**بسمه تعالي**

**تاریخ: ۰۱/۰۳/۱۴۰۳**

**شماره: ۵**

**دانشکده مهندسی کامپیوتر**

**گروه هوش مصنوعی - نرم افزار**

**فرم پیشنهاده پروژه پاياني - دوره كارشناسی**

**عنوان پروژه(فارسي):**

**ارائه یک الگوریتم برای بهبود مدل‌های تفسیر‌پذیر بر پایه درخت تصمیم**

**عنوان پروژه (لاتين):**

**Presenting a New Algorithm for Improving Interpretable Models based on Decision Trees**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | نام دانشجو | شماره دانشجويي | گرايش | تعداد واحدهاي گذرانده شده | امضاء |
| 1 | سید حسین حسینی دولت آبادی | 4012363328 | هوش مصنوعی  نرم افزار | 85 |  |
| 2 |  |  |  |  |  |

|  |  |
| --- | --- |
| استاد راهنماي پروژه: دکتر حمید رضا برادران | امضاء استاد راهنما - تاریخ |
| نظر استاد راهنما: |

پیشنهاده در تاريخ در شوراي گروه مهندسی هوش مصنوعی - نرم افزار مطرح گرديد و

□ بدون تغییر مورد تصويب قرار گرفت.

□ با شرايط زير مورد تصويب قرار گرفت.

□ به دلايل زير مورد تصويب قرار نگرفت.

نام عضو هیئت علمی بررسی کننده: دکتر سید فخرالدین نوربهبهانی تاریخ و امضا

1. **معرفي موضوع و بيان اهداف پروژه:**

مدل‌های تفسیر‌پذیر، مخصوصا مدل‌های بر پایه درخت تصمیم، نقش حیاتی را در زمینه‌های مختلف مانند: اقتصاد، مراقبت‌های سلامت و امنیت سایبری ایفا می‌کند. مشخصه‌ی قابل بیان این مدل‌ها در فراهم آوردن یک بینش شفاف برای فرآیند تصمیم گیری می‌باشد. وجود تفسیر‌پذیری، حتی در سطح محدود، باعث افزایش اعتماد کاربر به مدل می‌شود. به طوری که در زمینه‌هایی که تصمیمات مستقیما بر جان و سلامت انسان‌ها یا اطلاعات حساس، اثر‌گذار است؛ نیاز به مدل‌های تفسیر‌پذیر بیش از پیش به چشم می‌آید.

قابلیت تفسیر‌پذیری و توضیح مدل، یک برتری برای آن محسوب می‌شود. فرقی نمی‌کند تصمیم‌گیری در حوزه‌ی پیش‌بینی نتایج آزمایش یک بیمار باشد، یا در حوزه‌ی تشخیص یک تراکنش جعلی و یا نشخیص تهدید‌های امنیتی احتمالی در زمینه امنیت سایبری. هدف ما از انجام این پژوهش، ارائه‌ی یک الگوریتم برای بهبود مدل‌های تفسیر‌پذیر بر پایه درخت تصمیم است.

امید است که با توسعه‌ی یک الگوریتم خلاق که هم تفسیر‌پذیری و هم دقت پیش‌بینی را بهبود ببخشد، فاصله‌ی بین پیچیدگی تصمیم‌گیری مدل و درک انسان کاهش یابد.

به طور خلاصه، پروژه‌ی ما بر آن است تا با استفاده از پتانسیل زیاد درختان تصمیم، شفافیت و کارآیی مدل‌های مبتنی بر هوش مصنوعی را افزایش دهیم.

1. **روش انجام پروژه را توضيح دهيد.**

روش انجام این پروژه، به طور کلی به هفت بخش تقسیم می‌شود، که به شرح زیر است:

1. بررسی و مرور ادبیات موضوع:

* بررسی مقالات و تحقیقات موجود در زمینه‌های مدل‌های تفسیر‌پذیر، مخصوصا مدل‌های بر پایه‌ی درخت‌های تصمیم.
* مشخص کردن تحقیقات کلیدی در این زمینه.
* تعیین مشکلات عمومی و محدودیت‌هایی که در روش‌های فعلی وجود دارد.

1. مشخص کردن خصوصیات جزئی مسئله:

* تعیین چالش ها و محدودیت‌های خاص مربوط به مسئله.
* تعریف محدوده و اهدافی که الگوریتم باید بتواند آن‌ها را به‌طور موثر پاسخ دهد.

1. توسعه چارچوب مفهومی:

* فرمول‌بندی یک چارچوب مفهومی برای الگوریتم جدید با جمع‌آوری داده‌ها از مرحله‌ی یک.
* تعریف قوانین راهنما برای الگوریتم با در نظر گرفتن تفسیر‌پذیری، دقت و مقیاس پذیری.

