

Contraceptive Method Used

Chi Ting Low & Vidya Viswanathan

Email: s3611774@student.rmit.edu.au

s3613547@student.rmit.edu.au

Date: 22/05/2018

Table of Contents

<i>Abstract</i>	<i>3</i>
<i>Introduction</i>	<i>4</i>
<i>Methodology</i>	<i>5</i>
<i>Results</i>	<i>8</i>
<i>Model Selection</i>	<i>15</i>
<i>Results.....</i>	<i>16</i>
<i>Decision Tree Classification</i>	<i>16</i>
<i>K Nearest Neighbor Classification.....</i>	<i>17</i>
<i>Random Forest</i>	<i>18</i>
<i>Discussion</i>	<i>19</i>
<i>Conclusion.....</i>	<i>20</i>
<i>References.....</i>	<i>21</i>

Abstract

The aim of the current assignment is to identify the best classification method that will predict the choice of contraceptive method. By using decision tree, K nearest neighbor and random forest classification method, it shows that decision tree has the highest accuracy rate compared to others classification method. In addition, using method Grid Search the accuracy has increase from 42% to 62.9% in decision tree.

Introduction

Contraception is a method used by individuals to prevent pregnancy (Department of Health Western Australia, 2018). There are many different methods of birth control and it is available in different forms, such as hormonal, barrier, intrauterine devices or behavioral ("Birth Control Methods & Options | Types of Birth Control", 2018). Hormonal refers to method such as birth control pill; Barrier refers to condom or any method that prevent pregnancy; Intrauterine devices refers to birth control device that implant in women or vasectomy on male; Behavioural refers to methods such as abstinence or withdrawal method ("Birth Control Methods & Options | Types of Birth Control", 2018).

According to Guttmacher Institute demographic and health surveys (Sedgh, Ashford & Hussain, 2016) in 52 countries between 2005 and 2014 estimated that 220 million women does not using any contraceptive method or using a traditional method to prevent pregnancy. It was estimated that if modern contraceptive methods were introduced to these countries around 52 million unwanted pregnancy and 70000 pregnancy related deaths can be prevented (Sedgh, Ashford & Hussain, 2016). Therefore, it is important to understand what factors can influence the choice and used of contraceptive methods.

World Health Organization suggested that there are many benefits using contraceptive method (World Health Organization, n.d). World Health Organization suggested that by using contraceptive method, it can prevent women from pregnancy-related health risk, reduce infant mortality rate, prevent HIV or AIDS and reduce teenagers pregnancies rates (World Health Organization, n.d). There are many factors influence the choice and used of different contraceptive method, social-psychological factors (Fekadu, 2017), lack of education and

information (Biggs et. al, 2014), different preference between racial and ethnic group (Jackson, Karasek, Dehlendorf & Foster, 2015), peer pressure among adolescents (Melo, Peters, Teal, Guiahi, 2015), gender and health education (Ringheim, 1993) and other factors that influence the uses of contraceptive methods.

The aim of the current project is to using two different classification methods; decision tree (Quinlan, 1986) and K Nearest Neighbor Classifier (Cover & Hart, 1967) to classify the dataset and understand which factor is more important factor that influence the choice of contraceptive method. Decision tree is used for this project is because it is easy to understand, robust and help to provides an evaluation framework to help to determine which features are important (books citation). In addition, K Nearest Neighbor Classifier (KNN) is chosen for this project is to understand the pattern of the contraceptive method choice based on the given variables (Hassanat, Abbadi, Altarawneh & Alhasanat, 2014).

Methodology

The dataset used in current project is Contraceptive Method Choice from UCL Machine Learning Repository. There are 1474 observations and 10 variables in the current dataset. The variables information is provided below:

1. Wife's age (numerical)
2. Wife's education (categorical, 1 = low, 4 = high)
3. Husband's education (categorical, 1 = low, 4 = high)
4. Number of children ever born (numerical)
5. Wife's religion (binary/categorical, 0 = Non-Islam, 1 = Islam)
6. Wife's employment (binary/categorical, 0 = Yes, 1 = No)

7. Husband's occupation (categorical, 1 = low, 4 = high)
8. Standard of living (categorical, 1 = low, 4 = high)
9. Media exposure (binary/categorical, 0 = Good, 1 = Not Good)
10. Contraceptive Method Used (class attribute/categorical, 1 = No-use, 2 = Long-term, 3 = Short-term)

Prior to the exploratory analysis the categorical variables are recoded from numerical to string and converted into categorical data types based on Harris (2002) naming method on the variables. Table 1 shows the descriptive statistics of wife's and number of children they have. It shows that the mean age is 32.54 years old and average number of children are 3.26.

