مجموعه چکیده مقالات

هجدهمین

کنفرانس آمار ایران

دانشکده ریاضی، دانشگاه صنعتی خواجه نصیرالدین طوسی

۷ تا ۹ شهریور ۱۴۰۵ | تهران، ایران

نسخه فارسی

اطلاعات دبیرخانه

وبگاه: https://isc18.kntu.ac.ir/fa

رایانامه: isc18@saba.kntu.ac.ir

تلفن: ۰۲۱-۷۷۱۲۵۳۶۵ (داخلی ۱۳۴ تا ۴۰۲) | ۰۲۱-۷۷۱۲۵۲۵۴

نشانی: انتهای بزرگراه زین‌الدین شرق، خیابان وفادار شرقی، بلوار دانشگاه خواجه نصیرالدین طوسی، پردیس شهید رضائی‌نژاد

صندوق پستی: ۳۳۸۱-۱۶۷۶۵

دانشگاه صنعتی خواجه نصیرالدین طوسی

انجمن آمار ایران

به نام خداوند بخشنده مهربان

دانشکده ریاضی
دانشگاه صنعتی خواجه نصیرالدین طوسی

مجموعه چکیده‌ها (فارسی)

از

هجدهمین کنفرانس آمار ایران

به اهتمام:

دکتر راضیه خودسیانی (مسؤول دبیرخانه کنفرانس)

دکتر احد ملک‌زاده (دبیر کنفرانس)

دکتر محمد آرشی (دبیر کمیته علمی)

دکتر مریم عبدالعلی (عضو کمیته اجرایی)

دکتر جواد کوشکی (عضو کمیته اجرایی)

تابستان ۱۴۰۵

پیشگفتار

به لطف پروردگار، هجدهمین کنفرانس آمار ایران به میزبانی گروه علوم کامپیوتر و آمار، دانشکده ریاضی، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، در تاریخ ۷ تا ۹ شهریور ۱۴۰۵ (۲۹ تا ۳۱ اوت ۲۰۲۶) برگزار می‌شود. این رویداد علمی با هدف فراهم کردن بستر مناسب برای ارائه آخرین دستاوردهای پژوهشی، بحث و تبادل نظر میان اعضای هیأت علمی، پژوهش‌گران و کارشناسان آمار، و نیز ارتقای فرهنگ آماری، سواد داده‌ای و کاربردهای علم آمار در حوزه‌های گوناگون برگزار می‌گردد.

کمیته‌های برگزاری و علمی کنفرانس نهایت تلاش خود را برای برگزاری آن در سطحی علمی و شایسته به‌کار بسته‌اند. با دبیری علمی دکتر محمد آرشی و دبیری اجرایی دکتر احد ملک‌زاده، سخنرانی‌های کلیدی، نشست‌های تخصصی و مقالات پذیرفته‌شده، با حمایت انجمن آمار ایران و دانشگاه میزبان، برنامه‌ریزی شده است. چکیده‌های فارسی مقالات منتخب در کتابچه حاضر گردآوری شده‌اند تا در اختیار جامعه آماری قرار گیرند.

بدین‌وسیله از همه همکاران و مسئولینی که مستقیم یا غیرمستقیم در برگزاری این کنفرانس نقش داشته‌اند سپاسگزاری می‌شود: ریاست و معاونت‌های دانشگاه صنعتی خواجه نصیرالدین طوسی، جناب آقای دکتر علی ذاکری رئیس دانشکده ریاضی و همکاران محترم دانشکده، هیأت مدیره انجمن آمار ایران، اعضای کمیته‌های علمی و اجرایی، دبیرخانه، داوران، حامیان علمی و مالی، و همه عزیزانی که به هر شکل یاری‌گر کنفرانس بوده‌اند.

کمیته علمی هجدهمین کنفرانس آمار ایران،
تابستان ۱۴۰۵.

کمیته اجرایی

دکتر احد ملک‌زاده
دبیر کنفرانس - آمار (استنباط)
دانشگاه صنعتی خواجه نصیرالدین طوسی

دکتر راضیه خودسیانی
مسؤول دبیرخانه - آمار (طرح و تحلیل آزمایش‌ها)
دانشگاه صنعتی خواجه نصیرالدین طوسی

دکتر مریم عبدالعلی
علوم کامپیوتر (هوش مصنوعی)
دانشگاه صنعتی خواجه نصیرالدین طوسی

دکتر جواد کوشکی
ریاضی (بهینه‌سازی)
دانشگاه صنعتی خواجه نصیرالدین طوسی

کمیته علمی

حامیان علمی


دانشگاه صنعتی خواجه نصیرالدین طوسی


IRANIAN STATISTICAL SOCIETY 1996


دانشگاه گیلان


جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری


پژوهشگاه رویان
جهاد دانشگاهی


JASLDA
Journal of Advances in Statistical Learning and Data Analysis


پژوهشکده‌ی آمار


همه با هم
جهاد
کشاورزی
۱۳۷۹


دانشگاه خوارزمی

حامیان مالی


وزارت علوم تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران
IRANDOC


پژوهشکده بیمه


مرکز آمار ایران
Statistical Centre of Iran


ریاست جمهوری
سازمان اداری و استخدامی کشور


پژوهشکده‌ی آمار


بانک کشاورزی


FARAN
sustain it ...

هجدهمین کنفرانس آمار ایران

مجموعه چکیده‌ها (فارسی)

فهرست مطالب

مرتب شده بر اساس نام خانوادگی نویسنده اول

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

استنباط روی چندک توزیع پارتو

همراز ابراهیم دوست[۱*]، احد ملک زاده[۱]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی

چکیده: توزیع پارتو یکی از توزیع‌های دم‌سنگین و چوله به راست است که در مدل‌های پدیده‌هایی مانند توزیع ثروت‌سازی، شدت زلزله، خسارت‌های بیمه‌ای و مهندسی قابلیت اعتماد کاربرد گسترده دارد. در این مقاله روش‌های برآورد نقطه‌ای پارامترهای این توزیع شامل برآوردگرهای ماکزیمم درستنمایی (MLE) و برآوردگرهای یکتای ناریب با کمترین واریانس (UMVUE) مرور می‌گردد. سپس با استفاده از تبدیل لگاریتمی، مسئله استنباط روی چندک توزیع پارتو تسهیل می‌شود. یک کمیت محوری دقیق معرفی می‌شود که توزیع آن مستقل از پارامترهای مجهول بوده و امکان ساخت فاصله اطمینان دقیق را فراهم می‌کند. با استفاده از این کمیت محوری، فاصله اطمینان عمومی و کوتاهترین فاصله اطمینان برای چندک توزیع پارتو ساخته می‌شوند. نتایج شبیه‌سازی مونت‌کارلو نشان می‌دهد که کوتاهترین فاصله اطمینان عملکرد بهتری نسبت به فاصله اطمینان عمومی دارد و با افزایش حجم نمونه، طول فواصل کاهش می‌یابد. روش ارائه شده در این مقاله، اولین روش دقیق برای استنباط روی چندک توزیع پارتو است که هیچ تقریب مجانبی در آن استفاده نشده است.

واژه‌های کلیدی: توزیع پارتو، چندک، برآوردگر یکتای ناریب با کمترین واریانس، کمیت محوری دقیق، کوتاهترین فاصله اطمینان، شبیه‌سازی مونت کارلو.

*نویسنده مسئول: ebrahimdoosthamraz@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

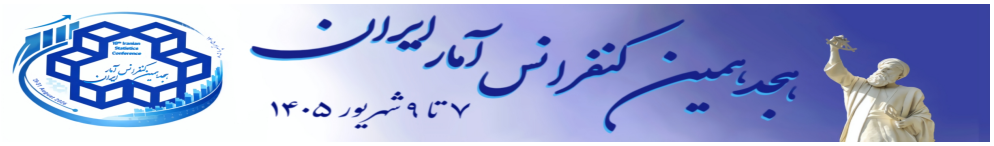
شبکه عصبی خودرمزگذار مقابل تحلیل مؤلفه‌های اصلی: یک مطالعه مقایسه‌ای در کاهش ابعاد و بازسازی داده‌ها

میلاد آراسته نیا[1]*، سیدرضا هاشمی[1]
[1]دانشگاه رازی

چکیده: امروزه با پیشرفت تکنولوژی و دسترسی بیشتر به اطلاعات با حجم بیشتری از داده‌ها، به خصوص از نظر بعد، مواجه هستیم. داده با بعد بیشتر همیشه به معنای بهتر بودن نیست و گاهی اوقات این بعد بالا مترادف با نویز بیشتر، محدودیت‌های محاسباتی و خطر بیش‌برازشی در مدل می‌شود. هدف از این مقاله مطالعه دو روش تحلیل مؤلفه‌های اصلی و شبکه عصبی خودرمزگذار جهت کاهش ابعاد داده به نحوی که اطلاعات مهمی از مسئله از دست نرود و منجر به سریع‌تر و عملکرد بالاتر شود. با کمک این رویکردها فضای اصلی مسئله را به یک فضای جدید نگاشت می‌کنیم که در فضای جدید ویژگی‌هایی مورد نظر خود را اختیار می‌کنیم. نگاشتی که در ابرصفحه جدید با کمک روش مؤلفه‌های اصلی انجام می‌شود یک ترکیب خطی از ویژگی‌های اولیه می‌باشد. حال این‌که در خیلی از مسائل دنیای واقعی ساختارهای غیرخطی وجود دارد که روش مؤلفه‌های اصلی از عهده حل این مسائل برنخواهد آمد در نتیجه شبکه عصبی خودرمزگذار به عنوان روشی که توانایی یادگیری نگاشت‌ها غیرخطی را دارد را معرفی می‌کنیم. نتایج نشان می‌دهند که شبکه عصبی خودرمزگذار عملکرد بهتری در استخراج ویژگی‌های معنادار و همچنین بازسازی داده‌ها با پیچیدگی غیرخطی دارد.

واژه‌های کلیدی: استخراج ویژگی،کاهش ابعاد،تحلیل مؤلفه‌های اصلی،شبکه عصبی خودرمزگذار

*نویسنده مسئول: r.hashemi@razi.ac.ir



تنوع داده، تنوع مدل: اهمیت شناخت ساختار داده در مدیریت ریسک بیمه‌ای با مطالعه موردی بیمه آتش‌سوزی

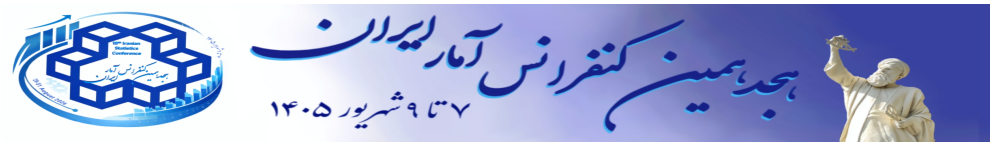
شیما آراء[۱]، زهرا برزگر[۱*]
[۱]شرکت بیمه سامان

چکیده: مدیریت ریسک مدرن در صنعت بیمه، بیش از آنکه به حجم داده‌ها وابسته باشد، به شناخت دقیق «ساختار داده» و انتخاب مدل متناسب با آن نیازمند است. فرض کلاسیک استقلال مشاهدات (i.i.d.) که سنگ بنای بسیاری از مدل‌های اکتشافی و اکچوئری سنتی است، در حالیکه این فرضیه هواره برقرار نسیت و تنوع ساختار داده در صنعت بیمه، لزوم توجه همه‌جانبه در انتخاب روش مدل‌بندی را ایجاب می‌کند. پژوهش حاضر، با هدف ارائه‌ی مثالی از ساختارهای پیچیده و متنوع، به بیمه‌ی آتش‌سوزی می‌پردازد که با ساختاری معمول هم‌خوانی ندارد و دارای ماهیت جغرافیایی و مکانی است. در این راستا، با به‌کارگیری پنج ساختار متنوع برای ماتریس همسایگی فضایی (باینری، استاندارد شده سطری، کل، واحد و تثبیت واریانس) و تحلیل آزمون موران، ابتدا وجود ساختار وابستگی فضایی در نسبت خسارت به حق‌بیمه تایید شد. در گام بعد، با برازش تنوعی از مدل‌های خودبازگشتی همزمان (SAR)، تأثیر شاخص کلان اقتصادی «نرخ مشارکت اقتصادی استان‌ها» بر میانگین خسارت تحلیل گردید. یافته‌ها نشان می‌دهد که مدل‌های مبتنی بر استانداردسازی‌های واریانسی و سطری، با هموارسازی بهینه خطاها، برازش دقیق‌تری از ساختار ریسک ارائه می‌دهند و نادیده گرفتن این تنوع ساختاری، به کژبرآوردی ریسک منجر می‌شود. ازینرو پیشنهاد می‌شود روش‌های متنوع مدل‌بندی ساختارهای مختلف‌منطبق با ماهیت داده‌ها در اولویت قرار داشته باشد.

واژه‌های کلیدی: : مدیریت ریسک بیمه‌ای، ساختار داده، تنوع مدل، مدل‌های ناحیه‌ای، مدل خودبازگشتی همزمان (SAR)، بیمه آتش‌سوزی.

---

*نویسنده مسئول: z.barzegar@samaninsurance.ir



انتخاب متغیر استوار مضاعف در رگرسیون چندک نیمه‌پارامتری با پاسخ‌ها و متغیرهای کمکی گمشده

احسان ارمز [۱*]

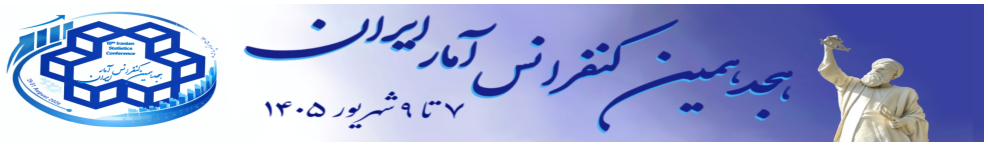
[۱]گروه ریاضی و آمار، واحدمشهد، دانشگاه آزاد اسلامی، مشهد، ایران

چکیده: این مقاله چارچوبی برای انتخاب متغیر و برآورد هم‌زمان در رگرسیون چندک نیمه‌پارامتری (مدل‌های خطی جزئی) با پاسخ‌ها و متغیرهای کمکی گمشده، تحت فرض گم‌شدگی تصادفی (MAR) معرفی می‌کند. روش پیشنهادی بر پایه معادلات برآورد احتمال معکوس وزن‌دار شده‌ی افزوده (AIPW) بنا شده است که از ویژگی استواری مضاعف برخوردار است. برای انجام انتخاب متغیر، از توابع جریمه SCAD یا MCP در بخش پارامتری و از تابع جریمه group-SCAD برای ضرایب B –اسپلاین در بخش ناپارامتری استفاده شده است. نتایج نظری این روش شامل شناسایی‌پذیری، سازگاری یکنواخت، سازگاری n، نرمالیتی مجانبی، ویژگی اوراکل و نمایش خطی بهادر استخراج و اثبات شده‌اند.

واژه‌های کلیدی: وزن‌دهی احتمال معکوس افزوده، نمایش بهادر، گم‌شدگی تصادفی، ویژگی اوراکل، مدل خطی جزئی، رگرسیون چندک جریمه‌شده، کارایی نیمه‌پارامتری، انتخاب متغیر.

---

*نویسنده مسئول: ehsan.ormoz@iau.ac.ir



ارزیابی حساسیت برآوردهای رگرسیونی به چارچوب تحلیلی در داده‌های پیمایشی پیچیده: مطالعه‌ای کاربردی با داده‌های NHANES 2017-2018

شیما آزادیان[1*]، عرفان صادقی[1]
[1]دانشگاه علوم پزشکی شیراز

چکیده: نتایج رگرسیونی در داده‌های پیمایشی، به‌شدت به تصمیمات تحلیلی مانند وزن‌دهی، ساختار نمونه‌گیری و تعریف متغیرها وابسته‌اند، نادیده‌گرفتن این موارد منجر به برآوردهای ناپایدار می‌شود. این مطالعه با استفاده از داده‌های NHANES 2017-2018 ، حساسیت برآوردها را در سه مدل (بدون وزن، وزن‌دار و طراحی کامل) ارزیابی کرد. ارتباط میان BMI و فشار خون سیستولیک (SBP) با تعاریف مختلف پیامد، به‌عنوان مطالعه موردی استفاده شد. در مدل بدون وزن، ضریب BMI و خطای استاندارد به‌ترتیب 0.3894 و 0.0366 بود. با اعمال وزن‌ها، برآورد به 0.4245 رسید. در مدل طراحی کامل، برآورد نقطه‌ای ثابت ماند، اما خطای استاندارد به 0.0501 افزایش و فاصله اطمینان گسترش یافت. نتایج نشان می‌دهد وزن‌دهی عمدتا بر برآورد نقطه‌ای و لحاظ طراحی بر عدم‌قطعیت اثرگذار است، لذا تحلیل حساسیت باید بخشی استاندارد از گزارش نتایج باشد.

واژه‌های کلیدی: داده‌های پیمایشی پیچیده، NHANES، تحلیل حساسیت، رگرسیون وزن‌دار، طراحی نمونه‌گیری و فشار خون سیستولیک.

---

*نویسنده مسئول: shima_azadian@sums.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

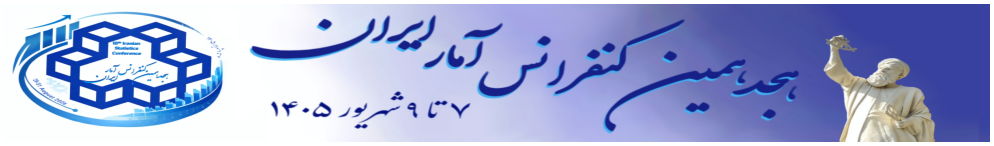
تخمین ظرفیت بدهی شرکت‌ها با استفاده از الگوریتم‌های یادگیری ماشین مبتنی بر شاخص‌های جریان نقد

سپیده اشرفی[۱*]، احمدرضا یزدانیان[۱]
[۱]دانشگاه خوارزمی

چکیده: هدف این پژوهش، تخمین ظرفیت بدهی شرکت‌ها با استفاده از الگوریتم‌های یادگیری ماشین و تمرکز بر متغیرهای جریان نقدی است. برخلاف مطالعات گذشته که عمدتا از مدل‌های رگرسیونی سنتی استفاده می‌کردند، این تحقیق شاخص‌های عملکرد مالی، ساختار سرمایه و نسبت‌های جریان نقد را به‌منظور آموزش مدل‌ها به‌کار گرفت. بدین منظور، عملکرد سه الگوریتم Random Forest، XGBoost و LightGBM در پیش‌بینی ظرفیت بدهی مقایسه شد. یافته‌ها نشان داد که مدل XGBoost بالاترین دقت را دارد و متغیرهای مبتنی بر جریان نقد، دقت تخمین را به‌طور معناداری بهبود می‌دهند. نتایج این پژوهش چارچوبی داده‌محور و کارآمد برای مدیریت ریسک اعتباری، رتبه‌بندی و تصمیم‌گیری‌های بهینه تأمین مالی در اختیار مدیران و بانک‌ها قرار می‌دهد.

واژه‌های کلیدی: ظرفیت بدهی،الگوریتم‌های یادگیری ماشین،جریان نقد،نسبت‌های مالی

*نویسنده مسئول: sepideashrafi78@gmail.com



درخت تصمیم مبتنی بر شاخص آتکینسون و نقش پارامتر حساسیت در معیارهای شکافت

هدیه افتخاری مودی[۱*]، محمد خراشادی زاده[۱]، غلامرضا محتشمی برزادران[۲]، حمیدرضا نیلی ثانی[۱]

[۱]دانشگاه بیرجند

[۲]دانشگاه فردوسی مشهد

چکیده: درخت تصمیم یکی از روش‌های مهم و پرکاربرد در داده‌کاوی و یادگیری ماشین است که به دلیل سادگی تفسیر، توانایی کار با داده‌های عددی و طبقه‌ای، و قابلیت استخراج قواعد تصمیم‌گیری، جایگاه ویژه‌ای در مسائل طبقه‌بندی و پیش‌بینی دارد. عملکرد این درخت‌ها به شدت وابسته به معیارهایی است که برای انتخاب ویژگی و ایجاد گره‌های شکاف به کار می‌روند. معیارهایی مانند آنتروپی، بهره اطلاعات، ضریب جینی و نسبت بهره اطلاعات از شناخته‌شده‌ترین روش‌ها در این زمینه هستند. با وجود این، این معیارها در بسیاری از موارد از یک حساسیت ثابت نسبت به توزیع داده‌ها پیروی می‌کنند و امکان تنظیم میزان توجه به عدم‌توازن یا نابرابری داده‌ها را به‌صورت مستقیم فراهم نمی‌سازند. در این مقاله، ضمن مروری کوتاه بر مفاهیم پایه درخت تصمیم، شاخص آتکینسون به‌عنوان یک معیار پارامتریک برای سنجش نابرابری معرفی می‌شود. سپس ایده استفاده از پارامتر در شاخص آتکینسون برای تنظیم حساسیت درخت تصمیم مورد بحث قرار می‌گیرد. این پارامتر می‌تواند رفتار معیار شکافت را نسبت به نمونه‌های کم‌تراکم یا کم‌درآمد در داده‌های عددی کنترل کند و در نتیجه به‌عنوان یک اهرم تنظیم‌گر برای ساخت درخت تصمیم حساس‌تر یا محافظه‌کارتر عمل نماید. در پایان، مزایا و ظرفیت‌های این ایده برای توسعه روش‌های جدید در ساخت درخت تصمیم بررسی می‌شود.

واژه‌های کلیدی: شاخص آتکینسون

---

*نویسنده مسئول: hedieh.eftekhari73@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

برآوردگرهای انقباضی بیز تجربی نوع استاین در مدل رگرسیون لجستیک بعد بالا

سیده زهرا آقامحمدی[۱]*
[۱]دانشگاه آزاد اسلامی واحد اسلامشهر

چکیده: در این مقاله مدل بیز تجربی در چارچوب رگرسیون لجستیک بعد بالا بر اساس توزیع‌های پیشین مبتنی بر همبستگی برای هر دو ضرایب رگرسیون و فضای مدل در نظر گرفته می‌شود. از آنجا که فرم بسته‌ای برای توزیع پسین با پارامترهای بعد بالا وجود ندارد، از محاسبات عددی برای تقریب توزیع آن استفاده می‌شود. همچنین، بر اساس اطلاع پیشین غیر قطعی که به صورت محدودیت روی فضای پارامتر اعمال می‌شود، استراتژی برآوردگرهای انقباضی بیز تجربی نوع استاین را پیشنهاد می‌دهیم. در انتها در قالب یک مطالعه‌ی شبیه‌سازی و مثال واقعی کارایی روش‌های ارائه شده مورد ارزیابی قرار می‌گیرد

واژه‌های کلیدی: بیز تجربی،بعد بالا،رگرسیون لجستیک،توزیع پیشین،برآورد انقباضی استاین

*نویسنده مسئول: zahraaghamohamadi@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

پیش‌بینی شاخص کیفیت هوا با استفاده از مدل‌های یادگیری ماشین، یادگیری عمیق و شبکه‌های عصبی گرافی-زمانی

مریم اکبری[۱*]، امید کریمی[۱]، فاطمه حسینی[۱]
[۱]دانشگاه سمنان

چکیده: پیش‌بینی شاخص کیفیت هوا یکی از مسائل مهم در تحلیل داده‌های محیط‌زیستی و مطالعات فضایی–زمانی است که نقش مهمی در مدیریت شهری، حفاظت از سلامت عمومی و تصمیم‌گیری‌های زیست‌محیطی ایفا می‌کند. با توجه به ماهیت پیچیده، غیرخطی و وابسته به زمان و مکان داده‌های آلودگی هوا، استفاده از روش‌های نوین یادگیری ماشین و یادگیری عمیق می‌تواند به بهبود دقت پیش‌بینی کمک کند. در این مقاله، عملکرد چند مدل پیش‌بینی شامل رگرسیون خطی، جنگل تصادفی، XGBoost، شبکه عصبی بازگشتی و شبکه عصبی گرافی فضایی–زمانی برای پیش‌بینی شاخص کیفیت هوا بر روی داده‌های واقعی شهر تهران مورد بررسی قرار گرفت. در مرحله آماده‌سازی داده‌ها، عملیات تکمیل داده‌های گمشده، استخراج ویژگی‌های زمانی و لحاظ نمودن الگوهای فصلی انجام شد. همچنین برای مدل‌سازی وابستگی‌های مکانی، ساختار گراف مبتنی بر فاصله جغرافیایی میان ایستگاه‌ها برای مدل گرافی تعریف گردید. نتایج نشان داد که مدل جنگل تصادفی با کمترین میزان خطا، بهترین عملکرد را در میان مدل‌های مورد بررسی ارائه می‌دهد و مدل XGBoost نیز عملکردی نزدیک به آن دارد. یافته‌های این مطالعه نشان می‌دهد که در مجموعه داده مورد بررسی، مدل‌های مبتنی بر درخت تصمیم نسبت به مدل‌های پیچیده‌تر یادگیری عمیق و گرافی، کارایی بالاتری در پیش‌بینی شاخص کیفیت هوا دارند.

واژه‌های کلیدی: شاخص کیفیت هوا، داده‌های فضایی–زمانی، یادگیری ماشین، جنگل تصادفی، شبکه عصبی گرافی.

*نویسنده مسئول: omid.karimi@semnan.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

چشم‌انداز: داده‌کاوی مواد و داده‌های عظیم؛ تحقق چهارمین پارادایم علم در علم مواد

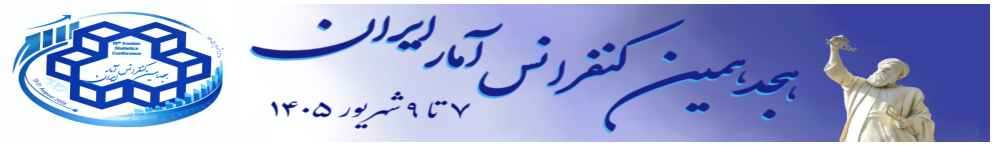
علی اکبریان آذر[۱*]، حمید خرسند[۱]، شیرین تیموری غرب[۲]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی
[۲]دانشگاه تبریز

چکیده: توانایی بشر در گردآوری حجم عظیمی از داده‌ها بسیار سریع‌تر از توانایی او در تحلیل و بهره‌برداری از این داده‌ها رشد کرده است. این موضوع اهمیت رویکردی نوین در پژوهش‌های علمی را نشان می‌دهد که بر کشف دانش از طریق تحلیل داده‌ها استوار است. اهمیت استفاده از روش‌های نوین تحلیل داده در برنامه «ژنوم مواد» نیز مورد تأکید قرار گرفته و همین امر به گسترش حوزه‌ای نوظهور به نام «داده‌کاوی مواد» کمک کرده است. این حوزه از داده‌ها و روش‌های هوشمند برای مطالعه، طراحی و توسعه مواد جدید بهره می‌گیرد. در این مقاله بررسی می‌شود که چگونه روش‌های مبتنی بر داده در شناسایی ارتباط میان مراحل ساخت مواد، ساختار داخلی آن‌ها، خواص مهندسی و عملکرد نهایی‌شان نقش مؤثری ایفا می‌کنند. برای روشن‌تر شدن این موضوع، نمونه‌هایی از مدل‌های پیش‌بینی خواص مواد و همچنین روش‌هایی برای کشف و طراحی مواد جدید ارائه شده است. به‌کارگیری این روش‌های تحلیلی باعث می‌شود پژوهشگران در زمان کوتاه‌تری به نتایج ارزشمند دست یابند و بتوانند مواد جدید را با هزینه کمتر و سرعت بیشتر طراحی و تولید کنند. این هدف یکی از مهم‌ترین برنامه‌های پژوهشی در علم مواد به شمار می‌آید و می‌تواند روند نوآوری و توسعه مواد پیشرفته را به‌طور چشمگیری تسریع کند.

واژه‌های کلیدی: ماشین لرنینگ،داده بزرگ،مهندسی مواد،آنالیز داده

---

*نویسنده مسئول: a.ostadakbrianazar@email.kntu.ac.ir



حکمرانی داده‌های سلامت در عصر هوش مصنوعی: چالش‌ها، الزامات و راهبردها

سارا امامقلی پور[۱*]، قدرت طاهری[۲]
[۱]دانشگاه علوم پزشکی تهران
[۲]مرکز تحقیقات نیکاپویش

چکیده: تحول دیجیتال و گسترش هوش مصنوعی، داده‌های سلامت را به دارایی راهبردی برای تصمیم‌گیری، سیاست‌گذاری و نوآوری تبدیل کرده است. این مطالعه مروری ـ تحلیلی، ابعاد مفهومی، نهادی و سیاستی حکمرانی داده‌های سلامت در عصر هوش مصنوعی را بررسی می‌کند. نتایج نشان می‌دهد بهره‌گیری ایمن و عادلانه از هوش مصنوعی بدون حکمرانی مؤثر داده‌ها با چالش‌هایی نظیر نقض حریم خصوصی، کاهش اعتماد عمومی، عدم شفافیت الگوریتمی و سوگیری‌های داده‌ای مواجه خواهد شد. ارزیابی وضعیت ایران با استفاده از شاخص ODIN نشان می‌دهد که مهم‌ترین موانع، نه محدودیت‌های فنی، بلکه ضعف‌های ساختاری در حکمرانی داده شامل سیلوهای اطلاعاتی، ضعف فرادادهها، فقدان راهبرد ملی داده و نبود چارچوب‌های شفاف برای استفاده ثانویه از داده‌هاست. بر این اساس، چارچوبی سه‌لایه شامل زیرساخت و فناوری، صیانت و قانون‌گذاری، و اخلاق و مشارکت جمعی پیشنهاد می‌شود. یافته‌ها تأکید می‌کنند که تحقق هوش مصنوعی مسئولانه در نظام سلامت ایران مستلزم گذار به حکمرانی مشارکتی، استانداردمحور و عدالت‌گرا است.

واژه‌های کلیدی: حکمرانی داده‌های سلامت، هوش مصنوعی، حریم خصوصی، سوگیری الگوریتمی، تعامل‌پذیری

---

*نویسنده مسئول: s-emamgholipour@tums.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

شاخصی جدید برای مولفه های قابل استفاده مجدد یک سیستم

ابراهیم امینی سرشت[۱]، حامد لروند[۲]*
[۱]دانشگاه بوعلی سینا
[۲]دانشگاه صنعتی اصفهان

چکیده: در حوزه قابلیت اعتماد یک سیستم منسجم ممکن است بهصورت برگشتناپذیردچار شکست شود، در حالی که برخی از اجزای آن همچنان عملیاتی و بالقوه قابل استفاده مجدد در سیستمهای دیگر باشند. این اجزای باقیمانده نه جدید هستند و نه خراب، بلکه فرسوده شده اند و سطح تخریب آنها باید قبل از استفاده مجدد بهدرستی کمی سازی شود. در این مقاله، یک شاخص جدید مبتنی بر اطلاعات برای اندازهگیری میزان تخریب اجزای استفادهشده معرفی میشود. شاخص پیشنهادی بهعنوان کاهش نرمالشده اطلاعات تعریف میشود که از طریق نسبت آنتروپی باقیمانده یک جزء استفادهشده به آنتروپی شانونی عمر اولیه آن اندازهگیری میشود. در این مقاله دو ویژگی اساسی از این شاخص بررسی شده است. ویژگی اول آن افزایش میزان این شاخص با افزایش میزان کارکرد مولفه و ویژگی دیگر شرایط مرزی این شاخص را نشان می دهد که شاخص تخریب مولفه کاملا جدید برابر با صفر است و با تخریب کامل مولفه به یک نزدیک میشود.

واژههای کلیدی: قابلیت اطمینان،آنتروپی شانون،آنتروپی باقیمانده،شاخص مبتنی بر اطلاعات

*نویسنده مسئول: lorvandhamed@iut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌بندی داده‌های فضایی-زمانی حجیم با استفاده از مشبکه‌ی کمکی

علی انصاری[۱*]، محسن محمدزاده[۱]، کیومرث مترجم[۱]
[۱]دانشگاه تربیت مدرس

چکیده: وقتی مجموعه داده‌های فضایی-زمانی حجیم باشند، معمولا بار محاسباتی سنگین تحلیل داده‌ها می‌تواند منجر به توقف یا دشواری اجرای روش‌های متعارف آمار فضایی-زمانی شود. در این مطالعه، یک مدل فضایی-زمانی سلسله‌مراتبی بیزی محاسباتی پیشنهاد می‌شود که در آن برای غلبه بر چالش‌های محاسباتی، از یک مشبکه‌ی کمکی بهره گرفته شده است. در این راستا، همبستگی فضایی داده‌ها توسط یک میدان تصادفی مارکوفی گاوسی بر روی این مشبکه‌ی کمکی تقریب زده شده و همبستگی زمانی آن‌ها با یک مدل اتورگرسیو برداری بیان می‌شود. ابتدا با تعیین مراحل و میزان پیچیدگی محاسبات روش پیشنهادی، نشان داده خواهد شد که چگونه می‌توان این روش را برای تحلیل داده‌های فضایی-زمانی حجیم با زمان‌های اجرایی قابل‌قبول به کار گرفت. در نهایت، عملکرد مدل پیشنهادی ارزیابی شده و نحوه کاربست آن در تحلیل مجموعه داده‌های بارش باران نشان داده می‌شود.

واژه‌های کلیدی: داده‌های فضایی-زمانی، داده‌های حجیم، مشبکه‌ی کمکی، میدان تصادفی مارکوفی گاوسی، مدل سلسله‌مراتبی بیزی.

*نویسنده مسئول: a.ansari@modares.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

خانواده‌ای ناوردا از پیشین‌های هندسی–آماری

کیمیا باوفای سمیرمی[۱*]، بهزاد نجفی سقزچی[۱]، امیرحسین قطاری[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: در این مقاله، یک خانواده‌ی ناوردا از پیشین‌های بیزی در چارچوب هندسه‌ی اطلاعات معرفی می‌شود. این ساختار جدید با افزودن یک مؤلفه‌ی هندسی ناوردا، مبتنی بر نرم تانسور آماری–چنتسوف ، به پیشین جفریز به دست آمده است. از این‌رو، پیشین پیشنهادی را می‌توان به‌عنوان یک پیشین کم‌آگاهنده (یا نیمه‌آگاهنده) در نظر گرفت که عملا نقش یک عامل تنظیم‌گر را ایفا می‌کند. رویکرد پیشنهادی، ضمن حفظ ویژگی ناوردایی نسبت به بازپارامترسازی و تبدیلات دوگان، تعمیمی هندسی از پیشین جفریز ارائه می‌دهد.بررسی‌های نظری نشان می‌دهد که در مدل‌های آماری که نرم تانسور آماری–چنتسوف در آن‌ها ثابت است، پیشین پیشنهادی دقیقا منطبق بر پیشین جفریز خواهد بود. در نهایت، عملکرد این رویکرد از طریق مطالعات عددی بر روی توزیع‌های برنولی، پواسون و چندجمله‌ای ارزیابی شده است. یافته‌های شبیه‌سازی حاکی از آن است که در برخی موارد، اعمال این تعدیل هندسی، به‌ویژه در نمونه‌هایی با اندازه‌ی کوچک و متوسط، به کاهش واریانس برآوردگرها و بهبود میانگین مربعات خطا منجر می‌شود.

واژه‌های کلیدی: استنباط بیزی،هندسه اطلاعات،پیشین جفریز،تانسور آماری–چنتسوف

*نویسنده مسئول: kimia.bavafa@aut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌بندی توام با اثرات مختلط غیرخطی مبتنی بر رگرسیون چندکی بیزی برای داده‌های طولی و زمان تا وقوع رخداد

مهدیه برنور[۱]*، سیدرضا هاشمی[۱]
[۱]دانشگاه رازی کرمانشاه

چکیده: هنگامی که استنباط آماری برای یک مجموعه داده طولی با خطای اندازه‌گیری با مکان غیرمرکزی، غیرخطی، غیرنرمال و داده‌های گم‌شده و همچنین زمان رخداد با سانسور فاصله‌ای انجام می‌شود، مهم است که به‌طور همزمان این ویژگی‌ها را در نظر گرفت تا نتایج استنباطی درست‌تری حاصل شود. در این مقاله رویکرد مدل‌سازی توام بیزی را برای تحلیل یک مجموعه داده بالینی مربوط به بیماری ایدز را در نظر گرفته و عملکرد مدل‌های توام و روش پیشنهادی را ارزیابی کنیم.

واژه‌های کلیدی: مدل‌سازی توام طولی و بقا، مدل اثرات مختلط، مدل زمان شکست شتابیده، رگرسیون چندکی، سانسور فاصله‌ای.

*نویسنده مسئول: mbarnoor@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مقایسه عملکرد الگوریتمهای جنگل تصادفی و ماشین بردار پشتیبان در پیش‌بینی دیابت با استفاده از داده‌های **Pima**

محمد بلبلیان قالیباف[۱*]
[۱]دانشگاه حکیم سبزواری

چکیده: دیابت از شایع‌ترین بیماری‌های متابولیک است که تشخیص زودهنگام آن نقش مهمی در پیشگیری از عوارض دارد. در این مطالعه، عملکرد دو الگوریتم پرکاربرد یادگیری ماشین، یعنی جنگل تصادفی و ماشین بردار پشتیبان با هسته تابع پایه شعاعی، برای پیش‌بینی ابتلا به دیابت بر روی مجموعه داده استاندارد Pima ارزیابی و مقایسه شد. ارزیابی با استفاده از اعتبارسنجی متقاطع ۱۰-تایی لایه‌بندی شده و معیارهای دقت، حساسیت، ویژگی، امتیاز F_1 و سطح زیر منحنی ROC (AUC) انجام گرفت. نتایج نشان داد که الگوریتم ماشین بردار پشتیبان با AUC معادل 0.837 ± 0.053 عملکردی کمی بهتر از الگوریتم جنگل تصادفی با AUC معادل 0.829 ± 0.038 دارد، اما آزمون t زوجی اختلاف معناداری آماری نشان نداد. در داده های مورد بررسی مدل جنگل تصادفی مزیت تفسیرپذیری بیشتری از طریق اهمیت ویژگی‌ها دارد که نشان داد گلوکز، BMI، سن و تابع سابقه دیابت مهم‌ترین پیش‌بینی‌کننده‌ها هستند. به طور کلی، هر دو روش برای غربالگری دیابت مناسب بوده و انتخاب بین آنها به نیاز بین تفسیرپذیری و دقت نهایی بستگی دارد.

واژه‌های کلیدی: ماشین بردار پشتیبان،جنگل تصادفی،داده‌های دیابت Pima

*نویسنده مسئول: m.bolbolian@hsu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

برآورد بیزی پارامتر تنش–مقاومت توزیع لیندلی‌توانی تحت داده‌های سانسورشده هیبرید نوع‌دوم با استفاده از نمونه‌گیری مجموعه‌رتبه‌دار

نسرین بلوچ رودباری[۱*]، ایمان مخدوم[۲]، مهدی بسی خواسته[۳]، محمدرضا قلانی[۱]
[۱]دانشگاه آزاد واحد اهواز
[۲]دانشگاه پیام نور
[۳]دانشگاه آزاد واحد دزفول

چکیده: در این مقاله، برآورد بیزی قابلیت اطمینان تنش–مقاومت $R = P(Y < X)$ برای توزیع لیندلی‌توانی تحت داده‌های سانسورشده هیبرید نوع دوم و با استفاده از نمونه‌گیری مجموعه‌رتبه‌دار بررسی‌می‌شود. فرض می‌شود متغیرهای مقاومت X و تنش Y مستقل بوده و از توزیع لیندلی توانی با پارامترهای α و β پیروی می‌کنند. ابتدا تابع درستنمایی متناظر با طرح نمونه‌گیری مجموعه‌رتبه‌دار و سانسور هیبرید نوع دوم استخراج شده و سپس برآوردگرهای ماکسیمم درستنمایی پارامترها و قابلیت اطمینان R به‌دست آمده اند. در چارچوب بیزی، با در نظر گرفتن پیشین‌های گاما، توزیع پسین مشترک پارامترها تشکیل شده است و به دلیل پیچیدگی محاسباتی، الگوریتم متروپولیس–هستینگز برای تولید نمونه‌های پسین به‌کار رفت. برآوردگر بیزی R تحت تابع زیان مربع خطا محاسبه شده است. عملکرد روش پیشنهادی از طریق مطالعه شبیه‌سازی ارزیابی گردد. نتایج نشان می‌دهد که ترکیب نمونه‌گیری مجموعه‌رتبه‌دار با رویکرد بیزی، دقت برآورد قابلیت اطمینان را در حضور داده‌های سانسورشده افزایش می‌دهد.

