International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Personality Prediction using Machine Learning

Maithili Kulkarni, Isha Kulkarni, Pratiksha Shelar, Vikrant Patil

MMCOE,Pune

ABSTRACT

The corporate world today does not focus just on the skills a potential employee possesses but also their personality. Personality is what helps one be successful in professional as well as personal life. Hence, the recruiter must be aware of the personality traits a person has. With an exponential increase in job seekers but a decrease in the number of jobs, it is difficult to manually shortlist the best fit candidate for a suitable job by looking at the CV.

Introduction

Identification of a person's personality based on their nature is an old practise. Previously, these were done manually, requiring a significant amount of time to estimate the person's personality. Companies with a strong consumer emphasis, such as retail, financial, communication, and marketing, are the most common users of data mining today.Surveys, interviews, questionnaires, classroom activities, shopping website data, and social network data on user experiences and difficulties were all used to analyse the data. Traditional approaches, on the other hand, are time-consuming and scale-limited. Our proposed technology will reveal details on the user's personality.The practise of determining a person's personality based on their nature is not new.

Personality can also influence how he or she interacts with the outside world and his or her surroundings. Personality can also be used as a differentiator during the hiring process, career counselling, and health counselling, among other things. Analysing a person's behaviour to predict their personality is an old practise. This manual approach of personality prediction took a long time and a lot of effort.

Analysing personality based on one's nature was a time-consuming activity that required a great deal of human work. Furthermore, when analysing a user's personality based on their nature and behaviour, this manual analysis did not produce reliable findings. Because the analysis was done by hand, it has an impact on the correctness of the results because humans are prone to prejudice and see things in a certain way.

Literature Survey

This project provides a number of new research avenues in the area of Automated Personality Classification (APC). To begin, we look into ways to improve existing APC solutions, which we do by combining APC corpora, psychological trait measures, and learning algorithms. Following that, we look into APC extensions and related tasks, such as dynamical APC and detecting personality discrepancy in a text.

The entire study was conducted in the setting of social networks and the data mining methods that go along with them.

One of the issues addressed by personality psychology, a discipline of psychology, is personality classification. The study of personality and individual characteristics is the emphasis of this field. Personality, according to that study, can be defined as a dynamic and ordered set of a person's qualities that have a unique influence on that person's cognition, motivation, and behaviour. The subject of automated personality classification is explored in this research based on information from the following material: textual content written by the person and meta information about a person obtained on request, via social networks or other ways. There are studies that involve speech, facial analysis, gestures, and other aspects of behaviour, but these are not the subjects of our investigation. The following steps describe the conventional technique to fixing the APC problem using the aforementioned content: A) Collecting data from the corpus. B) Determining the individuals' personality characteristics, and C) constructing the model

Personality can also influence how he or she interacts with the outside world and his or her surroundings. Personality can also be used as a differentiator during the hiring process, career counselling, and health counselling, among other things. Analysing a person's behaviour to predict their personality is an old practise. This manual approach of personality prediction took a long time and a lot of effort.

Analysing personality based on one's nature was a time-consuming activity that required a great deal of human work. Furthermore, when analysing a user's personality based on their nature and behaviour, this manual analysis did not produce reliable findings. Because the analysis was done by hand, it has an impact on the correctness of the results because humans are prone to prejudice and see things in a certain way.

Some studies applied supervised machine learning approaches for the classification of psychopathy. For example, Kishtwar et al. [14] exploited different machine learning algorithms, namely, (i) SVM, (ii) NB, and (iii) DT, where the main emphasis of the proposed system was to automatically identify the personality traits of students in an educational game. The experimental results showed that the n-gram gave the best performance compared with the other feature sets. dimensionally. If you must use mixed units, Equations

- System Architecture

System Architecture   (1)
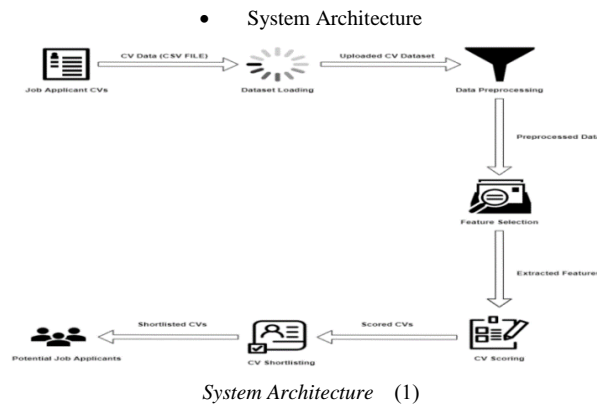
System Architecture

The above diagram shoes the system architecture. It demonstrates the flow of the program, Job applicants CV's are upload in the system with will get parsed and then the important points in them will be collected. These Keywords will be checked with the dataset uploaded. The data uploaded will be parsed and is checked for missing values or any outliers and if found any they will be fixed. After the dataset is cleaned we will select the features for our prediction.

Algorithmic strategy.

