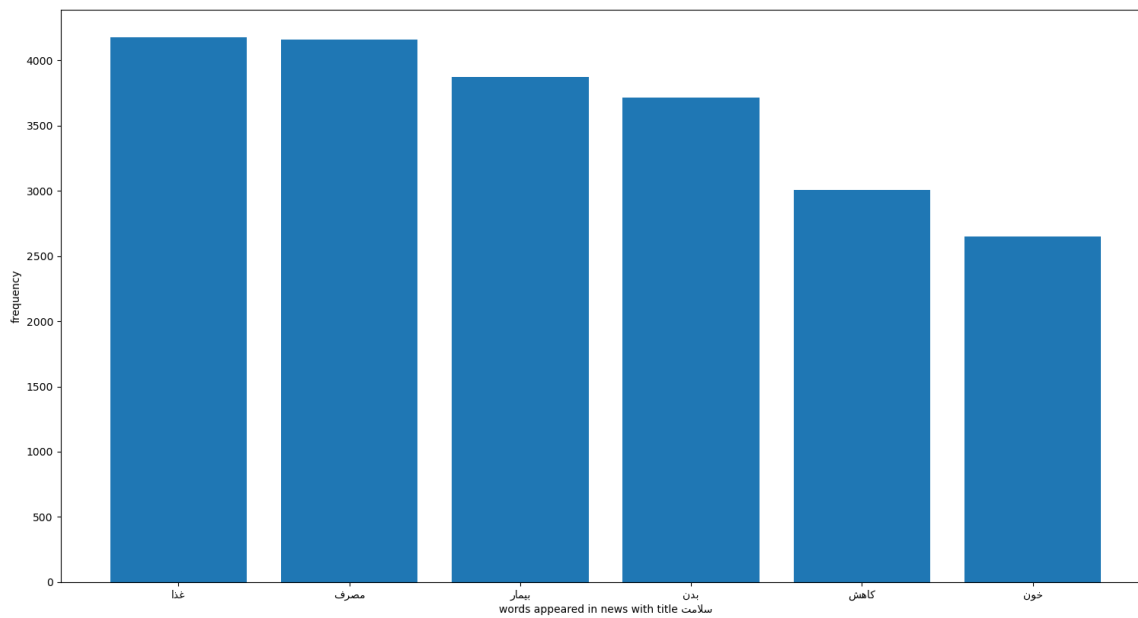
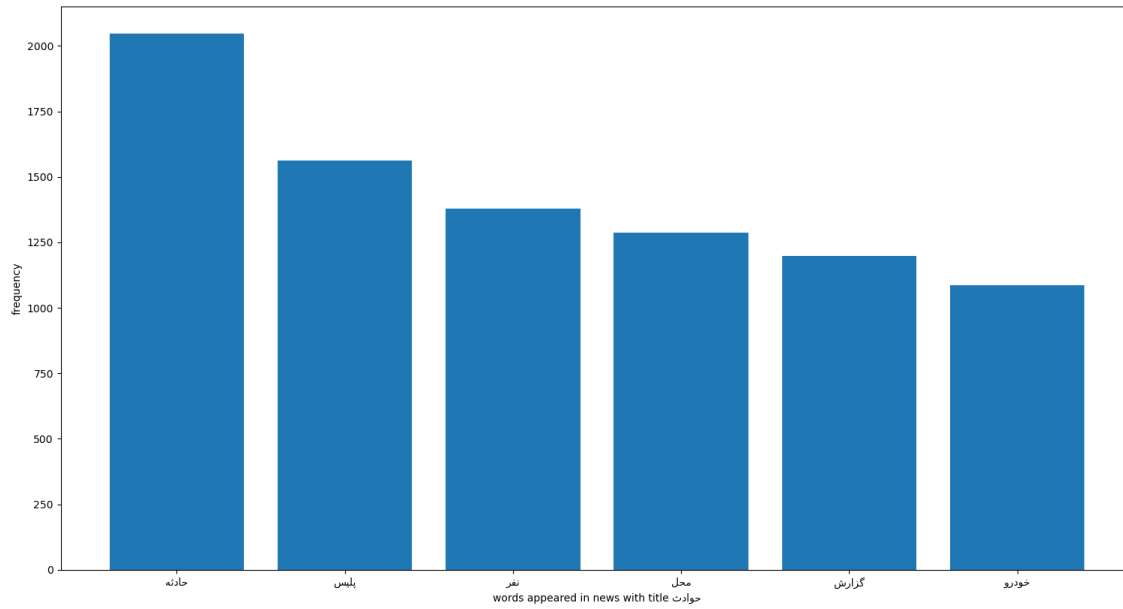


