

EARLY DIVORCE PREDICTION IN ABU DHABI

IAOS Conference 2016 Abu Dhabi, UAE

Authors: Razan Ismail Ahmed Mohamed, Mohammed Khalifa Alkhyeli

Statistics Centre – Abu Dhabi
P.O. Box: 6036 Abu Dhabi, UAE
T: +971.2.81.00.000

Abu Dhabi, August 2016

Abstract:

This paper analyse the demographic variables and status at time of marriage for the couples that can predict divorce before completing three years of marriage in Abu Dhabi Emirate. It uses administrative data available at time of marriage and predict divorce through applying data mining techniques for classification, namely logistic regression and Decision trees.

Key words: Predict divorce, demographic reasons for divorce, Data mining, Abu Dhabi, logistic regression, Decision tree, administrative records

1. Introduction

Divorce phenomena has increased in the recent decades affecting the cohesion of the family, the nuclear of the community. Although divorce rate has declined slightly in Abu Dhabi Emirate, Early divorce (from zero to three years) has rapidly increased. It has social and economic implications and that is why is useful to analyse the reasons behind it. Suggested explanations range from economic trends to cultural shifts and legal changes. There are many dimensions lead couples to the risk of divorce, such as demographic characteristics, Socio economic characteristics and couples' behaviours. In this paper, I will concentrate on the influence of demographic indicators on early divorce. It would be examined through data mining techniques to predict whether a couple will be in the risk of getting divorce in the first three years. This study will be based on recent administrative data of marriage and divorce of Abu Dhabi Emirate.

Data mining techniques applied to social issues are new compared to business applications. It is exciting that this study can help us understand application of data mining techniques to administrative records of divorces. Finding predictors of divorce cases can provide an enhancement to divorce statistics. Furthermore, Many government organizations in Abu Dhabi Emirate can benefit from this model and prediction, such as judicial department of Abu Dhabi and Family foundation to draw policies that can limit early divorce.

1.1 Abu Dhabi

Abu Dhabi is the federal capital of the United Arab Emirates (UAE) and the largest of the seven emirates. Geographically, Abu Dhabi lies on the borders with the Kingdom of Saudi Arabia, the Sultanate of Oman, and the Arabian Gulf. Over the past 40 years, Abu Dhabi has experienced significant population growth and economic development.

To manage the growth and prosperity of the Emirate, the Government of Abu Dhabi required an official agency that could provide statistics for decision-making and policy setting.

1.2 Statistics Centre – Abu Dhabi

Statistics Centre – Abu Dhabi (SCAD) was established in accordance with Law #7 for the year 2008. SCAD is responsible for the collection, classification, storage, analysis and dissemination of official statistics covering social, demographic, economic, environmental and cultural indicators.

As a young statistical office, SCAD is in the fortunate position of being able to implement best practices from international bodies and leading National Statistical Organisations (NSOs). SCAD is aiming to be a world leader in innovative and efficient methods for data collection, analysis and dissemination.

2. Literature review

UAE is alarmed by the increasing number of divorcee each year (Bromfield, 2014). Although one can see that divorce rate in the U.A.E. is very low compared to international rate (0.9 divorce case per 1000 Emirati), 60% of these cases happened before completing three years of marriage in 2014 in Abu Dhabi Emirate (SCAD,2014). There are many reasons why divorce occurs. Mahir Faysal stated that sociological research has focused on structural and life course predictors of marital disruption, while psychological research has focused on dimensions of marital interaction or on personality characteristics, chronic negative affect (Faysal, 2008). In the study on hand, I will focus on demographic indicators that might lead to marriage dissolution.

3. Methodology

The main purpose of this research is to implement data mining techniques for classification and predicting in social science problem through administrative records. The benefit on this approach is that it will save time and money, benefit from available resources and data, and use intelligent statistical and computational techniques to get the knowledge that is not explicitly available. Although data mining is widely introduced in business, medical, educational and economic area, it is fairly used in sociology and official statistics. Upon my search on social science studies that applied data mining tools, I see that sociologist rarely use the ready tools and programs implemented for this field. In this paper, I will study the possibility to predict early divorce (before completing three years of marriage) through demographic indicators registered at time of marriage in Abu Dhabi Emirate. I am going to implement classification and predicting methods, which are logistic regression and decision trees to build some models based on demographic indicator -indicators that are available on administrative records at the time of marriage-, and then I will identify the most significant predictor variables –if any- from all available variables containing information about marriage that can predict divorce.

The reason why I choose demographic indicator only is that I do not have any other data available on marriage and divorce data sets. The idea is to be able to predict divorce without conducting surveys or qualitative analysis that is costly and available in one run only. Administrative records are available continuously without any additional cost. From the literature review, it seems that psychological, behaviour and demographic factors are all affecting each other's.