واژه‌های کلیدی: قابلیت اطمینان تنش–مقاومت ، توزیع لیندلی توانی، سانسور هیبرید نوع دوم، نمونه گیری مجموعه رتبه دار، برآورد بیزی، الگوریتم متروپولیس–هستینگز

---

*نویسنده مسئول: n.balochroodbari@iau.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

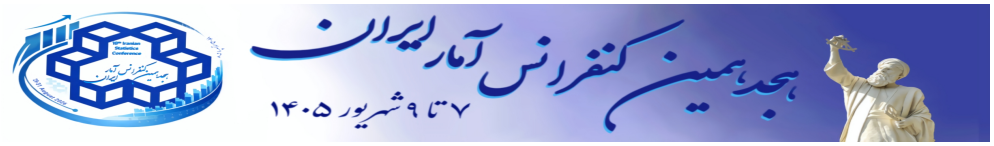
مدل‌سازی داده‌های بقا ناهمگن با مدل عمیق آمیخته کاکس

سمیرا بلوچی[۱*]، کیومرث مترجم[۱]
[۱]دانشگاه تربیت مدرس

چکیده: مدل‌سازی داده‌های بقا به دلیل وجود ویژگی‌های خاص مانند چولگی، سانسور و ناهمگنی جمعیت، اغلب با چالش‌هایی به ویژه در داده‌های پیچیده و بعد بالاروبرو است. از طرفی بکارگیری روش‌های کلاسیک در تحلیل بقا با محدودیت‌هایی مانند نیاز به فرضیات توزیعی یا برقراری فرض مخاطرات متناسب رو به رو هستند. این مقاله با هدف ارزیابی عملکرد روش‌های کلاسیک و مدل‌های یادگیری عمیق در پیش‌بینی زمان بقا انجام شده است. یافته‌ها نشان می‌دهند که مدل‌های یادگیری عمیق با توانایی ذاتی در پردازش داده‌های پیچیده و غیرساختاریافته، از دقت پیش‌بینی بالاتری برخوردارند. این مدل‌ها با توانایی یادگیری ویژگی‌های پنهان و الگوهای پیچیده در داده‌ها، به بهبود دقت پیش‌بینی زمان وقوع رویدادها کمک شایانی می‌کنند. نتایج بیشتر نشان می‌دهد اگرچه مدل‌های کلاسیک از نظر تفسیرپذیری و انطباق با داده‌های واقعی قابلیت بالایی دارند، ولی در تشخیص دقیق زیرگروه‌ها در داده‌های بقای ناهمگن عملکرد ضعیف‌تری دارند. در مقابل، مدل‌های بقای عمیق در تمایزگذاری بین گروه‌ها موفق‌ترند با این حال، پیچیدگی این مدل‌ها، نیاز به داده‌های حجیم و با کیفیت بالا محدودیت‌های اصلی آنها است.

واژه‌های کلیدی: تحلیل بقا،مدل مخاطرات متناسب،شبکه‌های عصبی،یادگیری عمیق

*نویسنده مسئول: s.balochi@modares.ac.ir



مقایسه‌ی کوتاه سه برآورد بازه‌ای با تاکید بر بازه‌های درست‌نمایی

امیررضا بهرامی[۱]، ترانه بنا[۱]، سید محمود طاهری[۱*]
[۱]دانشگاه تهران

چکیده: سه رویکرد برآورد بازه‌ای (بازه اطمینان، بازه اعتبار و بازه درست‌نمایی) مرور و بررسی می‌شوند. توضیح داده می‌شود که بازه‌های درست‌نمایی به دلیل ناوردایی (پایایی)، عدم نیاز به توزیع پیشین، تفسیر مبتنی بر شواهد، سادگی و منطق مبتنی بر تابع درست‌نمایی برتری‌های قابل‌توجهی نسبت به روش‌های فراوانی‌گرا و بیزی دارند. این سه رویکرد برپایه‌ی مثال‌های عددی مرتبط با برآورد بازه‌ای برای میانگین توزیع نرمال مقایسه می‌شوند.

واژه‌های کلیدی: برآورد بازه‌ای،استنباط بیزی،رویکرد درست‌نمایی،بازه اطمینان،بازه اعتبار

---

*نویسنده مسئول: sm_taheri@ut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تبیین مؤلفه‌های سواد آماری مدرسه‌ای

احسان بهرامی سامانی[۱*]
[۱]دانشگاه شهید بهشتی

چکیده: سواد آماری یکی از موضوعات مهم در حوزه‌ی آموزش آمار مدرسه‌ای محسوب می‌شود و نقش بسیار مهمی را در تربیت دانش‌آموزان برای درک، تفسیر و نقد انتقادی اطلاعات آماری را ایفا می‌کند. تبیین مؤلفه‌های سواد آماری مدرسه‌ای می‌تواند تصمیم‌سازان را درحوزه آموزش ریاضیات مدرسه‌ای به‌منظور تولید محتواهای مناسب برای رسیدن به هدف سواد آماری یاری نماید. در این راستا، با بهره‌گیری از رویکرد فراترکیب کیفی، پژوهش‌های انجام‌شده در بازه سال‌های ۱۹۹۰ تا ۲۰۲۲ مورد بررسی و تحلیل قرار گرفته‌اند. یافته‌ها نشان می‌دهد که مؤلفه‌های اصلی سواد آماری مدرسه‌ای شامل داده‌محوری و سواد نموداری: شناخت نیاز به داده، جمع‌آوری، بازنمایی و تبدیل انواع داده‌ها و داستان‌سرایی داده، احتمال و تعبیر شانس در پدیده‌های تصادفی، تغییرپذیری: درک تغییرات ذاتی در پدیده‌ها، استنباط آماری شامل کاربرد مدل‌های آماری، تحلیل نمودارها و تعمیم نتایج از نمونه به جامعه، و تحلیل زمینه‌ای شامل تلفیق مسئله آماری با بستر و زمینه آن است. همچنین، نتایج بر لزوم بازنگری در برنامه‌های درسی آمار، ارتقای دانش و مهارت معلمان و ایجاد استانداردهای آموزشی متناسب با نیازهای روز تأکید دارد.

واژه‌های کلیدی: آموزش آمار

*نویسنده مسئول: e_bahrami@sbu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌سازی متغیرهای مؤثر در بازگشت بیماری اسکیزوفرنیا با استفاده از مدل اندرسن-گیل برای پیشامدهای بازگردنده (مطالعه موردی بیماران بستری در مرکزآموزشی-درمانی فارابی کرمانشاه)

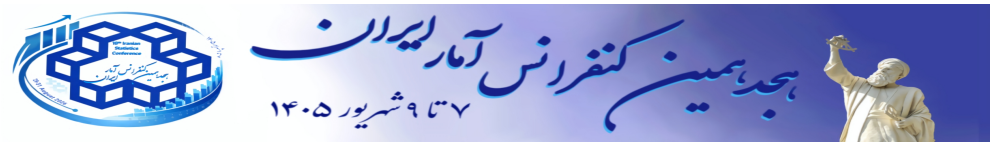
آزاده بهروش[۱*]، سیدرضا هاشمی[۲]، عمران داوری نژاد[۳]
[۱]دانشگاه رازی
[۲]دانشگاه رازی کرمانشاه
[۳]دانشگاه علوم پزشکی کرمانشاه

چکیده: در بررسی عوامل مؤثر بر وقوع بازگشت بیماری اسکیزوفرنیا، متغیرهای زیادی می‌بایست مورد کنکاش قرار گیرند، اما در مطالعات قبلی با نوعی پیش‌داوری تعداد کمی از این عوامل در مدل‌ها واردشده‌اند. در این مطالعه که بر روی بیماران باسابقه بستری در مرکز آموزشی- درمانی فارابی کرمانشاه انجام‌شده است ابتدا با استفاده از الگوریتم مختصات کاهشی در مدل اندرسون-گیل که برای مدل‌های پیشامدهای بازگردنده به کار می‌رود تعداد متغیرهای موجود در مطالعه را کاهش داده سپس مدل بهینه را برای این متغیرها برازش داده‌ایم. نتایج حاکی از تأثیر معنی‌دار متغیرهای تعداد بازگشت‌ها، تعداد شوک‌های الکتریکی، شوک الکتریکی، سابقه خودکشی، شغل بیمار، جنسیت بیمار و نحوه آغاز بیماری پرمخاطره وقوع بازگشت بیماری اسکیزوفرنیا می‌باشد.

واژه‌های کلیدی: پیشامدهای بازگردنده، مدل اندرسون- گیل، انتخاب متغیر، مدل رگرسیون جریمه‌دار، الگوریتم مختصات کاهشی، بازگشت بیماری اسکیزوفرنیا.

---

*نویسنده مسئول: azibehravesh1400@gmail.com



مروری بر چند راهکار نوین برای انتخاب متغیر در تحلیل داده‌های بعد بالا

رها بهزادی فرد[۱*]، موسی گلعلی زاده[۱]
[۱]دانشگاه تربیت مدرس

چکیده: بسیاری از مجموعه داده‌ها در دنیای امروز همانند تصاویر، سیگنال‌های پردازش گفت‌وگو و به ویژه داده‌های ژنتیکی طبیعتی بعد بالا دارند، بدان معنا که در این مجموعه‌ها تعداد متغیرها اغلب بسیار بیشتر از نمونه‌های موجود است. تحلیل داده‌های بعد بالا نیازمند روش‌های مختص به خود است، زیرا روش‌های کلاسیک در مواجهه با چنین داده‌هایی دارای ایرادات نظری هستند. بهره‌گیری از اصل تنکی یکی از راه حل‌های موجود برای برخورد با چالش‌های ناشی از بعد بالا بودن داده‌ها است. بر اساس این اصل، فرض می‌شود که تنها نسبت معقولی از متغیرهای پیشگو با متغیر پاسخ مرتبط بوده و به بیان دیگر اغلب متغیرهای پیشگو دارای ضریب صفر هستند. افزایش سرعت محاسباتی و بهبود تفسیرپذیری از مزایای استفاده از اصل تنکی است. روش‌های مختلفی برای ایجاد مدل‌های تنک وجود دارند. با این حال، تعدادی از رویکردهای محبوب تنک‌سازی مانند لاسو برخلاف کاربرد گسترده خود در برخورد با همبستگی میان متغیرها دارای ضعف هستند. بنابراین، روش‌های نوین به دنبال رفع این چالش هستند. در این پژوهش چند رویکرد انتخاب متغیر با تمرکز بر همبستگی میان متغیرها مرور شده، سپس مقایسه عملکرد آن‌ها با کاربست این روش‌ها بر روی یک مجموعه داده ژنتیکی برای پیشگویی سطح کلسترول به کمک متغیرهای مهم صورت می‌گیرد که در قیاس با تعداد متغیرهای اصلی بسیار کمتر هستند.

واژه‌های کلیدی: داده‌های بعد بالا،انتخاب متغیر،لاسو،تورکشسان،ژن

*نویسنده مسئول: raha.behzadifard@modares.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تحلیل رفتار کاربران در کامنت های فارسی بسترهای خبری با تمرکز بر انواع محتوا و مدل های ترنسفورمر

یونس بیات[۱*]، حسین قانعی بافقی[۱]
[۱]دانشگاه میبد

چکیده: تحلیل رفتار کاربران در بستر شبکه‌های اجتماعی و پلتفرم‌های محتوایی، به‌ویژه در زبان فارسی یکی از موضوعات مهم در حوزه پردازش زبان طبیعی و داده‌کاوی است. در این پژوهش، ۵۰۱،۶۱ رکورد زنده شامل کامنت‌های کاربران و پست‌های فارسی که از طریق یک وب‌سرویس جمع‌آوری شده‌است مورد تحلیل قرار گرفته که داده‌ها شامل چهار نوع محتوا در ستون media_type، شامل post، photo، video و text هستند و نقش آن‌ها در الگوهای رفتاری کاربران بررسی شده است. برای سنجش شباهت معنایی کامنت‌ها از یک مدل چندزبانه و برای تحلیل رفتار کاربران از شاخص‌هایی نظیر تعداد کامنت‌ها، نرخ تکراری بودن و زمان غالب فعالیت استفاده شده است. نتایج نشان می‌دهد که کاربران از نظر سطح فعالیت در سه گروه کم، متوسط و زیاد قرار می‌گیرند و فعالیت آن‌ها در ساعات مشخصی از شبانه‌روز متمرکز است. همچنین نوع محتوا و زمان فعالیت می‌تواند بر الگوهای نوشتاری و میزان تکرار کامنت‌ها تأثیرگذار باشد.

واژه‌های کلیدی: رفتار کاربران ، پردازش زبان طبیعی ، یادگیری ماشین ، انواع محتوا ، تحلیل کامنت ، بسترهای خبری

*نویسنده مسئول: bayat@stu.meybod.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

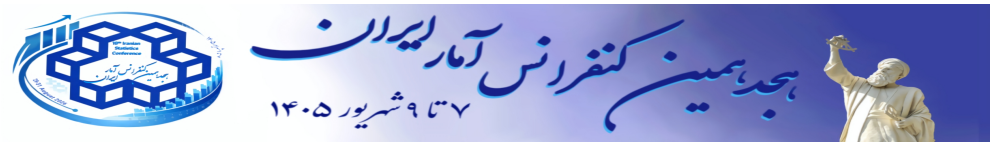
کاربرد روش های یادگیری ماشین و یادگیری ژرف برای دسته بندی هوشمند اخبار تلگرام و مقایسه آن ها

یونس بیات[۱*]، حسین قانعی بافقی[۱]
[۱]دانشگاه میبد

چکیده: با گسترش پیام رسان ها تولید اخبار کوتاه و نویزی افزایش یافته است این پژوهش سه مدل Logistic Regression ، SVM و Lstm را برای طبقه بندی اخبار تلگرام در سه کلاس (سیاسی/ اقتصادی، اجتماعی/ حوادث، جهان/ سایر) مقایسه می‌کند. داده‌ها از یک کانال خبری واقعی با وب سرویس جمع آوری شده و ۲۰۰ نمونه برچسب گذاری دستی شدند و توزیع کلاس ها نامتوازن است (۵۹% سیاسی ، ۳۲% اجتماعی ، ۹% جهان) که نتایج نشان می‌دهد مدل SVM با دقت ۸۵۳۶/۰ و نمره=۴۴۳۴/۰ نسبت به رگرسیون لجستیک دقت ۸۳۰۵/۰ ،نمره = ۳۳۰۷/۰ برتری دارد LSTM . بدون اعتبارسنجی متقاطع دقت ضعیف (حدود ۶۵/۰) دارد اما با تکرار ۱۰ تایی به دقت متوسط ۷۴۳۹/۰ و نمره وزنی =۷۱۵۵/۰ می رسد که همچنان از SVM پایین تر است در نتیجه برای مجموعه داده کوچک و نامتوازن تلگرام SVM گزینه قابل اعتمادتری است.

واژه‌های کلیدی: اخبار تلگرام،LSTM،SVM،LR،طبقبندی متن،عدم توازن کلاس

*نویسنده مسئول: bayat@stu.meybod.ac.ir



برآورد عدم قطعیت نمای هارست در سری‌های زمانی با حافظه بلندمدت: رویکرد بوت‌استرپ بلوکی

معصومه بی باک[۱*]، احمد نادری بنی[۱]، سجاد مری[۲]، بهروز میرزا[۲]

[۱]دانشگاه ملی مهارت

[۲]دانشگاه صنعتی اصفهان

چکیده: نمای هارست (Hurst exponent) شاخصی کلیدی برای تشخیص حافظه بلندمدت در سری‌های زمانی است. با این حال، ارزیابی عدم قطعیت آن کمتر مورد توجه قرار گرفته است. در این پژوهش، روش بوت‌استرپ بلوکی ایستا برای محاسبه فاصله اطمینان نمای هارست (برآورد شده با روش DFA) ارائه می‌شود. اعتبارسنجی با شبیه‌سازی ARFIMA(0,d,0) نشان می‌دهد که نرخ پوشش فواصل اطمینان ۹۵% نزدیک به سطح اسمی است (۹۲%-۹۴%). کاربرد روش در داده‌های دمای هفتگی شهرکرد (2003-2025) نشان می‌دهد که پس از حذف روند و فصلی، نمای هارست برابر 0.474 با فاصله اطمینان ۹۵% صدکی [0.294، 0.683] و BCa [0.320، 0.729] است که حاکی از عدم وجود حافظه بلندمدت معنادار است. روش ارائه‌شده می‌تواند در تحلیل سری‌های مالی، اقلیمی و زیستی به کار رود.

واژه‌های کلیدی: نمای هارست، تحلیل نوسان‌زدایی، بوت‌استرپ بلوکی، فاصله اطمینان، حافظه بلندمدت، داده‌های اقلیمی.

---

*نویسنده مسئول: mbibak@tvu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تحلیل علیت گرنجر کسری و همبستگی جزئی روندزدایی شده در داده‌های اقلیمی شهرکرد (1403-1375)

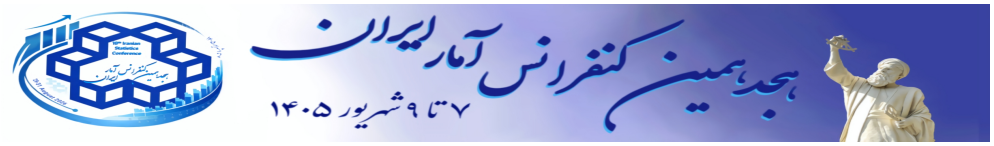
معصومه بی باک[۱*]، احمد نادری بنی[۱]، سجاد مری[۲]، بهروز میرزا[۲]
[۱]دانشگاه ملی مهارت
[۲]دانشگاه صنعتی اصفهان

چکیده: در این پژوهش، رابطه علی و همبستگی ذاتی بین سرعت باد و دمای میانگین روزانه شهرکرد در بازه ۱۹۹۶ تا ۲۰۲۴ با رویکردهای نوین علیت گرنجر کسری و همبستگی جزئی روندزدایی شده (DPCCA) مورد تحلیل قرار گرفته است. با استفاده از داده‌های روزانه، سری هفتگی تمیز (۳۵۷ هفته معتبر) ساخته شد و دو دوره 2010-1996 (۳۹ هفته) و 2024-2011 (۳۱۸ هفته) تفکیک گردید. نتایج علیت گرنجر کسری نشان می‌دهد که در دوره اول، تنها علیت یک‌طرفه از دما به باد وجود دارد (0.0272=p-value)، در حالی که در دوره دوم، علیت دوطرفه معناداری بین باد و دما مشاهده می‌شود (p-value به ترتیب 0.0108 و 0.0141). همچنین تحلیل DPCCA با حذف اثر فشار هوا نشان داد که همبستگی جزئی باد–دما در مقیاس‌های کوتاه (کمتر از ۳۰ هفته) مثبت و قوی (نزدیک به ۱) و در مقیاس‌های بلند (بیشتر از ۳۰ هفته) به شدت منفی (نزدیک به ۱-) است. این یافته‌ها بیانگر تغییر رژیم اقلیمی منطقه و وابستگی شدید رابطه باد–دما به مقیاس زمانی است.

واژه‌های کلیدی: علیت گرنجر کسری، همبستگی جزئی روندزدایی شده، داده‌های اقلیمی شهرکرد، تغییر اقلیم، سری‌های زمانی فرکتالی.

---

*نویسنده مسئول: anaderi@tvu.ac.ir



رویکردی مبتنی بر ژرفای داده برای رتبه‌بندی چندمتغیره داوطلبان در آزمون سراسری ایران

سمانه تات[۱]، مریم پارسائیان[۲]، ابراهیم خدائی[۲]، مه سیما اربابی[۲*]

[۱]دانشگاه شاهد

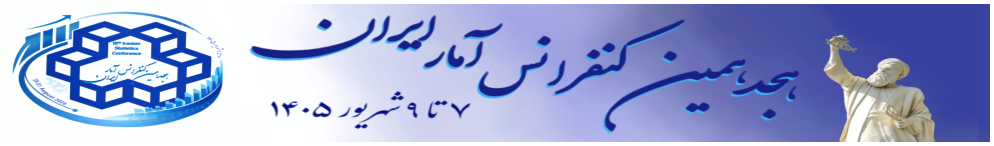
[۲]دانشگاه تهران

چکیده: رتبه‌بندی داوطلبان در آزمون‌های چندموضوعی از چالش‌های اصلی تحلیل داده‌های آموزشی چندمتغیره است. روش‌های سنتی با تقلیل عملکرد چندبعدی به شاخص‌های تجمیعی یا میانگین‌های وزنی، اغلب ناهمگنی و همبستگی میان نمرات را نادیده می‌گیرند. این پژوهش چارچوبی نوین مبتنی بر ژرفای داده برای رتبه‌بندی منصفانه و مقاوم در ارزیابی‌های بزرگ‌مقیاس ارائه می‌کند. داده‌های 110550 داوطلب گروه ریاضی-فیزیک آزمون سراسری ایران (INUEE، سال ۱۴۰۰) شامل نمرات دروس و معدل دبیرستان تحلیل شد. در این چارچوب، تحلیل مولفه‌های اصلی همراه با برآوردگر قوی حداقل کوواریانس (MCD) جهت محاسبه ژرفای ماهالانوبیس به‌کار گرفته شد و با یک نگاشت مختصات استوانه‌ای به‌منظور افزایش قدرت تفکیک ترکیب گردید. نتایج نشان داد دو مولفه اصلی نخست، بخش عمده واریانس کل را تبیین کرده و رتبه‌بندی حاصل همبستگی اسپیرمن بالای ۰/۹۱۲- با رتبه‌بندی رسمی با ثبات بیشتر (0.994) و گره‌خوردگی کمتر دارد. همچنین همپوشانی دهک برتر در دو رتبه‌بندی به ۹۰/۹۶درصد رسید. این یافته‌ها بیانگر آن است که رویکرد مبتنی بر ژرفای داده، دقت و عدالت رتبه‌بندی را در آزمون‌های رقابتی ارتقا داده و قابلیت تعمیم به ارزیابی‌های گسترده‌ای چون SAT و Gaokao را داراست.

واژه‌های کلیدی: ژرفای داده،رتبه‌بندی چندمتغیره،تابع ژرفای ماهالانوبیس،آزمون‌های مقیاس بزرگ

---

*نویسنده مسئول: mahsima.arbabi@ut.ac.ir



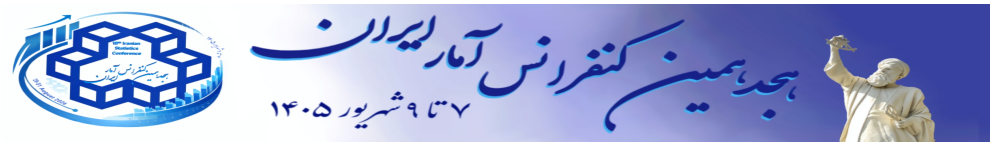
شبیه‌سازی مونت‌کارلو در پایتون برای تحلیل فرایندهای تصادفی و داده‌های بازار سهام

محبوبه تاتاری[۱*]
[۱]دانشگاه پیام نور

چکیده: فرایندهای تصادفی به عنوان ساختارهای بنیادی در نظریه امار و احتمال، نقش مهمی در مدل‌سازی پدیده های همراه با عدم‌قطعیت در حوزه‌های مختلف علمی و مهندسی ایفا می‌کنند. کاربرد روش شبیه‌سازی مونت‌کارلو در پایتون به عنوان یک ابزار کارآمد آموزشی و تحلیلی برای درک و مدل‌سازی پدیده‌های تصادفی است. محیط پایتون، به دلیل دارا بودن کتابخانه‌های قدرتمند و با کیفیت برای محاسبات آماری، بستری مناسب برای شبیه‌سازی آزمایش‌های احتمالاتی فراهم می کند. در این مقاله فرایندهای تصادفی پرتاب سکه، پرتاب دو تاس و بازده بازار سهام بر اساس داده‌های واقعی شاخص بازار سهام ایران مورد بررسی قرار گرفته است. این پژوهش بر نقش مهم پایتون به‌عنوان ابزاری قدرتمند در آموزش تحلیل‌های آماری و مدل‌سازی پدیده های مالی و تصادفی تأکید دارد.

واژه‌های کلیدی: فرایندهای تصادفی، شبیه‌سازی مونت‌کارلو، پایتون، مدل‌سازی تصادفی، بازار سهام ایران

*نویسنده مسئول: m-tatari@pnu.ac.ir



نقش آمار رسمی در آینده مطالعات جمعیتی ایران در عصر هوش مصنوعی

فاطمه ترابی [۱*]
[۱]دانشگاه تهران

چکیده: آمار رسمی همواره زیربنای مطالعات جمعیتی بوده و بخش زیادی از تحلیل‌های مربوط به باروری، مرگ‌ومیر، ازدواج، طلاق و مهاجرت بر پایه داده‌های حاصل از سرشماری‌ها، ثبت وقایع حیاتی و طرح‌های آماری استوار است. در سال‌های اخیر، گسترش فناوری‌های دیجیتال، مه‌داده‌ها و هوش مصنوعی افق‌های تازه‌ای برای تولید و تحلیل این داده‌ها گشوده است. این مقاله با هدف تبیین نقش آمار رسمی در آینده مطالعات جمعیتی ایران در عصر هوش مصنوعی و بررسی ظرفیت‌های نوین نظام آماری کشور، به‌صورت مرور تحلیلی ادبیات و بررسی تجربیات موجود انجام شده است. یافته‌ها نشان می‌دهد که بهره‌گیری از الگوریتم‌های یادگیری ماشین می‌تواند برخی محدودیت‌های سنتی نظام آماری، از جمله فاصله زمانی انتشار داده‌ها و دشواری تحلیل‌های پیچیده را کاهش دهد. ترکیب داده‌های ثبت احوال با روش‌های هوشمند در پیش‌بینی روندهای باروری و سن ازدواج، کاربرد مدل‌های یادگیری ماشین در تحلیل الگوهای اقتصادی و مهاجرتی طرح‌های آماری و نقش هوش مصنوعی در گذار به سرشماری‌های ثبتی‌مبنا از جمله ظرفیت‌های مهم این تحول است. با وجود این، چالش‌هایی همچون کیفیت و یکپارچگی داده‌ها، محدودیت دسترسی، کمبود زیرساخت‌های فنی و ملاحظات اخلاقی همچنان پابرجاست. در مجموع، آینده مطالعات جمعیتی ایران وابسته به هم‌افزایی میان آمار رسمی، زیرساخت‌های داده‌ای و ابزارهای هوش مصنوعی است.

واژه‌های کلیدی: آمار رسمی،مطالعات جمعیتی،هوش مصنوعی

[*]نویسنده مسئول: fatemeh_torabi@ut.ac.ir

واکاوی بحران آب کشور

پریا ترابی کهلان[1*]
[1]پژوهشکده‌ی آمار

چکیده: بحران آب در کشور به یکی از مهم‌ترین چالش‌های راهبردی در دهه‌های اخیر تبدیل شده است. کاهش بارندگی، افزایش دما و تداوم خشکسالی‌های بلندمدت از یک‌سو، و رشد جمعیت، توسعه‌ی صنعت و اراضی کشاورزی، افزایش مصرف و هدررفت آب، الگوهای ناپایدار کشت و ضعف حکمرانی آب از سوی دیگر، فشار بی‌سابقه‌ای بر منابع آبی کشور وارد کرده‌اند. بحران آب موضوعی چندبعدی است که مدیریت آن نیازمند نگاه جامع، هماهنگ و مبتنی بر شواهد است. در چنین شرایطی، بازنگری در سیاست‌ها و رویکردهای موجود و حرکت به‌سوی مدیریت پایدار منابع آب ضرورتی اجتناب‌ناپذیر است. اجرای برنامه‌های مدیریت یکپارچه‌ی منابع آب، اصلاح حکمرانی آب و تقویت نظام‌های نظارتی، ایجاد هماهنگی بیش‌تر در سیاست‌ها، ارتقای بهره‌وری در بخش کشاورزی، توسعه‌ی طرح‌های آبخیزداری بخشی از راهکارهای کلیدی برای مواجهه با بحران آب کشور به شمار می‌آیند.

واژه‌های کلیدی: بحران آب،توسعه‌ی پایدار،تنش آبی،راهکارهای مدیریت منابع آب

*نویسنده مسئول: torabiparya@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

پیش‌بینی مصرف گاز طبیعی با رویکرد ترکیبی Prophet، XGBoost و Bootstrap
مطالعه موردی: استان تهران

مهرانه تیراندازی[۱*]، امیر حسین فاکهی[۱]
[۱]موسسه مطالعات بین‌المللی انرژی

چکیده: پیش‌بینی دقیق مصرف گاز طبیعی به‌ویژه در بخش خانگی، یکی از چالش‌های اساسی در برنامه‌ریزی انرژی ایران است. در این پژوهش، با استفاده از داده‌های ماهانه مصرف گاز خانگی شهرهای استان تهران در بازه زمانی ۱۳۹۳ تا ۱۴۰۱، عملکرد مدل Prophet (متا/فیسبوک) به‌عنوان یک مدل سری‌زمانی خودکار، با مدل ترکیبی Prophet+XGBoost+Bootstrap مقایسه شده است. نتایج نشان می‌دهد که مدل ترکیبی ضمن کاهش میانگین خطای مطلق درصدی (MAPE) در اغلب شهرها (برای نمونه شهر تهران از ۱۹.۸۵% به ۱۳.۱۳%)، بازه عدم قطعیت پیش‌بینی را نیز به‌طور معناداری کاهش می‌دهد. ویژگی مهم این روش ترکیبی، عدم افزایش بازه عدم قطعیت با افزایش افق زمانی پیش‌بینی (تا ۱۰ سال) است که آن را برای برنامه‌ریزی بلندمدت انرژی مناسب می‌سازد. روش ارائه‌شده قابلیت تعمیم به کل استان‌ها و سطح ملی را دارد.

واژه‌های کلیدی: پیش‌بینی مصرف گاز، Prophet، XGBoost، Bootstrap، مدل ترکیبی، سری‌های زمانی، یادگیری ماشین، استان تهران، علم داده، کلان داده

---

*نویسنده مسئول: mehraneh.tirandari@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

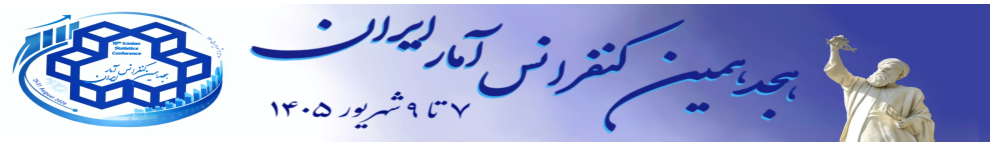
پیش‌بینی داده‌های فضایی-زمانی پیچیده با استفاده از شبکه‌های عصبی عمیق

فاطمه ثانی[۱*]، فاطمه حسینی[۱]، امید کریمی[۱]
[۱]دانشگاه سمنان

چکیده: پیش‌بینی داده‌های فضایی-زمانی یکی از مسائل بنیادین در آمار کاربردی، علوم محیطی و تحلیل داده‌های وابسته است. روش‌های کلاسیک زمین‌آماری، به‌ویژه کریگینگ به دلیل ویژگی بهترین پیش‌بین خطی نااریب، از ابزارهای متداول برای برآورد مقادیر در نقاط مشاهده‌نشده فضایی-زمانی محسوب می‌شوند. با این حال، در بسیاری از کاربردهای واقعی که ساختار وابستگی داده‌ها پیچیده، غیرخطی یا دور از فروض کلاسیک مانند نرمال بودن و ایستایی است، عملکرد این روش‌ها ممکن است با محدودیت همراه باشد. در این پژوهش، یک چارچوب پیش‌بینی نوین مبتنی بر شبکه‌های عصبی عمیق برای داده‌های فضایی-زمانی ارائه می‌شود که با بهره‌گیری از قابلیت یادگیری الگوهای پیچیده، وابستگی‌های غیرخطی موجود در داده را بهتر مدل‌سازی می‌کند. همچنین علاوه بر متغیرهای کمکی، از توابع پایه فضایی-زمانی برای غنی‌سازی فضای ویژگی استفاده شده و ساختار یادگیری به‌صورت ترکیبی طراحی شده است تا ظرفیت مدل در تقریب روابط پیچیده افزایش یابد. به‌منظور ارزیابی تجربی، عملکرد مدل بر روی یک مجموعه داده شبیه‌سازی بررسی و با روش کریگینگ کلاسیک مقایسه شده است. اعتبارسنجی مدل با استفاده از راهبرد مناسب داده‌های وابسته فضایی-زمانی انجام شده و نتایج نشان می‌دهد که روش پیشنهادی، به‌ویژه در داده‌های پیچیده و غیرگاوسی، می‌تواند دقت پیش‌بینی بالاتری نسبت به روش‌های کلاسیک ارائه دهد.

واژه‌های کلیدی: داده‌های فضایی-زمانی، کریگینگ فضایی-زمانی، بهترین پیش‌بین خطی نااریب، شبکه‌های عصبی عمیق، نظریه یادگیری آماری، تقریب تابع.

*نویسنده مسئول: sanifatemeh04@semnan.ac.ir, fatemeh.hoseini@semnan.ac.ir



تحلیل اثر شوک‌های نرخ ارز و تورم بر هزینه تولید محصولات کشاورزی در ایران با رویکرد رگرسیون **PLS** و توابع کنش-واکنش آنی (مورد مطالعه: محصول برنج، گندم آبی و جو آبی)

محمد جریره[1*]

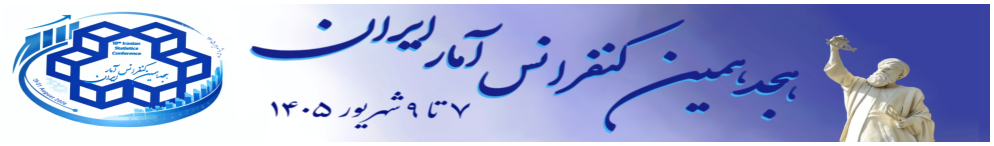
[1]مرکز آمار وزارت جهاد کشاورزی

چکیده: این پژوهش با هدف بررسی اثر شوک‌های نرخ ارز و تورم بر هزینه تولید محصولات کشاورزی (برنج، گندم آبی و جو آبی) در ایران طی سال‌های ۱۳۸۱ تا ۱۴۰۱ انجام شده است.برای تحلیل داده‌ها از روش رگرسیون حداقل مربعات جزئی با اثرات ثابت، توابع کنش-واکنش آنی و تجزیه واریانس استفاده شده است. نتایج رگرسیون نشان می‌دهد که تورم و نرخ ارز تأثیر مثبت و معناداری بر هزینه تولید دارند. ضریب تعیین مدل 0.75 است که بیانگر تبیین ۷۵ درصد از تغییرات هزینه تولید توسط متغیرهای مدل می‌باشد. نتایج توابع کنش-واکنش نشان می‌دهد که شوک‌های نرخ ارز و تورم ماهیت دائمی دارند و هزینه تولید پس از شوک به حالت تعادل بازنمی‌گردد، بلکه با گذشت زمان واگراتر می‌شود. شوک نرخ ارز سریع‌ترین واگرایی و شوک تورم واگرایی ملایم‌تر اما پایدار ایجاد می‌کند. همچنین، پیش‌بینی مدل نشان می‌دهد که شوک‌های بزرگ در سال‌های ۱۳۸۱ و ۱۴۰۱ باعث گسترش شدید فاصله اطمینان و افزایش عدم قطعیت در دوره‌های بحرانی شده است. در نتیجه، شوک‌های نرخ ارز و تورم اثر دائمی و واگرا بر هزینه تولید محصولات کشاورزی دارند و یک چرخه معیوب از افزایش هزینه‌ها ایجاد می‌کنند که سیستم را از تعادل دورتر می‌سازد. این یافته‌ها لزوم مدیریت پایدار نرخ ارز و کنترل تورم را برای کاهش فشار بر هزینه‌های تولید در بخش کشاورزی آشکار می‌سازد.

واژه‌های کلیدی: تورم،نرخ ارز،هزینه‌ی تولید،محصولات کشاورزی

---

*نویسنده مسئول: mohamad-jarire@yahoo.com



تحلیل محتوای فصل‌نامه علمی‌پژوهشی اندیشه‌ی آماری طی سال‌های 1401-1398

محمد جریره[۱*]، کیانوش فتحی وارجاگاه[۱]، پروین اژدری[۱]
[۱]دانشگاه آزاد اسلامی واحد تهران شمال

چکیده: هدف این مطالعه، شناخت ویژگی‌های محتوایی مقالات منتشرشده در شماره‌های ۴۷ تا ۵۴ دوفصلنامه‌ی «اندیشه آماری» طی سال‌های ۱۳۹۸ تا ۱۴۰۱ است. جامعه‌ی آماری پژوهش شامل ۹۶ مقاله بوده که با روش تحلیل محتوا بررسی شده‌اند. یافته‌ها نشان می‌دهد اکثریت نویسندگان مرد بوده و بیشتر مقالات به‌صورت گروهی نگارش شده‌اند. از نظر رتبه‌ی علمی، دانشیاران بیشترین سهم مشارکت را داشته‌اند و از نظر وابستگی سازمانی، دانشگاه سمنان و چند دانشگاه دیگر در رتبه‌های نخست قرار دارند. اکثر مقالات با روش شبیه‌سازی یا تحلیل عددی انجام شده و از نظر موضوعی، آمار کاربردی (به‌ویژه برآوردیابی، رگرسیون و نظریه اعتماد) سهم بیشتری نسبت به آمار نظری داشته است.

واژه‌های کلیدی: تحلیل محتوا،نشریه علمی و پژوهشی،مجله اندیشه آماری،مقالات آماری

*نویسنده مسئول: mohamad-jarire@yahoo.com

مطالعه‌ای بر روی خرابی‌های یک سیستم منسجم از زمان نظارت تا زمان خرابی

محمد جریره[۱]*، کیانوش فتحی واجارگاه[۲]، پروین اژدری[۲]
[۱]مرکز امار و فناوری وزارت خانه جهاد کشاورزی
[۲]دانشگاه آزاد اسلامی واحد تهران شمال

چکیده: بردار علامت سیستم‌های منسجم یک ابزار مفید برای محاسبه توابع قابلیت اعتماد سیستم، متوسط طول عمر سیستم و مقایسه سیستم‌های مختلف با استفاده از ترتیب تصادفی می باشد. در این مقاله سعی شده است با استفاده از بردار علامت یک فرمول ترکیبی برای توزیع تعداد خرابی‌های در یک سیستم منسجم از زمان نظارت t تا زمان خرابی سیستم [مادامی که سیستم در زمان t سالم باشد] ارائه شود. بعلاوه نشان داده شد که توزیع احتمال شرطی آن دارای لگ-مقعر است که این نتیجه با یک مثال عددی نیز مورد تایید قرار داده شد. در نهایت امیدریاضی آن در یک سیستم k از n: G محاسبه شده و برخی از ویژگی‌های آن مورد مطالعه قرار گرفته است.

واژه‌های کلیدی: سیستم منسجم، بردار علامت سیستم k از n : F، سیستم k از n : G.

