

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Winter 11-7-2019

Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis

Gagandeep Kaur

SIT, gagandeep.kaur@sitpune.edu.in

Shruti Maheshwari

SIT, shruti.maheshwari@sitpune.edu.in

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#), and the [Other Computer Engineering Commons](#)

Kaur, Gagandeep and Maheshwari, Shruti, "Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis" (2019). *Library Philosophy and Practice (e-journal)*. 3614.

<https://digitalcommons.unl.edu/libphilprac/3614>

Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis

Gagandeep Kaur¹, Shruti Maheshwari²

¹Assistant Professor, IT Dept, Symbiosis Institute of Technology, Pune, India

²Assistant Professor, Mechanical Dept, Symbiosis Institute of Technology, Pune, India

¹gagandeep.kaur@sitpune.edu.in

²shruti.maheshwari@sitpune.edu.in

Abstract— A recruitment process requires an eligibility check, an aptitude evaluation and a psychometric analysis of prospective candidates. The work puts forward an application where the system allows employers to post new job offerings and registered candidates can apply. The application estimates applicant's emotional aptitude through a psychometric analysis based on a test whereas the professional standard is verified via a technical aptitude test. OCEAN Model is used to assess emotional quotient and predict the personality traits. Machine learning techniques such as Logistic Regression are used for modelling the personality predictor. The details of the candidates are kept secure by using a password encryption algorithm. The passwords are only known to the required individuals. The system outputs whether the candidate has been selected for the interview process via a dashboard and SMS alerts. A list of candidates is generated for the employer in order to keep track of the shortlisted candidates along with their scores.

Keywords— *Machine Learning, Password Encryption, Big Five Model, Recruitment.*

I. INTRODUCTION

Selection of candidates in a recruitment drive from a vast pool of candidates is a common issue. Traditional techniques consist of organizing personality and technical aptitude tests, group discussions and interviews. With the advancement in technology, the way recruitments are being conducted has shifted. Knowledge management systems (Online Recruitment Systems) are used to employ candidates, by using the medium, Web 2.0 and various social media websites. Using social media as a tool here poses newer, powerful

challenges and opportunities for employers. It also presents high efficiency, speed and attracts and targets specific, that is, relevant job applicants from the humongous number of candidates. [1]. A set of unique differences that are manipulated by the development of an individuals' personal memories, values, social relationships, attitude, skills and habits. Characteristics revealed in a certain pattern of behavior in a different bunch of situations form a personality trait.

Finding an individual personality trait and intelligence from his or her face plays a crucial role in inter-personal relationships but it is non-reliable. Results show personality traits such as social interaction, capacity of mutual respect, creativity, and many other traits cannot be just estimated by personal interactions [2]. In this approach, the escalating social media data has been used in multiple methods composed from scratch. A system proposed by Kessler et al. [12] separates applicants based on their relevance by analyzing unstructured text documents (job offers) and categorizes.

Another XML-based multi-agent recommender system, proposed by Pasquale De Meo [13] uses rich user profiles for support. The proposed system was a multi-agent recommender system which used XML for exploiting user profiles to enhance recruitment services in a personalized manner.

Another approach uses NLP for standardizing resumes through a modelling language approach. Despite in great usage, these techniques have disparities related to structure, inconsistent CV formats and contextual information. Additionally,

the applicants may show themselves in a well-behaved manner as an online questionnaire's responses can be manipulated for personality inference. Also, social networking sites contain data that is usually irrelevant for recruitment and thus shall not consist sufficient supplementary information regarding the candidate [3].

In this research paper, a system is proposed, which automates the eligibility check and estimates the emotional intelligence by leveraging the potentials of the data found in the test scripts. Various attributes of the test are processed for evaluating the candidate's personality in the system. The professional eligibility of a candidate is checked based on the entries in the online CV submitted by the applicants. Credibility is assured from the mandatory declaration of the users and also resolves the standardization issue. The prime intention is the reduction in the time spent on the initial recruitment phases keeping the end-goal of making the procedure more effective at a higher stage. Overriding of the decision-making capabilities of employers does not take place by the system. Rather, the proposed system helps in removing the time-consuming phases and shortens the tedious process.

Section 2 discusses architecture and various other aspects related to the system. Section 3 will look into the details of the underlying methodologies of the system.

II. SYSTEM FRAMEWORK

The system built consists of three prime components, which are:

Job Application Component: Registration by submitting online CV and extraction of required information and then performing an eligibility check with respect to the job description of the position being offered.

Evaluation Component: Estimation of the emotional and technical aptitude of the user by evaluating the test scripts. The standard used to mine the personality traits is OCEAN.

