

Comparison of Classification rules for Predicting Personality

Mixymol V.K.,

Assistant Professor, Department of Computer Science,
Kanchi Mamunivar Centre for Post Graduate Studies, P
uducherry, Lawspet- 605 008.

E-mail : mail2mixymol@yahoo.co.in

Abstract: Personality refers to an individual's characteristic patterns of thought, emotion and behaviour, together with the psychological mechanisms. It has been a long-term goal for psychologists to understand human personality and its impact on human behaviour. Behavior involves an interaction between a person's underlying personality traits and situational variables. Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. It is evident that there is a strong correlation between users' personality and the way they behave on online social network and a large number of studies are made on this topic.

Objective of this paper is to classify the personality traits into Extraversion and Introversion. Dataset for this purpose is collected from 'http://personality-testing.info/_rawdata/' which stores data on personality surveys for the purposes of academic research and personal use. The dataset consists of 140 instances with 17 attributes. Three types of classification rules like J48 Decision Tree classifier, Naive Bayes Classifier and ZeroR Rules available on Weka data mining tool are used to predict Personality traits. Performances of each of these rules are measured by the values of the Precision, Recall, and Accuracy. Comparative study of these methods shows that J48 Decision tree Classifier performs better than Naive Bayes classifier and ZeroR Rule for personality traits recognition on the Social Media.

Keywords- Personality trait, Social Media, Data Mining, Classification.

I. Introduction

Social Media is a place collecting large volumes of data related to user. The availability of Social Media data has been growing fast, and there is a need to analyze huge amounts of data generated from Social Media [1]. People are becoming more interested in and relying on social network for information, news and opinion of other users on diverse subject matters. Personality is the most complex of all the human attributes and it also characterizes the uniqueness of a person. The situation, that a person finds himself or herself in, plays a major role on his or her reaction. However, in most of the cases, people respond with respect to their underlying personality traits [2].

The "Big Five" model of personality dimensions has emerged as one of the well-researched measures of personality structure in recent years. This model describes a personality structure which is divided into five components known as "OCEAN": Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism[3]. It describes the human personality as a vector of five values corresponding to bipolar traits. This is a popular model among the language and computer science researchers and it has been used as a framework for both personality traits identification and simulations.

II. Problem Statement

People spend a considerable amount of effort in order to manage impressions, especially in the initial stage of social interactions. Nowadays, this fundamental process has been modified by the usage of new communication technologies. Moreover, social psychology research has also highlighted that personality plays an important role in the way people manage day to day situation they conveying self-presentations. Personality has been shown to be relevant to many types of interactions.[4] It has been shown to be useful in predicting job satisfaction, professional and relationship

success, and even preference for different interfaces. Personality is one of the fundamental aspects, by which we can understand behavioral dispositions. It is evident that there is a strong correlation between users' personality and the way they behave on online social network. a relationship among the personalities of users communicating within a social network, and analysing how personality affects the way in which each individual communicates with others in such a context[5.]The objective of this paper is to study various classification rules like J48 Decision Tree classifier, Naïve Baye's Classifier and ZeroR rule to predict a Personality trait as Extraversion and Introversion by collecting a suitable dataset[6],[7]. By using these algorithms various measures like Precision, Recall, Accuracy are calculated.

III. Methodology

Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It comprises tools for regression, classification, clustering, association rules, visualization, and data pre-processing. Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data.

IV. Dataset Description

Dataset for this study is collected from 'http://personality-testing.info/_raw data', which stores data on personality surveys for the purposes of academic research and personal use. The dataset is in the form of questionnaire and samples of 10 questions are used to predict the

personality. Dataset consists of 140 instances with 17 attributes. The attributes referring to the Person's characteristics include personal details and personality test details. The attributes includes basic information about Country, Hand used, Job Category, English Native and Source by which gives the user response like, Twitter, Face book, Website. The answers to the questions are in a gradient form, ranging from Agree, Disagree and Neutral.

Table 1 gives the details of the dataset.

