



The Prediction of Employee Turnover and its Personal and Organizational Determinants Using Machine Learning Methods

Soheila Mohammadi¹ , Abbas Khandan^{*},

1. Faculty of Economics, University of Kharazmi, Tehran, Iran.

^{*} Corresponding author

Article Info	Abstract
Article Type: Research Article Article History: Received: 29 July 2024 Revised: 14 May 2025 Accepted: 07 June 2025 Published: 12 Sep. 2025 Keywords: <i>Classification, Determinants, Employee Turnover, Individual and Occupational Machine Learning.</i> JEL Classification: <i>C63, J63, L53.</i>	This study analyzing the chain stores' personnel data intends to predict the employees' turnover rate and examine its personal and organizational (occupational) determining factors. A dataset of 17542 records including information of personnel's activity status (active or quit) and 12 personal and occupational characteristics in the period of November 2018 to 2014 are used in machine learning algorithms of decision tree and support vector machine (SVM) to predict the workforce turnover rate across the country in a pilot chain store. The results show that job characteristics have a greater effect on employees' turnover. Specifically, among different characteristics, four factors of personnel transfers, their organizational position (line or staff), years of service and working hours have the most importance and influence on job turnover. Among the individual characteristics, it was also observed that job quit is more prevalent among young people with less than 30 years old. Based on these characteristics, support vector machine (SVM) model with 91% accuracy and F1-score above 90% and decision tree algorithm with 83% accuracy and F1-score have shown a good performance in classification and prediction of employees' turnover. Based on personal and job characteristics and using data mining and machine learning methods, organizations can set policies to preserve human resources, which will reduce costs and also maintain their competitive advantages necessary for the progress and development of the company.

Mohammadi, S., & Khandan, A. (2025). The Prediction of Employee Turnover and its Personal and Organizational Determinants Using Machine Learning Methods. *Journal of Economic Research*, 60(2), 1054-1097.



©The Authors retain the copyright and full publishing rights.

Publisher: The University of Tehran Press.

DOI: [10.22059/jte.2025.380061.1008925](https://doi.org/10.22059/jte.2025.380061.1008925)

پیش‌بینی ترک شغل و عوامل فردی و سازمانی مؤثر بر آن با استفاده از

روش‌های یادگیری ماشین

سهیلا محمدی^۱✉، عباس خندان^{۱*} 

۱. دانشکده اقتصاد، دانشگاه خوارزمی، تهران، ایران.

* نویسنده مسئول

اطلاعات مقاله	چکیده
نوع مقاله: پژوهشی	این مقاله در نظر دارد تا با تجزیه و تحلیل داده‌های مرتبط با نیروی کار یک
تاریخ دریافت: ۱۴۰۳/۰۵/۰۸	فروشگاه‌های زنجیره به پیش‌بینی ترک شغل نیروها و بررسی عوامل فردی و سازمانی
تاریخ بازنگری: ۱۴۰۴/۰۲/۲۵	(شغلی) مؤثر بپردازد. تعداد ۱۷۵۴۲ رکورد اطلاعاتی منحصر به فرد شامل اطلاعات
تاریخ پذیرش: ۱۴۰۴/۰۳/۱۸	وضعیت فعالیت فرد (ادامه فعالیت یا ترک شغل) و ۱۲ مشخصه فردی و سازمانی
تاریخ انتشار: ۱۴۰۴/۰۶/۲۲	(شغلی) در بازه زمانی آبان ۱۳۹۸ الی ۱۴۰۱ بکار گرفته شد و سپس به روش داده‌کاوی
	و با استفاده از الگوریتم‌های یادگیری ماشین درخت تصمیم و ماشین بردار پشتیبان
	(SVM) به پیش‌بینی ترک شغل نیروی کار در این فروشگاه زنجیره‌ای به صورت
کلیدواژه‌ها:	پایلوت در سراسر کشور پرداخته شد. نتایج بدست آمده نشان دادند که ویژگی‌ها و
ترک شغل،	مشخصه‌های شغلی تأثیر بیشتری بر ترک شغل دارند. به طور مشخص، از بین
عوامل فردی و سازمانی،	ویژگی‌های مختلف تعداد جابه‌جایی نیروها، سمت سازمانی (صف یا ستادی)، سنوات
طبقه‌بندی،	خدمت و ساعت کار بیشترین اهمیت و تأثیرگذاری را بر ترک شغل دارند. از میان
یادگیری ماشین.	مشخصه‌های فردی نیز مشاهده شد که ترک شغل در میان جوانان و افراد کمتر از ۳۰
	سال سن بیشتر است. بر اساس این ویژگی‌ها، مدل‌های ماشین بردار پشتیبان (SVM)
طبقه‌بندی JEL:	با دقت ۹۱ درصد و امتیاز-F1 بالای ۹۰ درصد و الگوریتم درخت تصمیم نیز با دقت
J63, L53, C63.	۸۳ درصد و امتیاز-F1 به همین اندازه از عملکرد مناسبی در دسته‌بندی و پیش‌بینی
	موارد ترک شغل برخوردار شدند. می‌توان بر اساس مشخصه‌های فردی و شغلی و با
	استفاده از روش‌های داده‌کاوی و یادگیری ماشینی سیاست‌هایی در جهت حفظ و
	نگه‌داشت منابع انسانی تنظیم کرد که موجب کاهش هزینه‌ها و همچنین حفظ
	مزیت‌های رقابتی و پیشرفت و توسعه بنگاه خواهد شد.

محمدی، سهیلا، و خندان، عباس. (۱۴۰۴). پیش‌بینی ترک شغل و عوامل فردی و سازمانی مؤثر بر آن با استفاده از روش‌های یادگیری ماشین. *تحقیقات اقتصادی*، ۶۰(۲)، ۱۰۵۴-۱۰۹۷.



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: [10.22059/jte.2025.380061.1008925](https://doi.org/10.22059/jte.2025.380061.1008925)

۱- مقدمه

سازمان را می‌توان فرایندهای نظام یافته از روابط متقابل افراد برای دست یافتن به هدف‌های معین دانست جایی که عنصر اصلی و مؤثر آن را منابع انسانی تشکیل می‌دهد. در دنیای پر چالش امروز، جذب استعدادها و حفظ و نگهداری منابع انسانی مستعد بسیار مشکل است (فیلیپس و روپر^۱، ۲۰۰۹). سازمان‌هایی که نمی‌توانند سازوکارها و استراتژی‌های مناسبی برای حفظ و نگهداری استعدادها طراحی نمایند، آن‌ها را از دست خواهند داد. منابع انسانی مهمترین مزیت رقابتی و سرمایه سازمانی هستند و از این‌رو، کاهش نرخ ترک شغل آنان برای باقی ماندن سازمان در شرایط رقابتی و دستیابی به سطوح عالی عملکرد حیاتی است (پیرایش و همکاران، ۱۳۹۹). با توجه به هزینه‌های بالای استخدام و توانمندسازی کارکنان، خروج کارکنان از سازمان علاوه بر وارد آمدن هزینه‌های زیاد به سازمان، موجب ناکامی در دستیابی به اهداف و برنامه‌های سازمانی می‌شود (دباشی و همکاران، ۱۳۹۵). از این رو، یکی از مهمترین مسائل مهم سازمان‌ها در دنیای امروز مسئله ترک شغل یا خروج کارکنان به‌ویژه کارکنان ماهر یا همان استعدادها است. آگاهی از این موضوع که چه عواملی باعث می‌شود یک کارمند تصمیم به ترک سازمان بگیرد، در توانایی سازمان برای حفظ کارمندان بسیار حیاتی است (سبک‌رو و همکاران، ۱۳۹۰). مطالعات نشان می‌دهند که ترک شغل برای سازمان‌ها قابل پیش‌بینی است. ترک شغل عملی آگاهانه و هدفمند است؛ کارکنان سازمان به یکباره تصمیم به ترک سازمان نمی‌گیرند، بلکه این تصمیم به تدریج گرفته می‌شود و افراد پس از بررسی تمامی جوانب به‌خصوص فرصت‌های استخدام در سایر سازمان‌ها، تصمیم به ترک شغل می‌گیرند (عابدین و همکاران، ۱۳۹۹). تمایل به ترک شغل انعکاسی از علاقه کارکنان به ترک سازمان و جستجوی مشاغل جایگزین است (چن^۲، ۲۰۲۰). بنابراین، با به‌کارگیری اصول و راهکارهای مؤثر می‌توان دلایل پنهان در تمایل کارکنان به ترک شغل را شناسایی و آن‌ها را برطرف کرد (آروکیاسامی^۳، ۲۰۱۳).

1. Phillips and Roper

2. Chen

3. Arokiasamy

با توجه به اهمیت این موضوع در سازمان‌ها و در نتیجه در اقتصاد، این مقاله قصد دارد تا با استفاده از روش‌های یادگیری ماشین^۱ به پیش‌بینی نرخ خروج و ترک شغل کارکنان بپردازد. حوزه مورد مطالعاتی این مقاله به فروشگاه‌های زنجیره‌ای مصرف محدود است اما نتایج آن در دیگر حوزه‌های اقتصادی نیز می‌تواند مورد استفاده قرار گیرد. دلیل توجه به فروشگاه‌های زنجیره‌ای تحولات گسترده‌ای است که در دهه‌های اخیر در صنعت خرده‌فروشی در ایران اتفاق افتاده است. این تحولات که در راستای افزایش بهره‌وری و کارآیی اتفاق افتاده به گونه‌ای است که امروزه مراجعه به فروشگاه‌های زنجیره‌ای مصرف به کاری روزمره و عادی برای تمامی اقشار جامعه مبدل شده و این صنعت به سرعت در حال رقابتی شدن است. گستره مکانی تحقیق شعب سراسر کشور یک فروشگاه زنجیره‌ای مصرف را شامل می‌شود. با بررسی ترک شغل نیروی انسانی شغل مختلف این فروشگاه در بازه زمانی ۱۳۹۸ تا ۱۴۰۱، این مقاله قصد دارد به این سؤال اصلی بپردازد که چه عوامل اصلی سازمانی یا مشخصه‌های فردی وجود دارند که با ترک شغل افراد مرتبط بوده و می‌توانند آن را پیش‌بینی کنند.^۲ برای این منظور از دو الگوریتم یادگیری ماشین بردار پشتیبان^۳ (SVM) و درخت تصمیم^۴ (DT) استفاده شده است. استفاده از دو الگوریتم در عین این که به نتایج بدست آمده اطمینان بیشتری می‌بخشد، امکان مقایسه به لحاظ عملکرد مدل‌ها در پیش‌بینی را نیز فراهم می‌آورد. از این رو، می‌توان هدف فرعی این مقاله را مقایسه عملکرد این دو الگوریتم یادگیری ماشین را در پیش‌بینی دانست.

ساختار مقاله به این صورت است که ابتدا و در بخش دوم به ادبیات موضوع پرداخته و پیشینه پژوهش مرور خواهد شد. بخش سوم به روش‌شناسی پژوهش اختصاص داشته که در آن ابتدا متغیرهای پژوهش و آمار توصیفی از آن‌ها ارائه خواهد شد. در این بخش همچنین شیوه آماده‌سازی داده‌ها و همچنین دو الگوریتم ماشین بردار پشتیبان (SVM) و درخت تصمیم به عنوان دو الگوریتم مورد استفاده برای پیش‌بینی ترک شغل معرفی خواهند شد. بخش چهارم به ارائه یافته‌های پژوهش می‌پردازد. و در نهایت، بخش ششم به جمع‌بندی و نتیجه‌گیری اختصاص دارد.

1. Machine Learning

۲. اگرچه ترک شغل از عوامل اقتصادی و کلان نیز متأثر می‌گردد اما بررسی این عوامل نیازمند در اختیار داشتن یک مجموعه داده کلان سری زمانی است که در این مطالعه و با استفاده از این مجموعه داده خرد از شاغلین فروشگاه امکان‌پذیر نیست.

3. Support Vector Machine

4. Decision Tree

۲- ادبیات موضوع و پیشینه پژوهش

ترک شغل برای سازمان‌ها هزینه‌های مختلف مستقیم و غیرمستقیم دارد. هزینه‌های اداری قطع حقوق، پرداخت پاداش پایان کار و حق سنوات و ... هزینه‌های مستقیمی است که با ترک شغل به سازمان وارد می‌شود. علاوه بر این، استخدام افراد جدید و جایگزین نیز برای سازمان هزینه دارد؛ هزینه آگهی و تبلیغات، هزینه مربوط به بررسی سوابق کاری، مصاحبه استخدامی، اجرای آزمون‌های استخدامی برای بررسی مهارت‌ها و استعدادها و هزینه اداری استخدام (تشکیل پرونده استخدامی اداری، صدور کارت شناسایی) از جمله این هزینه‌های مستقیم به‌کارگیری افراد جایگزین است. اما شاید از همه مهمتر هزینه‌های آموزش کارکنان تازه وارد باشد. هر سازمان برای آموزش، تربیت و آماده‌سازی کارکنان خود تا مرحله بهره‌دهی و کارایی مطلوب هزینه‌های بسیاری را صرف می‌کند و با از دست دادن نیروهای ارزشمند، متحمل از دست دادن مهارت‌ها و تجربیاتی می‌شود که طی سال‌ها تلاش بدست آمده است (زبانی و همکاران، ۱۳۹۶). والدمن^۱ و همکاران (۲۰۰۴) هزینه ترک شغل در سازمان‌های دانشی را تقریباً بین ۳/۴ تا ۵/۸ درصد کل بودجه عملیاتی آن‌ها برآورد کرده است. البته هزینه‌های ترک شغل تنها به این هزینه‌های مستقیم محدود نمی‌شود و بخش مهمی از تأثیرات و هزینه‌های ترک شغل در بلندمدت اتفاق می‌افتد. به‌عنوان مثال می‌توان به کاهش رضایت شغلی در کل مجموعه اشاره کرد. از دست دادن کارکنان خوب تأثیر مخربی بر بقیه دارد و آن‌ها را دچار بی‌انگیزگی می‌کند و در نتیجه بهره‌وری در سازمان کاهش می‌یابد. از جمله هزینه‌های بلندمدت دیگر ترک شغل می‌توان به مواردی از جمله احتمال دستیابی رقبا به اطلاعات کلیدی کارکنان، افت فروش، از دست دادن پتانسیل نیروهای کارآمد، شکل‌گیری تصور در کارکنان مبنی بر وجود فرصت‌های شغلی بهتر در بیرون سازمان، به‌هم خوردن شبکه اجتماعی درون سازمان و ... اشاره کرد (وینترتون^۲، ۲۰۰۴).

عوامل مختلفی که می‌توانند بر ترک شغل تأثیر بگذارند را می‌توان در سه دسته عوامل کلان (اقتصادی، اجتماعی و سیاسی)^۳، دوم عوامل فردی و شخصی، و سوم عوامل مدیریتی و

^۱. Waldman

^۲. Winterton

^۳. عوامل کلان اقتصادی و اجتماعی با استفاده از یکسری متغیرهای دامی استانی و زمانی کنترل خواهند شد که در بخش سوم و معرفی متغیرها شرح داده شده است.

