نشریه علمی تخصصی شباک
سال پنجم، شماره ۴۷، آبان ۱۳۹۸
ISSN: 2476-3667

ارزیابی عوامل موثر بر کاهش تولید گاز نسبت به برنامه تکلیفی سکوی فاز ۱۲ میدان پارس جنوبی با رویکرد داده‌کاوی

شکیبا خادم القرانی[۱]، سعید یوسفیان[۲]

تاریخ دریافت: ۹۸/۰۷/۲۳
تاریخ پذیرش: ۹۸/۰۸/۳۱
کد مقاله: ۷۲۵۴۹

چکـیده

نقش و جایگاه تأثیرگذار مدیریت داده در تمام عرصه‌ها برکسی پوشیده نیست. اهمیت این موضوع زمانی خود را بیشتر نمایان می‌کند که در صنعت بزرگی مانند نفت و گاز نقش آفرینی نماید. میزان استحصال گاز از سکوی فاز ۱۲ پارس جنوبی طی سال‌های ۱۳۹۲ لغایت ۱۳۹۶ نشان می‌دهد که ۸۴ درصد از برنامه مصوب تولید محقق گردیده است. عدم تحقق کامل برنامه مصوب تولید، به طور قطع تبعات منفی در زمینه تأمین انرژی مورد نیاز کشور در مصارف خانگی و صنعتی را بدنبال خواهد داشت. در این پژوهش با توجه به حجم بالای داده در زمینه عوامل موثر بر کاهش تولید گاز از رویکرد داده کاوی و سه الگوریتم درخت تصمیم، قواعد انجمنی و شبکه‌عصبی ارزیابی وضعیت سکو صورت پذیرفته است. نتایج این پژوهش نشان می‌دهد، محدودیت‌های دریافت پایین دست (پالایشگاه‌ها) به عنوان یک عامل مهم برون سازمانی، همچنین پتانسیل پایین تولید برخی چاه‌ها و نیز خرابی برخی تجهیزات اعم از تجهیزات برق و ابزاردقیق، مکانیک و ... به عنوان عوامل مهم درون سازمانی در بروز کاهش تولید گاز و در نتیجه عدم تحقق کامل برنامه‌های مصوب برداشت از این سکو دخیل بوده‌اند. از این‌رو با استفاده از نتایج حاصل از این پژوهش، مدیران و کارشناسان مجموعه قادر خواهند بود، در آینده نسبت به اتخاذ تصمیم‌ها و سیاست‌های صحیح در جهت اولویت‌بندی و پیشگیری از وقوع عوامل کاهش تولید و در نتیجه پوشش کامل برنامه مصوب تولید، اقدامات موثر را بعمل آورند.

واژگـان کلـیدی: میدان پارس‌جنوبی، داده‌کاوی، درخت تصمیم‌گیری، قواعد انجمنی، شبکه عصبی

۱- عضو هیئت علمی و استاد دانشگاه شیخ بهایی، *shakiba_kh@shbu.ac.ir*
۲- کارشناس ارشد مهندسی صنایع

www.SID.ir

۱- مقدمه

کاوش داده‌های سازمانی با استفاده از تکنیک‌های داده‌کاوی، و آشکارسازی روابط پنهان میان داده‌های سازمان به ما کمک خواهد کرد تا مسیری برای رشد و تعالی سازمان که استراتژی پیشبرد نامیده می‌شود، مشخص سازیم. بطور کلی استفاده وسیع از سیستم‌های مکانیزه و اتوماسیون های صنعتی و اداری، سازمان‌ها را با حجم زیادی از اطلاعات روبرو می‌نماید. رشد روز افزون در داده‌های ذخیره شده نیاز مبرم به وجود تکنولوژی‌های جدید و ابزارهای خودکاری را ایجاد کرده که بصورت هوشمند به انسان یاری رسانند تا این حجم عظیم داده را به اطلاعات و دانش تبدیل نماید (تایوغ و همکاران، ۱۳۹۳).

تأکید بر نظارت و حصول اطمینان از عملیات تولید در طول بهره‌برداری از صنایع نفت و گاز موجب گردیده که در طول دو دهۀ گذشته، وجود فناوری اطلاعات در مدیریت "داده‌های بزرگ" در این صنعت نقش برجسته‌تری ایفا نماید. حجم انبوهی از داده‌های حاصل از نظارت بر تولید چاه‌های نفت و گاز و فرایندهای عملیاتی، مدیریت و استفاده از این اطلاعات را برای این صنعت، سهام‌داران و سرمایه‌گذران آن از اهمیت برخوردار نموده است. ادغام و تحلیل داده‌ها برای کارشناسان صنعت نفت و گاز بمنظور ایجاد دیدگاهی جهت حرکت به سمت تصمیم‌گیری‌های مبتنی بر داده‌ها ضروری می‌باشد (ابوسید، ۲۰۱۲).

با توجه به آنکه داده‌کاوی فرآیندی است که از انواع تکنیک‌های مدل‌سازی و آنالیز داده برای کشف الگوها و ارتباطات در داده‌ها برای انجام پیش‌بینی‌های دقیق بهره می‌برد، این امر صنعت نفت را در موضوعات متنوعی یاری می‌دهد. بگونه‌ای که استفاده از روش‌های جدید آنالیز داده؛ مانند داده‌کاوی برای پایگاه‌های داده این صنعت بسیار حیاتی بوده و بدین جهت امروزه دانش داده‌کاوی در تمامی موضوعات و برنامه‌ریزی‌های خرد و کلان صنعت نفت نقشی موثر ایفا می‌نماید.

