



الگوی تهیه گزارش پایانی پروژه های علم داده

شرکت رایانش سریع هزاره ایرانیان



شناسنامه سند

عنوان سند	الگوی تهیه گزارش پایانی پروژه های علم داده
شناسه ی سند	

تأیید و تصویب سند


نام و نام خانوادگی	سمت	تاریخ	امضاء
تهیه کننده	دکتر امین نظارات	۴ بهمن ماه ۱۳۹۹	
	مهدیه معمارزاده	۴ بهمن ماه ۱۳۹۹	
تصویب کننده	دکتر امین نظارات	۱۵ بهمن ماه 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
تلفن تماس	۰۳۵۳۶۲۳۲۱۷۶
وبسایت	www.astek.ir
پست الکترونیک	info@astek.ir



	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شماره ثبت: ۱۹۰۰۴۵۵۵۲۴۶۰ شرکت رایانش سریع هزاره ایرانیان طراحی و توسعه سیستم های هوشمند طراحی و توسعه سیستم های داده های بزرگ طراحی و توسعه سیستم های یادگیری ماشین</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

گزارش پروژه "نام پروژه"

۱. تعریف

بررسی اجمالی پروژه


در این بخش، به دنبال ارائه یک نمای کلی از سطح پروژه به زبان ساده باشید. سؤالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا نمای کلی از پروژه مانند دامنه مشکل، ریشه پروژه و مجموعه داده های مرتبط یا داده های ورودی ارائه شده است؟
- آیا اطلاعات پیش زمینه به اندازه کافی داده شده است که یک خواننده بی اطلاع دامنه مشکل را درک کند و شرح مشکل را دنبال نماید؟

بیان مسأله پروژه

در این بخش، شما می خواهید مشکلی را که در تلاش برای حل آن هستید، از جمله استراتژی (رئوس مطالب) که برای دستیابی به راه حل مورد نظر استفاده می کنید را به وضوح مشخص نمایید. همچنین باید به طور کامل در مورد راه حل مورد نظر برای این مشکل بحث و تبادل نظر کنید. سؤالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا بیان مسئله به وضوح تعریف شده است؟ آیا خواننده متوجه خواهد شد که شما چه چیزی را می خواهید حل کنید؟
- آیا به طور کامل در مورد چگونگی تلاش برای حل مشکل بحث کرده اید؟
- آیا راه حل پیش بینی شده به وضوح تعریف شده است؟ آیا خواننده می فهمد که به دنبال چه نتایجی هستید؟

	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شماره ثبت: ۱۳۰۰۴۵۵۵۲۴۴۰ شرکت رایانش سریع هزاره ایرانیان ارائه دهنده راهکارهای نوین در زمینه رایانش سریع و هوش مصنوعی</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

متریک ها


در این بخش، باید متریک ها یا محاسباتی که برای اندازه گیری عملکرد یک مدل یا نتیجه در پروژه خود مورد استفاده قرار می دهید را به روشنی تعریف کنید. این محاسبات و متریک ها باید براساس ویژگی های دامنه مسئله و خود مسئله باشند. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا متریک هایی که برای سنجش عملکرد مدل های خود انتخاب کرده اید به طور واضح مورد بحث و تعریف قرار گرفته اند؟
- آیا توجیه منطقی برای متریک های انتخاب شده براساس مشکل و راه حل ارائه داده اید؟
- آیا متریک های انتخاب شده می توانند رابطه منطقی را با KPI های تجاری تعیین شده برای پروژه برقرار کنند؟
- کدام متریک(ها) به کدام KPI(ها) مرتبط است؟

۲. تحلیل و بررسی(آنالیز)

شناسایی داده

در این بخش، انتظار می رود داده هایی را که برای این مشکل مورد استفاده قرار می دهید، تجزیه و تحلیل کنید. این داده ها می توانند به صورت یک مجموعه داده (یا مجموعه داده ها)، داده ورودی (یا فایل های ورودی) یا حتی یک محیط باشند. نوع داده ها باید کاملاً توصیف شده و در صورت امکان، آمار و اطلاعاتی اساسی ارائه شود (مانند بحث در مورد ویژگی های ورودی یا تعریف خصوصیات مربوط به ورودی یا محیط). هر گونه ناهنجاری یا خصوصیات جالب در مورد داده هایی که ممکن است نیاز به پرداختن داشته باشند شناسایی می شوند (مانند ویژگی هایی که باید تغییر شکل داده شوند یا Outlier هستند). سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

