

דוח הגשה פרויקט סופי בפייתון

מגישות: ורד כהן 315322677, אביבה הלר 015972045

הקדמה

לפרויקט שלנו בחרנו בדאטה "Sleepy Driver EEG Brainwave Data". מאגר הנתונים מכיל סיגנלי EEG שנאספו מארבעה נהגים בזמן שהיו במצב ערות ושינה, ומטעמי בטיחות הם לא נהגו בפועל בזמן איסוף האותות. סיגנלי ה-EEG כוללים אותות המייצגים פעילות חשמלית במוח, שנמדדו ברצועות תדרים של גלים שונים (כגון אלפא, בטא, דלתא) וכן הערכה כוללת של המכשיר בנוגע למצב הנהג אם הוא במצב של attention או במצב meditation (לא תואם או קשור להיותו ער או ישן בזמן המדידה). כל דגימה שנלקחה בכל כ-שנייה נרשמה בשורה נפרדת בקובץ CSV. כמו כן הדאטה כלל עמודה (בשם: classification שבו צויין 0 אם הקריאה נלקחה מנהג ישן או 1 אם נלקח מנהג ער). למען הבנה כללית יוצרי המאגר בנו אותו במטרה ליצור ממשק מוח-מחשב שיתריע בפני נהגים מנומנמים ובכך למנוע תאונות.

שאלת המחקר שהובילה אותנו בפרויקט הנוכחי היא מהו הקשר בין אורכי גל שונים או דפוסים משולבים שלהם שנאספו על ידי ה-EEG לבין ההקצאה של "attention" או "meditation" שהוקצתה למשתתפים על ידי המכשיר. הקביעה attention או meditation הם המהווים את המשתנה התלוי, ואורכי הגל השונים מהווים את המשתנה הבלתי תלוי.

לשם כך היה עלינו לנקות את הנתונים בהתאם לצרכינו, לעשות אנליזות הכוללות חקירה של מערך הנתונים והדמיית קשרים בין משתנים ולהיעזר בהן על מנת להפיק תובנות על מגמות ודפוסים במטרה למצוא תשובה.

שיטות

תהליך

כדי שנוכל לעבוד בצורה היעילה ביותר התחלנו בשרטוט קווים כלליים ומבנה פרויקט לפיו נרצה לעבוד:

- src - תיקייה שתכיל 3 קבצים שונים של הקוד: data visualization.py, data cleaning.py, data analyzing.py. בנוסף תכיל את קובץ init שיאתחל את החבילה.
- tests - יכיל 3 קבצים שונים שבודקים תקינות של כל אחד מקבצי הSRC בנוסף לקובץ init שיאתחל את החבילה.
- main.py - קובץ נקי שייבא מקבצי הsrc את הפונקציות הרלוונטיות ויריץ את הניקוי, האנליזה והוויזואליזציה.
- READ ME - קובץ טקסט שיתאר את הרציונל לפרויקט והשיטות בהן השתמשנו.
- toml - הקובץ יגדיר את המידע הבסיסי על הפרויקט, תלויות חיצוניות לפיתוח, ותצורת בניית הפרויקט באמצעות poetry.
- workspace - הקובץ יגדיר את ההגדרות עבור סביבת העבודה בשביל להפוך את הפרויקט ליותר נוח לניהול.

צורת העבודה הייתה לרוב יחדיו, וכאשר כל אחת עבדה בנפרד על קבצים מסוימים נעזרנו במשיכה ודחיפה מ-
.git

אקספלורציה וניקיון

כעת, לפני ניקיון הנתונים וכל מה שמגיע אחריו, עשינו אקספלורציה של המאגר כדי שבין כיצד הוא בנוי- כמה שורות, כמה עמודות, האם יש NaNs, האם משהו שגוי וכן הלאה.

תוך כדי בנינו את ה- TOML והתקנו לאורך מהלך העבודה את החבילות כאשר היה צורך. לדוגמא, בשביל שנוכל להשתמש ב- pandas ב- DATA.HEAD הוספנו את pandas. לצורך בדיקת מבנה ה- DATA, נעזרנו ב- INFO.DATA בשביל לראות מה גודל המאגר, ובעזרתו בדקנו את מצב ה- NaNs. הממצאים הראו כי הם לא קיימים, ולכן לא היה צורך בהשלמה. בדקנו DUPLICATE ROWS, ולא היו. במטרה לבדוק ה- OUTLIERS ולהבטיח שערכים קיצוניים יוסרו, בנינו פונקציה remove_outliers שעוברת על הדאטא ומוחקת את הנתונים באמצעות שיטת Z-Score עם סף של 3.5.

