**דו"ח פרויקט גמר פייתון**

**הקדמה:**

תחום עיבוד הסיגנל המוחי התפתח בשנים האחרונות יחד עם שאר הטכנולוגיה. דבר שהביא ליצירתם של כלים פשוטים ונוחים לשימוש, נגישים לציבור הרחב ולמטפלים, יותר מאשר המכשיר הקיים במעבדות. בעקבות כך רצינו לבדוק האם איכותם של מכשירים אלו טובה מספיק על מנת לשרת משתמשים או מטפלים ביום יום. בניסוי זה השתמשנו בנתונים של מכשיר ה-MUSE ולו 5 אלקטרודות מתוכן 4 מקליטות את הסיגנל המוחי, מה שגורם לו להיות ברזולוציה נמוכה באופן יחסי אל מול מכשיר מדידת ה-EEG המוכר והנפוץ בשימוש במעבדות.

התמקדנו בניסיון לזהות 3 מצבים מנטליים – רגוע, ניטרלי ומרוכז. זאת באמצעות ניתוח נתוני EEG שנמדדו במשך 60 שניות במכשיר MUSE בארבעה נבדקים: שני גברים שתי נשים. כל נבדק נמדד פעמיים בכל מצב מנטלי. המטרה הייתה לראות האם ניתן להבחין בין המצבים השונים, כדי בהמשך, ליצור ממשק המתחבר למכשיר המדידה ויכול לאבחן למשתמשים, או למטפלים שלהם, את המצב המנטלי בו הם מצויים.

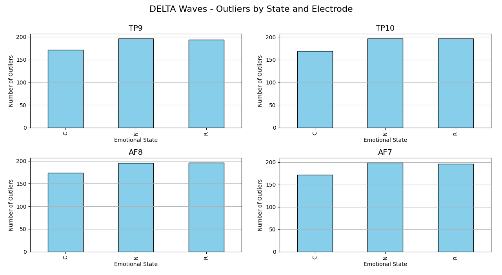
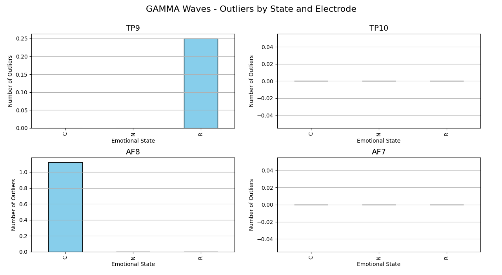
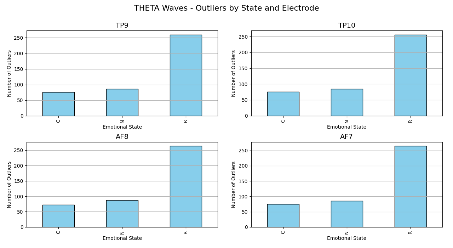
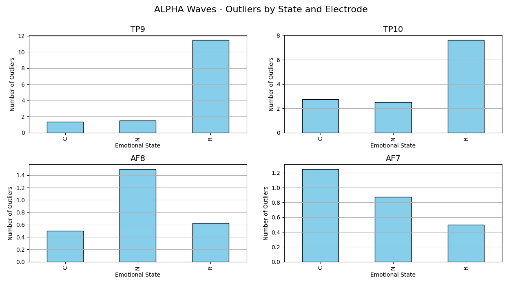
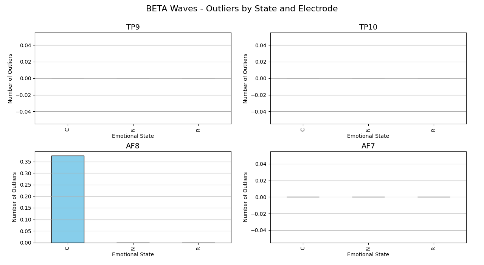
הצלחה בפרויקט כזה והבנה משמעותית של המצבים המנטליים בהם אדם יכול להיות מצוי, היא בעלת חשיבות עצומה. תחילה ניתן לפתח את האלגוריתם לסווג עוד מצבים מנטליים שונים כמו סטרס או חרדה. שנית ניתן להשתמש בסיווג המצבים באופן יומיומי ופשוט ביותר. למשל, נוכל להשתמש בסיווג בבתי הספר כדי לבחון האם תלמידים נמצאים במצב מרוכז ומוכנים לקלוט ידע, כך נוכל לשפר את הלמידה ולהפוך אותה לאפקטיבית ויעילה יותר. כך נוכל להשתמש בכלי הזה גם לעוד תחומים רבים ולשפר את יעילות ואיכות החיים.

