



دانشگاه علوم پزشکی
و خدمات بهداشتی درمانی تهران

معاونت آموزشی

مرکز مطالعات و توسعه آموزش علوم پزشکی

واحد برنامه‌ریزی آموزشی

چارچوب طراحی «طرح دوره»

اطلاعات درس:

گروه آموزشی ارابه دهنده درس: مدیریت اطلاعات و انفورماتیک سلامت

عنوان درس: تحلیل داده های سلامت

کد درس: ۱۴

نوع و تعداد واحد^۱: نظری - عملی

نام مسؤؤل درس: دکترزهراسادات آزاد

مدرس/ مدرسان: دکترزهراسادات آزاد

پیش‌نیاز/ هم‌زمان: ندارد

رشته و مقطع تحصیلی: ارشد فناوری اطلاعات سلامت

اطلاعات مسؤؤل درس:

رتبه علمی: استادیار

رشته تخصصی: مدیریت اطلاعات سلامت

محل کار: دانشکده پیراپزشکی

تلفن تماس: ۰۹۱۲۸۱۱۸۸۲۷

نشانی پست الکترونیک: azadm@tums.ac.ir

^۱ مشتمل بر: نظری، عملی و یا نظری- عملی به تفکیک تعداد واحدهای مصوب. (مثال: ۲ واحد نظری، ۱ واحد عملی)

توصیف کلی درس (انتظار می‌رود مسؤول درس ضمن ارائه توضیحاتی کلی، بخش‌های مختلف محتوایی درس را در قالب یک یا دو بند، توصیف کند):

در این درس، مباحث مربوط به علم داده، انواع تحلیل‌ها، ویژگی‌ها و تفاوت‌های کلان داده‌ها با داده‌های سنتی، فرایند داده‌کاوی، تفاوت داده‌کاوی و آمار، انواع ساختارهای داده‌خطی و غیرخطی، مفهوم تابع و الگوریتم‌های خطی و غیرخطی، وجه تمایز انواع یادگیری‌های ماشینی، عمیق، تقویتی در حوزه هوش مصنوعی، مفاهیم و اصطلاحات استاندارد درباره مجموعه داده در یادگیری ماشینی، تفاوت الگوریتم و مدل، مفاهیم ریاضی ماتریس و تنسور در هوش مصنوعی، انواع مسائل و Task‌های یادگیری ماشینی، انواع مسائل طبقه‌بندی (Binary vs multiclass vs multilabel)، روش‌های نظارتی و غیرنظارتی، علل محبوبیت زبان پایتون در حوزه علم داده برای تحلیل داده، کتابخانه‌های پرکاربرد پایتون برای تحلیل داده، معرفی IDE‌های پایتون بصورت تئوری مطرح می‌شود.

سپس نحوه نصب، حذف و ارتقا Anaconda، محیط کار Jupyter notebook، نحوه فراخوانی کتابخانه‌ها و لودینگ انواع فرمت‌های مجموعه داده در پایتون، ساختار دیتافریم پانداس، اجرای متدهای رایج دیتافریم پانداس برای تحلیل‌های توصیفی، روش‌ها و کاربردهای Slicing و Grouping داده‌ها در تحلیل داده، روش‌ها و مراحل تحلیل اکتشافی EDA، مصورسازی داده‌ها با کتابخانه‌های Matplotlib و Seaborn، تکنیک‌های ترنسفورم متغیرهای طبقه‌ای، تکنیک‌های شناسایی و مدیریت Missing و Outliers و Duplications بصورت عملی، تکنیک‌ها و نکات data balancing و scaling، روش‌های Data Split / Cross-validation، روش‌ها و شاخص‌های ارزیابی عملکرد مدل‌ها، آموزش الگوریتم‌های پایه و ensemble یادگیری ماشینی، روش‌های Optimization مدل‌سازی، بصورت عملی به دانشجویان آموزش داده می‌شود.

