



**הטכניון**  
מכון טכנולוגי  
לישראל

# אתגר האינטרדיסציפלינריות במדעי הנתונים

שיעור 2 – 27.3.2022

# תוכן

- רקע: X-תחומיות
- אתגרים:
- תפיסת היחס בין מדעי הנתונים למדעי המחשב ולסטטיסטיקה
- תפיסת היחס בין מדעי הנתונים לתחום הדעת
- אתגר הוראת תחום הדעת
- אתגר יישום למידה מבוססת פרויקטים

# רקע: X-תחומיות

סדרו את המונחים הבאים בסדר היררכי

על תחומיות

רב תחומיות

בין תחומיות

# האם מדעי המחשב הם

1. בין תחומיים?

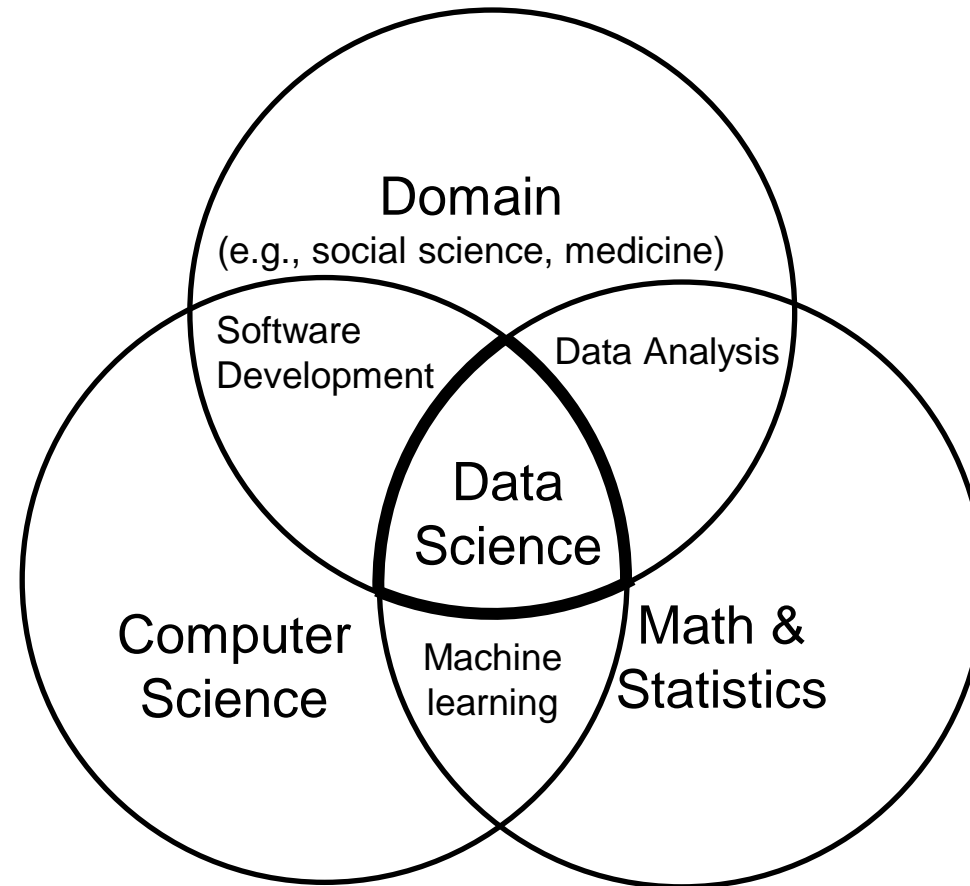
2. רב תחומיים?

3. על תחומיים?

# האם מדעי הנתונים הם

1. בין תחומיים?
2. רב תחומיים?
3. על תחומיים?

# בינתחומיות של מדעי הנתונים



# תפיסת היחס בין מדעי הנתונים למדעי המחשב ולסטטיסטיקה

\* עד כמה אתה מסכים עם ההיגדים הבאים

\* עד כמה אתה מסכים עם ההיגדים הבאים

	במידה רבה מאוד	במידה רבה	במידה בינונית	במידה מועטה	לא מסכים כלל		במידה רבה מאוד	במידה רבה	במידה בינונית	במידה מועטה	לא מסכים כלל
מדעי הנתונים ומדעי המחשב הם שני מדעים נפרדים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	מדעי הנתונים וסטטיסטיקה הם שני מדעים נפרדים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
מדעי הנתונים כוללים את מדעי המחשב	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	מדעי הנתונים כוללים את מדע הסטטיסטיקה	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
מדעי המחשב כוללים את מדעי הנתונים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	סטטיסטיקה כוללת את מדעי הנתונים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
מדעי הנתונים ומדעי המחשב הם תחומים חופפים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	מדעי הנתונים וסטטיסטיקה הם תחומים חופפים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
מדעי הנתונים נסמכים על מדעי המחשב	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	מדעי הנתונים נסמכים על סטטיסטיקה	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
מדעי המחשב נסמכים על מדעי הנתונים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	סטטיסטיקה נסמכת על מדעי הנתונים	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

# תפיסת היחס בין מדעי הנתונים למדעי המחשב ולסטטיסטיקה

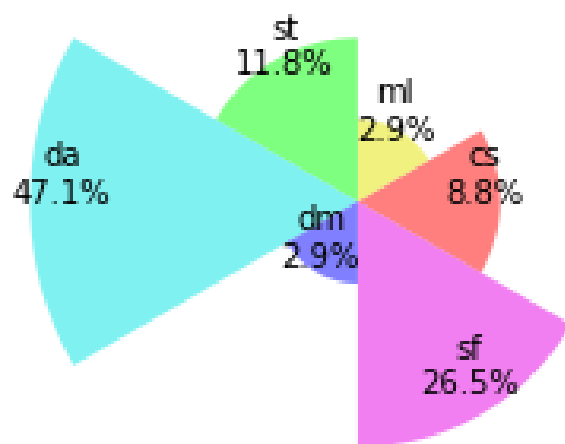
יחס בין מדעי הנתונים ל..	מדעי המחשב	סטטיסטיקה	t-test
מדעי הנתונים הוא תחום דעת נפרד מ..	M=2.95 SD=0.83	M=2.73 SD=1.01	t(106)=1.88,p=0.06
מדעי הנתונים כוללים את ..	M=2.88 SD=1.14	M=3.70 SD=0.96	t(105)=-6.26,p=0.00
מדעי הנתונים נכללים ב..	M=3.12 SD=1.06	M=2.70 SD=1.05	t(105)=3.11,p=0.00
מדעי הנתונים הוא תחום חופף ל..	M=3.07 SD=1.02	M=3.20 SD=0.96	t(103)=-0.85,p=0.40
מדעי הנתונים נסמכים על..	M=3.77 SD=0.94	M=4.04 SD=0.78	t(103)=-2.04,p=0.04
מדעי הנתונים הינם בסיס ל..	M=2.45 SD=1.05	M=2.42 SD=1.00	t(102)=-0.09,p=0.93



