CHARACTER ANALYSIS OF A PERSON

By
K.HEMA ANMISHA 17MIS1059
R.APARNA   17MIS1065
P.NITYASREE   17MIS1007

A project report submitted to

Dr. NAGARAJ S V

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

in partial fulfilment of the requirements for the course of

SWE2009 -   Data Mining Techniques

in

M.Tech. (Software Engineering)

WIN SEM 2018-19

VIT UNIVERSITY, CHENNAI

Vandalur – Kelambakkam Road Chennai – 600127

1. INTRODUCTION

Personality identification of a human being by their nature an old technique. Earlier these were done manually by spending lot of time to predict the nature of the person. Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations.

2.OBJECTIVES:

This project's objective is all about predicting and analysing a particular person's character based on their answers for the given set of questions. This process is carried out using the naive bayes technique used in data mining. The user is asked to answer a set of questions which are helpful for finding their character. This can be done for a large crowd too.
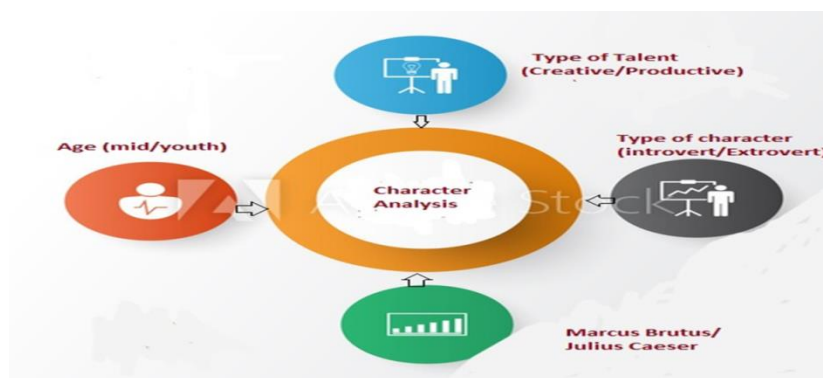
3.DATA STRUCTURES AND ALGORITHMS USED:

- Naive Bayes Algorithm
- Decision Tree Algorithm

4.ARCHITECTURE DIAGRAM:
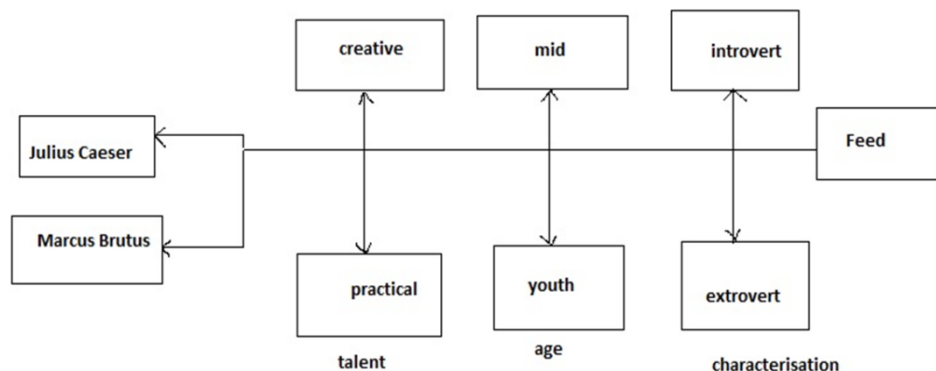

5.DETAILED DESIGN:

6.PERFORMANCE EVALUATION:

- This project comes across areas where it has access to large amounts of person behavioral data. This data can be helpful to classify persons using Automated personality classification (APC).

- In this project, the system proposes an advanced APC – automated personality classification system. The system uses learning algorithms like Naive Bayes and SVM, Decision tree along with advanced data mining to mine user characteristics data and learn from the patterns.

- This learning can now be used to classify/predict user personality based on past classifications. The system analyses vast user characteristics and behaviors and based on the patterns observed, it stores its own user characteristics patterns in a database.

- The system now predicts new user personality based on personality data stored by classification of previous user data.

- This system is useful to social networks as well as various ad selling online networks to classify user personality and sell more relevant ads. Also the system is useful for government agencies to observe user personality and predict new user personality on a large scale.

7.APPLICATIONS IN SOME DOMAIN:

- It is useful to social networks as well as various ad selling online networks to classify user personality and sell more relevant ads. Also the system is useful for government agencies to observe user personality and predict new user personality on a large scale.

