

عنوان : تشخیص نویسنده ی یک متن با استفاده از یادگیری ماشین و مخصوصا ترکیب طبقه بندها خواهد بود

<https://www.kaggle.com/c/spooky-author-identification/overview>

Overview*

Description-1

همانطور که من در دفتر کار، دست نوشته های دست داشتم، فکر کردم که آن را ساخته ام. هیچ چیز دیگر نمی تواند به من آسیب بزند. کمی می دانستم که آخرین ترس در اطراف گوشه وجود دارد.

دینگ! تلفن من با یک زنگ خوردن بی موقع ، من را از جا پراند . یک ایمیل، ناامید کننده مدیران Kaggle بود، به اشتراک گذاشتن اخبار مربوط به نشست اطلاعات دیگر. "ph'nglui mglw'nafh Cthulhu R'lyeh wgah'nagl fhtagn!"

من گریه کردم و همه چیز را پرت کرده بودم.. صفحات پراکنده در سراسر طبقه محوطه. چگونه می توانم با توجه به نویسندگان که آنها را نوشتم، چگونه آنها را با هم جمع کنم؟ یا برای همیشه از دست داده اند؟ صبر کن، فکر کردم ... من می دانم، یادگیری ماشین!

در سال جاری در مسابقات بازی های هالووین، شما به چالش کشیده شده است که پیش نویس نویسنده گزیده ای از داستان های ترسناک ادگار آلن پو، مری شلی و HP Lovecraft باشد. ما شما را تشویق (با جوایز نقدی!) برای به اشتراک گذاشتن نظرات خود در بحث و گفتگو در مسابقه و کد در هسته ها. ما جوایزی را برای جایزه دادن نویسندگان هسته ها و موضوعات بحث و گفتگو که مخصوصا برای جامعه ارزشمند است، تعیین کردیم. برای کسب اطلاعات بیشتر روی تب "Prizes" روی این صفحه مرور کلیک کنید.

ما برخی از هسته های شروع کننده را در Python و R قرار داده ایم تا به شما کمک می کنیم تا زمین را اجرا کنید.

<https://www.kaggle.com/rtatman/beginner-s-tutorial-python/notebook>

Evaluation-2

ارسال ها با استفاده از الگوریتم های مولتی کلاس ارزیابی می شوند. هر id یک کلاس واقعی دارد. برای هر id، باید یک احتمال پیش بینی شده برای هر نویسنده ارسال کنید.

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

فرمول :

که N تعداد مشاهدات در تست دیتا است، M تعداد لیبل کلاس است (3 کلاس)، y_{ij} یک است در صورتی که مشاهدات i ی متعلق به j باشد، در غیر این صورت y_{ij} برابر صفر است. P_{ij} احتمال این که پیش بینی شده مشاهدات i متعلق به کلاس j است.

احتمال های ارسال شده برای جملات داده شده لازم نیست که با یک جمع بشوند چرا که آنها از قبل مجددا مجزا می شوند (هر ردیف بر مجموع ردیف تقسیم می شود). به منظور اجتناب از log تابع ورودی، احتمالات پیش بینی شده با

$$\max(\min(p, 1 - 10^{-15}), 10^{-15}).$$

جا به جا می گردد.

شما باید فایل CSV را با شناسه و احتمال هر یک از سه کلاس ارسال کنید. ترتیب ردیف مهم نیست. فایل باید هدر داشته باشد و باید مانند موارد زیر باشد:

```
id,EAP,HPL,MWS
id07943,0.33,0.33,0.33
...
```

DataSet*

دیتاست این رقابت حاوی متن از آثار داستانی نوشته شده توسط نویسندگان رمان های وحشتناک **spooky** از دامنه عمومی:

Edgar Allan Poe, HP Lovecraft and Mary Shelley.

داده ها توسط CoreNLP's MaxEnt sentence tokenizer، متون بزرگتر را به جملات تکه تکه شده تبدیل کرده است. بنابراین شما ممکن است عبارات غیر عادی را اینجا و آنجا مشاهده کنید. هدف شما این است که نویسنده جملات را در مجموعه **test set** شناسایی کنید.