Table 1. Descriptive statistics of Wife's age and Number of Children

Variables	Wife Age	Number of Children
Count	1473.00	1473.00
Mean	32.54	3.26
Standard Deviation	8.23	2.36
Minimum	16.00	0.00
25%	26.00	1.00
50%	32.00	3.00
75%	49.00	4.00

Max	49.00	16.00
-----	-------	-------

Figure 1 and *Figure 2* shows the density plot of the wife's age and the number of children owned. It shows that wife's age is following close to normal distribution. Majority of the distribution fall between the age of 20 and 40. However, in *Figure 2* there is a positive skew of the density plot where majority of point skews towards left side of the plot.

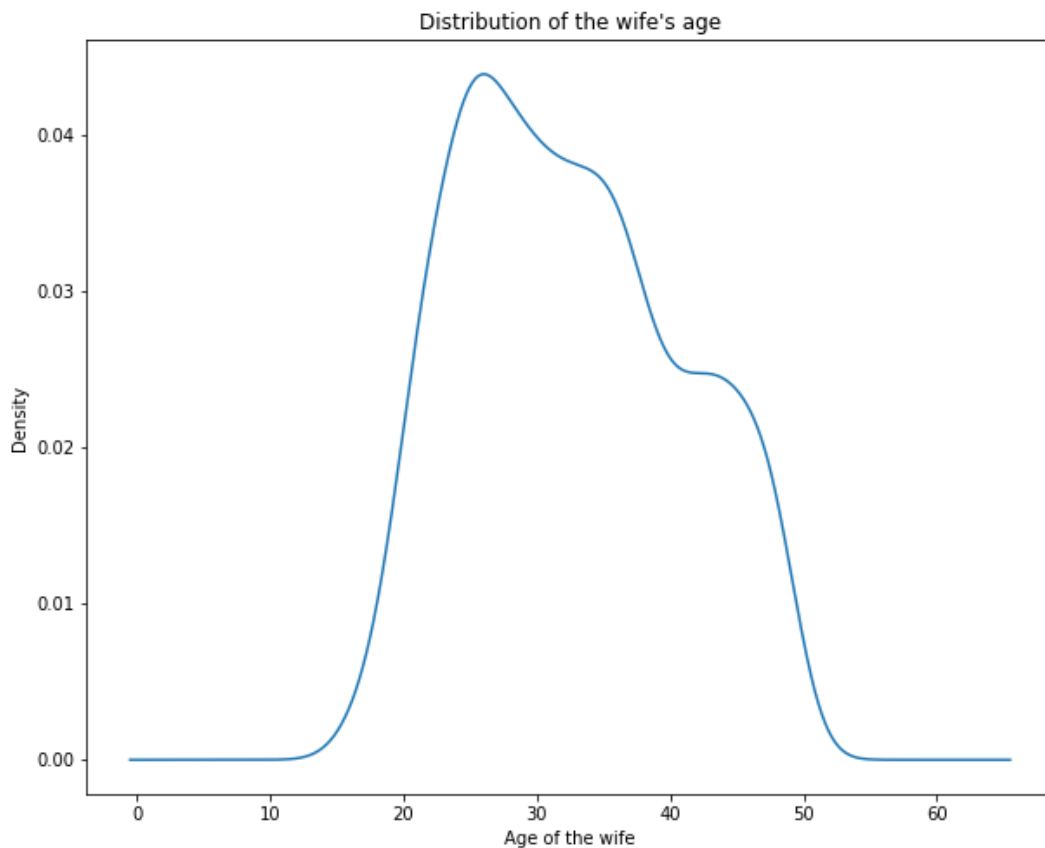


Figure 1. The Density Plot of Wife Age

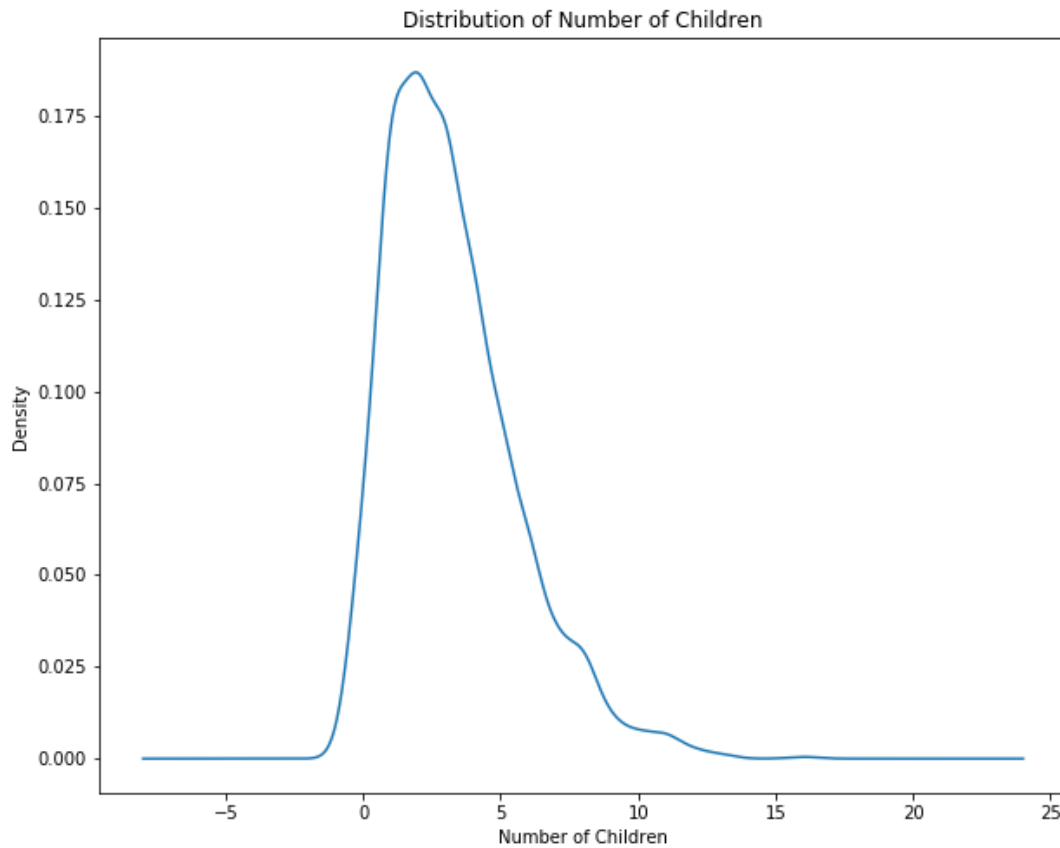


Figure 2. The Density Plot of Number of Children

Results

Preliminary analysis shows that there is no difference of using contraceptive method among the group with high standard of living. It shows that the choice of contraceptive methods does not change when they have high standard of living. However, *Figure 3* shows that majority of the women does not using any contraceptive methods regardless of their standard of living status.

