



رابطه‌ی بین طبقات اجتماعی و شبکه‌های اجتماعی

طناز حافظی لای

استاد راهنما: دکتر عبدالله صفری

پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

تابستان ۱۴۰۱

چکیده

در این پروژه می‌خواهیم با استفاده از اطلاعات درباره‌ی رابطه‌ی بین طبقات اجتماعی و شبکه‌های اجتماعی اطلاعات کسب کنیم. برای رسیدن به هدف پروژه از دو مجموعه داده استفاده کردیم که یکی از آن‌ها مجموعه داده آماده موجود در سایت www.kaggle.com و مجموعه داده دیگر با استفاده از پرسشنامه جمع‌آوری شده است.

بعد از بررسی‌های انجام شده به صورت کلی می‌توانیم بگوییم که در قشر مرفه مدت زمان بیشتری صرف شبکه‌های اجتماعی می‌شود و هرچقدر افراد از نظر اقتصادی ضعیف‌تر باشند مدت زمان کمتری را صرف شبکه‌های اجتماعی خواهند کرد. البته که این اختلاف چشم‌گیر نیست و می‌توان گفت شبکه‌های اجتماعی برای تمامی طبقات اجتماعی قابل دسترس و مورد استفاده است. دلایل استفاده از شبکه‌های اجتماعی در طبقات مختلف تقریباً یکسان است و موضوعات محبوب نیز در طبقات مختلف تفاوت چندانی ندارند.

سپاس‌گزاری

با تشکر از استاد محترم دکتر عبدالله صفری که مرا در این پروژه یاری کردند.

پیشگفتار

علوم داده‌های اجتماعی یکی از حوزه‌های نوظهور در کل دنیا است. در این حوزه سعی می‌شود تا با استفاده از اطلاعات مشکلات و مسائل اجتماعی را واکاوی کرد و با تجزیه و تحلیل داده‌های اجتماعی، مسائل و مشکلات اجتماعی را کشف و در جهت بهبود آن حرکت کرد. در این پروژه می‌خواهیم با استفاده از داده‌های موجود و همچنین جمع‌آوری اطلاعات، رابطه‌ی بین طبقات اجتماعی که زیر مجموعه‌ای از ساختار هویت اجتماعی است را با شبکه‌های اجتماعی موجود بررسی کنیم.

شبکه‌های اجتماعی که امروزه عضو جدانشدنی از زندگی انسان مدرن است، تاثیرات بسزایی در رفتارهای اجتماعی افراد دارد. این تاثیرات می‌تواند مفید یا مضر باشد اما قطعاً در سبک زندگی افراد جامعه تاثیر به‌سزایی خواهند داشت. نکته حائز اهمیت درباره‌ی شبکه‌های اجتماعی این است که رفتار گروه‌های مختلف در آن می‌تواند متفاوت و یا همگن باشد. به عبارتی افراد موجود در طبقات اجتماعی مختلف می‌توانند رفتارهای متفاوتی و یا یکسانی در آن داشته باشند، می‌توانند به شکل‌ها و اهداف متفاوتی یا یکسانی از شبکه‌های اجتماعی استفاده کنند. این موضوع نیاز به بررسی دارد که ما در ادامه به آن خواهیم پرداخت. از طریق این لینک زیر نیز می‌توانید به کدها و مجموعه داده‌های مورد استفاده در این پروژه دسترسی پیدا کنید.

لینک :

(<https://drive.google.com/drive/folders/1h2b-dgZtxqa2dYi4ECe6MA3f-dHPduL0?usp=sharing>)

فهرست مطالب

۶	۱	هویت اجتماعی
۶	۲	طبقات اجتماعی
۶	۳	علوم داده‌های اجتماعی
۶	۴	داده‌های پروژه
۶	۱.۴	بخش اول
۷	۲.۴	بخش دوم
۷	۱.۲.۴	طراحی پرسشنامه
۷	۲.۲.۴	جمع‌آوری اطلاعات
۷	۵	مراحل تحلیل داده‌ها
۷	۱.۵	بخش اول
۱۴	۲.۵	بخش دوم
۲۶	۶	نتیجه‌گیری
۲۶	۱.۶	بخش اول
۲۸	۲.۶	بخش دوم
		واژه نامه ۳۰ و واژه نامه ۳۱

۱ هویت اجتماعی

هویت‌های اجتماعی برچسب‌هایی هستند که افراد برای دسته‌بندی یا شناسایی خود و یا دیگران به عنوان اعضای گروه‌های خاص استفاده می‌کنند. برخی از هویت‌های اجتماعی رایج عبارتند از: نسل، قومیت، نژاد، مذهب، جنسیت، گرایش جنسی، ملیت، وابستگی سیاسی، موقعیت اجتماعی-اقتصادی، طبقه‌ی اجتماعی و

۲ طبقات اجتماعی

وقتی جامعه‌شناسان از طبقه اجتماعی صحبت می‌کنند، به گروهی از افراد اشاره می‌کنند که موقعیت مشابه‌ای در نظام اقتصادی تولید دارند یا گروه‌بندی وسیعی از افراد که دارای منابع اقتصادی مشترکی هستند. به عبارتی از نظر اقتصادی شبیه به یکدیگر هستند؛ در این سیستم شغل بسیار مهم و تعیین کننده است. البته توجه داشته باشید که در اغلب نوشته‌های مارکس مفهوم طبقه به صورت آزادانه و بدون تصریح تعریفی رسمی به کار رفته است. تقسیم بندی‌های متفاوتی برای طبقات اجتماعی در طول تاریخ ارائه شده است؛ کشور‌های مختلف از معیار‌های متفاوتی برای تقسیم بندی افراد در طبقات متفاوت اجتماعی استفاده می‌کنند. در این پروژه از دو نوع رایج آن استفاده می‌کنیم:

طبقه: مرفه - متوسط به بالا - متوسط به پایین - فقیر

طبقه: مرفه - متوسط - فقیر

۳ علوم داده‌های اجتماعی

علوم داده‌های اجتماعی رشته‌ای است که علوم اجتماعی و علوم کامپیوتر را ترکیب می‌کند و در آن تجزیه و تحلیل داده‌ها به نظریه و تحلیل علمی اجتماعی مرتبط است.

۴ داده‌های پروژه

در این پروژه از دو مجموعه داده مجزا استفاده شده است. در قسمت اول از مجموعه داده موجود در سایت kaggle استفاده شده است و در قسمت دوم با تنظیم یک پرسشنامه سعی در جمع آوری اطلاعات کرده‌ام.

۱.۴ بخش اول

این مجموعه داده از یکی از شهرهای موجود در هند جمع آوری و در راستای ارزیابی آموزش آنلاین انجام شده است. با توجه به متغیرهایی که در این مجموعه داده موجود است؛ ما نیز توانستیم به هدف مسئله‌ی خود برسیم. وجود دو متغیر طبقه‌ی اجتماعی افراد (Economic status) و مدت زمانی که از شبکه‌های اجتماعی استفاده می‌کنند (Time spent on social media) به ما کمک کرد تا به هدف پروژه نزدیک شویم. این مجموعه داده شامل 1033 تا مشاهده و 23 تا متغیر است.

مجموعه داده‌ی این بخش را می‌توانید از طریق [لینک](#) زیر مشاهده کنید.

(<https://www.kaggle.com/datasets/sujaradha/online-education-system-review>)

۲.۴ بخش دوم

۱.۲.۴ طراحی پرسشنامه

همانطور که در بالا اشاره شد ، اطلاعات این بخش را با استفاده از پرسشنامه به دست آوردیم . لازم به ذکر است جامعه هدفی که در این قسمت در نظر گرفتیم ؛ دانشجویان مشغول به تحصیل دانشگاه تهران هستند . برای این کار ابتدا پرسشنامه را طراحی و سپس توزیع کردیم . برای بخش طراحی 3 فاکتور اصلی را در نظر گرفتیم :

□ سوالات عمومی مثل : سن ، جنسیت ، مقطع تحصیلی و

□ اطلاعات مربوط به شبکه های اجتماعی افراد

□ اطلاعات مربوط به طبقه اجتماعی افراد

در ادامه سعی کردیم بهترین پلتفرم را برای جمع آوری اطلاعات استفاده کنیم که در آخر پلتفرم پرسلاین (www.porsline.com) انتخاب شد . پرسشنامه طراحی شده را میتوانید از طریق [لینک](https://survey.porsline.ir/s/dVSLRye) زیر مشاهده کنید .