طرح توسعه فاز ۱۲، بزرگترین طرح توسعۀ میدان پارس جنوبی است که بمنظور برداشت روزانه ۳ میلیارد فوت مکعب گاز از مخزن پارس جنوبی طراحی و در دست اجرا میباشد. اهداف اجرای این طرح شامل:

تولید روزانه ۷۵ میلیون متر مکعب گاز شیرین و غنی جهت تزریق به خط لوله ششم سراسری گاز کشور

تولید روزانه ۱۱۰هزار بشکه میعانات گازی

تولید روزانه ۷۵۰ تن گوگرد دانه بندی شده

فاز ۱۲ میدان گازی پارس جنوبی با مساحتی در حدود ۲۰۵ کیلومتر مربع، روی خط مرزی مشترک ایران و قطر در عمق حدود ۳۰۰۰ متری کف خلیج فارس و در ۱۰۵ کیلومتری سواحل ایران قرار گرفته است، این فاز با برخورداری از ذخایر درجای معادل ۶۰۰ میلیارد مترمکعب گاز حدود ۵ درصد از ذخایر میدان گازی پارس جنوبی را به خود اختصاص داده است (وبگاه شرکت نفت و گاز پارس[1]).

۲- ابزار داده‌کاوی مورد استفاده در این پژوهش

با توجه به اینکه بانک اطلاعاتی SQL Server به‌عنوان یکی از قویترین سیستمهای مدیریت بانک اطلاعات رابطه‌ای، مخصوصاً در حجم بالای داده شناخته شده است، استفاده از SQL Server در پروژه‌های داده‌کاوی به‌منظور ذخیره‌سازی و مدیریت حجم بالای داده‌ها و استفاده از امکانات Business Intelligence Development Studio که قابلیت‌های برنامه نویسی را در SQL Server فراهم آورده است، یکی از بهترین انتخاب‌ها می‌باشد (شهرابی، ۱۳۹۲، ۱۶).

از مزایای استفاده از نرم افزار Microsoft SQL Server بمنظور اجرای فرایند داده‌کاوی آن است که؛ این برنامه نسبت به ارائه یک محیط یکپارچه برای ایجاد مدل‌های داده‌کاوی و کار با آنها می‌پردازد. این محیط شامل SQL Server Development Studio است که با بهره‌مندی از الگوریتم‌های داده‌کاوی و ابزارهای Query موجب تسهیل در روند ایجاد یک راهکار جامع برای بسیاری از پروژه‌ها و SQL Server Management Studio می‌شود که ابزارهایی برای جستجوی مدل‌ها و مدیریت اشیاء داده‌کاوی می‌باشند. مهمترین ویژگی موتور SQL Server، یکپارچگی آن با Microsoft.Net Framwork می باشد که امکان ساخت رویه های ذخیره شده، توابع تعریف شده کاربر و تریگرها را با هریک از زبان های برنامه نویسی Net. میسر می‌سازد. استفاده از موتور بانک اطلاعاتی SQL Server امکان مدیریت همزمان و یکپارچه داده‌های گسسته و پیوسته را پدید می‌آورد که دغدغۀ همیشگی در بکارگیری الگوریتم‌های داده‌کاوی بوده است (شهرابی، ۱۳۹۲، ۱۵).

در پژوهش حاضر تمام داده‌های تولید سکوها و نیز عوامل کاهش تولید وابسته به آنها در محیط SQL Server تجمیع و ذخیره‌سازی گردیده‌اند. استفاده از قابلیت‌های موجود در محیط Analysis Services این امکان را به پژوهشگر و کارشناسان می‌دهد تا با برقراری یک ارتباط با محیط پایگاه داده نسبت به ساخت کوئری‌های مورد نظر اقدام و با اجرای الگوریتم‌های داده‌کاوی در محیط Analysis Services نسبت به بررسی و تحلیل آنها اقدام نماید.

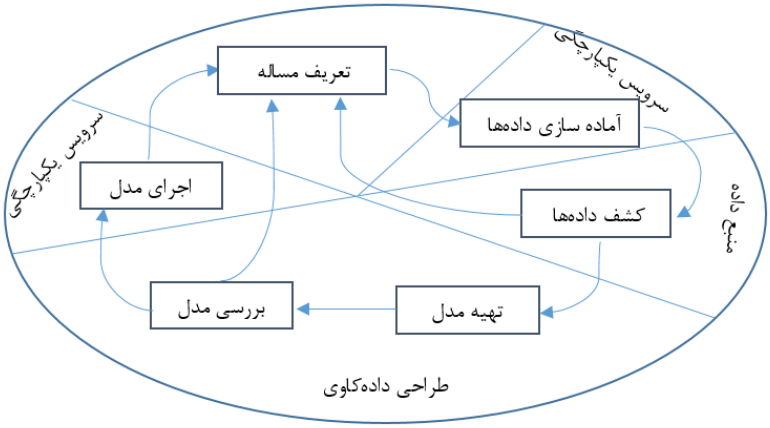
[1] www.pogc.ir

www.SID.ir

بنابراین ابزار مورد استفاده در این پژوهش، SQL Server Data Tools (SSDT) می‌باشد که در محیط Microsoft SQL Server Analysis Services (SSAS) قابلیت اجرایی پیدا می‌کند.

مراحل مختلف ایجاد پروژه در محیط SQL Server Data Tools، شامل: ایجاد منبع داده[1]، ایجاد پرس و جو[2] و در نهایت ایجاد ساختار داده‌کاوی[3] جهت کار بر روی الگوریتم‌های آن می‌باشد (مک لنن، ۲۰۰۸، ۱۸۷).

شکل ۱، رابطه میان هر یک از مراحل فوق را در این فرایند و تکنولوژی‌های موجود در Microsoft SQL Server توصیف می‌نماید.



