4- تجزیه و تحلیل داده ها

در این فصل، مدل مفهومی در ارتباط با بررسی تاثیر عوامل اصلی اثر گذار بر روی فرآیند آموزش منابع انسانی مورد تجزیه و تحلیل قرار می گیرد. مدلی که ارائه شد توسط روش داده کاوی با استفاده از نرم افزارهای SPSS و کلمنتاین مورد تجزیه و تحلیل قرار می گیرد. مدل مفهومی که ارائه می شود در واقع شامل متغیرهای پنهان می باشند. برای اندازه گیری این متغییرها از متغیرهای قابل مشاهده استفاده می شود. برای این منظور از باک اطلاعاتی موجود در بخش منابع انسانی سازمان صدا و سیما استفاده گردیده است. در این فصل به تجزیه و تحلیل داده های این تحقیق پرداخته می شود.

4.1 ساخت مدل

در این تحقیق جهت ارزیابی دوره های ضمن خدمت و سنجش رضایت شرکت کنندگان یک مدل ترکیبی داده کاوی ارائه شده است.مدل فوق از 4 بخش تشکیل شده است که در شکل 3 نشان داده شده است.

داده های منابع انسانی

بانک اطلاعاتی کار طرحکنان

بانک اطلابانک اطلاعاتی سازمانعاتی کارکنان

بانک اطلاعاتی سازمان بانک اطلاعاتی سامانه ثبت سامانه tms

انتخاب ویژگی ، پیش پردازش و آماده سازی داده ها

گام 1:به کاربردن تکنیک خوشه بندی جهت طبقه بندی

ویژگی ها یه خوشه های کم،متوسط و زیاد

گام 2:به کار بردن درخت تصمیمCHAID برای ویژگی های فراگیران

میزان رضایت از دوره های برگذار شده

گام 3:به کاربردن شبکه عصبی پرسپترون چندلايه جهت مشخص

کردن ویژگی هایی که بر نتیجه آزمون شرکت کنندگان تاثیر دارد

شکل 3. چارچوب فرآیند داده کاوی.

4.2تحلیل خوشه ای

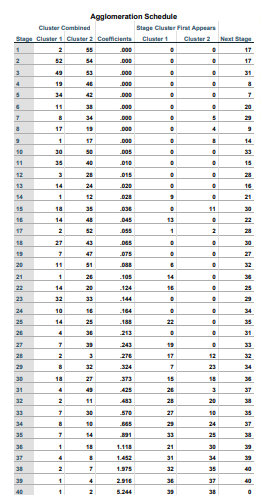
تحلیل خوشه‌بندی (Cluster Analysis) یا بطور خلاصه خوشه‌بندی، فرآیندی است که به کمک آن می‌توان مجموعه‌ای از اشیاء را به گروه‌های مجزا افراز کرد. هر افراز یک خوشه نامیده می‌شود. اعضاء هر خوشه با توجه به ویژگی‌هایی که دارند به یکدیگر بسیار شبیه هستند و در عوض میزان شباهت بین خوشه‌ها کمترین مقدار است. در چنین حالتی هدف از خوشه‌بندی، نسبت دادن برچسب‌هایی به اشیاء است که نشان دهنده عضویت هر شیء به خوشه است.

به این ترتیب تفاوت اصلی که بین تحلیل خوشه‌بندی و تحلیل طبقه‌بندی (Classification Analysis) وجود دارد، نداشتن برچسب‌های اولیه برای مشاهدات است. در نتیجه براساس ویژگی‌های مشترک و روش‌های اندازه‌گیری فاصله یا شباهت بین اشیاء، باید برچسب‌هایی بطور خودکار نسبت داده شوند. در حالیکه در طبقه‌بندی برچسب‌های اولیه موجود است و باید با استفاده از الگوی‌های پیش‌بینی قادر به برچسب گذاری برای مشاهدات جدید باشیم. نکته مهم دیگر آن است که خوشه بندی به طور آشکار از مجموعه های آموزشی استفاده نمی کند ، در حالی که طبقه بندی برای شناسایی ویژگی های مشابه به طور ضروری نیاز به مجموعه های آموزشی دارد.