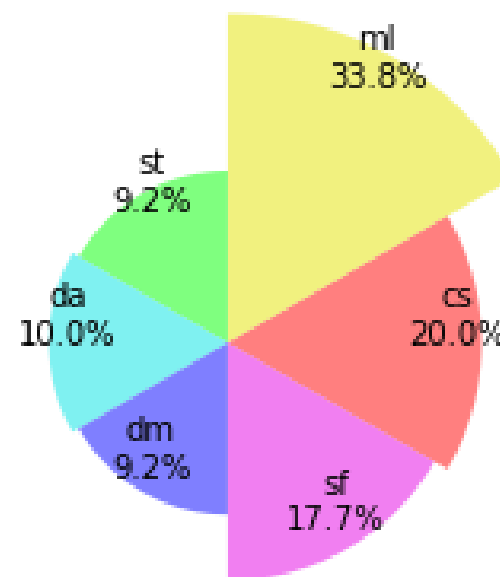
# תפיסת היחס בין מדעי הנתונים לתחום הדעת

# קורס מבוא למדעי הנתונים – ציפיות מהקורס

מבוא לניתוח נתונים פתיחה  
ציין/י 5-3 נושאים שאת/ה מצפה ללמוד בקורס



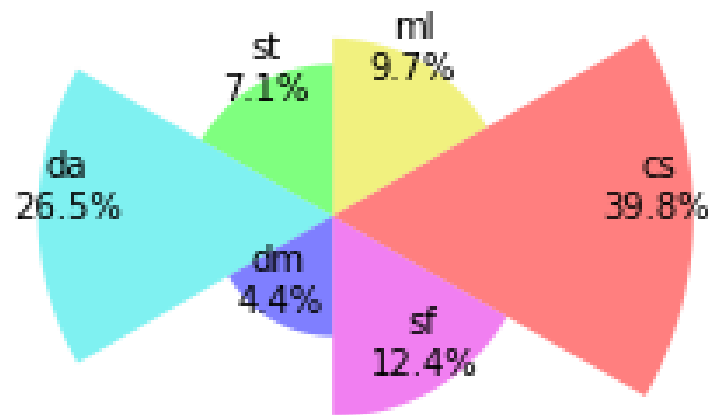
מבוא להנדסת נתונים ומידע פתיחה  
ציין/י 5-3 נושאים שאת/ה מצפה ללמוד בקורס



# קורס מבוא למדעי הנתונים – מושגים הקשורים למדעי הנתונים

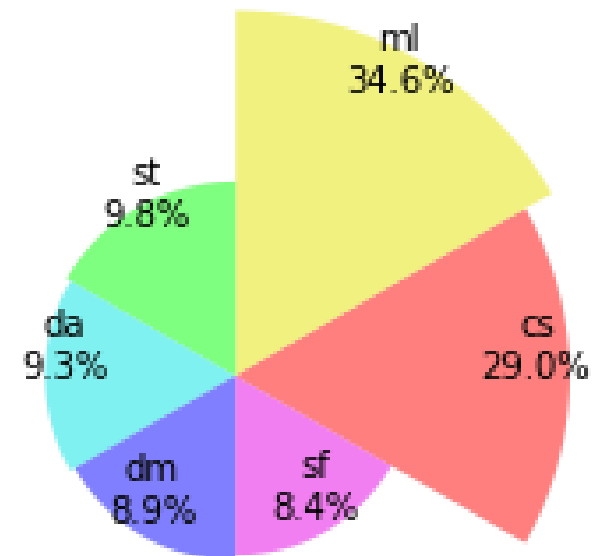
מבוא לניתוח נתונים פתיחה

צינף/י בבקשה עד 5 מושגים הקשורים לדעתך למדעי הנתונים



מבוא להנדסת נתונים ומידע פתיחה

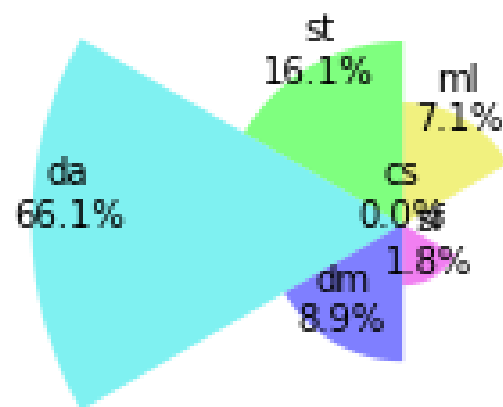
צינף/י בבקשה עד 5 מושגים הקשורים לדעתך למדעי הנתונים



# קורס מבוא למדעי הנתונים – עקרונות חשובים במדעי הנתונים

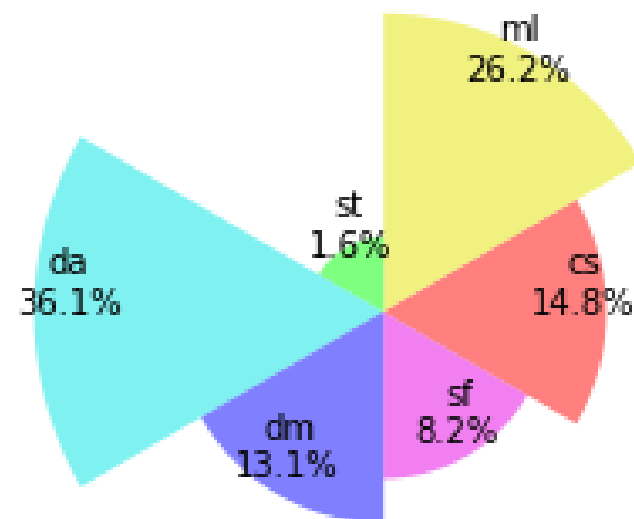
מבוא לניתוח נתונים סיכום

מהם לדעתך 3-5 העקרונות החשובים של מדעי הנתונים?



מבוא להנדסת נתונים ומידע סיכום

מהם לדעתך 3-5 העקרונות החשובים של מדעי הנתונים?



# שאלון שילוב תחום הדעת בלמידת מכונה

<https://docs.google.com/forms/d/1GYsvuWE5BtUQx1wBlv51VMVBkj6T9kxMrg3gw7CDUyg/edit>

# שאלת הרמזורים

חוקרים בחנו מכונות לומדות שמטרתן לזהות את צבע האור ברמזור.

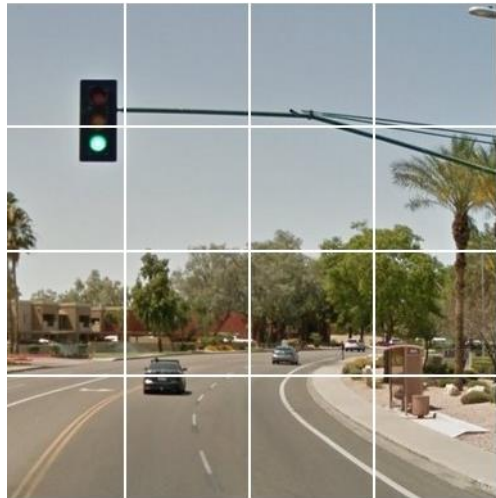
מינימום סיכוי  
לתאונה

מכונה A אינה טועה כאשר האור אדום;  
אולם טועה בהסתברות 0.001 כאשר האור ירוק

מכונה B אינה טועה כאשר האור ירוק;  
אולם טועה בהסתברות 0.0001 כאשר האור אדום.

