



الگوی ارسال پیشنهاد پروژه های علم داده

شرکت رایانش سریع هزاره ایرانیان



شناسنامه سند

عنوان سند	الگوی ارسال پیشنهاد پروژه های علم داده
شناسه ی سند	

تأیید و تصویب سند


نام و نام خانوادگی	سمت	تاریخ	امضاء
تهیه کننده	دکتر امین نظارات	۶ بهمن ۱۳۹۹	
	مهدیه معمارزاده	۶ بهمن ۱۳۹۹	
تصویب کننده	دکتر امین نظارات	۱۵ دیماه ۱۳۹۹	


سوابق ویرایش

شماره ویرایش	تاریخ انتشار	خلاصه تغییرات
۱،۰،۰	۱۵ بهمن ماه ۱۳۹۹	آماده سازی اولیه سند

تماس با ما

آدرس پستی	یزد، خیابان آیت الله کاشانی، کوچه ۲۹، شرکت رایانش سریع هزاره ایرانیان، کدپستی ۸۹۱۵۶۶۵۴۸۵
تلفن تماس	۰۳۵۳۶۲۳۲۱۷۶
وبسایت	www.astek.ir
پست الکترونیک	info@astek.ir



	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره تماس: ۰۲۱-۸۸۸۸۸۸۸۸ شرکت رایانش سریع هزاره ایرانیان ارائه دهنده راهکارهای نوین و تخصصی در زمینه های مختلف</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

طرح پیشنهادی "{نام پروژه}"

نکات برجسته طرح پیشنهادی

الگوی طرح پیشنهادی پروژه تهیه شده است تا تمامی افراد و تیمهایی که قصد ارائه پیشنهاد اجرای **پروژه های علم داده** را دارند مطابق با این سند اقدام به تهیه و ارسال پیشنهاد خود نمایند.


شرح دلیل ارائه طرح پیشنهادی

بر اساس مستند درخواست برای پیشنهاد پروژه که در آن مشخصاتی از یک مشکل در سازمان شرح داده شده است این پیشنهادیه را تنظیم نماییم یا اینکه یک مسئله موجود را در دامنه فعالیت سازمان انتخاب کنید که می توانید با استفاده از الگوریتم ها و تکنیک های یادگیری ماشین آن را حل کنید. مطمئن باشید که تمام منابع مورد نیاز (مانند مجموعه داده ها، ورودی ها و تحقیقات) برای تکمیل این پروژه وجود داشته یا جمع آوری شده اند و هر جا لازم باشد در پیشنهاد خود نقل قول های مناسب را ذکر کنید.

الزامات نرم افزار

پروژه پیشنهادی شما باید در پایتون ۳ به بالا نوشته شود.

با توجه به آزاد بودن فرمت پروژه های یادگیری ماشین، نرم افزار و کتابخانه هایی که برای تکمیل موفقیت آمیز کار خود به آنها نیاز خواهید داشت، بسته به حوزه برنامه انتخاب شده و تعریف مسئله متفاوت خواهند بود. به همین دلیل، ضروری است که کلیه نرم افزارها و کتابخانه های لازم که در نظر دارید در پروژه استفاده کنید، به طور مستند قابل دسترسی بوده و ذکر شوند. لطفاً توجه داشته باشید که از نرم افزار اختصاصی، نرم افزاری که به مجوزهای خصوصی نیاز دارد باید اجتناب کنید. توجه داشته باشید که تکنولوژی ها، ابزارها و کتابخانه های مورد استفاده یا بایستی از میان لیست درج شده در


	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره ثبت: ۱۹۰۰۴۵۵۵۲۶۶۰ شرکت رایانش سریع هزاره ایرانیان مرکز تخصصی راهکارهای هوش مصنوعی راهکارهای نوین در زمینه علم داده</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

مستند درخواست برای ارائه پیشنهاد انتخاب شوند یا اینکه با آنها سازگار باشند، در صورتیکه پیشنهادی به جز لیست ذکر شده دارید لازم است که با ذکر دلیل، شرحی از آن را بیان کنید.

