تعیین خصوصیات اقتصادی- اجتماعی مسافران در ترافیک شهری با استفاده از هوش مصنوعی

مقاله علمی-پژوهشی

*شهریار افندی‌زاده (نویسنده مسئول)، استاد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
محمود احمدی نژاد، دانشیار، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
محمد قایدی، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
حمید بیگدلی‌راد، دانشجوی دکتری، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
*پست الکترونیکی نویسنده مسئول: zargari@iust.ac.ir

دریافت: ۱۴۰۵/۰۳/۲۸- پذیرش: ۱۴۰۵/۰۸/۱۰
صفحه ۵۰-۳۱

چکیده

در مدل‌های پیش‌بینی تقاضای سفر ویژگی‌های اقتصادی – اجتماعی افراد به عنوان ورودی مدل استفاده می‌شود. در این مطالعه به کمک مدل‌های هوش مصنوعی و بخصوص مدل‌های یادگیری عمیق سعی شده است تا از روی الگوی سفرهای افراد این ویژگی‌ها استخراج شود. مدلی که انتخاب شد مدل LSTM می‌باشد. سن، درآمد و جنسیت از مهم‌ترین ورودی‌های این مدل می‌باشد. گروه‌های سنی، جنسیتی و درآمدی مختلف ممکن است ترجیحات رفتار سفر خاصی داشته باشند که سعی شده با بررسی رفتارهای سفر آنها این ویژگی‌ها کشف شود. یکی از منابع مهم برای مشخص کردن الگوی سفر افراد داده‌های تلفن همراه است که جابجایی روزانه آن‌ها را نشان می‌دهد. با توجه به دردسترس نبودن داده‌های تلفن همراه مسافران، دراین تحقیق از مجموعه داده نظرسنجی بررسی ملی مسافرت خانگی NHTS استفاده شده است. با استفاده ازنتایج به‌دست آمده، به کمک داده‌های تلفن همراه مسافران، ویژگی‌های اقتصادی و اجتماعی آنها برای استفاده در مدل‌های تقاضای سفربه دست می‌آیند. میانگین مسافت پیموده شده برای مردان در هرسفر ۶/۷۱ مایل است ولی میانگین مسافت طی شده توسط خانم‌ها ۳/۴۶ مایل می‌باشد. بنابراین زنان سفرهای کوتاه‌تری نسبت به زنان دارند. افراد بین ۳۰-۲۰ سال دارای الگوی سفرهای ساده‌تری هستند. بسیاری از این افراد هنوز مشغول به تحصیل در دانشگاه هستند یا شاغل نیستند. با بررسی الگوی سفرهای افراد روی داده‌ها مشخص شد که طبقه متوسط درآمدی سفرهای بیشتری دارند و مسافت‌های بیشتری را طی می‌کنند بنابراین الگوی سفر پیچیده‌تری دارند.

واژه‌های کلیدی: مدل تقاضای سفر، پارامترهای اقتصادی – اجتماعی، یادگیری عمیق، مدل LSTM

۱- مقدمه

در مدل‌های حمل و نقل، ورودی‌های مدل شامل پارامترهای اقتصادی– اجتماعی افراد جامعه، کاربری زمین و مالکیت اتومبیل و متغیرهای دیگر می‌باشند. برای ساخت مدل گام اول جمع آوری داده‌های ورودی می‌باشد و از آنجایی که جمع‌آوری این داده‌ها وقت‌گیر و هزینه بر است و به‌خصوص در مورد پارامترهای اقتصادی– اجتماعی افراد نظیر جنسیت، درآمد، اشتغال نیاز به سرشماری و پرسش مستقیم از افراد وجود دارد. به همین دلیل سوال اینجا است که آیا روش‌های جدیدی هست تا بتوان سریع، راحت و کم هزینه و دقیق این پارامترهای اقتصادی– اجتماعی افراد را بدست آورد و مدل‌های حمل و نقلی دقیق‌تری بدست آید؟ دراین تحقیق با توجه به پیشرفت‌های هوش مصنوعی سعی شده است که از راهکارهای جدید برای دستیابی به این پارامترها بدون استفاده از روش‌های مستقیم سرشماری و پرسشنامه استفاده شود. در این پژوهش، مدل‌های حمل و نقلی به صورت معکوس دیده شده است. به این معنی که از روی الگوی سفر افراد در یک

روز پارامترهای اقتصادی-اجتماعی افراد مشخص شده است(Afandizadeh & Bigdeli Rad, 2021).

ویژگی‌های اجتماعی-اقتصادی از جمله عناصری است که در مدل‌سازی حمل و نقل از اهمیت بالایی برخوردار است. این ویژگی‌ها می‌تواند شامل سن، درآمد، تحصیلات، جنسیت و وضعیت تاهل باشد و معمولاً از طریق بررسی‌های مرسوم سفر با استفاده از ترکیب جمعیت یا داده‌های سرشماری جمع آوری می‌شود. با توجه به مدل‌سازی تقاضای سفر، ویژگی‌های خانوار مانند درآمد و وضعیت اجتماعی-اقتصادی می‌تواند یک عامل تأثیرگذار اصلی بر الگوهای سفر و در نتیجه در انتخاب سیاست‌های مدیریت ترافیک کارآمد برای برنامه‌ریزان حمل‌ونقل باشد(Zhou et al., 2019).

۲- پیشینه تحقیق

گروه‌های جمعیتی برای اجرای انواع سیاست‌های مؤثر حمل و نقل بر اساس ویژگی‌های اجتماعی- اقتصادی خود هدف قرار می‌گیرند. به عنوان مثال، گروه‌های کم درآمد به احتمال زیاد انگیزه بیشتری برای کاهش استفاده از خودروی شخصی و روی آوردن به حمل و نقل عمومی در صورت اعمال قیمت گذاری جاده‌ای دارند (Zargari & Rad, 2023). از نظر سیاست حمل و نقل، اطلاعات جمعیت شناختی در ترسیم استراتژی‌های بهینه حمل و نقل شهری مانند سطوح خدمات حمل و نقل و کرایه‌ها، با توجه به منابع محدود موجود مهم است (Bing-quan et al., 2020). شواهد زیادی وجود دارد که نشان می‌دهد افراد با ویژگی‌های اجتماعی- اقتصادی متفاوت رفتار سفر متفاوتی دارند. الگوهای رفتار سفر بر اساس جنسیت، سن و نژاد متمایز می‌شوند (Gopalakrishna et al., 2019).

در مطالعات سفرها و الگوی سفرها در این چند سال از مدل‌های ماشین لرنینگ استفاده زیادی شده است.

آنیل کیوشیک و همکاران رفتار سفر افراد و فعالیت را با مدل‌های ML مورد تجزیه و تحلیل قرار دادند. الگوریتم‌های یادگیری ماشین، که تلاش می‌کنند با استفاده از در دسترس بودن مقادیر زیادی داده، هوشمندی ایجاد کنند، به عنوان ابزار قدرتمندی در زمینه‌های تشخیص الگو و تجزیه و تحلیل کلان داده‌ها ظاهر شده‌اند. این تکنیک‌ها از اوایل دهه ۹۰، زمانی که شبکه‌های عصبی مصنوعی برای مدل‌سازی تصمیم‌گیری‌های انتخاب حالت به کار گرفته شدند، در مطالعات رفتاری فعالیت‌ها به کارگرفته شدند. ازآن زمان محققان تکنیک‌های ML مانند ماشین‌های بردار پشتیبان، درخت‌های تصمیم، شبکه‌های عصبی، طبقه‌بندی‌کننده‌های بیزی، را برای مدل‌سازی و پیش‌بینی رفتار سفر در فعالیت‌ها بررسی کرده‌اند (Khatri et al., 2021).

در تحقیقی که توسط نیوکیو وانگ انجام شده است از داده‌های تلفن همراه اشخاص برای تشخیص وضعیت سلامت جسمی و بررسی زوال عقل در رانندگان استفاده شده است. به‌طور خاص، پیشنهاد شده است که از یادگیری عمیق با شبکه‌های حافظه کوتاه‌مدت (LSTM) به‌عنوان طبقه‌بندی‌کننده برای تشخیص الگوهای سفر استفاده شود.

نتایج تجربی روی یک مجموعه داده واقعی جمع‌آوری‌شده از ۱۴ نفر نشان می‌دهد که طبقه‌بندی کننده‌های عمیق LSTM بهتر از طبقه‌بندی کننده‌های یادگیری ماشین سنتی (ML) عمل می‌کنند. در این تحقیق نشان داده شده که ممکن است برای رسیدن به دقت بهتر ممکن است برای هر بخش مدل های مختلفی را امتحان کنیم. که مثلاً برای حالات سرعت دور زدن یا قدم زدن افراد LSTM بهترین نتایج را می دهد (Chu et al., 2019).

رفتار سفر افراد با توجه به درآمد آنها متفاوت است، افراد متعلق به گروه کم مسافت‌های کوتاه‌تری را طی می‌کنند. بنابراین، این دسته از افراد بار کمتری را بر جامعه و محیط تحمیل می‌کنند.

مطالعات مختلف، تغییرپذیری در رفتار سفر را توضیح داده‌اند. هم در کشورهای توسعه‌یافته و هم در کشورهای در حال توسعه، افراد متعلق به گروه‌های درآمد پایین‌تر بیشتر به سفرهای غیرموتوری وابسته هستند، جابجایی کمتری دارند و هزینه کمتری برای سیستم حمل‌ونقل عمومی دارند. البته مطالعات نشان داده که در بسیاری کشورها درآمد تأثیر قابل توجهی بر انتخاب مد ندارد، با این حال درآمد بر تعداد وسایل نقلیه متعلق به خانوارها تأثیر می‌گذارد (Bulusu et al., 2021).

در مطالعه دیگری استنباط وضعیت اشتغال مسافران را به کمک داده‌های کارت هوشمند بررسی کردند. درک وضعیت اشتغال مسافران در سیستم‌های حمل و نقل عمومی برای اپراتورهای حمل و نقل در بسیاری از کاربردهای واقعی مانند پیش بینی تقاضای سفر و ارائه خدمات حمل و نقل شخصی حائز اهمیت است. این تحقیق یک رویکرد یادگیری عمیق را برای استنباط وضعیت اشتغال مسافر با استفاده از داده‌های کارت هوشمند با نظرسنجی خانوار توسعه می‌دهد (Xie et al., 2021).

