

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی معدن، نفت و ژئوفیزیک  
گروه مهندسی اکتشاف معدن

مدل‌سازی و شبیه‌سازی محتوای کربن آلی با استفاده از چاه نمودارهای پتروفیزیکی به وسیله  
الگوریتم لولیموت در شیل‌های گازی

محمد جودی نصرالله بیگلو

اساتید راهنما :

دکتر بهزاد تخم‌چی

دکتر محمدرضا رضایی

استاد مشاور:

دکتر محمد کنشلو

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۱۳۹۰

## دانشگاه صنعتی شاهرود

دانشکده : مهندسی معدن، نفت و ژئوفیزیک

گروه : اکتشاف معدن، نفت و ژئوفیزیک

پایان نامه کارشناسی ارشد آقای محمد جودی نصرالله بیگلو

تحت عنوان: مدل سازی و شبیه سازی محتوای کربن آلی با استفاده از چاه نمودارهای

پتروفیزیکی به وسیله الگوریتم لولیموت در شیل های گازی

در تاریخ ۱۳۹۱/۱۱/۲۸ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد

مورد ارزیابی و با درجه ----- مورد پذیرش قرار گرفت.

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	دکتر محمد کنشلو		دکتر بهزاد تخمچی
	-		دکتر محمدرضا رضایی

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	دکتر سوسن ابراهیمی		دکتر علی مرادزاده
			دکتر سیدعلی سلیمانی ایوری

با احترام تقدیم به

پدر و مادر مهربانم

و برادر، همراهم

# من لم يشكر المخلوق، لم يشكر الخالق

## تشکر و قدردانی

خداوند متعال را سپاسگزارم از این که توانستم این پایان‌نامه را به اتمام برسانم. همچنین خدا را شاکرم که در حین انجام این پایان‌نامه با افراد خیره و شایسته‌ای آشنا شدم و همکاری داشتم تا با راهنمایی و مشاوره آن‌ها این پروژه در مسیر مناسب خود قرار گیرد. به رسم ادب و احترام، در ابتدای این پایان‌نامه باید از زحمات بزرگوارانی که بنده را در انجام این پایان‌نامه یاری نمودند، تشکر و قدردانی نمایم.

از جناب آقای **دکتر بهزاد تخم‌چی** که به عنوان استاد راهنما در زمان‌هایی که روند پایان‌نامه با مشکل مواجه می‌شد با راهنمایی‌های کارساز راه را روشن می‌ساخت، آقای **دکتر محمد کنشلو** که مانند برادری بزرگوار مشاوره‌های کارآمدی به بنده ارائه دادند، آقای **دکتر محمدرضا رضایی** که در تهیه داده‌ها بنده را یاری نمودند، آقای **محمدجواد مجلسی** که در پردازش داده بنده را یاری نمودند، خانم **مهرنوش علیپور** که در تهیه کدها مرا همراهی نمودند و آقایان **امیر ملاجان** و **یوسف عسگری نژاد** و خانم **مینا دلنوا** که در مراحل مختلف مرا از راهنمایی‌های خود بهره‌مند کردند.

به امید زندگی در حضور امام زمان (عج)

محمد جودی نصرالله بیگلو

## تعهد نامه

اینجانب **محمد جودی نصرالله بیگلو** دانشجوی دوره کارشناسی ارشد رشته **مهندسی اکتشاف معدن** دانشکده **مهندسی معدن، نفت و ژئوفیزیک** دانشگاه صنعتی شاهرود نویسنده پایان نامه **مدل سازی و شبیه سازی محتوای کربن آلی با استفاده از چاه نمودارهای پتروفیزیکی به وسیله الگوریتم لولیموت در شیل های گازی** تحت راهنمایی **دکتر بهزاد تخمچی** و **دکتر محمدرضا رضایی** متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهش های محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه 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 می باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «*Shahrood University of Technology*» به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده ( یا بافت های آنها ) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، 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 ، برنامه های رایانه ای ، نرم افزارها و تجهیزات ساخته شده است ) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

## چکیده

تخمین پارامترهای مختلف مخازن نفتی یکی از مهم‌ترین مراحل در اکتشاف منابع هیدروکربوری است. تخلخل، تراوایی، سطح جدایش نفت، گاز و آب از جمله این پارامترهای مخزنی هستند. محتوای کربن آلی نیز یکی از پارامترهای مخازن نفتی و گازی است که در ارزیابی مخازن اهمیت دارد. برای اندازه‌گیری این پارامترها روش‌های مختلفی وجود دارد که به دو دسته مستقیم و غیرمستقیم تقسیم می‌شوند. در روش‌های مستقیم پارامتر مورد نظر را از طریق آزمایش بر روی نمونه‌های مغزه و خرده‌های حفاری بدست می‌آورند و در روش‌های غیرمستقیم پارامترهای مربوطه به وسیله چاه نمودارهای پتروفیزیکی محاسبه می‌شود. در پایان نامه حاضر روشی جدید بر پایه الگوریتم لولیموت از دسته روش‌های هوشمند برای تخمین محتوای کربن آلی ارائه شده است. پس از آن با استفاده از چاه نمودارهای موجود از ۹ حلقه چاه حفاری شده در حوزه پرت واقع در جنوب غربی استرالیا، پارامتر محتوای کربن آلی تخمین زده شده است. در نهایت مقادیر تخمینی با مقادیر واقعی مقایسه شده است که اعتبار سنجی انجام شده همبستگی بین ۴۰ تا ۸۰ درصد را در موارد مختلف نشان می‌دهد.

**کلمات کلیدی:** تخمین، محتوای کربن آلی، لولیموت، خوشه‌بندی، کلاسه‌بندی، پتروفیزیک

## لیست مقالات مستخرج از این پایان نامه

۱. جودی م.، تخم‌چی ب.، رضایی م. کنشلو م. علیپور م.، مجلسی کوپایی

م. ۱۳۹۱، "بررسی کارایی روش DBSCAN به منظور خوشه‌بندی و کلاسه‌بندی داده‌ها

در علوم زمین: مطالعه موردی میدان پرت در جنوب غربی استرالیا"، سی و یکمین

گردهمایی علوم زمین، تهران، ایران



## فهرست مطالب

۲.....	۱ کلیات
۳.....	۱-۱ محتوای کربن آلی
۳.....	۲-۱ تاریخچه تخمین TOC
۷.....	۳-۱ لولیموت
۷.....	۴-۱ معرفی فصول پایان نامه
۱۰.....	۲ روش شناسی تحقیق
۱۰.....	۱-۲ روش میانگین K
۱۲.....	۲-۲ روش گوستافسون-کسل
۱۳.....	۱-۲-۲ الگوریتم گوستافسون-کسل
۱۵.....	۳-۲ معرفی روش DBSCAN
۱۵.....	۱-۳-۲ ساختار الگوریتم DBSCAN
۲۰.....	۲-۳-۲ الگوریتم DBSCAN برای کلاسه بندی
۲۱.....	۴-۲ الگوریتم لولیموت
۲۲.....	۱-۴-۲ ساختار الگوریتم لولیموت
۲۴.....	۲-۴-۲ روند کار الگوریتم لولیموت
۲۵.....	۳-۴-۲ الگوریتم لولیموت
۲۹.....	۴-۴-۲ ساختار الگوریتم لولیموت
۳۰.....	۵-۴-۲ روش ساخت یک LLM
۳۲.....	۶-۴-۲ بررسی انواع شرط توقف الگوریتم
۳۵.....	۳ داده ها
۳۵.....	۱-۳ داده های واقعی
Error! Bookmark not defined.....	۱-۱-۳ موقعیت منطقه
۳۸.....	۲-۱-۳ داده های اولیه

۳۱.....	۳-۱-۳	گزینش داده‌های مورد استفاده
۴۰.....	۳-۱-۴	تاریخچه ای از چاه‌های مورد مطالعه
۵۱.....	۳-۱-۵	سازندهای هدف
۵۲.....	۳-۲-۲	داده‌های مصنوعی
۵۲.....	۳-۲-۱	دسته اول
۵۳.....	۳-۲-۲	دسته دوم
۵۴.....	۳-۲-۳	دسته سوم
۵۴.....	۳-۲-۴	دسته چهارم
<b>۵۷.....</b>	<b>۴</b>	<b>مدل‌سازی محتوای کربن آلی</b>
۵۷.....	۴-۱-۱	بررسی کارایی روش‌های مورد استفاده
۵۷.....	۴-۱-۱	خوشه‌بندی DBSCAN
۶۱.....	۴-۱-۲	کلاسه‌بندی DBSCAN
۶۲.....	۴-۱-۳	لولیموت
۶۵.....	۴-۲-۲	آماده‌سازی داده‌ها
۶۶.....	۴-۲-۱	هماهنگ‌سازی داده‌ها
۶۷.....	۴-۲-۲	خوشه‌بندی و کلاسه‌بندی داده‌ها
۷۳.....	۴-۳-۳	اجرای الگوریتم لولیموت بر روی داده‌ها
۷۵.....	۴-۴	نحوه انجام پروژه
۸۳.....	۴-۵	روند آموزش و آزمون
<b>۱۰۲.....</b>	<b>۵</b>	<b>نتیجه‌گیری و پیشنهادات</b>
۱۰۲.....	۵-۱	نتیجه‌گیری
۱۰۳.....	۵-۲	پیشنهادات
<b>۱۰۴.....</b>	<b>۶</b>	<b>مراجع</b>

## فهرست اشکال

- شکل ۱-۱ مثالی از نحوه کار روش  $\Delta \text{LOG } R$  برای تخمین مقدار TOC با استفاده از چاه نمودارهای گاما، مقاومت و صوتی. برهم نهی چاه نمودارهای مقاومت و صوتی مقدار جدایش  $\Delta \text{LOG } R$  را در مقاطع غنی از مواد آلی بدست می‌دهد. (PASSEY, CREANEY ET AL. 1990)..... ۵
- شکل ۱-۲ نمودار تبدیل  $\Delta \text{LOG } R$  به TOC با توجه به درجه بلوغ به همراه نمودارهای تعیین درجه بلوغ با استفاده از مقادیر TOC و  $S_2$  با توجه به نوع کروژن (کروژن نوع ۲ شکل سمت چپ و کروژن نوع ۳ شکل سمت راست) ..... ۶
- شکل ۱-۲ روش کار الگوریتم DBSCAN با توجه به تراکم نقاط برای خوشه‌بندی نقاط مختلف ..... ۱۸
- شکل ۲-۲ الف) تخمین خطی مدل غیرخطی در حالت جهانی برای مدل، ب) تخمین خطی مدل غیرخطی در منطقه محدود A ..... ۲۳
- شکل ۲-۲ الف) وزن دهی مدل به صورت مطلق، ب) وزن دهی مدل به صورت فازی ..... ۲۴
- شکل ۲-۴ روند کار الگوریتم لولیموت در یک محیط دو بعدی (NELLES 2000) - مناطق خاکستری نشان دهنده قسمتی است که دارای بیشترین خطا است و احتیاج به تقسیم دارد. .... ۲۸
- شکل ۲-۵ ساختار الگوریتم لولیموت - هر نرون شامل یک مدل خطی و یک قسمت برای محاسبه تابع اعتبار است. .... ۳۰
- شکل ۳-۱ نقشه چاه‌های حفاری شده و موقعیت منطقه نسبت به استرالیا ..... ۳۷
- شکل ۳-۲ نمایی از چینه شناسی، لایه‌ها و سازندهای موجود در حوزه پرت (MARSHAL, MITCHELL ET AL. 1983) ..... ۴۲
- شکل ۳-۳ نمایی از ستون چینه شناسی چاه راکینگهام ۱ (MARSHAL, MITCHELL ET AL. 1983) ..... ۵۰
- شکل ۳-۴ داده‌های مصنوعی تولید شده برای آزمودن کارایی روش DBSCAN ..... ۵۳
- شکل ۳-۵ داده‌های تولید شده برای مقایسه کارایی روش‌های مختلف در حالت داده‌های منظم تر ..... ۵۳
- شکل ۳-۶ داده‌های تصادفی تولید شده برای بررسی کارایی روش کلاسه‌بندی بر پایه روش خوشه‌بندی DBSCAN ..... ۵۴
- شکل ۳-۷ داده مصنوعی تولید شده برای نشان دادن گام‌های اجرای الگوریتم لولیموت: الف) نمودار تابع اصلی، ب) شکل نمونه‌های برداشته شده ..... ۵۵
- شکل ۴-۱ نتیجه خوشه‌بندی با روش DBSCAN خوشه‌های مختلف با رنگ‌های متفاوت نشان داده شده است ..... ۵۸
- شکل ۴-۲ نتیجه حاصل از خوشه‌بندی با روش میانگین K خوشه‌های مختلف با رنگ‌های متفاوت نشان داده شده است ..... ۵۸
- شکل ۴-۳ نتیجه حاصل از خوشه‌بندی با روش گوستافسون - کسل خوشه‌های مختلف با رنگ‌های متفاوت نشان داده شده است ..... ۵۹

- شکل ۴-۴ نتیجه حاصل از خوشه‌بندی با روش DBSCAN در داده‌های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است ..... ۵۹
- شکل ۴-۵ نتیجه حاصل از خوشه‌بندی با روش میانگین K در داده‌های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است ..... ۶۰
- شکل ۴-۶ نتیجه حاصل از خوشه‌بندی با روش گوستافسون-کسل در داده‌های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است ..... ۶۰
- شکل ۴-۷ نتیجه کلاسه‌بندی با استفاده از روش ابداعی بر پایه روش خوشه‌بندی DBSCAN ..... ۶۲
- شکل ۴-۸ نحوه حل مسئله یک بعدی با استفاده از الگوریتم لولیموت برای ۷ گام ابتدایی ..... ۶۴
- شکل ۴-۹ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند کاکلیشیل گولی ..... ۷۶
- شکل ۴-۱۰ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ گیج ..... ۷۷
- شکل ۴-۱۱ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند لیدرویل ..... ۷۷
- شکل ۴-۱۲ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ لزوتر ..... ۷۸
- شکل ۴-۱۳ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند پارملیا ..... ۷۸
- شکل ۴-۱۴ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند شیل پرت جنوبی ..... ۷۹
- شکل ۴-۱۵ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند یاراگادی ..... ۷۹
- شکل ۴-۱۶ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند کاکلیشیل گولی پس از حذف داده‌های زاید ..... ۸۰
- شکل ۴-۱۷ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ گیج پس از حذف داده‌های زاید ..... ۸۰
- شکل ۴-۱۸ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند لیدرویل پس از حذف داده‌های زاید ..... ۸۱
- شکل ۴-۱۹ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ لزوتر پس از حذف داده‌های زاید ..... ۸۱
- شکل ۴-۲۰ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند پارملیا پس از حذف داده‌های زاید ..... ۸۲
- شکل ۴-۲۱ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند شیل پرت جنوبی پس از حذف داده‌های زاید ..... ۸۲
- شکل ۴-۲۲ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند یاراگادی پس از حذف داده‌های زاید ..... ۸۳
- شکل ۴-۲۳ میزان خطای آموزش و آزمون برای سازند کاکلیشیل گولی-حالت انتخاب شده ۳ مدل ..... ۸۴
- شکل ۴-۲۴ میزان خطای آموزش و آزمون برای سازند ماسه سنگ گیج-حالت انتخاب شده ۶ مدل ..... ۸۴
- شکل ۴-۲۵ میزان خطای آموزش و آزمون برای سازند لیدرویل-حالت انتخاب شده ۳ مدل ..... ۸۵

- شکل ۴-۲۶ میزان خطا طی آموزش و آزمون برای سازند ماسه سنگ لزوئر-حالت انتخاب شده ۴ مدل ..... ۸۵
- شکل ۴-۲۷ میزان خطا طی آموزش و آزمون برای سازند پارملیا-حالت انتخاب شده ۵ مدل ..... ۸۶
- شکل ۴-۲۸ میزان خطا طی آموزش و آزمون برای سازند شیل پرت جنوبی-حالت انتخاب شده ۶ مدل ..... ۸۶
- شکل ۴-۲۹ میزان خطا طی آموزش و آزمون برای سازند یاراگادی-حالت انتخاب شده ۸ مدل ..... ۸۷
- شکل ۴-۳۰ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند کاکلیشیل گولی ..... ۸۷
- شکل ۴-۳۱ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند ماسه سنگ گیج ..... ۸۸
- شکل ۴-۳۲ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند لیدرویل ..... ۸۸
- شکل ۴-۳۳ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند ماسه سنگ لزوئر ..... ۸۹
- شکل ۴-۳۴ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند پارملیا ..... ۸۹
- شکل ۴-۳۵ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند شیل پرت جنوبی ..... ۹۰
- شکل ۴-۳۶ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند یاراگادی ..... ۹۰
- شکل ۴-۳۷ TOC تخمینی برای چاه گیجروود ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۲
- شکل ۴-۳۸ TOC تخمینی برای چاه گیجروود ۲ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۳
- شکل ۴-۳۹ TOC تخمینی برای چاه لیک پرستون ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۴
- شکل ۴-۴۰ TOC تخمینی برای چاه مولالو ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۵
- شکل ۴-۴۱ TOC تخمینی برای چاه پل ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۶
- شکل ۴-۴۲ TOC تخمینی برای چاه کویینزراک ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۷
- شکل ۴-۴۳ TOC تخمینی برای چاه راکینگهام ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۸

شکل ۴-۴۴ TOC تخمینی برای چاه شوگرلوف ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه

گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۹۹

شکل ۴-۴۵ TOC تخمینی برای چاه وارنبرو ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه

گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی ..... ۱۰۰

## فهرست جداول

- جدول ۳-۱ موقعیت گوشه‌های منطقه انتخاب شده برای پردازش ..... ۳۹
- جدول ۳-۲ خلاصه ای از وضعیت چاه‌های مورد استفاده در پروژه ..... ۴۰
- جدول ۳-۳ تفکیکی از چینه شناسی چاه کویینزراک ۱ (BOZANIC 1969) ..... ۴۳
- جدول ۳-۴ داده‌های مربوط به چینه شناسی چاه گیجرود ۱ (BOZANIC 1969) ..... ۴۴
- جدول ۳-۵ داده‌های مربوط به چینه شناسی چاه گیجرود ۲ (MOYES 1971) ..... ۴۵
- جدول ۳-۶ داده‌های مربوط به چینه شناسی چاه وارنبرو ۱ (MOYES 1971) ..... ۴۶
- جدول ۳-۷ داده‌های مربوط به چینه شناسی چاه شوگرلوف ۱ (BIRD AND MOYES 1971) ..... ۴۷
- جدول ۳-۸ داده‌های مربوط به چینه شناسی چاه لیک پرینستون ۱ (YOUNG AND JOHANSON 1973) ..... ۴۸
- جدول ۳-۹ داده‌های مربوط به چینه شناسی چاه پل ۱ (PHILLIPS-AUSTRALIAN-OIL-COMPANY 1978) ..... ۴۹
- جدول ۴-۱ وزن‌های محاسبه شده برای LLMها در هر مرحله ..... ۶۵
- جدول ۴-۲ نتیجه کلاسه‌بندی داده‌های چاه گیجرود ۱ ..... ۶۸
- جدول ۴-۳ نتیجه کلاسه‌بندی داده‌های چاه گیجرود ۲ ..... ۶۸
- جدول ۴-۴ نتیجه کلاسه‌بندی داده‌های چاه لیک پرینستون ۱ ..... ۶۹
- جدول ۴-۵ نتیجه کلاسه‌بندی داده‌های چاه مولالو ۱ ..... ۶۹
- جدول ۴-۶ نتیجه کلاسه‌بندی داده‌های چاه پل ۱ ..... ۷۰
- جدول ۴-۷ نتیجه کلاسه‌بندی داده‌های چاه کویینزراک ۱ ..... ۷۰
- جدول ۴-۸ نتیجه کلاسه‌بندی داده‌های چاه راکینگهام ۱ ..... ۷۱
- جدول ۴-۹ نتیجه کلاسه‌بندی داده‌های چاه شوگرلوف ۱ ..... ۷۱
- جدول ۴-۱۰ نتیجه کلاسه‌بندی داده‌های چاه وارنبرو ۱ ..... ۷۲
- جدول ۴-۱۱ مجموع داده‌های مورد استفاده برای پردازش ..... ۷۲
- جدول ۴-۱۲ مجموع داده‌های اصلاح شده برای پردازش ..... ۷۳
- جدول ۴-۱۳ نتیجه تقسیم داده‌ها در چاه‌های موجود بر اساس نسبت ۷۰٪ به ۳۰٪ به ترتیب برای آموزش و آزمون ..... ۷۴
- جدول ۴-۱۴ نتیجه طبقه بندی داده‌ها با استفاده از روش DBSCAN با شعاع‌های مختلف ..... ۷۵
- جدول ۴-۱۵ خلاصه‌ای از وضعیت آموزش و آزمون الگوریتم لولیموت برای سازندهای مورد مطالعه ..... ۹۱

فصل اول

کلیات



# ۱ کلیات

با توجه به نیاز روز افزون بشر به انرژی، یافتن منابع جدید انرژی جزء لاینفک زندگی انسان شده است. در این بین کشف ذخایر نفت و گاز که رایج‌ترین شکل انرژی هستند بسیار حائز اهمیت است. در استخراج نفت و گاز از دل زمین حفاری این منابع پر هزینه‌ترین قسمت کار است. هزینه حفاری یک چاه نفتی بین ۱۵۰۰۰۰ تا ۴۲۰۰۰۰ دلار در روز است (<http://www.rigzone.com/data/dayrates>) که با توجه به شرایط مختلف از جمله منطقه حفاری، کارایی ادوات حفاری، آب و هوای منطقه و ... متغیر است. با احتساب حدود ۱۰۰ روز حفاری برای هر چاه و هزینه دیگر موارد لازم برای حفاری چاه نفتی هزینه حفاری هر چاه نفتی بین ۱۵ تا ۱۰۰ میلیون دلار متغیر است. با توجه به هزینه بالای حفاری به خصوص در عمق‌های زیاد یافتن نقاط بهینه برای حفاری از اهمیت خاصی برخوردار می‌شود.

برای کاهش هزینه استخراج هیدروکربور قبل از حفاری کارهای سطحی مانند گرانی سنجی، مغناطیس سنجی و لرزه نگاری انجام می‌شود و پس از آن نقاط حفاری اکتشافی پیشنهاد داده می‌شود. در مرحله حفاری اکتشافی سعی می‌شود مجموعه کاملی از چاه نمودارها برداشت شود. در مرحله بعد این چاه نمودارها با داده‌های قبلی تلفیق شده و برای حفاری‌های تولیدی استفاده می‌شود. چاه نمودارهای بدست آمده بسیار کارآمد هستند و برای تخمین موارد مختلف از جمله تخلخل، تراوایی، درصد اشباع نفت و آب و ... استفاده می‌شوند.

## ۱-۱ محتوای کربن آلی

سنگ‌های مخزن نفتی و گازی دارای مقدار قابل توجهی ماده آلی هستند که معمولاً به عنوان محتوای کربن آلی<sup>۱</sup> (TOC) شناخته می‌شود (Khoshnoodkia, Mohseni et al. 2011). محتوای کربن آلی موجود در سنگ مخزن طی بلوغی که به علت حرارت و فشار با گذشت زمان حاصل می‌شود، می‌تواند منجر به تولید هیدروکربور شود (Tissot and Welte 1978).

برای اندازه‌گیری مقدار TOC مخازن نفتی راه‌های گوناگونی وجود دارد. راه‌های اندازه‌گیری مستقیم عبارتند از: سوزاندن مستقیم نمونه و استفاده از روش پیرولیز<sup>۲</sup>. روش اول یعنی سوزاندن مستقیم نمونه امروزه کمتر مورد استفاده قرار می‌گیرد (رضایی ۱۳۸۰). روش دوم با استفاده از دستگاه پیرولیز راک-اول<sup>۳</sup> انجام می‌شود.

روش‌های غیرمستقیم برای اندازه‌گیری TOC اندازه‌گیری آن از روی چاه نمودارهای پتروفیزیکی است. در پایان نامه حاضر هدف تخمین و مدل‌سازی عامل محتوای کربن آلی با استفاده از چاه نمودارهای معمول با استفاده از ترکیب یک روش هوشمند و آماری است.

## ۲-۱ تاریخچه تخمین TOC

روش‌های معمول برای تشخیص غنی‌شدگی و بلوغ سنگ مادر آلی، تنوعی از آنالیزهای آزمایشگاهی مانند پیرولیز، آنالیز عنصری، ضریب بازتابندگی ویترونیات، شاخص آلتراسیون دمایی، کروماتوگرافی گازی و ... است (Meyer and Nederlof 1984). در مراجع مختلف از تأثیر ماده آلی بر روی ابزارهای چاه پیمایی رایج در صنعت نفت بحث شده است (Fertl and Chilingar 1988) و تلاش‌های بسیاری

---

<sup>1</sup> Total Organic Carbon

<sup>2</sup> Pyrolysis

<sup>3</sup> Rock-Eval Pyrolysis

برای ارائه روشی ساده و آسان از بر هم نهی نمودارهای درون چاهی بیان شده است. در سنگ‌های فقیر از ماده آلی و اشباع از آب، تغییرات نمودارهای مقاومت ویژه و صوتی به علت پاسخ هر دو نمودار به تغییرات تخلخل سازند می‌توانند با یکدیگر موازی باشند و یا با تغییرات مناسب مقیاس بر روی هم قرار گیرند (Schmoker 1981). در مقابل در سنگ مخزن دارای هیدروکربن و سنگ مادر دارای کروژن و هیدروکربن، جدایشی بین این دو نمودار اتفاق خواهد افتاد که با استفاده از نمودار پرتو گاما می‌توان مناطق مربوط به سنگ مخزن را تشخیص داد. جدایی بین دو نمودار اخیر به دو دلیل عمده صورت می‌گیرد (Dellenbach, Espitalie et al. 1983, Schmoker and Hester Schmoker 1979),  
(1983):

اول- پاسخ نمودار تخلخل (صوتی، چگالی، نوترون) به چگالی و سرعت پایین کروژن.

دوم- پاسخ نمودار مقاومت ویژه به سیال سازند.

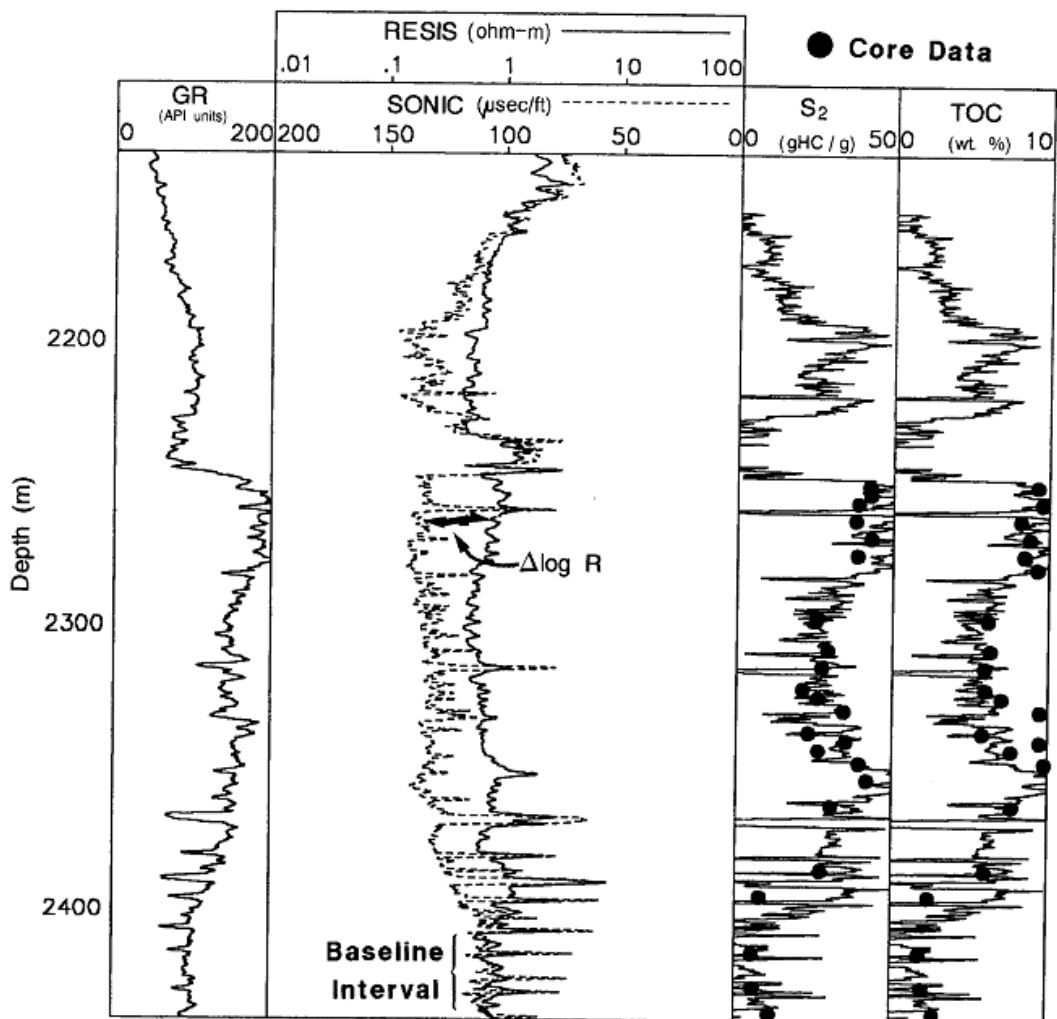
در یک سنگ مادر نابالغ غنی از ماده آلی، هیچ هیدروکربنی تولید نشده، جدایی مشاهده شده (بین دو نمودار اخیر) مربوط به پاسخ نمودار تخلخل (صوتی، نوترون، چگالی) است. در صورتی که در سنگ مادر بالغ آلی، علاوه بر پاسخ نمودار تخلخل، مقاومت ویژه به دلیل حضور هیدروکربن‌های تولید شده، افزایش می‌یابد (Dellenbach, Espitalie et al. 1983, Schmoker and Schmoker 1979),  
(Hester 1983).

در سال ۱۹۹۰، پسی<sup>۱</sup> روش عملی  $\Delta \log R$  را برای پیش‌بینی غنی‌شدگی آلی سنگ‌های مادر با استفاده از نمودارهای درون چاهی ارائه داد. این روش در دو نوع سنگ کربناته و آواری به قدر کافی موثر واقع می‌شود و در پیش‌بینی TOC در محدوده وسیعی از درجه بلوغ صحیح است (Passey, Creaney et al. 1990). در این روش از بر هم نهی دو نمودار درون چاهی مقاومت ویژه و تخلخل (به ویژه صوتی) با مقیاس نسبی مناسب (۱۰۰ میکروثانیه بر فوت در دو سیکل لگاریتمی مقاومت ویژه) استفاده می‌شود. جدایی بین این دو نمودار در هر عمقی به صورت  $\Delta \log R$  قابل محاسبه است (شکل

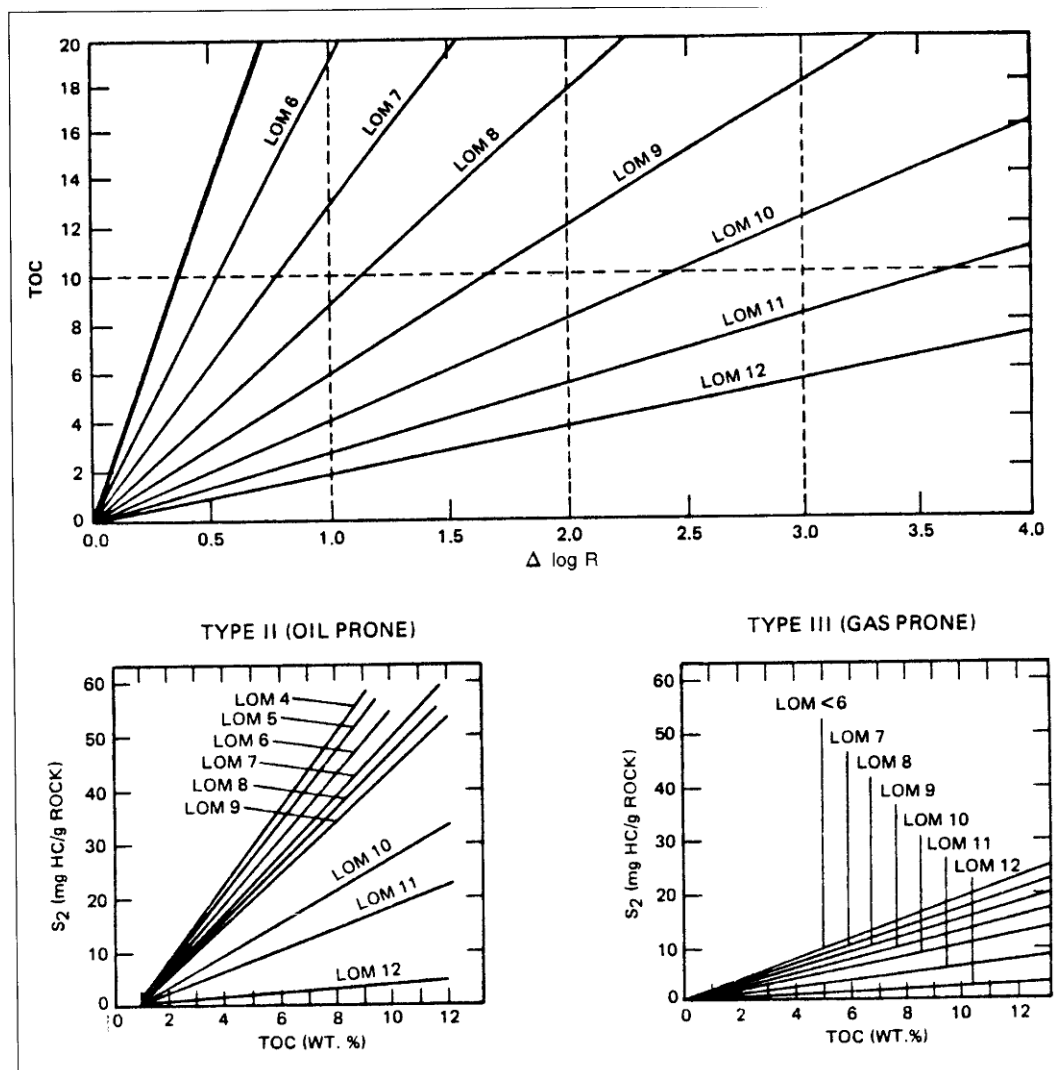
---

<sup>1</sup> Passey

۱-۱) (Passey, Creaney et al. 1990). پسی نشان داد که جدایی  $\Delta \log R$  به طور خطی با مقدار کل TOC ارتباط دارد و تابعی از درجه بلوغ است و با استفاده از دیاگرام  $\Delta \log R$  (شکل ۱-۲) این جدایی می‌تواند به طور مستقیم به TOC تبدیل شود. نمودارهای درون چاهی مقاومت ویژه و صوتی در مناطق غنی از ماده آلی دارای جدایی مثبت هستند. جدایی مثبت زمانی اتفاق می‌افتد که چاه نمودار صوتی در سمت چپ قرار بگیرند.



