دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر
گروه کنترل

طراحی سیستم طبقه بند مجتمع پویا به منظور تصمیم سازی هوشمند درآنالیز کسب و کارها

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق
گرایش کنترل

علی حافظی

استاد راهنما

دکتر بهزاد مشیری

شهریور ۱۴۰۱

بسم الله الرحمن الرحيم

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر
گروه کنترل

طراحی سیستم طبقه بند مجتمع پویا به منظور تصمیم سازی هوشمند درآنالیز کسب و کارها

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق
گرایش کنترل

علی حافظی

استاد راهنما

دکتر بهزاد مشیری

شهریور ۱۴۰۱

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده مهندسی برق و کامپیوتر

گواهی دفاع از پایان‌نامه کارشناسی ارشد

هیأت داوران پایان‌نامه کارشناسی ارشد آقای / خانم علی حافظی به شماره دانشجویی ۸۱۰۱۹۷۲۱۳ در رشته مهندسی برق - گرایش کنترل را در تاریخ ..... با عنوان « طراحی سیستم طبقه بند مجتمع پویا به منظور تصمیم سازی هوشمند درآنالیز کسب و کارها »

	به عدد	به حروف
با نمره نهایی		

و درجه [ ] ارزیابی کرد.

ردیف	مشخصات هیئت داوران	نام و نام خانوادگی	مرتبه دانشگاهی	دانشگاه یا مؤسسه	امضـــــــــــا
۱	استاد راهنما	دکتر بهزاد مشیری	استاد	دانشگاه تهران	
۲	استاد داور داخلی	دکتر احمد کلهر	دانشیار	دانشگاه تهران	
۳	استاد مدعو	دکتر امین رمضانی	دانشیار	دانشگاه تربیت مدرس	
۴	نماینده تحصیلات تکمیلی دانشکده	دکتر احمد کلهر	دانشیار	دانشگاه تهران	

نام و نام خانوادگی معاون تحصیلات تکمیلی و پژوهشی دانشکده / گروه:

تاریخ و امضاء:

نام و نام خانوادگی معاون آموزشی و تحصیلات تکمیلی پردیس دانشکده‌های فنی:

تاریخ و امضاء:

تعهدنامه اصالت اثر

باسمه تعالی

اینجانب علی حافظی تائید می‌کنم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نشده است.

نام و نام خانوادگی دانشجو: علی حافظی

تاریخ و امضای دانشجو:

کلیه حقوق مادی و معنوی این اثر
متعلّق به دانشگاه تهران می‌باشد.

تقدیم به:

پدر و مادرم

قدردانی

سپاس خداوندگار حکیم را که با لطف بی‌کران خود، آدمی را به زیور عقل آراست.

در آغاز وظیفه خود می‌دانم از زحمات بی‌دریغ اساتید راهنمای خود، جناب آقای دکتر بهزاد مشیری صمیمانه تشکر و قدردانی کنم که در طول انجام این پایان‌نامه با نهایت صبوری همواره راهنما و مشوق من بودند و قطعاً بدون راهنمایی‌های ارزنده ایشان، این مجموعه به انجام نمی‌رسید.

از جناب آقای دکتر احمد کلهر که مطالب ارزنده‌ای را در طول تحصیل از ایشان آموختم قدردانی و تشکر می‌کنم، چرا که آموزه‌های ایشان کمک شایانی به بنده در انجام این پایان‌نامه کرد.

از جناب آقای دکتر امین رمضانی که زحمت داوری این رساله را به عهده داشتند سپاس فراوان دارم.

و در پایان، بوسه می‌زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می‌کنم وجود مقدس‌شان را و تشکر می‌کنم از خانواده عزیزم به پاس عاطفه سرشار و گرمای امیدبخش وجودشان، که بهترین پشتیبان من بودند.

علی حافظی

شهریور ۱۴۰۱

چکیده

همانطور که مفهومی به نام بهترین مدل وجود ندارد، مفهوم بهترین نوع مدلسازی نیز بصورت عمومی قابل تعریف نیست. بنابراین در این پژوهش، بسته به نوع مجموعه‌داده در دسترس، یک سیستم طبقه‌بند مجتمع پویا مبتنی بر خوشه‌بندی و فاصله اقلیدسی نمونه‌ها تا مرکز هر خوشه، طراحی شده است. این مدل به منظور پیش‌بینی رخداد عدم موجودی کالا به‌کار می‌رود که برای کل فضای ویژگی داده‌ها معتبر است. رویکردهای سنّتی و ابزارهای فیزیکی که تا کنون برای پیش‌بینی رخدادهای عدم موجودی کالا استفاده شده‌اند، به دلیل خطاهای انسانی و هزینه‌های بالای پرسنل و ابزار فیزیکی دارای محدودیت‌های زیادی می‌باشند. به علاوه در صورتیکه مجموعه‌داده در دسترس، مشاهدات در دامنه وسیع‌تر را پوشش ندهند و یا توزیع داده‌های منبع را بصورت ناقص نمایندگی کنند، مدل‌های یادگیری ماشینی و مبتنی بر هوش مصنوعی تکی نیز دارای عملکرد مطلوبی نخواهند بود. از طرفی مدل‌های یادگیری مجتمع ایستا هم برای چنین داده‌هایی، توانایی لازم برای طبقه‌بندی اختصاصی داده‌ی جدید را ندارد. مدل‌های یادگیری مجتمع پویا این توانایی را دارند اما برای هر داده‌ی جدید صرفاً یک مدل طبقه‌بند را انتخاب، با همان مدل طبقه‌بندی داده‌ی جدید را انجام داده و از ویژگی‌ها و تنوع سایر طبقه‌بندها استفاده نمی‌کنند. بنابراین در این پژوهش، بر اساس ساختار و پراکندگی داده‌ها در فضای ویژگی، با ترکیب چندین طبقه‌بند بهینه که در فضاهای ویژگی تقسیم‌بندی شده آموزش داده شده‌اند، مدل بهینه‌ای موسوم به "سیستم طبقه‌بند مجتمع پویا مبتنی بر خوشه‌بندی و فاصله اقلیدسی" طراحی شده است تا محدودیت‌های موجود در روش‌های قبلی را پوشش داده و دارای عملکرد مناسب‌تری باشد. با استفاده از این رویکرد، محدودیت‌های موجود در روش‌های قبلی رفع می‌شود و می‌توان از ویژگی‌های هر طبقه‌بند و تنوع آن‌ها استفاده کرد. به علاوه اینکه می‌توان با استفاده کردن از مدل‌های پایه با ساختار ساده، به مدلی با تفسیرپذیری بالا دست یافت. مدل پیشنهادی با روش‌های دیگر از جمله روش‌های یادگیری ماشینی تکی و همچنین سایر روش‌های یادگیری مجتمع پویا که تا کنون مورد استفاده قرار گرفته‌اند توسط شاخص‌های استاندارد اندازه‌گیری عملکرد مدل، مقایسه شده و برتری آن احراز شده است.

واژگان کلیدی: سیستم طبقه‌بند مجتمع پویا، خوشه‌بندی، آنالیز کسب و کارها، رخداد عدم موجودی کالا

فهرست مطالب

فهرست تصاویر ث

فهرست جداول چ

فصل ۱: مقدمه ۱

۱-۱ تعریف مسئله ۱

۱-۲ ضرورت انجام پژوهش ۲

۱-۳ هدف و پرسش های تحقیق ۵

۱-۴ کلیات پیاده‌سازی مدل پیشنهادی برای پیش‌بینی کالاهای ناموجود ۶

۱-۴-۱ تحلیل توصیفی مجموعه‌داده ۶

۱-۴-۲ مدلسازی برای پیش‌بینی موجودی کالا در انبار ۶

۱-۵ ساختار فصل‌های پایان‌نامه ۱۱

فصل ۲: مروری بر پیشینه پژوهش و روش‌های شناسایی و پیش‌بینی رخدادهای عدم موجودی ۱۳

۲-۱ مقدمه ۱۳

۲-۲ مشکلات و عوامل ایجاد رویداد رخداد عدم موجودی کالا ۱۴

۲-۳ روش‌های سنتی و فیزیکی شناسایی و پیش‌بینی رویدادهای عدم موجودی کالا ۱۷

۲-۴ محدودیت‌های روش‌های سنتی و فیزیکی تشخیص و پیش‌بینی رخدادهای عدم موجودی . ۱۹

۲-۵ روش‌های یادگیری ماشین و هوش‌مصنوعی جهت پیش‌بینی رخداد عدم موجودی کالا . . ۲۰

۲-۶ ترکیب طبقه‌بند مجتمع ایستا برای پیش‌بینی رخداد عدم موجودی ۲۶

۲-۷ جمع‌بندی ۲۷

فصل ۳: نحوه طراحی طبقه‌بند مجتمع پویا به‌منظور پیش‌بینی رخداد عدم موجودی کالا ۲۹
۳-۱ مقدمه ۲۹
۳-۲ روش جمع آوری مجموعه داده ۳۰
۳-۳ تحلیل توصیفی داده‌ها ۳۰
۳-۴ پیش‌پردازش داده‌ها ۳۰
۳-۴-۱ داده گمشده ۳۱
۳-۴-۲ داده پرت ۳۲
۳-۴-۳ نرمال‌سازی داده‌ها ۳۲
۳-۴-۴ کاهش ابعاد ۳۳
۳-۵ طبقه‌بندهای پایه برای طراحی سیستم طبقه‌بند مجتمع پویا ۴۰
۳-۵-۱ مدل طبقه‌بند بیز ساده ۴۱
۳-۵-۲ مدل طبقه‌بند رگرسیون لجستیک ۴۳
۳-۵-۳ مدل طبقه‌بند درخت تصمیم ۴۴
۳-۵-۴ مدل طبقه‌بند ماشین بردار پشتیبان ۴۵
۳-۵-۵ مدل طبقه‌بند جنگل تصادفی ۴۶
۳-۵-۶ مدل‌های طبقه‌بند مبتنی بر شبکه عصبی مصنوعی ۴۷
۳-۶ سیستم‌های طبقه‌بند مجتمع مرسوم ۴۹
۳-۶-۱ مدل طبقه‌بند مجتمع ۴۹
۳-۶-۲ مدل طبقه‌بند مجتمع قدرتی ۵۰
۳-۶-۳ مدل طبقه‌بند مجتمع پشته‌سازی ۵۰
۳-۷ مدل طبقه‌بند مجتمع پویا ۵۱
۳-۸ ارزیابی مدل‌های یادگیری ماشینی ۵۳
۳-۸-۱ معرفی ماتریس "درهم ریختگی" ۵۴
۳-۸-۲ شاخص‌های ارزیابی مدل‌های یادگیری ماشینی ۵۴
۳-۸-۳ توابع زیان برای ارزیابی مدل‌های یادگیری ماشینی ۵۶

فصل ۴: ارائه و توصیف داده‌ها و مقایسه نتایج حاصل از پیاده‌سازی روش پیشنهادی با سایر روش‌ها ۵۸

۴-۱ مقدمه ۵۸

۴-۲ بررسی و دستکاری اولیه داده‌ها ۵۸

۴-۲-۱ حذف داده‌های بدون استفاده ۵۹

۴-۲-۲ حذف و یا تغییر نام متغیرها ۵۹

۴-۲-۳ یکسان‌سازی واحد قیمت‌ها ۵۹

۴-۲-۴ یکسان‌سازی واحد وزن ۵۹

۴-۲-۵ تبدیل متغیرهای کیفی به کمّی و خلق ویژگی‌های جدید ۶۰

۴-۳ تحلیل توصیفی داده‌ها ۶۱

۴-۳-۱ تحلیل توصیفی و مصورسازی متغیرهای مستقل ۶۱

۴-۳-۲ تحلیل و مقایسه فروش برندهای مختلف در مجموعه داده ۶۱

۴-۳-۳ تحلیل و مصورسازی میزان فروش و تخفیف بصورت ماهیانه ۶۴

۴-۳-۴ تحلیل و مقایسه فروش و توزیع آن در فروشگاه‌های موجود در داده‌ها ۶۵

۴-۳-۵ شاخص‌های آماری متغیرهای عددی ۶۸

۴-۳-۶ همبستگی بین متغیرهای عددی ۶۹

۴-۳-۷ رفع داده‌های پرت ۷۳

۴-۳-۸ تحلیل توصیفی و مصورسازی متغیر کلیدی ۷۶

۴-۴ نتایج حاصل از مدل‌سازی و تحلیل پیش‌بین ۷۸

۴-۴-۱ نتایج پیاده‌سازی طبقه‌بندهای واحد ۷۸

۴-۴-۲ نتایج پیاده‌سازی طبقه‌بندهای مجتمع ایستا ۸۰

۴-۴-۳ نتایج پیاده‌سازی طبقه‌بند پیشنهادی مجتمع پویا ۸۲

۴-۵ پاسخ به سوالات تحقیق ۸۸

فصل ۵: بحث و نتیجه‌گیری ۹۰

۵-۱ مقدمه ۹۰

۵-۲ جمع‌بندی پژوهش ۹۱
۵-۳ نوآوری ۹۱
۵-۴ پیشنهادها ۹۲

مراجع ۹۳

واژه نامه فارسی به انگلیسی **دوم**

واژه نامه انگلیسی به فارسی **هفتم**

فهرست تصاویر

۱-۱ مدلسازی با رویکرد تجمیع ناهمگن [۱] ۷
۱-۲ مدلسازی با رویکرد تجمیع ناهمگن [۲] ۸
۱-۳ سیستم طبقه‌بند مجتمع به روش ایستا [۲] ۸
۱-۴ سیستم طبقه‌بند مجتمع به روش پویا [۲] ۹
۱-۵ سه نمونه تجزیه تحلیل برای داده‌ها ۱۱

۱-۲ روش مشارکتی پیشنهاد شده در [۳] ۲۴

۳-۱ مراحل کاهش بعد با استفاده از PCA [۴] ۳۶
۳-۲ ساختار شبکه خودرمزگذار [۵] ۳۸
۳-۳ نمایش داده‌های کاهش بعد داده شده توسط روش‌های کاهش بعد UMAP، T-SNE [۶] . ۳۹
۳-۴ انواع مدل‌های یادگیری ماشینی ۳-۴ . . . . . ۴۱
۳-۵ مدل‌های رگرسیون لجستیک ساده و چند متغیره [۷] ۴۴
۳-۶ نمونه‌ای از درخت تصمیم سخت و نرم [۸] ۴۵
۳-۷ نحوه عملکرد مدل SVM در فضای دو بعدی [۹] ۴۶
۳-۸ نمونه‌ای از شبکه عصبی پیشخور [۱۰] ۴۸
۳-۹ مقایسه ساختار سه مدل تجمیع مختلف [۱۱] ۵۰
۳-۱۰ مراحل شکل‌گیری خوشه‌ها توسط الگوریتم خوشه‌بندی K-Means [۱۲] ۵۲
۳-۱۱ شماتیک مربوط به طبقه‌بند مجتمع پویای مبتنی بر فاصله [۱۲] ۵۳
۳-۱۲ نمونه‌ای از ماتریس درهم‌ریختگی دوکلاسه ۵۵

۴-۱ مقایسه فروش و میزان کالاهای تخفیف خورده و بدون تخفیف برای برندهای مختلف . . . ۶۲

۲-۴ مقایسه بین برندهای مختلف از نظر فروش و همچنین رقابت آن‌ها در فروش و تخفیف کالاها ۶۳
۳-۴ میزان کل فروش بدون تخفیف و با تخفیف برندها و همچنین درآمد کل بصورت ماهانه . . ۶۴
۴-۴ نمودار میله‌ای مربوط به میزان فروش ماهیانه‌ی فروشگاه‌های مختلف ۶۵
۵-۴ توزیع میزان فروش در فروشگاه‌های مختلف مجموعه داده ۶۶
۶-۴ توزیع میزان فروش بر حسب قیمت قیمت در ماه‌های مختلف هر فروشگاه ۶۷
۷-۴ توزیع میزان فروش بر حسب قیمت ماه‌های مختلف در فروشگاه Best Boy ۶۸
۸-۴ ماتریس همبستگی بین متغیرهای عددی ۷۰
۹-۴ توزیع متغیرهای عددی و همبستگی بین آن‌ها ۷۲
۱۰-۴ توزیع متغیرهای عددی و همبستگی بین آن‌ها بعد از حذف تاثیر منفی داده‌های پرت . . . ۷۴
۱۱-۴ نمودار میله‌ای مربوط به فراوانی متغیر IsSale ۷۵
۱۲-۴ نمودار میله‌ای مربوط به فراوانی متغیر موجودیت کالا ۷۷
۱۳-۴ نمودار جعبه‌ای توزیع صحت آزمایش، در طبقه‌بندها ۷۹
۱۴-۴ نمودار جعبه‌ای توزیع صحت آزمایش، در طبقه‌بندهای مجتمع ایستا ۸۱
۱۵-۴ مراحل ایجاد توده‌های محلی با استفاده از روش خوشه‌بندی K-means ۸۲
۱۶-۴ مقایسه توزیع شاخص ROC-AUC مربوط به انواع طبقه‌بندهای مجتمع پویا ۸۷

فهرست جداول

۲-۱ نتایج مربوط به مقاله [۱۳] ۲۷

۴-۱ معرفی متغیرها و نوع آن‌ها ۶۰

۴-۲ مقادیر شاخص‌های آماری متغیرهای عددی در مجموعه داده ۶۹

۴-۳ میانه و میانگین توزیع صحت در طبقه‌بندهای مختلف ۷۹

۴-۴ میانه و میانگین توزیع صحت در طبقه‌بندهای مجتمع مختلف ۸۰

۴-۵ میانه و میانگین توزیع ROC-AUC در طبقه‌بندهای مجتمع مختلف ۸۶

فصل ۱

مقدمه

۱-۱ تعریف مسئله

در سال‌های اخیر افزایش روز افزون منابع اطلاعاتی از جمله شبکه‌های اجتماعی، وبسایت‌های خبر رسانی، وبسایت‌های تبلیغاتی، رسانه‌ها، اطلاعات درون‌سازمان‌ها شرکت‌ها و ... سبب بوجود آمدن داده‌های عظیم در حوزه‌های متعدد شده‌است. این گستردگی اطلاعات، فرصت مناسبی برای نظارت و کنترل هر چه بهتر فرآیندهای مختلف اقتصادی، اجتماعی، سیاسی و...، فراهم آورده است این امر مستلزم تحلیل داده‌ها بصورت مؤثر و استخراج اطلاعات مفید از داخل داده‌های عظیم می‌باشد [۱۴، ۱۵]. تحلیل داده‌های مربوط به فرآیندهای اقتصادی، چه در شرایط عادی و چه در شرایط اضطراری مانند شیوع اپیدمی های مختلف، جنگ، قحطی و ...، که ممکن است منجر به رکود اقتصادی شوند ، به منظور نظارت و کنترل سریع این فرآیند ها، حائز اهمیت است.در همین راستا مفهومی به نام تجزیه و تحلیل کسب و کارها[1] (BA)، تعریف شده است. عبارت است از استفاده از داده در مقیاس وسیع ، تحلیل‌های کمی و آماری ، مدل‌های پیشبینی کننده و توضیحی و مدیریت مبتنی بر واقعیات در راستای هدایت تصمیمات و اقدامات لازم [۱۶]. در فرآیند تجزیه و تحلیل کسب و کارها تحلیل داده‌های مربوط به یک شرکت و یا سازمان با هدف کمک به سرمایه‌گذاری‌های بلند و کوتاه‌مدت مناسب در کسب و کارشان و بازاری که در آن فعال هستند انجام می‌پذیرد.

[1] Business Analytics

از طرفی با توجه به گسترش بازار رقابتی، تحلیل بازار و رفتار مصرف‌کنندگان به منظوراخذ و اعمال سیاست گذاری های بهینه سیار مورد تاکید است. به عبارتی تحلیل داده‌های مربوط به یک شرکت و یا سازمان با هدف کمک به سرمایه‌گذاری‌های بلندمدت و کوتاه‌مدت مناسب در کسب و کارشان و بازاری که در آن فعال هستند انجام می‌پذیرد. تمرکز بر روی یافتن پاسخ سوال‌های : '' چه اتفاقی افتاده است '' ، '' چند وقت'' ، '' کجا '' برای یافتن پاسخ به سوالات : '' چرا '' ، ''چه می شود اگر این روند ادامه داشته باشد '' ، '' در آینده چه خواهد شد '' و '' سناریوی ایده آل چه خواهد بود '' به منظور تحلیل کسب و کارها الزامی است. برای یافتن پاسخ به این سوال ها تحلیل های آماری، مدل های پیشبینی کننده و بهینه‌سازها باید بکار برده شوند. بینش بدست آمده از توسعه و پیاده سازی این ابزارهابا محوریت روش‌شناسی در کسب و کارها و با هدف اتخاذ تصمیم‌گیری‌های به‌موقع در شرایط مختلف مورد استفاده قرار می‌گیرد [۱۷] فرآیند جمع‌آوری، تحلیل، توصیف و مدلسازی بر روی داده های حجیم به منظور تصمیم‌گیری‌های به‌موقع در سازمان‌ها و کسب و کارهایی که تمایل به توسعه برند و بهبود شاخص رقابت‌پذیری خود را دارند در سطح کلان مورد استفاده انجام می‌پذیرد [۱۸]. ضمن اینکه به علت وجود منابع مختلف داده و تنوع در نوع داده‌ها و حجم عظیم داده استفاده از روش‌های سنتی و معمول تحلیل داده امکانپذیر نمی‌باشد. در نتیجه باید از ابزارهای قدرتمند و بروز داده‌کاوی به منظور تحلیل داده‌های مربوط به کسب و کارها و سازمان‌ها در مقیاس‌های مختلف استفاده کرد.

۱-۲ ضرورت انجام پژوهش

برنامه‌ریزی برای موجودی کالا بخش اساسی در آنالیز کسب‌وکارها می‌باشد. برای دهه‌ها عدم موجودی کالا مسئله مهم برای خرده‌فروشان و تولیدکنندگان بوده‌است. مدیریت موجودی روشی برای پیگیری روند موجودی کالا در کسب‌وکار است، به عبارتی سیستم مدیریت موجودی به کسب‌وکار کمک می‌کند تا میزان موجودی شرکت ردیابی و کنترل شود تا موجودی را بهینه کرده و بدون صرف وقت و هزینه اضافی، مدیریت شوند. مدیریت موجودی، فرآیند سفارش، ذخیره و تحویل محصولات به‌صورت کالاهای نهایی یا مواد اولیه برای تولید کنندگان است و بر اصل تعادل بنا شده است، یعنی داشتن یک محصول مناسب، در مقدار مناسب، در زمان مناسب، در مکان مناسب و با هزینه‌ی مناسب [۱۳]. این روش، نحوه‌ی سفارش، نگهداری و تحویل موجودی را نشان می‌دهد و همچنین جزء اصلی عملیات توزیع کارآمد سازمان است.

موجودی یا عدم موجودی می‌تواند تاثیر گسترده‌ای بر روی بقیه زنجیره تامین داشته باشد. سه قسمت اصلی مدیریت موجودی کالا عبارتند از:

- خرید: به دست آوردن کالاهای تمام شده در انتظار فروش یا مواد اولیه مورد نیاز برای تولید محصول که این مرحله آغاز روند مدیریت موجودی است.
- ذخیره سازی: مشاغل از شیوه‌های کنترل موجودی کالا برای مدیریت سطح موجودی کالا و بهینه‌سازی سطح عرضه استفاده می‌کنند. ابزارهایی مانند شمارش منظم موجودی، بهینه‌سازی مکان‌های حمل‌ونقل و اتوماسیون روش‌هایی برای بهبود تلاش‌های ذخیره‌سازی موجودی استفاده می‌شوند.
- تحویل: آخرین مرحله مدیریت موجودی، ارسال محصول به مشتریان است. لجستیک در اینجا نقشی اساسی دارد، زیرا شواهد نشان می‌دهند که یکی از سریع‌ترین راه‌های از دست‌دادن مشتریان، تحویل غیرقابل اعتماد یا دیررس است.

از دیدگاه محصول، اهمیت مدیریت موجودی کالا در درک این است که چه موجودی در انبار خود دارید شما و نحوه ورود و خروج آن چیست. مشاغلی که موجودی بیش از حد داشته باشد، دارای اضافه‌فروشی است. مشاغلی که موجودی بیش از حد دارند، جریان نقدی را محدود می‌شود و آن عامل به طور بالقوه کسری بودجه ایجاد می‌کند. این موجودی مازاد که به آن موجودی مرده نیز گفته می‌شود، غالباً در انبار قرار می‌گیرد، قادر به فروش آن نیستند و حاشیه سود مشاغل را می‌خورد، اما اگر موجودی کسب‌وکار کافی نباشد، می‌تواند بر خدمات مشتری تأثیر منفی بگذارد. کمبود موجودی کالا به معنای از دست دادن فروش یک تجارت است. اینکه به مشتریانتان بگویید موجودی ندارید، می‌تواند باعث شود مشتریان سفارششان را از جای دیگری تامین کنند. یک سیستم مدیریت موجودی می‌تواند به مشاغل کمک کند تا بین کمبود و بیش از حد را برای کارآیی و سودآوری مطلوب تعادل برقرار کنند.

از طرفی مدیریت موجودی‌ها شامل تدوین و اجرای خط مشی‌ها، سیستمها و روش‌هایی است که اتخاذ آنها می تواند جمع هزینه های مرتبط با تصمیمات مربوط به موجودیها و وظایفی نظیر تامین نیازهای مشتریان، برنامه ریزی تولید، خرید، حمل و نقل و امثالهم را به حداقل برساند. بنابراین زمینه مدیریت موجودیها وسیع است و بر بسیاری از فعالیتهای واحد تجاری اثر می گذارد. از طرف دیگر، کنترل موجودیها اساسا در ارتباط با اجرای خط مشی ها، سیستمها و روشهای استقرار یافته می‌باشد. مثلا فعالیت‌های کنترل موجودی‌ها ممکن است شامل نگاهداری مدارک و گزارشهای مربوط به موجودیها، مراحل تهیه و صدور درخواستهای خرید مواد و شمارش موجودی‌ها باشد.

تکنیکهای مختلف کنترل موجودی‌ها که تا کنون مورد استفاده قرار گرفته‌اند را می‌توان از جمله تکنیکهای عملی نام برد که در بسیاری از واحدهای تجاری به طور موفقیت آمیزی مورد استفاده قرار گرفته اند. در حالیکه برخی از تکنیکهای کنترل موجودی‌ها برای بسیاری از واحدهای تجاری بزرگ با ارزش و واجد اهمیت است لزوما نسبت به وضعیت کلی و عمومی موجود بها قابل اعمال و یا از نظر اقتصادی عملی نیست زیرا این تکنیک‌ها غالبابر برخی مفروضات و نیازهای خاص مبتنی می‌باشند. باید توجه داشت که به‌کارگیری تکنیکهای پیچیده الزاما به خودی خود به سیستم مفید و موثر مدیریت موجودی‌ها منجر نخواهد شد. در بیشتر موارد نبود اطلاعات صحیح و به موقع مزایای بکارگیری تکنیکهای کنترل موجودی را از بین می برد. علاوه بر این، مشکل بالقوه این است که تکنیکهای پیچیده کنترل موجودیها ممکن است به دلیل عدم درک مفاهیم اصلی این تکنیکها توسط مدیریت یا کارکنانی که با کنترل موجودیها سروکار دارند، به نحوی نادرست بکار گرفته و اعمال شود. با پیشبینی به‌هنگام موجودی کالاهای شرکت را می‌توان گام مهم و اساسی در مدیریت موجودی کالا برداشت.

۱-۳ هدف و پرسش های تحقیق

هدف اصلی این پژوهش، ایجاد رویکردی مناسب جهت رفع یکی از مشکلات موجود در آنالیز کسب و کارها، در حوزه پیش‌بینی رخداد عدم موجودی[۲] می‌باشد. در واقع با پیش‌بینی و تشخیص صحیح رخداد ناموجودی کالا، سعی بر تصمیم‌سازی هوشمند مبتنی بر مدل‌های ترکیبی یادگیری ماشینی و هوش‌مصنوعی، به منظور رفع تاثیرگذاری منفی مشکل ناموجودی کالا بر روی کسب و کارهای فروشگاهی است. پرسش کلی دربرگیرنده این هدف این است که: چگونه در یک سازمان می‌توان با بکار بردن رویکردهای مبتنی بر هوش مصنوعی برای استفاده حداکثری از داده‌های در دسترس و بهنیه سازی فرآیند BA، تصمیمات و سیاست‌های مؤثر را اتخاذ کرد. همچنین پرسش‌های اصلی این پژوهش که طی فصل‌های آینده به آن‌ها پاسخ داده می‌شود عبارت اند از:

- آیا می‌توان بدون استفاده از ابزارهای گران‌قیمت فیزیکی و با صرف هزینه معقول، رخدادهای عدم موجودی کالا را با صحت و دقت بالا پیش‌بینی کرد؟
- آیا استفاده از روش‌های یادگیری ماشینی و هوش‌مصنوعی نسبت به روش‌های سنتی در پیش‌بینی رخدادهای عدم موجودی کالا عمکرد مناسب‌تری دارند؟
- آیا می‌توان علاوه بر پیش‌بینی هر چه دقیق‌تر رخدادهای عدم موجودی کالا در زمان مناسب، به تفسیرپذیری مناسبی برای روش و مدل پیشنهادی دست یافت؟
- آیا استفاده از مدل طبقه‌بند یکتای مبتنی بر یادگیری ماشینی در هر شرایطی عملکرد بهینه دارد؟
- مزایای استفاده از ترکیب طبقه‌بندها (یادگیری مجتمع) چیست؟
- با چه ترکیبی از طبقه‌بندها یه بهینه‌ترین مدل مجتمع می‌توان دست یافت؟
- مدل طبقه‌بند مجتمع پویا چه مزایایی نسبت به طبقه‌بند مجتمع ایستا دارد؟

یکی از چالش‌های موجود در آنالیز کسب و کارها پیشبینی موجودی کالاها می‌باشد. با پیشبینی موجودی کالا می‌توان سیاست‌های بهینه از جمله میزان تولید یا تهیه کالا برای جلوگیری از کمبود و یا باقی ماندن کالای اضافی در انبار را اتخاذ کنیم. در این پژوهش با پیاده‌سازی یک طبقه‌بند مجتمع پویای بهینه مبتنی بر هوش مصنوعی، به این مسئله پرداخته می‌شود.

[2] Out of stock

۱-۴ کلیات پیاده‌سازی مدل پیشنهادی برای پیش‌بینی کالاهای ناموجود

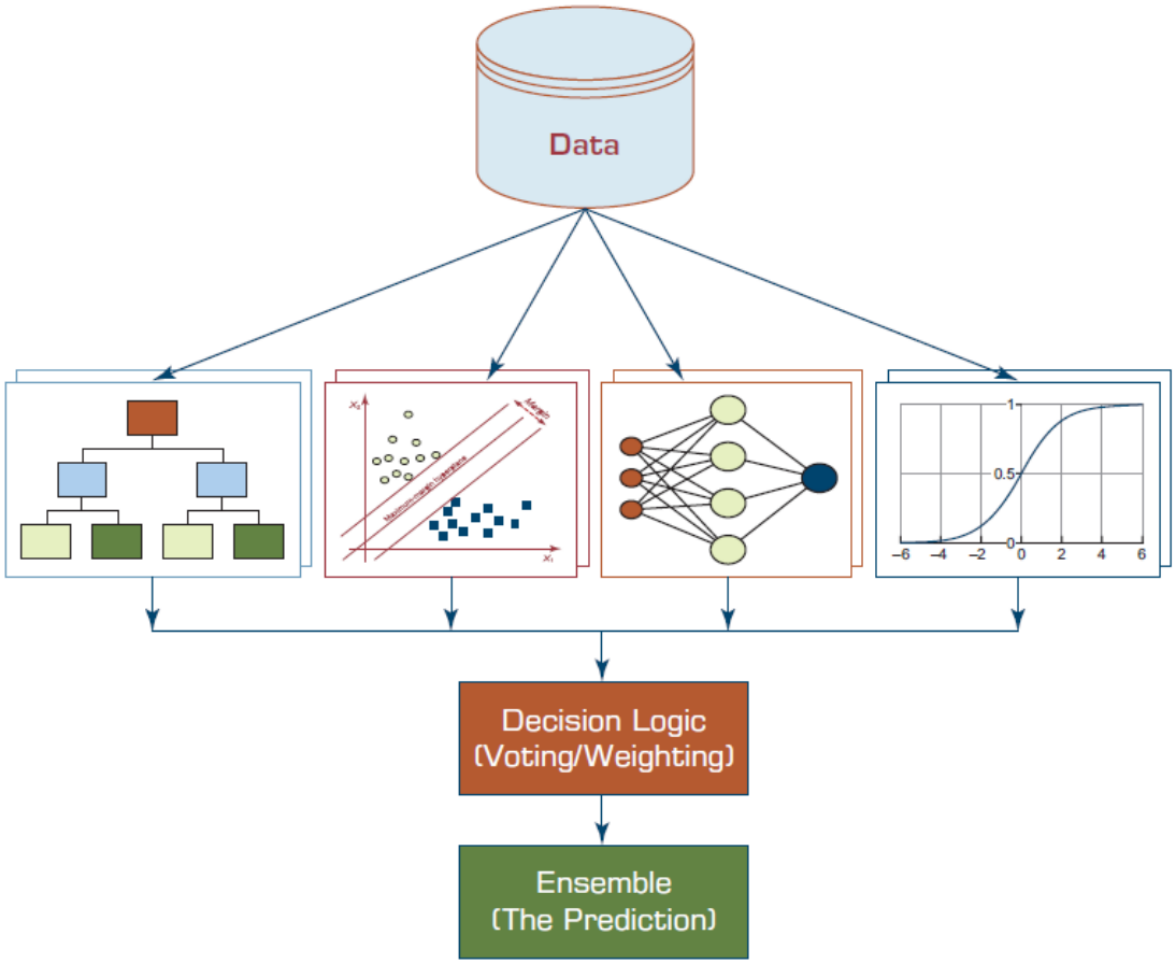
۱-۴-۱ تحلیل توصیفی مجموعه‌داده

در مرحله اول به تحلیل توصیفی داده‌ها پرداخته می‌شود. تحلیل توصیفی داده‌ها تمرکز بر اطلاع از اتفاقات داخل سازمان و فهم برخی از روندهای مهم و اساسی و علت رخداد چنین اتفاقاتی دارد. به منظور تحلیل داده‌ها با سرعت و دقت بالا، ابتدا باید به یک بینش و دید اولیه نسبت به آنها رسید.به علاوه فهم داده‌هابه منظور پیش پردازش پیش‌پردازش[۳] و تمیزسازی داده‌ها تمیزسازی داده‌ها[۴] امری ضروریست. بررسی نوع متغیرها، ارتباط بین متغیرهای ورودی با یکدیگر، ارتباط بین متغیرهای ورودی با متغیر خروجی و مصورسازی داده‌ها از روش هاییست که برای فهم داده‌ها مورد استفاده قرار می‌گیرند.

