

مبدل چندوجهی برای پیش‌بینی تقاضای خرده‌فروشی بر پایه معناشناسی توصیفات محصول و تاریخچه خرید

فاطمه زارع باقی‌آباد*

گروه مهندسی صنایع، دانشکده فنی مهندسی، دانشگاه اردکان، اردکان، ایران

خلاصه

پیش‌بینی دقیق تقاضای روزانه در سطح مشتری-محصول، همچنان یک مشکل حیاتی اما چالش‌برانگیز در زنجیره تأمین خرده‌فروشی است که به دلیل کمبود داده‌ها، رفتار متغیر مصرف‌کننده و عدم استفاده از اطلاعات ساختار نیافته محصول، با مشکل مواجه است. این مطالعه یک چارچوب جدید مبتنی بر ترنسفورمر معنایی چندوجهی ارائه می‌کند. این چارچوب، جاسازی‌های معنایی محصول را که از پردازش زبان طبیعی توصیفات محصول به‌دست می‌آیند، با تاریخچه خرید ساختاریافته مشتری ادغام می‌کند. همچنین، ویژگی‌های زمانی چندمقیاسی را نیز وارد مدل می‌سازد تا این شکاف پژوهشی پوشش داده شود. پایگاه داده مورد استفاده، خرده‌فروشی آنلاین دو است. نتایج این مدل با روش‌های حافظه بلندمدت-کوتاه، ماشین‌های تقویت گرادیان و یک ترنسفورمر تک‌وجهی مقایسه شد. نتایج نشان داد که چارچوب ترنسفورمر معنایی چندوجهی به‌طور قابل توجهی از همه معیارها بهتر عمل می‌کند و در مقایسه با بهترین حالت پایه، به کاهش ۱۵/۵ درصدی در میانگین مربعات خطا دست یافته است. همچنین یافته‌های تحقیق نشان داد که ادغام معنایی، سیگنال حیاتی برای محصولات کم فروش ارائه داده و تعبیه‌های زمانی با شرطی‌سازی توجه پویا، برای مدل‌سازی فصلی بودن فروش در زنجیره تأمین خرده‌فروشی ضروری هستند. این مطالعه تأیید کرد که ادغام چندوجهی عمیق، رویکردی برای پیش‌بینی تقاضای جزءنگر مشتری-محصول است و یک راه‌حل مقیاس‌پذیر و قابل تفسیر برای مدیریت موجودی سفارشی و زنجیره‌های تأمین خرده‌فروشی مشتری‌محور و انعطاف‌پذیرتر ارائه می‌دهد.

اطلاعات مقاله

تاریخچه مقاله:

دریافت ۱۴۰۴/۰۹/۱۸

بازنگری ۱۴۰۴/۱۲/۱۸

پذیرش ۱۴۰۴/۱۲/۲۰

(مقاله پژوهشی)

کلمات کلیدی:

ترنسفورمر چندوجهی،

پیش‌بینی تقاضا،

ادغام معنایی،

تعبیه‌های پردازش زبان

طبیعی،

زنجیره تأمین خرده‌فروشی،

۱. مقدمه

پیش‌بینی دقیق تقاضا، جزء حیاتی مدیریت مدرن زنجیره تأمین خرده‌فروشی است که مستقیماً بر بهینه‌سازی موجودی، تخصیص منابع و برنامه‌ریزی استراتژیک تأثیر می‌گذارد [۱]. با این حال، مدل‌های مرسوم اغلب با نوسانات ذاتی و پیچیدگی چندبعدی رفتار مصرف‌کننده، به‌ویژه در سطح جزئی مشتری-محصول، دست و پنجه نرم می‌کنند. در حالی که داده‌های سری زمانی تراکنش‌ها یک مبنای تاریخی ارائه می‌دهند، تا حد زیادی اطلاعات معنایی غنی موجود در ویژگی‌ها و توصیفات محصول را

که به‌طور قابل توجهی بر تصمیمات خرید تأثیر می‌گذارند، نادیده می‌گیرند. به‌عنوان مثال، چالش مداوم عدم تطابق تقاضا-عرضه همچنان منجر به ناکارآمدی‌های عملیاتی قابل توجه و ضررهای مالی در سراسر بخش خرده‌فروشی جهانی می‌شود و نیاز مبرم به راه‌حل‌های پیش‌بینی پیشرفته‌تر و یکپارچه‌تر را برجسته می‌کند [۲]. پیشرفت‌های اخیر در یادگیری عمیق، به‌ویژه معماری‌های ترنسفورمر، موفقیت چشمگیری در ثبت وابستگی‌های زمانی پیچیده نشان داده‌اند [۳]. علاوه بر این، عوامل کلان مانند نوسانات قیمت محصول و بحران‌های حوزه فروش نیز بر دقت

* نویسنده مسئول: فاطمه زارع باقی‌آباد

تلفن: ۰۹۱۳۲۵۷۴۶۸۳؛ پست الکترونیکی: f.zare@ardakan.ac.ir

مدلهایی که متن، سری‌های زمانی و تصاویر را برای پیش‌بینی ادغام می‌کنند و پتانسیل اطلاعات معنایی محصول را برای توضیح و پیش‌بینی تقاضای خاص مشتری برجسته می‌کنند، اگرچه این حوزه هنوز کمتر مورد بررسی قرار گرفته است [۱۱].

در زمینه پیش‌بینی تقاضای مبتنی بر داده، توصیفات محصول به داده‌های متنی بدون ساختار همراه اقلام در یک کاتالوگ یا سیستم فروش اشاره دارد که ویژگی‌ها، مشخصات و کاربرد مورد نظر را شرح می‌دهد. در چارچوب‌های تحلیلی پیشرفته، این متون از طریق پردازش زبان طبیعی پردازش می‌شوند تا جاسازی‌های معنایی، نمایش‌های برداری متراکم و کم‌بعدی ایجاد کنند که معنای اساسی، زمینه و ویژگی‌های رابطه‌ای کلمات را دربر می‌گیرند. این تبدیل قابل توجه است زیرا شکاف معنایی بین اطلاعات محصول قابل خواندن توسط انسان و داده‌های قابل تفسیر توسط ماشین را پر می‌کند و مدل‌ها را قادر می‌سازد تا شباهت‌های رابطه‌ای بین اقلام را درک کنند. ابعاد کلیدی این فرآیند شامل انتخاب مدل جاسازی، سطح دانه‌بندی معنایی و مکانیسم ادغام در یک مدل پیش‌بینی پایین‌دستی است. مطالعات قبلی بر ارزش استفاده از چنین داده‌های متنی در چارچوب‌های چندوجهی تأکید می‌کنند. به‌عنوان مثال، تحقیقات در مورد ترکیب ویژگی‌های مکانی برای پیش‌بینی تقاضای کالای تجارت الکترونیک، ارزش هدایت داده‌های چندوجهی را نشان داده است [۱۲]. اخیراً، دیدگاه نظری به‌سمت یادگیری عمیق و ادغام چندوجهی تغییر یافته است، و مطالعات نشان می‌دهد که مدل‌های ادغام مبتنی بر ترنسفورمر می‌توانند با ارائه یک الگوی معنایی، به‌ویژه برای محصولات جدید، دقت پیش‌بینی را به‌طور قابل توجهی افزایش دهند [۱۰]. به‌طور مشابه، تحقیقاتی که با هدف بهبود دقت پیش‌بینی فروش از طریق روش‌های تجزیه تانسور انجام می‌شوند، از سیگنال‌های معنایی بهره می‌برند [۱۳]. این مطالعات در مجموع یک یافته مهم را بیان می‌کنند: گنجاندن اطلاعات معنایی محصول، یک سیگنال قوی و قابل تعمیم ارائه داده که داده‌های کمی تراکنشی را به‌ویژه برای مدل‌سازی محرک‌های انتخاب مصرف‌کننده و تعمیم به اقلام جدید یا به ندرت خریداری شده تکمیل می‌کند. با این حال، ادغام این جاسازی‌ها در مدل‌های پیش‌بینی سری زمانی، همچنان یک مرز نوظهور محسوب می‌شود.

ورودی اساسی برای مدل‌های پیش‌بینی تقاضای مدرن اغلب به‌صورت یک ماتریس سری زمانی چند متغیره، $X = [x_1, x_2, \dots, x_T] \in R^{d \times T}$ فرموله می‌شود که نشان‌دهنده مشاهدات متوالی T است که در آن هر بردار x_t مجموعه‌ای d بعدی از ویژگی‌های مهندسی شده مشتق شده از داده‌های خام تراکنشی است. این نمایش ساختار یافته از آن جهت اهمیت دارد که سوابق خرید جزئی را به قالبی مناسب برای یادگیری ماشین تبدیل می‌کند و وابستگی‌های زمانی، الگوهای تاریخی و محرک‌های زمینه‌ای تقاضا را به صراحت رمزگذاری می‌کند. ابعاد کلیدی x_t با دقت براساس اصول اقتصادسنجی و یادگیری ماشینی تثبیت شده ساخته شده‌اند. متغیرهای تقاضای با تأخیر (به‌عنوان مثال، فروش از روزهای قبل) برای ثبت اثرات خودرگرسیون و تکانه کوتاه مدت در

پیش‌بینی تأثیر می‌گذارند [۴]. با این حال، کاربرد آن‌ها در زنجیره‌تأمین خرده‌فروشی عمدتاً تک‌وجهی بوده است و بر داده‌های تراکنشی ساختاریافته تمرکز دارد و در عین حال پتانسیل پیش‌بینی در داده‌های متنی بدون ساختار را نادیده می‌گیرد [۵]. این تحقیق با ارائه یک چارچوب پیش‌بینی تقاضای بهبود یافته معنایی، این شکاف مهم را پر می‌کند. در این مقاله یک مدل ترنسفورمر چندوجهی جدید معرفی می‌شود که به‌صورت هم‌افزایی، توصیفات محصول پردازش شده از طریق پردازش زبان طبیعی^۱ را برای استخراج جاسازی‌های معنایی با سابقه خرید دقیق مشتری ادغام می‌کند تا به پیش‌بینی جامع‌تر و دقیق‌تری از تقاضای آینده دست یابد. با یکپارچه‌سازی این جریان‌های داده ناهمگن، این مطالعه قصد دارد فراتر از تحلیل سری‌های زمانی سنتی حرکت کند و به حوزه نوظهور هوش مصنوعی چندوجهی در مدیریت عملیات کمک کند و یک راه‌حل مقیاس‌پذیر برای افزایش دقت پیش‌بینی، کاهش هزینه‌های عملیاتی و فعال کردن زنجیره‌های تأمین واقعاً مشتری‌محور ارائه دهد.