- By conducting quiz, we can know the character of the students which helps the teachers analyze how they can teach a set of students.

- In grocery shops we can predict by this method , the kind of goods/stationeries that are purchased by the people frequently.

- We can know how the student approaches to a problem(Practically or Creatively)

- We can understand the character of a person,whether he is an extrovert or an introvert.

8.EXPERIMENT RESULTS:

intelligence_analysis

status_analysis
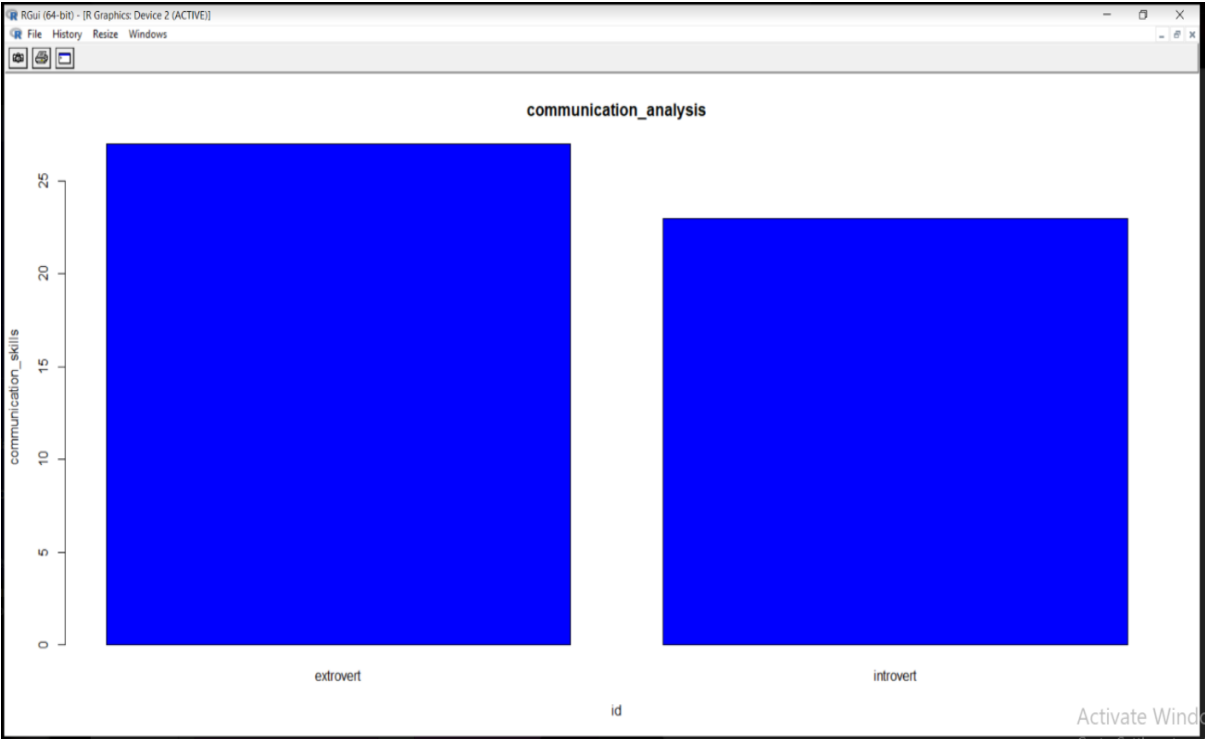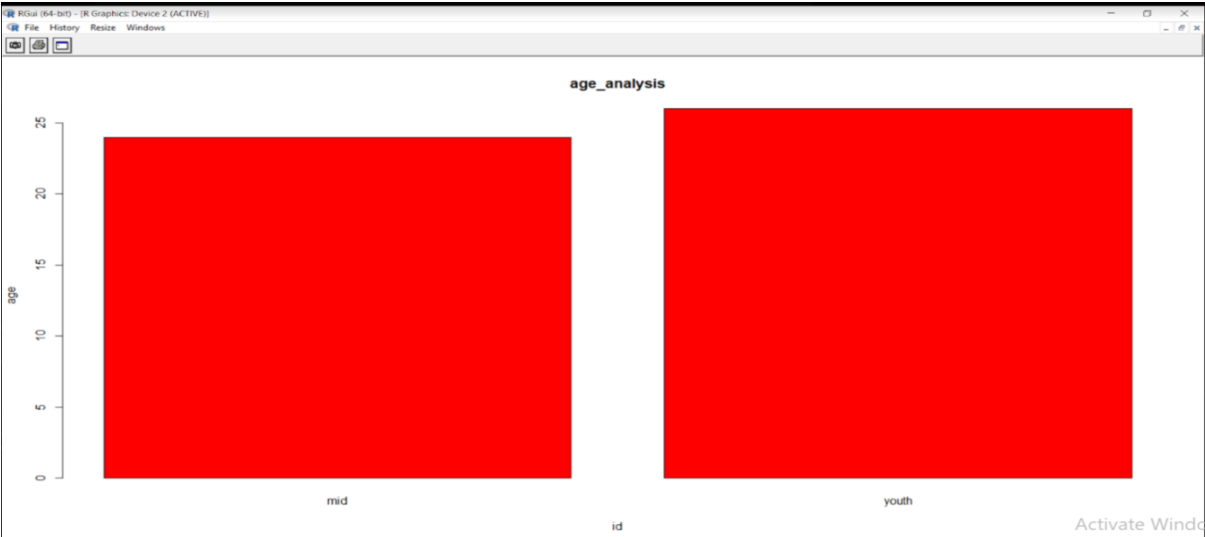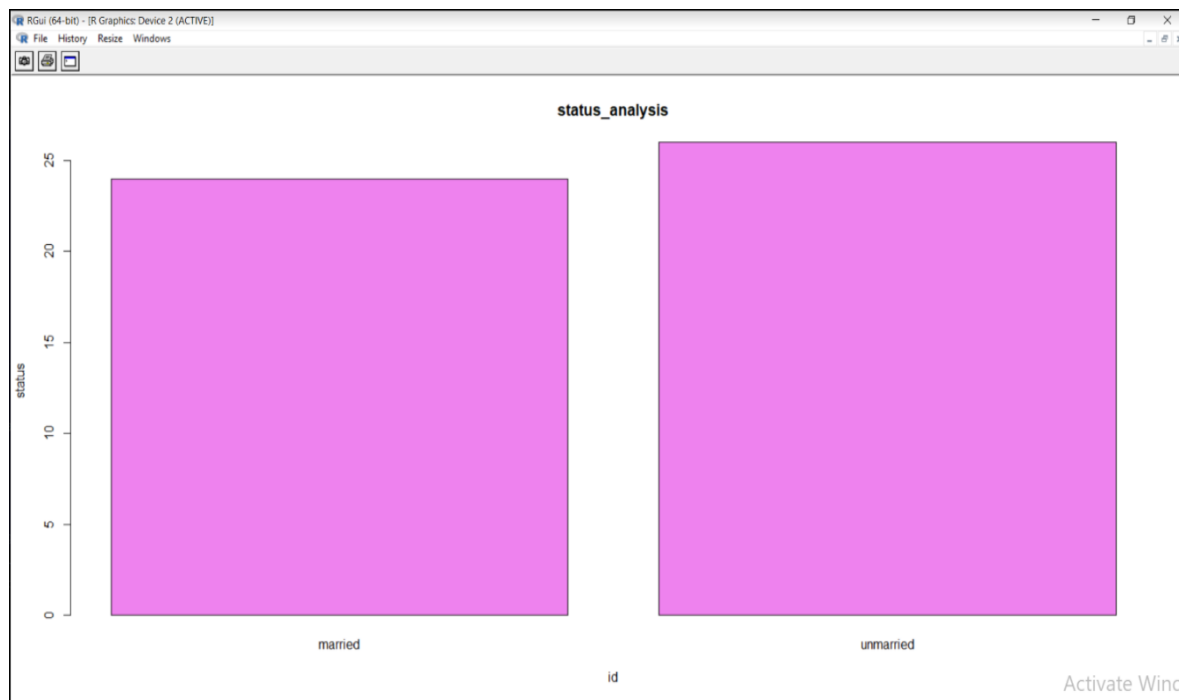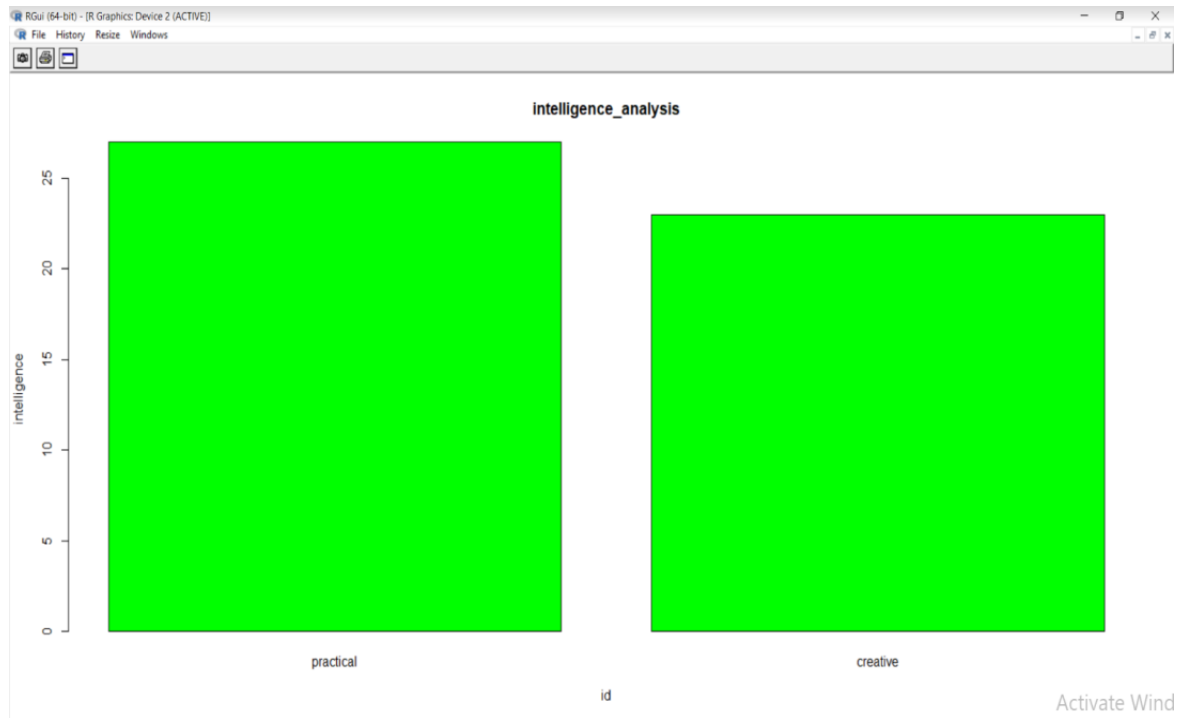
9.LIMITATIONS

