

פרויקט בבינה מלאכותית (236502) סמסטר חורף 2022-23

צוות הקורס							
						שאל מרקוביץ	מרצה
					רוי שפיר	דין צדוק	מנחי פרוייקטים

תיאור כללי

במסגרת הקורס תבצעו פרויקט בנושא הקשור לבינה מלאכותית. תוכלו לבחור את אחד הנושאים המוצעים על ידינו, או להציע נושא בעצמכם. הצעת הפרויקט והדו"ח המסכם ישלחו באופן אלקטרוני **בפורמט PDF**. הציונים יקבעו בעיקר על פי הדו"ח הסופי כולל נספח הקוד. הנכם רשאים להשתמש בשפת תכנות לפי בחירתכם.

קדמים

הקדם של הקורס הזה הוא 236501 - מבוא לבינה מלאכותית.

צורת עבודה

העבודה של כל קבוצה חייבת להיות עצמאית לחלוטין. אין להעתיק או להסתמך על מקורות אחרים ללא הפניה מפורשת למקור.

העבודה תעשה בזוגות או בבודדים (באישור). קבוצות שיעבדו בשלוש ידרשו לעבודה גדולה יותר משמעותית מכאלה שיעבדו בזוגות או בודדים. בשלב אישור הפרויקט הקבוצה תעלה הצעה להרחבה של כ 50% והיא תאושר ע"י המנחה.

מהלך הפרויקט

לוח זמנים	
פגישה של כל הכתה עם המרצה.	6/11/22 16:30-17:30
פגישה עם המנחה ובחירת נושא. שלב זה כולל חשיבה עמוקה על הבעיה ועל הדרכים לפתרון. במידה ומעורבים נתונים, עליכם לבדוק שניתן להשיג אותם.	6/11/22-20/11/22
הגשת הצעת פרויקט לפי ההנחיות	21/11/22
פגישות והתייעצות עם המנחה	בהמשך
הגשת הפרויקט לפי ההנחיות.	20/3/23

בחירת נושא

כללית, הקורס נועד לאמן את הסטודנט בהפעלת טכניקות אותן למד בקורס "מבוא לבינה מלאכותית" על בעיה מעניינת. אנחנו מעדיפים נושא שמציעים הסטודנטים - הנסיון שלנו מראה שאלה הפרויקטים המעניינים ביותר. ניתן גם לבחור מתוך ההצעות כאן.