شکل ۱- فرایند ایجاد یک مدل داده‌کاوی در *Microsoft SQL Server*

۳- پایگاه داده

پایگاه داده ایجاد شده در محیط SQL Server دارای دو جدول تحت عناوین جدول تولید و نیز جدول عوامل کاهش تولید نسبت به برنامه مصوب سکوها می‌باشد که بصورت رابطه‌ای و از طریق یک فیلد مشترک با یکدیگر ارتباط یافته‌اند. جدول تولید شامل ۶ فیلد (ستون) تحت عناوین: نام سکو، تاریخ، برنامه مصوب، میزان تولید گاز، میزان تولید میعانات‌گازی و یک فیلد یکتا جهت ارتباط با جدول عوامل کاهش تولید می‌باشد که مقادیر تولید واقعی و برنامهٔ تولید سکوها را بصورت روزانه در خود ذخیره نموده است. نوع داده‌های ذخیره شده در فیلدهای نام سکو و تاریخ از نوع متنی و در مابقی فیلدها از نوع عددی در نظر گرفته شده است.

جدول عوامل کاهش تولید نیز شامل ۷ فیلد (ستون) تحت عناوین: تاریخ، میزان کاهش تولید، ماهیت کاهش تولید، حوزه‌های مسئولیتی، حوزه‌های رخداد و علت بروز کاهش تولید می‌باشد که توسط فیلد یکتای مشترک به جدول تولید ارتباط داده شده است. نوع داده‌های ذخیره شده در فیلد مقدار کاهش تولید از نوع عددی و در سایر فیلدها از نوع متنی در نظر گرفته شده است.در واقع بمنظور انجام مدل‌سازی و اجرای فرایند داده‌کاوی، نسبت به ساخت کوئری‌های مورد نیاز در SQL Server Database Engine و برقراری ارتباط با محیط Analysis Services از این پایگاه داده، اقدام گردیده است. همچنین تعداد رکوردهای ثبت گردیده متعلق به سکوی SPD12 در جدول عوامل کاهش تولید برابر با ۲۲۵۸ رکورد می‌باشد که برهمین اساس و با توجه به لزوم تفکیک کل داده‌ها به ۲ بخش مجموعه داده‌های آموزشی[4] که برای ساخت مدل و داده‌های آزمایشی[5] که برای ارزیابی مدل ساخته شده بکار گرفته می‌شود و نیز با احتساب اختصاص ۳۰ درصد از کل داده‌ها به مجموعه آزمایشی، تفکیک تعداد رکوردها در جدول عوامل کاهش تولید به شرح جدول ۱ خواهد بود.

جدول ۱- مجموعه داده‌های آزمایشی و آموزشی

عناوین	درصد متعلقه	تعداد
مجموعه داده های آزمایشی	٪۳۰	۶۷۷
مجموعه داده های آموزشی	٪۷۰	۱۵۸۱
کل داده ها	٪۱۰۰	۲۲۵۸

1 Data Source
2 Data Source Views
3 Mining Structure
4 Training Data Set
5 Testing Data Set

www.SID.ir

۴-الگوریتم‌های مورد استفاده در پژوهش

در این پژوهش بمنظور اجرای پروژه داده‌کاوی از سه الگوریتم مطرح در این حوزه تحت عناوین؛ الگوریتم درخت تصمیم[1]، قواعد انجمنی[2] و شبکه های عصبی[3] به شرح زیر استفاده و سپس به تفسیر نتایج حاصل از آن پرداخته شده است.

۴-۱- درخت تصمیم

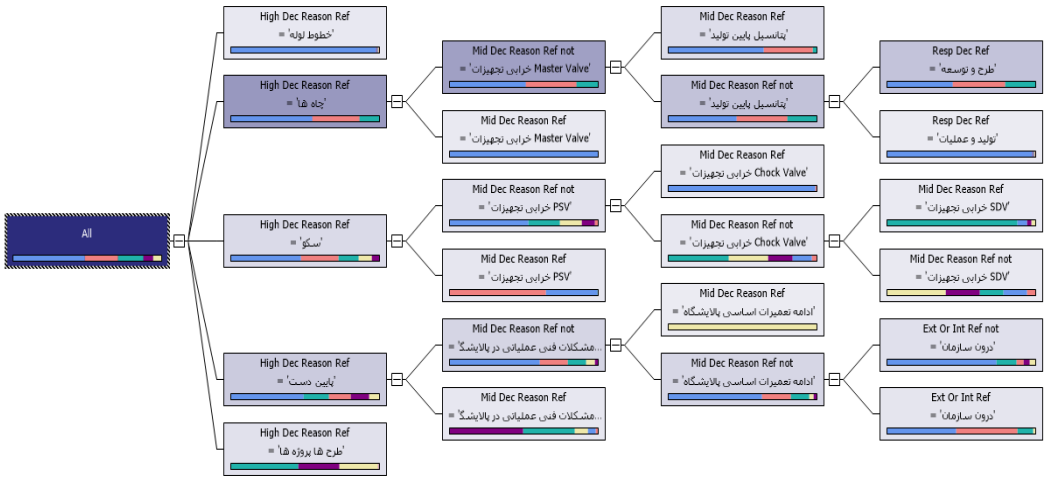
ساختار درخت تصمیم یک ساختار درختی شبیه به فلوچارت است. بالاترین گره در درخت، گره ریشه است و گره‌های برگ، دسته‌ها یا توزیع دسته‌ها را نشان می‌دهد. درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته بندی و پیش بینی می‌باشد و بر خلاف شبکه‌های عصبی به تولید قاعده می‌پردازد. در ساختار درخت تصمیم، پیش بینی بدست آمده از درخت در قالب یکسری قواعد توضیح داده می‌شود. در حالیکه در شبکه‌های عصبی تنها نتیجه پیش بینی بیان می‌گردد و چگونگی بدست آمدن آنها در خود شبکه پنهان می‌ماند (غضنفری، ۱۳۹۵).