Language Specific Limitations	Technological Limitations	Data Specific Limitations
Words have different meanings	Extraction of Entities	Noisy data
Language is a barrier	Cannot identify the root cause of the review	Videos and Images
Comparative and Complex statements	Entity level vs Article level Sentiment.	Large data sets and lots of spam
Scarsam statements. Ex:Irony	Human Accuracy is missing.	Not available for all domains
Short forms and many representations of the words.	Only three categories to categorize the reviews.	Fails to classify with respect to others prespective
Co–relationship between the sentences.	Focused on explicit opinion.	Knowledge of the data

10.CONCLUSION/DISCUSSION:

- It has access to large amounts of person behavioral data. Helps to classify people using Automated personality classification (APC).

- In this project, the system proposes an advanced APC – automated personality classification system.

- The system uses learning algorithms like Naive Bayes and SVM, Decision tree along with advanced data mining to mine user characteristics data and learn from the patterns.

11.REFERENCES

- [1]. Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and trends in information retrieval 2.1-2 (2008)

- [2]. Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on Vol 1 IEEE, 2001

- [3]. Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis. "IEEE Intelligent Systems 28.2(2013): 15-21

- [4]. Golbeck, Jennifer, Cristina Robles, and Karen Turner. "Predicting personality with social media." CHI'11 extended abstracts on human

factors in computing systems. ACM, 2011.

- [5]. Barrick, Murray R., and Michael K. Mount. "The Big Five personality dimensions and job performance: A meta-analysis." (1991).

- [6]. Judge, Timothy A., et al. "The big five personality traits, general mental ability, and career success across the life span." Personnel psychology 52.3 (1999): 621-652.

- [7]. Shaver, Phillip R., and Kelly A. Brennan. "Attachment styles and the" Big Five" personality traits: Their connections with each other and with romantic relationship outcomes." Personality and Social Psychology Bulletin 18.5 (1992): 536-545.

