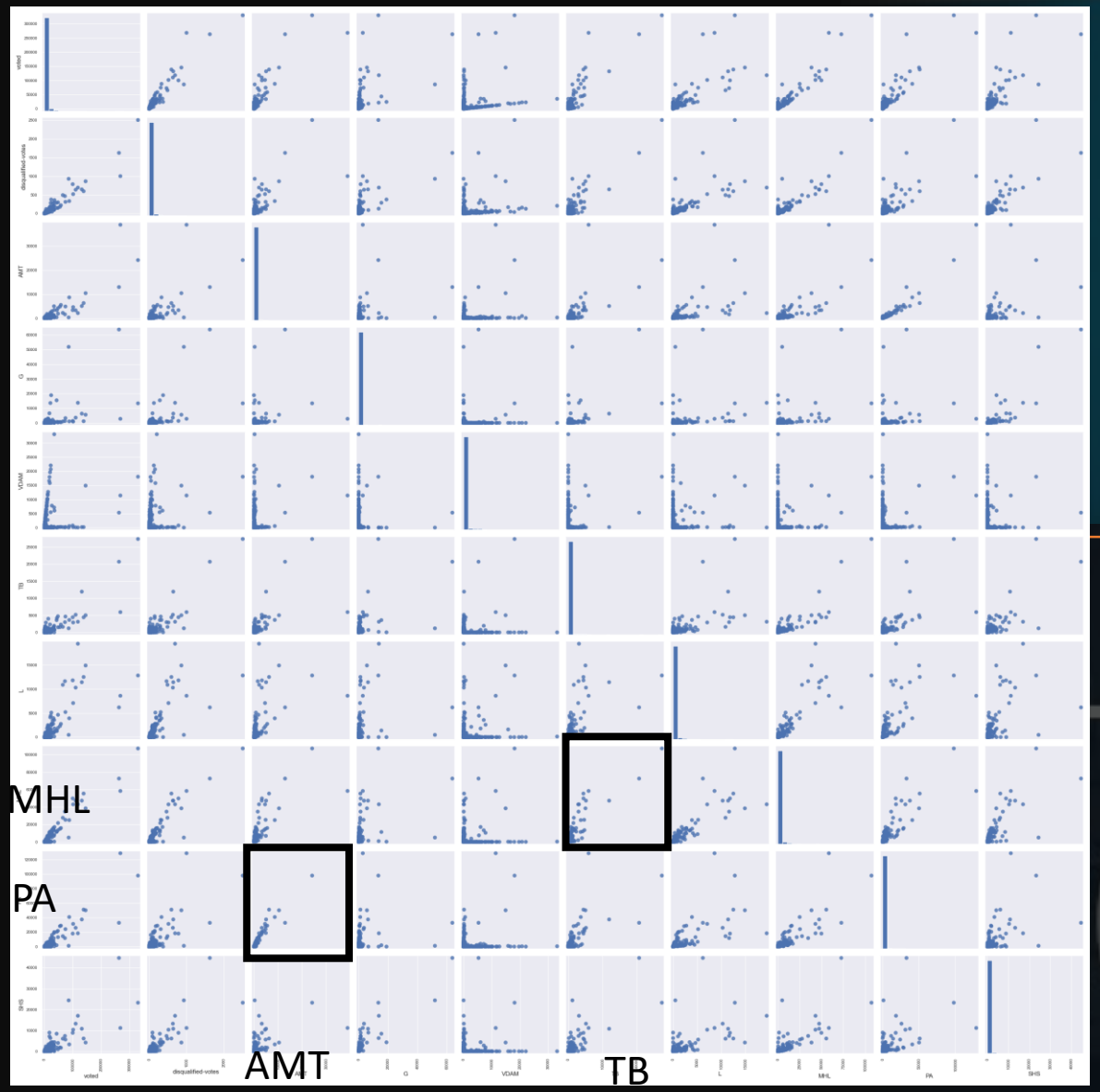
# Data Analysis Of Israel's Elections

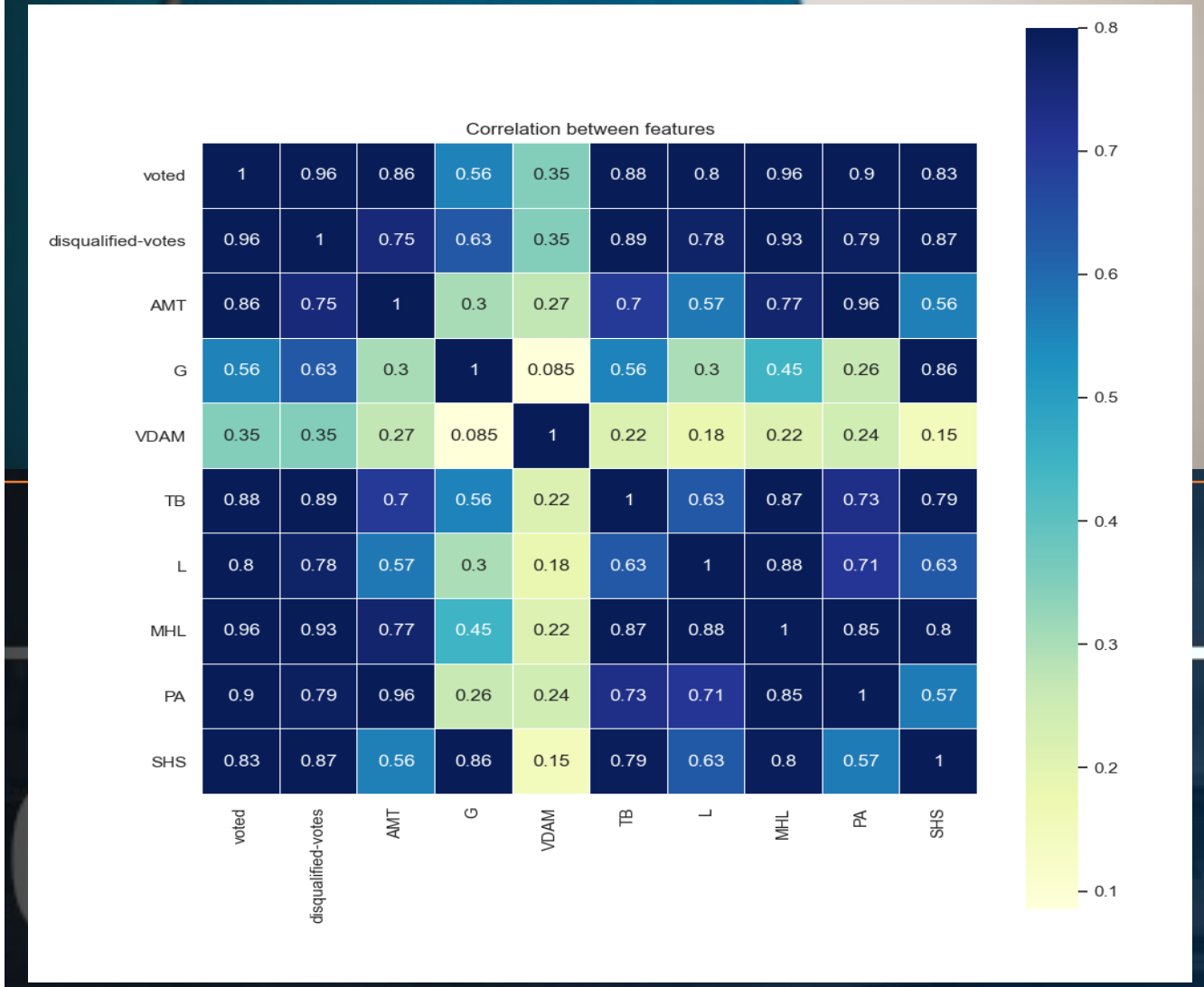
Or Yagol

Saghi Kungurov

# Data Discription:

- We found a correlation and conections between parties, voting patterns
- We used that knowledge to predict valid/disqualified votes and combined parties to political blocks (right, center-left, orthodox, Arabs, Liberman)
- We used that knowledge to decide a logical number of clusters to work with in order to get conclusions
- We Used  the connection between votes for parties to valid votes to predict valid/disqualified votes feature
- We decided to normalize the votes in a town to a percentage (votes for party, votes that were disqualified, general voting percentage) That way each town will have the same algorithmic weight