الگوریتم درخت تصمیم در سکوی فاز ۱۲ (SPD12) با هدف طبقه‌بندی داده‌های موجود در پایگاه داده، بر اساس۴ ستون (فیلد)، ماهیت کاهش تولید، حوزهٔ مسئولیت، حوزهٔ رخداد و نیز علت اصلی بروز عوامل کاهش تولید، به عنوان ورودی مدل، صورت پذیرفته است. بنابراین ابتدا در نرم افزار SQL Server Data Tools (SSDT) متغیرهای ورودی و نیز متغیر پیش‌بینی مطابق جدول ۲، مشخص گردیده است.

جدول ۲- متغیرهای ورودی و خروجی مدل

نام متغیر (فیلدها)	نوع متغیر	نوع داده های ذخیره شده
ماهیت کاهش تولید	ورودی	متنی
حوزهٔ مسئولیت	ورودی	متنی
حوزهٔ رخداد	ورودی	متنی
علت رخداد	ورودی	متنی
میزان کاهش تولید	پیش‌بینی	عددی

بر اساس مدل‌سازی انجام شده، قوانینی استخراج گشته است که در ادامه در مورد آنها بحث خواهد شد. شمای کلی الگوریتم درخت تصمیم‌گیری در این سکو در شکل ۲، قابل مشاهده است.



شکل ۲- ساختار الگوریتم درخت تصمیم در محیط SSDT

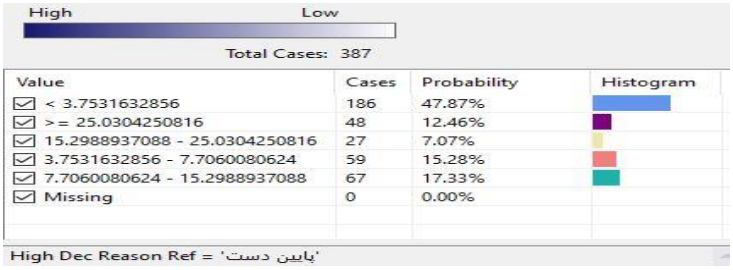
[1] Decision Trees

[2] Association Rules

[3] Neural Network

www.SID.ir

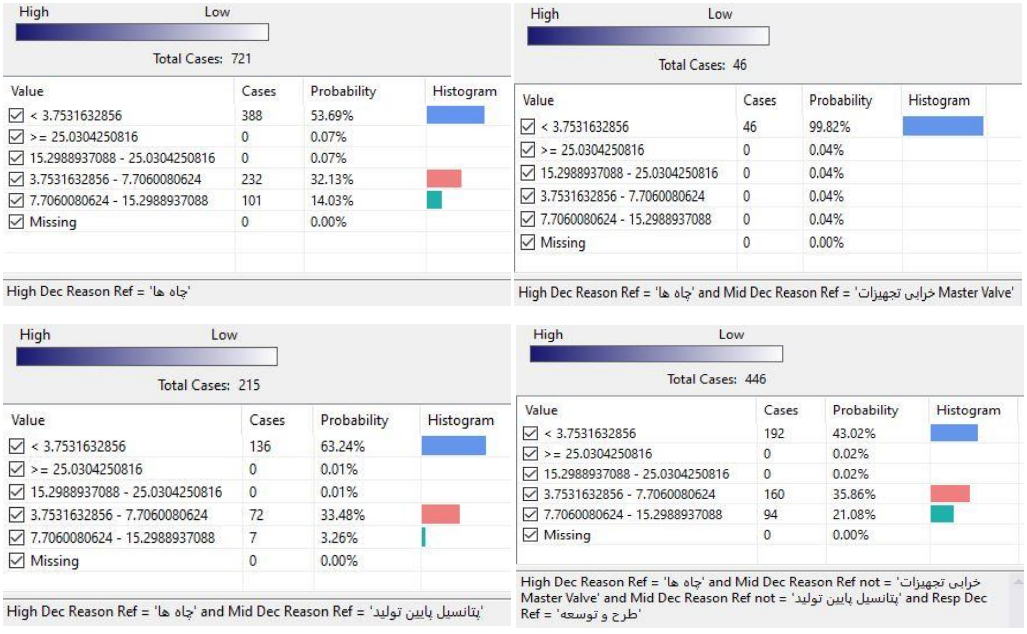
قانون شمارهٔ ۱: مطابق با شکل ۳، در محیط SSDT، در ۳۸۷ رخداد، کاهش تولید در حوزه پایین دست با مسئولیت پالایشگاه بوده است که در ۴۸ درصد تراکم آن با کاهش تولید کمتر از ۳٫۸ و در ۱۷ درصد با کاهش تولید بین ۷٫۷ تا ۱۵٫۳ میلیون متر مکعب مواجه گردیده است.



