

מצרף נתוני צפייה (ושמע) – נייר עמדה

התבקשתי על ידי הוועדה הישראלית למדרוג לחוות דעתי על היוזמה הבאה של משרד התקשורת: איסוף נתוני צפייה מגופים המודדים צפייה ושמע ויצירה אפשרית של מצרף נתונים אלו.

מצרף נכון ומהימן של נתונים ממקורות שונים (תלויים או בלתי תלויים) לכדי קובץ אחד אינטגרטיבי נחשב לאתגר מדעי-סטטיסטי. אתגר זה, הידוע בשמו המדעי DATA FUSION (DF), נוגע כמעט בכל המדעים מבוססי נתונים. גופים מסחריים גדולים עושים שימוש בטכניקות שונות של DF. כמי שחקר בנושא (ראה מקורות מדעיים בסוף מסמך זה), נתקלתי לא אחת בבורות מדעית נוקבת במתודולוגיה בה משתמשים חלק מגופים אלה. תוצאת שימוש לקוי זה הוא מצרף נתונים שגוי שאינו מייצג מציאות כמותית ריאלית.

נתונים גולמיים אינם עומדים בפני עצמם ולכן נדרש Fusion עם נתונים נוספים המבוססים על פאנל מקצועי. במקומות בעולם בהם עושים שימוש בנתונים מצרפיים תמיד הדבר נעשה בשילוב עם נתוני פאנל של חברה בעלת טכנולוגיה מתאימה כדוגמת Kantar, Nielsen, GFK.

אחת מהבעיות הידועות ביותר בנושא DF היא שהמידע הכמותי הנאסף ממקורות שונים אינו מודד למעשה תוכן דומה. אי ידיעה מוקדמת על המשמעות השונה של מדידות אלו פוסל לחלוטין את משמעות המצרף הסופי.

במקרה בו אנו דנים, נתוני צפייה גולמיים הנאספים מחברות כבלים עלולים להיות בעל משמעות שונה מאלה הנאספים על ידי חברות מדרוג פורמליות. כך למשל, חברות הכבלים מודדות על איזה ערוץ מכוון הממיר בכל זמן ולא האם מכשיר הטלוויזיה פתוח ומי מבני הבית צופה בו. כך עלולה להתקבל מדידת יתר של צפייה ממכשיר סגור המחובר לממיר.

קיימים בעולם תהליכים סטטיסטיים המאפשרים בחינת כל ממיר וממיר בנפרד על פי תנועת השלט בדקות שקדמו לדגימתו, ו"החלטה" האם מדובר בממיר שיש או אין מאחוריו טלוויזיה פעילה והעברת המידע אך ורק לגבי טלוויזיות פעילות. הליך זה חייב להיעשות בשילוב ובתאום בין מערכת המדרוג למערכת פלטפורמת התוכן. רק מערכות מדרוג מובילות בעולם כדוגמת המערכת של KANTAR או של NIELSEN או של GFK יודעות ומסוגלות בזהירות רבה לבצע אינטגרציה זו.

כל יומרה לבצע הליך זה באופן עצמאי, קל וחומר לצרף נתונים גולמיים של הפלטפורמות ולהציגם כנתוני צפייה, הינו פסול ויעוררו ביקורת שלא לומר לעג בקהילה המדעית בתחום הסטטיסטיקה ומדידת הרייטינג בעולם.

ככל שקיימת כוונה לעשות שימוש בנתוני מצרף לצורך מדידת הצפייה, ראוי שהביצוע יעשה כדלקמן:

1. בתהליך סדור ומבוקר סטטיסטית של מעבר נתונים באופן מבוקר בין ועדת מידרוג הפועלת בטכנולוגיה מקובלת (כגון הטכנולוגיה של KANTAR) לבין הפלטפורמה שתשתף פעולה בתהליך הסטטיסטי שימנע העברת מידע שגוי.
2. חיבור קוד תכנותי מבוקר ומאושר מדעית ליצירת מצרף רב-מקורות המבוסס על מיטב הספרות המקצועית המתאימה.

7.8.2023

תאריך



פרופסור צבי גילולה,

יועץ מדעי-סטטיסטי של ועדת המדרוג הישראלית

Gilula, Z, McCulloch, R, and Rossi, P (2006): Direct Data Fusion. **Journal of Marketing Research**, March, 1-15.

Gilula Z, and McCulloch R,: (2013): Multi level categorical data fusion using partially fused data. *Quantitative marketing and Economics*, Volume 11, pages 353-377

Gilula, Z., McCulloch, R., Ritov Y., and Urminsky, O. (2020): A study into mechanisms of conversion of ordinal attitudinal scales. **Quantitative Marketing and Economics**, 17, 32-51.