

## مقایسه شبکه‌های عصبی با روش‌های داده‌کاوی به منظور شبیه‌سازی عنصر مس؛ مطالعه موردی: پرکام کرمان

سید سعید قنادپور<sup>۱</sup>، اردشیر هزارخانی<sup>۲\*</sup>، ترانه رودپیما<sup>۳</sup>

۱- دانشجوی دکتری اکتشاف، دانشکده مهندسی معدن و متالورژی، دانشگاه صنعتی امیرکبیر

۲- استاد دانشکده مهندسی معدن و متالورژی، دانشگاه صنعتی امیرکبیر

۳- دانشجوی کارشناسی ارشد اکتشاف، دانشکده مهندسی معدن و متالورژی، دانشگاه صنعتی امیرکبیر

پذیرش مقاله: ۱۳۹۴/۱۱/۱۸

تأیید نهایی مقاله: ۱۳۹۵/۶/۱۰

### چکیده

تجزیه و تحلیل داده‌ها به ما کمک می‌کند تا بدانیم چگونه می‌بایست به نتایج مورد انتظار دست‌یابیم، بنابراین برای دستیابی به پردازش‌هایی دقیق‌تر، لازم است تا از بین تمام روش‌های تحلیل اطلاعات، هر یک که برای موضوع تحت بررسی مان مناسب‌تر است را انتخاب نماییم. بدین منظور جهت آنالیز نمونه‌های حاصله از عملیات نمونه‌برداری سطحی سیستم مس پورفیری پرکام واقع در استان کرمان، تحت چهار مقدار طول و عرض نقاط نمونه‌برداری، عیار عناصر مس و مولیبدن، از سه روش پرکاربرد K-نزدیکترین همسایه (KNN)، K میانگین (K-Means) و شبکه‌های عصبی بهره خواهیم گرفت. یکی از دیدگاه‌های مهم در علم داده‌کاوی برای تحلیل و بررسی روی حجم زیادی از داده‌ها و نمونه‌ها با مشخصه‌های گوناگون، دیدگاه خوشه‌بندی می‌باشد. از معروف‌ترین الگوریتم‌های خوشه‌بندی، الگوریتم KNN و K-Means می‌باشد که الگوریتم KNN بر اساس تخمین پیش‌می‌رود و روشی غیر پارامتری جهت کلاسه‌بندی و رگرسیون-گیری و به دست آوردن روابط چندین متغیر می‌باشد در حالی که K-Means بر اساس یک معیار فاصله، داده‌ها را به K خوشه تقسیم می‌کند و پس از کلاسه‌بندی داده‌ها، رفتار آنها نسبت به یکدیگر را مورد تحلیل قرار می‌دهد. شبکه‌های عصبی در تشخیص الگوها و نیز زمانی که اطلاعات در دسترس برای تفسیر کافی نیستند، می‌توانند ابزاری سودمند باشند. به منظور شبیه‌سازی و تخمین عیار مس، الگوریتم‌های یاد شده با یکدیگر مورد مقایسه واقع شده و در نهایت نتایج ارائه شده‌اند. در مقاله پیش‌رو، هدف مقایسه نتایج این سه روش به منظور تعمیم آن برای سایر پژوهش‌ها در مواجهه با تعداد داده‌های محدود و هموار ساختن مسیر برای محققین می‌باشد. نتایج حاصله نشان می‌دهد که روش KNN با ضریب همبستگی مناسب‌تر نسبت به شبکه‌های عصبی و K-Means برای تخمین عیار عنصر مس، مؤثر واقع شده است. امتیاز استفاده از روش KNN نسبت به دیگر روش‌های تخمینی در مقاله پیش‌رو، ارائه‌گر الگویی مشخص و دقیق به منظور تخمین عیار در مواجهه با تعداد داده‌های محدود به تصمیم‌گیران این صنعت می‌باشد.

واژه‌های کلیدی: پرکام، داده‌کاوی، شبکه‌های عصبی.

## مقدمه

که به‌طور کلی در حالت شناسایی، الگوریتم KNN روشی برای کلاسه‌بندی هدف مورد نظر، بر اساس نزدیکترین نمونه‌های تعلیم در فضای مشخص می‌باشد. کلاسه‌بندی توسط KNN زمانی توسعه یافت که نیاز به تجزیه و تحلیل‌های دقیق و مشخص احساس گردید. در واقع در شرایطی که تخمین‌های پارامترهای قابل اعتماد تراکم‌های احتمالی، ناشناخته‌اند و حتی برای مشخص کردن دشوارند، نیاز به این روش به مراتب بیشتر احساس می‌شود. در یک گزارش پزشکی منتشر نشده از گروه نیروی هوایی ایالات متحده حمل‌ونقل هوایی در سال ۱۹۵۱، یک متد غیر پارامتری برای الگوی کلاسه‌بندی که امروزه با قاعده K-نزدیکترین همسایه می‌شناسیم، معرفی شده است. آنها روشی غیر پارامتری و بدیعی جهت طبقه‌بندی مبتنی بر فاصله بین نقاط یا توزیعات را معرفی کردند. KNN مزایای عمده بسیاری دارد از جمله: سادگی، کارایی، شهودی و دارای عملکرد طبقه‌بندی رقابتی در بسیاری از حوزه‌ها. بسیار مقاوم در برابر داده‌های تعلیم پر هرج و مرج است و در زمانی که داده‌های آموزشی بزرگ هستند، مؤثر واقع می‌شود. از جمله فعالیت‌های علمی حول روند KNN می‌توان به مطالعه جهت حل چالش عمده طبقه‌بندی متن چینی با فضایی با ابعاد زیاد توسط الگوریتم بدیع KNN جهت کلاسه‌بندی متن (زو، لو و جو، ۲۰۱۳)، طراحی یک الگوریتم فازی قابل انطباق و کلاسه‌بندی متن جهت حل چالش طبقه‌بندی متن با چندین کلاس (شنگ، هوانگ، جو، لین، کو و همکاران، ۲۰۰۶)، پیشنهاد یک تابع جدید پیوستگی برای اندازه‌گیری فاصله و رسیدن به نزدیکترین فاصله توسط الگوریتم KNN (باتاچاریا، گوش و چادری، ۲۰۱۲)، مقایسه عملکرد متدهای کاهش ابعاد که می‌تواند کلاسه‌بندی الگوریتم

مفهوم یک الگوریتم یا برنامه کامپیوتری برای ما آزمون‌های بسیار کارآمدی را برای رسیدن به عمق دانش در ارتباط با هر موضوع داده شده، فراهم می‌کند؛ به این ترتیب مجموعه‌های داده‌ها پس از کشف، مورد تجزیه و تحلیل قرار می‌گیرند. پس شناخت مناسب‌ترین روش برای آنالیز داده‌های در اختیارمان، چالشی بنیادین برای محققین می‌باشد. چرا که داده‌ها می‌توانند نشان دهند که آیا با هر تغییر قابل توجهی در متغیر وابسته می‌تواند امید به اثرگذاری باشد یا خیر. با استفاده از تجزیه داده‌ها، به روابط بین چندین فاکتور متعدد تأثیرگذار بر نتایج ارزیابی‌ها، پی می‌بریم. با پردازش‌های کمی و کیفی، به تأثیرگذار بودن یا نبودن یک عملکرد و حتی علت آن می‌توان پی برد و همچنین با فراهم آوردن مدارک معتبر نشان می‌دهند که آیا برنامه ما موفقیت‌آمیز بوده و به نوعی آگاهی دهنده در رابطه با محدودیت‌ها می‌تواند باشند. اولین و البته ساده‌ترین مرحله در پردازش اطلاعات، توصیف توسط خلاصه‌سازی داده‌ها مانند میانه و انحراف استاندارد است. داده‌کاوی مزایای بسیاری به‌خصوص در هوش مصنوعی و در آمار دارد. در هر دو، این علم بر روی چالش شناسایی الگوها و کلاسه‌بندی تمرکز دارد. قدرت روز افزون کامپیوترها و هزینه کم، با نیاز به تجزیه و تحلیل مجموعه اطلاعات زیاد، زمینه را جهت توسعه تکنیک‌های نو و بدیع بر اساس اکتشاف راه‌حل‌های احتمالی فراهم می‌کند (بکس، ۲۰۰۰؛ اودیبرت و سایاکوو، ۲۰۰۷). روش‌های مورد بررسی در مقاله حاضر، KNN و K-Means و روش شبکه‌های عصبی می‌باشد که برتری یکی بر دیگری را در مواقعی که با تعداد داده محدود روبرو هستیم، نشان خواهیم داد. از روش‌های معروف خوشه‌بندی در داده‌کاوی، روش K-نزدیکترین همسایه می‌باشد،

