



הוראת מדעי הנתונים בכל תחום ידע באוניברסיטה

דוח שאושר במועצת האקדמיה ביום כ"ח בכסלו תשפ"א / 14 בדצמבר 2020

**הוראת מדעי הנתונים
בכל תחום ידע באוניברסיטה**

עריכת לשון: יהודית ידלין
עיצוב ועימוד: נאוי קצמן עיצוב גרפי
נדפס בדפוס כספית, ירושלים

© כל הזכויות שמורות לאקדמיה הלאומית הישראלית למדעים



הוראת מדעי הנתונים בכל תחום ידע באוניברסיטה

דוח שאושר במועצת האקדמיה ביום כ"ח בכסלו תשפ"א / 14 בדצמבר 2020

תוכן העניינים

7.....	תקציר הדוח
9.....	חשיבותם של מדעי הנתונים
10.....	הקמת הוועדה
10.....	מטרותיה של הוראת מדעי הנתונים בקמפוס
11.....	תוכני ההוראה והיקפם
12.....	אתיקה וחשיבה ביקורתית בנייתוח נתונים
12.....	המלצות ליישום גמיש בדיסציפלינות שונות
13.....	א. מרכיב הליבה
13.....	ב. מרכיבים דיסציפלינריים נוסף על מרכיב הליבה
13.....	שיתופי פעולה בהוראת מדעי הנתונים
15.....	הקדמה
19.....	פרק 1 חשיבותם של מדעי הנתונים
22.....	1.1 חשיבותם במדע
23.....	1.2 חשיבותם בחברה ובתעשייה
24.....	1.3 הסיבה לעליית הדיון כעת
26.....	1.4 הקמת הוועדה
26.....	1.5 המצב בארצות הברית (מסמכי NAS)
28.....	1.6 מדעי הנתונים בוועדת המועצה להשכלה גבוהה
29.....	1.7 מדעי הנתונים בוועדת ההיגוי של הו"ת
31.....	פרק 2 עיקרי הגישה המוצעת להוראת מדעי הנתונים בקמפוס
33.....	2.1 מטרות ההוראה
35.....	2.2 תוכני ההוראה והיקפם
35.....	2.2.1 שלבי מחזור הנתונים

38.....	2.2.2	ארגז הכלים שעומד לרשות כל שלב
39.....	2.2.3	אתיקה וחשיבה ביקורתית בניתוח נתונים
	2.2.4	הצורך להציג גישות שונות ולתת דגשים שונים
42.....		לדיסציפלינות שונות
42.....	2.2.5	המלצות כלליות
45.....		פרק 3 המלצות יישום בתחומי ידע שונים
47.....	3.1	מבוא
49.....	3.2	מדעי הרוח
51.....	3.3	מדעי החברה וניהול
53.....	3.4	משפטים
54.....	3.5	מדעי החיים ורפואה
56.....	3.6	מדעים מדויקים
58.....	3.7	הנדסה
61.....		פרק 4 המלצות בתחומים משיקים
64.....	4.1	שיטות הוראה ועקרונות בהוראת מדעי הנתונים
65.....	4.2	פיתוח קובצי נתונים המיוחדים לישראל והנגשתם
	4.2.1	משאבים למחקר במדעי הנתונים בשפות המקומיות:
65.....		עברית וערבית
66.....	4.2.2	נתוני משפט
67.....	4.2.3	מקורות לנתונים למדעי החברה
68.....	4.3	סיווגייה עם ספריות אקדמיות
70.....	4.4	מדעי הנתונים בחינוך הטרומ־אוניברסיטאי
74.....		רשימת הכותבים והיועצים

תקציר הדוח

תקציר הדוח

חשיבותם של מדעי הנתונים

מדעי הנתונים עוסקים באיסוף, בניהול, בעיבוד ובניתוח של נתונים. כמו כן הם עוסקים בהוצאת מסקנות, בתחזיות ובפיתוח כלים הנשענים על כל אלה. הנתונים מגיעים ממגוון רחב של דיסציפלינות אקדמיות ויישומים מסחריים. התפתחויות רבות תרמו בשנים האחרונות לזינוק במדעי הנתונים, ובהן השימוש העצום ברשתות חברתיות, פיתוחים בתחום נתוני עתק (big data) וחישובים הנעשים עליהם, אגירת נתונים בענן ולמידת מכונה על ענפיה השונים, ובכלל אלה ראייה ממוחשבת ועיבוד שפה טבעית. חשיבות מדעי הנתונים לחברה ולמדינה הולכת וגוברת. האינטראקציה בין המדינה לאזרחיה הופכת לדיגיטלית ויוצרת נתונים שאפשר לגזור מהם מסקנות וחיזויים. הנתונים נדרשים לעיתים לביטחון המדינה, למעקב אחר פשיעה ולהתמודדות עם מגפות. לתעשייה, ובייחוד לתעשיית ההיי-טק, יש דרישה גבוהה לכוח אדם איכותי במדעי הנתונים, והחוסר הקיים כעת צפוי לגדול.

הקמת הוועדה

חשיבותם של מדעי הנתונים הביאה את האקדמיה הישראלית למדעים להקים ועדה זו, על מנת להעריך את הצורך בהכשרה במדעי הנתונים בלימודים האקדמיים ולעמוד על הדרכים לעשות זאת. הוועדה סבורה שהאקדמיה הישראלית מחויבת להעניק לבוגריה כישורים בסיסיים במדעי הנתונים, כפי שהיא מחויבת להכיר להם גישות מדעיות שונות ולהכשירם לסגל לעצמם חשיבה ביקורתית. הוועדה ממליצה ללמד את מדעי הנתונים בכל היחידות האקדמיות בארץ אגב מתן גמישות לכל יחידה לממש את התוכנית לפי צרכיה האקדמיים ולפי אופייה של תוכנית הלימודים הנהוגה בה. המלצה זו היא הזדמנות לכל היחידות האקדמיות בארץ להתחדש ולגייס משאבים לתוכנית הלימודים המתחדשת.

מטרותיה של הוראת מדעי הנתונים בקמפוס

א. תלמידי תואר ראשון במדעי הנתונים

תוכנית כזו מתנהלת בדרך כלל באחד או יותר מן החוגים האלה: מדעי המחשב, סטטיסטיקה, הנדסת תעשייה וניהול, הנדסת חשמל, ניהול ומערכות מידע ומדעי המידע. הדוח הנוכחי איננו עוסק בקבוצה זו.

ב. תלמידי חוגים בעלי נגיעה במדעי הנתונים שאינם בתוכנית ייחודית

לקבוצה זו ניתן לצרף תלמידי חוגים שבהם הדרישות החישוביות והאנליטיות של התואר מהוות בסיס לעיסוק מקצועי בתחום. עבור תלמידים אלו התוכנית לתואר ראשון תשמש לשתי מטרות: א. לאפשר תעסוקה במגזר העסקי והציבורי בתפקידים הכרוכים בנייתוח מידע; ב. לאפשר מעבר ללימודים מתקדמים במדעי הנתונים בהשקעה מעטה יחסית.

ג. תלמידי תואר ראשון בכל יתר הקמפוס

מטרות הלימוד לתלמידים אלה הן א. הקניית יכולת לזהות צורך בנתונים ולהשתמש בהם לצורכי לימודים, מחקר ותעסוקה; ב. הקניית הבנה של היתרונות והמגבלות של השימוש בשיטות עתירות נתונים, ליד גישה ביקורתית לתוצאות; ג. פיתוח רגישות וביקורתיות לתהליכי מחקר נתונים ולבעיות של אתיקה העולות בתהליך זה.

תוכני ההוראה והיקפם

הדרך המקובלת לקבל תמונה כוללת על תוכנם של מדעי הנתונים ועל היקפם היא באמצעות "מחזור הנתונים" (the data cycle). מחזור הנתונים הוא מודל סכמתי המאפיין כל תהליך של מחקר או קבלת החלטות מבוסס־נתונים, והוא תהליך החוזר על עצמו.

בתרשים להלן מתואר מחזור הנתונים תיאור גראפי. מרכיביו מקובלים על כל העוסקים במדעי הנתונים, אך מספר השלבים בו יכול לנוע בין 6 ל-9, לפי ההגדרות השונות. כאן יוצג מבנה פשוט יחסית:



הדוח המלא מציג פירוט של הכלים המיושמים בכל אחד משלביו של מחזור הנתונים ואשר על התלמידים להכירם, ואת חלקם גם להפעיל. אנו מפרידים בין כלים מינימליים שראוי שכל סטודנט וסטודנטית בקמפוס יוכלו להפעילם לבין כלים מתקדמים יותר שנועדו לסטודנטים בעלי רקע מתאים.

אתיקה וחשיבה ביקורתית בניתוח נתונים

בתהליך ניתוח הנתונים עולות סוגיות אתיות הנוגעות לפרט ולחברה, וחלקן אף נושא השלכות משפטיות. ההכשרה במדעי הנתונים מחייבת מודעות לסוגיות אלה והכרת הדרכים להתמודד איתן. יש להעלות על הפרק את שאלת הפרטיות כשמדובר בנתונים אישיים כגון מידע על הליכים רפואיים או משפטיים. במהלך ההוראה יש לדון באיזון בין היתרון לחברה מגילוי מלא של הנתונים לבין הפגיעה האפשרית בפרט; בין היתרון שצוברות חברות מסחריות גדולות לבין האינטרסים של החברה ושל היחיד. יש לדון גם בשאלות של זכויות יוצרים. יש להקפיד על שקיפות התהליך ועל הנגשת הנתונים. ולבסוף, יש להצביע על החשיבות שבהתנהגות אתית של החוקרים אגב גילוי נאות של אינטרסים צולבים, ללא עיוות של התהליך וללא הטיות.

המלצות ליישום גמיש בדיסציפלינות שונות

הוועדה ממליצה שכל תלמיד בקמפוס ייחשף במהלך לימודיו לתואר ראשון לכל הנושאים שבמחזור הנתונים. ניתן לעשות זאת באמצעות פיתוח קורס מבוא למדעי הנתונים או בהעשרת קורסים קיימים בשיטות מחקר ובסטטיסטיקה, כדי שיקיפו את כל הנושאים שבמחזור הנתונים. ייתכן גם שילוב של שתי דרכים אלו.

הוועדה ממליצה על גמישות באימוץ ההמלצות של הדוח הנוכחי ביחידות האקדמיות השונות, לפי המשאבים העומדים לרשותן ולפי צורכיהן ואופי תוכנית הלימודים בכל אחת מהן. עם זאת ראוי שהמערכת תאפשר העמקה נוספת לסטודנטים מכל התחומים החפצים בכך. הוועדה ממליצה שכל סטודנט ילמד את מודל מחזור החיים של נתונים המתאים לתחום לימודו. מומלץ שהקורס הייעודי המוקדש למטרה זו יהיה מודולרי: תחילה ליבה עקרונית וגנרית המבוססת על מחזור חיי הנתונים ואחריה נושאים המותאמים לצרכים של תחומי הלימוד השונים.

א. מרכיב הליבה

היקף הקורס:

2 שעות הרצאה שבועיות, 2 שעות תרגול שבועיות

נושאי הקורס:

1. מחזור חיי הנתונים
2. סקירת כלים התומכים בשלבים השונים של מחזור חיי הנתונים
3. סוגי נתונים במרחב הדיגיטלי: מספריים, טקסטואליים, מובנים ולא-מובנים (structured, unstructured)
4. יישום בסיסי של איסוף נתונים, אינטגרציה של נתונים ממקורות שונים, ניתוח הנתונים, למידה אלגוריתמית וויזואליזציה של נתונים
5. חשיבה ביקורתית לאורך מחזור חיי הנתונים
6. שיתוף נתונים (open data, data repositories) והכללים לציטוט אקדמי שלהם

ב. מרכיבים דיסציפלינריים נוסף על מרכיב הליבה

הדוח מציג דוגמאות של סילבוס לקורס המבוא עבור הדיסציפלינות האלה: מדעי הרוח, מדעי החברה וניהול, משפטים, מדעי החיים ורפואה, מדעים מדויקים והנדסה.

שיתופי פעולה בהוראת מדעי הנתונים

מרכזי מחקר אוניברסיטאיים למדעי הנתונים: בעקבות הצעדים שנקטו מל"ג-ות"ת קמו באוניברסיטאות המחקר בישראל מרכזי מחקר למדעי הנתונים. הצלחתה של היוזמה הנכוחית תלויה במידה רבה בשיתוף פעולה עם המרכזים האוניברסיטאיים האלה.

פיתוח קובצי נתונים המיוחדים לישראל והנגשתם: הוראה של מדעי הנתונים כפי שהומלץ בדוח זה מחייבת קיומם של משאבים מספיקים בכל אחד מהתחומים. יצירת המשאבים האלה היא חובה לאומית, כחלק מכינון של תשתיות לאומיות אחרות למדע. הדוח דן במשאבים בשפות המקומיות: עברית וערבית; בנתוני משפט; בנתונים ממדעי החברה. כל אחד מן הסעיפים סוקר את המצב הקיים ומציע כיוונים לפיתוח עתידי.

סינוגייה עם ספריות אקדמיות: שירותי הספריות כוללים פיתוח והנגשה של תשתיות מידע ותמיכה בתהליכי לימוד ומחקר. צמיחה מהירה במיוחד חלה בעשור האחרון בתחום מדע הנתונים ובתחום נתוני העתק, ועל הספריות לתת לה מענה. דרישה זו היא הזדמנות לספריות לפתוח בתנופה חדשה של עשייה וחדשנות. הדוח ממליץ לספריות להקים מרכזי הדרכה במדעי הנתונים ולפתח תשתיות פיזיות ומתחמים של ציוד מתקדם לפי צורכי המחקר וההוראה.

מדעי הנתונים בחינוך הטרומ־אוניברסיטאי: הדרך הנכונה להטמיע יכולות עבודה עם נתונים איננה בשיעורים ייעודיים לכך אלא בעבודה יום־יומית של שימוש בנתונים בכל אחד מן המקצועות הנלמדים בבתי הספר, מהיסטוריה ועד מתמטיקה, ובכל הגילים. בתי הספר שעברו ללמידה באמצעות פרויקטים רב־תחומיים יכללו בהם מיומנויות של מדעי הנתונים. כמו כל מיומנויות תשתית אחרות, אין להותיר את השימוש בנתונים ללימוד מרוכז בבית הספר התיכון לקראת בחינות הבגרות. לשם הטמעת מדעי הנתונים במערכת החינוך יש להכשיר מורים בנושא זה, הן אלה שהם כבר חלק מן המערכת והן אלה שעוברים כעת את ההכשרה במכללות למורים. משרד החינוך הקים ועדת מקצוע לנושא מידע ונתונים אשר מפתחת בימים אלה התמחות בבתי ספר תיכוניים. מומלץ לשתף פעולה עם ועדה זו.

הקדמה

הקדמה

בחודש דצמבר 2018 מינתה פרופ' נילי כהן, נשיאת האקדמיה הלאומית למדעים, ועדה לבחינת האפשרות לשלב לימוד מדעי הנתונים בכל תחום ידע באוניברסיטה. עם חברי הוועדה נמנו פרופ' ניב אחיטוב (אוניברסיטת תל אביב), פרופ' יונתן בן־דב (אוניברסיטת חיפה ואוניברסיטת תל אביב – מרכז הוועדה), פרופ' יואב בנימיני (אוניברסיטת תל אביב וחבר האקדמיה למדעים), פרופ' יגאל ברונר (האוניברסיטה העברית בירושלים), פרופ' ידין דודאי (מכון ויצמן למדע, חבר האקדמיה למדעים, ומשמש בה יו"ר החטיבה למדעי הטבע), פרופ' דפנה רבן (אוניברסיטת חיפה) ופרופ' רווד שרן (אוניברסיטת תל אביב).

הוועדה התכנסה חמש פעמים בין החודשים ינואר 2019 וינואר 2020. המפגשים כללו גם פגישות עם מומחים חיצוניים: פרופ' אביגדור גל (הטכניון – מכון טכנולוגי לישראל), ד"ר רננה קידר (האוניברסיטה העברית בירושלים), ד"ר יוסי בן־דב (בית הספר הריאלי העברי בחיפה), ד"ר אבי כהן (משרד החינוך) וגב' רונית נחמיה (משרד החינוך). חברי הוועדה, שנחלקו לתת־צוותים לפי תחומי עיסוקם, התייעצו עם מומחים נוספים בתחומם, כפי שמפורט ברשימת הכותבים והיועצים להלן.

