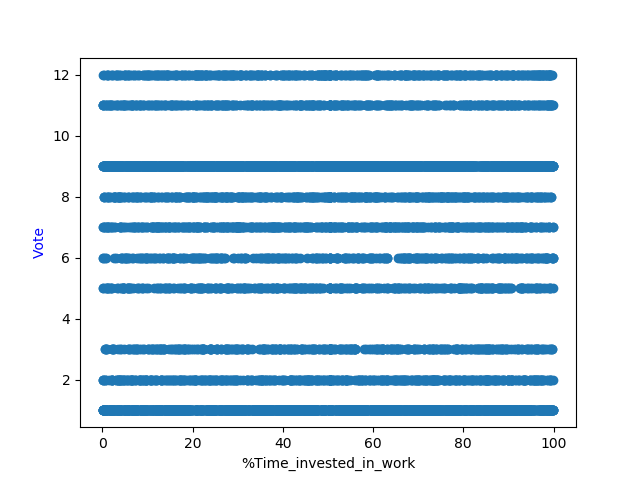
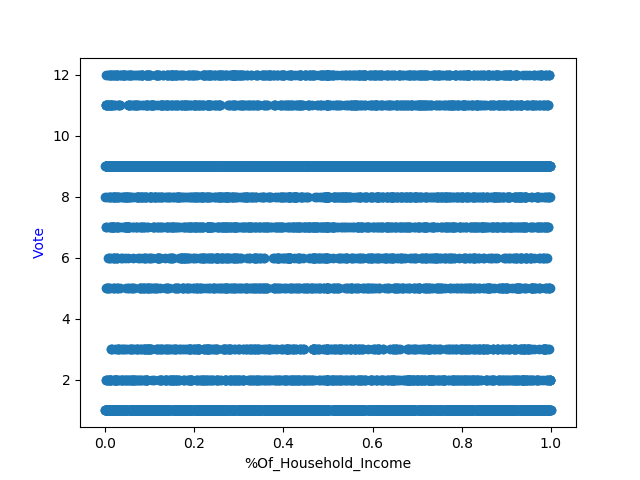
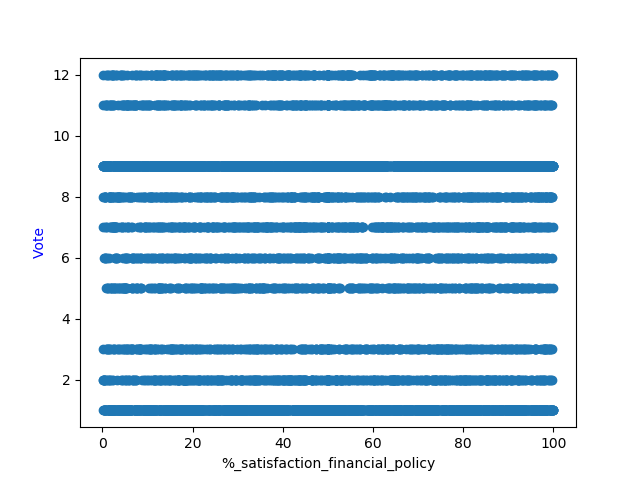
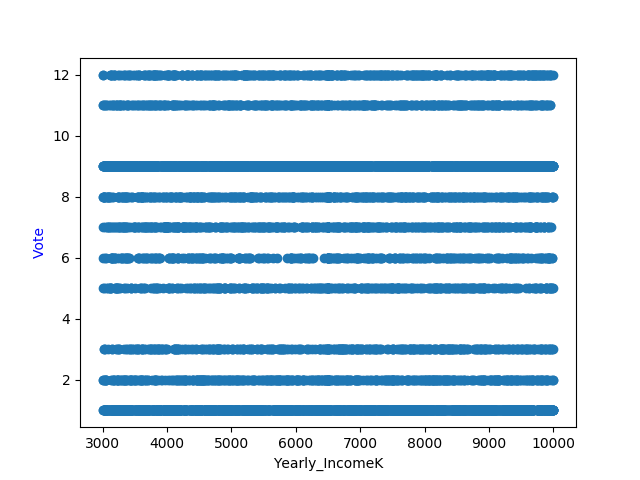
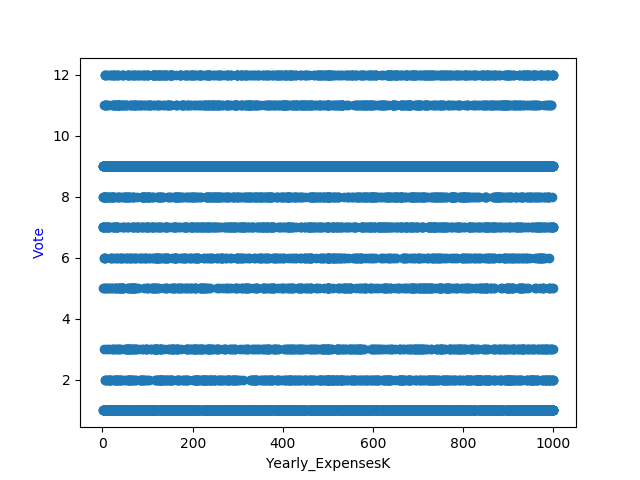
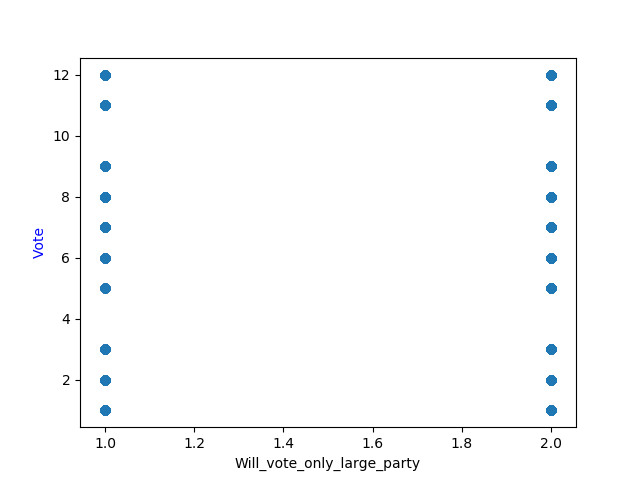
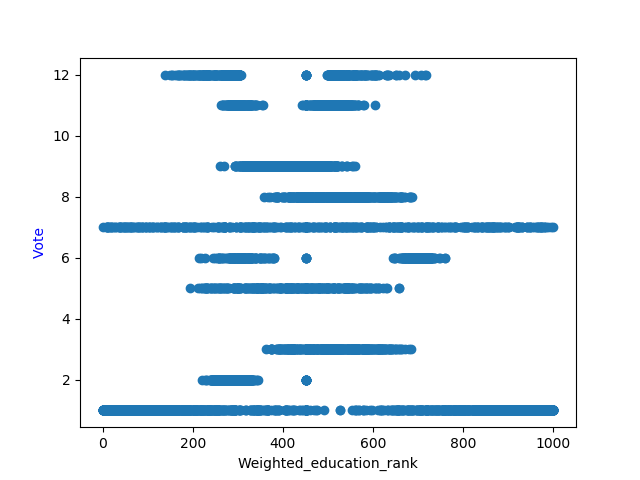
