

کاربرد تکنیک های داده کاوی در پیش بینی زمان مساعد برداشت محصول زراعی

زهرا زیران^۱ و شیرین حسینیان^۲

^۱ کارشناسی ارشد مهندسی فناوری اطلاعات (تجارت الکترونیک)، دانشگاه گیلان،
^۲ کارشناسی ارشد مهندسی فناوری اطلاعات (تجارت الکترونیک)، دانشگاه گیلان،

چکیده - نظر به اینکه تولید مواد غذایی بهتر و بیشتر یکی از مسائل مهم دنیای کنونی بشمار می‌رود و با توجه به اینکه تولید محصول و قابلیت‌های کشاورزی هر منطقه به هوا و مشخصات اقلیمی آن بستگی دارد لذا مطالعه در زمینه پیش بینی عوامل جوی موثر بر کشاورزی از اهمیت ویژه‌ای برخوردار است. برداشت غلات باید به موقع صورت گیرد. برداشت زودتر یا دیرتر از موقع محصول، موجب کاسته شدن کیفیت آن می‌شود. تأخیر در برداشت غلات دانه ریز، سبب ریزش دانه، خوابیدگی یا شکستن بوته‌ها در کشتزار و بالاخره کاهش میزان ماده خشک می‌گردد. برداشت زود هنگام محصول هم موجب پایین آمدن کیفیت غلات می‌گردد.. ازاینرو در این تحقیق به اطلاع رسانی به موقع درخصوص زمان کاشت و برداشت محصول با توجه به داده های هواشناسی و دستورالعمل فنی دراین خصوص اشاره شده است. منطقه‌ی مطالعاتی این تحقیق استان گیلان است. در ابتدا داده های ۱۵ سال اخیر مورد تحلیل قرار گرفته و پارامترهای تاثیرگذار استخراج شده است. تکنیک های مورد استفاده در این بحث شامل الگوریتم های رده بندی در داده کاوی می باشد. در انتهای بین تکنیک های استفاده شده مقایسه ای انجام شده و بهترین روش برای پیش بینی زمان مناسب برداشت برنج انتخاب گردیده است. این تحقیق می تواند هزینه های ناشی از برداشت بدنهنگام برنج را کاهش دهد.

کلمات کلیدی: پیش بینی، درخت تصمیم، داده کاوی، شبکه عصبی و طبقه بندی

اما هواشناسی بعنوان خدمات و حرفه، شاخه‌ای است از علوم فنی و مهندسی که با اندازه گیری، جمع آوری داده و پردازش داده با هدف تجزیه و تحلیل اتمسفر زمین سرو کار دارد و با فن آوری های جدید همواره ارتباط داشته است [۱]. کنگاوا انسان برای فهم علت تغییرات هوا سابقه ای تاریخی دارد. تغییرات هوا انسان را همواره شگفت زده می کرد و هنوز هم مایه اعجاب انسان است. هیچ نیرویی در طبیعت به اندازه نیروهای اتمسفر زمین که گاهی لجام گسیخته رها می شوند قوی و قدرمند نیستند. طوفان های شدید با وزش باد شدید و بارندگی های سیل آسا و ایجاد سیل های مهیب انسان را همواره مروع و مجبور به پناه گیری و فرار نموده است و امروزه هم به نحوی این ماجرا ادامه دارد. بدلیل پیشرفت در شناخت تغییرات اقلیمی گاهی به سوالات اقلیمی در گذشته پاسخ های درست داده می شد. مثلاً اگر سؤال می شد که آب و هوای منطقه ای چگونه است، جواب کیفی نسبتاً نزدیکی به این سؤال داده می شد. اما این پرسش در گذشته پاسخ کمی دریافت نمی کرد. زیرا هیچ عامل جوی در گذشته شناخته شده نبود و اندازه گیری نمی شد [۲]. و این سؤال

۱- مقدمه

علیرغم توسعه های زیاده فناوریهای کشاورزی در چند دهه اخیر، تولیدات کشاورزی هنوز هم به وضعیت جوی و اقلیمی منطقه بستگی دارد شناخت شرایط جوی و در نظر گرفتن اوضاع اقلیم در منطقه مورد نظر در برنامه های توسعه اقتصادی و اجتماعی از شرایط لازم است. چرا که دانش هواشناسی در تمام مراحل برنامه ریزی، اجرا و مرحله بازدهی فعالیتهای کشاورزی نقش مؤثری را ایفا می کند و بدون استثناء با استفاده از خدمات هواشناسی و اطلاعات اقلیمی در تمام بخش‌های کشاورزی، می توان هزینه تولید را کاهش و بهره وری را افزایش داد.

