

הנחיות לפרויקט הסיום – סטטיסטיקה למדע הנתונים (90% מהציון הסופי)

מטרה כללית

הפרויקט נועד לבדוק את יכולתכם לישם באופן עצמאי ומקצועי את רוב הכלים והחישיבה הסטטיסטית שנלמדו בקורס על מנת נתונים אמיתי – מהשאלה ועד למסקנות בעלות משמעות.

נושא נתונים

- כל סטודנט מקבל נושא מוגדר מראש + קובץ נתונים אחד
- אין אפשרות לבחור נושא אחר או להחליף / להוציא קובץ נתונים

מה נדרש להגיש

- מסמך / קובץ / מצגת / דוח – בכל פורמט שנראה לכם מתאים
- המטרה היא להראות הבנה عمוקה, חשיבה סטטיסטית נcona, קבלת החלטות מושכלות ויכולת להעביר מסר ברור
- אין מבנה חובה – אתם מחייבים איך לארכן את החומר (רצף לוגי, סיפור, דוח קלאסי, מצגת מורחבת, נוטבוק עם הסברים וכו')

אלמנטים שכדי שיויעו (בצורה כלליה)

- הצגת השאלה / המטרה העסקית-מחקרית
- תיאור ראשוני של הנתונים (גודל, סוג, משתנים, בעיות בולטות)
- מה עשיתם עם הנתונים לפני הניתוח (חסרים? חריגים? טרנספורמציות? משתנים חדשים?)
- אילו שיטות ניתוח בחרתם – ולמה דוקא אותן
- תוצאות עיקריות (טבלאות / גרפים / מדדים)
- אבחון ואיומות (שאריות? הנחות? השוואת מודלים? over-fitting?)
- מסקנות ישירות לשאלה + פרשנות בעלת משמעות
- מגבלות + מה לא הצליחם לעשות / מה היה יכול לשפר את העבודה

דרישות טכניות מינימליות

- קוד נקי ומתווד (Python או R או שילוב) – חייב להיות ניתן להריץ אותו
- גרפים קריים עם כוורות ותוויות
- טבלאות מסודרות (לא צילומי מסר מבולגנים)

קריטריוני הערכה עיקריים

- טיפול עמוק ומנומך בנתונים + EDA אינט'
- בחירה, יישום ובדיקה מודלים
- השוואה בין מודלים והסקת מסקנות מבוססות
- פרשנות עסקית/מדעית + מודעות למוגבלות
- אינט' הכתיבה, הייזואלייזציה והארגון

הפרויקט זהה הוא המרכיב המרכזי של הציון בקורס.

הוא נועד לדמות עבודה אמיתית של מדען/ת נתונים – מהשאלה ועד המלצתה המעשית.

בהצלחה רבה!