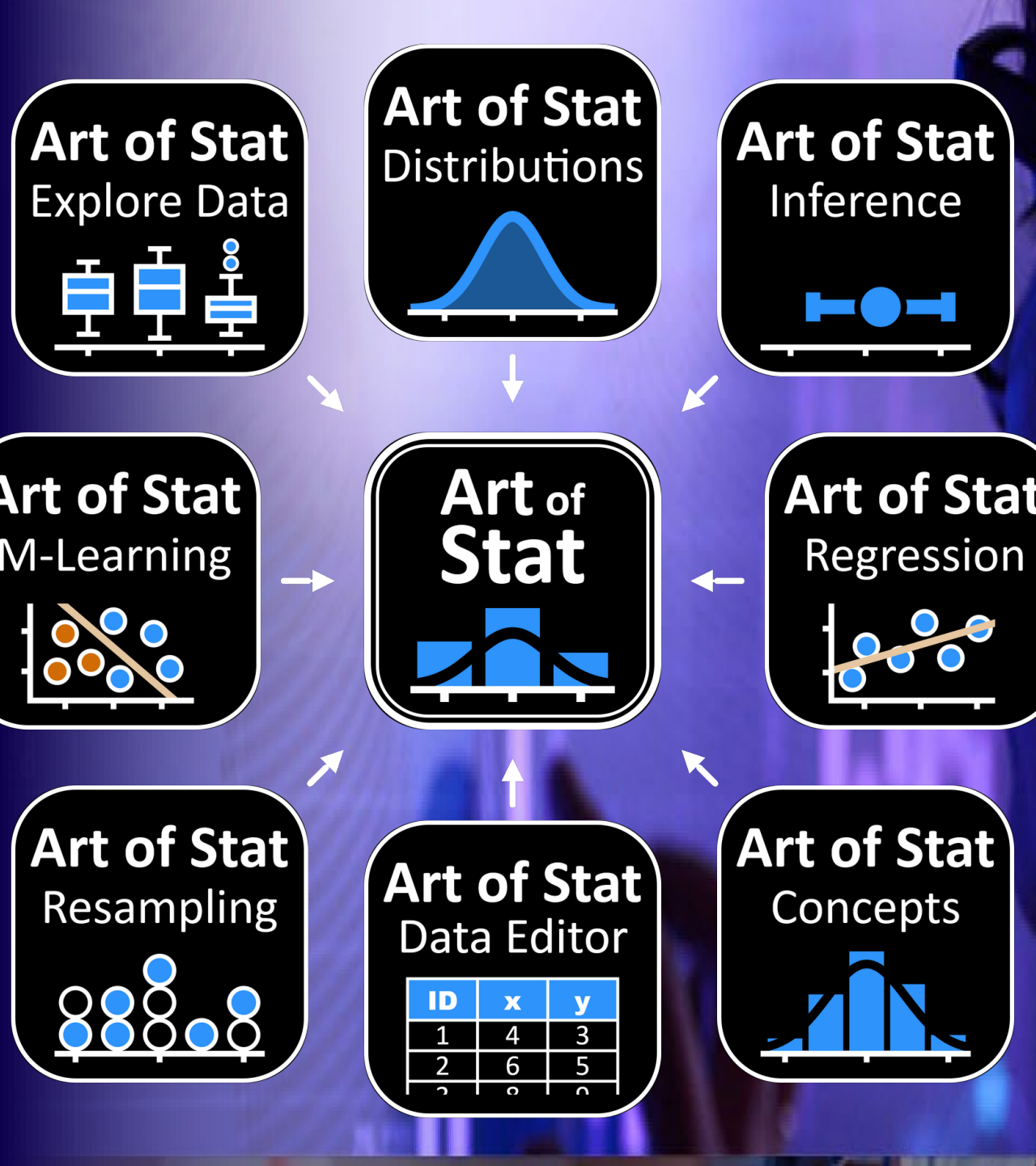
Machine Learning with No Code

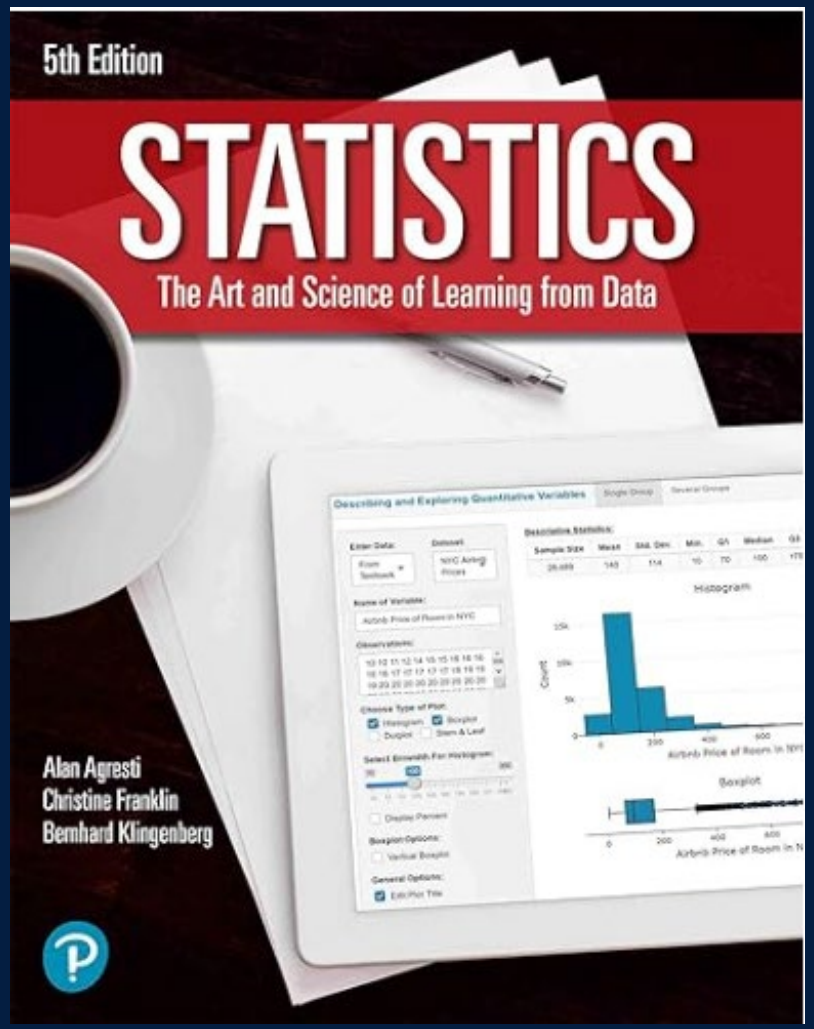
Bernhard Klingenberg
Director, Masters in Applied Data Science
New College Florida

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



Introduction

- Ph.D. in Statistics (@UF in 2004)
- Worked at Williams College for 20 years
- Research in Categorical Data Analysis, ML for Solar Flare identification
- Co-author of: “Statistics, the Art & Science of Learning from Data” (with Alan Agresti, Chris Franklin, Pearson 2021)
- Online and Mobile Apps: “Art of Stat”



