



بسمه تعالی

تاریخ: ۱۴۰۳/۰۳/۰۱

شماره: ۵

دانشکده مهندسی کامپیوتر

گروه هوش مصنوعی - نرم افزار

فرم پیشنهاد پروژۀ پایانی - دوره کارشناسی

عنوان پروژه (فارسی):

ارائه یک الگوریتم برای بهبود مدل های تفسیر پذیر بر پایه درخت تصمیم

عنوان پروژه (لاتین):

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

نام دانشجو	شماره دانشجویی	گرایش	تعداد واحدهای گذرانده شده	امضاء
1	سید حسین حسینی دولت آبادی	4012363328	هوش مصنوعی نرم افزار	85
2				

استاد راهنمای پروژه: دکتر حمید رضا برادران

نظر استاد راهنما:

امضاء استاد راهنما - تاریخ

پیشنهاد در تاریخ

در شورای گروه مهندسی هوش مصنوعی - نرم افزار مطرح گردید و

☐ بدون تغییر مورد تصویب قرار گرفت.

☐ با شرایط زیر مورد تصویب قرار گرفت.

☐ به دلایل زیر مورد تصویب قرار نگرفت.

نام عضو هیئت علمی بررسی کننده: دکتر سید فخرالدین نوربهبهانی

تاریخ و امضا

## 1- معرفی موضوع و بیان اهداف پروژه:

مدل‌های تفسیرپذیر، مخصوصاً 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 تصمیم‌گیری می‌باشد. وجود تفسیرپذیری، حتی در سطح محدود، باعث افزایش اعتماد کاربر به مدل می‌شود. به طوری که در زمینه‌هایی که تصمیمات مستقیماً 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 تشخیص 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 کارایی مدل‌های مبتنی بر هوش مصنوعی را افزایش دهیم.

## 2- روش انجام پروژه را توضیح دهید.

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، مخصوصاً مدل‌های بر پایه‌ی درخت‌های تصمیم.
- مشخص کردن تحقیقات کلیدی در این زمینه.
- تعیین مشکلات عمومی و محدودیت‌هایی که در روش‌های فعلی وجود دارد.

### 2- مشخص کردن خصوصیات جزئی مسئله:

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

### 3- 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 مقیاس‌پذیری.

### 4- 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.

### 5- پیاده‌سازی:

- پیاده‌سازی الگوریتم طراحی شده با به کارگیری برنامه‌نویسی و فریم‌ورک‌های مناسب.
- اطمینان از کارایی و مقیاس‌پذیری پیاده‌سازی به منظور تطبیق‌پذیری با مجموعه داده‌های مقیاس بزرگ. کاربردهای دنیای واقعی.

### 6- ارزیابی:

- ارزیابی کارایی 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

### 7- آنالیز نتایج:

- آنالیز نتایج مرحله‌ی ارزیابی با توجه به کارایی الگوریتم در دو بخش تفسیرپذیری و دقت پیش‌بینی.
- تفسیر مفهوم اطلاعات به‌دست آمده برای کاربردهای بالقوه در سناریوهای دنیای واقعی.

### 3- آیا این پروژه و یا مشابه آن قبلاً انجام شده است؟ در صورت مثبت بودن جواب، تفاوت‌های این پروژه با پروژه‌های قبلی را بیان نمایید.

مدل‌های قابل تفسیر، به‌ویژه مدل‌هایی که بر پایه درخت تصمیم ساخته شده‌اند، مورد توجه تحقیقاتی در زمینه یادگیری ماشین بوده‌اند. مطالعات قبلی، با تمرکز بر بهبود شفافیت، قابلیت استفاده و کاربرد عملی در کاربردهای واقعی، روش‌های مختلفی را برای بهبود تفسیرپذیری مدل‌های بر پایه درخت تصمیم بررسی کرده‌اند.

موارد زیر تفاوت‌های کلیدی بین پروژه‌ی ما و پروژه‌های قبلی را برجسته می‌کند:

#### 1- نوآوری الگوریتم پیشنهادی:

پروژه ما الگوریتمی نوآورانه را معرفی می‌کند که از ابتدا طراحی شده است تا دقت پیش‌بینی را حفظ کرده و در عین حال تفسیرپذیری را افزایش دهد. با بهره‌گیری از پیشرفت‌هایی در زمینه یادگیری ماشین، بهینه‌سازی و نظریه اطلاعات، الگوریتم ما، ویژگی‌ها و روش‌های منحصر به فردی را در بر می‌گیرد؛ که آن را از روش‌های موجود متمایز می‌کند. به‌طور خاص، ما بر روی ترکیبی از تکنیک‌های نوآورانه برای استخراج ویژگی، بهینه‌سازی ساختار درخت و یادگیری گروهی تأکید می‌کنیم.