۲.۲.۴ جمع آوری اطلاعات

با توجه به جامعه ی هدفی ("دانشجویان مشغول به تحصیل دانشگاه تهران") که در نظر گرفتیم ؛ ابتدا سعی شد به واسطه ی آموزش هر دانشکده یا آموزش کل دانشگاه تهران به ایمیل دانشجویان دسترسی پیدا کرد و سپس به هر یک ایمیلی حاوی لینک پرسشنامه فرستاده شود اما با این درخواست موافقت نشد و در آخر از طریق توزیع پرسشنامه از طریق کانال های انجمن های علمی هر دانشکده به اطلاعات لازم دسترسی پیدا کردیم . این روند نزدیک به یک ماه به طول انجامید و در آخر از بین 229 نفر 165 نفر به پرسشنامه پاسخ دادند . (نرخ پاسخ دهی : تقریباً 71 درصد است.) در نهایت بعد از مرحله جمع آوری اطلاعات ، شروع به تمیز کردن اطلاعات کردیم.

۵ مراحل تحلیل داده ها

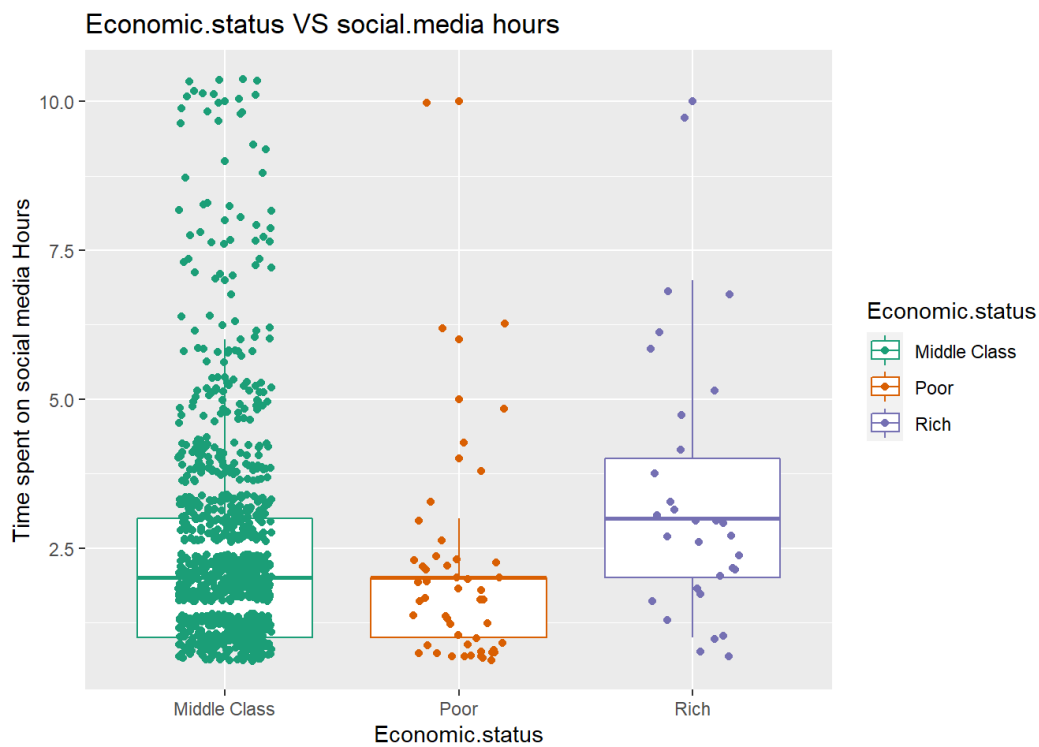
۱.۵ بخش اول

ابتدا EDA انجام دادیم و با ساختار داده ها آشنا شدیم . در این مرحله تک تک متغیر ها و مشاهدات را بررسی کردم . از نمودار های موجود در کتابخانه ggplot کمک گرفتم و هم به صورت تک متغیره^۱ و هم به صورت دو متغیره^۲ ، متغیر ها را بررسی کردم . سعی کردم تک

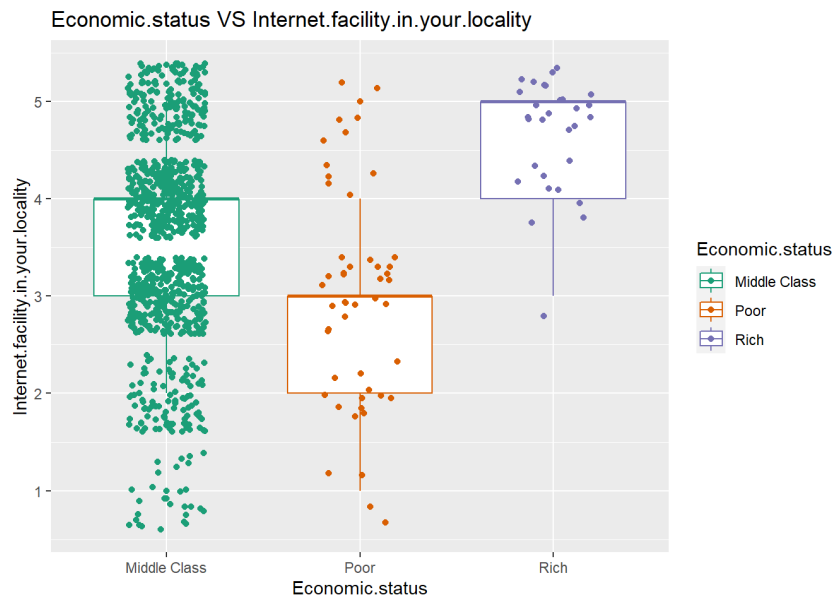
¹univariate

²bivariate

تک متغیرها را در مقابل متغیر Economic.status رسم و بررسی کنیم تا به هدف مسئله نزدیک شویم.
 با توجه به نمودار زیر میتوان مشاهده کرد مدت زمانی که در فضای مجازی سپری می شود در طبقات اجتماعی مختلف متفاوت است :

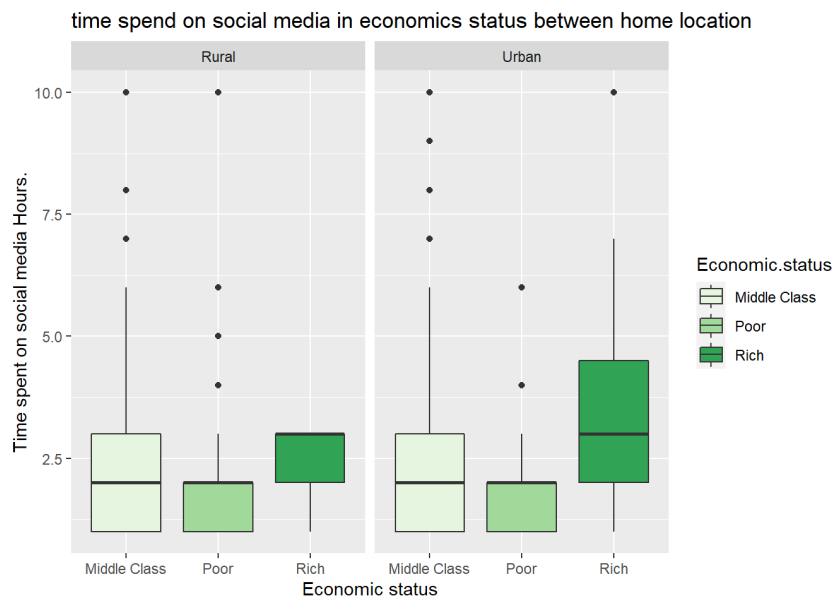


شکل ۱: نمودار طبقات اجتماعی در مقابل مدت زمان صرف شده در شبکه های اجتماعی
 با توجه به نمودار زیر میتوان مشاهده کرد دسترسی به اینترنت در طبقات اجتماعی مختلف متفاوت است :