1. طراحی الگوریتم:

* توسعه معماری الگوریتم بر پایه‌ی اطلاعات مرحله‌ی سوم.
* استفاده از تکنیک‌های یادگیری ماشین، بهینه‌سازی و نظریه‌ی اطلاعات برای افزایش تفسیر‌پذیری در عین حفظ دقت پیش‌بینی.

1. پیاده‌سازی:

* پیاده‌سازی الگوریم طراحی شده با به کارگیری برنامه‌نویسی و فریم‌ورک‌های مناسب.
* اطمینان از کارآیی و مقیاس‌پذیری پیاده‌سازی به منظور تطبیق‌پذیری با مجموعه داده‌های مقیاس بزرگ . کاربرد‌های دنیای واقعی.

1. ارزیابی:

* ارزیابی کارآیی الگوریتم با استفاده از مجموعه داده‌های معیار (benchmark datasets) و معیار‌های مناسب.
* مقایسه تفسیر‌پذیری و دقت پیش‌بینی الگوریتم جدید با روش‌های موجود؛ شامل درخت‌های تصمیم سنتی و پیشرفته‌ترین مدل‌های تفسیرپذیر

1. آنالیز نتایج:

* آنالیز نتايج مرحله‌ي ارزيابي با توجه به كارايي الگوريتم در دو بخش تفسیرپذيري و دقت پیش‌بیني.
* تفسیر مفهوم اطلاعات به‌دست آمده براي كاربردهاي بالقوه در سناريوهاي دنیاي واقعي.

1. **آيا اين پروژه و يا مشابه آن قبلاً انجام شده است؟ در صورت مثبت بودن جواب، تفاوت‌هاي اين پروژه با پروژه‌هاي قبلي را بيان نماييد.**

مدل‌هاي قابل تفسیر، به‌ويژه مدل‌هايي كه بر پايه درخت تصمیم ساخته شده‌اند، مورد توجه تحقیقاتي در زمینه يادگیري ماشین بوده‌اند. مطالعات قبلي، با تمركز بر بهبود شفافیت، قابلیت استفاده و كاربرد عملي در كاربردهاي واقعي، روش‌هاي مختلفي را براي بهبود تفسیرپذيري مدل‌هاي بر پايه درخت تصمیم بررسي كرده‌اند.

موارد زير تفاوت‌هاي كلیدي بین پروژه‌ي ما و پروژه‌هاي قبلي را برجسته مي‌كند:

1. نوآوری الگوریتم پیشنهادی:

پروژه ما الگوريتمي نوآورانه را معرفي مي‌كند كه از ابتدا طراحي شده است تا دقت پیش‌بیني را حفظ كرده و در عین حال تفسیرپذيري را افزايش دهد. با بهره‌گیري از پیشرفت‌هايي در زمینه يادگیري ماشین، بهینه‌سازي و نظريه اطلاعات، الگوريتم ما، ويژگي‌ها و روش‌هاي منحصر به فردي را در بر مي‌گیرد؛ كه آن را از روش‌هاي موجود متمايز مي‌كند. به‌طور خاص، ما بر روي تركیبي از تکنیک‌هاي نوآورانه براي استخراج ويژگي، بهینه‌سازي ساختار درخت و يادگیري گروهي تاكید مي‌كنیم.

1. بهبود عملکرد:

بنا بر پايه برداشت‌هاي به‌دست آمده از تحقیقات پیشین، الگوريتم ما براي رفع مشکلات و محدود يت‌هاي خاص موجود در رويکردهاي

كنوني توسعه مدل قابل تفسیر ميشود. با اولويت‌بندي قابلیت تفسیرپذيري بدون از دست دادن دقت پیش‌بیني، الگوريتم ما هدف دارد با

چالش‌هايي مانند اورفیت، پیچیدگي مدل و قابلیت مقیاس‌پذيري مقابله كند. اين اقدام منجر به بهبود كارايي مدل‌هاي بر پايه‌ي درخت

تصمیم در موارد عملي مي‌شود.

1. اعتبارسنجی تجربه:

پروژه ما تاكید زيادي بر اعتبارسنجي تجربي دارد، با يک چارچوب ارزيابي طراحي شده كه اثربخشي الگوريتم پیشنهادي را در كاربردهاي

واقعي ارزيابي ميكند. از طريق آزمايش‌هاي گسترده با استفاده از مجموعه داده‌هاي مرجع و مطالعات موردي از حوزه‌هاي مختلف، ما قصد

داريم كه كارايي و كلیت الگوريتم خود را در كاربردهاي عملي ثابت كنیم.

