

کد درس: ۱۶

نام درس: داده‌کاوی و استخراج دانش در پزشکی

پیش‌نیاز یا همزمان: هوش مصنوعی در علوم پزشکی (کد ۱۵)

تعداد واحد: ۲ واحد (۱/۵ واحد نظری - ۰/۵ واحد عملی)

نوع واحد: نظری - عملی

هدف کلی درس

یادگیری مبانی داده‌کاوی و فرایند استخراج دانش

شرح درس و رئوس مطالب (۲۶ ساعت نظری - ۱۷ ساعت عملی)

انتظار می‌رود دانشجو با دانش قبلی که در مورد حوزه سلامت و انواع داده‌های موجود در آن دارد، بتواند مسائلی را تعریف نماید و با رویکرد داده‌کاوی آن را حل کند. به همین دلیل، دانشجو باید دانش کافی در زمینه اکتشاف داده و پایگاه‌داده‌های پزشکی، پیش‌پردازش داده‌ها، شناسایی داده‌های پرت، روش‌های داده‌کاوی از جمله روش‌های دسته‌بندی، خوشه‌بندی، کاوش الگوهای مکرر و همبستگی‌ها و همچنین روش‌های ارزیابی دانش کشف شده و مصورسازی آن کسب نماید.

رئوس مطالب نظری

- مفاهیم اکتشاف دانش از طریق داده و داده‌کاوی
- شناخت داده‌ها (انواع داده‌های عددی، اسمی، دودویی، ترتیبی، گسسته، پیوسته، شناخت روش‌های مصورسازی داده‌ها (روش‌های مصورسازی پیکسل‌گرا، هندسی، مبتنی بر شمایل و مصورسازی داده‌ها و روابط پیچیده)، محاسبه تشابه و عدم تشابه داده‌ها)
- پیش‌پردازش داده‌ها (روش‌های بررسی کیفیت و پاکسازی داده‌ها، یکپارچه‌سازی داده‌ها، کاهش داده‌ها، تبدیل داده‌ها و روش‌های مقابله با نویز، داده‌های ناموجود و داده‌های پرت)
- مفاهیم و روش‌های دسته‌بندی داده‌ها (درخت تصمیم، شبکه‌های عصبی مصنوعی، دسته‌بند بیز، ماشین بردار پایه، یادگیرنده‌های کند، استدلال مبتنی بر مورد، دسته‌بند k نزدیکترین همسایه)
- مفاهیم و روش‌های خوشه‌بندی و تحلیل خوشه (روش‌های افزاز، روش‌های سلسله مراتبی، روش‌های مبتنی بر چگالی، روش‌های مبتنی بر گرید، روش‌های ارزیابی خوشه‌ها)
- ارزیابی و انتخاب مدل (سنجه‌های ارزیابی کارآیی دسته‌بند، اعتبارسنجی متقابل، آزمون‌های آماری، منحنی ROC)
- روش‌های مصورسازی دانش کشف شده (مصورسازی شعاعی، نگاشت‌های خودسازمانده)

رئوس مطالب عملی

- دانشجو باید بر اساس داده‌های آزمایشی (که برای امور تحقیقاتی منتشر می‌شوند) یا داده‌های واقعی کلیه فرایندهای اکتشاف دانش و داده‌کاوی مطرح شده را به شکل عملی تمرین کند.
- دانشجو باید بتواند از یکی از ابزارهای داده‌کاوی (Matlab, R, Weka, Rapid miner, IBM modeler,...) برای مدلسازی الگوریتم‌ها استفاده نماید.

منابع اصلی درس

1. Han, J. Pei, J. and Kamber, M., Latest edition. Data mining: concepts and techniques. Latest edition. Elsevier.



2. Kantardzic, M. Data mining: concepts, models, methods, and algorithms. Latest edition. John Wiley & Sons.
3. Fernández-Llatas, C. and García-Gómez, J.M. eds. Data mining in clinical medicine. Latest edition. Humana Press.

شیوه ارزیابی دانشجو

ارزشیابی نظری: این درس بشکل تراکمی (امتحان پایان‌ترم) و تکوینی (امتحان میان‌ترم، ارزیابی‌های طول ترم) می‌باشد. بررسی‌های طول ترم ۲۰ درصد، امتحان میان‌ترم ۲۰ درصد و امتحان پایان‌ترم ۶۰ درصد خواهد بود.

ارزشیابی عملی: ارزشیابی عملی به صورت ارزشیابی فعالیت‌های عملی دانشجو در طول ترم (۵۰ درصد) و پروژه پایان ترم (۵۰ درصد) خواهد بود.

