



## کاوش داده‌گان انبوه

دستیاران آموزشی  
مهدی صادقی (mahdisadeghi@ut.ac.ir)  
شیوا پارسا راد (shiva.parsarad@ut.ac.ir)  
ایمان برازنده (barazandeh.iman@ut.ac.ir)

دکتر سامان هراتی زاده  
دانشگاه تهران - دانشکده علوم و فنون  
ترم اول سال تحصیلی - پاییز ۱۳۹۹

## مجموعه تمرین ششم: سبد سهام خود را بسازید!

مهلت تحویل این تمرین روز جمعه ۱۷ بهمن ساعت ۲۳:۵۹ می‌باشد.

هدف تمرین آشنایی شما با تکنیک‌های خوشه‌بندی است. در این تمرین تحلیل ساده‌ای روی بازار سهام نزدیک<sup>۱</sup> انجام می‌دهید. تحلیل بازار سهام، یکی از زمینه‌های پژوهشی بسیاری از دانشجویان مقطع کارشناسی ارشد و دکتری در آزمایشگاه داده‌کاوی و کشف دانش<sup>۲</sup> در دانشکده می‌باشد. دانشجویانی که این زمینه را برای برای کار و تحقیق انتخاب می‌کنند سعی دارند با استفاده از تکنیک‌های داده‌کاوی و یادگیری ماشین به سرمایه‌گذاران کمک کنند تا سود بیشتری از بازار کسب کنند.

### ۱- (طراحی و پیاده‌سازی الگوریتم)

دیتاست تمرین بازده ماهیانه سهام بازار نزدیک در بازه زمانی ۲۰۰۰ تا ۲۰۰۹ را نشان می‌دهد. با استفاده از تکنیک‌های خوشه‌بندی k-means و BFR این بازار را تحلیل کنید و یک سبد سهام برای سرمایه‌گذاری به ما پیشنهاد دهید.

می‌توانید ابعاد این دیتاست را با استفاده از تکنیک تحلیل مولفه‌های اصلی<sup>۳</sup> (PCA) به دو بعد کاهش دهید. این الگوریتم [در کتابخانه scikit-learn پایتون](#) قرار دارد. همچنین فرایند انتخاب k مناسب را گزارش کنید. در گزارش خود خوشه‌ها را بصورت گرافیکی نمایش داده و مقایسه و تحلیل کنید.

۲- (پژوهش) مقاله‌ای از یک منبع معتبر که از خوشه‌بندی برای تحلیل بازار سهام استفاده کرده است را با جزئیات کافی شرح دهید.

<sup>1</sup> NASDAQ

<sup>2</sup> Knowledge Discovery and Data Mining Laboratory

<sup>3</sup> Principal Component Analysis (PCA)



## کاوش داده‌گان انبوه

دستیاران آموزشی  
مهدی صادقی (mahdisadeghi@ut.ac.ir)  
شیوا پارسا راد (shiva.parsarad@ut.ac.ir)  
ایمان برازنده (barazandeh.iman@ut.ac.ir)

دکتر سامان هراتی زاده  
دانشگاه تهران - دانشکده علوم و فنون  
ترم اول سال تحصیلی - پاییز ۱۳۹۹

## نکات تکمیلی

- از زبان برنامه سازی پایتون کنید.
- کدها با نامگذاری مناسب در پوشه SOURCES قرار گیرد.
- گزارش کامل تمرین را با ساختار مناسب، بدون ابهام و به زبان ساده نوشته و با فرمت PDF در پوشه دیگری قرار دهید. برای نگارش گزارش‌ها نیز امتیاز مثبت و منفی در نظر گرفته شده است، پس برای نگارش گزارش خود وقت مناسبی اختصاص دهید.
- پاسخ تمرین خود را به ایمیل [barazandeh.iman@ut.ac.ir](mailto:barazandeh.iman@ut.ac.ir) ارسال نمایید. هر دو پوشه را در یک فایل ZIP با نام MMDS\_STID\_FullName\_HW# مثلا MMDS\_830123451\_ImanBarazandeh\_HW1 قرار داده و در ایمیلی با همین عنوان نیز ارسال کنید.
- در صورت مشخص شدن هر نوع تقلب انفرادی یا مشترک نمره تمرین صفر در نظر گرفته می‌شود.
- به ازای هر روز تاخیر ۲۵٪ از امتیاز تمرین از دست خواهید داد.
- هر گونه ابهام را با دستیاران آموزشی مطرح کنید.