



# DV Program Run Instructions

By: CWU Visual Knowledge Discovery Lab



# Overview

- Creating a Project
- Interacting with Data
- Analytics
- 3+ Class Visualizations
- Saving a Project



# Preprocessing

- Must be a csv file or formatted similarly.
- Must have a header row and class column.
- The class column must be last.
- ID columns are allowed, but they must be the first column in the dataset.

ID	feature1	feature2	feature3	class
1	5	1	1	dog
2	5	4	1	dog
3	3	1	1	dog
4	6	8	1	cat
5	4	1	1	cat
6	8	10	1	cat
7	1	1	1	bird
8	2	1	1	bird
9	2	1	5	bird



# Creating a Project

- Select “File” then “Create New Project.”
  - Or press Alt + N for a keyboard shortcut.
- Questions about the dataset will then appear.
  - Is the first column for ID?
  - Is the last column for classes?
  - Min-Max or z-Score Min-Max normalization?
- Answer the questions according to the dataset being visualized.

File	Help
Create New Project	Alt+N
Open Saved Project	Alt+O
Save Project	Alt+S
Save Project As	Alt+A
Import Data	Alt+I
Validation Data	Alt+V



# UCI Machine Learning Repository Dataset

- Wisconsin Breast Cancer
  - Add header row to dataset.
  - Answer “Yes” to the first column being for ID.
  - Answer “Yes” to the last column being for classes.
  - Select “z-Score Min-Max Normalization”

# Setting up Wisconsin Breast Cancer Dataset

1.

ID	feature1	feature2	feature3	feature4	feature5	feature6	feature7	feature8	feature9	class
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2

2.

ID Column

Does this project use the first column to designate ID?

3.

Classes

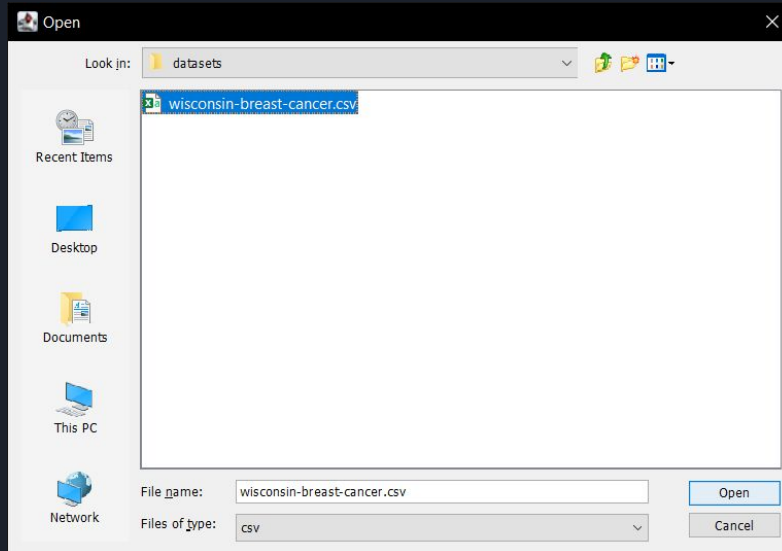
Does this project use the last column to designate classes?

4.

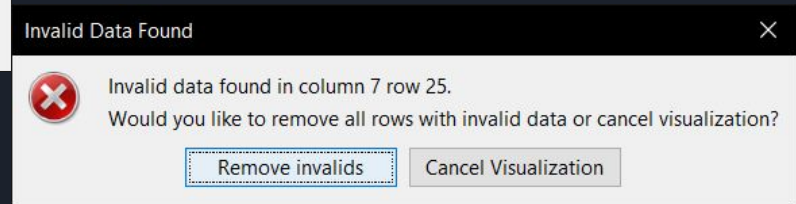
Normalization Style

Choose a normalization style or click "Help" for more information on normalization styles.