در این مطالعه، داده‌های کارت هوشمند با داده‌های نظرسنجی خانوار ترکیب می‌شود تا ابتدا قدرت تمایز کارت هوشمند برای پیش‌بینی وضعیت اشتغال را بررسی کند، که می‌تواند به عنوان یک مسئله طبقه‌بندی فرموله شود. این مقاله داده‌های خام کارت هوشمند را به عنوان یک ساختار تصویر سه بعدی برای ثبت رفتار سفر مسافران نشان می‌دهد. با بهره‌گیری از آن، کار استنتاج وضعیت اشتغال را می‌توان با استفاده از یک مدل مبتنی بر شبکه عصبی کانولوشن انجام داد. شبکه عصبی کانولوشن یک تکنیک یادگیری عمیق محبوب است که موفقیت زیادی در زمینه طبقه‌بندی تصاویر به‌دست آورده است (Hörcher et al., 2022).

پس از بررسی فراوانی که در مقالات متعدد انجام شد نشان می‌دهد که امروزه روش‌های جمع‌آوری داده‌های بزرگ بسیار متفاوت از گذشته است و با استفاده از داده‌های تلفن همراه و همچنین خروجی‌های GPS بسیاری از اطلاعات موردنیاز جمع‌آوری می‌شوند. نکته بعدی پارامترهای اقتصادی– اجتماعی

اشخاص سفرکننده است که در مدل‌های حمل و نقلی بسیار مهم هستند و بدست آوردن دقیق، سریع و ارزان آنها نقش مهمی در ساخت مدل‌های حمل و نقلی دارند، چون این خصوصیات معمولاً به عنوان ورودی مدل‌ها استفاده می‌شود و مدل‌سازان بیشتر ترجیح می‌دهند که با این خصوصیات مدل‌های سفر بسازند و در اکثر موارد این خصوصیات را از سرشماری برداشت می‌کنند ولی در این تحقیق سعی خواهد شد مسیر برعکس عمل شود یعنی با استفاده از سفرهای روزانه اشخاص پارامترهای اقتصادی- اجتماعی استنباط شود که دیگر نیازی به جمع‌آوری این شاخص‌ها با روش‌های سنتی نباشد (Chen et al., 2020). همچنین در مطالعاتی هم که تاکنون انجام شده است نتایجی به دست آمده است که سعی می‌شود در تحقیق پیش رو با نگاهی جدید مدل‌های قدرتمندی ساخته شود و نتایج بهتری به دست آید. در جدول ۱ به بررسی مطالعات پیشین در موضوع مورد مطالعه پرداخته می‌پردازد.

جدول ۱. خلاصه‌ای از مهم‌ترین مطالعات بررسی شده

ردیف	عنوان تحقیق	نام محققان و سال تحقیق	هدف تحقیق	نتایج تحقیق
۱	الگوهای سفر و اثرات زیست محیطی در حال حاضر و آینده: پیامدهای تفاوت در مصرف انرژی در میان گروه‌های اجتماعی-اقتصادی	Carlsson et al. (1999)	کاهش مصرف انرژی و اثرات زیست محیطی با تغییر در الگوی سفر افراد	افزایش بازده انرژی خودرو وتغییر در الگوی سفر افراد بر اساس سن و جنسیت آنها
۲	عوامل تعیین کننده پذیرش قیمت‌گذاری جاده‌ها توسط کاربران خودروهای شخصی	Jakobsson et al. (2000) ,	به‌دست‌آوردن عوامل پذیرش قیمت گذاری	پذیرش منفی قیمت‌گذاری باعث احساس نقض آزادی و گسترش بی عدالتی است.
۳	جنسیت و تفاوت در چرخه سفر	Collins and Tisdell (2002)	تأثیر جنسیت و نقطه اوج سن افراد در الگوی سفر آنها	مردان بیشترسفرهای کاری می روند و زنان سفرهای تفریحی و دیدار با دوستان دارند و سن اوج سفر در زنان زودتر اتفاق می‌افتد.
٤	اصول یکپارچه سازی در استراتژی‌های حمل و نقل شهری	May et al. (2006)	تعیین یک استراتژی یکپارچه موثر درراستای رسیدن به اهداف توافق شده حمل ونقلی	با هم‌افزایی و ادغام بسیاری ازشاخص‌ها نتایج بهتری دررسیدن به اهداف می‌گیریم.
٥	تشخیص الگو و یادگیری ماشین	Bishop and Nasrabadi (2006)	کاربرد روش‌های بیزی در الگوریتم‌ها	بهبود چشمگیر عملکرد الگوریتم‌ها با استفاده از روش‌های بیزی دریادگیری ماشین
٦	مدیریت تقاضای سفر با هدف کاهش استفاده از خودروی شخصی: اثربخشی، اقبال عمومی و امکان پذیری سیاسی	Gärling and Schuitema (2007)	ارائه روش‌های قابل انجام مدیریت تقاضای سفر برای کاهش استفاده خودروهای شخصی	ترکیب اقدامات اجباری و غیراجباری مدیریت تقاضای سفرمانند افزایش هزینه و ممنوعیت به همراه روش‌های مشوق و جذاب
۷	تفاوت‌های سفر بر اساس جنسیت	Crane and Takahashi (2009)	جنسیت افراد چه تأثیری در سفرهای آنها دارد؟	درگذشته مردان بیشتر از زنان سفر می‌کردند ولی با رشد جوامع سفرهای زنان در حال افزایش است ولی نژاد افراد تأثیر زیادی در سفرهای آنها ندارد.
۸	پیش‌بینی نقش اجتماعی- جمعیت شناختی فردی بر اساس تغییر رفتار سفر با استفاده از داده‌های بلندمدت GPS	Zhu et al. (2017)	تعیین نقش اجتماعی افراد به استفاده از الگوی سفرهای آنها	با استفاده از داده‌های GPS و نوع سفر افراد الگوی سفر آنها مشخص شده و نقش افراد تعیین می‌شود.
۹	تبیین رفتار سفر با داده‌های اجتماعی-اقتصادی محدود: مطالعه موردی ویشاخاپاتنام، هند	Jain and Tiwari (2019)	بررسی تفاوت‌های رفتار سفر افراد بر اساس سطح درآمد آنها	افراد کم درآمد سفرهای کمتر و کوتاه‌تری دارند و بارکمتری را به سیستم حمل و نقل جامعه وارد می‌کنند.

ردیف	عنوان تحقیق	نام محققان و سال تحقیق	هدف تحقیق	نتایج تحقیق
۱۰	رویکرد یادگیری عمیق برای استنباط وضعیت اشتغال مسافران با استفاده از داده‌های کارت هوشمند	Zhang and Cheng (2019)	درک وضعیت اشتغال مسافران برای پیش‌بینی تقاضای سفر و ارائه خدمات حمل و نقل	همبستگی زیادی بین سفرهای افراد و وضعیت اشتغال آنها وجود دارد.
۱۱	تجزیه و تحلیل ویژگی‌های رفتار سفر مسافران مرکز حمل و نقل با استفاده از داده‌های تلفن همراه	Zhong et al. (2019)	با استفاده از اطلاعات جابجایی تلفن همراه مسافران رفتار سفر آنها چگونه است.	داده‌های تلفن همراه روش بسیار مفید برای جمع‌آوری اطلاعات سفر افراد است و منطقه جامع هاب نقش مهمی در سفرهای افراد دارد.
۱۲	کاربردهای یادگیری ماشین در تحقیقات رفتاری سفر	Koushik et al. (2020)	تشخیص الگو و تجزیه و تحلیل داده‌های سفر افراد به کمک روش‌های یادگیری ماشین	کاربرد زیاد یادگیری عمیق و مدل‌های شبکه‌های عصبی و همچنین عدم تفسیر پذیری نتایج
۱۳	یادگیری عمیق با شبکه‌های حافظه کوتاه‌مدت برای طبقه‌بندی الگوهای سفر مرتبط با زوال عقل	Vuong et al. (2020)	به‌دست آوردن الگوهای رفتاری خاص و عجیب از داده‌های تلفن همراه افراد	شناسایی افراد مبتلا به آلزایمر از روی داده‌های تلفن همراه به کمک مدل‌های یادگیری عمیق
۱۴	شبکه‌های عصبی برای پیش‌بینی اهداف سفر حمل‌ونقل عمومی از داده‌های کارت هوشمند فردی	Sari et al. (2021)	به کمک داده‌های کارت هوشمند فردی هدف سفرافراد اعم از کاری، خانه و سرگرمی مشخص می‌شود.	به کمک شبکه عصبی مصنوعی اهداف سفر افراد با دقت بالا پیش‌بینی شده است.
۱۵	استنباط ویژگی‌های اجتماعی و اقتصادی از الگوهای سفر	Bakhtiari et al. (2023)	شناسایی گروه‌های مختلفی سنی، جنسیتی و درآمدی به کمک الگوی سفرافراد	با استفاده از شبکه عصبی مصنوعی و سایر روش‌ها با استفاده از الگوی سفرها سن و جنسیت و درآمدافراد مشخص می‌شود.

۳- روش‌شناسی تحقیق



شکل ۱. فلوچارت روش تحقیق

۳-۱- مفاهیم اولیه در طراحی مدل

در تحقیقات قبلی برای مشخص کردن خصوصیات اجتماعی– اقتصادی افراد سعی شده است از روش‌های مختلف یادگیری ماشین و شبکه‌های عصبی استفاده شود. ازشبکه‌های عصبی مکرر برای پیش‌بینی خصوصیات اقتصادی– اجتماعی افراد استفاده شده است. در این تحقیق هدف این است که وابسته بودن سفرهای افراد به هم در یک روز مورد بررسی قرا گیرد و مشخص شود که این وابستگی در سفرها تا چه میزان می‌تواند در پیش‌بینی خصوصیات افراد موثر باشد. برای این منظور از مدل‌های دنباله‌ای و شبکه عصبی LSTM استفاده شده است. در این مطالعه نحوه مرتب کردن داده‌ها با یک روند جدید آغاز شده است. به این صورت که داده‌های ورودی مدل براساس نوع فعالیت به صورت دنباله‌ای مرتب شدند. ایده اولیه هم این گونه است که همواره سفرها ونوع فعالیت افراد در طول روز به فعالیت‌ها و سفرهای قبلی آنها وابسته است و چون مدل‌های دنباله‌ای به جملات قبلی دائماً بازگشت دارند بنابراین از این نوع مدل استفاده می‌شود. سعی می‌شود از شبکه‌های عصبی بازگشتی که حافظه طولانی مدت و کوتاه مدت دارند استفاده شود تا رفتارهای قبلی افراد درنظر گرفته شود و در رفتارهای جدید هم استفاده شود.

