

נתונים אודות מיקומו של משתמש בטלפון סלולרי על פני מספר שבועות. העמודות מייצגות ימים,

# שביל קליפות המסרונים

איסוף מידע דיגיטלי על אדם אחד הוא פגיעה חמורה בפרטיות; שימוש במידע כזה שנאסף ממיליוני אנשים, באופן אנונימי, הוא העתיד

מאת ערן דינר

ומהם ניתן להפיק תובנות על התנהגות ותכניות תזוזה של אנשים, מספר בן גל. "למשל, קיימים נתונים לכל המשתמשים בתל אביב על פני כמה חודשים. כמוכן שאין מידע מיהם האנשים. כל מה שאנחנו מקבלים אלו נקודות שבהן אנשים היו וזו ברחבי העיר. אנחנו בונים על פי המידע הזה מודלים של התנהגות; מהן תכניות התזוזה שמאפיינות את האנשים האלה בעיר לאורך שעות היום והלילה. אחר כך אפשר להעלות על המודלים האלו כל מיני שכבות גיאוגרפיות. למשל, אם מישוה נמצא במקום מסוים בצהריים, חשוב לדעת אם זה מקום של מסחר, עבודה או אולי פארק."

**מה אפשר ללמוד מהנתונים האלה?**  
"אפשר להבין למשל אם אדם עובד במ"

קום קבוע או נע ממקום למקום כמו סוכן מכירות, או אם הוא יוצא לאכול, יוצא לפארק או הולך למשרדים אחרים, ואפשר לדעת היכן הוא גר, קונה ומבלה. עד רמה מסוימת, אפשר לדעת אפילו מהי הדרגה של אותו עובד בחברה. יש מחקרים שמראים איך מובילות משתנה כשעובד עולה בדרגה בחברה. אם היית יודע את שמו של אותו משתמש, היית נכנס לשאלות של פרטיות, אבל יש הרבה יישומים שאינם כרוכים בחידרה לפרטיות, כמו יישומים של אופטימיזציה

היסטוריה החלה באמצע השנה שעברה, עם הגילויים הראשונים מהמסמכים שהדליף אדוארד סנודן על המעקבים שביצעה הסוכנות האמריקאית לביטחון לאומי (NSA) באמצעות כלים שכולנו משתמשים בהם מדי יום, כמו גוגל ופייסבוק. אם תמיד ידענו שאי אפשר לבלוש אחרינו ברשת, מסמכי סנודן הוכיחו שאכן, האח הגדול בהחלט עלול לפשפש בתיבת המייל שלנו, בסמסים ואיפה לא.

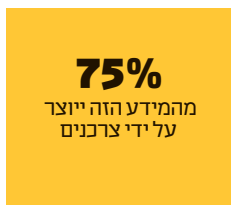
העניין הוא שהאח הגדול לא באמת צריך לקרוא על החוויות שלנו מהנופש באילת כדי לדעת איפה היינו ומה עשינו. גם אם נצפין את כל תכני תובות המייל ונכתוב סמסים בשפת הבי"ת, עדיין ניתן יהיה להפיק אודותינו מידע רב רק מתוך השובר הדיגיטלי שמוותרות אחריו של הפי"עולות היומיומיות שאנו מבצעים, כמו תשלומים בכרטיסי אשראי, שימוש בכספומטים, שליחה וקבלה של שיחות ומסרונים וכמוכן, עצם נשיאת הטלפון הסלולרי, שמשרד ללא הרף נתונים על מיקומו. מידע כזה שנאסף מהמוני משתמשים על פני זמן עשוי להיות בעל ערך גדול במיוחד למגוון שימושים. במחלקה להגדרת תעשייה וניהול באוניברסיטת תל אביב פיתח צוות חוקר רים בראשותו של פרופ' עירד בן-גל אלגוריתמים שמאפשרים לנתח מידע ולנסח מתוכו תובנות שימושיות עבור גופים ממשלתיים ומסחריים.

"אנחנו משאירים אחרינו חותם דיגיטלי בכל מקום, כל הזמן", אומר פרופ' בן-גל. "לא מדובר רק במחשבים או בסמארטפונים. בכל מקום שבו אתה הולך בעיר גדולה, בסופו של דבר יש מצלמות, כלומר גם כשאתה לא במחשב או משתמש בטלפון שלך, אתה חשוף. צריך להחליט גם מתי מקריבים את הפרטיות למען צרכים מוצדקים, למשל כדי לקבל שירות טוב יותר או הגנה טובה יותר", הוא אומר. "מקום אחד שבו אתה מקריב במידה מסוימת את הפרטיות הוא כשאתה מבצע ניתוח אנליטי של הרבה אנשים יחד."