File descriptions-1

- train.csv - the training set
- test.csv - the test set
- sample_submission.csv - a sample submission file in the correct format یک فایل ارسالی نمونه در فرمت صحیح

1	id	EAP	HPL	MWS
2	id02310	0.403494	0.287808	0.308698
3	id24541	0.403494	0.287808	0.308698
4	id00134	0.403494	0.287808	0.308698
5	id27757	0.403494	0.287808	0.308698
6	id00001	0.403494	0.287808	0.308698

Data fields-2

- id - a unique identifier for each sentence یک شناسه منحصر به فرد برای هر جمله است
- text - some text written by one of the authors برخی از متن نوشته شده توسط یکی از نویسندگان
- author - the author of the sentence (EAP: Edgar Allan Poe, HPL: HP Lovecraft; MWS: Mary Wollstonecraft Shelley) نویسنده جمله

	A	B	C	D	E
1	id	text	author		
2	id26305	This process, however, afforded me no means of ascertaining the dimensions of my dungeon; as	EAP		
3	id17569	It never once occurred to me that the fumbling might be a mere mistake.	HPL		
4	id11008	In his left hand was a gold snuff box, from which, as he capered down the hill, cutting all manner	EAP		
5	id27763	How lovely is spring As we looked from Windsor Terrace on the sixteen fertile counties spread be	MWS		
6	id12958	Finding nothing else, not even gold, the Superintendent abandoned his attempts; but a perplexe	HPL		
7	id22965	A youth passed in solitude, my best years spent under your gentle and feminine fosterage, has sc	MWS		
8	id09674	The astronomer, perhaps, at this point, took refuge in the suggestion of non luminosity; and here	EAP		
9	id13515	The surcingle hung in ribands from my body.	EAP		
10	id19322	I knew that you could not say to yourself 'stereotomy' without being brought to think of atomies,	EAP		
11	id00912	I confess that neither the structure of languages, nor the code of governments, nor the politics of	MWS		
12	id15737	He shall find that I can feel my injuries; he shall learn to dread my revenge!" A few days after he si	MWS		

یک تذکر در مورد مسابقات:

در Kaggle ، روح مسابقات این است که از آن لذت ببرید و با هم یاد بگیرید. امتیازات شما روی لیدربرد ، نمی تواند نمره واقعی شما باشد.