۳-۲- زبان برنامه نویسی

در این تحقیق در همه مراحل آماده‌سازی و پردازش اولیه داده‌ها و همچنین مدل‌سازی اصلی و تحلیل خروجی‌ها از زبان برنامه‌نویسی پایتون استفاده شده است. این زبان به دلیل کارایی سطح بالا و قدرت تحلیل قوی و همچنین استفاده ساده و راحت از بهترین زبان‌های برنامه‌نویسی در این سال‌ها است.

۳-۲- جمع‌آوری داده‌ها

با توجه به اینکه دراین تحقیق از داده های موبایل استفاده می‌شود و چنین داده‌هایی در دسترس وجود ندارد بنابراین از روش دیگری استفاده می‌شود. در دسترس‌ترین و کامل‌ترین این داده‌ها ازنظرسنجی ملی مسافرت خانگی قابل استفاده است. نظرسنجی ملی مسافرت خانگی منبع اطلاعات کشور آمریکا در مورد سفر ساکنان ایالات متحده در تمام ۵۰ ایالت و ناحیه کلمبیا است. این فهرست رفتار سفر، شامل سفرهایی است که توسط همه شیوه‌های سفر (وسیله‌نقلیه شخصی، حمل و نقل عمومی، عابر پیاده و دوچرخه‌سواری) و برای همه مقاصد (سفر به محل کار، خانه، تفریحی، درمانی، مدرسه و دانشگاه) انجام می‌شود. آژانس‌های دولتی و منطقه‌ای از داده‌ها برای حمایت از مدل‌سازی تقاضای سفر و برنامه‌ریزی حمل و نقل دوربرد استفاده می‌کنند. داده‌های NHTS در گزارش‌های گسترده‌تر و دوسالانه به کنگره در مورد عملکرد سیستم حمل و نقل سطحی گنجانده شده است. داده‌های نظرسنجی نیز در خارج از حمل و نقل در زمینه‌های بهداشت عمومی، تجزیه و تحلیل محیطی، مطالعات استفاده از زمان، مصرف انرژی و رفاه اجتماعی استفاده می‌شود.

داده‌های NHTS به طور مستقیم از یک نمونه تصادفی طبقه بندی شده از خانواده‌های ایالات متحده جمع‌آوری می‌شود. تنوع اطلاعات موردنیاز برای استفاده در طیف وسیعی از تصمیمات حمل و نقل در محیط کنونی بر اهمیت داده‌های قابل اعتماد برای تصمیم‌گیری‌های مبتنی بر شواهد تأکید دارد. NHTS داده‌هایی را در مورد روندهای رفتار مسافرتی فردی و خانگی مرتبط با عوامل اقتصادی، جمعیتی و جغرافیایی که بر تصمیمات سفر تأثیر می‌گذارد و برای پیش بینی تقاضای سفر استفاده می‌شود، ارائه می‌دهد. این نظرسنجی به‌عنوان نمونه‌گیری بر اساس آدرس طراحی شد تا نمونه‌ای با احتمال مساوی از خانوارها به دست آید. فایل نمونه ملی برای تولید ۲۶۰۰۰ خانوار تکمیل شده طراحی و انتخاب شد. نمونه‌ای از آدرس‌ها در دو مقطع زمانی انتخاب شد: ابتدا در ژانویه ۲۰۲۱ قبل از شروع جمع‌آوری داده‌های مطالعه اصلی و دوباره در سپتامبر ۲۰۲۱ تقریباً در نیمه راه جمع‌آوری داده‌های مطالعه اصلی، با استفاده از آخرین چارچوب نمونه‌گیری که توسط گروه سیستم بازاریابی به روز شده است. این طرح امکان تنظیم اندازه نمونه را فراهم می‌کند. بسته‌های پرسشنامه و بسته‌های یادآوری برای بیش از ۹۰۰۰۰۰ خانوار پست شد. بیش از ۲۳۹۰۰۰ خانوار (۲۷٪) پرسشنامه کاغذی را تکمیل و برگرداندند. ۱۳۰۰۰ (۱٫۴٪) پرسشنامه اضافی از طریق وب یا تلفن تکمیل شد. تقریباً ۷٫۵ درصد از بسته‌های پرسشنامه پست شده به‌عنوان «تحویل‌نشده» توسط اداره پست بازگردانده شد.

پس از بررسی هر نظرسنجی، به طور خودکار تاریخ سفر به خانواده اختصاص داده شد. تاریخ سفر تعیین شده از طریق یک بسته پستی گزارش سفر به هر خانواده اطلاع داده شد. گزارش سفر برای همه افراد خانوار از سن ۵ سال و بالاتر تکمیل می‌شود. تاریخ سفر خانواده بر اساس روز از پیش انتخاب شده هفته بود

که به طور تصادفی به خانواده اختصاص داده شد. تاریخ سفر حداقل ۱۰ روز پس از تاریخی که بسته پرسشنامه اولیه پردازش شد، تعیین شد، که زمان را برای دریافت بسته اطلاعات سفر قبل از تاریخ سفر به خانوار می‌داد. این به خانواده زمان می‌دهد تا روند را درک کنند و اگر قبل از تاریخ سفر سؤالی داشتند، با خط تلفن مرکز کمک رایگان پروژه تماس بگیرند(NHTS Data User Guide, 2022).

پس از روز سفر افراد پیام‌های یادآوری دریافت کردندکه مشخصات سفر را ارسال کنند و کار پرسشنامه تکمیل شود. درپایان از شرکت کنندگان تشکر و ۲۰ دلار به عنوان تشویق به آنها پرداخت شد.

پس از مرتب کردن داده‌ها در پایان کل اطلاعات خانوار و سفر آنها در ٤ فایل آماده سازی و مرتب شد:

فایل خانوار: داده‌هایی که یک بار برای خانواده جمع آوری می‌شود (یک رکورد در هر خانوار).

فایل شخص: اطلاعات جمع آوری شده برای هرشخص (یک رکورد برای هر عضو خانواده).

فایل خودرو: اطلاعات مربوط به وسایل نقلیه خانوار (یک رکورد برای هر وسیله نقلیه خانگی).

فایل سفر یک روزه: اطلاعات جمع آوری شده برای هر سفری که هر یک از اعضای خانواده بالای ۵ سال در روز سفر خانواده انجام داده است (یک رکورد برای هر سفری که هر فرد بالای ۵ سال انجام داده است) (NHTS Data User Guide, 2022).

داده‌ها در ٤ دسته فوق تنظیم شد و این چهار دسته از وب‌سایت NHTS دریافت شد و سپس مرحله بعدی پیش پردازش داده‌ها انجام شده است.

٤-۲- روش پیش پردازش داده‌ها

پس از مطالعه کامل این ٤ فایل و بررسی دقیق آنها، مرحله پیش پردازش داده‌ها آغاز گردید. در این مرحله ازبین کل داده‌های جمع‌آوری شده، آن داده‌هایی که نیاز است را انتخاب کرده و مرتب می‌شوند. با توجه به اینکه هدف اصلی در این تحقیق استفاده از اطلاعات داده‌های موبایل افراد است بنابراین صرفاً از داده‌های خاصی استفاده شده است. داده‌هایی که در ابتدا انتخاب شدند به ترتیب به صورت زیر هستند.

۱- شناسه خانوار: در این ستون به هر خانوار یک شناسه داده می‌شود که در فایل‌ها از عدد ۳۰۰۰۰۰۰۷ شروع می‌شود.

۲- شناسه شخص: عددی که برای معرفی هر یک از افراد یک خانوار تعریف می شود و ازیک شروع می شود.

۳- زمان شروع سفر: زمان شروع سفر شخص در روز سفر را نشان می‌دهد.

٤- زمان پایان سفر: زمان پایان سفر شخص در روز سفر را نشان می‌دهد.

۵- زمان سفر شخص: زمان سفر یک شخص را مشخص می‌کند.

٦- مسافت: مسافتی که شخص طی می کند را بر حسب واحد مایل نشان می‌دهد.

۷- نوع وسیله نقلیه: نوع وسیله نقلیه‌ای که شخص با آن سفر می‌کند را نشان می‌دهد.

۸- هدف سفر: در این ستون هدف سفر ازجمله سفرخانه مبنای کاری، سفرخانه مبنای اجتماعی / تفریحی، سفرخانه مبنای خرید، سفرخانه مبنای دیگر و سفرهای غیرخانه مبنا مشخص شده‌اند.

۹- زمان ماندن در مقصد: زمانی که شخص در مقصد سپری می‌کند و پس از آن سفر بعدی را شروع می‌کند.

۱۰ – مبدا سفر : محل شروع سفر را مشخص می‌کند که با توجه به هدف سفر مکان‌های مختلفی می‌تواند باشد. جدول ۲ انواع مبدا سفر با کدهای مشخص کننده آنها را نشان می‌دهد.

۱۱- مقصد سفر: محل مقصد سفر را نشان می‌دهد. جدول ۱ انواع مقصد سفر با کدهای مشخص کننده آنها را نشان می‌دهد.

در ستون های دیگر برچسب‌های مدل به منظور آموزش و تست مدل قرار داده شده است. با توجه به اینکه قراراست جنسیت، سن و درآمد افراد تخمین زده شود این موارد به عنوان برچسب در سه ستون قرار داده شده است. البته در مورد سن افراد، در نهایت برای مدل سازی، افراد زیر ۲۰ سال حذف شدند چون این افراد الگوی سفر ثابتی دارند.

در سن‌های کودکی افراد با پدر و مادر سفر می‌کنند و در حالت‌های دیگر هم در طول روز فقط یک سفر اصلی به مدرسه دارند، بنابراین این افراد از مدل حذف شده‌اند. موارد فوق فقط شامل اطلاعاتی است که برای مدل سازی استفاده شده و آنها از NHTS استخراج شده‌اند. سپس یکی از مراحل اصلی کار مدل‌سازی یعنی آماده کردن داده‌ها شروع می‌شود.