پیشنهاد پروژه بر اساس پیشنهاد کتبی ارائه شده ارزیابی می شود. علاوه بر این، بسته به پروژه پیشنهادی، مواد دیگر مانند داده های مورد استفاده مورد ارزیابی قرار خواهد گرفت. انتظار می رود که این پیشنهاد شامل جزئیات، مستندات، تجزیه و تحلیل و بحث کافی باشد تا بتواند کار را برای انجام پروژه منعکس کند. به همین دلیل، بسیار مهم است که پروپوزال به صورت حرفه ای و استاندارد نوشته شود، بنابراین کسانی که پروپوزال پروژه را بررسی می کنند، می توانند به طور واضح هر یک از اجزای پروژه را در گزارش شناسایی نمایند. بدون یک پیشنهاد صحیح نوشته شده، پروژه نمی تواند به اندازه کافی واکاوی و ارزیابی شود. در ادامه **یک الگوی طرح پیشنهادی پروژه** فراهم شده است که بر اساس طرح کلی آن پیشنهاد پروژه تنظیم و برای شرکت ارسال شود.

یک پیشنهاد پروژه شامل **هفت** نکته اساسی است:

- پیشینه دامنه پروژه: زمینه تحقیقاتی که از آن پروژه استخراج شده است.
- بیان مسئله: یک مشکل مورد بررسی قرار گرفته است که راه حلی برای آن تعریف خواهد شد.
- مجموعه داده ها و ورودی ها: داده ها یا ورودی هایی که برای حل مشکل استفاده می شوند.
- بیان راه حل: راه حل پیشنهادی برای مسئله ارائه شده.
- مدل بنچمارک: چند مدل ساده یا تاریخی برای مقایسه راه حل تعریف شده.
- مجموعه ای از معیارهای ارزیابی: نمایش عملکردی برای چگونگی اندازه گیری راه حل.
- طرح کلی پروژه: چگونگی توسعه راه حل و نتایج بدست آمده.

	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره تماس: ۰۲۱-۴۵۵۵۲۴۴۰ شرکت رایانش سریع هزاره ایرانیان مرکز تخصصی آموزش و مشاوره در زمینه های مختلف رایانش سریع - هوش مصنوعی - علم داده</p>
تهیه کننده: دکتر امین نظارات - شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

پیشینه دامنه مسئله


در این بخش، جزئیات مختصری درباره اطلاعات پیشینه دامنه ای که پروژه از آن پیشنهاد شده است، ارائه دهید. اطلاعات تاریخی مربوط به پروژه باید درج شود. باید روشن شود که چگونه یا چرا یک مشکل در دامنه قابل حل است یا باید حل شود. تحقیقات آکادمیک مرتبط باید در این بخش به درستی ذکر شود، از جمله این که چرا این تحقیقات مرتبط هستند.

بیان مسأله

در این بخش، مشکلی که قرار است حل شود را به روشنی توصیف کنید. مسئله توصیف شده باید کاملاً مشخص باشد و حداقل یک راه حل بالقوه مرتبط داشته باشد. علاوه بر این، مسئله را کاملاً توصیف کنید به گونه ای که مشخص شود مسئله قابل سنجش (مسئله را می توان با اصطلاحات ریاضی یا منطقی بیان کرد)، قابل اندازه گیری (مسئله را می توان با استفاده از برخی معیارها اندازه گیری و به وضوح نتایج را مشاهده کرد) و قابل تکرار است (مسئله می تواند دوباره تولید شده و بیش از یک بار رخ دهد).