معمولا ۴ گروه اصلی برای الگوریتم‌های خوشه‌بندی وجود دارد. الگوریتم‌های خوشه‌بندی تفکیکی، الگوریتم‌های خوشه‌بندی سلسله مراتبی، الگوریتم‌های خوشه‌بندی برمبنای چگالی و الگوریتم‌های خوشه‌بندی برمبنای مدل. در ادامه ما خوشه بندی سلسله مراتبی را بکار می بریم.

خوشه‌بندی سلسله مراتبی (Hierarchical Clustering)، در هر سطح از فاصله، نتیجه خوشه‌بندی را نشان می‌دهد. این سطوح به صورت سلسله مراتبی (Hierarchy) هستند.برای نمایش نتایج خوشه‌بندی به صورت سلسله مراتبی از درختواره (Dendrogram) استفاده می‌شود. سه تکنیک خوشه بندی linkage Single ، Ward و Flexible-β مربوط به روش سلسله مراتبی هستند. روش Ward در این روش به عنوان دقیقترین روش سلسله مراتبی، هر عضو در گروهی جای می گیرد که مجموع مربعات انحرافات درون گروهی به حداقل برسد. برای خوشه بندی با این روش، ماتریس های ورودی همان ماتریس های ورودی به روش تجزیه به مؤلفه های اصلی است .

با توجه به اینکه داد ه های ما در مورد ویژگی های رضایت از اجرا،رضایت از محتوا، رضایت از مدرس و ضریب یادگیری که در دیتابیس به صورت پیوسته ذخیره شده اند، ابتدا توسط با روش خوشه بندی سلسله مراتبی آن ها را به سه خوشه با برچسب های کم،متوسط و زیاد طبقه بندی می کنیم تا خوانایی داده ها افزایش یابد. در ادامه روش خوشه بندی سلسله مراتبی برای ویژگی رضایت از اجرا توضیح می دهیم برای بقیه هم به همین صورت عمل می کنیم.



جدول 3 فرآیند تجمع خوشه ای برای ویژگی رضایت از اجرا

در شکل 4 ، نمودار درختی را با نتیجه روند تجمع و ترتیب همگرایی بالا اجازه می دهد کاربران به یک تجزیه و تحلیل واضح ترکمک می کند، سپس داده های داده های تجزیه و تحلیل شده در جدول 4 سازماندهی مجدد می شود. سه مورد زیر برای خوشه های میزان رضایت از اجرا وجود دارد.

خوشه ها:

(1) خوشه 1 شامل دوره هایی با میزان رضایت اجرا متوسط است .

(2) خوشه 2 دوره دوره هایی با میزان رضای اجرا بالا را شامل می شود

(3) خوشه 3 دوره هایی با رضایت اجرا کم را شامل می شود.