It is important to note that the aim of applying classification models for prediction is mainly to know what are the demographic variables that can be consider as divorce predictors, then to have a model that can predict divorce on record base. From this point, I come to choose what classification techniques I have to apply for the data. The main classification methods are Decision trees, Neural network, Logistic regression and Naïve Bayes. I decide to choose logistic regression and decision trees. The main reason why I choose them is they will explicitly show me what are the important variables that have significant contribution to the change in the output

variable. Both logistic regression and decision tree are interpreting methods. I can see the variables that has more influence to a change on the predictors. In logistic regression, I can examine it through the p-values and the odd ratio. On the other hand, from decision tree I can see what are the variables that come first to split the tree of the models. I will consider the first variables appear as the most important variables ignoring repetition. Decision tree results are easy to understand and interpret since readable rules can be easily derived from the models. One can see the growth of the tree in decision tree to see how the data splits to leaves and nodes and how variables interact with each other's. Furthermore, both techniques are based on statistical hypothesis. Decision trees use statistical methods to purify the model using pruning methods. In logistic regression I can test the model statistically with hypothesis testing and p – value . I cannot do it on neural network as example that lack classical statistical properties.

Before applying classification task, I have to go through many steps of pre-processing to reach my goal with good quality and accuracy. I have to prepare and clean the data, transform data and design derived variables, apply over sampling technique and select the variables to be included in the model, then I have to build the model with the classification methods. In building the model I have to add and remove the variables, as well as design new interaction variables and derived variables that reclassify the original variables on the dataset when needed until I reach the model that have the best performance in both training and validation data sets. After that, I will analyse and evaluate the model through data mining evaluation measurements.

I will separate the study in two sections. First, I will study the citizen population, where Emirati man married Emirati women. Then I will study The Emirati mixed marriage where Emirati men married Non-Emirati or Non-Emirati man married Emirati women. For each of them I will have one logistic regression model and one decision tree model. I would like to see if there is difference in the predictors of divorce among them or not.

The program used for sampling, feature selection, Partitioning datasets into training and validation datasets, building and evaluating the model is xlminer. Front-line solver develops this program. I use it because it can work under Excel add-in.

4. Results

4.1 Data collection

The examined data of this research is marriage contracts registered in judicial department of Abu Dhabi Emirate in 2011, which contain 5,708 contracts. To follow those cases to see if a divorce occurs during the first three years of marriage, I examined divorce data sets of the years 2011, 2012, 2013 and 2014 registered in Abu Dhabi Emirate. Although there were, 1,049 divorce contracts registered in divorce datasets from judicial department that got married in 2011, I managed to link 679 (65%) cases only between marriage and divorce datasets. I assume many reasons why I could not link all the 1,049 cases; the most important one is the missing of key variables that can link between them. In other word, there is no reference id that can be referred to the marriage contract as well as the absence of husband and wife names. The data received from judicial department is confidential; therefore, any identifiers are omitted from the dataset. The only way I could manage to link between them is through date of marriage contract as well as data of births of husband and wife. I assume that it is unlikely that three of them will be match to more than one contract (although it could happened mainly when the same couples got married and divorces several times as Islam permit). Therefore, the result is as table 1. I use Microsoft access to merge datasets.

4.2 Citizen Model

4.2.1. Logistic regression for Citizen Model

Table 2 present logistic regression output table for citizen model. There are five effects in the model: Husband number of wives, Abu Dhabi Region, Age difference, same education level and Husband education above secondary. The tests for parameters suggest that each of the effects in the model is significant at the .001 level (p-values < :001).

The equation of the logistic model for early divorce for citizens are as follow:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.75 + 1.06 \text{ Husband number of wives} + 0.59 \text{ Abu Dhabi} + 0.26 \text{ age difference} \\ - 0.18 \text{ Same education level} - 0.24 \text{ husband education above secondary}$$

Where \hat{p} the estimated probability of a couple divorced.

I can say that having more than one wife, living in Abu Dhabi region, and increasing the age gap between the couples increase the probability of getting divorce while sharing the same education level between the couples and attaining more than secondary education degree for the husband decrease the probability of getting divorce for Emirati males who got married to Emirati females. Moreover, I see that an increase in husband number of wives increase the odds of getting divorce almost 3 times, living in Abu Dhabi increase it two times, one year decrease in age deference also increase it 1.2 times while sharing the same education level and attaining more that secondary level for the husband decrease it almost one time.

Regarding the model evaluation, it was generally significant. Figure 1 shows that ROC curve and the Area under Curve (AUC) achieved 60%, value confirm the good predictive model. The result shows that the overall error rate is 41.8%, it is able to correctly classify 58% of divorce cases.

4.2.2. Decision tree for Citizen Model

I build a singletree classification and find that the top variables that appear to split the data were Wife age at marriage, Region, Husband number of wives, Husband age at marriage and Wife education as presented in table 3.

Figure 2 present the pruned three after refining the model with validation data set model with minimum errors and misclassification. The reason why I might not see all significant predictable variables on the below chart such as husband number of wives is that it did not work as a splitting variable. I can say that whenever a husband married extra wife he will face the risk of divorce.

We can extract from decision tree rules predictable knowledge. Age of both the husband and wife are the most influence predictor of divorce. If husband age at marriage is less than 20.5 years old or greater than 36.5 years old, couples are likely to divorce. The median age of marriage in Abu Dhabi Emirate for citizen males and females in 2014 are 26.6 and 23.9 years old respectively (Statistical year book of Abu Dhabi Emirate, 2015). Citizens are normally getting married in their 20th. Getting married relatively late for emirate women might lead her to sacrifice her demands. It is unlikely for men to remain unmarried until 36 years old, on the other hand if the husband got married very young are likely face the inference of the family and unlikely has his own income and might face communication issues with his wife. In addition, region seems the second important factor predicting divorce; nevertheless, it is mainly combined with the husband age at marriage. There were no explicit rule defined regarding the region. Education level for the wife on the other hand, shows that if the education level of the women is less than university-degree, it is

likely to stabilize the married, but if it is university or above she need to get married to a man whose age is more than 25 years old in order not get the risk of being divorce.

The model was able to predict well in the validation data thus confirming the accuracy of the model. Moreover, The graph of ROC curve and the Area under Curve (AUC) 72%, value confirm the very good predictive model” as figure 3 states. the result shows that the overall error rate is 33%.

4.2.3. Conclusion of citizen model

From the result above it is confirmed from all performance indicator, that Decision tree model work better than logistic regression model. Nevertheless, all the variables that was selected on logistic regression was highly significant and rather than decision tree I can see the effect of each variable separately on divorce decision, while decision tree show me these meaningful interaction between variables. I can say I accept both models and reveal that number of wives, region, education level, and age are the key demographic variables relevant to early divorce.

4.3 Mixed-Citizen Model

4.3.1. Logistic regression for Mixed-Citizen Model

Logistic regression model reveals that the previous experience of marriage from the husband as well as the wife increase the probability of getting divorce, the odd ratio is 2.55 and 1.62 respectively (P-value 0.000 and 0.003 respectively). Additionally, as the Citizen model reveals, having more than one wife at the moment of marriage increase the probability of marriage dissolution (odd ratio – 1.54 and p-value 0.011). Moreover, Marring to Arab nation person - who are not a GCC nation - increase the risk of divorce with odd ratio equal to 1.46 (p-value 0.008). The last variable that raise the risk of early divorce is when the husband got married late after reaching 35 years old (odd ratio 1.63 and p-value 0.021). On the other side, I see that an increment in wife age at marriage and husband number of children has stabilizing effect on marriage life for the first three years with an odd ratio 0.98 and 0.84 respectively and p-value 0.022 and 0.003 respectively. Moreover, having a kinship relation between the couples has also positive effect on remaining marriage life half time those who does not have relation between them. The model built has 32% overall error and the roc curve figure 4 shows 72% strength which make us accept the model.

The equation of the logistic regression for Mixed marriages is as the following:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.32 + 0.94 \text{ Husband married before} \\ + 0.49 \text{ husband married before and age above 35 years} + 0.48 \text{ Wife married before} \\ + 0.43 \text{ Husband number of wives} + 0.38 \text{ Arab marriage} - 0.02 \text{ Wife age at marriage} \\ - 0.1 \text{ husband number of children} - 0.6 \text{ Relative relation yes}$$

Where \hat{p} the estimated probability of a couple divorced.

4.3.2. Decision tree for Mixed-Citizen Model

Decision tree on the other hand, present that both couples marital status, age at marriage for the husband and wife, number of wives and children for the husband and having relative relation are the key demographic indicator predicting divorce on mixed marriage. Table 5 present Full-Grown Tree Rules for training dataset. Again as the citizen model state, Decision tree validation model present better performance than Logistic regression with an overall error rate equal to 28%, 70% precision and 76% area under the curve shown in figure 6.

4.3.3. Conclusion of mixed-citizen model

From the result above it is confirmed from all performance indicator, that Decision tree model work better than logistic regression model. I can say I accept both models and reveal that number of wives, previous marital status, age of the couples, education level, and previous marital status, relative relation, Arab marriage are the key demographic variables relevant to early divorce for mixed citizen marriages.

5. Limitation and further steps

The most obvious limitation is that I thought that if I collect marriage dataset from Abu Dhabi emirate followed by divorce datasets for the following years I can catch and follow all marriages and know if they resume on marriage or proceed divorce. This assumption was not valid. Almost half of divorce cases that were collected from the years 2011 to 2014 could not be linked to 2011 marriage data sets. A recommendation that can reduce this affect is to study divorce for the U.A.E. as a whole and not only Abu Dhabi Emirate with availability of identifiers such as names.

Regarding data processing and modelling, I have to reach to up to 50% of over sampling divorce cases in order to build a model with reasonable performance. Although this is an accepted ratio, it will change population composition a lot and that will affect the rules mentioned. I think there should be more advance techniques that allow you to stabilize population. Moreover, although logistic regression models show significant variables that can predict divorce, the adjusted R-square for all the models did not exceed 10%. I think one need to see how independence are the demographic variables from each other's. For example, number of wives' variable is correlated with number of children; also, education level is correlated to employment and so on. So further analysis for the variables dependency need to be investigated through testing collinearity diagnostics and covariance matrix of coefficient.

