

DJILLALI LIABES UNIVERSITY OF SIDI BEL ABBES
FACULTY OF EXACT SCIENCES
DEPARTMENT OF COMPUTER SCIENCES



Module : Data Mining
1ST YEAR OF MASTER'S DEGREE IN
NETWORKS, SYSTEMS & INFORMATION SECURITY (RSSI)
2021/2022

Data Mining with Weka

TP-02

Student:
HADJAZI Mohammed
Hisham
Group: 01 / RSSI

Module Instructor:
Pr. ELBERRICHI Zakaria
TP Instructor:
Dr. FAHSI. Mahmoud

A paper submitted in fulfilment of the requirements for the
Data Mining TP-02

March 24, 2022

Contents

List of Figures	iv
1 Dataset Preparation	1
1.1 Introduction	1
1.2 Original Article	2
1.3 Dataset Preperation	3
2 Evaluation Process	6
2.1 Introduction	6
2.2 Algorithms used	6
2.2.1 One Rule Algorithm	6
2.2.2 Naive Bayes Algorithm	7
2.2.3 K-nearest Neighbour Algorithm	7
2.2.4 Decision Trees (PART) Algorithm	8
2.2.5 Decision Trees (C4.5) Algorithm	10
2.2.6 Decision Trees (Random Forests) Algorithm	11
3 Evaluation Results	13
3.1 One Rule "default settings"	13
3.1.1 Cross Validation (5 Folds)	13
3.1.2 Cross Validation (10 Folds)	14
3.1.3 Cross Validation (15 Folds)	14
3.1.4 Cross Validation (20 Folds)	15
3.1.5 Cross Validation (25 Folds)	16
3.1.6 Cross Validation (30 Folds)	17
3.1.7 Percentage Split 50%	17
3.1.8 Percentage Split 60%	18
3.1.9 Percentage Split 66%	19
3.1.10 Percentage Split 70%	20
3.1.11 Percentage Split 80%	20
3.1.12 Percentage Split 90%	21
3.1.13 Leave one out Fold	22
3.1.14 Conclusion	23
3.2 Naive Bayes "default settings"	23
3.2.1 Cross Validation (5 Folds)	23
3.2.2 Cross Validation (10 Folds)	25
3.2.3 Cross Validation (15 Folds)	26
3.2.4 Cross Validation (20 Folds)	27
3.2.5 Cross Validation (25 Folds)	29
3.2.6 Cross Validation (30 Folds)	30
3.2.7 Percentage Split 50%	32
3.2.8 Percentage Split 60%	33
3.2.9 Percentage Split 66%	35

3.2.10	Percentage Split 70%	36
3.2.11	Percentage Split 80%	38
3.2.12	Percentage Split 90%	39
3.2.13	Leave one out Fold	41
3.2.14	Conclusion	42
3.3	K-nearest Neighbour " K-value = 3 and default settings"	43
3.3.1	Cross Validation (5 Folds)	43
3.3.2	Cross Validation (10 Folds)	43
3.3.3	Cross Validation (15 Folds)	44
3.3.4	Cross Validation (20 Folds)	45
3.3.5	Cross Validation (25 Folds)	45
3.3.6	Cross Validation (30 Folds)	46
3.3.7	Percentage Split 50%	47
3.3.8	Percentage Split 60%	47
3.3.9	Percentage Split 66%	48
3.3.10	Percentage Split 70%	49
3.3.11	Percentage Split 80%	50
3.3.12	Percentage Split 90%	50
3.3.13	Leave one out Fold	51
3.3.14	Conclusion	52
3.4	C4.5 "default settings"	52
3.4.1	Cross Validation (5 Folds)	52
3.4.2	Cross Validation (10 Folds)	53
3.4.3	Cross Validation (15 Folds)	54
3.4.4	Cross Validation (20 Folds)	55
3.4.5	Cross Validation (25 Folds)	56
3.4.6	Cross Validation (30 Folds)	57
3.4.7	Percentage Split 50%	58
3.4.8	Percentage Split 60%	60
3.4.9	Percentage Split 66%	61
3.4.10	Percentage Split 70%	62
3.4.11	Percentage Split 80%	63
3.4.12	Percentage Split 90%	64
3.4.13	Leave one out Fold	65
3.4.14	Conclusion	66
3.5	PART "default settings"	68
3.5.1	Cross Validation (5 Folds)	68
3.5.2	Cross Validation (10 Folds)	69
3.5.3	Cross Validation (15 Folds)	70
3.5.4	Cross Validation (20 Folds)	71
3.5.5	Cross Validation (25 Folds)	73
3.5.6	Cross Validation (30 Folds)	74
3.5.7	Percentage Split 50%	75
3.5.8	Percentage Split 60%	76
3.5.9	Percentage Split 66%	78
3.5.10	Percentage Split 70%	79
3.5.11	Percentage Split 80%	80
3.5.12	Percentage Split 90%	81
3.5.13	Leave one out Fold	83
3.5.14	Conclusion	84
3.6	Random Forest "default settings"	84

3.6.1	Cross Validation (5 Folds)	84
3.6.2	Cross Validation (10 Folds)	85
3.6.3	Cross Validation (15 Folds)	86
3.6.4	Cross Validation (20 Folds)	87
3.6.5	Cross Validation (25 Folds)	87
3.6.6	Cross Validation (30 Folds)	88
3.6.7	Percentage Split 50%	89
3.6.8	Percentage Split 60%	89
3.6.9	Percentage Split 66%	90
3.6.10	Percentage Split 70%	91
3.6.11	Percentage Split 80%	92
3.6.12	Percentage Split 90%	92
3.6.13	Leave one out Fold	93
3.6.14	Conclusion	94
3.7	General Conclusion	94
4	Application part	96
4.1	Cross Validation implementation	96
4.2	Percentage Split implementation	97
5	Conclusion	99
A	Source Code	101
A.1	apigui.java	101

List of Figures

1.1	Star Wars	1
1.2	ModernCSV	3
1.3	arff file view	5
1.4	arff file view 2	5
1.5	arff file view 3	5
2.1	Naive Bayes Algorithm	7
2.2	Distance Functions	8
2.3	Hamming Distance	8
2.4	Random Forests	11
3.1	Generated Tree for C4.5	67
4.1	Cross Validation Execution	96
4.2	Percentage Split Execution	98

Chapter 1

Dataset Preparation

1.1 Introduction



FIGURE 1.1: Star Wars

The idea was to find the popularity of the star wars franchise among people and the relation of this addiction to some other factors. it was hard to find reliable sources of data or surveys, however i was able to find a survey done by **abc News** on the **22 of July 2014**. the survey had another aim, as it was searching for best Star Wars movies, least favourite (most hated) characters among Star Wars characters. The title of the survey was **America's Favorite 'Star Wars' Movies (And Least Favorite Characters)** [8].

However I re-utilised the survey to use it for finding the relation between Star Wars fans and for example **Gender, Age, Income, Education** and many other factors.

1.2 Original Article

This week, I caught a sneak peek of the X-Wing fighter from the new “Star Wars” films in production. The forthcoming movies — and the middling response to the most recent trilogy — provide a perfect excuse to examine some questions I’ve long wanted answers to: How many people are “Star Wars” fans? Does the rest of America realize that “The Empire Strikes Back” is clearly the best of the bunch? Which characters are most well-liked and most hated? And who shot first, Han Solo or Greedo?

We ran a poll through SurveyMonkey Audience, surveying 1,186 respondents from June 3 to 6 (the data is available on GitHub). Seventy-nine percent of those respondents said they had watched at least one of the “Star Wars” films. This question, incidentally, had a substantial difference by gender: 85 percent of men have seen at least one “Star Wars” film compared to 72 percent of women. Of people who have seen a film, men were also more likely to consider themselves a fan of the franchise: 72 percent of men compared to 60 percent of women.

We then asked respondents which of the films they had seen. With 835 people responding, here’s the probability that someone has seen a given “Star Wars” film given that they have seen any Star Wars film:

So we can see that “Star Wars: Episode V — The Empire Strikes Back” is the film seen by the most number of people, followed by “Star Wars: Episode VI — Return of the Jedi.” Appallingly, more people reported seeing “Star Wars: Episode I — The Phantom Menace” than the original “Star Wars” (renamed “Star Wars: Episode IV — A New Hope”).

So, which movie is the best? We asked the subset of 471 respondents who indicated they have seen every “Star Wars” film to rank them from best to worst. From that question, we calculated the share of respondents who rated each film as their favorite.

We can also drill down and find out, generally, how people rate the films. Overall, fans broke into two camps: those who preferred the original three movies and those who preferred the three prequels. People who said “The Empire Strikes Back” was their favorite were also likely to rate “A New Hope” and “Return of the Jedi” higher as well. Those who rated “The Phantom Menace” as the best film were more likely to rate prequels higher.

This chart shows how often each film was rated in the top third (best or second best), the middle third (third or fourth) or the bottom third (second worst or worst). It’s a more nuanced take on the series:

Finally, we took a boilerplate format used by political favorability polls — “Please state whether you view the following characters favorably, unfavorably, or are unfamiliar with him/her” — and asked respondents to rate characters in the series.

You read that correctly. Jar Jar Binks has a lower favorability rating than the actual personification of evil in the galaxy.

And for those of you who want to know the impact that historical revisionism can have on a society:

1.3 Dataset Preperation

The original survey dataset was obtained from github[5]. it was a CSV type which included 1186 results. after converting it to ARFF format which was very challenging, as some special characters had to be removed, as it wasn't compatible to UTF-8 standard. anothe problem I had was replacing all missing values with '?' without missing up the data. it must be noted that it was very helpful to utilise an online converter from CSV to ARFF format <https://ikuz.eu/csv2arff/>.

	0	1	2	3
	Have-you-seen-any-of-the-6-films-in-the-Star-Wars-franchise	Which-character-shot-first	Are-you-familiar-with-the-Expanded-Universe	Do-you-conside
0	Yes	none	Yes	No
1	No	?	?	?
2	Yes	none	No	?
3	Yes	none	No	?
4	Yes	Greedo	Yes	No
5	Yes	Han	Yes	No
6	Yes	Han	Yes	No
7	Yes	Han	No	?
8	Yes	Han	No	?
9	Yes	none	No	?
10	Yes	?	?	?
11	Yes	?	?	?
12	No	?	?	?
13	Yes	Greedo	No	?
14	Yes	none	No	?
15	Yes	Han	No	?
16	Yes	Han	Yes	Yes
17	Yes	Han	Yes	Yes
18	Yes	Greedo	No	?
19	Yes	Han	No	?
20	Yes	Greedo	No	?
21	Yes	Han	No	?
22	Yes	none	No	?
23	Yes	Greedo	Yes	Yes
24	Yes	Han	Yes	?

FIGURE 1.2: ModernCSV

Since handling CSV files is much easier, I did the renaming of attributes and affectation of missing values from **null** to **"?"** to be in weka's arff specifications. I used a third party tool **ModernCSV** to fix the problems in the dataset.

There were many missing values and the worst part many instances with only a single value which were useless and didn't help in the evaluation part. but I decided to keep the dataset as intact as possible. I have only removed the data that was unrelated to the search subject. The original CSV survey dataset had these attributes :

1. RespondentID
2. Have you seen any of the 6 films in the Star Wars franchise?
3. Do you consider yourself to be a fan of the Star Wars film franchise?
4. Which of the following Star Wars films have you seen? Please select all that apply.
5. Please rank the Star Wars films in order of preference with 1 being your favorite film in the franchise and 6 being your least favorite film.
6. Please state whether you view the following characters favorably, unfavorably, or are unfamiliar with him/her.
7. Which character shot first?

8. Are you familiar with the Expanded Universe?
9. Do you consider yourself to be a fan of the Expanded Universe?
10. Do you consider yourself to be a fan of the Star Trek franchise?
11. Gender
12. Age
13. Household Income
14. Education
15. Location (Census Region)

I have used the **Do you consider yourself to be a fan of the Star Wars film franchise?** as the **class** attribute and removed the attributes that i think were not useful for my question. after modifying i ended with the following list of attributes :

1. Have you seen any of the 6 films in the Star Wars franchise?
2. Which character shot first?
3. Are you familiar with the Expanded Universe?
4. Do you consider yourself to be a fan of the Expanded Universe?
5. Gender
6. Age
7. Household Income
8. Education
9. Location (Census Region)
10. Do you consider yourself to be a fan of the Star Trek franchise?
11. **Class** : Do you consider yourself to be a fan of the Star Wars film franchise?

Now after we are done with the modifications we convert the file to arff format to use it in weka :

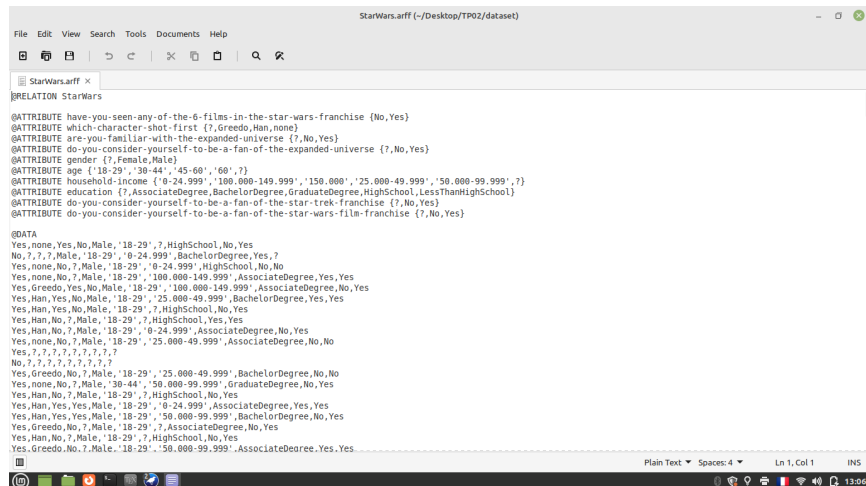


FIGURE 1.3: arff file view

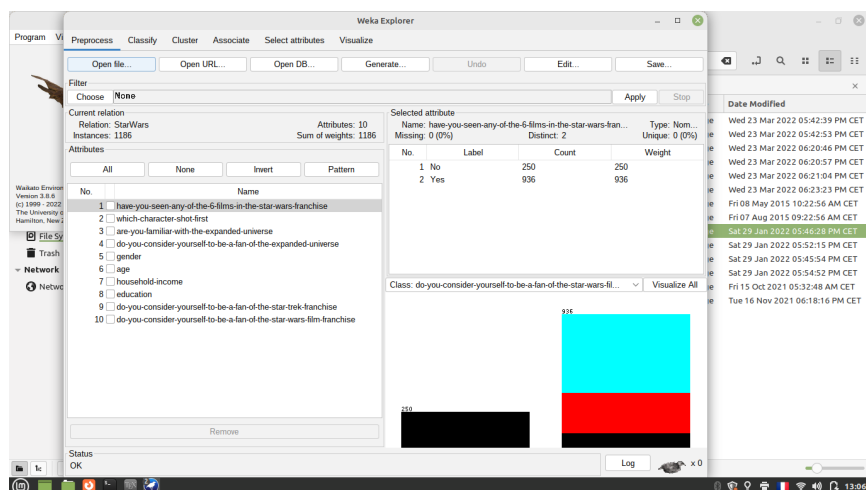


FIGURE 1.4: arff file view 2

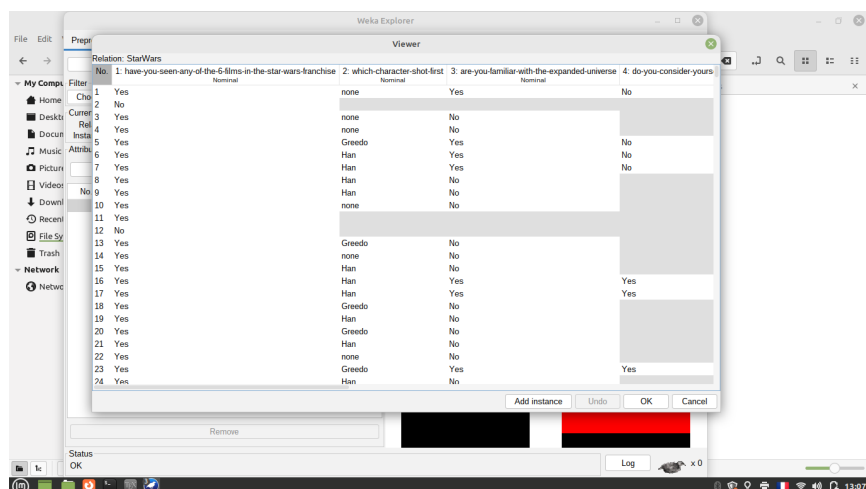


FIGURE 1.5: arff file view 3

Chapter 2

Evaluation Process

2.1 Introduction

The choice of algorithms was classic as all the major classification algorithms have been used except for ID3, which was excluded due to 2 reasons first the need to update the dataset to be able to work with ID3, second C4.5 can replace it as its mainly the same algorithm with improvements to make it able to deal with numerical data.

2.2 Algorithms used

We have chosen 6 classification algorithms for the experiments, some are rule based some are tree based, as we will see some are fast and some requires some computation time to execute.

2.2.1 One Rule Algorithm

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, we construct a frequency table for each predictor against the target. It has been shown that OneR produces rules only slightly less accurate than state-of-the-art classification algorithms while producing rules that are simple for humans to interpret.[\[3\]](#)

OneR Algorithm

1. For each predictor,
2. For each value of that predictor, make a rule as follows;
3. Count how often each value of target (class) appears
4. Find the most frequent class
5. Make the rule assign that class to this value of the predictor
6. Calculate the total error of the rules of each predictor
7. Choose the predictor with the smallest total error.

2.2.2 Naive Bayes Algorithm

The Naive Bayesian classifier is based on Bayes' theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.[3]

Naive Bayes Algorithm

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.[3]

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$P(c|x)$ is labeled **Posterior Probability**
 $P(x|c)$ is labeled **Likelihood**
 $P(c)$ is labeled **Class Prior Probability**
 $P(x)$ is labeled **Predictor Prior Probability**

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

FIGURE 2.1: Naive Bayes Algorithm

- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

2.2.3 K-nearest Neighbour Algorithm

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.[3]

KNN Algorithm

A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a

distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. [3]

Distance functions

Euclidean

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

Minkowski

$$\left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{1/q}$$

FIGURE 2.2: Distance Functions

It should also be noted that all three distance measures are only valid for continuous variables. In the instance of categorical variables the Hamming distance must be used. It also brings up the issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset.[3]

Hamming Distance

$$D_H = \sum_{i=1}^k |x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

FIGURE 2.3: Hamming Distance

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3-10. That produces much better results than 1NN.[3]

2.2.4 Decision Trees (PART) Algorithm

PART is a partial decision tree algorithm, which is the developed version of C4.5 and RIPPER algorithms. The main speciality of the PART algorithm is that it does

not need to perform global optimisation like C4.5 and RIPPER to produce the appropriate rules. However, decision trees are sometime more problematic due to the larger size of the tree which could be oversized and might perform badly for classification problems.[2]

PART is a separate-and-conquer rule learner. The algorithm producing sets of rules called "decision lists" which are planned set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the class of the first matching rule. PART builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule.

2.2.5 Decision Trees (C4.5) Algorithm

C4.5 is an algorithm developed by Ross Quinlan that generates Decision Trees (DT), which can be used for classification problems. It improves (extends) the ID3 algorithm by dealing with both continuous and discrete attributes, missing values and pruning trees after construction. Its commercial successor is C5.0, See5, a lot faster than C4.5, more memory efficient and used for building smaller decision trees.[6]

Being a supervised learning algorithm, it requires a set of training examples and each example can be seen as a pair: input object and a desired output value (class). The algorithm analyzes the training set and builds a classifier that must be able to correctly classify both training and test examples. A test example is an input object and the algorithm must predict an output value (the example must be assigned to a class).[6]

C4.5 Algorithm :

1. Check for the above base cases.
2. For each attribute a , find the normalised information gain ratio from splitting on a .
3. Let a_{best} be the attribute with the highest normalized information gain.
4. Create a decision node that splits on a_{best} .
5. Recur on the sublists obtained by splitting on a_{best} , and add those nodes as children of node.

The classifier used by C4.5 is a decision tree and this tree is built from root to leaves by respecting the Occam's Razor. This razor says that given two correct solutions for a certain problem, we should choose the simpler solution.[6]

The advantages of the C4.5 are:

- Builds models that can be easily interpreted
- Easy to implement
- Can use both categorical and continuous values
- Deals with noise

The disadvantages are:

- Small variation in data can lead to different decision trees (especially when the variables are close to each other in value)

- Does not work very well on a small training set

C4.5 is used in classification problems and it is the most used algorithm for building DT. It is suitable for real world problems as it deals with numeric attributes and missing values. The algorithm can be used for building smaller or larger, more accurate decision trees and the algorithm is quite time efficient. Compared to ID3, C4.5 performs by default a tree pruning process, which leads to smaller trees, more simple rules and more intuitive interpretations.[6]

2.2.6 Decision Trees (Random Forests) Algorithm

Random forest classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement.[1]

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.[12]

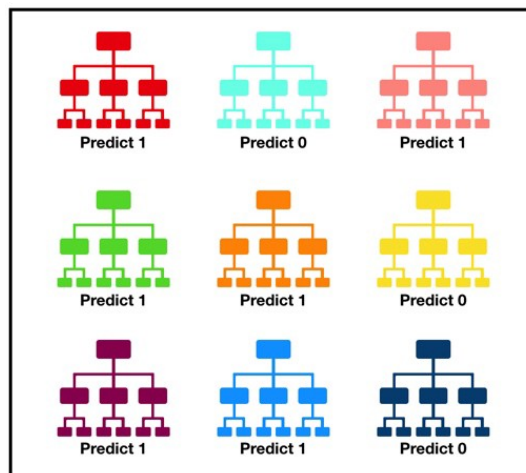


FIGURE 2.4: Random Forests

The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the random forest model works so well is:

“A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.”