شکل 3. نمودار درختی نتیجه طبقه بندی تجزیه و تحلیل خوشه سلسله مراتبی

جدول 4 نتیجه طبقه بندی

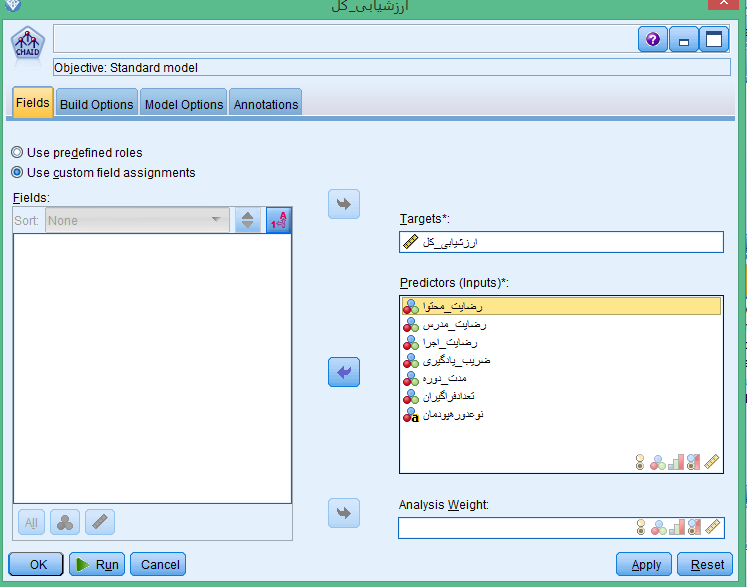
|  |  |  |  |
| --- | --- | --- | --- |
| با سطح رضایت از اجرا کم | با سطح رضایت از اجرا بالا | با سطح رضایت از اجرا متوسط | خوشه |
| آداب پذیرایی و تشریفات(2) | vMWARE vSPHERE OPTIMIZE AND SCALE (O&S)-مجازی سازی سرور پیشرفته | SNGوپیکره بندی چند برنامه ای مراکز |  |
| آنونس خوانی و آیتم خوانی، ویژه گویندگان | آداب پذیرایی و تشریفات(1) | با تمرکز بر تکنیک های عملیات روانی و کنترل ذهن در رسانه | دوره های آموزشی |
| اصول ارزیابی محتوایی برنامه های تلویزیونی | آشنایی با قوانین، مقررات و فرایند های اداری | خبرجعلی و راهکارهای مقابله با آن(2مورد) |  |
| تعمیر و نگهداری از فرستنده های تلویزیونی کم قدرت فناموج | اصول شبکه و شبکه های کامپیوتری ( NETWORK+) با رویکرد مانیتورینگ و سیگنال رسانی | خبرنویسی سطح یک |  |
| فنون خلاقیت در اجرای برنامه های رادیویی(1و2و3) | بهره برداري و نگهداري از دستگاه راك صوتي | سوژه سازی و سوژه یابی در خبر |  |
| فنون طراحی صحنه و دکور | دوربین استودیویی SONY HSC-100 | شبکه تلفنیvoip |  |
| مجری گری گزارش خبری (انگلیسی) | زبان بدن | فنون خلاقیت در اجرای برنامه های رادیویی(4) |  |
| نامه نگاری و گزارش نویسی اداری | زیبایی شناسی تصویر خبری | گام دوم انقلاب اسلامی و رسانه ملی |  |
| کاربرد پژوهش در تولید برنامه های رادیویی | صدابرداری خبری | مستندسازی تلویزیون سطح یک |  |
|  | فرايندهاي تامين نيروي انساني | میکسر تصویر kula |  |
|  | فنون کاربردی تولید افکت در برنامه های رادیو |  |  |
|  | قوانین کاربردی مالیات در سازمان |  |  |
|  | میکسر صدای SX – DHD(2مورد) |  |  |
|  | نمایه سازی منابع دیداری شنیداری |  |  |
|  | نویسندگی رادیو |  |  |
|  | کاربرد فنون جدید تهیه کنندگی در تولید برنامه های ورزشی رادیو |  |  |
|  | کاربرد موسيقی ايرانی در رسانه |  |  |

4.3 الگوریتم درخت تصمیم CHAID در داده های مربوط به دوره های آموزشی

الگوریتم درخت تصمیم قادر است علاوه بر متغیر های کمی،متغیرهای کیفی را نیز پیش بینی کند. نتیجه پیاده سازی الگوریتم درخت تصمیم مجموعه ای از شرط های منطقی(if-then وconditions) با ساختار درختی است که برای پیش بینی یک ویژگی هدف برچسپ می خورند. این مدل به دلیل سهولت در تفسیر نتایج و ناپارامتری و غیر خطی بودن،نیاز به پیش فرض رابطه خطی بین متغیرهای مستقل و وابسته ندارد. الگوریتم های درخت تصمیم به گونه ای عمل می کنند که سعی دارند گوناگونی و یا تنوع(از نظر ویژگی هدف)را در گره ها به حداقل ممکن برساند.اغلب تفاوت انواع درخت های تصمیم در همین معیار اندازه گیری عدم خلوص،شیوه شاخه بندی و هرس کردن شاخه های درخت می باشد.انواع مختلفی از درخت تصمیم مانند CART، CHAID وC5.0 وجود دارندکه در این تحقیق از درخت CHAID استفاده شده است.

در روش CHAID، که به «شناسایی اثرات متقابل خودکار کای ۲» (Chi-squared Automatic Interaction Detection) نیز شهرت دارد، در هر گام، متغیر پیشگویی که بیشترین میزان ارتباط با متغیر وابسته را دارد در مدل و درخت تصمیم به کار می‌رود. سطوح یا طبقه‌های هر متغیر پیشگو ممکن است در این حالت با یکدیگر ادغام شوند زیرا سطح معنی‌داری ممکن است در هر طبقه کمتر از مقدار مورد انتظار باشد.