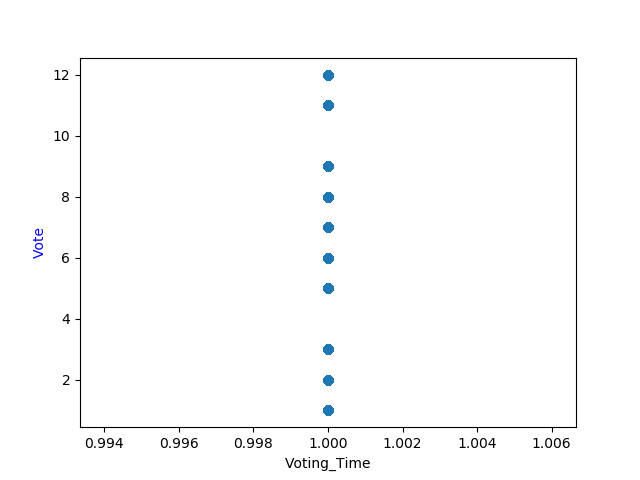
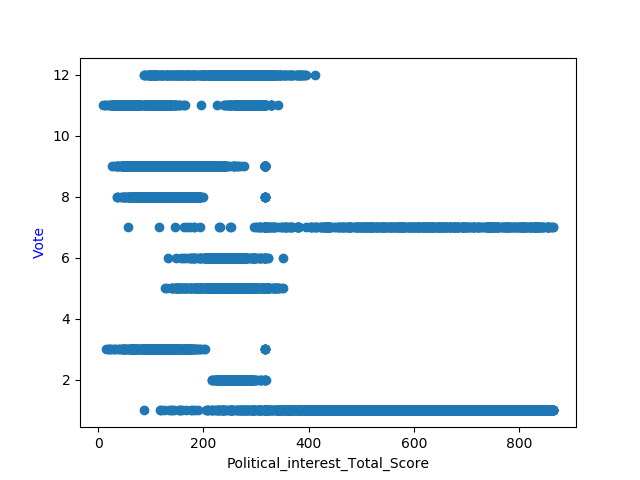
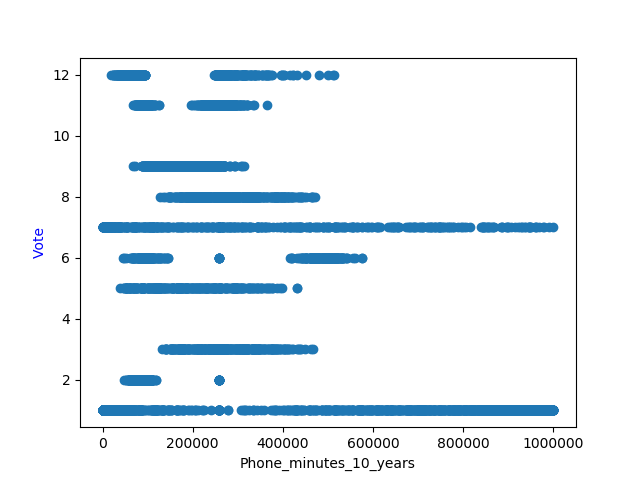
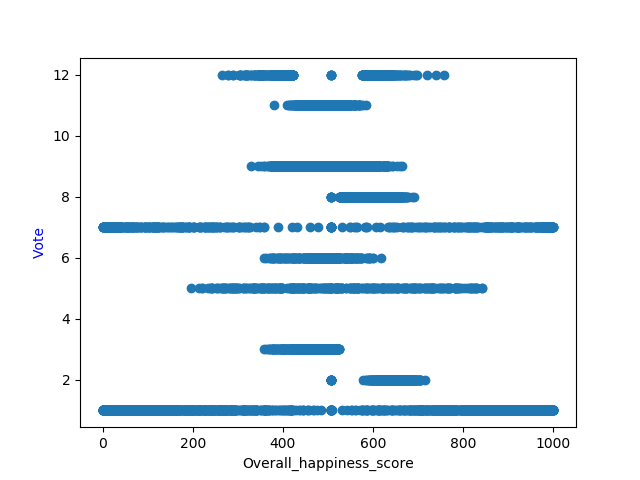
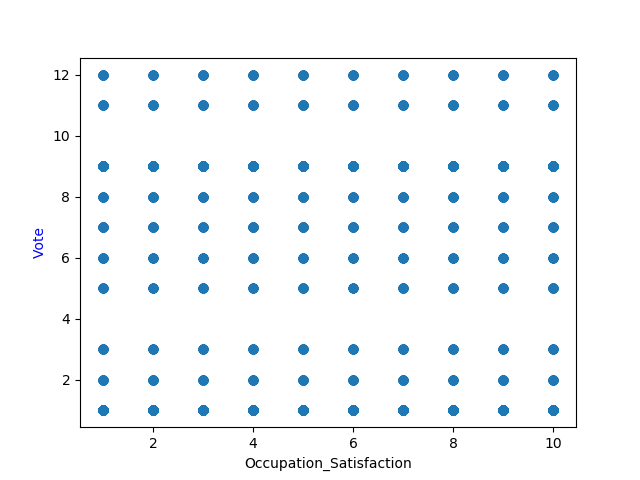
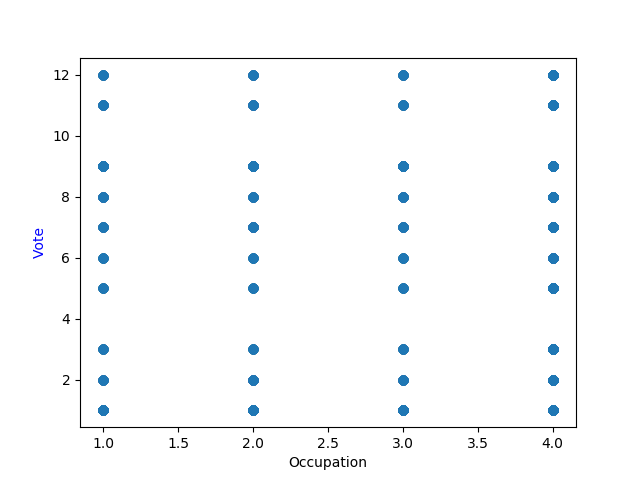
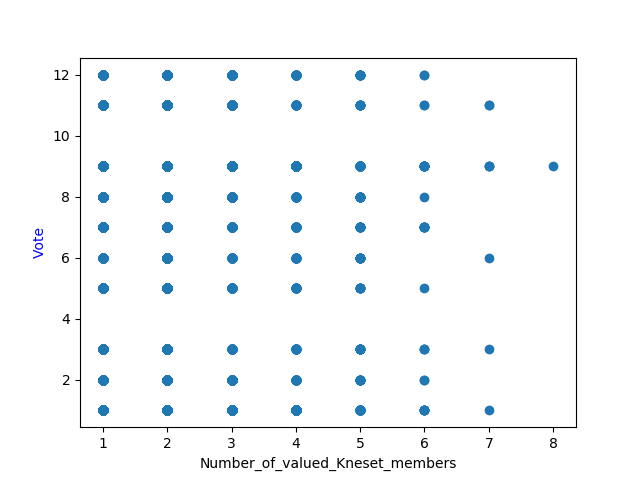
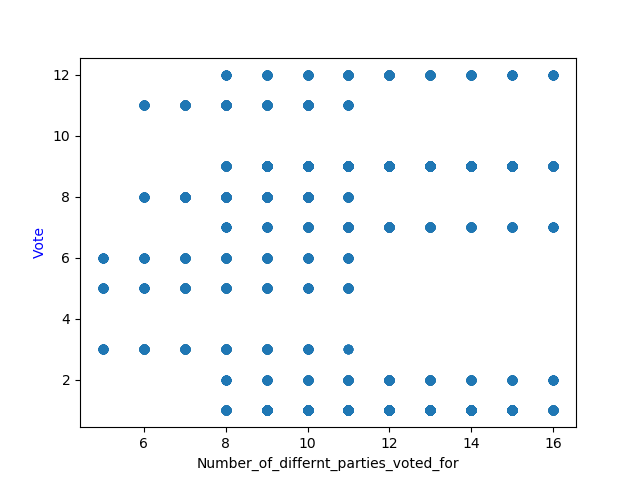
