

## "דימות נתונים" – עבודת גמר

**הקדמה:**

במסגרת הקורס, מצ"ב פרטים על עבודת הגמר אשר יש להגיש. מטרת העבודה היא להכיר לסטודנט בצורה אישית ובלתי אמצעית את מגוון הנושאים אשר נלמדו בקורס. בין היתר, יש להבין את הנושאים שעברנו עליהם בקורס הקודם בצורה משמעותית וטובה.

**משימות על:**

במסגרת עבודת הגמר, על כל סטודנט לבחור DATASETS אחוד מתוך האתר KAGGLE. ניתן לעבוד עם הDATASET שעבדתם איתם בסמסטר הקודם. בנוסף, יש dataset המצורף לעבודה אשר חובה לעבוד איתו.

**פירוט המשימות:**

* נא לעבור על הקורס של git באתר udacity. . זו מטלה חובה. בנוסף, יש לשנות את המחברות בגיטהב שלכם דרך commits שונים.
* יש לבחור שני datasets, אחד שעושה קלסיפיקציה ואחד שעושה רגרסיה. חשוב להציג את הנתונים **בצורה גרפית טובה שניתן ללמוד ממנה תובנות לגבי ה-DATA.** **זה חלק משמעותי מהציון**
* יש להשתמש בין היתר באלגוריתמים שלמדנו במהלך הסמסטר. בין היתר: PCA, ICA, K-MEANS, NAÏVE BAYES, ADABOOST וכו'
* מה מידת הדיוק של המודל שלכם יחסית למנצחים בתחרות KAGGLE? במה הוא יותר טוב ממודל 0?
* בנוסף, יש לבצע קלסיפיקציה של DATA אמיתי בדיוק כמו שראינו בתרגול מספר 2. פרטים לגבי מטלה זו מרוכזים בצורה מסודרת בנספח שבסוף מסמך זה. מטרת השאלה היא להגיע לתוצאה שניתן להשוות אחוזי דיוק בין כל הסטודנטים.



בהצלחה לכולם

**נספח א – פירוט עבודת סיווג בקורס "דימות נתונים"**

**מטרה**

מטרת העבודה היא לסווג בין שלושה מצבים שונים באופן בו אנשים מתקשרים אחד עם השני. הראשון הינו מצב ספונטני (אוטונומי) בו שני אנשים מזיזים את הידיים שלהם בצורה חופשית אחד מול השני. השני הוא תנועה סינכרונית בה שני האנשים מזיזים את הידיים ביחד והשלישי הוא תנועה במצב לבד. בו רק הצד אחד מזיז את הידיים.

הרעיון הוא להסתכל על דפוסי הידיים ולנסות להסיק מהם האם מדובר במצב לבד, ספונטני או סינכרוני.

כפי שראיתם בתרגיל בית מספר 3, יש לכם כבר את כל המידע ב-DATAFRAME מסודר.

יש גם להציג את ההבדלים בצורה גרפית יפה.

בונוס יינתן ל-3 סטודנטים שיגיעו לרמות הדיוק הגבוהות ביותר. בנוסף, מצ"ב מאמר מדעי שפורסם על העבודה הראשונית הזו. כדאי מאד לקרוא אותו על מנת להבין מה נצרך מכם בעבודה זו.

שימו לב שבתיקיה המצורפת, יש תת-תיקייה שנקראת TRAINING ורק עליה יש לעבוד באימון הרשת. בנוסף, יש תת תיקיה הנקראת TESTING ורק עליה יש לעבוד בוולידציה.

**הערות נוספות:**

* במצב ALONE, קיימת רק הקלטה של יד אחת (יד שמאל). יש לבדוק את המצב הזה אל מול קובץ שנקרא HandRight ובו הקלטה של תנועות ביד ימין. שימו לב שאת כל מצבי ה-ALONE יש לבדוק אל מול הקובץ הזה.
* יש ספריות אשר יש שם 2 מקבצי הקלטות. במקרה כזה, יש לקחת את המקבץ האחרון.
* חשוב מאד לא לקחת את ה-7 שניות הראשונות של הקובץ.
* הרעיון פה הוא ליצצר פרמטר שבודק סינכרון בין שני אנשים. חשבו בעצמכם אילו פרמטרים חשובים ולמה.
* מטרת העבודה היא לסווג תנועות ידיים בין שני אנשים למצב (ALONE, AUTONMOUS ו—סינכרוני). ניתן לחשוב על כל SAMPLE כעל שנייה של הקלטה המורכבת מ-5 timestamps.
* חשוב להבין מה אחוז הדיוק ומהו הf-score של האלגוריתם שלכם. ניתן לייצר אלגוריתמים שונים אבל בשום מקרה אסור לאמן את המערכת על המידע מספריית הוולידציה.

בהצלחה לכולם.