על בסיס האלגוריתמים שפותחו באוניברסיטה הקימו ב־2007 בן-גל ושות' פו, ד"ר גונן זינגר, את הסטארט אפ C-B4 (Context Based 4casting), שפיתח תוכנה לזיהוי וניתוח דפוסי נתונים נסתרים במאגרי מידע גדולים, במטרה לתרגם למידע עסקי עבור חברות קמעוניות, חברות טלקום וארגוני ביטחון פנים. המידע שמשמש את בן גל וצוותו הוא אינדיבידואלי, אבל לא בהכרח פרסונלי – כלומר, הוא נאסף מאנשים אמיתיים, אך אינו מזהה אותם בשמותיהם ולא מכיל מידע אודות תוכן המסרים או השיחות שלהם בטלפון הנייד.

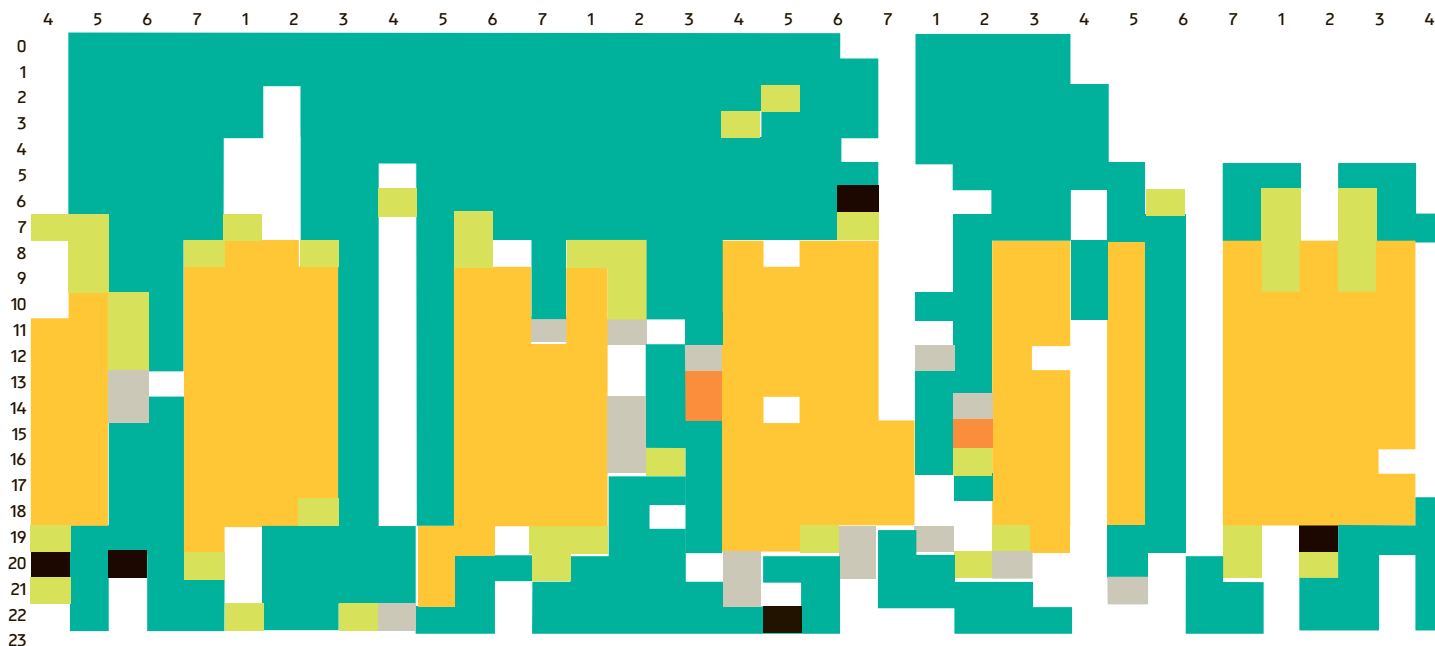
"בפרויקט מחקר אחר שלנו, ששותפים בו ד"ר ערן טוך ופרופ' בועז לרנר וקבוצת סטודנטים דנטים לתארים מתקדמים, אנחנו מקבלים מחברה סלולרית גדולה נתוני מיקום ותזוזה של המוני משתמשים באזורים גדולים,

5 6 7 1 2 3 4 5 6 7 1 2 3 4 5 6 7 1 2



מקורות: IDC, Radicati Group, TR Research, Pew internet

בית ■ עבודה ■ נסיעה ■ מקום בילוי ■ ביקור אצל ההורים ■ קניות



**פרופ' עירד בן גל: "כדי להילחם בסכנת הפגיעה בפרטיות צריך להחליט גם מתי מקריבים אותה למען צרכים מוצדקים, למשל כדי לקבל שירות טוב יותר או הגנה טובה יותר"**