۲- سرگذشت علم هواشناسی

هواشناسی به عنوان علم شاخه ای از علم فیزیک است در گروه علوم زمین یعنی علومی که کره زمین را مطالعه و بررسی می نمایند.

در جو (رطوبت)، میزان ابر و نوع آن، میزان بارندگی، نوع پدیده جوی و سمت و سرعت باد و..... در گذشته بخوبی شناخته شده نبودند. هوا توضیحات کیفی مانند سردی و گرمی، خشکی و رطوبت و بادخیزی داشتند که همه جملات کیفی بودند که حد و مرزی مشخص نداشتند [4]. هواشناسی در کشاورزی در بخش های مختلف علوم زیست محیطی (کشاورزی، جنگلداری، منابع طبیعی و محیط زیست) اهمیت دارد در این مقاله برخی از تواناییهای این دانش معرفی شده که عبارتند از:

پیش بینی تاریخهای کشت مراحل نموگیاهی و عملکردمحصول. پیش بینی شرایط حرارتی در طی فصل رشد تعیین درجه حرارت حداقل کاشت و تعیین درجه حرارت بحرانی نقطه انجمادگیاه چهت پیش بینی بخ زدن محصول- توانایی در بکارگیری پیش بینی های بلندمدت برای کشاورزان چهت مراحل کاشت، داشت و برداشت برنج، تشریح و کاربردمدهای پیش بینی عملکردمحصول [5].

پیش بینی شرایط حرارتی در فصل رشد گیاهان گرمابسند مانند برنج و اطلاع رسانی به هنگام و به موقع که برای کشاورزان، از اهمیت فراوانی برخوردار است. به طور مثال هنگام برداشت برنج بایستی از وضعیت هوا اطلاعات کافی بدست آورده و آنگاه به برداشت محصول برنج اقدام کنیم. مثلاً اگر برنج درو نشود و در زمین باقی مانده و آنگاه باران ببارد، محصول برنج آسیب چندانی نخواهد دید، اما به محض اینکه این محصول درو شود و باران ببارد، برنج مرغوبیت خود را از دست داده و غیرقابل عرضه به بازار مصرف می گردد. حال پیش بینی هواشناسی با استفاده از علم داده کاوی در حجم اینبوهی از داده ها می تواند مشکلات ناشی از این قبیل نا آگاهی را تا حدودی برطرف نماید.

۲-۱- کاربرد علم داده کاوی در پیش بینی هواشناسی
امروزه با پیشرفت علم و تکنولوژی در صنایع گوناگون حجم داده ها روز به روز افزایش می یابد به طوری که توانایی های ما در تحلیل این داده ها با چالش جدی روپرورد شده است. همین مسئله موجب شده است تا نیاز به رویکردهای نوین در این زمینه به منظور تولید، جمع آوری و تحلیل داده ها پیش از پیش احساس شود. داده کاوی در واقع ابزاری است برای تحلیل داده ها که شامل رویکردهای جدید است، ابزاری که به طور بالقوه به منظور آشکارسازی اطلاعات نهان در داده ها بسیار مفید است [7,6]. داده کاوی تحلیل مجموعه داده های مشاهده ای (اغلب بزرگ)، به منظور پیدا کردن روابط و خلاصه کردن داده ها به بهترین نوع است به طوری که برای دارنده اطلاعات قابل فهم و مفید باشد. روابط و خلاصه های بودست آمده از طریق داده کاوی عموماً به عنوان مدل یا الگو در نظر گرفته می شوند. هرساله