The low correlation between models is the key. Just like how investments with low correlations (like stocks and bonds) come together to form a portfolio that is greater than the sum of its parts, uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. The reason for

this wonderful effect is that the trees protect each other from their individual errors (as long as they don't constantly all err in the same direction). While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction. So the prerequisites for random forest to perform well are:

1. There needs to be some actual signal in our features so that models built using those features do better than random guessing.
2. The predictions (and therefore the errors) made by the individual trees need to have low correlations with each other.

Random forests can be viewed as a generalization of the basic bagging method, as applied to decision trees. Random forests are defined as an ensemble of decision trees, in which randomness has explicitly been inserted into the model building process of each decision tree. While the bootstrapped sampling approach of bagging is also an indirect way of adding randomness to model-building, there are some disadvantages of doing so. The main drawback of using decision-trees directly with bagging is that the split choices at the top levels of the tree are statistically likely to remain approximately invariant to bootstrapped sampling. Therefore, the trees are more correlated, which limits the amount of error reduction obtained from bagging. In such cases, it makes sense to directly increase the diversity of the component decision-tree models. The idea is to use a randomized decision tree model with less correlation between the different ensemble components. The underlying variability can then be more effectively reduced by an averaging approach. The final results are often more accurate than a direct application of bagging on decision trees.[4]

The random-split selection approach directly introduces randomness into the split criterion. An integer parameter $q \leq d$ is used to regulate the amount of randomness introduced in split selection. The split selection at each node is preceded by the randomized selection of a subset S of attributes of size q . The splits at that node are then executed using only this subset. Larger values of q will result in correlated trees that are similar to a tree without any injected randomness. By selecting small values of q relative to the full dimensionality.[4]

Chapter 3

Evaluation Results

The results are taken straight out of weka's classifier output and made into text files (will be included with the supplementary files). at the end of the tests there will be a conclusion of the results along with a table that summarises the performance according to correctly classified instances.

3.1 One Rule "default settings"

3.1.1 Cross Validation (5 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   5-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762
Root mean squared error             0.4198
Relative absolute error             58.7986 %
Root relative squared error         108.5589 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.838	0.317	0.576	0.838	0.683	0.494	0.696	0.350	No
	0.683	0.162	0.891	0.683	0.773	0.494	0.709	0.620	Yes
Weighted Avg.	0.736	0.215	0.784	0.736	0.743	0.494	0.704	0.528	

```

=== Confusion Matrix ===

```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 238 46   |   b = No
0 175 377  |   c = Yes

```

3.1.2 Cross Validation (10 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762
Root mean squared error            0.4198
Relative absolute error             58.8124 %
Root relative squared error        108.5585 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        ?        ?          ?        ?          ?        ?        ?        ?
      0.838    0.317    0.576      0.838    0.683      0.494    0.696    0.350    No
      0.683    0.162    0.891      0.683    0.773      0.494    0.709    0.620    Yes
Weighted Avg. 0.736    0.215    0.784      0.736    0.743      0.494    0.704    0.528

```

```

=== Confusion Matrix ===

```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 238 46   |   b = No
0 175 377  |   c = Yes

```

3.1.3 Cross Validation (15 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income

```

```

education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 15-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      605          72.3684 %
Incorrectly Classified Instances    231          27.6316 %
Kappa statistic                    0.4392
Mean absolute error                 0.1842
Root mean squared error             0.4292
Relative absolute error             61.4781 %
Root relative squared error        110.9881 %
Total Number of Instances          836
Ignored Class Unknown Instances    350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000    ?          ?       ?         ?       ?         ?         ?
          0.796    0.313    0.566    0.796    0.662    0.457    0.683    0.342    No
          0.687    0.204    0.867    0.687    0.766    0.457    0.691    0.601    Yes
Weighted Avg.    0.724    0.241    0.765    0.724    0.731    0.457    0.688    0.513

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 226  58 |   b = No
  0 173 379 |   c = Yes

```

3.1.4 Cross Validation (20 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    20-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762

```

```

Root mean squared error          0.4198
Relative absolute error          58.8178 %
Root relative squared error      108.5576 %
Total Number of Instances        836
Ignored Class Unknown Instances  350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000  ?           ?         ?           ?         ?         ?         ?
      0.838    0.317    0.576    0.838    0.683       0.494    0.696    0.350    No
      0.683    0.162    0.891    0.683    0.773       0.494    0.709    0.620    Yes
Weighted Avg.   0.736    0.215    0.784    0.736    0.743       0.494    0.704    0.528

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 238  46 |   b = No
  0 175 377 |   c = Yes

```

3.1.5 Cross Validation (25 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   25-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762
Root mean squared error             0.4198
Relative absolute error             58.8191 %
Root relative squared error         108.5576 %
Total Number of Instances          836
Ignored Class Unknown Instances    350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000  ?           ?         ?           ?         ?         ?         ?
      0.838    0.317    0.576    0.838    0.683       0.494    0.696    0.350    No
      0.683    0.162    0.891    0.683    0.773       0.494    0.709    0.620    Yes
Weighted Avg.   0.736    0.215    0.784    0.736    0.743       0.494    0.704    0.528

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 238  46 |   b = No
  0 175 377 |   c = Yes

```

3.1.6 Cross Validation (30 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   30-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762
Root mean squared error             0.4198
Relative absolute error             58.8183 %
Root relative squared error         108.555 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.838	0.317	0.576	0.838	0.683	0.494	0.696	0.350	No
	0.683	0.162	0.891	0.683	0.773	0.494	0.709	0.620	Yes
Weighted Avg.	0.736	0.215	0.784	0.736	0.743	0.494	0.704	0.528	

```

=== Confusion Matrix ===

 a   b   c   <-- classified as
0   0   0 | a = ?
0 238  46 | b = No
0 175 377 | c = Yes

```

3.1.7 Percentage Split 50%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 50.0% train, remainder test

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?

```

```

No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      303          72.315 %
Incorrectly Classified Instances    116          27.685 %
Kappa statistic                    0.3922
Mean absolute error                 0.1846
Root mean squared error            0.4296
Relative absolute error             61.467 %
Root relative squared error         110.1453 %
Total Number of Instances          419
Ignored Class Unknown Instances     174

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000  ?          ?        ?          ?        ?        ?        ?
          0.605    0.213    0.605    0.605    0.605    0.392    0.738    0.464    No
          0.787    0.395    0.787    0.787    0.787    0.392    0.532    0.475    Yes
Weighted Avg.    0.723    0.331    0.723    0.723    0.723    0.392    0.604    0.471

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0  89   58 |  b = No
  0   58 214 |  c = Yes

```

3.1.8 Percentage Split 60%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 60.0% train, remainder test

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      248          72.3032 %
Incorrectly Classified Instances     95          27.6968 %
Kappa statistic                    0.3973
Mean absolute error                 0.1846
Root mean squared error            0.4297
Relative absolute error             60.981 %
Root relative squared error         108.3917 %

```

```

Total Number of Instances          343
Ignored Class Unknown Instances    131

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000  ?          ?        ?          ?        ?        ?        ?
      0.591    0.199    0.636    0.591    0.612    0.398    0.733    0.485    No
      0.801    0.409    0.769    0.801    0.785    0.398    0.546    0.480    Yes
Weighted Avg.    0.723    0.332    0.720    0.723    0.721    0.398    0.615    0.482

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 |  a = ?
  0 75 52 |  b = No
  0 43 173 | c = Yes

```

3.1.9 Percentage Split 66%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      212          73.6111 %
Incorrectly Classified Instances     76          26.3889 %
Kappa statistic                     0.4179
Mean absolute error                  0.1759
Root mean squared error              0.4194
Relative absolute error              58.4354 %
Root relative squared error          107.592 %
Total Number of Instances           288
Ignored Class Unknown Instances      115

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000  ?          ?        ?          ?        ?        ?        ?
      0.614    0.198    0.626    0.614    0.620    0.418    0.746    0.481    No
      0.802    0.386    0.794    0.802    0.798    0.418    0.545    0.488    Yes
Weighted Avg.    0.736    0.320    0.735    0.736    0.735    0.418    0.615    0.485

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 |  a = ?
  0 62 39 |  b = No
  0 37 150 | c = Yes

```


3.1.10 Percentage Split 70%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 70.0% train, remainder test

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      190          73.6434 %
Incorrectly Classified Instances    68          26.3566 %
Kappa statistic                    0.4228
Mean absolute error                 0.1757
Root mean squared error            0.4192
Relative absolute error             58.2869 %
Root relative squared error        107.3686 %
Total Number of Instances          258
Ignored Class Unknown Instances      98

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        ?        ?          ?        ?          ?        ?        ?        ?
      0.626    0.204    0.626      0.626    0.626      0.423    0.749    0.488    No
      0.796    0.374    0.796      0.796    0.796      0.423    0.549    0.495    Yes
Weighted Avg.  0.736    0.314    0.736      0.736    0.736      0.423    0.620    0.493

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0  57  34 |   b = No
  0  34 133 |   c = Yes

```

3.1.11 Percentage Split 80%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 80.0% train, remainder test

=== Classifier model (full training set) ===

```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      127          76.0479 %
Incorrectly Classified Instances    40           23.9521 %
Kappa statistic                    0.5164
Mean absolute error                 0.1597
Root mean squared error             0.3996
Relative absolute error             53.2298 %
Root relative squared error        103.217 %
Total Number of Instances         167
Ignored Class Unknown Instances    70

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000    ?          ?        ?          ?        ?        ?        ?
          0.860    0.291    0.605    0.860    0.710    0.540    0.721    0.373    No
          0.709    0.140    0.907    0.709    0.796    0.540    0.717    0.624    Yes
Weighted Avg.   0.760    0.192    0.804    0.760    0.767    0.540    0.718    0.539

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 | a = ?
  0 49  8 | b = No
  0 32 78 | c = Yes

```

3.1.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 90.0% train, remainder test

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      65          76.4706 %
Incorrectly Classified Instances    20           23.5294 %
Kappa statistic                    0.5184
Mean absolute error                 0.1569

```

```

Root mean squared error          0.3961
Relative absolute error          53.4562 %
Root relative squared error      104.91 %
Total Number of Instances        85
Ignored Class Unknown Instances  34

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000  ?          ?          ?          ?         ?         ?         ?
      0.885    0.288    0.575    0.885    0.697    0.551    0.738    0.359    No
      0.712    0.115    0.933    0.712    0.808    0.551    0.723    0.658    Yes
Weighted Avg.   0.765    0.168    0.824    0.765    0.774    0.551    0.727    0.567

=== Confusion Matrix ===

 a  b  c  <-- classified as
0  0  0 | a = ?
0 23  3 | b = No
0 17 42 | c = Yes

```

3.1.13 Leave one out Fold

```

=== Run information ===

Scheme:      weka.classifiers.rules.OneR -B 6
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    1186-fold cross-validation

=== Classifier model (full training set) ===

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise:
? -> ?
No -> No
Yes -> Yes
? -> Yes
(615/836 instances correct)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      615          73.5646 %
Incorrectly Classified Instances    221          26.4354 %
Kappa statistic                    0.4693
Mean absolute error                 0.1762
Root mean squared error             0.4198
Relative absolute error             58.7542 %
Root relative squared error         108.4305 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000  ?          ?          ?          ?         ?         ?         ?
      0.838    0.317    0.576    0.838    0.683    0.494    0.696    0.350    No
      0.683    0.162    0.891    0.683    0.773    0.494    0.709    0.620    Yes
Weighted Avg.   0.736    0.215    0.784    0.736    0.743    0.494    0.704    0.528

=== Confusion Matrix ===

 a  b  c  <-- classified as
0  0  0 | a = ?
0 238 46 | b = No
0 175 377 | c = Yes

```

3.1.14 Conclusion

ONE RULE			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	73.56%	
	10 folds	73.56%	
	15 folds	72.36%	
	20 folds	73.56%	
	25 folds	73.56%	
	30 folds	73.56%	
Percentage Split	50.00%	72.31%	
	60.00%	72.30%	
	66.00%	73.61%	
	70.00%	73.64%	BEST
	80.00%	76.04%	ignored
	90.00%	76.47%	ignored
Leave One Out Fold		73.56%	

One Rule did very good here despite being last, and the generated rule was the following **do you consider yourself to be a fan of the star trek franchise ?** with *Yes* – > *Yes* and *No* – > *No*, not mentioning missing values.

The best performance was by **Percentage Split 70%** with score of **73.64%** Correctly Classified Instances. excluding 80% and 90% results due to over-fitting problem.

3.2 Naive Bayes "default settings"

3.2.1 Cross Validation (5 Folds)

=== Run information ===

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   5-fold cross-validation

```

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0

```

none                                1.0  181.0  127.0
[total]                             4.0  287.0  549.0

are-you-familiar-with-the-expanded-universe
?                                  1.0    1.0    1.0
No                                1.0  256.0  361.0
Yes                                1.0   29.0  186.0
[total]                           3.0  286.0  548.0

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
?                                  1.0    1.0    1.0
No                                1.0   23.0   93.0
Yes                                1.0    7.0   94.0
[total]                           3.0   31.0  188.0

gender
?                                  1.0    1.0    1.0
Female                            1.0  160.0  239.0
Male                              1.0  121.0  304.0
[total]                           3.0  282.0  544.0

age
18-29                             1.0   57.0  125.0
30-44                             1.0   58.0  151.0
45-60                             1.0   87.0  155.0
60                                 1.0   81.0  114.0
?                                  1.0    1.0    1.0
[total]                           5.0  284.0  546.0

household-income
0-24.999                          1.0   38.0   62.0
100.000-149.999                  1.0   37.0   80.0
150.000                          1.0   23.0   56.0
25.000-49.999                   1.0   64.0   85.0
50.000-99.999                   1.0   83.0  157.0
?                                  1.0    1.0    1.0
[total]                           6.0  246.0  441.0

education
?                                  1.0    1.0    1.0
AssociateDegree                  1.0   84.0  172.0
BachelorDegree                  1.0   91.0  173.0
GraduateDegree                  1.0   75.0  153.0
HighSchool                      1.0   31.0   42.0
LessThanHighSchool              1.0    1.0    4.0
[total]                           6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                                  1.0    1.0    1.0
No                                1.0  239.0  176.0
Yes                                1.0   46.0  371.0
[total]                           3.0  286.0  548.0

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
 === Summary ===

```

Correctly Classified Instances      655          78.3493 %
Incorrectly Classified Instances    181          21.6507 %
Kappa statistic                    0.4999
Mean absolute error                 0.1845
Root mean squared error             0.3129
Relative absolute error             61.554 %
Root relative squared error         80.9063 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.609	0.127	0.712	0.609	0.657	0.503	0.753	0.552	No
	0.873	0.391	0.813	0.873	0.842	0.503	0.859	0.856	Yes
Weighted Avg.	0.783	0.301	0.779	0.783	0.779	0.503	0.823	0.753	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0 |   a = ?
0 173 111 |   b = No
0   70 482 |   c = Yes

```

3.2.2 Cross Validation (10 Folds)

```
=== Run information ===
```

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   10-fold cross-validation

```

```
=== Classifier model (full training set) ===
```

```
Naive Bayes Classifier
```

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			
?	1.0	1.0	1.0
AssociateDegree	1.0	84.0	172.0
BachelorDegree	1.0	91.0	173.0
GraduateDegree	1.0	75.0	153.0
HighSchool	1.0	31.0	42.0
LessThanHighSchool	1.0	1.0	4.0
[total]	6.0	283.0	545.0
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise			
?	1.0	1.0	1.0
No	1.0	239.0	176.0


```

Yes                                     1.0    7.0   94.0
[total]                               3.0   31.0  188.0

gender
?                                     1.0    1.0    1.0
Female                               1.0  160.0  239.0
Male                                 1.0  121.0  304.0
[total]                             3.0  282.0  544.0

age
18-29                               1.0   57.0  125.0
30-44                               1.0   58.0  151.0
45-60                               1.0   87.0  155.0
60                                   1.0   81.0  114.0
?                                   1.0    1.0    1.0
[total]                             5.0  284.0  546.0

household-income
0-24.999                             1.0   38.0   62.0
100.000-149.999                     1.0   37.0   80.0
150.000                              1.0   23.0   56.0
25.000-49.999                       1.0   64.0   85.0
50.000-99.999                       1.0   83.0  157.0
?                                   1.0    1.0    1.0
[total]                             6.0  246.0  441.0

education
?                                     1.0    1.0    1.0
AssociateDegree                     1.0   84.0  172.0
BachelorDegree                     1.0   91.0  173.0
GraduateDegree                     1.0   75.0  153.0
HighSchool                         1.0   31.0   42.0
LessThanHighSchool                 1.0    1.0    4.0
[total]                             6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                                   1.0    1.0    1.0
No                                  1.0  239.0  176.0
Yes                                  1.0   46.0  371.0
[total]                             3.0  286.0  548.0

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
 === Summary ===

```

Correctly Classified Instances      658           78.7081 %
Incorrectly Classified Instances    178           21.2919 %
Kappa statistic                    0.5086
Mean absolute error                 0.185
Root mean squared error             0.3138
Relative absolute error             61.7256 %
Root relative squared error         81.1406 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.616	0.125	0.717	0.616	0.663	0.512	0.752	0.561	No
	0.875	0.384	0.816	0.875	0.844	0.512	0.858	0.854	Yes
Weighted Avg.	0.787	0.296	0.782	0.787	0.783	0.512	0.822	0.755	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0 |   a = ?
0 175 109 |   b = No
0   69 483 |   c = Yes

```

3.2.4 Cross Validation (20 Folds)

=== Run information ===

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe

```



```

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 20-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute                                     Class
                                           ?      No      Yes
                                           (0) (0.34) (0.66)
=====
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
No                                           1.0      1.0      1.0
Yes                                           1.0 285.0 553.0
[total]                                     2.0 286.0 554.0

which-character-shot-first
?                                           1.0      1.0      1.0
Greedo                                       1.0  47.0 152.0
Han                                           1.0  58.0 269.0
none                                         1.0 181.0 127.0
[total]                                     4.0 287.0 549.0

are-you-familiar-with-the-expanded-universe
?                                           1.0      1.0      1.0
No                                           1.0 256.0 361.0
Yes                                           1.0  29.0 186.0
[total]                                     3.0 286.0 548.0

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
?                                           1.0      1.0      1.0
No                                           1.0  23.0  93.0
Yes                                           1.0   7.0  94.0
[total]                                     3.0  31.0 188.0

gender
?                                           1.0      1.0      1.0
Female                                       1.0 160.0 239.0
Male                                         1.0 121.0 304.0
[total]                                     3.0 282.0 544.0

age
18-29                                       1.0  57.0 125.0
30-44                                       1.0  58.0 151.0
45-60                                       1.0  87.0 155.0
60                                           1.0  81.0 114.0
?                                           1.0      1.0      1.0
[total]                                     5.0 284.0 546.0

household-income
0-24.999                                    1.0  38.0  62.0
100.000-149.999                            1.0  37.0  80.0
150.000                                      1.0  23.0  56.0
25.000-49.999                              1.0  64.0  85.0
50.000-99.999                              1.0  83.0 157.0
?                                           1.0      1.0      1.0
[total]                                     6.0 246.0 441.0

education
?                                           1.0      1.0      1.0
AssociateDegree                            1.0  84.0 172.0
BachelorDegree                             1.0  91.0 173.0
GraduateDegree                             1.0  75.0 153.0
HighSchool                                 1.0  31.0  42.0
LessThanHighSchool                         1.0   1.0   4.0
[total]                                     6.0 283.0 545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                                           1.0      1.0      1.0
No                                           1.0 239.0 176.0
Yes                                           1.0  46.0 371.0
[total]                                     3.0 286.0 548.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      654      78.2297 %

```



```

30-44                1.0   58.0  151.0
45-60                1.0   87.0  155.0
60                   1.0   81.0  114.0
?                    1.0    1.0    1.0
[total]              5.0  284.0  546.0

household-income
0-24.999             1.0   38.0   62.0
100.000-149.999     1.0   37.0   80.0
150.000              1.0   23.0   56.0
25.000-49.999       1.0   64.0   85.0
50.000-99.999       1.0   83.0  157.0
?                    1.0    1.0    1.0
[total]              6.0  246.0  441.0

education
?                    1.0    1.0    1.0
AssociateDegree      1.0   84.0  172.0
BachelorDegree       1.0   91.0  173.0
GraduateDegree       1.0   75.0  153.0
HighSchool           1.0   31.0   42.0
LessThanHighSchool   1.0    1.0    4.0
[total]              6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                    1.0    1.0    1.0
No                   1.0  239.0  176.0
Yes                  1.0   46.0  371.0
[total]              3.0  286.0  548.0

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
 === Summary ===

```

Correctly Classified Instances      657           78.5885 %
Incorrectly Classified Instances    179           21.4115 %
Kappa statistic                    0.5089
Mean absolute error                 0.1848
Root mean squared error             0.3131
Relative absolute error             61.6645 %
Root relative squared error         80.9742 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.627	0.132	0.709	0.627	0.665	0.511	0.752	0.559	No
	0.868	0.373	0.819	0.868	0.843	0.511	0.858	0.855	Yes
Weighted Avg.	0.786	0.291	0.782	0.786	0.782	0.511	0.822	0.754	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0 | a = ?
0 178 106 | b = No
0   73 479 | c = Yes

```

3.2.6 Cross Validation (30 Folds)

=== Run information ===

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    30-fold cross-validation