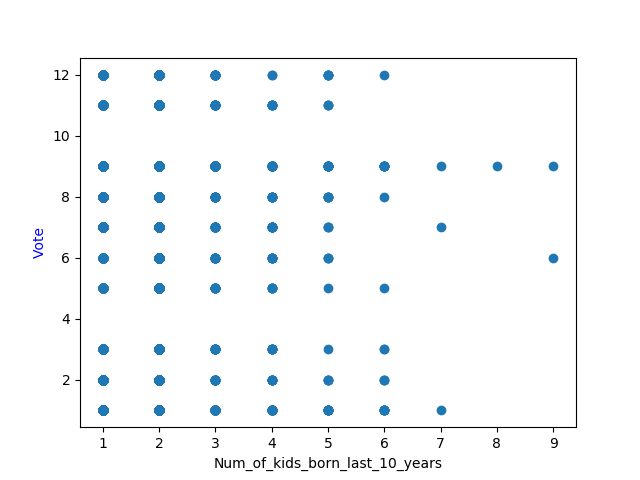
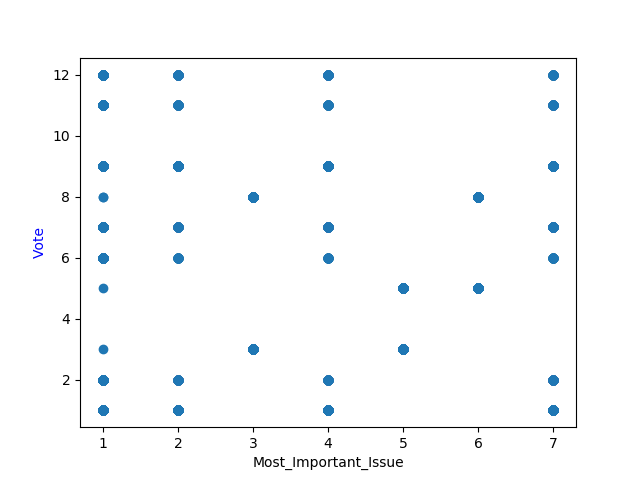
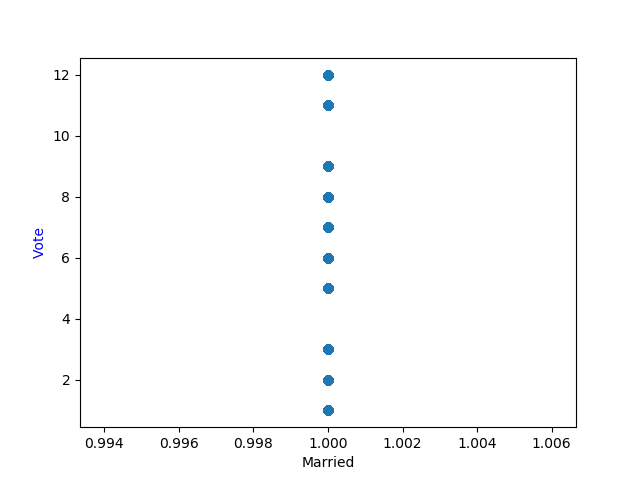
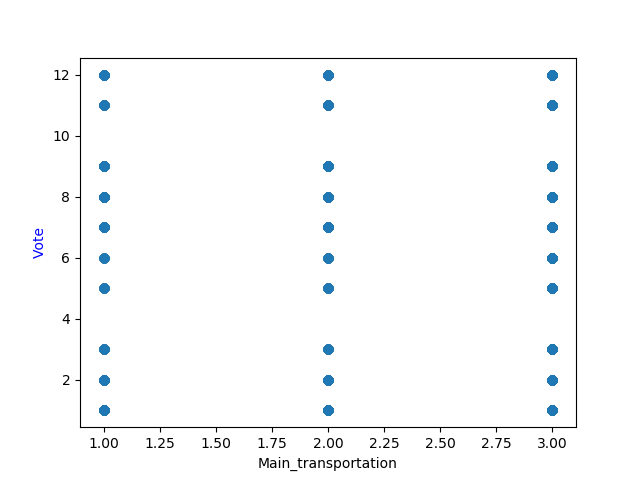
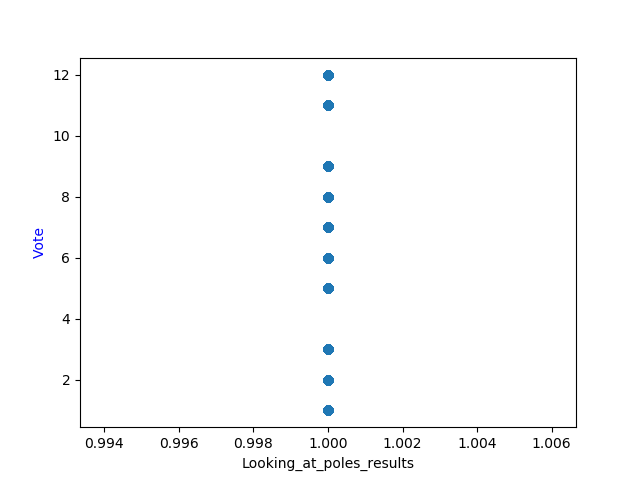
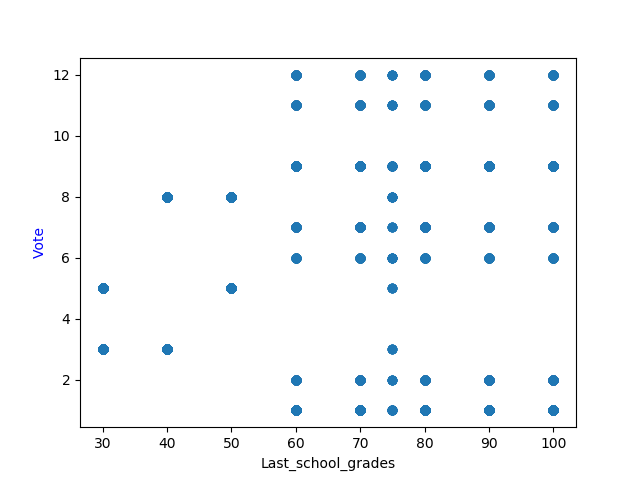
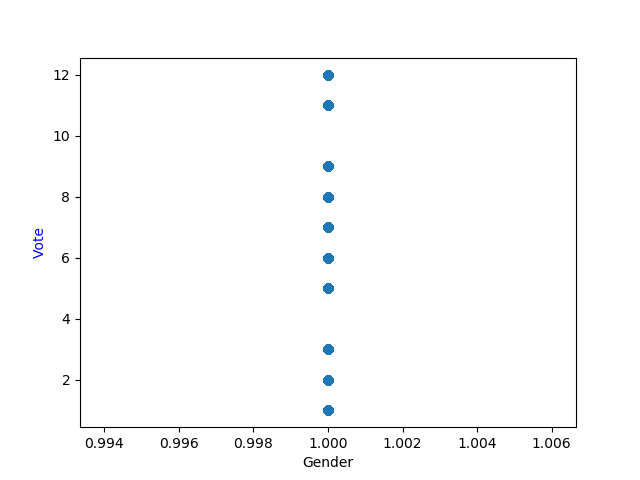