۱-۴-۲ مدلسازی برای پیش‌بینی موجودی کالا در انبار

همانطور که مفهومی به نام بهترین مدل وجود ندارد ، بهترین نوع مدلسازی بصورت عمومی هم تعریف نشده است. بنابراین بسته به نوع مسئله، ساختار و پراکندگی داده‌ها در فضای ویژگی داده‌ها، می‌توان با انتخاب نوع پارامترهای مدل منتخب، مدلسازی بهینه را برای داده‌های مورد مطالعه انجام داد [۱]. هر کدام از انواع طبقه‌بندها خصوصیات منحصربه‌فرد خودشان را دارند و با انتخاب یکی از آنها به عنوان بهترین تک‌مدل ، ویژگی‌های دیگر طبقه‌بندها را از دست داده ایم. بنابراین برای مدلسازی بهینه، رویکردی موسوم به سیستم طبقه‌بند چندگانه[۵] ،(MCS) به‌منظور پیاده‌سازی مدل‌های مجتمع را می‌توان به‌کار برد. مدل های مجتمع به دو دسته همگن همگن[۶] و ناهمگن ناهمگن[۷] تقسیم‌بندی می‌شوند. مدل همگن از اجتماع دو یا چند طبقه‌بند همنوع، تشکیل شده است. (مثلا دو یا چند طبقه بند شبکه عصبی) و مدل ناهمگن از اجتماع دو یا چند طبقه‌بند غیر هم‌نوع ساخته می‌شود. مدل های ناهمگن را مدل های مبتنی بر ترکیب اطلاعات نیز می نامند ، هدف استفاده از این نوع رویکرد مجتمع، طراحی مدل مجتمعی با دقت و قوام بالا نسبت به رویکرد تک مدل می‌باشد. در شکل ۱-۱ مدلسازی با رویکرد تجمیع ناهمگن به تصویر کشیده شده است. در این

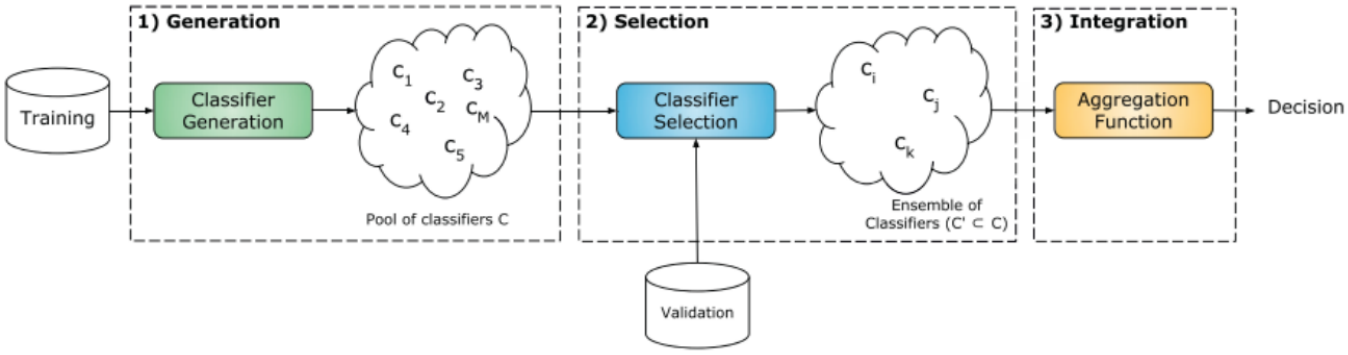
[3] Preprocessing
[4] Data Cleaning
[5] Multiple Classifier System
[6] Homogeneous
[7] Heterogamous



شکل ۱-۱: مدلسازی با رویکرد تجمیع ناهمگن [۱]

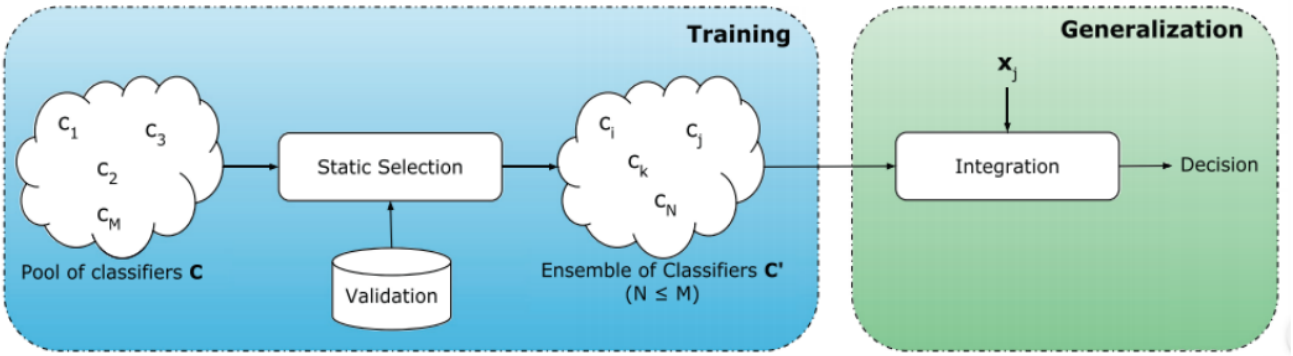
رویکرد مدلسازی در سه مرحله اصلی انجام می‌شود : تولید، انتخاب و ترکیب. در شکل ۱-۲ شماتیک مربوط به این سه مرحله قابل مشاهده می‌باشد .(فرض بر این است مدل های C_1 تا C_M انواع مختلفی از طبقه‌بندها می‌باشند.) در مرحله تولید، انواع مختلفی از طبقه‌بندها (شبکه عصبی، درخت تصمیم، بیز ساده، مدل‌های مبتنی بر هوش مصنوعی، ماشین‌های بردار پشتیبان، مدل کا نزدیکترین همسایه[۸] (KNN) آموزش داده می‌شوند. همچنین با تنظیم پارامترها و ساختار هر طبقه‌بند، تنوع طبقه‌بندها بیشتر می شود. در مرحله انتخاب با ارزیابی مدل‌هایی که در مرحله قبل آموزش داده شده اند، دو یا چند طبقه‌بند بهینه انتخاب شده و در مرحله آخر یعنی مرحله تجمیع، خروجی طبقه‌بندهای مجتمع با هم ترکیب شده و بر اساس این خروجی، تصمیم نهایی اتخاذ

[8] K-Nearest-Neighbors

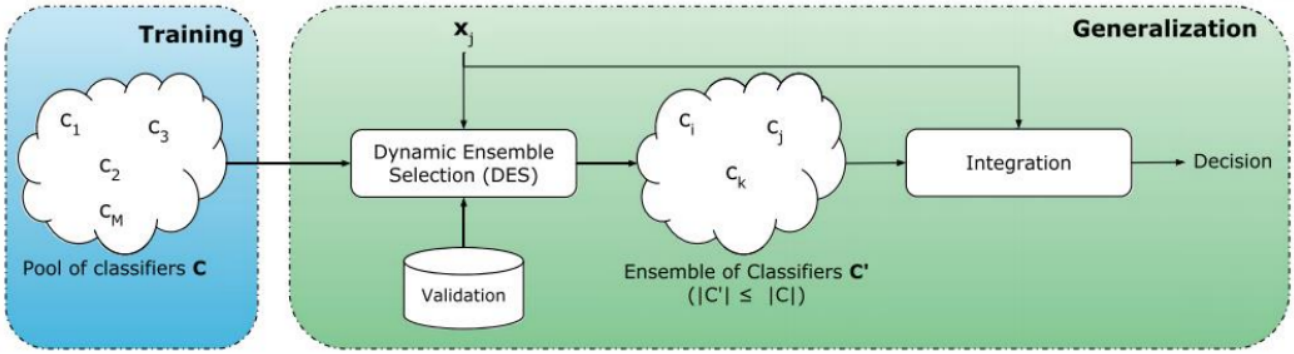


شکل ۱-۲: مدلسازی با رویکرد تجمیع ناهمگن [۲]

می‌شود. مدل های مجتمع ناهمگن/همگن به دو روش ایستا و یا پویا قابل پیاده سازی می باشند. سیستم طبقه بند مجتمع به روش ایستا درشکل ۱-۳ و به روش پویا در شکل ۱-۴ قابل مشاهده می‌باشند.



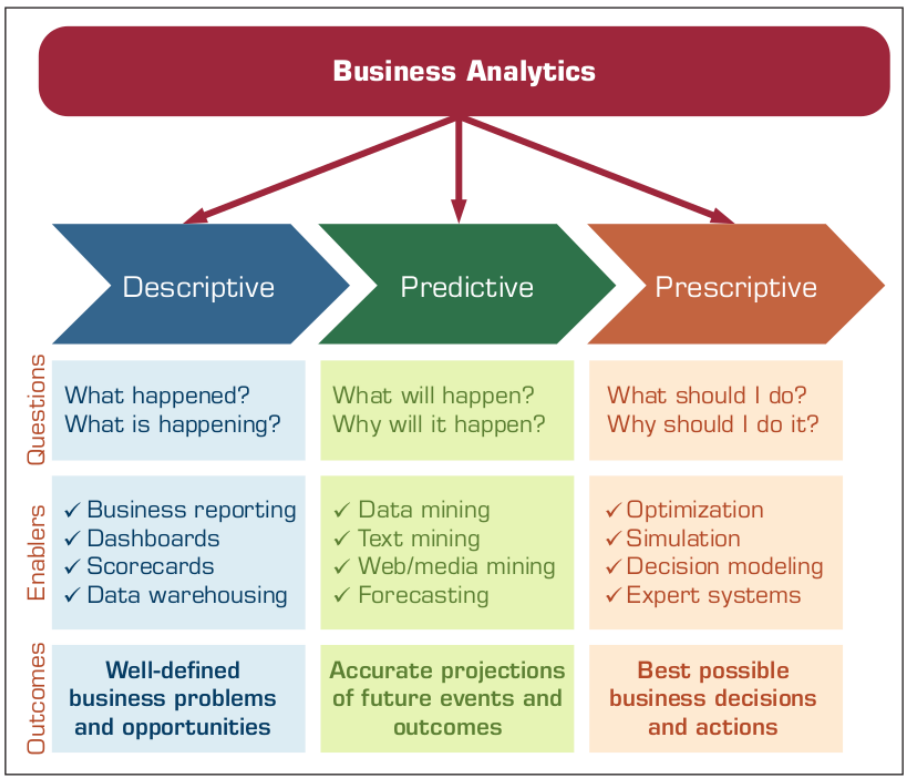
شکل ۱-۳: سیستم طبقه‌بند مجتمع به روش ایستا [۲]



شکل ۱-۴: سیستم طبقه‌بند مجتمع به روش پویا [۲]

مشـــاهده می‌شـــود تفاوت این دو روش در مرحله انتخاب می باشـــد. در روش ایســـتا با اســـتفاده از داده های ارزیابی، از میان N طبقه‌بند که در مرحله تولید بر روی داده‌های آموزش، آموزش داده شده‌اند، M طبقه‌بند بهینه انتخاب می‌شود و تمامی داده‌های آزمایش توسط این M کلاس طبقه‌بندی خواهند شد. اما روش پویا مبتنی بر پراکندگی و ســـاختار داده‌ها در فضای ویژگی اســـت. به این معنی که ابتدا داده‌ها خوشـــه‌بندی می شـــوند (یادگیری بدون نظارت) سپس به ازای هر دســـته از داده‌ها دو یا چند طبقه‌بند مختلف در همان ناحیه از فضای ویژگی که مختص خودشان است مورد آموزش و ارزیابی قرار می‌گیرند. همچنین در مرحله آزمایش، بسته به اینکه داده آزمایش مورد نظر در چه ناحیه‌ای از فضا قرار می‌گیرد، دو یا چند طبقه‌بند به داده مذکور اختصـاص داده شـده و توسـط آنها طبقه‌بندی می‌شـود و در نهایت خروجی این طبقه‌بندها در مرحله تجمیع، با هم ترکیب می‌شوند و نتیجه نهایی حاصل خواهد شد. این طبقه بند ها در مرحله تجمیع ، با هم ترکیب می شوند. سیستم یادگیری مجتمع پویا[۹] (DECS) روشی است که دراین پایان‌نامه مورد اسـتفاده قرارخواهد گرفت. در مرحله تولید طبقه‌بندهای درخت تصـمیم، ماشـین بردار پشـتیبان، شـبکه عصبی، بیز ساده و به منظور خوشه‌بندی ازروش KNN استفاده خواهد شد. فرآیند تجزیه و تحلیل، ترکیبی از تکنولوژی‌های رایانه‌ای، روش‌های علمی مدیریتی و آمار برای حل مسائل واقعی می‌باشد. منابع مختلف تفسیرهای متعددی برای عبارت تجزیه و تحلیل[۱۰] ارائه کرده‌اند. از منظر اصلی، تجزیه و تحلیل به عنوان فرآیند بهینه‌سازی تصمیمات عملی از طریق پیشنهادات مؤثر، بر اساس تاریخچه اطلاعات در دسترس می‌باشد [۱]. تحلیل های آماری فهم بهتری از الگوهای درون داده‌ها به ما می‌دهند در نتیجه‌ی این امر در مرحله بعد می‌توان مدل‌های پیشبینی کننده را برای تحلیل رفتار مشتریان توسعه داد. هنگامی که یک سازمان یک دید مناسب نسبت به اینکه چه اتفاقی در حال رخ دادن است و چه اتفاقاتی در آینده رخ خواهند داد، می‌تواند با در پی گرفتن سیاست‌های مناسب در شرایط مختلف مناسب‌ترین تصمیمات را اخذ کند. این سه مرحله به عنوان آنالیز توصیفی[۱۱] ، آنالیز مبتنی بر پیشبینی[۱۲] و آنالیز آینده‌نگر[۱۳] شناخته شده می‌شود که در ادامه به شرح این مراحل پرداخته می‌شود. شکل ۱-۵ نمایانگر شمای کلی این سه فاز از تجزیه تحلیل می‌باشد.

[9] Dynamic Ensemble Learning
[10] Analytics
[11] Descriptive Analytics
[12] Predictive Analytics
[13] Prespective Analytics



شکل ۱-۵: سه نمونه تجزیه تحلیل برای داده‌ها

۱-۵ ساختار فصل‌های پایان‌نامه

پژوهش پیش‌رو از پنج فصل تشکیل شده است.

- در فصل اول به مقدمه و کلیات پژوهش پرداخته شده است. همچنین به تشریح موضوع پژوهش و انگیز پرداختن به آن، طرح سوالات اصلی پژوهش، به علاوه اشاره‌ای کوتاه به روش و مراحل تحقیق پرداخته شده است.
- در فصل دوم به بررسی و طبقه‌بندی یافته‌های تحقیقات دیگر محققان و تعیین و شناسایی خلأهای تحقیقاتی آن‌ها در زمینه تشخیص و پیشبینی رخدادهای عدم موجودی کالا پرداخته شده است.
- در فصل سوم، روش پژوهش بصورت کامل تشریح شده است. همچنین به نحوه جمع‌آوری مجموعه‌داده، مراحل پیش‌پردازش داده‌ها، چکونگی تحلیل توصیفی داده‌ها و نحوه پیاده‌سازی مدل پیشنهادی پرداخته شده است.

- در فصل چهارم به ارائه داده‌ها، نتایج، تحلیل و تفسیر مدل پیشنهادی و مقایسه آن با سایر مدل‌ها پرداخته شده است.
- نهایتا در فصل پنجم به ارائه برداشت‌ها از نتایج و یافته‌های ناشی از پیاده‌سازی مدل پیشنهادی بر روی مجموعه‌داده پرداخته شده و همچنین به پرسش‌های پژوهش که در فصل یک طرح شده‌اند پاسخ داده می‌شود.

فصل ۲

مروری بر پیشینه پژوهش و روش‌های شناسایی و پیش‌بینی رخدادهای عدم موجودی

۲-۱ مقدمه

علیرغم تلاش‌ها برای کاهش تعداد کالاهای با برچسب عدم موجودی (OOS) عدم موجودی (تلاش برای همواره موجود نگه داشتن کالاهای مورد نظر) این رویدادها به طور گسترده‌ای رایج هستند [۱۳]. از منظر عملیاتی، رویدادهای OOS تأثیر بالقوه تبلیغات را کاهش داده و تخمین میزان تقاضای واقعی را نیز مخدوش می‌کنند. از دیدگاه استراتژیک، رویدادهای OOS بر وفاداری به برند نیز تأثیر می‌گذارند، برندهای رقیب را ترویج کرده و اثربخشی منابع تیم فروش را کاهش می‌دهند. یک روش معمول برای تولیدکنندگان به منظور شناسایی و پیش‌بینی رویدادهای OOS، شناسایی ناهنجاری‌ها در داده‌های معاملاتی ارائه شده توسط خرده‌فروشان می‌باشد. اخیرا شاهد هستیم که چنین فرآیندی را می توان با مدل‌های یادگیری ماشین و هوش مصنوعی، خودکار کرد. این رویکرد برای اولین بار توسط [۱۹] معرفی شد. در [۲۰] مدل های کنترل موجودی و روش های طبقه‌بندی مورد مطالعه قرار گرفت و در [۲۱] الگوریتم های طبقه بندی مختلف با هم مقایسه شدند و یک رویکرد یادگیری گروهی برای مقابله با عدم توازن کلاس‌ها در تعداد، استفاده شد.

دیگر روش‌ها برای شناسایی و پیش بینی رویدادهای OOS با استفاده از داده‌های تراکنش در [۲۲] [۲۳] گزارش شده‌اند. موضوع OOS همچنان زمینه مورد علاقه برای مطالعات تحقیقاتی است. مقالاتی که اخیر منتشر شده‌اند به مشکلات OOS از رویکردهای مختلف پرداخته‌اند از جمله روش‌های پردازش تصویر [۲۴]، سیستم رباتیک خودکار برای نظارت قفسه [۲۵]، ربات متحرک، دوربین های عمقی و شبکه های عصبی برای تعیین میزان اشغال قفسه [۲۶] تشخیص OOS با استفاده از یادگیری عمیق [۲۷] نظارت بر قفسه با استفاده از یادگیری نظارت شده [۲۸] و ترکیبی از یادگیری نیمه نظارت شده و یادگیری عمیق برای نظارت بر در دسترس بودن کالا در قفسه‌ها [۲۹]. در [۱۳] اثربخشی یادگیری ماشینی را در پیش‌بینی و تشخیص رویدادهای OOS تولیدکننده‌ای که چندین محصول را در یک خرده‌فروشی چند فروشگاهی می‌فروشد مطالعه شده‌است. این مقاله، مطالعه خود را بر روی یک تولید کننده آجیل و میوه های خشک متمرکز کرده است که محصولات بسته‌بندی شده خرده خود را به یک زنجیره مهم مواد غذایی در آمریکای لاتین می‌فروشد. همچنین در این مقاله ۶۸٪ از رویدادهای OOS با دقت ۷۲٪ شناسایی شده‌اند. مطالعه درباره OOS در پنجاه سال اخیر یک زمینه فعال بوده است. مقالات در این زمینه را می توان به طور کلی به سه گروه طبقه بندی کرد:

- آن‌هایی که عوامل ایجاد رویدادهای OOS را مطالعه می کنند.
- آن‌هایی که نحوه واکنش مصرف کنندگان به OOS را مطالعه می کنند.
- آن‌هایی که نحوه پیش بینی و تشخیص رویدادهای OOS را مطالعه می کنند.

۲-۲ مشکلات و عوامل ایجاد رویداد رخداد عدم موجودی کالا

در دهه‌های گذشته، موضوع رویدادهای OOS همواره یکی از موضوعات مهم برای مطالعات [۳۰] و بخشی از زمینه تحصیلی در عملیات خرده فروشی، بوده است [۳۱].

با توجه به اهمیت رویدادهای OOS، حوزه های مختلف تحقیقاتی به این مسئله پرداخته اند. تحقیق در بازاریابی و رفتار مصرف کننده، واکنش مصرف کننده را هنگام مواجهه با وضعیت OOS مورد مطالعه قرار داده است. برخی از پیامدهای احتمالی آن، واکنش های منفی مصرف کننده، تعویض فروشگاه، تعویض محصول یا به تعویق انداختن خرید می‌باشد [۳۲]. حوزه دیگری از تحقیقات، بر روی مطالعه عوامل بوجود آمدن OOS در سالهای اخیر می‌باشد.

[۳۳] یک بررسی سیستماتیک در مورد این عوامل ارائه کرده است. می توان آن‌ها را به دو گروه اصلی طبقه بندی کرد، شیوه های خرده‌فروشی و مشکلات بالادستی در زنجیره تامین خرده فروشی. چند مثال برای گروه اول عبارتند از عدم دقت در ثبت میزان موجودی، فشرده‌سازی و روش‌های اشتباه در پر کردن قفسه‌ها. گروه دوم هم فشرده‌سازی ناشی از جابجایی و انتقال محصول و عدم دقت در پیشبینی است [۳۴، ۳۵، ۳۶، ۳۰].

[۳۷] مشکلات ایجاد شده ناشی از شیوه‌های غلط بکار رفته درفروشگاه‌های خرده‌فروشی و اطلاعات نادقیق در سیستم ثبت موجودی را مورد بررسی قرار می‌دهد. در این تحقیق نویسندگان نشان می‌دهند که عدم دقت در فهرست موجودی کالاها، منجر به وضعیت OOS خواهد شد. [۳۸] تعیین می‌کند که یک وظیفه کلیدی برای خرده فروشان انتخاب محصول و فضای قفسه اختصاص داده شده به آن و ساختار محل قرارگیری محصول معمولا به عوامل خارج از انبار وابسته است.آنها برای کاهش تعداد رویدادهای OOS یک استراتژی برای تخصیص مجدد بهینه فضای قفسه پیشنهاد کردند.

در [۳۹] نویسندگان پیشنهاد کردند که عملیات خرده فروشی و تدارکات در فروشگاه، زمینه‌های کلیدی برای بهبود موجودیت کالا در قفسه است، اما این امر صرفا با اصلاح فرآیندهای جاری قابل اجرا است.

با پیشرفت تکنولوژی شناسایی فرکانس رادیویی[1] در دهه گذشته در عملیات خرده فروشی استفاده شده است. [۴۰] درباره چگونگی مفید واقع شدن RFID برای اتوماسیون تصمیمات بهینه برای چیدن کالاها در قفسه‌های فروشگاه خرده‌فروشی و نظارت بر جابجایی محصولات بین انبار و قفسه‌ها تحقیق کرده است. در مقاله مذکور یک استراتژی برای تخصیص بهینه فضای قفسه‌ها به‌منظور کاهش تعداد رویدادهای OOS پیشنهاد کرده است.

[1] Radio Frequency Identification

از روی شاخصه‌های سمت تقاضا، می توانیم توزیع کالا را به دو نوع دسته‌بندی کنیم:

- سنتی:

یعنی تولیدکنندگان محصولات را به مرکز توزیع خرده‌فروشی تحویل می دهند و از این طریق محصولات در فروشگاه‌ها توزیع می‌شوند.

- تحویل مستقیم به فروشگاه:

محصولات به طور مستقیم به فروشگاه‌های خرده فروشی تحویل داده می‌شوند.

در [۴۱] نویسندگان، مشکلات بوجود آمده در سمت عرضه و تقاضا را برای OOSهای مکرر گردآوری کرده‌اند. آنها در صورت مواجهه مکرر با شرایط OOS، یک برخورد منفی با تولید کننده و خرده‌فروش را پیشنهاد کردند. علاوه بر این، آنها دریافتند که از سمت عرضه، خرده‌فروش و تولیدکننده، می‌توان از استراتژی توزیع مستقیم در هنگام مواجهه با OOSهای مکرر بهره‌مند شد. با این حال، استراتژی توزیع مستقیم مستلزم صرف هزینه‌های بالاتر در مقایسه با استراتژی توزیع سنتی است. جنبه دیگر مطالعه مذکور بر مطالعه رابطه بین واحد انبارداری (SKU) واحد انبارداری[۲] و عملکرد OOS، تمرکز دارد. در [۴۲] مطرح شد که یکی از علت‌های عملکرد منفی در OOS، وجود اقلام با سرعت فروش بالا است. از سوی دیگر، اتوماسیون فرآیند سفارش کالا منجر به کاهش تعداد کالاهای OOS نیز شده‌است. همچنین موجودیت کالا به پیشبینی‌های مربوطه وابسته است.

[۴۳، ۴۴] کاربرد سیاست‌های مدیریت موجودی پلکانی[۳] را برای کاهش خطای پیش‌بینی سفارش‌های کوتاه‌مدت پیشنهاد کرده‌اند. در [۴۵] استفاده از یک شبکه عصبی مصنوعی ترکیبی را برای توسعه پیش‌بینی فروش غذای تازه پیشنهاد شده است. پیشبینی‌های این مدل، به تشخیص اینکه آیا در فروشگاه محصولات ناکافی هستند یا بیش از حد نیاز وجود دارند کمک می‌کند.

[2] Stock-Keeping Unit
[3] Echelon Inventory

[۴۶] بر روابط متقابل بین زنجیره تامین خرده‌فروشی و متغیرهای بازاریابی آنها تمرکز دارد. در مقاله مذکور آن‌ها اذعان دارند که پکیج‌های بزرگتر تعداد دفعات بازپرسازی بازپرسازی[۴] را کاهش می‌دهد که در نتیجه آن، احتمال رویداد OOS کاهش می‌یابد. یکی دیگر از نتایج مطالعه مذکور نشان می‌دهد که ترکیب بازاریابی خرده‌فروشی و تصمیمات مدیریت زنجیره تامین برای پایداری شرکت‌های مشتری‌مدار مانند خرده‌فروشان بسیار مهم است. با توجه به عوامل مختلف مرتبط با وقوع OOS [۳] یک رویکرد طراحی و اجرای مشارکتی را برای کاهش مداوم رویدادهای OOS پیشنهاد می‌دهد. این رویکرد شامل پارامترهای متنوعی مانند داده‌های سمت تقاضا، چرخه بازپرسازی و سیستم بسته‌بندی ارائه می‌دهند. نویسندگان این مقاله همچنین اظهار داشتند که ماهیت چندگانه رویدادهای OOS بر شرکت‌های مختلف در زنجیره غذایی از جمله تولیدکنندگان، خرده‌فروشان و فروشگاه‌ها تأثیر می‌گذارد.

فروشگاه‌های خرده‌فروشی در طول هفته با رنج‌های متفاوتی از کالاهای OOS روبرو هستند و حتی در ساعاتی از همان روز این نرخ‌ها متغیرند از دست رفتن درآمد ناشی از قفسه‌های خالی در فروشگاه تقریبا ۳/۹ درصد بازار کل را شامل می‌شود که از نظر ارزش پولی بیش از چند میلیارد یورو در سال بسته به اندازه بازار می‌باشد. [۴۷]

۲-۳ روش‌های سنتی و فیزیکی شناسایی و پیش‌بینی رویدادهای عدم موجودی کالا

بیشتر تحقیقات در مورد تشخیص OOS از دیدگاه خرده‌فروشی انجام شده است.

[۴۸] استراتژي گام‌های تعادل صفر را ارائه کرده‌است. در این استراتژی، کارکنان فروشگاه به‌صورت دوره‌ای برای بررسی موجودی انبارها (ممیزی های فیزیکی) قدم می‌زنند. [۴۹] یک سیتم RFID را در یک خرده‌فروشی برای نظارت بر موجودیت کالا در قفسه‌ها پیاده‌سازی کرده‌است.

در [۲۱، ۲۳] از داده‌های نقطه‌فروش (POS) نقطه‌فروش[۵] که به‌طور مستقیم از یک سیستم اطلاعات خرده‌فروشی برای تشخیص وقوع OOS به دست آمده، استفاده شده است. هزینه بالای تمام شده در حسابرسی‌های فیزیکی

[4] Replenishment
[5] Point Of Sale

حسابرسی‌های فیزیکی[۶] و RFID، برای شناسایی OOS وجه مشترک آنهاست.

تشخیص رویدادهای OOS بر اساس داده‌های POS (داده‌محور) مزایای مهمی مانند کاهش شدت بار فیزیکی کار و خطای انسانی در اندازه گیری دارد [۴۳] که این روش را مقیاس پذیر می‌کند و از نظر هزینه کلی تمام شده کارآمدتر است. برخی از چالش‌های این روش دسترسی به داده‌های POS و تعیین اینکه کدام متغیرها با وقوع OOS و توسعه مکانیزم‌هایی با عملکرد بهینه در شناسایی آنهاست می‌باشد. معمولا داده‌های POS در فروش محصولات و ثبت لیست اموال مورد استفاده قرار می‌گیرند.

مطالعات اندکی در مورد تشخیص OOS از دیدگاه تولیدکنندگان وجود دارد. در [۲۲] نویسندگان با یک تولیدکننده محصول و ارائه‌دهنده خدمات خرده‌فروشی با داده‌های به اشتراک گذاشته شده توسط خرده‌فروشی برای شناسایی رویدادهای OOS و تصحیح آنها با حسابرسی‌های فیزیکی همکاری کرده‌اند.

با توجه به تحقیقات انجام شده در زمینه OOS ایجاد شده در سال های اخیر، علم و آگاهی نسبت به عواقب و عوامل بوجود آمدن آنها افزایش یافته است. این امر منجر به توسعه زمینه تحقیقاتی در زمینه تشخیص و یا پیش‌بینی وقوع OOS شده است. در [۴۳] سه روش اصلی برای اندازه‌گیری OOS ارائه شده‌است:

- روش حسابرسی دستی

رویکرد سنتی است که در آن یک حسابرس به دنبال "حفره" های ایجاد شده توسط محصولاتی است که در قفسه توسط مصرف‌کنندگان قابل مشاهده نیستند.

- روش تخمین فروش POS

در این روش از داده‌های نقطه‌فروش (داده های POS) برای پیش‌بینی میزان درآمد از دست رفته ناشی از OOS استفاده می‌شود.

- روش تجمیع لیست موجودی دائمی

در این روش از داده‌های ثبت دائم موجودی (PI) ثبت دائم موجودی[۷] استفاده می‌شود. روش PI خریدها را پیگیری می‌کند بدین صورت که اگر تعداد فروش‌ها برابر با صفر ثبت شود مورد OOS خواهد بود.

[6] Physical Audit
[7] Perpetual Inventory

۲-۴ محدودیت‌های روش‌های سنتی و فیزیکی تشخیص و پیش‌بینی رخدادهای عدم موجودی

اخیراً فناوری‌ها و ابزارهای جدیدی برای اندازه‌گیری و تشخیص OOS مانند RFID، سیستم‌های شناسایی خودکار از طریق شناسایی تصویر، مدل‌های پیش‌بینی تصادفی، و روش‌های یادگیری ماشین، طراحی شده‌اند. همانطور که قبلا ذکر شد، [۴۸] روش قدم‌زنی تعادل صفر[8] را ارائه کرد، که در آن کارکنان به طور منظم در طبقات فروشگاه قدم می‌زنند تا به‌دنبال کالاهای OOS بگردند. برخی از معایب این روش، این است که خرده‌فروشان اغلب تعداد کمی از اقلام را انتخاب می‌کنند و به‌دلیل محدودیت‌های مالی و کمبود تعداد کارکنان لازم، برای انجام این حسابرسی‌ها زمان‌های خاصی تعیین می‌کنند. به‌عبارتی هزینه این اندازه‌گیری‌هایی که بصورت پیوسته باید انجام می‌شود بسیار زیاد است. در حال حاضر هر کدام از خرده‌فروش یا تولیدکننده، ممکن است اندازه OOS را شروع و هدایت کنند. مرجع [۴۹] طی یک پروژه آزمایشی‌ای، زیرساخت FRID ای را پیشنهاد کرده‌است که ذخیره کردن داده‌های آنی،فراهم آوردن امکان ردیابی OOS و نمایان کردن مزایای مدیریت موجودی کالا را در یک فروشگاه خرده فروشی ارائه می‌دهد.

با این‌حال، برخی از نویسندگان محدودیت‌هایی را در از این فناوری بیان کرده‌اند [۵۰]. به این موضوع اشاره کرده که با توجه به محدودیت‌های فیزیکی RFID، این رویکرد پتانسیل تولید خطای منفی کاذب (FN) خطای منفی کاذب[9] را دارد. از اینرو برای شناسایی دقیق‌تر OOS هر محصول باید شناسایی و بر آن نظارت شود.

پیاده‌سازی‌های RFID در سطح اقلام نیز بسیار پرهزینه هستند. به‌طوریکه این پیاده‌سازی‌ها هنوز از فناوری‌های تشخیص مرسوم، گرانتر می‌باشند [۵۱]. همین دلیل است که [۵۲] با تعجب می‌پرسند: «آیا باید خرده‌فروشی‌ها بر روی همه اقلام از جمله اقلام «ارزان» هم برچسب RFID بزنند؟».

در زمینه شناسایی تصویر، [۵۳] برخی مشکلات در زمینه تشخیص اقلام قفسه‌های فروشگاه توسط RFID را عنوان کرده‌است. به‌عنوان مثال به‌دلیل تغییرات مرتبط با اندازه، وضعیت قرارگیری، چشم‌انداز و... وضعیت ظاهری کالا بسیار متغیر است.

بر اساس پیش‌بینی داده‌های POS اسکنرهای فروشگاه (به عنوان مثال، آنهایی که در صندوق فروشگاه‌ها وجود داشته و فروش را ثبت می‌کنند) و داده‌های بدست آمده از سیستم اطلاعات خرده‌فروشی، در سیستم های

[8] Zero Balance Walk
[9] False Negative

تشخیص OOS استفاده می شوند. این رویکرد هنگامی‌که توسط حسابرسی‌های دستی آزمایش می شود، دارای دقت ۸۵-٪۹۰ در شناسایی موارد مثبت واقعی خطای مثبت واقعی[10] (مواردی که به درستی به عنوان OOS علامت گذاری شده بودند) می‌باشند. با این حال، نرخ بالای FN یک ایراد این مدل است (اقلامی که OOS هستند اما به عنوان OOS شناسایی نمی شوند). از دیگر محدودیت‌های این رویکرد می‌توان به عدم مدیریت SKU های با گردش مالی کم و نیاز به داده‌های POS دقیق اشاره کرد. با این وجود، این روش یک راه حل مناسب به‌منظور حسابرسی دستی مؤثر است، زیرا کارایی نیروی انسانی را بهبود بخشیده و میزان خطای نیروی انسانی را کاهش می دهد.

استفاده از داده‌های لیست ثبت موجودی دارای محدودیت‌هایی است، مخصوصاً عدم دقت در آن. [۵۴] نشان داد که تاثیر منفی عدم دقت در لیست ثبت موجودی، منجر به برآوردهای اشتباه درباره OOS خواهد شد. تقریب های احتمالاتی و آمار اخیرا برای تشخیص OOS پیشنهاد شده است.

۲-۵ روش‌های یادگیری ماشین و هوش‌مصنوعی جهت پیش‌بینی رخداد عدم موجودی کالا

[۵۵] استفاده از احتمال بیزی را برای تصحیح "فانتوم عدم موجودی" با استفاده از تاریخ ثبت داده‌هااز سوابق موجودی و فروش، به همراه سیاست های بازرسی موجودی ارائه کرد. نویسندگان همچنین "فرآیند انقباض برنولی" و یک آستانه را برای دوره های فروش صفر متوالی در نظر گرفتند.

مرجع [۲۲] از داده های POS ثبتی، برای شناسایی کالاهای OOS احتمالی قفسه در یک مطالعه موردی استفاده کرد. این امر برای فرستادن حسابرسان به فروشگاه‌ها به‌منظور اصلاح هر گونه مشکل OOS استفاده می شود. مقاله مذکور بیشتر توضیح می دهد که چگونه یک تولیدکننده باید با ارائه‌دهنده خدمات به‌منظور اجرای حسابرسی‌های فیزیکی خارجی برای افزایش موجودیت در قفسه همکاری کند. در مطالعه مذکور تقاضا با استفاده از یک مدل دو جمله‌ای منفی توصیف شده است و تشخیص OOS با استفاده از فروش صفر متوالی انجام می شود. به طور مشابه، [۵۶] یک استراتژی حسابرسی بهینه، با استفاده از اندازه‌گیری‌های متوالی فروش صفر در داده های POS پیشنهاد می‌دهد.

[10] True Positive

[۲۳] وقوع OOS از طریق استفاده از یک مدل پنهان مارکوف (HMM) را پیشنهاد می‌دهد. در این مدل، یک شرط OOS توسط یکی از حالت‌های مخفی نشان داده می‌شود. حالت‌های دیگر تغییرات ایجاد شده در الگوهای مربوط به تقاضا را تشخیص می دهند.همچنین داده‌های POS برای کالیبره کردن مدل و اعتبارسنجی استفاده می شود. [۵۷]. مرجع [۱۹] یکی از اولین مطالعاتی است که از تکنیک‌های یادگیری ماشین و داده‌های به‌دست‌آمده از خرده‌فروشان برای شناسایی محصولات قفسه و یک سیستم تصمیم پشتیبان که به طور خودکار OOS ها را در قفسه‌ها شناسایی می کند استفاده کرده است. بطور کلی داده‌ها به دو دسته تقسیم می شوند:

- داده‌های به‌دست آمده از سیستم اطلاعات داخلی خرده‌فروشی (POS-Data) که شامل مجموعه محصولات، دسته‌بندی محصولات، کاتالوگ محصولات و سفارشات می‌شوند.
- داده‌های به‌دست آمده از حسابرسی‌های فیزیکی

پیشبینی OOSها به‌عنوان یک مسئله طبقه‌بندی باینری مطرح می شود. با استفاده از داده‌های به‌دست آمده از سیستم اطلاعاتی، ویژگی های مختلفی محاسبه شده که بخشی از متغیرهای مستقل مسئله هستند. متغیر وابسته یا خروجی به‌دست آمده از حسابرسی‌های فیزیکی،به‌عنوان کلاس Exists (محصول موجود در قفسه) و یا OOS در نظر گرفته می‌شود. [۱۹] چالش‌های بزرگی را برای بهبود دقت در بده‌بستان بده‌بستان[11] کردن بین دقت داده‌ها و پشتبانی تعداد داده‌ها هر کلاس مطرح کرده‌است.

