



کتابچه راهنمای سومین سمینار تخصصی علم داده ها و کاربردهای آن

۲۱ و ۲۲ آذرماه ۱۴۰۳

دانشگاه فردوسی مشهد

کتابچه راهنمای سومین سمینار تخصصی علم داده‌ها و کاربردهای آن
تدوین، تنظیم و صفحه‌آرایی: مینا مقبول حامدی
ویراستار: ترنم ترابی نعمان
طراح جلد: علیرضا سحرخیز
تاریخ انتشار: آذر ۱۴۰۳

پیشگفتار

با افتخار، دانشکده علوم ریاضی دانشگاه فردوسی مشهد میزبان سومین سمینار تخصصی علم داده‌ها و کاربردهای آن است و بدین وسیله، از حضور ارزشمند پژوهشگران و علاقه‌مندان گرامی استقبال می‌نماید. این سمینار، نتیجه بیش از یک سال تلاش مداوم تیم‌های علمی و اجرایی است و امید داریم که برنامه‌های ارائه شده رضایت خاطر شما را فراهم آورند. برگزاری این رویداد علمی نشان‌دهنده قابلیت‌ها و پویایی دانشگاه فردوسی مشهد و تلاش خستگی‌ناپذیر مسئولان دانشگاه و دانشکده برای ایجاد محیطی علمی و پویا برای تبادل نظر و مشارکت علمی است.

در این دوره، از بین ۸۶ مقاله ارسالی به دبیرخانه، پس از بررسی‌های دقیق و تخصصی، تعداد ۷۵ مقاله برای ارائه شفاهی به صورت سخنرانی پذیرفته شده است. شایسته است از زحمات صادقانه و همکاری صمیمانه اعضای کمیته‌های علمی و اجرایی، استادان ارجمند، دانشجویان و کارمندان دانشکده علوم ریاضی که در برپایی این رویداد نقش برجسته‌ای ایفا کردند، قدردانی کنیم.

در پایان، از تمامی اعضای کمیته‌های راهبری، علمی، اجرایی، دانشجویی، داوران محترم، انجمن‌های آمار ایران و هوش مصنوعی ایران و سازمان‌های حامی که با همراهی و حمایت خود به برگزاری این سمینار یاری رساندند، تشکر ویژه داریم.

کمیته برگزاری سومین سمینار تخصصی علم داده‌ها و کاربردهای آن

دانشگاه فردوسی مشهد

آذر ماه ۱۴۰۳

فهرست مندرجات

صفحه

عنوان

۱.....	اعضای کمیته راهبری
۲.....	اعضای کمیته علمی.....
۳.....	اعضای کمیته اجرایی.....
۴.....	کمیته دانشجویی.....
۵.....	حامیان سمینار.....
۷.....	محورهای سمینار.....
۸.....	چکیده مقالات فارسی
۴۸.....	چکیده مقالات انگلیسی.....



اعضای کمیته راهبری

دانشگاه فردوسی مشهد (دبیر سمینار)

دکتر محمد آرشی

دانشگاه فردوسی مشهد

دکتر مهدی دوست‌پرست

دانشگاه فردوسی مشهد

دکتر عبدالحمید رضائی رکن آبادی

دانشگاه فردوسی مشهد

دکتر مجید سرمد

دانشگاه فردوسی مشهد

دکتر احمد عرفانیان مشیری‌نژاد



اعضای کمیته علمی

دانشگاه فردوسی مشهد(دبیر علمی)	دکتر محمد آرشی
دانشگاه فردوسی مشهد	دکتر سهیلا اشکذری طوسی
دانشگاه فردوسی مشهد/دانشگاه حکیم سبزواری	دکتر محمود امین طوسی
دانشگاه تهران	دکتر سید مرتضی امینی
دانشگاه علامه طباطبایی	دکتر فرزاد اسکندری
دانشگاه علوم پزشکی مشهد	دکتر حبیب الله اسماعیلی
دانشگاه بیرجند (نماینده انجمن آمار ایران)	دکتر حمیدرضا نیلی ثانی
دانشگاه خلیج فارس	دکتر حسین حق‌بین
دانشگاه سمنان	دکتر مهدی روزبه
دانشگاه تهران	دکتر هدیه ساجدی
دانشگاه فردوسی مشهد	دکتر هادی صدوقی یزدی
موسسه آموزش عالی الکترونیکی ایرانیان و مدیر مرکز هوش مصنوعی شرکت ملی انفورماتیک	دکتر الهام فرهانی
دانشگاه امام رضا مشهد	دکتر عادل قاضی خانی
دانشگاه فردوسی مشهد	دکتر رضا قنبری
دانشگاه تهران	دکتر زهرا رضایی قهرودی
دانشگاه فردوسی مشهد	دکتر حمیدرضا کوشا
دانشگاه فردوسی مشهد	دکتر جلال الدین نصیری



اعضای کمیته اجرایی

دانشگاه فردوسی مشهد	دکتر محمد آرشی
دانشگاه فردوسی مشهد	دکتر هادی جباری نوقابی
دانشگاه فردوسی مشهد (دبیر کمیته)	دکتر مصطفی رزمخواه
دانشگاه فردوسی مشهد	دکتر مهدی عمادی
دانشگاه خوارزمی تهران	دکتر خاطره قربانی مقدم
دانشگاه فردوسی مشهد	دکتر جلال الدین نصیری
دانشگاه فردوسی مشهد	آقای سید حسن احمدی
دانشگاه فردوسی مشهد	آقای مهندس رضا احمدئی



کمیته دانشجویی

- الهه اخلاقی نیا
- سید مرتضی پنج‌تنی
- ترنم ترابی نعمان
- سارا ترک‌پور
- احسان تقوی
- مینا رحیمی‌پور
- علیرضا سحرخیز
- امیرعلی سطوتی
- سید امیرحسین طباطبایی شیرازی
- طاهره عالمی
- عرشیا عنبریان
- یلدا فتحی
- آرمیتا کریمی
- سیده فاطمه مصطفوی
- مینا مقبول حامدی
- سمانه نظری
- الهه نیکنامی
- عماد هاشم‌آبادی



حامیان سمینار

	دانشگاه فردوسی مشهد		مرکز آمار ایران
	دانشگاه تهران		دانشگاه علوم پزشکی مشهد
	دانشگاه صنعتی جندی شاپور دزفول		دانشگاه علامه طباطبائی
	پایگاه علوم جهان اسلام		انجمن هوش مصنوعی ایران
	آستان قدس رضوی		سیولیکا
	انجمن صنفی کارفرمایی شرکت‌های دانش بنیان		شرکت نقشه و مسیریاب نشان
	انجمن علمی آمار دانشگاه فردوسی مشهد		بهیویان



اوج رایانه گستر شریف



آزمایشگاه علم داده
دانشکده علوم ریاضی
دانشگاه فردوسی مشهد



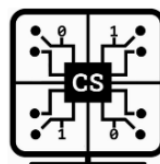
دفتر تحقیقات کاربردی فرماندهی
انتظامی استان خراسان رضوی



مدرسه داتین



موسسه آموزشی و تحقیقاتی صنایع
دفاعی



انجمن علمی علوم کامپیوتر
دانشگاه فردوسی مشهد



داتین



دانشگاه صنعتی مالک اشتر



پارت لاستیک



دانشکده علوم ریاضی
دانشگاه فردوسی مشهد



محورهای سمینار

- تحلیل داده‌های دیجیتال و امنیت سایبری
- علم داده‌ها در حوزه مالی و پولی
- آمار و داده‌کاوی
- هوش مصنوعی در علم داده‌ها
- بهینه‌سازی و علم داده‌ها
- علم داده‌ها و سیستم‌های کنترلی
- علم داده‌ها در داده‌های بزرگ
- احراز هویت دیجیتال
- علم داده‌ها در آمار زیستی و اپیدمیولوژی
- علم داده‌ها در پزشکی
- علم داده‌ها در صنعت، تجارت و بازرگانی
- هوش مصنوعی در شناسایی، ردیابی و پیش‌بینی جغرافیایی جرائم



چکیده مقالات فارسی

(به ترتیب حروف الفبا)

عنوان مقاله ۱: احراز هویت دیجیتال و امنیت

نویسنده/نویسندگان: سعیده حسینی نیا

چکیده:

احراز هویت دیجیتال به فرآیند تأیید هویت کاربران در سیستم‌های دیجیتال اشاره دارد که از اهمیت فزاینده‌ای در امنیت سایبری و مدیریت دسترسی برخوردار است. با پیشرفت‌های تکنولوژیک و افزایش نگرانی‌ها در مورد امنیت داده‌ها، روش‌های مختلف احراز هویت دیجیتال در حال توسعه و بهبود هستند. این مقاله مروری به بررسی روش‌های مختلف احراز هویت دیجیتال، چالش‌های مرتبط و روندهای جدید در این حوزه می‌پردازد.

تحول دیجیتال در بخش بانکداری عصر جدیدی از فرصت‌ها و چالش‌ها را به ارمغان آورده است، به‌ویژه در حوزه تأیید هویت. این مقاله به بررسی تحول و اهمیت مکانیزم‌های تأیید هویت دیجیتال، به‌ویژه در زمینه فرآیندهای KYC (شناخت مشتری) که برای تضمین امنیت و تمامیت تراکنش‌های مالی در عصر دیجیتال ضروری هستند، می‌پردازد. با استفاده از یک مرور جامع ادبیات، مقاله به همگرایی فناوری‌هایی مانند یادگیری ماشین، ارتباطات 5G و بلاک‌چین در شکل‌دادن به آینده تأیید هویت دیجیتال در بانکداری پرداخته است.

مطالعات موردی بینش‌های عملی در مورد پیاده‌سازی این فناوری‌ها ارائه می‌دهند و پتانسیل تحول‌آفرین و چالش‌های آن‌ها را برجسته می‌سازند. تحلیل یافته‌ها در چارچوب مرور ادبیات، درک عمیق‌تری از پیامدهای واقعی فناوری‌های تأیید هویت دیجیتال ارائه می‌دهد. مقاله به مزایای این فناوری‌ها، از جمله امنیت تقویت‌شده، کارایی بهتر و بهبود تجربه مشتری، تأکید می‌کند، در عین حال به چالش‌هایی مانند نگرانی‌های امنیتی، مشکلات تعامل‌پذیری و تغییرات مقرراتی نیز پرداخته می‌شود.



علاوه بر این، تحقیق بر اهمیت اعتماد، شفافیت و سهولت استفاده در سیستم‌های تأیید هویت دیجیتال تأکید می‌کند. نتیجه‌گیری مقاله به آینده امیدوارکننده تأیید هویت دیجیتال در بانکداری اشاره دارد و نیاز به نوآوری مستمر و سازگاری با چالش‌های نوظهور را مورد تأکید قرار می‌دهد. به‌طور کلی، این مقاله نمایی جامع از منظر تأیید هویت دیجیتال در بخش بانکداری ارائه می‌دهد و بینش‌های ارزشمندی برای دینفعان، محققان و سیاست‌گذاران علاقمند به تحول دیجیتال بانکداری و پیامدهای آن برای تأیید هویت فراهم می‌آورد.

کلمات کلیدی: تأیید هویت دیجیتال، بلاک‌چین، KYC، فناوری‌ها، فین‌تک



عنوان مقاله ۲: استفاده از روش‌های داده‌کاوی جهت بررسی روندهای مطالعاتی در حوزه منابع طبیعی

نویسنده/نویسندگان: امیررضا پورکاظمی سالکویه و جلال الدین نصیری

چکیده:

این مقاله با بهره‌گیری از روش‌های داده‌کاوی به بررسی روندهای پژوهشی در حوزه منابع طبیعی طی دو دهه اخیر می‌پردازد. هدف اصلی این پژوهش، شناسایی شکاف‌های تحقیقاتی و تعیین موضوعات داغ در این حوزه است. برای تحلیل داده‌های مقالات استخراج‌شده در این مطالعه، از مدل‌سازی موضوعی و تحلیل فراوانی کلیدواژه‌ها استفاده شده است. داده‌های مقالات شامل پراکندگی جغرافیایی، تعداد نویسندگان، زمان چاپ، و کلیدواژه‌های مرتبط با هر مقاله بررسی و تحلیل شده است تا روندهای پژوهشی شناسایی شوند. به علاوه، عملکرد نویسندگان و نشریات برتر نیز از طریق مقایسه مقالات پراستناد و تأثیرگذار مورد بررسی قرار گرفته است. نتایج نشان می‌دهد که موضوعاتی مانند حفاظت از خاک، کیفیت هوا، فرسایش بادی و مدیریت منابع آبی از مهم‌ترین زمینه‌های پژوهشی در این دوره بوده‌اند. همچنین، تغییرات قابل توجهی در فراوانی استفاده از کلیدواژه‌های مرتبط با این موضوعات در طول زمان مشاهده شده است. تحلیل جغرافیایی نشان داد که استان‌های خاصی مانند تهران و البرز فعالیت بیشتری در زمینه پژوهش‌های محیط‌زیستی دارند. این پژوهش با ارائه تصویری جامع از روندهای موجود، به تعیین اولویت‌های تحقیقاتی آینده کمک می‌کند.

کلمات کلیدی: داده‌کاوی، منابع طبیعی، روندهای پژوهشی، مدل‌سازی موضوعی، خلأ تحقیقاتی



عنوان مقاله ۳: استفاده از سری‌های زمانی جانشین در رده‌بندی سری‌های زمانی در دامنه طیفی

نویسنده/نویسندگان: ابوالقاسم بادلی، منوچهر بابانژاد و مجید عظیم محسنی

چکیده:

استنباط آماری برای سری‌های زمانی به دلیل وجود ارتباط سریالی بین داده‌ها پیچیده می‌باشد. تولید سری‌های زمانی جانشین، این امکان را فراهم می‌کند که بتوان روش‌های بازنمونه‌گیری را برای استنباط آماری در دامنه زمان و طیفی سری‌های زمانی را انجام داد. در این مقاله از سری‌های زمانی جانشین در رده بندی سری‌های زمانی در دامنه طیفی استفاده می‌شود. بر اساس مشاهدات واقعی از سری‌های زمانی بار رسوبی رودخانه گرگان‌رود به ارزیابی روش پرداخته می‌شود.

کلمات کلیدی: رده بندی، سری زمانی جانشین، دامنه طیفی، دوره نگار، فاصله کولبک-لیبلر، سری

زمانی بار رسوبی



عنوان مقاله ۴: استقرار یک چارچوب مبتنی بر هوش مصنوعی و علم داده در فاز طراحی مدیریت چرخه عمر محصول

نویسنده/نویسندگان: رشید فریدینیا

چکیده:

اخیراً فناوری هوش مصنوعی (AI) توجه گسترده‌ای را در زمینه صنعت و توسعه محصول به خود جلب کرده است. مدیریت چرخه عمر محصول (PLM) پارادیمی است که کلیه فعالیت‌های فکری مربوط به یک محصول در کل چرخه عمر آن - از آغاز یک مفهوم ناملموس تا بازیافت محصول فیزیکی نهایی، را در بر می‌گیرد. در زمینه بکارگیری هوش مصنوعی در فرآیند توسعه محصول، این مقاله با مرور نظریه‌ها، الگوریتم‌ها و فناوری‌های مختلف هوش مصنوعی در مراحل مختلف PLM (یعنی طراحی محصول، ساخت و خدمات)، مدلی را برای جهت بکارگیری هوش مصنوعی در چرخه عمر توسعه محصول ارائه می‌کند. و یک انگاشت و نقشه راه ساختاریافته برای کاربرد هوش مصنوعی در PLM و بیان فرصت‌های و چالش‌های بکارگیری هوش مصنوعی در فاز طراحی ارائه شده است.