I am the Life of the Party	{ Agree, Disagree, Neutral }
I get stressed out easily.	{ Agree, Disagree, Neutral }
I feel little concern for others.	{ Agree, Disagree, Neutral }
I have a rich vocabulary.	{ Agree, Disagree, Neutral }
I am always prepared.	{ Agree, Disagree, Neutral }
I don't talk a lot.	{ Agree, Disagree, Neutral }
I am relaxed most of the time.	{ Agree, Disagree, Neutral }
I am interested in people.	{ Agree, Disagree, Neutral }
I have difficulty understanding abstract ideas.	{ Agree, Disagree, Neutral }
I leave my belongings around.	{ Agree, Disagree, Neutral }
Type of Source	{ Twitter, Facebook, Website }

V. Classification Algorithm Used

J48 Decision tree classifier

Classification is the process of building a model of classes from a set of instances that contain class labels. Decision Tree Algorithm is to find out the way the attributes-vector behaves for a number of instances. Also on the bases of the training instances the classes for the newly generated instances are being found. This algorithm generates the rules for the prediction of the target variable. With the help of tree classification algorithm the critical distribution of the data is easily understandable. These algorithms analyse the training set and build a classifier that must be able to correctly classify both training and test examples. The basic algorithm for Decision Tree induction is a Greedy Algorithm that constructs Decision Trees in a top-down recursive Divide-and-Conquer manner.

ZeroR Rule

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

Naïve Bayes Classifier

The Naive Bayes algorithm is based on conditional probabilities. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data. Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. 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.

VI. Result Analysis

Dataset is trained on J48 Decision tree classifier which is available on Weka it provides the following information

Correctly Classified Instances	131
Incorrectly Classified Instances	9
Kappa statistic	0.8594
Mean absolute error	0.0608
Root mean squared error	0.2369
Relative absolute error	13.4717 %
Root relative squared error	49.8872 %
True Positive Rate	0.936
False Positive Rate	0.063
Precision	0.938
Recall	0.936
F-Measure	0.936
Accuracy	93.5714 %

While the dataset is trained on Naïve Bayes Classifier the following values are obtained.

Correctly Classified Instances	120
Incorrectly Classified Instances	20
Kappa statistic	0.6732
Mean absolute error	0.1994
Root mean squared error	0.3391
Relative absolute error	44.1605 %
Root relative squared error	71.4123 %
True Positive Rate	0.857
False Positive Rate	0.204
Precision	0.856
Recall	0.857
F-Measure	0.855
Accuracy	85.7143 %

When the dataset is trained on ZeroR Rule the following values are obtained.

Correctly Classified Instances	92
Incorrectly Classified Instances	48
Kappa statistic	0
Mean absolute error	0.4516

Root mean squared error	0.4749
Relative absolute error	100 %
Root relative squared error	100 %
True Positive Rate	0.657
False Positive Rate	0.657
Precision	0.432
Recall	0.657
F-Measure	0.521
Accuracy	65.7143 %

The following table2 represents the comparison of different Algorithms with Precision, Recall, Accuracy Values.

Table2: Precision , Recall, Accuracy Values using Various Methods

Algorithm	Precision	Recall	Accuracy
J48 Decision tree classifier	93.8	93.6	93.57
Naive Bayes Classifier	85.6	85.7	85.71
ZeroR Rule	43.2	65.7	65.71