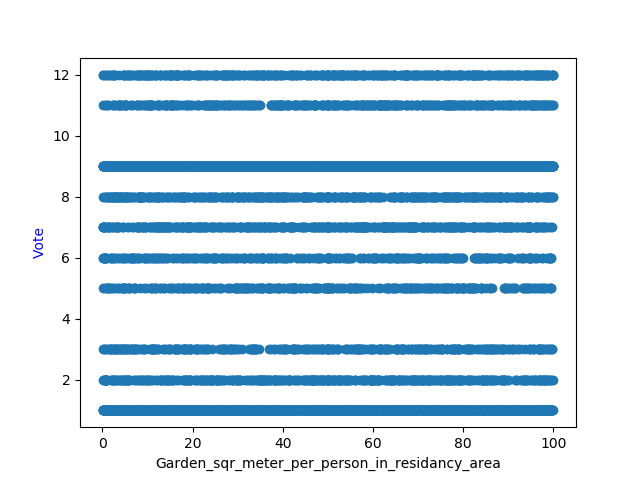
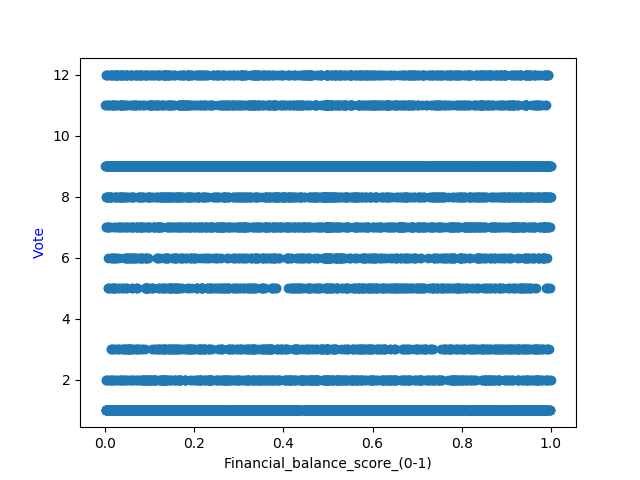
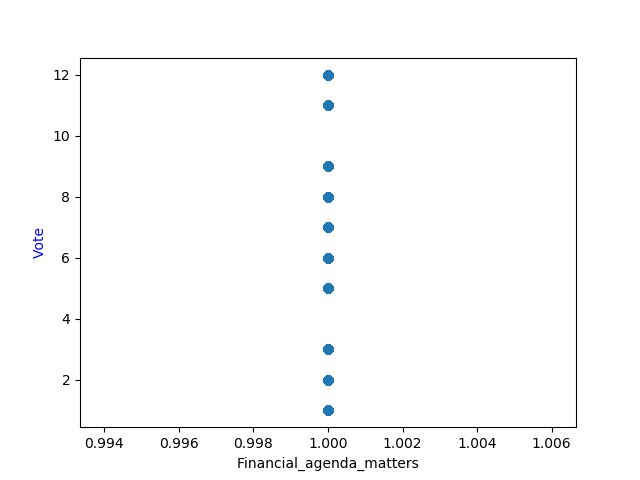
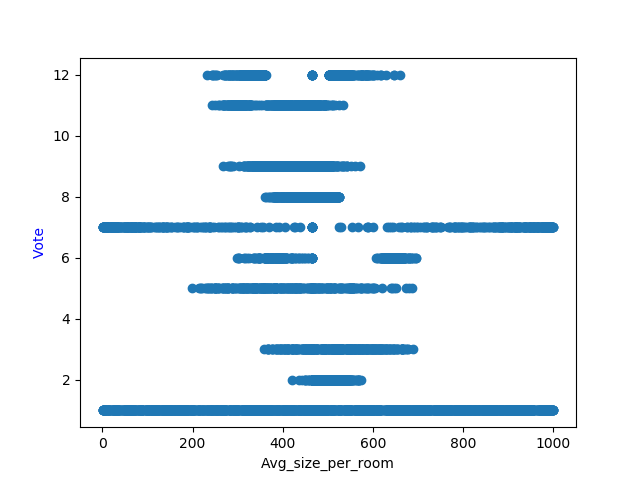
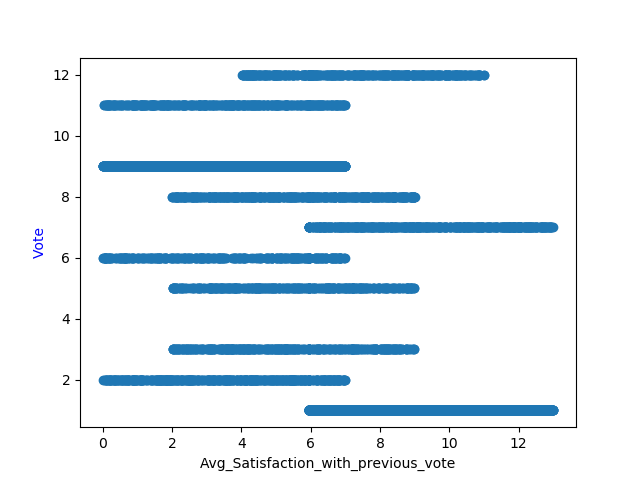
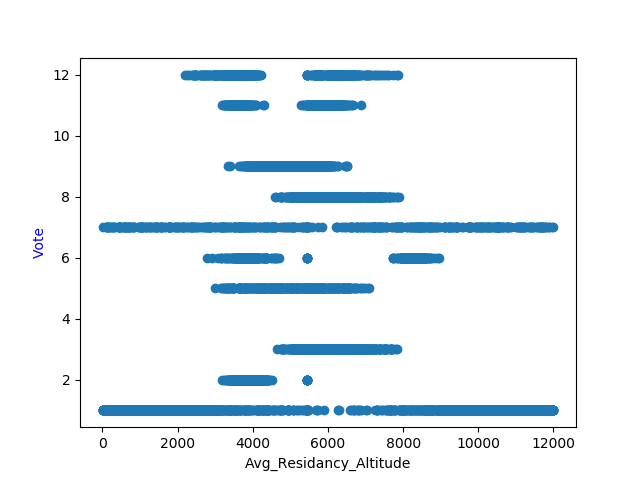
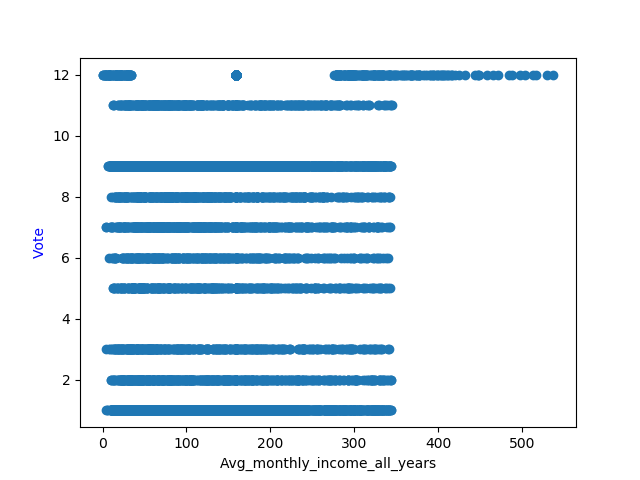
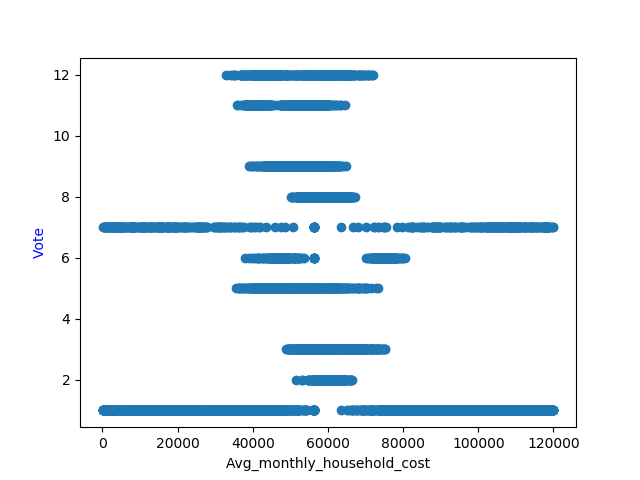