کلمات کلیدی: هوش مصنوعی، علم داده، چرخه عمر محصول، یادگیری ماشین، فازهای توسعه محصول، صنعت خودرو، داده‌های حجیم



عنوان مقاله ۵: انتخاب ویژگی‌های موثر سیگنال‌های مغزی به منظور تشخیص اختلال نارسایی توجه/ فزون کنشی با استفاده از روش انتخاب ویژگی ترکیبی

نویسنده/نویسندگان: وفا پورتو کلی، صفا پورتو کلی و کامراد خوشحال رودپشتی

چکیده:

اختلال نارسایی توجه/ فزون کنشی (ADHD) یک اختلال رفتاری مشترک است که ممکن است در ۵ تا ۸ درصد از کودکان مشاهده شود. تشخیص اولیه ADHD برای درمان بیماری و کاهش اثرات مضر آن بر آموزش، اشتغال، روابط و کیفیت زندگی بسیار مهم است. از طرفی روش‌های انتخاب ویژگی در پردازش سیگنال‌های الکتروانسفالوگرام (EEG) کاربرد گسترده‌ای دارند. ثابت شده است که فعالیت عصبی مغز و سیگنال‌های EEG مرتبط با آن دارای رفتاری پراکنده است. از اینرو، برای طبقه بندی سیگنال‌های EEG می‌توان از روش‌های انتخاب ویژگی ترکیبی استفاده کرد. در این مقاله، یک روش جدید بر اساس ترکیبی از ویژگی‌های استخراج شده شامل ۶۵ شرکت کننده که ۲۲ نفر مبتلا به ADHD و ۴۳ نفر سالم می‌باشند. روش‌های دسته‌بندی شامل ماشین بردار پشتیبان، K نزدیک‌ترین همسایه و ترکیبی برای مرحله ارزیابی استفاده شده است. آزمایش به دست آمده بهبود قابل توجهی در مقایسه با روش‌های دسته‌بندی در شناسایی و طبقه بندی افراد مبتلا به ADHD فراهم می‌کند.

کلمات کلیدی: اختلال نارسایی توجه/ فزون کنشی (ADHD)، سیگنال‌های مغزی، ماشین بردار پشتیبان (SVM)

عنوان مقاله ۶: آسیب‌شناسی حکمرانی داده‌های اقتصادی در ایران

نویسنده/نویسندگان: فاضل مریدی فریمانی و مهدی پهلوانی

چکیده:

هدف: شناسایی چالش‌های حکمرانی داده‌های اقتصادی و دلایل عدم بکارگیری این داده‌ها در امر سیاست‌گذاری از مهمترین اهداف این پژوهش است. ضرورت: با گسترش فرایندهای الکترونیکی در بدنه دولت، حجم عظیمی از داده‌ها تشکیل شده است که استفاده کافی از آن نمی‌شود؛ در حالیکه این داده‌ها، ظرفیت خوبی در گره‌گشایی امور حکمرانی اقتصادی دارند. سیاست‌گذاری کورکوانه، گسترش فساد در نتیجه عدم ردیابی فعالیت‌های اقتصادی، عدم شناسایی جریانهای درآمدی، ناتوانی در یافتن حد بهینه حریم خصوصی و انتفاع جامعه، ناتوانی در نگهداری صحیح داده و ... از مواردی هستند که در نتیجه حکمرانی داده‌ای غلط، اجتناب‌ناپذیر خواهد بود. روش شناسی: این پژوهش از نوع کاربردی و روش آن توصیفی-تحلیلی، مرور متون و پیمایش است. ابزارگردآوری داده‌ها از طریق مطالعه اسناد کتابخانه‌ای و همچنین مصاحبه با کارشناسان و مدیران دستگاه‌های دولتی، فعال در حوزه حکمرانی و مدیریت داده بوده است. تجزیه و تحلیل اطلاعات و نتایج گردآوری شده از منابع اطلاعاتی، به صورت تفسیر استنتاجی و نگاه آسیب‌شناسانه بوده است. یافته‌ها: مشکل در حکمرانی داده در اقتصاد ایران عمدتاً ریشه در مسائل قانونی، اداری و اجرایی دارد. چالش‌هایی مانند تقنین کم عمق، احساس مالکیت سازمانی بر داده، عدم شناخت داده‌های موجود و ... از جمله مواردی هستند که در صورت حل شدن، به بهره‌برداری بهینه از منابع داده‌ای کمک خواهند کرد. در صورت نگاه موشکافانه، مسائل دیگر مانند پایین بودن دانش نیروی انسانی نیز که در ظاهر جزو مسائل فنی موضوع تلقی می‌شوند، ریشه‌های اداری و قانونی دارند. نتیجه‌گیری: بررسی‌ها نشان می‌دهد که اقتصاد ایران در بکارگیری و استفاده کارا از داده‌های اقتصادی ناتوان بوده و مواردی مانند تخصیص نابهینه در حوزه‌های انرژی، بازار کار، یارانه‌های دولتی و ... شاهدی بر این ادعاست. عدم استفاده از داده نه ناشی از فقدان داده، بلکه منتج از ضعف در حکمرانی و استفاده از داده‌های موجود است و با توجه به جدید بودن موضوع حتی در سطح دنیا، نیاز است که دولت ایران با توجه به اقتضائات مربوط به خود به دنبال راه حل ویژه خود باشد.

کلمات کلیدی: داده، حکمرانی داده، داده‌های اقتصادی، داده‌های مالی، سیاست‌گذاری اقتصادی



عنوان مقاله ۷: آشنایی با مدل‌های سری زمانی شمارشی و کاربرد آن در علم داده

نویسنده/نویسندگان: سیده طیبه جعفری پور، وحید فکور و تکتّم اکبری خلج

چکیده:

سری‌های زمانی به عنوان یکی از ارکان اساسی آمار، نقش مهمی در تحلیل داده‌های انبوه ایفا می‌کنند. در عصر حاضر، با افزایش حجم داده‌ها، نیاز به مدل‌های پیشرفته آماری برای تجزیه و تحلیل این اطلاعات بیش از پیش احساس می‌شود. بدین منظور، مدل‌های GARMA و GLARMA به طور خاص برای پردازش داده‌های شمارشی طراحی شده‌اند و قابلیت‌های منحصر به فردی در شناسایی الگوها و پیش‌بینی روندها دارند. در این سخنرانی پس از آشنایی با اصول و ساختار این مدل‌ها، کاربرد آن‌ها در تحلیل داده‌های تنفسی مورد بررسی قرار خواهد گرفت. این مدل‌ها می‌توانند به عنوان ابزاری مؤثر در پژوهش‌های آینده در زمینه بهداشت و پزشکی مورد استفاده قرار گیرند.

کلمات کلیدی: سری زمانی، داده‌های شمارشی، GARMA، GLARMA



عنوان مقاله ۸: بررسی تاثیر مدیریت دانش بازاریابی در عملکرد بازاریابی فدراسیون فوتبال جمهوری اسلامی ایران

نویسنده/نویسندگان: مجید خرسندی فرد و حسین طهوریان

چکیده:

تحقیق حاضر با هدف بررسی تاثیر مدیریت دانش و اطلاعات بازاریابی در عملکرد بازاریابی فدراسیون فوتبال جمهوری اسلامی ایران انجام شد. جامعه آماری تحقیق شامل کارکنان فدراسیون فوتبال جمهوری اسلامی ایران می باشد که تعداد آنها ۸۵ نفر بوده و با توجه به محدودیت جامعه تمامی افراد به عنوان نمونه آماری تحقیق انتخاب شدند. ابزار تحقیق شامل پرسشنامه دانش بازاریابی آکروش و آل محمد (۲۰۱۰) و پرسشنامه عملکرد بازاریابی کاتلر (۲۰۰۷) می باشد که روایی آنان توسط ده نفر از اساتید مدیریت ورزشی مورد تایید قرار گرفت و پایایی آنها به وسیله آزمون آلفای کرونباخ محاسبه شد که برای پرسشنامه دانش بازاریابی ۰/۸۹ و برای پرسشنامه عملکرد بازاریابی ۰/۹۴ محاسبه گردید. برای تجزیه و تحلیل داده ها از آزمون های آمار توصیفی، آزمون کالموگروف اسمیرنف، آزمون ضریب همبستگی پیرسون و رگرسیون چندگانه استفاده گردید. نتایج نشان داد مدیریت دانش بازاریابی و مولفه های آن با عملکرد بازاریابی و مولفه های آن ارتباط مثبت و معنی دار دارند. با توجه به نتایج فدراسیون فوتبال جمهوری اسلامی ایران، باید با توسعه دانش بازاریابی در تمامی بخش های سازمان و به ویژه بخش بازاریابی، به طور مداوم بخش های مربوط به بازاریابی را مورد بازبینی قرار خواهد دهد و در تمامی فعالیت های بازاریابی از بهترین روش های موجود، با توجه به شرایط بازار ورزشی، استفاده نماید.

کلمات کلیدی: مدیریت دانش، عملکرد بازاریابی، فدراسیون فوتبال ایران



عنوان مقاله ۹: بررسی زنجیره ارزش و هوشمندی در صنعت اینترنت اشیا

نویسنده/نویسندگان: حمداله مهرآیین

چکیده:

هدف این مقاله ارزیابی زنجیره ارزش در صنعت اینترنت اشیا است. روش تحقیق این پژوهش از نظر شیوه گردآوری و تحلیل اطلاعات روشی آمیخته است. اینترنت اشیا بعد از اینترنت به عنوان جدیدترین فناوری اطلاعاتی دنیا شناخته می‌شود. طرح‌های پژوهشی و پروژه‌های کاربردی متعددی در خصوص اینترنت اشیا توسط دانشگاه‌ها و یا کنسرسیوم‌های مشترک صنعت و دانشگاه در سال‌های اخیر انجام شده است. با این حال، ابزارهایی که برای ایجاد ارتباط با اینترنت اشیا ساخته می‌شوند وسایل هوشمند تازه‌ای را برای منازل به ارمغان می‌آورند. اشیا و اپلیکیشن‌های لازم برای زندگی روزمره "هوشمند" اغلب به اتصال به خدمات ارتباطی مختلف، حسگرها و اجزای دیگر نیاز دارند. در عین حال، این اشیا باید بتوانند به سرعت در مقابل تغییرات و اقدامات کاربران واکنش نشان دهند. برای این کار باید هماهنگی سیستم به اندازه کافی کنترل شود. اینترنت اشیا به عنوان یک صنعت نوین که از دانش‌های پیشین و تجربه‌های صنایعی چون اینترنت، ارتباطات، فناوری اطلاعات، سخت‌افزار و نرم‌افزار استفاده کرده و مفهومی جدید را پایه‌گذاری نموده است، قابلیت آن را دارد که راه‌حلی برای مسائل دنیای امروز عرضه کند. امروزه دولت‌های جهان به دلیل فواید بسیاری که از آینده این صنعت سراغ دارند و آن را عاملی جهت توسعه شناسایی نموده‌اند، تمایلی منطقی به سمت استقرار این صنعت در کشور خود دارند. از آن جا که صنعت اینترنت اشیا در دوران کودکی خود قرار دارد، دارای مشکلاتی می‌باشد که یکی از این مشکلات عدم توسعه متعادل اجزای زنجیره می‌باشد. در واقع این مقاله، با بررسی و تحلیل مدل‌های زنجیره ارزش در این حوزه به ارائه زنجیره ارزشی که اینترنت اشیا ایجاد می‌کند پرداخته است و از طریق مرور سیستماتیک مقالات و روش فراترکیب به بررسی در این حوزه پرداخته است.

کلمات کلیدی: اینترنت اشیا، زنجیره ارزش، شبکه‌های توزیع شده صنعت، فناوری اطلاعات، فراترکیب



عنوان مقاله ۱۰: بررسی شاخص تحصیلات دانشگاهی در رفتار خریداران آنلاین با روش یادگیری عمیق

نویسنده/نویسندگان: میلاد فریدنیا و امیررضا استخریان حقیقی

چکیده:

تحصیلات دانشگاهی به عنوان یکی از عوامل تأثیرگذار بر رفتارهای مصرفی افراد، نقش مهمی در خرید آنلاین ایفا می‌کند. با توجه به رشد سریع تجارت الکترونیک و گسترش استفاده از فناوری‌های دیجیتال، نیاز به درک عمیق‌تری از ارتباط میان سطح تحصیلات و الگوهای خرید آنلاین وجود دارد. در این پژوهش، به بررسی تأثیر تحصیلات دانشگاهی بر میزان و نوع خرید آنلاین پرداخته شده است. برای تحلیل داده‌های مربوط به خرید آنلاین افراد با سطوح تحصیلات مختلف، از روش‌های یادگیری عمیق به ویژه شبکه‌های عصبی پیچشی (CNN) و شبکه‌های عصبی بازگشتی (RNN) استفاده شده است. این روش‌ها امکان شناسایی الگوهای پیچیده در رفتارهای خرید را فراهم کرده و به پیش‌بینی تأثیر تحصیلات بر تصمیم‌گیری خرید کمک می‌کنند.

داده‌های این پژوهش از فروشگاه‌های آنلاین جمع‌آوری شده و شامل اطلاعاتی نظیر میزان خرید، نوع کالاهای خریداری شده، روش‌های پرداخت، و سطح تحصیلات خریداران است. نتایج نشان می‌دهد که افراد با تحصیلات دانشگاهی بیشتر تمایل به خرید محصولات دیجیتال و تکنولوژیک، استفاده از روش‌های پرداخت الکترونیکی پیشرفته و انجام خریدهای با ارزش بالاتر دارند. همچنین، افرادی با تحصیلات پایین‌تر، بیشتر به خریدهای ساده‌تر و کالاهای مصرفی روزمره روی می‌آورند.

این تحقیق بیانگر آن است که تحصیلات دانشگاهی به عنوان عاملی کلیدی، بر اعتماد به خرید آنلاین، انتخاب کالاهای تخصصی و بهره‌برداری از فناوری‌های نوین در فرآیند خرید تأثیر بسزایی دارد. از سوی دیگر، استفاده از روش‌های یادگیری عمیق نشان داده که این مدل‌ها به خوبی قادر به پیش‌بینی الگوهای پیچیده خرید بر اساس سطح تحصیلات هستند. نتایج این مطالعه می‌تواند به فروشگاه‌های آنلاین و توسعه‌دهندگان فناوری‌های خرید اینترنتی کمک کند تا استراتژی‌های بازاریابی و طراحی سیستم‌های خود را با توجه به سطح تحصیلات کاربران بهبود بخشند.