[۲۰] دو روش برای توسعه یک سیستم تصمیم‌گیری ارائه کرده‌است که به مدیران فروشگاه امکان میزان کاهش نرخ کالاهای OOS را می‌دهد. یکی از این روش‌ها از مدل‌های کنترل موجودی استفاده کرده و دیگری روش طبقه‌بندی است. یکی از نتایج مهم مطالعه مذکور این است که توسعه ابزارهای تصمیم‌گیری در هر دو روش می‌تواند نقش کلیدی برای بهبود سیستم پشتیبانی تصمیم را ایفا کند. نتایج مهم این مطالعه با تمرکز بر روی متود معرفی شده در [۲۱، ۱۹] بدست آمده که الگوریتم‌های طبقه‌بندی مختلف برای طبقه‌بندی اقلام با برچسب OOS را با هم مقایسه کرده‌اند.

داده‌های واقعی دارای ایراد اقلیت تعداد در کلاس OOS هستند که موجب عدم‌تناسب تعداد کلاس‌ها برای الگوریتم طبقه‌بندی می‌باشد. برای غلبه بر این مشکل و بهبود عملکرد طبقه‌بندها، از رویکرد یادگیری گروهی یادگیری گروهی[12] استفاده می‌شود. یک محدودیت مدل‌هایی که تا به الان از این رویکرد استفاده کرده‌اند این

[11] Trade-Off
[12] Ensemble Learning

است که سیستم پیشبینی‌کننده بخش بزرگی از اقلام OOS را شناسایی نکرده‌اند. مشابه آنچه در مطالعات قبلی مورد بررسی گرفت، شناخت عوامل مؤثر بر دسترسی محصول برای توسعه ابزارهایی با ظرفیت تشخیص بهتر، حائز اهمیت است. زنجیره تأمینی که تضمین می‌کند دسترسی داخل قفسه (OSA) دسترسی داخل قفسه[۱۳] (یا رویدادهای OOS را کاهش می دهد) به‌درستی عمل کند، می‌تواند به عنوان بهبود عملیات عمل کند [۳۶، ۵۸، ۵۹].

با توجه به اینکه مسئله OOS با دیدگاه‌های مختلف در مطالعات مورد بررسی قرار گرفته‌اند، برای سادگی در تحلیل، [۳] به این مسئله از چهار دیدگاه نگاه می‌کند. اولین گروه از مقالات بر روی کشف تأثیر رویدادهای OOS بر روی مشتریان است که معمولاً جایگزینی (با برندی یکسان یا متفاوت) برای کالای OOS برمی‌گزینند یا تاخیر در خریدشان ایجاد می‌کنند و یا به سراغ فروشگاهی دیگر می‌روند [۳۴، ۶۰، ۶۱، ۶۲، ۶۳]. چنین رویدادهایی نه‌تنها دارای تأثیر منفی بر رضایت مشتری و سودآوری تولیدکنندگان و خرده‌فروشان دارد بلکه می‌توانند یک اثر آبشاری[۱۴] بر روی گزینه‌های جایگزین کالاهای ناموجود در قفسه‌ها گذاشته و حتی منجر به رویدادهای OOS جدید شوند. با این‌حال نویسندگان مقالات [۶۴، ۳۳] که جنبه های مثبت رویدادهای OOS را در صورت انجام برنامه‌ریزی برای آن‌ها، شناسایی کرده‌اند. به‌طور مثال برنامه‌ریزی مناسب می‌تواند فروش را از یک محصول با سود کم به محصولات دیگر با حاشیه سود بالاتر تبدیل کند یا حتی اجتناب از هر گونه تلفات یا تولید ضایعات را زمانی که هر دو محصولات قدیمی و جدید موجود در قفسه‌ها (مخصوصا برای محصولات فاسد شدنی) را رقم بزنند.

تمرکز گروه دوم مطالعاتی را بر روی تحقیق در مورد کشف رویدادهای OOS است که پیش‌تر به آن اشاره شد.

مقالات گروه سوم پیشنهاداتی را به‌منظور بهبود مدیریت قفسه در فروشگاه با استفاده از رویکرد مبتنی بر تحلیل و صحت‌سنجی، قفسه‌ها ارائه کرده است.

با توجه به ابعاد بسته‌بندی یک محصول، حداکثر تعداد واحدهایی که عملا مناسب هسند را می‌توان تعیین کرد. اگرچه باید همیشه در نظر گرفت که این محصولات به‌درستی روبه‌روی مشتری قرار گیرند. بنابراین، گروه سوم از این نوع مقالات نه‌تنها به‌دنبال کاهش تعداد کالاهای OOS بلکه همچنین در بهبود معیارهای مورد استفاده برای انتخاب مناسب ترین روش چینش محصولات در قفسه‌ها پیشنهاداتی را ارائه می‌دهند. [۶۵، ۶۶، ۶۷، ۶۸]

[13] On-Shelf Availability
[14] Cascade

در اکثر این مقالات، رویکردهای سازماندهی شده‌ای را برای کاهش سیستماتیک رویدادهای OOS ارائه نمی‌دهند که این امر یکی از انگیزه‌های اصلی مقاله [۳] می‌باشد.

دیدگاه نهایی که در مقالات مروری با آن مواجه می‌شویم، مسئله تشخیص و اندازه‌گیری عینی رویدادهای OOS است. برخی از نویسندگان مقالات مذکور، [۱۹، ۳۳] تحلیل‌های بیشتری را در مورد نحوه استفاده از شاخص های عملکرد کلیدی KPI شاخص های عملکرد کلیدی[۱۵] که با رویدادهای OOS مرتبط هستند را به‌عنوان روش برای بهبود فرآیند تشخیص و اندازه‌گیری این رویدادها ارائه کرده‌اند.

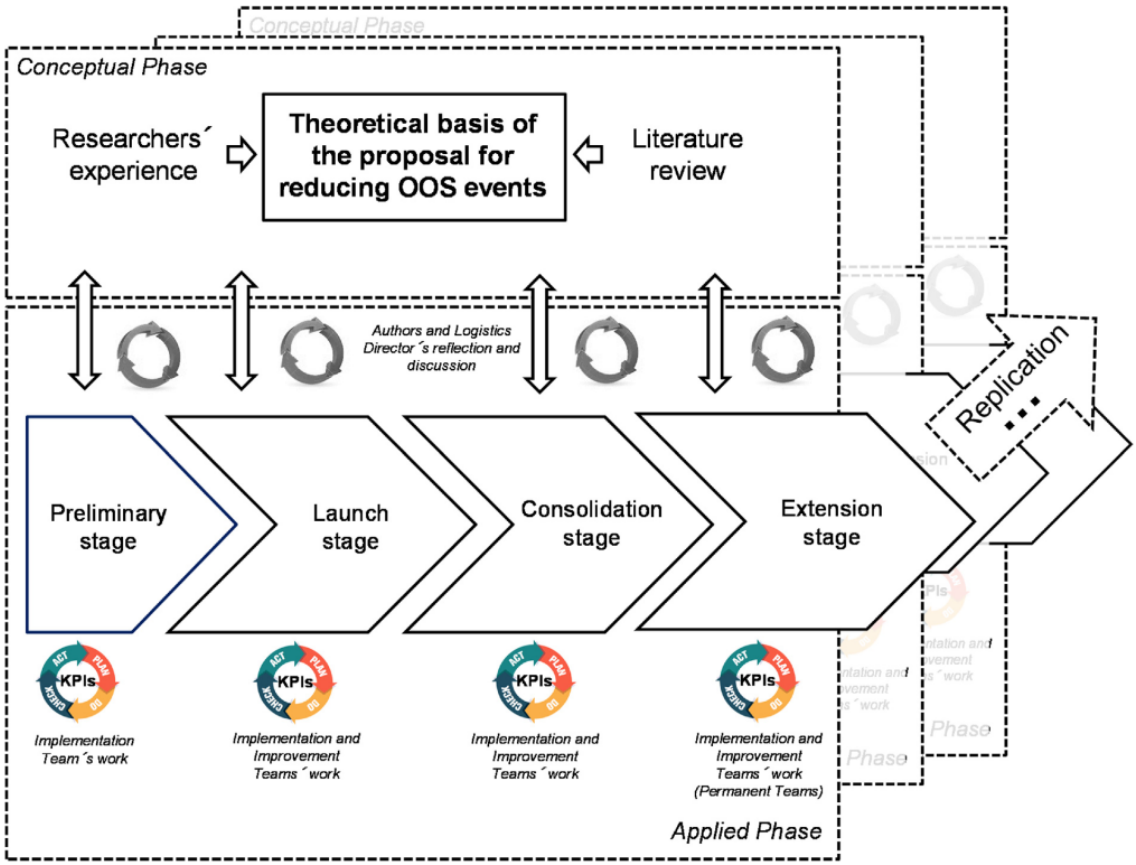
علی‌رغم استفاده از روش تشخیص رویدادهای OOS حتی در صورت اتمام موجودی در زمان پیشبینی شده، این روش ممکن است بهترین روش مناسب برای تشخیص رویداد OOS نباشد و وضعیت OOS تا زمانی ادامه خواهد یافت که محصول جدیدی روی قفسه قرار داده شود. (اگر فروشگاه موجودی دارد از اتاق پشتی یا از انبار خرده‌فروش فراهم شود). در این مرحله، مسئله فقط کشف اینکه آیا OOS وجود دارد یا نه نیست بلکه مدت زمان ادامه آن نیز اهمیت دارد. مطالعه [۲۲] پتانسیل نظارت بر رویدادهای OOS در قفسه‌ها را با استفاده از یک روش کمّی و با توسل به ممیزی‌های خارجی که به‌صورت دوره‌ای هستند معرفی می‌کنند.

مسئله اصلی مقاله [۳] تجزیه‌وتحلیل امکان طراحی و اجرای یک روش مشارکتی است که به‌طور سیستماتیک به دنبال کاهش رویدادهای OOS، حفظ تصویری اقلام یک فروشگاه و بهبود فرآیندهای ترابری با استفاده از رویکرد اقدام پژوهی است. طبق [۶۹، ۷۰، ۷۱، ۷۲] ارجاعات کمی به پذیرش رویکرد اقدام‌پژوهی در زمینه مدیریت زنجیره تأمین وجود دارد، به‌ویژه در جایی که زنجیره به فروشگاه می‌رسد. در عین حال، کمبود تحقیقات کاربردی را در مورد نقشی که افراد در زنجیره تأمین ایفا می‌کنند را معرفی می‌کند. که با استقرار سیستم‌های مشارکتی ساختاریافته در چارچوب بهبود پیاده‌سازی مستمر مرتبط است. ناسلوند و همکارانش، ۳ جنبه کلیدی که را در این خصوص را بیان می‌کنند:

- استقرار یک سیستم دقیق، ساختاریافته و مستند برای همکاری بین شرکت (یا سازمان) و محققان
- همکاری قابل توجه پژوهش در تولید دانش علمی
- علاقه خود شرکت یا سازمان در به دست آوردن نتایج حاصله از تحقیقات.

علاوه بر این، شرکت‌ها (به‌ویژه خرده‌فروشان) و زنجیره‌های‌تأمین نیز می‌توانند به روشی دست یابند که می‌تواند به کاهش رویدادهای OOS کمک کند.

[15] Key Performance Indicators



شکل ۲-۱: روش مشارکتی پیشنهاد شده در [۳]

به منظور اعمال رویکرد اقدام‌پژوهی، [۳] روشی را پیشنهاد می‌کند که با ساختار دوفازی تطبیق داده شده است. این متود در شکل ۲-۱ قابل مشاهده است.

اجرای روش مشارکتی منجر به کاهش ۳۰ درصدی رویدادهای OOS در خرده‌فروشی شده است.

همانطور که پیش‌تر اشاره شد، رویدادهای OOS ارتباط نزدیکی با مسئله عدم دقت در لیست ثبت موجودی دارد زیرا در موجودیت یک محصول در قفسه به‌طور مستقیم با کیفیت نحوه مدیریت در ثبت موجودی توسط خرده‌فروش در ارتباط است. در واقع، عدم دقت در لسیت ثبت موجودی، یک دلیل مهم برای رویدادهای OOS می‌باشد. [۷۳] یک ابزار مدیریت موجودی راتوسعه داده که با استفاده از آن عدم دقت ثبت موجودی را با استفاده از ابزار باوربیزی BB باور بیزی[۱۶] محاسبه می‌کند.

[16] Bayesian Belief

اخیراً از مدل مارکوف پنهان (HMM) مدل مارکوف مخفی[۱۷]

برای مدل‌سازی سیستم ثبت‌موجودی با در نظر گرفتن داده‌های معاملاتی استفاده شده‌است. [۷۴] یک رویکرد تحلیلی برای مدیریت موجودی را با استفاده از HMM پیشنهاد داده است. در مقاله مذکور، شرایط مورد نیاز برای کنترل بهینه موجودی در زمان مشاهده تقاضا بررسی شده‌اند.

[۷۵] خط مشی‌هایی برای سیستم‌هایی با عدم دقت در ثبت موجودی را پیشنهاد کرده است. در مقاله مذکور یک خط مشی مبتنی بر سیستم ثبت‌موجودی را پیشنهاد شده است که باعث می شود ممیزی زمانی که سوابق موجودی کمتر از حد معین باشد انجام شود.

[۵۵] دو مدل تصمیم مارکوف مارکوف مشاهده پذیر جزئی[۱۸] را برای کاهش تقاضا و اجرای مجدد پیشنهاد داده است. در واقع نویسنده یک مدل BB را درباره سطح موجودیت واقعی مبتنی بر سیستم ثبت موجودی و داده‌های فروش ثبتی، پیشنهاد داده‌است. در این رویکردهای BB برای تکمیل مدل ثبت موجودی، حالات پنهان با تعداد واحدهای موجودیت‌های مشاهده نشده مطابقت دارند. بنابراین نویسندگان سعی بر توصیف تکاملات مشاهده شده در تقاضا را به صورت تجربی ندارند زیرا تعداد مشاهده‌ها بسیار زیاد بوده و متغیرهای کمکی بالقوه دخیل در توصیف انتقالی بین حالات، می توانند تخمین HMM را پیچیده کنند. در واقع تعداد زیادی از حالت‌ها، با وجود تحقیقات گسترده در بررسی سیاست‌های تشخیص بهینه رویدادهای OOS، در قفسه، کمبود مطالعات تجربی محسوس است. با این حال، چند استثنا در این‌باره وجود دارد. [۱۹] یک سیستم پشتیبان تصمیمگیر قانون‌مند را به‌منظور تشخیص OOSهای مبتنی بر داده‌های فروش توسعه داد.

درحالی‌که [۷۶] یک الگوریتم یادگیری ماشینی پیشنهاد کرد که توسط تاریخچه داده‌های OOS آموزش داده شده و در هر زمان حالت آنی قفسه را طبقه‌بندی می‌کند. در این دو مورد، الگوریتم‌های توسعه داده شده که به‌صورت نظارت‌شده هستند. در [۲۳] روش نظارت‌شده‌ای معرفی شده که تنها به داده‌های POS نیاز دارد تا امکان نظارت بر تعداد زیادی از محصولات را فراهم کند. برخلاف [۲۲] که از یک مدل دوجمله‌ای برای توصیف تقاضا و رویکرد پردازش-کنترل برای دریافت هشدارها استفاده می‌کند در [۲۳] با استفاده از HMMای که تکمیل فروش‌ها و زنگ‌هشدارها را بصورت آنی، در حین رسیدن سیستم به حالت OOS می‌رسد، مدل می‌کند. علاوه بر این در مقاله مذکور، طبق تحقیقات قبلی،تقاضا به‌عنوان یک فرآیند زیربنایی مارکوف که حالات جهان را نشان می‌دهد مدل‌سازی می‌شود. [۷۷، ۷۸، ۷۹] به‌علاوه از مزایای HMM به‌منظور کشف OOSهای مشاهده نشده

[17] Hidden Markov Model
[18] Partially Observable Markov

استفاده می‌شود.

مرجع [۲۳] از داده‌های ثبتی و بازرسی بصری برای کالیبره کردن مدل استفاده کرده و به‌منظور اعتبارسنجی، برای هر محصول به طور مستقل برآورد سلسله مراتبی بیز را برای ۵۰۰۰۰۰ اجرا کرده است. عملکرد این روش توسط نظارت بر نمودارهای ردیابی خروجی ارزیابی شده‌است. همچنین خروجی تخمینی پارامترهای مدل برای هر محصول کمتر از دو ساعت به طول می‌انجامد و با توجه به اینکه این تخمین برای هر محصول مستقل است این امر می‌تواند به‌صورت موازی انجام شود. پس از برآورد مدل، استنباط حالت قفسه برای همه محصولات و فروشگاه‌ها در طول دوره مدت زمان اعتبارسنجی کمتر از یک دقیقه می‌باشد.

۲-۶ ترکیب طبقه‌بند مجتمع ایستا برای پیش‌بینی رخداد عدم موجودی

[۱۳] در سعی بر حل مشکلات مربوط به شناسایی کالاهای OOS کرده است. این مشکلات بر مسائل خرده‌فروشی و همچنین تولیدکنندگان سایه افکنده‌اند. همانطور که قبلا ذکر شد، پیشنهادات مختلفی برای تشخیص یا پیش‌بینی وقوع OOS مطرح شده‌است [۲۱، ۱۹، ۲۳]. در [۱۳] برای حل این مشکلات، استفاده از یک مدل مبتنی بر الگوریتم‌های یادگیری ماشین برای طبقه‌بندی و سیستم ترکیب طبقه‌بند مجتمع را پیشنهاد کرده‌است. مرحله اول، مدل‌سازی مطالعه موردی و تبدیل مسئله واقعی OOS به یک مدل ریاضیاتی است که توسط یک الگوریتم طبقه‌بندی قابل حل باشد. مطالعه موردی ارائه شده در مطالعه مذکور در یک شرکت تولیدی که کالاهای بسته‌بندی شده عرضه می‌کند انجام می شود. درواقع شناسایی OOS توسط داده‌های POS ثبت شده انجام می‌پذیرد. در این مقاله پس از انتخاب متغیرهای مناسب از درون داده‌ها، یک تحلیل اکتشافی داده انجام شده، سپس الگوریتم‌های طبقه‌بندی مختلف مورد آزمایش قرار گرفته و عملکرد آنها را با استفاده از معیارهای دقت دقت[19] ، صحت[20] ، حساسیت[21] ، ویژگی[22] و معیار اف[23] مقایسه می‌کنند.

نتایج به‌دست آمده ارائه شده است در جدول ۲-۱ قابل مشاهده است. تکرار تصادفی انجام شد که میانگین و انحراف استاندارد برای هر یک از معیارهای عملکرد. با توجه به نتایج به‌دست آمده، الگوریتم ترکیب طبقه‌بند

[19] Precision
[20] Accuracy
[21] Sensivity
[22] Specificity
[23] F-Score

جدول ۲-۱: نتایج مربوط به مقاله [۱۳]

مدل	صحت	حساسیت	ویژگی	معیار F
جنگل‌تصادفی	۰/۹۱۳۷۱	۰/۲۴۱۶۴	۰/۹۸۹۹۱	۰/۳۶۳۸۱
رگرسیون لجستیک	۰/۸۹۷۸۶	۰/۰۰۹۹۸	۰/۹۹۸۳۴	۰/۰۱۹۳۹
درخت تصمیم	۰/۸۹۷۱۹	۰/۱۰۰۷۶	۰/۹۸۷۳۹	۰/۱۶۴۲۵
بیز ساده	۰/۸۴۵۳۷	۰/۱۹۴۸۴	۰/۹۱۹۰۹	۰/۲۰۸۲۰
ماشین‌های بردار پشتیبان	۰/۸۹۷۶۸	۰/۰۲۰۳۱	۰/۹۹۷۰۱	۰/۰۳۸۹۳
شبکه عصبی	۰/۸۹۹۴۳	۰/۰۷۳۷۱	۰/۹۹۲۹۳	۰/۱۳۰۰۳
ترکیب طبقه‌بند مجتمع	۰/۹۱۶۶۷	۰/۳۱۴۶۷	۰/۹۸۴۹۷	۰/۴۳۵۶۱

مجتمع بهترین عملکرد را نسبت به سایر مدل‌ها دارد. معیار صحت آن بیشترین میزان را داشته‌است و همچنین میزان سبک‌سنگینی بین مهمترین معیارها یعنی حساسیت و معیار F، بهترین بوده‌است.

۲-۷ جمع‌بندی

مدیریت موجودی کالا به ارائه‌ی تجربه‌ی بی‌نقص مشتری، بهبود جریان نقدی و به حداکثر رساندن سود، جلوگیری از انقباض موجودی و اتلاف و هم‌چنین بهینه‌سازی فرآیند تکمیل کمک می‌کند و همه‌ی این‌ها مدیریت موجودی را به وظیفه‌ای ضروری برای خرده‌فروشی‌ها تبدیل می‌کنند.در برخی از مشاغل موجودی شامل مواد خام و اجزا می‌شود، در حالی که برخی دیگر ممکن است فقط با اقلام موجودی آماده برای فروش سروکار داشته‌باشند. در هر صورت، مدیریت موجودی به تعادل داشتن بستگی دارد، یعنی داشتن مقدار مناسبی کالای موجود در مکان مناسب و در زمان مناسب. شناسایی وضعیت‌های عدم موجودی از مهم‌ترین امور به منظور مدیریت موجودی کالا می‌باشد. روش‌های استخراج شده از مقالات مروری که در این فصل به آن‌ها پرداخته شد،به منظور شناسایی دقیق وضعیت‌های عدم موجودی کالا و بهبود سیستم‌های پیش‌بینی این وضعیت‌هابوده که در نتیجه آن با تصمیم‌گیری‌های مناسب می‌توان درصدد کاهش تأثیرهای منفی آن‌ها برآییم. همانطورکه مشاهده کردیم، حسابرسی‌های دستی و سنّتی روند بسیار کند به همراه خطای انسانی دارند. روش پیاده‌روی تعادل صفر که در مقالات مرووی ذکر شد دارای معایبی است از جمله اینکه خرده‌فروشان اغلب تعداد کمی از اقلام را انتخاب می‌کنند و به‌دلیل محدودیت‌های مالی و کمبود تعداد کارکنان لازم، برای انجام حسابرسی‌ها زمان‌های خاصی را

تعیین می‌کنند در نتیجه هزینه تمام شده،برای این حسابرسی هنگفت خواهد بود. سیستم‌های فیزیکی و RFID مشکل هزینه بالای نصب و نگه‌داری را داشتند.همچنین روش‌های یادگیری‌ماشینی و هوش‌مصنوعی که به منظور پیشبینی وضعیت‌های عدم‌موجودی، تا به‌حال بکار گرفته‌شده اند به‌دلیل نوع داده‌های مربوط به وضعیت‌های عدم موجودی دارای عملکرد مناسبی نمی‌باشند. مشاهده کردیم که روش ترکیب طبقه‌بند مجتمع نسبت به سایر روش‌های هوش‌مصنوعی و یادگیری ماشینی دارای عملکرد مناسب‌تری بوده است. ایراد این روش این است که این یادگیری ترکیبی، بصورت ایستا بوده و به توزیع داده‌ها در فضای ویژگی توجهی ندارد. روش پیشنهادی این سند که در فصول آتی به آن پرداخته خواهد شد یک روش طبقه‌بند مجتمع پویا است که با توجه به ساختار داده و توزیع آن در فضای ویژگی، عمل پیش‌بینی وضعیت‌های عدم موجودی را با عملکرد بهتری نسبت به سایر روش‌ها انجام می‌دهد.

فصل ۳

نحوه طراحی طبقه‌بند مجتمع پویا به‌منظور پیش‌بینی رخداد عدم موجودی کالا

۳-۱ مقدمه

در دو فصل گذشته به یکی از چالش‌های مهم در حوزه آنالیز کسب‌وکارها پرداخته شد. بر اساس مطالعات گذشته، مشخص شد که تشخیص و شناسایی وضعیت ناموجودی کالا (وضعیت OOS

در خرده‌فروشی، از مهم‌ترین عوامل تصمیم‌سازی‌های بهینه در مدیریت موجودی کالا می‌باشد. همچنین اشاره شد سیستم ترکیب طبقه‌بند مجتمع پویا که در این فصل به تشریح آن پرداخته خواهد شد از عملکرد مناسب‌تری نسبت به روش‌های سابق برخوردار است. در این فصل به روش جمع آوری داده‌های مورد استفاد در این پژوهش، روش‌های آماده‌سازی داده‌ها و همچنین مدل‌های مختلف یادگیری ماشین و مدل پیشنهادی پرداخته شده است.

۳-۲ روش جمع آوری مجموعه داده

مجموعه داده مورد استفاده در این پژوهش مربوط به اطلاعات قیمتی کالاهای دیجیتال می‌باشد. این مجموعه‌داده در سایت Kaggale موجود بوده و از پایگاه داده محصولاتی Datafiniti جمع‌آوری شده است.تیم Datafiniti زمینه کاملی برای تبدیل اطلاعات از وب و سایر منابع به داده‌های قابل استفاده و به‌موقع ایجاد کرده است. Datafiniti با هدف اصلی کمک به شرکت‌ها و کارهای پژوهشی برای دستیابی به داده‌های ارزشمند محصولات، مشاغل، املاک و ... فعالیت می کند. متغیرهای مستقل آن‌را اطلاعاتی مانند محدوده‌قیمتی، وزن، نام محصول، دسته‌بندی نوع کالا، سازنده کالا، برند تشکیل می‌دهند. قبل از طراحی، آموزش و آزمایش مدل مراحلی را باید طی کرد که بر روی داده‌ها انجام می‌شوند.

۳-۳ تحلیل توصیفی داده‌ها

تحقیق توصیفی، جامعه، شرایط و پدیده ها را به طور صحیح و سیستماتیک توصیف می کند. این نوع تحقیق به سوالاتی که شامل چه چیزی، چه زمانی، کجا، و چگونه هستند، پاسخ می دهد ولی برای پاسخ به سوالاتی که شامل چرا می گردند، کاربردی ندارد. برای پاسخ به این سوال از روش تحقیق آزمایشی استفاده می شود. در طراحی تحقیق توصیفی برای بررسی متغیرها انواع روش های کمّی و کیفی به کار گرفته می شود. برخلاف تحقیق آزمایشی، پژوهشگر متغیرها را کنترل یا دستکاری نمی‌کند، ولی آنها را مورد مشاهده و آزمون قرار می دهد.

۳-۴ پیش‌پردازش داده‌ها

در این مرحله داده‌ها برای ورود به مدل، آماده می‌شوند. پیش پردازش داده پردازش داده[1] به مراحلی گفته میشود که در آن داده ها برای داده کاوی آماده میشود. لازم به ذکر است که این مراحل جز مهم ترین گام ها در داده کاوی هستند. پس باید با دقت با مفهوم آن آشنا شد. در این بخش موارد مهمی که پیش پردازش داده‌ها در آن‌ها انجام می‌شود، بیان شده است. قبل از انجام موارد اصلی پیش‌پردازش، ابتدا در صورت نیاز بایستی بر روی مجموعه

[1]Data Preprocessing

داده‌ها، عمل دستکاری داده‌ها دستکاری داده[۲] را انجام که موارد آن بصورت زیر است:

- داده‌های تکراری: در صورتیکه داده داده‌های تکراری‌های تکراری درون مجموعه داده موجود باشد باید می‌توان داده اول را نگه داشته و سایر نمونه‌های تکراری را حذف نمود.
- تغییر اسم‌های طولانی و یا غیرمناسب متغیرها
- اضافه و یا حذف ویژگی‌های لازم: درصورتیکه متغیرهای بی‌اهمیتی که اطلاعات خاصی ندارند و حاوی اطلاعات یکسان برای تمامی نمونه‌ها هستند بایستی این متغیرها از داده‌ها حذف شوند. همچنین درصورتیکه نیاز باشد می‌توان از دل متغیرهای موجود در داده‌ها، متغیرهای جدیدی را استخراج نمود و به متغیرهای مجموعه داده اضافه کرد.
- اصلاح و یکسان‌سازی واحدهای اندازه‌گیری در متغیرها (در صورت نیاز)
- تبدیل متغیرهای وابسته‌ی کیفی به کمی و یا بالعکس.

پس از دستکاری داده‌ها بایستی اعمال اصلی پیش‌پردازشی بر روی داده‌ها را انجام داده شوند که در ادامه به آن‌ها پرداخته شده است.

۳-۴-۱ داده گمشده

وقتی با یک مجموعه داده کار می‌کنیم، امکان دارد به مشاهداتی برخوریم که یک یا چند متغیر یا ویژگی در آن، بدون مقدار باشد. این ایراد معمولا وقتی اتفاق می‌افتد که در هنگام جمع‌آوری داده‌ها دقت لازم وجود نداشته باشد. در این موارد می‌توان گفت که مشاهدات دارای «مقدار گمشده» داده گمشده[۳] می‌باشند. روش‌های متفاوتی برای مجموعه داده با مقدارهای گمشده وجود دارد که ممکن است هر یک از آن‌ها دارای معایب با مزایایی خاص خودشان باشند.

- حذف نمونه‌ها: در صورتی که مقادیر از دست رفته در برخی از متغیرها از نوع داده کاملا از دست یاشد به‌طوری‌که تمامی ویژگی‌های آن داده از دست رفت باشد و همچنین اگر تعداد مقادیر از دست رفته خیلی زیاد نباشد، می توان درعوض پر کردن و یا جایگزینی داده‌های گم شده، آن‌ها را به‌کلی حذف نمود.

[2]Data Manipulation
[3]Missing Value

- جایگزینی با ثوابت: در صورت تکرار زیاد این ثوابت، توزیع داده‌ها به‌هم ریخته و در یادگیری مشکل ایجاد می‌شود.
- جایگزینی با مقادیر نمونه‌های بعدی یا قبلی
- حذف ویژگی‌های از دست رفته (ستون): اگر یک ویژگی در تعداد زیادی از نمونه‌ها دارای مقدار گمشده باشد می‌توان آن‌را حذف نمود.
- روش‌های آماری جایگزینی مقادیر از دست رفته: با استفاده از روش‌های گوناگون آماری می‌توان مقدارهای ازدست رفته را با مقادیر جدید جایگزین نمود. جایگزینی با میانه، میانگین و یا مود در هر متغیر از روش‌های متداول پر کردن مقادیر گم‌شده است.
- روش‌های درونیابی جایگزینی مقادیر از دست رفته: برای داده‌های با متغیر پیوسته می‌توان با استفاده از روش‌هایی مانند رگرسیون و یا روش‌های مبتنی بر هوش‌مصنوعی مانند شبکه‌های بازگشتی مقادیر از دست‌رفته را جایگزین نمود.

۳-۴-۲ داده پرت

ممکن است برخی از نمونه‌ها دارای متغیر پرت باشند به این معنی که مقدار آن‌ها در محدوده مقادیر توزیع متغیر مربوطه نباشد و دارای مقدار بسیار زیاد یا بسیار کم‌تر نسبت به متغیر سایر نمونه‌ها باشد، در اینصورت باید این مقدار را مانند داده‌گمشده حذف و یا جایگزین کرد.

۳-۴-۳ نرمال‌سازی داده‌ها

نرمال‌سازی داده‌ها روشی برای یکنواخت کردن بازه مقادیر مربوط به متغیرهای مختلف پژوهش است و به بی‌مقیاس‌سازی داده‌ها نیز معروف است. اگر واحد سنجش متغیرهای موردمطالعه متنوع باشد با استفاده از روش‌های نرمال‌سازی می‌توان داده‌ها را بی‌مقیاس کرد. روش‌های مختلف استانداردسازی داده‌ها بصورت زیر است:

- روش خطی: یک روش ساده برای نرمال کردن اعداد توسط ساعتی مطرح شده که به محاسبه بردار ویژه نیز معروف شده است. در این روش کافی است هر مقدار در یک مجموعه بر مجموع عناصر آن مجموعه

تقسیم شود. در این‌صورت جمع کل عناصر پس از نرمال‌سازی یک خواهد بود. رابطه‌ی این روش به صورت فرمول ۳-۱ است.

$$n_{ij} = \frac{x_{ij}}{\sum_{۱}^{m} x_{ij}} \qquad (۱-۳)$$

- نرمال کردن برداری:اگر منطق پردازش داده‌ها بر فاصله اقلیدسی (گشتاور مرتبه دوم) بنا شده باشد، از روش برداری (نُرم اُقلیدسی) استفاده می‌شود. روش برداری برخلاف روش ساده نرمال سازی خطی به صورت فرمول ۳-۲ انجام می‌شود:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{۱}^{m} x_{ij}^{۲}}} \qquad (۲-۳)$$

- استانداردسازی: یک مفهوم دیگر از نرمال کردن که به استاندارد کردن نیز موسوم است که در تحلیل شبکه‌های عصبی مصنوعی و تحلیل پوششی داده‌ها کاربرد دارد. برای استاندارد کردن یک عنصر باید آن عنصر را منهای میانگین کرده و بر انحراف معیار تقسیم نمایید. در این روش برای نرمال کردن یک عنصر باید آن عنصر را منهای مینیمم کرده و بر دامنه تغییرات تقسیم نمایید. روش‌های دیگری هم در زمینه نرمال کردن داده‌ها وجود دارد. البته باید بین روش نرمال کردن داده‌ها در تصمیم‌گیری چندمعیاره با روش‌های آماری تفاوت قائل شد. روابط موجود در فرمول ۳-۳ برای این روش صدق می‌کنند.

$$\frac{\mathrm{x} - \mathrm{mean}}{\mathrm{s}}, \frac{\mathrm{x} - \mathrm{min}}{\mathrm{max} - \mathrm{min}} \qquad (۳-۳)$$

۳-۴-۴ کاهش ابعاد

در یادگیری ماشین، فاکتورهای زیادی وجود دارند که طبقه‌بندی نهایی بر اساس آن‌ها انجام می شود. این عوامل اساساً به‌عنوان متغیر شناخته می شوند. هرچه تعداد ویژگی‌ها بیشتر باشد، تجسم مجموعه آموزش و سپس کار بر روی آن سخت‌تر خواهد شد. در برخی اوقات، بیشتر این ویژگی‌ها با هم مرتبط بوده و لذا ویژگی زائد تلقی می شوند. در اینجاست که الگوریتم های کاهش ابعاد بکار گرفته می شوند. از اهداف کاهش بعد می‌توان به موارد

زیر اشاره نمود:

- وقتی با مسائل واقعی و داده های واقعی سروکار داریم ، اغلب با داده های با ابعاد بالا روبرو هستیم.
- داده ها ساختار اصلی خود را در ابعاد بالا، نشان می دهند. اگرچه، گاهی اوقات باید این ابعاد را کاهش داد.
- باید ابعاد را طوری کاهش داده شوند که قابل تجسم باشند.

مؤلفه‌های کاهش بعد دو نوع می‌باشند:

- انتخاب ویژگی: در این بخش، باید زیرمجموعه ای از مجموعه اصلی متغیرها را انتخاب شوند. همچنین، به زیرمجموعه‌ای نیاز است که برای مدل‌سازی مسئله بصورت بهینه و مناسب عمل کند. روش‌های انتخاب ویژگی[۴] به منظور مواجهه با داده‌های با ابعاد بالا، به مؤلفه‌ای جدایی ناپذیر از فرآیند یادگیری مبدل شده‌اند. یک انتخاب ویژگی صحیح می‌تواند منجر به بهبود یادگیرنده استقرایی از جهت‌های گوناگون از جمله سرعت یادگیری، ظرفیت تعمیم و سادگی مدل استنتاج شده شود. از طرفی انتخاب ویژگی را می‌توان به عنوان فرآیند شناسایی ویژگی‌های مرتبط و حذف ویژگی‌های غیر مرتبط و تکراری با هدف مشاهده زیرمجموعه‌ای از ویژگی‌ها که مسئله را به خوبی و با حداقل کاهش درجه کارایی تشریح می‌کنند تعریف کرد. از مهم‌ترین روش‌های مختلف انتخاب ویژگی بصورت زیر می‌باشد.