	الگوی تهیه گزارش پایانی پروژه های علم داده	
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir


- اگر یک مجموعه داده برای این مشکل وجود دارد، آیا شما در مورد برخی از ویژگی ها در مورد مجموعه داده کاملاً بحث کرده اید؟ آیا نمونه داده ای در اختیار خواننده قرار گرفته است؟
- اگر یک مجموعه داده برای این مشکل وجود دارد، آیا آمار مربوط به مجموعه داده محاسبه و گزارش شده است؟ آیا نتایج مرتبط با این محاسبه مورد بحث قرار گرفته است؟
- اگر یک مجموعه داده برای این مشکل وجود ندارد، آیا در مورد فضای ورودی یا داده های ورودی برای مشکل شما بحث شده است؟
- آیا در مورد فضای ورودی یا مجموعه داده ها ناهنجاری یا ویژگی هایی وجود دارد که باید برطرف شود؟ (*categorical variables, missing values, outliers, etc.*)

مصورسازی اکتشافی

در این بخش، شما باید نوعی بصری سازی را ارائه دهید که ویژگی یا خصیصه مربوط به داده ها را خلاصه یا استخراج کند. تجسم باید به اندازه کافی از داده های مورد استفاده پشتیبانی کند. در مورد چرایی انتخاب این تجسم و چگونگی ارتباط آن بحث کنید. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا مشخصه یا ویژگی مربوط به مجموعه داده یا داده های ورودی را مصورسازی کرده اید؟
- آیا مصورسازی انجام شده کاملاً تحلیل و بحث شده است؟
- در صورت ارائه پلات، آیا محورها، عنوان و داده ها به وضوح تعریف شده اند؟


الگوریتم ها و تکنیک ها

	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شماره ثبت: ۱۹۰۰۴۵۵۵۴۴۰ شرکت رایانش سریع هزاره ایرانیان ارائه دهنده راهکارهای نوین و تخصصی در زمینه های مختلف رایانش و هوش مصنوعی</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

- در این بخش، شما باید در مورد الگوریتم ها و تکنیک هایی که قصد دارید برای حل مسئله استفاده کنید بحث کنید. شما باید استفاده از هر یک را بر اساس ویژگی های مسئله و دامنه مشکل توجیه کنید. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:
- آیا الگوریتم های مورد استفاده شما، از جمله متغیرها / پارامترهای پیش فرض در پروژه، به وضوح مشخص شده اند؟
 - آیا تکنیک های مورد استفاده کاملاً بحث و توجیه می شوند؟
 - آیا مشخص شده است که داده های ورودی یا مجموعه داده ها چگونه توسط الگوریتم ها و تکنیک های انتخاب شده اداره می شوند؟

معیار (بنچمارک)


- در این بخش، شما باید یک نتیجه معیار یا آستانه مشخص برای مقایسه عملکردهای بدست آمده از راه حل خود ارائه دهید. استدلال پشت معیار (در موردی که نتیجه قطعی نداشته باشد) باید مورد بحث قرار گیرد. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:
- آیا نتیجه یا مقداری ارائه شده است که به عنوان معیار سنجش عملکرد باشد؟
 - آیا مشخص است که چگونه این نتیجه یا مقدار (چه با استفاده از داده و چه با فرضیه ها) بدست آمده است؟

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

۳. روش شناسی

پیش پردازش داده ها
<p>در این بخش، در صورت لزوم، تمام مراحل پیش پردازش شما باید به وضوح ثبت شود. از بخش قبلی، هر ناهنجاری یا خصوصیتی که در مورد مجموعه داده شناسایی کرده اید، در اینجا بحث و اصلاح می شود. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:</p> <ul style="list-style-type: none"> • اگر الگوریتم های انتخاب شده به مراحل پیش پردازش مانند انتخاب ویژگی یا تبدیل ویژگی نیاز دارند، آیا آنها به درستی مستند شده اند؟ • براساس بخش اکتشاف داده، اگر ناهنجاری ها یا خصوصیتی وجود داشت که باید برطرف و اصلاح شوند، آیا آنها به درستی اصلاح شده اند؟ • اگر نیازی به پیش پردازش نیست، مشخص شده است که چرا؟