Applicants Grading Component: Combines the personality assessment and the eligibility criteria from the evaluation component for assigning a

relevant score. Ranking is on the basis of the relevant scores and this function is extracted from supervised algorithms.

A. Architecture

The never-ending amount of applications in traditional hiring tend to stun in-charges of selecting personnel. Meticulous parameterization is required by the HR department's expert members for the efficient working of existing automated grading systems. Change in criteria of hiring can make this a cumbersome and error prone technique. Candidate ranking models in the proposed system utilize Machine Learning algorithms (Logistic Regression). It involves ample amount of input training data for model training consisting of previous candidate decision data.

As in Figure 1, an eligibility check is performed by the candidate evaluator system, based on the entries in the online CV. This enables the HR infrastructure to be associated to the system for further commercial use.

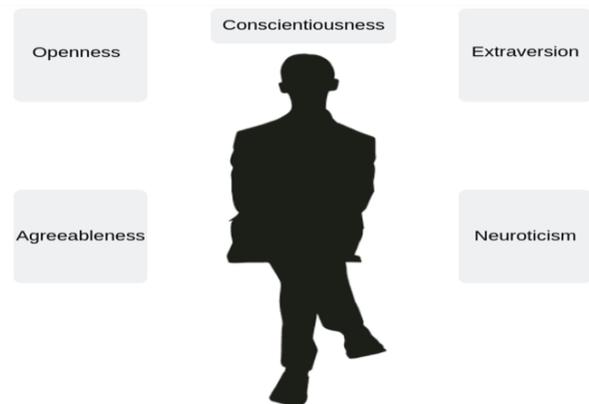


Figure 1: Big Five Model [1]

The main concerns here would be:

Professional Experience: It tallies the requirement mentioned by the recruiters and the candidate's previous work experience

Education: Courses perceived along their education period and the scores

Loyalty Index: Number of years spent on an average in the prior jobs

Co-curricular Activities: The hobbies and other activities help in the psychometric analysis of the candidate

The qualifications claimed including the estimated eligibility scores are stored in the database. The top candidates are shortlisted for interviews and the further process is handled manually.

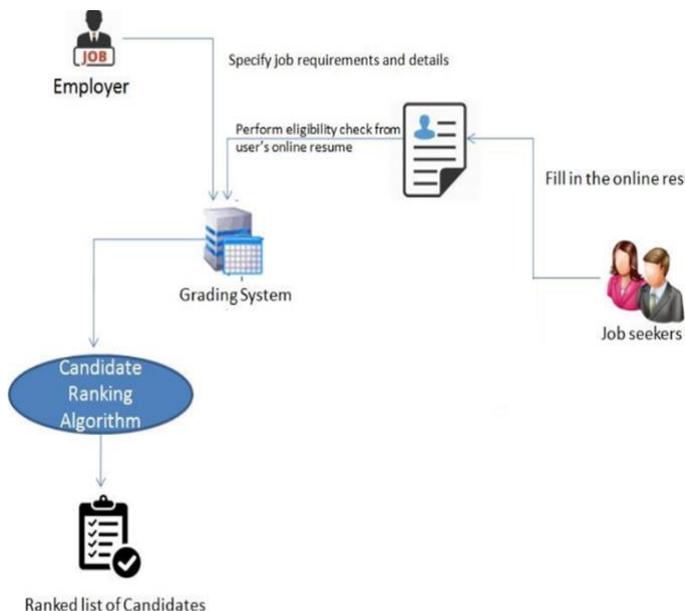


Figure 2: Proposed System Architecture

B. The Big Five Personality Model

Personality attributes of an individual are reflected in the way people present themselves on social networking platforms. The Five-Factor Model (FFM) uses such information for segregation into OCEAN traits, classifications which can be eventually utilized by other respective teams. Allegedly, deeper knowledge of personality data to extract specific information continues to be a hot research topic. The OCEAN model checks for the personality on the basis of agreeableness, openness, extraversion, conscientiousness and neuroticism. Although FFM is used vastly, it poses certain issues. Prediction and classification of individuals based on behavior is a research area and FFM(OCEAN) is frequently used. Complexity due to different interpretations is one of the challenges proposed [4]. The factors are explained below [1]:

Openness to experiences shows the degree of intellectual curiosity for creativity and acceptance of new ideology and liking for variety and novelty. High levels show novelty seeking and broad interests, while low levels say preference for conventionality and familiarity (data driven and expedient).

Conscientiousness is for people who are organized, disciplined, meticulous and punctual. Low level is perceived as spontaneous and flexible, but can be unreliable and tawdry. High level is seen as obsessive and stubborn.