اخیر با این الگوریتم می‌باشد. در بررسی ادبیات موضوع، مطالعات ارزشمند دیگری در مباحث زمین‌شناسی و علوم معدنی یافت می‌شود که با کمک روش‌های کلاسترینگ به خصوص روش K-Means به کلاس‌بندی پارامترهای مختلف در یک مطالعه پرداخته شده تا بتوان به کمک آن تحلیلی مناسبی از رفتار داده‌ها و پارامترهای موجود نسبت به یکدیگر داشت. در این مورد می‌توان به تقسیم‌بندی عوارض زمین‌شناسی (یانگ، ژنگ و وو، ۲۰۱۲)، تقسیم‌بندی تأثیر پوشش گیاهی و بازیابی سلامتی آب در جنگل‌های ساحل مدیترانه (مورا، آرماس هرا، گوئرا، رودریگز و آربلو، ۲۰۱۲)، ارائه برنامه شناسایی الگوهای ژئوشیمیایی در مناطق معدنی به کمک روش K-Means (مشکانی و همکاران، ۲۰۱۱) و تعیین و تحلیل زمین‌شناسی رفتار عناصر مختلف در سیستم‌های مس پرفیری (قنادپور و هزارخانی، ۱۳۹۱؛ ۲۰۱۵) با استفاده از روش K-Means اشاره نمود. شبکه‌های عصبی مصنوعی نیز یکی از پرکاربردترین ساختارهای مورد استفاده در هوش مصنوعی می‌باشند که حتی می‌توان آنها را به‌عنوان پایه و اساس هوش مصنوعی دانست. امروزه این ساختارها در علوم مهندسی و پایه در راستای خودکارسازی، طبقه‌بندی و تخمین توابع پیچیده استفاده می‌شوند. در این زمینه می‌توان به کارهای انجام شده مانند بررسی‌های ژئوشیمیایی به همراه عیار و تناژ آنها با استفاده از شبکه‌های عصبی احتمالی (سینگر، ۲۰۰۶)، طبقه‌بندی مواد ارگانیکی رسوبی (ولر، کرکران، هریس و وار، ۲۰۰۵)، اندازه‌گیری‌های شوری به دست آمده حاصله از تصاویر ذرات ارگانیکی رسوبی با استفاده از شبکه‌های عصبی طبقه‌بندی (اندرو، ولر، هریس و اندرو، ۲۰۰۷) و عددی کردن داده‌های ژئوشیمیایی سنگ آتشفشانی و طبقه‌بندی آنها با استفاده از شبکه‌های عصبی مصنوعی (لاکاسی،

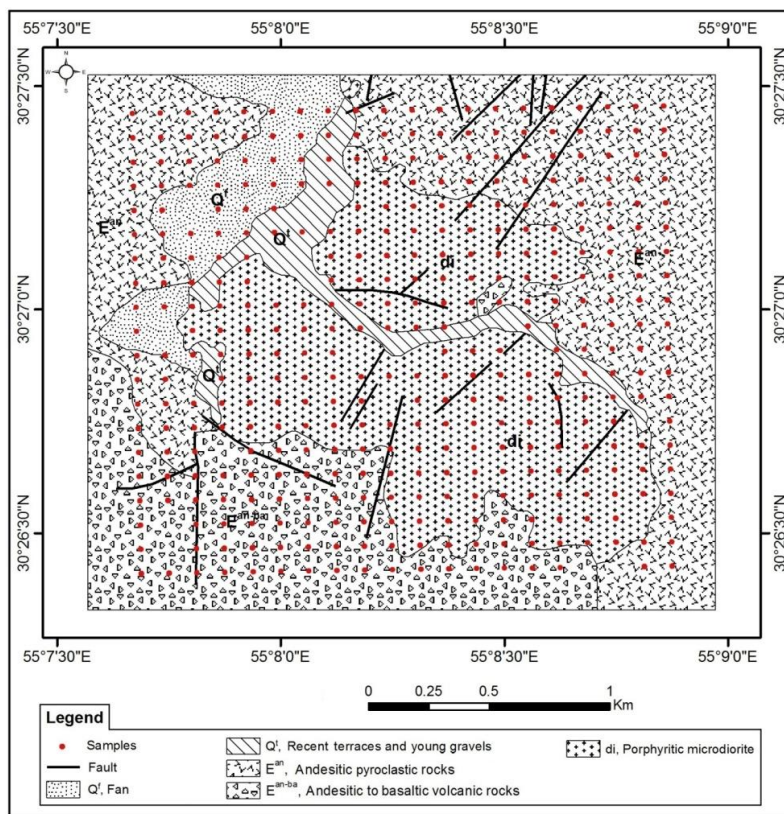
KNN را برای مجموعه اطلاعات با ابعاد بالا مانند داده‌های ریز آرایه، بهبود ببخشد (دیگالا و باستروم، ۲۰۰۷)، مطالعه رگرسیون KNN به‌عنوان فرایند نسبت دادن جهت پیش‌بینی مجموعه اطلاعات فضایی و زمانی باد (پولوچوک، تریبر و کرامر، ۲۰۱۴) و به‌عنوان تدوین سیاست‌های بهره‌برداری از مخزن در زمان وقوع سیلاب (زهراپی و تکشی، ۱۳۸۷)، اشاره کرد. دیگر روش مورد بررسی، روش K-Means می‌باشد که به‌طور وسیعی توسط پژوهشگران مختلف مورد بررسی قرار گرفته است و در تلاش است که به خوشه‌بندی تعدادی نمونه با تعداد دسته مشخص بیردازد (بی و ژانگ، ۲۰۱۳)، به نحوی که مجموع فواصل اقلیدسی هر یک از نمونه‌ها از مرکز دسته‌ای که به آن تخصیص یافته است، کمینه گردد. در مطالعات مختلفی دیده می‌شود که رفتار عناصر به روش‌های گوناگونی نسبت به یکدیگر سنجیده می‌شوند که می‌توان به‌عنوان مثال از مطالعه (تارکیان و استرایرنی، ۱۹۹۹)، نام برد. با افزایش پیچیدگی نرم‌افزارهای وب، تشخیص نواقص در این صنعت امری بحرانی تلقی گشت. استفاده از متد خوشه‌بندی در پیش-بینی نواقص می‌تواند برآورد آنها را در کد منبع صفحه وب تسهیل ببخشد؛ به این منظور در مطالعه‌ای (اوزترک، کاوسگلو و زنگین، ۲۰۱۵) از روش پرکاربرد و مؤثر K-Means بهره گرفته شده است. مطالعه K-Means بر روی داده‌های نفوذ شبکه (واراپراساد، ۲۰۱۲)، پیشنهاد یک الگوی اصلاح شده از K-Means مبتنی بر فاصله متریک (زو-فنگ، سیاوفان، کیائو و ژیکوانگ، ۲۰۱۴)، ارائه الگوریتم ترکیبی برای خوشه‌بندی مدرک وب (کمل، اوچن و بالی، ۲۰۱۴) و مطالعاتی در جهت ادغام الگوریتم K-Means و الگوریتم تکامل دیفرانسیل جهت انتخاب تعداد خوشه‌ها (کیو، سوریانی و یاسید، ۲۰۱۳)، از جمله فعالیت‌های

- آتشفشانی ائوسن می‌باشد که میزبان توده‌های ساب ولکانیک و کانی‌سازی پورفیری است. در این ناحیه مجموعه سنگ‌های آتشفشانی، توف و آذر آواری دارای ترکیب آندزیت بازالیت و آندزیت هستند که گاه لایه‌بندی دارند. توده‌های نفوذی ساب ولکانیک منطقه شامل دیوریت و میکروکوآرتزدیوریت پورفیری می‌باشد (شکل ۱) که با سامانه دگرسانی و کانی‌سازی ارتباط نشان می‌دهند (قنادپور و هزارخانی، ۱۳۹۲) این واحدهای سنگی توسط دایک‌هایی از نوع دیوریت قطع شده‌اند. دگرسانی در محدوده پرکام نسبتاً شدید است و شامل پتاسیک، فیلیک، آرژیلیک و پروپلیتیک می‌باشد.