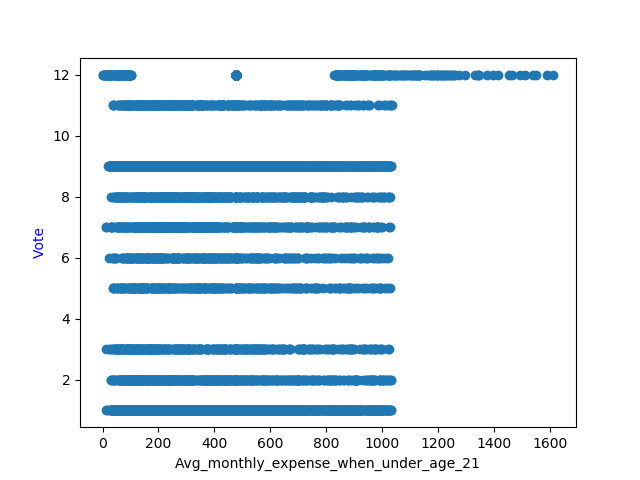
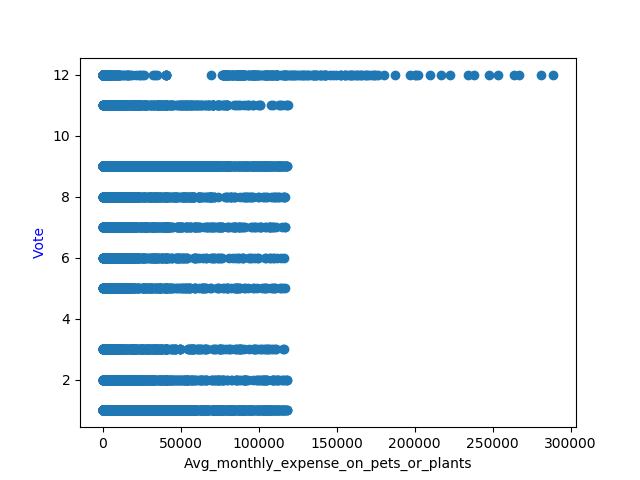
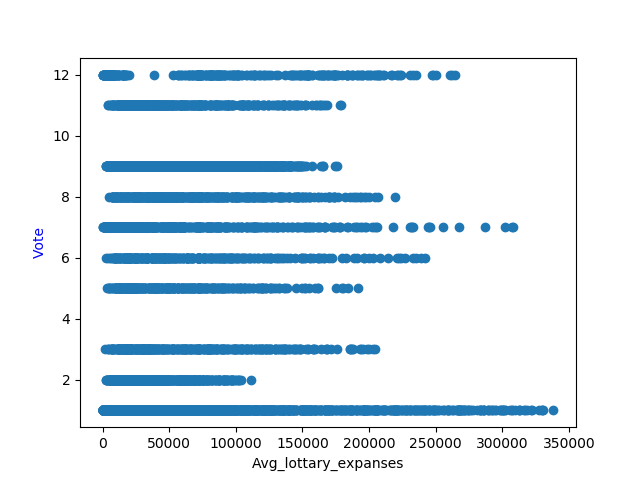
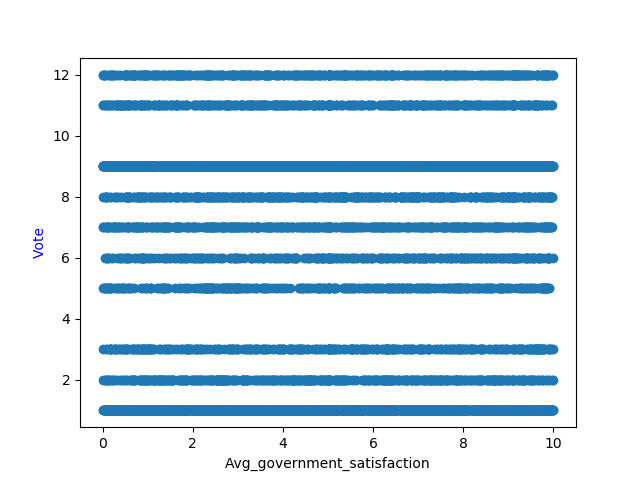
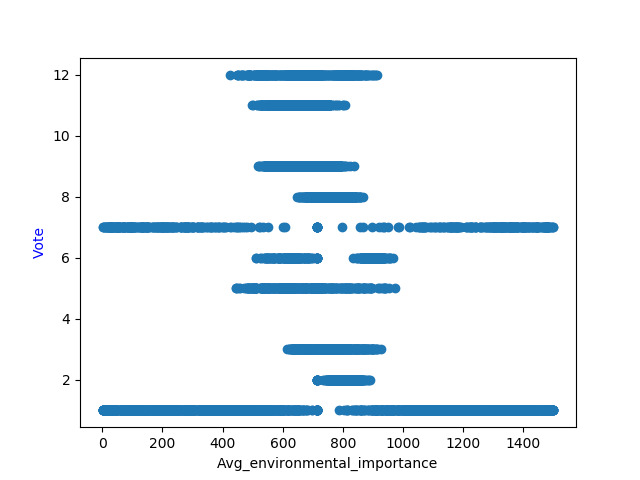
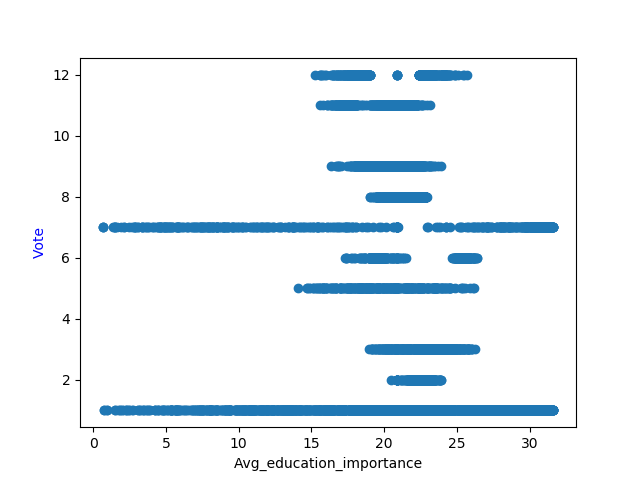
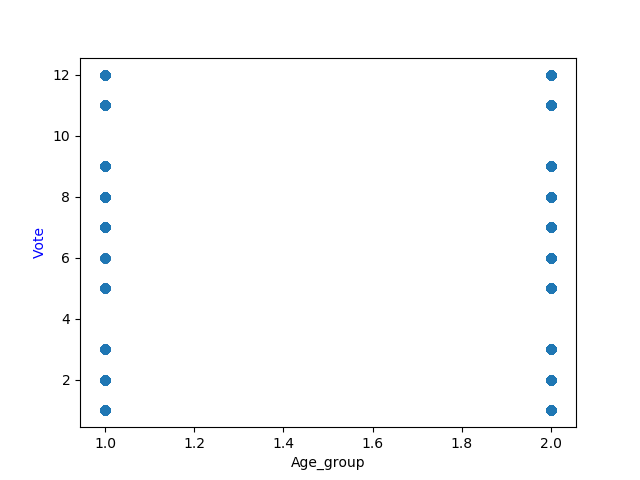
***Report hw2***