شکل ۱-۱ مثالی از نحوه کار روش  $\Delta \log R$  برای تخمین مقدار TOC با استفاده از چاه نمودارهای گاما، مقاومت و صوتی. برهم نهی چاه نمودارهای مقاومت و صوتی مقدار جدایش  $\Delta \log R$  را در مقاطع غنی از مواد آلی بدست می‌دهد. (Passey, Creaney et al. 1990)



شکل ۱-۲ نمودار تبدیل  $\Delta \log R$  به TOC با توجه به درجه بلوغ به همراه نمودارهای تعیین درجه بلوغ با استفاده از مقادیر TOC و  $S_2$  با توجه به نوع کروژن (کروژن نوع ۲ شکل سمت چپ و کروژن نوع ۳ شکل سمت راست)

پس از آن تلاش برای معرفی و ابداع روش‌های هوشمند برای تشخیص TOC از روی چاه نمودارهای پتروفیزیکی انجام شد. هانگ<sup>۱</sup> و ویلیامسون<sup>۲</sup> (Huang and Williamson 1994), Williamson 1996) با استفاده از شبکه عصبی روشی برای مدل سازی محتوای کربن آلی با استفاده از چاه نمودارهای پرتو گاما، صوتی و مقاومت ویژه ارائه کردند.

<sup>1</sup> Huang  
<sup>2</sup> Williamson

در سال ۲۰۰۴ آقای کمالی و میرشادی (Kamali and Allah Mirshady 2004) با استفاده از دو روش  $\Delta \log R$  و نرو فازی محتوای کربن آلی را در دو چاه واقع در سازندهای پابده و گورپی به وسیله چاه نمودارهای معمول (صوتی، چگالی، نوترون و مقاومت) تخمین زدند. آقای کدخدایی و همکاران در سال ۲۰۰۹ با ترکیب سیستم‌های هوشمند برای تخمین مقدار TOC به وسیله داده‌های پتروفیزیکی روشی ارائه دادند (Kadkhodaie-Ilkhchi, Rahimpour-Bonab et al. 2009). (Kadkhodaie-Ilkhchi, Rezaee et al. 2009).

## ۳-۱ لولیموت

روش استفاده شده برای تخمین پارامتر TOC در پایان نامه حاضر الگوریتم لولیموت<sup>۱</sup> است که تا به حال از این الگوریتم در کارهای مربوط به علوم زمین استفاده نشده است و هدف از انجام این پایان نامه وارد کردن این الگوریتم در علوم زمین است. الگوریتم لولیموت ترکیبی از دو روش مدل‌سازی خطی و نرو فازی است که از جمله روش‌های هوشمند به حساب می‌آید. الگوریتم لولیموت با توجه به این نکته که مدل‌سازی خطی برای هر ناحیه با روش رگرسیون خطی انجام می‌شود، در روند آموزش سریع است. سرعت بالای آموزش الگوریتم لولیموت یکی از محاسن این روش است.

## ۴-۱ معرفی فصول پایان نامه

در فصل اول این پایان نامه کلیاتی در مورد موضوع پایان نامه گفته شده است. پس از آن روش‌های انجام شده برای تخمین مقدار TOC و کارهای انجام شده در این زمینه آورده شده است. در ادامه الگوریتم مورد استفاده توضیح داده شده است.

---

<sup>1</sup> Lolimot (LOcal LInear MOdel Tree)

در فصل دوم در مورد روش‌های مورد استفاده در پایان نامه توضیح داده شده‌اند. روش‌های استفاده شده برای خوشه‌بندی داده‌ها عبارتند از روش میانگین  $K$ ، روش گوستافسون-کسل و الگوریتم DBSCAN<sup>1</sup> که مقایسه‌ای بین این روش‌ها انجام شده است و روش بهینه انتخاب شده است. روش‌های دیگری که استفاده شده است عبارتند از روش‌های زمین آماری برای مدل‌سازی TOC منطقه و روش لولیموت برای تخمین TOC که توضیح داده شده‌اند.

در فصل سوم داده‌های مورد استفاده مورد بررسی قرار گرفته‌اند. منطقه داده‌های موجود و نحوه انتخاب داده‌ها از بین تعداد بسیار زیاد داده آورده شده است. در کنار داده‌های واقعی، داده‌های مصنوعی ساخته شده برای سنجش روش‌ها نیز آورده شده است.

در فصل چهارم تخمین و مدل‌سازی انجام شده بر روی داده‌ها آورده شده است و در فصل پنجم نتیجه‌گیری‌ها و پیشنهادات در راستای انجام پایان نامه آورده شده است.

---

<sup>1</sup> Density Based Spatial Clustering of Application with Noise

# فصل دوم

## روش‌شناسی تحقیق



## ۲ روش شناسی تحقیق

### ۱-۲ مقدمه

در فصل حاضر روش‌های مورد استفاده برای خوشه‌بندی و کلاسه‌بندی داده‌ها، تخمین و مدل‌سازی آورده شده است. در بخش‌های اول و دوم روش‌های خوشه‌بندی میانگین K و گوستافسون-کسل برای مقایسه با روش اصلی آورده شده است. در بخش سوم روش خوشه‌بندی و کلاسه‌بندی DBSCAN توضیح داده شده است. برای بخش مدل‌سازی از روش زمین آماری استفاده شده است که توضیح آن در بخش چهارم آورده شده است و در نهایت در بخش پنجم روش لولیموت که برای تخمین مورد استفاده قرار گرفته است، آورده شده است.

### ۲-۲ روش میانگین K

این روش علی‌رغم سادگی آن یک روش پایه برای بسیاری از روش‌های خوشه‌بندی دیگر (مانند خوشه‌بندی فازی) محسوب می‌شود. این روش، روشی انحصاری و مسطح محسوب می‌شود (Alpaydin 2004). برای این الگوریتم شکل‌های مختلفی بیان شده است. ولی همه آن‌ها دارای روالی تکراری هستند که برای تعدادی ثابت از خوشه‌ها سعی در تخمین موارد زیر دارند:

- بدست آوردن نقاطی به عنوان مراکز خوشه‌ها، این نقاط در واقع همان میانگین نقاط متعلق به هر خوشه هستند.
- نسبت دادن هر نمونه داده به یک خوشه که آن داده کمترین فاصله تا مرکز آن خوشه را دارا باشد.

در نوع ساده‌ای از این روش ابتدا به تعداد خوشه‌های مورد نیاز نقاطی به صورت تصادفی انتخاب می‌شود. سپس داده‌ها با توجه به میزان نزدیکی (شباهت) به یکی از این خوشه‌ها نسبت داده می‌شوند و بدین ترتیب خوشه‌های جدیدی حاصل می‌شود. با تکرار همین روال می‌توان در هر تکرار با میانگین‌گیری از داده‌ها مراکز جدیدی برای آن‌ها محاسبه کرد و مجدداً داده‌ها را به خوشه‌های جدید نسبت داد. این روند تا زمانی ادامه پیدا می‌کند که دیگر تغییری در داده‌ها حاصل نشود. تابع زیر به عنوان تابع هدف مطرح است.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad ۱-۲$$

که  $\| \cdot \|$  معیار فاصله بین نقاط و  $c_j$  مرکز خوشه  $j$ ام است.

الگوریتم زیر الگوریتم پایه برای این روش محسوب می‌شود:

الف) در ابتدا  $K$  نقطه به عنوان مراکز خوشه‌ها انتخاب می‌شوند.

ب) هر نمونه داده به خوشه‌ای که مرکز آن خوشه کمترین فاصله تا آن داده را داراست، نسبت داده می‌شود.

پ) پس از تعلق تمام داده‌ها به یکی از خوشه‌ها، برای هر خوشه یک نقطه جدید به عنوان مرکز محاسبه می‌شود. (میانگین نقاط متعلق به هر خوشه)

ت) مراحل «ب» و «پ» تکرار می‌شوند تا زمانی که دیگر هیچ تغییری در مراکز خوشه‌ها حاصل نشود.

## ۳-۲ روش گوستافسون-کسل

گوستافسون-کسل الگوریتم میانگین فازی C را با استفاده از ضابطه‌ی فاصله‌ی قابل تطبیق بسط داده‌اند تا بتوان اشکال مختلف هندسی را در یک مجموعه‌ی داده، خوشه‌بندی کرد. هر خوشه ماتریس اندیس ضابطه‌ی  $A_i$  دارد که ضابطه داخلی زیر را می‌دهد:

$$D_{ikA}^2 = (x_k - v_i)^T A_i (x_k - v_i), \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad 2-2$$

ماتریس  $A_i$  به عنوان متغیر بهینه‌سازی در تابع میانگین فازی C استفاده می‌شود. بنابراین به هر خوشه اجازه می‌دهند تا ضابطه فاصله را به ساختار توپولوژیکی محلی داده‌ها نسبت بدهد.  $A$  مشخص کننده‌ی  $c$  گانه‌ی ماتریس اندیس ضابطه‌ی است:  $A = [A_1, A_2, \dots, A_c]$  عملکرد اصلی الگوریتم گوستافسون-کسل به صورت زیر تعریف می‌شود:

$$J(X; U, V, A) = \sum_{i=1}^c \sum_{k=1}^N [\mu_{ik}]^m D_{iKA}^2 \quad 3-2$$

برای یک  $A$  ثابت، شرایط فازی می‌تواند به طور مستقیم استفاده شود. تابع اصلی نمی‌تواند به طور مستقیم متناسب با  $A_i$  کمینه گردد چرا که رفتار آن با  $A_i$  به طور خطی است. این بدان معنی است که می‌توان  $J$  را به راحتی با تبدیل کردن  $A_i$  به یک مقدار معین مثبت کوچک‌تر، به هر اندازه‌ی مد نظر کوچک نمود. برای رسیدن به جوابی مطلوب،  $A_i$  بایستی به گونه‌ای محدود گردد. راه معمول نیل به این مقصود، محدود نمودن دترمینان  $A_i$  است. با اجازه دادن به ماتریس  $A_i$  برای تغییر با دترمینان خودش، پاسخ‌ها را نسبت به بهینه‌سازی شکل خوشه ثابت می‌کند (Gustafson and Kessel 1978).

$$\|A_i\| = \rho_i, \quad \rho > 0 \quad 4-2$$

که در آن  $\rho_i$  برای هر خوشه ثابت است. با استفاده از روش ضرایب لاگرانژ، ۲-۵ را برای  $A_i$  بدست می‌دهد:

$$A_i = [\rho_i \det[F_i]]^{\frac{1}{n}} F_i^{-1} \quad 5-2$$

که در آن  $F_i$  ماتریس کوواریانس فازی نامین خوشه بوده و به صورت زیر تعریف می‌شود:

$$F_i = \frac{\sum_{k=1}^N (\mu_{ik})^m [x_k - v_i][x_k - v_i]^T}{\sum_{k=1}^N [\mu_{ik}]^m} \quad 6-2$$

دقت شود که با جانشینی 5-2 و 6-2 در 2-2، اندازه‌ی مربع فاصله‌ی ماهالونوبیس کلی بین  $x_k$  و  $v_i$  میانگین خوشه را بدست می‌دهد در حالی که کوواریانس با درجات عضویتی در  $U$  وزن دار شده‌اند.

در استفاده از الگوریتم گوستافسون-کسل باید موارد زیر را در نظر گرفت:

الف) اگر هیچ دانش اولیه ای موجود نیست،  $\rho_i$  برای هر خوشه برابر 1 است بنابراین الگوریتم

گوستافسون-کسل تنها می‌تواند خوشه‌هایی را بیابد که تقریباً حجم‌های برابر دارند.

ب) در حالتی که خوشه‌ها به شدت در مسیر بزرگ‌ترین مقادیر ویژه پخش شده‌اند، ماتریس

کوواریانس محاسبه شده نمی‌تواند توزیع داده‌های زیرین را تخمین بزند.

## 2-3-1 الگوریتم گوستافسون-کسل

مجموعه داده‌ی  $X$  را بگیر، تعداد خوشه‌ها  $1 < c < N$ ، توان وزندهی  $m > 1$ ، شرط خاتمه  $\epsilon > 0$  و ماتریس

اندیس ضابطه‌ی  $A$  را انتخاب کن. ماتریس تقسیم بندی را به صورت تصادفی شروع کن.

برای  $l=1, 2, \dots$  تکرار شود:

گام اول: محاسبه مراکز خوشه‌ها

$$v_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m x_k}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, \quad 1 \leq i \leq c \quad 7-2$$

گام دوم: محاسبه ماتریس کوواریانس خوشه‌ها

$$F_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{[l-1]})^m (x_k - v_i^{(l)}) (x_k - v_i^{(l)})^T}{\sum_{k=1}^N (\mu_{ik}^{[l-1]})^m}, \quad 1 \leq i \leq c \quad ۸-۲$$

یک ماتریس یکه با مقیاس مناسب اضافه می شود

$$F_i = (1-\gamma) F_i + \gamma [F_0]_i^{-1} I \quad ۹-۲$$

مقادیر ویژه  $\lambda_{ij}$  و بردارهای ویژه  $\phi_{ij}$  استخراج می شوند،  $\lambda_{i,\max} = \max_j \lambda_{ij}$  را یافته و تنظیم می شود

$$\lambda_{i,\max} = \lambda_{ij} / \beta, \quad \forall j \quad ۱۰-۲$$

که در آن

$$\lambda_{i,\max} / \lambda_{ij} \geq \beta \quad ۱۱-۲$$

با استفاده از رابطه ۲-۱۵  $F_i$  بازسازی می شود:

$$F_i = [\phi_{i,1} \dots \phi_{i,n}] \text{diag}(\lambda_{i,1} \dots \lambda_{i,n}) [\phi_{i,1} \dots \phi_{i,n}] \quad ۱۲-۲$$

گام سوم: محاسبه ی فواصل

$$D_{ikA_i}^2(x_k, v_i) = (x_k - v_i^{(l)})^T \left[ (\rho_i(F_i))^{-1} F_i^{-1} \right] (x_k - v_i^{(l)}) \quad ۱۳-۲$$

گام چهارم ماتریس تقسیم بندی را تا زمانی که  $\|U^{[1]} - U^{[l-1]}\| < \epsilon$  به روز رسانی می شود.

$$\mu_{ik}^{[l]} = \frac{1}{\sum_{j=1}^c [D_{ikA_i} [x_k, v_i] / D_{jk} [x_k, v_j]]^{2/[m-1]}}, \quad 1 \leq i \leq c, 1 \leq k \leq N \quad ۱۴-۲$$

## ۴-۲ معرفی روش DBSCAN

در اغلب علوم مهندسی، خوشه‌بندی<sup>۱</sup> و کلاسه‌بندی<sup>۲</sup> داده‌ها گامی مهم در پردازش داده‌ها می‌باشد (Matheus, Chan et al. 1993). برای خوشه‌بندی داده‌ها روش‌های بسیاری مانند میانگین K<sup>۳</sup> (O.Duda, E.Hart et al. 2003)، گوستافسون-کسل<sup>۴</sup> (Balasko, Abonyi et al. 2007) و... پیشنهاد شده و توسعه یافته است که هر یک دارای معایب و محاسنی هستند. همه روش‌های خوشه‌بندی زیر مجموعه دو حالت کلی هستند: تقسیم بندی و سلسله مراتبی (Kaufman and Rousseeuw 1990).

روش استفاده شده در نوشته حاضر برای خوشه‌بندی و کلاسه‌بندی داده‌ها روش خوشه‌بندی فضایی بر پایه تراکم داده‌های دارای نوفه (DBSCAN) است که در این بخش به معرفی و توضیح این روش پرداخته می‌شود.

## ۱-۴-۲ ساختار الگوریتم DBSCAN

الگوریتم DBSCAN نخستین بار توسط مارتین استر و همکاران معرفی شد (Ester, Kriegel et al. 1996). این الگوریتم روشی برای خوشه‌بندی است که بر پایه چگالی داده‌ها عمل می‌کند. در این روش به منظور تخمین چگالی توزیع نقاط از دو پارامتر شعاع همسایگی و حداقل نقاط مورد نیاز برای تشکیل یک خوشه استفاده می‌شود که پارامتر شعاع همسایگی را با  $\epsilon$  و پارامتر دوم یعنی حداقل نقاط مورد نیاز برای تشکیل یک خوشه را با  $\text{minPts}$  نشان می‌دهند.

---

<sup>1</sup> Clustering  
<sup>2</sup> Classification  
<sup>3</sup> K-means  
<sup>4</sup> Gustafson-Kessel

ابتدا فاصله تمام نقاط از یکدیگر محاسبه می‌شود تا در گام‌های بعدی نیازی به محاسبه چندین باره فاصله نقاط نباشد و در زمان مورد نیاز فاصله‌ها از جدولی که ابتدا محاسبه شده خوانده شود. پس از آن الگوریتم از یک نقطه اختیاری شروع می‌شود. سپس تعداد نقاطی که در همسایگی این نقطه هستند و فاصله آن‌ها کمتر از  $\epsilon$  است، شمارش می‌شوند. اگر تعداد نقاط بیشتر از پارامتر  $\minPts$  بود شرایط ایجاد یک خوشه برقرار است و یک خوشه جدید ایجاد می‌شود و تمام نقاط در این همسایگی در یک خوشه قرار می‌گیرند. در غیر این صورت نقطه مورد بررسی به عنوان نوفه شناخته می‌شود.

پس از تعیین اولین نقطه همین کار برای نقاط بعدی تکرار می‌شود. یعنی یک نقطه انتخاب می‌شود و تعداد نقاط در همسایگی شمارش می‌شود اگر تعداد نقاط بیشتر از پارامتر  $\minPts$  بود خوشه تشکیل می‌شود و اگر کمتر از پارامتر  $\minPts$  بود خوشه ایجاد نمی‌شود. نکته مهم این است که اگر شرایط ایجاد یک خوشه برقرار بود و در همسایگی نقطه مورد نظر نقطه ای قبلاً خوشه‌بندی شده بود شاخص جدیدی برای این خوشه ایجاد نمی‌شود و نقطه مورد بررسی به همان خوشه اضافه می‌شود و دارای شاخص آن خوشه می‌شود ولی اگر هیچ یک از نقاط در همسایگی خوشه‌بندی نشده بودند خوشه جدیدی ایجاد می‌شود و تمام نقاط در این خوشه قرار می‌گیرند. به این ترتیب از ازدیاد تعداد خوشه‌ها جلوگیری می‌شود.

نکته مهم دیگری که در این روش وجود دارد این است که به علت بررسی چندین باره نقاط شاید هر نقطه ای در ابتدا نوفه شناخته شود ولی در گام‌های بعدی امکان دارد به عنوان قسمتی از یک خوشه دسته بندی شود و حسن دیگر این روش امکان تشخیص و تفکیک نوفه از دیگر داده‌ها است.

## ۲-۴-۱-۲ مثالی از الگوریتم DBSCAN

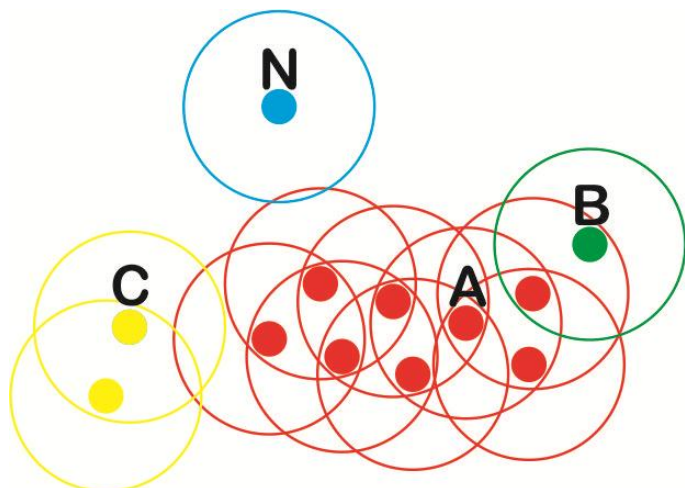
مراحل کار الگوریتم در شکل ۱-۲ بهتر دیده می‌شود. فرض می‌شود که نقطه A برای شروع الگوریتم انتخاب شود (با توجه به عدم وابستگی به ترتیب نقاط در این الگوریتم این فرض اشتباه نیست). با توجه به این که تعداد نقاط در همسایگی نقطه A بیشتر از معیار minPts است (در اینجا مقدار minPts برابر ۳ است) نقطه A به عنوان یک نقطه مرکزی انتخاب می‌شود. چون نقاطی که در همسایگی این نقطه هستند هیچ یک شاخصی از یک خوشه ندارند در نتیجه یک خوشه جدید تشکیل می‌شود (خوشه قرمز رنگ).

پس از آن دیگر نقاط نیز به همین ترتیب انتخاب می‌شود و الگوریتم ادامه می‌یابد. اگر نقاط بعدی که انتخاب می‌شوند نقاط در همسایگی نقطه A باشند با توجه به اینکه در همسایگی نقطه A هستند که خوشه‌بندی شده پس آن‌ها هم به همین خوشه اضافه می‌شوند.

پس از ادامه روند وقتی نوبت به بررسی نقطه B می‌رسد اتفاقی که رخ می‌دهد این است که تعداد نقاط در همسایگی کمتر از تعداد minPts است پس نقطه B یک نقطه مرکزی نیست. اما با توجه به اینکه نقطه ای که در شعاع همسایگی نقطه B است قبلاً خوشه‌بندی شده است این نقطه نیز به آن خوشه اضافه می‌شود و در خوشه نقاط قرمز قرار می‌گیرد (نقطه B به علت بهتر دیده شدن و تفاوت آن با دیگر نقاط این خوشه سبز رنگ شده است).

در ادامه بررسی نوبت به نقطه C می‌رسد. تعداد نقاط در همسایگی نقطه C کمتر از minPts است و چون هیچ یک از نقاط در همسایگی این نقطه خوشه‌بندی نشده‌اند پس این نقطه نیز خوشه‌بندی نمی‌شود و به عنوان نوفه در نظر گرفته می‌شود (نقطه C و نقطه دیگر در همسایگی این نقطه که زرد رنگ هستند هر دو نوفه هستند). نقاطی مانند نقطه N که هیچ نقطه ای در همسایگی خود ندارند به عنوان نوفه شناخته می‌شوند.





شکل ۱-۲ روش کار الگوریتم DBSCAN با توجه به تراکم نقاط برای خوشه بندی نقاط مختلف

### ۳-۱-۴-۲ الگوریتم DBSCAN

در این بخش الگوریتم DBSCAN برای خوشه بندی مرحله به مرحله که به صورت زیر آورده می شود (Ester, Kriegel et al. 1996):

- الگوریتم DBSCAN با ورودی های  $\epsilon$  و  $MinPts$  شروع شود.
- برای تمام نقاط موجود در  $SetOfPoints$  حلقه زیر انجام شود.
  - متغیر  $Point$  برابر با نقطه جاری از  $SetOfPoint$  قرار گیرد.
  - اگر نقطه  $Point$  خوشه بندی نشده بود آنگاه:
    - اگر خروجی تابع  $ExpandCluster$  به ازای ورودی های  $Point$ ،  $SetOfPoints$ ،  $ClusterId$ ،  $\epsilon$  و  $MinPts$  صحیح بود آنگاه مقدار  $ClusterId$  یک واحد افزایش یابد.
    - پایان شرط
      - پایان شرط
- پایان حلقه

در الگوریتم آورده شده SetOfPoint اطلاعات نقاط موجود برای خوشه‌بندی است که شامل مختصات نقاط و شاخص خوشه‌بندی آنها است که با ClusterId نشان داده می‌شود و در ابتدا برای تمام نقاط برابر خوشه‌بندی نشده است. تابع مهمی که در این الگوریتم به کار رفته تابع ExpandCluster است. این تابع پارامترهای SetOfPoints، Point، ClusterId،  $\epsilon$  و MinPts را به عنوان ورودی دریافت می‌کند و یک مقدار صحیح یا غلط را باز می‌گرداند. در این تابع ابتدا نقاط موجود در همسایگی نقطه Point که در شعاع کمتر از  $\epsilon$  هستند شمرده می‌شوند. اگر تعداد نقاط کمتر از MinPts بود نقطه به عنوان نوفه علامت گذاری می‌شود و مقدار غلط به تابع ExpandCluster باز گردانده می‌شود و کار تابع تمام می‌شود. ولی اگر تعداد نقاط بیشتر از MinPts بود نقطه Point با عنوان ClusterId علامت گذاری می‌شود و تمام نقاطی که در این بازه هستند نیز با عنوان ClusterId علامت گذاری می‌شوند و برای تمام نقاطی که به صورت زنجیر در محدوده نقاط علامت گذاری شده هستند نیز این کار ادامه پیدا می‌کند و در نهایت مقدار صحیح به تابع باز گردانده می‌شود. الگوریتم تابع ExpandCluster به صورت زیر است:

```
ExpandCluster (SetOfPoints, Point, CId, Eps, minPts) : Boolean;
  seeds :=SetOfPoints.regionQuery (Point,Eps);
  IF seeds.size < minPts THEN // no core point
    SetOfPoint.changeCId (Point,NOISE);
    RETURN False;
  ELSE // all points in seeds are density reachable from Point
    SetOfPoints.changeCIds (seeds,CId);
    seeds.delete (Point);
    WHILE seeds <> Empty DO
      currentP := seeds.first();
      result :=SetOfPoints.regionQuery (currentP,Eps);
      IF result.size >= minPts THEN
        FOR i FROM 1 TO result.size DO
          resultP :=result.get (i);
```

```

IF resultP.ClId IN (UNCLASSIFIED, NOISE) THEN
    IF resultP.ClId = UNCLASSIFIED THEN
        seeds.append (resultP);
    END IF;
    SetOfPoints.changeClId (resultP,ClId);
END IF; // UNCLASSIFIED or NOISE
END FOR;
END IF; // result.size >= minPts
seeds.delete (currentP);
END WHILE; // seeds <> Empty
RETURN True;
END IF
END; // ExpandCluster

```

## ۲-۴-۲ الگوریتم DBSCAN برای کلاسه‌بندی

روش کلاسه‌بندی ارائه شده، بر پایه روش خوشه‌بندی DBSCAN که در بخش قبل توضیح داده شد ایجاد شده و گسترش یافته است که در این بخش به توضیح این روش پرداخته می‌شود. برای کلاسه‌بندی، داده‌های ورودی عبارتند از: مجموعه نقاطی که کلاس آن‌ها معلوم است و مجموعه نقاطی که کلاس آن‌ها معلوم نیست که هدف الگوریتم ارائه شده کلاسه‌بندی این نقاط است و پارامتر  $\epsilon$  که شعاع جستجو است. برای هر نقطه از مجموعه نقاط هدف برای کلاسه‌بندی مراحل زیر انجام می‌شود:

- نقاط موجود در فاصله  $\epsilon$  نقطه هدف از مجموعه نقاط ورودی کلاس معلوم را تعیین کن و در مجموعه points قرار بده.
- برای هر کلاس عملیات زیر را انجام بده

○ مقدار  $f_j = \frac{1}{\sum_{i=1}^m d_i^2}$  را محاسبه کن که در آن  $j$  شماره کلاس،  $d$  فاصله نقاط و  $m$

تعداد نقاط موجود در مجموعه points که در کلاس  $j$  است.

- وزن فازی میزان تعلق نقطه مورد نظر به هر کلاس از رابطه  $fuzzy\_weight_j = \frac{f_j}{\sum_{i=1}^c f_i}$

محاسبه می‌شود که در آن  $c$  تعداد کلاس‌های مجموعه ورودی‌هاست.

پس از طی مراحل بالا درصد تعلق نقطه مورد نظر به هر کلاس تعیین می‌شود و برای رسیدن

به یک کلاس بزرگ‌ترین مقدار انتخاب می‌شود. این الگوریتم برای تمام نقاط انجام می‌شود.

## ۵-۲ الگوریتم لولیموت

الگوریتم لولیموت اولین بار توسط آقای الیور اُنیلز<sup>۱</sup> در راستای پایان نامه دکترای ایشان در سال ۱۹۹۹ مطرح شد (Nelles 1999). الگوریتم درخت مدل خطی محلی<sup>۲</sup> (که به اختصار الگوریتم لولیموت گفته می‌شود) یک الگوریتم فزاینده با ساختاری درختی است (Nelles 2000). الگوریتم لولیموت بر پایه رگرسیون چند متغیره و ساختار فازی عمل می‌کند که عملکرد آن در قسمت‌های بعدی مورد بررسی قرار می‌گیرد. چون این الگوریتم برای هر مدل یک نرون جدید می‌سازد از این جهت می‌توان آن را یک ساختار نرو فازی<sup>۳</sup> دانست.

در این بخش ابتدا ساختار الگوریتم لولیموت توضیح داده می‌شود که شامل توضیحی برای مدل خطی و نحوه وزن دهی به مدل‌ها است. پس از آن توضیح مختصری درباره روند کار داده شده است. در بخش بعد الگوریتم لولیموت آورده شده است و روند کار الگوریتم بررسی شده است. پس از شرح

<sup>1</sup> Oliver O.Nelles

<sup>2</sup> local liner model tree (LOLIMOT)

<sup>3</sup> neuro-fuzzy

ساختار الگوریتم، مثالی برای روشن شدن کارکرد الگوریتم آورده شده است. در ادامه روش ساخت یک مدل خطی آمده است و در پایان انواع شروط توقف الگوریتم نیز آورده شده است.

## ۱-۵-۲ ساختار الگوریتم لولیموت

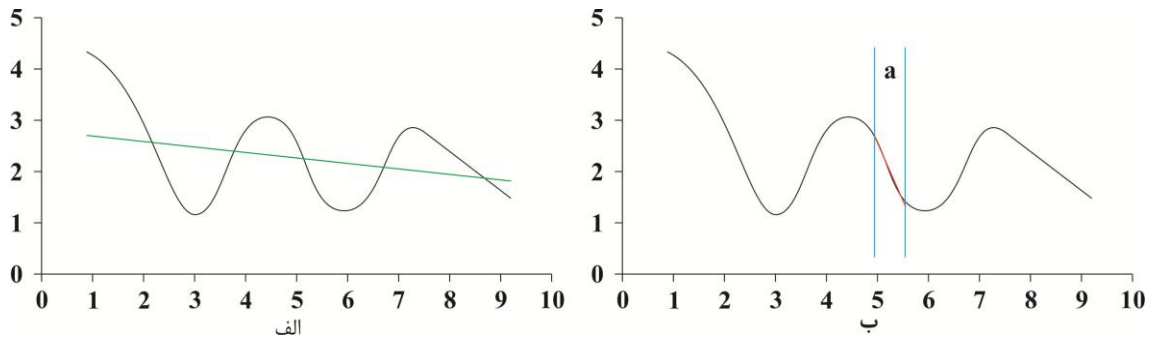
همان‌طور که قبلاً گفته شد الگوریتم لولیموت ساختاری فزاینده دارد؛ بدین معنی که ساختار ابتدا از یک مدل ساده شروع می‌شود و به تدریج رشد می‌کند (مانند یک درخت) و تعداد مدل‌ها افزایش می‌یابد. در این الگوریتم مدل‌های جدید با استفاده از تقسیمات متوالی بدست می‌آیند. الگوریتم لولیموت از دو قسمت عمده تشکیل شده است: الف) مدل خطی محلی<sup>۱</sup>، ب) وزن دهی فازی مدل‌ها.

### ۱-۱-۵-۲ مدل خطی محلی

مدل‌های پیچیده غیر خطی را می‌توان با استفاده از مدل‌های ساده خطی تقریب زد اما مشکل این تقریب خطای بالای این کار می‌باشد. حال اگر فضای مد نظر برای تقریب کوچک‌تر شود مقدار خطای ناشی از تقریب در فضای کوچک‌تر در اکثر موارد کاهش می‌یابد و مدل خطی به شکل تابع نزدیک‌تر می‌شود. شکل ۲-۲ این مطلب را که چگونه یک مدل غیرخطی را می‌توان با یک تابع خطی تقریب زد نشان می‌دهد.

---

<sup>1</sup> local liner model (LLM)

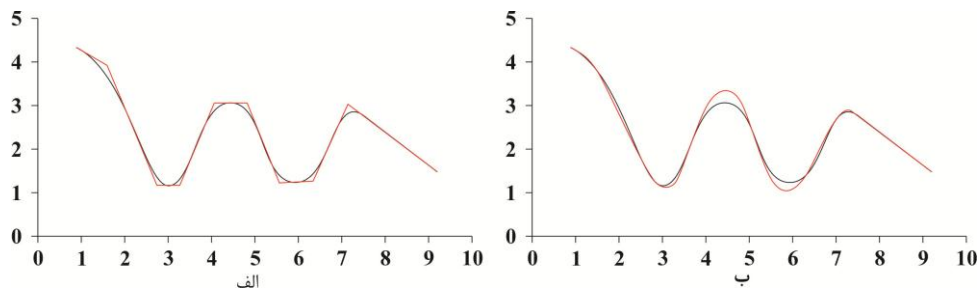


شکل ۲-۲ الف) تخمین خطی مدل غیرخطی در حالت کلی برای مدل، ب) تخمین خطی مدل غیرخطی در منطقه محدود a

همان‌طور که در شکل ۲-۲ دیده می‌شود وقتی یک مدل را بخواهیم با یک تابع خطی تقریب بزنیم مقدار خطا در اکثر موارد زیاد خواهد بود. اما وقتی ناحیه محدودی از یک مدل را بخواهیم با یک تابع خطی مقدار خطا کاهش خواهد یافت. برای دستیابی به دقت بالا در تخمین باید مدل به قسمت‌های کوچک‌تری تقسیم شود و در نتیجه تعداد توابع خطی افزایش خواهد یافت. مشکلی که با افزایش تعداد نواحی به وجود می‌آید این است که مدل ارائه شده جامعیت خود را از دست داده و دیگر قابلیت تعمیم نخواهد بود. در شکل ۲-۲ سمت راست مدلی که برای ناحیه a ارائه شده یک مدل خطی محلی (LLM) است.

### ۲-۱-۵-۲ وزن دهی فازی مدل‌ها

وقتی یک مدل که از چندین LLM تشکیل شده برای تخمین یک پدیده ارائه می‌شود؛ زمانی که محدوده اعمال یک LLM تمام شده و نوبت به اعمال LLM بعدی می‌شود نقطه شکستگی به وجود می‌آید که نامطلوب است و مقدار خطا را افزایش می‌دهد. برای تصحیح این مورد به جای اینکه LLMها به صورت مطلق وزن دهی شوند، از وزن دهی فازی استفاده می‌شود. این موضوع در شکل ۳-۲ نشان داده شده است.



