

האבולוציה של אנליטיקה בארגון

(מתוך הבלוג של STKI)

Data is the new Oil

כך לפחות נהוג להגיד Data. הפכה להיות שיחת היום בקרב הנהלות ארגונים, והנושא מתחיל לקבל את החשיבות הראויה לו כ Enabler של מהלכים עסקיים, כאמצעי לעמידה ברגולציות מחייבות, וגם כתחום משמעותי לצורך הישרדות בשנים הקרובות.

חברות כיום מוערכות כלכלית יותר ויותר על סמך הדאטה שברשותן, וגם על סמך הדאטה הפוטנציאלי שעשוי להיות ברשותן (המנגנונים שהן מצליחות לייצר ולקלוט דאטה/ "קולטני מידע") ועל סמך היכולת שלהן להעשיר ולנהל את הנתונים שבבעלותן.

אולם ככל שארגונים אוגרים ואוספים יותר ויותר דאטה, וככל שיכולות איסוף, אחסון, עיבוד וניהול הדאטה (Big Data) משתכללות, מתבהרת גם ההבנה שאיסוף דאטה זהו רק תנאי הבסיס.

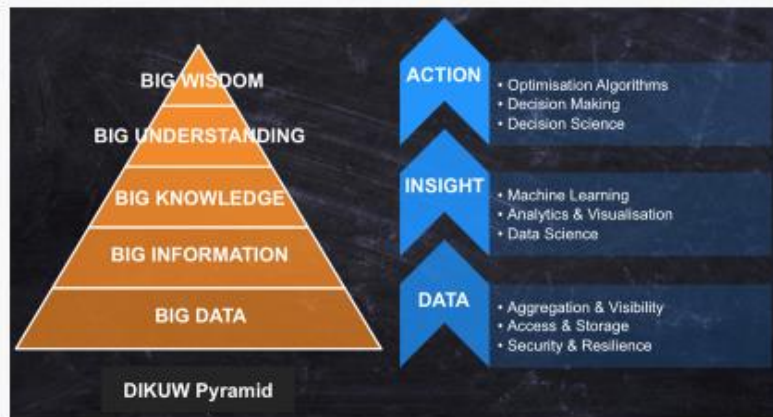
Data Driven Businesses

ארגונים שואפים להפוך ולהיות Data-Driven, מה זה בדיוק אומר?

Data -> Insights -> Action

1. **DATA**: להיות מסוגלים לנהל היטב דאטה, לאסוף, לאגור, לחבר תשתיות נתונים שונות, להעשיר את הדאטה כל הזמן ולייצר שכבת דאטה אמינה, מנהלת, עם מעטפת של Data Governance.
2. **INSIGHTS**: לייצר תובנות על סמך אותה שכבת דאטה, יצירת מודלים אנליטיים ואלגוריתמים על ידי אנליסטים ו Data Scientists תוך שימוש בכלי אנליטיקה ואנליטיקה מתקדמת.
3. **ACTION**: אין ספק שזהו השלב המאתגר והחמקמק ביותר. לייצר מנגנונים תפעוליים שיאפשרו לארגון לפעול כל הזמן על סמך אותן תובנות (ולא על סמך "תחושת בטן" לדוגמה), כל יום ובכל נקודת החלטה.

Just having a lot of data isn't enough!



STKI's work Copyright©2016. Do not remove source or attribution from any slide, graph or portion of graph

4

לכולנו ברור שאנחנו לא ממצים את פוטנציאל הדאטה במובן של ההחלטות והפעולות שאנו נוקטים, אשר אינן מתבססות על דאטה אלא יותר על רגשות, כוח האירציה (כך פועלים בארגון שלנו), העדפות פרסונליות ועוד. אבל עד כמה אנחנו לא מספיק פועלים על סמך הדאטה ש*פוטנציאלית* יכול היה להוות הבסיס לפעולה? על פי מחקר של חברת Pivotal, רק כ-0.5% מה Data בעולם הנו **Operationalized Data!** הפוטנציאל הוא כמעט אינסופי (כמות הדאטה הפוטנציאלית היא כמעט אינסופית, והאופן בו נוכל להסתמך על דאטה בקבלת החלטות תלוי בעיקר בדמיון שלנו, ביצירתיות המחשבתית, ובהבנה כיצד דאטה ותובנות יכולים להשפיע בארגון שלנו).