سازمانی تقسیم کرد. در خصوص عوامل خرد شخصی و سازمانی (شغلی) که این مقاله قرار است مورد بررسی قرار دهد، نظریات مختلفی وجود دارند. نخست نظریه انتظار^۱ که توسط روم^۲ (۱۹۶۴) مطرح شد و بیان می‌دارد که اگر اهداف سازمان با انتظارات کارکنان منطبق باشد، اثربخشی سازمان بالا خواهد رفت چون که کارکنان با میل و رغبت بیشتری فعالیت کرده و خلاقیت و توانشان را به نحو شایسته‌تری به کار خواهند گرفت. در غیراین صورت، تعارض بین اهداف فردی و سازمان عاملی مؤثر در ترک شغل است. نظریه دیگر نظریهٔ برابری^۳ است که توسط آدامز^۴ (۱۹۶۵) مطرح شد. این نظریه بیان می‌دارد که افراد داده‌ها و سرمایه‌گذاری‌های خود برای سازمان (تحصیلات و تجربیاتی که از آن‌ها چشم پوشیده، زمان، موقعیت اجتماعی و سازمانی، میزان تلاش) و ستاده‌های خود از سازمان (دستمزد، مقام و مرتبه سازمانی، پاداش و ترفیع و سایر امکانات رفاهی) را با دیگران مقایسه کرده و بر اساس قضاوت شخصی خود احساس برابری یا نابرابری می‌کند. بر اساس این نظریه افراد خواستار برابری هستند، حتی وقتی که احساس نابرابری مثبت می‌کنند هم می‌کوشند تا به برابری برسند. اما وقتی که افراد احساس نابرابری منفی دارند مثلاً تصور می‌کنند که در مقایسه با دیگران بیشتر کار کرده ولی دستمزد کمتری دریافت داشته‌اند، نیز از طریق کاهش کیفیت تولید یا کم کاری تلاش می‌کنند تا به برابری مورد نظر برسند. در صورتی که تلاش‌های کارکنان برای دستیابی به برابری نتیجه بخش نباشد، دست به اقدامات منفی می‌زنند که هیچگاه به نفع سازمان نخواهد بود. نمونه این اقدامات منفی کم کاری، خرابکاری، کاهش تعهد سازمانی و ترک شغل است (چشم براه، ۱۳۹۷).

نظریه معروف «جان هالند»^۵ در سال ۱۹۱۹ در باره تناسب شغل با شخصیت مطرح شد و بیان می‌دارد که رغبت یا علاقه فرد متناسب با محیط کارش است. هالند در نظریه خود شش نوع شخصیت ارائه می‌کند و معتقد است که رضایت شغلی و میل به ترک شغل به درجه‌ای بستگی دارد که فرد می‌تواند به صورتی موفقیت‌آمیز شخصیت خود را با یک محیط شغلی وفق دهد. اگر شخصیت افراد و نوع شغل آن‌ها با هم متناسب باشد، رضایت شغلی به بالاترین حد می‌رسد و در

^۱. Expectancy Theory

^۲. Victor Vroom

^۳. Equity Theory

^۴. Adams

^۵. John Holland

نتیجه جابه جایی کارکنان و ترک شغل به پایین‌ترین میزان تنزل پیدا می‌کند. بنابراین، طراحی مناسب مشاغل اهمیت زیادی دارد و برای این منظور الگوی «هاکمن»^۱ که در سال ۱۹۷۶ مطرح شد، می‌تواند بسیار مؤثر باشد. این الگو پنج ویژگی‌ها اصلی یعنی (۱) تنوع مهارت، (۲) معنی‌دار بودن وظایف محوله، (۳) اهمیت وظایف، (۴) داشتن استقلال و اختیار و (۵) قابلیت بازخورد گرفتن برای مشاغل بر می‌شمارد و معتقد است که این ویژگی‌های پنجگانه شغلی بر مثلث بهره‌وری، انگیزش و رضایت شغلی کارکنان بسیار مؤثرند. هر چه این مثلث تقویت شده یعنی مشاغل با انگیزش، عملکرد و رضایت بیشتری همراه باشد، غیبت در کار و احتمال ترک شغل نیز کمتر خواهد شد. هر قدر دانش و توانایی کارکنان افزایش یابد، آن‌ها همچنین در مسیر سلسله مراتب نیازهای مازلو^۲ ارتقا یافته و ماهیت مشاغل برای آن‌ها اهمیت بیشتری پیدا خواهد کرد. بنابراین مشاغل باید به گونه‌ای طراحی شوند که برای تمامی سطوح کارکنان و برای افراد با تحصیلات و استعدادهای مختلف چالشی باشد.

۲-۱- پیشینه پژوهش

از معدود مطالعات داخلی که به موضوع مهم ترک شغل و عوامل مؤثر بر آن پرداخته‌اند، می‌توان به موارد زیر اشاره کرد. هاشم زهی و همکاران (۱۴۰۰) مدل ترک شغل کارکنان دانشی را در شرکت پخش فرآورده‌های نفتی تحلیل کرده‌اند. در این پژوهش به منظور جمع‌آوری داده‌ها از تکنیک سه مرحله‌ای دلفی، مصاحبه‌های نیمه‌ساختاریافته و در بخش کمی نیز از معادلات ساختاری استفاده گردید. بر اساس یافته‌های بخش کیفی هفت عامل محیطی، سازمانی و ساختاری، شغلی، فردی، قانونی و حقوقی، ارزشی-عقیدتی و مدیریتی شناسایی شدند که به همراه ۳۵ مؤلفه و ۱۶۷ گویه در ترک شغل کارکنان دانشی مؤثرند. نتایج حاصل از معادلات ساختاری در بخش کمی نیز نشان داد که از بین عوامل هفتگانه عامل ساختاری یا سازمانی مهم‌ترین عامل در ترک شغل کارکنان دانشی بوده و عامل قانونی-حقوقی آخرین رتبه را به خود اختصاص داد.

¹. Hachman

². Maslow

عابدین و همکاران (۱۳۹۹) یک مدل مفهومی برای توضیح ترک شغل افراد مستعد از سازمان‌های ایران طراحی کرده‌اند. این مطالعه با استفاده از مصاحبه‌های عمیق^۱ با ۱۳ فرد با استعدادی که شغل خود را ترک کرده بودند و پس از آن تحلیل داده‌های حاصل از مصاحبه‌ها، استراتژی پژوهشی داده بنیاد^۲ به کار گرفته شده است. مدل مفهومی بدست آمده، عوامل اصلی تأثیرگذار بر ترک شغل افراد با استعداد از جمله عوامل علی- محیطی همچون رکود اقتصادی، عوامل زمینه‌ای از جمله ویژگی‌های سازمانی، عامل تغییر در احساس زندگی کاری مثل عدم رضایت، عامل ویژگی‌های فردی مانند شخصیت و عوامل مداخله‌گر همچون بی‌عدالتی ساختاری در کشورهای توسعه‌نیافته را نمایان می‌سازد.

پیرایش و همکاران (۱۳۹۹) عوامل مؤثر بر قصد ترک شغل کارکنان و تأثیر آن بر کارایی آن‌ها را مورد مطالعه قرار داده‌اند. جامعه آماری این پژوهش کارکنان شاغل در شرکت زیرین روی زنجان را شامل می‌شود که ۶۲ نفر از آن‌ها با نمونه‌گیری تصادفی به عنوان نمونه آماری تحقیق انتخاب شدند. برای بررسی فرضیه‌های تحقیق ضریب همبستگی پیرسون محاسبه شد. به منظور سنجش عوامل، پرسشنامه استاندارد به کار گرفته شد. نتایج تجزیه و تحلیل داده‌ها حاکی از تأیید فرضیه‌های تحقیق است و نشان می‌دهند که بین ویژگی‌های سرپرست، ویژگی‌های شغلی، احساس و عواطف در شغل، نگرش شغلی و شرایط کاری با کاهش ترک شغل و جابه‌جایی کارکنان رابطه معنادار و مستقیمی وجود دارد و تأثیرگذارترین عامل نیز احساسات و عواطف در شغل و همچنین کمترین تأثیر را نگرش شغلی بر قصد ترک شغل و جابه‌جایی کارکنان دارند.

عارفی و همکاران (۱۳۹۸) به بررسی عوامل مؤثر بر ترک شغل پرستاران پرداخته‌اند. این مطالعه از نوع کیفی بوده که به شیوه پدیدارشناسی^۳ در سال ۱۳۹۷ انجام شد. داده‌ها از طریق مصاحبه‌های نیمه‌ساختاریافته^۴ با افراد یکی از بیمارستان‌های خصوصی شهر تهران که محیط کار خود را ترک کرده بودند، جمع‌آوری شد. عوامل مؤثر بر ترک شغل پرستاران در ۴ کد اصلی شامل عوامل سازمانی، مالی و اقتصادی، فردی و خانوادگی و عوامل مرتبط با محیط کار و ۱۳ کد فرعی

1. In-depth interviews

2. Foundation data method

3. Phenomenology

4. Semi-structured interview

شناسایی شدند. طبق نتایج بدست آمده اجرای قانون بهره‌وری در بخش خصوصی، انعقاد قراردادهای بلندمدت با کارکنان، بهبود مقررات مالی بیمارستان‌های خصوصی و نظارت اثربخش بر اجرای آن‌ها و توجه بیشتر به مزایای شغلی مناسب و انگیزشی در بیمارستان می‌تواند منجر به احساس امنیت روانی و شغلی بیشتر، کاهش فرسودگی شغلی، احساس مطلوب در محیط کار و نهایتاً باعث کاهش ترک شغل در پرستاران شاغل در بخش خصوصی شود.

طالقانی و همکاران (۱۳۹۷) در تحقیقی عوامل فردی مؤثر بر قصد ترک شغل کارکنان را بررسی کرده‌اند. روش پژوهش مدل‌یابی معادلات ساختاری است. برای تعیین پایایی ابزار پژوهش از آلفای کرونباخ استفاده شد که این مقدار برای تمام پرسشنامه‌ها بیشتر از ۰/۷۵ بدست آمد و نشان داد ابزارهای پژوهش پایایی مناسبی دارند. داده‌ها در دو سطح توصیفی و استنباطی تحلیل شدند. نتایج بدست آمده نشان دادند که اثر متغیرهای رقابت‌پذیری (توانایی و عملکرد شرکت در فروش و عرضه کالا و خدمات در مقایسه با دیگر شرکت‌ها در همان بازار)، خوداثربخشی (قضاوت شخصی از چگونگی انجام کارهای که تأثیرات آن‌ها به آینده مرتبط می‌شود) و تلاش بر عملکرد کارکنان معنادار هستند. متغیرهای خوداثربخشی و تلاش بر رضایت شغلی کارکنان نیز اثری معنادار دارند. و در نهایت، متغیرهای عملکرد و رضایت شغلی نیز بر قصد ترک شغل اثر معناداری داشته است.

چشم براه (۱۳۹۷) در یک پایان‌نامه کارشناسی ارشد به عوامل مؤثر بر ترک شغل در بین کارکنان اداری پرداخته است. عوامل مؤثر بر ترک شغل طبق این پژوهش به دو دسته عوامل فردی (سن، نوع قرارداد استخدامی، سطح تحصیلات، طبقه شغلی) و سازمانی (تعهد سازمانی، خشنودی شغلی، تناسب بین ارزش‌های فردی و سازمانی) تقسیم می‌شوند که در برهه‌های سنی افراد شکل‌های متفاوتی داشته و لذا در صورت کنترل نشدن، به تدریج منجر به ترک شغل می‌شوند. بر اساس نتایج بدست آمده از این پژوهش مشاهده شد که در حوزه عوامل فردی، ترک شغل با متغیرهای سن و استخدام رسمی رابطه منفی و با سطح تحصیلات رابطه مثبتی وجود دارد. در واقع، با افزایش سن از میزان تمایل به ترک شغل کاسته می‌شود و بیشترین میزان ترک شغل نیز در یک سال اول استخدام اتفاق می‌افتد و کارکنانی که سازمان را ترک می‌کنند معمولاً سابقه استخدام کوتاهی دارند. افراد با تحصیلات بالاتر هم بیشتر از افراد با تحصیلات پایین

تمایل دارند که شغل خود را ترک کنند. همچنین مشاهده شد ترک شغل در افرادی که به لحاظ شغلی در طبقه غیرمدیریتی قرار دارند در مقایسه با افرادی که در این طبقه قرار دارند (طبقه مدیریتی) بیشتر است. نتایج در حوزه عوامل سازمانی نیز نشان دادند که کاهش تعهد سازمانی کارکنان می‌تواند موجبات تمایل به ترک شغل را در آنان فراهم آورد. بین خشنودی شغلی و ترک شغلی نیز رابطه منفی وجود دارد. افرادی که ارزش‌های فردی‌شان با ارزش‌های سازمانی تجانس بیشتری دارد، تمایل دارند که خشنودی شغلی و تعهد سازمانی بیشتری از خود نشان دهند و در مقابل، اگر میزان تجانس بین ارزش‌های فردی و سازمانی کاهش پیدا کند، کارکنان تمایل بیشتری به ترک سازمان پیدا می‌کنند.

دیهیم پور دولتی (۱۳۹۶) عوامل ایجادکننده آسیب‌های سازمانی بر میزان ترک شغل کارکنان نظامی را مورد بررسی قرار داده‌اند. ابزار گردآوری اطلاعات در این مقاله پرسشنامه است که مشتمل بر ۳۴ گویه و بر اساس طیف لیکرت^۱ تنظیم شده بود. جهت محاسبه روایی^۲ پرسشنامه از طریق روایی محتوای و استفاده از نظریات خبرگان و جهت محاسبه پایایی^۳ پرسشنامه از روش آلفای کرونباخ^۴ استفاده شد که ضریب مذکور ۰/۸۱ بدست آمد. برای تحلیل داده‌ها از اطلاعات جمعیت‌شناختی (آمار توصیفی) و در بخش استنباطی نیز برای تحلیل داده‌ها از آزمون همبستگی، ضریب رگرسیون چندمتغیره، آزمون t برابری و نابرابری واریانس‌ها و آنالیز واریانس استفاده گردید. نتایج حاکی از آن بود که بین ابعاد ایجادکننده آسیب‌های سازمانی و ترک شغل کارکنان رابطه معنی‌داری وجود دارد و سهم تأثیرگذاری هریک از این عوامل متفاوت است. همچنین این تأثیر بر اساس جنسیت، نوع قرارداد استخدامی، سنوات خدمت و پست سازمانی بر کارکنان با یکدیگر متفاوت است. به این ترتیب که میزان ترک شغل در مردان بیشتر از زنان، در کارکنان با استخدام موقت بیشتر از کارکنان رسمی، در کارکنان با سنوات کمتر، بیشتر از سایر کارکنان و در کارکنان با پست‌های سازمانی پایین، بیشتر است.

۱. Likert

۲. Validity

۳. Reliability

۴. Cronbachs alpha

علیرحیمی و همکاران (۱۳۹۶) مدل کاهش نرخ ترک شغل سازمانی را طراحی و تدوین کرده‌اند. این پژوهش به روش دلفی^۱ انجام گرفته است که بر آن اساس از ۱۲ متخصص مرتبط با موضوع تحت عنوان پانل دلفی خواسته شد تا طی دوره‌های متوالی دلفی، اقدام به اعتباریابی شاخص‌های شناسایی شده برای کاهش نرخ ترک شغل سازمانی نمایند. انجام روش دلفی پس از انجام دور سوم و دستیابی به اتفاق نظر مطلوب پایان یافت. نهایتاً مدل مفهومی بر اساس دلفی خبرگان در خصوص شاخص‌های بومی مؤثر بر کاهش ترک شغل سازمانی کارکنان شرکت پاپانه‌های نفتی ایران (مستخرج از مجموعه عوامل و شاخص‌های احصا شده از تحقیقات و پژوهش‌های پیشین) طراحی گردید.

