عوامل موثر بر وضعیت تحصیلی دانش آموزان

فاطمه سلیمانی(9924803), سحرتوکلی(9920613)

*دانشکده برق و کامپیوتر,دانشگاه صنعتی اصفهان*

[fsoleimai697@gmail.com](mailto:fsoleimai697@gmail.com)

sahartavakoli2002@gmail.com

**مقدمه:**

در این مسئله عواملی که بر نتیجه ی 3 امتحان مختلف تاثیر دارد بررسی میشود و با توجه به نتیجه ی به دست آمده میتوان عواملی که تاثیر بیشتری روی نتیجه ی امتحانات دارد را بررسی کرده و بر این اساس برای بهبود نتایج امتحانات دانش اموزان از این فاکتور ها بهره ببریم.

نتیجه ی به دست آمده از این پژوهش، برای استفاده در مراکز آموزشی به منظور بهبود در وضعیت تحصیلی دانش آموزان می تواند بسیار کاربردی باشد.

**روش:**

**داده ها:**

دیتاست شامل 1000 رکورد از اطلاعات دانش آموزان است که برای هر دانش آموز اطلاعاتی از قبیل جنسیت،  قومیت، وضعیت تحصیلی والدین ، نوع ناهار مصرف شده و آمادگی برای امتحان وجود دارد و تاثیر این عوامل  بر سه نمره ریاضی و املا و روخوانی بررسی می شود.

وضعیت تحصیلی والدین دانش آموزان در 6 سطح از some collage  تا masters degree مقداردهی شده ، قومیت یا نژاد دانش آموزان نیز در 5 سطح در قالب گروه های A تا E مقداردهی شده است که هر کدام مربوط به نژاد خاصی هستند به عنوان مثال گروه A مربوط به سفید پوستان بریتانیایی هستند و … . نوع ناهار با دو مقیاس standard و free/reduced  مقداردهی شده و امادگی قبل از

امتحان نیز با دو مقیاس none  و completed بررسی میشوند.

**نحوه انجام کار:**

با استفاده از نمودار های مختلف، تاثیر عوامل مختلف بر نمره دانش آموزان بررسی شده و روابط بین ویژگی ها و نمره ها بررسی و تحلیل می شوند.

درانتها به منظور پیش بینی پاس شدن یا نشدن دانش اموز به طور کلی مدل های مختلف را بررسی و مدل با دقت بیشتر انتخاب میشود.

**پردازش:**

قبل از انجام هرکاری نیاز هست که داده های پرت و نوبزی  از دیتاست حذف شوند برای این منظور با کشیدن boxplot مربوط به

هر ویژگی داده های پرت  را شناسایی کرده و با در نظر گرفتن محدوده ای برخی از داده ها که نویزی شناخته شده اند حذف میشوند.

در دیتاست موجود تعداد رکوردها 1000 بوده و به همین دلیل در حذف داده ها باید با دقت عمل کرد. به همین دلیل با بررسی و تست بازه های مختلف در نهایت بازه ی مطلوب برای حذف داده ها به دست آمد. و پس از حذف داده های پرت تعداد کل آن ها به 939 رکورد رسید.

در قدم بعد به دلیل نیاز به سبد بندی به منظور انجام برخی بررسی ها دو روش مختلف سبد بندی انتخاب و بررسی شده و در نهایت روش بهینه تر انتخاب شده و در ازمایشات بعدی مورد استفاده قرار می گیرد.

**آزمایشات:**

به دلیل تحلیلی بودن ماهیت مسئله و ضرورت کشف ارتباط بین ستون ها مناسب ترین ابزار مورد استفاده رسم نمودارها و بصری سازی خواهد بود، به همین منظور با رسم نمودارهای مختلف سعی بر کشف ارتباط بین feature ها است.

در قدم اول به منظور یک بررسی کلی و مشاهده ی تاثیر هر ستون یک ستون جدید به نام percentage که شامل میانگین این 3 نمره است اضافه می شود .

در قدم بعدی با کشیدن نمودارهای مختلف توزیع نمرات در هر درس به تفکیک و همچنین روی میانگین بررسی می شود که در نتیجه مشاهده می شود در نمره ی writing بهترین عملکرد و بعد از آن در reading و بدترین عملکرد دانش آموزان در درس ریاضی دیده میشود.

