# מבוא

הוראה מרחוק והוראה היברידית הן שיטות הוראה שבדרך כלל משתמשות במרחב מלווה מקצוע שמוגדר כסביבה אישית לארגון, תכנון וניהול הוראה-למידה-הערכה. ניתן להקים את המרחב במערכת לניהול למידה מודלMoodle או בסביבות הענן לחינוך גוגל, מיקרוסופט ועוד. בתקופה האחרונה בתי ספר תיכון רבים משתמשים במערכות לניהול למידה כגון מודל, והתלמידים משתמשים במערכות אלה ומשאירים שם "עקבות דיגיטליים" שניתן להשתמש בהם לייעול ושיפור תהליך ההוראה, הלמידה וההערכה וגם למטרות חיזוי הצלחה או נשירה של תלמידים ועוד.

כריית נתונים ידועה בשם KDD-Knowledge Discovery in Database, היא ידועה כבעלת תפקיד חשוב ועוצמתי בחשיפת מידע חבוי מכמויות גדולות של נתונים וניתוחם. ניתן לסווג כריית נתונים ברשת Web Mining לשלושה סוגים: כריית נתוני משתמש, כריית נתוני מבנה וכריית נתוני תוכן. כריית נתונים בחינוך EDM-Educational Data Mining מוגדרת כתחום מחקר מדעי הממורכז סביב פיתוח שיטות לחקר נתונים מסביבות חינוכיות ולימודיות ושימוש בשיטות אלו להבנה טובה יותר לתלמידים ולתהליכי הלמידה שלהם. היא יעילה לצורך בחינת הפעילויות הלימודיות של התלמידים בסביבה מקוונת. המחקר המתואר כאן עוסק בכריית נתונים בחינוך, ומתמקד בכריית נתוני משתמש, תוך כדי התבססות על קובץ יומן Log file להשגת שלוש מטרות עיקריות.

מחקר זה מתבסס על עיבוד וניתוח נתונים ראשוניים שנלקחו מקובץ יומן מאתר מודל שמלווה את מקצוע הביוטכנולוגיה לתלמידי כיתה י, מבית ספר תיכון שנמצא במרכז הארץ. מטרות המחקר הן: (א) לאפיין ולזהות את תבניות השימוש של הלומדים ואת הפעילויות שלהם באתר: צפייה בתכנים, הגשת מטלות, השתתפות בדיונים בפורום, ומספר כולל של הפעילויות שנעשות פר שבוע ופר חודש בתקופה של ששה חודשים מנובמבר 2017 ועד אפריל 2018. (ב) לבנות מודל לחיזוי נשירה תוך כדי התבססות על הנתונים המעובדים. (ג) לעשות אשכול ללומדים לצורכי יצירת קבוצות למידה. המחקר משתמש בתוכנת Rapid Miner ובטכניקת עץ החלטה Decision Tree לסיווג Classification וחיזוי נשירה predicting, ובאשכול Clustering בתוכנת spss לצורך יצירת קבוצות למידה.

תרומתו העיקרית של המחקר היא ביכולתו לאתר בטרם עת, תלמידים שעלולים לנשור בעתיד. חיזוי נשירה בשלבים מוקדמים עשוי לסייע באיתור וזיהוי תלמידים שנמצאים בסכנת נשירה, ואז ניתן להכין עבורם תכניות לימוד, או תכניות מעקב ייחודיות שעשויות לסייע להם ולהוציא אותם מסכנת נשירה. כאנשי חינוך, וכמורים בבתי ספר תיכון, הקטנת ממדי הנשירה נחשב לאחד היעדים הבסיסיים שלנו ושל מערכת החינוך באופן כליי. אמנם יש למחקר מגבלה עיקרית והיא אי יכולתו להיות מוכלל באופן רחב היקף לכל אתר אחר שמלווה קורס, אך ניתן בהחלט לאמץ את הרעיון הבסיסי ולעשות אדפטציה למודל ולהתאימו לקורסים אחרים.