پیش‌بینی کل تقاضای روزانه در سطح محصول-مشتری به‌عنوان وظیفه پیش‌بینی مقدار دقیق یک محصول خاص که یک مشتری خاص در یک روز آینده درخواست خواهد کرد، تعریف می‌شود. این معیار جزءنگر، سنگ بنای مدیریت خرده‌فروشی و زنجیره‌تأمین خرده‌فروشی مدرن و مبتنی بر داده است که استراتژی‌های فوق شخصی‌سازی شده را در تکمیل موجودی، قیمت‌گذاری پویا و تبلیغات هدفمند امکان‌پذیر می‌سازد [۶]. اهمیت آن در تأثیر عملیاتی مستقیم آن نهفته است. پیش‌بینی‌های دقیق در این سطح خرد بخش‌بندی می‌تواند به‌طور چشمگیری هزینه‌های کمبود موجودی و موجودی بیش از حد را کاهش دهد و در عین حال رضایت مشتری را از طریق بهبود در دسترس بودن محصول افزایش دهد. با این حال، این وظیفه پیش‌بینی به دلیل ابعاد کلیدی خود، چالش‌های منحصربه‌فردی را به همراه دارد. این چالش‌ها شامل پراکندگی بالا (موارد متعدد تقاضای صفر)، نوسانات شدید ناشی از تصمیم‌گیری فردی و تعاملات پیچیده بین ترجیحات خاص مشتری، ویژگی‌های محصول و عوامل زمانی مانند اثرات روز هفته می‌باشد [۷]. مدل‌های آماری سنتی، مانند چارچوب‌های میانگین متحرک یکپارچه خودهمبسته^۲، اغلب با این ویژگی‌ها، به‌ویژه در محیط‌های خرده‌فروشی در مقیاس بزرگ، دست و پنجه نرم می‌کنند، همان‌طور که در بررسی‌های سیستماتیک این حوزه اشاره شده است [۸]. در پاسخ، ادبیات اخیر به‌سمت رویکردهای یادگیری ماشینی و عمیق گرایش پیدا کرده است. برای مثال، تجزیه و تحلیل پیش‌بینی مبتنی بر گروه، اثربخشی خود را در ثبت روابط غیرخطی از داده‌های تراکنشی و مشتری با ابعاد بالا برای خرده‌فروشی چندکاناله نشان داده است [۹]. همزمان، مدل‌های اصلاح‌شده ترنسفورمر برای مدل‌سازی الگوهای خرید متوالی و وابستگی‌های زمانی با دقت و کارایی بالاتر به کار گرفته شده‌اند [۱۰]. یک دیدگاه نظری غالب، فرض می‌کند که ادغام منابع داده ناهمگن، به‌ویژه داده‌های بدون ساختار مانند توصیفات محصول، برای پیشرفت فراتر از تاریخچه تراکنش بسیار مهم است. مطالعات اخیر، کاوش در مدل‌های ترکیبی چندوجهی را برای این منظور آغاز کرده‌اند، مانند

ماسک پویا برای تزریق سیگنال‌های زمینه‌ای و خاص مشتری همسواست [۱۵]. این مطالعات در مجموع یک یافته مهم را بیان می‌کنند: عملکرد معماری‌های ترنسفورمر در پیش‌بینی سری‌های زمانی نه تنها توسط مکانیسم توجه، بلکه توسط اجزای تخصصی که به صراحت سوگیری‌های القایی را برای زمان و زمینه رمزگذاری می‌کنند، به‌طور قابل توجهی افزایش می‌یابد و شکاف بین قابلیت پردازش توالی عمومی مدل و دانش ساختاریافته خاص حوزه پیش‌بینی را پر می‌کند.

در حالی که تحقیقات موجود، اثربخشی ترنسفورمرها را برای پیش‌بینی سری‌های زمانی اثبات کرده و ارزش داده‌های چندوجهی را برجسته ساخته است، اما همچنان شکاف قابل توجهی در پیش‌بینی تقاضای مشتری محور و مبتنی بر معنا در دقیق‌ترین جزئیات زنجیره تأمین خرده‌فروشی وجود دارد. به‌طور خاص، کارهای قبلی هنوز به‌طور کامل از داده‌های متنی غنی و بدون ساختار توصیفات محصول برای بهبود مدل‌سازی ترجیحات مشتریان در یک معماری یکپارچه و چندوجهی ترنسفورمر که به‌طور خاص برای زوج محصول-مشتری طراحی شده است، استفاده نکرده‌اند. برای پرداختن به این موضوع، مطالعه حاضر سه سوال اصلی تحقیق را مطرح می‌کند: (سوال ۱) چگونه می‌توان جاسازی‌های معنایی از توصیفات محصول را به‌طور مؤثر با داده‌های ساختاریافته تراکنشی و مشتری در یک چارچوب ترنسفورمر ادغام کرد؟ (سوال ۲) آیا این ادغام چندوجهی منجر به بهبود آماری قابل توجه در پیش‌بینی کل تقاضای روزانه در سطح محصول-مشتری زنجیره تأمین خرده‌فروشی در مقایسه با معیارهای تک‌وجهی می‌شود؟ (سوال ۳) چگونه اجزای پیشرفته مانند جاسازی‌های زمانی چندمقیاسی و ماتریس ماسک پویا در این زمینه به عملکرد و تفسیرپذیری مدل کمک می‌کنند؟ در نتیجه، اهداف این مقاله عبارتند از: (۱) ارائه یک چارچوب جدید ترنسفورمر معنایی چندوجهی^۴ که معناشناسی محصول مبتنی بر پردازش زبان طبیعی، تاریخچه خرید مشتری و ویژگی‌های سری زمانی مهندسی شده را ادغام می‌کند؛ (۲) ارزیابی دقیق عملکرد آن در برابر مدل‌های پیشرفته با استفاده از پایگاه داده خرده‌فروشی آنلاین دو^۵؛ و (۳) ارائه یک تحلیل قابل تفسیر از چگونگی هدایت پیش‌بینی‌ها توسط عوامل معنایی و مشتری. برای این منظور، مقاله به شرح زیر ساختار یافته است. مرور ادبیات، کارهای مرتبط در پیش‌بینی تقاضا، یادگیری چندوجهی و کاربردهای ترنسفورمر را ارائه می‌کند و شکاف تحقیقاتی شناسایی شده را برجسته می‌سازد. بخش روش‌شناسی، معماری ترنسفورمر معنایی چندوجهی، از جمله پیش‌پردازش داده‌ها، مهندسی ویژگی و طراحی مکانیسم ادغام را به تفصیل شرح می‌دهد. بخش آزمایش‌ها و نتایج، تنظیمات تجربی، مقایسه‌های معیار و مطالعات حذف را برای پاسخ به سوالات اصلی تحقیق ارائه می‌دهد. در نهایت، بخش بحث و نتیجه‌گیری، یافته‌ها را تفسیر می‌کند، پیامدهای عملی را مورد بحث قرار داده و راه‌هایی را برای تحقیقات آینده پیشنهاد می‌دهد. بخش بعدی این تحقیق با بررسی جامع ادبیات مرتبط آغاز می‌شود.

زنجیره تأمین خرده‌فروشی بسیار مهم هستند. آمارهای متحرک، مانند میانگین‌های متحرک، به هموارسازی نویز و برجسته کردن روندهای اساسی کمک می‌کنند. ویژگی‌های زمانی شامل شاخص‌های روز هفته، ماه و تعطیلات برای مدل‌سازی اثرات فصلی و تقویمی ضروری هستند. در نهایت، متغیرهای خارجی ساختاریافته مانند قیمت یک واحد محصول، شناسه مشتری و کد کشور (پس از کدگذاری مناسب) اطلاعات بیرونی حیاتی را که تغییرات تقاضا را فراتر از الگوهای تاریخی محض توضیح می‌دهند، ادغام می‌کنند. مطالعات قبلی به‌طور مداوم اهمیت این مرحله مهندسی ویژگی را برای ساخت مدل‌های مؤثر، چه مبتنی بر درخت، چه متوالی یا ترکیبی، نشان داده‌اند [۱۴]. دیدگاه نظری زیربنای این عمل این است که تقاضا صرفاً یک سری مستقل نیست، بلکه تابعی از گذشته، چرخه‌های قابل پیش‌بینی و شرایط خارجی قابل اندازه‌گیری است. بنابراین، یک ماتریس X مناسب، پیش‌شرط ضروری برای هر مدل مبتنی بر داده است تا فرآیند تولید تقاضای واقعی را بیاموزد، اصلی که حتی هنگام تطبیق مدل‌ها با زنجیره تأمین خرده‌فروشی جدید نیز صادق است [۱۵].

برای مدل‌سازی مؤثر دینامیک‌های زمانی پیچیده و اثرات زمینه‌ای در پیش‌بینی تقاضای مدرن مبتنی بر ترنسفورمر، دو مؤلفه معماری پیشرفته تعبیه‌های زمانی چندمقیاسی^۱ و ماتریس ماسک پویا^۲ ضروری هستند. تعبیه‌های زمانی چندمقیاسی، نمایش‌های برداری یادگیری‌شده‌ای هستند که زمان را در دانه‌بندی‌های مختلف (مثلاً هفتگی، ماهانه) مستقیماً در فضای ورودی مدل کدگذاری می‌کنند و به آن اجازه می‌دهند تا الگوهای فصلی و دوره‌ای را بدون تکیه بر مراحل پیش‌پردازش صریح و اغلب شکننده مانند تجزیه سری‌های زمانی کلاسیک، به‌طور ذاتی ثبت کند. اهمیت آن‌ها در این است که مدل را قادر می‌سازد تا الگوهایی را که به‌طور همزمان در چرخه‌های مختلف عمل می‌کنند، تفکیک کرده و به آن‌ها توجه کند، که یک بعد کلیدی، طراحی توابع تعبیه برای دوره‌های مختلف است. همزمان، ماتریس ماسک پویا، مکانیسمی است که برای تعدیل وزن‌های خود-توجهی با تزریق مستقیم تأثیر متغیرهای خارجی خاص و متغیر با زمان مانند قیمت در زمان واقعی (P_t)، مشخصات مشتری (C_i) و منطقه جغرافیایی (G_j) طراحی شده است. این ماتریس ماسک پویا به‌صورت $M = f(P_t, C_i, G_j)$ تعریف می‌شود، که در آن f یک تابع پارامتری است. اهمیت ماسک پویا در توانایی آن در مشروط کردن تمرکز مدل بر داده‌های تاریخی براساس متغیرهای خارجی فعلی است که مکانیسم توجه را آگاه از زمینه و قابل تفسیرتر می‌کند. یک دیدگاه نظری کلیدی این است که خودتوجهی محض، اگرچه برای ثبت وابستگی‌ها قدرتمند است، اما در مورد محرک‌های خارجی بی تفاوت است. ماسک پویا روشی اصولی برای معرفی این مشروط‌سازی ارائه می‌دهد. مطالعات قبلی، کاربرد هر دو مفهوم را در معماری‌های پیش‌بینی پیشرفته تأیید کرده‌اند. ادغام همجوشی^۳ زمانی چندوجهی، همان‌طور که در ترنسفورمرهای اخیر نشان داده شده است، به رمزگذاری مؤثر زمان متکی است [۵]. علاوه بر این، نیاز به گنجاندن چارچوب‌های رفتاری و محیطی در پیش‌بینی با هدف

