# PERSONALITY PREDICTION SYSTEM THROUGH CV ANALYSIS

**BY**

**Afroja Khatun Monalisa**

**ID: 171-15-9503**

**And**

**Md. Omar Kaiser Mahin**

**ID: 171-15-9487**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Ms. Nazmun Nessa Moon**
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

**Most. Hasna Hena**
Assistant Professor
Department of CSE
Daffodil International University


**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

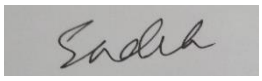
**9th September 2021**

# APPROVAL

This Project titled "**Personality Prediction System Through CV Analysis**", submitted by **Afroja Khatun Monalisa**, ID:**171-15-9503**, **Md. Omar Kaiser Mahin** , ID: **171-15-9487** to the Department of CSE, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in CSE and affirmed concerning its style and substance. The presentation has been held on 9 September 2021.

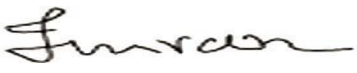
## BOARD OF EXAMINERS

**Dr. Touhid Bhuiyan** — **Chairman**
**Professor and Head**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



**Md. Sadekur Rahman**
**Assistant Professor** — **Internal Examiner**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Gazi Zahirul Islam** — **Internal Examiner**
**Assistant Professor**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



**Shah Md. Imran** — **External Examiner**
**Industry Promotion Expert**
LICT Project, ICT Division, Bangladesh

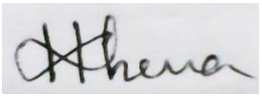
# DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon, Assistant Professor, and Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**

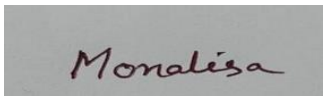
**Ms. Nazmun Nessa Moon**
Assistant Professor
Department of CSE
Daffodil International University

**Co-Supervised by:**

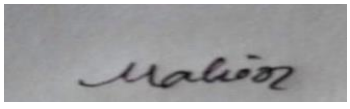


**Most. Hasna Hena**
Assistant Professor
Department of CSE
Daffodil International University

**Submitted by:**



**Afroja Khatun Monalisa**
Id: 171-15-9503
Department of CSE
Daffodil International University



**Md. Omar Kaiser Mahin**
Id: 171-15-9487
Department of CSE
Daffodil International University

# ACKNOWLEDGEMENT

First, we express my heartiest thanks and gratefulness to Almighty God for his divine blessing that makes us possible to complete the final year project successfully.

We are truly thankful and wish my significant and obligation to **Ms. Nazmun Nessa Moon, Assistant Professor,** Department of Computer Science and Engineering at Daffodil International University, Dhaka. Her interminable persistence, academic direction, consistent support, steady and vivacious management, useful analysis, important guidance, perusing

Finally, we must acknowledge with due respect the continuous support and patients of our parents.

# ABSTRACT

This project titled "Personality Prediction System through CV Analysis", which is built to solve the CV analysis problem. This project is a construction of solutions to CV analysis problems. In this system, in order to make the recruitment process more effective and efficient, we are building a personality prediction model. In this system, we will implement the model based on the resume, ability and personality test . The concept behind the model is to add candidates to the shortlist according to the administrator's request. Therefore, the workload of the and the time required to select the best horsepower will be reduced. First, candidates will be shortlisted based on of their resumes, and then they will get ability and personality test links. Based on scores and HR requirements, candidates will be selected. To carry out this system, random forest algorithm, support vector machine, weighted majority voting algorithm are studied

# TABLE OF CONTENT

# LIST OF FIGURES

# LIST OF TABLES

**TABLE NAME** **PAGE**

# CHAPTER 1
# INTRODUCTION

## 1.1 Introduction

The proposed framework has two sides: it will be either the candidate lying or the lying organization. In the first case, the manager will offer the candidate a list of jobs more suited to his abilities. In the current situation, Scouting will disseminate the determinations and prerequisites of accessible positions and candidates will be able to apply for the same position by submitting their CV. The existing recruiting framework essentially filters submitted CVs and the candidate waiting list, in which the framework recommends performing physical and identity checks on the web and then predicting the candidate's identity as well. How to select a candidate based on their skills and decision-making