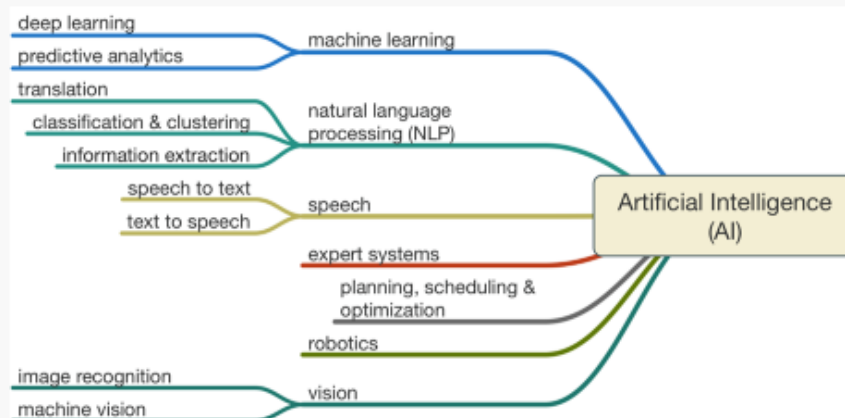
AI, AI, AI! מה הייתה השאלה?

למרות שהמונח AI הפך להיות המונח השיווקי החם ביותר כיום (מחליף את המונח ביג דאטה שעד לא מזמן כיכב בראש מצעד הבאזזוורדים), טכנולוגיות בינה מלאכותית (AI) ומחשוב קוגניטיבי הן תוצאה של התפתחות אבולוציונית רבת שנים ולא מהפכה שהתרחשה רק בשנים האחרונות. אולם כן ישנה הרגשה שטכנולוגיות AI הגיעו לנקודת ציון משמעותית לאחרונה, בשל צירוף של שתי תופעות טכנולוגיות:

- זמינות גבוהה של כמויות נתונים משמעותיות (Data Sets) גדולים (שהפכו להיות זמינים לנו רק לאחרונה. קיומם של Data Sets עצומים הנו תנאי בסיס להתקדמות משמעותית בתחום.
- יכולות עיבוד וכוח חישובי זמין וזול יחסית

AI הינו מונח רחב וגמיש מאוד, תחתיו נכנסים עוד הרבה תתי-תחומים, חלקם קיבלו דחיפה משמעותית וזכו לפריצות דרך של ממש לאחרונה, כמו תחום העיבוד החזותי, עיבוד שפה ועוד.

Artificial Intelligence is a VERY broad term



STKI's work Copyright©2016. Do not remove source or attribution from any slide, graph or portion of graph

9

בתוך עולם ה-AI-טכנולוגיות מבוססות רשתות נוירונים, ובמיוחד תחום ה-Deep Learning- DL, מהווה **התחום הצומח ביותר** עם אחוזי גידול של 200-300% בקצב האימוץ על ידי Data Scientists. על פי סקר של KD Nuggets, ב-2016, 18% מ-Data Scientists כבר עשו שימוש ב-DL לפחות בפרויקט עובד אחד. על פי תחזית גרטנר, עד 2018 אצל כ-80% מה-Data Scientists בעולם, השימוש ב-Deep Learning יהפוך להיות עוד כלי סטנדרטי בארגז הכלים שלהם.

היכן אנו מצויים ברצף ההתפתחות האבולוציונית של עולם האנליטיקה?