גב' סימה דניאל ניהלה ביעילות את עבודת הוועדה בכל שלביה, וגב' גליה פינצי, מנכ"לית האקדמיה, ליוותה בעצה ובתמיכה. על כך נתונה לשתיהן תודתנו.

פרק 1

חשיבותם של מדעי הנתונים

חשיבותם של מדעי הנתונים

מדעי הנתונים עוסקים באיסוף נתונים, באחסונם ובניתוחם, בהתייחס להיבטים אנושיים וחברתיים. הנתונים משמשים לחיזוי ידע, ליצירת ידע ולמיונו. מדעי הנתונים עוסקים גם בפיתוח כלים הנשענים על הנתונים הנאספים. הנתונים מגיעים ממגוון רחב של דיסציפלינות אקדמיות ויישומים מסחריים. התהליך שהנתונים עוברים מביא לידי ביטוי את מקורם המיוחד של הנתונים ואת השימוש המיועד למסקנות שיעלו מהם. התפתחויות רבות תרמו בשנים האחרונות לזינוק במדעי הנתונים, ובהן פיתוחים בתחום בסיסי הנתונים וחישובים הנעשים עליהם ולמידת מכונה על ענפיה השונים, ובכלל אלה ראייה ממוחשבת ועיבוד שפה טבעית.

אנו נסמכים בדוח על מודל "מחזור החיים" של השימוש בנתונים, שהוא תהליך גנרי שנגדירו ונפרטו להלן בסעיף 2.2.1. איננו משייכים את מדעי הנתונים לדיסציפלינה מסוימת או לדיסציפלינות מסוימות, שכן השימוש בנתונים נרחב, וגם ההתפתחויות המתודולוגיות נעשות (לפעמים בד בבד) בדיסציפלינות שונות. הסטטיסטיקה על ענפיה השונים עוסקת כולה במחזור חיים זה. מדעי המחשב לתחומיהם השונים עוסקים גם הם במחזור זה, בשימת דגש על למידת מכונה, על בינה מלאכותית ועל חישוב במערכות מבוזרות. הדיסציפלינה של מערכות מידע, שמופיעה לעיתים כדיסציפלינה עצמאית ולעיתים כחלק מניהול, עוסקת כולה בתחום. חלקים מן התהליך מוצאים ביטוי בהנדסת תעשייה וניהול שבבתי הספר להנדסה, ואילו

התחום של עיבוד אותות, השייך היסטורית להנדסת חשמל, עוסק ישירות במדעי הנתונים. גם במתמטיקה שימושית הדגש על נושאים במדעי הנתונים התחזק. נוסף על זה, אנו רואים מקצועות אינטרדיסציפלינריים, כמו ביואינפורמטיקה ומדעי המוח, שלא רק משתמשים במדעי הנתונים אלא אף מקדמים אותם.

סקירה קצרה זו מבהירה שמדעי הנתונים אינם עיסוק בין-תחומי, במובן זה שהוא חיתוך של תחומים מדעיים, כי אם נושא רב-תחומי. בשלב זה קיימת אי-בהירות באשר לעתיד מדעי הנתונים – אם יתפתחו לדיסציפלינה שלמה העומדת בפני עצמה, כפי שמסתמן מהמאמצים הנעשים בחמש השנים האחרונות בארץ ובעולם לפתח תוכניות לתואר ראשון בתחום, או שישארו מפוזרים בין תחומי ידע רבים. הדיונים בנושא מרובים, אך לענייננו בדוח זה אין חשיבות להכרעה בכך. סביר להניח ששני התהליכים יתקיימו זה לצד זה.

1.1 חשיבותם במדע

על חשיבותם של מדעי הנתונים במדעים הניסויים אין צורך להכביר מילים. מאז המצאתם של מכשירי המדידה, כמד-חום ומד-לחץ, היה האיסוף השיטתי של נתונים, ניתוחם והוצאת מסקנות מתוכם חלק מרכזי במתודולוגיה המדעית. מתודולוגיה זו קיבלה את צורתה הנוכחית בהשפעתו של רוברט בויל עם תחילת תקופת הנאורות. העלאת שאלה מדעית, תכנון הניסוי, ביצועו המתועד, ניתוחו בשיטות סטטיסטיות, הוצאת מסקנות מתוכו, שיתוף הקהילה המדעית בתוצאותיו והצגת התובנות העולות ממנו באמצעות טקסט וויזואליזציה – כל אלה משותפים לכל המדעים הניסויים. את המחקר העכשווי ניתן לכוון "עתיר נתונים". כך למשל במחקר הרפואי, על העבודה הניסויית שבו, בשלבים השונים של פיתוח מושכל של טיפולים ותרופות. המחקר האפידמיולוגי הנוכחי למשל, הקושר את הגנומיקה, את הסביבה ואת ההתנהגות עם התחלואה והמתקדם לעבר רפואה מותאמת אישית, הוא מחקר עתיר נתונים. גם המחקר הפסיכולוגי הניסויי נעשה ברובו בשיטות אלו. לכן לא מפתיע שחלק ניכר מהכשרתם של תלמידי תואר ראשון בתחומים הניסויים הוא בנושאים סטטיסטיים, שהם בליבת מדעי הנתונים. גם במדעי החברה והכלכלה יש עיסוק רב בשיטות הנשענות על ניתוח נתונים, אם כי פחות מבאלה הנשענות על ניסויים.

במדעי הרוח ובחקר המשפט העיסוק בנתונים דיגיטליים עבר מהפכה. הענף המכונה "מדעי הרוח הדיגיטליים" יצר שפע של מידע שניתן לשבץ באלגוריתמים של למידה על מנת להפיק תובנות מחקריות חדשות. נתונים אלה כוללים כמות רבה של מידע ויזואלי, כגון תצלומים מתקדמים של כתבי יד, יצירות אומנות ועיתונות היסטורית. לצד אלה נמצא כמובן מידע טקסטואלי רב שעבר דיגיטציה בעשור האחרון, והוא צופן בחובו שפע של נתונים חשובים. פרויקטים שונים בלימודים קלאסיים, בארכאולוגיה, בפפירולוגיה, במדעי היהדות ובהיסטוריה אזורית ברחבי העולם כבר מפעילים למידה כזו בהצלחה. למידת מכונה כבר מסוגלת ליצור קריאה אוטומטית של כתבי יד במבחר שפות העולם, והיא העשירה במידה ניכרת את מאגר הנתונים הזמין לקריאה ולמחקר. כך למשל התפתח ניתוח גאוגרפי של הנתונים וקישורם לרשתות חברתיות, לשם חקר התכתובת בין הוגים וכן בין דמויות היסטוריות שונות.

1.2 חשיבותם בחברה ובתעשייה

חשיבותם של מדעי הנתונים לחברה ולמדינה הולכת וגוברת. מערכת קבלת ההחלטות במדינה, שתמיד הייתה מבוססת על נתונים סטטיסטיים, הולכת ומתעדכנת בעוד ועוד מידע הנאסף ממקורות שונים. נתונים מקרר-כלכליים, נתוני עסקות ועסקים, נתוני בריאות ועוד – כולם משמשים לפעילות היום-יומית של המדינה, של גופי הביטחון והמשטרה ושל תאגידים גדולים כגון בנקים, רשויות מקומיות ועוד. בכל אלה הדרישה לכוח אדם מקצועי בתחום היא רבה. יתר על כן, האינטראקציה בין המדינה לאזרחיה הופכת לדיגיטלית ויוצרת נתונים שמהם אפשר לגזור מסקנות וחיוזיים. נתונים אלה חשובים לעיתים לביטחון המדינה כנגד אויביה ולמעקב אחר פשיעה, אבל גם להתמודדות עם מגפות. ככל שהשימוש בכלים אלה נעשה נפוץ יותר, כן הוא דורש הבנה רבה יותר של יכולותיהם, יתרונותיהם ומגבלותיהם גם ממי שאינו מדען.

בתעשייה, ובייחוד בתעשיית ההיי־טק, יש דרישה גבוהה לכוח אדם מקצועי בעל יכולות גבוהות במדעי הנתונים.

שינוי מהותי נוסף שאירע בעשור האחרון הוא שהפרטים הפכו להיות יוצרי הנתונים הגדולים ביותר. כמעט כל איש ואישה משאירים אחריהם שובל של נתונים שנאספים, נשמרים וניתנים לניתוח: נתוני מיקום מהטלפון, נתוני התקשרויות, דואר אלקטרוני, שימוש ברשתות חברתיות ולאחרונה גם נתוני התנהגות בשיחות וידאו וברשתות השונות. נתונים אלו חשובים לענקי המידע ולארגונים כלכליים גדולים. הם מנתחים אותם ומשתמשים בתוצאות הניתוחים כדי להשפיע על הפרט ועל התנהגותו הכלכלית והפוליטית. מצב זה הופך כל אזרח ואזרח בחברה המודרנית למי שמשפיע על פעילותם של מדעני נתונים ומושפע ממנה.

1.3 הסיבה לעליית הדין כעת

הדין העכשווי והצורך בפעולת הוועדה עלו בעקבות שתי התפתחויות שחלו זו לצד זו:

א. אינטנסיביות הולכת וגוברת של שימוש במדעי הנתונים בדיסציפלינות מדעיות רבות, ממדעים מדויקים ומדעי הטבע, לאורך כל ה"קמפוס" ועד למדעי הרוח ומדעי החברה.

ב. ניצול גובר והולך של נתוני עתק (big data) ועיבוד מתוחכם של נתונים בסקטור העסקי והציבורי, דבר שמחייב הכשרת כוח אדם לניהול מידע ולניתוחו.

ניגע תחילה בקצרה בצרכים המדעיים ולאחר מכן בצרכים בתעשייה.

במחקר המדעי התפתחו מדעי הנתונים תחת מטריית הסטטיסטיקה במשך יותר ממאתיים שנה. מאז שנות השבעים של המאה העשרים חלו שינויים במדעי הסטטיסטיקה ובמדעי המחשב, שזכו לשמות שונים אך כיוונם אחד. בסטטיסטיקה הושם הדגש מחדש על ניתוח נתונים כמטרה בפני עצמה בלי להידרש גם לדיון מתמטי, ובמדעי המחשב, כעשור מאוחר יותר, החלו לפתח אלגוריתמים מיוחדים לפתרון בעיות עתירות נתונים, בשימת דגש על מטרות מיון וחיזוי. בתוך זמן מה קיבל התחום הזה במדעי המחשב את השם "למידת מכונה". נושאי המחקר בשני התחומים היו דומים אבל שמרו כל אחד על טרמינולוגיה שונה. השיפורים ביכולתם של האלגוריתמים נעשו עקב בצד אגודל ובהתמדה.

בשני העשורים האחרונים השתלבה מגמת ההאצה במהירות המחשוב בשינוי נוסף ביכולות המחשוב: מחיר אחסון הנתונים ירד במידה ניכרת. כבר לפני עשרים שנה ירד המחיר עד כדי כך שהיה אפשר לשמור נתונים שנאספו לצרכים מסוימים, כמו למשל הנהלת חשבונות או מעקב אחר חולים לצורך הטיפול בהם, ואחר כך לנתח אותם למטרות אחרות. תחום זה, שנקרא כריית מידע (data mining), הוסיף להתפתח ולגדול, במדעי הנתונים שנאספו ובמספרם, אגב שימוש בטכנולוגיית "אחסון בענן", עד שזכה לכינוי "נתוני עתק" (big data). בד בבד חלו פיתוחים בתוכנות, אגב שימוש בקוד פתוח, שהקלו את יכולת התכנות בקובצי נתונים גדולים כאלה, הפזורים על פני מקומות ושיטות אחסון שונות. הרכיב האחרון היה הפיתוח האלגוריתמי שנקרא "למידה עמוקה" (deep learning), שקידם במידה ניכרת יכולות שונות, כגון זיהוי תמונות, זיהוי קולות, תרגום, זיהוי תנועה (למשל מכשולים בנהיגה) ומספר רב של מטרות חדשות. מטרות אלו, שמדמות בינה אנושית, זכו לשם "בינה מלאכותית".

הבינה המלאכותית הייתה נושא מאתגר עוד בטרם נבנה המחשב האלקטרוני הראשון, והעסיקה חוקרים רבים בשלבים הראשונים של מדעי המחשב. למידה עמוקה, שהשיגה קרובים הרבה יותר למטרה הנכספת, משתמשת בפרדיגמה של מדעי הנתונים: בעזרת נתונים רלוונטיים רבים וכושר מחשוב גבוה יקבע מודל שיבצע את המשימה כהלכה אגב ויתור על הבנת הדרך שבה המודל משיג את המטרה.

בתעשייה ההישגים במדעי הנתונים עומדים מאחורי הזינוק בתחום הבינה המלאכותית ושימושיה במגזר העסקי והציבורי. החסך בכוח אדם בתעשיית ההיי־טק, בין בחברות גדולות כאינטל, מיקרוסופט, גוגל ודומיהן, ובין בחברות הזנק צעירות, הוא מכשול רציני בפני התפתחות נוספת בתחום. אחד המקצועות המבוקשים ביותר כיום בהיי־טק ובחברות שיווק בין־לאומיות מכונה data analyst, שהוא שם גנרי לכל עובד מקצועי שיועד לאתר ולהנגיש נתונים רלוונטיים, לכרות אותם, להפיק מהם תובנות ולהציגם (ראו: **מחזור הנתונים** להלן בסעיף 2.2). עובדים בתחום זה חייבים לקבל את הכשרתם בתוכנית אקדמית.

1.4 הקמת הוועדה

חשיבותם העולה של מדעי הנתונים כמעט בכל דיסציפלינה מדעית וחשיבותם בתעשייה, בביטחון ובמשל העלו את הצורך לבחון את הדרך שבה ניתן להורות את מדעי הנתונים לתואר ראשון לכל רוחב הקמפוס. בחינה שכזו צריכה להעריך את סוגו ואת היקפו של הידע שכדאי ללמד בדיסציפלינות השונות ואת הכישורים שיש להעניק לסטודנטים, ולהצביע על הדרכים להשיג זאת. האקדמיה הישראלית מחויבת להעניק לכל בוגריה כישורים בסיסיים במדעי הנתונים, כפי שהיא מחויבת להכשירם להכיר גישות מדעיות שונות ולסגל לעצמם חשיבה ביקורתית. מטרת אלו הביאו את האקדמיה הישראלית למדעים להקים ועדה זו, הפועלת לצד יוזמות פעילות מרובות באוניברסיטאות השונות ובמוסדות המדיניות האקדמיים בישראל. שילוב מדעי הנתונים בהוראה בכל תחום אוניברסיטאי מהווה הזדמנות עבור כל היחידות האקדמיות בארץ להתחדש ולגייס משאבים לתוכנית הלימודים.

1.5 המצב בארצות הברית (מסמכי NAS)

בשנת 2017 הקימו האקדמיות המדעיות הלאומיות בארצות הברית למדעים, להנדסה ולרפואה ועדה שנשאה את השם: Data Science for Undergraduates: Opportunities and Options

מטרותיה של הוועדה הוגדרו כלהלן:

A National Academies of Sciences, Engineering, and Medicine study will set forth a vision for the emerging discipline of data science at the undergraduate level. It will emphasize core underlying principles, intellectual content, and pedagogical issues specific to data science, including core concepts that distinguish it from neighboring disciplines. It will focus on the undergraduate level, addressing related issues at the middle and high school as well as community colleges as appropriate, and will draw on experiences in creating master's-level programs. הדוח התפרסם בספר ב־2018.

The National Academies Press, ISBN 978-0-309-47559-4 DOI 10.17226/25104

זוהו בו מיומנויות היסוד הדרושות לעוסקים בתחום, שנידונו בפירוט בדוח: Data description and curation, mathematical foundations, computational thinking, statistical thinking, data modeling, communication, reproducibility, ethics, and computational skills.