شکل ۳- قانون اول درخت تصمیم در محیط SSDT

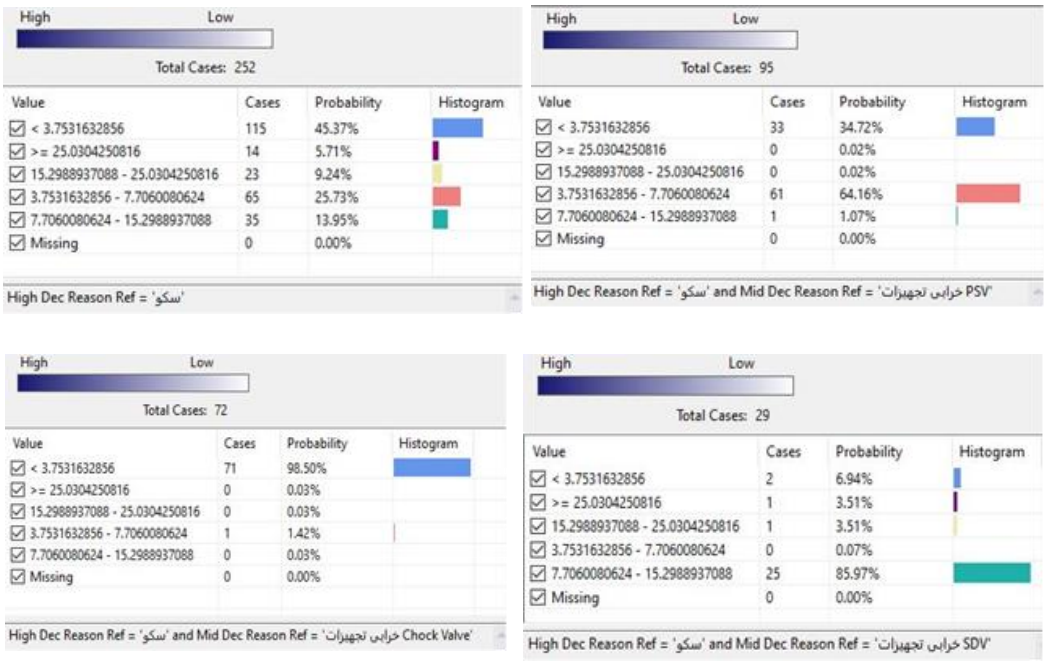
قانون شمارهٔ ۲: مطابق با شکل ۴، در محیط SSDT، ۷۲۱ رخداد کاهش تولید در حوزه چاه‌ها با مسئولیت معاونت طرح و توسعه بوده است که در ۵۴ درصد تراکم آن با کاهش تولید کمتر از ۳٫۸ و در ۱۴ درصد با کاهش تولید بین ۷٫۷ تا ۱۵٫۳ میلیون متر مکعب میلیون متر مکعب مواجه گردیده است. بر این اساس، ۴۶۰ رخداد مربوط به انجام فعالیت‌های حفاری چاه‌ها بوده است که در اختیار معاونت طرح و توسعه می‌باشد. در ۲۱۵ رخداد نیز بدلیل پتانسیل پایین تولید چاه‌ها و در ۴۶ رخداد نیز بدلیل خرابی تجهیزات مشخصاً خرابی تجهیز Master Valve می‌باشد.

قانون شمارهٔ ۳: مطابق با شکل ۵، در محیط SSDT، ۲۵۲ رخداد کاهش تولید در حوزهٔ سکو اتفاق افتاده که در ۴۵ درصد تراکم آن، با کاهش تولید کمتر از ۳٫۸ و در ۱۴ درصد با کاهش تولید بین ۷٫۷ تا ۱۵٫۳ میلیون متر مکعب مواجه گردیده است. در این میان عامل خرابی تجهیزات و مشخصاً خرابی تجهیز PSV با ۹۵ رخداد، خرابی تجهیز Chock Valve با ۸۵ رخداد و خرابی تجهیز SDV با ۲۹ رخداد بیشترین تاثیرگذاری، در بروز کاهش‌های تولید را داشته‌اند.



شکل ٤- قانون دوم درخت تصمیم در محیط SSDT

www.SID.ir



شکل ۵- قانون سوم درخت تصمیم در محیط SSDT

۴-۲-الگوریتم قوانین انجمنی

استخراج قواعد انجمنی نوعی عملیات داده‌کاوی است که به جستجو برای یافتن ارتباط بین ویژگی‌ها در مجموعه داده‌ها می‌پردازد. به عبارتی قوانین وابستگی (انجمنی) اتفاق و وقوع یک شیء را براساس وقوع سایر اشیاء توصیف می‌کنند. برای مثال در یک سوپر مارکت هدف در کاوش قوانین انجمنی؛ یافتن نظم حاکم بر سبد خرید می‌باشد، در این کاربرد به ازای هر سبد؛ یک قانون پیدا می‌شود و بررسی خواهد شد که این قانون در چه تعداد از سبدها صدق می‌کند و در نهایت یک مجموعه قوانین که در بیشترین تعداد از سبدها صدق می‌کند به عنوان مجموعه قوانین انجمنی خروجی ارائه می‌شود. به بیان دیگر در این کاربرد به دنبال پیدا کردن یک مجموعه از قوانین وابستگی هستیم تا براساس آن قوانین بتوانیم نتیجه‌گیری کنیم وجود کدامیک از مجموعه اشیاء بر وجود چه مجموعه اشیاء دیگری تاثیر گذار است (غضنفری، ۱۳۹۵).

با پیاده سازی این الگوریتم بر روی سکوی SPD12 و با در نظر گرفتن متغیرهای ورودی و پیش بینی مشابه الگوریتم درخت تصمیم به قوانینی به شرح زیر دست خواهیم یافت.

البته باید توجه داشت که در مدل قوانین انجمنی، تعداد بسیار زیادی قوانین استخراج شده که با اعمال فیلترینگ و انتخاب شاخص؛ حداقل احتمال رخداد به میزان ۶۰ درصد قوانینی معنا دار به شرح زیر استخراج گردیده است:

قانون شماره ۱: چنانچه حوزهٔ رخداد کاهش تولید سکو با مسئولیت معاونت تولید و عملیات و علت بروز آن نشت گاز از اتصالات لوله‌ها باشد، آنگاه سکو به احتمال ۱۰۰٪ با کاهش تولید بین ۱۵,۳ تا ۲۵ میلیون متر مکعب مواجه خواهد شد.