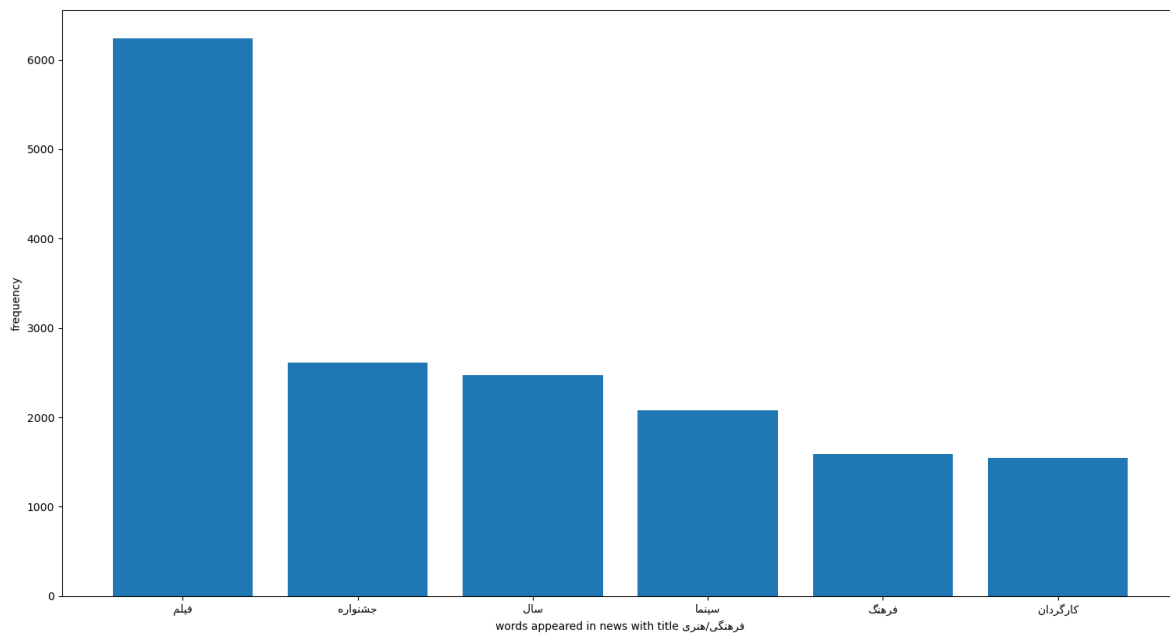
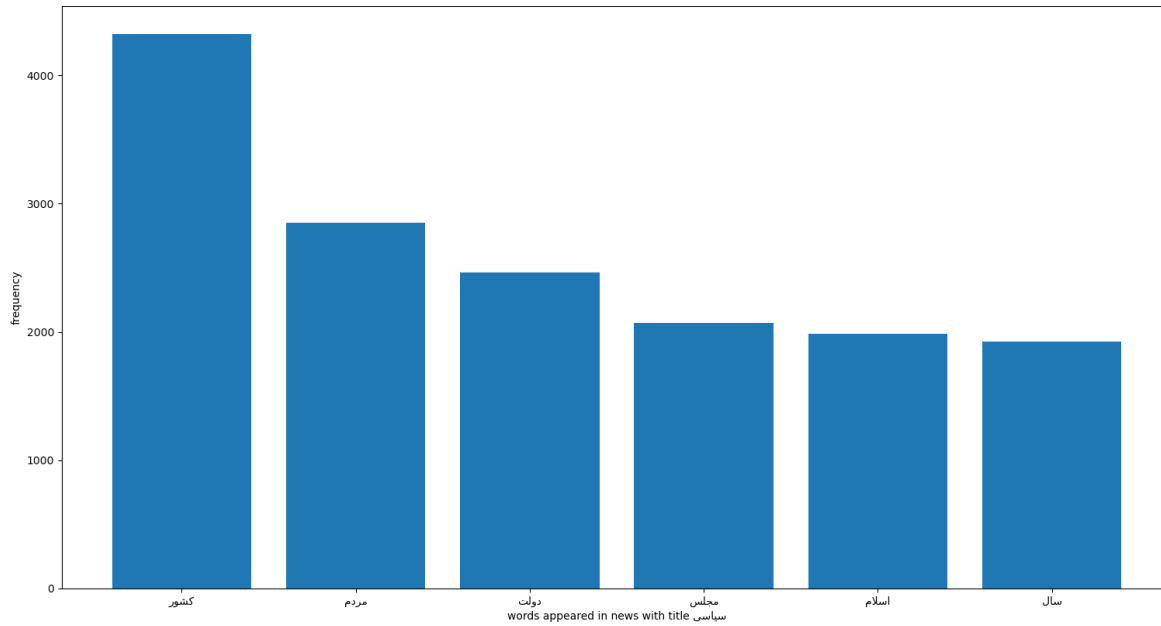




با توجه به نمودارهای تعداد کلمات پر تکرار در هر عنوان خبری می‌توان گفت بهتر است دو کلمه «ایران» و «توانست» را از لیست کلمات حذف کنیم. (به لیست ایست واژه‌ها اضافه‌شان می‌کنیم و دوباره کلمات پرتکرار در هر موضوع را پیدا می‌کنیم.) در ادامه نمودارهای جدید مربوط به تعداد کلمات پر تکرار هر عنوان خبری آورده شده است.







Bigrams

- شیر گران شد.
- شیر آهوپی را شکار کرد.

در جمله اول شیر به معنای شیر خوراکی و در جمله دوم شیر به معنای شیر جنگل است و می‌توان تا حدودی از روی کلمه بعدی، معنای شیر را در دو جمله فهمید.

استفاده از bigram با چک کردن تعداد تکرارهای کلمات و ترکیب آنها می‌تواند منجر به افزایش دقت classifier شود. در این مثال bigram برای مشخص کردن معنای کلمه کافیت و نیازی به n-gram طولانی نیست.

در طول پروژه فقط از unigram استفاده می‌شود و توجهی به بافت متن نشده است.

Additive Smoothing

طبق رابطه $p(c|x) = \frac{p(x|c)p(c)}{p(x)}$ در صورتی که مجموعه کلمات x در عنوان خبری c دیده نشود، $p(x|c)$ صفر می‌شود. در نتیجه $p(c|x)$ نیز صفر خواهد شد. تفسیر این موضوع این است که اگر مجموعه کلمات x فقط در یک عنوان خبری خاص مانند y آمده باشند، مدل با قاطعیت تشخیص خواهد داد که عنوان خبری هر محتوایی که مجموعه کلمات x در آن دیده شوند، y است که چنین چیزی نمی‌تواند همواره درست باشد.

برای رفع این مشکل روش additive smoothing پیشنهاد می‌شود که در این روش اگر فرض کنیم رابطه زیر برقرار است : $p(x|c) = \frac{f(x)}{s(c)}$ ، آنگاه به جای $\frac{f(x)}{s(c)}$ داریم :

$$\frac{f(x) + \alpha}{s(c) + \alpha \times \text{unique words}}$$

که $s(c)$ برابر تعداد کل کلماتی است که در عناوین خبری با عنوان c آمده‌اند و unique words برابر تعداد کلمات متفاوت متن است.

در این صورت اگر مجموعه کلمات x در عنوان خبری c وجود نداشته باشد، $p(x|c)$ و در نتیجه آن $p(c|x)$ صفر نخواهد شد. پس مدل نخواهد توانست با قاطعیت رد کند که عنوان خبری c نباشد.

فاز سوم : ارزیابی

برای راحتی در محاسبات به جای ضرب احتمالات از جمع لگاریتم آنها استفاده میکنیم:

```

1 def calc_probabilities(alpha, text) :
2     newsCount = len(trainData)
3     words = preprocessor(text)
4     probabilities = {}
5     #the loop below calculates p(c) for each news title (c : news title)
6     for title in titles.keys() :
7         probabilities[title] = np.log(titles[title] / newsCount)
8     # what has been calculated till here is p(c) and we put p(c|x) = p(c) --> In the following to calculate the correct value of
9     #p(c|x), we'll have : p(c|x) = p(c|x) * p(x|c) / p(x)
10    #to have easier calculations we use log, so * becomes +
11    for word in words :
12        for title in titles.keys() :
13            if word not in wordsCountInTitles[title] :
14                wordsCountInTitles[title][word] = 0
15            #if alpha == 0, additive smoothing is off else is on
16            if alpha == 0 :
17                probabilities[title] += np.log(wordsCountInTitles[title][word] / totalWordsOfTitles[title])
18            else :
19                probabilities[title] += np.log((wordsCountInTitles[title][word] + alpha) / (totalWordsOfTitles[title] + alpha * distinctWordsCount))
20    return probabilities
21
22
23 def predict_title(alpha, text) :
24     probabilities = calc_probabilities(alpha, text)
25     title = 'فناوری'
26     max_probability = -1 * inf
27     for key, value in probabilities.items() :
28         if value > max_probability :
29             title = key
30             max_probability = value
31    return title