הדוח עסק גם בדרכים לשלב הקניית מיומנויות יסוד אלה במהלך הרגיל של הלימודים ובקורסים נפרדים מתוך ההכרה שבוגרים בתחומים שונים ימצאו עצמם עוסקים במדעי הנתונים בתחומם, ומתוך הבנת החשיבות בהעמקה שתאפשר לבוגרים להמשיך ולהתפתח. הודגשו שיטות הלמידה, ובהן סמינרים כבר בשנה הראשונה ללימודים, קורס ליבה בסיסי משותף, יצירת קהילות לומדות, קורסים עתירי כתיבה, מטלות ופרויקטים משותפים, לימוד באמצעות שירות לקהילה ומסלולי הכשרה בתעשייה (internship).

יש לזכור שמערכת החינוך הגבוה בארצות הברית היא בת ארבע שנים, והיא כוללת ואינה מופרדת למקצועות (למעט הנדסה), ולפיכך יישום המלצותיה נעשה במסלולי הדגש (major), לעומת ההשכלה הגבוהה בארץ, המוכוונת כבר מהתואר הראשון. לפיכך הדוח האמריקאי עומד על חשיבות ההכשרה בתחום מדעי הנתונים כהכשרת יסוד חיונית לכל בוגר אוניברסיטה, בכל תחום שיבחר לעסוק בו בעתיד.

כמה המלצות בדוח האמריקאי רלוונטיות גם למערכת הישראלית: 1. על המוסדות האקדמיים לפתח טווח של מסלולים חינוכיים על מנת להכשיר סטודנטים למשרות בתחום מדעי הנתונים; 2. כדי להכין סטודנטים לעולם הנשען על נתונים, על המוסדות להקנות הבנה בסיסית בתחום לכל הסטודנטים; 3. ככל שתוכניות במדעי הנתונים מתפתחות, כן הסטודנטים שעליהן לנסות ולמשוך צריכים להיות בעלי רקע אקדמי רחב יותר; 4. מאחר שאנו בשלבים המוקדמים של התפתחות התחום, על המוסדות האקדמיים להיות מוכנים לשינויים שיחולו עם הזמן, לגלות גמישות ולפתח מערכת שתעודד שיתוף של תכנים ושיטות בין המוסדות.

1.6 מדעי הנתונים בוועדת המועצה להשכלה גבוהה

המועצה להשכלה גבוהה (להלן: המל"ג) הקימה ועדה שבסמכותה לאשר תוכניות לתואר ראשון במדעי הנתונים. הוועדה נאלצה להגדיר לראשונה מהם תחומי הידע שאותם יש להקיף בתואר ראשון במדעי הנתונים, ולשם כך פרסמה קווים מנחים. לדעת הוועדה, נוסף על ראייה בנתונים כלי לניתוח או משאב שיש לארגנו ולנהלו, יש גם להבין את מחזור החיים השלם שלהם, החל מאיסופם, דרך ניהולם וניתוחם וכלה בהנגשתם (דרך ויזואליזציה למשל ושימוש בהנמקות למשתנים מסבירים [features]). מחזור חיים זה ניזון מהתקדמות טכנולוגית בכל התחומים, כולל internet of things, מחשוב ענן, סטטיסטיקה חישובית, תורת המשחקים האלגוריתמית, עיבוד שפה טבעית וראייה ממוחשבת. תוכנית לימודים בהנדסת נתונים צריכה לחנך את הסטודנט מיזמו הראשון לחשיבה באמצעות נתונים, להבנת אופיים (נכונותם ושלמותם), את דרך התפתחותם על פני זמן, את ההטיות הנגזרות מאיסופם (הן דרך סנסורים והן מאנשים) ועוד. הכשרת מדען נתונים ברמת תואר ראשון דורשת מיומנויות של עבודה עם נתונים רבים בכל שלבי מחזור החיים אגב שימוש באשכול מחשוב (cluster) ובמשאבי ענן מגוונים.

הוועדה הסתמכה על תואר ארבע־שנתי שאושר לראשונה בטכניון – מכון טכנולוגי לישראל, ובפניה עמדה השאלה באיזו מתכונת ליצור את התואר הראשון בתחום זה – חד־חוגית או דו־חוגית בשילוב חוג נוסף. הוועדה הדגישה את הצורך להקנות בתואר הראשון הכשרה מתמטית בסיסית, חישוב אלגוריתמיקה ואופטימיזציה, אי־ודאות, סטטיסטיקה ולמידת מכונה. הודגש כי כל זכאותה של תוכנית להשתמש במדעי הנתונים להעניק תואר תיבחן לפי ארבעה מאפיינים: שלבי מחזור החיים של הנתונים הנלמדים, רמת הקרבה לנתונים גולמיים, סוגי הנתונים שאיתם מתמודדים וישומים קונקרטיים.

נכון לספטמבר 2020 תוכניות לתואר הראשון במדעי הנתונים והנדסת נתונים אשר אושרו במל"ג קיימות באוניברסיטת בן־גוריון בנגב, בטכניון – מכון טכנולוגי לישראל, באוניברסיטת חיפה ובמכללה למינהל. תוכניות רבות אחרות מציעות התמחות במדעי הנתונים.

1.7 מדעי הנתונים בוועדת ההיגוי של הו"ת

מל"ג-ות"ת הקימו את הוועדה על מנת לרכז את המאמצים בפיתוח המחקר במדעי הנתונים. עם מטרותיה העיקריות נמנו תחימה מדעית של מדעי הנתונים ומיפויים ואפיון הצרכים המחקריים של הקהילה האקדמית הרלוונטית בארץ ותייעודם. דוח הוועדה פורסם **כאן**. מסקנותיה העיקריות הרלוונטיות לדוח הנוכחי, העוסק בהוראה, הן כלהלן:

תחומי הליבה של המחקר במדעי הנתונים (מסודרים לפי מחזור החיים של נתונים) הם אלה: שיטות לאיסוף נתונים באמצעות ניסויים, דגימה, כרייה ומנועי חיפוש; שיטות לאחסון נתונים, להנגשתם, לניתוח וחישוב (מבוזר) במערכות של בסיסי נתונים/מידע ושל נתוני עתק; פיתוח מתודולוגיות לניתוח נתונים, כגון רשתות עצביות ולמידה עמוקה, למידה בסיוע חיזוקים ובקרה מסתגלת, למידה סטטיסטית והסקה סטטיסטית; בניית אלגוריתמים ומודלים למיון ולחיזוי, לתיקופם של הנתונים ולניצולם ליצירת ידע ולהנגשתו (ויזואליזציה); התייחסות לנתונים מיוחדים כגון ניתוח תמונות וראייה ממוחשבת, עיבוד שפה טבעית, דיבור ושמיעה; ניתוח רשתות, נתונים אורכיים, נתונים עיתיים ומרחביים ונתוני הישרדות – כל אלה בהתייחס להיבטים אתיים הנוגעים לפרט, כגון פרטיות וחסיון, ולחברה, כגון הדירות ותקפות השימוש בתוצאות.

למדעי הנתונים יש נגיעה כמעט בכל תחומי התוכן של המרחב האקדמי, ובהם כאלה שיש להם זיקה עמוקה לתחומי הליבה ואפילו חפיפה חלקית עימם, וניתן להתייחס אליהם כאל מעטפת התחום. **תחומי המעטפת** הם בינה מלאכותית (החלק שאינו מכוסה בתחומי הליבה), אופטימיזציה, חקר ביצועים, רובוטיקה, תורת המשחקים, תהליכים סטוכסטיים, אינפורמציה, בקרה, ניתוח אותות, ביראינפורמטיקה, ביולוגיה חישובית, רפואה דיגיטלית/מותאמת-אישית, אפידמיולוגיה, אקונומטריקה, פסיכומטריקה, סייבר, חישוב קוונטי, בנייה וניהול של בסיסי נתונים/מידע. תחומים אלו טומנים בחובם פוטנציאל חדשנות ופריצת דרך לא רק במדעים המדויקים אלא גם במדעי החברה והרוח.

גיבוש חזון ארוך-טווח: יש לקדם את מעמדה ואת יכולותיה המחקריות של האקדמיה במדינת ישראל במדעי הנתונים, הן בתחומי הליבה והן בתחומי

המעטפת. יש להניח תשתית אקדמית ויסודות לשיתופי פעולה, הן בין חוקרים מדיסציפלינות אקדמיות שונות והן בין האקדמיה למגזר הציבורי ולתעשייה, לשם הבטחת שגשוגה של מדינת ישראל בתחומים הנשענים על מדעי הנתונים. הוועדה תקדם שיתופי פעולה בתחומי הליבה על מנת לאפשר התפתחות התחום לדיסציפלינה עצמאית, אגב גיבוש מסה קריטית שתאפשר הובלה בין-לאומית בכיוונים פורצי דרך.

מרכזי מחקר ותפקידיהם: מרכזי-גג מוסדיים יתכללו את הפעילות בתחום בכל מוסד ויאפשרו יצירת מסה קריטית. המרכזים ישולבו במסגרת מוסד על-לאומי (Israel Data Sciences Initiative) שתפקידו לסנכרן את פעילות המרכזים המוסדיים וליזום שיתופי פעולה עם גופי ממשלה, עם ארגוני חוץ ועם ענקי התעשייה. המרכזים עשויים להקצות משאבים לפיתוח קורסים על מנת להרחיב את הכשרת הסטודנטים במדעי הנתונים.

הכשרת סטודנטים: הוועדה מעוניינת בהגדלת מספר הסטודנטים לתארים מתקדמים, ובמיוחד לתואר שלישי, הן לצורך השוק הפרטי והן למגזר האקדמי, מתוך הבנה שמעסיקים מעוניינים בתלמידים שהתנסו בביצוע עבודות מחקר עצמאיות, ועל מנת לתת מענה למחסור באנשי סגל בשנים הקרובות. הרחבת ההכשרה במדעי הנתונים נדרשת לכל רוחב האוניברסיטה על מנת לחשוף תלמידים מחוגים שאינם עוסקים בליבת התחום לנושא, ובכך להגדיל את סיכויי ההשתלבות שלהם בחברות ובארגונים המשתמשים בנתונים שימוש אינטנסיבי.

נוסף על הדוח הנוכחי האקדמיה הלאומית למדעים וות"ת משתתפות בפורום תל"מ, פורום וולונטרי העוסק באיגום משאבים לתשתיות מחקר לאומיות. דוח של הפורום על תשתית מחקר לאומית בתחום הבינה המלאכותית מצוי עתה בשלבים הראשונים של יישומו. מסקנותיו של דוח זה ישפיעו גם על המשאבים הקיימים בתחום שבו עוסקת הוועדה הנוכחית.

עיקרי הגישה
המוצעת להוראת
מדעי הנתונים
בקמפוס

עיקרי הגישה המוצעת להוראת מדעי הנתונים בקמפוס

2.1 מטרות ההוראה

חשיבותה של הוראת מדעי הנתונים כבר נדונה בפרק הקודם, ולפיכך נעבור למטרות השונות של הוראתם בקמפוס. לשם כך נחלק את אוכלוסיית הקמפוס לשלוש קבוצות:

א. תלמידי תואר ראשון במדעי הנתונים

תוכנית כזו מתנהלת בדרך כלל באחד מהחוגים האלה או בשילוב של כמה חוגים מהם: מדעי המחשב, סטטיסטיקה, הנדסת תעשייה וניהול, הנדסת חשמל, ניהול ומדעי המידע. ייתכן שכחלק מההתפתחות הטבעית יוקמו עם הזמן מחלקות נפרדות למדעי הנתונים. על כל פנים, הדוח הנוכחי איננו עוסק בקבוצה זו, שכן עבורה הוקמה ועדה של המל"ג (לעיל).

ב. תלמידי חוגי הליבה או המעטפת שאינם בתוכנית ייחודית למדעי הנתונים

מטבע הדברים חלק מהתלמידים הללו ירצה להתמקצע במדעי הנתונים בעבודתם או בלימודים לתואר גבוה במדעי הנתונים. לקבוצה זו ניתן לצרף תלמידי חוגים שהדרישות החישוביות והאנליטיות של התואר בהם – במתמטיקה, בחישוב ובסטטיסטיקה והסתברות – מהוות בסיס לעיסוק עתידי מקצועי בתחום. חלקם יהיה ממדעים מדויקים, כמו מתמטיקה, פיזיקה והנדסות שונות, וחלקם יכול להיות ממדעי החברה, כגון כלכלה, ניהול ופסיכולוגיה, וממדעי החיים וכן מרפואה (לאלה ממדעי החברה וממדעי החיים תידרשנה ככל הנראה השלמות חישוביות ואנליטיות כמתואר לעיל). עבור תלמידים אלו התוכנית לתואר ראשון תשמש לשתי מטרות:

1. לאפשר להם תעסוקה מקצועית במגזר העסקי והציבורי בתפקידים הכרוכים בניתוח מידע מנתונים
2. לאפשר להם מעבר ללימודים מעמיקים לתארים מתקדמים במדעי הנתונים בהשקעה קטנה יחסית

מטרות אלה יאפשרו לתלמידים לעסוק במחקר באקדמיה או לעבוד בתעשייה במקצועות הדורשים מיומנויות גבוהות מאוד, כפי שנדרש במכוני מחקר או ביחידות מו"פ של חברות טכנולוגיות מידע מתקדמות.

ג. תלמידי תואר ראשון בכל יתר הקמפוס

קבוצה זו כוללת את תלמידי מדעי הרוח, תלמידי משפטים ואת החלק הנותר ממדעי החברה. עבור תלמידים אלו המטרות הן אלה:

1. הקניית יכולת לזהות צורך בנתונים ולהשתמש בהם לצורכי לימודים ומחקר בתחומי הלימוד שלהם, או לאחר הלימודים – בעיסוקיהם; הקניית היכולת לזהות את מקורות הנתונים האפשריים, ליצור או לאחזר אותם, לנתחם ולהציג אותם ולשתף בהם.
2. הקניית הבנה של היתרונות ושל המגבלות בשימוש בשיטות עתירות נתונים אגב הקניית יכולת גישה ביקורתית לתוצאות.
3. פיתוח רגישות וביקורתיות לבעיות של אתיקה בנוגע לחברה ולפרט, העולות בעת ניתוח נתונים, והשימוש בתוצאותיו.

2.2 תוכני ההוראה והיקפם

הדרך המקובלת לקבל תמונה כוללת של מדעי הנתונים היא באמצעות "מחזור הנתונים" (the data cycle). מחזור הנתונים הוא מודל סכמתי המאפיין כל תהליך של מחקר או קבלת החלטות מבוסס-נתונים, אם זה מחקר מדעי, אם זו קבלת החלטות בארגון ואם החלטה אישית. זהו מודל גנרי המתאים לכל תהליך שמסתיים בהחלטה, בהסקת מסקנות או בגילוי מדעי, ולכן הוא תקף לכל דיסציפלינה או בעיה. בכל היוזמות שתוארו בפרק הקודם מחזור הנתונים הוא הציר המרכזי בדיון. גם כאן התכנים המוצעים והכלים התומכים בהם מוצגים בהלימה למחזור זה.

2.2.1 שלבי מחזור הנתונים

א. הגדרת הבעיה: הגדרה ראשונית של בעיית ההחלטה, המשימה, המטרה או המחקר במרחב הידע שלשמש נועדו הנתונים. ניסוח הבעיה צריך להבהיר אילו סוגי נתונים עשויים לסייע בפתרונה. הבעיה יכולה להיות של תיאור, חיזוי, הבנה, סיווג או הוכחה של קשר סיבתי בין משתנים מסבירים לתוצאות.

ב. איסוף הנתונים הרלוונטיים: זיהוי מקורות אפשריים לנתונים ואיתורם; סוגי הנתונים הדרושים, בחינת התאמתם לבעיה מבחינת נגישות, עדכניות, כיסוי, הטיות אפשריות בדרכי יצירתם וההשלכות של הטיות אלה על אמינות המחקר; יצירת נתונים רלוונטיים באמצעות אחזורם ממסדי נתונים קיימים או באמצעות ניסויים, ניסויי מחשב וסקרים; איסוף הנתונים ושמירתם בדרך נוחה לניתוח או יצירת מערכת המאפשרת שליפת נתונים חוזרת ממקור מידע מתעדכן שזוהה; הבטחת שקיפות התהליך והאפשרות לשחזרו.