جدول ۱. مبدا و مقصد سفر

کد	هدف سفر
-۹	Not ascertained
-۸	I don't know
-۷	I prefer not to answer
۱	Regular home activities (chores, sleep)
۲	Work from home (paid)
۳	Work
۴	Work-related meeting / trip
۵	Volunteer activities (not paid)
۶	Drop off /pick up someone
۷	Change type of transportation
۸	Attend school as a student
۹	Attend child care
۱۰	Attend adult care
۱۱	Buy goods (groceries, clothes, appliances, gas)
۱۲	Buy services (dry cleaners, banking, service a car, pet care)
۱۳	Buy meals (go out for a meal, snack, carry-out)
۱۴	Other general errands (post office, library)
۱۵	Recreational activities (visit parks, movies, bars, museums)
۱۶	Exercise (go for a jog, walk, walk the dog, go to the gym)
۱۷	Visit friends or relatives
۱۸	Health care visit (medical, dental, therapy)
۱۹	Religious or other community activities
۹۷	Something else

۳- نتایج و بحث

پس از کار و اجرای تعداد زیادی مدل روی داده‌ها و نمونه‌های مختلف مدل نهایی به‌دست آمد و این مدل روی کل داده‌ها اجرا شد. در این بخش همه داده‌هایی که از نظرسنجی مسافرت ملی آمریکا در سال ۲۰۲۲ برداشت شده بود، در نظر گرفته شد. تعداد کل داده‌ها پس از تمیز کردن ۲۱۹۱۵۵ داده می‌باشد. این افراد کسانی هستند که اطلاعات کاملی از فعالیت آنها در طول روز موجود بود. این اطلاعات در فایل edited_data_Fainal بارگذاری شد و به عنوان ورودی به مدل نهایی داده شد.

۳-۱- مدل جنسیت

پس از اجرای مدل روی کل داده‌ها نتایج روی داده‌های آموزشی به صورت جدول ۲ داده شده است. این نتایج روی داده‌های تست به صورت جدول ۳ داده شده است.

جدول ۲. نتایج مدل جنسیت روی داده‌های آموزشی

Evaluation method	نتایج داده های آموزشی
accuracy_score	۰,۵۵۹۱۷٤۰۸۵۵۱٦۳۳۳٦
precision_score	۰,۵۸۹٦۵۵۸۲۸۱۰٤۲۰۲٤
recall_score	۰,۲۳۸۵۳۳۸۷٤۳٤۸۵۷۰۲
f1_score	۰,۳۳۹٦٦۳٤۵۸۲٤۷۱۸۹۸

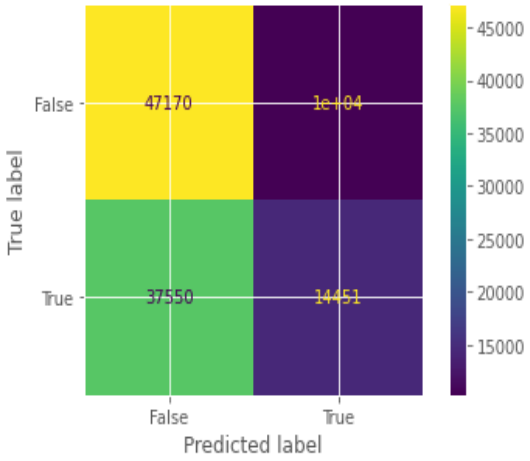
جدول ۳. نتایج مدل جنسیت روی داده‌های تست

Evaluation method	نتایج داده های تست
accuracy_score	۰,۵۶۲۹۷۱۲۹۰۶۸۵۲۳۹۵
precision_score	۰,۵۹۸۴۴۸۵۱۹۰۴۰۹۰۲۷
recall_score	۰,۲۴۴۷۷۹۰۴۶۹۵۹۷۳۲۳۲
f1_score	۰,۳۴۷۴۴۵۱۳۵۹۳۱۸۷۰۳

ماتریس درهم ریختگی برای داده‌های آموزشی و تست به صورت زیر است.

ماتریس درهم ریختگی داده‌های آموزشی

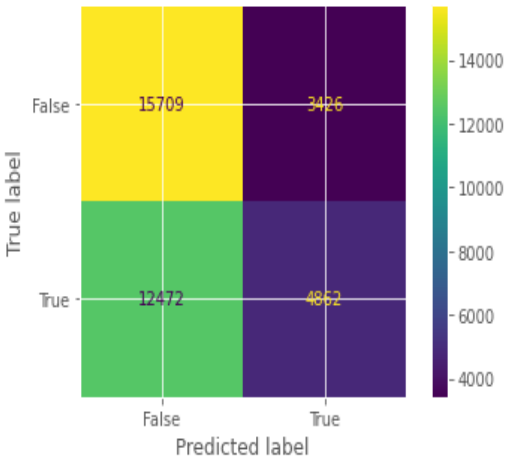
با توجه به خروجی مدل داده‌هایی که درست یا به صورت نادرست پیش‌بینی شده‌اند در ماتریس شکل ۲ نشان داده شده است.



شکل ۲. ماتریس درهم ریختگی داده‌های آموزشی

ماتریس درهم ریختگی داده‌های تست

برای داده‌های تست نتایج پیش بینی داده‌ها در شکل ۳ بررسی شده است.



شکل ۳. ماتریس در هم ریختگی داده‌های تست

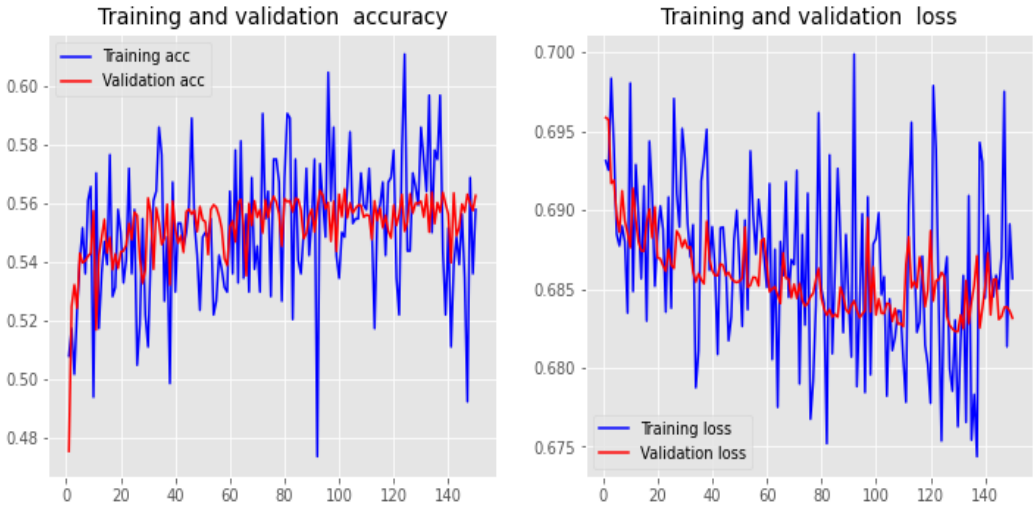
حال با بررسی وضعیت خطا در داده‌های آموزشی و اعتبارسنجی می‌بینیم که پس از ۱۵۰ تکرار خطا درحال کاهش است و وضعیت نمودار روی داده‌های اعتبارسنجی بهتر بوده و نوسان کمتری دارد. همچنین میزان دقت مدل هم در حال افزایش است.

با توجه به نتایج فوق مدل ساخته شده در تخمین زنان خیلی بهتر عمل می کند یعنی وابستگی‌های سفرهای زنان را بهتر تشخیص داده است ولی درمورد مردان با توجه به اینکه سفرهای طولانی‌تر و متنوع‌تری دارند در تشخیص وابستگی سفرهای آنها دقت کمتری دارد.

درحالت کلی با بررسی سفرهای افراد روی نمونه ۵۰۰۰ تایی از زنان و ۵۰۰۰ تایی از مردان این نتیجه حاصل شد که مسافت پیموده شده توسط مردان خیلی بیشتر از زنان است. میانگین مسافت پیموده شده برای مردان در هرسفر ٦/٧١ مایل است ولی میانگین مسافت طی شده توسط خانم ها ٣/٤٦ مایل می‌باشد. بنابراین زنان سفرهای کوتاهتری نسبت به زنان دارند. تعداد سفرها هم برای مردان بیشتر از زنان است. با توجه به اینکه سفرهای مردان از زنان بیشتر است و مسافت بیشتری طی می‌کنند، بنابراین الگوی سفرهای مردان پیچیدگی بیشتری نسبت به زنان دارد و درک این الگو توسط مدل مشکل‌تر است. به همین دلیل دقت مدل برای تشخیص مردان پایین‌تر از زنان می‌باشد.

دراین حالت هم مانند سایر مدل‌های یادگیری ماشین یکی از موارد مهم، بررسی مقدار خطا است که برای هر دو دسته داده‌های آموزشی و داده‌های اعتبارسنجی بررسی می‌شود. این کار درهر تکرار در روند آموزش به منظور بررسی مسائل بیش برازش و کم برازشی انجام می شود. برای این کار نمودار زیان در هر دو دسته داده‌های آموزشی و داده‌های اعتبارسنجی به عنوان یک از خروجی‌های مدل در نظرگرفته شده است.

در این مدل برای روند آموزش تعداد ۲۰۰ تکرار در نظر گرفته شده است. بررسی نمودارها به خوبی نشان می‌دهد که ابزارهای بیش برازشی مانند تابع حذف زودهنگام استفاده شده در مدل به خوبی کار کرده است و مدل را در تکرار ۱۳۵ متوقف کرده است و مانع از بیش برازشی مدل شده است. دراین شرایط بهترین وزن‌ها برای پارامترهای قابل آموزش درنظر گرفته شده است و تا این مرحله دقت مدل هم درحال افزایش می‌باشد.