# Setting up Wisconsin Breast Cancer Dataset

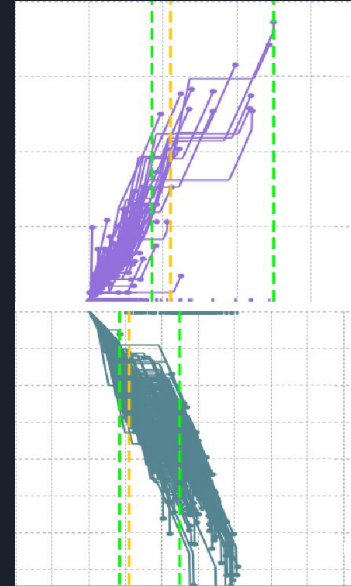
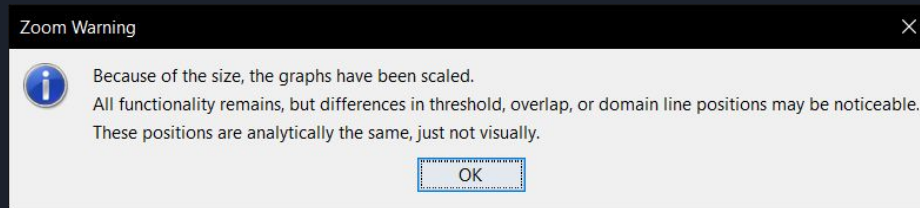


- Load from Examples directory in the same folder where the program is located.
- WBC has invalid/incomplete data.
- Select “Remove Invalids” to visualize valid data.



# Interacting with Data

- Zoom Warning will appear signifying one or more graphs have been scaled.
- In this case, the scaled graph will have smaller grid squares than the non-scaled graph.



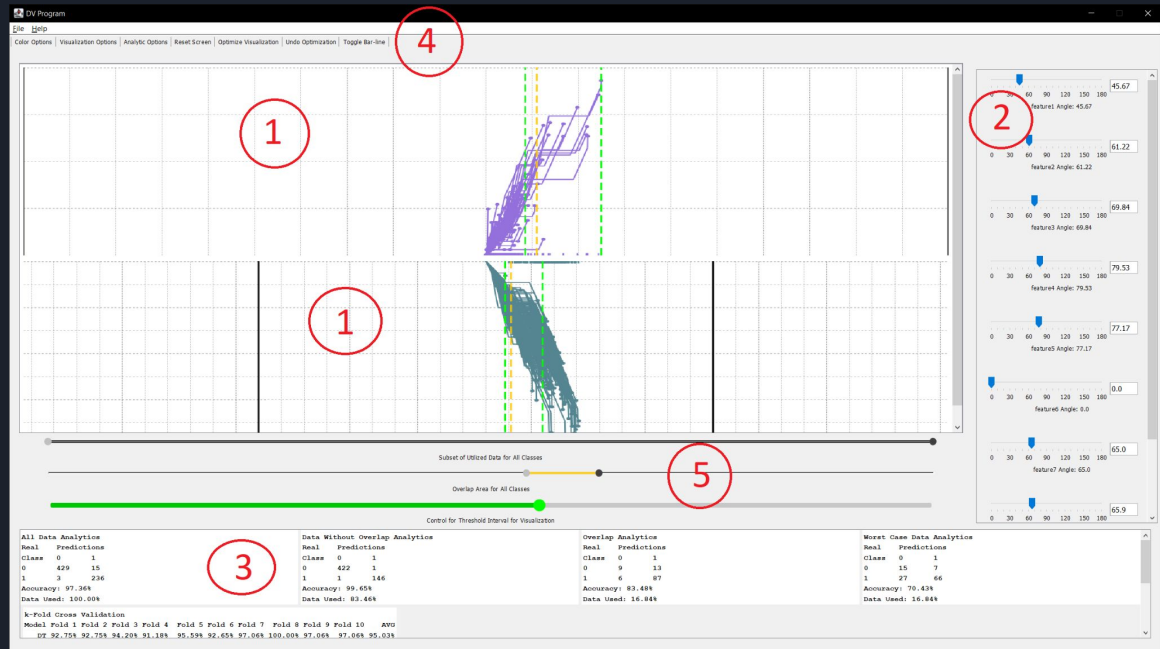


# Interacting with Data

Once the project is created, the program will be the same as below.

