

המכללה האקדמית כנרת עתודה טכנולוגית

קורס: בינה מלאכותית
תשפ"ד – עבודה סופית
תאריך אחרון להגשה 04/03/2024

1. מוכנות לבית הספר

העבודה תתמקד בבדיקת מוכנות לבית הספר של ילדי גן חובה, הבדיקות נעשות ע"י שני מחקרי שדה, כל מחקר בודק שדות אחרים (מאפיינים)

דרישות

שלב א'

עבור שני המחקרים לשמור את קובץ האקסל לקובץ מצורת CSV המפריד בפסיקים בין כל ערך וערך. לעבד את הנתונים שיצרתם בקובץ ה-SV לפי השלבים שנלמדו בהרצאה:

1. Importing the dataset
2. Taking care of missing data
3. Encoding categorical data
4. Splitting the dataset into the Training set and Test set
5. Feature Scaling

שלב ב'

לעבד את הנתונים ע"י שימוש באחת משפות הסקריפטים (Python או R) ולבחור שלושה תחומים שונים מטכניקות של כריית הנתונים Techniques of Data Mining, בכל תחום צריך לבחור אלגוריתם מתאים. התחומים השונים הם:

1. רגרסיה Regression
2. סיווג Classification
3. ניתוח אשכולות Clustering
4. לימוד חוקיות אסוציאטיבי Association Rule Learning
5. לימוד עם חיזוקים Reinforcement Learning

עליכם להעלות לאתר את:

1. קבצי הנתונים של האקסל הלא מעובד
2. קבצי הנתונים של ה-CSV לפני עיבוד ואחרי עיבוד
3. קבצי המקור של הקוד של שפת/ות הסקריפט שבחרתם הכולל את שלב העיבוד ושלושת האלגוריתמים שבחרתם
4. גרפים כקבצי jpeg שקיבלתם אחרי הפעלת האלגוריתמים

מוכנות לביה"ס

No	מס. סידורי
S	מגדר (1-בת, 2-בן)
Arith	חשבון
Reading	קריאה
Dictation	הכתבה
Exercise	תרגול
Havanat	הבנת הנקרא
Analogy	אנלוגיה
Arithmetic Operations	פעולות חשבוניות
Balloon Counting	ספירת בלונים
Chose the form	בחירת צורה
Click the ...	לבחור
Count and numbers	ספירה ומספרים
Faces	פרצופים
Incomplete Shadow	צל לא שלם
Magic Circle	מעגל קסם
More or less	יותר או פחות
Remember location	זכור מיקום
Set order	קבעו סדר
Triangles	משולשים
R4s	מוכן לבית ספר (1=מוכן, 2=נשאר כיתה)

2. רצח נשים ערביות ויהודיות

העבודה תתמקד באתר של עיתון הארץ, "רצח מדרגה ראשונה", לפי האתר, "מדי שנה נרצחות בישראל כעשרים נשים על ידי הקרובים אליהן ביותר", האתר ממין לפני שנים, משנת 2011 עד שנת 2019. האתר מכיל נתוני משטרה עבור הנשים מהמגזר היהודי והערבי.

<https://www.haaretz.co.il/st/c/prod/heb/global/homicide/>

אתם מתבקשים לכנס לאתר הנ"ל, לבחור איזה מגזר אתם רוצים לחקור, קיימות שלוש אפשרויות:

1. להתמקד בחקירת רצח נשים ערביות בלבד
2. להתמקד בחקירת רצח נשים יהודיות בלבד
3. להתמקד בחקירת רצח הנשים בשני המגזרים

אתם מתבקשים לעבור על כל מקרה ומקרה ולנתח את הנתונים של כל קורבן ע"פ מה שלמדנו בכתה, אחרי בחירת כל המאפיינים בכל מקרה, אתם נדרשים לתת לכל מאפיין ערך מתאים: מספר בדיד, מספר רציף או קטגוריה מסוימת.

דרישות

שלב א'

הקלדת הנתונים לתוך קובץ אקסל, כל משתנה (מאפיין) יש להקיש לתוך עמודה נפרדת. שמות הנשים ישארו בעילום שם, עם כל הכאב, לכל מקרה ינתן מספר סידורי בלבד.

שלב ב'

לשמור את קובץ האקסל לקובץ מצורת CSV המפריד בפסיקים בין כל ערך וערך. לעבד את הנתונים שיצרתם בקובץ ה-SV לפי השלבים שנלמדו בהרצאה:

6. Importing the dataset
7. Taking care of missing data
8. Encoding categorical data
9. Splitting the dataset into the Training set and Test set
10. Feature Scaling

שלב ג'

לעבד את הנתונים ע"י שימוש באחת משפות הסקריפטים (Python או R) ולבחור שלושה תחומים שונים מטכניקות של כריית הנתונים Techniques of Data Mining, בכל תחום צריך לבחור אלגוריתם מתאים. התחומים השונים הם:

6. רגרסיה Regression
7. סיווג Classification
8. ניתוח אשכולות Clustering
9. לימוד חוקיות אסוציאטיבי Association Rule Learning
10. לימוד עם חיזוקים Reinforcement Learning

שלב ד'

להכין דו"ח סופי (לפחות 10 עמודים לא כולל מקורות) לפי המבנה הבא:

1. תקציר Abstract
2. הקדמה Introduction
3. על האלימות ורצח נשים
 - a. בכלל
 - b. על רצח נשים במגזר הערבי, להסתמך על מקורות
 - c. על רצח נשים במגזר היהודי, להסתמך על מקורות
 - d. על רצח נשים בישראל
4. החשיבות של שימוש בטכניקות של כריית מידע בכלל ובנושא בפרט
5. על הנתונים
 - a. איסוף נתונים Data Collecting
 - b. ייבוא נתונים Importing the dataset
 - c. טיפול בנתונים חסרים Taking care of missing data
 - d. קידוד נתונים קטגוריים Encoding Categorical data
 - e. פיצול הנתונים לקבוצת בקרה וקבוצת אימון Splitting data to training and test set
 - f. הבאת הנתונים לקנה מידה אחד Feature Scaling
6. חיזוי מקרה רצח (לפי המגזר) ע"י שימוש בלמידת מכונה Predicting murder of Machine Learning women by using
7. שימוש בשלושה אלגוריתמים שונים משלושה תחומים שונים מטכניקות של כריית הנתונים, כולל תיאור קצר של כל אלגוריתם, צירוף גרפים וקטעי קוד מתאימים לשלבי העיבוד והשימוש באלגוריתם המתאים
8. תוצאות ודיון Results and discussion
9. מסקנות Conclusion
10. מקורות References

עליכם להעלות לאתר את:

5. העבודה הכתובה בשני עותקים אלקטרוניים שונים: קובץ WORD וקובץ PDF
6. קובץ הנתונים של האקסל הלא מעובד
7. קובץ הנתונים של ה-CSV
8. קבצי המקור של הקוד של שפת/ות הסקריפט שבחרתם
9. הגרפים כקבצי תמונה (png או jpeg)

דוגמא חלקית לאיסוף נתונים: להלן תיק של אלמונית עם הנתונים שנאספו ע"י המשטרה



אלמונית

נרצחה 02/08/2018

בת

מקום מזרח ירושליים

מוצא ערבייה ישראלית

הרוצח לא ידוע

כלי הרצח לא ידוע

מוכר לרשויות לא ידוע

סטטוס חקירה לא ידוע



גופת צעירה תושבת מזרח ירושלים נמצאה ללא רוח חיים ברכב נטוש בחברון ולצידה גופת גבר. החשד הוא שמדובר ברצח על רקע מגדרי אולם לא פורסמו בכלי התקשורת פרטים נוספים על הרצח

נתונים חלקיים בהסתמך על הנתונים באתר

#	שם משתנה	סוג	ערכים אפשריים
1	שנת הרצח	מספר רציף	שנה
2	גיל הנרצחת	מספר רציף	מספר שלם
3	יישוב	טקסט	שם עיר או כפר בישראל
4	איזור	טקסט	הגליל, מרכז, צפון, נגב, חיפה, ...
5	מוצא	טקסט	ערביה, יהודיה, אחר
6	דת	בדיד	יהודיה, נוצרית, מוסלמית, אחר
7	עדה	בדיד	אשכנזית, ספרדית, אתיופית
8	תחום הגיל	תחום	1-20, 21-30, 31-40, 41-50, 51-60, 61-70, מעל 70
9	מספר רוצחים	מספר רציף	1 או יותר
10	הרוצח	בינארי	מוכר/ים לנרצחת, זר
11	אופן הרציחה	בדיד	נחנקת, נדקרה, נורתה, פיצוץ, הוטבעה, נשרפה, אחר
12	כלי הרצח	בדיד	סכין, אקדח, רימון, מטען, שריפה, חומצה, אחר
13	מוכר לרשויות	בינארי	0 או 1
14	מעורבות נשים ברצח	בינארי	0 או 1
15	סטטוס חקירה	בדיד	איסור פרסום, בחקירה, הוגש כתב אישום, הואשם, הורשע
16	איפה נמצאה הגופה	טקסט	שם איזור
17	הרקע של הרצח	בדיד	מגדרי, כבוד המשפחה, פלילי, קנאה, נקמה, חיסול חשבונות, לא ידוע, אחר
18	מעורבות בני המשפחה ברצח	בינארי	0 או 1
19	הזמן שעבר עד שנמצאה הגופה	מספר רציף	ימים