



پژوهشکده‌ی آمار

بسم الله الرحمن الرحيم

شناسنامه نشریه	
SO-1-11-1401-29	کد گزارش
شناسایی الگوهای فعالیت کاربران اینترنت	فارسی
Identifying The Internet Users Activity Patterns	انگلیسی
کاربران اینترنت، یادگیری ماشین، خوشه‌بندی، رده‌بندی	واژه‌های کلیدی
مریم میرزایی و دکتر محمد شیری	نویسندگان
دکتر علیرضا رضایی	ناظر علمی
پژوهشکده‌ی آمار،	ناشر
پژوهشکده‌ی آمار، تهران، خیابان دکتر فاطمی، خیابان باباطاهر، خیابان شهید فکوری، شماره‌ی ۱۴۵. کد پستی: ۱۴۱۳۷۱۷۹۱۱	نشانی
فروردین ۱۴۰۲	زمان انتشار
۳-۴۴۰-۸۸۶۳۰	تلفن
research@srtc.ac.ir	پست الکترونیک

♦ مسئولیت آرا و نظرهای ارائه‌شده در گزارش بر عهده‌ی نویسنده یا نویسندگان است.

♦ حق مالکیت معنوی این گزارش تحلیلی متعلق به پژوهشکده‌ی آمار است و نقل مطالب فقط با ذکر مأخذ مجاز است.

صفحه‌بندی:

طاهره امینی

طرح جلد:

ساناز مهندسی



شناسایی الگوهای فعالیت کاربران اینترنت

مریم میرزایی و محمد شیری



* مقدمه

مجازی و یافتن الگوهای استفاده از آن است؛ به این صورت که افراد با چه خصوصیات و ویژگی‌هایی چه نوع و چه مدت زمانی چه فعالیت‌هایی را دنبال می‌کنند. برای این منظور و دستیابی به این اهداف روش‌های یادگیری ماشین^۱ مورد استفاده قرار گرفته است. به طور کلی می‌توان گفت در طول تحقیق به سؤالات زیر پاسخ داده شده است:

- ◆ کدام عوامل جمعیت‌شناختی بر نوع استفاده از اینترنت و فضای مجازی تأثیرگذار هستند؟
- ◆ الگوی استفاده از اینترنت و فضای مجازی به چه صورت است؟
- ◆ به طور کلی افراد با چه ویژگی‌هایی مانند سن، تحصیلات، وضعیت تأهل، وضعیت اشتغال و ... از چه الگوی مصرفی اینترنت تبعیت می‌کنند؟
- ◆ کدام روش یادگیری ماشین برای تحلیل داده‌های اینترنت و فضای مجازی مطرح شده‌اند؟
- ◆ کدام روش برای استخراج الگوی استفاده از فضای مجازی و اینترنت مفید است؟

گذران وقت در فضای مجازی و اینترنت یکی از موضوع‌های مهمی است که امروزه در زندگی بشر جا باز کرده است و بسیاری از افراد وقت زیادی از زندگی روزمره خود را صرف فعالیت در فضای مجازی و اینترنت می‌کنند. به طوری که بسیاری از افراد برای رفع بسیاری از نیازهای خود به اینترنت و فضای مجازی رو می‌آورند. حال ممکن است این فعالیت‌ها منجر به اتلاف وقت و یا صرفه‌جویی در وقت فرد شود. فعالیت‌هایی مانند خرید، آموزش، سرگرمی، دستیابی به اطلاعات عمومی، تماشا کردن برنامه‌ها یا گوش کردن به موسیقی، جستجو در شبکه‌های اجتماعی، دانلود و آپلود کردن اطلاعات، چک کردن ایمیل و ... می‌تواند فرد را ترغیب به استفاده از رایانه و اینترنت کند. ویژگی‌های جمعیت‌شناختی افراد مانند سن، جنسیت، تحصیلات، وضعیت تأهل، شغل، شهری یا روستایی بودن، تعداد اعضای خانوار و ... می‌توانند بر نوع استفاده افراد از اینترنت و فضای مجازی تأثیرگذار باشد. در واقع هدف این مطالعه شناسایی و بررسی عوامل مؤثر بر اینترنت و فضای

مریم میرزایی کارشناس گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار هستند.