Art of Stat Distributions

Art of Stat Explore Data

Art of Stat Inference

Art of Stat Regression

Art of Stat Concepts

Art of Stat Resampling

Art of Stat M-Learning

Download on the App Store

GET IT ON Google Play

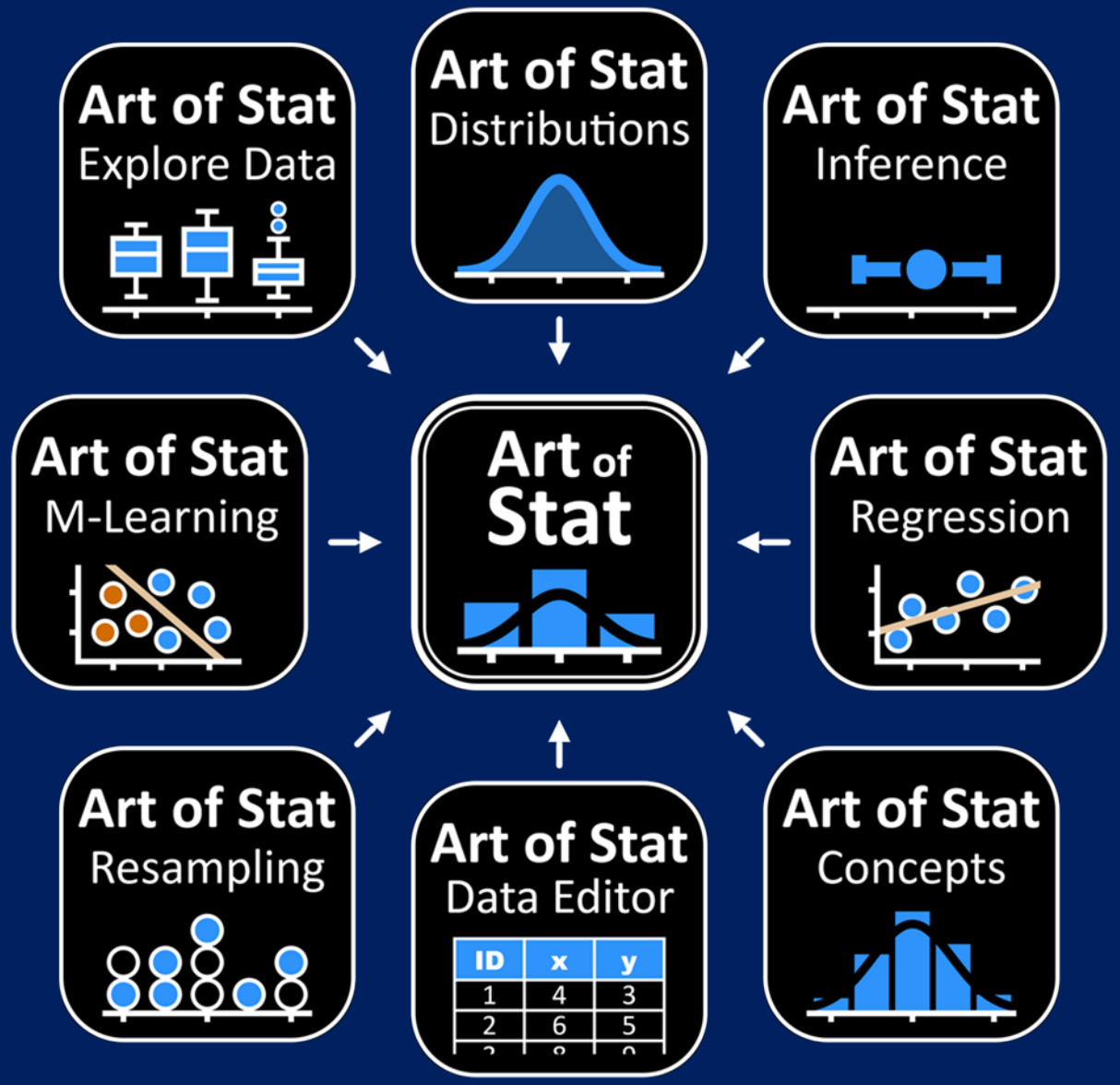
Single Art of Stat Mobile App

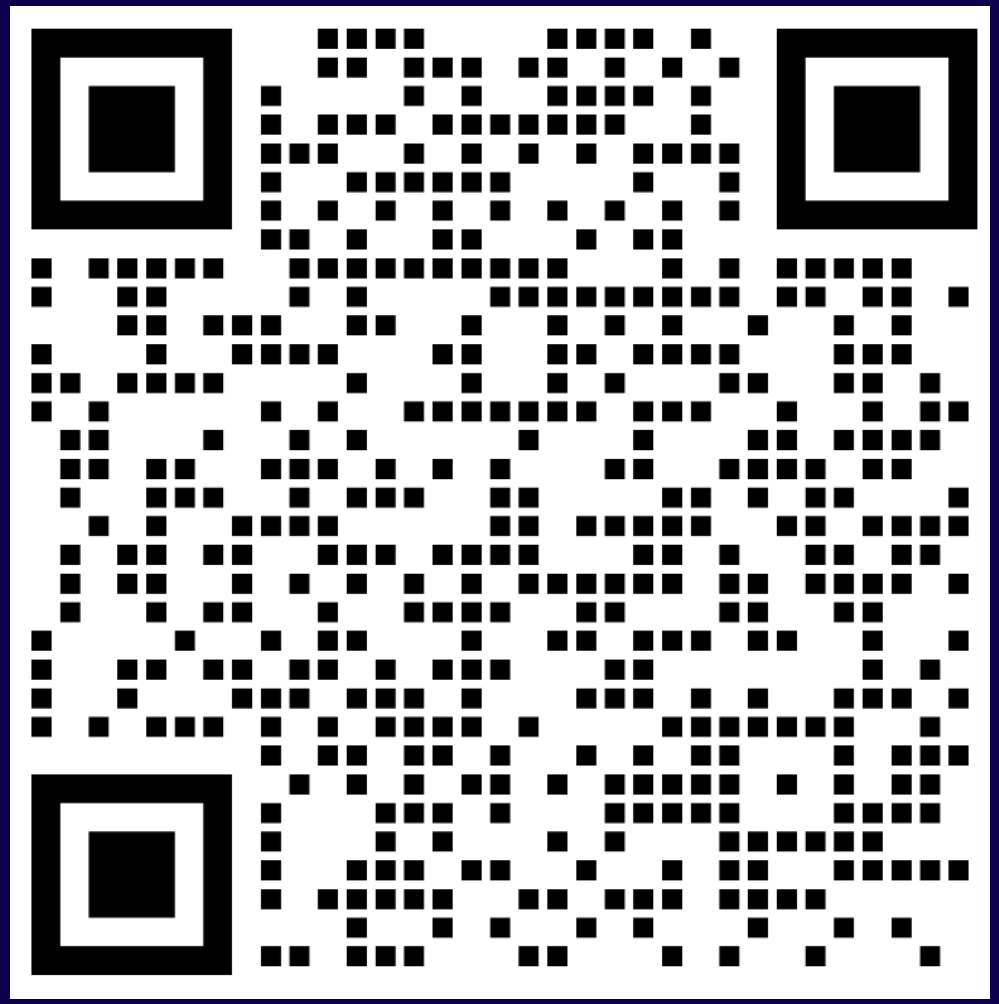
Art of Stat

Download on the App Store

GET IT ON Google Play

Contains 8 Modules





Download on the App Store

GET IT ON
Google Play

6:29 5G 36

ART OF STAT

APPS FOR STATISTICS & DATA SCIENCE

Explore Data

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

Distributions

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

Concepts

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

Inference

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

Regression

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

Machine Learning

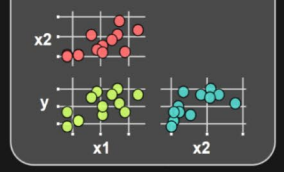
Supervised & unsupervised learning with train/test split: Linear & Logistic Regession, Naive Bayes, K-Means, ...

4:26 5G 64

ART OF STAT
Machine Learning

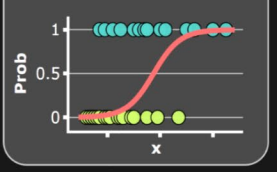
Supervised Learning

Linear Regression



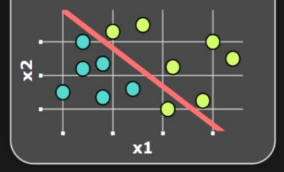
Multiple Linear Regression

Logistic Regression



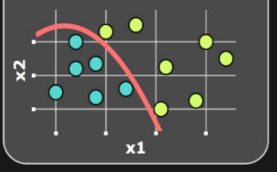
Multiple Logistic Regression

Linear LDA



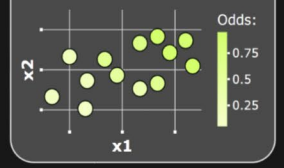
Linear Discriminant Analysis

Quadratic LDA



Quadratic Discriminant Analysis

Naive Bayes



Naive Bayes Classifier

5:10 5G 61

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)

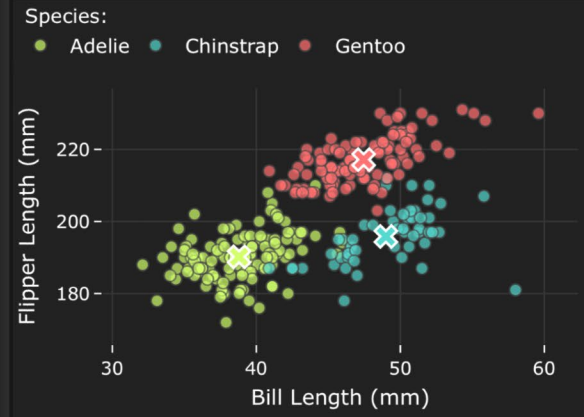
Split into Train and Test Data

ENTER DATA | PLOTS AND CLASSIFICATIONS | ACCURACY & PREDICTIONS

5:12 5G 61

Naive Bayes

Plot of Training Data (Observed Labels)