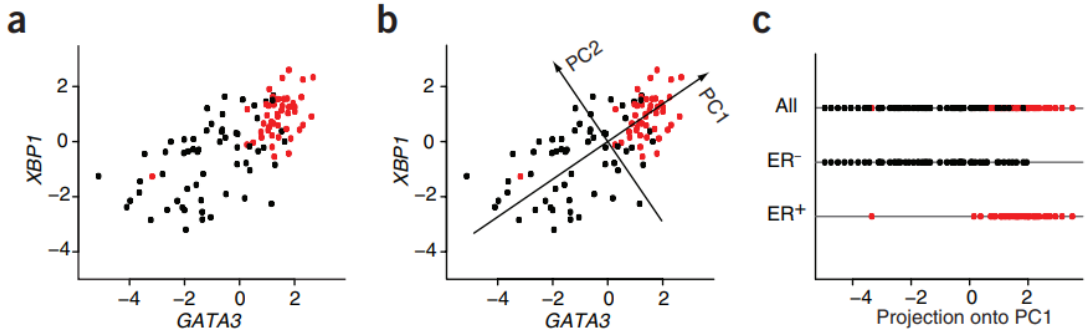
  - انتخاب ویژگی مستقیم: در روش های انتخاب ویژگی مستقیم[۵] ابتدا یک زیر مجموعه تهی از ویژگی‌ها ساخته می‌شود. سپس در هر مرحله، ویژگی‌هایی که بهترین عملکرد را برای مدل یادگیری به ارمغان می‌آورند، به این زیر مجموعه اضافه می‌شوند.
  - انتخاب ویژگی معکوس: در انتخاب ویژگی معکوس[۶] ابتدا تمامی ویژگی‌ها در زیر مجموعه حضور دارند. سپس در هر مرحله، بدترین ویژگی‌ها از زیر مجموعه حذف می‌شوند (ویژگی‌هایی که حذف آن‌ها، باعث ایجاد کمترین کاهش در عملکرد، دقت و کارایی روش یادگیری ماشین می‌شوند).

[4]Feature Selection
[5]Forward Feature Selection
[6]Backward Feature Selection

- ترکیب و استخراج ویژگی: در این فرآیند برای کاهش ابعاد، داده‌ها از یک فضای یعد بالا به یک فضای بعد پایین ، یعنی یک فضای با تعداد ابعاد کمتر اعمال می کنیم. در واقع ویژگی ها ترکیب شده و به یک فضای پایین‌تر نگاشت داده می‌شوند. روش های مختلف مورد استفاده برای کاهش ابعاد عبارتند از:

  - روش آنالیز مؤلفه‌های اصلی (PCA) آنالیز اجزای اصلی[۷]: در این روش بردارهایی که داده‌ها در آن‌ها بیشترین پراکندگی را دارند شناسایی شده و داده‌ها بر روی آن بردارها نگاشت داده می‌شوند. به این ترتیب راستاهایی که داده‌ها پراکندگی کمی در جهت آن‌ها دارند حذف می‌شوند و داده‌ها بر روی راستاهایی با بیشترین پراکندگی داده نگاشت داده می‌شوند که به این راستاها مؤلفه‌های اصلی گویند. تحلیل مؤلفه اصلی به بیان ساده، روشی برای استخراج متغیرهای مهم (به شکل مؤلفه) از مجموعه بزرگی متغیرهای موجود در یک مجموعه داده است. تحلیل مؤلفه اصلی در واقع یک مجموعه با بُعد پایین از ویژگی‌ها را از یک مجموعه دارای بُعد بالا استخراج می‌کند تا به ثبت اطلاعات بیشتر با تعداد کمتری از متغیرها کمک کند. بدین شکل، بصری‌سازی داده‌ها نیز معنادارتر می‌شود. تحلیل مؤلفه اصلی هنگامی که با داده‌های دارای سه یا تعداد بیشتری بُعد سروکار داشته باشید، کاربردپذیرتر است. تحلیل مؤلفه اساسی همیشه روی ماتریس کوواریانس یا همبستگی اعمال می‌شود. این یعنی داده‌ها باید عددی و استاندارد شده باشند. لازم به ذکر است که قبل از انجام PCA بایستی عمل نرمال‌سازی داده‌ها را انجام دهیم که پیش‌تر دراین‌باره صحبت شد. در شکل ۳-۱ می‌توان مراحل کاهش بعد انجام PCA را از دو بعد به یک بعد مشاهده کرد. [۴].

[7] Principal Component Analysis



شکل ۳-۱: مراحل کاهش بعد با استفاده از PCA [۴]

- روش تحلیل تفکیک کننده خطی (LDA) جدایی پذیری خطی[۸]: روش تحلیل تفکیک کننده خطی اکثرا به‌عنوان متود کم‌کردن بعدها در مرحله پیش‌پردازش برای دسته بندی الگوهای کاربردی یادگیری ماشین مورد استفاده قرار می‌گیرد. مقصود این است که یک داده‌ها بر روی یک فضای بعدی با جداپذیری مناسب دسته‌بندی شوند که از "نفرین ابعاد" بیش‌برازش[۹] جلوگیری شود و همچنین هزینه‌های محاسباتی کم شوند. تحلیل تفکیک کننده خطی تا حد زیادی به آنالیز واریانس و آنالیز رگرسیونی میل می‌کند؛ در هر سه این روش‌های آماری متغیر وابسته به صورت یک ترکیب خطی از متغیرهای دیگر مدل‌سازی می‌شود. با این وجود دو روش آخر متغیر وابسته را از نوع فاصله‌ای در نظر می‌گیرند در حالی که آنالیز افتراقی خطی برای متغیرهای وابسته اسمی یا رتبه‌ای به کار می‌رود. در نتیجه آنالیز افتراقی خطی به رگرسیون لجستیک شباهت بیشتری دارد. LDA همچنین با تحلیل مؤلفه‌های اصلی و تحلیل عاملی هم شباهت دارد؛ هر دوی این روش‌های آماری برای ترکیب خطی متغیرها به شکلی که داده را به بهترین نحو تفسیر کند بکار می‌روند یک کاربرد عمده هر دوی این روش‌ها، کاهش تعداد ابعاد است. با این وجود این روش‌ها تفاوت بزرگی با یکدیگر دارند: در تحلیل تفکیک کننده خطی، تفاوت کلاس‌ها اهمیت دارند در صورتیکه که در تحلیل مؤلفه‌های اصلی تفاوت کلاس‌ها نادیده گرفته می‌شود. LDA ارتباط نزدیکی با تحلیل واریانس و تحلیل رگرسیون دارد که سعی دارند یک متغیر مستقل را به عنوان ترکیبی خطی از ویژگی های

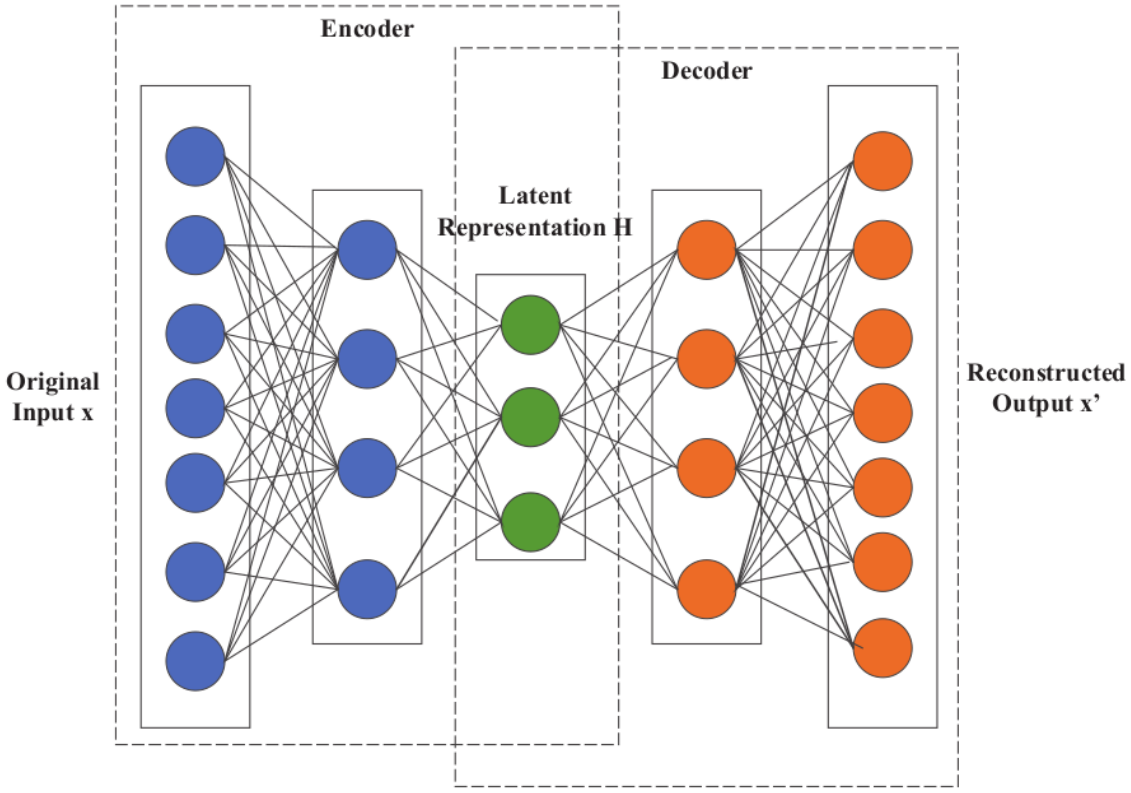
[8] Liear Discriminant
[9] Overfitting

دیگر بیان کنند. این متغیر مستقل در LDA به شکل برچسب یک کلاس است. به علاوه LDA ارتباطی نزدیکی با تحلیل مولفه های اصلی PCA دارد. زیرا هر دو روش بر روی ترکیبی خطی از متغیرهایی می‌باشند که بصورت مناسب‌ترین شکل داده‌ها را نشان می‌دهند. به علاوه اینکه LDA سعی در تفکیک کلاس‌های مختلف مجموعه داده را دارد. LDA وقتی مورد استفاده قرار می‌گیرد که اندازه‌های نمونه‌ها، مقدارهای پیوسته‌ای داشته باشند. ضمنت اینکه LDA داده‌های دارای توزیع نرمال، ویژگی‌هایی را که از نظر آماری مستقل می‌باشند و ماتریس کوواریانس یکسان را برای هر دسته مفروض است.

- خودرمزگذار: شبکه خودرمزگذار خودرمزگذار[۱۰] یک روش یادگیری ماشینی بی‌نظارت است که یک بردار ویژگی را به‌عنوان ورودی دریافت کرده و با تعداد ویژگی کمتر آن را به خروجی می‌دهد. در نتیجه کاهش بعد مبتنی بر شبکه هوش‌مصنوعی بصورت هوشمند و خودکار انجام می‌شود که در واقع سعی دارد بهترین کاهش بعد را در خروجی تحویل دهد. این مدل دارای دو بخش رمزگذار و رمزگشا دارد. که مانند سایر شبکه‌های عصبی این شبکه هم آموزش داده شده و تست می‌شود با این تفاوت که ورودی و خروجی در این شبکه یکی است یه این معنی که بردار ورودی وارد شبکه رمزگذار شده و خروجی شبکه رمزگذار که در واقع بردار کاهش بعد داده شده از ورودی اصلی است وارد ورودی شبکه رمزگشا می‌شود و خروجی شبکه رمزگشا هم دارای نورون‌هایی به تعداد بردار ویژگی ورودی اصلی است به این ترتیب شبکه خودرمزگذار آموزش و تست شده و در نهایت از قسمت شبکه رمزگذار به منظور کاهش بعد استفاده می‌کنیم. در شکل ۳-۲ ساختار شبکه خودرمزنگار قابل مشاهده می‌باشد [۵].

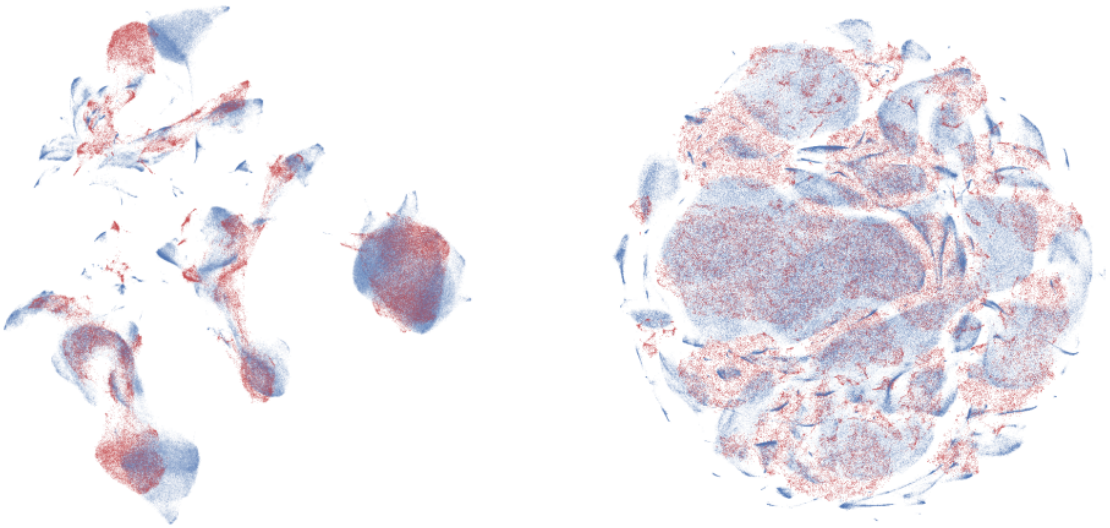
- روش تقریب چندگانه یکنواخت و تصویر[۱۱] (UMAP): این روش مبتنی بر کاهش بعد غیر خطی می‌باشد. ایده این روش مشابه با به مدل T-SNE می‌باشد اما مزیت روش مذکور سرعت بالای آن بر روی داده‌های با حجم و ابعاد بالا است. به‌علاوه این روش تا حد زیادی قادر است تا الگوهای کلی را حفظ کند به این معنی که توزیع و شکل کلی داده‌ها وقتی از بعد بالا به بعد پایین کاهش داده شود، حفظ می‌شود. با این روش می‌شود تا با رویکردهای باناظر و بی‌نظارت مدل کاهش بعد زا آموزش داد و از آن به‌منظور کم کردن ابعاد برای داده‌های جدید بهره برد. این روش به‌وسیله‌ی دو پارامتر بهینه‌سازی می‌شود که تعداد همسایه

[10] Auto-Encoder
[11] Uniform Manifold Approximation and Projection



شکل ۳-۲: ساختار شبکه خودرمزگذار [۵]

که تعادل بین الگوهای کلی و محلی را کنترل می‌کند و همچنین کمترین فاصله که نحوه متحد کردن توده‌های داده را در کنار هم کنترل می‌کند. مقدارهای کم برای تعداد همسایه باعث تمرکز الگوریتم روی ساختارهای محلی شده و در مقابل مقدارهای زیاد برای آن‌ها منجر به تمرکز بر روی ساختارهای کلی‌تر می‌شود. به‌علاوه مقدارهای زیاد برای کمترین فاصله منجر به مرتبط شدن نقاطی می‌شود که درواقع ارتباط کمی با هم دارند اما در عوض تمرکز شبکه بر روی حفظ ساختارهای گسترده‌تر خواهد شد. در شکل ۳-۳ داده‌های کاهش بعد داده شده از فضای بالاتر به فضای دو بعدی با استفاده از روش‌های UMAP ، T-SNE قابل مشاهده می‌باشند [۶] .



(a) UMAP (b) t-SNE

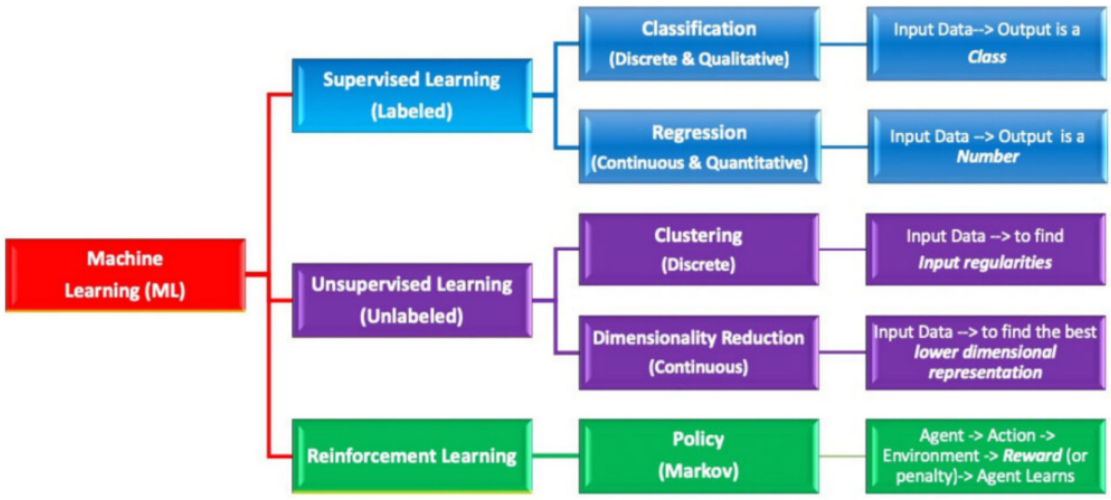
شکل ۳-۳: نمایش داده‌های کاهش بعد داده شده توسط روش‌های کاهش بعد UMAP، T-SNE [۶]

۳-۵ طبقه‌بندهای پایه برای طراحی سیستم طبقه‌بند مجتمع پویا

برای طراحی سیستم طبقه‌بند مجتمع پویا، به طبقه‌بندهای پایه نیاز است. در این قسمت به معرفی و تشریح انواع مدل‌های یادگیری‌ماشین یادگیری ماشینی[۱۲] و همچنین مدل‌های مبتنی بر هوش‌مصنوعی پرداخته خواهد شد.

یادگیری ماشین یکی از روش‌های مورد استفاده در آنالیز داده‌ها است که محور اساسی آن نحوه یادگیری خودکار ماشین‌ها برای تشخیص الگوهای درهم و اتّخاذ تصمیم‌های هوشمندانه مبنی بر داده‌های سابق می‌باشد. همچنین یادگیری ماشین زیر مجموعه‌ای از هوش مصنوعی بوده و تمرکز آن بر آموزش رایانه می‌باشد. تکنیک‌های یادگیری ماشین توان حل موضوعات بانظارت و نیز بدون نظارت و رسیدن به دانش مخفی موجود در داده‌ها را دارا می‌باشند. این تکنیک‌ها بوسیله‌ی آموزش مدل‌های ریاضی، امکان تشخیص الگوها و رسیدن به پیش‌بینی دقیق‌تر از آینده را فراهم می‌آورند. موضوعات یادگیری ماشینی از دو منظر دسته‌بندی می‌شوند. نگاه اول بر اساس وجود یا عدم وجود الگوهای برچسب‌دار است که تکنیک‌ها را به دو دسته بانظارت و نظارت‌نشده تقسیم می‌کند. دسته سوم هم تکنیک‌های نیمه‌نظارت شده هستند که کاربرد آنها در مواقعی است که الگوهای برچسب‌دار موجود اما محدود است. نگاه دوم تکنیک‌ها را بر اساس عملکرد و هدف آنها در حل مسئله طبقه بندی می‌کند. از این جهت برخی تکنیک‌ها به منظور طبقه‌بندی الگوها مورداستفاده قرار می‌گیرند. دسته بعدی با هدف برازش و پیش‌بینی سیگنال به‌کار می‌روند که با عنوان موضوعات رگرسیونی شناخته می‌شوند. دسته سومی از تکنیک‌ها نیز وجود دارند که با هدف خوشه‌بندی الگوهای شبیه به‌هم در دسته‌های کوچک‌تر استفاده می‌شوند. این‌گونه موضوعات که در داده‌های بدون‌برچسب تعریف می‌شوند از لحاظ هدف، شبیه به موضوعات طبقه‌بندی هستند ولی از نظر تکنیک متفاوت می‌باشند. تعدادی از متودهای یادگیری ماشینی با توجه به نوع موضوع در شکل ۳-۴ قابل مشاهده هستند. همانطور که پیش‌تر ذکر شد الگوریتم‌های یادگیری ماشین بر سه نوع هستند که در ادامه به تشریح آن‌ها پرداخته می‌شود. غالب روش‌های یادگیری ماشین یادگیری بانظارت را مورد استفاده قرار می‌دهند. در یادگیری ماشین بانظارت، مدل تلاش می‌کند تا از الگوهای قبلی که در اختیار آن قرار دارند، آموزش ببیند. به عبارت دیگر، در این روش یادگیری، مدل تلاش می‌کند تا الگوها را بر مبنای الگوهای ارائه شده به آن یاد بگیرد. همچنین در یادگیری ماشینی مجموعه داده به الگوریتم ارائه شده و ماشین منطق خود را بر مبنای آن مجموعه داده شکل می‌دهد. مجموعه داده مذکور سطرها و ستون‌هایی را دارد. سطرها که آن‌ها را رکورد و الگو داده می‌نامند، الگو داده‌ها را نمایندگی می‌کنند. موضوعات یادگیری ماشین بانظارت به دو دسته تقسیم می‌شوند:

[12] Machine Learning



شکل ۳-۴: انواع مدل‌های یادگیری ماشینی ۳-۴

- طبقه‌بندی: مسئله‌ای طبقه‌بندی نامیده می‌شود که متغیر خروجی، دسته و یا گروه باشد. برای نمونه، می‌توان به تعلق یک الگو به دسته‌های سیاه یا سفید و یک رایانامه به دسته‌های هرزنامه یا غیر هرزنامه نام برد. در حین آموزش، داده‌هابه یک الگوریتم دسته‌بندی وارد می‌شوند که هر کدام پیش‌تر در یک دسته یا کلاس خاصی طبقه‌بندی‌شده‌اند. وظیفه‌ی یک الگوریتم دسته‌بندی این است که بعد از پروسه آموزش یک ورودی بگیرد و آن را به یک دسته خاص از قبل تعیین شده اختصاص دهد که بر اساس روندی که در اثنای آموزش فرا گرفته است، آن داده بصورت واقعی در آن دسته قرار بگیرد.
- رگرسیون: مسئله‌ای رگرسیون نامیده میشود که متغیر خروجی یک مقدار پیوسته حقیقی مانند وزن باشد. دراصل در طبقه‌بندی با متغیرهای گسسته و در رگرسیون با متغیرهای پیوسته سروکار داریم.

در ادامه به بررسی انواع طبقه‌بندهای مرسوم می‌پردازیم.

۳-۵-۱ مدل طبقه‌بند بیز ساده

در حوزه «یادگیری ماشین»، روش طبقه‌بندی بیز ساده[۱۳] با استفاده از قضیه بیز و با فرض مستقل بودن متغیرها از یکدیگر، یکی از زیرروش‌های «دسته‌بندهای برمبنای احتمال» می‌باشد. در این روش با استفاده از بیشینه

[13] Naive Bayes

کردن تابع صحیح‌نمایی حداکثرسازی احتمال[۱۴] پارامترهای مدل تخمین زده می‌شوند. در زمینه‌های آماری و علم کامپیوتر، روش بیز ساده اسم‌های دیگری از جمله «بیز ساده» و یا «بیز مستقل» هم دارا می‌باشد. در واقع بیز ساده را یک مدل برپایه احتمال شرطی است که اگر X را برداری حاوی n ویژگی مستقل از هم در نظر گرفته شود، احتمال رخداد C_k را که از حالات دسته رویدادهای گوناگون است را می‌توان بصورت فرمول ۳-۴ نمایش داد.

$$p\left(C_k \mid X\right) = \frac{p\left(C_k\right) p\left(X \mid C_k\right)}{p(X)} \tag{۳-۴}$$

برای محاسبه این احتمال لازم می‌باشد که از «احتمال توام» استفاده کرده و همچنین با استفاده از احتمال شرطی، بصورت زیر ساده شود:

(۳-۵)

$$\begin{aligned}
&p\left(C_k, x_1, \ldots, x_n\right) = p\left(x_1, \ldots, x_n, C_k\right) \\
&= p\left(x_1 \mid x_2, \ldots, x_n, C_k\right) p\left(x_2, \ldots, x_n, C_k\right) \\
&= p\left(x_1 \mid x_2, \ldots, x_n, C_k\right) p\left(x_2 \mid x_3, \ldots, x_n, C_k\right) p\left(x_3, \ldots, x_n, C_k\right) \\
&= \ldots \\
&= p\left(x_1 \mid x_2, \ldots, x_n, C_k\right) p\left(x_2 \mid x_3, \ldots, x_n, C_k\right) \ldots p\left(x_{n-1} \mid x_n, C_k\right) p\left(x_n \mid C_k\right) p\left(C_k\right)
\end{aligned}$$

همچنین با فرض مستقل بودن ویژگی‌های X_i ها از هم، احتمالات را به شکل ساده‌تری قابل بازنویسی است:

$$p\left(x_i \mid x_{i+1}, \ldots, x_n, C_k\right) \approx p\left(x_i \mid C_k\right) \tag{۳-۶}$$

[14] Likelihood Maximization

نتیجتا احتمال توأم را به‌صورت حاصلضرب احتمال شرطی می‌توان نوشت:

$$\begin{aligned} p\left(C_k \mid x_1, \ldots, x_n\right) &\propto p\left(C_k, x_1, \ldots, x_n\right) \\ &\approx p\left(C_k\right) p\left(x_1 \mid C_k\right) p\left(x_2 \mid C_k\right) p\left(x_3 \mid C_k\right) \cdots \\ &= p\left(C_k\right) \prod_{i=1}^{n} p\left(x_i \mid C_k\right) \end{aligned} \tag{۳-۷}$$

در نهایت با استفاده از مدل احتمالاتی بیز و روابط بدست آمده طبقه‌بند را براساس مدل احتمالاتی با بیشینه‌سازی تابع زیر می‌توان طراحی کرد:

$$\hat{y} = \underset{k \in \{1, \ldots, K\}}{\operatorname{argmax}} p\left(C_k\right) \prod_{i=1} p\left(x_i \mid C_k\right) \tag{۳-۸}$$

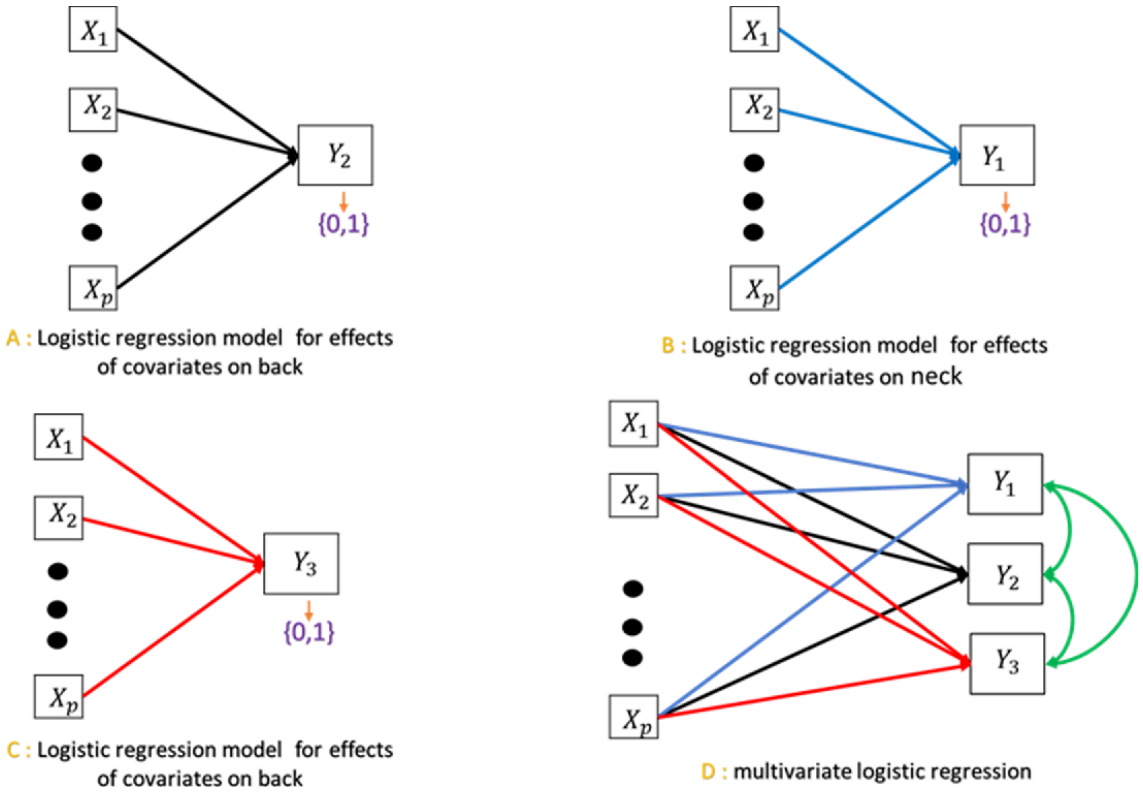
۳-۵-۲ مدل طبقه‌بند رگرسیون لجستیک

نحوه عملکرد مدل طبقه‌بند رگرسیون لجستیک[15] تا حد زیادی مشابه با مدل شبکه عصبي مصنوعي بایک لایه می‌باشد. در واقع رگرسیون لجستیک روشی مبتنی بر آمار می‌باشد که به‌منظور نشان دادن تأثیر ویژگی‌های کمی و یا کیفی بر متغیر خروجی دو کلاسه می‌باشد. آنالیز رگرسیون لجستیک مشابه با آنالیز رگرسیون خطی می‌باشد اما فرقش این است که در رگرسیون خطی متغیر خروجی متغیری کمی و پیوسته است منتها در رگرسیون لجستیک متغیر خروجی متغیر گسسته و دو کلاسه می‌باشد. رابطه رگرسیون لجستیک در فرمول ۳-۹ قابل مشاهده است.

$$p = \frac{e^{B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_j X_j + \cdots + B_k X_k}}{1 + e^{B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_j X_j + \cdots + B_k X_k}} \tag{۳-۹}$$

اگر این روش را بصورت مدل شبکه عصبی با یک لایه در نظر گیریم، به طوریکه ورودی به‌صورت بردار ویژگي X با ضرب برداری در بردار وزن و جمع با یک مقدار بایاس در تابع فعال‌ساز که به‌عنوان نورورن شبکه عصبی در نظر گرفته می‌شود وارد شود خروجی مورد نظر به‌صورت، مقدار باینری خواهد بود که در واقع خروجی هدف می‌باشد. البته مدل رگرسیون لجستیک بصورت چند کلاسه و چند متغیره در خروجی هم قابل تعریف است. در شکل ساختار مدل رگرسیون لجستیک و رگرسیون لجستیک چند متغیره قابل مشاهده است. [۷].

[15] Logistic Regression



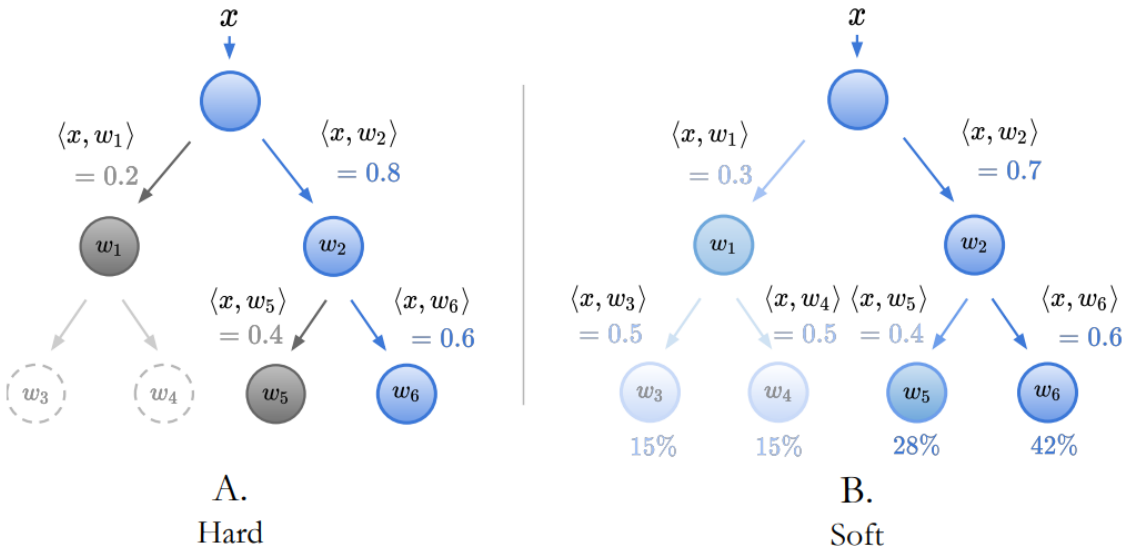
شکل ۳-۵: مدل‌های رگرسیون لجستیک ساده و چند متغیره [۷]

۳-۵-۳ مدل طبقه‌بند درخت تصمیم

مدل طبقه‌بند درخت تصمیم[۱۶] نمونه ای از مدل‌های پر کابرد در بین روش‌های داده‌کاوی می‌باشد. این مدل به‌صورت یک درخت عمل می‌کند ببه این صورت که از ریشه به سمت پایین (برگ) رشد کرده است. در مدل درخت تصمیم نمونه‌ها گروه‌بندی می‌شوند پس گروه‌ها در آخر گره‌های برگ جای دارند. درخت تصمیم در مواردی مورد استفاده قرار می‌گیرد که امکان تعریف نام یک دسته یا کلاس به عنوان پاسخ واحد وجود داشته باشد. پس در واقع متودهای طراحی درخت تصمیم اغلب از بالا به پایین رفتار می‌کنند یعنی اول فضای ورودی به فضاهای کوچکتر تبدیل شده، پس از آن مرحله بخش‌بندی برای هر کدام از این بخش‌ها تکرار می‌شود. به عبارت دیگر در زمان طراحی درخت، اول از همه ریشه ساخته شده، بعد از آن هر کدام از زیر شاخه‌ها به شاخه‌های دیگری تقسیم

[16] Decision Tree

شد و این مراحل تا به انتها تکرار می‌شوند. نمونه‌ای از درخت تصمیم سخت و نرم در شکل ۳-۶ قابل مشاهده است [۸].

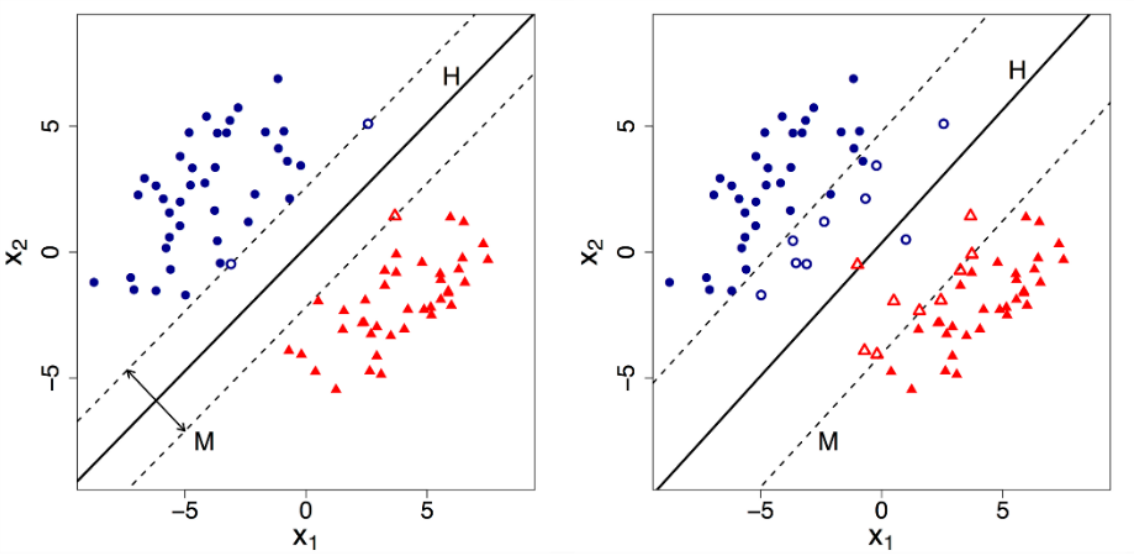


شکل ۳-۶: نمونه‌ای از درخت تصمیم سخت و نرم [۸]

۳-۵-۴ مدل طبقه‌بند ماشین بردار پشتیبان

مدل طبقه‌بند ماشین بردار پشتیبان[۱۷]، یک روش نظارت‌شده یادگیری ماشین می‌باشد که هم برای طبقه‌بندی و هم رگرسیون مورد استفاده قرار می‌گیرد، با این‌وجود اکثرا در طبقه‌بندی از آن بهره برده می‌شود. در مدل مذکور، هر نمونه به منزله یک نقطه در یک فضای n بعدی روی نمودار گستردگی داده‌ها رسم کرده و مقدار هر متغیر مربوط به داده‌ها را یکی از مؤلفه‌های مختصات نقطه روی نمودار را مشخص می‌کند. پس از آن، با رسم کردن یک خط راست، داده‌های مختلف و متمایز از هم را طبقه‌بندی می‌کند. به‌عبارتی داده‌ها توسط یک خط راست از هم جدا شده به علاوه اینکه می‌توان با استفاده از حاشیه اطمینان نرم، مدل ارتقا یافته از SVM را به منظور طبقه‌بندی نرم استفاده کرد. در شکل ۳-۷ نحوه عملکرد این مدل در فضای دو بعدی قابل مشاهده است [۹].