پیاده سازی
<p>در این بخش، فرایندی که برای معیارها، الگوریتم ها و تکنیک هایی که برای داده های داده شده پیاده سازی کرده اید، باید به صورت واضح مستندسازی شوند. باید چگونگی اجرای آن کاملاً روشن باشد و در مورد هرگونه عارضه ای که در طی این فرآیند رخ داده است باید بحث شود. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:</p> <ul style="list-style-type: none"> • آیا مشخص شده است که الگوریتم ها و تکنیک ها با مجموعه داده های داده شده یا داده های ورودی چگونه پیاده سازی شده اند؟ • آیا معیارها یا تکنیک های اصلی قبل از دستیابی به راه حل تغییراتی داشتند؟ • آیا بخشی از فرایند کدنویسی (به عنوان مثال، نوشتن توابع پیچیده) وجود دارد که باید مستند شود؟

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

اصلاح

در این بخش، شما باید در مورد روند بهبودی که در الگوریتم ها و تکنیک هایی که در اجرای خود استفاده کرده اید، صحبت کنید. به عنوان مثال، تنظیم پارامترها برای مدل های خاص جهت دستیابی به راه حل های بهبود یافته در گروه پالایش قرار می گیرد. راه حل های اولیه و نهایی شما و همچنین نتایج متوسط قابل توجه در صورت لزوم باید گزارش شوند. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:


- آیا راه حل اولیه پیدا شده و به طور واضح گزارش شده است؟
- آیا روند بهبود به روشنی مستند شده است، مانند اینکه چه تکنیک هایی استفاده شده اند؟
- آیا با بهبود روند، راه حل های متوسط و نهایی به طور واضح گزارش می شوند؟

۴. نتایج

ارزیابی و اعتبارسنجی مدل

در این بخش، مدل نهایی و هرگونه کیفیت مورد انتظار برای مدل باید به طور دقیق ارزیابی شود. باید مشخص شود که مدل نهایی چگونه استخراج شده و چرا این مدل انتخاب شده است. علاوه بر این، برای تأیید قوت این مدل و راه حل آن، باید از نوعی آنالیز مانند دستکاری داده های ورودی یا محیط برای دیدن چگونگی تأثیر راه حل مدل (به این روش تحلیل حساسیت می گویند) استفاده شود. در اینجا بایستی در مورد نحوه ارتباط و سنجش نتایج حاصله و اثر آن بر KPI های تجاری پروژه شرح داده شود. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا مدل نهایی منطقی و مطابق با انتظارات راه حل است؟ آیا پارامترهای نهایی مدل مناسب هستند؟
- آیا مدل نهایی با ورودی های مختلف آزمایش شده است تا ارزیابی شود که آیا این مدل به خوبی داده های دیده نشده (*unseen*) را تعمیم می دهد یا خیر؟


	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شرکت رایانش سریع هزاره ایرانیان راهبر و ارائه دهنده راهکارهای نوین در زمینه های نوین</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

- آیا این مدل برای این مشکل به اندازه کافی قوی است؟ آیا اغتشاشات (تغییرات) کوچک در داده های آموزشی یا فضای ورودی تأثیر زیادی بر نتایج دارند؟
- آیا می توان به نتایج بدست آمده از مدل اعتماد کرد؟
- از چه راه حلی برای دریافت بازخورد از کاربران و ذینفعان استفاده شود؟ آیا این راه حل می تواند مکانیزه بوده و اثر خود را در مدل بگذارد؟
- چگونه می توان تفسیر مناسبی از نتایج و پیش بینی های مدل پیاده سازی شده را در اختیار ذینفع قرار داد؟ آیا برای تفسیر پذیری روشی پیاده سازی شده است که به مرور زمان هم برای ذینفعان قابل استفاده باشد و هم امکان ارسال بازخوردها را به مدل داشته باشد (Explainability)؟

توجیه

در این بخش، راه حل نهایی مدل شما و نتایج آن باید با معیاری که قبلاً در پروژه ایجاد کرده اید با استفاده از نوعی تجزیه و تحلیل آماری مقایسه شود. همچنین باید توجیه کنید که آیا این نتایج و راه حل به اندازه کافی قابل توجه است که بتواند مشکل پیش آمده در پروژه را برطرف کند یا خیر. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا نتایج نهایی قوی تر از نتیجه معیار گزارش شده قبلی هستند؟
- آیا شما به طور کامل راه حل نهایی را تحلیل و بحث کرده اید؟
- آیا راه حل نهایی آنقدر قابل توجه است که بتواند مشکل را حل کند؟

	الگوی تهیه گزارش پایانی پروژه های علم داده	
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