## 1.2 Motivation

Our country is an unemployment country. Where finding a job is a lot of trouble. We find a system that will help people get jobs. This will empower a more successful way to brief list submitted candidate CVs from a huge number of candidates giving a reliable and reasonable CV positioning approach, which can be lawfully defended. Framework will rank the encounter and key aptitudes required for specific work position. Than framework will rank the CV’s based on the involvement and other key abilities which are required for specific work profile. This framework will offer assistance the HR office to effectively waitlist the candidate based on the CV positioning arrangement. This framework will center not as it were in capability and involvement but moreover centers on other critical viewpoints which are required for specific work position. This framework will offer assistance the human asset office to choose right candidate for specific work profile which in turn give master workforce for the organization.

## 1.3 Project Objectives

The objectives of the project are as stated below:

• To create a system that will provide a more efficient method of short-listing candidates.

• To identify the important skill characteristic, each expert's preferences and ranking decisions must be defined.

• To simplify the process of defining requirements and ranking applicants.

• To administer an online skill and aptitude tests.

• To make ranking decisions that are more consistent than those made by human experts.

## 1.4 Rationale of the Study

The current e-recruitment system scans the presented CV simply and lists the candidates for an online aptitude test and personality assessment based on their ability and decision-making, thus predicting the personalities of the candidate and shortlisting the candidate. The system proposed is two-sided: candidate-oriented or organizational. In the first instance, a list of jobs would be recommended to the candidate that best suits his abilities. The recruiter would publish the requirements and specifications of available jobs in the second scenario, and by presenting their CVs candidates can apply for the same.

## 1.5 Expected Outcome

In our CV analysis system is a system that helps to generate an expected result based on the given dataset. In this system, we used 77% of the training to get more accurate predictions. After completing all the essential procedure of the proposed system, our CV analysis system has been ready for preparing out on the given dataset. We have applied various strategies to achieve our desired results. We got 77% accuracy from SVM, 68% from LR, 71% from the K-Nearest Neighbors (KNN), 60% from the Decision Tree among all that we have used.

## 1.6 Report Layout

This research paper consist with some chapter-wise contents which are given below at a glance:

**Chapter 1:** This chapter is the representation of the main objective and motivation of us for doing this research. It also represents what we are looking for from this project.

**Chapter 2:** This chapter is the representation of the background studies we have done for this research and also it shows the obstacles we had faces during working on this project.

**Chapter 3:** This chapter is the representation of the research subject, what instrument we needed, data collection procedure, processing data, and statistical analysis along with the implementation.

**Chapter 4:** This chapter is the representation of the experimental result, project descriptive analysis, and our findings in summary.

**Chapter 5:** This chapter is the last chapter which represents the outcome of this project, some of our recommendations, and a further study plan of us about this research with our finishing counsels.

# CHAPTER 2
# BACKGROUND

## 2.1 Introduction

The proposed system tries to design a plan to integrate the job specification model into the EHR system in order to find a new effective operating model of human resource management in the Internet age. In this project, we present a set of techniques that make the entire recruitment process more efficient and effective. We have established a system that ranks applicants according to the weight policy as well as the aptitude test.

## 2.2 Related Works

In this part examined about some comparable related works with this project-

- Personality Prediction System using AI - Dhanashree Sonaje et. al. described in her paper they had implemented system which is useful for any recruiter for recruitment process where the applied candidates are in a huge number. This system will reduce the workload of recruiter or Human Resource Department.
- Evaluating Attributed Personality Traits from Scene Perception Probability- Hancheng Zhu et. Al. described in his paper by letter purposed. This card proposes a personality prediction method based on the perceptual probability of image scenes. The convolutional neural network is first used to identify the scene contained in the image that the user likes, and the linear regression model is trained on the image using the associated attributable personality scores
- Personality Prediction through Curriculum Vitae Analysis involving Password Encryption and Prediction Analysis- Gagandeep Kaur et. Al. described in his paper by Machine Learning, Password Encryption, Big Five Model, Recruitment.
- Personality Prediction System through CV Analysis - Allan Robey et. Al. described in his paper they had implemented an organization oriented

recruitment system that would assist the human resource department in short listing the right candidate for a specific job profile. The system would be used in many business sectors that will require expert candidate, thus reducing the work load on the human resource department.