در گام بعد تاثیر هر عامل را روی میانگین نمرات و همچنین به تفکیک روی هر نمره بررسی میکنیم:

فاکتور اول جنسیت است که با توجه به نمودارهای رسم شده مشاهده میکنیم خانم ها به طور میانگین و همچنین در نمره ی reading و writing عملکرد بهتری نشان دادند و در امتحان ریاضی اقایان عملکرد بهتری داشتند در نگاهی دیگر با رسم نمودار kdeplot جنبه ی دیگری از این تاثیر را مشاهده میکنیم که در نمره ی reading و writing تفاوت زیادی بین میانگین عملکرد خانم ها و اقایان مشاهده میشود در صورتیکه این اختلاف در نمره ی ریاضی کمتر به چشم می آید.

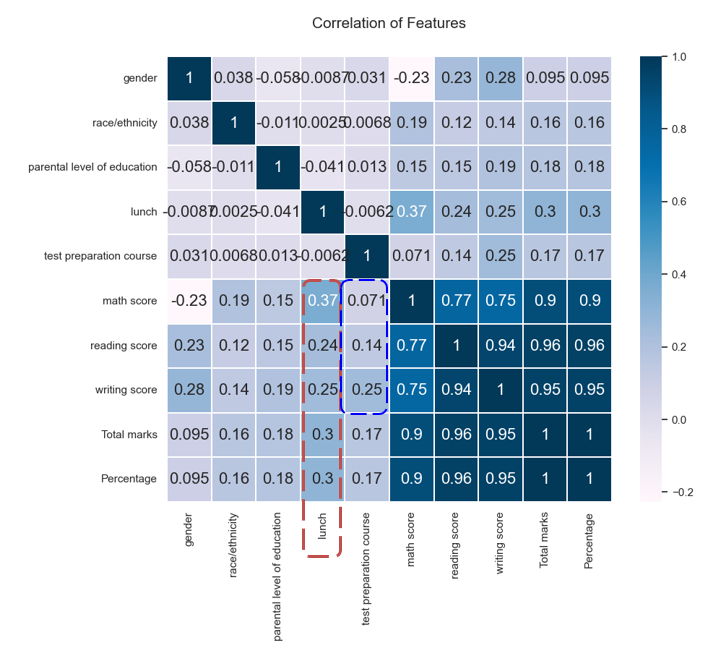
فاکتور بعدی وضعیت ناهار دانش آموزان است که مشاهده میشود افرادی که ناهار استاندارد داشتند علاوه بر میانگین نمرات به تفکیک در هر سه امتحان هم بهتر عمل کردند. با رسم نمودار kdeplot هم دیده میشود که تاثیر ناهار دانش آموزان روی نمره ی ریاضی بسیار چشمگیرتر از دو نمره ی دیگر است هرچند در دو نمره ی reading و writing هم تاثیر زیادی دارد.

فاکتور بعد آمادگی قبل از امتحان دانش آموزان است که افرادی که امادگی کامل داشتند نسبت به افراد بدون آمادگی عملکرد بهتری در میانگین و به تفکیک روی هر 3 نمره داشتند و هم چنین این فاکتور کمترین تاثیر را روی نمره ی ریاضی و بیشترین تاثیر را روی نمره writing داشته است.

به روش مشابهی که برای عوامل دیگر تشریح شد در بررسی سطح تحصیلات والدین دانش آموزان افرادی که والدین انها در مقطع masters degree تحصیل کردند در هر 3 نمره عملکرد بهتری نشان دادند هر چند تاثیر خیلی چشمگیری دیده نمیشود.

در بررسی فاکتور آخر یعنی قومیت افراد ، گروه D در writing score عملکرد بهتری نشان دادند و در reading score و math score گروه E بهتر عمل کردند.

پس از این مرحله لازم است نمودار heatmap به منظور بررسی correlation رسم شود که برای اینکار نیاز است داده های categorical استاندارد شده و به عدد تبدیل شوند بعد از اینکار نمودار correlation را رسم کرده و مشاهده میشود که ارتباط reading score و writing score تقریبا به صورت خطی است یعنی هرکس در یکی از ان ها عملکرد خوبی داشته باشد انتظار می رود در دیگری نیز خوب عمل کند و نرخ coef بین ان ها 0.94 است.



با توجه به نمودار بیشترین رابطه مربوط به ناهار و نمره ی ریاضی دانش آموز است و در کل مشاهده میشود، ناهار تاثیر به نسبت چشمگیری روی نمرات و به خصوص نمره ی ریاضی داشته است ، فاکتور بعدی که بررسی میشود امادگی قبل از امتحان دانش آموزان است که این عامل بیشترین تاثیر را روی نمره ی writing و حداقل تاثیر را روی نمره ی ریاضی داشته است.

مدل سازی: