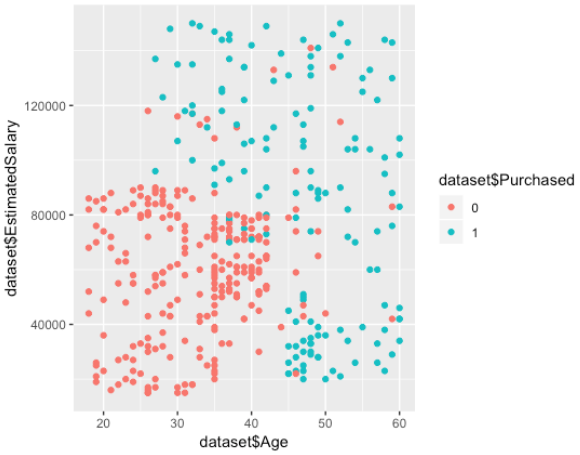
Q1 – Classifier Performance Comparison

Q1a –Analyze the data set Social_Network_Ads.csv and create the plot with correct titles on axes:

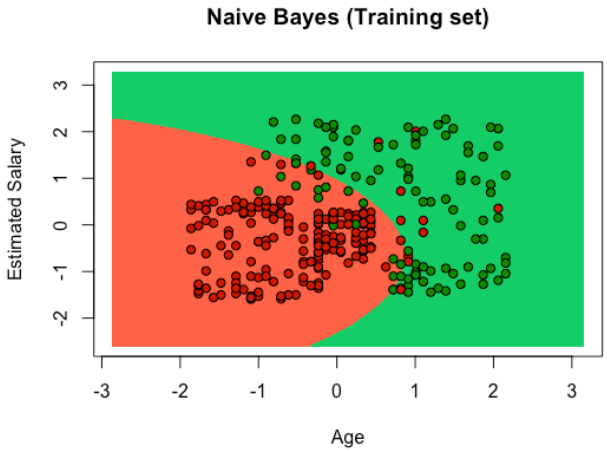


Q1b Use the following classifiers

I. Naïve Bayes
II. Logistic Regression
III. KNN

For each classifier show

- The classifier boundary for training and test
- Printout your 1st name on all graphs



Q1c Compare the confusion matrix in the following table for the above data set

	TP	TN	FP	FN	Accuracy
Naïve Bayes					90%
Logistic Regression					
KNN					

Q2 Iris Data Set and Classification (iris.csv)

The Iris dataset was used in R.A. Fisher's classic 1936 paper. It includes three iris species with 50 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other. The columns in this dataset are:

- Id
- Sepal Length Cm
- Sepal Width Cm
- Petal Length Cm
- Petal Width Cm
- Species

a. Plot the iris dataset – i) "Sepal Length vs Sepal Width" ii) "Petal Length vs Petal Width"
Split into Training / Test and
b. Apply Naïve Bayes Classifier to classify species with the decision boundaries
c. Apply logistic regression to classify species with the decision boundaries
d. Apply KNN algorithm to classify species with the decision boundaries
e. Compare the "Truth matrix" and Accuracy of the three algorithms

	TP	TN	FP	FN	Accuracy
Naïve Bayes					
Logistic Regression					
KNN					

Hint
Naïve Bayes - https://xavierbourretsicotte.github.io/Naive_Bayes_Classifier.html
Logistic Regression –
https://scikit-learn.org/stable/auto_examples/linear_model/plot_iris_logistic.html
https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python
KNN Algorithm –
https://www.ritchieng.com/machine-learning-k-nearest-neighbors-knn/

Q3 Titanic Data Set and Classification (titanic.zip – already separated as test, train)

a. Perform Exploratory Data Analysis
b. Do Feature Engineering
c. Apply logistic regression
d. Apply KNN algorithm

Hint
https://www.kaggle.com/angps95/basic-classification-methods-for-titanic

Q4. How does k-fold cross validation and grid search on the Social Ads Network data?
Use Q4 Dataset to analyze the difference
Use the references the explain how the two work together to evaluate a model
https://scikit-learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html
https://sebastianraschka.com/faq/docs/evaluate-a-model.html