



دانشگاه شهید بهشتی

Shahid Beheshti University

پروژه پایانی

درس شناسایی الگو

مدرس: دکتر آرمین سلیمانی بدر

پاییز ۱۴۰۳



مقدمه

با پیشرفت سریع الگوریتم‌های هوش مصنوعی، یادگیری ماشین و یادگیری عمیق، صنایع مختلف به دنبال بهره‌گیری از این تکنولوژی‌ها برای بهبود فرآیندها و افزایش بهره‌وری هستند. صنعت خودروسازی نیز از این قاعده مستثنی نیست و در سال‌های اخیر شاهد تحولاتی چشمگیر در این حوزه بوده است. هوش مصنوعی به خودروسازان این امکان را می‌دهد که از داده‌های گسترده‌ای که خودروها در طول کارکرد خود تولید می‌کنند، برای بهینه‌سازی مصرف سوخت، کاهش آلایندگی، افزایش ایمنی و بهبود تجربه رانندگی استفاده کنند.

یکی از چالش‌های مهم در صنعت خودرو، مصرف سوخت و آلایندگی برآمده از آن است. با توجه به نگرانی‌های زیست‌محیطی و قوانین سخت‌گیرانه‌تر در زمینه انتشار گازهای گلخانه‌ای، تولید خودروهای با بهره‌وری انرژی بالاتر و آلایندگی کمتر به یک اولویت اساسی تبدیل شده است. در این راستا، الگوریتم‌های هوش مصنوعی می‌توانند در بهینه‌سازی موتورهای احتراق داخلی، طراحی خودروهای هیبریدی و الکتریکی، و حتی پیش‌بینی رفتار رانندگی کاربران برای کاهش مصرف سوخت، نقش کلیدی ایفا کنند.

همچنین، استفاده از یادگیری ماشین می‌تواند به تحلیل و پیش‌بینی الگوهای مصرف سوخت و رفتار رانندگی کمک کند و از داده‌های واقعی خودروها برای ارائه پیشنهادات بهینه به رانندگان و خودروسازان بهره‌برداری کند. یادگیری عمیق نیز با تحلیل داده‌های پیچیده می‌تواند در کاهش خطاهای انسانی و افزایش کارایی سیستم‌های مدیریت انرژی خودروها تأثیرگذار باشد.

در کنار مصرف سوخت و آلایندگی، تکنولوژی‌های هوش مصنوعی و یادگیری ماشین می‌توانند در ارتقای سیستم‌های خودران، بهبود سیستم‌های ناویگی و حتی مدیریت ترافیک شهری به کمک داده‌های جمع‌آوری شده از خودروها مؤثر باشند. این فناوری‌ها نویدبخش آینده‌ای هستند که در آن خودروها نه تنها هوشمندتر، بلکه دوستدار محیط زیست خواهند بود.

در این پژوهه تمرکز ما استفاده از داده‌های واقعی یک شرکت خودروسازی است که از طریق کامپیوتر مرکزی خودرو (ECU) جمع‌آوری شده است. ما با استفاده از این داده‌ها می‌توانیم نوع رفتار رانندگی را مشخص کنیم، تعویض دنده را برای افزایش لذت رانندگی و کاهش مصرف سوخت بهینه سازی کنیم، مصرف سوخت خودرو را در سیکل‌های استاندارد اروپا و آسیا پیش‌بینی کنیم. البته مسائل بهینه سازی می‌تواند با استفاده از یادگیری تقویتی انجام شود که در این درس مدنظر نیست.



هنر شرکت‌های بزرگ خودروسازی استفاده از الگوریتم‌های کلاسیک یادگیری ماشین برای انجام این موارد است زیرا پیاده سازی این نوع الگوریتم‌ها راحت‌تر است و همچنین می‌توان از این الگوریتم‌ها در کامپیوتر داخلی خودرو در آینده نزدیک برای بهینه سازی رانندگی استفاده کنند.



دادگان

در قسمت قبل در مورد اهمیت استفاده از داده‌های واقعی برای آموزش و پژوهش صحبت شد. همان طور که اشاره شد این داده‌های واقعی یک خودرو است که از طریق ECU داده برداری شده است.

این داده‌ها از یک ماشین خاص ولی با چندین راننده متفاوت داده برداری شده است، نوع رانندگی این راننده‌ها به صورت کم مصرف، مصرف متوسط و پر مصرف می‌باشند. داده‌های موجود در پیوست شامل این چند رانندگی و یک داده دیگر خودرویی با سوخت عالی می‌باشد که در ادامه کارکرد هر یک از این داده‌ها گفته شده است.

داده‌های چندین راننده شامل ستون‌های زیر می‌باشد:

Time: زمان رانندگی (ثانیه)

ES: سرعت دور موتور خودرو (دور بر دقیقه)

CT: گشتاور خودرو (نیوتن متر)

TFAR: نسبت هوا به سوخت

TP: موقعیت دریچه گاز خودرو (درصد از ۰ تا ۱۰۰)

APP: موقعیت پدال گاز خودرو (درصد از ۰ تا ۱۰۰)

C: دمای مایع خنک کننده خودرو (درجه سانتی گراد)

S: سرعت خودرو (کیلومتر بر ساعت)

G: دندنه فعلی خودرو

CM: مسافت پیموده شده به صورت تجمعی (کیلومتر)

TPF: سوخت مصرفی خودرو به صورت تجمعی (لیتر)

FQ: کیفیت سوخت خودرو

Slope: شیب مسیر

همان طور که گفته شد این داده به صورت فایل اکسل و دارای ستون‌های گفته شده است.



ابتدا پیش پردازش‌های لازم مانند رسم هیستوگرام و مشخص کردن توزیع هر ویژگی، محاسبه میانگین، انحراف معیار و... انجام دهید، انجام پیش‌پردازش‌های اولیه روی داده‌ها کمک می‌کند تا درک بهتری از مسئله پیدا کنید.

سپس با استفاده از این داده‌ها میزان همبستگی هر یک نسبت به مصرف سوخت لحظه‌ای را بیابید، محاسبه مصرف سوخت لحظه‌ای دست خودتان است و می‌توانید مصرف لحظه‌ای را در ۱۰۰ کیلومتر محاسبه کنید یا فقط از اختلاف لحظه‌ای مصرف تجمعی (با شیفت به بالا یا پایین و یا بدون شیفت) محاسبه کنید. برای پیدا کردن بهترین حالت می‌توانید از ماتریس همبستگی استفاده کنید. توجه داشته باشید این پارامتر در قسمت‌های بعدی مورد استفاده است پس بهتر است بهترین حالت را در نظر بگیرید.

برای محاسبه مصرف سوخت در ۱۰۰ کیلومتر می‌توانید از فرمول زیر استفاده کنید.

$$\text{مصرف سوخت در } 100 \text{ کیلومتر} = \frac{\text{مسافت طی شده (کیلومتر)}}{\text{مقدار سوخت مصرف شده (لیتر)}}$$



بخش اول

رفتار رانندگی مبتنی بر داده از طریق تجزیه و تحلیل داده های جمع آوری شده از سنسورها و سیستم های پیشرفته مانند GPS شتاب سنج ها، دوربین ها و سیستم های کنترل خودر (OBD)، الگوهای رانندگی رانندگان را شناسایی و تحلیل می کند. داده های کلیدی در قسمت قبل مورد بررسی قرار گرفته است. با تحلیل این اطلاعات، می توان الگوهای رفتاری رانندگان را به دسته هایی مانند رانندگی تهاجمی، معمولی و محافظه کارانه تقسیم کرد. البته ممکن است تعداد خوش های بھینه بیشتر از این تعداد باشد.

شناسایی راننده (Driver Identification) با استفاده از تکنیک های یادگیری ماشین، هوش مصنوعی و داده کاوی، به طور دقیق هویت راننده را بر اساس سبک رانندگی، فشار پا روی پدال، فرمان دهی و سایر ویژگی های رانندگی تشخیص می دهد. این روش با استفاده از بیومتریک رفتاری، یک نوع امضا منحصر به فرد برای هر راننده ایجاد می کند که امکان تشخیص افراد مختلف حتی در یک خودرو مشترک را فراهم می کند، و اگر خودرویی ربوده شود با استفاده از این داده ها خودرو به صورت خودکار اجازه رانندگی به سارق را نمی دهد.

همچنین شناسایی رفتارهای خطرناک مانند شتاب های ناگهانی، تغییر مسیر های غیرمنتظره و ترمز های شدید به سامانه های پیشرفته امکان می دهد هشدارهای لازم را به راننده ارسال کنند یا حتی خودرو را به طور خودکار کنترل کنند. این موضوع در کاهش تصادفات و ارتقای ایمنی نقش اساسی دارد.

