

به نام خدا

مبانی داده‌کاوی

ترم دوم ۱۳۹۹-۱۴۰۰

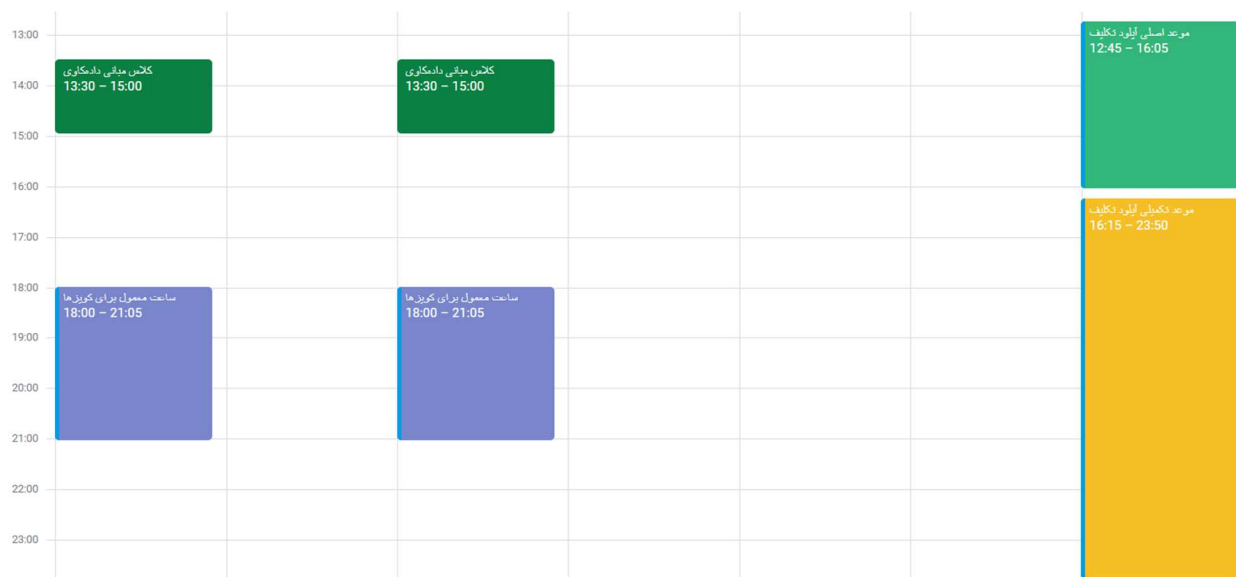
راهنما و قواعد درس

نسخه ۱/۴

در این سند مطالب مورد نیاز در مورد درس مبانی داده‌کاوی در ترم جاری بیان شده است. با توجه به اهمیت آنها و جهت بهره‌برداری هرچه بهتر خودتان از این درس خواهشمندم آن را با دقت مطالعه فرمایید. توضیحات درباره فایل را نیز ان‌شالله در اولین جلسه بیان خواهیم نمود.

محتوی و تشکیل جلسات کلاس

روزهای کلاس شنبه و دوشنبه از ساعت ۱۳:۳۰ تا ۱۵ در برنامه تعیین شده است. جلسات بصورت ترکیبی از محتوای ضبط شده و جلسات آنلاین و رفع اشکال در همین ساعات برنامه‌ریزی می‌شود. سعی خواهیم کرد برای کمک به شما در برنامه‌ریزی ساعات هفتگی خودتان در طول ترم، صرفاً در ساعات تعیین شده در خدمت تان باشیم و بتوانید دروس و برنامه‌های دیگر خود را نیز بخوبی تنظیم نمایید.



نکته: با توجه به محتوای برنامه‌ریزی شده و جهت کسب نتیجه بهتر، توصیه می‌شود صرفاً دانشجویانی در این درس ثبت‌نام نمایند که درس پایگاه داده‌ها را با تسلط گذرانده باشند.

نکته: برای آشنایی بهتر شما با کارهای عملی در این حوزه، طوری برنامه‌ریزی شده است که تکالیف کلاس را با زبان پایتون و با استفاده از امکانات گسترده ارائه شده روی آن انجام داده و کسب مهارت کنید.

توجه: بخش زیادی از این کسب مهارت، به کارهای خارج کلاس خود شما و مطالعات و پیاده‌سازیها روی این بسترها برمی‌گردد. این ترم فرصتی است برای کار بیشتر در این زمینه‌ها که می‌تواند جزو پرکاربردترین حوزه‌ها و مهارت‌ها هنگام ورود شما به بازار کار یا در طول ادامه تحصیلات شما باشد.