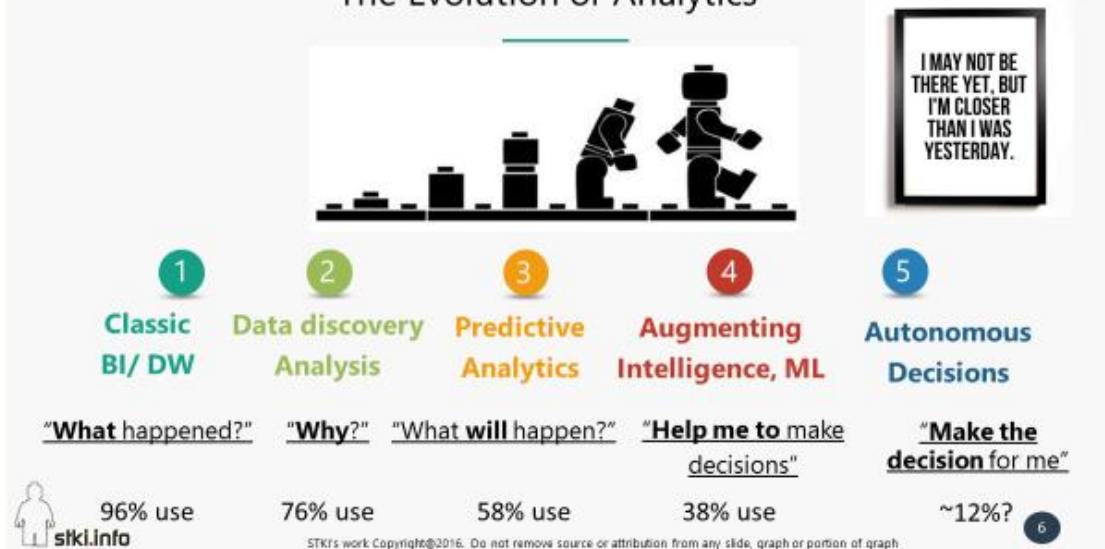
כבר ברור כי שלב ה-BI "הקלאסי" מוצה ברוב הארגונים בתעשיות השונות, ובתחומים השונים בארגון (פיננסים, שיווק, לוגיסטיקה, תפעול, משאבי אנוש – התחום האחרון שנכנס לנושא רק לאחרונה), כלומר, אנו לרוב כן יכולים לשאול "מה קרה בעבר"? ולקבל תשובה טובה. רובנו גם התקדמנו ליכולת לבצע Analysis חיתוכים והסתכלות מעמיקה על הנתונים בחתכים שונים כדי לענות גם על השאלה "מדוע זה קרה"? חלקנו עושים זאת באמצעות קוביות, חלקנו באמצעות סביבות של ויזואליזציה ו-Data Discovery אולם עדיין מדובר על נתוני עבר והסתכלות רטרופקטיבית.

השלב הבא באבולוציה היווה קפיצה משמעותית מבחינת הערך והתועלת העסקית בארגונים, והוא תחום ה: Predictive Analytics היכולת לשאול "מה עתיד לקרות"? זהו תחום ותיק שניתן למצוא שימושים מסוימים שלו בלא מעט ארגונים, אולם המפגש של תחום זה עם עולם הביג דאטה ועם תחום ה ML-(השלב האבולוציוני הבא) מאפשר לנו להבין יותר טוב את הדאטה, וזו תועלת משמעותית ככל שאנו עובדים עם מגוון יותר גדול של נתונים (חלקם לא מובנים) ונפחי נתונים משמעותיים. כתוצאה מהבנת הדאטה, הישויות, הנושאים, הקשרים, ה Features, אנו יכולים להבין איזה סוג שאלות אנחנו יכולים בכלל לשאול ומה כדאי לנו להמשיך ולחקור. כאן סוג המחקר הוא יותר Exploratory: אנו פחות נדע מראש מה השאלה, נגלה אותה תוך כדי מחקר ושיטוטים בדאטה ("Data Journalist" אמן נתונים שעוסק בוויזואליזציה והצגה מתוחכמת של נתונים בדרכים המובילות לפעולה (טען ש Data is not the new Oil, לדעתו הגדרה מדויקת יותר היא Data is the new Soil (אדמה). בדאטה יש לשחק, לחפור, עד שמגיעים למשהו מעניין ומתוכו ניתן להצמיח נבטים של תובנות משמעותיות מאוד.