محمد شیری عضو هیئت علمی و مدیر گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار هستند.

* بیان مسئله و روش‌شناسی

بسته به نوع داده‌هایی که تحلیل می‌شوند، مورد استفاده قرار می‌گیرند. برای نمونه داده‌هایی که تنها شامل متغیرها یا ویژگی‌های عددی هستند، روش k -میانگین، k -نماینده و k -میانه، برای داده‌هایی که شامل تنها متغیرهای رسته‌ای هستند روش k -مد و برای داده‌هایی که شامل هر دو نوع متغیر عددی و رسته‌ای هستند، k -نمونه اولیه مورد استفاده قرار می‌گیرند. از آنجایی که هدف در تحلیل داده‌های اینترنت و فضای مجازی کشف الگوهای پنهان موجود در داده‌ها است و داده‌های مورد مطالعه فاقد برچسب رده است برای این منظور مناسب‌ترین روش، به کارگیری روش‌های خوشه‌بندی است.

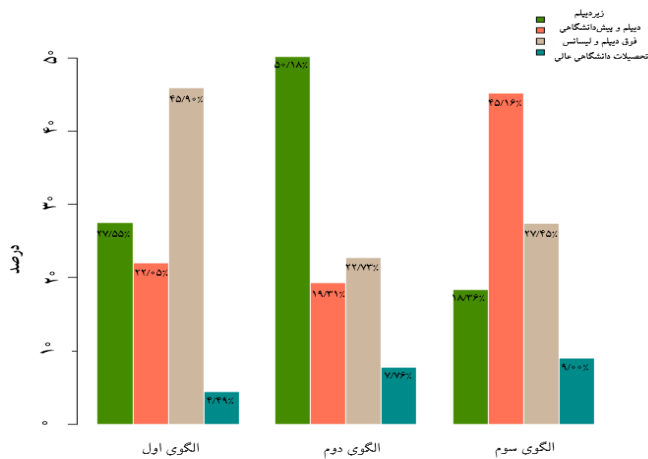
* نتایج

با اجرای خوشه‌بندی k -نمونه اولیه با تعداد خوشه ۳ (تعداد خوشه با روش اعتبارسنجی متقابل حاصل شده است) روی داده‌های گذران وقت و فضای مجازی استخراج‌شده از داده‌های طرح گذران وقت مرکز آمار ایران برای سال ۱۳۹۹ الگوهای حاصل از نتایج خوشه‌بندی را به طور کلی می‌توان به سه الگوی زیر بیان کرد (هر الگو بر اساس بیش‌ترین درصد‌های استفاده از اینترنت و فضای مجازی ساخته شده است برای نمونه با توجه به شکل‌های ۱ و ۲ در الگوی اول سطح تحصیلات فوق دیپلم و لیسانس و وضعیت اشتغال در حال تحصیل بیش‌ترین درصد استفاده از اینترنت و فضای مجازی را دارد برای این منظور به‌عنوان ویژگی‌های الگوی اول آورده شده‌اند، سایر الگوها نیز به همین صورت ساخته شده‌اند. لازم به ذکر است امکانات رفاهی شامل موارد خودرو شخصی، موتورسیکلت، دوچرخه، رادیو، تلویزیون، انواع ویدیو، انواع رایانه و تبلت، تلفن همراه، تلفن ثابت، یخچال، فریزر و ... هستند که بر اساس تعداد امکانات رفاهی در اختیار خانوار به سه رده‌ی کم، متوسط و زیاد تقسیم شد):