طرح است که اگر سازوکار آب و هوا را درک کنیم آیا این امکان وجود دارد تا شرایط آب و هوایی آینده کره ای زمین را کنترل نماییم؟ خوشبختانه با وجود فراهم بودن اطلاعات بهتری نسبت به گذشته و پیشرفت ها در علم که اخیراً اتفاق افتاده است، پیش بینی های آب و هوایی به صورت مساله ای کاملاً با اهمیت درآمده است. کاربرد عملی این دانش با قابلیت دسترسی و صحت پیش بینی های وضعیت جوی بالگوهای جوی موردنظر بسته به مقیاس زمانی، درارتباط می باشد. محدوده نیازهای پیش بینی هوا برای کشاورزی از پیش بینی کوتاه مدت دقیق تفصیلی هوا (یک الی سه روزه) تا پیش گویی های فصلی الگوهای هوایی تغییر می کند [3].

۲-۲- اهمیت اقتصادی و اجتماعی خدمات هواشناسی
با توجه به اهمیت و اثراتی که آب و هوا در زندگی و نحوه فعالیت جوامع بشری دارد، انسان ها از زمانهای خیلی دور بدان توجه داشته اند، بنابراین بر حسب تجارب، پیشینیان نسبت به ثبت پدیده های جوی به سبک زمان خود اقدام نموده اند، که با توجه به اقلیم هر منطقه، اقوام پدیده های مهمتر را مد نظر قرارداده اند، این نوع فعالیت ها در حقیقت پایه گذار دانش هواشناسی در میان نسل بشر بوده است. در این میان با پیشرفت علم و فن آوری و بکارگیری نیروهای متخصص، هواشناسی به جایگاه واقعی خود نزدیک می شود و با ارائه پیش بینی ها و پیش آگاهی ها نقش واقعی خود را به معرض ظهور می رساند. از زمان آغاز شکل گیری جو تاکنون، آب و هوا در مقیاس های زمانی و به دلایل فراوان دایما در حال تغییر و تحول بوده است. به موازات تغییرات آب و هوایی، واکنش ها در برابر اثرات متقابل همراه با آنها دایماً الگوهای حیات را دستخوش تحول نموده اند. در کوتاه زمانی از قرن، استفاده گسترده از امکانات هواشناسی جایگزین حافظه بشری و فهرست رویدادهای کیفی به همراه منابع اطلاعات کمی در زمینه آب و هوا تغییر یافته است و چگونه تغییرات آب و هوایی گذشته که فرهنگ های کره زمین را آراسته اند و آنها تحت تاثیر قرار داده اند جایگزین شده است. اکنون مساله این است که مفهوم تاریخ غنی طبیعی را به منظور پیش بینی حوادث آب و هوایی آینده و آینده نگری برای آنها کاملاً درک کنیم. این مساله امری بسیار مهم برای بشریت است اگر تصمیم به بقا در میان عصر یخچالی آینده بدون از دست دادن ثمرات عظیم تمدن خود داشته باشد. چنانچه بخواهیم نتایجی درباره سیر تمایلات آب و هوا در زمینه تغییر یافتن آن ترسیم کنیم باشد جریانات دیگری را نیز که از اهمیت برخوردار است به خاطر داشته باشیم. عوامل جوی یعنی دما، فشار، میزان بخار آب موجود