- We have used the data of the parties that got above 1% of the votes
- Each party name was defined by translation of the voting note like: Likud was translated to “MHL" (in Hebrew מחל)


MHL
PA
AMT
TB


Correlation between features
voted 1 0.96 0.86 0.56 0.35 0.88 0.8 0.96 0.9 0.83
disqualified-votes 0.96 1 0.75 0.63 0.35 0.89 0.78 0.93 0.79 0.87
AMT 0.86 0.75 1 0.3 0.27 0.7 0.57 0.77 0.96 0.56
G 0.56 0.63 0.3 1 0.085 0.56 0.3 0.45 0.26 0.86
VDAM 0.35 0.35 0.27 0.085 1 0.22 0.18 0.22 0.24 0.15
TB 0.88 0.89 0.7 0.56 0.22 1 0.63 0.87 0.73 0.79
L 0.8 0.78 0.57 0.3 0.18 0.63 1 0.88 0.71 0.63
MHL 0.96 0.93 0.77 0.45 0.22 0.87 0.88 1 0.85 0.8
PA 0.9 0.79 0.96 0.26 0.24 0.73 0.71 0.85 1 0.57
SHS 0.83 0.87 0.56 0.86 0.15 0.79 0.63 0.8 0.57 1
voted
disqualified-votes
AMT
G
VDAM
TB
L
MHL
PA
SHS
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

