

ANALYZING LATEST ISRAEL ELECTIONS



Course – Machine Learning
Assignment – 2

Nareed Hashem 209243526 | Osama Hariri 312375462



Section 01 – Data



Drop irrelevant or missing features and fill out missing data.



Choose which features to keep based on all research questions.



Inspect the features looking for any correlation.



03 – Clustering

The research question

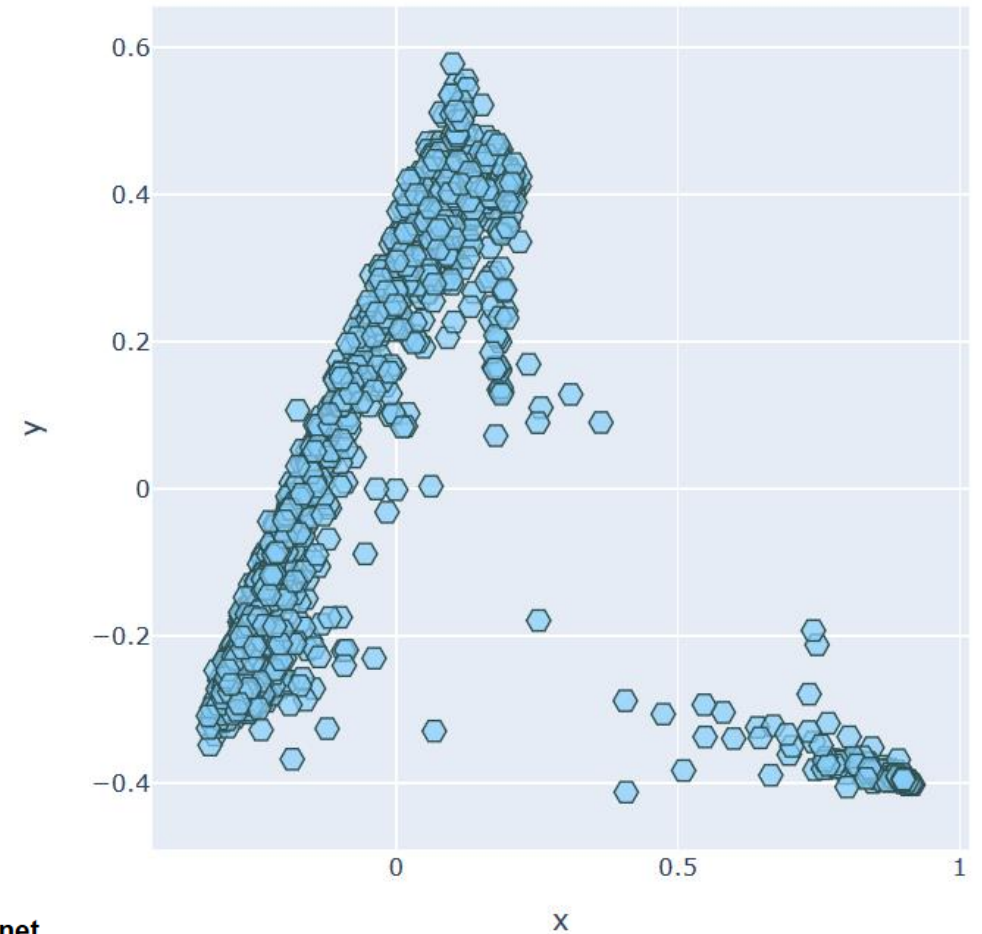
Is Israel's district distribution correlated to which parties each settlement votes for?

The data we choose to fit is a normalized table of the first elections for parties with more than 1k votes which looks like the table below, then using sklearn's PCA we lowered the dimension to 2D and draw a scatter plot.

Ha Avoda	Degel hatora	Joint-list	Zehoot	New yameen	Israel beteno	Likude	N	Natz	Blue- white	Internet pirates	Kets	Shas
-------------	-----------------	------------	--------	---------------	------------------	--------	---	------	----------------	---------------------	------	------

0.022184	0.000000	0.945392	0.0	0.000000	0.001706	0.006826	0.0	0.000000	0.013652	0.003413	0.003413	0.003413
----------	----------	----------	-----	----------	----------	----------	-----	----------	----------	----------	----------	----------

0.028090	0.001498	0.807865	0.0	0.001873	0.001873	0.104869	0.0	0.001124	0.046067	0.000000	0.001498	0.005243
----------	----------	----------	-----	----------	----------	----------	-----	----------	----------	----------	----------	----------