مراجع

سه مرجع برای این درس استفاده می‌شود که در ادامه آمده است. علاوه بر مباحث پایه، پیاده‌سازی به زبان پایتون نیز جزو کارهای دانشجویان بخصوص در تکالیف و پروژه خواهد بود:

- BK1 – Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V. (2018). *Introduction to Data Mining*. Pearson Higher Education.
- BK2 – Larose, C. D., & Larose, D. T. (2019). *Data science using Python and R*. John Wiley & Sons.
- BK3 – Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining, 2nd Edition*. John Wiley & Sons.

هر سه مرجع در فولدر درس روی iutbox که آدرس آن در همین سند آمده در دسترس شما قرار گرفته است.

سرفصل‌ها

- Introduction
- Data preparation
- Exploratory data analysis
- Preparing to model the data
- Model evaluation and validation
- Classification
 - Decision trees
 - Naïve Bayes
 - Neural networks
 - K-Nearest neighbors
 - Regression
- Clustering
- Dimension Reduction
- Generalized linear models
- Association rule mining

ارزشیابی

زمینه اولیه «ارزشیابی مستمر» در طول ترم بخصوص در دوران آموزش مجازی مورد تاکید وزارت علوم بوده و کمک می‌کند تا با توجه به آنلاین بودن آزمون پایان‌ترم و محدودیت‌های آزمون آنلاین، فشار کمتری بر دانشجو در یک آزمون خاص وارد شود. به جای آن با صرف وقت بیشتر توسط استاد برای طراحی و دانشجو برای شرکت در آزمون‌ها، ارزیابی در طول ترم پخش می‌شود تا در صورت بروز مشکل برای یک دانشجو در آزمون اصلی، کل نمره او تحت تاثیر قرار نگیرد.

با توجه به زمینه فوق، در ساعات کلاس یا عصر همان روزهای کلاس شنبه یا دوشنبه (بسته به میزان پیشرفت درس که موضوع یا فصل خاص در چه روزی تمام شود) کویزهایی برگزار خواهد شد. سوال این کویزها ممکن است روی سامانه باشد، یا با ارجاع به سوال یا مطلبی که در محتوای ضبط شده یا آنلاین مطرح شده، فقط محل وارد کردن پاسخ برای شما در نظر گرفته شود و صورت سوال به کلیپ ارجاع داده شود.

بارم بندی کلی تقریباً به صورت زیر است:

- کویزها و امتحان (شامل امتحان پایان ترم و میان ترم که برحسب نیاز بصورت کتبی و شفاهی برگزار می شود): ۱۲ الی ۱۳ نمره
- تکالیف و پروژهها: ۵ الی ۶ نمره
- حضور فعال در کلاس، آزمونکها و کارهای مرتبط: ۲ نمره

سهم دقیق هر قسمت برحسب شرایط ترم و پیشرفت کار و وضعیت کلی درس در انتهای ترم مشخص می گردد.

نکته: توجه داریم که «نمره» در واقع چیزی است که خود ما در طول ترم با فعالیتهای تقسیم شده و منظم آن را «می سازیم»، و وقتی خوب کار کنیم و از درس و فعالیتهای مرتبط با آن بهره کافی همراه با فهم عمیق و کسب مهارت را ببریم، نمره نیز بعنوان یک ارزش افزوده همراه آن فعالیتهای خواهد آمد. از طرف دیگر نمره چیزی نیست که بتوان ناگهان در انتهای ترم تامین کرد.

تکلیف

زمینه اولیه «همیشه هدف از تکلیف درسی، همگام شدن دانشجو با موضوع یا فصل تدریس شده، پیش از شروع موضوع یا فصل بعدی است. این موضوع بخصوص در شرایط آموزش مجازی و احتمال جا ماندن دانشجو از یک درس در یک مقطع زمانی، و به دنبال آن جا ماندن از کل بقیه ترم و انبار شدن همه چیز برای روز آخر، مهم تر می شود. توجه فرمایید که انجام دیر هنگام و تمدید مهلت با این هدف سازگار نیست و لذا درخواستهای تمدید قابل اجرا نخواهد بود.

با در نظر داشتن زمینه فوق، ۷ تکالیف برای شما در نظر گرفته شده است که با پایان یک موضوع یا یک فصل، کار روی آن را باید تا روز ثابتی در هفته (ظهر روزهای جمعه در هفته هایی که تکلیف داشته باشید) روی سامانه آپلود کنید. برنامه زمان بندی تکالیف بصورت زیر خواهد بود (توجه فرمایید که زمان بندی با توجه به برنامه پیشرفت درس و میان ترمها و سایر شرایط تنظیم شده و قابل تغییر نمی باشد):

به روزرسانی (۵ اسفند ۹۹): با توجه به اعلام عدم دسترسی به سامانه ها از ۲۰ تا ۲۲ اسفند از طرف مرکز فناوری اطلاعات، دو تاریخ ۲۲ اسفند

در جدول زیر به ۲۳ اسفند تغییر می یابد.