شکل ٤. میزان خطا و دقت روی داده‌های آموزشی و اعتبارسنجی

۳-۱-۱- تحلیل نتایج مدل جنسیت با confusion matrix

در مدل نهایی همان‌طور که قبلاً گفته شد ما کل داده‌ها را مورد بررسی قرار دادیم و تعداد داده‌ها در دو گروه زنان و مردان یکسان نیستند. با توجه به این موضوع همه داده‌ها در مدل وارد شد و تعداد ۱۰۹٤۰٦ داده آموزشی و ۳٦٤٦۹ داده تست در مدل وارد شد. در داده‌های تست تعداد ۱۹۱۳۵ زن و تعداد ۱۷۳۳٤ مرد مورد بررسی قرار گرفتند. طبق ماتریس درهم ریختگی (confusion matrix) درگروه مردان ٤۸٦۲ نفر به درستی پیش بینی شده‌اند و در گروه زنان ۱۵۷۰۹ نفر به درستی پیش‌بینی شده‌اند. یعنی مدل نسبت به زنان پیش بینی درست‌تری داشته و حساسیت بهتری نسبت به گروه زنان داشته است.

با بررسی داده‌های استفاده شده در مدل، این نتیجه حاصل شد که مسافت پیموده شده توسط مردان خیلی بیشتر از زنان است. میانگین مسافت پیموده شده برای مردان به ازای هرسفر ۸/۲۱ مایل است ولی میانگین مسافت طی شده توسط خانم‌ها ٤/۳۵ مایل می‌باشد. بنابراین زنان سفرهای کوتاه‌تری نسبت به زنان دارند. این سفرها بیشتر خانه مبنا هستند. تعداد سفرها برای مردان بیشتر از زنان است. با توجه به اینکه سفرهای مردان از زنان بیشتر است و مسافت بیشتری طی می‌کنند بنابراین الگوی سفرهای مردان پیچیدگی بیشتری نسبت به زنان دارد و درک این الگو توسط مدل LSTM مشکل‌تر است. به همین دلیل دقت مدل برای تشخیص مردان پایین‌تر از زنان می‌باشد.

۳-۱-۲- تحلیل نتایج مدل جنسیت با معیارصحت

همان‌طور که در بخش قبل توضیح داده شد، پارامتر صحت (Accuracy)، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بندی است و عبارت است از میزان تشخیص صحیح دسته‌بندی در مجموع دو دسته که از رابطه (۱) به دست می‌آید.

(۱)

$$Accuracy = (TP+TN) / (TP+FN+FP+TN)$$

در دسته‌بندی فوق این عدد برابر ۰,۵٦۲۹ معادل ۵٦,۲۹٪ به دست آمد. همان‌طور که در قسمت قبل هم بررسی شد مدل در دسته مردان دقت خوبی ندارد و به همین دلیل دقت نهایی به این عدد رسید. درگروه زنان دقت بیشتر از ۸۵٪ داریم.

۳-۱-۳- تحلیل نتایج مدل جنسیت با معیاردقت (Precision)

حداکثر مقدار این معیار یک و یا ۱۰۰ درصد و حداقل مقدار آن صفر است و هرچه مواردی که برنامه به غلط پیش‌بینی کرده است که به آن False Positive گفته می‌شود نسبت به پیش‌بینی‌های درست یا True Positive بیشتر باشد مقدار Precision کمتر خواهد شد. درواقع معیار پرسیژن، نشان می‌دهد که چقدر از نتایجی که توسط مدل مثبت اعلام شده‌اند، واقعا مثبت هستند. مقدار آن از رابطه (۲) به دست می‌آید.

(۲)

Precision = TP / (TP + FP)

این عدد در مدل ۵۹٫۸٤٪ است و نشان می‌دهد که ازبین همه پیش‌بینی‌هایی که در مدل مرد پیش‌بینی شده است ۵۹٫۸٤ درصد درست پیش‌بینی شده است.

۳-۱-٤- تحلیل نتایج مدل جنسیت با معیار Recall

معیار ریکال، نشان می‌دهد که چقدر از موارد مثبت داده شده به مدل، پیدا شده و به عنوان نمونه مثبت اعلام شده‌اند. یعنی از بین گروه مردان چقدر به درستی پیش‌بینی شده است.

(۳)

Recall = TP / (TP + FN)

دراین قسمت با توجه به اینکه تعداد مردان ٤۲٤۳ نفر از ۱۷۳۳٤ نفر پیش‌بینی شده بود بنابراین این عدد ۲٤٫٤۷٪ پیش‌بینی شده است که خود نشان‌دهنده این است که مردان به خوبی پیش‌بینی نمی‌شوند.

۳-۱-۵- تحلیل نتایج مدل جنسیت با معیارF1_score

زمانی که می‌خواهند معیار ارزیابی میانگینی از دو مورد قبلی باشد یعنی همان Recall و Precision می‌توان از میانگین هارمونیک این دو معیار استفاده کنید که به آن معیار f1-score می‌گویند.

(٤)

$$F_1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = \frac{2TP}{2TP + FP + FN}$$

در این قسمت میانگین هارمونیک این دو معیار ۳٤٫۷٤٪ می‌باشد. با توجه به اینکه دراین مدل برای زنان پیش‌بینی بهتری دارد، روی این نمونه نتیجه می‌گیریم که مدل‌های سری زمانی برای کسانی که عموما سفرهای ساده‌تری انجام می‌دهند، پیش‌بینی جنسیتی بهتری دارد. یعنی این مدل‌ها زمانی مفید هستند که الگوی سفرافراد پیچیدگی کمتری داشته باشد.

۳-۳- مدل سن افراد

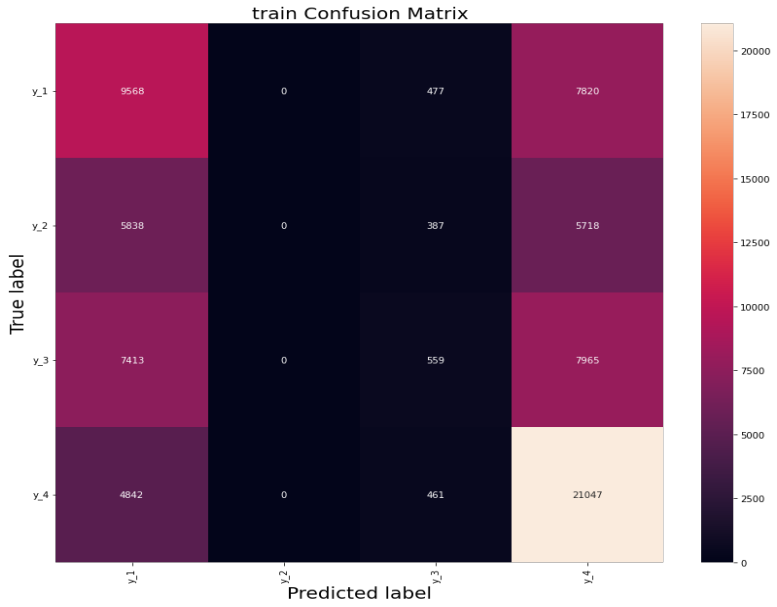
همان طور که در قسمت‌های قبلی توضیح داده شد برای سن افراد ۷ دسته در نظر گرفتیم و مدل را بر اساس این ۷ دسته اجرا کردیم. با توجه به نوع پراکندگی سن کل افراد شرکت‌کننده در تحقیق در بعضی از گروه‌ها تعداد افراد بسیار کم بود و مانند تحقیقات قبلی سن افراد در ٤ دسته بررسی شد. دسته اول شامل ٤۰ – ۲۰ سال و دسته دوم شامل ۵۰–٤۰ سال و دسته سوم شامل ٦۰ – ۵۰ سال و دسته آخر بیشتر از٦۰ سال درنظر گرفته شد.

جدول ٤. نتایج نهایی دسته‌بندی سن افراد روی داده‌های آموزش

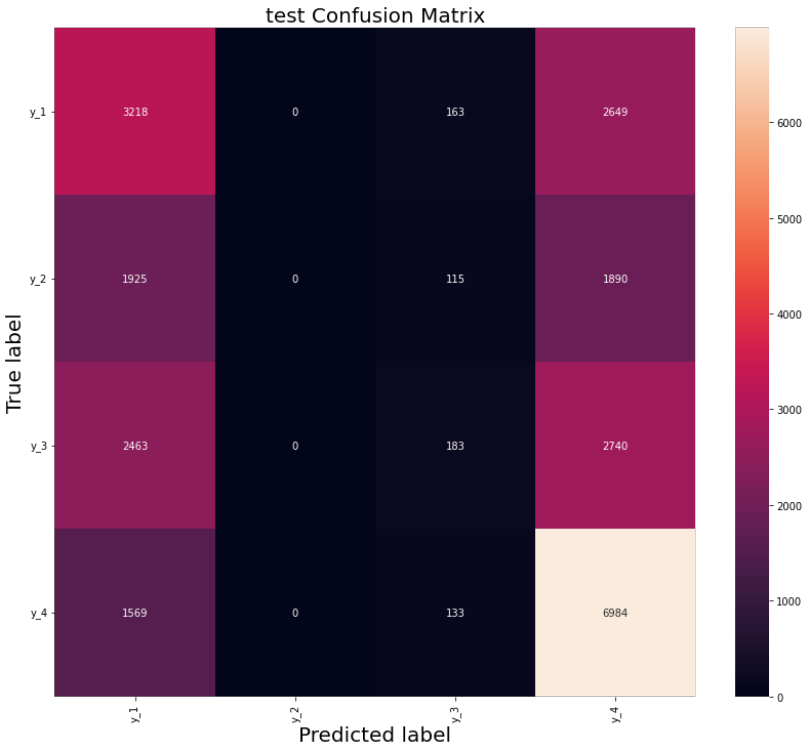
train	y_1	y_2	y_3	y_4
accuracy_score	۰٫۵۳۵۵	۰	۰٫۰۳۵	۰٫۷۹۸۷
train_precision	۰٫۳٤۵۹	۰	۰٫۲۹٦۷	۰٫٤۹٤٦
train_recall	۰٫۵۳۵۵	۰	۰٫۰۳۵	۰٫۷۹۸۷
train_f1_score	۰٫٤۲۰۳۳	۰	۰٫۰٦۲۷۳	۰٫٦۱۰۹

جدول ۵. نتایج نهایی دسته‌بندی سن افراد روی داده‌های تست

test	y_1	y_2	y_3	y_4
accuracy_score	۰٫۵۳۳٦	۰	۰٫۰۳۳۹۷	۰٫۸۰٤۰
test_precision	۰٫۳۵۰۷	۰	۰٫۳۰۸۰	۰٫٤۸۹٦
test_recall	۰٫۵۳۳٦	۰	۰٫۰۳۳۹۷	۰٫۸۰٤۰
test_f1_score	۰٫٤۲۳۲	۰	۰٫۰٦۱۲۰	۰٫٦۰۸٦