CHAID،برای ساخت یک درخت تصمیم ،داده ها را به طور متناوب به زیر مجموعه های مشابه افراز می کند. تا آنجا که هر زیر مجموعه دارای تعداد مشخصی نمونه شود. این الگوریتم می تواند درختی تولید کند که در برخی مواقع به صورت غیر دودوئی عمل کند. در واقع از روش جداکردن چندتائی به جای جدا کردن دودوئی ایتفاده می کند. به این صورت که می توان نود پدر را به بیش از دو تقسیم نماید.این الگوریتم از آزمون کای دو برای تصمیم گیری در هر تقسیم برای مشخص کردن نودهای فرزند برای مشخص کردن نودهای فرزند اسنفاده می کند. سپس شاخه های درخت ساخته شده تا تحقق معیار توقف یا رسیدن به سطح پیچیدگی خواسته شده هرس می شود. به بیان دیگر CHAID ابتدا تفاوت های هر نمونه را با سایر نمونه ها می یابد و درخت مورد نظر را تولید می کند.هرس کردن درخت از طریق یافتن تفاوت های مشابه انجام می شود.ابتدا داده ها را به دو قسمت 80 درصد برای آموزش و 20 درصد برای تست تقسیم می کنیم.



شکل 4.ویژگی های بکار رفته در ساخت درخت CHAID مربوط به دوره های آموزشی

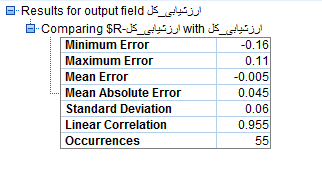
در نمودار درختی در شکل 5،در سطح اول رضایت از محتوای دوره های آموزشی برگزار شده را می بینیم که به با روش خوشه بندی به سه دسته میزان رضایت کم( 3.55 تا 4.09)،متوسط(4.24تا4.6) و زیاد(بیش از 4.6) تقسیم بندی شده است که گره اول مربوط میزان رضایت متوسط با حدود 63 درصد است و از بیشترین میزان برخوردار است سپس رضایت زیاد که حدود 20 از کل را داراست وسطح رضایت کم حدود 16.6 درصد را داراست.که نشن دهنده این است که سطح رضایت از محتوا تاثیر زیادی بر رضایت کلی از دوره برگزار شده است. از آنجا که مقدار p هرسه کمتر از استاندارد قابل توجه است که دستی مقدار 05/0 تنظیم شده است ، شاخه های بیشتری در سطح دوم برای هر سه گره خواهیم داشت. . در گره اول ، رضایت از اجرا به دو دسته تقسیم شده است.گره چهارم میزان رضایت اجرا کم و متوسط(از 3.96 تا4.5) که شامل 30 درصد آمار گره اول است. و گره پنجم میزان رضایت از اجرا زیاد(از 4.5) است که 33 درصد بقیه موارد موجود در گره اول است. . در گره دوم نیز ، میزان رضایت از اجرا به دو دسته تقسیم شده است.گره ششم میزان رضایت اجرا کم و متوسط(از 3.96 تا4.5) که شامل 13.3درصد آمار گره دوم است. و گره هفتم میزان رضایت از اجرا زیاد(بیشتراز 4.5) است که 3.3 درصد بقیه موارد موجود در گره دوم است.

همچنین در گره سوم ، رضایت از اجرا به دو دسته تقسیم شده است.گره هشتممیزان رضایت اجرا زیاد(بیشتر از 4.5) که شامل 16.6 درصد آمار گره سوم است. و گره نهم میزان رضایت از اجرا کم(کمتر از 4.3) است که 3.3 درصد بقیه موارد موجود در گره سوم را شامل می شود.

