

## להלן קורסי ההקבץ:

- חשיבה סטטיסטית למדעי הרוח 3 ש"ס.
- מבוא למדעי המחשב ושפת פייתון 3 ש"ס.
- קורס מדעי הנתונים 1 (היכרות עם נתונים מסוגים שונים וניהולם) 3 ש"ס.
- קורס במדעי הנתונים 2 (ניתוח והסקת מסקנות מנתונים, כולל נושאים בסיסיים בלימוד מכונה) 3 ש"ס.
- פרזיקט במדעי הנתונים - יישום הידע והמיומניות במדעי הנתונים באחד מתחומי הדעת של מדעי הרוח המפורטים להלן 4 ש"ס.

### סילבוס לקורס " חשיבה סטטיסטית למדעי הרוח" (Math1) - 3 ש"ס

ד"ר ריקרדו טרש וד"ר ארנון הרשקוביץ

מטרת הקורס : הקניית מושגי יסוד במתמטיקה למדעי הנתונים. התלמידים יכירו מושגים יסוד באלגברה, הסתברות וסטטיסטיקה ואת שימושיהם לניתוח נכון של נתונים במאגרי מידע של תחומים הומאניסטיים וחברתיים.

נושאי הקורס :

שיעור	נושא
1.	סוגי משתנים
2.	הגדרות בסיסיות בהסתברות - מרחב מדגם, מאורע, מאורע משלים, הסתברות במרחב מדגם אחיד ולא אחיד
3.	חיתוך ואיחוד מאורעות, מאורעות זרים ומכילים
4.	כלל המכפלה, ההכלה וההפרדה
5.	הסתברות מותנית במרחב דגימה אחיד ולא אחיד
6.	נוסחת בייס
7.	סטטיסטיקה תיאורית - הצגה של נתונים.
8.	סטטיסטיקה תיאורית - מדדי מיקום מרכזי ומדדי פיזור. טרנספורמציות לינאריות ולא לינאריות
9.	המשתנה המקרי הבדיד - פונקצית הסתברות, תוחלת, שונות וסטיית תקן - דוגמאות להתפלגויות בדידות מיוחדות - אחידה, בינומית וכו
10.	המשתנה המקרי הרציף (ברמת אינטואיציה בלבד , דרך דוגמאות) - פונקצית התפלגות מצטברת, תוחלת, שונות - דוגמאות להתפלגויות רציפות מיוחדות – מעריכית, אחידה, נורמלית וכו - ציוני תקן
11.	מדדי קשר - רגרסיה ליניארית (הגדרה ואינטואיציה בלבד)
12.	הסקה סטטיסטית (ברמת אינטואיציה בלבד, הדגמה של הסקה מבוססת דגימה ללא נוסחאות)
13.	אלגברה : מטריצות ווקטורים - הגדרה ופעולות בסיסיות : חיבור ומכפלה אלגברה : מטריצות ווקטורים - קשר למשוואות לינאריות ונקודות במרחב, מרחק בין וקטורים

מטלת סיום הקורס : בחינת כיתה.

## סילבוס לקורס "מבוא למדעי המחשב ושפת פייתון" (CS1) - 3 ש"ס

מרצה: ד"ר אמיר רובינשטיין

מטרות הקורס: ללמד מיומנות התחלתית בשפת התכנות פייתון, תוך היכרות עם מושגים מרכזיים במדעי המחשב והבנה כיצד מתכננים אלגוריתמים לפתרון בעיות באמצעות מחשבים.

נושאי הקורס:

שיעור	נושא
1.	מושגי יסוד במדעי המחשב ושפת פייתון
2.	רשימות תנאים ופונקציות
3.	רשימות תנאים ופונקציות (המשך)
4.	לולאות
5.	לולאות
6.	מפגש אמצע סמסטר פיזי לתרגול וחיזוק הנושאים שנלמדו עד כה
7.	אלגוריתמים בסיסיים לחיפוש ומיון
8.	אלגוריתמים בסיסיים לחיפוש ומיון (המשך)
9.	תוכנה פוגשת בעיה 1: קודים לגילוי ותיקון שגיאות
10.	תוכנה פוגשת בעיה 2: מבוא לקריפטוגרפיה
11.	תוכנה פוגשת בעיה 3: יצוג ועיבוד תמונה דיגיטלית
12.	שאלות מרכזיות במדעי המחשב
13.	מפגש סוף סמסטר פיזי לתרגול וחיזוק

מטלת סיום הקורס: בחינת כיתה.