```

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			
?	1.0	1.0	1.0
AssociateDegree	1.0	84.0	172.0
BachelorDegree	1.0	91.0	173.0
GraduateDegree	1.0	75.0	153.0
HighSchool	1.0	31.0	42.0
LessThanHighSchool	1.0	1.0	4.0
[total]	6.0	283.0	545.0
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise			
?	1.0	1.0	1.0
No	1.0	239.0	176.0
Yes	1.0	46.0	371.0
[total]	3.0	286.0	548.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	656	78.4689 %
Incorrectly Classified Instances	180	21.5311 %
Kappa statistic	0.5048	
Mean absolute error	0.1848	
Root mean squared error	0.3136	
Relative absolute error	61.6918 %	
Root relative squared error	81.0885 %	
Total Number of Instances	836	
Ignored Class Unknown Instances	350	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.620	0.130	0.710	0.620	0.662	0.507	0.751	0.563	No
	0.870	0.380	0.816	0.870	0.842	0.507	0.858	0.855	Yes
Weighted Avg.	0.785	0.295	0.780	0.785	0.781	0.507	0.822	0.756	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 176 108   |   b = No
0   72 480   |   c = Yes
```

3.2.7 Percentage Split 50%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:    1186
Attributes:   10
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
which-character-shot-first
are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 50.0% train, remainder test
```

```
=== Classifier model (full training set) ===
```

```
Naive Bayes Classifier
```

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0

```

50.000-99.999          1.0   83.0  157.0
?                      1.0    1.0    1.0
[total]                6.0  246.0  441.0

education
?                      1.0    1.0    1.0
AssociateDegree        1.0   84.0  172.0
BachelorDegree         1.0   91.0  173.0
GraduateDegree         1.0   75.0  153.0
HighSchool             1.0   31.0   42.0
LessThanHighSchool     1.0    1.0    4.0
[total]                6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                      1.0    1.0    1.0
No                     1.0  239.0  176.0
Yes                    1.0   46.0  371.0
[total]                3.0  286.0  548.0

```

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

```

Correctly Classified Instances      331          78.9976 %
Incorrectly Classified Instances    88           21.0024 %
Kappa statistic                    0.513
Mean absolute error                 0.1827
Root mean squared error             0.3087
Relative absolute error             60.8618 %
Root relative squared error         79.1334 %
Total Number of Instances          419
Ignored Class Unknown Instances      174

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.585	0.099	0.761	0.585	0.662	0.522	0.754	0.578	No
	0.901	0.415	0.801	0.901	0.848	0.522	0.870	0.860	Yes
Weighted Avg.	0.790	0.304	0.787	0.790	0.782	0.522	0.829	0.761	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0  86  61   |   b = No
0   27 245   |   c = Yes

```

3.2.8 Percentage Split 60%

=== Run information ===

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 60.0% train, remainder test

```

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0

```

Yes                                     1.0  285.0  553.0
[total]                               2.0  286.0  554.0

which-character-shot-first
?                                     1.0    1.0    1.0
Greedo                               1.0   47.0  152.0
Han                                   1.0   58.0  269.0
none                                  1.0  181.0  127.0
[total]                               4.0  287.0  549.0

are-you-familiar-with-the-expanded-universe
?                                     1.0    1.0    1.0
No                                    1.0  256.0  361.0
Yes                                    1.0   29.0  186.0
[total]                               3.0  286.0  548.0

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
?                                     1.0    1.0    1.0
No                                    1.0   23.0   93.0
Yes                                    1.0    7.0   94.0
[total]                               3.0   31.0  188.0

gender
?                                     1.0    1.0    1.0
Female                               1.0  160.0  239.0
Male                                  1.0  121.0  304.0
[total]                               3.0  282.0  544.0

age
18-29                                 1.0   57.0  125.0
30-44                                 1.0   58.0  151.0
45-60                                 1.0   87.0  155.0
60                                    1.0   81.0  114.0
?                                     1.0    1.0    1.0
[total]                               5.0  284.0  546.0

household-income
0-24.999                             1.0   38.0   62.0
100.000-149.999                     1.0   37.0   80.0
150.000                              1.0   23.0   56.0
25.000-49.999                       1.0   64.0   85.0
50.000-99.999                       1.0   83.0  157.0
?                                     1.0    1.0    1.0
[total]                               6.0  246.0  441.0

education
?                                     1.0    1.0    1.0
AssociateDegree                     1.0   84.0  172.0
BachelorDegree                     1.0   91.0  173.0
GraduateDegree                     1.0   75.0  153.0
HighSchool                         1.0   31.0   42.0
LessThanHighSchool                 1.0    1.0    4.0
[total]                               6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                                     1.0    1.0    1.0
No                                    1.0  239.0  176.0
Yes                                    1.0   46.0  371.0
[total]                               3.0  286.0  548.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          268           78.1341 %
Incorrectly Classified Instances        75           21.8659 %
Kappa statistic                        0.4994
Mean absolute error                     0.1813
Root mean squared error                 0.3114
Relative absolute error                 59.8708 %
Root relative squared error            78.5474 %
Total Number of Instances              343
Ignored Class Unknown Instances         131

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         ?         ?         ?         ?         ?         ?         ?         ?
0.551  0.083  0.795  0.551  0.651  0.517  0.746  0.577  No
0.917  0.449  0.776  0.917  0.841  0.517  0.885  0.874  Yes
Weighted Avg.  0.781  0.313  0.783  0.781  0.771  0.517  0.834  0.764

```

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0  70  57   |   b = No
0  18 198   |   c = Yes

```

3.2.9 Percentage Split 66%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 66.0% train, remainder test

```

```
=== Classifier model (full training set) ===
```

```
Naive Bayes Classifier
```

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			


```

?
AssociateDegree      1.0      1.0      1.0
BachelorDegree      1.0      84.0     172.0
BachelorDegree      1.0      91.0     173.0
GraduateDegree      1.0      75.0     153.0
HighSchool          1.0      31.0      42.0
LessThanHighSchool  1.0      1.0      4.0
[total]             6.0     283.0    545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?
No                  1.0      239.0    176.0
Yes                 1.0      46.0     371.0
[total]            3.0     286.0    548.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      233              80.9028 %
Incorrectly Classified Instances    55              19.0972 %
Kappa statistic                    0.5596
Mean absolute error                 0.1734
Root mean squared error             0.2974
Relative absolute error             57.5813 %
Root relative squared error        76.2906 %
Total Number of Instances          288
Ignored Class Unknown Instances      115

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
?            0.000    ?           ?           ?           ?           ?           ?           ?
0.624        0.091    0.788     0.624     0.696     0.568     0.768     0.595     No
0.909        0.376    0.817     0.909     0.861     0.568     0.892     0.888     Yes
Weighted Avg. 0.809    0.276    0.807     0.809     0.803     0.568     0.849     0.785

=== Confusion Matrix ===

 a   b   c  <-- classified as
0   0   0 |  a = ?
0  63  38 |  b = No
0  17 170 |  c = Yes

```

3.2.10 Percentage Split 70%

```
=== Run information ===  
  
Scheme:      weka.classifiers.bayes.NaiveBayes  
Relation:    StarWars  
Instances:   1186  
Attributes:  10  
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise  
             which-character-shot-first  
             are-you-familiar-with-the-expanded-universe  
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe  
             gender  
             age  
             household-income  
             education  
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise  
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise  
  
Test mode:   split 70.0% train, remainder test  
  
=== Classifier model (full training set) ===  
  
Naive Bayes Classifier  
  
Attribute                                     Class  
                                              ?       No     Yes  
                                              (0) (0.34) (0.66)  
=====
```

	?	No	Yes
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0

```

Greedo                1.0   47.0  152.0
Han                   1.0   58.0  269.0
none                  1.0  181.0  127.0
[total]               4.0  287.0  549.0

are-you-familiar-with-the-expanded-universe
?                     1.0     1.0   1.0
No                    1.0  256.0  361.0
Yes                   1.0   29.0  186.0
[total]               3.0  286.0  548.0

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
?                     1.0     1.0   1.0
No                    1.0   23.0   93.0
Yes                   1.0    7.0   94.0
[total]               3.0   31.0  188.0

gender
?                     1.0     1.0   1.0
Female                1.0  160.0  239.0
Male                  1.0  121.0  304.0
[total]               3.0  282.0  544.0

age
18-29                 1.0   57.0  125.0
30-44                 1.0   58.0  151.0
45-60                 1.0   87.0  155.0
60                    1.0   81.0  114.0
?                     1.0     1.0   1.0
[total]               5.0  284.0  546.0

household-income
0-24.999              1.0   38.0   62.0
100.000-149.999      1.0   37.0   80.0
150.000               1.0   23.0   56.0
25.000-49.999        1.0   64.0   85.0
50.000-99.999        1.0   83.0  157.0
?                     1.0     1.0   1.0
[total]               6.0  246.0  441.0

education
?                     1.0     1.0   1.0
AssociateDegree       1.0   84.0  172.0
BachelorDegree        1.0   91.0  173.0
GraduateDegree        1.0   75.0  153.0
HighSchool            1.0   31.0   42.0
LessThanHighSchool    1.0    1.0    4.0
[total]               6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                     1.0     1.0   1.0
No                    1.0  239.0  176.0
Yes                   1.0   46.0  371.0
[total]               3.0  286.0  548.0

```

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

```

Correctly Classified Instances      206          79.845 %
Incorrectly Classified Instances    52           20.155 %
Kappa statistic                    0.5426
Mean absolute error                 0.1748
Root mean squared error             0.2979
Relative absolute error             57.9698 %
Root relative squared error         76.3087 %
Total Number of Instances          258
Ignored Class Unknown Instances      98

```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.637	0.114	0.753	0.637	0.690	0.547	0.774	0.604	No
	0.886	0.363	0.818	0.886	0.851	0.547	0.891	0.890	Yes
Weighted Avg.	0.798	0.275	0.795	0.798	0.794	0.547	0.850	0.789	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0 |   a = ?

```

```

0  58  33 |   b = No
0  19 148 |   c = Yes

```

3.2.11 Percentage Split 80%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 80.0% train, remainder test

```

```
=== Classifier model (full training set) ===
```

Naive Bayes Classifier

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			
?	1.0	1.0	1.0
AssociateDegree	1.0	84.0	172.0
BachelorDegree	1.0	91.0	173.0
GraduateDegree	1.0	75.0	153.0
HighSchool	1.0	31.0	42.0

```

LessThanHighSchool          1.0    1.0    4.0
[total]                     6.0  283.0  545.0

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?                            1.0    1.0    1.0
No                           1.0  239.0  176.0
Yes                           1.0   46.0  371.0
[total]                     3.0  286.0  548.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      137          82.0359 %
Incorrectly Classified Instances    30          17.9641 %
Kappa statistic                    0.5791
Mean absolute error                0.1776
Root mean squared error            0.308
Relative absolute error            59.2057 %
Root relative squared error        79.5489 %
Total Number of Instances         167
Ignored Class Unknown Instances      70

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        ?        ?          ?        ?          ?        ?        ?        ?
0.632    0.082    0.800    0.632    0.706    0.588    0.745    0.565    No
0.918    0.368    0.828    0.918    0.871    0.588    0.875    0.868    Yes
Weighted Avg.    0.820    0.271    0.818    0.820    0.814    0.588    0.831    0.765

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |   a = ?
  0   36   21 |   b = No
  0    9  101 |   c = Yes

```

3.2.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 90.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

              Class
Attribute      ?      No      Yes
              (0) (0.34) (0.66)
=====
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
No              1.0    1.0    1.0
Yes            1.0  285.0  553.0
[total]        2.0  286.0  554.0

which-character-shot-first
?              1.0    1.0    1.0
Greedo         1.0   47.0  152.0
Han            1.0   58.0  269.0
none           1.0  181.0  127.0
[total]        4.0  287.0  549.0

```

```

are-you-familiar-with-the-expanded-universe
?
No
Yes
[total]

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
?
No
Yes
[total]

gender
?
Female
Male
[total]

age
18-29
30-44
45-60
60
?
[total]

household-income
0-24.999
100.000-149.999
150.000
25.000-49.999
50.000-99.999
?
[total]

education
?
AssociateDegree
BachelorDegree
GraduateDegree
HighSchool
LessThanHighSchool
[total]

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
?
No
Yes
[total]

```

are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			
?	1.0	1.0	1.0
AssociateDegree	1.0	84.0	172.0
BachelorDegree	1.0	91.0	173.0
GraduateDegree	1.0	75.0	153.0
HighSchool	1.0	31.0	42.0
LessThanHighSchool	1.0	1.0	4.0
[total]	6.0	283.0	545.0
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise			
?	1.0	1.0	1.0
No	1.0	239.0	176.0
Yes	1.0	46.0	371.0
[total]	3.0	286.0	548.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	71	83.5294 %
Incorrectly Classified Instances	14	16.4706 %
Kappa statistic	0.5854	
Mean absolute error	0.1616	
Root mean squared error	0.2821	
Relative absolute error	55.0644 %	
Root relative squared error	74.7112 %	
Total Number of Instances	85	
Ignored Class Unknown Instances	34	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.615	0.068	0.800	0.615	0.696	0.595	0.766	0.650	No
	0.932	0.385	0.846	0.932	0.887	0.595	0.890	0.878	Yes
Weighted Avg.	0.835	0.288	0.832	0.835	0.829	0.595	0.852	0.808	

=== Confusion Matrix ===

```

a  b  c  <-- classified as
0  0  0  |  a = ?
0 16 10  |  b = No
0  4 55  |  c = Yes

```

3.2.13 Leave one out Fold

```
=== Run information ===
```

```
Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   1186-fold cross-validation
```

```
=== Classifier model (full training set) ===
```

```
Naive Bayes Classifier
```

Attribute	Class		
	?	No	Yes
	(0)	(0.34)	(0.66)
=====			
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise			
No	1.0	1.0	1.0
Yes	1.0	285.0	553.0
[total]	2.0	286.0	554.0
which-character-shot-first			
?	1.0	1.0	1.0
Greedo	1.0	47.0	152.0
Han	1.0	58.0	269.0
none	1.0	181.0	127.0
[total]	4.0	287.0	549.0
are-you-familiar-with-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	256.0	361.0
Yes	1.0	29.0	186.0
[total]	3.0	286.0	548.0
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe			
?	1.0	1.0	1.0
No	1.0	23.0	93.0
Yes	1.0	7.0	94.0
[total]	3.0	31.0	188.0
gender			
?	1.0	1.0	1.0
Female	1.0	160.0	239.0
Male	1.0	121.0	304.0
[total]	3.0	282.0	544.0
age			
18-29	1.0	57.0	125.0
30-44	1.0	58.0	151.0
45-60	1.0	87.0	155.0
60	1.0	81.0	114.0
?	1.0	1.0	1.0
[total]	5.0	284.0	546.0
household-income			
0-24.999	1.0	38.0	62.0
100.000-149.999	1.0	37.0	80.0
150.000	1.0	23.0	56.0
25.000-49.999	1.0	64.0	85.0
50.000-99.999	1.0	83.0	157.0
?	1.0	1.0	1.0
[total]	6.0	246.0	441.0
education			
?	1.0	1.0	1.0
AssociateDegree	1.0	84.0	172.0
BachelorDegree	1.0	91.0	173.0
GraduateDegree	1.0	75.0	153.0
HighSchool	1.0	31.0	42.0
LessThanHighSchool	1.0	1.0	4.0
[total]	6.0	283.0	545.0
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise			
?	1.0	1.0	1.0
No	1.0	239.0	176.0

```

Yes                                1.0   46.0  371.0
[total]                           3.0  286.0  548.0

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      657           78.5885 %
Incorrectly Classified Instances   179           21.4115 %
Kappa statistic                    0.5071
Mean absolute error                 0.1853
Root mean squared error             0.314
Relative absolute error             61.7834 %
Root relative squared error         81.1101 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000    ?          ?          ?          ?        ?         ?         ?
0.620    0.129    0.713    0.620    0.663    0.510    0.750    0.560    No
0.871    0.380    0.817    0.871    0.843    0.510    0.857    0.854    Yes
Weighted Avg.   0.786    0.295    0.781    0.786    0.782    0.510    0.821    0.754

=== Confusion Matrix ===

 a   b   c  <-- classified as
0   0   0 |  a = ?
0 176 108 |  b = No
0   71 481 |  c = Yes

```

3.2.14 Conclusion

NAIVE BAYES			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	78.34%	
	10 folds	79.06%	
	15 folds	78.70%	
	20 folds	78.22%	
	25 folds	78.58%	
	30 folds	78.46%	
Percentage Split	50.00%	78.99%	
	60.00%	78.13%	
	66.00%	80.90%	BEST
	70.00%	79.84%	
	80.00%	82.96%	ignored
	90.00%	83.52%	ignored
Leave One Out Fold		78.58%	

Naive Bayes executed and performed very well with good results being the second best performing algorithm in our tests. the generated result was **do you consider yourself to be a fan of the star trek franchise ?** with 0.66% yes and 0.34% no.

The best result was from **Percentage Split 66%** with impressive **80.90%** Correctly Classified Instances. excluding 80% and 90% results due to to over-fitting problem.

3.3 K-nearest Neighbour "K-value = 3 and default settings"

3.3.1 Cross Validation (5 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\""
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    5-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      660           78.9474 %
Incorrectly Classified Instances    176           21.0526 %
Kappa statistic                    0.5275
Mean absolute error                 0.1911
Root mean squared error            0.322
Relative absolute error             63.7702 %
Root relative squared error        83.2745 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.676	0.152	0.696	0.676	0.686	0.528	0.734	0.458	No
	0.848	0.324	0.836	0.848	0.842	0.528	0.826	0.792	Yes
Weighted Avg.	0.789	0.266	0.788	0.789	0.789	0.528	0.795	0.679	

```

=== Confusion Matrix ===

 a   b   c   <-- classified as
0   0   0 | a = ?
0 192  92 | b = No
0   84 468 | c = Yes

```

3.3.2 Cross Validation (10 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\""
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier

```


using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	660	78.9474 %
Incorrectly Classified Instances	176	21.0526 %
Kappa statistic	0.5242	
Mean absolute error	0.192	
Root mean squared error	0.323	
Relative absolute error	64.0833 %	
Root relative squared error	83.5214 %	
Total Number of Instances	836	
Ignored Class Unknown Instances	350	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.662	0.145	0.701	0.662	0.681	0.525	0.727	0.431	No
	0.855	0.338	0.831	0.855	0.843	0.525	0.827	0.801	Yes
Weighted Avg.	0.789	0.272	0.787	0.789	0.788	0.525	0.793	0.675	

=== Confusion Matrix ===

a	b	c	<-- classified as
0	0	0	a = ?
0	188	96	b = No
0	80	472	c = Yes

3.3.3 Cross Validation (15 Folds)

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
 \"weka.core.EuclideanDistance -R first-last\""

Relation: StarWars

Instances: 1186

Attributes: 10

have-you-seen-any-of-the-6-films-in-the-star-wars-franchise

which-character-shot-first

are-you-familiar-with-the-expanded-universe

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe

gender

age

household-income

education

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise

do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise

Test mode: 15-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
 using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	660	78.9474 %
Incorrectly Classified Instances	176	21.0526 %
Kappa statistic	0.5234	
Mean absolute error	0.1923	
Root mean squared error	0.3245	
Relative absolute error	64.1777 %	
Root relative squared error	83.9075 %	
Total Number of Instances	836	
Ignored Class Unknown Instances	350	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.658	0.143	0.703	0.658	0.680	0.524	0.725	0.435	No
	0.857	0.342	0.830	0.857	0.843	0.524	0.829	0.805	Yes
Weighted Avg.	0.789	0.274	0.787	0.789	0.788	0.524	0.794	0.679	

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 187  97   |   b = No
0   79 473   |   c = Yes

```

3.3.4 Cross Validation (20 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:    StarWars
Instances:   1186
Attributes:  10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   20-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      664           79.4258 %
Incorrectly Classified Instances    172           20.5742 %
Kappa statistic                    0.5358
Mean absolute error                 0.1911
Root mean squared error             0.3211
Relative absolute error             63.7775 %
Root relative squared error         83.0383 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.673	0.143	0.707	0.673	0.690	0.536	0.729	0.434	No
	0.857	0.327	0.836	0.857	0.846	0.536	0.830	0.803	Yes
Weighted Avg.	0.794	0.265	0.792	0.794	0.793	0.536	0.796	0.677	

```

=== Confusion Matrix ===

a   b   c   <-- classified as
0   0   0   |   a = ?
0 191  93   |   b = No
0   79 473   |   c = Yes

```

3.3.5 Cross Validation (25 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:    StarWars
Instances:   1186
Attributes:  10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise

```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:      25-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      657           78.5885 %
Incorrectly Classified Instances    179           21.4115 %
Kappa statistic                    0.5148
Mean absolute error                 0.1927
Root mean squared error            0.3228
Relative absolute error            64.3086 %
Root relative squared error        83.4802 %
Total Number of Instances         836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000   ?          ?          ?          ?        ?        ?        ?
          0.651   0.145   0.698     0.651    0.674     0.516    0.727    0.428    No
          0.855   0.349   0.827     0.855    0.841     0.516    0.831    0.811    Yes
Weighted Avg.   0.786   0.279   0.783     0.786    0.784     0.516    0.796    0.681

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 185   99 |  b = No
  0   80 472 |  c = Yes

```

3.3.6 Cross Validation (30 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    30-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      662           79.1866 %
Incorrectly Classified Instances    174           20.8134 %
Kappa statistic                    0.528
Mean absolute error                 0.1915
Root mean squared error            0.3221
Relative absolute error            63.9036 %
Root relative squared error        83.2787 %
Total Number of Instances         836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000   ?          ?          ?          ?        ?        ?        ?
          0.651   0.145   0.698     0.651    0.674     0.516    0.727    0.428    No
          0.855   0.349   0.827     0.855    0.841     0.516    0.831    0.811    Yes
Weighted Avg.   0.786   0.279   0.783     0.786    0.784     0.516    0.796    0.681

```

	?	0.000	?	?	?	?	?	?	?
	0.658	0.139	0.708	0.658	0.682	0.529	0.727	0.429	No
	0.861	0.342	0.830	0.861	0.845	0.529	0.832	0.810	Yes
Weighted Avg.	0.792	0.273	0.789	0.792	0.790	0.529	0.797	0.680	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0 |   a = ?
0 187  97 |   b = No
0   77 475 |   c = Yes