בשלב זה אנחנו כבר מתחילים להתקרב לשלב ה Prescriptive-תנסו להגיד את המלה הזו 5 פעמים ברציפות, נראה אתכם Prescriptive Analytics. (עונה על השאלה: "כיצד יש לפעול על סמך הדאטה"? מה צריך לעשות? דאטה שמוביל לפעולה.

מהי הקפיצה הבאה, אליה עדיין לא הגענו – Autonomous Decisions? מכונות שלא רק ישאלו את השאלות בעצמן, אלא גם יחליטו על סמך התשובות בעצמן ויבצעו את הפעולה בעצמן Data-insights-actions. באופן אוטונומי. ישנם תהליכים מסוימים ב Domain מצומצם שכבר עובדים בתצורת Autonomous decisions, אולם באופן כללי לא הגענו לרמת חוכמת" AI כללית (General AI) "שאינה צריכה שאנו, האנשים, נאכיל אותה – במבנה הדאטה, במשמעות ובמנגנון ההסקה והפעולה.

The Evolution of Analytics



במפגש שולחן עגול שקיימנו לאחרונה בנושא, עלה כי התחום בישראל מעניין ארגונים רבים אבל כמעט כולם מרגישים ש"הם עוד לא שם", ה Use Case כלל לא ברור (החסם העיקרי בעולם כיום לאימוץ AI בארגונים – לכל מי שעוסק בתחום הביג דאטה זה יישמע מאוד מוכר מהשלב ההתחלתי שאפיין את השוק לפני שנתיים). עלתה התחושה הכללית שפוטנציאל הדאטה לא ממומש, כי הארגון "עוד לא שם" ושהם עדיין די רחוקים מהדגמת איזשהו ערך עסקי מוחשי. רוב הארגונים כבר עברו (או עוברים בימים אלה) מ BI קלאסי להוספת ממד ה Data Discovery (Self Service בנוסף לדוחות הקלאסיים). כמחצית מהארגונים כבר מתנסים כיום ב Predictive Analytics אולם באזורים מאוד ספציפיים בארגון (ניהול סיכונים/ מניעת נטישה/ גילוי הונאה וכד.).

כלים טכנולוגיים:

בעוד שבעולם ה BI-הקלאסי ישנה שאיפה לסטנדרטיזציה עם כלי ארגוני סטנדרטי אחד (או לפחות ניסיון משמעותי להפחתת מספר הכלים השונים, בניסיון להשיג יותר שליטה על השכבה הסמנטית ועל אופן השימוש בנתונים), בתחום ה – Data Science-היותר חדש, ישנה נכונות גדולה יותר לנסות מגוון של כלים ולתת למשתמשים יותר עצמאות באופן ניתוח במידע, גם אם זה אומר שימוש במספר כלים שונים ויותר מסוים על רמת השליטה באופן השימוש בנתונים לעת עתה.

כלומר ישנו שינוי תפיסתי בהתייחסות לכלים הטכנולוגיים – בשכבת אחזור המידע.

על פי סקר של KD Nuggets, כ-18% מ Data Scientists שנשאלו בסקר ענו כי הם כבר התנסו בפועל בפרויקט המערב יכולות מבוססות) Deep Learning עלייה של 200% לעומת השנה שעברה DS. (משתמשים במוצע בכ-3 כלים שונים לוויזואליזציה וניתוח המידע).