4. Multimodal Semantic Transformer (MST)

5. Online Retail II

1. Multi-scale temporal embeddings

2. Dynamic mask matrix

3. Fusion

۲. مرور ادبیات

دستیابی به پیش‌بینی دقیق تقاضا، چالشی مداوم و پیچیده در عملیات زنجیره‌تأمین خرده‌فروشی است که با مقیاس، سرعت و ابعاد اکوسیستم‌های تجارت الکترونیک معاصر، تشدید می‌شود. مدل‌های آماری سنتی، مانند میانگین‌های متحرک، اگرچه چارچوبی بنیادی ارائه می‌دهند، اما به‌طور فزاینده‌ای برای ثبت ابعاد بالا، غیرخطی بودن و نوسانات ذاتی که مشخصه جریان‌های داده خرده‌فروشی مدرن هستند، ناکافی شناخته می‌شوند [۸]. بررسی‌های جامع در این زمینه، تغییر الگوی واضحی را به سمت روش‌های مبتنی بر یادگیری ماشین و مبتنی بر داده نشان می‌دهد که نشان‌دهنده یک حرکت قاطع از تکنیک‌های صرفاً خودرگرسیون به سمت مدل‌هایی است که قادر به یادگیری الگوهای پیچیده و غیرخطی به‌طور مستقیم از مجموعه ویژگی‌های متنوع و غنی هستند [۷]. این تکامل ناشی از ضرورت حاد عملیاتی و مالی برای مدیریت کمبود موجودی و موجودی بیش از حد است، دو روی یک سکه پرهزینه که با پیش‌بینی‌های نادرست و در سطح کل تشدید می‌شوند [۲].

در پاسخ به این چالش‌ها، معماری‌های یادگیری عمیق به خط مقدم تحقیقات پیش‌بینی صعود کرده‌اند. در این حوزه، مدل‌های ترنسفورمر که براساس مکانیسم خود-توجهی ساخته شده‌اند، اخیراً یک فناوری جدید و پیشرفته برای وظایف سری زمانی ایجاد کرده‌اند. توانایی آن‌ها در پردازش کل توالی‌ها به‌صورت موازی و مدل‌سازی وابستگی‌ها بدون محدودیت‌های متوالی شبکه‌های بازگشتی مانند حافظه‌های بلندمدت-کوتاه، یک چارچوب نظری برتر را فراهم می‌کند. مطالعات تجربی به‌طور خاص در زمینه‌های خرده‌فروشی، مانند کار الیورا و راموس [۳]، اثربخشی ترنسفورمرهای سری زمانی را تأیید کرده و معیارهای عملکرد جدیدی را تعیین کرده‌اند. این موفقیت، نوآوری‌های بیشتری را تسریع کرده و منجر به مدل‌های ترکیبی پیچیده‌ای شده است که نقاط قوت ترنسفورمر را با معماری‌های دیگر ترکیب می‌کنند تا دقت، کارایی و مقیاس‌پذیری را افزایش دهند، همان‌طور که توسط مدل ترکیبی شبکه پیچشی زمانی^۱ و ترنسفورمر ارائه شده [۱۴]، نشان داده شده است.

به موازات پیشرفت‌های معماری، هدف استراتژیک پیش‌بینی از تخمین‌های کلی و در سطح محصول به بینش‌های جزئی و مشتری‌محور تکامل یافته است. تحقیقات در مورد پیش‌بینی ارزش طول عمر مشتری مانند سیستم با عملکرد بالا که توسط یان و رسنیک [۶] توسعه داده شده است، بر ارزش اقتصادی عظیم درک و پیش‌بینی رفتار فردی مشتری تأکید می‌کند. این امر با روند صنعتی گسترده‌تر به سمت شخصی‌سازی بیش از حد همسو است، جایی که مهندسی تجربیات مشتری متناسب با نیاز مشتری به‌عنوان منبع اصلی مزیت رقابتی شناخته می‌شود [۱۶]. در نتیجه، پیش‌بینی در پیوند با یک محصول خاص و یک مشتری خاص، پیش‌بینی نه تنها آنچه که فروخته خواهد شد، بلکه به چه کسی و چه زمانی فروخته خواهد شد، به یک قابلیت حیاتی برای فعال کردن مدیریت موجودی شخصی، قیمت‌گذاری پویا و تبلیغات هدفمند تبدیل شده است [۱۵]. با اذعان به اینکه سیگنال پیش‌بینی در انواع داده‌های متنوع توزیع

شده است، این حوزه قاطعانه به سمت ادغام چندوجهی حرکت کرده است. ادغام داده‌های ناهمگن شامل متن، سری‌های زمانی، تصاویر بصری و فراداده‌های متنی به‌طور گسترده به‌عنوان یک استراتژی قدرتمند برای ساخت مدل‌های پیش‌بینی قوی‌تر و قابل تعمیم‌تر شناخته می‌شود. یک کار پیشگامانه در این زمینه، ارزش هدایت ادغام ویژگی‌های مکانی با داده‌های چندوجهی را برای پیش‌بینی کالاهای تجارت الکترونیک نشان داد [۱۶]. این الگو به سرعت در حال تکامل است و چارچوب‌های پیشرفته‌ی معاصر به صراحت به‌عنوان سیستم‌های چندوجهی معماری شده‌اند. به‌عنوان مثال، تحقیق سوکل و وورینگ، اعتبارسنجی تجربی مستقیمی را برای این رویکرد اصلی ارائه می‌دهد [۵].

پیشرفت‌های تحقیقات چندوجهی فراتر از تلفیق ساده ویژگی‌ها به سمت یکپارچه‌سازی هوشمند و آگاه از زمینه در حال پیشرفت است. مدل‌های پیشرفته در حال ترکیب اصول استدلال علی برای تشخیص همبستگی‌های صرف از محرک‌های علی واقعی تقاضا هستند. وانگ [۱۱] این مسیر را با یک ترنسفورمر چندوجهی آگاه از علت نشان می‌دهد که متن، سری‌های زمانی و تصاویر ماهواره‌ای را برای پیش‌بینی زنجیره‌تأمین ادغام می‌کند و هدف آن جداسازی سیگنال‌های تقاضای واقعی است. به‌طور مشابه، روش‌های پیچیده گروهی و رویکردهای تجزیه تانسور برای افزایش آگاهی از تقاضا و بهبود دقت پیش‌بینی در طیف وسیعی از محصولات، همان‌طور که در کار [۹] و [۱۳] مشاهده می‌شود، در حال اصلاح هستند.

گذار از عملکرد مدل نظری به استقرار عملی و در مقیاس بزرگ، چالش‌های مهمی را ایجاد می‌کند. از جمله این چالش‌ها، نیاز به مقیاس‌بندی کارآمد پیش‌بینی‌ها در میلیون‌ها واحد نگهداری موجودی و تطبیق مدل‌ها با حوزه‌های خرده‌فروشی جدید با توزیع داده‌های متفاوت است. علاوه بر این، مدل‌ها باید در برابر تغییرات ناگهانی و ساختاری بازار مانند تغییرات ناشی از بیماری‌های همه‌گیر جهانی، راه‌اندازی قالب‌های جدید فروشگاه یا معرفی دسته‌بندی‌های جدید محصول که می‌توانند الگوهای تقاضا را اساساً تغییر داده و عملکرد مدل‌های آموزش‌دیده براساس داده‌های قبل از تغییر را کاهش دهند، مقاومت خود را حفظ کنند [۱۷]. تحقیقات انجام شده توسط [۱۵] مستقیماً از طریق تکنیک‌های نوآورانه تطبیق دامنه برای پیش‌بینی تقاضای خرده‌فروشی به مورد اخیر می‌پردازد. در همین حال، طراحی معماری‌های ذاتاً مقیاس‌پذیر همچنان یک محور اصلی تحقیق است، تمرکزی که به وضوح در توسعه مدل‌های ترکیبی کارآمد مشهود است [۱۴].

با وجود پیشرفت‌های قابل‌توجه در حوزه ادغام چندوجهی، بررسی انتقادی ادبیات پژوهش نشان می‌دهد که همچنان توجه کافی به بهره‌گیری از محتوای معنایی عمیق متن محصول در پیش‌بینی تقاضای خرده‌فروشی معطوف نشده است. به‌تازگی، مطالعه [۱۸] گامی در این مسیر برداشته و چارچوبی مبتنی بر ترنسفورمر را برای لحاظ کردن هم‌زمان معناشناسی محصول و اثرات شبکه‌ای ارائه کرده است. با این حال، رویکرد مذکور دارای سه محدودیت اساسی است که مقاله حاضر با هدف

استخراج کرده و آن‌ها را با تاریخچه خرید ساختاریافته مشتریان تلفیق می‌کند. همچنین، دو مؤلفه‌ی معماری نوین یعنی تعبیه‌های زمانی چندمقیاسی و ماتریس ماسک پویا برای نخستین بار در حوزه پیش‌بینی تقاضای خرده‌فروشی داخلی به‌کار گرفته شده‌اند.

سوم، در سطح کاربردی، این پژوهش با ارائه مدلی دقیق و قابل تفسیر برای پیش‌بینی تقاضا در سطح مشتری-محصول، ابزاری کارآمد برای شخصی‌سازی مدیریت موجودی، قیمت‌گذاری پویا و کاهش هزینه‌های عملیاتی در اختیار مدیران زنجیره‌تأمین قرار می‌دهد که تا پیش از این در پژوهش‌های داخلی مشابهی ارائه نشده بود.

۳. روش

این مطالعه از یک طرح تحقیقاتی تجربی که در یک محیط محاسباتی کنترل‌شده انجام شده است، برای توسعه و اعتبارسنجی یک چارچوب یادگیری عمیق چندوجهی جدید استفاده می‌کند. این روش‌شناسی از یک رویکرد کمی و داده‌محور رایج در مهندسی یادگیری ماشین پیروی می‌کند که بر توسعه مدل، آموزش و ارزیابی مقایسه‌ای متمرکز است. طراحی تحقیق توسط اصول نظری مدل‌سازی توالی مبتنی بر ترنسفورمر و ادغام داده‌های چندوجهی هدایت می‌شود. رویکرد کلی شامل پیش‌پردازش سیستماتیک داده‌ها، مهندسی ویژگی، طراحی معماری ترنسفورمر معنایی چندوجهی پیشنهادی (MST) و معیارسنجی دقیق عملکرد در برابر مدل‌های پیش‌بینی تثبیت‌شده است. تجزیه و تحلیل داده‌ها در درجه اول شامل مقایسه آماری معیارهای دقت پیش‌بینی و مطالعات حذف برای جداسازی سهم اجزای کلیدی مدل است.