- [8]. Golbeck, Jennifer, et al. "Predicting personality from twitter." Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011.

12.RESULTS:
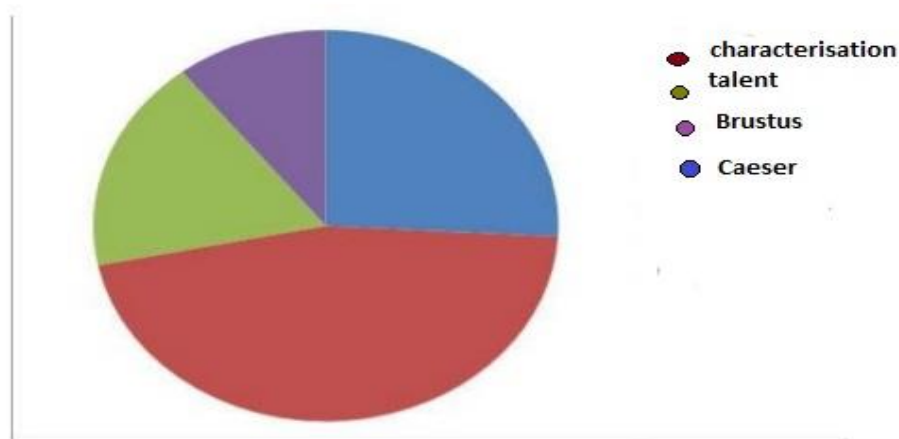
The outcome expected is a character of a Person. The user character is examined and which personality of the movie does the user resemble is the expected outcome.