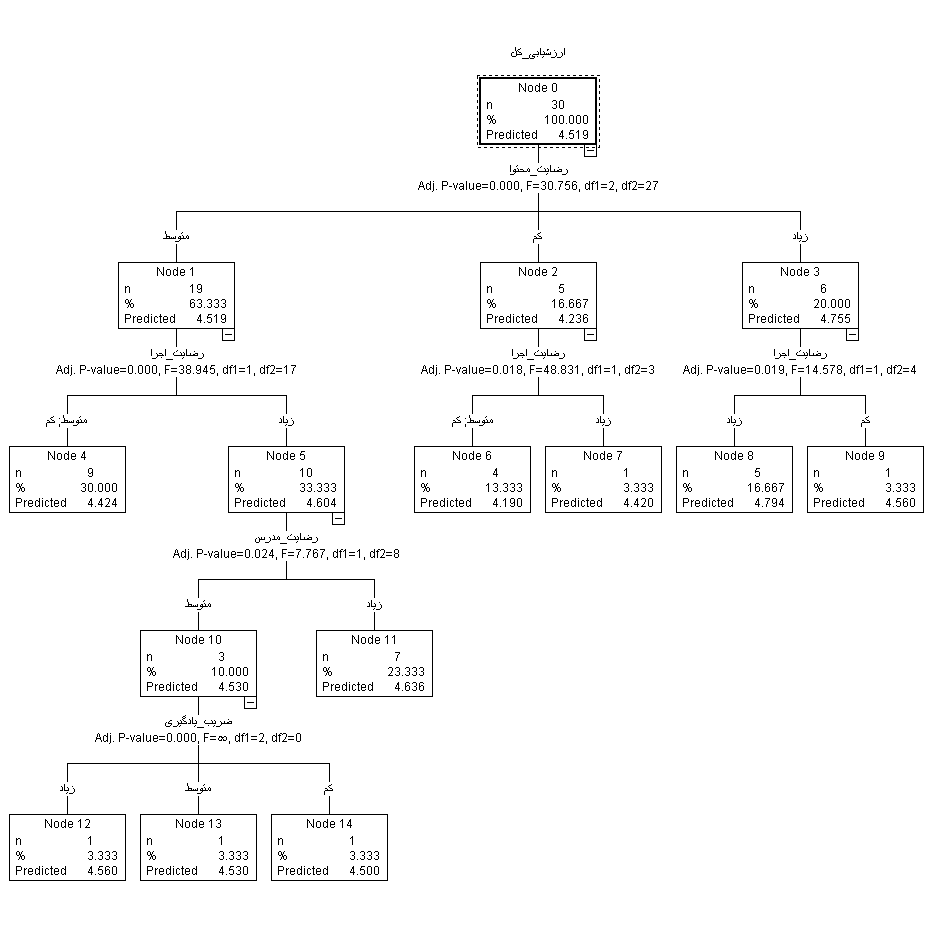
در سطح سوم گره پنجم به دو گره که سطح سطح رضایت از مدرس را نشان می دهدکه گره دهم مقادیر متوسط(از 4.25 تا4.65) با میزان 10 درصد و گره یازدهم زیاد(بیشتر از 4.65) با مقدار 23.3 درصد از آمار گره پنجم است.

گره دهم در سطچ چهارم متوسط ضریب یادگیری را به 3 گره دوازدهم زیاد(مقدار ضریب یادگیری بیش از 89) با3.33 درصد وگره سیزدهم متوسط(بین 50 تا80) با3.33 درصد و کم(کمتراز 50)با 3.33 درصد آمار کلی گره دهم می باشد.

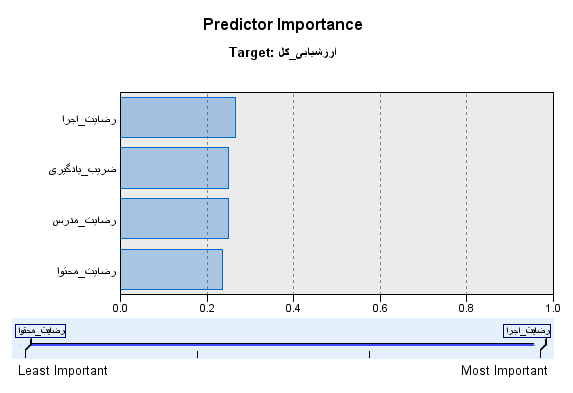
در شکل 6 مربوط به نمودار درختی فوق ارزش متغیر هااز لحاظ اهمت برای ساخت درخت ذکر شده است که به ترتیب اهمیت عبارت اند از:میزان رضایت از اجرا،ضریب یادگیری ،میزان رضایت از مدرس و در نهایت میزان رضایت از محتوا ذکر شده است و از سایر ویژگی ها استفاده نشده است. با توجه به جدول 5 ، در می یابیم که میزان دقت تحلیلی داده ها تا 95.5٪ است.



