

الحمد لله الذي
خلقنا من
الحمم



دانشکده مهندسی صنایع و مدیریت

پایان نامه کارشناسی ارشد مدیریت صنعتی

بررسی و پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی

(مطالعه موردی: شرکت‌های ارائه‌دهنده خدمات توزیع برق استان خوزستان)

نگارنده: فاطمه نظری

استاد راهنما

دکتر سید محمدحسن حسینی

مهرماه ۱۳۹۹

تقدیم به پدر و مادر عزیز و مهربانم

که هر لحظه وجودم را از چشمه سار پر از عشق چشمانشان سیراب می کنند....

تشکر و قدردانی

پاس خدای بزرگ را که مرا یاری رساند تا بتوانم این مقطع تحصیلی را به پایان رسانده و گامی در

راستای اعتلای علم بردارم.

از استاد راهنمای گران قدرم جناب آقای دکتر حسینی که وجودشان همیشه قوتی برای انجام کارهایم

بوده است و بدون شک انجام این پایان نامه بدون کمک و راهنمایی های ارزنده ایشان امکان پذیر

نبوده است، کمال تشکر را دارم.

اینجانب **فاطمه نظری** دانشجوی دوره کارشناسی ارشد رشته مدیریت صنعتی دانشکده مهندسی صنایع و مدیریت دانشگاه صنعتی شاهرود نویسنده پایان نامه «بررسی پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی (نمونه موردی: شرکت‌های ارائه‌دهنده خدمات توزیع برق استان خوزستان)» تحت راهنمایی جناب آقای دکتر حسینی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج بانام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.

چکیده

امروزه اندیشمندان سازمان مدیریت بیان می‌کنند که مهم‌ترین و ارزشمندترین سرمایه سازمان سرمایه نیروی انسانی آن است. پس تلاش برای توسعه تعهد وفاداری در کارکنان و ایجاد حس تعلق در آن‌ها امر مهمی است و با انتخاب مناسب کارکنان گام بزرگی در جهت پیشرفت سازمان خود برمی‌دارند. با توجه به پیشرفت فناوری در دهه‌های اخیر، حجم اطلاعات ذخیره‌شده درون سازمان‌ها روزبه‌روز در حال افزایش است. در نتیجه می‌توان با استفاده از این اطلاعات وسیع، به کشف الگوها قواعد میان داده‌های مختلف از جمله داده‌های منابع انسانی پرداخت که این امر با استفاده از علم داده‌کاوی امکان‌پذیر است.

هدف پژوهش حاضر بررسی پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی است. این پژوهش از نظر ماهیت اکتشافی و از نظر هدف کاربردی است. امروزه روش‌های داده‌کاوی به طور گسترده در کشف الگوها و قواعد بین داده‌ها استفاده می‌شوند. کاربرد این روش‌ها می‌تواند نقش موثری در پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی در نمونه موردی شرکت‌های ارائه‌دهنده خدمات توزیع برق استان خوزستان باشد. در این پژوهش، داده‌های بدست آمده توسط نرم افزار Rapid Miner مورد تجزیه و تحلیل قرار گرفت و عملکرد سه الگوریتم از روش‌های داده‌کاوی شامل الگوریتم درخت تصمیم، ماشین بردار پشتیبان و k نزدیک‌ترین همسایه مورد بررسی قرار گرفته است. همچنین در این پژوهش از ترکیب دو الگوریتم برتر بعنوان نوآوری پژوهشی و کسب نتیجه مطلوب‌تر استفاده شده است.

کلمات کلیدی: وفاداری کارکنان، تکنیک داده‌کاوی، درخت تصمیم، ماشین بردار پشتیبان، نزدیک‌ترین همسایگی

فهرست مطالب

صفحه	عنوان
۱	فصل اول - کلیات پژوهش
۱-۱	مقدمه
۲-۱	بیان مسئله
۳-۱	ضرورت و توجیه انجام پژوهش
۴-۱	هدف پژوهش
۱-۴-۱	هدف اصلی پژوهش
۲-۴-۱	اهداف فرعی پژوهش
۵-۱	سؤالات و فرضیات پژوهش
۱-۵-۱	سؤال فرعی تحقیق
۶-۱	روش پژوهش
۷-۱	قلمرو پژوهش
۸-۱	نوآوری پژوهش
۹-۱	توصیف اصطلاحات و واژگان کلیدی
۱۰-۱	ساختار پایان نامه
۱۱	فصل دوم - مبانی نظری و پیشینه پژوهش
۱-۲	مقدمه
۲-۲	مبانی نظری
۱-۲-۲	تعریف وفاداری
۲-۲-۲	تحقق وفاداری
۳-۲-۲	نتایج و پیامدهای وابستگی و تعهد نسبت به سازمان
۴-۲-۲	وفاداری کارکنان به هدف سازمان
۵-۲-۲	وفاداری به خود سازمان
۶-۲-۲	مفهوم وفاداری کارکنان
۷-۲-۲	وابستگی سازمانی و نقش آن در ایجاد تعهد و وفاداری به سازمان
۸-۲-۲	راهکارهای وفاداری کارکنان به سازمان
۳-۲	پیشینه تحقیق
۱-۳-۲	مطالعات داخلی
۲-۳-۲	مطالعات خارجی
۳-۳-۲	خلاصه پیشینه تحقیقات داخلی در حوزه وفاداری کارکنان
۴-۳-۲	خلاصه پیشینه تحقیقات خارجی در حوزه وفاداری کارکنان
۴-۲	جمع بندی

فصل سوم - روش‌شناسی پژوهش	۴۰
۱-۳) مقدمه	۴۱
۲-۳) روش تحقیق	۴۱
۳-۳) داده‌کاوی	۴۲
۱-۳-۳) تکنیک‌های مختلف داده‌کاوی	۴۳
۲-۳-۳) طبقه‌بندی در داده‌کاوی	۴۴
۳-۳-۳) مراحل داده‌کاوی	۴۴
۴-۳) روش‌های طبقه‌بندی داده‌کاوی	۴۶
۱-۴-۳) درخت تصمیم	۴۶
۱-۱-۴-۳) جذابیت درختان تصمیم	۴۹
۲-۱-۴-۳) چند الگوریتم درخت تصمیم	۴۹
۳-۱-۴-۳) بازنمایی درخت تصمیم	۵۰
۴-۱-۴-۳) معیارهای انتخاب صفت خاصه	۵۱
۵-۱-۴-۳) طراحی یک دسته‌بندی کننده درخت تصمیم	۵۳
۶-۱-۴-۳) روش‌های اصلی برای طراحی دسته‌بندی کننده درخت تصمیم	۵۴
۷-۱-۴-۳) انواع یادگیری در درخت تصمیم‌گیری	۵۵
۲-۴-۳) دسته‌بند با ماشین بردار پشتیبان	۵۶
۳-۴-۳) طبقه‌بند KNN	۵۹
۱-۳-۴-۳) مراحل آموزش الگوریتم	۶۱
۲-۳-۴-۳) ساختار الگوریتم	۶۱
۵-۳) جامعه آماری	۶۲
۶-۳) نمونه آماری	۶۲
۷-۳) ابزار گردآوری اطلاعات	۶۳
۸-۳) متغیرهای مورد بررسی	۶۳
۹-۳) ابزار و روش تجزیه و تحلیل داده‌ها	۶۴
۱۰-۳) مقایسه الگوریتم‌ها	۶۵
۱۱-۳) الگوریتم پیشنهادی	۶۷
۱-۱۱-۳) درخت تصمیم	۶۷
۲-۱۱-۳) الگوریتم SVM	۷۰
۳-۱۱-۳) الگوریتم KNN	۷۰
۱۲-۳) جمع‌بندی	۷۱
فصل چهارم - تحلیل و بررسی نتایج	۷۲
۱-۴) مقدمه	۷۳
۲-۴) آمار توصیفی	۷۳
۳-۴) مدل‌سازی	۸۳
۱-۳-۴) پیش‌پردازش داده‌ها	۸۳
۲-۳-۴) درخت تصمیم ID3	۸۵

۹۰.....	درخت تصمیم C4.5 (۳-۳-۴)
۹۸.....	درخت تصمیم CART (۴-۳-۴)
۱۰۳.....	ماشین بردار پشتیبان (۵-۳-۴)
۱۰۷.....	الگوریتم KNN (۶-۳-۴)
۱۱۰.....	الگوریتم ترکیبی (۷-۳-۴)
۱۱۳.....	مقایسه نتایج مدل‌ها (۴-۴)
۱۱۵.....	نتیجه‌گیری (۵-۴)
۱۱۶.....	فصل پنجم - نتیجه‌گیری و پیشنهادات
۱۱۷.....	مقدمه (۱-۵)
۱۱۷.....	مقایسه نتایج با نتایج پژوهش‌های پیشین (۱-۱-۵)
۱۱۸.....	نتیجه‌گیری (۳-۵)
۱۲۰.....	محدودیت‌های پژوهش (۴-۵)
۱۲۰.....	محدودیت‌های پژوهشگر (۱-۴-۵)
۱۲۱.....	پیشنهادات (۵-۵)
۱۲۱.....	پیشنهادات مدیریتی (۱-۵-۵)
۱۲۱.....	پیشنهادات تحقیقات آتی (۲-۵-۵)
۱۲۲.....	منابع

فهرست شکل‌ها

صفحه	عنوان
۴۳	شکل (۱-۳) تکنیک‌های داده‌کاوی
۴۵	شکل (۲-۳) مراحل داده‌کاوی
۵۱	شکل (۳-۳) نمونه درخت تصمیم
۵۷	شکل (۴-۳) دسته‌بند ماشین بردار پشتیبان
۵۸	شکل (۵-۳) SVM با حاشیه نرم
۶۴	شکل (۶-۳) صفحه اصلی نرم‌افزار RAPIDMINER
۸۴	شکل (۱-۴) نرمال سازی پایگاه داده
۸۵	شکل (۲-۴) مراحل ساخت درخت تصمیم ID3
۹۱	شکل (۳-۴) مراحل ساخت درخت تصمیم C4.5
۹۸	شکل (۴-۴) مراحل ساخت درخت تصمیم CART
۱۰۳	شکل (۵-۴) مراحل ساخت ماشین بردار پشتیبان
۱۰۸	شکل (۶-۴) مراحل ساخت KNN

فهرست جدول‌ها

صفحه	عنوان
۳۴.....	جدول (۱-۲) خلاصه پیشینه تحقیقات داخلی در حوزه وفاداری کارکنان
۳۷.....	جدول (۲-۲) خلاصه پیشینه تحقیقات خارجی در حوزه وفاداری کارکنان
۶۳.....	جدول (۱-۳) متغیرهای مورد بررسی
۶۶.....	جدول (۲-۳) مقایسه الگوریتم‌ها
۶۸.....	جدول (۳-۳) مقایسه درخت‌های تصمیم
۶۸.....	جدول (۴-۳) مشخصات الگوریتم درخت تصمیم
۷۰.....	جدول (۵-۳) پارامترهای مدل
۷۴.....	جدول (۱-۴) کدگذاری رشته‌های تحصیلی
۷۵.....	جدول (۲-۴) کدگذاری مقاطع تحصیلی
۷۶.....	جدول (۳-۴) کدگذاری سن استخدام کارکنان
۷۷.....	جدول (۴-۴) کدگذاری محل اشتغال
۷۸.....	جدول (۵-۴) کدگذاری سمت کارکنان
۷۹.....	جدول (۶-۴) کدگذاری نوع استخدام
۸۰.....	جدول (۷-۴) کدگذاری میزان حقوق
۸۱.....	جدول (۸-۴) کدگذاری جنسیت
۸۱.....	جدول (۹-۴) کدگذاری وضعیت تاهل
۸۶.....	جدول (۱۰-۴) اهمیت متغیرها
۸۹.....	جدول (۱۱-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم ID3
۹۰.....	جدول (۱۲-۴) ماتریس اغتشاش درخت تصمیم ID3
۹۲.....	جدول (۱۳-۴) ضریب اهمیت متغیرها
۹۷.....	جدول (۱۴-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم C4.5
۹۷.....	جدول (۱۵-۴) ماتریس اغتشاش درخت تصمیم C4.5
۱۰۰.....	جدول (۱۶-۴) ضریب اهمیت متغیرها
۱۰۲.....	جدول (۱۷-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم CART
۱۰۲.....	جدول (۱۸-۴) ماتریس اغتشاش درخت تصمیم CART
۱۰۵.....	جدول (۱۹-۴) ضریب اهمیت متغیرها
۱۰۶.....	جدول (۲۰-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم
۱۰۶.....	جدول (۲۱-۴) ماتریس اغتشاش SVM
۱۰۹.....	جدول (۲۲-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم KNN
۱۰۹.....	جدول (۲۳-۴) ماتریس اغتشاش الگوریتم KNN
۱۱۱.....	جدول (۲۴-۴) ضریب اهمیت متغیرها
۱۱۲.....	جدول (۲۵-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم ترکیبی

۱۱۲	جدول (۴-۲۶) ماتریس اغتشاش الگوریتم ترکیبی
۱۱۳	جدول (۴-۲۷) مقایسه نتایج الگوریتم‌ها
۱۱۷	جدول (۵-۱) مقایسه نتایج پژوهش‌های پیشین
۱۱۸	جدول (۵-۲) مقایسه نتایج ضرایب الگوریتم‌ها

۷۴.....	نمودار (۱-۴) پراکندگی رشته‌های تحصیلی
۷۵.....	نمودار (۲-۴) پراکندگی مقاطع تحصیلی
۷۶.....	نمودار (۳-۴) پراکندگی سن کارکنان
۷۷.....	نمودار (۴-۴) پراکندگی محل اشتغال
۷۸.....	نمودار (۵-۴) پراکندگی سمت کارکنان
۷۹.....	نمودار (۶-۴) پراکندگی نوع استخدام
۸۰.....	نمودار (۷-۴) پراکندگی نوع حقوق
۸۶.....	نمودار (۸-۴) نمودار اهمیت متغیرها در ساخت درخت تصمیم ID3
۸۸.....	نمودار (۹-۴) شماتیک درخت تصمیم ID3
۹۱.....	نمودار (۱۰-۴) نمودار اهمیت متغیرها در ساخت الگوریتم درخت تصمیم C4.5
۹۶.....	نمودار (۱۱-۴) شماتیک درخت تصمیم C4.5
۹۹.....	نمودار (۱۲-۴) نمودار اهمیت متغیرها در ساخت الگوریتم درخت تصمیم CART
۱۰۱.....	نمودار (۱۳-۴) شماتیک درخت تصمیم CART
۱۰۴.....	نمودار (۱۴-۴) نمودار اهمیت متغیرها در ساخت ماشین بردار پشتیبان
۱۰۵.....	نمودار (۱۵-۴) شماتیک ماشین بردار پشتیبان
۱۰۸.....	نمودار (۱۶-۴) شماتیک KNN
۱۱۰.....	نمودار (۱۷-۴) نمودار اهمیت متغیرها در ساخت الگوریتم ترکیبی
۱۱۳.....	نمودار (۱۸-۴) مقایسه نتایج الگوریتم‌ها معیارهای مورد بررسی
۱۱۴.....	نمودار (۱۹-۴) مقایسه نتایج الگوریتم‌ها در دقت آموزش و تست
۱۱۹.....	نمودار (۱-۵) مقایسه نتایج ضرایب الگوریتم‌ها

فصل اول:

کلیات پژوهش

رقابت تجاری و اقتصادی، دسترسی سازمان‌ها به قابلیت‌های استراتژیک و سرمایه‌های کلیدی را برای سبقت از رقبا، به یک ضرورت تبدیل کرده است. به گفته آدلر و آدلر^۱ در سال ۱۹۸۸، وفاداری سازمانی^۲ یک حس تعلق و اشتیاق به ارتباط با یک سازمان است و بر احساس وابستگی، اشتیاق شدید به عضویت در یک گروه، آمادگی برای تشریک مساعی، حس اعتماد، هم‌سویی داوطلبانه با یک گروه و تمایل به پیروی از رهنمودهای سازمانی دلالت دارد و می‌تواند تأثیر شایان توجهی در موفقیت سازمان‌ها در صحنه‌های رقابت داشته باشد، به‌طوری‌که پیتر دراگر، داشتن کارکنان وفادار را مزیتی رقابتی برای سازمان‌ها می‌داند و پژوهش‌های بسیاری مویده نقش مثبت وفاداری کارکنان در افزایش بهره‌وری و موفقیت سازمان‌ها هستند. پژوهش کوئل - شپیرو، کسلر و پرسل^۳ در سال ۲۰۰۳، مشخص کرد که ارتباط مثبتی بین وفاداری کارکنان و رفتار شهروندی سازمان وجود دارد. کنوسکی و پابف^۴ نیز در سال ۱۹۹۴ در پژوهش خود اعلام کردند، کارکنان وفادار، انگیزه بیشتری برای حضور در سازمان داشته و برای رسیدن به هدف‌های سازمان تلاش بیشتری می‌کنند. با توجه به نقش مؤثری که کارکنان وفادار می‌توانند در تحقق اهداف سازمان ایفا کنند، مطالعات بسیاری در راستای شناسایی مؤلفه‌های تأثیرگذار بر وفاداری کارکنان انجام شده است و عوامل متنوعی در این خصوص شناسایی شده‌اند و اجرای عدالت در آن به‌منزله تحقق عدالت در سطح جامعه است. بی‌عدالتی و توزیع ناعادلانه دستاوردها موجب کاهش روحیه تلاش در پرسنل می‌شود، پس رعایت عدالت، ضامن پیشرفت سازمان است. با بیان الوانی و پور عزت در سال ۱۳۸۲، با توجه به اهمیت تأثیرگذار عدالت سازمانی و وفاداری کارکنان، سازمان‌ها جهت ایجاد انگیزه اثربخش در کارکنان خود، گام‌های زیادی برمی‌دارند. تقویت احساس مفید بودن، چگونگی روابط رهبر-پیرو، همبستگی سازمانی، هویت‌بخشی و هم‌سویی اهداف فردی و سازمانی از دیگر عوامل تقویت‌کننده حس وفاداری در سازمان‌ها هستند.

^۱. Adler & Adler

^۲. Organizational Loyalty

^۳. Coyle - Shapiro - Kessler & Purcell

^۴. Konovsky & Pugh

۱- ۲) بیان مسئله

امروزه منابع انسانی به عنوان ارزشمندترین سرمایه‌های سازمانی محسوب می‌شوند که می‌توانند تأثیر به‌سزایی در موفقیت و تحقق اهداف سازمان‌ها داشته باشند؛ لذا حفظ و توسعه این منابع ارزشمند یکی از اهداف تمام سازمان‌ها به شمار می‌آید. از سوی دیگر افزایش حقوق، دیگر برای تأمین نیاز مردم امروز کافی نیست و سازمان‌ها باید مجموعه‌هایی از مزایا و برنامه‌های آموزشی را ارائه دهند تا راه ارتقاء آن‌ها در مسیر دستیابی به وفاداری کارکنان هموار شود. وفاداری کارکنان از مباحثی است که امروزه مدیران مخصوصاً در ساختارهای خدماتی، به علت فضای رقابتی، توجه ویژه‌ای به آن می‌کنند. نیروی انسانی بی تردید مهم‌ترین عامل بین عوامل و منابع تولید است. نیروی انسانی قادر است بر اساس رفتار متعهدانه از بروز زیان جلوگیری کرده و برای سازمان ارزشمند باشد. با بیان امیران در سال ۱۳۸۴، بهبود و پیشرفت عملکرد سازمان‌ها در مسیر رسیدن به اهداف تعیین‌شده مستلزم تعهد و وفاداری کارکنان نسبت به اهداف سازمانی است.

بر اساس پژوهش کاروانا، در سال ۲۰۰۲^۱ وفاداری کارکنان موجب می‌شود که نیروی کار نسبت به موفقیت سازمان احساس تعهد نموده و سازمان متبوعش را به‌عنوان بهترین قلمداد کند. یکی از عوامل مهم ایجاد انگیزه و اشتیاق به فعالیت در کارکنان سازمان، احساس وفاداری است که موجب می‌شود بهره‌وری سازمان بالا رفته و منجر به تعلق، تعهد و رضایت شغلی گردد. در پژوهش تستا^۲ در سال ۲۰۱۱، یک کارمند وفادار در واقع نیروی کاری عنوان شده است که نسبت به موفقیت سازمان احساس تعهد نموده و به سازمان احساس تعلق دارد و توجهی به سایر پیشنهادها نمی‌کند. وفاداری سازمانی به معنی اقدام به انجام وظایف بدون در نظر داشتن علائق فردی، گروه‌های کاری و یا بخش‌ها است. طرفداری و دفاع پیش روی تهدیدات بیرونی و حفظ تعهد حتی در شرایط نامطلوب می‌تواند به‌عنوان وفاداری سازمانی نگریسته گردد. وفاداری و تخصص مهم‌ترین ویژگی‌های کارکنان و به عنوان سرمایه سازمان‌ها محسوب می‌شوند، این گونه کارکنان بر فعالیت‌های خود در سازمان و آنچه باید انجام دهند اشراف کامل دارند.

^۱. Caruana

^۲. Testa

در راستای حفظ وضعیت این دست از سرمایه‌ها باید کوشید تا روزبه‌روز بر حوزه کاری خود تسلط بیشتری پیدا کنند. وفاداری و ماندگاری طولانی مدت کارکنان در سازمان، موجب تقویت بخش نیروی انسانی شده و منابعی از تجربیات را در بر خواهد داشت. کارکنان وفادار، سیاست‌ها، استراتژی‌ها و اهداف سازمان و چگونگی حرکت در راستای دستیابی به منافع سازمان را می‌دانند و نیازی به آموزش و توجیه ندارند. در مقابل تغییر مداوم کارکنان و عدم وفاداری هزینه‌های زیادی را برای سازمان در بر خواهد داشت و منابع سازمان اتلاف می‌شود (برومند ۱۳۸۶).

از نظر گرونهولد^۱، در سال ۲۰۰۱، از طرفی وفاداری منابع انسانی در یک سازمان بر نحوه ارائه خدمات آن‌ها مؤثر است به‌نحوی که عملکرد فردی آن‌ها را بهبود می‌دهد و در نتیجه بر عملکرد سازمان نیز مؤثر است. وفاداری مشتریان در گرو تضمین وفاداری کارکنان به سازمان است که به معنای این است که یعنی تمامی ارکان یک مجموعه باید خود را نسبت به مراجعه کننده متعهد بدانند. توجه صرف به بخش بازاریابی در خدمت‌رسانی به مشتری کافی نیست و برای رسیدن به وفاداری مشتری باید تمامی بخش‌های سازمان در این راستا آموزش ببینند و سازمان نیز، رضایت کارمندان خود را تأمین کرده تا با وفاداری همراه با تعهد کارکنان به سازمان، رضایت مشتری و وفاداری مشتریان را در پی داشته باشد. مدل زنجیره خدمت و سود بیان می‌کند کیفیت خدمات به چه ترتیبی منجر به افزایش وفاداری و رضایت کارکنان و افزایش بهره‌وری سازمان و در نتیجه افزایش رضایت و وفاداری مشتری و در نتیجه سودآوری شرکت‌ها می‌شود.

با توجه به اهمیت وفاداری کارکنان و تأثیر آن بر عملکرد یک سازمان، در تحقیق حاضر موضوع پیش‌بینی وفاداری کارکنان با استفاده از تکنیک‌های پیش‌بینانه داده‌کاوی مورد بررسی قرار می‌گیرد. در صورت تهیه و ارائه مدل مناسب برای پیش‌بینی وفاداری کارکنان تازه استخدام‌شده، سازمان می‌تواند برنامه‌ریزی بهتری برای سازمان‌دهی کارکنان، آموزش و ارتقاء ایشان، و توسعه توانمندی و به‌کارگیری مناسب‌تر برای این افراد داشته باشد و در مجموع منابع سازمان به‌صورت اثربخش‌تری مورد استفاده قرار خواهد گرفت. بر این اساس ویژگی‌هایی از افراد استخدام‌شده مانند جنسیت، تأهل، تحصیلات و غیره به‌عنوان متغیرهای موثر در نظر گرفته و وفاداری کارکنان به‌عنوان متغیر تابع اندازه‌گیری می‌شود.

^۱ . Grønholdt

۱- ۳) ضرورت و توجیه انجام پژوهش

وفاداری کارکنان یکی از عوامل اصلی تداوم و تعالی یک سازمان است. وفاداری، موجب مسوولیت کارکنان نسبت به اهداف سازمان شده و آنها را سمت انجام وظیفه سالم و متعهدانه و به دنبال آن احساس نزدیکی و همبستگی و در نتیجه ماندگاری در سازمان سوق می‌دهد. وفاداری در سازمان جنبه‌های متفاوتی دارد که از جمله آنها می‌توان به وفاداری کارکنان به سازمان و بالعکس و همچنین تأثیرات این دو بر وفاداری مشتریان به سازمان نام برد. بقا و تعالی یک سازمان به وفاداری سازمانی به عنوان یک ارزش ضروری برای کارکنان در سازمان بسیار وابسته است. احساس وفاداری از طرف سازمان به کارکنان موجب می‌شود با انگیزه و اشتیاق بیشتری فعالیت کرده و تمام توجه و توان خود را معطوف کار کنند که موجب افزایش بهره‌وری عامل انسانی خواهد شد. علم به نقش محوری نیروهای انسانی متخصص شایسته و درعین حال وفادار در بازی‌های رقابتی، از یکسو و تغییرات صورت گرفته در انتظارات نیروی کار، رشد فناوری‌های اطلاعاتی و ارتباطی و پویایی‌های محیطی از طریق مطرح کردن مباحثی نوین مانند نیروی سیال از سوی دیگر، وفاداری کارکنان را به یکی از دغدغه‌های اصلی امروز مدیران و پژوهشگران سازمانی تبدیل کرده است. بر اساس تحقیقات صورت گرفته، قضیه ثابت شده و پایه این است که وفاداری کارکنان از سطح رضایت آنها تأثیر می‌پذیرد. تعداد زیادی از کارهای صورت گرفته در امریکا نشان می‌دهد که رضایت کارکنان رابطه مثبتی با وفاداری آنان در شرکت‌های مطبوعشان و نیز رابطه منفی در قصد آنان در ترک خدمتشان دارد. سطح پایینی از قصد ترک خدمت در میان کارکنان راضی از این حقیقت ناشی می‌شود که کارکنان راضی در مقایسه با کارکنان ناراضی بر این باورند که منافع بیشتری به واسطه ماندن در سازمان عاید آنها می‌شود.

۱- ۴) هدف پژوهش

هر پژوهشی برای دستیابی به هدف‌های خاصی انجام می‌شود که در مسئله آن، پنهان است و باید آشکار شود. هدف‌ها به دودسته تقسیم می‌شوند. هدف کلی، پاسخ به مسئله اصلی پژوهش است و مشخص می‌کند در نهایت پژوهشگر، به چه چیزی دست می‌یابد و باید چه نتیجه‌ای را در پایان پژوهش انتظار کشید.

همچنین هدف‌های فرعی پژوهش از هدف کلی آن، برمی‌خیزد. هدف‌های فرعی پژوهش، راهنمایی برای نوع و تدوین ابزار گردآوری اطلاعات است. هدف پژوهش حاضر بررسی و پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی از شرکت‌های ارائه‌دهنده خدمات توزیع برق استان خوزستان است.

۱-۴-۱) هدف اصلی پژوهش

پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک‌های داده‌کاوی.

۱-۴-۲) اهداف فرعی پژوهش

۱. بررسی و شناسایی ویژگی‌ها و متغیرهای کارکنان شرکت‌های خصوصی.

۲. بررسی و مطالعه تکنیک‌های داده‌کاوی پیش‌بینانه.

۳. انتخاب مناسب‌ترین تکنیک داده‌کاوی جهت پیش‌بینی وفاداری کارکنان.

۴. ارزیابی و اولویت‌بندی ویژگی‌های مؤثر کارکنان بر وفاداریشان.

۱-۵-۵) سؤالات و فرضیات پژوهش

سؤال اصلی تحقیق: چگونه می‌توان با استفاده از تکنیک‌های پیش‌بینانه داده‌کاوی، وفاداری کارکنان تازه استخدام‌شده در یک سازمان خصوصی را پیش‌بینی نمود؟

۱-۵-۱) سؤال فرعی تحقیق

- عوامل شخصی مؤثر بر وفاداری کارکنان در شرکت‌های خصوصی کدام‌اند؟
- آیا با مطالعه تکنیک‌های داده‌کاوی می‌توان میزان وفاداری کارکنان را پیش‌بینی کرد؟
- مناسب‌ترین تکنیک داده‌کاوی جهت پیش‌بینی وفاداری کارکنان کدام است؟
- اولویت‌بندی عوامل مؤثر بر وفاداری کارکنان کدام است؟

۱-۶) روش پژوهش

پایه هر علمی، روش شناخت آن است و اعتبار و ارزش قوانین هر علمی به روش شناختی مبتنی است که در آن علم به کار می‌رود. از اصطلاح "روش تحقیق" معانی خاص و متمایزی در متون علمی استنباط می‌شود ولی آنچه در اینجا مطرح است یک فرایند نظام‌مند برای یافتن پاسخ یک پرسش یا راه‌حل یک مسئله است.

این تحقیق از لحاظ طبقه‌بندی بر مبنای هدف، از نوع تحقیقات کاربردی است. هدف تحقیق کاربردی توسعه دانش کاربردی در یک زمینه خاص است. به عبارت دیگر تحقیق کاربردی به سمت کاربرد عملی دانش هدایت می‌شود. تحقیقات کاربردی تحقیقاتی هستند که نظریه‌ها، قانونمندی‌ها، اصول و فنونی که در تحقیقات پایه تدوین می‌شوند را برای حل مسائل اجرایی و واقعی به کار می‌گیرد. این تحقیقات با استفاده از زمینه و بستر شناختی و معلوماتی که از طریق تحقیقات بنیادی فراهم شده برای رفع نیازمندی‌های بشر و بهبود و بهینه‌سازی ابزارها، روش‌ها، اشیاء و الگوها در جهت توسعه رفاه و آسایش و ارتقای سطح زندگی انسان مورد استفاده قرار می‌گیرند.

اطلاعات مورد نیاز این پژوهش از مطالعات میدانی، کتابخانه‌ای و مرور داده‌های گذشته به دست آمده است. گام‌های انجام پژوهش عبارت‌اند از:

۱. انجام مطالعات کتابخانه‌ای پیرامون تکنیک‌های داده‌کاوی و انجام مطالعات کتابخانه‌ای پیرامون وفاداری کارکنان.
۲. انتخاب مناسب‌ترین تکنیک داده‌کاوی برای پیش‌بینی وفاداری کارکنان.
۳. جمع‌آوری داده‌ها (اطلاعات کارکنان، نتیجه وفاداری آنان).
۴. توسعه و بومی‌سازی تکنیک‌های داده‌کاوی.
۵. بسط و بکارگیری مدل پیشنهادی بر روی داده.
۶. تحلیل نتایج و ارزیابی عملکرد.

۷-۱) قلمرو پژوهش

قلمرو موضوعی: بررسی پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی است و می‌تواند در بخش استخدام کارکنان سازمان مورد استفاده قرار گیرد.

قلمرو زمانی: این مطالعه از مهرماه ۱۳۹۷ شروع شده و پیش‌بینی می‌شود تا پایان سال ۱۳۹۸ انجام شود.

قلمرو مکانی: در شرکت خصوصی توزیع برق انجام می‌شود.

جامعه آماری: تحقیق حاضر شامل همه کارکنان شرکت‌های خصوصی توزیع برق استان خوزستان می‌باشد که از ابتدای جذب تا پایان سال ۱۳۹۸ مورد مطالعه قرار گرفته است. با توجه به اینکه در استفاده از تکنیک‌های داده‌کاوی، اندازه بیشتر نمونه موجب افزایش دقت می‌گردد، لذا کل اعضا جامعه تحقیق به صورت سرشماری مورد بررسی قرار می‌گیرد. جمع‌آوری داده‌های کارکنان از پرونده‌های پرسنلی جامعه مورد تحقیق انجام شده است.

۸-۱) نوآوری پژوهش

مهم‌ترین نوآوری تحقیق حاضر استفاده از تکنیک‌های خوش‌بینانه داده‌کاوی در پیش‌بینی وفاداری کارکنان استخدام‌شده مبتنی بر ویژگی‌ها و خصوصیات شخصی ایشان است که در شرکت‌های خصوصی مورد بررسی قرار می‌گیرد.

۹-۱) توصیف اصطلاحات و واژگان کلیدی

داده‌کاوی چیست^۱: به فرایند استخراج و کشف قواعد و الگوهای مفید از میان حجم انبوهی از داده‌های خام که با استفاده از الگوریتم و ساختارهای هوشمند انجام می‌گیرد داده‌کاوی نام دارد، به زبان ساده‌تر، استخراج دانش از میان مجموعه‌ای از داده‌ها را داده‌کاوی می‌نامند.

^۱ . Data Mining

وفاداری کارکنان^۱: در واقع وابستگی قلبی و ذهنی که عملیاتی شده و در نتیجه آن فرد با روحیه‌ای حمایتگر و پایدار نسبت به خود، خانواده، سازمان، اجتماع و غیره متعهد می‌شود.

طبقه‌بندی در داده‌کاوی^۲: هدف از طبقه‌بندی، مشخص کردن ویژگی‌هایی است که بتوان توسط آن، کلاس‌های مختلف را از یکدیگر متمایز کرد طبقه‌بندی در داده‌کاوی طی دو مرحله انجام می‌گیرد. ابتدا از روی داده‌های قدیمی، کلاس‌های مختلف تشخیص داده‌شده و سپس تعلق داشتن داده‌های جدید به کلاس‌های موجود، پیش‌بینی می‌شود. طبقه‌بندی جزو تکنیک‌های یادگیری با ناظر است زیرا با در اختیار داشتن یک مجموعه داده به‌عنوان راهنما، داده‌های جدید را طبقه‌بندی می‌کند. این روش جزو روش‌های پیش‌بینی‌کننده نیز به شمار می‌آید.

درخت‌های تصمیم^۳: بر اساس یک سری قوانین تصمیم‌گیری برای پیش‌بینی و رده‌بندی مورد استفاده قرار می‌گیرند که مزایای متعددی دارد، اولاً بعد از ساخت درخت به راحتی می‌توان علت استنتاج قوانین به دست آمده را مشاهده نمود؛ ثانیاً درخت تصمیم امکان می‌دهد تا دیدگاه خوبی نسبت به صفات مهم‌تر به دست آوریم، زیرا در درخت تصمیم به طور خودکار صفات مهم‌تر به گره‌های تصمیم بالایی درخت انتقال می‌یابند و همچنین، صفات کم‌اهمیت را کنار می‌گذارد.

نزدیک‌ترین همسایگی^۴: یک روش نا پارامتری است که در داده‌کاوی، یادگیری ماشین و تشخیص الگو مورد استفاده قرار می‌گیرد. نزدیک‌ترین همسایگی یکی از ده الگوریتمی است که بیشترین استفاده را در پروژه‌های گوناگون یادگیری ماشین داده‌کاوی، هم در صنعت هم در دانشگاه داشته است.

ماشین بردار پشتیبان^۵: این روش در سال‌های اخیر کارایی خوبی نسبت به روش‌های کلاسیک طبقه‌بندی نشان داده است. اساس کار طبقه‌بند SVM جداسازی دسته‌های داده به شکل خطی است. هنگام دسته‌بندی خطی داده‌ها تلاش بر این است که معادله خطی شود که حاشیه اطمینان بالاتری نسبت به دیگر تفکیک‌کننده‌های احتمالی داشته باشد.

^۱ . Employees Loyalty

^۲ . Classification in Data Mining

^۳ . Decision Tree

^۴ . K-Nearest Neighbors

^۵ . Support Vector Machine

۱-۱۰ ساختار پایان نامه

در فصل اول دانستیم هدف پژوهش، پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک‌های داده‌کاوی است و در فصل دوم ابتدا به تعاریف لازم پرداخته خواهد شد از جمله این تعاریف، تعریف وفاداری است: وفاداری به معنای احساس هویت و پیروی از مدیر سازمان و به‌صورت کلی، و فراتر رفتن از علایق گروهی افراد است و خواهیم خواند که وفاداری و تعهد در رابطه با یکدیگر دیده می‌شود از این‌رو نتایج، پیامدهای وابستگی و تعهد نسبت به سازمان را بررسی می‌کنیم و پس از آن شاخص‌های اندازه‌گیری وفاداری کارکنان را برمی‌شمایم و بعد از تعریف مفهوم وفاداری کارکنان به مطالعاتی که در این مورد در گذشته تابه‌حال انجام‌شده اشاره می‌کنیم. در فصل سوم تکنیک داده‌کاوی مورد مطالعه تشریح شده و دلایل انتخاب این تکنیک و چگونگی عملکرد و مزایا و معایب آن را نسبت به سایر روش‌های داده‌کاوی مورد بررسی قرار خواهیم داد. در فصل چهارم به تجزیه و تحلیل داده‌ها می‌پردازیم و در فصل پنج این پایان‌نامه به نتیجه‌گیری این روش‌ها و پیشنهادها برای آینده پرداخته خواهد شد.

فصل دوم:

مبانی نظری و پیشینه پژوهش

۲-۱) مقدمه

وفاداری یکی از ارزش‌های اجتماعی بوده و در زندگی و در روابط مابین افراد از اهمیت بالایی برخوردار است. مهم‌ترین سرمایه سازمان‌ها کارکنان وفادار و متخصص هستند. تحقیقات مختلف نشان داده که امکان ندارد یک سازمان مدعی وفاداری و رضایت مشتریانش باشد بدون اینکه کارمندان راضی و وفاداری داشته باشد. در مطالعه‌ای که در ایالات متحده آمریکا بر روی کارکنان بانک شهر لابلان در سال ۲۰۰۴ صورت گرفت، تأثیر متغیرهای سن، جنس، داشتن منزل شخصی و تحصیلات بر روی رضایت کارکنان بررسی شد. این تحقیق بر روی ۲۰۲ کارمند در مناطق مختلف جغرافیایی شمال، جنوب، مرکز شهر صورت گرفت و جمعیت هدف کارمندانی بودند که سمت مدیریتی نداشتند. نتایج این تحقیق نشان می‌دهد که کارکنان زن با پرداخت دستمزد بیشتر احساس رضایت شغلی می‌کنند. این مطلب استدلال‌های اسپکتور را اثبات می‌کند که، چون زنان انتظار کمتری از کاردارند بنابراین با کمترین دستمزد راضی می‌شوند ولی کارکنان مرد به‌وسیله ارتقاء به پست‌های سرپرستی از کار راضی می‌شوند. شاید این مطلب به این خاطر می‌باشد که مردان فرصت بیشتری برای ابراز وجود و تأثیرگذاری برای اتخاذ تصمیمات مهم در سازمان‌ها را دارند. کارکنانی که تحصیلات پایین‌تری دارند اغلب کمتر از کار رضایت دارند. این ممکن است به دلیل آن باشد که این کارمندان مهارت لازم را برای تغییر ماهیت صنعت به سمت مدرنیته شدن ندارند و یا کمتر مورد توجه مدیران قرار می‌گیرند کارمندانی که مستأجر بودند نیز رضایت شغلی خیلی کمتری نسبت به صاحبان منازل شخصی داشتند. گرچه پرداخت یک حقوق عالی و افزایش سخاوتمندانه حقوق می‌تواند منجر به تأمین نیازهای مادی کارکنان شود ولی منافع غیر ملموس و نامشهودی نظیر محیط کار خوب، ساعات کار قابل‌انعطاف و غیره نیز می‌تواند بر نگرش و دیدگاه کارکنان تأثیر بگذارد و آنان را برای تلاش بیشتر برانگیزاند. درحالی‌که بسیاری از سازمان‌ها مزایایی مانند مرخصی با حقوق، رفتن به مسافرت، مراقبت‌های پزشکی، فعالیت‌های ورزشی و غیره را به کارکنان تخصیص داده‌اند ولی ممکن است در یک محیط رقابتی این مزایا کافی نباشد. با توجه به

اهمیت روز افزون متعادل سازی کار و زندگی برای کارکنان نیاز است تا سازمان‌ها متفاوت عمل کرده و برنامه‌های منعطفی برای کارکنان در محیط ارائه دهند.

۳-۱) مبانی نظری

۲-۲-۱) تعریف وفاداری

وفاداری یعنی داشتن حس موجودیت در یک سازمان به همراه تبعیت از مدیر سازمان، به عبارت ساده‌تر گذر از علایق گروهی افراد است (ارگان و همکاران^۱، ۲۰۰۶). شخص وفادار دارای ویژگی‌هایی چون تبلیغات برای سازمان، حمایت و دفاع از سازمان و احساس تعهد حتی در شرایط نامطلوب نسبت به سازمان است (پودوسکووف و همکاران^۲، ۲۰۰۰). (پاورز^۳، ۲۰۰۰) بیان می‌کند که وفاداری به صورت ارادت و دل‌بستگی عاطفی نسبت به چیز خاصی است که ممکن است مربوط به یک شخص، یک گروه، یک وظیفه و یا یک هدف باشد.