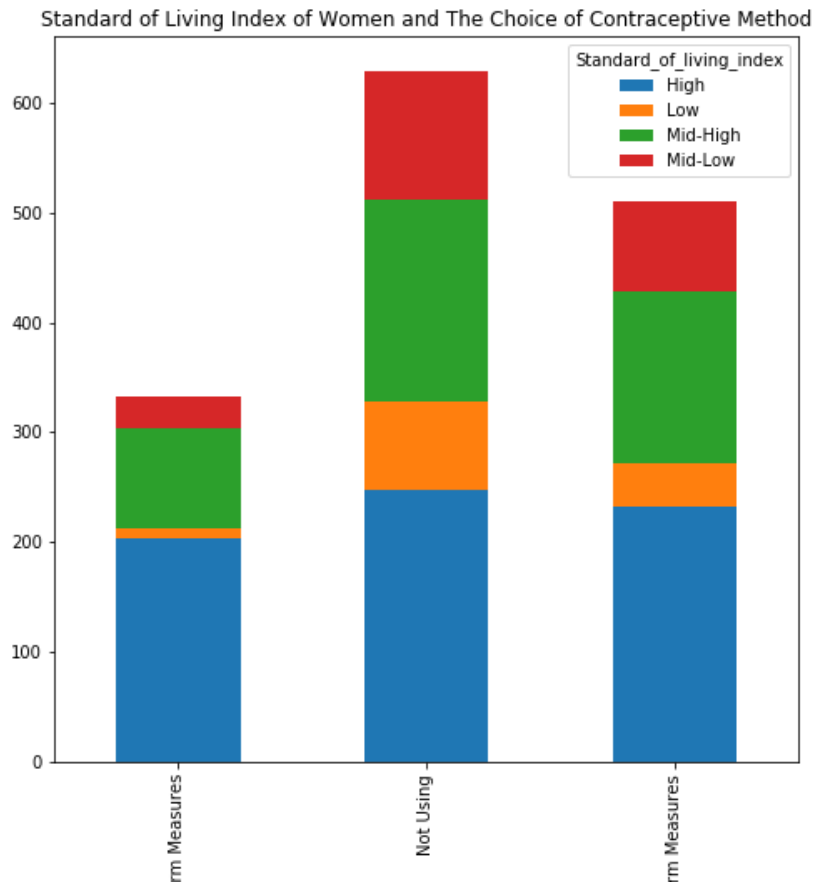


Figure 3. Standard of Living of Women and The Choice of Contraceptive Method

Table 2 shows the relationship between the used of contraceptive method among working wife. It shows that women who is not working prefers to used long term and short-term contraceptive. However, there are still majority of the women not using any contraceptive methods.

Table 2. The Use of Contraceptive Method Among Working Wife.

Working Wife	No	Yes
--------------	----	-----

Contraceptive Method Used		
Long Term	0.73	0.27
Not Using	0.73	0.27
Short Term	0.78	0.22

Figure 3 shows the relationship between husband occupation and the used of contraceptive method. It shows that majority of the husbands do not use any contraceptive method. However, some with significant high-level occupation choose to use some form of short term contraceptive method. Furthermore, minority of them are using long term contraceptive method.