این مطالعه از پایگاه داده عمومی خرده‌فروشی آنلاین دو، به‌عنوان نمونه اولیه خود استفاده می‌کند که شامل داده‌های تراکنشی از یک خرده‌فروش آنلاین مستقر در بریتانیا است. این مجموعه داده‌ها شامل ۵۲۵،۴۶۱ رکورد تراکنش انفرادی از دسامبر ۲۰۱۰ تا دسامبر ۲۰۱۱ است که نشان دهنده جمعیت کامل سفارشات لغو نشده در این دوره است. معیارهای ورود برای فیلتر کردن داده‌های خام اعمال شدند و فقط تراکنش‌هایی با کدهای کالای معتبر، شناسه‌های مشتری و مقادیر مثبت برای اطمینان از یکپارچگی تحلیلی حفظ شدند. این نمونه مجموعه‌ای غنی از ویژگی‌های حیاتی برای مطالعه، از جمله توصیفات محصول، قیمت واحد، تاریخ خرید، شناسه‌های مشتری و اطلاعات کشور مشتری را در زنجیره‌تأمین خرده‌فروشی ارائه می‌دهد که امکان ساخت سری‌های زمانی جزئی محصول-مشتری را فراهم می‌کند. این مجموعه داده‌ها به دلیل واقع‌گرایی، مقیاس و دسترسی عمومی آن انتخاب شده است و معیاری قوی برای ارزیابی مدل‌های پیش‌بینی تقاضای مشتری محور ارائه می‌دهد. این تحقیق کاملاً در یک محیط محاسباتی و با استفاده از کتابخانه‌ها و چارچوب‌های نرم‌افزاری تخصصی انجام شد. توسعه و مدل‌سازی اصلی با استفاده از پایتون ۳.۹+ با چارچوب یادگیری عمیق PyTorch انجام شد که به دلیل انعطاف‌پذیری آن در پیاده‌سازی معماری‌های سفارشی ترنسفورمر انتخاب شده بود. اصلاح و پیش‌پردازش داده‌ها به Pandas و NumPy متکی بود، در حالی که از scikit-learn برای مقیاس‌بندی

پوشش آن‌ها طراحی شده است: نخست، مدل یاد شده عمدتاً در سطح کلان محصول عمل می‌کند و توانایی پیش‌بینی در سطح خرد و شخصی‌سازی‌شده‌ی مشتری-محصول را ندارد؛ در حالی که پژوهش حاضر برای اولین بار به این سطح از جزئی‌نگری در ادبیات داخلی وارد شده است. دوم، رویکرد پیشین، متن محصول را صرفاً به‌عنوان یک سیگنال مکمل و کمکی در کنار سایر ویژگی‌ها در نظر گرفته، در حالی که مقاله حاضر با معماری چندوجهی عمیق خود، جاسازی‌های معنایی را به‌عنوان یک مؤلفه‌ی اصلی و قابل تفسیر در هسته مدل قرار داده و آن را با تاریخچه خرید فردی مشتری تلفیق می‌کند تا وابستگی‌های ظریف میان ترجیحات مشتری و ویژگی‌های محصول را به‌صورت صریح مدل‌سازی نماید. سوم، در حالی که مطالعه [۱۸] صرفاً بر داده‌های ساختاریافته و متکی بر روابط شبکه‌ای میان محصولات است، مقاله حاضر با استخراج جاسازی‌های معنایی از توصیفات متنی محصولات با استفاده از مدل‌های پیش‌آموزش‌دیده پردازش زبان طبیعی و ترکیب آن با تعبیه‌های زمانی چندمقیاسی و ماتریس ماسک پویا، گامی فراتر نهاده و مدلی ارائه می‌دهد که علاوه بر دقت بالاتر، از تفسیرپذیری بیشتری نیز برخوردار است.

با ترکیب مسیرهای ذکر شده در بالا، یک شکاف تحقیقاتی متمایز و مهم پدیدار می‌شود. در حالی که ادبیات موضوع به‌طور قطعی (۱) قابلیت برتر معماری‌های ترنسفورمر برای مدل‌سازی توالی، (۲) ضرورت ادغام چندوجهی برای پیش‌بینی‌های قوی، و (۳) ضرورت استراتژیک پیش‌بینی دقیق و مشتری‌محور را اثبات می‌کند، هیچ چارچوب موجودی به صراحت برای ادغام عمیق و قابل تفسیر تعبیه‌های معنایی محصول با تاریخچه خرید مشتری در یک معماری ترنسفورمر یکپارچه و کارآمد برای وظیفه خاص پیش‌بینی تقاضای روزانه محصول-مشتری طراحی نشده است. مقالات قبلی تمایل دارند بر سطوح کلی تمرکز کنند، متن را به‌صورت سطحی در نظر بگیرند، یا به چالش‌های منحصر به فرد کمبود داده‌ها و نوسانات در سطح دوتایی محصول-مشتری نپردازند. این دقیقاً همان شکافی است که مطالعه حاضر با ارائه چارچوب ترنسفورمر معنایی چندوجهی قصد دارد آن را پر کند و از این طریق مستقیماً به پیشرفت‌ها در هوش مصنوعی چندوجهی برای عملیات، مدل‌سازی تقاضای قابل تفسیر و تجزیه و تحلیل خرده‌فروشی شخصی‌سازی‌شده کمک کند.

براساس شکاف‌های شناسایی شده در ادبیات پژوهش، نوآوری‌های مقاله حاضر را می‌توان در سه سطح طبقه‌بندی کرد:

نخست، در سطح نظری، این پژوهش برای اولین بار در ادبیات داخلی، چارچوبی یکپارچه برای پیش‌بینی تقاضا در سطح خرد و جزءنگر مشتری-محصول ارائه می‌دهد که در آن، سه مؤلفه کلیدی یعنی وابستگی‌های زمانی با تأخیر، اثرات شبکه‌ای میان محصولات و معناشناسی عمیق داده‌های متنی به‌طور هم‌زمان مدل‌سازی می‌شوند. این رویکرد، خلأ مدل‌های تک‌بعدی پیشین را که عمدتاً در سطح کلان محصول عمل می‌کردند، پوشش می‌دهد.

دوم، در سطح روش، توسعه یک معماری ترنسفورمر چندوجهی عمیق که با بهره‌گیری از مدل‌های پیش‌آموزش‌دیده پردازش زبان طبیعی (BERT)، جاسازی‌های معنایی را از توصیفات متنی محصولات

داده و مدل را در برابر خطاهای بزرگ گاه به گاه مقاوم کرده و آموزش را تثبیت می‌کند. ابرپارامتر $\alpha = 0.7$ سهم هر زبان را متعادل می‌کند و MSE را کمی ترجیح داده تا دقت کلی را در اولویت قرار دهد و در عین حال استحکام را حفظ کند.

منظم‌سازی $\|\theta\|_2^2$ ، λ یک منظم‌سازی L2 (کاهش وزن) است که بر روی تمام پارامترهای مدل θ اعمال می‌شود. این تابع مقادیر پارامترهای بزرگ را جریمه می‌کند تا از بیش‌برازش جلوگیری کند و مدل را به یادگیری الگوهای ساده‌تر و تعمیم‌پذیرتر تشویق کند. قدرت منظم‌سازی $\lambda = 0.001$ تأثیر این جریمه را در طول بهینه‌سازی کنترل می‌کند.

- y_i : تقاضای واقعی روزانه برای زوج مشتری-محصول i
- $\hat{y}_i$: تقاضای پیش‌بینی شده از مدل MST
- متغیرهای تصمیم، شامل

$\theta = \{W_Q, W_K, W_V, W_O, W_{\text{feat}}, W_{\text{text}}, E_{\text{daily}}, E_{\text{monthly}}, E_{\text{pos}}, \gamma, \beta\}$
می‌شود که در آن:

- $W_Q, W_K, W_V \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$: ماتریس‌های پرس‌وجو، کلید، مقدار برای خود-توجه
- $W_O \in \mathbb{R}^{d_{\text{model}} \times d_{\text{model}}}$: ماتریس پیش‌نگاشت خروجی
- $W_{\text{feat}} \in \mathbb{R}^{d_{\text{num}} \times d_{\text{model}}}$: پیش‌نگاشت ویژگی عددی
- $W_{\text{text}} \in \mathbb{R}^{d_{\text{text}} \times d_{\text{model}}}$: پیش‌نگاشت جاسازی متن
- $E_{\text{daily}} \in \mathbb{R}^{7 \times \frac{d_{\text{model}}}{3}}$: جاسازی‌های زمانی روزانه (هفتگی)
- $E_{\text{monthly}} \in \mathbb{R}^{12 \times \frac{d_{\text{model}}}{3}}$: جاسازی‌های زمانی ماهانه
- $E_{\text{pos}} \in \mathbb{R}^{L_{\text{max}} \times \frac{d_{\text{model}}}{3}}$: جاسازی‌های موقعیتی
- γ, β : پارامترهای نرمال‌سازی لایه‌ای هستند.

ورودی‌ها شامل ویژگی‌های عددی و جاسازی‌های متنی هستند. ویژگی عددی در رابطه (۲) نشان داده شده است.

$$X_{\text{num}} = [x_1, x_2, \dots, x_T] \in \mathbb{R}^{T \times d_{\text{num}}} \quad (2)$$

که در آن x_t شامل تقاضای تأخیردار $\{q_{t-1}, q_{t-2}, q_{t-3}, q_{t-7}\}$ و آماره‌های متحرک $\{\bar{q}_{7,t}, \sigma_{7,t}\}$ و ویژگی‌های زمانی $\{d_t, m_t, w_t\}$ و متغیرهای خارجی $\{p_t, c_t, g_t\}$ می‌باشد.

جاسازی‌های متنی در رابطه (۳) نشان داده شده است.

$$X_{\text{text}} = \text{BERT}(\text{Description}) \in \mathbb{R}^{T \times d_{\text{text}}} \quad (3)$$

این مؤلفه، تبدیل متن توصیفات محصول بدون ساختار را به جاسازی‌های معنایی متنی متراکم با استفاده از یک مدل بازنمایی دوسویه‌گذار از ترنسفورمر^۱ از پیش آموزش‌دیده نشان می‌دهد و توالی از گام‌های زمانی T را تولید می‌کند که هر کدام توسط یک بردار d_{text} بعدی نمایش داده می‌شوند.

