



A Review of Educational Data Mining and Analysis of Learner Behavior in MOOCs

Hossein Ghaffari ^{1*}, Shohreh Ajoudanian ²

¹ Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
hossein.ghaffari68{at}sco.iaun.ac.ir

² Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
shajoudanian{at}pco.iaun.ac.ir

Abstract

In recent years, data mining has been widely used in the analysis of educational processes and the analysis of learners' behaviors in MOOCs. Therefore, in this study, we investigate different methods for analyzing learners' behavior in MOOCs as well as investigating factors of satisfaction and dissatisfaction and providing appropriate solutions for them. Through this study the most important methods used for data analysis are classification and clustering methods. Also the most important factors that make the students satisfied and dissatisfied with these courses are the reputation of the universities, the quality of the courses, the flexibility of the course, the time bruisse, the lack of motivation, the isolation and the lack of interaction in the course and sufficient scientific background. The results of this study can help to gain a holistic view of the various elements of effective MOOCs and increase the rate of completion of the course.

Keywords: Educational Data Mining, Learning Analysis, MOOC, MOOCs Educational Analysis, Online Learning, E-Learning.



مروری بر داده کاوی آموزشی و تحلیل رفتار فراگیران در موبک ها

حسین غفاری^۱، شهره آجودانیان^۲

^۱ دانشجوی کارشناسی ارشد، دانشکده مهندسی کامپیوتر، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد
hossein.ghaffari68{at}sco.iaun.ac.ir

^۲ عضو هیئت علمی، دانشکده مهندسی کامپیوتر، واحد نجف آباد، دانشگاه آزاد اسلامی، نجف آباد
shajoudanian{at}pco.iaun.ac.ir

چکیده

در سال های اخیر از داده کاوی در تحلیل روند آموزشی و تحلیل رفتارهای فراگیران در موبک ها استفاده زیادی شده است. از این رو در این مطالعه روش های مختلف تحلیل داده ها جهت تحلیل رفتار فراگیران در موبک و همچنین بررسی عوامل رضایت و عدم رضایت و ارائه راهکارهای مناسب برای آنها بررسی می شود. از طریق این مطالعه مهمترین روشهایی که برای تحلیل داده ها استفاده می شود روش های طبقه بندی و خوشه بندی است. همچنین مهمترین عواملی که باعث رضایت و عدم رضایت فراگیران به این دوره ها می شود عبارتند از شهرت دانشگاه ها، کیفیت دوره ها، انعطاف پذیری درس، کبود زمان، نداشتن انگیزه، انزوا و نداشتن تعامل در درس و پس زمینه علمی کافی می باشد. نتایج این بررسی و تحقیق می تواند در به دست آوردن دیدگاه جامع در مورد عناصر مختلف موبک مؤثر و بر افزایش نرخ تکمیل دوره کمک کند.

کلمات کلیدی

داده کاوی آموزشی، تجزیه و تحلیل یادگیری، موبک، تحلیل آموزشی در موبک، آموزش آنلاین، آموزش الکترونیکی.

در سال های اخیر با افزایش چشمگیر محبوبیت موبک، میلیون ها نفر در سرتاسر دنیا از آنها برای یادگیری آنلاین استفاده می کنند. با این حال، نگرانی عمده ای که اغلب درباره موبک وجود دارد این است که اگرچه هزاران نفر برای دوره ها ثبت نام می کنند، اما بخش بسیار کمی در نهایت این دوره ها را می گذرانند. به طور متوسط کمتر از ۱۰ درصد شرکت کنندگان گواهی تکمیل دوره را دریافت می کنند [۱]. بنابراین بررسی عواملی که باعث رضایت و عدم رضایت دانشجویان به این دوره ها می شود می تواند تاثیر زیادی در بهبود این نرخ تکمیل داشته باشد.

۲- داده کاوی آموزشی

آلدوها و همکاران [۲] نشان می دهد که ابعاد تحقیقاتی محققان داده کاوی آموزشی در ارتباط با فعالیت های آموزشی به شکل های مختلفی است: یادگیری مشارکتی کامپیوتری با هدف کاوش الگوهای مشارکتی در مباحث آموزشی، پشتیبانی از معلمان در مدلسازی مشارکت فراگیران، ارزیابی منابع آموزشی دانشگاه و بهبود برنامه تحصیلی، شناسایی عوامل مربوط به موفقیت، شکست و قصد ترک تحصیل، استراتژی ها و برنامه ریزی سازمانی، درک پشتیبانی معلمان و تصمیم گیری اداری. در این خصوص تکنیک داده کاوی

۱- مقدمه

موبک (Massive Open Online Course) که از نظر لغوی به معنی دوره های انبوه یا همگانی درونخطی آزاد است، یک ابتکار آموزشی نسبتاً جدید و امیدوارکننده در زمینه وب است که به دنبال گسترش و ارائه خدمات آموزشی در مقیاس گسترده ای می باشد. موبک الگویی برای ارائه مطالب یادگیری بصورت آنلاین و بدون هیچ محدودیتی برای تقریباً هر شخصی است که می خواهد دوره را بگذراند. شرکت کنندگان می توانند دانشجویانی باشند که در موسسه میزبان موبک ثبت نام کرده اند یا هرکسی در سراسر جهان که با دسترسی به اینترنت ثبت نام کرده اند. موبک درهای یک دوره را باز می کند و هر کسی را به ورود دعوت می کند، در نتیجه پویایی یادگیری جدید، فرصتی برای همکاری و مکالمه قابل توجه برای فراگیران برای گردآوری و بحث درباره محتوای دوره فراهم می کند. همچنین یک بستر یادگیری غنی شده شامل منابع مختلف یادگیری از جمله سخنرانی های ویدیویی، متن ها، تکالیف، امتحانات، انجمن های بحث و گفتگو موضوعی و سایر فن آوری های وب ۲.۰ را ارائه می دهد. این دوره های آنلاین هیچ محدودیتی در اندازه کلاس ندارند.



دهند ولی در عوض ویدیو تماشا می کنند، Disengaging: فراگیرانی که در ابتدای دوره تکالیف را انجام داده اند اما پس از آن دوره را ترک می کنند، Sampling: فراگیرانی که فقط یک یا دو ویدیو را به صورت آزمایشی نگاه می کنند. طراحان و مجریان موبک ها می توانند از این دسته بندی برای تطبیق و توسعه ویژگی های موبک برای فراگیران استفاده کنند.

خلیل و ابنر [۱۰] با بررسی مقالات چاپ شده در این حوزه و بررسی نتایج ۴۲ موبک که بر روی پلتفرم های مختلفی اجرا شده اند به دو سوال مهم پاسخ می دهند. اول اینکه چرا فراگیران موبک را ترک می کنند؟ و دوم اینکه چه تکنیک ها و استراتژی هایی می تواند نرخ ماندگاری در موبک ها را افزایش دهد؟ آنها نشان می دهند که از قابل توجه ترین عوامل، کمبود وقت، عدم انگیزه، احساس انزوا و عدم تعامل در موبک، سابقه و مهارت ناکافی و در نهایت هزینه های پنهان می باشند. و از جمله تکنیک هایی که برای افزایش نرخ ماندگاری و رضایت فراگیران بایستی استفاده شوند می توان به قرار دادن فراگیران در جدول زمانی مختلف، افزایش تعامل بین فراگیر و فراگیر، مدرس و مدرس، و افزایش مهارت های یادگیری آنلاین مانند مهارت های فنی، یادگیری خودمحور و مدیریت زمان اشاره کرد.

کلفت و همکاران [۴] رویکردی ارائه می دهند که روی داده های کلیک-استریم کار می کند و الگوریتم های یادگیری ماشین بر روی داده های هفتگی فراگیران اعمال می شود و تغییرات رفتار فراگیران در طول هفته را بررسی می کنند. داده ها از یک درس روانشناسی جمع آوری شده اند که تعداد ۱۱۶۰۷ فراگیر در ابتدای هفته اول در آن شرکت کرده اند و در پایان هفته آخر ۳۶۸۱ فراگیر باقی مانده اند. با به کار گیری روش ماشین بردار پشتیبان بر روی داده های هفتگی، مدلی ارائه می شود که پیش بینی می کند که آیا فراگیر در هفته بعد فعالیتی خواهد داشت یا نه. بنابراین با این مدل، براساس داده های هفته جاری و هفته های گذشته می تواند ترک تحصیل فراگیران را برای هفته بعد پیش بینی کند.