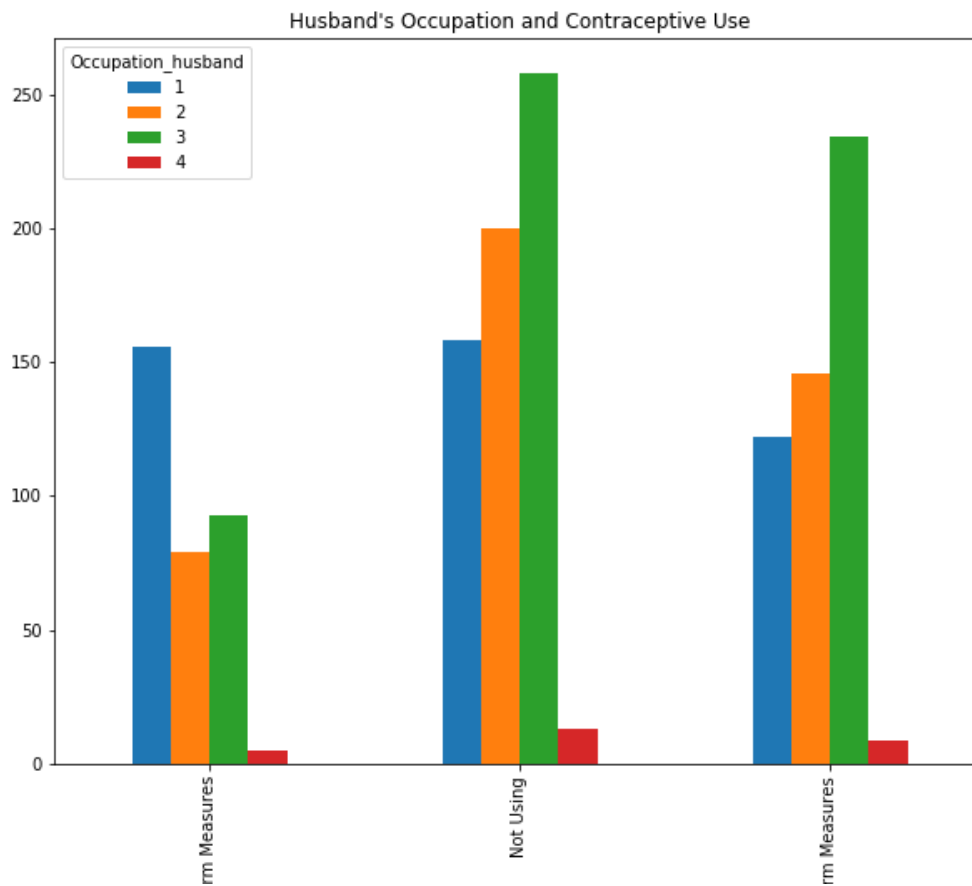


Figure 4. Husband's Occupation and Contraceptive Use

Figure 5 shows that high level of education tends to use some form of contraceptive method whether it is long-term or short-term. However, it shows that majority of husbands with high level of education are identify as not using any form of contraceptive method.