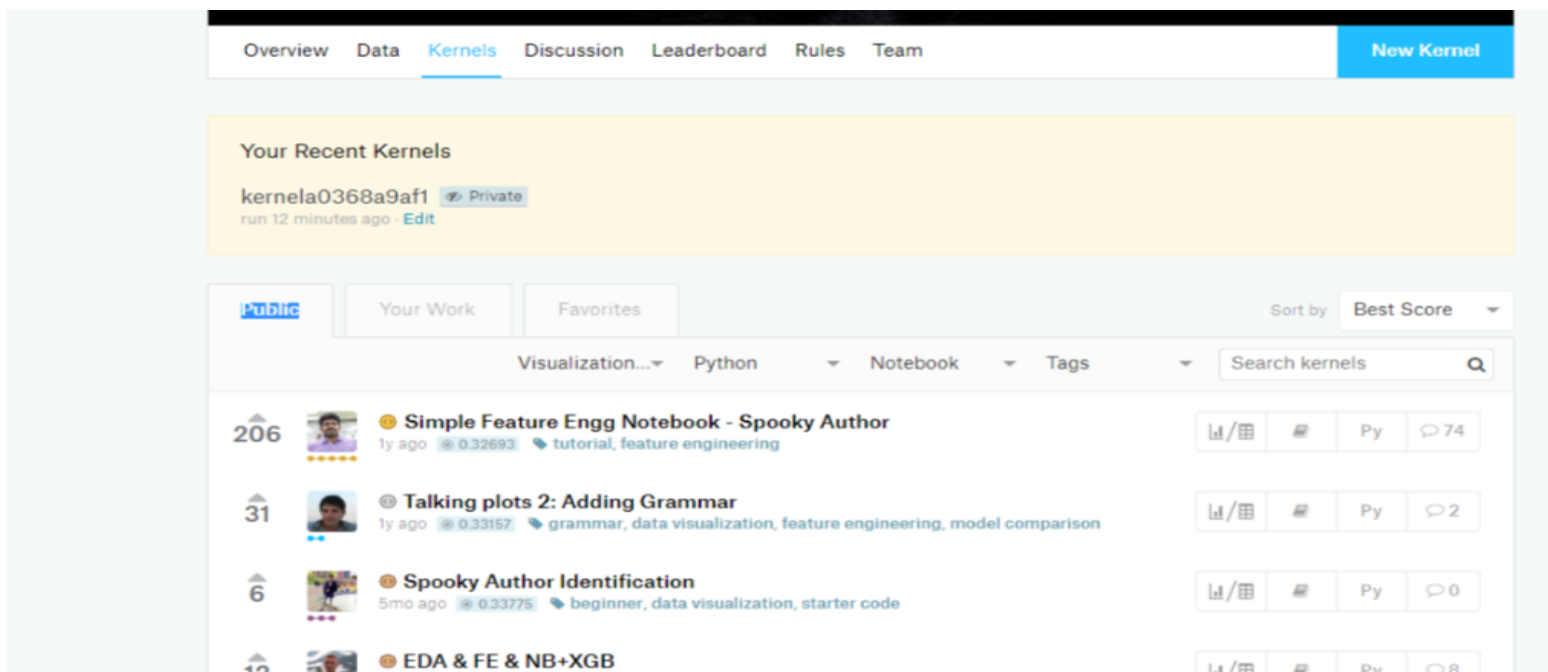
اما شما می توانید با به اشتراک گذاشتن کد خود در هسته ها و مشارکت در بحث ها (رقابت ها برای هر دو!) یک رقابت پرطرفدار برای همه باشد. به طور سطحی ، لطفا جوابها را ننویسید.

kaggle competitions download -c spooky-author-identification

Kernels*

Public-1

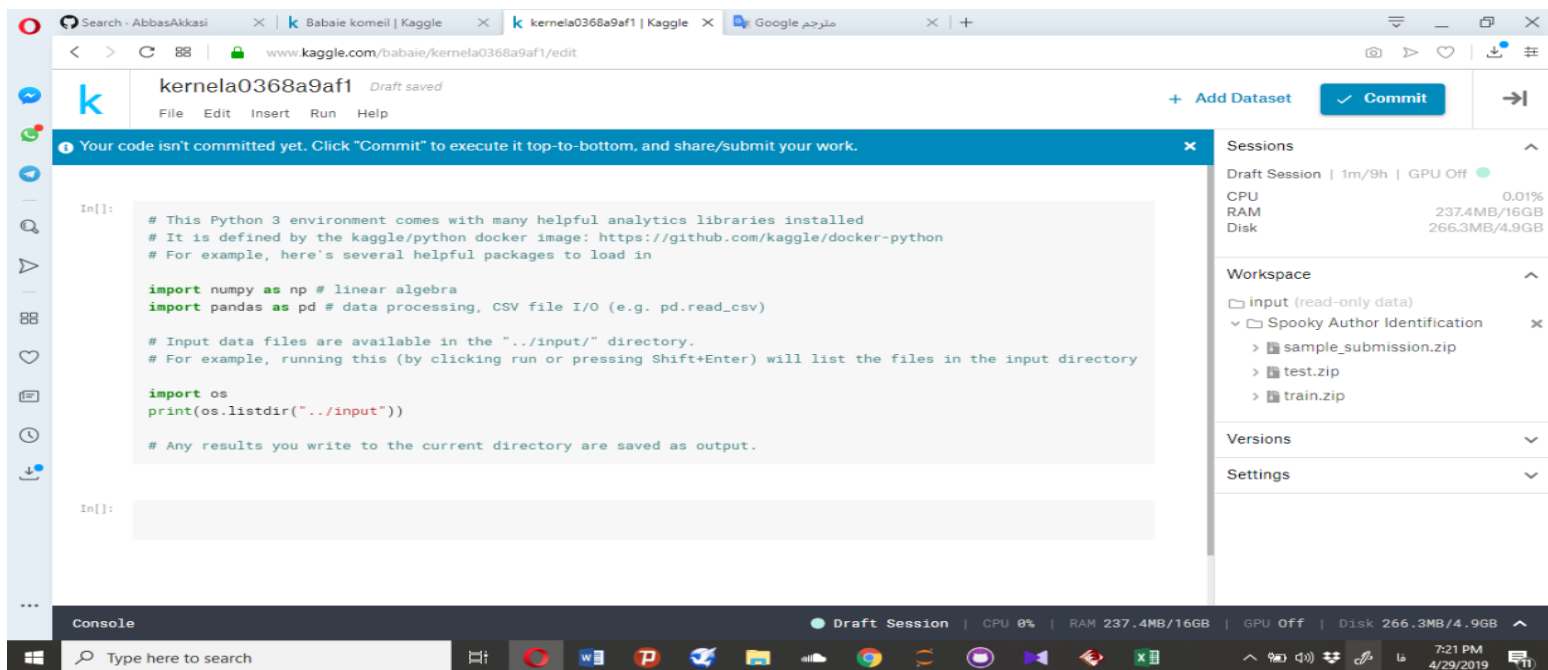
عملکرد سایر تیم ها به همراه کدهای که به اشتراک گذاشتند را می توانید مشاهده کنید. البته می توانید براساس نوع زبان و امتیاز و یا بیشترین کامن و... فیلتر انجام دهید.