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%

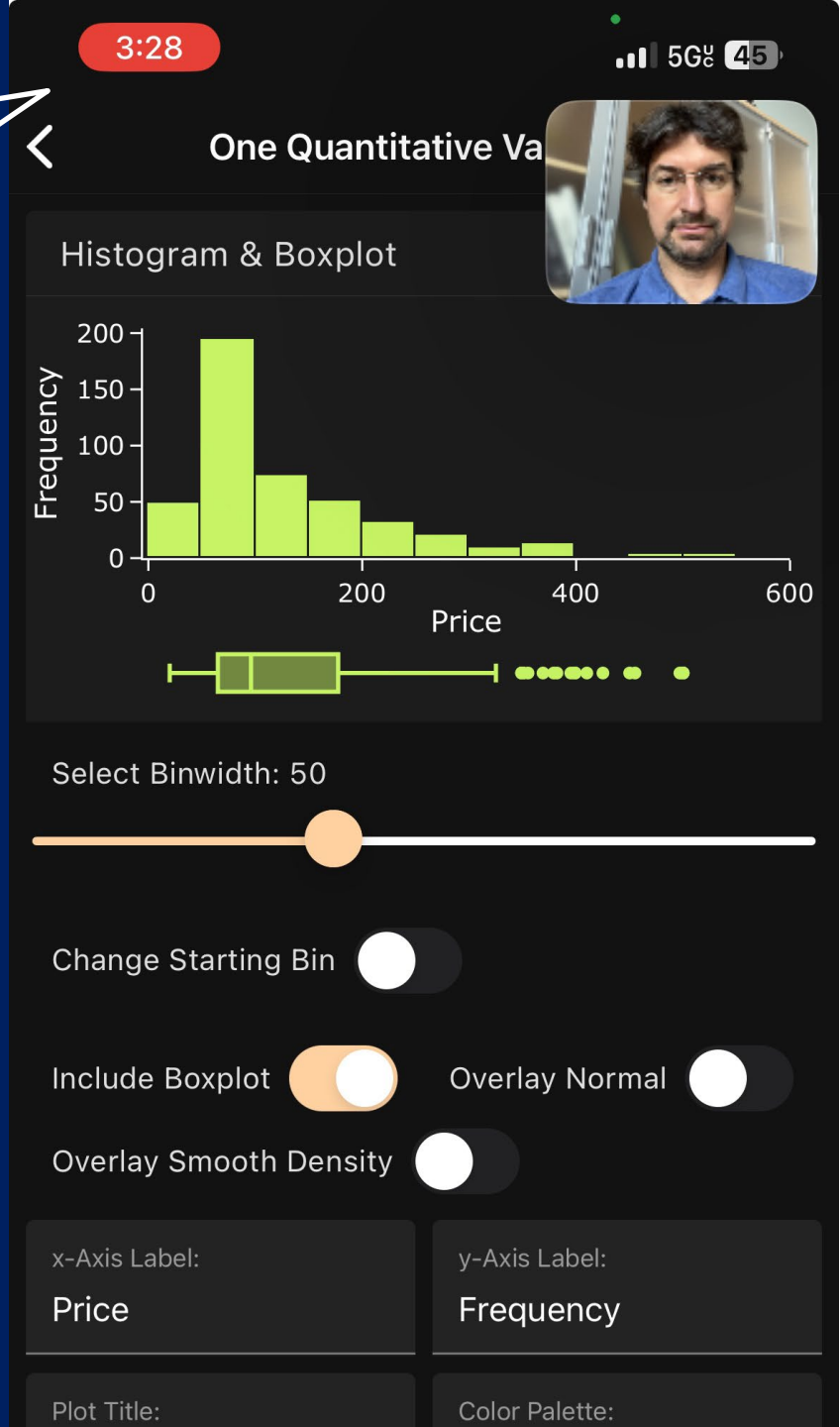
ENTER DATA | PLOTS AND CLASSIFICATIONS | ACCURACY & PREDICTIONS

Copyright 2025 P.Bruce. All rights reserved.

Technology

- 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)

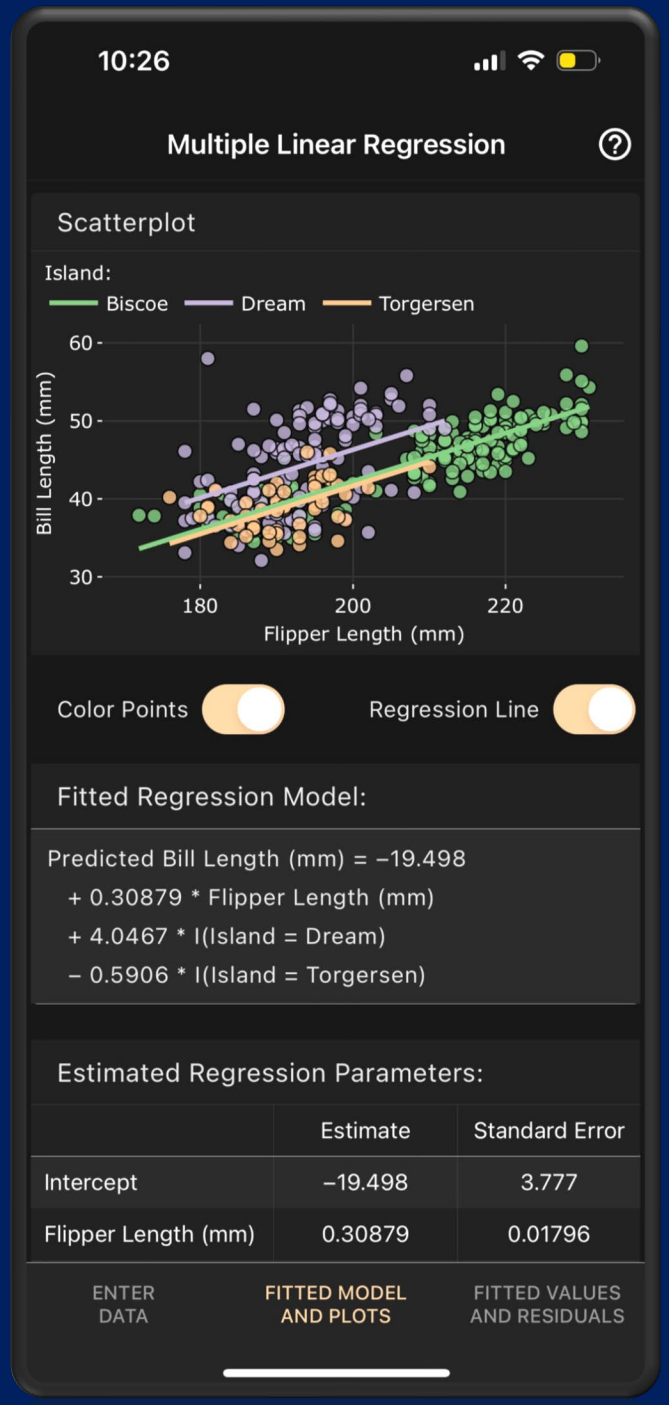
Me in a Zoom call sharing my screen with the class