ג. ניקוי, סינון, גיבוי ואינטגרציה של נתונים (data cleansing, filtering, backup and integration): ניפוי נתונים שגויים; מיון נתונים לפי הרלוונטיות שלהם; יצירת גיבוי; אינטגרציה בין נתונים ממקורות שונים בעלי הגדרות שונות; סטנדרטיזציה וטרנספורמציה של מדדים (למשל צלסיוס ופרנהייט, ק"מ ומייל, הכנסה, מספר נשאים מאומתים על סקאלה לוגריתמית או תמותה למיליון נפש); הגדרת תוצאות עיקריות ומשניות לפי הרלוונטיות לבעיה.

ד. עיבוד נתונים, ניתוחם וכרייתם (processing, analysis and mining):

עיבוד וניתוח של הנתונים הרלוונטיים. הניתוח תלוי בסוג הנתונים, במספר המאפיינים המתארים ובגודל קובץ הנתונים. ניתוח הנתונים יתקדם בדרך כלל מהפשוט ביותר, כגון השוואת תמציות מספריות של נתונים ובחינת קשר בין מאפיינים, ועד כלים מתקדמים המבוססים על מודלים סטטיסטיים ולמידת מכונה.

ה. כלים לויזואליזציה ולהוצאת מסקנות מניתוח הנתונים: הוצאת המסקנות

מהניתוח, הערכת אי-הוודאות שבהן ומגבלותיהן האחרות לפי הדרך שבה נאספו הנתונים; הצגת תוצאות העיבוד, הניתוח והמסקנות בדימוי, בכתב ובעל פה לפני מקבלי ההחלטות או החוקרים. הצגה יכולה להיות ויזואליזציה גראפית או טבלאית, דינמית או סטטית, בתקציר מנהלים או בהרצאה וכד'. יש חשיבות רבה להצגה ברורה המתאימה לתחום העיסוק של קהל היעד של תוצאות הניתוח. בכך עוסק התחום של ממשק אדם-מכונה.

ו. השלכות ניתוח הנתונים על הידע בבעיה, והיזון חוזר לתהליך: מסקנות

מנתונים, למרות חשיבותן, הן תמיד חלקיות וכוללות מרכיב של אי-ודאות וחוסר רלוונטיות. בצעד האחרון נבחנת השפעת תוספת הידע שנרכש על הבעיה המקורית בתחום הידע. זהו שלב קריטי, בין שמדובר במסקנות מדעיות, בין בהחלטות, בין בהתוויית כיווני פעולה ובין במחקרי המשך. לעיתים תודגם חשיבות הפעילות כתהליך מתמשך, והמסקנה תהיה בכיוון של פיתוח כלים ייחודיים להטמעת תהליכי ניתוח נתונים, למשל באמצעות טבלאות דינמיות או לוח מחוונים (dashboard). ניהול הידע הנצבר בארגון או אצל הפרט נמנה גם הוא על השלב הזה.

מחזור הנתונים, כפי שמעיד שמו, הוא תהליך החוזר על עצמו. בדרך כלל השלב האחרון במחזור הנתונים מאיר את הבעיה המקורית בתחום הידע באור חדש. סוגיות חדשות, או פנים חדשות בבעיה המקורית, עולות ודורשות הגדרה של בעיית המשך, איתור נתונים וכן הלאה – סיבוב חדש של פעילות ניתוח נתונים. במקרים אלו אפשר גם שהפעילות בשלבים קודמים תשתכלל בעקבות שיפור הביצועים בסבבים הבאים של המחזור.

בתרשים הבא מתואר מחזור הנתונים תיאור גראפי. מרכיביו מקובלים על כל העוסקים במדעי הנתונים, אך מספר השלבים שבו יכול לנוע בין 6 ל-9. כל אחד מהשלבים מפורק לעיתים בידי קהילה מסוימת של מדעני נתונים לשלבי משנה, פירוק המדגיש את תחום עיסוקם. לצורך תוכני ההוראה יוצג כאן מבנה פשוט יחסית:



מדובר במודל גנרי אשר מיושם הלכה למעשה הן במחקר אקדמי והן במגזר העסקי. לדוגמה למחקר אקדמי ניתן לעיין ב**דוח של מוסד נאמן מתאריך יולי 2018**. הבעיה שאיתה ביקש המחקר להתמודד הייתה להבין תהליכים הנוגעים לפער דיגיטלי, דהיינו פער הבקאות במיומנויות הדיגיטליות בין קבוצות שונות באוכלוסייה, ובין מדינות שונות בעולם. לשם כך הוצעה דרך לאסוף נתונים ממקורות שונים, למזג את הנתונים ולהפיק מהם תובנות. הנתונים כללו נתוני גלישה באתרים שונים, צריכת אקטואליה, פעילות ברשתות חברתיות ונתונים דמוגרפיים. שלבי הטיפול בנתונים המתוארים בדוח של מוסד נאמן חופפים את מחזור הנתונים

שתואר לעיל. בתחום העסקי מוכרות דוגמאות כגון מערכות ההמלצה של אמזון, נטפליקס, ספוטיפיי ואחרות כיישומים מובהקים של מחזור הנתונים לצורך הפקת תובנות עסקיות.

2.2.2 ארגז הכלים שעומד לרשות כל שלב

בארגז כלים זה אנחנו מפרידים בין כלים מינימליים, שראוי שכל סטודנט וסטודנטית בקמפוס יוכלו להפעילם (יופיעו בסעיף 1 להלן), לבין כלים מתקדמים יותר, המתאימים לסטודנטים בעלי רקע מתאים (יופיעו בסעיף 2 להלן).

א. כלים להגדרת הבעיה

1. שיטות גישות איכותניות להגדרת בעיות ולניסוחן (למשל סיעור מוחות וקבוצות מיקוד); מודלים כמותיים קיימים (למשל סקירת ידע קיים ומטא-אנליזה); הגדרת שאלות מחקר או השערות מחקר וגיבוש השערות לאור נתונים (exploratory data analysis)
2. כלי לניסוח פורמלי של בעיה עסקית

ב. כלים לאיסוף הנתונים הרלוונטיים

1. מנועי חיפוש; שימוש במאגרי מידע לאומיים ובין-לאומיים כגון לשכות לסטטיסטיקה באמצעות אינדקסים, מאגרי מידע של ארגונים מקומיים ובין-לאומיים (למשל האו"ם, ארגון הבריאות העולמי, אונסק"ו ועוד); טכנולוגיות של העברת קובצי נתונים גדולים; ניסויים מבוקרים; סקרים
2. שיטות עבודה מול ענן

ג. כלים לניקוי נתונים, לסינונם, לגיבויים ולאינטגרציה שלהם

1. ויזואליזציה של נתונים; ניתוח לוגי וניתוחי סבירות לשם בקרת איכות של נתונים
2. פילטרים, אונטולוגיות ומטא-דאטה; הסבה ודאגה לגיבוי

ד. כלים לעיבוד נתונים, לניתוחם ולכרייתם

1. תוכנות מבוססות-תפריטים לניתוח נתונים (ללא תכנות) ושימוש בהן לתיאור, להשוואה ולבחינה של קשרים מבוססי-מודלים סטטיסטיים; עצי החלטה ויערות מקריים לחיזוי

2. שפות תכנות לשימוש באלגוריתמים קיימים כגון R ו־Python ושימוש בהן ליישום שיטות סטטיסטיות ולמידת מכונה, ובתוכן רשתות נוירונים ולמידה עמוקה; כלים לעיבוד תמונה, שפה וקול, נתוני תחלואה והישרדות ונתונים גנטיים

ה. כלים לוויזואליזציה ולהוצאת מסקנות מניתוח הנתונים

1. ויזואליזציה של תוצאות, דיווח במצגת ובטקסט; הערכת אי־ודאות סטטיסטית; קשר אדם-מכונה
2. סימולציות; מערכות לגישה אינטראקטיבית ותשאול

ו. כלים להשלכות ניתוח הנתונים על הידע בבעיה ולהיזון חוזר לתהליך

1. אלגוריתמים של אם-אז; הכלים שהוגדרו עבור שלב 1
2. אפיון כלי החלטה לפיתוח, תוכנות לוח מחוונים dashboard

2.2.3 אתיקה וחשיבה ביקורתית בנייתוח נתונים

בתהליך ניתוח הנתונים עולות סוגיות אתיות הנוגעות לפרט ולחברה. חלק מן הסוגיות אף נושא השלכות משפטיות. הכשרת מדעני נתונים מחייבת מודעות לסוגיות אלה והכרת דרכים שונות להתמודד איתן. הסוגיות האתיות הן מהותיות, ויש להתחשב בהן כאשר קובעים את שאלות המחקר ואת שיטת הפעולה. מקצת הבעיות ייסקרו כאן.

בנתונים העוסקים במידע על אנשים, כגון נתונים רפואיים, נתונים על הליכים משפטיים או דיווחי הכנסה, סוגיות של שמירת חשאיות ופרטיות הן חשובות במיוחד: האם יש הסכמה מודעת לפרסום המידע? אם לא, מה כן ניתן לפרסם כדי לא לפגוע בה? האם מאן דהו עלול להיחשף לאלימות או לפגיעה אחרת בעקבות הפרסום? כיצד מאזנים בין היתרון לחברה מגילוי מלא של הנתונים לבין הפגיעה האפשרית בפרט?

ראוי לשקול גם את ערכו של היתרון שצוברות חברות מסחריות גדולות לעומת האינטרסים של החברה ושל היחיד – עד כמה ראוי להעניק לענקי מידע את היכולת להשפיע על התנהגות אנושית. נוסף על זה, יצירת מידע במאגרים גדולים תעורר גם שאלות של זכויות יוצרים, שלהן השלכות משפטיות רבות.

מצד הפרוצדורה החישובית, יש חשיבות להתנהגותו האתית של החוקר העוסק בניתוח נתונים אגב גילוי נאות של אינטרסים צולבים וללא עיוות של התהליך. עיוות שכזה יכול לקרות למשל בניפוי נתונים והשמטתם או בבחירה מודעת של תהליכי ניתוח לאחר ידיעת תוצאותיהם, וכל זאת לצורך הוצאת מסקנות הרצויות לחוקרים בשל אינטרס זר.

כאשר מפעילים אלגוריתמים מורכבים לשם קבלת החלטה בנושא בעל חשיבות חברתית-לאומית, כגון חלוקת תמיכה בעת מגפה או כל חלוקת משאבים אחרת, יש לבחון אם האלגוריתם מסתיר בחובו הטיה מגדרית, גילנית או אחרת. כמובן יש להעדיף מודלים ללא הטיית. הטיית שייקבעו בעת עיצוב האלגוריתם עלולות להתעצם בהמשך פיתוחן באמצעות מערכות אוטונומיות, ועל כן יש להציע דרכים לפיקוח על מערכות אלה.

שקיפות תהליך הניתוח והנגשת הנתונים הן חובות אתיות, שנועדו לאפשר למדענים אחרים להעריך את הדירות המסקנות ואת תקפותן. השקיפות מחייבת שיתוף בכל שלבי הניתוח (open source) ויצירת מאגרים (repositories) להנגשה ולשיתוף נוחים. בהתמודדות עם מגפת הקורונה ראינו לאחרונה התנגשות בין שני עקרונות אלו, כאשר ניתוחים מדעיים שהובילו למסקנות שהשפיעו על חיי רבים נותרו עלומים, ורק המסקנות פורסמו. יש להקפיד על גילוי נאות של הטיית אפשריות באיסוף הנתונים ושל מגבלות אפשריות של השיטות ששימשו לניתוחם, כמו גם על ניתוח נכון וצנוע של המסקנות העולות מניתוח הנתונים בציון מגבלותיהן ותוקפן.

רבות מן הסוגיות האתיות והמשפטיות שהעלינו כרוכות בחשיבה ביקורתית על המתודה, שהיא לב ליבה של עבודת נתונים איכותית. אין מדובר בעניין צדדי אלא במקצועיות של כל חוקר או חוקרת. יש לפתח אצל התלמידים חשיבה ביקורתית בעת עבודתם לאורך מחזור הנתונים ובעת הערכת תוצאות הניתוח שנעשה בידי אחרים. חשיבה שכזו ראוי שתירכש הן בעת ניתוח הנתונים והן בקריאה ביקורתית של מחקרים המבוססים על מדעי הנתונים. התלמידים יהיו מסוגלים לא רק לשפוט את הפרוצדורה המתמטית אלא גם להבחין בהשלכות אתיות של איסוף המידע והשימוש בו, במשמעותם החברתית ובסוגיות משפטיות שעשויות להתעורר.

חלק ניכר מענייני האתיקה שנמנו לעיל הוצגו בקורס אקדמי, ככל הידוע לנו בפעם הראשונה בארץ, בידי פרופ' אביגדור גל מהטכניון – מכון טכנולוגי לישראל ופרופ' ניבה אלקין-קורן מאוניברסיטת חיפה בשנת תש"ף. הדגש בקורס זה הוא על הצד המשפטי של הנושא. בבסיסו של הקורס קיימת התובנה שמערכות בינה מלאכותית מעלות סוגיות ייחודיות וכבודת משקל אשר משלבות בין עולמות המשפט והאתיקה לבין עולם הטכנולוגיה. על מנת לזהות סוגיות אלו, וכמובן על מנת לנתחן ולהציע להן פתרונות, נדרשת היכרות עם שני סוגי העולמות. מטבע הדברים, להרבה מאנשי מדעי הנתונים אין היכרות מעמיקה עם תחום המשפט, ולהרבה מאנשי המשפט אין רקע טכנולוגי, והדבר מקשה על דיאלוג בין הצדדים. לפיכך אנשי מדעי הנתונים עלולים לפתח טכנולוגיות שמעוררות קשיים אתיים וחוקיים, אך יהיה אפשר למזערן מראש אם פיתוחן ייתמך בהבנה מעמיקה יותר של עולם המשפט, ומנגד אנשי המשפט עלולים לצפות לפתרונות אשר אינם מתיישבים עם הכלים הקיימים במציאות. מטרתה העל של הקורס היא לייצר דיאלוג בין שני העולמות, שיניח את הבסיס לשפה שהסטודנטים יהיו מסוגלים לזהות באמצעותה בשלב מוקדם את נקודות המפגש הפוטנציאליות ביניהם.

הקורס נשען על ניתוח מקרים. לכל מקרה יש צד משפטי (הדורש ניתוח משפטי) וצד של מדעי הנתונים (אשר דורש הרצת מודלים והצגת תוצאות בצורה המובנת למשפטנים). בכל שיעור נידון נושא אתי אחר (לדוגמה: פרטיות, אפליה וכו'), והסטודנטים מבצעים מטלה ומציגים אותה. לשם ביצוע המטלה הסטודנטים מתחלקים לצוותים מעורבים.

שיטת ההוראה

הקורס כלל מפגש פתיחה וסיום ועוד חמישה מפגשים בני ארבע שעות כל אחד, שכל אחד מהם הוקדש לנושא אחר: אחריות נזיקית, הוגנות ואפליה, שקיפות והנמקה בקבלת החלטות, פרטיות ואנונימיות והתערבות בתכנים.

כל שיעור (לבד מהשיעור האחרון) התחלק לשלושה חלקים: מבוא לנושא הנידון, עבודה על מטלה קבוצתית וסיכום המטלה באמצעות פוסטרים ומצגות. הנושאים שנועדו לדיון בשיעורים וחומרי הקריאה ניתנו מראש. קריאת החומר לפני השיעור הייתה באחריות הסטודנט.

המטלות נועדו לעודד הבנה מעמיקה של החומר. ביצוע המטלות בקורס (כולל הפרויקט) היה בקבוצות בנות ארבעה סטודנטים, כל אחת, ואת מבנה הקבוצות קבעו מורי המקצוע בשיתוף פעולה עם הסטודנטים. המטרה הייתה לייצר קבוצות הטרוגניות ולהעניק לכל חבר בקבוצה ערך מוסף משלו.

במהלך הקורס ניתנו חמש מטלות (50% מהציון) ופרויקט (40% מהציון). נוסף על זה, כל סטודנט נדרש להציג מצגת קצרה (10%).

2.2.4 הצורך להציג גישות שונות ולתת דגשים שונים לדיסציפלינות שונות

אין להציג את נושאי הלימוד שנמנו בסעיף הקודם ואת ארגז הכלים שמלווה אותם באותה דרך לפני כל הסטודנטים בקמפוס אקדמי. ברור שהידע שנדרש בנושא מתלמידי מדעי הרוח שונה מן הידע הנדרש מתלמידי משפטים, ואלה שונים ממדעי החברה ומניהול. גם בתוך כל פקולטה עשוי להיות שוני בין הדיסציפלינות שכלולות בה. למשל: התמחות במימון והתמחות בהתנהגות ארגונית (בתוך הפקולטה לניהול) שונות זו מזו באופי איסופם ועיבודם של הנתונים שהן נדרשות להם; במדעי החברה סוציולוגיה שונה מכלכלה, ושתיהן שונות מפסיכולוגיה או מאנתרופולוגיה. קל וחומר כאשר משווים בין מדעי הטבע ורפואה למדעים המדויקים ולהנדסה, ואת כל אלה למדעי החברה והרוח.

2.2.5 המלצות כלליות

הוועדה ממליצה שכל תלמיד בקמפוס ייחשף במהלך לימודיו לתואר ראשון לכל הנושאים שבמחזור הנתונים וילמד אותם לשם פיתוח היכרות עם מדעי הנתונים והבנתם, וכן לפיתוח יכולת ביצוע במדעי הנתונים. ניתן לעשות זאת בכמה דרכים: הדרך האחת היא פיתוח קורס מבוא למדעי הנתונים; הדרך השנייה היא העשרת קורסים קיימים בשיטות מחקר, בסטטיסטיקה ובתכנות, כדי שיקיפו את כל הנושאים שבמחזור הנתונים. מובן שיייתכן גם שילוב של שתי דרכים אלו, עם כיסוי חלקי בקורסים קיימים והשלמה בקורס מבוא קצר יותר.

הקורס "מבוא למדעי הנתונים" אינו צריך להיות זהה לכל הסטודנטים אלא להיות מותאם לדיסציפלינה הספציפית. עם זאת לכל הקורסים האלה יהיה חומר

ליבה משותף, שיכלול עקרונות של מחזור הנתונים שתואר לעיל. בבניית הקורס, בחלק מהדיסציפלינות נסתפק בהכרת עקרונות וכלים, בייחוד אלו שהוגדרו כמינימום, ובאחרות נשאף להגיע למצב שבו ידע הסטודנט לפעול בתחום ולהפעיל את הכלי, ויהיה בעל יכולת ביצוע בכל נושא וכלי. סטודנט למדעי הרוח, למשל, צריך להכיר את מחזור הנתונים ולצידו כלי מדף מרכזיים בתחומי הנתונים הרלוונטיים לו, וגם לדעת להפעיל את הכלים ולהיות מסוגל להעריך בעין ביקורתית את עבודתם של אחרים המשתמשים בהם. סטודנט לסוציולוגיה או לכלכלה צריך לדעת, נוסף על אלה, על קיומם של מאגרי נתונים רלוונטיים וכן לדעת להפעיל תוכנה סטטיסטית. הוא הדין בתלמידי ניהול. לעומת זאת תלמידי מדעי הטבע, רפואה והנדסה צריכים לדעת להשתמש בכמה תוכנות מתקדמות, להכיר אלגוריתמים ומערכות תומכות-החלטה.

ראוי שהמערכת תאפשר גמישות לסטודנטים להעמיק בתחום, למשל לאפשר לתלמידי מדעי הרוח המעוניינים בכך לקחת קורס מעמיק יותר במדעי הנתונים הניתן בפקולטה אחרת.

פרק 3 מפרט אילו נושאים מתוך רשימת הנושאים והכלים שהובאה כאן כדאי שייכללו בכל תחום אקדמי, מהם הנושאים שיודגשו, ולעיתים גם אילו דוגמאות משמעותיות יכולות לשמש בהוראה.

פרק 3

המלצות יישום בתחומי ידע שונים

המלצות יישום בתחומי ידע שונים

3.1 מבוא

הוועדה ממליצה שכל תלמיד ילמד את מודל מחזור החיים של נתונים המתאים לתחום לימודו. לימוד זה יכול להיות בקורס ייעודי "מבוא למדעי הנתונים" או בהוספת תכנים לקורסים קיימים או בקורס משולב של הדיסציפלינה עם מדע הנתונים. אם יוחלט על קורס ייעודי, מומלץ לתכנן קורס מודולרי כדי שיהיה אפשר לעצבו לפי דרישות שונות: תחילה ליבה עקרונית וגנרית המבוססת על מודל מחזור החיים של נתונים (ראו לעיל סעיפים 2.2.1-2.2.2), ולאחריה נושאים המותאמים לצרכים של תחומי המחקר והלימוד השונים (לפי פקולטות או חוגים). מתווה מוצע לליבה הגנרית של הקורס מתואר להלן. ליבה זו מכילה נושאים עקרוניים, חוצי-תחומים. סעיפים 4-6 של מרכיב הליבה מיועדים להקשר ולדוגמאות הספציפיות של תחומי הלימוד השונים. בהמשך הפרק נעסוק בהרחבה בהמלצות היישום בתחומי הלימוד והידע השונים. כל אוניברסיטה וכל פקולטה תבחר לעצמה את מה שלדעתה דרוש לה ומתאים לה, וכן תחליט אילו נושאים יילמדו בקורס נפרד ואילו ישולבו בקורסים קיימים.

כדי לקדם את הנושא שיהיה שוויוני בכל המוסדות האקדמיים הוועדה ממליצה לפתח את החלק הגנרי של הקורס במתכונת של קורס מקוון כלל-ארצי שיעמוד לרשות כל המוסדות האקדמיים ויכל את הליבה העקרונית המבוססת על מודל מחזור החיים של נתונים. הקורס המקוון יכלול את סעיפים 1-3 המופיעים במסגרת להלן.

מטרות מרכיב הליבה "מבוא למדעי הנתונים"

- א. הבנת האפשרויות הגלומות במדעי הנתונים לצורך מחקר אקדמי ומעשי בתחומי הידע השונים
- ב. הכרת מושגים בסיסיים במדעי הנתונים
- ג. רכישת כלי ניתוח ותכנות בסיסיים (בעיקר לתלמידי מדעי הטבע וההנדסה)
- ד. יכולת לתקשר עם מומחים בתחום מדעי הנתונים ולהיעזר בהם
- ה. הכשרת בוגרים לעבודה במשרות שדורשות היכרות עם מדעי הנתונים ושיתוף פעולה עם מומחי נתונים

תיאור מרכיב הליבה

משך הקורס:

2-1 סמסטרים

היקף הקורס:

2 שעות הרצאה שבועיות, 2 שעות תרגול שבועיות

נושאי הקורס:

1. מחזור חיי הנתונים
2. סקירת כלים התומכים בשלבים השונים של מחזור חיי הנתונים
3. סוגי נתונים במרחב הדיגיטלי: מספריים, טקסטואליים, מובנים ולא-מובנים (structured, unstructured)
4. יישום בסיסי של איסוף נתונים, אינטגרציה של נתונים ממקורות שונים, ניתוח הנתונים, למידה אלגוריתמית וויזואליזציה של נתונים (מותאם לתחום הלימוד)
5. חשיבה ביקורתית לאורך מחזור חיי הנתונים (מותאם לתחום הלימוד):
 - * אמינות מקור הנתונים
 - * מהימנות ותוקף של נתונים
 - * מהימנות של אלגוריתמים
 - * הלימות (adequacy) והדירות (reproducibility) של שיטת המחקר

- * אתיקה בתהליך איסוף הנתונים, גילוי נאות
 - * שמירה על פרטיותם של יחידים ושל ארגונים
 - * בדיקת הטיות בנתונים ובמצאים
 - * קריאה ביקורתית של מחקרים המבוססים על מדעי הנתונים
6. שיתוף נתונים (open data, data repositories) וכללי הציטוט של מקורות נתונים בעבודה אקדמית מבוססת-נתונים

3.2 מדעי הרוח

הוועדה ממליצה להקים קורס ליבה ורכיבים מתקדמים לשימוש בנתונים במדעי הרוח, לפי הפירוט להלן. הקורס יילמד באוריינטציה ייעודית למדעי הרוח, מפי מומחים מן התחום, ויילמדו בו דוגמאות ומתודות המתאימות למדעי הרוח. הקורס עשוי להשתלב בתוכניות הלימודים ההולכות ונבנות במדעי הרוח הדיגיטליים, אך מידת השילוב של השניים פתוחה להחלטת כל פקולטה.

לצד הכשרתם העיקרית כהומניסטים, על תלמידי מדעי הרוח להכיר את האפשרויות הנפתחות בפניהם בזכות שימוש בנתונים ועיבודם. רוב התלמידים בתחום זה אינם בעלי המיומנויות המתמטיות והסטטיסטיות הנדרשות כדי להפעיל מחקר עצמאי במדעי הנתונים, אך הם חיים בעולם של דאטה, והם מדברים את שפת הנתונים שלא במודע. המסגרת הכללית של הקורס במדעי הרוח מוגדרת אפוא אחרת: הכרת המחקר במדעי הנתונים; הטמעת התפיסה כי חומר הומניסטי עשוי להיות מיוצג גם כדאטה, ומתי הדבר עשוי להועיל; האפשרות ליצירת שיח עם אנשי מדעי הנתונים על מנת לפעול במשותף; פיתוח חשיבה ביקורתית על עולם הדאטה והשימוש בו.

התלמידים יתנסו בחשיבה מושגית על האובייקט ההומניסטי כדאטה ויתנסו גם מעשית בכלים ובשיטות מחקר הנוגעים למחקר מונחה-הדאטה במדעי הרוח. מחזור הנתונים יהיה המסגרת המושגית העיקרית כנדרש, אך יודגשו השלבים מן המחזור שיש בהם חשיבות מיידית למדעי הרוח. יודגשו פחות השלבים שתלמידי מדעי הרוח יידרשו בהם לעזרת מומחים. במידת האפשר יכירו התלמידים כלי מדף לביצוע השלבים השונים במחזור הנתונים, ולא יידרשו ללמוד תכנות פעיל. כלים רבים כאלה למחקר במדעי הרוח זמינים ברשת ופעילים זה מכבר.

השילבים העיקריים במחזור הדאטה שיודגשו במהלך הקורס יכללו:

* **שלב ייחודי בקורס המיועד למדעי הרוח:** היסודות המתמטיים של מדעי הנתונים ויישומם בכלים בסיסיים, כגון אקסל, אוריינות בקריאת גרפים והיכרות עם רשתות (social network analysis)

* **בני אדם כנתונים:** אתיקה הומניסטית של מדע הנתונים; איסוף מידע פרטי ושימוש בו; גישה לארכיונים דיגיטליים; מגבלות וחסמים משפטיים

* **מעבר מבעיה לנתונים:** כיצד מתרגמים שאלת מחקר הומניסטית לנתונים, מדוע זה עשוי להיות חשוב, ומתי יש להשתמש בכלים מסורתיים; הכרת סוגי נתונים וזיהוי מקורות דאטה במדעי הרוח

* **מתוצאה מספרית למסקנה היסטורית:** ההליך הפרשני של המרת תוצאות המחקר למסקנה מחקרית בתחום מדעי הרוח

* **מטקסט חופשי למידע מובנה:** הבחנה בין דאטה למטא-דאטה; החשיבות שבמשאבי מידע פתוחים והיכולת לקשר ביניהם קישור חופשי (linked open data)

* **בין ביקורת הומניסטית לענן המילים:** הכרת כלי ניתוח וויזואליזציה; פרשנות ומתן משמעות לתוצאות הניתוח

* **שימור הנתונים (קיימות דאטה):** הבטחת המשכה של נגישות הדאטה

לשם השלמת התמונה הכוללת של מחזור הנתונים יילמדו בקורס גם השלבים הנוספים של המחזור:

* מהחומר למסך המחשב: דיגיטציה; איסוף וכרייה של נתונים; בניית מסד נתונים

* אחסון, שימור, עדכון ותחזוקה של מאגר הנתונים

* עיבוד נתונים וניתוחם (עיבוד שפה טבעית, ניתוח נושאי, ניתוח רשתות חברתיות)

באמצעות קבוצות התרגול ייתן הקורס ביטוי לתחומי התמחות שונים לנתונים במדעי הרוח: **מרחב** (מיפוי היסטורי וגאוגרפי), **חומר וכתב יד** (ארכאולוגיה, תולדות האומנות ומדע כתבי היד), **שפה** (מבנה, דקדוק, שינויים בשפה ואוצר מילים), **טקסטואליות** (ניתוח אוטומטי של טקסטים הומניסטיים). לצד התאוריות והמתודולוגיות יושם דגש לאורך הקורס על ניתוח ובחינה ביקורתיים של פרויקטים, אתרים ומסדי נתונים קיימים בתחומי מדעי הרוח השונים.

3.3 מדעי החברה וניהול

עקרונות:

- א. הוועדה תציין רשימת נושאים רלוונטיים וחשובים, אך לא תציע תוכנית לימודים אחידה ומוכתבת.
- ב. כל אוניברסיטה וכל פקולטה למדעי החברה תבחר לעצמה את מה שלדעתה דרוש ומתאים לה.
- ג. כל אוניברסיטה וכל פקולטה תחליט אילו נושאים יילמדו בקורס נפרד ואילו ישולבו בקורסים קיימים.

מטרות:

- א. הבנת האפשרויות הגלומות במדעי הנתונים לצורך מחקר אקדמי ומעשי במדעי החברה
- ב. הכרת מושגים בסיסיים במדעי הנתונים
- ג. יכולת לתקשר עם מומחים בתחום מדעי הנתונים ולהיעזר בהם
- ד. הכשרת בוגרים במדעי החברה לעבודה במשרות שדורשות היכרות עם מדעי הנתונים

הוועדה ממליצה על פיתוח קורס של "מבוא למדעי הנתונים למדעי החברה" – מתווה לקורס מתואר בפרק הקודם. חלקו של הקורס מכיל נושאים עקרוניים, חוצי-תחומים, וחלקו מיועד להתאמה נושאית על פי תחומי הלימוד (סוציולוגיה, כלכלה, מדע המדינה, תקשורת וכו').

בשל היותו של הקורס מכנה משותף רחב וכדי לקדם את הנושא שיהיה שוויוני בכל המוסדות האקדמיים, אנו ממליצים שהאקדמיה למדעים תשקול לפתח ביוזמתה, או להמליץ למל"ג, לפתח קורס מקוון כלל-ארצי שיעמוד לרשות כל המוסדות האקדמיים.

נוסף על הקורס הבסיסי אנו ממליצים להלן גם על רשימת נושאים עקרוניים וחוצי-תחומים שמתאימים לקורסים מגוונים ומתקדמים בתחום מדע הנתונים. ניתן לבנות קורסים/שלמים/מנושאים אלה או לשלב מבחר נושאים בקורסים קיימים.

הקונספט שלפיו ייבחרו הנושאים הוא מחזור החיים של נתונים, החל מבריאתם וכלה בשימוש בהם לקבלת החלטות או למחקר.