روزر، رویز-دل-سولار، زوزر و هرو، ۲۰۰۶)، تقریب برداشت‌های پیوسته و نقشه‌های آن (فوناهاشی، ۱۹۸۹)، تخمین ذخیره، مقایسه نتایج حاصله از کریجینگ و شبکه عصبی (وو و جو، ۱۹۹۳) کارهای پیش‌بینی در معادن اشاره نمود.

#### محدوده مورد مطالعه

ذخیره پورفیری پرکام واقع در ۲ کیلومتری معدن مس پورفیری میدوک در نقشه‌های زمین‌شناسی ۱:۲۵۰۰۰۰ انار و ۱:۱۰۰۰۰۰ شهر بابک قرار دارد. این محدوده بخشی از زون ارومیه - دختر (سهند - بزمان) است. واحدهای سنگی در برگیرنده در محدوده مورد مطالعه عمدتاً شامل مجموعه رسوبی



شکل ۱: نقشه زمین‌شناسی محدوده پرکام به همراه واحدهای لیتولوژی نقاط نمونه‌برداری شده.

(۱۳۹۱) مورد بررسی قرار گرفته است. دگرسانی در محدوده پرکام گسترده و شامل پروپلیتیک، فیلیک،

واحدهای زمین‌شناسی موجود در این محدوده به‌طور مفصل در مطالعه قنادپور و هزارخانی

زمین‌شناسی در شکل ۱ به نمایش درآورده شده است.

شبکه‌های عصبی: شاید یکی از بارزترین ویژگی‌های مغز توانایی فراگیری آن باشد. مغز می‌تواند به خود آموزش دهد. یادگیری از طریق مثال همان شیوه‌ای است که توسط آن اطفال زبان را فرا می‌گیرند. نوشتن، خوردن و آشامیدن را می‌آموزند. مجموعه معیارها و نکات اخلاقی را کسب می‌کنند. این عمل در سیستم‌های کامپیوتری متعارف و عمومی مشاهده نمی‌شود. کامپیوترها اغلب از برنامه‌های پیش‌نوشته شده پیروی می‌کنند که قدم به قدم دستورات مشخصی را در کلیه مراحل عملیاتی به آنها می‌دهند (البرزی، ۱۳۸۹). در عوض شبکه‌های عصبی، به‌طور کلی سیستم‌های ریاضی یادگیر غیرخطی هستند. طرز کار این شبکه‌ها از روش یک مدل که می‌توان آن را به وسیله سخت‌افزار یا نرم‌افزار شبیه‌سازی کرد و عملکردی شبیه مغز انسان دارند پیروی می‌کند. یک شبکه عصبی بر خلاف کامپیوترهای رقمی که نیازمند دستورات کاملاً صریح و مشخص است، به مدل‌های ریاضی محض نیاز ندارد بلکه مانند انسان قابلیت یادگیری به وسیله تعدادی مثال مشخص را دارد. هر شبکه عصبی سه مرحله آموزش، اعتبارسنجی و اجرا را پشت سر می‌گذارد. در واقع شبکه‌های عصبی را می‌توان در حل مسائلی که روابط دقیق ریاضی بین ورودی‌ها و خروجی‌های آن برقرار نیست به کار برد. این شبکه‌ها از عناصر ساده‌ای تشکیل شده‌اند که به‌طور موازی عمل می‌کنند و با الهام از شبکه‌های بیولوژیک، هر یک از این عناصر را نرون می‌نامند. آموزش دیدن شبکه‌های عصبی در واقع چیزی جز تنظیم وزن‌های ارتباطی این نرون‌ها به ازای دریافت مثال‌های مختلف نیست تا خروجی شبکه به سمت خروجی مطلوب همگرا شود. شبکه‌های عصبی را می‌توان با اغماض زیاد، مدل‌های الکترونیکی از

آرژیلیک و انواع پتاسیک از نظر وسعت می‌باشد. دگرسانی فیلیک تشکیل دهنده قسمت مرکزی سامانه پرکام است که با دگرسانی پتاسیک در شرق و غرب محدوده مورد مطالعه و دگرسانی آرژیلیک در جنوب محدوده در ارتباط می‌باشد. دگرسانی پروپلیتیک نیز احاطه کننده دگرسانی‌های مذکور می‌باشد. در محدوده مورد مطالعه کانی‌سازی به صورت مالاکیت گسترش قابل ملاحظه‌ای دارد ولی آغشتگی توسط مالاکیت گاه بصورت اغوا کننده می‌باشد. این بدان معنی است که اگرچه گسترش مالاکیت زیاد است ولی کانی‌سازی عیار بالا نبوده و بخش عمده مالاکیت به صورت پرکننده درز و شکاف‌ها می‌باشد. پیریت مهم‌ترین کانی سولفیدی است که در حال حاضر در سطح به صورت اکسیدی یافت می‌شود. چرخش آب‌های سطحی و سوپرژن باعث گسترش مالاکیت در درز و شکاف‌ها شده است. کانی‌سازی عمدتاً در بخش دگرسانی میانی و در شمال شرقی دره پرکام در منطقه دگرسانی فیلیک مشاهده می‌شود ولی گاه در درز و شکاف‌های مناطق مختلف دیده می‌شود. در درز و شکاف‌های منطقه پروپلیتیک که عمدتاً آندزیت می‌باشد نیز مالاکیت گسترش دارد ولی اهمیت اقتصادی چندانی ندارد (قنادپور و هزارخانی، ۱۳۹۲).

### مواد و روش‌ها

نمونه‌برداری در محدوده سیستم پورفیری پرکام در یک شبکه منظم صورت گرفته است. شبکه نمونه‌برداری به صورت مربعی با ابعاد ۱۰۰ متر می‌باشد. تعداد نمونه‌ها ۳۷۷ عدد بوده و آنالیز توسط دستگاه پلاسمای جفت شده القایی (ICP) صورت گرفته و مقدار غلظت ۴۵ عنصر در این تجزیه و تحلیل گزارش شده است. موقعیت قرارگیری نمونه‌ها نسبت به یکدیگر بر روی نقشه

۳- محاسبه تابع هدف ناشی از دسته‌بندی  $\{C_1, C_2, \dots, C_k\}$  بر اساس رابطه ۲ که به محاسبه مجموع فواصل نمونه‌ها از مرکز دسته‌ها می‌پردازد.