امیری و محمودزاده (۱۳۹۴) عوامل مؤثر بر کاهش نرخ ترک شغل کارکنان در سازمان‌های دولتی ایران را بررسی کرده‌اند. در این تحقیق عواملی چون ویژگی‌های شغلی، ویژگی‌های سرپرست شغل، هویت سازمانی، احساسات و عواطف در شغل، شرایط کاری و نگرش شغلی به عنوان عوامل مؤثر بر کاهش ترک شغل انتخاب شدند. جامعه آماری تحقیق شامل کارکنان شاغل در ستاد مرکزی وزارت راه و شهرسازی است که با استفاده از روش نمونه‌گیری تصادفی ۱۸۰ نفر به عنوان نمونه آماری تحقیق انتخاب شدند. در این پژوهش به منظور بررسی نرمال بودن داده‌ها از آزمون کولموگروف اسمیرنوف^۲ و به منظور بررسی و آزمون فرضیه‌های تحقیق نیز از روش‌های آمار استنباطی از جمله ضریب همبستگی پیرسون استفاده گردید. نتایج تحقیق نشان می‌دهد که بین ویژگی‌های شغلی، ویژگی‌های سرپرست، هویت سازمانی، احساسات و عواطف در شغل، شرایط کاری و نگرش شغلی با کاهش نرخ ترک شغل کارکنان رابطه مستقیم و معنی‌داری وجود دارد. از این میان نیز عامل احساسات و عواطف در شغل دارای بیشترین تأثیر و هویت سازمانی و نگرش شغلی کمترین تأثیر را بر کاهش نرخ ترک شغل کارکنان داشته‌اند.

از جدیدترین مطالعات خارجی که به موضوع ترک شغل پرداخته‌اند نیز می‌توان به موارد زیر اشاره کرد. ال آکاشه^۳ و همکاران (۲۰۲۴) در یک مطالعه مروری در بازه ۲۰۱۲ تا ۲۰۲۳ به بررسی پژوهش‌هایی می‌پردازند ترک شغل در سازمان‌های مختلف را با استفاده از الگوریتم‌های

¹. Delphi

². Kolmogorov-Smirnov

³. Al Akasheh

یادگیری ماشین مورد مطالعه قرار داده‌اند. این مقاله نشان می‌دهد که بیش از ۲۰ الگوریتم یادگیری ماشین برای پیش‌بینی ترک شغل مورد استفاده قرار گرفته که عمدتاً الگوریتم‌های مبتنی بر نظارت بوده است و در این میان الگوریتم جنگل تصادفی پرکاربردترین آن بوده و از میان عوامل مؤثر بر ترک شغل نیز موضوع حقوق و درآمد مهمترین عامل اثرگذار بوده است.

پارک^۱ و همکاران (۲۰۲۴) بر اهمیت استفاده از الگوریتم‌های یادگیری ماشینی در پیش‌بینی ترک شغل کارکنان تأکید می‌کنند چراکه این مدل‌ها با پیش‌بینی یک احتمال ترک شغل برای هر فرد یا مشاهده می‌توانند به مدیران در شناسایی سریع این موارد و مدیریت ترک شغل کمک کنند. در این راستا، این مقاله با استفاده از اطلاعات بدست آمده از نیروی کار کره جنوبی به بررسی آن با رگرسیون OLS در کنار مدل‌های یادگیری و دسته‌بندی رگرسیون لجستیک، K-نزدیکترین همسایگی و XGBoost می‌پردازد. نتایج این مطالعه نشان داد که امنیت شغلی مهمترین عامل و پس از آن عوامل دیگری از جمله حجم کار و ارتباط شغل با حوزه تحصیلی فرد بر ترک شغل تأثیرگذارند. دقت پیش‌بینی مدل XGB بیش از ۷۸/۵ درصد بوده که بالاتر از مدل‌های سنتی اقتصادسنجی بوده است.

چاودوری^۲ و همکاران (۲۰۲۳) به این موضوع می‌پردازد که علی‌رغم کاربرد بسیار زیاد الگوریتم‌های یادگیری ماشین در پیش‌بینی ترک شغل نیروی کار، اما نتایج این مدل‌ها به دلیل عدم شفافیت و عدم قابلیت تفسیرپذیری چندان نمی‌تواند مورد استفاده مدیران قرار گیرد. در این راستا، این مقاله هدف خود را ارائه یک چارچوب شفاف از نتایج هوش مصنوعی با استفاده از تکنیک‌های LIME^۳ معرفی می‌کند که به دنبال تقریب مدل‌های غیرشفاف یادگیری ماشین با مدل‌های موضعی قابل تفسیر هستند. این مقاله نشان می‌دهد که این تکنیک‌ها راهنمای خوبی برای مدیران برای درک نتایج مدل‌های یادگیری ماشینی هستند و قابلیت کاربردی بیشتری به آن‌ها می‌بخشند.

^۱. Park

^۲. Chowdhury

^۳. Local Interpretable Model-Agnostic Explanations

یون و یو^۱ (۲۰۲۱) راهبردهای کاهش جابه‌جایی پرستاران بیمارستانی در کره جنوبی را بررسی کرده‌اند. به بیان آن‌ها با وجود تلاش‌های اخیر دولت برای بهبود شرایط کاری، نرخ بالای جابه‌جایی پرستاران بیمارستانی همچنان یک چالش جدی برای سیستم مراقبت‌های بهداشتی کره جنوبی است. این مطالعه یک مطالعه کیفی با استفاده از مصاحبه گروهی متمرکز با ۳۳ پرستار از ۱۱ بیمارستان کره جنوبی است. بر اساس نتایج بدست آمده، سه موضوع اصلی مرتبط با کاهش جابه‌جایی پرستاران بیمارستان شناسایی شد که عبارتند از شرایط محیط کاری، ویژگی‌های سرپرست و نوع خدمت‌رسانی به بیماران. نتایج همچنین نشان می‌دهند که برای رسیدگی به کمبود پرستار در بیمارستان‌ها در کره جنوبی، رسیدگی به محیط‌های کاری ضعیف مهم است.

باسوک^۲ و همکاران (۲۰۲۱) شواهد جدیدی در مورد ترک شغل معلمان ارائه می‌کنند. نتایج شواهد جدیدی را در مورد شیوع ترک شغل ارائه می‌کند و بررسی می‌کند که آیا معلمانی که شغل خود را ترک می‌کنند با سایر معلمان، تفاوت دارند یا خیر. نتایج نشان می‌دهد که بیش از یک سوم معلمان، شغل خود را هر سال ترک می‌کنند و اکثریت قریب به اتفاق معلمانی که شغل خود را ترک می‌کنند، در سال بعد تدریس نمی‌کنند. همچنین معلمانی که شغل خود را ترک می‌کنند، کیفیت تعامل پایین‌تری دارند. نتایج همچنین نشان می‌دهد که شرایط کاری مهمترین عامل مؤثر بر ترک شغل معلمان است.

مک‌درمید^۳ و همکاران (۲۰۲۰) در یک مطالعه مروری به عوامل مؤثر بر نرخ ترک شغل بالای پرستاران اورژانس پرداخته‌اند. این مطالعه مروری با استفاده از تمام مطالعاتی که در پایگاه‌های اطلاعاتی الکترونیکی، تجزیه و تحلیل ادبیات پزشکی و سیستم بازیابی آنلاین، فهرست تجمعی ادبیات پرستاری و مرتبط با سلامت به طور سیستماتیک بین سال‌های ۲۰۰۶ و ۲۰۱۸ منتشر شده‌اند نشان می‌دهد که در مجموع ۲۰ مقاله، شامل ۱۶ مطالعه کمی و چهار مطالعه کیفی، عواملی را شناسایی کردند که به نرخ ترک شغل بالا منجر می‌شود. این عوامل را می‌توان در سه موضوع اصلی مؤثر بر ترک شغل طبقه‌بندی کرد؛ پرخاشگری و خشونت در محیط کار، بروز حوادث بحرانی در محیط کار و نامناسب بودن محیط کار. این مطالعه بیان می‌کند که چالش‌های متعددی

¹. Yun and Yu

². Bassok

³. McDermid

برای پرستاران شاغل در اورژانس وجود دارد که ممکن است منجر به استرس شغلی، فرسودگی شغلی، خستگی و اختلال استرس پس از سانحه شود که به نرخ ترک شغل کمک می‌کند.

ژائو^۱ و همکاران (۲۰۱۸) در تحقیقی با استفاده از ۱۰ روش یا الگوریتم یادگیری ماشین به پیش‌بینی تمایل به ترک شغل کارکنان پرداخته‌اند. نتایج مقایسه عملکرد الگوریتم‌ها نشان دادند که برای مجموعه داده کوچک الگوریتم‌های مختلف کارایی دارند و اما برای مجموعه داده‌های بزرگ، الگوریتم افزایش گرادیان افراطی^۲ به‌عنوان مطمئن‌ترین الگوریتم توصیه می‌شود. این روش به کمترین پیش‌پردازش داده نیاز دارد، قدرت پیش‌بینی مناسبی داشته و به طور اتوماتیک و قابل اطمینان اهمیت ویژگی را درجه‌بندی می‌کند. در این تحقیق، تأثیر عواملی از قبیل شرایط زندگی فرد (محل زندگی و وضعیت تأهل)، ویژگی‌های فردی و شخصیتی (سن و جنسیت)، شرایط کار (طبقه شغلی، نوع قرارداد استخدامی و سنوات خدمت) بر ترک شغل بررسی شده است. بر اساس نتایج بدست آمده مشخص شد که افراد متأهل تمایل کمتری به ترک شغل در یک سازمان دارند. همچنین بین سن و ترک شغل نیز یک رابطه منفی مشاهده شد به این معنی که افراد با سن بالاتر تمایل کمتری به ترک شغل دارند. بین جنسیت افراد و ترک شغل اما رابطه معینی یافت نشد. همچنین از عوامل مرتبط با شرایط کار نیز مشاهده شد که هر سه عامل یعنی طبقه شغلی، نوع قرارداد استخدامی و سنوات بر ترک شغل مؤثرند. افراد با پست شغلی بالاتر تمایل کمتری به ترک شغل دارند؛ ترک شغل در استخدام‌های رسمی کمتر از قراردادی است؛ و هرچه سنوات خدمت بیشتر باشد، ترک شغل کمتر است.

با مرور این پیشینه از مطالعات داخلی و خارجی می‌توان دید که اغلب مطالعات خارجی و تمامی مطالعات داخلی که تاکنون در بررسی عوامل مؤثر بر ترک شغل در حوزه‌های مختلف انجام شده مطالعاتی کیفی بر مبنای مصاحبه و پرسشنامه بوده‌اند. این مطالعه از آنجایی که در بررسی خود از داده‌های واقعی استفاده کرده و یک تحلیل کمی با استفاده از روش‌های یادگیری ماشین ارائه می‌دهد، می‌تواند تا حد زیادی درک ما از دلایل ترک شغل نیروی کار را بهبود بخشد و این مهمترین نوآوری این پژوهش است. علاوه بر این، تاکنون در هیچ تحقیقی به بررسی و پیش‌بینی

¹. Zhao

². Extreme gradient increase algorithm

نرخ ترک شغل در فروشگاه‌های زنجیره‌ای پرداخته نشده که این موضوع نیز یکی از نوآوری‌های این مقاله حاضر است.

۳- روش پژوهش و آمار توصیفی

برای تحلیل تجربی و آماری، نقطه شروع با معرفی متغیرهای پژوهش و بررسی آمار توصیفی داده‌هاست که این مهم در زیربخش اول از این بخش انجام می‌شود. پس از درک و شناخت داده، زیربخش دوم به پیش پردازش^۱ یا آماده‌سازی داده‌ها اختصاص دارد. داده‌ها در این مرحله برای پردازش و تجزیه و تحلیل در مدل آماده می‌شوند. در زیربخش سوم نیز انواع الگوریتم‌های مورد استفاده معرفی خواهند شد.

۳-۱- داده‌های آماری و متغیرهای پژوهش

جامعه آماری مورد استفاده در این پژوهش اطلاعات شخصی و شغلی کارکنان تمام شعب یک فروشگاه‌های زنجیره‌ای در سراسر کشور در بازه زمانی سال ۱۳۹۸ الی ۱۴۰۱ را در بر می‌گیرد که شامل ۱۷۵۴۰ نفر می‌باشد. این اطلاعات از سیستم مدیریت منابع سازمانی^۲ (SAP) که از سال ۱۳۹۸ در این فروشگاه استقرار یافته بدست آمده و اطلاعات پیش از این تاریخ طبیعتاً در دسترس نیست. از آنجایی که اطلاعات پرسنل در طول دوره همکاری از بدو استخدام تا قطع رابطه در سیستم SAP به روزرسانی و ثبت می‌گردد، لذا منبع مناسبی جهت استفاده می‌باشد. طبق بررسی انجام شده از این جامعه آماری حدود ۵۱/۷ درصد ترک شغل کرده و ۴۸/۳ درصد در شغل خود ادامه فعالیت می‌دهند و به این ترتیب با یک مجموعه داده متوازن روبرو هستیم. این مقاله در نظر دارد تا به پیش‌بینی و تعیین عوامل مؤثر فردی و سازمانی بر ترک شغل کارکنان بپردازد. برای این منظور باید از یکسری مشخصه‌های فردی و سازمانی (شغلی) که ثبت شده استفاده کرد. برای هر فرد در این سیستم ۳۲ مشخصه ثبت شده که از این میان تعداد ۱۲ مشخصه به عنوان

^۱. Dart Preprocessing

^۲. Enterprise Resource Planning (ERP)

متغیرهای اصلی انتخاب و مورد استفاده قرار گرفته‌اند. آمار توصیفی مربوط به این مشخصه‌های دوازده‌گانه در جدول ۱ ارائه شده است.

جدول ۱. آمار توصیفی متغیرهای پژوهش

نام متغیر	میانگین	کمترین	بیشترین
وضعیت خدمت (فعال=۰، ترک کار=۱)	۵۱/۷	۰	۱
جنسیت (مرد=۰، زن=۱)	۰/۴	۰	۱
سن (سال)	۳۱/۱	۱۸	۷۰
تأهل (متأهل=۱، مجرد=۰)	۰/۶۰۷	۰	۱
تعداد فرزندان	۰/۲۸۷	۰	۵
مقطع تحصیلی	۱/۵۶	۰	۵
سنوات خدمت (سال)	۲/۳	۰	۹
نوع قرارداد استخدامی	۰/۳	۰	۲
سمت سازمانی (عملیاتی=۰، ستادی=۱)	۰/۲۳	۰	۱
ساعت کار ماهانه (موظف و اضافه کار)	۱۴۵	۱۰۶	۲۲۰
تعداد جابه‌جایی	۱/۷	۰	۶
حوزه (استان) و زیرحوزه (شهرستان) کاری			

منبع: یافته‌های پژوهش.

