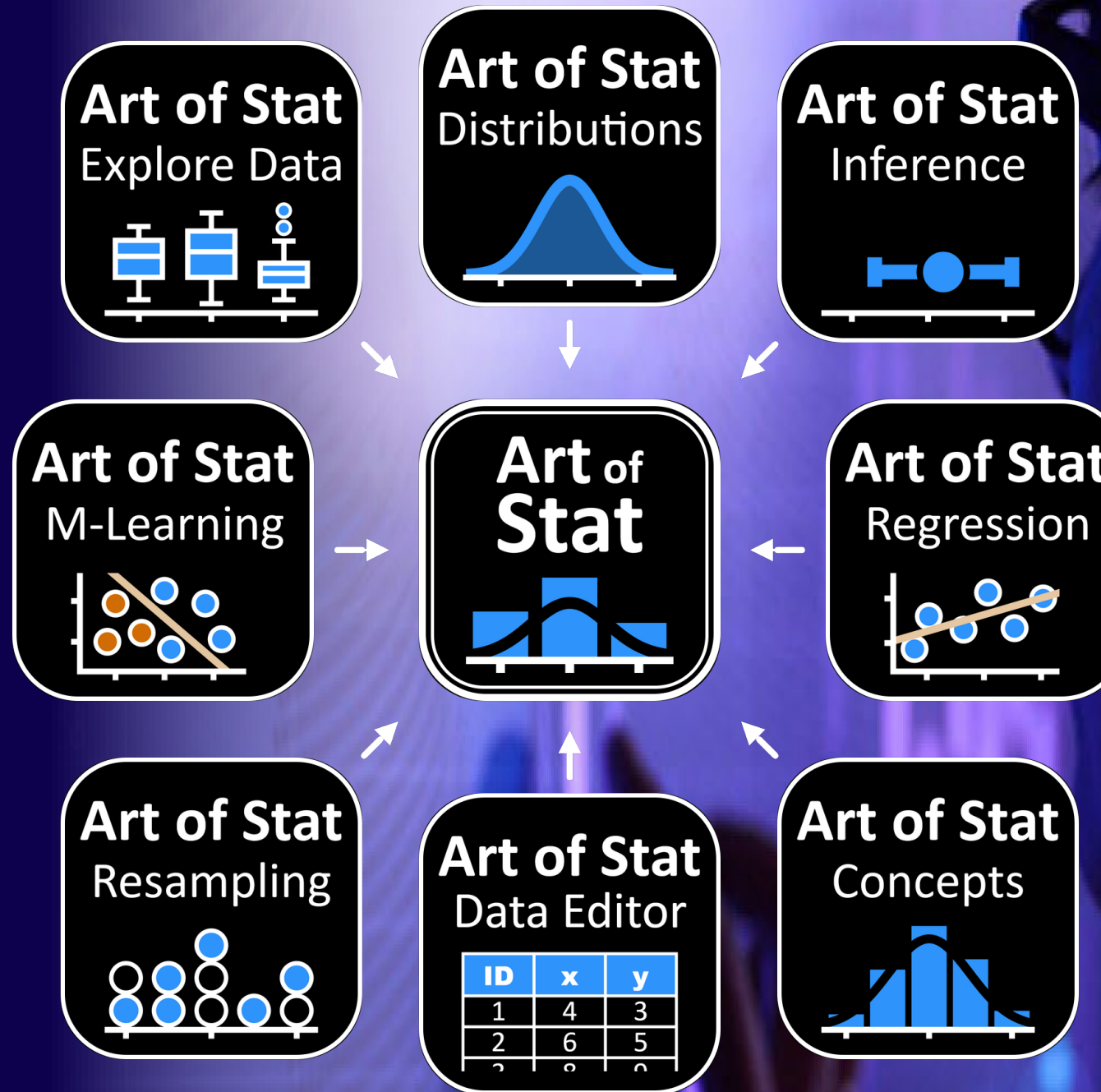


Machine Learning with No Code

Bernhard Klingenberg
Director, Masters in Applied Data Science
New College Florida

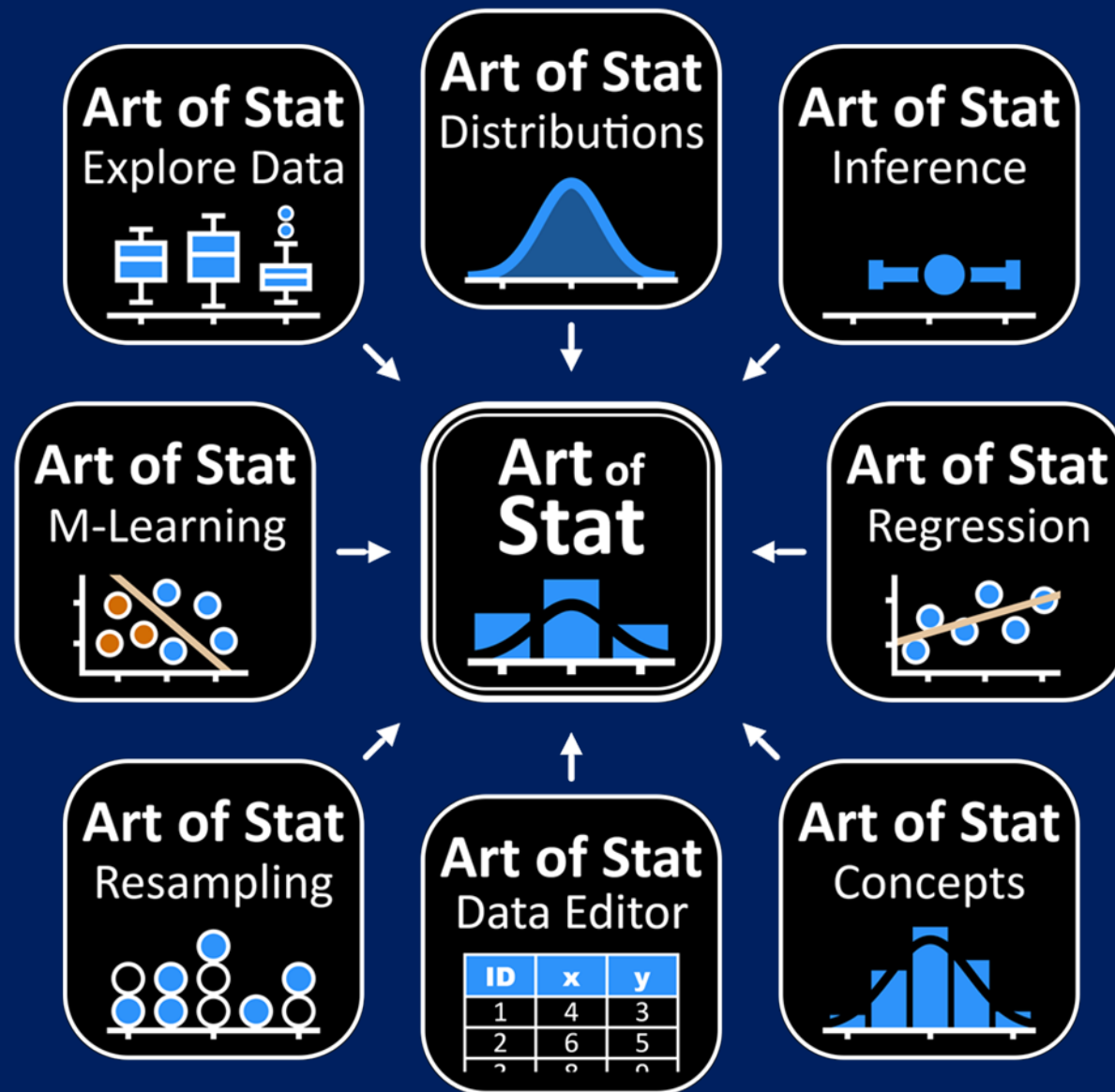
Symposium on Data Science & Statistics
SDSS 2026 | Milwaukee , WI

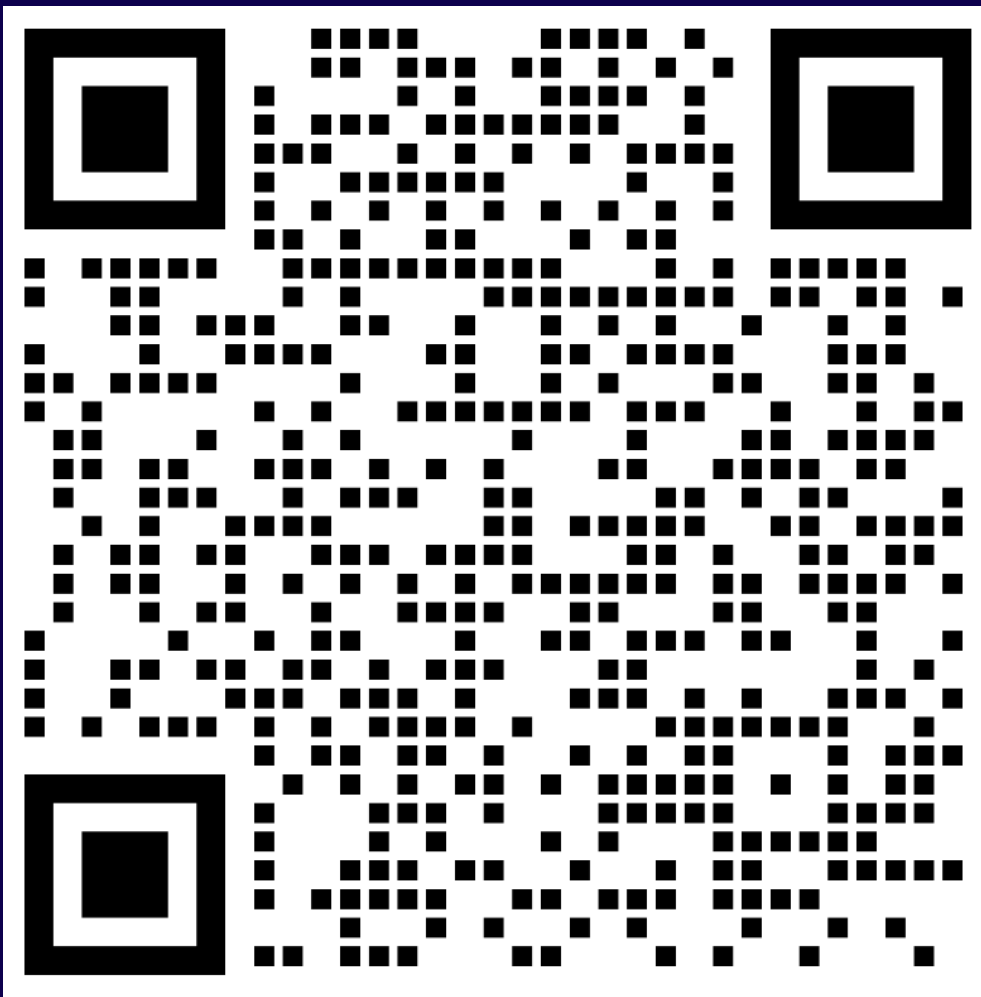


Art of Stat Mobile App



Contains 8 Modules





6:29

5G%



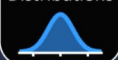
ART OF STAT

APPS FOR STATISTICS & DATA SCIENCE

Art of Stat
Explore Data

Explore Data

Perform descriptive and exploratory data analysis: Summary statistics, Barcharts, Boxplots, Scatterplots.

Art of Stat
Distributions

Distributions

Explore probability distributions and visualize their properties: Normal, t-distribution, Binomial, Chi-squared.

Art of Stat
Concepts

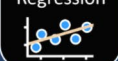
Concepts

Visualize fundamental statistical concepts: Central Limit Theorem, Coverage, Power, Correlation.

Art of Stat
Inference

Inference

Confidence intervals and hypothesis testing (P-values) for proportions and means: One- and Two Samples.

Art of Stat
Regression

Regression

Build linear or logistic regression models: Estimation, Visualization, Assessment, Prediction.

Art of Stat
M-Learning

Machine Learning

Supervised & unsupervised learning with train/test split: Linear & Logistic Regression, Naive Bayes, K-Means, ...

11:48

5G%

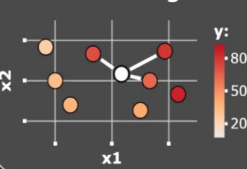
ART OF STAT
Machine Learning

Supervised Learning: Regression

Linear Regression

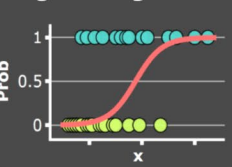
Multiple Linear
Regression

K-Nearest Neighbors

K-Nearest Neighbors
(Quantitative Resp.)

Supervised Learning: Classification

Logistic Regression

Multiple Logistic
Regression

Naive Bayes



Naive Bayes Classifier

Linear LDA



Linear Discriminant

Quadratic LDA



Quadratic Discriminant

5:10

5G%



Naive Bayes



Data Source:

Open Example Datasets

Choose from several example datasets. Then, select the response (=target) variable which provides the labels and the explanatory variables (=features) you want to include in the naive bayes classification.

Select Dataset:

Palmer Penguins

Penguin	Species	Island	Bill Length (mm)
1	Adelie	Torgersen	39.1
2	Adelie	Torgersen	39.5
3	Adelie	Torgersen	40.3
4	Adelie	Torgersen	36.7
5	Adelie	Torgersen	39.3

Showing all 342 rows, and all 8 columns (swipe left).

Response Variable (Class Labels):

Species

Two Features Selected:

Bill Length (mm)

Flipper Length (mm)

Split into Train and Test Data

ENTER
DATAPLOTS AND
CLASSIFICATIONSACCURACY &
PREDICTIONS

5:12

5G%



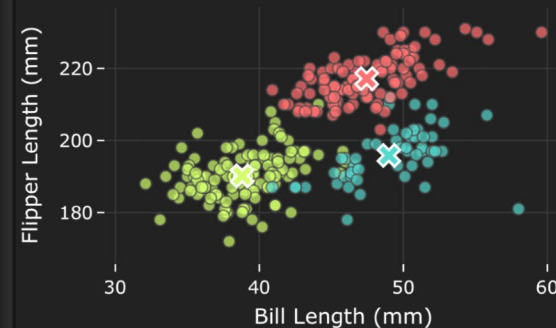
Naive Bayes



Plot of Training Data (Observed Labels)

Species:

● Adelie ● Chinstrap ● Gentoo



Jitter Points



Plot Test Data



Predicted Labels



Posterior Probabilities



Heatmap of Posterior Probabilities



Class Means



Normal Contours



Prior Class Probabilities:

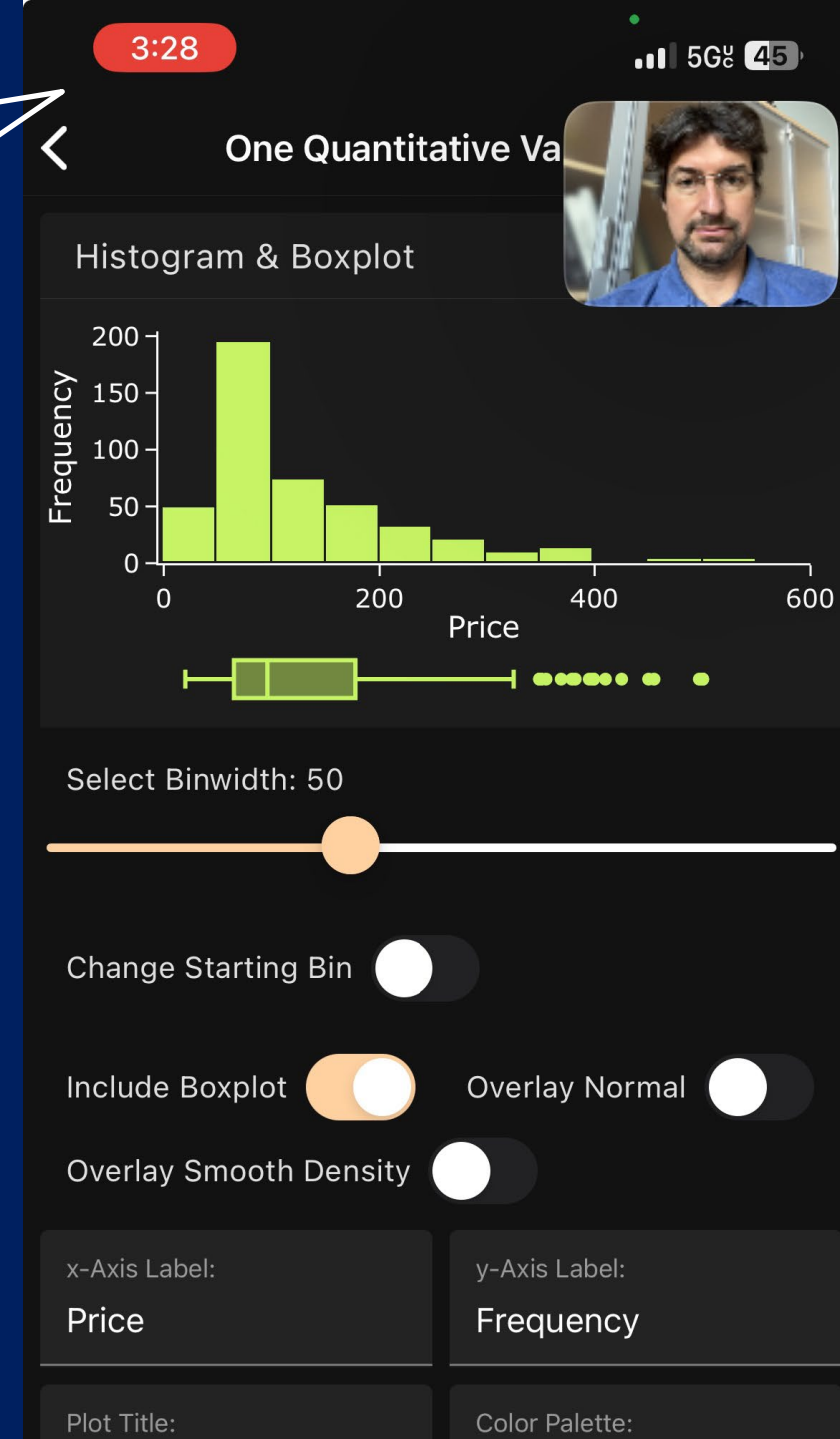
Species	Prior Probabilities
Adelie	43.8%
Chinstrap	19.3%

ENTER
DATAPLOTS AND
CLASSIFICATIONSACCURACY &
PREDICTIONS

Technology

Me in a Zoom call sharing my screen with the class

- Connect Phone/Tablet to Screen
- Share Screenshots, Join Zoom Calls
- Screen Recording (Videos)
- Upload your own data (.csv, .xls)
- Available 24/7, no lab needed
- Work Off-Line (Airplane Mode)