شکل ۲-۳ الف) وزن دهی مدل به صورت مطلق، ب) وزن دهی مدل به صورت فاززی

همان‌طور که در شکل ۲-۳ قسمت الف دیده می‌شود پدیده موجود با ۹ تابع خطی مدل شده است که در نواحی نزدیک به محدوده شکست‌ها خطای مدل افزایش یافته است که این وضعیت به علت وزن دهی مدل به صورت مطلق به وجود آمده است. در شکل ۲-۳ قسمت ب همان مدل با همان توابع ولی با وزن دهی فاززی دیده می‌شود. در این قسمت شکستگی‌های موجود در قسمت قبل از بین رفته و در مواردی خطا کاهش یافته است. وزن دهی فاززی مدل باعث بهبود نتیجه می‌شود و حتی امکان کاهش تعداد توابع را فراهم می‌کند که این وضعیت مطلوب است؛ زیرا امکان تعمیم مدل را افزایش می‌دهد.

## ۲-۵-۲ روند کار الگوریتم لولیموت

روند کار الگوریتم لولیموت به این صورت است که ابتدا برای پدیده مد نظر یک مدل خطی کلی<sup>۱</sup> که در تمام پدیده از آن استفاده می‌شود، ساخته می‌شود. سپس مدل ساخته شده در تمام ابعاد به دو نیم تقسیم می‌شود و در قسمت‌های به وجود آمده LLMها جدید محاسبه می‌شود. پس از آن بهترین تقسیم بندی (دارای کمترین خطا در کل مدل) انتخاب شده و کار ادامه پیدا می‌کند. سپس بدترین LLM (دارای بیشترین خطا در محدوده تأثیر) انتخاب شده و در تمام ابعاد به دو نیم تقسیم می‌شود و بهترین تقسیم بندی انتخاب می‌شود و همین روند تا رسیدن به شرط توقف الگوریتم ادامه پیدا می‌کند. شرط توقف الگوریتم می‌تواند به چند صورت تعیین شود. یک نوع شرط توقف رسیدن به حداکثر تعداد مدل محلی است. نوع دیگری از شرط توقف رسیدن به میزان مطلوبی

<sup>۱</sup> Global liner model

از خطا است که در کل مدل محاسبه می‌شود و انواع دیگری از شرط توقف که می‌توان از آن به روند کاهش خطا و ... اشاره کرد که در پایان نامه حاضر تنها از دو نوع شرط توقف اول استفاده می‌شود.

## ۲-۵-۳ الگوریتم لولیموت

الگوریتم لولیموت شامل یک حلقه خارجی است که در آن ساختار مدل تعیین می‌شود و یک حلقه تودرتوی داخلی که در آن ضرایب ثابت پارامترها با تخمین محلی بهینه می‌شود.

**الف) با یک مدل اولیه شروع کن:** تابع اعتبار را برای فضاهای اولیه ورودی بساز و پارامترهای LLM را با روش کمترین مربعات وزن دهی شده محلی تعیین کن. مقدار  $M$  را برابر تعداد LLM های اولیه قرار بده. اگر تقسیم بندی اولیه در فضا وجود نداشت  $M=1$  قرار بده و با یک LLM شروع کن که در واقع یک مدل کلی است و تابع اعتبار برای این LLM چون در تمام فضا استفاده می‌شود برابر  $\Phi_1(\underline{u})=1$  است.

**ب) بدترین LLM را پیدا کن:** تابع خطای محلی را برای LLM ها به ازای هر  $i=1, \dots, M$  محاسبه کن. تابع خطای محلی را می‌توان با وزن دهی مربع خطای مدل به وسیله تابع اعتبار بر طبق رابطه زیر محاسبه کرد:

$$I_i = \sum_{j=1}^N e^{-2} (j) \Phi_i(\underline{u}(j)) \quad ۱۵-۲$$

بدترین LLM دارای بیشترین مقدار  $I$  می‌باشد.

**ج) تمام تقسیمات را بررسی کن:** LLM،  $l$ ام برای تغییرات بعدی بررسی می‌شود. فرا سطح مستطیل این LLM به دو نیم تقسیم می‌شود. این تقسیمات برای تمام ابعاد انجام می‌شود. برای تمام تقسیمات در بعد  $dim=1, \dots, p$  مراحل زیر انجام می‌شود:



آ) مقدار تابع عضویت<sup>۱</sup> چند بعدی برای هر دو فرا سطح مستطیل محاسبه می‌شود.

ب) همه توابع عضویت محاسبه می‌شود.

پ) برای هر دو LLM تولید شده، ضرایب پارامتر محلی محاسبه می‌شود.

ت) مقدار تابع خطای کلی برای مدل جاری محاسبه می‌شود.

د) بهترین تقسیم بندی را پیدا کن: از بین  $p$  گزینه بررسی شده در مرحله «ج» کم خطاترین

قسمت انتخاب می‌شود. توابع اعتبار محاسبه شده در مرحله «ج آ» و LLMهای محاسبه شده در

مرحله «ج پ» برای مدل تعیین می‌شود. تعداد LLMها افزایش می‌یابد  $M \rightarrow M+1$ .

ه) همگرایی بررسی می‌شود: اگر شرایط توقف حاصل شده بود، کار را متوقف کن در غیر این

صورت برگرد به مرحله «ب».

شرایط توقف همان‌طور که در قسمت‌های قبلی گفته شد می‌تواند به چند گونه تعیین شود

مانند بیشترین تعداد مدل‌های ساخته شده و یا مقدار خطای کل مدل. مقدار خطایی که در مرحله

دوم محاسبه می‌شود، مقدار خطا برای کل مدل است و نه خطای هر یک از LLMها که محاسبه شده

است و کاربرد آن در شرط توقف است. کاربرد خطای محاسبه شده برای هر یک از LLMها برای

انتخاب بهترین و بدترین قسمت در تقسیم‌بندی‌هاست.

## ۲-۵-۳-۱ بررسی روند کار الگوریتم لولیموت

شکل ۲-۴ کارکرد الگوریتم لولیموت را در ۵ گام (تکرار) اول بهتر نشان می‌دهد. در گام اول با توجه

به اینکه فضا تقسیم بندی نشده است یک مدل اولیه ساخته می‌شود که در تمام فضا مورد استفاده

است. سپس گام دوم آغاز می‌شود. در این گام، فضا در دو جهت عمود بر  $u_1$  و عمود بر  $u_2$  به صورت

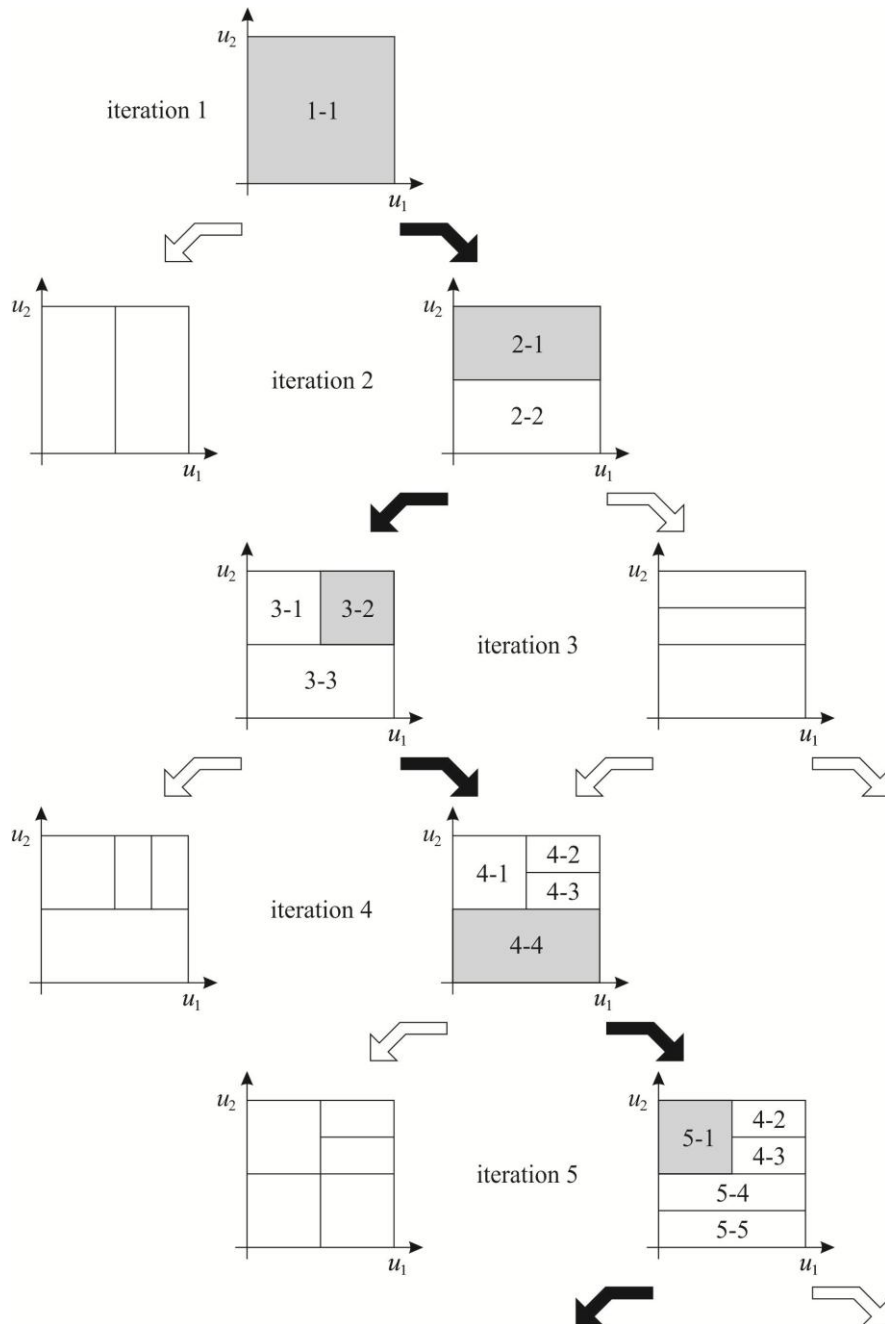
جداگانه تقسیم می‌شود و خطای مدل بدست آمده در دو حالت محاسبه می‌شود و چون مقدار خطا در

حالت سمت راست کمتر است این مدل انتخاب می‌شود. پس از این گام بدترین LLM انتخاب

---

<sup>1</sup> member ship function (MSF)

می‌شود که در این مرحله LLM شماره 1-2 است. حالا این LLM در دو جهت تقسیم می‌شود و مانند مراحل قبل خطای مدل حاصل در هر دو حالت محاسبه می‌شود که تقسیم بندی مناسب‌تر انتخاب شود. مدل بهینه در این گام مدل سمت چپ است. حالا بدترین LLM موجود انتخاب می‌شود که در این مرحله LLM شماره 2-3 بدترین مورد است. این مراحل تا جایی که به شرط توقف حاصل شود ادامه می‌یابد. در شکل فلش‌های سیاه رنگ جهت انتخاب مدل حاصل را نشان می‌دهد و LLM‌های مشخص شده با رنگ خاکستری بدترین مدل انتخاب شده در هر مرحله را مشخص می‌کنند.



شکل ۲-۴ روند کار الگوریتم لولیموت در یک محیط دو بعدی (Nelles 2000) - مناطق خاکستری نشان دهنده قسمتی است که دارای بیشترین خطا است و احتیاج به تقسیم دارد.

## ۴-۵-۲ ساختار الگوریتم لولیموت

شکل ۵-۲ ساختار الگوریتم لولیموت را نشان می‌دهد که در ادامه به توضیح هر یک از بخش‌های این ساختار پرداخته می‌شود. هر نرون ابتدا بردار ورودی  $\underline{u}$  را دریافت می‌کند. سپس خروجی LLM با استفاده از رابطه ۱۶-۲ محاسبه می‌شود.

$$\hat{y}_i = w_{i0} + w_{i1}u_1 + w_{i2}u_2 + \dots + w_{ip}u_p = \sum_{j=1}^p w_{ij}u_j \quad ۱۶-۲$$

در رابطه ۱۶-۲  $i$  شماره نرون مربوطه و  $w$  وزنه ای خطی LLMهاست که با استفاده از روش رگرسیون خطی محاسبه می‌شوند. به علت نرمال شدن توابع اعتبار، توابع اعتبار در رابطه ۱۷-۲ صدق می‌کنند.

$$\sum_{i=1}^M \Phi_i(\underline{u}) = 1 \quad ۱۷-۲$$

مقدار توابع اعتبار از رابطه ۱۸-۲ محاسبه می‌شود.

$$\Phi_i(\underline{u}) = \frac{\mu_i(\underline{u})}{\sum_{j=1}^M \mu_j(\underline{u})} \quad ۱۸-۲$$

در رابطه ۱۸-۲ مقدار  $\mu_i$  از رابطه ۱۹-۲ محاسبه می‌شود.

$$\begin{aligned} \mu_i(\underline{u}) &= \exp\left(-\frac{1}{2}\left(\frac{(u_1 - c_{i1})^2}{\sigma_{i1}^2} + \dots + \frac{(u_p - c_{ip})^2}{\sigma_{ip}^2}\right)\right) \\ &= \exp\left(-\frac{1}{2}\frac{(u_1 - c_{i1})^2}{\sigma_{i1}^2}\right) \times \dots \times \exp\left(-\frac{1}{2}\frac{(u_p - c_{ip})^2}{\sigma_{ip}^2}\right) \end{aligned} \quad ۱۹-۲$$

در رابطه ۱۹-۲ دو متغیر  $c$  و  $\sigma$  به ترتیب بیانگر میانگین ورودی‌های هر LLM و پارامتر دوم تعیین کننده میزان تأثیر دامنه فازی تابع اعتبار است که از رابطه ۲۰-۲ بدست می‌آید.

$$\sigma_{ij} = k_{\sigma} \cdot \Delta_{ij}$$

۲۰-۲

که در آن پارامتری است که توسط کاربر تعیین می‌شود و معمولاً مطابق رابطه ۲-۲۱ است.

$$k_{\sigma} = \frac{1}{3}$$

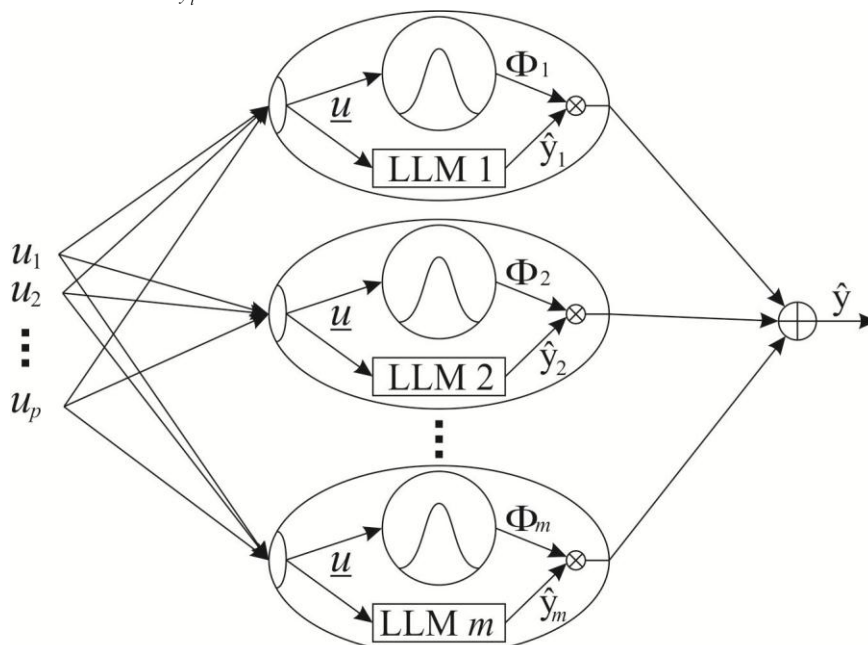
۲۱-۲

مقدار  $\Delta_{ij}$  نیز اندازه ضلع فرا سطح LLM شماره  $i$  در راستای بعد  $j$ ام است. پس از محاسبه خروجی هر

نرون خروجی نهایی مدل از رابطه ۲-۲۲ زیر محاسبه می‌شود.

$$\hat{y} = \sum_{i=1}^M \underbrace{(w_{i0} + w_{i1}u_1 + w_{i2}u_2 + \dots + w_{ip}u_p)}_{\hat{y}_i} \Phi_i(\underline{u})$$

۲۲-۲



شکل ۲-۵ ساختار الگوریتم لولیموت-هر نرون شامل یک مدل خطی و یک قسمت برای محاسبه تابع اعتبار است.

## ۲-۵-۵ روش ساخت یک LLM

برای ساخت یک LLM می‌توان از روش‌های مختلفی استفاده کرد، روش رگرسیون ساده، رگرسیون چندگانه و استفاده از شبکه عصبی از آن جمله هستند. در این نوشته روشی که برای ساخت LLMها

استفاده شده، روش رگرسیون چندگانه است. در این بخش به توضیح نحوه عملکرد این روش پرداخته می‌شود.

اگر فضای ورودی دارای  $p$  پارامتر باشد (فضای  $p$  بُعدی) و بردار ورودی  $x$  باشد که نتیجه آن خروجی  $y$  باشد می‌توان آن‌ها را به شکل زیر نشان داد:

$$\underline{x} = [1 \ x_1 \ x_2 \ \dots \ x_p] \quad , y \quad 23-2$$

حال اگر  $n$  بردار ورودی و خروجی داشته باشیم می‌توان بردارهای مربوط به ورودی و خروجی را به صورت نشان داده شده در رابطه ۲۴-۲ نمایش داد.

$$\underline{X} = \begin{bmatrix} 1 & x_1(1) & x_2(1) & \dots & x_p(1) \\ 1 & x_1(2) & x_2(2) & \dots & x_p(2) \\ 1 & x_1(3) & x_2(3) & \dots & x_p(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1(n) & x_2(n) & \dots & x_p(n) \end{bmatrix} \quad , \underline{y} = \begin{bmatrix} y(1) \\ y(2) \\ y(3) \\ \vdots \\ y(n) \end{bmatrix} \quad 24-2$$

معادله مربوط به سری نام ورودی و خروجی به صورت رابطه ۲۵-۲ خواهد بود.

$$y_i = \sum_{j=0}^n w_{ij} x_{ij} \quad \text{with } x_0 = 0 \quad 25-2$$

در رابطه ۲۵-۲  $w_{ij}$  وزن یا ضریب مربوطه است که در صورت نوشتن معادلات به صورت ماتریسی معادله ۲۶-۲ حاصل می‌شود.

$$[\underline{y}]_{n \times 1} = [\underline{X}]_{n \times p} \times [w]_{p \times 1} \quad 26-2$$

حال هدف تعیین ماتریس مربوط به وزن‌ها یا همان ماتریس  $[w]_{p \times 1}$  در معادله ۲۶-۲ است. برای

محاسبه ماتریس وزن به شکل ماتریسی می‌توان از عملیات ماتریسی آورده شده در رابطه ۲۷-۲

استفاده کرد که باعث افزایش سرعت محاسبات می‌شود:

$$[w]_{p \times 1} = \left( [\underline{X}]^T \times [\underline{X}] \right)^{-1} \times \left( [\underline{X}]^T \times [\underline{y}] \right) \quad 27-2$$

حالا با ضرب ماتریس وزن به دست آمده از رابطه ۲۷-۲ در ماتریس ورودی، ماتریس  $[\hat{y}]$  که همان ماتریس خروجی است بدست می‌آید. رابطه ۲۸-۲ این عملیات را نشان می‌دهد.

$$[\hat{y}]_{n \times 1} = [X]_{n \times p} \times [w]_{p \times 1} \quad 28-2$$

پس از محاسبه خروجی تخمین زده شده نوبت به محاسبه میزان خطای LLM ساخته شده می‌رسد. برای محاسبه مقدار خطای LLM ساخته شده از رابطه ۲۹-۲ استفاده می‌شود:

$$e = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad 29-2$$

و در اینجا عملیات مربوط به محاسبه LLM به پایان می‌رسد.

## ۶-۵-۲ بررسی انواع شرط توقف الگوریتم

شروط توقف دارای انواع مختلف است از جمله: رسیدن به تعداد معینی از مدل، رسیدن به مقدار معینی از خطا، شرط توقف زمانی و ... . در نوشته حاضر شروط توقف استفاده شده به دو شکل هستند: اول شرط توقف با رسیدن به یک تعداد معین مدل و دوم رسیدن به میزان مطلوبی از خطا که با حاصل شدن هر یک از این دو شرط الگوریتم متوقف می‌شود.

با توجه به اینکه شرط اول شمارش تعداد مدل‌های ساخته شده را شامل می‌شود، بررسی این شرط ساده است. بررسی شرط دوم یعنی رسیدن به میزان مطلوبی از نسبت به شرط اول پیچیده تر است.

با توجه به اینکه در این الگوریتم یک خطا، خطای هر LLM و خطای دیگری که وجود دارد خطای کلی مدل است، شرط توقف الگوریتم خطای نوع دوم یعنی، خطای کلی مدل است. برای محاسبه خطای کلی مدل میانگین خطا بین خروجی مدل و مقدار واقعی محاسبه می‌شود. این خطا از رابطه ۳۰-۲ بدست می‌آید.

$$e = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

۳۰-۲

خطای بدست آمده از رابطه ۳۰-۲ معیار توقف الگوریتم خواهد بود.



# فصل سوم

## معرفی و بررسی داده‌های مورد مطالعه

## ۳ داده‌ها

### ۱-۳ مقدمه

داده‌های مورد استفاده در پایان نامه حاضر به دو دسته تقسیم می‌شوند. دسته اول داده‌های واقعی که مربوط به منطقه مورد مطالعه هستند و دسته دوم داده‌های مصنوعی که برای سنجش کارایی روش‌های مورد استفاده ساخته شده‌اند. در این فصل به معرفی و بررسی داده‌های مورد استفاده پرداخته می‌شود.

### ۲-۳ داده‌های واقعی

ناحیه مورد مطالعه در این پایان نامه ناحیه پرت<sup>۱</sup> واقع در غرب استرالیا است. وسعت زیاد این منطقه و تعداد زیاد داده باعث ایجاد مشکل در انجام مراحل این پایان نامه است. در ادامه ابتدا توصیفی از منطقه و موقعیت چاه‌های انتخابی برای مطالعه آورده می‌شود. پس از آن نحوه گزینش داده‌ها و چگونگی محدود کردن داده‌ها برای ساده کردن پردازش توضیح داده می‌شود.

### ۳-۳ موقعیت منطقه

شکل ۱-۳ نشان دهنده موقعیت ناحیه پرت است. ناحیه پرت، حوزه ای رسوبی است که در منطقه‌ی جنوب غربی استرالیا واقع شده و از شمال تا جنوب منطقه کشیده شده است. رسوبات این

---

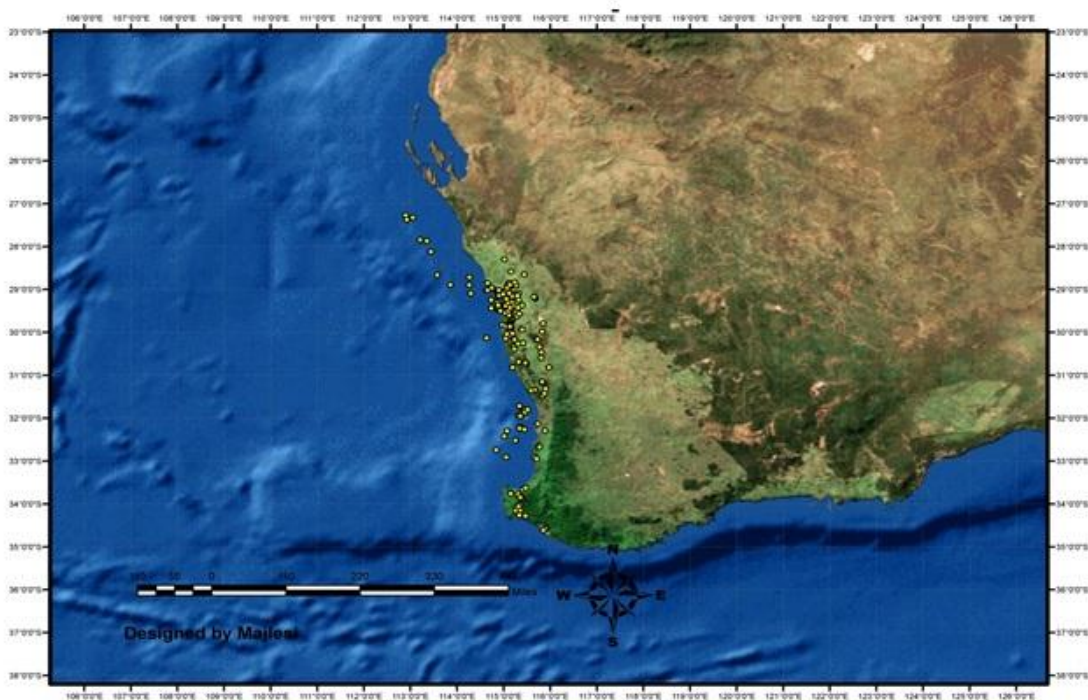
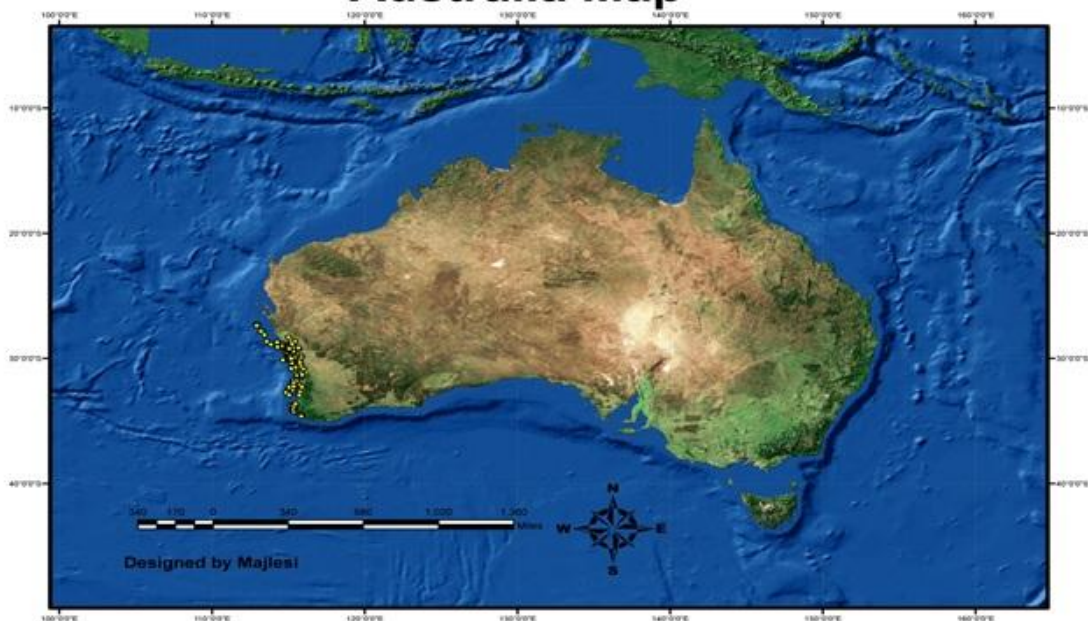
<sup>1</sup> Perth

منطقه شامل رسوبات دوره سیلورین و پلیوسن است. این منطقه از شرق به گسل دارلین محدود می‌شود و از غرب شامل ساحل تا مرز قاره ای-اقیانوسی می‌شود.

منطقه مورد مطالعه در نزدیکی قسمت مرکزی متمایل به جنوب ساحل پرت واقع شده است. موقعیت منطقه بین عرض جغرافیایی ۳۱ تا ۳۳ جنوبی و بین طول جغرافیایی ۱۱۵ تا ۱۱۶ قرار گرفته است.

مساحت منطقه مورد مطالعه ۱۳۲۵۱ کیلومتر مربع است. این منطقه شامل ۲۳ چاه است که طی مراحلی که در قسمت‌های بعدی توضیح داده می‌شود چاه‌ها محدود می‌شود تا پردازش ساده تر گردد.

## Australia Map



شکل ۱-۳ نقشه چاه‌های حفاری شده و موقعیت منطقه نسبت به استرالیا

## ۴-۳ داده‌های اولیه

داده‌هایی که برای انجام این پروژه در اختیار بود اطلاعات مربوط به تعداد تقریبی ۳۵۰ چاه بود که این مورد خود یکی از مشکلات در ادامه راه بود. مشکلی که در این باره به وجود آمد از این قرار بود که با توجه به تمرکز این پایان نامه برای تخمین ویژگی TOC لازم بود که در تعدادی از چاه‌ها این ویژگی اندازه‌گیری شده باشد. تعداد چاه‌هایی که این ویژگی در آن‌ها اندازه‌گیری شده بود ۷۷ چاه بود.

تعداد داده‌های مربوط به چاه‌ها و تعداد چاه‌هایی که ویژگی مورد نظر را دارا بودند زیاد بود ولی متأسفانه در موارد بسیاری داده‌ها با هم هماهنگی نداشتند. یعنی در برخی از چاه‌ها که داده TOC موجود بود، داده‌های مربوط به چاه نمودار مربوطه موجود نبود و یا بالعکس که این مشکل منجر به چشم پوشی از حجم عظیمی از داده‌ها شد. با توجه به مطالب گفته شده داده‌ها مورد گزینش قرار گرفتند که روش و ترتیب گزینش در زیر آمده است.

## ۱-۴-۳ گزینش داده‌های مورد استفاده

برای کاهش تعداد داده‌ها ابتدا محدوده مطالعه کوچک‌تر می‌شود. موقعیت گوشه‌های منطقه در جدول ۱-۳ آمده است.

جدول ۱-۳ موقعیت گوشه‌های منطقه انتخاب شده برای پردازش

گوشه مورد نظر	عرض جغرافیایی	طول جغرافیایی
شمال غربی	۶۴۸۹۲۵۶,۸۷	۲۹۷۸۴۰,۸۶۶۵
شمال شرقی	۶۴۸۹۲۵۶,۸۷	۳۹۵۵۲۹,۳۴۱۵
جنوب غربی	۶۳۵۳۶۱۱,۰۹۴	۲۹۷۸۴۰,۸۶۶۵
جنوب شرقی	۶۳۵۳۶۱۱,۰۹۴	۳۹۵۵۲۹,۳۴۱۵

در منطقه انتخاب شده ۲۳ چاه وجود دارد. از ۲۳ چاه موجود در این منطقه در ۱۹ چاه ویژگی TOC اندازه گیری شده است ولی تنها ۱۵ چاه دارای چاه نمودار هستند که با توجه به نکات گفته شده از بین این چاه‌ها ۱۴ چاه برای ادامه کار شرایط لازم را دارند.

در مرحله بعدی به علت نا هماهنگی چاه نمودارها تعداد دیگری از چاه‌ها حذف شدند تا به هماهنگی لازم بین چاه‌ها برای پردازش بدست آید. نتیجه این مرحله باقی ماندن تنها ۹ چاه برای پردازش بود. در بین داده‌های این ۹ چاه برای رسیدن به یک روش ثابت در پردازش‌ها چاه نمودارهای اضافه و داده‌های مربوط به اعماقی که داده‌های یک یا چند چاه نمودار موجود نبودند نیز حذف شدند. در نتیجه این عملیات داده‌های چاه نمودارهای باقیمانده به شکلی که در جدول ۲-۳ آمده باقی ماند که برای ادامه پردازش‌ها از آن‌ها استفاده شد.

جدول ۲-۳ خلاصه ای از وضعیت چاه‌های مورد استفاده در پروژه

تعداد نمونه‌های اندازه‌گیری شده برای TOC	تعداد کل داده‌های چاه نمودارهای باقیمانده	موقعیت چاه		نام چاه
		طول جغرافیایی	عرض جغرافیایی	
۴۴	۲۰۶۷۲	۳۴۶۸۶۵,۶۵۸۴	۶۴۶۳۴۴۸,۶۳۲	گیجرود <sup>۱</sup>
۵	۸۰۶۴	۳۴۵۳۳۴,۹۴۴۵	۶۴۶۳۹۱۸,۴۱۱	گیجرود <sup>۲</sup>
۲۲	۱۴۴۶۱	۳۷۴۸۹۰,۶۰۵۸	۶۳۵۶۹۰۷,۷۲۳	لیک پرینستون <sup>۳</sup>
۴۸	۷۰۸۰	۳۵۴۵۸۳,۲۶۱۲	۶۴۷۳۳۵۵,۵۰۷	مولالو <sup>۴</sup>
۸	۱۳۶۹۰	۳۵۳۶۶۲,۸۳۲۵	۶۴۲۹۴۵۶,۶۸۶	پل <sup>۵</sup>
۱۴	۸۸۵۵	۳۵۹۵۸۰,۶۹۵۹	۶۴۸۰۵۰۹,۴۳۳	کوئیز راک <sup>۶</sup>
۱۴	۶۶۱۰	۳۹۵۵۲۹,۳۴۱۵	۶۴۲۶۵۸۲,۷۲۹	راکینگهام <sup>۷</sup>
۱۲	۱۷۳۲۴	۳۱۸۱۹۲۹۵۸۸	۶۳۵۶۴۳۸,۷۹۷	شوگرلوف <sup>۸</sup>
۴۴	۱۹۵۸۵	۳۴۴۳۸۳,۶۶۹۲	۶۴۳۱۸۹۹,۲۶۶	وارنبرو <sup>۹</sup>

تاریخچه ای از چاه‌های مورد مطالعه در ادامه آمده است.

### ۲-۴-۳ تاریخچه ای از چاه‌های مورد مطالعه

در قسمت‌های مختلف حوزه پرت لایه‌های مربوط به دوران‌های اوردوئیسین<sup>۱۰</sup>، پرمین<sup>۱۱</sup>، تریاس<sup>۱۲</sup>،

ژوراسیک<sup>۱۳</sup>، کرتاسه<sup>۱</sup> و ترشیری<sup>۲</sup> دیده شده است (Bozanic 1969). شکل ۲-۳ نمای کلی از چینه

<sup>1</sup> Gage road 1

<sup>2</sup> Gage road 2

<sup>3</sup> Lake Preston 1

<sup>4</sup> Mullaloo 1

<sup>5</sup> Peel 1

<sup>6</sup> Quinnes rock 1

<sup>7</sup> Rockingham 1

<sup>8</sup> Sugarloaf 1

<sup>9</sup> Warnbro 1

<sup>10</sup> Ordovician

<sup>11</sup> Permian

<sup>12</sup> Triassic

<sup>13</sup> Jurassic

شناسی، لایه‌ها و سازندهای موجود در حوزه پرت را نشان می‌دهد. چاه‌های انتخابی برای ادامه مطالعه بین سال‌های ۱۹۶۸ تا ۱۹۸۴ حفاری شده‌اند. که در زیر خلاصه ای از مشخصات هر یک از چاه‌ها آمده است.