شکل ٤. ماتریس درهم ریختگی داده‌های آموزشی

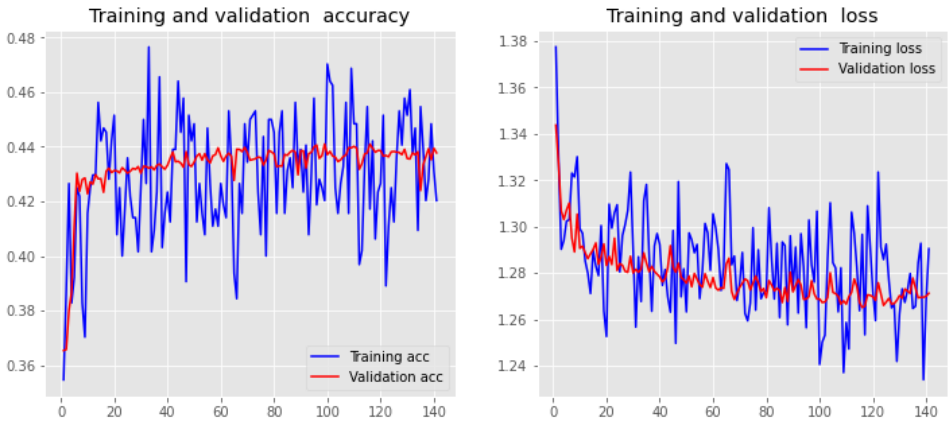


شکل ۵. ماتریس درهم ریختگی داده‌های تست

در این حالت هم وضعیت خطا در داده‌های آموزشی و اعتبارسنجی بررسی می‌شود. این کار در هر تکرار در روند آموزش به منظور بررسی مسائل بیش برازش و کم برازشی انجام می‌شود. برای این کار نمودار زیان در هر دو دسته داده‌های آموزشی و داده های اعتبارسنجی به عنوان یک از خروجی‌های مدل مطابق شکل ٦ در نظرگرفته شده است. در این مدل برای روند آموزش تعداد ۲۰۰ تکرار در نظر گرفته شده است. بررسی نمودارها به خوبی نشان می‌دهد که ابزارهای بیش برازشی مانند

تابع حذف زودهنگام استفاده شده در مدل به خوبی کار کرده است و مدل را در تکرار۱۴۱ متوقف کرده است و مانع از بیش برازشی مدل شده است. دراین شرایط بهترین وزن‌ها برای پارامترهای قابل آموزش در نظر گرفته شده است. در حالتی که تعداد داده‌ها مساوی بودند میانگین دقت روی داده‌های تست برابر ۲۵,۲۸ درصد می‌باشد که این مدل روی داده‌های افراد بین ۲۰-۳۰ سال و افراد بالای ۸۰ سال بهتر عمل می‌کند. با توجه به اینکه در هر طبقه سنی ۵۴۳ عضو در نظر گرفته شد، با بررسی داده‌ها مشخص شد که دربازه سنی ۳۰ تا ۷۰ سال تعداد سفرهای افراد بیشتر است. افراد بین ۳۰-۲۰ سال دارای الگوی سفرهای ساده‌تری هستند. بسیاری از این افراد هنوز مشغول به تحصیل در دانشگاه هستند یا شاغل نیستند. میانگین سفرهای این افراد ۲/۳ سفر می باشد، بنابراین الگوی سفر بسیار ساده‌تری دارند و تشخیص آن برای مدل LSTM به سادگی اتفاق می‌افتد. ولی در سنین بالاتر با توجه به اینکه افراد مشغله‌های کاری فراوان و دارای فرزند و خانواده هستند تعداد سفرهای کاری، خرید، تفریحی و سفرهای مرتبط با فرزندان مانند سفر به مدرسه یا کلاس‌های مختلف برای فرزندان را دارند. بنابراین الگوی سفر پیچیده‌تری دارند. طبق بررسی صورت گرفته روی افراد بین ۴۰-۵۰ مشخص شد که این افراد میانگین تعداد سفر ۴/۷ در روز دارند و میانگین مسافت پیموده شده به ازای هر سفر توسط آنها ۱۳/۴ مایل می‌باشد. بنابراین این افراد وضعیت سفرهای پیچیده‌تری داشته‌اند که درک الگوی سفر آنها توسط مدل LSTM مشکل‌تر می‌باشد.

در این حالت هم وضعیت خطا در داده‌های آموزشی و اعتبارسنجی بررسی می‌شود. این کار درهر تکرار در روند آموزش به منظور بررسی مسائل بیش برازش و کم برازشی انجام می‌شود. برای این کار نمودار زیان در هر دو دسته داده‌های آموزشی و داده‌های اعتبارسنجی به عنوان یک از خروجی‌های مدل در نظرگرفته شده است. دراین مدل برای روند آموزش تعداد ۲۰۰ تکرار در نظر گرفته شده است. بررسی نمودارها به خوبی نشان می‌دهد که ابزارهای بیش برازشی مانند تابع حذف زودهنگام استفاده شده در مدل به خوبی کار کرده است و مدل را در تکرار۶۹ متوقف کرده است و مانع از بیش برازشی مدل شده است. دراین شرایط بهترین وزن‌ها برای پارامترهای قابل آموزش در نظر گرفته شده است و تا این مرحله دقت مدل هم در حال افزایش‌می باشد.



شکل ٦. میزان خطا و دقت روی داده‌های آموزشی و اعتبارسنجی

۳-۲-۱- تحلیل نتایج مدل سن با confusion matrix

در مدل نهایی همانطور که قبلاً گفته شد ما کل داده‌ها را مورد بررسی قرار دادیم و تعداد داده‌ها در ٤ گروه بررسی شد. با توجه به این موضوع همه داده‌ها در مدل وارد شد و تعداد ۱۰۹٤۰٦ داده آموزشی و ۳٦٤٦۹ داده تست در مدل وارد شد. طبق ماتریس درهم ریختگی بهترین پیش‌بینی‌ها در گروه‌های اول و چهارم اتفاق افتاده است. یعنی گروه‌هایی در سنین ٤۰-۲۰ و بیشتر از ٦۰ بوده‌اند، بهترین پیش‌بینی را داشته‌اند.

۳-۲-۲- تحلیل نتایج مدل سن با معیارصحت

ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بندی است و عبارت است از میزان تشخیص صحیح دسته‌بندی در مجموع دو دسته که از رابطه (۵) به دست می‌آید.

(۵)

$$Accuracy = (TP+TN) / (TP+FN+FP+TN)$$

برای گروه ۱ مقدار ۵۳,۳۶٪ و برای گروه ٤ مقدار ۸۰,٤۰٪ را در نظر گرفتیم.

۳-۲-٤- تحلیل نتایج مدل سن با معیاردقت

برای گروه ۱ برابر ۳۵ درصد و برای گروه ٤ برابر ٤۸,۹٦ درصد می‌باشد.

۳-۲-۵- تحلیل نتایج مدل سن با معیار Recall

معیار ریکال، نشان می‌دهد که چقدر از موارد مثبت داده شده به مدل، پیدا شده و به عنوان نمونه مثبت اعلام شده‌اند. یعنی ازبین گروه مردان چقدر به درستی پیش بینی شده است.

(٦)

$$Recall = TP / (TP + FN)$$

این عدد برای برای گروه ۱ برابر ۵۳,۳٦ درصد و برای گروه ٤ برابر ۸۰,٤۰ می باشد که نشان‌دهنده پیش بینی خوب در این دسته‌ها می‌باشد.

۳-۲-٦- تحلیل نتایج مدل سن با معیارF1_score

زمانی که می‌خواهند معیار ارزیابی میانگینی از دو مورد قبلی باشد یعنی همان Recall و Precision می‌توان از میانگین هارمونیک این دو معیار استفاده کنید که به آن معیار f1-score می‌گویند.

(۷)

$$F_1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = \frac{2TP}{2TP + FP + FN}$$

در این قسمت میانگین هارمونیک این دو معیار ۳۳,۹۹٪ می‌باشد. این مدل برای داده‌های بین ٤۰ – ۲۰ سال و سن بالای ٦۰ سال پیش بینی خوبی داشته است و در بقیه حالات نتایج دقیقی نداده است. با بررسی داده‌ها در بین گروه‌های مختلف سنی این نتیجه حاصل شد که تعداد سفرها و همچنین مسافت پیموده شده برای افراد گروه‌های سنی بین ٦۰-٤۰ بیشتراز سایر گروه‌ها است. بنابراین این افراد الگوی سفر پیچیده‌تری داشته‌اند که توسط مدل‌های سری زمانی درک این الگو سخت‌تر بوده است. این افراد معمولاً کسانی در جامعه هستند که دارای خانواده و فرزند هستند و با توجه به سفرهایی که برای فرزندان و خرید و تفریحی با خانواده انجام داده اند، تعداد سفرهای بیشتری داشته‌اند.

۳-۳- مدل درآمد افراد

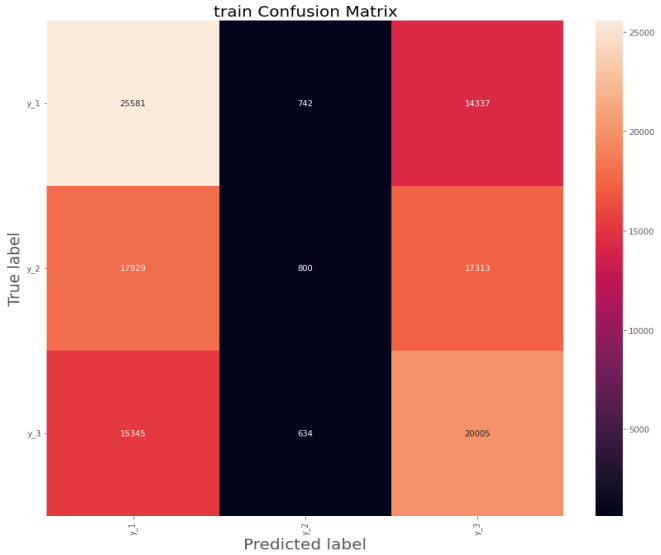
در این مدل برای درآمد افراد مدل اجرایی شد. با توجه به اینکه در دسته‌بندی‌های قبلی که برای درآمد ۱۱ طبقه درنظر گرفته بودیم نتایج قایل قبولی در مدل‌ها نداشتیم، برای درآمد افراد هم مانند تحقیقات قبلی سه دسته درنظر گرفتیم که شامل دسته اول درآمد سالیانه کمتر از ۵۰۰۰۰ دلار و دسته دوم درآمد سالیانه بین ۵۰۰۰۰ تا ۱۰۰۰۰۰ دلار و دسته سوم درآمد سالیانه بیشتر از ۱۰۰۰۰۰ دلاررا شامل می شوند. با اجرا کردن مدل‌ها نتایج برای داده‌های آموزشی و تست به ترتیب در جداول ٦ و ۷ دیده می‌شود.

