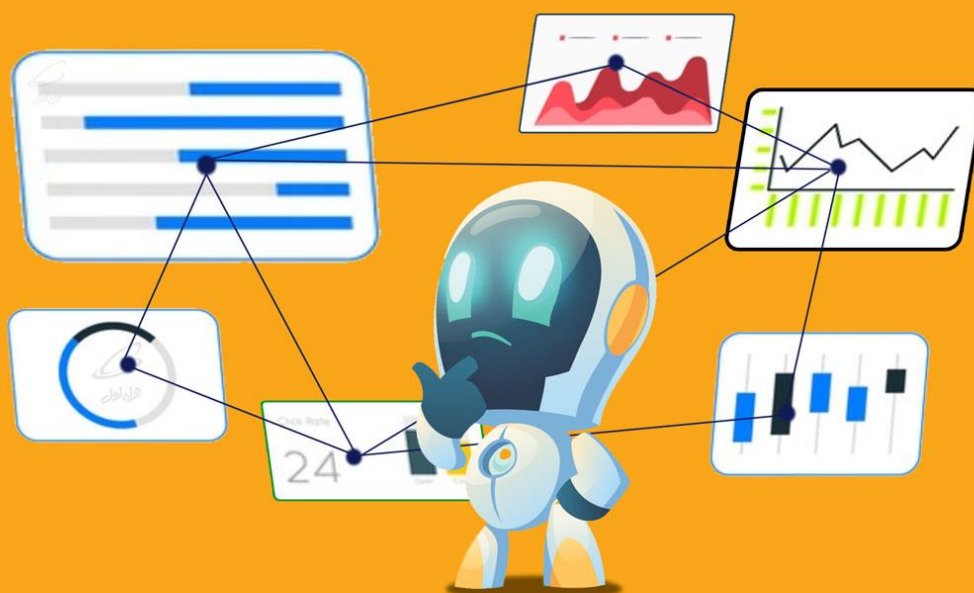


دانشگاه صنعتی امیرکبیر با همکاری مرکز تحقیق و توسعه همراه اول برگزار می‌کند

چالش پیش‌بینی میزان مصرف اینترنت مشترکین همراه اول

رقابت متخصصین هوش مصنوعی در پیش‌بینی سری‌های زمانی

ششمین دوره مسابقات هوش مصنوعی امیرکبیر (مهما)



تا ۲۸ بهمن تمدید شد

مهلت ثبت نام: ۲۸ بهمن ماه

مرحله آزمایشی	مرحله اول	مرحله دوم
۱۶ بهمن ماه	۳۰ بهمن ماه	۱۴ اسفند ماه

تیم اول	۱۰ میلیون تومان
تیم دوم	۷ میلیون تومان
تیم سوم	۳ میلیون تومان

هزینه شرکت در مسابقه رایگان است

✉ aaic@aut.ac.ir
🌐 aaic.aut.ac.ir
📱 @aaic_aut
☎ ۰۲۱ - ۶۴۵۴۲۷۰۰



شرکت ارتباطات سیار ایران
همراه اول



مرکز تحقیق و توسعه
همراه اول



دانشکده مهندسی
کامپیوتر



دانشگاه صنعتی
امیرکبیر



مسابقات هوش مصنوعی امیرکبیر (مهما)
چالش پیش‌بینی میزان مصرف روزانه اینترنت
مشترکین همراه اول

Amirkabir Artificial Intelligence Competitions (AAIC)
Challenge of forecasting daily internet usage of MCI subscribers
(User Internet Usage Forecasting)
MCI UIUF Challenge

باسمه تعالی

۱. مقدمه

باتوجه به رشد فناوری و افزایش ابزارهای متعدد دیجیتالی، استفاده از اینترنت در ابزارهای مختلف متداول شده است. شرکت ارتباطات سیار ایران (همراه اول) به عنوان بزرگترین اپراتور خاورمیانه که به طیف وسیعی از مردم خدمات متنوع ارتباطی نظیر اینترنت ارائه می دهد در نظر دارد با شناسایی بهترین استراتژی های بر پایه هوش مصنوعی، به افزایش بهره وری، ارائه خدمات متنوع و بهینه و ایجاد تجربه کاربری عالی در سطوح کاربری مختلف بپردازد.

