iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

A Statistical Study of Gender Identification

R. H. Waliv[1]
[1]*Asst. Professor Kisan Veer Mahavidyalaya, Wai (Maharashtra)*

Abstract: This paper is attempted to identify gender of person using different measurement of body. This has importance in forensic science. Sample of 400 persons is collected to study problem. Data is analysed by using data mining and regression techniques.

Keyword: Gender, Data mining, Regression

I. INTRODUCTION

Gender identity is defined as a personal conception of oneself if as a male or female. This concept is intimately related to the concept of gender role, which is defined as outward manifestation of personality that reflects the gender identity. Gender identity is one's personal experience of once own gender. Gender is our social & legal status as a girls and boys, women & men.

Gender identification is very important thing in forensic. There are many techniques in today's world to identify the gender. Such as from various prints like fingerprints, footprints we can identify the gender Biologically from pelvic girdle or from hip bone we can identify that the individual is either male or female but in some cases whenever these parts are not available or whenever we cannot use this techniques then from available body measurements like shoulder, Thumb, middle finger, nose, waist, height, weight etc. We can identify the gender which is of great importance in forensic.

II. OBJECTIVE

The main objective of the project is to identify the gender from body measurements.

III. VARIABLES OF THE STUDY

Dependent Variable: nose, ear, forehead, lip, chin, throat, shoulder, fingers, height, weight, waist, leg
Independent Variable: Gender

IV. TOOLS USED

Study of gender identification can be done by using a technique of data mining. Data mining is about explaining the past & predicting the future by means of data analysis. Data mining is a multi-disciplinary field. Classification is a data mining task of predicting the value of a categorical variable by building the model based on one or more numerical and or categorical variables.

A. Zero R
B. One R
C. Naive Bayesian
D. Logistic Regression

The sample is collected on 400 individuals of which 200 are males and 200 are females. The data analysed using R software.
Statistical analysis of data:

E. Naive Bayesian Classifier:

The Naive Bayesian classifier is based on Bayes' theorem with 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.

Confusion Matrix		Gender			
		Male	Female		
Naïve Bayesian	Male	54	14	Positive Predictive Value	0.794118
	Female	8	44	Negative Predictive Value	0.846154
		Sensitivity	Specificity	Accuracy	
		0.870968	0.758621		0.816666

©IJRASET (UGC Approved Journal): All Rights are Reserved

F. Logistic Regression Classifier

Logistic regression predicts the probability of an outcome that can only have two values (i.e. a dichotomy). The prediction is based on the use of one or several predictors (numerical and categorical). To perform the logistic regression I use binomial model with logic link for classifying the response variable.

Confusion Matrix		Gender			
		Male	Female		
Logistic Regression	Male	62	6	Positive Predictive Value	0.911765
	Female	3	49	Negative Predictive Value	0.942308
		Sensitivity	Specificity	Accuracy	0.941666
		0.953846	0.890909		

G. One-R Classifier

One-R, 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 One-R produces rules only slightly less accurate than state-of-the art classification algorithms while producing rules that are simple for humans to interpret.

Confusion Matrix		Gender			
		Male	Female		
One-R	Male	68	0	Positive Predictive Value	1
	Female	52	0	Negative Predictive Value	0
		Sensitivity	Specificity	Accuracy	0.566666
		0.566667	0		

H. Zero-R Classifier

Zero-R is the simplest classification method which relies on the target and ignores all predictors. Zero-R classifier simply predicts the majority category(class). Although there is no predictability power in Zero-R, it is useful for determining a baseline performance as a benchmark for other classification methods.

Confusion Matrix		Gender			
		Male	Female		
Zero-R	Male	0	68	Positive Predictive Value	0
	Female	0	52	Negative Predictive Value	1
		Sensitivity	Specificity	Accuracy	0.433333
		0	0. 433333		

I. Binary Logistic Regression Model

Here independent variable has two categories. Categories are male and female. Here independent variable is binary so for prediction of gender use of logistic regression is important. The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest and a set of independent variables. Logistic regression generates the coefficients of a formula to predict a logit transformation of the probability of presence of the characteristic of interest.

1) Model: Logit (p) = $b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$

Where, p is the probability of presence of the characteristic of interest.

R Output:

Call

glm(formula = Gender ~ Shoulder + Waist + Leg + F3 + Height +

©IJRASET (UGC Approved Journal): All Rights are Reserved

```
Weight, family = binomial (link = logit), data = Datag1)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6646 -0.3842   0.0194   0.3182   3.1723
Coefficients:
 Estimate Std. Error z valuePr(>|z|)
(Intercept)  -42.02856   5.39431  -7.791  6.63e-15 ***
Shoulder      0.30980    0.06384   4.853   1.22e-06 ***
Waist        -0.06793    0.02980   -2.280  0.02262 *
Leg           0.74857    0.17711    4.227     2.37e-05 ***
F3           -0.95358    0.32580  -2.927     0.00342 **
Height        0.12463    0.03072  4.057     4.96e-05 ***
Weight        0.09866    0.03170   3.112     0.00186 **
Signif.codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 554.52  on 399  degrees of freedom
Residual deviance: 214.04  on 393  degrees of freedom
AIC: 228.04
Number of Fisher Scoring iterations: 7
```

Model Log it (p) = -42.02856 + 0.30980(Shoulder) + (-0.06793)(Waist) + 0.74857(Leg) + (-0.95358)(F3) + 0.12463(Height) +0.09866(Weight)Above model gives probability of person belonging to female category. If p=0.95 means person is female.

