

# Generative Modelling and Clustering

הערה: כאשר אנו מתייחסים לציון הומוגניות אנו מתכוונים לזה המתקבל ע"י `sklearn.metrics.homogeneity_score`

## בחירת קואליציה יציבה

### Clustering

על מנת למצוא קואליציה עבור מפלגות בעלות מצביעים דומים, השתמשנו ב-GaussianMixture כדי לחלק את הדאטה לקלאסטרים. כיוונו את פרמטר מספר הקלאסטרים, החל מחצי מספר המפלגות ועד לפי שניים ממספר המפלגות, על מנת לאפשר גמישות מירבית, את טיב התוצאות מדדנו באמצעות כרוס-וולידישן עבור:

- Calinski Harabaz score
- Silhouette score

השתמשנו בציונים אלו על מנת לקבל קלאסטרים שהם יחסית צפופים ורחוקים אחד מהשני.

```
-----  
Number of components: 9  
Calinski Harabaz score: 556.8030194040691  
Silhouette score: 0.24622246612392878  
-----
```

על פי התוצאות בחרנו לבחון את הקואליציות עבור 9 קלאסטרים (בפועל בדקנו עבור מספרים שונים כדי לחזק את ההערכות שלנו).

יצרנו היסטוגרמת הצבעות עבור כל קלאסטר:

```
-----  
{'Purples': 348, 'Whites': 50, 'Greens': 151, 'Browns': 259, 'Pinks': 68}  
Cluster 0  
Size = 876  
Purples = 40 %  
Whites = 6 %  
Greens = 17 %  
Browns = 30 %  
Pinks = 8 %  
-----  
{'Blues': 42, 'Turquoises': 15, 'Yellows': 23}  
Cluster 1  
Size = 80  
Blues = 52 %  
Turquoises = 19 %  
Yellows = 29 %  
-----  
{'Purples': 367, 'Greens': 142, 'Browns': 218, 'Whites': 44, 'Pinks': 55}  
Cluster 2  
Size = 826  
Purples = 44 %  
Greens = 17 %  
Browns = 26 %  
Whites = 5 %  
Pinks = 7 %  
-----  
{'Greens': 298, 'Purples': 48, 'Whites': 6, 'Pinks': 4, 'Yellows': 9, 'Browns': 20, 'Turquoises': 6, 'Blues': 3}  
Cluster 3  
Size = 394  
Greens = 76 %  
Purples = 12 %  
Whites = 2 %  
Pinks = 1 %  
Yellows = 2 %  
Browns = 5 %  
Turquoises = 2 %  
Blues = 1 %  
-----
```

```

-----
{'Purples': 329, 'Yellows': 83, 'Turquoises': 49, 'Browns': 214, 'Greens': 152, 'Pinks': 117, 'Whites': 45, 'Blues': 131}
Cluster 4
Size = 1120
Purples = 29 %
Yellows = 7 %
Turquoises = 4 %
Browns = 19 %
Greens = 14 %
Pinks = 10 %
Whites = 4 %
Blues = 12 %
-----
{'Pinks': 287}
Cluster 5
Size = 287
Pinks = 100 %
-----
{'Greys': 438, 'Oranges': 341, 'Reds': 222}
Cluster 6
Size = 1001
Greys = 44 %
Oranges = 34 %
Reds = 22 %
-----
{'Pinks': 115, 'Browns': 491, 'Whites': 81, 'Purples': 693, 'Greens': 5}
Cluster 7
Size = 1385
Pinks = 8 %
Browns = 35 %
Whites = 6 %
Purples = 50 %
Greens = 0 %
-----
{'Yellows': 292, 'Blues': 526, 'Turquoises': 213}
Cluster 8
Size = 1031
Yellows = 28 %
Blues = 51 %
Turquoises = 21 %

```

ומטריצת מרחקים בין תוחלות הקלאסטרים:

```

[[[0.          0.8321478  0.65545816 0.65103621 0.67135404 0.87992144 0.61209788 0.569406  0.79421354]
 [0.8321478  0.          0.84028005 0.83515437 0.79217697 1.          0.74111336 0.77397516 0.21811525]
 [0.65545816 0.84028005 0.          0.64836549 0.68674044 0.88259413 0.6327204  0.57861083 0.78154874]
 [0.65103621 0.83515437 0.64836549 0.          0.67232575 0.86355508 0.69157294 0.41841755 0.79639845]
 [0.67135404 0.79217697 0.68674044 0.67232575 0.          0.89288147 0.58263892 0.59529232 0.75694346]
 [0.87992144 1.          0.88259413 0.86355508 0.89288147 0.          0.84486545 0.81162905 0.99921578]
 [0.61209788 0.74111336 0.6327204  0.69157294 0.58263892 0.84486545 0.          0.49681975 0.70035623]
 [0.569406  0.77397516 0.57861083 0.41841755 0.59529232 0.81162905 0.49681975 0.          0.73357406]
 [0.79421354 0.21811525 0.78154874 0.79639845 0.75694346 0.99921578 0.70035623 0.73357406 0.          ]]]

```

ביצענו חלוקה של המפלגות לקואליציה ואופוזיציה, עבור הרכבים נבחרים, ובדקנו את ציון ההמוגוניות של הקלאסטרים עבורם:

```

0.8479539497074186 ['Purples', 'Greens', 'Browns', 'Whites', 'Pinks']
0.6577488628222546 ['Purples', 'Greens', 'Browns', 'Whites']
0.6928930540531851 ['Purples', 'Greens', 'Browns', 'Pinks']
0.5744193121893825 ['Purples', 'Greens', 'Browns']

```

(ציון, הרכב הקואליציה)

בחנו את התוצאות והגענו למסקנה כי הקואליציה: Purples, Greens, Browns, Whites היא העדיפה ביותר.

לקואליציה זו צפוי 56.6 אחוזים מהקולות והיא מתפרסת על פני מספר קלאסטרים יחסית קרובים מה שמצביע על כך שהמצביעים למפלגות אלו יחסית דומים. כמו כן מפלגות האופוזיציה אינן חולקות בעיקרן את הקלסטרים עם הקואליציה ומרחק הקלסטרים בהן הן נמצאות רחוק יחסית מאלו של הקואליציה. ביצענו בדיקה על סט הטסט וראינו כי ההתפלגות המפלגות לקלסטרים נשמרת.

הערות:

(Purples, Greens, Browns) מהוות קואליציה, לא שונים מהאופוזיציה כי מצביעי ה-Whites דומים להם. לכן לא בחרנו בקואליציה זו)

(התלבטנו אם להוסיף את Pink לחיזוק הקואליציה אך לבסוף החלטנו שלא מכיוון שלמרות שמחצית מהמצביעים למפלגה זו הינם בעלי תכונות דומות לקואליציה לעיל, החצי האחר נמצא בקלסטר נפרד המרוחק משאר הקלאסטרים, מה שמעיד על השונות של מצביעים אלו)

## תכונות מובילות לכל מפלגה ומנצח חדש

את משימה זו ביצענו באופן הבא:

- אימנו מסווג גאוסיאן נאיב בייס על קבוצת האימון
  - ביצענו שינויים בערכי תכונות בקבוצת המבחן
  - בחנו כיצד השינויים משפיעים על מספר המצביעים עבור כל מפלגה
- השתמשנו במודל גנרטיבי כיוון שבהינתן דגימה חדשה המודל יסווג תוך התחשבות בהתפלגות הנראית של המרחב המקורי ובאופן זה אנו נראה כיצד השינויים משפיעים על כמה מפלגות.

לצורך כך חילקנו את התכונות ל-3 קבוצות:

- תכונות המתפלגות בין 1- ל-1 (נכנה אותן מתוחמות).
- תכונות המתפלגות בין מינוס אינסוף לאינסוף עם תוחלת אפס (נכנה אותן נורמליות).
- תכונות המקבלות או את הערך 1 או את הערך 1- (נכנה אותן בינאריות).

