

به نام خدا

پاسخنامه تکلیف اول درس مبانی داده کاوی – بخش اول

ترم دوم ۱۴۰۰-۱۳۹۹

• سوال یک

- فاز درک مسئله: ابتدا باید بررسی شود روند مصرف دارو در بیمارستان‌ها طی یک سال گذشته به چه عواملی بستگی دارد و سپس داروهای مصرف شده را به منظور یافتن نتایج دقیقتر در دسته‌های بزرگ تر دسته‌بندی کنیم به طور مثال درحال حاضر همه‌گیری کرونا موجب مصرف بیش از اندازه داروهای ضدعفونی در بیمارستان‌ها شده است. در ادامه باید بررسی کنیم در فصل زمستان چه بیماری‌هایی بیشتر بروز پیدا میکنند تا یکی از عوامل موثر بر مصرف دارو را در این فصل پیدا کنیم.
- فاز درک داده: در این فاز باید بررسی کنیم چه نوع داده‌هایی در دسترس داریم. به طور مثال بررسی کنیم در سیستم‌های اطلاعاتی بیمارستان چه اطلاعاتی از بیماران و روند بهبود و خدمات انجام‌شده برای آن موجود است. کیفیت و میزان این داده‌ها در این فاز مشخص میشود.
- فاز آماده‌سازی داده: در این فاز باید داده‌های جمع‌آوری شده از مرحله‌ی قبل را بررسی نموده و فرایندهای پاکسازی داده و تبدیلات مورد نظر را بر روی آن‌ها انجام میدهیم. به طور مثال نرمالایز کردن داده‌های عددی مانند میزان مصرف دارو
- فاز ایجاد مدل: در این فاز براساس داده‌های بدست آمده و همچنین نیاز مسئله مدل‌های مختلفی با پارامترها و متغیرهای موجود ایجاد میشوند. به طور مثال چندین مدل پیش‌بینی با استفاده از الگوریتم‌های مختلف مانند درخت تصمیم و غیره ایجاد میشوند.
- فاز ارزیابی مدل: در این فاز مدل‌های ایجاد شده از مرحله قبل را براساس معیارهای توصیف شده در مسئله بررسی میکنیم تا میزان دقت و کارایی هر مدل مشخص شده و بهترین مدل به عنوان مدل نهایی انتخاب شود.
- فاز استقرار مدل: در این فاز از مدل نهایی شده گزارشاتی استخراج شده و یا مدل بر روی داده‌های چندین بیمارستان دیگر اجرا میشود تا درنهایت مدل نهایی به مرور تکمیل و توسعه یابد.

• سوال دو

تخمین (Estimation): تخمین نمره درس دیفرانسیل براساس نمرات کسب شده در دروس ریاضی ۱ و ریاضی ۲ و فیلد دانشکده و رشته‌ی دانشجو

دسته‌بندی (Classification): دسته‌بندی دانشجویان به دسته‌های "نیاز به مراقبت" و "عدم نیاز به مراقبت" براساس نمرات کسب شده و تعداد مشروطی‌های دانشجو

خوشه‌بندی (Clustering): خوشه‌بندی دانشجویان براساس میزان نمرات کسب شده در دروس مختلف به منظور پیشنهاد رشته تخصصی به آن‌ها