اهداف کلی / محورهای توان‌مندی:

توانایی اجرای پروژه‌های تحلیل داده یادگیری ماشینی

اهداف اختصاصی / زیرمحورهای هر توان‌مندی:

- درک و توضیح مفاهیم مربوط به تحلیل داده‌های سلامت، انواع ساختارهای داده، مفاهیم ریاضی مرتبط
- کسب مهارت استفاده از زبان برنامه‌نویسی پایتون برای انجام تحلیل داده
- توانایی تشخیص روشها و تکنیک‌های صحیح باتوجه به ویژگی‌های داده‌ها، برای خلاصه‌سازی، تحلیل توصیفی و مصورسازی، Scaling, encoding و متعادل‌سازی آنها
- توانایی تشخیص نوع مساله، نوع یادگیری مورد نیاز برای تحلیل و مدل‌سازی‌ها
- توانایی انتخاب صحیح الگوریتم‌ها و اجرای درست فرایند آموزش برای تحلیل و مدل‌سازی
- توانایی ارزیابی دقیق مدل‌ها و تفسیر نتایج

رویکرد آموزشی^۱:

ترکیبی^۲

حضوری

مجازی^۲ (باتوجه به شرایط)

روش های یاددهی - یادگیری با عنایت به رویکرد آموزشی انتخاب شده:

رویکرد حضوری

- سخنرانی تعاملی (پرسش و پاسخ، کوئیز، بحث گروهی و ...)
- بحث در گروههای کوچک
- ایفای نقش
- یادگیری اکتشافی هدایت شده
- یادگیری مبتنی بر تیم (TBL)
- یادگیری مبتنی بر حل مسئله (PBL)
- یادگیری مبتنی بر سناریو
- استفاده از دانشجویان در تدریس (تدریس توسط همتایان)
- یادگیری مبتنی بر بازی

تقویم درس:

جلسه	عنوان مبحث	روش تدریس	فعالیت های یادگیری / تکالیف دانشجوی	نام مدرس / مدرسان
۱	مباحث مربوط به علم داده، انواع تحلیل ها در حوزه سلامت، ویژگی ها و تفاوت های کلان داده ها با داده های سنتی، چرخه و فرایند داده کاوی، تفاوت داده کاوی و آمار، انواع ساختارهای داده خطی و غیرخطی	سخنرانی تعاملی و ارائه اسلاید	مطالعه منابعی آموزشی تهیه شده و رفرنس ها	دکترزهراسادات آزاد
۲	مفهوم تابع و الگوریتم های خطی و غیر خطی، وجه تمایز انواع یادگیری های ماشین، عمیق، تقویتی در حوزه هوش مصنوعی، مفاهیم و اصطلاحات استاندارد درباره مجموعه داده در یادگیری ماشین، تفاوت الگوریتم و مدل، علل محبوبیت زبان پایتون در حوزه علم داده برای تحلیل داده، کتابخانه های پرکاربرد پایتون برای تحلیل داده.	سخنرانی تعاملی و ارائه اسلاید	مطالعه منابعی آموزشی تهیه شده و رفرنس ها	دکترزهراسادات آزاد
۳	توضیح مفاهیم ریاضی ماتریس و تنسور و کاربرد آنها در تحلیل های هوش مصنوعی، انواع مسائل و Task های یادگیری ماشین، انواع مسائل طبقه بندی (Binary vs	سخنرانی تعاملی و ارائه اسلاید	مطالعه منابعی آموزشی تهیه شده و رفرنس ها	دکترزهراسادات آزاد

1. Educational Approach

2. Virtual Approach

3. Blended Approach: Blended learning is an approach to education that combines online educational materials and opportunities for interaction online with traditional place-based classroom methods.