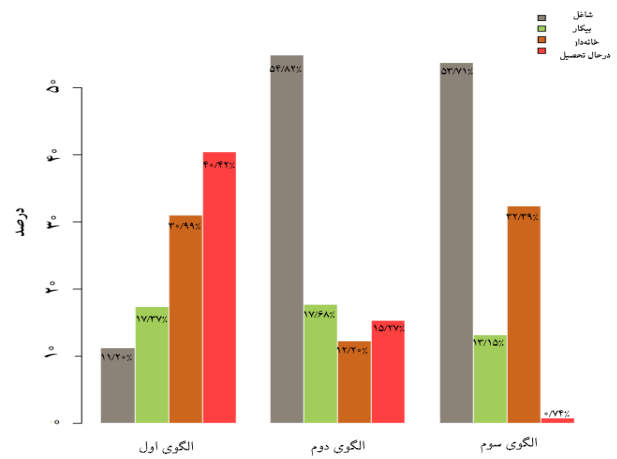
الگوی اول: جنس زن، میانگین گروه سنی ۲۴/۷، سطح

به منظور پاسخ به سوالات مطرح شده در بخش مقدمه لازم است الگوهای استفاده افراد از اینترنت و فضای مجازی را به دست آوریم. به‌منظور کشف این الگوها می‌توان از روش‌های جدید مطرح شده از جمله روش‌های یادگیری ماشین استفاده کرد. از مزایای این روش‌ها این است که در حین اینکه تجربه به دست می‌آورد، دقت و کارایی‌شان نیز افزایش می‌یابد. این به آن‌ها امکان تصمیم‌گیری بهتر می‌دهد. یکی دیگر از مزایای یادگیری ماشین این است که قابلیت کار با هر نوع داده‌ای را دارد. با استفاده از الگوریتم‌های یادگیری ماشین می‌توان روندها و الگوهای خاصی را که برای انسان‌ها ناشناخته هستند، کشف کرد و با آموزش مدل با استفاده از داده‌های پیشین می‌توان برای داده جدید که مشاهده‌نشده پیش‌بینی‌هایی را انجام داد.

در مطالعه‌ی حاضر روش‌های مختلف یادگیری ماشین از جمله روش‌های رده‌بندی^۲ و خوشه‌بندی^۳ مورد مطالعه قرار گرفتند و نهایتاً روشی که مناسب تحلیل و کشف الگوهای استفاده از اینترنت و فضای مجازی بود، مورد استفاده قرار گرفت. از جمله روش‌های رده‌بندی که مورد مطالعه قرار گرفتند شامل درخت تصمیم، بیز ساده، رگرسیون لوجیت چندجمله‌ای و ماشین بردار پشتیبان هستند، روش‌های رده‌بندی مناسب داده‌هایی است که دارای برچسب رده هستند و مدل بر مبنای داده‌های آموزشی که دارای برچسب رده هستند ساخته می‌شود و در نهایت اگر مدل از دقت کافی برخوردار بود برای پیش‌بینی رده‌ی داده جدید مورد استفاده قرار می‌گیرد. از جمله روش‌های خوشه‌بندی که مورد مطالعه گرفتند شامل k -میانگین، فازی، k -نماینده، k -میانه، k -مد و k -نمونه اولیه هستند. روش‌های خوشه‌بندی مناسب داده‌هایی است که بدون برچسب هستند و به‌منظور کشف الگوهای پنهان موجود در داده‌ها مورد استفاده قرار می‌گیرند. استفاده از هر یک از این روش‌ها