מניפולציה על תכונה יכולה להיות העלאת או הורדת ערכה לפי סוגה:

- עבור תכונות מתוחמות – רצינו להשאר בתחום הקבוע לכן עבור מצביע מסוים ביצענו ממוצע של ערך התכונה עבורו ו-1 (או עם 1- עבור הורדת ערך תכונה).
- עבור תכונות נורמליות – הוספנו או חסרנו 0.5 בהתאמה עבור העלאה והורדה מערך התכונה עבור הבוחר.
- עבור בינאריות – על מנת להעלות תכונה בינארית נרצה לשים בה ערך אחד וערך מינוס אחד בכל התכונות שלא ייתכן שיחזיקו גם בערך 1 (כיוון שבפועל הן מייצגות את אותה התכונה – למשל, את התכונה MostImportantIssue פיצלנו למספר תכונות שונות שכל אחת מייצגת תשובה אפשרית, כיוון שרק תשובה אחת אפשרית לא ייתכן שלשתי תכונות יהיה ערך 1). על מנת להוריד ערך תכונה נשים 1- עבור המצביע ונגריל תכונה אחרת מהתכונות האפשריות שתקבל ערך 1. עבור העלאה והורדה בהתאמה ביצענו פעולות אלו עבור כל מצביע בהסתברות 0.8, כלומר, בהנתן שנרצה להעלות נבצע בהסתברות 0.8 את הפעולה הראשונה ובהנתן שנרצה להוריד נבצע בהסתברות 0.8 את הפעולה השנייה.

Number\_of\_valued\_Kneset\_members  
Were scaled DOWN and the new winner prediction is Browns

Overall\_happiness\_score  
Were scaled DOWN and the new winner prediction is Browns

## תכונות מובילות

Blues	
Avg_Satisfaction_with_previous_vote	( down , voters changed: -47 , % changed: -31.33% )
Weighted_education_rank	( down , voters changed: -42 , % changed: -28.0% )
Overall_happiness_score	( up , voters changed: -40 , % changed: -26.67% )
Will_vote_only_large_party	( down , voters changed: -38 , % changed: -25.33% )
Browns	
Number_of_valued_Kneset_members	( up , voters changed: -155 , % changed: -60.08% )
Number_of_valued_Kneset_members	( down , voters changed: 135 , % changed: 52.33% )
Overall_happiness_score	( down , voters changed: 88 , % changed: 34.11% )
Overall_happiness_score	( up , voters changed: -67 , % changed: -25.97% )
Greens	
Overall_happiness_score	( down , voters changed: -34 , % changed: -21.25% )
Weighted_education_rank	( down , voters changed: -32 , % changed: -20.0% )
Greys	
Overall_happiness_score	( up , voters changed: -87 , % changed: -92.55% )
Weighted_education_rank	( up , voters changed: -66 , % changed: -70.21% )
Yearly_IncomeK	( down , voters changed: -66 , % changed: -70.21% )
Number_of_valued_Kneset_members	( up , voters changed: -65 , % changed: -69.15% )
Garden_sqr_meter_per_person_in_residency_area	( down , voters changed: -59 , % changed: -62.77% )
Oranges	
Weighted_education_rank	( up , voters changed: 65 , % changed: 89.04% )
Overall_happiness_score	( up , voters changed: 49 , % changed: 67.12% )
Military	( up , voters changed: 40 , % changed: 54.79% )
Yearly_IncomeK	( down , voters changed: 39 , % changed: 53.42% )
Garden_sqr_meter_per_person_in_residency_area	( down , voters changed: 38 , % changed: 52.05% )
Pinks	
Weighted_education_rank	( down , voters changed: -23 , % changed: -16.67% )
Will_vote_only_large_party	( up , voters changed: -18 , % changed: -13.04% )
Number_of_valued_Kneset_members	( up , voters changed: -17 , % changed: -12.32% )
Purples	
Number_of_valued_Kneset_members	( up , voters changed: 183 , % changed: 47.78% )
Number_of_valued_Kneset_members	( down , voters changed: -151 , % changed: -39.43% )
Overall_happiness_score	( down , voters changed: -57 , % changed: -14.88% )
Reds	
Weighted_education_rank	( down , voters changed: 48 , % changed: 102.13% )
Foreign_Affairs	( up , voters changed: 41 , % changed: 87.23% )
Number_of_valued_Kneset_members	( up , voters changed: 39 , % changed: 82.98% )
Overall_happiness_score	( up , voters changed: 38 , % changed: 80.85% )
Turquoises	
Weighted_education_rank	( up , voters changed: -11 , % changed: -18.03% )
Will_vote_only_large_party	( up , voters changed: 11 , % changed: 18.03% )
Garden_sqr_meter_per_person_in_residency_area	( up , voters changed: -7 , % changed: -11.48% )
Whites	
Overall_happiness_score	( up , voters changed: 53 , % changed: 108.16% )
Weighted_education_rank	( down , voters changed: 43 , % changed: 87.76% )
Garden_sqr_meter_per_person_in_residency_area	( up , voters changed: 43 , % changed: 87.76% )
Yearly_IncomeK	( up , voters changed: 38 , % changed: 77.55% )
Weighted_education_rank	( up , voters changed: 36 , % changed: 73.47% )
Yellows	
Will_vote_only_large_party	( up , voters changed: 54 , % changed: 62.07% )
Avg_Satisfaction_with_previous_vote	( down , voters changed: 46 , % changed: 52.87% )
Overall_happiness_score	( up , voters changed: 40 , % changed: 45.98% )
Weighted_education_rank	( down , voters changed: 35 , % changed: 40.23% )
Garden_sqr_meter_per_person_in_residency_area	( up , voters changed: 33 , % changed: 37.93% )