رابطه ۲)

$$f(C_1, C_2, \dots, C_k) = \sum_{j=1}^k \sum_{x \in C_j} |x - z_j|^2$$

۴- کمینه نمودن تابع هدف رابطه ۲ و یافتن دسته‌بندی مناسب روی مجموعه  $M$  با تعداد دسته  $K$ . لازم به ذکر است که به منظور تعیین  $K$  مناسب جهت تعیین تعداد کلاس‌ها، تعداد دسته‌ها از  $K=3$  الی  $K=10$  تغییر داده شده است و گروه‌بندی‌های حاصل شده مورد تحلیل قرار گرفته‌اند. به منظور ارزیابی گروه‌های تعیین شده به ازای مقادیر مختلف  $K$ ، از رابطه مطلوبیتی که در مطالعه قنادپور و هزارخانی (۱۳۹۱) آورده شده است، بهره گرفته‌ایم. مطابق با این رابطه که در مطالعات قنادپور و هزارخانی (۱۳۹۱ و ۱۳۹۲) در مورد آن مفصل بحث شده است، میزان مطلوبیت هر نمونه در یک کلاس مورد نظر می‌تواند بین  $(-1)$  تا  $(+1)$  تغییر نماید. عدد  $(-1)$  بیان‌کننده این می‌باشد که نمونه مورد نظر در دسته‌بندی نامناسب قرار گرفته است. عدد صفر به این معنا می‌باشد که حضور نمونه مورد نظر در دسته فعلی و یا دسته دیگری دارای اهمیت چندانی نمی‌باشد و عدد  $(+1)$  نیز بیان‌کننده دسته‌بندی مناسب است. بنابراین برای هر نمونه مقدار مطلوبیتی محاسبه می‌گردد و سپس با محاسبه میانگین اعداد حاصل شده تحت عنوان میانگین مطلوبیت دسته‌بندی انجام شده، به تحلیل نتایج پرداخته می‌شود.

روش **K**- نزدیکترین همسایه:  $K$ - نزدیکترین همسایه (KNN) یک الگوریتم یادگیری می‌باشد که در روش بازشناسی الگو طی چندین دهه مطالعه شده است (یانگ و لیو، ۱۹۹۹). KNN به عنوان یکی از کاراترین روش‌ها شناخته شده است، و مطالعات زیادی KNN را روی اسناد آزمایشی

ساختار عصبی مغز انسان نامید. مکانیسم فراگیری و آموزش مغز اساساً بر تجربه استوار است (البرزی، ۱۳۸۹).

الگوریتم **K-Means**: همان‌طور که در گذشته نیز بیان گردید، یکی از روش‌های معروفی که در دیدگاه خوشه‌بندی مطرح می‌باشد، الگوریتم **K-Means** است. این روش با مقداری مشخص برای  $K$  (تعداد دسته‌ها) شروع کرده و سعی می‌نماید که مجموعه‌ای از یک نمونه‌های مشخص را در این  $K$  گروه دسته‌بندی نماید به طوری که فرضیات بیان شده در روابط ۱ و ۲ رعایت گردند (یی و ژانگ، ۲۰۱۳). همچنین لازم به ذکر است که معیار تخصیص هر نمونه به یک کلاس نیز در این روش حداقل مسافت اقلیدسی هر نمونه از نقطه میانی (مرکزی یا نماینده) هر کلاس یا دسته می‌باشد. مراحل مهمی که در این الگوریتم طی می‌گردد به طور خلاصه به شرح ذیل می‌باشد (قنادپور و هزارخانی، ۱۳۹۲):

- ۱- تشکیل  $K$  کلاس و یا دسته به صورت  $\{C_1, C_2, \dots, C_k\}$  برای خوشه‌بندی  $m$  نمونه از مجموعه  $M$ .
- ۲- محاسبه بردار  $Z_j$  بر اساس رابطه ۱ که نشان‌دهنده مرکز یا نماینده هر دسته  $C_j$  می‌باشد.

رابطه ۱)

$$Z_j = \frac{\sum_{x \in C_j} x}{\#C_j} \quad \text{for } j = 1, 2, \dots, K$$

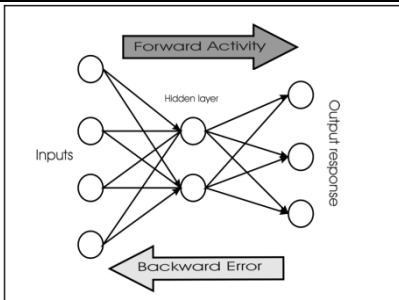
که در آن  $x$  نشان‌دهنده بردار یک نمونه‌ای است که در  $C_j$  عضو می‌باشد و  $\#C_j$  نشان‌دهنده تعداد نمونه‌هایی می‌باشد که در کلاس  $C_j$  عضو می‌باشند. لازم به ذکر است که رابطه ۱ برای محاسبه مرکز هر دسته در خلال حل استفاده می‌گردد و در شروع الگوریتم معمولاً به صورت تصادفی  $K$  نمونه انتخاب می‌گردند و به عنوان مرکز هر دسته در نظر گرفته می‌شوند.

استفاده می‌شود. اگر بیشتر از یک همسایه به دسته‌های مشابه تعلق داشته باشد، مجموع امتیاز آنها به‌عنوان وزن آن دسته استفاده می‌شود و دسته با بالاترین امتیاز به داده مورد آزمایش انتساب می‌یابد، که اگر از یک مقدار آستانه تجاوز کند، بیشتر از یک دسته می‌تواند به سند آزمایشی انتساب یابد. یک مشکل در روش KNN، تعیین مقدار K می‌باشد و برای تعیین آن باید یک سری از آزمایشات با مقادیر مختلف K انجام شود، تا بهترین مقدار برای K را تعیین کند. عیب دیگر KNN پیچیدگی زمانی محاسباتی مورد نیاز برای پیمایش همه داده‌های آموزشی می‌باشد (هی، تان و تان، ۲۰۰۰). تئوری و روابط مربوط به سه روش اشاره شده در فوق، به شکل خلاصه در جدول ۱ آورده شده است.

Reuters بکار برده‌اند، که این مطالعات پیشنهاد می‌کنند که KNN با Support Vector Machine (SVM) از روش‌هایی مانند تقریب خطی کوچکترین مربعات، Naïve Bayes و شبکه‌های عصبی بهتر عمل می‌نماید (یانگ و لیو، ۱۹۹۹؛ یانگ، ۱۹۹۹). KNN هنوز یک روش کارآ و ساده برای دسته‌بندی متن می‌باشد. ایده KNN طبق زیر است:

یک داده آموزشی برای دسته‌بندی وجود دارد، الگوریتم K همسایه نزدیک در میان داده‌های آموزشی پیش دسته‌بندی شده، بر اساس یک معیار شباهت پیدا کرده و دسته‌های این K همسایه نزدیک برای پیش‌بینی دسته داده آزمایشی به وسیله امتیازدهی داده‌های هر دسته منتخب،

جدول ۱: شرح مختصری از تئوری مربوط به روش‌های مورد استفاده (شبکه عصبی، K-Means و KNN).

| شرح مختصری از تئوری روش                                                                                                                                                                                                                                                                                                                                                                                   | روش تحت بررسی           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|
|  <p>– استفاده از قانون یادگیری پس انتشار خطا (BP) برای آموزش شبکه‌های عصبی.<br/>– این قانون متشکل از دو مسیر اصلی است: مسیر رفت: اعمال الگوی آموزشی و انتشار تأثیرات آن به لایه خروجی (در این مرحله اوزان ثابت است).<br/>مسیر برگشت: تغییر و تنظیم پارامترهای شبکه چند لایه (تغییر اوزان) بر اساس قانون تصحیح خطا.</p> | <p><b>شبکه عصبی</b></p> |
| <p>– تشکیل K کلاس پس از انتخاب K بهینه.<br/>– محاسبه مرکز یا نماینده هر دسته:</p> $z_j = \frac{\sum_{x \in C_j} x}{\#C_j} \quad \text{for } j = 1, 2, \dots, K$ <p>– محاسبه تابع هدف:</p> $f(C_1, C_2, \dots, C_k) = \sum_{j=1}^k \sum_{x \in C_j}  x - z_j ^2$ <p>– کمینه نمودن تابع هدف جهت دستیابی به دسته‌بندی مناسب.</p>                                                                             | <p><b>K-Means</b></p>   |

|            |                                                                                                                                                                                                                                                                                                                          |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>KNN</b> | <p>- انتخاب مقدار K بهینه.</p> <p>- به دست آوردن فواصل، بر اساس مشخصات ورودی.</p> <p>- تشکیل K کلاس بر حسب نزدیکترین فاصله (بیشترین شباهت) و سپس محاسبه فاصله رکورد جدید از تمام رکوردهای آموزشی.</p> <p>- انتخاب نزدیکترین همسایه.</p> <p>- استفاده از برجسب دسته K نزدیکترین همسایه، جهت پیش‌بینی دسته رکورد جدید.</p> |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### بحث و نتایج