Extroversion is for amicable, adventurous and talkative people while introversion is for reserved and timid ones.

Agreeableness shows friendly nature, trusting and how helpful a person is. Low levels suggest competitiveness while high levels show submissive or naïve nature.

Neuroticism judge emotions handling ability. Low levels show better grip on one's emotions whereas high levels show sensitivity and anxiety.

C. Security regime

The proposed system uses its own password encryption algorithm. When the user registers, the password is sent to the database after encryption. The encryption algorithm is a function which accepts a string parameter. The function consists of a complex combination of Pre-order, In-order and Post-order traversals of a Binary Search Tree (BST) consisting of cryptic combination of the string entered. The function returns a string which is then stored in the database. Thus, this function almost makes it impossible to decipher thus taking the data out of harm. Another specialty of this algorithm is that whatever is the length of the password, the database will store the password of a specific length only, i.e. all the passwords in the database will be of the same length. The proposed system has used a security regime as the data of the employers as well as the job seekers is sensitive and should not be exposed and used for any for any undesirable reasons.

D. SMS notifications

After the evaluation procedures, tests and ranking by the system, all the selected and unselected candidates are sent a notification via SMS regarding the application status. This can be achieved by libraries in python. The mobile numbers can be identified by retrieving them from the CVs of the registered candidates. This causes ease of dispatch of the application status than the traditional methodologies used.

III. PASSWORD ENCRYPTION

As mentioned earlier, the work has developed a password encryption algorithm using binary search trees (Figure 3). The algorithm is named as Tree XI as whatever is the size of the password entered by the user, the length of the password stored in the database will be 11 characters long. This algorithm is one way, i.e. there is no decryption available for the same. This adds more security to the whole system and difficult to crack into.

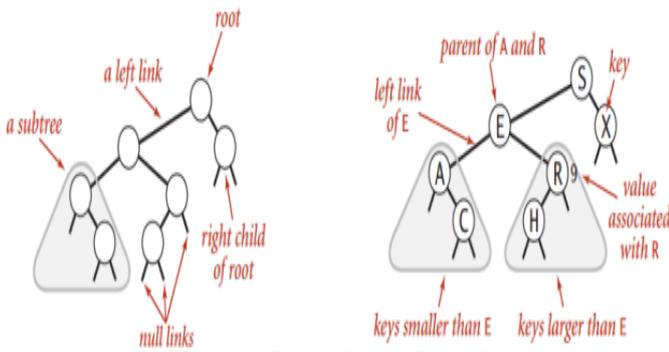


Figure 3: Binary Search Tree [5]

Following are the steps of the proposed Tree XI Password Encryption algorithm in detail:

- Input the desired password of the user.
- Find the length of the password.
- Convert each character of the password into its corresponding ASCII value.
- Divide each ASCII value by the square of the password length and store the remainder.
- Enter this remainder value into a binary search tree
- Perform pre-order, in-order and post-order operations on the binary search tree.
- Find the maximum and minimum values of the binary search tree and store them.
- Use the first 3 values of each of the traversals and the max and min values and store in an array, making the size 11.
- Convert these ASCII values into characters and store in a string.
- These 11 characters are the ones that form the encrypted password.

As this is a new encryption algorithm, it does have a few limitations that need to be taken care of, they are:

Limitations:

- For this algorithm to function a minimum of 4 characters as input are required from the user as a binary search tree uses distinct characters
- This algorithm will fail for large lengths of passwords

The idea of this algorithm is derived from one-way hash functions [8] [9]. This can also be a notion in public-key cryptography. Some sample inputs and their encrypted outputs would be:

```
Enter the password
Hello
|The encrypted password is - !!(6!(+(66
```

Figure 4: Encrypted Output 1

```
Enter the password
hello123
The encrypted password is - %%(,(%,%323|
```

Figure 5: Encrypted Output 2

```
Enter the password
123hello
The encrypted password is - %%(,1(%%/ ,3
|
```

Figure 6: Encrypted Output 3

It is evident from the above samples that even if all the characters are same, and they differ only by location in input string still there is a significant change in the output encrypted password.

Properties

- Deterministic – Always fixed length of output
- Fast – This uses binary search trees and hence are very fast
- Irreversible – Can't be decrypted
- Collision-Resistant – Difficult to find two strings that produce same encrypted password

From the mentioned properties, it can be assumed that the algorithm works similar to the SHA-1 algorithm.