Key:

1. Graphs
2. Angles
3. Analytics
4. Toolbar
5. Sliders



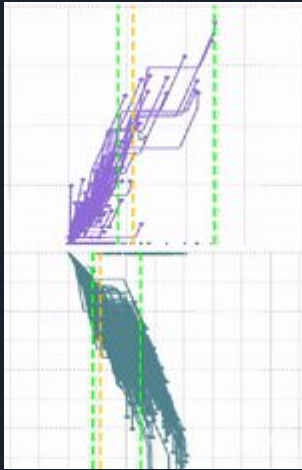


# Experiment with Interacting with Data

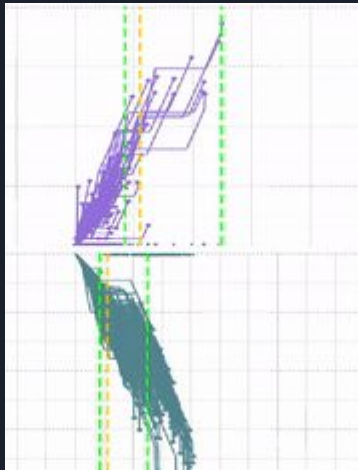
- Basic User Tools:
  - Panning: ctrl + left click drag allows for panning across the graphs.
  - Zooming: the scroll wheel allows both zooming in and out
  - Scaling:
    - left click drag down and to the right gives a scaled zoom of the selected area
    - left click drag down and left or up scales the graph outwards for full viewing angles
  - Reset Screen: resets screen to initial viewing position

# Interacting with Data

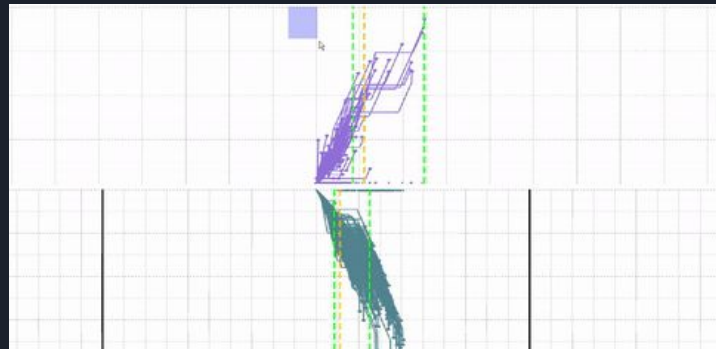
## Basic User Tool Examples:



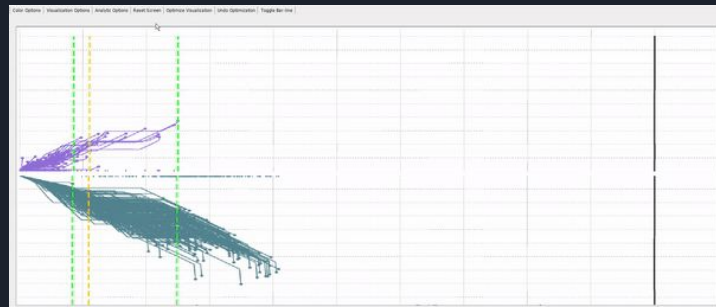
Panning



Zooming



Scaling

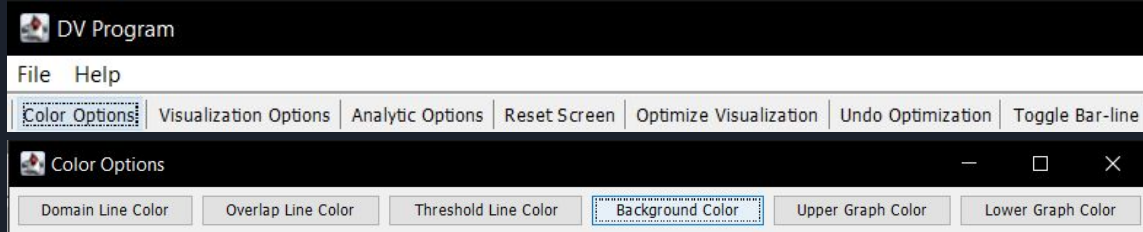


Reset Screen

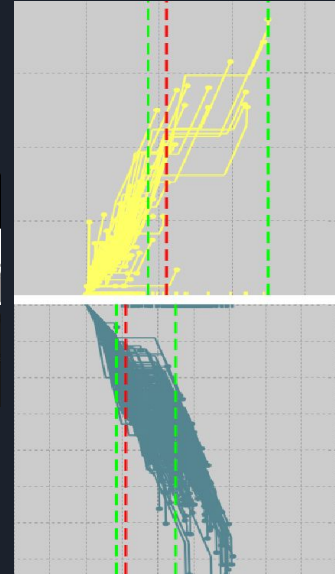
# Interacting with Data

- Colors

- To change the colors of any part of the visualization, select the “Color Options” button on the toolbar.

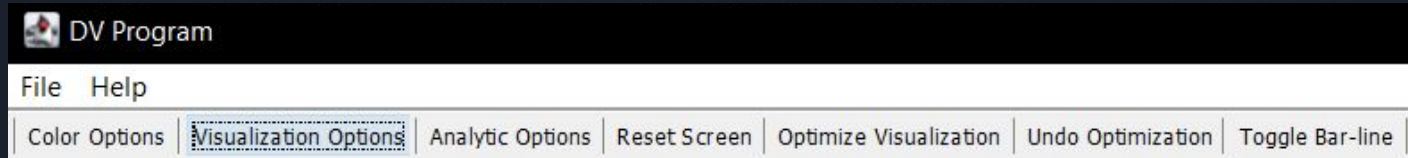


- Changing the “Upper Graph,” “Threshold,” and “Background” colors to yellow, red, and grey results in the shown visualization.



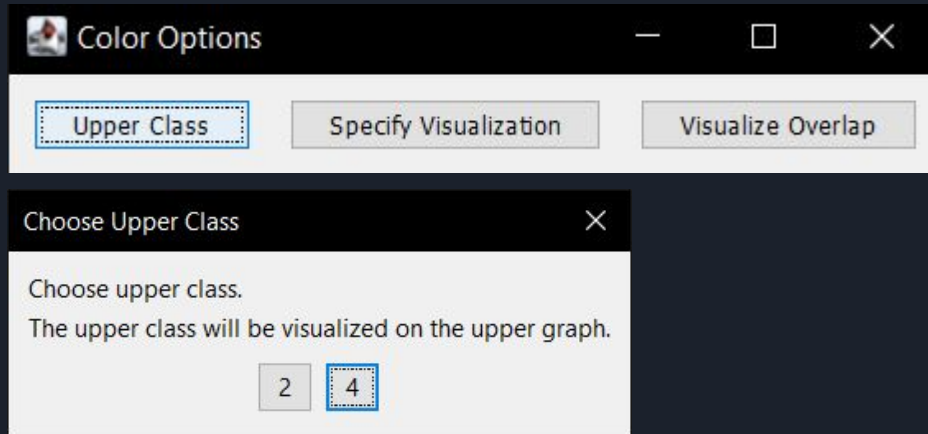
# Interacting with Data: data visualization

- Class Graphs
  - By selecting “Visualization Options” we can change which class is visualized on which graph.
  - We can also choose to “Visualize Overlap” to visualize only the overlapping datapoints.



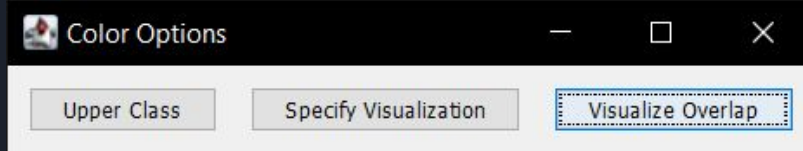
# Interacting with Data

- Changing which class is on the Upper Graph.
  - Changing the class to “4” results in the shown visualization.