---

*نویسنده مسئول: mohamad_jarire@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

واکاوی تغییرات بی‌نظمی سری‌های زمانی دمای روزانه در دوره همه‌گیری کووید-۱۹ با کاربرد آنتروپی نمونه: مطالعه موردی شهر بیرجند

علیرضا چاجی[1*]
[1]دانشگاه بزرگمهر قائنات

چکیده: پژوهش پیش‌رو با هدف کمی‌سازی تأثیر همه‌گیری کووید-۱۹ بر نظم‌پذیری سری‌های زمانی دمای روزانه، به تحلیل رفتار این سری‌ها در شهر بیرجند طی پنج دوره زمانی متمایز می‌پردازد: دوره الف (دوره قبل از همه‌گیری: ۲۱ اسفند ۱۳۹۷ تا ۲۱ اسفند ۱۳۹۸)، دوره ب (سال اول همه‌گیری: ۲۱ اسفند ۱۳۹۸ تا ۲۱ اسفند ۱۳۹۹)، دوره ج (سال دوم همه‌گیری: ۲۱ اسفند ۱۳۹۹ تا ۲۱ اسفند ۱۴۰۰)، دوره د (سال سوم همه‌گیری: ۲۱ اسفند ۱۴۰۰ تا ۲۱ اسفند ۱۴۰۱) و دوره ه (دوره پساگیر: ۲۱ اسفند ۱۴۰۱ تا ۲۱ اسفند ۱۴۰۲). معیار به‌کاررفته، آنتروپی نمونه است که به دلیل برخورداری از مزایایی چون مقاومت ذاتی در برابر نویز، استقلال نسبی از طول سری و حذف سوگیری خودهمسانی، معیاری کارآمد محسوب می‌شود. یافته‌های کمی حاکی از آن است که مقادیر آنتروپی نمونه در دوره همه‌گیری به طور معناداری کاهش یافته است که حکایت از افزایش نظم‌پذیری داخلی و کاهش مؤلفه تصادفی در نوسانات دمایی دارد. این دستاورد برای مدل‌سازی‌های اقلیمی و پیش‌بینی‌های کوتاه‌مدت دما در منطقه، کاربردهای راهبردی قابل توجهی فراهم می‌آورد.

واژه‌های کلیدی: آنتروپی نمونه، کووید-۱۹، دمای روزانه، بیرجند، سری‌های زمانی، آنتروپی تقریبی.

*نویسنده مسئول: a.chaji@buqaen.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

نظریه اعتمادی

رحیم چینی‌پرداز[۱*]
[۱]گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهیدچمران اهواز

چکیده: در بین ابداعات زیاد فیشر در آمار مدرن، نظریه اعتمادی بیشترین مخالفت‌ها را داشته است. یکی از مهمترین نقدها بر این نظریه نزدیک شدن دیدگاه کلاسیک به اندیشه بیزی است. با این وجود این نقدها، فیشر تا پایان عمر از این نظریه پشتیبانی کرد. در این مقاله ضمن بررسی نظریه اعتمادی و تاثیر آن در استنباط علمی، به پارهای از نقدها و مشکلاتی که با پذیرش آن مواجه می شویم، پرداخته می‌شود.

واژه‌های کلیدی: استنباط اعتمادی ، فلسفه آمار، استقرای علمی

*نویسنده مسئول: chinipardaz_r@scu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

طراحی و قیمت‌گذاری اوراق بهادار بیمه‌ای شاخص‌محور با استفاده از مدل مارکوف‌سوئیچینگ و تابع پرداخت مبتنی بر ضریب خسارت

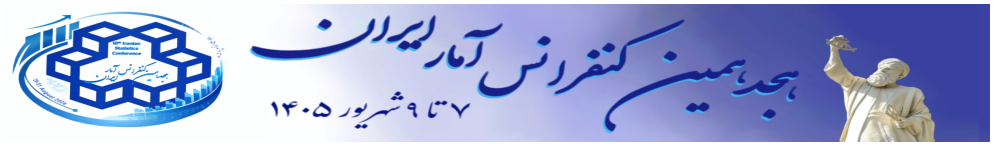
رضا حاجی‌پور فرسنگی[۱]، پوریا گودرز[۲]*، امید نقشینه ارجمند[۱]، وحید محمودی[۲]
[۱]دانشگاه صنعتی امیرکبیر
[۲]دانشگاه تهران

چکیده: این مقاله چارچوبی آماری-اکچوئری برای طراحی و قیمت‌گذاری اوراق بهادار بیمه‌ای شاخص‌محور در رشته نفت و انرژی ارائه می‌کند. شاخص مبنا، ضریب خسارت سالانه این رشته است که با توجه به مثبت بودن، راست‌چولگی و نوسان‌پذیری بالا، با مدل مارکوف‌سوئیچینگ دورژیمی و توزیع‌های مثبت مدل‌سازی می‌شود. توزیع پیش‌بینی سال آینده برای تعیین آستانه‌های پرداخت، تعریف تابع پرداخت و محاسبه حق‌بیمه فنی به کار می‌رود. نتایج نشان می‌دهد مدل‌های وایبول و گاما عملکرد مناسب‌تری دارند و مبنای سناریوی پایه قرار می‌گیرند. همچنین، تابع پرداخت خطی از نظر هزینه برای بیمه‌گر کاراتر است، در حالی که تابع غیرخطی پوشش محافظه‌کارانه‌تری برای خسارت‌های شدید فراهم می‌کند. مقایسه ساختارهای مالی نیز نشان می‌دهد ساختار اصل سرمایه در معرض ریسک ظرفیت پوشش بیشتری دارد، اما ساختار اصل سرمایه محفوظ دامنه حمایت بیمه‌ای را محدود می‌سازد.

واژه‌های کلیدی: اوراق بهادار بیمه‌ای، ضریب خسارت، بیمه نفت و انرژی، مارکوف‌سوئیچینگ، قیمت‌گذاری اکچوئری.

---

*نویسنده مسئول: rehajipour@aut.ac.ir



برآورد پارامتر و انتخاب متغیر در مدل‌های خطی آمیخته سانسور شده: یک مرور روایتی

علی اصغر حائری مهریزی[۱*]، علی شادرخ[۱]، پرویز نصیری[۱]
[۱]دانشگاه پیام نور

چکیده: مدل‌های خطی آمیخته از مهم‌ترین ابزارهای تحلیل داده‌های طولی و خوشه‌ای هستند و در بسیاری از مطالعات پزشکی با داده‌های سانسورشده مواجه می‌شوند. حضور همزمان سانسور و اثرات تصادفی، برآورد پارامترها و انتخاب متغیر را با چالش‌های محاسباتی و نظری همراه می‌کند. این مطالعه به صورت مرور روایتی انتقادی و با بررسی مقالات نمایه‌شده در پایگاه‌های معتبر علمی انجام شد. مطالعات بر اساس نوع مدل، روش برآورد، نوع سانسور و رویکرد آماری طبقه‌بندی و مقایسه شدند. نتایج نشان داد الگوریتم‌های مبتنی بر EM به‌ویژه SAEM در مدل‌های آمیخته سانسورشده از کارایی محاسباتی بالاتری برخوردارند. در انتخاب متغیر، روش‌های تاوانیده مانند LASSO ، Adaptive LASSO، SCAD و MCP عملکرد مناسبی در داده‌های با بعد بالا نشان دادند. همچنین روش‌های بیزی انقباضی نظیر Spike-and-Slab و Horseshoe توانایی مطلوبی در مدیریت عدم‌قطعیت و همخطی متغیرها دارند. با وجود مزایای روش‌های بیزی، هزینه محاسباتی آن‌ها نسبت به روش‌های فراوانی‌گرا بیشتر است. مرور مطالعات نشان داد شواهد تجربی درباره عملکرد این روش‌ها در مدل‌های خطی آمیخته دارای سانسور چپ هنوز محدود است. توسعه مطالعات مقایسه‌ای جامع می‌تواند به طراحی روش‌های کارآمدتر برای انتخاب متغیر در داده‌های پیچیده پزشکی و زیستی منجر شود.

واژه‌های کلیدی: مدل خطی آمیخته، داده‌های سانسورشده، LASSO، Adaptive LASSO، Spike-and-Slab، Horseshoe.

*نویسنده مسئول: a.haeri@acecr.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

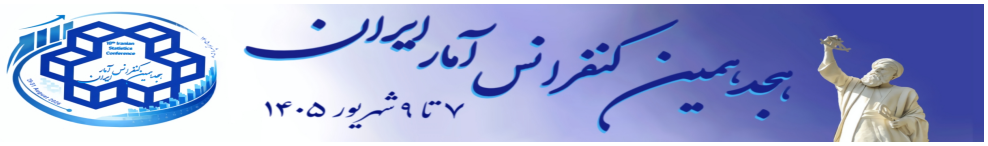
بررسی فضایی-زمانی داده‌های آلودگی هوای تهران با استفاده از میدان عصبی بیزی

فاطمه حسینی[۱*]، امید کریمی[۱]
[۱]دانشگاه سمنان

چکیده: این پژوهش به تحلیل داده‌های فضایی-زمانی آلودگی هوا در شهر تهران با تمرکز بر غلظت ماهانه $PM_{2.5}$ طی دوره ۲۰۱۴ تا ۲۰۲۴ می‌پردازد. داده‌های آلودگی تهران دارای ویژگی‌هایی مانند روندهای غیرخطی، ناایستایی، تفاوت‌های مکانی قابل‌توجه و چولگی مثبت ناشی از رخدادهای آلودگی شدید هستند که کارایی مدل‌های متداول گاوسی را محدود می‌کند. برای مواجهه با این چالش‌ها، یک چارچوب بیزی مبتنی بر میدان عصبی ارائه شده است که در آن ساختار میانگین به‌صورت تابعی از مختصات فضا-زمان مدل‌سازی شده و باقیمانده‌ها از توزیع چوله نرمال بسته منعطف پیروی می‌کنند. نتایج شبیه‌سازی و کاربرد واقعی نشان می‌دهد این مدل در مقایسه با روش‌های مرجع، از جمله میدان‌های عصبی گاوسی دقت پیش‌گویی بالاتری داشته و بازه‌های پیش‌گویی آن پوشش بهتری از عدم‌قطعیت واقعی، به‌ویژه در دوره‌های آلودگی شدید تهران، ارائه می‌دهد. همچنین استفاده از استنباط وارياسیونی باعث شده این چارچوب با وجود پیچیدگی، از نظر محاسباتی کارآمد و قابل‌استفاده برای داده‌های بزرگ شهری باشد.

واژه‌های کلیدی: مدل‌های فضایی زمانی،میدان عصبی بیزی،توزیع چوله نرمال بسته

*نویسنده مسئول: fatemeh.hoseini@semnan.ac.ir



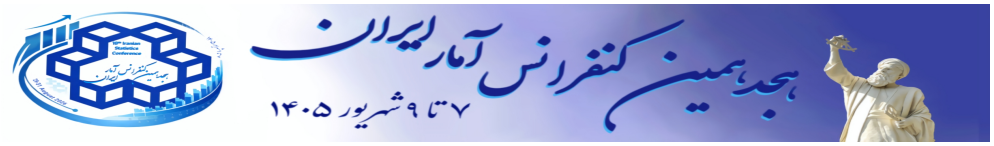
مدل‌سازی ریسک تجمعی با استفاده از تابع مفصل برنشتاین آمیخته و کاربرد آن در محاسبه ارزش در معرض خطر دمی

فاطمه حسینی[۱*]، صدیقه شمس[۱]
[۱]دانشگاه الزهرا(س)

چکیده: در این مقاله، یک مدل وابستگی انعطاف‌پذیر برای مدل‌سازی ریسک تجمعی معرفی می‌شود که بر پایه تابع مفصل برنشتاین آمیخته ساخته شده است. در این مدل، با وارد کردن یک عامل تصادفی مشترک، اثر ریسک سیستماتیک در کنار وابستگی میان ریسک‌های اختصاصی در نظر گرفته می‌شود. مدل پیشنهادی تعمیمی از ساختارهای ارشمیدسی است و امکان مدل‌سازی وابستگی‌های مثبت، منفی، تعویض‌پذیر و تعویض‌ناپذیر را فراهم می‌کند. بر اساس این ساختار، فرم‌های بسته‌ای برای تابع چگالی و تابع بقای ریسک تجمعی و همچنین معیار ارزش در معرض خطر دمی و سهم هر ریسک در تخصیص سرمایه ارائه می‌شود. نتایج شبیه‌سازی نشان می‌دهد که نوع تابع مفصل پایه و مرتبه چندجمله‌ای برنشتاین بر مقدار ارزش در معرض خطر و ارزش در معرض خطر دمی اثرگذار است.

واژه‌های کلیدی: تابع مفصل برنشتاین آمیخته،ریسک تجمعی،ارزش در معرض خطر،ارزش در معرض خطر دمی

*نویسنده مسئول: s.shams@alzahra.ac.ir



تحلیل شکاف میان ادراک ضرورت و آمادگی اجرایی فناوری زنجیره بلوکی در صنعت بیمه زندگی ایران: یک مطالعه پیمایشی

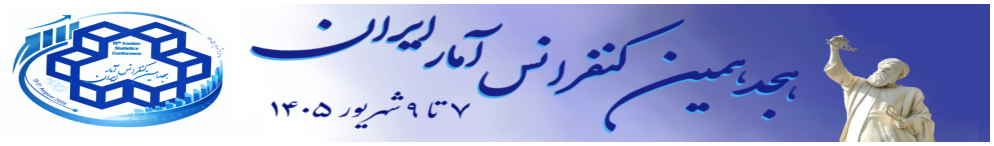
اسماء حمزه[1]*

[1]پژوهشکده بیمه

چکیده: فناوری زنجیره بلوکی به‌عنوان یکی از فناوری‌های تحول‌آفرین، ظرفیت بالایی برای بهبود شفافیت، کاهش تقلب، تسریع فرآیندهای بیمه‌ای و ارتقای مدیریت داده‌ها در صنعت بیمه زندگی دارد. با وجود مزایای بالقوه این فناوری، سطح پیاده‌سازی عملی آن در صنعت بیمه ایران همچنان محدود است. هدف این پژوهش، تحلیل آماری شکاف میان ادراک ضرورت و سطح آمادگی اجرایی فناوری زنجیره بلوکی در صنعت بیمه زندگی ایران است. پژوهش حاضر از نوع توصیفی-تحلیلی بوده و داده‌ها از طریق پرسشنامه مبتنی بر مقیاس پنج‌درجه‌ای لیکرت از ۲۴ نفر از خبرگان شامل مدیران، کارشناسان و متخصصان فعال در حوزه بیمه زندگی گردآوری شد. در تحلیل داده‌ها علاوه بر شاخص‌های توصیفی، از فاصله اطمینان و ضریب همبستگی برای پاسخ به سؤال‌های پژوهش استفاده شد. نتایج نشان داد برآورد میانگین ادراک ضرورت و مزایای فناوری زنجیره بلوکی در سطح نسبتا بالایی قرار دارد، در حالی که سطح آمادگی اجرایی و پیاده‌سازی عملی این فناوری پایین ارزیابی شد. همچنین نتایج بیانگر وجود شکاف قابل‌توجه میان ادراک مثبت نسبت به فناوری و میزان اجرای عملی آن در صنعت بیمه زندگی است. یافته‌ها نشان می‌دهد ضعف زیرساخت‌های فنی، ابهامات حقوقی و مقرراتی، و کمبود نیروی انسانی متخصص از مهم‌ترین عوامل محدودکننده توسعه این فناوری در صنعت بیمه زندگی ایران هستند.

واژه‌های کلیدی: زنجیره بلوکی،بیمه زندگی،تحول دیجیتال،پذیرش فناوری،صنعت بیمه ایران

*نویسنده مسئول: hamzeh@irc.ac.ir



ساختارهای جبری در مدل‌های آماری: رویکرد عملگرهای خطی برای کاهش بعد پایدار

پویان خامه چی[۱*]
[۱]دانشگاه ولایت

چکیده: کاهش بعد یکی از مسائل اساسی در تحلیل داده‌های با بعد بالا است که نقش مهمی در بهبود عملکرد مدل‌های یادگیری ماشین و کاهش پیچیدگی محاسباتی ایفا می‌کند. تحلیل مؤلفه‌های اصلی (PCA) به‌عنوان یکی از روش‌های کلاسیک، در شرایط نویزی و بدشرطی دچار ناپایداری عددی می‌شود. در این مقاله، یک چارچوب مبتنی بر نظریه عملگرهای خطی ارائه شده است که در آن ماتریس کوواریانس به‌عنوان یک عملگر خودالحاق در نظر گرفته شده و با افزودن جمله منظم‌سازی، پایداری طیفی بهبود می‌یابد. تحلیل نظری نشان می‌دهد که این رویکرد موجب کاهش عدد شرط، افزایش فاصله طیفی و کاهش حساسیت نسبت به نویز می‌شود. نتایج تجربی مبتنی بر اعتبارسنجی متقاطع نشان می‌دهد که روش پیشنهادی علاوه بر بهبود پایداری عددی، موجب افزایش دقت طبقه‌بندی نیز می‌شود.

واژه‌های کلیدی: کاهش بعد

*نویسنده مسئول: p.khamechi@velayat.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

برآورد پارامتری توابع کوپولا با استفاده از روش‌های بیزناپارامتری

سلیمان خزائی[۱]، کلثوم حسینی[۲*]
[۱]دانشگاه رازی
[۲]دانشگاه رازی کرمانشاه

چکیده: امروزه به‌کار بردن توابع کوپولا در انتخاب مدل آماری باعث افزایش دقت در ساختار نتایج می‌شود. روش‌های بیزناپارامتری کاربرد فراوانی در برآورد توابع کوپولا دارند. یکی از روش‌های بیزناپارامتری روش شکست چوب است که با استفاده از آن به نمونه‌گیری از آمیخته توابع کوپولا پرداخته می‌شود. در پایان از طریق روش‌های زنجیرمارکف مونت کارلو به مقایسه نتایج حاصل از داده‌های شبیه‌سازی شده و داده‌های واقعی برای خانواده‌ای از توابع کوپولاها پرداخته‌ایم.

واژه‌های کلیدی: بیز ناپارامتری،آمیخته کوپولا نامتناهی،روش شکست چوب،الگوریتم گیبز

*نویسنده مسئول: kolcomhossini@gmail.com

مثلث آماری تورم (MAT)

مهرناز خوشبخت[۱]، اکبر علیزاده اعتمادی[۲]*
[۱]دانشگاه تهران
[۲]دانشگاه جامع علمی ـ کاربردی

چکیده: تورم یکی از مهم‌ترین شاخص‌های اقتصاد کلان است که دقت و بهنگام بودن اندازه‌گیری آن، نقش تعیین‌کننده‌ای در سیاستگذاری پولی، مالی و تنظیم بازار دارد. در این پژوهش، چارچوبی نوین با عنوان «مثلث آماری تورم (MAT)» برای تلفیق آمار رسمی، داده‌های با بسامد بالا و مدل‌های اقتصادسنجی در سنجش تورم ایران ارائه می‌شود. این چارچوب بر این اصل استوار است که آمار رسمی، داده‌های فروشگاهی و سایر داده‌های نوظهور، به جای رقابت، می‌توانند در قالب ساختاری منسجم مکمل یکدیگر باشند. در این مقاله، ضمن مرور پیشینه داخلی و بین‌المللی، معماری مفهومی مثلث آماری تورم، مدل اقتصادسنجی مبتنی بر عامل نهفته، نحوه تعیین وزن‌های ترکیبی و فرآیند استانداردسازی و تلفیق داده‌ها تشریح شده است. همچنین یک مطالعه موردی مفهومی (اثبات مفهوم) برای نمایش مراحل اجرای چارچوب پیشنهادی ارائه شده و الزامات نهادی، حقوقی و فنی استقرار آن در نظام آماری کشور بررسی شده است. به دلیل محرمانه بودن داده‌های خام فروشگاه‌های زنجیره‌ای و محدودیت‌های حقوقی دسترسی به اطلاعات تراکنش‌های فروش، پژوهش حاضر بر طراحی و تبیین چارچوب روش‌شناختی تمرکز دارد و ارائه نتایج کمی به اجرای طرح‌های پایلوت و ایجاد سازوکارهای رسمی تبادل داده موکول شده است.

واژه‌های کلیدی: تورم، شاخص قیمت مصرف‌کننده، داده‌های با بسامد بالا، داده‌های فروشگاهی، اقتصادسنجی، فیلتر کالمن، مدل عامل نهفته، مثلث آماری تورم.

* نویسنده مسئول: mehrnazkhoshbakht@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

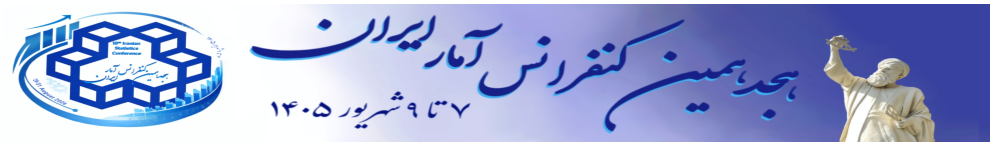
استنباط شواهدی در علم داده براساس نظریه دمپستر–شفر

علی دست برآورده[۱*]
[۱]دانشگاه یزد

چکیده: کمی‌سازی عدم‌قطعیت در مدل‌های یادگیری ماشین مدرن به عنوان یک چالش اساسی مطرح شده است، به‌ویژه برای کاربردهای پرریسک مانند رانندگی خودران، تشخیص پزشکی و ارزیابی ریسک مالی. در حالی که استنباط بیزی چارچوبی اصولی برای کمی‌سازی عدم‌قطعیت فراهم می‌کند، اتکای آن به توزیع‌های پیشین و دشواری محاسباتی برای مدل‌های بزرگ‌مقیاس، باعث شده است تا جستجو برای رویکردهای جایگزین ارزشمند باشد. این مقاله به مرور استنباط شواهدی مبتنی بر تابع درست‌نمایی می‌پردازد که تابع درست‌نمایی نسبی را به‌عنوان توزیع امکانی در چارچوب نظریه‌ی توابع باور دمپستر–شفر تفسیر می‌کند. در این راستا ضمن اشاره به بنیان‌های نظری این رویکرد، به کاربردهای آن در علم داده اشاره می‌شود.

واژه‌های کلیدی: استنباط مبتنی بر تابع درست‌نمایی، توابع باور، نظریه‌ی دمپستر–شفر، مجموعه‌های فازی تصادفی، یادگیری ماشین.

*نویسنده مسئول: dastbaravarde@yazd.ac.ir



شبکه عصبی عمیق مبتنی بر یادگیری باقیمانده و تابع درستنمایی افرون برای تحلیل داده‌های بقای با ابعاد بالا: یک مطالعه شبیه‌سازی

رضوان دلینو[1]*، حیدرعلی مردانی فرد[1]، مصطفی قادری زفره‌ای[1]
[1]دانشگاه یاسوج

چکیده: تحلیل داده‌های بقا با ابعاد بالا به دلیل وجود روابط پیچیده و غیرخطی میان متغیرها با چالش‌هایی مواجه است و مدل‌های خطی کلاسیک نظیر مدل کاکس در چنین شرایطی ممکن است کارایی محدودی داشته باشند. در این پژوهش، یک مدل شبکه عصبی عمیق مبتنی بر یادگیری باقیمانده همراه با تابع درستنمایی جزئی افرون برای مدل‌سازی داده‌های بقای پربعد ارائه می‌شود. در این چارچوب، ریسک نسبی افراد به صورت یک تابع غیرخطی از متغیرهای توضیحی مدل‌سازی شده و پارامترهای شبکه با استفاده از کمینه‌سازی منفی لگاریتم درستنمایی افرون برآورد می‌شوند. عملکرد روش پیشنهادی از طریق مطالعات شبیه‌سازی و با استفاده از شاخص توافق هارل با روش‌های کاکس ستیغی و جنگل تصادفی بقا مقایسه شد. نتایج نشان داد که مدل پیشنهادی در سناریوهای مختلف، به‌ویژه در شرایط افزایش ابعاد داده و وجود نویز، عملکرد پیش‌بینی بهتری ارائه می‌دهد.

واژه‌های کلیدی: یادگیری باقیمانده، تابع درستنمایی افرون، تحلیل بقا، داده‌های با ابعاد بالا، شبکه عصبی عمیق.

---

*نویسنده مسئول: rezvan.delino96@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

احتمال با مقادیر منفی: ریشه‌های تاریخی، فلسفه و کاربرد آن

علی دولتی [۱*]
[۱]گروه آمار، دانشکده ریاضی، دانشگاه یزد

چکیده: نظریه‌ی احتمال در صورت‌بندی کلاسیک خود بر اصول موضوعه‌ای استوار است که یکی از بنیادی‌ترین آن‌ها نامنفی بودن احتمال است. با این حال، در توسعه‌ی فیزیک کوانتومی در قرن بیستم، ساختارهایی ریاضی پدیدار شدند که از بسیاری جهات مشابه توزیع‌های احتمالی بودند، اما در برخی نواحی مقادیر منفی اختیار می‌کردند. این پدیده منجر به شکل‌گیری مفهوم «احتمال منفی» شد که نخستین بار به‌طور صریح در آثار دیراک و سپس در چارچوب دقیق‌تری توسط فاینمن مورد توجه قرار گرفت. این سخنرانی، به مفهوم احتمال منفی از منظر تاریخی، مفهومی و نقش آن در فیزیک کوانتومی می پردازد. به تفسیرهای مختلف در مورد ماهیت احتمال منفی نیز اشاره می‌شود.

واژه‌های کلیدی: احتمال منفی، توزیع‌شبه احتمال، اندازه علامت‌دار، مکانیک کوانتومی

*نویسنده مسئول: adolati@yazd.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

پیاده سازی مدل CTP-LN برای مدل بندی بارش استان تهران تا سال ۱۴۰۳

فاطمه دیده ور[۱*]، امیرحسین قطاری[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: مدل‌سازی همزمان فراوانی و حجم بارش از موضوعات مهم در هیدرولوژی آماری و تحلیل ریسک‌های اقلیمی است. در این پژوهش، کارایی مدل ترکیبی CTP-LN برای مدل‌سازی مشترک فراوانی و حجم بارش در استان تهران بررسی شد. بدین منظور، داده‌های بارش روزانه ۹۵ ایستگاه باران‌سنجی در استان تهران تحلیل گردید. عملکرد مدل پیشنهادی با چند توزیع رقیب بر اساس معیارهای AIC و BIC مقایسه شد. نتایج نشان داد در مدل‌سازی ارتفاع بارش، مدل CTP-LN عملکرد بسیار مناسبی دارد؛ به‌طوری که در 98.95 درصد ایستگاه‌ها کمترین مقدار AIC را کسب کرد. همچنین میانگین ΔAIC برابر با ۵۲۳۵ به دست آمد که بیانگر برتری آماری قابل توجه این مدل است. در مقابل، در تحلیل حجم بارش عملکرد مدل ضعیف‌تر بود؛ به‌گونه‌ای که تنها در 15.8 درصد ایستگاه‌ها بهترین برازش را ارائه داد و میانگین ΔAIC مقدار منفی بسیار بزرگی را نشان داد. به طور کلی، مدل ترکیبی پواسون برشی در صفر و لگ‌نرمال چارچوبی مناسب برای مدل‌سازی رابطه فراوانی و شدت بارش، به‌ویژه در تحلیل ارتفاع بارش، فراهم می‌کند. این چارچوب در مطالعات هیدرولوژیکی، ارزیابی خطر سیلاب و تحلیل اثرات تغییرات اقلیمی کاربرد دارد

واژه‌های کلیدی: داده‌های بارشی،مدل ترکیبی CTP-LN،توزیع پواسون برشی در صفر،توزیع لگ نرمال

*نویسنده مسئول: fatemeh.didehvar@aut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

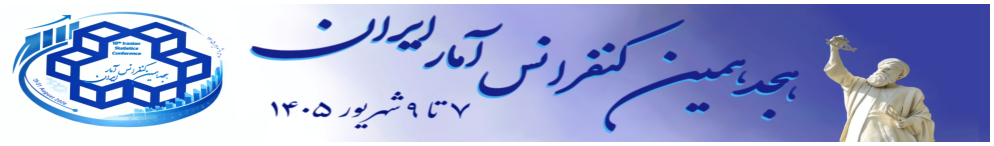
محاسبه تابع بقا در مدل δ-شوک تعمیم‌یافته با فواصل زمانی توزیع نمایی

لیلا رجبی[۱*]، سلیمان خزائی[۱]
[۱]دانشگاه رازی

چکیده: این مقاله به مطالعه یک مدل شوک تعمیم‌یافته می‌پردازد که در آن سیستم تحت تأثیر شوک‌هایی با فواصل زمانی تصادفی قرار دارد و سیستم زمانی از کار می‌افتد که فاصله زمانی بین دو شوک متوالی کمتر از مقدار آستانه δ_1 باشد. اگر این فاصله از δ_2 بیشتر شود ($\delta_1 < \delta_2$)، شوک تأثیری بر سیستم ندارد و در صورت قرار گرفتن فاصله زمانی بین δ_1 و δ_2، سیستم با احتمال θ از کار می‌افتد. تابع بقا در این مدل به‌صورت یک سری نامتناهی از انتگرال‌ها بیان می‌شود، که تحلیل کلی آن را دشوار می‌سازد. در این مقاله، تابع بقا برای حالت خاصی بررسی می‌شود که فواصل بین شوک‌ها دارای توزیع نمایی با پارامتر λ هستند. علاوه بر آن، نرخ خطر و طول عمر باقی‌مانده‌ی متوسط سیستم برآورد گردیده و سازگاری آن در حالت‌های حدی با مدل δ-شوک کلاسیک با شبیه‌سازی مونت‌کارلو بررسی شده است.

واژه‌های کلیدی: مدل شوک تعمیمیافته،تابع بقا،توزیع نمایی،طول عمر سیستم،نرخ خطر

*نویسنده مسئول: leilarajabi88@gmail.com



مقایسه عملکرد روش های ناپارامتری در برآورد تابع رگرسیون همراه با خطا

قدرت اله رحمتی[۱*]
[۱]دانشگاه پیام نور

چکیده: در برآورد تابع رگرسیون ممکن است مقدار واقعی متغیر پاسخ و پیش بین قابل مشاهده نباشند و همراه با خطا اندازه گیری شده باشند. در این حالت برآوردگرهای بدست آمده اریب خواهند بود. در این مقاله روش های ناپارامتری کرنل، اسپلاین و موجک برای برآورد تابع رگرسیون همراه با خطا با استفاده از شبیه سازی مونت کارلو و معیار میانگین مربعات خطا مورد مقایسه قرار گرفته اند. با توجه به نتایج شبیه سازی، روش موجک به مراتب نرخ همگرایی متناسب تری را ارایه می دهد.

واژه‌های کلیدی: رگرسیون همراه با خطا،رگرسیون کرنل،رگرسیون اسپلاین،روش موجک

*نویسنده مسئول: gh1.rahmati@pnu.ac.ir

مروری جامع بر خانواده‌ی معیارهای آنتروپی: ابزاری کلیدی برای سنجش پیچیدگی و پیش‌بینی‌پذیری در سری‌های زمانی

مریم رحیم زاده رنانی[۱*]، صفیه محمودی[۱]
[۱]دانشگاه صنعتی اصفهان

چکیده: امروزه با توجه به حجم فزاینده و ماهیت غیرخطی داده‌های زمانمند (مانند سیگنال‌های مغزی، نوسانات مالی و داده‌های صنعتی)، مدل‌های سنتی اغلب در درک و شناسایی کامل بی‌نظمی‌ها و الگوهای پنهان با محدودیت مواجه هستند. این مقاله به بررسی سیر تکامل و کاربرد مفهوم «آنتروپی» به عنوان یک معیار کمی، پویا و میان‌رشته‌ای برای سنجش عدم قطعیت، بی‌نظمی و پیچیدگی سیستم‌ها می‌پردازد. هدف از این مقاله، مروری بر خانواده‌ی معیارهای آنتروپی شامل آنتروپی شانون و تعمیم‌های آن (آنتروپی رنی و آنتروپی سالیس) و به ویژه برآوردگرهای نوین پیچیدگی مانند آنتروپی تقریبی و آنتروپی نمونه است. با تمرکز بر بهبود اساسی آنتروپی نمونه نسبت به آنتروپی تقریبی (مانند رفع اریبی و استقلال نسبی از طول داده‌ها)، اثربخشی این شاخص‌ها در تحلیل پیچیدگی سری‌های زمانی مورد ارزیابی قرار می‌گیرد. نتایج مقایسه‌ای و تحلیل مطالعه‌ی موردی نشان می‌دهد که آنتروپی نمونه حساسیت بالایی نسبت به تغییرات پارامترها و ویژگی‌های زمانی داده‌ها دارد و ابزاری قوی‌تر و واقعی‌تر برای ارزیابی اختلالات زمانی و بهبود مدل‌سازی سیستم‌های پیچیده است. این نتایج هم‌جهت و تأییدکننده‌ی یافته‌های ریچمن و مورمن است. در نهایت، این مقاله مسیرهای تحقیقاتی آتی در زمینه‌ی آنتروپی‌های چندمقیاسی را پیشنهاد می‌کند.

واژه‌های کلیدی: سری‌های زمانی، آنتروپی شانون، آنتروپی تقریبی، آنتروپی نمونه، آنتروپی رنی، آنتروپی سالیس.

*نویسنده مسئول: m.rahimzadeh@math.iut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مقایسه روش‌های آزمون عدم برازش در رگرسیون خطی با داده‌های بدون تکرار

زهرا رستمی[۱*]، احد ملک زاده[۱]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی

چکیده: یک مدل رگرسیونی به‌منظور تعیین تاثیر رفتار متغیر پاسخ بر اساس تعدادی متغیر توضیحی می‌باشد که به‌طور معمول روش‌های مختلفی در راستای برازش مدل وجود دارد. بعد از برازش انتظار داریم که رفتار متوسط متغیر پاسخ توسط متغیرهای توضیحی مدل شده باشند. سوال مهم اینجا است که آیا این هدف محقق شده است یا خیر؟ این بررسی تحت عنوان عدم برازش در مدل رگرسیونی مطرح می‌باشد و اولین بار (در حضور مشاهدات تکراری) روشی به‌منظور بررسی عدم برازش ارائه نمود. در واقعیت لزومی بر وجود تکرار در مشاهدات نیست و این امر باعث شد در ۸ دهه گذشته محققان زیادی به ارائه روش‌هایی بدین منظور بپردازند. در این مقاله ضمن طرح مفهوم عدم برازش روش‌های معروف به‌منظور بررسی آن‌را معرفی و به مقایسه آن‌ها تحت سناریوهای مختلف با بهره از شبیه‌سازی خواهیم پرداخت.

واژه‌های کلیدی: رگرسیون خطی،عدم برازش،داده‌های بدون تکرار،خوشه‌بندی

*نویسنده مسئول: malekzadeh@kntu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

کاربرد چارچوب تنش–مقاومت در ارزیابی ریسک دعاوی بانکی

امیر رضائی [۱*]
[۱]دانشگاه بوعلی سینا

چکیده: دعاوی بانکی به‌ویژه مطالبات معوق، اختلافات تسهیلاتی و ضمانت‌نامه‌ها، به دلیل حجم بالای تسهیلات، تأثیر تحریم‌ها و پیچیدگی مقررات، از پرریسک‌ترین حوزه‌های حقوقی نظام بانکی ایران هستند. پژوهش حاضر با اقتباس از مدل تنش–مقاومت در نظریه قابلیت اعتماد، چارچوبی کمی و شفاف برای ارزیابی این ریسک‌ها ارائه می‌دهد. در این چارچوب، عوامل تقویت‌کننده موقعیت بانک (کیفیت قراردادها، وثایق، مستندات و رویه قضایی) به عنوان «مقاومت حقوقی» (S) و عوامل تضعیف‌کننده (ابهامات قراردادی، ایرادات شکلی و گزارش‌های کارشناسی معارض) به عنوان «تنش حقوقی» (T) مدل‌سازی می‌شوند. بر این اساس، قابلیت اتکای ادعا به صورت $CR = P(S > T)$ و ریسک دعوی به صورت $LR = ۱ - CR$ تعریف می‌گردد. همچنین پژوهش حاضر با ارائه یک مطالعه موردی واقع‌گرایانه در حوزه مطالبات معوق بانکی (با CR0.52 و LR0.48)، نشان می‌دهد که چارچوب پیشنهادی قادر به تبدیل ارزیابی‌های کیفی خبرگان به شاخص‌های احتمالی قابل تفسیر و تصمیم‌گیری است. بنابراین این پژوهش می‌تواند با تحلیل حساسیت، محرک‌های اصلی ریسک را شناسایی کرده و به بانک‌ها در اولویت‌بندی پرونده‌ها، مذاکرات سازش و بهبود قراردادهای آینده کمک شایانی کند.

واژه‌های کلیدی: ریسک دعاوی بانکی،مدل تنش–مقاومت،مطالبات معوق،مدیریت ریسک حقوقی،بانکداری

*نویسنده مسئول: a.rezaei@basu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

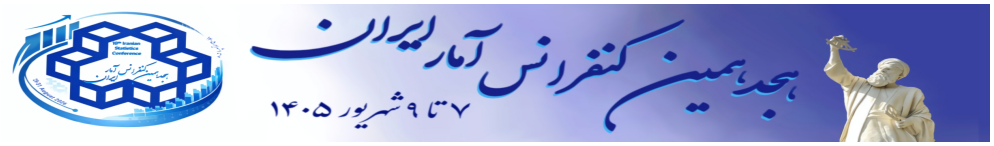
ارزیابی سوگیری جنسیتی در مدل‌های پیش‌بینی تکرار جرم با بهره‌گیری از تکنیک‌های هوش مصنوعی قابل توضیح: مطالعه‌ای بر داده‌های COMPAS

زهرا رضایی[۱]، نجمه اکبرپور[۱*]
[۱]پژوهشگاه قوه قضائیه

چکیده: روش‌های هوش مصنوعی قابل توضیح برای اطمینان از شفافیت و انصاف در مدل‌های یادگیری ماشین، به‌ویژه در حوزه‌های حساس مانند تصمیم‌گیری قانونی، حیاتی هستند. این مطالعه یک روش دو مرحله‌ای را برای افزایش تفسیرپذیری و کاهش تعصب جنسیتی در سیستم‌های حقوقی مبتنی بر هوش مصنوعی، با تمرکز بر مجموعه داده‌های COMPAS، پیشنهاد می‌کند. مرحله اول از یک مدل یادگیری عمیق کاملا متصل برای ثبت روابط غیرخطی پیچیده استفاده می‌کند و مرحله دوم از یک طبقه‌بند مبتنی بر تجمع بوت‌استرپی (بگینگ) برای بهبود تعمیم و استحکام بهره می‌گیرد. هر مرحله با چهار تکنیک XAI شامل ، LIME، SHAP اهمیت ویژگی با جایگشت و اثرات تجمعی محلی (ALE) برای ارزیابی اهمیت ویژگی و تشخیص سوگیری مرتبط با نژاد و جنسیت تحلیل شده است. نتایج نشان می‌دهد ویژگی‌های رفتاری مانند «تکرار جرم خشونت‌آمیز» و «تعداد سوابق کیفری پیشین» به‌طور مداوم بر پیش‌بینی‌ها غالب‌اند، در حالی که ویژگی‌های جمعیت‌شناختی نظیر نژاد و جنسیت تأثیر ناچیزی دارند. مدل بگینگ با میزان صحت 0.8156 (AUC) و دقت تست 73.80% به عملکرد برتر دست یافت. نتایج بر اهمیت کاربرد هوش مصنوعی اخلاقی و قابل تفسیر در سیستم‌های قضایی برای تضمین انصاف تأکید می‌کند.

واژه‌های کلیدی: هوش مصنوعی قابل توضیح (XAI)، عدالت جنسیتی، سیستم‌های حقوقی مبتنی بر هوش مصنوعی، پیش‌بینی تکرار جرم، مجموعه داده‌های COMPAS.

