

# توضیحات فاز اول پروژه‌ی داده‌کاوری

محمد کاهانی، علی گلدانی

عنوان پایگاه داده: SAP

۱۵ آذر ۱۳۹۸

## ۱ شرح کلی دیتاست

SAP یک دیتاست آموزشی می‌باشد که از یک سیستم مدیریت آموزش به نام ۳۶۰-Kalboard گردآوری شده است. این سیستم با هدف تسریع فرایند یادگیری، با استفاده از تکنولوژی‌های روز طراحی شده است و امکان دسترسی همگام‌سازی شده به منابع آموزشی را از طریق هر دستگاه متصل به اینترنت، به کاربران می‌دهد. این مجموعه داده از طریق یک ابزار ثبت فعالیت کاربران تهیه شده است. این ابزار رابط کاربری تجربه (xAPI) نام دارد. این ابزار امکان نظارت بر روی نحوه‌ی پیشرفت یادگیری کاربران را از طریق ثبت فعالیت‌هایی مانند خواندن یک مقاله یا مشاهده‌ی یک ویدیو فراهم می‌کند. این ابزار به مسئولین کمک می‌کند تا بتوانند با تحلیل روند استفاده و پیشرفت کاربران، یک تجربه‌ی آموزشی را توصیف کنند. این مجموعه داده شامل ۴۸۰ دانش‌آموز و ۱۶ ویژگی در رابطه با آن‌هاست. این ویژگی‌ها در ۳ دسته‌ی کلی طبقه‌بندی می‌شوند:

۱. اطلاعات جمعیتی مانند جنسیت و ملیت

۲. ویژگی‌های مربوط به سابقه‌ی تحصیلی مانند مقطع تحصیلی، سطح نمره و کلاس دانشجو

۳. ویژگی‌های رفتاری مانند تعداد دفعاتی که دانش‌آموز دستش را در کلاس بالا برده‌است، میزان استفاده از منابع درسی، پاسخگویی به پرسشنامه‌ها توسط والدین و رضایت از مدرسه

دانش‌آموزان بر اساس نمره‌ی کل آن‌ها به سه بازه‌ی عددی تقسیم می‌شوند:

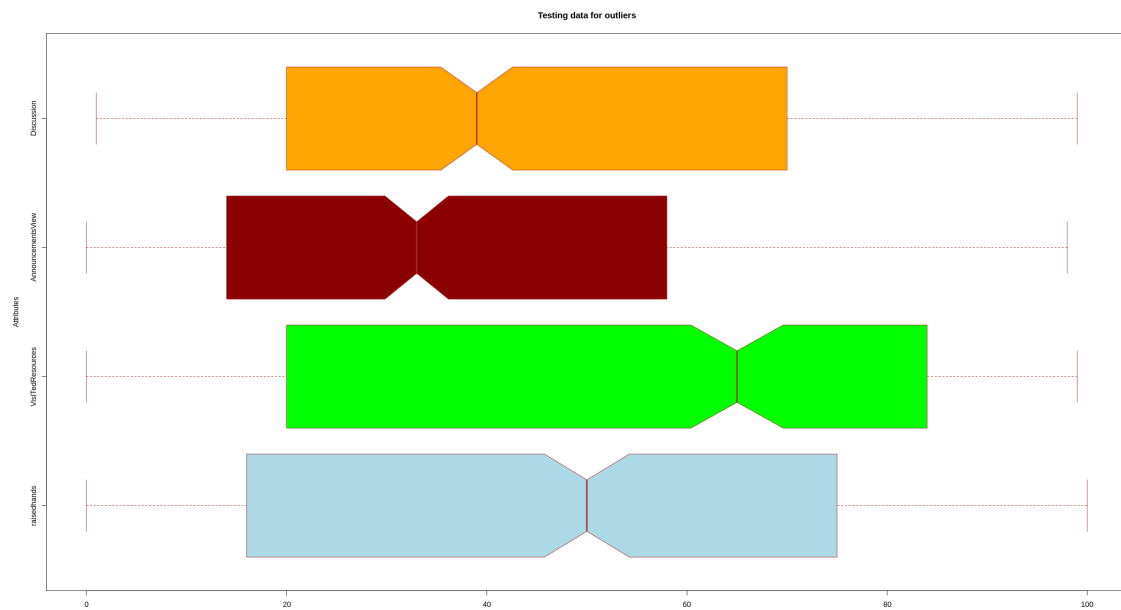
• سطح پایین: نمرات در بازه‌ی ۰ تا ۶۹

• سطح متوسط: نمرات در بازه‌ی ۷۰ تا ۸۹

• سطح بالا: نمرات در بازه‌ی ۹۰ تا ۱۰۰

توضیح ویژگی‌های مجموعه داده:

ویژگی	توصیف
gender	جنسیت دانش آموز
Nationality	ملیت دانش آموز
PlaceofBirth	محل تولد دانش آموز
StageId	مقطع تحصیلی دانش آموز
GradeId	سال تحصیلی دانش آموز
SectionId	کلاسی که دانش آموز در آن است
Topic	موضوع درس
Semester	نیم سال تحصیلی
Relation	پدر یا مادر مسئول دانشجو
raisedhands	میزان دفعاتی که دانش آموز دستش را در کلاس بالا برده است
VisitedResources	تعداد دفعاتی که دانش آموز منابع درس را مشاهده کرده است
AnnouncementsView	تعداد دفعاتی که دانش آموز اعلانات جدید را چک می‌کند
Discussion	تعداد دفعاتی که دانش آموز در گروه‌های مباحثه شرکت کرده است
ParentAnsweringSurvey	این که پدر یا مادر دانش آموز به پرسش‌نامه‌های مدرسه پاسخ داده است یا خیر
ParentschoolSatisfaction	این که پدر یا مادر از مدرسه رضایت داشته‌اند یا خیر
StudentAbsenceDays	این که دانش آموز بیش از ۷ جلسه غیبت داشته است یا خیر
Class	طبقه‌بندی دانش آموز بر اساس نمره‌ی کل



شکل ۱: BoxPlot Diagram

با رسم BoxPlot در شکل ۱ متوجه می‌شویم که داده‌ها بدون مقادیر نویز هستند.