مجموعه داده ها و ورودی ها

هر پروژه یادگیری ماشین مطمئناً به نوعی مجموعه داده یا ساختار داده ورودی (پرونده های متنی ورودی، تصاویر و غیره) نیاز دارد. مشابه نیازمندی های نرم افزاری، داده هایی که در نظر می گیرید باید در چارچوب موجودی داده ذکر شده توسط کارفرما باشد. داده های خصوصی یا اختصاصی نباید بدون

	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شرکت رایانش سریع هزاره ایرانیان فراتر از دانش و فناوری رایانش سریع</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

اجازه استفاده یا منتشر شوند. اگر داده های مورد نظر با معیارهای خیلی بزرگ بودن متناسب است، به این موضوع توجه کنید که آیا می توانید به جای آن با زیرمجموعه ای از داده ها کار کنید یا نمونه ای آزمایشی از داده ها را به کار برید.


در این بخش، مجموعه داده (ها) و / یا ورودی (های) در نظر گرفته شده برای پروژه باید کاملاً شرح داده شود، از جمله نحوه ارتباط آنها با مشکل و دلیل استفاده از آنها. اطلاعاتی مانند چگونگی بدست آوردن مجموعه داده یا ورودی و مشخصات مجموعه داده یا ورودی باید در صورت لزوم همراه با منابع و ارجاعات مربوطه ذکر شود، باید مشخص شود که مجموعه داده (ها) یا ورودی (ها) چگونه در پروژه استفاده خواهند شد و اینکه آیا با توجه به زمینه مساله استفاده از آنها مناسب است یا خیر.

بیان راه حل

در این بخش، راه حل مسئله را به روشنی شرح دهید. این راه حل باید برای دامنه پروژه قابل اجرا و برای مجموعه داده ها یا ورودی های داده شده مناسب باشد. علاوه بر این، راه حل را کاملاً توصیف کنید به گونه ای که مشخص شود راه حل قابل سنجش (راه حل را می توان با اصطلاحات ریاضی یا منطقی بیان کرد)، قابل اندازه گیری (راه حل را می توان با برخی معیارها اندازه گیری و به وضوح مشاهده نمود) و قابل تکرار است (راه حل می تواند دوباره تولید شده و بیش از یک بار رخ دهد).

مدل بنچمارک

در این بخش، جزئیات مربوط به یک مدل یا نتیجه بنچمارک را ارائه دهید که مربوط به دامنه، بیان مسئله و راه حل مورد نظر باشد. در حالت ایده آل، مدل معیار یا نتیجه، متدهای موجود یا اطلاعات شناخته شده در حوزه و مسئله داده شده را مبنا قرار می دهد، که می تواند به صورت عینی با

	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره تماس: ۰۲۱-۴۳۳۳۳۳۳۳ شرکت رایانش سریع هزاره ایرانیان ارائه دهنده راهکارهای نوین در زمینه های مختلف رایانش ابری - هوش مصنوعی - داده کاوی</p>
تهیه کننده: دکتر امین نظارات - شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

راه حل مقایسه شود. با جزئیات کامل توضیح دهید که چگونه مدل یا نتیجه معیار قابل اندازه گیری است (می توان آن را با یک معیار اندازه گیری و به وضوح مشاهده نمود).

معیارهای ارزیابی^۱


در این بخش، حداقل یک معیار ارزیابی را پیشنهاد دهید که می تواند برای تعیین کمیت عملکرد هر دو مدل معیار و مدل راه حل مورد استفاده قرار گیرد. معیار(های) ارزیابی پیشنهادی شما باید با توجه به زمینه داده ها، بیان مسئله و راه حل مورد نظر متناسب باشد. نحوه استخراج معیار(های) ارزیابی را توصیف کرده و مثالی از نمایش های ریاضی آنها (در صورت وجود) ارائه دهید. معیارهای ارزیابی پیچیده باید کاملاً مشخص و قابل اندازه گیری باشند (می توان آنها را با عبارات ریاضی یا منطقی بیان کرد). این معیارها باید متناسب با KPI و متریک های تجاری تعیین شده (در صورت وجود) در بخش سرفصل مسائل باشد. در این بخش لازم است که تشریح شود که برای ارزیابی مدل اجرا شده و الگوریتم نهایی متناسب با KPI تجاری توافق شده (یا مندرج در فراخوان) چه متریکی متناظر با آن در مدل لحاظ شده است و این نگاشت چگونه انجام شده است. فرآیند ارزیابی مدل می تواند شامل چندین مرحله باشد، مرحله ای برای اخذ تاییدیه اولیه با داده ها آزمایشی و الگوریتمهای ساده که به آن اثبات نظریه گفته می شود (Proof of Concept) و مرحله ای در بخش پایانی پروژه برای دریافت بازخوردهای ذینفعان و کاربران مدل به صورت مدت دار با امکان

