Archive of SID
DOI: 10.22059/ijbse.2018.248601.665021

مهندسی بیوسیستم ایران، دوره ۴۹، شماره ۳، پاییز ۱۳۹۷ (ص ۵۱۱-۵۰۱)

Investigating Effective Factors on Sugarcane Production Performance to Increase the Production of Sugarcane Using Data Mining

HASSAN ZAKIDIZAJI[*1], NASIM MONJEZI[2], MOHAMMAD JAVAD SHEIKHDAVOODI[3]

1. Assistant Professor, Biosystems engineering Dept., Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
2. Ph.D. Biosystems engineering Dept., Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
3. Professor, Biosystems engineering Dept., Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
(Received: Dec. 25, 2017- Revised: Apr. 5, 2018- Accepted: Apr. 11, 2018)

ABSTRACT

Sugarcane is an important industrial-agricultural crop in country. Considering the high cultivation level of this product in Khuzestan province and the high volume of data stored in these mechanized agricultural units, it is necessary to have a tool to process stored data. The data mining technique is well equipped to provide sugarcane manufacturers with the necessary information and patterns in modeling the yield of sugarcane. One of the most practical of these algorithms is decision trees. The main objective of this research is to predict yield sugarcane and to evaluate the factors affecting it using decision trees CART and CHAID. The present work was an analytical study conducted on a database containing 13211 records. Data were obtained from farms of Amir Kabir Agro-Industry, during 2013-2016. Data analysis was performed using IBM modeler software version 14.2 by CRISP methodology. The accuracy of decision tree (CART and CHAID) on the training data and test data were 90, 81, 85 and 79 percent, respectively. The results of this research can be used for sugarcane production and cultivation industries in Khuzestan Province in order to evaluate and optimize the sugarcane production process and predict the yield of sugarcane.

Key words: Data mining, Khuzestan, Optimization, Modeling, Sugarcane.

* Corresponding Author's Email: hzakid@scu.ac.ir

www.SID.ir

Archive of SID

بررسی عوامل موثر بر میزان عملکرد تولید نیشکر با هدف افزایش تولید با استفاده از رهیافت داده‌کاوی

حسن ذکی دیزجی[1*]، نسیم منجزی[2]، محمد جواد شیخ داودی[3]

۱. استادیار، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

۲. دانش آموخته دکتری، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

۳. استاد، گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران

(تاریخ دریافت: ۴/ ۱۰/ ۱۳۹۶- تاریخ بازنگری: ۱۶/ ۱/ ۱۳۹۷- تاریخ تصویب: ۲۲/ ۱/ ۱۳۹۷)

چکیده

نیشکر یکی از مهم‌ترین محصولات کشاورزی- صنعتی کشور است. با توجه به سطح زیر کشت بالای این محصول در استان خوزستان و حجم بالای داده‌های ذخیره شده در این واحدهای مکانیزه کشاورزی، نیاز به ابزاری است تا بتوان داده‌های ذخیره شده را پردازش کرد. تکنیک داده‌کاوی به خوبی قادر است تا در زمینه مدل‌سازی عملکرد محصول نیشکر، اطلاعات و الگوهای لازم را در اختیار تولیدکنندگان نیشکر قرار دهد. یکی از کاربردی‌ترین این الگوریتم‌ها درخت‌های تصمیم است. هدف اصلی این پژوهش، پیش‌بینی عملکرد نیشکر و بررسی عوامل موثر بر آن با استفاده از درخت‌های تصمیم CART و CHAID است. این پژوهش از نوع تحلیلی بوده و پایگاه داده آن شامل ۱۳۲۱۱ رکورد می‌باشد. داده‌های مورد نیاز این تحقیق طی سال‌های زراعی ۱۳۹۲ تا ۱۳۹۶ از کشت و صنعت امیرکبیر به‌دست آمده است. تجزیه و تحلیل با استفاده از نرم افزار IBM SPSS modeler 14.2 انجام شده است. بر اساس نتایج تحقیق، دقت مدل درخت‌های تصمیم CART و CHAID برای داده‌های آموزش و آزمایش به‌ترتیب ۹۰، ۸۱، ۸۵ و ۷۹ درصد می‌باشد. نتایج حاصل از این پژوهش می‌تواند برای کشت و صنعت‌های نیشکری استان خوزستان در راستای ارزیابی و بهینه‌سازی فرآیند تولید نیشکر و پیش‌بینی عملکرد محصول نیشکر راه گشا باشد.

واژه‌های کلیدی: بهینه‌سازی، خوزستان، داده‌کاوی، مدل‌سازی، نیشکر.

* ایمیل نویسنده مسئول: hzakid@scu.ac.ir

www.SID.ir

Archive of SID

مقدمه

نیشکر از لحاظ تولید شکر، تغذیه دام و تولید الکل، کاغذ و چوب اهمیت اقتصادی زیادی دارد. اما به لحاظ سرمایه‌گذاری سنگین در احداث کارخانه، جاده‌ها، کانال‌های آب‌رسانی و تامین ماشین‌های سنگین جهت تولید به‌صورت مکانیزه و متمرکز، به مدیریت دقیق نیاز دارد (Monjezi *et al.*, 2016; Monjezi *et al.*, 2015a,b). عملکرد نهایی نیشکر در فرآیند تولید این محصول به دلیل اهمیت زیادی که در تصمیم‌گیری در زمینه کاربرد نهاده‌های مختلف در شرایط مشخص کشت و صنعت‌های نیشکری دارا است، همواره مد نظر بوده است. لذا بهینه‌سازی و مدل‌سازی عملکرد در تولید نیشکر بسیار حائز اهمیت می‌باشد. از طرف دیگر تولیدکنندگان در این بخش با حجم بسیار زیادی از داده‌های جمع‌آوری شده با خصوصیات بسیار متنوع و با روابط پیچیده در بین آن‌ها مواجه هستند که آنالیز و مدیریت آن‌ها به‌وسیله تجربه و تحلیل‌های تجربی و آماری، امری دشوار می‌باشد. حال آن‌که، داده‌کاوی[1] یک فناوری توانمند در مدیریت و سازمان‌دهی اطلاعات با حجم بالا می‌باشد. داده‌کاوی فرآیندی است که با استفاده از تکنیک‌های هوشمند، دانش را از مجموعه‌ای از داده‌ها استخراج می‌کند. دانش استخراج شده در قالب مدل‌ها، الگوها یا قواعد ارائه می‌شود. این دانش می‌تواند ملاک تصمیم‌گیری‌های آتی، عملکردهای بعدی یا تغییرات لازم در سامانه قرار گیرند. در زمینه به‌کارگیری تکنیک‌های داده‌کاوی در کشاورزی تاکنون تحقیقات بسیاری انجام شده است که در ادامه به نمونه‌هایی ازآن اشاره می‌شود.