[17] Support Vector Machine



شکل ۳-۷: نحوه عملکرد مدل SVM در فضای دو بعدی [۹]

۳-۵-۵ مدل طبقه‌بند جنگل تصادفی

مدل طبقه‌بند جنگل تصادفی[۱۸] یک روش یادگیری بانظارت می‌باشد. همانطور که از اسم آن پیداست، این روش یک جنگل را بصورت تصادفی تشکیل می‌دهد. «جنگل» طراحی شده، در اصل مجموعه‌ای است از «درخت‌های تصمیم». وظیفه طراحی جنگل با استفاده از درخت‌ها معمولا به متودی به‌نام «کیسه گذاری» محول می‌شود. دستورکار متود کیسه‌گذاری به این صورت است که تلفیقی از روش‌های یادگیری، نتایج کلی مدل را بهبود می‌دهد. به عبارتی جنگل تصادفی تعدادی درخت تصمیم را طراحی کرده و با ادغام آن‌ها با یکدیگر، علکرد کلی مدل صحیح‌تر و پایدارتر خواهد شد. از مزیت‌های جنگل تصادفی این است که هم برای مسائل دسته‌بندی و هم رگرسیون قابل استفاده می‌باشد.

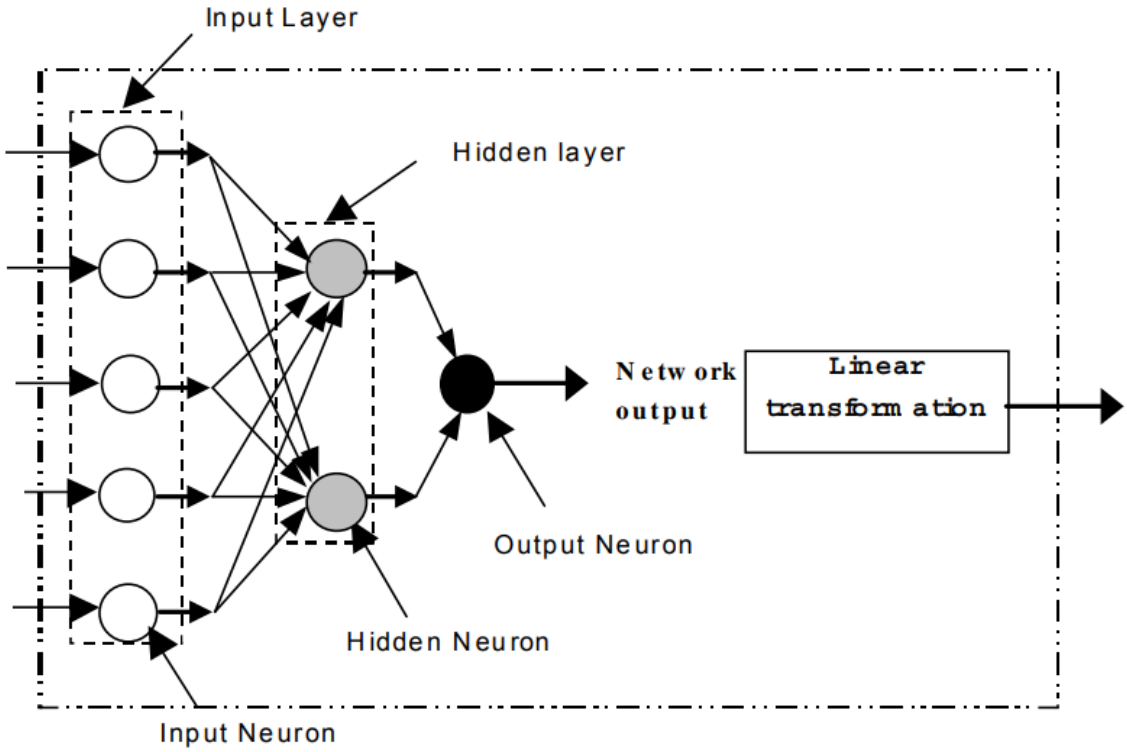
[18] Random Forest

۳-۵-۶ مدل‌های طبقه‌بند مبتنی بر شبکه عصبی مصنوعی

مدل‌های طبقه‌بند مبتنی بر شبکه عصبی مصنوعی[۱۹] متودی به منظور پردازش اطلاعات می‌باشد، این شبکه از ساختار عصبی زیستی الهام گرفته و مثل مغز داده‌ها را پردازش می‌کند. ساختار نوین سیستم پردازش اطلاعات از ویژگی‌های اصلی این متود است. این مجموعه از تعداد فراوانی از عناصر پردازشی تنیده در هم با اسم نورون‌ها تشکیل شده که به‌منظور حل یک مسئله با هم همکاری میکنند. شبکه‌های عصبی مصنوعی مثل مغز انسان با الگوهای مختلف یاد می گیرند و یک شبکه عصبی برای دستورالعمل‌های معین مثل تشخیص الگوها و طبقه‌بندی داده‌ها، در یک فرایند یادگیری تنظیم می‌شود. در مجموعه‌های زیستی، یادگیری با تنظیماتی در اتصالات سیناپسی که بین اعصاب قرار دارد انجام می‌پذیرد. از این روش در شبکه‌های عصبی هم استفاده می‌شود. شبکه های عصبی مصنوعی با پردازش داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل کرده که به این امر یادگیری می‌گویند. توانایی یادگیری مهمترین ویژگی یک مجموعه هوشمند می‌باشد. سیستمی که توانایی فراگیری الگوها را داراست، به علاوه منعطف تر است و ساده‌تر برنامه‌ریزی می‌شود، پس به‌صورت مناسب‌تر توانایی پاسخگویی در مورد مسائل و معادله‌های جدید را داشته باشد. انواع مدل‌های شبکه‌های عصبی را در ادامه مورد بررسی قرار می‌دهیم.

- شبکه عصبی پیشخور پیشخور[۲۰]: یک مدل عصبی مصنوعی است که در آن، اتصالات بین نورون‌ها حلقه تولید نمی‌کنند. در این مدل، داده‌ها فقط در جهت رو به جلو عبور می‌کنند به این معنی که تنها از نورون‌های ورودی به سمت نورون‌های خروجی و با عبور از نورون‌های مخفی، حلقه یا بازگشتی موجود نمی‌باشد. در شکل ۳-۸ نمونه ای از شبکه عصبی پیشخور قابل مشاهده است.

[19] Artificial Neural Network
[20] Feed Forward



شکل ۳-۸: نمونه‌ای از شبکه عصبی پیشخور [۱۰]

- شبکه پرسپترون تک لایه: این مدل، لایه مخفی نداشته و ساده‌ترین مدل از میان مدل‌های شبکه عصبی مصنوعی پیشخور می‌باشد که هیچ لایه مخفی هم دارا نیست. در اصل مدل مذکور، صرفا یک لایه ورودی،یک لایه مخفی و یک لایه خروجی دارد. در نتیجه به ایم مدل به شبکه پرسپترون تک لایه معروف است.

- شبکه پرسپترون چند لایه: این گروه از مدل‌ها، از چند عدد لایه نورون محاسباتی مخفی طراحی شده است که معمولا به‌طور پیشخور (در جهت رو به جلو) به هم اتصال دارند. هر نورون در یک لایه، بطور مستقیم به نورون‌های لایه بعدی متصل است. در شمار زیادی از موارد، نورون‌های این مدل‌ها، تابع سیگموید در فعالساز مورد استفاده قرار می‌گیرد. همچنین مدل پرسپترون چندلایه در توصیف‌های غیرخطی تا حد زیادی بادگیری را بطور مناسب انجام می‌دهد.

۳-۶ سیستم‌های طبقه‌بند مجتمع مرسوم

همانطور که در فصل اول ذکر شد، مدل‌های یادگیری ماشینی با رویکرد طبقه‌بندی مجتمع، با اهدافی مانند کاهش واریانس دقت خروجی، تفسیرپذیری بهتر، قابلیت تعمیم‌دهی بالاتر، کاهش بایاس خروجی استفاده می‌شود. در این نوع مدل‌ها مدل اصلی، تجمیعی از مدل‌های یادگیری ماشینی پایه ساده‌تر می‌باشد. همچنین این مدل‌ها در دو نوع ایستا و پویا طراحی می‌شوند. در مدل ایستا هر کدام از مدل‌های پایه به‌صورت جداگانه بر روی کل داده‌ها یا قسمتی از داده‌ها آموزش و تست می‌شوند و در نهایت با ترکیب مدل‌های پایه، مدل اصلی تشکیل می‌شود.

به عبارتی دیگر اصل "خرد جمعی" اثبات که دسته عظیمی از اشخاص با دانش متوسط در مورد یک موضوع قادر هستند تا به پرسش‌هایی مثل پیش‌بینی مقادیر، استدلال هندسی و دانش عمومی جواب‌هایی با قابلیت اطمینان بالاتری نسبت به یکم فرد پاسخ دهند. چنین اصلی می تواند در مورد مسائل هوش مصنوعی که بر یادگیری ماشین متکی هستند اعمال شود. در یادگیری ماشین، خرد جمعی به وسیله‌ی یادگیری گروهی به دست می آید. برای شمار زیادی از مسائل، نتایج حاصل از شبکه ترکیبی از نتایج هر یک از مدل‌های یادگیری ماشین پایه، دقیق‌تر می‌باشد. یادگیری تجمعی با تلفیق چند مدل به بهبود نتایج یادگیری ماشین کمک می‌کند. در [۸۰] نمونه‌ای از ترکیب طبقه‌بند مجتمع مبتنی بر رویکرد OWA پیاده‌سازی شده است. اصولا مدل‌های تجمعی به سه نوع در کیسه کردن[۲۱]، Boostingو Stacking تقسیم می‌شوند که در ادامه به توضیح آن‌ها پرداخته شده است:

۳-۶-۱ مدل طبقه‌بند مجتمع

در مدل طبقه‌بند مجتمع در کیسه کردن از مدل‌های پایه همگون استفاده می‌شود، این مدل‌ها به‌صورت مستقل از یکدیگر و به صورت موازی آموزش داده شده و خروجی آن‌ها با میانگین وزن‌دار با یکدیگر ترکیب شده و خروجی اصلی را منجر می‌شوند. این روش به منظور ساخت مدل تجمیعی پایدارتر و با واریانس کمتر در توزیع دقت خروجی به‌کار می‌روند. لازم به ذکر است که عملکرد مدل‌های پایه می‌توانند ضعیف باشند. مدل جنگل تصادفی نمونه‌ای از این مدل تجمعی می‌باشد که با هدف وایانس کمتر در خروجی طراحی می‌شوند.

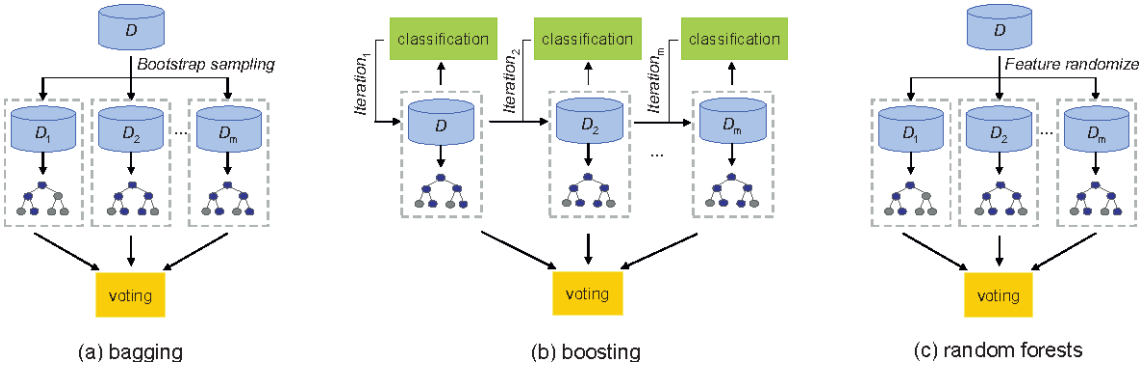
[21] Bagging

۳-۶-۲ مدل طبقه‌بند مجتمع قدرتی

در مدل طبقه‌بند مجتمع تقویت[۲۲] هم مدل‌های پایه ضعیف همگون مورد استفاده قرار می‌گیرند با این تفاوت که به جای ترکیب موازی، به‌صورت سری و تطبیقی آموزش داده می‌شوند و در نهایت یک مدل تجمیعی به‌مراتب قوی‌تر از هر یک از مدل‌های پایه ضعیف ساخته خواهد شد. توزیع دقت خروجی این مدل ترکیبی برخلاف مدل قبلی در عوض کاهش واریانس، کاهش بایاس را تجربه می‌کند. روش آدابوست متداول‌ترین در این نوع مدل ترکیبی می‌باشد که از مدل‌های پایه درخت تصمیم تشکیل می‌شود.

۳-۶-۳ مدل طبقه‌بند مجتمع پشته‌سازی

در مدل طبقه‌بند مجتمع پشته‌سازی[۲۳] مدل‌های پایه ناهمگون مورد استفاده قرار می‌گیرند و به‌طور موازی آموزش داده می‌شوند. خروجی این مدل‌های پایه در نهایت با هم ترکیب شده و مدل نهایی را تشکیل می‌دهند. [۱۱] در شکل ۳-۹ هر سه نوع مدل تجمیع ذکر شده قابل مشاهده می‌باشد.



شکل ۳-۹: مقایسه ساختار سه مدل تجمیع مختلف [۱۱]

[22] Boosting
[23] Stacking

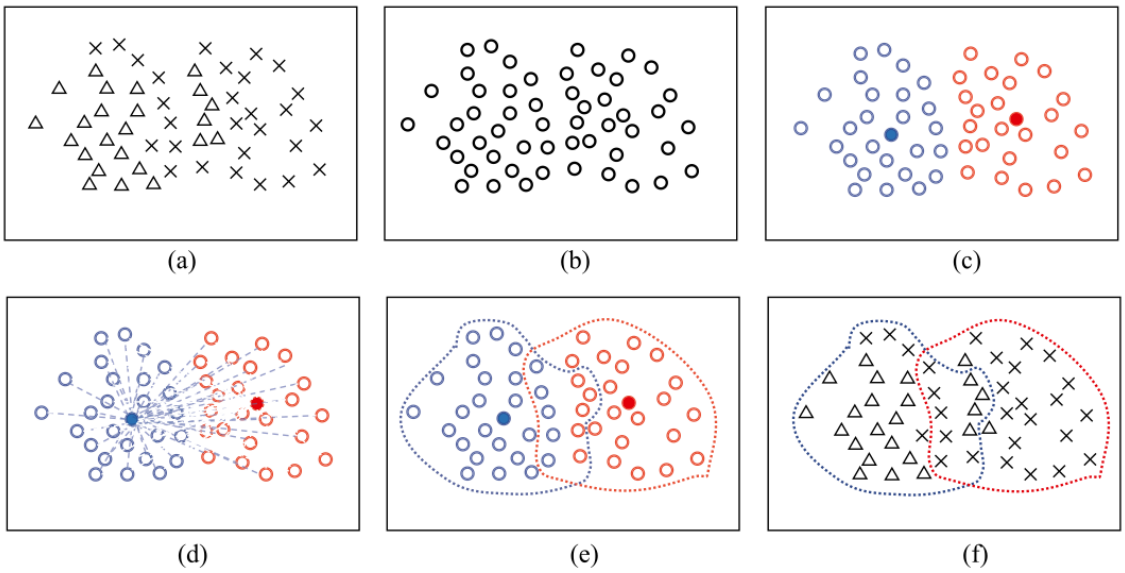
۳-۷ مدل طبقه‌بند مجتمع پویا

مدل طبقه‌بند مجتمع پویا: طبقه‌بندهای مجتمعی که پیش‌تر مورد بحث قرار گرفتند از نوع ایستا بودند. مدلی که در این پژوهش مورد بررسی و توسعه قرار می‌گیرد مدل مجتمع از نوع پویا می‌باشد. در این مدل بر اساس توزیع هر داده‌های ورودی در فضای ویژگی، طبقه‌بندی آن توسط مدل‌های پایه انجام می‌شود. این در صورتی است که در مدل ایستا تمامی داده‌های ورودی تست، مستقل از جایگاه آن‌ها در فضای ویژگی، طبقه‌بندی می‌شوند. در واقع آموزش و تست داده‌ها بصورت محلی و مطابق با جایگیری داده تست در فضای ویژگی انجام می‌پذیرد.

مدل‌های پایه مربوط به مدل مجتمع پویایی که در این پژوهش مورد بررسی قرار می‌گیرد بر اساس توزیع داده‌ها در فضای ویژگی،آموزش داده می‌شوند و داده‌های تست مطابق با محل قرارگیری آن‌ها در فضای ویژگی و فاصله اقلیدسی آن‌ها از مراکز خوشه‌ها طبقه‌بندی می‌شوند.

در واقع این روش از دو رویکرد خوشه‌بندی و طبقه‌بندی استفاده می‌کند. در ادامه مراحل انجام این روش توضیح داده شده است:

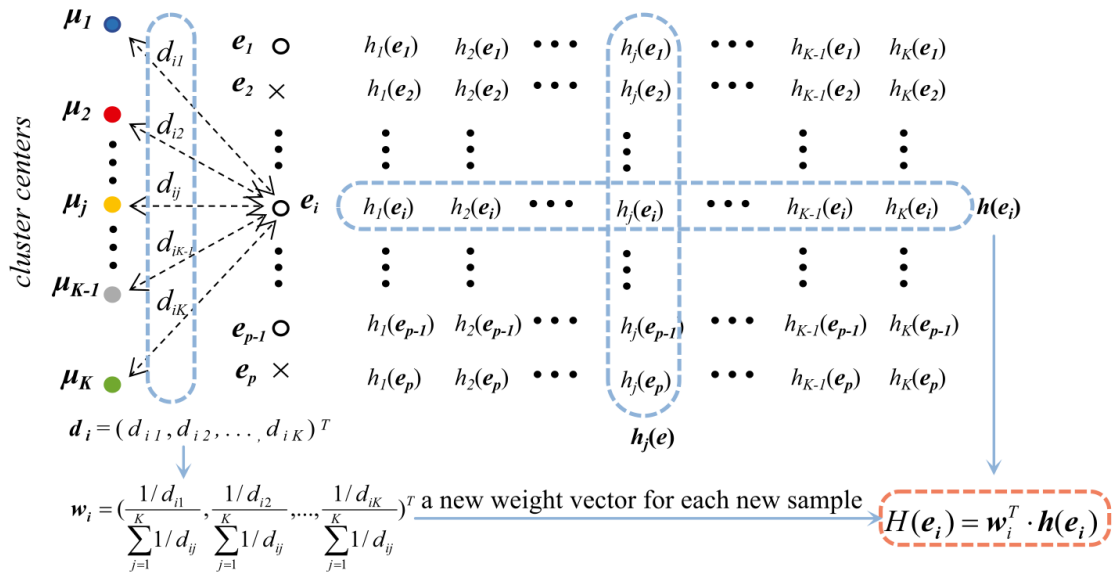
۱. خوشه‌بندی: در ابتدا داده‌ها توسط یک الگوریتم خوشه‌بندی در فضای ویژگی خوشه‌بندی می‌شوند (در صورت نیاز داده‌های هر کدام از خوشه‌ها برای مرحله بعد افزایش خواهد یافت.) در شکل ۳-۱۰ نحوه شکل‌گیری خوشه‌ها قابل مشاهده است [۱۲].



شکل ۳-۱۰: مراحل شکل‌گیری خوشه‌ها توسط الگوریتم خوشه‌بندی K-Means [۱۲]

۲. آموزش مدل‌های پایه: در این بخش مدل‌های پایه یادگیری ماشینی برای هر کدام از خوشه‌های نهایی بصورت جداگانه آموزش داده می‌شوند.

۳. ارزیابی مدل تجمیعی پویا: در این بخش داده‌های آزمایشی وارد مدل شده و بر اساس فاصله هر کدام از داده‌های آزمایشی تا مراکز خوشه‌ها، توسط مدل‌های پایه‌ی آموزش داده شده مربوط به نزدیکترین خوشه، طبقه‌بندی می‌شوند. شماتیک مربوط به این مرحله که در واقع مبتنی بر فاصله بودن این الگوریتم تجمیعی پویا را نشان می‌دهد، در شکل ۳-۱۱ قابل مشاهده می‌باشد.



شکل ۳-۱۱: شماتیک مربوط به طبقه‌بند مجتمع پویای مبتنی بر فاصله [۱۲]

۳-۸ ارزیابی مدل‌های یادگیری ماشینی

مدل‌های یادگیري ماشيني بايد طوری طراحی شوند که قابلیت تعمیم دهی را با دقت مناسب به دنياي واقعي را نيز دارا باشند. در واقع مدل‌هایی که طراحی می‌شوند بر روی قسمتی از داده‌های واقعی آموزش و آزمایش می‌شوند و باید بر روی داده‌های دیگر دنیای واقعی نیز به‌درستی عمل کنند. به همین دلیل به شاخص‌های ارزیابی نیاز داریم تا علاوه بر ارزیابی مدل بر روی داده‌های آزمایشی، با داده‌های آزمایشی طوری ارزیابی شوند که این ارزیابی برای داده‌های واقعی‌ای که در دنیای واقعی نیز وجود دارند صدق کند. در ادامه به بررسی انواع شاخص‌های ارزیابی پرداخته می‌شود.

۳-۸-۱ معرفی ماتریس "درهم ریختگی"

ماتریس درهم ریختگی[۲۴] ماتریسی است که در آن عملکرد مدل‌های مربوط را به نمایش می‌گذارد. اغلب چنین گزارشی برای مدل‌های یادگیری بانظارت مورد استفاده قرار می‌گیرد ، هرچند در یادگیری بی‌نظارت هم می‌تواند استفاده شود. در واقع این ماتریس تعداد نمونه‌های موجود در هر کلاس و تعداد نمونه‌هایی که به‌درستی در کلاس مذکور پیشبینی شده شده‌اند را نشان می‌دهد. با استفاده از آن می‌توان فهمید چند عدد نمونه از کلاس‌های دیگر به‌صورت اشتباه در یک کلاس طبقه‌بندی شده است و نیز چند عدد نمونه از این کلاس به‌صورت اشتباه در کلاس‌های دیگر طبقه‌بندی شده‌اند. به علاوه این ماتریس حاوی ۴ پارامتر اصلی می‌باشد:

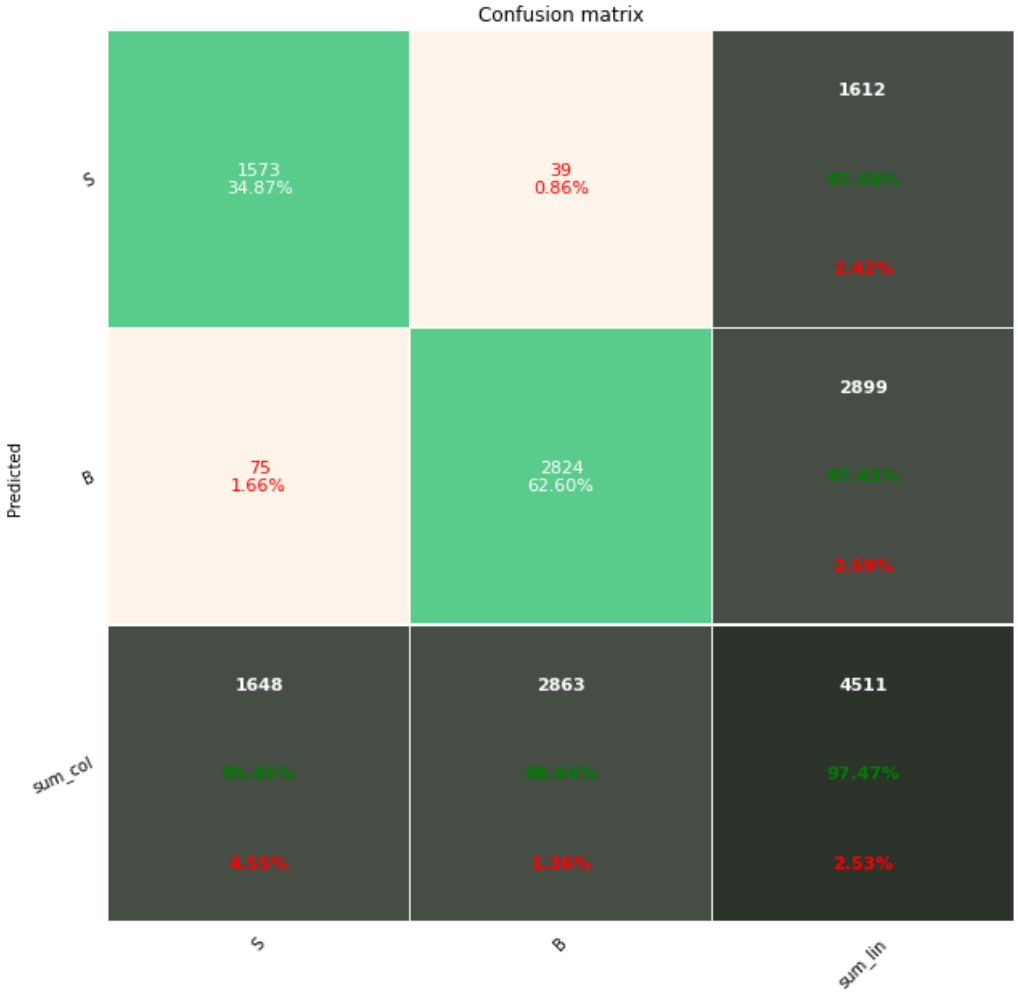
- مثبت صحیح خطای مثبت واقعی: اگر C_k را به‌عنوان یکی از کلاس‌ها در نظر بگیریم، تعداد نمونه‌هایی را که مدل به‌درستی در کلاس C_k طبقه‌بندی کرده است را این پارامتر نمایندگی می‌کند.
- منفی واقعی[۲۵]: تعداد نمونه‌هایی را که مدل به‌درستی در کلاسی غیر از C_k طبقه‌بندی کرده است را این پارامتر نمایندگی می‌کند.
- مثبت کاذب[۲۶]: تعداد نمونه‌هایی را که مدل به‌اشتباه در کلاس C_k طبقه‌بندی کرده است را این پارامتر نمایندگی می‌کند.
- خطای منفی کاذب: تعداد نمونه‌هایی را که مدل به‌اشتباه در کلاسی غیر از C_k طبقه‌بندی کرده است را این پارامتر نمایندگی می‌کند.

در شکل ۳-۱۲ نمونه‌ای از ماتریس درهم‌ریختگی قابل مشاهده می‌باشد.

۳-۸-۲ شاخص‌های ارزیابی مدل‌های یادگیری ماشینی

نتایج به‌دست‌آمده مورد ارزیابی شده و برای موارد مختلف تفسیر و استفاده می‌شود. در ارزیابی اغلب معیارهای زیر در نظر گرفته می‌شوند:

[24] Confusion Matrix
[25] True Negative
[26] False Positive



شکل ۳-۱۲: نمونه‌ای از ماتریس درهم‌ریختگی دوکلاسه

- دقت Accuracy: با استفاده از این معیار می‌توان عملکرد کلی مدل را در طبقه‌بندی تمامی کلاس‌های مورد نظر مجموعه‌داده ارزیابی کرد. رابطه مربوط به این معیار را در فرمول ۳-۱۰ زیر مشهود است.

$$Accuracy = (TP + TN)/(TP + FN + FP + TN) \qquad (۳-۱۰)$$

- صحت Precision: با استفاده از این معیار می‌توان ارزیابی کرد که مدل در طبقه‌بندی داده‌هایی که به

اشتباه از سایر کلاس‌ها در هر کلاس طبقه‌بندی می‌شوند به چه صورت عمل می‌کند. رابطه مربوط به این معیار را در فرمول ۳-۱۱ زیر مشهود است.

$$Precision = TP/(TP + FP) \quad (۱۱-۳)$$

- حساسیت Recall: با استفاده از این معیار می‌توان ارزیابی کرد که مدل در طبقه‌بندی داده‌هایی که به اشتباه از هر کلاس در کلاس‌های دیگر چه صورت عمل می‌کند. رابطه مربوط به این معیار را در فرمول ۳-۱۲ زیر مشهود است.

$$Recall = TP/(TP + FN) \quad (۱۲-۳)$$

- معیار F: میانگین وزن‌دار معیارهای صحت و حساسیت را معیار F گویند.

۳-۸-۳ توابع زیان برای ارزیابی مدل‌های یادگیری ماشینی

تابع زیان، شاخصی به‌منظور ارزیابی عملکرد مدل از دید قابلیت و توانایی در پیشبینی مقادیر جدید است. از روش‌های متداول برای یافتن کمترین مقدار تابع زیان، استفاده از مشتق و الگوریتم «گرادیان کاهشی» می‌باشد. مدل‌های زیان متداول بصورت زیر می‌باشند.

- تابع زیان خطای میانگین مربعات[۲۷] (MSE): این تابع زیان، میانگین مربعات فاصله بین مقدار پیش‌بینی و واقعی را محاسبه می‌کند. رابطه برای محاسبه این تابع بصورت فرمول

  است.

- میانگین قدرمطلق خطا (MAE) خطای قدر مطلق مربعات[۲۸]: این معیار خطا، مشابه MSE، از فاصله میان مقادیر پیش‌بینی و واقعی برای ارزیابی مدل استفاده کرده با این تفاوت که جهت این تفاضل را در نظر

[27] Means Square Error
[28] Mean Absolute Error

نمی‌گیرد. پس در محاسبه خطا MAE فقط اندازه فاصله و نه جهت فاصله به کار می‌رود.

فصل ۴

ارائه و توصیف داده‌ها و مقایسه نتایج حاصل از پیاده‌سازی روش پیشنهادی با سایر روش‌ها

۴-۱ مقدمه

در فصل گذشته به تشریح روش پیشنهادی این پژوهش و همچنین مدل‌های یادگیری ماشینی متداولی که در این پژوهش به‌کار می‌رود پرداخته شد. در این فصل با تحلیل توصیفی و تحلیل پیش‌بین بر روی داده‌های در دسترس که مربوط به مدیریت موجودی کالا می‌باشند و همچنین مقایسه نتایج روش‌های مختلف، به اثبات برتری روش پیشنهادی در این پژوهش پرداخته شده است.

۴-۲ بررسی و دستکاری اولیه داده‌ها

در ابتدا نیاز است یک سری از عملیات دستکاری را بر روی داده‌ها انجام دهیم تا تحلیل توصیفی و همجنین تحلیل پیش‌بین، ممکن و راحت‌تر انجام شوند. عملیاتی که بر روی مجموعه‌داده مذکور انجام شد است در ادامه بیان شده اند.

۴-۲-۱ حذف داده‌های بدون استفاده

پس از بارگزاری دادها با استفاده از دستورهای لازم در نرم افزار پایتون، داده‌های تکراری یعنی داده‌هایی که عینا در مجموعه داده تکراری هستند حذف شدند. همچنین نمونه‌ها و نیز ویژگی‌هایی را که دارای داده‌ی گمشده زیادی بودند از مجموعه داده حذف شدند. پس از انجام این حذفیات تعداد نمونه‌ها به ۱۴۱۰۰ رسید.

۴-۲-۲ حذف و یا تغییر نام متغیرها

متغیرهایی که برای تحلیل مورد نظرمان اهمیتی ندارند مانند شماره سریال کالاها، آدرس اینترنتی عکس کالا، شماره سریال مخصوص سازنده و ... از مجموعه داده حذف شدند. همچین نام متغیرهایی که مناسب نبودند و یا نام متغیرهای طولانی اصلاح شد.

۴-۲-۳ یکسان‌سازی واحد قیمت‌ها

در این مجموعه داده قیمت کالاها دارای واحدهای پولی مختلف بود که با انجام تبدیلات لازم واحد پولی تمام قیمت‌ها تبدیل به دلار آمریکا شد.

۴-۲-۴ یکسان‌سازی واحد وزن

یکسان‌سازی واحد وزن: در این مجموعه داده وزن کالاها دارای واحدهای وزنی مختلف بود که با انجام تبدیلات لازم واحد وزنی تمام کالاها تبدیل به پوند[1] شد. مجموعه داده‌ها بعد از عملیات دستکاری، شامل ۱۴۱۰۰ نمونه و ۸ ویژگی اصلی برای هر نمونه می‌باشد. متغیر availability (وضعیت موجودی کالاها) که متغیر وابسته و کلیدی برای برای مرحله تحلیل پیش‌بین است، یک متغیر عددی کیفی سه‌سطحه می‌باشد که به منظور تشخیص موجودیت کالا به عنوان خروجی مدل در نظر گرفته می‌شود. در جدول ۴-۱ اطلاعات مربوط به متغیرها و انواع آن‌ها قابل مشاهده می‌باشد.

[1] Pound

جدول ۴-۱: معرفی متغیرها و نوع آن‌ها

نام متغیر	نوع متغیر	تشریح
availability	Object	وضعیت موجودیت کالا در انبار
maxprice	Float	حداکثر قیمت فروش کالا
minprice	Float	حداقل قیمت فروش کالا
isSale	Bool	آیا کالا تخفیف خورده است یا خیر
prices.merchant	Object	نام فروشگاه
brand	Object	نام برند کالا
dateAdded	Datatime	اولین تاریخی که کالا به پایگاه داده محصولات اضافه شده است
dateUpdated	Datatime	آخرین تاریخ بروزرسانی محصول شده
weight	Float	وزن کالا

۴-۲-۵ تبدیل متغیرهای کیفی به کمّی و خلق ویژگی‌های جدید

به منظور انجام تحلیل توصیفی و همچنین پیش‌بین بر روی مجموعه داده در دسترس، بایستی متغیرهای کیفی تبدیل به متغیرهای کمی شوند. با استفاده از روش کدگذاری OneHot متغیرهای کمی مجموعه داده تبدیل به متغیرهای عددی می‌شوند. به این‌صورت که به تعداد سطوح سطوح[۲] هر متغیر کیفی، یک متغیر شامل مقادیر صفر و یکی به مجموعه داده افزوده می‌شود. همانطور که در جدول ۴-۱ پیداست، متغیرهای isSale و prices.merchant و brand متغیرهای کیفی بوده که پس از کمی‌سازی آن‌ها، تعداد متغیرهای مستقل به تعداد ۱۴۷ عدد رسید. علاوه بر ویژگی‌های اضافه شده به مجموعه داده، با استفاده از متغیرهای date Added و date Updated اطلاعات سال، ماه، روز برای هر تاریخ استخراج شده و به‌عنوان ویژگی جدید به مجموعه داده اضافه می‌شوند. پس از انجام این امر، تعداد ویژگی‌های مجموعه داده جدید ۱۵۵ عدد می‌شود. به همین ترتیب متغیر وابسته‌ی availability که یک متغیر کیفی است با استفاده از روش OneHot برای مدل‌هایی که نیاز به کدگذاری برای مدل‌سازی هستند، تبدیل به متغیر کمی می‌شود.

[2]Levels

۴-۳ تحلیل توصیفی داده‌ها

در فصل گذشته درباره نحوه دسترسی به داده‌های مورد استفاده در این پژوهش صحبت شد. در این قسمت به تحلیل توصیفی مجموعه داده پرداخته شده است.

۴-۳-۱ تحلیل توصیفی و مصورسازی متغیرهای مستقل

از آنجایی که متغیرهای مستقل بصورت مستقیم و ترکیبی بر روی متغیر بر روی متغیر کلیدی (متغیر وابسته) تاثیر و همبستگی دارند در نتیجه نقش اصلی را در پیشبینی وضعیت موجودی بازی می‌کنند در نتیجه لازم است متغیرهای مستقل به عنوان ویژگی‌های مجموعه داده بصورت مناسب توصیف و مصور سازی شده و رفتار هر یک از این متغیرها و ارتباطشان با هم دیگر را بررسی کنیم.