```

3.3.7 Percentage Split 50%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 50.0% train, remainder test

```

```
=== Classifier model (full training set) ===
```

```

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

```

```
Time taken to build model: 0 seconds
```

```
=== Evaluation on test split ===
```

```
Time taken to test model on test split: 0.05 seconds
```

```
=== Summary ===
```

Correctly Classified Instances	323	77.0883 %
Incorrectly Classified Instances	96	22.9117 %
Kappa statistic	0.4906	
Mean absolute error	0.193	
Root mean squared error	0.3261	
Relative absolute error	64.2684 %	
Root relative squared error	83.5959 %	
Total Number of Instances	419	
Ignored Class Unknown Instances	174	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.646	0.162	0.683	0.646	0.664	0.491	0.737	0.446	No
	0.838	0.354	0.814	0.838	0.826	0.491	0.817	0.784	Yes
Weighted Avg.	0.771	0.286	0.768	0.771	0.769	0.491	0.789	0.665	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0 |   a = ?
0  95  52 |   b = No
0   44 228 |   c = Yes

```

3.3.8 Percentage Split 60%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:     StarWars
Instances:    1186
Attributes:   10

```

```

have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
which-character-shot-first
are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:      split 60.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.09 seconds

=== Summary ===

Correctly Classified Instances      270          78.7172 %
Incorrectly Classified Instances    73           21.2828 %
Kappa statistic                    0.5259
Mean absolute error                 0.1943
Root mean squared error             0.3319
Relative absolute error             64.159 %
Root relative squared error        83.7239 %
Total Number of Instances          343
Ignored Class Unknown Instances      131

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000    ?          ?        ?          ?        ?        ?        ?
          0.622    0.116    0.760      0.622    0.684      0.532    0.734    0.483    No
          0.884    0.378    0.799      0.884    0.840      0.532    0.827    0.784    Yes
Weighted Avg.    0.787    0.281    0.785      0.787    0.782      0.532    0.793    0.672

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0  79  48 |   b = No
  0  25 191 |   c = Yes

```

3.3.9 Percentage Split 66%

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\""
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.06 seconds

```

```

=== Summary ===

Correctly Classified Instances      229      79.5139 %
Incorrectly Classified Instances    59      20.4861 %
Kappa statistic                    0.5429
Mean absolute error                 0.1895
Root mean squared error             0.3243
Relative absolute error             62.9318 %
Root relative squared error         83.1797 %
Total Number of Instances          288
Ignored Class Unknown Instances    115

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000  ?          ?        ?          ?        ?        ?        ?
      0.673    0.139  0.723     0.673   0.697     0.544    0.733    0.450    No
      0.861    0.327  0.830     0.861   0.845     0.544    0.826    0.789    Yes
Weighted Avg.  0.795    0.261  0.793     0.795   0.793     0.544    0.794    0.670

=== Confusion Matrix ===

 a   b   c  <-- classified as
0   0   0 |  a = ?
0  68  33 |  b = No
0  26 161 |  c = Yes

```

3.3.10 Percentage Split 70%

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
\weka.core.EuclideanDistance -R first-last\"
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 70.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correctly Classified Instances      203      78.6822 %
Incorrectly Classified Instances    55      21.3178 %
Kappa statistic                    0.5272
Mean absolute error                 0.1936
Root mean squared error             0.3267
Relative absolute error             64.2131 %
Root relative squared error         83.6936 %
Total Number of Instances          258
Ignored Class Unknown Instances    98

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000  ?          ?        ?          ?        ?        ?        ?
      0.670    0.150  0.709     0.670   0.689     0.528    0.732    0.459    No
      0.850    0.330  0.826     0.850   0.838     0.528    0.826    0.814    Yes
Weighted Avg.  0.787    0.266  0.785     0.787   0.785     0.528    0.793    0.689

=== Confusion Matrix ===

 a   b   c  <-- classified as
0   0   0 |  a = ?

```

```

0 61 30 | b = No
0 25 142 | c = Yes

```

3.3.11 Percentage Split 80%

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:    StarWars
Instances:   1186
Attributes:  10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 80.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correctly Classified Instances      132           79.0419 %
Incorrectly Classified Instances    35           20.9581 %
Kappa statistic                    0.5197
Mean absolute error                 0.1938
Root mean squared error             0.3322
Relative absolute error             64.6186 %
Root relative squared error         85.8103 %
Total Number of Instances          167
Ignored Class Unknown Instances      70

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        0.000    ?          ?         ?          ?        ?         ?         ?
              0.632    0.127    0.720      0.632    0.673      0.522    0.702    0.424    No
              0.873    0.368    0.821      0.873    0.846      0.522    0.824    0.789    Yes
Weighted Avg.   0.790    0.286    0.786      0.790    0.787      0.522    0.783    0.664

=== Confusion Matrix ===

a  b  c  <-- classified as
0  0  0 | a = ?
0 36 21 | b = No
0 14 96 | c = Yes

```

3.3.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:    StarWars
Instances:   1186
Attributes:  10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise

```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:      split 90.0% train, remainder test

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.04 seconds

=== Summary ===

Correctly Classified Instances      72          84.7059 %
Incorrectly Classified Instances    13          15.2941 %
Kappa statistic                    0.6194
Mean absolute error                 0.1798
Root mean squared error             0.3023
Relative absolute error             61.2659 %
Root relative squared error         80.0827 %
Total Number of Instances          85
Ignored Class Unknown Instances      34

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000    ?         ?         ?         ?         ?         ?         ?
          0.654    0.068    0.810    0.654    0.723    0.626    0.731    0.426    No
          0.932    0.346    0.859    0.932    0.894    0.626    0.847    0.806    Yes
Weighted Avg.    0.847    0.261    0.844    0.847    0.842    0.626    0.811    0.690

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 | a = ?
  0 17  9 | b = No
  0  4 55 | c = Yes

```

3.3.13 Leave one out Fold

```

=== Run information ===

Scheme:      weka.classifiers.lazy.IBk -K 3 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
              \"weka.core.EuclideanDistance -R first-last\"
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    1186-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 3 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      662          79.1866 %
Incorrectly Classified Instances    174          20.8134 %
Kappa statistic                    0.528
Mean absolute error                 0.1924
Root mean squared error             0.3232
Relative absolute error             64.1533 %
Root relative squared error         83.4779 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

```



```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.658	0.139	0.708	0.658	0.682	0.529	0.723	0.425	No
	0.861	0.342	0.830	0.861	0.845	0.529	0.830	0.808	Yes
Weighted Avg.	0.792	0.273	0.789	0.792	0.790	0.529	0.794	0.678	

```
=== Confusion Matrix ===
```

```

a   b   c  <-- classified as
0   0   0 |  a = ?
0 187  97 |  b = No
0   77 475 |  c = Yes

```

3.3.14 Conclusion

KNN			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	78.94%	
	10 folds	78.94%	
	15 folds	78.94%	
	20 folds	79.42%	
	25 folds	78.58%	
	30 folds	79.18%	
Percentage Split	50.00%	77.08%	
	60.00%	78.71%	
	66.00%	79.51%	BEST
	70.00%	78.31%	
	80.00%	79.04%	ignored
	90.00%	84.70%	ignored
Leave One Out Fold		79.18%	

KNN our 3rd best performing algorithm did very well as expected, it is noted that the K-value used during the tests was 3. best result was by **Percentage Split 66%** with **79.51%** Correctly Classified Instances. excluding 80% and 90% results due to over-fitting problem.

3.4 C4.5 "default settings"

3.4.1 Cross Validation (5 Folds)

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   5-fold cross-validation

```

```
=== Classifier model (full training set) ===
```

```

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      663          79.3062 %
Incorrectly Classified Instances   173          20.6938 %
Kappa statistic                    0.5151
Mean absolute error                 0.1959
Root mean squared error             0.3217
Relative absolute error             65.3627 %
Root relative squared error         83.179 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000    ?          ?         ?          ?        ?         ?         ?
          0.595    0.105    0.744    0.595    0.661    0.522    0.723    0.553    No
          0.895    0.405    0.811    0.895    0.851    0.522    0.827    0.773    Yes
Weighted Avg.    0.793    0.303    0.789    0.793    0.787    0.522    0.792    0.698

=== Confusion Matrix ===

  a    b    c    <-- classified as
  0    0    0 |    a = ?
  0 169 115 |    b = No
  0  58 494 |    c = Yes

```

3.4.2 Cross Validation (10 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    10-fold cross-validation

```

```

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      653           78.11 %
Incorrectly Classified Instances    183           21.89 %
Kappa statistic                    0.4952
Mean absolute error                 0.1957
Root mean squared error             0.3211
Relative absolute error             65.3121 %
Root relative squared error        83.0234 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                ?         0.000   ?          ?         ?          ?        ?         ?         ?
                0.609    0.130   0.706     0.609    0.654     0.498    0.732    0.540    No
                0.870    0.391   0.812     0.870    0.840     0.498    0.820    0.781    Yes
Weighted Avg.   0.781    0.302   0.776     0.781    0.777     0.498    0.790    0.699

=== Confusion Matrix ===

  a   b   c   <-- classified as
0   0   0 |   a = ?
0 173 111 |   b = No
0   72 480 |   c = Yes

```

3.4.3 Cross Validation (15 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education

```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 15-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ? : No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ? : No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ? : Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ? : Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      658          78.7081 %
Incorrectly Classified Instances    178          21.2919 %
Kappa statistic                    0.5006
Mean absolute error                 0.1963
Root mean squared error             0.3196
Relative absolute error             65.5085 %
Root relative squared error         82.6352 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000    ?          ?         ?          ?         ?         ?         ?
          0.585    0.109    0.735    0.585    0.651    0.507    0.725    0.551    No
          0.891    0.415    0.807    0.891    0.847    0.507    0.825    0.775    Yes
Weighted Avg.    0.787    0.311    0.782    0.787    0.780    0.507    0.791    0.699

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 166 118 |  b = No
  0   60 492 |  c = Yes

```

3.4.4 Cross Validation (20 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender

```

```

age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 20-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ? : No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ? : No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ? : Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ? : Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      663          79.3062 %
Incorrectly Classified Instances    173          20.6938 %
Kappa statistic                    0.5203
Mean absolute error                 0.1932
Root mean squared error             0.3164
Relative absolute error             64.4757 %
Root relative squared error         81.8066 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?      0.000    ?           ?           ?           ?           ?           ?           ?
      0.616   0.116   0.732     0.616     0.669     0.524     0.734     0.547     No
      0.884   0.384   0.817     0.884     0.849     0.524     0.823     0.773     Yes
Weighted Avg.   0.793   0.293   0.788     0.793     0.788     0.524     0.793     0.696

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 175 109 |   b = No
  0  64 488 |   c = Yes

```

3.4.5 Cross Validation (25 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first

```

```

are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 25-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      653          78.11 %
Incorrectly Classified Instances    183          21.89 %
Kappa statistic                    0.4889
Mean absolute error                 0.1958
Root mean squared error             0.3204
Relative absolute error              65.3638 %
Root relative squared error          82.8428 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         ?         ?         ?         ?         ?         ?         ?         ?
      0.585    0.118    0.719    0.585    0.645    0.494    0.720    0.532    No
      0.882    0.415    0.805    0.882    0.842    0.494    0.815    0.760    Yes
Weighted Avg.   0.781    0.314    0.776    0.781    0.775    0.494    0.783    0.682

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 166 118 |   b = No
  0  65 487 |   c = Yes

```

3.4.6 Cross Validation (30 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186

```

```

Attributes: 10
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
which-character-shot-first
are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 30-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      663          79.3062 %
Incorrectly Classified Instances    173          20.6938 %
Kappa statistic                    0.5142
Mean absolute error                 0.1946
Root mean squared error             0.3171
Relative absolute error             64.9425 %
Root relative squared error         82.0056 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000    ?          ?          ?          ?         ?         ?         ?
      0.592    0.103    0.747    0.592    0.660    0.521    0.722    0.527    No
      0.897    0.408    0.810    0.897    0.851    0.521    0.821    0.766    Yes
Weighted Avg.   0.793    0.305    0.789    0.793    0.786    0.521    0.787    0.685

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 168 116 |   b = No
  0  57 495 |   c = Yes

```

3.4.7 Percentage Split 50%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise

Test mode:    split 50.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      334          79.7136 %
Incorrectly Classified Instances    85          20.2864 %
Kappa statistic                    0.5193
Mean absolute error                 0.2024
Root mean squared error            0.3129
Relative absolute error             67.4117 %
Root relative squared error        80.2223 %
Total Number of Instances          419
Ignored Class Unknown Instances      174

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?      0.000    ?          ?          ?          ?          ?          ?          ?
      0.551    0.070    0.810    0.551    0.656    0.539    0.744    0.592    No
      0.930    0.449    0.793    0.930    0.856    0.539    0.858    0.816    Yes
Weighted Avg.   0.797    0.316    0.799    0.797    0.786    0.539    0.818    0.737

=== Confusion Matrix ===

a   b   c   <-- classified as
0   0   0 |   a = ?
0  81  66 |   b = No
0  19 253 |   c = Yes

```


3.4.8 Percentage Split 60%

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise

Test mode:   split 60.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
|   are-you-familiar-with-the-expanded-universe = ? : No (0.0)
|   are-you-familiar-with-the-expanded-universe = No
|   |   which-character-shot-first = ? : No (0.0)
|   |   which-character-shot-first = Greedo: Yes (65.63/30.08)
|   |   which-character-shot-first = Han
|   |   |   household-income = 0-24.999
|   |   |   |   age = 18-29: Yes (12.26/1.44)
|   |   |   |   age = 30-44: No (4.11/0.89)
|   |   |   |   age = 45-60: Yes (2.45/1.22)
|   |   |   |   age = 60: No (0.22/0.0)
|   |   |   |   age = ? : Yes (0.0)
|   |   |   household-income = 100.000-149.999: Yes (12.69/3.73)
|   |   |   household-income = 150.000: Yes (12.69/2.73)
|   |   |   household-income = 25.000-49.999: No (15.23/7.36)
|   |   |   household-income = 50.000-99.999: No (29.19/11.52)
|   |   |   household-income = ? : Yes (0.0)
|   |   which-character-shot-first = none: No (189.82/38.59)
|   are-you-familiar-with-the-expanded-universe = Yes
|   |   age = 18-29: Yes (30.29/4.04)
|   |   age = 30-44: Yes (19.18/1.02)
|   |   age = 45-60: Yes (11.11/3.01)
|   |   age = 60: No (12.12/1.1)
|   |   age = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves   :   21

Size of the tree   :   27

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      271           79.0087 %
Incorrectly Classified Instances    72           20.9913 %
Kappa statistic                    0.5136
Mean absolute error                 0.2018
Root mean squared error             0.3175
Relative absolute error             66.6419 %
Root relative squared error         80.0825 %
Total Number of Instances          343
Ignored Class Unknown Instances      131

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        0.000    ?          ?        ?          ?        ?        ?        ?
              0.535    0.060    0.840     0.535    0.654      0.540    0.742    0.612    No
              0.940    0.465    0.775     0.940    0.849      0.540    0.867    0.819    Yes
Weighted Avg.   0.790    0.315    0.799     0.790    0.777      0.540    0.821    0.742

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?

```

```

0  68  59 |   b = No
0  13 203 |   c = Yes

```

3.4.9 Percentage Split 66%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 66.0% train, remainder test

```

```
=== Classifier model (full training set) ===
```

```

J48 pruned tree
-----

```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
|   are-you-familiar-with-the-expanded-universe = ?: No (0.0)
|   are-you-familiar-with-the-expanded-universe = No
|   |   which-character-shot-first = ?: No (0.0)
|   |   which-character-shot-first = Greedo: Yes (65.63/30.08)
|   |   which-character-shot-first = Han
|   |   |   household-income = 0-24.999
|   |   |   |   age = 18-29: Yes (12.26/1.44)
|   |   |   |   age = 30-44: No (4.11/0.89)
|   |   |   |   age = 45-60: Yes (2.45/1.22)
|   |   |   |   age = 60: No (0.22/0.0)
|   |   |   |   age = ?: Yes (0.0)
|   |   |   household-income = 100.000-149.999: Yes (12.69/3.73)
|   |   |   household-income = 150.000: Yes (12.69/2.73)
|   |   |   household-income = 25.000-49.999: No (15.23/7.36)
|   |   |   household-income = 50.000-99.999: No (29.19/11.52)
|   |   |   household-income = ?: Yes (0.0)
|   |   |   which-character-shot-first = none: No (189.82/38.59)
|   |   are-you-familiar-with-the-expanded-universe = Yes
|   |   |   age = 18-29: Yes (30.29/4.04)
|   |   |   age = 30-44: Yes (19.18/1.02)
|   |   |   age = 45-60: Yes (11.11/3.01)
|   |   |   age = 60: No (12.12/1.1)
|   |   |   age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

```

```
Number of Leaves : 21
```

```
Size of the tree : 27
```

```
Time taken to build model: 0.01 seconds
```

```
=== Evaluation on test split ===
```

```
Time taken to test model on test split: 0 seconds
```

```
=== Summary ===
```

```

Correctly Classified Instances      231           80.2083 %
Incorrectly Classified Instances    57           19.7917 %
Kappa statistic                    0.5436
Mean absolute error                 0.193
Root mean squared error             0.304
Relative absolute error             64.1092 %
Root relative squared error         77.9879 %
Total Number of Instances          288
Ignored Class Unknown Instances      115

```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
?	?	0.000	?	?	?	?	?	?	?
0.614	0.096	0.775	0.614	0.685	0.551	0.768	0.639		No
0.904	0.386	0.813	0.904	0.856	0.551	0.875	0.839		Yes

```
Weighted Avg.    0.802    0.284    0.799    0.802    0.796    0.551    0.838    0.769
```

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0  62  39   |   b = No
0  18 169   |   c = Yes
```

3.4.10 Percentage Split 70%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 70.0% train, remainder test
```

```
=== Classifier model (full training set) ===
```

```
J48 pruned tree
-----
```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ?: No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ?: Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)
```

```
Number of Leaves : 21
```

```
Size of the tree : 27
```

```
Time taken to build model: 0.02 seconds
```

```
=== Evaluation on test split ===
```

```
Time taken to test model on test split: 0 seconds
```

```
=== Summary ===
```

```

Correctly Classified Instances      207          80.2326 %
Incorrectly Classified Instances    51          19.7674 %
Kappa statistic                    0.5479
Mean absolute error                 0.193
Root mean squared error             0.3021
Relative absolute error             64.015 %
Root relative squared error         77.3683 %
Total Number of Instances          258
Ignored Class Unknown Instances      98
```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.626	0.102	0.770	0.626	0.691	0.554	0.775	0.638	No
	0.898	0.374	0.815	0.898	0.855	0.554	0.884	0.860	Yes
Weighted Avg.	0.802	0.278	0.799	0.802	0.797	0.554	0.846	0.782	

```
=== Confusion Matrix ===
```

```

a   b   c  <-- classified as
0   0   0 |  a = ?
0  57  34 |  b = No
0  17 150 |  c = Yes

```

3.4.11 Percentage Split 80%

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 80.0% train, remainder test

```

```
=== Classifier model (full training set) ===
```

```
J48 pruned tree
-----
```

```

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ?: No (0.0)
| | are-you-familiar-with-the-expanded-universe = No
| | | which-character-shot-first = ?: No (0.0)
| | | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | | which-character-shot-first = Han
| | | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ?: Yes (0.0)
| | | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | | household-income = 150.000: Yes (12.69/2.73)
| | | | household-income = 25.000-49.999: No (15.23/7.36)
| | | | household-income = 50.000-99.999: No (29.19/11.52)
| | | | household-income = ?: Yes (0.0)
| | | which-character-shot-first = none: No (189.82/38.59)
| | are-you-familiar-with-the-expanded-universe = Yes
| | | age = 18-29: Yes (30.29/4.04)
| | | age = 30-44: Yes (19.18/1.02)
| | | age = 45-60: Yes (11.11/3.01)
| | | age = 60: No (12.12/1.1)
| | | age = ?: Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

```

```
Number of Leaves : 21
```

```
Size of the tree : 27
```

```
Time taken to build model: 0.01 seconds
```

```
=== Evaluation on test split ===
```

```
Time taken to test model on test split: 0 seconds
```

```
=== Summary ===
```

Correctly Classified Instances	133	79.6407 %
Incorrectly Classified Instances	34	20.3593 %
Kappa statistic	0.5314	
Mean absolute error	0.1935	

```

Root mean squared error          0.3158
Relative absolute error          64.4937 %
Root relative squared error      81.5672 %
Total Number of Instances       167
Ignored Class Unknown Instances 70

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000   ?          ?         ?          ?        ?         ?         ?
      0.632    0.118    0.735    0.632    0.679      0.535    0.759    0.593    No
      0.882    0.368    0.822    0.882    0.851      0.535    0.845    0.830    Yes
Weighted Avg.   0.796    0.283    0.792    0.796    0.792      0.535    0.815    0.749

=== Confusion Matrix ===

 a  b  c  <-- classified as
0  0  0 | a = ?
0 36 21 | b = No
0 13 97 | c = Yes