במטרה לאפשר גישה מכיוון אחר לניתוח הנתונים בנינו פונקציה נוספת שמפצלת את הנתונים לשני תתי קבוצות: מערך נתוני שינה הכולל תצפיות עם תווית של מצב שינה (0) ומערך נתוני ערות הכולל תצפיות עם תווית של מצב ערות (1). לאחר יצירת ה- datasets השונים, מחקנו את עמודות ה- classification מאחר שלא נצטרך אותם להמשך האנליזה. מאחר ובדקנו את כל הדברים האלו והכל תקין, לא נותר לנו הרבה לנקות ולכן קובץ ה- cleaning שלנו מכיל יותר אקספלורציה ופחות ניקיון.

אנליזה

במעבר לחלק האנליזה, ייבאנו את שלושת סטי הנתונים (dataset, dataset_0, dataset_1). ברצוננו היה לחפש בכל סט נתונים קשרים ליניאריים או לא ליניאריים בין אורכי הגל השונים לבין הקביעה על מצב attention או meditation כדי להבין מי מביניהם השפיע על הקביעה ובאיזה רמה.

לבחינת קיומם של קשרים ליניאריים בדקנו קורלציות בין גלים שונים לבין הקביעה האם האדם היה במצב meditation או attention באמצעות פונקציות מובנות. לאחר מכן בנינו פונקציה categorize_correlation שמדרגת כל ערך קורלטיבי שנמצא לפי weak/moderate/strong. ביצענו את האנליזה על הדאטא המלא וכן על הסטים המפורדים למצב ער וישן. על ידי סיווג הקורלציות הללו, זיהינו דפוסים ביחסים בין מאפייני גלי המוח ובחנו האם מצב המשתתפים (ערות או שינה) השפיע על יחסים אלו.

בשלב הבא חיפשנו מגמות ו- clusters הנראים לעין על ידי שימוש ב scatter plot.

בהמשך בחנו קיומם של קשרים שאינם ליניאריים (במודל ריבועי) ע"י הפונקציה analyze_r_squared הבודקת את הקשרים הללו בין גלי המוח לבין המשתנים attention או meditation. ערך R^2 , או מקדם הקביעה, מודד את שיעור השונות במשתנה התלוי (רמת attention או meditation) שמוסבר על ידי המשתנה הבלתי תלוי (מדד גלי המוח) באמצעות מודל הרגרסיה. כלומר, הפונקציה קובעת ערך R^2 המאפיין עד כמה מדדי גלי המוח מסבירים את רמות attention או meditation באמצעות מודל ריבועי. הסיבה שבחרנו להתחיל במודל ריבועי

(מתוך מגוון אפשרויות לא-לינאריים) הוא כי מודל ריבועי הינו פשוט להתאמה ולפירוש, ומתאר קשרים לא-לינאריים בסיסיים כמו עקומות בצורת U. קשרים מעין אלו רלוונטיים הרבה פעמים במערכות ביולוגיות והתנהגותיות שבהן הקשרים עוקבים אחר דפוסים פשוטים כמו שיאים, התמדה ואחר כך שחיקה, יחס תגמול-מוטיבציה ועוד.

ערכי R^2 סווגו לפי התאמה של low/medium/high.

לאחר שלא זוהו ממצאים משמעותיים עד כה, הרחבנו את הניתוח לשימוש בפונקציה הכוללת שימוש בחבילה של Random Forest. בשימוש של Random Forest לבעיות רגרסיה, הכוונה ליישום ספציפי של האלגוריתם לצורך פתרון בעיות רגרסיה, שבהן המטרה היא לחזות ערך מספרי רציף (continuous value). בשיטה זו האלגוריתם מחשב ממוצע של התחזיות מכל העצים על מנת לספק את הערך החזוי הסופי של משתנה המטרה (למשל, רמת attention או meditation הצפויים), ובמקביל מעריך את חשיבותן של התכונות השונות בקביעת משתנה המטרה. שיטה זו יעילה במיוחד בהתמודדות עם קשרים לא-לינאריים ובהדגשת המבנים המשמעותיים ביותר במאגר הנתונים. כמו כן, מעבר לזיהוי הקשרים הישירים בין כל גל מוח לרמת הקשב או המדיטציה, Random Forest מאפשר לחשוף דפוסים חבויים ושילובים מורכבים של גלים המשפיעים יחד על משתנה המטרה. גישה זו מספקת תובנות מעמיקות יותר על האופן שבו גלי מוח שונים יביאו לקביעת רמת attention ו meditation.