۲. شرح مسابقه

در این مسابقه به پیش بینی مصرف اینترنت گروهی خاصی از مشترکین منتخب همراه اول خواهیم پرداخت. در واقع نیاز است تا مدلی هوشمند و منعطف بر پایه علم هوش مصنوعی روز طراحی و توسعه یابد. در این مسابقه داده های اینترنت مصرفی روزانه آن کاربران به مدت سه ماه متوالی از سیم کارت های مختلف گمنام شده در اختیار شرکت کنندگان قرار می گیرد. شرکت کنندگان می بایست مدلی طراحی نمایند تا با استفاده از ۶۹ روز (دو ماه و یک هفته) داده دریافتی به بررسی رفتار داده ها پرداخته و در قالب سه مرحله، به پیش بینی میزان مصرف اینترنت مشترکین در سه هفته انتهایی اقدام نمایند. هفته اول به عنوان مرحله آزمایشی در نظر گرفته می شود و نتایج آن در ارزیابی تأثیر نخواهد داشت. در انتهای هر مرحله، مقدار واقعی مصرف اینترنت آن مرحله مشترکین در اختیار شرکت کنندگان قرار می گیرد تا در مراحل بعدی از آن استفاده شود.

۳. نحوه شرکت

ثبت نام از طریق سایت مسابقات انجام می شود. شرکت در این مسابقه به صورت انفرادی یا گروهی بوده و محدودیتی در تعداد افراد تیم ها وجود ندارد ولی نباید فردی در تیم های مشترک ثبت نام نماید.

پس از تایید ثبت نام، اطلاعات دادگان آموزشی و ارزیابی و توضیحات نحوه شرکت در مسابقه در اختیار شرکت کنندگان قرار خواهد گرفت. همچنین، تاریخ های مهم و نحوه بارگذاری پاسخ ها متعاقباً از طریق وب سایت و ایمیل ثبت نام اطلاع رسانی خواهد شد.

۴. قالب ورودی / خروجی کد

به منظور حفظ محرمانگی اطلاعات، در مجموعه داده ای که در اختیار شرکت کنندگان قرار می گیرد، مقادیر برخی ستون ها مقیاس شده است، به این معنی که تبدیل های خطی روی آن ها انجام شده و لذا مقادیر مقیاس شده می تواند شامل مقادیر مثبت یا منفی باشد. همچنین برخی ستون ها رمز

شده و برخی دیگر کد شده‌اند. داده‌های ورودی متشکل از n ویژگی در قالب یک فایل CSV در اختیار شرکت‌کنندگان قرار می‌گیرد. توضیحات ویژگی‌ها در بخش دادگان آورده شده است.

نمونه خروجی:

به ازای هر مشترک، می‌بایست مقدار پیش‌بینی حجم اینترنت مصرفی در قالب یک فایل اکسل با سه ستون با نام‌های مشخص شده همانند جدول زیر گزارش گردد.

day	subscriber_id	data_usage_volume
(n)03	fKlZpykjcd	0.56

۵. آیتم‌های مورد نیاز برای شرکت در مسابقه

- گزارش فنی با استاندارد رسمی (Technical Description Paper) TDP
- به‌منظور ارزیابی خروجی‌ها نیاز است تا کلیه فرایندها در قالب کدو اطلاعات مورد نیاز برای بازتولید نتایج تحویل داده شود.
- فایل اکسل پیش‌بینی
- ارائه روش حل مساله و نتایج