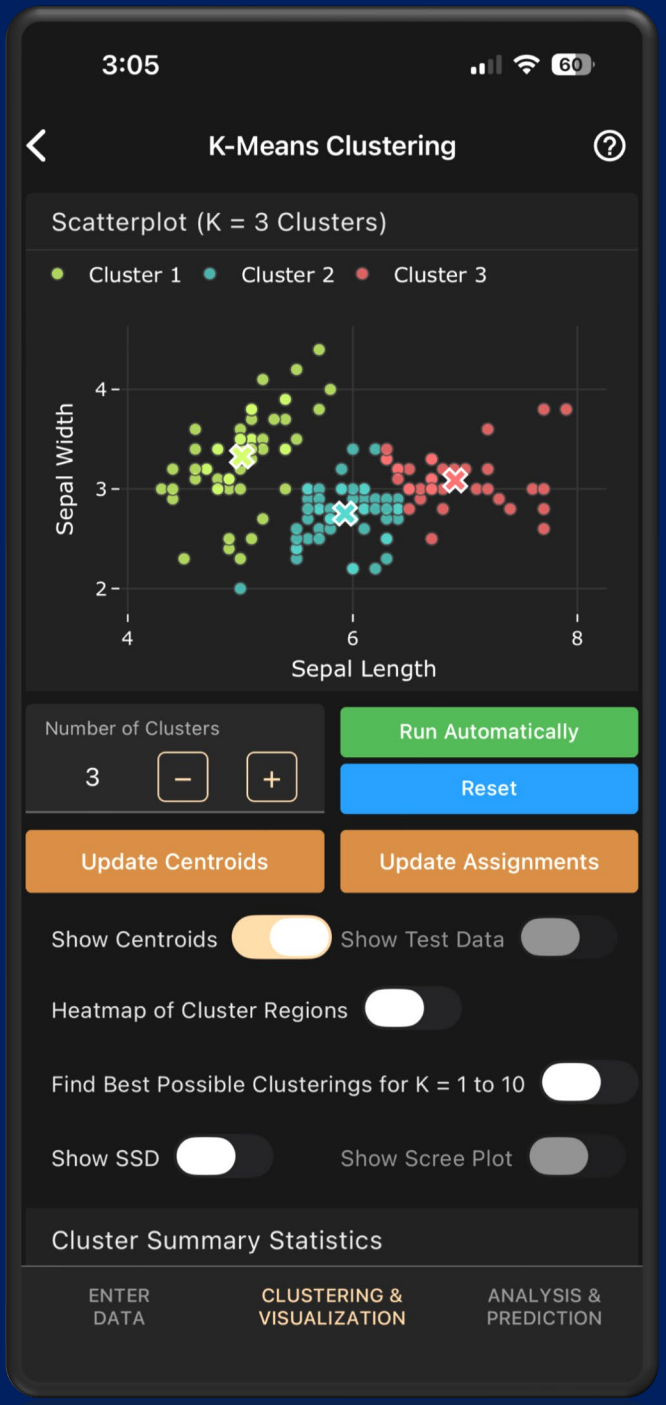


Art of Stat: Machine Learning

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

- **Unsupervised Learning**
  - 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 x1 and x2
- Finds Posterior Probability for observation falling in class k:

$$P(Class\ k \mid x1, x2) \propto P(x1, x2 \mid Class\ k) * P(Class\ k)$$
$$= P(x1 \mid Class\ k) * P(x2 \mid Class\ k) * P(Class\ k)$$

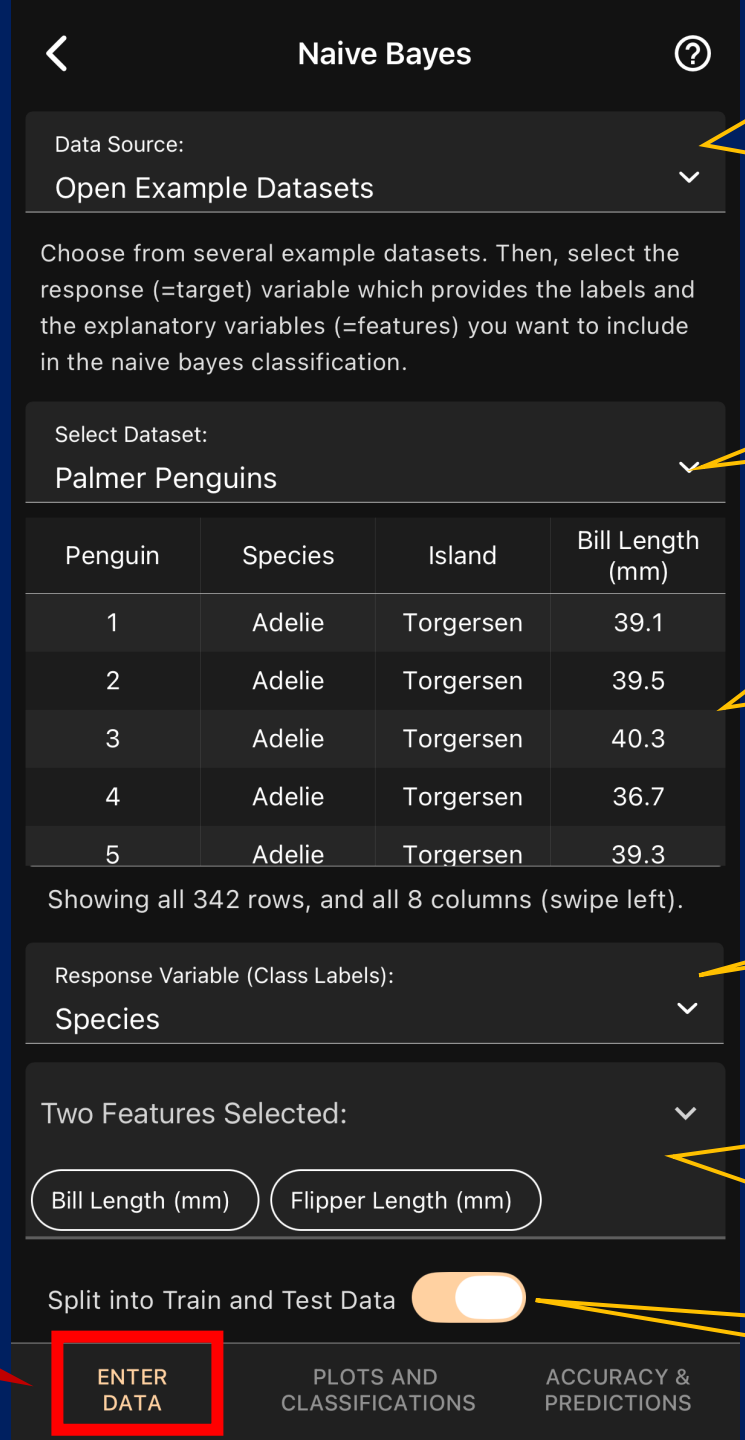
- Need prior class probabilities $P(Class\ k)$ and distribution of feature in a given class $P(x1 \mid Class\ k)$
- Predicted class is the one with largest posterior probability

Copyright 2025 Pearson. All rights reserved.

Example: Naïve Bayes

Tab 1:

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



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

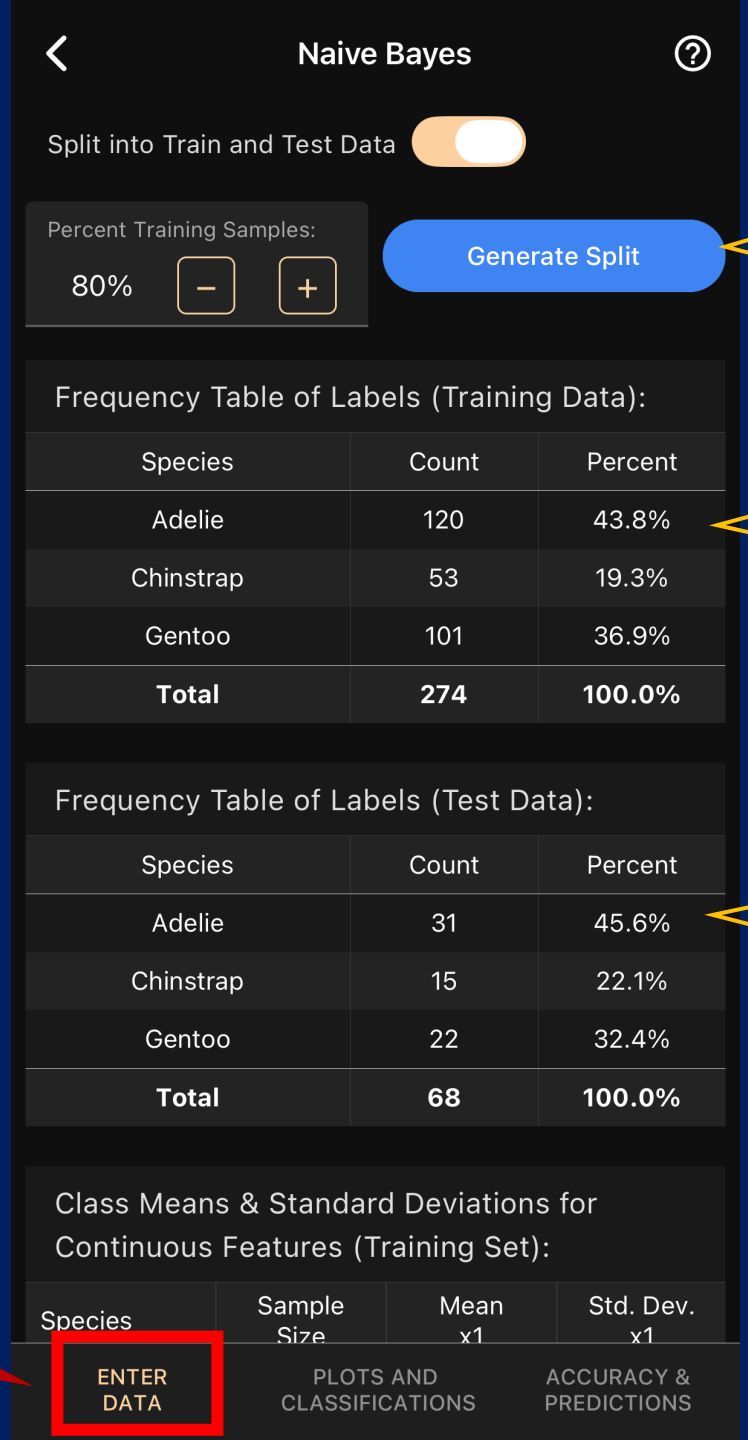
Split into testing & training set

Tab 1

Example: Naïve Bayes

Tab 1:

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



Generate 80/20 split by pressing button

See distribution of labels in training set

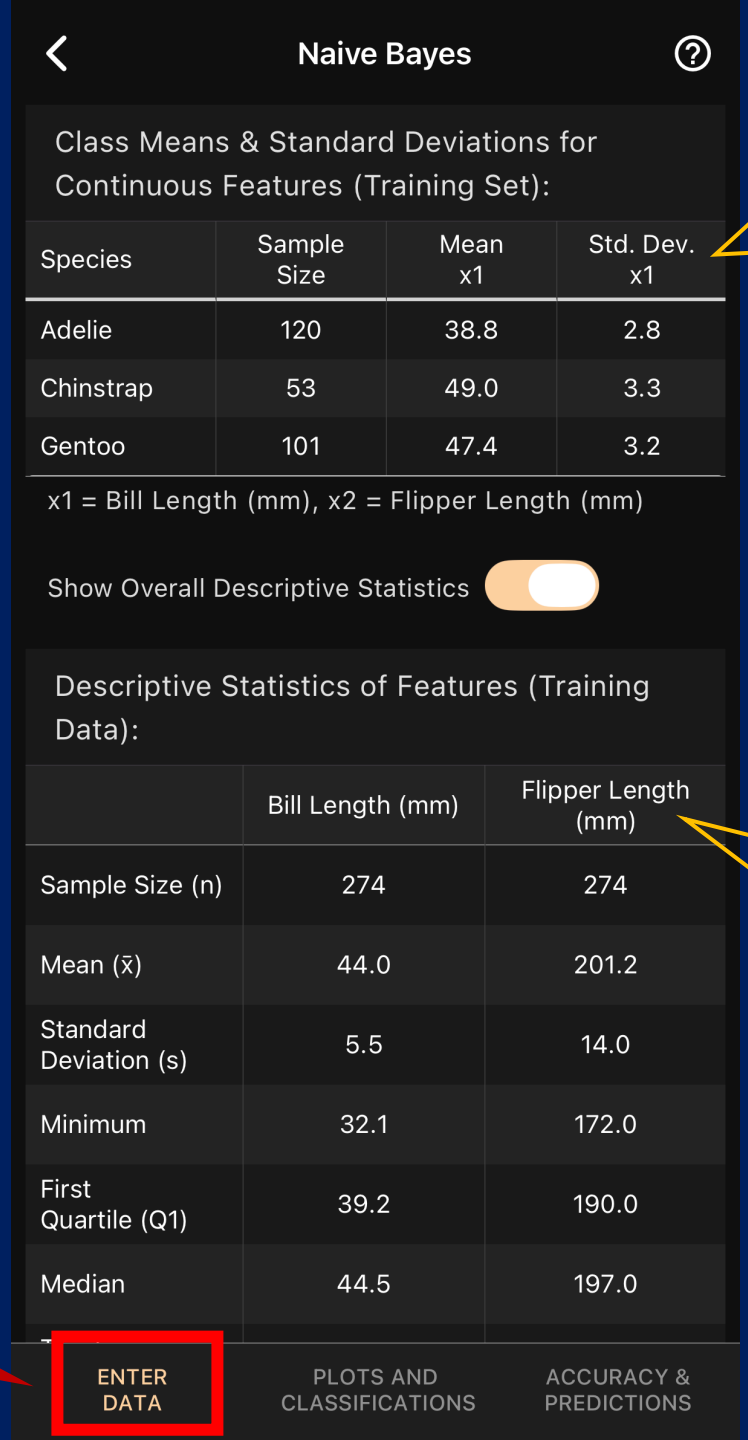
See distribution of labels in test set

Tab 1

Example: Naïve Bayes

Tab 1:

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



Tab 1

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

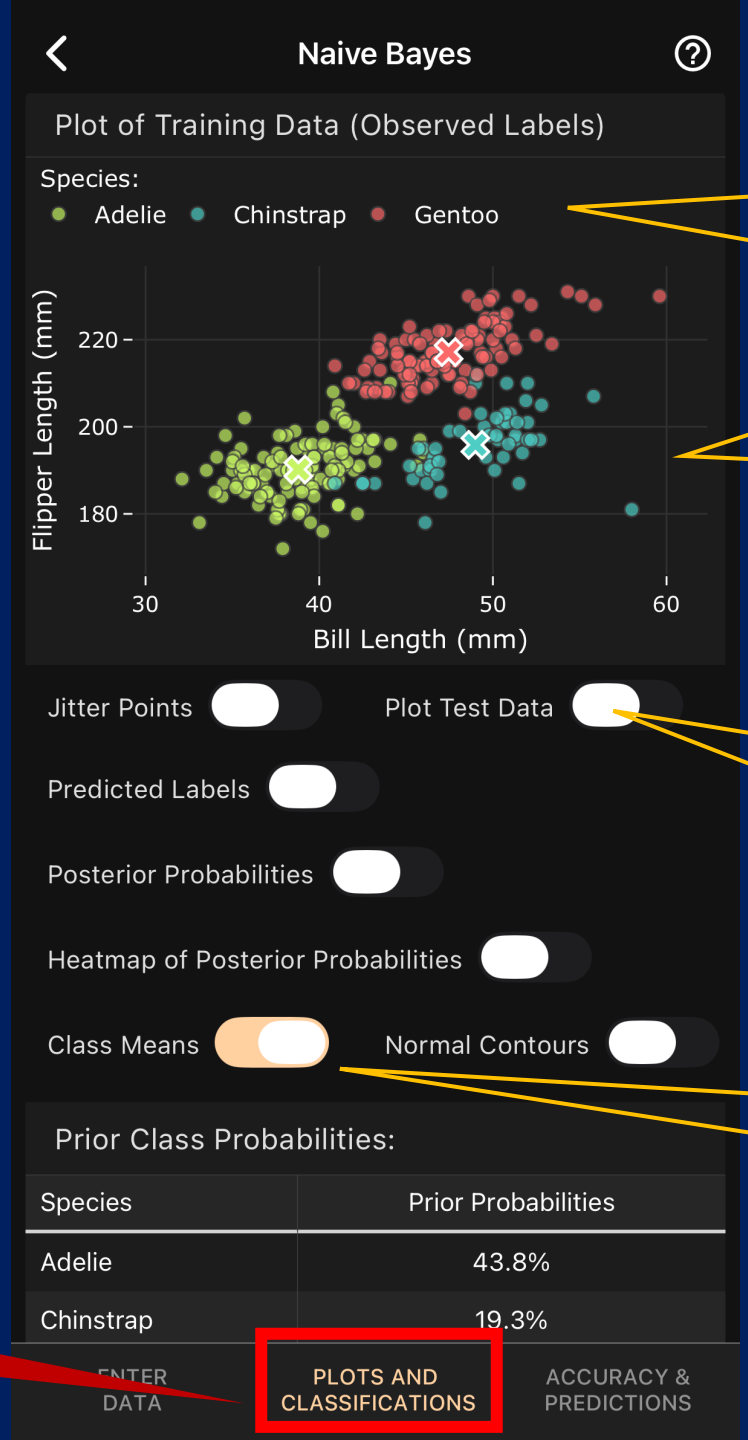
If we had categorical predictors (e.g., Sex or Island), app would show distribution of categories for each label.