על פי סקר של KDNuggets על הכלים בהם DS מעידים כי הם עושים שימוש:

Next table has the top 10 most popular tools in 2016 poll

Tool	2016 % share	% change	% alone
R	49%	+4.5%	1.4%
Python	45.8%	+51%	0.1%
SQL	35.5%	+15%	0%
Excel	33.6%	+47%	0.2%
RapidMiner	32.6%	+3.5%	11.7%
Hadoop	22.1%	+20%	0%
Spark	21.6%	+91%	0.2%
Tableau	18.5%	+49%	0.2%
KNIME	18.0%	-10%	4.4%
scikit-learn	17.2%	+107%	0%

18% use DL (up 200%)
Gartner: By 2018 DL will become a standard component in 80% of Data Scientists' toolbox

Source: KDNuggets



<http://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html/2>
STK's work Copyright©2016. Do not remove source or attribution from any slide, graph or portion of graph

Data Management, Data Governance ומה שביניהם

במפגש שערכנו בנושא Data Driven Businesses לאחרונה עסקנו בחשיבות של ניהול ומשילות נתונים באופן מרכזי בארגון.

CDOs – Chief Data Officers

מה עדיין חסר?

עם כל ההשקעה שנעשתה באיסוף דאטה וגזירת תובנות על מנת לשפר מהלכים והחלטות עסקיות, מה עדיין חסר? ולמה רוב הארגונים לא באמת פועלים על סמך הדאטה?

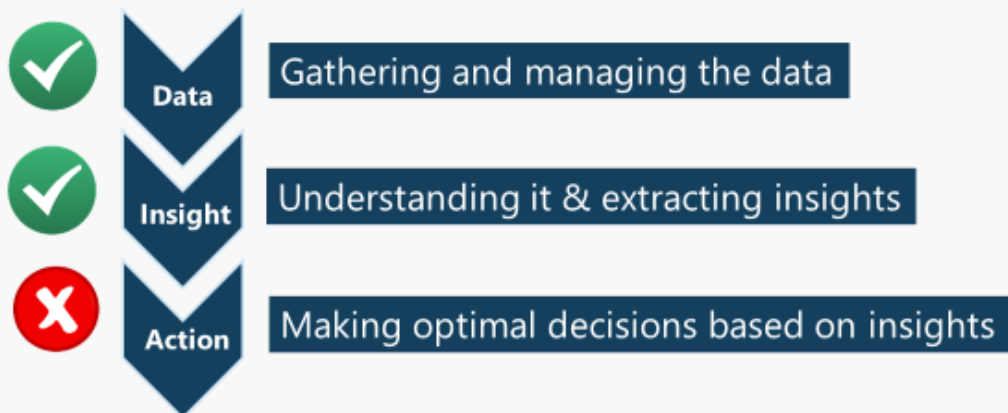
אין ספק שהפכנו להיות אלופים באיסוף דאטה. טכנולוגיות ביג דאטה מאפשרות לנו לעשות את זה אפילו יותר טוב, וארגונים מתארגנים כיום עם אנשים (CDOs, Data Governance Tools, EIM, Data Governance) כלי ניהול נתונים על מנת לנהל יותר טוב את שכבת הדאטה ולהפוך אותה לבסיס טוב לקבלת החלטות והסתמכות על נתונים.

אנחנו גם לא רעים ביכולת לגזור תובנות ומודלים אנליטיים על סמך הדאטה שברשותנו. וניתן לראות שאנו מתקדמים ונעים במסלול הזה בכיוון טוב. אנליסטים עוברים הכשרות להפוך להיות Data Scientists, יותר אנשים בארגון מעוניינים בגישה ה Self Service ובאופן כללי התחום נע בכיוון החיובי.