# Connection Between Demographic Characteristics to Voting Pattern:

- We checked the clustering algorithms: Kmeans (Chosen), GMM, Mean Shift, Hierarchical Agglomerative on the results of the 23rd Knesset

- After executing the algorithms we created an excel file of all towns and their determined cluster

- Our features were the percentage of voting for each party in a town, we also checked by block division

- We analyzed that 6 clusters are the best division and got these conclusion:
  - Cluster of Arab towns that voted VDAM
  - Cluster of Orthodox towns or peripheral towns that voted the Orthodx Parties
  - Cluster of settlements that voted mostly to TB and a bit for MHL
  - Cluster of Kibuzim and Moshavim that voted PA or AMT
  - Cluster of mostly peripheral towns that voted primarily for MHL
  - Cluster of Gush Dan area, Sharon area and additional large cities mostly in the center of Israel that voted primarily for PA


| 0 – הצבעה לרשימה המשותפת | 1 – הצבעה לכחול לבן | 2 – הצבעה לליכוד | 3 – הצבעה לימינה | 4 – הצבעה לש"ס/יהדות התורה | 5 – הצבעה ל-העבודה-גשר-מרצ |
|---|---|---|---|---|---|
| אום אל פאחם | יהוד | אופקים | אלקנה | ביתר עילית | איילת השחר |
| אכסאל | מודיעין מכבים רעות | אשקלון | יצהר | בני ברק | באר טוביה |
| טייבה | רעננה | באר שבע | סוסיה | רכסים | בנימינה גבעת עדה |
| כאבול | תל אביב יפו | דימונה | עתניאל | בית שמש | גינוסר |
| סחנין | תל מונד | בת ים | קרני שומרון | אלעד | כפר ורדים |

# Predict Valid Votes:

- Train data Is based on the 21, 22 Knesset elections. (every town instance is repeated twice by the data of that election)
- Features: percentage of voting by blocks(right, center-left, Arabs, orthodox, liberman), city size, total percentage of voting (of the eligble votes), and the disqualified votes percentage
- Test data as required by the results of the 23rd Knesset
- We used the regression random forest to predict disqualified votes with mse index, and then we did 1-score to predict the valid votes.

| עיר | כמות אמיתית | ניבוי |
|---|---|---|
| חיפה | 144625 | 144135 |
| איילת השחר | 585 | 585 |
| אילת | 17085 | 17015 |
| סחנין | 22636 | 22562 |
| קצרין | 3498 | 3492 |

# Predict Votes For Parties:

- Train data Is based on the 21, 22 Knesset elections. (every town instance is repeated twice by the data of that election)

- Features: percentage of voting per party, percentage of total votes from eligble votes, city size, religion, after optimization and checking features importance we considered
- Only voting for MHL, PA, VDAM and religion (important to predict VDAM)

