

طرح درس جهت ارائه در نیمسال تحصیلی دوم ۱۴۰۳-۱۴۰۴

| دانشکده       | مهندسی صنایع و سیستمها                   | گروه          | مهندسی فناوری اطلاعات                                                                                                                                                                                                                                                                               |
|---------------|------------------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| گرایش         | سامانه های شبکه ای                       | مقطع          | کارشناسی ارشد                                                                                                                                                                                                                                                                                       |
| نام درس       | داده کاوی، مدلها، الگوریتم ها و کاربردها | نوع درس       | <div> <div>پایه <input type="checkbox"/></div> <div>تخصصی <input type="checkbox"/></div> <div>اختیاری <input type="checkbox"/></div> </div> <div> <div>نظری <input checked="" type="checkbox"/></div> <div>عملی <input type="checkbox"/></div> <div>نظری-عملی <input type="checkbox"/></div> </div> |
| تعداد واحد    | ۳                                        | نام استاد     | بابک تیمورپور                                                                                                                                                                                                                                                                                       |
| دروس پیش نیاز | -                                        | تلفن دفتر کار | ۰۲۱۸۲۸۸۳۹۸۷                                                                                                                                                                                                                                                                                         |
| دروس هم نیاز  | -                                        | پست الکترونیک | <a href="mailto:babaktei@gmail.com">babaktei@gmail.com</a>                                                                                                                                                                                                                                          |

هدف کلی: هدف اصلی این درس آشنایی دانشجویان با داده کاوی، کاربردها، روشها و چهارچوبهای مرجع آن است. در این درس، دانشجویان با روشهای اصلی داده کاوی شامل مدلهای یادگیری بدون ناظر (مانند خوشه بندی) و با ناظر (مانند دسته بندی و رگرسیون) آشنا شده و پیاده سازی الگوریتمها و روشهای مختلف آنها را در قالب تمرینهای مختلف و پروژه پایانی درس تمرین می کنند. در طول درس بنا بر مبحث، مطالعات موردی از دنیای واقعی ارائه می شود. علاوه براین دانشجویان باید قادر شوند تا روشهای داده کاوی را در سیستمهای کمک تصمیم به کار گیرند.

اهداف ویژه:

۱. -کسب دانش و توانایی لازم در پیاده سازی روشهای مختلف داده کاوی در داده های واقعی و مجموعه داده های استاندارد به صورت عملی
۲. - آشنایی با کاربردهای مختلف داده کاوی به صورت مطالعه موردی در حین هر مبحث و به ویژه در حوزه سیستمهای کمک تصمیم اهداف درس:

✓ رئیس مطالب و برنامه ارائه در کلاس: (در صورتی که واحد عملی یا نظری-عملی بود، نوع آموزش در توضیحات بیان شود)

| شماره جلسه   | موضوع جلسه درس                                                                                                                           | توضیحات |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------|---------|
| جلسه ۱ و ۲   | مقدمه و معرفی داده کاوی، ضرورت داده کاوی، چهارچوبها، کاربردها                                                                            |         |
| جلسه ۳ و ۴   | انواع داده، ماتریس فاصله، نفرین بعد، پیش پردازش: مقادیر مفقوده، نقاط پرت، گسسته سازی نرمال سازی                                          |         |
| جلسه ۵ و ۶   | ۱.۳. مصورسازی داده: نمودار فراوانی، نمودار پراکندگی، محورهای موازی، MDS<br>۱.۱. کاهش بعد و ساخت ویژگی ترکیبی، PCA                        |         |
| جلسه ۷ و ۸   | نمونه گیری و متوازن سازی، انتخاب ویژگی: فیلتر، embedded.wrapper<br>معیارهای ارزیابی: دسته بندی، رگرسیون (متغیر وابسته پیوسته)، بیش برازش |         |
| جلسه ۹ و ۱۰  | درخت تصمیم                                                                                                                               |         |
| جلسه ۱۱ و ۱۲ | ماشین بردار پشتیبان                                                                                                                      |         |
| جلسه ۱۳ و ۱۴ | بیز ساده و شبکه بیزی، یادگیرنده تنبل و knn                                                                                               |         |
| جلسه ۱۵ و ۱۶ | یادگیری ترکیبی Ensemble<br>یادگیری فعال                                                                                                  |         |
| جلسه ۱۷ و ۱۸ | خوشه بندی kmeans<br>خوشه بندی سلسه مراتبی، خوشه بندی ترکیبی (Ensemble)                                                                   |         |

|              |                                                                                                                   |
|--------------|-------------------------------------------------------------------------------------------------------------------|
| جلسه ۱۹ و ۲۰ | خوشه بندی بر پایه چگالی Dbscan<br>شاخصهای ارزیابی خوشه بندی: نمودار آرنج، سیلوئت، ارزیابی بیرونی با Rand          |
| جلسه ۲۱ و ۲۲ | رگرسیون لجستیک، تابع فعال سازی سیگموئید، تابع زیان Cross Entropy<br>شبکه عصبی و پس انتشار خطا                     |
| جلسه ۲۳ و ۲۴ | مشکلات زیاد شدن تعداد لایه ها و راه حل، تابع ReLU<br>شبکه های عصبی پیچشی CNN                                      |
| جلسه ۲۵ و ۲۶ | شبکه های عصبی RNN                                                                                                 |
| جلسه ۲۷ و ۲۸ | یادگیری نیمه نظارتی و خودنظارتی<br>ساخت سیستمهای کمک تصمیم                                                        |
| جلسه ۲۹ و ۳۰ | واکافت داده های پیچیده: متن، تصویر، سری زمانی، گراف، توالی، ژنتیک<br>اریبی الگوریتم، تفسیرپذیری، محرمانگی و انصاف |
| جلسه ۳۱ و ۳۲ | ارائه دانشجویان                                                                                                   |

✓ روش ارزشیابی:

۱. یک پروژه درسی تا حداکثر ۲ ماه پس از شروع درس تحت نظارت استاد راهنما تعریف می گردد. برخی از تمرینهای درسی بر روی این پروژه اجرا می شوند. در هفته آخر درس جمع بندی پروژه توسط دانشجویان ارایه می گردد.

۲. - آزمون پایان نیمسال: ۸ نمره

۳. - تمرینهای کلاسی: ۷ نمره

۴. - پروژه درس: ۳ نمره

✓ منابع:

1. Tan PN, Steinbach M, Karpatne A., Kumar V. 2018, Introduction to data mining, 2nd ed., Pearson.
2. James G., Witten D., Hastie T., Tibshirani R., Taylor J. 2023, An Introduction to Statistical Learning with Applications in Python, 2nd ed, Springer.
3. Han, J, Kamber, M & Pei, J 2012, Data Mining: Concepts and Techniques, 3rd ed., Elsevier.
4. Zaki M.J. and JR W.M., 2020, Data mining and machine learning fundamental concepts and algorithms, 2nd Ed., Cambridge University Press.