```

3.4.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 90.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
| are-you-familiar-with-the-expanded-universe = ? : No (0.0)
| are-you-familiar-with-the-expanded-universe = No
| | which-character-shot-first = ? : No (0.0)
| | which-character-shot-first = Greedo: Yes (65.63/30.08)
| | which-character-shot-first = Han
| | | household-income = 0-24.999
| | | | age = 18-29: Yes (12.26/1.44)
| | | | age = 30-44: No (4.11/0.89)
| | | | age = 45-60: Yes (2.45/1.22)
| | | | age = 60: No (0.22/0.0)
| | | | age = ? : Yes (0.0)
| | | household-income = 100.000-149.999: Yes (12.69/3.73)
| | | household-income = 150.000: Yes (12.69/2.73)
| | | household-income = 25.000-49.999: No (15.23/7.36)
| | | household-income = 50.000-99.999: No (29.19/11.52)
| | | household-income = ? : Yes (0.0)
| | which-character-shot-first = none: No (189.82/38.59)
| are-you-familiar-with-the-expanded-universe = Yes
| | age = 18-29: Yes (30.29/4.04)
| | age = 30-44: Yes (19.18/1.02)
| | age = 45-60: Yes (11.11/3.01)
| | age = 60: No (12.12/1.1)
| | age = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

```

```

=== Summary ===

Correctly Classified Instances      72           84.7059 %
Incorrectly Classified Instances    13           15.2941 %
Kappa statistic                    0.6105
Mean absolute error                 0.1678
Root mean squared error             0.2689
Relative absolute error             57.1693 %
Root relative squared error         71.219 %
Total Number of Instances          85
Ignored Class Unknown Instances      34

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000  ?           ?         ?           ?         ?         ?         ?
      0.615    0.051  0.842    0.615    0.711    0.624    0.815    0.687    No
      0.949    0.385  0.848    0.949    0.896    0.624    0.897    0.874    Yes
Weighted Avg.  0.847    0.283  0.847    0.847    0.839    0.624    0.872    0.817

=== Confusion Matrix ===

 a  b  c  <-- classified as
0  0  0 | a = ?
0 16 10 | b = No
0  3 56 | c = Yes

```

3.4.13 Leave one out Fold

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    1186-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = No
|  are-you-familiar-with-the-expanded-universe = ? : No (0.0)
|  are-you-familiar-with-the-expanded-universe = No
|  |  which-character-shot-first = ? : No (0.0)
|  |  which-character-shot-first = Greedo: Yes (65.63/30.08)
|  |  which-character-shot-first = Han
|  |  |  household-income = 0-24.999
|  |  |  |  age = 18-29: Yes (12.26/1.44)
|  |  |  |  age = 30-44: No (4.11/0.89)
|  |  |  |  age = 45-60: Yes (2.45/1.22)
|  |  |  |  age = 60: No (0.22/0.0)
|  |  |  |  age = ? : Yes (0.0)
|  |  |  household-income = 100.000-149.999: Yes (12.69/3.73)
|  |  |  household-income = 150.000: Yes (12.69/2.73)
|  |  |  household-income = 25.000-49.999: No (15.23/7.36)
|  |  |  household-income = 50.000-99.999: No (29.19/11.52)
|  |  |  household-income = ? : Yes (0.0)
|  |  which-character-shot-first = none: No (189.82/38.59)
|  are-you-familiar-with-the-expanded-universe = Yes
|  |  age = 18-29: Yes (30.29/4.04)
|  |  age = 30-44: Yes (19.18/1.02)
|  |  age = 45-60: Yes (11.11/3.01)
|  |  age = 60: No (12.12/1.1)
|  |  age = ? : Yes (0.0)
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (419.01/45.5)

Number of Leaves : 21

Size of the tree : 27

```

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

```

Correctly Classified Instances      674          80.622 %
Incorrectly Classified Instances    162          19.378 %
Kappa statistic                    0.5496
Mean absolute error                0.187
Root mean squared error            0.3108
Relative absolute error            62.3589 %
Root relative squared error        80.2686 %
Total Number of Instances         836
Ignored Class Unknown Instances    350

```

=== Detailed Accuracy By Class ===

```

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000   ?          ?         ?          ?        ?         ?         ?
          0.630    0.103   0.758     0.630    0.688     0.555    0.705    0.584     No
          0.897    0.370   0.825     0.897    0.859     0.555    0.803    0.809     Yes
Weighted Avg.    0.806    0.279   0.802     0.806    0.801     0.555    0.770    0.733

```

=== Confusion Matrix ===

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 179 105   |   b = No
0   57 495   |   c = Yes

```

3.4.14 Conclusion

C4.5			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	79.30%	
	10 folds	78.11%	
	15 folds	78.70%	
	20 folds	79.30%	
	25 folds	78.11%	
	30 folds	79.30%	
Percentage Split	50.00%	79.71%	
	60.00%	79.00%	
	66.00%	80.20%	
	70.00%	80.23%	
	80.00%	79.64%	ignored
	90.00%	84.70%	ignored
Leave One Out Fold		80.62%	BEST

The favourite of algorithm by many including our professor which explains why it is in almost every exam. well the results speak for themselves as the best performing algorithm in our tests. the generated Tree can viewed in [3.1](#).

best performance results from our evaluation methods came from **Leave One Out Fold** with **80.62%** Correctly Classified Instances. excluding 80% and 90% results due to to over-fitting problem. it has to be noted that **Percentage Split 66%** and **Percentage Split 70%** also did well with **80.20%** and **80.23%** Correctly Classified Instances respectively.

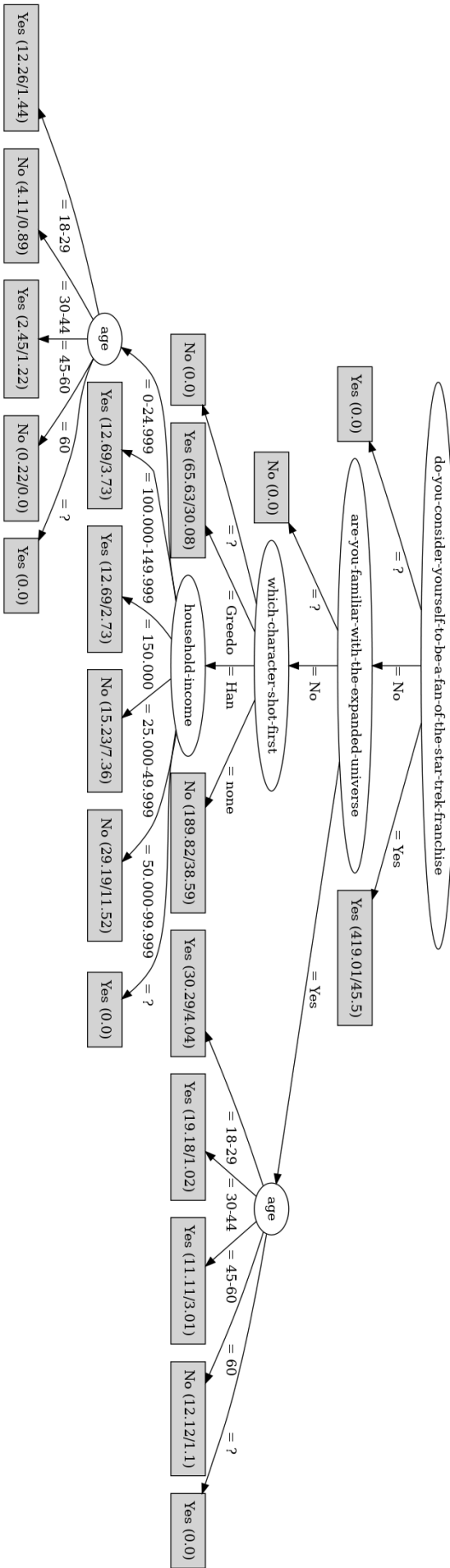


FIGURE 3.1: Generated Tree for C4.5

3.5 PART "default settings"

3.5.1 Cross Validation (5 Folds)

```

=== Run information ===

Scheme:      weka.Classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise

Test mode:    5-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      662           79.1866 %
Incorrectly Classified Instances    174           20.8134 %
Kappa statistic                    0.5188
Mean absolute error                 0.1951

```

```

Root mean squared error          0.3237
Relative absolute error          65.1064 %
Root relative squared error      83.7103 %
Total Number of Instances       836
Ignored Class Unknown Instances  350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000  ?           ?         ?           ?         ?         ?         ?
          0.620    0.120    0.727    0.620    0.669       0.522    0.727    0.528    No
          0.880    0.380    0.818    0.880    0.848       0.522    0.832    0.806    Yes
Weighted Avg.    0.792    0.292    0.787    0.792    0.787       0.522    0.796    0.712

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 176 108 |  b = No
  0  66 486 |  c = Yes

```

3.5.2 Cross Validation (10 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND

```

```

household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.13 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      667          79.7847 %
Incorrectly Classified Instances    169          20.2153 %
Kappa statistic                    0.5305
Mean absolute error                 0.1888
Root mean squared error             0.3143
Relative absolute error             63.0119 %
Root relative squared error         81.2736 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                ?         0.000   ?          ?         ?          ?        ?         ?         ?
                0.620    0.111   0.743     0.620    0.676     0.535    0.745     0.557     No
                0.889    0.380   0.820     0.889    0.853     0.535    0.840     0.817     Yes
Weighted Avg.   0.798    0.289   0.794     0.798    0.793     0.535    0.808     0.728

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 176 108 |  b = No
  0   61 491 |  c = Yes

```

3.5.3 Cross Validation (15 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    15-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

```

```

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      664          79.4258 %
Incorrectly Classified Instances    172          20.5742 %
Kappa statistic                    0.5301
Mean absolute error                 0.19
Root mean squared error             0.3161
Relative absolute error             63.4025 %
Root relative squared error         81.7394 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000    ?          ?          ?          ?         ?         ?         ?
      0.648    0.130    0.719    0.648    0.681    0.532    0.739    0.543    No
      0.870    0.352    0.828    0.870    0.848    0.532    0.842    0.798    Yes
Weighted Avg.  0.794    0.277    0.791    0.794    0.791    0.532    0.807    0.712

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 184 100 |   b = No
  0   72 480 |   c = Yes

```

3.5.4 Cross Validation (20 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:     StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    20-fold cross-validation

=== Classifier model (full training set) ===

```

```

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      656           78.4689 %
Incorrectly Classified Instances    180           21.5311 %
Kappa statistic                    0.5083
Mean absolute error                 0.1922
Root mean squared error             0.3191
Relative absolute error             64.1596 %
Root relative squared error         82.511 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000   ?          ?         ?          ?        ?         ?         ?
          0.634    0.138   0.703     0.634    0.667     0.510    0.743    0.554    No
          0.862    0.366   0.821     0.862    0.841     0.510    0.831    0.810    Yes
Weighted Avg.    0.785    0.289   0.781     0.785    0.782     0.510    0.801    0.723

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 180 104 |   b = No
  0  76 476 |   c = Yes

```

3.5.5 Cross Validation (25 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   25-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      655           78.3493 %
Incorrectly Classified Instances    181           21.6507 %
Kappa statistic                    0.5025
Mean absolute error                 0.1951
Root mean squared error             0.3227
Relative absolute error             65.1043 %
Root relative squared error         83.4351 %

```

```

Total Number of Instances          836
Ignored Class Unknown Instances    350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000   ?           ?         ?           ?         ?         ?         ?
          0.620    0.132   0.707    0.620    0.660    0.505    0.725    0.520    No
          0.868    0.380   0.816    0.868    0.841    0.505    0.830    0.791    Yes
Weighted Avg.   0.783    0.296   0.779    0.783    0.780    0.505    0.795    0.699

=== Confusion Matrix ===

  a   b   c  <-- classified as
  0   0   0 |  a = ?
  0 176 108 |  b = No
  0   73 479 |  c = Yes

```

3.5.6 Cross Validation (30 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:     StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    30-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

```

```

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      658           78.7081 %
Incorrectly Classified Instances    178           21.2919 %
Kappa statistic                    0.5103
Mean absolute error                 0.1895
Root mean squared error             0.3133
Relative absolute error             63.2495 %
Root relative squared error         81.0253 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        0.000   ?          ?        ?          ?        ?        ?        ?
0.623      0.129      0.714      0.623      0.665      0.513      0.746      0.555      No
0.871      0.377      0.818      0.871      0.844      0.513      0.843      0.808      Yes
Weighted Avg.  0.787      0.292      0.783      0.787      0.783      0.513      0.810      0.722

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 177 107 |  b = No
  0   71 481 |  c = Yes

```

3.5.7 Percentage Split 50%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 50.0% train, remainder test

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

```



```

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.03 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      328          78.2816 %
Incorrectly Classified Instances    91           21.7184 %
Kappa statistic                    0.5298
Mean absolute error                 0.1998
Root mean squared error             0.3324
Relative absolute error             66.5314 %
Root relative squared error         85.2277 %
Total Number of Instances          419
Ignored Class Unknown Instances      174

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        ?        ?          ?        ?          ?        ?        ?        ?
      0.721    0.184    0.679      0.721    0.700      0.530    0.726    0.477    No
      0.816    0.279    0.844      0.816    0.830      0.530    0.819    0.801    Yes
Weighted Avg.    0.783    0.246    0.786      0.783    0.784      0.530    0.786    0.687

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0   |   a = ?
  0 106  41   |   b = No
  0  50 222   |   c = Yes

```

3.5.8 Percentage Split 60%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 60.0% train, remainder test

=== Classifier model (full training set) ===

```

```

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      264          76.9679 %
Incorrectly Classified Instances    79           23.0321 %
Kappa statistic                    0.4817
Mean absolute error                 0.1927
Root mean squared error             0.3373
Relative absolute error             63.6387 %
Root relative squared error         85.0812 %
Total Number of Instances          343
Ignored Class Unknown Instances     131

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000    ?          ?        ?          ?        ?        ?        ?
          0.575    0.116    0.745     0.575    0.649     0.491    0.712    0.544    No
          0.884    0.425    0.780     0.884    0.829     0.491    0.840    0.813    Yes
Weighted Avg.    0.770    0.311    0.767     0.770    0.762     0.491    0.793    0.714

=== Confusion Matrix ===

  a   b   c  <-- classified as
  0   0   0 |  a = ?
  0  73  54 |  b = No
  0  25 191 |  c = Yes

```

3.5.9 Percentage Split 66%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 66.0% train, remainder test

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      231           80.2083 %
Incorrectly Classified Instances    57           19.7917 %
Kappa statistic                    0.55
Mean absolute error                 0.1837

```

```

Root mean squared error          0.3143
Relative absolute error          61.0212 %
Root relative squared error      80.6328 %
Total Number of Instances       288
Ignored Class Unknown Instances  115

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000   ?           ?         ?           ?         ?         ?         ?
          0.644    0.112   0.756     0.644    0.695     0.554    0.731    0.539    No
          0.888    0.356   0.822     0.888    0.853     0.554    0.856    0.828    Yes
Weighted Avg.    0.802    0.271   0.799     0.802    0.798     0.554    0.812    0.727

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0   65   36 |  b = No
  0   21  166 |  c = Yes

```

3.5.10 Percentage Split 70%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 70.0% train, remainder test

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND

```

```

household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      206          79.845 %
Incorrectly Classified Instances    52           20.155 %
Kappa statistic                    0.563
Mean absolute error                 0.1913
Root mean squared error             0.3163
Relative absolute error             63.461 %
Root relative squared error         81.0103 %
Total Number of Instances          258
Ignored Class Unknown Instances     98

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        ?        ?          ?        ?          ?        ?        ?        ?
              0.736    0.168    0.705      0.736    0.720      0.563    0.743    0.566     No
              0.832    0.264    0.853      0.832    0.842      0.563    0.843    0.822     Yes
Weighted Avg.   0.798    0.230    0.801      0.798    0.799      0.563    0.808    0.732

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0   67   24 |  b = No
  0   28  139 |  c = Yes

```

3.5.11 Percentage Split 80%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income
              education
              do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
              do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 80.0% train, remainder test

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

```

```

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      128          76.6467 %
Incorrectly Classified Instances    39           23.3533 %
Kappa statistic                    0.4404
Mean absolute error                 0.1918
Root mean squared error             0.3372
Relative absolute error             63.9343 %
Root relative squared error         87.1093 %
Total Number of Instances          167
Ignored Class Unknown Instances     70

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      ?        ?        ?          ?        ?          ?        ?        ?        ?
      0.509    0.100    0.725      0.509    0.598      0.454    0.667    0.473    No
      0.900    0.491    0.780      0.900    0.835      0.454    0.823    0.765    Yes
Weighted Avg.  0.766    0.358    0.761      0.766    0.754      0.454    0.770    0.666

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0  |  a = ?
  0 29 28  |  b = No
  0 11 99  |  c = Yes

```

3.5.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:    1186
Attributes:   10
              have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
              which-character-shot-first
              are-you-familiar-with-the-expanded-universe
              do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
              gender
              age
              household-income

```

```

education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: split 90.0% train, remainder test

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      68          80      %
Incorrectly Classified Instances   17          20      %
Kappa statistic                    0.5239
Mean absolute error                 0.1769
Root mean squared error             0.311
Relative absolute error             60.2696 %
Root relative squared error        82.3866 %
Total Number of Instances          85
Ignored Class Unknown Instances      34

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000   ?          ?          ?          ?        ?        ?        ?
      0.654    0.136    0.680     0.654     0.667     0.524    0.744    0.504    No
      0.864    0.346    0.850     0.864     0.857     0.524    0.858    0.863    Yes
Weighted Avg.    0.800    0.282    0.798     0.800     0.799     0.524    0.823    0.753

=== Confusion Matrix ===

```

```

a  b  c  <-- classified as
0  0  0  |  a = ?
0 17  9  |  b = No
0  8 51  |  c = Yes

```

3.5.13 Leave one out Fold

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -C 0.25 -M 2
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   1186-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Han: Yes (198.9/13.24)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
which-character-shot-first = Greedo: Yes (112.07/8.13)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes AND
are-you-familiar-with-the-expanded-universe = No AND
age = 60: Yes (38.79/5.47)

are-you-familiar-with-the-expanded-universe = Yes AND
age = 18-29: Yes (31.3/4.04)

do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise = Yes: Yes (68.23/18.66)

which-character-shot-first = none AND
education = AssociateDegree: No (55.82/5.75)

which-character-shot-first = none: No (142.08/35.91)

do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe = Yes: Yes (61.08/22.07)

age = 18-29: Yes (32.8/10.71)

age = 60 AND
household-income = 50.000-99.999: No (6.97/0.56)

household-income = 100.000-149.999: Yes (13.53/3.83)

are-you-familiar-with-the-expanded-universe = Yes AND
household-income = 50.000-99.999: Yes (7.06/0.94)

education = GraduateDegree AND
are-you-familiar-with-the-expanded-universe = No: Yes (12.12/4.05)

education = HighSchool: No (7.56/1.95)

education = AssociateDegree AND
household-income = 50.000-99.999 AND
gender = Female: No (5.73/2.13)

education = BachelorDegree: No (27.82/12.71)

education = AssociateDegree: Yes (10.49/3.32)

: No (3.64/1.05)

Number of Rules : 18

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

```



```

=== Summary ===

Correctly Classified Instances      661          79.067 %
Incorrectly Classified Instances    175          20.933 %
Kappa statistic                    0.5232
Mean absolute error                 0.1937
Root mean squared error             0.3181
Relative absolute error             64.5843 %
Root relative squared error         82.1625 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?         0.000   ?           ?         ?           ?         ?         ?         ?
      0.648    0.136   0.710      0.648    0.678      0.524    0.732    0.580    No
      0.864    0.352   0.827      0.864    0.845      0.524    0.824    0.816    Yes
Weighted Avg.   0.791    0.279   0.787      0.791    0.788      0.524    0.793    0.736

=== Confusion Matrix ===

 a   b   c   <-- classified as
0   0   0 |   a = ?
0 184 100 |   b = No
0   75 477 |   c = Yes

```

3.5.14 Conclusion

PART			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	79.18%	
	10 folds	79.78%	
	15 folds	79.42%	
	20 folds	78.46%	
	25 folds	78.34%	
	30 folds	78.70%	
Percentage Split	50.00%	78.28%	
	60.00%	76.96%	
	66.00%	80.20%	BEST
	70.00%	79.84%	
	80.00%	76.66%	ignored
	90.00%	80.00%	ignored
Leave One Out Fold		79.06%	

PART was one of the surprising algorithms as it was able to win 2 tests and position it self in the 4th place defeating more established algorithms like Random Forest. It had generated **18 rules** which can be viewed in weka's report above.

Best result came from **Percentage Split 66%** with **80.20%** Correctly Classified Instances. the weird thing is that **Percentage Split 66%** performed better than the higher percentages even with over-fitting which i don't have an explanation for.

3.6 Random Forest "default settings"

3.6.1 Cross Validation (5 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     StarWars
Instances:    1186

```

```

Attributes: 10
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
which-character-shot-first
are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: 5-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 2.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      644           77.0335 %
Incorrectly Classified Instances    192           22.9665 %
Kappa statistic                    0.4718
Mean absolute error                 0.1937
Root mean squared error             0.3334
Relative absolute error             64.64 %
Root relative squared error        86.2242 %
Total Number of Instances         836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000    ?          ?          ?          ?        ?        ?        ?
          0.599    0.141    0.685     0.599    0.639     0.474    0.731    0.543    No
          0.859    0.401    0.806     0.859    0.832     0.474    0.820    0.824    Yes
Weighted Avg.    0.770    0.313    0.765     0.770    0.766     0.474    0.790    0.729

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0 |   a = ?
  0 170 114 |   b = No
  0   78 474 |   c = Yes

```

3.6.2 Cross Validation (10 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:    1186
Attributes:  10
have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
which-character-shot-first
are-you-familiar-with-the-expanded-universe
do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
gender
age
household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.63 seconds