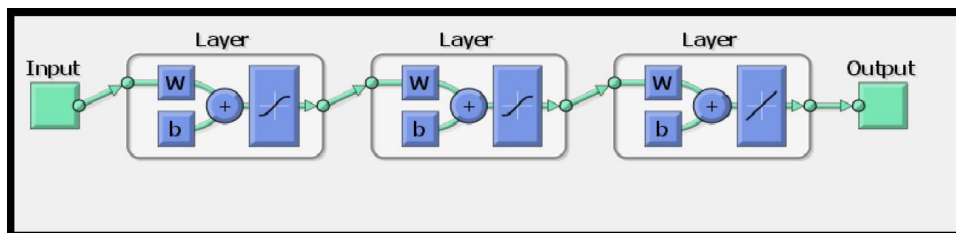
شبیه‌سازی و بررسی عنصر مس با استفاده از شبکه‌های عصبی: اولین و مهم‌ترین گام در طراحی شبکه‌های عصبی تأمین داده‌های اولیه می‌باشد به گونه‌ای که این داده‌ها خصوصیات و ویژگی‌های مورد نظر در رابطه با شبیه‌سازی را شامل شوند. تعداد و نوع داده‌ها از جمله پارامترهایی هستند که در رسیدن به هدف مطلوب ما را یاری می‌کنند. همان‌طور که گفته شد، داده‌های مورد استفاده در این بخش از تحقیق شامل ۳۷۷ نمونه آنالیز شده در مورد نمونه‌های سطحی محدوده پرکام می‌باشد. با توجه به پارامترهای انتخاب شده در شبیه‌سازی عیار مس، مختصات جغرافیایی (طول و عرض نقاط نمونه‌برداری شده) و عیار مولیبدن در هر نمونه به‌عنوان داده‌های اولیه شبکه انتخاب شده‌اند. به منظور دستیابی بر شبکه‌ی بهینه و مقایسه آنها، انواع الگوریتم‌های آموزشی زیر مجموعه الگوریتم BP به همراه شبکه‌های دارای تعداد نرون‌های مخفی متفاوت به منظور تعیین بهترین الگوریتم-های آموزشی و ساختارهای بهینه به کار رفته و عملکرد مدل‌های حاصل تحت بررسی و آنالیز قرار گرفت. با توجه به تحقیقاتی که در زمینه شبکه‌های عصبی صورت گرفته است، مشاهده می‌شود که برای اکثر مدل‌سازی‌ها (خطی و غیرخطی) وجود یک یا دو لایه میانی کفایت می‌کند. لذا در مطالعه حاضر انواع شبکه یک لایه (لایه مخفی) بررسی گردید و سرانجام شبکه‌هایی با دو لایه میانی بر

اساس آزمون‌های اولیه طراحی گشتند و در نهایت با بررسی پارامترهای مقایسه که بیان شدند، شبکه بهینه به منظور تخمین عیار مس به دست آمد. در این مرحله به منظور بررسی تأثیر توابع مختلف آموزشی بر نتایج حاصل شده از شبکه عصبی دو نوع تابع انتقال سیگموئید (تانژانتی و لگاریتمی) بررسی شده است که طبق مطالعات صورت گرفته این توابع تأثیر به‌سزایی در کارایی شبکه دارند. مطابق با آنچه بیان شد ساختار و اجزای برخی از شبکه‌های آزمایش شده به شرح جدول ۲ می‌باشند. مطابق با شرح جدول ۲ در ابتدا شبکه‌هایی با یک لایه مخفی طراحی و آزموده شدند. از میان توابع انتقال مورد استفاده و الگوریتم‌های آموزشی، الگوریتم LM در میان سایر الگوریتم‌ها برای شبکه‌های دولایه در توابع انتقال سیگموئیدی (تانژانتی و لگاریتمی) بهترین پاسخ را با مقایسه ضریب همبستگی هر شبکه، می‌دهد. بنابراین تصمیم گرفته شد تا این الگوریتم آموزشی در ۳ شبکه ۳ لایه (دو لایه مخفی) به کار گرفته شوند و از میان آنها شبکه بهینه برگزیده شود. در نهایت شبکه‌ای با ۳ لایه حاوی ۹۰ نرون در لایه اول و ۸۰ نرون در لایه دوم و یک نرون در لایه خروجی با توابع انتقال تانژانت سیگموئید به عنوان شبکه بهینه انتخاب گردید. در شکل ۲ و ۳ می‌توان ساختار شماتیک شبکه و میزان همبستگی میان داده‌ها را برای شبکه بهینه مشاهده نمود.

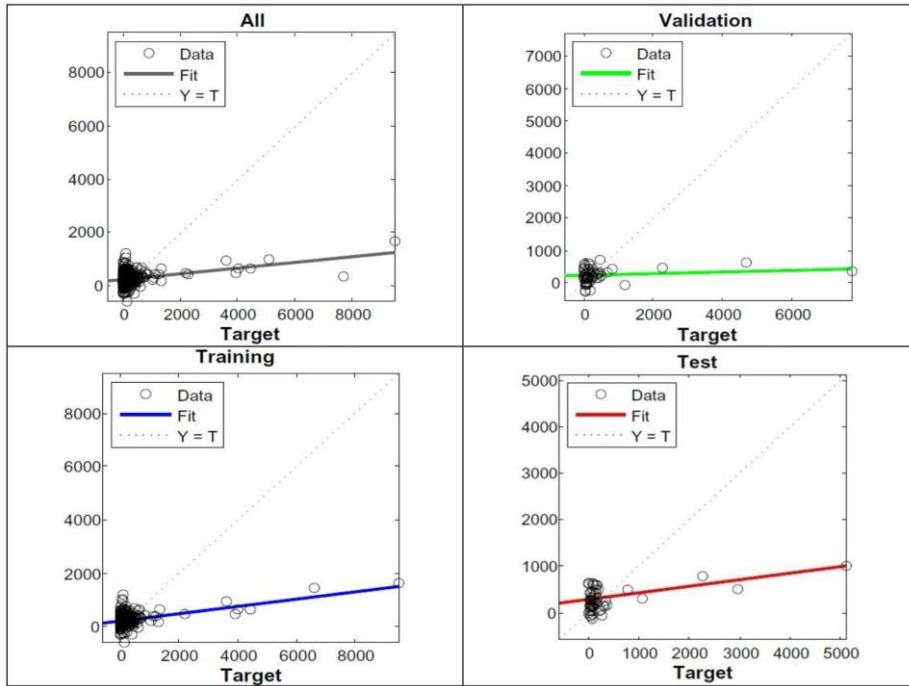


جدول ۲: خصوصیات شبکه‌های آزمایش شده.