۳^۱ کلید برای اندازه گیری موفقیت آمیز ROI

تعریف شاخص های اصلی عملکرد (KPI): اغلب اوقات، این امر در مرحله آموزش ML اجباری است زیرا معمولاً معیاری وجود دارد که می خواهید آن را بهینه کنید. با این حال، به عنوان مثال در مورد سگمنتیشن، باید بدانید که بخشها برای چه مواردی استفاده می شوند و معیارهای موفقیت بر اساس آن تعریف می شوند. معیارهای تعریف شده ممکن است بر اساس بخش نیز متفاوت باشند (به عنوان مثال، افزایش متوسط هزینه برای بخش بهترین مشتریان و بهبود نرخ نگهداری برای بخش مشتریان متوسط).

داشتن بنچمارک مقایسه: پس از مشخص شدن معیارهای عملکرد، رصد گروه های دارای خطا (به عنوان مثال مشتریانی که بهترین محصول یا خدمت بعدی پیشنهاد شده توسط آن الگوریتم ML را انتخاب نمی کنند) برای درک تأثیر واقعی مدل های ML و AI مهم هستند. ده ها سال است که بازاریابان موفقیت در مواجهه با مشتری را با این روش می سنجند و اصول هنوز هم پابرجا هستند.

مانیتورینگ در طی زمان: بسیاری از الگوریتم ها به طور منظم با در دسترس قرار دادن اطلاعات آموزش بیشتر به روز می شوند. این موضوع می تواند یک مزیت و یا حتی یک نقص باشد وقتی که صحبت از اجرای یک مدل ML می شود، زیرا مدل ها می توانند به مرور تغییر کنند و برای اطمینان از دستیابی به اهداف و معیارهای موفقیت برای مراقبت از تغییرات داده ها، باید آنها را کنترل کرد. حتماً ممیزی داده ها را انجام دهید. اگر داده های اصلی در حال تغییر باشد، الگوریتم های شما تغییر کرده و به طور بالقوه ممکن است تخریب شوند. این امر به ویژه پس از آنکه مدل های ML را در مقیاس کامل راه اندازی کرده اید، صادق است. برای نظارت بر مدل ها و آموزش مجدد و اندازه گیری تأثیر مکرر، یک فرآیند دقیق در نظر بگیرید. رویکرد دو گام به جلو، یک گام به عقب برای اطمینان از اینکه KPI شما در مسیر درست حرکت می کند معتبر است.


	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شرکت رایانش سریع هزاره ایرانیان فراتر از دانش و فناوری رایانش سریع</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

دریافت بازخورد و افزودن بازخوردهای دریافتی به مدل در قالب ویژگی جدید یا تغییر در پارامترهای مدل. از آنجا که سنجش معیارها بر اساس شاخصهای مورد نظر واحدهای ذینفع مبتنی بر KPI انجام می شود و برای دستیابی به سطح قابل قبولی از متریک تعریف شده نیازمند طی زمان مشخصی می باشد، در این پیشنهاد لازم است که به دقت به فرآیند اخذ بازخوردها اعم از روشهای ماشینی یا دستی و فرایند باز-آموزش مدل مبتنی بر دریافت بازخوردها اشاره دقیق شده و رابطه KPI با معیار ارزیابی الگوریتم که می تواند یکی از موارد زیر باشد و مقدار مورد انتظار آن در ابتدا و انتهای دوره بازخوردگیری پرداخته شود:

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient
8. Concordant – Discordant Ratio
9. Root Mean Squared Error
10. Mean absolute error (MAE)
11. Cross Validation (Not a metric though!)
12. Specificity
13. ...