6. Conclusion

To conclude the paper, I can say that most of the demographic variables significantly affect early divorce in the first three years in Abu Dhabi Emirate, such as age at marriage, previously marriage experience, education level, number of children, number of wives for the husband and kinship. Employment indicators are absents in all models. Husband number of wives at time of marriage shown to be the highest variable affect divorce. Husband age at marriage also appear to be a highly predictable model. For more than one model, it is been shown that if the husband age at marriage is less than 20 years or more than 35 years, it is likely to dissolute marriage life. When comparing the predictors of both Citizen only marriage and Citizen mixed marriage population, I see that husband' number of wives strongly appear on all models indicating how crucial this predictor to divorce. Moreover, age appear on a way or another on the models. Furthermore, attaining education level higher than secondary stabilize marriage life. On the other side, Region predictor appear only on citizens model and was one of the strongest predictors with a favorable to less urbanized region for marriage stability. In addition relative relation shows that it decrease the risk of getting divorce for mixed marriage. GCC countries share many tribes and families between their countries.

Decision trees and logistic regression classification methods both succeed to predict early divorce. In addition, Decision trees consistently has better results than logistic regression method, which might be due to the natural of the problem at hand where demographic indicators interact with each other's. In this research, I approve that data mining techniques can be applied to predict and classify social problem from administrative records, which can be considered as state of the art.

7. References

- 1) Michael Gähler, 1998. Life After Divorce: Economic, Social and Psychological Well-Being Among Swedish Adults and Children Following Family Dissolution. Swedish Institute for Social Research Reports No 32.
- 2) Härkönen ,2013, Divorce: Trends, Patterns, Causes, Consequences. Stockholm University Linnaeus Center on Social Policy and Family Dynamics in Europe, SPaDE
- 3) Abu Dhabi Development Statistics 1960-2010, 2010, Statistics Centre – Abu Dhabi
- 4) Orna Cohen and Rivka Savaya Cohen, O, Savaya, R, 2003, Lifestyle differences in traditionalism and modernity and reasons for divorce among Muslim Palestinian citizens of Israel, Journal of Comparative Family Studies, 283-302
- 5) Nicole Bromfield, 2014. Interviews with divorced women from the United Arab Emirates: a rare glimpse into lived experiences. United Arab Emirates
- 6) Al Munajjed, M, 2010, Divorce in the Gulf Cooperation Council countries: Risks and implications, Dubai, United Arab Emirates: Ideation Center, Booz & Company.
- 7) Al Gharaibeh, F, Bromfield, N, 2012, An analysis of divorce cases in the United Arab Emirates: a rising trend, Journal of Divorce and Remarriage, 53 (6), 436-52
- 8) Marriage and Divorce Report ,2014, Statistics Center – Abu Dhabi
- 9) Mahir Faysal, 2008, Causes of increasing divorce rate: In the context of World & Bangladesh
- 10) Labour Force report, 2014, Statistics Center – Abu Dhabi
- 11) Getting married, Abu Dhabi government portal
- 12) Education System in Abu Dhabi, Abu Dhabi government portal
- 13) Gulf News Article, 2008, Facts and divorce proceedings: An analytical study, Sharjah, United Arab Emirates: Sharjah Supreme Council of Family Affairs.
- 14) Amy E. Rodrigues, Julie H. Hall, SUNY Buffalo, 2013, What predicts divorce and relationship dissolution?
- 15) Zvika Neeman, Andrew F. Newman, Claudia Olivetti, 2006, Are Working Women Good for Marriage?
- 16) Animasahun, 2015, Predictive Potentials of Some Psycho-Socio-Personal Variables on Divorce among Couples in Ibadan Nigeria
- 17) Scott M. Stanley, Howard J. Markman and Sarah W. Whitton, 2002, Communication, Conflict, and Commitment: Insights on the Foundations of Relationship Success from a National Survey
- 18) Becker G, Landes EM, Michael RT, 1977, An Economic Analysis of Marital Instability. Journal of Political Economy 85(6): 1141-87