وقتی ما نسبت به شخص، سازمان و یا خود وفاداریم، به‌نوعی خویش را با موضوعی درگیر کرده و هویت بودن را وجدان می‌کنیم. بودن، خود یک احساس ارزشمند برای ادامه حیات یا میل به زیستن است. بنابراین خود وفاداری به انسان هویت می‌بخشد. انسان‌ها نسبت به چیزهایی وفادارند که با پوست و خونشان ممزوج شده و آن را از خودشان می‌دانند؛ بنابراین به بی‌راهه نرفته‌ایم اگر بگوییم اصولاً در وفاداری، فرد احساس مالکیت خود را تقویت می‌کند و چون مالکیت، القاء داشتن امکانات بیشتر است، احساس خوشایندتری برایش تولید می‌کند، احساس قدرت. ازاین‌رو است که افراد به دفاع از موضوعات دلخواهشان می‌پردازند چراکه موضوعات دلخواه آن‌ها با حیثیتشان آمیخته می‌شود و وقتی شما موضوعاتشان را نقد می‌کنید، بدون تحلیل نقد، به دفاع برخاسته و موضع‌گیری احساسی می‌کنند، زیرا شما ناخواسته هویت او را زیر سؤال برده‌اید. انسان‌ها چیزی جز اندیشه نیستند، از آنجایی که ابتدا در پرستش خویشتن پیشگام‌اند، موضوعات مورد علاقه‌شان را نه تنها موضوعات، بلکه خودیت خودشان دانسته و نقد را به مثابه حمله به هویت تلقی می‌کنند و اینجاست

¹. Organ et al

². Podsakoff et al

³. Powers

که چشم و گوش را بسته، گاه لبخند می‌زنند، گاه سکوت می‌کنند و گاه فریاد کشان از خود دفاع می‌کنند و بالکل فراموش می‌کنند که شما موضوع یا سازمان را مورد انتقاد قرار داده بودید، نه حیثیت او را.

حال اگر انسان‌ها ناخودآگاه یا خودآگاه، خودیت خود را با موضوعات مخلوط می‌کنند، چه شیرین است که این اختلاط با سازمان و اهداف آن باشد و اما چگونه؟ پاسخ عملیاتی به این سؤال، چالش‌برانگیزترین موضوع بازار امروز است. به این معنی که وقتی در بازار پرتلاطم موجود می‌خواهیم سهمی داشته باشیم می‌بایست سهمی از قلب‌ها باشد نه سهمی از دلار. در حقیقت سرعت رسیدن به دلار از راه دل میسرتر است. ارزش‌ها و هدف‌هایی که تصمیم‌های فرد را در سازمان هدایت می‌کند بیشتر به صورت هدف‌های سازمانی هستند. یعنی هدف‌های خدمت کردن و حفظ خود سازمان در اصل، این هدف‌ها از راه اعمال اقتدار بر فرد تحمیل می‌شوند، ولی تا اندازه بسیاری ارزش‌ها اندک‌اندک "درونی" می‌گردند، و در روان‌شناسی و نگرش‌های عضو سازمان جای می‌گیرند. فرد یک پیوستگی و وفاداری به سازمان پیدا می‌کند که به طور خودکار یعنی بدون ضرورت محرک‌های بیرونی تضمین می‌کند که تصمیم‌هایش سازگار با هدف‌های سازمان خواهند بود. این وفاداری خود دارای دو جنبه است: نخست اینکه می‌تواند نوعی دل‌بستگی به هدف‌های خدمت سازمان را در برداشته باشد، و دوم آنکه می‌تواند نوعی دل‌بستگی به پایایی و رشد خود سازمان را در برداشته باشد (هربرت ای سایمون^۱، ۱۳۷۱).

۲-۲-۲) تحقق وفاداری

وفادار بودن در سازمان، زمانی اتفاق می‌افتد که کارکنان به احساسی دست یابند که فراتر از شرح وظایف مکتوب، خود را ملزم به حرکت‌دادن چرخه‌ای از آن می‌دانند. به این معنی که آن‌ها نگران به خطر افتادن جایگاه سازمان چه در بازار، چه از نظر سهم محبوبیت و مقبولیت عمومی، هستند و دلسوزانه در راستای به ثمر نشاندن اهداف سازمان تلاش می‌کنند. و تفکر اینکه، این کاروظیفه من نیست نمودی نداشته و کارکنان در تلاش هستند کارهای ارجاع شده به سازمان را در اسرع وقت انجام داده و رضایت مشتریان را بیش‌ازپیش حاصل کنند. این مرحله مستقیماً مرتبط با طی مراحل فردی یک کارمند در هرم نیازهای مازلو است یعنی فرد از نیازهای زیستی و مسکن فراتر رفته و نسبت به کار و تعامل اجتماعی

^۱. Herbert simon

خود را ایمن می‌داند و در مرحله بالاتر به دنبال مقبولیت و درخشش شغلی است. اینجا همان جایی است که کارمند برای کسب معاش تلاش نکرده و هدف والاتری را پیگیری می‌کند.

پس مهم‌ترین عامل طی فرآیند این هرم است؛ بنابراین عملکرد سازمان مستقیماً با نحوه عملکرد ایشان مرتبط بوده و خود کارمند به عنوان اولین مشتری سازمان می‌بایست جزو اهداف سازمان باشد و در مقابل، بین سازمان و کارمند، احساس امنیت شغلی، رضایت شغلی، امنیت روانی نسبت به جایگاه سازمان در اجتماع که احساس کند در یک خانواده بزرگ و مشهور اجتماعی کار می‌کند منتقل شود.

در پیشینه علم روان‌شناسی، وفاداری و تعهد در ارتباط با دیگران مفهوم پیدا می‌کند و نشان دهنده مقدار انرژی مازادی است که یک شخص احتمالاً برای دیگران یا سازمان مطبوع خود صرف می‌کند. اشتیاق و میل بیش از حد برای حفظ این رابطه و قبول ارزش‌های شخصی دیگر یا سازمان نشانه تعهد و وفاداری است. وفاداری با تداوم تماس و نیز حس استقلال شخص تقویت خواهد شد. تغییر محیط کار ممکن است رخ دهد، ولی وفاداری به محیط از بین نخواهد رفت زیرا پیوندهای مابین افراد بسیار قوی هستند.

۲-۳ نتایج و پیامدهای وابستگی و تعهد نسبت به سازمان

- وابستگی و تعهد موجب افزایش بهره‌وری می‌شود.
- وابستگی و تعهد موجب سازگاری بهتر فرد با سازمان می‌شود.
- وابستگی و تعهد فرد را خلاق و در برابر تغییرات سازمانی انعطاف‌پذیر می‌کند.
- وابستگی سازمانی و تعهد باعث مشارکت در تصمیم‌گیری می‌شود و مدیر یک سازمان معمولاً کارکنانی را در تصمیم‌گیری مشارکت می‌دهد که وابستگی و تعهد بیشتری در آن‌ها نسبت به سازمان سراغ داشته باشد.
- وابستگی سازمانی و تعهد باعث افزایش سطح اطلاعات کارکنان برای حفظ هرچه بیشتر منافع سازمان می‌شود.
- وابستگی سازمانی و تعهد باعث تقویت حس کار گروهی و همگرایی بیشتر کارکنان در جهت حفظ منافع خود و سازمان می‌شود و فرهنگ سازمانی را ارتقاء می‌بخشد.

– وابستگی سازمانی و تعهد با کاستن تنش‌ها باعث افزایش بهره‌وری سازمان و کاهش هزینه‌ها می‌شوند (هربرت سایمون، ۱۳۷۱).

یافته‌های برخی از مطالعات نشان داده است که نگرش شغلی از جمله تعهد سازمانی با بعضی از رفتارهای کار از جمله تمایل به ترک شغل، قصد جستجوی شغلی دیگر، غیبت و عملکرد شغلی رابطه معنی‌دار دارند (ماتیو و زاجاک^۱، ۱۹۹۰، (کخ و استیرز^۲، ۱۹۷۸)، (پورتر^۳، ۱۹۷۴)، (مؤدی و همکارانش^۴، ۱۹۷۴)، بر این عقیده‌اند که کارمندان دارای تعهد سازمانی بالا ممکن است نسبت به کسانی که تعهد کمتری دارند، عملکرد بهتری داشته باشند.

مؤدی و همکارانش معتقد بودند که هر سازمان نیازمند اعضای با وابستگی بیش از قرارداد مبادله‌ای رسمی است. به مفهوم دیگر، سازمان‌ها نیازمند اشخاصی هستند که به نفع سازمان و بیش از چارچوب مقررات وظایف سازمانی متعهد باشند. این امر مخصوصاً در جایگاه‌های مهم بیشتر نمایان می‌شود.

۲-۲-۴) وفاداری کارکنان به هدف سازمان

گرچه هدف سازمان در پیوند با رفتار آن دسته از شرکت‌کنندگان که نام مشتری خوانده می‌شوند، دارای بزرگ‌ترین اهمیت است، به تقریب همه اعضاء سازمان، کم یا زیاد، به هدف سازمان وابسته می‌شوند و رفتارشان از آن تأثیر می‌پذیرد. به این موضوع از این پیش درباره سازمان‌های داوطلبانه اشاره شد؛ این موضوع درباره بنگاه‌های دولتی و سازمان‌های بازرگانی، گرچه به درجات کمتر نیز درست است. این یک جزء و جزئی بسیار مهم، از وفاداری سازمانی است. اگر هدف سازمان به ظاهر سودمند باشد، اعضای سازمان که توجهشان از راه کار روزانه‌شان پیوسته به‌سوی آن است، نسبت به اهمیت و ارزش آن احساس قدردانی خواهند کرد. گاهی قدرشناسی بیش از اندازه در حد گزافه، و دستیابی به آن هدف، در آن درجه جهت آنان دارای ارزش مشخصی خواهد شد. وفاداری کارکنان به هدف سازمان، وفاداری از نوع بسیار متفاوتی

^۱. Mathiew&zajac

^۲. koch & steers

^۳. porter

^۴. Movday et al

می‌تواند باشد. وفاداری به خود سازمان و دلبستگی به پایایی و رشد آن عاملی که مشارکت فرد را در سازمان برمی‌انگیزد از اندازه و رشد سازمان سرچشمه می‌گیرد. این‌ها را می‌توان ارزش‌های "بقا و پایداری" نامید.

در میان گروه، برجسته‌ترین کسی که این ارزش‌ها برای وی مهم هستند کارفرما است این سخن درست است که کارفرما تا جایی که یک مرد اقتصادی است به منافع علاقه‌مند است نه به اندازه و رشد سازمان. در عمل این اعتراض جدی بشمار نمی‌آید. نخست به آن دلیل که منافع به طور معمول به اندازه و رشد پیوستگی نیاز دارد. دوم آن که بسیاری از کارفرمایان به ارزش‌های غیرمادی مانند منزلت و قدرت نیز همانند سود علاقه‌مندند. این پیوستگی به هدف‌های بقا و پایداری حتی برای گروه مدیرحرفه‌ای، که نظارت فعال بر بیشتر سازمان‌های بزرگ دارند به چشم می‌خورد. هدف‌های پاینده ماندن شاید ارزش‌های مهمی را برای دیگر کارکنان سازمان به‌ویژه آنان که پویا و پرتحرک‌اند، نیز فراهم آورند. سازمانی که در حال رشد و نیک‌فرجامی است بیش از سازمانی که راکد یا در حال نزول است فرصت‌های نیکوی مهم‌تری را زمینه منزلت و پیشرفت عرضه می‌دارد. فردی که به هدف‌های سازمان وفادار است، در برابر هرگونه اصلاح آن هدف‌ها ایستادگی می‌کند، و حتی ممکن است اگر آن هدف‌ها به طور ریشه‌ای تغییر یابند، وی از ادامه مشارکتش در سازمان رو بتابد. فردی که به سازمان وفادار است، از دگرگونی‌های فرصت‌طلبانه در هدف‌های سازمان که پیش‌بینی می‌شود پایداری و رشد آن را افزایش دهند پشتیبانی می‌کند (هربرت ای سایمون، ۱۳۷۱).

۲-۵) وفاداری به خود سازمان

در حال حاضر این موضوع مطرح است که چرا کارکنان به سازمان وفادار می‌مانند؟ شاید دلیل این است که برخی عاشق کار خود و نتیجتاً به سازمان وفادار هستند و برخی دیگر تنها به علت همسویی اهداف خود با اهداف سازمان، به آن وفادار می‌مانند. تعدادی هم به علت ترس از دست رفتن موقعیت می‌مانند و دست آخر بخشی به دلیل تعهد به سازمان یا رئیس خود، وفادارند. پس برخی از عنوان‌های تعهد نام برده تأثیری منفی بر سلامت روانی، احترام به خود و رضایت شغلی در کارکنان دارند. پس روشی مناسب ایجاد تعهد کدام است؟ در این جا سه نوع متداول از انواع تعهد کارکنان تشریح و ابزاری

برای چگونگی تغییرات جهت بهبود میزان تعهد و وفاداری اعضای سازمان ارایه خواهد شد. ابزار مدل سه جزیی تعهد برای اولین بار توسط مایر^۱ و آلن^۲ بیان شد.

ابزار نشان دهنده این واقعیت است که وفاداری کارکنان نسبت به سازمان، یک رفتار روانشناختی است. با استفاده از این ابزار قادر به افزایش احساس وفاداری در میان کارکنان خود هستیم. این ابزار تعهد افراد را در سه جایگاه تعریف می‌کند:

۱. تعهد عاطفی به کار: تعلق به کار هنگامی ایجاد خواهد شد که فرد احساس عاطفی زیادی نسبت به سازمان یا کار خود داشته باشد. شخص متعهد، اهداف و ارزش‌های سازمان را به درستی شناخته و با تمام وجود برای تحقق آن‌ها تلاش می‌کند. این شخص از کار لذت برده و احساس خوبی نسبت به آن دارد و این موضوع سبب افزایش رضایت شغلی و در نهایت افزایش وفاداری به سازمان خواهد شد.

۲. تعهد ناشی از ترس: این گونه از تعهد زمانی واقع می‌شود که فرد جوانب مثبت و منفی ترک کار خود را بررسی کند. در این حالت چون فرد حس می‌کند با حضور در کاری جدید تمامی تجاربش از دست می‌رود به کار خود وفادار می‌ماند. در این حالت فرد بسیاری از مزایای ناشی از تغییر محل کار را از دست می‌دهد و با افزایش سابقه کاری زیان‌های ناشی از این نوع تعهد برای فرد به علت گرفتاری به روزمرگی و سازمان به علت مواجه شدن با فردی فاقد انگیزه و نوآوری، هر دو بیشتر خواهد شد.

۳. تعهد اخلاقی: فردی با این نوع تعهد حتی اگر از شرایط خود ناراضی باشد، نسبت به سازمان احساس وفاداری دارد، چرا که این کار درست‌ترین تصمیم حرفه‌ای اوست. این نوع تعهد در افراد ریشه در عواملی مانند احساس دین به سازمان، درک سرمایه‌گذاری سازمان در آموزش وی یا دریافت پاداش مناسب و یا حتی روش تربیت فرد در خانواده دارد.

شناسایی جنس تعهد افراد به عنوان یک رهبر سازمانی بسیار مهم است. پس از شناسایی توصیه می‌شود که به تدریج تعهد ناشی از ترس و تعهد اخلاقی را کاهش داده و با افزایش وابستگی عاطفی افراد، تعهد عاطفی را رشد دهید؛ چرا که وقتی جنس تعهد کارکنان از انواع دوم و سوم باشد، ممکن است با کمترین سطح روحیه و انگیزش به کار خود ادامه دهند

¹ J.Meyer

² N.Allen

و این روی تعهد دیگر اعضای سازمان نیز تاثیر نامطلوب دارد. برای تشویق افراد به تغییرات مثبت، مطمئن شوید که اهداف فردی و سازمانی همخوانی داشته و نقش‌های افراد متناسب با مهارت‌ها و علایق آن‌ها باشد. یادمان باشد که تنها با احساس تجارب مثبت، تعهد عاطفی حاصل می‌شود پس کارکنان را مورد تشویق قرار دهید (آلن و مایر ۱۹۹۶).

۲-۲-۶) مفهوم وفاداری کارکنان

وفاداری کارکنان یعنی تعهد برای موفقیت سازمان و اجرای بهترین کار برای سازمان (پینا و کان ها^۱، ۲۰۰۲). بررسی مطالعات پیشین نشان می‌دهد که ارتباطی منسجم بین تعهد سازمانی و وفاداری کارکنان، وجود دارد طوری که تشخیص آن‌ها از هم به آسانی مقدور نیست. محققان دیدگاه‌های مختلفی در خصوص مفهوم وفاداری دارند. برخی وفاداری را به شکل یک فرایند شرح داده و معتقدند که اگر فردی با وجود آگاهی از شرایط بهتر یک شغل، حقوق بیشتر و شرایط مناسب‌تر، از پذیرش شغل جدید امتناع کند، وفاداری فرد به سازمان پنداشته می‌شود (اچ بکر^۲، ۱۹۶۰). از دیدگاه (آلن و مایر^۳، ۱۹۹۷)، وفاداری با سه مشخصه تعیین می‌شود: اول اعتقاد بالا به ارزش‌های سازمان، دوم تلاش فراوان جهت تحقق اهداف سازمان و در نهایت تمایل فراوان جهت ماندن در سازمان. در این مدل دلیل وفاداری به سازمان، متعهد بودن نسبت به سازمان و اهداف آن است (ساوارکین^۴، ۲۰۰۹).

بقای و تعالی سازمان به وفاداری سازمانی به‌عنوان یک از ارزش لازم وابسته است. هنگامی که کارکنان احساس امنیت شغلی دارند و مزایا و حقوقشان مورد تایید مدیریت سازمان است، با تمایل و اشتیاق بیشتر و بدون توجه به مسائل حاشیه‌ای با تمام توجه به فعالیت می‌پردازند که نتیجه آن افزایش بهره‌وری عامل نیروی انسانی خواهد بود (طاهری روز بهانی، نعمتی، ۱۳۹۱).

کارکنان وفادار و متخصص مشرف بر فعالیت‌های خود سرمایه اصلی سازمان‌ها هستند. وفاداری و ماندگاری طولانی مدت کارکنان در سازمان، موجب افزایش تجربه و مهارت نیروی انسانی سازمان می‌شود (وانگ و پگی^۵، ۲۰۰۹). کارکنان وفادار

^۱. Pina & Cunha

^۲. H. J. Becker

^۳. Allen & Meyer

^۴. Savareikiene

^۵. Wang & Peggy

بسیاست‌ها، استراتژی‌ها و رسالت سازمان آشنا و همسو بوده و می‌دانند برای دستیابی به منافع سازمان چگونه قدم بردارند، در مقابل تغییر مدام کارکنان و عدم وفاداری موجب ایجاد هزینه‌های جدید یا نوعی دوباره‌کاری شده و منابع سازمان را هدر می‌دهد. انگیزه در کارکنان باعث ایجاد وفاداری، احساس تعهد، تعلق و رضایت شغلی خواهد شد.

۲-۲-۷) وابستگی سازمانی و نقش آن در ایجاد تعهد و وفاداری به سازمان

تجربه نشان داده است که منابع انسانی عامل اصلی و عمده پیشرفت‌های جوامع بشری در زمینه‌های مختلف بوده است. بر این اساس عدم رضایت و عدم احساس تعلق و وفاداری کارکنان به سازمان، علاوه بر اینکه باعث افزایش هزینه‌های سازمانی می‌شود ممکن است ترک سازمان را به همراه داشته و بر روحیه سایر کارکنان نیز تأثیر منفی داشته باشد. یکی از ویژگی‌ها و مشکلات سازمان‌ها در کشورهای جهان سوم به نظر می‌رسد عدم احساس تعلق و وفاداری مدیران و کارکنان به سازمان متبوع خود باشد. مؤدی و همکارانش بر این عقیده‌اند که هر سازمان نیازمند اعضای با وابستگی بیشتر از قرارداد مبادله‌ای رسمی است.

۲-۲-۸) راهکارهای وفاداری کارکنان به سازمان

آنچه در موارد زیر ذکر می‌شود بیست و شش روش، برای جذب، نگه داشتن و باانگیزه کردن کارکنان به کار می‌رود.

• قدردانی

تعیین زمانی برای قدردانی از کارکنان نه به شیوه پیچیده‌ای بلکه با جملات ساده، تلاش‌های ارزشمند کارکنان را در سازمان انعکاس دهید حتی به صورت خبری کوچک.

• مزایا

ایجاد تعادل بین کار و زندگی برای کارکنان اهمیت به سزایی دارد و باعث می‌شود سازمان شما متفاوت عمل کند.

• مراقبت

احساس مسوولیت دلسوزانه در برابر مشکلات کارکنان خود تا در عوض آن‌ها احساس وفاداری و وابستگی شدیدی به سازمان پیدا کنند.

• خواسته‌های کارکنان

کارکنان به اشکال مختلف پاداش، واکنش نشان نمی‌دهند. مثلاً همه کارکنان مایل نیستند در جمع معرفی شوند ولی برخی دیگر ممکن است از تشویق در جمع خوشحال شوند.

• برابری

برابری لزوماً به معنای عملکرد یکسان نیست و کارکنان انتظار ندارند به علت شرایط و نیازهای متفاوت با همه مشابه رفتار شود ولی آنچه نیاز دارند برخورد درست است.

• چشم‌پوشی

هنگام خطا واکنش سازمان باید طوری باشد که موجب تغییرات زیاد در احساس وفاداری و احترام کارکنان نسبت به سازمان نشود. با توجه به این نکته که ممکن است خطا از فرآیند ناشی شود کمک کنید از اشتباهات درس گرفته شود.

• بازخورد

بازخورد مستمر جهت بازنگری عملکرد کارکنان در طول سال باعث می‌شود تفاوت عملکرد و انتظاری که از آن‌ها می‌رود را دریابید.

• صداقت

صداقت در رفتار شما حتی اگر منجر به دریافت اطلاعات ناخوشایند باشد ستایش می‌شود. بهتر است اخبار مستقیماً از خود شما شنیده شود تا غیرمستقیم، این رفتار در عملکرد کارکنان تأثیر می‌گذارد.

• مرتبط بودن

به کارکنان فرصت مشارکت در اتخاذ تصمیم‌های موثر در سازمان را بدهید. این امر در مسائل کلی و حتی جزئی موجب ایجاد انگیزه در آنها می‌شود.

• سرعت

رسیدگی با سرعت به درخواست کارکنان از وظایف مهم شما است. حتی اگر موفق به حل مسئله نشدید باید تعهد خود را نسبت به حل مشکلات آنها ولو با اطلاع رسانی سریع و ساده نشان دهید.

• فرآیندها و ابزار ساده و مختصر

کارکنان دوست دارند که مؤثر و کارآمد باشند و در صورتی که فرآیندها و ابزار کارشان ناکارآمد و غیر مؤثر باشد نمی‌توانند به خوبی فعالیت کنند. پس وظیفه شماست که فرآیندها و ابزار مناسب در اختیار آنها بگذارید.

• شنیدن

با مکانیزم‌های رسمی و غیررسمی شنیدن نظرات کارکنان به صورت‌های متفاوت این فرصت را به آنها می‌دهید که نظرات و عقاید خود را بیان کنند و این روش، کلیدی برای ارتباط کارکنان با محتوای فعالیت‌هایشان است.

• مدیریت مؤثر

ثابت شده است که مدیر مستقیم یک فرد بیشترین تأثیر را بر کارایی، اثربخشی، وفاداری و ایجاد انگیزه در وی نسبت به سازمان دارد.

• فشار آوردن

کارکنان خود را وادار کنید تمام سعی خود را بکنند تا کار خود را کمی بهتر از قبل ارائه کنند.

• فرصت‌ها

فرصت‌هایی برای رشد کارکنان علاقه‌مند، در طول فعالیت سازمان فراهم کنید تا بتوانند به سطوح بالاتر برسند یا نقش‌های پرچالش‌تری را ایفا کنند.

• پرداخت

پرداخت مهم‌ترین عامل انگیزه کارکنان نیست ولی بسیار حائز اهمیت است. مطمئن شوید پرداخت‌ها با میانگین پرداخت‌های درون سازمان و نیز بیرون سازمان یعنی بازاری که کارکنان را از آن جذب می‌کنید تناقض نداشته باشد.

• سؤال

با پرسش از کارکنان خود، مطالب زیادی از آن‌ها یاد خواهید گرفت. هر نیرو با توجه به تخصص خود می‌تواند موجب افزایش بهره‌وری حتی به میزانی اندک گردد.

• پاداش

باید برای تلاش‌های فردی و گروهی در راستای مأموریت سازمان، به کارکنان خود پاداش دهید. در این زمینه باید نوآور باشید زیرا نتیجه موجب افزایش انگیزه کارکنان می‌شود.

• مشارکت در اطلاعات

کارکنان باید چیزهایی را بدانند که به آن‌ها کمک کند مشارکت با ارزش‌تری در سازمان خود داشته باشند. این مشارکت به‌سادگی باعث می‌شود کارکنان اطلاعات مثبت در مورد محصولات و خدمات شرکت به اطلاع مشتریان برسانند.

• صحبت کردن

صحبت کردن با کارکنان نشان می‌دهد که شما به آن‌ها و عقایدشان علاقه دارید. کارکنان نیز در فرصتی که ضمن صحبت با شما به دست می‌آورند سؤالاتی را مستقیماً مطرح کرده و جواب بگیرند.

• درک نکات مهم هر بخش

شما باید مطلع باشید که در هر بخش چه مواردی حائز اهمیت است.

• ارزش

در عین حالی که باید سازمانی باشید که موجب مباحثات کارکنان باشد و حقوق و مزایای سخاوتمندانه‌ای بپردازید و امکانات مناسب رفاهی برای آن‌ها فراهم کنید ولی هر چند وقت یکبار درباره آن‌ها فکر کنید یا به کارکنان یادآوری کنید که چنین ارزش‌هایی وجود دارند.

• چرا

وقتی شما تصمیم‌گیری می‌کنید یا سیاست جدیدی باید به اجرا درآید یا سیاست قبلی باید تغییر کند قبل از اجرای تصمیم‌گیری دلیل آن را نیز به کارکنان بگویید. ارتباط با کارکنان در مورد دلایل تصمیم‌های اتخاذ شده ذهن آن‌ها را به تغییرات یکپارچه می‌کند.

• تعالی

کارکنان دوست دارند احساس کنند بخشی از سازمان هستند که نزد دوستان و آشنایان خود می‌توانند به آن مباحثات کنند و با آن‌ها دوباره کارهای مهم و بزرگی که سازمان انجام می‌دهد و نحوه مشارکت خودشان صحبت کنند، وقتی سازمان به اهداف خود برسد، محصولات جدید تولید کند و جامعه آن را بشناسد، کارکنان بر خود خواهند بالید.

• بلی

گفتن نه آسان است ولی پیدا کردن راه‌هایی برای رفع نیاز کارکنان سخت و زمان‌بر است هیچ‌کس دوست ندارد «نه» بشنود ولی همه ما از اینکه جواب مساعد بشنویم خوشحال خواهیم شد.

• رغبت

رغبت داشتن به کار لازم است، وقتی کارکنان از شغل و محیط کار خود لذت ببرند، حضور در محل کار برای آن‌ها لذت بخش خواهد بود، پس راه‌هایی برای ایجاد محیطی بانشاط فراهم کنید تا آن‌ها با رغبت سرکار برگردند و به بهترین نحو کار کنند.

۳-۲) پیشینه تحقیق

پس از مطالعه و بررسی منابع خارجی و داخلی (مقالات، کتاب‌ها، مجلات و سایت‌ها و ...) پیرامون تحقیق حاضر ابتدا به تعدادی از تحقیقاتی که در داخل ایران و سپس مواردی از تحقیقات که در دیگر کشورها انجام شده، پرداخته می‌شود.

۲-۳-۱) مطالعات داخلی

مقاله ای تحت عنوان تعیین اولویت عوامل موثر بر تعهد سازمانی کارکنان (مطالعه موردی در شرکت های پالایش گاز فجر و بیدبلند) ارائه شده است که نتایج آن حاکی از این است که از میان عوامل مختلفی که بر تعهد سازمانی تأثیر می‌گذارند، متغیرهایی چون درک حمایت سازمانی قوی‌ترین رابطه مستقیم و مثبت و متغیر احساس عدالت سازمانی همبستگی نسبتاً قوی با تعهد سازمانی دارد. متغیر احساس امنیت شغلی همبستگی نسبتاً ضعیفی با تعهد سازمانی داشته است و روش اصلی این پژوهش پیمایشی و مقطعی است. جامعه آماری این پژوهش را کارکنان دو پالایشگاه گاز فجر و بید بلند از شرکت‌های تابعه شرکت گاز ملی ایران تشکیل می‌دهد. به دلیل حجم بالای نمونه از تحلیل عاملی و رگرسیون استفاده شده است. (مدنی و زاهدی، ۱۳۸۴).

مقاله‌ای تحت عنوان "بررسی رابطه مؤلفه‌های عدالت سازمانی با تعهد سازمانی در کارکنان شهروندی شهر اصفهان" توسط غفوری و گل پرور ارائه شده است. نتایج حاکی از آن بود که عدالت سازمانی و سه مؤلفه آن همبستگی مثبت و معنی‌داری با هر یک از حیطه‌های تعهد سازمانی یعنی تعهد سازمانی عاطفی، تعهد سازمانی هنجاری و تعهد سازمانی

مستمر داشتند و به منظور تجزیه و تحلیل داده‌ها از روش‌های آماری ضریب همبستگی ساده و چندگانه (تحلیل رگرسیون) استفاده شده است (محمد رضا غفوری و محسن گل پرور ۱۳۸۷).

قاضی زاده در سال ۱۳۸۷ در تحقیق خود به بررسی عوامل مؤثر بر ماندگاری سازمانی نیروی انسانی پرداخته است و به این نتیجه رسیده که کارکنانی که به سازمان وفادار بوده و در طول سال‌های متمادی برای سازمان فعالیت می‌کنند، با سیاست‌ها، استراتژی‌ها و رسالت سازمان به خوبی آشنا بوده و می‌دانند چگونه باید در راستای منافع سازمان قدم بردارند و در نتیجه نیازی به آموزش، توجیه و مسائلی از این قبیل ندارند.

در مقابل، تغییر پیاپی کارکنان و عدم وفاداری کارکنان به سازمان، موجب وارد آمدن هزینه‌های زیادی به سازمان شده و در واقع نوعی دوباره کاری در سازمان رخ می‌دهد و منابع سازمان اتلاف می‌شود. او در این تحقیق از روش توصیفی و از نوع پیمایشی استفاده کرده است. مطالعه موردی کلیه مدیران و کارکنان یکی از شرکت‌های بزرگ تحت پوشش بنیاد است (مصطفی قاضی زاده، ۱۳۸۷).

در تحقیقی که توسط یعقوبی و همکاران وی انجام شد، رابطه عدالت سازمانی با رضایت شغلی و تعهد سازمانی در بین کارکنان بیمارستان‌های منتخب دانشگاه علوم پزشکی اصفهان بررسی شده است. این بررسی به روش توصیفی - همبستگی صورت گرفته که نمونه مورد پژوهش آن، کارکنان بیمارستان‌های منتخب شهر اصفهان می‌باشند. ابزار جمع‌آوری داده‌ها پرسشنامه بوده و تحلیل داده‌ها با استفاده از نرم‌افزار SPSS انجام شده است. نتایج به دست آمده نشان می‌دهد که میزان عدالت سازمانی با اختلاف کمی در بیمارستان‌های خصوصی بیشتر است. البته از طرف دیگر میزان رضایت شغلی کارکنان بیمارستان خصوصی از بیمارستان دولتی بیشتر بود، ولیکن در بیمارستان‌های دولتی تعهد سازمانی بیشتری نسبت به بیمارستان خصوصی مشاهده گردید (مریم یعقوبی، سکینه سقاییان نژاد اصفهانی، ابوالقاسم گرگی، محسن نوروزی، فاطمه رضایی، ۱۳۸۸).

در تحقیقی که توسط حسینی و همکاران انجام شد، موضوع تاثیر تصویر ذهنی برند بر وفاداری کارکنان مورد ارزیابی قرار گرفت. در تحقیق ایشان پرسشنامه‌ای شامل ۴۵ سؤال، برگرفته از منابع که برای شرایط بومی طراحی شده بود، در سه برند برتر سال ۸۷ کشور در زمینه مواد غذایی، مورد آزمون قرار گرفته است. پایایی پرسشنامه از روش آلفای کرونباخ و روایی آن نیز با نظر کارشناسان، از طریق روایی محتوا صورت گرفته است. ۱۶۰ پرسشنامه در میان جامعه آماری توزیع، و داده‌ها با استفاده از، تحلیل مسیر و تکنیک مدل‌یابی معادلات ساختاری مورد تجزیه و تحلیل قرار گرفت. یافته‌های تحقیق نشان می‌دهد که، تصویر ذهنی برند، منجر به افزایش رضایت و در نهایت وفاداری کارکنان می‌گردد (میرزا حسن حسینی، وحید رضا حلوائی‌ها، سعید رضانی، ۱۳۸۹).

قره‌چه و همکاران در مطالعه دیگری به بررسی وفاداری کارکنان در تعامل با وفاداری مشتریان صنایع خدماتی پرداختند. در مطالعه ایشان، کیفیت خدمات و رضایت مشتری به عنوان متغیرهای اثرگذار در نظر گرفته شده و وفاداری کارکنان نیز متغیر وابسته بوده است. همچنین ایشان از روش معادلات ساختاری در تحقیق خود استفاده کرده‌اند و جامعه آماری آن را فروشگاه‌های زنجیره‌ای شهروند و رفاه در نظر گرفتند. نتایج به دست آمده نشان می‌دهد که وفاداری کارکنان و وفاداری مشتری، قویاً به هم وابسته‌اند (قره‌چه و همکاران، ۱۳۹۰).

رفعتی و همکاران در سال ۱۳۹۱، پژوهشی درباره بررسی تأثیر وفاداری بر خرید مجدد مشتریان فرآورده‌های لبنی پرداخته بودند که متغیرهای اثرگذار آن اعتماد مشتریان، درگیری ذهنی، تعهد و تبلیغات شفاهی بوده و همچنین متغیر وابسته آن رفتار خرید مجدد است و در پژوهش خود از روش معادلات ساختاری استفاده کرده است و همچنین جامعه آماری آن را مشتریان فروشگاه رفاه شهر تهران تشکیل داده است و نتیجه به دست آمده حاکی از آن است که اعتماد مشتریان، درگیری ذهنی، تعهد و تبلیغات شفاهی بر رفتار خرید مجدد آنان تأثیر مستقیم ندارد ولی رضایت مشتری بر خرید مجدد مشتریان تأثیر مستقیم دارد. همچنین درگیری ذهنی مشتریان مستقیماً بر رضایتمندی آن‌ها و از طرف دیگر رضایت مشتریان بر تبلیغات شفاهی، اعتماد و تعهد تأثیر مستقیم دارد که از این میان، رضایت بیشترین تأثیر را دارد. (رفعتی و همکاران، ۱۳۹۱)

در تحقیقی که توسط کرمپور و همکاران وی در مورد بررسی و اولویت‌بندی علل و عوامل ترک خدمت کارکنان شرکت پتروشیمی مبین در پارس جنوبی انجام شده است. این پژوهش تلاش کرده است تا مهمترین عواملی که موجب ترک خدمت کارکنان شده‌اند، شناسایی کند و با درک دقیق و اولویت‌بندی آنها، پیشنهادهایی برای حذف یا کاهش آنها ارائه دهد. و در این پژوهش از روش توصیفی استفاده شده است. یافته‌های پژوهش نشان می‌دهد که عدم امنیت شغلی در اولویت اول، تفاوت در حقوق و مزایا در مرتبه دوم، عدم ارتقای مناسب در مرتبه سوم و کاهش تعهد سازمانی در مرتبه چهارم، در ترک خدمت کارکنان تأثیرگذار است. (عبدالحسین کرمپور، محمد صفری کهره، حیدر احمدی، الهه اسدالهی دهکردی، ۱۳۹۲).

در تحقیقی که توسط بوژمهرانی و همکاران در مورد پیش‌بینی وفاداری کارکنان از طریق تصویر ذهنی برند مورد مطالعه قرار گرفت، متغیر اثرگذار، تصویر ذهنی برند و متغیر وابسته وفاداری کارکنان در نظر گرفته و از روش تحلیل همبستگی استفاده شد، همچنین جامعه آماری را آکادمی ملی المپیک و پارالمپیک تشکیل داد. نتایج این آزمون نشان داد که همبستگی مثبت و معناداری بین تصویر ذهنی برند و وفاداری کارکنان وجود دارد (بوژمهرانی و همکاران، ۱۳۹۶).

در تحقیق دیگری که توسط اسماعیلی و صیدزاده در مورد تأثیر رضایت شغلی بر عملکرد با نقش میانجی وفاداری سازمانی و با روش معادلات ساختاری مورد مطالعه قرار گرفت، متغیر اثرگذار رضایت شغلی و متغیر وابسته وفاداری سازمانی و جامعه آماری سازمان جهاد کشاورزی ایلام بود. نتیجه نشان می‌دهد که رضایت شغلی به طور مستقیم با ضریب مسیر ۰.۵۳ و به طور غیرمستقیم با نقش واسطه‌ای وفاداری سازمانی با ضریب مسیر ۰.۶۲ بر عملکرد شغلی تأثیر مثبت دارد. بعلاوه، رضایت شغلی با ضریب مسیر ۰.۷۴ بر وفاداری سازمانی و وفاداری سازمانی نیز با ضریب مسیر ۰.۷۶ بر عملکرد شغلی تأثیر مثبت دارد (اسماعیلی و صیدزاده، ۱۳۹۷).

ایجاد وفاداری در دراز مدت در پایداری کسب و کارها بسیار مهم است و تعیین ارزش وفاداری می‌تواند به شرکت‌ها در توسعه برنامه‌های راهبردی کمک کند. اگرچه مطالعات و پیشنهادهای جهت پیش‌بینی و کمی کردن وفاداری پیشنهاد

شده است اما در آن‌ها وضعیت سازمان و تاثیر آن در میزان وفاداری در نظر گرفته نشده است. هدف از این مقاله ارایه مدلی جهت پیش بینی وفاداری بر پایه وضعیت سازمان می باشد. در این مدل از سیستم های یادگیر و تکنیک های داده کاوی جهت کمی کردن وضعیت سازمان و متعاقبا مدل کردن وفاداری به همراه پارامتر های موثر دیگر استفاده شده است. مدل پیشنهادی بر روی داده های جمع آوری شده از کارمندان بانک های خصوصی و دولتی شهر تبریز اعمال شده است. نتایج به دست آمده نشان دهنده دقت مدل پیشنهادی و تاثیر وضعیت سازمان در پیش بینی وفاداری است (آقالاری، صمصام و امید سجودی شیجانی، ۱۳۹۵).

با در اختیار داشتن اطلاعات کافی از گروه های مختلف و خوشه بندی و طبقه بندی آنان می توان فعالیت هایی از جمله برنامه ریزی بلند مدت را به شکلی هدفمند طراحی و اجرا نمود که سازمان تا حد امکان به بهره وری برسد. هدف از انجام این پژوهش، خوشه بندی افراد به گروه های وفادار و غیر وفادار می باشد. داده های مورد بررسی در این پژوهش به اطلاعات ۱۰۳۰۰ نفر از اعضای بانک ملت مربوط می شود. روش پژوهش، توصیفی پیمایشی است و برای تحلیل اطلاعات از داده کاوی استفاده شده است که با توجه به اهداف پژوهش و بررسی الگوریتم های مختلف موجود در این تکنیک، از الگوریتم های C5.0 و Interactive CHAID استفاده شده است. (کیماسی مسعود، اسفیدانی محمدرحیم، اسماعیلیان امیربهن، ۱۳۹۵)

نظریه های متفاوتی انگیزه افراد از مشارکت در خلق مشترک ارزش را بررسی کرده اند و در تمام این پژوهش ها درباره اهمیت و سودمندی خلق مشترک ارزش به وسیله مشتریان توافق نظر وجود دارد. باوجود توافق درباره اهمیت و سودمندی خلق مشترک ارزش با مشتریان، همچنان کمبود دانش راجع به چرایی و چگونگی مشارکت افراد، انگیزه های آن ها در خلق مشترک ارزش در سازمان ها و وجود نظریه های مختلف برای تبیین این فرایند دیده می شود. تحقیق حاضر با بررسی انگیزه های اولیه مشارکت مصرف کنندگان در فرایند خلق مشترک با توجه به دیدگاه های مختلف و شناسایی مدل های رایج برای جلب مشارکت و خلق مشترک ارزش، در جست و جوی مدل مناسب در فرایند خلق مشترک ارزش است. برای

دستیابی به مدل، در این تحقیق از روش فراترکیب استفاده شده و مقالات مرتبط در بازه زمانی ۲۰۰۶ تا ۲۰۱۷ شناسایی و بررسی شده‌است. پس از استخراج مفاهیم و موضوعات مرتبط با مشارکت در خلق مشترک ارزش، ۶۶ کد گزینشی و ۱۳ شاخص شناسایی شد. این شاخص‌ها شامل انگیزه‌های رفتاری، هیجانی، منافع مالی، اجتماعی، ریسک ادراک‌شده، فنی، جمع‌گرایی، عملکرد ادراک‌شده، یادگیری اجتماعی، عوامل اطلاعاتی، هدفمندی شخصی، روانی و انتظار مالی هستند. هم‌چنین با استفاده از روش مدل‌سازی ساختاری تفسیری این عوامل در چهار سطح دسته‌بندی شدند. نتایج نشان می‌دهد انگیزه‌های روانی، اجتماعی و هیجانی بیشترین تأثیرگذاری را روی سایر انگیزه‌ها دارند؛ به گونه‌ای که ده انگیزه دیگر به صورت مستقیم یا غیرمستقیم از این ویژگی‌ها تأثیر می‌پذیرند. (امیرالسادات، نسیم، شفيعی، حاجی پور، سلطانی، ۱۳۹۸)

۲-۳-۲) مطالعات خارجی

تحقیقی دیگر توسط تستا و مارک آر در سال ۲۰۰۱، در مورد رضایت شغلی و وفاداری سازمانی با استفاده از تکنیک معادلات ساختاری انجام شد، که متغیر اثرگذار آن وفاداری کارکنان و متغیر وابسته وفاداری مشتری است. جامعه آماری را کارمندان دو سازمان خدماتی تشکیل دادند. نتایج نشان می‌دهد بین رضایت شغلی و وفاداری سازمانی رابطه‌ای بحث‌انگیز وجود دارد که کزالسکی^۱ (۱۹۸۰)، کاسپی و لازار^۲ (۱۹۹۱) و ماتیو زاجاک^۳ (۱۹۹۰) نیز به تفصیل به این رابطه اشاره نموده و معتقدند که بین رضایت شغلی و وفاداری سازمانی رابطه وجود دارد (تستا و مارک آر^۴، ۲۰۰۱).