کلمات کلیدی: خرید آنلاین، تحصیلات دانشگاهی، یادگیری عمیق، شبکه عصبی، تجارت الکترونیک



عنوان مقاله ۱۱: بررسی مدل‌های شبکه عصبی عمیق برای تحلیل داده‌های ساختار یافته و غیر ساختار یافته و کاربردهای موردی مختلف

نویسنده/نویسندگان: ابودر نظری نیا

چکیده:

شبکه‌های عصبی عمیق (DNNs) با پردازش پیشرفته داده‌های ساختار یافته و بدون ساختار، حوزه تجزیه و تحلیل داده‌ها را متحول کرده‌اند. این مقاله ادغام این انواع داده‌های ناهمگن را از طریق معماری‌های شبکه‌های عصبی عمیق چندوجهی بررسی می‌کند و مروری جامع از کاربردهای آنها در صنایع مختلف، از جمله مراقبت‌های بهداشتی، امور مالی و پردازش زبان طبیعی ارائه می‌دهد. شبکه‌های عصبی عمیق چندوجهی با ترکیب داده‌های ساختار یافته، مانند متغیرهای عددی و دسته‌ای، با داده‌های بدون ساختار مانند متن و تصاویر، دقت پیش‌بینی برتر و قابلیت‌های تصمیم‌گیری پیشرفته را نشان داده‌اند. این مطالعه همچنین پیشرفت‌های اخیر در معماری‌های شبکه‌های عصبی عمیق، از جمله استفاده از ترانسفورمرها و شبکه‌های عصبی گراف را برجسته می‌کند که کاربرد شبکه‌های عصبی عمیق را به انواع و وظایف پیچیده‌تر داده‌ها گسترش داده است. این مقاله به تفصیل به این چالش‌ها می‌پردازد و روش‌هایی مانند هرس ساختاریافته، یادگیری انتقالی و یادگیری خودنظارتی را برای آنها بررسی می‌کند. یافته‌ها پتانسیل عظیم شبکه‌های عصبی عمیق را در صنایع داده محور برجسته می‌کند و جهت‌گیری‌های تحقیقاتی آینده را برای افزایش بیشتر قابلیت‌های آنها روشن می‌سازد.

کلمات کلیدی: شبکه‌های عصبی عمیق، داده‌های ساختار یافته، داده‌های بدون ساختار، یادگیری چندوجهی، مدل‌های ترانسفورمر



عنوان مقاله ۱۲: بهبود پیش‌بینی قیمت مسکن با بازنمایی اطلاعات مکانی مبتنی بر قدم‌زنی تصادفی

نویسنده/نویسندگان: محمود امین طوسی

چکیده:

در این مقاله، روشی نوین برای بهبود پیش‌بینی قیمت مسکن با استفاده از قدم‌زنی تصادفی ارائه شده است. به صورت معمول تنها اطلاعات مکانی که در زمره ویژگی‌ها، در پیش‌بینی قیمت لحاظ می‌شود، طول و عرض جغرافیایی خانه‌ها هستند. در این پژوهش به منظور بهره‌گیری بیشتر از اطلاعات مکانی، از شیوهی قدم‌زنی تصادفی بهره برده شده است. به این منظور، ابتدا یک گراف غیر جهت‌دار با استفاده از داده‌های جغرافیایی منازل ایجاد شد. دو خانه در صورتی در این گراف مجاور در نظر گرفته شدند که فاصله‌ی آن‌ها کمتر از حدی مشخص باشد. سپس، با استفاده از روش قدم‌زنی تصادفی و بازنمایی رؤس گراف در الگوریتم DeepWalk، برای هر خانه، یک بردار معنایی ۳۰ بعدی استخراج گردید. این بردارها که اطلاعات مجاورت مکانی ضمنی خانه‌ها را در خود دارند، به همراه سایر ویژگی‌های خانه‌ها (مانند مساحت، تعداد اتاق و سن بنا) به عنوان ورودی، به مدل پیش‌بینی قیمت مسکن داده شدند. نتایج آزمایشات بر روی داده‌های قیمت مسکن در شهر مشهد، با یک مدل مبتنی بر تقویت گرادیان نشان داد که بردارهای معنایی حاصله، عملکرد مدل را بهبود می‌بخشند. به عنوان مثال، ضریب تعیین (R^2) از ۰.۸۲ در روش پایه به ۰.۸۵ در روش پیشنهادی افزایش یافت و خطای تقریب یک درصد کاهش پیدا کرد.

کلمات کلیدی: قدم‌زنی تصادفی، آنالیز داده‌ها، رگرسیون، گراف، یادگیری عمیق



عنوان مقاله ۱۳: بهبود طبقه‌بندی الگوهای شمعی با استفاده از شبکه‌های عصبی پیچشی بر اساس بلوک‌بندی ویژگی‌ها: مطالعه موردی بورس اوراق بهادار تهران

نویسنده/نویسندگان: پرستو فیض‌آبادی، سید محمدمهدی کاظمی و محمدعلی جعفری

چکیده:

پیش‌بینی روند بازارهای مالی همواره یکی از چالش‌های مهم در تحلیل تکنیکال بوده است. الگوهای شمعی به‌عنوان یکی از ابزارهای محبوب برای شناسایی تغییرات احتمالی در بازارهای مالی به کار گرفته می‌شوند، اما تشخیص دقیق این الگوها نیازمند دانش تخصصی و تجربه بالای تحلیل‌گران است. در این مقاله، از شبکه‌های عصبی پیچشی برای شناسایی و طبقه‌بندی دو الگوی شمعی مهم، یعنی چکش و مرد حلق‌آویز، در بورس اوراق بهادار تهران استفاده شده است. داده‌های قیمتی مربوط به چهار نماد شپدیس، خودرو، فولاد و فملی جمع‌آوری شده و پس از پیش‌پردازش به کمک تبدیل میدان زاویه‌ای گرمین به تصاویر تبدیل شده‌اند. از آنجا که تحلیل‌گران مالی تمایل دارند نمودارها و الگوها را به صورت یکپارچه مشاهده کنند، در این پژوهش برای نزدیک کردن نتایج به دیدگاه انسانی، تصاویر کوچک مربوط به سری‌های زمانی با تکنیک بلوک‌بندی ویژگی‌ها به یکدیگر چسبانده و به تصاویر بزرگ‌تر 20×20 تبدیل شدند. مدل شبکه‌ی عصبی پیچشی بر اساس داده‌های آموزش و آزمون، با دقت بالا توانسته است الگوهای شمعی را تشخیص دهد. نتایج نشان داد که افزایش ابعاد ورودی از $10 \times 10 \times 4$ به 20×20 در برخی از نمادها، منجر به بهبود عملکرد مدل شد. این مطالعه نشان می‌دهد که استفاده از تکنیک‌های تصویری و مدل یادشده می‌تواند در شناسایی الگوهای قیمتی و تصمیم‌گیری بهتر در بازارهای مالی مؤثر باشد.

کلمات کلیدی: الگوی شمعی، میدان زاویه‌ای گرمین، شبکه‌های عصبی پیچشی



عنوان مقاله ۱۴: بهبود کسب و کار الکترونیکی با استفاده از سیستم های هوشمند و رمزنگاری

نویسنده/نویسندگان: محمدرضا حسین زاده مقدم

چکیده:

با استفاده روزافزون اینترنت در زندگی روزمره و استفاده عموم مردم از اینترنت به جهت انجام کسب و کار و ذخیره و نگهداری داده ها در وب سایت ها برای جلوگیری از فاش شدن اطلاعات نیاز به افزایش امنیت و سیاستگذاری ها می باشد زیرا اینترنت جزو محیط های غیر متمرکز می باشد باید از روشهایی استفاده کنیم که بتوانیم حملاتی که در اینترنت در زمان انجام کسب و کار اتفاق می افتد را بگیرد به همین دلیل در این پژوهش از تکنیک های رمزنگاری و سیستم های هوشمند برای شناسایی و طبقه بندی داده ها استفاده می کنیم. در این مقاله با استفاده از خوشه بندی و رمزنگاری های متقارن باعث جلوگیری از حملاتی از جمله حملات منع سرویس و منع سرویس توزیع شده است می شود و این تکنیک پیشنهادی باعث امن شدن برنامه های کاربردی وب می شود.

کلمات کلیدی: داده کاوی، کسب و کار الکترونیکی، مدلسازی، قابلیت اطمینان، امنیت



عنوان مقاله ۱۵: پزشکی محاسباتی در تحول سیستم‌های بهداشتی و درمانی: از تحلیل داده‌های

زیستی تا تشخیص و درمان شخصی‌سازی

نویسنده/نویسندگان: وحیدرضا سبزواری و سهیلا اشکذری طوسی

چکیده:

پزشکی محاسباتی به‌عنوان حوزه‌ای میان‌رشته‌ای در تقاطع علوم پزشکی و علوم کامپیوتر، به توسعه و به‌کارگیری ابزارها و روش‌های محاسباتی جهت تحلیل دقیق‌تر و جامع‌تر داده‌های زیستی و پزشکی می‌پردازد. این حوزه شامل استفاده از مدل‌سازی‌های ریاضی و شبیه‌سازی‌های دقیق برای درک عمیق‌تر از فرآیندهای زیستی، مکانیسم‌های بیماری‌ها، و مسیرهای سلولی می‌باشد که امکان شناخت بهتر از بیماری‌ها را فراهم می‌سازد. همچنین، داده‌های حجیم و پیچیده‌ای مانند اطلاعات ژنتیکی، پروتئومیک، و متابولومیک از طریق تکنیک‌های پیشرفته‌ی محاسباتی و آماری پردازش و تحلیل می‌شوند تا الگوها و روابط پنهان در این داده‌ها آشکار شود. علاوه بر این، الگوریتم‌های یادگیری ماشین و هوش مصنوعی به‌طور گسترده در پزشکی محاسباتی به‌کار گرفته می‌شوند تا به پیش‌بینی دقیق‌تر و تشخیص سریع‌تر بیماری‌ها کمک کنند. این روش‌ها قادرند با تحلیل داده‌های متعدد و متنوع، الگوهای خاصی را شناسایی کنند که در داده‌های بالینی یا زیستی عادی ممکن است به‌راحتی قابل مشاهده نباشند. پزشکی محاسباتی با ارائه‌ی ابزارهای پیش‌بینی‌کننده و پشتیبان تصمیم، به متخصصان بهداشت و درمان کمک می‌کند تا پاسخ به درمان را پیش‌بینی کرده، روش‌های درمانی را بهینه‌سازی نموده و درمان‌های شخصی‌سازی شده را طراحی کنند که بر اساس ویژگی‌های فردی بیماران تنظیم شده است. این دستاوردها پتانسیل بالایی در بهبود دقت و کیفیت تشخیص‌ها، کاهش هزینه‌های درمان، و ارتقاء سطح سلامت عمومی جامعه دارند. با توسعه‌ی بیشتر این رویکردها و ادغام آن‌ها در فرآیندهای تشخیصی و درمانی، انتظار می‌رود پزشکی محاسباتی بتواند به تحولی عمیق در سیستم‌های بهداشتی و درمانی دست یابد و دسترسی به مراقبت‌های پزشکی با کیفیت بالا را افزایش دهد.

کلمات کلیدی: پزشکی محاسباتی، تحلیل داده‌های زیستی، مدل‌سازی ریاضی، یادگیری ماشین در

پزشکی، تشخیص و درمان شخصی‌سازی شده



عنوان مقاله ۱۶: پیش بینی خودکشی منجر به مرگ بیماران به کمک تحلیل داده های سیستم مراقبت های بهداشتی

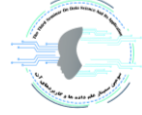
نویسنده/نویسندگان: نسرين خوش کيش، امير سيددانش و فرید فیضی

چکیده:

هدف از این مطالعه مقایسه عملکرد مدل های تکی یادگیری ماشین (شبکه عصبی بازگشتی و شبکه عصبی کانولوشن) با ترکیب این تکنیک های یادگیری ماشین (ترکیب شبکه عصبی بازگشتی و شبکه عصبی کانولوشن) روی داده های سیستم مراقبت های بهداشتی جهت پیش بینی مرگ در خودکشی بیماران است. مجموعه داده تحقیق مطابق با منبع است. برای اینکه ارائه دهندگان خدمات مراقبت های بهداشتی و ارائه دهندگان خط مشی مراقبت های بهداشتی اقداماتی را برای کاهش خطر خودکشی انجام دهند، باید بتوانند خطر خودکشی را کمیت سازی کنند. متأسفانه، کمی کردن خطر خودکشی دشوار است، زیرا خودکشی نادر است و عوامل خطر برای خودکشی رایج هستند. علاوه بر این، بسیاری از عوامل خطر برای خودکشی در طول زمان تغییر نمی کنند که تعیین کمیت حاد خطر خودکشی را دشوارتر می کند. داده های مراقبت های بهداشتی اداری به دلیل حجم و گستردگی داده ها به طور بالقوه برای توسعه مدل های پیش بینی ارزشمند هستند و به این دلیل که جمع آوری مداوم آن به این معنی است که مدل های پیش بینی توسعه یافته با این داده ها می توانند در آینده برای اطلاع رسانی خدمات و سیاست های مراقبت های بهداشتی مورد استفاده قرار گیرند. با در دسترس بودن سخت افزار و نرم افزار یادگیری ماشین برای افراد غیرمتخصص و با افزایش حجم داده های مراقبت های بهداشتی اداری، فرصتی وجود دارد برای ارزیابی اینکه آیا مدل های ترکیبی یادگیری ماشین می توانند بر روی مدل های تکی یادگیری ماشین برای تعیین کمیت خطر خودکشی بهبود یابند.

کلمات کلیدی: داده کاوی، داده های سیستم مراقبت های بهداشتی، فراگیری ماشین، خودکشی بیمار،

پیش بینی مرگ



عنوان مقاله ۱۷: پیش‌بینی ترافیک شهری با بهره‌گیری از شبکه‌های پیچشی گراف

نویسنده/نویسندگان: هدی مهرباقرپور، محمود امین طوسی و محمد آرشی

چکیده:

در پژوهش‌های اخیر، پیش‌بینی ترافیک به‌عنوان یک ابزار کلیدی در سیستم‌های حمل و نقل هوشمند مورد توجه قرار گرفته است. در این راستا، داده‌های جمع‌آوری شده از حسگرها در جاده‌ها به‌عنوان گره‌هایی در یک گراف تصور می‌شوند. تحقیقات اخیر با معرفی و به‌کارگیری مدل‌های پیشرفته بر روی مجموعه داده‌های واقعی، موفق به دستیابی به دقت بالاتری در پیش‌بینی ترافیک شده و قادر به شناسایی وابستگی‌های پیچیده‌ی زمانی و فضایی گردیده‌اند. شبکه‌های پیچشی گراف به‌عنوان روشی مؤثر در شناسایی همبستگی‌های بین گره‌های شبکه به‌کار رفته‌اند. با این حال، بسیاری از مدل‌های موجود به ماتریس‌های مجاورت ثابت یا تطبیقی متکی هستند که ممکن است نتوانند روابط واقعی بین سیگنال‌ها را به‌درستی منعکس کنند. به همین دلیل، روش‌های مبتنی بر گراف‌های پویا به کار گرفته می‌شوند که با در نظر گرفتن تغییرات زمانی، روابط بین گره‌ها را به‌طور مؤثرتری بازتاب می‌کنند. تولید گراف‌های پویا با ترکیب شبکه‌های پیچشی و مدل‌های مبتنی بر شبکه‌های عصبی بازگشتی، امکان ایجاد گراف‌هایی که با تغییرات زمانی سیگنال‌های ترافیکی سازگار هستند را فراهم می‌آورد. این رویکرد نه تنها امکان تحلیل دقیق‌تری برای پیش‌بینی فراهم می‌کند، بلکه تغییرات الگوهای ترافیکی در زمان‌های مختلف (روزانه، هفتگی و اخیراً) را نیز به‌خوبی شبیه‌سازی می‌نماید. پژوهش‌های انجام‌شده نشان‌دهنده‌ی اهمیت رویکردهای نوین در پیش‌بینی ترافیک هستند و می‌توانند به بهبود کیفیت مدیریت ترافیک شهری و کاهش مشکلات ناشی از آن کمک کنند.

