

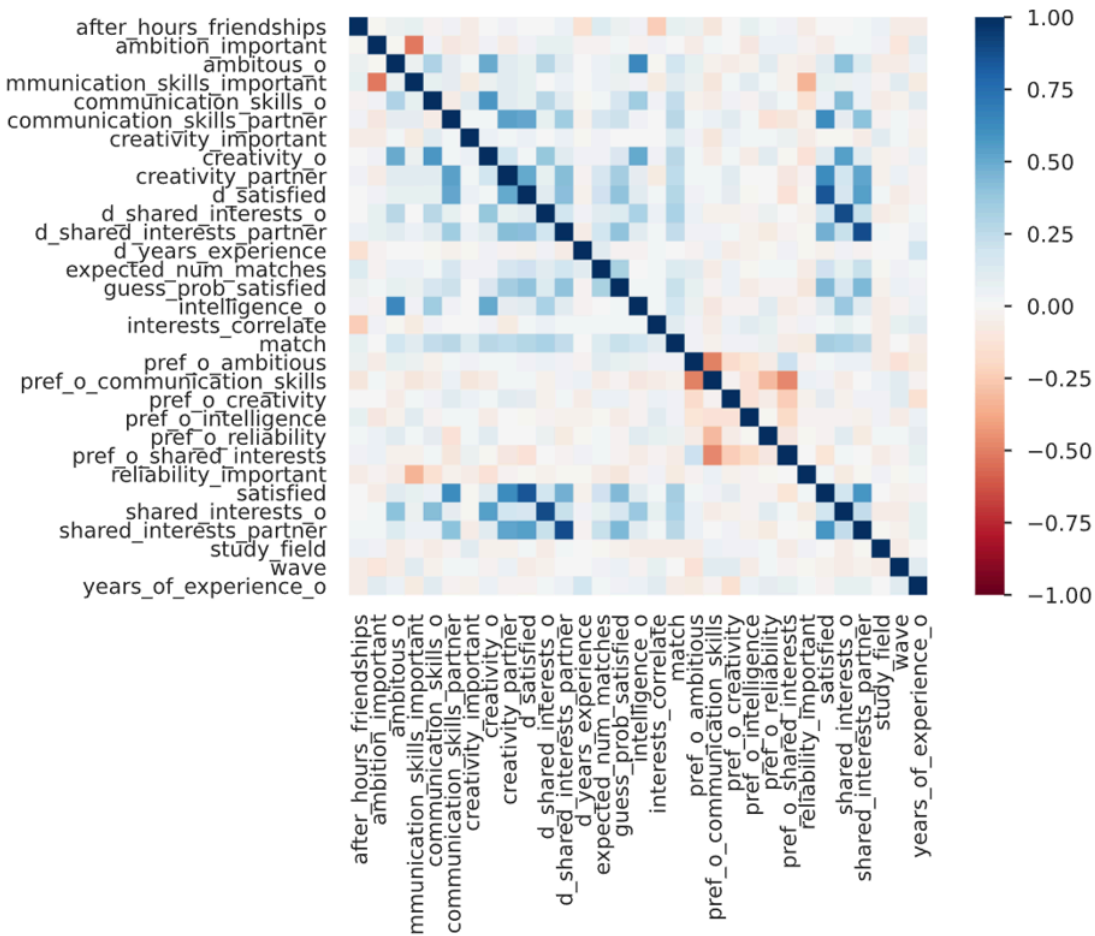
Conclusions & Suggestions

יאיר מחפוד 207807082, יונתן קרטור 213920432
עמית צור 214926214, ירין אוהיון 214169336

1. מסקנות

1.1. סעיף ראשון - פרדיקציה של התאמות בין משתתפים

בסעיף זה הגשנו את הפתרון הטוב ביותר מבין האפשרויות שמצוינות ב-project.pdf, והוא שימוש ברשתות נוירונים. בדרך לפתרון זה התחלנו עם feature evaluation וזאת עשינו באמצעות כמה דרכים אחת הדרכים היא לחשב את ה importance של כל אחד מה features באמצעות random forest בשביל לקבל רעיון של אילו features יכולים לעזור לנו לחזות את match, בעזרת זה השתמשנו בספריית pandas profiling על ה features החשובים ביותר שחישבנו בנוסף ל match, pandas profiling, heatmap של קורלציה בין ה features:

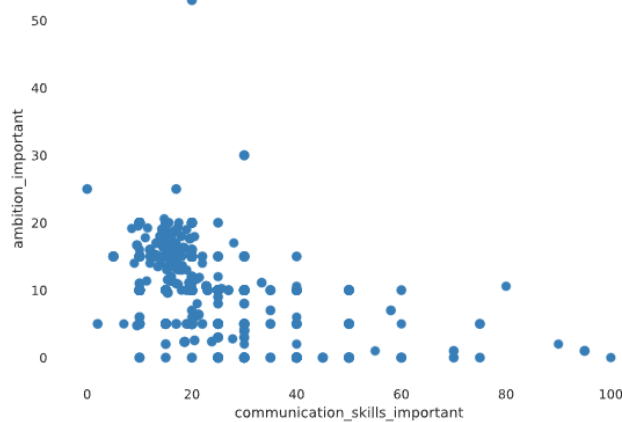


ניתן לראות על פי הגרף כי יש קורלציה חזקה בין האם יש התאמה בין שני משתתפים (match) לבין:

d_satisfied, communication_skills_partners, satisfied, shared_interests_o, d_shared_interests

מזה ניתן להסיק שרוב הדברים החשובים ביותר לקביעת *match* הם דברים יותר עמוקים ולא שטחיים כמו גזע, כמה ימים עובדים מהבית וכו', בנוסף חלק מה *features* החשובים נראה שהם דברים שנגרמו מכך שהיה התאמה לדוגמא אם הייתה התאמה אז ככל הנראה המשתתף יהיה מסופק מהמשתתף השני ויחשוב שה *communication_skills* שלו טובים.