در مقاله دیگری که توسط الیزابت و همکاران وی در سال ۲۰۱۳، در مورد رضایت کارکنان مجارستانی و وفاداری آن‌ها انجام شد و داده‌های بیش از ۶۰۰ کارمند هتل در سراسر مجارستان با استفاده از پرسشنامه جمع‌آوری شد که مؤلفه‌های اصلی آن در رضایت شغلی که شامل (رضایت شخصی، مشارکت اجتماعی، پیشرفت شغلی و حقوق و مزایا) و مؤلفه‌های اصلی وفاداری کارکنان (تعهد، شرایط شغلی، مزایای شخصی، عنصر خدمات و مکان، و شغل و وضعیت)، است. هم‌چنین

^۱ . Kalsalki

^۲ . Caspian & Lazar

^۳ . Mathiew&zajac

^۴ . Testa& M. R

بررسی روابط متقابل نشان داد که رفتار مدیران با کارمندان و مزایای مثبت درگیری اجتماعی در محیط کار تأثیر بیشتری بر وفاداری کارمندان نسبت به پاداش‌های پولی دارد (الیزابت م. اینسون^۱، استر بنک^۲، جوزف لازلو^۳، ۲۰۱۳).

در پژوهش مورد مطالعه رابرت و گرینبرگ که بر وفاداری کارکنان و عملکرد بوده است. آن‌ها در این پژوهش متغیر اثرگذار را وفاداری کارکنان و متغیر وابسته را کیفیت خدمات در نظر گرفته‌اند و از روش همبستگی اسپیرمن استفاده کرده‌اند و جامعه آماری آن را کارکنان شرکت تشکیل داده است. نتیجه نشان می‌دهد که افزایش وفاداری سازمانی باعث عملکرد بهتر افراد خواهد شد و نهایتاً از رضایت شغلی بالاتری هم برخوردار شوند (رابرت بارون و جerald گرینبرگ^۴، ۲۰۱۴).

در مقاله‌ای که توسط تورهان و سنگر در مورد سطح شناسایی عملکرد معلمان بر اساس تعهد سازمانی و میزان رضایت شغلی آن‌ها پیش‌بینی شده است. در این پژوهش پیش‌بینی با روش‌های محبوب K-NN انجام می‌شود. این مجموعه داده‌ها از ۱۲۶ معلم جمع‌آوری شده است که در آن از ۱۴ مقدار و ۱۱ سطح شناسایی عملکرد در کارهای تجربی استفاده شده است. در آثار تجربی ۱۴ مقدار به‌عنوان ورودی استفاده می‌شود و از هر سطح شناسایی عملکرد به‌عنوان خروجی استفاده می‌شود؛ بنابراین، آزمایش‌ها ۱۱ بار تکرار می‌شوند. نتایج نشان می‌دهد که بهترین پیش‌بینی برای سطح شناسایی اقدامات ششم انجام می‌شود و بدترین دقت در سطح شناسایی عملکرد دهم ثبت می‌شود.

به‌طور کلی، رویکرد K-NN خوب بدترین نمره را به دست می‌آورد و روش K-NN متوسط بهترین امتیازات را ایجاد می‌کند. نتایج همچنین نشان می‌دهد که داده‌کاوی می‌تواند به‌عنوان ابزاری مناسب برای تعیین سطح شناسایی عملکرد مورد استفاده قرار گیرد (محمد تورهان، سنگر ۲۰۱۸).

در تحقیق تانگ یئو و همکاران که در مورد حفظ کارکنان هتل به‌عنوان مشتریان داخلی و تأثیر تعهد سازمانی بر نگرش و رفتار وفاداری کارکنان مورد مطالعه قرار گرفته است. متغیر تأثیرگذار آن را وفاداری کارکنان و متغیر وابسته را تعهد سازمانی تشکیل داده است. تکنیک مورد استفاده معادلات ساختاری است و جامعه آماری آن هتل‌های شهرهای چین

¹. Ineson, Elizabeth M

². Ester benke

³. Jozsef lazlo

⁴. Robert B & gerinberg

است. نتیجه نشان می‌دهد ایجاد تعهد عاطفی در کارکنان موجب افزایش وفاداری نگرشی و رفتاری در کارکنان می‌شود (تانگ یئو همکاران^۱، ۲۰۱۹).

مطالعه دیگری توسط بهاتی و همکاران نقش واسطه‌ای اعتماد عاطفی و شناختی و نقش تعدیل کننده تعهد مداوم بر رهبری مشارکتی و روابط شهروندی سازمانی را بررسی می‌کند. چهارصد کارمند از صنعت هتلداری در پاکستان استخدام شدند. از روش بوت استرپینگ برای برآورد اثر میانجیگری توسط ماکرو فرآیند استفاده شد. این مطالعه از تحلیل عاملی تأییدی و مدل سازی معادلات ساختاری استفاده می‌کند. (حیات بهاتی^۲ و همکاران، ۲۰۱۹)

بر اساس تحقیق دیگری توسط فاروخ، در هر صنعت، وفاداری کارمندان از جمله عوامل اصلی رشد و پایداری تجارت محسوب می‌شود، بنابراین، بررسی عواملی که می‌توانند بر وفاداری کارکنان تأثیر بگذارند، به یکی از زمینه‌های تحقیقاتی تبدیل شده است و بنابراین هدف از این مطالعه بررسی تأثیر رضایت اعتماد و پشتیبانی رهبری از وفاداری کارمندان در صنعت هتلداری عربستان سعودی- سایپرز، داده‌ها با کمک پرسشنامه ساختار جمع آوری شد. یافته‌های تجزیه و تحلیل داده‌ها نشان داد که کل فرضیه پذیرفته شده است. در واقع این تحقیق به بررسی رابطه کلی اعتماد متقابل، رضایت شغلی، حمایت رهبر و وفاداری کارکنان پرداخته است. نتایج نشان می‌دهد که وفاداری کارکنان تحت تأثیر عناصر رضایت شغلی قرار دارد. هرچند اعتماد متقابل تأثیر کمی بر وفاداری کارمندان دارد. سهم دوم نظری این است که حمایت رهبر برای ایجاد وفاداری کارمندان قابل توجه است. وفاداری کارمندان مستقیماً با این عوامل ارتباط دارد. این سهم نظری از آن جهت قابل توجه است که یک رهبر نماینده سازمان است و نقشی حیاتی در پیشرفت شرکت دارد، این شامل حرکت نیروی انسانی است. این سهم نظری به سازمان‌ها امکان می‌دهد تا تفاوت بین نقش رهبران و نقش کارکنان برای پایداری سازمانی را دریابند (فاروخ^۳ و همکاران، ۲۰۱۹).

¹. Tang yao et al

². Hayat Bhatti et al

³. Farrukh, M. et al

تحقیق دیگری به بررسی تأثیرات مستقیم و غیرمستقیم پاداش، شخصیت و رضایت شغلی بر تعهد سازمانی معلم در دبیرستان‌های حرفه ای دولتی در شهر بوگور، جاوه غربی، اندونزی می‌پردازد. داده‌های این مطالعه با استفاده از یک پرسشنامه از طریق یک نظر سنجی، با استفاده از نمونه متناسب تصادفی انتخاب شده از ۱۵۴ معلم، جمع آوری شد. برای تجزیه و تحلیل داده ها از تحلیل مسیر پشتیبانی شده توسط آمار توصیفی استفاده شد. نتایج تأیید می‌کند که پاداش، شخصیت و رضایت شغلی تأثیر مستقیم مستقیم بر تعهد سازمانی داشتند. پاداش و شخصیت تأثیر مستقیمی بر رضایت شغلی داشتند. پاداش و شخصیت با اندازه گیری رضایت شغلی تأثیرات غیرمستقیم قابل توجهی بر تعهد سازمانی داشتند (ویدودو^۱ و دامایانتی^۲، ۲۰۲۰).

بر اساس نتایج تحقیق ساهنی، ارزشی که کیفیت زندگی کاری برای محل کار به ارمان می آورد مدت ها است که توسط دانشمندان مشخص شده است، با این حال، تأثیر آن بر میزان تعامل و تعهد کارکنان تحت بررسی است. مطالعه تجربی حاضر به بررسی کیفیت زندگی کاری در صنعت مخابرات و بررسی ارتباط آن با مشارکت کارمندان و تعهد سازمانی اختصاص یافته است. داده های جمع آوری شده از یک نمونه تصادفی از ۳۱۲ کارمند شاغل در صنعت مخابرات عربستان سعودی، از طریق آمار توصیفی و مدل سازی معادلات ساختاری مورد تجزیه و تحلیل قرار گرفت. شواهد از رابطه معنادار کیفیت زندگی کاری و تعهد سازمانی پشتیبانی می‌کند، با این وجود، رابطه ضعیفی از کیفیت زندگی کاری با مشاغل کارکنان وجود دارد. این اولین مطالعه برای بررسی رابطه بین سه سازه مهم محل کار است. کیفیت زندگی کاری، تعهد سازمانی و درگیری کارکنان. بنابراین، مطالعه به طور قابل توجهی به ادبیات موجود در مورد کیفیت زندگی کاری کمک می‌کند (ساهنی^۳، ۲۰۱۹).

هدف از این مطالعه دیگری توسط لی و همکاران، بررسی (الف) بررسی تأثیر مدیریت تنوع نیروی کار بر تطابق شغلی، رضایت شغلی و عملکرد شغلی است. (ب) نقش تأثیرگذاری تطابق شغلی بر رضایت شغلی و عملکرد شغلی و (ج) بررسی نقش واسطه ای شغل یک شخص در ارتباط بین مدیریت تنوع نیروی کار و نتایج کارکنان (رضایت شغلی و عملکرد

² Widodo, W.

³ Damayanti, R.

¹ Sahni, J.

شغلی) است. پرسشنامه های ساختاری به کارمندان شاغل در هتل های پنج ستاره چین ارسال شده است. در مجموع، ۳۲۴ پاسخ معتبر مورد تجزیه و تحلیل قرار گرفت. نتایج نشان داد که مدیریت تنوع نیروی کار با تطابق شغلی، رضایت شغلی و عملکرد شغلی فرد رابطه مثبت دارد. در مرحله بعدی، تطابق شغلی یک فرد رابطه مثبتی با رضایت شغلی و عملکرد شغلی دارد، به ویژه تطابق شغلی یک شخص واسطه رابطه بین مدیریت تنوع نیروی کار و نتایج کارمندان است. بیشتر مطالعات در زمینه مدیریت تنوع نیروی کار بر مدیریت تنوع از جمله سن، جنس، نژاد و قومیت از دیدگاه آمریکا متمرکز بود. این می تواند از جمله مطالعات نادری باشد که جنبه دیگری از مدیریت تنوع نیروی کار را بررسی می کند، مانند مدیریت تنوع بر اساس مهارت ها، دانش، علاقه و ترجیحات کارکنان از دیدگاه چینی (لی و همکاران^۱، ۲۰۲۰).

۳-۳-۲ خلاصه پیشینه تحقیقات داخلی در حوزه وفاداری کارکنان

جدول (۱-۲) خلاصه پیشینه تحقیقات داخلی در حوزه وفاداری کارکنان

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۱	مدنی و زاهدی (۱۳۸۴)	حمایت سازمانی	تعهد سازمانی ^۲	تحلیل عاملی و رگرسیون	پالایشگاه گاز فجر و بید بلند
۱	از میان عوامل مؤثر بر تعهد سازمانی، متغیرهایی چون درک حمایت سازمانی قوی ترین رابطه مستقیم و مثبت و متغیر احساس عدالت سازمانی همبستگی نسبتاً قوی با تعهد سازمانی دارد. متغیر احساس امنیت شغلی همبستگی ضعیفی با تعهد سازمانی دارد.				
۲	غفوری و گل پرور (۱۳۸۷)	عدالت سازمانی	تعهد سازمانی	تحلیل رگرسیون	کارکنان شهروندی شهر اصفهان
۳	نتایج حاکی از آن بود که عدالت سازمانی و سه مؤلفه آن همبستگی مثبت و معنی داری با هر یک از حیطه های تعهد سازمانی یعنی تعهد سازمانی عاطفی، تعهد سازمانی هنجاری و تعهد سازمانی مستمر داشتند.				

^۲ Li, W. et al

^۲. تعهد سازمانی یک نگرش درباره وفاداری کارمندان به سازمان و یک فرایند مستمری است که از طریق آن اعضای سازمان علاقه خود را به سازمان و موفقیت و کارایی پیوسته آن نشان می دهند.

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۳	قاضی زاده (۱۳۸۷)	وفاداری کارکنان	عوامل انگیزشی محیط	توصیفی - پیمایشی	شرکت تحت پوشش بنیاد
توضیحات	کارکنان وفادار در طول سال‌های متمادی برای سازمان فعالیت می‌کنند و با سیاست‌ها و رسالت سازمان آشنا هستند و نیازی به آموزش ندارند. در مقابل، تغییر کارکنان و عدم وفاداری کارکنان به سازمان، موجب وارد آمدن هزینه‌های زیادی به سازمان می‌شود.				
۴	یعقوبی و همکاران (۱۳۸۸)	عدالت سازمانی	رضایت شغلی تعهد سازمانی	توصیفی - همبستگی	بیمارستان علوم پزشکی اصفهان
توضیحات	میزان عدالت سازمانی با اختلاف کمی در بیمارستان‌های خصوصی بیشتر و میزان رضایت شغلی کارکنان بیمارستان خصوصی از بیمارستان دولتی بیشتر است. ولیکن در بیمارستان‌های دولتی تعهد سازمانی بیشتری نسبت به بیمارستان خصوصی مشاهده گردید.				
۵	حسینی و همکاران (۱۳۸۹)	تصویر ذهنی برند	وفاداری کارکنان رضایت کارکنان	معادلات ساختاری	سه برند برتر مواد غذایی ۸۷
توضیحات	یافته‌های تحقیق نشان می‌دهد که، تصویر ذهنی برند، منجر به افزایش رضایت و درنهایت وفاداری کارکنان می‌گردد.				
۶	قره‌چه و همکاران (۱۳۹۰)	کیفیت خدمات رضایت مشتری	وفاداری کارکنان	معادلات ساختاری	فروشگاه زنجیره‌ای رفاه
توضیحات	وفاداری کارکنان و وفاداری مشتری قویاً به هم وابسته‌اند.				
۷	رفعتی و همکاران (۱۳۹۱)	اعتماد مشتری تعهد تبلیغات شفاهی رضایت مشتری	وفاداری مشتریان	معادلات ساختاری	مشتریان فروشگاه رفاه تهران
توضیحات	اعتماد مشتریان، درگیری ذهنی، تعهد و تبلیغات شفاهی بر رفتار خرید مجدد آنان تأثیر مستقیم ندارد ولی رضایت مشتری بر خرید مجدد مشتریان تأثیر مستقیم دارد. همچنین درگیری ذهنی مشتریان مستقیماً بر رضایتمندی آن‌ها و از طرف دیگر رضایت مشتریان بر تبلیغات شفاهی، اعتماد و تعهد تأثیر مستقیم دارد که از این میان، رضایت بیشترین تأثیر را دارد.				

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۸	کرمپور و همکاران (۱۳۹۲)	استرس شغلی، پرداخت حقوق	ترک خدمت	روش توصیفی	کارکنانی که از سال ۸۰ تا ۸۵ ترک خدمت کرده‌اند
نتیجه	نتیجه نشان می‌دهد که عدم امنیت شغلی در اولویت اول، تفاوت در حقوق و مزایا در مرتبه دوم، عدم ارتقای مناسب در مرتبه سوم و کاهش تعهد سازمانی در مرتبه چهارم، در ترک خدمت کارکنان تأثیرگذار است.				
۹	آقالاری و همکاران (۱۳۹۵)	وضعیت سازمان	میزان وفاداری	تکنیک‌های داده کاوی	بانک‌های خصوصی و دولتی شهر تبریز
نتیجه	هدف از این مقاله ارایه مدلی جهت پیش بینی وفاداری بر پایه وضعیت سازمان می باشد. در این مدل از سیستم های یادگیر و جهت کمی کردن وضعیت سازمان و متعاقبا مدل کردن وفاداری به همراه پارامتر های موثر دیگر استفاده شده است. نتایج به دست آمده نشان دهنده دقت مدل پیشنهادی و تاثیر وضعیت سازمان در پیش بینی وفاداری است.				
۱۰	کیماسی و همکاران (۱۳۹۵)	رضایت شغلی	میزان وفاداری	CHAID، C5.0	بانک ملت
نتیجه	با در اختیار داشتن اطلاعات کافی از گروه‌های مختلف و خوشه‌بندی و طبقه بندی آنان می‌توان فعالیت هایی از جمله برنامه ریزی بلند مدت را به شکلی هدفمند طراحی و اجرا نمود که سازمان تا حد امکان به بهره وری برسد.				
۱۱	بوژمهرانی و همکاران (۱۳۹۶)	تصویر ذهنی برند	وفاداری کارکنان	تحلیل همبستگی	آکادمی ملی المپیک
نتیجه	این آزمون نشان داد که همبستگی مثبت و معناداری بین تصویر ذهنی برند و وفاداری کارکنان وجود دارد.				
۱۲	اسماعیلی و صید زاده (۱۳۹۷)	رضایت شغلی	وفاداری سازمانی	معادلات ساختاری	سازمان جهاد کشاورزی ایلام
نتیجه	رضایت شغلی به طور مستقیم با ضریب مسیر ۵۳ صدم و به طور غیرمستقیم با نقش واسطه‌ای وفاداری سازمانی با ضریب مسیر ۶۲ صدم بر عملکرد شغلی تأثیر مثبت دارد. بعلاوه، رضایت شغلی با ضریب مسیر ۷۴ صدم بر وفاداری سازمانی و وفاداری سازمانی نیز با ضریب مسیر ۷۶ صدم بر عملکرد شغلی تأثیر مثبت دارد.				

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۱۳	امیرالسادات و همکاران (۱۳۹۸)	انگیزه‌های رفتاری، روانی و انتظار مالی	مشارکت در خلق مشترک ارزش	مدل‌سازی ساختاری تفسیری	-
نتیجه	عوامل موثر در چهار سطح دسته‌بندی شدند. نتایج نشان می‌دهد انگیزه‌های روانی، اجتماعی و هیجانی بیشترین تأثیرگذاری را روی سایر انگیزه‌ها دارند؛ به گونه‌ای که ده انگیزه دیگر به صورت مستقیم یا غیرمستقیم از این ویژگی‌ها تأثیر می‌پذیرند.				
۱۴	فاطمه نظری (۱۳۹۸)	جمعیت شناختی	وفاداری کارکنان	تکنیک داده‌کاوی	کارکنان شرکت خصوصی توزیع برق خوزستان
نتیجه	نتیجه نشان داد که بیشترین ضریب تأثیر مربوط به متغیرهای وضعیت تاهل، جنسیت، محل اشتغال و سن استخدام بوده و کم اهمیت‌ترین متغیرها رشته تحصیلی، نوع استخدام، سمت و سنوات است.				

۲-۳-۴) خلاصه پیشینه تحقیقات خارجی در حوزه وفاداری کارکنان

جدول (۲-۲) خلاصه پیشینه تحقیقات خارجی در حوزه وفاداری کارکنان

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۱	تستا و مارک آر (۲۰۰۱)	وفاداری کارکنان	وفاداری مشتری	معادلات ساختاری	کارمندان دو سازمان خدماتی
نتیجه	بین رضایت شغلی و وفاداری سازمانی رابطه‌ای بحث‌انگیز وجود دارد و اشاره می‌کند که کزالسکی (۱۹۸۰)، کاسپی و لازار (۱۹۹۱) و ماتیو زاجاک (۱۹۹۰) نیز به تفصیل به این رابطه اشاره نموده و معتقدند که بین رضایت شغلی و وفاداری سازمانی رابطه وجود دارد.				
۲	الیزابت و همکاران (۲۰۱۳)	وفاداری کارکنان	رضایت شغلی	ضریب همبستگی پیرسون	۶۰۰ کارمند هتل در مجارستان
نتیجه	بررسی روابط متقابل نشان داد که رفتار مدیران با کارمندان و مزایای مثبت درگیری اجتماعی در محیط کار تأثیر بیشتری بر وفاداری کارمندان نسبت به پاداش‌های پولی دارد.				

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۳	بارون و گرینبرگ (۲۰۱۴)	وفاداری کارکنان	عملکرد شرکت	همبستگی اسپیرمن	کارکنان شرکت
۳	افزایش وفاداری سازمانی باعث می‌شود که افراد تلاش کنند و در نتیجه عملکرد بهتری ارائه دهند نهایتاً از رضایت شغلی بالاتری هم برخوردار شوند.				
۴	تورهان، سنگر (۲۰۱۸)	سطح شناسایی معلمان	تعهد سازمانی رضایت شغلی	داده کاوی	۱۲۶ معلم
۴	طرح پیشنهادی در پیش‌بینی سطوح شناسایی عملکرد کاملاً موفق است. به‌خصوص، برای برخی از سطوح شناسایی عملکرد، نمرات با دقت به‌دست‌آمده به ترتیب ۸۸.۱٪، ۸۹.۷٪ و ۹۳.۶٪ است که موفقیت ایده پیشنهادی را نشان می‌دهد.				
۵	تانگ یی و همکاران (۲۰۱۹)	وفاداری کارکنان	تعهد سازمانی	معادلات ساختاری	هتل‌های شهرهای چین
۵	از آنجاکه اعتماد و رضایت کارکنان در بخش هتل بسیار حیاتی است، ایجاد تعهد عاطفی در کارکنان موجب افزایش وفاداری نگرشی و رفتاری در کارکنان می‌شود.				
۶	بهائی و همکاران (۲۰۱۹)	اعتماد عاطفی و شناختی	تعهد مداوم بر رهبری مشارکتی	روش بوت استرپینگ	صنعت هتلداری در پاکستان
۶	نقش واسطه‌ای اعتماد عاطفی و شناختی و نقش تعدیل‌کننده تعهد مداوم بر رهبری مشارکتی و روابط شهروندی سازمانی را بررسی می‌کند. این مطالعه از تحلیل عاملی تأییدی و مدل سازی معادلات ساختاری استفاده می‌کند.				
۷	فاروخ و همکاران (۲۰۱۹)	رضایت اعتماد و پشتیبانی رهبری	وفاداری کارکنان	مدل سازی معادلات ساختاری	هتلداری عربستان سعودی- سایپرز
۷	نتایج نشان می‌دهد که وفاداری کارکنان تحت تأثیر عناصر رضایت شغلی قرار دارد. هرچند اعتماد متقابل تأثیر کمی بر وفاداری کارمندان دارد. سهم دوم نظری این است که حمایت رهبر برای ایجاد وفاداری کارمندان قابل توجه است. وفاداری کارمندان مستقیماً با این عوامل ارتباط دارد. این سهم نظری از آن جهت قابل توجه است که یک رهبر نماینده سازمان است و نقشی حیاتی در پیشرفت شرکت دارد، این شامل حرکت نیروی انسانی است. این سهم نظری به سازمان‌ها امکان می‌دهد تا تفاوت بین نقش رهبران و نقش کارکنان برای پایداری سازمانی را دریابند				
۸	ساهنی (۲۰۱۹)	کیفیت زندگی کاری	میزان تعامل و تعهد کارکنان	معادلات ساختاری	صنعت مخابرات عربستان سعودی

ردیف	محقق سال تحقیق	متغیر اثرگذار	متغیر وابسته	روش تحقیق	مطالعه موردی
۳		شواهد از رابطه معنادار کیفیت زندگی کاری و تعهد سازمانی پشتیبانی می‌کند، با این وجود، رابطه ضعیفی از کیفیت زندگی کاری با مشاغل کارکنان وجود دارد. این اولین مطالعه برای بررسی رابطه بین سه سازه مهم محل کار است. کیفیت زندگی کاری، تعهد سازمانی و درگیری کارکنان. بنابراین، مطالعه به طور قابل توجهی به ادبیات موجود در مورد کیفیت زندگی کاری کمک می‌کند.			
۹	ویدودو و دامایانتی (۲۰۲۰)	شخصیت و رضایت شغلی	تعهد سازمانی	تحلیل مسیر پشتیبانی	معلم در دبیرستان‌های حرفه ای دولتی در شهر بوگور، جاوه غربی، اندونزی
۳		نتایج تأیید می‌کند که پاداش، شخصیت و رضایت شغلی تأثیر مستقیم مستقیم بر تعهد سازمانی داشتند. پاداش و شخصیت تأثیر مستقیمی بر رضایت شغلی داشتند. پاداش و شخصیت با اندازه گیری رضایت شغلی تأثیرات غیرمستقیم قابل توجهی بر تعهد سازمانی داشتند			
۱۰	لی و همکاران (۲۰۲۰)	مدیریت نیروی کار	عملکرد شغلی	تحلیل آماری	کارمندان شاغل در هتل‌های پنج ستاره چین
۳		نتایج نشان داد که مدیریت تنوع نیروی کار با تطابق شغلی، رضایت شغلی و عملکرد شغلی فرد رابطه مثبت دارد. در مرحله بعدی، تطابق شغلی یک فرد رابطه مثبتی با رضایت شغلی و عملکرد شغلی دارد، به ویژه تطابق شغلی یک شخص واسطه رابطه بین مدیریت تنوع نیروی کار و نتایج کارمندان است. بیشتر مطالعات در زمینه مدیریت تنوع نیروی کار بر مدیریت تنوع از جمله سن، جنس، نژاد و قومیت از دیدگاه آمریکا متمرکز بود.			

۴-۲) جمع‌بندی

با مطالعه مبانی نظری همچون مفهوم وفاداری دریافتیم که چه عواملی می‌تواند بر وفاداری کارکنان یک شرکت خصوصی تأثیر بگذارد وفاداری کارکنان به سازمان، احساس جهت‌گیری مثبت کارکنان در رابطه با مسیر آینده و توسعه سازمان است که بر عملیات جاری سازمان تأثیر می‌گذارد. حال در فصل‌های آینده با استفاده از تکنیک‌های پیش‌بینانه داده‌کاوی، وفاداری کارکنان تازه استخدام‌شده در یک سازمان خصوصی را پیش‌بینی می‌کنیم.

قلمرو موضوعی این طرح بررسی پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی است و می‌تواند در بخش استخدام کارکنان سازمان مورد استفاده قرار گیرد. این مطالعه از مهرماه ۱۳۹۷ شروع شده و پیش‌بینی

می‌شود تا پایان سال ۱۳۹۸ انجام شود. جامعه آماری تحقیق حاضر شامل همه کارکنان شرکت‌های خصوصی توزیع برق استان خوزستان می‌باشد که از ابتدای جذب تا پایان سال ۱۳۹۸ مورد مطالعه قرار گرفته است.

با توجه به اینکه در استفاده از تکنیک‌های داده‌کاوی، اندازه بیشتر نمونه موجب افزایش دقت می‌گردد، لذا کل اعضا جامعه تحقیق به صورت سرشماری مورد بررسی قرار می‌گیرد. جمع آوری داده‌های کارکنان از پرونده‌های پرسنلی جامعه مورد تحقیق انجام شده است.

فصل سوم:

روش‌شناسی پژوهش

۳-۱) مقدمه

فرایند جستجوی منظم برای تشخیص یک موقعیت نامعین را تحقیق می‌گویند؛ بنابراین تحقیق روالی است که به موجب آن به جستجو در ارتباط با ناشناخته‌ها می‌پردازیم و نسبت به آن‌ها شناخت لازم را کسب می‌کنیم. در حین اجرای این فرایند روش‌شناسی یعنی چگونگی گردآوری شواهد و تبدیل آن‌ها به یافته‌ها (سرمد و همکاران، ۱۳۸۹).

در این پژوهش سعی بر این است که با استفاده از مقایسه چند الگوریتم داده‌کاوی به ارائه بهترین مدل برای بررسی و پیش‌بینی وفاداری کارکنان پرداخته شود. در این فصل ابتدا با توجه به هدف تحقیق و نحوه جمع‌آوری داده‌ها، نوع و روش تحقیق تعیین می‌شود. جامعه آماری و حجم نمونه در چارچوب موردنیاز مشخص می‌شود، سپس روش و ابزار جمع‌آوری داده‌ها و روش تجزیه و تحلیل داده‌ها مورد بررسی قرار می‌گیرد. در پایان در مورد تکنیک داده‌کاوی بحث می‌شود.

۳-۲) روش تحقیق

ملاک تقسیم‌بندی تحقیق‌ها بر دو نوع است. اول بر اساس هدف سپس بر اساس نیاز داشتن به فرضیه. پژوهش‌های علمی بر اساس هدف به سه دسته تقسیم می‌شوند. بنیادی، کاربردی، تحقیق و توسعه. و بر اساس نیاز داشتن به فرضیه شامل تحقیقات اکتشافی و توصیفی و آزمایشی است. در تحقیقات بنیادی هدف اصلی، آزمون نظریه‌ها، تبیین روابط بین پدیده‌ها و افزودن به مجموعه دانش موجود در یک موضوع خاص است؛ لذا تحقیقات بنیادی به دنبال گسترش مجموعه دانسته‌های کشف شده درباره اصول و قوانین علمی است. اما در مطالعات تحقیق و توسعه، هدف اساسی، تدوین با تهیه برنامه‌ها، طرح‌ها و امثال آن است. در این گونه تحقیقات ابتدا موقعیت نامعین ویژه‌ای تعیین شده و بر اساس نتایج پژوهشی طرح تدوین می‌شود.

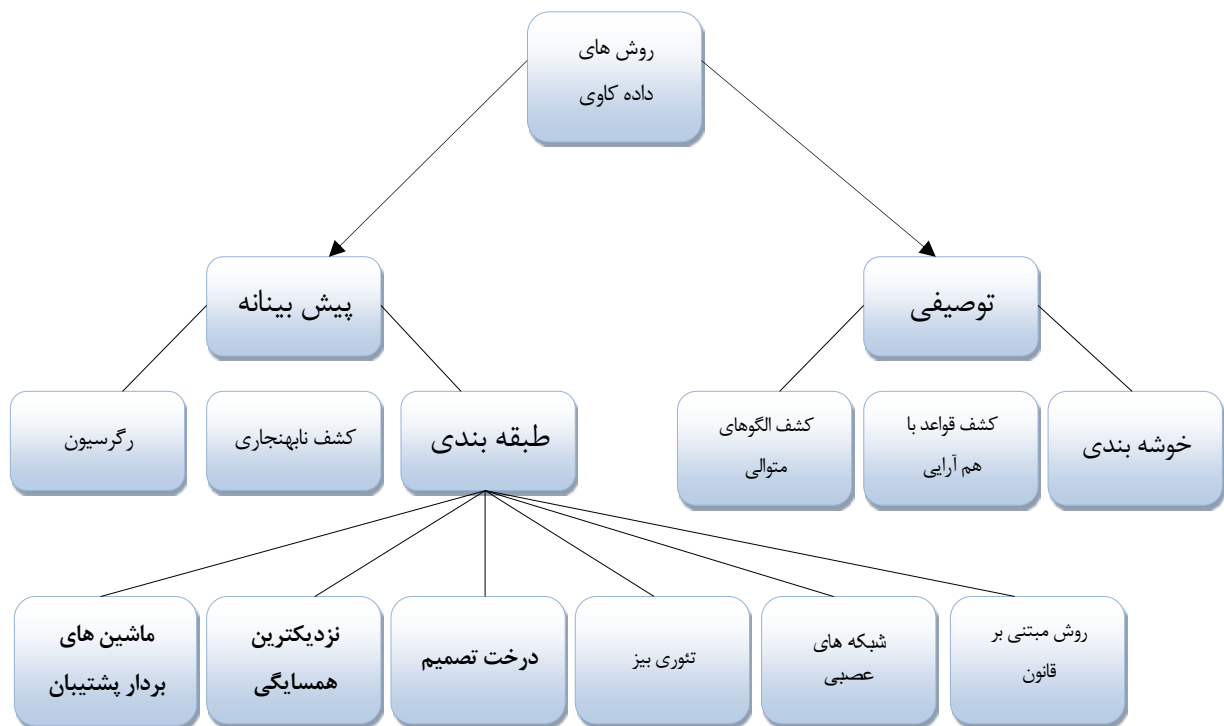
پژوهش پیش رو به دنبال بررسی و پیش‌بینی وفاداری کارکنان شرکت‌های خصوصی با استفاده از تکنیک داده‌کاوی است. این پژوهش از منظر هدف، کاربردی است. در پژوهش‌های کاربردی، نظریه‌ها، قانونمندی‌ها، اصول و فنونی که در تحقیقات پایه تدوین می‌شوند در مسائل اجرایی به کار گرفته می‌شوند. پژوهش‌های کاربردی با تکیه بر نتایج تحقیقات بنیادی و برای بهبود و ارتقاء رفتارها، روش‌ها، ابزارها، وسایل، تولیدات، ساختارها و الگوهای مورد استفاده جوامع انسانی اجرا می‌شوند؛ بنابراین این تحقیق در پی آن است که نتیجه تحقیق را برای حل مسائل اجرایی و واقعی به کار گیرد. این پژوهش از نظر ماهیت اکتشافی است. زمانی از مطالعه اکتشافی استفاده می‌شود که درباره وضع موجود یک پدیده اطلاعات چندانی وجود ندارد یا این که درباره چگونگی حل مسائلی که در گذشته پدید آمده‌اند و با وضعیت موجود شباهت دارند. اطلاعات درستی در دسترس نیست در این گونه موارد باید بررسی گسترده برای آشنایی با پدیده‌های مربوط به وضعیت مورد نظر صورت گیرد، تا بتواند مدلی را پدید آورد و برای تکمیل مطالعه طرح منسجمی را پیاده کرد. این گونه پژوهش‌ها درصدد پاسخ به چرایی‌ها یا توصیف یک پدیده در اجتماع است. در واقع مطالعات اکتشافی برای درک بهتر ماهیت مسئله‌ای انجام می‌گیرد که لازم باشد برای درک وضعیت و پدیده‌های آن مصاحبه‌های گسترده با افراد بسیار زیادی انجام شود. برخی از پژوهشگران این پژوهش را در طیف پژوهش‌های توصیفی قرار می‌دهند.

۳-۳) داده‌کاوی

درواقع داده‌کاوی عبارت است از فرایند یافتن و استخراج اطلاعات پنهان، الگوها و روابط مشخص در حجم انبوهی از داده‌ها با هدف پیش‌بینی رویدادها و نتایج آتی. داده‌کاوی، پایگاه‌ها و مجموعه حجیم داده‌ها را برای کشف و استخراج، مورد تحلیل قرار می‌دهد. در واقع داده‌کاوی فرآیندی خودکار برای بازیابی دانش به‌وسیله استخراج الگوهای مختلف است. داده‌کاوی به طور هم‌زمان از چندین رشته علمی نظیر تکنولوژی پایگاه داده، هوش مصنوعی، یادگیری ماشین، شبکه‌های عصبی، آمار، شناسایی الگو، سیستم‌های مبتنی بر دانش، بازیابی اطلاعات، و ... بهره می‌برد (دیوید ۲۰۰۱).

۳-۱-۳) تکنیک‌های مختلف داده‌کاوی

بر اساس نوع عملیات تکنیک‌های داده‌کاوی به دودسته «توصیفی» و «پیش‌بینانه» تقسیم می‌شوند. ساختار شکل (۳-۱) نمای کلی آن را نشان می‌دهد. داده‌کاوی توصیفی برای تعیین ویژگی‌های عمومی داده‌های یک پایگاه داده مورد استفاده قرار می‌گیرد. وظیفه داده‌کاوی پیش‌گوینه استنتاج داده‌های فعلی جهت پیش‌بینی در آینده است. رویکرد توصیفی شامل روش‌هایی چون خوشه‌بندی، کشف قواعد، کشف الگوها است و رویکرد پیش‌گوینه از روش‌هایی چون طبقه‌بندی (درخت تصمیم، شبکه‌های عصبی، تئوری بیز، روش مبتنی بر قانون، ماشین‌های بردار پشتیبان و نزدیک‌ترین همسایگی)، کشف ناپهنجاری و رگرسیون تشکیل شده است. با توجه به هرکدام از این روش‌های تکنیک داده‌کاوی ما روش پیش‌بینانه را برای کارکنان شرکت خصوصی توزیع برق مورد مطالعه قرار می‌دهیم.



شکل (۳-۱) تکنیک‌های داده‌کاوی

۳-۳-۲) طبقه‌بندی در داده‌کاوی

هدف از طبقه‌بندی مشخص کردن ویژگی‌هایی است که بتوان توسط آن، کلاس‌های مختلف را از یکدیگر متمایز کرد. طبقه‌بندی در داده‌کاوی طی دو مرحله انجام می‌گیرد. ابتدا از روی داده‌های قدیمی، کلاس‌های مختلف تشخیص داده شده و سپس تعلق داشتن داده‌های جدید به کلاس‌های موجود، پیش‌بینی می‌شود. طبقه‌بندی جزو تکنیک‌های یادگیری با ناظر است زیرا با در اختیار داشتن یک مجموعه داده آموزشی (به‌عنوان راهنما)، داده‌های جدید را طبقه‌بندی می‌کند. این روش جزو روش‌های پیش‌بینی‌کننده نیز به شمار می‌آید.