کلمات کلیدی: پیش‌بینی ترافیک، سیستم‌های حمل و نقل هوشمند، گراف پویا، شبکه پیچشی گراف



عنوان مقاله ۱۸: پیش‌بینی نرخ کلیک در صنعت تبلیغات آنلاین با داده‌های واقعی و چالش‌های آن

نویسنده/نویسندگان: محمدرضا فاتحی نیا، سید مسیح سجادی و محمود امین طوسی

چکیده:

پیش‌بینی نرخ کلیک (CTR)، یکی از چالش‌های اساسی و کلیدی در تبلیغات آنلاین محسوب می‌شود. با توجه به گسترش بازار تبلیغات آنلاین در پلتفرم‌های مختلف از جمله موتورهای جستجو، وبسایت‌های شخص ثالث، شبکه‌های اجتماعی و اپلیکیشن‌های موبایل، رقابت ارائه دهندگان تبلیغات آنلاین به شدت افزایش یافته است. با توجه به این که مدل درآمدی اکثر ارائه دهندگان تبلیغات آنلاین بر اساس تعداد کلیک‌های انجام شده بر روی تبلیغ است، پیش‌بینی دقیق نرخ کلیک به ازای هر تبلیغ و نمایش تبلیغات مبتنی بر آن برای بهبود اثربخشی تبلیغات بسیار مورد توجه قرار گرفته است. ما در این پژوهش سعی کردیم با تلفیق روش‌های پیشین در این حوزه، الگوریتمی ارائه دهیم که علاوه بر عملکرد مناسب، نیاز کمتری به منابع سخت‌افزاری داشته باشد و بتواند بر روی سیستم‌های با توان پردازشی متوسط بر روی کلان داده نیز به خوبی اجرا شود. در این پژوهش، داده‌ها از یک شرکت تبلیغات آنلاین واقعی دریافت شده‌اند و ساختگی نیستند. با توجه به حجم بالای داده‌ها (حدود ۵۰ گیگابایت)، با چالش کلان داده مواجه هستیم. همچنین، چالش‌های مختلفی از جمله نرخ پایین داده‌های کلاس مثبت و مسئله شروع سرد (Cold Start) را مورد بررسی قرار دادیم و راهکارهایی برای بهبود این مسائل ارائه کردیم.

کلمات کلیدی: پیش‌بینی نرخ کلیک، تبلیغات آنلاین، رتبه بندی، سیستم‌های پیشنهادگر، کلان داده



عنوان مقاله ۱۹: تاثیر رشد عمیق سازی سرمایه و آهنگ رشد تولید بر بهره وری کل عوامل (TFP) در چرخه اقتصادی

نویسنده/نویسندگان: سنا سیلانی، سیاوش احمدی چهره برق، مزده افشار کرمانی و توفیق الهویرنلو

چکیده:

این مقاله به بررسی تاثیر رشد عمیق سازی سرمایه و آهنگ رشد تولید بر بهره وری کل عوامل در چرخه اقتصادی می پردازد. با استفاده از داده های جمع آوری شده از سایت های معتبر و تحلیل های آماری پیشرفته، تاثیرات مثبت و منفی عمیق سازی سرمایه و رشد تولید بر بهره وری کل عوامل در دوره های مختلف اقتصادی (رونق و رکود) تحلیل می شود. برای این منظور داده های پیش فرض در جداول تحلیلی استفاده شده و مدل های ریاضی و اقتصادی به کار گرفته شده و داده ها با استفاده از " نرم افزار متلب " تحلیل می شوند. نتایج نشان می دهند که سرمایه گذاری های به موقع و رشد تولید پایدار می توانند به طور قابل توجهی بهره وری کل عوامل را بهبود بخشند و به رشد اقتصادی پایدار کمک کنند. این مطالعات به سیاست گذاران اقتصادی و مدیران صنایع راهکارهایی برای ارتقای بهره وری در دوره های مختلف اقتصادی ارائه می دهد.

کلمات کلیدی: چرخه رونق و رکود، آهنگ رشد تولید، عمیق سازی سرمایه، بهره وری کل عوامل



عنوان مقاله ۲۰: تحلیل احساسات متون مروری با استفاده از مدل BERT

نویسنده/نویسندگان: شهرام جمالی، مهناز آزماینده و بهنام پرستار

چکیده:

علم داده (Data Science) حوزه‌ای بین رشته‌ای است که با استفاده از آمار، یادگیری ماشین و سایر روش‌ها به استخراج معانی و الگوهای موجود در بطن داده‌ها می‌پردازد. یکی از نمودهای مهم علم داده، تحلیل احساسات یا Sentiment Analysis است که احساسات موجود در متون مختلف را بررسی و استخراج می‌نماید. تحلیل احساسات یکی از روش‌های موثر برای سنجش نظرات مشتریان و مخاطبان در حوزه‌های مختلف می‌باشد. این تحلیل برای درک تجربه مشتری و شناسایی نگرش افراد نسبت به یک محصول یا خدمت مورد استفاده قرار می‌گیرد. این مقاله تحلیل احساسات متون مروری را با استفاده از مدل زبانی BERT انجام می‌دهد و یک معماری جدید با هدف بهبود دقت تحلیل، پیشنهاد شده است. برای ارزیابی روش پیشنهادی، چندین آزمایش با الگوهای چند دامنه‌ای و دامنه خاص و با مجموعه داده‌های واقعی انجام شده است. نتایج ارزیابی نشان می‌دهد که مدل پیشنهادی در مقایسه با سایر مدل‌ها مانند LSTM، BIGRU، BIGRU-Attention عمل بهتری داشته و امتیاز F1-score آن برابر ۰.۸۷ می‌باشد.

کلمات کلیدی: علم داده، تحلیل احساسات، پردازش زبان طبیعی



عنوان مقاله ۲۱: تحلیل و دسته بندی داده های متنی حاصل از شبکه های اجتماعی با استفاده از رگرسیون لاسو

نویسنده/نویسندگان: ابوالفضل آدرسی، امیرحسین امیری و امیر آبیاری

چکیده:

دسته بندی پیام‌های عادی و ناهنجار یکی از موضوعات حیاتی در شبکه‌های اجتماعی است. با افزایش حجم و پیچیدگی داده‌های تولید شده در پلتفرم‌های مختلف، روش‌های آماری مخصوصاً روش‌های مبتنی بر یادگیری ماشین به عنوان ابزاری مؤثر برای شناسایی و دسته بندی پیام های عادی و ناهنجار به کار گرفته شده‌اند. این روش‌ها توانایی تحلیل و دسته بندی حجم عظیمی از داده‌ها را دارند و می‌توانند به طور هوشمند دسته بندی پیام های عادی و ناهنجار را انجام دهند. داده‌های متنی شبکه‌های اجتماعی به دلیل پیچیدگی زبانی و تنوع شیوه‌های بیان کاربران، چالش‌های بسیاری را در تشخیص و دسته بندی پیام های عادی و ناهنجار ایجاد می‌کنند. در این مقاله، مدل رگرسیون لاسو برای دسته بندی پیام های عادی و ناهنجار در داده‌های متنی به کار گرفته شده است. عملکرد مدل با استفاده از معیارهایی مانند صحت، بازخوانی و دقت ارزیابی شده است تا اطمینان حاصل شود که مدل پیشنهادی توانایی شناسایی و دسته بندی پیام های عادی و ناهنجار را با دقت و کارایی بالا دارد. صحت مدل پیشنهادی در شناسایی و دسته بندی پیام‌های ناهنجار و عادی به ترتیب 86.50% و 72.09% بوده که نشان دهنده این است که این مدل به‌طور مؤثری قادر به شناسایی و دسته بندی پیام‌های عادی و ناهنجار می‌باشد.

کلمات کلیدی: شبکه‌های اجتماعی، داده‌های متنی، دسته‌بندی ناهنجاری، رگرسیون لاسو، یادگیری ماشین



عنوان مقاله ۲۲: تشخیص سندروم بیش‌فعالی یا اختلال ADHD براساس الگوریتم خوشه‌بندی k-means تحت نظریه آمارشکل

نویسنده/نویسندگان: زهرا دلشاد و عباس رسولی

چکیده:

در دنیای امروز، تحلیل داده‌های بزرگ و پیچیده نقش کلیدی در شناسایی الگوها و روندهای پنهان موجود در الگوهای علمی دارد. با رشد سریع حجم داده‌ها در حوزه‌های مختلف، استفاده از روش‌های نوین یادگیری ماشین و هوش مصنوعی به ابزاری حیاتی برای پردازش و تحلیل داده‌ها تبدیل شده است. در این پژوهش، بر تحلیل آماری شکل تمرکز داریم، حوزه‌ای که به تمایز اشیاء هندسی از دیدگاه آماری و بدون تأثیر ترجمه، مقیاس‌بندی و چرخش می‌پردازد. بررسی کاربرد روش خوشه‌بندی K-means در تشخیص اختلال ADHD یا سندروم بیش‌فعالی، با استفاده از تصاویر MRI مغز افراد انجام گردیده است. داده‌ها شامل اشیاء هندسی است که از نقاط شاخص روی خطوط پیرامونی آنها استخراج شده سپس برای ارزیابی نتایج از شاخص‌های معتبر نظریه شکل آماری و خوشه‌بندی استفاده گردیده است.

کلمات کلیدی: شکل، نقاط شاخص، خوشه‌بندی، ADHD, k-means



عنوان مقاله ۲۳: تفکیک الکتروکاردیوگرافی از نوک انگشت به عنوان یک شناسه بیومتریک توسط

شبکه عصبی مصنوعی

نویسنده/نویسندگان: امیرحسین خیرخواه، علیرضا گائینی، امیرحسین مقام دوست و بابک رضایی

افشار

چکیده:

الکتروکاردیوگرافی می‌تواند از نوک انگشتان نیز انجام شود و با توجه به اینکه سیگنال‌های دریافتی از الکتروکاردیوگرافی هر فرد با افراد دیگر متفاوت بوده می‌توان از این تفاوت‌ها در صنعت و حوزه‌های احراز هویت و بیومتریک بهره‌حداکثری برد. در این مطالعه ۶ فرد، با سن 1 ± 22 سال با شاخص توده بدنی در محدوده 17 ± 24 کیلوگرم بر متر مربع به صورت داوطلبانه و آگاهانه شرکت کردند و از آنها سیگنال‌های الکتروکاردیوگرافی از بخش قفسه‌ی سینه و نوک انگشتان اخذ گردید و ویژگی‌های استخراج شده در یک طبقه‌بند شبکه عصبی مصنوعی مورد آموزش، آزمایش و بررسی قرار گرفت. نتایج نشان داده شده توسط شبکه عصبی مصنوعی از صحت تفکیک، برای الکتروکاردیوگرافی از انگشت برابر با ۷۳.۶۴ درصد برخوردار بوده و همچنین برای الکتروکاردیوگرافی از قفسه‌ی سینه با مقدار صحت ۸۲.۷۳ درصد تعیین گردید. نتایج طبقه‌بندی با شبکه عصبی نشان داد علیرغم افت محسوس صحت در زمان اخذ داده از انگشت امکان بهره‌برداری از این ابزار در هر دو حالت به عنوان یک شناسه بیومتریک فراهم است لذا پیشنهاد می‌شود با توجه به داده‌های اخذ شده و نتایج بدست آمده از این مطالعه، سیستم احراز هویت توسط الکتروکاردیوگرافی در صنعت و مراکز امنیت محور مورد توجه و استفاده قرار گیرد.

کلمات کلیدی: الکتروکاردیوگرافی، ECG، بیومتریک، احراز هویت، شبکه عصبی مصنوعی



عنوان مقاله ۲۴: چالش‌های فنی استفاده از علم داده در اقتصاد

نویسنده/نویسندگان: محمدعلی مقصودپور

چکیده:

پیشرفت‌های شگرفی که در دهه گذشته در محاسبات و فناوری اطلاعات رخ داده است، حجم وسیعی از داده‌ها را در حوزه‌های مختلف در دسترس قرار داده که از آن به عنوان داده‌های بزرگ یاد می‌شود. به‌ویژه در اقتصاد و امور مالی، بهره‌برداری از این داده‌ها، تحقیقات و کسب‌وکار را به هم نزدیک‌تر می‌کند، زیرا داده‌های تولید شده در فعالیت‌های اقتصادی معمولی، می‌توانند برای مدل‌های مؤثر و شخصی‌سازی شده استفاده شوند. در این زمینه، استفاده اخیر از فناوری‌های علم داده برای اقتصاد و امور مالی، منافع متقابلی را برای دانشمندان و متخصصان فراهم می‌کند و پیش‌بینی را بهبود می‌بخشد. این مقاله چالش‌های فنی اساسی استفاده از داده‌ها مانند مدیریت و حفاظت داده‌ها، مدل‌سازی، یکپارچه‌سازی و تفسیر داده‌ها را معرفی می‌کند. همچنین برخی از مسائل رایج در مدل‌سازی اقتصادی با فناوری‌های علم داده را تشریح می‌کند و راه‌حل‌های مدیریت کلان داده و راه‌حل‌های تحلیلی مربوطه را بررسی می‌کند و انگیزه استفاده از روش‌های علم داده در اقتصاد و امور مالی را ایجاد می‌کند.

کلمات کلیدی: اقتصاد مالی، داده‌های اقتصادی، کمیت و کیفیت داده‌ها



عنوان مقاله ۲۵: حذف نویز بهینه در تصاویر رنگی با ترکیب تحلیل بردارهای ویژه و نمونه‌گیری بوت استرپ وزنی

نویسنده/نویسندگان: رضا قاسمی نجف آبادی و علیرضا صفریان

چکیده:

حذف نویز از تصاویر رنگی همواره چالش‌برانگیز است، زیرا نیازمند تفکیک دقیق نویز از جزئیات اصلی تصویر و هموارسازی مناسب آن است. علاوه بر این، این فرآیند باید همبستگی بین کانال‌های رنگی را نیز در نظر گیرد. در این پژوهش، روشی نوین برای کاهش نویز در تصاویر رنگی ارائه شده که از تحلیل بردارهای ویژه ماتریس داده‌ای ناشی از مقادیر رنگی همسایگی هر پیکسل، همراه با تخصیص وزن‌های نمایی، برای بهبود حذف نویز بهره می‌گیرد. این تحلیل سه هدف اصلی را دنبال می‌کند: (۱) مدیریت همبستگی میان کانال‌های رنگی، (۲) تشخیص نواحی صاف در مقابل لبه‌ها و بافت‌ها، و (۳) تعیین سطح مناسب برای هموارسازی. نتایج آزمایش‌ها نشان می‌دهد که این رویکرد، به‌ویژه در ترکیب با بوت‌استرپ وزنی، نسبت به روش‌های سنتی و جدید در حذف نویز عملکرد رقابتی و بهبود چشمگیری ارائه می‌دهد.

