פרוייקט סיום בקורס ביג דאטה: הנחיות

כללי

מטרת הפרוייקט היא תרגול כל השלבים הכרוכים בתכנון וביצוע פרוייקט ניתוח מידע. במסגרת הפרוייקט עליכם לשאול שאלת מחקר ולמצוא את המידע אשר בעזרתו ניתן יהיה לענות עליה. לאחר מכן עליכם לבצע את כל הניתוח הנדרש, תוך שימוש בשיטות אותן למדתם בקורס או שיטות אחרות לבחירתכם, על מנת לענות על השאלה אותה בחרתם. לסיכום עליכם להגיש דו״ח אשר מתאר את שאלת המחקר, את האסטרטגיה בה השתמשתם בכדי לענות על השאלה, התוצאות אותן השגתם וכן הקוד בו השתמשתם.
היקף הפרוייקט המצופה הוא של כ-3-4 ימי עבודה עבור כל אחד מהמשתתפים (כלומר כ-7-8 ימי עבודה עבור פרוייקט המבוצע בזוג).

שלבים בביצוע הפרוייקט

1. הגשת תוכנית לפרוייקט, עד ל-18/1/2017 בחצות
2. הגשת דו״ח סיכום לפרוייקט – לא יאוחר מתחילת הסמסטר הבא - 4/3/2018

תכנון הפרוייקט

בשלב ראשון עליכם להגיש תיאור של הפרוייקט אותו ברצונכם לבצע. התיאור צריך לכלול את החלקים הבאים:

1. **שאלת המחקר – מה אתם רוצים להשיג?** שאלת המחקר יכולה להיות שאלה עליה אתם רוצים לענות, או היפותזה אותה אתם רוצים לבחון, או שהיא יכולה להיות מטרה אותה אתם רוצים להשיג על ידי ניתוח מידע והפעלת אלגוריתמים. חשוב שהשאלה עליה אתם מנסים לענות תהיה מוגדרת היטב ושניתן יהיה לענות עליה בעזרת מידע קיים. על שאלות כלליות מדי או לא מוגדרות היטב לא ניתן יהיה לענות בצורה מספקת.
דוגמאות:
   - **האם ניתן למצוא מאפייני טקסט שמפרידים אימיילים שהם ספאם מאימיילים שהם לא ספאם?** המטרה ברורה, קיים מידע שבו ניתן להשתמש.
   - **האם קיים שינוי במספר ימי השלג באמריקה הצפונית לאורך השנים בתוצאה משינויי האקלים בעולם?** קיים מידע, אנחנו מגדירים בדיוק את מה שאנחנו רוצים למדוד.
   - **האם ניתן לבנות מנבא לתוצאות משחקי כדורגל בליגה האנגלית לכדורגל שיעשה שימוש בתוצאות קודמות ושיהיה טוב יותר מניחוש אקראי?** המטרה מוגדרת היטב (להיות טובים יותר מניחוש אקראי), סוג המידע מוגדר וזמין (תוצאות קודמות).

השאלות הבאות הן כלליות מדי או שאינן מוגדרות היטב:

- **מיהו שחקן הכדורגל הטוב בעולם?** לפי אילו קריטריונים?
- **האם בני אדם גורמים לשינויי אקלים עפנ״י כדה״א?** השאלה כללית מדי. מענה על שאלה כזו דורש הוכחת מספר שלבים שכל אחד מהם דורש העלאת היפוטזות והוכחתן. אם מניחים שבני אדם גורמים לשינויי אקלים על ידי פליטת גזי חממה אז צריך להוכיח שבני אדם אכן פולטים כמות משמעותית של גזי חממה, שגזי חממה אכן גורמים להתחממות, ושגזי החממה הם הגורם המשמעותי ביותר מכל הגורמים האפשריים שגורמים להתחממות.

- **האם סטודנטים במכללת תל חי שמגיעים למבחנים אחרי שינה של 7 שעות בלילה לפחות משיגים ציונים טובים יותר?** השאלה מוגדרת היטב אבל רב הסיכויים שלא ניתן יהיה למצוא מידע בכדי לענות על השאלה בזמן סביר.

2. **באיזה מידע תשתמשו?** עליכם לספק לינק למידע בו תשתמשו, או להסביר איך תאספו את המידע. כמובן שזמינות המידע גם מכתיבה את סוג השאלות אותן ניתן לשאול. תארו איך נאסף המידע, מה הוא מכיל ואיך הוא מתאים לצורך מענה על השאלה אותה בחרתם.
3. **אילו ניתוחים תבצעו?** תארו אילו בדיקות תבצעו ואיך הן עוזרות לכם בכדי להשיג את מטרת המחקר.
4. **באילו שיטות תשתמשו?** על השיטות להתאים לשאלה אותה אתם שואלים. אתם מצופים להשתמש בכמה מהשיטות אותן למדתם (ומוזמנים להשתמש בשיטות נוספות). המחקר צריך לכלול אלמנט כלשהו של בדיקת מובהקות סטטיסטית, או של הרצת אלגוריתמים לניבוי או clustering בנוסף לטכניקות exploratory data analysis.

היקף מסמך תכנון הפרוייקט צריך להיות כ-2-1 עמודים.

הגשת דו"ח סיכום

דו"ח הסיכום צריך לתאר את הפרוייקט כולו:

1. רקע על הנושא: תיאור התחום, מה ידוע על הנושא אותו רציתם לחקור. המנעו מהעתקה של טקסט ממקורות המידע ותארו את הנושא במילים שלכם.
2. שאלת המחקר - יכולה להיות זהה לשאלת המחקר בתכנון הפרוייקט אם לא היה שינוי
3. תיאור המידע בו השתמשתם – יכול להיות זהה לתיאור אותו נתתם במסמך תכנון הפרוייקט
4. תיאור כללי של הצעדים בביצוע הפרוייקט ואיך הם תרמו להשגת מטרתכם.
5. תיאור תוצאות – כולל פענוח התוצאות, טבלאות, איורים וכו. הוסיפו תיאור לכל איור או טבלה: מה מייצגים הצירים/עמודות? איזה מידע מוצג? התאמצו ליצור איורים אסתטים ככל האפשר.
6. דיון בתוצאות: האם הצלחתם להשיג את מטרת הפרוייקט? מה מצאתם ואיך התוצאות מוכיחות זאת? על אילו חלקים בשאלת המחקר שלכם לא הצלחתם לענות?
7. צרפו את הקוד בו השתמשתם. על הקוד להיות מתועד כנדרש. ציינו באיזה חלק מהקוד השתמשתם בדו"ח עצמו.

הדו"ח צריך להיות ברור ולהכיל את כל המידע הנדרש. מניסיון העבר סד"ג של הדו"ח הסופי הוא כ-10-5 עמודים (אבל אין מגבלה).

הערות

סטודנטים אשר מעוניינים ברעיונות לפרוייקטים מתחומי העניין שלי (גנומיקה של חיידקים, ספורט, מדעי הטבע באופן כללי) מוזמנים ליצור אתי קשר.

בהצלחה!