جاسازی‌های زمانی چندمقیاسی در رابطه (۴) نشان داده شده است.

$$E_t = \text{Concat}[E_{\text{daily}}(d_t), E_{\text{monthly}}(m_t), E_{\text{pos}}(t)] \in \mathbb{R}^{d_{\text{model}}} \quad (4)$$

استاندارد و معیارسنجی مدل پایه استفاده شد. پردازش زبان طبیعی برای توصیفات محصول با استفاده از کتابخانه Transformers توسط Hugging Face اجرا شد، به‌طور خاص از یک مدل BERT از پیش آموزش دیده برای تولید جاسازی‌های معنایی استفاده شد.

جمع‌آوری داده‌ها شامل تهیه پایگاه داده عمومی خرده‌فروشی آنلاین دو از مخزن یادگیری ماشین UCI بود که استفاده از یک معیار استاندارد و غیر اختصاصی را تضمین می‌کرد. روش تجزیه و تحلیل از یک خط‌مشی ساختار یافته پیروی می‌کرد: ابتدا، داده‌های تراکنشی خام پاکسازی شدند (با مدیریت مقادیر گمشده، فیلتر کردن بازگشتی‌ها) و برای ایجاد توالی‌های تقاضای روزانه محصول-مشتری در زنجیره تأمین خرده‌فروشی تجمیع شدند. سپس مهندسی ویژگی، تأخیرهای زمانی، آماره متحرک و متغیرهای دسته‌بندی کدگذاری‌شده را ایجاد کرد. روش تحلیلی اصلی شامل آموزش مدل پیشنهادی ترنسفورمر معنایی چندوجهی (MST) و مدل‌های معیار (مانند LSTM، ماشین‌های تقویت گردایان) بر روی یک تقسیم زمانی آموزش-آزمون، با استفاده از میانگین مربعات خطا و میانگین خطای مطلق به‌عنوان معیارهای ارزیابی اولیه بود. عملکرد مدل از نظر آماری مقایسه شد و مطالعات حذف داده‌ها برای ارزیابی سهم جاسازی‌های معنایی و اجزای معماری انجام شد.

به منظور ارزیابی قدرت تعمیم‌پذیری مدل، داده‌های پیش‌پردازش شده به دو مجموعه آموزش (۸۰ درصد معادل ۶۴,۴۴۱ نمونه) و آزمون (۲۰ درصد معادل ۱۶,۱۱۱ نمونه) تقسیم شدند. معیارهای اصلی ارزیابی شامل میانگین مربعات خطا (MSE) و میانگین قدرمطلق خطا (MAE) بودند که به ترتیب در روابط (۱۶) و (۱۷) تعریف شده‌اند. برای اطمینان از اعتبار نتایج، عملکرد مدل پیشنهادی (MST) در مجموعه آزمون با سه مدل مرجع شامل ترنسفورمر تک‌وجهی (به‌عنوان قوی‌ترین مدل پایه)، شبکه‌های حافظه بلندمدت-کوتاه (LSTM) و ماشین‌های تقویت گردایان مقایسه گردید. همچنین، جهت بررسی سهم هر یک از مؤلفه‌های کلیدی مدل، مطالعه حذف با حذف سیستماتیک هر یک از سه مؤلفه اصلی (جاسازی‌های متنی، تعبیه‌های زمانی چندمقیاسی و ماتریس ماسک پویا) انجام شد و عملکرد هر حالت با مدل کامل مقایسه گردید.

۳-۱. مدل ریاضی چارچوب تبدیل معنایی چندوجهی (MST) پیش‌بینی تقاضای زنجیره تأمین خرده‌فروشی

تابع هدف:

تابع هدف شامل به حداقل رساندن خطای پیش‌بینی ترکیبی و منظم‌سازی برای جلوگیری از بیش‌برازش است:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N [\alpha \cdot \text{MSE}(y_i, \hat{y}_i) + (1 - \alpha) \cdot \text{MAE}(y_i, \hat{y}_i)] + \lambda \|\theta\|_2^2 \quad (1)$$

خطای پیش‌بینی ترکیبی، دو تابع زیان رگرسیون را ترکیب می‌کند:

۱- میانگین مربعات خطا (MSE) که خطاهای بزرگتر را به‌صورت درجه دوم جریمه می‌کند و بر پیش‌بینی دقیق مقادیر تقاضای شدید یا پرت تأکید می‌کند. ۲- میانگین مطلق خطا (MAE) که یک جریمه خطی ارائه

1. Bidirectional Encoder Representations from Transformers (BERT)

وزن‌های توجه را براساس متغیرهای خارجی تعدیل می‌کند. توجه چندوجهی با اعمال ماسک پویا از طریق رابطه (۱۰) محاسبه می‌شود:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + M_t\right)V \quad (10)$$

با $K = H^{l-1}W_K^l$ و $V = H^{l-1}W_V^l$ ، $Q = H^{l-1}W_Q^l$ مکانیزم توجه، مجموع وزن‌داری از بردارهای مقدار V را محاسبه می‌کند، که در آن وزن‌ها از سازگاری بین پرس‌وجو Q و کلید K پس از مقیاس‌بندی مشتق می‌شوند و مدل را قادر می‌سازد تا براساس پیش‌بینی‌های آموخته‌شده از توالی ورودی، روی مراحل زمانی مربوطه تمرکز کند.

پیش‌بینی خروجی:

$$\hat{y}_t = \text{ReLU}(H_t^L W_1 + b_1) W_2 + b_2 \quad (11)$$

که در آن $H_t^L \in \mathbb{R}^{2d_{\text{model}}}$ نمایش آخرین گام زمانی است. تقاضای پیش‌بینی شده نهایی $\hat{y}_t$ با اعمال یک شبکه پیش‌خور دو لایه با فعال‌سازی ReLU به نمایش چندوجهی H_t^L از آخرین لایه ترنسفورمر محاسبه می‌شود و آن را برای هر گام زمانی به یک خروجی اسکالر تبدیل می‌کند. ReLU یک تابع فعال‌سازی غیرخطی است که فقط مقادیر مثبت را عبور می‌دهد و مقادیر منفی را صفر می‌کند.

محدودیت‌ها و منظم‌سازی شامل محدودیت‌های پارامتری، برش گرادیان، محدودیت سازگاری زمانی و محدودیت غیرمنفی می‌باشد که به ترتیب در رابطه‌های (۱۲) الی (۱۵) نشان داده شده است.

$$\|\Theta\|_2 \leq C \quad \text{where } C = 10.0 \quad (12)$$

$$\|\nabla_{\Theta} \mathcal{L}\|_2 \leq 1.0 \quad \text{during training} \quad (13)$$

$$|\hat{y}_t - \hat{y}_{t-1}| \leq \delta \cdot \sigma_y \quad \text{where } \delta = 2.0 \quad (14)$$

$$\hat{y}_t \geq 0, \quad \forall t \quad (15)$$

محدودیت پارامتر، تضمین می‌کند که نرم L2 تمام پارامترهای مدل زیر ۱۰/۰ باقی بماند تا از وزن‌های بیش از حد بزرگ جلوگیری شود. برش گرادیان، نرم گرادیان را در طول آموزش به ۱/۰ محدود می‌کند تا فرآیند بهینه‌سازی را تثبیت و تسریع کند. محدودیت سازگاری زمانی، نوسانات روزانه در پیش‌بینی‌ها را محدود می‌کند تا در محدوده دو برابر انحراف معیار تقاضای هدف باشد و پیش‌بینی‌های روان را ارتقا دهد. محدودیت غیرمنفی بودن، تضمین می‌کند که تمام مقادیر تقاضای پیش‌بینی‌شده غیرمنفی باشند و منعکس‌کننده ماهیت فیزیکی واقع‌بینانه تقاضا باشند.

پارامترهای معماری:

- $d_{\text{model}} = 64$: بعد مدل
- $d_{\text{num}} = 16$: بعد ویژگی عددی
- $d_{\text{text}} = 50$: بعد جاسازی متن
- $L = 2$: تعداد لایه‌های تبدیل‌کننده
- $h = 4$: تعداد سرهای توجه
- $T = 20$: طول توالی ورودی

پارامترهای یادگیری:

- نرخ یادگیری: $\eta = 0.001$
- اندازه دسته: $B = 16$
- نرخ خروج از سیستم: $\rho = 0.2$
- کاهش وزن: $\lambda = 0.001$

که در آن، $d_t \in \{0, \dots, 6\}$ روز هفته و $m_t \in \{0, \dots, 11\}$ ماه است. این معادله با ترکیب سه بردار جاسازی آموخته‌شده، یک جاسازی زمانی چندمقیاسی می‌سازد. E_{daily} جاسازی زمانی روزانه، E_{monthly} جاسازی زمانی ماهانه و E_{pos} موقعیت مطلق را در دنباله ورودی کدگذاری می‌کند. بردار حاصل، الگوهای زمانی چرخه‌ای را با اطلاعات ترتیبی متوالی ترکیب می‌کند و بایاس‌های استقرایی زمانی صریح را برای مدل فراهم می‌کند تا وابستگی‌های فصلی و متوالی را بهتر ثبت کند.

ویژگی‌های پیش‌بینی شده در دو رابطه (۵) و (۶) نشان داده شده است:

$$H_{\text{num}} = X_{\text{num}} W_{\text{feat}} + E_t \left[:, \frac{d_{\text{model}}}{3} \right] \in \mathbb{R}^{T \times d_{\text{model}}} \quad (5)$$

$$H_{\text{text}} = X_{\text{text}} W_{\text{text}} + E_t \left[:, \frac{d_{\text{model}}}{3} : \frac{2d_{\text{model}}}{3} \right] \in \mathbb{R}^{T \times d_{\text{model}}} \quad (6)$$

برای ویژگی عددی H_{num} ، این معادله ماتریس ویژگی عددی X_{num} را با استفاده از ماتریس وزن W_{feat} به فضای جاسازی مدل تصویر می‌کند، سپس یک سوم اول جاسازی زمانی چندمقیاسی E_t را اضافه می‌کند تا زمینه زمانی (الگوهای روزانه و هفتگی) را مستقیماً به نمایش ورودی عددی تزریق کند.