- 19) Mădălina Cărbureanu, 2007, The Divorce Rate Prediction using Data Mining Techniques
- 20) LYNN K. WHITE, 1991, Divorce over the life course The Role of Marital Happiness, Journal of Family Issues
- 21) Torkild Hovde Lyngstad and Marika Jalovaar, 2010, A review of the antecedents of union dissolution
- 22) Booth, Alan, and John N. Edwards, 1985, Age at Marriage and Marital instability
- 23) Amato, Paul R., Booth, Alan , 1996, A prospective study of divorce and parent-child relationships. Journal of Marriage and Family. 58, (2), 356-365.
- 24) Thornton, A. and W.L. Rodgers. 1987, The Influence of Individual and Historical Time on Marital Dissolution, Demography 24:1-22.
- 25) TIM B. HEATON, 2002, Factors contributing to increasing stability in the United States
- 26) John Mordechai Gottman, 1994, What predict divorce, book
- 27) Dana Rotz, 2012, Why Have Divorce Rates Fallen? The Role of Women's Age at Marriage
- 28) Kurdek, L.A. (1993). Predicting marital dissolution: A 5-year prospective longitudinal study of newlywed couples. Journal of Personality and Social Psychology, 64, 221-242.
- 29) Abdul Jalil and Sibai, 2006, factor influencing the rise in divorce between the Saudi families
- 30) Bumpass LL, Sweet JA, Cherlin A, 1991, The role of cohabitation in declining rates of marriage. Journal of Marriage and Family. 53(4):913–927.
- 31) Groom, J., 2001, What Works in Premarital Counselling. Journal of Pastoral Counselling, 36, 46-64.
- 32) T. H. Lyngstad, Poortman, 2013, Labor and Love: Wives' Employment and Divorce Risk in its Socio-Political Context, Social Politics 1–28.
- 33) Shahzad Iqbal Sham, 2014, Some Aspects of Marriage and Divorce in Muslim Family Law
- 34) Abu Sanad, 2016, An Interview with Consultant in Family Guide Abu Dhabi
- 35) Stephane Tuffery, 2008, Data Mining and Statistics for Decision Making, book,
- 36) Samuel Daniel, 2014, Predictive modeling of trust to Social Media content
- 37) Shmueli G., Patel N. Bruce P., 2010, Data Mining for Business Intelligent
- 38) Saed Sayad, 2010, An Introduction to Data Mining
- 39) Frontline Solver guideline, 2011, V11.5
- 40) Razan Mohamed, 2016, Master thesis of EARLY DIVORCE PREDICTION IN ABU DHABI

8. Tables

Table 1: Divorce cases on marriage dataset by type of match

Category	Match between marriage contract date	Match in data of births of husband	Match in data of births of wife	Number of cases
1	*	*	*	327
2	*	*		139
3	*		*	142
4		*	*	71

Table 2: Logistic regression output – Citizen Model

Input Variables	Coefficient	Std. Error	P-Value	Odds	CI Lower	CI Upper
Intercept	-0.75	0.11	9.22E-12	0.47	0.38	0.59
Husband.no.wives	1.06	0.13	2.18E-17	2.90	2.27	3.71
abu dhabi Region	0.59	0.08	6.44E-13	1.80	1.53	2.11
age.difference	0.26	0.06	1.56E-05	1.29	1.15	1.45
same_Education_level	-0.18	0.08	2.80E-02	0.83	0.71	0.98
husband above secondary	-0.24	0.09	5.08E-03	0.79	0.67	0.93

Table 3: Decision tree Full-Grown Tree Rules (Using Training Data) - Citizen Model

NodeID	Level	Left Child	Right Child	SplitVar	SplitValue /Set	Cases	Misclassifications	Class	Node Type
0	0	1	2	wife.age.at.marriage.	29.5	2700	0.47	0	Decision
1	1	3	4	Region	1.5	2195	0.34	0	Decision
2	1	5	6	Husband.no.wives	0.5	505	0.06	1	Decision
3	2	7	8	husban.age.at.marriage.	20.5	1092	0.20	1	Decision
4	2	9	10	husban.age.at.marriage.	36.5	1103	0.14	0	Decision
5	2	11	12	Wife.education.coded	6.5	292	0.05	1	Decision
6	2	N/A	N/A	N/A	N/A	213	0.01	1	Terminal
7	3	N/A	N/A	N/A	N/A	108	0.01	1	Terminal
8	3	13	14	wife.age.at.marriage.	26.5	984	0.18	0	Decision
9	3	15	16	husban.age.at.marriage.	20.5	1051	0.12	0	Decision
10	3	N/A	N/A	N/A	N/A	52	0.01	1	Terminal
11	3	N/A	N/A	N/A	N/A	197	0.02	1	Terminal
12	3	N/A	N/A	N/A	N/A	95	0.01	0	Terminal
13	4	17	18	Wife.education.coded	6.5	811	0.14	0	Decision
14	4	N/A	N/A	N/A	N/A	173	0.02	1	Terminal
15	4	N/A	N/A	N/A	N/A	102	0.02	1	Terminal
16	4	19	20	Husband.education.coded	6.5	949	0.10	0	Decision
17	5	21	22	Wife.education.coded	5.5	533	0.08	0	Decision
18	5	23	24	husban.age.at.marriage.	25.5	278	0.05	1	Decision
19	5	25	26	husban.age.at.marriage.	24.5	757	0.09	0	Decision
20	5	N/A	N/A	N/A	N/A	192	0.01	0	Terminal