۳-۳-۳) مراحل داده‌کاوی

فرآیند داده‌کاوی دارای یازده مرحله نشان داده شده در شکل است:

۱) تبدیل موضوع به یک مسئله داده‌کاوی

۲) انتخاب داده‌های مناسب

۳) شناخت داده‌ها

۴) تهیه یک مدل اولیه

۵) تثبیت مسائل با توجه به داده‌ها

۶) تبدیل داده‌ها به اطلاعات مشخص

۷) ساخت مدل‌ها

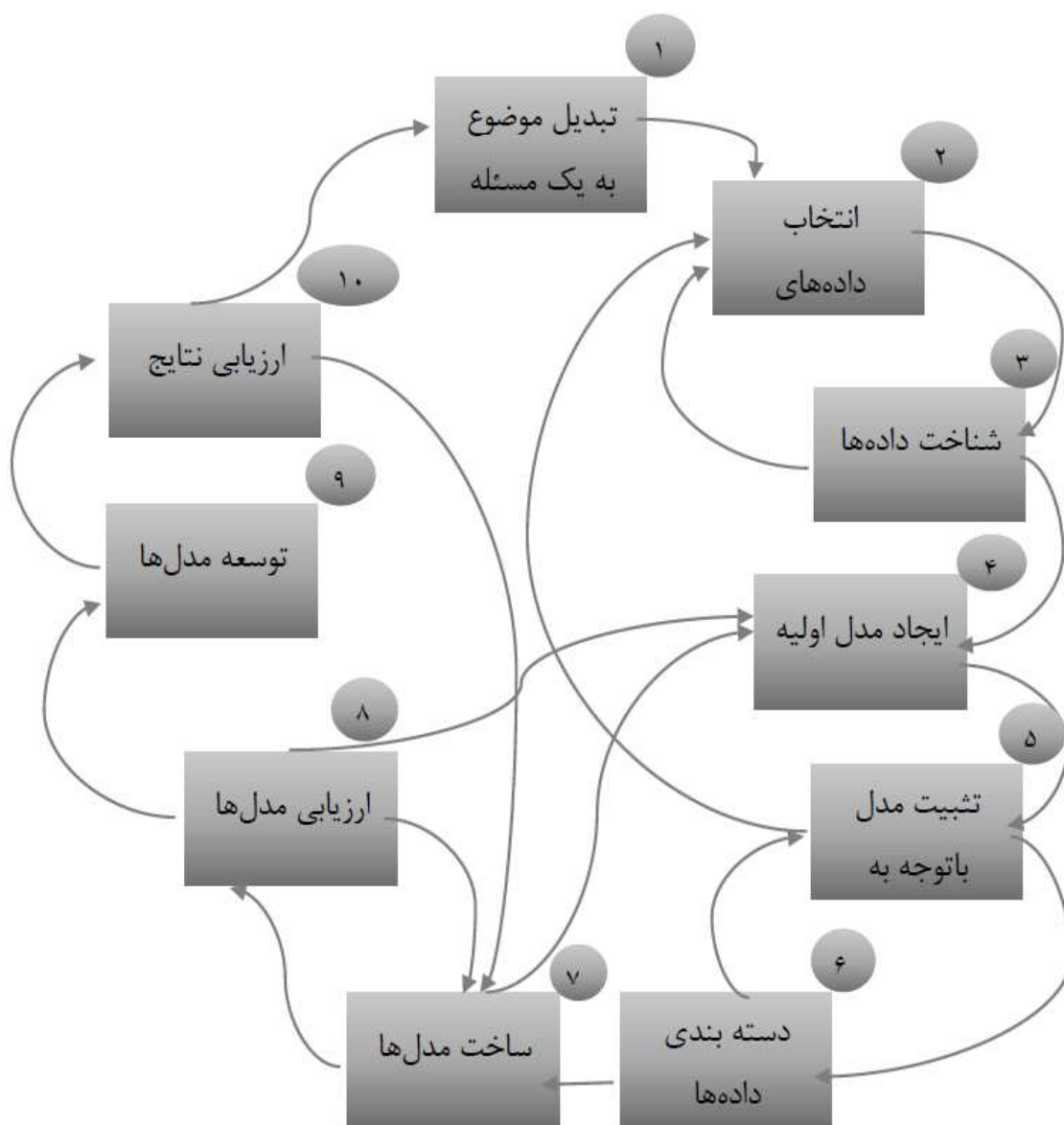
۸) ارزیابی مدل‌ها

۹) توسعه مدل‌ها

۱۰) ارزیابی نتایج

۱۱) شروع دوباره

گام یازدهم یک سیکل در فرآیند ایجاد می‌کند و فرآیند داده‌کاوی را به فرایند گردشی در حال توسعه تبدیل می‌کند در نتیجه همان‌طور که در شکل (۲-۳) نشان داده شده فرآیند داده‌کاوی به‌هیچ‌عنوان خطی نیست (لینوف ۲۰۱۱).



شکل (۲-۳) مراحل داده‌کاوی

در واقع داده‌کاوی در بسیاری از شاخه‌ها مانند بازاریابی، امور مالی مدیریت بحران، تحلیل شکست، بانکداری، مدیریت ارتباط با مشتری، بیمه، حمل‌ونقل، پزشکی، هوش مصنوعی، پیش‌بینی خرابی‌ها، و آموزش سازمانی کاربرد دارد (کاو ۲۰۰۵).

۳-۴) روش‌های طبقه‌بندی داده‌کاوی

در ادامه انواع روش‌های طبقه‌بندی داده‌کاوی که در این پژوهش برای بررسی و پیش‌بینی وفاداری کارکنان استفاده شده است. معرفی می‌شوند و شرح مختصری از هر یک داده می‌شود.

۳-۴-۱) درخت تصمیم

ساختار درخت تصمیم یک مدل پیش‌بینی‌کننده جهت نقش کردن حقایق مشاهده شده یک پدیده به استنتاج‌هایی در مورد مقدار هدف آن پدیده است و از رایج‌ترین روش‌های داده‌کاوی است.

در ساختار درخت تصمیم هر گره معادل یک متغیر و هر کمان یک فرزند است، فرزندها نشان دهنده مقادیر ممکن برای هر متغیر هستند. یک گره برگ یا انتهایی، با تعیین مقادیر متغیرهایی که در مسیر از ریشه درخت تا آن گره برگ مشخص می‌شود، ارزش پیش‌بینی‌شده متغیر مورد بررسی را نمایان می‌کند. در یک درخت تصمیم برگ‌ها نشانه دسته‌بندی و شاخه‌ها ترکیبات فصلی صفاتی هستند که در نهایت دسته‌بندی‌های مربوطه را برای ما می‌سازد. آموزش درخت با جداسازی مجموعه اصلی به زیرمجموعه‌ها و بر اساس آزمایش یک مقدار صفت برای هر متغیر انجام می‌شود و البته یک فرآیند بازگشتی است که در هر زیرمجموعه حاصل از تفکیک تکرار می‌شود. تکمیل عمل بازگشت زمانی است که تفکیک بیشتر اطلاعات بیشتری را برای یک دسته‌بندی جدید به همه نمونه‌های موجود در زیرمجموعه مورد بررسی اعمال نکند.

ساختار تصمیم‌گیری درخت تصمیم با عنوان تکنیک‌های ریاضی و محاسباتی برای توصیف و دسته‌بندی یک مجموعه از داده‌ها نیز معرفی می‌شوند. برای این منظور داده‌ها با ساختار $(x, y) = (x_1, x_2, x_3, \dots, x_k, y)$ داده می‌شوند و با استفاده از متغیرهای $x_1, x_2, \dots, x_k$ سعی در شناسایی و دسته‌بندی متغیر وابسته Y داریم. درخت تصمیم دارای دو نوع صفات دسته‌ای و حقیقی است. صفات دسته‌ای، دو یا چند مقدار گسسته یا سمبلیک می‌پذیرند درحالی‌که صفات حقیقی مقادیر حقیقی دریافت می‌کنند.

درختان تصمیم، ابزاری قدرتمند و پرتطرفدار برای دسته‌بندی و پیش‌بینی هستند. در حقیقت جذابیت و قدمت روش‌های مبتنی بر شاخه موجب تمایل به تکنیک‌های درخت تصمیم شده است. جداسازی مناسب داده‌ها در روش درخت تصمیم، یکی از مهم‌ترین مراحل استفاده از این روش شده است (پنگ ۲۰۰۴).

مدل‌های طبقه‌بندی متعددی در طی سال‌ها گزارش شده‌اند، به‌عنوان مثال، شبکه‌های عصبی، مدل‌های آماری مانند جداکننده‌های خطی/مربعی، درخت‌های تصمیم و مدل‌های ژنتیکی. در بین این مدل‌ها، درخت‌های تصمیم به‌ویژه برای داده‌کاوی مناسب می‌باشند. درخت‌های تصمیم در مقایسه با سایر روش‌ها می‌توانند نسبتاً سریع ساخته شوند. مزیت دیگر این است که مدل‌های درخت تصمیم برای فهم ساده و آسان می‌باشند. به‌علاوه، درخت‌ها می‌توانند به راحتی به عبارت‌های SQL تبدیل شوند که می‌تواند برای دسترسی مؤثر به پایگاه داده استفاده شوند. نهایتاً، درخت تصمیم در مقایسه با سایر روش‌ها صحتی مشابه و گاهی اوقات بهتر حاصل می‌کند. در نهایت درخت حاصل در انتهای الگوریتم دارای ویژگی‌های زیر می‌باشد:

- ❖ هر گره غیر از برگ با یک ویژگی که سؤالی در رابطه با مثال ورودی مطرح می‌کند، مشخص می‌شود.
- ❖ در هر گره غیر از برگ به تعداد جواب‌های ممکن برای این ویژگی شاخه با مقدار جواب مشخص دارد.
- ❖ برگ‌های درخت با برچسب کلاس برای یک دسته از جواب‌ها مشخص می‌شوند.
- ❖ گره ریشه، بالاترین گره در درخت تصمیم است.

در کل درختان تصمیم دارای مزایا و معایبی هستند. مزایای درختان تصمیم عبارت‌اند از:

۱. قابلیت به کارگرفتن و ایجاد مجموعه قوانین قابل دسترسی و درک.

۲. قابلیت استفاده از داده‌های پیوسته و گسسته.

۳. به کاربردن محدوده‌های اولیه اتخاذ تصمیم.

۴. حذف قیاس‌های غیر لازم.

۵. بهره از عوامل جداگانه برای نمونه‌های متمایز.

۶. عدم نیاز به تخمین تابع توزیع.
۷. عدم نیاز به آماده‌سازی داده‌های اولیه. (مثل نرمال‌سازی یا حذف مقادیر تهی)
۸. امکان توصیف ساده شرایط با منطق بولی.
۹. امکان نمایش قابلیت اطمینان مدل با آزمون‌های آماری.
۱۰. قابلیت تحلیل داده‌های بزرگ در زمان کوتاه.
۱۱. قابلیت کشف روابط غیرمنتظره یا نامعلوم.
۱۲. قابلیت شناسایی اختلاف در زیرگروه‌ها.

معایب درخت تصمیم عبارت‌اند از:

۱. نامناسب جهت تخمین تابع با مقادیر پیوسته.
۲. افزایش احتمال خطا با تعدد دسته‌ها و داده آموزشی کم.
۳. دارای هزینه محاسباتی بالا.
۴. دارای هزینه هرس کردن بالا.
۵. عملکرد نامناسب هنگام تفکیک نادرست یا هم‌پوشانی دسته‌ها.
۶. افزایش گره‌های پایانی در صورت هم‌پوشانی.
۷. انباشته شدن خطای لایه‌ها در درخت بزرگ.
۸. طراحی بهینه دشوار.
۹. امکان تولید روابط نادرست.
۱۰. بازنمایی دشوار.
۱۱. افزایش زمان جستجو و فضای حافظه با افزایش گره‌های پایانی به دلیل دسته‌های زیاد.

۴-۳-۱-۱) جذابیت درختان تصمیم

در ساختاری متفاوت با دسته‌بندی‌های تک مرحله‌ای در یک درخت، یک نمونه فقط روی زیرمجموعه‌های خاصی از دسته‌ها آزمایش می‌شود تا بدین ترتیب از محاسبات غیرضروری بپرهیزیم. همچنین برخلاف دسته‌بندی‌های تک مرحله‌ای، که معمولاً با یک معیار بهینه سراسری انتخاب می‌شوند برای جداسازی دسته‌ها، انعطاف‌پذیری در انتخاب زیرمجموعه به شکل بهینه که بهبودی در کارایی را ایجاد می‌کند، وجود دارد. هنگام تحلیل با تعداد صفات و دسته‌های زیاد، نیاز به تخمین توزیع‌های ابعاد یا پارامترهای خاصی از توزیع‌های دسته همانند احتمالات اولیه از یک مجموعه داده‌های آموزشی کوچک می‌باشد. در این حالت مشکل ابعاد بالا وجود دارد که امکان دارد در درخت دسته‌بندی کننده، با بکار بردن تعداد کمتری از صفات در هر گره داخلی بدون افت شدید کارایی، این مسئله حل شود.

۴-۳-۱-۲) چند الگوریتم درخت تصمیم

الگوریتم‌های متعددی برای ساخت درخت تصمیم وجود دارد که به برخی از آنها اشاره می‌شود:

۱) الگوریتم ID3

۲) الگوریتم C4.5

۳) الگوریتم CART

۱) الگوریتم ID3: این الگوریتم یکی از ساده‌ترین الگوریتم‌های درخت تصمیم است که از معیار Information Gain استفاده می‌کند. در اجرای این الگوریتم دو شرط توقف وجود دارد. یکی این که کلیه نمونه‌های باقیمانده متعلق به یک کلاس باشند و یا اینکه پس از محاسبه مقدار Information Gain بهترین آن بزرگ‌تر از صفر نباشند. این ساختار درخت روش هرس ندارد و قادر به دریافت صفات عددی و داده‌های ناقص به عنوان ورودی خواهد بود. فضای فرضیه ID3، یک فضای کامل از توابع مقدار گسسته متناهی، مربوط به صفات موجود است. این الگوریتم توانایی تعیین درختان تصمیم متعدد سازگار با داده‌های آموزشی یا ساخت پرسش‌های نمونه جدید که به شکل بهینه باعث انتخاب یک فرضیه از بین فرضیات رقیب می‌شوند را ندارد.

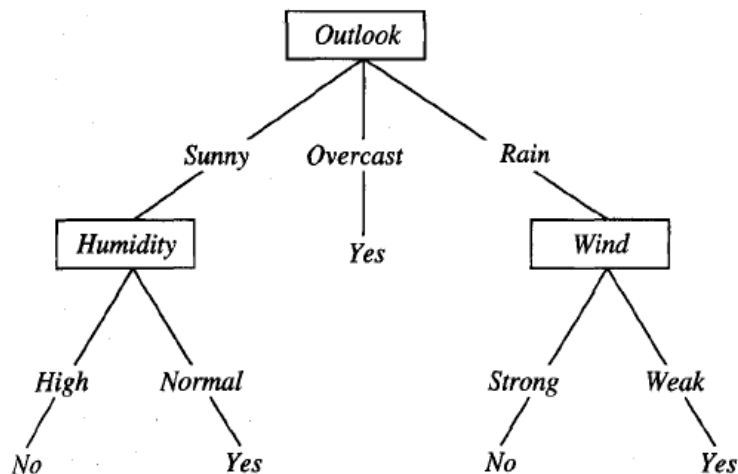
الگوریتم ID3 بازگشت به عقب ندارد و هیچگاه پس از انتخاب صفتی برای آزمایش، دوباره آن را بررسی نمی‌کند؛ بنابراین مستعد خطر جستجوی Hill-Climbing می‌باشد که یعنی همگرایی به راه‌حل‌های بهینه محلی که بهینه سراسری نمی‌باشند. همچنین الگوریتم ID3، تمام مثال‌ها را در هر گام از جستجو با این مزیت که جستجوی نهایی، کمتر به خطاهای داده‌های آموزشی حساس خواهد بود، استفاده می‌کند.

۲) الگوریتم C4.5: این الگوریتم یکی از تعمیم‌های الگوریتم ID3 است که از معیار Gain Ratio جهت انتخاب صفت خاصه استفاده می‌کند. الگوریتم هنگامی متوقف می‌شود که تعداد نمونه‌ها کمتر از مقدار مشخص شده‌ای باشد. این الگوریتم از تکنیک پس هرس استفاده می‌کند با این کار می‌توان مدل را به یک نقطه بهینه رساند که زیاد پیچیده نباشد. و همانند الگوریتم‌های قبلی داده‌های عددی را نیز می‌پذیرد با کمی تغییر هم می‌توان برای داده‌های ناقص از آن استفاده کرد. در این الگوریتم بحث وزن دهی به ویژگی‌ها است، یعنی به این نتیجه می‌رسند که با دادن وزن دهی بیشتر به یک ویژگی می‌تواند در انتخاب آن تاثیر بیشتری داشته باشد.

۳) الگوریتم CART: نتیجه این الگوریتم یک درخت تصمیم دودویی است. بدین معنی که هر گره داخلی به طور دقیق دارای دو انشعاب است. از معیار Gini Index استفاده می‌کند و روشی را نیز برای هرس کردن دارد. یکی از مهم‌ترین ویژگی‌های مهم Cart توانایی تولید درختان رگرسیون است. برگ‌ها در چنین درختی یک عدد واقعی را به جای برچسب کلاس تخمین می‌زنند.

۳-۴-۱) بازنمایی درخت تصمیم

درختان تصمیم، نمونه‌ها را با مرتب کردن آنها در درخت از گره ریشه به سمت گره‌های برگ دسته‌بندی می‌کنند. هر گره داخلی در درخت، صفتی از نمونه را آزمایش می‌کند و هر شاخه‌ای که از آن گره خارج می‌شود متناظر یک مقدار ممکن برای آن صفت می‌باشد. درختان تصمیم در واقع نمایش دهنده یک ترکیب فصلی از ترکیبات عطفی قیود روی مقادیر صفات نمونه‌ها را بازنمایی می‌کنند. برای مثال شکل (۳-۳) نمونه‌ای از یک درخت تصمیم متناظر با عبارت را نشان می‌دهد.



شکل (۳-۳) نمونه درخت تصمیم

۴-۱-۴-۳ معیارهای انتخاب صفت خاصه

اغلب هر گره داخلی در درخت تصمیم بر اساس مقدار یک صفت خاصه منشعب می‌شود. در نتیجه الگوریتم به دنبال بهترین انتخاب خود در میان صفات خاصه می‌گردد. بهترین معیارها برای انتخاب صفات خاصه وجود دارد که به چند مورد اشاره می‌کنیم.

❖ معیار Information Gain

این معیار یکی از معروف‌ترین معیارهایی است که برای ساخت درخت تصمیم از آن استفاده می‌شود که خود از معیار دیگری به نام آنتروپی^۱ استفاده می‌کند.

$$InformationGain(A) = Entropy(D) - Entropy_A(D) \quad (۱-۳)$$

این فرمول Information Gain را برای صفت خاصه A محاسبه می‌کند که D دلالت بر مجموعه داده آموزشی دارد:

$$Entropy(D) = - \sum_{i=1}^c P_i \times \log_2(P_i) \quad (۲-۳)$$

$$Entropy_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Entropy(D_j) \quad (۳-۳)$$

^۱. Entropy

C : تعداد برچسب کلاس‌های موجود در داده‌های آموزشی.

P_i : احتمال اینکه نمونه‌ای از داده‌ها متعلق به کلاس i ام باشد.

V : تعداد اعضای دامنه صفت خاصه A.

D_j : قسمتی از داده‌های اولیه که مقدار صفت خاصه آنها v_j است.

$|D|$: دلالت بر اندازه داده‌های D دارد.

صفت خاصه که بیشترین مقدار را داشته باشد به عنوان ریشه درخت تصمیم انتخاب می‌شود.

❖ معیار Gain Ratio

این معیار در واقع Information Gain را نرمال‌سازی می‌کند و به صورت زیر بیان می‌شود:

$$\text{GainRatio}_A(D) = \frac{\text{InformationGain}(A)}{\text{Entropy}_A(D)} \quad (4-3)$$

در صورتی که مخرج کسر مقدار صفر داشته باشد، این معیار قابل تعریف نمی‌باشد.

❖ معیار Gini index

جهت محاسبه این معیار برای داده‌های D از فرمول زیر استفاده می‌کنیم.

$$\text{Gini}(D) = 1 - \sum_{i=1}^c p_i^2 \quad (5-3)$$

C : تعداد کلاس‌های موجود در داده‌ها

P_i : احتمال تعلق نمونه‌ای از داده‌ها به کلاس i ام

اگر مجموعه داده D برای صفت خاصه A به دو زیر مجموعه D_1 و D_2 تقسیم شود پس داریم:

$$\text{Gini}_A(D) = \frac{|D_1|}{|D|} \times \text{Gini}(D_1) + \frac{|D_2|}{|D|} \times \text{Gini}(D_2) \quad (6-3)$$

برای هر یک از صفات خاصه، تمام حالت‌های دسته‌بندی دودویی در نظر گرفته می‌شوند. پس از محاسبه Gini Index برای همه حالات، مقدار حداقل انتخاب می‌شود. به عبارت دیگر در نهایت از میان صفات خاصه، هر کدام که مقدار ضریب Gini Index آن کوچک‌تر باشد، برای گره جاری درخت تصمیم در نظر گرفته می‌شود و می‌توان درجه ناخالصی را از فرمول زیر محاسبه و هر صفتی که آن را به حداکثر برساند، برگزید.

$$Gini(A) = Gini(D) - Gini_A(D) \quad (7-3)$$

۳-۴-۱-۵) طراحی یک دسته‌بندی کننده درخت تصمیم

اهداف اصلی دسته‌بندی کننده‌های درخت تصمیم:

- ✓ دسته‌بندی صحیح نمونه‌های آموزشی تا حد امکان
- ✓ عام سازی نمونه‌های آموزشی به شکلی که نمونه‌های نادیده بتوانند با دقت هر چه بیشتر دسته‌بندی شوند.
- ✓ به‌روزرسانی آسان درخت، در صورت فراهم شدن نمونه‌های آموزشی جدید
- ✓ داشتن یک ساختار تا سرحد امکان ساده.
- ✓ طراحی یک دسته‌بندی کننده درخت تصمیم به وظایف زیر تقسیم می‌شود:
- ✓ انتخاب مناسب ساختار درخت.
- ✓ انتخاب زیرمجموعه‌های ویژگی برای استفاده در هر گره داخلی.
- ✓ انتخاب قانون تصمیم یا استراتژی استفاده شده در هر گره داخلی.
- ✓ گام‌های لازم برای طراحی یک درخت تصمیم‌گیری:
- ✓ انتخاب مناسبی برای ساختار درخت.
- ✓ انتخاب ویژگی‌هایی مورد نظر برای تصمیم‌گیری در هر یک از گره‌های میانی.
- ✓ انتخاب قانون تصمیم‌گیری یا استراتژی مورد استفاده در هر یک از گره‌های میانی.

۳-۴-۱-۶) روش‌های اصلی برای طراحی دسته‌بندی کننده درخت تصمیم

۱. روش‌های پایین به بالا : طراح، با دسته‌های اطلاعاتی آغاز کرده و ترکیب کردن دسته‌ها را به شکل مداوم تا رسیدن به یک گره که دربردارنده تمام دسته‌ها می‌باشد ادامه می‌دهد.

۲. روش‌های بالا به پایین : طراح، از گره ریشه شروع و با استفاده از یک قانون تفکیک، دسته‌ها را تا زمان برقراری یک شرط توقف تقسیم می‌کند. مسائل اصلی در این روش شامل موارد زیر می‌باشد:

✓ انتخاب معیار تقسیم گره‌ها

✓ قوانین توقف

✓ انتساب برچسب به گره‌های پایانی

(اکثر کارهای انجام شده در زمینه درخت‌های تصمیم‌گیری روش‌های بالا به پایین هستند. نمونه

الگوریتم‌های بالا به پایین، الگوریتم‌های ID3, ID4, ID5R, C4.5 و CART می‌باشند.)

۳. روش‌های رشد دهنده - هرس کننده : این ساختار به منظور اجتناب از رویارویی با مشکلات انتخاب یک

شرط توقف، در ابتدا با استفاده از دستیابی به یکی از روش‌های استاندارد دیگر مانند یک روش بالا به

پایین، درخت را تا بیشترین اندازه ممکن می‌سازد و سپس در مرحله دوم با بهره‌گیری از یک الگوریتم

هرس، درخت به شکل انتخابی هرس خواهد شد. (الگوریتم CART از جمله این الگوریتم‌ها بوده و

الگوریتم C4.5 نیز دارای یک الگوریتم هرس است.)

۴. روش‌های ترکیبی: این روش‌ها زمانی کاربرد دارد که طراح بخواهد روال پایین به بالا را برای هدایت و

کمک یک روال بالا به پایین به کار گیرد. در این روش‌ها از هر دو روش بالا به پایین و پایین به بالا

استفاده می‌شود.

۳-۴-۱-۷) انواع یادگیری در درخت تصمیم‌گیری

در رابطه با نمونه‌های آموزشی دو امکان وجود دارد، امکان اول آنکه تمام نمونه‌های آموزشی در زمان طراحی درخت تصمیم فراهم باشند و امکان دوم آنکه به شکل یک جریان وارد شوند. در حالت اول، در یک مرحله درخت طراحی شده و کار پایان می‌یابد که این الگوریتم را غیرافزایشی می‌نامیم.

در حالت دوم دو انتخاب وجود دارد: هرگاه یک نمونه آموزشی جدید وارد شد، درخت قبلی را حذف کرده و درختی جدید با استفاده از مجموعه آموزشی افزایش‌یافته می‌سازیم یا اینکه درخت موجود را بر اساس اطلاعات جدید، اصلاح می‌کنیم که این الگوریتم‌ها را افزایشی می‌نامیم.

در واقع این دو روش را می‌توان به طور خلاصه اینگونه تعریف کرد :

✓ غیرافزایشی: الگوریتم درخت یک مرتبه با داده‌های آموزشی مورد آموزش قرار می‌گیرد.

✓ افزایشی: دریافت هر نمونه آموزشی جدید موجب بازبینی ساختار یادگیری درخت شده و ممکن است آنچه را که یاد گرفته است بهبود بخشد.

مطلوب است که الگوریتم افزایشی همان درختی را تولید کنند که در صورت در دسترس بودن تمام نمونه‌ها توسط الگوریتم غیر افزایشی تولید می‌شد. ویژگی‌های یک الگوریتم افزایشی خوب عبارتند از :

✓ حافظه مورد نیاز کم.

✓ سرعت بازسازی بالای درخت.

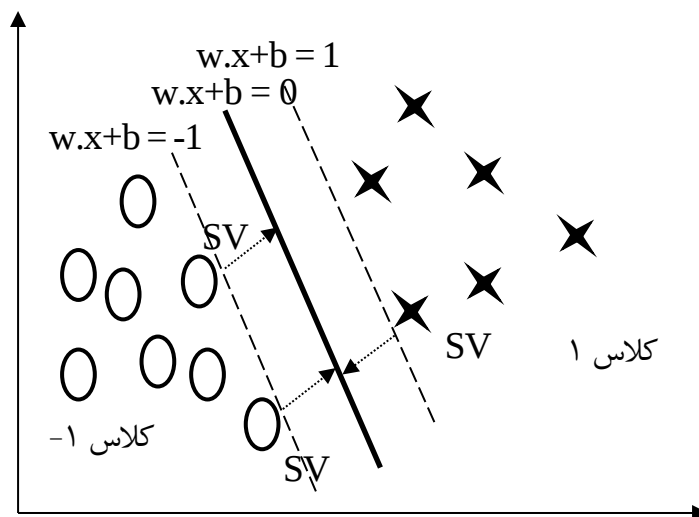
✓ تولید درختی مناسب نسبت به روش‌های غیر افزایشی.

الگوریتم‌های CART و ID3 از نوع الگوریتم‌های غیر افزایشی و الگوریتم‌های ID3 و ID4 و ID5R نمونه‌هایی از الگوریتم‌های افزایشی ساخت درخت می‌باشند.

۳-۴-۲) دسته‌بند با ماشین بردار پشتیبان

این روش یک ساختار دسته‌بندی مبتنی بر هسته^۱ و در قالب یادگیری ماشین است که از آن با عنوان ماشین بردار پشتیبان یاد می‌کنند. وپنیک در سال ۱۹۹۲ و بر مبنای نظریه آماری یادگیری این روش را معرفی کرده است. الگوریتم قادر است هم‌زمان حاشیه‌های هندسی را بیشینه و خطای تجربی جداسازی را کمینه کند به همین دلیل با عنوان دسته‌بند حداکثر حاشیه^۲ مورد بررسی قرار می‌گیرد. در ساختار دسته‌بندی که هدف مورد مطالعه نهایتاً در دو دسته قرار می‌گیرد خطوط زیادی هستند که معرف دسته‌بند باشند ولی تنها یک خط بیشترین تفکیک را حاصل می‌کند. در این بین، تفکیک کننده‌ای که حاشیه داده‌های آموزشی را به بیشترین حد ممکن می‌رساند موجب حداقل شدن خطای تعمیم خواهد شد. نقاط داده‌ای ممکن است ضرورتاً نقاط داده‌ای در فضای R^2 نباشند و ممکن است در فضای چندبعدی R^n باشند. دسته‌بندی‌های خطی بسیاری ممکن است این خصوصیت را ارضا کنند اما SVM به دنبال جداکننده‌ای است که حداکثر جداسازی را انجام می‌دهد. همان‌طور که در شکل (۳-۴) مشاهده می‌شود فرا صفحه‌هایی که از نزدیکی داده‌های آموزش می‌گذرند حساس به خطا می‌باشند و احتمال اینکه برای داده‌های خارج از مجموعه آموزش قدرت تعمیم دهی خوبی داشته باشند بسیار کم است. در عوض، به نظر می‌رسد فرا صفحه‌ای که بیشترین فاصله را از تمام نمونه‌های آموزشی دارد قابلیت‌های تعمیم دهی مناسبی را فراهم آورد. نزدیک‌ترین داده‌های آموزشی به فرا صفحه‌های تفکیک کننده را بردار پشتیبان نامیده می‌شوند. اگر مجموعه داده به صورت $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ نشان داده شود، Y_i می‌تواند مقدار ۱ و -۱ دریافت کند که توسط این ثابت‌ها دسته‌های نقاط X_i مشخص می‌گردد که هر یک بردار Π بعدی است. هنگامی که داده‌های آموزشی دسته‌بندی شده در دسته‌های صحیح را داریم، SVM توسط تقسیم‌بندی فرا صفحه‌ای آن‌ها را در دسته‌های جداگانه قرار می‌دهد طوری که $W^T X + b = 0$ ، بردار W نقاط عمودی فرا صفحه‌ها را جدا می‌کند و b میزان حاشیه را مشخص می‌کند.

^۱. Kernel Methods
^۲. Maximum Margi



شکل (۳-۴) دسته‌بند ماشین بردار پشتیبان

فرا صفحه‌های موازی را می‌توان به صورت $W \cdot X + b = 1$ و $W \cdot X + b = -1$ تعریف کرد. اگر داده‌های آموزشی به صورت خطی جدایی‌پذیر باشند، می‌توان فرا صفحه‌ها را به‌طوری انتخاب نمود که هیچ نمونه‌ای میان آن‌ها نباشد و سپس تلاش کرد تا فاصله آن‌ها را به حداکثر رسانید. برای هر نمونه i از داده‌ها رابطه‌های (۳-۸) و (۳-۹) را داریم.

$$w \cdot x_i + b \geq 1 \quad (۳-۸)$$

$$w \cdot x_i + b \leq -1$$

$$1 \leq i \leq n \quad y_i(w \cdot x_i + b) \geq 1, \quad (۳-۹)$$

فاصله بین دو فرا صفحه را از طریق تحلیل هندسی با رابطه $\frac{2}{\|w\|}$ می‌توان به دست آورد. بنابراین مسئله بهینه‌سازی ما به صورت زیر خواهد بود.

$$\min \frac{1}{2} \|w\|^2 \quad \max \frac{2}{\|w\|} \quad (۳-۱۰)$$

هنگام توزیع داده‌های دسته‌های مختلف این امکان وجود دارد که جدایی‌پذیری به راحتی اتفاق نیفتد و تداخل ایجاد گردد. در این صورت، تفکیک سازی دقیق نمونه‌ها ممکن است سبب تعمیم دهی ضعیف گردد. یک راه حل این است که مقداری خطا در دسته‌بندی را بپذیریم.

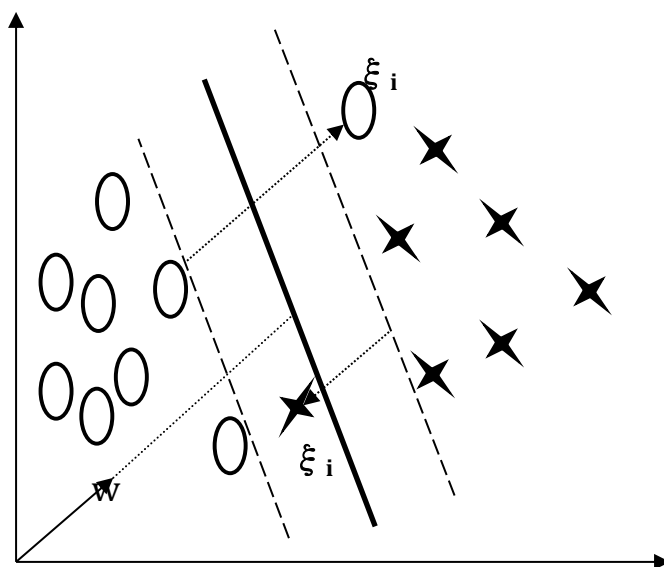
این کار با معرفی متغیر بی‌دقت^۱ (ξ_i) انجام می‌شود که نشانگر نمونه‌هایی است که توسط تابع $w^T x + b = 0$ غلط ارزیابی می‌شوند. این روش که به SVM با حاشیه‌ی نرم^۲ معروف است که اجازه می‌دهد بعضی از نمونه‌ها در ناحیه اشتباه قرار گیرند سپس آن‌ها را جریمه می‌کند، لذا این روش برخلاف SVM حاشیه‌ی سخت^۳ برای مواردی که نمونه‌های آموزشی به‌صورت خطی جدایی‌پذیر نیستند قابل استفاده است. با معرفی متغیر ξ_i محدودیت‌های قبلی ساده‌تر شده و رابطه $w^T x + b = 0$ به‌صورت زیر تغییر می‌کند.

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (۱۱-۳)$$

در این صورت مسئله بهینه‌سازی تبدیل می‌شود به یافتن W به‌نحوی که معادله زیر مینیمم شود:

$$\min \left\{ \frac{1}{2} \|w\|^2 + c \sum_i \xi_i \right\} \quad (۱۲-۳)$$

ماشین بردار پشتیبان مطابق شکل (۵-۳) با حاشیه نرم تلاش می‌کند ξ_i را صفر نگه دارد درحالی که حاشیه‌های دسته‌بند را حداکثر می‌کند.



شکل (۵-۳) svm با حاشیه نرم

^۱. Slack Variable

^۲. Soft-Margin

^۳. Hard-Margin

این الگوریتم تعداد نمونه‌هایی که به اشتباه دسته‌بندی شده‌اند را کمینه نمی‌کند بلکه سعی دارد مجموع فواصل از حاشیه فرا صفحه‌ها را کمینه نماید. مقادیر بزرگ برای C سبب می‌شود که رابطه (۳-۱۲) مانند روش با حاشیه سخت عمل کند. ماشین بردار پشتیبان با حاشیه نرم همیشه یک‌راه حل پیدا می‌کند و در مقابل مجموعه داده‌هایی که دارای یک عضو جدا هستند مقاوم است و به‌خوبی عمل می‌کند.

۳-۴-۳ طبقه‌بند KNN

الگوی الگوریتم K -نزدیک‌ترین همسایگان (K -NN) روشی غیر پارامتری است که برای طبقه‌بندی و رگرسیون استفاده می‌شود. در هر دو حالت، ورودی شامل K نزدیک‌ترین مثال‌های آموزشی در فضای ویژگی است. خروجی بر این اساس که از الگوریتم برای طبقه‌بندی یا رگرسیون استفاده می‌شود، مشخص خواهد شد. در K -NN، خروجی نشانه عضویت در یک کلاس مشخص است. یک عضو توسط بیشترین رأی همسایگان طبقه‌بندی شده و معمولاً به متداول‌ترین کلاس در نزدیک‌ترین همسایگان خود اختصاص داده می‌شود (K عدد صحیح مثبت است، معمولاً کوچک). در صورتی که مقدار K برابر عدد یک باشد، اختصاص شیء به‌سادگی امکان پذیر می‌شود. اما هنگام رگرسیون در K -NN، خروجی الگوریتم یک مقدار ویژگی، برای جسم است. این مقدار، میانگین مقادیر K نزدیک‌ترین همسایگان است. الگوریتم K -NN نوعی یادگیری مبتنی بر مثال یا یادگیری تنبل است، جایی که عملکرد فقط به‌صورت محلی تقریبی می‌شود و تمام محاسبات تا زمان طبقه‌بندی به تعویق می‌افتد. هم برای طبقه‌بندی و هم برای رگرسیون، یک روش مفید می‌تواند تعیین وزن به سهم همسایگان باشد، به‌طوری‌که همسایگان نزدیک‌تر بیشتر از افراد دورتر به طور متوسط مشارکت کنند. یک وزنه‌برداری مشترک شامل دادن وزن $1/d$ به هر همسایه است، جایی که d فاصله با آن همسایه است. ویژگی الگوریتم K -NN این است که به ساختار محلی داده حساس است. مثال‌های آموزش بردارهایی هستند که در یک فضای ویژگی چندبعدی وجود دارند، هر یک دارای برچسب کلاس هستند. مرحله آموزش الگوریتم فقط از ذخیره ناقل ویژگی‌ها و برچسب‌های کلاس نمونه‌های آموزشی تشکیل شده است.

در مرحله طبقه‌بندی، K یک عدد ثابت تعریف‌شده توسط کاربر است و یک بردار بدون برچسب (یک پرس‌وجو یا یک نقطه تست) با اختصاص برچسب طبقه‌بندی می‌شود که بیشترین تعداد نمونه‌های آموزش K را که نزدیک‌ترین نقطه به آن پرس‌وجو هستند، دارا باشد. متریک مسافتی که معمولاً برای متغیرهای مداوم استفاده می‌شود، فاصله اقلیدسی است. برای متغیرهای گسسته، مانند طبقه‌بندی متن، می‌توان از متریک دیگری مانند متریک هم‌پوشانی (یا فاصله چکش زدن) استفاده کرد. برای مثال، در داده‌های ریزآرایه بیان ژن، K -NN با ضریب همبستگی مانند پیرسون و اسپیرمن به‌عنوان یک متریک به‌کارگرفته‌شده است. اغلب، اگر طبقه‌بندی مسافت با الگوریتم‌های تخصصی مانند بزرگ‌ترین حاشیه نزدیک‌ترین همسایه یا تجزیه و تحلیل اجزای محلی، آموخته شود، می‌توان دقت طبقه‌بندی K -NN را به میزان قابل‌توجهی بهبود داد.

اشکال در طبقه‌بندی اصلی "رأی اکثریت" اساسی وقتی اتفاق می‌افتد که توزیع کلاس کم‌رنگ باشد. یعنی نمونه‌هایی از یک کلاس ممکن بیشتر تمایل دارند که بر پیش‌بینی نمونه جدید تسلط داشته باشند، زیرا به دلیل تعداد زیاد، آن‌ها در بین نزدیک‌ترین همسایگان متداول هستند. یکی از راه‌های غلبه بر این مشکل وزن‌گیری طبقه‌بندی با در نظر گرفتن فاصله از نقطه آزمون تا هر یک از نزدیک‌ترین همسایگان خود است. کلاس (یا مقدار، در مشکلات رگرسیون) از هر نزدیک‌ترین نقطه با وزن متناسب با معکوس فاصله از آن نقطه تا نقطه آزمون ضرب می‌شود. راه دیگر برای غلبه بر برنده بودن، انتزاع در نمایش داده‌ها است. به‌عنوان مثال، در یک نقشه خودسازمان‌دهی (SOM)، هر گره بدون توجه به چگالی آن‌ها در داده‌های اصلی آموزش، نماینده (مرکز) یک خوشه از نقاط مشابه است. الگوریتم K -NN می‌تواند در SOM اعمال شود. بهترین انتخاب K به داده بستگی دارد؛ به‌طور کلی، مقادیر بزرگ‌تر K تأثیر نویز را در طبقه‌بندی کاهش می‌دهد، اما مرزهای بین طبقات را کمتر مشخص می‌کند. با استفاده از تکنیک‌های مختلف اکتشافی، می‌توان k را انتخاب کرد.

دقت الگوریتم K -NN با وجود ویژگی نویز یا بی‌ربط یا هنگامی که مقیاس ویژگی با اهمیت آن‌ها سازگار نیست، می‌تواند به‌شدت تخریب شود. تلاش‌های زیادی در زمینه انتخاب یا مقیاس‌گذاری برای بهبود طبقه‌بندی صورت گرفته است. یک رویکرد استفاده از الگوریتم‌های تکاملی برای بهینه‌سازی مقیاس‌پذیری ویژگی است.

۳-۴-۱) مراحل آموزش الگوریتم

در این الگوریتم نمونه‌های آموزشی با ساختار بردارهایی در فضای ویژگی چند بعدی مطرح شده و فضا به ناحیه‌هایی با نمونه‌های آموزشی بخش‌بندی می‌شود. هر نقطه k در فضای مجموعه به کلاسی متعلق خواهد بود که بیشترین نقاط آموزشی در آن کلاس در داخل نزدیک‌ترین نمونه آموزشی به k در آن باشند. برای سنجش فاصله در این الگوریتم معمولاً از فاصله اقلیدسی با فرمول، $\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ استفاده می‌شود. در فاز تفکیک، نمونه تستی به عنوان یک بردار در فضای ویژگی نمایش داده می‌شود و فاصله اقلیدسی یا تشابه کوسینوسی بردار تستی با کل بردارهای آموزشی محاسبه می‌شود و نزدیک‌ترین نمونه آموزشی به k انتخاب می‌شود. سه فاکتور مهم در این الگوریتم به شرح زیر می‌باشد:

- ✓ معیار فاصله، برای پیدا کردن k همسایه نزدیک استفاده می‌شود.
- ✓ K تعداد همسایه‌های نزدیک است.
- ✓ قانون تصمیم‌گیری برای شناسایی یک کلاس برای سند تستی از k همسایه نزدیک می‌باشد.