- Test data as required by the 23 Knesset results

- We used the adaboost regressor to predict disqualified votes with mse index

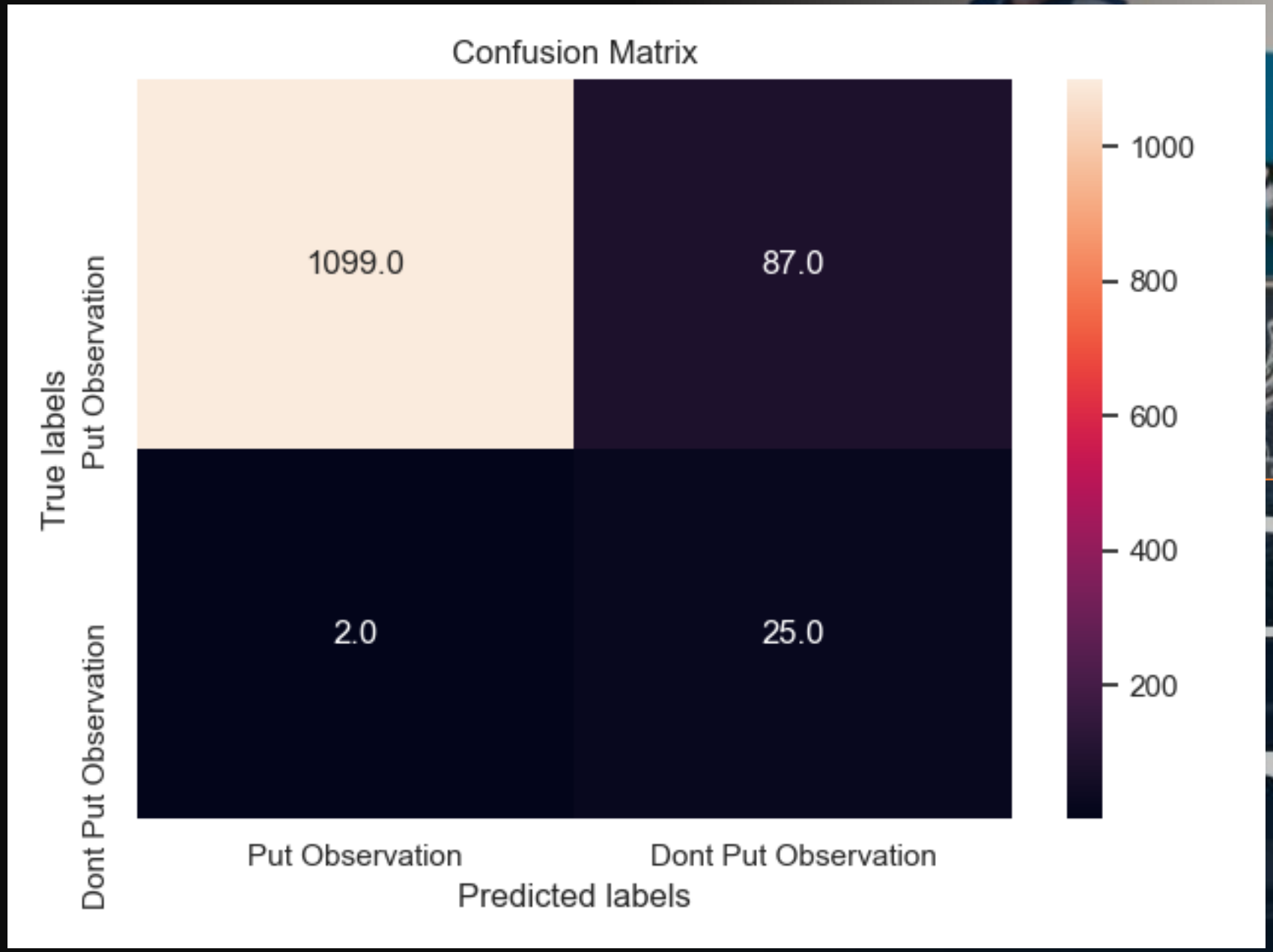
- We tried to predict for each party with out voting percentage of other parties, didn’t improve much


| עיר | מצביעי ליכוד בפועל | ניבוי מצביעי ליכוד | מצביעי כחול לבן בפועל | ניבוי מצביעי כחול לבן | מצביעי רשימה משותפת בפועל | ניבוי מצביעי רשימה משותפת |
|---|---|---|---|---|---|---|
| ירושלים | 72601 | 73718 | 32800 | 30964 | 5321 | 3291 |
| בני ברק | 4951 | 3874 | 1133 | 1812 | 22 | 1072 |
| סחנין | 39 | 553 | 120 | 362 | 16610 | 16402 |
| כרמיאל | 8879 | 8800 | 6627 | 6677 | 468 | 307 |
| דלית אל כרמל | 407 | 323 | 5200 | 5265 | 650 | 557 |

# Predict Where To Enforce Supervision:

<u>Assumptions and Process:</u>

- Unifying parties to voting 5 blocks based on political collaboration
- Train data is based on the voting data for of the 21,22 Knesset :
- Features used: percentage of voting by blocks(right, center-left, Arabs, orthodox, liberman), city size, total percentage of voting (of the eligble votes), and the disqualified votes percentage
- Chosen Algorithm: Regression Random Forest.
  - Predicted very well the valid votes on section 4
  - To predict invalid votes is the opposite of predict valid votes
- We decided to enforce supervision for the 10% of towns with the highest percentage of disqualified votes and also that the town got at least 20 disqualified votes
  - After checking, the percentage of disqualified votes that split the data to 10% and 90% is 0.008 percentage
- After we divided by enforce supervision and not enforce supervision we created a confusion matrix of our predictions


Confusion Matrix
1099.0
87.0
2.0
25.0
1000
800
600
400
200
True labels
Put Observation
Dont Put Observation
Put Observation
Dont Put Observation
Predicted labels