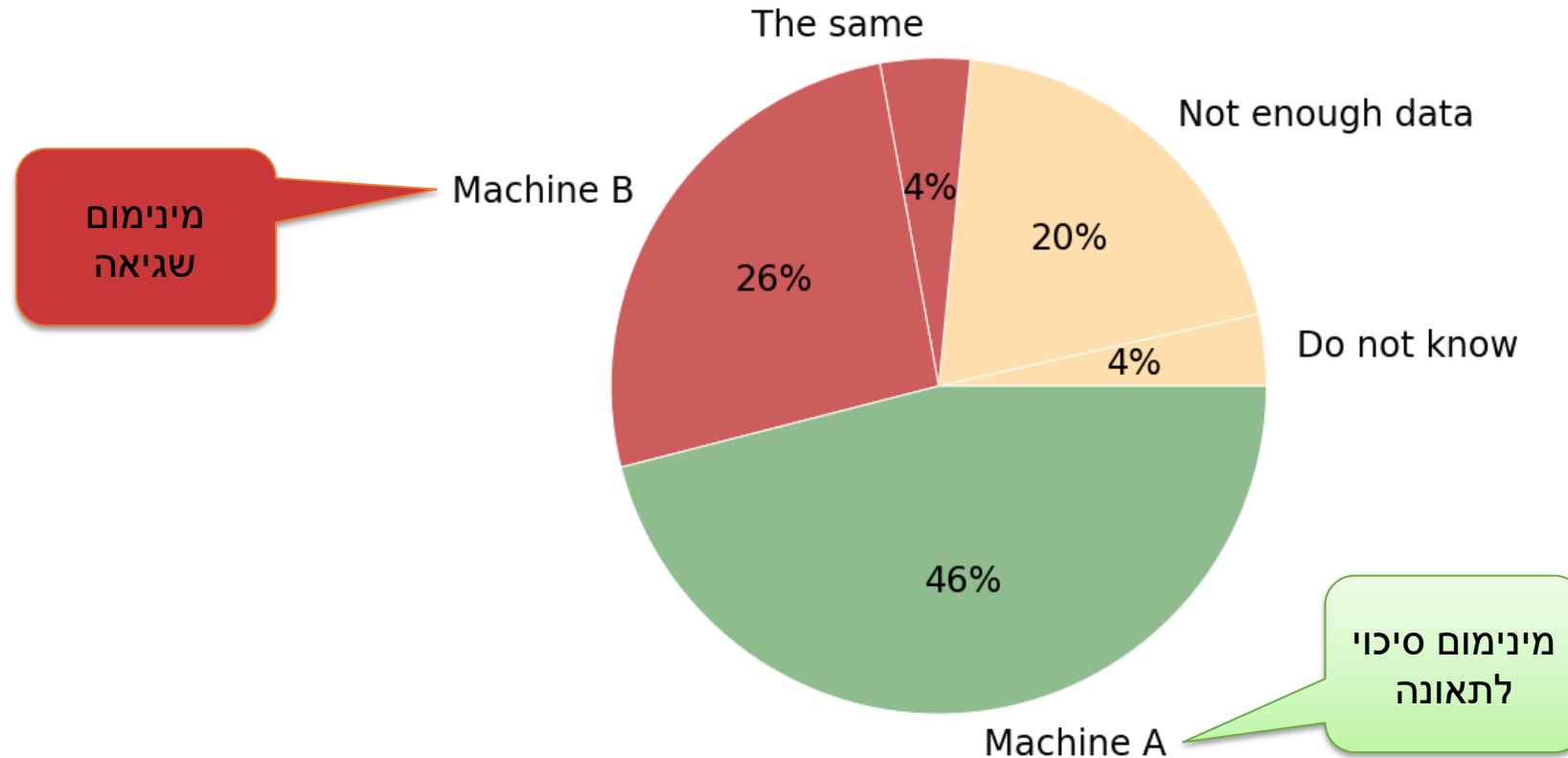
מינימום  
שגיאה

איזו מכונה טובה יותר לדעתך?

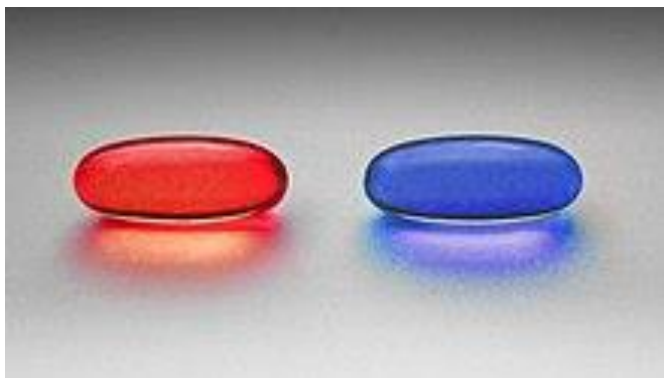


# שאלת הרמזורים - תשובות

Traffic light classification (all ,N=226)



# שאלת הגידולים



חוקרים בחנו מכוונות לומדות שמטרתן לזהות גידולים סרטניים:

- Basal cell carcinoma (BCC)
- Squamous cell carcinoma (SCC)

מסוכן יותר

מכונה A אינה טועה כאשר הגידול הוא SCC;  
אולם טועה בהסתברות 0.001 כאשר הגידול הוא BCC.

שגיאה זהה

מכונה B אינה טועה כאשר הגידול הוא BCC;  
אולם טועה בהסתברות 0.001 כאשר הגידול הוא SCC.