```

Precision : در این معیار توجهی به خبرهایی که عنوان شان فرضاً c بوده است ولی عنوان شان به درستی تشخیص داده نشده است، نداریم. (به عبارت دیگر از در نظر گرفتن false negative ها صرف نظر می کنیم) به عنوان مثال فرض کنید ۵۰ خبر با عنوان سلامت داریم و فقط یکی از آن ها را درست تشخیص داده ایم، در این حالت با وجود بالا بودن مقدار precision، کاملاً واضح است که دقت پیش بینی بسیار پایین است.

```

1 def calc_precision(predictedTitles, label) :
2     correctPredictionsCount, allPredictionsCount = 0, 0
3     for i in range(len(correctTitles)) :
4         if predictedTitles[i] == label :
5             allPredictionsCount += 1
6             if predictedTitles[i] == correctTitles[i] :
7                 correctPredictionsCount += 1
8     if allPredictionsCount == 0 : return 0
9     else : return correctPredictionsCount / allPredictionsCount

```

Recall : به نوعی برعکس precision عمل می کند یعنی از خبر هایی که عنوان شان c نبوده است ولی به اشتباه عنوان شان c تشخیص داده شده است، صرف نظر می کند. (به عبارت دیگر از در نظر گرفتن false positive ها صرف نظر می کند) به عنوان مثال فرض کنید که عنوان تمام اخبار را «سلامت» تشخیص داده باشیم که مشخصا غلط است ولی recall بالایی دارد.

```
1 def calc_recall(predictedTitles, label) :
2     correctPredictionsCount, totalPredictionsCount = 0, 0
3     for i in range(len(correctTitles)) :
4         if correctTitles[i] == label :
5             totalPredictionsCount += 1
6             if predictedTitles[i] == correctTitles[i] :
7                 correctPredictionsCount += 1
8     if totalPredictionsCount == 0 : return 0
9     else : return correctPredictionsCount / totalPredictionsCount
```

F1 : از آنجایی که این معیار از harmonic mean دو معیار قبلی استفاده می کند، هم به false negative ها و هم به false positive ها توجه می کند. در مقایسه با accuracy، accuracy زمانی بهتر است که false positive ها و false negative ها هزینه تقریباً یکسانی برای ما داشته باشند، اگر تفاوت هزینه برای این دو مورد زیاد باشد، استفاده از F1 بهتر است.

```
1 def calc_f1(predictedTitles, label) :
2     precision = calc_precision(predictedTitles, label)
3     recall = calc_recall(predictedTitles, label)
4     if (precision + recall) == 0 : return 0
5     else : return (2 * precision * recall) / (precision + recall)
```

```
1 def calc_accuracy(predictedTitles) :
2     correctPredictionsCount = 0
3     for i in range(len(correctTitles)) :
4         if predictedTitles[i] == correctTitles[i] :
5             correctPredictionsCount += 1
6     return correctPredictionsCount / len(correctTitles)
```

macro : در این روش از F1-score تمام کلاس‌ها میانگین می‌گیریم.

```
1 def calc_macro(predictedTitles) :
2     f1Score, counter = 0, 0
3     for title, _ in titles.items() :
4         f1Score += calc_f1(predictedTitles, title)
5         counter += 1
6     return f1Score / counter
```

weighted : در روش macro وزن یکسانی به F1-score تمام عناوین داده می‌شد ولی در weighted-F1 به F1-score هر عنوان، متناسب با تعداد دیتای موجود از هر عنوان وزن می‌دهیم و در ادامه میانگین‌گیری وزن‌دار انجام می‌دهیم.

```
1 def calc_weighted(predictedTitles) :
2     weights, f1 = {}, {}
3     for title, _ in titles.items() :
4         weights[title] = 0
5         f1[title] = calc_f1(predictedTitles, title)
6     for title in correctTitles :
7         weights[title] += 1
8     f1score = 0
9     for title, value in f1.items() :
10         f1score += value * weights[title]
11     return f1score / sum(weights.values())
```

micro : این معیار برابر با همان accuracy است :

$$micro - F1 = \frac{CorrectDetected}{AllDetected} \rightarrow micro - F1 = \frac{F1_1 + F1_2 + \dots + F1_n}{n}$$

تشخیص عنوان خبری خبر های test data

در تابع زیر وقتی $\alpha = 0$ باشد، از additive smoothing استفاده نمی‌شود و در غیر این صورت از additive smoothing استفاده خواهد شد:

```
1 def calc_probabilities(alpha, text) :
2     newsCount = len(trainData)
3     words = preprocessor(text)
4     probabilities = {}
5     #the loop below calculates p(c) for each news title (c : news title)
6     for title in titles.keys() :
7         probabilities[title] = np.log(titles[title] / newsCount)
8     # what has been calculated till here is p(c) and we put p(c|x)
9     # = p(c) --> In the following to calculate the correct value of
10    #p(c|x), we'll have : p(c|x) = p(c|x) * p(x|c) / p(x)
11    #to have easier calculations we use log, so * becomes +
12    for word in words :
13        for title in titles.keys() :
14            if word not in wordsCountInTitles[title] :
15                wordsCountInTitles[title][word] = 0
16            #if alpha == 0, additive smoothing is off else is on
17            if alpha == 0 :
18                probabilities[title] += np.log(wordsCountInTitles[title][word] / totalWordsOfTitles[title])
19            else :
20                probabilities[title] += np.log((wordsCountInTitles[title][word] + alpha) / (totalWordsOfTitles[title] + alpha * distinctWordsCount))
21    return probabilities
```

```
1 predictedTitlesWithSmoothing, predictedTitlesWithoutSmoothing = [], []
2
3 for _,row in testData.iterrows() :
4     if row.isnull().values.any() :
5         continue
6     predictedTitlesWithSmoothing.append(predict_title(1, row['content']))
7     predictedTitlesWithoutSmoothing.append(predict_title(0, row['content']))
8
9 testData['predicted titles with smoothing'], testData['predicted titles without smoothing'] = predictedTitlesWithSmoothing, predictedTitlesWithoutSmoothing
10
11 correctTitles = testData['label']
```