قانون شمارهٔ ۲: چنانچه حوزهٔ رخداد کاهش تولید سکو با مسئولیت معاونت تولید و عملیات و علت بروز آن خرابی تجهیزات و مشخصاً خرابی تجهیز ESD باشد، آنگاه سکو به احتمال ۱۰۰٪ با کاهش تولید بین ۳,۷ تا ۷,۷ میلیون متر مکعب مواجه خواهد شد.

قانون شمارهٔ ۳: چنانچه حوزهٔ رخداد کاهش تولید خطوط لوله دریایی با مسئولیت معاونت طرح و توسعه و علت بروز آن افزایش فشار خط باشد، آنگاه سکو به احتمال ۱۰۰٪ با کاهش تولید کمتر از ۳,۷ میلیون متر مکعب مواجه خواهد شد.

قانون شمارهٔ ۴: چنانچه حوزهٔ رخداد کاهش تولید سکو با مسئولیت معاونت تولید و عملیات و علت بروز آن خرابی تجهیزات و مشخصاً خرابی تجهیز Chock Valve باشد، آنگاه سکو به احتمال ۹۸٪ با کاهش تولید کمتر از ۳,۷ میلیون متر مکعب مواجه خواهد شد.

قانون شمارهٔ ۵: چنانچه حوزهٔ رخداد کاهش تولید پایین دست با مسئولیت پالایشگاه و علت بروز آن تاخیر در اتمام بموقع تعمیرات اساسی باشد، آنگاه سکو به احتمال ۱۰۰٪ با کاهش تولید بین ۱۵,۳ تا ۲۵ میلیون متر مکعب مواجه خواهد شد.

www.SID.ir

قانون شمارهٔ ۶: چنانچه حوزهٔ رخداد کاهش تولید سکو با مسئولیت معاونت تولید و عملیات و علت بروز آن خرابی تجهیزات و مشخصاً خرابی تجهیز SDV باشد، آنگاه سکو به احتمال ۸۶٪ با کاهش تولید بین ۷,۷ تا ۱۵,۳ میلیون متر مکعب مواجه خواهد شد.

۴-۳-الگوریتم شبکهٔ عصبی

شبکه‌های عصبی همانند، روش کار مغز انسان برای حل مسائلی است که با آن مواجه می‌شود. ابتدا حقایق مساله را در چند سطح تحلیل کرده و می‌سنجد. سپس این حقایق، وارد نرون‌های عصبی می‌شوند. این نرون‌های عصبی مانند فیلترهایی که براساس الگوهای معلوم قبلی عمل می‌کنند، شروع به فیلتر کردن حقایق می‌نمایند. درنهایت این موضوع سبب استنتاج می‌گردد که ممکن است منجر به پیدا کردن راه حلی برای مساله شود. در حقیقت شبکهٔ عصبی مصنوعی یک مدل محاسباتی انتزاعی از مغز انسان است. الگوریتم شبکه عصبی مایکروسافت، نرون‌های عصبی مصنوعی را بین ورودی‌ها و خروجی‌ها، برقرار می‌سازد و از آنها به عنوان الگو برای پیشبینی‌های آینده استفاده می‌نماید. مزیت این الگوریتم نسبت به الگوریتم‌های دیگر، کشف روابط خیلی پیچیده بین ورودی‌ها و خروجی‌ها است. البته نسبت به الگوریتم‌های دیگر زمان بیشتری را جهت ساخت و آموزش مدل استفاده می‌کند (غضنفری، ۱۳۹۵). در فاز اجرای الگوریتم شبکه عصبی بر روی سکوی SPD12 نیز با در نظر گرفتن متغیرهای ورودی و پیش بینی مشابه الگوریتم های قبلی نتایجی مشابه جدول.... حاصل گردیده که در واقع نسبت به اولویت بندی هر یک از عوامل کاهش تولید اقدام نموده است.

جدول ۳- نتایج حاصل از اجرای الگوریتم شبکه عصبی

ماهیت کاهش	حوزه رخداد	علت اصلی	احتمال میزان کاهش تولید کمتر از ۳,۷	احتمال میزان کاهش تولید بین ۳,۷ تا ۷,۷	احتمال میزان کاهش تولید بیشتر از ۷,۷
برون سازمانی	پالایشگاه	محدودیت در خطوط شبکه سراسری انتقال گاز	٪۸۰	٪۸	٪۱۲
		اشکال در کمپرسورهای Export Gas پالایشگاه	٪۱۵	٪۲۳	٪۶۲
		اشکال در شیرهای GOV پالایشگاه	٪۴۲	٪۷	٪۵۱
		مشکلات فنی و عملیاتی در پالایشگاه	٪۸	٪۱۵	٪۷۷
		تاخیر در اتمام تعمیرات اساسی پالایشگاه	٪۷۰	٪۹	٪۲۱
		محدودیت در مخازن ذخیره سازی	٪۵۹	٪۱۰	٪۳۱
درون سازمانی	چاه ها	انجام فعالیت های تعمیر و نگهداشت چاه ها	٪۱۸	٪۱۸	٪۶۴
		گریس‌کاری تجهیزات	٪۲۲	٪۳۷	٪۴۱
		پتانسیل پایین تولید چاه ها	٪۳۲	٪۲۵	٪۴۳
		خرابی تجهیزات سرچاهی DHSV	٪۱۵	٪۱۶	٪۶۹
		تست و نشت یابی چاه ها	٪۶۳	٪۱۱	٪۲۶
	سکو	نشت گاز از اتصالات لوله ها	٪۴	٪۴	٪۹۲
		خرابی تجهیزات Master Valve	٪۵۸	٪۵	٪۳۷
		خرابی تجهیزات ESD	٪۱۵	٪۱۰	٪۷۵
		خرابی تجهیزات Chock Valve	٪۲۷	٪۹	٪۶۴
		خرابی تجهیزات SDV	٪۲۸	٪۱۴	٪۵۸
		خرابی تجهیزات PSV	٪۲۱	٪۴۴	٪۳۵
		خرابی تجهیزات UPS	٪۸	٪۳۱	٪۶۱
		خرابی تجهیزات Control Valves	٪۲۷	٪۹	٪۶۴
	خط لوله	افزایش فشار	٪۳۱	٪۱۶	٪۵۳