באמצעות הפונקציה random_forest_analysis, נוכל להעריך את תרומתו של כל גל מוח לציון ה- R^2 באמצעות ניתוח חשיבות התכונות. ניתוח זה ימדוד עד כמה כל תכונה מפחיתה את השגיאה במהלך שלב האימון. התובנות שיתקבלו יאפשרו לזהות אילו חלקים במאגר הנתונים (למשל, גלי מוח מסוימים) משפיעים על דיוק התחזיות ומהי מידת תרומתם לציון ה- R^2 .

בחלק היוזאליזציה בנינו את הפונקציות הבאות:

plot_correlation_heatmap – מוציאה תרשים heatmap שמציג את כל הקורלציות שנמצאו בדאטא סט מסויים.

plot_brainwave_relationships – מוציאה scatter plots שמציגים את הפיזור של כל מדידות אורכי הגלים השונים בטווח של מצב meditation/attention.

visualize_r_squared – מוציאה heatmap להצגת ערכי הרגרסיה שנמצאו.

visualize_feature_importance – מוציאה היסטוגרמה עם אינדקס של חשיבות כל גל לקביעת מצב ה-attention או meditation.

טסטים

בתיקיית tests כתבנו קוד לפונקציות שבדקות את תקינות כל אחת מן הפונקציות בקבצי ה src.

בקובץ test_cleaning:

- טסטים לפונקציה split_dataset_by_classification:

- ✓ `test_split_dataset_invalid_column` - בונה מיני דאטא ומוודאת שאם חסרה עמודה הפונקציה מעלה שגיאה.
- ✓ `test_split_dataset_invalid_values` - בונה מיני דאטא עם ערכים החורגים מהטווח של ערכי הסיווג של שינה/ערוץ (0-1) ומוודאת שהפונקציה לא רצה ומעלה `Exception`.
- ✓ `test_split_dataset_NaNs_numeric` - בודקת שהפונקציה מעלה שגיאה אם היא מקבלת נתונים שאינם מספריים או `NaNs`.

בקובץ `test_analysis`:

- טסטים לפונקציה `categorize_correlation`:

- ✓ `test_categorize_correlation_positive` - בונה מיני דאטא המכיל מטריקס קורלציה ומוודאת שהפונקציה עובדת בצורה חלקה ולא מעלה `Errors`.
- ✓ `test_categorize_correlation_correctness` - בונה מיני דאטא המכיל מטריקס קורלציה ומדפיסה את הסיווג לפי הערכים, כך שבצורה ויזואלית נוכל לראות את תקינות הפונקציה לפי הסיווג המוצג.

- טסט לפונקציה `analyze_r_squared`:

- ✓ `test_no_variance_in_data` - בונה מיני דאטא המכיל עמודות עם ערכים ומוודאת שהפונקציה מדלגת על עמודות ללא שונות בהם.

- טסט לפונקציה `random_forest_analysis`:

- ✓ `test_random_forest_analysis` - בודק את תקינות הפונקציה (positive).

בקובץ `test_visualization`:

- טסט לפונקציה `plot_correlation_heatmap`:

- ✓ `test_plot_correlation_heatmap_invalid_values` - בודקת שבמידה והפונקציה תקבל מאגר נתונים עם ערכים לא תקינים ולא צפויים לקלט (מספר גבוה מ-1) היא תעלה שגיאה ולא תרוץ.

- טסט לפונקציה `plot_brainwave_relationships`:

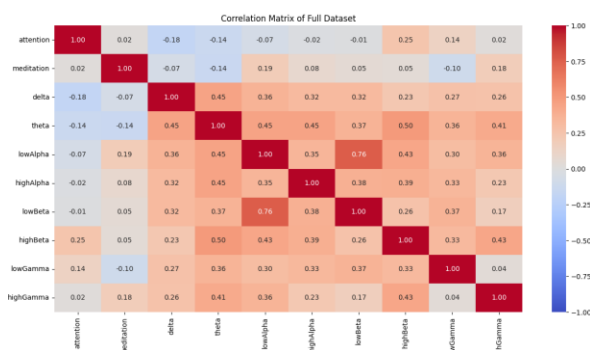
- ✓ `test_plot_brainwave_relationships_positive` - מוודאת שהפונקציה עובדת בצורה חלקה ולא מעלה שגיאות.

אם כך, לאחר ניקוי הנתונים, יצירת פונקציות של ניתוח וויזואליזציה ובדיקת תקינותם באמצעות טסטים- כתבנו קוד ב main.py שמייבא מקבצי SRC את הפונקציות הרלוונטיות ואת שלושת סטי הנתונים עליהם נרצה לעשות את האנליזה.