۶. نحوه ارزیابی

هر تیم شرکت‌کننده نهایتاً امتیازی بین ۰ و ۱۰۰ کسب می‌کند. تیم‌ها بر اساس امتیاز نهایی رتبه‌بندی می‌شوند. ۱۰ نمره از امتیاز نهایی برای گزارش فنی، ۱۰ نمره برای ارائه فنی و ۸۰ نمره نیز برای مدل یادگیری هر تیم در نظر گرفته شده است. جزئیات بیشتر در خصوص مجازی یا حضوری بودن ارائه‌ها متعاقباً اعلام می‌گردد. نحوه محاسبه ۸۰ امتیاز مدل یادگیری در ادامه بیان شده است.

توابع ارزیابی

- Symmetric Mean Absolute Percentage Error (SMAPE)

فرمول محاسبه این معیار به شکل زیر است.

$$SMAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}$$

که y سری واقعی و \hat{y} سری پیش‌بینی شده است و طول هر دو N است.

• R2

فرمول محاسبه این معیار

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

است که $\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$

هر تیم شرکت کننده که به ترتیب برای معیارهای $SMAPE$ و $R2$ ، مقادیر $smape$ و $r2$ را در مرحله i کسب کرده است ($i \in \{1, 2\}$)، مطابق فرمول زیر در آن مرحله امتیازی بین صفر و صد کسب می کند.

$$score_i = \left(\frac{\max_{teams} SMAPE - smape}{\max_{teams} SMAPE - \min_{teams} SMAPE} \times 0.5 + \frac{r2 - \min_{teams} R2}{\max_{teams} R2 - \min_{teams} R2} \times 0.5 \right) \times 100, \quad i = 1, 2$$

که بیشینه ها و کمینه ها بین تمامی تیم ها گرفته می شود. در انتهای مسابقه، هر تیم شرکت کننده دو امتیاز خواهد داشت. هشتاد درصد از امتیاز نهایی تیمی با دو امتیاز $score_1$ ، $score_2$ در دو مرحله، به شکل زیر محاسبه می شود.

$$\frac{1}{2} \left(\frac{score_1 - \min_{teams} score_1}{\max_{teams} score_1 - \min_{teams} score_1} + \frac{score_2 - \min_{teams} score_2}{\max_{teams} score_2 - \min_{teams} score_2} \right) \times 80,$$

که بیشینه ها و کمینه ها بین تمامی تیم ها در هر مرحله به طور مجزا گرفته می شود.

۷. دادگان

دادگان آموزشی مسابقه توسط مرکز تحقیق و توسعه همراه اول فراهم شده است. این دادگان بخش کوچکی از داده های مصرف اینترنت گروهی از مشترکین منتخب همراه اول در مدت زمان چند ماه متوالی بوده که با حفظ محرمانگی و گمنام بودن در اختیار شرکت کنندگان محترم قرار خواهد گرفت. توصیف ستون های داده در ادامه آورده شده است: لازم به ذکر است که ستون `data_usage_volume` متغیر هدف است.

توصیف	مثال	عنوان ستون در داده
روز ثبت داده (برای نمونه روز دوازدهم از ماه n)	12(n)	day
شناسه مشترک (رمز شده)	3HKsAfC47I	subscriber_ecid
جنسیت مشترک	0	subscriber_gender
سن مشترک	25	subscriber_age