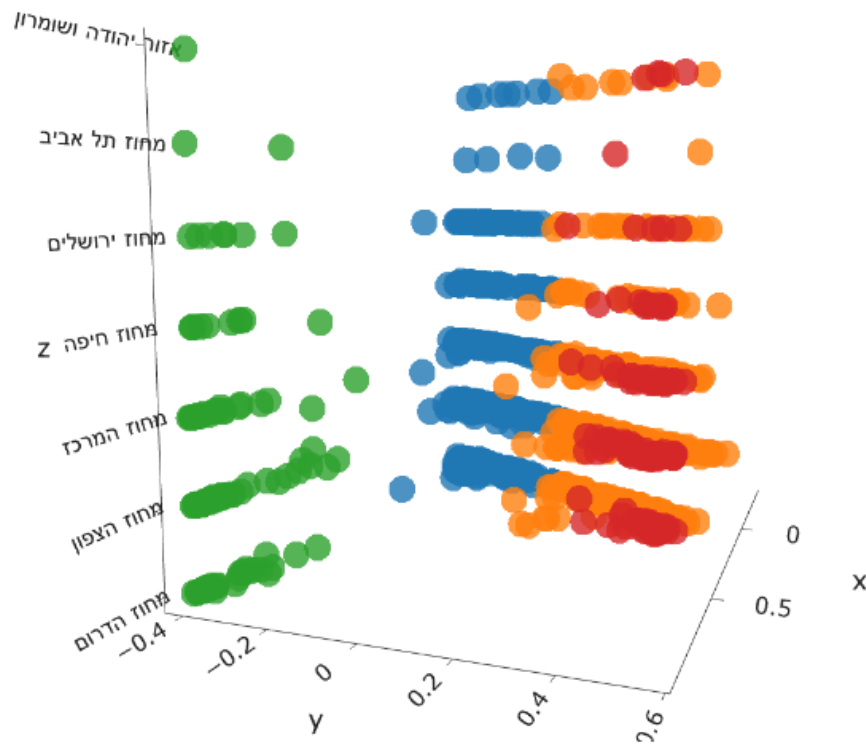




03 – Clustering Methods

KMEANS

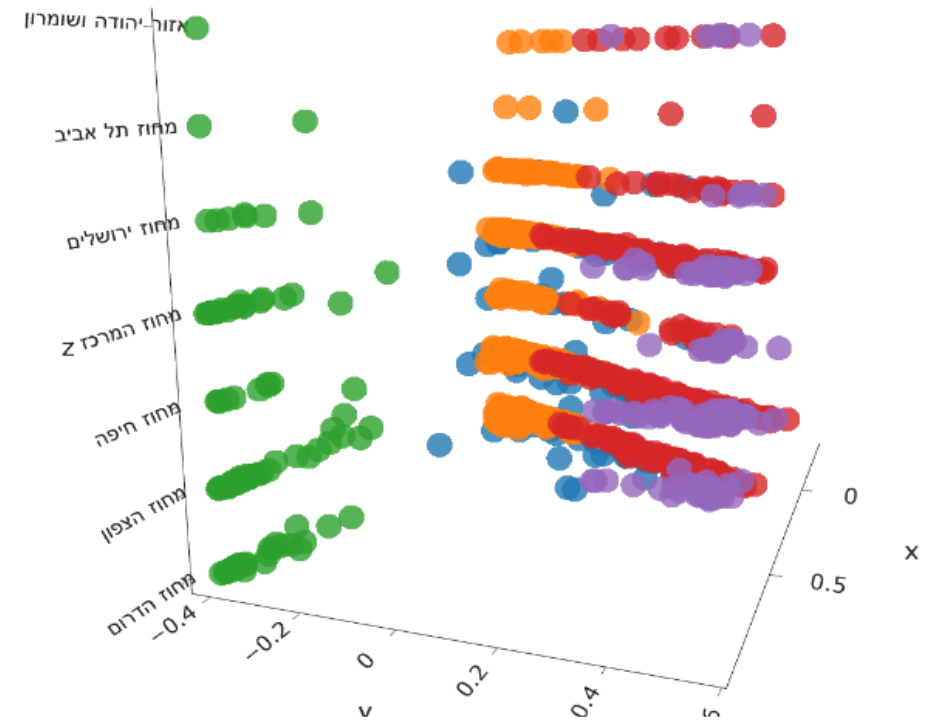
Silhouette Coefficient 0.50621



- Cluster1: Blue-white-0.58, Ha Avoda-0.19, Likude-0.15, New yameen-0.03,
- Cluster2: Likude-0.5, Blue-white-0.18, Shas-0.11, New yameen-0.06,
- Cluster3: Joint-list-0.92, Blue-white-0.03, Ha Avoda-0.03, Likude-0.01,
- Cluster4: New yameen-0.55, Likude-0.26, Blue-white-0.06, Degel hatora-0.05,

GMM

Silhouette Coefficient 0.37593



- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

04 – Random Forest – Predict number of legal votes

Features

- Almost all bycode columns.
- Num of legal votes in last election.