جلسه	عنوان مبحث	روش تدریس	فعالیت‌های یادگیری / تکالیف دانشجو	نام مدرس / مدرسان
	غیرنظارتی، (multiclass vs multilabel) روش های نظارتی و			
۴	معرفی IDE های پایتون، مزیت و علت محبوبیت Jupyter notebook، معرفی، نحوه نصب، حذف و ارتقا Anaconda، آشنایی با بخش هی محیط کار Jupyter notebook. کرنر، مدهای اجرایی، روش ها و کلید های میانبرها برای اجرای عملیات های مختلف روی سلول ها، معرفی انواع داده های رایج پایتون شامل float, integer, Dictionary, list, string و ساختار آرایه در کتابخانه NumPy	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح	مطالعه منابعی آموزشی تهیه شده، مشاهده فیلم های آموزشی ضبط شده توسط مدرس - اجرای دستورات و تمرین	دکترزهراسادات آزاد
۵	نحوه فراخوانی کتابخانه ها و لودینگ انواع فرمت های مجموعه داده در پایتون،تجمیع، ادغام از منابع مختلف، کار با ساختار دیتافریم پانداس، اضافه کردن عناوین متغیرها از طریق لیست در شرایط خاص، دستور تغییر عناوین متغیرها با دیکشنری، اجرای متدهای رایج دیتافریم پانداس برای تحلیل های توصیفی (محاسبه شاخص های مرکزی و پراکندگی برای متغیرها، محاسبه چولگی، بررسی correlation)، روش ها و کاربردهای Slicing و Grouping داده ها در تحلیل اکتشافی داده	آموزش عملی با نرم افزار با یک مجموعه داده واقعی- نوشتن و اجرای کد، نمایش به دانشجو و توضیح	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۶	روش ها و مراحل تحلیل اکتشافی EDA، مصورسازی داده ها با کتابخانه های Matplotlib و Seaborn، شیوه های مختلف نوشتن کد ترسیم نمودارهای هیستوگرام، بارچارت، تراکم، باکس پلات، نمودار حرارتی،تفسیر نمودارها، تعریف پارامترهای مربوطه برای تنظیم استایل و ظاهر نمودارها	آموزش عملی با نرم افزار با یک مجموعه داده واقعی نوشتن و اجرای کد، نمایش به دانشجو و توضیح	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۷	اهمیت و تکنیک های تبدیل متغیرهای کتگوریکال به عددی، مفاهیم missing، داده پرت، نویز، آشنایی با تکنیک های شناسایی و مدیریت Missing و Outliers، محاسبه Z-scores و IQR در برای Outliers معرفی الگورتیم ها حساس و غیرحساس به آنها، بررسی اثر استفاده از هر کدام از روشها بر عملکرد مدل و شناسایی و حذف Duplications بصورت عملی	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح + استفاده از اسلاید برای مباحث ارائه تئوری	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۸	مفهوم Feature scaling و اهمیت آن، تفاوت روش های نرمال سازی و استانداردسازی برای Feature scaling و شرایط استفاده از هر کدام، اجرای مقیاس بندی با ساب پکیج sklearn.preprocessing، کاربرد هر کدام از	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح + استفاده از اسلاید	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد

جلسه	عنوان مبحث	روش تدریس	فعالیت‌های یادگیری / تکالیف دانشجو	نام مدرس / مدرسان
	تکنیک Min-Max Scaling، MaxAbsScaler، RobustScaler، Power Transform برای نرمال سازی، اجرای تکنیک StandardScaler، پیشگیری از data leakage، مفهوم مجموعه داده نامتعادل، نحوه چک کردن، تکنیک های نمونه گیری مجدد under sampling و oversampling برای متعادل سازی داده ها (Tomek links, SMOTE, NearMiss)	برای مباحث ارائه تئوری		
۹	روش های Cross-validation / Data Split، شاخص های ارزیابی عملکرد مدل های رگرسیونی و طبقه بندی، مفاهیم کانفیوژن ماتریس، تحلیل نمودارهای لازم برای ارزیابی شامل ROC و AUC و Learning curve و غیره	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح + استفاده از اسلاید برای مباحث ارائه تئوری	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۰	مفهوم نفرین ابعاد، تمایز مفهوم Feature selection و Feature engineering، اهمیت مرحله انتخاب ویژگی، روشهای غیرنظارتی انتخاب شامل بررسی واریانس، مفهوم High multicollinearity. روشهای فیلتر، روشهای Intrinsic Wrapper، کاربرد هر کدام	آموزش عملی با نرم افزار- استفاده از اسلاید برای مباحث ارائه تئوری	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۱	معرفی و نحوه آموزش الگوریتم های پایه رگرسیون خطی، درخت تصمیم KNN، و NB	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۲	معرفی شبکه عصبی، معماری، انواع الگوریتم های شبکه عصبی و آموزش آن	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۳	معرفی و نحوه آموزش مدل های ensemble یادگیری ماشین، RF، XGBoost و Adaboost	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۴	تمایز مفهوم Hyperparameter و پارامتر، مفهوم hyperparameter tuning، روش های Optimization (Grid search, Random search vs Bayesian optimization) و روش تفسیر مدل	آموزش عملی با نرم افزار- نوشتن و اجرای کد، نمایش به دانشجو و توضیح + استفاده از اسلاید برای مباحث ارائه تئوری	اجرای عملی دستورات آموخته شده روی نمونه داده توسط دانشجو و تمرین	دکترزهراسادات آزاد
۱۵				
۱۶				
۱۷				

وظایف و انتظارات از دانشجو:

انتظار می رود دانشجو در کلاس حضور منظم و فعالانه داشته و در انجام تکالیف محوله هر جلسه، مطالعه محتوای آموزشی ارائه شده مدرس و منابع اصلی در موعد مقرر اهتمام داشته باشد و پروژه پایانی را در مهلت مقرر برای مدرس ارسال نماید.

روش ارزیابی دانشجو:

- موارد موثر در ارزیابی :
- حضور سر کلاس : به علت درجه سختی واحد درسی بیش از یک جلسه غیبت نمره منفی دارد - حضور کامل ۱ نمره اضافه تشویقی
- مشارکت در پرسش و پاسخ ها ۱ نمره اضافه تشویقی
- انجام تمرین ها و کارهای کلاسی هر جلسه ¼ نمره پایانی
- آزمون میان ترم خودارزیابی / فقط تاثیر مثبت
- پروژه پایان ترم ¼ نمره آزمون پایانی
- آزمون کتبی پایانی ½ نمره پایانی

منابع:

Book:

1. Machine Learning with Python. 2019. Tutorials Point
2. Python for data analysis. Michael Brothers. Portilla's Udemy course Learning Python for Data Analysis and Visualization. available on <http://github.com/mikebrothers/data-science/>

۳. یادگیری ماشین و علم داده ، مبانی و مفاهیم الگوریتم ها و ابزارها. میلاد وزان. نشر میعاد اندیشه. سال ۱۴۰۰ تهران

Cheat sheet:

- Numpy
- Pandas
- Jupyter Notebook

- matplotlib

Website:

<https://pandas.pydata.org>

<https://numpy.org>

<http://scikit-learn.org>

<https://matplotlib.org>

چک لیست ارزیابی طرح دوره

چگونگی پردازش طرح با توجه به معیارها			معیارهای ارزیابی	آیتم	نام درس	رشته مقطع	گروه
توضیحات در خصوص موارد نیازمند اصلاح	نیازمند اصلاح	قابل قبول					
			به اطلاعات کلی درس اعم از گروه آموزشی ارایه دهنده درس، عنوان درس، کد درس، نوع و تعداد واحد، نام	اطلاعات درس			

			مسئول درس و سایر مدرسان، دروس پیش نیاز و هم‌زمان و رشته و مقطع تحصیلی اشاره شده است.			
			اطلاعات مسئول درس اعم از رتبه علمی، رشته تخصصی، اطلاعات تماس و ... درج شده است.	اطلاعات مسئول درس		
			بخش‌های مختلف محتوایی درس در حد یک یا دو بند معرفی شده است.	توصیف کلی درس		
			اهداف کلی/ محورهای توانمندی با قالب نوشتاری صحیح درج شده‌اند.	اهداف کلی/ محورهای توانمندی		
			اهداف اختصاصی/ زیرمحورهای هر توانمندی با قالب نوشتاری صحیح درج شده‌اند.	اهداف اختصاصی/ زیرمحورهای هر توانمندی		
			رویکرد آموزشی مورد نظر در آرایه دوره اعم از حضوری، مجازی و ترکیبی مشخص شده است.	رویکرد آموزشی		
			روش‌های یاددهی و یادگیری درج شده‌اند.	روش‌های یاددهی- یادگیری		
			جدول مربوط به تقویم درس، به طور کامل تکمیل شده است.	تقویم درس		
			وظایف و انتظارات از دانشجویان نظیر حضور منظم در کلاس درس، انجام تکالیف در موعد مقرر، مطالعه منابع معرفی شده و مشارکت فعال در برنامه‌های کلاس و ... تعریف شده و درج گردیده است.	وظایف و انتظارات از دانشجو		
			نحوه ارزیابی دانشجو با ذکر نوع ارزیابی (تکوینی/تراکمی)، روش ارزیابی و سهم هر نوع/ روش ارزیابی در نمره نهایی دانشجو، درج شده است.	نحوه ارزیابی دانشجو		
			کتاب‌های درسی، نشریه‌های تخصصی، مقاله‌ها و نشانی وبسایت‌های مرتبط، معرفی شده‌اند	منابع		