Ramesh & Vardhan (2013) پیش‌بینی عملکرد محصولات کشاورزی را با استفاده از تکنیک‌های مختلف داده‌کاوی مانند k-means، k-Nearest Neighbor، support vector machines و شبکه‌های عصبی مصنوعی انجام دادند. آن‌ها به دنبال یافتن مدلی برای پیش‌بینی عملکرد محصولات کشاورزی بودند که دقت بالا و توانایی زیادی داشته باشد.

Jeysenthil *et al.* (2014) با استفاده از تکنیک خوشه‌بندی داده‌کاوی (k-means) یک سامانه پشتیبانی برای پایگاه داده‌های خاک مزارع نیشکر طراحی و پیشنهاد دادند.

Goktepe *et al.* (2015) با استفاده از ۱۲۰ نمونه خاک و ورش K-means و خوشه‌بندی فازی، خاک‌های منطقه آنتالیا را طبقه‌بندی کردند. نتایج این تحقیق نشان داد روش فازی خوشه‌بندی نتیجه مطلوب‌تری داشته است.

Noorzadeh *et al.* (2011) کارآیی روش خوشه‌بندی را برای خوشه‌بندی غلظت مس با استفاده از ۲۱۳ نمونه خاک در اراضی کشاورزی استان همدان بررسی کردند.

در ادامه منابعی که در داخل پرانتز آورده شده‌اند نیز توانمندی و لزوم به کارگیری تکنیک‌های داده‌کاوی در کشاورزی را تصدیق می‌نمایند (Kalpana et al., 2014a,b; Geetha, 2015; Sharma & Mehta, 2012; Yethiraj, 2012; Rajesh, 2011; Raorane & Kulkarni, 2013).

بنابراین در این تحقیق با استفاده از تکنیک‌های داده‌کاوی (درخت‌های تصمیم)[2] به مدل‌سازی عملکرد تولید نیشکر و بررسی عوامل موثر بر میزان آن با هدف افزایش تولید پرداخته شد.

مواد و روش‌ها

موقعیت جغرافیایی محل اجرای طرح

این تحقیق با استفاده از داده‌های سال‌های زراعی ۹۶ـ۱۳۹۲ در مزارع کشت و صنعت امیرکبیر یکی از واحدهای هفت‌گانه شرکت توسعه نیشکر و صنایع جانبی خوزستان اجرا شد. این کشت و صنعت در ۴۵ کیلومتری جنوب اهواز و در غرب رودخانه کارون و شرق جاده اهواز به خرمشهر و در طول جغرافیایی ۴۸ درجه و ۱۲ دقیقه تا ۴۸ درجه و ۳۰ دقیقه و عرض جغرافیایی ۳۱ درجه و ۱۵ دقیقه تا ۳۱ درجه و ۴۰ دقیقه قرار گرفته است (Monjezi & Zakidizaji, 2017). این منطقه دارای میانگین بارندگی سالیانه ۱۴۷/۱ میلی‌متر، میانگین دمای روزانه هوا ۲۵ درجه سانتی‌گراد، میانگین دمای خاک ۲۱/۲ درجه سانتی‌گراد و دارای میانگین ارتفاع ۷ متر از سطح دریا می‌باشد. مساحت این واحد ۱۴۰۰۰ هکتار و پتانسیل سطح زیر کشت آن ۱۲۰۰۰ هکتار می‌باشد و مابقی کانال، جاده، ساختمان و کارخانه می‌باشد. این واحد دارای ۴۸۰ مزرعه ۲۵/۵ هکتاری است. شکل ۱ موقعیت مزارع کشت و صنعت امیرکبیر (منطقه مورد مطالعه) را در استان خوزستان نشان می‌دهد.

جمع‌آوری داده‌ها

شناسایی صحیح منابع داده‌ای و جمع‌آوری اطلاعات خام مورد نیاز از فرآیندهای مختلف عملیات تولید نیشکر با انجام مصاحبه با مدیران و کارشناسان تولید کشاورزی، بازدید از مدیریت‌های تولید کشاورزی کشت و صنعت، گردآوری اطلاعات و داده‌های

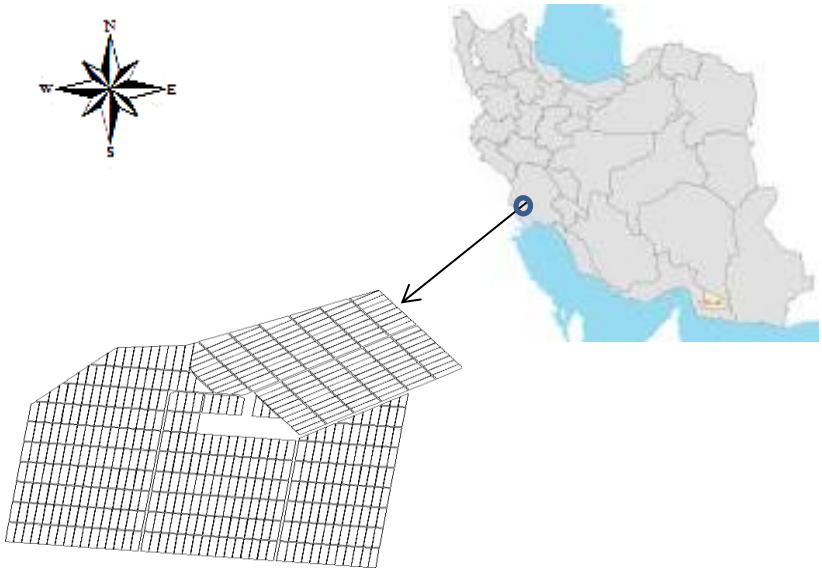
1. Data mining

2. Decision Tree

www.SID.ir

Archive of SID

خام عملیات‌های مختلف تهیه زمین، کاشت، داشت، برداشت، حمل نی و بازرویی نیشکر و بررسی و ارزیابی کلیه منابع داده‌ای موجود و مرتبط با عملکرد محصول نیشکر در مزارع کشت و صنعت امیرکبیر صورت پذیرفت.