- جنسیت: رابطه بین جنسیت و ترک کار در پژوهش‌های گذشته همواره مبهم بوده است. به عنوان مثال در مطالعه‌ی مارش و منری^۱ (۱۹۷۷) زنان نرخ بالاتری نسبت به مردان در ترک کار دارند، حال آن که هیلی^۲ و همکاران (۱۹۹۵) نشان دادند که بر خلاف مطالعات قبل رابطه مشخصی بین جنسیت و ترک کار نیست و یا الاین^۳ (۱۹۹۷) و سامرس و هندریکس^۴ (۱۹۹۱) به این نتیجه رسیدند که مردان بیشتر از زنان شغل خود را ترک می‌کنند به این علت که مردان نان‌آور خانواده هستند و انگیزه پیشرفت بیشتری نسبت به زنان دارند و در صورت نارضایتی شغلشان را جهت یافتن یک شغل جذابتر ترک

^۱. Marsh and Mannari

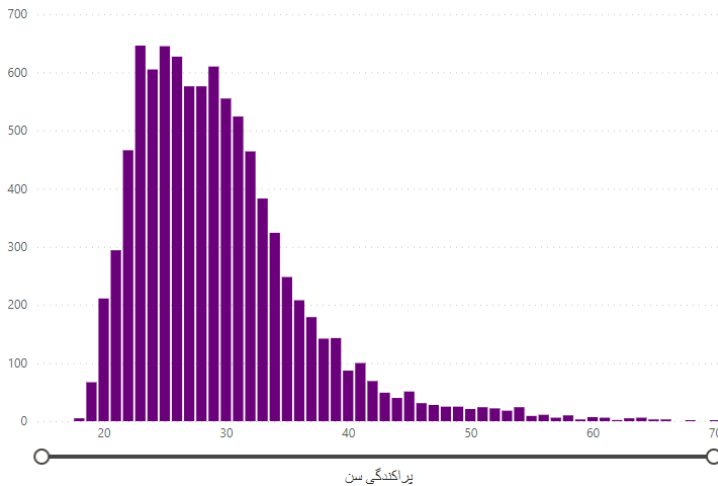
^۲. Healy

^۳. Elaine

^۴. Summers and Hendrix

کنند. در نمونه مورد بررسی فراوانی مردان در میان شاغلین حدود ۵۷ درصد و در میان افراد ترک کار کرده حدود ۶۲ درصد است که نشانگر بالاتر بودن احتمال ترک شغل در میان مردان است.

- سن: از متغیرهای تأثیرگذار بر ترک شغل و جابه‌جایی داوطلبانه سن فرد است. دونگان^۱ (۱۹۹۳) توضیح می‌دهد که جوانان چشم‌انداز شغلی و جای رشد بیشتری نسبت به افراد میانسال و مسن دارند. بعلاوه جوانان در شناسایی شغل‌های دارای تمام معیارهای مقبولیت مشکل بیشتری داشته و در نتیجه تمایل به جابه‌جایی شغلی و حرکت به سمت فرصت‌های شغلی جدید در آن‌ها بیشتر است. جوانان نسبت به افراد میانسال در پذیرش چالش‌ها پویاتر بوده و به دنبال آن ریسک ترک کار به منظور به چالش کشیدن خود بیشتر در آن‌ها اتفاق می‌افتد حال آن که قدرت ریسک با افزایش سن معمولاً در همه زمینه‌ها خصوصاً جابه‌جایی شغل کمتر است. مطابق شکل ۱ عمده افرادی که از ادامه کار انصراف داده و شغل خود را ترک نموده‌اند در بازه سنی ۲۳ الی ۳۰ سال قرار دارند.



شکل ۱. توزیع سنی ترک کار

منبع: یافته‌های پژوهش.

^۱. Dunegan

- تأهل و تعداد فرزندان: دو متغیر وضعیت تأهل و تعداد فرزندان به دنبال مفهوم رابطه کار و خانواده احتمالاً بر تمایل به ماندگاری در شغل و یا ترک کار مؤثر باشند در نمونه مورد بررسی حدود ۶۱ درصد افراد متأهل بوده‌اند. در میان افراد فعال نسبت افراد متأهل ۵۱ درصد بوده است که نشان می‌دهد فراوانی افراد متأهل در میان افراد ترک شغل کرده بسیار بیشتر است. بلحاظ تعداد فرزندان نیز عمده افراد در نمونه مورد بررسی یعنی حدود ۸۱ درصد بدون فرزند بوده‌اند و ۱۱/۵ درصد دارای ۱ و ۶/۵ درصد نیز دارای ۲ فرزند بوده‌اند.
- تحصیلات: مطالعات نشان‌دهنده آن است که افراد با تحصیلات بالا در صورت عدم رضایت از شغل و درآمد دریافتی تمایل بیشتری به ترک شغل خود دارند. در نمونه مورد بررسی در میان شاغلین فعال حدود ۵۹ درصد افراد تحصیلات دیپلم و زیر داشته‌اند در حالی که این نسبت در میان افراد ترک شغل کرده بالاتر و حدود ۶۶ درصد بوده است. این امر نشان می‌دهد که ترک شغل کارکنان فروشگاه‌های زنجیره‌ای ماهیتاً ارتباط مستقیمی با تحصیلات نداشته است. میانه تحصیلات در هر دو گروه شاغلین و افراد تحصیلکرده برابر دیپلم است. متغیر تحصیلات به صورت ترتیبی در شش سطح (زیر دیپلم=۰، دیپلم=۱، فوق دیپلم=۲، لیسانس=۳، فوق لیسانس=۴ و دکتری=۵) ثبت شده است.
- سنوات خدمت: سنوات خدمت نشان‌دهنده تجربه است. مطالعات نشان‌دهنده آن است که بیشترین درصد خروج و ترک سازمان در سال‌های ابتدایی اشتغال و در سنوات پایین اتفاق می‌افتد. نیروهای آماده به کار جهت کسب تجربه و سابقه بیمه با امکان جذب آسان‌تر وارد مجموعه‌هایی با مدل کار قرارداد ساعتی شده و پس از مدتی به منظور ارتقا مجموعه را به قصد یافتن شغل مناسب تر ترک خواهند کرد. در نمونه مورد بررسی و در میان افراد ترک شغل کرده حدود ۵۲ درصد سابقه کمتر از ۱ سال، ۲۳ درصد ۱ سال، ۱۲ درصد تا ۲ سال سابقه کار داشته‌اند. این امر نشان می‌دهد که تنها حدود ۱۰ درصد از موارد ترک شغل مربوط به افراد با سابقه کار بیش از ۲ سال بوده است.

- نوع قرارداد استخدامی: به طور خاص صناعی که نیاز به پرسنل پاره وقت دارند، نرخ ترک شغل بالاتری نیز دارند (اداره آمار امور کار ایالات متحده، ۲۰۲۳). این امر به دلایل مختلفی می‌تواند اتفاق بیافتد از جمله اینکه افرادی که کار پاره‌وقت را انتخاب می‌کنند به دنبال انعطاف‌پذیری در زمان کاری، تعادل کار-زندگی و... می‌باشند. در مقابل کارمندان با قراردادهای تمام وقت رسمی یا موقت معمولاً بهره‌وری بالاتری دارند و تعهدات بیشتری نیز به سازمان می‌دهند. نوع قرارداد استخدامی به صورت یک متغیر ترتیبی در سه سطح (ساعتی=۰، تمام وقت=۱ و رسمی=۲) ثبت شده است. در نمونه مورد بررسی و در میان افراد ترک شغل کرده حدود ۸۵ درصد ساعتی بوده‌اند در حالی که فراوانی قراردادهای ساعتی در میان افراد فعال و شاغل کمتر و حدود ۶۴ درصد بوده است که مؤید بالاتر بودن ترک شغل در قراردادهای ساعتی و پاره‌وقت است.
- سمت سازمانی: کارکنان در یک دسته بندی کلی به دو دسته پرسنل صف (کف فروشگاه) و ستادی دارای پست های نظارتی و مدیریتی تقسیم بندی شدند. در نمونه مورد بررسی، کادر عملیاتی یا پرسنل صف ۷۳ درصد جمعیت مورد مطالعه را تشکیل دادند.
- متوسط ساعت کار: به طور کلی ساعت کار درج شده در قرارداد و اضافه کار بالا الزاماً به معنای رضایت شغلی نیست چراکه ساعات طولانی کار بدون جبران مناسب خود می‌تواند منجر به خستگی و عدم رضایت و در نتیجه افزایش نرخ ترک شغل بشود. در فروشگاه‌های زنجیره‌ای، قانون کار برای کارکنان با قرارداد ساعتی طبق حداکثر سقف ساعت کارکرد ۲۲۰ ساعت می‌باشد و اضافه کار پرداخت نخواهد شد و برای کارکنان تمام وقت نیز سقف پرداخت اضافه کار در گروه‌های مختلف متفاوت است.
- تعداد جابه‌جایی: جابه‌جایی شغلی می‌تواند نشانگر عدم ثبات شغلی یا عدم انطباق شغل با توانمندی‌های فردی باشد و بر این اساس انتظار می‌رود که رابطه مستقیمی با ترک شغل داشته باشد. در نمونه مورد بررسی افراد به طور متوسط ۱/۷ بار جابه‌جایی شغلی را تجربه کرده‌اند.
- حوزه و زیرحوزه‌های کاری: منظور از حوزه کاری همان استان و منظور از زیرحوزه‌ها شهرستان محل خدمت است. دلیل ورود این مشخصه در مدل این است که بسیاری از

متغیرهای کلان مانند بیکاری، تورم، سطح کلی درآمدی و رفاهی منطقه و ... ممکن است بر ترک کار تأثیرگذار باشند اما از آنجایی که در این مقاله این عوامل کلان اقتصادی و اجتماعی به دلیل محدودیت داده مورد بررسی قرار نخواهند گرفت و تمرکز بر متغیرهای فردی و سازمانی است، با این وجود این متغیرها نیاز به کنترل دارند. در واقع محل خدمت پرسنل به عنوان متغیر کنترلی در مدل وارد شده است. در نمونه مورد بررسی حدود ۴۸/۲ درصد از افراد ساکن استان تهران بوده‌اند که بیشترین فراوانی را به خود اختصاص داده‌اند و پس از آن استان‌های اصفهان، فارس و خراسان رضوی در جایگاه بعدی قرار دارند.

۳-۲- آماده‌سازی داده‌ها

برای مدل‌سازی ابتدا باید یکسری اقدامات لازم تحت عنوان پیش‌پردازش داده‌ها، انتخاب ویژگی^۱ و هم‌مقیاس‌سازی^۲ به منظور آماده‌سازی داده‌های مورد نیاز انجام شود. پیش‌پردازش داده‌ها خود شامل چهار اقدام کنترل داده‌های طبقه‌بندی شده^۳، کنترل داده‌های گم‌شده^۴، کنترل داده‌های پرت^۵ و در آخر کنترل داده‌های تکراری^۶ می‌باشد. منظور از کنترل داده‌های طبقه‌بندی تبدیل همه متغیرها به متغیرهای عددی است. برای این منظور همانطور که در زیربخش قبلی گفته شد، متغیرهای جنسیت، وضعیت تأهل، تحصیلات، نوع قرارداد استخدامی و سمت سازمانی که کیفی بودند با اعداد جایگذاری شد. داده‌های متعلق به استان محل خدمت نیز با یک برچسب منحصر به فرد کدگذاری و برای هر استان (به جز استان تهران که به عنوان حالت پایه در نظر گرفته شده) یک متغیر مجازی دودویی ساخته شد. اقدام دوم، کنترل مقادیر گم‌شده‌ای است که اغلب در داده‌های دنیای واقعی وجود دارند. مدیریت داده‌های گم‌شده مهم است زیرا وجود مقادیر گم‌شده در یک مجموعه داده می‌تواند باعث ایجاد خطا در برخی از الگوریتم‌های یادگیری ماشین شود. در این پژوهش از بین ۱۷۶۴۴ رکورد داده استخراج شده

1. Feature Selection

2. Feature Scaling

3. Categorical data

4. Missing values

5. Outlier data

6. Duplicated data

۱۳۴ رکورد حاوی مقادیر گم شده در متغیر تاریخ استخدام بوده که این داده‌ها حذف گردید. یکی دیگر از اقدامات مهم در پیش‌پردازش، تمیز کردن داده‌ها از مقادیر پرت است. داده‌های پرت خارج از محدوده مورد انتظار بوده و بر خلاف سایر داده‌ها هستند و می‌تواند به دلایل مختلفی از جمله خطا در اندازه‌گیری یا ورودی و انحراف داده رخ دهد. برای شناسایی داده‌های پرت معمولاً از نمودار جعبه‌ای^۱ و محاسبه چارک‌ها استفاده می‌شود. با بررسی انجام شده روی متغیرها تنها در مشخصه ساعت اضافه کار و میانگین ساعت کار بود که تعداد ۲۰ مشاهده از مجموع کل داده‌ها با مرتب‌سازی و محاسبه چارک اول تا سوم، به عنوان داده پرت شناسایی و حذف گردید. و در نهایت، یک مرحله دیگر از پیش‌پردازش حذف داده‌های تکراری است که باعث خواهد شد تا الگوریتم بهتر عمل خواهد کرد. از دیدگاه ارزیابی الگوریتم، ردیف‌های تکراری منجر به عملکرد گمراه کننده می‌شود (براون‌لی^۲، ۲۰۲۰). در این مجموعه داده‌ها با توجه به اینکه هر فرد یک کد پرسنلی منحصر به فرد داشته و کدهای ایجاد شده در سیستم SAP^۳ به صورت اتوماتیک تولید خواهند شد، هیچ ردیف تکراری در آن وجود نداشت.

انتخاب ویژگی یکی دیگر از مراحل به منظور آماده‌سازی داده‌هاست. انتخاب ویژگی‌های ورودی که از یک سو از یکدیگر استقلال بیشتری داشته و از سوی دیگر مهم باشند به این معنی که با متغیر هدف بیشترین ارتباط را داشته باشند می‌تواند در ساده‌سازی و سرعت بخشیدن به فرآیند مدل‌سازی، کاهش میزان محاسبات مورد نیاز و همچنین توسعه مدل اهمیت زیادی داشته باشد. این مقاله یک متغیر هدف دارد و آن پیش‌بینی ترک کار است. یکی از روش‌های انتخاب ویژگی، روش فیلتر^۴ است که مستقل از الگوریتم مورد استفاده بوده و از معیارهای آماری مثل همبستگی بین متغیرهای ورودی برای شناسایی ویژگی‌های مهم استفاده می‌کند. یکی از مشهورترین شیوه‌های اندازه‌گیری وابستگی بین دو متغیر، محاسبه ضریب همبستگی پیرسون^۵ است که بر اساس کوواریانس قابل محاسبه است. با توجه به نتایج بدست آمده مشخص شد که

^۱. Box plot

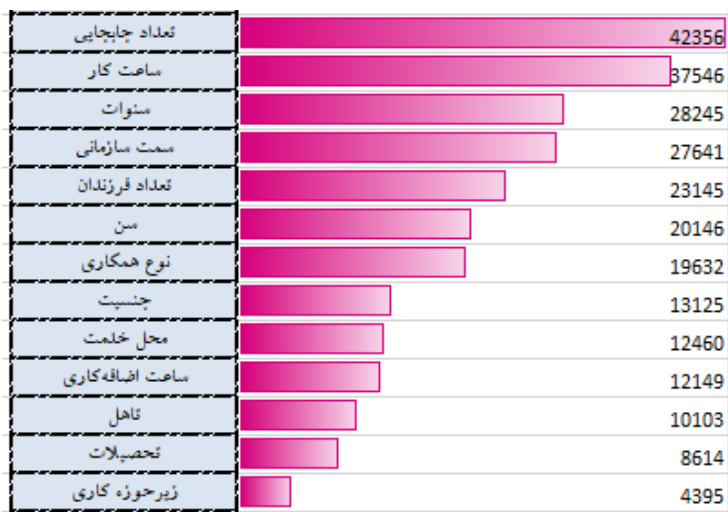
^۲. Brownlee

^۳. یک نرم‌افزار سازمانی ارائه شده توسط شرکت آلمانی SAP که به منظور مدیریت عملیات تجاری و قابلیت تحلیل و گزارش دهی کارکرد نیروها توسط شرکت‌ها استفاده می‌شود.