**שיטות:**

כדי להפוך נתונים גולמיים של EEG לניתוח מצבים מנטליים המיועדים לשימוש בממשק מתאים, היו דרושים שלבים רבים. תחילה התעסקנו בולידציה של נתונים. שלבי הולידציה כללו:  
בדיקה שהקובץ מסוג CSV, ערכים מתאימים בעמודות הקובץ, ערכים מספריים בלבד בתאי המדידות, הכללת קבצים להם מעל 11000 מדידות (45 שניות מדידה בערך) וגם קבצים בעלי פחות מ-2200 קבצים חסרים (5% אחוז נתונים חסרים מתוך 45 שניות), השלמת ערכים חסרים – במידה ואכן חסרים, ושמירת קובץ חדש להמשך עיבוד.  
לאחר מכן, התעסקנו בעיבוד הנתונים שקיבלנו:

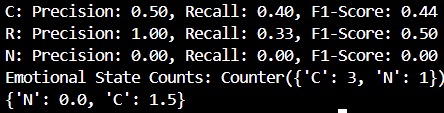
* **Resampling:** הבאת המדידות לאחידות דגימה של 200 הרץ (הערכים המקוריים הם בין 150-200 הרץ).
* **החלקת הנתונים:** באמצעות החלקה אקספוננציאלית עם אלפא של 0.1.
* **הסרת Outliers:** בסף של מעל לשתי סטיות תקן (אלו הם הרעשים שרצינו להסיר).
* **השלמת נתונים**: באמצעות השלמה קדימה ואחורה ליצירת רציפות במדידה.
* **התמרה פורייה:  
  -** לאחר ההתמרה מצאנו שוב Outliers והשארנו את 5% העליונים של עוצמת סיגנל לפי תדר.  
  **-** ספירה של Outliers וסיווגם לפי טווחי תדרים אופייניים במוח (ALPHA, BETA, THETA, GAMMA, DELTA).
* **יצירת קובץ חדש:** שמסווג לנבדק ספציפי, פר אלקטרודה ופר כמות Outliers בכל טווח תדר.

מהקובץ החדש עברנו לשלב של ניתוח סטטיסטי, שכלל שימוש במספר מבחנים סטטיסטיים על מנת לבדוק הבדלים בין הסינגלים בכל מצב מנטלי. ניתוח ANOVA חד כיווני שבוצע על סוגי סיגנלים ואלקטרודות במטרה לבדוק אם קיימת שונות מובהקת בין המצבים המנטליים. במבחן בדקנו את היחס בין השונות בתוך הקבוצות לשונות בין הקבוצות, תוך בדיקת מובהקות. בנוסף ביצענו מבחני טי טסט בין זוגות של קבוצות (C vs R, C vs N, R vs N), בליווי חישוב גודל אפקט. לאחר מכן בדקנו את הקשרים בין כל זוגות האלקטרודות לכל מצב רגשי ע"י מתאמי Pearson (ליניארי) ו-Spearman (מונוטוני), כך הצלחנו לזהות דפוסים וקשרים בתוך הנתונים תוך התמודדות עם ערכים חריגים. הצגנו את התוצאות הקורלציה בגרפים של מפת חום וביצענו באיור 1 ויזואליזציה של ה-Outliers של כל אלקטרודה בכל גל לפי מצב מנטלי.  
לבסוף, כתבנו קוד שתפקידו לחזות מצב רגשי של נבדק ע"י למידת מכונה. השתמשנו במודל Random forest תוך שימוש ב-200 עצים וזיהוי מצבים רגשיים ע"פ מאפייני גלים ואלקטרודות. הכנסנו גם מסקנות שלנו מהגרפים על מנת להשוות ולוודא שיש הלימה בין השניים.

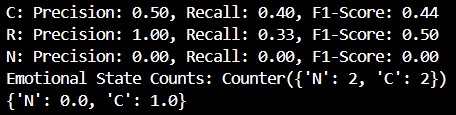


איור 1: הגרפים מתארים את כמות ה- Outliers שנמצאת בכל אלקטרודה, בכל גל, מחולק על פי המצבים הרגשיים.