Takeaways

- Pre-image Resistance – If the encrypted password is given and the algorithm is also given, it is difficult to find the input string
- Second Pre-image Resistance – If the input string, algorithm and the encrypted password is given, it is difficult to find another input string that will give the same encrypted password

- c) Unbreakable without using Brute Force Approach
- d) One-way

IV. EMPLOYEE GRADING MODULE

Generally, recruitment drives have job seeking candidates handing in physical resumes and employers checking the credentials. As outpour of applicants has increased recently, the amount of proposed CVs astounded the recruiters. Machine learning (ML) algorithms are used to build the results of the entire process until the interview phase. It requires a certain amount of input training data, which primarily consists of candidate selection decisions of previous candidates apportioned by experts in the field [1]. Eligibility scores generated for the analysis describes the candidate’s fitness for the job. The analysis features are saved as Boolean variables (an emotional assessment) or as numerical values (work experience). A learning algorithm is fed with the data that constructs the assessment when the candidate’s stated features are passed as input. The system, finally, generates the final ranked list to sort out the candidates [6] [7]. Suggested work shows the training data set as a set of input employee selection patterns. The work is sufficiently performed by the employee grading procedure and the evaluation tests using learning to rank algorithms such as Naïve Bayes and Support Vector Regression (SVR).

A. Naïve Bayes

In classification issues, the hypothesis could be assigning the class for a new data instance whereas in ML, a person selects the best hypothesis and data. Bayes’ Theorem provides an easy way of choosing the most probable hypothesis given data which can be used as prior knowledge, using which they can calculate the probability.

Bayes’ Theorem is stated as:

$$P(h|d) = (P(d|h) * P(h))/P(d)$$

where,

- P(h) - prior probability of h
- P(d) - probability of the data
- P(h|d) - posterior probability

P(d|h) - probability of data d given that the hypothesis h is true.

Selection of the hypothesis with the highest probability which is the Maximum a Posteriori (MAP) hypothesis can be done after finding various posterior probabilities for a variety of hypotheses. This can be written as:

$$\text{MAP}(h) = \max(P(h|d))$$

or

$$\text{MAP}(h) = \max((P(d|h) * P(h))/P(d))$$

or

$$\text{MAP}(h) = \max(P(d|h) * P(h))$$

In case of classification, the probability of each class (e.g.P(h)) will be equal if training data consists of an even number of instances in each class and it ends up with:

$$\text{MAP}(h) = \max(P(d|h))$$

B. Support Vector Regression

A set of methods for supervised learning called Support Vector Machines (SVMs) can be applied to both regression and classification problems. One of the major power sources for SVMs is nonlinear mapping of input space to a higher dimensional feature space in the kernel representation. Their objective is to discover a function that affects the data’s integral of a certain loss function and its unknown probability distribution. The empirical risk by which the original (undecided yet) one and the estimated function differs is reduced [10].

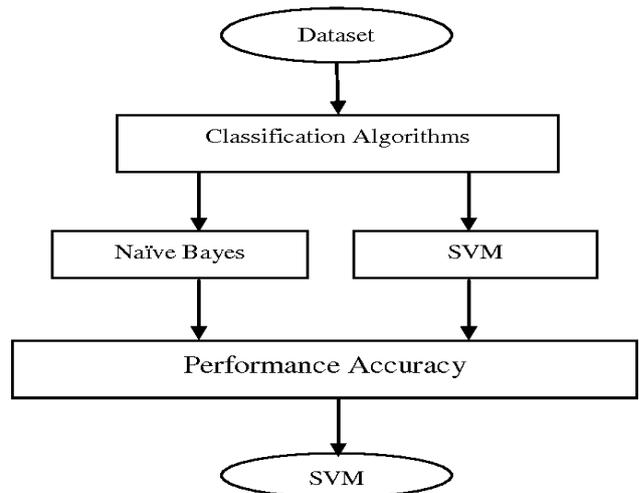


Figure 4. Machine Learning Algorithm [11]

C. Data set collection

The data set collection was done through a lot of websites and personal interactions with job seekers. The questions and responses were recorded and stored in a CSV file for easy data training and retrieval. The data set can be stored in any form: XML, JSON, etc.

V. RESULT ANALYSIS

From studies on proposed system, Table I was derived. Previous implemented studies using Naïve Bayes and Support Vector regression algorithms showed a higher coefficient of correlation. Thus, the system has improved accuracy than the previous existing systems but only to a certain extent i.e. a decrease of only 0.02 and 0.07 for Logistic Regression and SVR respectively.

The password encryption algorithm adopted here, was successfully implemented to store the password of the registered candidates and proved to provide a good security cover. The encryption process took a little longer time delaying the working the training of the system by a few minutes.