^۴. Filter method

^۵. Pearson correlation coefficient

مقدار ضریب همبستگی میان پارامترهای ورودی همبستگی بسیار کمی با هم داشته و از نظر استقلال می‌توان هر یک را به عنوان ورودی به صورت مجزا در نظر گرفت. میزان رابطه این متغیرها با متغیر تصمیم (ترک کار) نیز به صورت مرتب شده در شکل ۲ نمایش داده شده است. این شکل بر اساس امتیازبندی روش فیلتر تنظیم شده است. امتیاز هر مشخصه میزان آماره مربع کای است که محاسبات آن با توجه به فراوانی در هر گروه از مشاهدات و طبقه‌بندی آن و متغیر هدف کارمندان فعال و قطع رابطه شده صورت گرفته است. از میان ویژگی‌های ورودی و متغیرهای مورد استفاده تنها متغیر مربوط به زیرحوزه‌های کاری (شهرستان‌ها) رابطه معناداری با متغیر هدف یعنی ترک کار نداشته و از آنجایی که احتمالاً به شدت به زمان اجرای مدل می‌افزود حذف گردید و تنها به ویژگی استان محل خدمت اکتفا گردید.



شکل ۲. امتیاز و رتبه بندی متغیرهای ورودی در فرآیند انتخاب ویژگی

منبع: یافته‌های پژوهش.

هم‌مقیاس‌سازی ویژگی‌ها یک مرحله از آماده‌سازی داده‌ها است که توصیه می‌شود هنگام کار با بسیاری از الگوریتم‌های یادگیری ماشین انجام شود. منظور از هم‌مقیاس‌سازی این است بازه تغییرات اعداد در هر ویژگی یکسان شود تا تأثیرات غیردلخواهی بر روی یکدیگر و الگوریتم

نگذارند. یک روش متداول در هم مقیاس‌سازی داده‌ها استانداردسازی^۱ است که بر اساس رابطه ۱ مقیاس متغیرها را بین ۰ و ۱ قرار می‌دهد.

$$y = \frac{x - \min}{\max - \min} \quad (۱)$$

۳-۳- الگوریتم‌های مدل‌سازی

در این مقاله برای پیش‌بینی ترک کار از دو الگوریتم مدل‌سازی ماشین بردار پشتیبان^۲ و درخت تصمیم^۳ استفاده شد که در ادامه با جزئیات بیشتر این دو الگوریتم معرفی می‌شوند. استفاده از دو الگوریتم در کنار یکدیگر امکان مقایسه عملکرد آن‌ها در پیش‌بینی را فراهم می‌آورد. علت انتخاب الگوریتم درخت تصمیم سادگی و تفسیرپذیری آن است. عمده الگوریتم‌های یادگیری ماشین قابلیت تفسیرپذیری نداشته و مانند یک جعبه سیاه عمل می‌کنند اما از آنجایی که این مقاله در کنار پیش‌بینی به دنبال شناسایی عوامل مؤثر بر ترک شغل نیز بود، ترجیح داده شد تا از این الگوریتم استفاده شود.

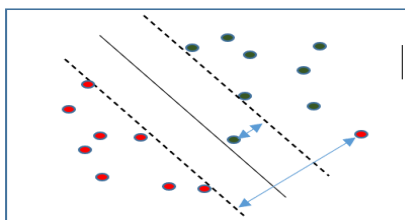
۳-۳-۱- الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. مبنای کاری دسته‌بندی‌کننده ماشین بردار پشتیبان (SVM) دسته‌بندی خطی داده‌ها است به این صورت که سعی می‌شود ابرصفحه‌ای انتخاب گردد که حاشیه اطمینان بیشتری داشته باشد. ابعاد ابرصفحه به تعداد ویژگی‌ها بستگی دارد به عنوان مثال اگر تعداد ویژگی‌های ورودی دو باشد، آنگاه ابرصفحه فقط یک خط است. شکل ۳ دسته‌بندی ماشین بردار پشتیبان را در این حالت ساده به تصویر می‌کشد.

¹. Standardization

². Support Vector Machine

³. Decision Tree Algorithm



شکل ۳. بیان دسته‌بندی بردار پشتیبان در حالت ساده با دو ویژگی

منبع: یافته‌های پژوهش.

ماشین‌های بردار پشتیبان به دنبال یافتن بهترین معادله جداکننده دسته‌ها (کلاس‌های تابع هدف) از یکدیگر هستند. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های برنامه‌سازی غیرخطی که روش‌های شناخته شده‌ای در حل مسائل مقید هستند، صورت می‌گیرد. قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند، داده‌ها به وسیله تابعی به فضای با ابعاد خیلی بالاتر^۱ برده می‌شود. برای اینکه بتوان مسئله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کرد از قضیه دوگانی لاگرانژ^۲ برای تبدیل مسئله مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده برای انتقال به فضایی با ابعاد بالا از تابع ساده‌تری به نام تابع هسته (کرنل) استفاده می‌شود که یک ضرب برداری است. از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموئید می‌توان استفاده نمود. ماشین بردار پشتیبان با ترسیم یک خط راست، داده‌های مختلف و متمایز از یکدیگر را دسته‌بندی می‌کند. به بیان ساده، بردارهای پشتیبان در واقع مختصات یک مشاهده منفرد هستند. ماشین بردار پشتیبان مرزی است که به بهترین شکل دسته‌های داده‌ها را از یکدیگر جدا می‌کند.

الگوریتم بردار پشتیبان مزایای زیادی به‌عنوان یک الگوریتم طبقه‌بندی دارد. نخست این که در این الگوریتم حاشیه جداسازی برای دسته‌های مختلف کاملاً واضح است. دوم این که این الگوریتم در فضاهای با ابعاد بالاتر کارایی بیشتری دارد. این الگوریتم در شرایطی که تعداد ابعاد بیش از تعداد نمونه‌ها باشد نیز کار می‌کند. سوم این که این الگوریتم از یک زیرمجموعه از نقاط تست در تابع تصمیم‌گیری استفاده می‌کند (که به آن‌ها بردارهای پشتیبان گفته می‌شود) و همین

^۱. High Dimensional Space

^۲. Lagrange Duality Theorem

باعث می‌شود در مصرف حافظه نیز به صورت بهینه عمل می‌کند. از دیگر مزیت‌های روش بردار پشتیبان می‌توان به یادگیری آسان، قابلیت استفاده و عملکرد خوب برای همه مدل‌های ساده و پیچیده، عملکرد خوب با تعداد نمونه‌های آموزشی کم اشاره کرد و همچنین این‌که در کمینه محلی گیر نمی‌افتد و کمینه سراسری را محاسبه می‌کند. با وجود این مزیت‌ها، اما الگوریتم بردار پشتیبان (SVM) دارای معیایی نیز است که باید نسبت به آن‌ها آگاهی داشت. نخست این‌که هنگامی که مجموعه داده‌ها بسیار بزرگ باشد، عملکرد خوبی ندارد، زیرا نیازمند زمان آموزش بسیار زیاد است. دوم این‌که هنگامی که مجموعه داده نوفه (نویز) زیادی داشته باشد، عملکرد خوبی ندارد و کلاس‌های هدف دچار همپوشانی می‌شوند. و سوم این‌که ماشین بردار پشتیبان به طور مستقیم تخمین‌های احتمالاتی را فراهم نمی‌کند و این موارد با استفاده از یک اعتبارسنجی متقابل پرهزینه پنج‌گانه انجام می‌شوند. احتمال رخ دادن بیش‌برازش^۱ نیز در این روش بالاست (چوهان^۲ و همکاران، ۲۰۱۹). برای جلوگیری از بیش‌برازش در ماشین بردار پشتیبان ضریب C یا منظم‌سازی^۳ که یکی از پارامترهای اصلی^۴ این الگوریتم است نباید مقدار بالایی به خود بگیرد. ضریب C همان هزینه خطای دسته‌بندی است. یک C بزرگ به این معنی است که خطای طبقه بندی باید صفر باشد که به تشدید قیدهای مسأله^۵ بهینه‌سازی منجر می‌شود. چنین مسأله بهینه‌سازی به دو مشکل عدم همگرایی و بیش‌برازش منجر خواهد شد.

۳-۳-۲- الگوریتم درخت تصمیم

درخت تصمیم یک نوع از الگوریتم یادگیری تحت نظارت است که در مسائل طبقه‌بندی استفاده می‌شود. یک نمایش ساده از درخت تصمیم در دسته‌بندی در شکل ۴ آورده شده است. یک ویژگی هدف به نام «طبقه‌بند» وجود دارد که در اینجا ویژگی هدف همان ترک شغل است. به هر گره داخلی یکی از ویژگی‌های برچسب زده می‌شود. یال‌های خارج شده از گره داخلی به یک گره برگ برای دسته‌بندی داده‌ها یا یک گره داخلی برای دسته‌بندی داده‌ها برحسب یک ویژگی دیگر هدایت

¹. Over fitting

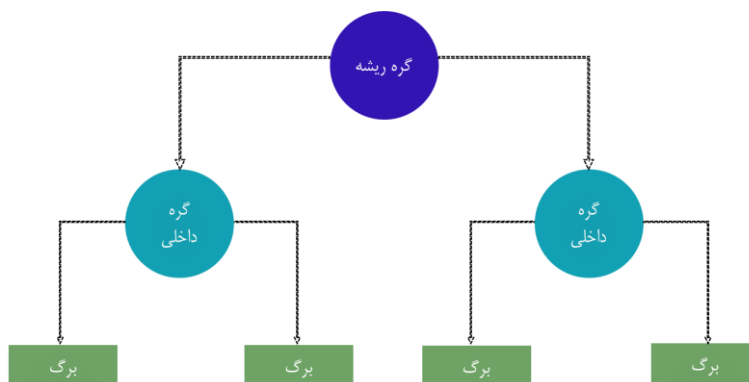
². Chauhan

³. Regularization

⁴. Hyperparameter

⁵. Hard margins

می‌شوند. درنهایت، هر گره برگ برحسب یکی از مقادیر در دامنه طبقه‌بند برچسب زده می‌شوند و همه داده‌هایی که در آن گره قرار بگیرند، به عنوان عضو آن کلاس پیش‌بینی خواهند شد.



شکل ۴. یک مثال از ساختار درخت تصمیم

منبع: یافته‌های پژوهش.

روند ایجاد درخت تصمیم از تقسیم مجموعه تمام داده‌های در دسترس (که در واقع ریشه درخت را تشکیل می‌دهند)، به زیر مجموعه‌هایی (که به هر گره فرزند تبدیل می‌شوند) است. شرایط تقسیم داده‌های هر گره به گره‌های فرزند آن براساس یک شرط تقسیم‌بندی در هر گره داخلی می‌باشد. تقسیم هر گره بدست آمده به گره‌های فرزند آن در یک روند بازگشتی انجام می‌گیرد که به آن تقسیم‌بندی بازگشتی نیز گفته می‌شود. این تقسیم‌بندی تا زمانی ادامه پیدا خواهد کرد که داده‌های در هر گره برگ متعلق به یک کلاس باشند یا زمانی که دیگر ویژگی‌ای برای تقسیم داده‌ها در یک گره وجود نداشته باشد. به این فرایند القای درخت تصمیم از بالا به پایین گفته می‌شود که یک مثال از الگوریتمی حریصانه است. الگوریتم‌های مختلفی برای ساخت درخت تصمیم وجود دارد که اغلب بر مبنای جستجو بالا به پایین در درخت تصمیم عمل می‌کنند. الگوریتم هانت^۱ قدیمی‌ترین الگوریتم یادگیری در درخت تصمیم است که در سال ۱۹۶۶ معرفی شد و پس از آن الگوریتم‌های دیگری از جمله C4,5, ID3 و CART نیز معرفی شدند (الصغیر^۲ و همکاران، ۲۰۱۷).

^۱. Hunt

^۲. Alsagheer

یکی از نکات کلیدی در الگوریتم درخت تصمیم، معیار انتخاب مشخصه برای انشعاب درخت تصمیم است. دو معیار انتخاب مشخصه شناخته شده عبارتند از کسب اطلاعات (آنتروپی)^۱ و شاخص جینی^۲ است. آنتروپی در رابطه ۲ نمایش داده شده است. جایی که عبارت‌های S مجموعه داده که آنتروپی آن محاسبه می‌شود، C کلاس‌های مجموعه S و $P(c)$ نسبت نقاط داده‌ای متعلق به کلاس C به تعداد کل نقاط داده در مجموعه S است. مقادیر آنتروپی می‌تواند بین ۰ و ۱ قرار گیرد. اگر تمام نمونه‌های مجموعه داده S متعلق به یک کلاس باشند، آنتروپی برابر صفر خواهد بود. اگر نیمی از نمونه‌ها به عنوان یک کلاس و نیمی دیگر در کلاس دیگری طبقه‌بندی شوند، آنتروپی به بالاترین حد خود یعنی ۱ خواهد رسید. برای انتخاب بهترین ویژگی جهت تقسیم و یافتن درخت تصمیم بهینه نیز باید از ویژگی (مشخصه یا متغیر) با کمترین مقدار آنتروپی استفاده شود. به تفاوت آنتروپی قبل و بعد از تقسیم در یک ویژگی معین نیز اطلاعات بدست آمده گفته می‌شود. ویژگی با بالاترین کسب اطلاعات بهترین تقسیم را ایجاد می‌کند. کسب اطلاعات با رابطه ۳ نشان داده می‌شود جایی که α یک ویژگی یا برچسب کلاس خاص بوده و $\frac{|S_v|}{|S|}$ نسبت مقادیر در مجموعه داده S_v به تعداد مقادیر در مجموعه داده S است. ابتدا آنتروپی گره والد محاسبه می‌شود. سپس آنتروپی هر گره جداگانه در تقسیم (انشعاب) فرموله و در انتها میانگین وزنی تمام گره‌ها محاسبه می‌شود.

$$Entropy(S) = - \sum_{c \in C} p(c) \log_2(p(c)) \quad (2)$$

$$Information\ Gain(S, \alpha)$$

$$= Entropy(S) - \sum_{v \in values(\alpha)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (3)$$