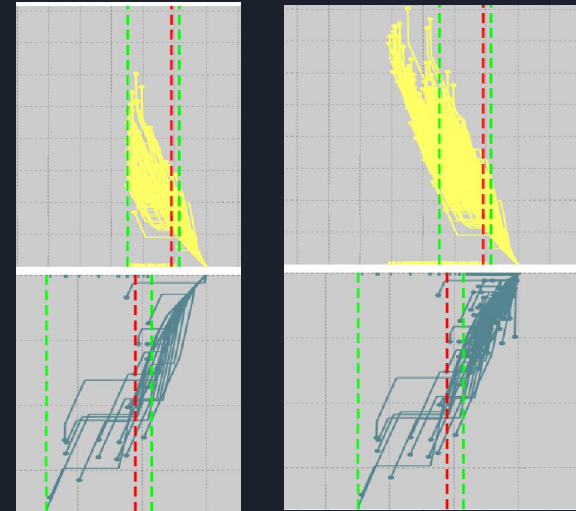
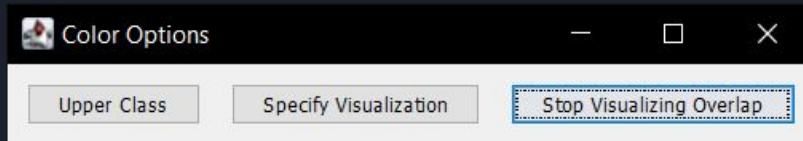


# Interacting with Data

- Visualizing overlap
  - Selecting “Visualize Overlap” will result in the shown visualization.



- Selecting “Stop Visualizing Overlap” will display all data again.

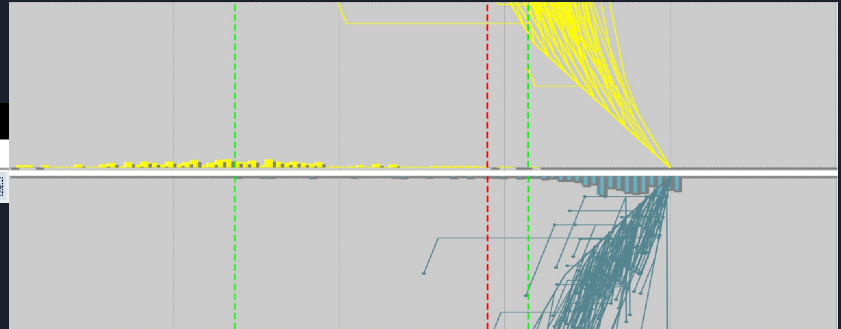
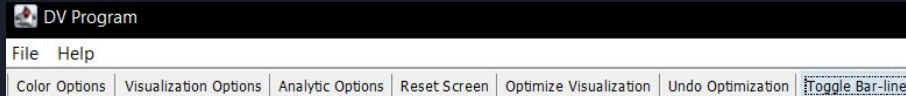


Overlap

All Data

# Interacting with Data

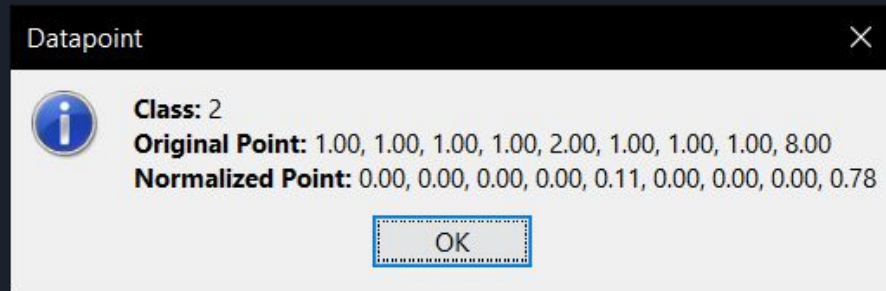
- Bar-line
  - The option is to show grouped frequency bars instead of bars for individual values when data are packed in the small area.
  - By zooming in and scaling the graph, we can get the visualization below.





