

טכנולוגיות עסקיות בניהול

שיעור 11

Big data

BI – הוא תחום בטכנולוגיות מידע העוסק בבניית מאגרי מידע, העוזרים לארגון להפיק מידע משמעותי מבחינה עסקית מתוך מכלול הנתונים הנאספים על ידם.

מערכות הבינה העסקית מספקות מידע היסטורי, עכשווי ותחזיות בנוגע לפעילות העסקית, ובעזרתן ניתן לאתר דפוסים משמעותיים לניהול של ארגון וליצור דוחות, התראות ואיתותים לצורך קבלת החלטות ניהוליות.

Data mining (כריית מידע) – היא תכנית מחשב לצורך גילוי מידע הטמון בבסיסי נתונים קיימים והסקת מסקנות מהצלבתו. גילוי ידע בבסיסי נתונים הוא תהליך שנועד לחקור ולנתח כמות גדולה של מידע ככל שניתן באמצעים אוטומטיים, כדי לגלות דפוסים חדשים ושימושיים. כלומר, המנהלים יוכלו להשתמש בדפוסים שהתגלו בכדי לשפר תהליכים קיימים ולתפוס את המידע לידע.

דוגמא – אלגוריתם כריית מידע לאיתור דואר זבל אשר מבחין בין דואר רצוי ללא רצוי. הוא יאומן על סדרת דוגמאות דואר, ולאחר האימון התבניות שאותרו יופעלו על מכתבים חדשים. מידת הדיוק של האלגוריתם תיקבע על פי כמות הפריטים שאותרו נכון.

דוגמא – הפיכת המידע לידע – בתחתית פירוט האשראי שמקבלים בכל חודש ישנם קופונים ומבצעים. לכל אדם בפירוט האשראי יש קופונים אחרים, לפי הקניות של אותו אדם. כך חברת האשראי הופכת את המידע לידע על הלקוחות.

Big data – הוא מונח שמשמעותו לרוב הוא יישום של כלי בינה מלאכותית, כמו למידה ממוחשבת של סט נתונים רחב היקף, מעבר לאלו שאוכסנו עד עתה בבסיסי נתונים אופייניים.

מקורות נתונים אלו כוללים, בין השאר, נתונים המגיעים מעקבות דיגיטליות של גלישה באינטרנט, תשדורות ברשתות החברתיות, נתונים חיישניים ונתונים מאמצעי מעקב טכנולוגיים.

המונח big data מציין מגמה מתקדמת בטכנולוגיה שפותחת דלת לגישה חדשה להבנת העולם וביצוע החלטות. חוקרים מאוניברסיטת ברקלי העריכו שמשחר האנושות ועד לשנת 2003 האנושות יצרה בספרים, מכתבים, סרטים וכו' כמות מידע שמיוצרת כיום **ביומיים** באינטרנט.

כמעט כל המידע שבעולם שמור בשרתי מחשבים, בעיקר של ממשלות וארגונים גדולים.

העולם טובע במידע. **מעל מיליארד** משתמשים מעלים ל-Facebook מעל 10 מיליארד תמונות מידי שעה. Youtube מזינים **מידי שנייה כשעה** של תכנים.

עיקר הגידול ב-Big Data הוא בנתונים פראים. כלומר, מושגים לא מובנים כמו – מילים, תמונות וסרטוני וידאו שמועלים לרשת. נתונים אלה נקראים **מידע בלתי מובנה** והוא ניתן לעיבוד כמו מסדי הנתונים המסורתיים.

טכנולוגיות עסקיות בניהול

שיעור 11

מימדי Big Data

1. **Volume – נפח** – Big Data מגיע רק במידה אחת XXL. עסקים כיום מוצפים מרגע לרגע בהרי מידע שצריך לנתח.
2. **Velocity – מהירות** – על מנת להפיק מהנתונים את המרב, השימוש בהם חייב להיעשות במהירות, ברגע שבו הם זורמים אל הארגון.
3. **Variety – מגוון** – המידע היום כולל מידע מובנה ולא מובנה (כמו טקסט, שמע, וידאו, לחיצות עכבר וכו') Big Data הוא הזדמנות לחשוף תובנות חדשות מתוך סוגים חדשים של מידע, להפוך עסקים לגמישים יותר, תחרותיים יותר ולפתור שאלות שבעבר נותרו ללא מענה.
4. **Value – ערך** – הערך של הלקוח. כלומר, מיקוד בלקוח ושיווק פרסונלי ע"י שימוש במידע מובנה ומידע בלתי מובנה. ניתוח של נתונים שמביא למצב של מיצוי מידע. הדבר יכול לעזור לארגון להגיע למסקנות עסקיות ולהבנות חדשות ומדויקות יותר מכפי שהיה אפשרי עד היום.

Hadoop – היא תוכנת ניהול של הרבה מחשבים שעובדים בו זמנית. למעשה, זו רשת של מחשבים המחוברים אחד לשני ואדם מנהל אותם. הרבה חברות בעולם כמו Google, Facebook משתמשות בו. הרעיון הוא שאם יש המון מידע, הוא מפוזר על פני הרבה מחשבים, והם מחלקים בינם לבין עצמם את העיבוד של הנתונים. כלומר, במקום שכל מחשב יעבוד בנפרד, העיבוד מתחלק בין כל המחשבים ואותה שאילתה מורדת בו זמנית בכל המחשבים.

התהליך ש-Hadoop עושה מאחורי הקלעים נקרא – **Map Reduce**

Map – מפזרים את המידע בין המחשבים השונים כדי להקל על העיבוד במקביל. ה-Map מציב את אותה מילה במחשב אחד, הוא עובר על המילים סופר אותם בכל מחשב ומחשב Hadoop מעביר את התוצאה למחשב אחר שהוא בוחר.

Reduce – כל מחשב לוקח את התוצאות שקיבל ועושה עיבוד. למשל, השאילתה שמורדת במקרה שלנו היא ספירת המילים המופיע בגוגל. הרעיון הוא ש-50 אלף מחשבים עובדים מהר הרבה יותר במקביל ממחשב אחד.