* ***Relation or lack of relation between the features and the labels features:***



'%\_satisfaction\_financial\_policy' – no correlation

'%Of\_Household\_Income' - no correlation

'%Time\_invested\_in\_work' - no correlation

'Age\_group' - no correlation

'Avg\_education\_importance' - people who care about their education vote to Browns and Purples.

'Avg\_environmental\_importance' - same for the environment

'Avg\_government\_satisfaction' - no correlation

'AVG\_lottary\_expanses' – people who spent a lot of money on the lottery is a vote for Browns

'Avg\_monthly\_expense\_on\_pets\_or\_plants' - people who spent a lot of money on pets or plants is a vote for Yellows

'Avg\_monthly\_expense\_when\_under\_age\_21' – some is before

'Avg\_monthly\_household\_cost' – people who care about household vote for Browns and Purples.

'Avg\_monthly\_income\_all\_years' – if the income greater from 350 then vote for Yellows, otherwise no correlate.

'Avg\_Residancy\_Altitude' – if the number of residences greater from 10000 or less then 2000, then vote for Browns or Purples.

'Avg\_Satisfaction\_with\_previous\_vote' – if satisfaction greater then 8 vote for Browns, Purples, or Yellows.

'Avg\_size\_per\_room' - if the size of the room greater from 700 or less then 200, then vote for Browns or Purples.

'Financial\_agenda\_matters' - no correlation

'Financial\_balance\_score\_(0-1)' - no correlation

'Garden\_sqr\_meter\_per\_person\_in\_residancy\_area' - no correlation

'Gender' - no correlation

'Last\_school\_grades' – people who get grades less than 55, votes for Greys, Oranges or Reds.

'Looking\_at\_poles\_results' - no correlation

'Main\_transportation' - no correlation

'Married' - no correlation

'Most\_Important\_Issue' – have a little correlation, if the important is number 3, 5 or 6 we vote for Greys, Oranges or, Reds.

'Num\_of\_kids\_born\_last\_10\_years' - if someone has greater then 7 kids is a vote for Turquoises or Pinks.

'Number\_of\_differnt\_parties\_voted\_for' – if the number of different parties voted is less then 6, then is a vote for Greys, Oranges, or Pinks.

'Number\_of\_valued\_Kneset\_members' – if greater then 7, then vote for Turquoises.

'Occupation' - no correlation

'Occupation\_Satisfaction' - no correlation

'Overall\_happiness\_score' - if the score is greater from 800 or less then 200, vote for Browns or Purples.

'Phone\_minutes\_10\_years' - if the score is greater from 600000, then vote for Browns or Purples.

'Political\_interest\_Total\_Score' - if the score is greater from 420, then vote for Browns or Purples.