הבעייתיות נעוצה בשלב ה Action. וזה נובע לטעמנו משתי סיבות שונות:

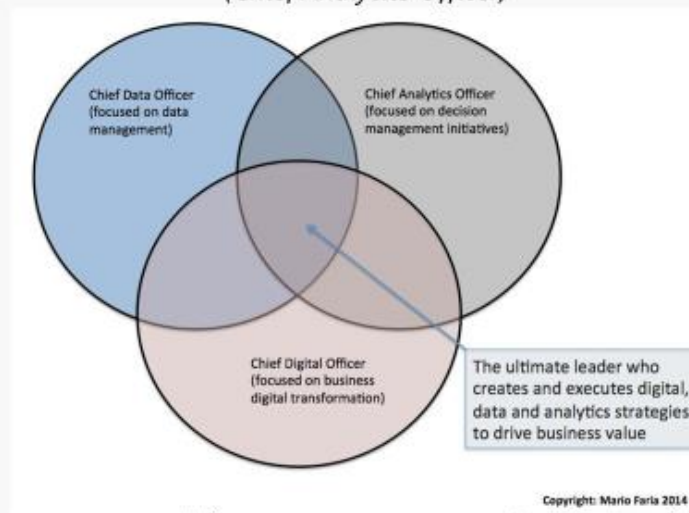
1. סיבה "פסיכולוגית" – אנשים מרגישים יותר בנוח להסתמך על תחושת הבטן שלהם לקבלת החלטות מאשר על נתונים. להפוך להיות ארגון Data Driven זה לא רק טכנולוגיה, זו תפיסה, תרבות, וזה תהליך שלוקח זמן.
2. סיבה תפעולית – בארגונים כיום לא קיימת התשתית התהליכית שמתרגמת תובנות לפעולות. האוטומציה נעצרת איפשהו בשלב גזירת התובנות. כיצד נוודא שתובנה מסוימת מפעילה פעולה בצורה אוטומטית? בתחום ה AI אחד התחומים המעניינים ביותר כיום הוא תחום ה – Decision Management תחום שכולל בתוכו כלים בסגנון מנועי CEP – Complex Event Processing, מנועי חוקה וכלים נוספים המניעים לכיוון יישום התובנות בקבלת החלטות וביצוע פעולות. בנוסף, ספקי מערכות עסקיות (ERP, CRM, Marketing automation, Supply Chain Management) יקלו על החיבור הזה בעתיד מכיוון שבבסיס של מערכות אלה תהיה שכבת AI שתסייע להניע תהליכים "יותר חכמים" תוך חיבור בין עולם התוכן התפעולי ואנליטי Embedded analytics – בתוך תהליכים.

The problem: Turning data into actionable decisions

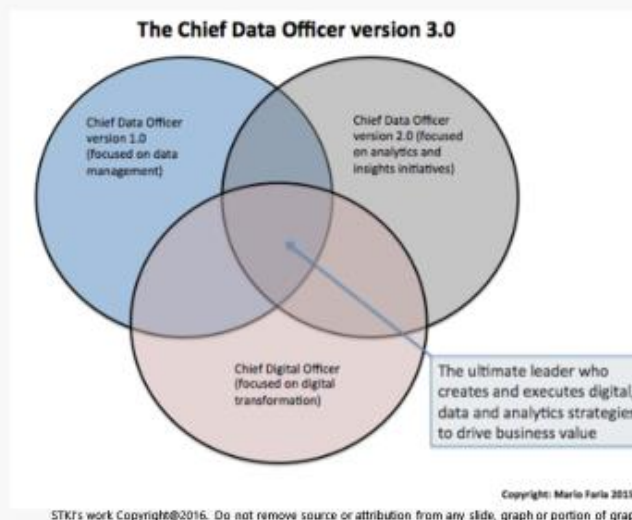


ואם מישהו מרגיש שלא שמענו על מספיק תפקידי מטה חדשים ברמת C-Level אל דאגה! ה CAOs (Chief Analytics Officers) בדרך. ותפקידם לוודא שאנליטיקה "תעלה מדרגה" ברמת השימוש שלה בקבלת ההחלטות בארגון. יש כאלה שיגידו שזה לא תפקיד נוסף אלא התקדמות בתפקידו של ה CDO (ה Chief Data Officer הארגוני) או פונקציה שקיימת במקום אחר בארגון, אבל מה שיותר חשוב הוא שמדובר על סוג של "אנליסט/ית בכיר/ה" שמקבל כיסא בחדר הדיונים של ההנהלה, היא/הוא מעורים במטרות ובאתגרים העסקיים ויודעים למנף את היכולות האנליטיות בארגון באופן אקטיבי ולפעול לכך שקבלת החלטות יומיומיות יישענו על תובנות אנליטיות, ולא על תחושות בטן, במיוחד ברמת ההנהלה הבכירה.