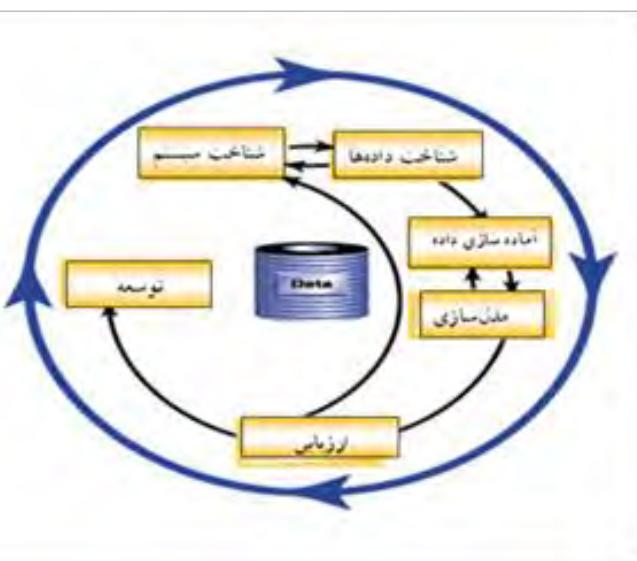
۳-۲ - متدلوژی داده کاوی

متدلوژی مورد استفاده در این مقاله، مدل *CRISP1* می باشد، این مدل، فعالیت های داده کاوی به شش مرحله، شناخت سیستم، شناخت داده، آماده سازی داده ها، مدلسازی و ارزیابی و توسعه ی سیستم تشکیل شده است. هر کدام از این گامها به زیر بخش هایی تقسیم می شوند. شکل (۱) روند اجرای مراحل مختلف در مدل فرایندی *CRISP* را نمایش می دهد [۱۰].

رخدادهای در هم گسیخته آب و هوایی ویرانگری رخ می دهد که حمل و نقل و در نتیجه از دست دادن زندگی و دارایی مردم را سبب می شود. با این وجود تحقیقات اندکی در زمینه وضعیت آب و هوایی مناطق گزارش شده است. دانش داده کاوی یکی ازده علم برتر قرن اخیر است. این دانش با بهره گیری از علوم مختلف همچون آمار، ریاضیات و هوش مصنوعی پاسخ گوی نیازهای هواشناسی است و به عنوان بینی و کشف الگوهای پنهان در داده های هواشناسی فرآگیری پل ارتباطی میان علم آمار، هوش مصنوعی، الگو شناسی، فرآگیری ماشین و بازنمایی بصری شناخته می شود. کاربردهای این علم در زمینه های مختلف مورد توجه دانشمندان قرار گرفته و نتایج بدست آمده از آن باعث تشویق پژوهش گران به استفاده گسترده تر از علم داده کاوی شده است. روش های پیش بینی کننده می توانند ممکن ترین حادثه را برای نزدیک ترین بازه زمانی پیش بینی کنند به عنوان مثال می توان در یک مجموعه داده در یک یا چندسال، میزان اثرگذاری تک تک پارامترها را بر وقوع حوادث بررسی کرد و از آنها برای پیش بینی و جلوگیری از وقوع رخدادها در آینده استفاده نمود [۸]. تکنیک های داده کاوی به طور تاریخی به گونه ای گسترش یافته اند که به سادگی می توان آنها را بر ابزارهای نرم افزاری امروزی تطبیق داده و از اطلاعات جمع آوری شده بهترین بهره را برد. در صورتی که سیستم های داده کاوی بر روی سکوهای سرور قوی نصب شده باشد و دسترسی به بانک های اطلاعاتی بزرگ فراهم باشد، می توان به سوالاتی از قبیل در کدام منطقه دمای هوا کمتر است و چرا، در کدام مقطع زمانی و بسیاری از موارد مشابه پاسخ داد [۹].

امر کاشت، داشت و برداشت محصولات کشاورزی بدون در نظر گرفتن پارامترهای هواشناسی دارای بازده اندک و همراه با خسارت احتمالی زیاد است. هواشناسی علاوه بر تحقیقات در مورد سازگاری گونه های مختلف گیاهی با آب و هوا در تعیین مناسب ترین زمان کاشت، داشت و برداشت محصول، سمپاشی، کوددهی و آبیاری و کاهش خسارت جوی بوبیزه سرمزدگی و پیش بینی آفات گیاهی می تواند مسؤولان و کشاورزان را یاری دهد. حال پیش بینی کوتاه مدت هواشناسی گاه نمی تواند برای کشاورزان مفید واقع شود و برای تصمیم گیری نیاز به پیشینه هواشناسی بلند مدت دارند که در این پژوهش جهت رفع مشکلات فوق الذکر و برای تسهیل روند برنامه ریزی بهنگام کاشت برنج در استان گیلان و برای پیش بینی احتمال سرمای دیررس در فصل کاشت، یک مدل داده کاوی مبتنی بر درختان تصمیم گیری توسعه داده شده است.