registration_province	17	مکان ثبت نام سیم کارت مشترک (کد شده)
months_of_subscription	100	تعداد ماه های همراهی مشترک
most_used_province	21	مکان بیشترین استفاده مشترک (کد شده)
is_usage_nonzero	1	فلگ فعال بودن مشترک (صفر و یک)
subscriber_total_expenses	0.34778	هزینه کل مشترک (مقیاس شده)
nonpackage_voice_expenses	0.1751	هزینه مکالمات صوتی غیربسته ای مشترک (مقیاس شده)
package_voice_noncash_expenses	1.0236	هزینه بسته های مکالمات صوتی مشترک که هزینه آن ها در صورت حساب مشترک قرار می گیرد (مقیاس شده)
is_voice_expenses_nonzero	1	فلگ نا صفر بودن هزینه مکالمات صوتی مشترک (صفر و یک)
is_voice_usage_nonzero	0	فلگ نا صفر بودن مکالمات مشترک (صفر و یک)
call_in_network_duration	0.6478	مدت زمان مکالمات درون شبکه مشترک (مقیاس شده)
call_off_network_duration	3.049	مدت زمان مکالمات برون شبکه مشترک (مقیاس شده)
nonpackage_call_in_network_expenses	0.8412	هزینه مکالمات غیربسته ای درون شبکه مشترک (مقیاس شده)
nonpackage_call_off_network_expenses	0.4123	هزینه مکالمات غیربسته ای برون شبکه مشترک (مقیاس شده)
total_call_duration	10.2544	مجموع مدت زمان مکالمات مشترک (مقیاس شده)
#inter_operator_calls	20	تعداد تماس های بین اپراتوری مشترک
xyz_score	0.2158	معیار سنجش XYZ مشترک (مقیاس شده)
hxr_score	4.6544	معیار سنجش hxr مشترک (مقیاس شده)
data_cash_expenses	0.034	هزینه بسته های اینترنت مشترک که هزینه آن ها هنگام فعال سازی توسط مشترک پرداخت می شود (مقیاس شده)
nonpackage_data_expenses	12.6574	هزینه اینترنت مصرفی غیربسته ای مشترک (مقیاس شده)
package_data_noncash_expenses	6.1581	هزینه بسته های اینترنت مشترک که هزینه آن ها در صورت حساب مشترک قرار می گیرد (مقیاس شده)
subscriber_data_expenses	51620	مجموع هزینه اینترنت مصرفی مشترک (مقیاس شده)
subscriber_nondata_expenses	4.1974	مجموع هزینه مصرفی غیر اینترنتی مشترک (مقیاس شده)
is_data_expenses_nonzero	1	فلگ نا صفر بودن هزینه اینترنت مشترک (صفر و یک)
is_data_usage_nonzero	0	فلگ نا صفر بودن مصرف اینترنت مشترک (صفر و یک)
#activated_monthly_data_packages	3	تعداد بسته های ماهیانه فعال کاربر
#activated_short_term_data_packages	7	تعداد بسته های کوتاه مدت فعال کاربر
#activated_type_one_data_packages	1	تعداد بسته های نوع یک فعال کاربر
#activated_type_two_data_packages	1	تعداد بسته های نوع دو فعال کاربر
#activated_type_three_data_packages	2	تعداد بسته های نوع سه فعال کاربر
data_usage_volume	5.1694	حجم اینترنت مصرفی مشترک (مقیاس شده)

تذکر: به دلیل اختصاصی بودن دادگان، تمامی شرکت کنندگان موظف به حفظ و نگهداری دقیق از آن بوده و می بایست از به اشتراک گذاری آن پرهیز کنند. ثبت نام در مسابقه به منزله پذیرش مسئولیت نگهداری و عدم انتشار دادگان خواهد بود.

جوایز مسابقه برای تیم‌های برگزیده اول تا سوم به شرح زیر است:

- تیم برگزیده اول: مبلغ ۱۰ میلیون تومان
- تیم برگزیده دوم: مبلغ ۷ میلیون تومان
- تیم برگزیده سوم: مبلغ ۳ میلیون تومان

۹. راه‌های ارتباطی

برای کسب آخرین اطلاعات و اخبار مرتبط با مسابقه، صفحه مسابقه در وب سایت را مطالعه کنید. سوال‌ها، انتقادات و پیشنهادهای خود درباره مسابقات را می‌توانید از طریق ایمیل یا تلفن با ما در میان بگذارید.

Email: aaic@aut.ac.ir

Website: <http://aaic.aut.ac.ir/>

Tel: +98-21-64542700