شکل ۲: نمودار امکانات اینترنت محلی در طبقات اجتماعی مختلف

با توجه به نمودار زیر میتوان میزان زمانی که در شبکه های اجتماعی هر طبقات اجتماعی صرف می کند را در مناطق روستایی و شهری مشاهده کرد :



شکل ۳: مدت زمان صرف شده در شبکه های اجتماعی برای هر طبقه ی اجتماعی در دو منطقه روستایی و شهری

در مرحله ی بعد با توجه به اینکه 90 درصد اطلاعات در طبقه ی middle class حضور دارند ، لازم دیدیم که این طبقه را به دو طبقه ی Upper Middle و Lower Middle تقسیم کنیم . این تقسیم بندی توسط ساختن متغیری به نام Score انجام شد . برای ساختن متغیر Score ابتدا یک نمونه ی ۵۰ تایی از طبقه ی متوسط گرفتیم و در کنار طبقه ی فقیر و مرفه قرار دادیم . برای این مجموعه داده از مدل رگرسیون لوژستیک ترتیبی استفاده کردیم و با معیار قرار دادن پی-مقدار^۳ های هر یک از متغیر های مجموعه داده ؛ دانه دانه متغیر ها را حذف کردم تا به یک مدلی برسیم که تمامی متغیر هایش معنادار هستند . از ضرایب این متغیر ها استفاده کردم و یک متغیر به نام Score ساختم . در ادامه میتوانید این متغیر را مشاهده کنید :

$$\begin{aligned} \text{score} = & 1.788 * \text{Home.Location} + 1.541 * (\text{Level.of.Education} == 2) + \\ & 1.888 * (\text{Level.of.Education} == 3) - 0.189 * \text{Study.time..Hours.} - \\ & 1.038 * \text{Device.type.used.to.attend.classes} + \\ & 0.743 * \text{Internet.facility.in.your.locality} + \\ & 1.012 * \text{Are.you.involved.in.any.sports.} + \\ & 0.291 * \text{Time.spent.on.social.media..Hours.} - \\ & 0.982 * \text{Interested.in.Gaming.} + \\ & 0.729 * \text{Have.separate.room.for.studying.} \end{aligned}$$

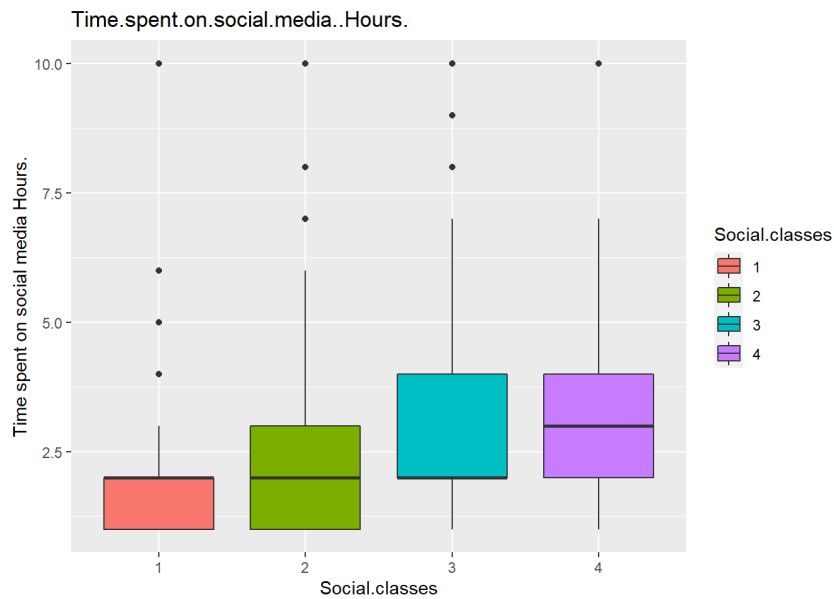
برای ارزشیابی این متغیر ، میانگین Score را مبنا قرار دادیم ؛ هر مشاهده ای که مقدار متغیر Score اش بیشتر از میانگین Score بود را به عنوان Upper Middle و هر مشاهده ای که مقدار متغیر Score اش کمتر از میانگین Score بود را به عنوان Lower Middle شناسایی کردم و تفاوت متغیر های دیگر را در برابر Upper Middle و Lower Middle مشاهده کردم . تا از کارا بودن این متغیر اطمینان حاصل کنیم . لازم به ذکر است که برای تقسیم بندی طبقه ی متوسط از الگوریتم های خوشه بندی^۴ ، KNN و naive bayes نیز استفاده کردم اما نتوانستند به خوبی داده های مسئله را خوشه بندی کنند.

بعد از این تقسیم بندی ، از متغیر Score برای کل مشاهدات موجود در middle class استفاده کردیم و متوجه شدیم که به خوبی طبقه ی متوسط را تقسیم کرده است . بعد از این مرحله متغیری به نام Social.classes تعریف کردم که دارای ۴ سطح است : فقیر - متوسط به پایین - متوسط به بالا - مرفه در ادامه سعی کردم با استفاده از نمودار باز هم متغیر های دیگر را در برابر متغیر Social.classes قرار دهم تا تفاوت متغیر های مجموعه داده در طبقات مختلف را واکاوی کنم .

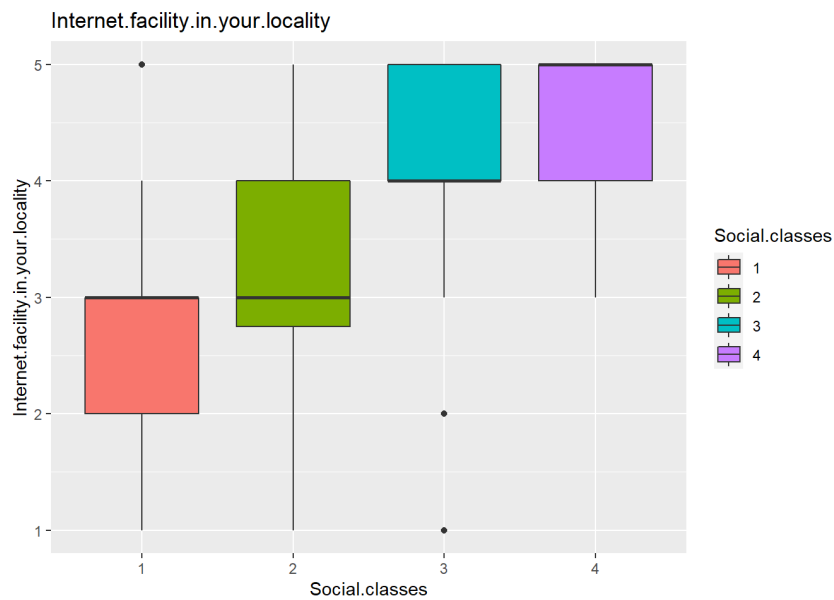
در نمودار زیر می توانید میزان زمان صرف شده در شبکه های اجتماعی برای هر ۴ طبقه را مشاهده کنید :

³p-value

⁴clustering



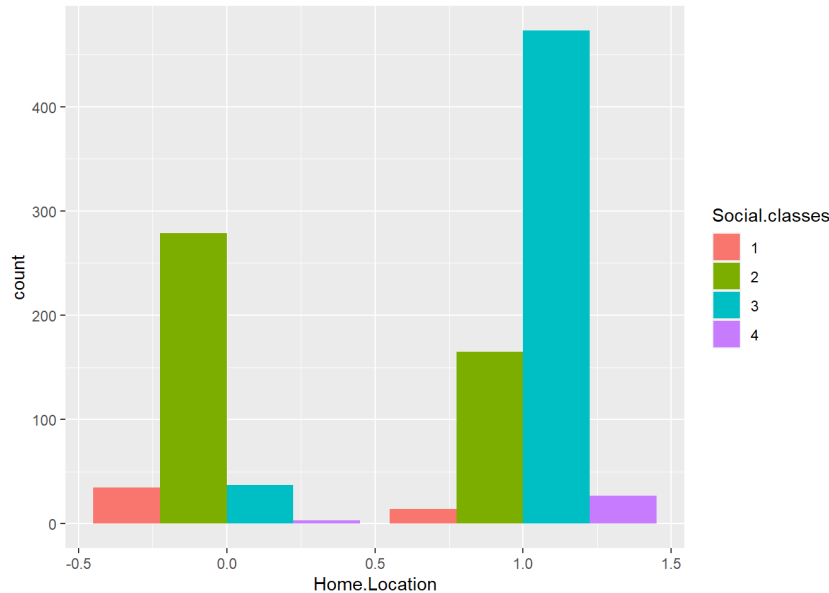
شکل ۴: نمودار میزان زمان صرف شده در شبکه های اجتماعی برای هر طبقه ی اجتماعی در نمودار زیر میتوانید میزان دسترسی به اینترنت را برای هر ۴ طبقه اجتماعی را مشاهده کنید :