در ادامه به پیش‌بینی عناوین خبری می‌پردازیم:

```
1 def print_result(predictions) :
2     newsTitles = ['سلامتی', 'سیاسی', 'ورزشی', 'فناوری', 'حوادث', 'فرهنگی/هنری']
3     precision, recall, f1 = {}, {}, {}
4     for title in newsTitles :
5         precision[title] = calc_precision(predictions, title)
6         recall[title] = calc_recall(predictions, title)
7         f1[title] = calc_f1(predictions, title)
8     accuracy = calc_accuracy(predictions)
9     macro = calc_macro(predictions)
10    micro = accuracy
11    weighted = calc_weighted(predictions)
12
13    print('precision:')
14    for item in list(precision.items()) :
15        print(item, end = ' ')
16    print()
17    print('recall:')
18    for item in list(recall.items()) :
19        print(item, end = ' ')
20    print()
21    print('f1:')
22    for item in list(f1.items()) :
23        print(item, end = ' ')
24    print()
25    print(f'accuracy: {accuracy}')
26    print(f'macro: {macro}')
27    print(f'micro: {micro}')
28    print(f'weighted: {weighted}')
29
30 print(f'results without additive smoothing:')
31 print_result(predictedTitlesWithoutSmoothing)
32 print()
33 print(f'result with additive smoothing:')
34 print_result(predictedTitlesWithSmoothing)
```

نتایج پیش‌بینی خبرها بدون استفاده از additive smoothing :

```
results without additive smoothing:
precision:
('فرهنگی/هنری', '0.8125') ('حوادث', '0.8611111111111112') ('فناوری', '0.18687430478309233') ('ورزشی', '0.9565217391304348') ('سیاسی', '0.8360655737704918') ('سلامت', '0.96')
recall:
('فرهنگی/هنری', '0.21666666666666667') ('حوادث', '0.155') ('فناوری', '1.0') ('ورزشی', '0.23157894736842105') ('سیاسی', '0.255') ('سلامت', '0.13559322033898305')
f1:
('فرهنگی/هنری', '0.2627118644067797') ('حوادث', '0.31490159325210876') ('فناوری', '0.3728813559322034') ('ورزشی', '0.3908045977011494') ('سیاسی', '0.2376237623762376') ('سلامت', '0.34210526315789475')
accuracy: 0.3201793721973094
macro: 0.32017140613772893
micro: 0.3201793721973094
weighted: 0.32115925652383304
```

نتایج پیش‌بینی خبرها با استفاده از additive smoothing :

```
result with additive smoothing:
precision:
('فرهنگی/هنری', '0.9238095238095239') ('حوادث', '0.9570552147239264') ('فناوری', '0.9945652173913043') ('ورزشی', '0.8853211009174312') ('سیاسی', '0.976878612716763') ('سلامت', '0.9520958083832335')
recall:
('فرهنگی/هنری', '0.8833333333333333') ('حوادث', '0.97') ('فناوری', '0.9285714285714286') ('ورزشی', '0.9631578947368421') ('سیاسی', '0.965') ('سلامت', '0.9548022598870056')
f1:
('فرهنگی/هنری', '0.9463414634146342') ('حوادث', '0.9425981873111783') ('فناوری', '0.9786096256684491') ('ورزشی', '0.923444976076555') ('سیاسی', '0.9657142857142856') ('سلامت', '0.9164265129682997')
accuracy: 0.9452914798206278
macro: 0.945522508525567
micro: 0.9452914798206278
weighted: 0.9454150790576176
```

همان‌طور که مشاهده می‌شود استفاده از additive smoothing دقت پیش‌بینی‌ها را به شدت افزایش داد (از ۳۲ درصد به ۹۴ درصد).

در حالتی که از additive smoothing استفاده نمی‌کنیم، از آن جایی که عنوان پیش‌فرض خبرها را «فناوری» در نظر گرفته ایم، این عنوان recall بالایی دارد و این عنوان به بسیاری از خبرها نسبت داده شده است.

در ادامه ۵ مورد از خبر هایی که حتی additive smoothing قادر به تشخیص صحیح عنوان آن ها نبوده است را می آوریم :

```
1 def print_mistakes() :
2     mistakesCount = 0
3     for index, row in testData.iterrows() :
4         if mistakesCount >= 5 :
5             return
6         if row.isnull().values.any() :
7             continue
8         if row['label'] != predictedTitlesWithSmoothing[index] :
9             print('#####')
10            print('real title :')
11            print(row['label'])
12            print('predicted title :')
13            print(predictedTitlesWithSmoothing[index])
14            print(row['content'])
15            mistakesCount += 1
16
17 print_mistakes()
```

```
#####
real title :
فرهنگی/هنری
predicted title :
سیاسی
```