شکل ۱- موقعیت مزارع کشت و صنعت نیشکر امیرکبیر در استان خوزستان (یافته‌های تحقیق)

درخت‌های تصمیم

درخت تصمیم یکی از روش‌های داده‌کاوی و از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد که به تولید قانون می‌پردازد. یعنی درخت تصمیم پیش‌بینی خود را در قالب یکسری قوانین توضیح می‌دهد. علاوه بر آن در درخت تصمیم‌گیری، می‌توان از داده‌های غیرعددی نیز استفاده نمود. در این تحقیق از الگوریتم CART[1] و CHAID[2] به عنوان یکی از انواع درخت‌های تصمیم رگرسیونی جهت پیش‌بینی عملکرد محصول نیشکر استفاده شد. ساخت این درختان بر سه اصل استوار است (Ayman *et al.*, 2015):

۱- مجموعه‌ای از سوالات به شکل ?x≤d که در آن x یک متغیر مستقل و d یک مقدار ثابت است و جواب هر سوال بله/خیر است.

۲- بهترین معیار شاخه‌زدن جهت انتخاب بهترین متغیر مستقل برای ایجاد شاخه جدید.

۳- ایجاد آمار خلاصه برای گره انتهایی.

معیارهای متفاوتی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، اما از آنجا که پژوهش حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل که انحراف حداقل مربعات[3] (LSD) نام دارد که این معیار به صورت زیر تعریف می‌شود (Ayman *et al.*, 2015):

$$ss\ (t) = \sum_{i=1}^{N_{tt}} (y_i(t) - \bar{y}(t))^2 \quad \text{(رابطه ۱)}$$

N_{tt} : تعداد رکوردها (داده‌ها) در گره برگ t.

$y_i(t)$: مقدار خروجی (متغیر هدف در گره برگ).

$\bar{y}(t)$: میانگین مقادیر متغیر هدف برای همه گره‌ها.

حال متغیر ورودی s زمانی بهترین متغیر برای ایجاد شاخه در گره t می‌باشد که مقدار Q(s,t) را بیشینه نماید (Ayman *et al.*, 2015).

$$Q(s,t)= ss\ (t)- ss\ (t\ R)-ss\ (t\ L) \quad \text{(رابطه ۲)}$$

که در آن ss (t R) و ss (t L) به ترتیب میزان ss (t) در شاخه سمت راست و سمت چپ گره t می‌باشد.

بنابراین در این تحقیق، از مدل‌های CART و CHAID در مدل‌سازی عملکرد و بررسی عوامل موثر بر آن در فرآیند تولید نیشکر استفاده شده است. مراحل تحقیق در شکل ۲ نمایش داده شده است.

تجزیه و تحلیل داده‌ها

در این تحقیق از نرم افزار داده‌کاوی IBM SPSS Modeler 14.2 برای مدل‌سازی و اعتبارسنجی نتایج بهره گرفته شد. داده‌ها به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم

1. Classification and Regression Trees
2. Chi-squared Automatic Interaction Detection
3. Least squares deviation

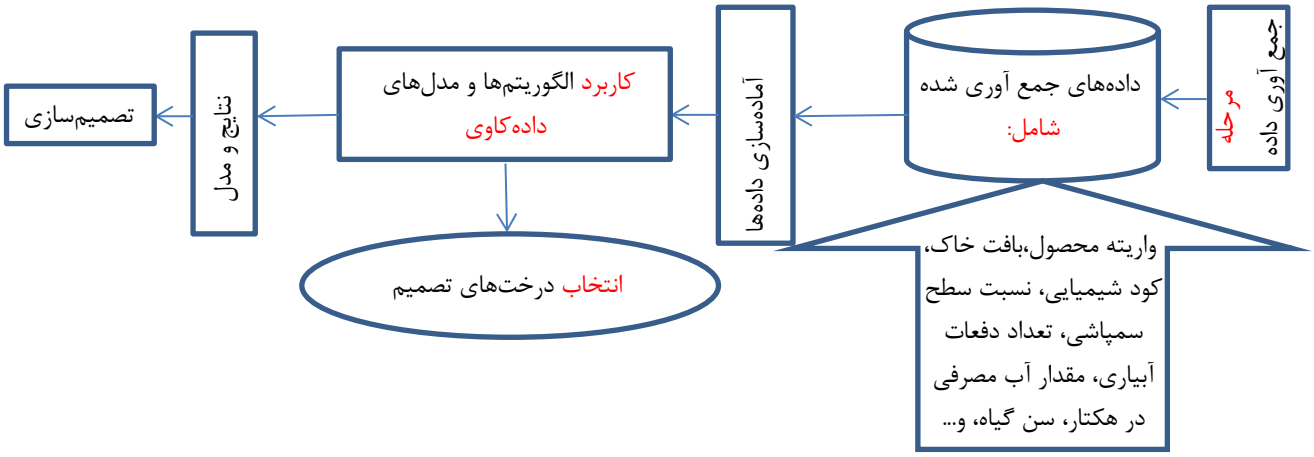
www.SID.ir

Archive of SID

شدند که سهم داده‌های آموزشی ۷۰ درصد داده‌ها و سهم داده‌های آزمایش ۳۰ درصد داده‌ها است.

ارزیابی

در این مرحله مدل خروجی از داده‌های آموزش برای داده‌های آزمون به‌کار گرفته می‌شود، به بیان دیگر در این قسمت داده‌های آزمون-که حدود ۳۰ درصد از داده‌های این پژوهش را تشکیل می‌دهند- با مدل‌های استخراجی دسته‌بندی می‌شوند و نتایج دسته‌بندی به‌وسیله این مدل‌ها با واقعیت مقایسه می‌گردد.