#### 2- بهبود عملکرد:

بنا بر پایه برداشت‌های به‌دست آمده از تحقیقات پیشین، الگوریتم ما برای رفع مشکلات و محدودیت‌های خاص موجود در رویکردهای کنونی توسعه مدل قابل تفسیر می‌شود. با اولویت‌بندی قابلیت تفسیرپذیری بدون از دست دادن دقت پیش‌بینی، الگوریتم ما هدف دارد با چالش‌هایی مانند اورفیت، پیچیدگی مدل و قابلیت مقیاس‌پذیری مقابله کند. این اقدام منجر به بهبود کارایی مدل‌های بر پایه درخت تصمیم در موارد عملی می‌شود.

#### 3- اعتبارسنجی تجربه:

پروژه ما تأکید زیادی بر اعتبارسنجی تجربی دارد، با یک چارچوب ارزیابی طراحی شده که اثربخشی الگوریتم پیشنهادی را در کاربردهای واقعی ارزیابی می‌کند. از طریق آزمایش‌های گسترده با استفاده از مجموعه داده‌های مرجع و مطالعات موردی از حوزه‌های مختلف، ما قصد داریم که کارایی و کلیت الگوریتم خود را در کاربردهای عملی ثابت کنیم.

### 4- طرح تجاری خود در رابطه با به سودرسانی پروژه انجام شده را در صورت وجود توضیح دهید؟

- ارائه یک مدل درآمدی بر اساس اخذ مجوز برای الگوریتم از کسب و کارها، موسسات تحقیقاتی و توسعه دهندگان نرم‌افزاری.
- در نظر گرفتن ارائه دسترسی مبتنی بر اشتراک به الگوریتم، با قیمت‌گذاری بر اساس مصرف یا ویژگی‌ها. بررسی فرصت‌هایی برای تأمین درآمد از طریق خدمات مشاوره، کارگاه‌های آموزشی، یا توسعه مدل‌های سفارشی برای صنایع یا کاربردهای خاص.
- تنظیم بازار هدف روی صنایع و سازمان‌هایی که بیشترین بهره را از افزایش قابلیت تفسیرپذیری در مدل‌های بر پایه درخت تصمیم به دست می‌آورند، مانند مؤسسات مالی، ارائه دهندگان خدمات سلامت و سازمان‌های نظارتی.
- در نظر گرفتن استراتژی‌هایی برای پایداری بلندمدت، مانند توسعه مداوم محصول، گسترش به بازارها یا صنایع جدید، و تلاش برای نگهداری مشتری.

## 5- امکانات مورد نیاز جهت اجرای پروژه را ذکر نمایید.

- منابع محاسباتی:  
منابع محاسباتی با عملکرد بالا برای توسعه و آزمایش الگوریتم جدید ضروری است. این مورد شامل دسترسی به سرورهای قدرتمند یا پلتفرمهای محاسبات ابری می شود که قادر به پردازش مجموعه داده های بزرگ و وظایف محاسباتی پیچیده به صورت کارآمد هستند. امکانات محاسباتی پیشرفته مجهز به شتاب دهنده های سخت افزاری تخصصی، مانند GPU ها یا TPU ها، می توانند فرآیند آموزش و بهینه سازی الگوریتم را سریع تر انجام دهند و همچنین بهره وری و قابلیت مقیاس پذیری را افزایش دهند.
- ابزارها و کتابخانه های نرم افزاری:  
دسترسی به یک مجموعه جامع از ابزارها و کتابخانه های نرم افزاری برای یادگیری ماشین و تجزیه و تحلیل داده برای اجرای پروژه بسیار حیاتی است. این مورد شامل زبان های برنامه نویسی مانند پایتون یا R، و همچنین فریم ورک های یادگیری ماشین مانند TensorFlow، PyTorch، یا scikit-learn می شود. کتابخانه ها و بسته های تخصصی برای مدل سازی درخت تصمیم، مهندسی ویژگی ها، و تفسیر مدل برای تسهیل در توسعه و ارزیابی الگوریتم ضروری است.
- ذخیره و مدیریت داده:  
امکانات کافی برای ذخیره و مدیریت داده ها برای ذخیره و پردازش مجموعه داده های استفاده شده برای آموزش و ارزیابی الگوریتم لازم است. این مورد شامل استقرار سیستم های پایگاه داده یا راه حل های ذخیره سازی توزیعی می باشد که قادر به مدیریت حجم های بزرگ از داده های ساختارمند و بی ساختار به طور کارآمد هستند. روش های مدیریت داده قوی، از جمله پیش پردازش داده، تمیز کردن و تبدیل داده ها ضروری است، تا از کیفیت و اصالت داده های ورودی مورد استفاده در پروژه اطمینان حاصل شود.
- محیط آزمون و اعتبارسنجی:  
یک محیط آزمون و اعتبارسنجی اختصاصی برای ارزیابی عملکرد و کارایی الگوریتم توسعه یافته ضروری است. این مورد شامل راه اندازی محیط های آزمایشی برای شبیه سازی سناریوهای واقعی و ارزیابی پایداری الگوریتم تحت شرایط مختلف می باشد. چارچوب های آزمون خودکار و پروتکل های اعتبارسنجی فرآیند آزمون را تسریع می بخشند و اطمینان از نتیجه هایی قابل تکرار و سازگار در طول چند آزمایش را فراهم می کنند.

- 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.