طراحی کلی پروژه

در این بخش نهایی، با توجه به مساله یک گردش کار نظری را برای نزدیک شدن به یک راه حل خلاصه کنید. بحث کاملی در مورد اینکه چه راهکارهایی ممکن است استفاده کنید، چه تجزیه و تحلیل داده هایی ممکن است قبل از استفاده مورد نیاز باشد یا کدام الگوریتم ها برای پیاده سازی توسط شما در نظر گرفته شده است.

	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره تماس: ۰۲۱-۴۵۵۵۴۴۴۴ شرکت رایانش سریع هزاره ایرانیان گروه مشاوران و پیمانکاران رایانش و هوش مصنوعی</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

روند کار و بحثی که ارائه می دهید باید با مباحث بخشهای قبلی مطابقت داشته باشد. علاوه بر این، به شما توصیه می شود که برای کمک به توصیف طراحی پروژه، مصورسازی های اولیه، شبه کد یا نمودارها را در اختیار داشته باشید، اما الزامی نیست. بحث باید به وضوح گردش کار مورد نظر شما را برای پروژه "نام پروژه" مشخص کند.

ارسال پیشنهادیه پروژه


ارزیابی

پیشنهاد پروژه توسط یک تیم ارزیابی در شرکت در مقایسه با نیازمندیهای مطرح شده در سرفصلهای مسائل و سایر موارد فنی بررسی خواهد شد. حتماً این مبحث را کاملاً بررسی کرده و قبل از ارسال، پروژه خود را ارزیابی نمایید. تمام معیارهای ذکر شده در مستندات پروژه که توسط کارفرما تهیه شده اند باید در پیشنهاد پروژه شما لحاظ شده باشد و روشهایی برای ارضا کردن آن نیازمندیها ارائه شده باشد.

فایل های ارسالی

حداقل بایستی فایل های زیر در لیست ارسالی وجود داشته باشد. اگر روش انتخابی شما برای تحویل، آرشیو است (* .zip)، لطفاً اندازه کلی فایل را در نظر بگیرید. فایل باید شامل موارد زیر باشد:

- یک طرح پیشنهادی پروژه، فقط در قالب PDF و Doc، با نام **Proposal.{pdf,doc,docx}**، که به هر یک از هفت نکته اصلی یک طرح پیشنهادی می پردازد.
- هرگونه ارقام پشتیبانی کننده اضافی مانند مجموعه داده ها، مشخصات ورودی که برای پروژه و پیشنهاد لازم است (در صورت لزوم ضمیمه شوند یا مشخصات آنها ذکر شود- برای دیتاست هایی که در اختیار سازمان است).

	الگوی ارسال پیشنهاد پروژه های علم داده	 <p>شماره ملی: ۱۹۰۰۴۳۵۵۲۴۶۶ شرکت رایانش سریع هزاره ایرانیان طراحی و توسعه سیستم های هوشمند راهکارهای نوین در زمینه های نوین</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

قبل از ارائه طرح پیشنهادیتان، موارد زیر را چک کنید. . .

- آیا پیشنهادی که نوشته اید از ساختار منظمی شبیه به الگوی پروژه پیروی می کند؟
- آیا هر بخش (به ویژه بیان راه حل و طراحی پروژه) به صورت واضح، مختصر و مشخص نوشته شده است؟ آیا اصطلاحات یا عباراتی مبهم وجود دارد که نیاز به توضیح داشته باشد؟
- آیا مخاطبان مورد نظر پروژه شما می توانند پیشنهاد شما را درک کنند؟
- آیا پیشنهاد خود را به درستی تصحیح کرده اید تا اطمینان حاصل کنید که حداقل اشتباهات دستوری و املایی وجود دارد؟
- آیا تمام منابع استفاده شده برای این پروژه به درستی ذکر و ارجاع داده شده اند؟