۵. نتیجه گیری

مصورسازی با فرم آزاد


در این بخش، شما باید نوعی مصورسازی را ارائه دهید که بر کیفیت مهمی در مورد پروژه تأکید دارد. این فرم بسیار آزادتر از مصورسازی در مراحل قبل است و با هر روش بصری سازی قابل ارائه است، اما باید منطقی از یک نتیجه یا مشخصه قابل توجه در مورد مسئله ای که می خواهید در مورد آن بحث کنید، باشد. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا شما یک کیفیت مهم یا مناسب را در مورد مشکل، مجموعه داده، داده های ورودی یا نتایج مصورسازی کرده اید؟
- آیا مصورسازی کاملاً تحلیل و بحث شده است؟
- در صورت ارائه طرح، آیا محورها، عنوان و داده ها به وضوح مشخص شده اند؟

بازتاب

در این بخش، شما به صورت End-to-End کل مسیر دستیابی به راه حل را خلاصه خواهید کرده و یک یا دو جنبه خاص پروژه که جالب یا دشوار بود را مورد بحث قرار خواهید داد. انتظار می رود که در کل پروژه تأمل کنید تا نشان دهید که درک کاملی از کل فرآیند کار دارید. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:

- آیا کل فرآیندی که برای این پروژه استفاده کرده اید را به طور کامل خلاصه کرده اید؟
- آیا جنبه های جالبی از این پروژه وجود داشت؟
- آیا جنبه های دشواری از این پروژه وجود داشت؟

	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شرکت رایانش سریع هزاره ایرانیان راهکارهای نوین در زمینه علم داده www.astek.ir</p>
تهیه کننده: دکتر امین نظارات – شرکت رایانش سریع هزاره ایرانیان		www.astek.ir


- آیا مدل و راه حل نهایی انتظارات شما را برای این مشکل برآورده می کند و آیا باید برای حل این نوع مشکلات در یک شرایط کلی استفاده شود؟

بهبود

- در این بخش، باید در مورد چگونگی بهبود یک جنبه از پیاده سازی که طراحی کرده اید، بحث شود. به عنوان مثال، روش هایی را که می توان پیاده سازی را به صورت کلی تر انجام داد و مواردی که باید اصلاح شود را در نظر بگیرید. نیازی به انجام این بهبود نیست، اما راه حل های بالقوه حاصل از این تغییرات در نظر گرفته شده و با راه حل فعلی تطبیق/مقایسه می شود. سوالاتی که می توانید هنگام نوشتن این بخش از خود بپرسید:
- آیا بهبودهای بیشتری وجود دارد که می تواند در الگوریتم ها یا تکنیک هایی که در این پروژه استفاده کرده اید انجام شود؟
 - آیا الگوریتم ها یا تکنیک هایی که در مورد آنها تحقیق کردید مواردی وجود داشته است که نمی دانید چگونه آنها را پیاده سازی کنید، اما اگر می دانستید چگونه استفاده می کردید؟
 - اگر از راه حل نهایی خود به عنوان معیار جدید استفاده کرده اید، فکر می کنید راه حل بهتری هم وجود دارد؟

قبل از ارسال، از خود بپرسید . .

- آیا گزارش پروژه ای که نوشتید از ساختار منظمی شبیه ساختار پروژه پیروی می کند؟
- آیا هر بخش (به ویژه تحلیل و روش) به صورت واضح، مختصر و خاص نوشته شده است؟ آیا اصطلاحات یا عباراتی مبهم وجود دارد که نیاز به توضیح داشته باشد؟

	الگوی تهیه گزارش پایانی پروژه های علم داده	 <p>شماره ملی: ۱۹۰۰۴۳۵۵۲۴۶۰ شرکت رایانش سریع هزاره ایرانیان توسعه دهنده خدمات رایانش سریع و ابر رایانش ابری - ابر - ابر</p>
تهیه کننده: دکتر امین نظارات - شرکت رایانش سریع هزاره ایرانیان		www.astek.ir

- آیا مخاطبان مورد نظر پروژه شما می توانند تجزیه و تحلیل، روش ها و نتایج شما را درک کنند؟
- آیا گزارش پروژه خود را به درستی خوانده 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 است؟
- آیا کدی که راه حل شما را پیاده سازی می کند به راحتی قابل خواندن و توضیح هست؟
- آیا کد بدون خطا اجرا می شود و نتایج مشابه نتایج گزارش شده تولید می کند؟