See descriptive stats for each continuous predictor (not stratified)

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

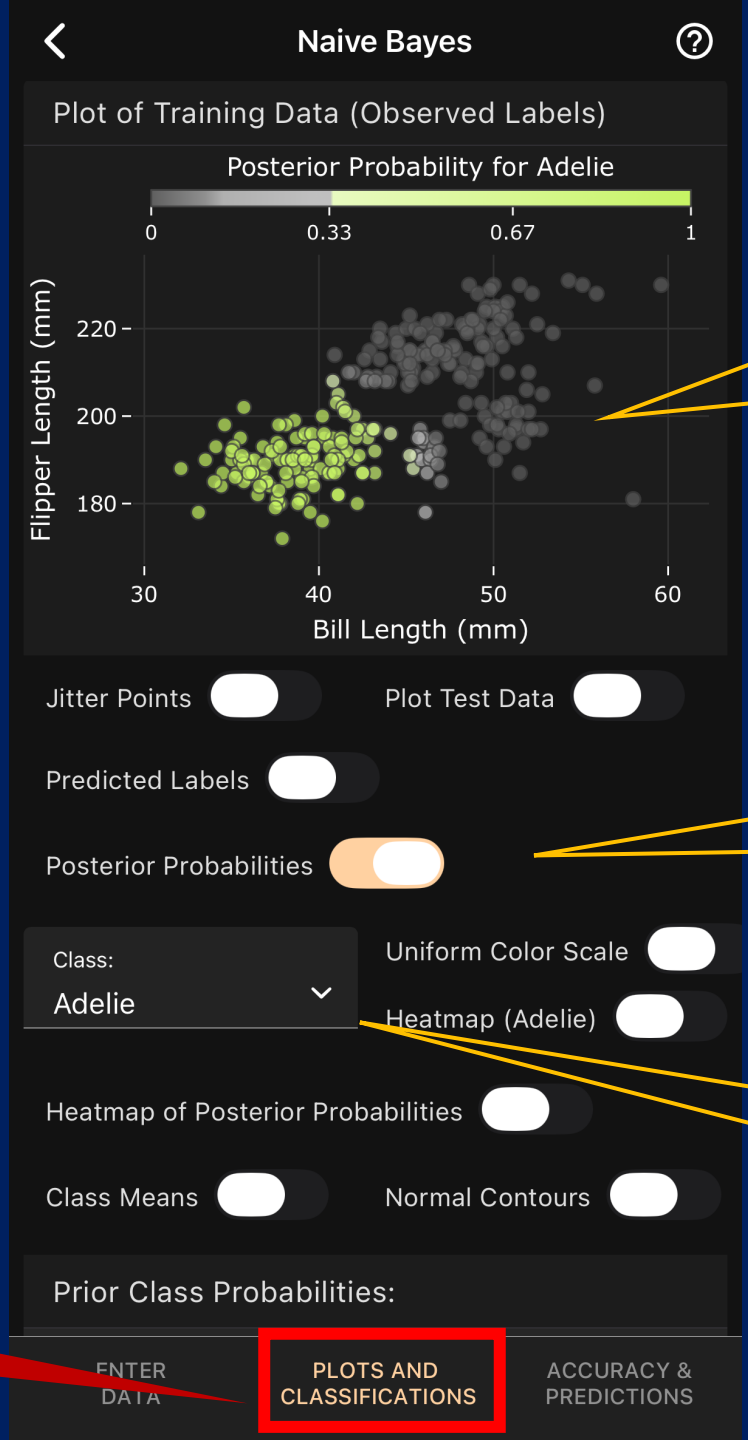
Show the mean for each target label

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

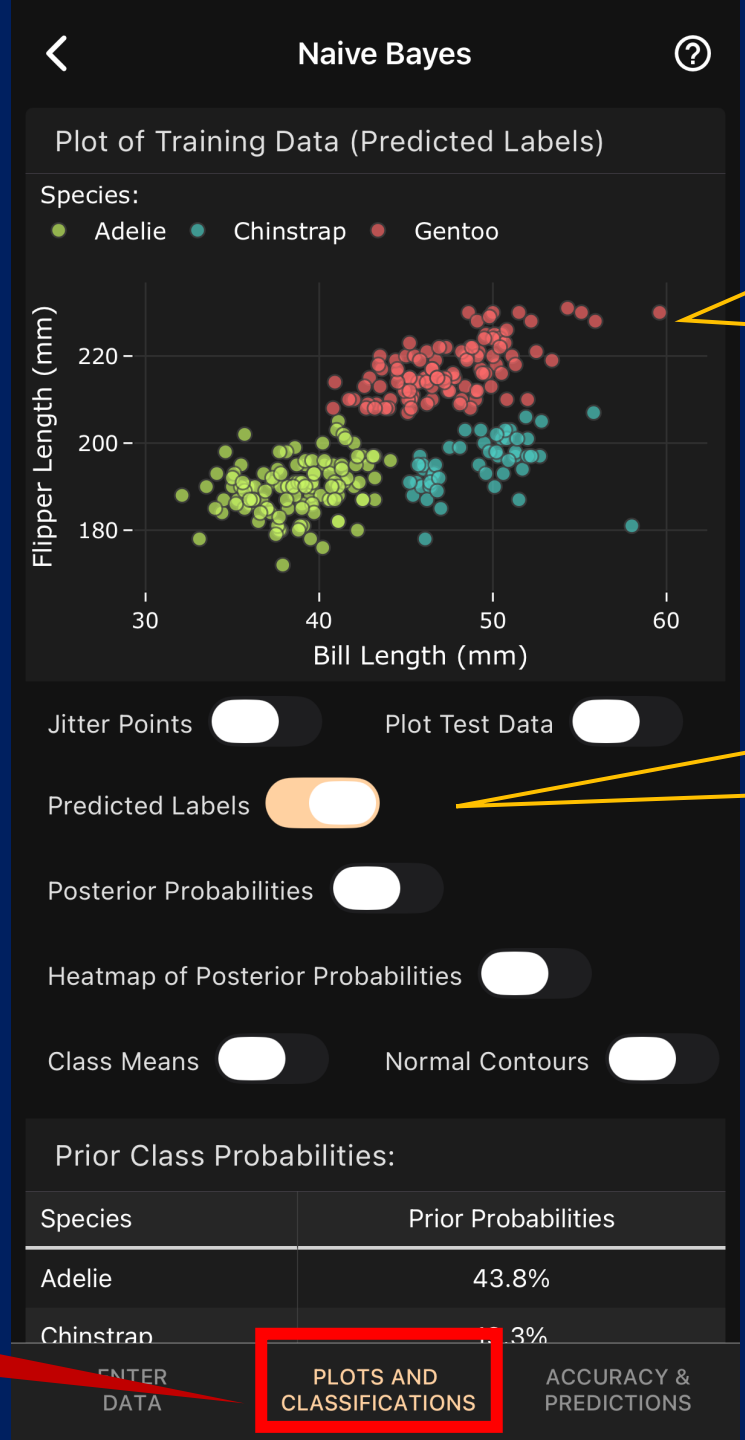
Select which class you want to show posterior odds for

Tab 2

Example: Naïve Bayes

Tab 2:

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



Scatterplot with points colored by predicted label (training set)