TABLE I. ELIGIBILITY SCORE OF DIFFERENT REGRESSION ALGORITHMS AND THEIR CORRELATION COEFFICIENTS

Job Position	Coefficient of Correlation	
	Logistic Regression	Support Vector Regression
Associate Programmer	0.63	0.65
Senior Manager	0.62	0.69

VI. CONCLUSION

In the work presented, an efficient and effective approach is used to rank and evaluate candidates through psychometric analysis for calculating emotional quotient. Technical eligibility criteria from the online CVs and emotional aptitude by leveraging responses in evaluations are processed by the proposed system. The OCEAN model performs the linguistic and personality analysis of the candidates. 65% and 87% average accuracy levels were obtained by the algorithm dependent and independent approaches respectively. The

accuracy obtained by the proposed system is higher than the previously implemented systems which had 45% average accuracy levels by the algorithm dependent approach. The Tree XI algorithm created is adding an extra layer of security to the existing ones. There is also scope for improvement in the algorithm as currently it can't handle larger lengths of input data. Employee grading process uses supervised algorithms which are trained with previous recruitment data. The future scope of the project could be to add a video and image processing elements. Also, the CV standardization issue can be dealt with by uploading CVs in a certain format. This would increase the efficiency of a system by a certain level.

REFERENCES

- [1] Vishnu M. Menon and Rahulnath H A, "A Novel Approach to Evaluate and Rank Candidates in a Recruitment Process by Estimating Emotional Intelligence through Social Media Data", in Big Five Model, 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 978-1-5090-0870-4/16
- [2] Gayatri Vaidya, Pratima Yadav, Reena Yadav, Prof. Chandana Nighut, "Personality Prediction By Discrete Methodology", in Personality traits, IOSR Journal of Engineering (IOSRJEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719, Volume 14, PP 10-13
- [3] Mohammad Mehrad Sadra, "The role of personality traits predicting emotion regulation strategies", International Academic Journal of Humanities Vol. 3, No. 4, 2016, pp. 13-24, ISSN 2454-2245
- [4] Muhammad Fahim Uddin, Jeongkyu Lee, "Proposing Stochastic Probability-based Math Model and Algorithms Utilizing Social Networking and Academic Data for Good Fit Students Prediction", IEEE International Conference – 2018, DOI 10.1007/s13278-017-0448-z
- [5] <https://medium.com/omarelgabrys-blog/binary-search-trees-598b657a779b>
- [6] Iftekhar Naim, M. Iftekhar Tanveer, Daniel Gildea and Mohammed (Ehsan) Hoque, "Automated Analysis and Prediction of Job Interview Performance", IEEE Transactions on Affective Computing, Volume: 9, Issue: 2, April-June 1, 2018
- [7] Muhammad Fahim Uddin, Soumita Banerjee, Jeongkyu Lee, "Personality Based Recommendation Engine (PBRE)", 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Electronic ISBN: 978-1-5090-3207-5, Print on Demand (PoD) ISBN: 978-1-5090-3208-2
- [8] Carolyn Winslow, Xiaoxiao Hu, Seth Kaplan, Yi Li, "Accentuate the Positive: Which Discrete Positive Emotions Predict Which Work Outcomes?", The Psychologist-Manager Journal – 2017, The Psychologist-Manager Journal, 20(2), 74-89
- [9] N. & M. S. Iyer, "Implementation of Secure Hash Algorithm-1 using FPGA," International Journal of Information and Computation Technology, Volume III, Issue 3, pp. 50-60, 2013.

[10] Shaza D. Rihan, Ahmed Khalid and Saife Eldin F. Osman, "A Performance Comparison of Encryption Algorithms AES and DES", International Journal of Engineering Research and Technology (IJERT) ISSN : 2278-0181 Vol. 4 Issue 12, December-2015.

[11] Tommy Tandra, Hendro, Derwin Suhartono, Rini Wongso, and Yen Lina Prasetio, "Personality Prediction System from Facebook Users", 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia, Vol. No. 10.1016/j.procs.2017.10.016

[12] Kessler, R., Béchet, N., Torres-Moreno, J.-M., Roche, M., and El-Bèze, M., "Job Offer Management: How Improve the Ranking of Candidates", in Foundations of Intelligent Systems, 431-441.doi:10.1007/978-3-642-04125-9_46

[13] Pasquale De Meo, Giovanni Quattrone, Giorgio Terracina, and Domenico Ursino, "An XML-based Multi-Agent System for Supporting Online Recruitment Services", IEEE Transactions on Systems, Man, And Cybernetics, Vol. 37, No. 4, 2007.