(השארנו את התכונות שהובילו לשינוי הכי גדול)

## חיזוק הקואליציה

לפי הנתונים שראינו כשבחנו את הקואליציה היציבה, כמחצית מהמצביעים של המפלגה 'Pinks' דומים בתכונותיהם למצביעי מפלגות הקואליציה, לכן נוכל לחזק את הקואליציה אם נצליח לגרום למצביעים אלו להצביע למפלגות הקואליציה. בנוסף לחיזוק הקואליציה מבחינת אחוז המצביעים, נצליח גם להגדיל את השונות בין הקואליציה לאופוזיציה. כלומר, הקלאסטרים של הקואליציה יכילו פחות מצביעי 'Pink' שהם באופוזיציה.

בנוסף, אנו נרצה לחזק את מפלגות הקואליציה ולכן נבחן גם תכונות מובילות שמוסיפות להן מצביעים.

לשם כך, הסתכלנו על רשימת התכונות המובילות שראינו בסעיף הקודם, ובחנו מספר צירופים באופן הבא:

1. יצרנו את הקלאסטרים ע"פ סט האימון
2. בחנו את התפלגות הקואליציה-אופוזיציה של קבוצת המבחן בקלסטרים (לקבלת רפרס עבור השינויים שנבצע)
3. ביצענו שינויים כמתואר בסעיף קודם על תכונות קבוצת המבחן
4. תייגנו את קבוצת המבחן בעזרת מסווג גאוסיאן נאיב בייס שאומן על סט האימון.
5. בחנו את התפלגות הקואליציה-אופוזיציה של קבוצת המבחן החדשה בקלסטרים (להשוואה עם הרפרנס לעיל)

הערכנו את טיב השינויים בין שלב 2 לשלב 5 על פי:

- השינוי בגודל הקואליציה
- השינוי במספר מצביעי 'Pinks'
- השינוי בציון ההומוגניות של הקלאסטרים.

חלקם היו פחות טובים מכיוון שהשינויים לא באו על חשבון האופוזיציה אלא בתוך הקואליציה בלבד, לבסוף בחרנו את:

- Number\_of\_valued\_Kneset\_members – העלאת ערך התכונה תגרום להפחתת מצביעי 'Pinks'
- Weighted\_education\_rank – הורדת ערך התכונה תגרום להפחתת מצביעי 'Pinks'
- Garden\_sqr\_meter\_per\_person\_in\_residency\_area – העלאת ערך התכונה תגרום להעלאת מצביעי 'Whites' ובכך לחיזוק הקואליציה.

שההשפעה שלהם הייתה:

לפני:

```
-----  
Homogeneity score: 0.7617366850278416  
Coalition size 850  
Pinks party size 138  
-----
```

אחרי:

```
-----  
Homogeneity score: 0.911400543065044  
Coalition size 902  
Pinks party size 87  
-----
```

כלומר, קואליציה צפויה חזקה והומוגנית יותר.

## מגישים

אביאל שמחי, 305376063

עמית סולומון, 305025785