چهارشنبه اول تیر ۱۴۰۱ اتفاق نادر و جالبی رخ داد که قطعاً به صورت تصادفی و بدون برنامه‌ریزی و اطلاع قبلی بوده است: این که به شکل کاملاً هم‌زمان در محل انجمن صنفی روزنامه نگاران استان تهران از رونمایی شد و در صداوسیما و در جمع معاونان سازمان هم از کلیات «...» اگرچه سه واژه در هر دو مشترک‌اند اما هیچ ربط و شابهتی به هم ندارند، چرا که در اولی منظور از سند است و در دومی سند (در حالی که به تعبیر دکتر وب2 را هم پشت سر گذاشته و به زودی وارد وب3 خواهیم شد) و خود رادیو و تلویزیون دیگر رسانه مدرن نیست و به تعبیری رسانه تلویزیون در معنی عام و جهانی نه مشخصاً صدا و سیما هر چند در ایران همچنان رایج‌ترین و گسترده‌ترین امکان وفه‌گذاری و ارتباطی است. (رایگان از جهت پرداخت حق اشتراک و گرنه بودجه آن از جیب ملت تأمین می‌شود). در اولی منظور از رسانه واقعاً رسانه است، یعنی بدون بودجه حکومتی و مخاطب‌محور و در دومی به جز تعبیر «رسانه ملی» در باقی موازین رد ترجیح می‌دهند از مفاهیم ایدئولوژیک استفاده کنند و آن را در واقع عرصه جنگ میدانند با واژه‌های جنگ و فرارگاه حجم عظیم پرسنل و بودجه‌های کلان و انحصار و جنگ انداختن به شبکه نمایش خانگی را توجیه می‌کنند. در اولی، مراد از واژه ملی واقعاً ملی است چون معطوف به ایران است و در دومی مراد از ملی معطوف به نظام سیاسی است و گرنه اگر ملی بود به زبان ملی، سینمای ملی، موسیقی ملی و مانند اینها بها میداد نه آن که چندان که در نوشته‌های آمدن آیین اولی دور از هر وابستگی دولتی و حکومتی و شاید تبلیغاتی و در حضور چهره‌های شاخص رسانه‌ای و استادان شاخص ارتباطات و با معیارهای کاملاً حرفه‌ای و جهانی و شناخته‌شده برگزارد شد و دومی اگر چه ذیل عنوان «رسانه ملی» اما با اهداف سیاسی و ایدئولوژیک و جدای معیارهای شناخته شده رسانه‌ای و می‌توان حدس زد بی‌حضور عالمان این رشته و در بنسبتی کاملاً اداری هر چند محتمل است برخی از معاونان تحصیلات مرتبط داشته باشند اما به این صفت حاضر نبودند. این اتفاق نمادین اما نشان‌دهنده ریشه مشکل رسانه‌ها در ایران هم هست و آن هم نگاه غیر علمی به موضوع رسانه در عرصه رسمی و برخوردار از بودجه‌های کلان چند هزار میلیارد تومانی در سال و با قریب ۵۰ هزار کارمند است حال آن که آن طرف با انواع محدودیتها و مصیقه‌ها دست به گریبان است و چرخ اقتصاد تحقیر آن با بارانه هم نمی‌گردد. جدای این زبان مشترک را هم دچار اختلال می‌کنند، چون ما از سه واژه سند و رسانه و ملی یک مراد داریم و کارمندان اداره سازمانی به نام صدا و سیما که همه باید حول این ۱۲ محور فعالیت کنند یک منظور. حاصل این چالش و کشاکش البته این شده که مردم هم از رسانه‌های چاپی و کاغذی رویگردان شده‌اند تا جایی که ایران به پایین‌ترین نرخ روزنامه‌خوانی در جهان - بله در جهان - سقوط کرده و هم از اخبار صداوسیما تا جایی که برای انبیا قُردن حسین عبدالباقی ناچار شدند کشتی سردخانه را باز کنند و بگویند اینها الناس! باور کنید این جنازه متعلق به مالک و سازنده، متروپل آبادان است و ما دروغ نگفتم هر چند روز دیگر گرفتار یک افتتاح با رسوایی دیگر شدند. این که به جای کارشناس واقعی به کارمند خودشان گفته بودند انگلیسی حرف بزن و بنده خدا وسط ماجرا با فارسی حرف زدن لو داد! نویسنده، این سطور، خود یکی از سخنرانان و حاضران نشست رونمایی از سند ملی رسانه در محل انجمن صنفی روزنامه نگاران بود اما از استادان ارتباطات خصوصاً دکتر محسین‌پوراد بسیار آموخت و همین مقدمه شد را با موازین حرفه‌ای بسیار منطبق یافت و می‌توان پیش بینی کرد متن کامل و مضمحل آن که حاصل چند گزارش تخصصی پژوهشی است کاری بسیار درخور باشد. در آن طرف هم آقای پیمان جلیلی ۱۲ محور تحول در محتوای برنامه‌های صدا و سیما را به این شرح اعلام کرده است: هر قدر ادبیات سند ملی رسانه (روزنامه نگاران) عاری از مفاهیم سیاسی و ایدئولوژیک و کاملاً منطبق با علوم ارتباطات و نظریات مدرن رسانه‌ها و مبتنی بر مشاوره‌های بر جسته ترین استادان ارتباطات و روزنامه نگاری است سند تحول صدا و سیما بر است از نگاه ایدئولوژیک و سیاسی و بسیار دور از مفاهیم تخصصی رسانه‌ای و آدم را بیشتر یاد اهداف سازمان تبلیغات اسلامی می‌اندازد. اگر صدا و سیما یک دستگاه وسیع و جیم تبلیغاتی - پروپاگاندایی است نقدی به رویکرد صدا و سیما نیست اما چرا اصرار دارند خود را ذیل عنوان «رسانه ملی» توصیف کنند؟ حال آن که این دو واژه تعاریف کاملاً خود را دارد. طبیعی است که آقای جلیلی مأموریت دارد این سازمان را با این اهداف متحول کند اما این محورها به علم رسانه‌ها چه ارتباطی دارند و چگونه می‌توانند با رسانه‌های دیگر رقابت با مقابله کنند؟ عین این ۱۲ محور را سازمان تبلیغات اسلامی هم می‌تواند اعلام کند و بر این اساس می‌توان پیش‌بینی کرد چنانچه این اهداف محقق شود سازمان صدا و سیما به سازمان تبلیغات اسلامی تبدیل خواهد شد و هر قدر این سازمان موفق بوده این سازمان هم موفق خواهد بود. (البته به لحاظ اداری و سازمانی در واقع شورای هماهنگی تبلیغات اسلامی، منظور تشکلات تبلیغاتی است نه یک عنوان خاص). منتها سازمان تبلیغات در طول سال با تعداد شرکت کنندگان در چند راهپیمایی مهم و سراسری و نوع برنامه‌ریزی برای آنها با جشن‌های ده فجر ارزیابی می‌شود حال آن که سازمان صدا و سیما لحظه به لحظه زیر ذره بین مخاطبان داخلی و خارجی است و قاعده‌ها با بدبختی‌ها و سراسری و نوع برنامه‌ریزی شدن تبلیغاتی. اگر سه کارکرد مهم صدا و سیما را ایدئولوژیک، سیاسی و فرهنگی ۱۳ سال گذشته و از زبان رؤسای مختلف سازمان بارها اعلام شده مشخص نیست کدام یک از این ۱۲ محور به اطلاع‌رسانی و سرگرمی و حتی آموزش معطوف می‌شود؟ مگر آن که آن سه رویکرد مربوط به مافیل سند بوده باشند و من بعد فرار است از این قرار باشد. اگر قرار بر تحولی این گونه است و با این اوصاف شاید بهتر باشد که صدا و سیما در سازمان تبلیغات اسلامی ادغام و نام آن به صدا و سیما و تبلیغات رسمی تغییر یابد چرا که هر چه باشد دیگر رسانه نخواهد بود آن هم از نوع ملی، کما این که فوتیبال، باید فوتیبال باشد حتی وقتی برای کمک به زلزله‌زده‌ها و دوستانه بازی می‌کنند نمی‌توان گفت چون دوستانه است بازیکنان می‌توانند با دست، توپ را لمس کنند، چون آن وقت دیگر فوتیبال نیست و هندبال است. با این سند تحول هم صدا و سیما سازمان تبلیغات می‌شود. البته این ۱۲ رویکرد اگر محور شایع با برخی شبکه‌ها باشد مرتبط و خیلی هم منطقی است اما مشخص نیست چگونه می‌توان تمام صدا و سیما را ذیل آن برد؟ هر چند برخی معتقدند اگر قرار بر تحول با سند تحول بود آموزش و پرورش باید خیلی زودتر با سند تحول متحول می‌شد با دانشگاه‌ها که مدام در حال تحول آنها هستند و چندان نباید جدی گرفت. روی کاغذ می‌نویسند و در عمل واقعیتها خود را تحمیل می‌کنند. نیکولو ماکیاولی در توصیه به شهریار می‌گفت پیمان‌ها برای شکست‌شدن بسته می‌شوند. به این سباق می‌توان گفت سندهای تحول هم برای متحول شدن نوشته می‌شوند! این صدا و سیما با این همه کابال و هزینه به خوراکهای متنوع نیاز دارد و صبح تا شب که نمی‌توان به غرب فحش داد و تبلیغ مذهبی پخش کرد. طالبان اگر توانستند با ساز و کاری دیگر بودی تعطیل کردند! چنان که در آغاز گفته شد تقارن عجیب این بود که درست در همان ساعتی که پیمان جلیلی در جمع معاونان صدا و سیما از سند تحول در سازمان متبوع با رویکرد تمام ایدئولوژیک رونمایی می‌کرد استاد راهنمای پایان نامه او بر ضرورت به کارگیری معیارهای حرفه‌ای رسانه‌ای در سند ملی رسانه در جمع روزنامه نگاران پای می‌فشرد تا مشخص شود ت