---

<sup>1</sup> Cretaceous  
<sup>2</sup> Tertiary



PERIOD	EPOCH	FORMATION	LITHOLOGY	ENVIRONMENT
QUATERNARY	PLEISTOCENE	COASTAL LIMESTONE		Marine
TERTIARY	PALAEOCENE	KINGS PARK SHALE		Marine
CRETACEOUS	UPPER	GINGIN CHALK		Marine
		OSBORNE FM.		Marine
	LOWER	LEEDERVILLE FM.		Marine to continental
		SOUTH PERTH SHALE		
JURASSIC	UPPER	YARRAGA DEE FM.		Continental
	MIDDLE	CADDA FM.		Fluvio-deltaic marginal marine
	LOWER	COCKLESHELL GULLY FM.	CATTAMARRA FM.	
ENEABBA (equiv.) FM.				
TRIASSIC	UPPER	LESUEUR SANDSTONE		
	MIDDLE	WOODADA FM.		Marine
	LOWER	KOCKATEA SHALE		Marine
PERMIAN	UPPER	Undifferentiated		Mainly marine
	LOWER			
SILURIAN to ORDOVICIAN		Undifferentiated		Glacio-continental
PRECAMBRIAN		Basement		

شکل ۲-۳ نمایی از چینہ شناسی، لایه ها و سازندهای موجود در حوزه پرت ( Marshal, Mitchell et al. 1983)

### ۳-۴-۱- چاه کویینزراک ۱

چاه کویینزراک ۱ قدیمی‌ترین چاه بین چاه‌های انتخابی برای پردازش است. موقعیت این چاه عبارت است از  $31^{\circ} 48' 08.7''$  جنوبی و  $115^{\circ} 30' 50.0''$  شرقی است. این چاه توسط شرکت نفت غرب استرالیا<sup>۱</sup> حفاری شد. کل عمق حفاری شده ۷۲۴۸ فوت<sup>۲</sup> معادل ۲۲۰۹ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Bozanic 1969).

مطالعات لرزه ای مربوط به چاه توسط شرکت وپت<sup>۳</sup> در سال ۱۹۶۵ شروع شد و تا سال ۱۹۶۸ ادامه یافت. مطالعات ژئوفیزیکی حوزه پرت توسط زمین‌شناسان لامنت<sup>۴</sup>، منابع معدنی بوارو<sup>۵</sup> و دانشگاه نیو سوات ولز<sup>۶</sup> انجام گرفته است (Hawkins 1965). مطالعات زمین شناسی تکمیلی مختلف برای منطقه حوزه پرت انجام شد (Smith 1967).

خلاصه ای از سازندهای موجود در چاه کویینزراک ۱ در جدول ۳-۳ آمده است.

جدول ۳-۳ تفکیکی از چینه شناسی چاه کویینزراک ۱ (Bozanic 1969)

دوره	سازند	عمق شروع حفاری (فوت)	عمق زیر آب (فوت)	ضخامت حفاری (فوت)	ضخامت لایه (فوت)
ترشیری	بدون نام	۲۱۲	۱۳۳-	۹۵۷	؟
کرتاسه پایینی	سازند پرت جنوبی ۷				۱۵۱۴
	عضو ماسه سنگ‌های لدرویل ۱	۱۱۶۹	۱۰۹۰-	۸۳۶	۷۸۶
	عضو شیل پرت جنوبی ۲	۲۰۰۵	۱۹۲۶-	۷۷۵	۷۲۹
ژوراسیک بالایی	سازند یاراگادی ۳	۲۷۸۰	۲۷۰۱-	۴۴۶۸	۴۲۰۰
	کل عمق	۷۲۴۸	۷۱۶۹-		

<sup>1</sup> West Australian Petroleum Pty. Limited

<sup>2</sup> feet

<sup>3</sup> Wapet

<sup>4</sup> Lamont Geological Observatory

<sup>5</sup> Bureau of Mineral Resources

<sup>6</sup> University of New South Wales

<sup>7</sup> South Perth Fm.

### چاه گیجرود ۱ ۲-۲-۴-۳

این چاه در سال ۱۹۶۸ حفاری شد. موقعیت این چاه ۱۲,۲۱' ۵۷° ۳۱' جنوبی و ۳۸,۲۸' ۲۲' ۱۱۵° شرقی است. این چاه توسط شرکت نفت غرب استرالیا حفاری شد. کل عمق حفاری شده ۱۲۰۰۹ فوت معادل ۳۶۶۰ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد. (Bozanic 1969)

خلاصه ای از سازندهای موجود در چاه گیجرود ۱ در جدول ۳-۴ آمده است.

جدول ۳-۴ داده‌های مربوط به چینه شناسی چاه گیجرود ۱ (Bozanic 1969)

دوره	سازند	عمق شروع حفاری (فوت)	عمق زیر آب (فوت)	ضخامت حفاری (فوت)
پلیستوسن <sup>۴</sup>	سنگ آهک کوستال	۲۶۱	۱۹۱-	۱۱۳ یا کمتر
میوسن <sup>۵</sup>	بی نام			۱۷۶۹
کرتاسه پایینی	سازند پرت جنوبی			۳۸۶۰
	عضو ماسه سنگ‌های لدرویل	۲۰۳۰	۱۹۶۰-	۲۱۷۰
	عضو شیل پرت جنوبی	۴۲۰۰	۴۱۷۰-	۱۰۰۸
	عضو ماسه سنگی گیج <sup>۶</sup>	۵۲۰۸	۵۱۳۸-	۶۹۸
ژوراسیک بالایی	سازند یاراگادی	۵۹۰۶	۵۸۳۶	۶۱۰۳
	عضو شیلی کوپینز <sup>۷</sup>	۸۴۱۰	۸۳۴۰-	
	کل عمق	۱۲۰۰۹	۱۱۹۳۹-	

<sup>1</sup> Leederville Sandstone Member

<sup>2</sup> South Perth Shale Member

<sup>3</sup> Yarragadee Fm.

<sup>4</sup> Pleistocene

<sup>5</sup> Miocene

<sup>6</sup> Gage Sandstone Member

<sup>7</sup> Quinns Shale Member

### ۳-۴-۳-۳ چاه گیجروود ۲

این چاه در سال ۱۹۷۱ حفاری شد. موقعیت این چاه  $31^{\circ} 57' 05''$  جنوبی و  $115^{\circ} 41' 35''$  شرقی است. این چاه توسط شرکت نفت غرب استرالیا حفاری شد. کل عمق حفاری شده ۹۷۵۰ فوت معادل ۲۹۷۱ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Moyes 1971).

خلاصه ای از سازندهای موجود در چاه گیجروود ۲ در جدول ۳-۵ آمده است.

جدول ۳-۵ داده‌های مربوط به چینه شناسی چاه گیجروود ۲ (Moyes 1971)

دوره	سازند	عمق شروع حفاری (فوت)	عمق زیر آب (فوت)	ضخامت حفاری (فوت)
پلیستوسن	سنگ آهک کوستال	۳۳۹	۲۴۰-	۱۵۰
میوسن <sup>۱</sup>	بی نام	۴۸۹	۳۹۰-	۱۴۰۶
کرتاسه پایینی	سازند پرت جنوبی			۲۶۱۵
	عضو ماسه سنگ‌های لدرویل	۱۸۹۵	۱۷۹۶-	۱۸۷۰
	عضو شیل پرت جنوبی	۳۷۶۵	۳۶۶۶-	۶۶۵
	عضو ماسه سنگی گیج	۴۴۳۰	۴۳۳۱-	۸۰
ژوراسیک بالایی	سازند یاراگادی	۴۵۱۰	۴۴۱۱-	۵۲۴۰
	عضو شیلی کویینز	۹۳۴۳	۹۲۴۴-	۲۷۵
	یاراگادی بالایی ژوراسیک	۹۶۱۸	۹۵۱۹-	۱۳۲
	کل عمق	۹۷۵۰	۹۶۵۱-	

<sup>1</sup> Miocene

### چاه وارنبرو ۱ ۴-۲-۴-۳

این چاه در سال ۱۹۷۰ حفاری شد. موقعیت این چاه "۲۰' ۱۴' ۳۲° جنوبی و "۴۵' ۲۰' ۱۱۵° شرقی است. این چاه توسط شرکت نفت غرب استرالیا حفاری شد. کل عمق حفاری شده ۱۲۰۰۹ فوت معادل ۳۶۶۰ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Moyes 1971). خلاصه ای از سازندهای موجود در چاه وارنبرو ۱ در جدول ۳-۶ آمده است.

جدول ۳-۶ داده‌های مربوط به چینه شناسی چاه وارنبرو ۱ (Moyes 1971)

سازند	عمق شروع حفاری (فوت)	عمق زیر آب (فوت)	ضخامت حفاری (فوت)
سنگ آهک کوستال	۲۳۷	۱۵۵-	
بی نام	؟	؟	؟
سازند کینگز پارک	۱۱۰۳	۱۰۲۱-	۷۰۹
گچ گینجین، پویزن هیل و معادل ماسه سبز مولکاپ <sup>۱</sup>	۱۸۱۲	۱۷۳۰-	۷۸۸
سازند آزبورن <sup>۲</sup>	۲۶۰۰	۲۵۱۸-	۶۹۰
سازند پرت جنوبی	۳۲۹۰	۳۲۰۸-	۳۹۴۰
عضو ماسه سنگ‌های لدرویل	۳۲۹۰	۳۲۰۸-	۱۳۳۰
عضو شیلی پرت جنوبی	۴۶۲۰	۴۵۳۸-	۱۷۶۰
عضو ماسه سنگی بازال <sup>۳</sup>	۶۳۸۰	۶۲۹۸-	۸۵۰
یاراگادی نوکومین <sup>۴</sup>	۷۲۳۰	۷۱۴۸-	۴۷۷۹
عضو شیلی کویینز	۹۶۲۰	۹۵۳۸-	۹۷
یاراگادی ژوراسیک	۹۷۱۷	۹۶۳۵-	
کل عمق	۱۲۰۰۹	۱۱۹۲۷-	

<sup>1</sup> Gingin Chalk, Poison Hill & Molecap Greensand Equivalents

<sup>2</sup> Osborne Formation

<sup>3</sup> Basal Sandstone Member

<sup>4</sup> Neocomian Yarragadee

### ۵-۲-۴-۳ شوگرلوف ۱

این چاه در سال ۱۹۷۱ حفاری شد. موقعیت این چاه  $32^{\circ} 54' 55''$  جنوبی و  $115^{\circ} 03' 09''$  شرقی است. این چاه توسط شرکت نفت غرب استرالیا حفاری شد. کل عمق حفاری شده ۱۲۰۰۰ فوت معادل ۳۶۵۷ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Bird and Moyes 1971). خلاصه ای از سازندهای موجود در چاه شوگرلوف ۱ در جدول ۷-۳ آمده است.

جدول ۷-۳ داده‌های مربوط به چینه شناسی چاه شوگرلوف ۱ (Bird and Moyes 1971)

دوره	سازند	عمق شروع حفاری (فوت)	عمق زیر آب (فوت)	ضخامت حفاری (فوت)
ترشیری تا کواترنری	کربنات‌های بی نام	۲۵۰	۱۵۱-	۵۸۰
کرتاسه بالایی	سازند آذربورن	۸۳۰	۷۳۱-	۹۲۰
کرتاسه پایینی	سازند پرت جنوبی	۱۷۵۰		
	عضو ماسه سنگ‌های لدرویل	۱۷۵۰	۱۶۵۱-	۵۸۰
	سازند یاگادی	۲۳۳۰	۲۲۳۱-	۴۵۵۲
	عضو شیلی کویینز	۶۸۸۲	۶۷۸۳-	۹۰
ژوراسیک بالایی	سازند یاراگادی	۶۹۷۲	۶۸۷۳-	۵۰۲۸
	عضو شیلی کویینز	۹۳۴۳	۹۲۴۴-	۲۷۵
	کل عمق	۱۲۰۰۰	۱۱۹۰۱-	

### ۶-۲-۴-۳ چاه لیک پرینستون ۱

این چاه در سال ۱۹۷۲ حفاری شد. موقعیت این چاه  $32^{\circ} 55' 12,46''$  جنوبی و  $115^{\circ} 39' 38,85''$  شرقی است. این چاه توسط شرکت نفت غرب استرالیا حفاری شد. کل عمق حفاری شده ۱۴۹۷۷ فوت معادل ۴۵۶۴ متر است. این چاه در قسمت ساحلی قرار داد (Young and Johanson 1973).

خلاصه ای از سازندهای موجود در چاه لیک پرینستون ۱ در جدول ۳-۸ آمده است.

جدول ۳-۸ داده‌های مربوط به چینه شناسی چاه لیک پرینستون ۱ (Young and Johanson 1973)

دوره	سازند	عمق شروع سازند (فوت)	ضخامت حفاری (فوت)
هولوسین <sup>۱</sup>	ماسه‌های سطحی	۱۵	۶۵
پولیستوسن	سنگ آهک کواستال	۸۰	۵۴
کرتاسه بالایی	گروه وارنبرو		
	عضو ماسه سنگ‌های لدرویل	۱۳۸	۲۳۰
ژوراسیک	سازند کاکلیشیل گولی <sup>۲</sup>		
	عضو انبأ	۳۶۸	۳۶۳۲
تریاسیک	ماسه سنگ لزوئر <sup>۳</sup>	۴۰۰۰	۷۳۹۹
پرمین	ماسه سنگ سابینا <sup>۴</sup>	۱۱۳۹۹	۱۸۴۰
	زغال سو <sup>۵</sup>	۱۳۲۳۹	۱۷۳۸
	عمق کل	۱۴۹۷۷	

### ۳-۴-۲-۷ چاه پل ۱

این چاه در سال ۱۹۷۷ حفاری شد. موقعیت این چاه ۴۷،۸' ۱۵" ۳۲° جنوبی و ۴۳،۰' ۲۶" ۱۱۵° شرقی است. این چاه توسط شرکت نفت فیلیپ استرالیا<sup>۶</sup> حفاری شد. کل عمق حفاری شده ۳۷۱۴ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Phillips-Australian-Oil-Company 1978).

خلاصه ای از سازندهای موجود در چاه پل ۱ در جدول ۳-۹ آمده است.

<sup>1</sup> Holocene

<sup>2</sup> Cockle

<sup>3</sup> Lesueur Sandstone

<sup>4</sup> Sabina Sandstone

<sup>5</sup> Sue Coal Measures

<sup>6</sup> Phillips Australian Oil Company

جدول ۳-۹ داده‌های مربوط به چینه شناسی چاه پل ۱ (Phillips-Australian-Oil-Company 1978)

دوره	سازند	عمق شروع حفاری (متر)	عمق زیر آب (متر)	ضخامت حفاری (متر)
ترشیری	بی نام	۷۲	۴۲-	۱۹۳
کرتاسه بالایی	گروه کولینا <sup>۱</sup>	۲۶۵	۲۳۵-	۵۹۴
کرتاسه پایینی	گروه وارنبرو	۸۵۹	۸۲۹-	۷۶۶
	سازند کارناک <sup>۲</sup>	۱۶۲۵	۱۵۹۵-	۱۴۳۹
	عضو ماسه سنگ کوارسی <sup>۳</sup>	۱۶۲۵	۱۵۹۵-	۷۸۴
	عضو ماسه سنگی/شیلی	۲۴۰۹	۲۳۷۹-	۶۵۵
	سازند یاراگادی	۳۰۶۴	۳۰۳۴-	۶۵۰
	عضو ماسه سنگی بالایی	۳۰۶۴	۳۰۳۴-	۶۵۰
ژوراسیک بالایی	عضو شیلی کویینز	۳۵۰۵	۳۴۷۵-	۴۶
	عضو شیلی قبل از کویینز	۳۵۵۱	۳۵۲۱-	۱۶۳
	کل عمق	۳۷۱۴	۳۶۸۴-	

### ۳-۴-۱-۸ چاه راکینگهام ۱

این چاه در سال ۱۹۸۳ حفاری شد. موقعیت این چاه ۳۸' ۱۷" ۳۲° جنوبی و ۲۱,۳" ۵۳' ۱۱۵° شرقی است. این چاه توسط شرکت نفت و گاز ققنوس<sup>۴</sup> حفاری شد. کل عمق حفاری شده ۱۵۶۳,۲ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد. نمایی از ستون چینه شناسی این چاه در شکل ۳-۳ آمده است (Marshall, Mitchell et al. 1983).

<sup>1</sup> Coolyena Group

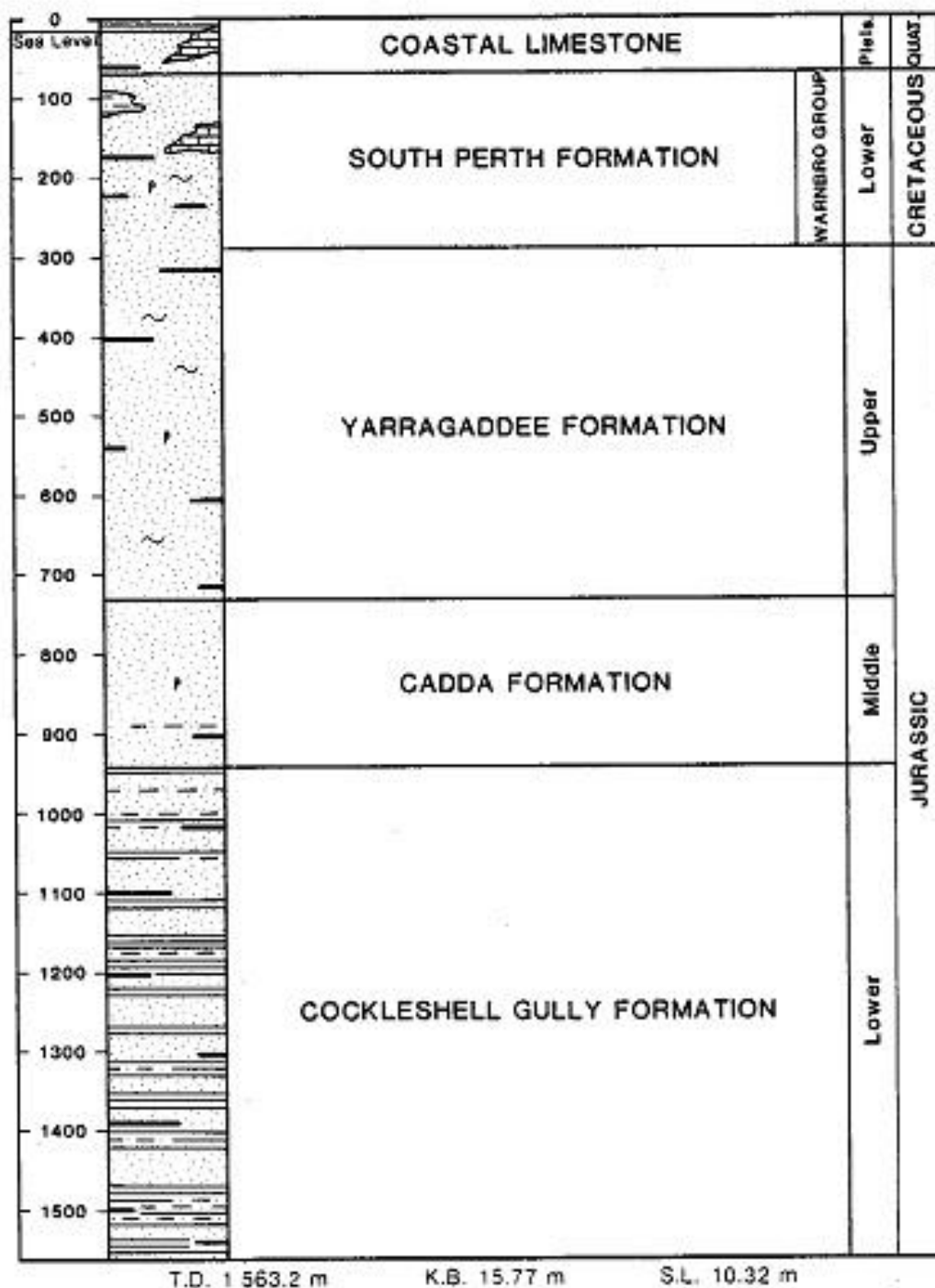
<sup>2</sup> Carnac Formation

<sup>3</sup> Coarse Sandstone Member

<sup>4</sup> Phoenix Oil & Gas N.L.



### Rockingham No. 1 STRATIGRAPHIC COLUMN



GEOGRAPHICS

June 1981

شکل ۳-۳ نمایی از ستون چینه شناسی چاه راکینگهام (Marshall, Mitchell et al. 1983)

### چاه مولالو ۱ ۹-۴-۳

این چاه در سال ۱۹۸۴ حفاری شد. موقعیت این چاه  $31^{\circ} 52' 03.02''$  جنوبی و  $115^{\circ} 27' 42.18''$  شرقی است. این چاه توسط شرکت اکتشاف و تولید اسو استرالیا<sup>۱</sup> حفاری شد. کل عمق حفاری شده ۲۰۳۰ متر است. این چاه در مرز قاره ای اقیانوسی قرار داد (Brooks 1984).

### سازندهای هدف ۳-۴-۳

در بخش ۳-۴-۲ سازندهای موجود در منطقه و سازندهای مشاهده شده در هر چاه آورده شد. با توجه به چاه نمودارهای موجود برای هر چاه و وجود یا عدم وجود نمونه‌های مغزه آنالیز شده برای اندازه گیری TOC در سازندها و چاه‌ها برخی از سازندها حذف شدند. با توجه به این مسئله سازندهای باقی عبارتند از: کاکلیشیل گولی، ماسه سنگ گیچ، لدرویلا، لزوتر، پارملیا، پرت جنوبی و سازند یاراگادی.

### چاه نمودارهای مورد استفاده ۱-۳-۴-۳

در چاه‌های مختلف با توجه به تقدم و تأخر زمانی بین حفاری‌ها چاه نمودارهای یکسانی تهیه نشده است و در غالب موارد عدم هماهنگی بین چاه نمودارها مشاهده می‌شود. برای یکسان سازی چاه نمودارها برای پردازش، تعدادی از چاه نمودارها برای برخی از چاه‌ها حذف شدند و در بعضی موارد داده‌هایی که در بعضی از اعماق نقص داشتند نیز حذف شدند. پس از عملیات حذف داده‌های ناکارآمد و حصول مجموعه ای مشابه بین چاه‌ها چاه نمودارهای باقی مانده برای ادامه پردازش‌ها عبارتند از: DRHO، DT، GR، IND، RHOB و SP. این مجموعه در بین تمام داده‌های موجود مشترک هستند و در بعضی از داده‌ها مقدار TOC نیز موجود می‌باشد که برای آموزش و آزمودن الگوریتم مورد استفاده قرار می‌گیرد.

<sup>1</sup> Esso Exploration and Production Australia Inc.

## ۳-۵ داده‌های مصنوعی

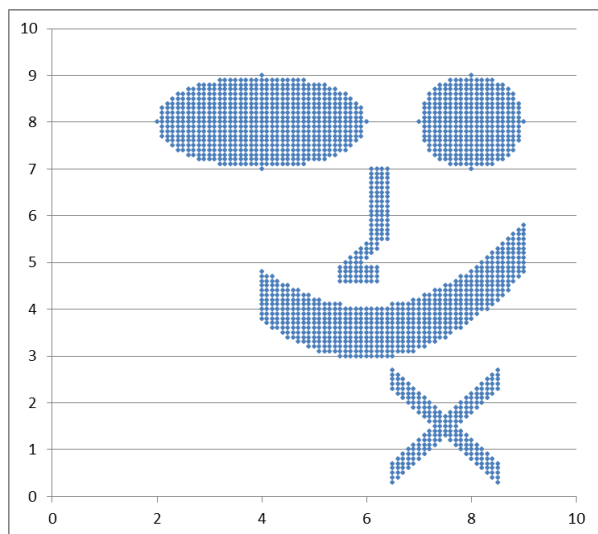
چهار دسته داده‌های مصنوعی ساخته شده است که ۳ دسته برای بررسی کارایی روش‌های خوشه‌بندی و کلاسه‌بندی ابداعی بر پایه الگوریتم خوشه‌بندی DBSCAN هستند و ۱ دسته برای بررسی و توضیح مراحل اجرای الگوریتم لولیموت است.

### ۳-۵-۱ دسته اول

مشکل روش‌های معمول خوشه‌بندی مانند میانگین K و روش گوستافسون-کسل وابستگی شدید این موارد به شکل و نحوه قرارگیری داده‌ها است. به طوری که داده‌هایی با شکل غیر معمول را نمی‌توان به وسیله این روش تفکیک کرد، به عبارت بهتر کارایی این روش‌ها در خوشه‌بندی داده‌های غیر توده ای کم است.

مشکل دیگر روش‌های معمول نیاز به تعیین تعداد خوشه‌ها به وسیله کاربر است. به گونه ای که لازم است کاربر از محیط داده‌ها یک دید داشته باشد. روش DBSCAN نسبت به دو روش دیگر کمتر دچار این مشکلات است. در ساخت داده‌های مصنوعی این موارد مد نظر قرار گرفته است.

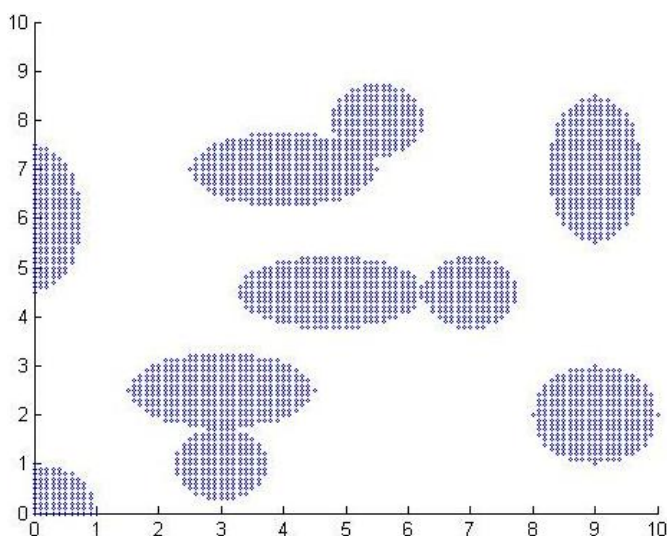
در ساخت داده‌های مصنوعی انواع شکل‌های معمول و غیر معمول با هم در یک جا ایجاد شدند تا اطمینان از کارایی این روش بدست آید. برای بررسی کارایی روش DBSCAN ابتدا داده‌هایی مصنوعی ساخته شد. در شکل ۳-۴ داده‌های مصنوعی ساخته شده نمایش داده شده‌اند.



شکل ۳-۴ داده‌های مصنوعی تولید شده برای آزمون کارایی روش DBSCAN

### ۲-۵-۳ دسته دوم

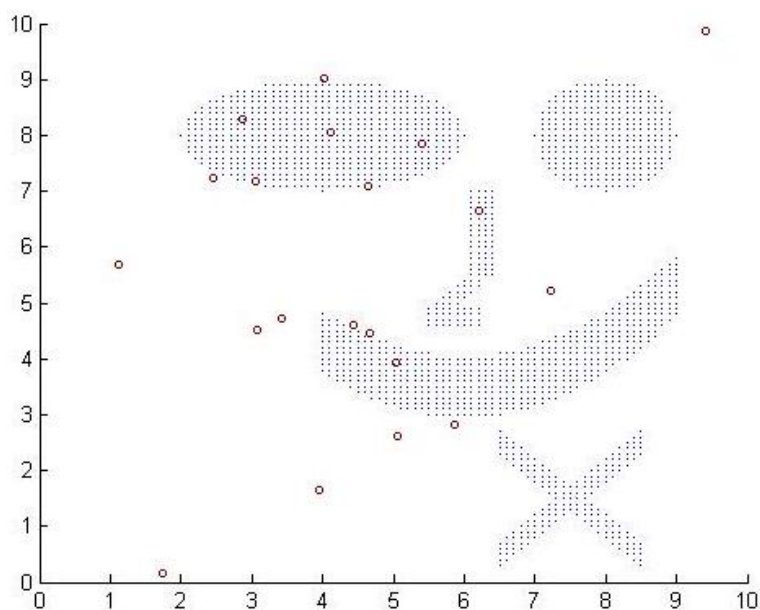
زمانی که داده‌ها نسبتاً منظم‌تر باشند و اشکال مشابهی در داده‌ها وجود داشته باشد کارایی دو روش میانگین K و گوستافسون-کسل برای خوشه‌بندی بهبود می‌یابد. برای مقایسه حالتی که شکل داده‌ها منظم‌تر است سری دیگری از داده‌های مصنوعی با شکل دیگر تولید می‌شود و کارایی روش‌ها روی این سری نیز بررسی می‌شود. شکل ۳-۵ داده‌های مصنوعی سری دوم را نشان می‌دهد.



شکل ۳-۵ داده‌های تولید شده برای مقایسه کارایی روش‌های مختلف در حالت داده‌های منظم‌تر

### ۳-۵-۳ دسته سوم

برای بررسی کارایی روش ارائه شده برای کلاسه‌بندی، بر پایه روش خوشه‌بندی DBSCAN به داده‌های مصنوعی تولید شده سری اول (شکل ۳-۴) ۲۰ داده به صورت تصادفی تولید شد. موقعیت این داده‌ها در شکل ۳-۶ آورده شده است.



شکل ۳-۶ داده‌های تصادفی تولید شده برای بررسی کارایی روش کلاسه‌بندی بر پایه روش خوشه‌بندی DBSCAN

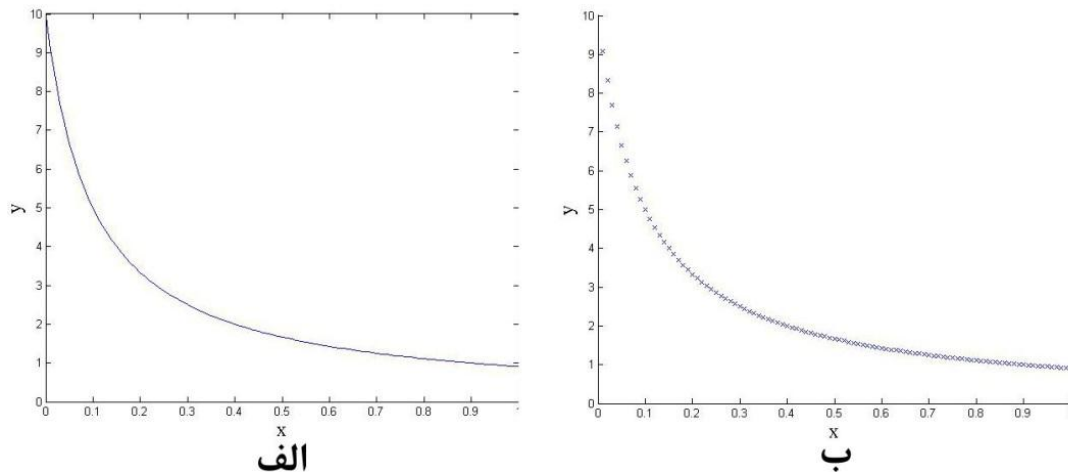
### ۳-۵-۴ دسته چهارم

داده‌های ساخته شده برای بررسی روند کار الگوریتم لولیموت، نمونه‌های برداشت شده از تابع آورده شده در رابطه ۳-۱ با فاصله ۰,۰۱ در راستای محور x است.

$$y = \frac{1}{0.1+x}$$

۳-۱

شکل ۷-۳ قسمت الف شکل رابطه ۱-۳ را نشان می‌دهد و شکل ۷-۳ قسمت ب نمونه‌های برداشته شده از تابع اصلی را نشان می‌دهد.



شکل ۷-۳ داده مصنوعی تولید شده برای نشان دادن گام‌های اجرای الگوریتم لولیموت: الف) نمودار تابع اصلی، ب) شکل نمونه‌های برداشته شده

## فصل چهارم

### مدل سازی محتوای کربن آلی

## ۴ مدل‌سازی محتوای کربن آلی

### ۴-۱ مقدمه

در این فصل ابتدا روش‌های مورد استفاده برای خوشه‌بندی و کلاسه‌بندی با داده‌های مصنوعی ساخته شده سنجیده می‌شوند. پس از آن روند کار الگوریتم لولیموت بر روی داده مصنوعی ساخته شده در فصل قبل نشان داده شده است.

در ادامه روند تخمین مقدار TOC برای چاه‌ها و سازندهای مختلف آورده شده است. پس از آن مدل‌سازی انجام شده برای مخزن آورده شده است.

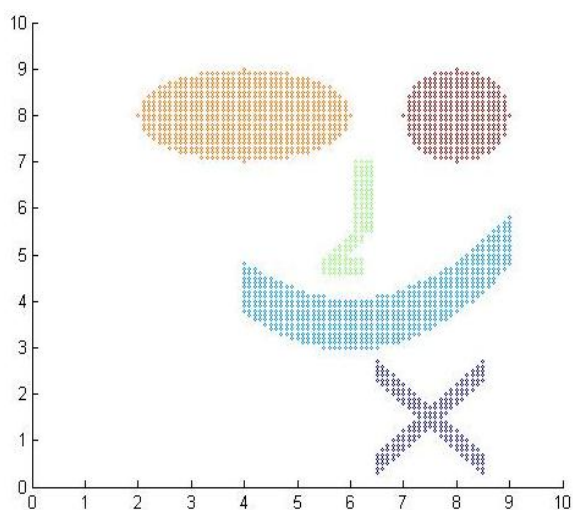
### ۴-۲ بررسی کارایی روش‌های مورد استفاده

#### ۴-۲-۱ خوشه‌بندی DBSCAN

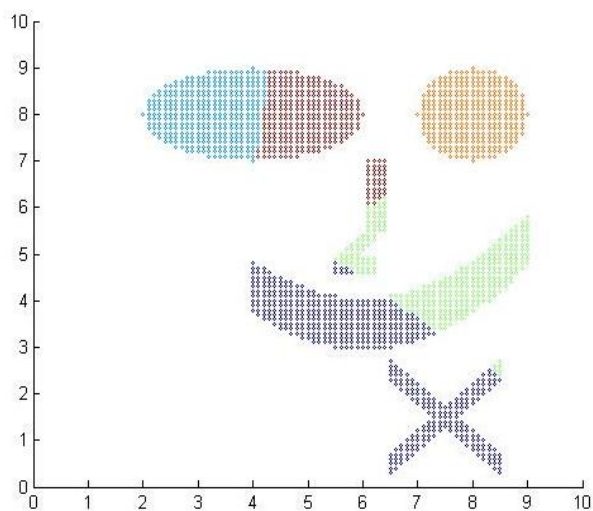
روش‌های خوشه‌بندی از قبیل میانگین K و گوستافسون-کسل به شکل داده‌ها واکنش نشان نمی‌دهند و شکل‌های غیرمعمول داده را نمی‌توانند تشخیص دهند. در واقع کارایی این روش‌ها اغلب برای داده‌های توده‌ای است.