معیار دوم یعنی ناخالصی جینی اندازه‌گیری می‌کند که با چه احتمالی یک عنصر از یک مجموعه می‌تواند به اشتباه برچسب‌گذاری (دسته‌بندی) شود اگر که دسته‌بندی عناصر در مجموعه به‌طور تصادفی براساس یک توزیع احتمال انجام شده باشد. ناخالصی جینی را می‌توان با جمع کردن احتمال‌های P_i برای یک عنصر که برچسب i برای آن انتخاب شده در احتمال دسته‌بندی

¹. Entropy

². Gini Impurity

اشتباه آن که برابر است با $\sum_{k \neq i} P_k = 1 - P_i$ بدست آورد. کمترین مقدار ممکن برای این معیار برابر صفر است که در این حالت تمام المان‌های در یک مجموعه به یک کلاس تعلق دارند. رابطه ۴ مقدار ناخالصی جینی را برای یک مجموعه از المان‌ها با J کلاس که $i \in \{1, 2, \dots, J\}$ نشان داده است جایی که P_i نسبتی از داده‌ها است که با برچسب i در این مجموعه برچسب زده شده‌اند:

$$\begin{aligned} I_G(p) &= \sum_{i=1}^J \left(p_i \sum_{k \neq i} p_k \right) = \sum_{i=1}^J p_i (1 - p_i) \\ &= \sum_{i=1}^J p_i - \sum_{i=1}^J p_i^2 = 1 - \sum_{i=1}^J p_i^2 \end{aligned} \quad (۴)$$

اینکه همه نقاط داده در مجموعه‌های همگن جداگانه طبقه‌بندی شوند یا نه، تا حد زیادی به پیچیدگی درخت تصمیم بستگی دارد. با افزایش اندازه درخت، حفظ این خلوص به طور فزاینده‌ای دشوار می‌شود و معمولاً باعث می‌شود که داده‌های بسیار کمی در یک زیردرخت معین قرارگیرند که اغلب می‌تواند منجر به بیش برآزش شود. برای کاهش پیچیدگی و جلوگیری از برآزش بیش از حد، با استفاده از هرس^۱ یا معیارهای توقف انجام می‌شود. هرس فرآیندی است که شاخه‌هایی را که بر روی ویژگی‌های با اهمیت کم تقسیم می‌شوند را حذف می‌کند. در واقع، هرس عکس فرایند تقسیم کردن است. یکی روش معیار توقف نیز به این صورت است که مرحله رشد درخت و تقسیم (انشعاب) گره‌ها به زیرگره‌های بیشتر ادامه می‌یابد تا زمانی که یک معیار توقف فعال شود. یکی از مهم‌ترین معیارهای توقف حداکثر عمق درخت می‌باشد که توسط کاربر مشخص می‌شود.

الگوریتم درخت تصمیم مزایای زیادی به عنوان یک الگوریتم طبقه‌بندی دارد. بر اساس ویژگی‌های برشمرده شده برای این الگوریتم توسط روکاخ و مائمون^۲ (۲۰۱۴) می‌توان از مهمترین نقاط قوت و ضعف الگوریتم درخت تصمیم به موارد زیر اشاره کرد. نخست این که این الگوریتم درک و تفسیر ساده‌ای دارد. دوم این که سایر تکنیک‌ها معمولاً در تجزیه و تحلیل مجموعه‌ای از

^۱. Pruning

^۲. Rokach and Maimon

داده‌ها با فقط یک نوع متغیر کاربرد دارند در حالی که الگوریتم درخت تصمیم در نسخه‌های جدیدتر مانند C4.5 این محدودیت را نداشته و قادر به رسیدگی به داده‌های عددی و اسمی است. سوم این که در الگوریتم درخت تصمیم نیاز به آماده‌سازی داده‌ها کم است در حالی که سایر تکنیک‌ها اغلب به نرمال‌سازی داده‌ها نیاز دارند. چهارم این که این الگوریتم از یک جعبه سفید یا جعبه باز استفاده می‌کند به این معنی که اگر یک موقعیت معین در یک مدل قابل مشاهده باشد، توضیح این شرط به راحتی با منطق بولی توضیح داده می‌شود. در مقابل، سایر الگوریتم‌ها مثلاً شبکه عصبی مصنوعی یک مدل جعبه سیاه است که توضیح نتایج آن معمولاً به سختی قابل درک است. پنجم این که در قبال مجموعه داده‌های بزرگ عملکرد خوبی دارد به این معنی که در الگوریتم درخت تصمیم مقادیر زیادی از داده‌ها را می‌توان با استفاده از منابع محاسباتی استاندارد در زمان معقول تجزیه و تحلیل کرد. و در نهایت، مزیت ششم این که در انتخاب ویژگی خود ساخته‌است. از ویژگی‌های نامربوط اضافی کمتر استفاده می‌شود تا بتوان آن‌ها را در اجراهای بعدی حذف کرد. سلسله مراتب صفات در درخت تصمیم نشان دهنده اهمیت صفات است. این بدان معنی است که ویژگی‌های در بالای درخت آموزنده‌ترین هستند. با این وجود، الگوریتم درخت تصمیم نواقص و معایبی نیز دارد. نخست این که درختان تصمیم می‌توانند بسیار غیرمستحکم باشند به گونه‌ای که یک تغییر کوچک در داده‌های آموزشی می‌تواند منجر به تغییر بزرگ در درخت و در نتیجه پیش‌بینی‌های نهایی شود. دوم این که برازش بیش از حد از رایج‌ترین مشکلات مدل‌های درخت تصمیم است اگرچه این مشکل را می‌توان با استفاده از محدودیت در پارامترهای مدل و همچنین هرس حل کرد. و سوم این که الگوریتم درخت تصمیم اگرچه می‌تواند با متغیرهای پیوسته نیز کار کند اما برای این متغیرها مناسب نیست.

۴- یافته‌های پژوهش

پس از آماده‌سازی داده‌ها، دو الگوریتم یادگیری ماشین درخت تصمیم و ماشین بردار پشتیبان به منظور طبقه‌بندی و پیش‌بینی ترک شغل یا ادامه فعالیت کارکنان یک فروشگاه‌های زنجیره‌ای به کار گرفته می‌شوند. برای اجرای این دو الگوریتم، مجموعه داده‌ها شامل ۱۷۵۴۰ نفر به دو قسمت ۱۲۲۷۸ نفر مجموعه داده آموزش (۷۰ درصد) و ۵۲۶۲ نفر مجموعه داده تست یا آزمون (۳۰

درصد) تقسیم شد. از مجموع داده‌های آموزش نیز ۱۷۵۴ نفر (۱۰ درصد) برای اعتبارسنجی و تنظیم پارامترهای اصلی استفاده شده است.

۴-۱- پیاده‌سازی و ارزیابی عملکرد مدل‌ها

نتایج و یافته‌های پیاده‌سازی دو الگوریتم ماشین بردار پشتیبان و درخت تصمیم برای پیش‌بینی ترک شغل یا ادامه فعالیت کارکنان یک فروشگاه‌های زنجیره‌ای در ادامه ارائه شده است.

۴-۱-۱- یافته‌های الگوریتم ماشین بردار پشتیبان

در این مقاله از تابع کرنل خطی در الگوریتم دسته‌بند ماشین بردار پشتیبان استفاده شده است. در کرنل خطی تنها یک پارامتر اصلی وجود دارد که باید به نحو مناسبی تعیین شود و بر نتایج تأثیرگذار است. پارامتر اصلی کرنل خطی ضریب C یا همان هزینه خطای دسته‌بندی است که مقدار بالای آن به مشکل عدم همگرایی و بیش‌برازش و مقدار کم آن نیز به مشکل حساسیت نتایج به داده‌های پرت^۱ منجر خواهد شد. در تنظیم این پارامتر به این شیوه عمل شد که برای رسیدن به بهترین عملکرد، الگوریتم بارها اجرا گردید و نهایتاً بهترین مدل یا مقدار بهینه این پارامتر برابر $C = 0.01$ انتخاب شد تا بیش‌برازش اتفاق نیفتد. پس از ایجاد مدل و تنظیم پارامترها، مجموعه داده آزمایش به کار گرفته شد تا درستی و دقت آن مورد ارزیابی قرار گیرد. نتایج و یافته‌های الگوریتم ماشین بردار پشتیبان در جدول ۲ که ماتریس اغتشاش^۲ نام دارد، نمایش داده شده است.

جدول ۲. نتایج ماتریس اغتشاش الگوریتم ماشین بردار پشتیبان

		پیش‌بینی مدل	
		ترک شغل	ادامه فعالیت
موارد واقعی	ترک شغل	۲۳۴۰ (TP)	۴۰۱ (FN)
	ادامه فعالیت	۶۵ (FP)	۲۴۵۶ (TN)

منبع: یافته‌های پژوهش.

^۱. Outlier

^۲. Confusion matrix

مطابق آنچه گفته شد ۳۰ درصد داده‌های آزمایش ۵۲۶۲ نمونه بوده است. از این میان تعداد ۲۳۴۰ نفر ترک شغل داشته (مثبت یا مقدار ۱) که مدل به درستی آن‌ها را پیش‌بینی و در گروه کارکنان ترک شغل قرار داده است. به این دسته در اصطلاح مثبت واقعی (TP) گفته می‌شود. تعداد ۲۴۵۶ نفر نیز در سازمان مانده و به فعالیت خود ادامه داده‌اند (منفی یا مقدار ۰) که مدل به درستی آن‌ها را پیش‌بینی و در گروه افرادی که در سازمان مانده‌اند دسته‌بندی کرده است. به این دسته نیز در اصطلاح منفی واقعی (TN) گفته می‌شود. در مقابل، تعداد ۶۵ نفر از مجموعه داده آزمایش در سازمان مانده و به فعالیت خود ادامه داده‌اند (منفی یا مقدار ۰) که مدل آن‌ها را به اشتباه پیش‌بینی و به اشتباه در گروه ترک شغل دسته‌بندی کرده است. به این دسته در اصطلاح مثبت کاذب (FP) گفته می‌شود. تعداد ۴۰۱ نفر نیز ترک شغل داشته‌اند (مثبت یا مقدار ۱) که مدل به اشتباه آن‌ها را پیش‌بینی و به اشتباه در گروه افرادی که در سازمان مانده‌اند دسته‌بندی کرده است. به این دسته در اصطلاح منفی کاذب (FN) گفته می‌شود.

آیا عملکرد مدل در پیش‌بینی مناسب بوده است؟ معیارهای مختلفی برای ارزیابی عملکرد مدل وجود دارد. نخستین معیار دقت^۱ است. شاخص دقت موارد پیش‌بینی درست از کل موارد یا مجموعه داده آزمایش را اندازه می‌گیرد. در واقع این معیار درصد طبقه‌بندی صحیح را از میان تمام مقادیر پیش‌بینی شده اندازه می‌گیرد. مقدار شاخص دقت طبق رابطه ۵ برای این الگوریتم برابر ۹۱/۱ درصد بدست می‌آید.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{2340 + 2456}{5262} = 0.911 \quad (5)$$

البته معیار دقت به تنهایی کافی نیست خصوصاً هنگامی که در مجموعه داده عدم توازن بین گروه اصلی و فرعی در متغیر هدف وجود داشته باشد یا هزینه پیش‌بینی‌های مثبت یا منفی کاذب متفاوت باشد باید از سایر معیارها نیز کمک گرفت. معیار دوم معیار صحت^۲ است. این معیار به دسته مثبت متمرکز است و پیش‌بینی‌های درست در این دسته را اندازه می‌گیرد. این شاخص طبق رابطه ۶ از تقسیم تعداد نمونه‌های تشخیصی درست مثبت (TP) به کل نمونه‌های مثبت

¹. Accuracy

². Precision

پیش‌بینی شده (TP+FP) بدست می‌آید. در واقع از میان افرادی که مدل آن‌ها را مثبت یا ترک شغل پیش‌بینی کرده، بیش از ۹۷ درصد این پیش‌بینی‌ها درست بوده است. بالاتر بودن معیار صحت از دقت نشان می‌دهد که مدل عملکرد پیش‌بینی بهتری برای دسته مثبت یا ترک شغل داشته است.

$$Precision = \frac{TP}{TP + FP} = \frac{2340}{2340 + 65} = 0.973 \quad (۶)$$

معیار سوم، معیار بازخوانی^۱ است. تمرکز اصلی این معیار بر روی داده‌هایی است که در واقع «مثبت» بوده‌اند. بازخوانی معیاری است که به ما می‌گوید از مواردی که واقعاً «مثبت» بوده‌اند، چند درصد به درستی مثبت تشخیص داده شده (TP) و چند درصد به اشتباه توسط مدل منفی پیش‌بینی (FN) شده‌اند. طبق رابطه ۷ عملکرد مدل به لحاظ معیار بازخوانی برابر ۸۵/۴ درصد بدست می‌آید. در واقع این نتیجه نشان می‌دهد که از مجموع افراد ترک شغل کرده، مدل در ۸۵/۴ موارد تشخیص درست داشته است.

$$Recall = \frac{TP}{TP + FN} = \frac{2340}{2340 + 401} = 0.854 \quad (۷)$$

دو معیار صحت و بازخوانی بیانگر خطای نوع اول و دوم بوده و بین آن‌ها یک بده‌بستان وجود دارد. بر این اساس، ساخت یک معیاری ترکیبی و تمرکز بر آن معیار به جای بررسی همزمان این دو معیار صحت و بازخوانی مناسب‌تر خواهد بود. امتیاز-F1 میانگین هارمونیک بین دو معیار صحت و پوشش است. معیار امتیاز-F1 که طبق رابطه ۸ نمایش داده شده، خود در بازه ۰ و ۱ قرار می‌گیرد. اگر یکی از دو معیار صحت یا بازخوانی عددی کوچک یا نزدیک به صفر باشد، نتیجه ترکیبی امتیاز-F1 عددی کوچک و نزدیک به صفر خواهد بود و اگر هر دو عددی بزرگ نزدیک ۱ باشند، نتیجه نهایی به سمت یک حرکت خواهد کرد.

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{1.662}{1.827} = 0.901 \quad (۸)$$

^۱. Recall

۴-۱-۲- یافته‌های الگوریتم درخت تصمیم

الگوریتم درخت تصمیم دو پارامتر اصلی دارد که باید از پیش مشخص شوند. این دو پارامتر اصلی عبارتند از حداکثر عمق درخت و معیار انشعاب. برای تنظیم این دو پارامتر اصلی به این شیوه عمل شد که با استفاده از داده‌های آموزشی ابتدا یکبار الگوریتم درخت تصمیم با استفاده از شاخص جینی و یکبار هم با استفاده از شاخص آنتروپی در عمق‌های مختلف اجرا شد. به این ترتیب، در هر عمق یک مدل با استفاده از داده‌های آموزش ساخته و سپس از داده‌های آزمون اعتبارسنجی برای سنجش دقت آن استفاده شد. در نهایت مدلی با بالاترین دقت بهینه‌ترین مقدار برای پارامترهای اصلی انتخاب می‌شوند. بر این اساس، درخت تصمیم با معیار آنتروپی به عنوان شاخص انشعاب و در حداکثر عمق ۴ به عنوان مقدار پارامترهای اصلی تعیین شد.

پس از پیاده‌سازی الگوریتم درخت تصمیم، تعداد ۵ شاخص برای تقسیم درخت در حالت بهینه بدست آمد که درخت آن در شکل ۵ نمایش داده شده است. عملکرد این مدل سپس با داده‌های آزمایش بررسی گردید که نتایج آن به صورت ماتریس اغتشاش در جدول ۳ ارائه شده است.