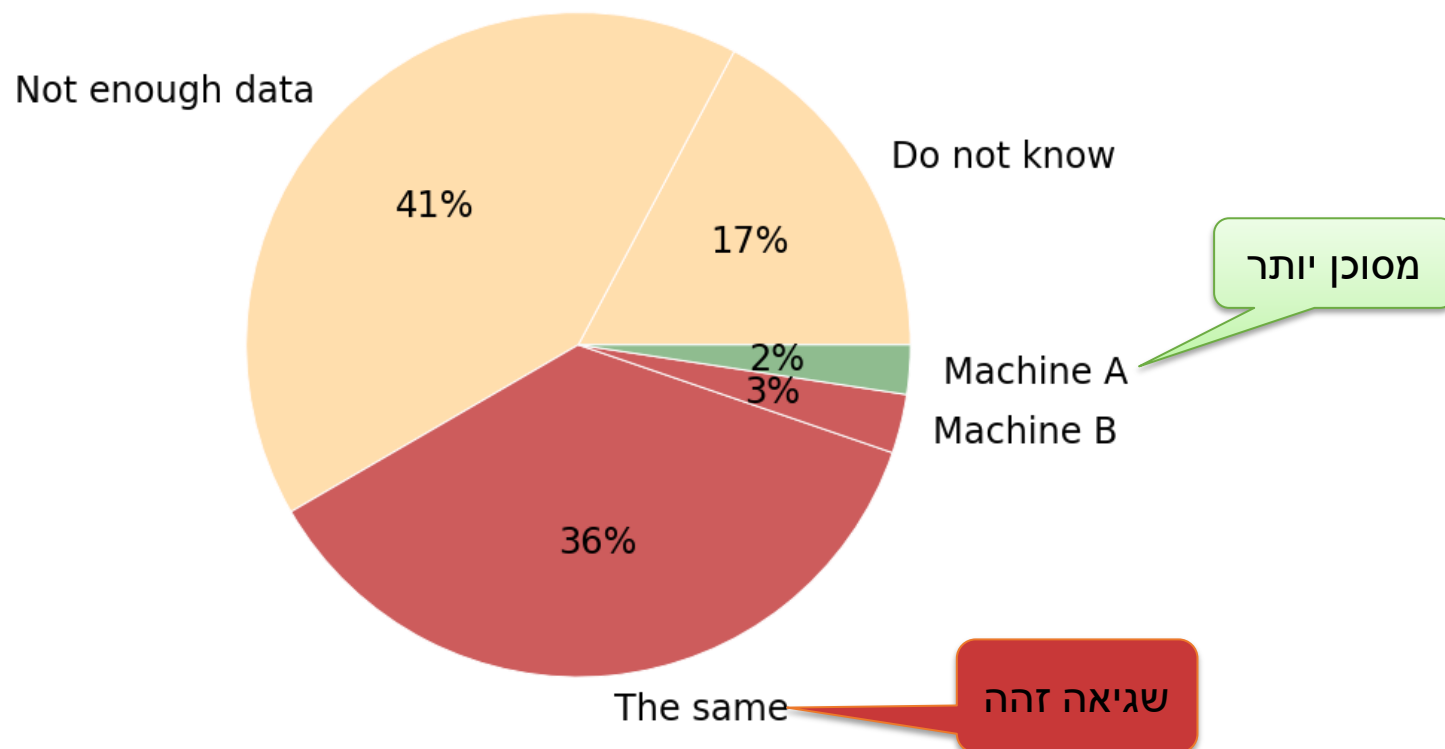
שגיאה זהה

איזו מכוונה טובה יותר לדעתך?



# שאלת הגידולים - תשובות

Carcinoma classification (all ,N=209)



# ניתוח הנימוקים לתשובות

נמצאו שלוש קטגוריות של הסברים:

- **Domain**: מבוססים על שיקולים מתחום הדעת

"רמת המסוכנות של הגידולים לא נתונה בשאלה"

- **Mathematical**: מבוססים על התייחסות לנתונים כאל מספרים ללא הקשר

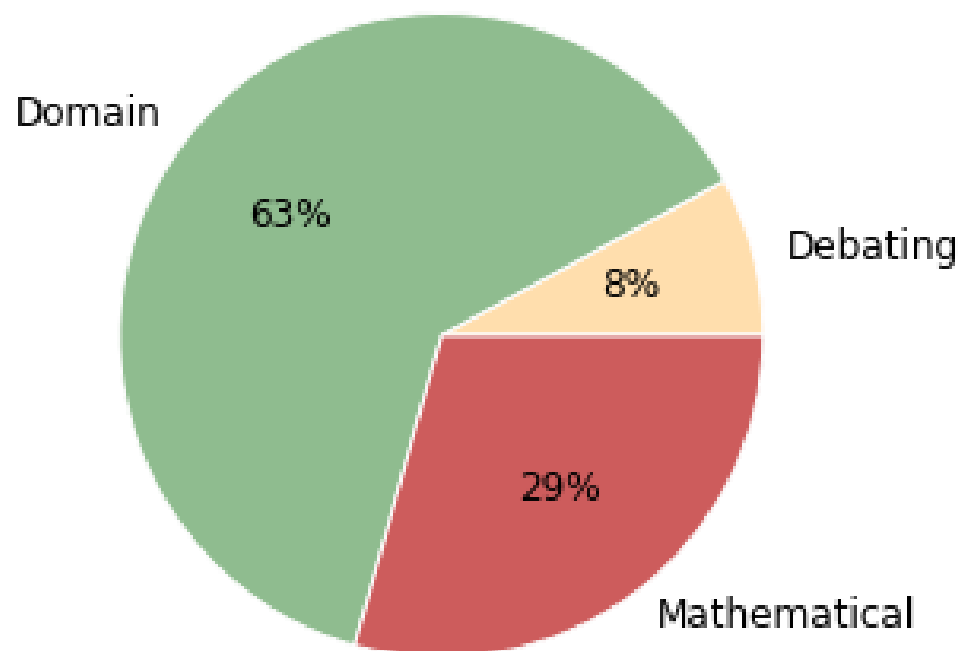
"ההסתברות לטעות זהה בשני המסווגים"

- **Debating**: מציגים גם שיקולים מבוססי תחום דעת וגם שיקולים מתמטיים ומתלבטים מי חשוב יותר

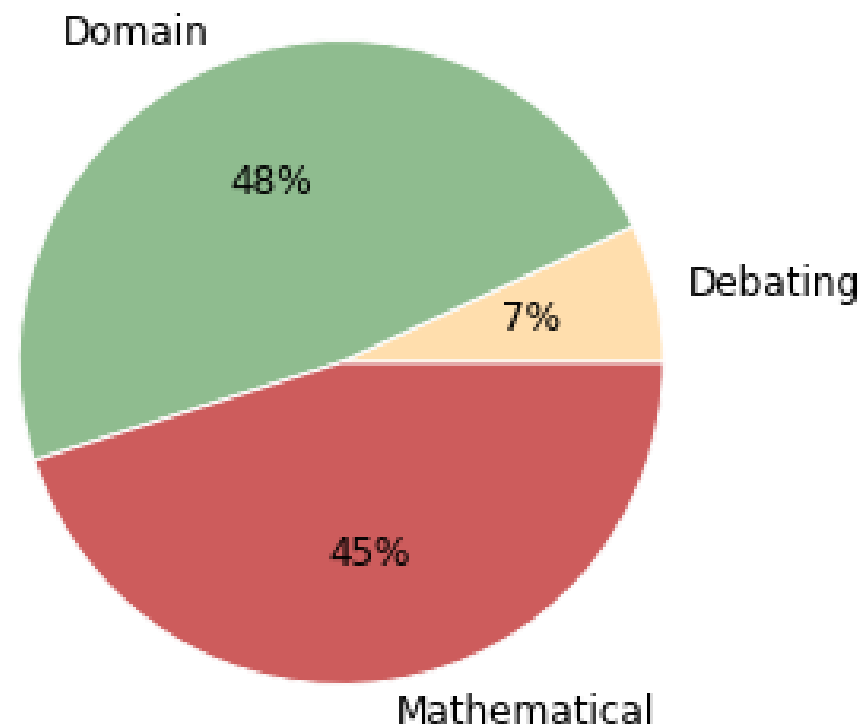
"אני לא בטוחה כיצד ניתן לעשות את המדדים שלמדנו לפי הנתונים הללו, כי מרגיש שחסרים לי נתונים - ניסיתי לחשוב כיצד לעשות זאת עם חישוב של  $100-0.001$  אבל זה פחות מתאים. מבחינת חיי אדם, אני מעריכה שההיגיון אומר שהטעות עם רמזורים אדומים לירוקים הרבה יותר מסוכנת בפועל, מה שאולי יכול להעיד שאם התוצאות יוצאות דומות זה עדיין לא לוקח בחשבון את הסכנה."