جدول 5 تجزیه و تحلیل میزان دقت داده ها در نمودار درختCHAID



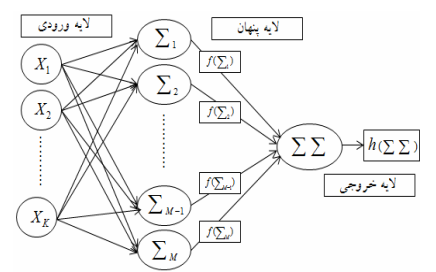
شکل 5. نمودار درختی ارزشیابی دوره های آموزشی برگزار شده



شکل 6 متغیر های موثر به کار رفته در درخت CHAID فوق به ترتیب اهمت

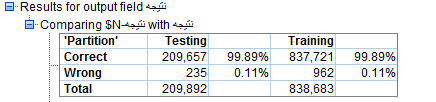
4.4شبکه عصبی پریستون چند لایهMLP در داده های مربوط به نتیجه دوره های آموزشی

یکی ازانواع شبکه های عصبی، شبکه عصبی پرسپترون می باشد وبصورت های پرسپترون تک لایه(SLP) پرسپترون چند لایه موجودهستند. شبکه عصبی پرسپترون جزء شبکه های عصبیپ یشخور طبقه بندی می شود. پرسپترون تکلایه تنها میتواند مسائل مجزای خطی رادسته بندی کند وبرای مسائل پیچیده تر لازم است که ازتعدادبیشتری لایهاستفاده کنیم. شبکه های عصبی پرسپترون چند لایه ازیک ویا تعدادبیشتری لایه های میانی تشکیل شده اند. سیگنال های ورودی به وسیله ضریبهای بهنجارکننده نرمالیزه شده وبعد ازمحاسبات، خروجی به مقدارواقعی برگردانده می شود. مقادیراولیه وزنها نیزبه صورت اتفاقی درنظرگرفته می شوند. شکل 7 یک شبکه عصبی پرسپترون چندلایه را نمایش می دهد. دراین شبکه شامل مجموعه ای از نورون ها است که در لایه مختلفی پشت سر هم قرار گرفته اند. مقادیر ورودی پس از ضرب در وزن های موجود در گذرگاه های بین لایه ها به نورون بعدی رسیده و در آن جا با هم جمع می شوند و پس از عبور از تابع شبکه مربوطه، خروجی نورون ها را تشکیل می دهند. در پایان خروجی به دست آمده با خروجی مورد نظر مقایسه شده و خطای به دست آمده جهت اصلاح وزن های شبکه به کار می رود، این امر در اصطلاح آموزش شبکه عصبی نامیده می شود. این چنین، شبکه ها معمولا با روش پس انتشار خطا انجام می شود. نمونه ای از یک شبکه پرسپترون چند لایه در شکل 7 نمایش داده شده است.



شکل 7. ساختار شبکه های عصبی پرسپترون چندلايه

شبكه هاي عصبي پرسپترون چندلايه برخلاف روشهاي سـنتي پيش بيني همچون رگرسيون، داراي محدوديت تعداد متغيرهـاي ورودي هستند. در روشهايي همچون رگرسيون عملکرد مـدل با تعداد متغيرهاي ورودي توضيح دهنده رابطه مسـتقيم غيراکيـد دارد، به عبارت ديگر با افزايش تعداد متغيرهاي توضـيح دهنـده، عملکرد مدل کاهش پيدا نخواهد کرد. اما در شبکههاي عصـبي چندلايه پيشخور همچون پرسپترونهـاي چندلايـه ايـن مسـئله برقرار نميباشد. لذا مسئله تعيـين تعـداد متغيرهـاي ورودي بـه شبکه به مسئلهاي مهم و تأثيرگذار در طراحي شبکه هاي عصبي تبديل شده است. هر چند قانون تعريـف شـدهاي بـراي تعـداد وروديهاي شبكه، در مسائل مختلف تعيين نشده است،امـا درحالت کلي تعداد وروديهاي شبکه به ساختار شبكه، الگـوريتم آموزش، پيچيدگي مسـئله مـورد نظـر و اخـتلالات موجـود در داده ها وابسته مي باشد.در این تحقیق از شبکه عصبی پریستون چند لایه با 80 درصد داده آموزشی و 20 درصد داده آزمون استفاده شده است.در لایه ورودی ، 5متغیر وجود دارد (نوع دوره/پودمان ، عنوان شغلی،مقطع تحصیلی ، رشته و گرایش تحصیلی ، حوزه کاری) که می تواند بر نتیجه کسب شده دوره ها توسط کارکنان تأثیر بگذارد و در شکل 8 به ترتیب اهمیتی که در ساخت مدل شبکه عصبی داشته اند، نشان داده شده است. در مورد خروجی متغیر نتیجه که یک متغیرچند مقداری با مقادیر حاضر، غیبت غیر مجاز،غیبت مجاز،قبول،مردود و نامشخص است به کار رفته است.در لایه مخفی ، تعداد 10 لایه مخفی به کار برده شده است.و دقت مدل همان طور که در جدول 6 نشان داده شده است، 99.89 درصد می باشد.