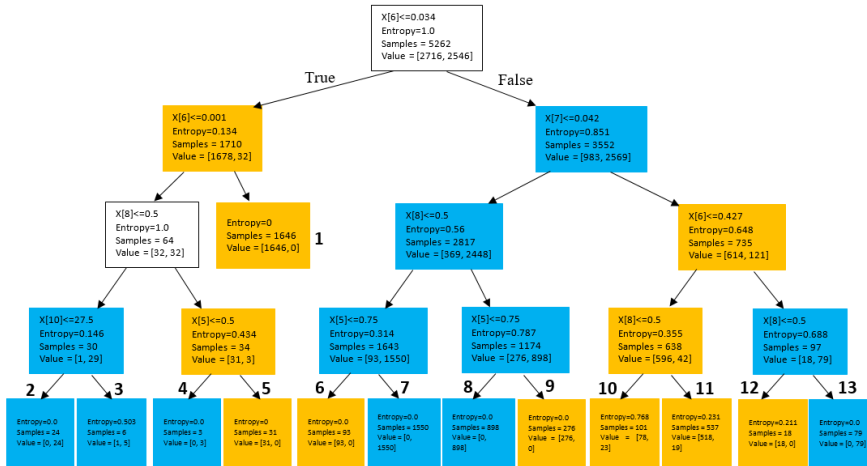
جدول ۳. نتایج ماتریس اغتشاش الگوریتم درخت تصمیم

		پیش‌بینی مدل	
		ترک شغل	ادامه فعالیت
موارد	ترک شغل	۲۲۰۴ (TP)	۵۳۷ (FN)
واقعی	ادامه فعالیت	۳۴۲ (FP)	۲۱۷۹ (TN)

منبع: یافته‌های پژوهش.

از مجموعه ۵۲۶۲ داده آزمایشی، تعداد ۲۲۰۴ نفر ترک شغل داشته (مثبت یا مقدار ۱) که مدل به درستی آن‌ها را پیش‌بینی و در گروه کارکنان ترک شغل قرار داده است که به آن‌ها در اصطلاح مثبت واقعی (TP) گفته می‌شود. تعداد ۲۱۷۹ نفر نیز در سازمان مانده و به فعالیت خود ادامه داده‌اند (منفی یا مقدار ۰) که مدل به درستی آن‌ها را پیش‌بینی و در گروه افرادی که در سازمان مانده‌اند دسته‌بندی کرده است. این دسته نیز همچنان که گفته شد در اصطلاح منفی واقعی (TN) نامیده می‌شود. در مقابل، تعداد ۳۴۲ نفر از مجموعه داده آزمایش در سازمان مانده و به فعالیت خود ادامه داده‌اند (منفی یا مقدار ۰) که مدل آن‌ها را به اشتباه در گروه ترک شغل دسته‌بندی کرده

است؛ در اصطلاح مثبت کاذب (FP). تعداد ۵۳۷ نفر نیز ترک شغل داشته‌اند (مثبت یا مقدار ۱) که مدل به اشتباه آن‌ها را پیش‌بینی و به اشتباه در گروه افرادی که در سازمان مانده‌اند دسته‌بندی کرده است. به این دسته در اصطلاح منفی کاذب (FN) گفته می‌شود.



شکل ۵. درخت تصمیم منتج شده

منبع: یافته‌های پژوهش.

آیا عملکرد مدل در پیش‌بینی مناسب بوده است؟ چهار معیار مختلفی برای ارزیابی عملکرد مدل معرفی شد. نخستین معیار دقت است که موارد پیش‌بینی درست از کل موارد یا مجموعه داده آزمایش را اندازه می‌گیرد. مقدار شاخص دقت برای این الگوریتم درخت تصمیم برابر

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{2204 + 2179}{5262} = 0.833$$

صحت است که به دسته پیش‌بینی شده مثبت متمرکز است و پیش‌بینی‌های درست در این دسته را اندازه می‌گیرد. این شاخص برای الگوریتم درخت تصمیم برابر

$$Precision = \frac{TP}{TP + FP} = \frac{2204}{2204 + 342} = 0.866$$

است. معیار سوم، معیار بازخوانی است که تمرکز آن بر روی داده‌هایی است که در واقع «مثبت» بوده‌اند. بازخوانی معیاری است که به ما می‌گوید از مواردی که واقعاً «مثبت» بوده‌اند، چند درصد به درستی مثبت تشخیص داده شده (TP) و چند درصد به اشتباه توسط مدل منفی پیش‌بینی (FN) شده‌اند. مقدار معیار بازخوانی برای الگوریتم درخت تصمیم برابر

$$Recall = \frac{TP}{TP + FN} = \frac{2204}{2204 + 537} = 0.804$$

یا ۸۵/۴ درصد بدست می‌آید. گفته شد که بین دو معیار

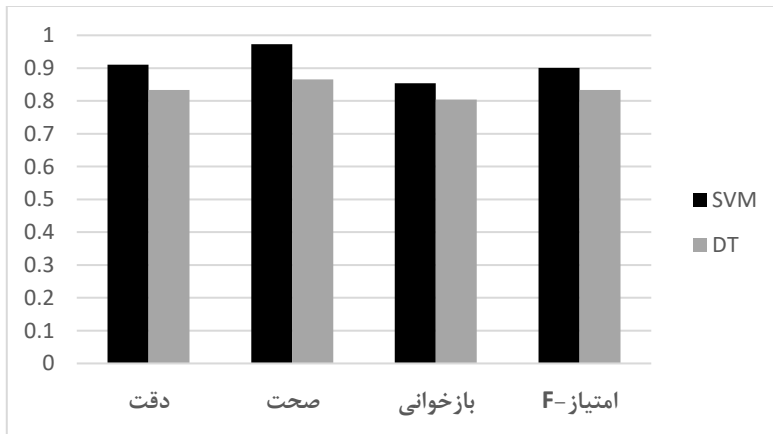
صحت و بازخوانی یک بده‌بستان وجود دارد و بهتر است به معیار امتیاز-F1 که میانگین هامونیک بین آن‌هاست توجه شود. معیار امتیاز-F1 برای الگوریتم درخت تصمیم نیز برابر

$$F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{1.392}{1.67} = 0.833$$

بدست می‌آید.

۴-۲- مقایسه عملکرد الگوریتم‌ها در دسته‌بندی مشتریان

شکل ۶ رتبه‌بندی عملکرد هردو مدل درخت تصمیم‌گیری (DT) و ماشین بردار پشتیبان (SVM) را نشان می‌دهد. می‌توان دید که به لحاظ هر چهار معیار دقت، صحت، بازخوانی و امتیاز-F1، مدل ماشین بردار پشتیبان عملکرد بهتری از درخت تصمیم‌گیری در پیش‌بینی موارد ترک کار داشته است.



شکل ۶. مقایسه عملکردی مدل‌های درخت تصمیم و ماشین بردار پشتیبان

منبع: یافته‌های پژوهش.

۴-۳- تفسیرپذیری نتایج

با وجود عملکرد بهتر ماشین بردار پشتیبان (SVM) در پیش‌بینی، متأسفانه این مدل یک جعبه سیاه است و به ما امکان تفسیر نتایج و بررسی شیوه دسته‌بندی را نمی‌دهد. در مقابل، درخت تصمیم به خواننده امکان این را می‌دهد که ببیند مدل از چه مشخصه‌ها و ویژگی‌هایی برای دسته‌بندی استفاده کرده است. این کار نیز به روش‌های مختلفی می‌تواند انجام شود که یکی از

آن‌ها روش اهمیت ویژگی^۱ است. در این روش به ویژگی‌های ورودی امتیازی نسبت داده می‌شود که نشان می‌دهد آن ویژگی تا چه اندازه در پیش‌بینی مشخصه هدف مفید بوده است. از آنجایی که در این مقاله از شاخص آنتروپی برای تقسیم داده‌ها استفاده شد، امتیاز اهمیت ویژگی در واقع مقداری است که با انتخاب آن ویژگی به عنوان دسته‌بندی مقدار آنتروپی کاهش می‌یابد. بر این اساس و طبق نتایج بدست آمده از الگوریتم درخت تصمیم، مشاهده شد که از میان ویژگی‌های ورودی به طور مشخص ۵ ویژگی تعداد جابه‌جایی، ساعت کار، سمت سازمانی، سنوات خدمت و سن بیشترین تأثیرگذاری را بر ترک شغل داشته‌اند.

علاوه بر این، می‌توان به درخت تصمیم بدست آمده در شکل ۵ توجه کرد که چگونه دسته‌بندی و پیش‌بینی بر اساس ویژگی‌ها انجام شده است. این شکل و نحوه دسته‌بندی مشاهدات توسط الگوریتم درخت تصمیم در جدول ۴ خلاصه شده است. مهمترین مشخصه در بالای درخت به عنوان گره ریشه نشان داده شده است. مقدار Value بیان می‌کند که مدل به چه تعداد ۲۷۱۶ مشاهده مقدار ۰ (نیروی کار فعال) و تعداد ۲۵۴۶ مشاهده مقدار ۱ (ترک کار) اختصاص داده است. در واقع به طور کل احتمال ترک شغل را در مجموعه داده آزمون حدود ۴۸ درصد برآورد کرده است. از گره ریشه طبق شکل ۵ بر اساس مشخصه $X[6]$ که تعداد جابه‌جایی است، نخستین انشعاب انجام می‌گیرد. طبق نتایج بدست آمده اگر تعداد جابه‌جایی صفر باشد، به سمت چپ شکل برگ‌های ۱ تا ۵ منتهی می‌شویم. تعداد ۱۷۱۰ مشاهده جابه‌جایی صفر داشته‌اند که مدل برای تنها ۲ درصد آن‌ها یا به عبارتی برای ۳۲ نفر ترک شغل را پیش‌بینی کرده است. به عبارت دیگر، مهمترین مشخصه پیش‌بینی‌کننده ترک شغل تعداد جابه‌جایی است که رابطه مستقیم با آن دارد. اگر تعداد جابه‌جایی صفر باشد می‌توان گفت که احتمال ترک شغل به شدت کم است. مدل سپس به دنبال توضیح با استفاده از سایر عوامل رفته و ترک شغل را با عامل سمت سازمانی $X[8]$ توضیح می‌دهد. احتمال ترک شغل در میان اعضای صف بسیار بالاتر از اعضای ستاد است. در واقع از این ۳۲ نفری که علی‌رغم جابه‌جایی صفر به ترک شغل مبادرت کرده‌اند، تعداد ۲۹ نفر از اعضاء صف و تنها ۳ نفر از اعضای ستاد بوده‌اند. در مرتبه بعدی این دو متغیر سن $X[10]$ و سنوات خدمت $X[5]$ هستند که تعیین‌کننده‌اند. بخش بزرگی از اعضای صف که ترک شغل

^۱. Feature importance

داشته‌اند، افراد جوان زیر ۲۸ سال (برگ ۲) بوده‌اند. تعداد ۳ نفر از اعضای ستادی که ترک شغل داشته‌اند نیز همگی سنوات خدمت کمتر از ۵ سال (برگ ۴) داشته‌اند. در واقع طبق پیش‌بینی مدل، در میان افراد با جابه‌جایی صفر احتمال ترک شغل بسیار کم است و اگر فرد عضو ستاد بوده و سنوات خدمت بیشتر از ۵ داشته باشد، این احتمال به صفر می‌رسد.

جدول ۴. خلاصه نتایج درخت تصمیم برای گروه ترک شغل

تعداد جابه‌جایی	ساعت کار	سمت سازمانی	سنوات	سن	برگ
۱ یا ۲	-	صف	-	کمتر از ۲۷	۲
			-	بیشتر از ۲۷	۳
		ستاد	کمتر از ۵	-	۴
			بیشتر از ۵	-	۵
۳ یا ۴	کمتر از ۱۷۶	صف	-	-	۱
			کمتر از ۱۰	-	۶
			بیشتر از ۱۰	-	۷
		ستاد	کمتر از ۱۰	-	۸
			بیشتر از ۱۰	-	۹
			-	-	۱۰
۵ و بیشتر	بیشتر از ۱۷۶	ستاد	-	-	۱۱
		ستاد	-	-	۱۲
		صف	-	-	۱۳

منبع: یافته‌های پژوهش.

برای آن‌هایی که جابه‌جایی بالا برابر ۳ یا ۴ داشته‌اند درخت تصمیم به سمت راست و برگ‌های ۶ تا ۱۱ منتهی می‌شود. تعداد ۳۵۵۲ مشاهده جابه‌جایی بالا داشته‌اند که مدل برای ۷۲ درصد آن‌ها یا به عبارتی برای ۲۵۶۹ نفر ترک شغل را پیش‌بینی کرده است. برای توضیح ترک شغل در این گروه با جابه‌جایی بالا، مدل از سایر عوامل استفاده کرده و به این نتیجه می‌رسد که ویژگی ساعت کار $X[7]$ تعیین کننده است. اگر متوسط ساعت کار کمتر از ۴۴ ساعت هفتگی باشد، احتمال ترک شغل به بیش از ۸۶ درصد می‌رسد. در واقع از میان ۲۵۶۹ نفری با جابه‌جایی شغلی بالا که ترک شغل کرده‌اند، تعداد ۲۴۴۸ نفر افراد با ساعت کاری پایین و تنها ۱۲۱ نفر افراد

با ساعت کاری بالا بوده‌اند. متغیر مهم دیگر سمت سازمانی است. اگر این افراد با جابه‌جایی بالا و ساعت کار پایین جزو اعضای صف باشند، احتمال ترک شغل به بیش از ۹۴ درصد می‌رسد (برگ ۶ و ۷). برای اعضای ستادی هم اگرچه وضع کمی بهتر است اما باز هم آن‌ها به یک سطح سنوات شغلی بالاتر برابر ۱۰ سال سابقه نیاز دارند تا در شرکت ماندگار شوند (برگ ۹) و گرنه و در صورت سنوات خدمت پایتینتر باز هم مدل برای آن‌ها ترک شغل را پیش‌بینی کرده است (برگ ۸). برای افرادی که بلحاظ ساعت کاری وضعیت مناسبی داشته‌اند، می‌توان دید که جابه‌جایی شغلی به تعداد ۵ و بیشتر است که شرایط را برای ماندگاری آن‌ها سخت‌تر کرده و احتمال ترک شغل را در آن‌ها مجدد افزایش می‌دهد (برگ ۱۲ و ۱۳).

این نتایج مجدد نشان می‌دهد که از میان عوامل تأثیرگذار بر ترک شغل بیش از همه تعداد جابه‌جایی شغلی، ساعت کار، سمت سازمانی (صف یا ستاد) و سنوات از سایر عوامل اهمیت بیشتری دارند. در مرتبه بعدی می‌توان دید که سن نیز اهمیت می‌یابد.