A.**Decision Trees** : are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split. which are outcomes like either 'fit', or 'unfit'. In this case this was a binary classification problem (a yes no type problem). There are two main types of Decision Trees:
We'll build a decision tree to do that using ID3 algorithm. ID3 Algorithm will perform following tasks recursively

1. Create root node for the tree
2. If all examples are positive, return leaf node 'positive'
3. Else if all examples are negative, return leaf node 'negative'
4. Calculate the entropy of current state H(S)
5. For each attribute, calculate the entropy with respect to the attribute 'x' denoted by H(S, x)
6. Select the attribute which has maximum value of IG(S, x)
7. Remove the attribute that offers highest IG from the set of attributes
8. Repeat until we run out of all attributes, or the decision tree has all leaf nodes.

B. Logistic Regreesion.

Logistic regression models the probability of the default class (e.g. the first class). For example, if we are modeling people's sex as male or female from their height, then the first class could be MMCOE, Department of Computer Engineering 2021-22 38 male and the logistic regression model could be written as the probability of male given a person's height, or more formally: P(sex=male height) Written another way, we are modelling the probability that an input (X) belongs to the default class (Y=1), we can write this formally as: $P(X) = P(Y=1|X)$ We're predicting probabilities. I thought logistic regression was a classification algorithm. Note that the probability prediction must be transformed into a binary values (0 or 1) in order to actually make a probability prediction. More on this later when we talk about making prediction. Logistic regression is a linear method, but the predictions are transformed using the logistic function. continuing on from above, the model can be stated as: $p(X) = e^{(b0 + b1*X)} / (1 + e^{(b0 + b1*X)})$ I don't want to dive into the math too much, but we can turn around the above equation as follows (remember we can remove the e from one side by adding a natural logarithm (ln) to the other): $\ln(p(X) / 1 - p(X)) = b0 + b1 * X$ This is useful because we can see that the calculation of the output on the right is linear again (just like linear regression), and the input on the left is a log of the probability of the default class. This ratio on the left is called the odds of the default class (it's historical that we use odds, for example, odds are used in horse racing rather than probabilities). Odds are calculated as a ratio of the probability of the event divided by the probability of not the event, e.g. 0.8/(1-0.8) which has the odds of 4. So we could instead write: $\ln(odds) = b0 + b1 * X$ Because the odds are log transformed, we call this left hand side the log-odds or the profit. It is possible to use other types of functions for the transform (which is out of scope_, but as such it is common to refer to the transform that relates the linear regression equation to the probabilities as the link function, e.g. the profit link function. We can move the exponent back to the right and write it as: $odds = e^{(b0 + b1 * X)}$.

Conclusion and Future work

This project, we discuss about how the personality is identified using different classification algorithms. Here we study relationship between user and his/her personality. In this we used Decision Trees because it gives best accuracy around 86.53% while compare to other algorithms that are used previously like naive Bayes , SVM , Logistic regression is fast and give accurate results compared to other algorithms.

Personality system is used in E-commerce sites, in Competitive exams , Psychometric tests , matrimonial sites , Government sectors like army, navy, Air force . Thus the personality is automatically classified by the system after user attempts the survey by the data set provided in the back end . Personality analysis and prediction is more in recent times so further in future more personality traits can be added. Further any improvement can be done using the data set and algorithms to improve the accuracy and can be helpful for career guidance module, if user has good speaking and convincing skills. This project discuss about personalitAnalysis and Prediction.

Output Screenshots

References

[1] Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C. Graesser,Using Data Mining Techniques to Detect the Personality of Players in an Educational Game

2] R. Wald,T. M. Khoshgoftaar,A. Napolitano Using Twitter Content to Predict Psychopathy

[3]YagoSaez , Carlos Navarro , Asuncion Mochon and Pedro Isasi, A system for personality and happiness detection.

[5] Golbeck, J., Robles, C., and Turner, K. 2011a. Predicting Personality with Social Media. In Proc of the 2011 annual conference extended abstracts on Human factors in computing systems.

[6] DURGESH K.SRIVASTAVA, LEKHA BHAMBHU, "DATA Classification using Support Vector Machine," Journal of Theoretical and Applied Information Technology

[7] YILUN WANG, "Understanding Personality through social media," International of computer Science stand ford University.

[8] . MAHESH KINI, SAROJA DEVI, PRASHANT G DESAI, NIRANJAN CHIPLUNKAR,"Text Mining approach to classify Technical Research Document using naive Bayes", International journal of Advanced Research in computer and communication engineering.

[9]. Cantandir, I. Fernandez-Tobia] z, A. Belllogin, "Relating personality types with user preferences in multiple entertainment domains," EMPIRE 1st Workshop on Emotions and Personality in Personalized Services, 2013.

[10]. P.T. Costa, R.R. McCrae(1992).Revised NEO personality inventory (NEO-PIR) and NEO five-factor inventory (NEOFFI),Psychological Assessment Resource. IEEE conference templates