شکل ۵: نمودار میزان دسترسی به اینترنت برای هر ۴ طبقه اجتماعی

با توجه به نمودار زیر میتوان وضعیت طبقات را در منطق شهری و روستایی مشاهده

کنید : همانطور که مشاهده می‌کنید در مناطق شهری نسبت افراد مرفه و متوسط رو به بالا بیشتر است و در مناطق روستایی نسبت افراد متوسط به پایین و فقیر بیشتر است .



شکل ۶: نمودار وضعیت طبقات اجتماعی در منطق شهری و روستایی

همچنین با بعضی از روش‌هایی که برای Feature Selection وجود دارد متغیرهای مهم در این مجموعه داده را مشخص کنیم . روش‌هایی که برای این کار استفاده کردیم عبارت‌اند از : Boruta و MARS(Multivariate Adaptive Regression Splines) - Random Forest .

Random Forest و Boruta هر دو از درخت‌های تصمیم‌گیری استفاده می‌کنند و مهم‌ترین متغیرهای مستقلی که به خوبی متغیر پاسخ را توضیح می‌دهند را جدا می‌کنند . الگوریتم MARS که توسط Friedman در سال ۱۹۹۱ به دست آمده است ، برای حل مسائل رگرسیون^۵ و طبقه‌بندی^۶ با متغیر پاسخ چندگانه ارائه شده است . متغیرهای مهم انتخاب شده از این مجموعه داده عبارت‌اند از :

- Home.Location
- Internet.facility.in.your.locality
- Device.type.used.to.attend.classes
- Study.time..Hours.

⁵regression

⁶classification

- Have.separate.room.for.studying.
- Level.of.Education
- Time.spent.on.social.media..Hours.

در انتها نیز سعی کردم با استفاده از الگوریتمی که تفسیر پذیری بالایی (مثل رگرسیون لوژیستیک ترتیبی) دارد، به هدف پروژه برسم .
از مدل رگرسیون لوژیستیک ترتیبی استفاده کردم و با استفاده از تمامی متغیر هایی که در مجموعه داده حضور دارند مدل سازی را انجام دادم :
با استفاده از تابع polr موجود در کتابخانه MASS مدل رگرسیون لوژیستیک ترتیبی را اجرا کردم و با دستور summary به ضرایب زیر رسیدم :

Table 1

	Value	std. Error	t value	p value
Gender1	0.32	0.18	1.80	0.07
Home.Locaiton1	2.91	0.22	13.16	0.00
Level.of.Education2	2.15	0.30	7.17	0.00
Level.of.Education3	2.22	0.40	5.62	0.00
Age.Years.	0.04	0.03	1.39	0.16
Number.of.Subjects	-0.01	0.03	-0.46	0.65
Device.type.used.to.attend.classes2	-1.28	0.53	-2.43	0.02
Device.type.used.to.attend.classes3	-3.00	0.55	-5.45	0.00
Family.size	-0.01	0.7	-0.16	0.87
Internet.Facility.in.your.locality2	0.30	0.49	0.63	0.53
Internet.Facility.in.your.locality3	1.11	0.46	2.43	0.02
Internet.Facility.in.your.locality4	2.44	0.47	5.17	0.00
Internet.Facility.in.your.locality5	3.56	0.51	6.92	0.00
Do.elderly.people.monitor.you.1	0.7	0.17	0.39	0.70
Study.time..Hours	-0.27	0.04	-6.52	0.00
Sleep.time..Hours	-0.05	0.06	0.82	0.41
Time.spent.on.social.media..Hours	0.36	0.05	7.69	0.00
Interested.in.Gaming.1	-1.31	0.18	-7.34	0.00
Have.seprate.room.for.studying.1	1.17	0.18	6.50	0.00
Engaged.in.group.studies.1	0.20	0.17	1.18	0.24
Clearing.doubts.with.faculties.in.online.mode	-0.03	0.09	-0.36	0.72
Interested.in.2	0.14	0.19	0.75	0.45
Interested.in.3	-0.03	0.23	-0.15	0.88
Performance.in.online	0.13	0.05	2.41	0.02
Your.level.of.satisfaction.in.online.Education2	-0.09	0.23	-0.39	0.70
Your.level.of.satisfaction.in.online.Education3	0.13	0.29	0.43	0.67
1 2	-0.71	1.10	-0.64	0.52
2 3	5.17	1.13	4.59	0.00
3 4	11.45	1.19	9.59	0.00

و در آخر با استفاده از ضرایب به دست آمده نتیجه گیری کردیم .

۲.۵ بخش دوم

اولین مرحله برای تحلیل داده‌های پرسشنامه ، تمیز کردن اطلاعات جمع‌آوری شده است .فایل ارسال شده توسط سایت پرسلاین قابل اجرا در R نبود و باید یک فایل .csv خوانا و مرتب را آماده می‌کردیم . بعد از این مرحله EDA را شروع کردم . بعد تحلیل اولیه اطلاعات

جمع آوری شده ، هر کدام از متغیر های چند گزینه ای^۷ را به چندین متغیر دودویی تبدیل کردم تا مدل سازی و تحلیل آن راحت تر باشد .
در ادامه وضعیت تک تک متغیر های جمع آوری شده را مشاهده می کنید :

جدول ۲: جنسیت

گزینه	فراوانی پاسخ	درصد فراوانی
زن	91	55.2
مرد	72	43.6
دیگر	2	1.2
جمع	165	100

جدول ۳: گروه های سنی

گزینه	فراوانی پاسخ	درصد فراوانی
> 15	0	0
[15, 24)	148	89.7
[25, 44)	16	9.7
[45, 64)	1	0.6
> 65	0	0
جمع	165	100

جدول ۴: وضعیت تاهل

گزینه	فراوانی پاسخ	درصد فراوانی
مجرد	152	92.1
متاهل	13	7.9
جمع	165	100

⁷multiple choice

جدول ۵: تحصیلات

گزینه	فراوانی پاسخ	درصد فراوانی
سیکل	0	0
دیپلم	51	30.9
لیسانس	96	58.2
فوق لیسانس	14	8.5
دکتر	4	2.4
جمع	165	100

جدول ۶: تعداد اعضای خانواده

گزینه	فراوانی پاسخ	درصد فراوانی
خانواده سه نفری	32	19.4
خانواده چهار نفری	88	53.3
خانواده پنج نفری	38	23
خانواده شش نفری	5	3
بیشتر	2	1.2
جمع	165	100

جدول ۷: درآمد

گزینه	فراوانی پاسخ	درصد فراوانی
از صفر تا ۱۰ میلیون تومان	152	93.3
از ۱۰ تا ۲۰ میلیون تومان	10	6.1
از ۲۰ تا ۳۰ میلیون تومان	1	0.6
از ۳۰ تا ۴۰ میلیون تومان	0	0
بیش از ۴۰ میلیون تومان	0	0
جمع	163	100