## 2.3 Research Summary

In this system, to make the recruitment process more helpful and efficient, we are building the personality prediction model. In this system, we are going to implement the model which is based on CV, aptitude and personality test. The concept system behind this model is to shortlist the candidates according to the admin requirement process.

## 2.4 Scope of the problem

We have faced many problems to complete this paper. But we have overcame these difficulties. Where we need to know about the process of the job opportunities. So, we collected information about the CV analysis by personality prediction system.

## 2.5 Challenges

Here is some challenges which we have faced are mentioned below:
Data collection is one of the big challenges for getting predicting accuracy. Without data, the prediction is not possible and it can't predict. After that, another challenge is preprocessing. After doing preprocessing our data set has no zero value and helps us to get a good prediction. Next, Feature scaling helps to take all feature values into the same scale with respect to value. Therefore, different algorithm has been applied to the proposed architecture. Finally, the implementation process has been established to get accurate predict

# CHAPTER 3

# RESEARCH METHODOLOGY

## 3.1 Introduction

In this proposed system, we have implemented system which is useful for any recruiter for recruitment process where the applied candidates are in a huge number. This system will reduce the workload of recruiter or Human Resource Department. In Figure 3.1 we have shown about our research Methodology at a glance.


Figure 3.1: Research Methodology at a glance

## 3.2 Research subject and instrumentation

Machine learning calculations have recently grown in popularity. Machine Learning Algorithms provide computers the ability to learn from data using measured techniques. For instance, a machine can discover inner information and, as a result, deliver a choice or foresight learning without the need of express. Coding is regarded as the most valuable component. In this manner, a similar calculation can be linked to datasets from various areas without requiring a change in its internal structures. There are various types of machine learning calculations, but we used some of them to our framework. We have used four type of algorithm – SVM, LR, KNN, Decision Tree.

### 3.2.1 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is an isolation hyperplane formally described as a discriminatory classification classifier. At the end of the day, the calculation produces an ideal hyperplane which gives new precedents, given the significant preparation information (managed input). This hyperplanet is a line in two-dimensional space that divides a plane into 2 sections, in which the numeric info factors (x), in your information, form an n-dimensional space on either side of every class. This would frame a two-dimensional space, for example, on the chance you had two factors for information. A hyperplane is an information variable space dividing line. The hyperplane of SVM is chosen to better distinguish the focuses by class 0 or class 1 within the info variable space. You can see this as a line in two measures, and we should expect that this line will totally insulate the majority of our data centres. Examples are:

B0+(B1*X1)+(B2*X2)= 0……(2)
In Equation 2, the learning calculations are determined by the coefficients (B1 and B2) which decide the incline and the capture (B0), and both X1 and X2 are factors.

### 3.2.2 Logistic Regression

Regression logistics, falling within the Machine Learning Supervised. It deals with classification issues (to settle on forecasts or take choices dependent on past information). Two times the results for a certain array of free factors are forecasted. The result of the predictor variables is discreet. Another method acquired in the statistical area is logistical recurrence through the machine. It's the strategy to deal with paired problems with characteristics. Logistic relapse uses a condition, especially as a direct relapse. Info estimates (x) are directly combined using weights or coefficient appreciation(s) to predict a yield appreciation (y).
A strategic retrospective model is as follows:
y =e^(b0 + b1*x)/(1+e^(b0 +b1*x)) ……. (1)

From Equation1, the expected y output is y, b0 is the inclination or collecting term and b1 is the single information value coefficient (x).

### 3.2.3 K-nearest neighbor (KNN)

In the adjustment group managed, KNN falls. Incidentally, this implies that a marked dataset consists of preparing perceptions (x, y) and that the connection between x and y might be desired. More formally, it is likely to be h:X vs. Y with the purpose of unconsciously predicting the corresponding y output given a disguised perception x, h(x). A non-parametrical and occasional learning calculation is also the KNN classification. It does not mean that it presupposes clearly the utilitarian type h, and does not risk mismodeling the basic dispersion of the data. For example, assume that our information is exceptionally non-Gaussian, but we accept a Gaussian form as a learning model. All of this would make our calculation a poor forecast to a great extent. Case-based learning does not expressly take a model into our calculation. It instead retains the preparation events, which are thus used for the prediction phase as 'learning.' This means that only when an inquiry is submitted to our database (i.e., if we ask for the name given to you), does the calculation take advantage of the opportunities to make a reply available?