۵- تفسیرپذیری نتایج

در این مقاله تلاش شد تا با تجزیه و تحلیل داده‌های مرتبط با نیروی کار یک فروشگاه‌های زنجیره در سراسر کشور در بازه زمانی آبان ۱۳۹۸ الی ۱۴۰۱ به پیش‌بینی ترک شغل نیروها و بررسی عوامل فردی و سازمانی (شغلی) مؤثر بپردازد. برای این منظور تعداد ۱۷۵۴۲ رکورد اطلاعاتی منحصر به فرد به کار گرفته شد که برای هر مورد وضعیت فعالیت فرد (ادامه فعالیت یا ترک شغل) و ۱۲ مشخصه فردی و شغلی از جمله جنسیت، وضعیت تأهل و تعداد فرزندان، سطح تحصیلات، سن، نوع قرارداد استخدامی، سمت سازمانی، سابقه سنوات، تعداد جابه‌جایی، ساعت کار (موظف و اضافه کار) و استان محل خدمت ثبت شده است. روش تحقیق مبتنی بر داده‌کاوی و یادگیری ماشین بود و از دو الگوریتم تحت نظارت ماشین بردار پشتیبان (SVM) و درخت تصمیم استفاده شد.

نتایج بدست آمده نشان دادند که از بین ویژگی‌های فردی و سازمانی (شغلی) عامل تعداد جابه‌جایی به طور چشمگیری مورد اهمیت است. این مشخصه در فروشگاه‌های زنجیره‌ای در میان گروه پرسنل صف و عملیاتی که درصد بالایی از مجموع کل کارمندان را تشکیل می‌دهند،

توضیح‌دهندگی بالایی دارد. معنی‌دار بودن این عامل قابل انتظار است چرا که می‌توان تعداد جابه‌جایی را نشانگری از عدم رضایت شغلی و عدم انطباق شغل با مهارت یا مشخصه‌های فردی دانست. به طور مشخص، درخواست‌های جابه‌جایی و انتقال در گروه پرسنل صف عمدتاً به علت جنس کار تکراری و مسأله مسافت بسیار رایج است و در صورتی که اتفاق نیفتد موجب ایجاد نارضایتی و کاهش بازده و عدم تمایل به کار خواهد شد. عامل مهم دیگر سمت سازمانی (صف یا ستاد) بود که طبیعتاً بر ترک شغل افراد تأثیرگذار است. از دیگر عوامل تأثیرگذار دیگر می‌توان به سنوات خدمت، تعداد ساعت کار (موظف و اضافه‌کار) و سن اشاره کرد. مشاهده شد که ترک کار در میان پرسنل جوان با سابقه کار کم و پاره وقت (تعداد ساعت کاری کم) بیشتر است.

مدل‌های یادگیری ماشین بر اساس این مشخصه‌ها توانستند عملکرد مناسبی از خود در دسته‌بندی و پیش‌بینی موارد ترک شغل داشته باشند. ماشین بردار پشتیبان (SVM) از دقت ۹۱ درصد و امتیاز-F1 بالای ۹۰ درصد برخوردار شد که نشان از عملکردی دقیق دارد. الگوریتم درخت تصمیم نیز با دقت ۸۳ درصد و امتیاز-F1 به همین اندازه نیز عملکرد مناسبی داشت. بر این اساس می‌توان گفت که با توجه به مشخصات و ویژگی‌های تأثیرگذار شناسایی شده و خروجی مدل‌های پیش‌بینی، این فروشگاه زنجیره‌ای باید سیاست‌های نگه‌داشت منابع انسانی و به‌عنوان سیاست‌های جذب را طوری تنظیم نماید تا موجب کاهش هزینه‌های سربار با ترک شغل بالاتر از حالت معمول شود. به‌عنوان مثال می‌توان برنامه‌های مدیریت منابع انسانی خاصی جهت نگه‌داشت گروه پرسنل جوان با سنوات خدمت کم در نظر گرفت که اتفاقاً از نظر بازدهی و انگیزه کار بسیار پویا هستند. حفظ پرسنل و کاهش کار نه فقط به لحاظ هزینه‌های سربار اهمیت دارد بلکه در صنعت پر رقابت خرده‌فروشی که در حال حاضر رو به پیشرفت و توسعه است، فروشگاه‌ها می‌توانند با حفظ پرسنل از منابع انسانی خود به عنوان مزیت رقابتی در ایجاد حس تعامل با مشتریان و همچنین اهرم اصلی گردش کار و پول حفاظت و بهره‌برداری کنند. یافته‌های این پژوهش همچنین می‌تواند در بحث بودجه‌ریزی و پیش‌بینی‌های مالی مجموعه نیز در نظر گرفته شود. البته شایان ذکر است که این نتایج با محدودیت‌هایی در خصوص داده‌ها و عدم مطالعه تأثیر عوامل کلان اقتصادی و اجتماعی بدست آمده است که امید است و اکیداً توصیه می‌شود در مطالعات آتی با جمع‌آوری یک مجموعه داده جامع‌تر به این عوامل توجه شود. همچنین برای

مطالعات آتی توصیه می‌گردد که پژوهشگران از الگوریتم‌های مختلف و متنوع دیگر که نتایج را قوام بخشیده و امکان مقایسه روشی را فراهم می‌کنند نیز بهره بگیرند.

تعارض منافع

نویسندگان اعلام کردند که هیچگونه تضاد منافع برای این پژوهش وجود ندارد.

مشارکت نویسندگان

نویسندگان در مفهوم‌سازی و نگارش مقاله مشارکت داشتند. همه نویسندگان محتوای مقاله را تأیید کردند و در مورد تمام جنبه‌های کار توافق داشتند.

منابع

امیری، قاسم، و محمودزاده، سیدمجتبی. (۱۳۹۴). بررسی عوامل مؤثر بر کاهش ترک خدمت کارکنان در سازمان‌های دولتی ایران (مطالعه موردی: ستاد وزارت راه و شهرسازی). مدیریت فرهنگ سازمانی، ۱۳(۲)، ۵۵۹-۵۷۹.

پیرایش، رضا، محمدی، مهدی خان، و بادفر، عمران. (۱۳۹۹). عوامل مؤثر بر قصد ترک خدمت کارکنان و تأثیر آن بر کارایی کارکنان شرکت زرین روی زنجان. مطالعات کاربردی در علوم مدیریت و توسعه، ۵(۱)، ۷-۱۹.

چشم براه، الهام. (۱۳۹۷). بررسی عوامل مؤثر بر ترک خدمت در بین کارکنان (پایان‌نامه کارشناسی ارشد). دانشگاه آزاد اسلامی واحد بندرعباس، بندرعباس.

دباشی، فرزانه، نوری، ابوالقاسم، عریضی، حمیدرضا، و دیباجی، سیدمیثم. (۱۳۹۵). پیش‌بینی تمایل به ترک شغل کارکنان بر اساس عوامل فردی، شغلی و سازمانی. دانش و پژوهش در روان‌شناسی کاربردی، ۱۷(۲)، ۴۵-۵۴.

دیهیم پور، مهدی، و دولتی، حسن. (۱۳۹۶). تأثیر عوامل ایجادکننده ترومای سازمانی بر میزان ترک خدمت کارکنان نظامی. پژوهش‌های مدیریت منابع انسانی، ۹(۴)، ۸۱-۱۰۶.

زبانی شادباد، محمدعلی، حسنی، محمد، قاسم‌زاده علیشاهی، ابوالفضل، و قاسم‌زاده، ابوالفضل. (۱۳۹۶). رابطه درگیری شغلی و تناسب شغلی با اخلاق حرفه‌ای و تمایل به ترک خدمت. / *اخلاق در علوم و فناوری*، ۱۲(۲)، ۷۷-۸۴.

طالقانی، غلامرضا، عبدالملکی، جمال، و غفاری، علی. (۱۳۹۷). بررسی عوامل فردی مؤثر بر قصد ترک شغل کارکنان اداره کل آموزش و پرورش استان کردستان. *مدیریت دولتی*، ۱(۱)، ۲۱۹-۲۳۲.

عابدین، بهاره، اکبری، شهناز، و عباس‌نژاد، طیبه. (۱۳۹۹). طراحی مدل مفهومی ترک خدمت افراد مستعد از سازمان‌های ایران: پژوهشی اکتشافی در صنعت ICT. *مدیریت منابع انسانی پایدار*، ۲(۲)، ۴۹-۶۴.

عارفی، صدیقه، سرخوش، سمانه، و رئوفی، سمیرا. (۱۳۹۸). تبیین عوامل مؤثر بر ترک شغل از دیدگاه پرستاران: یک مطالعه کیفی. *پژوهش‌های سلامت محور*، ۵(۱)، ۱۳-۲۸.

علیرحیمی، محمدمهدی، امیرخانی، امیرحسین، و رسولی، رضا. (۱۳۹۶). طراحی مدل کاهش ترک خدمت سازمانی (مورد مطالعه: شرکت پایانه‌های نفتی ایران). *مدیریت منابع انسانی در صنعت نفت*، ۹(۳۴)، ۵۳-۸۲.

هاشم زهی، ریحانه، نجفیگی، رضا، و ذبیحی، محمدرضا. (۱۴۰۰). طراحی مدل ترک خدمت کارکنان دانشی در شرکت پخش فراورده‌های نفتی خراسان رضوی. *مدیریت منابع انسانی در صنعت نفت*، ۱۲(۴۷)، ۳۹-۲۲.

References

Abedin, B., Akbari Emami, S., & Abbasnejad, T. (2020). Designing a Conceptual Model of Talent Turnover among Iranian Organizations: An Exploratory Research in ICT industry. *Journal of Sustainable Human Resource Management*, 2(2), 49-64. [In Persian]

Al Akasheh, M., Faisal Malik, E., Hujran, O., & Zaki, N. (2024). A Decade of Research on Machine Learning Techniques for Predicting Employee Turnover: A Systematic Literature Review. *Expert Systems with Applications*, 238, Retrieved from <https://doi.org/10.1016/j.eswa.2023.121794>

Alirahimi M. M., Amirkhani A., & Rasouli R. (2018). Designing a Model for Turnover Reduction in Iranian Oil Terminals Company (IOTC). *Strategic Studies in the Oil and Energy Industry*, 9(34), 53-82. [In Persian]

Alsagheer, R. H., Alharan, A. F., & Al-Haboobi, A. S. (2017). Popular Decision Tree Algorithms of Data Mining Techniques: A Review. *International Journal of Computer Science and Mobile Computing*, 6(6), 133-142.

Amiri, G. and Mahmoudzadeh, S. M. (2015). The Examination Factors Affecting Reduce of Employees Turnover in the Iranian Public Organizations. Case Study: Ministry of Road and Urban Development (Center Staff). *Organizational Culture Management*, 13(2), 559-579. [In Persian]

Arefi S., Sarkhosh S., & Raoofi S. (2019). Explanation of the Effective Factors of Job Leave from Nurses' Point of View: A Qualitative Study. *Health Based Research*, 5(1), 13-28. [In Persian]

Arokiasamy, A. R. A. (2013). A Qualitative Study on Causes and Effects of Employee Turnover in the Private Sector in Malaysia. *Middle-East Journal of Scientific Research*, 16(11), 532-1541.

Bassok, D., Markowitz, A. J., Bellows, L., & Sadowski, K. (2021). New Evidence on Teacher Turnover in Early Childhood. *Educational Evaluation and Policy Analysis*, 43(1), 172-180.

Brownlee, J. (2020) Data Preparation for Machine Learning-Data Cleaning, Feature Selection and Data Transformation in Python: Machine Learning Mastery. Retrived from <https://www.machinelearningmastery.com/data-preparation-for-machine-learning/>

Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem Formulations and Solvers in Linear SVM: A Review. *Artificial Intelligence Review*, 52(2), 803-855.

Cheshmbarah, E. (2018). *Study of Factors Affecting Employee Turnover* (Master's Thesis). Islamic Azad University, Bandar Abbas. [In Persian]

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2022). Embedding Transparency in Artificial Intelligence Machine Learning Models: Managerial Implications on Predicting and Explaining Employee Turnover. *The International Journal of Human Resource Management*, 34(14), 2732-2764.

Dabbashi, F., Nouri, A., Oreyzi, H., & Dibaji, S. M. (2015). Predicting Employees' Turnover Intention by Individual, Occupational and Organizational Factors. *Knowledge & Research in Applied Psychology*, 17(2), 45-54. [In Persian]

Dulaty, H., & deyhimpur, M. (2018). An Assessment of Factors Affecting Organizational Trauma on Leaving Service by Military Personnel. *Journal of Research in Human Resources Management*, 9(4), 81-106. [In Persian]

Dunegan, K. (1993). Framing, Cognitive Modes, and Image Theory: Toward an Understanding of a Glass Half Full, *Journal of Applied Psychology*, 78(3), 491-503.

Elaine, M. (1997). Job Tenure Shift for Men and Women. *HR Magazine*, 42(50), 1-20.

Hashemzahi R., Najafbeygi R., & Zabihi M. (2021). Designing the Model of Knowledge Workers Turnover in Khorasan Razavi Oil Products Distribution Company. *Strategic Studies in the Oil and Energy Industry*, 12(47), 22-39. [In Persian]

Healy, M. C., Lehman, M., & McDaniel, M. A. (1995). Age and Voluntary Turnover: A Quantitative Review. *Personnel Psychology*, 48(2), 335-345.

Marsh, R. M., & Mannari, H. (1977). Organizational Commitment and Turnover: A Prediction Study. *Administrative Science Quarterly*, 22(1), 57-75.

McDermid, F., Mannix, J., & Peters, K. (2020). Factors Contributing to High Turnover Rates of Emergency Nurses: A Review of the Literature. *Australian Critical Care*, 33(4), 390-396.

Park, J., Feng, Y. & Jeong, SP. Developing an Advanced Prediction Model For New Employee Turnover Intention Utilizing Machine Learning Techniques. *Scientific Reports*, 14, Retrieved from <https://doi.org/10.1038/s41598-023-50593-4>

Phillips, D. R., & Roper, K. O. (2009). A Framework for Talent Management in Real Estate. *Journal of Corporate Real Estate*, 11(1), 7-16.

Pirayesh, R., Mohammadi, M., & Badfar, I. (2019). Factors Affecting Employees' Intention to Leave and Its Impact on Employee Performance in Zarrin Roy Zanzan Company. *Applied Studies in Management and Development Sciences*, 5(1), 7-19. [In Persian]

Rokach, L., & Maimon, O. Z. (2014). *Data Mining With Decision Trees: Theory and Applications*. Singapore: World Scientific Publishing Co.

Summers, T. P., & Hendrix, W. H. (1991). Modelling the Role of Pay Equity Perceptions: A Field Study. *Journal of Occupational Psychology*, 64, 145-157

Taleghani, G., Abdolmaleki, J. and Ghafari, A. (2016). The Investigation of the Individual Factors on Turnover Intention of Employees in Education Administration of Kurdistan Province. *Journal of Public Administration*, 8(1), 219-232. [In Persian]

Waldman, J. D., Kelly, F., Aurora, S., & Smith, H. (2004). The Shocking Cost of Turnover in Health Care, *Health Care Management Review*, 29(1), 2-7.

Yun, M. R., & Yu, B. (2021). Strategies for Reducing Hospital Nurse Turnover in South Korea: Nurses' Perceptions and Suggestions. *Journal of Nursing Management*, 29(5), 1256-1262.

Zabani Shadabad, M. A., Hassani, M., & Ghasemzadeh, A. (2017). The Relationship between Job Engagement & Job Propriety with Professional Ethics & Intent to Leave. *Ethics in Science and Technology*, 12 (2). 77-84. [In Persian]

Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2018). Employee Turnover Prediction with Machine Learning: A Reliable Approach. In *Proceedings of SAI Intelligent Systems Conference*. Cham: Springer.