شکل ۲- الگوی اول تا سوم استفاده از اینترنت و فضای مجازی برای سطح تحصیلات



شکل ۱- الگوی اول تا سوم استفاده از اینترنت و فضای مجازی برای وضعیت اشتغال

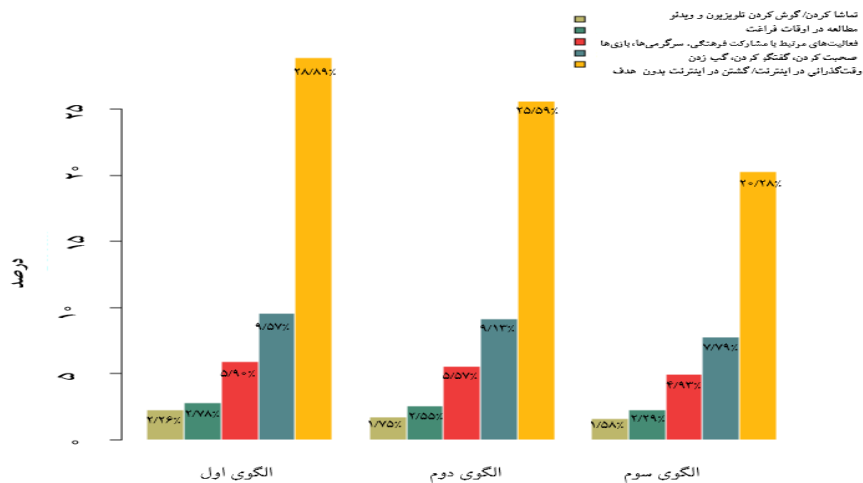
سه رده‌ی کم، متوسط و زیاد تقسیم شد، سپس برای هر الگو به صورت جداگانه درصد استفاده از اینترنت و فضای مجازی محاسبه شد و نهایتاً رده‌ی مصرف زیاد جهت مقایسه الگوهای ذکرشده در نظر گرفته شد (شکل ۳):

♦ بیش‌ترین علاقه‌مندی افراد در سه الگوی اول، دوم و سوم در اینترنت به وقت‌گذرانی در اینترنت، گشتن در اینترنت بدون هدف اشاره دارد (این امر را به وضوح می‌توان در شکل ۳ مشاهده کرد؛ درصد وقت‌گذرانی و گشتن در اینترنت بدون هدف نسبت به سایر فعالیت‌ها در هر سه الگوی ذکرشده بیش‌تر است) که لزوم توجه برنامه‌ریزان را به آن دسته افرادی با ویژگی‌های جمعیت‌شناختی الگوی اول، دوم و سوم مشخص می‌نماید. این روند مخصوصاً در افرادی با خصوصیات و ویژگی‌های الگوی اول که شامل جوانان و محصلین است، مشاهده می‌شود و به دلیل حساسیت‌های خاص این گروه سنی، شناسایی نیازهای این گروه و همچنین برنامه‌ریزی در راستای استفاده‌ی بهینه و مناسب از اینترنت، می‌تواند یکی از موضوعات مورد توجه سیاست‌گذاران و برنامه‌ریزان باشد. لازم به ذکر است با توجه به شکل ۳ اختلاف معناداری بین درصد استفاده از اینترنت در زمینه وقت‌گذرانی و گشتن در اینترنت بدون هدف به وضوح در سه الگوی ذکرشده دیده می‌شود. همچنین اختلاف بین درصد استفاده از اینترنت بین سه الگوی ذکرشده در موارد صحبت کردن/گفتگو

تحصیلات فوق دیپلم و لیسانس، بدون همسر، در حال تحصیل، امکانات رفاهی کم، نوع مسکن غیر آپارتمانی و تعداد اعضای خانوار چهار و بیش‌تر.

الگوی دوم: جنس مرد، میانگین گروه سنی ۲۸/۸۲، سطح تحصیلات زیر دیپلم، بدون همسر، شاغل، امکانات رفاهی زیاد، نوع مسکن غیر آپارتمانی و تعداد اعضای خانوار سه. الگوی سوم: جنس مرد، میانگین گروه سنی ۴۰/۶۳، سطح تحصیلات دیپلم و پیش‌دانشگاهی، دارای همسر، شاغل و امکانات رفاهی زیاد، نوع مسکن آپارتمانی و تعداد اعضای خانوار چهار و بیش‌تر.