*نویسنده مسئول: najmeh.akbarpor@gmail.com



طراحی چارچوب داده‌محور برای پیش‌بینی دعاوی قراردادی مبتنی بر هوش مصنوعی و مبانی فقه امامیه

زهرا رضائی[۱*]، عبدالکریم حقیقی[۱]
[۱]پژوهشگاه قوه قضائیه

چکیده: افزایش دعاوی قراردادی ناشی از ابهام در مفاد قراردادها، عدم تقارن اطلاعاتی میان طرفین و ضعف در ارزیابی ریسک‌های حقوقی، یکی از چالش‌های مهم نظام قضایی و اقتصادی ایران به شمار می‌رود. پیشرفت‌های اخیر در حوزه هوش مصنوعی، امکان بهره‌گیری از سامانه‌های هوشمند برای تحلیل قراردادها، شناسایی عوامل اختلاف‌زا و پیش‌بینی ریسک بروز دعاوی را فراهم ساخته است. پژوهش حاضر با هدف طراحی یک چارچوب داده‌محور برای پیش‌بینی دعاوی قراردادی مبتنی بر هوش مصنوعی و مبانی فقه امامیه انجام شده است. روش تحقیق از نوع تحلیلی-استنباطی با رویکرد میان‌رشته‌ای بوده و ضمن بررسی منابع معتبر فقهی از جمله اللمعة الدمشقیة، مکاسب، تحریرالوسیله و جواهرالکلام، ظرفیت قاعده اشتراط موامره در انطباق با سامانه‌های هوشمند تحلیل شده است. در چارچوب پیشنهادی، هوش مصنوعی به‌عنوان «موامر دیجیتال» با بهره‌گیری از داده‌های قراردادی، شاخص‌های ریسک حقوقی و قواعد تصمیم‌گیری، نقش مشاور تخصصی در ارزیابی و پیش‌بینی احتمال بروز اختلافات ایفا می‌کند. یافته‌های پژوهش نشان می‌دهد که فقه امامیه، با اتکا به قواعد عقلایی، اصل لزوم شروط و اعتبار رجوع به خبره، مانعی برای استفاده از سامانه‌های هوشمند در جایگاه مشاور قراردادی قائل نیست، مشروط بر آنکه نقش این سامانه‌ها مخل قصد و رضایت طرفین نباشد. همچنین نتایج بیانگر آن است که بهره‌گیری از چارچوب‌های داده‌محور در مرحله پیش از انعقاد قرارداد می‌تواند به افزایش شفافیت، کاهش عدم قطعیت، مدیریت ریسک‌های قراردادی و در نهایت کاهش احتمال شکل‌گیری دعاوی حقوق خصوصی کمک کند.

واژه‌های کلیدی: هوش مصنوعی، پیش‌بینی دعاوی قراردادی، تحلیل داده، ارزیابی ریسک حقوقی، موامره، فقه امامیه.

*نویسنده مسئول: z.rezaei@jri.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

ارزیابی سازوکار هدفمندسازی در برنامه پرداخت یارانه نقدی به خانوارهای کشور

علیرضا رضایی[۱*]، ریحانه رضایی[۲]
[۱]سازمان هدفمندسازی یارانه‌ها
[۲]دانشگاه شهید بهشتی

چکیده: یکی از برنامه‌های مهم کمک اجتماعی در کشور پرداخت یارانه نقدی به خانوارها است. این برنامه با توجه به پوشش کل خانوارهای سراسر کشور و اعتبار قابل توجه آن از اهمیت ویژه‌ای در بین برنامه‌های حمایت اجتماعی کشور برخوردار است. برنامه پرداخت یارانه نقدی در راستای اجرای قانون هدفمندکردن یارانه‌ها از ماه دی سال ۱۳۸۹ آغاز گردید. هدفمندسازی در این برنامه موضوع بسیار مهمی است، زیرا شناسایی خانوارهای غیر واجد شرایط حمایت دولت و حذف یارانه آنان همواره مورد نظر سیاستگذاران در طول سال‌های اجرای آن بوده است و بنا براین ارزیابی سازوکار هدفمندسازی برای سنجش حصول اهداف آن اهمیت زیادی دارد. شاخص‌های متعددی برای ارزیابی سازوکار هدفمندسازی در جنبه‌های مختلف ارائه شده است که در بین آن‌ها خطای طرد و شمول دو شاخص مهم نشان‌دهنده میزان موفقیت در غربالگری جامعه هدف به شمار می‌روند. آمارگیری هزینه و درآمد خانوار مرکز آمار ایران با توجه به مشخصات آن و همچنین استقلال فرآیند جمع‌آوری داده‌های آمارگیری از فرآیند گروه‌بندی درآمدی خانوارها در دستگاه اجرایی متولی، می‌تواند امکان محاسبه دو شاخص خطای طرد و شمول را برای ارزیابی سازوکار هدفمندسازی برنامه پرداخت یارانه نقدی فراهم کند. با این وجود مواردی در محاسبه خطاهای مذکور با استفاده از داده‌های آمارگیری وجود دارد که باید مورد ملاحظه قرار گیرد. در این مقاله ضمن بررسی این موارد قابل ملاحظه به محاسبه خطاهای طرد و شمول با استفاده از داده‌های آمارگیری مذکور و تحلیل آن پرداخته شده است.

واژه‌های کلیدی: خطای شمول،خطای طرد،فقر،کمک اجتماعی،هدفمندسازی

*نویسنده مسئول: alireza.re97@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

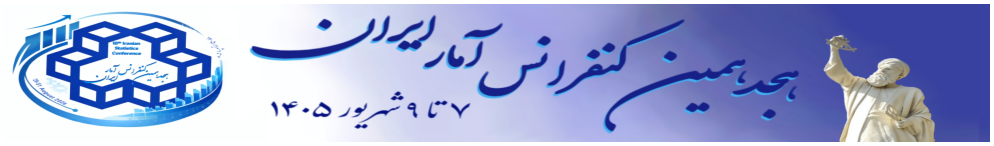
مکان‌یابی بهینه سرورهای کش در شبکه‌های توزیع با استفاده از خوشه‌بندی

مرضیه رضائی[۱*]، عباس پرچمی[۱]، ایوب شیخی[۱]
[۱]دانشگاه شهید باهنر کرمان

چکیده: تمامی روش‌های خوشه‌بندی موجود برای داده‌های حاصل از جوامع دقیق طراحی شده‌اند، در حالی که در بسیاری از مسائل واقعی، جوامع مورد مطالعه ماهیتی فازی دارند و هر مشاهده علاوه بر مقدار عددی، دارای درجه‌ای از عضویت در جامعه‌ی فازی نیز است. در این مقاله، یک الگوریتم خوشه‌بندی مبتنی بر میانه برای داده‌های مستخرج از جوامع فازی ارائه می‌شود. همچنین با بکارگیری روش Elbow در چارچوب الگوریتم پیشنهادی، تعداد بهینه خوشه‌ها تعیین می‌شود. در نهایت، کارایی روش ارائه‌شده در تحلیل داده‌های مرتبط با تعیین مکان‌های مناسب برای نصب سرورهای کش مورد بررسی و ارزیابی قرار می‌گیرد.

واژه‌های کلیدی: خوشه‌بندی،تابع عضویت،داده‌های مستخرج از جوامع فازی،وزن دهی

*نویسنده مسئول: Rezaei.Marzieh2021@math.uk.ac.ir



تحلیل الگوهای مکانی-زمانی تحرک جمعیت و شناسایی مکان‌های معنادار بر اساس داده‌های تلفن همراه

زهرا رضائی قهرودی[۱*]، یاسین محمدآقایی[۱]
[۱]دانشگاه تهران

چکیده: در چند دهه اخیر، گسترش استفاده از تلفن‌های همراه، به‌ویژه گوشی‌های هوشمند، موجب شده است بخش قابل توجهی از فعالیت‌های روزمره افراد در بستر شبکه‌های ارتباطی ثبت شود. بنابراین، داده‌های مکانی تلفن‌های هوشمند به یکی از منابع کلیدی برای درک الگوهای جابه‌جایی شهری و یکی از حوزه‌های مهم پژوهشی در علوم داده و مطالعات جمعیتی تبدیل شده‌اند. یکی از مراحل اساسی در این فرایند، شناسایی محل سکونت و محل کار افراد است که مبنای تحلیل‌هایی مانند الگوهای رفت‌وآمد، وضعیت اشتغال، دسترسی به خدمات و الگوهای اجتماعی-اقتصادی قرار می‌گیرد. اگرچه این داده‌ها امکان تحلیل تحرک جمعیت، شناسایی الگوهای جابه‌جایی، و تولید شاخص‌های پویای جمعیتی را با تفکیک مکانی و زمانی بالا فراهم می‌کنند، اما، ماهیت این داده‌ها شامل چالش‌هایی همچون پراکندگی داده، اریبی نمونه‌گیری و محدودیت‌های دقت مکانی است که استفاده مؤثر از آن‌ها را نیازمند روش‌های پیشرفته پردازش و اعتبارسنجی می‌سازد. در این مقاله، با مرور ادبیات و چارچوب‌های بین‌المللی، نقش داده‌های تلفن همراه به‌عنوان منبعی مکمل در تولید آمارهای رسمی در حوزه‌هایی نظیر جابه‌جایی جمعیت در شرایط بحران، نقشه‌برداری جمعیت پویا، مهاجرت و گردشگری بررسی شده است. سپس، دو رویکرد مهم برای استخراج الگوهای مکانی افراد شامل الگوریتم‌های تشخیص محل سکونت و محل کار و همچنین رویکرد «فضای فعالیت» معرفی و پیاده‌سازی شده‌اند. نتایج نشان می‌دهد که ترکیب این روش‌ها می‌تواند دقت بالایی در شناسایی رفتارهای مکانی افراد و تحلیل جابه‌جایی جمعیت ارائه دهد. همچنین، استفاده از مفهوم فضای فعالیت نسبت به رویکردهای سنتی مبتنی بر محل سکونت، امکان تحلیل جامع‌تر و واقع‌بینانه‌تری از تحرک انسانی فراهم می‌کند. این مطالعه بر اهمیت استفاده مکمل از داده‌های تلفن همراه در کنار منابع سنتی آمار رسمی تأکید دارد.

واژه‌های کلیدی: داده‌های تلفن همراه، تحرک جمعیت، آمارهای رسمی، نقاط توقف، الگوریتم‌های تشخیص محل زندگی.

*نویسنده مسئول: z.rezaeigh@ut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌بندی ترکیبی آماری و یادگیری ماشین داده‌های علم اعصاب‌شناختی

محدثه رفیقی[۱*]، زهرا رضائی قهرودی[۱]
[۱]دانشگاه تهران

چکیده: ساختار داده‌های علم اعصاب به دلیل مشاهدات مکرر از مکان‌های مختلف مغز در افراد، دارای ساختار آشیانه‌ای است. در روش‌های تحلیل این نوع داده‌ها، باید ساختار وابستگی بین اندازه‌گیری مکرر در نظر گرفته شود. با این حال، فرض وابستگی در ادبیات تحلیل داده‌های علوم اعصاب شناختی عمدتا نادیده گرفته می‌شود. در این مقاله، از روش‌های آماری و روش‌های یادگیری ماشین که برای تحلیل داده‌های مکرر توسعه یافته است، استفاده می‌شود و مزایا و معایب الگوریتم‌های مختلف مقایسه می‌شود. این پژوهش داده‌های علوم اعصاب را با در نظر گرفتن ساختار وابستگی با چندین روش آماری و یادگیری ماشین و روش‌های ترکیبی آن‌ها تحلیل می‌کند. علاوه بر این، عملکرد برازش الگوریتم‌های مختلف با استفاده از مجموعه داده‌های آلوده و رویکرد اعتبارسنجی متقابل مقایسه می‌شود. نتایج این پژوهش نشان می‌دهد که در تحلیل داده‌های علم اعصاب‌شناختی با ساختار آشیانه‌ای، استفاده از الگوریتم درخت امیدگیری بیشینه‌سازی اثرهای تصادفی، همراه با در نظر گرفتن ساختار آشیانه‌ای کانال‌های مغزی درون افراد، بالاترین دقت پیش‌بینی را ارائه می‌دهد. همچنین، روش‌های مدل خطی اثرهای آمیخته و بوستینگ فرایند گاوسی نسبت به آلودگی داده‌ها مقاومت بیشتری دارند، درحالی‌که درخت اثرهای آمیخته خطی تعمیم‌یافته، عملکرد بهتری در پیش‌بینی داده‌های آزمایشی در اعتبارسنجی متقابل نشان می‌دهد.

واژه‌های کلیدی: یادگیری ماشین، مدل آمیخته، مطالعات شناختی ، طیف‌شناسی عملکردی فروسرخ نزدیک (fNIRS)

*نویسنده مسئول: mohadese.rafighi@ut.ac.ir

مدل سری زمانی اتورگرسیو میانگین متحرک لگ لجستیک و کاربرد آن درتحلیل شاخص اقتصادی ضریب جینی

بهزاد ریحانی نیا[۱*]، فرشین هرمزی نژاد[۱]، محمد خدامرادی[۲]، محمد رضا قلانی[۱]
[۱]دانشگاه آزاد اسلامی واحد اهواز
[۲]دانشگاه آزاد اسلامی واحد ایذه

چکیده: یکی از توزیعهای آماری پرکاربرد در علوم مختلف، توزیع لگ لجستیک است، که در اقتصاد به تابع فیسک معروف است و مقدار عددی معکوس پارامتر اصلی آن برابر با ضریب جینی است.در الگوهای سری زمانی این امکان وجود دارد که متغیر وابسته به زمان از توزیع لگ لجستیک پیروی کند بر همین اساس در این مقاله، یک مدل سری زمانی جدید، که مدل لگ لجستیک اتورگرسیو میانگین متحرک (L-LARIMA) نامیده می شودمعرفی شده است که با استفاده تابع حداکثر درستنمایی شرطی و روش عددی نیوتن–رافسون پارامترهای آن برآورد شده است.آزمون فرضیه، ماتریس اطلاع و تابع پیش بینی نیز برآورد شده است. با استفاده از این مدل جدید، داده های اقتصادی شاخص جینی در ایران از سال ۱۳۹۰ تا ۱۴۰۳مورد بررسی و مدلبندی و تابع پیش بینی آن تعیین گردید.

واژه‌های کلیدی: توزیع لگ لجستیک،تابع درستنمایی شرطی،تابع پیش‌بینی،ماتریس اطلاع،ضریب جینی

---

*نویسنده مسئول: Farshin.Hormozinejad@iau.ac.ir

تحلیل فضایی-زمانی مدل‌های اشغال گونه‌های زیستی

زینب ستاری‌مهر[۱*]، محسن محمدزاده[۱]
[۱]دانشگاه تربیت مدرس

چکیده: مدل‌سازی توزیع گونه‌های زیستی مستلزم بهره‌گیری از داده‌های مشاهده‌شده‌ی حضور یا فراوانی در ابعاد زمان و مکان است. یکی از چالش‌های جدی در این زمینه، پدیده‌ی تشخیص ناقص است. جهت پیشگیری از سوگیری‌های آماری، ضروری است که احتمال تشخیص در فرآیند مدل‌سازی ادغام شود. مدل‌های اشغال به رویکردی استاندارد در تحلیل داده‌های حضور-غیاب بدل گشته‌اند؛ با این وجود، توسعه‌ی ابزارهای کارآمد برای بسط این مدل‌ها به چارچوب‌های فضایی-زمانی محدود است. پژوهش حاضر به واکاوی مدل‌های اشغال پرداخته و تحلیل بیزی آن‌ها را از طریق روش تقریب لاپلاس آشیانه‌ای یکپارچه با رویکرد معادلات دیفرانسیل جزئی تصادفی ارائه می‌دهد. به منظور ارزیابی کارایی چارچوب پیشنهادی، این مدل‌ها بر روی داده‌های شبیه‌سازی‌شده و داده‌های واقعی گونه‌ی پرنده‌ی ویروی چشم‌سرخ پیاده‌سازی شدند. نتایج ارزیابی بر اساس DIC و WAIC نشان داد که مدل‌های دربرگیرنده‌ی متغیرهای تبیینی فضایی-زمانی پویا عملکرد بهتری دارند. همچنین مشخص شد چارچوب تقریب لاپلاس آشیانه‌ای یکپارچه با رویکرد معادلات دیفرانسیل جزئی تصادفی چالش‌های محاسباتی الگوریتم‌های سنتی MCMC را برطرف کرده و زمان پردازش در داده‌های بزرگ‌مقیاس را به‌طور چشم‌گیری کاهش می‌دهد.

واژه‌های کلیدی: مدل اشغال، مدل‌سازی فضایی-زمانی، تقریب لاپلاس آشیانه‌ای یکپارچه، معادلات دیفرانسیل جزئی تصادفی.

*نویسنده مسئول: zeinab.satarimehr@modares.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

توسعه نمودار کنترل مبتنی بر رگرسیون بردار پشتیبان برای پایش پروفایل‌های پواسون با تابع پیوند لگاریتمی

سوگند سلطان محمدی[۱*]، فیروزه حقیقی[۱]
[۱]دانشگاه تهران

چکیده: پایش پروفایل‌های مدل خطی تعمیم‌یافته در شرایطی که کیفیت فرآیند به رابطه میان متغیرهای وابسته و مستقل وابسته است، از اهمیت ویژه‌ای برخوردار است. در این پژوهش، چارچوبی نوآورانه مبتنی بر مدل یادگیری ماشین به نام رگرسیون بردار پشتیبان برای طراحی و پایش نمودارهای کنترل پروفایل‌های پواسون ارائه می‌شود. در این راستا، تنها تابع پیوند لگاریتمی برای مدل‌سازی داده‌ها در نظر گرفته شده و مدل رگرسیون بردار پشتیبان برای ساخت نمودار کنترل به کار رفته است. عملکرد روش پیشنهادی از طریق مطالعه شبیه‌سازی و با استفاده از شاخص میانگین طول اجرا ارزیابی شده است. در این مطالعه، یک روش آموزشی مشخص برای مدل رگرسیون بردار پشتیبان و یک اندازه تغییر معین از نوع همزمان در پارامترهای فرآیند بررسی شده است تا کارایی نمودار کنترل در شرایط تعریف‌شده ارزیابی گردد. نتایج شبیه‌سازی نشان می‌دهد که نمودار کنترل مبتنی بر رگرسیون بردار پشتیبان قادر است تغییرات فرآیند را در شرایط تعیین‌شده با دقت مناسبی شناسایی کند.

واژه‌های کلیدی: نمودارهای کنترل، یادگیری ماشین، مدل خطی تعمیم‌یافته، پایش پروفایل، تابع پیوند، میانگین طول اجرا.

*نویسنده مسئول: sogandmohammadi@ut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

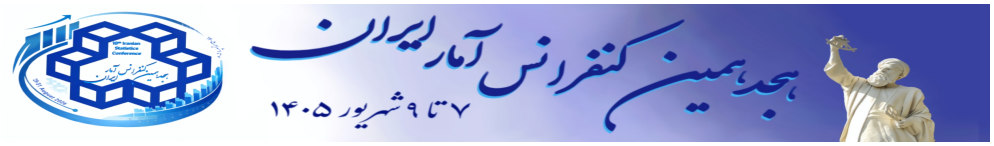
الگوریتم‌های بهینه‌سازی نوین برای رگرسیون SCAD با پاسخ‌های چندمتغیره

سیدعلی شاه‌صاحبی[۱]، امیرحسین قطاری[۱*]، فرشته اکبری[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: بهینه‌سازی توابع هدف در رگرسیون‌های تاوانیده همواره یکی از چالش‌های اساسی در آمار محاسباتی بوده است، زیرا دستیابی به برآوردی دقیق و پایدار از بردار ضرایب، نقشی کلیدی در عملکرد نهایی مدل ایفا می‌کند. در این مقاله، دو الگوریتم کارآمد به نام‌های کاهش مختصات (CD) و تقریب خطی موضعی (LLA) را برای برآورد بردار ضرایب در رگرسیون SCAD با پاسخ‌های چندمتغیره (Multi-SCAD) ارائه می‌دهیم. نتایج مطالعات عددی نشان می‌دهد که الگوریتم پیشنهادی CD در مقایسه با روش مرسوم نیوتن-رافسون، به ویژه در سناریوهای پرابعاد با همبستگی کم تا متوسط، عملکرد بهتری داشته و خطای پیش‌بینی کمتری به همراه دارد. همچنین الگوریتم LLA پایداری بالاتری را در انتخاب متغیرها نشان می‌دهد و حساسیت کمتری به ساختار همبستگی بین پیش‌بینی‌کننده‌ها دارد. این یافته‌ها حاکی از کارایی بالای الگوریتم‌های پیشنهادی برای مسائل رگرسیون تاوانیده با پاسخ‌های چندمتغیره است.

واژه‌های کلیدی: رگرسیون SCAD، پاسخ چندمتغیره، الگوریتم کاهش مختصات، تقریب خطی موضعی، انتخاب متغیر.

*نویسنده مسئول: seyedalishahsahebi@aut.ac.ir



حکمرانی داده و نقش نظام آماری ایران: از تولید آمار رسمی تا متولی‌گری داده

اشکان شباک[۱]*، مرضیه خاکستری[۲]
[۱]پژوهشکده آمار
[۲]دانشگاه خوارزمی

چکیده: تحولات شتابان فناوری‌های دیجیتال، گسترش داده‌های ثبتی و اداری، ظهور کلان‌داده‌ها و توسعه هوش مصنوعی، مفهوم و کارکرد نظام‌های آماری ملی را با تحولاتی بنیادین مواجه ساخته است. در این شرایط، نظام آماری دیگر صرفا متولی تولید و انتشار آمار رسمی نیست، بلکه به یکی از ارکان اصلی حکمرانی داده‌محور و تصمیم‌گیری مبتنی بر شواهد تبدیل شده است. مسئله اساسی این پژوهش آن است که نظام آماری ایران تا چه اندازه از ظرفیت نهادی، حقوقی، فنی و انسانی لازم برای ایفای چنین نقشی برخوردار است و چه الگویی می‌تواند گذار آن را از یک نظام عمدتا تولیدمحور به یک نظام متولی‌محور و داده‌محور امکان‌پذیر سازد. این مقاله با رویکردی تحلیلی ـ توصیفی، ضمن تبیین مبانی نظری حکمرانی داده، مفهوم چرخه عمر داده، هرم دانایی (DIKW) و چارچوب DAMA-DMBOK، به بررسی تحول مأموریت سازمان‌های ملی آمار و تجربه کشورهای پیشرو می‌پردازد. سپس با تمرکز بر وضعیت ایران، چالش‌هایی همچون معماری اسپاگتی در تبادل داده، سیلوهای داده، ضعف جایگاه قانونی و نهادی مرکز آمار ایران، کاهش سرمایه انسانی و سرمایه اجتماعی نظام آماری را تحلیل می‌کند. در ادامه، الگوی مطلوب هاب‌محور، نقش مرکز آمار ایران به عنوان «متولی ملی داده» و «دریاچه داده ملی»، مفهوم اتاق امن، تعامل‌پذیری مبتنی بر شناسه‌های یکتا و نقش استانداردهای SDMX، GSBPM و GSIM در ایجاد یک زیست‌بوم داده‌ای یکپارچه تبیین می‌شود. یافته‌های تحلیلی مقاله نشان می‌دهد که تحول نظام آماری ایران صرفا یک پروژه فنی یا فناوری اطلاعات نیست، بلکه مستلزم بازتعریف جایگاه حقوقی و نهادی مرکز آمار ایران، استقرار سازوکارهای حکمرانی داده، ایجاد معماری هاب‌محور، تضمین دسترسی امن به داده‌های اداری و ثبتی، تقویت زیرساخت‌های فراداده و تعامل‌پذیری و توسعه سرمایه انسانی تخصصی است. در این چارچوب، سازمان ملی آمار باید از جایگاه «تولیدکننده آمار» به جایگاه «متولی و هماهنگ‌کننده زیست‌بوم داده ملی» ارتقا یابد؛ جایگاهی که می‌تواند پیش‌شرط تحقق حکمرانی داده‌محور و آمادگی نظام آماری کشور برای عصر هوش مصنوعی باشد.

واژه‌های کلیدی: حکمرانی داده،نظام آماری،متولی‌گری داده،چرخه عمر داده،استانداردهای آماری

---

*نویسنده مسئول: a.shabbak@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

بررسی نقش آمار و داده‌کاوی در توسعه فناوری‌های هوشمند و کاربردهای مهندسی

حسین شریعت پناه[۱*]، حمید خرسند[۱]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی

چکیده: در سال‌های اخیر، رشد سریع فناوری‌های دیجیتال و افزایش حجم داده‌ها موجب شده است که آمار و تحلیل داده به‌یکی از ارکان اصلی توسعه فناوری‌های نوین تبدیل شوند. هوش مصنوعی به عنوان یکی از مهم‌ترین شاخه‌های علوم رایانه، وابستگی زیادی به روش‌های آماری و تحلیل داده دارد. بسیاری از الگوریتم‌های یادگیری ماشین، شبکه‌های عصبی و سیستم‌های پیش‌بینی بر پایه مفاهیم آماری توسعه‌یافته‌اند. همچنین در علوم مهندسی، استفاده از تحلیل داده و مدل‌های آماری نقش مهمی در کنترل کیفیت، پیش‌بینی خرابی تجهیزات، بهینه‌سازی فرآیندها و تصمیم‌گیری هوشمند ایفا می‌کند. در این مقاله، نقش آمار و تحلیل داده در توسعه هوش مصنوعی و کاربردهای مهندسی مورد بررسی قرار گرفته است. همچنین برخی از مهم‌ترین روش‌های آماری مورد استفاده در سامانه‌های هوشمند و چالش‌های موجود در تحلیل داده‌های بزرگ بررسی شده‌اند. نتایج نشان می‌دهد که تلفیق روش‌های آماری با فناوری‌های هوشمند می‌تواند موجب افزایش دقت، سرعت و کارایی سیستم‌های مهندسی و صنعتی شود.

واژه‌های کلیدی: آمار،تحلیل داده،هوش مصنوعی،یادگیری ماشین،علوم مهندسی،داده‌کاوی

*نویسنده مسئول: hosseinshariat72@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

روش‌های شبیه‌سازی برای مدل‌های بقا با متغیرهای تبیینی زمان وابسته

حنان شریفی[۱*]، کیومرث مترجم[۱]
[۱]دانشگاه تربیت مدرس

چکیده: با گسترش مطالعات طولی، نیاز به روش‌هایی که بتوانند متغیرهای زمان وابسته را در تحلیل بقا لحاظ کنند افزایش یافته است. مدل مخاطرات متناسب کاکس در چنین شرایطی کاربرد گسترده‌ای دارد، اما شبیه‌سازی داده برای این مدل، به‌ویژه در حضور متغیرهای پویا، همچنان چالش‌برانگیز است. در این پژوهش، به بررسی روشی پرداخته شده است که امکان تولید زمان‌های بقا را تحت مدل کاکس با متغیرهای زمان‌وابسته فراهم می‌سازد و تغییرات متغیرها را در طول دوره پیگیری بازتاب می‌دهد. این روش بر پایه تبدیل متغیرهای تصادفی مشتق شده از توزیع نمایی تکه‌ای بریده شده عمل می‌کند و انعطاف‌پذیری لازم برای تعریف تعداد دلخواهی از متغیرهای زمان‌وابسته و بازه‌های زمانی را فراهم می‌آورد. همچنین امکان شبیه‌سازی تغییرات متغیرها در گام‌های زمانی گسسته فراهم شده است تا ساختار داده‌های طولی بهتر بازنمایی شود. نتایج مطالعات شبیه‌سازی نشان داد که روش مورد بررسی قادر است داده‌هایی منطبق با مفروضات مدل کاکس تولید کند و رفتار الگوی خطر را در حضور متغیرهای زمان‌وابسته بازسازی نماید. این رویکرد می‌تواند ابزاری مؤثر برای پژوهشگران در طراحی مطالعات شبیه‌سازی و ارزیابی عملکرد مدل‌های بقا باشد.

واژه‌های کلیدی: تحلیل بقا، مدل مخاطرات متناسب کاکس، متغیرهای زمان‌وابسته، شبیه‌سازی داده‌های بقا.

*نویسنده مسئول: sharifi.h@modares.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

یک برآوردگر ناپارامتری برای ضریب وابستگی چندکی

الهام شریفی آبدر[۱]*، علی دست‌برآورده[۱]، علی دولتی[۱]، سیدمحسن میرحسینی[۱]
[۱]یزد

چکیده: ضریب وابستگی دمی، ابزاری کلاسیک برای اندازه‌گیری وابستگی در رویدادهای فرین است، اما از کمی‌سازی وابستگی در سایر نقاط توزیع، مانند چندک‌ها، ناتوان است. به‌تازگی، مفهوم ضریب وابستگی چندکی به‌عنوان تعمیمی از ضریب وابستگی دمی معرفی شده است که شدت وابستگی را در هر نقطه از توزیع، به‌صورت موضعی، اندازه‌گیری می‌کند. با وجود این، برآورد آماری آن تاکنون بررسی نشده است. در این مقاله، یک برآوردگر ناپارامتری برای این ضریب ارائه شده است و ویژگی‌های مجانبی آن شامل سازگاری، نرمال‌بودن مجانبی و نااریبی مجانبی نشان داده شده است.

واژه‌های کلیدی: مفصل،ضریب وابستگی چندکی،برآورد ناپارامتری،سازگاری

*نویسنده مسئول: sharifiabdar@stu.yazd.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

کاربرد توابع مفصل در یادگیری تقویتی: رویکردی مبتنی بر مدل‌سازی وابستگی

صدیقه شمس[۱*]
[۱]الزهرا

چکیده: یادگیری تقویتی یکی از شاخه‌های مهم یادگیری ماشین است که در آن عامل از طریق تعامل با محیط، سیاست تصمیم‌گیری مناسب را در مسائل ترتیبی و تحت عدم قطعیت فرا می‌گیرد. با وجود موفقیت‌های گسترده این روش در حوزه‌هایی مانند رباتیک، مالی، بیمه، سلامت و سیستم‌های هوشمند، بسیاری از مدل‌های متداول یادگیری تقویتی بر فرض استقلال میان مؤلفه‌های عمل، اعمال عامل‌های مختلف یا مؤلفه‌های حالت آینده استوار هستند. این فرض در بسیاری از کاربردهای واقعی، به‌ویژه در فضاهای عمل چندبعدی، محیط‌های چندعامله و سیستم‌های دارای رخدادهای حدی، چندان واقع‌بینانه نیست. توابع مفصل چارچوبی انعطاف‌پذیر برای جداسازی توزیع‌های حاشیه‌ای از ساختار وابستگی فراهم می‌کنند و امکان مدل‌سازی وابستگی‌های غیرخطی، نامتقارن و دمی را فراهم می‌سازند. در این مقاله، نقش توابع مفصل در مدل‌سازی وابستگی در فضای عمل، سیاست‌های تصادفی، تابع هدف، تابع ارزش و دینامیک انتقال حالت بررسی می‌شود. همچنین رویکردهای مختلف انتخاب تابع مفصل، شامل انتخاب مبتنی بر دانش مسئله، برازش آماری، تابع ارزش، میانگین‌گیری مدل و مفصل‌های عصبی مرور می‌گردد. این بررسی نشان می‌دهد که توابع مفصل می‌توانند چارچوبی نظری و قابل توسعه برای طراحی سیاست‌های وابسته، مدل‌سازی دقیق‌تر محیط و توسعه روش‌های یادگیری تقویتی حساس به ریسک فراهم کنند.

واژه‌های کلیدی: یادگیری تقویتی،تابع مفصل،مدل‌سازی وابستگی،فضای عمل،یادگیری چندعامله

*نویسنده مسئول: s.shams@alzahra.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

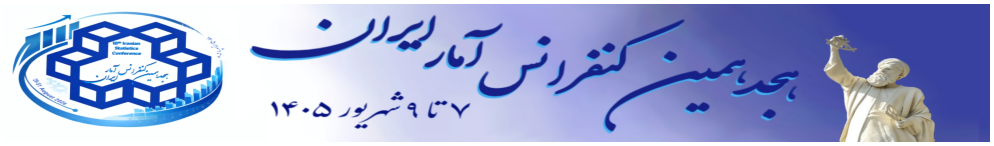
معرفی و بررسی چند مدل نیمه‌طیفی در برآورد توابع کوواریانس فضایی-زمانی

علی شهنواز[۱*]، جواد ناصریان[۱]
[۱]دانشگاه آزاد اسلامی واحد زنجان

چکیده: در این مقاله، رویکرد مدل‌سازی نیمه‌طیفی در تحلیل داده‌های فضایی-زمانی مورد واکاوی قرار می‌گیرد. این مقاله ضمن بررسی مبانی نظری و مرور سیر تحول این مدل‌ها، به معرفی دقیق رده‌های جدیدی از مدل‌های نیمه‌طیفی می‌پردازد. در ادامه، ویژگی‌های کلیدی این مدل‌ها استخراج شده و رده‌ای از مدل‌های نیمه‌طیفی با ساختارهای تفکیک‌پذیر و تفکیک‌ناپذیر با رویکرد تقارن کامل در قالب مثال‌های کاربردی ارائه می‌گردد. نتایج این پژوهش ضمن تبیین انعطاف‌پذیری مدل‌های معرفی‌شده در برازش تابع کوواریانس، راهکارهای کارآمدی را برای مدل‌سازی فرآیندهای فضایی-زمانی پیچیده فراهم می‌آورد

واژه‌های کلیدی: کوواریانس فضایی-زمانی،مدل فضایی-زمانی،مدل نیمه طیفی

*نویسنده مسئول: shahnavaz_ali@iau.ac.ir



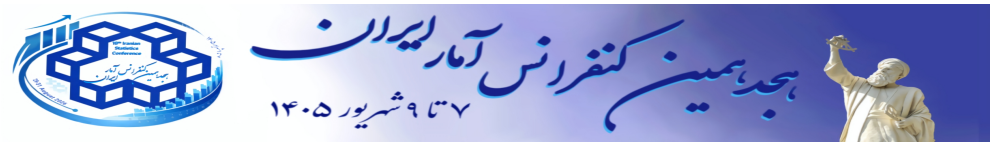
طبقه‌بندی **k** نزدیک‌ترین همسایگی با استفاده ازهمبستگی محلی

ایوب شیخی[۱]، زهرا ابولی[۱]، فاطمه ابولی[۱]*
[۱]دانشگاه شهید باهنر کرمان،

چکیده: یادگیری نیمه‌نظارتی رویکردی قدرتمند برای وظایف طبقه‌بندی است که در آن‌ها داده‌های برچسب‌گذاری‌شده کمیاب یا گران‌قیمت هستند. با این حال، روش‌های نیمه‌نظارتی موجود، اغلب با ساختارهای خوشه‌ای پیچیده و همپوشانی خوشه ها دست و پنجه نرم می‌کنند. علاوه بر این، این روش ها به فرضیات هندسی قوی‌ای متکی هستند که ممکن است در مجموعه داده‌های دنیای واقعی صدق نکنند. در این مقاله، یک چارچوب طبقه‌بندی نیمه‌نظارتی جدید پیشنهاد می‌کنیم که رگرسیون خطی محلی را با معیارهای همبستگی نزدیک‌ترین همسایه ترکیب می‌کند تا این محدودیت‌ها را برطرف کند. رویکرد پیشنهادی، طبقه‌بندی مبتنی بر باقیمانده را با معیارهای همبستگی محلی از طریق یک مکانیسم امتیازدهی وزنی ادغام می‌کند و الگوهای عملکردی و وابستگی‌های همسایگی را در نظر می‌گیرد. روش‌های خود را از طریق مطالعات شبیه‌سازی تحت سطح نویز و کاربردها در داده واقعی ارزیابی می‌کنیم. نتایج نشان می‌دهد که روش پیشنهادی به طور مداوم از تکنیک‌های نیمه‌نظارتی سنتی بهتر عمل می‌کند.

واژه‌های کلیدی: یادگیری نیمه‌نظارتی،رگرسیون خطی محلی،نزدیک‌ترین همسایه،همبستگی محلی

*نویسنده مسئول: faboli@math.uk.ac.ir



وب‌اپلیکیشن تعاملی تحلیل گرافیکی و اکتشافی داده‌ها: ابزاری برای ارتقای سواد آماری دانش‌آموزان و دانشجویان

سید مهدی صالحی[۱*]

[۱]دانشگاه نیشابور

چکیده: در عصر داده‌محور امروز، تقویت سواد آماری از سطح دبیرستان تا دانشگاه، ضرورتی انکارناپذیر برای درک پدیده‌های اجتماعی، اقتصادی و علمی است. با این حال، بسیاری از دانش‌آموزان مقطع متوسطه، دانشجویان تازه‌وارد رشته آمار و نیز دانشجویان رشته‌های غیرآماری، در مواجهه‌ی اول با مفاهیم پایه‌ای آمار و دیدارسازی داده‌ها دچار فاصله‌ی میان نظریه و عمل می‌شوند. در پاسخ به این چالش آموزشی، یک وب‌اپلیکیشن تعاملی با عنوان «وب‌اپلیکیشن آموزشی تحلیل گرافیکی و اکتشافی داده‌ها (https://mahdisalehi.shinyapps.io/stat_plots) » مبتنی بر بستر R-Shiny و به زبان فارسی طراحی و پیاده‌سازی شده است. این ابزار آموزشی به کاربران امکان می‌دهد که با مجموعه‌داده‌های واقعی بومی برگرفته از مرکز آمار ایران، داده‌های شناخته شده در حوزه‌ی علم داده و همچنین داده‌های قابل‌ویرایش دلخواه، به دیدارسازی و تحلیل توصیفی داده‌ها بپردازند. قابلیت‌های کلیدی سامانه عبارت‌اند از ترسیم انواع نمودارهای پرکاربرد، محاسبه‌ی شاخص‌های آماری مهم و ویرایش داده‌ها با مشاهده‌ی آنی تغییرات در نمودارها و جداول. همچنین، سامانه با ارائه‌ی پیام‌های راهنما هنگام انتخاب نمودار نامتناسب با نوع متغیر، فرایند یادگیری انتخاب نمودار درست را پشتیبانی می‌کند. نتایج به‌کارگیری این وب‌اپلیکیشن نشان می‌دهد که محیط تعاملی آن می‌تواند یادگیری آمار پایه را به تجربه‌ای عملی و پایدار تبدیل کند و نقش مؤثری در ارتقای سواد آماری دانش‌آموزان و دانشجویان ایفا نماید.

واژه‌های کلیدی: سواد آماری، تحلیل اکتشافی داده‌ها، دیدارسازی داده، آموزش آمار، وب‌اپلیکیشن تعاملی، R-Shiny

*نویسنده مسئول: salehi.sms@neyshabur.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

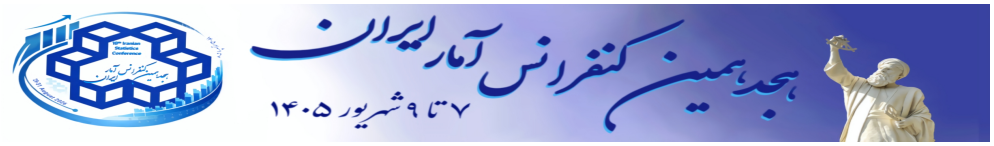
تحلیل مقایسه‌ای رویکردهای یادگیری آماری و یادگیری عمیق در رده‌بندی تصاویر حروف عربی دست‌نویس

زهرا صفایی عارف[۱*]، موسی گلعلی‌زاده[۱]
[۱]دانشگاه تربیت مدرس

چکیده: میراث مکتوب بشری، به‌ویژه نسخه‌های خطی مذهبی، از جایگاه معنوی و علمی برجسته‌ای برخوردار است و کوشش برای دیجیتال‌سازی و تحلیل این منابع، به یکی از جریان‌های اصلی پژوهش‌های معاصر بدل شده است. پیچیدگی ساختار خط عربی و تنوع چشمگیر در شیوه‌های نگارش، چالش‌های اساسی در بازشناسی خودکار متون دست‌نویس ایجاد می‌کند. در مقاله حاضر، رده‌بندی تصاویر حروف عربی با دو رویکرد یادگیری آماری مبتنی بر انتخاب ویژگی و رویکرد یادگیری عمیق مبتنی بر شبکه عصبی پیچشی مورد ارزیابی قرار گرفته است. به‌منظور تبیین رفتار و کارکرد مدل‌ها، از مجموعه‌ای از مصورسازی‌ها و تحلیل‌های تصویری در بخش‌های مختلف مقاله بهره گرفته شد. این تحلیل‌ها در کنار ارزیابی‌های کمی، دیدگاه دقیق‌تری نسبت به نحوه استخراج و تفسیر الگوها در هر دو روش فراهم ساخت. یافته‌ها نشان می‌دهد رویکرد یادگیری آماری در مواجهه با تنوع و بعد بالای داده‌ها با محدودیت‌های قابل توجهی روبه‌روست، حال آنکه مدل مبتنی بر یادگیری عمیق با توانایی پردازش الگوهای چندلایه، دقت به‌مراتب بالاتری ارائه می‌دهد. با وجود این برتری، تفاوت در سطح تفسیرپذیری دو روش، ضرورت انتخاب مدل متناسب با اهداف تحلیلی پژوهش را برجسته می‌سازد.