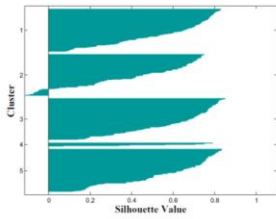
| شماره ردیف | تعداد لایه | تعداد نرون در لایه مخفی ۱ | تعداد نرون در لایه مخفی ۲ | تابع انتقال لایه ۱ | تابع انتقال لایه ۲ | الگوریتم آموزش | همبستگی بیشینه | همبستگی کلی |
|------------|------------|---------------------------|---------------------------|--------------------|--------------------|----------------|----------------|-------------|
| ۱          | ۲          | ۵۰                        | —                         | tansig             | —                  | LM             | ۰/۴۶۲          | ۰/۴۱۶       |
| ۲          | ۲          | ۵۰                        | —                         | tansig             | —                  | BFG            | ۰/۳۸۶          | ۰/۲۵۴       |
| ۳          | ۲          | ۵۰                        | —                         | tansig             | —                  | CGB            | ۰/۳۵۳          | ۰/۱۷۶       |
| ۴          | ۲          | ۵۰                        | —                         | tansig             | —                  | GDX            | ۰/۱۹۱          | ۰/۱۳۳       |
| ۵          | ۲          | ۵۰                        | —                         | tansig             | —                  | SCG            | ۰/۲۲۲          | ۰/۲۰۵       |
| ۶          | ۲          | ۵۰                        | —                         | logsig             | —                  | LM             | ۰/۴۳۷          | ۰/۳۹۱       |
| ۷          | ۲          | ۵۰                        | —                         | logsig             | —                  | BFG            | ۰/۳۴۳          | ۰/۱۹۸       |
| ۸          | ۲          | ۵۰                        | —                         | logsig             | —                  | CGB            | ۰/۳۱۸          | ۰/۱۰۳       |
| ۹          | ۲          | ۵۰                        | —                         | logsig             | —                  | GDX            | ۰/۲۷۸          | ۰/۱۳۴       |
| ۱۰         | ۲          | ۵۰                        | —                         | logsig             | —                  | SCG            | ۰/۳۲۳          | ۰/۲۰۸       |
| ۱۱         | ۳          | ۵۰                        | ۴۰                        | tansig             | tansig             | BFG            | ۰/۴۶۴          | ۰/۳۹۹       |
| ۱۲         | ۳          | ۵۰                        | ۵۰                        | tansig             | tansig             | LM             | ۰/۴۸۷          | ۰/۴۲۵       |
| ۱۳         | ۳          | ۶۰                        | ۵۰                        | tansig             | tansig             | LM             | ۰/۵۷۸          | ۰/۴۱۱       |
| ۱۴         | ۳          | ۷۰                        | ۶۰                        | tansig             | tansig             | LM             | ۰/۵۵۴          | ۰/۴۴۹       |
| ۱۵         | ۳          | ۸۰                        | ۷۰                        | tansig             | tansig             | LM             | ۰/۶۶۳          | ۰/۵۷۱       |
| ۱۶         | ۳          | ۹۰                        | ۸۰                        | tansig             | tansig             | LM             | ۰/۶۹۵          | ۰/۶۸۹       |



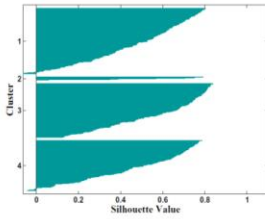
شکل ۲: معماری و طرح شبکه بهینه.



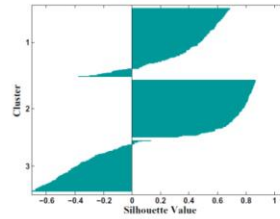
شکل ۳: نمودارهای همبستگی داده‌ها و خروجی‌ها.



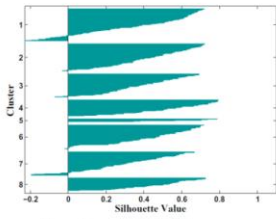
دسته‌بندی با  $K=5$  با مقدار مطلوبیت ۰/۵۳۴۶



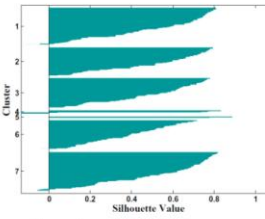
دسته‌بندی با  $K=4$  با مقدار مطلوبیت ۰/۵۰۵۵



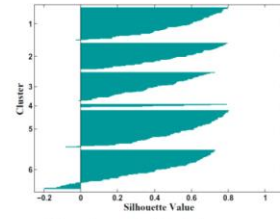
دسته‌بندی با  $K=3$  با مقدار مطلوبیت ۰/۲۸۲۸



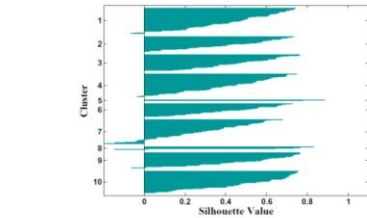
دسته‌بندی با  $K=8$  با مقدار مطلوبیت ۰/۳۹۷۹



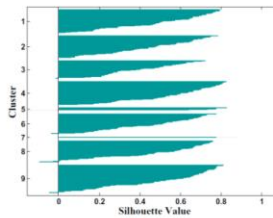
دسته‌بندی با  $K=7$  با مقدار مطلوبیت ۰/۴۹۱۳



دسته‌بندی با  $K=6$  با مقدار مطلوبیت ۰/۴۷۱۱



دسته‌بندی با  $K=10$  با مقدار مطلوبیت ۰/۴۲۷۱



دسته‌بندی با  $K=9$  با مقدار مطلوبیت ۰/۴۸۲۴

شکل ۴: نیم‌رخ کلاس‌ها و مقادیر مطلوبیت با  $K=3$  تا  $K=10$  در مورد عناصر مس و مولیبدن مربوط به نمونه‌های سطحی با در نظر گرفتن مختصات نمونه‌ها.

مطلوبیت مشاهده می‌گردد، دسته‌بندی روی نمونه‌های تعریف شده با مشخصه‌های عیار عنصر مس، عیار عنصر مولیبدن، طول و عرض نمونه‌های برداشت شده، به ازای  $K=5$  دارای بهترین دسته‌بندی می‌باشد.

تعیین رابطه بین عیار مس و مولیبدن: در این قسمت با استفاده از مفهوم رگرسیون چند متغیره و به کمک نرم‌افزار SPSS به تعیین رابطه در مورد مس و مولیبدن با توجه به طول و عرض نمونه‌ها در محدوده پراکندگی، پرداخته می‌شود. مقادیر مراکز دسته‌بندی عنصر مس به‌عنوان متغیر وابسته و مقادیر مراکز دسته‌بندی عنصر مولیبدن، طول و عرض نقاط به‌عنوان متغیرهای مستقل به نرم‌افزار معرفی می‌گردند. سپس نتایج موجود در جدول ۳ به‌عنوان مشخصات و ضرایب رگرسیون چند متغیره محاسبه و گزارش شده‌اند.

با توجه به جدول ۲، شبکه بهینه، شبکه‌ای است با سه لایه و الگوریتم آموزشی LM که همبستگی کلی آن در مورد داده‌های خروجی از شبکه و خروجی‌های واقعی،  $0/689$  گزارش شده است که میزان قابل توجهی می‌باشد. این شبکه در ۱۳ دور آموزش (epochs) به مقدار تابع کارایی  $0/075$  می‌رسد. بررسی رفتار مس و مولیبدن با استفاده از الگوریتم K-Means: در این قسمت رفتار دو عنصر مس و مولیبدن تنها در مورد داده‌های مربوط به نمونه‌های سطحی محدوده پراکندگی، با توجه به پارامترهای مختصاتی (x و y) مورد مطالعه قرار می‌گیرند. در حقیقت هدف اصلی در این قسمت تخمین عیار عنصر مس به کمک عنصر مولیبدن، طول و عرض نقاط مورد نظر می‌باشد. نیمرخ کلاس‌ها و میزان مطلوبیت دسته‌بندی‌های صورت گرفته با توجه به مقادیر  $K=3$  تا  $K=10$  در مورد مس و مولیبدن (با در نظر گرفتن طول و عرض نقاط) در شکل ۴ آورده شده است.

در نتیجه بر اساس نتایج حاصل شده به ازای مقادیر متفاوت برای K از ۳ الی ۱۰، همان‌طور که در میزان

جدول ۳: مشخصات و ضرایب مربوط به خط رگرسیون چندگانه.

| $b_3$     | $b_2$  | $b_1$ | A           | $r^2$ | R     |
|-----------|--------|-------|-------------|-------|-------|
| -۶۵۲۴/۷۹۶ | -۶/۷۷۱ | ۵/۰۷۲ | ۲۱۲۲۱۵۸۵/۵۵ | ۰/۷۴۸ | ۰/۸۶۵ |

است و بنابراین تقریباً  $0/75$  تغییرات میزان مس (y) به واسطه xها یعنی (مولیبدن، طول و عرض) است. به عبارت دیگر تقریباً  $0/25$  از تغییرات y به وسیله رگرسیون به دست نمی‌آید.