شکل ۲- مراحل اجرای تحقیق در مدل‌سازی عملکرد و بررسی عوامل موثر بر آن در فرآیند تولید نیشکر

نتایج و بحث

ماهیت داده‌های مورد استفاده در مدل‌های درخت تصمیم

متغیرهای ورودی مدل شامل ۱۰ متغیر است. جدول‌های ۱ و ۲ این متغیرهای پیوسته و گسسته را که شامل واریته محصول، کود شیمیایی نیتروژن، کود شیمیایی فسفر، سن گیاه، تعداد دفعات آبیاری، نسبت سطح سمپاشی محصول (نسبت مساحت سمپاشی شده به مساحت کل مزرعه)، بافت خاک، هدایت الکتریکی خاک (EC) و مصرف آب در هکتار می‌باشد، نشان داده است.

مدل درخت‌های تصمیم CART و CHAID

الگوریتم CART

در این الگوریتم ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد از داده‌ها برای آزمایش و آزمایش استفاده شدند (شکل ۳). از شاخص Gini برای ساخت درخت استفاده شد. الگوریتم درخت تصمیم به‌گونه‌ای عمل می‌کند که سعی دارد گوناگونی و یا تنوع (از نظر ویژگی هدف) را در گره‌ها به حداقل ممکن برساند. این عدم یکنواختی در گره‌ها با استفاده از معیارهای عدم خلوص[1] قابل اندازه‌گیری است که مهم‌ترین و پرکاربردترین آن شاخص جینی[2] می‌باشد (Yoneyama *et al.*, 2002).

درخت حاصل دارای ۵ سطح می‌باشد.

ساختار درخت CART ایجاد شده در شکل ۴ نمایش داده شده است. این درخت از ۲۶ گره تشکیل شده شده است. ۱۴ تا از این گره‌های، گره‌های نهایی می‌باشند. متغیر کود شیمیایی در ساختار درخت ظاهر نشده است. اولین متغیر انتخاب شده برای ایجاد شاخه در درخت، واریته نیشکر با شاخص Gini، ۱۲۴/۴۵ است. تقسیم‌بندی بعدی نیز با استفاده از واریته و رقم محصول نیشکر با شاخص Gini، ۲۶/۱۴ است که به دو واریته CP69-1062 و IRC99-02 تقسیم شده است. شاخه دیگر نیز با شاخص Gini، ۱۹/۳۸به واریته‌های CP48-103, CP57-614 و واریتهSP70-1143 تقسیم شده است. سپس واریته IRC99-02 (گره ۴) با بیشترین عملکرد (مقدار پیش‌بینی شده= ۱۱۵/۳۶ تن در هکتار) بهترین گروه با (تعداد داده‌ها=۱۱ و ۱/۸۱ درصد از کل مزارع) است. در جایی‌که واریته CP69-1062 است، تقسیم‌بندی بعدی بر اساس شاخص کود شیمیایی (نیتروژن) با Gini برابر ۱۴/۹۱ می‌باشد که به گره ۷، ≥۳۲۶/۵ و گره ۸، <۳۲۶/۵ تقسیم شده است. همچنین در واریته CP48-103 و CP57-614 مقدار آب مصرفی در هکتار با شاخص Gini ۴/۴۶ به دو گره ۹ (≥۱۳۴۱/۰۰) و گره ۱۰ (<۱۳۴۱/۰۰) تقسیم شده است. سایر اطلاعات به‌همین ترتیب از روی ساختار درخت قابل استخراج می‌باشد.

1. Impurity measure
2. Gini

www.SID.ir

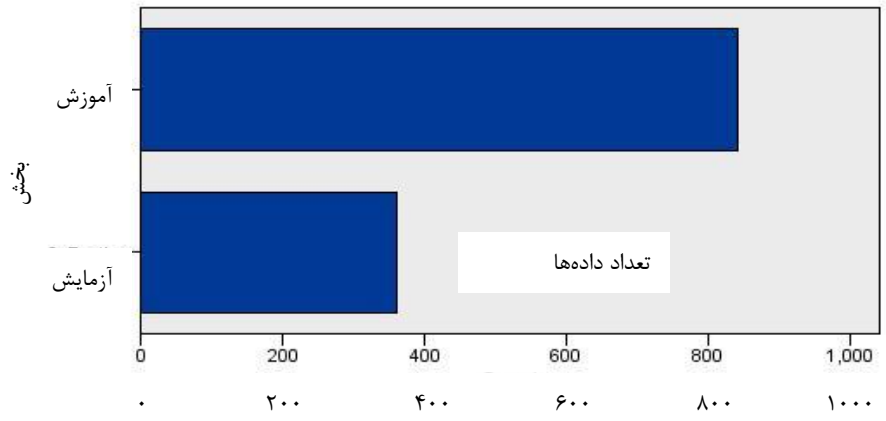
Archive of SID

جدول ۱- توصیف متغیرهای پیوسته مدل

نام متغیر	واحد	نوع متغیر	نقش متغیر	توصیف				
				کمترین مقدار	بیشترین مقدار	میانگین	انحراف معیار	تعداد رکوردهای در دسترس
کود شیمیایی (نیتروژن)	$kg\ ha^{-1}$	پیوسته	ورودی	۱۶۰	۵۵۳	۳۴۰/۷۶	۴۶/۸۸	۱۲۰۱
کود شیمیایی (فسفر)	$kg\ ha^{-1}$	پیوسته	ورودی	۰	۲۵۰	۶۴/۵۵	۹۹/۸۶	۱۲۰۱
تعداد دفعات آبیاری	-	پیوسته	ورودی	۱۲	۳۴	۲۱/۲۹	۳/۹۵	۱۲۰۱
هدایت الکتریکی خاک (EC)	$dS\ m^{-1}$	پیوسته	ورودی	۲	۱۶	۵/۷۳	۲/۱۳	۱۲۰۱
مقدار آب مصرفی در هکتار	m^3ha^{-1}	پیوسته	ورودی	۱۰۰۵	۱۹۰۰	۱۳۹۵/۵۵	۱۳۳/۵۸	۱۲۰۱
عملکرد (ساقه)	$t\ ha^{-1}$	پیوسته	هدف	۱۹	۱۳۸	۶۷/۷۴	۲۰/۹۷	۱۲۰۱