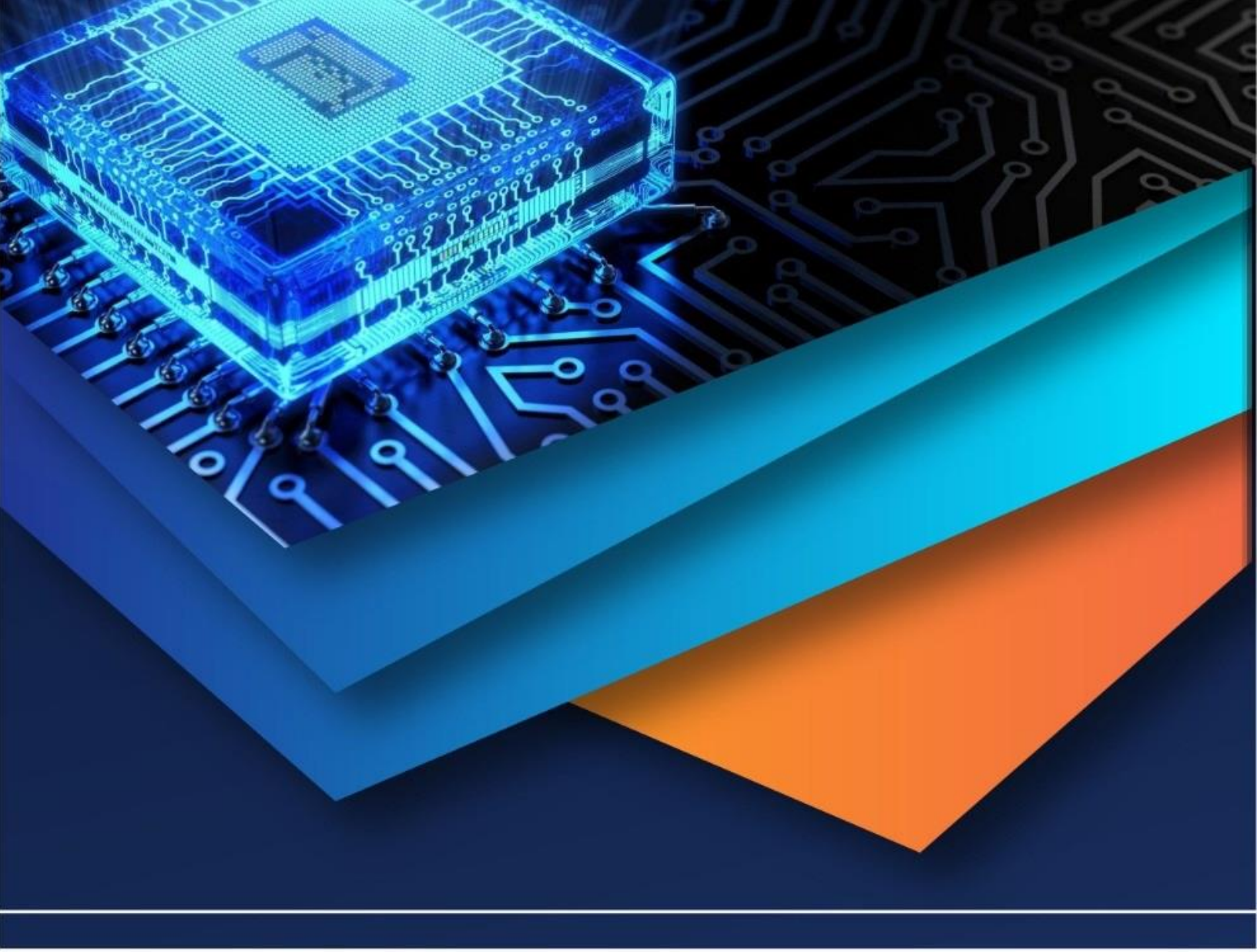
V. CONCLUSION

From above analysis we can see that, Logistic regression classifier classifies gender more accurately than zero-R, one-R and Naive bayes classifiers. So it preferable to use logistic regression classifier to classify the gender.

REFERENCES

[1] Evaluation of Evidence for Forensic scientist John Wiley and sons
[2] Han Y., Ryu C., Moon J., Kim H., Choi H. A Study on Evaluating the Uniqueness of Fingerprints Using Statistical Analysis. Information Security and Cryptology–ICISC; 2004-05: 467–477.
[3] Hsieh C.T., Shyu S.R., Hu C.S. An effective method of fingerprint classification combined with AFIS. Embedded and Ubiquitous Computing; 2005: 1107–1122.
[4] Nithin M.D., Balaraj B.M., Manjunatha B. &Mestri S.C. Study of Fingerprint Classification and their Gender Distribution among South Indian Population. 2009, J. Forensic. Leg. Med. 16(8): 460-463.
[5] Nandy A. Identification. Principles of Forensic Medicine 2nd Edition. New Central Book Agency (P) LTD. Calcutta. 2001:47-109.
[6] Subrahmanyam BV. Personal Identity. Modi’s Medical Jurisprudence and Toxicology 22nd Edition. Butterworths, New Delhi. 2001:37-90
[7] Lalit Kumar and et al. Gender Determination Using Fingerprints In the Region of Uttarakhand 2013,J Indian Acad Forensic Med. October-December 35(4), 308-311

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)