نتیجه خوشه‌بندی داده‌های مصنوعی ساخته شده در فصل قبل که در شکل ۳-۴ آمده است با استفاده از روش‌های مختلف آورده شده در فصل ۲ به ترتیب زیر است.

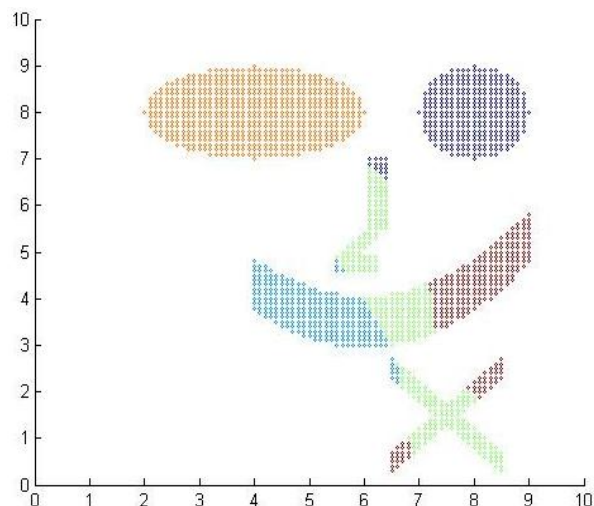




شکل ۱-۴ نتیجه خوشه بندی با روش DBSCAN خوشه های مختلف با رنگ های متفاوت نشان داده شده است



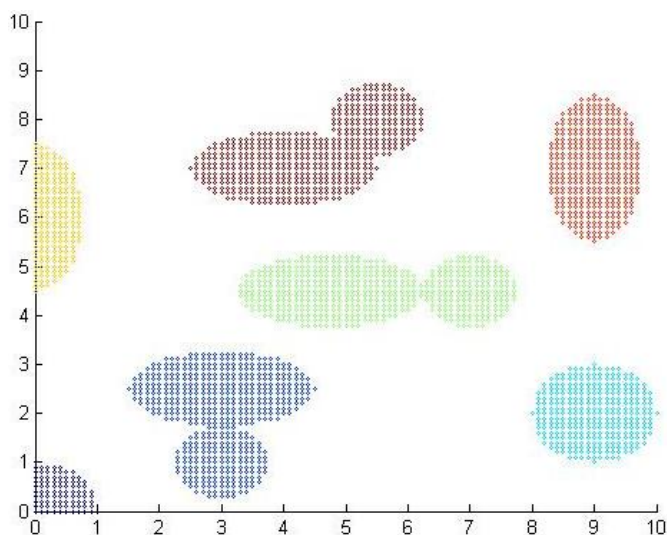
شکل ۲-۴ نتیجه حاصل از خوشه بندی با روش میانگین K خوشه های مختلف با رنگ های متفاوت نشان داده شده است



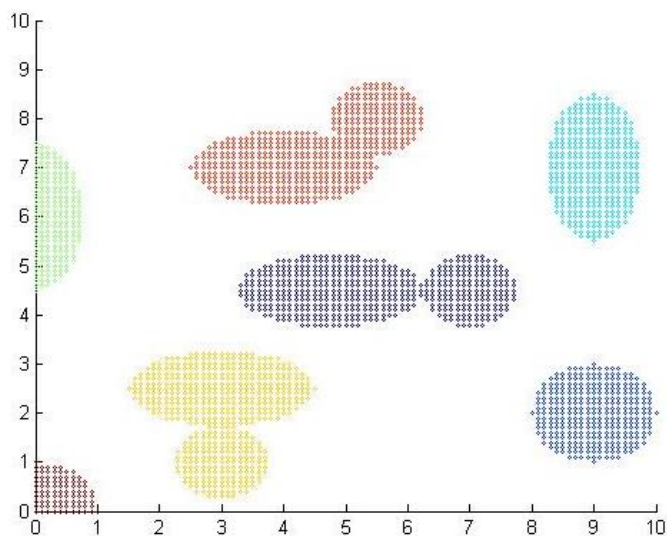
شکل ۳-۴ نتیجه حاصل از خوشه بندی با روش گوستافسون-کسل خوشه های مختلف با رنگ های متفاوت نشان داده شده است

همان طور که در شکل ۱-۴ تا شکل ۳-۴ دیده می شود کارایی روش DBSCAN زمانی که اشکال پیچیده در داده ها وجود دارد یا شکل های مختلف در یک سری داده وجود داشته باشد، بهتر از دو روش دیگر است؛ به گونه ای که دو روش دیگر تقریباً ناکارآمد می شوند.

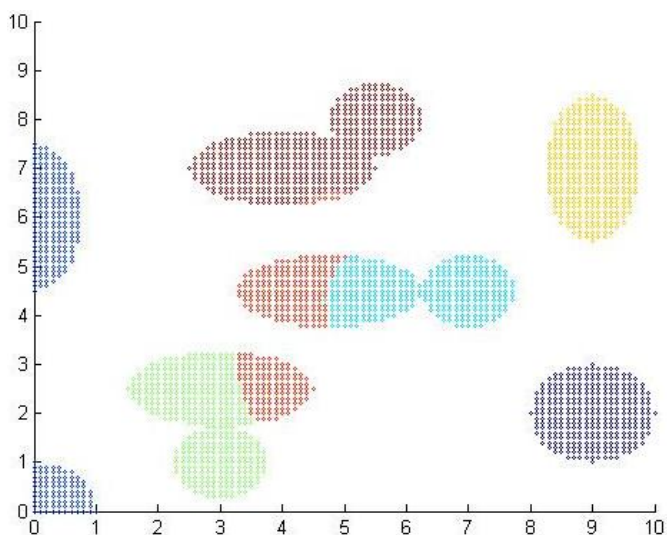
نتیجه خوشه بندی برای داده های مصنوعی ساخته شده سری دوم که در شکل ۳-۵ آمده است نیز به ترتیب در شکل ۴-۴ تا شکل ۶-۴ آورده شده است.



شکل ۴-۴ نتیجه حاصل از خوشه بندی با روش DBSCAN در داده های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است



شکل ۴-۵ نتیجه حاصل از خوشه‌بندی با روش میانگین K در داده‌های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است



شکل ۴-۶ نتیجه حاصل از خوشه‌بندی با روش گوستافسون-کسل در داده‌های سری دوم هر خوشه با رنگی متفاوت نشان داده شده است

همان‌گونه که در شکل ۴-۵ دیده می‌شود نتیجه خوشه‌بندی داده‌های سری دوم با روش میانگین K بسیار خوب است ولی نتیجه خوشه‌بندی به وسیله گوستافسون-کسل باز هم ضعیف است. البته ذکر این نکته لازم است که روش‌های خوشه‌بندی میانگین K و گوستافسون-کسل به ترتیب داده‌ها وابسته هستند و امکان دارد با تغییر ترتیب داده‌ها نتایج تغییر کند.

نکته دوم درباره خوشه‌بندی با استفاده از روش‌های میانگین K و گوستافسون-کسل این است که نتیجه این روش‌های همیشه یکسان نیست و امکان دارد با اجرای دوباره این الگوریتم‌ها نتیجه

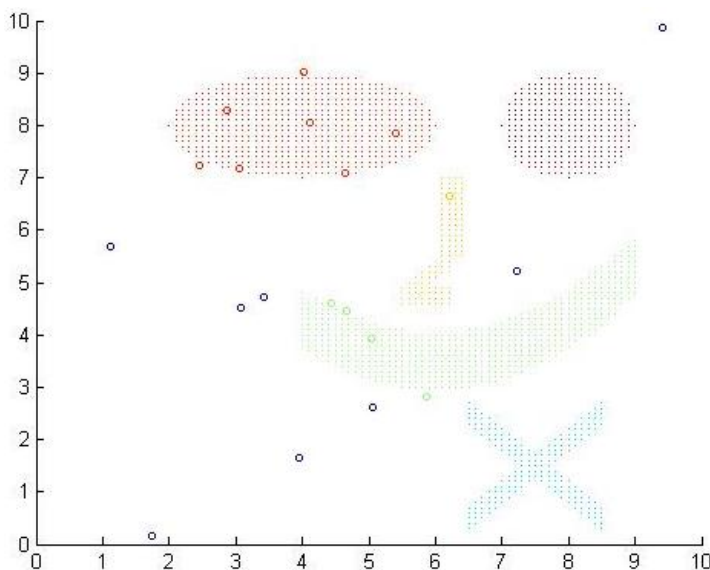
تغییر کند. اما روش خوشه‌بندی DBSCAN این‌گونه نیست و نتیجه همیشه یکسان است. سرعت اجرای این الگوریتم‌ها نسبت به هم نیز متفاوت است. سرعت اجرای الگوریتم میانگین  $K$  نسبت به دو الگوریتم دیگر بسیار سریع‌تر است و الگوریتم DBSCAN نسبت به دو الگوریتم دیگر سرعت کمتری دارد و زمان بیشتری برای اجرا لازم دارد.

با توجه به داده‌های مصنوعی ساخته شده و اشکال آورده شده می‌توان نتیجه گرفت که در داده‌هایی با ساختار مشابه (شکل‌های پیچیده و فاصله بین خوشه‌ای نسبتاً زیاد) روش DBSCAN بسیار خوب عمل می‌کند. با توجه به کارایی این روش الگوریتم کلاسه‌بندی با استفاده از روش DBSCAN ایجاد شد که در قسمت بعدی نتیجه اجرای آن بر روی داده‌های مصنوعی تولید شده به منظور بررسی کارایی روش ابداعی توضیح داده شده است.

#### ۲-۲-۴ کلاسه‌بندی DBSCAN

برای بررسی کارایی کلاسه‌بندی با روش ابداعی داده‌های مصنوعی ساخته شده شکل ۳-۶ مورد استفاده قرار گرفته است. کلاس‌های مورد استفاده خوشه‌های حاصل از خوشه‌بندی به روش DBSCAN (شکل ۴-۱) است. نتیجه کلاسه‌بندی با روش ارائه شده در شکل ۴-۷ نشان داده شده است.

همان‌طور که در شکل ۴-۷ دیده می‌شود نقاطی که در خارج از محدوده هر یک از خوشه‌ها هستند به عنوان نوفه شناخته می‌شوند. لازم به ذکر است که با تغییر پارامتر  $\epsilon$  می‌توان تعداد نوفه‌ها را کاهش یا افزایش داد. به عبارت دیگر می‌توان میزان اطمینان از تعلق یا عدم تعلق یک داده را به یک کلاس با تغییر پارامتر  $\epsilon$  بررسی کرد.



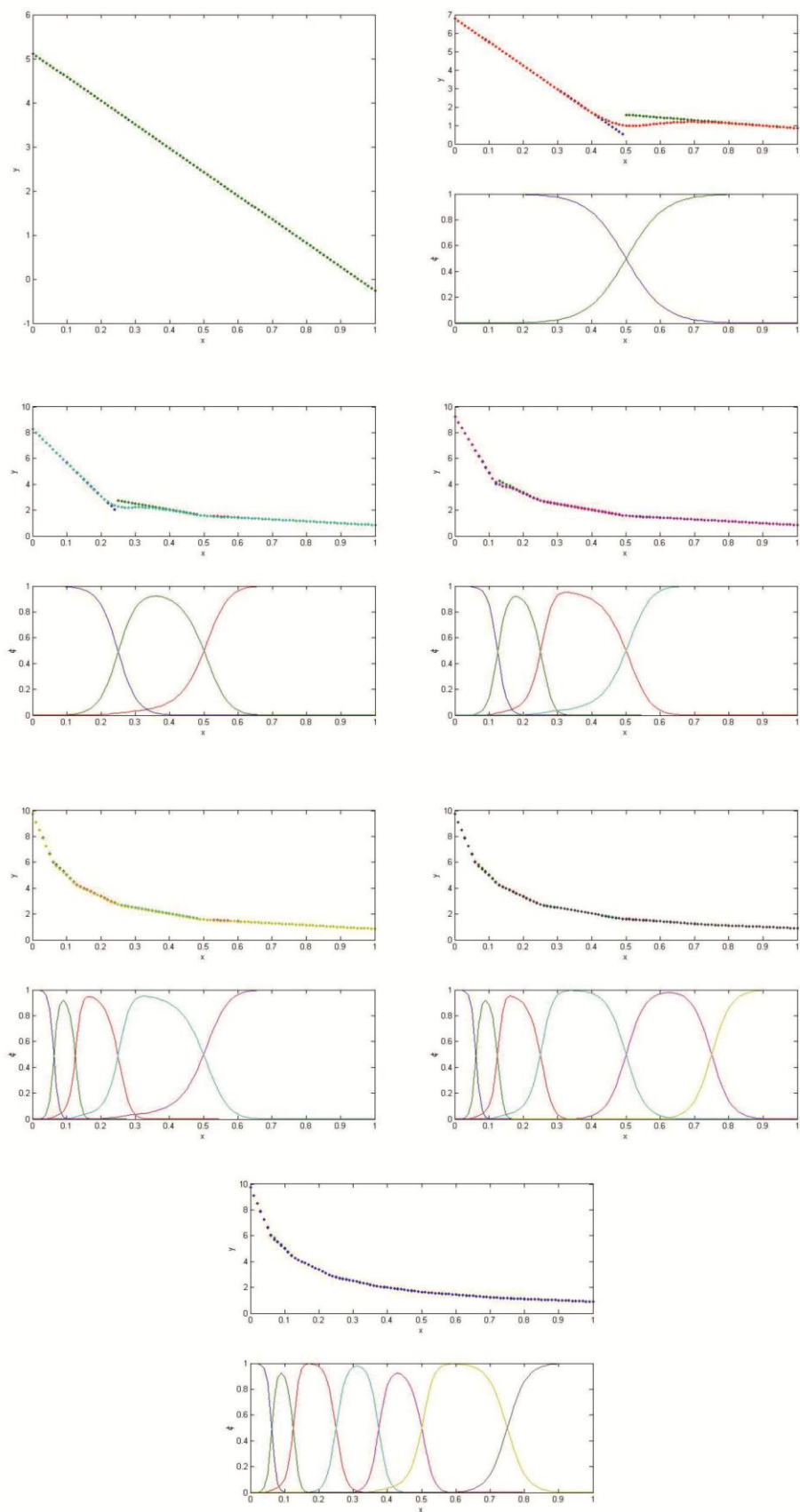
شکل ۴-۷ نتیجه کلاسه بندی با استفاده از روش ابداعی بر پایه روش خوشه بندی DBSCAN

### ۳-۲-۴ لولیموت

بررسی کارایی الگوریتم لولیموت بر روی داده مصنوعی سری چهارم که در بخش ۳-۵-۴ آورده شده بررسی می شود. مراحل مدل سازی به صورت گام به گام توضیح داده می شود. در گام اول از تمام ۱۰۱ نمونه برداشته شده برای ساخت یک مدل استفاده می شود که نتیجه آن خطی با معادله  $y = -5.381x + 5.119$  و چون این مدل در تمام فضای نمونه برداری معتبر است تابع اعتبار آن  $\Phi_1 = 1$  است. در گام بعدی با توجه به اینکه تنها یک متغیر وجود دارد فضا به دو نیم تقسیم می شود (از مقدار  $x=0.5$ ). حالا در هر دو نیمه دو مدل خطی مستقل از هم ساخته می شود. مدل های ساخته شده عبارتند از LLM 2-1:  $y = -12.755x + 6.793$  و LLM 2-2:  $y = -1.464x + 2.312$  و مقدار ضریب رگرسیون که نشان دهنده خطا برای مدل ها است به ترتیب برابر  $R_{2-1}^2 = 0.8039$  و  $R_{2-2}^2 = 0.9749$  است. با توجه به مقدار پارامتر  $R^2$  بدترین مدل موجود LLM 2-1 است پس در گام بعدی LLM 2-1 تقسیم می شود. حاصل تقسیم بندی در این گام ۳ مدل است که معادلات آن به این صورت است: LLM 3-1:  $y = -25.872x + 8.262$  و LLM 3-2:  $y = -4.729x + 3.930$  و LLM 3-3:  $y = -1.464x + 2.312$  و ضریب

رگرسیون آن‌ها عبارتند از:  $R_{3-1}^2=0.897$ ,  $R_{3-2}^2=0.981$  و  $R_{3-3}^2=0.975$ . پیشروی این روند تا مرحله

هفتم در شکل ۸-۴ آمده است.



شکل ۴-۸ نحوه حل مسئله یک بعدی با استفاده از الگوریتم لولیموت برای ۷ گام ابتدایی

معادله خطوط مدل‌های محاسبه شده در هر مرحله به صورت  $y_i = a_i x + b_i$  است که مقدار وزن‌های محاسبه شده برای مثال آورده شده در مراحل پیشروی در جدول ۱-۴ آورده شده است.

جدول ۱-۴ وزن‌های محاسبه شده برای LLMها در هر مرحله

شماره مرحله	پارامتر $a_i$	پارامتر $b_i$	شماره مرحله	پارامتر $a_i$	پارامتر $b_i$	
۱	-۵,۳۸۱۰	۵,۱۱۹۵	۶	-۱۲,۷۷۵	۶,۷۹۳۴	
	-۱,۴۶۳۹	۲,۳۱۱۷		-۴,۷۲۸۵	۳,۹۲۹۵	
	-۲۵,۸۷۱۷	۸,۲۶۱۶		-۱,۴۶۳۹	۲,۳۱۱۷	
۲	-۱۲,۷۷۵	۶,۷۹۳۴		-۴,۷۲۸۵	۳,۹۲۹۵	
	-۱,۴۶۳۹	۲,۳۱۱۷		-۱,۴۶۳۹	۲,۳۱۱۷	
	-۴,۷۲۸۵	۳,۹۲۹۵		-۱,۴۶۳۹	۲,۳۱۱۷	
۳	-۴,۷۲۸۵	۳,۹۲۹۵		۷	-۴۳,۳۸۰۲	۹,۲۳۲۴
	-۱,۴۶۳۹	۲,۳۱۱۷			-۱۲,۶۴۳۹	۵,۹۰۰۸
	-۲۵,۸۷۱۷	۸,۲۶۱۶			-۴,۷۲۸۵	۳,۹۲۹۵
۴	-۴۳,۳۸۰۲	۹,۲۳۲۴	-۱,۴۶۳۹		۲,۳۱۱۷	
	-۱۲,۶۴۳۹	۵,۹۰۰۸	-۶۴,۷۴۶۳		۹,۹۷۳۴	
	-۴,۷۲۸۵	۳,۹۲۹۵	-۲۶,۶۵۳۱		۷,۷۰۰۱	
۵	-۱,۴۶۳۹	۲,۳۱۱۷	-۱۲,۶۴۳۹		۵,۹۰۰۸	
	-۴,۷۲۸۵	۳,۹۲۹۵	-۱,۰۳۳	۱,۹۶۲۱		
	-۶۴,۷۴۶۳	۹,۹۷۳۴				
	-۲۶,۶۵۳۱	۷,۷۰۰۱				
	-۱۲,۶۴۳۹	۵,۹۰۰۸				

## ۳-۴ آماده سازی داده‌های واقعی

Equation Section 5 پس از حذف داده‌های زاید و انتخاب چاه نمودارهای مناسب که مراحل آن در بخش ۳-۴-۱ آورده شده، برای تخمین مقدار TOC می‌بایست داده‌ها مورد پردازش قرار گیرد. در این بخش به مراحل پردازش پرداخته می‌شود و مشکلات موجود در این مرحله و راه کارهای ارائه شده برای رفع مشکلات آورده شده است.



## ۱-۳-۴ هماهنگ سازی داده‌ها

پس از آماده سازی داده‌ها زمان پردازش داده‌ها فرا رسید. پس از آماده سازی داده‌ها، داده‌های خام (بدون هیچ پیش پردازشی بر روی داده‌ها) توسط الگوریتم لولیموت مورد پردازش قرار گرفت. اما با شناختی که نسبت به عملکرد روش بر روی داده‌ها بدست آمده بود، به وجود آمدن خطای بسیار زیاد پیش‌بینی شده بود. این خطای بسیار زیاد به دلیل ساختار داده‌ها و نحوه پردازش الگوریتم لولیموت است. چون داده‌ها ساختاری توده ای دارند و الگوریتم لولیموت در تخمین و پردازش داده‌هایی با شکل توده ای ناتوان است. به دلیل مشکل به وجود آمده باید بر روی داده‌ها عملیات پیش پردازش انجام گیرد که در ادامه توضیح داده شده است.

اولین گام برای پیش پردازش داده‌ها هماهنگ سازی داده است تا داده‌ها در تمام چاه‌ها هماهنگ گردد. برای هماهنگ سازی داده‌ها ابتدا داده‌های خارج از ردیف برای هر یک از چاه نمودارها و مقادیر اندازه گیری شده TOC از میان داده‌ها حذف شدند. روش استفاده شده برای شناسایی داده‌های خارج از ردیف ترکیبی از روش‌های آزمون مقادیر خارج از ردیف برای داده‌هایی با گسترش زیاد و روش فانوپ<sup>۱</sup> است (حسنی‌پاک and شرف‌الدین ۱۳۸۴).

پس از تعیین و حذف داده‌های خارج از ردیف، داده‌های باقی مانده استاندارد شدند تا در مراحل بعدی که شامل استفاده از روش‌های ارائه شده است نتایج بهتری حاصل شود (حسنی‌پاک and شرف‌الدین ۱۳۸۴). روش‌های استانداردسازی متفاوتی وجود دارد که نتایج آن‌ها تفاوت چندانی با هم ندارند چرا که در نهایت همه داده‌ها در محدوده [۰،۱] قرار می‌گیرند. برای استانداردسازی داده‌ها از رابطه زیر استفاده شد:

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad ۱-۴$$

<sup>۱</sup> Funop

که در آن  $z_i$  مقدار استاندارد شده یک داده،  $x_i$  مقدار واقعی داده،  $x_{\min}$  کمترین مقدار برای یک ویژگی و  $x_{\max}$  بیشترین مقدار برای یک ویژگی است (Bishop 1995). لازم به ذکر است که در استانداردسازی داده‌ها به این شکل عمل شد که عملیات استانداردسازی داده‌های مربوط به یک چاه نمودار برای تمام چاه‌ها با هم انجام شد تا مقیاس تمام داده‌ها یکسان باشد که در زمان تحلیل نتایج دید یکسانی بدست آید.

پس از حذف و استانداردسازی داده‌ها، الگوریتم لولیموت دوباره بر روی داده‌ها اجرا شد ولی باز هم نتیجه مطلوبی حاصل نشد که به علت اختلاط زون‌ها و سازندها بود.

## ۲-۳-۴ خوشه‌بندی و کلاسه‌بندی داده‌ها

پس از استانداردسازی داده‌ها، داده‌ها با روش‌های مختلف از جمله میانگین  $K$  و گوستافسون-کسل خوشه‌بندی شدند ولی با توجه به نتایج ضعیف این روش‌ها در خوشه‌بندی داده‌ها تصمیم گرفته شد که از روش‌های دیگر استفاده شود. در این بین روش خوشه‌بندی DBSCAN با توجه به کارایی حاصل برای داده‌های مصنوعی انتخاب شد. استفاده از روش DBSCAN نیز با توجه به ساختار داده‌ها نسبت به دیگر روش‌ها کارایی بهتری داشت ولی باز هم نتیجه چشم‌گیر نبود.

پس از آزمودن تمام روش‌های گفته شده برای خوشه‌بندی داده‌ها و عدم حصول نتیجه مطلوب، تصمیم گرفته شد که به جای خوشه‌بندی داده‌ها با استفاده از روش‌های گفته شده، سازندهای بدست آمده از آنالیزهای انجام شده بر روی مغزه‌ها به عنوان کلاس در نظر گرفته شود و سایر داده‌ها به وسیله روش کلاسه‌بندی ارائه شده در بخش ۲-۴-۲ کلاسه‌بندی شدند.

در استفاده از روش ارائه شده برای کلاسه‌بندی ورودی‌ها عبارتند از: داده‌های بدون نشانگر سازند به عنوان داده‌هایی که هدف کلاسه‌بندی آن‌ها است، داده‌های با نشانگر سازند به عنوان داده‌هایی که برای آموزش روش به کار گرفته می‌شوند و پارامتر  $\epsilon$  در بازه  $[0.1, 0.25]$  به فاصله  $0.05$ .

انتخاب شد، بدین معنی که کلاسه‌بندی چندین مرتبه با مقادیر مختلف  $\varepsilon$  انجام شد تا بهترین نتیجه انتخاب شود. معیار انتخاب بهترین نتیجه برای کلاسه‌بندی تعداد داده‌های نوفه تشخیص داده شده و حفظ نسبی برای داده‌های کلاسه‌بندی شده به نسبت داده‌های آموزشی هر سازند است.

نتیجه حاصل از کلاسه‌بندی داده‌های مربوط به هر چاه در جدول ۴-۲ تا جدول ۴-۱۰ آورده شده است.

جدول ۴-۲ نتیجه کلاسه‌بندی داده‌های چاه گیجرود ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۰	* ۹۳۴
ماسه سنگ گیج	۳	۱۱۵
لیدرویل	۴	۱۲۶۶
ماسه سنگ لزوئر	۰	* ۳۳۴
پارملیا	۰	* ۲۴
شیل پرت جنوبی	۷	۲۵۵۴
یاراگادی	۳۰	۱۲۴۴۵
مجموع	۴۴	۲۰۶۷۲

جدول ۴-۳ نتیجه کلاسه‌بندی داده‌های چاه گیجرود ۲

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۰	* ۰
ماسه سنگ گیج	۰	* ۲۲۶
لیدرویل	۰	* ۰
ماسه سنگ لزوئر	۰	* ۱
پارملیا	۰	* ۲۴۷۰
شیل پرت جنوبی	۳	۱۷۷۵
یاراگادی	۲	۳۵۹۲
مجموع	۵	۸۰۶۴

جدول ۴-۴ نتیجه کلاس‌بندی داده‌های چاه لیک پرینستون ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاس‌بندی شده
کاکلیشیل گولی	۴	۱۸۰
ماسه سنگ گیج	۰	* ۳۴
لیدرویل	۰	* ۱۰۵
ماسه سنگ لزوئر	۱۸	۵۵۲۹
پارملیا	۰	* ۷۷
شیل پرت جنوبی	۰	* ۴۶۶
پاراگادی	۰	* ۸۰۷۰
مجموع	۲۲	۱۴۴۶۱

جدول ۵-۴ نتیجه کلاس‌بندی داده‌های چاه مولالو ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاس‌بندی شده
کاکلیشیل گولی	۰	* ۶۳
ماسه سنگ گیج	۱۴	۹۸۸
لیدرویل	۰	* ۹۲
ماسه سنگ لزوئر	۰	* ۹۷
پارملیا	۱۷	۲۸۳۷
شیل پرت جنوبی	۱۷	۲۸۶۴
پاراگادی	۰	* ۱۳۹
مجموع	۴۸	۷۰۸۰

جدول ۴-۶ نتیجه کلاسه‌بندی داده‌های چاه پل ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۰	* ۷۰۹
ماسه سنگ گیج	۰	* ۲۶۸
لیدرویل	۰	* ۵۱
ماسه سنگ لزوئر	۰	* ۵۵۰
پارملیا	۰	* ۳۱
شیل پرت جنوبی	۴	۱۳۴۸
پاراگادی	۴	۱۰۷۳۳
مجموع	۸	۱۳۶۹۰

جدول ۴-۷ نتیجه کلاسه‌بندی داده‌های چاه کویینزراک ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۰	* ۸۰
ماسه سنگ گیج	۰	* ۹۹
لیدرویل	۰	* ۷۶۰
ماسه سنگ لزوئر	۰	* ۲۶۵
پارملیا	۰	* ۲۶
شیل پرت جنوبی	۰	۸۰۲
پاراگادی	۱۴	۶۸۲۳
مجموع	۴۴	۸۸۵۵

جدول ۴-۸ نتیجه کلاسه بندی داده های چاه راکینگهام ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه بندی شده
کاکلیشیل گولی	۱۴	۱۸۰۹
ماسه سنگ گیج	۰	* ۲۰
لیدرویل	۰	* ۹۴
ماسه سنگ لزوئر	۰	* ۴۴
پارملیا	۰	* ۰
شیل پرت جنوبی	۰	* ۳۳۵
پاراگادی	۰	* ۴۳۰۸
مجموع	۱۴	۶۶۱۰

جدول ۴-۹ نتیجه کلاسه بندی داده های چاه شوگرلوف ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه بندی شده
کاکلیشیل گولی	۰	* ۶۹۶
ماسه سنگ گیج	۰	* ۱۵۵
لیدرویل	۰	۳۵۹
ماسه سنگ لزوئر	۰	* ۱۷۲
پارملیا	۰	* ۱۱
شیل پرت جنوبی	۰	* ۲۴۹
پاراگادی	۱۲	۱۵۶۸۲
مجموع	۱۲	۱۷۳۲۴

جدول ۴-۱۰ نتیجه کلاسه‌بندی داده‌های چاه وارنبرو ۱

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۰	۷۹۱ *
ماسه سنگ گیج	۱	۴۹
لیدرویل	۶	۲۶۲۷
ماسه سنگ لزوئر	۰	۵۴۰ *
پارملیا	۰	۳۱ *
شیل پرت جنوبی	۱۳	۳۵۸۲
پاراگادی	۲۴	۱۱۹۶۵
مجموع	۴۴	۱۹۵۸۵

پس از بررسی نتایج و با توجه به اینکه در آنالیزهای انجام شده در هر چاه بعضی سازندها دیده نشده‌اند اما در زمان خوشه‌بندی با توجه به روش اعمال شده و با توجه به اینکه تمام داده‌ها همزمان برای کلاسه‌بندی استفاده شده‌اند، برخی کلاس‌ها بعد از پردازش حذف شده‌اند تا در داده‌ها به هماهنگی بیشتری دست یافته شود. کلاس‌های حذف شده در داخل جدول مربوط به هر چاه با علامت \* مشخص شده‌اند.

در مجموع تعداد داده‌های آموزشی و داده هدف برای کلاسه‌بندی در جدول ۴-۱۱ آمده است.

جدول ۴-۱۱ مجموع داده‌های مورد استفاده برای پردازش

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۱۸	۵۴۶۲
ماسه سنگ گیج	۱۸	۱۹۵۴
لیدرویل	۱۰	۵۳۵۴
ماسه سنگ لزوئر	۱۸	۷۵۳۲
پارملیا	۱۷	۵۵۰۵
شیل پرت جنوبی	۴۴	۱۳۹۷۵
پاراگادی	۸۶	۷۶۷۵۷
مجموع	۲۱۱	۱۱۶۳۴۱

با توجه به جدول ۴-۱۱ می‌توان گفت که به علت تعداد زیاد داده‌های آموزشی برای سازند یاراگادی تعداد داده‌هایی که هنگام کلاسه‌بندی در این کلاس قرار گرفته‌اند نیز زیاد است که نتیجه ای قابل پیش‌بینی است.

پس از حذف مواردی که در جدول ۴-۲ تا جدول ۴-۱۰ با علامت \* مشخص شده‌اند، جدول ۴-۱۱ به جدول ۴-۱۲ تبدیل شد.

جدول ۴-۱۲ مجموع داده‌های اصلاح شده برای پردازش

نام سازند	تعداد داده آموزشی	تعداد داده کلاسه‌بندی شده
کاکلیشیل گولی	۱۸	۱۹۸۹
ماسه سنگ گیج	۱۸	۱۱۵۲
لیدرویل	۱۰	۴۲۵۲
ماسه سنگ لزوئر	۱۸	۵۵۲۹
پارملیا	۱۷	۲۸۳۷
شیل پرت جنوبی	۴۴	۱۲۹۲۵
یاراگادی	۸۶	۶۴۲۴۰
مجموع	۲۱۱	۹۲۹۲۴

پس از این پیش‌پردازشی که بر روی داده‌ها انجام شد، داده‌ها آماده پردازش به وسیله الگوریتم تخمینی لولیموت شدند.

## ۴-۴ اجرای الگوریتم لولیموت بر روی داده‌ها

در متون مرجع مرتبط با الگوریتم‌های هوشمند و روش‌های طبقه‌بندی (Duda, Hart Bishop 1995, et al. 1995) نسبت ۷۰٪ به ۳۰٪ به ترتیب برای آموزش و آزمون انتخاب گردیده است. بنابراین در پروژه حاضر نیز با توجه به منابع مذکور از این نسبت استفاده شده است. پس از اعمال این تقسیم بندی بر روی داده‌ها نتایج به صورت مندرج در جدول ۴-۱۳ در می‌آید.



جدول ۴-۱۳ نتیجه تقسیم داده‌ها در چاه‌های موجود بر اساس نسبت ۷۰٪ به ۳۰٪ به ترتیب برای آموزش و آزمون

نام سازند	تعداد داده‌های هر چاه	آموزش	آزمون
کاکلیشیل گولی	۱۸	۱۲	۶
ماسه سنگ گیج	۱۸	۱۲	۶
لیدرویل	۱۰	۷	۳
ماسه سنگ لزوئر	۱۸	۱۲	۶
پارملیا	۱۷	۱۱	۶
شیل پرت جنوبی	۴۴	۳۰	۱۴
یاراگادی	۸۶	۶۰	۲۶

پس از انجام تمام مراحل پیش پردازش داده که انجام شد، اکنون داده‌ها آماده اجرای الگوریتم لولیموت هستند. با توجه به اینکه ترتیب اولیه داده‌ها به گونه ای است که در آموزش لولیموت تأثیر می‌گذارند باید آن‌ها را بر زده تا ترکیب بهتری حاصل شود.

لازم به ذکر است که الگوریتم لولیموت به ترتیب ورود داده‌ها وابسته نیست. در واقع علت بُر زدن داده‌ها این است که با توجه به کمبود داده‌های آموزش (در اکثر موارد کمتر از ۲۰ داده آموزش) و نیاز به سنجش نتیجه الگوریتم بر روی داده‌ها باید تعدادی داده برای سنجش استفاده شود. حال اینکه کدام داده‌ها برای آموزش باشد و کدام داده‌ها برای آزمون علت بُر زدن داده‌ها است.

در نتیجه روند کار به این شکل شد که در هر بار داده‌ها ابتدا بر زده می‌شوند، سپس الگوریتم بر روی داده‌های بُر زده شده اجرا می‌شود. ملاک مقایسه بین خروجی‌ها میزان خطای حاصل برای داده‌های آموزش و آزمون است. در بخش بعد نتایج حاصل آورده شده و نمایش کاملی از خروجی‌ها و مقایسه آن‌ها آورده شده است.