واژه‌های کلیدی: رده‌بندی،یادگیری عمیق،داده‌های بعد بالا،مصورسازی،پردازش تصویر

*نویسنده مسئول: z.safaei@modares.ac.ir



تلفیق منطق شواهدی با رگرسیون: رویکردی نو در آزمون آماری و استنباط پارامتری

سید امیرحسین طباطبائی شیرازی[۱]، مهدی عمادی[۱*]
[۱]دانشگاه فردوسی مشهد

چکیده: این مقاله به بررسی و بسط چارچوب «رگرسیون شواهدی» به عنوان رویکردی نوین و مقیاس‌پذیر برای جایگزینی استنباط پارامتری کلاسیک می‌پردازد. چالش بنیادین در مدل‌سازی‌های استاندارد آماری، محدودیت در برآورد همزمان و تفکیک‌یافته‌ی عدم قطعیت تصادفی (ناشی از نویز داده‌ها) و عدم قطعیت معرفتی (ناشی از کمبود دانش مدل) است. در این راستا، با بهره‌گیری از توزیع‌های پیشین مزدوج، به‌ویژه توزیع نرمال-وارون-گاما، و ادغام منطق شواهدی با رگرسیون، روشی برای پیش‌بینی مستقیم توزیع‌ها به جای مقادیر نقطه‌ای ارائه می‌گردد. برخلاف روش‌های پرهزینه نمونه‌گیری بیزی، این چارچوب با استفاده از یک تابع زیان و منظم‌ساز شواهدی، شبکه‌های عصبی را برای تخمین عدم قطعیت‌ها آموزش می‌دهد. علاوه بر این، مقاله حاضر به منظور ارزیابی فرض‌های آماری، رویکرد تحلیل شواهدی مبتنی بر «درست‌نمایی نمایه» را جایگزین آزمون‌های فرض کلاسیک و مقادیر احتمال (p-value) می‌نماید. در نهایت، با استفاده از ماتریس اطلاعات فیشر و قضیه ویلکس، هم‌ارزی مجانبی میان فواصل شواهدی (مانند فاصله درست‌نمایی 1/8) و فواصل اطمینان کلاسیک اثبات شده است که نشان‌دهنده استواری، دقت بالا و مقاومت ذاتی این روش در برابر بیش‌برازش و داده‌های پرت می‌باشد.

واژه‌های کلیدی: استنباط شواهدی، رگرسیون پارامتری، عدم قطعیت تصادفی و معرفتی، درست‌نمایی نمایه، قضیه ویلکس، فواصل اطمینان

*نویسنده مسئول: emadi@um.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

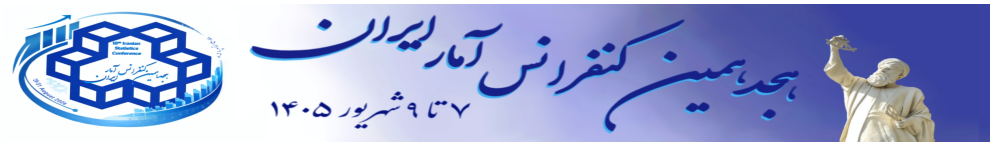
مقایسه عملکرد روش‌های ریج، لاسو و الستیک نت در مدل‌های رگرسیونی با خطاهای AR(1) در حضور هم‌خطی شدید و داده‌های دورافتاده: یک مطالعه شبیه‌سازی

علی ظاهرزاده[1*]
[1]دانشگاه صنعتی جندی شاپور دزفول

چکیده: در بسیاری از مسائل رگرسیونی، فرض‌های کلاسیک مدل خطی نظیر استقلال خطاها و عدم هم‌خطی بین متغیرهای توضیحی برقرار نیست. در این مقاله، عملکرد روش‌های ریج، لاسو و الستیک نت در مدل‌های رگرسیونی تحت هم‌خطی شدید، خطاهای خودهمبسته از نوع AR(1) و حضور داده‌های دورافتاده مورد بررسی قرار می‌گیرد. برای این منظور، یک مطالعه شبیه‌سازی مونت‌کارلو در سطوح مختلف هم‌خطی، خودهمبستگی و درصد داده‌های دورافتاده انجام شده است. معیار مقایسه، میانگین مربعات خطا (MSE) است. نتایج نشان می‌دهد که روش ریج در بسیاری از سناریوها، به‌ویژه در حضور هم‌خطی شدید و داده‌های دورافتاده، عملکرد پایدارتر و خطای کمتری نسبت به لاسو و الستیک نت دارد.

واژه‌های کلیدی: رگرسیون ریج، رگرسیون لاسو، رگرسیون الستیک نت، هم‌خطی، رگرسیون با خطای AR(1)، داده‌های دورافتاده

*نویسنده مسئول: Zaherzadeh@jsu.ac.ir



بررسی استفاده از داده‌های کمی در دیوان‌سالاری ایرانیان: از هخامنشیان تا صفویان

مسعود عجمی[۱*]

[۱]گروه آمار، دانشکده علوم ریاضی، دانشگاه ولی عصر رفسنجان

چکیده: جمع‌آوری و ثبت سیستماتیک داده‌های کمی، برای اهداف حکمرانی یکی از بنیادی‌ترین کارکردهای دولت‌های سازمان یافته در تاریخ بشر بوده است. این موضوع، نه یک پدیده دوره‌ای، بلکه یک ضرورت ساختاری تاریخی بوده است. از الواح گلی تخت جمشید تا طومارهای مالیاتی صفوی، یک خط تکاملی منسجم قابل ردیابی است که در آن هر دوره با افزودن دقت، گستره، و نهادینگی بر میراث پیشین، نظام مالیاتی کارآمدتری ایجاد کرده است. این میراث مستند، به ویژه از منظر تاریخ علم آمار و حکمرانی، پایه‌ای اساسی برای مطالعات بین رشته‌ای آینده فراهم می‌آورد. در ایران، از دوران هخامنشیان تا صفویان، در بازه‌ای بیست وسه قرنی، نهاد دیوان‌سالاری همواره از داده‌های کمی برای مدیریت مالیات، منابع انسانی، تولیدات کشاورزی و ساختار نظامی بهره می‌برد. هدف این مقاله، بررسی تطبیقی سیر تکامل این سیستم در چهار دوره هخامنشی، اشکانی، ساسانی و صفوی بر پایه منابع مستند و قابل تأیید است.

واژه‌های کلیدی: الواح تخت جمشید، داده‌های کمی در حکمرانی، دیوان‌سالاری، سیستم مالیاتی، سلسله‌های ایرانی

*نویسنده مسئول: m.ajami@vru.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

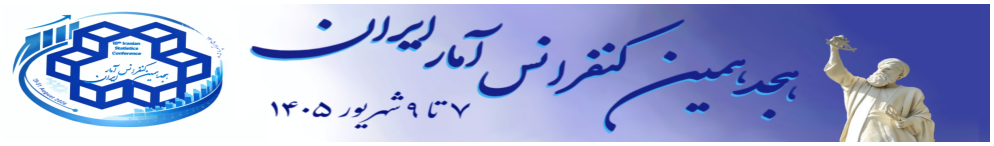
تبیین چارچوب‌های مدل‌سازی برای تحلیل داده‌های ارزیابی لحظه‌ای بوم‌شناختی

حمیدرضا عریضی سامانی [۱*]
[۱]دانشگاه اصفهان

چکیده: روش ارزیابی لحظه‌ای بوم‌شناختی (EMA) با فراهم کردن امکان ثبت پویای تجربه‌های افراد در محیط طبیعی زندگی، تحولی اساسی در روش‌های گردآوری داده‌های طولی ایجاد کرده است. با این حال، تحلیل داده‌های متراکم به دست آمده از این روش، پژوهشگران را با چالش‌های روش‌شناختی پیچیده‌ای روبه‌رو می‌کند. مهم‌ترین چالش در این زمینه، پدیده وابستگی دنباله‌ای است که به دلیل فاصله‌های زمانی کوتاه و پی‌درپی میان ارزیابی‌ها رخ می‌دهد. نادیده گرفتن این پدیده در رویکردهای چندسطحی رایج، مانند مدل‌های خطی سلسله‌مراتبی (HLM) و مدل‌های خطی مختلط (LMM) ، باعث می‌شود برآوردهای آماری با خطای چشمگیری همراه شوند و در نهایت به نتیجه‌گیری‌های نامعتبر بینجامند. مقاله حاضر، با هدف رفع این محدودیت‌ها، به بررسی و توسعه چارچوب‌های پیشرفته مدل‌سازی برای تحلیل داده‌های EMA می‌پردازد. در این پژوهش، ضمن ارائه راهکارهایی برای اصلاح خودهمبستگی در شرایط چالش‌برانگیز فاصله‌های زمانی نامنظم، معادلات رگرسیون مختلط از طریق افزودن اثرهای تعاملی غیرخطی، به‌ویژه تعامل میان مشاهدات تأخیری و فاصله‌های زمانی، توسعه یافته‌اند. نتایج این بررسی نشان می‌دهد که استفاده از روش‌های نوین آماری برای تفکیک دقیق سطوح واریانس و مدل‌سازی درست متغیرهای وابسته به زمان، نه‌تنها یک نیاز محاسباتی است، بلکه پیش‌شرطی ضروری برای افزایش اعتبار و دقت در پژوهش‌های مبتنی بر داده‌های مکرر به شمار می‌رود.

واژه‌های کلیدی: ارزیابی لحظه‌ای بوم‌شناختی (EMA)، مدل‌سازی خطی سلسله‌مراتبی (HLM)، وابستگی دنباله‌ای، مدل‌های خطی مختلط (LMM)، رضایت شغلی.

*نویسنده مسئول: oreizyh@gmail.com



مدل‌سازی مستقیم تابع بروز تجمعی در داده‌های ریسک رقیب با کاربرد در بیماران سوختگی

حبیب الله عزت آبادی پور[۱*]، زهرا شایان[۲]

[۱]دانشگاه علوم پزشکی شیراز

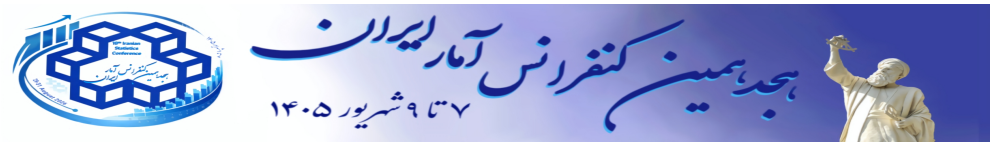
[۲]دانشگاه علوم پزشکی شیرز

چکیده: در تحلیل داده‌های بقا با حضور ریسک‌های رقیب، تابع بروز تجمعی (CIF) ابزاری کلیدی برای برآورد احتمال وقوع هر پیشامد است. در این مطالعه، عملکرد مدل نیمه‌پارامتریک فاین-گری با مدل‌های پارامتریک مستقیم (بر پایه توزیع گامپرتز) در پیش‌بینی CIF مرگ بیماران سوختگی مقایسه شده است. مطالعه شبیه‌سازی با ۱۰۰۰ تکرار و ۶ سناریو طراحی شد. سپس مدل‌ها روی داده‌های ۲۰۴۲ بیمار بستری در بیمارستان سوانح سوختگی امیرالمؤمنین شیراز (سال ۱۴۰۰) پیاده‌سازی شدند. نتایج شبیه‌سازی نشان داد که در سناریوهای با ساختار خاص ریسک‌های رقیب (به‌ویژه خطر صعودی)، مدل مستقیم از نظر شاخص MSE خطای کمتری دارد، اما مدل فاین-گری در اکثر سناریوها از نظر شاخص برایر عملکرد بهتری داشت، بنابراین هیچ‌کدام از دو رویکرد به‌طور مطلق بر دیگری برتری ندارد و انتخاب بین آن‌ها به معیار ارزیابی و هدف پژوهش بستگی دارد. در داده‌های واقعی، همه مدل‌ها نقش معنادار سن، درصد سوختگی، آلبومین، اوره و کراتینین را تأیید کردند. مدل‌های مستقیم به دلیل قابلیت مدل‌سازی کسر درمان‌شدگان، برای اهدافی مانند پیش‌بینی بلندمدت گزینه قابل توجهی به شمار می‌روند، هرچند از نظر دقت کالیبراسیون کوتاه‌مدت (شاخص برایر) مدل فاین-گری در این مطالعه برتری نسبی داشت.

واژه‌های کلیدی: ریسک رقیب،تابع بروز تجمعی،مدل فاین-گری،مدل مستقیم،بیماران سوختگی

---

*نویسنده مسئول: habibezati99@gmail.com



تحلیل روند شاخص‌های مرگ و میر در ایران طی سال‌های **1395-1402** با تاکید بر تاثیر همه‌گیری کووید ۱۹

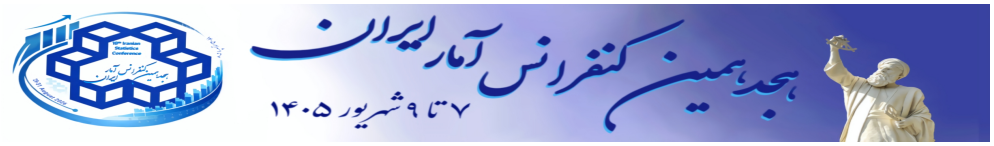
الهام فتحی[۱*]
[۱]مرکز آمار ایران

چکیده: مرگ‌ومیر به عنوان یکی از متغیرهای اصلی تغییرات جمعیت، منعکس‌کننده وضعیت سلامت عمومی و تاب‌آوری نظام‌های بهداشتی است. در میان بحران‌های اخیر، همه‌گیری جهانی کووید ۱۹ به عنوان چالشی بزرگ، شوک سنگینی به ساختارهای جمعیتی وارد کرده است که ابعاد آن نیازمند واکاوی دقیق آماری است. در ایران، این پاندمی با ایجاد نوسانات شدید، تغییرات معناداری را در شاخص‌های حیاتی رقم زد که بررسی روند آن‌ها در بازه زمانی ۱۳۹۵ تا ۱۴۰۲، برای درک اثرات کوتاه‌مدت و میان‌مدت این بحران بر سلامت جامعه ضروری است. این پژوهش با استفاده از داده‌های ثبتی سازمان ثبت احوال کشور و منحصر به جمعیت ایرانی، روند شاخص میزان خام مرگ و میر و میزان مرگ و میر ویژه سنی را تحلیل می‌کند. داده‌های ثبتی پس از ارزیابی و تسطیح با روش‌های علمی، مورد تحلیل قرار گرفتند. نتایج نشان می‌دهد میزان خام مرگ‌ومیر به‌شدت تحت تأثیر کووید ۱۹ قرار گرفته است؛ به‌گونه‌ای که شدت این تأثیر میان دو جنس ودر سنین مختلف تفاوت‌های معناداری داشته است.

واژه‌های کلیدی: تغییرات جمعیت، تسطیح داده‌های فوت، کووید ۱۹، شاخص میزان خام مرگ‌ومیر، شاخص میزان مرگ و میر ویژه سنی.

---

*نویسنده مسئول: fathiel@yahoo.com



شبیه‌سازی و استنباط مبتنی بر درست‌نمایی مدل فضایی دومتغیره با یک رهیافت طیفی

فتانه فخری[۱*]، حسین باغیشنی[۱]، احمد نزاکتی[۱]
[۱]دانشگاه صنعتی شاهرود

چکیده: آمار فضایی شامل مدل‌سازی و تحلیل داده‌هایی است که به مکان و جهت فضایی وابسته‌اند. در سال‌های اخیر، افزایش تقاضا برای اطلاعات و رشد سریع مجموعه‌داده‌های بزرگ و چندمتغیره، چالش‌های جدیدی را در توسعه مدل‌های آماری فضایی ایجاد کرده است. این مسئله به‌ویژه در مدل‌های آمیخته فضایی که متشکل از اثرات ثابت و تصادفی فضایی است، نمود پیدا می‌کند. برای مدل‌سازی اثرات تصادفی، می‌توان از روش‌هایی مبتنی بر توابع پایه استفاده کرد. روش‌های طیفی، از توابع مثلثاتی به‌عنوان توابع پایه استفاده می‌کنند. اما به دلیل پیچیدگی محاسباتی این روش‌ها، به‌ویژه در حالت‌های چندمتغیره، توجه محدودی به آن‌ها شده است. در این مقاله، ساختار کوواریانس منعطف و پرکاربرد مترن، برای میدان‌های تصادفی گاوسی در $\mathbb{R}^2$ در نظر گرفته می‌شود. سپس برآورد پارامترهای مدل با بهره از رویکردهای ماکسیمم درست‌نمایی (ML) و ماکسیمم درست‌نمایی محدودشده (ReML)، با استفاده از مقادیر اولیه مبتنی بر نیم‌واریوگرام تجربی، در چارچوب بهینه‌سازی L-BFGS مورد بررسی قرار می‌گیرد. همچنین عملکرد مدل در یک مطالعه‌ی شبیه‌سازی گسترده ارزیابی می‌شود. یافته‌ها حاکی از آن است که در برآورد ضرایب رگرسیونی، روش ML از خطای کمتری برخوردار است. همچنین در برآورد پارامترهای کوواریانس و پیشگویی متغیرها رویکرد ReML عملکرد بهتری دارد.

واژه‌های کلیدی: مدل آمیخته فضایی، روش‌های طیفی، کوواریانس مترن، ماکسیمم درست‌نمایی محدودشده و ماکسیمم درست‌نمایی.

*نویسنده مسئول: fatanefakhri@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

پیش‌بینی بیماری دیابت با استفاده از رگرسیون لوژستیک و الگوریتم جنگل تصادفی

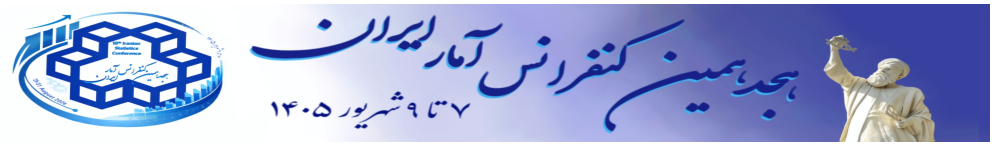
محدثه سادات فراهانی[۱]، ودود کرامتی[۲*]
[۱]دانشگاه علامه طباطبایی
[۲]موسسه عالی آموزش و پژوهش برنامه ریزی

چکیده: دیابت یک بیماری مزمن شایع با بار بهداشتی و اقتصادی قابل توجه در سراسر جهان است. پیش‌بینی و تشخیص زودهنگام می‌تواند به مدیریت مؤثر و پیشگیری از عوارض آن کمک کند. در این مطالعه به بررسی استفاده از رگرسیون لوژستیک و الگوریتم جنگل تصادفی برای پیشبینی دیابت پرداختیم. داده‌های مورد استفاده از مجموعه داده دیابت هندی پیما استخراج شده‌اند که شامل هشت متغیر ورودی و یک متغیر پاسخ دودویی است و تمام افراد موجود در پرونده‌ها زن هستند. پس از پیش‌پردازش داده‌ها و مدیریت مقادیر گمشده، مدل‌ها آموزش داده شدند و عملکرد آن‌ها با استفاده از معیارهایی نظیر دقت، حساسیت، ویژگی و مساحت زیر منحنی ROC ارزیابی شد. نتایج نشان داد که مدل رگرسیون لوژستیک با دقت حدود ۷۶ درصد و مدل جنگل تصادفی با دقت حدود ۷۷ درصد عملکرد قابل قبولی در پیش‌بینی دیابت دارند. این یافته‌ها ظرفیت روش‌های یادگیری ماشین در پیشبینی دیابت را برجسته می‌کند و بهبودهای آینده را از طریق تکنیک‌های انتخاب ویژگی و یادگیری گروهی پیشنهاد می‌دهد.

واژه‌های کلیدی: جنگل تصادفی،داده‌کاوی،دیابت،رگرسیون لوژستیک

---

*نویسنده مسئول: mohadese_farahani@atu.ac.ir



ادغام روش چولسکی با مدل MPGARCH برای مدل‌سازی نوسانات چندمتغیره

فاطمه فرج الهی[۱*]، سعید رضاخواه ورنوسفادرانی[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: در این پژوهش یک چارچوب نوین برای مدل‌سازی نوسان‌پذیری چندمتغیره ارائه می‌شود که در آن ساختار چولسکی با مدل‌های (MPGARCH) گارچ دوره ای آمیخته ترکیب شده است. در مدل پیشنهادی، ماتریس کوواریانس شرطی به‌صورت یک عامل چولسکی پایین‌مثلثی بازنویسی می‌شود و پویایی واریانس‌ها از طریق یک ساختار آمیخته-دوره‌ای مدل‌سازی می‌گردد. این ساختار اجازه می‌دهد که وضعیت های مختلف نوسان‌پذیری با الگوهای دوره‌ای متفاوت به‌طور هم‌زمان حضور داشته باشند و وزن هر وضعیت به‌صورت احتمالات آمیخته تعیین شود. ادغام ساختار چولسکی با فرآیند آمیخته-دوره‌ای موجب تضمین مثبت‌معین‌بودن ماتریس کوواریانس شرطی، کاهش قابل‌توجه تعداد پارامترها و افزایش تفسیرپذیری رفتار نوسان‌پذیری می‌گردد. نتایج تجربی بر روی داده‌های مالی نشان می‌دهد که مدل چلوسکی گارچ دوره ای آمیخته توانایی بالایی در شناسایی الگوهای دوره‌ای، وضعیت های متنوع نوسان‌پذیری و تغییرات ناگهانی همبستگی داشته و نسبت به مدل‌های کلاسیک GARCH–BEKK و نسخه‌های غیردوره‌ای چولسکی عملکرد بهتری در پیش‌بینی، پایداری عددی و برازش مدل دارد. چارچوب ارائه‌شده ابزاری قدرتمند برای تحلیل رفتارهای پیچیده بازار، مدیریت ریسک و مدل‌سازی ساختار وابستگی پویا در محیط‌های دارای الگوهای دوره‌ای و تغییر وضعیت محسوب می‌شود.

واژه‌های کلیدی: Mixture Periodic GARCH–Cholesky، مدل‌های آمیخته-دوره‌ای، نوسان‌پذیری چندمتغیره، کوواریانس شرطی، تحلیل همبستگی پویا، مدل‌سازی مالی

*نویسنده مسئول: karamollahi@aut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

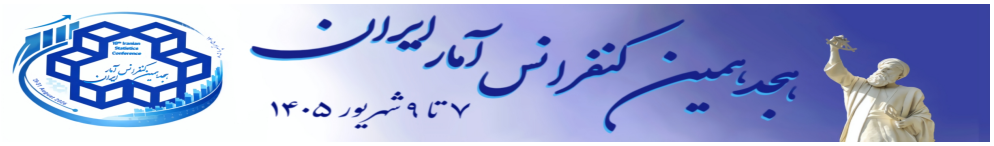
پیش‌بینی ماتریس‌های کوواریانس شرطی بر پایه تجزیه چولسکی اصلاح‌شده و مدل‌سازی ضرایب با شبکه LSTM

فاطمه فرج الهی [۱*]، سعید رضا خواه ورنوسفادرانی [۱]
[۱] دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: تخمین و پیش‌بینی ماتریس کوواریانس شرطی بردار بازده دارایی‌های مالی، سنگ‌بنای مدیریت ریسک مدرن، قیمت‌گذاری دارایی‌های مشتق و بهینه‌سازی سبد سهام محسوب می‌شود. در حوزه اقتصادسنجی مالی چندمتغیره، مدل‌های کلاسیک خانواده GARCH (نظیر BEKK و DCC) علی‌رغم کاربرد گسترده، در مواجهه با ابعاد بالا با چالش‌هایی همچون "نفرین ابعاد" و دشواری در تضمین ساختاری مثبت‌معین بودن ماتریس کوواریانس روبرو هستند. این پژوهش با هدف غلبه بر این محدودیت‌ها، یک چارچوب ترکیبی نوین مبتنی بر تجزیه چولسکی اصلاح‌شده (MCD) و شبکه‌های عصبی حافظه طولانی کوتاه-مدت (LSTM) ارائه می‌دهد. تجزیه MCD با بازپایدارسازی ماتریس کوواریانس به پارامترهای رگرسیونی و واریانس‌های پسماند بدون قید، فضای جستجو را برای یادگیری عمیق بهینه می‌کند. در مقابل، معماری LSTM با بهره‌گیری از گیت‌های حافظه، توانایی منحصربه‌فردی در استخراج وابستگی‌های زمانی غیرخطی و خوشه‌بندی نوسانات نشان می‌دهد. نتایج حاکی از آن است که تلفیق این دو رویکرد، منجر به پیش‌بینی‌هایی پایدارتر و دقیق‌تر در بازارهای مالی متلاطم می‌گردد.

واژه‌های کلیدی: کوواریانس شرطی،تجزیه چولسکی اصلاح‌شده،LSTM،اقتصادسنجی مالی،پیش‌بینی

* نویسنده مسئول: karamollahi@aut.ac.ir



اطلس جرم از توصیف به تحلیل فضایی: بازخوانی ساختار مکانی جرائم در ایران

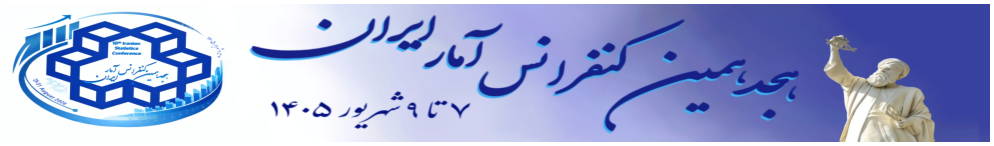
محدثه السادات فرزام مهر [۱*]

[۱]پژوهشکده قوه قضائیه

چکیده: در دهه‌های اخیر اطلس‌های جرایم به‌عنوان یکی از ابزارهای اصلی نمایش الگوهای مکانی جرم مورد استفاده قرار گرفته‌اند. این اطلس‌ها عمدتا بر نرخ وقوع یا تراکم جرائم در واحدهای جغرافیایی استوار بوده و تصویری توصیفی از پراکنش فضایی جرم ارائه می‌کنند. با این حال، این رویکرد غالبا جرم را پدیده‌ای مستقل از زمینه‌های فضایی و روابط میان مناطق در نظر می‌گیرد و از تحلیل وابستگی‌های فضایی میان واحدهای جغرافیایی غفلت می‌ورزد. در مقابل، نظریه‌های نوین جغرافیای جرم و جرم‌شناسی محیطی بر این نکته تأکید دارند که جرم دارای ساختار فضایی بوده و می‌تواند تحت تأثیر ویژگی‌ها و شرایط مناطق همجوار شکل گیرد. بر این اساس، الگوهای جرم در قالب خوشه‌های فضایی و نواحی ناهنجار قابل شناسایی هستند که تنها از طریق تحلیل‌های فضایی قابل آشکارسازی‌اند. پژوهش حاضر با هدف گذار از اطلس‌های صرفا توصیفی به اطلس‌های تحلیلی، به بررسی ساختار فضایی جرائم در سطح مناطق جغرافیایی می‌پردازد. در این راستا از ساختار همسایگی ملکه مانند، شاخص خودهمبستگی فضایی موران جهانی، تحلیل لکه‌های فضایی محلی و شاخص انحراف فضایی استفاده شده است. نتایج مورد انتظار این رویکرد نشان می‌دهد که وجود خودهمبستگی فضایی معنادار، بیانگر ناکارآمدی اطلس‌های مبتنی بر نرخ خام جرم در تبیین واقعی ساختار فضایی جرائم بوده و ضرورت لحاظ روابط فضایی در تحلیل و تفسیر داده‌های جرم را برجسته می‌سازد. این پژوهش چارچوبی برای توسعه اطلس‌های تحلیلی جرم ارائه می‌دهد که علاوه بر نمایش شدت وقوع جرم، روابط فضایی، کانون‌های جرم‌خیز و نواحی ایمن را نیز آشکار می‌سازد.

واژه‌های کلیدی: اطلس جرم، ساختار فضایی جرم، خودهمبستگی فضایی، جغرافیای جرم، سرریز فضایی جرم، جرم‌شناسی محیطی.

*نویسنده مسئول: ma.farzammehr@gmail.com



کاربرد تحلیل حساسیت طرح‌های D-بهینه بیزی در فارماکوکینتیک لیپوآمید

حمیدرضا فریدپور[۱]، حبیب جعفری[۱]، الناز کسانی[۱*]
[۱]دانشگاه رازی کرمانشاه

چکیده: در بسیاری از کاربردهای پزشکی و دارویی، داده‌ها دارای ساختار همبستگی هستند و مدل‌های غیرخطی برای توصیف فرآیندهای جذب و حذف دارو به کار می‌روند. طراحی بهینه آزمایش‌ها در چنین مدل‌هایی به دلیل وابستگی ماتریس اطلاع فیشر به پارامترهای مجهول، با چالش مواجه است. در این مقاله، با استفاده از رویکرد بیزی، تأثیر سه توزیع پیشین یکنواخت، گاما و لوگ‌نرمال بر روی طرح‌های D-بهینه برای مدل فارماکوکینتیک دو محفظه‌ای (داده‌های غلظت لیپوآمید) بررسی می‌شود. سه ساختار همبستگی (نمایی، کرنل نوع ۱ و کوشی نوع ۱) برای جمله خطا در نظر گرفته شده است. معیارهای اطلاع AIC و BIC ساختار نمایی را به عنوان بهترین ساختار انتخاب می‌کنند. نتایج نشان می‌دهد که در ساختار نمایی، هر سه توزیع پیشین طرح‌های تقریبا یکسان و کارایی نزدیک به یک تولید می‌کنند. مطالعه شبیه‌سازی با ۱۰۰۰ تکرار نشان می‌دهد که انتخاب توزیع پیشین نامناسب (گاما با واریانس سه برابر) باعث افزایش شدید واریانس برآورد پارامتر همبستگی می‌شود.

واژه‌های کلیدی: مشاهدات همبسته،توزیع پیشین،ماتریس اطلاع فیشر،فارماکوکینتیک

*نویسنده مسئول: Hamidrezafarid@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

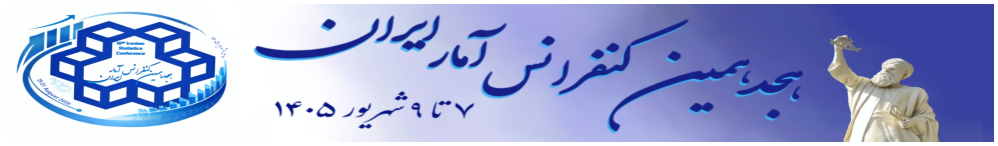
پیش‌بینی ترافیک شبکه تلفن همراه با استفاده از شبکه‌های عصبی

عاطفه قاسمی [۱]*، زهرا رضائی قهرودی [۱]
[۱]دانشگاه تهران

چکیده: با افزایش سریع حجم داده‌ها و گسترش خدمات تلفن همراه، نقش پیش‌بینی ترافیک در مدیریت، بهینه‌سازی و برنامه‌ریزی ظرفیت شبکه‌ها بیش از پیش برجسته شده است. با این حال، پیش‌بینی دقیق و به‌موقع ترافیک به‌دلیل ماهیت پیچیده و وابستگی‌های فضایی-زمانی آن همچنان چالش‌برانگیز است. بسیاری از روش‌های مبتنی بر یادگیری عمیق، عمدتا بر وابستگی‌های موضعی تمرکز دارند و از درک همبستگی‌های پنهان و فراموضعی غافل می‌مانند. در این راستا، ترکیبی از روش‌های یادگیری عمیق، به‌ویژه شبکه‌های عصبی بازگشتی و پیچشی نتایج امیدوارکننده‌ای در مدل‌سازی وابستگی‌های زمانی و فضایی ترافیک از خود نشان داده‌اند. در این مقاله به بررسی و مقایسه دو مدل پیشرفته‌ی شبکه‌های عصبی پیچشی با اتصال متراکم فضایی-زمانی و مدل عمیق دیگری با نام شبکه عصبی میان‌دامنه‌ای فضایی-زمانی پرداخته می‌شود. نتایج نشان می‌دهد که شبکه عصبی میان‌دامنه‌ای فضایی-زمانی با بهره‌گیری از معماری‌های نوین، توانایی بالایی در استخراج الگوهای پیچیده و در نظر گرفتن تأثیر عوامل خارجی در پیش‌بینی ترافیک دارد.

واژه‌های کلیدی: داده‌های تلفن همراه،ترافیک شبکه تلفن همراه،شبکه‌های عصبی

*نویسنده مسئول: at.ghasemi@ut.ac.ir



بهبود فیلترهای کلاسیک کاهش نویز تصویر با استفاده از نمایش فازی شدت پیکسل‌ها

رضا قاسمی نجف ابادی[۱*]، محمدرضا ربیعی[۲]

[۱]دانشگاه پیام نور

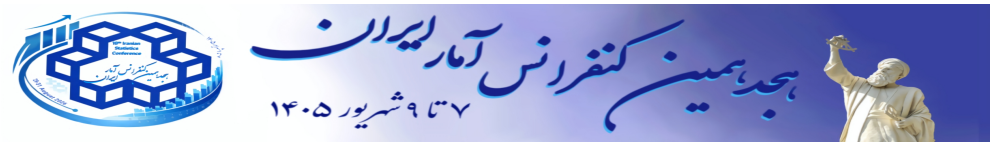
[۲]دانشگاه صنعتی شاهرود

چکیده: نویز ضربه‌ای، به‌ویژه نویز نمک و فلفل، یکی از رایج‌ترین و مخرب‌ترین انواع نویز در تصاویر دیجیتال است که موجب تخریب جزئیات، لبه‌ها و ساختارهای محلی تصویر می‌شود. اغلب فیلترهای کلاسیک کاهش نویز، شدت هر پیکسل را به‌صورت یک مقدار قطعی در نظر گرفته و فرآیند بازسازی را بر اساس روابط عددی دقیق میان پیکسل‌های همسایه انجام می‌دهند. با این حال، در شرایط نویزی، مقدار مشاهده‌شده هر پیکسل الزاما بیانگر مقدار واقعی آن نبوده و دارای نوعی عدم قطعیت است. در این مقاله، رویکردی مبتنی بر فازی‌سازی مقادیر پیکسل‌ها برای افزایش کارایی فیلترهای کلاسیک کاهش نویز ارائه می‌شود. در این چارچوب، هر پیکسل به‌جای یک مقدار قطعی، به‌صورت یک عدد فازی شامل مقدار مرکزی و پهناهای عدم قطعیت چپ و راست مدل‌سازی می‌شود که این پهناها از اطلاعات آماری همسایگی محلی استخراج می‌گردند. هدف اصلی این پژوهش معرفی یک فیلتر جدید مستقل نیست، بلکه نشان دادن این موضوع است که نگاه فازی به مقادیر پیکسل‌ها تا چه اندازه می‌تواند عملکرد روش‌های متداول را بهبود دهد. به‌منظور ارزیابی این ایده، دو روش کلاسیک کاهش نویز به‌عنوان نمونه انتخاب شده و نسخه فازی آن‌ها پیاده‌سازی گردیده است. نتایج آزمایش‌ها بر روی تصاویر استاندارد آلوده به نویز نمک و فلفل در سطوح مختلف نشان می‌دهد که نسخه‌های فازی در معیارهای کمی نظیر PSNR و SSIM و نیز کیفیت بصری، به‌طور معناداری بهتر از نسخه‌های اصلی عمل می‌کنند. یافته‌های این پژوهش نشان می‌دهد که نمایش فازی پیکسل‌ها می‌تواند به‌عنوان یک چارچوب عمومی برای ارتقای بسیاری از فیلترهای کاهش نویز مورد استفاده قرار گیرد.

واژه‌های کلیدی: فیلتر مکانی،فیلتر فازی،کاهش نویز تصاویر،شدت پیکسل فازی

---

*نویسنده مسئول: r.ghasemi62@pnu.ac.ir



سواد آمار مدنی برای شهروندی

آزاده قاهری[۱*]

[۱]پژوهشگاه رویان، مرکز تحقیقات اپیدمیولوژی باروری جهاد دانشگاهی، گروه تحقیقات پایه و جمعیتی بیماری های غیرواگیر، تهران، ایران

چکیده: این روزها در دنیای پیرامون ما بیش از هر زمان دیگری، حجم بسیار زیادی از داده‌های متنوع در دسترس است اما تصمیمهای روزانه ما شهروندان در بسیاری موارد، مبتنی بر شواهد و حقایق نیست. در این میان نقش مدرسان علم آمار این است که کمک کنند سواد آماری شهروندان برای درک مسایل اجتماعی ارتقا یابد. این جنبه از سواد آماری با عنوان «آمار مدنی» به دانش، مهارتها و تمایلاتی گفته می‌شود که شهروندان هر روزه برای درک، ارزیابی نقادانه، گفتگو و تعامل با آمارهای مربوط به مسائل جامعه نیاز دارند. سواد آمار مدنی کمک می‌کند مباحثات عمومی به جای احساسات و تمایلات فردی در بستر حقایق و شواهد شکل بگیرد. نگاه به مسایل آمار مدنی در قالب پدیده‌های چندمتغیره و در نظر گرفتن پیچیدگی این پدیده‌ها می‌بایست در آموزش آمار مدنی گنجانده شود. مدل مفهومی آمار مدنی در سه بعد تعریف و ارزیابی می‌شود: بعد اول: درگیر شدن و کنش، بعد دوم: دانش و بعد سوم: اجرا کردن فرآیندها. این ابعاد مجموعا یازده جنبه دارند که آموزش و ارزیابی آمار مدنی برای شهروندان را در چارچوب مناسبی صورتبندی می‌کند.

واژه‌های کلیدی: آمار مدنی،سواد آماری

---

*نویسنده مسئول: azadeh.ghaheri@gmail.com

آزمون نیمن-پیرسون وزنی برای فرضیه‌های فازی

ابوذر قربانی شیخ نشین[۱*]، رحیم چینی پرداز[۱]، جلال چاچی[۱]
[۱]دانشگاه شهید چمران اهواز

چکیده: تاکنون، روش‌های آزمون فرضیه‌های فازی با داده‌های قطعی مبتنی بر نمونه‌گیری تصادفی مورد توجه بوده‌اند. با این حال، توسعه یک روش جدید برای آزمون فرضیه‌های فازی که در آن‌ها داده‌های قطعی از یک طرح نمونه‌گیری وزنی به دست می‌آیند، بررسی نشده است. در این مقاله، هدف ما ترکیب فرضیه‌های فازی در چارچوب توزیع‌های وزنی به منظور ایجاد ارتباط معنادار بین حوزه‌های منطق فازی و آمار استنباطی کلاسیک است. هدف اصلی این مقاله، توسعه یک چارچوب آماری جدید با تطبیق پایه‌های نظریه آزمون فرضیه‌های فازی با توزیع‌های وزنی است. در شرایطی که روش‌های استنتاج آماری کلاسیک به دلیل «عدم قطعیت» مفاهیم یا «اریبی» در مکانیسم نمونه‌گیری با چالش‌هایی روبرو هستند، این تحقیق به دنبال پر کردن این شکاف نظری است. در این راستا، با تعمیم روش‌هایی مانند آزمون نیمن-پیرسون و تطبیق توابع عضویت فازی (برای مدل‌سازی ابهام در فرضیه‌ها) و توابع وزن (برای اصلاح اریبی در داده‌ها)، یک روش‌شناسی جامع ارائه شده است. این رویکرد یکپارچه، قابلیت تصمیم‌گیری آماری منسجم و واقع‌بینانه‌تری را هنگام مواجهه با داده‌های قطعی ناشی از محیط‌های پیچیده و غیر ایده‌آل فراهم می‌کند. در نهایت، نتایج را با ارائه یک مثال تجزیه و تحلیل می‌کنیم.