צילום: ניר קידר

רק על איך לאגור ולארגן את מאגרי הנתונים כך שיהיו מוכנים לשליפה. אחר כך הגיעו מער"כות הבינה העסקית (BI), אבל הן היו טובות כדי לתת תשובות לשאלות שאתה רוצה לשאול: אם אני חושב שלקוחות שלי בשנות ה-20 לחייהם, שגרים במקום מסוים, לא משתמשים במוצר A, אני יכול לאמת או להפריך את ההשערה הזאת עם מערכת בינה עסקית. אבל כדי לגלות תבניות שימוש לא מוכרות זה לא מספיק. עכשיו מגיע השלב של חיזוי אנליטי (Predictive Analytics), שבו המערכות שואלות בשבילך את השאלות. אתה שואל שאלות גדולות כמו 'מיהו לקוח טוב?' או 'איך מגינים על הרשת', והן מוצאות עבורך את השאלות הפרטניות. "הדור הבא, שחלקו כבר קיים, לא רק שואל את השאלות אלא גם מציג המלצות יישומיות. בשביל זה צריך להבין עוד משהו: מהם הגורמים שאתה שולט בהם בארגון. למשל, אני יכול לדעת שלקוח מסוים במקום מסוים קונה מוצר אם יורד גשם, אבל אני לא יכול להמליץ שיירד גשם. אם מסתכלים על הטרנד הזה, רואים שבסופו של דבר יותר ויותר מההמלצות האלה גם ימור משו אוטומטית. גם כאן הסיפור יהיה לאזן בין הפרטיות לבין אנליטיקה, שתהיה פרטנית עד לרמה מבחינה. ואז השאלה היא מה אנחנו נעשה. אולי נלך לים".

eran.dinar@themarket.com

על השעות ומשך ההתקשרויות. פרופיל השימוש התקשורתי שלהם הוא שונה מאוד. בדרך כלל הם משתמשים ביותר ממכשיר אחד, שעות העבודה שלהם אחרות, השיחות שלהם קצרות ושעות הפעילות הן בלילה. הם גם משתמשים הרבה יותר בסמסים. עוד משהו שקפץ בצורה מאוד מובהקת הוא היחס בין רקות השיחה לסמסים. בזמן נסיעה למשל, אנשים 'רגילים' מדברים הרבה יותר משהם מסמסים. אצל החשודים, זה לא בהכרח כך. "פה נכנסת השאלה של חדירה לפרטיות כדי להגן על האוכלוסייה. אצל 99% מהאוכלוסייה לא רואים תבניות מובהקות כאלה, ולגבי האחוז הבודד שאצלו נמצאות תבניות מחשירות, אפשר ללכת לבית משפט ולבקש צו, כי אולי במקרה כזה החדירה לפרטיות מוצדקת".

**לאן עוד עשוי התחום הזה להתפתח?**  
 "תחום האנליטיקה של כמויות גדולות של מידע (ביג דאטה) ללא ספק ילווה אותנו בשנים הקרובות. על פי נתונים של גרטנר, זה תחום שכמעט 78% מהחברות מעוניינות להיכנס אליו, ורק 8% מהן כבר שם. אנב, השנה נפתחה אצלנו במחלקה תוכנית התמחות חדשה לסטודנטים לתארים מתקדמים, שתעסוק ביישומים של אניליטיקה עסקית בהיבט של תהליכים ארגוניים ובשיתוף עם חברות מובילות. "כדי לדבר על הכיוון שאליו המדע הזה הולך, צריך להבין איך הוא התפתח. עשר שנים עברו

של תחבורה, קאר-פולינג ואפילו תמחור שונה למסלולים בכביש על פי עומס צפוי. אם יש בידוך מיליון יומנים כאלה של אנשים שונים, היכולת לנהל את זה ביעילות עולה מאוד.

"מידע כזה עשוי לשמש גם את רשויות המדינה. בסוף 2012, במבצע עמוד ענן, אספנו מידע שמראה איך עיר מתנהגת בזמן התראת 'צבע אדום'. אינפורמציה כזאת יכולה להיות חשובה מאוד לפיקוד העורף, ולשמש אותו כדי לדעת למשל איך למקם בצורה אידיאלית מקלטים או מחסות. מעבר לכך יש גם מידע שישמש גופים מסחריים. נניח שיש אנשים שידוע שבצהריים הם תמיד פנויים או שהם עובדים בדרך כלל מחוץ לעיר, אבל בימים מסוימים הם שוהים בתוך העיר. על בסיס המידע הזה אפשר להציע להם שירותים שונים, החל בשירותי תשתית כמו סמארט סיטיז וכלה בקופונים, נניח, שמבוססים על הלייפסטייל המסוים הזה".

**זה בדיוק המקום שבו אנשים חשים מאוימים מיישומים כאלה. אתה מיד חושב "איך הם יודעים מה אני צריך בדיוק עכשיו?".**  
 "מצד אחד זו יכולה להיות חדירה איזומה לפרטיות, מצד שני אפשר לחשוב על הרבה יישומים שמאפשרים להשתמש במידע כזה בלי להגיע לאדם עצמו.

"הנה דוגמה אחרת שנוגעת לשאלת הפרטיות: עברנו עם חברה שעוסקת בתחום הביטחון על נתונים סטטיסטיים של משתמשים וקיבלנו מהם מידע על מורשעים בפלילים (לא בישראל). די בקלות אפשר היה לסמן מאפיינים שמתארים את ההתנהגות של אנשים חשודים, ללא כל מידע על התוכן של השיחות או הסמסים שלהם, אלא רק