با توجه به سه الگوی استفاده از اینترنت و فضای مجازی به دست آمده، نتایج زیر را می‌توان استخراج نمود (در تحلیل نتایج در هر الگو یک سری رده‌های ایده‌آل از هر ویژگی یا متغیر بر اساس بیش‌ترین درصدها به‌عنوان ملاک برای مقایسه الگوهای مختلف در نظر گرفته شده است؛ مدت زمان استفاده از اینترنت و فضای مجازی به‌منظور انجام هر یک از فعالیت‌های وقت‌گذرانی در اینترنت به عبارتی گشتن در اینترنت بدون هدف؛ صحبت کردن/گفتگو کردن و گپ زدن؛ مشارکت‌های فرهنگی، سرگرمی و بازی‌ها؛ مطالعه در اوقات فراغت؛ تماشا کردن/ گوش کردن تلویزیون و ویدئو به



شکل ۳- الگوی اول تا سوم استفاده از اینترنت و فضای مجازی.

چندانی از اینترنت در این زمینه نداشته باشد و افرادی که دارای سن بالاتری هستند نیز به دلیل عادتی که از گذشته به همراه دارند، تمایلی به استفاده از اینترنت و فضای مجازی در این زمینه ندارند.

♦ افرادی با الگوی اول به نسبت الگوی دوم و سوم علاقه‌مندی بیشتری به استفاده از اینترنت جهت تماشا کردن/گوش کردن تلویزیون و ویدئو دارند. وضعیت اشتغال، سن و وضعیت تأهل در استفاده از اینترنت در این زمینه بسیار تأثیرگذار است؛ افراد در حال تحصیل و خانه‌دار به نسبت دیگر گروه‌ها استفاده بیشتری در این زمینه دارد و این حاکی از آن است که ممکن است افراد در حال تحصیل و خانه‌دار دارای اوقات فراغت بیشتری باشند و در نتیجه به نسبت در این زمینه فعالیت بیشتری دارند.

♦ افراد با الگوی اول به نسبت الگوهای دوم و سوم از اینترنت جهت مطالعه در اوقات فراغتشان بیشتر استفاده می‌کنند. به عبارتی زنان جوان بدون همسر و در حال تحصیل با تحصیلات فوق دیپلم و لیسانس علاقه و انگیزه بیشتری نسبت به سایر گروه‌ها به استفاده از اینترنت در زمینه مطالعه در اوقات فراغت دارند. این ممکن است به این دلیل باشد که افرادی با چنین ویژگی‌هایی دارای اوقات فراغت نسبتاً بیشتری نسبت به سایر گروه‌ها باشند.

♦ افرادی با الگوی اول به نسبت الگوهای دوم و سوم استفاده

کردن و گپ زدن؛ مشارکت‌های فرهنگی، سرگرمی و بازی‌ها؛ مطالعه در اوقات فراغت؛ تماشا کردن/گوش کردن تلویزیون و ویدئو زیاد نیست، اما دارای الگوهای جمعیت‌شناختی بسیار متفاوت است. همچنین به طور کلی می‌توان استنتاج نمود که اولویت‌های استفاده از اینترنت و فضای مجازی در افرادی که دارای هر یک از سه الگوی ذکر شده هستند به ترتیب عبارتند از: وقت‌گذرانی در اینترنت به عبارتی گشتن در اینترنت بدون هدف؛ صحبت کردن/گفتگو کردن و گپ زدن؛ مشارکت‌های فرهنگی، سرگرمی و بازی‌ها؛ مطالعه در اوقات فراغت؛ تماشا کردن/گوش کردن تلویزیون و ویدئو.

♦ افرادی که دارای الگوی جمعیت‌شناختی سوم هستند به نسبت الگوهای اول و دوم علاقه‌مندی کمتری نسبت به استفاده از اینترنت جهت صحبت کردن، گفتگو کردن و گپ زدن که شامل فعالیت‌های چت کردن، صحبت تلفنی و تصویری با هدف معاشرت با دیگران است، دارند (در شکل ۳ می‌توان مشاهده کرد که درصد استفاده از اینترنت جهت صحبت کردن، گفتگو کردن و گپ زدن در الگوی سوم نسبت به سایر فعالیت‌ها بیشتر است) که به‌ویژه وضعیت تأهل و سن افراد در استفاده از اینترنت و فضای مجازی در این زمینه تأثیرگذار است به عبارتی افراد بدون همسر و جوان علاقه‌مندی نسبتاً بیشتری را نسبت به افراد متأهل و مسن جهت استفاده از اینترنت در راستای تماس‌های اینترنتی دارند که این احتمال وجود دارد افراد متأهل به دلیل مشغله‌ای که دارند فرصت