۴-۳-۲ تحلیل و مقایسه فروش برندهای مختلف در مجموعه داده

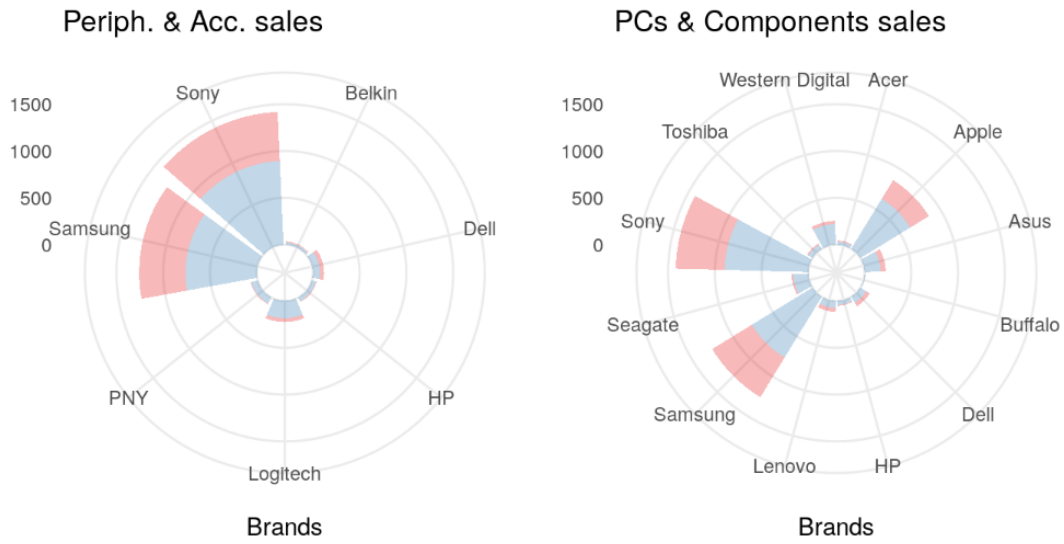
دو شاخه اصلی از برندهای مختلف این مجموعه داده بصورت زیر می‌باشند:

۱. برندهای مربوط به فروش لوازم جانبی کالاهای دیجیتال:

Dell, HP Inc., Logitech, TRANSCEND, Samsung, SanDisk, Sony, Belkin, PNY

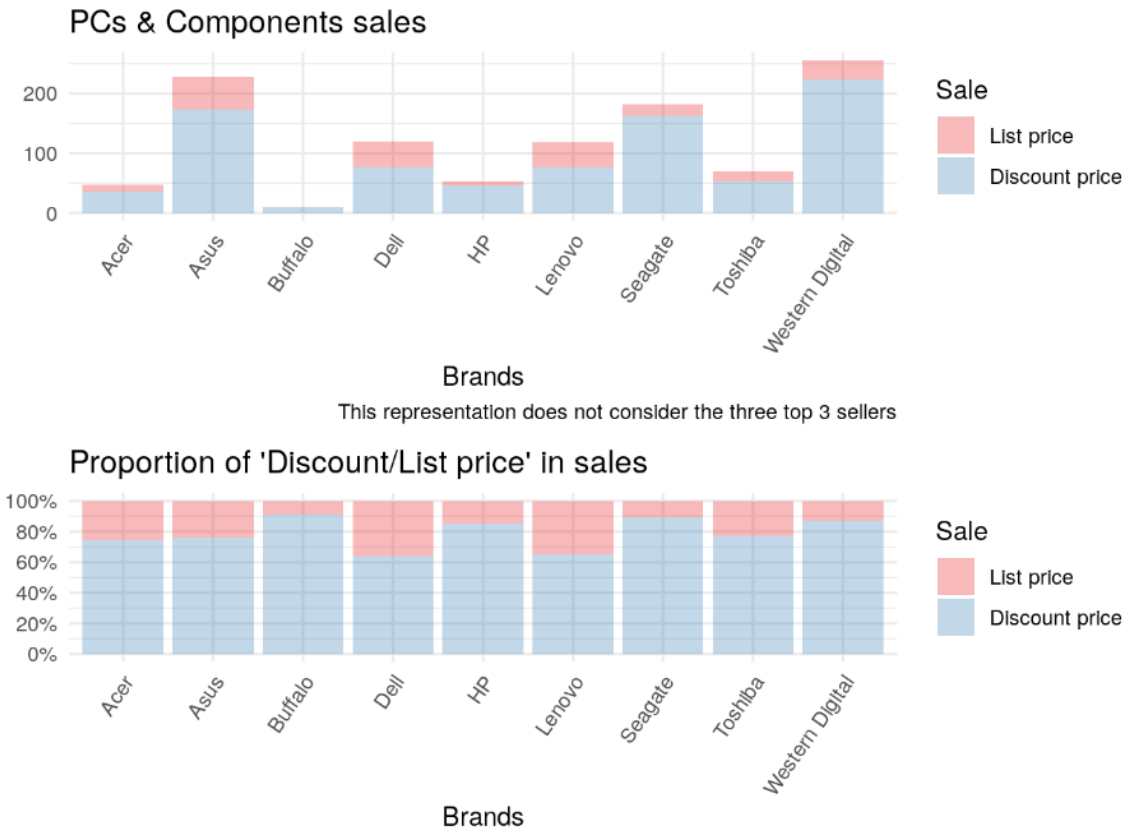
۲. برندهای مربوط به فروش کامپیوتر و سخت افزار:

Acer, Apple, Asus, Dell, HP Inc, Lenovo, Samsung, Seagate Technology, Sony, Western Digital, Buffalo Technology, Toshiba. در شکل ۴-۱ اطلاعات مربوط به مقایسه میزان فروش و همچنین میزان کالاهای تخفیف خورده و بدون تخفیف برای برندهای مختلف بصورت جداگانه برای هر دو دسته با هم مقایسه شده اند.



شکل ۴-۱: مقایسه فروش و میزان کالاهای تخفیف خورده و بدون تخفیف برای برندهای مختلف

مطابق با شکل ۴-۱ برندهای سونی، سامسونگ و اپل بیشترین میزان فروش کالاها را در بین سایر برندها دارا می‌باشند و خارج از مقیاس برندهای دیگر می‌باشند. همچنین مشهود است که برندهای سونی و سامسونگ علاوه بر کامپیوتر و اجزای آن، اکسسوری‌های خانگی و شخصی هم عرضه می‌کند اما برند اپل بازار و تکنولوژي مخصوص به خود را داشته و دستگاه‌های تخصصی خود مانند گوشی‌های هوشمند را دارد. در شکل ۴-۲ دو نوع نمودار میله‌ای مختلف به منظور مقایسه بین برندهای مختلف (غیر از سه برند خارج از مقیاس ذکر شده) از نظر فروش و همچنین رقابت آن‌ها در فروش و تخفیف کالاها، قابل مشاهده و بررسی است.

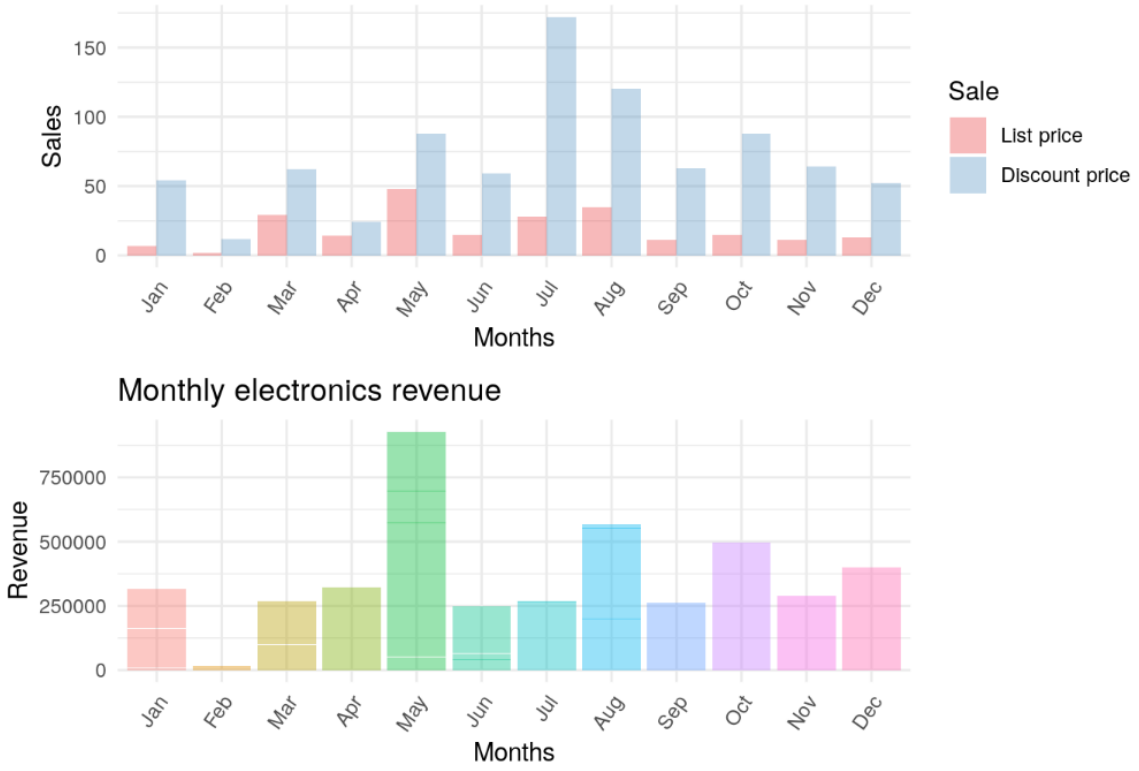


شکل ۴-۲: مقایسه بین برندهای مختلف از نظر فروش و همچنین رقابت آن‌ها در فروش و تخفیف کالاها

مطابق با نمودار میله‌ای نخستین مصور شده در شکل ۴-۲ مشهود است برندهای Western Digital و Asus بیشترین میزان فروش و برند Buffalo کمترین میزان فروش را نسبت به سایر برندها دارا هستند. همچنین مطابق با دومین نمودارمیله‌ای که در شکل مذکور مصور شده، برندهای Dell و Lenovo بیشترین و برندهای Buffalo ، Seagate و Western Digital کمترین میزان درصد تخفیف را در کل فروش‌هایی که داشته‌اند نسبت به سایر برندها دارا می‌باشند .

۴-۳-۳ تحلیل و مصورسازی میزان فروش و تخفیف بصورت ماهیانه

در شکل ۴-۳ نیز میزان کل فروش به ازای کالاهای با تخفیف و بدون تخفیف و همچنین درآمد برای تمامی برندها را بصورت ماهانه در بازه زمانی ۱۲ ماهه، به صورت نمودار میله‌ای سری زمانی، مصورسازی شده‌اند.



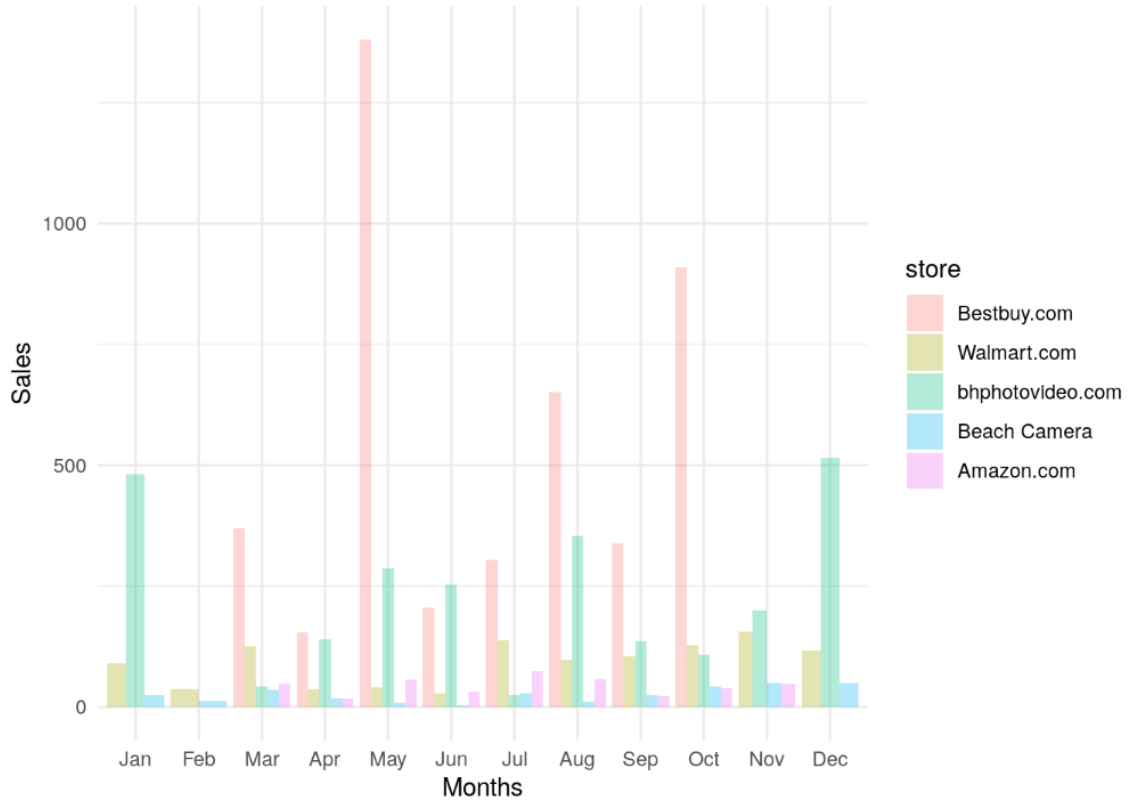
شکل ۴-۳: میزان کل فروش بدون تخفیف و با تخفیف برندها و همچنین درآمد کل بصورت ماهانه

مطابق با نمودارهای مصور شده در شکل ۴-۳ مشهود است که بیشترین میزان درآمد در ماه May بیشترین درآمد کل ماهیانه و در ماه‌های جولای و آگوست که شروع فصل مدارس محسوب می‌شوند بیشترین میزان فروش به‌خصوص فروش با تخفیف ثبت شده است.

۴-۳-۴ تحلیل و مقایسه فروش و توزیع آن در فروشگاه‌های موجود در داده‌ها

در شکل ۴-۴ نمودار میله‌ای مربوط به میزان فروش ماهیانه‌ی فروشگاه‌های مختلف از میان مجموعه داده مصور شده است. مشاهده می‌شود که فروشگاه Best Boy در اکثر ماه‌ها بیشترین فروش را نسبت به سایر فروشگاه‌ها داشته است.(به‌خصوص در ماه‌های می، آگوست و اکتبر که نسبت به سایر فروشگاه‌ها مقیاس فروش بسیار بزرگتری را داراست).

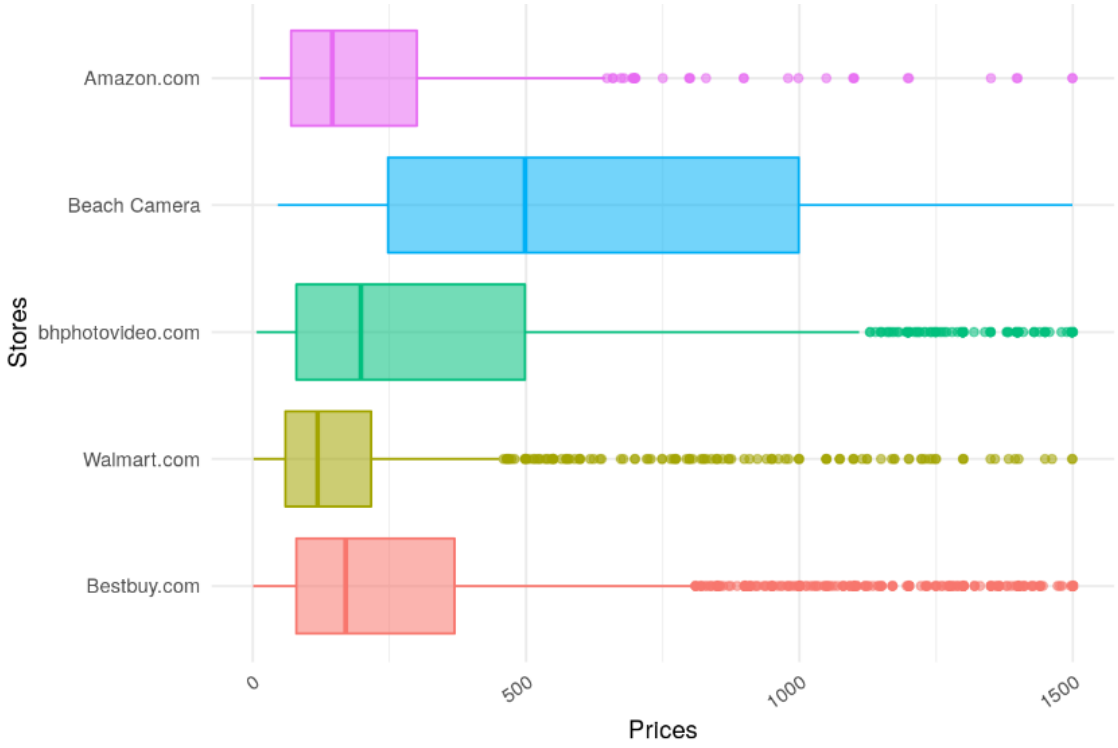
نکته قابل توجه این است که سطح فروش اینترنتی لوازم الکترونیکی در فروشگاه اینترنتی آمازون بسیار کمتر از رقبای خود است. به نظر می رسد که فروش فیزیکی در بین مشتریان هنگام خرید لوازم الکترونیکی ترجیح داده می شود.



شکل ۴-۴: نمودار میله‌ای مربوط به میزان فروش ماهیانه‌ی فروشگاه‌های مختلف

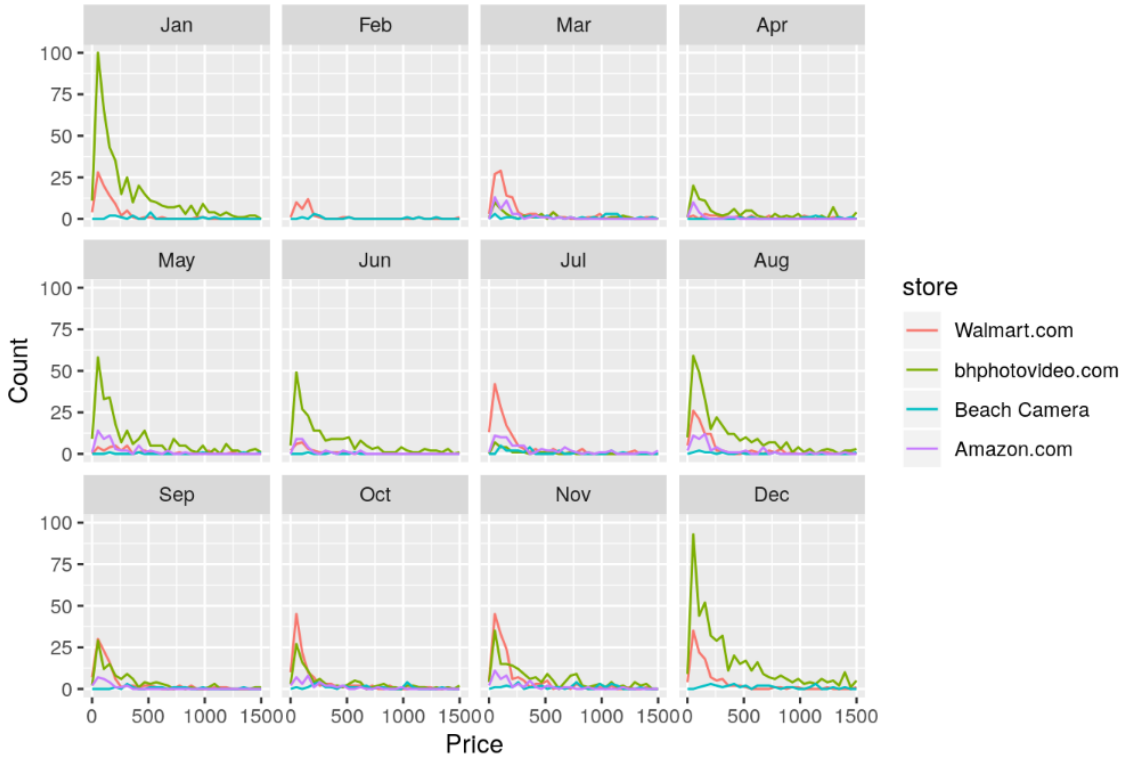
حال مشاهده خواهیم کرد که در بین فروشگاه‌های مختلف توزیع متغیر قیمت به چه صورت می‌باشد. در

شکل ۴-۵ توزیع میزان فروش در فروشگاه‌های مختلف مجموعه داده مصورسازی شده است.

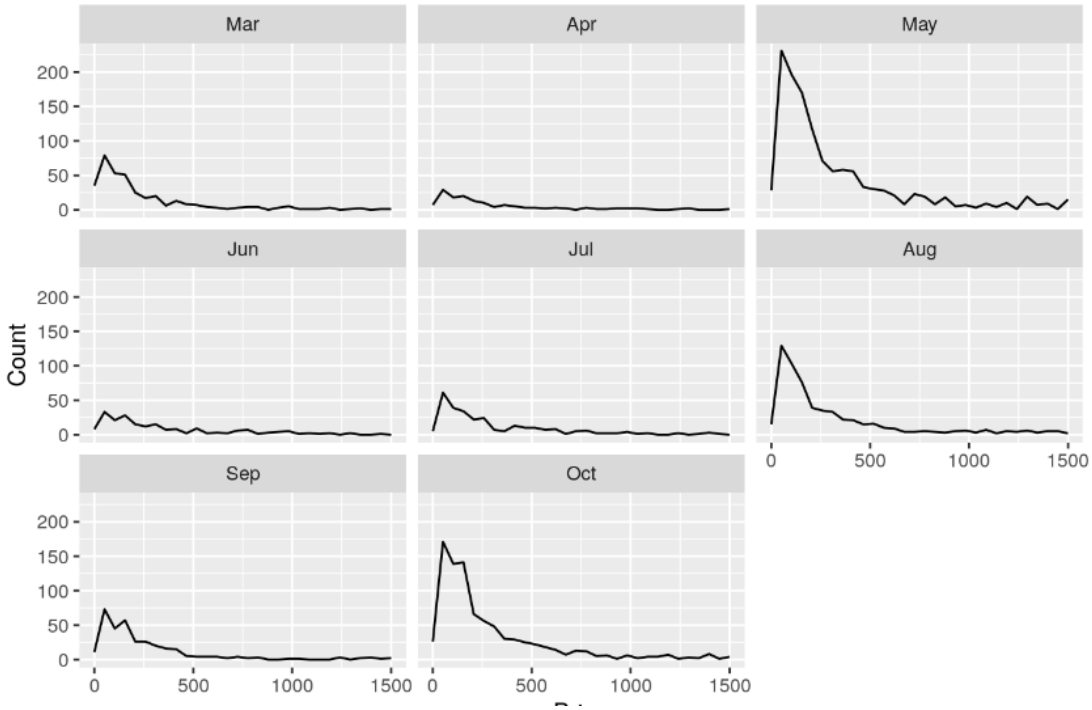


شکل ۴-۵: توزیع میزان فروش در فروشگاه‌های مختلف مجموعه داده

مطابق با شکل ۴-۵ مشاهده می‌شود که فروشگاه Beach Camera گران ترین محصولات خریداری شده را دارند و پس از آن (با فاصله زیاد) فروشگاه BH Photovideo قرار دارد. توزیع عمده بقیه فروشگاه‌ها در لوازم الکترونیکی با قیمت پایین می‌باشد. همچنین در شکل ۴-۶ مقایسه توزیع میزان فروش بر اساس قیمت در ماه‌های مختلف هر فروشگاه قابل مشاهده و مقایسه می‌باشد. (به دلیل مقیاس بالای فروش در فروشگاه Best Boy در شکل ۴-۷ بصورت جداگانه مصورسازی شده است.)



شکل ۴-۶: توزیع میزان فروش بر حسب قیمت قیمت در ماه‌های مختلف هر فروشگاه



شکل ۴-۷: توزیع میزان فروش بر حسب قیمت ماه‌های مختلف در فروشگاه Best Boy

همانطور که انتظار می‌رفت، در قیمت پایین مستقل از ماه فروش و فروشنده، میزان فروش زیاد است.

۴-۳-۵ شاخص‌های آماری متغیرهای عددی

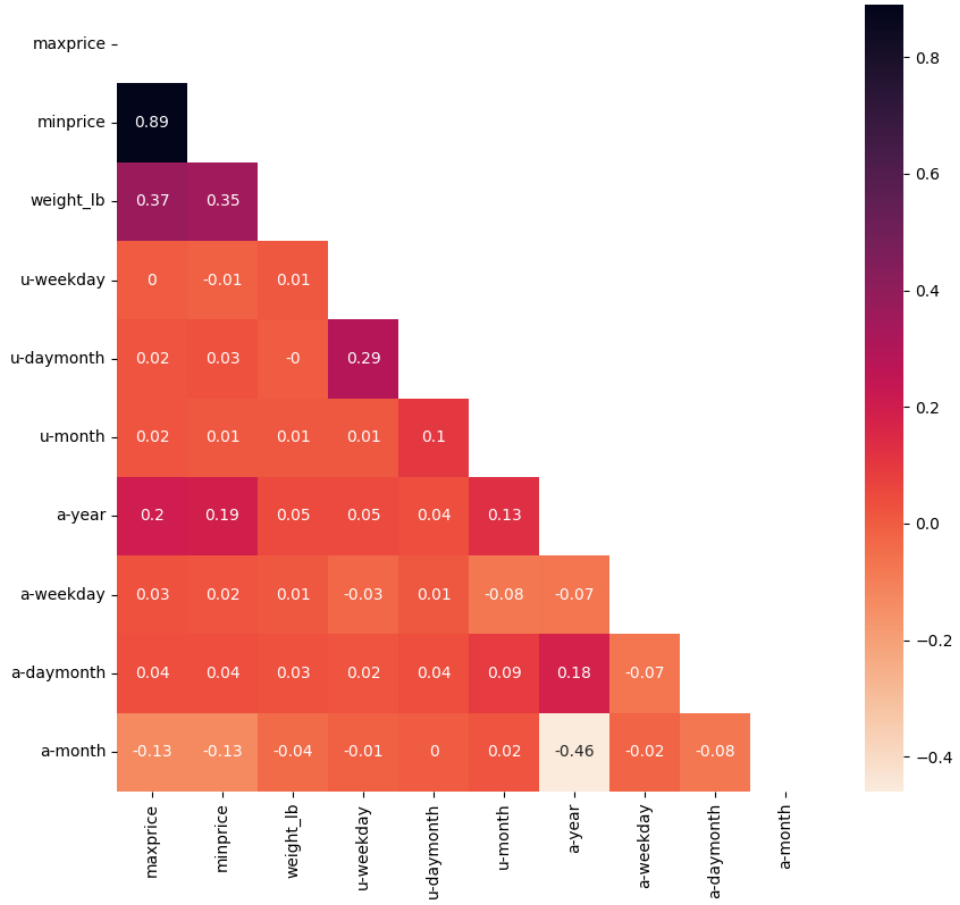
شاخص‌های آماری متغیرهای عددی مجموعه داده به‌صورت جدول ۴-۲ می‌باشد.

جدول ۴-۲: مقادیر شاخص‌های آماری متغیرهای عددی در مجموعه داده

شاخص آماری	حداکثر قیمت	حداقل قیمت	وزن
تعداد	۱۴۱۰۰	۱۴۱۰۰	۱۴۱۰۰
میانگین	۵۰۸٫۳	۴۷۴	۸٫۸
انحراف معیار	۷۹۰	۷۲۱٫۲	۱۷٫۱
حداقل مقدار	۱	۱	۲
چارک اول	۹۳	۸۵	۰٫۸۶
میانه	۲۱۸	۳.۲۱۲	۲
چارک سوم	۵۴۹	۴۸۷	۷
حداکثر مقدار	۲۰۰۰۰۰	۲۰۰۰۰۰	۳۵۰

۴-۳-۶ همبستگی بین متغیرهای عددی

ماتریس همبستگی بین متغیرهای عددی حداکثر قیمت، حداقل قیمت، وزن و متغیرهای مربوط به ماه و سال و روز فروش محصول از مجموعه داده در شکل ۴-۸ مصورسازی شده است.



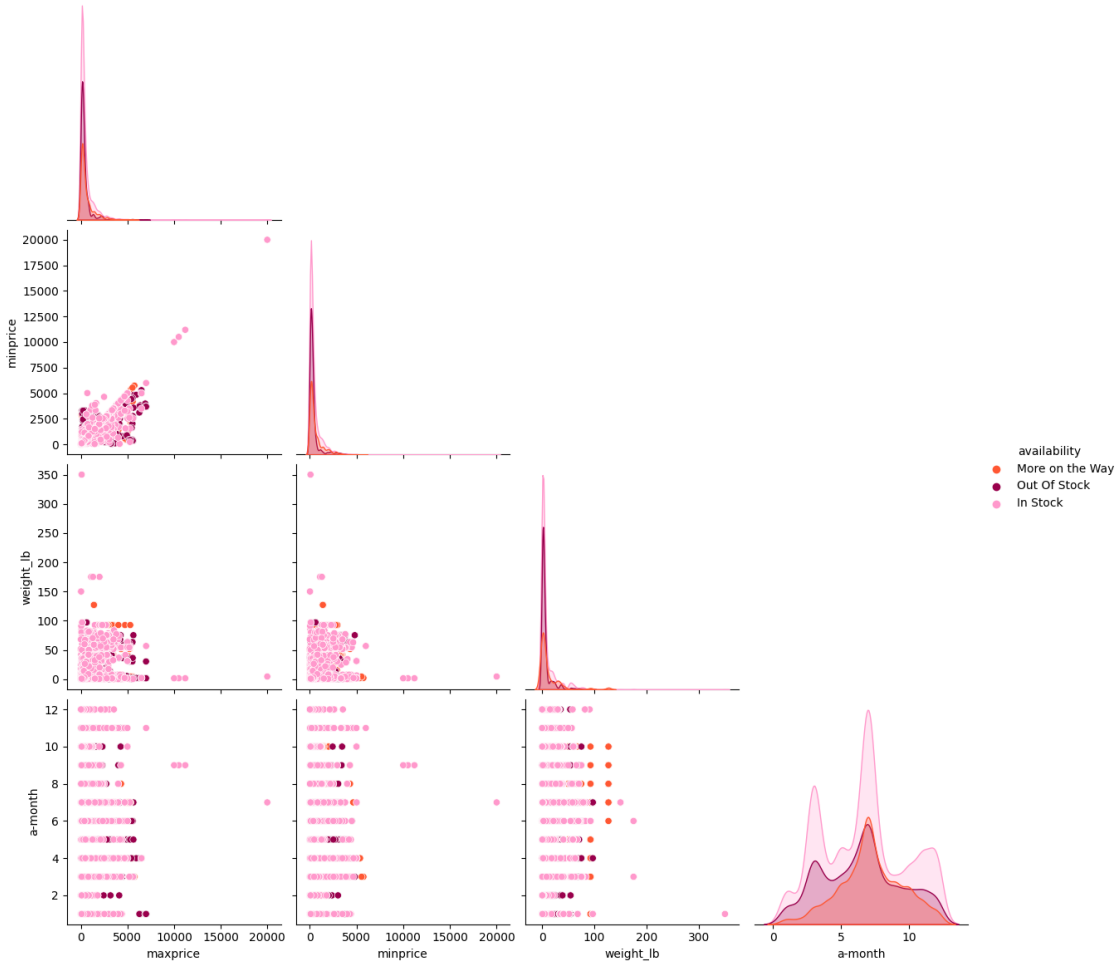
شکل ۴-۸: ماتریس همبستگی بین متغیرهای عددی

همانطور که از ماتریس همبستگی در شکل ۴-۸ پیداست مطابق با انتظار همبستگی بین متغیرهای maxprice و minprice مقدار مثبت و زیاد ۸۹٪ می‌باشد، به این معنی که اگر مقدار یکی از متغیرها افزایش/کاهش یابد، مقدار دیگری هم افزایش/کاهش می‌یابد که نرخ این همبستگی ۰/۸۹ است. بعد از آن همبستگی بین متغیر وزن و قیمت نسبت به همبستگی سایر متغیرها با هم بیشتر بوده و تقریبا ۳۶٪ (میانگین بین همبستگی حداقل قیمت و حداکثر قیمت با متغیر وزن) می‌باشد، به این معنی که اگر مقدار یکی از متغیرها افزایش/کاهش یابد، مقدار دیگری هم افزایش/کاهش می‌یابد که نرخ این همبستگی ۰/۳ است.

از دیگر همبستگی‌های تقریبا معنی‌دار، همبستگی میان a-month و price می‌باشد. میزان همبستگی بین این دو ویژگی ضعیف بوده و تقریباً ۱۳ (میانگین بین همبستگی حداقل قیمت و حداکثر قیمت با متغیر ماه اضافه شدن کالا) و منفی می‌باشد. به این معنی که با افزایش/کاهش ماه اضافه شدن کالا به انبار، قیمت آن کاهش می‌یابد که نرخ این همبستگی ۰/۱۳ می‌باشد.

نمودار دیگری که برای نشان دادن نحوه همبستگی بین متغیرها را به‌خوبی نشان می‌دهد گراف همبستگی موسوم به نمودار جفتی[۳] است. این نمودار برای متغیرهای عددی در شکل ۴-۹ مصورسازی شده است. در این گراف، توزیع هر کدام از متغیرها نیز رسم شده است.

[3]Pair Plot

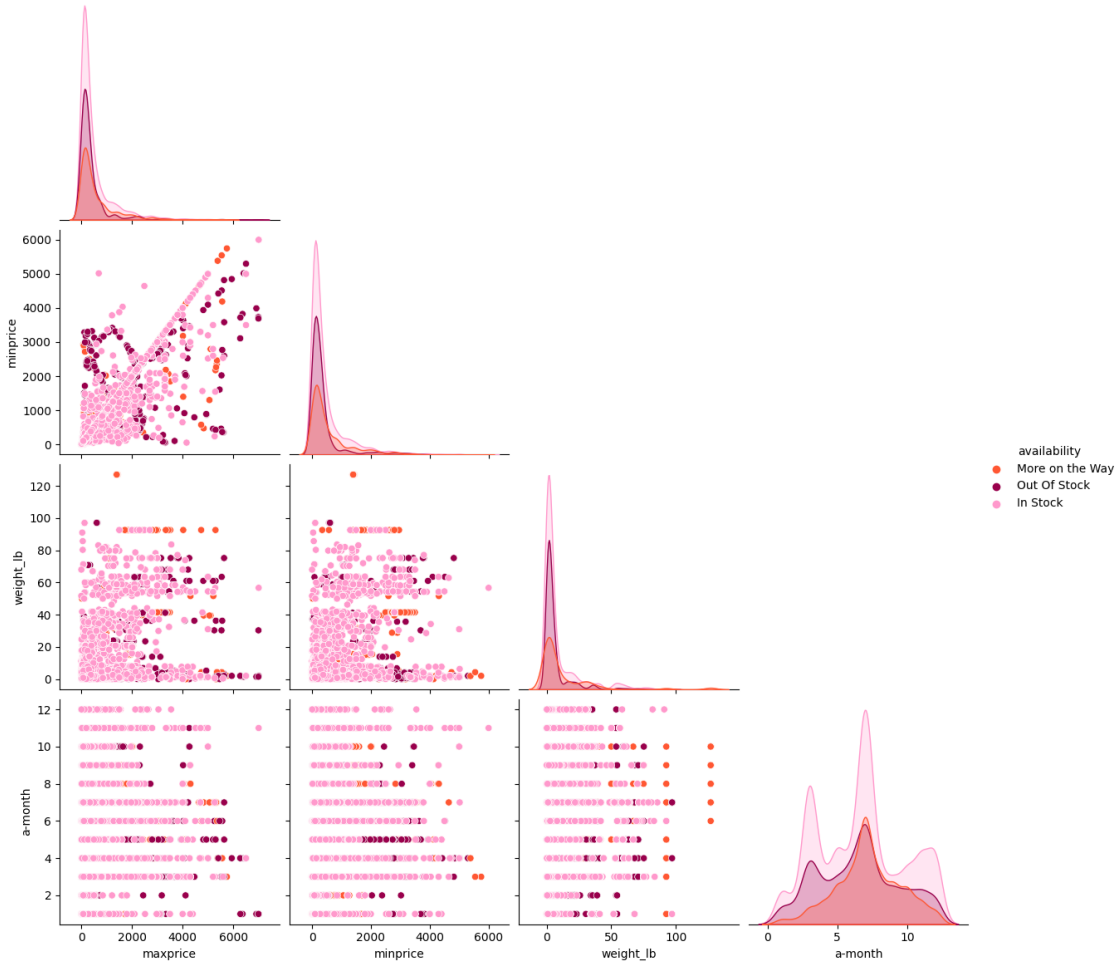


شکل ۴-۹: توزیع متغیرهای عددی و همبستگی بین آن‌ها

همانطور که از گراف همبستگی مصورشده در شکل ۴-۹ پیداست همبستگی خطی و مثبت قوی بین متغیرهای maxprice و minprice وجود دارد ضمن این‌که بین متغیرهای price و weight رابطه همبستگی خطی ضعیف و مثبت وجود دارد. همچنین توزیع متغیرها در شکل نیز مشهود است. لازم به ذکر است به دلیل وجود داده پرت، توزیع متغیرها و همچنین رابطه همبستگی بین آن‌ها از نظر ظاهری و مقیاس به‌خوبی واضح نیست. در گام بعدی پس از رفع داده‌های پرت، گراف همبستگی مجددا مصورسازی شده و گراف مذکور دقیق‌تر مورد بررسی قرار می‌گیرد.