۳-۴-۲) ساختار الگوریتم

همچنین لازم به ذکر است که تمام سه شرط اندازه‌گیری فاصله تنها برای متغیرهای پیوسته معتبر است. در مورد متغیرهای گسسته باید از فاصله همینگ (Hamming Distance) استفاده شود که مسئله استانداردسازی متغیرهای عددی بین ۰ و ۱ را به وجود می‌آورد درحالی‌که مخلوطی از متغیرهای عددی و گسسته در مجموعه داده وجود دارد. در روش K نزدیک‌ترین همسایه، انتخاب بهترین مقدار برای K بهتر است با اولین بررسی داده‌ها انجام شود. به‌طور کلی، یک مقدار بزرگ K دقیق‌تر است زیرا نویز کلی را کاهش می‌دهد اما هیچ تضمینی برای اعتبار آن وجود ندارد. اعتبار سنجی متقاطع یکی دیگر از راه‌های به‌دست آوردن یک K خوب با استفاده از یک مجموعه داده مستقل برای اعتبار سنجی K می‌باشد. به لحاظ تجربه، K مطلوب برای بیشتر مجموعه داده‌ها بین ۳ تا ۱۰ است. مراحل الگوریتم KNN شامل موارد زیر خواهد بود:

۱. داده‌ها را بارگیری کنید.

۲. K به عنوان تعداد نزدیک‌ترین همسایگان انتخاب کنید.

۳. برای هر یک از داده‌های اولیه:

۱. فاصله بین داده مورد سؤال و هر یک داده‌های اولیه را محاسبه کنید.

۲. فاصله و اندیس نمونه را به یک مجموعه اضافه کنید.

۳. مجموعه را بر اساس فاصله از کوچک به بزرگ مرتب کنید.

۴. نقاط K عضو اول مجموعه مرتب شده را انتخاب کنید.

۵. بسته به حالت یا حالت طبقه‌بندی، خروجی را اعلام کنید.

۳-۵) جامعه آماری

جامعه آماری عبارت‌اند از مجموعه‌ای از افراد، رویدادها و پدیده‌ها که دارای حداقل یک صفات مشترک باشند. هر جامعه آماری دارای چهارچوبی متشکل از فهرست همه عناصر جامعه است که نمونه پژوهش از آن استخراج می‌شود. تعریف جامعه آماری باید مانع و جامع باشد، به این معنی که تعریف باید چنان بیان شود که از نقطه نظر زمانی و مکانی همه‌ی واحدهای مورد مطالعه را در بر گرفته و از مدنظر قرار دادن واحدهایی که نباید به مطالعه آن‌ها پرداخته شود، جلوگیری به عمل آورد (حافظ نیا، ۱۳۸۹). جامعه آماری تحقیق حاضر شامل همه کارکنان شرکت‌های خصوصی توزیع برق استان خوزستان می باشد که از ابتدای جذب تا پایان سال ۱۳۹۸ مورد مطالعه قرار گرفته است.

۳-۶) نمونه آماری

نمونه‌گیری فرایند انتخاب کردن تعداد کافی از میان اعضای جامعه آماری است، به گونه‌ای که با مطالعه گروه نمونه و درک خصوصیات و ویژگی‌های آزمودنی‌های گروه نمونه، قادر خواهد بود این خصوصیات و ویژگی‌ها را به اعضای جامعه آماری تعمیم دهد.

با توجه به اینکه در استفاده از تکنیک های داده کاوری، اندازه بیشتر نمونه موجب افزایش دقت می گردد، لذا کل اعضا جامعه تحقیق به صورت سرشماری مورد بررسی قرار می گیرد.

۷-۳) ابزار گردآوری اطلاعات

پژوهش حاضر جهت دستیابی به اطلاعات دانش نظری و اصولی پژوهش از روش کتابخانه ای بهره گرفته است روش کتابخانه ای در این تحقیق، به منظور شناخت ادبیات و سوابق موضوع و بررسی روند مقوله مورد بحث از روش کتابخانه ای استفاده گردیده است. برای جمع آوری داده های مربوط به کارکنان از بانک های اطلاعاتی و پرونده های پرسنلی جامعه مورد تحقیق استفاده شده است.

۸-۳) متغیرهای مورد بررسی

عدم همکاری شرکت ها جهت در اختیار گذاشتن اطلاعات نیروی انسانی به دلیل مسایل امنیتی و زمان بر بودن یافتن سازمانی که حاضر به همکاری شود و در عین حال اطلاعات مورد نیاز را داشته باشد و مساله مهم تر این که اکثر فرم ها و پرونده ها به صورت غیر سیستمی و دستی ذخیره شده اند، باعث شد تا در نهایت به مجموعه نمایش داده شده در جدول (۱-۳) به عنوان متغیرهای مورد بررسی برسیم.

جدول (۱-۳) متغیرهای مورد بررسی

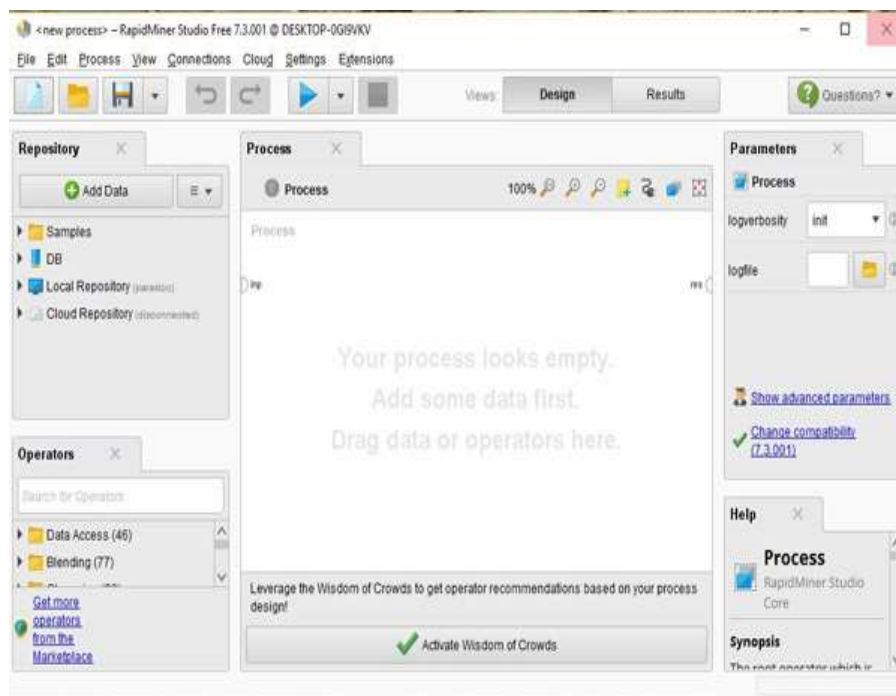
متغیر
جنسیت
وضعیت تاهل
مدرک تحصیلی (ورود به سازمان)
رشته تحصیلی (ورود به سازمان)
سن استخدام
نوع استخدام
محل اشتغال
سمت
حقوق سال آخر خدمت
سنوات

۳-۹) ابزار و روش تجزیه و تحلیل داده‌ها

پژوهش حاضر به دنبال بررسی و پیش‌بینی وفاداری کارکنان با استفاده از تکنیک داده کاوی است. در این پژوهش به منظور تجزیه و تحلیل داده‌ها از نرم‌افزار Rapidminer بهره گرفته شده است. داده کاوی با Rapidminer یکی از محبوب‌ترین و پرتعدادترین شیوه‌های داده کاوی است که بدون نیاز به هیچ‌گونه دانش برنامه‌نویسی می‌توانید از آن استفاده کنید. صفحه اصلی نرم‌افزار در شکل ۳-۶ نشان داده شده است. اجزای این صفحه عبارت‌اند از :

پنجره Repository : در این پنجره فایل‌های مدل‌سازی قرار خواهد گرفت و شامل پوشه‌هایی است که پروژه‌های نرم‌افزار در آن ذخیره می‌گردد.

پنجره عملگرها یا Operators : در سمت چپ محیط نرم‌افزار پنجره‌ای قرار دارد که تمام کنترل‌ها و الگوریتم‌های قابل استفاده در نرم‌افزار را به صورت تفکیک شده ارائه می‌دهد. این فضا امکان جستجوی یک کنترل یا الگوریتم خاص را فراهم کرده است و برای این کار وارد کردن ابتدای نام الگوریتم مورد نظر مثل Read در فضای جستجو کفایت می‌کند تا تمام کنترل‌هایی که شامل عبارت Read هستند نمایان گردد.



شکل (۳-۶) صفحه اصلی نرم‌افزار Rapidminer

پنجره Process : این پنجره از مهم‌ترین پنجره‌های نرم‌افزار است و برای اضافه کردن کنترل‌های متنوع جهت تهیه مدل‌های مورد نظر، به کار می‌رود.

پنجره پارامترها یا Parameters : هر کنترل در نرم‌افزار دارای مجموعه‌ای از آرگومان‌های ورودی و ویژگی است که تمام آن‌ها از طریق پنجره Parameters قابل دسترسی هستند. این پنجره در سمت راست محیط نرم‌افزار قرار دارد. آرگومان‌های موجود در این نرم‌افزار به راحتی و با یک کلیک ساده بر روی هر کنترل قابل استفاده می‌باشند.

پنجره Help : این پنجره زمانی که کاربران نیازمند کسب اطلاعات درباره یک الگوریتم یا کنترل واحدهای آن هستند، مورد استفاده قرار می‌گیرد. اطلاعات موجود در فضای این بخش مطالبی در ارتباط با ورودی‌ها و خروجی‌ها و آرگومان‌های یک کنترل را در اختیار کاربر قرار می‌دهد.

پنجره Design یا طراحی : این پنجره به شکل یک دکمه در بالای پنجره Process قرار دارد و با استفاده از آن هر لحظه می‌توان به پنجره اصلی نرم‌افزار مراجعه کرد.

پنجره Result : این پنجره جهت نمایش محیط نتایج و خروجی مدل‌های تهیه شده استفاده خواهد شد.

۳-۱۰) مقایسه الگوریتم‌ها

کسب دانش برای انسان مسیرهای گوناگونی دارد. یکی از روش‌های یادگیری در علم شناسایی الگو از طریق یادگیری نظارتی است. در این روش هر الگو یک برچسب (Label) دارد که مشخص کننده خروجی مطلوب آن می‌باشد. هدف این نوع از الگوریتم‌ها تشخیص تابعی است که بتواند بر اساس ساختاری مشخص، ورودی را به برچسب‌های متناظر داده شده نگاشت کند. این فاز تحت عنوان آموزش انجام می‌گیرد. فاز بعدی تست است که در آن الگوهایی با برچسب نامشخص به سیستم وارد و به کمک تابع آموزش داده شده خروجی را پیش‌بینی می‌کند. حال اگر خروجی سیستم یادگیری گسسته در نظر گرفته شود مساله طبقه بندی و تابعی که ورودی را به خروجی نگاشت می‌کند طبقه بندی کننده نامیده می‌شود.

در جدول (۲-۳) به مقایسه الگوریتم‌های مورد استفاده جهت دسته‌بندی کارمندان با برچسب وفاداری و بررسی نکات ضعف و قوت و ویژگی‌های آن‌ها خواهیم پرداخت.

جدول (۲-۳) مقایسه الگوریتم‌ها

الگوریتم	ویژگی‌ها	مزایا	معایب	نکات
ID3	• یک الگوریتم پایه یادگیری • قدرت شناسایی بهترین ویژگی در حجم اطلاعات زیاد • یک الگوریتم از بالا به پایین	• استفاده از انتخاب ویژگی (بعد) کمک می‌کند تا سطح کمتری داشته باشد و ساده تصمیم بگیرد. • Entropy به درد داده‌هایی می‌خورد که قسمت‌های کوچک زیادی دارند و مقادیر یکتا در آن‌ها بیشتر است.	• فقط برای ویژگی‌های گسسته • عدم ساخت درخت تصمیم در صورت مقادیر ناموجود • پیچیدگی در ساخت درخت تصمیم به علت دادن شاخه و برگ زیاد	آزمون Information Gain استفاده می‌شود تا مشخص گردد که تا چه حد قادر است مثال‌های آزمایشی را دسته‌بندی کند. ID3 برای اینکه تصمیم بگیرد چگونه گره‌های درخت را انتخاب کند از معیاری به نام معیار Entropy استفاده می‌کند.
C4.5	• می‌تواند مقادیر گسسته یا پیوسته را در ویژگی‌ها درک کند. • قادر است تا مقادیری که موجود نیستند را هم تحمل کند. • این قابلیت را دارد که وزن‌های مختلف و غیر یکسانی را به ویژگی‌ها بدهد.	• استفاده برای مقادیر پیوسته و گسسته • ساخت درخت با توجه به مقادیر ناموجود • وزن دهی به ویژگی‌ها • دارای عملیات هرس جهت جلوگیری از Overfitting است.	• اگر تعداد رکوردها کم باشد نرخ خطا بالا می‌رود	با عملیات هرس درخت که انجام می‌دهد، می‌توان مدل را به یک نقطه بهینه رساند که زیاد پیچیده نباشد برای اینکه تصمیم در مورد چگونگی انتخاب گره‌های درخت از معیار Ratio Gain استفاده می‌کند.
CART	• کاربرد زیادی در طبقه‌بندی و رگرسیون دارد.	• یکی از ساده‌ترین الگوریتم‌های درخت تصمیم است. • بر اساس درخت دودویی (باینری) بنا شده است. • شاخص جینی برای داده‌هایی که دارای قسمت بزرگتر هستند به درد می‌خورد. • جهت جلوگیری از Overfit شدن درخت تصمیم CART می‌توان از یک شرط توقف استفاده کرد.	• برای تعیین مقدار بهینه جهت نقطه‌ای جداساز نیاز به آزمون و خطا دارد.	مخفف Classification and Regression Tree است. CART برای اینکه تصمیم بگیرد چگونه گره‌های درخت را انتخاب کند از معیار شاخص Gini استفاده می‌کند. برای هر ویژگی (بعد) هر چقدر شاخص جینی کمتر باشد، یعنی آن ویژگی اطلاعات بیشتری را به ما می‌دهد

الگوریتم	ویژگی ها	مزایا	معایب	نکات
SVM	ماشینهای مبتنی بر بردار پشتیبان به محاسبات پیچیده و زمان بر نیاز دارند و به دلیل پیچیدگی محاسباتی، حافظه زیادی نیز مصرف می کنند.	دارای یک شالوده نظری منسجم بوده و جواب های تولید شده سراسری و یکتا می باشند.	- محدودیت های ذاتی دارند مثلاً هنوز مشخص نشده است که به ازای یک تابع نگاشت، پارامترها را چگونه باید تعیین کرد. - داده های گسسته و غیر عددی هم با این روش سازگار نیستند و باید تبدیل شوند.	امروزه ماشینهای بردار پشتیبان، به متداول ترین تکنیک های پیش بینی در داده کاوی تبدیل شده اند.
KNN	- KNN یک مورد خاص از پهنای باند متغیر است. - یکی از روش های یادگیری با نظارت است. - برای اندازه گیری فاصله یا میزان شباهت از روش اقلیدسی استفاده می کند.	- ساده است. - نتایجی با سازگاری (Consistency) قوی دارد. - دقت نسبتاً بالا دارد.	- در الگوریتم KNN ابتدا بایستی فاصله ی هر نمونه ی جدید را با نمونه های گذشته مقایسه کنیم. - برای مجموعه اولیه های بزرگ، این محاسبات بسیار سنگین خواهد بود.	در مسائل دنیای واقعی ممکن است ابعاد بسیار بیشتری وجود داشته باشد. این مسائل را دیگر نمی توان روی بردارهای عمودی و افقی به صورت ساده رسم کرد. در دسته بندی های دو کلاس، بهتر است که k یک عدد فرد انتخاب شود؛ زیرا این امر از گره خوردن آرا جلوگیری می کند

۳-۱۱) الگوریتم پیشنهادی

۳-۱۱-۱) درخت تصمیم

هر گره در درخت تصمیم با عنوان یک آزمایش منظور می شود و در کل یک نمودار گردش عملیات را شبیه سازی می کند و هر شاخه نیز نشان دهنده نتایج این آزمایش است. پس در نهایت برگ های هر شاخه شامل برچسب کلاس بندی خواهد بود. مسیر از ریشه به برگ ها بیانگر قوانین طبقه بندی است. به منظور توسعه و رشد درخت، تکنیک های مختلفی وجود دارد که در ادامه در جدول (۳-۳) به برخی از آنها اشاره می شود. توجه داشته باشید که درخت تصمیم در این متن به منظور تعیین ارتباط بین دو یا چند متغیر کمی و کیفی به کار رفته است.

جدول (۳-۳) مقایسه درخت‌های تصمیم

ردیف	شرح	C4.5	CART	ID3
۱	برمبنای کای ۲	بله	—	—
۲	جایگزینی متغیرهای مستقل (پیش‌گو)	—	بله	بله
۳	امکان هرس درخت	—	بله	بله
۴	تفکیک گره چند راهه	بله	—	—
۵	تفکیک گره دو دویی	—	بله	بله
۶	متغیرهای تاثیرگذار	بله	بله	—
۷	استفاده از احتمال پیشین	—	بله	بله
۸	تابع هزینه بد رده‌بندی	بله	بله	بله
۹	محاسبه سریع	بله	—	بله

جدول (۳-۴) پارامترهای اجرایی هر الگوریتم را به تفکیک نشان خواهد داد.

جدول (۳-۴) مشخصات الگوریتم درخت تصمیم (<https://docs.rapidminer.com>)

الگوریتم	پارامتر	حدود تغییر	مقادیر مجاز	بهترین مقدار ^۱	شاخص مورد استفاده
ID3	minimal_size_for_split	integer	2-4-6	4	information_gain accuracy
	minimal_leaf_size	integer	6-8-10	8	
	minimal_gain	real	0.3-0.9	0.7	
C4.5	c45_filestem	filename	2-4-6	4	gain_ratio accuracy
	Datamanagement	selection	2	2	
	decimal_point_character	char	10	10	
	Encoding	selection	0.1-0.5	0.25	
CART	minimal_size_for_split	integer	2-4-6	4	gini_index accuracy
	minimal_leaf_size	integer	2-4-6	6	
	minimal_gain	real	0.1-0.5	0.1	
	maximal_depth	integer	8	8	
	Confidence	real	0.1-0.3	0.25	
	number_of_prepruning_alternatives	integer	3	3	
	no_prepruning	boolean	no	no	
	no_pruning	boolean	no	no	
	guess_subset_ratio	boolean	no	no	
	subset_ratio	real	0.7-0.9	0.9	
	use_local_random_seed	boolean	no	no	
	local_random_seed	integer	2001	2001	

^۱ مقادیر به دست آمده به عنوان بهترین مقدار بر اثر برآورد مقادیر به دست آمده در کارهای پیشین و مشابه، راهنمای نرم افزار رپیدماینر (<https://docs.rapidminer.com>) و البته آزمون روی الگوریتم انتخابی بین مقادیر ثبت شده حاصل شده اند.

۳-۱۱-۲) الگوریتم SVM

پس از آماده سازی ماتریس الگو و انتخاب تابع کرنل، پارامتر تابع کرنل و مقدار محدودیت حافظه کرنل را انتخاب خواهیم کرد. اطلاعات پارامترها در جدول (۳-۵) نمایش داده شده است.

جدول (۳-۵) پارامترهای مدل

پارامتر	ارزش ^۱
تابع کرنل	Sigmoid
حداکثر تعداد تکرار	۵۰۰۰
محدودیت حافظه کرنل	۱۰۰۰
مقدار اولیه	Random
مجموعه داده آموزش	۹۰٪
مجموعه داده های تست	۱۰٪

۳-۱۱-۳) الگوریتم KNN

در هنگام اجرای الگوریتم، K به عنوان یک ثابت توسط کاربر تعریف خواهد شد و بردار بدون برچسب (بردار تست) از دسته ای در نظر گرفته می شود که بیشترین تعداد را در K نزدیک ترین همسایه آن نقطه داشته باشد. بهترین انتخاب برای K بستگی به موارد متعددی دارد؛ ولی در کل باید در نظر داشت که مقادیر بزرگ برای آن موجب کاهش خطا در طبقه بندی شده ولی وضوح تفکیک کلاس ها را کمتر می کند. میزان K مناسب را می توان با استفاده از تکنیک های مختلف انتخاب کرد. مورد خاص زمانی است که دسته پیش بینی شده برای عضو جدید، همان دسته نزدیکترین نمونه باشد. در دسته بندی های هنگامی که با دو کلاس مواجه هستیم، بهتر این است که K یک عدد فرد باشد، این حالت موجب جلوگیری از پیچیدگی خواهد شد. در این طرح میزان K را از مقادیر ۱، ۳ و ۵ آزمایش کردیم که بهترین مقدار ۳ به دست آمد، نتایج حاصل در بخش بعد آمده است (اوریت^۲ و همکاران، ۲۰۱۱).

^۱ ارزش پارامترهای مورد استفاده برای الگوریتم ماشین بردار پشتیبان به جز تفکیک داده های آموزش و تست که با آزمون و خطا و البته از بازه ۷۰٪ به ۳۰٪ آغاز شد، از برآیند موارد مشابه و پیشینه پژوهش استفاده شده است.

^۲ Everitt, B. S. et al

۳-۱۲) جمع‌بندی

در این فصل به روش‌شناسی پژوهش و زیرشاخه‌های آن پرداخته شد. سپس در ادامه ابزار جمع‌آوری اطلاعات و داده‌های مورد نیاز، و ابزار تحلیل و بررسی داده‌ها معرفی شد. همچنین در این فصل جامعه آماری پژوهش و روش تجزیه و تحلیل شرح داده شد.

فصل چہارم:

تخلیل و بررسی نتائج

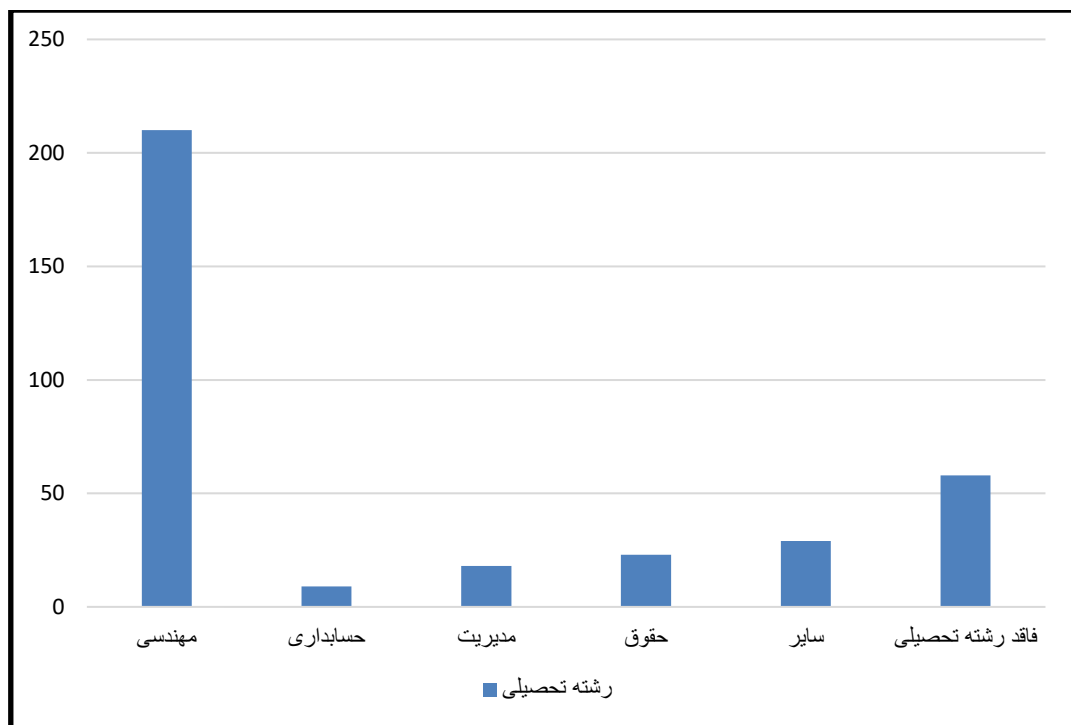
همان طور که در بخش های قبلی بیان شد در این پژوهش با بررسی و استفاده از ۳۴۹ پرونده از پرونده های کارکنان شرکت های ارائه دهنده خدمات توزیع برق استان خوزستان سعی در کشف الگوهای انتخاب موثر برای کشف میزان وفاداری کارکنان به سازمان شده است. در ادامه پژوهش، به بررسی و تحلیل نتایج حاصل از روش های معرفی شده در فصول قبل می پردازیم. در این فصل خروجی های حاصل از اجرای هر الگوریتم در نرم افزار آورده شده است و هریک از الگوریتم ها به تفکیک مورد تحلیل و بررسی قرار می گیرند و در انتها دو الگوریتمی که دارای نتیجه قابل قبول تر بودند با هم ترکیب می شوند و نتایج حاصل از ترکیب دو الگوریتم نیز مورد تحلیل قرار می گیرند و در انتها بهترین الگوریتم، به عنوان الگوریتم موثر و کارا جهت تعیین کارکنان وفادار معرفی می شود و مورد استفاده سازمان قرار می گیرد.

۴-۲) آمار توصیفی

این پژوهش از کارکنان شرکت های ارائه دهنده خدمات توزیع برق استان خوزستان انجام شده است و تعداد کل کارکنانی که از اطلاعات ایشان جهت ارائه الگوریتم مناسب برای تعیین کارکنان وفادار استفاده شد برابر با ۳۴۹ نفر بود. از بین این افراد ۴۰ نفر معادل با ۱۱.۵ درصد زن و ۳۰۹ نفر معادل با ۸۸.۵ درصد مرد بودند. تعداد افراد مجرد در بین این کارکنان ۲۲ نفر معادل با ۶.۳ درصد و تعداد افراد متاهل ۳۲۷ نفر معادل با ۹۳.۷ درصد بود. همچنین توزیع فراوانی دیگر متغیرها نظیر تحصیلات، سن و رشته تحصیلی افراد در جدول های (۴-۱)، (۴-۲) و (۴-۳) که مربوط به کدگذاری این متغیرها می باشد آورده شده است. همانطور که در این جداول مشاهده می شود بیشترین تعداد افراد استخدام شده در سازمان در رشته های مدیریت و مهندسی تحصیل کرده اند، تعداد قابل توجهی از افراد دارای مدرک تحصیلی دیپلم و لیسانس هستند و سن اکثر افراد به هنگام ورود به سازمان در بازه ۲۰ تا ۴۰ سالگی بوده است. از بین ۳۴۹ نفری که در ۳۸ سال گذشته در این سازمان استخدام شده اند تعداد ۲۷۹ نفر (۸۰ درصد) در سازمان مشغول هستند و تعداد ۷۰ نفر (۲۰ درصد) از سازمان خارج شده اند.

جدول (۱-۴) کدگذاری رشته‌های تحصیلی

کد	تعداد افراد	رشته تحصیلی
۱	۲۱۰	مهندسی مهندسی عمران، کامپیوتر، مکانیک، برق
۲	۹	حسابداری حسابداری صنعتی، حسابداری مالی، حسابداری عمومی، آمار
۳	۱۸	مدیریت مدیریت برنامه ریزی و آموزش، مدیریت صنعتی، مدیریت بازرگانی، مدیریت اجرایی، مدیریت دولتی
۴	۲۴	حقوق علوم انسانی
۵	۳۰	سایر کشاورزی، تاریخ و علوم تجربی، جغرافیا
۶	۵۸	فاقد رشته تحصیلی ^۱
	۳۴۹	مجموع

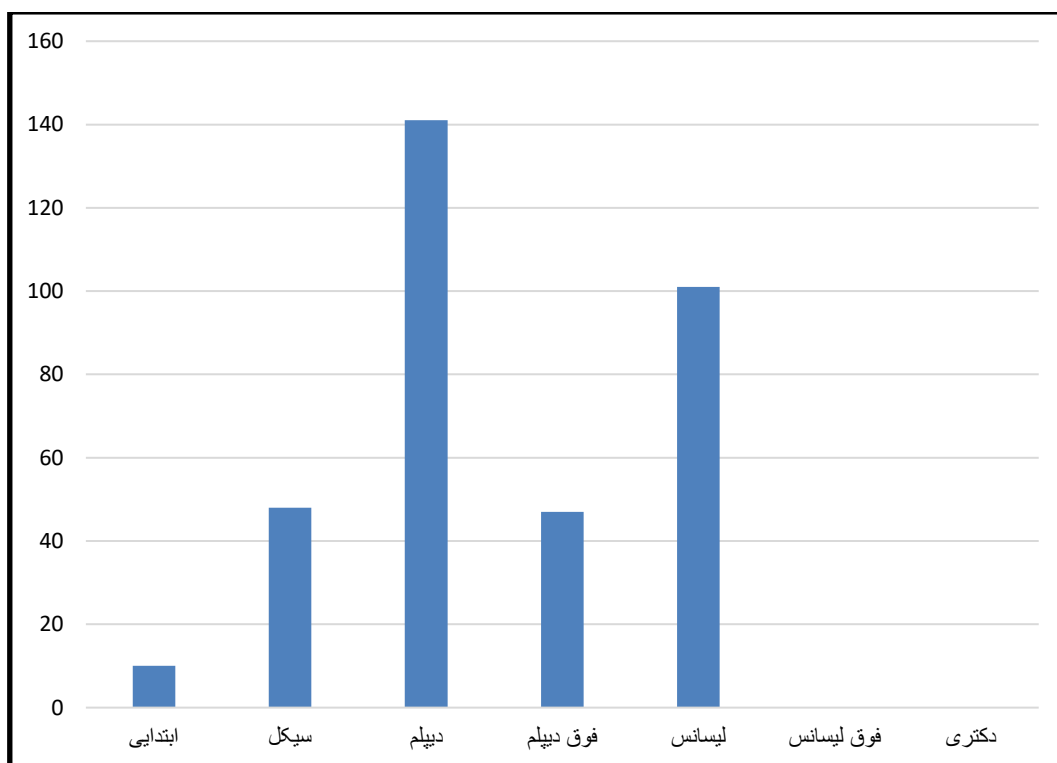


نمودار (۱-۴) پراکندگی رشته‌های تحصیلی

^۱ رشته‌های تحصیلی مقطع دبیرستان در این رسته قرار نمی‌گیرند و صرفاً مقاطع تحصیلی پایین‌تر فاقد رشته در نظر گرفته شده‌اند.

جدول (۲-۴) کدگذاری مقاطع تحصیلی

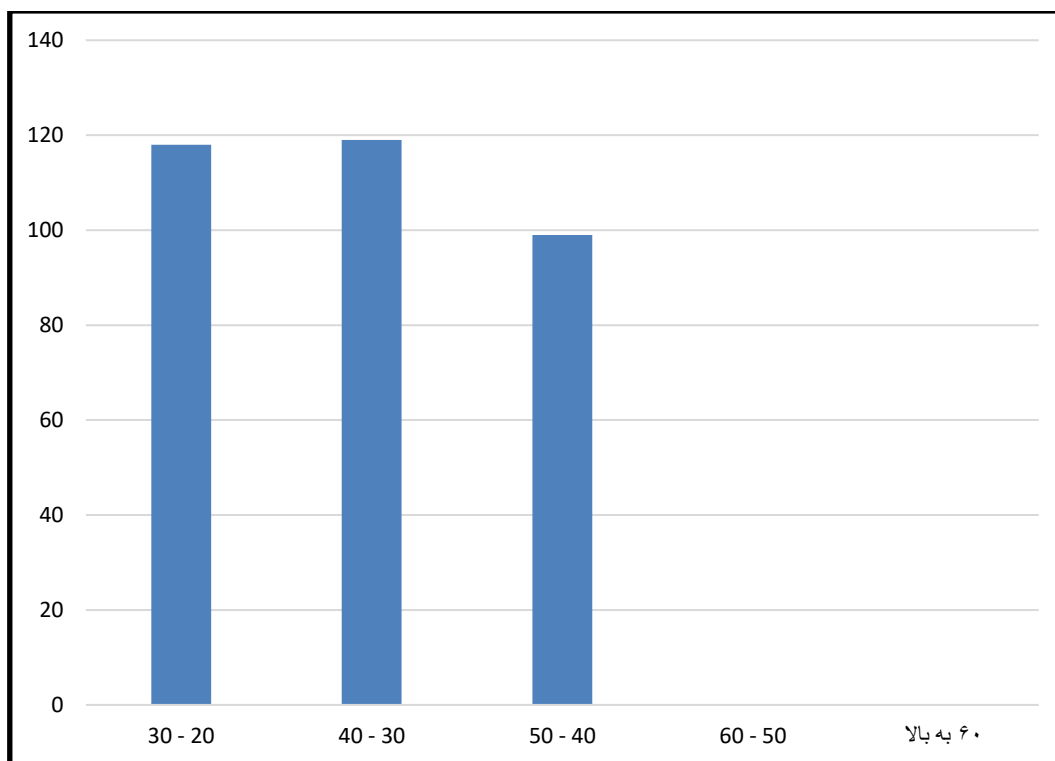
کد	تعداد افراد	مقاطع تحصیلی
۱	۱۰	ابتدایی
۲	۴۸	سیکل
۳	۱۴۱	دیپلم
۴	۴۸	فوق دیپلم
۵	۱۰۲	لیسانس
۶	۰	فوق لیسانس
۷	۰	دکتری
	۳۴۹	مجموع



نمودار (۲-۴) پراکندگی مقاطع تحصیلی

جدول (۳-۴) کدگذاری سن استخدام کارکنان

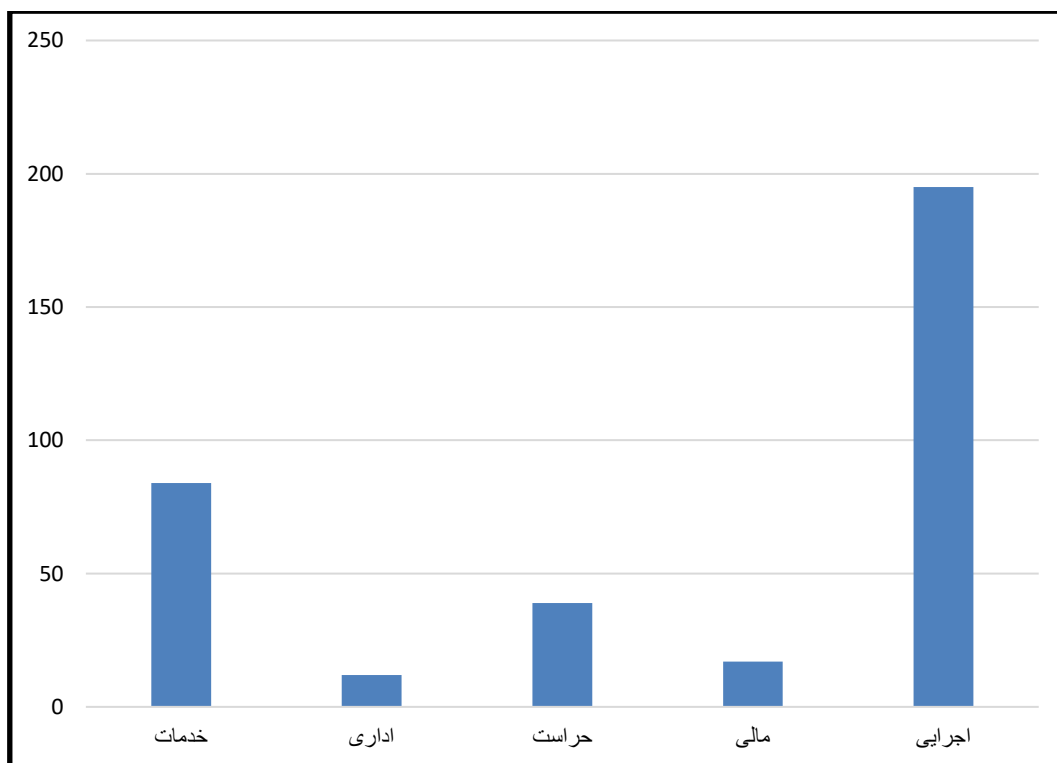
کد	تعداد افراد	سن استخدام کارکنان
۱	۱۴۱	۲۰ - ۳۰
۲	۱۰۳	۳۰ - ۴۰
۳	۱۰۵	۴۰ - ۵۰
۴	۰	۵۰ - ۶۰
۵	۰	۶۰ به بالا
	۳۴۹	مجموع



نمودار (۳-۴) پراکندگی سن کارکنان

جدول (۴-۴) کدگذاری محل اشتغال

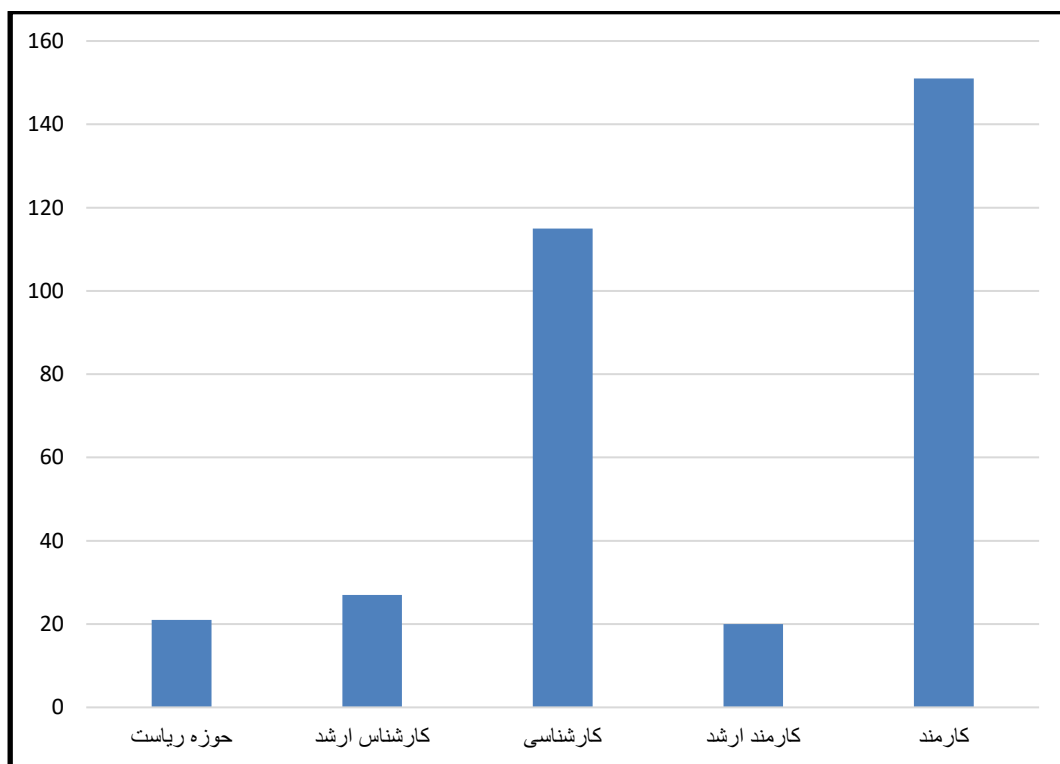
کد	تعداد افراد	محل اشتغال
۱	۸۴	خدمات امور مشترکین، اداره مشترکین، خدمات مشترکین
۲	۱۳	اداری امور اداری، دفتر حقوقی، ریاست
۳	۳۹	حراست اداره حراست، نگهبانی، حراست و امور محرمانه
۴	۱۸	مالی حسابداری، امور مالی، اداره مالی، مالی پشتیبانی
۵	۱۹۵	اجرایی اداره مهندسی، اداره عملیات، اداره بهره برداری، انبار
	۳۴۹	مجموع