محمد جواد بشارتی

شماره دانشجویی: ۸۱۰۱۹۹۳۸۶

پروژه سوم

م ایدئولوژیک رونمایی می کرد استاد راهنمای پایان نامه او بر ضرورت به کارگیری معیارهای حرفه ای رسانه ای در سند ملی رسانه در جمع روزنامه نگاران پای می فشرد تا مشخص شود ت هاوت نگاه از زمین تا آسمان است. این یادداشت را به سه پرسش به پایان می‌برم:
اول: آیا پیش‌بینی می‌کنند اگر تمام شبکه‌های صداوسیما از مضامینی چون «عدالت گستری، روایت پیشرفت همه جانبه جامعه ایرانی، ایثار، مقاومت شهادت و توسعه سبک زندگی ایثارگرانه، سبک زندگی و خانواده با محوریت سبک زندگی انقلابی، هنر انقلاب با معرفی شایسته آثار برجسته، نقد غرب و هجو غربزدگی، تقویت هویت جهانی امتی با بازنمایی قدرت نرم ایران اسلامی در جهان، روایت جوان مومن انقلابی با نگاه پیشرفت آینده، الگوی زن انقلاب اسلامی، مقابله با جریان ارتجاع و تظہیر پهلوی، روایت هوشمندانه و صحیح تاریخ اسلام و ایران، بسط الگوی حکمرانی مردمی با بازخوانی نهادهای مردمی – انقلابی» پر شود شمار مخاطبان و اعتبار ا بن سازمان افزایش می‌یابد؟ (اگر زیاد کند تمام این نوشته را پس می‌گیرم).
دوم: صدا ووسیما ذیل عنوان رسانه قابل تعریف است یا قرارگاه فرهنگی با سازمان تبلیغات اسلامی و اگر نمی‌خواهد رسانه باشد چرا دود سازمان‌های دیگر ادغام نشود؟
پرسش سوم اما مهم‌تر این است: چرا شر را از سر گشاد آن می‌نوازید؟ این که مردم پای ب صدا و سیما ایدئولوژیک مثل ب با اجرای ، با ، با ، با و مانند آن در ساعات پربیننده شب بنشینند آنان را از شبکه‌های فارسی زبان ماهواره‌ای دور می‌کند یا اگر مدام برنامه های ایدئولوژیک پخش کنید تا او راد، قید داخل را بزنند و بروند سراغ آن طرفی‌ها؟
آنچه با نگاه رسانه‌ای – امنیتی با هوشمندی مفهیم د و مدیران دیگر سازمان کمتر می‌فهمند این است که غیرایدئولوژیک‌ترین برنا مه صداوسیما مثل برنامه‌های مدیری و فردوسی‌پور چون مردم را پای تلویزیون جمهوری اسلامی می‌نشانند و از م‌وتو و ایران اینترنت‌شال دور می‌کنند به نفع جمهوری اسلامی است اما ایدئول وژیک‌ترین و شعاری‌ترین آنها چون سبب می‌شود مخاطب عام کانال را تغییر دهد به سود رفقااست! یکی از این محورها است اما وقتی برنامه سرگرم‌کننده نداشته باشید و مردم را به م‌وتو سو ق دهید از این فرصت بهره می‌برند و با مستندهای خود به تظہیر می‌پردازند و در واقع نقش غرض می‌شود. (نمیدانم می‌توانم مطورم را برسانم یا نه!چه کنیم زبان مشترک مختل شده است.)
سال 88 در سفر به هرات افغانستان وقتی محری مراسم مدام به تعداد روزنامه ها و آزادی آنها می‌بالد تا به روزنامه‌های ایران و روزنامه‌نگاران میهمان طعنه بزند که دستشان زیر ساطور بود یکی از همکاران روزنامه‌نگار خانم -که شوربخانه بعدتر کوچید و رفت- در واکنش گفت: وقتی 70 درصد مردم سواد نداشته باشند با روزنامه نخواهند چه تفاوت دارد که این د عداد روزنامه داشته باشند یا نه! حالا حکایت صدا ووسیما و سند تحول آفای جلیی است. اگر دوشنبه شب خلق‌الله پای نود بنشینند می‌توانید چند تا شعار ایدئولوژیک را هم زیربوس پ کنید و قبل و بعد آن مفاهیمی را به خورد خلق‌الله بدهید اما وقتی نود را حذف می‌کنید و آن پسر هر قدر هم تلاش کند نتواند جای عادل را بگیرد ببینند، عام یک دکمه را فشار می‌دهد و در همان ساعت پای شبه مزدک میرزایی در ایران اینترنت‌شال می‌نشیند و آن وقت آنها هستند که از افکار عمومی جهت می‌دهند. تمام طرافت سیاسی، فرهنگی و امنیتی و نگو رسانه‌ای و در همین است.
این را خیلی خوب می‌فهمید و شنیدم هر وقت به قم میرفت و عالمی انتقاد می‌کرد که چرا در فلان سربال مثلا جاب خانم کامل نیست یا وعط و خطا به کم پخش می‌کنید توص یح میداد حاج آقا! حالا مثل قبل نیست. اگر بخوایم فقط شما را نشان بدهیم آن وقت تلویزیون را خاموش می‌کنند. اما اگر برنامه سرگرم‌کننده داشته باشیم هم آن را می‌بینند هم شما !