1. Cross Industry Standard Process for Data Mining



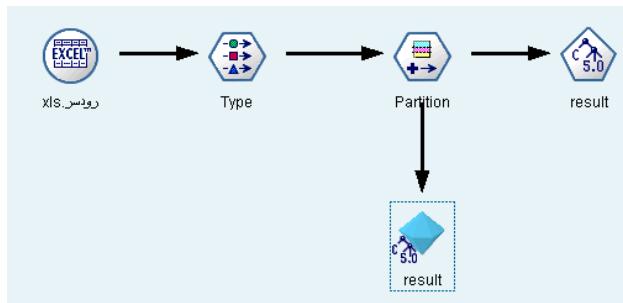
شکل(۱): روند اجرای مراحل داده کاوی در مدل فرایندی *CRISP*

۴-۲ - داده های جمع آوری شده

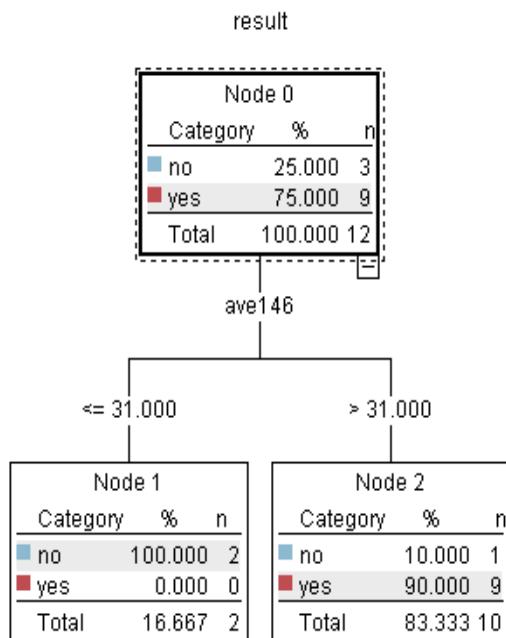
داده های جمع آوری شده ۲ ویژگی را شامل می شوند، که با استفاده از این ویژگی ها میانگین دما به دست آمده است. این داده ها از ۲۱۹۰ رکورد تشکیل شده اند که مربوط به سال های ۱۳۷۶ الی ۱۳۹۰ می باشند. در جدول (۱) داده ها معرفی شده اند.

جدول (۱) مشخصات فیلد های پایگاه داده مورد استفاده

عنوان فیلد	مقادیر
سال	از ۱۳۷۶ تا ۱۳۹۰ سال



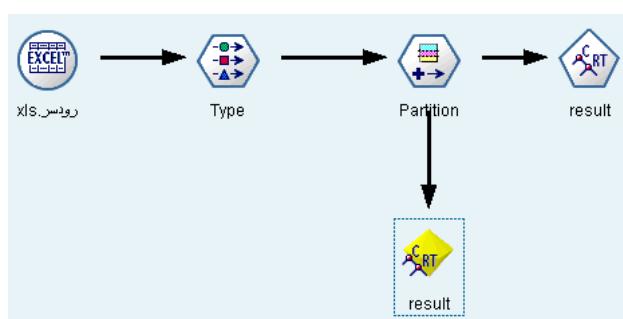
شکل(۲): مراحل داده کاوی در مدل C5.0



شکل(۳): نتیجه اجرای الگوریتم C5.0 بر روی داده های آزمون

۳-۳- پیش بینی براساس روش C&R Tree

-۴-۳



از ۱ تا ۳۱ روز در هرماه از اوایل سال (فروردين الى مرداد)	دماي هوا AVE
بله / خير	شرابيط مساعد برداشت