In the schedule setting, the K-closest neighbor calculation basically results in the formation of a large part of the K vote for some "disguised" perception. Similarity is distinguished by a divider between the two focuses of information. The Euclidean separation given by is an important decision.

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \ldots + (x_n - x'_n)^2} \quad \ldots\ldots\ldots\ldots (4)$$

In equation 4, the KNN classifier plays out the two accompanying phases, given the positive total number K, an unclear perception x and an equivalence metric d: It traverses the complete d-registered data set between x and each perception. In the preparation information closest to the set, we will call the K focuses. Note that K is usually strange in the circumstances of a tie. The separation of another information point from all other information preparation focuses is essentially determined. It can be separated by every type, eg. Euclidean or Manhattan, etc. It chooses the K-closest data focus at that point, where K may be any number. Now, it is likely that the

variable K and its impact on that classifier are taken into consideration. All things taken into account, the K in KNN is a hyperparameter, like most machine calculations, which you, as a creator, must select with the aim of fitting most effectively into the information index. It can, of course, consider K monitoring the state of the choice limit we previously discussed. When K is small, we restrict the range of a certain forecast to "more visually impaired" general transmission. A small incentive to K gives the most adapted fit, which is low but highly fluctuating. In graphic terms, our limit of choice is jagged. On the other hand, in every forecast, a higher K-medium is more votes and therefore stronger than the exceptions. Bigger K estimates are smoother in choice, meaning change is reduced but predispositions are expanded. The group quantity, K, must be fixed in advance. Its inconvenience is that with each run, the results are not similar, since subsequent bundles rely on the irregular assignments underlying them. We never know the true group, using the same information, assuming that they are provided in an alternative request. If the amount of information is small, they may be different. As far as we know, datasets for the KNN display building are very well arranged. Since KNN is a calculation which is non-parametric, we will not acquire the model parameters. A vector with elements of characterization of the test set is restored by the work of KNN ().

### 3.2.4 Decision tree (CART)

Decision trees are essential for precise machine learning calculations.
Traditional decision tree calculation has been around for a long time now, and today's varieties, such as arbitrary forests, are among the most advanced methods.
A humble calculation of a decision tree known by the current CART name, representi ng Trees of Classification and Regression. The Decision Tree technique is used as the most comprehensive asset for the machine to be used as soon as possible. The tree chosen is computed into different types: Cart, ID3, C 4.5, CHH, and H48, respectively. J48 is used among them and the mainstream algorithm is exceptionally. J48 uses the technique of pruning for tree construction. This calculation continues to be a recursive procedure until the normal results are identified. It provides great accuracy and adjustment. The following equations make this formula available.

$E = \sum_{i=1}^{k} Pi\ log2\ Pi$ .....(4)

In Equation 4, K sets the number of target attribute classes,

Pi sets the number of class occurrences,

and I sets the number of instances.

This calculation is traditionally referred to as "decision trees," but it is referred to in a few phases, such as R in the later term CART.

3.5 Selected Algorithm

In order to achieve maximum exactness from our datasets, we use different algorithms. We are showing in this figure 3.2 which algorithm, among other algorithms, is the most precise.