The screenshot shows the Kaggle Kernels interface. At the top, there are tabs for Overview, Data, Kernels (selected), Discussion, Leaderboard, Rules, and Team. A 'New Kernel' button is on the right. Below this is a section for 'Your Recent Kernels' showing a private kernel 'kernela0368a9af1' run 12 minutes ago. The main section displays a list of public kernels. The top kernel is 'Simple Feature Engg Notebook - Spooky Author' by Babaie kornell, with a score of 0.32693 and 206 upvotes. Other kernels include 'Talking plots 2: Adding Grammar' and 'Spooky Author Identification'.

Your Work-2

میحت را برای ما آماده کردند. 😊



The screenshot shows the Kaggle Jupyter Notebook interface for the kernel 'kernela0368a9af1'. The notebook contains Python code for loading libraries (numpy, pandas, os) and listing files in the input directory. The right sidebar shows the workspace with files like 'sample_submission.zip', 'test.zip', and 'train.zip'. The bottom status bar shows the draft session details: CPU 0%, RAM 237.4MB/16GB, GPU Off, and Disk 266.3MB/4.9GB.

تالار گفتگو

Public Leaderboard-1

Overview Data Kernels Discussion **Leaderboard** Rules Team My Submissions **Late Submission**




Public Leaderboard Private Leaderboard

This leaderboard is calculated with approximately 30% of the test data.

The final results will be based on the other 70%, so the final standings may be different.

[Raw Data](#) [Refresh](#)

■ **In the money**

#	Team Name	Kernel	Team Members	Score 🏆	Entries	Last
1	banana			0.03224	50	1y
2	Newbee			0.03521	28	1y
3	AlexanderHades			0.08105	16	1y

این لیدر مورد با حدود 70٪ از testdata محاسبه می شود. این رقابت تکمیل شده است این لیدر مورد نشان دهنده جدول نهایی است.

Overview Data Kernels Discussion **Leaderboard** Rules Team My Submissions **Late Submission**

Public Leaderboard Private Leaderboard

The private leaderboard is calculated with approximately 70% of the test data.

This competition has completed. This leaderboard reflects the final standings.