لازم به ذکر است همبستگی به دست آمده بیانگر همبستگی صفحه عبور داده شده با مراکز دسته‌ها است، لذا به منظور مقایسه این روش با دیگر روش‌ها نیاز به همبستگی بین مقادیر واقعی و تخمین زده مس می‌باشد. به این منظور همبستگی مورد

به این ترتیب در رابطه ۳ خواهیم داشت:

رابطه ۳)

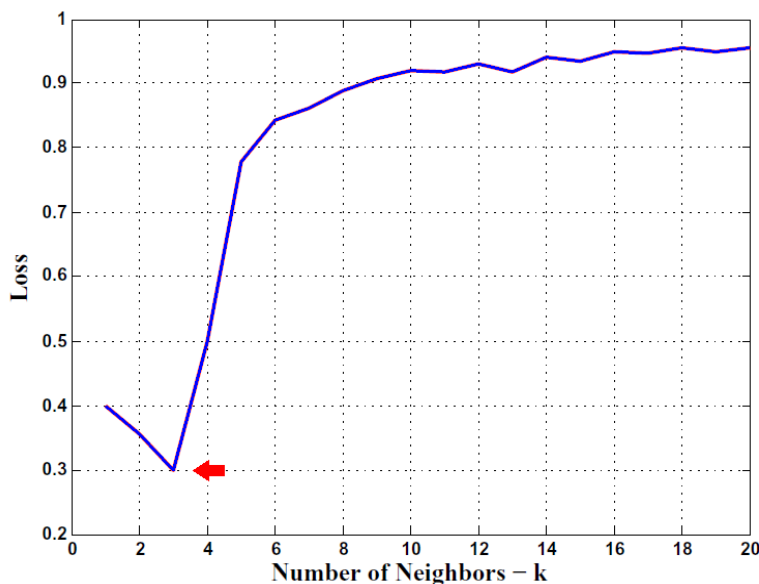
$$Cu = 21221585.55 + 5.072X - 6.771Y - 6524.796Mo$$

میزان  $r^2$  حاصل شده نشان‌دهنده سهمی است که مدل رگرسیون توانسته تغییرات را نسبت به y (یعنی Cu) توضیح دهد. در اینجا  $r^2 = 0/748$  شده

بر اساس داده‌های آموزشی می‌باشد. در مطالعه حاضر، خطای مذکور، با شمارش تعداد نمونه‌های به اشتباه پیش‌بینی شده و تقسیم آنها به تعداد داده‌های کل، به سادگی قابل محاسبه می‌باشد. با اجرای برنامه مذکور نمودار خطا نسبت به تعداد  $K$ ، تهیه شده و در شکل ۵ قابل مشاهده است. با توجه به شکل مشاهده می‌شود که بهترین مقدار برای  $K=3$  در نظر گرفته شده است. زیرا کمترین خطای Resubstitution در کلاس‌بندی با مقدار ۳ رخ داده است. لذا با مقدار ۳ شبکه مورد نظر ساخته شده و مقدار مس با توجه به سه پارامتر طول، عرض جغرافیایی و عیار عنصر مولیبدن به دست آمده است. خطای این شبکه با توجه به مقدار  $K$  برابر با ۳،  $0/303$  به دست آمده است. لذا در این روش، همبستگی بین مقادیر واقعی و تخمین زده مس تقریباً برابر با  $0/7$  گزارش شده است.

نظر با استفاده از فرمول و جایگذاری مقادیر  $X$ ،  $Y$  و عیار مولیبدن محاسبه و برابر با  $0/608$  گزارش شده است.

شبیه‌سازی عنصر مس با استفاده از الگوریتم KNN: به منظور اعمال روش KNN بر روی داده‌ها و همچنین برای تعیین بهترین مقدار  $K$ ، الگوریتم روش KNN به شکل برنامه‌ای در نرم‌افزار متلب نوشته شده است که برای مقادیر  $K$  از مقدار ۱ تا ۲۰ برنامه به تعداد ۲۰ بار اجرا می‌شود و در هر بار از اجرای آن، مقدار خطای Resubstitution اندازه‌گیری می‌شود. همچنین در مورد هر مقدار  $K$  برنامه ۵ بار الگوریتم را اجرا نموده و مقدار میانگین خطا در هر پنج مرحله را به‌عنوان خطا به  $K$  ی مورد نظر اختصاص می‌دهد. خطای Resubstitution، تفاوت بین پاسخ داده‌های آموزشی و پاسخ پیش‌بینی‌های صورت گرفته توسط شبکه طراحی شده مورد نظر



شکل ۵: نمودار مقدار خطای میانگین حاصل از ۵ بار اجرای الگوریتم KNN برای مقادیر مختلف  $K$ .

## نتیجه‌گیری

همان‌طور که مشاهده گردید، در تحقیق فوق به منظور تعیین بهترین روش در امر شبیه‌سازی برای مواقعی که با تعداد داده‌های محدودی روبرو هستیم، سه روش شبکه‌های عصبی، روش پرکاربرد K- نزدیکترین همسایه (KNN) و K میانگین (K-Means) مورد استفاده قرار گرفتند. با استفاده از شبکه‌های عصبی، در نهایت پس از بررسی و مقایسه‌های صورت گرفته، شبکه‌ای با سه لایه و الگوریتم آموزشی LM که همبستگی کلی داده‌های خروجی از شبکه و خروجی‌های واقعی آن ۰/۶۸۹ است به‌عنوان شبکه بهینه انتخاب گردید. این شبکه در ۱۳ دور آموزش به مقدار تابع کارایی ۰/۰۷۵ می‌رسد. در بخش دیگر، در ابتدا روش K-Means به‌عنوان یکی از روش‌های رایج کلاسه‌بندی و در نهایت روش KNN به‌عنوان یکی از روش‌های معروف و کارآمد پیش‌بینی در مبحث داده‌کاوی بر روی داده‌ها اعمال شدند. محدودیت اصلی در انتخاب و استفاده از این دو روش انتخاب K بهینه و سپس اجرای الگوریتم آنها با توجه به مقدار K می‌باشد. در روش اول پس از انتخاب K بهینه (K=۵) به کمک رابطه مطلوبیت گفته شده در متن مقاله و سپس با استفاده از مفهوم رگرسیون چند متغیره،

رابطه  $Cu = 21221585.55 + 5.072X - 6524.796Mo - 6.771Y$  با ضریب همبستگی ۰/۷۴۸ به منظور تخمین عیار عنصر مس بیان گردید. در این روش همبستگی داده‌های خروجی از رابطه مذکور و مقدار واقعی مس، برابر با ۰/۶۰۸ گزارش شده است. سپس با اجرای ۱۰۰ مرتبه از الگوریتم روش KNN بر روی داده‌ها (برای ۲۰ مقدار K و برای هر K ۵ بار اجرا) به منظور دسترسی به کمترین خطای ممکن، مقدار K=۳ به‌عنوان بهترین مقدار K انتخاب و با توجه به آن روش مذکور بر روی داده‌ها اعمال شد. در نهایت نیز با توجه به نتایج حاصله مشاهده گردیده که روش KNN نسبت به دو روش دیگر، همان‌طور که بیان شد برای مجموعه‌هایی با تعداد محدودی داده، با نظر به اینکه همبستگی به دست آمده از آن، مقدار بیشتری نسبت به شبکه عصبی و K-Means را نشان می‌دهد، مناسب‌تر می‌باشد. همچنین می‌توان اظهار داشت که اگر به جای یک صفحه، یک رویه فضایی منطبق با چیدمان مراکز داده‌ها به آنها برآزش شود، به این ترتیب روش K-Means نتایج مطلوب‌تری فراهم خواهد کرد و ارائه‌گر الگویی مشخص و دقیق‌تر به منظور تخمین عیار به تصمیم‌گیران این صنعت خواهد بود.