جدول ۲- توصیف متغیرهای گسسته مدل

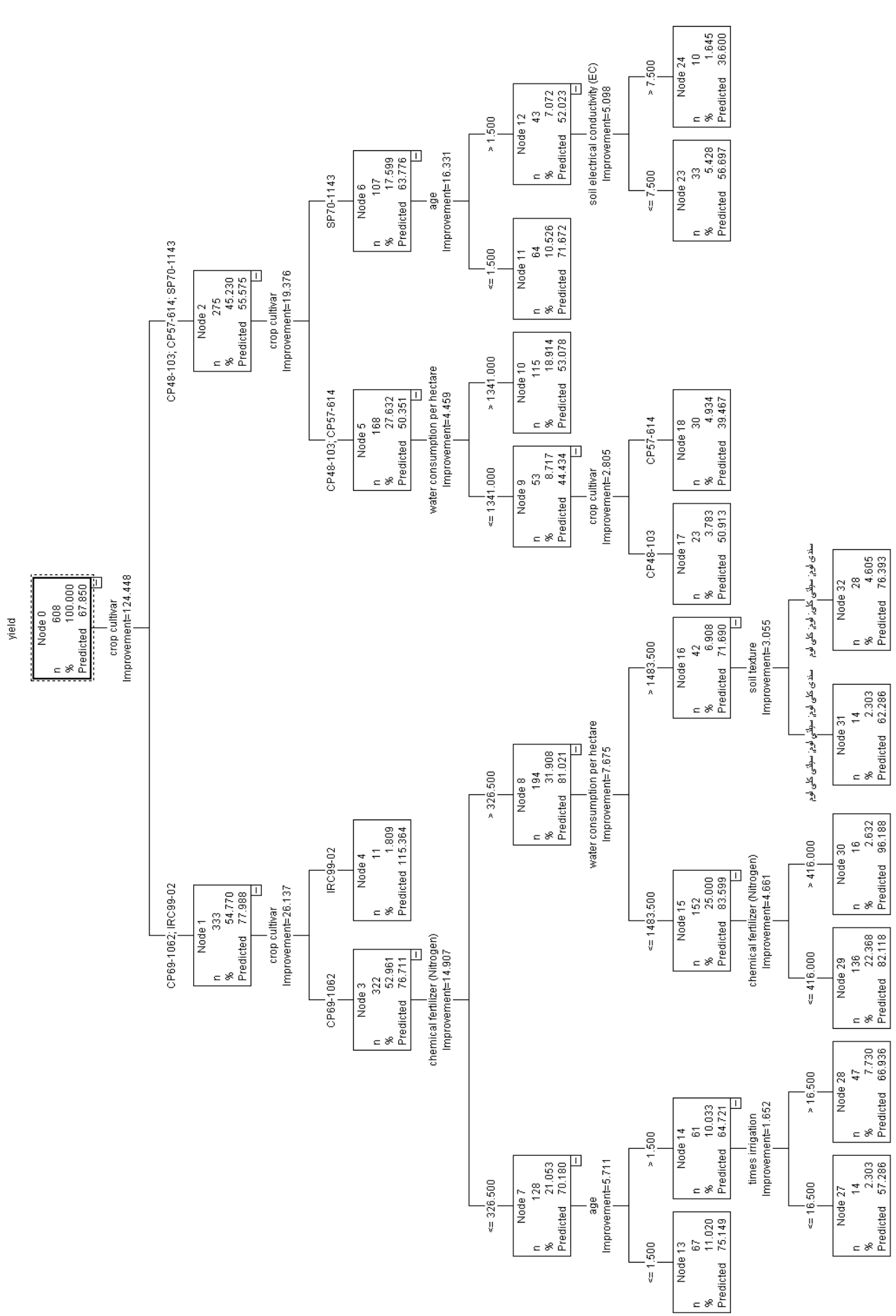
نام متغیر	نوع متغیر	نقش متغیر	توصیف	تعداد مزارع (n)	سطح (٪)
واریته محصول	گسسته	ورودی	SP70-1143	۲۳۴	۱۹/۴۸
			IRC99-02	۲۲	۱/۸۳
			CP69-1062	۶۱۹	۵۱/۵۴
			CP57-614	۱۵۳	۱۲/۷۴
			CP48-103	۱۷۳	۱۴/۴۰
بافت خاک	گسسته	ورودی	لومی شنی	۷۶	۶/۳۳
			لومی رسی شنی	۳۶	۳/۰۰
			لومی سیلتی	۲۴	۲/۰۰
			رسی سیلتی	۷۴	۶/۱۶
			لومی رسی سیلتی	۲۵۵	۲۱/۲۳
			لوم	۴۱۴	۳۴/۴۷
			لومی رسی	۳۲۲	۲۶/۸۱
سن گیاه	گسسته	ورودی	کشت اول	۳۵۹	۲۹/۸۹
			راتون اول	۳۵۸	۲۹/۸۱
			راتون دوم	۲۹۳	۲۴/۴۰
			راتون سوم	۱۹۱	۱۵/۹۰
نسبت سمپاشی	گسسته	ورودی	۱	۲۷۱	۲۲/۵۶
			۲	۳۷۱	۳۰/۸۹
			۳	۲۹۶	۲۴/۶۵
			۴	۲۶۳	۲۱/۹۰