'Voting\_Time' - no correlation

'Weighted\_education\_rank' - if the rank is greater from 800 or less then 180, vote for Browns or Purples.

'Will\_vote\_only\_large\_party' - no correlation

'Yearly\_ExpensesK' - no correlation

'Yearly\_IncomeK' - no correlation

***Summary:***

18 - no correlation

19 – yes correlation

no\_correlation = [

'%\_satisfaction\_financial\_policy' ,

'%Of\_Household\_Income' ,

'%Time\_invested\_in\_work',

'Age\_group' ,

'Avg\_government\_satisfaction' ,

'Financial\_agenda\_matters' ,

'Financial\_balance\_score\_(0-1)',

'Garden\_sqr\_meter\_per\_person\_in\_residancy\_area' ,

'Gender',

'Looking\_at\_poles\_results' ,

'Main\_transportation' ,

'Married' ,

'Occupation' ,

'Occupation\_Satisfaction' ,

'Voting\_Time' ,

'Will\_vote\_only\_large\_party' ,

'Yearly\_ExpensesK' ,

'Yearly\_IncomeK']

yes\_correlation = [

'Avg\_education\_importance' ,

'Avg\_environmental\_importance',

'AVG\_lottary\_expanses' ,

'Avg\_monthly\_expense\_on\_pets\_or\_plants' ,

'Avg\_monthly\_expense\_when\_under\_age\_21' ,

'Avg\_monthly\_household\_cost' ,

'Avg\_monthly\_income\_all\_years' ,

'Avg\_Residancy\_Altitude' ,

'Avg\_Satisfaction\_with\_previous\_vote' ,

'Avg\_size\_per\_room' ,

'Last\_school\_grades' ,

'Num\_of\_kids\_born\_last\_10\_years' ,

'Number\_of\_differnt\_parties\_voted\_for' ,

'Number\_of\_valued\_Kneset\_members' ,

'Overall\_happiness\_score' ,

'Phone\_minutes\_10\_years' ,

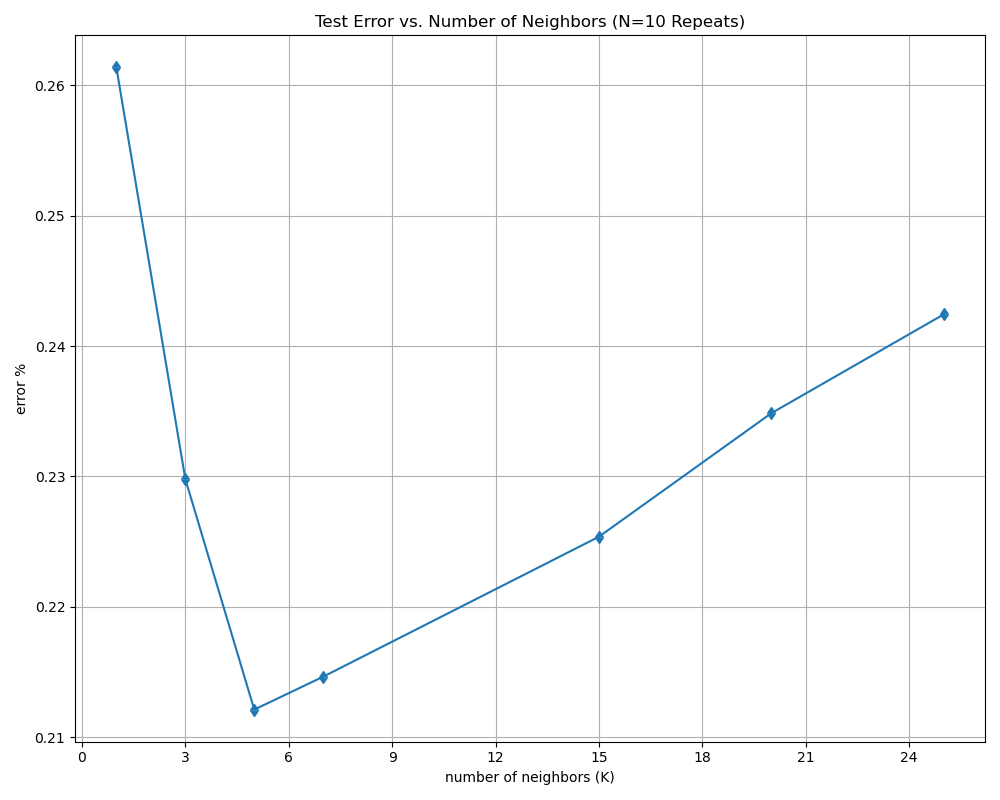
'Political\_interest\_Total\_Score' ,

'Weighted\_education\_rank',

'Most\_Important\_Issue' ]

* ***Relief algorithm:***

Their strengths are that they are not dependent on heuristics, they run in low-order polynomial time, and they are noise-tolerant and robust to feature interactions, as well as being applicable for binary or continuous data; however, it does not discriminate between redundant features, and low numbers of training instances fool the algorithm.



['Vote' 'Avg\_monthly\_expense\_when\_under\_age\_21' 'Avg\_lottary\_expanses'

'Avg\_environmental\_importance' 'Financial\_balance\_score\_(0-1)'

'Avg\_Residancy\_Altitude' 'Yearly\_ExpensesK' '%Time\_invested\_in\_work'

'Avg\_Satisfaction\_with\_previous\_vote' 'Avg\_government\_satisfaction'

'Last\_school\_grades' 'Number\_of\_differnt\_parties\_voted\_for'

'Political\_interest\_Total\_Score' 'Overall\_happiness\_score']

'Avg\_lottary\_expanses', 'Financial\_balance\_score\_(0-1)', 'Yearly\_ExpensesK',

'%Time\_invested\_in\_work', 'Avg\_government\_satisfaction': appers in Relief but not apper in SBS.