واژه‌های کلیدی: فرضیههای فازی،نمونه‌گیری وزنی،توابع عضویت،توابع وزن،آزمون نیمن-پیرسون

*نویسنده مسئول: abouzarstat@gmail.com

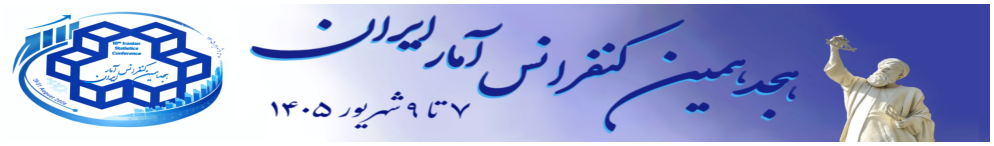
تحلیل ریزش بیمه‌نامه‌های زندگی در یک چارچوب داده‌محور

میترا قنبرزاده[۱*]
[۱]پژوهشکده بیمه

چکیده: هدف از پژوهش حاضر شناسایی عوامل مؤثر بر ماندگاری و ریزش بیمه‌نامه‌های زندگی و مقایسه عملکرد مدل‌های آماری کلاسیک و روش‌های یادگیری ماشین در پیش‌بینی رفتار بیمه‌گذاران است. بدین منظور، داده‌های شبیه‌سازی‌شده مربوط به ۳۰۰۰ بیمه‌گذار بر اساس الگوهای صنعت بیمه ایران تولید و مورد تحلیل قرار گرفت. در این مجموعه داده، متغیرهای دموگرافیک (سن و درآمد)، ویژگی‌های قرارداد (مدت قرارداد، سرمایه فوت و مزایای پایان قرارداد، نحوه پرداخت حق بیمه و ...) و متغیرهای رفتاری–مالی شامل دریافت وام بیمه‌نامه، تأخیر در پرداخت حق‌بیمه و نسبت حق‌بیمه به درآمد در نظر گرفته شد. محاسبات اکچوئری تعهدات نیز بر مبنای جدول مرگ‌ومیر ILT1400 و ساختار نرخ بهره فنی پلکانی انجام شد. برای تحلیل ریزش بیمه‌نامه‌ها، ابتدا مدل مخاطرات متناسب کاکس برآورد گردید و سپس به‌منظور شناسایی روابط غیرخطی و تعامل متغیرها، الگوریتم جنگل تصادفی بقا به کار گرفته شد. نتایج نشان داد هر دو مدل از قدرت پیش‌بینی قابل قبولی برخوردارند و مقدار شاخص هم‌خوانی در حدود 0.71 به دست آمد. همچنین تحلیل اهمیت متغیرها در مدل جنگل تصادفی بقا نشان داد که نسبت حق‌بیمه به درآمد مهم‌ترین عامل مؤثر بر ریزش بیمه‌نامه‌ها است و پس از آن متغیرهای دریافت وام بیمه‌نامه و تأخیر در پرداخت حق‌بیمه بیشترین تأثیر را بر افزایش مخاطره ریزش دارند. یافته‌های این پژوهش می‌تواند در بهبود مدل‌های ارزیابی ریسک، طراحی محصولات بیمه زندگی و سیاست‌های مدیریت ماندگاری بیمه‌گذاران مورد استفاده قرار گیرد.

واژه‌های کلیدی: بیمه زندگی،ریزش مشتری،تحلیل بقا،جنگل تصادفی

*نویسنده مسئول: ghanbarzadeh@irc.ac.ir



مطالعه‌ی نمودارهای کنترل جمع تجمعی و جمع تجمعی دوگانه و ارزیابی عملکرد آنها

طیبه کارکن [۱]*، ابراهیم صالحی [۱]، محمد کاظمی [۱]

[۱]دانشگاه صنعتی بیرجند

چکیده: نمودار کنترل جمع تجمعی یکی از ابزارهای بسیار رایج برای نظارت بر فرآیندهای تولیدی است. تشخیص سریع تغییر در فرآیند (در صورت وقوع)، از ویژگی‌های شاخص این نوع نمودار کنترل به شمار می‌رود. در این مقاله، به مطالعه و بررسی نمودارهای کنترل مبتنی بر جمع تجمعی برای حالتی که مشخصه کیفی از توزیع نرمال پیروی می‌کند، پرداخته‌ایم. همچنین عملکرد این نمودارهای کنترل برای مقادیر مختلف پارامترهایشان، با استفاده از معیارهایی مبتنی بر طول دنباله مورد ارزیابی قرار داده‌ایم.

واژه‌های کلیدی: کنترل کیفیت آماری،طول دنباله،نمودار کنترل جمع تجمعی،شبیه‌سازی مونت کارلو

---

* نویسنده مسئول: tybah.karkne_stu@birjandut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

ارزیابی تطبیقی مدل‌های یادگیری ماشین مبتنی بر بهینه‌سازی بیزی و تفسیرپذیری SHAP برای پیش‌بینی مقاومت جانبی ستون‌های بتن‌آرمه تحت بارگذاری چرخه‌ای

سپهر کاشانی[1*]، محمدحسن دانشوری[1]، برات مجردی[1]
[1]دانشگاه علم و صنعت ایران

چکیده: در این پژوهش، عملکرد چندین الگوریتم یادگیری ماشین شامل DT، RF، KNN، XGBoost، LightGBM و CatBoost برای پیش‌بینی مقاومت برشی آرماتور عرضی در ستون‌های بتن‌آرمه تحت بارگذاری چرخه‌ای مورد ارزیابی قرار گرفت. به‌منظور بهبود عملکرد مدل‌ها، فرآیند تنظیم ابرپارامترها با استفاده از بهینه‌سازی بیزی مبتنی بر Optuna انجام شد. همچنین، از روش SHAP برای تفسیرپذیری مدل‌ها و تحلیل اهمیت متغیرهای ورودی استفاده گردید. نتایج نشان داد که مدل‌های مبتنی بر گرادیان بوستینگ بالاترین دقت پیش‌بینی را ارائه داده و رویکرد پیشنهادی قادر است پیش‌بینی دقیق و قابل‌اعتمادی از مقاومت جانبی ستون‌های بتن‌آرمه فراهم کند.

واژه‌های کلیدی: یادگیری ماشین، مقاومت برشی ستون بتن آرمه، بهینه‌سازی بیزی، SHAP، تفسیرپذیری مدل.

*نویسنده مسئول: mrskykashani@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

به‌کارگیری شبکه‌های آرنولد-کولموگروف در برآورد احتمال نکول اعتباری برپایه مدل ساختاری مرتون

سیدمحمدمهدی کاظمی[۱]، امیر حسین زینال نژاد[۱]*
[۱]دانشگاه خوارزمی

چکیده: احتمال نکول یکی از کمیت‌های کلیدی در مدیریت ریسک اعتباری است و مدل مرتون چارچوبی ساختاری برای تعریف آن ارائه می‌دهد.به‌دلیل غیرخطی بودن نگاشت متغیرهای شرکتی به احتمال‌نکول، تقریب دقیق و پایدار دشوار است.این مقاله از شبکه‌های کولموگروف-آرنولد به‌عنوان یک روش یادگیری ماشین تفسیرپذیر برای تقریب احتمال‌نکول مدل مرتون استفاده می‌کند.

واژه‌های کلیدی: مدل بلک-شولز،مدل مرتون،ریسک اعتباری شبکه عضبی،شبکه‌های کولموگروف-آرنولد

*نویسنده مسئول: Smm.kazemi@khu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

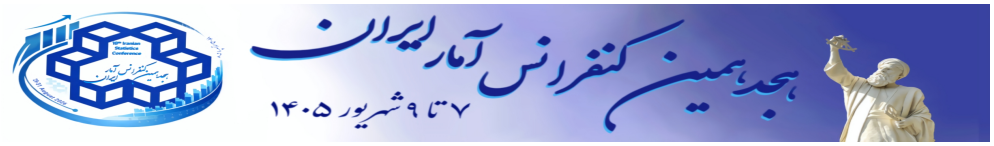
مدل‌سازی ریسک اعتباری با رویکرد احتمالاتی بر اساس شبکه‌های عصبی بیزی

سید محمد مهدی کاظمی[۱]، رضوان دلاور[۱]*
[۱]دانشگاه خوارزمی

چکیده: شبکه‌های عصبی عمیق علی‌رغم عملکرد بالا، معمولا فاقد مدل‌سازی صریح عدم‌قطعیت بوده و در بسیاری از کاربردهای حساس مانند مدل‌سازی ریسک اعتباری دچار بیش‌اعتمادی در پیش‌بینی می‌شوند. در این مقاله، چارچوب شبکه‌های عصبی بیزی به‌منظور مدل‌سازی عدم‌قطعیت وزن‌ها و بهبود قابلیت اطمینان پیش‌بینی‌ها در مسئله ریسک اعتباری بررسی می‌شود. برای تقریب توزیع پسین وزن‌ها از استنتاج واریاسیونی و الگوریتم بیز با پس انتشار استفاده شده و پارامترها با بیشینه‌سازی کران پایین شواهد بهینه‌سازی می‌شوند. همچنین توزیع پیش‌بینی پسین با استفاده از نمونه‌گیری مونت‌کارلو محاسبه می‌گردد. ارزیابی کالیبراسیون مدل با بهره‌گیری از نمودار قابلیت اطمینان و معیار خطای کالیبراسیون مورد انتظار انجام شده است. نتایج نشان می‌دهد مدل بیزی نسبت به روش‌های استاندارد، کالیبراسیون بهتری در ارزیابی ریسک اعتباری ارائه می‌دهد.

واژه‌های کلیدی: شبکه عصبی بیزی،استنتاج واریاسیونی،کران پایین شواهد،کالیبراسیون

*نویسنده مسئول: rezvandelavar6@gmail.com Smm.kazemi@khu.ac.ir,



مدل‌های خطی با اثر ترکیبی از منظر عدم‌تعادل هاردی وینبرگ

مهسا کاظمی[۱*]، سعید رضاخواه ورنوسفادرانی[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: در مطالعات ژنتیکی مرتبط با افراد، مدل خطی اثر مختلط رایج‌ترین روش مورد استفاده است. این مدل ترکیبی از متغیر پاسخ (فنوتیپ) و متغیر کمکی (ژنوتیپ) را درنظر می گیرد. در این گزارش نشان می‌دهیم که برخلاف تصور رایج، این روش استاندارد می‌تواند از دو جهت نسبت به خروج از تعادل هاردی وینبرگ (منظور همان عدم‌تعادل هاردی وینبرگ است) حساس باشد. یک جهت زمانی‌که صفت وراثت‌پذیری به‌عنوان پارامتر مزاحم درنظر گرفته می‌شود. علی رغم اینکه آزمون مربوط به مدل خطی اثر مختلط دارای خطای نوع I کنترل شده است، براورد وراثت‌پذیری می‌تواند درصورت عدم‌تعادل هاردی وینبرگ ، اغلب دارای بیش اریبی (اریبی زیاد) باشد. از جهت دیگر اگر وراثت‌پذیری واقعی در مدل خطی اثر مختلط استفاده شود، آزمون مربوط به مدل خطی اثر مختلط در عدم‌تعادل هاردی وینبرگ، به‌طور قطع اریب می‌شود. در اینجا برخی از دیدگاه‌های تحلیلی را با پشتیبانی نتایج تجربی حاصل از شبیه‌سازی و نیز مطالعات کاربردی ارائه می‌دهیم. یک رویکرد جایگزین، جابه جایی نقش فنوتیپ و ژنوتیپ است که این گونه آزمون‌ها شامل آزمون‌های شبه درستنمایی می‌شود. در پایان ،پس از بررسی روش‌های موجود، رگرسیون معکوس را پیشنهاد می‌دهیم که می‌تواند برای مطالعه عدم تعادل هاردی وینبرگ بکار رود و انعطاف پذیری دیگر مدل پیشنهادی را در براورد فراوانی آللی نمایش می‌دهیم. همچنین رابطه آن را با روش قبلی یعنی بهترین براوردگر نااریب خطی آللی بررسی می‌کنیم.

واژه‌های کلیدی: مطالعه‌ی ارتباط ژنومی، براورد وراثت‌پذیری، رگرسیون، رگرسیون معکوس، عدم‌تعادل هاردی وینبرگ

*نویسنده مسئول: rezakhah@aut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تحلیل داده‌های فضایی–زمانی آلودگی هوای PM10 با مدل چوله کم‌بعد و الگوریتم AHMC

امید کریمی[۱*]، فاطمه حسینی[۱]
[۱]دانشگاه سمنان

چکیده: تحلیل داده‌های فضایی–زمانی آلودگی هوا به دلیل وجود وابستگی‌های پیچیده مکانی–زمانی، ابعاد بالای داده‌ها و رفتارهای غیرنرمال مانند چولگی، از مسائل مهم در مدل‌سازی آماری محیط‌زیستی به شمار می‌رود. مدل‌های فضایی–زمانی چوله، با وجود توانایی مناسب در بازنمایی داده‌های نامتقارن، معمولا به دلیل پیچیدگی ساختار وابستگی، ابعاد بالای پارامترهای پنهان و هزینه محاسباتی زیاد، در تحلیل مجموعه‌داده‌های حجیم با چالش‌های قابل توجهی مواجه هستند. در این مقاله، یک چارچوب بیزی برای تحلیل داده‌های فضایی–زمانی آلودگی هوای PM10 مبتنی بر یک مدل چوله منعطف کم‌بعد پیشنهاد می‌شود. در این رویکرد، برای کاهش هزینه‌های محاسباتی، میدان فضایی–زمانی پنهان به کمک توابع پایه مکانی شعاعی و توابع پایه زمانی فوریه در یک نمایش کم‌بعد بازنویسی می‌شود، به‌گونه‌ای که ساختار وابستگی اصلی داده‌ها حفظ گردد. برای استنباط پارامترهای مدل، از الگوریتم مونت‌کارلوی همیلتونی تطبیقی استفاده شده است که کارایی مناسبی در نمونه‌گیری از توزیع‌های پسین با ابعاد نسبتا بالا فراهم می‌کند. کارایی روش پیشنهادی با استفاده از داده‌های واقعی آلودگی هوای PM10 در آلمان شامل (۵۵۰،۲۵) مشاهده متناظر با (۳۵) ایستگاه پایش طی (۷۳۰) روز مورد بررسی قرار گرفت. نتایج نشان دادند که مدل پیشنهادی توانایی مناسبی در بازنمایی ساختار فضایی–زمانی و چولگی داده‌ها دارد.

واژه‌های کلیدی: میدان تصادفی چوله منعطف،مدل‌سازی فضایی–زمانی،استنباط بیزی

*نویسنده مسئول: omid.karimi@semnan.ac.ir

کاربرد آمار دایره‌ای برای الگوهای زمانی: یک مطالعه موردی در حوزه امنیت

نجیب‌الله کریمی[۱*]، علی دست‌برآورده[۱]، سیدمحسن میرحسینی ابرندآبادی[۱]
[۱]دانشگاه یزد

چکیده: این مطالعه، کاربرد عملی آمار دایره‌ای را در شناسایی الگوهای زمانی پنهان در یک مجموعه داده امنیتی واقعی نشان می‌دهد. با تبدیل زمان وقوع رویدادها به مختصات دایره‌ای و به‌کارگیری معیارهایی توصیفی دایره‌ای، الگوهای معنادار در ساعت‌های شبانه‌روز، روزهای هفته و ماه‌های سال استخراج شدند. نمودارهای دایره‌ای نیز درک بصری روشنی از این الگوها ارائه دادند. یافته‌ها وجود الگوهای چندوجهی (مانند سه اوج در ساعت‌های صبح، ظهر و عصر) و تک‌وجهی (تمرکز در روزها و ماه‌های خاص) را آشکار کرد که می‌تواند مبنای علمی برای تدوین راهبردهای پیش‌گیرانه و تخصیص بهینه منابع امنیتی قرار گیرد. این مطالعه موردی، اثربخشی چارچوب آمار دایره‌ای را به‌عنوان ابزار قدرت‌مند برای تصمیم‌گیری‌های مبتنی بر داده در حوزه‌های حساس تأیید می‌کند. برای استخراج نتایج کاربردی و رسم نمودارها از بسته circular در نرم‌افزار R استفاده شده است.

واژه‌های کلیدی: آمار دایره‌ای،الگوهای زمانی،امنیت،برنامه‌ریزی امنیتی،داده‌های دایره‌ای

*نویسنده مسئول: n.karimi@stu.yazd.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌سازی رگرسیون بتا با پراکندگی متغیر برای نمره ترکیبی سواد سلامت دهان مادران و رفتار سلامت دهان کودکان

امید کریمی پورباصری[۱]*، لیلی تاپاک[۱]، مریم افشاری[۱]، طیب محمدی[۱]
[۱]دانشگاه علوم پزشکی همدان

چکیده: سلامت دهان کودکان به طور چشمگیری تحت تأثیر سواد سلامت مادران و رفتارهای بهداشتی آن‌ها قرار دارد. در پژوهش حاضر، با استفاده از داده‌های ۴۷۰ زوج مادر–کودک در شهر همدان، مدل رگرسیون بتا با پراکندگی متغیر برای تحلیل نمره ترکیبی سواد سلامت مادران و رفتار سلامت دهان کودکان به کار گرفته شد. یافته‌ها نشان داد که تحصیلات دانشگاهی مادر و عادت رفتاری نوع ۱ (تعداد دفعات مسواک زدن کودک در روز) اثر مثبت و معنی‌داری بر نمره ترکیبی دارند. در بخش دقت مدل، منبع کسب اطلاعات نوع ۳ (خانواده) به طور معنی‌داری ناهمگونی پاسخ‌ها را افزایش می‌دهد؛ در حالی که افزایش سن کودک، همگنی پاسخ‌ها را افزایش می‌دهد. همچنین، آزمون نسبت درستنمایی برتری معنادار مدل پراکندگی متغیر را نسبت به مدل با پراکندگی ثابت تأیید نمود.

واژه‌های کلیدی: رگرسیون بتا،پراکندگی متغیر،سواد سلامت دهان،سلامت دهان کودکان

*نویسنده مسئول: tayeb_mohamadi@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

بهبود عملکرد تصفیه پساب نساجی با بهرهگیری از روشهای آماری مبتنی بر طراحی آزمایش

امیرحسین کریمی رهنما[۱*]، مهرداد مظفریان[۱]، نیکا حامدی سنگری[۱]، امیرحسین باقری[۱]
[۱]دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

چکیده: امروزه، با توجه به گسترش روز افزون وابستگی علوم کاربردی به داده، علم آمار نقش بسیار برجستهای در بین سایر علوم پیدا کرده است. یکی از مفاهیم آماری که در مباحث مهندسی توجه بسیاری از پژوهشگران را به علت صرفهجویی در وقت و هزینه به خود جلب کرده است، مفاهیم طراحی آزمایش میباشد. در این پژوهش، به منظور بهبود عملکرد فرایندهای اکسیداسیون پیشرفته که یک فرایند تصفیهی آب و پساب به حساب میآید و همچنین با بهره بردن از روش طراحی آزمایش CCD اقدام به گسترش یک مدل بر مبنای پارامترهای آزمایشگاهی شد. نتایج بدست آمده حاکی از این امر است که مدل سهمی پیشنهادی تطابقپذیری بسیار خوبی (در حدود ۹۰%) با دادههای آزمایشگاهی از خود نشان داده است. همچنین نتایج حاکی از آنالیز ANOVA حاکی از این امر است که پارامتر pH در بین سایر پارامترها دارای بیشتری ن تاثیر گذاری بر سطح پاسخ میباشد. در انتها نیز حالات بهینهی پیشنهادی توسط مدل مورد صحتسنجی آزمایشگاهی قرار گرفت و نتایج آزمایشگاهی تایید کنندهی نتایج مدلسازی هستند که این نشان از توانمندی قابل قبول مدل دارد.

واژههای کلیدی: روش Central Composite Design،فتوکاتالیست،طراحی آزمایش،سنتز هیدروترمال

*نویسنده مسئول: kamir1376@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

اهمیت معیارهای انتخاب نمونه در یادگیری فعال با کاربرد در غربالگری پوکی استخوان

الناز کسانی[۱*]، محمد مرادی[۱]
[۱]دانشگاه رازی کرمانشاه

چکیده: غربالگری همگانی پوکی استخوان به دلیل هزینه بالای آزمایش‌های تخصصی، معمولا با محدودیت بودجه مواجه است. در این پژوهش، کارایی پنج معیار انتخاب نمونه در چارچوب یادگیری فعال مبتنی بر جنگل تصادفی برای کاهش هزینه برچسب‌زنی ارزیابی شده است. روش‌های انتخاب شامل نمونه‌های با انتروپی بالا، حاشیه بزرگتر، کمترین اطمینان، انتخاب تصادفی و معیار پیشنهادی که ترکیب حاشیه-چگالی است در مجموعه داده ۱۵۰۰ نفری و با شیوع ۳۷ درصدی پوکی استخوان مقایسه شده‌اند. معیار پیشنهادی با بهره‌گیری همزمان از عدم قطعیت و چگالی هندسی، به اندازه ۳۰ درصد کاهش هزینه داشته است، یعنی تنها با ۲۲۰ نمونه به دقت حاصل از روش تصادفی با ۳۲۰ نمونه رسیده است. روش‌های مبتنی بر حاشیه و انتروپی بالا نیز عملکردی نسبتا خوب و معادل کاهش هزینه ۲۰ درصدی نسبت به روش تصادفی ارائه دادند. یافته‌های این پژوهش مجددا بر کارایی بالای یادگیری فعال در شرایط محدودیت بودجه دلالت دارند و همچنین زمانی که مجموعه داده بیانگر کل جامعه نیست، چون نرخ شیوع پوکی در جامعه قطعا کمتر از ۳۷ درصد است.

واژه‌های کلیدی: آنتروپی،حاشیه،کمترین اطمینان،شاخص توده بدنی،معیار ترکیبی،چگالی-هندسی

*نویسنده مسئول: e.kasani@razi.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

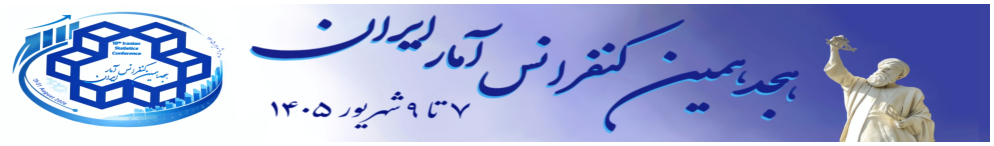
مدل‌سازی ناهمگنی فضایی برپایه‌ی رگرسیون چندکی بیزی با خطای لاپلاس نامتقارن تعمیم‌یافته

کیمیا کلائی[۱*]، فیروزه ریواز[۱]
[۱]دانشگاه شهید بهشتی

چکیده: در بسیاری از مطالعات فضایی، روابط بین متغیر پاسخ و متغیرهای کمکی الگوهای پیچیده‌ای را نشان می‌دهند. رویکردهای متداول، نظیر رگرسیون میانگین (مانند مدل‌های خطی معمولی یا رگرسیون وزنی جغرافیایی) عمدتا بر تغییرات مکانی در میانگین شرطی پاسخ متمرکز هستند و فرض می‌کنند که تأثیر متغیرهای کمکی بر کل توزیع پاسخ یکسان است. این در حالی است که ناهمگنی فضایی ممکن است در بخش‌های مختلف توزیع شرطی پاسخ (مانند چندک‌های پایین، میانه و چندک‌های بالای پاسخ) به شکل‌های متفاوتی ظاهر شود و این تفاوت به سادگی در مدل‌های میانگین محو می‌شود. برای رفع این مشکل، مقاله‌ی حاضر با بهره‌گیری از رگرسیون چندکی، به مدل‌سازی ناهمگنی پنهان در کل توزیع شرطی پاسخ می‌پردازد. به‌طور دقیق‌تر، یک مدل رگرسیون چندکی فضایی بیزی با توزیع خطای لاپلاس نامتقارن تعمیم‌یافته معرفی می‌شود که قادر به مدل‌سازی همزمان چندک‌های مختلف توزیع پاسخ، وابستگی فضایی و چولگی/دم‌سنگینی داده‌ها است. برخلاف توزیع لاپلاس نامتقارن، توزیع لاپلاس نامتقارن تعمیم‌یافته با داشتن پارامتر شکل اضافی، امکان کنترل مستقل چولگی و رفتار دم‌ها را در هر چندک فراهم می‌کند. استنباط بیزی با استفاده از الگوریتم مونت‌کارلوی همیلتونی انجام شده و مدل پیشنهادی به داده‌های اجاره‌بهای ۲۲ منطقه شهر تهران برازش داده می‌شود. نتایج نشان می‌دهد که پارامتر شکل در چندک‌های انتهایی 0.1 و 0.9 معنی‌دار است. همچنین نقشه‌های دامنه‌ی میان‌چارکی آشکار می‌سازد که ناهمگنی فضایی اجاره‌بها در مناطق شمالی تهران به طور معناداری بیشتر از سایر مناطق است.

واژه‌های کلیدی: ناهمگنی فضایی، رگرسیون چندکی، توزیع لاپلاس نامتقارن تعمیم‌یافته، مدل اتورگرسیو فضایی، استنباط بیزی.

*نویسنده مسئول: kimia.kalaei@gmail.com



تجربه کشورها در به‌کارگیری چارچوب جهانی آماری-فضایی: چالش‌ها و راهکارهای پیشنهادی برای ایران

لیدا کلهری [۱]*

[۱] پژوهشکده‌ی آمار

چکیده: اطلاعات فضایی تا حد زیادی منطبق بر اطلاعات محیط طبیعی و یا محیط‌های ساخته شده توسط انسان که زمینه فعالیت‌های انسانی را فراهم می‌کند متمرکز بوده است. اما امروز نیاز به تطبیق داده‌های اجتماعی-اقتصادی در چارچوب داده‌های فضایی در قالب دسته‌بندی مرزهای اداری یا مرزهای آماری ایجاد شده است که همچنان یک مسئله پیچیده است. افزایش آگاهی در رابطه با نقش زیرساختی اطلاعات فضایی در توسعه ملی و دستیابی به اهداف توسعه پایدار، باعث شده است تا این زیرساخت در برخی از کشورهای جهان مانند استرالیا و آفریقای جنوبی با کاربست چارچوب جهانی آماری-فضایی در سطح ملی اجرا شود. این چارچوب، هماهنگی، تبادل، دسترسی و به اشتراک‌گذاری داده و اطلاعات فضایی بین کاربران جامعه فضایی را تسهیل می‌نماید. با توجه به اهمیت موضوع در این مقاله ضمن معرفی اصول چارچوب جهانی آماری-فضایی، تجربه کشورهای استرالیا و افریقای جنوبی در کاربست این چارچوب ارائه خواهد شد. همچنین ضرورت، چالش‌ها و راهکارهای استقرار زیرساخت‌های داده‌های فضایی در ایران مطرح خواهد شد.

واژه‌های کلیدی: اطلاعات فضایی،چارچوب جهانی آماری-فضایی،توسعه پایدار

* نویسنده مسئول: lidakalhori@yahoo.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

طرح D -بهینه برای مدل رگرسیونی خطی ساده فازی

مریم کیانی مرام[۱*]، حبیب جعفری[۱]، مجتبی کاشانی[۲]
[۱]دانشگاه رازی
[۲]دانشگاه گنبد کاووس

چکیده: در این پژوهش، مسئله طراحی آزمایش‌های بهینه برای مدل رگرسیون خطی ساده فازی با هدف افزایش پایداری آماری و دقت برآورد پارامترها در حضور داده‌های مبهم بررسی شده است (در تحقیقات حال حاضر صرفا برای مدل های رگرسیونی غیر فازی این مبحث مورد تحقیق قرار گرفته است). برخلاف رویکردهای کلاسیک که بر فرض قطعی بودن مشاهدات استوار هستند، این مطالعه با بازنگری در ساختار ماتریس اطلاعات فیشر و بهره‌گیری از تفکیک سه‌بانده اعداد فازی مثلثی (چپ، مرکز و راست)، فرمول‌بندی جدیدی از ماتریس اطلاعات به صورت بلوک-قطری ارائه می‌دهد. با به‌کارگیری معیار D -بهینگی و قضیه هم‌ارزی کیفر-ولفویتز، نقاط بهینه طرح استخراج شده و کارایی آنها از طریق تحلیل تابع حساسیت اثبات می‌گردد. نتایج عددی نشان می‌دهد که متمرکز کردن وزن‌های آزمایش در نقاط مرزی فضای طرح، دترمینان ماتریس اطلاعات را بیشینه می‌سازد. همچنین، ارزیابی پایداری طرح‌ها بر اساس معیار D-کارایی تفکیک‌شده تحت فرضیه ی مقایسه طرح قطعی کلاسیک با مؤلفه‌های فازی (شامل بخش مرکز به عنوان بدترین حالت اطلاعاتی و باندهای چپ و راست)، نشان‌دهنده برتری مطلق و پایداری بالای رویکرد فازی پیشنهادی در مواجهه با شرایط عدم قطعیت است.

واژه‌های کلیدی: رگرسیون خطی فازی، طرح D -بهینه، ماتریس اطلاعات فیشر، اعداد فازی مثلثی، قضیه هم‌ارزی عمومی، طرح آزمایش‌ها.

---

*نویسنده مسئول: m.kiani@stu.razi.ac.ir

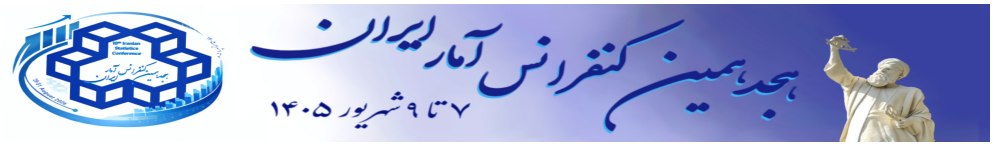
استفاده از رویکردهای تصحیح و توازن بخشی در آمارگیری آمیخته‌مد هزینه و درآمد خانوار برای براورد مجموع هزینه غیرخوراک خانوارهای روستایی

پریسا گوانجی[۱]*، روشنک علی اکبری صبا[۲]
[۱]دانشگاه علامه طباطبایی
[۲]پژوهشکده‌ی آمار

چکیده: از آنجا که در آمارگیری های آمیخته‌مد از مدهای چندگانه برای گردآوری داده‌های واحدهای نمونه استفاده می‌شود، همه واحدها به یک شکل اندازه‌گیری نمی‌شوند و ممکن است اندازه‌گیری آنها هم‌ارز نباشد. هنگام اعمال روش‌های وزن‌دهی استاندارد برای داده‌های آمیخته‌مد، برآوردگرهای حاصل، با ترکیب موزون مشاهدات گردآوری‌شده با مدهای مختلف محاسبه می‌شوند. اگر در واقع هیچ هم‌ارزی اندازه‌گیری بین مدهای گردآوری داده‌ها وجود نداشته باشد، برآوردگرهای به دست آمده به روش معمول شامل مؤلفه‌های خطای اندازه‌گیری افتراقی خواهند بود. در این حالت اگر یکی از مدها به عنوان مد مبنا در نظر گرفته شود، این مؤلفه‌های خطا به عنوان اریبی در نظر گرفته خواهند شد. در این تحقیق، دو رویکرد برای کاهش خطر اریبی برآوردها مورد بحث قرار می‌گیرد. همچنین، با استفاده از روش‌های رگرسیونی، جانهی و پیش‌بینی تلاش می‌شود اریبی اندازه گیری از مشاهدات حذف گردد. سپس روش‌های ارائه‌شده برای پیش‌بینی هزینه ها در طرح هزینه و درآمد خانوار که یک طرح آمیخته‌مد است به کار گرفته می‌شود و نتایج مقایسه می‌شوند. در پایان، برای براورد خطای پیش بینی هزینه‌های غیرخوراک خانوار از روش بوت‌استرپ استفاده می‌شود.

واژه‌های کلیدی: آمارگیری آمیخته‌مد،اریبی اندازه‌گیری افتراقی،کالیبدن مدها،مد مبنا

*نویسنده مسئول: parisa.gavanji@atu.ac.ir



الکندی اولین ریاضی‌دان رمزشکن

فرید مالک قائینی[۱*]

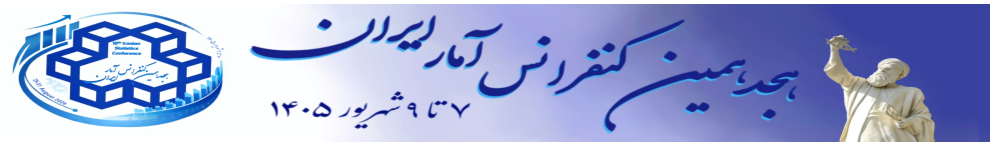
[۱]استاد بازنشسته گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه یزد

چکیده: تاریخنگاری رایج در حوزه رمزنگاری، به‌ویژه در سنت علمی غربی، معمولا خاستگاه روش‌های ریاضی رمزگشایی را به آثار کلود شانون در اواسط قرن بیستم نسبت می‌دهد. این روایت خطی، اگرچه از جهت صورت‌بندی نظری اطلاعات و رمزنگاری مدرن درست است، اما از لحاظ تاریخی، پیشینه‌های چشمگیری را نادیده می‌گیرد که در تمدن اسلامی شکل گرفته‌اند. این سخنرانی با اتکا به اسناد تاریخی و تحلیل محتوای رساله‌های بازمانده، نشان می‌دهد که روایت‌های متعارف در تاریخ رمزنگاری، نیازمند بازنگری اساسی هستند و در واقع ابویوسف یعقوب بن اسحاق الکندی (متوفی ۲۵۶ هجری قمری) در قرن سوم هجری، یعنی نزدیک به یازده قرن پیش از شانون، نخستین اندیشمندی است که به‌صورت نظام‌مند و با استفاده از تحلیل بسامدی (که امروزه به عنوان یکی از بنیادی‌ترین روش‌های آماری در رمزگشایی شناخته می‌شود) به شناسایی ساختار زبان‌های رمز شده پرداخته است.

واژه‌های کلیدی: الکندی، تاریخ علم، تحلیل بسامدی، رمزگشایی تاریخی.

---

*نویسنده مسئول: maalek@yazd.ac.ir



توسعه مدل استوار و تنک ماشین یادگیری کرانگین با استفاده از وزن‌دهی تطبیقی و منظم‌سازی کشسان

صبا مبشر[۱*]، سمانه افتخاری مهابادی[۱]
[۱]دانشگاه تهران

چکیده: ماشین یادگیری کرانگین (Extreme Learning Machine, ELM) به دلیل سرعت بالای آموزش، سادگی ساختار و توانایی مناسب در مدل‌سازی روابط غیرخطی، در مسائل رگرسیون و رده‌بندی کاربرد گسترده‌ای یافته است؛ با این حال، عملکرد این روش در حضور داده‌های پرت و نویزهایی با توزیع ناشناخته به شدت کاهش می‌یابد، زیرا تابع هزینه درجه دوم مورد استفاده در ELM استاندارد نسبت به نمونه‌های غیرعادی حساس است. مدل ماشین یادگیری کرانگین استوار (R-ELM) با استفاده از ترکیب توزیع‌های گاوسی و الگوریتم بیشینه‌سازی امید ریاضی (EM) گامی مؤثر در جهت افزایش پایداری مدل در برابر نویز و داده‌های پرت برداشته است، اما فاقد سازوکار مناسب برای ایجاد ساختار تنک در ضرایب خروجی و کنترل تطبیقی شدت کاهش اثر داده‌های پرت است؛ مدل Elastic R-ELM نیز با افزودن منظم‌سازی کشسان ساختار تنک‌تری ایجاد می‌کند، اما نرخ کاهش اثر داده‌های پرت در آن ثابت و غیرتطبیقی است. در این مقاله، مدل جدیدی با عنوان ماشین یادگیری کرانگین استوار کشسان تطبیقی (AER-ELM) ارائه می‌شود که در آن یک تابع وزن‌دهی تطبیقی مبتنی بر پارامتر توان (p) برای کنترل مستقیم و پیوسته نرخ کاهش وزن نمونه‌های پرت طراحی شده و به‌طور هم‌زمان با منظم‌سازی کشسان در چارچوب مدل ترکیبی گاوسی ادغام گردیده است؛ این ترکیب، تا جای اطلاع نگارندگان، پیش از این در پژوهش‌های مرتبط ارائه نشده و برای نخستین‌بار در این مقاله معرفی می‌شود. الگوریتم بهینه‌سازی مدل نیز بر پایه ترکیب EM و روش کمینه‌سازی بازوزن‌دهی تکراری (IRLS) توسعه یافته است. نتایج آزمایش‌های عددی نشان می‌دهد که مدل پیشنهادی در مقایسه با روش‌های ELM، R-ELM و Elastic R-ELM دارای خطای پیش‌بینی کمتر، پایداری عددی بیشتر و مقاومت بالاتری در برابر داده‌های پرت است؛ علاوه بر این، استفاده از منظم‌سازی کشسان موجب ایجاد ساختار تنک و کاهش پیچیدگی مدل شده است.

واژه‌های کلیدی: ماشین یادگیری کرانگین، یادگیری استوار، منظم‌سازی کشسان، داده‌های پرت، وزن‌دهی تطبیقی، رگرسیون.

---

*نویسنده مسئول: seftekhari@ut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

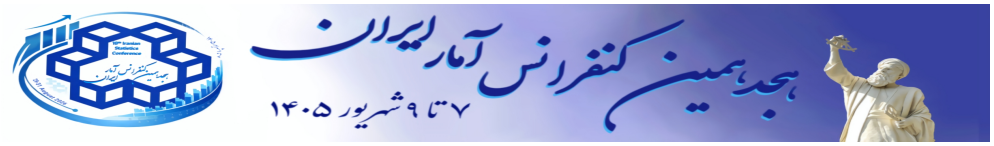
مدل‌سازی بقا با جنگل تصادفی مبتنی بر همبستگی فضایی

کیومرث مترجم [۱*]
[۱]دانشگاه تربیت مدرس

چکیده: تحلیل بقای فضایی به بررسی زمان تا وقوع یک رخداد با در نظر گرفتن موقعیت فضایی آن می‌پردازد و کاربردهای زیادی در اپیدمیولوژی، پزشکی و محیط زیست دارد. مساله اصلی داده‌های بقای فضایی، همبستگی بین مشاهدات مجاور است که فرض استقلال در روش‌های سنتی را نقض می‌کند. در این مقاله سه رویکرد برای تحلیل داده‌های بقای فضایی مورد مقایسه و ارزیابی قرار می‌گیرد. مدل کاکس با اثر تصادفی اتورگرسیو شرطی به عنوان روش پارامتری مرسوم، جنگل تصادفی بقا به عنوان روش ناپارامتری بدون در نظر گرفتن اثرات فضایی و جنگل تصادفی بقای فضایی که در آن همبستگی فضایی بین مشاهدات به مدل اضافه شده است. برای ارزیابی این رویکردها، یک مطالعه شبیه‌سازی انجام شده است که در آن داده‌ها از یک مشبکه 5050 با سه متغیر کمکی و اثرات فضایی حاصل از مدل اتورگرسیو شرطی تولید شده‌اند. ارزیابی با سه معیار شاخص هماهنگی، امتیاز بریر یکپارچه و سطح زیر منحنی وابسته به زمان انجام شده و برای جلوگیری از نشت اطلاعات فضایی از اعتبارسنجی بلوکی فضایی استفاده شده است. نتایج نشان می‌دهد که جنگل تصادفی بقای فضایی با اضافه کردن مختصات فضایی عملکرد بهتری نسبت به سایر مدل‌ها دارد. این یافته حاکی از آن است که افزودن موقعیت فضایی به مدل‌های یادگیری ماشین می‌تواند جایگزین مؤثری برای مدل‌سازی پیچیده اثرات تصادفی فضایی باشد.