# 2 סקירת ספרות

## 2.1 אתרים מלווי מקצוע

האפשרות שמזמנת הטכנולוגיה בת ימינו ובעיקר התפתחות האינטרנט מניחות בידי המורה ובידי מפתחי תכניות לימוד כלים ומשאבים המיועדים לשפר את הלמידה. האינטרנט כובש בהדרגה את מקומו בשדה ההוראה והחינוך גם כאמצעי תקשורת וגם כמרכז משאבים מסוגים שונים. משרד החינוך שם לו למטרה לקדם תהליכי פיתוח מקצועי משמעותיים בסביבה עתירת טכנולוגיה. אחד האמצעים לכך הוא המעבר לשימוש במערכת לניהול למידה המהווה תשתית טכנולוגית ללמידה משמעותית בתהליכי פיתוח מקצועי. השימוש במערכת מאפשר ומעודד תהליכי למידה דיגיטליים מתמשכים ושיתופיים ויישום עקרונות של למידה שיתופית ולמידת עמיתים.

שימוש הלומדים במערכות לניהול למידה, או אתרים מלווי מקצוע כגון Moodle משאיר "עקבות דיגיטליים", שניתן להשתמש בהם, לנתחם ולהסיק מהם מסקנות שונות. כריית נתונים בחינוךEDM: Educational Data Mining מוגדרת כתחום מחקר מדעי הממורכז סביב פיתוח שיטות לחקר נתונים מסביבות חינוכיות, לימודיות, אתרים ושימוש בשיטות אלו להבנה טובה יותר לתלמידים ולתהליכי הלמידה שלהם בסביבת למידה מקוונת. הדבר כולל חיזוי משתנים שונים כגון הצלחה, נשירה מאפייני אישיות כמו למידה במכוונות עצמית, מוטיבציה ועוד, וגם לצורכי אשכול הלומדים ועוד (Baker, 2010). בסקירה מקיפה שנעשתה על ידי מוחמד ותסיר נמצא שרב המחקרים בתחום כריית נתונים בחינוך נותנים תשומת לב רבה לשימוש בסביבת למידה מקוונת כגון Moodle, Blackboard, WebCT. ממצאי המחקר מראים שהטכניקות הנפוצות ביותר בכריית נתונים בחינוך הן אשכול, סיווג, תבניות עוקבות וחיזוי לפי הסדר ( Mohamad & Tasir, 2013).

מחקרים בתחום ניתוח למידהLA וכריית נתונים בחינוך EDM, משתמשים בכלים מיוחדים לתיאור תבניות התנהגות הלומדים והפעילויות שלהם בסביבת למידה מתוקשבת כגון טבלאות פיבות pivot של אקסל, שנותנות אינפורמציה בשלושה ממדים: מי עשה, מה עשה ומתי עשה? (Dierenfeld & Merceron, 2012). ניתן ליצור "לרנוגרמות" – גרפים מיוחדים שמתארים את התקדמות תהליך הלמידה על ציר הזמן מזוויות שונות (Nachmias & Hershkovitz, 2007). שיטות אלו לייצוג נתונים מתבססות על עיבוד וניתוח הנתונים מקובץ יומן (log file ) של אתר מודל שמלווה את הקורס.

## 2.2 אשכול לומדים

באשכול המטרה היא למצוא נקודות נתונים שמתחברות יחדיו, וכך מפצלות את מערך הנתונים למערך של אשכולות. ניתן לעשות אשכול לפעילויות התלמידים כדי לחקור תבניות או התנהגות. מחקרים שונים מגדירים משתנים שונים לצורכי אשכול הלומדים: (1) האשכול עשוי להתבסס על התנהגות הלומדים באתר הלמידה על ציר הזמן לדוגמא: (א) באתרים מלווי למידה – מודל, של קורסים אקדמאיים באוניברסיטת תל אביב, במחקרם של הרשקוביץ ונחמיאס (Hershkovitz & Nachmias, 2011) נמצאו 5 אשכולות: משתמשים בהיקף קטן, משתמשים מאחרים, נושרים מקוונים, משתמשים מואצים, משתמשים מואטים (Low-extent Users, Late Users, Online Quitters, Accelerating Users, Decelerating Users). (ב) בקורסים מקוונים, פתוחים ורבי משתתפים-מוקים, במחקר של (Kizilcec, Piech & Schneider, 2013) נמצאו 4 אשכולות שמתייחסות להתנהגות הלומד פר שבוע: תלמידים אשר מגישים את המטלות בזמן, סטודנטים אשר מגישים את המטלות אך באיחור, תלמידים שלא מגישים את המטלות אך צופים בסרטוני ההרצאות, תלמידים שלא השתתפו בקורס באותו שבוע ( [T] on track, [B] behind, [A] auditing, [O] Out) (2). אשכול הלומדים בקורסים מקוונים, פתוחים ורבי משתתפים-מוקים, לפי מידת המעורבות של הלומדים ולפי תבניות האינטראקציה שלהם עם סרטוני ההרצאות, המטלות ועוד רכיבים באתר של מוק, על ציר הזמן. לדוגמא, במחקר של (Ferguson & Clow, 2015) נמצאו 4 אשכולים: תלמידים שהשלימו את רב המשימות, תלמידים שלא השלימו את רב המשימות אם בכלל, אך צפו בסרטונים, תלמידים שבהתחלה ביצעו מטלות אך מידת מעורבותם פחתה עם הזמן, תלמידים שצפו בסרטון אחד או שניים בלבד (Completing, Auditing, Disengaging, Sampling). (3) אשכול הלומדים בקורסי מוק (An, Krauss & Meceron, 2017) לפי סוג הרכיבים שהתלמידים צופים בהם באתר, זמן שהייה ברכיב, ציונים במטלות הערכה, או הכלאה של ציונים וזמן שהייה ברכיב (Clicks only, Elapsed time, Assessment scores, Elapsed time and Assessment score).