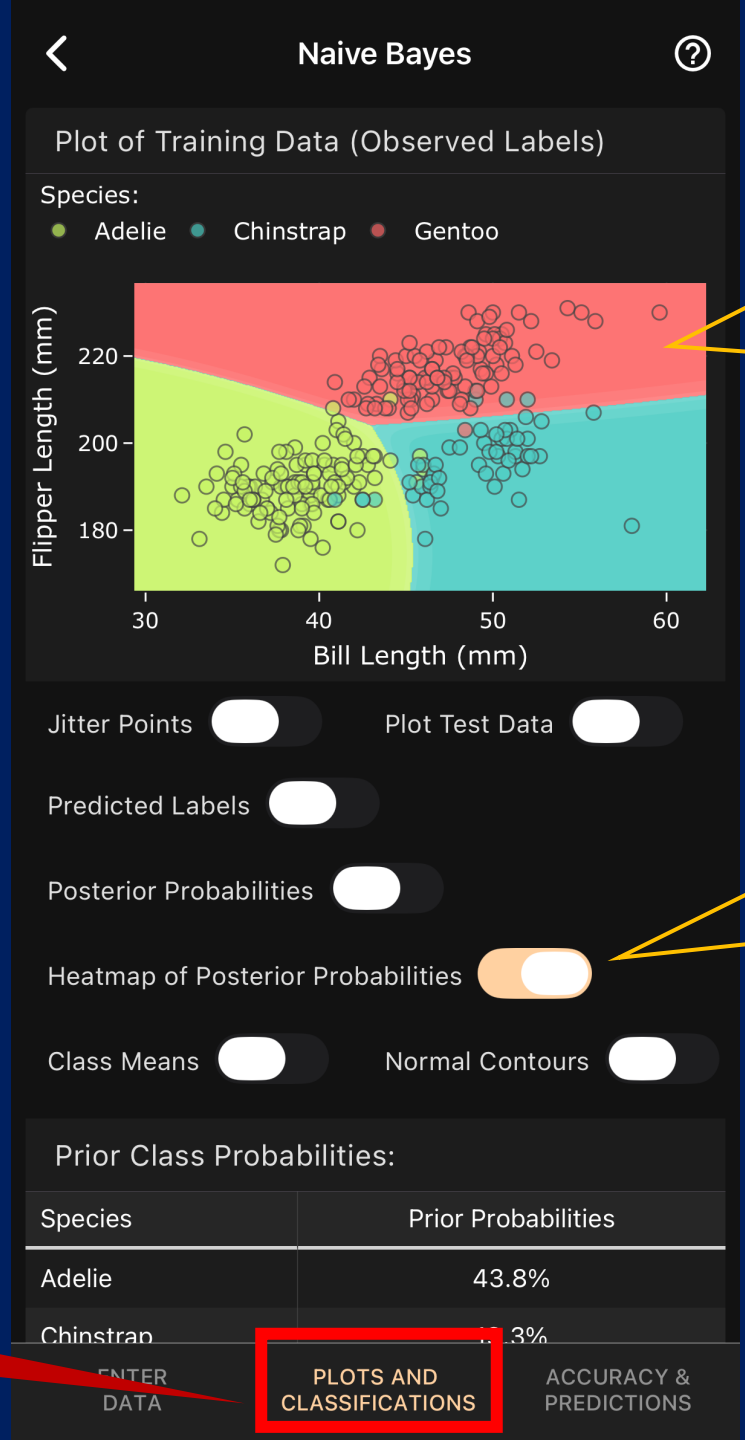
Switch to show the predicted label, based on maximum posterior odds

Tab 2

Example: Naïve Bayes

Tab 2:

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



Tab 2

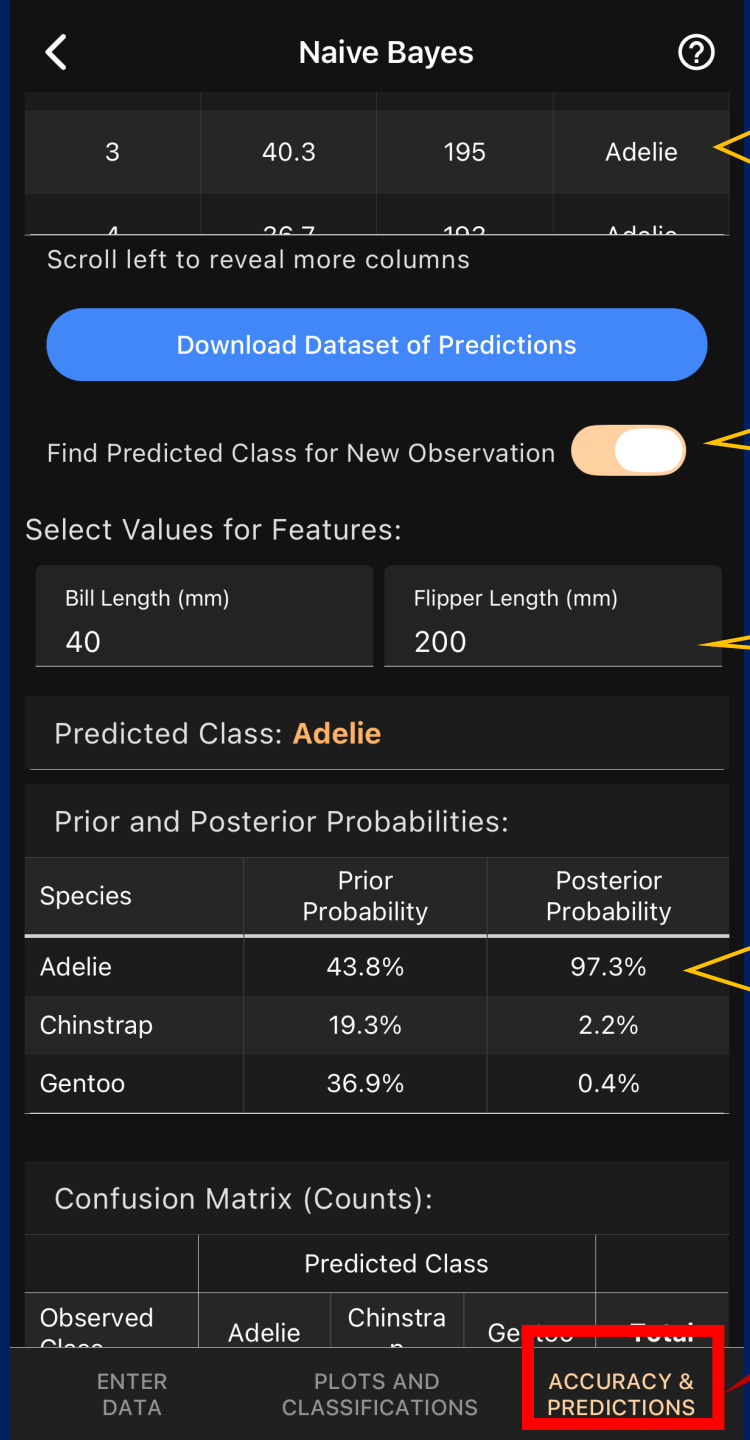
Scatterplot with points colored by original label but area colored by predicted label. (Training set)

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

Example: Naïve Bayes

Tab 3:

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



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

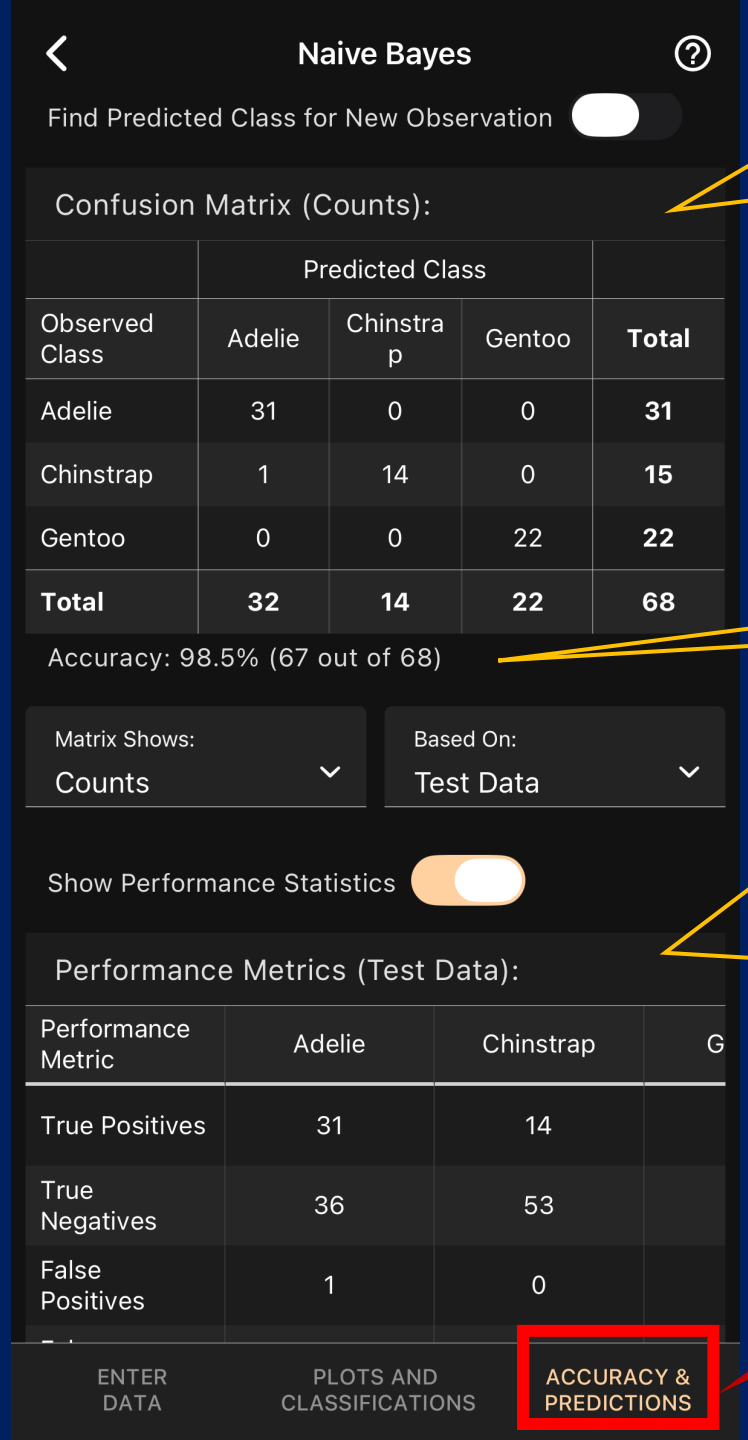
Table showing, for each class, the posterior probabilities for each class, based on the supplied feature values