# ניתוח הנימוקים לתשובות

Traffic lights explanations, all (N=201)



Carcinoma explanations, all (N=173)



# שאלת גיוס חוקר.ת למידת מכונה

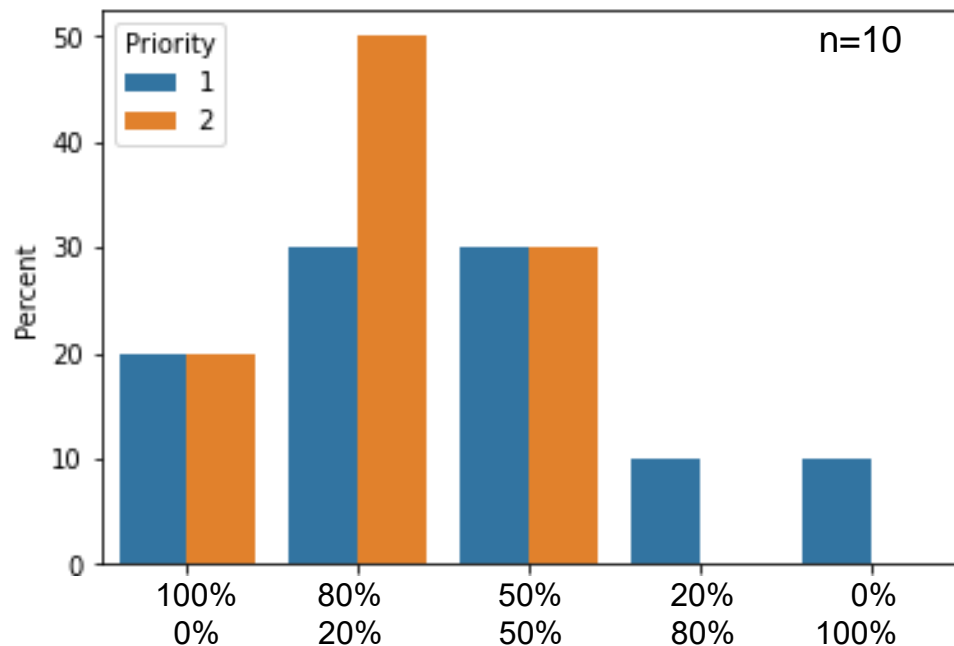
עליך לגייס עובד.ת לצוות למידת מכונה לחברה המפתחת מוצרים ל **בדיקות רפואיות**. מבין 5 המועמדים.ות הבאות עליך לבחור 2. סמני את העדפתך.

- מומחיות בלמידת מכונה: 100%, מומחיות ברפואה : 0%
- מומחיות בלמידת מכונה: 80%, מומחיות ברפואה : 20%
- מומחיות בלמידת מכונה: 50%, מומחיות ברפואה : 50%
- מומחיות בלמידת מכונה: 20%, מומחיות ברפואה : 80%
- מומחיות בלמידת מכונה: 0%, מומחיות ברפואה : 100%