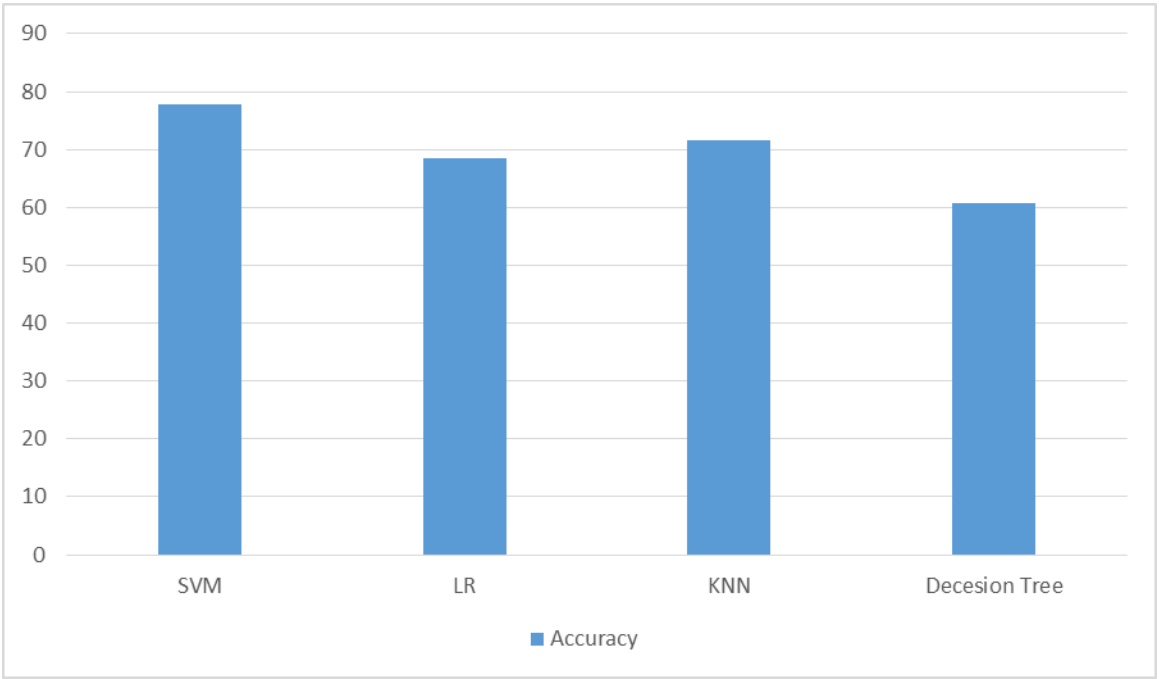



Figure 3.2: The accuracy of the estimates for Four models

In Figure 3.2, we have four models of logistic regression: k-classifier, Decision Tree Classifier, SVM and estimates of the accuracy for each. We must compare the models and select the most accurate SVM.

## 3.3 Data Collection Procedure

We have attempted heart and soul to gather great quality of information. To begin with of all we attempted to gather information by online. The medium computer program was Google Frame. At that point we had collected a few of the information which tuned into a total catastrophe. Since individuals are not interested to share their private information with us. At that point we chosen to our boss and examined approximately this matter. After a long considering we have come to an conclusion that we'll collect information physically. And we begun collecting information by a scripts where the respondent had to fair tick a few yes or no questions. This time we were able to urge great quality of information. We had attempted to gather great quality of information.

## 3.4 Statistical Analysis

In our system dataset, we are doing 77% SVM algorithm, 71% KNN algorithm, 68% LR algorithm, 60% Decision Tree algorithm. In this system we try to find out accuracy from the dataset and use KNN method to predict CV analysis. In Figure 3.43we have shown our system for statistical analysis.