## ۵-۴ نحوه انجام پروژه

تمام الگوریتم‌ها و برنامه‌های مورد استفاده به وسیله نرم افزار MATLAB 2011a تهیه شدند. پس از آماده سازی داده‌ها (حذف داده‌های خارج از ردیف و زاید و استانداردسازی) روشی که برای طبقه بندی استفاده شد، روش کلاسه‌بندی DBSCAN بود که در بخش ۲-۴-۲ کاملاً توضیح داده شد. در استفاده از روش DBSCAN شعاع همسایگی ( $\epsilon$ ) دقیقی وجود ندارد. به همین دلیل روش DBSCAN با شعاع‌های ۰,۱، ۰,۱۵، ۰,۲ و ۰,۲۵ اعمال شد که نتیجه حاصل از طبقه بندی در جدول ۱۴-۴ آمده است که با توجه به شرایط موجود در نهایت شعاع ۰,۲ با سعی و خطا و با توجه به تعداد داده‌ی نوفه ایجاد شده انتخاب گردید.

جدول ۱۴-۴ نتیجه طبقه بندی داده‌ها با استفاده از روش DBSCAN با شعاع‌های مختلف

سازند/ شعاع اعمال شده	۰,۱	۰,۱۵	۰,۲	۰,۲۵
نوفه	۵۰۴۹۹	۲۱۱۲۹	۸۴۰۵	۲۵۰۱
کاکلیشیل گولی	۴۱۵۵	۶۰۷۱	۵۲۶۲	۳۸۴۹
ماسه سنگ گیج	۲۳۴۲	۲۲۳۵	۱۹۵۴	۱۵۸۲
لیدرویل	۳۲۴۷	۵۱۷۹	۵۳۵۴	۴۷۳۳
ماسه سنگ لزوئر	۹۵۸۵	۹۷۸۳	۷۵۳۲	۵۶۸۳
پارملیا	۲۸۲۶	۳۰۶۷	۵۵۰۷	۷۴۴۲
شیل پرت جنوبی	۸۰۲۶	۱۱۵۶۱	۱۳۹۷۵	۱۵۸۷۴
یاراگادی	۴۴۰۶۶	۶۵۷۲۱	۷۶۷۵۷	۸۳۰۸۲

پس از اعمال کلاسه‌بندی، داده‌ها به وسیله الگوریتم لولیموت مورد پردازش قرار گرفتند. با توجه به این نکته که تعداد بهینه LLM در دسترس نیست باید با استفاده از تکرار الگوریتم برای تعداد مختلف حداکثر LLM به تعداد بهینه دست یافت. نکته مهم دیگری که در مورد الگوریتم لولیموت باید در

نظر داشت این است که با توجه به خطی بودن این الگوریتم و تعداد کم داده‌های آموزش در اکثر موارد نمی‌توان تعداد زیاد مدل را اعمال کرد زیرا نتایج حاصل غیر قابل قبول می‌شود.

در اجرای الگوریتم لولیموت پارامترهای انتخاب شده برای الگوریتم عبارتند از شرط توقف

$$e=0.1, \text{ شرط توقف حداکثر تعداد مدل } 15 \text{ و } \sigma = \frac{1}{3}.$$

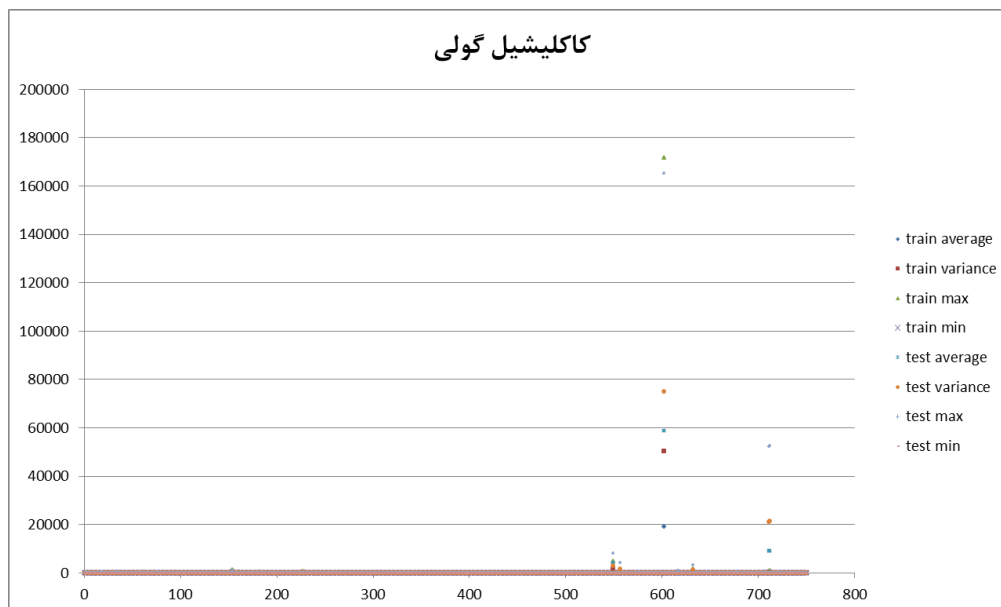
برای مقایسه نتایج در حالات مختلف ملاک را میزان میانگین خطا، واریانس خطا، بیشینه و

کمینه خطای موجود در قسمت آموزش و آزمون قرار داده می‌شود. شکل ۴-۹ تا شکل ۴-۱۵ میزان

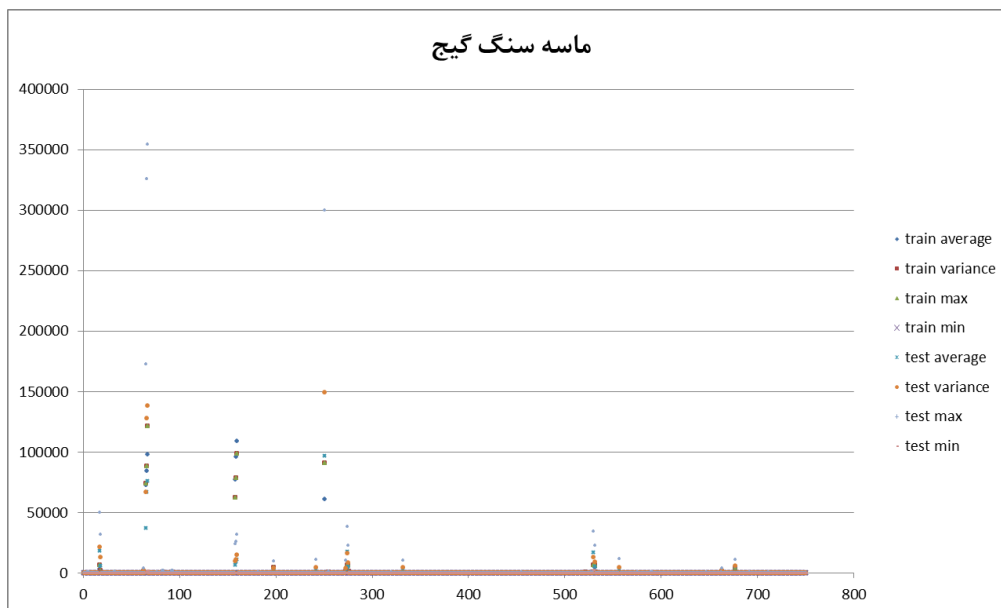
خطای پردازش هر یک از ۷ سازند انتخاب شده را برای تعداد مدل‌ها و حالات مختلف بر زدن نشان

می‌دهد. تعداد حداکثر مدل ۱۵ و تعداد حالات بر زدن ۵۰ است که در مجموع موجب به وجود آمدن

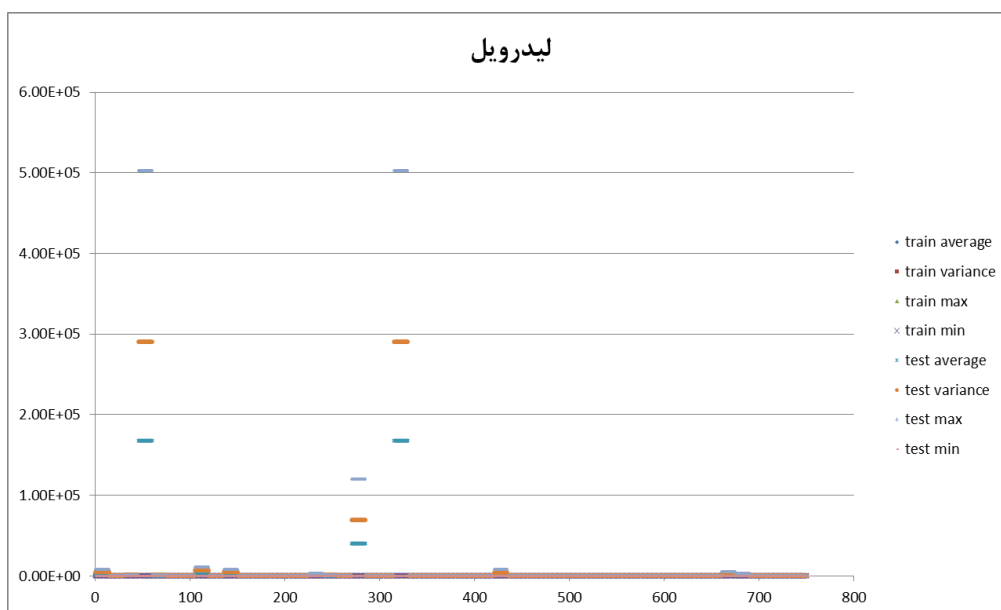
۷۵۰ حالت می‌شود.



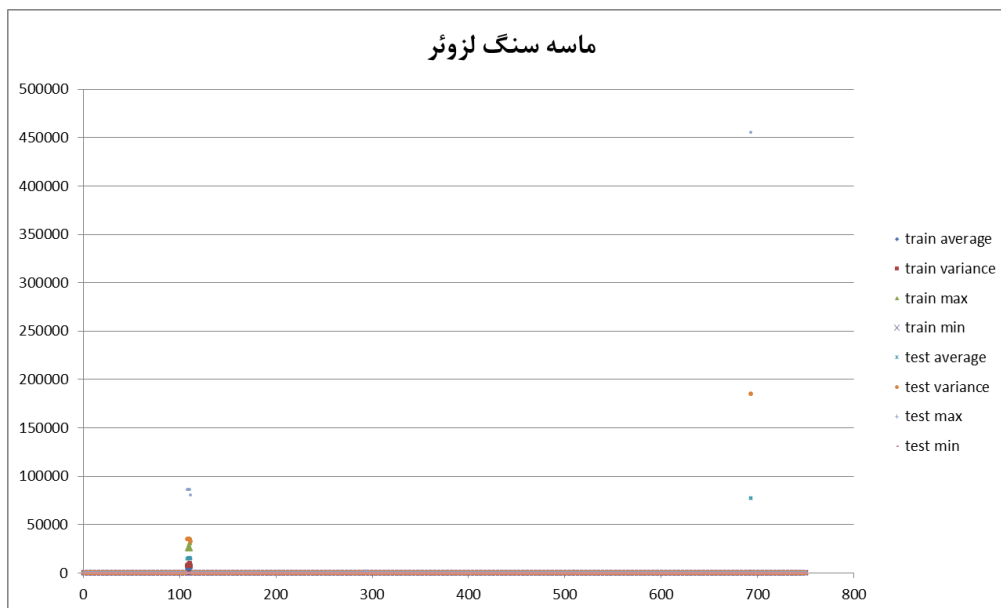
شکل ۴-۹ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند کاکلیشیل گولی



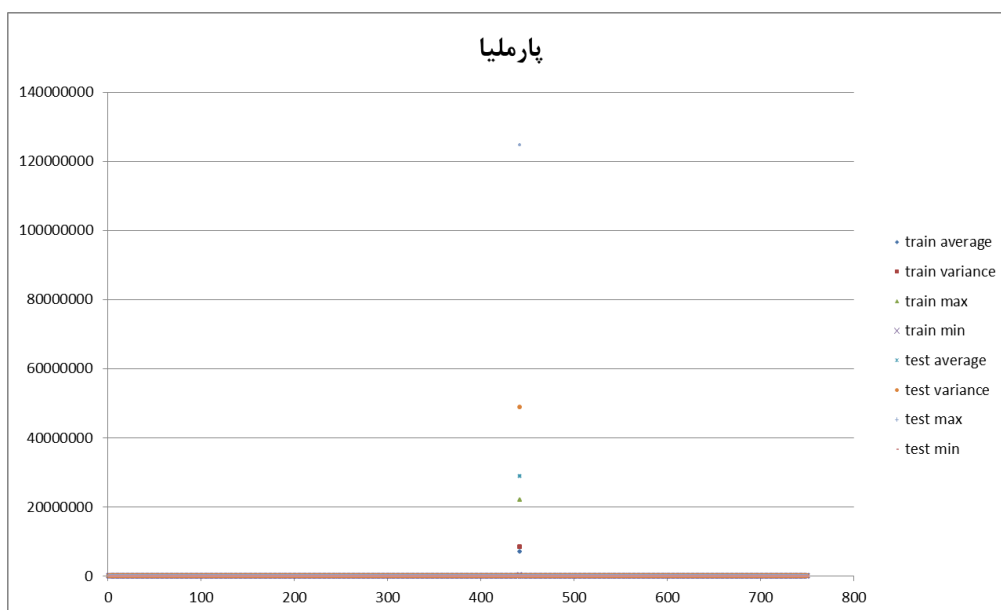
شکل ۴-۱۰ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ گيچ



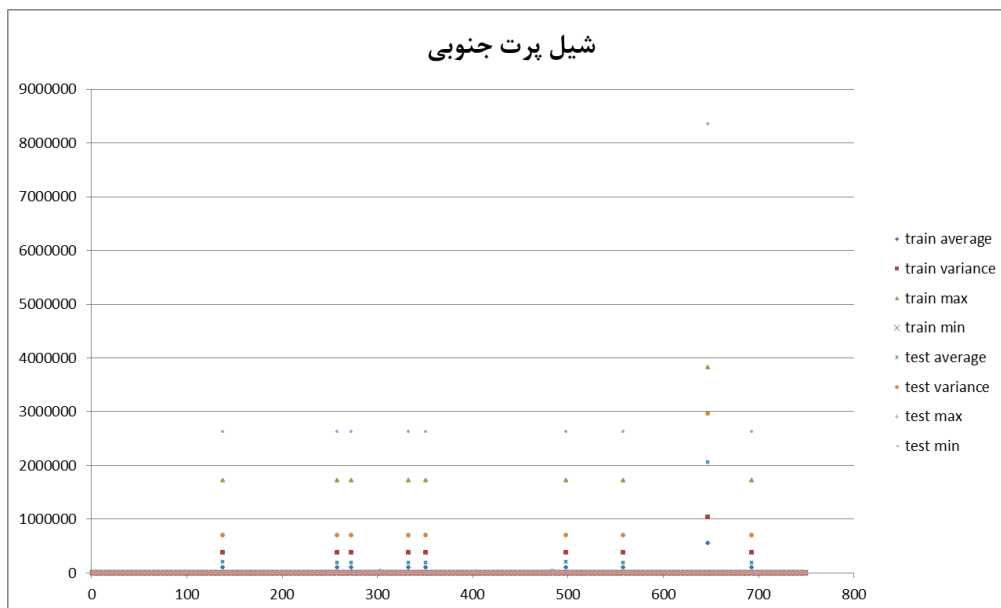
شکل ۴-۱۱ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ليدرويل



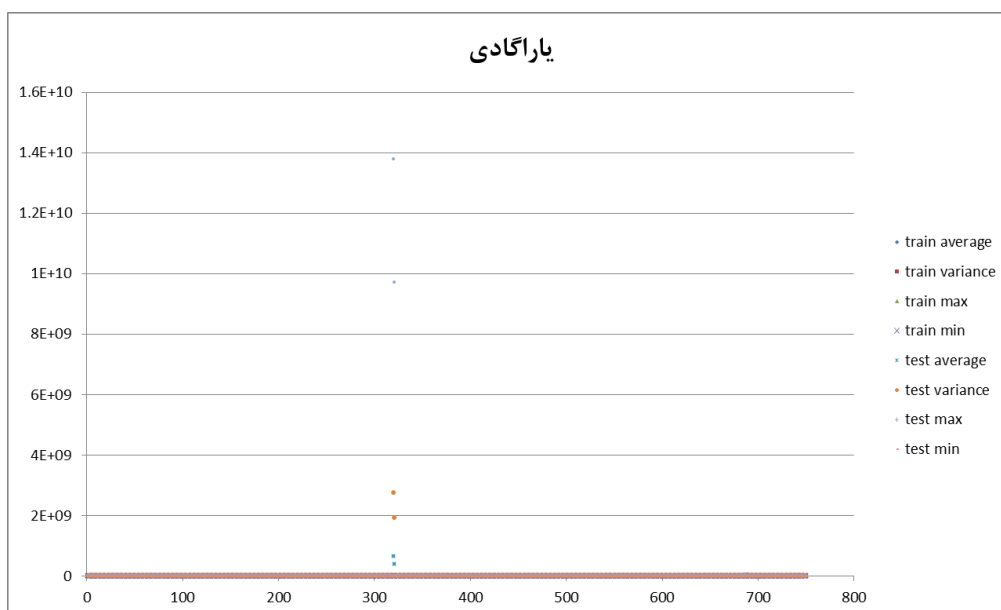
شکل ۴-۱۲ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ لزوئر



شکل ۴-۱۳ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند پارملیا



شکل ۴-۱۴ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند شیل پرت جنوبی

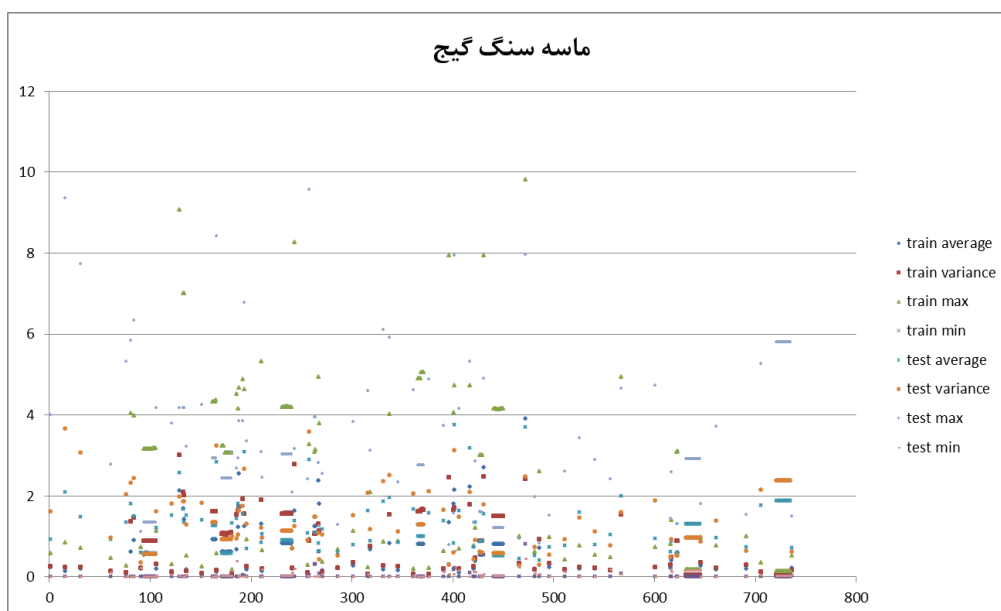


شکل ۴-۱۵ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند یاراگادی

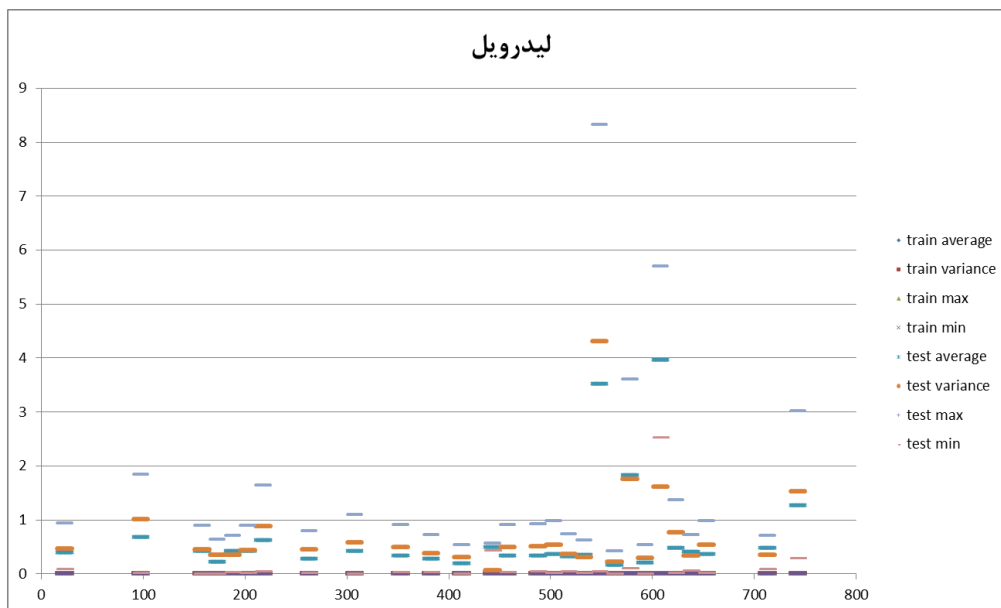
با توجه به مقدار بالای خطا در بعضی حالات اشکال حاصل مفهوم نبوده و خوانا نیستند و میزان خطا به دقت دیده نمی‌شود. برای حل این مشکل و وضوح بیشتر تصاویر، حالات نامطلوب با مقدار خطای زیاد از بین نتایج حاصل حذف شدند. در نتیجه شکل ۴-۹ تا شکل ۴-۱۵ به صورت شکل ۴-۱۶ تا شکل ۴-۲۲ تغییر می‌کنند.



شکل ۴-۱۶ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند کاکلیشیل گولی پس از حذف داده‌های زائد



شکل ۴-۱۷ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ گیج پس از حذف داده‌های زائد

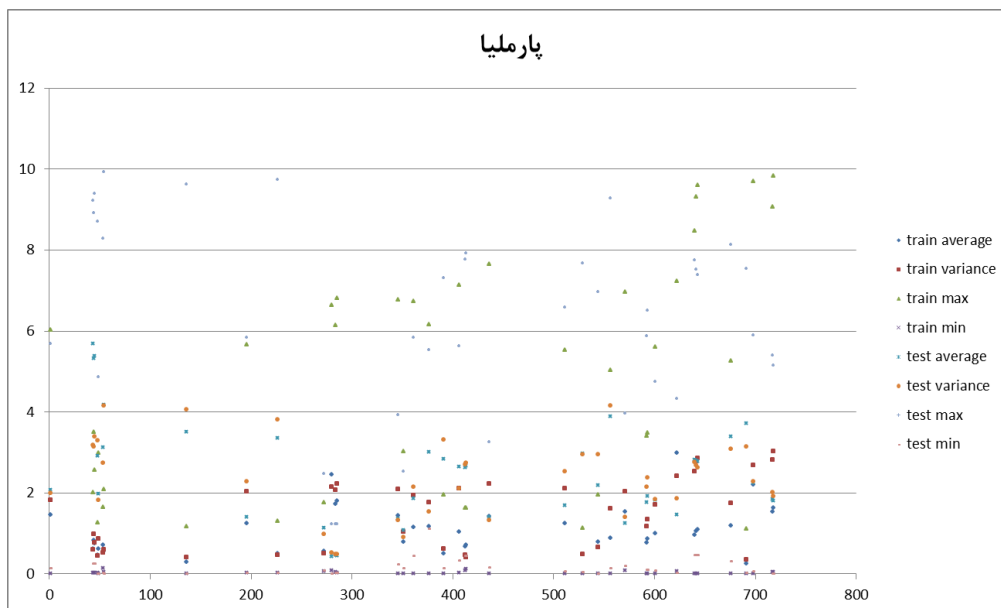


شکل ۴-۱۸ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند لیدرویل پس از حذف داده‌های زاید

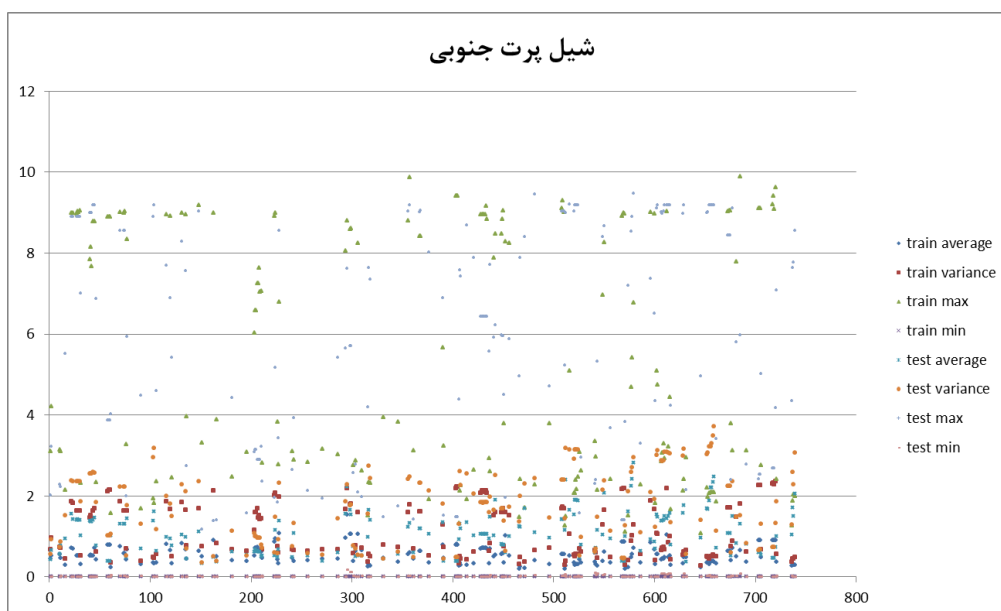


شکل ۴-۱۹ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند ماسه سنگ لزوئر پس از حذف داده‌های زاید

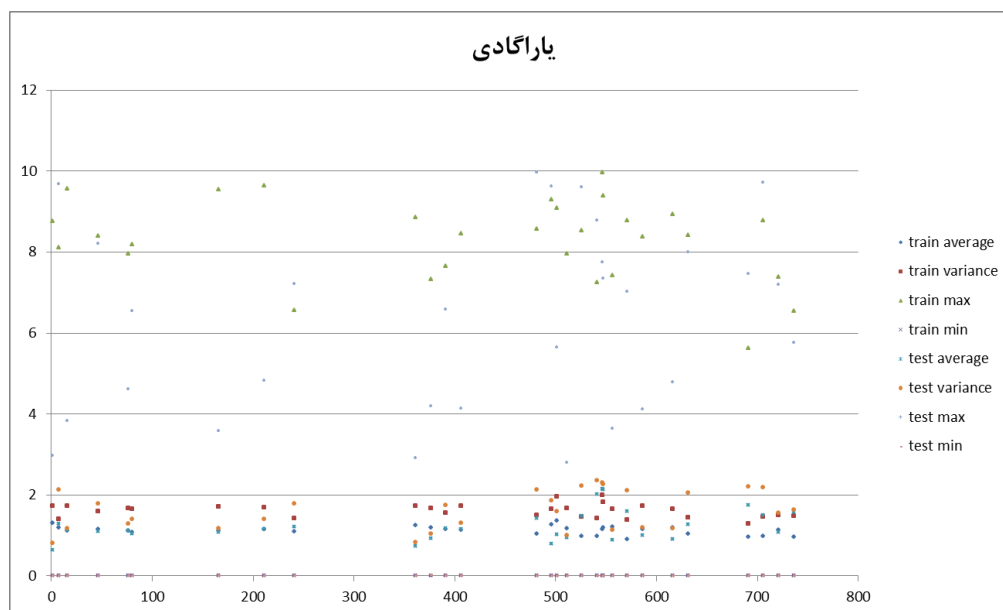




شکل ۴-۲۰ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند پارملیا پس از حذف داده‌های زاید



شکل ۴-۲۱ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند شیل پرت جنوبی پس از حذف داده‌های زاید

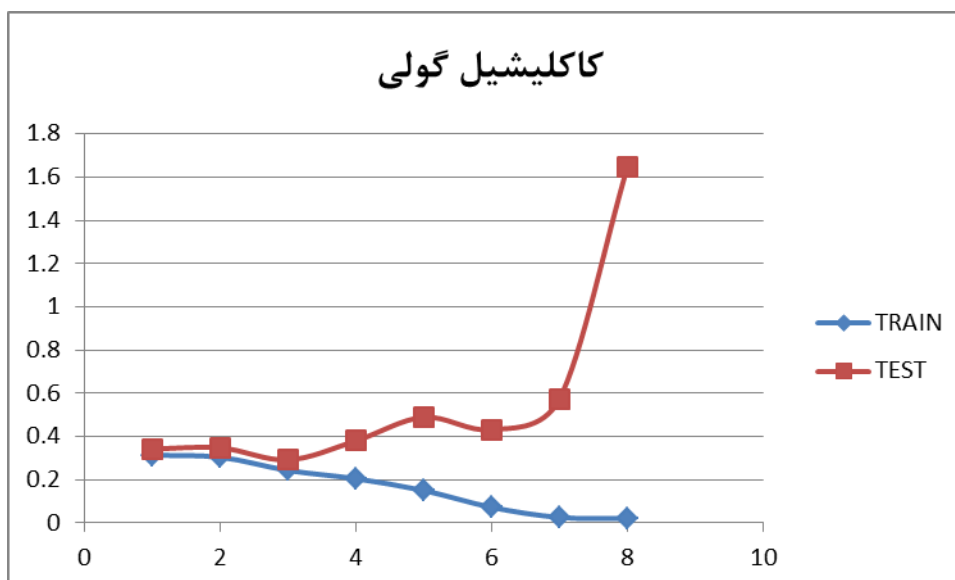


شکل ۴-۲۲ نتیجه پردازش داده‌ها با الگوریتم لولیموت بر روی سازند یاراگادی پس از حذف داده‌های زائد

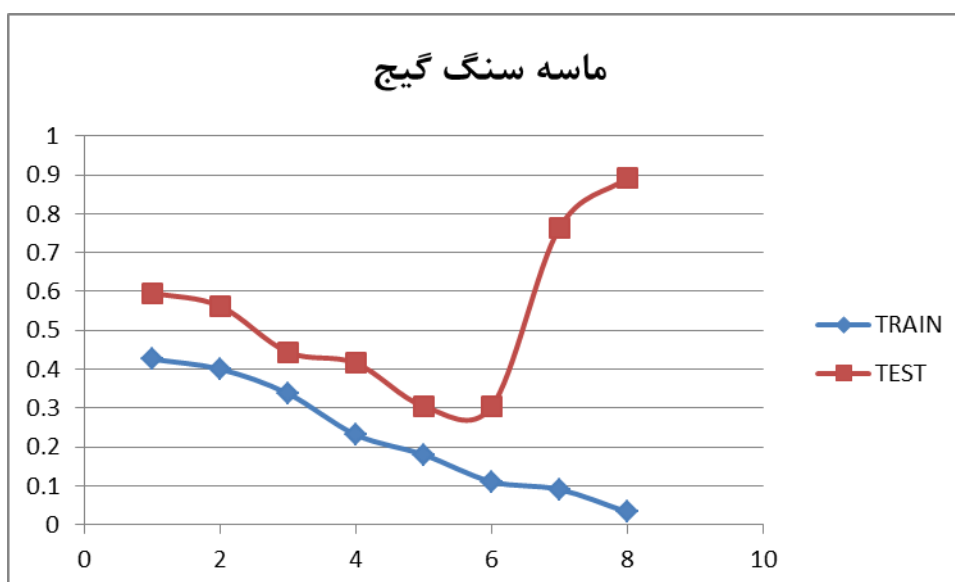
در فصول قبل الگوریتم‌های استفاده شده، نحوه انتخاب داده‌ها از بین تعداد زیاد داده موجود، موقعیت چاه‌های انتخابی، وضعیت سازندها، نحوه آماده سازی داده‌ها و نحوه پردازش داده‌ها مطالبی گفته شد. در این فصل نتایج به دست آمده تفسیر شده و مقادیر تخمینی پارامتر مورد نظر در هر یک از چاه‌ها آورده می‌شود.

## ۶-۴ روند آموزش و آزمون

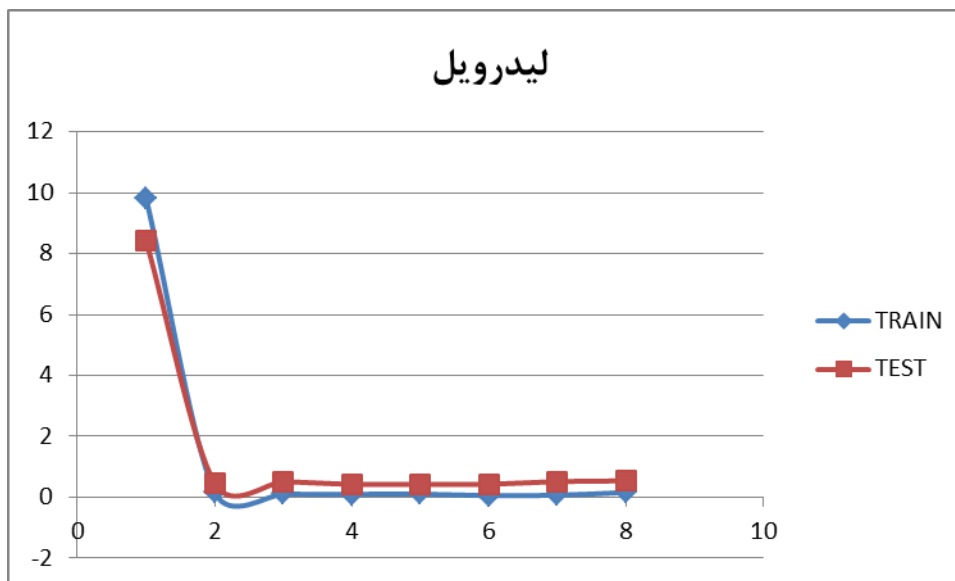
روند آموزش داده‌ها برای سازندهای مختلف در حالتی که ترتیب مناسب از بین حالات مختلف انتخاب شده در شکل ۴-۲۳ تا شکل ۴-۲۹ آورده شده است.