کلمات کلیدی: حذف نویز، بردار ویژه، میانگین وزنی، بوت استرپ



عنوان مقاله ۲۶: خلاصه‌سازی انتزاعی متن با استفاده از مدل‌های زبانی از پیش آموزش دیده

نویسنده/نویسندگان: حدیثه نجفی قره‌جلو و مصطفی قاضی‌زاده احسائی

چکیده:

امروزه خلاصه‌سازی متن با توجه به افزایش حجم مطالب در رسانه‌های گوناگون، به یکی از مهم‌ترین فعالیت‌ها در حوزه پردازش متن تبدیل شده است. رویکردهای خلاصه‌سازی متنوعی وجود دارد اما خلاصه‌سازی انتزاعی نسبت به استخراجی موثرتر و خروجی آن به خلاصه‌های انسانی نزدیک‌تر است. مدل‌های زبانی از پیش آموزش‌دیده، در حوزه‌های مختلفی مورد استفاده قرار می‌گیرند. با ظهور الگوریتم‌های یادگیری تقویتی، قدرت ترنسفورمرهای از پیش آموزش‌دیده به همراه این الگوریتم‌ها افزایش می‌یابد. این پژوهش، استفاده از مدل زبانی از پیش آموزش‌دیده مبتنی بر معماری ترنسفورمر همراه با اعمال الگوریتم بهینه‌سازی اولویت مستقیم یا DPO را برای تولید خلاصه‌های بهتر پیشنهاد می‌کند. این پژوهش، استفاده از دو مدل زبانی از پیش آموزش‌دیده مبتنی بر معماری ترنسفورمر mT5 و ARMAN و الگوریتم یادگیری تقویتی بهینه‌سازی اولویت مستقیم را پیاده‌سازی کرده و سپس نتایج را بررسی و تحلیل می‌کند. این رویکرد در مدل زبانی ARMAN با کسب امتیاز ۵۳/۳۶ در Rouge-1، ۵۰/۰۳ در Rouge-2 و ۵۲/۶۵ در Rouge-L نویدبخش پیشرفت روش‌های خلاصه‌سازی در زبان فارسی است.

کلمات کلیدی: پردازش زبان طبیعی، یادگیری عمیق، مدل‌های زبانی از پیش آموزش‌دیده، خلاصه‌سازی انتزاعی متن، یادگیری تقویتی



عنوان مقاله ۲۷: خوشه‌بندی داده‌ها با استفاده از یک رویکرد نیمه‌پارامتری مبتنی بر چگالی

نویسنده/نویسندگان: سید مهدی صالحی، آندریت بکر و محمد آرشی

چکیده:

خوشه‌بندی چگالی-محور برخی مزایا نسبت به روش‌های خوشه‌بندی مبتنی بر فاصله و مدل دارد. در سال‌های اخیر، برخی محققان این رویکرد را با پیشنهاد یک روش مبتنی بر نظریه‌ی گراف برای شناسایی مدهای موضعی تابع چگالی داده‌ها توسعه داده‌اند. آن‌ها از روش ناپارامتری برآورد چگالی هسته (KDE) با هسته‌های نرمال و t برای این منظور استفاده کردند. در پژوهش حاضر قصد داریم الگوریتم ناپارامتری خوشه‌بندی چگالی-محور بالا را به یک الگوریتم نیمه‌پارامتری گسترش دهیم. به‌طور دقیق‌تر، ما در این مقاله برآوردگرهای نیمه‌پارامتری KDE با خانواده‌ای از هسته‌های انعطاف‌پذیرتر از نرمال از جمله چوله-نرمال و چوله- t پیشنهاد می‌کنیم. شواهد نشان می‌دهند که برآوردگرهای پیشنهادی نسبت به برآوردگرهای معمولی، که از هسته‌های نرمال و t استفاده می‌کنند، نه تنها میزان اریبی کمتری در نقاط مرزی تکیه‌گاه چگالی داده‌ها دارند بلکه به چگالی واقعی جامعه نیز نزدیک‌تر هستند. پیدا کردن پهنای باند بهینه تحت حالت‌های یک‌بعدی و چندبعدی نتیجه اصلی دیگری از این مقاله است. در نهایت، از طریق یک مطالعه عددی جامع، عملکرد الگوریتم خوشه‌بندی چگالی-محور نیمه‌پارامتری پیشنهاد شده را با استفاده از برخی مجموعه داده‌های شبیه‌سازی شده و واقعی ارزیابی خواهیم داد.

کلمات کلیدی: برآورد چگالی هسته، خوشه‌بندی چگالی-محور، شاخص ارزیابی سیلوئت چگالی-محور، هسته‌های نامتقارن



عنوان مقاله ۲۸: روش‌های تجزیه و تحلیل کلان داده‌ها در اقتصاد و امور مالی

نویسنده/نویسندگان: محمدعلی مقصودپور

چکیده:

چکیده نظارت بر وضعیت فعلی و آینده اقتصاد برای دولت‌ها، سازمان‌های بین‌المللی و بانک‌های مرکزی در سراسر جهان از اهمیت اساسی برخوردار است. سیاستگذاران برای طراحی سیاست‌های مؤثری که می‌توانند رشد اقتصادی و حفظ و ارتقای سطح رفاه اجتماعی را تقویت کنند، به اطلاعات کلان اقتصادی در دسترس نیاز دارند. با این حال، شاخص‌های کلیدی اقتصادی، که در طول فرآیند تصمیم‌گیری بر آن‌ها تکیه می‌کنند، با فرکانس پایین تولید می‌شوند و با تأخیر قابل توجهی منتشر می‌شوند، که در نتیجه آن، با چنین مجموعه ناقصی از اطلاعات، اقتصاددانان تنها می‌توانند به طور تقریبی شرایط اقتصادی واقعی، آینده و حتی اخیر را بسنجند، و این امر، پیش‌بینی اقتصادی را به وظیفه‌ای بسیار چالش‌برانگیز تبدیل می‌کند. پیشرفت‌های شگرفی که در دهه گذشته در محاسبات و فناوری اطلاعات رخ داده است، حجم وسیعی از داده‌ها را در حوزه‌های مختلف در دسترس قرار داده که از آن به عنوان داده‌های بزرگ یاد می‌شود. به‌ویژه در اقتصاد و امور مالی، بهره‌برداری از این داده‌ها، تحقیقات و کسب‌وکار را به هم نزدیک‌تر می‌کند، زیرا داده‌های تولید شده در فعالیتهای اقتصادی معمولی، می‌توانند برای مدل‌های مؤثر و شخصی‌سازی شده استفاده شوند. این مقاله دو روش مهم و پرکاربرد برای تجزیه و تحلیل و استفاده از کلان داده‌ها در اقتصاد و امور مالی را به صورت اجمالی بررسی کرده است. نتایج نشان می‌دهد که استفاده از روش‌هایی مانند الگوریتم‌های یادگیری ماشین و فن‌آوری‌های وب‌معنایی، علاوه بر اینکه می‌توانند داده‌های جدید را استخراج کنند، قادرند به اقتصاددانان در مدل‌سازی سیستم‌های پویا پیچیده و به هم پیوسته کمک کنند.

کلمات کلیدی: کلان داده‌ها، یادگیری ماشین، داده‌های اقتصادی و مالی



عنوان مقاله ۲۹: شبکه‌های عصبی گراف در پیش‌گویی خواص مولکولی

نویسنده/نویسندگان: امیر جلوگیر باغستان، محمود امین طوسی و محمد آرشی

چکیده:

پیش‌گویی خواص مولکولی نقش محوری در زمینه‌های مختلفی از جمله کشف دارو و طراحی مواد دارد که به هدف پیش‌گویی طیف وسیعی از خواص مولکول‌ها، مانند حلالیت، سمی بودن و فعالیت بیولوژیکی انجام می‌شود.

حوزه کشف دارو و پیش‌گویی خواص مولکولی به دلیل ظهور تکنیک‌های یادگیری ماشین پیشرفته تحول چشمگیری داشته است. از آنجایی که ساختارهای مولکولی به طور ذاتی پیچیده هستند و شامل ترتیبات متنوعی از اتم‌ها و پیوندها هستند، این پیچیدگی چالش‌هایی را برای تکنیک‌های یادگیری ماشین سنتی در پیش‌گویی دقیق خواص مولکولی ایجاد می‌کند.

روش‌های سنتی مبتنی بر اثرانگشت‌ها و توصیف‌گرهای مولکولی، هرچند اطلاعاتی ارزشمند ارائه می‌دهند، اما غالباً در به‌تصویر کشیدن روابط ساختاری پیچیده درون مولکول‌ها دچار مشکل می‌شوند. پیشرفت‌های اخیر در شبکه‌های عصبی گراف توانایی آن‌ها در رفع این محدودیت‌ها را از طریق نمایش مؤثر مولکول‌ها به عنوان گراف‌ها نشان داده است.

در این پژوهش ما به معرفی رویکردهای مبتنی بر شبکه‌های عصبی گراف برای پیش‌گویی خواص مولکولی می‌پردازیم.

یافته‌های ما کارآمدی این رویکردها را در مواجهه با چالش‌های پیش‌گویی خواص مولکولی نشان می‌دهد و پتانسیل آن برای کاربردهای گسترده‌تر در شیمی انفورماتیک را برجسته می‌کند.

کلمات کلیدی: پیش‌گویی خواص مولکولی، شبکه‌های عصبی گراف، طراحی محاسباتی دارو، یادگیری ماشین مولکولی



عنوان مقاله ۳۰: شناسایی و حذف حفره‌های انرژی شبکه‌های بی‌سیم با استفاده از الگوریتم‌های ابتکاری

نویسنده/نویسندگان: احمد غفورنیا و حسین بلوچیان

چکیده:

شبکه‌های حسگر بی‌سیم شامل صدها تا هزاران حسگر کوچک می‌شوند که اطلاعات از محیط پیرامون مانند دما، رطوبت، نور، ارتعاش و صدا را جمع‌آوری و پیش‌پردازش می‌کنند و سپس داده‌های مفید به گره‌های دیگر یا ایستگاه پایه ارسال می‌شود. این شبکه‌ها از برنامه‌های متعددی مانند سیستم هشداردهنده‌ها، مراقبت از سلامتی، نظارت بر محیط‌زیست و امنیت بهره‌مندند، اما مصرف انرژی اصلی‌ترین چالش آنها است. در این تحقیق، تلاش برای شناسایی و رفع حفره‌های انرژی در شبکه‌های حسگر بی‌سیم با استفاده از الگوریتم‌های ابتکاری جهت بهینه‌سازی مصرف انرژی صورت می‌گیرد. این پژوهش به دنبال پاسخ به این سوال است که آیا روش پیشنهادی تحقیق می‌تواند منجر به بهینه‌سازی مصرف انرژی شبکه شود؟ در تحقیق حاضر، از روش ترکیبی CHD-CR-ND و الگوریتم رقابت استعماری برای شناسایی و ترمیم حفره‌های پوششی در شبکه‌های حسگر بی‌سیم استفاده می‌شود. این روش‌ها بهبود مصرف انرژی شبکه را ارتقا می‌بخشند و بر طول عمر پوشش شبکه تأثیر مثبتی دارند. الگوریتم CHD-CR-ND مسئول تشخیص و ترمیم حفره‌ها است و الگوریتم رقابت استعماری پارامترهای آن را بهینه‌سازی می‌کند. این روش‌ها با هدف افزایش بهره‌وری و کارایی شبکه حسگر بی‌سیم، به صورت گام به گام و از مرحله تجزیه و تحلیل تا مراحل پیاده‌سازی و اجرا در تحقیق مورد بررسی قرار می‌گیرند. مکانیسم پیشنهادی برای حذف حفره‌های انرژی نیز پیاده‌سازی شده است که منجر به بهبود مصرف انرژی و افزایش طول عمر شبکه شده است. نتایج تحقیق نشان می‌دهد که گره‌های حسگر به توازن در مصرف انرژی دست پیدا کرده و این باعث افزایش طول عمر شبکه می‌شود.

کلمات کلیدی: شناسایی، حذف حفره‌های انرژی، شبکه‌های بی‌سیم، الگوریتم‌های ابتکاری، رقابت استعماری



عنوان مقاله ۳۱: کاربرد نمونه‌گیری مجموعه-رتبه‌دار در تحلیل داده‌های بزرگ

نویسنده/نویسندگان: محمد حسین زرین کلاه، هادی جباری نوقابی و محمد مهدی صابر

چکیده:

یکی از اساسی‌ترین چالش‌ها در مدل‌سازی و تحلیل داده، برآورد دقیق پارامترهای مدل است. در تحلیل داده‌های بزرگ، به دلیل حجم بالا و پیچیدگی، پردازش کل مجموعه داده‌ها از نظر محاسباتی و زمانی غیرعملی یا بسیار پرهزینه است. در چنین شرایطی، استفاده از نمونه‌های معرف به‌عنوان جایگزینی برای کل داده‌ها، علاوه بر کاهش هزینه‌های محاسباتی و افزایش سرعت آموزش مدل‌ها، به مدیریت محدودیت‌های حافظه و منابع نیز کمک می‌کند. در این مقاله، از روش نمونه‌گیری مجموعه-رتبه‌دار میانه‌ای (Median Ranked-Set Sampling) برای انتخاب نمونه‌های نماینده و برآورد پارامترهای مدل استفاده شده است. برای برآورد پارامترهای مدل در داده‌های درآمدی مرکز آمار ایران، روش‌های نمونه‌گیری تصادفی ساده (Simple Random Sampling) و نمونه‌گیری مجموعه-رتبه‌دار میانه‌ای (MRSS) به کار گرفته شد. نتایج نشان داد که در مواردی که داده‌ها نامتقارن هستند، استفاده از MRSS منجر به برآوردهایی با دقت بیشتر نسبت به SRS می‌شود.

کلمات کلیدی: نمونه‌گیری مجموعه-رتبه‌دار، داده‌های بزرگ، بررسی‌های نمونه‌ای



عنوان مقاله ۳۲: کاربرد یادگیری ماشین و داده کاوی در صنعت تولید

نویسنده/نویسندگان: زهره محمدی

چکیده:

یکی از اهداف حیاتی صنعت تولید، تولید محصولات با کیفیت بالا با حداقل هزینه است. از سال ۲۰۱۶ که جهان وارد چهارمین انقلاب صنعتی با عنوان "صنعت ۴" شده است، محیط‌های تولید هوشمند شده و در نتیجه، حجم وسیعی از داده‌ها از همه قسمت‌های درگیر از جمله: طراحی محصول و فرآیند، مونتاژ، برنامه‌ریزی، کنترل کیفیت، نگهداری و غیره جمع‌آوری می‌شوند. در این راستا، فناوری‌های جدید مانند یادگیری ماشین و داده کاوی نقش کلیدی در کسب دانش از داده‌ها و پردازش آنها ایفا می‌کنند. با استفاده از آنها می‌توان هزینه‌های عملیاتی و زمان خرابی را کاهش داد و در عین حال بهره‌وری و کارایی را افزایش داد. همچنین، قابلیت تکنیک‌های یادگیری ماشین در ارائه بینش‌های پیش‌بینی‌کننده، امکان شناسایی الگوهای پیچیده تولید را فراهم کرده و راهی برای یک سیستم پشتیبانی تصمیم‌گیری هوشمند در انواع وظایف ارائه می‌دهد. اگرچه تکنیک‌های مختلف یادگیری ماشین در برنامه‌های تولیدی متنوعی استفاده شده‌اند، اما سؤالات و چالش‌های بسیاری باقی مانده است. بنابراین در این مطالعه، نگاهی دقیق‌تر به نقش یادگیری ماشین و داده کاوی در صنعت تولید خواهیم داشت و یک مرور جامع از آخرین دستاوردهای محققان در یادگیری ماشین و کاربردهای آن در صنعت تولید را ارائه خواهیم کرد.