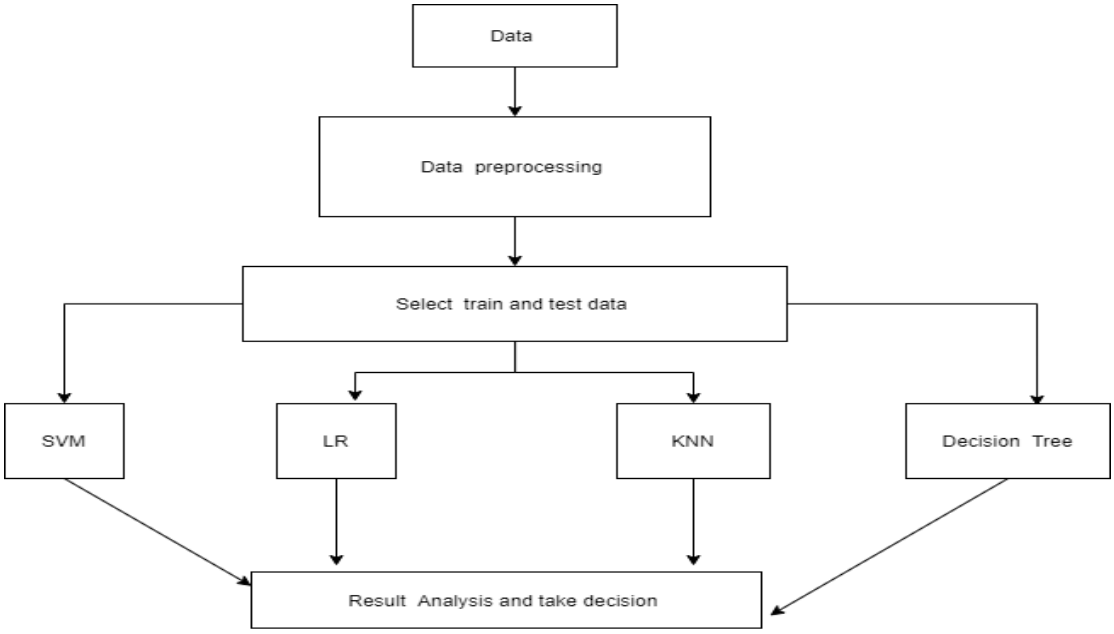



Figure 3.3: System for statistical analysis

## 3.5 Implementation Requirements

We are using this algorithm-

SVM algorithm: SVM is a supervised **machine learning algorithm** that is commonly used for classification and regression challenges. Common applications of the SVM algorithm are Intrusion Detection System, Handwriting Recognition, Protein Structure Prediction, Detecting Steganography in digital images, etc.

KNN algorithm: KNN is one of the simplest forms of machine learning algorithms mostly used for **classification**. It classifies the data point on how its neighbor is classified.

Naïve Bayes: Naive Bayes Algorithm is one of the popular **classification machine learning algorithms** that helps to classify the data based upon the conditional probability values computation. It implements the Bayes theorem for the computation and used class levels represented as feature values or vectors of predictors for classification.

Decision tree: A **decision tree algorithm can handle both categorical and numeric data and is much efficient compared to other algorithms.** Any missing value

present in the data does not affect a decision tree which is why it is considered a flexible algorithm.

# CHAPTER 4
# EXPERIMENTAL RESULT & DISCUSSION

## 4.1Experimental Results

In accordance with our requirements, we update our models and data sets. Through this modification, we can realize that the classifier we use is fully suitable for a wide range of purposes according to our data set. As we expected, we achieved an 77 % accuracy from our proposed model, which is a fruitful result. This performance of the model creates a way of thinking about experimental results which is shown in table 3.1.

Table 3.1: Experimental Result

| Algorithm | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 0.77894 | 0.7894 | 0.84 | 0.8139 |
| LR | 0.68421 | 0.7826 | 0.4675 | 0.5853 |
| KNN | 0.71578 | 0.8775 | 0.5972 | 0.7107 |
| Decision Tree | 0.60736 | 0.4905 | 0.5777 | 0.5306 |

In Figure 4.1 we have discussed about accuracy of our four types algorithm.