## סילבוס לקורס "מבוא למדעי הנתונים 1" (DS1) 3 ש"ס

מרצה: ד"ר סלבה נובגורודוב

מטרת הקורס: הכרת סוגי מאגרי מידע ונתונים מסוגים שונים. לימוד אופני הגישה, השימוש והניהול של מאגרים אלה.

דרישות קדם: CS1 או כל קורס תכנות בפיתון, Math1 או כל קורס בהסתברות וסטטיסטיקה (את החומר באלגברה ניתן להשלים לבד ע"י צפייה עצמאית בשתי ההרצאות הראשונות של Math1)

דרישת סיום: בחינת כיתה.

שיעור	נושא
1.	מבוא – ניתוח נתונים מהו? מה זה מדעי הנתונים? סוגיות אתיות. מטרת הקורס וסקירת הנושאים שיועברו
2.	ריענון תיכנות בפיתון – חזרה על פקודות שימושיות, עבודה עם Jupiter notebooks
3.	סוגי נתונים – structured, unstructured, טקסט, תמונה, צליל. שיטות כלליות לעבודה עם נתונים כאלה
	[רקע דרוש מ Math1: מטריצות, וקטורים]
4.	עבודה עם נתונים בפיתון – טעינת נתונים, טרנספורמציות, ניקיון והכנה
5.	מבוא לעבודה עם structured data – יצירת טבלאות, תשאול בסיסי ב SQL, הכנסה, עדכון, ומחיקת נתונים
6.	מבוא לעבודה עם טקסט – tokenization, פיצול משפטים, דמיון טקסט, שיטות מתקדמות
	[רקע דרוש מ Math1: מרחק בין וקטורים]
7.	משימות שימושיות בעבודה עם טקסט – sentiment analysis, relations extraction
	[רקע דרוש מ Math1: חוק בייס]
8.	ויזואליזציה של נתונים – שיטות שימושיות
	[רקע דרוש מ Math1: ממוצע, חציון, היסטוגרמות]
9.	עבודה עם נתונים גולמיים באינטרנט – Web scraping, Web crawling
10.	ניתוח של נתונים אינטרנטיים – חילוץ תובנות מנתוני Web (מבוא לניתוח נתונים)
11.	בניית פתרונות קצה לקצה על סמך הנתונים שחולצו מהאינטרנט, ארגון של הנתונים (structuring), עדכון נתונים על בסיס איחוד/שינוי/ניקוי של הנתונים המקוריים.
12.	דוגמאות לשימושים פרקטיים שניתן לממש על סמך החומר הנלמד עד כה
13.	סיכום וחזרה על החומר שנלמד בקורס

## סילבוס לקורס "מבוא למדעי הנתונים 2" (DS2) 3 ש"ס

מרצה: ד"ר סלבה נובגורודוב

דרישות מקדימות: DS1  
דרישת סיום: בחינת כיתה.