نمودار (۴-۴) پراکندگی محل اشتغال

جدول (۴-۵) کدگذاری سمت کارکنان

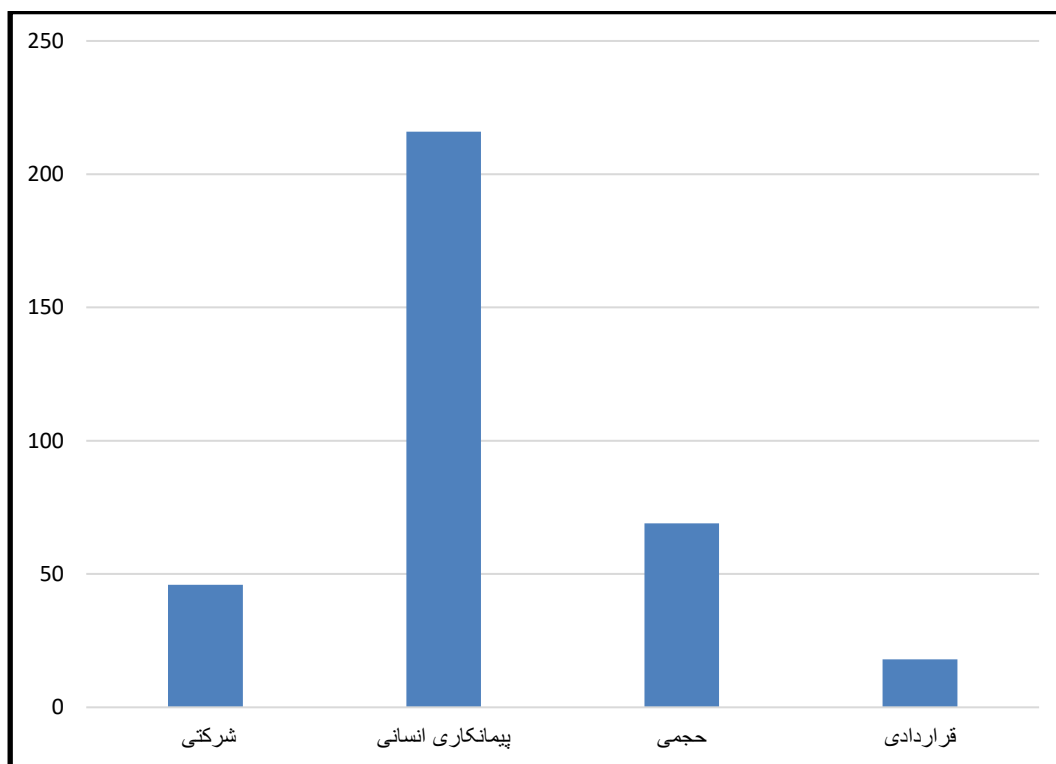
کد	تعداد افراد	سمت کارکنان
۱	۲۱	حوزه ریاست حراست، مشترکین، بهره برداری، امور اداری، مالی و پشتیبانی، برق شعیه، مهندسی
۲	۲۷	کارشناس ارشد ایمنی، کنترل پروژه، ناظر مهندسی، طراح، ناظر بهره برداری، ناظر، کارشناس فنی
۳	۱۲۷	کارشناسی کنترل بار، برقکار، ناظر مشترکین، بازرس فنی مشترکین، مسئول دیماند، کاردانی فنی، مامور پایش
۴	۲۲	کارمند ارشد مسئول دفتر مدیریت، خدمات مشترکین، مسئول درآمد، کارشناس حسابداری، مسئول توسعه و نوسازی، نماینده حقوقی
۵	۱۵۲	کارمند راننده، مسئول انبار، آبدارچی، نگهبان، باجه مشترکین، متصدی خدمات اداری، مسئول کنترل، بایگانی، مامور قرائت توزیع، اپراتور
	۳۴۹	مجموع



نمودار (۴-۵) پراکندگی سمت کارکنان

جدول (۴-۶) کدگذاری نوع استخدام

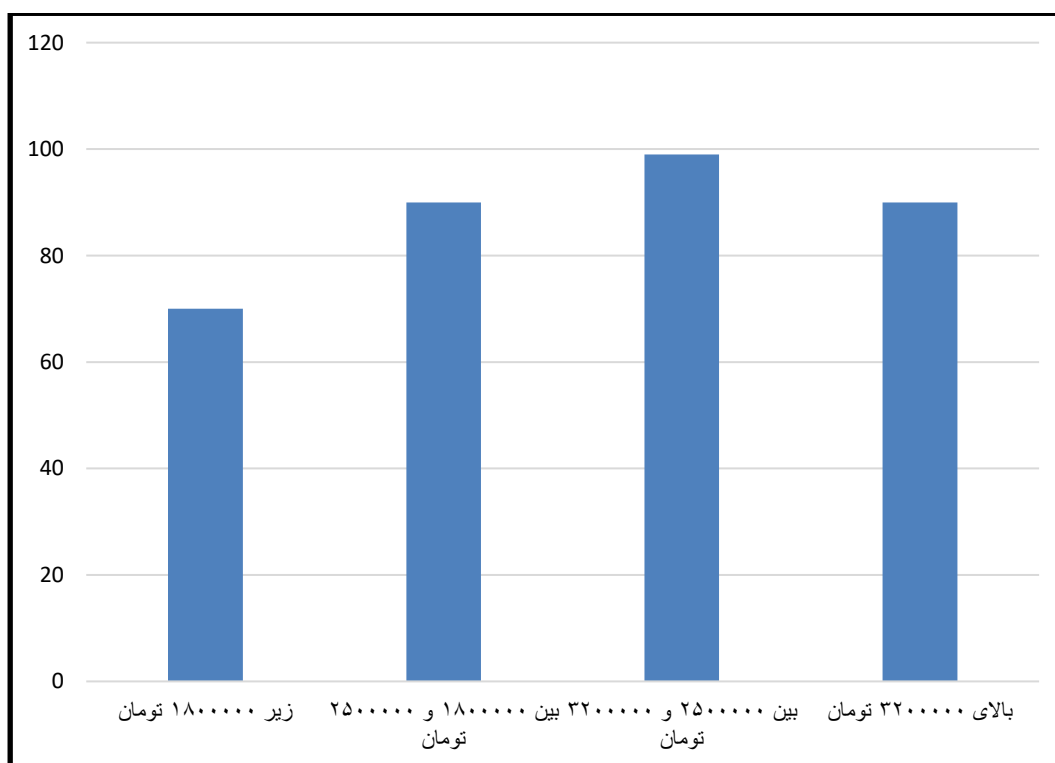
کد	تعداد افراد	نوع استخدام
۱	۴۶	شرکتی
۲	۲۱۶	پیمانکاری انسانی
۳	۶۹	حجمی
۴	۱۸	قراردادی
	۳۴۹	مجموع



نمودار (۴-۶) پراکندگی نوع استخدام

جدول (۷-۴) کدگذاری میزان حقوق

کد	تعداد افراد	میزان حقوق
۱	۷۰	زیر ۱۸۰۰۰۰۰ تومان
۲	۹۰	بین ۱۸۰۰۰۰۰ و ۲۵۰۰۰۰۰ تومان
۳	۹۹	بین ۲۵۰۰۰۰۰ و ۳۲۰۰۰۰۰ تومان
۴	۹۰	بالای ۳۲۰۰۰۰۰ تومان
	۳۴۹	مجموع



نمودار (۷-۴) پراکندگی نوع حقوق

جدول (۴-۸) کدگذاری جنسیت

کد	تعداد افراد	نوع جنسیت
۱	۳۰۹	مرد
۰	۴۰	زن
	۳۴۹	مجموع

جدول (۴-۹) کدگذاری وضعیت تاهل

کد	تعداد افراد	وضعیت تاهل
۱	۳۲۷	متاهل
۰	۲۲	مجرد
	۳۴۹	مجموع

در این مرحله داده‌ها و متغیرهایی لازم جهت دسته‌بندی کدگذاری شدند، متغیرهایی که در این قسمت وجود ندارند، نیازی به کدگذاری نداشتند و آماده ورود به نرم افزار بوده اند. توجه به این نکته ضروری است که همیشه در جمع آوری اطلاعات و وارد کردن داده‌ها به کامپیوتر، به دلیل دخالت انسان و وجود هرگونه خطای انسانی امکان از دست رفتن برخی اطلاعات و یا اشتباه وارد شدن آن‌ها وجود دارد. به همین دلیل برای انجام پروژه‌های داده‌کاوی باید این داده‌های از دست رفته و نویزها مدیریت شوند. همچنین نوع داده‌های ورودی باید تعیین شوند که آیا از نوع باینری یا اعداد متوالی و یا انواع دیگر داده‌ها هستند. مدیریت نقاط از دست رفته و داده‌های پرت اولین گام در پیش پردازش داده‌ها می‌باشد که در نرم افزار Rapid Miner انجام می‌شود.

همچنین در فاز پیش پردازش داده علاوه بر تعیین نوع داده‌ها و مدیریت نویزها و ... عملیات مختلف مورد نیاز بر روی داده‌ها را می‌توان انجام داد. در پروژه‌های مختلف، با توجه به هدف پژوهش کارهای مختلفی بر روی داده‌های ورودی انجام می‌شود. یکی از کارهایی که در این پژوهش بر روی داده‌های ورودی انجام شد به دست آوردن سن افراد به هنگام استخدام ایشان بود. در داده‌های اولیه سن فعلی افراد وجود داشت که تأثیری بر مدلسازی و عملکرد افراد نداشت. به همین جهت این مقدار از سال‌های خدمتشان کم شد تا سن ورود به سازمان بدست آید.

در حین پیش پردازش داده‌های اولیه که مهمترین قسمت در پروژه‌های داده‌کاوی است، باید به این نکته توجه داشته باشیم که هدف باید تک متغیری باشد. با توجه به این موضوع که در سنجش عملکرد کارکنان متغیرهای مختلفی دخیل است و نمی‌توان افراد را مثلاً بر حسب نمره ارزیابی عملکرد یا فقط تعداد ارتقا به دو گروه وفادار یا غیروفادار تقسیم کرد، پس باید متغیری تحت عنوان متغیر هدف تعریف شود که شامل تمام موارد تأثیرگذار بر انتخاب فرد به عنوان فرد وفادار یا غیروفادار باشد.

از آنجایی که هدف این پژوهش انتخاب افراد جهت استخدام در سازمان است، باید متغیر به این منظور تعریف شود تا به عنوان متغیر هدف به نرم‌افزار معرفی شود. متغیرهای زیادی در ساختن فیلد هدف دخیل هستند که با توجه به نظر خبرگان سازمان این متغیر به صورت فرمول زیر تعریف شد.

Target = 'Last Job Position' - 'Initial Job Position' >= 1 and 'In Work' = 1 and 'Rank' <= 4 and 'Last Education Position' - 'Initial Education Position' >= 1 and ('Age' >= 2 and 'Age' <= 3) and ('Years of Service' <= 10 and 'Type of Employment' <= 4) or ('Years of Service' > 10 and 'Type of Employment' <= 3) and 'Payment' >= 2 .

شرح فرمول بدین صورت است که اگر فردی شرایط زیر را دارا بود وفادار به سازمان است:

- ✓ حداقل اختلاف جایگاه شغلی ورود به سازمان و جایگاه فعلی او یک واحد باشد.
- ✓ در حال حاضر مشغول به کار در سازمان باشد.
- ✓ حداکثر سمت کارمند ۴ باشد.
- ✓ حداقل اختلاف مقطع تحصیلی ورود به سازمان و فعلی او یک واحد باشد.
- ✓ سن استخدام افراد بین ۲ و ۳ باشد.
- ✓ اگر سنوات از ۱۰ سال کمتر است حداکثر نوع استخدام ۴ یا اگر سنوات خدمت بیشتر از ۱۰ سال است حداکثر نوع استخدام ۳ باشد.
- ✓ حداقل جایگاه حقوقی ۲ باشد.

با توجه به این موضوع که پس از ایجاد مدل قرار است از آن جهت میزان وفاداری افراد استفاده شود داده‌ها جهت آموزش (ساخت مدل)، تست و ارزیابی باید پارتیشن بندی شوند و دقت مدل در پارتیشن تست مورد ارزیابی و تحلیل قرار بگیرد تا پس از مدل‌سازی به بهترین دقت در تست دست یافته شود و میزان خطا به هنگام استفاده از مدل پایین باشد. با در نظر گرفتن این موضوع که تعداد داده‌های اولیه در این پژوهش نسبت به اکثر پژوهش‌های داده‌کاوی کمتر است و تعداد داده‌های اولیه ۳۴۹ رکورد بود، درصد بیشتری از داده‌ها بعنوان داده‌های آموزش جهت ساخت مدل قرار داده شد تا مدل نسبتاً دقیق‌تری ساخته شود. داده‌های آموزشی این پژوهش ۹۰ درصد، داده‌های تست ۵ درصد و ۵ درصد باقی‌مانده از داده‌ها جهت ارزیابی مدل تقسیم‌بندی شدند. انتخاب داده‌ها برای قسمت‌های آموزش، ارزیابی و تست به صورت تصادفی توسط نرم افزار انجام می‌شود.

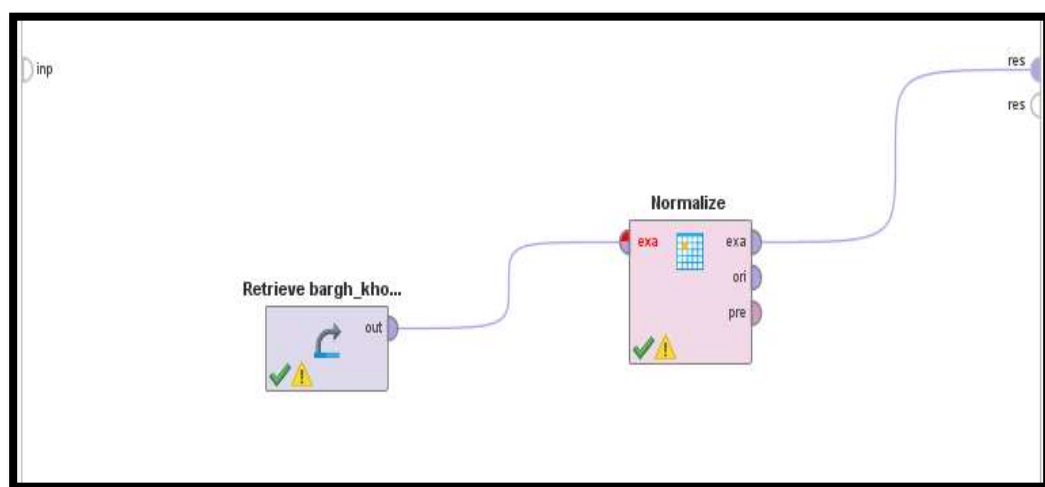
۴-۳) مدل‌سازی

همانطور که پیش‌تر بیان شد در این پژوهش از الگوریتم‌های درخت تصمیم ID3، C4.5 و CART، ماشین بردار پشتیبان و KNN برای مدل‌سازی استفاده می‌شود که در ادامه به بیان چگونگی مدل‌سازی هریک پرداخته می‌شود.

۴-۳-۱) پیش‌پردازش داده‌ها

اولین اقدام در کار با مجموعه داده‌ها تعیین وضعیت مقادیر از دست رفته می‌باشد. با توجه به اینکه در این مجموعه داده، مقادیر از دست‌رفته با صفر مقداردهی شده است آن را در مجموعه داده مشخص می‌کنیم و مقادیر را با علامت سوال «؟» جایگزین می‌کنیم. این کار باعث می‌شود که بتوان عملیات‌های بعدی برای حذف یا جایگذاری این مقادیر را انجام داد. در این مرحله تعداد مقادیر از دست‌رفته برای هر رکورد مشخص می‌شود. پس از مشخص شدن تعداد آن برای هر رکورد، آن‌هایی که ۲ و بیشتر از ۲ مشخصه‌ی آن‌ها مقداری نداشته باشد حذف می‌شود. پس از یافتن مقادیر از دست رفته و جایگذاری آن‌ها با علامت سوال بر اساس مجموعه‌ای از شرط‌های منطقی داده‌هایی که شرایط ما را نقض می‌کنند (۴ مورد) را حذف می‌کنیم.

ساختار نرمال سازی داده ها در نرم افزار در شکل (۴-۱) نشان داده شده است.



شکل (۴-۱) نرمال سازی پایگاه داده

شرایط زیر برای فیلتر کردن اطلاعات استفاده شده است:

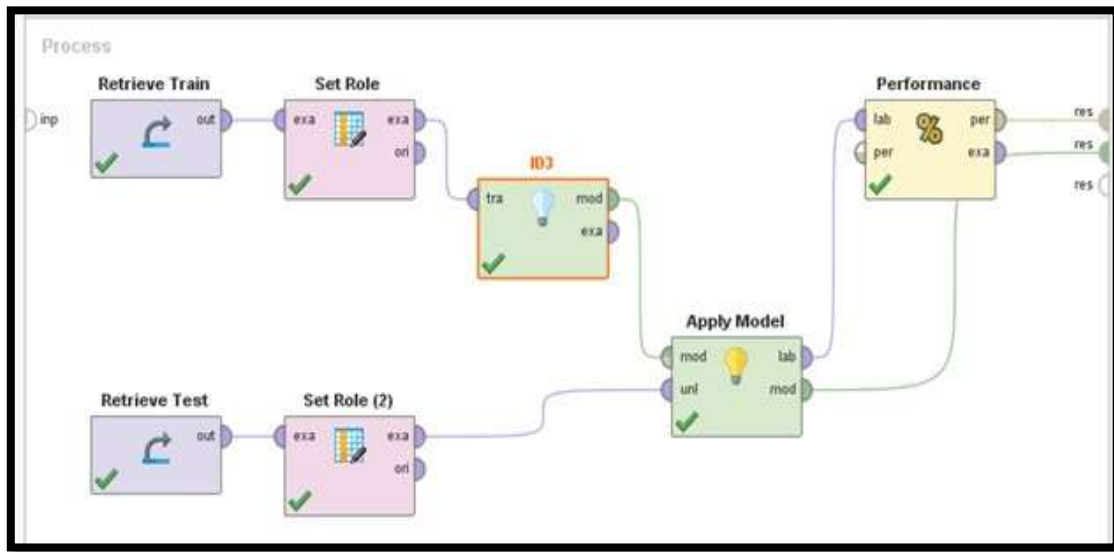
۱. شخص فاقد رشته تحصیلی باشد.
۲. نوع استخدامی کمتر یا مساوی ۲ باشد.
۳. سن شخص کمتر یا مساوی ۲۵ باشد.

اجرای فیلتر باعث حذف ۱۱ مورد از مجموعه داده های انتخابی می شود. پس در مجموع ۱۵ رکورد از مجموعه انتخابی حذف می شود و ۳۳۴ رکورد باقی می ماند. مقادیر نامشخص در داده های باقیمانده را با مقدار میانگین همان رده جایگذاری می کنیم. پس از جایگذاری مقادیر نامشخص برای حذف تاثیر وزن ویژگی های عددی آنها را در بازه ۰ تا ۱ نرمال سازی می کنیم. با استفاده از فاصله ی اقلیدسی داده های پرت شناسایی می شوند. در این مورد هیچ رکورد به عنوان داده ی پرت شناسایی و حذف نشد. در پایان ۳۳۴ داده برای بررسی انتخاب می شوند.

لازم به ذکر است که از دیگر متغیرهای مورد نظر جهت بررسی، متغیر بومی بودن بود که به علت اینکه تمام نیروهای به کار رفته در سازمان از استان بودند و صد در صد جامعه آماری را پوشش می داد، در نظر گرفته نشد.

۴-۳-۲) درخت تصمیم ID3

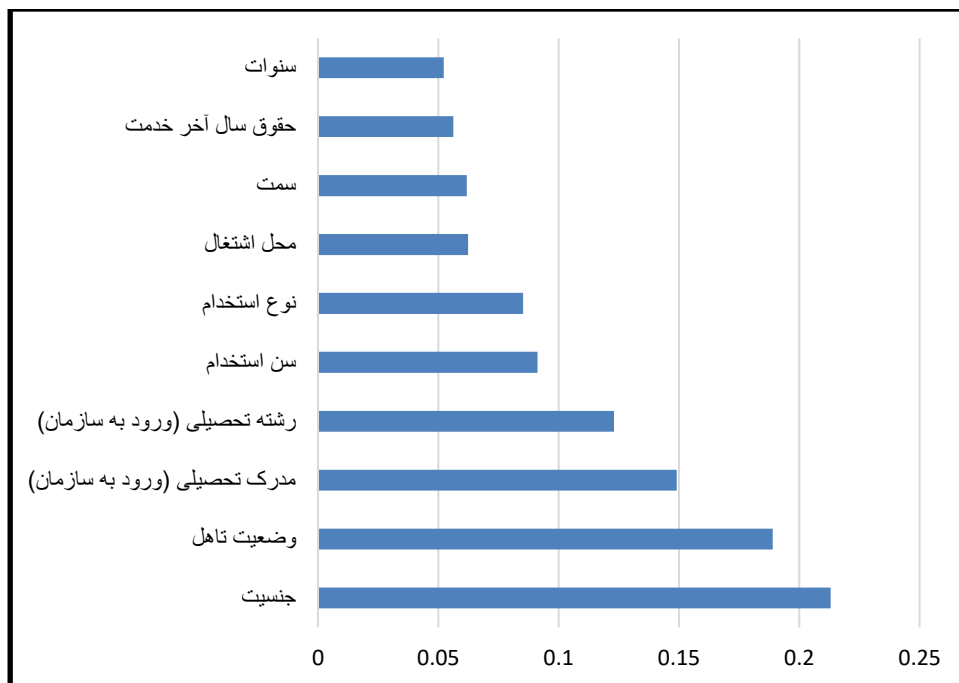
درخت‌های تصمیم روشی برای نمایش یک سری از قوانین هستند که منتهی به یک رده یا مقدار می‌شوند. براساس الگوریتم، ممکن است دو یا تعداد بیشتری شاخه داشته باشد. هر شاخه منجر به نود تصمیم دیگر یا یک نود برگ می‌شود. با پیمایش یک درخت تصمیم از ریشه به پایین به یک مورد یک رده یا مقدار می‌رسیم. هر نود از داده‌های یک مورد برای تصمیم‌گیری درباره آن انشعاب استفاده می‌کند.



شکل (۴-۲) مراحل ساخت درخت تصمیم ID3

شکل (۴-۲) نودهای استفاده شده به هنگام پیش پردازش داده‌ها و عملیات انجام شده جهت آماده سازی داده‌های ورودی به الگوریتم را نشان می‌دهد. الگوریتم به کار رفته در این قسمت درخت تصمیم از نوع ID3 است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد.

در نمودار (۴-۸) میزان تأثیر و اهمیت هر یک از متغیرهای ورودی در ساخت مدل نشان داده شده است. طبق این نمودار سه متغیر جنسیت و وضعیت تاهل و مدرک تحصیلی ترتیب بیشترین تأثیر را در ساخت مدل داشته اند. پس از آن‌ها متغیرهای، رشته تحصیلی، سن استخدام، نوع استخدام اهمیت بیشتری داشتند و میزان تأثیر متغیرهای دیگر تقریباً یک اندازه بوده و اهمیت کمتری در ساخت مدل داشته اند.



نمودار (۴-۸) نمودار اهمیت متغیرها در ساخت درخت تصمیم ID3

جدول (۴-۱۰) اهمیت متغیرها

متغیر گره	ضریب اهمیت
جنسیت	۰.۲۱۳
وضعیت تاهل	۰.۱۸۹
مدرک تحصیلی (ورود به سازمان)	۰.۱۴۹
رشته تحصیلی (ورود به سازمان)	۰.۱۲۳
سن استخدام	۰.۰۹۱۲
نوع استخدام	۰.۰۸۵۲
محل اشتغال	۰.۰۶۲۳
سمت	۰.۰۶۱۸
حقوق سال آخر خدمت	۰.۰۵۶۳
سنوات	۰.۰۵۲۲

جدول (۴-۱۰) نشان دهنده ضریب اهمیت هر کدام از متغیرها می باشد در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت ترین متغیر به سمت کم اهمیت ترین متغیر مرتب سازی شده اند. مطابق این جدول ضریب تأثیر متغیرهای کم اهمیت شامل حقوق سال آخر خدمت و سنوات و متغیرها با بیشترین تأثیرگذاری متغیرهای جنسیت و وضعیت تاهل می باشند.

```

sex = 0
| mar = 0: Yes {Yes=3, No=1}
| mar = 1
| | edu = 3: Yes {Yes=9, No=0}
| | edu = 4: Yes {Yes=5, No=2}
| | edu = 5
| | | maj = 1
| | | | age = 1: Yes {Yes=2, No=1}
| | | | age = 2: Yes {Yes=7, No=0}
| | | | age = 3: Yes {Yes=2, No=0}
| | | maj = 3: Yes {Yes=4, No=0}
| | | maj = 4: Yes {Yes=1, No=0}
| | | maj = 5: Yes {Yes=2, No=2}
sex = 1
| mar = 0
| | edu = 3: Yes {Yes=9, No=0}
| | edu = 4: Yes {Yes=1, No=0}
| | edu = 5
| | | maj = 1
| | | | age = 2: Yes {Yes=3, No=3}
| | | | age = 3: Yes {Yes=4, No=2}
| mar = 1
| | edu = 2: Yes {Yes=8, No=0}
| | edu = 3
| | | maj = 1
| | | | age = 1
| | | | | emplo = 2
| | | | | | dep = 3: Yes {Yes=3, No=1}
| | | | | | dep = 5
| | | | | | job = 3
| | | | | | | payment = 1: Yes {Yes=3, No=1}
| | | | | | | payment = 2: Yes {Yes=4, No=0}
| | | | | | | payment = 3: Yes {Yes=1, No=0}
| | | | | | | payment = 4: Yes {Yes=8, No=0}
| | | | | | job = 5: Yes {Yes=6, No=1}
| | | | | emplo = 3
| | | | | | dep = 1
| | | | | | | job = 3: No {Yes=0, No=1}
| | | | | | | job = 5
| | | | | | | payment = 2: No {Yes=1, No=1}
| | | | | | | payment = 3: Yes {Yes=2, No=1}
| | | | | | | payment = 4: Yes {Yes=2, No=1}
| | | | age = 2
| | | | | emplo = 2
| | | | | | dep = 3: Yes {Yes=2, No=0}
| | | | | | dep = 5
| | | | | | job = 3
| | | | | | | payment = 1: Yes {Yes=3, No=1}
| | | | | | | payment = 2: Yes {Yes=3, No=1}
| | | | | | | payment = 3: Yes {Yes=1, No=0}
| | | | | | | payment = 4: Yes {Yes=4, No=0}
| | | | | | job = 5: Yes {Yes=6, No=1}
| | | | | emplo = 3
| | | | | | payment = 1: No {Yes=0, No=1}
| | | | | | payment = 2: No {Yes=1, No=1}
| | | | | | payment = 3: Yes {Yes=5, No=0}
| | | | | | payment = 4: Yes {Yes=2, No=0}
| | | age = 3
| | | | emplo = 2

```

						dep = 3: Yes {Yes=4, No=1}
						dep = 5
						job = 3
						payment = 1: No {Yes=1, No=2}
						payment = 2: Yes {Yes=2, No=1}
						payment = 3: Yes {Yes=3, No=2}
						payment = 4: No {Yes=0, No=1}
						job = 5: Yes {Yes=8, No=0}
						emplo = 3: No {Yes=2, No=2}
					maj = 4	
					age = 1	
					emplo = 2: Yes {Yes=6, No=1}	
					emplo = 3: Yes {Yes=1, No=0}	
					age = 2: Yes {Yes=2, No=1}	
					age = 3: Yes {Yes=4, No=1}	
					maj = 5	
					age = 1: Yes {Yes=6, No=1}	
					age = 2: Yes {Yes=3, No=3}	
					age = 3: Yes {Yes=2, No=1}	
					edu = 4	
					maj = 1	
					job = 2: Yes {Yes=2, No=1}	
					job = 3: Yes {Yes=11, No=0}	
					job = 4: No {Yes=0, No=1}	
					job = 5: Yes {Yes=15, No=0}	
					maj = 2: No {Yes=0, No=1}	
					maj = 5: No {Yes=1, No=2}	
					edu = 5	
					maj = 1	
					age = 1	
					emplo = 2: No {Yes=2, No=2}	
					emplo = 3: Yes {Yes=1, No=0}	
					emplo = 4: Yes {Yes=6, No=0}	
					age = 2	
					emplo = 2: Yes {Yes=8, No=0}	
					emplo = 3: Yes {Yes=1, No=0}	
					emplo = 4: Yes {Yes=5, No=2}	
					age = 3	
					emplo = 2	
					dep = 3: Yes {Yes=2, No=0}	
					dep = 5	
					job = 2: Yes {Yes=3, No=1}	
					job = 3: Yes {Yes=2, No=0}	
					job = 5: Yes {Yes=2, No=2}	
					emplo = 3: Yes {Yes=2, No=0}	
					emplo = 4: Yes {Yes=2, No=1}	
					maj = 2: Yes {Yes=2, No=1}	
					maj = 3	
					age = 1: No {Yes=1, No=5}	
					age = 2: Yes {Yes=2, No=0}	
					maj = 4: Yes {Yes=2, No=1}	
					maj = 5	
					age = 1: No {Yes=1, No=1}	
					age = 2: Yes {Yes=2, No=0}	
					age = 3: Yes {Yes=4, No=0}	

نمودار (۹-۴) شمایک درخت تصمیم ID3

نمودار (۴-۹) نشان دهنده درخت ساخته شده توسط نرم افزار است و نوع تقسیم بندی شاخه ها و برگ ها را به خوبی نشان می دهد. در درخت ایجاد شده گره ریشه نشان دهنده درصد افرادی است که در سازمان مانده اند یا نسبت به سازمان وفادار بوده اند. مطابق اطلاعات موجود در این گره درصد افرادی که وفادار بوده اند ۸۸.۲۳ درصد کارکنان را شامل می شود. اولین متغیری که به ازای آن انشعاب صورت گرفته و گره های فرزند ایجاد شده اند متغیر جنسیت می باشد. طبق این مدل رشته های تحصیلی موجود در گروه ۱ و ۶ که شامل رشته های مهندسی و فاقد رشته تحصیلی می شوند در یک شاخه قرار دارند که ۱۴.۲ درصد آن ها غیروفا دار و ۸۵.۸ درصدشان وفادار بوده اند. همچنین رشته های تحصیلی موجود در گروه های ۲، ۳، ۴ و ۵ نیز در یک شاخه قرار گرفته اند که ۴۰.۹ درصدشان غیروفا دار و ۵۹.۱ درصد وفادار بوده اند. بر طبق این تقسیم بندی ۸۱.۸ درصد افراد مجرد دارای رشته تحصیلی ۱ یا ۶ غیروفا دار بوده اند و تنها ۱۸.۲ درصد آن ها وفادار بوده اند. همچنین مطابق این تقسیم بندی ۱۴.۰۷ درصد افراد متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۸۵.۹۳ درصد وفادار به سازمان بوده اند. در ادامه این تقسیم بندی افراد متاهل براساس جنسیت تقسیم بندی شده اند که ۸ درصد خانم های متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۹۲ درصدشان وفادار بوده اند و درمورد آقایان ۱۵ درصد غیروفا دار و ۸۵ وفادار بوده اند. سپس خانم ها بر این اساس به شاخه های دیگر تقسیم شده اند و این روند تا رسیدن به برگ ها ادامه پیدا می کند. بر اساس دقت آنالیز مدل در قسمت آموزش ۸۷ درصد است. همچنین عمق درخت تصمیم ساخته شده ۸ است.

جدول (۴-۱۱) مقایسه دقت در مدل های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم ID3

ارزیابی		تست		آموزش		دسته
۱۵	۸۸.۲۳٪	۱۴	۸۲.۳۵٪	۲۶۱	۸۷٪	درست
۲	۱۲.۷۷٪	۳	۱۷.۶۵٪	۳۹	۱۳٪	نادرست
۱۷	-	۱۷	-	۳۰۰	-	مجموع

در جدول (۴-۱۱) نشان داده شده که دقت مدل در قسمت تست ۸۲.۳۵ درصد و دقت در قسمت ارزیابی ۸۸.۲۳ درصد می‌باشد. تعداد داده‌هایی که در قسمت تست و ارزیابی مورد سنجش قرار گرفته اند داده‌های مربوط به ۳۴ نفر بوده است که برای ۲۹ نفر آن‌ها به درستی تشخیص داده شده که برای کار در سازمان وفادار هستند یا خیر.

جدول (۴-۱۲) ماتریس اغتشاش درخت تصمیم ID3

دسته	Negative	Positive	نوع تشخیص
آموزش	۲۴	۱۹۵	True
	۱۵	۶۶	False
تست	۲	۱۰	True
	۱	۴	False
ارزیابی	۱	۱۳	True
	۱	۲	False

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 74.71\%$$

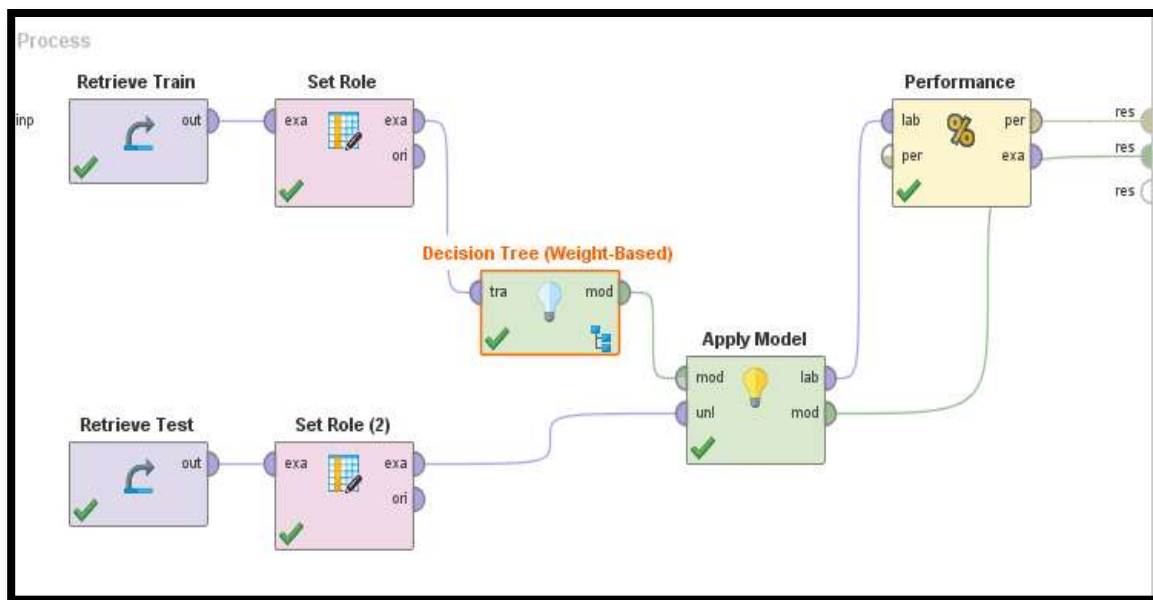
$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 92.85\%$$

۴-۳-۳) درخت تصمیم C4.5

الگوریتم C4.5 می‌تواند مقادیر گسسته یا پیوسته را در ویژگی‌ها درک کند. در صورتی که الگوریتم ID3 نمی‌تواند تفاوت مقادیر عددی پیوسته را درک کند. همچنین الگوریتم C4.5 قادر است تا مقادیری که موجود نیستند را هم تحمل کند. به این داده‌ها، مقادیر ناموجود^۱ نیز می‌گویند. سومین موردی که باعث بهینه شدن الگوریتم C4.5 نسبت به ID3 می‌شود، عملیات هرس کردن^۲ جهت جلوگیری از Overfitting است. عملیات هرس در الگوریتم C4.5 مدل را به نقطه بهینه که زیاد پیچیده نباشد، می‌رساند. مورد آخر که C4.5 را از دیگر الگوریتم‌ها متمایز می‌کند، وزن‌دهی به ویژگی‌ها است. C4.5 این قابلیت را دارد که وزن‌های غیریکسانی به برخی از ویژگی‌ها بدهد.

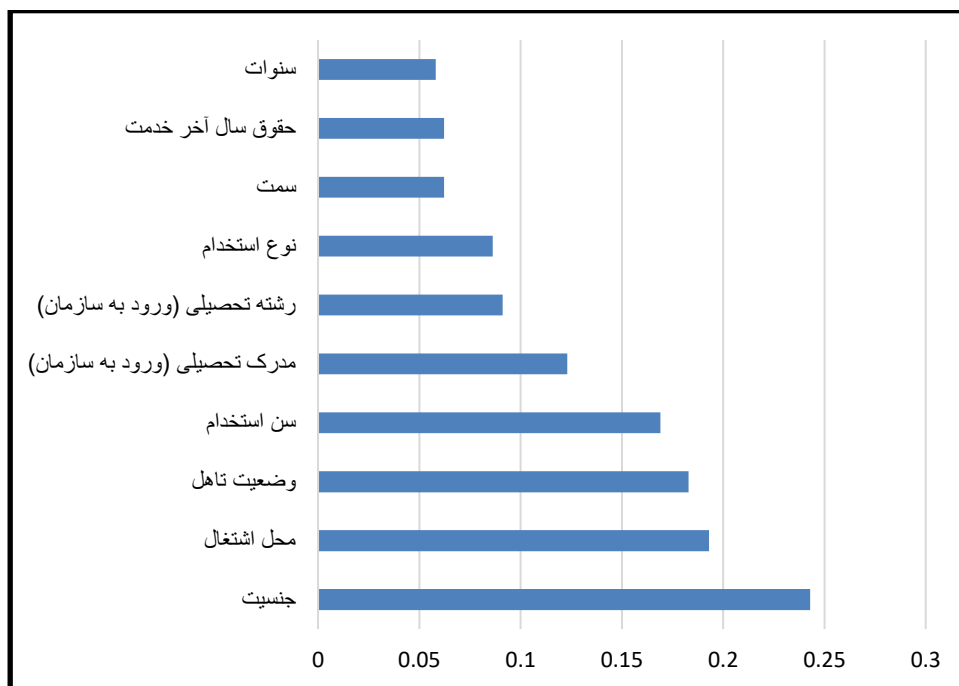
^۱ Missing Values

^۲ Pruning



شکل (۳-۴) مراحل ساخت درخت تصمیم C4.5

شکل (۳-۴) نودهای استفاده شده به هنگام پیش پردازش داده‌ها و عملیات انجام شده جهت آماده سازی داده‌های ورودی به الگوریتم را نشان می‌دهد. الگوریتم به کار رفته در این قسمت درخت تصمیم از نوع C4.5 است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد.



نمودار (۴-۱) نمودار اهمیت متغیرها در ساخت الگوریتم درخت تصمیم C4.5

در نمودار (۴-۱۰) میزان تأثیر و اهمیت هر یک از متغیرهای ورودی در ساخت مدل نشان داده شده است. طبق این نمودار سه متغیر جنسیت و محل اشتغال و وضعیت تاهل به ترتیب بیشترین تأثیر را در ساخت مدل داشته اند. پس از آن‌ها متغیرهای سن استخدام، مدرک و رشته تحصیلی و نوع استخدام اهمیت بیشتری داشتند و میزان تأثیر متغیرهای دیگر تقریباً یک اندازه بوده و اهمیت کمتری در ساخت مدل داشته اند.