תוצאות

1. קורלציה לינארית

אנליזה של מערכת נתונים מלאה:

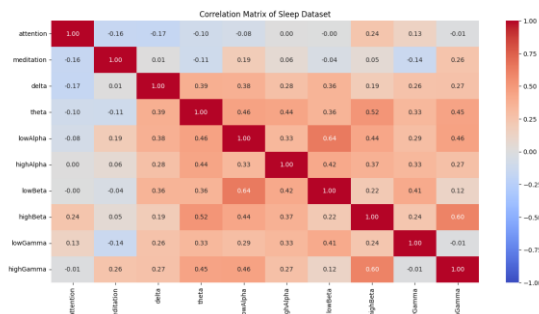


Weak Correlation (0.0 - 0.3):
 attention - meditation: 0.02
 attention - delta: 0.18
 attention - theta: 0.14
 attention - lowAlpha: 0.07
 attention - highAlpha: 0.02
 attention - lowBeta: 0.01
 attention - highBeta: 0.25
 attention - lowGamma: 0.14
 attention - highGamma: 0.02
 meditation - attention: 0.02
 meditation - delta: 0.07
 meditation - theta: 0.14
 meditation - lowAlpha: 0.19
 meditation - highAlpha: 0.08
 meditation - lowBeta: 0.05
 meditation - highBeta: 0.05
 meditation - lowGamma: 0.10
 meditation - highGamma: 0.18

הקורלציות החזקות ביותר נמצאו בין רצועות גלי המוח השונים לבין עצמם, מה שמעיד על תלות הדדית בין תחומי התדרים השונים (לדוגמא, קורלציות גבוהות בין lowBeta ל-highBeta). אולם, קורלציות אלו אינן רלוונטיות לשאלת המחקר שלנו.

קורלציות חלשות נצפו בין רמות attention ו- meditation למרבית תחומי הגלים המוחיים, המצביעים על קשר ישיר מינימלי בין משתנים אלה לפעילות גלי מוח.

אנליזה על מערכות הנתונים המופרדים - מצב שינה:

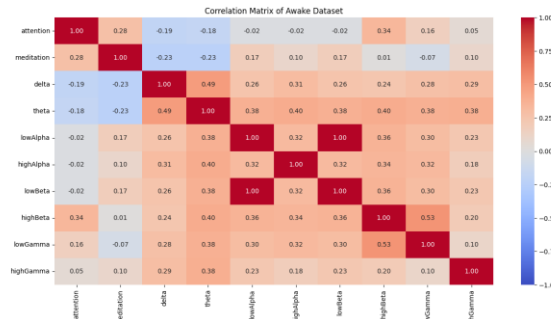


Weak Correlation (0.0 - 0.3):
 attention - meditation: 0.16
 attention - delta: 0.17
 attention - theta: 0.10
 attention - lowAlpha: 0.08
 attention - highAlpha: 0.00
 attention - lowBeta: 0.00
 attention - highBeta: 0.24
 attention - lowGamma: 0.13
 attention - highGamma: 0.01
 meditation - attention: 0.16
 meditation - delta: 0.01
 meditation - theta: 0.11
 meditation - lowAlpha: 0.19
 meditation - highAlpha: 0.06
 meditation - lowBeta: 0.04
 meditation - highBeta: 0.05
 meditation - lowGamma: 0.14
 meditation - highGamma: 0.26

קורלציות חלשות נצפו בין רמות attention ו meditation למרבית תחומי הגלים המוחיים, המצביעים על קשר ישיר מינימלי בין משתנים אלה לפעילות גלי מוח.

אנליזה על מערכות הנתונים המופרדים - מצב ערות:

Moderate Correlation (0.3 - 0.7):
attention - highBeta: 0.34



Weak Correlation (0.0 - 0.3):
attention - meditation: 0.28
attention - delta: 0.19
attention - theta: 0.18
attention - lowAlpha: 0.02
attention - highAlpha: 0.02
attention - lowBeta: 0.02
attention - highBeta: 0.34
attention - lowGamma: 0.16
attention - highGamma: 0.05
meditation - attention: 0.28
meditation - delta: 0.23
meditation - theta: 0.23
meditation - lowAlpha: 0.17
meditation - highAlpha: 0.10
meditation - lowBeta: 0.17
meditation - highBeta: 0.01
meditation - lowGamma: 0.07
meditation - highGamma: 0.10

קורלציות חלשות נצפו בין רמות attention ו meditation למרבית תחומי הגלים המוחיים, המצביעים על קשר ישיר מינימלי בין משתנים אלה לפעילות גלי מוח. עם זאת, יש לציין כי באנליזה זו נמצאה הקורלציה המשמעותית ביותר שזוהתה עד כה – קורלציה בין רמת ה-attention ל-highBeta (0.34~). אולם, יש להדגיש שערך זה רק חצה במעט את הסף המגדיר קורלציה בינונית (0.3–0.7), ולכן קשה להתייחס אליו כבעל משמעות רבה.