MSE = 1,105,1438.11

Log(MSE) = 7.043

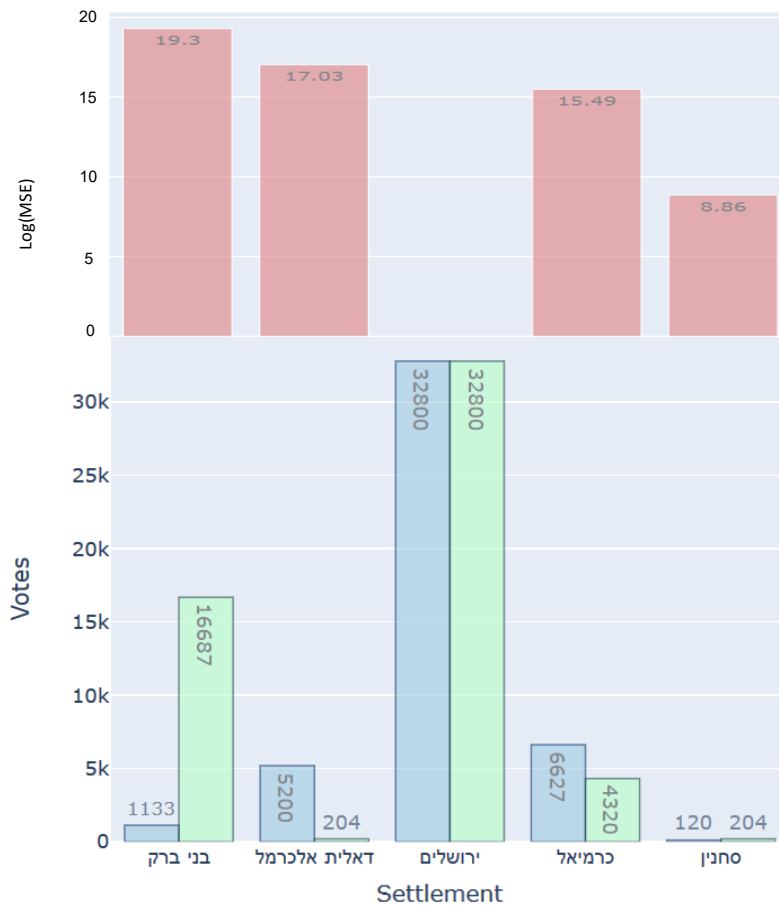




05 – Adaboost – Predict number of voters

Blue-white

Features - Population 2018, Jews, Arabs



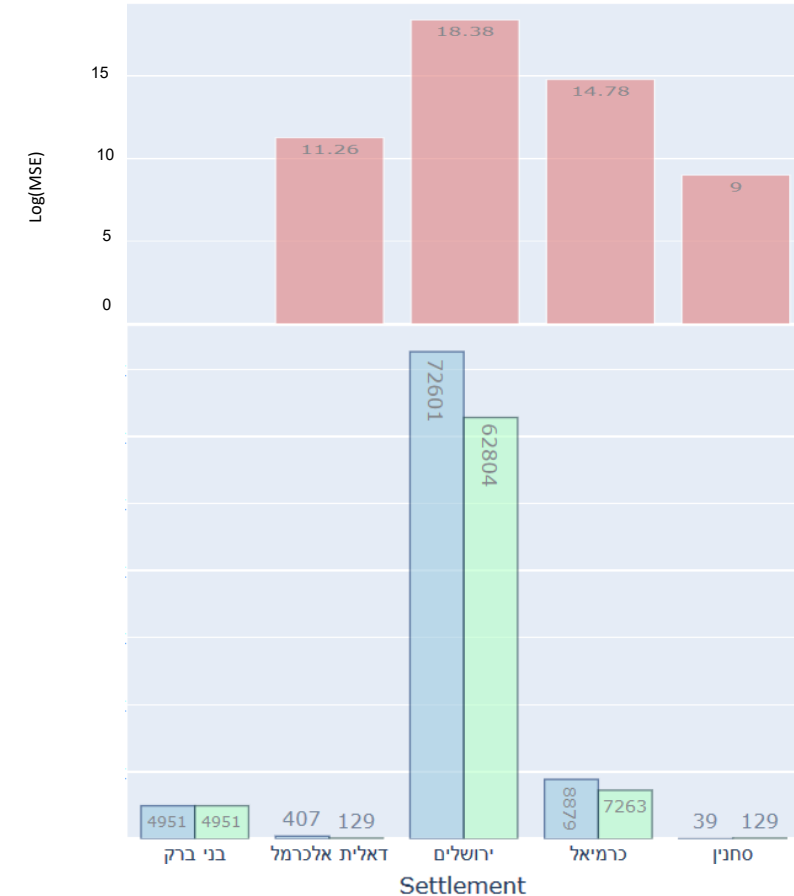
Joint-list

Features - Arabs



Likud

Features - Population 2018, Jews





06 – Where to place representatives?