#####

real title :
فرهنگی/هنری
predicted title :
سیاسی
اعتراض گسترده کاربران توئیتر به اظهارات وزیر امورخارجہ آمریکا هشتک انگلیسی برشین گلف (خلیج فارس) را در توئیتر ترند کرد. به گزارش خبرآنلاین، رکس تیلرسون وزیر امورخارجہ آمریکا در اظهارات اخیرش با به کار بردن واژه‌ای معمول به جای خلیج فارس موجی از اعتراضات را در شبکه‌های اجتماعی برانگیخت. بسیاری از کاربران شبکه‌های اجتماعی به ویژه توئیتر گر همه می‌تواند شبکه استقلال را ببیند. ما از تیم، قهرمانی می‌خواهیم و سند تحول آفای جلیی است. اگر دوشنبه شب خلق‌الله پای نود بنشینند می‌توانید چند تا شعار ایدئولوژیک را هم زیربوس پ کنید و قبل و بعد آن مفاهیمی را به خورد خلق‌الله بدهید اما وقتی نود را حذف می‌کنید و آن پسر هر قدر هم تلاش کند نتواند جای عادل را بگیرد ببینند، عام یک دکمه را فشار می‌دهد و در همان ساعت پای شبه مزدک میرزایی در ایران اینترنت‌شال می‌نشیند و آن وقت آنها هستند که از افکار عمومی جهت می‌دهند. تمام طرافت سیاسی، فرهنگی و امنیتی و نگو رسانه‌ای و در همین است.
این را خیلی خوب می‌فهمید و شنیدم هر وقت به قم میرفت و عالمی انتقاد می‌کرد که چرا در فلان سربال مثلا جاب خانم کامل نیست یا وعط و خطا به کم پخش می‌کنید توص یح میداد حاج آقا! حالا مثل قبل نیست. اگر بخوایم فقط شما را نشان بدهیم آن وقت تلویزیون را خاموش می‌کنند. اما اگر برنامه سرگرم‌کننده داشته باشیم هم آن را می‌بینند هم شما !

#####

real title :
سیاسی
predicted title :
سیاسی
حوادث
وکیل مدافع نازنین زاغری متهم ایرانی - بریتانیایی از برگزاری دادگاه و اعلام ختم رسیدگی به دومین پرونده موکلش به اتهام فعالیت تبلیغی علیه نظام در شعبه ۱۵ دادگاه انقلاب . به ریاست قاضی صلواتی خبر داد #####

real title :
فرهنگی/هنری
predicted title :

ورزشی

روز گذشته در حاشیه معرفی ساینیتو سر مربی باشگاه استقلال مصطفی آجورلو از یک خبر داد، مدیرعامل استقلال اعلام کرد شبکه ماهواره‌ای استقلال به صورت آزمایشی فعالیت خود را - آغاز کرد. آجرولو گفت: «می‌خواهم به همه هواداران یک خبر خوب و ویژه بدهم. شبکه استقلال از طریق ماهواره، هم اکنون فعالیت خود را به‌صورت آزمایشی آغاز کرد. ما از طریق ماهوار ه، کار شبکه تلویزیونی استقلال را به صورت آزمایشی آغاز کردیم. درباره تلویزیون باید بگویم که منافع استقلال که است گم شده است و من باید این مبلغ را از صدا و سیما بگیرم، من دست‌پوش همه عوامل صدا وسیما هستم اما این پول باید داده شود اما هنوز هیچ اقدامی انجام نشده است. من ولی کار خود را می‌کنم و شبکه استقلال را راه‌اندازی می‌کنم. تا هفته دی گر همه می‌تواند شبکه استقلال را ببیند. ما از تیم، قهرمانی می‌خواهیم و سند تحول آفای جلیی است. اگر دوشنبه شب خلق‌الله پای نود بنشینند می‌توانید چند تا شعار ایدئولوژیک را هم زیربوس پ کنید و قبل و بعد آن مفاهیمی را به خورد خلق‌الله بدهید اما وقتی نود را حذف می‌کنید و آن پسر هر قدر هم تلاش کند نتواند جای عادل را بگیرد ببینند، عام یک دکمه را فشار می‌دهد و در همان ساعت پای شبه مزدک میرزایی در ایران اینترنت‌شال می‌نشیند و آن وقت آنها هستند که از افکار عمومی جهت می‌دهند. تمام طرافت سیاسی، فرهنگی و امنیتی و نگو رسانه‌ای و در همین است.
این را خیلی خوب می‌فهمید و شنیدم هر وقت به قم میرفت و عالمی انتقاد می‌کرد که چرا در فلان سربال مثلا جاب خانم کامل نیست یا وعط و خطا به کم پخش می‌کنید توص یح میداد حاج آقا! حالا مثل قبل نیست. اگر بخوایم فقط شما را نشان بدهیم آن وقت تلویزیون را خاموش می‌کنند. اما اگر برنامه سرگرم‌کننده داشته باشیم هم آن را می‌بینند هم شما !

#####

real title :
سلامت
predicted title :
سیاسی

وزیر بهداشت، طی نامه‌ای به معاون بهداشت این وزارتخانه دستور داد که هر چه سریع‌تر مرکز خدمات جامع سلامت در روستای چنار محمودی ایجاد و تیم تخصصی و فوق‌تخصصی در این مرکز مس تقرر شود. به گزارش ایسنا و بنا بر اعلام وبدا، متن نامه دکتر سعید نمکی خطاب به دکتر علیرضا ربیسی- معاون بهداشت وزارت بهداشت به شرح زیر است: «اینک که با هوشمندی مردم عذب ناکام ماندند و ضمن تنگن از بیگنری‌های مستمر HIV از و تلاش بی وقفه مسئولین محترم، گروه‌های ضد انقلاب و معاند در ایجاد آشوب به بهانه آلودگی تعدادی از هوطنان‌مان به ویروس ننابعالی، جناب آقای دکتر گویا و افراد فعال در حدود ۲۰ تیم اعزامی در ماه‌های اخیر به استان چهارمحال بختیاری و بررسی‌های دقیق کارشناسی همکاران استانی و ستادی، ضروریست موارد ذر بر مورد توجه قرار گیرد. گروهی فتنه انگیز جهت تشویش آذهان عمومی تلاش کرده اند از یکسو کلبه مردم شریف روستای چنارمحمودی را آلوده به ویروس قلمداد نمائند که آخرین بررسی بها خوشبختانه حاکی از آلودگی درصد کمی از مردم و سلامتی اکثر اهالی روستاست. از سوی دیگر سعی کردند از این ابزار کذب جهت بدنامی نظام بهداشتی درمانی کشور و نهایتاً "سربه بازی‌های لیک، پوشش تیم‌های پایه و ... همگی آیتمه‌های جذاب و البته مورد نیاز برای باشگاهی چون استقلال است. چه نیکوست که باشگاه‌های برطرفدار دیگر چون پرسپولیس، تراکتور، س باهان و ... این راه را ادامه داده و هر کدام برای خود شبکه تلویزیونی درست کنند. قطعاً چنین اتفاقی برای همه مثبت است، باشگاه ، هوادار، بازیکنان، و ... یک نکته مثبت دیگر آن در فوتبال ایران است که سالهای مدیدی است حق پخش تلویزیونی مسابقات، نحوه پوشش اخبار ، اطلاع رسانی و ... مناقشه ای بی پایان برای فوتبال ایران را رقم زده است. این نکت ه مثبت البته در چنین شرایطی احتمال دارد که جالش میان استقلال و صداسیما را بیشتر کند. در همین گفته‌ها آجرولو اشاره کرده است که باشگاه استقلال هزار میلیارد تومان از صداوسیما طلب دارد (کم شده است) و تئولی هیچ اقدامی برای آن صورت نگرفته است. طبیعتاً اگر شبکه‌های تلویزیونی دیگری وجود داشته باشد، صداوسیما از کنار چنین اتفاقاتی چنین چراغ خاموش و بی صدا عبور نخواهد کرد. امیدواریم که این اقدام شجاعانه آجرولو سرکوب نشود و تلویزیون باشگاه استقلال همچنان ادامه یابد، همچنین امیدواریم صداوسیما هم چوب لای چرخ آن نگ ذارد (مانند اتفاقاتی که برای پلتفرم‌های نمایش خانگی در حال رخ دادن است) ، این اتفاق حسن مصافع است. اساساً وجود در کشور یکی از نکات مغفول مانده است که قطعاً جای خالی آن ا حساس می‌شود، در هر پدیده‌ای وجود رقیب و فضای باز و سالم برای رقابت منجر به رشد و شکوفایی خواهد شد. در تمام دنیا باشگاه‌های ورزشی برطرفدار چنین شبکه‌های تلویزیونی دارند ، حیف است که ورزش ایران و علی الخصوص فوتبال ایران تا به امروز از چنین امکانی محروم مانده است. اگر آجرولو این اتفاق را به نتیجه برساند و بتواند تداوم آن را رقم بزند، قطعاً . اسم او به عنوان آغازکننده راهی جدید در ورزش و رسانه کشور خواهد ماند