نتایج حاصل از اجرای الگوریتم شبکه عصبی مطابق جدول ۳، میزان اهمیت هر یک از عوامل بروز کاهش تولید را در سکوی مذکور را نشان می‌دهد؛ به عبارت دیگر با استفاده از این نتایج می‌توان نسبت به اولویت‌بندی عوامل کاهش تولید بر اساس میزان

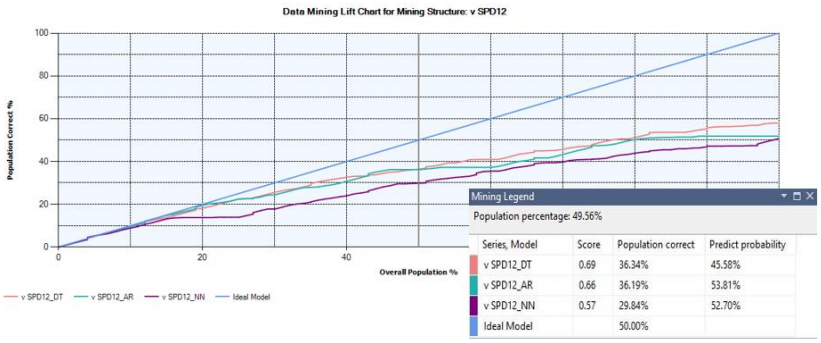
www.SID.ir

تأثیرگذاری آنها بر متغیر تولید و برنامه‌ریزی جهت تعدیل آنها استفاده نمود. به عنوان مثال چنانچه عامل کاهش تولید، نشت گاز از اتصالات لوله‌ها، خرابی تجهیزات ESD یا خرابی تجهیزات سرچاهی باشد، سکو بترتیب با احتمال ۹۲٪ ، ۷۵٪ و ۶۹٪ با کاهش تولید بیش از ۷,۷ میلیون متر مکعب مواجه خواهد گردید؛ که نشان دهندهٔ بیشترین تأثیر در میان سایر عوامل کاهش تولید می‌باشد.

۵-ارزیابی مدل

در این مرحله می‌بایست مدل تهیه شده مورد ارزیابی قرار گرفته و بررسی شود که آیا این مدل پاسخگوی نیاز کسب و کار هست یا خیر؟ آیا اهداف پروژه را تعیین می کند یا خیر؟ نتایج حاصل از ارزیابی باعث بهبود مدل شده و آنرا قابل استفاده می‌نماید. در این گام اعتبار مدل بررسی شده و گزارشی از کل فرایند تهیه خواهد شد. به عبارت دیگر، در مسائل داده‌کاوی یکی از مهمترین مواردی که برای ارزیابی موفقیت یا عدم موفقیت مدل مورد استفاده قرار می‌گیرد دقت و قدرت تفکیک مدل ایجاد شده است (شکوهیار، ۱۳۹۵).

در نمودار ۱، تحت عنوان (Lift Chart) ارزیابی مدل تولید شده مورد بررسی قرار می‌گیرد. همانطور که قابل مشاهده است، دقت مدل حاصل از اجرای الگوریتم‌های درخت تصمیم، قوانین انجمنی و شبکه عصبی به ترتیب ۶۹٪، ۶۶٪ و ۵۷٪ می‌باشد که حاکی از آن است که مدل درخت تصمیم از مطلوبیت نسبتاً بیشتری برخوردار می‌باشد. همچنین در نمودار زیر ستون (Population Correct) نشان دهندهٔ آن است که رکوردها تا چه میزان صحت و دقت مورد پیشبینی واقع گردیده‌اند. در این حالت، مطلوبیت مدل بصورت پیشفرض برابر با ۵۰٪ در نظر گرفته شده است و مدل درخت تصمیم با میزان ۳۶,۳۴٪ بیشترین مطلوبیت را دارا می‌باشد.



نمودار ۱- ارزیابی الگوریتم‌های اجرا شده در محیط *SSDT*

۵-۱- ماتریس درهم ریختگی[1] در ارزیابی کارایی الگوریتم

در این ماتریس چگونگی عملکرد الگوریتم دسته‌بندی با توجه به مجموعه داده‌های ورودی و نحوه تفکیک آنها در دسته بندی‌های مناسب و نامناسب مورد تحلیل و نمایش قرار می‌گیرد (فتحی و محمدی،۱۳۹۶). جدول ۴ ماتریس درهم ریختگی را نشان می‌دهد.

جدول ۴- ماتریس درهم ریختگی

		تخمین یا پیش بینی‌ها	
		درست T	نادرست F
نمونه‌های واقعی	مثبت P	TP	FN
	منفی N	FP	TN

ماتریس درهم ریختگی دارای چهار عنصر می‌باشد:

True Positive (TP): نمونه‌های مثبتی که به درستی مثبت تشخیص داده شده است.

True Negative (TN): نمونه‌های منفی که به در ستی منفی تشخیص داده شده است.

[1] Confusion Matrix

www.SID.ir

False Negative (FN): نمونه‌های مثبتی که به صورت نادرست منفی تشخیص داده شده است.

False Positive (FP): نمونه‌های منفی که به صورت نادرست مثبت تشخیص داده شده است.