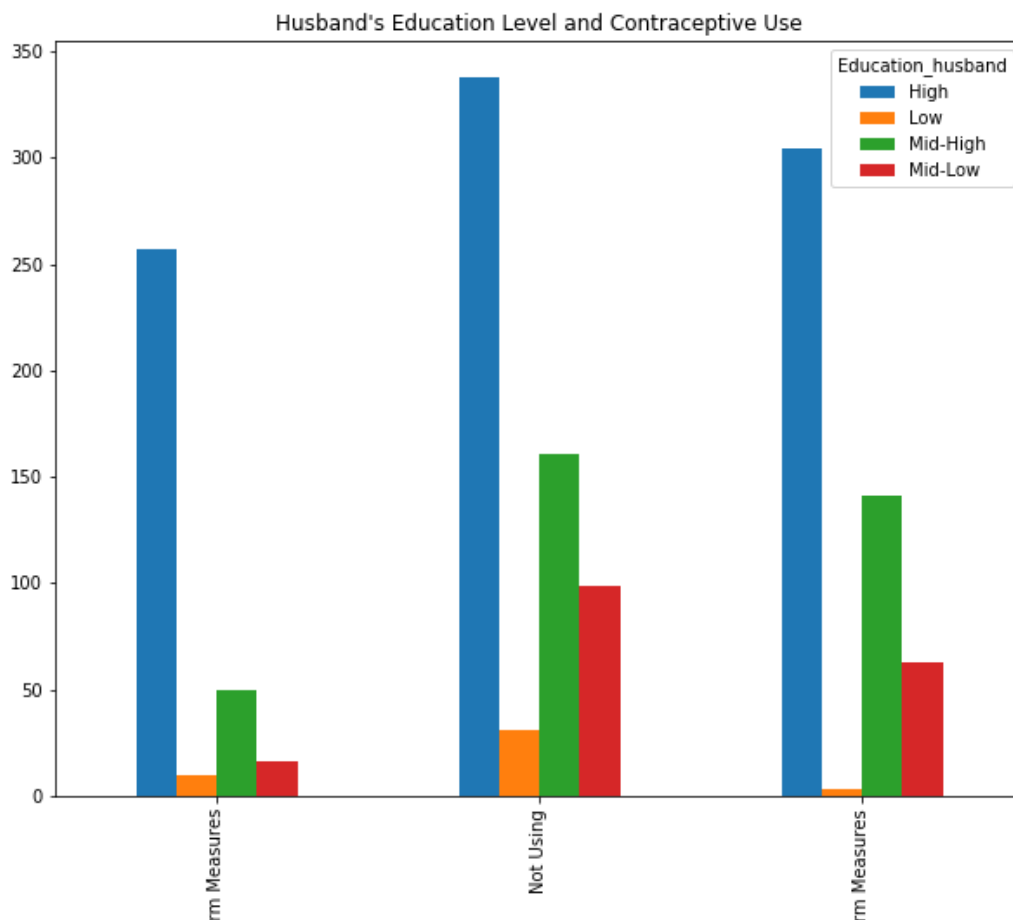


Figure 5. Husband's Education Level and The Used of Contraceptive Method

Figure 6 shows the relationship between wife's education and the choice of contraceptive methods. It shows that wife with higher education are more likely to use short-term and long-term contraceptive methods. However, majority of them are choosing not to use any form of contraceptive method.

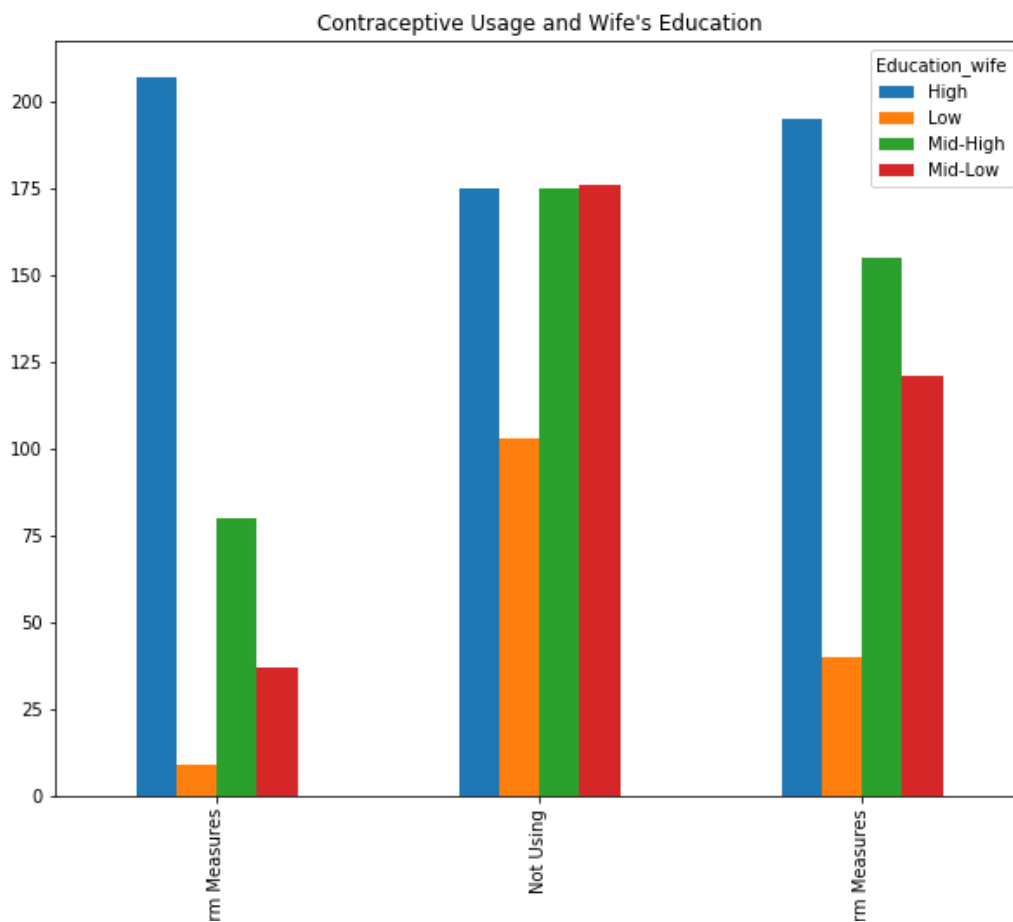


Figure 6. Wife's Education Level and The Choice of Contraceptive Method.