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

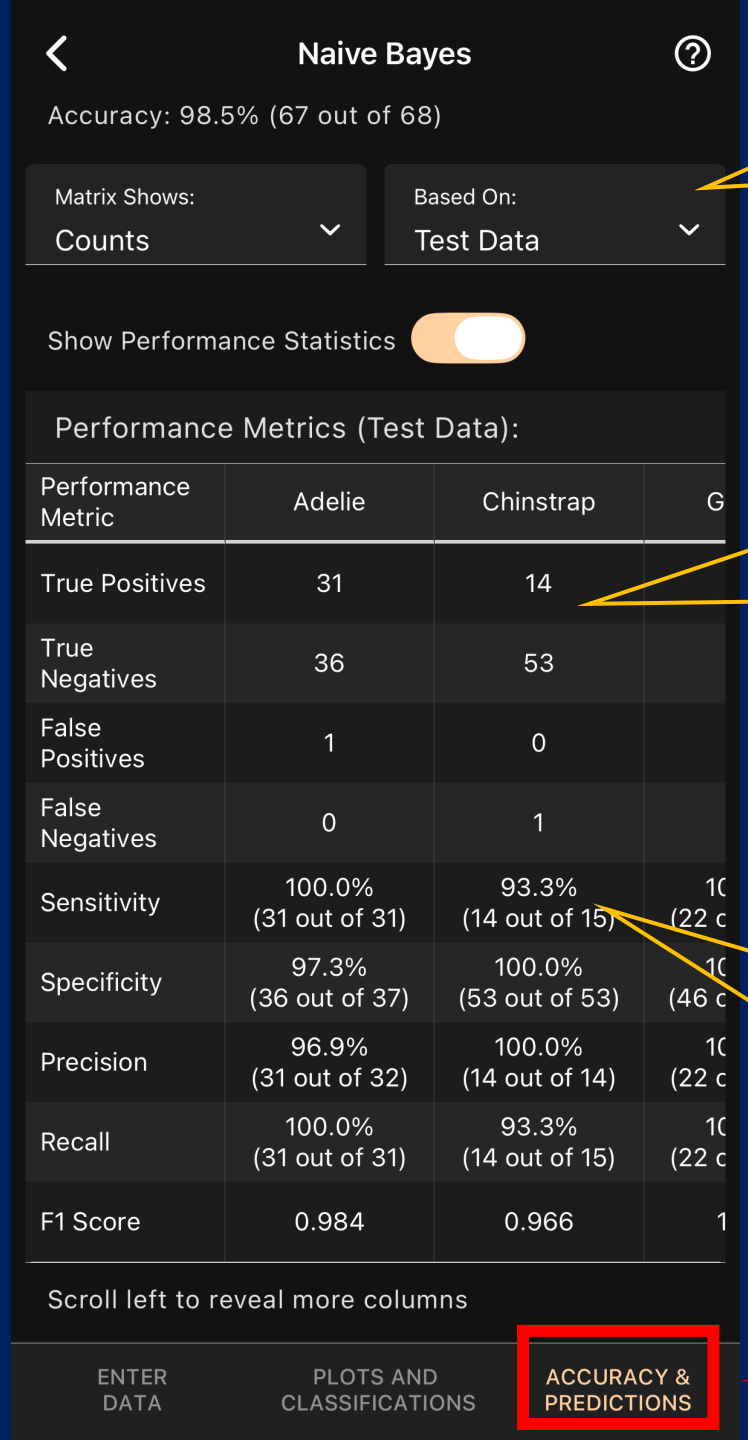
Table showing, for each class, true positives, true negatives, false positives and false negatives, sensitivity, specificity, precision, recall and F1 score.

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.

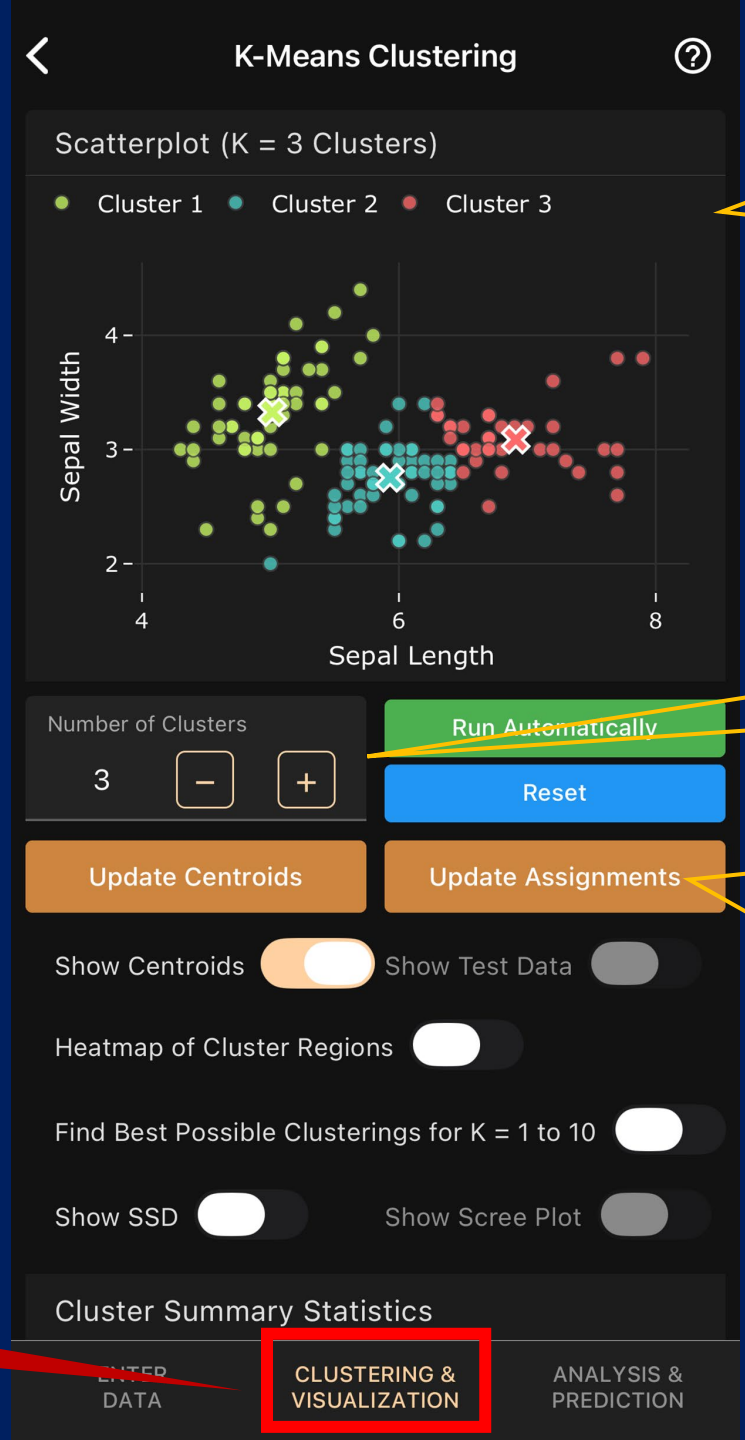
Also showing resulting sensitivity, specificity, precision, recall and F1 score.

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