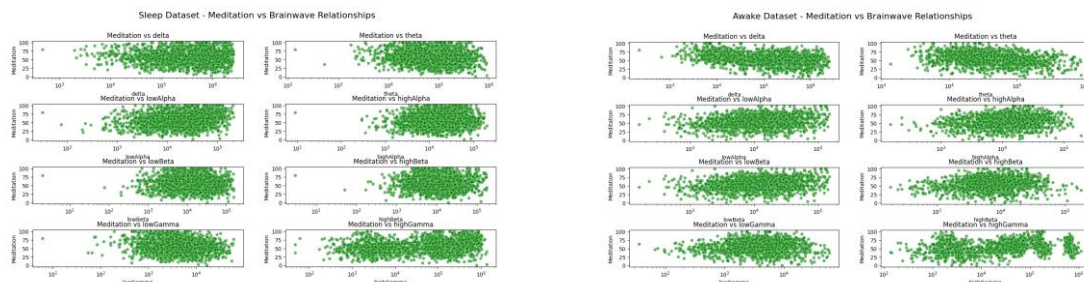
הרציונל לביצוע אנליזה גם על הדאטה הנפרד, נבעה בעיקר מהחשש מערבוב אותות. כאשר הנתונים משני המצבים מעורבבים, ייתכן שחלק מהדפוסים יוסתרו או יהיו פחות ברורים. לחלופין, עשויה להיווצר קורלציה מדומה שתציג קשר מובהק לכאורה שאינו נכון. ניתוח הנתונים בנפרד מבטיח שהתוצאות ממצב אחד לא יטשטשו או ייצרו מסקנות שגויות לגבי המצב השני.

בנוסף, ניתוח נפרד יאפשר גילוי והדגשה של קשרים ייחודיים. על ידי בדיקת מערכי הנתונים של ערות ושינה בנפרד, יכולנו לראות כיצד גלי המוח קשורים לרמות attention או meditation בכל מצב. הבדלים אלו מדגישים כיצד המוח פועל בצורה שונה במצב ערות לעומת מצב שינה.

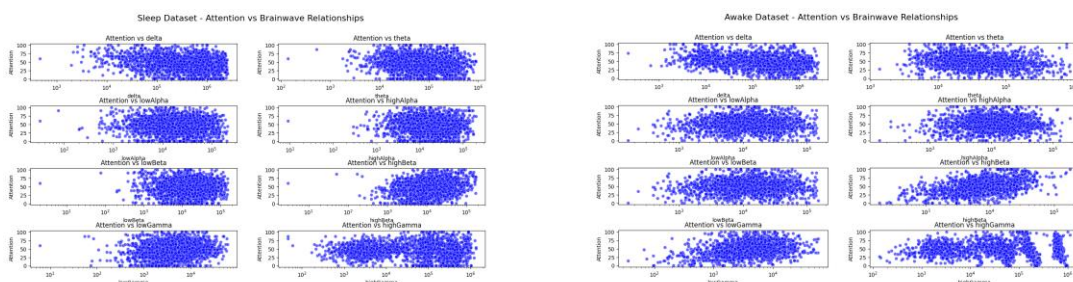
באמצעות שילוב ניתוחים אלה, גם המלא וגם המופרד, ניסינו שהממצאים שלנו היו מצד אחד רחבים ומצד שני מפורטים, תוך פירוט מגמות כלליות והתנהגויות ספציפיות מכל מצב.

לסיכום, נצפתה קורלציה לינארית חלשה בין רמות attention או meditation למרבית מאפייני הגלים המוחיים. ממצא זה מצביע על קשר ישיר מינימלי בין משתנים אלה לפעילות גלי מוח בערכות הנתונים שנבדקו. כתוצאה מכך, לא ניתן להסיק באופן חד-משמעי כי ערך מסוים של גל מוח מהווה בסיס מהימן לקביעת רמת attention או meditation של המכשיר.