Figure 7 shows the relationship between media exposure and the choice of contraceptive method. It shows that with good media exposure they are more likely to use short-term and long-term contraceptive methods. However, figure 7 also shows that even with good media exposure

majority of them are choosing not to use any contraceptive method.

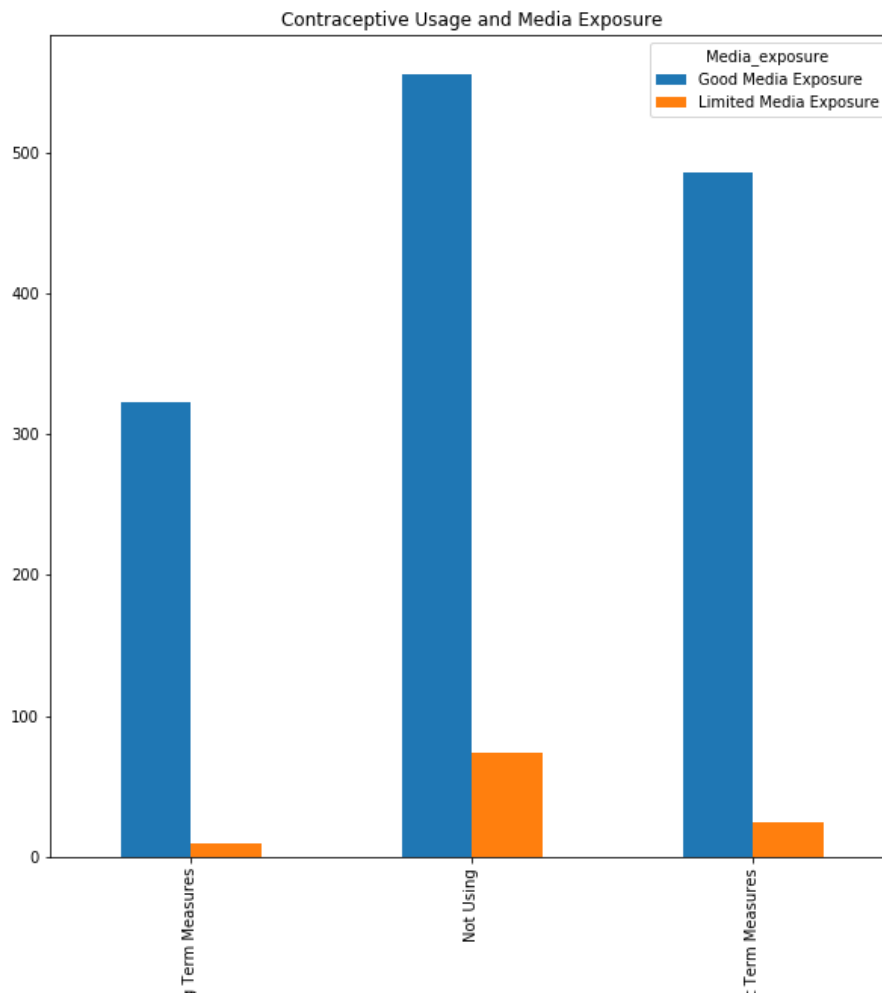


Figure 7. Media Exposure and the Choice of Contraceptive Method.

According to *Figure 8*, majority of the women who chose not to use any contraceptive measures indicated Islam as their religion. The number of women who did not practice Islam were evenly distributed over the 3 contraceptive usage groups.