جدول (۴-۱۳) ضریب اهمیت متغیرها

متغیر گره	ضریب اهمیت
جنسیت	۰.۲۴۳
محل اشتغال	۰.۱۹۳
وضعیت تاهل	۰.۱۸۳
سن استخدام	۰.۱۶۹
مدرک تحصیلی (ورود به سازمان)	۰.۱۲۳
رشته تحصیلی (ورود به سازمان)	۰.۰۹۱۲
نوع استخدام	۰.۰۸۶۲
سمت	۰.۰۶۲۳
حقوق سال آخر خدمت	۰.۰۶۲۲
سنوات	۰.۰۵۸۲

جدول (۴-۱۳) نشان دهنده ضریب اهمیت هر کدام از متغیرها می‌باشد در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت‌ترین متغیر به سمت کم اهمیت‌ترین متغیر مرتب سازی شده اند. مطابق این جدول ضریب تأثیر متغیرهای کم اهمیت شامل حقوق سال آخر خدمت و سنوات است که نشان دهنده اهمیت کم این متغیرها در ساخت مدل C4.5 می‌باشد. همچنین با توجه به جدول میزان اهمیت متغیرهای چهارم تا هشتم که شامل متغیرهای سن استخدام، مدرک و رشته تحصیلی، نوع استخدام و سمت می‌باشد در محدوده ۰.۰۶۲ تا ۰.۱۷ است. طبق جدول (۴-۱۳) بیشترین متغیرهای تأثیرگذار یعنی متغیرهای جنسیت با ضریب اهمیت ۰.۲۴۳ و محل اشتغال با ضریب ۰.۱۹۳ می‌باشند که نشان دهنده اهمیت بالای این دو متغیر نسبت به سایر متغیرهای ورودی در ساخت مدل است.

```

sex = 0
|
|   dep = 1
|   |
|   |   mar = 0
|   |   |
|   |   |   age = 1: Yes {Yes=1, No=0}
|   |   |   age = 2: Yes {Yes=1, No=0}
|   |   |   age = 3
|   |   |   |   edu = 5
|   |   |   |   |   maj = 1
|   |   |   |   |   |   emplo = 3
|   |   |   |   |   |   |   job = 5
|   |   |   |   |   |   |   |   payment = 1: No {Yes=1, No=1}
|   |   |   |   |   |   |   |   |   mar = 1: Yes {Yes=18, No=0}
|   |   |   |   |   |   |   |   |   |   dep = 2: Yes {Yes=1, No=0}
|   |   |   |   |   |   |   |   |   |   |   dep = 4
|   |   |   |   |   |   |   |   |   |   |   |   edu = 3: Yes {Yes=5, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   edu = 4: Yes {Yes=3, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   edu = 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 2: Yes {Yes=1, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 3
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   payment = 2: No {Yes=0, No=2}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   payment = 3: Yes {Yes=1, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   dep = 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   edu = 4: No {Yes=0, No=2}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   edu = 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   mar = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   maj = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   emplo = 4
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   job = 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   payment = 2: No {Yes=1, No=1}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 2: Yes {Yes=1, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 3: Yes {Yes=1, No=0}
sex = 1
|
|   mar = 0
|   |
|   |   edu = 3: Yes {Yes=9, No=0}
|   |   edu = 4: Yes {Yes=1, No=0}
|   |   edu = 5
|   |   |   emplo = 2
|   |   |   |   payment = 1
|   |   |   |   |   age = 2: No {Yes=0, No=2}
|   |   |   |   |   age = 3
|   |   |   |   |   |   maj = 1
|   |   |   |   |   |   |   dep = 5
|   |   |   |   |   |   |   |   job = 5: Yes {Yes=1, No=1}
|   |   |   |   |   |   |   |   |   payment = 2: No {Yes=0, No=1}
|   |   |   |   |   |   |   |   |   payment = 4: Yes {Yes=3, No=0}
|   |   |   |   |   |   |   |   |   |   emplo = 3: No {Yes=0, No=1}
|   |   |   |   |   |   |   |   |   |   |   emplo = 4: Yes {Yes=3, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   mar = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   edu = 2: Yes {Yes=8, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   edu = 3
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   maj = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   age = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   emplo = 2
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   payment = 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   dep = 3: Yes {Yes=1, No=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   dep = 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   job = 3: Yes {Yes=3, No=1}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   payment = 2

```



```

| | | | | payment = 2
| | | | | | age = 1
| | | | | | | dep = 3
| | | | | | | | job = 5: Yes {Yes=4, No=1}
| | | | | | | age = 2: Yes {Yes=1, No=0}
| | | | | | | age = 3: Yes {Yes=2, No=0}
| | | | | | payment = 3
| | | | | | | age = 1: Yes {Yes=2, No=0}
| | | | | | | age = 2: No {Yes=0, No=1}
| | | | | | | age = 3: No {Yes=0, No=1}
| | | | | | emplo = 3: Yes {Yes=4, No=0}
| | | | | maj = 5
| | | | | | emplo = 2
| | | | | | | age = 1
| | | | | | | | payment = 1: Yes {Yes=1, No=0}
| | | | | | | | payment = 2: No {Yes=0, No=1}
| | | | | | | | payment = 3: Yes {Yes=1, No=0}
| | | | | | | | payment = 4: Yes {Yes=2, No=0}
| | | | | | | age = 2
| | | | | | | | dep = 3
| | | | | | | | | payment = 3
| | | | | | | | | | job = 5: Yes {Yes=1, No=1}
| | | | | | | | | payment = 4: No {Yes=0, No=1}
| | | | | | | | dep = 4: No {Yes=0, No=1}
| | | | | | | age = 3
| | | | | | | | payment = 3: Yes {Yes=1, No=0}
| | | | | | | | payment = 4: No {Yes=0, No=1}
| | | | | | | emplo = 3: Yes {Yes=5, No=0}
| | | | | edu = 4
| | | | | | maj = 1
| | | | | | | job = 2
| | | | | | | | age = 1
| | | | | | | | | emplo = 2
| | | | | | | | | | dep = 5
| | | | | | | | | | | payment = 1: No {Yes=1, No=1}
| | | | | | | | | | age = 2: Yes {Yes=1, No=0}
| | | | | | | | job = 3: Yes {Yes=11, No=0}
| | | | | | | | job = 4: No {Yes=0, No=1}
| | | | | | | | job = 5: Yes {Yes=15, No=0}
| | | | | | | maj = 2: No {Yes=0, No=1}
| | | | | | | maj = 5
| | | | | | | | payment = 1: No {Yes=0, No=2}
| | | | | | | | payment = 3: Yes {Yes=1, No=0}
| | | | | | edu = 5
| | | | | | | age = 1
| | | | | | | | emplo = 2
| | | | | | | | | job = 2: No {Yes=0, No=1}
| | | | | | | | | job = 3: Yes {Yes=1, No=0}
| | | | | | | | | job = 4
| | | | | | | | | | maj = 4
| | | | | | | | | | | dep = 2
| | | | | | | | | | | | payment = 1: No {Yes=1, No=1}
| | | | | | | | | | job = 5
| | | | | | | | | | | maj = 1
| | | | | | | | | | | | dep = 5
| | | | | | | | | | | | | payment = 4: No {Yes=1, No=1}
| | | | | | | | | | | maj = 2
| | | | | | | | | | | | payment = 1: No {Yes=0, No=1}
| | | | | | | | | | | | payment = 2: Yes {Yes=1, No=0}

```


بر طبق این تقسیم‌بندی ۹۰.۹ درصد افراد مجرد دارای رشته تحصیلی ۱ یا ۶ غیروفا دار بوده اند و تنها ۹.۱ درصد آن‌ها وفادار بوده اند. همچنین مطابق این تقسیم بندی ۷.۹۶ درصد افراد متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۹۲.۰۴ درصد وفادار بوده اند. در ادامه این تقسیم‌بندی افراد متاهل براساس جنسیت تقسیم‌بندی شده اند که ۵ درصد خانم‌های متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۹۵ درصدشان وفادار بوده اند و در مورد آقایان ۱۷ درصد غیروفا دار و ۸۳ درصد وفادار بوده اند. سپس خانم‌ها بر اساس به شاخه‌های دیگر تقسیم شده اند و این روند تا رسیدن به برگ‌ها ادامه پیدا می‌کند. بر اساس دقت آنالیز مدل در قسمت آموزش ۸۸.۲۳ درصد است.

جدول (۴-۱۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم C4.5

دسته	آموزش		تست		ارزیابی	
درست	۸۹٪	۲۶۷	۸۸.۲۳٪	۱۵	۹۴.۱۱٪	۱۶
نادرست	۱۱٪	۳۳	۱۲.۷۷٪	۲	۵.۸۹٪	۱
مجموع	-	۳۰۰	-	۱۷	-	۱۷

در جدول (۴-۱۴) نشان داده شده است که دقت مدل در قسمت تست ۸۸.۲۳ درصد و دقت در قسمت ارزیابی ۹۴.۱۱ درصد می‌باشد. تعداد داده‌هایی که در قسمت تست و ارزیابی مورد سنجش قرار گرفته اند داده‌های مربوط به ۳۴ نفر بوده است که برای ۳۱ نفر به درستی تشخیص داده شده که برای کار در سازمان وفادار هستند یا خیر.

جدول (۴-۱۵) ماتریس اغتشاش درخت تصمیم C4.5

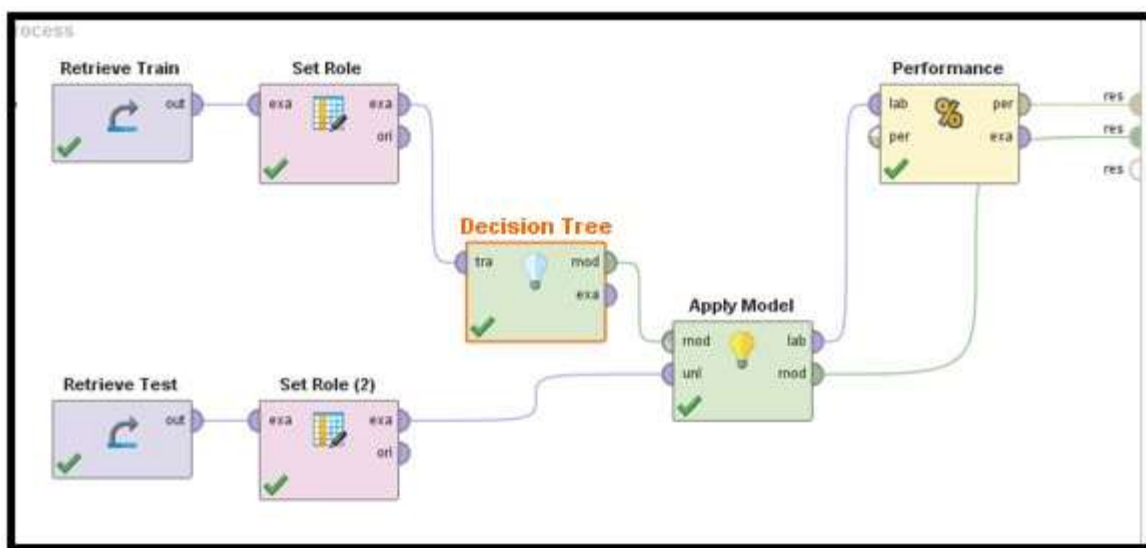
دسته	Positive		Negative		نوع تشخیص
آموزش	۲۰۵	۲۶	۶۲	۷	True
	۱۳	۲	۲	۰	False
تست	۱۶	۰	۰	۱	True
	۰	۰	۰	۰	False
ارزیابی	۰	۰	۰	۰	True
	۰	۰	۰	۰	False

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 76.77\%$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 96.69\%$$

۴-۳-۴) درخت تصمیم CART

یکی از محبوب‌ترین و در عین حال ساده‌ترین الگوریتم‌های درخت‌های تصمیم، درخت تصمیم CART است که کاربردهای زیادی در طبقه‌بندی و رگرسیون دارد. CART بر اساس درخت‌های دودویی بنا نهاده شده است. الگوریتم درخت تصمیم CART برای ساخت درخت تصمیم، داده‌ها را به قسمت‌های دوتایی تقسیم کرده و بر اساس آن‌ها درخت دودویی را می‌سازد. CART برای تصمیم‌گیری چگونگی انتخاب گره‌های درخت از معیاری به نام معیار شاخص جینی^۱ استفاده می‌کند. هر چقدر یک ویژگی شاخص جینی کمتری داشته باشد آن ویژگی اطلاعات بیشتری را دارد و می‌تواند در درخت ساخته شده، بالاتر (یعنی نزدیک به ریشه) قرار بگیرد. تفاوت شاخص جینی و آنترپی^۲ در این است که معمولاً شاخص جینی برای داده‌هایی که دارای قسمت بزرگتر هستند به درد می‌خورد این در حالی است که آنترپی به درد داده‌هایی می‌خورد که قسمت‌های کوچک زیادی دارند که مقادیر یکتا در آن‌ها بیشتر است. جهت جلوگیری از Overfit شدن درخت تصمیم CART، می‌توان از یک شرط توقف استفاده کرد.

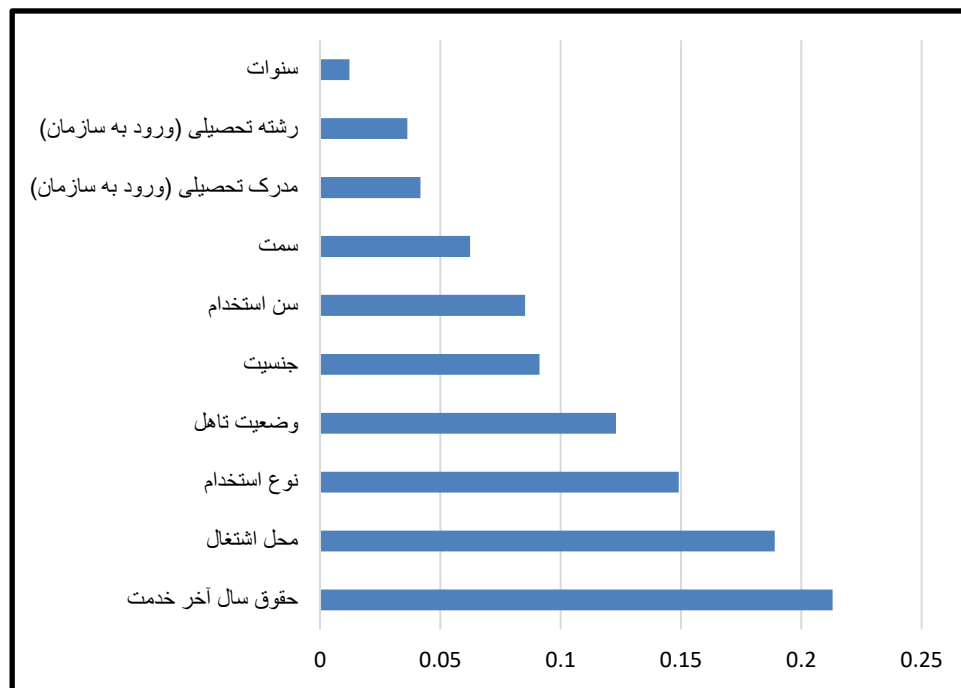


شکل (۴-۴) مراحل ساخت درخت تصمیم CART

^۱ Gini

^۲ Entropy

شکل (۴-۴) نودهای استفاده شده به هنگام پیش پردازش داده‌ها و عملیات انجام شده جهت آماده سازی داده‌های ورودی به الگوریتم را نشان می‌دهد. الگوریتم به کار رفته در این قسمت درخت تصمیم از نوع CART است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد.



نمودار (۴-۱۲) نمودار اهمیت متغیرها در ساخت الگوریتم درخت تصمیم CART

در نمودار (۴-۱۲) میزان تأثیر و اهمیت هر یک از متغیرهای ورودی در ساخت مدل نشان داده شده است. طبق این نمودار سه متغیر حقوق سال آخر خدمت و محل اشتغال و نوع استخدام به ترتیب بیشترین تأثیر را در ساخت مدل داشته اند. پس از آن‌ها متغیرهای وضعیت تاهل، جنسیت، سن استخدام، سمت و مدرک و رشته تحصیلی اهمیت بیشتری داشتند و میزان تأثیر متغیر سنوات بسیار ناچیز بود و اهمیت کمتری در ساخت مدل داشته اند. جدول (۴-۱۶) نشان دهنده ضریب اهمیت هر کدام از متغیرها می‌باشد در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت‌ترین متغیر به سمت کم اهمیت‌ترین متغیر مرتب سازی شده اند. مطابق این جدول ضریب تأثیر متغیرهای کم اهمیت شامل سنوات، رشته و مدرک تحصیلی است که نشان دهنده اهمیت کم این متغیرها در ساخت مدل CART می‌باشد.

جدول (۴-۱۶) ضریب اهمیت متغیرها

متغیر گره	ضریب اهمیت
حقوق سال آخر خدمت	۰.۲۱۳
محل اشتغال	۰.۱۸۹
نوع استخدام	۰.۱۴۹
وضعیت تاهل	۰.۱۲۳
جنسیت	۰.۰۹۱۲
سن استخدام	۰.۰۸۵۲
سمت	۰.۰۶۲۳
مدرک تحصیلی (ورود به سازمان)	۰.۰۴۱۸
رشته تحصیلی (ورود به سازمان)	۰.۰۳۶۳
سنوات	۰.۰۱۲۲

همچنین با توجه به جدول میزان اهمیت متغیرهای سوم تا هشتم که شامل متغیرهای نوع استخدام، وضعیت تاهل، جنسیت، سن استخدام، سمت و مدرک تحصیلی می‌باشد در محدوده ۰.۰۴۲ تا ۰.۱۵ است. طبق جدول (۴-۱۶) بیشترین متغیرهای تأثیرگذار یعنی متغیرهای حقوق سال آخر خدمت با ضریب اهمیت ۰.۲۱۳ و محل اشتغال با ضریب ۰.۱۸۹ می‌باشند که نشان دهنده اهمیت برالای این دو متغیر نسبت به سایر متغیرهای ورودی در ساخت مدل است.

نمودار (۴-۱۳) نشان دهنده درخت ساخته شده توسط نرم افزار است و نوع تقسیم بندی شاخه‌ها و برگ‌ها را به خوبی نشان می‌دهد. ساختار کلی درخت CART ساخته شده به صورت نمایش داده شده می‌باشد. مطابق اطلاعات موجود در این گره درصد افرادی که وفادار بوده اند ۷۶.۴۷ درصد کارکنان را شامل می‌شود. اولین متغیری که به ازای آن انشعاب صورت گرفته و گره‌های فرزند ایجاد شده اند متغیر حقوق سال آخر خدمت می‌باشد. طبق این مدل رشته‌های تحصیلی موجود در گروه ۱ و ۶ که شامل رشته‌های مهندسی و فاقد رشته تحصیلی می‌شوند که ۲۱.۲ درصد آن‌ها غیروفادار و ۷۸.۸ درصدشان وفادار بوده اند. همچنین رشته‌های تحصیلی موجود در گروه‌های ۲، ۳، ۴ و ۵ نیز در یک شاخه قرار گرفته اند که ۳۸.۵ درصدشان غیروفادار و ۶۱.۵ درصد وفادار بوده اند.

```

payment = 1
|   dep = 1
|   |   age = 1
|   |   |   job = 3: No {Yes=1, No=1}
|   |   |   job = 5: Yes {Yes=2, No=1}
|   |   age = 2: No {Yes=1, No=2}
|   |   age = 3: Yes {Yes=5, No=1}
|   dep = 2: No {Yes=1, No=1}
|   dep = 3: Yes {Yes=11, No=1}
|   dep = 5
|   |   age = 1
|   |   |   edu = 3: Yes {Yes=3, No=1}
|   |   |   edu = 4: Yes {Yes=4, No=2}
|   |   |   edu = 5: No {Yes=1, No=2}
|   |   age = 2
|   |   |   job = 2: Yes {Yes=3, No=0}
|   |   |   job = 3
|   |   |   |   edu = 3: Yes {Yes=3, No=1}
|   |   |   |   edu = 4: No {Yes=1, No=1}
|   |   |   job = 5: Yes {Yes=5, No=1}
|   |   age = 3
|   |   |   edu = 3: No {Yes=1, No=2}
|   |   |   edu = 4: Yes {Yes=2, No=0}
|   |   |   edu = 5: No {Yes=1, No=2}
payment = 2
|   maj = 1
|   |   emplo = 2
|   |   |   mar = 0: No {Yes=1, No=1}
|   |   |   mar = 1: Yes {Yes=22, No=5}
|   |   emplo = 3
|   |   |   sex = 0: Yes {Yes=6, No=0}
|   |   |   sex = 1
|   |   |   |   age = 1
|   |   |   |   |   edu = 3: No {Yes=1, No=2}
|   |   |   |   |   edu = 4: Yes {Yes=2, No=0}
|   |   |   |   age = 2: No {Yes=1, No=1}
|   |   |   |   age = 3: No {Yes=0, No=2}
|   |   emplo = 4: Yes {Yes=5, No=1}
|   maj = 2
|   |   sex = 0: Yes {Yes=2, No=0}
|   |   sex = 1: No {Yes=1, No=1}
|   maj = 3: No {Yes=1, No=4}
|   maj = 4: Yes {Yes=13, No=1}
|   maj = 5: No {Yes=2, No=3}
payment = 3: Yes {Yes=65, No=11}
payment = 4
|   dep = 1: Yes {Yes=25, No=1}
|   dep = 3: Yes {Yes=7, No=3}
|   dep = 4: Yes {Yes=4, No=0}
|   dep = 5
|   |   sex = 0: No {Yes=1, No=2}
|   |   sex = 1: Yes {Yes=36, No=4}

```

نمودار (۴-۱۳) شمایک درخت تصمیم CART

بر طبق این تقسیم‌بندی ۷۵.۲ درصد افراد مجرد دارای رشته تحصیلی ۱ یا ۶ غیروفا دار بوده اند و تنها ۲۴.۸ درصد آن‌ها وفادار بوده اند. همچنین مطابق این تقسیم بندی ۱۶.۸ درصد افراد متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۸۳.۲ درصد وفادار بوده اند. در ادامه این تقسیم‌بندی افراد متاهل براساس جنسیت تقسیم‌بندی شده اند که ۷ درصد خانم‌های متاهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۹۳ درصدشان وفادار بوده اند و درمورد آقایان ۱۸ درصد غیروفا دار و ۸۲ وفادار بوده اند. بر اساس دقت آنالیز مدل در قسمت آموزش ۸۵.۶۶ درصد است. همچنین عمق درخت تصمیم ساخته شده ۴ است.

جدول (۱۷-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم CART

ارزیابی		تست		آموزش		دسته
۱۳	۷۶.۴۷٪	۱۲	۷۰.۵۸٪	۲۵۷	۸۵.۶۶٪	درست
۴	۲۳.۵۳٪	۵	۲۹.۴۲٪	۴۳	۱۴.۳۴٪	نادرست
۱۷	-	۱۷	-	۳۰۰	-	مجموع

در جدول (۱۷-۴) نشان داده شده است که دقت مدل در قسمت تست ۷۰.۵۸ درصد و در قسمت ارزیابی ۷۶.۴۷ درصد می‌باشد. تعداد داده‌هایی که در قسمت تست و ارزیابی مورد سنجش قرار گرفته اند داده‌های مربوط به ۳۴ نفر بوده است که برای ۲۵ نفر آن‌ها به درستی تشخیص داده شده که برای کار در سازمان وفادار هستند یا خیر.

جدول (۱۸-۴) ماتریس اغتشاش درخت تصمیم CART

نوع تشخیص	Positive	Negative	دسته
True	۱۹۵	۳۲	آموزش
False	۶۲	۱۱	
True	۱۰	۴	تست
False	۲	۱	
True	۱۲	۲	ارزیابی
False	۱	۲	

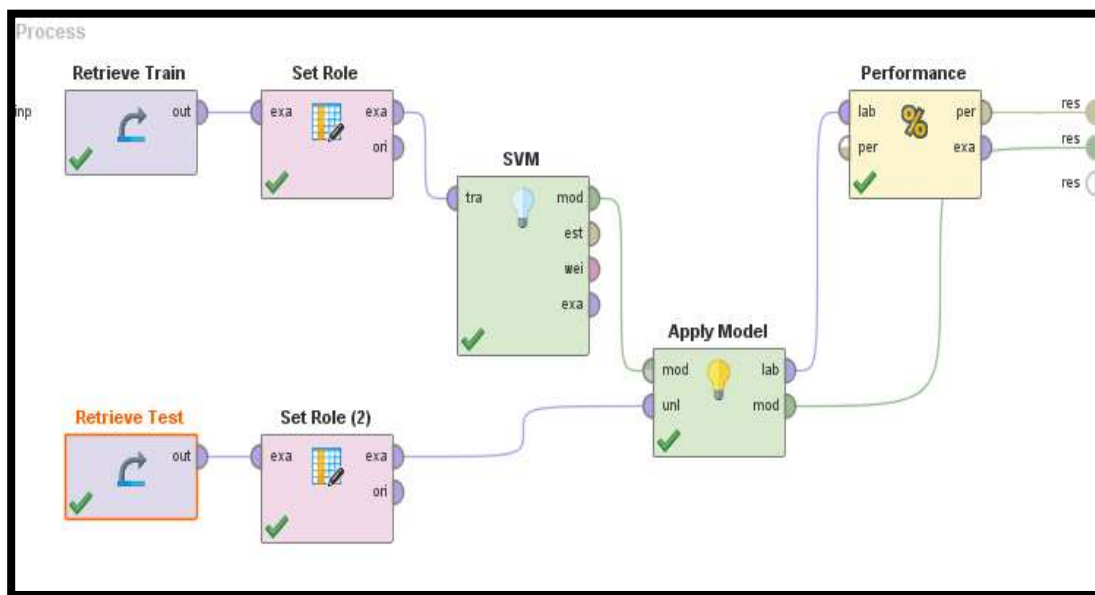
$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 75.87\%$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 94.66\%$$

۴-۳-۵) ماشین بردار پشتیبان

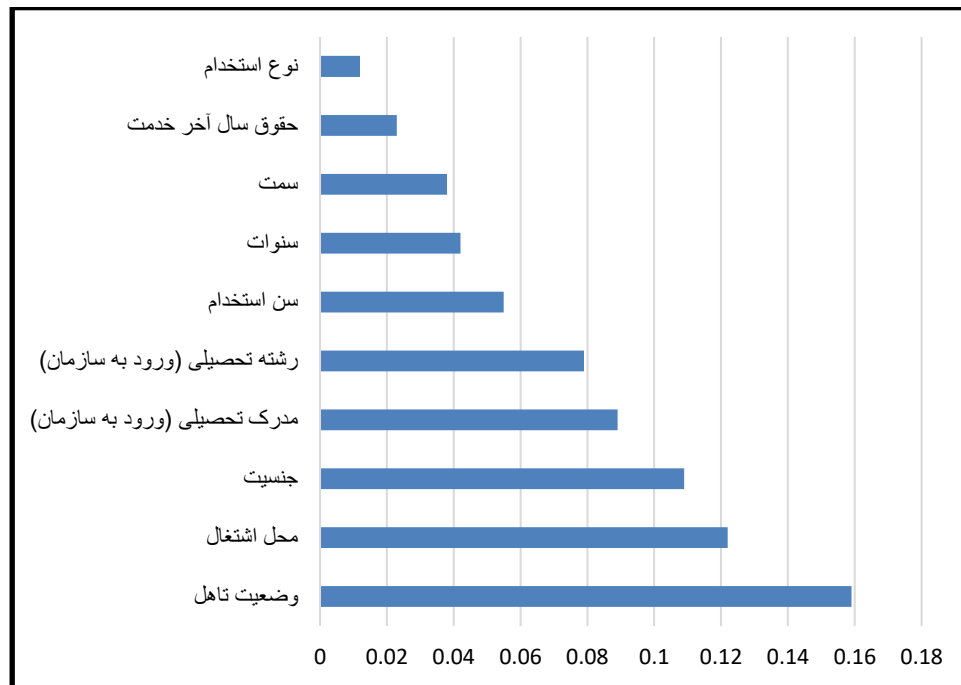
ماشین بردار پشتیبان یا SVM داده‌ها را با توجه به دسته‌های از پیش تعیین شده آن‌ها به یک فضای جدید می‌برد به گونه‌ای که داده‌ها به صورت خطی (یا ابر صفحه) قابل تفکیک و دسته‌بندی باشند و سپس با یافتن خطوط پشتیبان (صفحات پشتیبان در فضای چند بعدی)، سعی در یافتن معادله خطی دارد که بیشترین فاصله را بین دو دسته ایجاد می‌کند. بردارهای پشتیبان هر کدام یک فرمول مشخصه دارند که خط مرزی هر دسته را توصیف می‌کند. برای استفاده از SVM در مورد داده‌های واقعی، چندین نکته را باید رعایت کرد:

- ✓ داده‌ها را پالایش کنید (نقاط پرت، داده‌های ناموجود و ...)
- ✓ داده را عددی و نرمال کنید. داده‌هایی مانند جنسیت، رشته تحصیلی و ... را عددی و مقادیر همه صفات بین $[-1, 1]$ نرمال شوند تا بزرگ یا کوچک بودن مقادیر یک ویژگی، ماشین را تحت تاثیر قرار ندهد.
- ✓ کرنل‌های مختلف را امتحان و به ازای هر کدام، با توجه به مجموعه داده آموزشی که در اختیار دارید و دسته‌بندی داده‌های آن‌ها مشخص است، دقت SVM را اندازه‌گیری کنید و در صورت نیاز پارامترهای توابع تبدیل را تغییر دهید تا جواب‌های بهتری بگیرید.



شکل (۴-۵) مراحل ساخت ماشین بردار پشتیبان

شکل (۴-۵) نودهای استفاده شده به هنگام پیش پردازش داده‌ها و عملیات انجام شده جهت آماده سازی داده‌های ورودی به الگوریتم را نشان می‌دهد. الگوریتم به کار رفته در این قسمت SVM است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد.



نمودار (۴-۱۴) نمودار اهمیت متغیرها در ساخت ماشین بردار پشتیبان

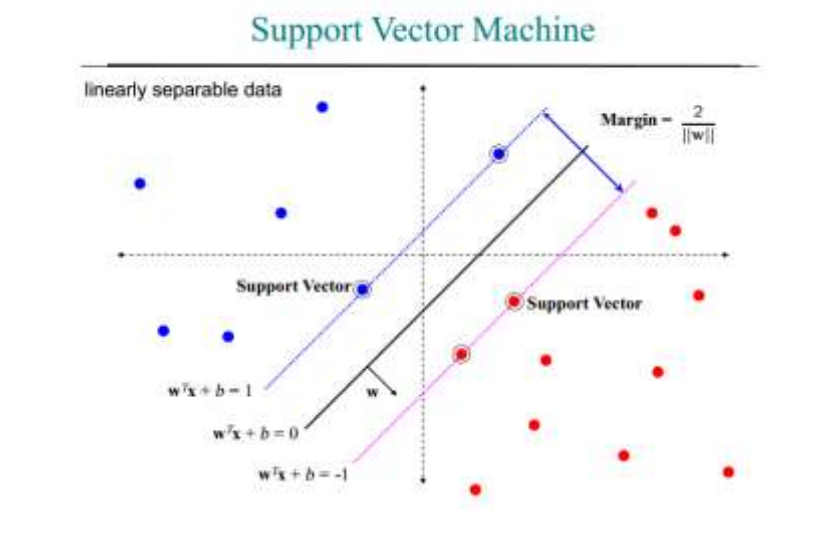
در نمودار (۴-۱۴) میزان تأثیر و اهمیت هر یک از متغیرهای ورودی در ساخت مدل نشان داده شده است. طبق این نمودار دو متغیر وضعیت تاهل و محل اشتغال بیشترین تأثیر را در ساخت مدل داشته اند. میزان تأثیر متغیرهای دیگر که شامل جنسیت و سطح تحصیلات می‌باشد تقریباً به یک اندازه بوده و اهمیت کمتری نسبت به متغیرهای اولیه در ساخت مدل داشته اند.

جدول (۴-۱۹) نشان دهنده ضریب اهمیت هر کدام از متغیرها می‌باشد در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت‌ترین متغیر به سمت کم اهمیت‌ترین متغیر مرتب سازی شده اند. مطابق این جدول ضریب تأثیر متغیرهای کم اهمیت شامل حقوق سال آخر خدمت و نوع استخدام است که نشان دهنده اهمیت کم این متغیرها در ساخت SVM می‌باشد.

جدول (۴-۱۹) ضریب اهمیت متغیرها

متغیر گره	ضریب اهمیت
وضعیت تاهل	۰.۱۵۹
محل اشتغال	۰.۱۲۲
جنسیت	۰.۱۰۹
مدرک تحصیلی (ورود به سازمان)	۰.۰۸۹
رشته تحصیلی (ورود به سازمان)	۰.۰۷۹
سن استخدام	۰.۰۵۵
سنوات	۰.۰۴۲
سمت	۰.۰۳۸
حقوق سال آخر خدمت	۰.۰۲۳
نوع استخدام	۰.۰۱۲

همچنین با توجه به جدول میزان اهمیت متغیرهای چهارم تا هفتم که شامل متغیرهای سطح تحصیلات، سن استخدام، سنوات و سمت می‌باشد در محدوده ۰.۰۹ تا ۰.۰۴ است. طبق جدول (۴-۱۹) بیشترین متغیرهای تأثیرگذار یعنی متغیرهای وضعیت تاهل با ضریب اهمیت ۰.۱۵۹ و محل اشتغال با ضریب ۰.۱۲۲ می‌باشند که نشان دهنده اهمیت برالای این دو متغیر نسبت به سایر متغیرهای ورودی در ساخت مدل است.



نمودار (۴-۱۵) شمایک ماشین بردار پشتیبان

شکل (۴-۱۵) نشان دهنده ساختار شماتیک ماشین بردار پشتیبان است. مطابق اطلاعات موجود در این ساختار درصد افرادی که وفادار بوده اند ۸۸.۲۳ درصد کارکنان را شامل می‌شود. طبق این مدل ۱۲.۵ درصد افراد با رشته‌های تحصیلی موجود در گروه ۱ و ۶ نسبت به سازمان غیروفا دار و ۸۷.۵ درصدشان وفادار بوده اند. همچنین رشته‌های تحصیلی موجود در گروه‌های ۲، ۳، ۴ و ۵ نیز ۳۸.۹ درصدشان غیروفا دار و ۶۱.۱ درصد وفادار بوده اند. بر طبق این تقسیم‌بندی ۴۲.۳ درصد افراد مجرد دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و تنها ۵۷.۷ درصد آن‌ها وفادار بوده اند. همچنین مطابق این تقسیم بندی ۱۱.۸ درصد افراد متأهل دارای رشته تحصیلی ۱ یا ۶ غیروفا دار و ۸۸.۲ درصد وفادار بوده اند.

جدول (۴-۲۰) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم درخت تصمیم

ارزیابی		تست		آموزش		دسته
۱۵	۸۸.۲۳٪	۱۴	۸۲.۳۵٪	۲۴۶	۸۲٪	درست
۲	۱۱.۷۷٪	۳	۱۷.۶۵٪	۵۴	۱۸٪	نادرست
۱۷	-	۱۷	-	۳۰۰	-	مجموع

در جدول (۴-۲۰) نشان داده شده که دقت مدل در قسمت تست ۸۲.۳۵ درصد و دقت در قسمت ارزیابی ۸۸.۲۳ درصد می‌باشد. تعداد داده‌هایی که در قسمت تست و ارزیابی مورد سنجش قرار گرفته اند داده‌های مربوط به ۳۴ نفر بوده است که برای ۲۹ نفر آن‌ها به درستی تشخیص داده شده که برای کار در سازمان وفادار هستند یا خیر.

جدول (۴-۲۱) ماتریس اغتشاش svm

دسته	Negative	Positive	نوع تشخیص
آموزش	۳۲	۱۹۹	True
	۲۲	۵۵	False
تست	۲	۱۰	True
	۱	۴	False
ارزیابی	۱	۱۱	True
	۳	۲	False

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 78.34\%$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 90.04\%$$

جدول (۴-۲۱) ماتریس اغتشاش یا ماتریس درهم ریختگی را نشان می‌دهد. بر اساس این ماتریس تعداد و نوع خطاهای صورت گرفته در دسته‌بندی به خوبی مشخص است و همچنین نشان می‌دهد چه تعداد از داده‌ها در دسته مناسب خود قرار گرفته اند. در ادامه با استفاده از ماتریس اغتشاش و فرمول‌های بیان شده دو معیار Precision و Recall در داده‌های تست که مقدار آن‌ها به ترتیب برابر با ۷۸.۳۴ درصد و ۹۰.۰۴ درصد شد، که تایید کننده اطمینان مدل و مفید بودن آن می‌باشد.

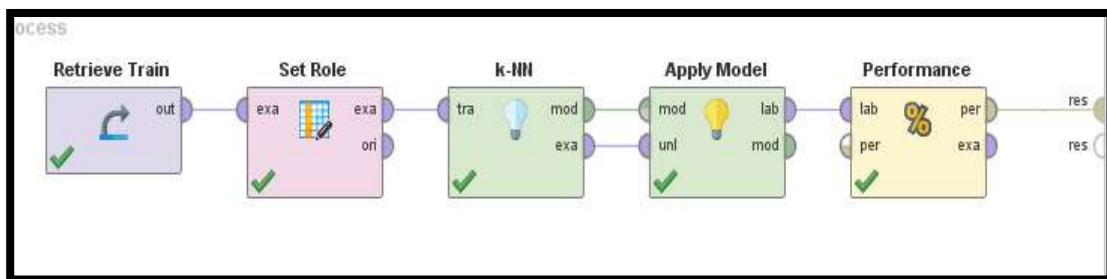
۴-۳-۶ الگوریتم KNN

این الگوریتم یک نمونه تستی را بر اساس k همسایه نزدیک دسته‌بندی می‌کند. نمونه‌های آموزشی به عنوان بردارهایی در فضای ویژگی چند بعدی مطرح می‌شوند. فضا به ناحیه‌هایی با نمونه‌های آموزشی پارتیشن‌بندی می‌شود. یک نقطه در فضا به کلاسی تعلق می‌یابد که بیشترین نقاط آموزشی متعلق به آن کلاس در داخل نزدیک‌ترین نمونه‌ی آموزشی به k در آن باشد. معمولاً فاصله‌ی اقلیدسی یا تشابه کوسینوسی در این روش استفاده می‌شود. در فاز دسته‌بندی K Nearest Neighbors، نمونه تستی به عنوان یک بردار در فضای ویژگی نمایش داده می‌شود و فاصله‌ی اقلیدسی یا تشابه کوسینوسی بردار تستی با کل بردارهای آموزشی محاسبه می‌شود و نزدیک‌ترین نمونه‌ی آموزشی به k انتخاب می‌شود. البته راه‌های زیادی برای دسته‌بندی بردار تستی وجود دارد و بنابراین الگوریتم k همسایه نزدیک کلاسیک، یک نمونه تستی را بر اساس بیشترین آراء از k همسایه‌ی نزدیکش تعیین می‌کند. سه فاکتور مهم در این الگوریتم روش k همسایه‌ی نزدیک، به شرح زیر می‌باشد:

✓ معیار فاصله یا شباهت، برای پیدا کردن k همسایه نزدیک استفاده می‌شود.

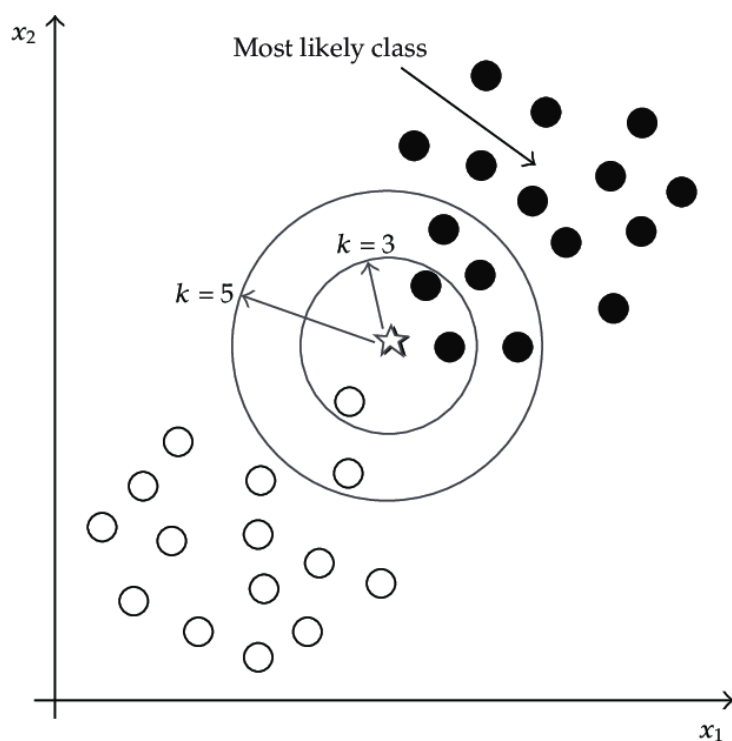
✓ K تعداد همسایه‌های نزدیک است.

✓ قانون تصمیم‌گیری برای تعیین (شناسایی) یک کلاس برای سند تستی از k همسایه نزدیک می‌باشد.



شکل (۴-۶) مراحل ساخت KNN

شکل (۴-۶) نودهای استفاده شده به هنگام پیش پردازش داده‌ها و عملیات انجام شده جهت آماده‌سازی داده‌های ورودی به الگوریتم را نشان می‌دهد. الگوریتم به کار رفته در این قسمت KNN است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد.



نمودار (۴-۱۶) شمایک KNN

نمودار (۴-۱۶) نشان دهنده ساختار کلی الگوریتم KNN می‌باشد. مطابق اطلاعات موجود در این الگوریتم درصد افرادی که به سازمان وفادار بوده اند ۹۴.۱۱ درصد کارکنان را شامل می‌شود.

جدول (۲۲-۴) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم KNN

ارزیابی		تست		آموزش		دسته
۱۶	۹۴.۱۱٪	۱۶	۹۴.۱۱٪	۲۹۷	۹۹٪	درست
۱	۵.۸۹٪	۱	۵.۸۹٪	۳	۱٪	نادرست
۱۷	-	۱۷	-	۳۰۰	-	مجموع

در جدول (۲۲-۴) نشان داده شده است که دقت مدل در قسمت تست ۹۴.۱۱ درصد و دقت در قسمت آموزش ۹۹ درصد می‌باشد.

جدول (۲۳-۴) ماتریس اغتشاش الگوریتم KNN

نوع تشخیص	Positive	Negative	دسته
True	۲۴۵	۱	آموزش
False	۵۲	۲	
True	۱۵	۰	تست
False	۱	۱	
True	۱۴	۱	ارزیابی
False	۲	۰	

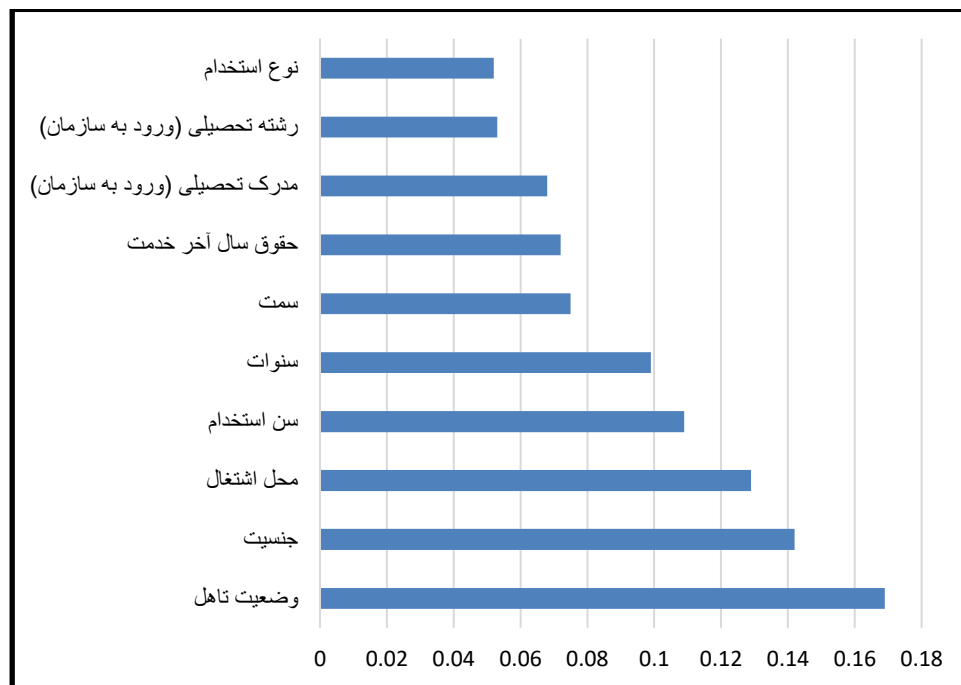
$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 82.49\%$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 99.19\%$$

جدول (۲۳-۴) ماتریس اغتشاش یا ماتریس درهم ریختگی را نشان می‌دهد. بر اساس این ماتریس تعداد و نوع خطاهای صورت گرفته در دسته‌بندی به خوبی مشخص است و همچنین نشان می‌دهد چه تعداد از داده‌ها در دسته مناسب خود قرار گرفته اند. در ادامه با استفاده از ماتریس اغتشاش و فرمول‌های بیان شده دو معیار Precision و Recall در داده‌های تست که مقدار آن‌ها به ترتیب برابر با ۶۵.۶۶ و ۹۸.۹۸ درصد شد، که تایید کننده اطمینان مدل و مفید بودن آن می‌باشد.