جدول ۸: وضعیت محل سکونت

گزینه	فراوانی پاسخ	درصد فراوانی
مالک	122	73.9
مستاجر	43	26.1
جمع	165	100

جدول ۹: شغل

درصد فراوانی	فراوانی پاسخ	گزینه
86.1	142	در حال تحصیل
16.4	27	رشته‌ی امور آموزشی و فرهنگی و هنری شامل کلیه‌ی مشاغل معلمی، تدریس، تولید ادبی و هنری (روابط عمومی، تبلیغات، مطبوعات و رسانه‌ها)
5.5	9	رشته‌ی اداری و مالی شامل کلیه‌ی مشاغل حسابداری، حسابرسی) ،اداری، مالی، بیمه، بانکداری، بورس (آمار اقتصادی و ... مدیریت‌ها و برنامه‌ریزی‌های مربوط)
7.3	12	رشته‌ی امور اجتماعی شامل کلیه‌ی مشاغل پژوهش‌های میدانی روان‌شناسی) و جامعه‌شناسی و علوم سیاسی و روابط بین‌الملل و (... تحلیل‌گری و آمار اجتماعی و مشاوره و علوم تربیتی و
2.4	4	رشته‌ی بهداشتی و درمانی شامل کلیه‌ی مشاغل پزشکی و) دندان پزشکی و پیراپزشکی و (... پرستاری و مامایی و داروسازی و
3	5	رشته‌ی خدمات شامل کلیه‌ی مشاغل دلالی و) واسطه‌گری و فروشنده‌گی و بازرگانی و توزیع کالا و خدمات و ورزش و توریسم و سرگرمی و تغذیه و ساخت و (... ساز بنا و حرفه‌های اصلاح و تعمیر و
1.8	3	رشته‌ی کشاورزی و محیط زیست) شامل مشاغل کشت و داشت و برداشت محصول و) پژوهش‌های مربوط به هر مرحله پرورش انواع جانوران سودمند، دامپزشکی، گیاه‌شناسی، (... مطالعات جنگل و آب و هوا، بیابان‌زدایی و
15.2	25	رشته‌ی فنی و مهندسی شامل کلیه‌ی مشاغل سخت‌افزار و نرم‌افزار صنعت و تکنولوژی) اختراع و تولید صنعتی، کارگری صنعت (کارگاه‌ها و آزمایشگاه‌ها و... برنامه‌ریزی‌ها و مدیریت‌های مربوط
2.4	4	رشته‌ی فرآوری داده‌ها شامل کلیه‌ی مشاغل مربوط به اطلاع‌رسانی) ،کتاب‌داری، تولید نرم‌افزارهای اطلاعاتی ،گسترش اطلاعات، شبکه‌ی اینترنت ،انتقال و تلفیق اطلاعات (ارزیابی داده‌ها، حفاظت داده‌ها
100	165	جمع

جدول ۱۰: وضعیت اجاره مستاجرین

گزینه	فراوانی پاسخ	درصد فراوانی
از ۰ تا ۵ میلیون	32	74.4
از ۵ تا ۱۰ میلیون	6	14
از ۱۰ تا ۱۵ میلیون	2	4.7
از ۱۵ تا ۲۰ میلیون	1	2.3
از ۲۰ تا ۲۵ میلیون	1	2.3
از ۲۵ تا ۳۰ میلیون	1	2.3
از ۳۰ تا ۴۰ میلیون	0	0
بیش از ۴۰ میلیون	0	0
جمع	43	100

جدول ۱۱: شبکه‌های اجتماعی خارجی

گزینه	فراوانی پاسخ	درصد فراوانی
Instagram	133	80.6
Twitter	77	46.7
Facebook	10	6.1
Linkedin	61	37
Youtube	127	77
Pintrest	64	38.8
Clubhouse	15	9.1
Tiktok	4	2.4
همه‌ی موارد	0	0
هیچ‌کدام	6	3.6
جمع	165	100

جدول ۱۲: شبکه‌های اجتماعی داخلی

گزینه	فراوانی پاسخ	درصد فراوانی
آپارات	58	35.2
روبوکا	4	2.4
کمدآ	4	2.4
هورسا	0	0
روبینو	0	0
گپ	1	0.6
دیگر شبکه‌های اجتماعی داخل	10	6.1
هیچ‌کدام	103	62.4
جمع	165	100

جدول ۱۳: دلایل استفاده از شبکه‌های اجتماعی

گزینه	فراوانی پاسخ	درصد فراوانی
کسب درآمد	15	9.1
پر کردن اوقات فراغت	137	83
ارتباط نزدیک با خانواده و اطرافیان	96	58.2
به روز بودن، دریافت آخرین اخبار و در جریان قرار گرفتن اتفاقات روز دنیا	120	72.7
در اختیار داشتن شبکه گسترده و جامع با دیگران	93	56.4
دلایل دیگر	33	20
جمع	165	100

جدول ۱۴: میزان خرید با کمک شبکه‌های اجتماعی

گزینه	فراوانی پاسخ	درصد فراوانی
از ۰ تا ۲۵ درصد	121	73.7
از ۲۵ تا ۵۰ درصد	23	13.9
از ۵۰ تا ۷۵ درصد	17	10.3
از ۷۵ تا ۱۰۰ درصد	4	2.4
جمع	165	100

جدول ۱۵: مدت زمان استفاده از شبکه‌های اجتماعی (در یک روز)

گزینه	فراوانی پاسخ	درصد فراوانی
یک الی دو ساعت در روز	43	26.1
دو الی سه ساعت در روز	47	28.5
سه الی چهار ساعت در روز	49	29.7
چهار الی پنج ساعت در روز	14	8.5
پنج الی شش ساعت در روز	7	4.2
بیش از ۶ ساعت	5	3
جمع	165	100

جدول ۱۶: آیا بلافاصله بعد از بیدار شدن از خواب شبکه‌های اجتماعی خود را چک می‌کنید؟

گزینه	فراوانی پاسخ	درصد فراوانی
همیشه	38	23
اکثر مواقع	58	35.2
گاهی اوقات	45	27.3
خیلی کم	19	11.5
هرگز	5	3
جمع	165	100

جدول ۱۷: محبوبیت موضوعات در شبکه‌های اجتماعی

گزینه	فراوانی پاسخ	درصد فراوانی
رسانه‌ها	76	46.1
مد و فشن و زیبایی	33	20
موسیقی و هنر	91	55.2
ورزشی و تناسب اندام	41	24.8
تکنولوژی	54	32.7
انگیزشی	31	18.8
سرگرمی	107	64.8
آموزشی	109	66.1
دیگر موضوعات	44	26.7
جمع	165	100

با گروه بندی مشاهدات سعی کردم رفتار دانشجویان را بررسی کنم . به عنوان مثال دانشجویانی که درآمد بین ۰ تا ۱۰ میلیون تومان دارند در شبکه‌های اجتماعی چه رفتاری را دارند یا افرادی که درآمد بین ۱۰ تا ۲۰ میلیون تومان دارند در شبکه‌های اجتماعی چه

رفتاری دارند و ... در ادامه گزارشی از بعضی از سوالات کلیدی را میبینید :

سوال یک (دانشجویانی که درآمد بین ۰ تا ۱۰ میلیون تومان دارند در شبکه های اجتماعی چه رفتاری را دارند ؟
شبکه های اجتماعی (خارجی) این گروه به ترتیب زیر است :

- instagram (79.3 %)
- youtube (76 %)
- twitter (45.3 %)
- pinterest (40%)
- linkedin (35.3%)
- clubhouse (9%)
- facebook (5.3%)
- tiktok (2%)

شبکه های اجتماعی (ایرانی) این گروه به ترتیب زیر است :

- None (63.3 %)
- Aparat (37.7 %)
- Other (6 %)
- Robika (2.7 %)
- Komoda (2.7 %)
- Gap (6 %)

وضعیت خرید اینترنتی :
72.7 درصد افراد بین ۰ تا ۲۵ درصد ، 14 درصد افراد بین ۲۵ تا ۵۰ ، 10.7 درصد افراد بین ۵۰ تا ۷۵ درصد و 2.7 درصد افراد بین ۷۵ تا ۱۰۰ درصد خرید های خود را به کمک شبکه های اجتماعی انجام میدهند .