1. בעיות חיזוי. רוב הפרויקטים שנעשו במסגרת הקורס הזה בשנים האחרונות עסקו בבעיות חיזוי בתחומים שונים על ידי שימוש באלגוריתמי למידה. השלבים בפרויקט כזה הם בדרך כלל:
 1. מציאת מאגר מידע מעניין. דוגמאות למאגרים:
 1. מאגר מידע על קבוצות כדורגל באנגליה עם תוצאות משחקים ופרטים על השחקנים
 2. מאגר ערכים של מניות
 3. בסיס נתונים על סרטים (IMDB)
 4. מאגרים מטוויטר ומפייסבוק
 5. ויקיפדיה
 6. מאגרים הקשורים בתאונות דרכים
 7. מאגרים הקשורים בזיהום אוויר
 8. מאגרים רפואיים, לדוגמה בנושא הקורונה
 2. הגדרת בעיית חיזוי מעניינת על המאגר. דוגמאות לבעיות חיזוי מפרויקטים קודמים של הקורס:
 1. בהינתן חבר כנסת והצעת חוק - האם יצביע בעד או נגד?
 2. בהינתן שתי קבוצות כדורגל - איזו מהן תנצח?
 3. בהינתן מניה ספציפית - האם מחר היא עתידה לעלות או לרדת?
 4. בהינתן סרט חדש - מה הדירוג הממוצע הצפוי שלו?
 5. בהינתן טוויט, האם יזכה להרבה רטוויטים?
 3. הגדרת מאפיינים (תכונות) שעשויים להועיל לאלגוריתם הלמידה. כאן בדרך כלל מתבצעת העבודה העיקרית ומושקעת עיקר היצירתיות. **בעיה שבה אתם משתמשים בטבלה מוכנה של ערכי תכונות לא מובילה בדרך כלל לפרויקט מספק.** הבעיה הופכת למעניינת ומאתגרת אם אתם מגדירים ואוספים תכונות ממקורות מידע שונים.
 4. הפעלה של אלגוריתם למידה אחד או יותר עם סוגים שונים של תכונות
 5. עריכת ניסויי חיזוי עם קבוצות אימון וקבוצות מבחן. בדיקת השפעת פרמטרים שונים על יכולות החיזוי של המסווג בתנאים מגוונים. חשוב להתמקד יותר בפרמטרים יחודיים לבעיה ופחות בפרמטרים כלליים. לדוגמא, השוואה בין אלגוריתמי למידה שונים פחות מעניינת. השוואה בין משפחות שונים של מאפיינים או בין שיטות שונות לסימון דוגמאות מעניינות יותר.
 6. הערה חשובה: למצוא באינטרנט dataset מוכן ולהפעיל עליו אלגוריתם למידה - זה לא מעניין וממש לא בהיקף של פרויקט (אפילו תרגיל 3 בקורס בינה היה משמעותית מורכב יותר).
 2. בתרגיל הראשון בקורס נחשפנו למאגר המפות openstreetmap. המאגר מכיל מפות של מדינות רבות עם נתונים נוספים. הגדירו בעיות חיפוש/למידה מעניינות המבוססות על בסיס הנתונים הזה.
 3. פרויקטים הקשורים במשחקים. בדרך כלל רצוי על יותר ממשחק אחד:
 1. כתבו אלגוריתם ויוריסטיקות למשחקים בני יותר מ-2 שחקנים. לדוגמא: דמקה סינית. נסו את מינימקס המוכלל מול אלטרנטיבות.
 2. כתבו אלגוריתם הלומד מסווגים (למשל עצי החלטה) כדי להפוך את מינימקס לחיפוש סלקטיבי.
 3. כתבו אלגוריתם הלומד עצי החלטה כדי להפוך את expectimax לסלקטיבי (ממשו על שש בש).
 4. כתבו אלגוריתם הלומד עצי החלטה המשמשים כפונקציה יוריסטית לסידור מצבים.
 5. כתבו אלגוריתם להשתתפות בתחרות general game
 6. כתבו אלגוריתם להשתתפות ב robocup simulation
 7. כתבו אלגוריתם למשחקי זירה רבי משתתפים מסוג League of Legend
 4. פרויקטים הקשורים בסיווג טקסטים. בנוסף על מה שנאמר בסעיף 1, יש ספריות שונות לניתוח טקסטים ולעיבוד סמנטי שיכולות לעזור. בשנים האחרונות נוספו כלים חזקים מבוססי רשתות נוירונים לניתוח סמנטי של טקסטים. בנוסף, עבור עברית, יש את הספרייה של מרכז "מילה" לניתוח מלים.
 1. כתבו אלגוריתם המשתמש בשיטות סיווג כדי לזהות כותב
 2. פתחו מערכת לזיהוי מושג מעניין בטקסט (כמו המסווג למושג "סרקזם" שפתחו בירושלים)

3. חשבו על פרויקטים הקשורים בטובקים - למשל קיבוץ של מספר טובקים שכנראה נכתבו על ידי אותו כותב. או חיזוי של מספר הלייקים שטובקב יקבל או זיהוי טובקים אלימים או גזעניים.
5. פרויקטים אלגוריתמיים, לדוגמה:
 1. כתבו אלגוריתם למציאת תת קבוצה אופטימלית (או טובה):
 1. על ידי דגימה
 2. על ידי אלגוריתם גנטי כאשר פעולת ה crossover נותנת עדיפות נמוכה לאיברים בקבוצה שקשורים חיובית
 2. כמו 6 - רק סידור קבוצה
 3. כתבו מערכת לכוונון קבוצת פרמטרים עבור אלגוריתם למידה.
 4. כתבו תכנית ליצירת תכונות עבור אלגוריתם למידה על ידי חיפוש במרחב התכונות המורכבות.
 5. אלגוריתם מבוסס clustering ליצירת embedding ל טקסט

פגישה עם המנחה

קביעת נושא הפרויקט תעשה בפגישה עם המנחה. במהלך הפגישה ננסה למקד את הבעיה ואת שיטות הפתרון. אנחנו מבקשים שתגיעו מאוד מוכנים לפגישה. כלומר עליכם לבוא עם נושא מוכן אחר שכבר בדקתם איזה נתונים זמינים עבורכם ויש לכם קו כללי לפתרון. בסוף הפגישה, ברוב המקרים, מחליטים על נושא ודרך פעולה.