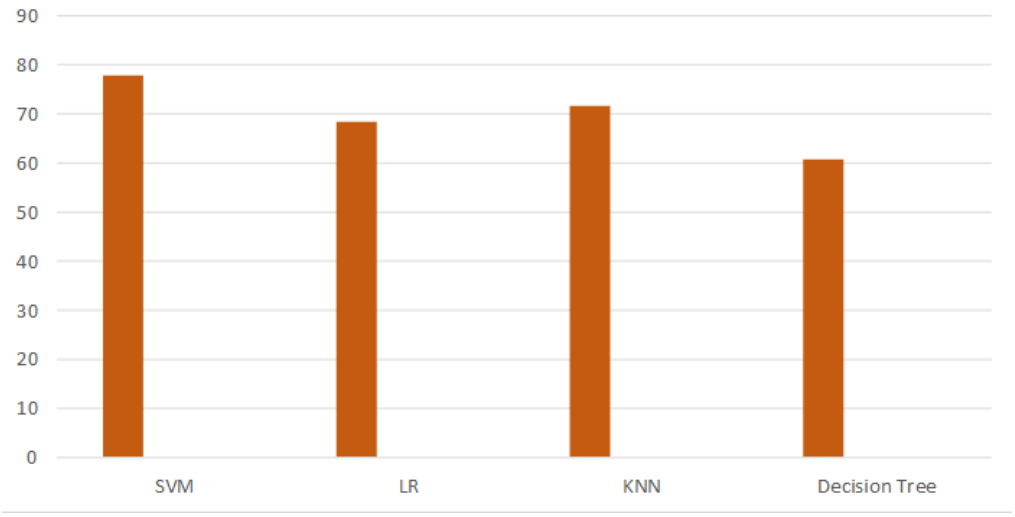



Figure 4.1: Accuracy

# CHAPTER 5

# SUMMARY, CONCLUSION, RECOMENDATION AND IMPLICATION FOR FUTURE

## 5.1 Summary of the Study

In this proposed system, we have implemented system which is useful for any recruiter for recruitment process where the applied candidates are in a huge number. This system will reduce the workload of recruiter or Human Resource Department.

## 5.2 Conclusion

In this project, we have actualized an organization situated enlistment framework that would help the human asset office in brief posting the proper candidate for a particular work profile. The framework would be utilized in numerous trade segments that will require master candidate, hence decreasing the work stack on the human asset division.

## 5.3 Recommendation

We Provided a more effective way for the candidate to develop the system. Then Determined the significant characteristics of the ability by defining the preference and classification decisions of each expert. After we Automates the process of the request specification and the classification of the applicant. We perform fitness and testing online characters. produces a classification decision that is more consistent than experts in human experts.

## 5.4 Implication for further study

In our paper, we have many features. Some are similar and some different comparing to the mentioned papers. Moreover, we have developed it with full information which is more user friendly to users, admin, authentic of job authority.

- Firstly, the candidate will be shortlisted on the basis of CV and then aptitude and personality test link will be given to him.
- According to the score and HR requirement candidate will be selected.
- For implementing this system, SVM Algorithm, Naïve Bayes, Decision Tree, KNN Algorithms are studied.

# REFERENCES

[1] Hancheng Zhu, Leida Li, Sicheng Zhao, Hongyan Jiang, **"**Evaluating Attributed Personality Traits from Scene Perception Probability" Pattern Recognition Letters (2018),doi: https://doi.org/10.1016/j.patrec.2018.09.027

[2] Dhanashree Sonaje1, Shrutika Alladwar2, Neha Bhise3, Vishakha Dhote4, Ketan Desale5. "Personality Prediction System using AI" Volume: 06 Issue: 11 | Nov 2019 www.irjet.net

[3] Vinay Desai, Dheeraj Bahl, Shreekumar Vibhandik,Isra Fatma,"Implementation of an Automated Job Recommendation System Based on Candidate Profiles"-Volume: 04 Issue: 05 | May -2017 www.irjet.net

[4] Gagandeep Kaur1, Shruti Maheshwari2, "Personality Prediction through Curriculam Vitae Analysis involving Password Encryption and Prediction Analysis"-Winter 11-7-2019

[5] Allan Robey1, Kaushik Shukla2, Kashish Agarwal3, Keval Joshi4, Professor Shalmali Joshi5,' "Personality Prediction System through CV Analysis"-Volume: 06 Issue: 02 | Feb 2019 www.irjet.net