شکل ۳- نسبت داده‌های آموزش و آزمایش

www.SID.ir

Archive of SID



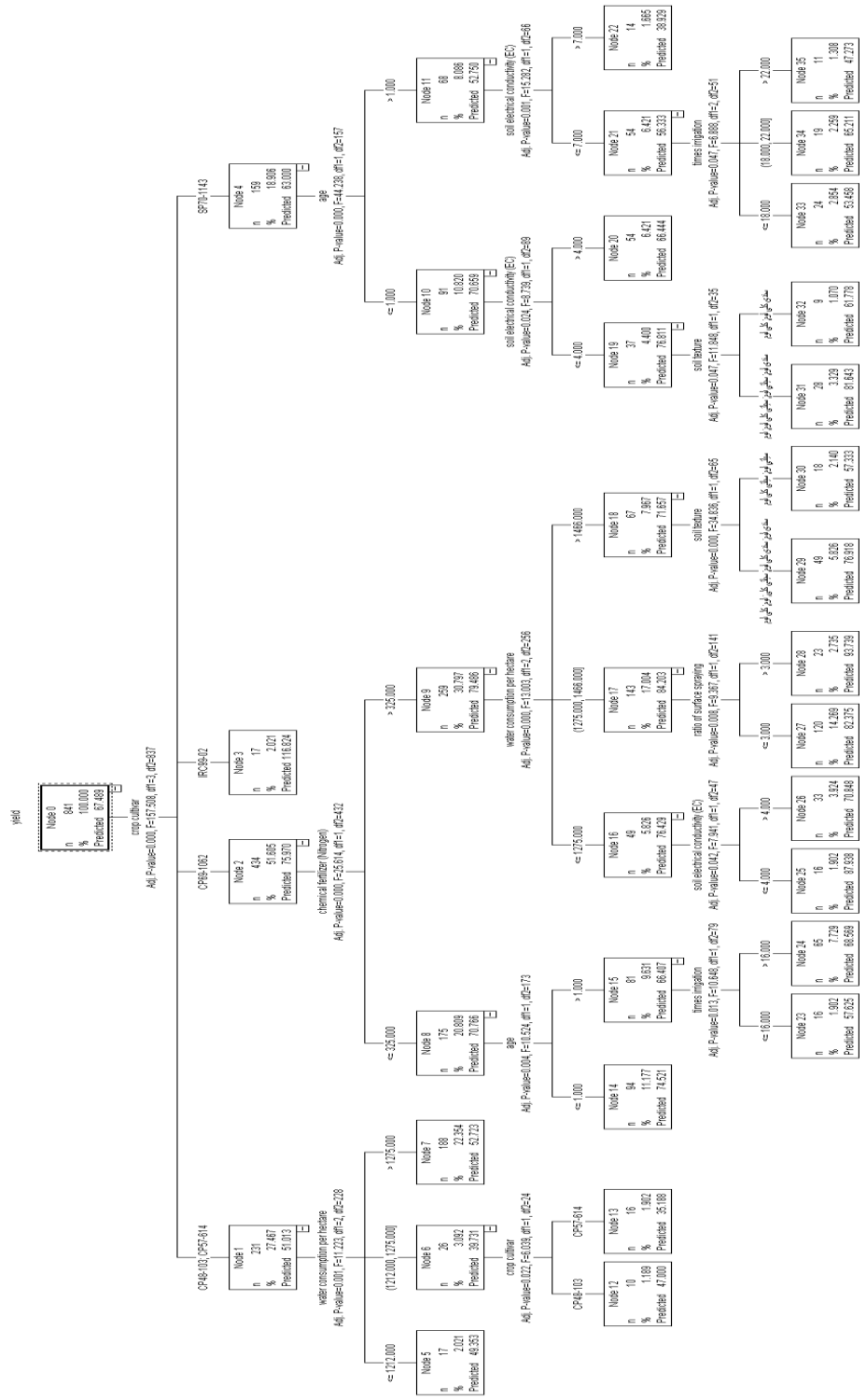
شکل ۴- درخت تصمیم CART

www.SID.ir

Archive of SID

الگوریتم CHAID

در این الگوریتم نیز، ۷۰ درصد از داده‌ها برای آموزش و ۳۰ درصد از داده‌ها برای آزمایش و آزمایش استفاده شدند (شکل ۳) و ساختار درخت حاصل در ۴ سطح به‌دست آمد. مدل درخت CHAID در شکل ۵ نمایش داده شده است. در این ساختار، ۳۵ گره شامل ۲۱ گره نهایی وجود دارد. اولین متغیر ایجاد شاخه در درخت، متغیر واریته محصول می‌باشد که درخت را به ۴ گره (CP57-614 &CP48-103، CP69-1062، IRC99-02، SP70-1143) تقسیم کرده است. دومین متغیر تقسیم کننده درخت، مقدار آب مصرفی در هکتار برای واریته‌هایCP48-103, CP57-614 و کود شیمیایی نیتروژن برای واریته CP69-1062 و سن گیاه برای واریته SP70-1143 می‌باشد. سایر اطلاعات و نتایج به‌همین صورت قابل استخراج می‌باشند.