Table 4: Logistic regression output – Mixed-Citizen Model

Input Variables	Coefficient	Std. Error	P-Value	Odds	CI Lower	CI Upper
Intercept	-0.32	0.26	2.15E-01	0.73	0.44	1.2
husband_married_before	0.94	0.2	3.26E-06	2.55	1.72	3.79
Husband.married.before.and.age.over.35	0.49	0.21	2.07E-02	1.63	1.08	2.46
wife_married_before	0.48	0.17	3.88E-03	1.62	1.17	2.24
Husband.no.wives	0.43	0.17	1.15E-02	1.54	1.1	2.15
Arab_marriage	0.38	0.14	8.78E-03	1.46	1.1	1.93
wife.age.at.marriage.	-0.02	0.01	2.22E-02	0.98	0.96	1
husband.number.of.children	-0.1	0.04	3.24E-03	0.9	0.84	0.97
relative_relation_yes	-0.6	0.19	1.38E-03	0.55	0.38	0.79

Table 5: Decision tree Full-Grown Tree Rules (Using Training Data)-Citizen Mixed Model

NodeID	Level	ParentID	LeftChild	RightChild	SplitVar	SplitValue/Setting	Cases	Misclassifications	Class	NodeType
0	0	N/A	1	2	Husband.marital.status	1.5	1200	0.443333	0	Decision
1	1	0	3	4	wife.age.at.marriage	31.5	639	0.155	0	Decision
2	1	0	5	6	husband.number.of.children	8.5	561	0.179167	1	Decision
3	2	1	7	8	wife.age.at.marriage	30.5	598	0.153333	0	Decision
4	2	1	N/A	N/A	N/A	N/A	41	0.001667	0	Terminal
5	2	2	9	10	Husband.marital.status	2.5	527	0.158333	1	Decision
6	2	2	N/A	N/A	N/A	N/A	34	0.0075	0	Terminal
7	3	3	11	12	wife.marital.status	2	568	0.136667	0	Decision
8	3	3	N/A	N/A	N/A	N/A	30	0.008333	1	Terminal
9	3	5	13	14	Husband.married.before.and.age.over.35	0.5	356	0.0925	1	Decision
10	3	5	15	16	wife.marital.status	2	171	0.065833	1	Decision
11	4	7	17	18	relative_relation_yes	0.5	532	0.121667	0	Decision
12	4	7	N/A	N/A	N/A	N/A	36	0.015	1	Terminal
13	4	9	19	20	wife.age.at.marriage.	21.5	164	0.049167	1	Decision
14	4	9	21	22	Husband.no.wives	1.5	192	0.043333	1	Decision
15	4	10	N/A	N/A	N/A	N/A	79	0.021667	0	Terminal
16	4	10	N/A	N/A	N/A	N/A	92	0.021667	1	Terminal
17	5	11	23	24	wife.age.at.marriage.	16.5	351	0.09	0	Decision
18	5	11	25	26	wife.age.at.marriage	17.5	181	0.031667	0	Decision
19	5	13	N/A	N/A	N/A	N/A	43	0.005	1	Terminal
20	5	13	27	28	relative_relation_yes	0.5	121	0.044167	1	Decision
21	5	14	29	30	wife.age.at.marriage	20.5	177	0.035833	1	Decision
22	5	14	N/A	N/A	N/A	N/A	15	0.005	0	Terminal
23	6	17	N/A	N/A	N/A	N/A	10	0.003333	1	Terminal
24	6	17	31	32	wife.age.at.marriage.	20.5	341	0.085	0	Decision
25	6	18	N/A	N/A	N/A	N/A	22	0.006667	0	Terminal
26	6	18	33	34	wife.age.at.marriage.	18.5	159	0.025	0	Decision
27	6	20	N/A	N/A	N/A	N/A	118	0.041667	1	Terminal
28	6	20	N/A	N/A	N/A	N/A	3	0	0	Terminal
29	6	21	N/A	N/A	N/A	N/A	10	0.003333	0	Terminal
30	6	21	35	36	Husband.no.wives	0.5	167	0.030833	1	Decision
31	7	24	N/A	N/A	N/A	N/A	84	0.015	0	Terminal
32	7	24	N/A	N/A	N/A	N/A	257	0.07	0	Terminal
33	7	26	N/A	N/A	N/A	N/A	12	0	0	Terminal
34	7	26	N/A	N/A	N/A	N/A	147	0.025	0	Terminal
35	7	30	N/A	N/A	N/A	N/A	1	0	0	Terminal
36	7	30	N/A	N/A	N/A	N/A	166	0.03	1	Terminal

9. Figures

Figure 1: Roc curve training dataset and validation dataset – Logistic regression – Citizen model