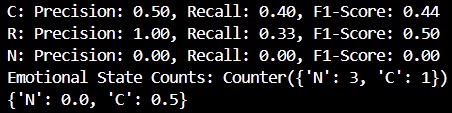
**תוצאות:**



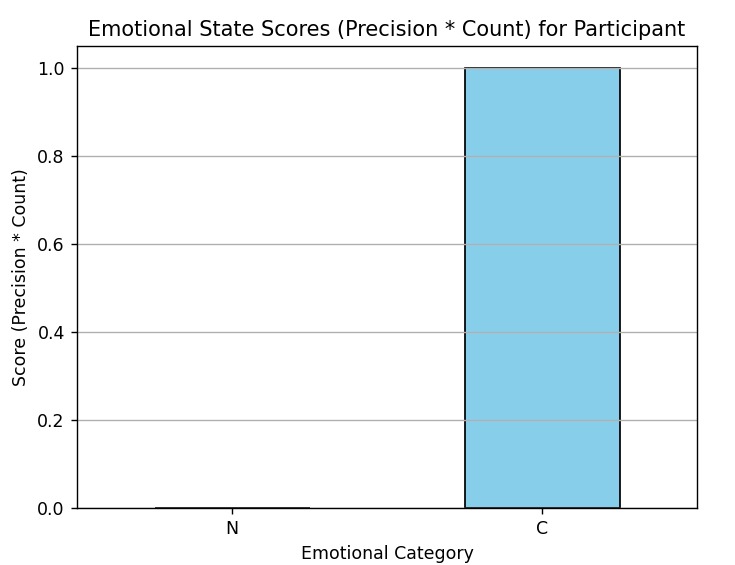
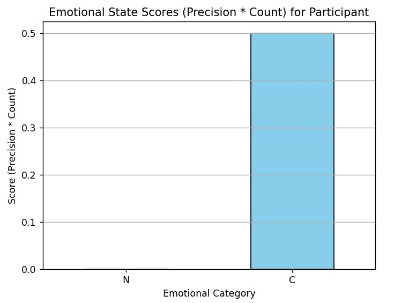
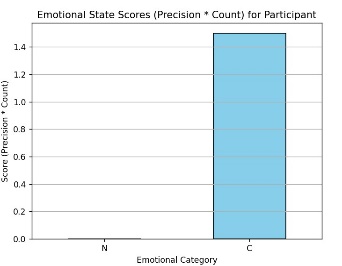
**A**



**C**



**B**



איור 2: מוצגת הרצה של נבדקים שונים: A- במצב מרוכז, B- במצב ניטרלי, C- במצב רגוע. לכל נבדק מוצג מילון וגרף.

התוצאות המוצגות הן בגרף והן במילון מייצגות את התחזית המשותפת של המודל (Random forest) ושל התנאים הבסיסיים שהצבנו לקוד. זאת בשביל לסווג לפי הסטטיסטיקות שנעשו על נתוני המשתתפים המקוריים את נתוני המשתתפים החדשים.

באיור 2 ניתן לראות את הניקוד שקיבל כל מצב מנטלי במבחנים שאנחנו ניסחנו, כפול רמת הדיוק לאותן תוצאות, פר מבחן, לפי החיזוי של המודל (המודל קבע איזה אחוז מהניקוד תואם את מסד הנתונים עליו הוא מתבסס כדי לסווג את הנתונים). כלומר, אנחנו הגדרנו את המבחנים שייקבעו כמה דומה הנבדק החדש לאחד המצבים המנטליים והמודל קבע כמה זה נכון.

לפי נבדק A באיור 2 ניתן לראות שהציון ואחוז הדיוק שלו אכן תואמים נבדק חדש שנמדד במצב מרוכז.   
לעומת זאת בנבדק B ניתן לראות כי המבחן אכן זיהה כי הנבדק (שנמדד במצב ניטרלי) הוא אכן ניטרלי אך המודל לא מאשר זאת מבחינת דיוק, ולכן לא ניתן לסווג בהתאם את הנבדק.   
בנבדק C ניתן לראות תופעה הפוכה, המודל אישר את נכונות הניקוד של מצב הנבדק בפועל (רגוע) אבל המבחנים לא נתנו למצב זה ניקוד כלל.

**מסקנות:**

הקוד יודע לסווג נבדק מרוכז כמרוכז, אך לא יודע להבדיל בין שאר המצבים למרוכז ומסווג גם אותם כמרוכז. עם זאת, ניכר כי המודל או ההתניות הבסיסיות (לפחות אחד מהם) כן מצליחים לזהות באיזה מצב מדובר אך מבטלים זה את זה.

**דיון:**

לפי התוצאות והמסקנות נראה כי אם היה מדגם גדול יותר, אז היה יותר מידע למודל ללמוד. ואם היו נמדדים מצבים מנטליים מגוונים יותר באופי הנתונים הנכנסים למערכת, זו הייתה מצליחה לסווג אותם טוב יותר. אם היו יותר נבדקים המודל היה מדויק יותר וכנראה לא היה מבטל את הניקוד מהמבחנים ולא פוסל תוצאה נכונה שהתקבלה מהמבחנים. כמו כן בגלל דמיון בין מצבים כמו ניטרלי ורגוע ורק 4 אלקטרודות שאותן מודד המכשיר, אין רזולוציה גבוהה מספיק ולכן מבחני הניקוד לא מדויקים, מה שגורם לסיווג מוטעה, דבר שמבטל זיהוי נכון של המודל.