=== Stratified cross-validation ===
=== Summary ===

```

```

Correctly Classified Instances      638          76.3158 %
Incorrectly Classified Instances    198          23.6842 %
Kappa statistic                    0.4591
Mean absolute error                0.195
Root mean squared error            0.3354
Relative absolute error             65.0754 %
Root relative squared error        86.7304 %
Total Number of Instances         836
Ignored Class Unknown Instances    350

```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.602	0.154	0.668	0.602	0.633	0.460	0.728	0.535	No
	0.846	0.398	0.805	0.846	0.825	0.460	0.814	0.810	Yes
Weighted Avg.	0.763	0.315	0.759	0.763	0.760	0.460	0.785	0.717	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 171 113   |   b = No
0   85 467   |   c = Yes

```

3.6.3 Cross Validation (15 Folds)

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    15-fold cross-validation

```

```
=== Classifier model (full training set) ===
```

```
RandomForest
```

```
Bagging with 100 iterations and base learner
```

```
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
```

```
Time taken to build model: 0.58 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

```

Correctly Classified Instances      637          76.1962 %
Incorrectly Classified Instances    199          23.8038 %
Kappa statistic                    0.4559
Mean absolute error                0.1939
Root mean squared error            0.3363
Relative absolute error             64.7254 %
Root relative squared error        86.9702 %
Total Number of Instances         836
Ignored Class Unknown Instances    350

```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.599	0.154	0.667	0.599	0.631	0.457	0.725	0.546	No
	0.846	0.401	0.804	0.846	0.824	0.457	0.816	0.814	Yes
Weighted Avg.	0.762	0.317	0.757	0.762	0.759	0.457	0.785	0.723	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 170 114   |   b = No
0   85 467   |   c = Yes

```

3.6.4 Cross Validation (20 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   20-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.51 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      644           77.0335 %
Incorrectly Classified Instances    192           22.9665 %
Kappa statistic                    0.4746
Mean absolute error                 0.1929
Root mean squared error             0.3346
Relative absolute error             64.3927 %
Root relative squared error         86.5333 %
Total Number of Instances          836
Ignored Class Unknown Instances      350

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?        0.000    ?          ?        ?          ?        ?        ?        ?
              0.609    0.147    0.681      0.609    0.643      0.476    0.727    0.546    No
              0.853    0.391    0.809      0.853    0.831      0.476    0.816    0.807    Yes
Weighted Avg.   0.770    0.308    0.766      0.770    0.767      0.476    0.786    0.719

=== Confusion Matrix ===

  a   b   c   <-- classified as
  0   0   0   |   a = ?
  0 173 111 |   b = No
  0  81 471 |   c = Yes

```

3.6.5 Cross Validation (25 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   25-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

```

```

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.59 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      640           76.555 %
Incorrectly Classified Instances    196           23.445 %
Kappa statistic                    0.4608
Mean absolute error                 0.1922
Root mean squared error             0.3344
Relative absolute error             64.1305 %
Root relative squared error         86.4616 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000  ?          ?        ?          ?        ?        ?        ?
          0.592    0.145    0.677     0.592    0.632     0.463    0.727    0.543    No
          0.855    0.408    0.803     0.855    0.828     0.463    0.820    0.816    Yes
Weighted Avg.   0.766    0.319    0.760     0.766    0.761     0.463    0.789    0.723

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0 168 116 |  b = No
  0  80 472 |  c = Yes

```

3.6.6 Cross Validation (30 Folds)

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    30-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.68 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      639           76.4354 %
Incorrectly Classified Instances    197           23.5646 %
Kappa statistic                    0.4623
Mean absolute error                 0.1928
Root mean squared error             0.3354
Relative absolute error             64.3417 %
Root relative squared error         86.7401 %
Total Number of Instances          836
Ignored Class Unknown Instances     350

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000  ?          ?        ?          ?        ?        ?        ?
          0.606    0.154    0.669     0.606    0.636     0.464    0.728    0.551    No
          0.846    0.394    0.807     0.846    0.826     0.464    0.819    0.814    Yes
Weighted Avg.   0.764    0.313    0.760     0.764    0.761     0.464    0.788    0.725

=== Confusion Matrix ===

```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 172 112 |   b = No
0   85 467 |   c = Yes

```

3.6.7 Percentage Split 50%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 50.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 1.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.35 seconds

=== Summary ===

Correctly Classified Instances      318              75.895 %
Incorrectly Classified Instances    101              24.105 %
Kappa statistic                    0.4439
Mean absolute error                 0.2035
Root mean squared error             0.3451
Relative absolute error             67.759 %
Root relative squared error         88.4744 %
Total Number of Instances          419
Ignored Class Unknown Instances      174

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              ?         0.000   ?          ?         ?          ?        ?         ?         ?
              0.551    0.129   0.698     0.551    0.616     0.450    0.703    0.517    No
              0.871    0.449   0.782     0.871    0.824     0.450    0.801    0.786    Yes
Weighted Avg.   0.759    0.337   0.753     0.759    0.751     0.450    0.767    0.692

=== Confusion Matrix ===

a   b   c   <-- classified as
0   0   0   |   a = ?
0  81  66 |   b = No
0  35 237 |   c = Yes

```

3.6.8 Percentage Split 60%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age

```

```

household-income
education
do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode: split 60.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.49 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.34 seconds

=== Summary ===

Correctly Classified Instances      262          76.3848 %
Incorrectly Classified Instances    81           23.6152 %
Kappa statistic                    0.4613
Mean absolute error                 0.2033
Root mean squared error             0.3454
Relative absolute error             67.138 %
Root relative squared error        87.1308 %
Total Number of Instances          343
Ignored Class Unknown Instances    131

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      ?        0.000   ?          ?        ?          ?        ?        ?        ?
0.535  0.102  0.756    0.535    0.627    0.476    0.705    0.528    No
0.898  0.465  0.767    0.898    0.827    0.476    0.814    0.793    Yes
Weighted Avg.  0.764  0.330  0.763    0.764    0.753    0.476    0.774    0.695

=== Confusion Matrix ===

 a   b   c   <-- classified as
0   0   0 |   a = ?
0  68  59 |   b = No
0  22 194 |   c = Yes

```

3.6.9 Percentage Split 66%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 66.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.48 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.3 seconds

=== Summary ===

Correctly Classified Instances      218          75.6944 %

```

```

Incorrectly Classified Instances      70          24.3056 %
Kappa statistic                     0.4408
Mean absolute error                  0.1936
Root mean squared error              0.3367
Relative absolute error              64.3092 %
Root relative squared error          86.3717 %
Total Number of Instances           288
Ignored Class Unknown Instances      115

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000   ?          ?        ?          ?        ?        ?        ?
          0.554    0.134    0.691     0.554    0.615      0.447    0.724    0.513     No
          0.866    0.446    0.783     0.866    0.822      0.447    0.827    0.824     Yes
Weighted Avg.    0.757    0.336    0.751     0.757    0.750      0.447    0.791    0.715

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0   56   45 |  b = No
  0   25  162 |  c = Yes

```

3.6.10 Percentage Split 70%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    split 70.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.47 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.24 seconds

=== Summary ===

Correctly Classified Instances      194          75.1938 %
Incorrectly Classified Instances     64          24.8062 %
Kappa statistic                     0.4281
Mean absolute error                  0.1982
Root mean squared error              0.338
Relative absolute error              65.7419 %
Root relative squared error          86.5679 %
Total Number of Instances           258
Ignored Class Unknown Instances      98

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?        0.000   ?          ?        ?          ?        ?        ?        ?
          0.538    0.132    0.690     0.538    0.605      0.435    0.723    0.508     No
          0.868    0.462    0.775     0.868    0.819      0.435    0.823    0.822     Yes
Weighted Avg.    0.752    0.345    0.745     0.752    0.744      0.435    0.788    0.711

=== Confusion Matrix ===

  a    b    c  <-- classified as
  0    0    0 |  a = ?
  0   49   42 |  b = No
  0   22  145 |  c = Yes

```


3.6.11 Percentage Split 80%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 80.0% train, remainder test

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.81 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.18 seconds

=== Summary ===

Correctly Classified Instances      127           76.0479 %
Incorrectly Classified Instances    40           23.9521 %
Kappa statistic                    0.4438
Mean absolute error                 0.1921
Root mean squared error             0.3425
Relative absolute error             64.0376 %
Root relative squared error         88.4552 %
Total Number of Instances          167
Ignored Class Unknown Instances      70

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          ?         0.000    ?           ?         ?           ?         ?         ?         ?
          0.561    0.136    0.681     0.561    0.615     0.448    0.697    0.456    No
          0.864    0.439    0.792     0.864    0.826     0.448    0.830    0.817    Yes
Weighted Avg.   0.760    0.335    0.754     0.760    0.754     0.448    0.785    0.694

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 | a = ?
  0 32 25 | b = No
  0 15 95 | c = Yes

```

3.6.12 Percentage Split 90%

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:   1186
Attributes:  10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:   split 90.0% train, remainder test

```

```

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.57 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.23 seconds

=== Summary ===

Correctly Classified Instances          67           78.8235 %
Incorrectly Classified Instances        18           21.1765 %
Kappa statistic                        0.4544
Mean absolute error                     0.1837
Root mean squared error                 0.3312
Relative absolute error                 62.6139 %
Root relative squared error             87.7419 %
Total Number of Instances              85
Ignored Class Unknown Instances          34

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               ?         0.000    ?         ?         ?         ?         ?         ?         ?
    0.500    0.085    0.722    0.500    0.591    0.468    0.692    0.477    No
    0.915    0.500    0.806    0.915    0.857    0.468    0.843    0.844    Yes
Weighted Avg.   0.788    0.373    0.780    0.788    0.776    0.468    0.797    0.731

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  0 | a = ?
  0 13 13 | b = No
  0  5 54 | c = Yes

```

3.6.13 Leave one out Fold

```

=== Run information ===

Scheme:      weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    StarWars
Instances:    1186
Attributes:   10
             have-you-seen-any-of-the-6-films-in-the-star-wars-franchise
             which-character-shot-first
             are-you-familiar-with-the-expanded-universe
             do-you-consider-yourself-to-be-a-fan-of-the-expanded-universe
             gender
             age
             household-income
             education
             do-you-consider-yourself-to-be-a-fan-of-the-star-trek-franchise
             do-you-consider-yourself-to-be-a-fan-of-the-star-wars-film-franchise
Test mode:    1186-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.65 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          640           76.555 %
Incorrectly Classified Instances        196           23.445 %
Kappa statistic                        0.4618
Mean absolute error                     0.194
Root mean squared error                 0.3364
Relative absolute error                 64.6795 %
Root relative squared error             86.8935 %
Total Number of Instances              836
Ignored Class Unknown Instances          350

```

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	?	0.000	?	?	?	?	?	?	?
	0.595	0.147	0.676	0.595	0.633	0.464	0.724	0.537	No
	0.853	0.405	0.804	0.853	0.828	0.464	0.817	0.816	Yes
Weighted Avg.	0.766	0.317	0.760	0.766	0.762	0.464	0.786	0.721	

```
=== Confusion Matrix ===
```

```

a   b   c   <-- classified as
0   0   0   |   a = ?
0 169 115   |   b = No
0   81 471   |   c = Yes

```

3.6.14 Conclusion

Random Forest			
Evaluation Process		Correctly Classified Instances	notes
Cross Validation	5 folds	77.03%	BEST
	10 folds	76.31%	
	15 folds	76.19%	
	20 folds	77.03%	BEST
	25 folds	76.85%	
	30 folds	76.43%	
Percentage Split	50.00%	75.89%	
	60.00%	76.38%	
	66.00%	75.69%	
	70.00%	75.19%	
	80.00%	76.04%	ignored
	90.00%	78.82%	ignored
Leave One Out Fold		76.55%	

If PART was the positive surprise Random Forest is the sad surprise as it wasn't able to win a single test along with One Rule despite being a very strong algorithm. it was only able to beat One Rule during our tests.

The surprise was also in terms of evaluation method as it was the only algorithm where Cross Validation performed better than Percentage Split. I have no clue why and if the bagging system has any effect here or if it was pure luck. never the less best results came from **Cross Validation 5 Folds** and **Cross Validation 20 Folds** with both scoring **77.03%** Correctly Classified Instances. excluding 90% results due to over-fitting problem.

3.7 General Conclusion

Correctly Classified Instances by Algorithm									
Evaluation Process		One Rule	Naive Bayes	KNN	C4.5	PART	Random Forest	AVG Methods	Rank Methods
Cross Validation	5 folds	73.56%	78.34%	78.94%	79.30%	79.18%	77.03%	77.73%	4
	10 folds	73.56%	79.06%	78.94%	78.11%	79.78%	76.31%	77.63%	6
	15 folds	72.36%	78.70%	78.94%	78.70%	79.42%	76.19%	77.39%	8
	20 folds	73.56%	78.22%	79.42%	79.30%	78.46%	77.03%	77.67%	5
	25 folds	73.56%	78.58%	78.58%	78.11%	78.34%	76.85%	77.34%	9
	30 folds	73.56%	78.46%	79.18%	79.30%	78.70%	76.43%	77.61%	7
Percentage Split	50.00%	72.31%	78.99%	77.08%	79.71%	78.28%	75.89%	77.04%	10
	60.00%	72.30%	78.13%	78.71%	79.00%	76.96%	76.38%	76.91%	11
	66.00%	73.61%	80.90%	79.51%	80.20%	80.20%	75.69%	78.35%	1
	70.00%	73.64%	79.84%	78.31%	80.23%	79.84%	75.19%	77.84%	3
	80.00%	76.04%	82.96%	79.04%	79.64%	76.66%	76.04%	78.40%	ignored
	90.00%	76.47%	83.52%	84.70%	84.70%	80.00%	78.82%	81.37%	ignored
Leave One Out Fold		73.56%	78.58%	79.18%	80.62%	79.06%	76.55%	77.93%	2
AVG Algorithms		73.70%	79.56%	79.27%	79.76%	78.84%	76.49%		
Rank Algorithms		6	2	3	1	4	5		

The 78 experiment we did resulted an interesting conclusion showing in terms of algorithms a domination performance from C4.5 which was able to win in 7 different evaluations out of the 13 tests which ranks it first with an average score of 79.76%. Naive Bayes was second winning in 3 tests and scoring an average of 79.56%. Third place was for KNN that also won in 3 tests and scored an average of 79.27%. forth place was for the dark sheep of the bunch PART that won in 2 tests and scored an average of 78.84%. fifth and sixth places were for Random Forest and One Rule that scored a 76.49% and 73.70% respective average scores, both were not able to win in any test.

In terms of evaluation methods Percentage Split 66% was first 3 times (Naive Bayes, KNN, PART) with an average of 78.35% making it the best overall method, followed by Leave One Out Fold which was first once in C4.5 with an average score of 77.93%. third best evaluation method was for Percentage Split 70% where it was first once in One Rule with average overall score of 77.84%. forth and fifth and sixth and seventh were all for Cross Validation where it won once in Random Forest by Folds and 20 folds.

Chapter 4

Application part

4.1 Cross Validation implementation

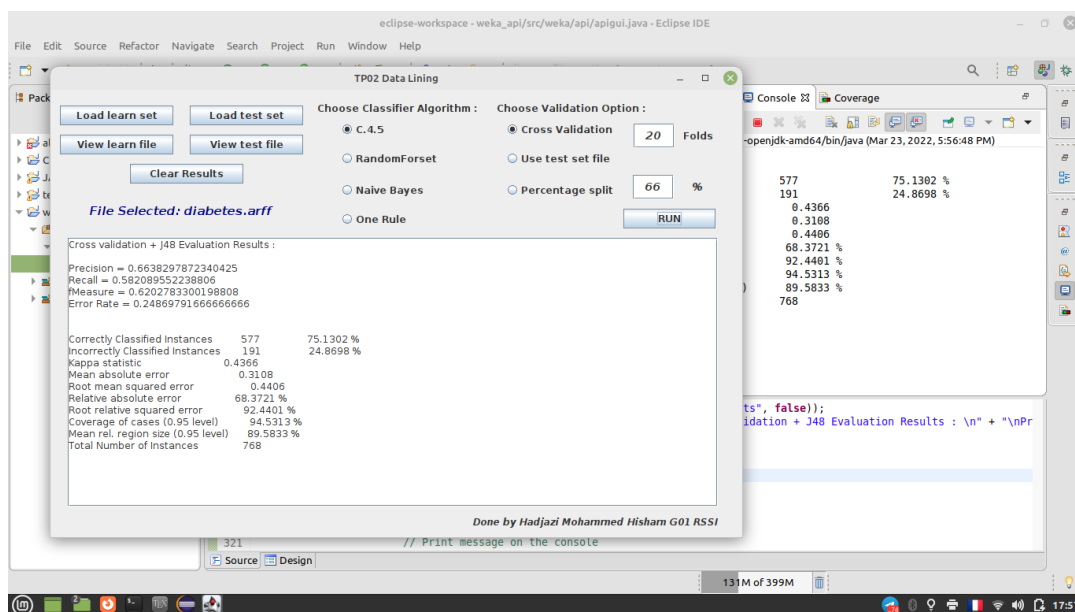


FIGURE 4.1: Cross Validation Execution

```

1 // getting CrossValidation_Value input
2
3     CrossValidation_Value = Integer.parseInt(
4         input1.getText());
5
6     // Create J48 classifier by
7     // creating object of J48 class
8     J48 j48Classifier = new J48();
9
10    // Dataset path
11    String DataStes = file.getPath();
12
13    // Creating bufferedreader to read the
    dataset
  
```

```
14         BufferedReader bufferedReader = new
            BufferedReader(new FileReader(DataStes))
            ;
15
16         // Create dataset instances
17         Instances datasetInstances = new Instances(
            bufferedReader);
18
19         // Set Target Class
20         datasetInstances.setClassIndex(
            datasetInstances.numAttributes() - 1);
21
22         // Evaluating by creating object of
            Evaluation
23         // class
24         Evaluation evaluation = new Evaluation(
            datasetInstances);
25
26         // Cross Validate Model with user selected
            folds
27         evaluation.crossValidateModel(j48Classifier
            , datasetInstances,
            CrossValidation_Value,new Random(1));
28
29         System.out.println(evaluation.
            toSummaryString("\nResults", false));
30         textArea.setText(evaluation.toSummaryString
            ("Cross validation + J48 Evaluation
            Results : \n" + "\nPrecision = "+
            evaluation.precision(1)+"\nRecall = "+
            evaluation.recall(1)+"\nfMeasure = "+
            evaluation.fMeasure(1)+"\nError Rate = "
            +evaluation.errorRate()+"\n\n", false));
```

4.2 Percentage Split implementation

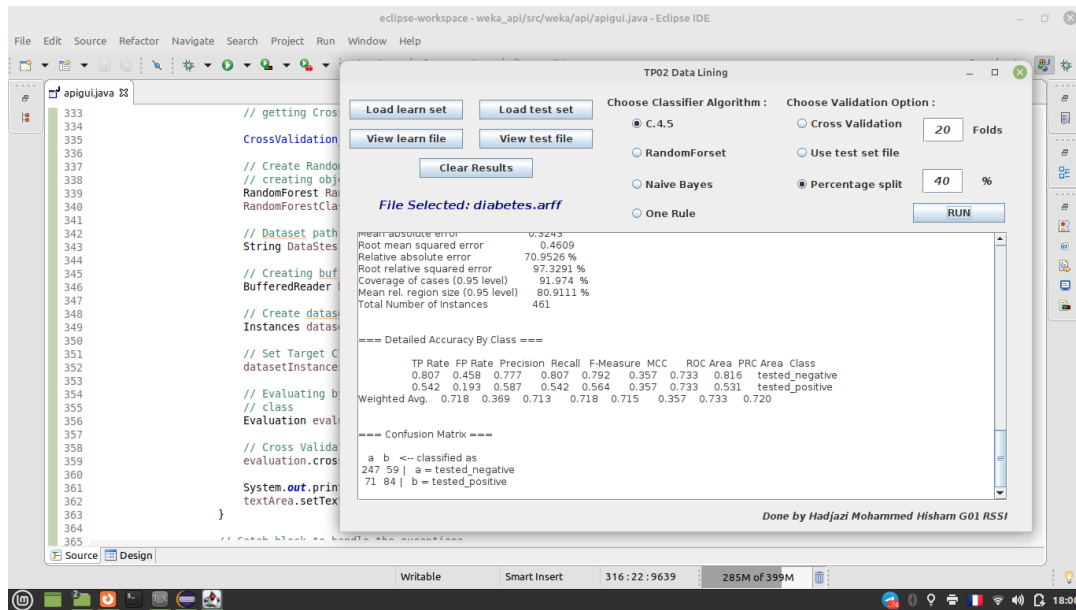


FIGURE 4.2: Percentage Split Execution

```

32 // Dataset path
33
34 String DataStes = file.getPath();
35
36
37
38 String[] options = new String[4];
39 options[0] = "-t";
40 options[1] = DataStes;
41 options[2] = "-split-percentage";
42 options[3] = String.valueOf(
43     Split_Percentage);
44
45
46
47
48 textArea.setText("Percentage Split + Naive
49                 Bayes Evaluation Results : \n"+
                    Evaluation.evaluateModel(new NaiveBayes
                    (), options));

```

Chapter 5

Conclusion

The conclusion of this TP showed how hard it is to make a dataset from information found online, and this is with the help of internet i can't imagine how hard it will be if the server was performed manually. despite that i was able to find the results of the survey online and in a very useful CSV format, just converting it to arff format was a another hard task were I had to do all the modifications and fixes in CSV since there are strong programs like the one we used ModernCSV that can perform hard tasks of searching and replacing using may techniques like finding exact words or using Regular Expressions for searching and replacing. most important changes were deleting columns that we didn't need and replace missing values with "?", fixing number intervals like age and income by adding the quotations them and other fixes to titles and names to be compatible with the ARFF format. after that thanks to another online tool that did the conversion automaticity <https://ikuz.eu/csv2arff/>.

Another lesson that was learned from the problems of TP01 was the copying of experiment results in tables, this step usually consumes a lot of time and was prone to errors, however this time we have 78 experiment so thanks to a built-in feature of Weka of save result buffer into text files. this saved me time that was better spent in creating comparison tables that were very informative and helped in deducing conclusions and rank best algorithms and best evaluation methods.

Again the most fun and interesting part was the coding part were this time we upgraded our program to be able include cross validation and percentage split and allowing the user to change manually the number of folds and the size of split. this time it was easy as the most of the hard work was already done in TP01, the challenge was in percentage split as I wasn't able to generate the same results as the ones in Weka until i used the options method, which used the same way that weka randomizes the data to be able to generate repetitive experiment results.

In terms of evaluation methods, despite that percentage split won, I wouldn't use it and rather use Leave one out fold or cross validation as I noticed had more stability which made it more credible to me. I would only use percentage split when the dataset is very large and execution time becomes more critical to finish the experiments.

The question marks that can be seen in the results data are manly because during the calculations we can't divide by zero. however it was not a problem for our evaluation as we used Correctly Classified Instances as a measuring criteria.

After running the results we find that the people who like Star trek also like Star wars. =)

Appendix A

Source Code

A.1 apigui.java