הצעת פרויקט

אחרי הפגישה ואחרי שהוחלט נושא הפרויקט, עליכם להגיש הצעת פרויקט. הצעת הפרויקט תוגש באימייל למנחה.

ההצעה היא מסמך PDF באורך של בין 1-3 עמודים המכיל את הסעיפים הבאים:

1. כותרת בעלת משמעות (לא "הצעת פרויקט" אלא, לדוגמה "חיזוי מהירות תנועה על פי נתונים היסטוריים"), משתתפים (שמות + מספר סטודנט + כתובות אלקטרוניות)
2. תיאור הבעיה אותה מנסים לפתור
3. תיאור הדרך אותה אתם מציעים לפתרון הבעיה - השלכת הבעיה לאחת מסוגי הבעיות שנלמדו בכיתה, תיאור מרכיבי הבעיה הנ"ל, והאלגוריתמים שתשתמשו לפתרון בעיה זו. יש לציין כל מידע שתנצלו לצורך הפתרון, ולהסביר איך אתם מתכננים לנצל אותו.
4. תיאור המערכת אותה אתם מציעים לבנות. תכנון מבנה המערכת, החלקים שעליה לכלול, כולל פירוט רכיבים חיצוניים שיש להיעזר בהם וחלקים שתתכנתו עצמאית.
5. תיאור הניסויים אותם אתם מתכננים לבצע לשם הערכת הפתרון המוצע. התיאור יכלול פרמטרים אותם אתם מתכננים לבדוק ומדדים בהם תשתמשו לשם הערכת ביצועי המערכת.
6. פרטים טכניים: שפת תכנות, סביבת תכנות, ספריות ומודולים קיימים בהם תשתמשו (למשל sklearn ללמידה).

תכנית ניסויים

אחרי שסיימתם את כתיבת המערכת ואחרי שערכתם ניסויים ראשוניים אתם מתבקשים להגיש למנחה תכנית ניסויים. התכנית תכלול את הסעיפים הבאים:

1. תיאור קצר מאוד של הבעיה.
 2. תיאור קצר של המערכת אותה בניתם.
 3. תיאור של הבעיות עליהם תריצו את הניסויים.
 4. פירוט של המדדים השונים בהם תשתמשו כדי להעריך את ביצועי האלגוריתמים.
 5. פירוט של פרמטרים שונים שתבדקו.
 6. פירוט של סדרת הניסויים שתערכו.
- כמובן שיתכן שבהמשך תעשו שינויים בתכנית הניסויים - אבל שלב זה חשוב מכיון שפעמים רבות הסטודנטים מבזבזים זמן רב בעריכת ניסויים לא מעניינים.

הגשת הפרויקט

אחרי שהמנחה הצמוד שלכם יאשר לכם הגשה, עליכם להגיש את הפרויקט בתיקיה מקודדת כקובץ zip המכילה:

1. דו"ח מסכם (**בפורמט PDF**) שיכתב לפי הסעיפים הבאים (בדרך כלל 15-30 עמודים):
 1. עמוד שער - כותרת אמיתית (המעידה על תוכן הפרויקט), שמות ומספרי זהות וכן כתובות דואר אלקטרוני.
 2. מבוא. רקע כללי והצגת הבעיה אותה ניסיתם לפתור.
 3. תיאור מפורט של הפתרון המוצע לבעיה. בדרך כלל יופיע כאן תיאור כללי של הדרך בה אתם פותרים את הבעיה וכן אלגוריתמים אותם פיתחתם. את האלגוריתמים יש לכתוב בצורה פורמלית ומדויקת (בפסאודוקוד) שתאפשר לקורא המאמר לממש אותו בקלות. כמו כן יש לצרף תאור מילולי שיסביר איך האלגוריתם פועל באופן אינטואיטיבי בליווי דוגמאות להמחשה. אין צורך לתת תיאור מפורט או פסאודוקוד של אלגוריתמים ידועים בהם השתמשתם.
 4. תיאור המערכת ששמשה למימוש הפתרון. תיאור החלקים הלוגיים שלה. רצוי לצרף שרטוטים מתאימים. אם השתמשתם במימושים שהורדתם מהרשת עליכם לציין במדויק מה בדיוק נלקח ממקורות אחרים (עם הפניה מתאימה) ומה כתבתם בעצמכם. כל זה - רק ברמה אלגוריתמית כללית. הדו"ח לא אמור להכיל מושגים תכנותיים (שמות פרוצדורות, משתנים, מחלקות וכו').
 5. מתודולוגיה ניסויית: תיאור הדרך הכללית שבה ערכתם ניסויים, הפרמטרים שניבדקו, המדדים שהשתמשתם בהם וכו'.
 6. תיאור הניסויים, תוצאות ומסקנות. פרק זה יתאר כל ניסוי, יציג את התוצאות בטבלאות או גרפים וינתח אותן.
 1. כל גרף חייב לכלול שמות משמעותיים לשני הצירים
 2. נא להציג מספרים בצורה קריאה. המספר הזה: 0.8078643864386439 הופיע בדו"ח של סטודנטים. בדרך כלל מספיקות 2 או 3 ספרות אחרי הנקודה.
 7. פרק מסכם: דיון בתוצאות, מה חסר במה שעשיתם, כיוונים להמשך המחקר וסיכום. זהו פרק חשוב ביותר. הקדישו לו את תשומת הלב הראויה.
 2. תיקייה עם כל קבצי התכנה שכתבתם בעצמכם וקבצי קוד שהורדתם ואתם משתמשים בהם. הנכם רשאים להוריד קטעי קוד מהאינטרנט. שיפוט הפרויקט יתייחס לעבודה שעשיתם בעצמכם.
 3. מסמך README עם הסבר קצר לכל קובץ ופירוט מהיכן הקובץ נלקח (במידה והורדתם אותו).
 4. מדריך למשתמש המפרט כיצד בדיוק יש להשתמש בתכנה. עליכם להוסיף מסמך `libs.txt`, המכיל את שמות כל החבילות שהתקנתם לצורך הפרויקט ושקוד המקור שלהן **אינו** מצורף לפרויקט עצמו (בפייתון, למשל, כל תיקייה שניתן להתקין באמצעות `pip install`). בכל מקרה בו התקנת החבילה אינה טריוויאלית, יש לציין לינק מפורש להורדה יחד עם הוראות התקנה מפורטות.
 5. מדריך לבדוק המסביר כיצד בדיוק ניתן לבצע שנית את הניסויים המתוארים במאמר.
 6. תיעוד חיצוני למתכנת שיפרט את מבני הנתונים הפנימיים, החלקים העיקריים בתכנית, המבנה הכללי שלה וצורת הקלט/פלט.

דגשים בקשר לדו"ח הסופי

1. הדו"ח הסופי הוא מרכיב עיקרי בקביעת הציין. אנחנו מאוד ממליצים לקרוא את ההנחיות האלה שוב לפני שאתם מתחילים לכתוב את הדו"ח.
2. קהל היעד של הדו"ח הם המנחה והמרצה. אפשר להניח שיש להם ידע כללי בבינה מלאכותית ולמידת מכונה ולכן אין לכלול בדו"ח רקע כללי בנושאים האלה.
3. אין לכלול בדו"ח תיאורים של אלגוריתמים שלמדתם בכיתה. אתם יכולים לכלול תיאור קצר של אלגוריתמים בהם השתמשתם שאינם בחומר של הכיתה.
4. אסור להשתמש בטקסט או שרטוטים שנלקחו ממקורות אחרים. בחומר הרקע ניתן להשתמש בשרטוטים כאלה אבל אז יש לציין בצורה ברורה את מקור השרטוט.
5. דו"ח הפרויקט צריך להיות דומה במבנהו למאמר מדעי. אין לכלול בדו"ח עצמו מונחים הקשורים לתכנות, לקבצים, לספריות, לשפות תכנות וכו'.

בדיקת הפרויקט

- בדיקת הפרויקט תתרכז בעיקר בדו"ח הסופי. השאלות שאנחנו שואלים את עצמנו בזמן הבדיקה:
1. כמה יצירתיות הפגינו הסטודנטים בבחירת הבעיה ובפתרון שלה?
 2. כמה בהירה הייתה הכתיבה? האם הסבירו היטב את הבעיה ואת דרכי פתרונה?
 3. האם מתדולוגיית הניסויים תקינה? האם נעשו הניסויים הנכונים?
 4. האם תוצאות הניסויים הוצגו ונותחו כראוי?
 5. האם פרק המסקנות כתוב בצורה עמוקה וברורה?
 6. האם ההגשה הייתה תקינה? האם ניתן להפעיל ולשחזר תוצאות?