1. **طرح تجاری خود در رابطه با به سودرسانی پروژه انجام شده را در صورت وجود توضیح دهید؟**

* ارائه يک مدل درآمدي بر اساس اخذ مجوز براي الگوريتم از كسب و كارها، موسسات تحقیقاتي و توسعه دهندگان نرم‌افزاري.
* در نظر گرفتن ارائه دسترسي مبتني بر اشتراک به الگوريتم، با قیمت‌گذاري بر اساس مصرف يا ويژگي‌ها. بررسي فرصت‌هايي براي تأمین درآمد از طريق خدمات مشاوره، كارگاه‌هاي آموزشي، يا توسعه مدل‌هاي سفارشي براي صنايع يا كاربردهاي خاص.
* تنظیم بازار هدف روي صنايع و سازمان‌هايي كه بیشترين بهره را از افزايش قابلیت تفسیرپذيري در مدل‌هاي بر پايه‌ي درخت تصمیم به دست مي‌آورند، مانند مؤسسات مالي، ارائه دهندگان خدمات سلامت و سازمان‌هاي نظارتي.
* در نظر گرفتن استراتژي‌هايي براي پايداري بلندمدت، مانند توسعه مداوم محصول، گسترش به بازارها يا صنايع جديد، و تلاش براي نگهداري مشتري .

1. **امكانات مورد نياز جهت اجراي پروژه را ذكر نماييد.**

* منابع محاسباتی:

منابع محاسباتي با عملکرد بالا براي توسعه و آزمايش الگوريتم جديد ضروري است. اين مورد شامل دسترسي به سرورهاي قدرتمند يا پلتفرم‌هاي محاسبات ابري مي‌شود كه قادر به پردازش مجموعه داده‌هاي بزرگ و وظايف محاسباتي پیچیده به صورت كارآمد هستند. امکانات محاسباتي پیشرفته مجهز به شتاب‌دهنده‌هاي سخت‌افزاري تخصصي، مانند GPU ها يا TPU ها، مي‌توانند فرآيند آموزش و بهینه‌سازي الگوريتم را سريع‌تر انجام دهند و همچنین بهره‌وري و قابلیت مقیاس‌پذيري را افزايش دهند.

* ابزارها و کتابخانه‌های نرم‌افزاری:

دسترسي به يک مجموعه جامع از ابزارها و كتابخانه‌هاي نرم‌افزاري براي يادگیري ماشین و تجزيه و تحلیل داده براي اجراي پروژه بسیار حیاتي است. اين مورد شامل زبان‌هاي برنامه‌نويسي مانند پايتون يا R، و همچنین فريم‌ورک‌هاي يادگیري ماشین مانندTensorFlow ، PyTorch، ياscikit-learn مي‌شود. كتابخانه‌ها و بسته‌هاي تخصصي براي مدل‌سازي درخت تصمیم، مهندسي ويژگي‌ها، و تفسیر مدل براي تسهیل در توسعه و ارزيابي الگوريتم ضروري است.

* ذخیره و مدیریت داده:

امکانات كافي براي ذخیره و مديريت داده‌ها براي ذخیره و پردازش مجموعه داده‌هاي استفاده شده براي آموزش و ارزيابي الگوريتم لازم است. اين مورد شامل استقرار سیستم‌هاي پايگاه داده‌ يا راه‌حل‌هاي ذخیره سازي توزيعي می‌باشد كه قادر به مديريت حجم‌هاي بزرگ از داده‌هاي ساختارمند و بي‌ساختار به طور كارآمد هستند. روش‌هاي مديريت داده قوي، از جمله پیش‌پردازش داده، تمیز كردن و تبديل داده‌ها ضروري است، تا از كیفیت و اصالت داده‌هاي ورودي مورد استفاده در پروژه اطمینان حاصل شود.

* محیط آزمون و اعتبارسنجی:

يک محیط آزمون و اعتبارسنجي اختصاصي براي ارزيابي عملکرد و كارايي الگوريتم توسعه يافته ضروري است. اين مورد شامل راه‌اندازي

محیط‌هاي آزمايشي براي شبیه‌سازي سناريوهاي واقعي و ارزيابي پايداري الگوريتم تحت شرايط مختلف مي‌باشد. چارچوب‌هاي آزمون خودكار و پروتکل‌هاي اعتبارسنجي فرآيند آزمون را تسريع مي‌بخشند و اطمینان از نتیجه‌هايي قابل تکرار و سازگار در طول چند آزمايش را فراهم مي‌كنند .

1. **صورت مهمترين منابع و ماخذ مربوط به موضوع پيشنهادي را قيد فرمائيد.**

* Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
* Molnar, C. (2021). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
* Interpretable Machine Learning with XGBoost" by Tarek Amr:

<https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27>

* "Explaining Machine Learning Models: A User's Guide to Interpretable Models" by Christoph Molnar:

<https://christophm.github.io/interpretable-ml-book>

* Interpretability Techniques for Machine Learning: A Guide for Data Scientists" by IBM Watson:

<https://www.ibm.com/cloud/learn/interpretability-techniques-machine-learning>

* Singh, G., Arora, A., & Singh, M. (2021). BonsaiNet: Re-Imagining Decision Trees with Learned Differentiable Splits.