רשימת הנושאים

מחזור החיים של נתונים	נושאים או קורסים מוצעים
הגדרת צורך	מודלים כמותיים, מודלים איכותניים, יסודות הסטטיסטיקה, מידע וקבלת החלטות
איתור מקורות נתונים (מאגרי נתונים קיימים, נתונים מהרשת, רשתות חברתיות, אינטרנט של הדברים)	הכרת תוכנות לסריקה ברשת, הכרת מאגרי מידע ונתונים חשובים ופתוחים לקהל (למשל הלשכה המרכזית לסטטיסטיקה, הסדנה לידע ציבורי)
איסוף/כריית נתונים וניקוי נתונים	תקשורת נתונים, ניהול מאגרי נתונים, פילטרים, יסודות התכנות (Python או R)
אינטגרציה של נתונים ממקורות שונים	תוכניות הסבה, מטא-דאטה, אינדקסים
כריית נתונים	פילטרים, טכניקות לאחזור מידע, כלים של בינה מלאכותית, יוריסטיקה, יסודות התכנות
עיבוד נתונים וניתוחם	אלגוריתמיקה, machine learning, יוריסטיקה, עיבוד מידע, סימולציות
ויזואליזציה של נתונים	dashboard software, כלים גראפיים, כלי דיווח, UX
הסקת מסקנות, קבלת החלטות	מערכות תומכות החלטה, סימולציות
משוב והגדרה מחדשת של צורך	מערכות דיווח, כלי agile, כלי DEVOPS, machine learning

דוגמה מכלכלה, ניהול וסוציולוגיה

להלן מבנה קורס המבוסס על המודול הגנרי ועל תוספות המכוונות למדעי החברה ולניהול

משך הקורס:

2-1 סמסטרים

היקף הקורס:

2 שעות הרצאה שבועיות, 2 שעות תרגול שבועיות

נושאי הקורס:

1. מחזור חיי הנתונים

2. סקירת כלים התומכים בשלבים השונים של מחזור חיי הנתונים

3. סוגי נתונים במרחב הדיגיטלי: מספריים, טקסטואליים, מובנים ולא-מובנים (structured, unstructured)
4. יישום בסיסי של איסוף נתונים, אינטגרציה של נתונים ממקורות שונים, ניתוח הנתונים, למידה אלגוריתמית וויזואליזציה של נתונים (מותאם לתחום הלימוד) – הכרת מקורות מידע הלקוחים מהלשכה המרכזית לסטטיסטיקה, מהמוסד לביטוח לאומי, ממוסדות בין-לאומיים כמו למשל סוכנויות של האו"ם, אונסק"ו, האיחוד האירופי, OECD, ממאגרי מידע של מדינות שונות וכד'; הצגת בעיות של אינטגרציה של נתונים המגיעים ממקורות שונים וחקר שאלה דמוגרפית; חקר שאלה מקרו-כלכלית, ניתוח מסחר בין-לאומי וניתוח פערים חברתיים
5. חשיבה ביקורתית לאורך מחזור חיי הנתונים (מותאם לתחום הלימוד):
 - * אמינות מקור הנתונים
 - * מהימנות ותוקף של נתונים
 - * מהימנות של אלגוריתמים
 - * הלימות (adequacy) והדירות (reproducibility) של שיטת המחקר
 - * אתיקה בתהליך איסוף הנתונים, גילוי נאות
 - * שמירה על פרטיותם של יחידים ושל ארגונים
 - * בדיקת הטיות בנתונים ובמצאים
 - * קריאה ביקורתית של מחקרים המבוססים על מדעי הנתונים
6. שיתוף נתונים (open data, data repositories) וכללי הציטוט של מקורות נתונים בעבודה אקדמית מבוססת-נתונים

3.4 משפטים

בשנים האחרונות גובר השימוש בנתונים בתחום המשפט בין באמצעות כלים סטטיסטיים ובין באמצעות אלגוריתמים של בינה מלאכותית. הדבר נעשה הן לצרכים מעשיים והן לצורכי מחקר. הממשק שבין משפט לנתונים ולטכנולוגיה נוגע כיום למכלול ההיבטים של העיסוק המשפטי של השופטים, של עורכי הדין, של המשפטנים במגזר הפרטי ובשירות הציבורי ושל חוקרי המשפט. על הפקולטות למשפטים מוטלת המשימה להכשיר את דור העתיד של השופטים, עורכי הדין והחוקרים להתמודדות עם העולם העשיר של מדעי הנתונים, הן לשם

חשיפה והכרה בסיסית של התכנים הנוגעים לשימוש בנתונים והן לשם רכישת תובנות תאורטיות וכלים מעשיים לניתוח נתונים. נוסף על זה, חשוב לפתח את יכולות התלמידים לקיים דיון מושכל בשימוש בנתונים ובאלגוריתמיקה בעולם המשפטי והערכה ביקורתית שלהם, על מנת להתמודד עם שאלות אתיות ומשפטיות הנוגעות לשימוש בנתונים בידי בני אדם ומכונות כאחד.

קורס בנושא משפט ומדעי הנתונים יחולק לכמה יחידות שידגישו היבטים שונים בממשק שבין שיפוט, עריכת דין, מחקר משפטי לנתונים. בכל אחת מהיחידות יעסקו הסטודנטים בהבנת המסגרת המשפטית הנוהגת ובזיהוי החסרים הנורמטיביים והמעשיים בהסדרה המשפטית של השימוש בנתונים לצד חשיפה למאגרי נתונים, לפרויקטים דיגיטליים ולכלי ניתוח נתונים והכרתם.

יחידה 1: שימוש בנתונים לקבלת החלטות במגזר הציבורי וזכויות אדם: חקיקה, רגולציה, אכיפה פלילית ושפיטה

יחידה 2: שימוש בנתונים במגזר הפרטי וזכויות הפרט: פרטיות, מעקב וחופש ביטוי

יחידה 3: שימוש בנתונים ובבינה מלאכותית

יחידה 4: שימוש בנתונים בפרקטיקה המקצועית – ליגל טק

יחידה 5: שימוש בנתונים במחקר משפטי

כבר כיום מלמדים במוסדות בארץ קורסים בנושא, וקיימים שיתופי פעולה מחקריים בין חוקרי משפט למרכזי המחקר במדעי הנתונים.

3.5 מדעי החיים ורפואה

תחומי הלימוד בפקולטות למדעי החיים ולרפואה בישראל מגוונים ונעים בין התחומים הקלאסיים של זואולוגיה ובוטניקה, לביולוגיה תאית ומולקולרית, לגנטיקה ולביולוגיה מערכתית וחישוביות עצבית. מטרתנו להצביע על הבסיס הראוי המשותף לתוכניות הלימוד השונות ולהדגישיהן בתחום מדעי הנתונים, כשיש להביא בחשבון את שאר מרכיביהן של תוכניות הלימוד בתואר הראשון בתחומים אלה מחד גיסא, ואת הצורך להעניק אוריינות בסיסית לבוגרים שיפנו לתחומי התמחות שונים מאידך גיסא.

ההתפתחויות במדעי החיים מעלות צורך גובר והולך באוריינות תאורטית ומעשית במדעי הנתונים, וצורך זה צפוי לגבור בעתיד. בסקר שעשינו בקרב המוסדות להשכלה גבוהה בישראל קיבלנו תשובות מ-15 אוניברסיטאות ומכללות, הן מן המוסד ממבט כולל שלו והן מפקולטות למדעי החיים ולרפואה. בתשובות צוינו 61 קורסים שלדעת המשיבים עוסקים במדעי הנתונים או מתייחסים אליהם. חלק מן המשיבים הצביעו גם על תוכניות להוספת קורסים בתחום וכן על צרכים בקורסים אחרים. השלמנו מידע זה בבירורים בעל פה. מתברר כי במרבית המוסדות קיימת התעניינות במרכיבי הכשרה שונים היכולים לעמוד יחד תחת המטרייה של "מדעי הנתונים" במובנו הרחב של המונח, אולם בדרך כלל תרגומה לקוריקולום איננו מספק בסיס מושגי ומעשי מגובש.

אנו ממליצים למוסדות השונים להנהיג קורס בסיסי במדעי הנתונים (קורס פרונטלי + תרגול) בלימודי תואר ראשון במדעי החיים וברפואה. קורס זה יתמשך על פני שנת הלימודים כולה במספר שעות שבועי מתאים לחשיבות הגדולה של הנושא ולמנעד הרחב של מרכיביו. תוכניתו צריכה להביא בחשבון רקע חיוני מינימלי, ולא דווקא אופטימלי, במדעים המדויקים.

אנו מציעים כי קורס זה יקנה הבנה של האפשרויות הגלומות במדעי הנתונים לצורך לימוד, ביקורת תוצאות ועשיית מחקר עצמאי במדעי החיים. לשם כך הוא צריך לכלול את אלה:

- א. הכרת מושגים בסיסיים במדעי הנתונים
- ב. יכולת לקרוא בביקורתיות שיטות נתונים ומסקנות המבוססות על מדעי הנתונים במחקרים במדעי החיים והרפואה
- ג. יכולת לתכנן מחקר עצמאי, לעשותו, לנתח את הנתונים ולהסיק מסקנות ביקורתיות – כל זאת בשימוש במושגים ובכלים של מדעי הנתונים
- ד. יכולת לתקשר עם מומחים במדעי הנתונים, להיעזר בהם ולשתף עימם פעולה

אם אין בתואר ראשון במוסד במגמה האמורה קורס חובה בתכנות בסיסי, תשולב הקדמה לתכנות כולל תרגול מטלה בסיסית בשפת תכנות נפוצה בקורס זה, כשיש להביא בחשבון את מספר השעות הנוסף הנדרש בשל הרחבה זו.

3.6 מדעים מדויקים

הפקולטה למדעים מדויקים כוללת באוניברסיטאות השונות תחומי ידע רבים ולא כולם זהים בין כל המוסדות. מתמטיקה, פיזיקה וכימיה הן חלק מפקולטה זו. נוסף על זה מדעי המחשב, סטטיסטיקה ומדעי הסביבה או מדעי כדור הארץ שייכים לפקולטה זו במוסדות שונים. בדרך כלל רמת ההכשרה הבסיסית של הסטודנטים במקצועות הרקע הדרושים לעיסוק במדעי הנתונים היא גבוהה.

אפשרי וראוי שתלמידי מדעים מדויקים יגיעו לרמה מספקת המאפשרת להם לעסוק עיסוק אקטיבי במדעי הנתונים עם סיום לימודיהם, כפי שהוא מתבטא בדיסציפלינה שאותה למדו ואף מחוצה לה. ראוי שהרמה המושגת תאפשר להם התמקצעות במדעי הנתונים לתארים מתקדמים, אם יבחרו בכך, בהשקעה מועטה יחסית. במרבית תחומי הידע בפקולטה ההכשרה במקצועות הרקע היא מספקת. הידע בסטטיסטיקה ובשפות תכנות למדעי הנתונים אינו מספיק בדיסציפלינות כמו כימיה ופיזיקה, אך ניתן להשלימו בקלות יחסית.

רשימת הנושאים והכלים הייעודיים

- * **הגדרת הבעיה:** כל הנושאים והכלים
- * **איסוף הנתונים הרלוונטיים:** כל הנושאים והכלים המוזכרים בפרק 2 (לבד מנושא השימוש במאגרי מידע מקומיים)
- * **שילוב, ניקוי, סינון, וגיבוי של נתונים:** הכרה והבנה של כל הנושאים ויכולת ביצוע ביצירת קובץ נתונים אחרי אינטגרציה; ויזואליזציה של קשר בין כמה משתנים, תרשימי צפיפות וקופסה והתאמת תרשימים לנתונים מממד גבוה
- * **עיבוד נתונים וניתוחם:** הבנה ויכולת ביצוע של מבחנים סטטיסטיים, רגרסיה ליניארית מוכללת ורגרסיה לא פרמטרית, עצי החלטה יער אקראי, הכרה והבנה לצד למידה עמוקה; יכולת ניתוח של נתונים כמותיים וקטגוריים, הכרת ניתוח נתוני תמונה וסוגי הנתונים האחרים; יכולת לנתח לפחות אחד מהם (למשל ניתוח תמונה באסטרופיזיקה, נתונים גנטיים בביואינפורמטיקה)
- * **הוצאת מסקנות מניתוח הנתונים והצגתן:** יכולת ליצור ויזואליזציות דינמיות, להעריך איודאות סטטיסטית וליצור סימולציות, להכיר מערכות לגישה אינטראקטיבית ולהבין עקרונות בקשר אדם-מכונה

* **השלכות ניתוח הנתונים על הידע:** בבעיה ובהיזון חוזר לתהליך. הבנה והכרה של כלי החלטה ופיתוח ושל תוכנות לוח מחוונים; בניית תחזיות אם-אז

* **שיקולי אתיקה לפרט ולחברה:** הכרת כל הנושאים והבנתם ויכולת להעמיק בתחום אחד לפחות

ראשית, אנו ממליצים **להעשיר את הקורסים הקיימים** במדעים המדויקים כדי שיוכלו לתת מענה לדרישות של מדעי הנתונים. העשרת הקורסים הנוכחיים בסטטיסטיקה, בהסתברות ובתכנות בנושאי הרקע החסרים, בצירוף חשיפת הסטודנטים לנושאים ייעודיים למדע הנתונים המופיעים מטה, יביאו את הסטודנטים לאוריינות המבוקשת.

שנית, אנחנו מציעים **קורס ייעודי** במדעי הנתונים שיכול להיות תחליף לקורסים הקיימים או תוספת להעשרתם כאמור לעיל. הקורס עוקב בעיקרו אחרי התוכנית שגיבש פרופ' סהרון רוסט באוניברסיטת תל אביב, ושגרסה ראשונה שלה נלמדה השנה.

* איסוף וארגון של נתונים: סקרים, טבלאות ו־web scraping

* ניקוי והבנה של נתונים: סיכומים, ויזואליזציה ו־principal component analysis

* חשיבה הסתברותית: מומנטים, התפלגויות מותנות ופרדוקס סימפסון

* הסקה סטטיסטית: אמידה ובדיקת השערות

* בניית מודלים על נתונים: רגרסיה ליניארית ולוגיסטית

* כלים ורעיונות בסיסיים בשיטות מודרניות: עצי החלטה, gradient descent,

רגולריזציה ופירוקי bias-variance

* שיטות מודרניות לבניית מודלים: random forest, boosting, deep learning

הקורס יכלול עבודה מעשית נרחבת עם נתונים: ניתוח אירועים מעשיים (חקרי מקרה – case studies) ילווה את הקורס אגב תכנות בסביבת Python (אין נדרשת הכרת Python; הנושא לא יילמד במפורט אלא יסופקו חומרים וסיוע לסגירת פערים עצמית). כמו כן ישולבו בקורס הרצאות אורח על חקרי מקרה נוספים.

הרקע הנדרש: ידע בסיסי באלגברה ליניארית, בחשבון אינטגרלי ובהסתברות; הכרה בסיסית של סביבת מחשב ועקרונות כלליים של תוכנה ותכנות ברמת קורס מבוא למחשבים.

3.7 הנדסה

פקולטה להנדסה היא יצור "מפוצל-אישיות" מבחינת הדרישות ללימודי מדעי הנתונים. מחד גיסא היא כוללת תחומים עתירי חומר במדעי הנתונים, כמו הנדסת תעשייה וניהול והנדסת מחשבים ואלקטרוניקה, ומאידך גיסא קיימים בה תחומים אשר עיבוד נתונים וניתוחם אינם תשתית לימודיהם, למשל הנדסה אזרחית, הנדסה מכנית וארכיטקטורה. לפיכך ראוי שהפקולטה תציע תוכנית מלאה לתואר ראשון (או שני) במדעי הנתונים, ולצד זה תעניק הכשרה מספקת גם לסטודנטים שניתוח נתונים אינו חלק מהכשרתם העיקרית.

כמו במדעים המדויקים, גם ברוב מקצועות ההנדסה ההכשרה הבסיסית מאפשרת התמקצעות במדעי הנתונים לאחר השלמות מסוימות. המתווה הנוכחי מומלץ כמתווה חובה בכל תחומי ההנדסה, להוציא את אלה שמדעי הנתונים קיימים בתשתית ההכשרה המקצועית שבהם.

מומלץ לבנות שני קורסים בהיקף של 4 נקודות זכות כל אחד, שיעסקו בתחומים שונים של מדעי הנתונים. רצוי שהקורסים יכללו שעות תרגול ופרויקט מעשי, שיותאם לתחום הידע וההתמחות בכל חוג או פקולטה (למשל ניתוח נתונים אורבניים יבוצע כפרויקט בלימודי ארכיטקטורה).