.1	מבוא ללימוד מכונה – סוגים שונים של לימוד מכונה (supervised/unsupervised), למה בכלל צריך? סוגיות אתיות
.2	מערכות למידת מכונה – הסבר בסיסי ודוגמאות על איך מלמדים מכונה [רקע דרוש מ Math1: הסתברות של משתנים בלתי תלויים, חישוב הסתברות]
.3	למידה unsupervised – סקירה כללית, הדגמה ע"י clustering, k-means [רקע דרוש מ Math1: מרחקים בין וקטורים]
.4	למידה supervised – סקירה כללית של סוגי supervised learning, רגרסיה לעומת שיטות קלסיפיקציה [רקע דרוש מ Math1: חוק בייס, רגרסיה]
.5	השגת נתונים מתוויגים - scrapping, crowdsourcing, data generation
.6	גרסיה לינארית – הסבר, שימושים, דוגמאות מהעולם האמיתי [רקע דרוש מ Math1: רגרסיה]
.7	נאיב בייס – הסבר, שימושים, דוגמאות מהעולם האמיתי [רקע דרוש מ Math1: חוק בייס]
.8	הערכה של איכות התוצאות – הערכה בעזרת נתונים מתוויגים, precision/recall, RMSE, AUC
.9	מבוא ל NLP – שימוש ב word embeddings, דוגמאות מהעולם האמיתי [רקע דרוש מ Math1: מטריצות, וקטורים]
.10	מבוא לרשתות ניורונים – סקירה כללית ברמה אינטואיטיבית [רקע מ Math1: מטריצות ווקטורים, פעולות על מטריצות, בעיקר מכפלה]
.11	שיטות למידה מתקדמות – סקירה של שיטות state-of-the-art ובעיות שנפתרות על ידיהן
.12	דוגמאות לשימושים פרקטיים שניתן לממש על סמך החומר שנלמד בשני הקורסים למדעי הנתונים יחדיו
.13	סיכום וחזרה על החומר שנלמד בקורס 2

## פרוייקט במדעי הנתונים (PROJ) 4 ש"ס

קדם : DS2, DS1

דרישת סיום : עבודה שמסכמת את פרויקט המחקר.

במסגרת הפרוייקט, התלמידים יעשו שימוש בשיטות ובכלים של מדעי הנתונים שנלמדו במסגרת ההקבץ על מנת לענות על שאלת בתחומי המחקר, שנחקרים בתוכנית הלימודים שבה הם לומדים. את הפרוייקט ינחו חברי/ות סגל מהפקולטה שבה לומדים התלמידים. חברי סגל של הפקולטה המעוניינים לשלב סטודנטים שלהם בסמינר מחקר שיש לו זיקה למדעי מידע יוכלו להציע פרויקטים או קליטה של תלמידים ממדעי הנתונים עם הדרכה ייחודית ועבודה משותפת בין קהילת חוקרי מדעי הנתונים של הפקולטה (קהילה שתצמח, כך אנו מקווים, ביחד עם ההיקבץ). הפרוייקט יאפשר למשתתפים בו לעשות שימוש במאגרי מידע גדולים על מנת לענות על שאלות מחקר מפרספקטיבה הומניסטית. מאגרי מידע העשויים לשמש את המשתתפים בפרוייקט, למשל, הם מאגרים של יחידות ממשלתיות, מועצות מקומיות, הלמ"ס או משרד החינוך, הספרייה הלאומית מאגר המידע של יד ושם ועוד, הפרוייקט יונחה על ידי חברת או חבר סגל המומחה בתחום מסוים יהיה גם בעל אופי יישומי אשר יאפשר לסטודנטים גם דריסת רגל תעסוקתית. הפרוייקט מתוכנן לשנה שלישית וישולב באופן מפורט בתוכנית הלימודים בשנת תשפ"ה. היקבץ זה מוצע לתלמידי הפקולטות של הצד המערבי של הקמפוס, וכל פקולטה אחראית על פרויקט המחקר במדעי הנתונים שלה. אנו עובדות על יצירת קהילה של חוקרות וחוקרים שיעבדו במשותף לבניית פרויקט במדעי הנתונים שיש בו גם היבט יישומי בעולם העבודה.