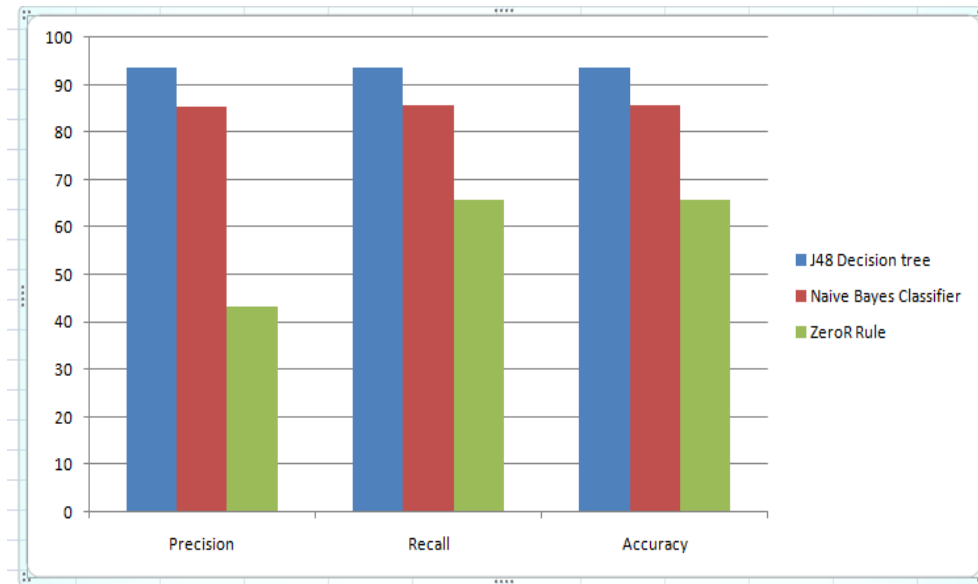


Figure1: Comparison chart for the three Algorithms namely to measure Precision, Recall and Accuracy value.

Figure 1 gives the pictorial representation of the these algorithm analysis. The comparative study shows that J48 Decision tree classifier (Precision=0.938, Recall=0.936, Accuracy=93.57%) performs better than Naive Bayes (Precision=0.856, Recall=0.857, Accuracy=85.71%) and ZeroR Rule (Precision=0.432, Recall=0.657, Accuracy=65.71%) for personality traits recognition on the social Media.

Conclusion

Personality plays a fundamental role in human interaction. Personality is one of the fundamental aspects, by which we can understand behavioral dispositions. Personality is the most complex of all the human attributes and it also characterizes the uniqueness of a person. Dataset is collected from 'http://personality-testing.info/_rawdata/' which stores data on personality surveys for the purposes of academic research and personal use is used to predict personality using various classification rules, J48 decision tree classifier, Naive Bayes Classifier and ZeroR. Comparative study shows that J48 Decision tree classifier performs better than Naive Bayes classifier and ZeroR Rule for personality traits recognition on the social Media.

References

[1] Vitaly Levashenko, Elena Zaitseva, Krzysztof Pancierz and Jerzy Gomula "Fuzzy Decision Tree Based Classification of

Psychometric Data" Federated Conference on Computer Science and Information Systems (2014).

[2] Firoj Alam, Evgeny A. Stepanov, Giuseppe Riccardi "Personality Traits Recognition on Social Network – Face book", AAAI Technical Report, WS 13-01.
 [3] S. Rotthmaan, E.P., Coetaer, "The Big Five Personality Dimensions and Job Performance", Journal of Industrial Psychology, 29(1), 2003
 [4] Basant Agarwal, " Personality Detection from Text: A Review", International Journal of Computer Systems, Volume 1, Issue 1, September 2014
 [5] Drs. R. Sudrajat, M.Si., Rudi Rosadi, S.Si., M.Kom, and Harits Muhammad "Implementation of Data Mining in Analyzing Social Media Users Personality with Naive Bayes Classifier: A Case Study of Instagram Social Media" (July 2016)
 [6] Syed Khalid Perwez , HamzaMohd. Zubair , Muhammad Rukunuddin Ghalib, KausarAhmedP, Mohammed Itekhara " Association Rule Mining Technique for Psychometric Personality Testing and Behaviour Prediction" International Journal of Engineering and Technology (IJET) ,Nov 2013
 [7] Celli, F.; Pianesi, F.; Stillwell, D.S.; and Kosinski, M. 2013 WCPR (Shared Task). The 7th International AAAI Conference On Weblogs And Social Media. Boston.