۳- پیش بینی بر اساس درخت تصمیم

درخت تصمیم یکی از تکنیک های داده کاوی است، که ابزاری قوی و متداول برای دسته بندی و پیش بینی می باشد. روش درخت تصمیم در تقسیم بندی داده ها به گروه های مختلف به گونه ای است که هیچ داده ای حذف نمی شود. دسته بندی که توسط درخت تصمیم انجام می شوند، از روی شباهت داده های ذخیره شده در پارامترهای پیش بینی کننده، قابل انجام می باشد. خروجی درخت تصمیم می تواند کمی یا وصفی باشد که در اینجا مقداری کمی است. مهمترین هدف از دسته بندی به دست آوردن مدلی برای پیش بینی می باشد. بدین منظور از مجموعه ای به نام داده های آموزشی که مجموعه متغیرهای داماست، استفاده می کنیم. هدف از این مدل پیش بینی شرایط حرارتی در فصل رشد برج و به تبع آن ارزیابی زمان مناسب برداشت محصول برای کشاورزان می باشد [11].

در بخش های بعدی، فرایند آماده سازی داده ها و سپس نحوه ایجاد روابط داده کاوی و تولید درختان تصمیم توضیح داده شده است و در ادامه نتایج به دست آمده مورد بررسی، مقایسه و تحلیل قرار گرفته اند.

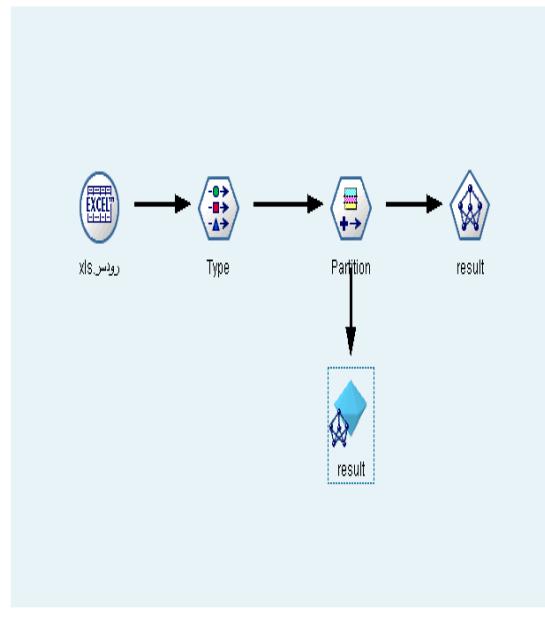
۳-۱- آماده سازی داده ها

در این فاز داده ها جهت اعمال تکنیک های داده کاوی بر آنها پالایش و آماده سازی شده اند. تبدیل فرمت، گسترش سازی و نرم افزار سازی داده ها و غیره از جمله فعالیت های این فاز می باشد.

۳-۲- پیش بینی براساس روش C5.0

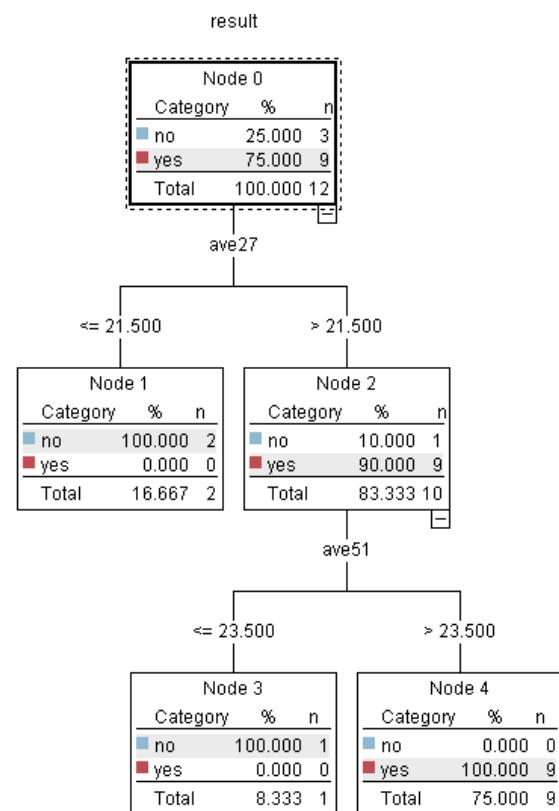
ایجاد درختان تصمیم با روش ها و الگوهای مختلفی از جمله QUEST، CHAID، C&R Tree، C5.0 ... ممکن بذیر می باشد با توجه به لزوم ارائه قوانین به صورت ساده و قابل فهم در این مقاله ۳ الگوریتم رادرنظرگرفته ایم، در این تحقیق از نرم افزار SPSS Clementine از دسته بندی برای ساخت درختان تصمیم گیری استفاده شده است.