کلمات کلیدی: یادگیری ماشین، داده کاوی، صنعت ۴، تولید



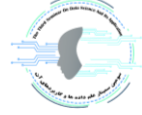
عنوان مقاله ۳۳: مدل بندی زمان حوادث ترافیکی با استفاده از تحلیل بقای عمیق

نویسنده/نویسندگان: فاطمه محمدی، محمد آرشی و آرزو حبیبی راد

چکیده:

حوادث ترافیکی یکی از چالش‌های اصلی در مدیریت ترافیک شهری هستند که میتوانند تأثیرات منفی بر ایمنی و کارایی سیستم‌های حمل و نقل داشته باشند. اقدامات نادرست آژانس‌های مدیریت ترافیک در برخورد با این حوادث میتواند منجر به تاخیر در عبور و مرور شود. آژانس‌های حمل و نقل برای مدیریت حوادث ترافیکی و به کارگیری استراتژی‌های مناسب نیاز به پیش بینی مدت زمان حادثه دارند. یکی از اهداف مدل بندی داده‌های حوادث ترافیکی، پیش بینی مدت زمان حادثه است. داده‌های زمان حوادث ترافیکی داده‌های زمان تا رویداد هستند. یکی از ویژگی‌های این داده‌ها وابستگی به مدت زمان است. یکی از مدل‌های رایج برای در نظر گرفتن این ویژگی، مدل‌های نرخ خطر مبنا است. مدل‌های نرخ خطر مبنا شامل مدل‌های زمان شکست شتابیده و مدل‌های خطرات متناسب پارامتری و نیمه پارامتری کاکس می‌باشد. عملکرد مدل‌های نرخ خطر مبنا در پیش بینی زمان حوادث ترافیکی، با توجه به فرضیاتی که دارند ممکن است دارای محدودیت‌هایی باشد. از جمله این فرضیات میتوان به فرض خطرات متناسب، در نظر گرفتن توزیع خاص برای داده‌ها، اثر ثابت با زمان عوامل تأثیرگذار و رابطه خطی بین متغیرهای مستقل و وابسته، اشاره کرد و ممکن است برخی از این فرضیات در داده‌های حوادث ترافیکی وجود نداشته باشد. یکی دیگر از روش‌های رایج برای مدل بندی زمان حوادث ترافیکی مدل‌های یادگیری ماشین می‌باشد. نقطه قوت این مدل‌ها این است که برخی از فرضیات مشکل ساز مدل‌های نرخ خطر مبنا را ندارند. یکی از روش‌های یادگیری ماشین، شبکه‌های عصبی عمیق می‌باشد. اگر هدف این مدل‌ها پیش بینی مدت زمان حادثه باشد، آنها نمیتوانند وابستگی به مدت زمان را در داده‌های حوادث ترافیکی در نظر بگیرند. لذا از مدل‌های شبکه‌های عصبی عمیق با هدف برآورد تابع بقا استفاده میکنیم. اما این مدل‌ها هم نمی‌توانند به صراحت مدت زمان یک حادثه ترافیکی را پیش بینی کنند. بنابراین در این مطالعه علاقه مندیم که مدل یادگیری عمیق مبتنی بر تحلیل بقا طراحی کنیم که هم تابع بقا را برآورد کند و هم زمان حوادث ترافیکی را پیش بینی کند.

کلمات کلیدی: تحلیل بقای عمیق، مدل‌های نرخ خطر مبنا، تابع بقا، یادگیری چند وظیفه‌ای



عنوان مقاله ۳۴: مدیریت امنیت داده‌های بزرگ در مراقبت‌های بهداشتی هوشمند

نویسنده/نویسندگان: مسعود فضلعلی پور و سیمین منصوری بروجنی

چکیده:

تحلیل داده‌های بزرگ در بخش پزشکی می‌تواند به متخصصان پزشکی در تسهیل بهبود مراقبت‌های بهداشتی کمک کند. با کمک تحلیل داده‌ها، می‌توان از تصاویر بالینی بیماران برای تشخیص برخی بیماری‌ها استفاده کرد. همچنین برای مدیریت یکپارچه مراقبت‌های بهداشتی، نیاز به امنیت داده‌های پزشکی امری ضروری می‌باشد. در این مقاله، یک چارچوب امنیتی مبتنی بر معادله لجستیک، پیشنهاد و سپس، یک روش به اشتراک گذاری تصویر محرمانه محاسباتی بدون از دست دادن اطلاعات (CSIS) برای تبدیل تصویر محرمانه رمزگذاری شده به اشتراک گذاری برای ذخیره‌سازی توزیع شده در سرورهای مبتنی بر داده‌های دیجیتالی استفاده می‌شود. علاوه بر این، اعداد شبه تصادفی (PRN) تولید شده توسط معادله لجستیک با توالی تصویر در دو مرحله با تغییر جزئی پارامترها XOR می‌شوند. سهم‌های تولید شده از نظر اندازه کوچک هستند و به منابع کمتری مانند ظرفیت ذخیره‌سازی و پهنای باند انتقال نیاز دارند که در سیستم‌های مبتنی بر اینترنت اشیا بسیار مطلوب است. در ادامه نشان می‌دهیم سیستم رمزنگاری در برابر حملات و هم چنین تداخل‌ها بسیار ایمن بوده و حساسیت کلیدی بسیار بالایی دارد.

کلمات کلیدی: داده‌های بزرگ، مراقبت‌های بهداشتی از راه دور، امنیت داده‌ها، رمزگذاری، اشتراک گذاری تصویر محرمانه



عنوان مقاله ۳۵: مروری بر کاربردهای شبکه عصبی گرافی در حوزه سلامت و پزشکی

نویسنده/نویسندگان: بنیامین عندالیب الشهدا، سهیلا اشکذری طوسی

چکیده:

پیشرفت‌های اخیر در هوش مصنوعی و یادگیری ماشین تحولات گسترده‌ای در پزشکی ایجاد کرده‌اند. یکی از تکنیک‌های پیشرو که به سرعت جایگاه خود را در این عرصه تثبیت کرده، شبکه‌های عصبی گرافی (Graph Neural Networks یا GNNs) است. این مدل‌ها که برای تحلیل ساختارهای پیچیده و غیرخطی طراحی شده‌اند، توانایی منحصر به فردی در مدل‌سازی داده‌های غیرساختار یافته و روابط پیچیده موجود در گراف‌ها دارند؛ موضوعی که مدل‌های یادگیری ماشین سنتی معمولاً در آن ناتوان‌اند. تحلیل داده‌های پزشکی به دلیل تنوع و پیچیدگی زیاد، از جمله داده‌های ژنتیکی، شبکه‌های پروتئینی، تعاملات سلولی و تصاویر پزشکی (مانند MRI و CT)، آن‌ها را به یکی از چالش‌های اصلی در علم داده تبدیل کرده است و تحلیل دقیق آن‌ها نیاز به ابزارهایی دارد که بتوانند این پیچیدگی‌ها را به خوبی مدل‌سازی کنند. از کاربردهای مهم GNNها در پزشکی، پیش‌بینی دقیق‌تر بیماری‌ها و تشخیص اختلالات ژنتیکی است. همچنین در زمینه‌هایی چون تحلیل تصاویر پزشکی، شبیه‌سازی‌های بیولوژیکی پیچیده و بهینه‌سازی درمان‌های شخصی‌سازی شده، از این تکنیک‌ها به خوبی استفاده می‌شود. علاوه بر این، GNNها با تکنیک‌های دیگر مانند پردازش زبان طبیعی (NLP) و یادگیری عمیق ترکیب می‌شوند تا تحلیل پرونده‌های الکترونیکی سلامت و دیگر داده‌های پزشکی را به طور چشمگیری بهبود دهند. با توجه به رشد سریع داده‌های پزشکی و نیاز به تحلیل سریع، دقیق و هوشمند این داده‌ها، شبکه‌های عصبی گرافی به عنوان ابزاری قدرتمند برای ارتقاء دقت تشخیص‌ها و بهبود درمان‌ها مطرح شده‌اند. هدف این مقاله بررسی پتانسیل‌ها و چالش‌های GNNها در حوزه پزشکی است.

کلمات کلیدی: شبکه عصبی گرافی، یادگیری ماشین در پزشکی، تحلیل داده‌های پیچیده، پیش‌بینی

بیماری، پرونده‌های الکترونیکی سلامت



عنوان مقاله ۳۶: مروری بر یافته‌های نوروفیزیولوژیک مبنی بر وجود اصول نظریه تشدید انطباقی (ART)

در فرآیندهای ادراکی-شناختی-هیجانی مغز

نویسنده/نویسندگان: هدی جلال کمالی

چکیده:

این مقاله به بررسی اصول نظریه تشدید انطباقی (ART) در فرآیندهای ادراکی-شناختی-هیجانی مغز می‌پردازد. یافته‌های نوروفیزیولوژیک نشان می‌دهند که چگونه یادگیری ART می‌تواند در انواع سیکل‌های ادراکی و شناختی که ویژگی بسیاری از رفتارهای پستانداران است، نقش داشته باشد. این مقاله به سوالاتی درباره چگونگی تحقق یادگیری ART در قشر مغز، تنظیم پارامتر مراقبت و تأثیر آن بر اختلال‌های عصبی مانند اوتیسم و فراموشی می‌پردازد. همچنین، فرآیند انطباق الگو، که جزئی اساسی از سیستم‌های ART است، و نقش آن در ادراکات بینایی، شناسایی اشیاء و کنترل حسی-حرکتی مورد بررسی قرار می‌گیرد. در نهایت، مقاله به تحلیل حلقه‌های بازخوردی و تأثیر آنها بر تجربه آگاهانه و توهمات شیزوفرنیایی می‌پردازد.

کلمات کلیدی: نظریه تشدید انطباقی (ART)، فرآیندهای ادراکی-شناختی-هیجانی مغز، یافته‌های

نوروفیزیولوژیک



عنوان مقاله ۳۷: نقش محوری علم داده در احراز هویت دیجیتال شهروندان در سازمان‌های وابسته به شهرداری

نویسنده/نویسندگان: فاطمه السادات عالم

چکیده:

مقاله حاضر با هدف بررسی نقش حیاتی علم داده در ارتقای فرآیند احراز هویت دیجیتال در سازمان‌های شهرداری به نگارش درآمده است. در این پژوهش، ضمن تبیین مفهوم احراز هویت دیجیتال و انواع روش‌های آن، مقایسه‌ای بین رویکردهای سنتی و دیجیتال در این حوزه صورت گرفته است. همچنین، کاربردهای متنوع علم داده در فرآیند احراز هویت و هشت مرحله کلیدی بهره‌گیری از آن به تفصیل مورد تحلیل قرار گرفته‌اند. علاوه بر این، چالش‌ها و راهکارهای موجود در این زمینه به همراه مطالعه موردی ارائه شده است. نتایج این پژوهش نشان می‌دهد که با استفاده از علم داده می‌توان به طور چشمگیری دقت و سرعت فرآیند احراز هویت را افزایش داد، ضمن آنکه سطح امنیت و اطمینان از صحت اطلاعات نیز بهبود می‌یابد. در نهایت، این مقاله به اهمیت پیاده‌سازی سیستم‌های احراز هویت دیجیتال مبتنی بر علم داده در سازمان‌های شهرداری به منظور افزایش بهره‌وری و ارائه خدمات بهتر به شهروندان تاکید دارد.

کلمات کلیدی: علم داده، احراز هویت دیجیتال، سازمان‌های شهرداری، رویکرد سنتی و دیجیتال



عنوان مقاله ۳۸: نقش هوش مصنوعی در پردازش داده‌های اینترنت اشیا

نویسنده/نویسندگان: سیامک نیرومند

چکیده:

صنعت جهانی اینترنت روز به روز بالغ تر می شود و پتانسیل توسعه و سرزندگی بالایی را نشان می دهد و به تدریج بخش مهمی از ارتقای توسعه اقتصادی و پالش ها تبدیل می شود. هوش مصنوعی بهترین راه حل برای مدیریت جریان های عظیم داده است. در آینده با کمک فناوری های جدید هوش مصنوعی، فناوری داده های بزرگ و افزایش استفاده از اینترنت تمرکز اتصال مردم به اینترنت اشیا تغییر خواهد کرد. پیشرفت ها و همگرایی در فناوری های حسگر، پردازش اطلاعات و ارتباطات، اینترنت اشیا امروزی را شکل داده است. افزایش سریع نیازهای داده و خدمات، چالش های جدیدی را برای اینترنت اشیا به ارمغان می آورد. فناوری های نوظهور و تکنیک های هوشمند می توانند نقش قانع کننده ای در توسعه معماری ها و سرویس های هوشمند در اینترنت اشیا داشته باشد.

کلمات کلیدی: هوش مصنوعی، پردازش داده ها، اینترنت اشیا



عنوان مقاله ۳۹: یادگیری تابع درست‌نمایی پواسن با استفاده از شبکه های عصبی عمیق

نویسنده/نویسندگان: ترنم ترابی نعمان، مهدی عمادی، محمد آرشی

چکیده:

برای تحلیل و مدلسازی هر داده ای ابزارهای متنوعی وجود دارد از جمله مدل‌های رگرسیونی و مدل های بیزی و درخت های تصمیم و یکی از به روزترین و دقیق ترین ابزار تحلیل داده ها استفاده از شبکه عصبی است. ما می توانیم شبکه عصبی را با توزیع های گسسته ای همچون پواسن ترکیب کنیم زیرا شبکه‌های عصبی به دلیل ساختار چندلایه‌ای خود قادر به یادگیری روابط غیرخطی میان ورودی‌ها و خروجی‌ها هستند. آنها می‌توانند حجم بالای داده‌ها را پردازش کرده و اطلاعات معنادار را استخراج کنند. شبکه‌های عصبی قادر به کار با انواع داده‌ها هستند. همچنین به طور مؤثری در وظایف پیش‌بینی و طبقه‌بندی داده‌ها مورد استفاده قرار می‌گیرند. به طور دقیق تر، در این طرح منفی تابع درست‌نمایی را به عنوان تابع زیان در شبکه عصبی پرسپترون چند لایه در نظر گرفته و در مرحله پس انتشار با استفاده از روش مشتقات نزولی، وزن های بهینه که همان پارامترهای تابع درست‌نمایی هستند را برآورد می کنیم.