۴-۳-۷) الگوریتم ترکیبی

الگوریتم به کار رفته در این قسمت یک الگوریتم ترکیبی از KNN و الگوریتم SVM است و نتایج حاصل از آن در ادامه پژوهش مورد بحث و بررسی قرار می‌گیرد. در نمودار (۴-۱۷) میزان تأثیر و اهمیت هر یک از متغیرهای ورودی در ساخت مدل نشان داده شده است. طبق این نمودار دو متغیر وضعیت تاهل و جنسیت بیشترین تأثیر را در ساخت مدل داشته‌اند. میزان تأثیر متغیرهای دیگر که شامل محل اشتغال، سن استخدام و سنوات می‌باشد تقریباً به یک اندازه بوده و اهمیت کمتری نسبت به متغیرهای اولیه در ساخت مدل دارند.



نمودار (۴-۱۷) نمودار اهمیت متغیرها در ساخت الگوریتم ترکیبی

جدول (۴-۲۴) نشان دهنده ضریب اهمیت هر کدام از متغیرها می‌باشد در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت‌ترین متغیر به سمت کم اهمیت‌ترین متغیر مرتب سازی شده‌اند. مطابق این جدول ضریب تأثیر متغیرهای کم اهمیت شامل نوع استخدام و رشته تحصیلی است که نشان دهنده اهمیت کم این متغیرها در ساخت مدل ترکیبی می‌باشد.

جدول (۴-۲۴) ضریب اهمیت متغیرها

متغیر گره	ضریب اهمیت
وضعیت تاهل	۰.۱۶۹
جنسیت	۰.۱۴۲
محل اشتغال	۰.۱۲۹
سن استخدام	۰.۱۰۹
نوع استخدام	۰.۰۹۹
سمت	۰.۰۷۵
حقوق سال آخر خدمت	۰.۰۷۲
مدرک تحصیلی (ورود به سازمان)	۰.۰۶۸
رشته تحصیلی (ورود به سازمان)	۰.۰۵۳
سنوات	۰.۰۵۲

همچنین با توجه به جدول میزان اهمیت متغیرهای سوم تا هفتم که شامل متغیرهای محل اشتغال، سن استخدام، نوع استخدام و سمت می‌باشد در محدوده ۰.۱۲۹ تا ۰.۰۷ است. طبق جدول (۴-۲۴) بیشترین متغیرهای تأثیرگذار یعنی متغیرهای وضعیت تاهل با ضریب اهمیت ۰.۱۶ و جنسیت با ضریب ۰.۱۴ می‌باشند که نشان دهنده اهمیت بالای این دو متغیر نسبت به سایر متغیرهای ورودی در ساخت مدل است. مطابق اطلاعات موجود میزان افرادی که به سازمان وفادار بوده اند ۹۴.۱۱ درصد کارکنان را شامل می‌شود. بر طبق این تقسیم‌بندی ۹۲.۳ درصد افراد مجرد وفادار بوده اند و تنها ۷.۷ درصد آن‌ها غیروفا دار بوده اند. همچنین مطابق این تقسیم بندی ۱۱.۸ درصد افراد متاهل دارای غیروفا دار و ۸۸.۲ درصد وفادار بوده اند. در ادامه این تقسیم‌بندی افراد متاهل براساس جنسیت تقسیم‌بندی شده اند که ۱۴ درصد خانم‌های متاهل غیروفا دار و ۸۶ وفادار بوده اند و درمورد آقایان ۲۳ درصد غیروفا دار و ۷۸ درصد وفادار بوده اند. بر اساس آن دقت آنالیز مدل در قسمت آموزش ۹۶ درصد است. در جدول (۴-۲۵) نشان داده شده است که دقت مدل در هر دو قسمت تست و ارزیابی ۹۴.۱۱ درصد می‌باشد.

جدول (۴-۲۵) مقایسه دقت در مدل‌های آموزش، تست و ارزیابی در الگوریتم ترکیبی

ارزیابی		تست		آموزش		دسته
۱۶	۹۴.۱۱٪	۱۶	۹۴.۱۱٪	۲۸۸	۹۶٪	درست
۱	۵.۸۹٪	۱	۵.۸۹٪	۱۲	۴٪	نادرست
۱۷	-	۱۷	-	۳۰۰	-	مجموع

جدول (۴-۲۶) ماتریس اغتشاش الگوریتم ترکیبی

نوع تشخیص	Positive	Negative	دسته
True	۲۶۲	۱۰	آموزش
False	۲۶	۲	
True	۱۵	۰	تست
False	۱	۱	
True	۱۶	۰	ارزیابی
False	۰	۱	

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) = 90.97\%$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) = 91.66\%$$

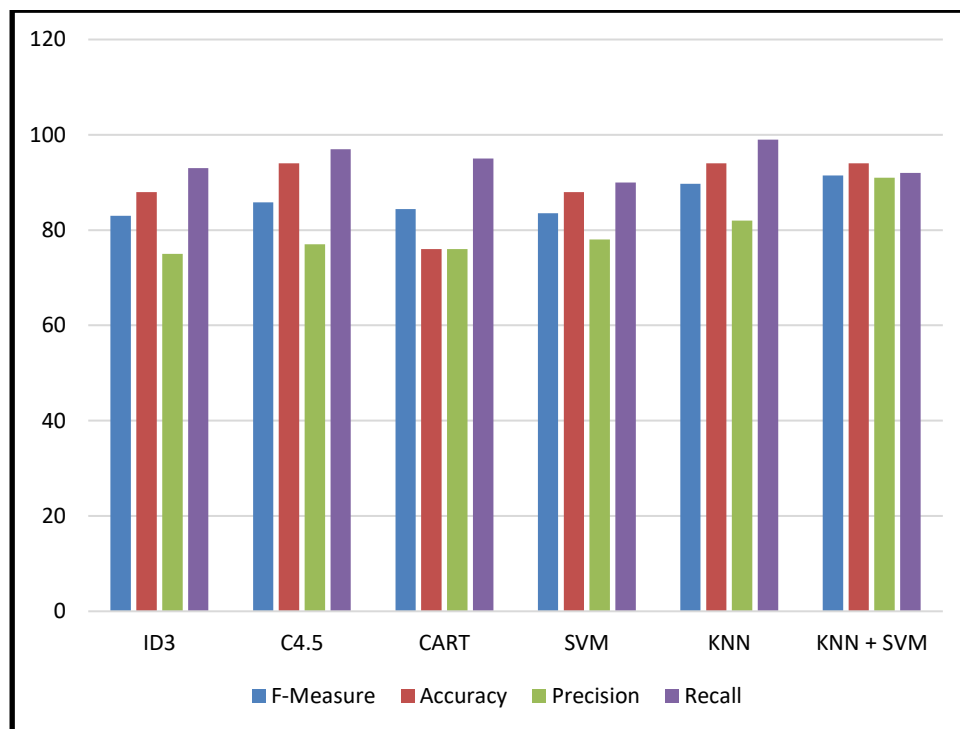
جدول (۴-۲۶) ماتریس اغتشاش یا ماتریس درهم ریختگی را نشان می‌دهد. بر اساس این ماتریس تعداد و نوع خطاهای صورت گرفته در دسته‌بندی به خوبی مشخص است و همچنین نشان می‌دهد چه تعداد از داده‌ها در دسته مناسب خود قرار گرفته‌اند. با استفاده از ماتریس اغتشاش معیارهای Precision و Recall جهت تحلیل مدل‌ها به دست می‌آیند. در ادامه با استفاده از ماتریس اغتشاش و فرمول‌های بیان شده دو معیار Precision و Recall در داده‌های تست که مقدار آن‌ها به ترتیب برابر با ۹۰.۹۷ و ۹۱.۶۶ درصد شد، که تایید کننده اطمینان مدل و مفید بودن آن می‌باشد.

۴-۴) مقایسه نتایج مدل‌ها

نتیجه‌گیری نهایی این پژوهش بر اساس مقایسه و تحلیل پارامترهای موجود در این جدول انجام می‌شود. مطابق با نتایج نشان داده شده در جدول (۴-۲۷) تمامی مدل‌های ساخته شده از دقت‌های قابل قبولی در مدل تست و آموزش برخوردار هستند.

جدول (۴-۲۷) مقایسه نتایج الگوریتم‌ها

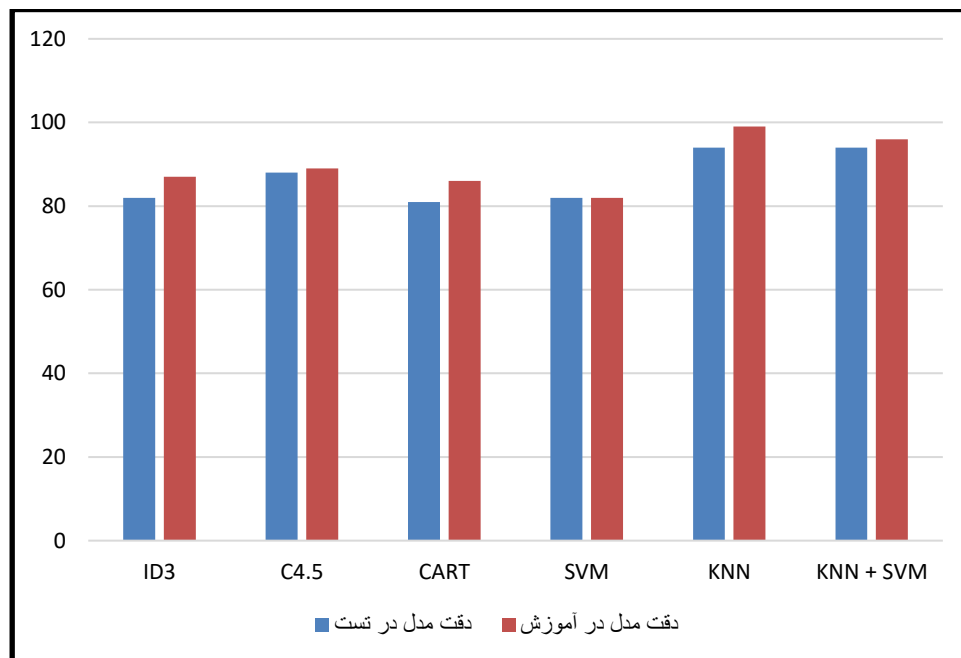
الگوریتم	دقت مدل در آموزش	دقت مدل در تست	Recall	Precision	Accuracy	F-Measure
ID3	۸۷	۸۲	۹۳	۷۵	۸۸	۸۳.۰۳
C4.5	۸۹	۸۸	۹۷	۷۷	۹۴	۸۵.۸۵
CART	۸۶	۸۱	۹۵	۷۶	۷۶	۸۴.۴۴
SVM	۸۲	۸۲	۹۰	۷۸	۸۸	۸۳.۵۷
KNN	۹۹	۹۴	۹۹	۸۲	۹۴	۸۹.۷۰
SVM + KNN	۹۶	۹۴	۹۲	۹۱	۹۴	۹۱.۴۹



نمودار (۴-۱۸) مقایسه نتایج الگوریتم‌ها معیارهای مورد بررسی

همان گونه که پیش تر در مورد معیارهای Recall و Precision توضیح داده شد، هر چه این معیارها به ۱ نزدیک تر باشند مدل قابل قبول تر است، در نتیجه مدل های ساخته شده تا حد بسیار زیادی قابل قبول می باشند. هدف اصلی این پژوهش ساخت الگوریتم موثر برای تعیین کارمندان وفادار بود. پس از تشریح کوتاه نتایج این مدل، مقایسه ای بین تمامی الگوریتم های ساخته شده انجام می شود و بهترین الگوریتم معرفی می شود. با توجه به جدول (۴-۲۷) در الگوریتم های SVM و KNN ساخته شده دقت در مدل آموزش بالای ۹۵ درصد است که دقت بسیار خوبی به نظر می رسد، همچنین دقت مدل در تست ۹۴ درصد می باشد که بیانگر کاربردی بودن مدل است.

معیارهای Recall هر دو تقریباً در یک حدود و بالای ۹۰ درصد می باشد ولی معیار Precision در الگوریتم ترکیبی به مراتب بهتر است. از آنجایی که قسمت تست در مدل سازی حایز اهمیت بیشتر می باشد این معیارها نشان دهنده تعداد خطای کم در قسمت تست است. همان طور که در جدول (۴-۱۸) نمایش داده شده است، دقت در مدل تست در تمام الگوریتم ها دقتی مناسب و موثر می باشد. با توجه به این موضوع که از دست دادن یک نیروی وفادار موثر، ضرر و زیان بسیاری به سازمان می زنند، در نتیجه معیارهای Precision و Recall اهمیت بیشتری نسبت به معیار دقت دارند و باید حتما در نظر گرفته شوند.



نمودار (۴-۱۹) مقایسه نتایج الگوریتم ها در دقت آموزش و تست

مطابق جدول معیارهای Precision و Recall نیز در الگوریتم SVM و KNN بیشتر از دیگر الگوریتم‌ها است و حداقل خطای ممکن را دارد. با توجه به جدول (۴-۱۹)، الگوریتم ترکیبی صحت بالاتری نسبت به الگوریتم‌های SVM و KNN، و معیار Precision حدوداً نزدیک به ۱۰۰ درصد دارد، که مقداری ایده آل است و نشان دهنده این که مطابق این مدل هرگز افراد وفادار و موثر به اشتباه از سازمان خارج نمی‌شوند و در نتیجه هزینه‌های اضافی سازمان که صرف آموزش نیروهای جایگزین می‌شود حذف خواهد شد.

۴-۵) نتیجه‌گیری

معیارهای Precision و Recall در مدل‌سازی حایز اهمیت بیشتر می‌باشد این معیارها نشان دهنده تعداد خطای کم در قسمت تست است. دقت در مدل تست در تمام الگوریتم‌ها دقتی مناسب و موثر می‌باشد. با توجه به این موضوع که از دست دادن یک نیروی وفادار موثر، ضرر و زیان بسیاری به سازمان می‌زند، در نتیجه معیارهای Precision و Recall اهمیت بیشتری نسبت به معیار دقت دارند و باید حتماً در نظر گرفته شوند. مطابق جدول معیارهای Precision و Recall نیز در الگوریتم SVM و KNN بیشتر از دیگر الگوریتم‌ها است و حداقل خطای ممکن را دارد.

فصل پنجم:

نتیجہ گیری و پیشہ نادات

پژوهش حاضر به دنبال ارایه الگوریتمی برتر جهت انتخاب نیروی انسانی وفادار به یک سازمان است. در واقع چنین مدلی باعث صرفه‌جویی در وقت و هزینه‌های بخش نیروی انسانی در یک سازمان می‌شود. در فصل چهارم مدل‌های معرفی شده در قسمت‌های قبل به طور کامل مورد تجزیه و تحلیل قرار گرفتند. هدف از این فصل نتیجه‌گیری کلی از نتایج مدل‌ها و تجزیه و تحلیل صورت گرفته بر روی آن‌ها می‌باشد. در ادامه فصل مدل برتر معرفی می‌شود و در پایان پیشنهادات سازنده‌ای برای آینده این گونه پژوهش‌ها و چگونگی به کارگیری این روش‌ها برای دیگر سازمان‌ها مطرح می‌شود.

۵-۱-۱) مقایسه نتایج با نتایج پژوهشهای پیشین

همانطور که پیش‌تر اشاره شد، از الگوریتم‌های نام برده شده در حوزه منابع انسانی استفاده زیادی نشده است و پژوهش‌هایی از این دست نیز نتایج مربوط به دقت و Precision و Recall آورده نشده بود و فقط قواعد حاصل از الگوریتم‌ها بیان شده بود که برای مقایسه نتایج مناسب نیستند. به همین علت در این بخش علاوه بر پژوهش‌های مدیریت منابع انسانی، پژوهش‌های دیگر علوم با استفاده از الگوریتم‌های داده‌کاوی نیز بررسی شد تا مشخص شود نتایج این پژوهش تا چه اندازه قابل استناد و اتکا می‌باشد.

جدول (۵-۱) مقایسه نتایج پژوهش‌های پیشین

سال	محقق	موضوع	معیار اندازه‌گیری	الگوریتم	میزان
۲۰۱۵	سیکارودی و قوسی	چرخش شغلی	دقت	جنگل تصادفی	۹۰.۶
				C&R	۸۰.۴
۲۰۱۵	هانگ و لوو	طبقه‌بندی تصاویر	دقت	جنگل تصادفی	۹۰.۳
				C&R-Bagging	۸۸.۵
				C&R	۸۳.۷
۲۰۱۱	جانتان و همدان	مدیریت استعداد	دقت	جنگل تصادفی	۹۵.۱
				C4.5	۷۴.۹

۵-۳) نتیجه گیری

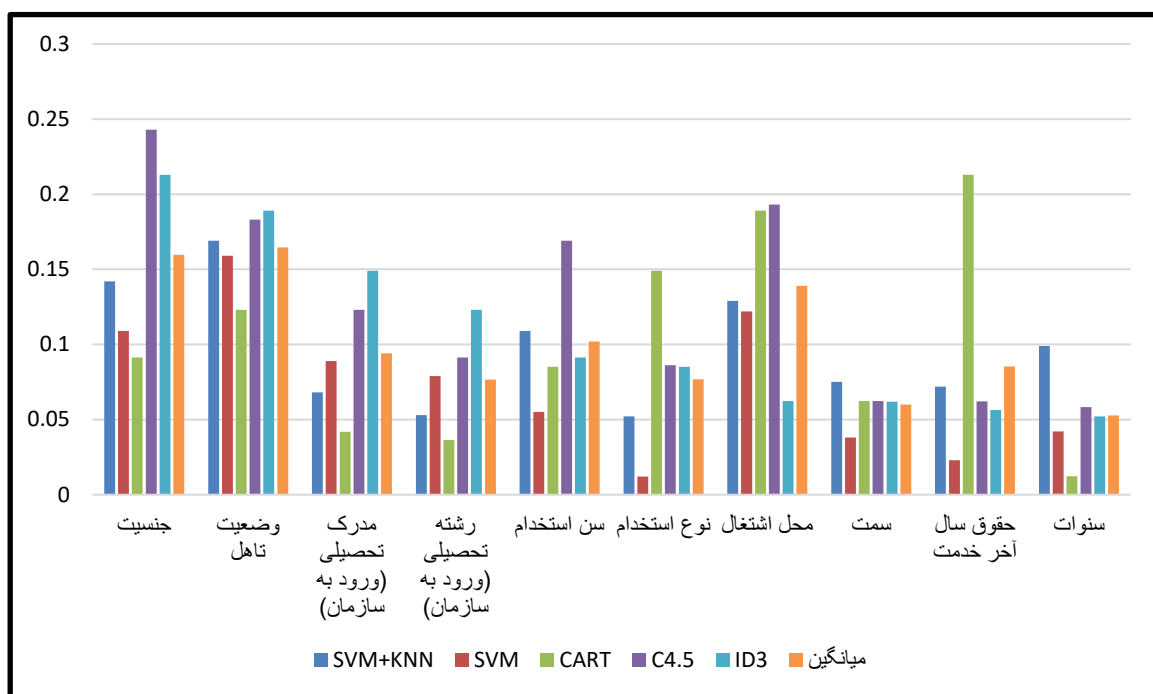
در دنیای امروز اغلب مسایلی که برای تصمیم گیری به مدیران عرضه می شود ابعاد متنوعی دارد و با چند معیار فرموله می شود. به بیان دیگر بیشتر تصمیم گیری های مدیران، از عوامل گوناگون کمی و کیفی تأثیر می پذیرد که اغلب این عوامل با یکدیگر در تعارض هستند و آنان سعی می کنند بین گزینه های موجود، بهترین گزینه را انتخاب کنند. اشتباه و بی دقتی در تصمیم گیری مستلزم پرداخت هزینه است. هرچه قدرت و اختیارات مدیریت بیشتر باشد هزینه تصمیم نادرست نیز بالاتر خواهد رفت. به همین دلیل ترکیبی از الگوریتم های SVM و KNN ساختیم تا در صورت امکان خطا و اشتباه مدیریتی به حداقل ممکن برسد.

جدول (۵-۲) مقایسه نتایج ضرایب الگوریتم ها

متغیر گره	ID3	C4.5	CART	SVM	SVM+KNN	میانگین
وضعیت تاهل	۰.۱۸۹	۰.۱۸۳	۰.۱۲۳	۰.۱۵۹	۰.۱۶۹	۰.۱۶۴
جنسیت	۰.۲۱۳	۰.۲۴۳	۰.۰۹۱۲	۰.۱۰۹	۰.۱۴۲	۰.۱۵۹
محل اشتغال	۰.۰۶۲۳	۰.۱۹۳	۰.۱۸۹	۰.۱۲۲	۰.۱۲۹	۰.۱۳۹
سن استخدام	۰.۰۹۱۲	۰.۱۶۹	۰.۰۸۵۲	۰.۰۵۵	۰.۱۰۹	۰.۱۰۱
مدرک تحصیلی (ورود به سازمان)	۰.۱۴۹	۰.۱۲۳	۰.۰۴۱۸	۰.۰۸۹	۰.۰۶۸	۰.۰۹۴
حقوق سال آخر خدمت	۰.۰۵۶۳	۰.۰۶۲۲	۰.۲۱۳	۰.۰۲۳	۰.۰۷۲	۰.۰۸۵
رشته تحصیلی (ورود به سازمان)	۰.۱۲۳	۰.۰۹۱۲	۰.۰۳۶۳	۰.۰۷۹	۰.۰۵۳	۰.۰۷۶
نوع استخدام	۰.۰۸۵۲	۰.۰۸۶۲	۰.۱۴۹	۰.۰۱۲	۰.۰۵۲	۰.۰۷۶
سمت	۰.۰۶۱۸	۰.۰۶۲۳	۰.۰۶۲۳	۰.۰۳۸	۰.۰۷۵	۰.۰۵۹
سنوات	۰.۰۵۲۲	۰.۰۵۸۲	۰.۰۱۲۲	۰.۰۴۲	۰.۰۹۹	۰.۰۵۲

با توجه به توضیحات فوق مدل ترکیبی بهترین مدل ساخته شده و دارای دقت بالاتر نسبت به الگوریتم‌های تکی بود. با استناد به این موضوع که امر حفظ نیروی انسانی یک امر پیچیده و تاثیرگذار در کارایی سازمان است در نتیجه سازمان باید با استفاده از روش‌هایی که حداقل خطای ممکن را در پی دارد و میزان اشتباهات را به حداقل می‌رساند در نگهداری از نیروی انسانی موثر و وفادار کوشش کند.

جدول (۵-۲) و نمودار (۵-۱) نشان دهنده ضریب اهمیت هر کدام از متغیرها برای هر یک از الگوریتم‌ها همچنین میانگین ضریب اهمیت هر متغیر برای تمام الگوریتم‌ها می‌باشد.



نمودار (۵-۱) مقایسه نتایج ضرایب الگوریتم‌ها

در این جدول میزان اهمیت متغیرها به ترتیب از با اهمیت‌ترین متغیر به سمت کم اهمیت‌ترین متغیر بر اساس میانگین ضریب اهمیت مرتب سازی شده اند. مطابق این جدول بیشترین ضریب تأثیر مربوط به متغیرهای، وضعیت تاهل، جنسیت، محل اشتغال و سن استخدام بوده و کم اهمیت‌ترین متغیرها رشته تحصیلی، نوع استخدام، سمت و سنوات است که نشان دهنده اهمیت کم این متغیرها در ساختار مورد بررسی می‌باشد.

۴-۵) محدودیت‌های پژوهش

✓ در پژوهش‌های مرتبط با مدیریت منابع انسانی و پژوهش‌های استخدای به دلیل اینکه معیارهای جذب و استخدام یا حفظ نیروهای کنونی در سازمان‌های مختلف، متفاوت است در نتیجه نتایج حاصل از این پژوهش را نمی‌توان برای سازمان‌های دیگر به کار برد و هر سازمان به صورت مجزا الگوریتم منحصر به فردی که فقط مخصوص همان سازمان است نیاز دارد.

✓ در این پژوهش از متغیرهایی که سازمان در اختیار پژوهشگر قرار داده است برای مدل‌سازی استفاده شده است. ممکن است متغیرهایی در حفظ افراد در این سازمان تأثیر بسیار زیادی بر مدل‌سازی داشته باشند که سازمان اطلاعات مربوط به آن‌ها را نداشته باشد یا به هر دلیلی در اختیار پژوهشگر قرار ندهد، در نتیجه امکان از دست رفتن اطلاعات ارزشمندی در چنین پژوهش‌هایی وجود دارد.

✓ علم داده‌کاوی شامل الگوریتم‌های متفاوتی می‌باشد که می‌توان از آنها برای ارایه مدل استفاده کرد. در این پژوهش فقط تعداد محدودی از این الگوریتم‌ها به کار گرفته شد و نتایج حاصل از هر کدام مقایسه شدند. در صورت نیاز برای یافتن بهترین مدل می‌توان تمام الگوریتم‌های موجود را مورد بررسی قرار داد و بهترین مدل را انتخاب کرد.

۱-۴-۵) محدودیت‌های پژوهشگر

✓ عدم همکاری سازمان‌های مختلف جهت در اختیار گذاشتن اطلاعات نیروی انسانی به دلیل مسایل امنیتی سازمان.

✓ زمان بر بودن یافتن سازمانی که حاضر به همکاری شود و در عین حال اطلاعات مورد نیاز را داشته باشد.

✓ اکثر فرم‌ها و پرونده‌ها به صورت غیر سیستمی و دستی ذخیره شده اند و تعداد کمی از آن‌ها به صورت الکترونیکی نگهداری می‌شوند. از این رو جستجوی دستی در این اسناد و سیستمی کردن آن‌ها بسیار زمان‌بر و دشوار است.

۵-۵) پیشنهادات

۵-۵-۱) پیشنهادات مدیریتی

به منظور کاربردی شدن دانش حاصل از داده کاوی در فرآیند حفظ نیروهای موثر، باید قواعد حاصله از داده کاوی به تصمیم گیری در امور مدیریت نیروی انسانی کمک کند. بدین منظور لازم است دانش و نتایج به دست آمده به صورت دستورالعمل های مدون و گویا تهیه شوند تا با قواعد شرطی به دست آمده، تصمیم گیری توسط مدیران به راحتی صورت گیرد. این قواعد معمولاً به صورت گزاره های شرطی بیان می شوند که در فصل ۴ نمونه هایی از آنها شرح داده شد. نوع دیگری از تصمیم گیری در این خصوص، تعیین استراتژی های حفظ و نگهداری نیروی انسانی جهت تسریع در تصمیم گیری ها است. البته باید به این نکته توجه داشت که داده کاوی یک ابزار است و به جهت اینکه مانند انسان قوه تشخیص ندارد نمی تواند همه کارها را انجام دهد و جایگزین افراد نخبه و تحلیلگر در سازمان شود بلکه فقط به تحلیلگران برای پیدا کردن الگوها و روابط بین داده ها کمک می کند و در این مورد نیز روابطی یافت می شود که باید بوسیله داده های واقعی دوباره بررسی و تست شوند. بر اساس نتایج حاصله استخدام افراد متاهل و استفاده بیشتر از نیروهای خانم بازدهی سازمانها را افزایش خواهد داد.

۵-۵-۲) پیشنهادات تحقیقات آتی

همچنین در به کارگیری این روش ها باید توجه داشت همان گونه که روش های داده کاوی در استخراج و تحلیل داده ها بسیار موثر هستند به همان نسبت نیز به کارگیری آنالیزها با داده های نامناسب می تواند ما را به سمت نتایج اشتباه هدایت نماید و برعکس، استفاده از روش های نامناسب با مجموعه ای از داده ها می تواند یک نگرش کاملاً متفاوتی را ایجاد نماید. از سوی دیگر مدل ها می توانند در راستای فرضیات فریبده استفاده گردند که گسترش همگی این موارد به زیان های گران قیمتی منجر می شود. همچنین برای تحقیقات آتی در این حوزه توصیه می شود که چالش های موجود بر سر راه سازمان ها و نوع عملکرد مدیران و تاثیر مستقیم آن برای افزایش میزان وفاداری کارکنان با مولفه های مدیریت دانش بررسی گردد.

منابع

- اسماعیلی، صید زاده، (۲۰۱۷). تأثیر رضایت شغلی بر عملکرد با نقش میانجی وفاداری سازمانی. مطالعات مدیریت بهبود و تحول، ۲۵(۸۳)، ۵۱-۶۸.
- افشار کاظمی، بیگدلی، منوچهری. (۲۰۱۳). پیش‌بینی الگوی ورود بیمار به بخش اورژانس بیمارستان با استفاده از تکنیک داده‌کاوی و مدل شبکه عصبی. مجله بیمارستان، ۱۲(۴)، ۷۳-۸۱.
- امیران، حیدر، (۱۳۸۴). ارزیابی و سنجش میزان وفاداری کارکنان به سازمان، ششمین کنفرانس بین‌المللی مدیریت کیفیت.
- برومند، زهرا (۱۳۸۶). مدیریت رفتار سازمانی. چاپ پانزدهم. تهران. انتشارات پیام نور.
- بوژمهرانی، سارا. کالته سیفری، معصومه، اکبری یزدی، حسین. (۱۳۹۶). پیش‌بینی وفاداری کارکنان از طریق تصویر ذهنی برند (مطالعه موردی: آکادمی ملی المپیک و پارالمپیک) فصلنامه علمی - پژوهشی مطالعات مدیریت رفتار سازمانی در ورزش ۸۶-۷۹، ۴ (۱).
- حافظ نیام، (۱۳۸۴)، "مقدمه‌ای بر روش تحقیق در علوم انسانی". چاپ دهم، انتشارات سمت، ص ۳۴۴ تهران.
- حسینی، م.، حلوائی‌ها، و.، رضائی، س.، (۱۳۸۹).، بررسی تأثیر تصویر ذهنی برند بر وفاداری کارکنان، مجله مدیریت توسعه و تحول، شماره ۵، ۷۵-۶۹.
- رفعتی، جواد، آقاموسی، رضا، تعامی، عبدالله، (۱۳۹۱). بررسی تأثیر وفاداری بر خرید مجدد مشتریان فراورده‌های لبنی. قاضی‌زاده، مصطفی، امانی، جمشید و کریمی خسرو، حسین. (۱۳۸۷). بررسی عوامل مؤثر بر ماندگاری سازمانی نیروی انسانی. دوماهنامه دانشور رفتار. سال پانزدهم. شماره ۲۹، صص ۴۰-۲۹.
- ستایش محمدحسین، ابراهیمی فهیمه، سیف سید مجتبی، ساریخانی مهدی. پیش‌بینی نوع اظهار نظر حساب‌برسان با رویکردی بر روش‌های داده‌کاوی.
- سرمد ز، حجازی آ و بازرگان ع، (۱۳۸۹) "روش تحقیق در علوم رفتاری"، چاپ سوم، انتشارات آگاه، تهران ص ۹۸.

- طاهری روزبهانی، م، نعمتی، م، ۱۳۹۱، بررسی تأثیر رضایت‌مندی شغلی بر وفاداری کارکنان ماهانه کار و جامعه، شماره ۳۸، ۱۵۴-۳۳.
- غفوری ورنوسفادرانی، گل‌پرور. بررسی رابطه مؤلفه‌های عدالت سازمانی با تعهد سازمانی در کارکنان شهرداری شهر اصفهان. مطالعات روان‌شناختی، ۲۰۰۹، ۵: ۱۲۹-۱۴۸.
- قاضی‌زاده، مصطفی، امانی، جمشید و کریمی خسرو، حسین. (۱۳۸۷). بررسی عوامل مؤثر بر ماندگاری سازمانی نیروی انسانی. دوماهنامه دانشور رفتار. سال پانزدهم. شماره ۲۹. صص ۲۹-۴۰.
- قره‌چه، منیژه، دایوئیان، منیره. (۱۳۹۰). وفاداری کارکنان در تعامل با وفاداری مشتریان صنایع خدماتی تحقیقات بازاریابی نوین، ۱ (۳): ۲۷-۴۶.
- مدنی، حسین و زاهدی، محمدجواد ۱۳۸۴ تعیین اولویت عوامل مؤثر بر تعهد سازمانی کارکنان، مجله جامعه‌شناسی ایران، دوره ششم، شماره ۱.
- هربرت ای. سایمون، ۱۳۷۱، رفتار اداری - بررسی در فراگردهای تصمیم‌گیری سازمان اداری، دکتر محمدعلی طوسی، مرکز آموزش مدیریت دولتی، چاپ اول ص ۱۴۱.
- یعقوبی مریم، سقاییان نژاد اصفهانی سکینه، ابوالقاسم گرجی حسن، نوروزی محسن، رضایی فاطمه. رابطه عدالت سازمانی به رضایت شغلی و تعهد سازمانی در بین کارکنان بیمارستان‌های منتخب دانشگاه علوم پزشکی اصفهان.
- آقالاری، صمصام و امید سجودی شیجانی، ۱۳۹۵، آرایه مدلی جدید برای پیش‌بینی وفاداری مشتری مبتنی بر وضعیت سازمان با روش‌های داده‌کاوی. مطالعه موردی بانک‌های خصوصی و دولتی تبریز، سومین کنفرانس بین‌المللی مدیریت و اقتصاد، تربت حیدریه، دانشگاه تربیت حیدریه.
- کیماسی مسعود، اسفیدانی محمدرحیم، اسماعیلیان امیربهمن. خوشه بندی وفاداری مشتریان باشگاه مشتریان بانک ملت بر اساس اطلاعات دموگرافیک با استفاده از روش داده کاوی.
- امیرالسادات، نسیم، شفیعی، حاجی پور، سلطانی (۲۰۱۹). شناسایی و الگوبردازی انگیزه‌های مصرف‌کنندگان برای مشارکت در خلق مشترک ارزش در صنعت نوشیدنی؛ شرکت زمزم ایران. تحقیقات بازاریابی نوین، ۹(۲)، ۵۳-۷۶.

- And p.s david hand, Heikki Mannila, "principles of Data Mining , "p.8290 ,2001.
- Adler, P. A., & Adler, P. (1988). Intense loyalty in organizations: A case study of college athletics. *Administrative Science Quarterly*, 401-417.
- Allen, N. J., & Meyer, J. P. (1996). Affective, continuance, and normative commitment to the organization: An examination of construct validity. *Journal of vocational behavior*, 49(3), 252-276.
- Allen,N.J. and J.P. Meyer, (1997), "Commitment in the workplace: Theory, research, and application", sage publications,inc.
- Becker, H.S., (1960), "Notes on the concept of commitment", *The American Journal of Sociology*, 66,1, JSTOR.
- Caruana, A. (2002). Service loyalty: The effects of service quality and the mediating role of customer satisfaction. *European Journal of marketing*, 36(7/8), 811-828.
- Grønholdt, L., &Martensen, A. (2001). Linking employee loyalty, customer loyalty and profitability. In *The 6th World Congress for Total Quality Management* (pp. 332-339).
- Ineson, E. M., Benke, E., & Laszlo, J. (2013). Employee loyalty in Hungariahotels. *International Journal of Hospitality Management*, 32, 31-39.
- Konovsky, M. A., & Pugh, S. D. (1994). Citizenship behavior and social exchange. *Academy of management journal*, 37(3), 656-669.
- LINOFF, Gordon S.; BERRY, Michael JA. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2011.
- Organ, W.P., Podsakoff, M.P., and Mackenzie,B.S. (2006). "Organizational citizenship bis nature antecedents and consequences". A sage publication series.research. *Journal of Management*, 26(3), 513-563.
- Peng, C. Y., & Chien, C. F. (2003). Data value development to enhance competitive advantage: A retrospective study of EDA systems for semiconductor fabrication. *International Journal of Services Technology and Management*, 4(4-6), 365-383.
- Pina e Cunha, M. (2002). Organizational improvisation: What, when, how and why. *Organizational improvisation*. KN Kamoche, M. Pina e Cunha and J. Vieira da Cunha. London and New York.
- Podsakoff, P.M., Mackenzie, S.B., Paine J., &Bachrach, D. (2000)..Organizational citizebehaviors: a critical review of the theoretical and empirical
- Powers,E. L. (2000). Employee loyalty in the new millennium. *S.A. M.AdvancedManageJournal*, 65(3),4-8.

- Robert, B. Gerinberg, J. (2014). Employee loyalty and firm performance. *Management Research News*, 30(3): 203-15.
- Šavareikienė, D., & Daugirdas, L. (2009). Forms of expression of employee loyalty. *Socialiniai tyrimai*, (2), 86-93.
- Şengür, D., & Turhan, M. (2018). Prediction of the action identification levels of teachers based on organizational commitment and job satisfaction by using k-nearest neighbors method. *Fırat University Turkish Journal of Science & Technology*, 13(2), 61-68.
- Testa, Mark.R. (2001) "Organizational Commitment, Job Satisfaction and Effort in the Service Environment". *Journal of Psychology*, 135(2): 226
- WANG, Guangping; LEE, Peggy D. Psychological empowerment and job satisfaction: An analysis of interactive effects. *Group & organization management*, 2009, 34.3: 271-296.
- Wu, C. H., Kao, S. C., Su, Y. Y., & Wu, C. C. (2005). Targeting customers via discovery knowledge for the insurance industry. *Expert Systems with Applications*, 29(2), 291-299.
- Yao, T., Qiu, Q., & Wei, Y. (2019). Retaining hotel employees as internal customers: Effect of organizational commitment on attitudinal and behavioral loyalty of employees. *International Journal of Hospitality Management*, 76, 1-8.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Hierarchical clustering. *Cluster analysis*, 5, 71-110.
- Hayat Bhatti, M., Ju, Y., Akram, U., Hasnat Bhatti, M., Akram, Z., & Bilal, M. (2019). Impact of participative leadership on organizational citizenship behavior: Mediating role of trust and moderating role of continuance commitment: Evidence from the Pakistan hotel industry. *Sustainability*, 11(4), 1170.
- Farrukh, M., Kalimuthuan, R., & Farrukh, S. (2019). Impact of Job Satisfaction and Mutual Trust on Employee Loyalty in Saudi Hospitality Industry: A Mediating Analysis of Leader Support. *International Journal of Business and Psychology*, 1(2), 30-52.
- Widodo, W., & Damayanti, R. (2020). Vitality of job satisfaction in mediation: the effect of reward and personality on organizational commitment. *Management Science Letters*, 10(9), 2131-2138.
- Sahni, J. (2019). Role Of Quality Of Work Life In Determining Employee Engagement And Organizational Commitment In Telecom Industry. *International Journal for Quality Research*, 13(2).

Abstrac

Today, the thinkers of the management organization state that the most important and valuable asset of the organization is its human resources. Therefore, it is important to strive to develop a commitment of loyalty in employees and create a sense of belonging in them, and with the proper selection of employees, they will take a big step towards the development of their organization. Due to technological advances in recent decades, the volume of information stored within organizations is increasing day by day. As a result, it is possible to use this vast amount of information to discover patterns and rules between different data, including human resource data, which is possible using data mining science.

The purpose of this study is to investigate the loyalty prediction of employees of private companies using data mining technique. This research is applied in terms of exploratory nature and purpose. Today, data mining methods are widely used in discovering patterns and rules between data. The application of these methods can be an effective role in predicting the loyalty of employees of private companies using data mining techniques in a case study of companies providing electricity distribution services in Khuzestan province. In this study, the obtained data were analyzed by Rapid Miner software and the performance of three algorithms from data mining methods including decision tree algorithm, backup vector machine and k nearest neighbor was investigated. Also, in this research, the combination of two top algorithms has been used as research innovation and obtaining a more desirable result.

Keywords: Employee Loyalty, Data Mining Technique, Decision Tree, Support Vector Machine, Nearest Neighborhood



Shahrood University of Technology
Faculty of Industrial Engineering and Management
Master's thesis MBA

Title

Investigate and anticipate the loyalty of private company employees using
data mining technique

(Case Study: Electricity Distribution Services Providers of
Khuzestan Province)

Written by: Fatemeh Nazari

Supervisor: Dr. M. H. Hosseini

1399