۴-۳-۷ رفع داده‌های پرت

در این مرحله بایستی داده‌های پرت موجود در متغیرهای عددی را حذف و یا جایگزین کنیم. با توجه به اینکه معیار آماری میانه نسبت به داده‌پرت مقاوم است داده‌های پرت موجود در متغیرهای عددی قیمت و نیز وزن، با میانه جایگزین می‌شوند. روش تشخیص این داده‌های پرت از طریق پارامتر $IQR = (Q_۳ - Q_۱)$ است به این صورت که داده‌هایی که مقادیر آن‌ها از $Q_۳ + \alpha * IQR$ بیشتر و با کمتر از $Q_۱ - \alpha * IQR$ باشد، به‌عنوان داده پرت تشخیص داده شده و با مقدار میانه جایگزین می‌شوند (مقدار α را معمولا برابر با ۱٫۵ در نظر می‌گیرند). به این صورت داده‌های پرت برای متغیرهای عددی جایگزین میانه هر کدام شدند. حال گراف دوتایی توزیع متغیرهای عددی به همراه همبستگی بین آن‌ها در شکل ۴-۱۰ مصورسازی شده است.

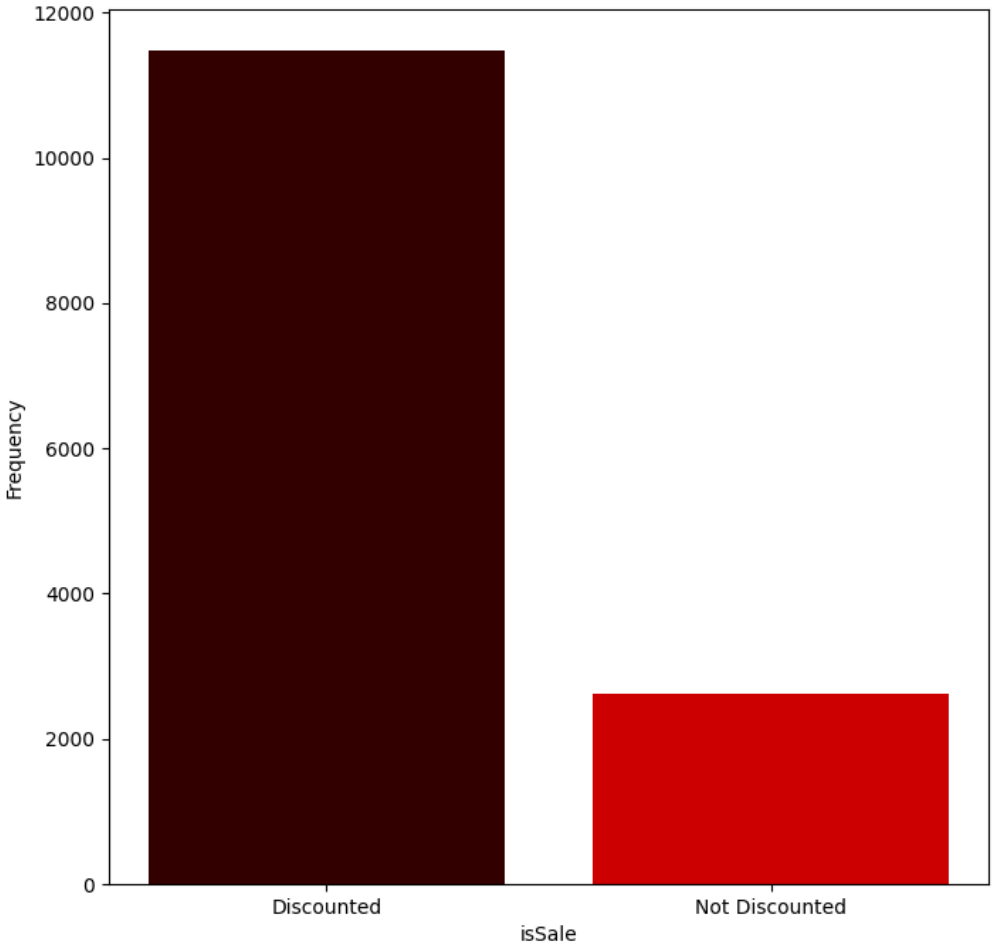


شکل ۴-۱۰: توزیع متغیرهای عددی و همبستگی بین آن‌ها بعد از حذف تاثیر منفی داده‌های پرت

با توجه به شکل ۴-۱۰ مشاهده می‌شود با جایگزینی داده‌های پرت با مقدار میانه آماری هر متغیر مربوط به خودشان، مصورسازی داده‌ها به‌صورت بهتری قابل تفسیر می‌باشد. مشاهده می‌شود توزیع قیمت و وزن برای هر سه کلاس مربوط به موجودی کالا، دارای چولگی سمت راست می‌باشند. همچنین با توجه به توزیع رسم شده در توزیع ماه‌افزوده شدن کالا به انبار، مشاهده می‌شود که کالاهای با برچسب موجودیت More On the way بصورت تقریبی توزیع نرمال با یک مود را دارد با این معنی که کالاهای با این برچسب در اطراف ماه‌های وسط سال فراوانی بیشتری نسبت به ماه‌های اولیه و پایانی دارند. همچنین کالاهای با بر چسب موجودی In stock و Out of stock دارای سه مود بوده و فراوانی آن‌های در ماه‌های اول (۳و۴) ماه‌های میانی و ماه آخر سال بیشتر

مقدار بوده است. همچنین همبستگی مثبت و قوی بین حداکثر قیمت و جداقل قیمت کالا در شکل مذکور بیشتر نمایان است.

پیش‌تر در مقایسه بین برندها و فروشگاه‌های مختلف در مجموعه داده نمودار میله‌ای مربوط به فراوانی متغیرهای کیفی برند و فروشگاه کالای مربوطه مصورسازی شدند. نمودار میله‌ای مربوط به فراوانی داده‌ها مربوط به متغیر IsSale نیز که تخفیف یا عدم تخفیف کالا را نمایندگی می‌کند در شکل ۴-۱۱ مصورسازی شده است.

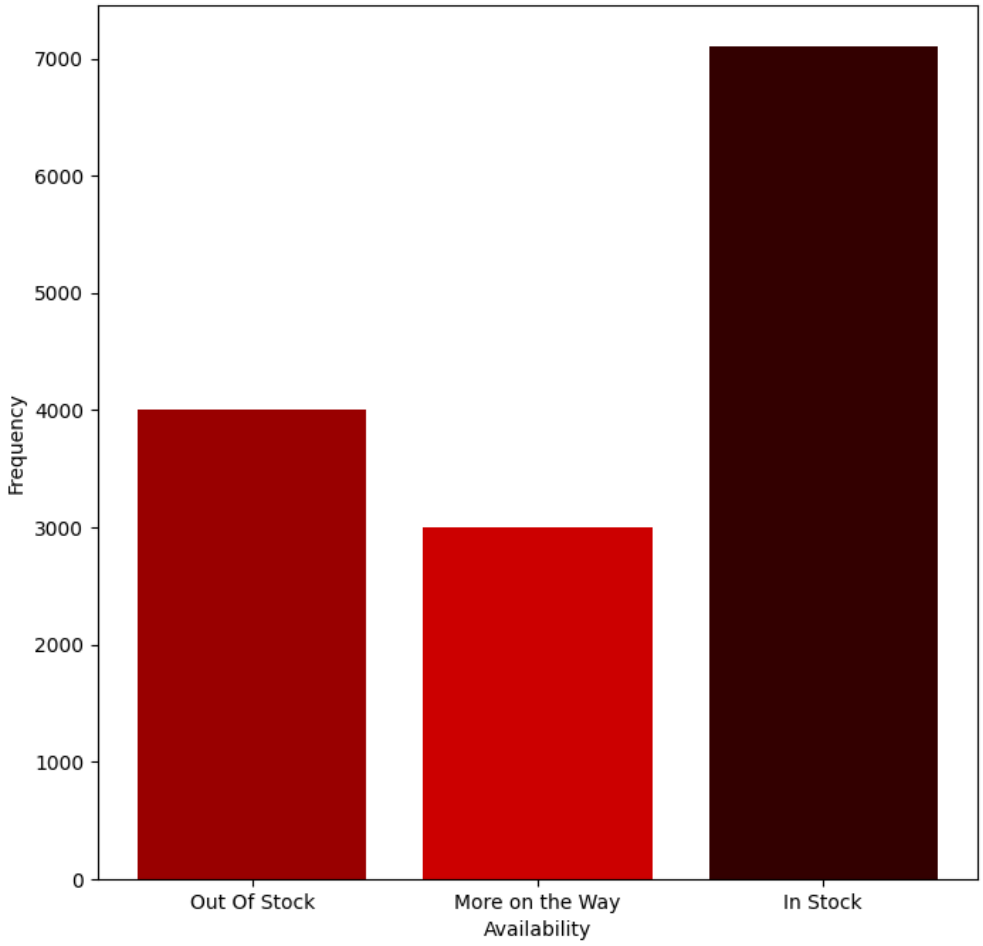


شکل ۴-۱۱: نمودار میله‌ای مربوط به فراوانی متغیر IsSale

مطابق با این نمودار، واضح است که کالاهای بدون تخفیف فراوانی بسیار بیشتری نسبت به کالاهای تخفیف خورده دارند.

۴-۳-۸ تحلیل توصیفی و مصورسازی متغیر کلیدی

تا اینجای کار متغیرهای مستقل را مورد بررسی و تحلیل قرار داده شده اند. در ادامه متغیر کلیدی Availability را مورد برررسی قرار می‌دهیم. نمودار میله‌ای مربوط به فراوانی متغیر مذکور در شکل ۴-۱۲ مصورسازی شده است.



شکل ۴-۱۲: نمودار میله‌ای مربوط به فراوانی متغیر موجودیت کالا

) همانطور که انتظار می‌رفت مشاهده می‌شود که کالاهای موجودی در انبار نسبت به سایر کلاس‌ها بیشتر است.

۴-۴ نتایج حاصل از مدل‌سازی و تحلیل پیش‌بین

همانطور که پیش‌تر ذکر شد هدف از این پژوهش ارائه الگوریتم مجتمع پویا مبتنی بر هوش مصنوعی و یادگیری ماشینی برای تشخیص موجودی در انبار است. در ادامه نتایج حاصل از پیاده سازی الگوریتم‌های یادگیری ماشینی و همچنین روش پیشنهادی،بصورت عددی و همچنین مصور شده ارائه می‌گردد.

۴-۴-۱ نتایج پیاده‌سازی طبقه‌بندهای واحد

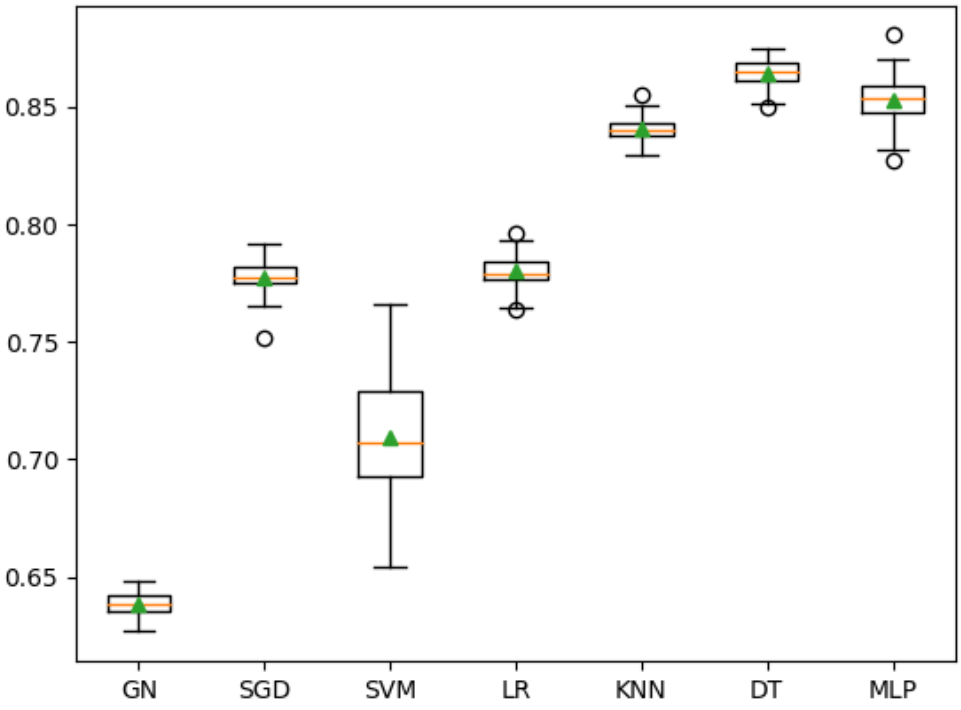
در ادامه نتایج مربوط به طبقه‌بندهای مرسوم که به صورت تکی بر روی داده‌های برازش و آزمایش شده‌اند آورده می‌شود. لازم به ذکر است فرآیند تقسیم داده‌ها به داده‌های آموزش و آزمایش، بصورت تصادفی انجام می‌شود تا داده‌ها بصورت تصادفی قسمت شوند به همین دلیل در هر بار اجرا شدن برنامه و استفاده از مدل، ممکن است نتایج تغییر کنند. پس با استفاده از به منظور رفع این مشکل، از روش K-fold استفاده می‌شود. به این ترتیب که داده‌ها به‌صورت تصادفی به k قسمت تبدیل شده و در هر بار اجرا شدن مدل، یکی از این قسمت‌ها به‌عنوان داده آزمایش و بقیه به‌عنوان داده آموزش در نظر گرفته شده و از نتایج در نهایت میانگین/میانگین وزن‌دار گرفته شده و به عنوان نتیجه نهایی در نظر گرفته می‌شود. همچنین می‌توان این عمل را به تعداد بیشتری انجام داد تا به توزیع بهتری از نتایج دست پیدا شود.

در این پژوهش تعداد k را برای K-fold Cross Validation برابر با پنج و تعداد تکرار عملیات K-fold را برابر با هفت در نظر می‌گیریم. نتایج مربوط به طبقه‌بندهای Gaussian NB ، SGD ، SVM ، Logistic Regression ، K-Neighbors

Decision-Tree و MLP در شکل ۴-۱۳ بصورت نمودار جعبه‌ای مصورسازی شده است. در این شکل توزیع صحت آزمایشی هر کدام از طبقه‌بندها قابل مشاهده است. همچنین میانگین و میانه برای هر کدام از آن‌ها به ترتیب به رنگ‌های سبز و نارنجی مشخص شده است. از نتایج مشخص است که مدل درخت تصمیم دارای بایاس کمتری نسبت به سایر طبقه‌بندها است. همچنین طبقه‌بند Gaussian NB دارای واریانس پایین اما بایاس بسیار زیاد نسبت به سایر است. در جدول ۴-۳ اطلاعات عددی مربوط به میانه و میانگین توزیع صحت در طبقه‌بندها قابل مشاهده است.

جدول ۴-۳: میانه و میانگین توزیع صحت در طبقه‌بندهای مختلف

نوع طبقه‌بند	میانگین توزیع صحت	واریانس توزیع صحت
Gaussian NB	۰/۶۳۹	۰/۰۰۵
SGD	۰/۷۷۷	۰/۰۰۸
Logistic Regression	۰/۷۸۰	۰/۰۰۷
KNN	**۰/۸۴۱**	۰/۰۰۵
Decision-Tree	**۰/۸۶۴**	۰/۰۰۶
SVM	۰/۷۰۹	۰/۰۲۴
MLP	۰/۸۵۳	۰/۰۱۱



شکل ۴-۱۳: نمودار جعبه‌ای توزیع صحت آزمایش، در طبقه‌بندها

۴-۴-۲ نتایج پیاده‌سازی طبقه‌بندهای مجتمع ایستا

در ادامه نتایج مربوط به طبقه‌بندهای مرسوم که به صورت مجتمع ایستا بر روی داده‌های برازش و آزمایش شده‌اند آورده می‌شود. در اینجا هم از K-fold Cross Validation استفاده می‌شود که مقدار K برابر با ۵ و تعداد تکرار برابر با ۷ در نظر گرفته شده است. طبقه‌بندهای تکی که به‌عنوان طبقه‌بندهای پایه برای طبقه‌بندی مجتمع در نظر گرفته شده‌اند عبارتند از:

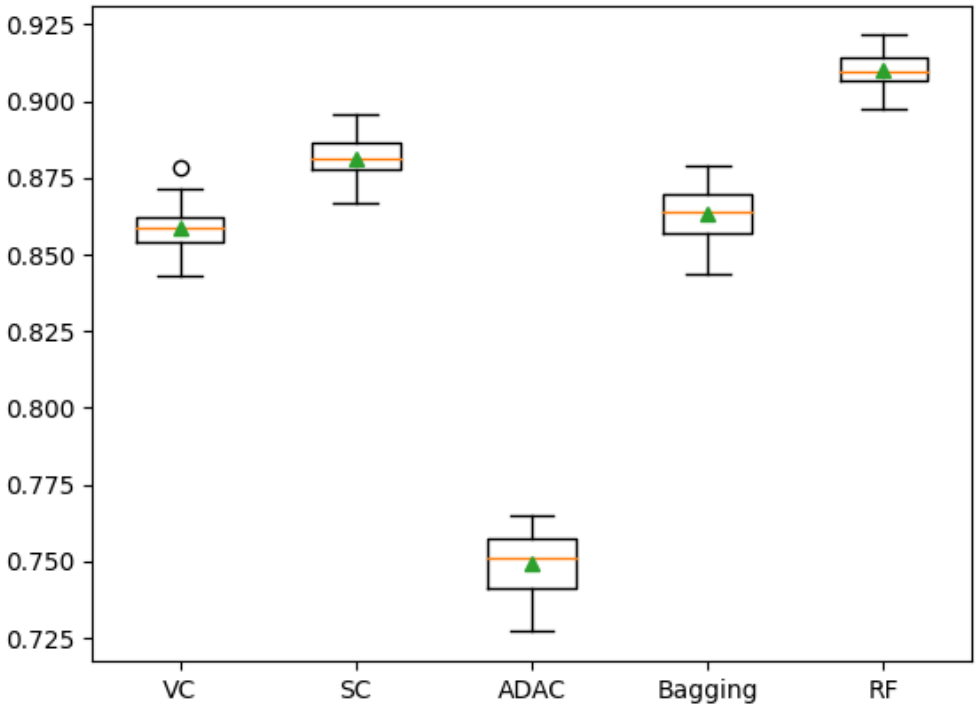
- Logistic Regression
- K-Neighbors
- SGD
- Decision-Tree

نتایج مربوط به طیقه‌بندهای مجتمع ایستای:
Voting ، Stacking ، AdaBoost ، Bagging ، K-Neighbors و Random Forest در شکل ۴-۱۴ بصورت نمودار جعبه‌ای مصورسازی شده است. در جدول ۴-۴ اطلاعات عددی مربوط به میانه و میانگین توزیع صحت در طبقه‌بندهای مجتمع ایستا قابل مشاهده است.

جدول ۴-۴: میانه و میانگین توزیع صحت در طبقه‌بندهای مجتمع مختلف

نوع طبقه‌بند مجتمع ایستا	میانگین توزیع صحت	واریانس توزیع صحت
Stacking	**۰٫۸۸۱**	۰٫۰۰۶
AdaBoost	۰٫۷۵۰	۰٫۰۰۸۹
Bagging	۰٫۸۶۳	۰٫۰۰۹
RF	**۰٫۹۱۰**	۰٫۰۰۶

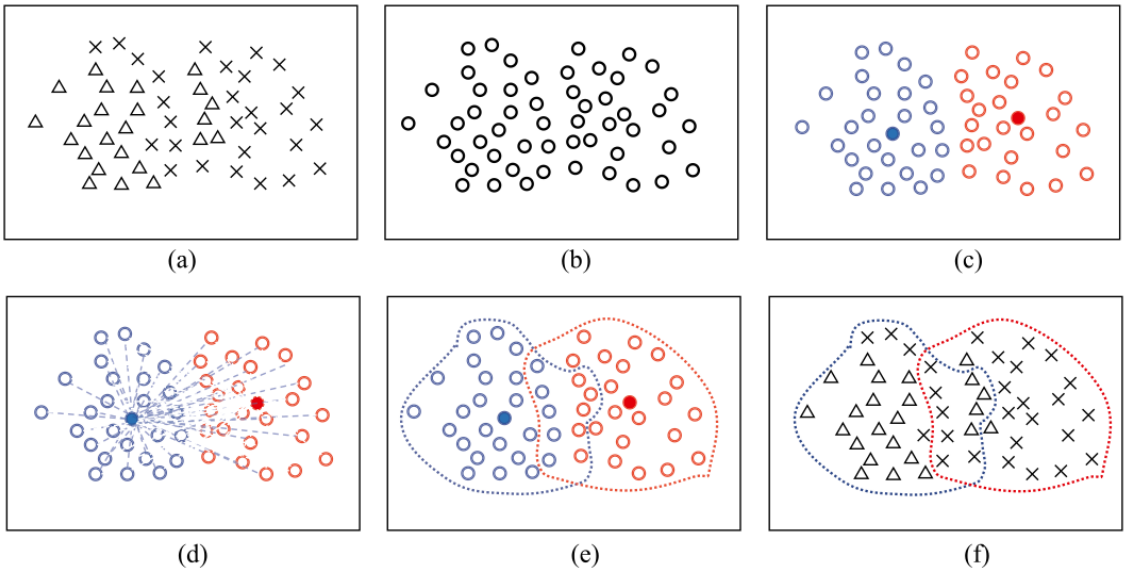
با توجه به نتایج مشاهده می‌شود که عملکرد کلی روش‌های مجتمع ایستا نسبت به مدل‌های تکی به مراتب بهتر می‌باشد. همچنین روش مجتمع ایستای RF بهترین عملکرد را نسبت به سایر مدل‌های مجتمع ایستا دارا می‌باشد.



شکل ۴-۱۴: نمودار جعبه‌ای توزیع صحت آزمایش، در طبقه‌بندهای مجتمع ایستا

۴-۴-۳ نتایج پیاده‌سازی طبقه‌بند پیشنهادی مجتمع پویا

همانطور که از نتایج قسمت قبل پیداست روش‌های ترکیب طبقه‌بند ایستا نسبت به مدل‌های تکی دارای صحت بیشتری بوده و همچنین مقاوم‌تر و با قابلیت اطمینان بیشتری می‌باشند. اما طبقه‌بندی داده‌های تست را بصورت سراسری در فضای ویژگی انجام می‌دهد و برای تمامی داده‌های تست که در فضای ویژگی در هر جایی ممکن حاضر شوند با طبقه‌بند یا طبقه‌بندهای یکسان طبقه‌بندی می‌شوند. با استفاده از طبقه‌بند مجتمع پویا می‌توان طبقه‌بندی را برای هر کدام از داده‌های آزمایشی ورودی شخصی‌سازی کرد به‌طوری که بنا بر جایگیری داده‌ی تست در فضای ویژگی، بصورت محلی طبقه‌بندی شود. در این نوع طبقه‌بند بر اساس فاصله داده‌ی آزمایش به توده‌های داده‌های آموزشی، توسط طبقه‌بندهای مختلف و با تاثیر گذاری مختلف، طبقه‌بندی داده‌ی تست انجام می‌شود. محدوده این توده‌ها با استفاده از خوشه‌بندی مناسب داده‌های آزمایشی تعیین می‌شوند. در [۱۲] روشی مبتنی بر خوشه‌بندی K-means برای تعیین توده‌های مورد نظر استفاده شده است که در این پژوهش برای ساختن توده‌های مذکور از آن استفاده می‌شود. در شکل ۴-۱۵ مراحل انجام این فرآیند در دو بعد قابل مشاهده است.



شکل ۴-۱۵: مراحل ایجاد توده‌های محلی با استفاده از روش خوشه‌بندی K-means

در واقع مراحل تعیین توده‌های مذکور به‌صورت زیر است:

- (a): آماده‌سازی داده‌ها و نمایش آن‌ها با برچسب مختص هر کدام
- (b): حذف برچسب داده‌ها برای انجام خوشه‌بندی
- (c): انجام خوشه‌بندی و نمایش خوشه‌ها به همراه مراکز آن‌ها
- (d): محاسبه فاصله اقلیدسی هر کدام از داده‌ها تا مراکز خوشه‌ها
- (e): نمونه‌برداری برای هر کدام از مراکز خوشه‌ها: اختصاص درصدی از کل‌داده‌هایی که نزدیکترین فاصله را نسبت هر کدام از خوشه‌ها دارند به هر کدام آز آن‌ها
- (f): بازگرداندن برچسب داده‌ها برای مرحله بعدی (طبقه‌بندی)

در مرحله بعدی بایستی به ازای هر کدام از توده‌های مشخص شده، طبقه‌بند پایه آموزش داد. پس برای انجام این روش سه پارامتر اصلی برای طراحی مدل بهینه قابل تنظیم و انتخاب هستند:

- تعداد خوشه‌ها: تعیین تعداد خوشه‌ها برای تعیین توده‌های مورد نظر (K)
- نسبت نرخ نمونه‌برداری: تعیین نرخ نمونه‌برداری از هر خوشه (ϕ)
- تعیین نوع طبقه‌بند پایه (F)

در اینجا این سه پارامتر با استفاده از روش جستجوی شبکه‌ای[۴]، (GS) بصورت بهینه تعیین می‌شوند.

در مرحله آخر با استفاده از فاصله اقلیدسی هر کدام از داده‌های آزمایشی تا هر کدام از مراکز خوشه‌ها، داده‌ی آزمایشی مذکور توسط ترکیبی توسط طبقه‌بندهایی که در مرحله قبل بر روی توده‌ها آموزش داده شده اند طبقه‌بندی می‌شود به این صورت که سهم طبقه‌بندهای اختصاص یافته به توده‌هایی که مراکز آن‌ها به داده آزمایش نزدیک‌تر است بیشتر می‌باشد.

به‌طور کلی ترکیب خروجی طبقه‌بندهایه‌صورت سنتی می‌تواند به چند شکل صورت بگیرد:

- میانگین‌گیری خروجی احتمالاتی هر کدام از طبقه‌بندها (AvgE) با استفاده از این روش خروجی طبقه‌بندها با میانگین‌گیری ترکیب می‌شوند:

[4] Grid Search

$$\boldsymbol{H}(\boldsymbol{e}) = \left(\sum_{\eta=1}^{K} \boldsymbol{h}_{\eta}(\boldsymbol{e}) \right) / K \qquad (۱-۴)$$

که $h(e)$ خروجی احتمالاتی نهایی طبقه‌بندها و K تعداد خوشه‌ها می‌باشد.

- بیشترین مقدار احتمالاتی برای خروجی هر کدام از طبقه‌بندها (MaxE) بیشترین خروجی احتمالاتی طبقه‌بندها به عنوان خروجی نهایی در نظر گرفته می‌شود:

$$\boldsymbol{H}(\boldsymbol{e}) = \max \{ \boldsymbol{h}_{\eta}(\boldsymbol{e}) \mid \eta = 1, 2, \ldots, K \} \qquad (۲-۴)$$

- کمترین مقدار احتمالاتی برای خروجی هر کدام از طبقه‌بندها (MinE) کمترین خروجی احتمالاتی طبقه‌بندها به عنوان خروجی نهایی مخالف کلاس مثبت در نظر گرفته می‌شود:

$$\boldsymbol{H}(\boldsymbol{e}) = \min \{ \boldsymbol{h}_{\eta}(\boldsymbol{e}) \mid \eta = 1, 2, \ldots, K \} \qquad (۳-۴)$$

- در این روش، ترکیب خروجی احتمالاتی طبقه‌بندها بر اساس اجتماع وزن‌دار مساحت زیر منحنی ROC انجام می‌پذیرد:

$$\boldsymbol{H}(\boldsymbol{e}) = \sum_{\eta=1}^{K} \frac{AUROC_{\eta}}{\sum_{j=1}^{K} AUROC_{j}} \boldsymbol{h}_{\eta}(\boldsymbol{e}) \qquad (۴-۴)$$

در این پژوهش با استفاده از روش‌های مختلف ترکیب پویای خروجی:

- MaxE(Maximum Ensemble)
- AvgE(Average Ensemble)
- WAUROCE(Weighted Area Under the ROC Curve Ensemble)

- DistE(Distance-Based Ensemble)

استفاده می‌کنیم. همچنین به منظور استفاده از ویژگی‌های هر دو روش ترکیبی WAUROCE و DistE با میانگین‌گیری وزن‌دار دو بردار وزنی اختصاص یافته به طبقه‌بندهای پایه که از دو روش مذکور بدست آمده‌اند، بردار وزنی نهایی بدست می‌آید. با استفاده از معیار فاصله، مبتنی بر روش K-means خروجی نهایی ترکیبی برای روش DistE(Distance-Based Ensemble) بدست می‌آید. با فرض اینکه هر کدام از توابع موجود در بردار زیر را به‌عنوان طبقه‌بند آموزش داده شده بر روی توده‌ها در نظر بگیریم:

$$h_۱(\cdot), h_۲(\cdot), \ldots, h_K(\cdot) \tag{۴-۵}$$

خروجی ناشی از داده آزمایشی به‌صورت بردار زیر خواهد بود:

$$\left(h_۱\left(\boldsymbol{e_i}\right), h_۲\left(\boldsymbol{e_i}\right), \ldots, h_K\left(\boldsymbol{e_i}\right)\right)^T \tag{۴-۶}$$

به علاوه اگر فاصله داده آزمایشی e_i را از هر کدام از توده‌ها بصورت بردار زیر در نظر بگیریم:

$$\boldsymbol{d_i} = \left(d_{i۱}, d_{i۲}, \ldots, d_{iK}\right)^T \tag{۴-۷}$$

بردار فاصله‌ي استاندارد شده برای هر کدام از داده‌ها به‌صورت زیر می‌باشد:

$$\left(\frac{۱/d_{i۱}}{\sum_{j=۱}^{K} ۱/d_{ij}}, \frac{۱/d_{i۲}}{\sum_{j=۱}^{K} ۱/d_{ij}}, \ldots, \frac{۱/d_{iK}}{\sum_{j=۱}^{K} ۱/d_{ij}}\right)^T \tag{۴-۸}$$

در نهایت خروجی مربوط به داده آزمایش به‌صورت زیر خواهد شد:

$$H\left(\boldsymbol{e_i}\right) = \boldsymbol{w_i}^T \cdot \boldsymbol{h}\left(\boldsymbol{e}_i\right) \tag{۴-۹}$$

تابع هدف در معادله ۴-۱۰ قابل مشاهده است.

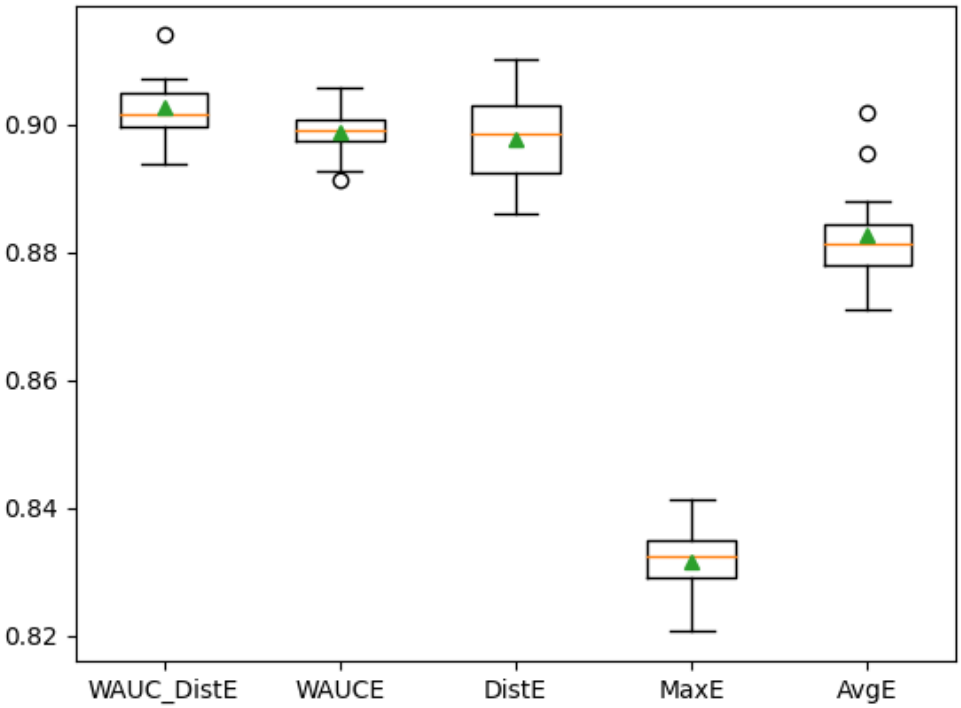
$$\max_{\phi,K,f} E\left(y_i, F\left(\boldsymbol{x_i} \mid \phi, K, F\right) \mid i = ۱, ۲, \ldots, m\right) \qquad (۴-۱۰)$$

که E مقیاس ارزیابی عملکرد کلی مدل، x_i بردار ویژگی y_i برچسب واقعی داده ارزیابی F طبقه‌بند پایه phi نرخ نمونه‌برداری برای هر توده و K تعداد توده‌ها می‌باشند. پس از انتخاب نوع طبقه‌بند پایه، پارامترهای phi، K و F را باید طوری در نظر گرفته شوند که تابع هدف بیشترین مقدار را اخذ کند. در این پژوهش با استفاده از جست‌وجوی شبکه‌ای طبقه‌بند شبکه‌عصبی چندلایه برای طبقه‌بند پایه، K برابر با پنج و phi برابر با ۰٫۵ به‌عنوان پارامترهای بهینه برای حداکثرسازی تابع هدف به‌دست آمدند. به علاوه طبقه‌بندهای شبکه عصبی چندلایه به همراه مدل درخت تصمیم به عنوان طبقه‌بندهای پایه انتخاب شدند. مقایسه نتایج نهایی حاصل از انواع روش‌های ترکیب طبقه‌بند مجتمع، با استفاده از پارامترهای بهینه در قالب نمودار جعبه‌ای در شکل ۴-۱۶ مصورسازی شده است. که میانگین توزیع شاخص ROC-AUC برای روش اجتماع پویای WAUC-DistE برابر با ۰٫۹۰۸ و واریانس آن برابر با ۰٫۰۰۴۶ می‌باشد. در نهایت با مقایسه نتایج مشهود است طبقه‌بند مجتمع پویا با روش پیشنهادی این پژوهش یعنی WAUC-DistE نسبت به سایر طبقه‌بندها دارای عملکرد بهتری می‌باشد. همچنین اطلاعات عددی مربوط به میانه و میانگین توزیع صحت در انواع طبقه‌بندهای مجتمع پویا در جدول ۴-۵ قابل مشاهده است.

جدول ۴-۵: میانه و میانگین توزیع ROC-AUC در طبقه‌بندهای مجتمع مختلف

نوع طبقه‌بند مجتمع ایستا	میانگین توزیع ROC-AUC	واریانس توزیع ROC-AUC
WAUC-DistE	**۰٫۹۰۸**	۰٫۰۰۴۶
WAUCE	۰٫۸۹۹	۰٫۰۰۳۷
DistE	۰٫۸۹۸	۰٫۰۰۷۶
MaxE	۰٫۸۳۲	۰٫۰۰۶
AvgE	۰٫۸۸۳	۰٫۰۰۷۵

همانطور که از توزیع معیار ROC-AUC در شکل ۴-۱۶ قابل مشاهده است، عملکرد روش پیشنهادی برای ترکیب طبقه‌بندی WAUC-DistE بهتر می‌باشد. این روش در واقع ترکیبی از روش‌های DistE و AUROC می‌باشد که پیش‌تر به توضیح آن پرداخته شد.