کلمات کلیدی: شبکه عصبی عمیق، پرسپترون چند لایه، تابع زیان، پس انتشار، توزیع های گسسته



چکیده مقالات انگلیسی

(به ترتیب حروف الفبا)

40th Article Title : A New Geometric -Based Clustering Algorithm for Medical Data Analysis

Author(s): Mohammad Mahdi Karimi & Hajar Ghahremani-Gol

Abstract: Many medical data sets are mathematically structured as graphs. For example, AIDS dataset consists of graph characterizing chemical compounds. In this paper, we address the clustering of a collection of graphs, where each data point is represented as a graph. Our proposed method employs the heat kernel algorithm as a geometric approach. We present an embedding technique based on the heat kernel to transform each graph into a feature vector. Subsequently, we cluster the data set of graphs using these feature vectors. Experimental results demonstrate that the proposed method achieves 98.85% accuracy in clustering the AIDS dataset.

Keywords: Clustering, Heat Kernel, Graph Embedding



41th Article Title : Adaptive Cluster Sampling in Large-Scale Data: Applications in Dynamic Social Network

Author(s): Amin Ferdosi Makan, Abdolhamid Rezaei Roknabadi & Mehdi Jabbari Nooghabi

Abstract: Increasing data complexity and volume demand efficient, accurate sampling strategies for large-scale, dynamic populations, such as social networks. This paper examines the application of adaptive cluster sampling (ACS) in these situations, emphasizing its effectiveness in capturing hidden and rare clustered populations with greater efficiency and smaller sample sizes than conventional methods. By incorporating adaptive link-tracing techniques, ACS improves access to hard-to-reach, dynamic populations. With a design tailored for sampling units selection in evolving populations, this approach shows potential to enhance data reliability and its applications in data science.

Keywords: Rare and Clustered Populations. Adaptive Cluster Sampling. Big Data. Dynamic Populations. Social Networks



42th Article Title : Advanced Data Science Techniques for Uncovering Hidden Patterns: Detecting Long-Memory Processes

Author(s): Leila Marvian & Hossein Hassani

Abstract: Long-memory models are essential in data science and finance for identifying long-term dependencies in time series. Accurate detection of long-memory processes is crucial for reliable predictions. This paper highlights a key limitation in using the sample autocorrelation function (ACF) for this purpose. Although ACF theoretically defines long memory, summing sample ACF values fails to indicate it accurately. Hassani's " $-1/2$ theorem" shows that the sum of sample ACFs is always $-1/2$ for any stationary time series, regardless of its length. This discovery challenges diagnostic methods relying on this sum. Recognizing this limitation is vital for accurate model development and forecasting.

Keywords: Data Pattern Recognition, Hassan's $-1/2$ Theorem in Data Science, Frequency Domain Analysis, Long-Term Dependencies Detection



43th Article Title : Advancing Automated Diagnosis of Knee Injuries with Deep Learning and MRI Data through Data-Driven Medical Image Analysis

Author(s): Mina Iman Shayan, Fatemeh Salehi & Toktam Khatibi

Abstract: The integration of data science in medical imaging has opened new avenues for improving diagnostic accuracy. In this study, we propose a deep learning model for the automated diagnosis of knee injuries using MRI data. Our model processes MRI scans and generates diagnostic reports for injuries to the Posterolateral Corner (PLC), Lateral Collateral Ligament (LCL), Posterior Cruciate Ligament (PCL), and knee cartilage. Preliminary results show promising accuracy in detecting these injuries, suggesting that this approach can enhance diagnostic efficiency and reduce human error in interpreting complex MRI data. This method could support clinicians by offering faster, more reliable insights into knee injury assessments.

Keywords: Medical Image Analysis, Deep Learning, MRI, Knee Injuries, Automated Diagnosis



44th Article Title : Analysis of the high-dimensional data sets using robust penalized least-squares versus support vector machines

Author(s): Mahdi Roozbeh

Abstract: In data science techniques, machine learning is a cutting-edge class of powerful technologies that can tackle many significant issues that they may encounter in the current world. Among these methods, support vector machine (SVM) is one that is well-known for classification. One method for creating a regression model that is a fantastic member of the machine learning family is support vector regression (SVR). High-dimensional data sets are now the most difficult issue to solve. The estimate of the coefficients and interpretation are the biggest challenges when dealing with high-dimensional data. Because there are a lot of predictor variables in high-dimension problems, classical techniques like the ordinary least-squares approach, which is the best approach if the key assumptions (such as normality and independency to the error terms as well as little to no multicollinearity in the explanatory variables), cannot be applied. SVR is one of the best methods to model the high-dimensional data sets. It is a really reliable and robust approach to have a good fit with high accuracy. In this research, SVR is used in a real high-dimension dataset about the gene expression in the Riboflavin Vitamin B2 production data and then compare it with well-known methods: least absolute shrinkage and selection operator and sparse least trimmed squared methods.

Keywords: High-dimensional data. Multicollinearity. Outlier data. Penalized least trimmed squares



45th Article Title : Decision Tree Analysis for Predicting Recovery in Patients with Sudden Sensorineural Hearing Loss

Author(s): Razieh Yousefi, Azadeh Saki & Mohsen Rajati

Abstract: There is growing interest in predicting treatment outcomes, aiding in risk classification and identifying significant factors in patient management. Sudden Sensorineural Hearing Loss (SSNHL) is a condition that requires immediate medical attention, and timely recognition and treatment can enhance patient hearing. This study employed decision tree modeling to forecast recovery outcomes in SSNHL patients treated at Ghaem Hospital in Mashhad. Data from 330 patients were analyzed, revealing an 84.5% recovery rate. The CART algorithm was utilized, identifying 11 variables, with significant predictors including age, time to treatment initiation, and gender. The model demonstrated a mean residual error of 0.44 and a classification error rate of 9%. Findings suggest that decision trees can effectively predict treatment outcomes and important factors for SSNHL.

Keywords: Decision Tree Modeling, Treatment Outcomes, CART Algorithm



46th Article Title : Designing Automated Diagnostic Processes for Knee Injuries in Healthcare Organizations Using Deep Learning and MRI Images

Author(s): Mina Iman Shayan, Fatemeh Salehi & Toktam Khatibi

Abstract: Recent technological advancements have significantly transformed the diagnostic processes in healthcare organizations. The creation of automated diagnostic systems for knee injuries using deep learning and MRI imaging represents a substantial improvement in healthcare service quality. These systems allow for swift and accurate analysis of MRI images, facilitating precise identification and classification of knee injuries. Given the vast amount of medical data managed by healthcare institutions, such automation enhances operational efficiency and reduces response times. However, the successful implementation of these systems requires effective integration with existing Hospital Information Systems (HIS) and Picture Archiving and Communication Systems (PACS). This integration should be designed to optimize workflows and ensure the secure and efficient management of patient data.

Keywords: Deep Learning, Knee MRI Imaging, Automated Diagnostic Processes, Healthcare Workflow Optimization



47th Article Title : Early Detection of Diabetic Retinopathy and Macular Conditions Using Deep Learning Models and SMOTE

Author(s): Alireza Safariyan & Reza Ghasemi

Abstract: Diabetic retinopathy and macular-related conditions are critical complications of diabetes, potentially leading to severe vision loss or blindness. Early detection and diagnosis of these conditions are vital to prevent their progression. In this study, we employed deep learning models, specifically DenseNet-121, ResNet101, and GoogleNet (InceptionV3), along with the SMOTE (synthetic minority over-sampling technique) technique to address class imbalance. The models were trained on fundus images to classify both diabetic retinopathy and macular conditions. Our results show that the GoogleNet model, enhanced with SMOTE, achieved an overall accuracy of 75% in diagnosing diabetic retinopathy, with a precision of 0.76 and an F1-score of 0.74. The DenseNet-121 model, when combined with SMOTE, performed the best, reaching an accuracy of 85%, precision of 0.85, and an F1-score of 0.85 for diabetic retinopathy. For macular classification, the DenseNet-121 model demonstrated an accuracy of 92%, with a macro average precision, recall, and F1-score of 0.92 across three macular status classes. These findings highlight the effectiveness of deep learning models in automating the diagnosis of diabetic retinopathy and macular disorders, offering a valuable tool for early detection and clinical decision-making and ensure the secure and efficient management of patient data.

Keywords: Diabetic Retinopathy Detection. Macular Condition Classification. Deep Learning Models. SMOTE. Fundus Imaging Analysis



48th Article Title : Empowering Data Science with StatAI Basic: AI Solutions for Statistical Analysis

Author(s): Emad Ashtari Nezhad

Abstract: Statistical software is crucial in data science, offering tools for data analysis, visualization, modeling, and hypothesis testing. Software such as R, Python, and SPSS facilitate the processing and analysis of large datasets, the construction of predictive models, and the visualization of insights, making data-driven decision-making more efficient and accurate. However, using these tools often requires statistical expertise, especially for software like R and Python, which also demand knowledge of statistical programming. Thus, developing tools that are user-friendly for both statisticians and non-statisticians is critical. With advancements in artificial intelligence, this is becoming more achievable.

In this context, we are designing a series of AI-based statistical software called StatAI. This article introduces the first in the Basic series, which allows users to conduct analyses on fundamental methods such as regression, correlation, and clustering by interacting with the software through chat in any language, including Persian.

Keywords: StatAI Basic. AI-based analysis. Statistical software. Artificial intelligence (AI)



49th Article Title : Enhancing Survival Analysis of Recurrent Events with Transformers: A Financial Prediction Approach

Author(s): Zahra Esfandiar & Vahid Ghanbarizadeh

Abstract: Recent deep learning models like WTT-RNN, Deep Recurrent Survival Analysis (DRSA) and CmpXRnnSurv-AE have advanced survival analysis but struggle with long-range dependencies and time-varying relationships in financial data, especially for irregular events like credit defaults. To overcome this, we propose a Transformer-based model for survival analysis of recurrent events, focusing on variable gap times between events. Using self-attention mechanisms, the model captures complex patterns in event sequences without predefined temporal assumptions. Time-encoding techniques account for dynamic gap times, and the model is trained with Cox partial likelihood and a custom survival loss to handle censored data. Applied to the Credit Card Payment Default Prediction dataset, our model outperforms WTT-RNN, Deep Recurrent Survival Analysis (DRSA) and CmpXRnnSurv-AE in predictive accuracy, validated by the Concordance Index (C-index), making it a promising solution for survival analysis in financial settings.

Keywords: Recurrent Event Data. Transformer Neural Network. Gap Times



50th Article Title : Enhancing Text Extraction from Scanned Medical Documents Using Large Language Models

Author(s): Mehdi Nemati & Mahmood Amintoosi

Abstract: Accurate text extraction from scanned medical documents is essential for data management and clinical decision-making. This study evaluates Large Language Models (LLMs) as an enhancement to traditional Optical Character Recognition (OCR) methods. By leveraging language and context, LLMs offer improved accuracy and relevance in text interpretation. We compared the EasyOCR model and the multimodal "gpt-4o-mini" LLM on a dataset of 110 medical transcript samples. Performance was assessed by comparing extracted texts against clinical data embeddings, using cosine similarity for semantic accuracy. The OCR model achieved an F1-score of 0.59, while the LLM scored 0.70, demonstrating LLMs' potential to advance text extraction in healthcare.

Keywords: Optical Character Recognition. Large Language Models. Text Extraction. Medical Documents



51th Article Title : Estimation of the effective reproduction number for Pulmonary Tuberculosis in Iran

Author(s): Vahid Fakoore, Maryam Rastegar & Mohammad Taghi Shakeri

Abstract: Tuberculosis (TB), caused by the bacterium "Mycobacterium tuberculosis", remains one of the most significant global health threats, especially in developing countries. The reproduction number represents the average number of secondary infections produced by a single infected individual in a susceptible population. The estimation of the reproduction number for TB is vital for public health authorities to design effective control and prevention strategies. In this talk, we focus on the application of a Bayesian and non-Bayesian approach to estimate the reproduction number for pulmonary tuberculosis in Iran.

Keywords: Bayesian approach, Reproduction number, Tuberculosis



52th Article Title : Evaluating the Impact of Nomination Sampling on Logistic Regression Performance Metrics in imbalanced Data Sets

Author(s): Razieh Yousefi, Hassan Doosti & Habibollah Esmaily

Abstract: Sampling methods are vital for enhancing predictive modeling, particularly in scenarios of class imbalance. This study investigates the performance of Nomination Sampling (NS) compared to Simple Random Sampling (SRS) for classifying imbalanced datasets using logistic regression. We assessed these methods on populations with binary responses featuring imbalanced ratio of 20 and 30% utilizing 10 predictors across various conditions: population sizes of 1000, 5000, and 10000; sample sizes of 100, 200, and 300; set sizes of 3, 5, and 10; and correlation coefficients of 0.5 and 0.7. Results showed NS consistently outperformed SRS in precision, recall, F1-score, and Pr-AUC, while SRS had higher specificity in all scenarios. Notably, the advantages of NS increased with larger sample sizes. These findings underscore the effectiveness of NS in addressing the challenges posed by imbalanced data.

Keywords: Nomination Sampling(NS). Simple Random Sampling(SRS). Logistic Regression. Imbalanced Datasets. Classification



53th Article Title : EXPLORING EMERGENCY MEDICAL RESPONSE IN STROKE CASES: A DIRECTIONAL STATISTICAL APPROACH

Author(s): Nooshin Akbari sharak, Mohammad Taghi Shakeri & Seyed Morteza Najibi

Abstract: Background and Aim: Circular data, characterized by its cyclical nature, arises when variables are measured on a circular scale, such as time, direction, or phase. Unlike linear data, circular data repeats at specific intervals, such as 24-hour cycles or seasonal periods, necessitating specialized statistical methods to accurately capture patterns and relationships. In healthcare, understanding cyclic patterns is crucial for identifying times of increased demand, particularly in emergency services. Materials and Methods: We collected data on patients exhibiting stroke symptoms who were transferred to Ghaem Hospital in Mashhad by the Emergency Medical Services (EMS) between April 2018 and March 2019. To analyze the arrival and admission times, descriptive directional statistics were performed using the circular package in R. Results: The average arrival time for stroke patients was approximately 2:00 PM. Analysis revealed substantial variability in arrival times, as indicated by high circular variance and standard deviation. Additionally, the low mean resultant length suggests that arrival times are not tightly clustered, indicating considerable diversity in the times at which stroke patients present to the emergency center. The average admission time for patients to the unit was around 5:00 PM. Similar variability was observed in admission times, with high circular variance and standard deviation, and a low mean resultant length suggesting that admission times are widely dispersed rather than closely aligned with the mean. Conclusion: There is an approximate three-hour delay between patient arrival and admission, emphasizing the importance of understanding the time intervals involved in patient care and the potential implications for resource allocation in emergency services.