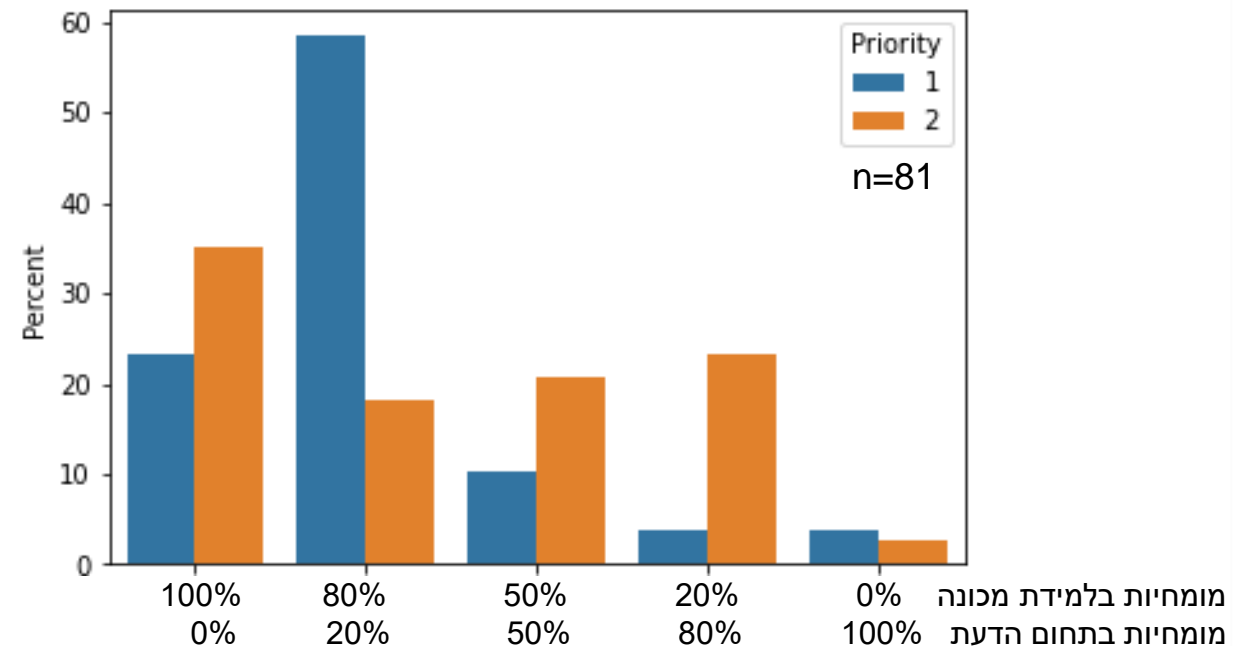


# שאלת הגיוס - תשובות

רפואה



תחברה



אוכלוסייה: סטודנטים.ות למדעים

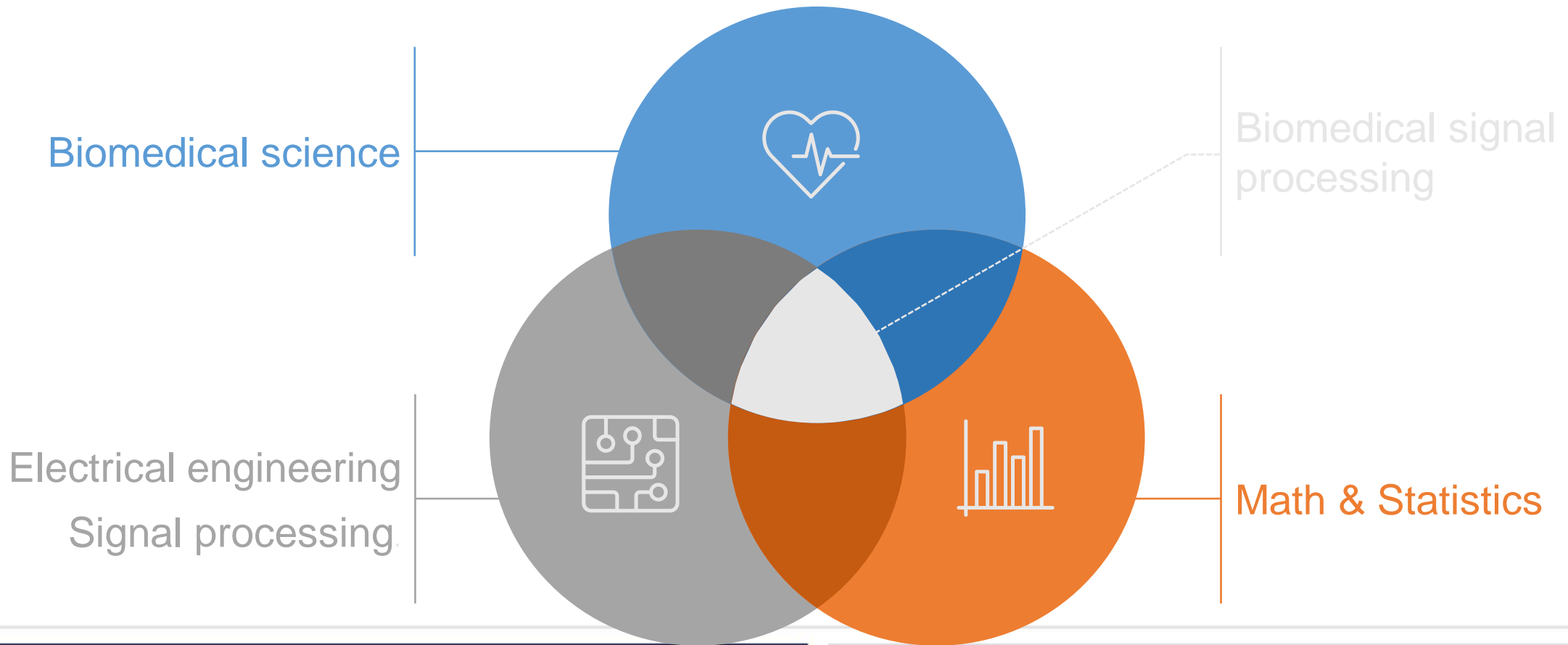
# אתגר הוראת תחום הדעת

תפקיד החוקר	תחומי דעת מהם מוצגים נתונים	קורס
סטודנט	אינטרנט, בידור, ביו-רפואה, בלשנות, היסטוריה, זיהוי תמונה, חינוך, חיפוש, כלכלה, מדיה חברתית, מדעי החברה, מדעי הסביבה, מחקר, ניתוח טקסט, ניתוח תמונה, ספורט, עיבוד תמונה, פוליטיקה, רפואה, תחבורה	The data science design manual, Skiena, 2017
סטודנט	בוטניקה, ביולוגיה, בריאות הציבור, גאוגרפיים, ולידציה, חברה, חינוך, כלכלה, כלכלה/בריאות, כללי, מגדר, מוזיקה, מטרולוגיה, מסחר, משפטים, ניהול ייצור ותחזוקה, ניהול שירות לקוחות, ניתוח תמונה, פוליטיקה, פסיכולוגיה, פרמצבטיקה, קולנוע, רפואה, רשתות חברתיות	ניתוח נתונים בשפת פיתון, טכניון, אביב תשע"ט
צופה	בידור, ביולוגיה, זיהוי כתב יד, חיפוש טקסטואלי, כלכלה, מזון, מסחר, ראייה, רפואה, תעופה	מבוא למדעי הנתונים, טכניון, אביב תשע"ט
מרצה	בוטניקה, זיהוי תמונה, יהלומנות, ספנות	השתלמות מורים למדעי המחשב בניתוח נתונים ולמידת מכונה, האוניברסיטה הפתוחה, יולי 2021