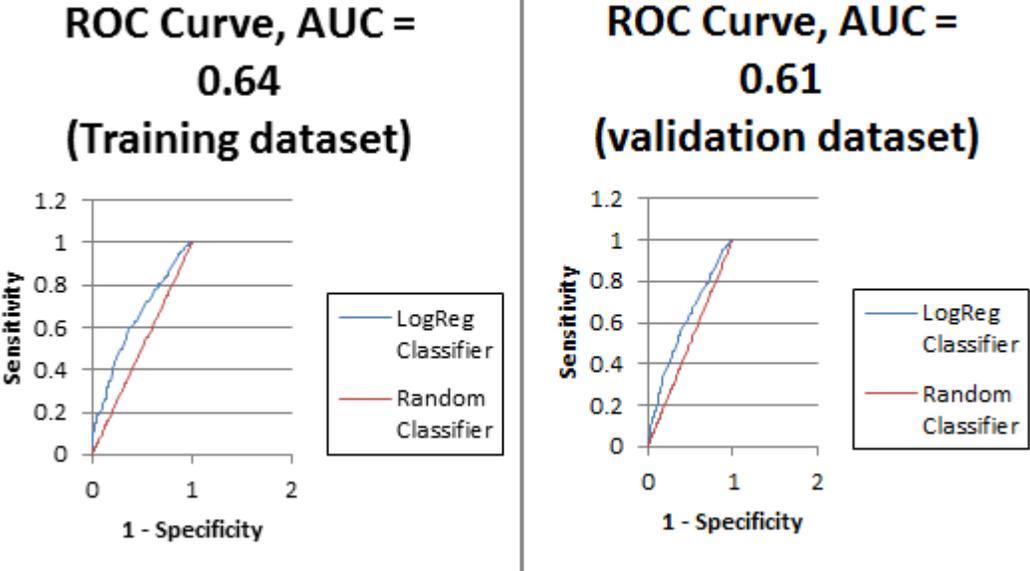


Figure 2: Decision tree Pruned chart – Citizen model

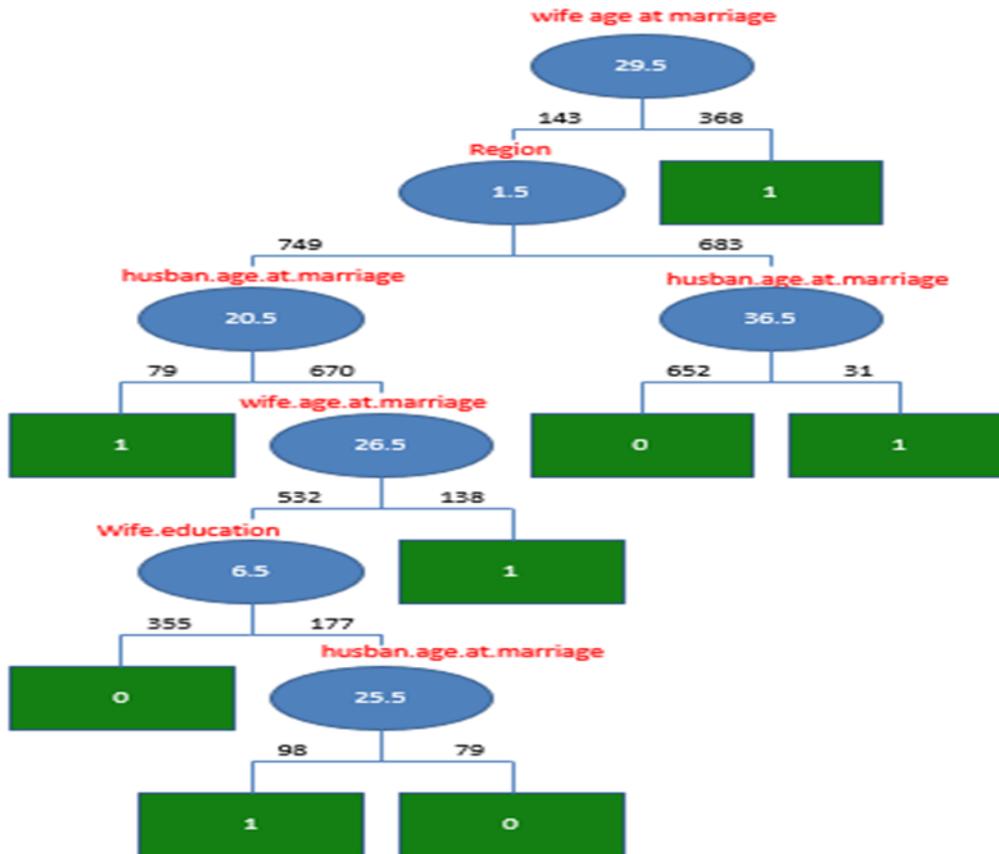


Figure 3: Roc curve training dataset and validation dataset – Decision tree – Citizen model