رفتارهای رانندگی پرمصرف مانند شتاب گیری های شدید، توقف های مکرر و حرکت های نامناسب در ترافیک، باعث افزایش مصرف سوخت می شوند. از طریق پایش و تحلیل داده های رانندگی، می توان عادات نادرست رانندگان را اصلاح و آن ها را به سمت رانندگی بھینه تر هدایت کرد. این موضوع نه تنها مصرف سوخت و هزینه ها را کاهش می دهد، بلکه انتشار آلاینده های زیست محیطی را نیز محدود می کند، همچنین می توان با نشان دادن امتیاز یا پیشنهاد در صفحه نمایش خودرو تأثیر مثبتی بر روی رانندگی فرد در طول زمان داشت.

در نتیجه، با ترکیب تحلیل داده های رانندگی و شناسایی راننده، می توان به بهبود ایمنی جاده ها، کاهش مصرف سوخت و بهینه سازی عملکرد خودروها دست یافت. این فناوری ها به خصوص در حوزه حمل و نقل هوشمند، خودروهای اشتراکی و ناوگان های حمل و نقل تجاری بسیار کاربرد دارند.

با توجه به توضیحات بالا از اهمیت خوش بندی یا دسته بندی رانندگان مطلع شدید. حال در این قسمت باید با استفاده از داده های موجود در پوشه Cluster. رانندگان را به ۳ یا ۵ دسته مختلف تقسیم و خوش بندی کنید. برای ۳ دسته رانندگی می توانید از برچسب های رانندگی Eco، Normal، Sport استفاده کنید که رانندگی Eco راننده کم مصرف و Sport راننده پر مصرف محسوب می شوند، برای ۵ دسته نیز از برچسب های Full Eco،



استفاده کنید که مصرف سوخت مانند دسته بندی قبلی Full Sport ، Semi Sport ، Normal ، Semi Eco است. فقط برای خوشبندی مجاز به استفاده از **ستون مصرف تجمعی یا لحظه‌ای خودرو نیستید**. در این قسمت شما می‌توانید از روش‌های بی‌نظرارت (unsupervised) یا نیمه‌نظرارتی (semi-supervised) برای خوشبندی (cluster) استفاده کنید. روش حل این مسئله و استفاده از ویژگی‌ها با خلاصت خودتان می‌باشد و پس از خوشبندی باید ویژگی‌های خوشبندی شده را با استفاده از PCA کاهش بعد داده (اگر بیشتر از ۲ بعد باشد) و در یک نمودار رسم کنید.

سپس در یک فایل پیوست (Filename_Label) نام فایل به همراه برچسب‌های ۳ تایی و ۵ تایی موجود است که شما بعد از خوشبندی باید با لیبل‌های داده شده در فایل پیوست دقت بالای ۸۵٪ درصد به دست آورید. برای دستیابی به دقت مورد انتظار شما باید از خلاصت خود و **مقالات‌های ضمیمه شده** استفاده کنید، برای نمونه استفاده از روش‌های رأی گیری و یا روش‌های Ensemble Clustering یا افزودن پارامترهای آماری به عنوان ویژگی می‌تواند مؤثر باشد.



بخش دوم

بعد از انجام مرحله اول نوبت به پیشビینی مصرف سوخت لحظه‌ای رانندگان می‌رسد. در این قسمت شما باید از الگوریتم‌های مختلف یادگیری ماشین مانند رگرسیون، جنگل تصادفی و... برای پیشビینی مصرف سوخت لحظه‌ای هر یک از فایل‌ها اقدام کنید. برای این‌کار نخست با استفاده از داده‌های پوشه train مدل‌های خود را آموزش داده و دقت مدل را با استفاده از داده‌های موجود در پوشه test ارزیابی کنید، در این قسمت بهترین مدل خود و همچنین ویژگی‌های مورد استفاده را همراه با دقت MAPE که به صورت فرمول زیر است برای هر فایل به صورت جداگانه گزارش کنید. انتظار می‌رود که خطای تست‌های شما کمتر از ۲۵٪ باشد. همچنین با استفاده از اثرگذاری هر ویژگی برای پیشビینی مصرف سوخت را توضیح SHapley Additive exPlanations(SHAP) دهید.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A - P}{A} \right|$$

که در این فرمول A مصرف لحظه‌ای واقعی و P مقدار پیشビینی شده توسط شما می‌باشد. فقط توجه داشته باشید که ممکن است مصرف لحظه‌ای در مواردی برابر ۰ شود که محاسبه فرمول امان پذیر نمی‌باشد پس در این موارد مقدار را none گذاشته و از تعداد n کم کنید.