# אתגר יישום למידה מבוססת פרויקטים

- מחקר הוראת דימות ביו-רפואי
- מחקר שילוב תחום הדעת בפרויקטים

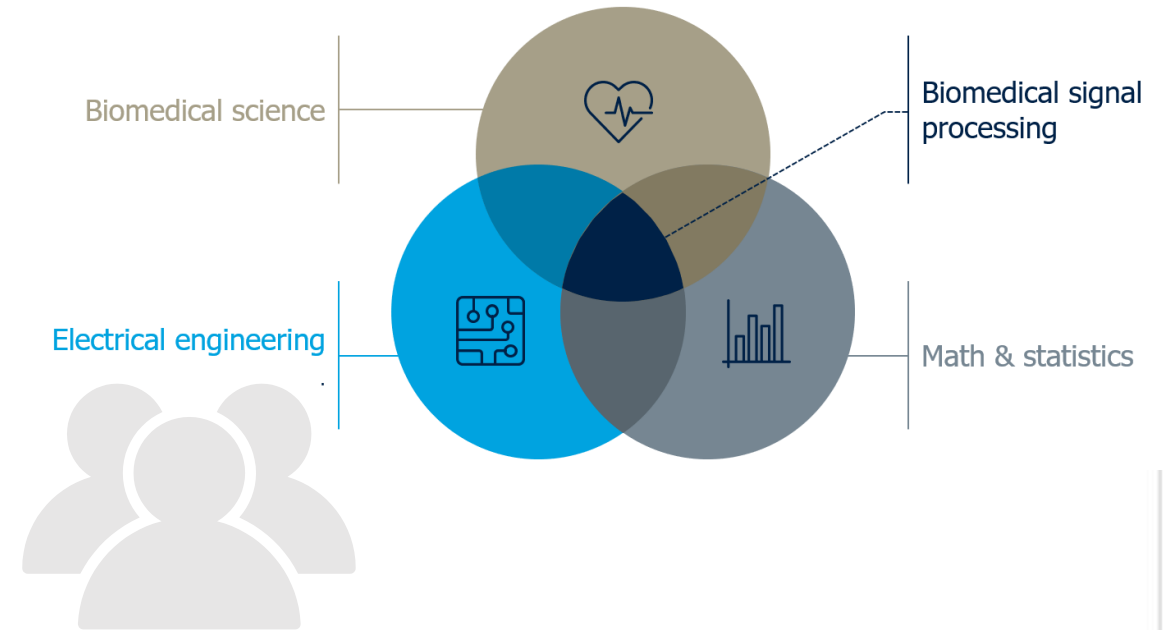
# מחקר הוראת דימות ביו-רפואי





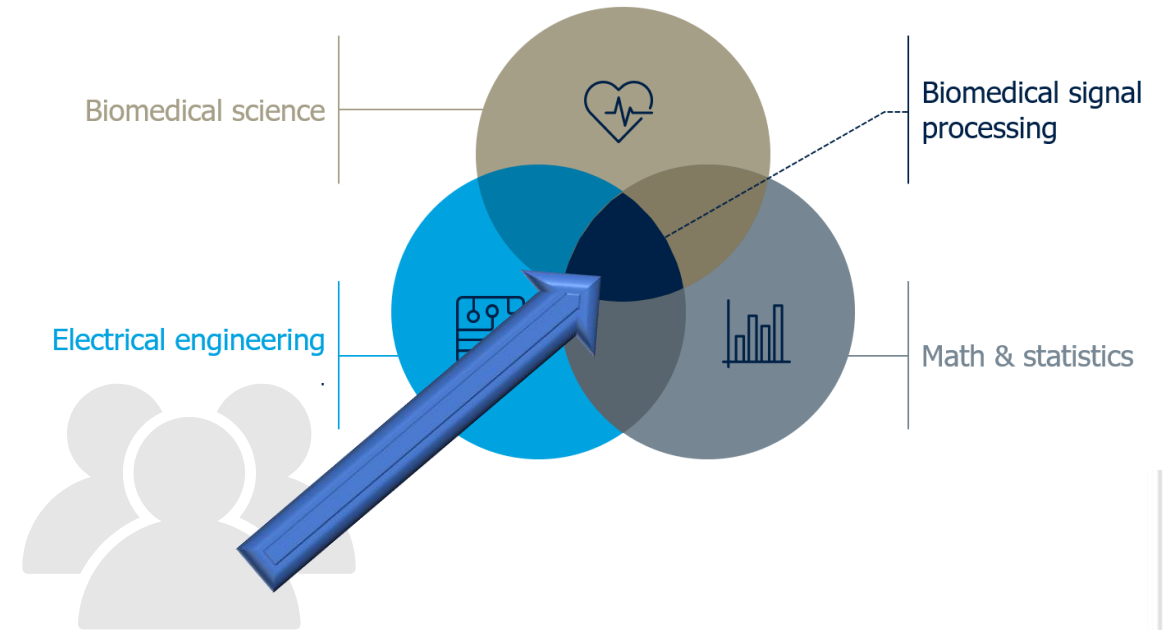
# Research Problem - Homogeneous Teams

- Lack of required knowledge
- Lack of awareness to the lack of knowledge
- Awareness appears only too late into the project
- Learning is only partially gained



# Research Target: Develop and Evaluate an Intervention Program

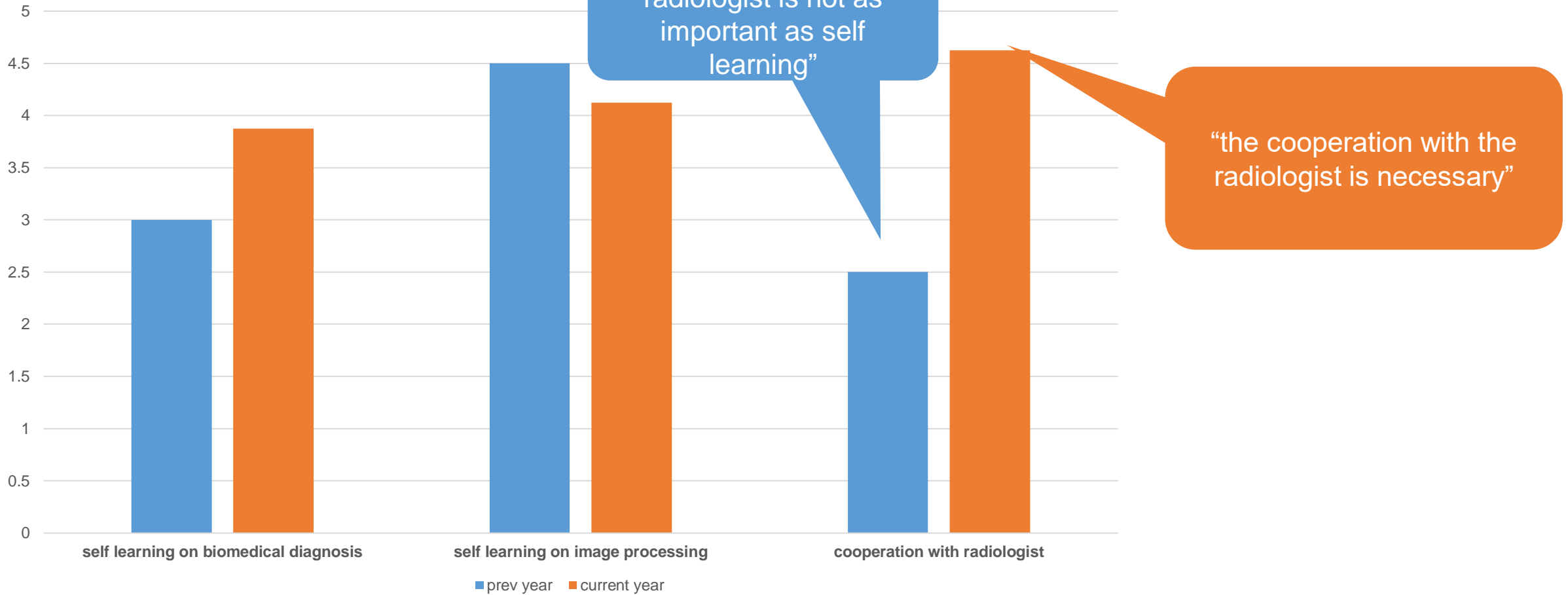
- Teach biomedical domain knowledge
- Convey the importance of medical domain knowledge
- Refer students to self-learning material