Threshold

MAX(5% of highest disqualified votes)

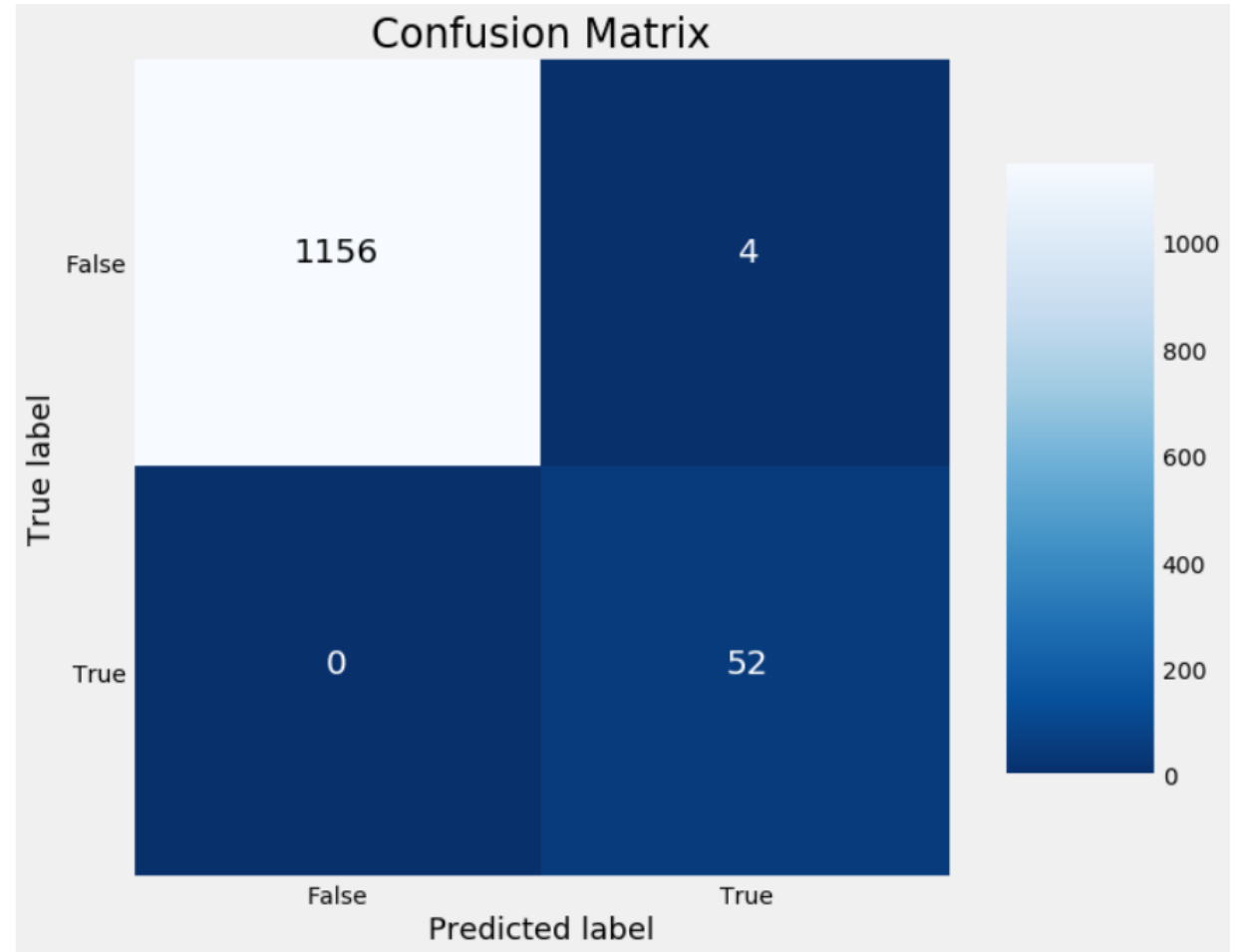
Threshold is not relative to the settlements size since its more important to stop 6k disq votes when they're 4% rather than 200 votes when they're a 20%.

Features

- Number of voters.
- Number of Jews.
- Number of Arabs.
- Religion.

Accuracy

0.99669





QUESTIONS