Art of Stat: Machine Learning

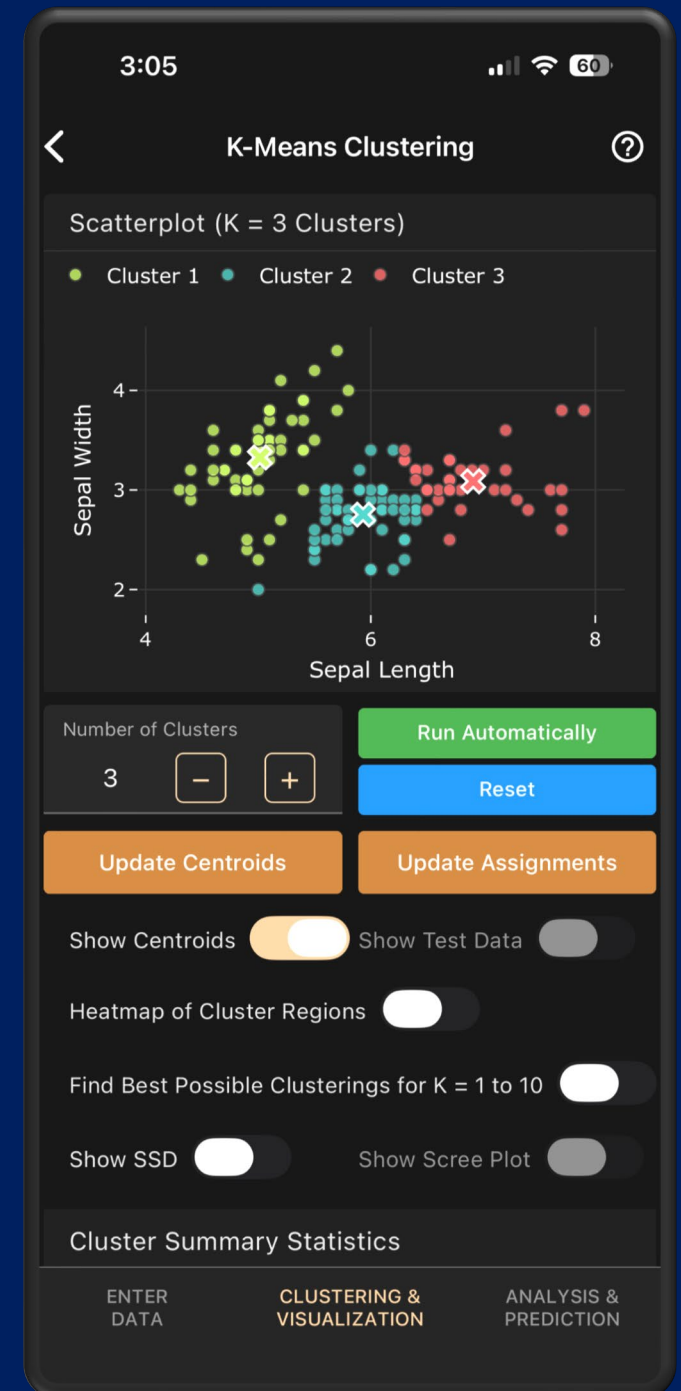
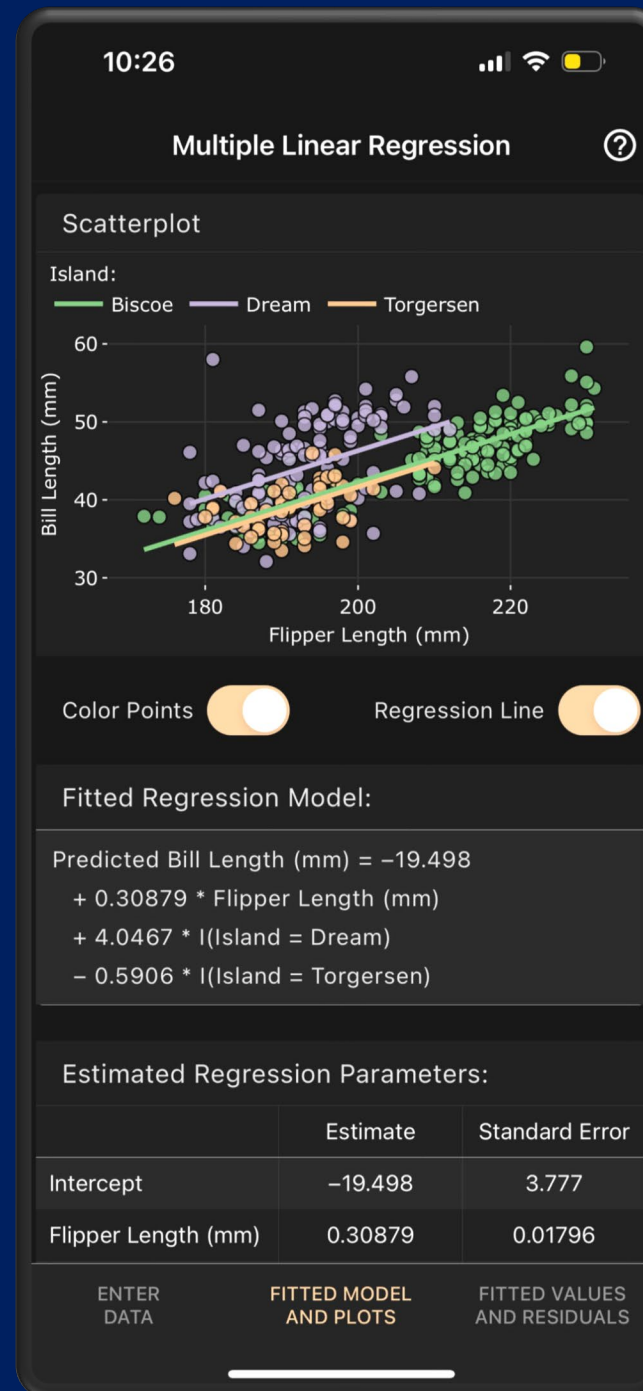


Supervised Learning

- Regression (Linear, Logistic)
- Discriminant Analysis
- Naïve Bayes
- K-Nearest Neighbors
- Trees, Random Forests

Unsupervised Learning

- Principle Components
- K-Means Clustering



Naïve Bayes

- Classification method, trying to predict class membership based on some features (supervised learning)
- Uses Bayes Theorem and “naïve” independence assumption about joint distribution of features x_1 and x_2
- Finds **Posterior Probability** for observation falling in class k :

$$\begin{aligned} P(\text{Class } k \mid x_1, x_2) &\propto P(x_1, x_2 \mid \text{Class } k) * P(\text{Class } k) \\ &= P(x_1 \mid \text{Class } k) * P(x_2 \mid \text{Class } k) * P(\text{Class } k) \end{aligned}$$

- Need **prior** class probabilities $P(\text{Class } k)$ and distribution of feature in a given class $P(x_1 \mid \text{Class } k)$
- Predicted class is the one with largest posterior probability

Example: Naïve Bayes

Tab 1:

- Load Data
- Select Target
- Select Predictors
- Train/Test Split
- Descriptive Stats

Naive Bayes

Data Source:
Open Example Datasets

Choose from several example datasets. Then, select the response (=target) variable which provides the labels and the explanatory variables (=features) you want to include in the naive bayes classification.

Select Dataset:
Palmer Penguins

Penguin	Species	Island	Bill Length (mm)
1	Adelie	Torgersen	39.1
2	Adelie	Torgersen	39.5
3	Adelie	Torgersen	40.3
4	Adelie	Torgersen	36.7
5	Adelie	Torgersen	39.3

Showing all 342 rows, and all 8 columns (swipe left).

Response Variable (Class Labels):
Species

Two Features Selected:

Bill Length (mm) Flipper Length (mm)

Split into Train and Test Data ☐

ENTER DATA PLOTS AND CLASSIFICATIONS ACCURACY & PREDICTIONS

Pick a Data Source:

- Pre- Implemented Examples
- Upload your own CSV/Excel
- Data Editor

Pick the particular dataset

Scroll through the dataset

Pick target variable

Select Features:

- continuous
- categorical

Split into testing & training set

Tab 1

Example: Naïve Bayes

Tab 1:

- Load Data
- Select Target
- Select Predictors
- Train/Test Split
- Descriptive Stats

Naive Bayes

Split into Train and Test Data

Percent Training Samples: 80%

Generate Split

Frequency Table of Labels (Training Data):

Species	Count	Percent
Adelie	120	43.8%
Chinstrap	53	19.3%
Gentoo	101	36.9%
Total	274	100.0%

Frequency Table of Labels (Test Data):

Species	Count	Percent
Adelie	31	45.6%
Chinstrap	15	22.1%
Gentoo	22	32.4%
Total	68	100.0%

Class Means & Standard Deviations for Continuous Features (Training Set):

Species	Sample Size	Mean x1	Std. Dev. x1
---------	-------------	---------	--------------

ENTER DATA

PLOTS AND CLASSIFICATIONS

ACCURACY & PREDICTIONS

Generate 80/20 split by pressing button

See distribution of labels in training set

See distribution of labels in test set

Tab 1

Example: Naïve Bayes

Tab 1:

- Load Data
- Select Target
- Select Predictors
- Train/Test Split
- Descriptive Stats

Naive Bayes

Class Means & Standard Deviations for Continuous Features (Training Set):

Species	Sample Size	Mean x1	Std. Dev. x1
Adelie	120	38.8	2.8
Chinstrap	53	49.0	3.3
Gentoo	101	47.4	3.2

x1 = Bill Length (mm), x2 = Flipper Length (mm)

Show Overall Descriptive Statistics

Descriptive Statistics of Features (Training Data):

	Bill Length (mm)	Flipper Length (mm)
Sample Size (n)	274	274
Mean ($\bar{x}$)	44.0	201.2
Standard Deviation (s)	5.5	14.0
Minimum	32.1	172.0
First Quartile (Q1)	39.2	190.0
Median	44.5	197.0

ENTER DATA

PLOTS AND CLASSIFICATIONS

ACCURACY & PREDICTIONS

Mean and standard deviation for each continuous predictor, stratified by label.