هالاوا و همکاران [۵] با استفاده از داده های ۱۲ موبک مختلف در زمینه ریاضیات، فیزیک، کشاورزی، علوم سیاسی و علوم کامپیوتر، یک مدل پیش بینی کننده ارائه می دهند تا با استفاده از ویژگی های فعالیت های فردی فراگیران پیش بینی کند که کدام فراگیران در معرض خطر بالای ترک تحصیل هستند. پیش بینی کننده موفق می شود ۶۰ درصد فراگیران را حداقل تا ۲ هفته قبل از ترک تحصیل شناسایی کند. این مدل می تواند به عنوان یک افزونه در سیستم های مدیریت یادگیری در نظر گرفته شود. در این مطالعه به فراگیرانی که در معرض خطر بالای ترک تحصیل قرار دارند اصطلاحاً پرچم قرمز داده می شود. فراگیری در معرض خطر ترک تحصیل قرار دارد که یکی از این دو شرط را داشته باشند: ۱- بیش از یک دوره یک ماهه در درس حضور نداشته باشد. ۲- کمتر از ۵۰ درصد ویدیوها را مشاهده کرده باشد. آنها با استفاده از الگوریتم های یادگیری ماشین دو مدل مختلف طراحی می کنند: یک مدل برای فراگیرانی که هنوز فعال هستند و از روی تحلیل فعالیت های فراگیر عدم انگیزه و رغبت فراگیر به درس را پیش بینی می کنند. و یک مدل دیگر برای فراگیران غایب که با استفاده از تعداد روزهای غیبت فراگیر احتمال ترک تحصیل او حساب می شود. مدل اول تا ۵۰ درصد،

انجمن ها مشارکت نمی کنند، ویدیو تماشا می کنند، در انجمن ها مشارکت می کنند اما در آزمونها شرکت نمی کنند، در آزمون ها شرکت می کنند، در انجمن ها مشارکت می کنند اما ویدیو تماشا نمی کنند، فعالیت در هر سه زمینه: تماشای ویدیو، شرکت در آزمونها و مشارکت در انجمن ها. آنها با استفاده از نظرسنجی های پس از دوره نشان دادند که فراگیران در هر الگویی میزان رضایت متفاوتی را بیان می کنند. قابل توجه است که در این تحقیق، فرض بر این است که فراگیران در همان الگوی مشارکت باقی می ماندند و حرکت طولی آنها بین الگوها و تأثیر آن بر رضایت آنها مورد توجه قرار نمی گیرد.

مور [۱۱] نشان می دهد که نظریه یادگیری و آموزش مستقل و سه نوع مدل تعامل، تعامل فراگیر-محتوا، تعامل فراگیر-مدرس و تعامل فراگیر-فراگیر بر تجربه فراگیر و رضایت از موبک بسیار مفید هستند. براساس نظریه یادگیری و آموزش مستقل، آموزش از راه دور یک سیستم آموزشی است که فراگیر و مدرس در آن از طریق چاپ الکترونیکی و سایر رسانه های غیر انسانی در ارتباط هستند. این سیستم آموزشی از سه زیر سیستم تشکیل می شود که عبارتند از: فراگیر، مدرس و حالت های ارتباطی.

سان و همکاران [۹] یک مدل یکپارچه با شش بعد توسعه می دهند، که شامل: فراگیران، مدرسین، دوره ها، تکنولوژی، طراحی و محیط می باشد. آنها عوامل مهم و موثر بر روی رضایت فراگیران در آموزش الکترونیکی را بررسی می کنند. نتایج نشان می دهد که اضطراب و تنش استفاده از کامپیوتر، نگرش مدرس نسبت به آموزش الکترونیکی، انعطاف پذیری درس، کیفیت درس، مفید بودن درس، راحتی در استفاده، تنوع تکالیف و ارزیابی ها عوامل مهمی هستند که بر روی رضایت کاربر تأثیر گذار هستند.

کیزیلس و همکاران [۳] علی رغم اینکه روی میزان نرخ تکمیل موبک تمرکز کرده اند، یک دید یکپارچه ای از عواملی که مشوق و باعث جدایی فراگیر از فرآیند یادگیری می شوند را نشان می دهد. داده های استفاده شده از سه موبک در سه سطح آموزشی مختلف مربوط به علوم کامپیوتر جمع آوری شده اند، سطح دبیرستان، سطح کارشناسی و سطح کارشناسی ارشد. آنها سه نوع ویژگی مختلف را از داده ها استخراج کرده که شامل ویژگی های دموگرافی، ویژگی های مربوط به تجربه کلی کاربر از موبک (که در قالب سوالات نظرسنجی در پایان دوره از کاربر پرسیده می شوند) و ویژگی های مربوط به فعالیت های اجتماعی کاربر در انجمن های بحث و گفتگو. آنها یک روش طبقه بندی ساده، مقیاس پذیر و آموزنده ارائه می دهند که تعداد کمی از مسیرهای تعامل طولی را در موبک شناسایی می کنند و فراگیران را بر اساس الگوهای تعامل خود با سخنرانی ها و ارزیابی های ویدیویی طبقه بندی می کنند. بدین صورت که در طول دوره، فراگیران براساس اینکه تکلیف را سر موقع انجام می دهند، تکلیف را دیر انجام می دهند، تکلیف انجام نمی دهند ولی ویدیو تماشا می کنند و اصلاً مشارکتی ندارند، برچسب گذاری می شوند و با استفاده از تکنیک های یادگیری ماشین آنها را به ۴ دسته مختلف، دسته بندی می کنند. بنابراین، چهار نوع از فراگیران موبک شناسایی می شوند: Completing: فراگیرانی که اکثر تکالیف ارائه شده در کلاس را تکمیل می کنند، Auditing: فراگیرانی که تکالیف ارائه شده را به ندرت انجام می-