2. מגמות (Trends)



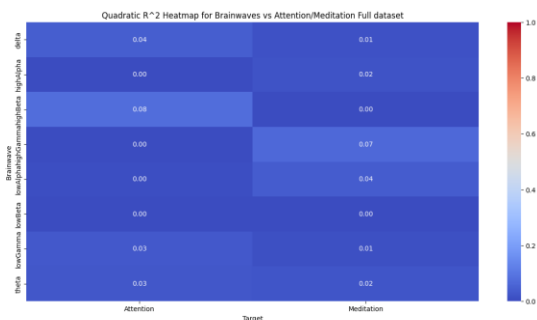
Meditation: נצפו מגמות חלשות, אם בכלל. מה שכן נצפה הוא שינוי בטווחי מאפיינים. עבור מרבית גלי המוח, הטווחים במהלך מצבי ערות היו צרים יותר (או מדוללים בקצוות) בהשוואה למערכת הנתונים של מצב שינה.



Attention: נצפו מגמות חלשות, אם בכלל. במצב ערות כן נצפה אשכול (cluster) ב-High Gamma. מה שכן נצפה הוא שינוי בטווחי מאפיינים. עבור מרבית גלי המוח, הטווחים במהלך מצבי ערות היו צרים יותר (או מדוללים בקצוות) בהשוואה למערכת הנתונים של מצב שינה.

3. קורלציה לא-ליניארית

אנליזה של מערכת נתונים מלאה:



LOW:

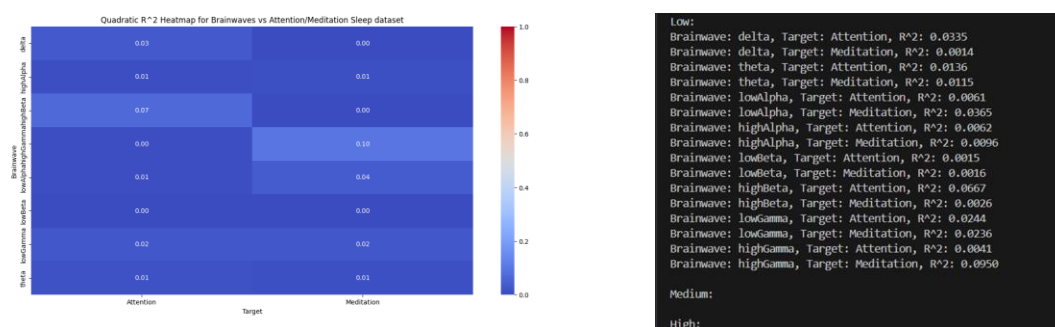
- Brainwave: delta, Target: Attention, R^2 : 0.0430
- Brainwave: delta, Target: Meditation, R^2 : 0.0134
- Brainwave: theta, Target: Attention, R^2 : 0.0256
- Brainwave: theta, Target: Meditation, R^2 : 0.0223
- Brainwave: lowAlpha, Target: Attention, R^2 : 0.0045
- Brainwave: lowAlpha, Target: Meditation, R^2 : 0.0369
- Brainwave: highAlpha, Target: Attention, R^2 : 0.0022
- Brainwave: highAlpha, Target: Meditation, R^2 : 0.0181
- Brainwave: lowBeta, Target: Attention, R^2 : 0.0001
- Brainwave: lowBeta, Target: Meditation, R^2 : 0.0030
- Brainwave: highBeta, Target: Attention, R^2 : 0.0769
- Brainwave: highBeta, Target: Meditation, R^2 : 0.0035
- Brainwave: lowGamma, Target: Attention, R^2 : 0.0304
- Brainwave: lowGamma, Target: Meditation, R^2 : 0.0110
- Brainwave: highGamma, Target: Attention, R^2 : 0.0016
- Brainwave: highGamma, Target: Meditation, R^2 : 0.0655

Medium:

High:

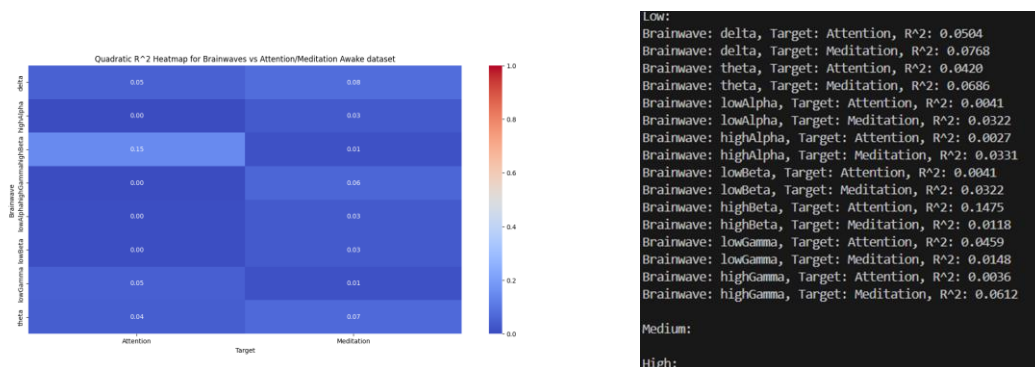
ערכי R^2 נמוכים מאוד. ערכיים אלו הצביעו על קשרים ריבועיים חלשים בין מרבית מאפייני הגלים המוחיים לרמות attention או meditation.