واژه‌های کلیدی: تحلیل بقای فضایی،اتورگرسیو شرطی،جنگل تصادفی بقا،شاخص هماهنگی

*نویسنده مسئول: k.motarjem@modares.ac.ir



الگوریتم بازگشتی تقریبی برای برآورد پارامترهای مدل خودرگرسیون فضایی یک‌طرفه با روند تصادفی

آزاده مجیری[۱*]
[۱]دانشگاه زابل

چکیده: مدل‌های خودرگرسیون فضایی یک‌طرفه از ابزارهای مهم برای مدل‌سازی داده‌های شبکه‌ای هستند. در بسیاری از کاربردها، داده‌ها علاوه بر وابستگی فضایی، دارای روند تصادفی نیز هستند که بر برآورد پارامترهای مدل اثر می‌گذارد. در این مقاله، یک الگوریتم بازگشتی تقریبی برای برآورد پارامترهای مدل خودرگرسیون فضایی یک‌طرفه مرتبه اول در حضور روند تصادفی فضایی ارائه می‌شود. روش پیشنهادی با استفاده از ساختار بازگشتی روند، به‌صورت متناوب مولفه روند و پارامترهای خودرگرسیونی را برآورد می‌کند. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم پیشنهادی از پایداری عددی مناسبی برخوردار بوده و در سناریوهای مختلف وابستگی فضایی، برآوردهای قابل قبولی برای پارامترهای مدل فراهم می‌کند.

واژه‌های کلیدی: مدل‌های خودرگرسیون فضایی یک‌طرفه، روند تصادفی فضایی، الگوریتم بازگشتی تقریبی، برآورد پارامتر، داده‌های شبکه‌ای.

---

*نویسنده مسئول: amojirii@uoz.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

کاربرد طرح مرکب مرکزی در مدل‌سازی رگرسیونی و بهینه‌سازی باززایی گیاه آویشن دنایی تحت تأثیر اکسین‌های

مهدی محب الدینی[۱*]، نسترن مهام[۱]، سوسن پناهی گلستان[۲]
[۱]دانشگاه محقق اردبیلی
[۲]دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران

چکیده: سامانه‌های زیستی در کشت درون‌شیشه‌ای، به دلیل وجود برهم‌کنش‌های غیرخطی میان تنظیم‌کننده‌های رشد گیاهی، نیازمند بهره‌گیری از رویکردهای آماری پیشرفته است. در پژوهش حاضر، روش سطح پاسخ (RSM) مبتنی بر طرح مرکب مرکزی وجه‌مرکز (CCD، (α = ۱)) در قالب طرح کاملا تصادفی و بدون بلوک‌بندی، به‌منظور مدل‌سازی و کمی‌سازی اثرات مستقل و متقابل غلظت‌های اندول-۳-استیک اسید (IAA؛ (۰-۳)میلی‌گرم بر لیتر) و نفتالین استیک اسید (NAA؛ (۰-1.25) میلی‌گرم بر لیتر) بر باززایی گیاه آویشن دنایی (Thymus daenensis Celak.) به‌کار گرفته شد. نتایج تجزیه واریانس (ANOVA) نشان داد که مدل‌های رگرسیونی درجه دوم برای هر دو متغیر پاسخ شامل وزن کالوس (P=0.001) و درصد باززایی (P<0.001) در سطح احتمال ۵ درصد معنی‌دار بودند. همچنین، آزمون عدم برازش (Lack of Fit) عدم معنی‌داری را نشان داد (P>0.05)که بیانگر کفایت و اعتبار مدل‌های پیشنهادی بود. اثر متقابل (IAA × NAA) نیز بر وزن کالوس (P=0.009) و درصد باززایی (P=0.023) معنی‌دار ارزیابی شد. بر اساس نمودارهای سطح پاسخ، غلظت‌های بالاتر IAA همراه با مقادیر پایین NAA (کمتر از (0.2)میلی‌گرم بر لیتر)، موجب افزایش باززایی و کنترل کالوس‌زایی شدند؛ در حالی که غلظت‌های بالای NAA اثر بازدارنده بر فرآیند باززایی داشتند.

واژه‌های کلیدی: طراحی آزمایش،روش سطح پاسخ،طرح مرکب مرکزی،مدل‌سازی رگرسیونی،اکسین

---

*نویسنده مسئول: nasim67maham@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

نگاهی به سرشماری در ایران

غلامرضا محتشمی برزادران[۱*]
[۱]دانشگاه فردوسی مشهد

چکیده: سابقه سرشماری در دنیا به هزاران سال قبل از میلاد مسیح بازمی‌گردد. مطالعه اسناد و مدارک تاریخی نشان می دهد بابلی ها ، مصری ها و چینی ها در سالهای 3000-4500-۵۰۰۰ قبل از میلاد سرشماری انجام دادند. این سرشماری‌ها اطلاعاتی از قبیل سن، جنس و شغل افراد جمع‌آوری شده است. در امپراطوری ایران باستان سرشماری های عمومی جمعیت برای دریافت مالیات سرانه وتعیین شماره مردان قادر به جنگ انجام می شده است. قدیمی ترین سند تاریخی که در آن به سرشماری نفوس اشاره شده است؛ در فصل مهاجرت از کتاب تورات است یکی از ابتدایی ترین اسناد سرشماری مربوط به دوره هخامنشی(کوروش و داریوش)،برای برنامه ریزی و مسائل نظامی و تعیین مالیات سرانه صورت گرفت. ساسانیان توجه بیشتری به آمار داشتند و امور مالی، کشاورزی و صنعتی و بازرگانی خود را بر اساس آمارها و اطلاعاتی که ماموران سرشماری جمع آوری می کردند، به انجام می رساندند. بعدها توسط امپراتوری‌های ایران، روم و همچنین در یونان باستان، روش‌هایی برای جمع‌آوری و تنظیم داده‌ها ابداع شد؛ با این حال ویلیام پتی در قرن هفدهم نخستین کسی بود که از روش‌های آماری برای تجزیه و تحلیل داده ها استفاده کرد و به نوعی پایه‌گذار دانش نوین آمار شد. نخستین‌بار در سال ۱۲۲۹ شمارش جمعیت، منازل و ساختمان‌های تهران انجام شد. اگر بخواهیم از دقیق‌ترین سرشماری ایران در دوره قاجار یاد کنیم، باید به سراغ سرشماری دقیق و خانه‌به‌خانه زین‌العابدین‌میرزای قاجار، دانش‌آموخته مدرسه دارالفنون در مشهد برویم که،کتاب «تعداد نفوس ارض اقدس یا مردم مشهد قدیم » معرف آمار خانه به خانه محلات شش‌گانه مشهد در سال ۱۲۹۵ ـ ۱۲۹۶ هجری قمری است. اولین سرشماری جمعیتی ایران به روش نوین در سال ۱۲۴۶ شمسی (البته در سطحی محدود) درزمان ناصرالدین شاه و همزمان با وزارت اعتضادالسلطنه بر وزارت علوم و معارف، توسط مهندس عبدالفخار، معلم ریاضی مدرسه دارالفنون به عمل آمد.در این سرشماری تهران به پنج محله ارگ، عوددلاجان، چاله میدان، سنگلج و بازار تقسیم شد و همه محله های خارج از شهر نیز در یک طبقه قرار گرفت. درآن زمان جمعیت تهران۱۵۵ هزار و ۷۳۶ نفر برآورد شده بود.درسال ۱۲۷۷ شمسی در چهارمین سال سلطنت مظفرالدین شاه نیز از ابنیه تهران آمارگیری به عمل آمد و با ایجاد تشکیلات جدید بلدیه در سال ۱۳۰۱ شمسی سرشماری دیگری در تهران صورت گرفت. تا این زمان سازمان خاصی که مسوولیت جمع آوری اطلاعات جمعیتی و تمرکز بخشیدن به آن ها را برعهده داشته باشد. درکشور وجود نداشت. درسال ۱۲۹۷ هجری شمسی به منظور ثبت وقایع چهارگانه ، اداره ثبت احوال کشور تاسیس شد. با ثبت اطلاعات مرتبط با تولد ، فوت ، ازدواج و طلاق توسط اداره مذکور، ضرورت اطلاع از جمعیت کشور و تعیین سازمانی که به جمع آوری این اطلاعات بپردازد مورد توجه قرار گرفت و منجر به آن شد که در سال ۱۳۰۳ هجری شمسی آیین نامه ای به تصویب برسد و براساس آن مسئولیت جمع آوری و متمرکز کردن آمارهای مورد نیاز به عهده وزارت کشور گذاشته شد. در خرداد ماه سال ۱۳۱۸ هجری شمسی اولین قانون سرشماری در مجلس شورای ملی تصویب شد. در اجرای این قانون سرشماری نفوس از دهم اسفندماه همان سال در شهر تهران و در سال ۱۳۱۹ و ۱۳۲۰ هجری شمسی در ۳۳ شهر کشور به تدریج به اجرا درآمد. دراسفند ماه سال ۱۳۳۱ هجری شمسی سازمان همکاری آمار عمومی تشکیل شد و در فروردین ماه سال ۱۳۳۲ هجری شمسی ، قانون آمار و سرشماری به تصویب رسید. دراین سال اداره آمار و سرشماری از اداره کل آمار و ثبت احوال منتزع و به سازمان همکاری آمار عمومی ملحق شد. به این ترتیب برای اولین بار سازمانی که منحصرا وظیفه جمع آوری آمار را به عهده داشت به وجود آمد که در سال ۱۳۳۴ هجری شمسی به اداره آمار عمومی ، وابسته به وزارت کشور ، تغییر نام یافت و این اداره درسال ۱۳۳۵ هجری شمسی اولین سرشماری عمومی نفوس را در کل کشور به اجرا درآورد. با تاسیس اداره آمار عمومی ، فعالیتهای آماری وارد مرحله جدیدی شد و همه ساله طرحهای گوناگون آماری در زمینه های

مختلف اجتماعی – اقتصادی به اجرا درآمد نیاز روزافزون دستگاه‌های برنامه ریزی کشور به آمار و اطلاعات و ضرورت همکاری بسیار نزدیک سازمان اصلی تولیدکننده آمار با دستگاه برنامه ریزی ، موجب شد تا براساس قانون ۱۳۴۴ هجری شمسی ، اداره آمار عمومی از وزارت کشور جدا و با نام مرکز آمار ایران به سازمان برنامه و بودجه وابسته شود که اولین رییس آن دکتر عباس جامعی(با انتخاب هیات وزیران) و دکتر خواجه نوری مشاور عالی مرکز شد و متولی سرشماری های زیادی شده است. اداره کردن شئون مختلف جامعه در عصر حاضر یک مسئله علمی شده است حتی سیاست مداران نیز در برنامه ها و قضاوت هایشان باید متکی به علم باشند و به همین سبب گروه کارشناس را همواره همراه دارند و کارشناس آمار نقش کلیدی در این امر دارد و دکتر خواجه نوری از معدود ممتاز خبرگانی در امر آمار است که باید از وی زمامداران بهره می گرفتند ولی او به عدم توجهشان در بدو امر وقعی نگذاشت و در مهیا کردن زیرساخت و فکر و دانش آماری در مملکت برای تربیت نسل جدید قدمهای استواری برداشت اولین آمارگیری نمونه‌ای به توصیه سازمان ملل در سال ۱۳۳۶ و پس از آن درتهران و دماوند به سر پرستی استاد خواجه نوری و حمایت متولیان وقت انجام شد که نیاز به مرکز آموزشی برای آموزش آمار برای آموزش کارکنان مرکز آمار و سایر ادارات تحت عنوان موسسه آموزش عالی آمار با هدف تربیت سه دوره فوق دیپلم، لیسانس (آمارشناسی درجه ۲) و فوق لیسانس (آمارشناسی درجه یک) تحت مدیریت دکتر خواجه نوری بنا نهاده شد. نظرایشان این بود که ضمن رعایت فهم و پایه خوب ریاضی با این آماری که می‌خوانید باید توان انجام کاری عملی آمار مفید درجامعه را داشته باشید. آقای دکتر خواجه نوری در سال ۱۳۳۵ اعلام داشت که یک سر شماری را که مخارج آن سی میلیون تخمین زده شده بود با یک‌دهم هزینه با آمارگیری نمونه‌ای سریع تر و ارزانتر و نزدیک به واقعیت می توان یافت که افراد ذینفع سد عملی شدن این ایده شدند ولی تلاش وی برای محقق شدن روشهای علمی برای تحلیل و برنامه ریزی در مملکت بر اساس آمار همچنان به راه خود ادامه داد. اولین سرشماری بر اساس این قانون سال ۱۳۱۸ هم در تابستان همان سال در کاشان انجام شد. بعد از آن شهرهای تبریز، مشهد، کرمان، شیراز، اردبیل، یزد، اصفهان، همدان، کرمانشاه، رشت و بندر انزلی(پهلوی) و در نهایت تهران، با همان ترتیب سرشماری شد. تا شهریور ۱۳۲۰، سرشماری در ۳۲ شهر ایران انجام شده بود، اما بحران سیاسی و اشغال ایران در جنگ جهانی دوم، توقف سرشماری را موجب شد. .در ۱۳۲۸، وقتی محمدرضا شاه دیگر جاگیر شده بود، تلاش‌هایی برای از سرگیری سرشماری جمعیت در تهران و ایران انجام و بودجه‌هایی به این کار اختصاص داده شد هزینه های سنگین سرشماری گاه دقت را نیز تحت الشعاع قرار می داد که ضرورت نمونه گیری ودر زمان حاضر توجه به آمارهای ثبتی را مورد توجه قرار داده که متولیان را به این سمت هدایت می کند در دنیای تکنولوژی پیش رو و سرعت تحولات ضرورت نگاه جدید به یافتن اطلاعات جامعه در سایه روشهای نوین و کم هزینه را می طلبد. هر چه تساهل و تغافل کنیم عقب ماندگی ازدنیای علم بیشتر خواهد شد دنیای وجود داده های حجیم چاره اندیشی عمیق را می طلبد و از جمله ملزومات و تبعات گام در آینده گذاشتن خواهد بود و تحولی را در آموختن و آموزاندن و یافتن مباحث نو، حل معضلات در راستای روند تکنولوژی و بهبود زندگی مردم را رقم خواهد زد. به‌طور کلی گذشته فقط در پرتو زمان حال برای ما قابل درک است آموزه های تاریخی بسیاری نکات را برجسته می‌نماید و ذهن را برای کشف و مدل نمودن حقایق آماده‌تر می‌نماید.

واژه‌های کلیدی: تاریخچه سرشماری ایران، قانون سرشماری، آمارگیری نمونه‌ای، مرکز آمار ایران، داده‌های ثبتی، کاربرد آمار در برنامه‌ریزی.

---

*نویسنده مسئول: grmohtashami@um.ac.ir

آمار کاربردی در عصر هوش مصنوعی: فرصت‌ها، چالش‌ها و چشم‌اندازها

عادل محمدپور [۱*]
[۱]گروه آمار، دانشگاه صنعتی امیرکبیر

چکیده: گسترش شتابان فناوری‌های هوش مصنوعی، ضمن دگرگونی بسیاری از حوزه‌های علمی و کاربردی، جایگاه آمار کاربردی را نیز بازتعریف کرده است. این مقاله با رویکردی مروری، نقش نوین آمار کاربردی را در عصر هوش مصنوعی بررسی می‌کند و نشان می‌دهد که تفکر و روش‌های آماری چگونه می‌توانند در کنار الگوریتم‌های یادگیری ماشینی در تحلیل داده‌ها، ارزیابی مدل، تفسیر نتایج و پشتیبانی از تصمیم‌گیری علمی نقش‌آفرین باشند. در این پژوهش، ضمن مرور تحول تاریخی آمار کاربردی و پیوند آن با نیازهای علمی، ظرفیت روش‌های کلاسیک آماری در مدل‌سازی، استنباط، کنترل عدم قطعیت و اعتبارسنجی نتایج بررسی می‌شود. همچنین، زمینه‌های نوظهوری مورد توجه قرار می‌گیرند که در نتیجه رشد مه‌داده‌ها، توان محاسباتی پیشرفته و الگوریتم‌های هوشمند شکل گرفته‌اند و افق‌های تازه‌ای برای توسعه آمار کاربردی گشوده‌اند. محور اصلی مقاله، رابطه دوسویه میان آمار کاربردی و هوش مصنوعی است. از یک سو، اصول آماری می‌توانند برای سنجش قابلیت اعتماد، استواری، تعمیم‌پذیری، تفسیرپذیری و کارایی مدل‌های هوش مصنوعی به کار روند؛ و از سوی دیگر، هوش مصنوعی می‌تواند ابزارهای نوینی برای مدل‌سازی پیچیده، تحلیل داده‌های بزرگ‌مقیاس، کشف الگوهای پنهان و خودکارسازی فرایندهای آماری فراهم سازد. این مقاله نشان می‌دهد که نقش آمارشناسان در عصر هوش مصنوعی از تحلیل داده فراتر رفته و به راهبری تصمیم‌سازی‌های داده‌محور، طراحی روش‌های ارزیابی سامانه‌های هوشمند و تفسیر نتایج پیچیده گسترش یافته است. از جمله پیامدهای مهم این پژوهش، برجسته‌سازی ضرورت بازنگری در برنامه‌های درسی رشته آمار و هم‌سوسازی آن‌ها با نیازهای نوظهور عصر هوش مصنوعی است.

*نویسنده مسئول: adel@aut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

جنگل تصادفی آگاه از ناهمگنی در پیش‌بینی حساسیت دارویی

پریا محمدی[۱]، سکینه دهقان[۱*]
[۱]دانشگاه شهید بهشتی

چکیده: پیش‌بینی پاسخ سلول‌های سرطانی به داروهای ضدسرطان یکی از مسائل مهم در پزشکی شخصی‌سازی‌شده است. وجود ناهمگنی میان انواع سرطان باعث می‌شود مدل‌های سراسری تفاوت‌های اختصاصی گروه‌های توموری را نادیده بگیرند، در حالی که مدل‌های جداگانه برای هر نوع سرطان نیز به دلیل محدود بودن تعداد نمونه‌ها با کاهش دقت مواجه می‌شوند. جنگل تصادفی آگاه از ناهمگنی، با آموزش یک جنگل تصادفی سراسری و انتخاب تطبیقی درخت‌های سازگار با هر نمونه، میان استفاده از کل داده‌ها و حفظ ساختار ناهمگن آن‌ها تعادل برقرار می‌کند. در این مقاله، ساختار آماری و الگوریتمی این روش بررسی شده و نتایج گزارش‌شده روی داده‌های بیان ژن و پاسخ دارویی پایگاه‌های CCLE و GDSC بازخوانی می‌شود. نتایج نشان می‌دهد که این روش، به‌ویژه در شرایطی که میانگین پاسخ دارویی انواع سرطان متفاوت است، می‌تواند خطای پیش‌بینی را نسبت به جنگل تصادفی معمول کاهش دهد.

واژه‌های کلیدی: پزشکی شخصی‌سازی‌شده، جنگل تصادفی، جنگل تصادفی آگاه از ناهمگنی، حساسیت دارویی، یادگیری ماشین.

*نویسنده مسئول: sa_dehghan@sbu.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تحلیل و بهبود توان عملیاتی سیستم جراحی قلب باز با استفاده از رویکرد شبیه‌سازی صف‌بندی

طیب محمدی[۱]*، قدرت اله روشنایی[۲]، جواد فردمال[۱]، مجید صادقی فر[۳]، بابک منافی[۱]، حسین محجوب[۱]
[۱]دانشگاه علوم پزشکی همدان
[۲]دانشگاه علوم پزشکی اراک
[۳]دانشگاه بوعلی سینا

چکیده: مدیریت صف و بهینه‌سازی ظرفیت در بخش‌های مراقبت پس از جراحی از چالش‌های اساسی در ارتقای کارایی و بهره‌برداری مؤثر از منابع بیمارستانی محسوب می‌شود. در این راستا، هدف پژوهش برآورد و تحلیل شاخص توان عملیاتی سیستم جراحی قلب باز با بهره‌گیری از رویکرد شبیه سازی صف‌بندی است. برای این منظور، عملکرد بخش جراحی قلب بیمارستان فرشچیان همدان به‌عنوان یک سیستم صف‌بندی چندمرحله‌ای و با استفاده از شبیه‌سازی گسسته‌پیشامد مدل‌سازی شد. در این مدل، بیماران کاندید جراحی قلب به‌عنوان مشتری و تخت‌های ICU و POW به‌عنوان سرویس‌دهنده در نظر گرفته شدند. نتایج شبیه‌سازی طی دوره ۸۰۰ روزه و با لحاظ محدودیت تعداد جراحی‌های روزانه نشان داد که بخش POW گلوگاه اصلی سیستم است و افزایش تعداد تخت‌های این بخش موجب بهبود توان عملیاتی سیستم می‌شود. نتایج نشان داد که نرخ اشغال تخت‌های POW و مدت زمان انتظار-بستری در ICU از شاخص‌های کلیدی عملکرد سیستم هستند و با انتخاب ترکیب بهینه ظرفیت دو بخش، می‌توان بدون نیاز به توسعه پرهزینه ICU، عملکرد کلی سیستم را بهبود داد. چارچوب پیشنهادی ابزار مؤثری برای برنامه‌ریزی ظرفیت و بهینه‌سازی تخصیص منابع، از جمله تخت‌ها و جراحان، در سیستم‌های جراحی قلب فراهم می‌کند.

واژه‌های کلیدی: توان عملیاتی، توزیع کاکسی فاز نوع، شبیه‌سازی مونت‌کارلو، جراحی قلب، سیستم صف‌بندی.

*نویسنده مسئول: mahjub@umsha.ac.ir

مقایسه روش‌های الگوریتم عقبگرد در مدل‌های جمعی

نیایش محمدی[۱*]، احد ملک زاده[۱]
[۱]دانشگاه خواجه نصیرالدین طوسی

چکیده: الگوریتم عقبگرد، هسته اصلی برازش مدل‌های جمعی و جمعی‌تعمیم‌یافته است. با وجود سادگی مفهومی، نحوه پیاده‌سازی جزئیاتی مانند «مرکزی کردن توابع» و «به‌روزرسانی عرض از مبدأ» تأثیر مستقیمی بر یکتایی، تفسیرپذیری و همگرایی برآوردها دارد. در این مقاله، چهار رویکرد متفاوت از الگوریتم عقبگرد بر اساس ترکیب اعمال یا عدم اعمال قید مرکزی کردن توابع و به‌روزرسانی یا ثابت نگهداشتن عرض از مبدأ معرفی و از نظر ریاضی و آماری مقایسه شدهاند.

واژه‌های کلیدی: مدل‌های جمعی،روش‌های برآورد،الگوریتم عقبگرد،اسپلاین

*نویسنده مسئول: niyayeshmohamadi77@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

بررسی خاصیت جمع‌پذیری واریانس فازی مبتنی بر آلفا-برش

محدثه محمدپور[۱*]، احمد نزاکتی رضازاده[۱]، محمدرضا ربیعی[۱]
[۱]دانشگاه صنعتی شاهرود

چکیده: در نظریه احتمال کلاسیک، واریانس یکی از مهم‌ترین معیارهای پراکندگی برای متغیرهای تصادفی است. با گسترش مفاهیم به محیط فازی، تعریف واریانس برای متغیرهای تصادفی فازی نیازمند رویکردهایی مبتنی بر آلفا-برش و اتحاد تجزیه زاده است. در این مقاله، ابتدا تعریف دقیقی از متغیر تصادفی فازی ارائه شده و واریانس فازی آن با استفاده از آلفا-برش‌ها تحت شرایط خاصی معرفی می‌گردد. سپس مفهوم استقلال دو متغیر تصادفی فازی بر اساس استقلال خانواده آلفا-برش‌های آن‌ها تعریف می‌شود. لم اصلی مقاله نشان می‌دهد که برای دو متغیر تصادفی فازی مستقل با شکل‌های فازی از نوع مثلثی، آلفا-برش واریانس فازی مجموع آن‌ها زیرمجموعه‌ای از مجموع آلفا-برش واریانس‌های فازی هر یک است. به عبارت دقیق‌تر،رابطه $\widetilde{\mathrm{Var}}_{[\alpha]}(\tilde{X}_1 + \tilde{X}_2) \subseteq \widetilde{\mathrm{Var}}_{[\alpha]}(\tilde{X}_1) + \widetilde{\mathrm{Var}}_{[\alpha]}(\tilde{X}_2), \ \forall \alpha \in [0,1]$ برقرار است. که در آن نشان‌دهنده جمع فازی براساس اصل تعمیم است. این نتیجه با ارائه یک مثال عددی برای متغیرهای تصادفی فازی مثلثی نامتقارن مورد تأیید قرار گرفته است. یافته‌های این مقاله می‌تواند مبنایی برای توسعه روش‌های برآورد و آزمون فرضیه‌ها در آمار فازی باشد.

واژه‌های کلیدی: متغیر تصادفی فازی،واریانس فازی،آلفا-برش

*نویسنده مسئول: rabie.mr @gmail.com

ابرروندهای نوظهور در نظام آماری: بازاندیشی چرخه تولید آمار رسمی

عباس مرادی[۱*]
[۱]پژوهشکده‌ی آمار

چکیده: نظام آمار رسمی در دهه اخیر با تحولات بنیادینی مواجه شده است؛ تحولاتی که از پیشرفت فناوری، گسترش منابع داده‌ای نوظهور و رشد تقاضا برای آمارهای زمانتزدیک سرچشمه می‌گیرند. مدل GSBPM همچنان چارچوب مرجع تولید آمار رسمی است، اما شیوه اجرای مراحل آن با ورود داده‌های ثبتی، داده‌های سکوها، یادگیری ماشین، پردازش خودکار، APIهای عمومی و رویکردهای جدید حفظ محرمانگی دگرگون شده است. در این مقاله، تحول مفهومی و عملی در مراحل گردآوری، پردازش، تحلیل، انتشار و ارزیابی بررسی می‌شود. همچنین مروری نظاممند بر ادبیات بین‌المللی شامل تجربه کشورهای عضو ESS، OECD و UNSD ارائه شده است. در پایان، پیامدهای این ابرروندها برای آینده نظام آماری ایران تحلیل می‌شود.

واژه‌های کلیدی: آمار رسمی،داده‌های ثبتی،مدرنسازی،محرمانگی،انتشار

*نویسنده مسئول: a_moradi@sci.org.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

تحلیل حساسیت و ارزیابی عدم قطعیت در سیستم‌های توصیه‌گر با رویکرد بیزی و یادگیری ماشین

مبینا مزین اصل[۱]، بهناز عباسی شوازی[۱*]، سیدمحسن میرحسینی ابرندآبادی[۱]، جمال زارع پور احمدآبادی[۱]
[۱]دانشگاه یزد

چکیده: این پژوهش چارچوبی ترکیبی مبتنی بر رویکرد بیزی و یادگیری ماشین برای تحلیل حساسیت و ارزیابی عدم قطعیت در سیستم‌های توصیه‌گر ارائه می‌دهد. با استفاده از شبکه‌های بیزین، رگرسیون فرایند گاوسی و سایر مدل‌های بیزی، عدم قطعیت پارامترها و داده‌ها مدل‌سازی شد. تحلیل حساسیت با روش‌های موریس (غربالگری) و شاخص‌های سوبل (مبتنی بر واریانس) بر پارامترهای کلیدی مانند تعداد عوامل نهفته، نرخ یادگیری، پارامتر نظم‌دهی و آستانه شباهت انجام گرفت. نتایج بر روی داده‌های LastFM، MovieLens و Amazon Reviews نشان داد مدل‌های بیزی بهبود 12 الی 18 درصدی در پوشش پیش‌بینی (PICP) و 25 درصدی در PINAW نسبت به روش‌های غیربیزی دارند. پارامتر «تعداد عوامل نهفته» حساس‌ترین (0.42) و «آستانه شباهت» کمترین (0.08) تأثیر را داشت. آزمون ویلکاکسون نشان داد اختلاف عملکرد مدل‌ها در سطح 0.05 معنادار است. این چارچوب قابلیت اطمینان توصیه‌ها را کمی کرده و راه را برای بهینه‌سازی هدفمند فراپارامترها هموار می‌سازد.

واژه‌های کلیدی: سیستم‌های توصیه‌گر، تحلیل حساسیت، رویکرد بیزی، شبکه بیزین، شاخص سوبل، بهینه‌سازی فراپارامتر.

*نویسنده مسئول: bhnaz.abc@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

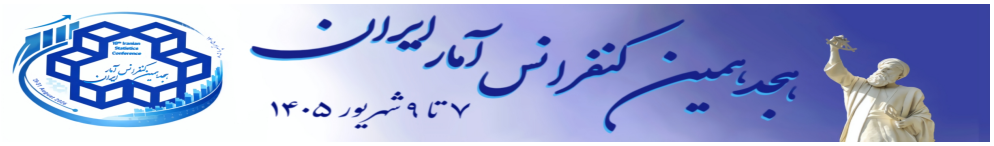
رده‌بندی داده‌های چنبره‌ای با استفاده از ستیغ‌های چگالی ناپارامتری

میثم مقیم بیگی[۱*]، محمد مهدی صابر[۲]
[۱]دانشگاه خوارزمی
[۲]مرکز آموزش عالی اقلید

چکیده: رده‌بندی داده‌های چنبره‌ای به دلیل ساختار نااقلیدسی و تناوبی بودن آن‌ها همواره با چالش روبروست. روش‌های رده‌بندی سنتی نیز اغلب در حفظ ویژگی‌های ذاتی فضاهای چنبره‌ای ناکام هستند. از این رو در این مقاله، رویکردی نوین ارائه شده است که در آن از برآورد ستیغ چگالی ناپارامتری برای رده‌بندی داده‌های چنبره‌ای استفاده می‌شود. همچنین برای یافتن ستیغ چگالی ناپارامتری از حاصلضرب هسته‌های فون‌میزس استفاده می‌شود. روش پیشنهادی بر روی مجموعه داده‌های شبیه‌سازی‌شده و واقعی چنبره‌ای اعمال شده است و عملکرد رده‌بندی به‌روش جدید با روش‌های استاندارد در یادگیری ماشین مقایسه خواهد شد.

واژه‌های کلیدی: داده‌های چنبره‌ای،ستیغ چگالی،رده‌بندی

*نویسنده مسئول: m.moghimbeygi@khu.ac.ir



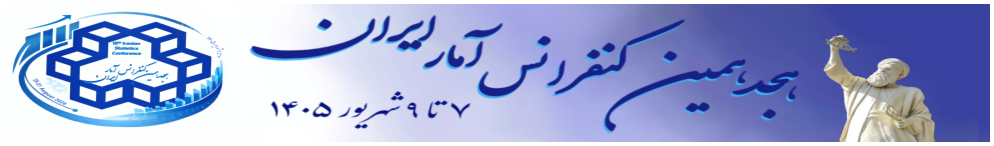
تحلیل آماری و رتبه‌بندی عوامل مؤثر بر انتخاب فروشگاه‌های مرکز تبادل کتاب از دید مشتریان (مطالعه موردی: کلان‌شهر تهران)

ابوالقاسم مقیمی اسفندآبادی[۱*]، داود شیشه بری[۱]
[۱]دانشگاه یزد

چکیده: در بازار رقابتی و پویای امروزی، شناخت دقیق نیازها و اولویت‌های مشتریان یکی از مهم‌ترین عوامل موفقیت کسب‌وکارها به شمار می‌رود. این تحقیق از نظر هدف، کاربردی و از لحاظ ماهیت و روش گردآوری داده‌ها، توصیفی-پیمایشی است. جامعه آماری پژوهش شامل کلیه مشتریان فعال فروشگاه‌های مرکز تبادل کتاب شهر تهران است که جمعیت آن حدود ۵۰،۰۰۰ نفر برآورد شده است. با استفاده از فرمول کرجسی و مورگان، حجم نمونه ۳۸۱ نفر تعیین گردید که به روش نمونه‌گیری تصادفی ساده انتخاب شدند. روایی و پایایی پرسشنامه تأیید شده و تجزیه‌وتحلیل داده‌ها با استفاده از نرم‌افزار SPSS و آزمون‌های آماری توصیفی، استنباطی و آزمون رتبه‌بندی فریدمن انجام گرفت. نتایج پژوهش ضمن شناسایی میزان تأثیر هر یک از متغیرهای آمیخته بازاریابی، اولویت‌بندی این عوامل را از دیدگاه مشتریان مشخص می‌سازد.

واژه‌های کلیدی: انتخاب فروشگاه تبادل کتاب،اس پی اس اس،قیمت‌گذاری،تنوع محصول،تبلیغات

*نویسنده مسئول: ab.moghimi@email.ac.ir



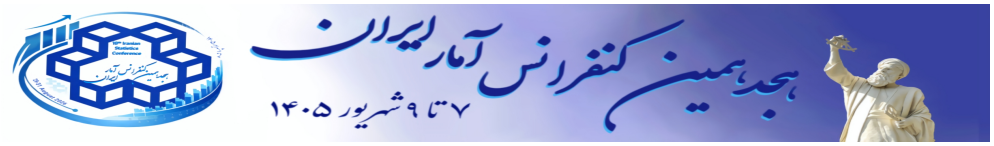
تحلیل مولفه‌های اصلی داده‌های تابعی تنک با همبستگی فضایی

مهرداد ملکی[۱*]، امید خادم نوع[۱]، علی محمدیان مصمم[۱]، عباس رسولی[۱]
[۱]دانشگاه زنجان

چکیده: تحلیل مؤلفه‌های اصلی تابعی یکی از روش‌های مهم در تحلیل داده‌های تابعی است که برای کاهش بعد و استخراج الگوهای غالب تغییرات داده‌ها به‌کار می‌رود. در بسیاری از مسائل کاربردی، داده‌ها به‌صورت توابعی ثبت می‌شوند که علاوه بر ماهیت تابعی، دارای وابستگی فضایی نیز هستند. در روش‌های کلاسیک تحلیل داده‌های تابعی عموما فرض می‌شود که توابع تصادفی از یک‌دیگر مستقل هستند این فرض در داده‌های فضایی برقرار نیست. در این مقاله، رهیافتی مفهومی برای تحلیل مؤلفه‌های اصلی تابعی در حضور همبستگی فضایی برای داده‌های تابعی، مورد مطالعه قرار می‌گیرد.

واژه‌های کلیدی: تحلیل داده‌های تابعی،همبستگی فضایی،تحلیل مولفه‌های اصلی تابعی،آمار فضایی

*نویسنده مسئول: mehrdadmlk9024@gmail.com



مدل‌های رگرسیون چندسطحی استوار برای داده‌های نسبتی

سحر موسوی چهره برق[۱*]، مجید جعفری خالدی[۱]
[۱]دانشگاه تربیت مدرس

چکیده: برای مدل‌سازی ارتباط میان یک متغیر پاسخ و متغیرهای تبیینی می‌توان از مدل‌های رگرسیونی استفاده نمود. در برخی مسائل کاربردی با حالت‌هایی مواجه می‌شویم که داده‌ها به صورت خوشه‌ای، هستند. این داده‌ها در چند سطح و در هر سطح متغیرهای متفاوتی را لحاظ می‌کنند. به منظور تحلیل داده‌هایی با ساختارهای پیچیده و سلسله‌مراتبی، می‌توان از مدل‌های رگرسیون چندسطحی استفاده کرد. در مدل‌های رگرسیون چندسطحی کلاسیک، اغلب فرض بر نرمال بودن توزیع متغیر پاسخ است؛ در حالی‌که در کاربردهای عملی، متغیر پاسخ ممکن است مقادیر خود را در بازه $(۰,۱)$ اختیار کند. در این حالت برای مدل‌بندی متغیر پاسخ، معمولا از مدل رگرسیون چندسطحی بتا استفاده می‌شود. اما این مدل در حضور داده‌های دورافتاده عملکرد مناسبی از خود نشان نمی‌دهد. هدف این مقاله، معرفی مدل‌های رگرسیونی چندسطحی استوار در برابر داده‌های دورافتاده از طریق بکارگیری تعمیم‌هایی از توزیع بتا است. برای استنباط بیزی مدل‌های معرفی شده، از الگوریتم‌های مونت کارلو زنجیر مارکوفی استفاده می‌شود. مدل‌های معرفی شده، از طریق آزمایش‌های شبیه‌سازی بررسی و ارزیابی می‌شوند.

واژه‌های کلیدی: استنباط بیزی، داده‌های نسبتی، داده‌های دورافتاده، توزیع بتای واریانس آماسیده، مدل‌های رگرسیونی استوار، رگرسیون چندسطحی.

*نویسنده مسئول: sahar.moosavi@modares.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

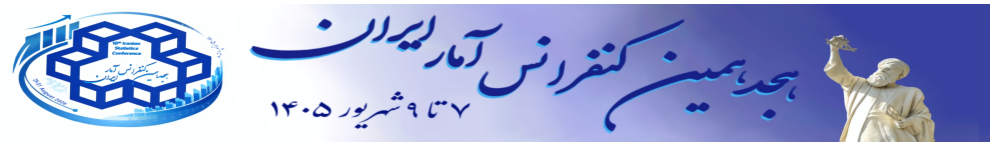
مدل‌سازی فراوانی-شدت خسارت در بیمه خودرو در حضور متغیرهای توضیحی

مرجان سادات موسوی ماسوله[۱*]، احد ملک‌زاده[۱]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی

چکیده: مدل‌سازی ریسک خسارت از مهم‌ترین مباحث آمار بیمه و مدیریت ریسک به شمار می‌رود و نقش کلیدی در قیمت‌گذاری بیمه‌نامه‌ها و محاسبه ذخایر فنی ایفا می‌کند. در این پژوهش، ریسک خسارت از طریق مدل مرکب فراوانی-شدت مورد بررسی قرار گرفته است؛ به‌طوری‌که فراوانی خسارت با توزیع پواسن و شدت خسارت با توزیع گاما مدل‌سازی شده‌اند. به منظور ارزیابی رفتار برآوردگرهای ماکزیمم درستنمایی پارامترهای مدل، مطالعه‌ای مبتنی بر شبیه‌سازی مونت‌کارلو در سناریوهای مختلف انجام شده است. نتایج نشان می‌دهد که برآوردگرهای مورد استفاده از ویژگی‌های آماری مطلوبی برخوردار بوده و مدل پواسن-گاما قادر است رفتار ریسک خسارت و توزیع خسارت کل را به‌طور مناسبی توصیف کند. این یافته‌ها می‌تواند مبنایی برای توسعه مدل‌های کمی ریسک در مسائل بیمه‌ای و اکچوئری باشد.

واژه‌های کلیدی: مدل‌سازی فراوانی-شدت خسارت، توزیع پواسن، توزیع گاما، شبیه‌سازی، حداکثر درستنمایی.