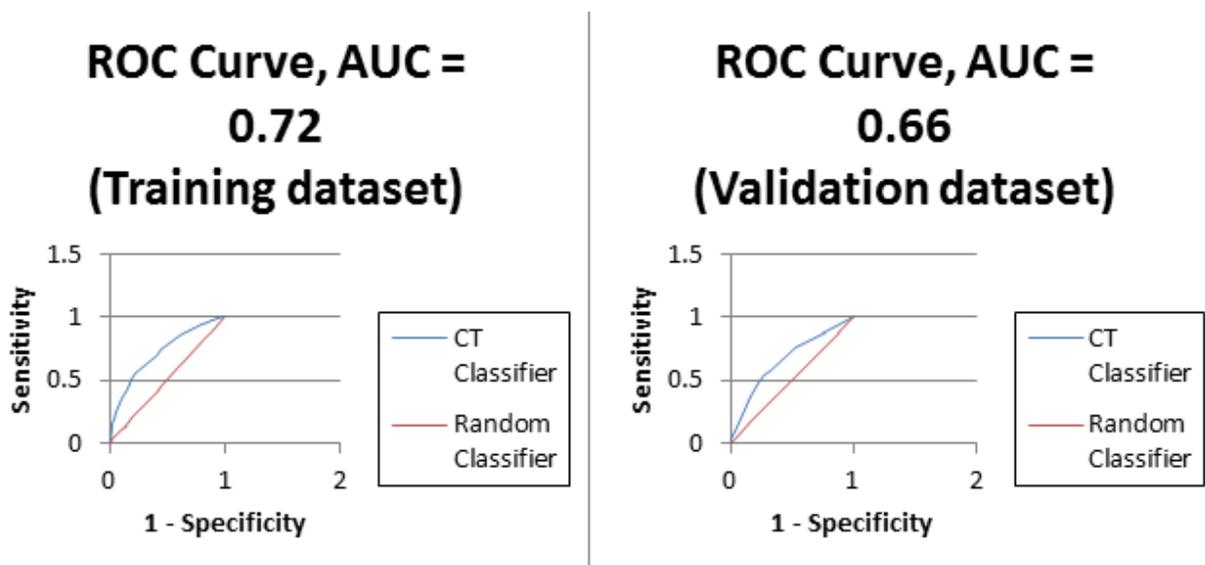


Figure 4: Roc curve training dataset and validation dataset – Logistic regression – Mixed-Citizen model

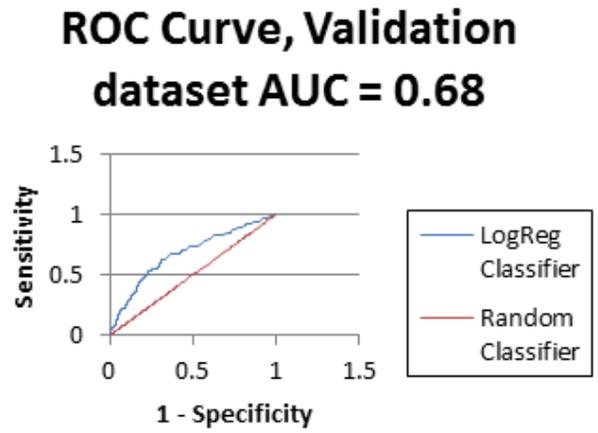
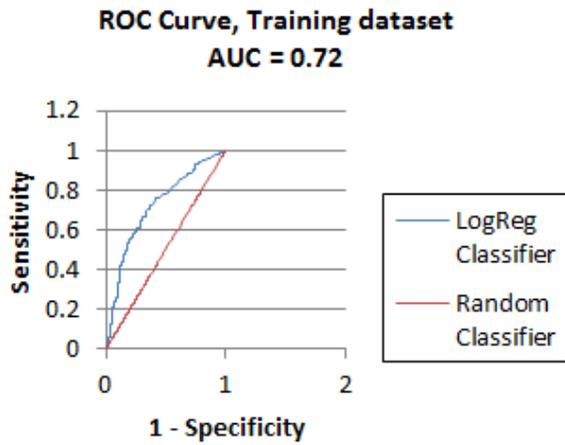


Figure 5: Decision tree Pruned chart – Mixed-Citizen model

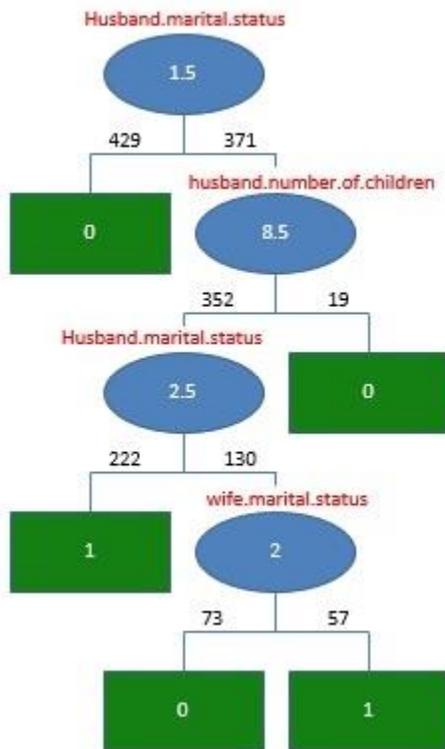


Figure 6: Roc curve training dataset and validation dataset – Decision tree – Mixed-Citizen model