پس از مطالعات انجام شده شاخص های تاثیرگذار بر روی زمان مناسب برداشت محصول در این منطقه شناسایی شده که با توجه به کارگیری الگوریتم های متفاوت این شاخص ها متغیر ممی باشند.



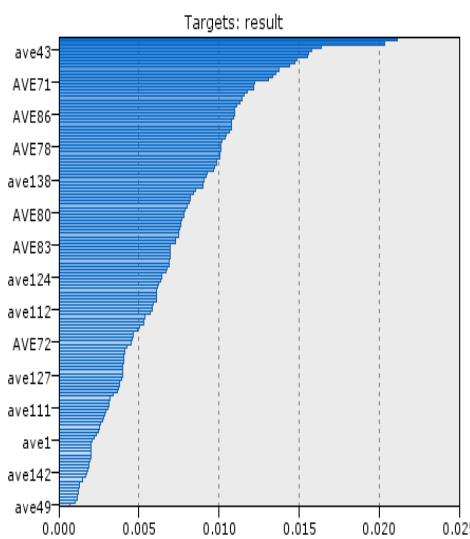
شکل(۶): مراحل داده کاوی براساس شبکه عصبی

شکل(۴): مراحل داده کاوی در مدل C&R Tree



شکل(۵): نتیجه اجرای الگوریتم C&R Tree بر روی داده های آزمون

Variable Importance



شکل(۷): نتیجه اجرای الگوریتم Neural net بر روی داده های آزمون

۳-۵- پیش بینی بر اساس شبکه عصبی

شبکه عصبی یکی از تکنیک های داده کاوی است، شبکه های عصبی مصنوعی در برخی از عملیات مانند پیش بینی و دسته بندی در مقایسه با سایر روش ها دارای مزایای نسبی بوده و معمولاً در کارهای اجرایی ترجیح داده می شوند [12,13]. به دلیل ساختار و معما ری خاص و الگوریتم های شبکه های مصنوعی، کلیه ویژگی های ارزشی در مدل این شبکه ها می باشد به صورت استاندارد تبدیل شوند. برای متغیرهای کمی پیوسته از روش های مناسب نرم افزاری داده ها به صورت زیر استفاده می شود:

$$x^* = \frac{x - \min(x)}{\text{range}(x)} \quad \text{معادله (۱)}$$

برای متغیر های کیفی و دسته ای معمولاً از متغیرهای شاخصی استفاده می شود که در این تحقیق متغیرهای ورودی از نوع کمی می باشند.

- [4] G. Grell, A. Baklanov, "Integrated modeling for forecasting weather and air quality", *Atmospheric Environment* 45, pp6845e6851, 2011, journal homepage: www.elsevier.com/locate/atmosenv

[5] Shaomin Cai, Hangjie Yu, "Analysis of different weather research and forecasting radiationschemes impact on the numerical simulation of a typical mesoscale convective weather in China", *Journal of Atmospheric and Solar-Terrestrial Physics* 80, pp 68–72, Available at www.elsevier.com/locate/jastp, 2012

[6] Jiawei, H. and Kamber, M. "Data Mining: Concepts and Techniques", Morgan Kaufmann, San Francisco, 2001 Ho T.B Dam H.C, "Introduction to Knowledge Discovery and Data Mining", Available from<URL:<http://www.jaist.ac.jp>, 2005

[7] Ho T.B Dam H.C, "Introduction to Knowledge Discovery and Data Mining", Available from<URL:<http://www.jaist.ac.jp>, 2005

[8] Lee, S.J. and Siau, K "A Review of Data Mining Techniques", *Industrial Management and Data Systems* 101/1 , MCB University Press, pp41 –46, (2001)