מודלים של אשכול וחיזוי שנבנו במחקרים בשדה כריית נתונים מקורסים מקוונים, פתוחים ורבי משתתפים (MOOCs), נבדקו ונבדקה אפשרות יישומם בקורסים שאינם מוקים. ממצאי המחקר העידו שמודל האשכול שנבנה מניתוח נתונים מקורסי מוקים, הינו בר יישום עבור קורס אחר שאינו מוק, אך אינו ישים ואינו מתאים עבור כל קורס שאינו מוק, כך שלא ניתן לעשות הכללה גורפת למודלים (An, Krauss & Meceron, 2017).

## 2.3 חיזוי נשירה

חיזוי של נשירה dropout prediction הפסקת פעילות, ושימור התלמידים retention הם נושאי מפתח בתחום כריית נתונים בחינוך. הגדלת שימור התלמידים הינה מטרה משותפת של מוסדות אקדמיים רבים, במיוחד ברמת האוניברסיטה ( Iam-On & Boongoen, 2017; Kang & Wang, 2018)‏. ההשפעות השליליות של נשירה של תלמידים ניכרות לתלמידים, להורים, לאוניברסיטה ולחברה כולה. נושא זה הינו חיוני ביותר גם ברמה של בתי ספר תיכון, כיוון שהוא משמש כמערכת אזהרה. מחקר שאסף דאטה מ 419 תלמידים מבתי ספר תיכון במקסיקו, הצליח לפתח מודל לחיזוי נשירה שנמצא כמספיק אמין לחזות נשירה כבר בשלבים מוקדמים, בשבועות 4-6 של הקורס (Márquez‐Vera et al.,2016).

פיקוח על פעילות הסטודנטים בעזרת ניתוח קבצי יומן (לרב ביצוע מטלות, ציונים ומבחנים), של מערכות לניהול למידה מאפשר לגלות ולאתר "סטודנטים בסכנה" בשלבים מוקדמים, וכך לפעול בהתאם מבעוד מועד (Lykourentzou et al., 2009; Romero-Zaldivar et al., 2012). מחקר שפיתח מודלים לחיזוי נשירה אקדמית, עיצב תכנית פעולה עם שיעורים פרטיים, ויישם אותה בפועל. פעולות אלו שננקטו שכללו פיתוח מודל לחיזוי נשירה, עיצוב תכנית התערבות מתאימה ויישומה בשטח, הקטינו את אחוז הנשירה ב 14% ביחס לשנים קודמות שבהן לא הופעל שום מנגנון למניעת נשירה (Burgos et al., 2018). תכניות מקוונות בתחום ההשכלה הגבוהה בדרך כלל סובלות מאחוז נשירה גדול יותר מאשר בקמפוסים, דבר שגרם לחוקרים Kang & Wang (2018) לפתח מודלים לחיזוי נשירה, במטרה לספק למנהלים, למורים ולעובדים, הזדמנות לנקוט פעולות כדי להתערב בתהליך הנשירה של התלמידים לפני שהוא מתרחש בפועל. מודל החיזוי התבצע בסמסטר הראשון ועוד לפני תחילת השני. הממצאים אומצו על ידי הנהלת המוסד האקדמי שאליו משתייכים החוקרים, כדי לשפר את שיעור השימור של התוכנית המקוונת .