אנליזה על מערכות הנתונים המופרדים - מצב שינה:



ערכי R^2 נמוכים מאוד. ערכיים אלו הצביעו על קשרים ריבועיים חלשים בין מרבית מאפייני הגלים המוחיים לרמות attention או meditation.

אנליזה על מערכות הנתונים המופרדים - מצב ערות:



ערכי R^2 נמוכים מאוד. ערכיים אלו הצביעו על קשרים ריבועיים חלשים בין מרבית מאפייני הגלים המוחיים לרמות attention או meditation.

בהתאם לשיקולים שפורטו קודם, בוצע תחילה ניתוח על הדאטה המלא ולאחר מכן על הדאטה המופרד.

לסיכום, ערך R^2 , או מקדם ההתאמה, מודד את יכולת המודל (הריבועי) להסביר את השונות במשתנה היעד (כגון attention או meditation) על בסיס מאפייני הקלט (כגון נתוני גלי מוח). ערך זה משקף עד כמה ניתן לחזות או להסביר את השינויים במשתנה היעד באמצעות המודל.

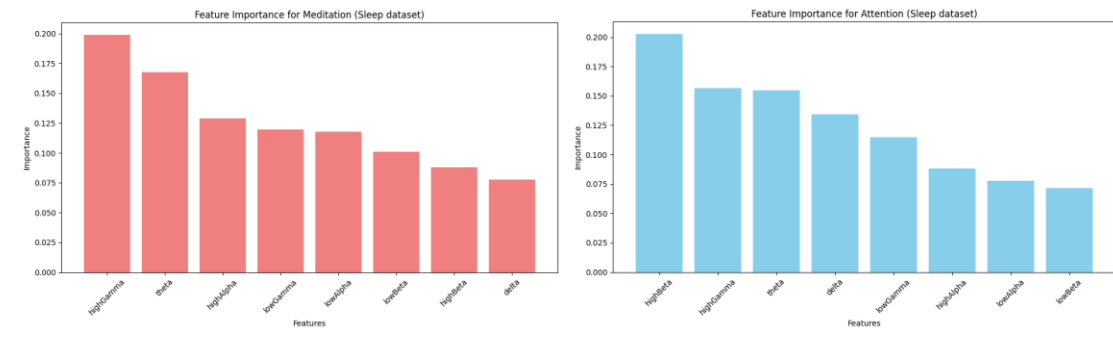
ערך R^2 משמש להערכת עוצמת הקשרים הריבועיים (לא-ליניאריים) בין מאפייני גלי מוח למשתני היעד. באופן ספציפי, הוא בוחן את עוצמת הקשר הלא-ליניארי בין פעילות גלי מוח למצבים קוגניטיביים כמו attention ו-meditation.

ערכי R^2 נמוכים מעידים על קשרים חלשים, ומראים שמאפייני גלי מוח מסוימים אינם מנבאים באופן משמעותי את משתנה היעד תחת המודל הריבועי, מה שמרמז על מגבלות המודל במקרים אלה.

4. Random Forest Regression Analysis

אנליזה על מערכות הנתונים המופרדים - מצב שינה:

R^2 score for Attention: 0.3266
 R^2 score for Meditation: 0.3407



בגרף עבור attention, ל-highBeta יש את החשיבות הגבוהה ביותר (~ 0.2), ולכן הוא התכונה המשפיעה ביותר על חיזוי attention. בגרף עבור meditation, ל-highGamma יש את החשיבות הגבוהה ביותר (~ 0.2), מה שהופך אותו לקריטי לחיזוי meditation.