[9] Mark S. Roulstone, Jerome Ellepola, Jost von Hardenberg, Leonard A. Smith, "Forecasting wave height probabilities with numerical weather prediction models", *Ocean Engineering* 32, pp 1841–1863, Available at www.elsevier.com/locate/oceaneng, 2011

[10] Shearer C. "The CRISP-DM Model: The New Blueprint for Data Mining". *Journal of Data Warehousing* 1000; 1: 1 48-11

[11] Gürbüz, Feyza., Özbakir, Lale., and Yapıcı, Hüseyin, "Classification rule discovery for the aviation incidents resulted in fatality", *Knowledge-Based Systems*, Volume22 , Issue8 , Pages622-632,2009

[12]

[13] [12] Changsong Chen, Shanxu Duan, Tao Cai, Bangyin Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network", *Solar Energy* 85, pp 2856–2870, Available at www.elsevier.com/locate/solener, 2011

[14] [13] Ghosh, Nag, Biswas, Singh, Sarkar, "Weather Data Mining Using Artificial Neural Network", *Intelligent Computational System RAICS*, Trivandrum, 2011

-۶-۳- مقایسهٔ دقت پیش‌بینی مدل درخت تصمیم با مدل شبکه عصبی

با توجه به نتایج به دست آمده از خروجی درخت تصمیم با استفاده از الگوریتم های متفاوت به این نتیجه رسیدیم که الگوریتم $C\&R$ بر اساس داده های موجود روابط شاخص های تاثیرگذار را تاحدود قابل قبولی تشخیص می دهد و نیز شبکه عصبی با متodelوژی *Tree Neural net* نتیجه قابل قبولی را بر اساس درجهی تاثیرگذاری شاخص ها نشان می دهد.

۴ - نتیجه‌گیری

برنج یکی از گیاهانی است که در مناطق مرطوب استوایی و مناطق نسبتاً گرم و یا معتدل خوب رشد می‌کنند. دمای محیط کشت در طول دوره‌ی رشد همواره باید مناسب با مراحل رشد گیاه باشد، دمای مذکور بویژه در زمان گل دادن باید مناسب و کافی باشد، زیرا در این دوره گیاه به حداقل درجه‌ی حرارت نیاز دارد. متوسط دمای مورد نیاز برنج حدود ۳۳ درجه سانتی‌گراد است. چنانچه در این مرحله از رشد برنج دمای محیط پیرامون بعلی پایین آمده باشد و رطوبت هوا نیز بیش از اندازه‌ی طبیعی یعنی بالاتر از ۸۰ درصد باشد تلقیح به خوبی انجام نشده و دانه‌های تشکیل نخواهد شد. به عبارت دیگر برنج پوک شده و در نتیجه میزان تولید محصول به اندازه‌ی قابل ملاحظه‌ای پایین خواهد بود در جه حرارت داخل خوشها نیز در رشد برنج اهمیت بسزایی دارد، لکن اهمیت درجه حرارت هوا بیش از حرارت آب می‌باشد. نتایج حاصل از پیاده‌سازی نشان می‌دهد درخت تصمیم می‌تواند به عنوان ابزاری مناسب جهت مدل سازی اطلاعات به منظور پیش‌بینی میانگین دمای محیط کشت برنج و زمان مناسب برداشت برنج در استان گیلان جهت بهبود کاشت و داشت برنج برای کشاورزان مورد استفاده قرار گیرد لذا با توجه به نتایج به دست آمده از درخت تصمیم ما می‌توانیم در پایان بهار هرسال احتمال وجود سال زراعی مطلوب را در سال جاری پیش‌بینی نماییم.

مراجع

- [1] عجیج لویت لی لُ، آب ای طران (شیت ج غرفه‌ای) (دفیگی پیام و رُت ران. 1331

[2] کاراًلوی، مح موضا، سازمان مطاع شدی دکته علی م بین‌لوی دفیگی، هشوی آب نزیلی، 1331

[3] Louise Bodri and Vladimir Cermak." new method how to reconstruct climate", International Geophysics, 2007