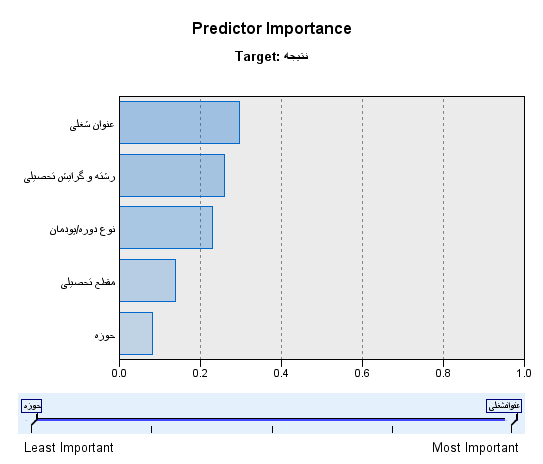


جدول 6 جزییات دقت مدل شبکه های عصبی برای پیش بینی نتیجه دوره های آموزشی برای داده های آموزش و آزمایش.

با نتیجه ، ما طبقه بندی را انجام می دهیم و کسب می کنیم میزان دقت طبقه بندی نتیجه آموزش (1197 نمونه) (جدول 7 را ببینید). پس از سازماندهی مجدد ، از 1197 نمونه 10 موردبرای تست و 53مورد آموزش در سطح اول یعنی حاضر طبقه بندی و پیش بینی کرده. 60 موردبرای تست و 202مورد آموزش در سطح دوم یعنی غیبت غیر مجاز طبقه بندی و پیش بینی شد، 5موردبرای تست و 243مورد آموزش در سطح سوم یعنی غیبت مجاز طبقه بندی و پیش بینی شد،154 موردبرای تست و 605مورد آموزش در سطح چهارم یعنی قبول طبقه بندی و پیش بینی کردهاست.همچنین6 موردبرای تست و 70مورد آموزش در سطح پنجم یعنی مردود طبقه بندی و پیش بینی شده است و مقدار 8 مورد با مقدار نامشخص در داده های آموزش پیدا شده است. با توجه به شبکه های عصبی پرسپترون چندلايه در این تحقیق ، متوجه می شویم که میزان دقت طبقه بندی شبکه های عصبی پرسپترون چندلايه در نتیجه آموزش تا 99.9٪ است.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | نتایج طبقه بندی و پیش بینی (آموزش) | | | |
| جمع | مقادیر خالی | تعداد |  |  |
| 10 | 3 | 7 | حاضر | داده آزمایش |
| 60 | 17 | 43 | غیبت غیرمجاز |
| 5 | 0 | 5 | غیبت مجاز |
| 154 | 19 | 135 | قبول |
| 6 | 0 | 6 | مردود |
| 53 | 0 | 53 | حاضر | داده آموزش |
| 202 | 2 | 200 | غیبت غیرمجاز |
| 24 | 11 | 13 | غیبت مجاز |
| 605 | 2 | 603 | قبول |  |
| 70 | 36 | 34 | مردود |  |
| 8 | 2 | 6 | نامشخص |  |
| 1197 |  |  | میزان دقت نتیجه آموزش: 99.9٪ | |

جدول 7 ماتریس انطباق برای نتیجه دوره آموزشی کارکنان



شکل 8 متغیر های موثر به کار رفته در شبکه عصبی فوق به ترتیب اهمیت