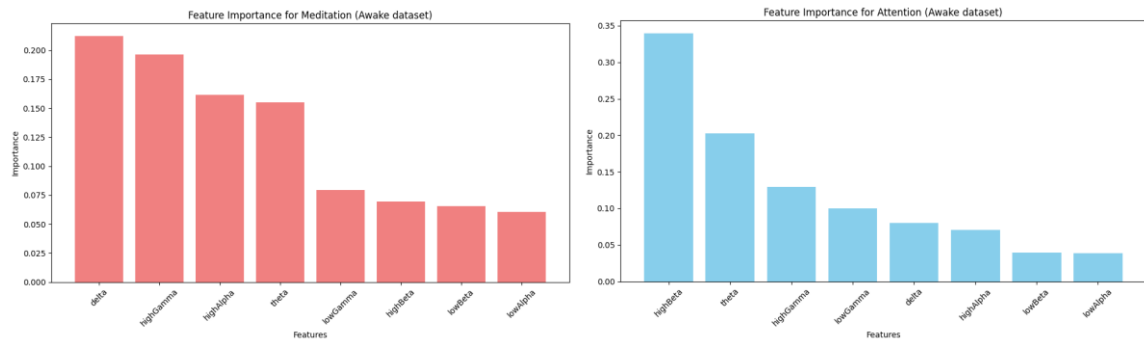
תכונה עם ציון חשיבות של 0.2 תורמת 20% מכוח קבלת ההחלטות של המודל, וזה משמעותי מכיוון שבעבור 8 תכונות, חלוקה שווה הייתה מעניקה לכל תכונה חשיבות ממוצעת של $\frac{1}{8} = 0.125$. מאחר ש-0.2 גדול מ-0.125, זה משקף תכונה בעלת השפעה גדולה במיוחד על תחזיות המודל.

ציוני ה- R^2 שמצאנו הם 0.3266 עבור attention ו-0.3407 עבור meditation. כלומר, 32.66% מהשונויות משמעו שהמודל הצליח להסביר כשליש מן השינויים וההבדלים בערכי היעד attention בהתבסס על נתוני גלי המוח. זהו המדד לכמה טוב המודל מצליח "להתאים" את התחזיות שלו לנתונים האמיתיים. עבור meditation, המודל מסביר כ-34.07% מהשונויות, כלומר, המודל מצליח לחזות חלק מתנודות ה meditation, אך עדיין נותר מקום להסבר נוסף.

ציונים אלו מעידים על יכולת ניבוי בינונית, מה שמעיד על כך שייתכן שיש גורמים נוספים המשפיעים על attention ו-meditation או שאולי יש צורך בשיפור המודל או הנתונים.

אנליזה על מערכות הנתונים המופרדים - מצב ערות:

R^2 score for Attention: 0.4282
 R^2 score for Meditation: 0.4388



בגרף עבור attention, ל-highBeta יש את החשיבות הגבוהה ביותר (~ 0.35), ולכן הוא התכונה המשפיעה ביותר על חיזוי attention. בגרף meditation, ל-highDelta ו-highGamma יש את החשיבות הגבוהה ביותר, עם ציוני חשיבות של ~ 0.21 ו- ~ 0.2 בהתאמה, מה שהופך אותם לחשובים עבור חיזוי meditation.

ציון חשיבות של 0.35 או אפילו 0.2 תורמת באופן משמעותי לכוח קבלת ההחלטות של המודל (< 0.125 כאמור לעיל).

ציוני ה- R^2 שמצאנו עבור סט הנתונים של ערות הם 0.4282 עבור attention ו-0.4388 עבור meditation. ציונים אלו מצביעים על כך שהמודל מסביר כ-42.82% מהשונות ב attention וכ-43.88% מהשונות ב meditation. בהשוואה לסט הנתונים של שינה, שבו ציוני ה- R^2 היו נמוכים יותר ($\sim 32.66\%$ ו- $\sim 34.07\%$), המודל מציג שיפור בביצועים החיזויים עבור סט הערות. עם זאת, חלק ניכר מהשונות עדיין לא מוסבר, דבר המצביע על פוטנציאל לשיפור המודל או הוספת תכונות רלוונטיות נוספות.

סיכום:

לאור התוצאות הנ"ל לא מצאנו קורלציה חזקה בין גל מוח מסוים שנמדד לבין קביעת ה EEG על מצב של attention או meditation של הנבדק, הן מבחינת קשר לינארי, הן מבחינת מגמה והן מבחינת קשר לא לינארי (מודל ריבועי). במודל הלינארי הרגיל, הקורלציה המשמעותית ביותר שנמצאה היתה בין attention ל-highBeta במצב ערות. ממצא זה עקבי עם תוצאות המודל הבוחן קשרים מורכבים בין אורכי הגל השונים, שם ראינו שבמצב שינה וערות highBeta השפיע הכי הרבה על קביעת attention. לצד זה, על קביעת meditation במצב שינה הגל highGamma השפיע ביותר ובמצב ערות היו אלו גם highGamma וגם highDelta. עם זאת, המודל הציג יכולת בינונית בלבד בחיזוי הקביעה של attention ו-meditation ולכן סביר שיש גורמים נוספים המשפיעים על הקביעה של המכשיר את רמת ה-attention ו-meditation.