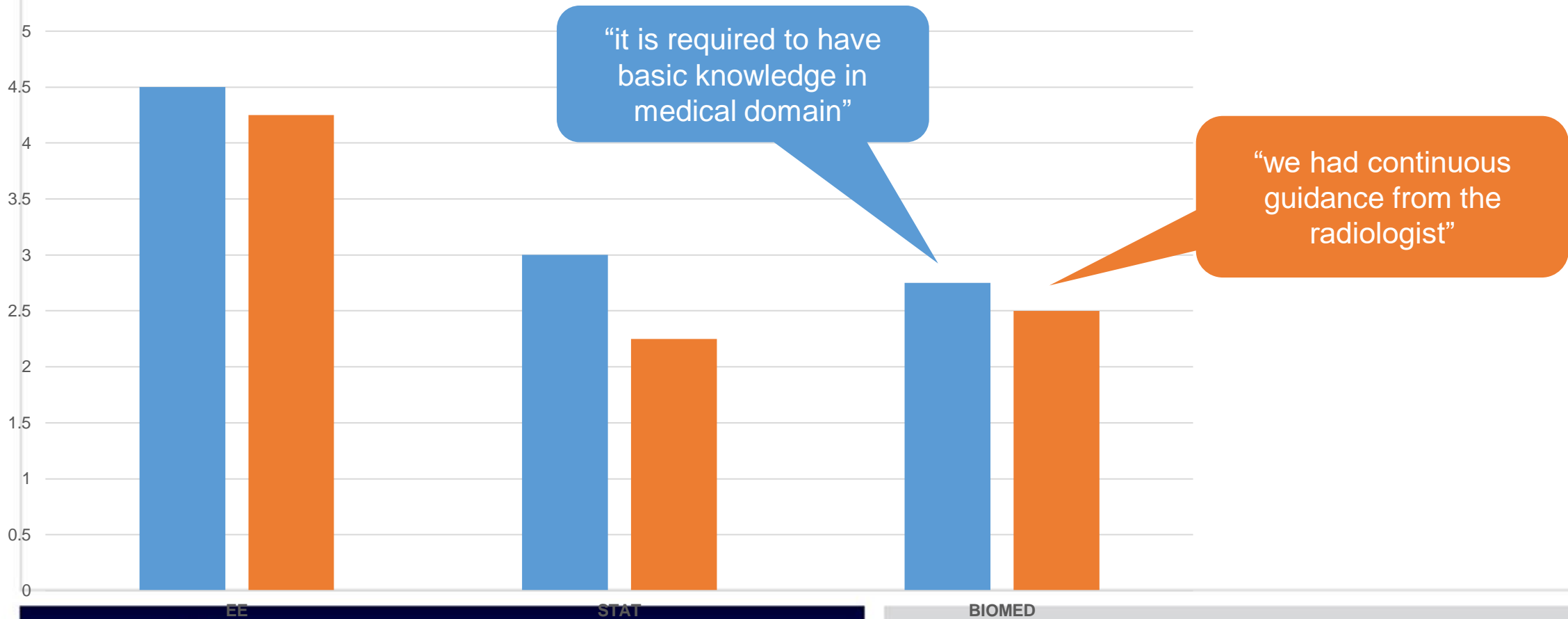
# The Intervention Program

Week	Theoretical background lectures	Meeting with radiologist
1	Ultrasound imaging	
4	The challenge of biomedical image diagnosis	Discussion and feedback on project goals
6	Fidelity criteria in biomedical images	
8		Feedback on intermediate results Resolving issues regarding medical diagnosis Refining project goals
9-13		Continuous remote communication for feedback on results
Final presentation (planned)	The biomedical industry	Feedback on final results

# Results: Students' Perception of Project Success Factors

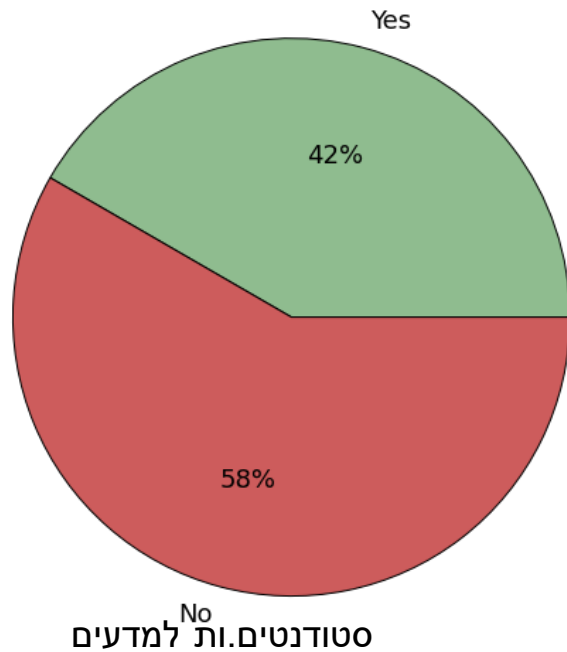


# Results: Students' Perception of Required Knowledge

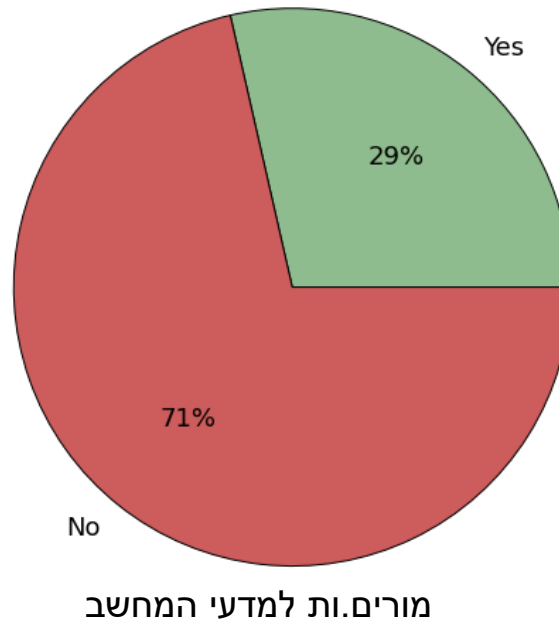


# התייחסות לתחום הדעת בביצוע פרויקט בלמידת מכונה

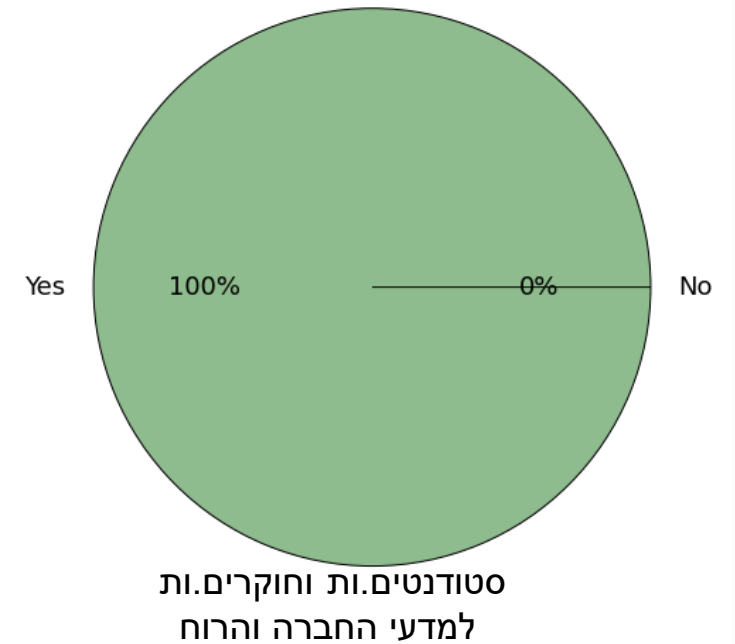
Using domain in PBL (Students n=12)



Using domain in PBL (Teachers n=28)



Using domain in PBL (SSDH n=30)





**הטכניון**  
מכון טכנולוגי  
לישראל

**שאלות?**