برای تعیین کارایی، الگوریتم دقت از مهم‌ترین معیارهایی است که مشخص می‌کند، دسته‌بندی طراحی شده چند درصد از کل مجموعهٔ رکوردهای آزمایشی را بدرستی دسته بندی کرده است.

ارزیابی نرخ مثبت درست در استخراج داده:

درستی اطلاعات به دست آمده را با فراخوانی یا TPR محاسبه می کنند. از طریق حاصل تقسیم تعداد مستندات بازیابی شده واقعا با ربط، بر تعداد کل مستندات مرتبط موجود.

$$TPR=TP/(TP+FN) \quad (۱)$$

ارزیابی نرخ مثبت نادرست در استخراج داده:

آن دسته از داده‌ای منفی که به اشتباه در دسته مثبت‌ها کلاسه بندی شده‌اند.

$$FPR=FP/(FP+TN) \quad (۲)$$

ارزیابی نرخ منفی درست در استخراج داده:

آن دسته از داده‌ای منفی که به درستی در دسته منفی‌ها کلاسه بندی شده‌اند.

$$TNR=TN/(TN+FP) \quad (۳)$$

ارزیابی نرخ منفی نادرست در استخراج داده:

داده‌ای مثبتی که به اشتباه در دسته منفی‌ها کلاسه بندی شده‌اند.

$$FNR=FN/(TP+FN) \quad (۴)$$

ارزیابی دقت الگوریتم:

حاصل تقسیم تعداد مستندات بازریابی شده واقعا با ربط بر تعداد کل مستندات بازریابی شده.

$$Precision=TP/(TP+FN) \quad (۵)$$

ارزیابی صحت الگوریتم:

نسبت تعداد کل پیش بینی‌های درست می‌باشد.

$$Accuracy=(TP+TN)/(TN+TP+FN+FP) \quad (۶)$$

شکل ۶ اجرای ماتریس طبقه‌بندی[1] در محیط SSDT بمنظور انجام ارزیابی مدل درخت تصمیم برای مجموعه داده‌های تست در سکوی SPD12 را نشان می‌دهد که با استفاده از نتایج آن، ماتریس در هم ریختگی و شاخص‌های ارزیابی بصورت جداول ۵ و ۶ حاصل می‌گردد.

جدول ۵- ماتریس درهم ریختگی

	پیش‌بینی‌ها	
نمونه‌های واقعی	۳۹۲	۳۹
	۲۴۱	۵

شکل ٦- ماتریس طبقه بندی در محیط SSDT

[1] Classification Matrix

www.SID.ir

جدول ٦- محاسبه شاخص‌های ارزیابی

ردیف	شاخص	مقدار	ردیف	شاخص	مقدار
۱	ارزیابی نرخ مثبت درست در استخراج داده	٪۹۱	۴	ارزیابی نرخ منفی اشتباه در استخراج داده	٪۹
۲	ارزیابی نرخ مثبت نادرست در استخراج داده	٪۹۸	۵	ارزیابی دقت الگوریتم	٪۹۱
۳	ارزیابی نرخ منفی درست در استخراج داده	٪۲	۶	ارزیابی صحت الگوریتم	٪۵۹

نتایج بدست آمده بر اساس جدول ۶ حاکی از آن است که مدل حدود ۶۰ درصد موارد را به درستی پیش بینی کرده و دقت آن در پیش بینی موارد منجر به کاهش تولید حدود ۹۰ درصد و مواردی که منجر به کاهش تولید نبوده اند ۹۵ درصد بوده است.

نتیجه‌گیری

در این پژوهش مدلی بمنظور شناسایی و ارزیابی عوامل مؤثر برکاهش تولید گاز و میعانات‌گازی در سکوی فراساحلی SPD12 میدان پارس جنوبی با رویکرد داده‌کاوی ارائه گردید. پیش‌فرض در این پژوهش بر این اصل استوار بود که داده‌کاوی می‌تواند ما را در ارزیابی عوامل مؤثر برکاهش تولید سکوی مذکور، یاری نماید. بدین منظور از سه الگوریتم مطرح در حوزهٔ داده‌کاوی تحت عناوین درخت تصمیم‌گیری، شبکه عصبی و قوانین انجمنی استفاده و نتایج حاصل از آن مورد ارزیابی قرار گرفت. اولویت بندی و میزان تأثیر هر یک از عوامل بر متغیر تولید و در نتیجه برنامه‌ریزی جهت پیشگیری و تعدیل آنها بمنظور تحقق برنامه مصوب تولید از نتایج مهم این پژوهش می‌باشد.

منابع

- تایوغ اصغر، حاتم لو علیرضا (۱۳۹۳). «داده کاوی در بانکهای اطلاعاتی سازمانی جهت تعین استراتژی و پیشبرد اهداف سازمان». اولین همایش ملی پژوهش‌های مهندسی رایانه.
- شکوهیار سجاد، رجا نیلوفر (۱۳۹۵)، «کاربردهای داده کاوی در نرم افزار مایکروسافت اکسل»، چاپ اول، تهران، موسسه کتاب مهربان نشر.
- شهرابی جمال، شکورنیاز ونوس (۱۳۹۲)، «داده کاوی در SQL Server»،چاپ دوم، تهران، جهاد دانشگاهی صنعتی امیرکبیر.
- غضنفری مهدی، علیزاده سمیه، تیمورپور بابک (۱۳۹۵)، «داده کاوی و کشف دانش»، چاپ پنجم، تهران، مرکز انتشارات دانشگاه علم و صنعت ایران.
- Abou-Sayed, A. (2012). Data Mining Applications in the Oil and Gas Industry.
- MacLennan, J., Crivat, B. (2008). Data Mining With Microsoft SQL Server, Wiley Publishing.
- SQL server 2012 Tutorial: Analysis Services Data Mining. (2012). Microsoft.

www.SID.ir