# Interacting with Data

- Individual datapoints
  - Selecting the endpoint of a line will result in the datapoint's information being shown.





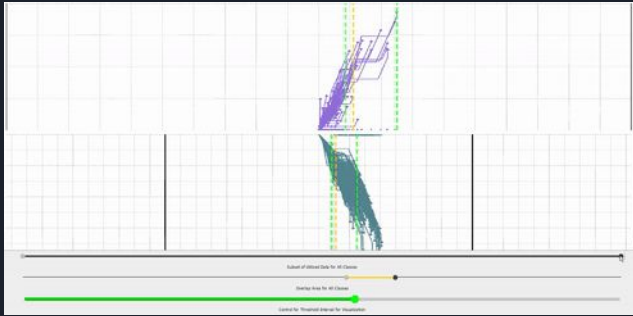
# Interacting with Data

- Sliders
  - Subset of Utilized Data:
    - changes the area of utilized data
  - Overlap Area for All Classes:
    - changes the overlap area
  - Control for Threshold Interval for Visualization:
    - Changes the threshold
  - Angles:
    - Changes the angle of the specified feature.

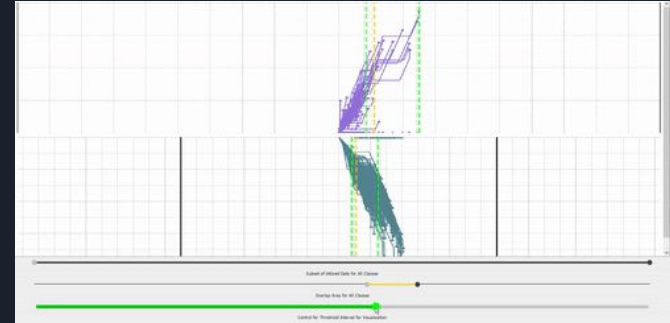
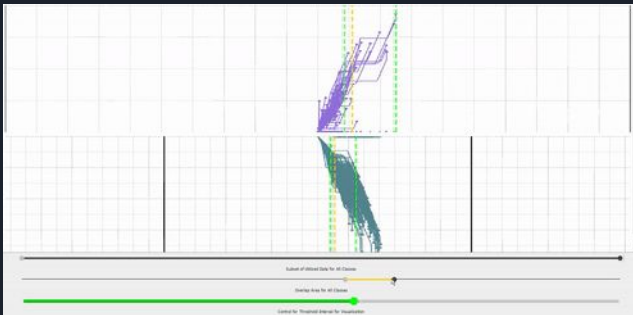
# Interacting with Data

## Slider Examples:

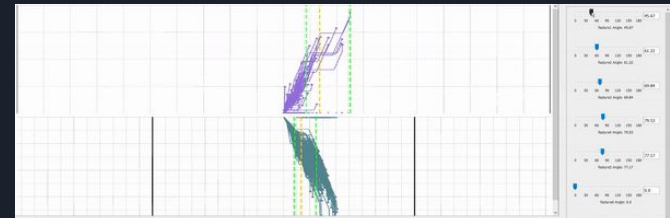
Subset of  
Utilized Data



Overlap Area  
of All Classes



Control for Threshold  
Interval for Visualization



Angles



# Analytics

Confusion Matrices for  
classifiers constructed on

- 1) All Data
- 2) Data without Overlap  
area
- 3) Overlap Data only
- 4) Worst Case Data  
validation set

All Data Analytics		
Real	Predictions	
Class	1	0
1	236	3
0	15	429
Accuracy: 97.36%		
Data Used: 100.00%		

Data Without Overlap Analytics		
Real	Predictions	
Class	1	0
1	129	1
0	1	421
Accuracy: 99.64%		
Data Used: 80.82%		

Overlap Analytics		
Real	Predictions	
Class	1	0
1	104	6
0	13	10
Accuracy: 85.71%		
Data Used: 19.47%		