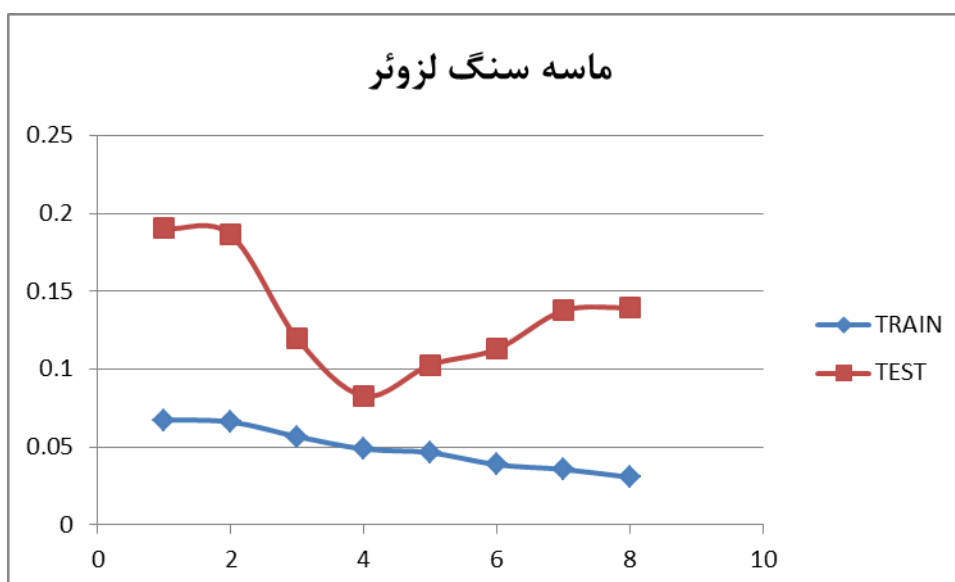
شکل ۴-۲۳ میزان خطای آموزشی و آزمون برای سازند کاکلیشیل گولی-حالت انتخاب شده ۳ مدل



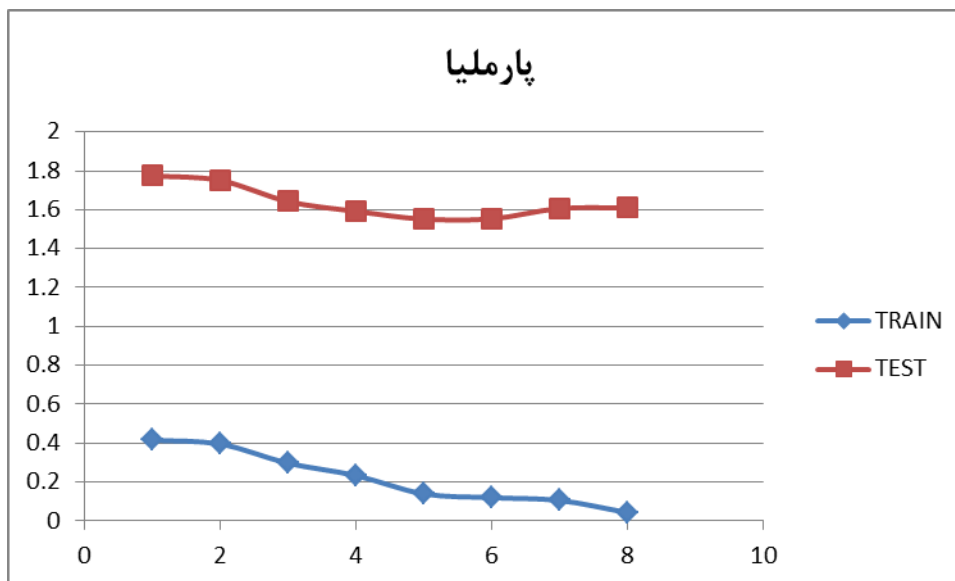
شکل ۴-۲۴ میزان خطای آموزشی و آزمون برای سازند ماسه سنگ گیج-حالت انتخاب شده ۶ مدل



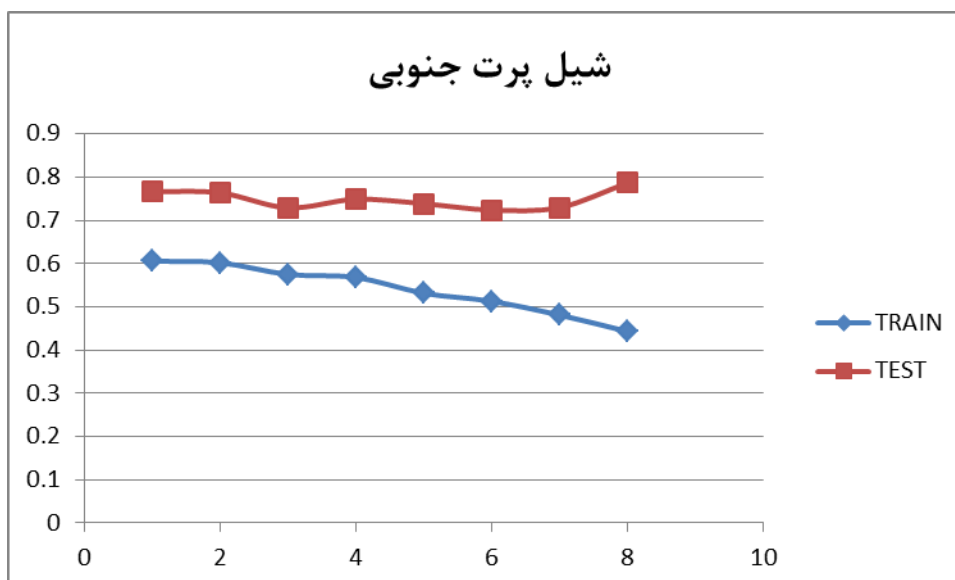
شکل ۴-۲۵ میزان خطای آموزشی و آزمون برای سازند لیدرویل-حالت انتخاب شده ۳ مدل



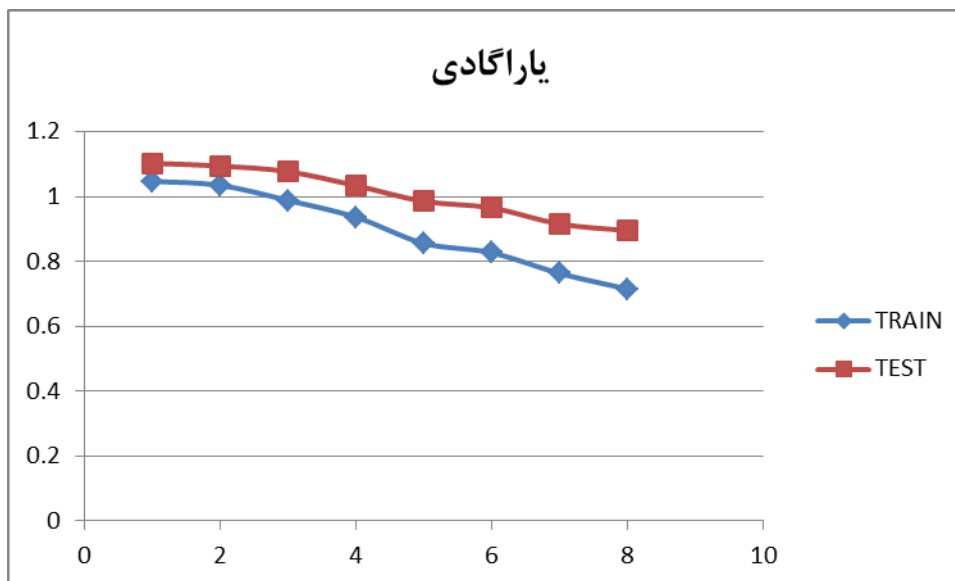
شکل ۴-۲۶ میزان خطای آموزشی و آزمون برای سازند ماسه سنگ لزوئر-حالت انتخاب شده ۴ مدل



شکل ۴-۲۷ میزان خطای آموزشی و آزمون برای سازند پارملیا-حالت انتخاب شده ۵ مدل

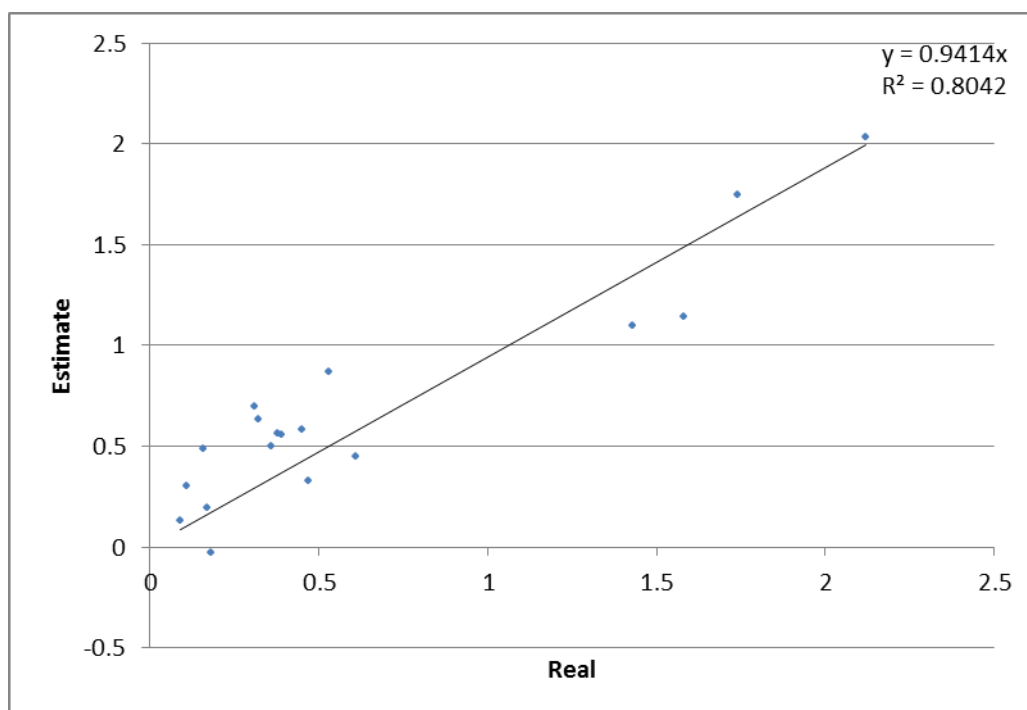


شکل ۴-۲۸ میزان خطای آموزشی و آزمون برای سازند شیل پرت جنوبی-حالت انتخاب شده ۶ مدل

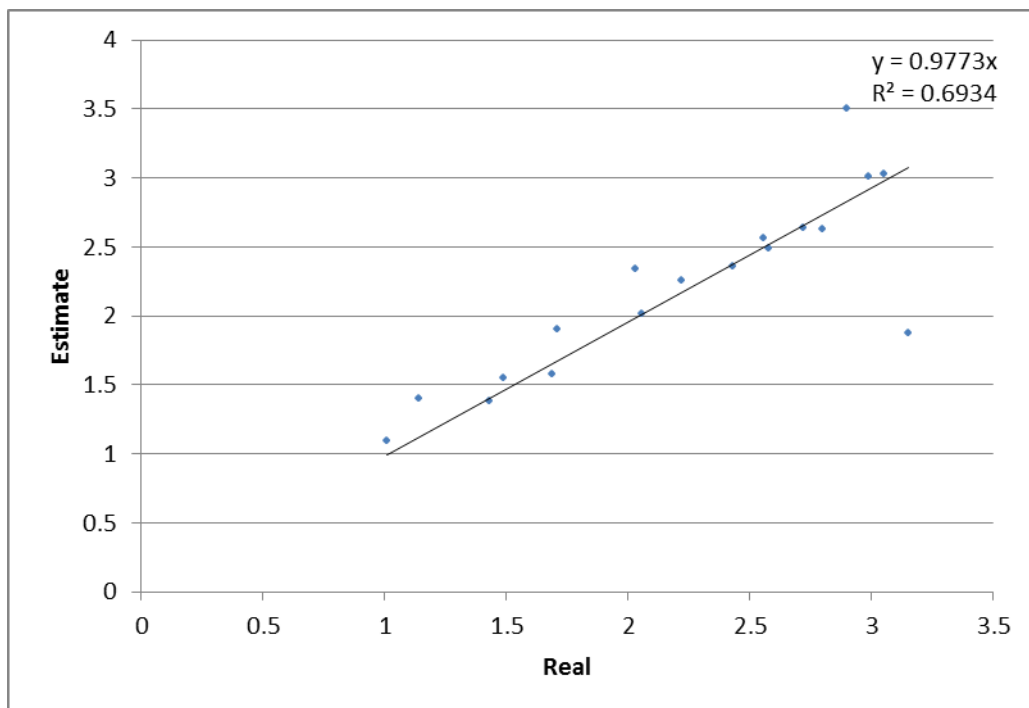


شکل ۴-۲۹ میزان خطای آموزشی و آزمون برای سازند یاراگادی-حالت انتخاب شده ۸ مدل

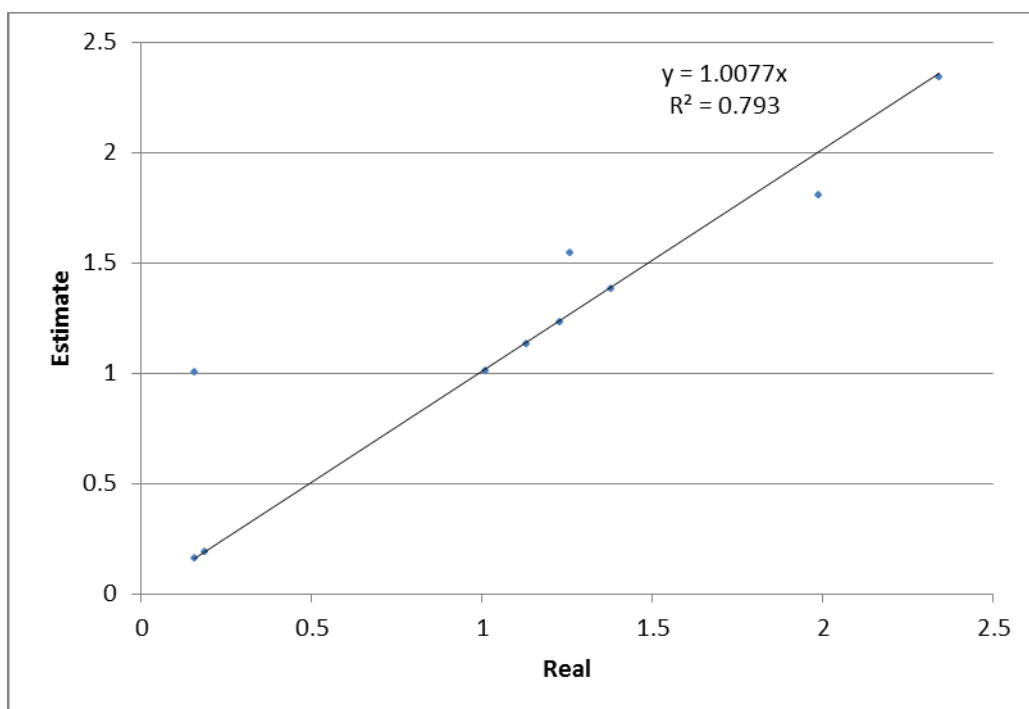
نمودارهای آورده شده در شکل ۴-۳۰ تا شکل ۴-۳۶ مقدار TOC اندازه گیری شده و تخمینی را برای هر یک از سازندها در حالت چینش انتخاب شده نشان می دهند.



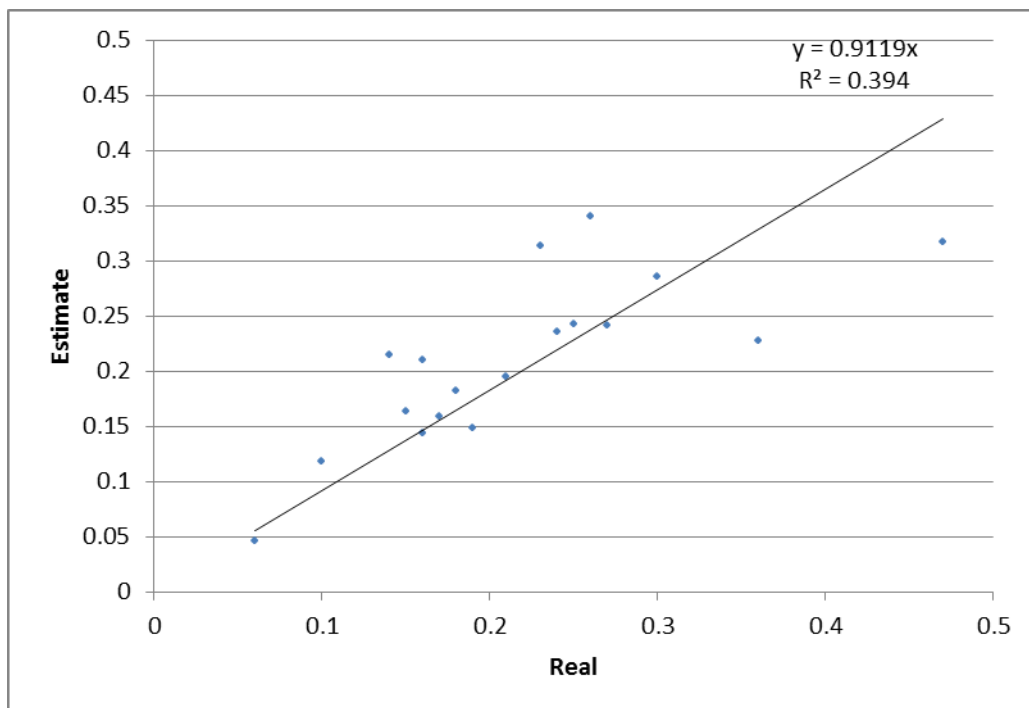
شکل ۴-۳۰ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند کاکلیشیل گولی



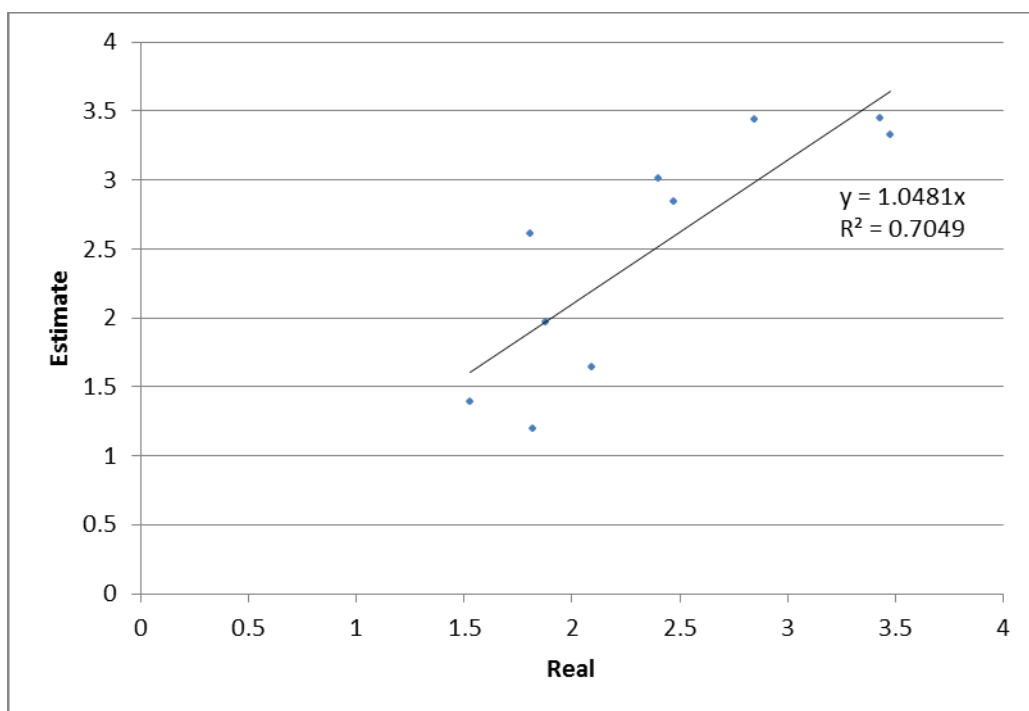
شکل ۳۱-۴ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند ماسه سنگ گیج



شکل ۳۲-۴ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند لیدرویل

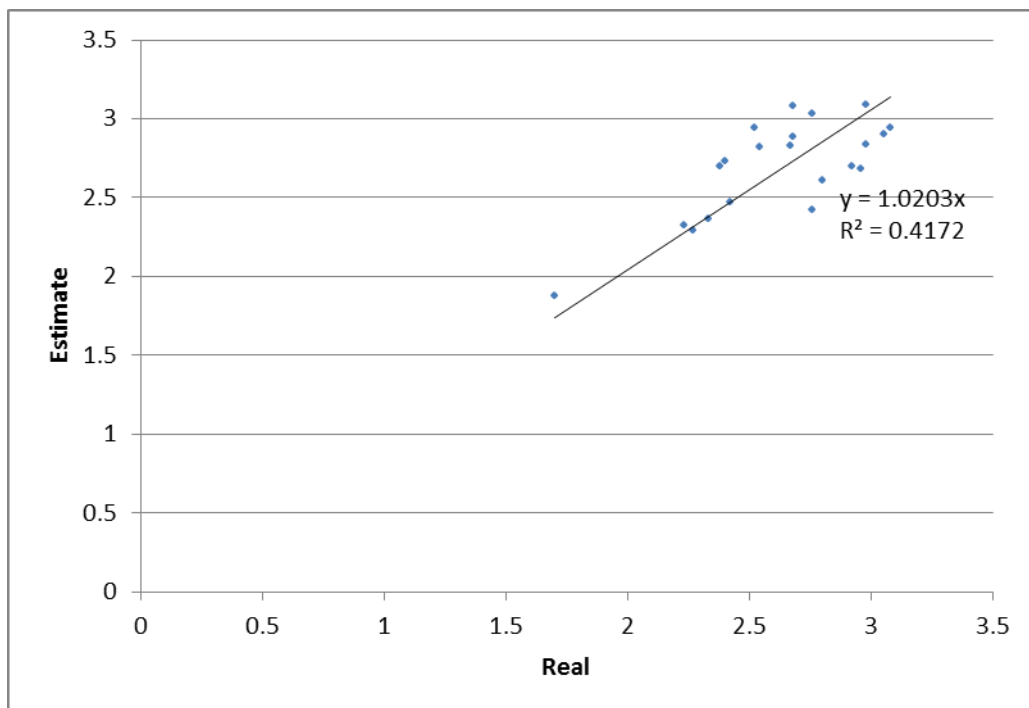


شکل ۴-۳ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند ماسه سنگ لزوئر

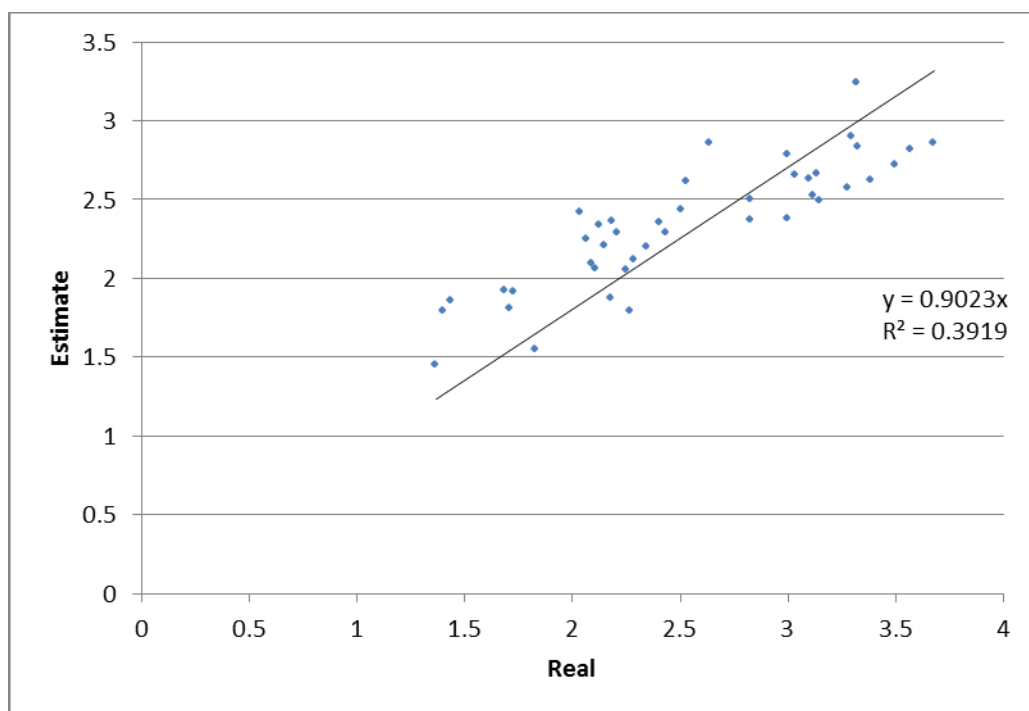


شکل ۴-۳ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند پارملیا





شکل ۴-۳۵ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند شیل پرت جنوبی



شکل ۴-۳۶ مقدار TOC اندازه گیری شده و تخمین زده شده برای سازند یاراگادی

خلاصه‌ای از میزان خطای آموزش و آزمون مدل‌های ساخته شده برای هر یک از سازندها در جدول

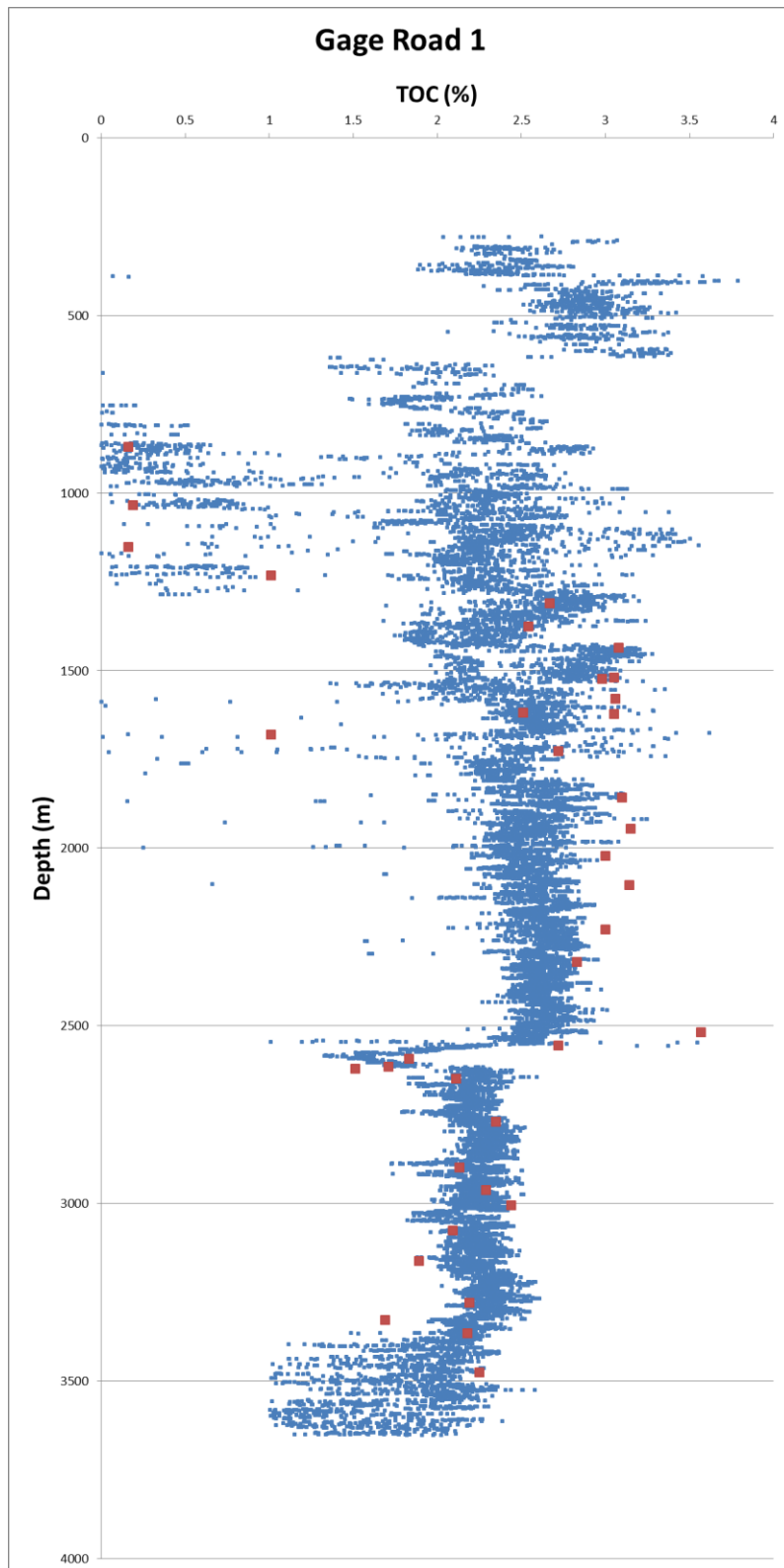
۴-۱۵ آورده شده است.

جدول ۴-۱۵ خلاصه ای از وضعیت آموزش و آزمون الگوریتم لولیموت برای سازندهای مورد مطالعه

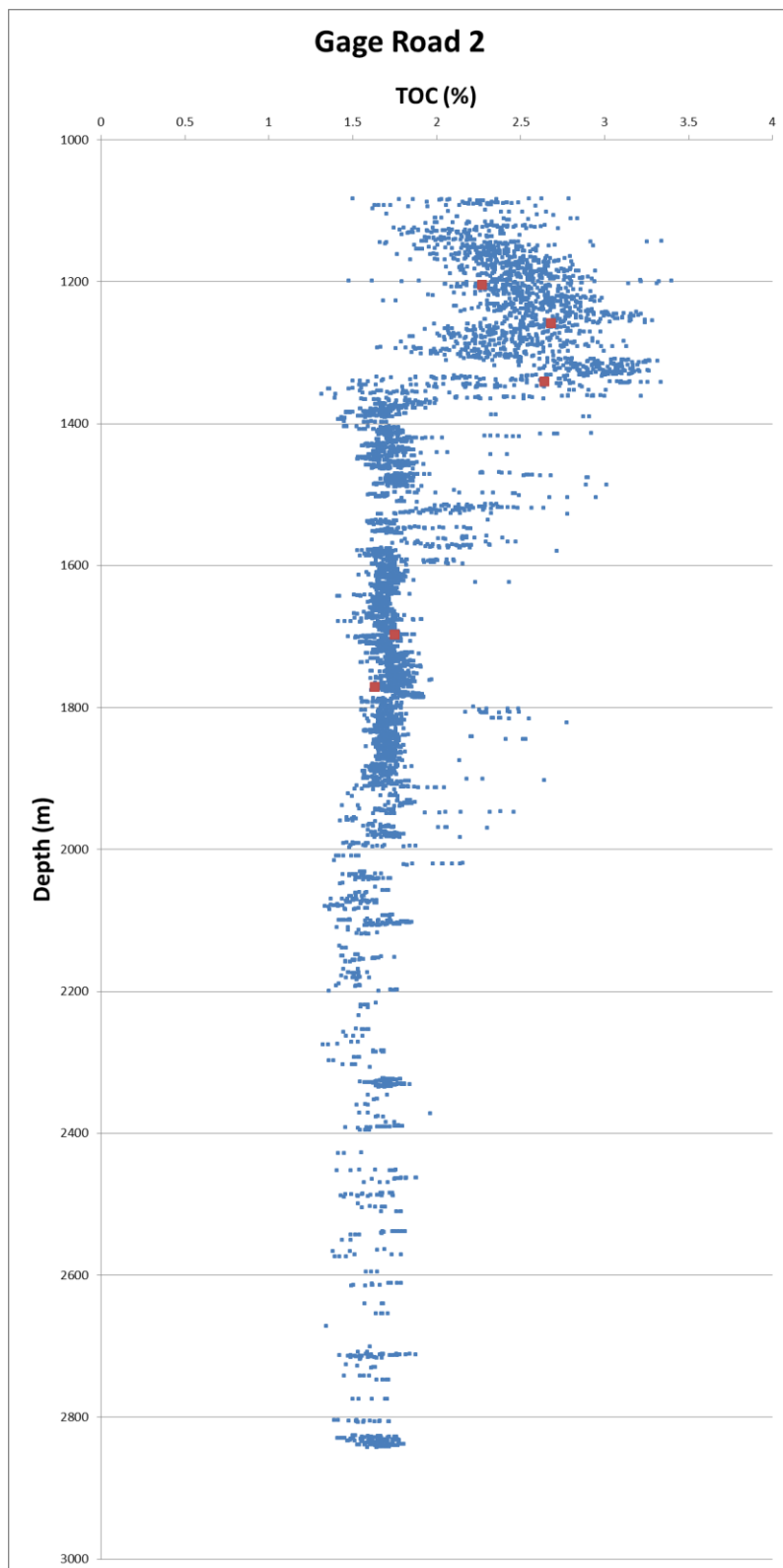
نام سازند	تعداد داده مغزه	تعداد داده آموزش	تعداد داده آزمون	مقدار $R^2$ (%)
کاکلیشیل گولی	۱۸	۱۲	۶	۸۰,۴۲
ماسه سنگ گیج	۱۸	۱۲	۶	۶۹,۳۴
لیدرویل	۱۰	۷	۳	۷۹,۳۰
ماسه سنگ لزوئر	۱۸	۱۲	۶	۳۹,۴۰
پارملیا	۱۷	۱۱	۶	۷۰,۴۹
شیل پرت جنوبی	۴۴	۳۰	۱۴	۴۱,۷۲
یاراگادی	۸۶	۶۰	۲۶	۳۹,۱۹

پس از انتخاب تعداد LLM برای هر سازند، داده‌های انتخاب شده از چاه‌های مختلف که متعلق به سازندهای فوق هستند برای تخمین وارد الگوریتم شدند، که در نتیجه آن مقدار TOC تخمینی برای هر چاه بدست آمد.