www.SID.ir

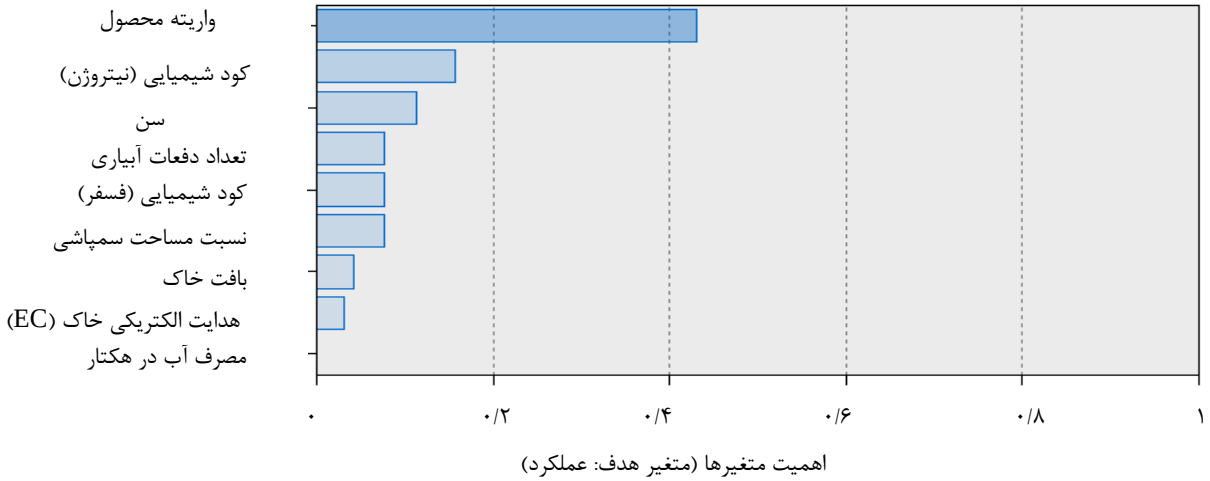
Archive of SID

شکل ۵- درخت تصمیم CHAID

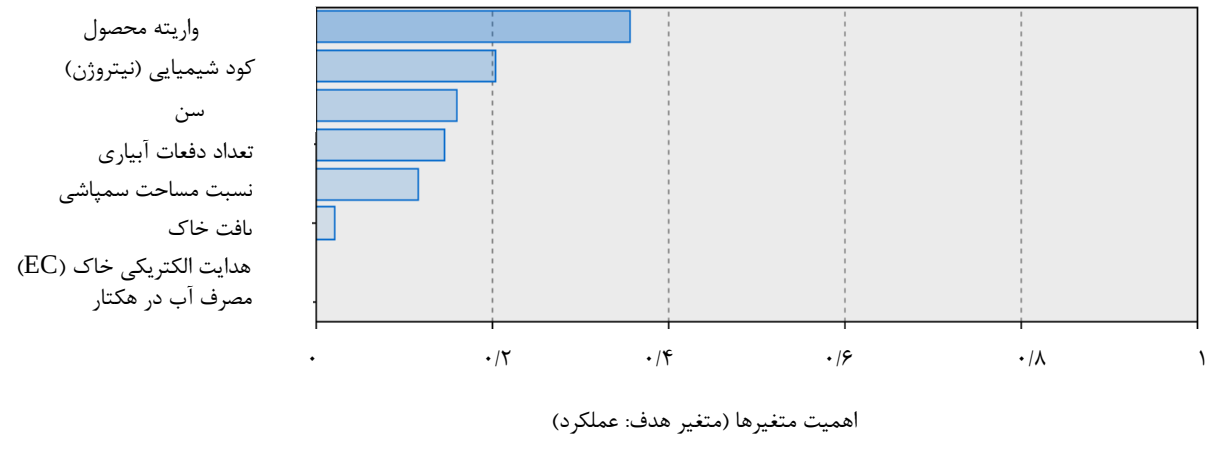
اهمیت متغیرها

متغیرهای واریته محصول، کود شیمیایی (نیتروژن)، سن، تعداد دفعات آبیاری، کود شیمیایی (فسفر)، نسبت سطح سمپاشی، بافت خاک و هدایت الکتریکی خاک (EC) مهم‌ترین و تاثیرگذارترین متغیرها در مدل درخت تصمیم CART می‌باشند (شکل ۶). در مدل درخت تصمیم CHAID، واریته محصول، کود شیمیایی نیتروژن، سن، تعداد دفعات آبیاری، نسبت سطح سمپاشی و بافت خاک از اهمیت بیشتری برخوردار می‌باشند (شکل ۷). بنابراین در دو مدل CART و CHAID متغیر واریته محصول مهم‌ترین متغیر تاثیرگذار بر پیش‌بینی عملکرد محصول نیشکر است. این نتایج با نتایج تحقیقاتی از تولی میل[1] استرالیا مطابقت دارد (Ellis *et al.*, 2001; Lawes *et al.*, 2002). همچنین، در مدل‌های CART و CHAID متغیرهای کود شیمیایی (نیتروژن)، سن گیاه و تعداد دفعات آبیاری گیاه از مهم‌ترین عوامل تاثیرگذار بر عملکرد محصول نیشکر می‌باشند.

1. Tully Mill



شکل۶- اهمیت متغیرها در الگوریتم CART



شکل ۷- اهمیت متغیرها در الگوریتم CHAID

دقت مدل‌سازی

با استفاده از رابطه همبستگی خطی بین نتایج پیش‌بینی شده توسط مدل و نتایج واقعی می‌توان به این نکته پی برد که مدل حاصل تا چه اندازه در پیش‌بینی عملکرد نیشکر موفق بوده است. با توجه به جدول ۳ دقت مدل CART در قسمت آموزش و آزمایش به ترتیب برابر ۹۰ و ۸۱ درصد بوده است. همچنین دقت مدل CHAID در جدول ۴ نشان داده شده است. دقت این مدل در بخش آموزش و آزمایش به ترتیب برابر ۸۵ و ۷۹

www.SID.ir

Archive of SID

درصد می‌باشد. بنابراین به‌طورکلی دقت پیش‌بینی در این پژوهش بین ۸۰ تا ۹۰ درصد است.

جدول ۳- نتایج مدل CART

قسمت	آموزش	آزمایش
خطای مینیمم	-۵۸/۱۲	-۴۳/۳۶
خطای ماکزیمم	۵۵/۹۲	۴۱/۳۳
میانگین خطا	-۰/۲۳	۰/۷۱
میانگین مطلق خطا	۱۱/۹۷	۱۲/۹۵
انحراف معیار	۱۵/۳۶	۱۵/۹۵
همبستگی خطی	۰/۹۰	۰/۸۱
نسبت داده‌ها	۸۴۱	۳۶۰

جدول ۴- نتایج مدل CHAID

قسمت	آموزش	آزمایش
خطای مینیمم	-۴۶/۸۵	-۴۸/۷۴
خطای ماکزیمم	۵۶/۲۸	۴۹/۲۲
میانگین خطا	۰/۰	۰/۶۶
میانگین مطلق خطا	۱۱/۶۹	۱۲/۹۷
انحراف معیار	۱۴/۸۹	۱۶/۱۷
همبستگی خطی	۰/۸۵	۰/۷۹
نسبت داده‌ها	۸۴۱	۳۶۰