Worst Case Data Analytics		
Real	Predictions	
Class	1	0
1	82	28
0	8	15
Accuracy: 72.93%		
Data Used: 19.47%		



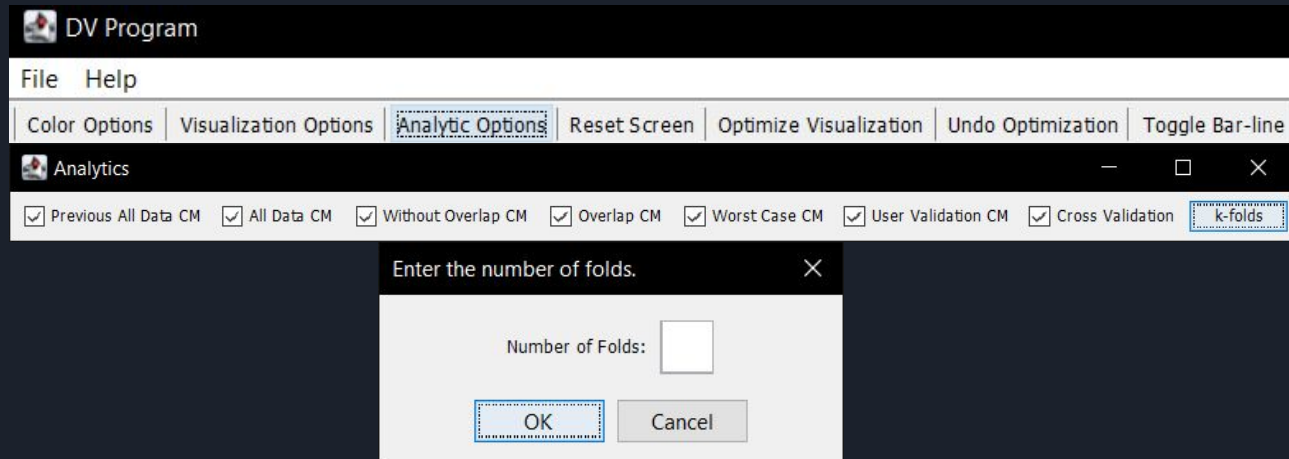
# Analytics

- k-fold Cross Validation

k-Fold Cross Validation											
Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	AVG
DT	91.30%	92.75%	94.20%	94.12%	95.59%	94.12%	92.65%	98.53%	95.59%	95.59%	94.44%
SGD	91.30%	94.20%	95.65%	94.12%	100.00%	97.06%	97.06%	98.53%	97.06%	100.00%	96.50%
NB	91.30%	95.65%	95.65%	94.12%	98.53%	95.59%	97.06%	97.06%	98.53%	97.06%	96.05%
SVM	89.86%	98.55%	95.65%	94.12%	98.53%	97.06%	97.06%	100.00%	98.53%	98.53%	96.79%
KNN	88.41%	98.55%	95.65%	94.12%	100.00%	97.06%	98.53%	100.00%	98.53%	98.53%	96.94%
LR	91.30%	92.75%	95.65%	94.12%	100.00%	97.06%	97.06%	98.53%	98.53%	100.00%	96.50%
LDA	88.41%	89.86%	97.10%	94.12%	100.00%	97.06%	97.06%	98.53%	98.53%	100.00%	96.07%
MLP	89.86%	94.20%	94.20%	94.12%	100.00%	97.06%	97.06%	100.00%	98.53%	100.00%	96.50%
RF	89.86%	92.75%	95.65%	94.12%	98.53%	98.53%	98.53%	100.00%	98.53%	98.53%	96.50%
SD	1.21%	2.85%	0.87%	0.00%	1.47%	1.23%	1.72%	1.04%	1.04%	1.55%	1.30%
AVG	90.18%	94.36%	95.49%	94.12%	99.02%	96.73%	96.90%	99.02%	98.04%	98.69%	96.25%

# Analytics to activate

- User Options:
  - Toggle on/off any analytic option
  - Choose number of folds for k-fold cross validation