If we had categorical predictors (e.g., Sex or Island), app would show distribution of categories for each label.

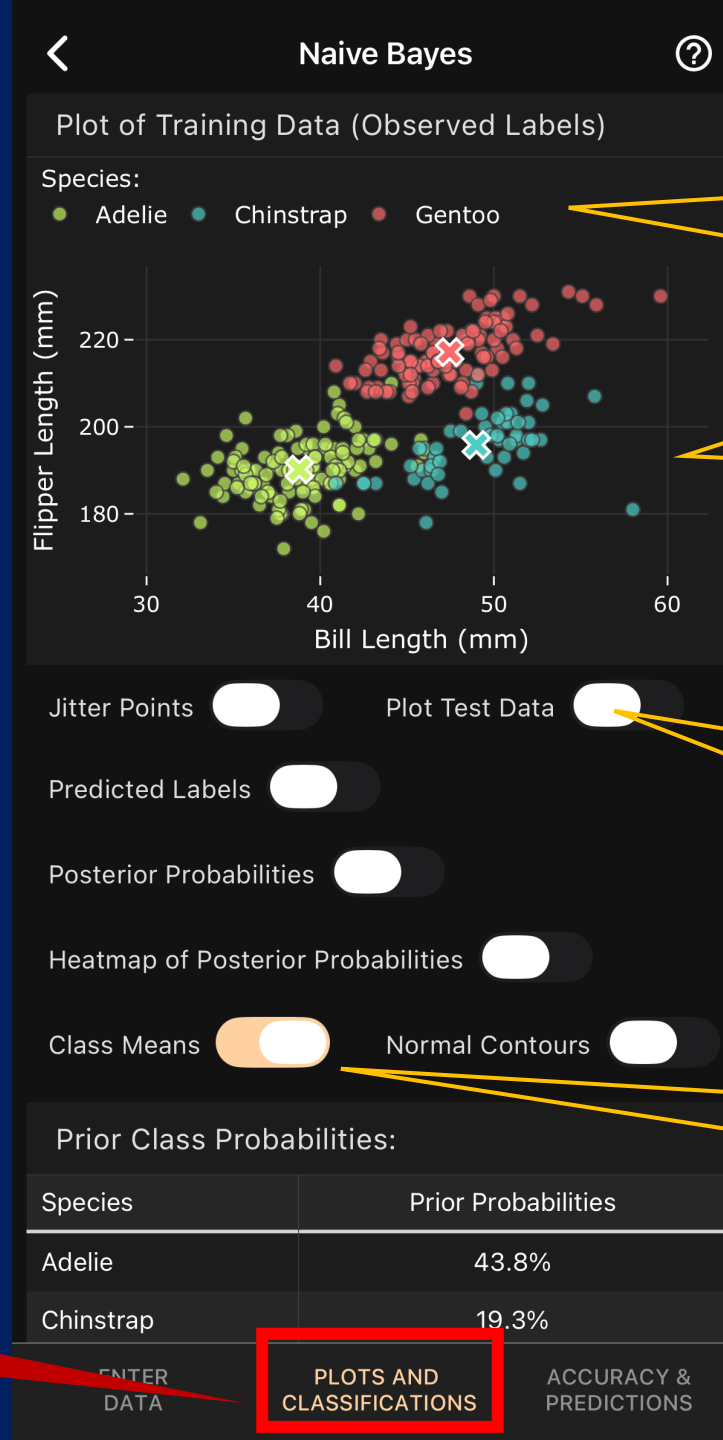
See descriptive stats for each continuous predictor (not stratified)

Tab 1

Example: Naïve Bayes

Tab 2:

- Scatterplot
- Class Means
- Posterior Odds
- Predictions
- Heatmap
- Contours



Different target labels are represented by different colors (plot interactive)

Scatterplot with the two features on the x - and y-axis

Request to plot test data instead of training data

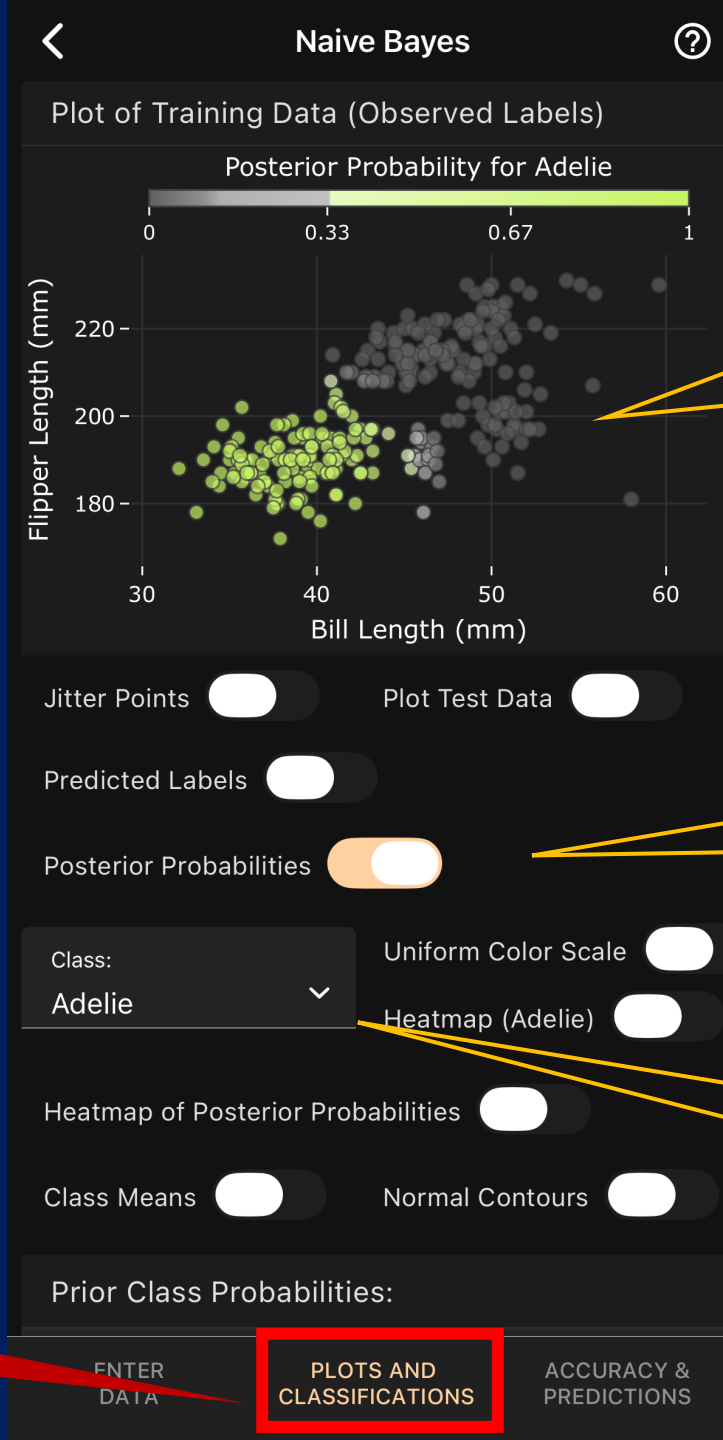
Show the mean for each target label

Tab 2

Example: Naïve Bayes

Tab 2:

- Scatterplot
- Class Means
- **Posterior Odds**
- Predictions
- Heatmap
- Contours



Dots colored by their posterior probability of belonging to species "Adelie"

Request to show posterior odds of class membership

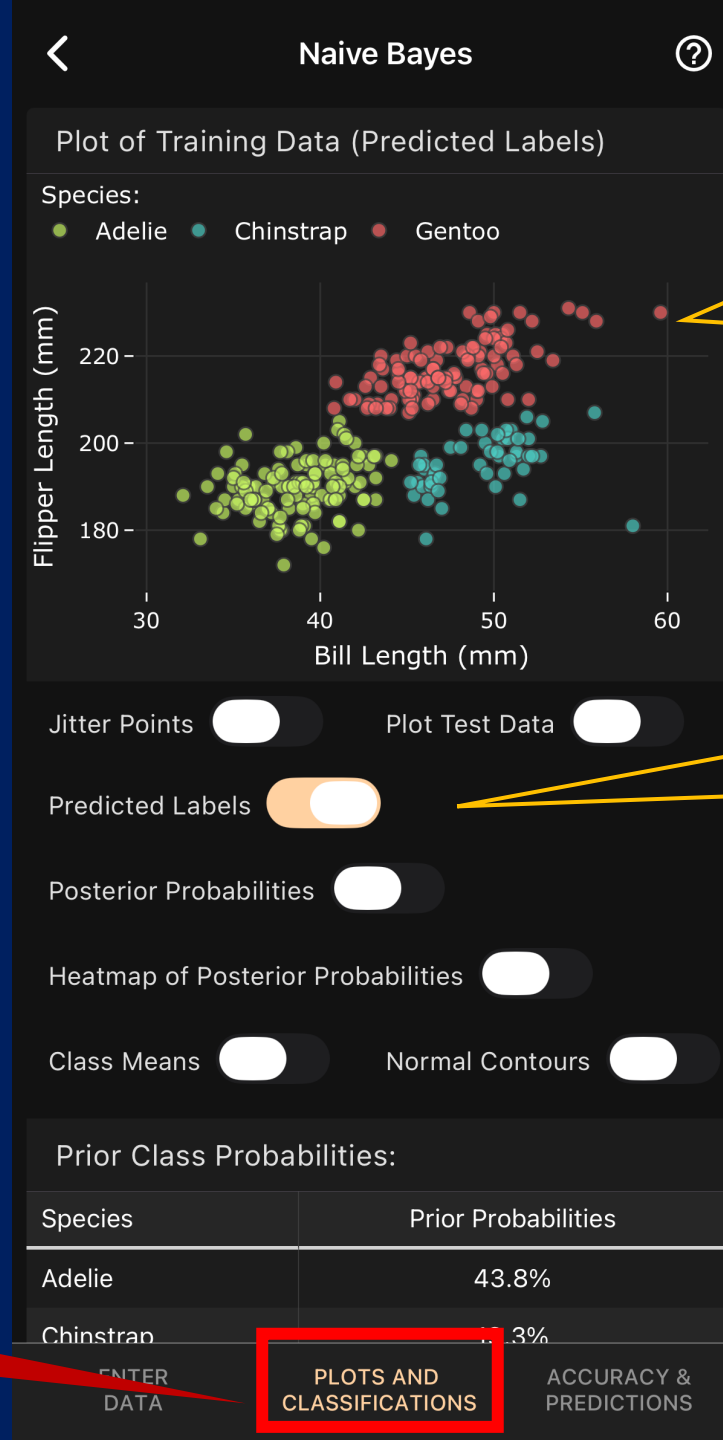
Select which class you want to show posterior odds for

Tab 2

Example: Naïve Bayes

Tab 2:

- Scatterplot
- Class Means
- Posterior Odds
- Predictions
- Heatmap
- Contours



Scatterplot with points colored by predicted label (training set)

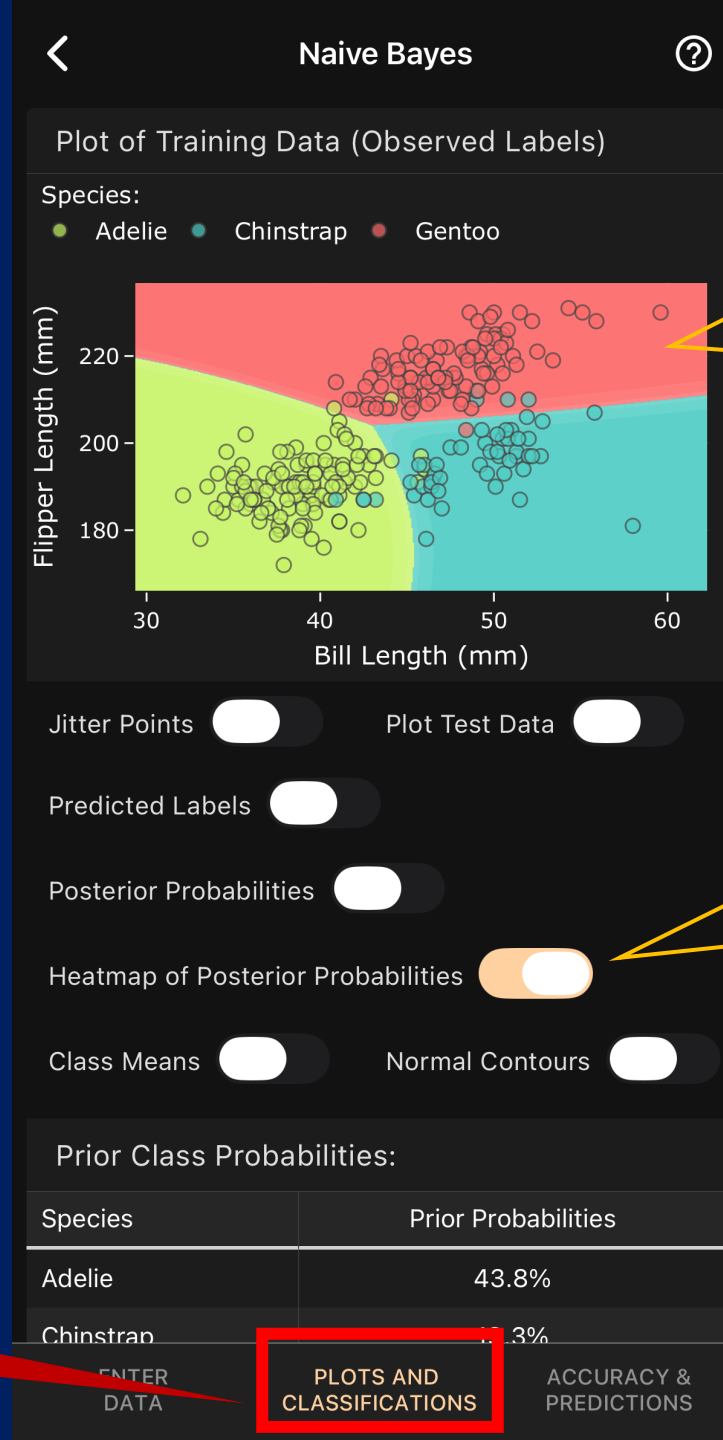
Switch to show the predicted label, based on maximum posterior odds