 Refresh

■ **In the money**

#	Δ pub	Team Name	Kernel	Team Members	Score 🏆	Entries	Last
1	—	banana			0.02467	50	1y
2	—	Newbee			0.02593	28	1y
3	▲ 1	东大老实人		 	0.07036	105	1y
4	▼ 1	AlexanderHades			0.07174	16	1y

یک حساب کاربری در هر شرکت کننده

شما نمیتوانید از چندین حساب به Kaggle ثبت نام کنید و بنابراین نمیتوانید از چند حساب ارسال کنید.

هیچ اشتراک خصوصی در تیم های خارجی وجود ندارد

خصوصی سازی کد یا داده های خارج از تیم مجاز نیست. اگر به همه اشتراک کنندگان در انجمن ها دسترسی پیدا کنید، به اشتراک گذاری کد خوب است.

ادغام تیم

ادغام تیم مجاز می باشد و می تواند توسط رهبر تیم انجام شود. به منظور ادغام، تیم ترکیبی باید تعداد ارسال کامل را کمتر یا برابر حداکثر مجاز به عنوان تاریخ ادغام داشته باشد. حداکثر مجاز تعداد ارسال ها در روز ضرب شده توسط تعداد روزهای رقابت در حال اجرا است.

محدودیت های تیم

حداکثر اندازه تیم وجود ندارد.

محدودیت های ارسال

شما می توانید حداکثر 5 پست در روز ارسال کنید.

شما می توانید حداکثر 2 نهایی را برای قضاوت انتخاب کنید.

گامشمار مسابقه

تاریخ شروع: 23/10/2014

مهلت تجمع: هیچکدام

مهلت ورودی: هیچ

تاریخ پایان: 15/12/2015 11:59 UTC

شرایط رقابت

مسابقات برای ساکنان ایالات متحده و سراسر جهان باز است، مگر اینکه شما ساکن کریمه، کوبا، ایران، سوریه، کره شمالی، سودان هستید یا تحت کنترل ایالات متحده یا کنترل تحریم قرار می گیرید، پس شما ممکن است وارد رقابت شوید. دیگر قوانین و مقررات محلی ممکن است برای شما اعمال شود، بنابراین لطفا قوانین محلی خود را بررسی کنید تا اطمینان حاصل شود که شما واجد شرایط مشارکت در مسابقات مبتنی بر مهارت هستید.

ورود به این رقابت، پذیرش شما از این قوانین رقابت رسمی است.

مسابقه که در بالا ذکر شد، یک رقابت مبتنی بر مهارت است تا زمینه دانش علمی را ارتقا دهد. شما باید از طریق وب سایت مسابقه برای ثبت نام وارد شوید. ارسال های رقابت شما ("ارسال ها") باید مطابق با الزامات مندرج در وب سایت مسابقه باشد. حامی مسابقه، مطالب خود را براساس معیار سنجش ارزیابی شده در وب سایت مسابقه، ضبط خواهد کرد. با توجه به انطباق با قوانین مسابقه، جوایز در وب سایت مسابقه شرح داده شده، در صورت وجود، به شرکت کنندگان با بالاترین نمره بر اساس محتویات ارائه شده و همچنین مدرک Kaggle اعطا می شود. برای قوانین مسابقه کامل زیر را ببینید.

("TITLE COMPETITION مسابقه"): شناسایی وحشیانه نویسنده

اسپانسر رقابت: Kaggle

وب سایت رقابت: <https://www.kaggle.com/c/spooky-author-identification>

جوایز: مجموع 25000 دلار

و.....

Browser tabs: Password sent! | Babaie komeil | Kaggle | Spooky Author Identification | Google مترجم | Google مترجم

Address bar: www.kaggle.com/c/spooky-author-identification/team

Spooky Author Identification

Share code and discuss insights to identify horror authors from their writings

\$25,000
Prize Money

Kaggle · 1,244 teams · a year ago

Overview Data Kernels Discussion Leaderboard Rules **Team** My Submissions Late Submission



Manage Team

Team Name

Babaie Komeil [Save Team Name](#)

This name will appear on your team's leaderboard position.

Team Members

		Babaie komeil (you)	Leader
---	---	----------------------------	--------

[Invite Others](#)

Windows taskbar: Type here to search | 8:30 PM 4/29/2019