13.COMPLETE CODE:

```java
package charec;

import java.util.*;
public class Charec
{

    public double count(String age[],String ms[],String intro[],String extro[],String c[],String pr[],String ch[],String td[],int n)
    {
        Scanner sc=new Scanner(System.in);
        String x,y,a,b,p,q,car;
        double y_cnt=0,m_cnt=0; // age count
        double ma_cnt=0,nma_cnt=0; //marriage status count
        double iry_cnt=0,irn_cnt=0; // introvert count
        double exy_cnt=0,exn_cnt=0; // extrovert count
        double cy_cnt=0,cn_cnt=0; // creative count
        double py_cnt=0,pn_cnt=0;int i; // practical count
        double cp_cnt=0,cc_cnt=0 ; // character count

        System.out.println("Enter the element to be counted in age : ");
        x=sc.next();
        System.out.println("Enter the element to be counted in marriage status : ");
        y=sc.next();
        System.out.println("Enter the element to be counted in Introvert : ");
        a=sc.next();
        System.out.println("Enter the element to be counted in Extrovert : ");
        b=sc.next();
        System.out.println("Enter the element to be counted in Creative : ");
        p= sc.next();
```

```java
p= sc.next();
System.out.println("Enter the element to be counted in Practical : ");
q=sc.next();
System.out.println("Enter the element to be counted in Character : ");
car=sc.next();
for(i=0;i<n;i++)
{
   if(age[i].equals(x))
   {
      y_cnt++;
   }

   if(ms[i].equals(y))
   {
      ma_cnt++;
   }

   if(intro[i].equals(a))
   {
      iry_cnt++;
   }
   if(extro[i].equals(b))
   {
      exy_cnt++;
   }
   if(c[i].equals(p))
   {
      cy_cnt++;
   }
```

```java
            if(intro[i].equals(a))
            {
                iry_cnt++;
            }
            if(extro[i].equals(b))
            {
                exy_cnt++;
            }           .
            if(c[i].equals(p))
            {
                cy_cnt++;
            }
            if(pr[i].equals(q))
            {
                py_cnt++;
            }
             if(ch[i].equals(car))
            {
                cp_cnt++;
            }

        }

    m_cnt=n-y_cnt;
    nma_cnt=n-ma_cnt;
    irn_cnt=n-iry_cnt;
    exn_cnt=n-exy_cnt;
    cn_cnt=n-cy_cnt;
```

```java
pn_cnt=n-py_cnt,
cc_cnt=n-cp_cnt;


System.out.println("AGE COUNT!!!");
System.out.println("Number of YOUTH in age: "+y_cnt);
System.out.println("Number of MID in age: "+m_cnt);
System.out.println("\n");
System.out.println("MARRIAGE STATUS COUNT!!!");
System.out.println("Number of M in Marriage Status: "+ma_cnt);
System.out.println("Number of NM in Marriage Status: "+nma_cnt);
System.out.println("\n");
System.out.println("INTROVERT COUNT!!!");
System.out.println("Number of y in introvert: "+iry_cnt);
System.out.println("Number of n in introvert: "+irn_cnt);
System.out.println("\n");
System.out.println("EXTROVERT COUNT!!!");
System.out.println("Number of y in extrovert: "+exy_cnt);
System.out.println("Number of n in extrovert: "+exn_cnt);
System.out.println("\n");
System.out.println("CREATIVE COUNT!!!");
System.out.println("Number of y in creative: "+cy_cnt);
System.out.println("Number of n in creative: "+cn_cnt);
System.out.println("\n");
System.out.println("PRACTICAL COUNT!!!");
System.out.println("Number of y in practical: "+py_cnt);
System.out.println("Number of n in practical: "+pn_cnt);
System.out.println("\n");
System.out.println("CHARACTER COUNT!!!");
```

```java
System.out.println("Number of C in character: "+cp_cnt);
System.out.println("Number of B in character: "+cc_cnt);



//Naive Bayes Calculation

double A,B,X,Y;
A=cp_cnt/n;
B=cc_cnt/n;


System.out.println("Probability of the population being julius caeser:"+A);
System.out.println("Probability of the population being marcus brutus:"+B);

// Calculation of test data with character julius caeser

double P1,P2,P3,P4,P5,P6;
if(td[0].equals("youth"))
{
    P1=y_cnt/cp_cnt;
}
else{P1=m_cnt/cp_cnt;}

if(td[1].equals("m"))
{
    P2=ma_cnt/cp_cnt;
}
else{P2=nma_cnt/cp_cnt;}
```

```java
}
else{P2=nma_cnt/cp_cnt;}

if(td[2].equals("y"))
{
   P3=iry_cnt/cp_cnt;
}
else{P3=irn_cnt/cp_cnt;}

if(td[3].equals("y"))
{
   P4=exy_cnt/cp_cnt;
}
else{P4=exn_cnt/cp_cnt;}

if(td[4].equals("y"))
{
   P5=cy_cnt/cp_cnt;
}
else{P5=cn_cnt/cp_cnt;}

if(td[5].equals("y"))
{
   P6=py_cnt/cp_cnt;
}
else{P6=pn_cnt/cp_cnt;}

//Calculation of test data with character as marcus brutus
```

```java
double C1,C2,C3,C4,C5,C6;
if(td[0].equals("youth"))
{
    C1=y_cnt/cc_cnt;
}
else{C1=m_cnt/cc_cnt;}

if(td[1].equals("m"))
{
    C2=ma_cnt/cc_cnt;
}
else{C2=nma_cnt/cc_cnt;}

if(td[2].equals("y"))
{
    C3=iry_cnt/cc_cnt;
}
else{C3=irn_cnt/cc_cnt;}

if(td[3].equals("y"))
{
    C4=exy_cnt/cc_cnt;
}
else{C4=exn_cnt/cc_cnt;}

if(td[4].equals("y"))
{
    C5=cy_cnt/cc_cnt;
}
```

```java
        else{C5=cn_cnt/cc_cnt;}

    if(td[5].equals("y"))
    {
        C6=py_cnt/cc_cnt;
    }
    else{C6=pn_cnt/cc_cnt;}

    //End calculation and comparison

    Y=C1*C2*C3*C4*C5*C6*A;
    X=P1*P2*P3*P4*P5*P6*B;

    System.out.println("X:" + X+ "Y:"+ Y);

    if(X<Y)
    {System.out.println("JULIUS CAESER!!!");}
    else{
        System.out.println("MARCUS BRUTUS!!!");
    }

    return 1;

}

public static void main(String args[])
{
```

```java
{
    int n,i,l=6;
    //String age;
    Scanner sc= new Scanner(System.in);
    System.out.println("Enter the number of records to be entered:");
    n=sc.nextInt();

    System.out.println("Way to enter the age:  10-30:youth age>30:mid ");
    System.out.println("Way to enter the status:  Married-M UnMarried-NM ");

    // Creating an array to store the values of age ,marriage status,introvert,extrovert,creative,practical,character !!!!
    String age[] = new String[n];
    String ms[]=new String[n];// marriage status
    String intro[]=new String[n];
    String extro[]=new String[n];
    String c[]=new String[n];
    String pr[]=new String[n];
    String ch[]=new String[n];
    String td[]=new String[50]; // TEST DATA

    for(i=0;i<n;i++)
    {


    System.out.println("Enter the age of person  "+(i+1)+" :");
    age[i]=sc.next();
    System.out.println("Enter the status of person  "+(i+1)+" :");
    ms[i]=sc.next();
age[i]=sc.next();
System.out.println("Enter the status of person  "+(i+1)+" :");
ms[i]=sc.next();
System.out.println("Introvert: Y/N?");
intro[i]=sc.next();
System.out.println("Extrovert: Y/N?");
extro[i]=sc.next();
System.out.println("Creative: Y/N?");
c[i]=sc.next();
System.out.println("Practical: Y/N?");
pr[i]=sc.next();
System.out.println("Character: Caeser/Brutus?");
ch[i]=sc.next();
System.out.println("\n");
}

System.out.println("Enter test data in the order- age,marriagestatus,introvert,extrovert,creative,practical");
for(i=0;i<l;i++)
{
    td[i]=sc.next();
}



for(i=0;i<n;i++)
{
    System.out.println("PERSON: "+(i+1));
    System.out.println("Age:"+age[i]);
```

```java
            System.out.println("Age:"+age[i]);
            System.out.println("Marriage Status:"+ms[i]);
            System.out.println("Introvert:"+intro[i]);
            System.out.println("Extrovert:"+extro[i]);
            System.out.println("Creative:"+c[i]);
            System.out.println("Practical:"+pr[i]);
            System.out.println("Character:"+ch[i]);

            System.out.println("\n");
        }


        System.out.println("TEST DATA!!!");

        for(i=0;i<6;i++)
        {
            System.out.println(td[i]);

        }



        Charec ca=new Charec();
        ca.count(age,ms,intro,extro,c,pr,ch,td,n);

    }

}
```