#####

real title :
سلامت
predicted title :
سیاسی
وزیر بهداشت، طی نامه‌ای به معاون بهداشت این وزارتخانه دستور داد که هر چه سریع‌تر مرکز خدمات جامع سلامت در روستای چنار محمودی ایجاد و تیم تخصصی و فوق‌تخصصی در این مرکز مس تقرر شود. به گزارش ایسنا و بنا بر اعلام وبدا، متن نامه دکتر سعید نمکی خطاب به دکتر علیرضا ربیسی- معاون بهداشت وزارت بهداشت به شرح زیر است: «اینک که با هوشمندی مردم عذب ناکام ماندند و ضمن تنگن از بیگنری‌های مستمر HIV از و تلاش بی وقفه مسئولین محترم، گروه‌های ضد انقلاب و معاند در ایجاد آشوب به بهانه آلودگی تعدادی از هوطنان‌مان به ویروس ننابعالی، جناب آقای دکتر گویا و افراد فعال در حدود ۲۰ تیم اعزامی در ماه‌های اخیر به استان چهارمحال بختیاری و بررسی‌های دقیق کارشناسی همکاران استانی و ستادی، ضروریست موارد ذر بر مورد توجه قرار گیرد. گروهی فتنه انگیز جهت تشویش آذهان عمومی تلاش کرده اند از یکسو کلبه مردم شریف روستای چنارمحمودی را آلوده به ویروس قلمداد نمائند که آخرین بررسی بها خوشبختانه حاکی از آلودگی درصد کمی از مردم و سلامتی اکثر اهالی روستاست. از سوی دیگر سعی کردند از این ابزار کذب جهت بدنامی نظام بهداشتی درمانی کشور و نهایتاً "سربه بازی‌های لیک، پوشش تیم‌های پایه و ... همگی آیتمه‌های جذاب و البته مورد نیاز برای باشگاهی چون استقلال است. چه نیکوست که باشگاه‌های برطرفدار دیگر چون پرسپولیس، تراکتور، س باهان و ... این راه را ادامه داده و هر کدام برای خود شبکه تلویزیونی درست کنند. قطعاً چنین اتفاقی برای همه مثبت است، باشگاه ، هوادار، بازیکنان، و ... یک نکته مثبت دیگر آن در فوتبال ایران است که سالهای مدیدی است حق پخش تلویزیونی مسابقات، نحوه پوشش اخبار ، اطلاع رسانی و ... مناقشه ای بی پایان برای فوتبال ایران را رقم زده است. این نکت ه مثبت البته در چنین شرایطی احتمال دارد که جالش میان استقلال و صداسیما را بیشتر کند. در همین گفته‌ها آجرولو اشاره کرده است که باشگاه استقلال هزار میلیارد تومان از صداوسیما طلب دارد (کم شده است) و تئولی هیچ اقدامی برای آن صورت نگرفته است. طبیعتاً اگر شبکه‌های تلویزیونی دیگری وجود داشته باشد، صداوسیما از کنار چنین اتفاقاتی چنین چراغ خاموش و بی صدا عبور نخواهد کرد. امیدواریم که این اقدام شجاعانه آجرولو سرکوب نشود و تلویزیون باشگاه استقلال همچنان ادامه یابد، همچنین امیدواریم صداوسیما هم چوب لای چرخ آن نگ ذارد (مانند اتفاقاتی که برای پلتفرم‌های نمایش خانگی در حال رخ دادن است) ، این اتفاق حسن مصافع است. اساساً وجود در کشور یکی از نکات مغفول مانده است که قطعاً جای خالی آن ا حساس می‌شود، در هر پدیده‌ای وجود رقیب و فضای باز و سالم برای رقابت منجر به رشد و شکوفایی خواهد شد. در تمام دنیا باشگاه‌های ورزشی برطرفدار چنین شبکه‌های تلویزیونی دارند ، حیف است که ورزش ایران و علی الخصوص فوتبال ایران تا به امروز از چنین امکانی محروم مانده است. اگر آجرولو این اتفاق را به نتیجه برساند و بتواند تداوم آن را رقم بزند، قطعاً . اسم او به عنوان آغازکننده راهی جدید در ورزش و رسانه کشور خواهد ماند

#####

علت این که با additive smoothing هم هنوز خبر هایی وجود دارند که عنوان شان اشتباه تشخیص داده شده است این است که فقط به وجود داشتن یک کلمه یا گروهی از کلمات در یک خبر توجه می کنیم و به بافت متن توجهی نداریم. به عنوان مثال در خبر چهارمی که آوردیم وجود کلماتی مانند باشگاه، استقلال، هوادار و ... باعث می شود فکر کنیم این خبر، ورزشی است. در حالی که اگر به بافت متن توجه کنیم، خواهیم فهمید این خبر، فرهنگی/هنری است.