* ***SBS algorithm:***

**Pros:**

Conservative, can choose how much features u want

**Cons:**

May take a lot of time if have a lot of features or the classifier is heavy

**KNN:**

knn **=** KNeighborsClassifier**(**n\_neighbors**=**3**)**

sbs **=** SFS**(**knn**,**

k\_features**=**8**,**

forward**=False,** # if forward = True then SFS otherwise SBS

floating**=False,**

verbose**=**2**,**

n\_jobs**=-**1**,**

scoring**=**'accuracy'**)**

**features:**

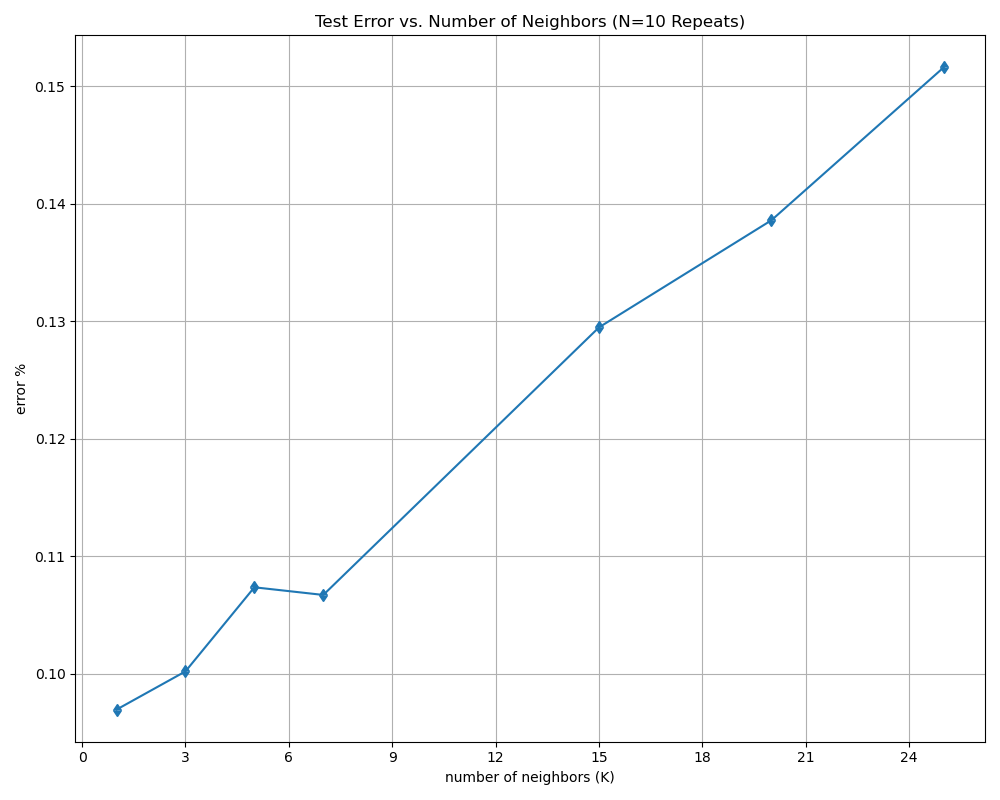
['Vote' 'Avg\_environmental\_importance' 'Yearly\_IncomeK'

'Avg\_Residancy\_Altitude' 'Avg\_Satisfaction\_with\_previous\_vote'

'Last\_school\_grades' 'Number\_of\_differnt\_parties\_voted\_for'

'Political\_interest\_Total\_Score' 'Overall\_happiness\_score']

give us 90% accuracy.



**GradientBoost**

clf\_gradient **=** GradientBoostingClassifier**(**n\_estimators**=**100**,** random\_state**=**0**)**

sbs **=** SFS**(**clf\_gradient**,**

k\_features**=**8**,**

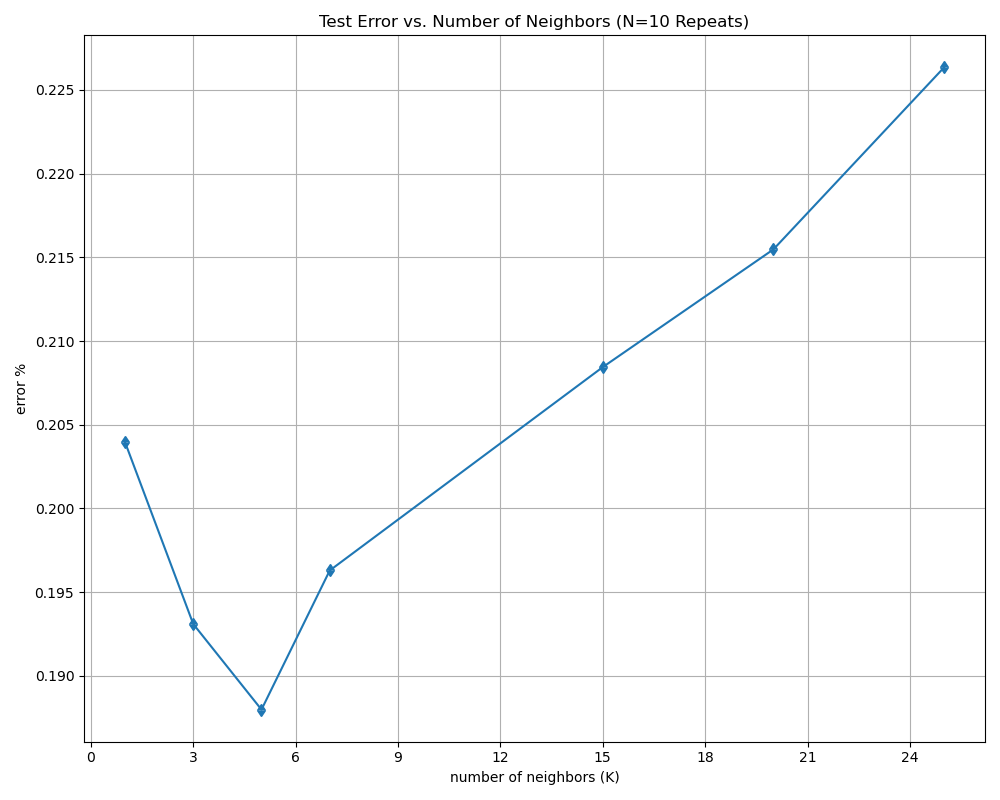
forward**=False,** # if forward = True then SFS otherwise SBS

floating**=False,**

verbose**=**2**,**

n\_jobs**=-**1**,**

scoring**=**'accuracy'**)**



['Vote' 'Avg\_monthly\_expense\_when\_under\_age\_21'

'Avg\_environmental\_importance' 'Avg\_Residancy\_Altitude'

'Avg\_Satisfaction\_with\_previous\_vote' 'Most\_Important\_Issue'

'Number\_of\_differnt\_parties\_voted\_for' 'Political\_interest\_Total\_Score'

'Overall\_happiness\_score']

'Avg\_monthly\_expense\_when\_under\_age\_21', 'Most\_Important\_Issue': appers in GradientBoost

but not appers in KNN.

'Yearly\_IncomeK', 'Last\_school\_grades': appers in KNN

but not appers in GradientBoost.

***Summey for features selection:***

**(SBS from KNN) + (SBS from GradientBoost) + (Relief) – (no\_correlation) =**

**final =**

**[**

**'Avg\_monthly\_expense\_when\_under\_age\_21'**

**'Avg\_environmental\_importance'**

**'Avg\_Residancy\_Altitude'**

**'Avg\_Satisfaction\_with\_previous\_vote'**

**'Most\_Important\_Issue'**

**'Number\_of\_differnt\_parties\_voted\_for'**

**'Political\_interest\_Total\_Score'**

**'Overall\_happiness\_score'**

**'Last\_school\_grades'**

**'Avg\_lottary\_expanses'**

**]**