```
50
51 package weka.api;
52
53 import java.awt.EventQueue;
54
55 import javax.swing.JFrame;
56 import javax.swing.JLabel;
57 import javax.swing.JScrollPane;
58 import javax.swing.JTextArea;
59 import javax.swing.filechooser.FileNameExtensionFilter;
60 import javax.swing.JButton;
61 import java.awt.event.ActionListener;
62 import java.io.BufferedReader;
63 import java.io.File;
64 import java.io.FileInputStream;
65 import java.io.FileReader;
66 import java.io.InputStreamReader;
67 import java.util.Random;
68 import java.awt.event.ActionEvent;
69 import javax.swing.JRadioButton;
70
71 import weka.classifiers.bayes.NaiveBayes;
72 import weka.classifiers.evaluation.Evaluation;
73 import weka.classifiers.rules.OneR;
74 import weka.classifiers.trees.J48;
75 import weka.classifiers.trees.RandomForest;
76 import weka.core.Instances;
77
78 import javax.swing.JFileChooser;
79 import javax.swing.ButtonGroup;
80 import javax.swing.SwingConstants;
81 import java.awt.Font;
82 import java.awt.SystemColor;
83 import javax.swing.JTextField;
84
85 public class apigui {
86
87
88
89     private JFrame frmTpDataLining;
90     private File file;
91     private File file2;
```

```

92     private JRadioButton A1;
93     private JRadioButton A2;
94     private JRadioButton A3;
95     private JRadioButton A4;
96     private JRadioButton B1;
97     private JRadioButton B2;
98     private JRadioButton B3;
99     private JTextField input1;
100    private JTextField input2;
101    private int CrossValidation_Value;
102    private float Split_Percentage;
103
104    /**
105     * Launch the application.
106     */
107    public static void main(String[] args) {
108        EventQueue.invokeLater(new Runnable() {
109            public void run() {
110                try {
111                    apigui window = new apigui();
112                    window.frmTpDataLining.setVisible(true);
113                } catch (Exception e) {
114                    e.printStackTrace();
115                }
116            }
117        });
118    }
119
120    /**
121     * Create the application.
122     */
123    public apigui() {
124        initialize();
125    }
126
127    /**
128     * Initialize the contents of the frame.
129     */
130    private void initialize() {
131
132
133
134        frmTpDataLining = new JFrame();
135        frmTpDataLining.setTitle("TP02 Data Lining");
136        frmTpDataLining.setBounds(100, 100, 881, 599);
137        frmTpDataLining.setDefaultCloseOperation(JFrame.
138            EXIT_ON_CLOSE);
139        frmTpDataLining.getContentPane().setLayout(null);
140
141        JLabel label = new JLabel("");
142        label.setForeground(SystemColor.activeCaption);
143        label.setFont(new Font("Dialog", Font.BOLD | Font.ITALIC,
144            15));
145        label.setHorizontalAlignment(SwingConstants.CENTER);
146        label.setBounds(12, 132, 309, 44);
147        frmTpDataLining.getContentPane().add(label);

```

```

147
148     JScrollPane scrollPane = new JScrollPane();
149     scrollPane.setBounds(22, 188, 826, 341);
150     frmTpDataLining.getContentPane().add(scrollPane);
151
152     JTextArea textArea = new JTextArea();
153     scrollPane.setViewportViewView(textArea);
154
155     // Button for opening the learning dataset
156
157     JButton btnNewButton = new JButton("Load learn set");
158     btnNewButton.addActionListener(new ActionListener() {
159         @Override
160         public void actionPerformed(ActionEvent e) {
161             JFileChooser fileChooser = new JFileChooser();
162             fileChooser.setFileFilter(new
163                 FileNameExtensionFilter("ARFF File", "arff"));
164             int option = fileChooser.showOpenDialog(
165                 frmTpDataLining);
166             if(option == JFileChooser.APPROVE_OPTION){
167                 file = fileChooser.getSelectedFile();
168                 label.setText("File Selected: " + file.
169                     getName());
170             }else{
171                 label.setText("Loading ARFF file canceled");
172             }
173         }
174     });
175
176     // Button for opening the learning testset
177
178     JButton btnLoadTestSet = new JButton("Load test set");
179     btnLoadTestSet.addActionListener(new ActionListener() {
180         @Override
181         public void actionPerformed(ActionEvent e) {
182             JFileChooser fileChooser2 = new JFileChooser();
183             fileChooser2.setFileFilter(new
184                 FileNameExtensionFilter("ARFF File", "arff"));
185             int option = fileChooser2.showOpenDialog(
186                 frmTpDataLining);
187             if(option == JFileChooser.APPROVE_OPTION){
188                 file2 = fileChooser2.getSelectedFile();
189                 label.setText("File Selected: " + file2.
190                     getName());
191             }else{
192                 label.setText("Loading TEST file canceled");
193             }
194         }
195     });
196     btnLoadTestSet.setBounds(177, 20, 144, 25);
197     frmTpDataLining.getContentPane().add(btnLoadTestSet);

```

```

198
199     btnNewButton.setBounds(12, 20, 144, 25);
200     frmTpDataLining.getContentPane().add(btnNewButton);
201
202     // Button for view the learning learn set
203
204     JButton btnViewArff = new JButton("View learn file");
205     btnViewArff.addActionListener(new ActionListener() {
206         public void actionPerformed(ActionEvent arg0) {
207             //textArea.setText("");
208             try {
209                 BufferedReader input = new BufferedReader(new
210                     InputStreamReader(
211                         new FileInputStream(file)));
212                 textArea.read(input, "READING FILE :-");
213             } catch (Exception e) {
214                 e.printStackTrace();
215             }
216         }
217     });
218     btnViewArff.setBounds(12, 57, 144, 25);
219     frmTpDataLining.getContentPane().add(btnViewArff);
220
221
222
223     input1 = new JTextField();
224     input1.setFont(new Font("Dialog", Font.BOLD | Font.ITALIC,
225         14));
226     input1.setHorizontalAlignment(SwingConstants.CENTER);
227     input1.setText("10");
228     input1.setBounds(741, 43, 52, 31);
229     frmTpDataLining.getContentPane().add(input1);
230     input1.setColumns(10);
231
232     input2 = new JTextField();
233     input2.setFont(new Font("Dialog", Font.BOLD | Font.ITALIC,
234         14));
235     input2.setText("66");
236     input2.setHorizontalAlignment(SwingConstants.CENTER);
237     input2.setColumns(10);
238     input2.setBounds(740, 108, 52, 31);
239     frmTpDataLining.getContentPane().add(input2);
240
241
242     JLabel lblChooseValidationOption = new JLabel("Choose
243         Classifier Algorithm :");
244     lblChooseValidationOption.setBounds(339, 16, 216, 15);
245     frmTpDataLining.getContentPane().add(
246         lblChooseValidationOption);
247
248     JLabel lblChooseValidationOption_1 = new JLabel("Choose
249         Validation Option :");
250     lblChooseValidationOption_1.setBounds(567, 16, 200, 15);

```

```

248         frmTpDataLining.getContentPane().add(
                lblChooseValidationOption_1);
249
250         B1 = new JRadioButton("Cross Validation");
251         B1.setBounds(577, 39, 150, 23);
252         frmTpDataLining.getContentPane().add(B1);
253
254         B2 = new JRadioButton("Use test set file");
255         B2.setBounds(577, 76, 150, 23);
256         frmTpDataLining.getContentPane().add(B2);
257
258         B3 = new JRadioButton("Percentage split");
259         B3.setBounds(577, 116, 150, 23);
260         frmTpDataLining.getContentPane().add(B3);
261
262         A1 = new JRadioButton("C.4.5");
263         A1.setBounds(367, 39, 200, 23);
264         frmTpDataLining.getContentPane().add(A1);
265
266         A2 = new JRadioButton("RandomForset");
267         A2.setBounds(367, 76, 200, 23);
268         frmTpDataLining.getContentPane().add(A2);
269
270         A4 = new JRadioButton("One Rule");
271         A4.setBounds(367, 153, 200, 23);
272         frmTpDataLining.getContentPane().add(A4);
273
274         A3 = new JRadioButton("Naive Bayes");
275         A3.setBounds(367, 116, 200, 23);
276         frmTpDataLining.getContentPane().add(A3);
277
278         ButtonGroup gpp = new ButtonGroup();
279         gpp.add(A1);
280         gpp.add(A2);
281         gpp.add(A3);
282         gpp.add(A4);
283
284         ButtonGroup gpp2 = new ButtonGroup();
285         gpp2.add(B1);
286         gpp2.add(B2);
287         gpp2.add(B3);
288
289
290         // Button for clearing screen
291
292         JButton btnClearResults = new JButton("Clear Results");
293         btnClearResults.addActionListener(new ActionListener() {
294             public void actionPerformed(ActionEvent arg0) {
295                 textArea.setText("");
296             }
297         });
298         btnClearResults.setBounds(101, 94, 144, 25);
299         frmTpDataLining.getContentPane().add(btnClearResults);
300
301         // Button for view the learning testset
302
303         JButton btnViewTestFile = new JButton("View test file");

```

```

304     btnViewTestFile.addActionListener(new ActionListener() {
305         public void actionPerformed(ActionEvent arg0) {
306             //textArea.setText("");
307             try {
308                 BufferedReader input = new BufferedReader(new
309                     InputStreamReader(
310                         new FileInputStream(file2)));
311                 textArea.read(input, "READING FILE :-");
312             } catch (Exception e) {
313                 e.printStackTrace();
314             }
315         });
316     btnViewTestFile.setBounds(177, 57, 144, 25);
317     frmTpDataLining.getContentPane().add(btnViewTestFile);
318
319     // most important button
320
321     JButton btnNewButton_2 = new JButton("RUN");
322     btnNewButton_2.addActionListener(new ActionListener() {
323         public void actionPerformed(ActionEvent arg0) {
324
325
326
327
328
329             // Cross validation + J48
330             if (A1.isSelected() && B1.isSelected()) {
331
332                 try {
333                     // getting CrossValidation_Value input
334
335                     CrossValidation_Value = Integer.parseInt(
336                         input1.getText());
337
338                     // Create J48 classifier by
339                     // creating object of J48 class
340                     J48 j48Classifier = new J48();
341
342                     // Dataset path
343                     String DataStes = file.getPath();
344
345                     // Creating bufferedreader to read the
346                     // dataset
347                     BufferedReader bufferedReader = new
348                         BufferedReader(new FileReader(DataStes))
349                         ;
350
351                     // Create dataset instances
352                     Instances datasetInstances = new Instances(
353                         bufferedReader);

```

```

354         // Evaluating by creating object of
           Evaluation
355         // class
356         Evaluation evaluation = new Evaluation(
           datasetInstances);
357
358         // Cross Validate Model with user selected
           folds
359         evaluation.crossValidateModel(j48Classifier
           , datasetInstances,
           CrossValidation_Value,new Random(1));
360
361         System.out.println(evaluation.
           toSummaryString("\nResults", false));
362         textArea.setText(evaluation.toSummaryString
           ("Cross validation + J48 Evaluation
           Results : \n" + "\nPrecision = " +
           evaluation.precision(1)+"\nRecall = " +
           evaluation.recall(1)+"\nfMeasure = " +
           evaluation.fMeasure(1)+"\nError Rate = "
           +evaluation.errorRate()+"\n\n", false));
363
364
365
366     }
367
368     // Catch block to handle the exceptions
369     catch (Exception e) {
370
371         // Print message on the console
372         System.out.println("Error Occurred!!!! \n"
           + e.getMessage());
373     }
374
375
376
377 }
378 // Cross validation + RandomForest
379 else if (A2.isSelected() && B1.isSelected()) {
380
381     try {
382
383         // getting CrossValidation_Value input
384
385         CrossValidation_Value = Integer.parseInt(
           input1.getText());
386
387         // Create RandomForest classifier by
388         // creating object of RandomForest class
389         RandomForest RandomForestClassifier = new
           RandomForest();
390         RandomForestClassifier.setNumTrees(100);
391
392         // Dataset path
393         String DataStes = file.getPath();
394

```



```

395         // Creating bufferedreader to read the
396         dataset
397         BufferedReader bufferedReader = new
398             BufferedReader(new FileReader(DataStes))
399             ;
400
401         // Create dataset instances
402         Instances datasetInstances = new Instances(
403             bufferedReader);
404
405         // Set Target Class
406         datasetInstances.setClassIndex(
407             datasetInstances.numAttributes() - 1);
408
409         // Evaluating by creating object of
410         Evaluation
411         // class
412         Evaluation evaluation = new Evaluation(
413             datasetInstances);
414
415         // Cross Validate Model with user selected
416         folds
417         evaluation.crossValidateModel(
418             RandomForestClassifier, datasetInstances
419             , CrossValidation_Value, new Random(1));
420
421         System.out.println(evaluation.
422             toSummaryString("\nResults", false));
423         textArea.setText(evaluation.toSummaryString
424             ("Cross validation + RandomForest
425             Evaluation Results : \n" + "\nPrecision =
426             "+evaluation.precision(1)+"\nRecall = "
427             +evaluation.recall(1)+"\nfMeasure = "+
428             evaluation.fMeasure(1)+"\nError Rate = "
429             +evaluation.errorRate()+"\n\n", false));
430     }
431
432     // Catch block to handle the exceptions
433     catch (Exception e) {
434
435         // Print message on the console
436         System.out.println("Error Occurred!!!! \n"
437             + e.getMessage());
438     }
439
440 }
441
442 // Cross validation + NaiveBayes
443 else if (A3.isSelected() && B1.isSelected()) {
444
445     try {
446
447         // getting CrossValidation_Value input
448
449         CrossValidation_Value = Integer.parseInt(
450             input1.getText());

```

```

433
434         // Create NaiveBayes classifier by
435         // creating object of NaiveBayes class
436         NaiveBayes NaiveBayesClassifier = new
            NaiveBayes();
437
438         // Dataset path
439         String DataStes = file.getPath();
440
441         // Creating bufferedreader to read the
442         // dataset
443         BufferedReader bufferedReader = new
444             BufferedReader(new FileReader(DataStes))
445             ;
446
447         // Create dataset instances
448         Instances datasetInstances = new Instances(
449             bufferedReader);
450
451         // Set Target Class
452         datasetInstances.setClassIndex(
453             datasetInstances.numAttributes() - 1);
454
455         // Evaluating by creating object of
456         // Evaluation
457         // class
458         Evaluation evaluation = new Evaluation(
459             datasetInstances);
460
461         // Cross Validate Model with user selected
462         // folds
463         evaluation.crossValidateModel(
464             NaiveBayesClassifier, datasetInstances,
465             CrossValidation_Value, new Random(1));
466
467         System.out.println(evaluation.
468             toSummaryString("\nResults", false));
469         textArea.setText(evaluation.toSummaryString(
470             "Cross validation + NaiveBayes
471             Evaluation Results : \n" + "\nPrecision =
472             "+evaluation.precision(1)+"\nRecall = "
473             +evaluation.recall(1)+"\nfMeasure = "
474             +evaluation.fMeasure(1)+"\nError Rate = "
475             +evaluation.errorRate()+"\n\n", false));
476     }
477
478     // Catch block to handle the exceptions
479     catch (Exception e) {
480
481         // Print message on the console
482         System.out.println("Error Occurred!!!! \n"
483             + e.getMessage());
484     }
485
486
487
488
489
490

```

```

471
472     }
473
474     // Cross validation + One Rule
475     else if (A4.isSelected() && B1.isSelected()) {
476
477
478         try {
479
480             // getting CrossValidation_Value input
481
482             CrossValidation_Value = Integer.parseInt(
483                 input1.getText());
484
485             // Create One Rule classifier by
486             // creating object of One Rule class
487             OneR OneRClassifier = new OneR();
488
489             // Dataset path
490             String DataStes = file.getPath();
491
492             // Creating bufferedreader to read the
493             // dataset
494             BufferedReader bufferedReader = new
495                 BufferedReader(new FileReader(DataStes)
496                     );
497
498             // Create dataset instances
499             Instances datasetInstances = new Instances(
500                 bufferedReader);
501
502             // Set Target Class
503             datasetInstances.setClassIndex(
504                 datasetInstances.numAttributes() - 1);
505
506             // Evaluating by creating object of
507             // Evaluation
508             // class
509             Evaluation evaluation = new Evaluation(
510                 datasetInstances);
511
512             // Cross Validate Model with user selected
513             // folds
514             evaluation.crossValidateModel(
515                 OneRClassifier, datasetInstances,
516                 CrossValidation_Value, new Random(1));
517
518             System.out.println(evaluation.
519                 toSummaryString("\nResults", false));
520             textArea.setText(evaluation.toSummaryString(
521                 "Cross validation + One Rule Evaluation
522                 Results : \n" + "\nPrecision = " +
523                 evaluation.precision(1) + "\nRecall = " +
524                 evaluation.recall(1) + "\nfMeasure = " +
525                 evaluation.fMeasure(1) + "\nError Rate = "
526                 + evaluation.errorRate() + "\n\n", false));

```

```
510         }
511
512         // Catch block to handle the exceptions
513         catch (Exception e) {
514
515             // Print message on the console
516             System.out.println("Error Occurred!!!! \n"
517                               + e.getMessage());
518
519         }
520
521     }
522
523     // Test File + J48
524     else if (A1.isSelected() && B2.isSelected()) {
525
526         try {
527
528             // Create J48 classifier by
529             // creating object of J48 class
530             J48 j48Classifier = new J48();
531
532             // Dataset path
533             String DataStes = file.getPath();
534             // testset path
535             String stastes2 = file2.getPath();
536
537             // Creating bufferedreader to read the
538             // dataset
539             BufferedReader bufferedReader = new
540                 BufferedReader(new FileReader(DataStes))
541                 ;
542
543             // Creating bufferedreader to read the
544             // testset
545             BufferedReader bufferedReader2 = new
546                 BufferedReader(new FileReader(stastes2))
547                 ;
548
549             // Create dataset instances
550             Instances datasetInstances = new Instances(
551                 bufferedReader);
552
553             // Create test instances
554             Instances testsetInstances = new Instances(
555                 bufferedReader2);
```

```
556
557
558     //build model
559
560     j48Classifier.buildClassifier(
        datasetInstances);
561
562     //use
563     Evaluation eval = new Evaluation(
        testsetInstances);
564     eval.evaluateModel(j48Classifier,
        testsetInstances);
565
566     System.out.println(eval.toSummaryString("\nResults", false));
567     textArea.setText(eval.toSummaryString("Test
        Set + J48 Evaluation Results : \n" + "\nPrecision = "+eval.precision(1)+"\nRecall = "+eval.recall(1)+"\nfMeasure = "+eval.fMeasure(1)+"\nError Rate = "+eval.errorRate()+"\n\n", false));
568 }
569
570 // Catch block to handle the exceptions
571 catch (Exception e) {
572
573     // Print message on the console
574     System.out.println("Error Occurred!!!! \n"
        + e.getMessage());
575 }
576
577
578
579
580 }
581
582
583 // Test File + RandomForest
584 else if (A2.isSelected() && B2.isSelected()) {
585
586
587     try {
588
589         // Create RandomForest classifier by
590         // creating object of RandomForest class
591         RandomForest RandomForestClassifier = new
            RandomForest();
592         RandomForestClassifier.setNumTrees(100);
593
594         // Dataset path
595         String DataStes = file.getPath();
596         // testset path
597         String stastes2 = file2.getPath();
598
599         // Creating bufferedreader to read the
            dataset
```

```

600         BufferedReader bufferedReader = new
            BufferedReader(new FileReader(DataStes))
            ;
601
602         // Creating bufferedreader to read the
            testset
603         BufferedReader bufferedReader2 = new
            BufferedReader(new FileReader(stastes2))
            ;
604
605         // Create dataset instances
606         Instances datasetInstances = new Instances(
            bufferedReader);
607
608         // Create test instances
609         Instances testsetInstances = new Instances(
            bufferedReader2);
610
611         // Set Target Class
612         testsetInstances.setClassIndex(
            testsetInstances.numAttributes() - 1);
613
614         datasetInstances.setClassIndex(
            datasetInstances.numAttributes() - 1);
615
616
617
618         //build model
619
620         RandomForestClassifier.buildClassifier(
            datasetInstances);
621
622         //use
623         Evaluation eval = new Evaluation(
            testsetInstances);
624         eval.evaluateModel(RandomForestClassifier,
            testsetInstances);
625
626         System.out.println(eval.toSummaryString("\nResults", false));
627         textArea.setText(eval.toSummaryString("Test
            Set + RandomForest Evaluation Results :
            \n" + "\nPrecision = "+eval.precision(1)
            + "\nRecall = "+eval.recall(1) + "\n
            nfMeasure = "+eval.fMeasure(1) + "\nError
            Rate = "+eval.errorRate() + "\n\n", false)
            );
628     }
629
630     // Catch block to handle the exceptions
631     catch (Exception e) {
632
633         // Print message on the console
634         System.out.println("Error Occurred!!!! \n"
            + e.getMessage());
635     }
636

```