Tab 2

Example: Naïve Bayes

Tab 2:

- Scatterplot
- Class Means
- Posterior Odds
- Predictions
- **Heatmap**
- Contours



Scatterplot with points colored by original label but area colored by predicted label. (Training set)

Switch to show heatmap of posterior odds. (Works when there are only continuous features.)

Tab 2

Example: Naïve Bayes

Tab 3:

- Dataset with predictions, posterior probs and odds (not shown)
- Find Prediction
- Confusion Matrix
- Performance

Naive Bayes

3	40.3	195	Adelie
4	26.7	102	Adelie

Scroll left to reveal more columns

Download Dataset of Predictions

Find Predicted Class for New Observation

Select Values for Features:

Bill Length (mm)
40

Flipper Length (mm)
200

Predicted Class: Adelie

Prior and Posterior Probabilities:

Species	Prior Probability	Posterior Probability
Adelie	43.8%	97.3%
Chinstrap	19.3%	2.2%
Gentoo	36.9%	0.4%

Confusion Matrix (Counts):

	Predicted Class			
Observed Class	Adelie	Chinstrap	Gentoo	Total
Adelie	195	10	5	210
Chinstrap	5	10	5	20
Gentoo	5	5	10	20
Total	210	25	15	250

ENTER DATA

PLOTS AND CLASSIFICATIONS

ACCURACY & PREDICTIONS

Dataset (train and test) with posterior probabilities and odds (only shown partially in this screenshot)

Ask for class prediction of new observation

Enter values for features

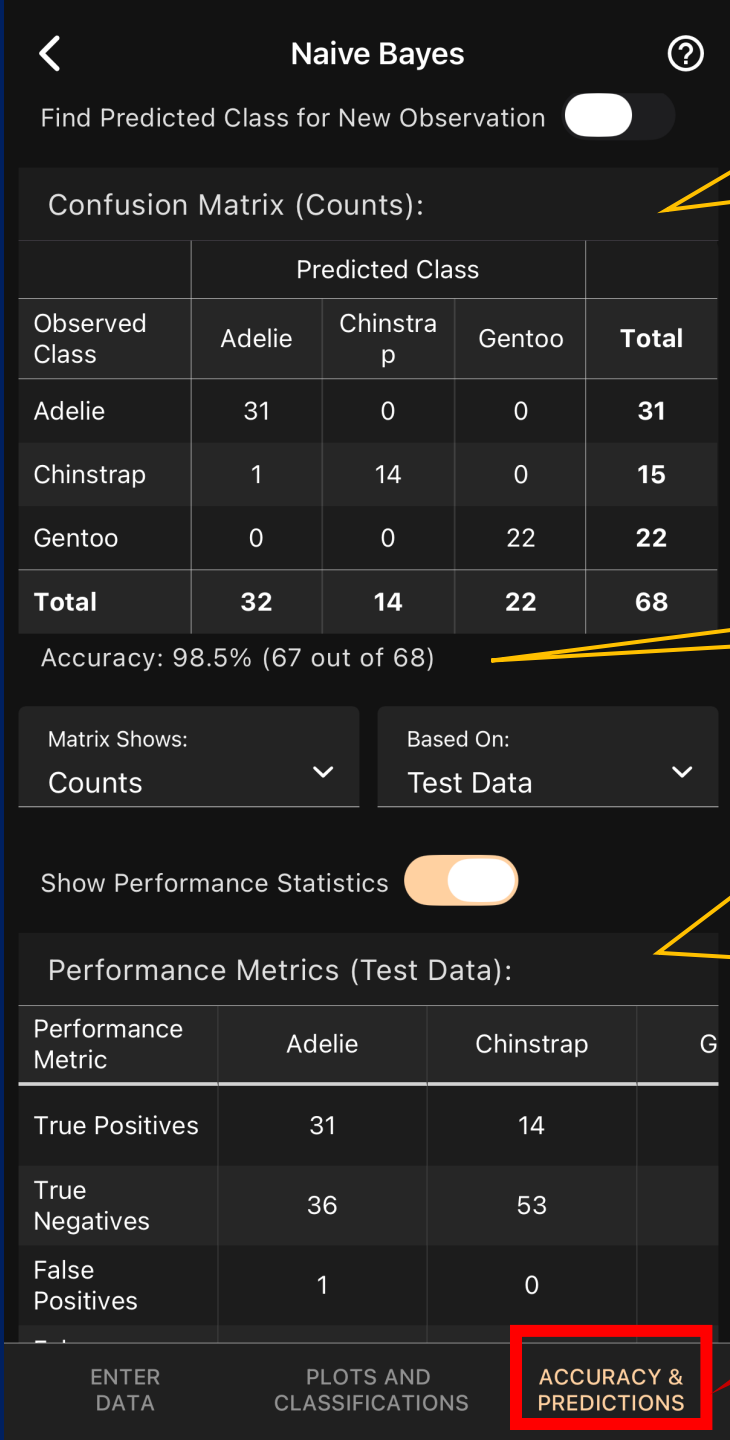
Table showing, for each class, the posterior probabilities for each class, based on the supplied feature values

Tab 3

Example: Naïve Bayes

Tab 3:

- Dataset with predictions, posterior probs. and odds
- Find Prediction
- Confusion Matrix
- Performance