شماره تکلیف	ارسال صورت	پوشش تا جلسه (دوشنبه)	موعد آپلود (جمعه)	تعداد هفته پوشش
HW1	13991201	13991204	13991208	2
HW2	13991215	13991218	13991222	2
HW3	13991222	13991225	14000106	1
HW4	14000120	14000123	14000127	2
HW5	14000210	14000213	14000217	3
HW6	14000224	14000227	14000231	2
HW7	14000314	14000317	14000321	3

نکته مهم - روش دریافت: برای نظم بیشتر و جلوگیری از ضایع شدن حق دانشجویان، هیچ نوع فایلی در طول ترم از هیچ راهی جز سامانه قابل دریافت توسط استاد یا همکاران او نخواهد بود و ارسال فایل با ایمیل و پیام رسان و بلافاصله حذف خواهد شد.

نکته مهم – تشویق منظم بودن در کنار مهلت کمی بیشتر: از نظر درسی حالت ایده‌آل این است که تا ساعت ۱۶ جمعه تکلیف خود را آپلود کنید (صرفاً در هفته‌هایی که تکلیف دارید). اما اگر به هر دلیل تا ساعت ۱۶ موفق به آپلود نشدید، تکلیف خود را تا نیمه شب جمعه می‌توانید آپلود کنید. در انتهای ترم کسانی که کلیه تکالیف خود را سرموقع یعنی قبل از ساعت ۱۶ آپلود کرده باشند از نمره تشویقی تکالیف برخوردار خواهند شد. همچنین پاسخنامه تکالیف تا صبح شنبه روی سامانه قرار خواهد گرفت.

پروژه

برای کسب بهترین تجربیات عملی توسط دانشجویان، پروژه نیز با موضوعات مرتبط تعریف خواهد شد تا دانشجویان در قالب گروه‌های دونفره انجام دهند. گروه تک نفره یا سه نفره قابل پذیرش نخواهد بود.

دسترسی، اطلاع‌رسانی و ارتباط

تشکیل جلسات: جلسات در اتاق ایجاد شده روی سامانه دروس دانشگاه (یکتا) تشکیل می‌شود. اگر نیاز به تشکیل جلسه‌ای در محیط دیگری ایجاد شد اطلاع‌رسانی خواهد گردید

سامانه دروس و ایمیل دانشگاهی: روش اصلی ارتباط با شماست که از طریق ارسال پیامها روی سامانه و اطلاع دادن سامانه به ایمیل شما قابل دسترسی است. لطفاً همه دانشجویان ایمیل خود را فعال کنند و هر روز به آن سر بزنند. ترجیحاً روی گوشی نیز به روشی که راهنمای آن در سایت مرکز فناوری اطلاعات آمده است ایمیل دانشگاهی خود را تنظیم نمایید:

<https://it.iut.ac.ir/services/email>

این کار با برنامه Gmail روی گوشی نیز بصورت افزودن یک اکانت اضافی قابل انجام است.

کانال درس: لینک کانال درس در ادامه آمده است:

<https://b2n.ir/fdm2021ch>

عضویت در کانال الزامی نیست، ولی توصیه می‌شود از آن استفاده کنید. ممکن است برای برخی نظرسنجی‌ها نیز در طول ترم در کانال برای مسائلی که پیش می‌آید در خدمت شما باشیم. کسانی که عضو می‌شوند لازم است اسم و فامیل پروفایل خود را کامل کنند و کاربران بدون اسم و فامیل از کانال حذف خواهند شد. روی پیام‌رسان امکان پاسخگویی به پیام‌های خصوصی وجود ندارد و لازم است از ایمیل استفاده کنید.

ایمیل به استاد: در موارد لازم می‌توانید به استاد ایمیل بزنید. آدرس زیر برای این درس در نظر گرفته شده است، و ایمیل‌های ارسالی به آدرسهای دیگر قابل پاسخگویی نخواهد بود. داخل ایمیل حتما خود را با اسم و فامیل معرفی کنید.

nghadiri+fdm@gmail.com

توجه: ایمیل‌ها با موضوعات زیر قابل بررسی نخواهد بود:

- تمدید مهلت‌ها
- درخواست تبعیض برای یک نفر در تخصیص نمره یا ارسال فایل‌ها در خارج سامانه
- هر نوع درخواست خارج از شرایط کلی مطرح شده در این راهنما

دسترسی به فضای فایل: فایل‌های حجیم درس مانند ویدئوها و مراجع روی سامانه iutbox با آدرس کوتاه زیر در دسترس شما قرار می‌گیرد:

<https://b2n.ir/fdm2021>

با آرزوی توفیق همه شما
قدیری