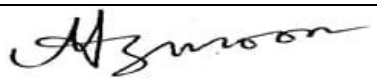
# APPENDICES

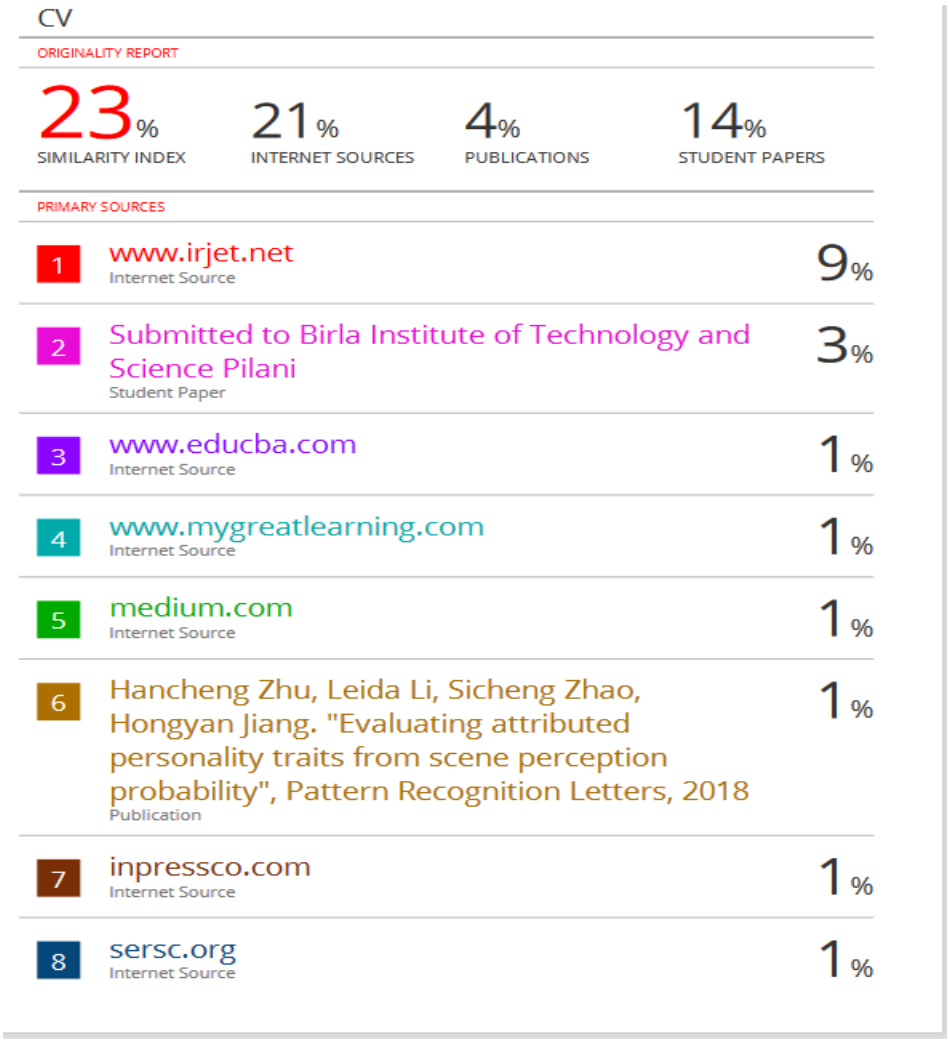
## Appendix :

## Research Reflection

At the exceptionally to begin with period of this investigate, we had nearly no thought with respect to AI and ML. Our executive was especially kind and sincere and he gave us imperative heading and concepts to fulfill this investigate. In this whole season of inquire about, we learned various modern strategies, collect information, figured out how to utilize calculation additionally how to work with different strategies. I moreover moreover found out almost Anaconda-navigator, Jupyter note pad, Google Colaboratory, and Python programming dialect. At to begin with, there were issues working with these, yet steadily we turned out to be increasingly familiar with Python. We too have confronted a parcel of issue whereas collecting information. Managing with thousands of individuals and persuading them for giving us about their secret information was not a simple errand. At last after a part of battle we accomplish this last yield

# Plagiarism Report

| | |
|---|---|
| Plagiarism Checked By<br>Ms Nazmun Nesa Moon<br>Assistant Professor, DIU |  <br>08-Sep-2021 |



CV

ORIGINALITY REPORT

| 23% | 21% | 4% | 14% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.irjet.net<br>Internet Source | 9% |
| 2 | Submitted to Birla Institute of Technology and Science Pilani<br>Student Paper | 3% |
| 3 | www.educba.com<br>Internet Source | 1% |
| 4 | www.mygreatlearning.com<br>Internet Source | 1% |
| 5 | medium.com<br>Internet Source | 1% |
| 6 | Hancheng Zhu, Leida Li, Sicheng Zhao, Hongyan Jiang. "Evaluating attributed personality traits from scene perception probability", Pattern Recognition Letters, 2018<br>Publication | 1% |
| 7 | inpressco.com<br>Internet Source | 1% |
| 8 | sersc.org<br>Internet Source | 1% |