3x3 Confusion Matrix comparing observed and predicted labels

Overall test accuracy

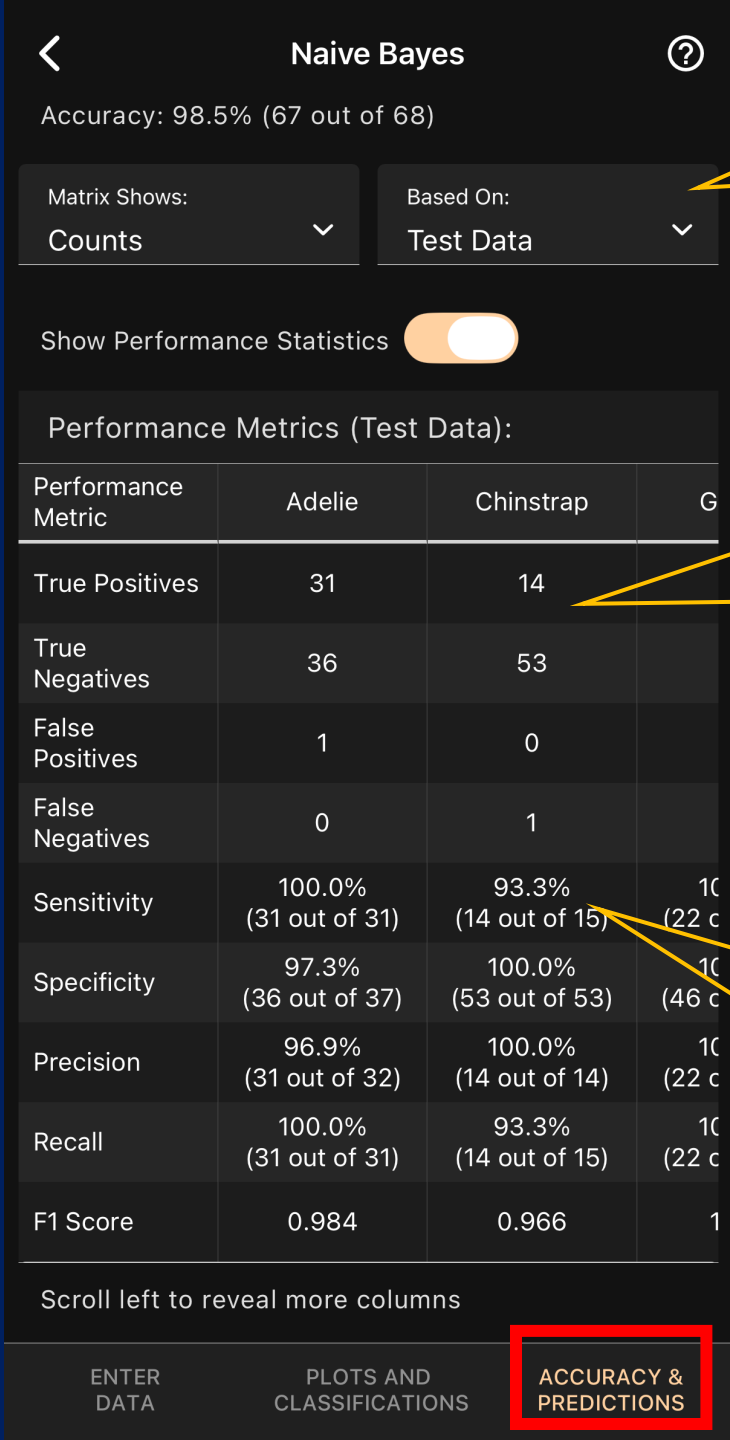
Table showing, for each class, true positives, true negatives, false positives and false negatives, sensitivity, specificity, precision, recall and F1 score.

Tab 3

Example: Naïve Bayes

Tab 3:

- Dataset with predictions, posterior probs. and odds
- Find Prediction
- Confusion Matrix
- Performance



Requested performance statistics for test data

Table showing, for each class, true positives, true negatives, false positives and false negatives.

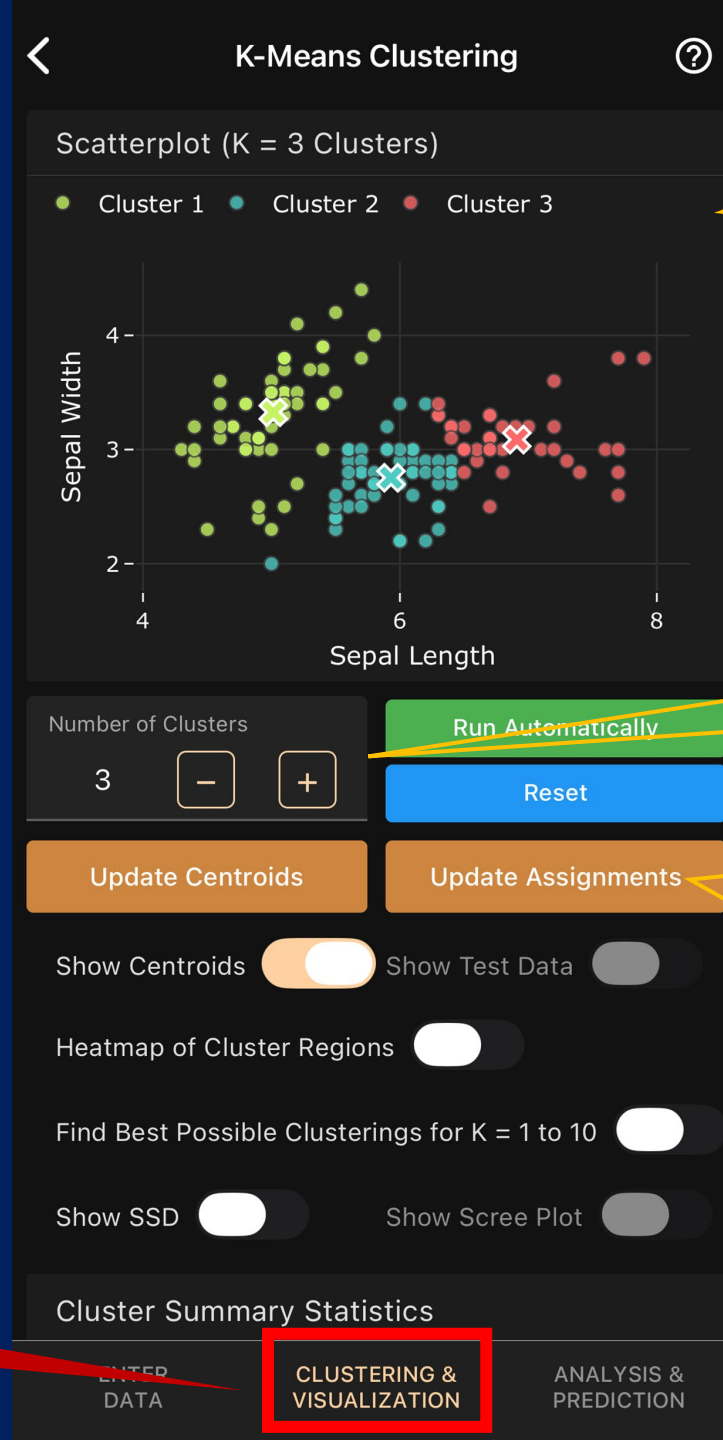
Also showing resulting sensitivity, specificity, precision, recall and F1 score.

Third tab

Example: K-Means Clustering

Tab 2:

- Scatterplot
- Select K
- Carry out clustering
- Heatmap
- Screeplot
- Cluster Summaries



Scatterplot of two features to be clustered on

Select the number of clusters (K)

Step through clustering step by - step, iterating between finding centroids and then updating assignment.

Tab 2

Thank you!
And please download
and try the
Art of Stat app!



Slides available at ArtofStat.com