مدت زمان استفاده از شبکه های اجتماعی (در یک روز) این گروه به ترتیب زیر است :

- [3,4) hour (32 %)
- [1,2) hour (28 %)

- [2,3) hour (25.3 %)
- [4,5) hour (8 %)
- [5,6) hour (3.3 %)
- >6 hour (3.3 %)

سوال دو (افرادی که درآمد بین ۱۰ تا ۲۰ میلیون تومان دارند در شبکه‌های اجتماعی چه رفتاری دارند؟
شبکه‌های اجتماعی (خارجی) این گروه به ترتیب زیر است :

- instagram (100 %)
- youtube (90 %)
- twitter (50 %)
- pinterest (30 %)
- linkedin (60 %)
- clubhouse (10 %)
- facebook (20 %)
- tiktok (10 %)

شبکه‌های اجتماعی (ایرانی) این گروه به ترتیب زیر است :

- None (60 %)
- Aparat (30 %)
- Other (10 %)
- Robika (0 %)
- Komoda (0 %)
- Gap (0 %)

وضعیت خرید اینترنتی :
80 درصد افراد بین ۰ تا ۲۵ درصد ، 20 درصد افراد بین ۲۵ تا ۵۰ درصد خرید های خود را
به کمک شبکه های اجتماعی انجام میدهند .

مدت زمان استفاده از شبکه‌های اجتماعی (در یک روز) این گروه به ترتیب زیر است :

- [2,3) hour (70 %)
- [3,4) hour (10 %)
- [4,4) hour (10 %)
- [5,6) hour (10 %)
- [1,2) hour (0 %)
- >6 hour (0 %)

سوال سه) افرادی که درآمد بین ۲۰ تا ۳۰ میلیون دارند در شبکه‌های اجتماعی چه رفتاری دارند؟ (توجه: با توجه به اطلاعاتی که جمع آوری کردیم تنها یک نفر در این دسته قرار می‌گیرد .)

شبکه‌های اجتماعی (خارجی) این فرد به صورت زیر است :

- instagram
- youtube
- twitter
- linkedin

شبکه‌های اجتماعی (ایرانی) این فرد به صورت زیر است :

- None

وضعیت خرید اینترنتی :

بین ۰ تا ۲۵ درصد خرید های این فرد به کمک شبکه های مجازی صورت می‌گیرد .

مدت زمان استفاده از شبکه‌های اجتماعی (در یک روز) این فرد به صورت زیر است :

- [1,2) hour

سوال ۴) بررسی افرادی که از شبکه‌های اجتماعی کسب درآمد میکنند :

شبکه های اجتماعی (خارجی) این گروه به ترتیب زیر است :

- instagram (78.6 %)
- youtube (78.6 %)
- pinterest (57.1 %)

- linkedin (50 %)
- twitter (42.9 %)
- clubhouse (14.3 %)
- facebook (0 %)
- tiktok (0 %)

شبکه‌های اجتماعی (ایرانی) این گروه به ترتیب زیر است :

- Aparat (42.9 %)
- None (50 %)
- Other (7.1 %)
- Robika (7.1 %)
- Komoda (7.1 %)
- Gap (0 %)

وضعیت خرید اینترنتی :

71.4 درصد افراد بین ۰ تا ۲۵ درصد ، 14.3 درصد افراد بین ۲۵ تا ۵۰ ، 7.1 درصد افراد بین ۵۰ تا ۷۵ درصد ، 7.1 درصد افراد بین ۷۵ تا ۱۰۰ درصد خرید های خود را به کمک شبکه های اجتماعی انجام میدهند .

مدت زمان استفاده از شبکه‌های اجتماعی (در یک روز) این گروه به ترتیب زیر است :

- [1,2) hour (35.7 %)
- [2,3) hour (21.4 %)
- [3,4) hour (21.4 %)
- [4,5) hour (14.3 %)
- [5,6) hour (7.1 %)
- >6 hour (0 %)

مسئله‌ی مهمی که بعد از تحلیل داده‌ی جمع‌آوری شده با آن مواجه می‌شویم، تعیین متغیر پاسخ برای مدل‌سازی است. به همین دلیل سعی کردم مشاهدات را به سه گروه اصلی طبقه‌ی فقیر، متوسط و مرفه تقسیم کنیم. برای انجام این دسته‌بندی از اطلاعات اقتصادی که در مجموعه داده حضور دارند مثل درآمد و وضعیت مسکن افراد استفاده کردیم. (لازم به ذکر است برای سادگی مسئله دو نفر از افرادی را که به سوال میزان درآمد پاسخ ندادند؛ حذف شدند). افرادی که درآمد بین ۰ تا ۱۰ میلیون تومان دارند و وضعیت مسکنشان اجاره هست در گروه فقیر قرار گرفتند، افرادی که درآمد بین ۰ تا ۱۰ میلیون دارند و صاحب خانه هستند را در گروه متوسط و در آخر افرادی که درآمد بین ۱۰ تا ۲۰ میلیون تومان یا ۲۰ تا ۳۰ میلیون تومان دارند در گروه مرفه قرار گرفته‌اند. با توجه به این تقسیم‌بندی 39 نفر در طبقه‌ی فقیر، 111 نفر در طبقه‌ی متوسط و 11 نفر در طبقه‌ی مرفه قرار می‌گیرند. بعد از مشخص کردن متغیر پاسخ، برای تعیین متغیرهای مهم در این مجموعه داده از الگوریتم‌های مشابه بخش قبل استفاده کردم که عبارت‌اند از:

- lodging
- income
- reason 3 (connection)
- age.group
- family.size
- aparat
- NONE (IranSocialMedia)
- Job8 (Engineering)
- Job3 (Administrative and financial)
- other (IranSocialMedia)
- Job6 (Services)
- Tenant

برای مدل‌سازی با توجه به متغیر پاسخ از مدل رگرسیون لوژیستیک ترتیبی^۸ استفاده کردیم. در این مجموعه داده بعضی از متغیرها بودند که همبستگی زیادی با یکدیگر داشتند، برای داشتن مدل‌سازی بهتر و قابل اعتمادتر این متغیرها را از مدل حذف کردیم و در آخر با استفاده از تابع polr مدل‌سازی را انجام دادیم. برای مشاهده‌ی ضرایب به پی‌دی‌اف با نام kable که در لینک زیر موجود است مراجعه کنید.

⁸Ordinal logistic regression

لینک :

(<https://drive.google.com/drive/folders/1h2b-dgZtxqa2dYi4ECe6MA3f-dHPduL0?usp=sharing>)

در آخر از این ضرایب استفاده میکنیم و نتیجه گیری را انجام می دهیم .

۶ نتیجه گیری

۱.۶ بخش اول

در نهایت با استفاده از مدل رگرسیون لوژیستیک ترتیبی و ضرایبی که در قسمت تحلیل داده‌ی این بخش آورده شده است میتوانیم تاثیر متغیر های پیشگوی (که مدت زمان استفاده از شبکه های اجتماعی نیز جزو آن است) مسئله را روی طبقات اجتماعی social.classes با استفاده از مدل های زیر ببینیم . در مدل هایی که آورده شده ، exp ضرایب را در نظر گرفتیم تا بتوانم از $odds(P(Y \leq j)) = \frac{P(Y \leq j)}{P(Y > j)}$ استفاده کنم .