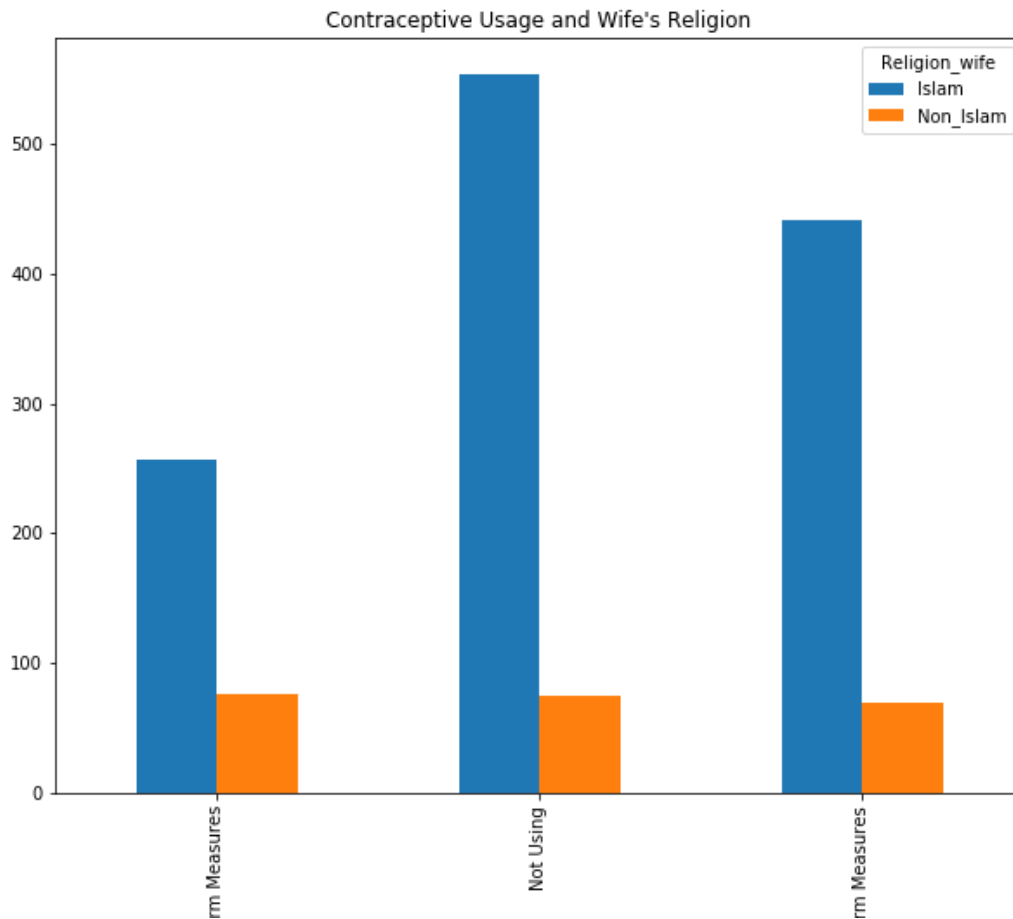


Figure 8: The Relationship Between Wife's Religion and Contraceptive Choice.

Model Selection

Prior to fitting the models, the categorical and the ordinal variables in the dataset were converted to numeric values using the 'one-hot-encoding' technique. The *pd.get_dummies()* function was used to perform this task. The Decision Tree classifier, K-Nearest Neighbor classifier and Random Forest were used to perform this classification task. To train the classifiers the data was split into training and test data.

The classifier was trained on the train data the learned model was generalized on the test data in order to test the classifier model's prediction. The *train_test_split()* function from *sklearn* was used to partition the data. The *test_size* was set at 0.3. The training data and the test data obtained was 70% and 30% of the dataset.

Results

The Confusion Matrix, F1-Score, accuracy, precision and recall for each of the 3 models is presented in *Tables 3, 4* and *5*. The metrics in these tables will be used to evaluate the best predictor under each model.

Decision Tree Classification

A decision tree classifier that builds classification models in the form of a tree structure was fit on the dataset to predict the target. Table 3 gives the scores for the model after tuning the parameters.

Predicted Contraceptive Choice	1	2	3
Actual Contraceptive Choice			
1	135	10	36
2	41	52	24
3	45	26	73

Figure 11: Confusion Matrix for the Decision Tree classifier

Table 3: Metrics for Decision Tree Classifier

Decision Tree

Metrics

Precision	0.59
Recall	0.59
F1-score	0.58
Support	442
Classification Error Rate	0.41

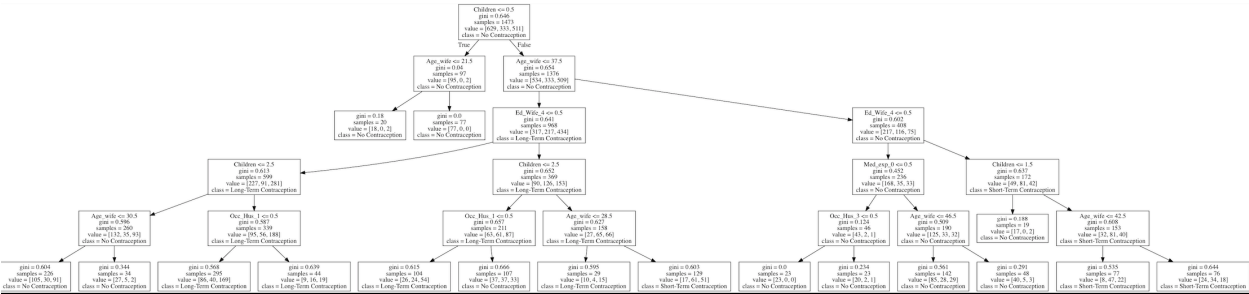


Figure 10: Decision Tree generated using the Decision Tree classifier

K Nearest Neighbor Classification

The K-nearest neighbor classifier algorithms predicts the target label by finding the nearest neighbor class which is identified using the distance measures like Euclidean distance. Table 4 gives the scores for the model after tuning the parameters.

Table 4: Metrics for K-Nearest Neighbor Classifier

K-Nearest Neighbor	
Metrics	
Precision	0.57
Recall	0.57
F1-score	0.57
Support	442
Classification Error Rate	0.43

Predicted Contraceptive Choice	1	2	3
Actual Contraceptive Choice			
1	123	19	39
2	35	55	27
3	40	30	74

Figure 12: Confusion Matrix for the KNN classifier

Random Forest

Random forest algorithm is an ensemble classification algorithm or a group of classifiers. Instead of using only one classifier to predict the target, this classification model uses multiple classifiers to predict the target. Table 5 gives the scores for the model after tuning the parameters.