Keywords: Circular data. directional statistics. Emergency services. Stroke patients



54th Article Title : Fraud Detection in Automobile Insurance based on a Deep Hybrid Approach

Author(s): Seyed MohammadAli Fazilat & Shahla Najafloo

Abstract: Insurance fraud causes massive financial damages to vehicle insurance companies every year. Despite the many advances in insurance fraud detection, the costs created for automobile insurance companies due to such frauds are still increasing. Therefore, the main goal of this work is to present and apply different methods with unbalanced dataset in the field of predicting the occurrence of fraud in automobile insurance. In the proposed approach, we propose a new hybrid model based on deep neural network and machine learning using a data-driven approach. A proposed approach is presented for the first time to predict car insurance fraud with the goal of greater accuracy. In addition, the feature selection technique with XGboost algorithm, feature extraction with multi-layer perceptron neural network and classification with SVM algorithm are used in the field of predicting the occurrence of fraud in automobile insurance. The obtained results show the effectiveness of our proposed hybrid method in accuracy and precision of 96.97% and 97.05%. The numerical experiments demonstrate that the proposed approach achieves promising results for detecting fake accident claims.

Keywords: Machine learning. Credit card fraud detection. Artificial Neural Networks. Random Forests. XGBoost



55th Article Title: Mixture of factor analysis model and outlier detection based on skew contaminated-normal distribution

Author(s): Farzane Hashemi

Abstract: Factor analysis (FA), as an unsupervised data mining technique, allows us to reduce technique in a variety of scientific fields, especially in psychology and educational testing. This article proposes a mixture multivariate skew contaminated-normal (SCN) factor analysis modeling approach for model-based clustering of high-dimensional data. The use of the SCN distribution can enrich the accuracy of inference since the model can detect outliers automatically and produce robust parameter estimates in each cluster in skewed data. The maximum likelihood (ML) estimation of model parameters are carried out by developing an expectation-maximization algorithm. We also derive the score vector and the empirical information matrix for calculating standard errors associated with the ML parameter estimates. The potential and applicability of our methodology are illustrated through analyzing a real dataset study.

Keywords: strongly leptokurtic, heavy-tail, factor analysis, Skew contaminated-normal distribution, EM-type algorithm



56th Article Title: Mixtures of common factor analyzers based mean-mixture of normal distribution

Author(s): Saeed Darijani & Farzane Hashemi

Abstract: This paper presents a novel extension of mixtures of common factor analyzers based on the multivariate mean-mixture of normal distribution for model based clustering of high-dimensional data with missing and asymmetrically heavy-tailed observations. Under the missing at random mechanism, we develop a computationally analytical expectation conditional maximization (ECM) algorithm to compute the maximum likelihood estimates of model parameters. Prediction of factor scores and imputations of missing values are automatically retrieved during the estimating procedure. The optimal numbers of factors and mixture components are identified via penalized likelihood criteria. Extensive simulation studies and two real data analyses illustrate the efficacy and performance of the proposed methodology.

Keywords: Mean-mixture of normal distribution. EM-type algorithm. Factor analysis. Heavy-tail. Missing data



57th Article Title: Optimal control of wastewater treatment based on the optimization algorithm in an urban wastewater treatment plant in order to effectively save electrical energy

Author(s): Mohammadreza Mohammadiyan Asiabar

Abstract: The wastewater treatment system is analyzed based on the particle swarm optimization algorithm to improve energy efficiency. From the experimental results of the wastewater treatment system, it plays an important role in realizing effective wastewater treatment and improving the economic and social benefits of wastewater treatment. Then the particle swarm optimization algorithm can play an effective role in the wastewater treatment system. Therefore, actively promoting the application of this algorithm can improve the energy efficiency of the wastewater treatment system. Finally, the solution of the wastewater treatment system is analyzed based on the particle swarm optimization algorithm to improve the energy efficiency. First, this article considers the particle swarm optimization algorithm and the activated sludge wastewater treatment method as the research objective, and with the help of the experimental analysis method, the analysis of the particle swarm optimization algorithm to improve the energy efficiency of the wastewater treatment system has achieved good results. . From the experimental results of the wastewater treatment system, it plays an important role in realizing effective wastewater treatment and improving the economic and social benefits of wastewater treatment. Then the particle swarm optimization algorithm can play an effective role in the wastewater treatment system. Therefore, actively promoting the application of this algorithm can improve the energy efficiency of the wastewater treatment system. The experimental results show that the solution designed in this paper can be applied to other wastewater power plants by using the activated sludge oxidation ditch process, and it has a special popularity value, which effectively saves electrical energy.

Keywords: Particle swarm optimization, Energy, Efficiency



58th Article Title: PLSe2: An Efficient Estimator for Partial Least Square

Author(s): Jamshid Jamali, Farideh Khosravi & Habibollah Esmaily

Abstract: Partial Least Square (PLS) and consistent PLS(PLSc) estimation methods as variance-based approaches in Structural Equations Modeling (SEM) have disadvantages. PLSe2 method, an optimal Generalized Least Squares (GLS) methodology using PLSc implied covariance matrix, allows formal testing of models via chi-square statistics and evaluation of parameter estimates by deriving standard error estimates, which are previously not directly available. The performance of this estimator under normal and abnormal conditions has been checked by Monte Carlo simulation. The results indicate that the proposed estimator provides estimates that are almost as good as the theoretically optimal ML estimator under normality. Also, it is demonstrated that the standard error estimates closely correspond to the empirical Monte Carlo variation. In particular, it is used from Satorra-Bentler's (1994) scaled chi-square statistic. In short, PLSe2 has the advantages of both ML and PLS, thereby suggesting it as the methodology of choice for model specification, estimation, and evaluation in social sciences empirical and health studies.

Keywords: Structural Equation Modeling (SEM). Partial Least Square (PLS). Consistent Partial Least Square (PLSc). Efficient Partial Least Square (PLSe2)



59th Article Title: Polynomial regression or polynomial meshless function approximation?

Author(s): Somayeh Jafari & Alireza Soheili

Abstract: THIS PAPER EXPLORES THE RELATIONSHIP BETWEEN MACHINE LEARNING ALGORITHMS AND POLYNOMIAL REGRESSION, ILLUSTRATING HOW THE STATISTICAL CONCEPT OF REGRESSION ALIGNS WITH MATHEMATICAL FUNCTION APPROXIMATION TECHNIQUES. SPECIFICALLY, WE INTERPRET POLYNOMIAL REGRESSION AS A TYPE OF MESHLESS FUNCTION APPROXIMATION. MESHLESS METHODS ADDRESS DIMENSIONALITY CHALLENGES; HOWEVER, BY APPLYING A DECENT GRADIENT TO THESE TECHNIQUES, WE ACHIEVE RESULTS ANALOGOUS TO THOSE FOUND IN POLYNOMIAL REGRESSION WITHIN DEEP .LEARNING FRAMEWORKS. THE DEFINITION OF THE SUPPORT DOMAIN IN MESHLESS METHODS SIGNIFICANTLY INFLUENCES NUMERICAL OUTCOMES.

IN OUR STUDY, WE GENERATED A STOCHASTIC DATASET CONSISTING OF 1100 SAMPLES, WHERE 700 WERE DESIGNATED AS TRAINING DATA AND 400 AS VALIDATION DATA. THE NUMERICAL FINDINGS REVEAL THAT WHILE THE APPLICATION OF A DECENT GRADIENT INCREASES TRAINING ERRORS, IT MARKEDLY REDUCES VALIDATION ERRORS. FOR INSTANCE, WHEN IMPLEMENTING A DECENT GRADIENT METHOD ON A QUADRATIC POLYNOMIAL MESHLESS FUNCTION APPROXIMATION, WE RECORDED VALIDATION AND TRAINING ERRORS OF 0.2 AND 0.266, RESPECTIVELY. IN CONTRAST, USING THE INVERSE MOMENT MATRIX .YIELDED ERRORS OF APPROXIMATELY $1.216E-13$ AND 51.896. SIMILAR TRENDS WERE OBSERVED FOR CUBIC POLYNOMIAL MESHLESS FUNCTION APPROXIMATION.

Keywords: Polynomial regression, Polynomial meshless function approximation



60th Article Title: Restricted Bayesian Lasso Regression: Addressing Linear Inequality Constraints in High-Dimensional Models

Author(s): Solmaz Seifollahi & Mohammad Arashi

Abstract: In this article, we explore the application of restricted Bayesian Lasso regression, specifically targeting high-dimensional models where linear inequality constraints are applied to the coefficients. The Lasso method, known for its efficacy in variable selection and regularization, is enhanced through a Bayesian framework that allows for the incorporation of prior knowledge and uncertainty in coefficients. We will discuss the theoretical foundations of the restricted Bayesian Lasso, including the specification of priors and the computational techniques employed to derive posterior distributions. Through simulation studies and real data applications, we will demonstrate the advantages of this approach over traditional Bayesian Lasso regression, particularly in terms of prediction accuracy.

Keywords: Bayesian Lasso regression. High dimensional models. linear inequality constraints, Gibbs sampler. Truncated multivariate normal distribution



61th Article Title: Stellar Spectra Interpolation Using Machine Learning Techniques

Author(s): Mahsa Sohrabi, Reza Mokhtari & Soroush Shakeri

Abstract: In this study, empirical stellar spectral libraries are used to interpolate stellar spectra based on three stellar atmospheric parameters. However, the irregular distribution of stars' data in these libraries often leads to inadequate results with traditional interpolation methods. After examining a radial basis function (RBF) network and its structure, a fully connected artificial neural network (ANN) model and a random forest (RF) model are trained to generate a stellar spectrum for a given set of atmospheric parameters. Comparisons reveal that the ANN and RF models significantly outperform RBF and TGM2 interpolations, with both offering comparable performance and an R^2 score of approximately 96 percent. A brief analysis shows that these machine learning models are robust against small input errors.

Keywords: Stellar Spectral Library, Radial Basis Function Network, Artificial Neural Network, Random Forest



62th Article Title: Stochastic multi-class support vector machine: behavior in dealing with outliers

Author(s): Tara Mohammadi, Hadi Jabbari Nooghabi & Sohrab Effati

Abstract: This study examines the effectiveness of stochastic multi-class support vector machine (SVM) in handling outliers. Outliers, which deviate significantly from most data, can negatively impact classification accuracy. In this paper, the constraints are in the probabilistic form, so to release the probabilistic state we use statistical tools and theorems. Hence, entering the median into the SVM structure enhances the algorithm's robustness and generalization capabilities, reducing sensitivity to noisy data. A real-world breast ductal carcinoma dataset demonstrates that stochastic multi-class SVM outperforms traditional deterministic SVM in rich environments. This research highlights the potential of stochastic approaches to improve the efficiency of SVM where data irregularities are common.

Keywords: Multi-class SVM. Median. Outlier. Probabilistic constraint



63th Article Title: Survival Analysis for Brain Stroke Prediction mortality: Hybrid Machine learning and statistical methods

Author(s): Seyedeh maryam Mousavi, Solmaz Norouzi, Samira Ahmadi, Farshid Farzipoor, Ramazan Fallah & Mohammad Asghari Jafarabadi

Abstract: Objective: Stroke is the leading cause of death globally and poses a significant health threat to many individuals. This study aimed to assess the predictive accuracy of mortality and survival rates in stroke patients using machine learning and statistical methods.

Method: We utilized data from stroke patients collected during a retrospective study conducted in Iran from 2004 to 2022. The analysis employed the Cox proportional hazards model as a statistical approach, alongside a hybrid model combining Random Survival Forest (RSF) and Support Vector Machine (SVM) techniques. Python and R were used for modeling, with the Area Under the Curve (AUC) serving as the primary metric for model evaluation and performance comparison.

Results: While Cox regression models can be developed for predicting stroke patient prognosis, their accuracy is lower than that of certain machine learning methods in various studies. The hybrid RSF-SVM model integrates two machine learning algorithms to enhance accuracy and efficiency in predicting time-to-event outcomes. In our findings, the Cox proportional hazards model achieved an accuracy of 74%, whereas the hybrid model reached 86% accuracy.

Conclusions: Analyzing the survival of stroke patients is crucial for effective health service planning and evaluation. Furthermore, identifying and validating prognostic factors is essential for guiding treatment protocols.

Keywords: Survival Analysis. random survival forest. support vector machine. Hybrid Model. Brain Stroke



64th Article Title: The best data fitting using by modified generalized Laguerre polynomials and particle swarm optimization algorithm

Author(s): Gholam Reza Zaki & Ali Asghar Joderie Akbarfam

Abstract: In this paper, for the given data in the plane, we construct a curve of degree M by using the linear combination of modified generalized Laguerre polynomials. These polynomials have two parameters Alpha and Beta. First, we show that the least squares error of this constructed curve for fitting the given data, depends on the parameters of its polynomials. Then, in order to find the best parameters that minimize the least squares error, we use the particle swarm optimization algorithm. The presented examples show the advantage of the proposed method.

Keywords: curve fitting, particle swarm optimization algorithm, modified generalized Laguerre polynomials



65th Article Title: Unlocking the Potential of Metaverse in Innovative and Immersive Digital Health

Author(s): Fatemeh Ebrahimzadeh & Ramin Safa

Abstract: The concept of Metaverse has attracted a lot of attention in various fields and one of its important applications is health and treatment. The Metaverse has enormous potential to transform healthcare by changing patient care, medical education, and the way teaching/learning and research are done. The purpose of this research is to provide an introduction to the basic concepts and fundamental technologies of the Metaverse. This paper examines the pros and cons of the Metaverse in healthcare context and analyzes its potential from the technology and AI perspective. In particular, the role of machine learning methods is discussed; We will explain how machine learning algorithms can be applied to the Metaverse generated data to gain better insights in healthcare applications. Additionally, we examine the future visions of the Metaverse in health delivery, by examining emerging technologies such as blockchain and also addressing privacy concerns. The findings of this study contribute to a deeper understanding of the applications of Metaverse in healthcare and its potential to revolutionize the delivery of medical services.

Keywords: Metaverse. Healthcare. Virtual Reality. Artificial Intelligence. Machine Learning



66th Article Title: Unraveling Temporal Dynamics: Utilizing Cross-Lagged Panel Models for Data Mining in Dental Health Research

Author(s): Reza Delavari, Jamshid Jamali & Fatemeh Moghaddam

Abstract: The Cross-Lagged Panel Model (CLPM) is a robust statistical tool widely employed in longitudinal data analysis, enabling researchers to assess reciprocal relationships between variables over time. CLPM has significant applications in data mining within dental science, particularly in elucidating the temporal relationships between psychological and behavioral variables related to oral health. This cohort study focused on individuals aged 60 years and older in Tehran, examining changes in self-efficacy scores, risk perception, and outcome expectations across three time periods: before the intervention and one and three months post-intervention. The study utilized the Health Action Process Approach (HAPA) model to guide the analysis. Both the standard lag-cross panel model and the random-origin lag-cross panel model were fitted to the data. The standard lag-cross panel model evaluates relationships between variables at different time points, capturing both lagged effects (influence of past values) and cross-lagged effects (inter-relationships between variables). This model typically employs fixed or random effects to control for unobserved individual heterogeneity but does not account for variations in initial conditions. In contrast, the random-origin lag-cross panel model enhances this framework by accommodating differences in baseline levels among individuals, thereby allowing for the examination of individual trajectories and variations in relationships from distinct starting points. This model is particularly advantageous in contexts characterized by substantial variability in initial conditions, as it effectively captures individual-specific dynamics. To assess the goodness of fit of the models, the study utilized several fit indices, including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA).

Keywords: Cross-Lagged Panel Model. Longitudinal Data Analysis. Data Mining. Dental Science. Psychological Variables