جدول ٦. نتایج نهایی دسته‌بندی درآمد افراد روی داده‌های آموزش

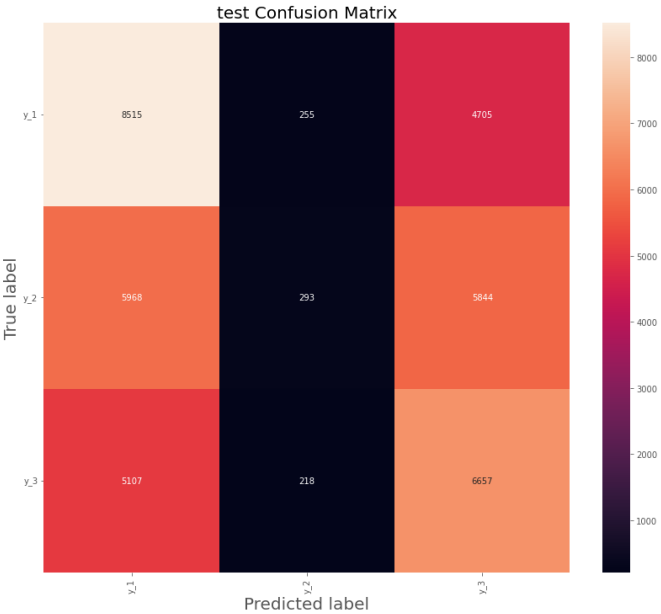
train	y_1	y_2	y_3
accuracy_score	۰,۷۳۷۷	۰,۰۳۲۲۹	۰,٤۲۲۸
train_precision	۰,٤۱۷۲	۰,۳۵۸٤	۰,٤۰۵۲
train_recall	۰,۷۳۷۷	۰,۰۳۲۲۹	۰,٤۲۲۸
train_f1_score	۰,۵۳۳۰	۰,۰۵۹	۰,٤۱۳۸

جدول ۷. نتایج نهایی دسته‌بندی درآمد افراد روی داده‌های تست

test	y_1	y_2	y_3
accuracy_score	۰,۷۴۲۱	۰,۰۳۱۷۲	۰,۴۳۳۰
train_precision	۰,۴۱۷۸	۰,۳۵۵۵	۰,۴۰۴۲
train_recall	۰,۷۴۲۱	۰,۰۳۱۷۲	۰,۴۳۳۰
train_f1_score	۰,۵۳۴۶	۰,۰۵۸۲۴	۰,۴۱۳۵



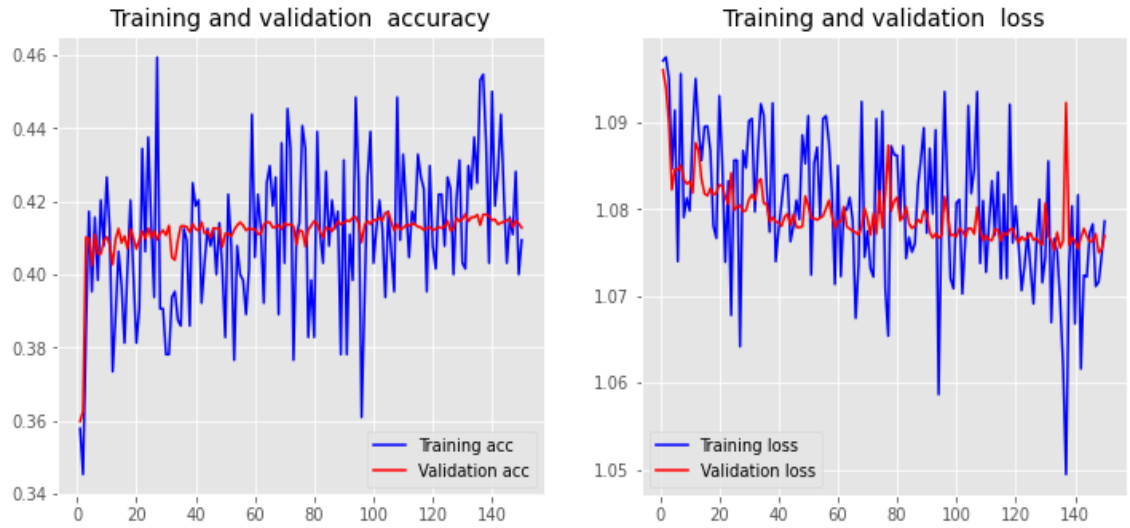
شکل ۷. ماتریس درهم ریختگی داده‌های آموزشی



شکل ۸. ماتریس درهم ریختگی داده‌های تست

در این حالت هم وضعیت خطا در داده‌های آموزشی و اعتبارسنجی بررسی می‌شود. این کار درهر تکرار در روند آموزش به منظور بررسی مسائل بیش برازش و کم برازشی انجام می‌شود. برای این کار نمودار زیان در هر دو دسته داده‌های آموزشی و داده‌های اعتبارسنجی به عنوان یک از خروجی‌های مدل مطابق شکل ۹ در نظرگرفته شده است. دراین مدل برای روند آموزش تعداد ۲۰۰ تکرار در نظر گرفته شده است. بررسی نمودارها به خوبی نشان می‌دهد که ابزارهای بیش برازشی مانند تابع حذف زودهنگام استفاده شده در مدل به خوبی کار کرده است و مدل را در تکرار ۱۵۰ متوقف کرده است و مانع از بیش برازشی مدل شده است. دراین شرایط بهترین وزن‌ها برای پارامترهای قابل آموزش در نظر گرفته شده است.

بنابراین خروجی‌های متوسط برای سه طبقه درآمدی به صورت جدول ۸ خواهد بود.



شکل ۹. میزان خطا و دقت روی داده‌های آموزشی و اعتبارسنجی

بنابراین T در این حالت میانگین درآمدی ٤١/١٧٪ می‌باشد. درمدل درآمد افراد نیز خروجی نهایی برای طبقه اول و سوم بهتر از طبقه دوم پیش بینی شده است. با بررسی الگوی سفرهای افراد روی داده‌ها مشخص شد که طبقه متوسط درآمدی سفرهای بیشتری دارند و مسافت‌های بیشتری را طی می‌کنند. بنابراین الگوی سفر پیچیده‌تری دارند. نتایج حاصل همه موارد به همراه حالات بهینه بررسی شد و خروجی‌های مدل ارزیابی گردید. حال نتایج مدل را با تحقیقات قبلی انجام شده بررسی می‌کنیم. در تحقیقات گذشته پارامترهای اقتصادی- اجتماعی با مدل شبکه عصبی بررسی شده است ولی در این تحقیق از مدل‌های دنباله‌ای استفاده شد که وابستگی زمانی سفرهای افراد در این مدلها مورد توجه قرار گرفته است. نتایج کلی نشان می‌دهد که مدل‌های LSTM برای درک الگوی سفرهای پیچیده افراد به صورت محسوسی ضعیف عمل می‌کنند. یعنی در شرایطی که افراد سفرهای بیشتری با مسافت‌های بیشتر طی کرده‌اند برای پیش بینی سن، جنسیت و درآمد افراد ضعیف عمل کردند ولی در شرایطی که افرادالگوی ساده‌تری برای سفرها دارند، پیش‌بینی این مدل‌ها بسیار خوب است و دقت بالایی دارند. دراین مورد یک تحقیق دیگر در مورد ویژگی‌های اقتصادی- اجتماعی افراد انجام گرفته است که نتایج را با هم مقایسه می‌کنیم.

جدول ۸ نتایج مدل درآمد روی داده های تست

test	average
accuracy_score	۰,٤۱۱۷
train_precision	۰,٤۰۱۳
train_recall	۰,٤۱۱۷
train_f1_score	۰,۳٤٤۹

عنوان تحقیق	تحقیق حاضر	نتیجه تحقیق	برتری و نوع‌آوری تحقیق حاضر
بررسی ویژگی‌های اقتصادی– اجتماعی	ویژگی‌های اقتصادی– اجتماعی مسافران با lstm	مشخص کردن سن، جنسیت و درآمد افراد سفر کننده	۱- مدل سازی با روش مدل‌های دنباله‌ای ۲- استفاده از فعالیت روزانه افراد به صورت یک جمله

٤- نتیجه‌گیری

در این پژوهش، برای استنباط ویژگی‌های اقتصادی- اجتماعی سفرهای افراد در یک روز بررسی شد از آنجایی که سفرهای هر فرد در طول یک روز به‌هم وابسته است بنابراین درک این وابستگی‌ها در مدل بسیار مهم است. مثلاً اگر شخصی ابتدای صبح یک سفر از منزل به محل کار خود دارد به ناچار سفر برگشتی هم از محل کار به منزل خواهد داشت. پس اگر از مدلی استفاده شود که این وابستگی‌ها را درک کند می‌توان انتظار داشت که نتایج بهتری بدست آید. مهم‌ترین نگاهی که در این تحقیق وجود داشت این این است که آیا می‌توان مدل هایی ساخت که وابستگی‌های بین سفرهای افراد را درک کند؟ ازجمله دست آوردهای این تحقیق می‌توان به قدرت پیش‌بینی بیش از ۸۰ درصدی مدل روی جنسیت زنان اشاره کرد که این بیانگر این واقعیت است که مدل‌ها توانسته‌اند درک خوبی روی الگوی سفر آنها داشته باشند. نوع آوری به کاررفته در این تحقیق این است که سفرهای افراد را مانند کلمات یک جمله درنظر گرفتیم و همانطور که مدل‌های دنباله‌ای کلمات یک جمله را ترجمه می‌کنند و وابستگی بین آنها را درک می‌کنند، دراین حالت هم سفرهای افراد و وابستگی آنها تشخیص داده شوند.