הסוגיות של הפסקת פעילות של הלומדים, ואיך לעורר ולהגביר את מעורבות התלמידים בפעילויות לימודיות וכתוצאה מכך להגביר את שביעות הרצון הכללית שלהם, נחקרו על ידי מספר חוקרים. המחקרים הראו שגורמים דמוגרפיים והישגים של הלומד והביצועים שלו, נמצאו כקשורים למוטיבציה של הלומד להישאר מעורב ופעיל (Dejaeger et al., 2012; Giesbers et al., 2013; Guo, 2010). זיהוי גורמי הצלחה דחף את Giesbers ושותפיו (2013) לחקור את הקשר בין התנהגות הלומד הנצפית, המוטיבציה של הלומד, והביצועים שלו בבחינה הסופית. החוקרים בדקו האם שימוש בכלים סינכרוניים מעלה את המוטיבציה של הלומדים להשתתף בקורסים שתומכים בכלים אלה. במחקר אחר נבדקו מדדי שביעות הרצון של התלמידים לשמירה על אוכלוסיית הלומדים, והקטנת הנשירה (Dejaeger et al., 2012). החוקרים בדקו מספר של מבנים לשביעות רצון כגון תועלת נתפסת ויעילות נתפסת, יחד עם עוד משתני כיתה קשורים.

מחקרים שונים השתמשו בטכניקות שונות כדי לחזות נשירה בשלב מוקדם. במחקר של דקר ושותפיו, החוקרים השתמשו באלגוריתמים שונים של סיווג כדי לחזות נשירה בהתבסס על נתונים מוקדמים של הסטודנטים, נתונים משנה ראשונה של הלימודים. החוקרים השוו את מידת הדיוק של אלגוריתמים שונים של סיווג, ומצאו שמודלים פשוטים לאשכול, יש להם רמת דיוק גדולה יותר בהשוואה לאלגוריתמים מורכבים יותר לאשכול (Dekker et al., 2009). מחקר אחר (Lykourentzou et al., 2009) בדק טכניקה אחרת לחיזוי נשירה בשלב מוקדם, החוקרים יישמו הכלאה של שלוש טכניקות על פרופילים מפורטים של סטודנטים בסביבה לניהול למידה. החוקרים השוו את הדיוק והרגישות של המתודות המוצעות עם אחרות מהספרות. ממצאי המחקר הראו ששימוש בהכלאה של שלוש טכניקות סיפק תוצאות אמינות יותר .

במחקר שסקר 209 מחקרים בתחום כריית נתונים בחינוך וניתוח למידה משנת 2008-2013, והשווה ביניהם מבחינת שאלות מחקר, מתודולוגיה וממצאים, נמצאו רק 7 מחקרים אשר עסקו בחקר וחיזוי נשירה של סטודנטים מאתרי למידה מתוקשבת (Papamitsiou & Economides, 2014). המחקר מצביע על פערים, חזקות, חולשות, איומים והזדמנויות בתחום כריית נתונים בחינוך. חלק מהאיומים המעניינים שהחוקרים מצביעים עליהם הם "ניתוח יתר", שאינו מאפשר הכללה של המודלים, דבר שעלול לפגיעה בתמונה ההוליסטית, ואז רואים את העץ אך מפספסים את היער. עוד איום חשוב הינו הבעיה האתית ופרטיות הנתונים והסטודנטים שנפרצת. ממצאים מעניינים נוספים הם חולשות בתחום כריית נתונים בחינוך כגון פירושים לקויים של הממצאים כתוצאה מגורמי שיפוט אנושיים, והתמקדות בדיווח ולא בדיונים או בהחלטות. חולשה חשובה נוספת היא העובדה שרב המחקרים בתחום כריית נתונים בחינוך מספקים ממצאים כמותיים, וקיים מחסור בסיפוק ממצאים חשובים ממתודות איכותניות. מחקר אחר ( Mohamad & Tasir, 2013) מצביע על עוד חולשה בתחום כריית נתונים בחינוך והיא מחסור והיעדר למידה שיתופית אמתית.