Method #2: Assign CAO (Chief Analytics Officer)



... Or is it CDO v2.0?



שיטה נוספת להידוק הלולאה בין Data – Insights – Action היא שימוש במתודולוגיית DataOps ששואבת השראה רבה מתחום ה DevOps. עצם העניין כאן הוא יצירת תצורת עבודה המבטיחה שיתופיות בין אנשי ה Data לאנשי ה Analytics ואנשי התהליכים, במקום שכל אחד מהם יפעל בנפרד.

Method #1: DataOps



DataOps bridges the gap between:

1. Those who collect/arrange the data
2. Those who analyze the data
3. And Those who will put the findings from that analysis to good use

It is all about fostering **collaboration**



STK's work Copyright©2016. Do not remove source or attribution from any slide, graph or portion of graph

15

הרבה שיטות לאותה המטרה, בדרך להפוך להיות Data Driven Business.

לסיכום:

תחום ה AI-כיום מראה תוצאות מיטביות בשילוב עם אנשים / Data Scientists / אנליסטים. אין לנו עדיין "אלגוריתמים אוטומטיים". הכלים כיום מסייעים לנו בהבנת הדאטה, ב"ראייה" של דפוסים ובמציאת קורלציות בדאטה.

החסמים העיקריים: הערך העסקי וה Use Cases עוד לא ברורים; מחסור משמעותי באנשים ובכישורים מתאימים בארגון; פלטפורמת דאטה לא מספיק "מוכנה" (אי אפשר להסתמך על הנתונים לקבלת החלטות "בשטח", אין מספיק "Governance" איכות נתונים נמוכה מדי, או תשתית נתונים מיושנת מדי).

יחד עם זאת, ברור שהפוטנציאל הוא עצום ושכדאי להתחיל לבחון וללמוד את הנושא, ללמד את האנליסטים בארגון ולהכשיר אותם לסביבות טכנולוגיות חדשות לחלוטין, לעתים נואלץ גם לגייס כ"א חיצוני שמגיע מתחום ה Data Science, לנהל את תשתית הדאטה ככזו

שאמורה לתמוך בתפעול היומיומי של הארגון ובקבלת החלטות קריטיות בזמן אמיתי / קרוב לכך.

התעלמות מתחום זה – יש בה סיכון של ממש, שכן האיומים החדשים על תעשיות הפכו להיות לא המתחרים הישירים שלהם אלא המתחרים הפוטנציאליים – אלה שיצליחו לאתר אלגוריתמים יותר חכמים ויעילים מהמערך הקיים בארגון כיום. מספיק לדמיין מה היה קורה אילו לאחרת מ"ענקיות הדאטה" היה מתחשק מחר בבוקר להיכנס כמתחרה בוורטיקל בו הארגון פועל, וזה תרחיש לחלוטין לא דמיוני.

אסור לשכוח שאת כל זה נושא ניהול הנתונים ו Data Governance חייב לעטוף. זהו תנאי מהותי להצלחה, בלעדיו לא ניתן יהיה לייצר תשתית אמינה וניתנת לניהול שנוכל להסתמך עליה בהחלטות התפעוליות של הארגון.

אז האם Data is the new Oil? לא בטוח, אנחנו יותר מתחברים להגדרה של David McCandless – Data is the new Soil (אדמה). באדמה צריך לשחק, להתלכלך, להתייחס לנושא לעתים כאל מגרש משחקים, להצטייד באנשים ובכלים שיכולים למצוא שם דברים בעלי ערך ולשתול את הזרעים למה שעתיד להשפיע מהותית על פעילות הארגון.