14.PROJECT POSTER:

Character analysis of a person

K.Hema Anmisha,Aparna.R,P.NityaSree | Nagaraj SV | SITE
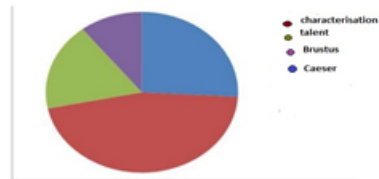
Motivation/Introduction

Results:

Personality identification of a human being by their nature an old technique. Earlier these were done manually by spending lot of time to predict the nature of the person. Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations.

The outcome expected is a character of a Person.The user character is examined and which personality of the movie does the user resemble is the expected outcome.

SCOPE of the Project

Social network analysis has increased tremendously in recent times. To extract the personality of the authors on the social networking websites is very useful for much application in various domain like including job success, attractiveness, marital and happiness. Personality detection from text means to extract the behavior characteristics of authors written the text.

Conclusion/Summary

It has access to large amounts of person behavioral data. Helps to classify people using Automated personality classification (APC). In this project, the system proposes an advanced APC – automated personality classification system. The system uses learning algorithms like Naive Bayes and SVM, Decision tree along with advanced data mining to mine user characteristics data and learn from the patterns.

Methodology

This project comes across areas where it has access to large amounts of person behavioral data. This data can be helpful to classify persons using Automated personality classification (APC). In this project, the system proposes an advanced APC – automated personality classification system. The system uses learning algorithms like Naive Bayes and SVM, Decision tree along with advanced data mining to mine user characteristics data and learn from the patterns. This learning can now be used to classify/predict user personality based on past classifications. The system analyses vast user characteristics and behaviors and based on the patterns observed, it stores its own user characteristics patterns in a database. The system now predicts new user personality based on personality data stored by classification of previous user data. This system is useful to social networks as well as various ad selling online networks to classify user personality and sell more relevant ads.

Contact :
paladugunitya@gmail.com
anmishaklvn@gmail.com
aparna99.ramesh@gmail.com

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]Yago Saez , Carlos Navarro , Asuncion Mochon and Pedro Isasi, A system for personality and happiness detection.
[4]Aleksandar Kartelj, Vladimir Filipović, Veljko Milutinović, Novel approaches to automated personality classification: Ideas and their potentials.