برای ویژگی متنی H_{text} ، این معادله ماتریس جاسازی متن X_{text} را از طریق W_{text} به فضای جاسازی مدل تصویر می‌کند، سپس یک سوم دوم جاسازی زمانی E_t را اضافه می‌کند تا معانی متنی را با سیگنال‌های زمانی ماهانه و موقعیتی همسو کند و مدل را قادر سازد تا توصیفات محصول را با زمینه‌های فصلی یا متوالی خاص مرتبط کند. همجوشی چندوجهی در رابطه (۷) نشان داده شده است.

$$H = \text{Concat}[H_{\text{num}}, H_{\text{text}}] \in \mathbb{R}^{T \times 2d_{\text{model}}} \quad (7)$$

این مرحله، جاسازی‌های عددی و متنی بهبودیافته زمانی را در امتداد بعد ویژگی به هم پیوند می‌دهد تا یک نمایش چندوجهی یکپارچه تشکیل دهد که به‌طور مشترک اطلاعات تاریخی، معنایی و زمانی را برای هر گام زمانی رمزگذاری می‌کند. لایه‌های رمزگذاری ترنسفورمر برای هر لایه $l = 1, \dots, L$ در رابطه (۸) نشان داده شده است.

$$H^l = \text{LayerNorm}(\text{Multi Head Attention}(H^{l-1}) + H^{l-1}) \quad (8)$$

$$H^l = \text{LayerNorm}(\text{FFN}(H^l) + H^l)$$

معادله اول، خودتوجه چند سره و به دنبال آن اتصال باقیمانده و نرمال‌سازی را اعمال می‌کند و به مدل اجازه می‌دهد تا گام‌های زمانی مختلف را براساس ارتباط آن‌ها وزن‌دهی کند. معادله دوم، یک شبکه پیش‌خور با یک اتصال باقیمانده و نرمال‌سازی را اعمال می‌کند که امکان تبدیل غیرخطی نمایش‌های مورد نظر را برای ثبت تعاملات پیچیده ویژگی‌ها فراهم می‌کند. به منظور شرطی‌سازی مکانیزم خودتوجهی براساس متغیرهای خارجی پویا، ماتریس ماسک پویا از طریق رابطه (۹) تعریف شده است.

$$M_t = \text{MLP}([P_t, C_i, G_j]) \in \mathbb{R}^{T \times T} \quad (9)$$

که در آن، P_t قیمت لحظه‌ای محصول، C_i ویژگی‌های مشتری (شامل شناسه، تاریخچه خرید و دسته‌بندی)، G_j کد موقعیت جغرافیایی، و MLP یک شبکه‌ی عصبی پیش‌خور با یک لایه پنهان است. این ماتریس،

جفت مشتری-محصول شناسایی شده، ۲۷۴,۱۸۷ جفت (معادل ۹۹/۹۱ درصد) به دلیل عدم برخورداری از الگوی خرید منظم و کافی، حذف و تنها ۲۵۵ جفت (معادل ۰/۰۹ درصد) که دارای تکرارپذیری زمانی معناداری بودند، برای تحلیل نهایی حفظ گردیدند. این رویکرد که با استانداردهای پژوهش‌های معتبر بین‌المللی در حوزه پیش‌بینی تقاضا همسو است، به‌جای کاهش کیفیت، با متمرکز شدن بر الگوهای قابل اعتماد تقاضا، بستر مناسبی برای آموزش مدل عمیق فراهم می‌آورد [۱۹]. به منظور تشکیل توالی‌های زمانی پیوسته با طول یکسان، کلیه روزهای بدون تراکنش در بازه زمانی بین اولین و آخرین تراکنش هر جفت مشتری-محصول، با مقدار تقاضای صفر پر شدند. این رویکرد استاندارد در مدل‌سازی سری‌های زمانی با داده‌های پراکنده است (مانند [۱۹]) و امکان تشکیل دنباله‌های پیوسته با طول یکسان را فراهم می‌آورد. سپس با استفاده از روش پنجره‌های لغزنده با طول توالی ورودی $T=20$ و گام زمانی ۱، نمونه‌های اولیه استخراج شدند. بدین ترتیب که برای هر جفت مشتری-محصول دارای M روز داده (شامل روزهای صفر و غیرصفر)، تعداد نمونه‌های قابل استخراج برابر با $(M - T + 1)$ خواهد بود. براساس محاسبات، میانگین تعداد روزهای دارای داده برای هر یک از ۲۵۵ جفت واجد شرایط، حدوداً ۳۳۶ روز بوده است؛ بنابراین هر جفت به‌طور میانگین $336 - 20 + 1 = 317$ نمونه خام تولید کرده است. با ضرب این عدد در ۲۵۵ جفت، به عدد ۸۰,۸۳۵ نمونه می‌رسد که با اعمال مهندسی ویژگی‌های پیشرفته (شامل متغیرهای تأخیری، میانگین‌های متحرک و ترکیب جاسازی‌های متنی) و حذف نمونه‌های لبه‌ای که داده‌ی کافی برای ساخت متغیرهای تأخیری نداشتند، ۸۰,۵۵۲ نمونه نهایی به‌دست آمد.



شکل (۱). نمودار جریان مراحل پیش‌پردازش، فیلترسازی و آماده‌سازی داده‌های تراکنشی برای مدل پیش‌بینی تقاضای MST

ماتریس‌های ارزیابی شامل تابع هدف اصلی، فرعی و ضریب تعیین می‌باشد که به ترتیب در رابطه‌های (۱۶)، (۱۷) و (۱۸) نشان داده شده است.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (17)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (18)$$

در انتخاب مدل، پارامترهایی انتخاب می‌شوند که MSE اعتبارسنجی را به حداقل برسانند و در عین حال، موارد زیر را حفظ کرده تا از بیش‌برازش جلوگیری شود:

$$\frac{\text{Validation MSE}}{\text{Training MSE}} \leq 1.3 \quad (19)$$

۴. نتایج

ارزیابی تجربی چارچوب پیشنهادی ترنسفورمر معنایی چندوجهی (MST) بر روی پایگاه داده خرده‌فروشی آنلاین دو، اثربخشی قابل توجه آن را برای پیش‌بینی تقاضای جزئی مشتری-محصول نشان می‌دهد. این مدل با موفقیت همگرا شد و به میانگین مربعات خطا برابر با ۰/۱۴۷ و میانگین مطلق خطا برابر ۰/۲۰۱ در آزمون نهایی دست یافت که نشان دهنده درجه بالایی از دقت پیش‌بینی با توجه به پراکندگی و نوسانات ذاتی پیش‌بینی است. این نتایج نشان‌دهنده بهبود قابل توجه نسبت به مدل‌های پایه تک‌وجهی است که برای ثبت الگوهای پیچیده و غیرخطی در داده‌ها تلاش می‌کردند. نتیجه کلی، فرضیه اصلی را تأیید می‌کند که یک معماری یکپارچه که قادر به ادغام اطلاعات معنایی محصول با داده‌های تراکنشی و زمانی ساختاریافته باشد، می‌تواند دقت پیش‌بینی را برای پیش‌بینی تقاضای خرده‌فروشی فوق‌العاده شخصی‌سازی شده به‌طور قابل توجهی افزایش دهد.

به منظور تضمین کیفیت داده‌ها و آماده‌سازی بستر مناسب برای آموزش مدل ترنسفورمر معنایی چندوجهی (MST)، یک مسیر چندمرحله‌ای برای پیش‌پردازش داده‌ها طراحی و اجرا شد (شکل (۱)). روند آماده‌سازی با کلان‌داده تراکنشی شامل ۵۲۵,۴۶۱ رکورد اولیه از پایگاه داده خرده‌فروشی آنلاین دو آغاز گردید. در گام نخست پاکسازی، تراکنش‌های نامعتبر شامل مقادیر منفی قیمت و مقادیر مفقود مربوط به شناسه مشتریان حذف شدند که این امر حجم داده‌ها را به ۴۰۷,۶۶۴ تراکنش تأییدشده کاهش داد. سپس، با تجمیع روزانه تراکنش‌ها، تعداد رکورد سری زمانی در سطح مشتری-محصول حاصل شد. از ۳۹۰,۳۷۰ میان کل، ۲۷۴,۴۴۲ جفت منحصر به فرد مشتری-محصول شناسایی شد. از آنجا که داده‌های خرده‌فروشی آنلاین ذاتاً با پراکندگی شدید مواجه هستند و اکثریت جفت‌های مشتری-محصول دارای کمتر از ۵ تراکنش در کل دوره یک‌ساله می‌باشند، برای آموزش مؤثر معماری ترنسفورمر که به توالی‌های زمانی با طول حداقل ۲۰ روزه نیاز دارد، آستانه حداقل ۱۵ مشاهده فعال به‌عنوان شرط ورود در نظر گرفته شد. از مجموع ۲۷۴,۴۴۲

جدول (۱). عملکرد مقایسه‌ای مدل‌های پیش‌بینی

مدل	MSE	MAE	R ²	بهبود در مقابل LSTM (%MSE)	بهبود در مقایسه با ترنسفورم تک‌وجهی %MSE	درصد کاهش MSE نسبت به مدل‌های پایه
گرادیان بوستینگ	۰/۲۱	۰/۲۷۵	۰/۴۸۵	۰/۶۳۱	-	۳۰/۱٪
LSTM	۰/۱۸۹	۰/۲۴۷	۰/۴۳۵	۰/۶۶۷	پایه	۲۲/۲٪
ترنسفورم تک‌وجهی	۰/۱۷۴	۰/۲۱۸	۰/۴۱۷	۰/۶۹۴	۷/۹٪	۱۵/۵٪ (بهبود اصلی مقاله)

نشان‌دهنده‌ی نقش حیاتی این مؤلفه در ثبت صریح الگوهای فصلی هفتگی و ماهانه است. در نهایت، حذف ماتریس ماسک پویا با افزایش خطا به ۰/۱۶۱ (کاهش ۹/۵ درصدی عملکرد)، کم‌ترین اما همچنان قابل توجه‌ترین تأثیر را داشته که نشان‌دهنده نقش آن در شرطی‌سازی توجه مدل براساس متغیرهای خارجی مانند قیمت و ویژگی‌های مشتری می‌باشد. علاوه بر این، تجزیه و تحلیل وزن‌های توجه آموخته‌شده، اجرای موفقیت‌آمیز مدل از اصل پوشش پویا را نشان داد. این مدل به‌صورت پویا تمرکز خود را بر روی نقاط داده تاریخی براساس مقادیر همزمان متغیرهای خارجی مانند قیمت، بخش مشتری و دسته محصول تنظیم می‌کند. به‌عنوان مثال، در دوره‌های با قیمت بالا، مدل توجه بیشتری به روندهای تقاضای اخیر نشان داد، در حالی‌که برای اقلام تبلیغاتی، دوره‌های تاریخی با تخفیف‌های قیمتی مشابه را در اولویت قرار داد. این مکانیسم توجه شرطی، یک دیدگاه علی-مانند و شفاف به فرآیند تصمیم‌گیری مدل ارائه می‌دهد و تأیید می‌کند که چارچوب MST نه تنها با دقت بیشتری پیش‌بینی می‌کند، بلکه این کار را به شیوه‌ای آگاه از زمینه و قابل تفسیر انجام می‌دهد و به‌طور مؤثر شکاف بین پردازش توالی خالص و داده‌های پیشین ساختاریافته مختص حوزه پیش‌بینی خرده‌فروشی را پر می‌کند.