دراین تحقیق از روش‌های مختلف ارزیابی نتایج مدل‌ها مانند ماتریس confusion، Precision ، Recall، f1_score استفاده شد و صحت نتایج بررسی شد.

پس از تحقیق روی مدل‌های مختلف ML و کار با آنها از جمله مدل‌های ساده رگرسیونی تا مدل‌های شبکه عصبی، این نتیجه به دست آمد که مدل‌های شبکه عصبی قدرت بیشتری در این زمینه دارند، بنابراین روی مدل‌های شبکه عصبی و بخصوص یادگیری عمیق کار شد و از بین آنها مدل‌های LSTM انتخاب شد. ویژگی این مدل‌ها این است که به کمک آنها می‌توان وابستگی بین سفرها را بهتر درک کرد. یکی از کاربردهای اصلی مدل های LSTM برای پردازش زبان طبیعی (NLP) است که در ترجمه یک جمله از یک زبان به زبان دیگر استفاده می‌شود. با توجه به اینکه در ترجمه زبان وابستگی کلمات به یکدیگر بسیار مهم است، بنابراین مدل‌های LSTM این وابستگی را به خوبی درک می‌کنند. با بررسی سفرهای افراد روی نمونه ۵۰۰۰ تایی از زنان و ۵۰۰۰ تایی از مردان این نتیجه حاصل شد که مسافت پیموده شده توسط مردان خیلی بیشتر از زنان است. میانگین مسافت پیموده شده برای مردان در هرسفر ٦/۷۱ مایل است ولی میانگین مسافت طی شده توسط خانم‌ها ۳/٤٦ مایل می‌باشد. بنابراین زنان سفرهای کوتاه‌تری نسبت به زنان دارند. تعداد سفرها هم برای مردان بیشتر از زنان است. با توجه به اینکه سفرهای مردان از زنان بیشتر است و مسافت بیشتری طی می کنند بنابراین الگوی سفرهای مردان پیچیدگی بیشتری نسبت به زنان دارد و درک این الگو توسط مدل مشکل تر است. به همین دلیل دقت مدل برای تشخیص مردان پایین تر از زنان می‌باشد. با توجه به اینکه در هر طبقه سنی ٥٤۳ عضو در نظر گرفته شد، با بررسی داده‌ها مشخص شد که دربازه سنی ۳۰ تا ۷۰ سال تعداد سفرهای افراد بیشتر است. افراد بین ۳۰-۲۰ سال دارای الگوی سفرهای ساده‌تری هستند.

بسیاری از این افراد هنوز مشغول به تحصیل در دانشگاه هستند یا شاغل نیستند. میانگین سفرهای این افراد ۲/۳ سفر می‌باشد، بنابراین الگوی سفر بسیار ساده‌تری دارند و تشخیص آن برای مدل LSTM به سادگی اتفاق می‌افتد. ولی در سنین بالاتر با توجه به اینکه افراد مشغله‌های کاری فراوان و دارای فرزند و خانواده هستند تعداد سفرهای کاری، خرید، تفریحی و سفرهای مرتبط با فرزندان مانند سفر به مدرسه یا کلاس‌های مختلف برای فرزندان را دارند. بنابراین الگوی سفر پیچیده‌تری دارند.

طبق بررسی صورت گرفته روی افراد بین ۵۰-۴۰ مشخص شد که این افراد میانگین تعداد سفر ٤/٧ در روز دارند و میانگین مسافت پیموده شده به ازای هر سفرتوسط آنها ۱۳/٤ مایل می‌باشد. بنابراین این افراد وضعیت سفرهای پیچیده‌تری داشته‌اند که درک الگوی سفر آنها توسط مدل LSTM مشکل‌تر می‌باشد. درمدل درآمد افراد نیز خروجی نهایی برای طبقه اول و سوم بهتر از طبقه دوم پیش بینی شده است. با بررسی الگوی سفرهای افراد روی داده‌ها مشخص شد که طبقه متوسط درآمدی سفرهای بیشتری دارند و مسافت‌های بیشتری را طی می‌کنند بنابراین الگوی سفر پیچیده‌تری دارند.

۵- مراجع

-Afandizadeh, S., & Bigdeli Rad, H. (2021). Developing a model to determine the number of vehicles lane changing on freeways by Brownian motion method. *Nonlinear Engineering,* 10(1), 450-460.

-Bakhtiari, A., Mirzahossein, H., Kalantari, N., & Jin, X. (2023). Inferring Socioeconomic Characteristics from Travel Patterns. *Journal of Regional and City Planning,* 34(1), 122-136.

-Bing-quan, L. I. U., & Wei, D. U. (2020). Model and method for traffic demand management with parking controlled on bimodal network. *Operations Research and Management Science,* 29(9), 218.

-Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning, New York: *Springer*. Vol. 4, No. 4, 738—739.

-Bulusu, V., Onat, E. B., Sengupta, R., Yedavalli, P., & Macfarlane, J. (2021). A traffic demand analysis method for urban air mobility. *IEEE Transactions on Intelligent Transportation Systems,* 22(9), 6039-6047.

-Carlsson-Kanyama, A., & Linden, A. L. (1999). Travel patterns and environmental effects now and in the future:: implications of differences in energy consumption among socio-economic groups. *Ecological Economics,* 30(3), 405-417.

-Chen, C., Wei, H., Xu, N., Zheng, G., Yang, M., Xiong, Y., & Li, Z. (2020, April). Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. *In Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 04, 3414-3421.

-Chu, T., Wang, J., Codecà, L., & Li, Z. (2019). Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Transactions on Intelligent Transportation Systems,* 21(3), 1086-1095.

-Collins, D., & Tisdell, C. (2002). Gender and differences in travel life cycles. *Journal of Travel Research,* 41(2), 133-143.

-Crane, R., & Takahashi, L. (2009). Sex changes everything: the recent narrowing and widening of travel differences by gender. *Public Works Management & Policy,* 13(4), 328-337.

-Gärling, T., & Schuitema, G. (2007). Travel demand management targeting reduced private car use: effectiveness, public acceptability and political feasibility. *Journal of Social Issues,* 63(1), 139-153.

-Gopalakrishna, D., Hsu, E., Kuhn, B., Schreffler, E., & Diggins, L. (2019). Strengthening Linkages between Transportation Demand Management and Traffic Management, No. FHWA-HOP-18-072. United States. *Federal Highway Administration.*

-Hörcher, D., Singh, R., & Graham, D. J. (2022). Social distancing in public transport: mobilising new technologies for demand management under the Covid-19 crisis. *Transportation,* 49(2), 735-764.

-Jakobsson, C., Fujii, S., & Gärling, T. (2000). Determinants of private car users' acceptance of road pricing. *Transport Policy,* 7(2), 153-158.

-Khatri, S., Vachhani, H., Shah, S., Bhatia, J., Chaturvedi, M., Tanwar, S., & Kumar, N. (2021). Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. *Peer-to-Peer Networking and Applications,* 14, 1778-1805.

-Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine learning applications in activity-travel behaviour research: a review. *Transport Reviews,* 40(3), 288-311.

May, A. D., Kelly, C., & Shepherd, S. (2006). The principles of integration in urban transport strategies. *Transport Policy,* 13(4), 319-327.

-NHTS Data User Guide.

-Sari Aslam, N., Ibrahim, M. R., Cheng, T., Chen, H., & Zhang, Y. (2021). Activity NET, Neural networks to predict public transport trip purposes from individual smart card data and POIs. *Geo-spatial Information Science,* 24(4), 711-721.

-Vuong, N. K., Liu, Y., Chan, S., Lau, C. T., Chen, Z., Wu, M., & Li, X. (2020, July). Deep learning with long short-term memory networks for classification of dementia related travel patterns. *In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (EMBC), IEEE., 5563-5566.

-Xie, G., Qian, Y., & Wang, S. (2021). Forecasting Chinese cruise tourism demand with big data: An optimized machine learning approach. *Tourism Management,* 82, 104208.

-Zargari, S. A., & Rad, H. B. (2023). Development of a gray box system identification model to estimate the parameters affecting traffic accidents. *Nonlinear Engineering,* 12(1), 20220218.

-Zhang, Y., & Cheng, T. (2019). A deep learning approach to infer employment status of passengers by using smart card data. *IEEE Transactions on Intelligent Transportation Systems,* 21(2), 617-629.

-Zhong, G., Yin, T., Zhang, J., He, S., & Ran, B. (2019). Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data. *Transportation,* 46, 1713-1736.

-Zhou, Z., Lin, S., Du, W., & Liang, H. (2019). Integration of regional demand management and signals control for urban traffic networks. *IEEE Access,* 7, 20235-20248.

-Zhu, L., Gonder, J., & Lin, L. (2017). Prediction of individual social-demographic role based on travel behavior variability using long-term GPS data. *Journal of Advanced Transportation*.

Determining the Socio-Economic Characteristics of Passengers in Urban Traffic Using Artificial Intelligence

Shahriar Afandizadeh, *Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.*

Mahmoud Ahmadinejad, Associate Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

Mohammad Ghaedi, M.Sc., Student, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

Hamid Bigdeli Rad, Ph.D., Candidate, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

E-mail: zargari@iust.ac.ir

Received: July 2026- Accepted: November 2026

ABSTRACT

In travel demand forecasting models, socio-economic characteristics of people are used as input to the model. In this study, with the help of artificial intelligence models and especially deep learning models, it has been tried to extract these features from the travel patterns of people. The model that was chosen is the LSTM model. Age, income and gender are the most important inputs of this model. Different age, gender, and income groups may have specific travel behavior preferences, which are tried to be discovered by examining their travel behaviors. One of the important sources for determining people's travel patterns is mobile phone data that shows their daily movements. Due to the unavailability of mobile phone data of passengers, in this research, the NHTS National Household Travel Survey data set has been used. Using the obtained results, with the help of mobile phone data of passengers, their economic and social characteristics are obtained for use in travel demand models. The average distance traveled by men is 6.71 miles per trip, but the average distance traveled by women is 3.46 miles. Therefore, women have shorter trips than men. People between 20-30 years old have simpler travel patterns. Many of these people are still studying at university or are not working. By examining the travel patterns of people on the data, it was found that the middle income class has more trips and travels longer distances, so they have a more complex travel pattern.

Keywords: Travel Demand Model, Socio-Economic Parameters, Deep Learning, LSTM Model