להלן סקירה קצרה של הנהוג במוסדות שונים, ובהם מוסדות שבחרו לתת קורס אחד בלבד המתמקד במדעי הנתונים. עם זאת עדיין לא נקבע אף לא באחד ממוסדות אלו קורס אחד (או שניים) כחובת לימודים לרוחב כל בית הספר להנדסה. השיטות השונות אפשריות, אך ראוי שהסילבוס ייבחן לאור הנושאים השונים במחזור הנתונים שנמנו לעיל בפרק 2. מוסד שיבחר ללמד קורס אחד בלבד, ראוי שישלב בו את כל שלביו של מחזור הנתונים, ולו בדגש המתאים למוסד. במקרה כזה ראוי לשלב בקורסים אחרים באופן מובנה פרויקטים העוסקים בניית נתונים (ברוח קורס הפרויקטים המוצע בטכניון – מכון טכנולוגי לישראל). בקורסים אלו יכולים להדריך אנשי התחום המתאים בהנדסה בשילוב עם מתרגלים ממדעי הנתונים. יש חשיבות להשתתפותם של כל סטודנט או סטודנטית בפרויקט אחד כזה לפחות, שעליו יוכלו להצביע כהישג אם יחפשו תעסוקה במדעי הנתונים בתחומם.

החלוקה להלן מוצעת באוניברסיטת תל אביב:

א. קורס שמתרכז בהקניית הנתונים

זיהוי מקורות הנתונים, crawlers, השגת הנתונים ואחסונם, ניקוי (cleansing) וברירת המוץ מהבר ואינטגרציה של נתונים ממקורות שונים (למשל metadata). קורס זה מתמקד בשלבים 1-3 שבמחזור הנתונים.

ב. קורס שמתרכז בעיבוד הנתונים

זיהוי הצרכים, זיהוי האלגוריתמים הנדרשים, תכנות בסיסי ועיבוד, אנליטיקה, ויזואליזציה וקבלת החלטות, למידת מכונה ולמידה עמוקה. קורס זה מתמקד בשלבים 3-6 שבמחזור הנתונים.

חלוקה אחרת מוצעת בטכניון – מכון טכנולוגי לישראל:

א. קורס כללי במדעי הנתונים

כולל איסוף וניקוי של נתונים גולמיים, ייצוג הנתונים ביעילות, ניתוח אקספלורטורי של הנתונים, ניתוח אשכולות ושיטות בסיסיות לסיווג. עוד אמור הקורס להקנות יכולת להבין את האתגרים בהסקת מסקנות וגם את היכולת לתקשר ממצאים באפקטיביות. כך נוגע הקורס בכל הנושאים שבשלבים 2-6 שבמחזור הנתונים. נושא האתיקה מועבר בהרצאה אחת מיוחדת (ומוצע גם כקורס בחירה נפרד; ראו לעיל סעיף 2.2.3 בדוח). קורס כללי זה נשען על ידע של הסטודנטים בשפת Python ומשתמש בחבילות תוכנה שונות להצגת נתונים ולניתוחם. הקורס פותח בידי ד"ר עופרה עמיר.

ב. קורס פרויקטים

מתמקד בבעיות וביישומים בתחומי הלימוד השונים בהנדסה, ובכך מכסה נושא נוסף אגב יישום המיומנויות שנרכשו בקורס הראשון.

ריכוז ההוראה בקורס אחד (3 שעות הוראה + תרגול) נהוג באוניברסיטת בן-גוריון בנגב. בקורס זה, אחרי מבוא כללי ובו דוגמאות מתחומי ההנדסה, עוברים לפרק ארוך יחסית העוסק בתכנון ובבנייה של מסדי נתונים ובשימוש בהם, ולאחר מכן עיבוד נתונים, ניקויים והתמרתם. כמחצית הקורס עוסקת בשיטות ניתוח שונות, מעצי החלטה ו-SVM, דרך רשתות עצביות ולמידה עמוקה ועד ניתוח אשכולות.

קורס זה מדגיש את הנושאים 2-4 שבמחזור הנתונים, ותוכניתו מסוכמת בטבלה להלן:

שיעור	נושא
2-1	מבוא לקורס: סקירה כללית של תחום מדעי הנתונים כולל הרקע ההיסטורי והטכנולוגי ודוגמאות של פרויקטים מעשיים בהנדסה
3	תכנון מסדי נתונים רלציוניים, מודל entity-relationship
4	שליפת מידע ממסדי נתונים רלציוניים, מבוא לשפת SQL, בסיסי נתונים לנתוני עתק ו־NoSQL
5	מחסני נתונים ובינה עסקית (business intelligence)
6	עיבוד נתונים (data manipulation), ניקוי נתונים (data cleaning), זיהוי חריגים (outlier detection) והמרת נתונים (data transformation)
9-7	בעיות סיווג (classification), גישת השכנים הקרובים ביותר (k-NN), עצי החלטה (decision trees), Random Forest ו־SVM
11-10	רשתות נוירונים ולמידה עמוקה
13-12	ניתוח אשכולות (clustering)

פרק 4

המלצות בתחומים משיקים

המלצות בתחומים משיקים

על תוכנית הלימודים להיות פשוטה ליישום כדי שלא תכביד על משאבי המוסדות שיאמצו אותה, ולאפשר לכל מוסד גמישות רבה ביישומה. לכן איננו מציעים תוכנית לימודים מלאה ומפורטת או סדרת סילבוסים מהודקים עד הפרט האחרון אלא אך ורק עקרונות יסוד במדעי הנתונים, המבוססים על מחזור החיים של הנתונים (כפי שהוסבר לעיל בפרק 2).

בעקבות הצעדים שנקטו מל"ג-ות"ת קמו באוניברסיטאות המחקר בישראל מרכזי מחקר למדעי הנתונים, ובחלק מן האוניברסיטאות קמו גם תוכניות לתארים (ראשון או מתקדם) במדעי הנתונים. יוזמות אלה ירכזו באוניברסיטאות את חברי וחברות הסגל בתחום זה ואף יגדילו את מספרם. המרכזים הם על-דיסציפלינריים, ובהם ייוצגו כל תחומי המחקר המיוצגים באוניברסיטה. מטבע הדברים יהיו המרכזים למדעי הנתונים שותפים להקמת הקורסים המומלצים בדוח הוועדה, ולמעשה הצלחתה של היוזמה תלויה במידה רבה בשיתוף פעולה עם המרכזים למדעי הנתונים.

שיתוף הפעולה יתבטא, ראשית, בהתייעצות עם מומחי נתונים בעת הקמת הסילבוס. נוסף על זה, הוועדה רואה ערך בכך שמנסחי הקורס לכל אחת מן הדיסציפלינות יהיו חוקרים או חוקרות המשתייכים לדיסציפלינה. מלבד זאת, הוועדה תותיר לכל מוסד או פקולטה ולכל מרכז מוסדי למדעי הנתונים להחליט על רמת שיתוף הפעולה: אך טבעי שאלו הגופים שיובילו את יישום התוכנית המוצעת ויסייעו בהבניית הקורסים הרלוונטיים, אם לא יעבירו אותם בעצמם. אנו יוצאים מתוך הנחה שהכנתם ושילובם בתוכנית הלימודים של קורסים במדעי הנתונים בכל מוסד ייעשו בין השאר מתוך הישענות על המשאבים של אותם המרכזים.

4.1 שיטות הוראה ועקרונות בהוראת מדעי הנתונים

לימוד על סמך ניתוח נתוני אמת

מדע הנתונים, מטבע הדברים, עוסק בנתונים. החשיפה לנתוני אמת, על האתגרים העולים בנייתוחם, היא רכיב מרכזי בהוראה בתחום. כפי שאי אפשר ללמוד שחייה מחוץ למים, וגם לא בבריכת תינוקות, כך אי אפשר ללמוד ניתוח נתונים ללא נתונים. השימוש בדוגמאות סינתטיות (כפי שנעשה בחלק מספרי הלימוד הוותיקים יותר) גם הוא לא יועיל, שכן יש חשיבות להלימה בין המטרות, שבדרך כלל אינן מוגדרות היטב, לבין הנתונים ושיטות הניתוח, וקשה ללמד זאת ללא התנסות של ממש. אומנם מדע הנתונים הוא מצד אחד דיסציפלינה מדעית בהתגבשותה, אך מצד אחר הוא תחום משולב התומך כמעט בכל דיסציפלינה מדעית אחרת. לכן יש לשלב במהלך ההוראה ניתוח נתונים העולים מבעיות של ממש בתחום הלימוד הראשי של הסטודנטים. כדי להתנסות בשלבים המובנים פחות של הגדרת הבעיה (זיהוי הנתונים, איתורם ואיסופם, ואחר כך ניקוי וסינון) רצוי לשלב במהלך ההוראה פרויקטים מעשיים. האפשרות לקחת פרויקט ניתוח נתונים שילווה את הקורס מתחילתו ועד סופו, כולל הצגת התוצאות בהרצאה ובכתב, היא שהופכת קורס תאורטי לחוויה לימודית יעילה.

לימוד בצוותים

בחיים המחקריים, כמו גם בחיים המקצועיים, ניתוח נתונים הוא עבודה של צוות ולא של איש או אישה העובדים בדד. במקרים כאלו יתרום כל אחד מהצוות את תרומתו העיקרית באחד השלבים, לפי מומחיותו, אך גם יהיה משתתף פעיל בכל

יתר השלבים. לכן כחלק מתהליך ההוראה רצוי לשלב צוותים לפרויקטים במקום התמקדות במטלות אישיות. רקע שונה לחברי הצוות רק יסיף לאפשרות ללמוד זה מזה, בהתמודדות משותפת, ולהגביר את יעילות הלימוד.

קורסים מקוונים

קיימים קורסים ממוחשבים רבים בשפה האנגלית, וכן מקודמות כעת תוכניות לפיתוח קורסים נוספים בשפה העברית. מאמצים מבורכים אלה יצליחו להעביר היטב את התכנים הטכניים יותר של הנושא. קורסים שכאלה יאפשרו גם השלמות למי שנחשפו לתחום במהלך לימודיהם ומבקשים להעמיק, הן במסגרת לימודיהם לתארים מתקדמים והן בעיסוקם המקצועי. עם זאת בשל ההבחנות הקודמות בפרק זה, איננו רואים לנכון שהקורסים הממוחשבים יחליפו את ההוראה האישית של קורסים בתחום. ההוראה יכולה להיות מקוונת, אך דרושה אינטראקטייה פרטנית עם סגל ההוראה כדי להפוך את עבודת הצוות לתהליך של לימוד.

4.2 פיתוח קובצי נתונים המיוחדים לישראל והנגשתם

4.2.1 משאבים למחקר במדעי הנתונים בשפות המקומיות: עברית וערבית

תשתית ראשונית נחוצה לצורך מחקר ישראלי במדעי הנתונים, הן בשיטות של עיבוד שפה טבעית (NLP) והן בשיטות בלשניות אחרות, והיא קורפוס מתויג של הלשון העברית, הן המודרנית והן ההיסטורית. קיומו של קורפוס כזה יאפשר איך־ספור פרויקטים של תלמידים בקורסי המבוא למדעי הנתונים המוצעים בדוח זה. למעשה, מדובר במשימה לאומית. קורפוסים דומים קיימים בלשונות אחרות והם מהווים בסיס למחקר ענף; למשל [Penn Parsed Corpora of Historical English](#) או [Penn Treebank](#), ובו טקסט אנגלי מתויג ברמות שונות. ראו גם [ריכוז של קורפוסים היסטוריים](#) ברשת CLARIN. [קבוצה מחקרית אירופית](#) נחלצה אף היא ליוזמה דומה של קורפוסים בשפות רבות. כאמור, הקורפוסים מתעדים טקסט ומתייגים אותו ברמות שונות: ניתוח בסיסי של שמות עצם, שמות מקומות, שמות אנשים וכדומה; ניתוח תחבירי; ניתוח מורפולוגי. הנגשתם דרך ממשק משתמש שאינו דורש תכנות היא תנאי להיותם שימושיים למחקר במדעי הרוח ולציבור הרחב כאחד.

בלשון העברית קיים מאגר ראשוני **קורפוס השפה העברית - תיוג מורפולוגי** ביזמתה של רשות התקשוב הממשלתית, ובשיתוף האקדמיה ללשון העברית. מאגר זה איננו עוסק בלשון היסטורית אלא רק בלשון ממשלתית עכשווית. מפעל ותיק יותר הוא **מיל"ה מאגרים** - מפעל המילון ההיסטורי ללשון העברית - מאגר מידע לשוני רב ומנותח, אולם הוא פתוח לקהל רק דרך הממשק של האתר ואינו פתוח למחקרים באמצעות קישור לפלטפורמות אחרות. יש לציין גם את הפרויקט בהתווית **Jerusalem Corpus of Emergent Modern Hebrew** של האוניברסיטה העברית בירושלים, המציע קורפוס מובנה חופשי לשימוש של השפה העברית המתחדשת.

הלשון העברית זוכה אף היא לקורפוסים מנותחים, הזמינים למחקר ממוחשב, כגון פרויקט Kitab של **Open ITI**. פרויקטים אחרים מוקדשים ליצירת קורפוסים מנותחים של ערבית קלאסית ומודרנית בדיאלקטים שונים, כולל קורפוס של **הערבית הפלסטינית**.

היזומה להטיל לימודי מידע על כל תלמידי האוניברסיטה בארץ מחייבת המשך פיתוחם וביסוסם של הקורפוסים הללו בעברית ובערבית. הוועדה ממליצה לרשויות לתמוך בהם ביתר שאת.

4.2.2 נתוני משפט

נתונים על הליכים משפטיים הם תחום פורה למחקרים עשירים בנתונים, והם גם בעלי ערך רב עבור החברה האזרחית בישראל, באשר הם עשויים להצביע על אפשרויות לשיפור המערכת וייעולה ולמניעת עוולות וכן על כיווני פיתוח חדשים. חוקרים רבים בתחום מדע המשפט משתמשים בגישות אמפיריות, דהיינו מחקרי משפט מבוססי-נתונים. עבור סטודנטים למשפטים מדובר באוצר בלום של חומר גלם לצורך המשימות במדעי הנתונים שיידרשו להן לפי הדוח הנוכחי. זהו חומר רגיש בכל הנוגע לשמירה על פרטיות האזרחים, ולכן יש לחדד את נושא הפרטיות בקורס של מדעי הנתונים בתחום המשפט ולהטמיע נורמות אתיות עם התרגול. מנגד, כיוון שרוב ההליכים המשפטיים הם פומביים על פי חוק, יש לעודד את חופש המידע גם בתחום זה.

לשם כך יש צורך בכמות רבה של מטא-דאטה על הליכים משפטיים. המאגרים הרלוונטיים לזה הם, ראשית כול, פסקי הדין וההליכים של בתי המשפט ובתי הדין השונים, ולצידם המערכת של משטרת ישראל, שירות בתי הסוהר, פרקליטות המדינה וכדומה. הצלבת נתונים ממקורות שונים – גם אם חלקם פומבי – עלולה להפוך נתונים תמימים ומותממים למידע רגיש ומזוהה. שיטות סטטיסטיות וטכנולוגיות שונות להתממת נתונים, בהקפדה על אבטחת מידע, ויישום והטמעה של כללים ברורים למי שנחשפים למידע – עשויים לצמצם את הסכנות שבעיבוד מידע אישי. מומלץ לבחון את הסוגיה באמצעות ועדות האתיקה המוסדיות.

מדינת ישראל מעמידה לרשות הציבור בלא הגבלה פסקי דין ומידע ראשוני על ההליכים שמהם ניתן לחלץ מטא-דאטה, אולם כיום המדינה אינה מציעה רשמית גישה למטא-דאטה שיש להניח שנוצר אצלה. הנושא רגיש, ורשויות החוק נותנות דעתן עליו בשנים האחרונות. יש מחלוקת גוברת באשר למידה שבה רצוי להנגיש מידע זה בשל עקרונות פומביות הדיון וחופש המידע מחד, והזכות לפרטיות מאידך. מידע על הליכים פליליים ועל החלטות בהליכים כאלו ניתן לקבל כיום גם מכוח "תקנות המרשם הפלילי ותקנת השבים (מסירת מידע לחוקר מדעי תשמ"ו 1986)", כשמדובר במידע על המרשם הפלילי, או באמצעות בקשות לחופש מידע.