```
637
638     }
639
640     // Test File + NaiveBayes
641     else if (A3.isSelected() && B2.isSelected()) {
642
643
644         try {
645
646             // Create NaiveBayes classifier by
647             // creating object of NaiveBayes class
648             NaiveBayes NaiveBayesClassifier = new
                NaiveBayes();
649
650             // Dataset path
651             String DataStes = file.getPath();
652             // testset path
653             String stastes2 = file2.getPath();
654
655             // Creating bufferedreader to read the
656             // dataset
657             BufferedReader bufferedReader = new
658                 BufferedReader(new FileReader(DataStes))
659                 ;
660
661             // Creating bufferedreader to read the
662             // testset
663             BufferedReader bufferedReader2 = new
664                 BufferedReader(new FileReader(stastes2))
665                 ;
666
667             // Create dataset instances
668             Instances datasetInstances = new Instances(
669                 bufferedReader);
670
671             // Create test instances
672             Instances testsetInstances = new Instances(
673                 bufferedReader2);
674
675             // Set Target Class
676             testsetInstances.setClassIndex(
677                 testsetInstances.numAttributes() - 1);
678
679             datasetInstances.setClassIndex(
680                 datasetInstances.numAttributes() - 1);
681
682
683             //build model
684
685             NaiveBayesClassifier.buildClassifier(
686                 datasetInstances);
687
688             //use
689             Evaluation eval = new Evaluation(
690                 testsetInstances);
```

```

680         eval.evaluateModel(NaiveBayesClassifier,
681                             testsetInstances);
682
683         System.out.println(eval.toSummaryString("\nResults", false));
684         textArea.setText(eval.toSummaryString("Test
685         Set + Naive Bayes Evaluation Results :
686         \n" + "\nPrecision = "+eval.precision(1)+
687         "\nRecall = "+eval.recall(1)+"\nfMeasure
688         = "+eval.fMeasure(1)+"\nError Rate = "+
689         eval.errorRate()+"\n\n", false));
690     }
691
692     // Catch block to handle the exceptions
693     catch (Exception e) {
694
695         // Print message on the console
696         System.out.println("Error Occurred!!!! \n"
697             + e.getMessage());
698     }
699 }
700
701 // Test File + One Rule
702 else if (A4.isSelected() && B2.isSelected()) {
703
704     try {
705
706         // Create One Rule classifier by
707         // creating object of One Rule class
708         OneR OneRClassifier = new OneR();
709
710         // Dataset path
711         String DataStes = file.getPath();
712         // testset path
713         String stastes2 = file2.getPath();
714
715         // Creating bufferedreader to read the
716         // dataset
717         BufferedReader bufferedReader = new
718             BufferedReader(new FileReader(DataStes))
719             ;
720
721         // Creating bufferedreader to read the
722         // testset
723         BufferedReader bufferedReader2 = new
724             BufferedReader(new FileReader(stastes2))
725             ;
726
727         // Create dataset instances
728         Instances datasetInstances = new Instances(
729             bufferedReader);
730
731         // Create test instances
732         Instances testsetInstances = new Instances(
733             bufferedReader2);

```



```

721
722         // Set Target Class
723         testsetInstances.setClassIndex(
724             testsetInstances.numAttributes() - 1);
725
726         datasetInstances.setClassIndex(
727             datasetInstances.numAttributes() - 1);
728
729         //build model
730
731         OneRClassifier.buildClassifier(
732             datasetInstances);
733
734         //use
735         Evaluation eval = new Evaluation(
736             testsetInstances);
737         eval.evaluateModel(OneRClassifier,
738             testsetInstances);
739
740         System.out.println(eval.toSummaryString("\nResults", false));
741         textArea.setText(eval.toSummaryString("Test
742             Set + One Rule Evaluation Results : \n"
743             + "\nPrecision = "+eval.precision(1)+"\n"
744             + "\nRecall = "+eval.recall(1)+"\n"
745             + "\nfMeasure = "+eval.fMeasure(1)+"\n"
746             + "\nError Rate = "+eval.errorRate()+"\n\n", false));
747     }
748
749     // Catch block to handle the exceptions
750     catch (Exception e) {
751
752         // Print message on the console
753         System.out.println("Error Occurred!!!! \n"
754             + e.getMessage());
755     }
756
757     }
758
759     // Percentage Split + J48
760     else if (A1.isSelected() && B3.isSelected()) {
761
762         try {
763
764             // Get Percentage from input
765
766             Split_Percentage = Integer.parseInt(input2.
767                 getText());
768
769             // Create J48 classifier by
770             // creating object of J48 class
771             J48 j48Classifier = new J48();
772
773             // Dataset path

```

```

765         String DataStes = file.getPath();
766
767
768         //          // Creating bufferedreader to read the
dataset
769         //          BufferedReader bufferedReader = new
BufferedReader(new FileReader(DataStes));
770         //
771         //
772         //          // Create dataset instances
773         //          Instances datasetInstances = new Instances(
bufferedReader);
774
775
776
777         String[] options = new String[4];
778         options[0] = "-t";
779         options[1] = DataStes;
780         options[2] = "-split-percentage";
781         options[3] = String.valueOf(
Split_Percentage);
782
783
784
785
786         //          // Randomize data
787         //          Randomize rand = new Randomize();
788         //          rand.setInputFormat(datasetInstances);
789         //          rand.setRandomSeed(42);
790         //          datasetInstances = Filter.useFilter(
datasetInstances, rand);
791         //
792         //          // Remove train percentage from data to
get the train set
793         //          RemovePercentage rp = new RemovePercentage
();
794         //          rp.setInputFormat(datasetInstances);
795         //          rp.setPercentage(Split_Percentage);
796         //          Instances test = Filter.useFilter(
datasetInstances, rp);
797         //
798         //          // Remove test percentage from data to get
the test set
799         //          rp = new RemovePercentage();
800         //          rp.setInputFormat(datasetInstances);
801         //          rp.setPercentage(Split_Percentage);
802         //          rp.setInvertSelection(true);
803         //          Instances train = Filter.useFilter(
datasetInstances, rp);
804
805
806         //
807         //
808         //          int trainSize = (int) Math.round(
datasetInstances.numInstances() * Split_Percentage / 100);
809         //          int testSize = datasetInstances.
numInstances() - trainSize;

```

```

810 //          Instances train = new Instances(
      datasetInstances, 0, trainSize);
811 //          Instances test = new Instances(
      datasetInstances, trainSize, testSize);
812
813
814 //          // Set Target Class
815 //          test.setClassIndex(test.numAttributes() -
      1);
816 //
817 //          train.setClassIndex(train.numAttributes() -
      1);
818 //
819 //          //build model
820 //
821 //          j48Classifier.buildClassifier(train);
822 //
823 //          //use
824
825 //System.out.println(Evaluation.
      evaluateModel(new J48(), options));
826 textArea.setText("Percentage Split + J48
      Evaluation Results : \n"+ Evaluation.
      evaluateModel(new J48(), options));
827     }
828
829 // Catch block to handle the exceptions
830 catch (Exception e) {
831
832     // Print message on the console
833     System.out.println("Error Occurred!!!! \n"
      + e.getMessage());
834 }
835
836
837
838
839 }
840
841
842 // Percentage Split + RandomForest
843 else if (A2.isSelected() && B3.isSelected()) {
844
845
846     try {
847
848
849
850         // Create RandomForest classifier by
851         // creating object of RandomForest class
852         RandomForest RandomForestClassifier = new
      RandomForest();
853         RandomForestClassifier.setNumTrees(100);
854         // Dataset path
855         String DataStes = file.getPath();
856
857

```

```

858 //                                // Creating bufferedreader to read the
dataset
859 //                                BufferedReader bufferedReader = new
BufferedReader(new FileReader(DataStes));
860 //
861 //
862 //                                // Create dataset instances
863 //                                Instances datasetInstances = new Instances(
bufferedReader);
864
865
866
867                                String[] options = new String[4];
868                                options[0] = "-t";
869                                options[1] = DataStes;
870                                options[2] = "-split-percentage";
871                                options[3] = String.valueOf(
Split_Percentage);
872
873
874 //                                // Randomize data
875 //                                Randomize rand = new Randomize();
876 //                                rand.setInputFormat(datasetInstances);
877 //                                rand.setRandomSeed(42);
878 //                                datasetInstances = Filter.useFilter(
datasetInstances, rand);
879 //
880 //                                // Remove train percentage from data to
get the train set
881 //                                RemovePercentage rp = new RemovePercentage
();
882 //                                rp.setInputFormat(datasetInstances);
883 //                                rp.setPercentage(Split_Percentage);
884 //                                Instances test = Filter.useFilter(
datasetInstances, rp);
885 //
886 //                                // Remove test percentage from data to get
the test set
887 //                                rp = new RemovePercentage();
888 //                                rp.setInputFormat(datasetInstances);
889 //                                rp.setPercentage(Split_Percentage);
890 //                                rp.setInvertSelection(true);
891 //                                Instances train = Filter.useFilter(
datasetInstances, rp);
892
893
894 //
895 //
896 //                                int trainSize = (int) Math.round(
datasetInstances.numInstances() * Split_Percentage / 100);
897 //                                int testSize = datasetInstances.
numInstances() - trainSize;
898 //                                Instances train = new Instances(
datasetInstances, 0, trainSize);
899 //                                Instances test = new Instances(
datasetInstances, trainSize, testSize);
900

```

```

901
902 //                      // Set Target Class
903 //                      test.setClassIndex(test.numAttributes() -
//                      1);
904 //
905 //                      train.setClassIndex(train.numAttributes() -
//                      1);
906 //
907 //
908 //
909 //                      //build model
910 //
911 //                      RandomForestClassifier.buildClassifier(
train);
912 //
913 //                      //use
914 //                      Evaluation eval = new Evaluation(test);
915 //                      eval.evaluateModel(RandomForestClassifier,
test);
916
917 //                      System.out.println(eval.toSummaryString("\n
nResults", false));
918 //                      textArea.setText(eval.toSummaryString("Test
Set + RandomForest Evaluation Results : \n"+ "\nPrecision = "+
eval.precision(1)+"\nRecall = "+eval.recall(1)+"\nfMeasure = "+
eval.fMeasure(1)+"\nError Rate = "+eval.errorRate()+"\n\n",
false));
919
//                      textArea.setText("Percentage Split + Random
//                      Forset Evaluation Results : \n"+
//                      Evaluation.evaluateModel(
//                      RandomForestClassifier, options));
920
//                      }
921
//                      // Catch block to handle the exceptions
922 catch (Exception e) {
923
//                      // Print message on the console
924
//                      System.out.println("Error Occurred!!!! \n"
+ e.getMessage());
925
//                      }
926
//                      }
927
//                      }
928
//                      }
929
//                      }
930
//                      }
931
//                      // Percentage Split + NaiveBayes
932 else if (A3.isSelected() && B3.isSelected()) {
933
//                      try {
934
//                      // Create NaiveBayes classifier by
//                      // creating object of NaiveBayes class
935
//                      NaiveBayes NaiveBayesClassifier = new
NaiveBayes();
936
//                      // Dataset path
937
//                      String DataStes = file.getPath();
938

```

```

944
945
946 //                      // Creating bufferedreader to read the
dataset
947 //                      BufferedReader bufferedReader = new
BufferedReader(new FileReader(DataStes));
948 //
949 //
950 //                      // Create dataset instances
951 //                      Instances datasetInstances = new Instances(
bufferedReader);
952
953
954
955
956 String[] options = new String[4];
957 options[0] = "-t";
958 options[1] = DataStes;
959 options[2] = "-split-percentage";
960 options[3] = String.valueOf(
Split_Percentage);
961
962
963 //                      // Randomize data
964 //                      Randomize rand = new Randomize();
965 //                      rand.setInputFormat(datasetInstances);
966 //                      rand.setRandomSeed(42);
967 //                      datasetInstances = Filter.useFilter(
datasetInstances, rand);
968 //
969 //                      // Remove train percentage from data to
get the train set
970 //                      RemovePercentage rp = new RemovePercentage
();
971 //                      rp.setInputFormat(datasetInstances);
972 //                      rp.setPercentage(Split_Percentage);
973 //                      Instances test = Filter.useFilter(
datasetInstances, rp);
974 //
975 //                      // Remove test percentage from data to get
the test set
976 //                      rp = new RemovePercentage();
977 //                      rp.setInputFormat(datasetInstances);
978 //                      rp.setPercentage(Split_Percentage);
979 //                      rp.setInvertSelection(true);
980 //                      Instances train = Filter.useFilter(
datasetInstances, rp);
981
982
983 //
984 //
985 //                      int trainSize = (int) Math.round(
datasetInstances.numInstances() * Split_Percentage / 100);
986 //                      int testSize = datasetInstances.
numInstances() - trainSize;
987 //                      Instances train = new Instances(
datasetInstances, 0, trainSize);

```

```

988 //          Instances test = new Instances(
          datasetInstances, trainSize, testSize);
989
990
991 //          // Set Target Class
992 //          test.setClassIndex(test.numAttributes() -
          1);
993
994 //          train.setClassIndex(train.numAttributes() -
          1);
995 //
996 //
997 //
998 //          //build model
999 //
1000 //          NaiveBayesClassifier.buildClassifier(
          datasetInstances);
1001 //
1002 //          //use
1003 //          Evaluation eval = new Evaluation(test);
1004 //          eval.evaluateModel(NaiveBayesClassifier,
          test);
1005
1006 //          System.out.println(eval.toSummaryString("\nResults", false));
1007 //          textArea.setText(eval.toSummaryString("Test
          Set + Naive Bayes Evaluation Results : \n"+ "\nPrecision = "+
          eval.precision(1)+"\nRecall = "+eval.recall(1)+"\nfMeasure = "+
          eval.fMeasure(1)+"\nError Rate = "+eval.errorRate()+"\n\n",
          false));
1008
          textArea.setText("Percentage Split + Naive
          Bayes Evaluation Results : \n"+
          Evaluation.evaluateModel(new NaiveBayes
          (), options));
1009
          }
1010
          // Catch block to handle the exceptions
1011 catch (Exception e) {
1012
          // Print message on the console
1013 System.out.println("Error Occurred!!!! \n"
          + e.getMessage());
1014
          }
1015
          }
1016
          }
1017
          }
1018
          }
1019
          }
1020
          // Percentage Split + One Rule
1021 else if (A4.isSelected() && B3.isSelected()) {
1022
          try {
1023
          //
          // Create One Rule classifier by
1024 // creating object of One Rule class
1025 OneR OneRClassifier = new OneR();
1026
          // Dataset path
1027
1028
1029
1030

```

```

1031         String DataStes = file.getPath();
1032
1033
1034         //          // Creating bufferedreader to read the
1035         dataset      BufferedReader bufferedReader = new
1036         BufferedReader(new FileReader(DataStes));
1037
1038         //          // Create dataset instances
1039         //          Instances datasetInstances = new Instances(
1040         bufferedReader);
1041
1042
1043         String[] options = new String[4];
1044         options[0] = "-t";
1045         options[1] = DataStes;
1046         options[2] = "-split-percentage";
1047         options[3] = String.valueOf(
1048             Split_Percentage);
1049
1050
1051
1052         //          // Randomize data
1053         //          Randomize rand = new Randomize();
1054         //          rand.setInputFormat(datasetInstances);
1055         //          rand.setRandomSeed(42);
1056         //          datasetInstances = Filter.useFilter(
1057         datasetInstances, rand);
1058
1059         //          // Remove train percentage from data to
1060         //          get the train set
1061         //          RemovePercentage rp = new RemovePercentage
1062         ();
1063         //          rp.setInputFormat(datasetInstances);
1064         //          rp.setPercentage(Split_Percentage);
1065         //          Instances test = Filter.useFilter(
1066         datasetInstances, rp);
1067
1068         //          // Remove test percentage from data to get
1069         //          the test set
1070         //          rp = new RemovePercentage();
1071         //          rp.setInputFormat(datasetInstances);
1072         //          rp.setPercentage(Split_Percentage);
1073         //          rp.setInvertSelection(true);
1074         //          Instances train = Filter.useFilter(
1075         datasetInstances, rp);
1076
1077         //          int trainSize = (int) Math.round(
1078         datasetInstances.numInstances() * Split_Percentage / 100);
1079         //          int testSize = datasetInstances.
1080         numInstances() - trainSize;

```



```

1076 //          Instances train = new Instances(
1077 //          datasetInstances, 0, trainSize);
1078
1079 //          Instances test = new Instances(
1080 //          datasetInstances, trainSize, testSize);
1081
1082 //          // Set Target Class
1083 //          test.setClassIndex(test.numAttributes() -
1084 //          1);
1085 //          train.setClassIndex(train.numAttributes() -
1086 //          1);
1087 //          //build model
1088 //          OneRClassifier.buildClassifier(train);
1089 //          //use
1090 //          Evaluation eval = new Evaluation(test);
1091 //          eval.evaluateModel(OneRClassifier, test);
1092 //          System.out.println(eval.toSummaryString("\nResults", false));
1093 //          textArea.setText(eval.toSummaryString("Test
1094 //          Set + One Rule Evaluation Results : \n"+ "\nPrecision = "+eval.
1095 //          precision(1)+"\nRecall = "+eval.recall(1)+"\nfMeasure = "+eval.
1096 //          fMeasure(1)+"\nError Rate = "+eval.errorRate()+"\n\n", false));
1097 //          textArea.setText("Percentage Split + One
1098 //          Rule Evaluation Results : \n"+
1099 //          Evaluation.evaluateModel(new OneR(),
1100 //          options));
1101 //          }
1102 //          // Catch block to handle the exceptions
1103 //          catch (Exception e) {
1104 //          // Print message on the console
1105 //          System.out.println("Error Occurred!!!! \n"
1106 //          + e.getMessage());
1107 //          }
1108 //          }
1109 //          });
1110 btnNewButton_2.setBounds(728, 151, 117, 25);
1111 frmTpDataLining.getContentPane().add(btnNewButton_2);
1112
1113 JLabel lblNewLabel = new JLabel("Done by Hadjazi Mohammed
1114 //          Hisham G01 RSSI");
1115 lblNewLabel.setFont(new Font("Dialog", Font.BOLD | Font.
1116 //          ITALIC, 12));
1117 lblNewLabel.setHorizontalAlignment(SwingConstants.RIGHT);
1118 lblNewLabel.setBounds(494, 541, 354, 15);
1119 frmTpDataLining.getContentPane().add(lblNewLabel);

```

```
1118
1119
1120
1121
1122     JLabel lblFolds = new JLabel("Folds");
1123     lblFolds.setHorizontalAlignment(SwingConstants.CENTER);
1124     lblFolds.setBounds(797, 51, 52, 15);
1125     frmTpDataLining.getContentPane().add(lblFolds);
1126
1127     JLabel lblFolds_1 = new JLabel("%");
1128     lblFolds_1.setFont(new Font("Dialog", Font.BOLD | Font.
1129         ITALIC, 14));
1129     lblFolds_1.setHorizontalAlignment(SwingConstants.CENTER);
1130     lblFolds_1.setBounds(796, 116, 52, 15);
1131     frmTpDataLining.getContentPane().add(lblFolds_1);
1132
1133
1134
1135
1136
1137
1138
1139     }
1140 }
1141
```

Bibliography

- [1] *7 Types of Classification Algorithms*. en-US. Jan. 2018. URL: <https://analyticsindiamag.com/7-types-classification-algorithms/> (visited on 03/05/2022).
- [2] Shawkat Ali and Kate A. Smith. "On learning algorithm selection for classification". en. In: *Applied Soft Computing* 6.2 (Jan. 2006), pp. 119–138. ISSN: 15684946. DOI: 10.1016/j.asoc.2004.12.002. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1568494605000049> (visited on 03/05/2022).
- [3] *Data Mining Map*. URL: <https://www.saedsayad.com/> (visited on 03/05/2022).
- [4] *Data mining: the textbook*. 1st edition. New York, NY: Springer Science+Business Media, 2015. ISBN: 9783319141411.
- [5] *data/star-wars-survey at master · fivethirtyeight/data*. en. URL: <https://github.com/fivethirtyeight/data> (visited on 03/23/2022).
- [6] *Decision Trees – C4.5*. en. Mar. 2011. URL: <https://octaviansima.wordpress.com/2011/03/25/decision-trees-c4-5/> (visited on 03/05/2022).
- [7] *Documentation - Weka Wiki*. URL: <https://waikato.github.io/weka-wiki/documentation/> (visited on 03/06/2022).
- [8] Walt Hickey. *America's Favorite 'Star Wars' Movies (And Least Favorite Characters)*. en-US. July 2014. URL: <https://fivethirtyeight.com/features/americas-favorite-star-wars-movies-and-least-favorite-characters/> (visited on 03/23/2022).
- [9] Uday Kamath and Krishna Choppella. *Mastering Java machine learning: a java developer's guide to implementing machine learning and big data architectures*. English. OCLC: 1059420113. 2017. ISBN: 9781785880513.
- [10] I. H. Witten, Eibe Frank, and Mark A. Hall. *Data mining: practical machine learning tools and techniques*. 3rd ed. Morgan Kaufmann series in data management systems. OCLC: ocn262433473. Burlington, MA: Morgan Kaufmann, 2011. ISBN: 9780123748560.
- [11] I. H. Witten and I. H. Witten, eds. *Data mining: practical machine learning tools and techniques*. Fourth Edition. Amsterdam: Elsevier, 2017. ISBN: 9780128042915.
- [12] Tony Yiu. *Understanding Random Forest*. en. Sept. 2021. URL: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (visited on 03/05/2022).