להלן שני פרויקטים לדוגמא שכבר התחילו להתגבש בפקולטה למדעי הרוח :

### **מדעי הנתונים וחקר ההיסטוריה האנושית: לחשוב בגדול**

ד"ר דפנה לנגוט – החוג לארכיאולוגיה ותרבויות המזרח הקדום

(קורס סמינר המוצע לתלמידי ב.א. בשנת תשפ"ה)

הקורס יעסוק בהיבטים שונים של מדעי הנתונים וחקר ההיסטוריה האנושית (אבולוציה, תרבות חומרית, חברה, בלשנות ועוד), תוך התמקדות במזרח הקדום. נכיר אפשרויות גישה ואופני שימוש עם מסדי נתונים גדולים קיימים, בניית מסדי נתונים חדשים, ניתוח הנתונים, פרשנות ויישומים מחקריים. הקורס ישלב הרצאות אורח של חוקרים. ות העושים שימוש מחקרי יומיומי במסדי נתונים גדולים והיסטוריה וארכיאולוגיה דיגיטאלית (כדוגמת ד.נ.א קדום, חומר ותלת מימד, אפיגרפיה וכרונולוגיה). נעסוק ביתרונות וחסרונות גישה כמותית זו ונשלב גישה ביקורתית בעת הטיפול בנתונים ובפרשנותם (למשל תת-ייצוג וייצוג-יתר כתוצאה משימוש בדיגטציה בסקרים ארכאולוגיים ובמהלך החפירות עצמן). ניבחן את השפעת המידע הדיגיטלי על פארדיגמות הרווחות כיום במחקר. כמו-כן נדון במגבלות פוליטיות-אזוריות המשפיעות על בנייה ופרשנות של מסדי הנתונים של המרחב הלבנטיני הקדום. תוצרי הקורס הינם עבודת חקר סמינריונית ייחודית מתחום מדעי

הנתונים ההיסטוריים המתבססת על ניתוח נתונים ותרומתם להארות סוגיות בחקר העבר. כך למשל ניתן לערוך השוואה צורנית כמותית בין מיכלולים של כלי אבן מאתרים שונים לאורך ציר הזמן באמצעות שימוש בשיטות סטטסטיות מרובות משתנים בכדי לזהות דמיון ושונות אשר יסייעו באפיון מגמות אנושיות התנהגותיות (העברה של ידע, קוגניציה והתפתחות חברתית-תרבותית). באופן דומה ניתן לבחון גם כלים קרמיים תוך שילוב מידע טיפולוגי, פטרוגרפי ותלת מימדי. בהיבטים ארכיאובוטאניים וארכיאוזאולוגיים מסדי נתונים גדולים עשויים לזהות מתי והיכן בוית לראשונה (או הובא במכוון) מין מסוים של בע"ח או צמח וכן לסייע במעקב אחר הפצת הרעיון במרחב. ניתן לחקור במסגרת העבודה הסמינריונית גם מסדי נתונים גדולים שהינם תוצרים של מערכות מידע גיאוגרפיות (GIS), תצלומי לוין ואויר וכן מתיעוד באמצעות רחפנים.

### **סיווג וניתוח טקסטים בעברית מהעת העתיקה באמצעים דיגיטליים**

ד"ר אשבל רצון

הקורס יתמקד בטקסטים עבריים מהעת העתיקה שכבר עברו דיגיטציה. לאחר מבוא כללי שמציג את הספרות והחומר האפיקרפי שהגיע לידינו בעברית מכל העת העתיקה, נבחן שאלות מחקריות מרכזיות שעוסק בהן המחקר כגון: תארוך הטקסטים, תהליך החיבור והעריכה שלהם, זיהוי המחברים וקשרים בין חיבורים. נערוך הכרות עם המאגרים והכלים הדיגיטליים הקיימים בתחום, ונלמד לעומק מספר פרויקטים חישוביים שכבר נעשו ופורסמו. בפרויקט הסופי התלמידים יוכלו לבחור שאלות שכבר נשאלו במחקר לגבי טקסט מסוים או לגבי קבוצה של טקסטים ונענו רק באמצעים קלאסיים של מדעי הרוח, להציע שיטות חישוביות כדי לענות על השאלות, ולהשוות את התוצאות הקודמות לתוצאות שלהם. לחילופין, הפרויקט יוכל לענות על שאלות חדשות בעלות משמעות שלא ניתן היה לבחון אותן ללא כלים חישוביים. הקורס ייכלל בין שיעורים פרונטליים להנחיה במהלך העבודה על הפרויקטים ופרזנטציות של התוצרים בסוף השנה.