نتیجه‌گیری

در این تحقیق داده‌های بخش کشاورزی کشت و صنعت امیرکبیر مورد استفاده قرار گرفت. پایگاه داده تهیه شده به دو بخش آموزش و آزمایش تقسیم گردید. مدل‌های CART و CHAID ساخته شد. در نهایت مدل‌های تولید شده با استفاده از مجموعه داده‌های آزمایش، مورد ارزیابی قرار گرفت و به دقت قابل قبولی برای تخمین عملکرد محصول رسید. دقت مدل CART در قسمت آموزش و آزمایش به ترتیب برابر ۹۰ و ۸۱ درصد بود. همچنین دقت مدل CHAID در بخش آموزش و آزمایش به ترتیب برابر ۸۵ و ۷۹ درصد می‌باشد. بنابراین به طور کلی دقت پیش‌بینی در این پژوهش بین ۸۰ تا ۹۰ درصد است. همچنین به خوبی مشاهده می‌شود که مدل درخت تصمیم تولید شده از دقت بسیار خوبی در تخمین مقادیر عملکرد خروجی مزارع برخوردار است. لذا زمانی که به حجم بالایی از اطلاعات دسترسی باشد و به‌صورت مناسب پیش‌پردازش‌های لازم انجام شده باشد، تکنیک‌های داده‌کاوی قادر خواهند بود که به پیش‌بینی بهتر از مسائل مختلف در مدیریت منابع و نهادها و همچنین تولید دانش سودمند در شناخت وضعیت پیچیده موجود بپردازند. نهایتاً نتایج این مطالعه نشان می‌دهد بهره‌گیری از روش‌های پیش‌بینی، با ارائه تصویری دقیق‌تر از وضعیت عملکرد محصول در کشت و صنعت نیشکر، امکان تخصیص مناسب منابع و افزایش بهره‌وری از نهاده‌ها را فراهم می‌نماید.

سپاسگزاری

این مقاله مستخرج از طرح پژوهشی به شماره ۱۲۷۳ از محل اعتبارات پژوهانه واحد پژوهشی دانشگاه شهید چمران اهواز می‌باشد. بنابراین نویسندگان از معاونت پژوهشی دانشگاه شهید چمران اهواز بابت تامین هزینه‌های این پژوهش سپاسگزاری می‌نمایند.

REFERENCES

Ayman E. K., Kadry, M. & Walid, G. (2015). Proposed framework for implementing data mining techniques to enhance decisions in agriculture sector, *Procedia Computer Science*, 65, 633

Ellis, R. N., Basford, K. E., Cooper, M., Leslie, J. K. & Byth, T. L. D. E. (2001). A methodology for analysis of sugarcane productivity trends. I. Analysis across districts. *Australian Journal of Agricultural Research*, 52, 1001–1009.

Geetha, M. C. S. (2015). A survey on data mining techniques in agriculture. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2), 887-892.

Goktepe, A. B., Altun, S. & Sezar, A. (2015). Soil clustering by fuzzy C-Means algorithm. *Advances in Engineering Software*, 36, 691-698.

Jeysenthil.K. M. S., Manikandan.T & Murali, E. (2014). Third generation agricultural support system development using data mining. *International Journal of Innovative Research in Science, Engineering and Technology*, 3 (3), 9923- 9930.

Kalpana, R., Shanthi, N. & Arumugam, S. (2014a). Data mining – An evolutionary view of agriculture. *International Journal of Application or Innovation in Engineering & Management*, 3 (3), 102- 105.

Kalpana, R., Shanthi, N. & Arumugam, S. (2014b). A survey on data mining techniques in agriculture. *International Journal of Advances in Computer Science and Technology*, 3(8), 426-431.

Lawes, R. A., Lawn, R. J., Wegener, M. K. & Basford, K. E. (2002). Understanding and managing the late time of ratooning effect on cane yield. *Proceedings of the Australian Society of Sugar Cane Technology*, 24, 160–165.

Monjezi, N., Sheikhdavoodi1, M. J., Zakidizaji1, H., Marzban, A. & Shomeili, M. (2016). Operations scheduling of sugarcane production using fuzzy GERT method (part II: preserve operations, harvesting and rationing). *Agricultural Engineering International: CIGR Journal*, 18 (3), 343- 349.

Monjezi, N., Sheikhdavoodi, M. J., Zakidizaji, H.

www.SID.ir

Archive of SID

Marzban, A. & Shomeili, M. (2015a). Operations scheduling of sugarcane production using classical GERT method (part I: land preparation, planting and preserve operations). *Journal of Agricultural Studies*, 3 (2), 85-96.

Monjezi, N., Sheikhdavoodi, M. J., Zakidizaji, H., Marzban, A. & Shomeili, M. (2015b). Operations scheduling of sugarcane production using classical GERT method (part II: preserve operations, harvesting and ratooning). *Journal of Agricultural Studies*, 3 (2), 85-96.

Monjezi, N. & Zakidizaji, H. (2017). Fuzzy approach to optimize overhaul time of sugarcane harvester using GERT network method. *Iranian Journal of Biosystem Engineering*, 48(1), 83-91. (In Farsi)

Noorzadeh, M., Khavazi, K., Malakooti, M. & Hashemi, S. (2011).Evaluation of the effectiveness of C-means and GK methods for fuzzy clustering of copper concentration in agricultural lands (Case study: Hamedan Province). *Journal of Agricultural Engineering*, 33 (1), 61-70. (In Farsi)

Rajesh, D. (2011). Application of spatial data mining for agriculture. I*nternational Journal of Computer Applications*, 15(2), 7-9.

Ramesh, D. & Vishnu Vardhan, B. (2013). Data mining techniques and applications to agricultural yield data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(9), 3477-3480.

Raorane, A.A. & Kulkarni R.V. (2013). Review- Role of data mining in agriculture. *International Journal of Computer Science and Information Technologies*, 4 (2), 270 – 272.

Sharma, L. & Mehta, N. (2012). Data mining techniques: A tool for knowledge management system in agriculture. *International Journal of Scientific and Technology Research*, 1(5), 67-73.

Yethiraj, N. G. (2012). Applying Data Mining Techniques in the field of agriculture and alliedsciences. *International Journalof Business Intelligents*, 1(2), 72-76.

Yoneyama, Y., Suzuki, S., Sawa, R., Yoneyama, K., Power, G. G. & Araki, T. 2002. Increased plasma adenosine concentrations and the severity of preeclampsia. *Obstet Gynecol*, 100(6), 1266-1270.

www.SID.ir