نکته ۱: استفاده از تمام ویژگی‌ها آزاد نیست و فقط می‌توانید از ۶ ویژگی برای پیشビینی مصرف سوخت استفاده کنید پس توجیه خود برای استفاده از هر یک از ویژگی‌ها را توضیح دهید. البته می‌توانید از این ویژگی‌ها، ویژگی‌های جدید ایجاد کنید، مثلاً شتاب که تفاصل سرعت در هر لحظه است یا تفاصل دور موتور یا همچنین شتاب خودرو در ۳ ستون توسط شیفت بالا و پایین تولید شده باشد. پس این قسمت نیز نیاز به خلاقیت دارد و عدم وجود آن ممکن است دقت مورد نظر را فراهم نکند.

نکته ۲: استفاده از مدل‌های یادگیری عمیق مجاز نمی‌باشد.

همه مدل‌ها باید ذخیره شده و در فایل ارسالی شما برای تست موجود باشند همچنین در یک بلوک امکان خواندن فایل جدید و تست با استفاده از مدل شما فراهم باشد.



بخش سوم

کیفیت بنزین در تولید آلایندگی و مصرف سوخت خودرو یک عامل بسیار تأثیرگذار می‌باشد. این عامل به ویژه در خودروهای پرخوران (Turbo) بسیار چشمگیر می‌باشد زیرا این نوع موتورهای بنزینی به کیفیت سوخت حساس هستند و حتی ممکن هست باعث تخریب موتور در طولانی مدت شوند.

در فایل Bad_Good_Fuel_part3 که در پیوست موجود است چندین سفر از یک خودرو موجود می‌باشد که این سفرها با استفاده از سوخت‌های مختلف تست و بررسی شده‌اند. شاخص کیفیت سوخت در این فایل در ستون FQ موجود است که مقدار این ستون بین ۰ و ۱ می‌باشد. عدد ۰ نشانگر کیفیت کم و ۱ نشانگر کیفیت خوب سوخت است. حال با استفاده از ویژگی‌هایی که در قسمت پیش استفاده کردید و ویژگی جدید FQ بهترین مدل خود را با این ویژگی‌ها آموزش داده و مدل خود را ذخیره کنید. ستون trip نشان دهنده شماره سفر می‌باشد و برای صحت سنجی کار خود باید ۴ سفر با کیفیت سوخت‌های متفاوت برای داده آزمایش خود جدا کنید و مقدار MAPE را برای هر یک گزارش کنید. حال فرض کنید این خودرو در همه قسمت‌ها یکسان بوده با فرض اینکه داده‌های قسمت پیش همین خودرو با کیفیت سوخت ۰ می‌باشد، مصرف جدید هر فایل در ۱۰۰ کیلومتر را با استفاده از سوخت خوب ($FQ=1$) به دست آورید فایل خروجی شما باید در یک فایل اکسل، ستون‌های نام فایل و مصرف خودرو در ۱۰۰ کیلومتر موجود باشد. انتظار می‌رود که مصرف پیش‌بینی شده شما کمتر از مقدار واقعی باشد (البته در مواردی ممکن است بیشتر باشد).

در این قسمت می‌توانید کیفیت سوخت را به صورت پیوسته یا گسسته انتخاب کنید برای نمونه کیفیت بد عددی بین ۰ تا ۰,۴ و کیفیت متوسط عددی بین ۰,۴ تا ۰,۸ و کیفیت عالی عددی بین ۰,۸ تا ۱ می‌باشد. همچنین باید مقدار همبستگی این پارامتر نسبت به مصرف سوختی که انتخاب کرده‌اید مشخص شده باشد.



نکات حائز اهمیت در تحویل پروژه

- ۱- تحویل پروژه از طریق سامانه درس افزار می‌باشد.
- ۲- کد باید در قالب نوتبوک و نتایج آن قابل مشاهده باشد، عدم توجه به این موضوع باعث کسر نمره می‌شود. همه روش‌های تست شده هم موجود باشد.
- ۳- پروژه را در قالب یک فایل فشرده بارگذاری کنید و نام فایل به صورت زیر باشد.
Project-StudentName1-StudentName2
- ۴- پروژه به صورت گروه دو نفره است و هرگونه تقلب از سایر گروه‌ها، استفاده از کدهای آماده و استفاده از AI‌های مرسوم نمره تمرین را صفر می‌کند.
- ۵- گزارش نویسی برای تحلیل و بررسی ایده‌های مورد استفاده بخشی از نمره پروژه را داراست، پس یک گزارش کامل شامل مقدمه، توضیح مختصر ایده‌های استفاده شده، تحلیل بخش‌های خواسته شده نه کم و نه زیاد است. پس وقت کافی برای گزارش نویسی صرف شود.
- ۶- در صورت داشتن هرگونه سؤال با ایده‌های زیر در ارتباط باشید:

@sirmmp

@ali_es779

پیروز باشید.