דבר מעניין נוסף הוא קורלציה שלילית בין *communication_skills_important* ו-*ambition_important*:



כלומר בדרך כלל, כאשר האחד גבוה - השני נמוך. זה מידע שלא צפינו לו, ומראה עד כמה חשוב להשתמש במידע זה כדי לשפר את המודלים, שלא פועלים מתוך אינטואיציה.

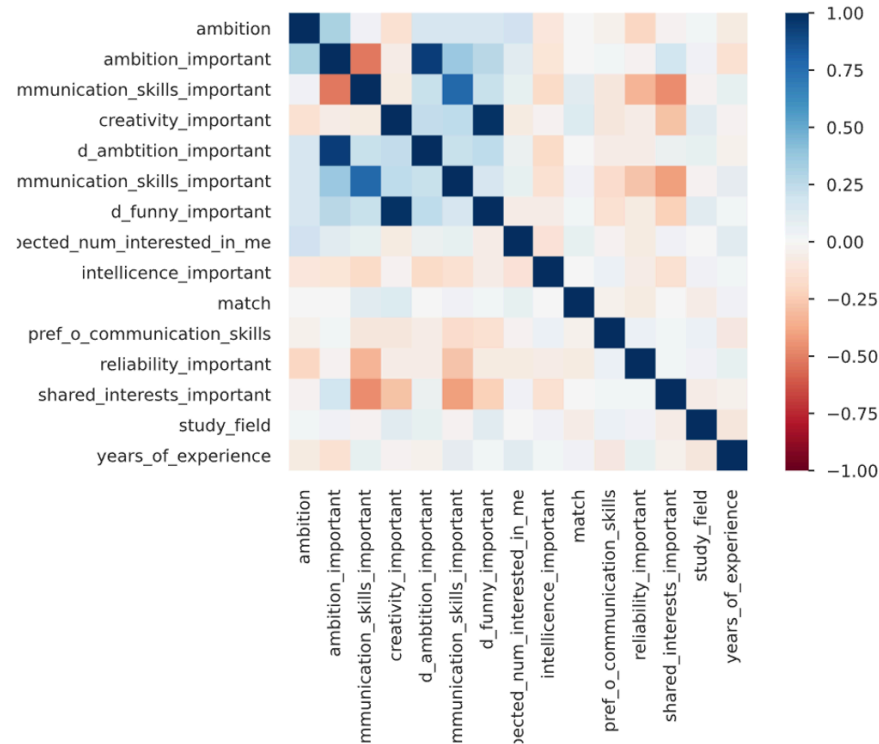
נראה עוד מידע מעניין בהקשר של *ambition_important* בסעיף הבא.

1.2. סעיף שני - פרדיקציה של העדפות משתתפים

בסעיף זה הגשנו את הפיתרון המשתמש בקונספט של Multioutput Regression, תוך שימוש בספריה שלמדנו אותה לראשונה עבור תרגיל זה - *xgboost*. השגנו פרדיקציה של שני *Labels*, האחד *ambition_important* והשני *creativity_important*.

אנו מצרפים למטה תמונה של heatmap של *correlations* של איחוד ה-*features* המתאימים ביותר עבור *ambition_important* ו-*creativity_important*:

5.7.24



ניתן לראות כי אין כמעט קורלציה בין `ambition_important` ו-`creativity_important`, ולכן יתכן שיכולנו להריץ מודל עבור כל אחד מהם בנפרד ולקבל תוצאות דומות למה שהגשנו.

מידע שהתווסף לעומת heatmap הקודם הוא כי יש קורלציה ישירה בין `creativity_important` לבין `d_funny_important`. כלומר השפעה על האחד, משפיעה באותו אופן על השני. בן אדם מצחיק בדרך כלל גם יצירתי, או לפחות עושה רושם כזה.

2. הצעות

נציע בהתאם למסקנות דרכי פעולה שונות שניתן להמליץ למשתתפים שנכחים באירועים מסוגים אלו, כדי להעלות הסיכוי ל-`match`:

- 2.1 להרחיב את תחומי העניין שלך כדי להעלות את הסיכויי שיהיה לך תחומי עניין משותפים עם משתתפים אחרים ולשוחח עם המשתתפים האחרים על תחומי העניין שלהם.
- 2.2 לשפר את יכולות הדיבור שלך, תתאמן על לדבר עם אנשים בביטחון ודיבור מול קהל כדי שאנשים אחרים יחשבו שה-`communication skills` שלך טובים יותר ויש סיכוי טוב יותר שהם יחשבו שאתה מתאים והסיכוי ל-`match` יהיה גבוהה יותר.
- 2.3 לפני ההגעה, לבדוק על כמה שיותר מידע שיכול להתקשר מנושאים שונים לנושא שלך - הדבר יגרום לאדם שאתה נפגש איתו להזדהות יותר עם נושא המחקר / ממנו אתה מגיע, ולגרום ל-`match` להיות יותר סביר. חשוב למצוא הקשרים אלו, שהרי הדבר יגרום לדיבור טוב יותר.
- 2.4 אל תפחד לדבר על הנושא שלך, כל עוד אתה מפגין שאפתנות כלפי נושא השיחה, הדבר יראה בביטחון ויגרום שיפור בסיכוי ל-`match` בהקשר לנושא המפגש.