(سرگرمی، افزایش مهارت، افزایش دانش و ...) ۳- انزوا و نداشتن تعامل در ماک: به خاطر طراحی نادرست ماک و نداشتن تعامل با محتوا، فراگیر از ادامه دادن به درس منصرف می شود. ۴- نداشتن پس زمینه علمی کافی: باعث عدم توانایی در حل مسائل و یادگیری مفاهیم می شود و فراگیر بی انگیزه می شود. ۵- به روز نبودن مطالب دوره: باعث عدم انگیزه برای تکمیل دوره می شود. ۶- کمبود تمرین و مثال در محتوای دوره: ایهامات موجود در یادگیری مفاهیم برطرف نمی شود. ۷- طولانی بودن دوره: طولانی بودن دوره نیاز به زمان بندی طولانی مدت دارد و به مرور باعث منصرف شدن فراگیر از ادامه دوره می شود.

از جمله روشهای پیشنهادی برای افزایش نرخ ماندگاری فراگیران در ماک ها می توان به موارد زیر اشاره کرد: ۱- قرار دادن فراگیران در جداول زمانی مختلف تا زمان کافی برای انجام ارزیابی ها و مشاهده ویدیوها داشته باشند. ۲- ترغیب دانشجو به تکمیل دوره مثلا با دادن گواهی پایان دوره جهت ارتقاء رزومه کاری، مشارکت استاد در انجمن های بحث و گفتگو و رفع مشکلات فراگیران ۳- محدود کردن طول ویدیوهای سخنرانی تا باعث خستگی فراگیر نشود. ۴- استفاده از شکل و تصاویر مختلف با کیفیت در ویدیوها جهت تعامل بیشتر فراگیر با مطالب مطرح شده. ۵- مطرح کردن مباحث عملی و تئوریک در کنار یکدیگر جهت کاربردی کردن دوره. ۶- قرار دادن عناوین مناسب برای سرفصل ها و محتوا ۷- قرار دادن مسائل کاربردی و به روز جهت ترغیب فراگیر به حل آنها.

مراجع

- [1] S. Shrader, M. Wu, D. Owens, and K. Santa Ana, "Massive open online courses (MOOCs): Participant activity, demographics, and satisfaction," *Online Learn.*, vol. 20, pp. 199–2016, 2016.
- [2] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37. Elsevier Ltd, pp. 13–49, 01-Apr-2019.
- [3] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses," 2013.
- [4] N. Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in *Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, 2014*, pp. 60–65.
- [5] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," *Proc. Second Eur. MOOC Stakehold. summit*, vol. 37, no. 1, pp. 58–65, 2014.
- [6] C. P. Yang, Diyi and Sinha, Tanmay and Adamson, David and Ros{\e}, "Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses," in *Proceedings of the 2013 NIPS Data-driven education workshop*, 2013, p. 14.
- [7] L. Rai and D. Chunrao, "Influencing factors of success and failure in MOOC and general analysis of learner

پیش بینی ترک تحصیل را برای فراگیران فعال به درستی انجام می دهد و مدل دوم بالای ۹۰ درصد پیش بینی ترک تحصیل را برای فراگیران غایب با آستانه غیبت ۱۴ روز به درستی انجام می دهد.

رای و همکاران [۷] با استفاده از تحلیل های آماری نشان می دهند که موفقیت و شکست در یادگیری آنلاین بیشتر به عوامل شخصی بستگی دارد تا عواملی که تحت تاثیر محیط اطراف یا محیط خارجی قرار می گیرند. رفتار فردی دانش آموزان نقش مهمی در یادگیری دارد. با این وجود، این ممکن است در مورد یادگیری مبتنی بر کلاس یا شیوه های سنتی یادگیری که تحت آزمایش قرار گرفته اند، اتفاق نیفتد، چون که چندین دانش آموز در کنار هم یاد می گیرند و همچنین معلمان بر عملکرد دانش آموزان کم انگیزه تاثیر می گذارند. بیشتر این فراگیران مجذوب شهرت دانشگاه ها، کیفیت دوره ها و سرگرمی که در حل چالش برانگیز تکالیف وجود دارد هستند.

۴- نتیجه

عوامل مختلفی در روند موفقیت و شکست ماکها تاثیر دارند. شهرت مدرس و دانشگاه یا موسسه ارائه دهنده دوره تاثیر بسزایی در جذب فراگیران خواهد داشت. معمولا خیلی از فراگیران فقط روی جمع کردن منابع و ویدیوهای آموزشی متمرکز هستند بدون اینکه هیچگونه استفاده ای از آنها بکنند. از طرفی آنها با جمع کردن منابع مطالعاتی مختلف و متنوع بدون اینکه در انتخاب آنها راهنمایی گرفته باشند (از نظر سطح علمی) باعث سردرگمی خود می شوند. بیشتر فراگیران علاقه زیادی به یادگیری از طریق ویدیوهای سخنرانی دارند و نسبت به مطالعه متن رغبت کمتری نشان می دهند. این امکان برای فراگیر که بتواند مساله یا مشکل مورد نظر خود را در یک انجمن بحث و گفتگو مطرح کند و آن را مورد بحث قرار دهد، امکان بسیار مطلوبی می باشد. دشواری دوره و دردسترس نبودن مدرس جهت راهنمایی کردن می تواند یکی از عوامل دلسرد شدن فراگیر نسبت به دوره باشد. از طرفی مثل دوره های سنتی، مدرسی به صورتی حضوری وجود ندارد که انگیزه دانش آموز را به ادامه دوره بیشتر کند. حجم مطالب بالا و کار زیاد در دوره میتواند فراگیرانی که وقت کافی ندارند را نسبت به ادامه دوره بی میل کند. گاهی اوقات نیاز است تا فراگیر در یک برنامه زمان بندی خاص قرار بگیرد تا بتواند خودش را با دوره تطبیق دهد. اگر زمان بندی دوره به شکلی نباشد که فراگیر در داخل یک چهارچوب قرار بگیرد می تواند بعضی فراگیران را نسبت به صورت ناخواسته نسبت به دوره ناراضی کند. گاهی اوقات طول ویدیوی سخنرانی بیش از اندازه زیاد است یا صدای مدرس کیفیت خوبی ندارد و می تواند باعث ناراضی فراگیر شود. برخی فراگیران برای افزایش مهارت در حرفه ی مورد نظر خود در این دوره ها شرکت می کنند و اگر محتوای دوره به جای مباحث کاربردی و عملی بیشتر شامل مباحث تئوری شود، فراگیر مایل به ادامه دوره نمی شود. علاوه بر موارد ذکر شده عوامل دیگری نیز موثر هستند که در ادامه برخی آنها را ذکر می کنیم: ۱- کمبود زمان: کامل کردن ماک وقت بسیار زیادی می گیرد و مشاهده ویدیوها، انجام تکالیف و ارزیابی ها نیازمند زمان زیادی هستند. ۲- نداشتن انگیزه: برآورده نکردن نیازهای فراگیر



- behavior,” *Int. J. Inf. Educ. Technol.*, vol. 6, no. 4, p. 262, 2016.
- [8] B. G. Gameel, “Learner Satisfaction with Massive Open Online Courses,” *Am. J. Distance Educ.*, vol. 31, no. 2, pp. 98–111, Apr. 2017.
- [9] P. C. Sun, R. J. Tsai, G. Finger, Y. Y. Chen, and D. Yeh, “What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction,” *Comput. Educ.*, vol. 50, no. 4, pp. 1183–1202, 2008.
- [10] H. Khalil and M. Ebner, “MOOCs completion rates and possible methods to improve retention-A literature review L3T Lehrbuch für Lernen und Lehren mit Technologien View project Successful transition from Secondary to Higher Education using Learning Analytic (STELA) View project,” 2014.
- [11] M. G. Moore, “Toward a Theory of Independent Learning and Teaching,” *J. Higher Educ.*, vol. 44, no. 9, pp. 661–679, Dec. 1973.