*نویسنده مسئول: marjanmoosavi1999@gmail.com



بهبود روش‌های جورسازی آماری با استفاده از مدل‌های خطی اثرات آمیخته

علیرضا موفقی اردستانی[۱*]، زهرا رضایی قهرودی[۱]
[۱]دانشگاه تهران

چکیده: علاقه سازمان‌های ملی آماری به ترکیب داده‌های حاصل از آمارگیری‌ها و سرشماری‌ها، با هدف تولید مجموعه‌داده‌های غنی‌تر و پوشش متغیرهای بیشتر، در سال‌های اخیر افزایش یافته است. جورسازی آماری یکی از مهم‌ترین رویکردهای ادغام منابع داده است که از طریق متغیرهای مشترک، امکان ترکیب مجموعه‌داده‌های متعلق به یک جامعه هدف را فراهم می‌کند. با این حال، در بسیاری از کاربردها، منابع داده مورد استفاده هم‌مقیاس نیستند و ممکن است در واحدهای آماری متفاوت، مانند افراد، خانوارها یا بنگاه‌ها جمع‌آوری شده باشند که این موضوع فرایند جورسازی را با چالش مواجه می‌کند. در آمارگیری نیروی کار، اطلاعات مرتبط با درآمد و هزینه خانوارها به‌صورت مستقیم گردآوری نمی‌شود، در حالی که وجود این اطلاعات برای بسیاری از تحلیل‌های اقتصادی و اجتماعی ضروری است. ازاین‌رو، در این پژوهش، جورسازی آماری آمارگیری نیروی کار و آمارگیری هزینه و درآمد خانوار با هدف جانهی درآمد خانوار در داده‌های نیروی کار بررسی شده است. برای این منظور، عملکرد چهار روش شامل رگرسیون خطی، جنگل تصادفی، مدل نیمه‌پارامتری مبتنی بر ترکیب رگرسیون خطی و جنگل تصادفی، و مدل رگرسیون خطی با اثرات آمیخته مقایسه شده است. ارزیابی عملکرد روش‌ها با استفاده از آزمون هم‌توزیعی کولموگروف–اسمیرنوف و نرخ پذیرش فرض هم‌توزیعی در سطح استان و به تفکیک مناطق شهری و روستایی انجام شده است. نتایج نشان می‌دهد که مدل رگرسیون خطی با اثرات آمیخته، در مقایسه با سایر روش‌ها، تطابق بیشتری میان توزیع درآمد واقعی و درآمد جانهی‌شده ایجاد کرده و دقت بالاتری در فرایند جورسازی آماری ارائه می‌دهد.

واژه‌های کلیدی: جورسازی آماری، جورسازی چندمقیاسی، مدل خطی اثرات آمیخته، آمارگیری نیروی کار، آمارگیری هزینه و درآمد خانوار

*نویسنده مسئول: alireza.movaffagh@ut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

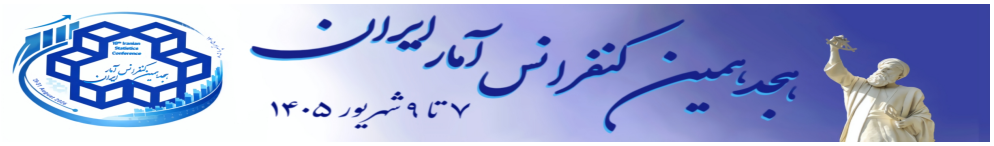
بهبود کیفیت پاسخ‌گویی پزشکی بدون تنظیم دقیق مدل: یک رویکرد بازیابی‌محور مبتنی بر UMLS و Cross-Encoder

احمدرضا مؤمنی[۱*]، مریم عبدالعلی[۱]
[۱]دانشگاه صنعتی خواجه نصیرالدین طوسی

چکیده: مدل‌های زبانی بزرگ در سال‌های اخیر توانایی بالایی در درک و تولید متن نشان داده‌اند، اما به‌کارگیری آن‌ها در حوزه‌های تخصصی و حساس مانند پزشکی همچنان با چالش‌هایی نظیر توهم، اتکا به دانش ایستا و کاهش قابلیت استنادپذیری همراه است. یکی از راهکارهای مؤثر برای کاهش این محدودیت‌ها، استفاده از معماری تولید تقویت‌شده با بازیابی است که در آن مدل زبانی به‌جای تکیه صرف بر دانش پارامتری، از اسناد بازیابی‌شده از یک منبع دانش معتبر نیز بهره می‌برد. با این حال، عملکرد این معماری به کیفیت مؤلفه بازیابی وابسته است. در این مقاله، یک رویکرد بازیابی‌محور برای پاسخ‌گویی پزشکی ارائه می‌شود که بدون نیاز به تنظیم دقیق مدل مولد، بر بهبود لایه بازیابی تمرکز دارد. روش پیشنهادی شامل گسترش معنایی پرسش با استفاده از بازیابی اولیه با SBERT و FAISS، و بازرتبه‌بندی نتایج به کمک رمزگذار متقاطع است؛ سپس پاسخ نهایی توسط مدل LLaMA 3.2 تولید می‌شود. ارزیابی انجام‌شده نشان می‌دهد که این معماری در مقایسه با مدل پایه مبتنی بر بازیابی تک‌مرحله‌ای، عملکرد بهتری در کیفیت بازیابی و پاسخ نهایی دارد. یافته‌های پژوهش بیانگر آن است که در سامانه‌های پاسخ‌گویی پزشکی مبتنی بر LLM، بهبود هدفمند بازیابی می‌تواند بدون تنظیم دقیق پرهزینه مدل، راهکاری عملی و مؤثر برای ارتقای کیفیت پاسخ‌ها فراهم آورد.

واژه‌های کلیدی: پاسخگویی پزشکی،مدل‌های زبانی بزرگ،RAG،UMLS،رمزگذار متقاطع،بازرتبه‌بندی،Reranking،Cross-Encoder

*نویسنده مسئول: ahmadrezamoemeni@gmail.com



اثر فاحش پاسخ نیابتی بر نتایج آمارگیری جهانی کار اجباری و چند روش برخورد با آن

فرهاد مهران[۱*]

[۱]دفتر بین‌المللی کار

چکیده: از سال ۲۰۰۵ میلادی، برای پنجمین بار دفتر بین‌المللی کار در حال تهیه برآورد جهانی کار اجباری می‌باشد. در حالی که کار اجباری به صورت بردگی و رعیتی از زمان‌های دور همیشه وجود داشته است، امروزه این پدیده در قسمت‌های مختلف دنیا به صورت‌های گوناگون دیگر مجددا ظاهر شده است. از نظر آماری، طبق تعاریف بین‌المللی، شخصی مشمول کار اجباری است که تحت تهدید مجازات غیرداوطلبانه وادار به انجام کاری شده باشد. «تحت تهدید مجازات» و «کار غیرداوطلبانه» دو شرط اساسی کار اجباری هستند. کار اجباری، پدیده بسیار نادری است که حدود ۶ در هزار فرد در سن کار، به آن می‌پردازند. به منظور بدست آوردن تعداد نمونه کافی، پوشش طرح نمونه‌گیری دفتر بین‌المللی کار شامل فرد پاسخگوی خانوار نمونه و همچنین تمامی اعضای خانواده مستقیم او می‌باشد. خانواده شامل پاسخگو، همسر، پدر و مادر، فرزندان و برادران و خواهران تنی است. در این طرح آمارگیری، اطلاعات مربوط به اعضای خانواده به‌طور مستقیم ار طریق پاسخگو به طور نیابتی به‌دست می‌آید. در این مقاله، در مرحله اول، اثر پاسخگویی نیابتی بر نتایج آمارگیری در کشورهای مختلف و ثبات آن در زمان نشان داده می‌شود. سپس، چند روش برای برخورد با اثر پاسخگویی نیابتی از جمله تصحیح وزن در قالب نمونه‌گیری غیرمستقیم ارائه می‌شود.

واژه‌های کلیدی: کار اجباری،پاسخ نیابتی،تصحیح وزن،نمونه‌گیری غیرمستقیم

*نویسنده مسئول: mehranxfarhad@yahoo.com

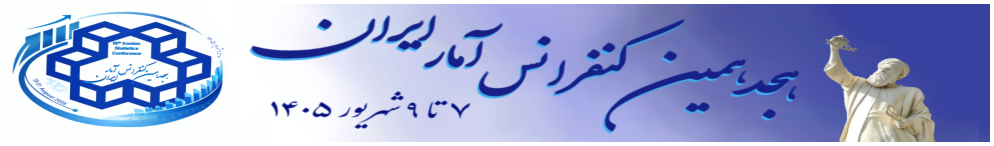
سواد آماری: تجربه کشورها و سازمانهای بین‌المللی

یاسر مهرعلی[۱*]
[۱]مرکز آما

چکیده: در این ارائه، ابتدا به پروژه بین‌المللی سواد آماری (ISLP) به عنوان یک ابتکار پیشگامانه از سوی انجمن بین‌المللی آموزش آمار (IASE) پرداخته می‌شود که با هدف دسترسی جهانی به منابع و ارتقای سواد آماری از سال ۲۰۰۱ آغاز به‌کار کرده است. سپس، نقش گروه اقتصادی و اجتماعی سازمان ملل برای آسیا و اقیانوسیه (UNESCAP) در برگزاری وبینارهای تخصصی با هدف یکپارچه‌سازی آمار رسمی در برنامه‌های درسی دانشگاهها و کاهش شکاف میان دانش نظری و نیازهای عملی بررسی می‌شود. بخش اصلی ارائه، به مطالعه تطبیقی چهار کشور فیلیپین، اندونزی، نیوزلند و فیجی اختصاص دارد و هر یک از آنها را از زوایای مختلف شامل ساختار آموزشی، همکاری بینهادی، محتوای دورهها و چالش‌های پیشرو بررسی می‌کند. در پایان، درسآموختههایی برای طراحی و اجرای برنامه‌های ارتقای سواد آماری در نظام آماری کشور ارائه می‌شود.

واژه‌های کلیدی: آموزش آمار،آمار رسمی،سیستم آماری ملی

*نویسنده مسئول: y_mehrali@sci.org.ir



کدگذاری هوشمند رشته تحصیلی با روش‌های یادگیری آماری

یاسر مهرعلی[۱*]
[۱]مرکز آمار ایران

چکیده: در این مقاله، یک چارچوب جامع برای کدگذاری هوشمند رشته‌های تحصیلی بر اساس طبقه‌بندی بین‌المللی ISCED با استفاده از روش‌های یادگیری آماری ارائه شده است. این مطالعه یک رویکرد سه مرحله‌ای شامل پاکسازی داده‌ها، تطبیق دقیق و کدگذاری هوشمند را پیشنهاد می‌کند. برای بخش کدگذاری، از مدل‌های پیشرفته‌ی یادگیری ماشین شامل جنگل تصادفی، ماشین بردار پشتیبان (SVM) و رگرسیون لوژستیک با استفاده از شبکه‌های عصبی استفاده شده و سپس یک روش ترکیبی مبتنی بر رأی اکثریت برای بهبود دقت نهایی طراحی شده است. نتایج تجربی نشان می‌دهد که مدل‌های ترکیبی قادرند دقت بسیار بالایی در کدگذاری رشته‌های تحصیلی به دست آورند. این چارچوب نه تنها قابلیت پیاده‌سازی در سامانه‌های بزرگ داده‌های آموزشی را دارد، بلکه می‌تواند به‌طور چشمگیری هزینه‌های عملیاتی و زمان مورد نیاز برای کدگذاری دستی را کاهش دهد و در عین حال، دقت و یکنواختی کدهای تولید شده را افزایش دهد.

واژه‌های کلیدی: کدگذاری هوشمند رشته‌های تحصیلی، طبقه‌بندی ISCED، ماشین بردار پشتیبان، جنگل تصادفی، شبکه‌های عصبی، پردازش زبان طبیعی.

*نویسنده مسئول: y_mehrali@sci.org.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

ارزیابی روش‌های برآورد پارامترهای توابع کوواریانس فضایی-زمانی با رویکرد مدل‌های نیمه‌طیفی

جواد ناصریان[۱*]، علی شهنواز[۱]
[۱]دانشگاه آزاد اسلامی واحد زنجان

چکیده: در این مقاله، به ارزیابی روش‌های مختلف برآورد پارامترهای توابع کوواریانس فضایی-زمانی در چارچوب مدل‌های نیمه‌طیفی پرداخته می‌شود. ابتدا مبانی نظری مدل‌سازی نیمه‌طیفی مرور و برخی از ویژگی‌های مهم مدل‌های نیمه‌طیفی اخیر معرفی و بررسی می‌گردد. سپس یک روش برآورد برای تابع کوواریانس فضایی-زمانی در حالت نیمه‌طیفی پیشنهاد می‌شود. به منظور سنجش کارایی این روش، دو مطالعه شبیه‌سازی طراحی شده است که در هر یک، روش برآورد پیشنهادی با چند روش متداول دیگر مقایسه می‌شود. نتایج شبیه‌سازی‌ها نشان می‌دهد روش پیشنهادی، به‌ویژه در اندازه نمونه‌های بزرگ، عملکرد برتری نسبت به رقبا دارد.

واژه‌های کلیدی: کوواریانس فضایی-زمانی،مدل فضایی-زمانی،مدل نیمه طیفی،درستنمایی وایتل

*نویسنده مسئول: nasseryan.javad@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

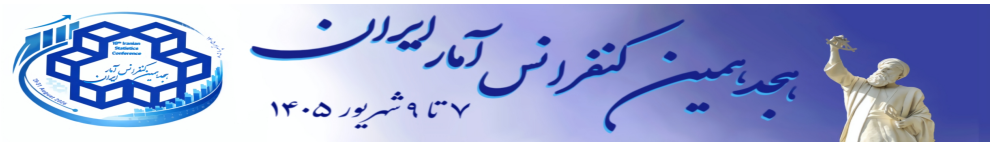
تحلیل بیز ناپارامتری مدل رگرسیون بتای آمیخته

اسماعیل نجفی[۱*]، مجید جعفری خالدی[۱]
[۱]دانشگاه تربیت مدرس

چکیده: برای مدل‌بندی ارتباط میان یک متغیر پاسخ با یک یا چند متغیر تبیینی از مدل‌های رگرسیونی استفاده می‌شود. در برخی از مسائل کاربردی با حالت‌هایی مواجه می‌شویم که متغیر پاسخ مقادیر خود را از بازه $(۰,۱)$ اختیار می‌کند. برای مدل‌بندی چنین متغیرهای پاسخی می‌توان از توزیع بتای بازپارامتریده در یک ساختار هماهنگ با مدل‌های خطی تعمیم یافته استفاده کرد. در برخی از مسائل علاوه بر متغیرهای تبیینی، عوامل پنهانی نیز بر متغیر پاسخ تأثیرگذار هستند. این عوامل را می‌توان در قالب اثرات تصادفی وارد مدل کرد. بدین ترتیب مدل رگرسیون بتای آمیخته شکل می‌گیرد. در این مقاله برای مدل‌بندی اثرات تصادفی از رهیافت بیزی ناپارامتری استفاده می‌شود. در این رهیافت، توزیع اثرات تصادفی نامعلوم و تصادفی فرض می‌شود. در این مقاله، از فرایند دیریکله آمیخته با هسته‌هایی مانند توزیع چوله‌نرمال و تعمیم آن یعنی توزیع چوله‌نرمال دومدی به‌عنوان پیشینی برای توزیع اثرات تصادفی مدل رگرسیون بتای آمیخته استفاده می‌شود. همچنین عملکرد مدل‌های معرفی شده در مطالعه شبیه‌سازی مورد ارزیابی قرار می‌گیرد.

واژه‌های کلیدی: رگرسیون بتای آمیخته، بیز ناپارامتری، فرایند دیریکله، فرایند دیریکله آمیخته، فرایند شکست چوب.

*نویسنده مسئول: esmaeil_najafi@modares.ac.ir



برخی روش‌های آموزش درس آمار در دانشگاه‌ها

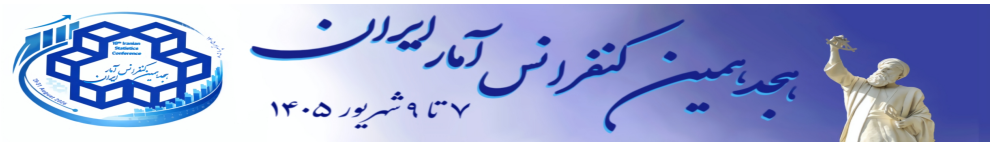
علیرضا نجفی زاده[۱*]

[۱]دانشگاه پیام نور

چکیده: روش‌ها و الگوهای متعدد آموزش درس آمار در دانشگاه‌ها و موسسه‌های مختلف آموزشی بر اساس نوع نیاز، کارآمدی، میزان اثر بخشی و غیره به کار گرفته می‌شوند. از جمله این روش‌ها می‌توان به الگوهای مفهوم‌محور، تمرین‌محور، فعال ترکیبی، روش‌های کلاسیک و غیره اشاره کرد که تحقیقات متعددی در این زمینه در سطوح و موضوع‌های مختلفی جهت اثبات کارآیی آن‌ها انجام شده است. هدف این سخنرانی، ارایه نتایج و مقایسه برخی از این تحقیقات در سال‌های اخیر توسط پژوهش‌گران مختلف است.

واژه‌های کلیدی: مفهوم‌محور،تمرین‌محور،فعال ترکیبی،کلاسیک،روش تدریس،عمل‌کرد

*نویسنده مسئول: najafizadeh@pnu.ac.ir



چارچوب یکپارچه یادگیری ماشین و بهینه‌سازی تصادفی برای قیمت‌گذاری ریسک‌آگاه در بیمه سلامت

امیرعلی نعمتی فرد[۱]*، مجتبی رنجبر[۱]، سامان وهابی[۱]
[۱]دانشگاه خوارزمی

چکیده: این پژوهش چارچوبی چهارلایه برای قیمت‌گذاری بیمه سلامت ارائه می‌دهد که شبکه عصبی عمیق با اتصالات میان‌بر، مدل ترکیبی با وزن‌دهی بهینه، شبیه‌سازی مونت‌کارلو لوگ‌نرمال و بهینه‌سازی تکامل دیفرانسیلی را یکپارچه می‌کند. آزمایش بر ۳۳۸،۱ رکورد واقعی نشان داد مدل ترکیبی ضریب تعیین ۹۴۵۳.۰ و خطای وزنی ۱۲.۸ درصد حاصل کرد. بهینه‌سازی تکاملی کمترین بار ریسک سازگار با قیود مالی را ۲۶.۵ درصد تعیین نمود و توانگری ۱۰۰ درصد با مازاد ۲۷.۲ میلیون دلار به‌دست آمد. تحلیل مقادیر شپلی همخوانی منطق مدل با دانش پزشکی را تأیید کرد.

واژه‌های کلیدی: قیمت‌گذاری بیمه سلامت، شبکه عصبی با اتصالات میان‌بر، فراگیری انباشته، شبیه‌سازی مونت‌کارلو، تکامل دیفرانسیلی، تفسیرپذیری شپلی.

---

*نویسنده مسئول: amiralinmtifrd@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

ارزیابی عدم قطعیت مدل انباشته با رویکرد ترکیبی در بیمه‌ی سلامت

مهرداد نیاپرست[۱]، طیبه کرمی[۱]*
[۱]دانشگاه رازی کرمانشاه

چکیده: عدم قطعیت یکی از مسائل مهم در صنعت بیمه‌ی سلامت است و نادیده گرفتن آن می‌تواند به قیمت‌گذاری اشتباه و کنارگذاشتن مشتریان کم‌ریسک منجر شود. در این مقاله، مدل انباشته برای رده‌بندی بیمه‌شدگان به دو رده‌ی پرریسک و کم‌ریسک بر روی مجموعه داده بیمه‌ی سلامت پیاده‌سازی شده است. سه نوع عدم قطعیت (بین‌مدلی، درون‌مدلی و خروجی مدل متا) به ترتیب با روش‌های واریانس پیش‌بینی مدل، جک‌نایف و آنتروپی شانون محاسبه گردید. نتایج نشان می‌دهد که مدل انباشته عملکرد مطلوبی دارد. عدم قطعیت کلی تحت تأثیر خروجی مدل متا قرار دارد. بنابراین رویکرد پیشنهادی می‌تواند در شناسایی مشاهدات مرزی و تصمیمات محافظه‌کارانه در بیمه سلامت مؤثر باشد.

واژه‌های کلیدی: عدم‌قطعیت،روش جک نایف،مدل انباشته،آنتروپی شانون

*نویسنده مسئول: tayebeh.karami20@gmail.com

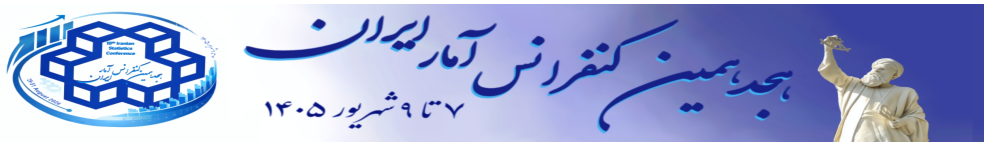
جنگل تصادفی فازی مبتنی بر رگرسیون فازی

فاطمه نیروئی وقفی[۱]*، محمدرضا ربیعی[۱]، سید محمود طاهری[۲]
[۱]دانشگاه صنعتی شاهرود
[۲]دانشگاه تهران

چکیده: روش‌های کلاسیک رگرسیون خطی فازی، با وجود توانایی در مدیریت عدم قطعیت، به دلیل ماهیت سراسری در شناسایی ساختارهای محلی و ناهمگن ناتوانند. در این پژوهش، مدل جنگل تصادفی فازی مبتنی بر رگرسیون فازی ارائه می‌شود. در هر درخت، فضای متغیرهای غیرفازی با استفاده از معیار میانگین مربعات خطای موزون به نواحی همگن تقسیم می‌شود و در هر برگ، یک مدل رگرسیون خطی فازی برازش می‌گردد. مدل پیشنهادی در یک مسئله کاربردی مرتبط با داده‌های قیمت مسکن به کار گرفته می‌شوند. نتایج نشان می‌دهد که این رویکرد ضمن حفظ تفسیرپذیری بالا، قادر به کشف روابط پنهان و ناهمگن بوده و پیش‌بینی‌های فازی همراه با عدم قطعیت ارائه می‌دهد.

واژه‌های کلیدی: ارزیابی اهمیت متغیرها،جنگل تصادفی فازی،فاصله دیاموند،اعداد فازی مثلثی

*نویسنده مسئول: Rabiei_stat@shahroodut.ac.ir



مقایسه الگوریتم‌های یادگیری ماشین برای پیش‌بینی مرگ‌ومیر بیماران سکته مغزی: انتخاب XGBoost به‌عنوان مدل نهایی و تفسیر نتایج با SHAP

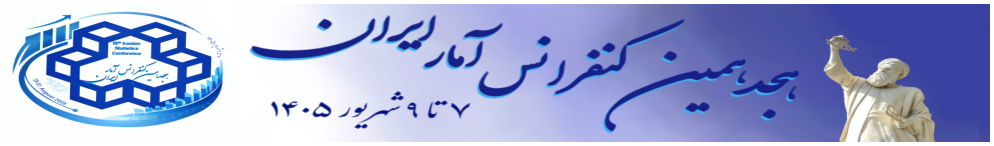
رحیم نیک بخت فرد[۱*]، الهه طالبی قانع[۱]، مجتبی خزایی[۱]، لیلی تاپاک[۱]
[۱]دانشگاه علوم پزشکی همدان

چکیده: این مطالعه با هدف ارزیابی و مقایسه عملکرد پنج الگوریتم یادگیری ماشین در پیش‌بینی مرگ‌ومیر بیماران مبتلا به سکته مغزی انجام شد. داده‌های ۲۶۵۳ بیمار مورد تحلیل قرار گرفت و پس از شناسایی متغیرهای مهم توسط الگوریتم XGBoost، از تحلیل SHAP برای تفسیر نتایج مدل استفاده شد. نتایج اعتبارسنجی متقاطع ۱۰-بخشی نشان داد که مدل XGBoost بهترین عملکرد را با مقادیر AUC = ۰.۸۹۵ ± ۰.۰۱۷، Accuracy = ۰.۸۷۹ ± ۰.۰۲۲، Specificity = ۰.۹۶۱ ± ۰.۰۱۲ و Sensitivity = ۰.۴۵۹ ± ۰.۰۹۷ ارائه می‌کند. مهم‌ترین عوامل مؤثر بر پیش‌بینی مرگ‌ومیر شامل شاخص‌های سطح هوشیاری، دسته‌های مقیاس کمای گلاسکو (GCS)، وضعیت شغلی و نحوه ورود بیمار به بیمارستان بودند. تحلیل SHAP نیز نقش برجسته متغیرهای نورولوژیک و شرایط اولیه پذیرش بیمار را در پیش‌بینی پیامد بیماری نشان داد. نتایج این مطالعه بیانگر ظرفیت مناسب مدل XGBoost برای پیش‌بینی مرگ‌ومیر پس از سکته مغزی است. با این حال، پیش از کاربرد بالینی، انجام مطالعات آینده‌نگر، اعتبارسنجی بیرونی و ارزیابی در مراکز درمانی متعدد ضروری است.

واژه‌های کلیدی: سکته مغزی،مرگ ومیر،یادگیری ماشین،XGBoost،داده‌های نامتعادل،تحلیل SHAP

---

*نویسنده مسئول: l.tapak@umsha.ac.ir



استراتژی ارتباطات و اتصالات در بهبود موفقیت در درس ریاضی

محمداسماعیل نیکفر[۱*]

[۱]دانش آموخته دکتری ریاضیات، دکترای تخصصی ریاضی محض، دبیر ریاضی، آموزش و پرورش منطقه ۴، تهران، ایران

چکیده: پژوهش حاضر با هدف بررسی تأثیر آموزش استراتژی ارتباطات و اتصالات در بهبود موفقیت در درس ریاضی دانش‌آموزان انجام گرفته است. این تحقیق از نظر روش ، شبه آزمایشی از نوع پیش آزمون و پسآزمون با گروه کنترل است . کلیه ی دانش‌آموزان پایه ششم ابتدایی شهر ورامین که در سال تحصیلی ۱۴۰۴ مشغول به تحصیل بودند به عنوان جامعهی آماری در نظر گرفته شده است. به روش نمونه گیری در دسترس ، ۶۰ دانش‌آموز به عنوان نمونه آماری انتخاب شدند. ابزار گردآوری اطلاعات، آزمون های ریاضی شامل ۵ سؤال متفاوت در دو مرحله پیش آزمون و پس آزمون بود. اطلاعات گردآوری شده براساس آمار توصیفی و استنباطی(آزمون t مستقل و وابسته) از طریق نرم افزارآماری spss ۲۲ مورد تجزیه وتحلیل قرار گرفت. یافته های پژوهش نشان داد استراتژی ارتباطات و اتصالات به تنهایی نیز بر بهبود موفقیت در درس ریاضی مؤثر بودند. با توجه به نتایج به دست آمده لزوم آموزش استراتژی ارتباطات و اتصالات در بهبود موفقیت در درس ریاضی دانش‌آموزان ضروری است.

واژه‌های کلیدی: آمار کاربردی،مطالعات میان‌رشته‌ای،تحلیل داده‌ها،روش‌های آماری،روش علمی،تصمیم‌گیری مبتنی بر داده

---

*نویسنده مسئول: m.s.nikfar@gmail.com

کاربرد بسط تیلور در تقریب نرخ انتروپی متقابل رنی

زهره نیکوروش[۱*]
[۱]دانشگاه صنعتی بیرجند

چکیده: انتروپی متقابل رنی بین دو توزیع احتمال متغیرهای تصادفی، تعمیمی از انتروپی متقابل شانون است که به ویژه در مبحث پردازش تصویر کاربردهای فراوان دارد. از جمله مواردی دیگری که در این مبحث اهمیت دارد، زنجیر های مارکف و مارکف پنهان هستند. در این مقاله سعی شده رابطه ای قابل محاسبه برای نرخ انتروپی متقابل رنی بین یک زنجیر مارکف و زنجیر مارکف پنهان متناظر آن، که در یک کانال با اغتشاش ایجاد می شود، به وسیله بسط تیلور (مرتبه اول) به دست آید.

واژه‌های کلیدی: انتروپی متقابل رنی،زنجیر مارکف پنهان،بسط تیلور

*نویسنده مسئول: nikooravesh@birjandut.ac.ir

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مدل‌سازی اکسیدها و کانی‌های معدنی با شبکه عصبی عمیق احتمالاتی

شاهین واعظی[۱*]، محسن محمدزاده[۲]
[۱]دانشگاه علم و فرهنگ تهران
[۲]دانشگاه تربیت مدرس

چکیده: پیش‌بینی خواص مواد بدون انجام آزمایش‌های پرهزینه، از چالش‌های مهم مهندسی مواد است. در این پژوهش، یک چارچوب یادگیری ماشین احتمالاتی برای پیش‌بینی الگوی پراش پرتو ایکس از روی داده‌های طیف‌سنجی پلاسمای جفت‌شده القایی ارائه شده است. هدف این تحقیق، مدل‌سازی رابطه غیرخطی میان ترکیب شیمیایی و ساختار مواد، همراه با برآورد عدم‌قطعیت پیش‌بینی‌ها است. مدل پیشنهادی ترکیبی از یک شبکه پرسپترون چندلایه برای استخراج ویژگی‌ها و یک شبکه چگالی آمیخته برای مدل‌سازی توزیع احتمالاتی خروجی است. این ساختار برخلاف مدل‌های قطعی مرسوم، امکان ارائه پیش‌بینی نقطه‌ای همراه با بازه اطمینان را فراهم می‌سازد. نتایج ارزیابی نشان می‌دهد که این چارچوب علاوه بر پیش‌بینی دقیق الگوهای پراش پرتو ایکس، عدم‌قطعیت ناشی از نویز داده‌ها و پیچیدگی ساختار را کمی‌سازی می‌کند. رویکرد ارائه‌شده ابزاری کارآمد برای کاهش هزینه‌های آزمایشگاهی، تسریع طراحی مواد و پشتیبانی از تصمیم‌گیری در تحقیقات این حوزه است.

واژه‌های کلیدی: یادگیری عمیق، رگرسیون، مدل‌سازی عدم قطعیت، پراش پرتو ایکس و طیف‌سنجی نشر پلاسمای القایی.

*نویسنده مسئول: shahinvaezi00@gmail.com

هجدهمین کنفرانس آمار ایران
۷ تا ۹ شهریور ۱۴۰۵

مقایسه الگوریتم های SARSA،VQ-QL و DQN با روش سنتی تمدید قیمت گذاری بیمه نامه ها با رویکرد بتا-حساسیت

سیدعلی ولایتی پاکرو[۱]*، رحمان فرنوش[۱]
[۱]دانشگاه علم و صنعت ایران

چکیده: شرکت‌های بیمه غیرزندگی هر سال با چالش تعیین حق‌بیمه تمدید مواجه هستند که باید درآمد را با حفظ مشتری متعادل کند. روش‌های سنتی مانند مدل‌های خطی تعمیم‌یافته (GLM) این مسئله را به‌صورت ایستا حل کرده و حساسیت قیمتی فردی و پویایی بازار را نادیده می‌گیرند. در این پژوهش، سه الگوریتم یادگیری تقویتی - SARSA، VQ-QL و DQN - برای قیمت‌گذاری پویای تمدید ،پیاده‌سازی و مقایسه شده‌اند. برخلاف پژوهش‌های پیشین که از مدل پذیرش ساده‌شده برنولی استفاده می‌کنند، یک مدل پذیرش بتا-حساسیت معرفی شده است که منحنی‌های پاسخ به قیمت واقع‌گرایانه و وابسته به مشتری تولید می‌کند. ارزیابی با استفاده از اعتبارسنجی متقابل ۵ لایه و آزمون‌های فریدمن و ویلکاکسون روی مجموعه‌داده واقعی بیمه سلامت خودرو (۱۰۹،۳۸۱ رکورد) انجام شده است. نتایج نشان می‌دهد VQ-QL با ضریب تغییرات ۰۱۴/۰ پایدارترین روش است. SARSA با نرخ نگهداری ۶۱/۱۶% محافظه‌کارانه‌ترین و DQN با پتانسیل بالا نیازمند داده و تنظیم بیشتر است. همه روش‌های RL انعطاف‌پذیری بیشتری نسبت به قیمت‌گذاری سنتی در مدیریت مبادله درآمد-نگهداری نشان می‌دهند.

واژه‌های کلیدی: فرایند تصمیم گیری مارکف ۱، قیمت‌گذاری تمدید بیمه۲، یادگیری تقویتی۳ و مدل بتا-حساسیت ۴.

*نویسنده مسئول: rfarnoosh@iust.ac.ir

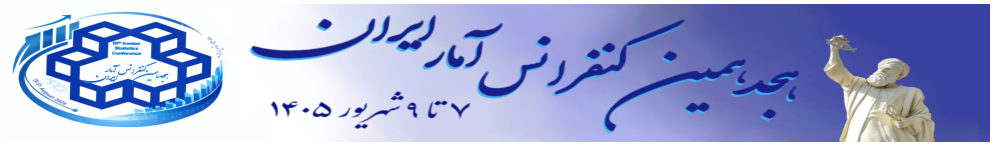
برآورد کوچک ناحیه‌ای با به‌کارگیری مدل پواسن آمیخته با ساختار صفر آماسیده

مطهره یوسفی[۱*]، موسی گلعلی‌زاده[۱]
[۱]دانشگاه تربیت مدرس

چکیده: در سال‌های اخیر مساله برآورد کوچک ناحیه‌ای، به‌دلیل نیاز به آمارهای قابل اعتماد، بسیار مورد توجه قرار گرفته است. مشکل عمده در برآورد کوچک ناحیه‌ای کافی نبودن حجم نمونه و یا حتی صفر بودن ویژگی مورد مطالعه است و همین امر باعث می‌شود که روش‌های سنتی آماری، برآوردهای ناپایدار و واریانس‌های بالایی را تولید نمایند. علاوه بر این، وجود صفرهای زیاد، ناهمگنی بین نواحی و عدم وجود اطلاعات کمکی مناسب، دقت و اعتبار مدل‌سازی را در این حوزه کاهش می‌دهد. اگرچه در مطالعات آماری که داده‌های مورد بررسی به‌صورت شمارشی هستند، به‌کارگیری یک مدل پواسن اجتناب ناپذیر است، اما زمانی که داده‌ها دارای صفرهای زیادی باشند، مدل پواسن استاندارد مناسب نیست. هدف اصلی مقاله حاضر ارائه یک روش آماری نوین برای برآوردگرهای کارا در ارتباط با نسبت و تعداد خانوارهای تک‌نفره در موقعیت‌های کوچک ناحیه‌ای است. بدین منظور مدل پواسن آمیخته با ساختار صفر آماسیده در سطح ناحیه به‌کار گرفته می‌شود. نتایج شبیه‌سازی نشان می‌دهد که رویکرد پیشنهادی عملکرد بهتری نسبت به مدل‌های رقیب دارد.

واژه‌های کلیدی: کوچک ناحیه‌ای، صفر آماسیده، مدل آمیخته‌ی پواسن.

*نویسنده مسئول: motahareh_yousefi@modares.ac.ir



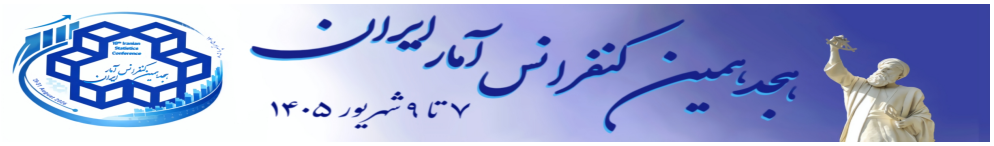
کنکاشی توصیفی-تحلیلی پیرامون جوانان ۱۵-۲۴ ساله‌ی غیرمحصل و غیرشاغل در ایران طی سال‌های 1399-1403

نریمان یوسفی[۱]، شهلا جعفری[۱*]
[۱]مرکز آمار ایران

چکیده: شاخص نیت که به جوانان خارج از چرخه‌ی آموزش، اشتغال و مهارت‌آموزی اشاره دارد، در دهه‌های اخیر به یکی از شاخص‌های کلیدی در تحلیل وضعیت بازارکار و رفاه اجتماعی تبدیل شده است. این شاخص نه تنها بیانگر میزان مشارکت اقتصادی نسل جوان است، بلکه نشان‌دهنده میزان کارآمدی نظام‌های آموزشی، اقتصادی و سیاست‌گذاری در هر جامعه نیز می‌باشد. تحلیل عوامل مؤثر بر آن نشان می‌دهد که ترکیبی از متغیرهای اقتصادی (نظیر بیکاری و رکود)، آموزشی (عدم انطباق مهارت‌ها با بازارکار) و اجتماعی (نابرابری‌ها و محدودیت‌های فرهنگی) در شکل‌گیری آن نقش دارند. همچنین، پیامدهای اقتصادی، اجتماعی، روانی و حتی سیاسی این مساله می‌تواند روند توسعه‌ی کشور را با چالش مواجه کند. هدف اصلی این پژوهش، بررسی وضعیت آماری جوانان نیت (جوانان غیر محصل و غیرشاغل) در ایران طی سال‌های 1399-1403 و تحلیل پیامدهای چند بعدی آن است. در این راستا، با استفاده از داده‌های آماری موجود، روند تغییرات این شاخص به تفکیک جنس و مناطق جغرافیایی (شهری و روستایی) مورد بررسی قرار گرفته است. نتایج بیانگر این واقعیت است، اگرچه در برخی سال‌ها روند کاهشی در نرخ نیت مشاهده می‌شود اما سطح این شاخص همچنان در مقایسه با کشورهای توسعه‌یافته و در حال توسعه بالا بوده و به‌ویژه در میان زنان شدت بیش‌تری دارد. در نهایت، با تاکید بر ضرورت مداخلات سیاستی، راهکارهایی جهت کاهش نرخ نیت و بهبود وضعیت مشارکت جوانان ارائه شده است.

واژه‌های کلیدی: جوانان نیت،آموزش،اشتغال،آسیب‌های اجتماعی،توسعه اقتصادی

*نویسنده مسئول: shahlajafari1366@gmail.com



پیش‌بینی بیماری‌زایی واریانت‌ها در سرطان پستان با استفاده از چارچوب یادگیری ماشین MAGPIE

ابوالفضل یونسی زیارانی[۱*]، زهرا رضائی قهرودی[۱]، کاوه کاوسی[۱]، لیلا میرصادقی[۱]
[۱]دانشگاه تهران

چکیده: شناسایی و تفسیر واریانت‌های ژنتیکی با اثر بیماری‌زایی در سرطان پستان، نقشی حیاتی در پزشکی شخصی‌سازی شده ایفا می‌کند. با این حال، حجم بالای داده‌های توالی‌یابی و عدم‌تعادل شدید میان واریانت‌های بیماری‌زا و خنثی، چالش‌های جدی در دقت پیش‌بینی ایجاد کرده است. در این پژوهش، چارچوب یادگیری ماشین MAGPIE به‌منظور پیش‌بینی بیماری‌زایی واریانت‌های تک‌نوکلئوتیدی بر روی مجموعه‌داده‌ای شامل ۱۴۵۸ جهش شناسایی‌شده در سطح ژنوم به‌کار گرفته شده است و با اعتبارسنجی متقاطع ارزیابی شد. این داده‌ها از پایگاه داده معتبر پروژه اطلس ژنوم سرطان (TCGA) و از نمونه‌های مرتبط با سرطان پستان استخراج و تحلیل شده‌اند. به منظور مقابله با چالش عدم‌تعادل داده‌ها و برای حفظ اصالت بیولوژیکی نمونه‌ها، از استراتژی وزن‌دهی به رده‌ها و یادگیری حساس به تابع هزینه استفاده شده است. نتایج ارزیابی نشان داد که این رویکرد در مقایسه با ابزارهای پیش‌بینی موجود براساس معیارهای score-F_1، فراخوانی، دقت و سطح زیر منحنی ویژگی عملگر گیرنده، عملکرد مناسبی در اولویت‌بندی واریانت‌های کلینیکی نشان می‌دهد.

واژه‌های کلیدی: سرطان پستان، یادگیری ماشین وزن‌دار، واریانت‌های ژنتیکی، چارچوب MAGPIE، عدم‌تعادل داده‌ها

---

*نویسنده مسئول: abolfazl.younesi@ut.ac.ir