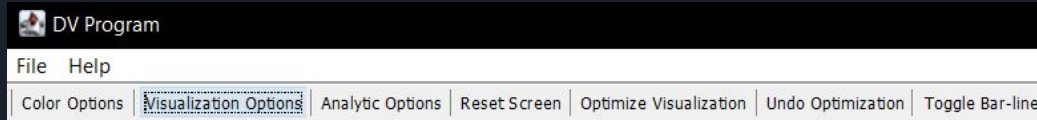


## 3+ Class Visualizations

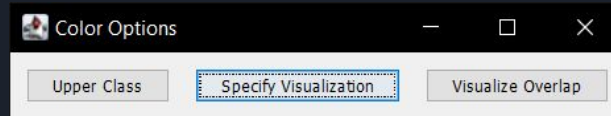
- The DV Program is only capable of visualizing two classes at once.
- For 3+ class visualizations the upper graph displays one class while the lower graph displays all others.
- Classes in the lower graph can then be removed one by one until eventually only two classes remain.

# 3+ Class Visualizations

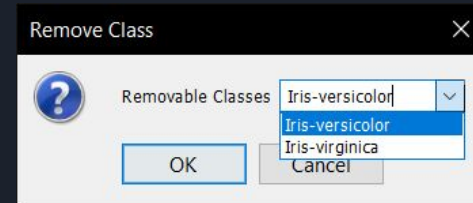
- To remove a class from the lower graph select “Visualization Options”



- Select “Specify Visualization”



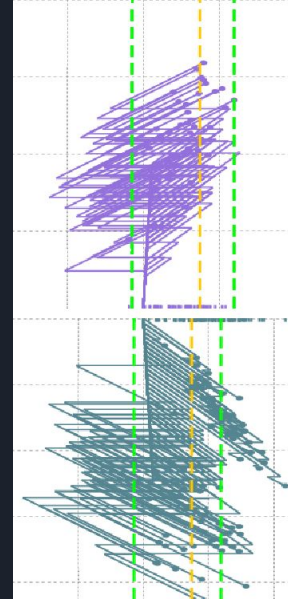
- Select the class you wish to remove



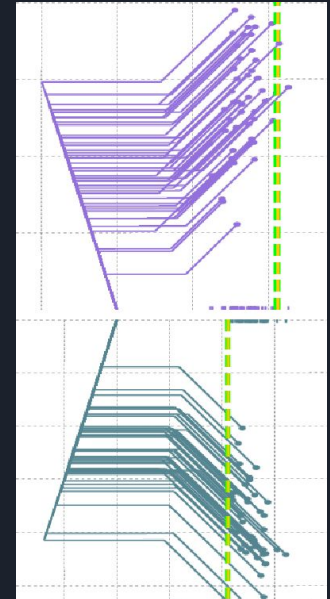


# 3+ Class Visualizations

- Initial Visualization:
  - Upper Graph: Iris-Versicolor
  - Lower Graph: Iris-Setosa & Iris-Virginica
- Iris-Setosa Class removed



Initial  
Visualization



Iris-Versicolor  
Class Removed

# 3+ Class Visualizations

- Analytics:
  - The “All Data Analytics” confusion matrix is saved when removing classes.
  - New “All Data Analytics” confusion matrix displays overall accuracy of the combined visualizations.

Iris-Versicolor  
vs  
Iris-Setosa &  
Iris-Virginica

All Data Analytics		
Real	Predictions	
Class	1	0,2
1	40	10
0,2	26	74
Accuracy: 76.00%		
Data Used: 100.00%		

Iris-Versicolor  
vs  
Iris Virginica

All Data Analytics		
Real	Predictions	
Class	1	2
1	48	2
2	1	49
Accuracy: 97.00%		
Overall Accuracy: 84.40%		
Data Used: 66.67%		

# Saving a Project

- Select “File” then “Save Project As.”
  - Or press Alt + A for a keyboard shortcut.
- Select a file location and name the project.
- To open the project simply select “Open Saved Project” and select the project save.
- Once the project save is established, simply select “Save Project” to save all subsequent edits.