## منابع

- البرزی، م.، ۱۳۸۹. آشنایی با شبکه‌های عصبی، انتشارات دانشگاه صنعتی شریف، ؟ ص.
- زهرایی، ب. و تکشی، آ.، ۱۳۸۷. کاربرد روش‌های الگوریتم ژنتیک و K- نزدیک‌ترین همسایه در تدوین سیاست‌های بهره‌برداری از مخزن در زمان وقوع سیلاب، نشریه تحقیقات منابع آب، نشریه تحقیقات منابع آب ایران، دوره ۳، شماره ۱۲، ص ۳۷-۲۷.
- قنادپور، س.س. و هزارخانی، ا.، ۱۳۹۱. بررسی رفتار سرب نسبت به روی و آهن در کانسار مس پورفیری پرکام، شهر بابک، کرمان با استفاده از روش گروه بندی، نشریه علمی- پژوهش‌های دانش زمین، سال سوم، شماره ۹، ص ۶۴ - ۷۷.
- قنادپور، س.س. و هزارخانی، ا.، ۱۳۹۲. بررسی چگونگی رفتار عنصر مس نسبت به عناصر مولیبدن، سرب و روی در کانسار مس پورفیری پرکام در استان کرمان، با استفاده از روش K-

ژئوشیمیایی مس و مولیدن در منطقه پرکام، کرمان. نشریه علمی- پژوهش‌های دانش زمین، سال ششم، شماره ۲۱، ص ۴۰ - ۵۰.

- Andrew, F., Weller, A.J., Harris, J. and Andrew, W., 2007. Two Supervised Neural Networks for Classification of Sedimentary Organic Matter Images from Palynological Preparations: *Mathematical Geology*, v. 39, p. 657-671.
- Audibert, J.Y. and Tsybakov, A.B., 2007. Fast learning rates for plug-in classifiers under the margin condition: *Annals of Statistics*, v. 35, p. 608-633.
- Bax, E., 2000. Validation of nearest neighbor classifiers: *Ieee transactions information theory*, v. 58, p. 2746-2752.
- Bhattacharya, G., Ghosh, K. and Chowdhury, A.S., 2012. An affinity-based new local distance function and similarity measure for kNN algorithm: *Pattern Recognition Letters*, v. 33, p. 356-363.
- Deegalla, S. and Boström, H., 2007. Classification of Microarrays with kNN: Comparison of Dimensionality Reduction Methods: *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, v. 4881, p. 800-809.
- Funahashi, K.I., 1989. On the approximate realization of continuous mappings by neural networks: *Neural Networks*, v. 2, p. 193.
- Ghannadpour, S.S. and Hezarkhani, A., 2015. Investigation of Cu, Mo, Pb and Zn Geochemical behavior and geological interpretations for Parkam Porphyry Copper system, Kerman, Iran: *Arabian Journal of Geosciences*, v. 8, p. 7273-7284.
- He, J., Tan, A. and Tan, C., 2000. Comparative Study on Chinese Text Categorization Methods, On the PRICAI

Means, مجله زمین‌شناسی کاربردی پیشرفته، شماره ۷، ص ۵۳ - ۶۳.

-قنادپور، س.س. و هزارخانی، ا.، ۱۳۹۲. برآورد مناطق امید بخش جهت تهیه نقشه‌های ناهنجاری

- 2000 Workshop on Text and Web Mining, Melbourne, p. 25-31.
- Kamel, N., Ouchen, I. and Baali, K., 2014. A Sampling-PSO-K-means Algorithm for Document Clustering, *Genetic and Evolutionary Computing: Springer International Publishing Switzerland*, 238 p.
- Kuo, R.J., Suryani, E. and Yasid, A., 2013. Automatic Clustering Combining Differential Evolution Algorithm and k - Means Algorithm: *Proceedings of the Institute of Industrial Engineers Asian Conference 2013*, Springer Singapore.
- Lacassie, J.P., Roser, B.P., Ruiz-del-Solar, J., Roser, B. and Hervé, F., 2006. Visualization of Volcanic Rock Geochemical Data and Classification with Artificial Neural Networks: *Math Genealogy*, v. 38, p. 697-710.
- Meshkani, S.A., Mehrabi, B., Yaghubpur, A. and Alghalandis, Y.F., 2011. The application of geochemical pattern recognition to regional prospecting: A case study of the Sanandaj-Sirjan metallogenic zone, Iran: *Journal of Geochemical Exploration*, v. 108, p. 183-195.
- Mora, J.L., Armas-Herrera, C.M., Guerra, J.A., Rodríguez, A. and Arbelo, C.D., 2012. Factors affecting vegetation and soil recovery in the Mediterranean woodland of the Canary Islands (Spain): *Journal of Arid Environments*, v. 87, p. 58-66.
- Ozturk, M.M. and Cavusoglu, U. and Zengin, A., 2015. A novel defect prediction method for web pages using k-means++: *journal of Export Systems with Applications*, v. 42, p. 6496-6506.

- Poloczek, J., Treiber, N.A. and Kramer, O., 2014. KNN Regression as Geo-Imputation Method for Spatio-Temporal Wind Data: International Joint Conference SOCO'14-CISIS'14-ICEUTE'14, v. 299, p. 185-193.
- Shang, W., Huang, H., Zhu, H., Lin, Y., Qu, Y. and Dong, H., 2006. An Adaptive Fuzzy kNN Text Classifier: Computational Science – ICCS 2006, p. 216-223.
- Singer, D.A., 2006. Typing mineral deposits using their associated rocks and grades and tonnages in a probabilistic neural network: Mathematics Genealogy, v. 38, p. 465-475.
- Tarkian, M. and Stribrny, B., 1999. Platinum-group elements in porphyry copper deposits: a reconnaissance study: Mineralogy and Petrology, Springer-Verlag, v. 65, p. 161-183.
- Varaprasad, M., 2012. Algorithm for Clustering with Intrusion Detection Using Modified and Hashed K – Means Algorithms: Engineering & Applications, Springer-New Delhi, India, v. 167, p. 737-744.
- Weller, A.F., Corcoran, J., Harris, A.J. and Ware, J.A., 2005. The semi-automated classification of sedimentary organic matter in palynological preparations: Computer and Geoscience, v. 31, p. 1213-1223.
- Wu, X. and Zhou, Y., 1993. Reserve estimation using neural network techniques: Computer and Geoscience, v. 19, p. 567-575.
- Xu, H., Lu, Sh. and Zhou, Sh., 2013. A Novel Algorithm for Text Classification Based on KNN and Chaotic Binary Particle Swarm Optimization: Proceedings of the 2012 International Conference on Information Technology and Software Engineering, v. 211, p. 619-627.
- Yang, Y., 1999. An evaluation of statistical approaches to text categorization: Journal of Information Retrieval, v. 1, p. 69-90.
- Yang, Y. and Liu, X., 1999. A re-examination of text categorization methods: In proceedings of the 22<sup>nd</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99), p. 42-49.
- Yang, J., Zhuang, Y. and Wu, F., 2012. ESVC- based extraction and segmentation of texture features: Computers and Geosciences, v. 49, p. 238-247.
- Yi, X. and Zhang, Y., 2013. Equally contributory privacy-preserving K-Means clustering over vertically partitioned data: Elsevier, Information Systems, v. 38, p. 97-107.
- Zhang, C., 2013. Study on the Application of Fuzzy KNN to Chinese and English Recognition: Proceedings of the 2012 International Conference of Modern Computer Science and Applications, v. 191, p. 327-332.
- Zu-Feng, W., Xiao-Fan, M., Qiao, L. and Zhi-guang, Q., 2014. Logical Symmetry Based K-means Algorithm with Self-adaptive Distance Metric: Advances in Computer Science and its Applications, Springer-Verlag, Berlin Heidelberg. v. 279, p. 929-936.