جدول (۲). نتایج مطالعه حذف برای بررسی سهم هر یک از

مؤلفه‌های مدل

مؤلفه حذف شده	MSE	افزایش خطا نسبت به مدل کامل	کاهش عملکرد (درصد)
مدل کامل (MST)	۰/۱۴۷	-	-
بدون جاسازی‌های متنی	۰/۱۸۰	+۰/۰۳۳	-۱۸/۳٪
بدون تعبیه‌های زمانی چندمقیاسی	۰/۱۶۵	+۰/۰۱۸	-۱۲/۲٪
بدون ماتریس ماسک پویا	۰/۱۶۱	+۰/۰۱۴	-۹/۵٪

۵. بحث

نتایج این مطالعه شواهد قانع‌کننده‌ای ارائه می‌دهد که ادغام معنایی عمیق توصیفات محصول در یک معماری ترنسفورم چندوجهی، پیشرفتی متحول‌کننده برای پیش‌بینی تقاضای مشتری‌محور فراهم می‌کند. افزایش قابل توجه عملکرد نسبت به خطوط پایه تک‌وجهی، فرضیه اصلی ما را تأیید می‌کند که داده‌های متنی بدون ساختار، سیگنال‌های حیاتی و نهفته‌ای در مورد ترجیح مصرف‌کننده و وابستگی محصول دارند که مدل‌های سری زمانی عددی سنتی نمی‌توانند به آن‌ها دسترسی داشته

این کسر ناچیز (حدود ۰/۳۵ درصد) مربوط به جفت‌هایی است که در ابتدا یا انتهای دوره‌ی فعالیت خود، داده‌ی کافی برای ساخت متغیرهای تأخیری و میانگین‌های متحرک نداشته‌اند. در نهایت، نمونه‌ها با رعایت توالی زمانی به دو بخش آموزش (۶۴,۴۴۱ نمونه، معادل ۸۰ درصد) و آزمون (۱۶,۱۱۱ نمونه، معادل ۲۰ درصد) تقسیم شدند.

در مورد اولین سوال تحقیق، نتایج تأیید می‌کنند که جاسازی‌های معنایی از توصیفات محصول را می‌توان به‌طور مؤثر با داده‌های ساختاریافته در چارچوب ترنسفورم ترکیب کرد. مطالعه حذف داده‌ها نشان داد که گنجاندن جاسازی‌های متنی، خطای پیش‌بینی را در مقایسه با مدلی که فقط از ویژگی‌های عددی استفاده می‌کند، تقریباً ۱۸/۳ درصد کاهش می‌دهد (میانگین مربعات خطا ۰/۱۸۰ در مقابل ۰/۱۴۷). مکانیسم توجه مدل با موفقیت یاد گرفت که وزن‌های بالاتری را به کلمات کلیدی توصیفی مرتبط با فصلی بودن (مانند کریسمس، چراغ‌ها) و ویژگی‌های محصول در دوره‌های زمانی مربوطه اختصاص دهد و مکانیسم ادغام را تأیید کند. این ادغام معنایی به‌ویژه برای محصولاتی با سابقه خرید پراکنده، که در آن اطلاعات متنی سیگنال مهمی برای تعمیم ارائه می‌داد، تأثیرگذار بود.

در پاسخ به سوال دوم تحقیق، ادغام چندوجهی، بهبود آماری معنی‌داری در پیش‌بینی تقاضای روزانه محصول-مشتری ایجاد کرد. همان‌طور که در جدول (۱) نشان داده شده است، مدل MST پیشنهادی از همه مدل‌های معیار بهتر عمل کرد. آزمون t زوجی تأیید کرد که کاهش MSE از نظر آماری معنی‌دار است ($p < 0.01$). به‌طور خاص، مدل MST در مقایسه با قوی‌ترین مدل پایه (ترنسفورم تک‌وجهی) که تنها از داده‌های تراکنشی استفاده می‌کند، خطا را ۱۵/۵ درصد کاهش داد. همچنین این مدل به‌ترتیب بهبود ۲۲/۲ درصدی نسبت به LSTM و ۳۰/۱ درصدی نسبت به ماشین تقویت گرادیان را نشان می‌دهد. این افزایش عملکرد، ضرورت ادغام اطلاعات معنایی و استفاده از مؤلفه‌های پیشرفته معماری را در پیش‌بینی تقاضای خرده‌فروشی برجسته می‌سازد.

در پاسخ به سوال سوم تحقیق، اجزای معماری پیشرفته شامل تعبیه‌های زمانی چندمقیاسی، جاسازی‌های معنایی و ماتریس ماسک پویا به‌عنوان عوامل حیاتی برای عملکرد و تفسیرپذیری مدل شناسایی شدند. نتایج مطالعه حذف در جدول (۲) نشان داد که حذف جاسازی‌های متنی، خطای مدل را از ۰/۱۴۷ به ۰/۱۸۰ افزایش می‌دهد که معادل کاهش ۱۸/۳ درصدی عملکرد است. این امر گویای نقش کلیدی سیگنال‌های معنایی استخراج‌شده از توصیفات محصول، به‌ویژه برای محصولات کم‌فروش، می‌باشد. همچنین حذف تعبیه‌های زمانی چندمقیاسی، خطا را از ۰/۱۴۷ به ۰/۱۶۵ افزایش داده است (کاهش ۱۲/۲ درصدی عملکرد) که

محصول با سابقه خرید مشتری و ویژگی‌های زمانی مهندسی‌شده، دقت پیش‌بینی تقاضای روزانه را در سطح جزئی محصول-مشتری به‌طور قابل توجهی افزایش می‌دهد. نتایج تجربی، پاسخ‌های مثبت و روشنی به سوالات اصلی تحقیق ارائه می‌دهند: اول، جاسازی‌های معنایی از توصیفات محصول می‌توانند به‌طور مؤثر در معماری ترنسفورمر از طریق مکانیسم‌های تخصصی پیش‌بینی و توجه ادغام شوند؛ دوم، این ادغام چندوجهی، بهبود آماری قابل توجهی نسبت به معیارهای تک‌وجهی ایجاد می‌کند، همان‌طور که با کاهش ۱۵/۵ درصدی در میانگین مربعات خطا نشان داده شده است؛ و سوم، اجزای پیشرفته‌ای مانند جاسازی‌های زمانی چندمقیاسی و شرطی‌سازی زمینه‌ای پویا برای عملکرد و تفسیرپذیری مدل بسیار مهم هستند و امکان ثبت روند فصلی پیچیده و زمینه شخصی‌سازی شده را فراهم می‌کنند. در مجموع، این یافته‌ها فرضیه اصلی پژوهش را تأیید می‌کنند؛ یعنی یک مدل یکپارچه و آگاه از معنا می‌تواند محدودیت‌های روش‌های سنتی را در مواجهه با پراکندگی داده‌ها و نوسانات شدید رفتار مصرف‌کننده فردی کاهش دهد.

سهم دانش‌افزایی این پژوهش را می‌توان در سه سطح خلاصه کرد: (۱) ارائه اولین چارچوب پیش‌بینی تقاضا در سطح مشتری-محصول در ادبیات داخلی که به‌طور هم‌زمان اثرات شبکه‌ای، تأخیرهای زمانی و معنانشناسی محصول را مدل‌سازی می‌کند؛ (۲) توسعه معماری ترنسفورمر چندوجهی با مؤلفه‌های نوین تعبیه‌های زمانی چندمقیاسی و ماسک پویا که قابلیت ثبت الگوهای فصلی و شرطی‌سازی توجه براساس متغیرهای خارجی را فراهم می‌آورد؛ و (۳) ارائه راهکاری عملی برای شخصی‌سازی مدیریت زنجیره‌ای تأمین که به‌طور مستقیم به کاهش هزینه‌های موجودی و افزایش کارایی عملیاتی می‌انجامد.

ارزش‌های گسترده‌تر این تحقیق هم برای تحقیقات دانشگاهی و هم برای عملکرد صنعت خرده‌فروشی قابل توجه است. برای مدیریت زنجیره‌تأمین، چارچوب MST ابزاری عملی برای دستیابی به عملیات مشتری‌محور واقعی ارائه می‌دهد که امکان برنامه‌ریزی موجودی بیش از حد شخصی‌سازی‌شده، قیمت‌گذاری پویا و تبلیغات هدفمند را فراهم می‌کند که می‌تواند ضررهای مالی عظیم مرتبط با عدم تطابق تقاضا-عرضه را کاهش دهد. این چارچوب با نشان دادن مسیری مناسب برای ادغام جریان‌های داده ناهمگن در یک سیستم واحد و قابل یادگیری سرتاسری، به حوزه هوش مصنوعی عملیاتی کمک می‌کند.

تحقیقات آینده می‌توانند بر افزایش مقیاس‌پذیری، تعمیم‌پذیری و پایداری چارچوب پیشنهادی در محیط‌های واقعی خرده‌فروشی تمرکز کنند. یکی از مسیرهای مهم، بررسی روش‌های یادگیری ائتلافی برای آموزش مدل بر داده‌های غیرمتمرکز و حساس مشتریان است؛ رویکردی که می‌تواند بدون نیاز به تجمع مستقیم داده‌ها، امکان بهره‌گیری از اطلاعات چند فروشگاه یا چند سازمان را فراهم کند. مسیر دیگر، به‌کارگیری استنتاج علی برای تفکیک اثر تبلیغات، تخفیف‌ها و مداخلات بازاریابی از تقاضای واقعی مشتریان است تا پیش‌بینی‌ها تنها بر همبستگی‌های مشاهده‌شده متکی نباشند. همچنین، ارزیابی قابلیت انتقال چارچوب پیشنهادی در بخش‌های مختلف خرده‌فروشی، مانند کالاهای

باشند. موفقیت مکانیسم ادغام معنایی، به‌ویژه برای محصولات پراکنده یا جدید، با نشان دادن اینکه جاسازی‌های زبان طبیعی می‌توانند به‌طور مؤثر با توالی‌های تراکنشی ساختاریافته همسو شوند تا یک نمایش منسجم و با ابعاد بالا از رابطه مشتری و محصول تشکیل دهند، سوال پژوهشی ۱ را پاسخ می‌دهد. علاوه بر این، توانایی نشان داده شده این مدل در مشروط کردن توجه زمانی خود به متغیرهای خارجی پویا (سوال پژوهش ۳) دقت برتر آن را توضیح می‌دهد. این مدل صرفاً تاریخچه را پردازش نمی‌کند، بلکه به‌طور هوشمندانه آن را براساس عوامل زمینه‌ای مانند تغییرات قیمت و هویت مشتری وزن‌دهی مجدد می‌کند و فراتر از تشخیص الگو به استدلال آگاه از زمینه می‌رود.