$$\begin{aligned} odds(P(Y < 1)) = & 0.49 + 1.38(\text{Gender1}) + 18.45(\text{Home.Location1}) + \\ & 8.56(\text{Level.of.Education2}) + 9.23(\text{Level.of.Education3}) + \\ & 1.04(\text{Age.Years.}) + 0.99(\text{Number.of.Subjects}) + \\ & 0.28(\text{Device.type.used.to.attend.classes2}) + \\ & 0.05(\text{Device.type.used.to.attend.classes3}) + \\ & 0.99(\text{Family.size}) + 1.36(\text{Internet.facility.in.your.locality2}) \\ & + 3.04(\text{Internet.facility.in.your.locality3}) + \\ & 11.52(\text{Internet.facility.in.your.locality4}) + \\ & 35.07(\text{Internet.facility.in.your.locality5}) + \\ & 1.07(\text{Do.elderly.people.monitor.you.1}) + \\ & 0.77(\text{Study.time..Hours.}) + 0.95(\text{Sleep.time..Hours.}) + \\ & 1.44(\text{Time.spent.on.social.media..Hours.}) + \\ & 1.22(\text{Engaged.in.group.studies.1}) + 0.27(\text{Interested.in.Gaming.1}) + \\ & 3.21(\text{Have.separate.room.for.studying.1}) + \\ & 0.97(\text{Clearing.doubts.with.faculties.in.online.mode}) + \\ & 1.15(\text{Interested.in.2}) + 0.97(\text{Interested.in.3}) + \\ & 1.14(\text{Performance.in.online}) + \\ & 0.92(\text{Your.level.of.satisfaction.in.Online.Education2}) + \\ & 1.13(\text{Your.level.of.satisfaction.in.Online.Education3}) \end{aligned}$$

$$\begin{aligned} \text{odds}(P(Y < 2)) = & 176.64 + 1.38(\text{Gender}1) + 18.45(\text{Home.Location}1) + \\ & 8.56(\text{Level.of.Education}2) + 9.23(\text{Level.of.Education}3) + \\ & 1.04(\text{Age.Years.}) + 0.99(\text{Number.of.Subjects}) + \\ & 0.28(\text{Device.type.used.to.attend.classes}2) + \\ & 0.05(\text{Device.type.used.to.attend.classes}3) + \\ & 0.99(\text{Family.size}) + 1.36(\text{Internet.facility.in.your.locality}2) \\ & + 3.04(\text{Internet.facility.in.your.locality}3) + \\ & 11.52(\text{Internet.facility.in.your.locality}4) + \\ & 35.07(\text{Internet.facility.in.your.locality}5) + \\ & 1.07(\text{Do.elderly.people.monitor.you.1}) + \\ & 0.77(\text{Study.time..Hours.}) + 0.95(\text{Sleep.time..Hours.}) + \\ & 1.44(\text{Time.spent.on.social.media..Hours.}) + \\ & 1.22(\text{Engaged.in.group.studies.1}) + 0.27(\text{Interested.in.Gaming.1}) + \\ & 3.21(\text{Have.separate.room.for.studying.1}) + \\ & 0.97(\text{Clearing.doubts.with.faculties.in.online.mode}) + \\ & 1.15(\text{Interested.in.2}) + 0.97(\text{Interested.in.3}) + \\ & 1.14(\text{Performance.in.online}) + \\ & 0.92(\text{Your.level.of.satisfaction.in.Online.Education}2) + \\ & 1.13(\text{Your.level.of.satisfaction.in.Online.Education}3) \end{aligned}$$

$$\begin{aligned} \text{odds}(P(Y < 3)) = & 94206.88 + 1.38(\text{Gender}1) + 18.45(\text{Home.Location}1) + \\ & 8.56(\text{Level.of.Education}2) + 9.23(\text{Level.of.Education}3) + \\ & 1.04(\text{Age.Years.}) + 0.99(\text{Number.of.Subjects}) + \\ & 0.28(\text{Device.type.used.to.attend.classes}2) + \\ & 0.05(\text{Device.type.used.to.attend.classes}3) + \\ & 0.99(\text{Family.size}) + 1.36(\text{Internet.facility.in.your.locality}2) \\ & + 3.04(\text{Internet.facility.in.your.locality}3) + \\ & 11.52(\text{Internet.facility.in.your.locality}4) + \\ & 35.07(\text{Internet.facility.in.your.locality}5) + \\ & 1.07(\text{Do.elderly.people.monitor.you.1}) + \\ & 0.77(\text{Study.time..Hours.}) + 0.95(\text{Sleep.time..Hours.}) + \\ & 1.44(\text{Time.spent.on.social.media..Hours.}) + \\ & 1.22(\text{Engaged.in.group.studies.1}) + 0.27(\text{Interested.in.Gaming.1}) + \\ & 3.21(\text{Have.separate.room.for.studying.1}) + \\ & 0.97(\text{Clearing.doubts.with.faculties.in.online.mode}) + \\ & 1.15(\text{Interested.in.2}) + 0.97(\text{Interested.in.3}) + \\ & 1.14(\text{Performance.in.online}) + \\ & 0.92(\text{Your.level.of.satisfaction.in.Online.Education}2) + \\ & 1.13(\text{Your.level.of.satisfaction.in.Online.Education}3) \end{aligned}$$

Time.spent.on.social.media..Hours. با توجه به مطالب بالا، به ازای یک واحد افزایش $\text{odds}(P(Y < j))$ ، $j = 1, 2, 3$ ، برابر می شود .

همچنین با استفاده از نمودارهایی که در بخش تحلیل داده آورده شده است می توان دید که مدت زمان استفاده شده در قشر مرفه بیشتر از سه گروه دیگر است و سه گروه متوسط به بالا و متوسط به پایین و فقیر تفاوت چندانی دیده نمیشه . با این همه در نمودار دسترسی به اینترنت در تمامی گروه ها یکسان نبودند و تفاوت بیشتری داشتند .

۲.۶ بخش دوم

ابتدا مانند بخش قبل عمل میکنیم و مدل های زیر را به دست آمده است : در این قسمت $\text{logit}(P(Y < 1))$ را نیز معرفی میکنیم :

$$\begin{aligned} \text{logit}(P(Y < 1)) = & -49.13 + -0.32(\text{sex2}) + 0.63(\text{sex3}) - 2.12(\text{marriage2}) - \\ & 0.90(\text{family.size}) + 0.87(\text{Job1}) + 0.90(\text{Job2}) + 1.49(\text{Job3}) - 4.48(\text{Job4}) - \\ & 3.72(\text{Job5}) - 3.47(\text{Job6}) + 12.77(\text{Job7}) + 2.99(\text{Job8}) - 0.07(\text{Job9}) - \\ & 46.22(\text{income1}) + 12.97(\text{income2}) - 0.03(\text{instagram}) - 1.65(\text{twitter}) - \\ & 2.35(\text{facebook}) - 0.12(\text{linkdin}) - 1.93(\text{youtube}) - 1.13(\text{pintrest}) + \\ & 0.61(\text{clubhouse}) - 0.36(\text{tiktok}) - 4.00(\text{none}) - 4.06(\text{robika}) + 1.79(\text{komoda}) + \\ & 2.71(\text{Gap}) + 1.37(\text{other}) + 0.22(\text{NONE}) - 2.52(\text{reason1}) + 1.41(\text{reason2}) + \\ & 0.41(\text{reason3}) + 0.38(\text{reason4}) - 0.08(\text{reason5}) + 1.33(\text{reason6}) + \\ & 1.46(\text{shopping2}) - 1.05(\text{shopping3}) - 0.63(\text{SocialmediaTime2}) + \\ & 0.17(\text{SocialmediaTime3}) - 0.99(\text{SocialmediaTime4}) + 9.54(\text{SocialmediaTime5}) - \\ & 3.99(\text{SocialmediaTime6}) + 1.11(\text{checkingSocial2}) + 2.02(\text{checkingSocial3}) - \\ & 1.04(\text{checkingSocial4}) + 3.50(\text{checkingSocial5}) + 1.54(\text{Subject1}) + \\ & 0.34(\text{Subject2}) + 0.16(\text{Subject3}) - 1.75(\text{Subject4}) + 0.99(\text{Subject5}) - \\ & 0.44(\text{Subject6}) + 1.85(\text{Subject7}) - 1.64(\text{Subject8}) + 0.02(\text{Subject9}) \end{aligned}$$