Table 5: Metrics for Random Forest Classifier

Random Forest

Metrics	
Precision	0.53
Recall	0.51
F1-score	0.49
Support	442
Classification Error Rate	0.49

Predicted Contraceptive Choice	1	2	3
Actual Contraceptive Choice			
1	129	5	47
2	44	27	46
3	62	12	70

Figure 13: Confusion Matrix for the KNN classifier

Discussion

The Decision Tree Classifier was first fit on the data with set parameters. This model yielded a low accuracy score of 42.76 with a high error rate of 0.57. A confusion matrix generated for this model had an average precision rate of 0.32. Grid Search was done to tune the *max_depth* and *min_samples_leaf* parameters for this model. The best parameters and the best estimators were generated, and a new decision tree classifier was built. The prediction based on the new hyper-tuned parameters gave an improved average Accuracy score of 62.9 %with an 0.41 error rate.

The decision tree generated for this model has a key split at the number of children for women after which it splits the tree at the Age of the woman. If a woman had no children and was under the age of 25, then she was most likely to not use any form of contraception.

For the K-Nearest Neighbor model, Grid Search was employed to sift through models with different parameters. In the Grid Search, the weights parameters were adjusted to test the default *Uniform* option that weighed all points in the neighborhood uniformly and using the *Distance* option that weighed nearest neighbors more heavily. The Grid Search yielded a model with the k adjusted to 15 with *weights* set to *Uniform*. The model gave an average precision of 0.57 with an error rate of 0.43. The accuracy score was 57.01% for the model.

In the Random Forest model, the *max_depth* and *min_samples_leaf* parameters were to get the best model. The Random Forest model with tuned parameters gave a Precision Score of 0.56 and an error rate of 0.45.

Conclusion

Based on the metrics obtained from the evaluation of the classifiers, the Decision Tree Classifier had the highest Accuracy and Precision and is the best model of the three to make predictions about the preferred contraceptive choices of the women in the Contraceptive dataset.

References

Birth Control Methods & Options | Types of Birth Control. (2018). *Plannedparenthood.org*.

Retrieved 27 April 2018, from <https://www.plannedparenthood.org/learn/birth-control>

Biggs, M. A., Harper, C. C., Malvin, J., & Brindis, C. D. (2014). Factors influencing the provision of long-acting reversible contraception in California. *Obstetrics & Gynecology*, 123(3), 593-602.

Department of Health Western Australia. (n.d.). Types of contraception. Retrieved from <https://www.getthefacts.health.wa.gov.au/condoms-contraception/types-of-contraception>

Fekadu, Z. (2017). Social-psychological factors associated with contraceptive attitudes of married women in the Kechene Community of Addis Ababa, Ethiopia. *The Ethiopian Journal of Health Development (EJHD)*, 10(3).

Harris, E. (2002). Information Gain Versus Gain Ratio: A Study of Split Method Biases. In *ISAIM*.

Jackson, A. V., Karasek, D., Dehlendorf, C., & Foster, D. G. (2016). Racial and ethnic differences in womens preferences for features of contraceptive methods. *Contraception*, 93(5), 406-411. doi:10.1016/j.contraception.2015.12.010

Melo, J., Peters, M., Teal, S., & Guiahi, M. (2013). Adolescent and young women's contraceptive decision-making processes: Choosing "the best method for her". *Contraception*, 88(3), 459. doi:10.1016/j.contraception.2013.05.111

Ringheim, K. (1993). Factors that Determine Prevalence of Use of Contraceptive Methods for Men. *Studies in Family Planning*, 24(2). doi:10.2307/2939202

Sedgh, G., Ashoford, L. S., & Hussain, R. (2016). Unmet need for contraception in developing countries: examine women's reasons for not using a method. The Guttmacher Institute.

World Health Organization. (n.d.). Family planning/Contraception. Retrieved April 28, 2018, from <http://www.who.int/en/news-room/fact-sheets/detail/family-planning-contraception>

Chris Albon. (2017, December 20). One-Hot Encode Nominal Categorical Features. Retrieved from https://chrisalbon.com/machine_learning/preprocessing_structured_data/one-hot_encode_nominal_categorical_features/

Tuning the hyper-parameters of an estimator. (n.d.). Retrieved May 01, 2018, from http://scikit-learn.org/stable/modules/grid_search.html

Tuning the hyper-parameters of an estimator. (n.d.). Retrieved May 01, 2018, from http://scikit-learn.org/stable/modules/grid_search.html#exhaustive-grid-search

Optimal Tuning Parameters. (n.d.). Retrieved from <http://www.ritchieng.com/machine-learning-efficiently-search-tuning-param/>