שתי יוזמות להנגשת מידע בולטות בתחום זה. אין מדובר ביוזמות ממלכתיות אלא במאגרים המשתמשים בנתונים פתוחים ורשמיים בדרך חוקית. יוזמה ראשונה היא [The Israel Supreme Court Database](#), הפועל בחסותה של האוניברסיטה העברית בירושלים. כלי חשוב נוסף הוא [תולעת המשפט](#), יוזמה עצמאית ופרטית של פלטפורמה להנגשת מידע שנאסף ממערכות מידע ממשלתיות.

4.2.3 מקורות לנתונים למדעי החברה

דוגמאות למקורות לנתוני עתק בישראל:

* [מאגרי מידע ממשלתיים בכל התחומים](#)

* [מאגר הנתונים המרכזי של הלשכה המרכזית לסטטיסטיקה](#)

* [מאגר נתוני האינטרנט בישראל](#)

* [מאגר הנתונים למדעי החברה](#)

* [מאגר נתוני הבורסה בת"א](#)

* [מאגר נתונים של המוסד לביטוח לאומי](#)

נוסף על מאגרי הנתונים הציבוריים (דוגמאות לעיל) קיימים מאגרי נתונים בבעלות פרטית, בתשלום. חשוב להדגיש כי גם אם אין מאגר נתונים, ניתן להקים מאגר כזה בעזרת מיומנויות בתחום מדע הנתונים בשילוב כוחות בין מומחי תוכן לבין מומחים בתחום מדע הנתונים. לדוגמה, מי שמעוניין לחקור היבטים סמנטיים של הזמר העברי יוכל לכתוב תוכנית מחשב קצרה (או לבקש ממומחה מדע הנתונים לבנות תוכנית כזאת) שתאסוף עבורו בשיטתיות נתונים מאתרים כגון שירונט. ואם חושבים על דוגמאות שאינן מוגבלות לישראל, למשל: מי שמתעניין בתועלת הנגזרת מתהליכי רישום פטנטים יוכל להכין תוכניות מחשב שיכרו נתונים ממאגרי פטנטים חופשיים ברשת וממאגרים כלכליים/עסקיים ולבצע מיוזג נתונים לניתוח כזה.

4.3 סינרגייה עם ספריות אקדמיות

הספריות האקדמיות משרתות שני קהלים עיקריים: סטודנטים וסגל אקדמי. שירותי הספריות כוללים פיתוח והנגשה של תשתיות מידע פיזיות ודיגיטליות וכן תמיכה בתהליכי לימוד ומחקר. התקדמות המחקר המדעי וההוראה האקדמית מחייבת את הספריות למעורבות עמוקה בתהליכים האקדמיים כדי להמשיך ולתת את השירות המיטבי בכל תחומי הדעת. צמיחה מהירה במיוחד חלה בעשור האחרון בתחום מדע הנתונים ובתחום יישומי יותר של ניתוח נתוני עתק. מרכזי מחקר ותוכניות לימוד ייעודיות קמו ומוסיפים וקמים בכל האוניברסיטאות בישראל. לפי המלצת הוועדה הנוכחית, על הספריות האקדמיות להיערך למתן שירותים מסורתיים ולפיתוח שירותים חדשים כדי לשרת את ההתקדמות המהירה של מדע הנתונים במחקר ובהוראה. להלן נתאר כיצד יכולות הספריות לספק את התשתיות ואת התמיכה הנדרשות:

פיתוח תשתיות דיגיטליות

א. פיתוח והנגשה של מקורות מידע אקדמי איכותיים על מדעי הנתונים (ספרים, כתבי עת וכינוסים): זהו תפקיד מסורתי של ספריות אקדמיות – להיות פילטר ממליץ, מכוון ומייעץ בנושא של מקורות מידע אקדמיים. בתחום של מדעי הנתונים יש עומס אדיר של מידע. חשוב שהספרייה תהיה פעילה בהכוונה פעילה של חוקרים וסטודנטים למקורות מעולים. עם הגידול בפרסומים

אקדמיים בגישה חופשית (open access) על הספריות להיערך למתן גישה לתחום מתרחב זה, ואף לתרום לו בהקמת כתבי עת וארכיבים חדשים בענפי ידע שבהם התחום עדיין לא התפתח.

ב. ריכוז מקורות מהעולם לנתונים למחקר והוראה בכל תחומי האקדמיה. פעילות של ספריות אוניברסיטאיות לדוגמה:

[Clemson Libraries, Social Science Data Sources: US Data Sources](#)

[Vanderbilt University, Data Sets for Social Sciences](#)

[California Polytechnic State University, Data Sources & Repositories](#)

ג. מתן תמיכה לבניית קורסים "רגילים" ומקוונים בנושאים השונים של מדעי הנתונים – חומרי הוראה, תוכנות והכוונה לבניית קורסים מתוקשבים.

ד. פיתוח מאגר לאיסוף שיטתי של קובצי נתונים וקוד שפותחו במסגרת מחקר באוניברסיטה.

מחקר עצמאי

א. מחקר ייחודי בתחום מדעי הרוח והחברה הדיגיטליים המבוסס על אוספים מיוחדים בספרייה, טקסטים ופרויקטי דיגיטציה: מחקר מסוג זה יכול להיעשות בידי חוקרים באוניברסיטה, בידי ספרנים או בשילוב של השניים. מרכזי מחקר (Center for Digital Scholarship) מתחילים לצוץ בספריות אקדמיות ברחבי העולם, לדוגמה:

[Columbia University Libraries, Digital Scholarship](#)

[Brown University Library, Center For Digital Scholarship](#)

[Princeton University, The Center for Digital Humanities](#)

ב. שימוש בנתוני הספרייה המצטברים לאורך שנים רבות כדי לשאול שאלות מעניינות הן מבחינה ניהולית (ניתוח ביקושים למגוון השירותים, ניתוח תמורה לרכישות, יעילות השירות וכו') והן מבחינה מחקרית (מחקרים בספרנות, במדעי הרוח הדיגיטליים, ביבליומטריה וכו'); שימוש במאגרי המידע הספרייתיים כגון WoS, Scopus, Proquest לצורך מחקרי טקסט ורשת רחבי-היקף.

פיתוח תשתיות פיזיות

- א. ייעוד מקום פיזי למעבדת נתונים במדעי הרוח והחברה הדיגיטליים – לתלמידי מחקר, יזמויות וסדנאות.
- ב. פיתוח מתחם עִשְׂיִינִים (מייקרים) הכולל מדפסות תלת־ממד, רובוטים או ציוד אחר לפי צורכי המחקר וההוראה.

פיתוח הדרכה

- א. בניית מרכז משאבי למידה עצמית של מדעי הנתונים הכולל קורסים אקדמיים מקוונים והשתלמויות מקצועיות איכותיות אונליין, מהארץ ומהעולם. במקום שכל אחד מהחוקרים ומתלמידי המחקר יפלוס לעצמו את הדרך ויחפש שוב ושוב את אותם חיפושיים, תוכל הספרייה להעניק שירות איכותי. כדאי שהמערכת שתיבנה תכיל המלצות. מובן שמומחי הספרייה יכולים (וצריכים) להבין את הצרכים של השטח בעניין זה לפני בניית המשאב.
- ב. שיתוף פעולה עם סגל אקדמי לבניית קורסים ייחודיים המשרתים את המחקר ואת ההוראה המקומיים (לדוגמה: פיתוח שאלות מחקר, סקירת ספרות, כתיבה אקדמית, שימוש בכלי ניתוח נתונים, ניתוח טקסטים וויזואליזציה).
- ג. שיתוף פעולה בקיום ימי עיון, כינוסים, סדנאות, האקתונים ואירועים אחרים המחברים בין קהילות שונות מהקמפוס ומחוצה לו סביב נושא מדעי הנתונים.
- ד. קשר עם ספריות בארץ ובעולם כדי לאגם ידע ומקורות שאינם זמינים בארץ.

4.4 מדעי הנתונים בחינוך הטרום־אוניברסיטאי

כדי לקדם את מטרת הוועדה – להטמיע את מדעי הנתונים כמיומנות בסיסית אצל כל סטודנט באקדמיה הישראלית – מצאה הוועדה לנכון לדון בשילובם של מדעי הנתונים גם במערכת החינוך הכללית, הטרום־אוניברסיטאית. לצורך זה נערך דיון בתאריך 26.1.2020 בהשתתפות ד"ר אבי כהן (מפמ"ר מדעי המחשב, משרד החינוך), גב' רונית נחמיה (מפמ"ר גילוי ואיתור מידע דיגיטלי, משרד החינוך) וד"ר יוסי בן־דב (מנכ"ל בית הספר הריאלי, חיפה). להלן המלצות הוועדה בעקבות הדיון:

הוועדה מודאגת מן הביצועים הגרועים של תלמידי ישראל במבחנים בין-לאומיים של אוריינות בטכנולוגיית מחשב ומידע. על מנת שהאקדמיה תוכל להקנות להם יכולות מתקדמות בתחום זה, יש לוודא כי היסודות ניתנים לתלמידים כבר בשלב מוקדם. משבר הקורונה שחוינו ב־2020 יכול להיות נקודת זינוק טובה, עם ההכרה כי הלמידה מרחוק מחייבת השגת מידע גם מחוץ לתהליך הפרונטלי.

המלצות הוועדה אינן נוגעות לתחום ההוראה של מדעי המחשב בבתי הספר התיכונים. ההמלצות אינן נוגעות לשימוש בשפות תכנות בקרב התלמידים שמתמחים בזה. תחת זאת הוועדה מתמקדת בגישה למידע ובשימוש ראשוני בו, ברמה של הפעלת גיליונות נתונים (למשל Excel), זיהוי מקורות מידע וכרייה של מידע, קישור בין נתונים בטבלה לבין גרפים, ויזואליזציה של מידע ושאר הגורמים הקשורים למחזור חיי הנתונים, שתוארו בפירוט לעיל. טוב עשה משרד החינוך שהקים תחום דעת לגילוי ולאיתור של מידע דיגיטלי ומינה מפמ"ר לצורך זה. המלצות הוועדה להלן נוגעות לעבודה בתחום דעת זה.

הדרך הנכונה להטמיע יכולות כאלה בקרב התלמידים איננה בשיעורים המיועדים למדעי המחשב או למדעי הנתונים אלא בעבודה יום-יומית בשימוש בנתונים בכל אחד מן התחומים הנלמדים בבתי הספר, מהיסטוריה ועד מתמטיקה. בתי הספר (המעטים) שעברו ללמידה באמצעות פרויקטים רב-תחומיים יכללו מיומנויות של מדעי הנתונים בתוך כל אחד מהפרויקטים.

יש ללמוד ולתרגל שימוש במידע כבר בבית הספר היסודי, ובוודאי בחטיבת הביניים. כמו כל מיומנויות תשתית אחרות, כגון אוריינות של קריאה או התבטאות בכתב ובעל פה, אין להותיר את השימוש בנתונים ללימוד מרוכז בבית הספר התיכון לקראת בחינות הבגרות. הוועדה תומכת במאמצי של הגב' רונית נחמיה, האמונה בשנים האחרונות על מאמצי משרד החינוך בכיוון זה, וכינסה ועדה מקצועית של מומחים כדי לגבות את מאמצייה.

לביצוע רציני של הטמעת מדעי הנתונים במערכת החינוך יש להכשיר מורים בנושא זה, הן אלה שהם כבר חלק מן המערכת והן אלה שעוברים כעת את ההכשרה להיכנס אליה במכללות למורים. הגב' נחמיה מדווחת כי היא מתקשה לאתר מורים הבקיאים במדעי הנתונים ואשר מסוגלים להפעיל את הכלים

הבסיסיים של טיפול בנתונים. לטעמנו, הקושי נובע – שוב – מהעובדה שהמגמות של "טכנולוגיה בחינוך" וכיו"ב נקלעות לאי בודד של מומחים לטכנולוגיה ואינן נחלתם של כלל המורים. כך נוצרים שוב ושוב מורים "טכנופוביים", המנחילים את הרתיעה לתלמידיהם. כל מורה, של כל קבוצת גיל, בכל תחומי הלימוד בארץ, צריך לעבור השתלמויות של שימוש במידע כדי שיוכל להנחיל זאת לתלמידיו.

דיונה של הוועדה לא הצליחו לברר עד כמה המכללות להוראה פועלות ברוח זו. ייתכן שכן. על כל פנים, אשר למורים שכבר מצויים בתוך המערכת, מתברר שעדיין אין גוף מרכזי שמחזיק בידע וביכולות הלוגיסטיות להטמיע את הנושא של מדעי הנתונים בקהילת המורים. הוועדה ממליצה להקנות לכל ציבור המורים הכרה בסיסית של הנושא באמצעות MOOC, קורס פתוח מקוון מרובה-משתתפים המיועד למטרה זו.

יש לחשוף את מדעי הנתונים לא רק למורים המקצועיים אלא גם לדרג הפיקוח והניהול, שיתנסו אף הם בכינוסים, בהרצאות, ב-MOOC ובלמידה אחרת בתחום זה.

למותר לציין כי על מנת שהשינוי יתרחש, יש לצייד את בתי הספר – המורים והתלמידים כאחד – בהתאם. ההצטיידות תכלול מחשבים, מחשבי לוח (טבלטים), חיבור למרשתת (אינטרנט) ואת התוכנות הנדרשות.



רשימת הכותבים והיועצים

- פרק 1 חשיבותם של מדעי הנתונים
פרופ' יואב בנימיני
- פרק 2 עיקרי הגישה המוצעת להוראת מדעי הנתונים בקמפוס
פרופ' ניב אחיטוב, פרופ' יואב בנימיני ופרופ' דפנה רבן
- סעיף 3.2 מדעי הרוח
פרופ' יונתן בן־דב ופרופ' יגאל ברונר
בהתייעצות עם ד"ר רננה קידר (האוניברסיטה העברית בירושלים)
וד"ר איתי מרינברג־מיליקובסקי (אוניברסיטת בן־גוריון בנגב)
- סעיף 3.3 מדעי החברה וניהול
פרופ' ניב אחיטוב ופרופ' דפנה רבן
- סעיף 3.4 משפטים
בהתייעצות עם פרופ' נילי כהן, נשיאת האקדמיה למדעים,
וד"ר רננה קידר (האוניברסיטה העברית בירושלים)
- סעיף 3.5 מדעי החיים ורפואה
פרופ' ידין דודאי ופרופ' רודד שרן
בהתייעצות עם פרופ' מרטין קופיק (אוניברסיטת תל אביב)
- סעיף 3.6 מדעים מדויקים
פרופ' יואב בנימיני ופרופ' רודד שרן
בהתייעצות עם פרופ' סהרון רוסט (אוניברסיטת תל אביב)
- סעיף 3.7 הנדסה
פרופ' ניב אחיטוב ופרופ' יואב בנימיני
בסיוע פרופ' עודד מימון (אוניברסיטת תל אביב)
-



- סעיף 4.1 שיטות הוראה ועקרונות בהוראת מדעי הנתונים
פרופ' יונתן בן־דב
- סעיף 4.2.1 משאבים למחקר במדעי הנתונים בשפות המקומיות: עברית וערבית
פרופ' יונתן בן־דב, בהתייעצות עם ד"ר סיני רוסינק
(מעבדת אליהו והתוכנית למדעי הרוח הדיגיטליים באוניברסיטת חיפה)
- סעיף 4.2.2 נתוני משפט
פרופ' יונתן בן־דב
בהתייעצות עם פרופ' אורן גזל־אייל (אוניברסיטת חיפה),
פרופ' מיכאל בירנהק (אוניברסיטת תל אביב),
אנדי וורמס וגיא זומר (התמנון – מידע ציבורי לכל [ע"ר])
- סעיף 4.2.3 מקורות לנתונים למדעי החברה
פרופ' ניב אחיטוב ופרופ' דפנה רבן
- סעיף 4.3 סינרגייה עם ספריות אקדמיות
פרופ' דפנה רבן
- סעיף 4.4 מדעי הנתונים בחינוך הטרנס־אוניברסיטאי
פרופ' יונתן בן־דב



האקדמיה הלאומית הישראלית למדעים
THE ISRAEL ACADEMY OF SCIENCES AND HUMANITIES

כיכר אלברט איינשטיין, ת"ד 4040, ירושלים 9104001
טלפון 02-5676222, דוא"ל info@academy.ac.il

www.academy.ac.il

