



## پیش بینی ژانر فیلم های کافه بازار

یادگیری ماشین - پاییز ۹۹ - دانشگاه تهران

### صورت مسئله



امیرحسین به تازگی مسئول نگه داری فیلم های کافه بازار شده و داره سعیش رو می کنه، که همه فیلم ها اطلاعات کاملی داشته باشن و مرتب باشن. توی این مرتب سازی فیلم ها میبینه که ژانر تعدادی از فیلم هایی که تو کافه بازار داریم معلوم نیست و از اونجایی که آدم خیلی دقیقی هست و اسش مهم شده که بدون این فیلم ها تو چه ژانری هستن تا بتونه اون ها رو به درستی مرتب کنه.



امیرحسین به فکرش رسید که بشینه فیلم ها رو ببینه و بفهمه هر کدوم چه ژانرین اما از اونجایی که قراره که تا به هفته دیگه این کارو انجام بده، فهمیده که وقت نمیشه. از اونجایی که یه چیزایی درباره علوم داده و ماشین لرنینگ میدونه یه ایده ای به ذهنش رسیده که بیاد با داده های بقیه فیلم هایی که قبلا مرتب کرده، ژانر اون فیلم هایی که ژانرشون معلوم نیست رو حدس بزنه. برای همین هم از شما می خواد که این کارو براش انجام بدید و نتایج رو براش بفرستید.

### لینک داده ی فیلم هایی که مرتب شده

لینک	عنوان داده
<a href="https://drive.google.com/file/d/1t_4fBp8N-B5rg_o1OR9cLT_7ev-lAjVt/view?usp=sharing">https://drive.google.com/file/d/1t_4fBp8N-B5rg_o1OR9cLT_7ev-lAjVt/view?usp=sharing</a>	داده ی آموزش فیلم ها
<a href="https://drive.google.com/file/d/1BaEnnhQOHSOU7oFGj1yJPcbgNCWUJRir/view?usp=sharing">https://drive.google.com/file/d/1BaEnnhQOHSOU7oFGj1yJPcbgNCWUJRir/view?usp=sharing</a>	پوستر فیلم ها در داده ی آموزش
<b>نکته مهم:</b> داده هایی که قراره شما پیش بینی کنید (داده های ارزیابی) رو ۲ روز قبل از ددلاین در اختیارتون قرار خواهیم داد.	

## مشخصات داده‌ی فیلم‌هایی که مرتب شده

ویژگی	توضیحات
video_id	آی دی هر کدوم از فیلم‌ها
title_fa	عنوان فیلم‌ها
description_fa	توضیحات هر کدوم از فیلم‌ها (حدس امیرحسین اینه که داده‌ی ارزشمندی برای پیش‌بینی باشه)
age_rating	حداقل سن مجاز برای تماشای فیلم مثلاً ۱۷ یعنی افراد بالای ۱۷ سال مجاز به دیدنش هستن.
country_fa	کشوری که فیلم توش ساخته شده
production_year	سالی که فیلم توش ساخته شده
duration	طول زمان فیلم
genre	ژانر فیلم
image	پوستر فیلم، در واقع اسم عکسی که توی پوشه «پوستر فیلم‌ها در داده‌ی آموزش» قرار داره. (امیرحسین اینجا هم فکر می‌کنه که با استفاده از این ویژگی میشه پیش‌بینی بهتری داشت)

## نحوه ارزیابی

مواردی که در این پروژه نمره دارند: ۱. گزارش کار ۲. پیش‌بینی شما روی داده‌ی ارزیابی

## گزارش کار

مراحلی که برای حل مسئله طی کردین رو مهمه که توی گزارش بیارین. مرحله‌ی مثل

- آشنایی با داده‌ها و فاز اکتشافی و تصویرسازی و تحلیل‌ها (Exploratory data analysis)

- به طور مثال تصویرسازی توزیع‌ها، شناسایی داده‌های پرت و ...

- پیش‌پردازش داده‌ها و انتخاب ویژگی

- به کارگیری روش‌هایی برای استخراج ویژگی از توضیحات فیلم‌ها مانند Bag of Words و ...

- به کارگیری روش‌هایی برای استخراج ویژگی از پوستر فیلم مانند SIFT و ...

■ برای مطالعه بیشتر روش‌های استخراج ویژگی از متن از این [لینک](#) و استخراج ویژگی از تصویر از این [لینک](#) استفاده کنید.

- به کارگیری روش‌هایی برای انتخاب ویژگی از داده‌های پیش‌پردازش شده مثلاً روش‌های فیلترینگ در اهمیت ویژگی و ...

- در صورت لزوم به کارگیری روش‌هایی برای کاهش ویژگی از داده‌های پیش‌پردازش شده مانند Encoding, PCA و ...

## • آموزش مدل طبقه‌بندی و ارزیابی عملکرد مدل

- استفاده یک یا چندتا از الگوریتم‌های طبقه‌بندی مانند SVM, Decision Tree, Logistic Regression, MLP و ... .  
(استفاده از مدل‌های یادگیری عمیق مانند CNN و ... نمره امتیازی دارد)
- ارزیابی عملکرد مدل
  - ارزیابی دقت با معیارهای رایج مانند Accuracy, F1 score و ...
  - سنجش زمان آموزش و پیش‌بینی
  - تحلیل خطا و شناسایی نقاط ضعف مدل

## • جمع‌بندی و نتیجه‌گیری

**نکته:** این انتظار می‌ره که گزارش و خروجی در قالب فایل‌هایی مثل Jupyter Notebook, R Markdown و ... ارائه بشن که قابلیت اجرای مجدد و رسیدن به همان نتایج رو داشته باشن. همینطور گزارش باید قابل فهم باشه و مراحل که برای حل مسئله طی شده، شفاف شده باشه.

**نکته:** اگه اسکریپتی به جز موارد بالا می‌فرستین، لازمه که همراهش یک گزارش که نتایج و نمودارها رو داشته باشه به همراه گزارش فرستاده بشه و الزاما نیاز نباشه کد اجرا بشه.

## پیش‌بینی شما روی داده‌ی ارزیابی

بعد از اینکه شما به مدل نهایی رسیدید، قراره داده‌های ارزیابی رو طبقه‌بندی کنید و از طریق لینک زیر بفرستید. خروجی شما بر اساس دقت (accuracy) پیش‌بینی شما از ژانر نمره‌ای داده خواهد شد. علاوه بر دقت، معیارهایی مانند f1-score و Balanced Accuracy هم مد نظر هستش.

**نکته:** شما حداکثر ۱۰ بار می‌تونید این پیش‌بینی‌ها رو بفرستید.

لینکی که شما باید پیش‌بینی‌های خود را به آن بفرستید: [به زودی در اینجا یا صفحه درس اطلاع‌رسانی خواهد شد]

اگر سوال یا ابهامی داشتید، از طریق ایمیل [Sepsad@cafebazaar.ir](mailto:Sepsad@cafebazaar.ir) می‌تونیم با هم در ارتباط باشیم.