شکل ۴-۱۶: مقایسه توزیع شاخص ROC-AUC مربوط به انواع طبقه‌بندهای مجتمع پویا

۴-۵ پاسخ به سوالات تحقیق

در این بخش با توجه به نتایج بدست آمده از پیاده‌سازی روش پیشنهادی، به سوالات اصلی مطرح شده در فصل اول پاسخ داده می‌شود که در ادامه به آن پرداخته شده است.

- آیا می‌توان بدون استفاده از ابزارهای گران‌قیمت فیزیکی و با صرف هزینه معقول، رخدادهای عدم موجودی کالا را با قابلیت اطمینان بیشتری بالا کرد؟ در فصل دوم با توجه به مقالات مروری، به مشکلات استفاده از روش‌های سنتی و استفاده از ابزار فیزیکی پرداخته شد. خطاهای انسانی و محدودیت‌های پرسنل در روش‌های سنتی مبتنی بر انسان و همچنین هزینه‌های زیاد استفاده از ابزار فیزیکی از جمله مشکلات این روش‌ها بودند. در این فصل با توجه به نتایج بدست آمده و مقایسه روش پیشنهادی با سایر روش‌های سنتی و فیزیکی، دریافتیم که روش‌های مبتنی بر یادگیری ماشینی و هوش‌مصنوعی روش‌های ارزانتری هستند که قابلیت اطمینان بیشتری نسبت به سایر روش‌های دیگر را دارا می‌باشند.

- آیا استفاده از روش‌های یادگیری ماشینی و هوش‌مصنوعی نسبت به روش‌های سنتی در پیش‌بینی رخدادهای عدم موجودی کالا عمکرد مناسب‌تری دارند؟ همانطور که در پاسخ به سوال قبلی ذکر شد، با توجه به هزینه پایین‌تر و همچنین قابلیت اطمینان بالاتر، این روش‌ها عملکرد بهتری نسبت به سایر روش‌ها دارا می‌باشند.

- آیا می‌توان علاوه بر پیش‌بینی هر چه دقیق‌تر رخدادهای عدم موجودی کالا در زمان مناسب، به تفسیرپذیری مناسبی برای روش و مدل پیشنهادی دست یافت؟ با استفاده از یادگیری مجتمع به جای استفاده از یک مدل پیچیده برای پیش‌بینی رخدادهای عدم موجودی کالا، از یادگیری مجتمع و گروهی مجموعه‌ای از مدل‌های ساده استفاده کرد. که این مدل‌های ساده مانند درخت تصمیم قابلیت تفسیر پذیری بالایی را دارند. در نتیجه کسب و کارها می‌توانند با تفسیر مدل، به درک بهتری نسبت به عملکرد آن در شرایط مختلف پی ببرند.

- آیا استفاده از مدل طبقه‌بند یکتای مبتنی بر یادگیری ماشینی در هر شرایطی عملکرد بهینه دارد؟ خیر. ممکن است حجم داده‌های در دسترس، به اندازه‌ای کافی نباشد که مسئله تحقیق مورد نظر را نمایندگی کنند. به عبارتی دیگر نمونه داده‌های در دسترس ممکن است تمامی اطلاعات توزیع مشاهدات موجود در مسئله را در برنگیرد. در این صورت مدل تکی، حتی با وجود صحت بالا بر روی داده‌های آموزشی و آزمایشی،

بر روی داده‌های دیده نشده[5] عملکردی ضعیف خواهد داشت.

- مزایای استفاده از ترکیب طبقه‌بندها (یادگیری مجتمع) چیست؟
  - افزایش قوام مدل در برابر داده‌های دیده نشده
  - همانطور که از نتایج پیداست، افزایش عملکرد مدل از نظر شاخص ROC-AUC پیش‌بینی و طبقه‌بندی داده‌ها
  - قابلیت تفسیرپذیری بالا
- با چه ترکیبی از طبقه‌بندها به بهینه‌ترین مدل مجتمع می‌توان دست یافت؟ با استفاده از روش جستوجوی شبکه‌ای اجتماع مجموعه‌ای از طبقه‌بندهای KNN برای این پژوهش دارای عملکرد مناسب‌تری نسبت به سایر طبقه‌بندها می‌باشد.
- مدل طبقه‌بند مجتمع پویا چه مزایایی نسبت به طبقه‌بند مجتمع ایستا دارد؟ با استفاده از روش طبقه‌بند مجتمع پویا، می‌توان برای هر داده‌ی آزمایشی بسته به اینکه در کجای فضای ویژگی قرار داشته باشد و چگونگی توزیع داده‌هایی که در همسایگی آن قرار گرفته‌اند، بصورت محلی طبقه‌بندی را انجام داد. در این صورت داده‌ها بسته به قرارگیری آن‌ها در فضای ویژگی طبقه‌بندی شده و عملکرد بهتر خواهد شد.

[5] Unseen

فصل ۵

بحث و نتیجه‌گیری

۵-۱ مقدمه

همانطور که پیش‌تر ذکر شد، مسئله رخداد عدم موجودی کالا یکی از مشکلات حوزه کسب و کارهای فروشگاهی است که می‌تواند پیامدهای منفی را در پی داشته باشد. در این پژوهش به تفصیل درباره علل بوجود آمدن این رخدادها و همچنین تصمیم‌سازی بهینه درباره این رخدادها با استفاده از مدلسازی پیشبین و تشخیصی بهینه پرداخته شده است.

- در فصل اول این پژوهش درباره طرح مسئله تحقیق، انگیزه و هدف از انجام آن و همچنین کلیات تحقیق صحبت شد.
- در فصل دوم مقالات مروری در رابطه با مسئله تحقیق و تمایز کار این پژوهش با آنها به تفصیل مورد بررسی قرار گرفت.
- در فصل سوم به نحوه تشریح، توصیف و پیش‌پردازش داده‌ها و همچنین شیوه و مراحل پیاده‌سازی مدل پیشنهادی بر روی مجموعه داده پرداخته شد.
- درفصل چهارم داده‌ها و نتایج حاصل از پیاده‌سازی مدل پیشنهادی این پژوهش با نمودارها، جداول و شکل‌های متعدد مصور سازی شدند و پاسخ به سوالات تحقیق داده شد.

در این فصل به بحث درباره یافته‌های برداشت شده از نتایج حاصله در فصل گذشته و فایده آنها پرداخته خواهد شد.

۵-۲ جمع‌بندی پژوهش

در فصل قبل با استفاده از معیارهای سنجش استاندارد و مصورسازی توزیع آن‌ها توسط نمودار جعبه‌ای، به مقایسه نتایج مدل‌های مختلف پرداخته شد. مدل طبقه‌بندی مجتمع پویای پیشنهادی در این پژوهش مبتنی بر فاصله اقلیدسی داده‌های آزمایشی را طبقه‌بندی می‌کند. این مدل علاوه بر اینکه برای طبقه‌بندی داده‌های دیده‌نشده عملکرد بهتری را دارد [۸۱] مطابق با نتایج بدست آمده که در فصل قبل ارائه شد، دارای معیار دقت بیشتری است. علاوه بر این قابلیت تفسیرپذیری بیشتری نسبت به مدل‌های پیچیده‌ی تکی دارد.

همچنین با مقایسه نتایج مربوط به پیاده‌سازی مدل طبقه‌بند مجتمع پویای پیشنهادی با مدل‌های مجتمع ایستا، مدل پیشنهادی دارای دقت بهتری نسبت به مدل طبقه‌بند مجتمع ایستا می‌باشد. به دلیل اینکه هر داده آزمایشی را بر اساس جایگیری آن در فضای ویژگی طبقه‌بندی می‌کند، در واقع طبقه‌بندی را برای هر داده آزمایشی بصورت اختصاصی انجام می‌دهد.

۵-۳ نوآوری

با توجه به مطالعات مروری بررسی شده در فصل دوم، علاوه بر مدل‌های سنتی و استفاده از ابزار فیزیکی، اکثر مدل‌هایی که در زمینه رخداد ناموجودی کالا انجام شده است روش‌های یادگیری ماشینی تکی است. به علاوه در موارد کمی هم مدل مجتمع ایستا در این زمینه مورد استفاده قرار گرفته است. در این پژوهش برای نخستین بار مدل مجتمع پویا با رویکرد فاصله اقلیدسی بر روی داده‌های رخداد عدم موجودی کالا پیاده‌سازی شده است.

۵-۴ پیشنهادها

با توجه به نتایج حاصل از پیاده‌سازی سیستم طبقه‌بند مجتمع پویای پیشنهادی بر روی مجموعه‌داده‌ی مربوط به اطلاعات رخداد عدم موجودی کالا، کسب و کارهای فروشگاهی می‌توانند در عوض استفاده از رویکردهای سنتی و استفاده از ابزار فیزیکی گران قیمت (که درباره محدودیت‌ها و مشکلاتشان پیش‌تر صحبت شد) و یا مدل‌های یادگیری ماشینی که بر روی حجم زیاد داده‌ها عملکرد مناسبی ندارند از مدل یادگیری گروهی (مجتمع) استفاده کنند. به این علت که هم از نظر شاخص‌های استاندارد اندازه‌گیری عملکرد، بهتر بوده و هم قابلیت اطمینان بیشتری برای داده‌های دیده نشده دارند. به علاوه اینکه می‌توانند از مدل‌های پایه‌ی ساده در یادگیری گروهی استفاده کرده و مدل را به راحتی تفسیر کنند. در واقع قابلیت تفسیرپذیری یک مدل از آنجایی اهمیت دارد که میتوان با تحلیل مناسب نحوه عملکرد آن بر روی مجموعه‌داده به درک درستی از رخدادها و علل انجام آن رسید و همچنین فهمید که مدل ممکن است در چه مواقعی و بر روی چه داده‌هایی در شرایط خاص پاسخ مناسب ندهد. همچنین از مقایسه نتایج پیداست که مدل یادگیری مجتمع پویا به دلیل طبقه‌بندی اختصاصی نمونه‌ها، دارای عملکرد بهتری نسبت به مدل‌های مجتمع ایستا دارد.

در نتیجه کسب و کارهای فروشگاهی با استفاده از مدل پیشنهادی می‌توانند با تشخیص و پیش‌بینی به‌موقع رخدادهای عدم موجودی، به تصمیم‌سازی بهینه در این حوزه دست یابند. همچنین این مدل قابل استفاده بر روی سایر داده‌ها و حوزه‌هایی است که طبقه‌بندی، پیش‌بینی و تشخیص پدیده یا رخدادی در آن اهمیت داشته باشد.

مراجع

[1] Sharda, Ramesh, Delen, Dursun, and Turban, Efraim. *Business intelligence, analytics, and data science: a managerial perspective*. Pearson, 2016.

[2] Cruz, Rafael MO, Sabourin, Robert, and Cavalcanti, George DC. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, 41:195–216, 2018.

[3] García-Arca, Jesús, Prado-Prado, J Carlos, and Garrido, A Trinidad González-Portela. On-shelf availability and logistics rationalization. a participative methodology for supply chain improvement. *Journal of Retailing and Consumer Services*, 52:101889, 2020.

[4] Bro, Rasmus and Smilde, Age K. Principal component analysis. *Analytical methods*, 6(9):2812–2831, 2014.

[5] Mac, Hieu, Truong, Dung, Nguyen, Lam, Nguyen, Hoa, Tran, Hai Anh, and Tran, Duc. Detecting attacks on web applications using autoencoder. in *Proceedings of the ninth international symposium on information and communication technology*, pp. 416–421, 2018.

[6] McInnes, Leland, Healy, John, and Melville, James. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[7] Basir, Ahmad Vakili, Fesharaki, Mohammad Gholami, and Rowzati, Mohsen. Multivariate logistic regression analysis using multilevel model. *Journal of Biostatistics and Epidemiology*, 4(3):119–123, 2018.

[8] Wan, Alvin, Dunlap, Lisa, Ho, Daniel, Yin, Jihan, Lee, Scott, Jin, Henry, Petryk, Suzanne, Bargal, Sarah Adel, and Gonzalez, Joseph E. Nbdt: neural-backed decision trees. *arXiv preprint arXiv:2004.00221*, 2020.

[9] Chauhan, Vinod Kumar, Dahiya, Kalpana, and Sharma, Anuj. Problem formulations and solvers in linear svm: a review. *Artificial Intelligence Review*, 52(2):803–855, 2019.

[10] Papatheocharous, Efi and Andreou, Andreas S. Size-based software cost modelling with artificial neural networks and genetic algorithms. *Artificial Neural Networks Application*, pp. 168–188, 2011.

[11] Yang, Pengyi, Hwa Yang, Yee, B Zhou, Bing, and Y Zomaya, Albert. A review of ensemble methods in bioinformatics. *Current Bioinformatics*, 5(4):296–308, 2010.

[12] Guo, Chonghui, Liu, Mucan, and Lu, Menglin. A dynamic ensemble learning algorithm based on k-means for icu mortality prediction. *Applied Soft Computing*, 103:107166, 2021.

[13] Andaur, Juan Manuel Rozas, Ruz, Gonzalo A, and Goycoolea, Marcos. Predicting out-of-stock using machine learning: An application in a retail packaged foods manufacturing company. *Electronics*, 10(22):2787, 2021.

[14] Han, Jing, Haihong, Ee, Le, Guan, and Du, Jian. Survey on nosql database. in *2011 6th international conference on pervasive computing and applications*, pp. 363–366. IEEE, 2011.

[15] Barik, Rabindra K, Dubey, Harishchandra, Samaddar, Arun B, Gupta, Rajan D, and Ray, Prakash K. Foggis: Fog computing for geospatial big data analytics. in *2016 IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON)*, pp. 613–618. IEEE, 2016.

[16] Davenport, Thomas and Harris, Jeanne. *Competing on analytics: Updated, with a new introduction: The new science of winning*. Harvard Business Press, 2017.

[17] Chen, Hsinchun, Chiang, Roger HL, and Storey, Veda C. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, pp. 1165–1188, 2012.

[18] López-Robles, José Ricardo, Otegi-Olaso, Jose Ramón, Gómez, I Porto, and Cobo, Manuel J. 30 years of intelligence models in management and business: A bibliometric review. *International journal of information management*, 48:22–38, 2019.

[19] Papakiriakopoulos, Dimitrios, Pramatari, Katerina, and Doukidis, Georgios. A decision support system for detecting products missing from the shelf based on heuristic rules. *Decision Support Systems*, 46(3):685–694, 2009.

[20] Papakiriakopoulos, Dimitris. Developing a mechanism to support decisions for products missing from the shelf. *Journal of decision systems*, 20(4):417–441, 2011.

[21] Papakiriakopoulos, Dimitris. Predict on-shelf product availability in grocery retailing with classification methods. *Expert Systems with Applications*, 39(4):4473–4482, 2012.

[22] Chuang, Howard Hao-Chun, Oliva, Rogelio, and Liu, Sheng. On-shelf availability, retail performance, and external audits: A field experiment. *Production and Operations Management*, 25(5):935–951, 2016.

[23] Montoya, Ricardo and Gonzalez, Carlos. A hidden markov model to detect on-shelf out-of-stocks using point-of-sale data. *Manufacturing & Service Operations Management*, 21(4):932–948, 2019.

[24] Geng, Zelong, Wang, Zhongze, Weng, Tiandong, Huang, Yuhui, and Zhu, Yu. Shelf product detection based on deep neural network. in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6. IEEE, 2019.

[25] Mehta, Rohit Nilesh, Joshi, Hitaishi Vijay, Dossa, Inziya, Yadav, Rahul Gyanch, Mane, Sarika, and Rathod, Mansing. Supermarket shelf monitoring using ros based robot. in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 58–65. IEEE, 2021.

[26] Crăciunescu, Mihai, Baicu, Diana, Mocanu, Ştefan, and Dobre, Cristian. Determining on-shelf availability based on rgb and tof depth cameras. in *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*, pp. 243–248. IEEE, 2021.

[27] Chen, Jun, Wang, Shu-Lin, and Lin, Hong-Li. Out-of-stock detection based on deep learning. in *International Conference on Intelligent Computing*, pp. 228–237. Springer, 2019.

[28] Higa, Kyota and Iwamoto, Kota. Robust shelf monitoring using supervised learning for improving on-shelf availability in retail stores. *Sensors*, 19(12):2722, 2019.

[29] Yilmazer, Ramiz and Birant, Derya. Shelf auditing based on image classification using semi-supervised deep learning to increase on-shelf availability in grocery stores. *Sensors*, 21(2):327, 2021.

[30] Aastrup, Jesper and Kotzab, Herbert. Forty years of out-of-stock research–and shelves are still empty. *The International Review of Retail, Distribution and Consumer Research*, 20(1):147–164, 2010.

[31] Mou, Shandong, Robb, David J, and DeHoratius, Nicole. Retail store operations: Literature review and research directions. *European Journal of Operational Research*, 265(2):399–422, 2018.

[32] Fitzsimons, Gavan J. Consumer response to stockouts. *Journal of consumer research*, 27(2):249–266, 2000.

[33] Moussaoui, Issam, Williams, Brent D, Hofer, Christian, Aloysius, John A, and Waller, Matthew A. Drivers of retail on-shelf availability: Systematic review, critical assessment, and reflections on the road ahead. *International Journal of Physical Distribution & Logistics Management*, 2016.

[34] Corsten, Daniel and Gruen, Thomas. Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management*, 2003.

[35] Dupre, Kyle and Gruen, Thomas W. The use of category management practices to obtain a sustainable competitive advantage in the fast-moving-consumer-goods industry. *Journal of Business & Industrial Marketing*, 2004.

[36] Ehrenthal, Joachim CF and Stölzle, Wolfgang. An examination of the causes for retail stockouts. *International Journal of Physical Distribution & Logistics Management*, 2013.

[37] Kang, Yun and Gershwin, Stanley B. Information inaccuracy in inventory systems: stock loss and stockout. *IIE transactions*, 37(9):843–859, 2005.

[38] Frontoni, Emanuele, Marinelli, Fabrizio, Rosetti, Roberto, and Zingaretti, Primo. Shelf space re-allocation for out of stock reduction. *Computers & Industrial Engineering*, 106:32–40, 2017.

[39] Reiner, Gerald, Teller, Christoph, and Kotzab, Herbert. Analyzing the efficient execution of in-store logistics processes in grocery retailing—the case of dairy products. *Production and Operations Management*, 22(4):924–939, 2013.

[40] Condea, Cosmin, Thiesse, Frédéric, and Fleisch, Elgar. Rfid-enabled shelf replenishment with backroom monitoring in retail stores. *Decision Support Systems*, 52(4):839–849, 2012.

[41] Rosales, Claudia R, Whipple, Judith M, and Blackhurst, Jennifer. The impact of distribution channel decisions and repeated stockouts on manufacturer and retailer performance. *IEEE transactions on engineering management*, 66(3):312–324, 2018.

[42] Avlijas, Goran, Simicevic, Ana, Avlijas, Radoslav, and Prodanovic, Marijana. Measuring the impact of stock-keeping unit attributes on retail stock-out performance. *Operations Management Research*, 8(3):131–141, 2015.

[43] Gruen, Thomas W, Corsten, Daniel S, et al. A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry. 2007.

[44] Williams, Brent D, Waller, Matthew A, Ahire, Sanjay, and Ferrier, Gary D. Predicting retailer orders with pos and order data: The inventory balance effect. *European Journal of Operational Research*, 232(3):593–600, 2014.

[45] Chen, Chen-Yuan, Lee, Wan-I, Kuo, Hui-Ming, Chen, Cheng-Wu, and Chen, Kung-Hsing. The study of a forecasting sales model for fresh food. *Expert Systems with Applications*, 37(12):7696–7702, 2010.

[46] Waller, Matthew A, Williams, Brent D, Tangari, Andrea Heintz, and Burton, Scot. Marketing at the retail shelf: an examination of moderating effects of logistics on sku market share. *Journal of the Academy of Marketing Science*, 38(1):105–117, 2010.

[47] Anderson, Eric T, Fitzsimons, Gavan J, and Simester, Duncan. Measuring and mitigating the costs of stockouts. *Management science*, 52(11):1751–1763, 2006.

[48] Fisher, Marshall and Raman, Ananth. *The new science of retailing: how analytics are transforming the supply chain and improving performance*. Harvard Business Review Press, 2010.

[49] Bottani, Eleonora, Bertolini, Massimo, Rizzi, Antonio, and Romagnoli, Giovanni. Monitoring on-shelf availability, out-of-stock and product freshness through rfid in the fresh food supply chain. *International Journal of RF Technologies*, 8(1-2):33–55, 2017.

[50] Metzger, Christian, Thiesse, Frédéric, Gershwin, Stanley, and Fleisch, Elgar. The impact of false-negative reads on the performance of rfid-based shelf inventory control policies. *Computers & Operations Research*, 40(7):1864–1873, 2013.

[51] Sarac, Aysegul, Absi, Nabil, and Dauzère-Pérès, Stéphane. A literature review on the impact of rfid technologies on supply chain management. *International journal of production economics*, 128(1):77–95, 2010.

[52] Piramuthu, Selwyn, Wochner, Sina, and Grunow, Martin. Should retail stores also rfid-tag 'cheap'items? *European Journal of Operational Research*, 233(1):281–291, 2014.

[53] Franco, Annalisa, Maltoni, Davide, and Papi, Serena. Grocery product detection and recognition. *Expert Systems with Applications*, 81:163–176, 2017.

[54] Fleisch, Elgar and Tellkamp, Christian. Inventory inaccuracy and supply chain performance: a simulation study of a retail supply chain. *International journal of production economics*, 95(3):373–385, 2005.

[55] Chen, Li. Fixing phantom stockouts: Optimal data-driven shelf inspection policies. *Production and Operations Management*, 30(3):689–702, 2021.

[56] Chuang, Howard Hao-Chun. Fixing shelf out-of-stock with signals in point-of-sale data. *European Journal of Operational Research*, 270(3):862–872, 2018.

[57] Gruen, Thomas W. A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry a research study conducted. 2007.

[58] Ettouzani, Younes, Yates, Nicola, and Mena, Carlos. Examining retail on shelf availability: promotional impact and a call for research. *International Journal of Physical Distribution & Logistics Management*, 2012.

[59] Peinkofer, Simone T, Esper, Terry L, and Howlett, Elizabeth. Hurry! sale ends soon: the impact of limited inventory availability disclosure on consumer responses to online stockouts. *Journal of Business Logistics*, 37(3):231–246, 2016.

[60] Sloot, Laurens M, Verhoef, Peter C, and Franses, Philip Hans. The impact of brand equity and the hedonic level of products on consumer stock-out reactions. *Journal of Retailing*, 81(1):15–34, 2005.

[61] McKinnon, Alan C, Mendes, Daniela, and Nababteh, M. In-store logistics: an analysis of on-shelf availability and stockout responses for three product groups. *International Journal of Logistics Research and Applications*, 10(3):251–268, 2007.

[62] Trautrims, Alexander, Grant, David B, Fernie, John, and Harrison, Tim. Optimizing on-shelf availability for customer service and profit. *Journal of Business Logistics*, 30(2):231–247, 2009.

[63] Azeem, Muhammad Masood, Baker, Derek, Villano, Renato A, Mounter, Stuart, and Griffith, Garry. Response to stockout in grocery stores: A small city case in a changing competitive environment. *Journal of Retailing and Consumer Services*, 49:242–252, 2019.

[64] Parker, Jeffrey R and Lehmann, Donald R. When shelf-based scarcity impacts consumer preferences. *Journal of retailing*, 87(2):142–155, 2011.

[65] Syam, Siddhartha S and Bhatnagar, Amit. A decision support model for determining the level of product variety with marketing and supply chain considerations. *Journal of Retailing and Consumer Services*, 25:12–21, 2015.

[66] Pizzi, Gabriele and Scarpi, Daniele. The effect of shelf layout on satisfaction and perceived assortment size: An empirical assessment. *Journal of Retailing and Consumer Services*, 28:67–77, 2016.

[67] Hübner, Alexander and Schaal, Kai. An integrated assortment and shelf-space optimization model with demand substitution and space-elasticity effects. *European Journal of Operational Research*, 261(1):302–316, 2017.

[68] Hekimoğlu, Mustafa, Sevim, İsmail, Aksezer, Çağlar, and Durmuş, İpek. Assortment optimization with log-linear demand: Application at a turkish grocery store. *Journal of Retailing and Consumer Services*, 50:199–214, 2019.

[69] Frankel, Robert, Naslund, Dag, and Bolumole, Yemisi. The “white space” of logistics research: A look at the role of methods usage. *Journal of Business logistics*, 26(2):185–209, 2005.

[70] Raelin, Joseph A and Coghlan, David. Developing managers as learners and researchers: Using action learning and action research. *Journal of Management Education*, 30(5):670–689, 2006.

[71] Näslund, Dag, Kale, Rahul, and Paulraj, Antony. Action research in supply chain management—a framework for relevant and rigorous research. *Journal of Business Logistics*, 31(2):331–355, 2010.

[72] Coughlan, Paul, Draaijer, Domien, Godsell, Janet, and Boer, Harry. Operations and supply chain management: The role of academics and practitioners in the development of research and practice. *International Journal of Operations & Production Management*, 2016.

[73] DeHoratius, Nicole, Mersereau, Adam J, and Schrage, Linus. Retail inventory management when records are inaccurate. *Manufacturing & Service Operations Management*, 10(2):257–277, 2008.

[74] Bayraktar, Erhan and Ludkovski, Michael. Inventory management with partially observed nonstationary demand. *Annals of Operations Research*, 176(1):7–39, 2010.

[75] rhan and Shang, Kevin H. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing & service operations management*, 9(2):185–205, 2007.

[76] Papakiriakopoulos, Dimitris and Doukidis, Georgios. Classification performance for making decisions about products missing from the shelf. *Advances in Decision Sciences*, 2011.

[77] Song, Jing-Sheng and Zipkin, Paul. Inventory control in a fluctuating demand environment. *Operations Research*, 41(2):351–370, 1993.

[78] Chen, Fangruo and Song, Jing-Sheng. Optimal policies for multiechelon inventory problems with markov-modulated demand. *Operations Research*, 49(2):226–234, 2001.

[79] Treharne, James T and Sox, Charles R. Adaptive inventory control for nonstationary demand and partial information. *Management Science*, 48(5):607–624, 2002.

[80] Kazemian, Majid, Moshiri, Behzad, Nikbakht, Hamid, and Lucas, Caro. Protein secondary structure classifiers fusion using owa. in *International Symposium on Biological and Medical Data Analysis*, pp. 338–345. Springer, 2005.

[81] MacKay, David JC. Ensemble learning for hidden markov models. tech. rep., Citeseer, 1997.

*واژه‌نامه فارسی به انگلیسی

ا

آبشاری . . . . . Cascade
آنالیز آینده‌نگر . . . . . Prespective Analytics
آنالیز اجزای اصلی . . . . . Principal Component Analysis
آنالیز توصیفی . . . . . Descriptive Analytics
آنالیز مبتنی بر پیشبینی . . . . . Predictive Analytics
انتخاب ویژگی . . . . . Feature Selection
انتخاب ویژگی مستقیم . . . . . Forward Feature Selection
انتخاب ویژگی معکوس . . . . . Backward Feature Selection

ب

بازپرسازی . . . . . Replenishment
باور بیزی . . . . . Bayesian Belief
بده‌بستان . . . . . Trade-Off
بیز ساده . . . . . Naive Bayes
بیش‌برازش . . . . . Overfitting

پ

پردازش داده . . . . . Data Preprocessing
پشته‌سازی . . . . . Stacking
پوند . . . . . Pound
پیش‌پردازش . . . . . Preprocessing
پیشخور . . . . . Feed Forward

ت

تجزیه و تحلیل . . . . . Analytics

تجزیه و تحلیل کسب و کارها . . . . . Business Analytics
تقریب چندگانه یکنواخت و تصویر . . . . . Uniform Manifold Approximation and Projection
تقویت . . . . . Boosting
تمیزسازی داده‌ها . . . . . Data Cleaning

ث

ثبت دائم موجودی . . . . . Perpetual Inventory

ج

جدایی پذیری خطی . . . . . Liear Discriminant
جستجوی شبکه‌ای . . . . . Grid Search
جنگل تصادفی . . . . . Random Forest

ح

حداکثرسازی احتمال . . . . . Likelihood Maximization
حسابرسی‌های فیزیکی . . . . . Physical Audit
حساسیت . . . . . Sensivity

خ

خطای قدر مطلق مربعات . . . . . Mean Absolute Error
خطای مثبت واقعی . . . . . True Positive
خطای منفی کاذب . . . . . False Negative
خطای میانگین مربعات . . . . . Means Square Error
خودرمزگذار . . . . . Auto-Encoder

د

داده گمشده . . . . . Missing Value

در کیسه کردن . . . . . Bagging
درخت تصمیم . . . . . Decision Tree
درهم ریختگی . . . . . Confusion Matrix
دسترسی داخل قفسه . . . . . On-Shelf Availability
دستکاری داده . . . . . Data Manipulation
دقت . . . . . Precision

ر

رگرسیون لجستیک . . . . . Logistic Regression

س

سطوح . . . . . Levels
سیستم طبقه‌بند چندگانه . . . . . Multiple Classifier System

ش

شاخص های عملکرد کلیدی . . . . . Key Performance Indicators
شبکه عصبی مصنوعی . . . . . Artificial Neural Network
شناسایی فرکانس رادیویی . . . . . Radio Frequency Identification

ص

صحت . . . . . Accuracy

ع

عدم موجودی . . . . . Out of stock

ق

قدم‌زنی تعادل صفر . . . . . Zero Balance Walk

ک

کا نزدیکترین همسایه . . . . . K-Nearest-Neighbors

م

مارکوف مشاهده پذیر جزئی . . . . . Partially Observable Markov
ماشین بردار پشتیبان . . . . . Support Vector Machine
مثبت کاذب . . . . . False Positive
مدل مارکوف مخفی . . . . . Hidden Markov Model
مدیریت موجودی پلکانی . . . . . Echelon Inventory
معیار اف . . . . . F-Score
منفی واقعی . . . . . True Negative

ن

ناهمگن . . . . . Heterogamous
نقطه‌فروش . . . . . Point Of Sale
نمودار جفتی . . . . . Pair Plot

و

واحد انبارداری . . . . . Stock-Keeping Unit
ویژگی . . . . . Specificity

ه

همگن . . . . . Homogeneous

ی

یادگیری گروهی . . . . . Ensemble Learning
یادگیری ماشینی . . . . . Machine Learning
یادگیری مجتمع پویا . . . . . Dynamic Ensemble Learning

*واژه‌نامه انگلیسی به فارسی

A

صحت . . . . . Accuracy
تجزیه و تحلیل . . . . . Analytics
شبکه عصبی مصنوعی . . . . . Artificial Neural Network
خودرمزگذار . . . . . Auto-Encoder

B

انتخاب ویژگی معکوس . . . . . Backward Feature Selection
در کیسه کردن . . . . . Bagging
باور بیزی . . . . . Bayesian Belief
تقویت . . . . . Boosting
تجزیه و تحلیل کسب و کارها . . . . . Business Analytics

C

آبشاری . . . . . Cascade
درهم ریختگی . . . . . Confusion Matrix

D

تمیزسازی داده‌ها . . . . . Data Cleaning
دستکاری داده . . . . . Data Manipulation
پردازش داده . . . . . Data Preprocessing
درخت تصمیم . . . . . Decision Tree
آنالیز توصیفی . . . . . Descriptive Analytics
یادگیری مجتمع پویا . . . . . Dynamic Ensemble Learning

E

مدیریت موجودی پلکانی . . . . . Echelon Inventory
یادگیری گروهی . . . . . Ensemble Learning

F

خطای منفی کاذب . . . . . False Negative
مثبت کاذب . . . . . False Positive
انتخاب ویژگی . . . . . Feature Selection
پیشخور . . . . . Feed Forward
انتخاب ویژگی مستقیم . . . . . Forward Feature Selection
معیار اف . . . . . F-Score

G

جستجوی شبکه‌ای . . . . . Grid Search

H

ناهمگن . . . . . Heterogamous
مدل مارکوف مخفی . . . . . Hidden Markov Model
همگن . . . . . Homogeneous

K

شاخص های عملکرد کلیدی . . . . . Key Performance Indicators
کا نزدیکترین همسایه . . . . . K-Nearest-Neighbors

L

سطوح . . . . . Levels

جدایی پذیری خطی . . . . . Liear Discriminant
حداکثرسازی احتمال . . . . . Likelihood Maximization
رگرسیون لجستیک . . . . . Logistic Regression

M

یادگیری ماشینی . . . . . Machine Learning
خطای قدر مطلق مربعات . . . . . Mean Absolute Error
خطای میانگین مربعات . . . . . Means Square Error
داده گمشده . . . . . Missing Value
سیستم طبقه‌بند چندگانه . . . . . Multiple Classifier System

N

بیز ساده . . . . . Naive Bayes

O

دسترسی داخل قفسه . . . . . On-Shelf Availability
عدم موجودی . . . . . Out of stock
بیش‌برازش . . . . . Overfitting

P

نمودار جفتی . . . . . Pair Plot
مارکوف مشاهده پذیر جزئی . . . . . Partially Observable Markov
ثبت دائم موجودی . . . . . Perpetual Inventory
حسابرسی‌های فیزیکی . . . . . Physical Audit
نقطه‌فروش . . . . . Point Of Sale
پوند . . . . . Pound
دقت . . . . . Precision
آنالیز مبتنی بر پیشبینی . . . . . Predictive Analytics

پیش‌پردازش . . . . . Preprocessing
آنالیز آینده‌نگر . . . . . Prespective Analytics
آنالیز اجزای اصلی . . . . . Principal Component Analysis

R

شناسایی فرکانس رادیویی . . . . . Radio Frequency Identification
جنگل تصادفی . . . . . Random Forest
بازپرسازی . . . . . Replenishment

S

حساسیت . . . . . Sensivity
ویژگی . . . . . Specificity
پشته‌سازی . . . . . Stacking
واحد انبارداری . . . . . Stock-Keeping Unit
ماشین بردار پشتیبان . . . . . Support Vector Machine

T

بده‌بستان . . . . . Trade-Off
منفی واقعی . . . . . True Negative
خطای مثبت واقعی . . . . . True Positive

U

تقریب چندگانه یکنواخت و تصویر . . . . . Uniform Manifold Approximation and Projection

Z

قدم‌زنی تعادل صفر . . . . . Zero Balance Walk

Abstract:

In the same way that there is no concept known as the best model, the concept of the best type of modelling cannot be defined in general. Therefore, in this research, depending on the type of available dataset, a Dynamic Ensemble Classifier System based on clustering and the Euclidean distance of samples to the center of each cluster has been designed. This model is used to predict the out of stock situation, which is valid for the entire data feature space. The traditional approaches and physical tools that have been used so far to predict out of stock situation have many limitations due to human errors and high costs of personnel and physical tools. It is possible that your training or test datasets are an unrepresentative sample of data from the domain. In addition, if the available dataset is unrepresentative sample of data from the domain or incompletely represent the distribution of source data, single machine learning and AI-based models will not perform well. On the other hand, for such data static ensemble learning models cant to make personalized predictions for each new patient. Dynamic ensemble learning models have this ability, but for each new data, they simply select a classification model, classify the new data with the same model, and do not use the features and diversities of other classifications. Also, in this research, based on the structure and distribution of data in the feature space, by combining several optimal classifiers that have been trained in segmented feature spaces, an optimal model called ”Dynamic Ensembel Classifier System based on clustering and Euclidean distance” has been designed to covers the limitations of the previous methods and has a better performance. This approach removes the limitations of previous methods and makes it possible to use the characteristics of individual classifiers and their diversity.Thus it is possible to produce a highly interpretable model using base models with a simple structure. The proposed model has been compared with other methods, including single machine learning methods and other dynamic ensemble learning methods that have been used so far, also its superiority has been confirmed by the standard indicators of measuring the performance of the model.

Keywords: Dynamic ensemble classifier system, Clustering, Business analytics, Out of stock situation

University of Tehran
College of Engineering
Faculty of Electrical and Computer Engineering
Control Department

Design of Dynamic Ensemble Classifier Fusion for intelligent Decision Making in Business Analytics

A Thesis submitted to the Graduate Studies Office
In partial fulfillment of the requirements for
The degree of Master of Science
in Electrical Engineering - Control

By:

Ali Hafezi

Supervisor:

Professor Behzad Moshiri

September 2022