$$\begin{aligned} \text{logit}(P(Y < 2)) = & -20.4766 + -0.32(\text{sex2}) + 0.63(\text{sex3}) - 2.12(\text{marriage2}) - \\ & 0.90(\text{family.size}) + 0.87(\text{Job1}) + 0.90(\text{Job2}) + 1.49(\text{Job3}) - 4.48(\text{Job4}) - \\ & 3.72(\text{Job5}) - 3.47(\text{Job6}) + 12.77(\text{Job7}) + 2.99(\text{Job8}) - 0.07(\text{Job9}) - \\ & 46.22(\text{income1}) + 12.97(\text{income2}) - 0.03(\text{instagram}) - 1.65(\text{twitter}) - \\ & 2.35(\text{facebook}) - 0.12(\text{linkdin}) - 1.93(\text{youtube}) - 1.13(\text{pintrest}) + \\ & 0.61(\text{clubhouse}) - 0.36(\text{tiktok}) - 4.00(\text{none}) - 4.06(\text{robika}) + 1.79(\text{komoda}) + \\ & 2.71(\text{Gap}) + 1.37(\text{other}) + 0.22(\text{NONE}) - 2.52(\text{reason1}) + 1.41(\text{reason2}) + \\ & 0.41(\text{reason3}) + 0.38(\text{reason4}) - 0.08(\text{reason5}) + 1.33(\text{reason6}) + \\ & 1.46(\text{shopping2}) - 1.05(\text{shopping3}) - 0.63(\text{SocialmediaTime2}) + \\ & 0.17(\text{SocialmediaTime3}) - 0.99(\text{SocialmediaTime4}) + 9.54(\text{SocialmediaTime5}) - \\ & 3.99(\text{SocialmediaTime6}) + 1.11(\text{checkingSocial2}) + 2.02(\text{checkingSocial3}) - \\ & 1.04(\text{checkingSocial4}) + 3.50(\text{checkingSocial5}) + 1.54(\text{Subject1}) + \\ & 0.34(\text{Subject2}) + 0.16(\text{Subject3}) - 1.75(\text{Subject4}) + 0.99(\text{Subject5}) - \\ & 0.44(\text{Subject6}) + 1.85(\text{Subject7}) - 1.64(\text{Subject8}) + 0.02(\text{Subject9}) \end{aligned}$$

در مدل بعدی از \exp ضرایب را در نظر گرفتیم تا بتوانیم از $\frac{P(Y \leq j)}{P(Y > j)}$ استفاده کنیم . در این بخش $\text{odds}(P(Y \leq 1)) = \text{odds}(P(Y \leq 2))$ است . چون \exp عرض از مبدا هر دو مدل صفر شده است .

$$\begin{aligned} \text{odds}(P(Y < 1)) = & 0.00 + 0.73(\text{sex2}) + 1.88(\text{sex3}) + 0.12(\text{marriage2}) + \\ & 0.40(\text{family.size}) + 2.39(\text{Job1}) + 2.46(\text{Job2}) + 4.42(\text{Job3}) + 0.01(\text{Job4}) + \\ & 0.02(\text{Job5}) + 0.03(\text{Job6}) + 349823.11(\text{Job7}) + 19.80(\text{Job8}) + 0.93(\text{Job9}) + \\ & 427338.61(\text{income2}) + 0.98(\text{instagram}) + 0.19(\text{twitter}) + 0.10(\text{facebook}) + \\ & 0.88(\text{linkdin}) + 0.14(\text{youtube}) + 0.32(\text{pintrest}) + 1.84(\text{clubhouse}) + \\ & 0.70(\text{tiktok}) + 0.02(\text{none}) + 0.02(\text{robika}) + 6.01(\text{komoda}) + 14.99(\text{Gap}) + \\ & 3.95(\text{other}) + 1.24(\text{NONE}) + 0.08(\text{reason1}) + 4.10(\text{reason2}) + 1.51(\text{reason3}) + \\ & 1.46(\text{reason4}) + 0.92(\text{reason5}) + 3.77(\text{reason6}) + 4.32(\text{shopping2}) + \\ & 0.35(\text{shopping3}) + 0.03(\text{shopping4}) + 0.53(\text{SocialmediaTime2}) + \\ & 1.19(\text{SocialmediaTime3}) + 0.37(\text{SocialmediaTime4}) + 13962.77(\text{SocialmediaTime5}) + \\ & 0.02(\text{SocialmediaTime6}) + 3.03(\text{checkingSocial2}) + 7.55(\text{checkingSocial3}) + \\ & 0.35(\text{checkingSocial4}) + 33.16(\text{checkingSocial5}) + 4.65(\text{Subject1}) + \\ & 1.41(\text{Subject2}) + 1.17(\text{Subject3}) + 0.17(\text{Subject4}) + 2.68(\text{Subject5}) + \\ & 0.64(\text{Subject6}) + 6.34(\text{Subject7}) + 0.19(\text{Subject8}) + 1.02(\text{Subject9}) \end{aligned}$$

همچنین میتوان گفت : در طبقه متوسط و مرفه، موضوعات محبوب در شبکه‌های اجتماعی سرگرمی ، آموزشی و موسیقی و هنر است و بسیار کم از شبکه‌های اجتماعی برای خرید هایشان کمک میگیرند. دلایل استفاده از شبکه های اجتماعی در این طبقات پرکردن اوقات فراغت ، به روز بودن است و علاقه ی چندانی به کسب درآمد از طریق فضای مجازی ندارند . تنها تفاوت این دو طبقه میزان وابستگی شان به شبکه های اجتماعی است ؛ طبقه متوسط وابستگی نسبتا زیادی به شبکه های اجتماعی دارند اما در طبقه مرفه وابستگی کمی وجود دارد .

در طبقه‌ی فقیر موضوعات محبوب مشابه دو طبقه ی دیگر است ، وابستگی کمی به شبکه های اجتماعی دارند و علاقه‌ای به خرید اینترنتی ندارند ، دلایل استفاده از شبکه‌های اجتماعی این طبقه ارتباط با دوستان و آشنایان و به روز بودن است .

شبکه‌های اجتماعی پر کاربرد و محبوب در تمامی طبقات instagram و youtube است .

واژه‌نامه فارسی به انگلیسی

social classes	طبقات اجتماعی
social media	شبکه‌های اجتماعی
social data science	علوم داده‌های اجتماعی
social identity	هویت اجتماعی
upper middle	طبقه‌ی متوسط به بالا
lower middle	طبقه‌ی متوسط به پایین
social sciences	علوم اجتماعی
computer science	علوم کامپیوتر
questionnaire	پرسشنامه
exploratory data analysis	تحلیل اکتشافی داده
univariate	تک متغیره
bivariate	دو متغیره
ordinal logistic regression	رگرسیون لوژستیک ترتیبی
p-value	پی-مقدار
clustering	خوشه‌بندی
regression	رگرسیون
classification	طبقه‌بندی
multiple choice	چندگزینه‌ای
binary variable	متغیر دودویی
response variable	متغیر پاسخ
predictor variable	متغیر پیشگو
correlation	همبستگی

واژه‌نامه انگلیسی به فارسی

Economic status	وضعیت اقتصادی
middle class	طبقه‌ی متوسط
KNN	K-نزدیک ترین همسایگی
naive bayes	بیز ساده
social classes	طبقات اجتماعی
feature selection	انتخاب ویژگی
random forest	جنگل تصادفی

مراجع

- [۱] کتاب ساختار طبقاتی جوامع پیشرفته - آنتونی گیدنز - انتشارات تیسا - سال انتشار : ۱۳۹۷
- [۲] Discovering Knowledge in Data: An Introduction to Data Mining, 2nd Edition - Daniel T. Larose, Chantal D. Larose - wiley - published : July 2014
- [۳] An Introduction to Statistical Learning: With Applications in R , 1nd Edition - Gareth M. James, Daniela Witten, Trevor Hastie, Robert Tibshirani - Springer - published : June 24, 2013

Abstract

In this project, we want to use data to get information about the relationship between social classes and social medias. To achieve the goal of the project, we used two data sets, one of them is the dataset on www.kaggle.com and the other one is collected data from a questionnaire.

After the investigations , we can say that wealthy strata spend more time on social medias , and the weaker people spend less time on social medias. Ofcourse , this difference isnot significant. in general , we can be said that social medias are accessible for all social classes. The reasons of using social medias in different classes are almost the same and the popular topics are not much different in different classes.



College of Science
School of Mathematics, Statistics, and Computer Science

The Relationship Between Social classes and Social medias

Tannaz Hafezi

Supervisor:
Dr. Abdollah Safari

A thesis submitted to Graduate Studies Office
in partial fulfillment of the requirements for the degree of
B.Sc. in
Applied Statistics

July 2022