توضیحات

^۱Machine Learning

^۲Classification

^۳Clustering

منابع

- چکیده نتایج طرح آمارگیری گذران وقت، تابستان (۱۳۹۹). تهران: مرکز آمار ایران.
- رضائی قهرودی، زهرا؛ رنجی، حسن؛ رضایی، علیرضا. (۱۴۰۰). استفاده از الگوریتم‌های یادگیری آماری رده‌بندی در آمار رسمی. مجله علوم آماری-نشریه علمی پژوهشی انجمن آمار ایران. ۱۵(۱). ۱۱۹-۱۴۶.

- Abe, S. (2012). Pattern classification: neuro-fuzzy methods and their comparison. Springer Science & Business Media.
- Abe, S., and Inoue, T. (2002, April). Fuzzy support vector machines for multiclass problems. In ESANN (pp. 113-118).
- Ahmad, I. (2014). K-Mean and K-Prototype Algorithms Performance Analysis. International Journal of Computer and Information Technology, 3, 823-828.
- Bengio, Y., Goodfellow, I., and Courville, A. (2017). Deep learning (vol. 1). MIT Press.
- Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Springer, Boston, MA.
- Zhou, H., Zhang, Y., and Liu, Y. A. (2017). Global-Relationship Dissimilarity Measure for the k-Modes Clustering Algorithm. Computational Intelligence and Neuroscience. Vol. 2017

بیش‌تری از اینترنت جهت مشارکت‌های فرهنگی، سرگرمی و بازی‌ها دارند. به عبارتی می‌توان اینطور تحلیل نمود که مردان متأهل و شاغل با تحصیلات دیپلم و پیش‌دانشگاهی با میانگین سنی ۴۰/۶۳ در زمینه مشارکت‌های فرهنگی، سرگرمی و بازی‌ها نسبت به سایر گروه‌ها کمترین استفاده از اینترنت را دارند. شاغل بودن و متأهل بودن افراد در کمبود وقت افراد و پرداختن به فعالیت‌های غیر ضروری می‌تواند تأثیرگذار باشد. همین‌طور می‌توان بیان کرد که افرادی با الگوی دوم نسبت به الگوی سوم استفاده بیش‌تری از اینترنت در این زمینه دارند. به طور کلی می‌توان گفت مردان متأهل و شاغل با تحصیلات دیپلم و پیش‌دانشگاهی با میانگین سنی ۴۰/۶۳ نسبت به سایر گروه‌ها استفاده‌ی هدفمندتری از اینترنت و فضای مجازی دارند.

* نتیجه‌گیری

در مطالعه حاضر سعی ما بر این است که با کشف الگوهای پنهان استفاده از اینترنت و فضای مجازی به سؤالات مطرح‌شده در بخش مقدمه پاسخ داده شود. برای این منظور روش خوشه‌بندی را مورد استفاده قرار دادیم و از آنجایی که داده‌های ما شامل هر دو نوع ویژگی‌های عددی و رسته‌ای هستند، روش خوشه‌بندی k-نمونه اولیه را روی داده‌های گذران وقت در فضای مجازی و اینترنت اجرا کردیم. نهایتاً سه الگوی استفاده از اینترنت و فضای مجازی حاصل شد که به طور کامل در بخش نتایج مورد بحث قرار گرفتند. با توجه به اهمیت موضوع اینترنت و فضای مجازی در عصر جدید، از نتایج حاصل از این مطالعه می‌توان به‌منظور اتخاذ تصمیمات و برنامه‌ریزی در زمینه فعالیت‌های اینترنتی مناسب برای گروه‌های مختلف افراد استفاده نمود، همچنین می‌توان از آن به‌عنوان پیش‌زمینه‌ای برای کارهای آینده در حوزه استفاده بهینه از اینترنت و فضای مجازی استفاده کرد.