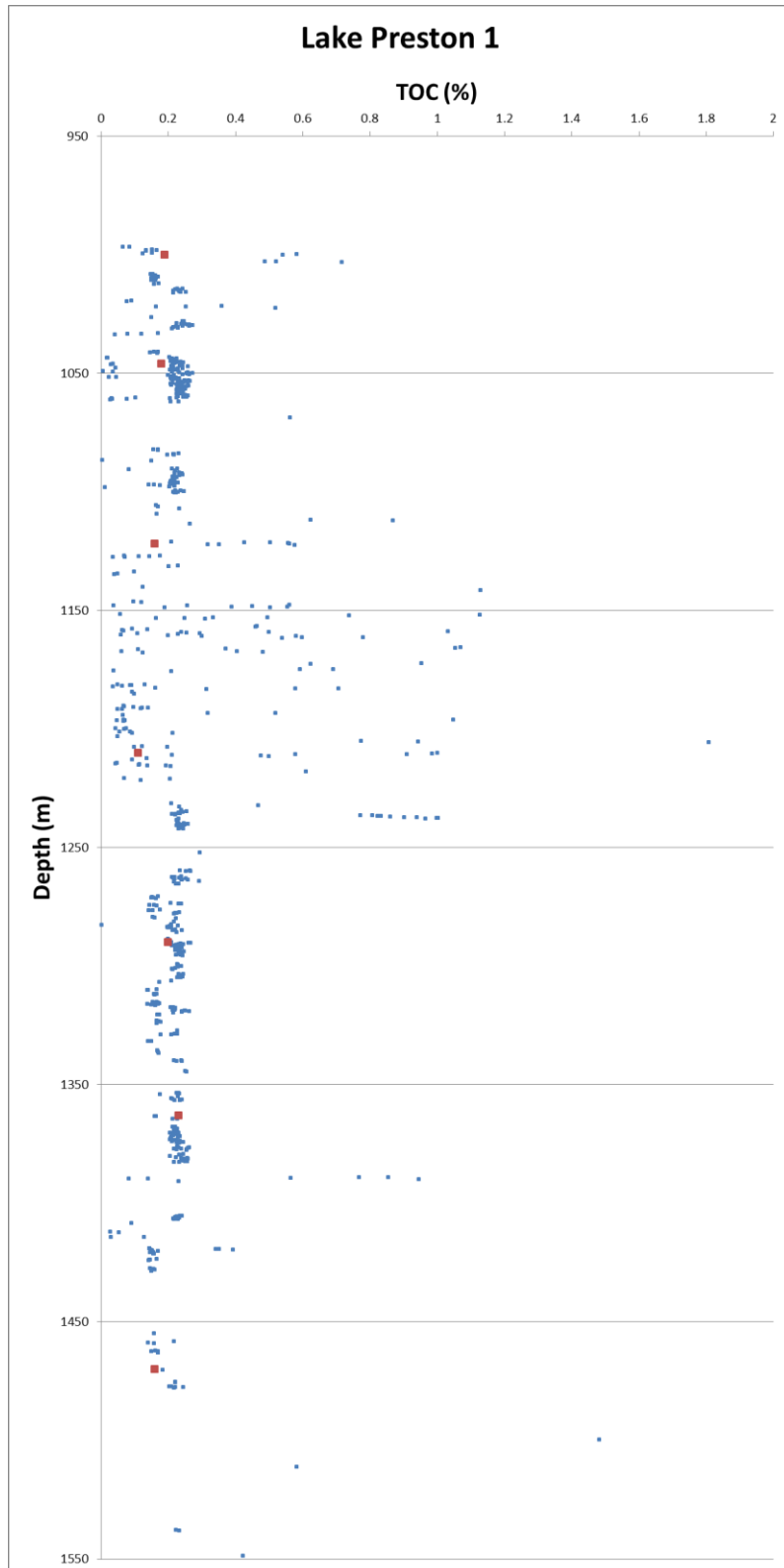
شکل ۴-۳۷ تا شکل ۴-۴۵ نتایج تخمین TOC را برای تمام چاه‌ها نمایش می‌دهند.



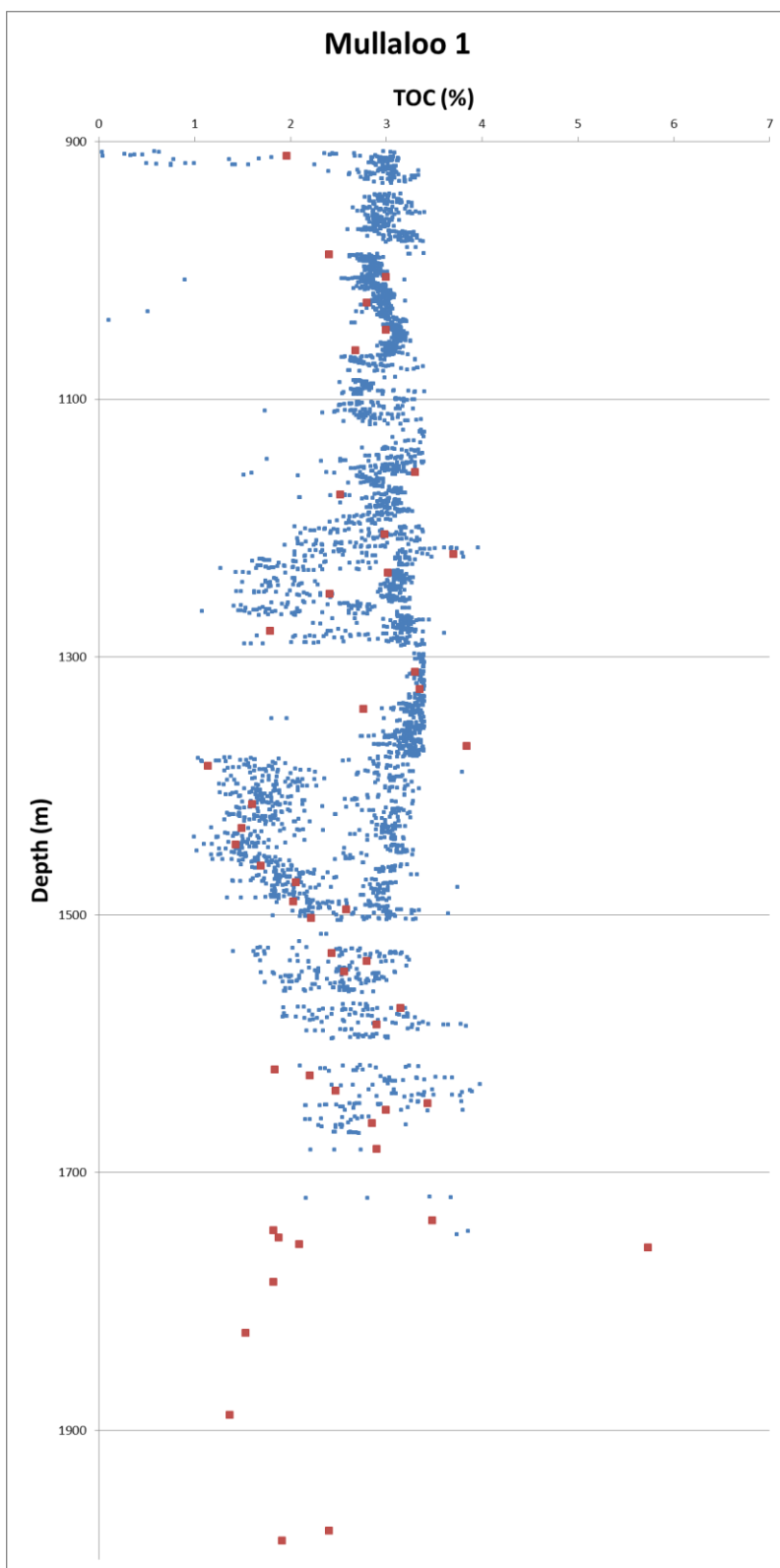
شکل ۴-۳۷ TOC تخمینی برای چاه گیجرود ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی



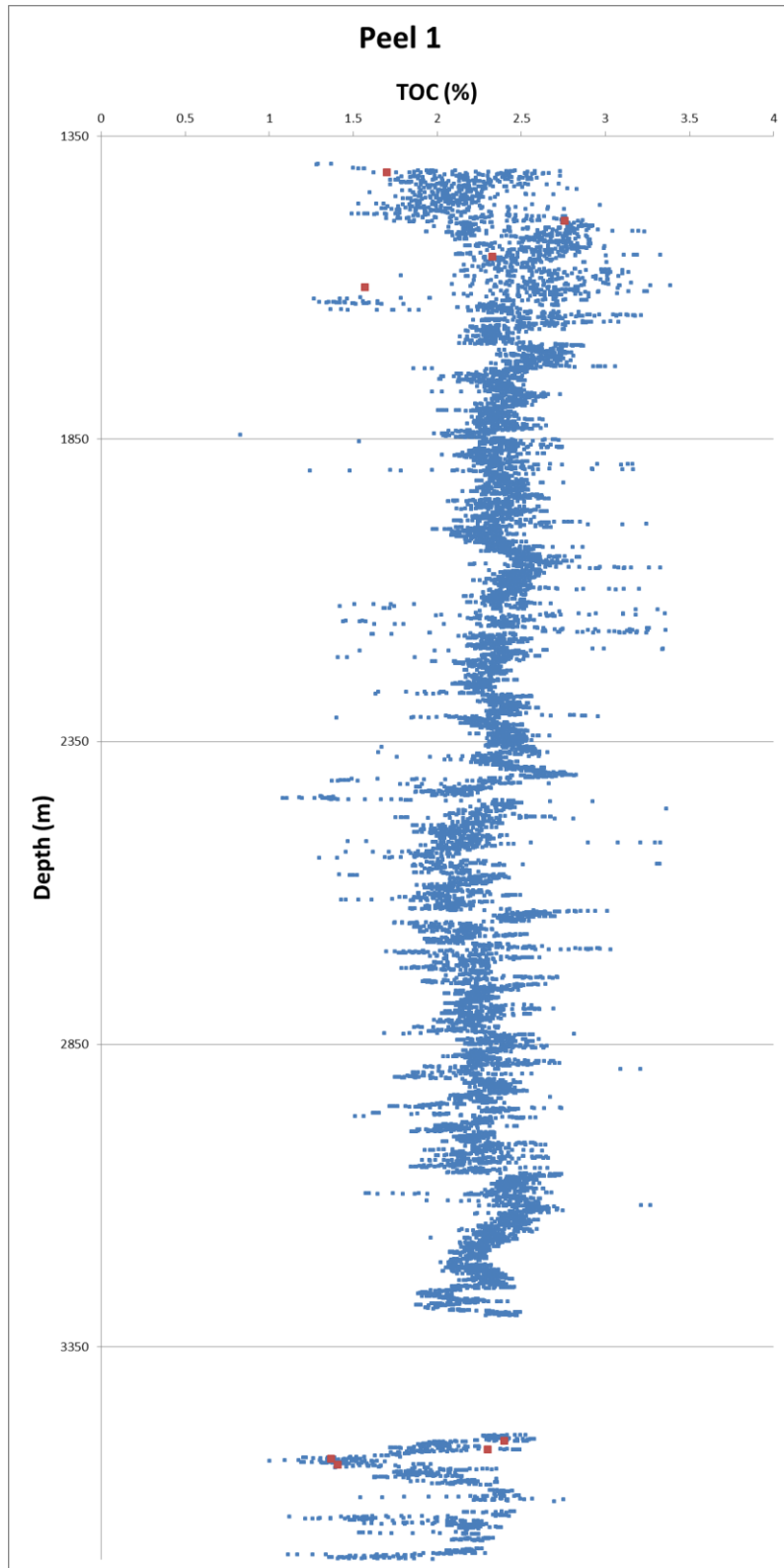
شکل ۴-۳۸ TOC تخمینی برای چاه گیجرود ۲ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی



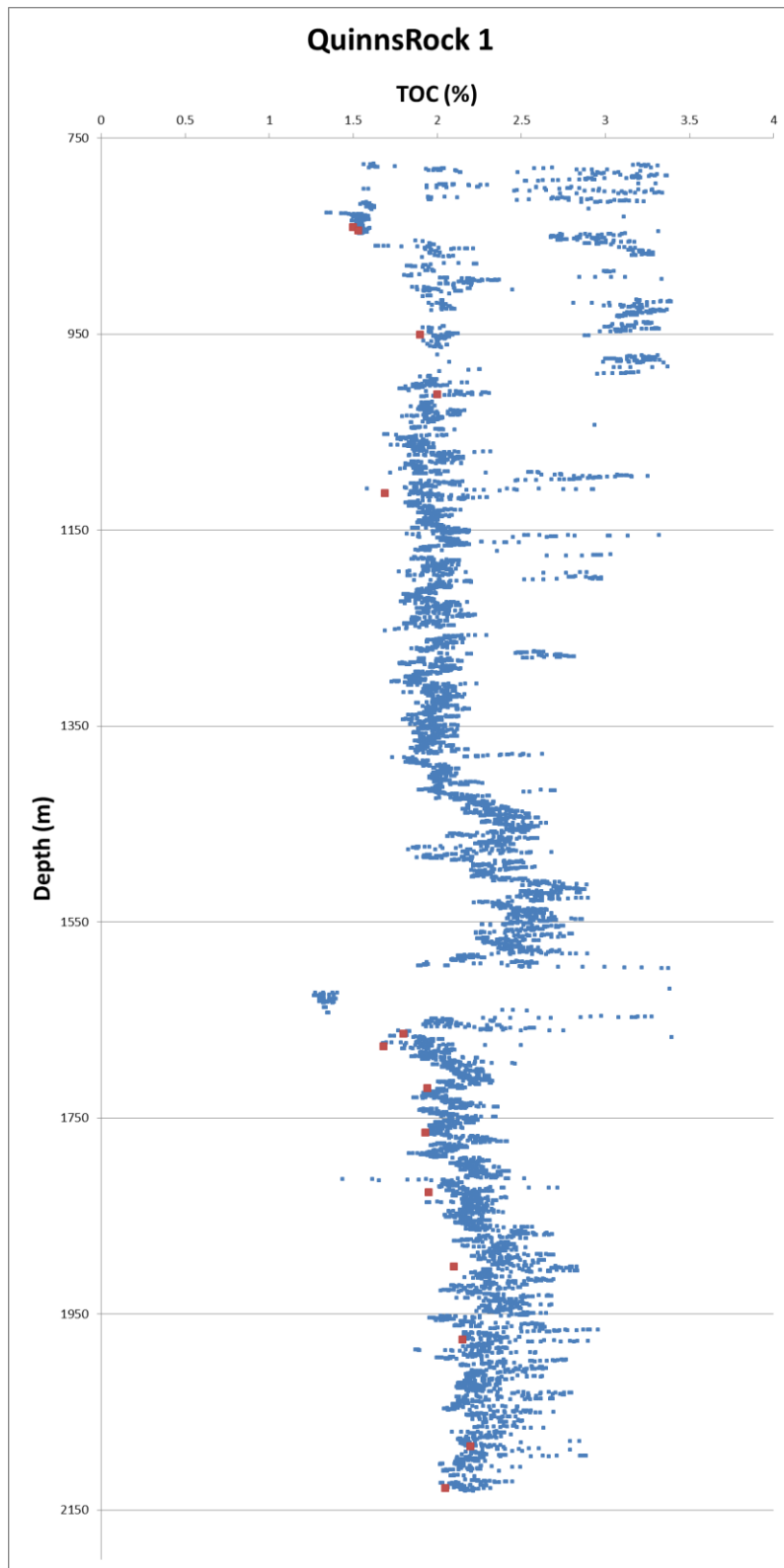
شکل ۴-۳۹ TOC تخمینی برای چاه لیک پریستون ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی



شکل ۴-۴ TOC تخمینی برای چاه مولالو ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی

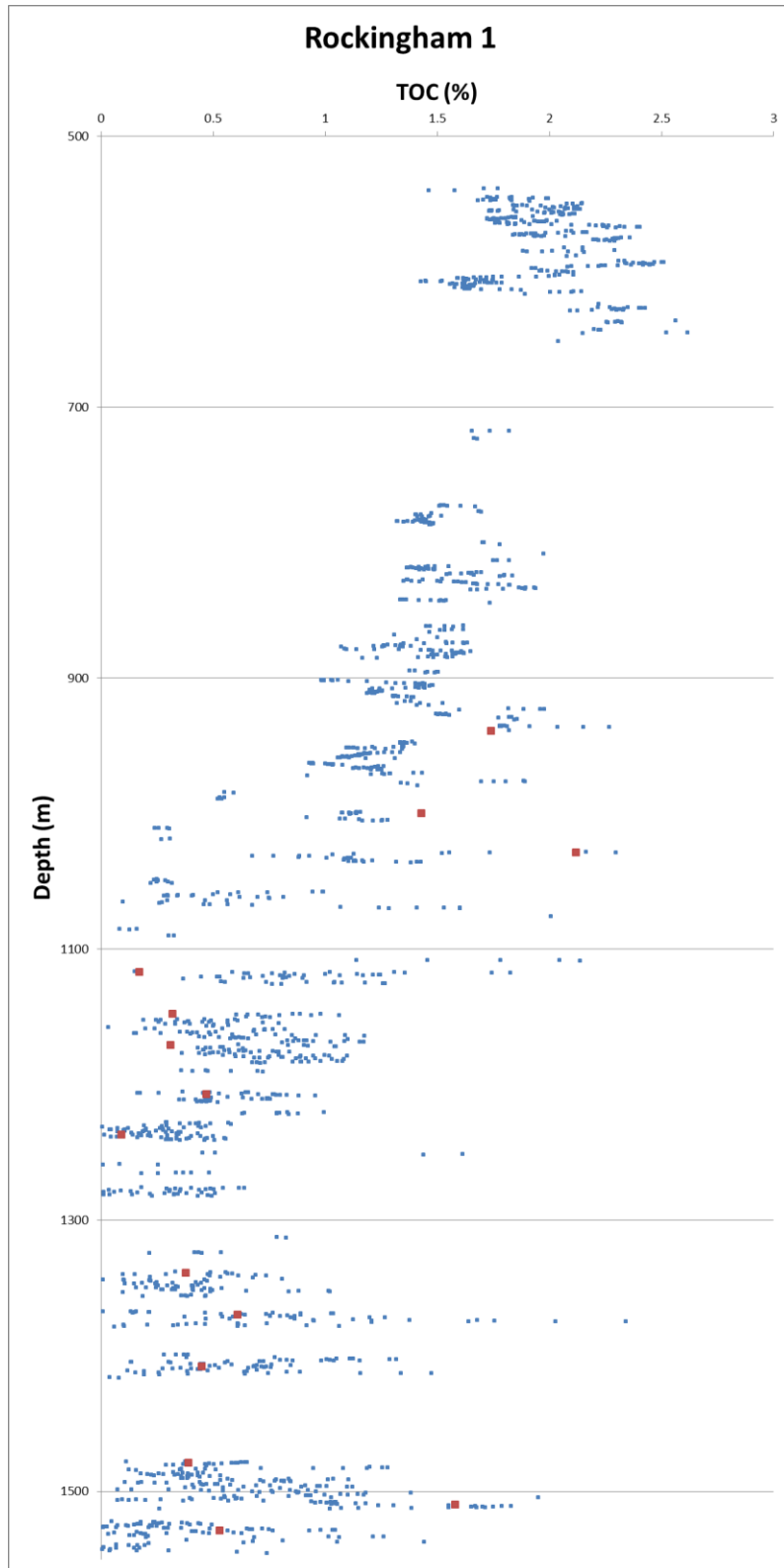


شکل ۴-۴۱ TOC تخمینی برای چاه پل ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی

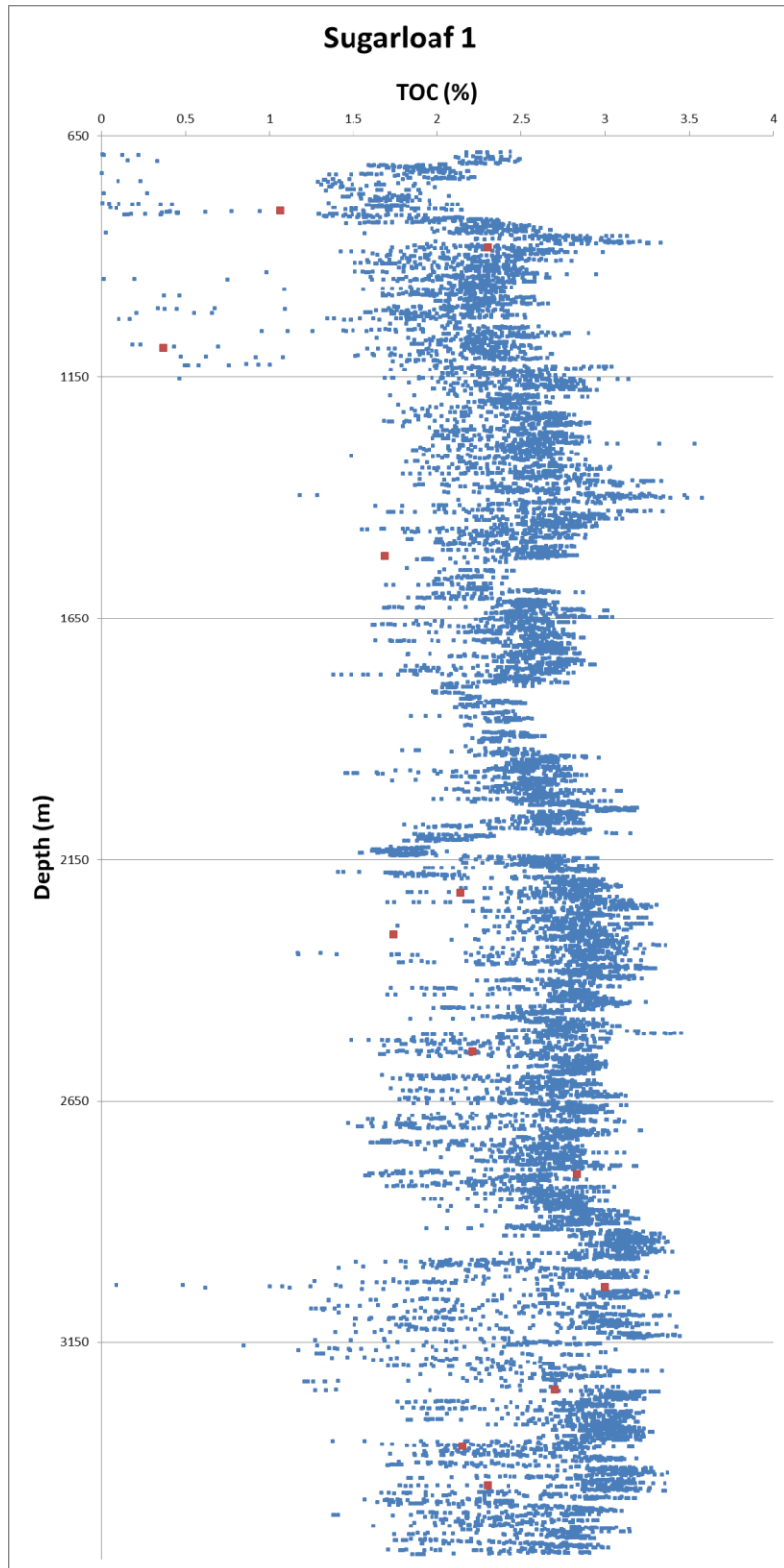


شکل ۴-۴۲ TOC تخمینی برای چاه کویینزراک ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی

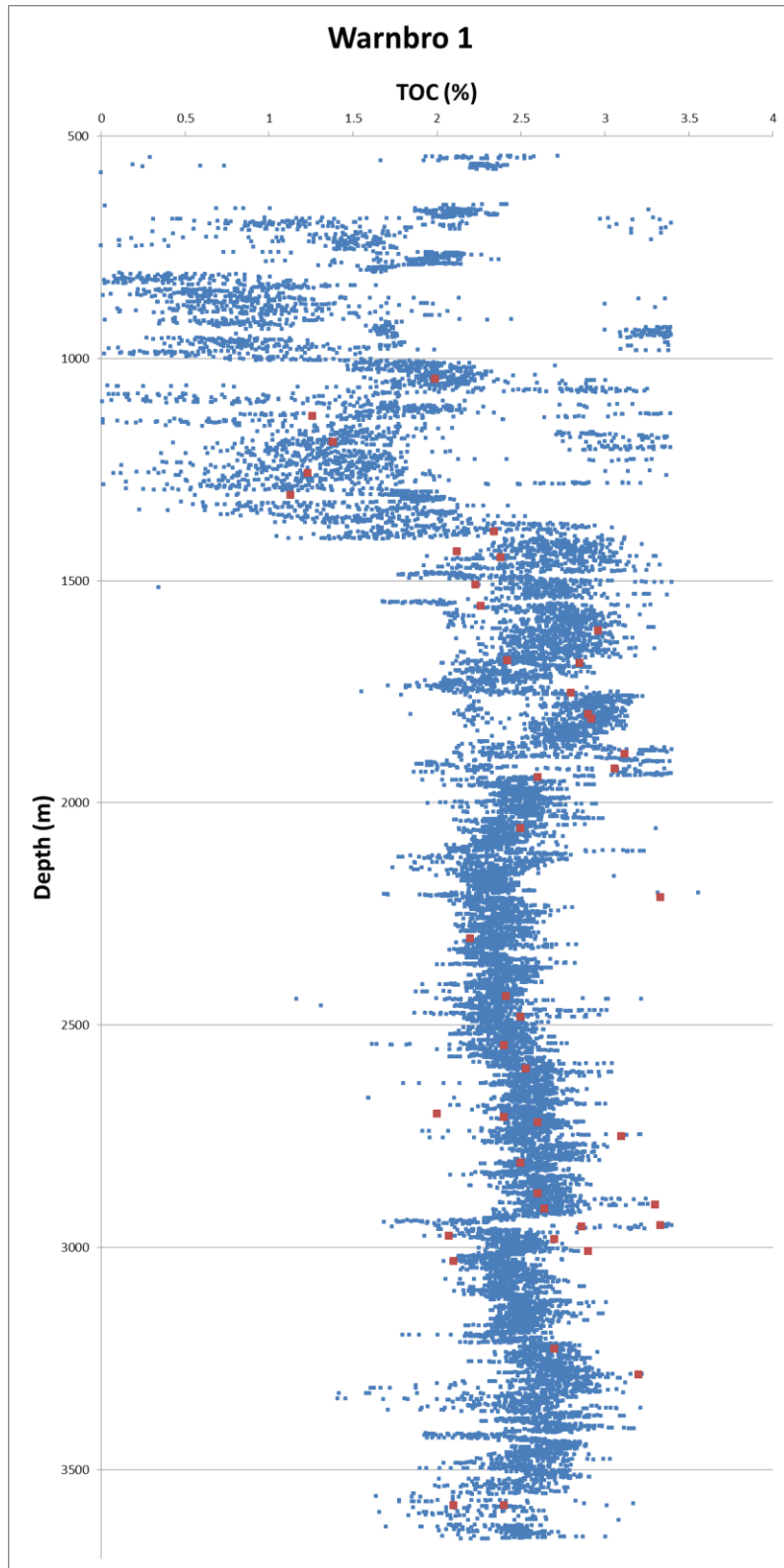




شکل ۴-۴۳ TOC تخمینی برای چاه راکینگهام ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی



شکل ۴-۴۴ TOC تخمینی برای چاه شوگرلوف ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی



شکل ۴-۴۵ TOC تخمینی برای چاه وارنبرو ۱ با استفاده از الگوریتم لولیموت-نقاط قرمز رنگ مقادیر موجود از اندازه گیری های آزمایشگاهی و نقاط آبی مقادیر تخمینی

# فصل پنجم

## نتیجہ گیری و پیشہادات

## ۵ نتیجه گیری و پیشنهادات

### ۵-۱ نتیجه گیری

- با توجه به نتایج حاصل از روش خوشه‌بندی DBSCAN بر روی داده‌های مصنوعی و واقعی، عدم وابستگی این الگوریتم به شکل داده‌ها تایید می‌شود که یکی از معضلات در خوشه‌بندی داده‌های مربوط به علوم زمین و دیگر علوم محسوب می‌شود را بر طرف می‌کند.
- روش ابداعی برای کلاسه‌بندی داده‌ها بر پایه الگوریتم خوشه‌بندی DBSCAN کارایی بسیار خوبی را در کلاسه‌بندی داده‌های مصنوعی از خود نشان داد که باعث استفاده از این روش برای داده‌های واقعی شد.
- بهترین نتیجه مدل‌سازی الگوریتم لولیموت بر روی داده‌هایی است که در قسمت‌های مختلف به شکل فرا سطح باشند و با توجه به این نکته که داده‌های مربوط به چاه‌های نفتی در اغلب موارد شکل توده‌ای دارند به هنگام استفاده از این الگوریتم باید پیش پردازش‌های کافی بر روی داده‌ها انجام شود.
- با توجه به نتایج کسب شده از مدل‌سازی به وسیله الگوریتم لولیموت که خلاصه آن‌ها در جدول ۴-۱۵ آمده است دو نتیجه حاصل می‌شود: اول) بیشترین شباهت به فرا سطح در داده‌ها به ترتیب متعلق به سازندهای کاکلیشیل گولی و پارملیا است، دوم) دلیل خطای پایین مدل‌سازی در سازند لیدرویل تعداد کم داده است.
- یافتن روش یا الگوریتمی که باعث خطی شدن داده‌ها در قسمتی از فضا شود بهبود نتیجه مدل‌سازی را به همراه خواهد داشت.

## ۲-۵ پیشنهادات

- با توجه به سرعت بالای مدل‌سازی با الگوریتم لولیموت پیشنهاد می‌شود تعداد داده و یا تعداد چاه نمودارها افزایش یابد و روندی ترتیب داده شود که از بین چاه نمودارها مواردی انتخاب شود که باعث ایجاد اشکال فرا سطحی شود.
- با توجه به الگوریتم‌های بهبود یافته نسبت به لولیموت مانند پولیموت برای مدل‌سازی داده‌ها از این الگوریتم‌ها استفاده شود.
- روش انجام شده در این پایان‌نامه برای علوم زمین کار جدید و نوبی محسوب می‌شود، پیشنهاد می‌شود از الگوریتم لولیموت در مواد دیگر نیز استفاده شود و با مقایسه نتایج راه‌های بهبود این روش بررسی شود.
- روش خوشه‌بندی انجام شده در این پایان‌نامه نسبت به روش‌های معمول در علوم زمین از کارایی بالاتری برخوردار است. پیشنهاد می‌شود از این روش برای خوشه‌بندی داده‌ها استفاده شود و روش‌هایی برای بهبود آن نیز به وجود آید.
- روش‌های خوشه‌بندی دیگری مانند CLARNAS، BIRCH، SUBCLU و OPTICS نیز برای پردازش داده‌ها وجود دارد که پیشنهاد می‌شود در علوم زمین وارد شده و کارایی آن‌ها نیز بررسی گردد.

## ۶ مراجع

حسنی‌پاک، ع. ا. و م. شرف‌الدین (۱۳۸۴). تحلیل داده‌های اکتشافی. موسسه انتشارات و چاپ دانشگاه تهران.

رضایی، م. (۱۳۸۰). زمین‌شناسی نفت. امین، علوی.

Alpaydin, E. (2004). Introduction to machine learning, MIT press.

Balasko, B., J. Abonyi and B. Feil (2007). Fuzzy Clustering and Data Analysis Toolbox, Department of Process Engineering University of Veszprem.

Bird, K. J. and C. P. Moyes (1971). Well Completion Report Sugarloaf NO.1. West Australian Petroleum PTY. Limited.

Bishop, C. M. (1995). "Neural networks for pattern recognition."

Bozanic, D. (1969). Gage Road NO.1 Well Completion Report. West Australian Petroleum PTY. Limited.

Bozanic, D. (1969). Quinns Rock NO.1 Well Completion Report. West Australian Petroleum PTY. Limited.

Brooks, D. (1984). Well Completion Report Mullaloo-1. ESSO AUSTRALIA LIMITED. 2.

Dellenbach, J., J. Espitalie and F. Lebreton (1983). Source rock logging. SPWLA 8th European Formation Evaluation Symposium.

Duda, R. O., P. E. Hart and D. G. Stork (1995). "Pattern Classification and Scene Analysis 2nd ed."

Ester, M., H.-P. Kriegel, J. Sander and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining, AAAI Press.

Fertl, W. and G. Chilingar (1988). "Total organic carbon content determined from well logs." SPE formation evaluation 3(2): 407-419.

Gustafson, D. E. and W. C. Kessel (1978). Fuzzy clustering with a fuzzy covariance matrix. Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference on, IEEE.

Hawkins, L. V. (1965). Geophysical Investigation in the Area of the Perth Basin. Western Australia. xxx: 1026-1052.

<http://www.rigzone.com/data/dayrates>.

Huang, Z. and M. A. Williamson (1994). "Geological pattern recognition and modelling with a general regression neural network."

Huang, Z. and M. A. Williamson (1996). "Artificial neural network modelling as an aid to source rock characterization." Marine and petroleum geology 13(2): 277-290.

Kadkhodaie-Ilkhchi, A., H. Rahimpour-Bonab and M. Rezaee (2009). "A committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: An example from Kangan and Dalan reservoirs in South Pars Gas Field, Iran." Computers & Geosciences 35(3): 459-474.

Kadkhodaie-Ilkhchi, A., M. R. Rezaee and H. Rahimpour-Bonab (2009). "A committee neural network for prediction of normalized oil content from well log data: An example from South Pars Gas Field, Persian Gulf." Journal of Petroleum Science and Engineering 65(1): 23-32.

Kamali, M. R. and A. Allah Mirshady (2004). "Total organic carbon content determined from well logs using  $\Delta \log R$  and neuro fuzzy techniques." Journal of Petroleum Science and Engineering **45**(3): 141-148.

Kaufman, L. and P. J. Rousseeuw (1990). Finding Groups in Data: an Introduction to Cluster Analysis, John Wiley.

Khoshnoodkia, M., H. Mohseni, O. Rahmani and A. Mohammadi (2011). "TOC determination of Gadvan Formation in South Pars Gas field, using artificial intelligent systems and geochemical data." Journal of Petroleum Science and Engineering **78**(1): 119-130.

Marshall, J. L. C., L. P. Mitchell and G. Jackman (1983). Rockingham NO.1 Completion Report. Phoenix Oil & Gas N.L.

Matheus, C. J., P. K. Chan and G. Piatetsky-Shapiro (1993). "Systems for Knowledge Discovery in Databases." IEEE Transactions on Knowledge and Data Engineering **5**: 903-913.

Meyer, B. L. and M. H. Nederlof (1984). "Identification of source rocks on wireline logs by density/resistivity and sonic transit time/resistivity crossplots." AAPG Bulletin **68**(2): 121-129.

Moyes, C. P. (1971). Gage Road NO.2 Well Completion Report. West Australian Petroleum PTY. Limited.

Moyes, C. P. (1971). Well Completion Report Warnbro NO.1. West Australian Petroleum PTY. Limited.

Nelles, O. (1999). Nonlinear System Identification with Local Linear Neuro-Fuzzy Model. Ph.D, TU Darmstadt.

Nelles, O. (2000). Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models, Springer.

O.Duda, R., P. E.Hart and D. G.Stork (2003). Pattern Classification, John Wiley.

Passey, Q. R., S. Creaney, J. B. Kulla, F. J. Moretti and J. D. Stroud (1990). "A practical model for organic richness from porosity and resistivity logs." AAPG Bulletin **74**(12): 1777-1794.

Phillips-Australian-Oil-Company (1978). Well Completion Report Peel NO. 1. Phillips Australian Oil Company.

Schmoker, J. W. (1979). "Determination of Organic Content of Appalachian Devonian Shales from Formation-Density Logs: GEOLOGIC NOTES." AAPG Bulletin **63**(9): 1504-1509.

Schmoker, J. W. (1981). "Determination of organic-matter content of Appalachian Devonian shales from gamma-ray logs." AAPG Bulletin **65**(7): 1285-1298.

Schmoker, J. W. and T. C. Hester (1983). "Organic carbon in Bakken formation, United States portion of Williston basin." AAPG Bulletin **67**(12): 2165-2174.

Smith, D. N. (1967). Cockburn NO.1 Well Completion Report. P.S.S.A Report.

Tissot, B. P. and D. H. Welte (1978). "Petroleum formation and occurrence: a new approach to oil and gas exploration."

Young, R. J. B. and J. N. Johanson (1973). Lake Princeton NO. 1 Well Completion Report Perth Basin E.P. 25. West Australian Petroleum PTY. Limited.