این یافته‌ها هم با ادبیات فعلی در مورد پیش‌بینی خرده‌فروشی همسو هستند و هم آن را گسترش می‌دهند. آن‌ها قویاً از الگوی نوظهوری که توسط محققانی مانند سوکل و وورینگ [۵] ارائه شده است، مبنی بر اینکه ادغام چندوجهی برای پیش‌بینی‌کنندگان نسل بعدی ضروری است، پشتیبانی می‌کنند. تحقیق حاضر با تحقیق لی [۱۰] در مورد ارزش الگوهای معنایی برای محصولات جدید همسو است، اما با تعبیه این قابلیت در یک چارچوب ترنسفورمر مقیاس‌پذیر و سرتاسری برای پیش‌بینی مداوم و جزئی به جای پیش‌بینی‌های معرفی یک‌باره، آن را گسترش می‌دهد. نتایج همچنین با تحقیقاتی که از جزئی‌نگری در سطح مشتری [۶] حمایت می‌کنند، همخوانی دارد و ابزار روش‌شناختی برای دستیابی به آن را ارائه می‌دهد. با این حال، رویکرد حاضر از مطالعاتی که انطباق دامنه را به‌عنوان یک فرآیند جداگانه و پس از وقوع در نظر می‌گیرند، متفاوت است [۱۷]. در عوض، شرطی‌سازی پویای چارچوب MST مسیری را به سمت سازگاری ذاتی با تغییرات بازار ارائه می‌دهد، که نشان می‌دهد استحکام را می‌توان در معماری اصلی طراحی کرد، نه اینکه به‌عنوان یک اصلاح خارجی اضافه نمود.

با وجود نتایج برتر مدل، این مطالعه با چند محدودیت همراه است. نخست، مدل تنها بر یک مجموعه داده تاریخی از یک حوزه مشخص خرده‌فروشی ارزیابی شده است؛ بنابراین، تعمیم‌پذیری یافته‌ها به بخش‌هایی با پویایی تقاضای متفاوت، مانند کالاهای تندمصرف یا اقلام لوکس، نیازمند احتیاط است. دوم، تولید جاسازی‌های BERT برای کاتالوگ‌های بزرگ محصولات هزینه محاسباتی قابل توجهی دارد و می‌تواند استقرار مدل را در مقیاس عملیاتی دشوار کند. سوم، ارزیابی مدل در این مطالعه عمدتاً بر داده‌های تاریخی انجام شده و پایداری آن در برابر شوک‌های ساختاری بازار، مانند بحران‌های اقتصادی یا اختلالات زنجیره‌تأمین، به‌طور مستقیم آزمون نشده است. همچنین، دامنه داده‌های بدون ساختار مورد استفاده به توصیفات متنی محصول محدود بوده و سایر منابع بالقوه اثرگذار بر تقاضا، مانند احساسات شبکه‌های اجتماعی یا تصاویر تبلیغاتی، در مدل لحاظ نشده‌اند.

۶. نتیجه‌گیری

این مطالعه با موفقیت چارچوب ترنسفورمر معنایی چندوجهی (MST) را توسعه و اعتبارسنجی کرد و نشان داد که ادغام جاسازی‌های معنایی

- [7] J. Rahikka and P. Mikkola, (2025). Modern time series methods for demand forecasting in retail.
- [8] A. R. Chowdhury, R. Paul, and F. Z. Rozony, (2025). A Systematic Review of Demand Forecasting Models for Retail E-Commerce Enhancing Accuracy in Inventory and Delivery Planning, *International Journal of Scientific and Interdisciplinary Research*, vol. 6, no. 1, 1–27.
- [9] T. Samal and A. Ghosh, (2025). Ensemble-based predictive analytics for demand forecasting in multi-channel retailing, *Expert Systems with Applications*, vol. 2025, 130212.
- [10] Q. Li, (2023). Achieving Sales Forecasting with Higher Accuracy and Efficiency: A New Model Based on Modified Transformer, *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 4, 1990–2006.
- [11] Y. Wang, (2025). Causal-Aware Multimodal Transformer for Supply Chain Demand Forecasting: Integrating Text, Time Series, and Satellite Imagery, *IEEE Access*, vol. 13, 176813–176829.
- [12] W. Cai, Y. Song, and Z. Wei, (2021). Multimodal Data Guided Spatial Feature Fusion and Grouping Strategy for E-Commerce Commodity Demand Forecasting, *Mobile Information Systems*, vol. 2021, 5568208.
- [13] X. Bi, G. Adomavicius, W. Li, and A. Qu, (2022). Improving Sales Forecasting Accuracy: A Tensor Factorization Approach with Demand Awareness, *INFORMS Journal on Computing*, vol. 34, no. 3, 1644–1660.
- [14] R. M. A. et al., (2025). A Hybrid Temporal Convolutional Network and Transformer Model for Accurate and Scalable, *IEEE Open Journal of the Computer Society*, vol. 6, 380–391.
- [15] S. Tripathi, R. Trigunait, and D. Chandra, (2025). A Behavioral and Environmental Framework for Sustainable Retail Forecasting and Decision-Making, *Circular Economy and Sustainability*, 1–29.
- [16] S. K. Rai, (2025). Data-Driven Retail: The Engineering Behind Personalized Customer Experiences, *Journal of Computer Science and Technology Studies*, vol. 7, no. 10, 571–581.
- [17] N. Tarighat et al., (2025). Domain Adaptation for Retail Demand Prediction, *IEEE Access*, vol. 13, 146267–146294.
- [۱۸] ف. زارع باقی‌آباد، (۱۴۰۵). پیش‌بینی پویای تقاضای خرده‌فروشی مبتنی بر رویکرد ترنسفورمر با تلفیق اثرات شبکه‌ای، تأخیرهای زمانی و معناشناسی محصول. مطالعات مدیریت صنعتی، دوره ۲۴، شماره ۸۱، ۱–۳۶.
- [19] A. Dolgui, M. Pashkevich, (2008). Demand forecasting for multiple slow-moving items with short requests history and unequal demand variance. *International Journal of Production Economics*, vol. 112, no. 2, 885–894.

زودگذردش، اقلام لوکس و محصولات فصلی، می‌تواند شواهد دقیق‌تری درباره تمم‌پذیری مدل فراهم سازد. افزون بر این، آزمون مدل در زمینه‌های فرهنگی و بازارهای متفاوت می‌تواند نشان دهد که الگوهای معنایی استخراج‌شده از توصیفات محصول تا چه اندازه در محیط‌های گوناگون پایدار باقی می‌مانند. از منظر توسعه مدل، استفاده از مدل‌های زبانی کارآمدتر، مدل‌های مختص حوزه، یا جاسازی‌های مبتنی بر نمودار دانش می‌تواند سربار محاسباتی تولید بازنمایی‌های معنایی را کاهش دهد. به‌کارگیری نظام‌مند روش‌های انطباق دامنه نیز می‌تواند برای سنجش و بهبود پایداری مدل در برابر شوک‌های ساختاری بازار، مانند بحران‌های اقتصادی یا اختلالات زنجیره‌تأمین، مورد توجه قرار گیرد. در نهایت، ادغام منابع داده بدون ساختار دیگر، از جمله احساسات شبکه‌های اجتماعی، تصاویر تبلیغاتی و محتوای تولیدشده توسط کاربران، می‌تواند شناخت مدل از محرک‌های تقاضا را غنی‌تر کند. چنین توسعه‌هایی مسیر را برای طراحی زنجیره‌های تأمین خرده‌فروشی هوشمندتر، سازگارتر و تاب‌آورتر هموار می‌سازد؛ زنجیره‌هایی که توانایی پاسخ‌گویی مؤثر به بازارهای ناپایدار، رقابتی و به‌شدت شخصی‌سازی‌شده را دارند.

تقدیر و تشکر

در این بخش از داوران محترم که با صرف وقت و دقت، موجب ارتقای کیفیت مقاله شده‌اند، کمال تشکر و سپاسگزاری می‌شود.

مراجع

- [1] R. Tiwari, (2025). Harnessing AI and Predictive Analytics for Enhanced Demand Forecasting in Retail Supply Chains, *Supply Chain & Retail Management*, vol. 8, no. 1, 41–52.
- [2] L. Saarinen and P. Huttunen, (2025). Revisiting the value of data sharing in retail supply chain demand planning, *International Journal of Operations & Production Management*, vol. 45, no. 11, 1910–1936.
- [3] J. Oliveira and P. Ramos, (2024). Evaluating the effectiveness of time series transformers for demand forecasting in retail, *Mathematics*, vol. 12, no. 17, 2728.
- [4] M. Eskandari, M. Darabi, and H. Asghari, (2025). Identification and Assessment of Risks in the Sales Barriers to Reduce Food Supply Chain Crises, *Journal of Supply Chain Management*, vol. 27, no. 1, 18–29.
- [5] M. Sukel and M. Worrington, (2024). Multimodal Temporal Fusion Transformers Are Good Product Demand Forecasters, *IEEE Multimedia*, vol. 31, no. 2, 48–60.
- [6] Y. Yan and N. Resnick, (2024). A high-performance turnkey system for customer lifetime value prediction in retail brands, *Quantitative Marketing and Economics*, vol. 22, no. 2, 169–192.



DOI: <https://doi.org/10.22084/ier.2026.32880.2238>

A Multimodal Transformer for Retail Demand Forecasting Based on the Semantics of Product Descriptions and Purchase History

Fatemeh Zare Baghiabad*

Department of Industrial Engineering, Faculty of Engineering, Ardakan University, P.O. Box 184, Ardakan, Iran

ARTICLE INFO

Article history:

Received 2025/12/08

Revised 2026/03/09

Accepted 2026/03/11

Keywords:

Multimodal Transformer,
Retail Demand Forecasting,
Semantic Integration,
Natural Language Processing
(NLP),
Deep Learning,
Inventory Management.

ABSTRACT

Accurate daily demand forecasting at the customer-product level remains a critical yet challenging task in retail supply chain management. This challenge is mainly driven by sparse transaction records, highly volatile consumer behavior, and the limited use of unstructured product information in forecasting models. To address these issues, this study proposes a novel multimodal semantic Transformer framework that integrates semantic product embeddings, extracted from product descriptions through natural language processing, with structured customer purchase history. In addition, the model incorporates multi-scale temporal features to capture seasonality, short-term demand fluctuations, and longer-term purchasing patterns more effectively. The proposed framework is evaluated on a real-world online retail dataset and compared with several benchmark models, including Long Short-Term Memory (LSTM) networks, Gradient Boosting Machines, and a unimodal Transformer architecture. The empirical results show that the proposed model consistently outperforms all baseline methods and achieves a 15.5% reduction in Mean Squared Error (MSE). The findings further reveal that semantic product information provides valuable predictive signals, particularly for low-velocity products where historical demand patterns alone are often insufficient. Moreover, temporal embeddings combined with dynamic attention conditioning improve the model's ability to represent seasonal variations and evolving customer preferences over time. Overall, the results confirm that deep multimodal integration offers an effective, scalable, and interpretable solution for fine-grained retail demand forecasting and can support more customized inventory decisions and more resilient, customer-centric retail supply chains.

* Corresponding author. F. Zare Baghiabad
Tel.: 09132574683; E-mail address: f.zare@ardakan.ac.ir