

Personality Prediction Using Decision Tree

- **Overview:** This application will predict the personality of an individual by asking the appropriate questions to the user. Personality is a complex and multifaceted construct, and accurately predicting it has a variety of practical applications, such as hiring and recruitment, marketing, and mental health diagnosis and treatment. Decision trees are a popular machine learning technique for classification problems, providing a transparent and interpretable model that can aid in determining which features are most important in predicting personality.
- **Why we need to predict personality?**
 - psychological research
 - clinical psychology
 - counseling
 - human resource management
 - marketing
- **Methodology:** The dataset used in the analysis will be described in the report, which may include personality assessment measures, demographic information, and other relevant variables. The decision tree algorithm used, as well as any pre-processing or feature engineering steps taken to prepare the data, will be described.
- **Data Set:**

	Q1A	Q1I	Q1E	Q2A	Q2I	Q2E	Q3A	Q3I	Q3E	Q4A	...	Q89E	Q90A	Q90I	Q90E	Q91A	Q91I	Q91E	gender	age	IE
0	5	51	7107	3	91	2522	1	56	6180	2	...	11548	3	40	4648	3	35	4609	2	23	3
1	5	39	6354	5	13	3092	1	12	5243	5	...	2827	4	28	3884	3	1	10409	1	25	2
2	3	17	5397	4	35	2747	5	40	5262	3	...	3337	1	87	1759	1	19	2691	1	19	1
3	5	41	3055	2	14	3348	1	13	5141	1	...	3198	3	15	2345	3	23	3697	1	23	1
4	1	76	2542	2	54	1878	1	15	5637	1	...	4024	5	86	6413	5	69	1662	1	18	2
...
7183	1	46	1328	4	82	3214	4	43	3360	5	...	3953	5	49	3608	4	62	3495	2	53	1
7184	2	5	8786	5	24	2233	5	10	10387	5	...	3483	3	76	2778	2	33	6088	1	20	1
7185	3	29	6618	5	44	2393	4	58	5768	5	...	3815	1	32	2345	1	21	3425	2	28	1
7186	4	15	8321	2	18	6179	5	60	5037	1	...	14867	5	78	1187	2	86	17416	2	19	1
7187	5	57	2950	2	66	2232	4	24	7095	4	...	2780	4	78	1843	4	40	1901	2	25	1

7188 rows × 276 columns

➤ Data Set Description:

The test contained 91 questions. The questions were presented one at a time in a random order. For each questions 3 values were recorded:

A - The user's selected response. 1=Disagree, 2=Slightly disagree, 3=Neutral, 4=Slightly agree, 5=Agree

I - The position of the question in the survey.

E - The time elapsed on that question in milliseconds.

The text of the questions were:

```
{
  "Q1" : "I would never audition to be on a game show.",
  "Q2" : "I am not much of a flirt.",
  "Q3" : "I have to psych myself up before I am brave enough to make a
phone call.",
  "Q4" : "I would hate living with room mates.",
  "Q5" : "I mostly listen to people in conversations.",
  "Q6" : "I reveal little about myself.",
  "Q7" : "I spend hours alone with my hobbies.",
  "Q8" : "I prefer to eat alone.",
  "Q9" : "I have trouble finding people I want to be friends with.",
  "Q10" : "I prefer to socialize 1 on 1, than with a group.",
  "Q11" : "I sometimes speak so quietly people sometimes have trouble
hearing me.",
  "Q12" : "I do not like to get my picture taken.",
  "Q13" : "I can keep a conversation going with anyone about anything.",
  "Q14" : "I want a huge social circle.",
  "Q15" : "I talk to people when waiting in lines.",
  "Q16" : "I act wild and crazy.",
  "Q17" : "I am a bundle of joy.",
  "Q18" : "I love excitement.",
  "Q19" : "I'd like to be in a parade.",
  "Q20" : "I am a flamboyant person.",
  "Q21" : "I am good at making impromptu speeches.",
  "Q22" : "I naturally emerge as a leader.",
  "Q23" : "I am spontaneous.",
  "Q24" : "I would enjoy being a sports team coach.",
  "Q25" : "I have a strong personality.",
  "Q26" : "I am excited by many different activities.",
  "Q27" : "I spend most of my time in fantasy worlds.",
  "Q28" : "I often feel lucky.",
  "Q29" : "I don't make eye contact when I talk with people.",
  "Q30" : "I have a monotone voice.",
  "Q31" : "I am a touchy feely person.",
  "Q32" : "I would like to try bungee jumping.",
  "Q33" : "I tend to be admired by others.",
  "Q34" : "I make big physical movements whenever I get excited.",
  "Q35" : "I am brave.",
  "Q36" : "I am always in the moment.",
  "Q37" : "I am involved with my community.",
  "Q38" : "I am good an entertaining children.",
  "Q39" : "I like formal occasions.",
  "Q40" : "I would have to be lost for a very long time before asking
help.",
```

```

"Q41" : "I do not care about sports.",
"Q42" : "I prefer individual sports to team sports.",
"Q43" : "My parents know nothing about my love life.",
"Q44" : "I mostly listen to people in conversations.",
"Q45" : "I never leave the door to my room open.",
"Q46" : "I make a lot of hand motions when I talk.",
"Q47" : "I take lots of pictures of my activities.",
"Q48" : "When I was a child, I put on fake concerts and plays with my
friends.",
"Q49" : "I really like dancing.",
"Q50" : "I would have difficulty describing myself to someone.",
"Q51" : "My life would not make a good story.",
"Q52" : "I am hesitant to give suggestions.",
"Q53" : "I tire out quickly.",
"Q54" : "I never tell people the important things about myself.",
"Q55" : "I avoid going to unknown places.",
"Q56" : "Going to the doctor is always awkward for me.",
"Q57" : "I have not kept up with my old friends over the years.",
"Q58" : "I have not been joyful for quite some time.",
"Q59" : "I hate to ask for help.",
"Q60" : "If I were to die, I would not want there to be a memorial for
me.",
"Q61" : "I hate shopping.",
"Q62" : "I love to do impressions.",
"Q63" : "I would be pleased if asked to speak at a funeral.",
"Q64" : "I would never go to a dance club.",
"Q65" : "I find it very hard to tell people I find them attractive.",
"Q66" : "I hate people.",
"Q67" : "I was an outcast in school.",
"Q68" : "I would enjoy being a librarian.",
"Q69" : "I am usually not single.",
"Q70" : "I am able to stand up for myself.",
"Q71" : "I would go surfing regularly if I lived on a beach.",
"Q72" : "I have wanted to be a stand-up comedian.",
"Q73" : "I am a high status person.",
"Q74" : "I work out regularly.",
"Q75" : "I laugh a lot.",
"Q76" : "I like pranks.",
"Q77" : "I am happy with my life.",
"Q78" : "I am never at a loss for words.",
"Q79" : "I feel healthy and vibrant most of the time.",
"Q80" : "I love large parties.",
"Q81" : "I am quiet around strangers.",
"Q82" : "I don't talk a lot.",
"Q83" : "I keep in the background.",
"Q84" : "I don't like to draw attention to myself.",
"Q85" : "I have little to say.",
"Q86" : "I often feel blue.",
"Q87" : "I am not really interested in others.",
"Q88" : "I make people feel at ease.",
"Q89" : "I don't mind being the center of attention.",
"Q90" : "I start conversations.",
"Q91" : "I talk to a lot of different people at parties."
}

```

After the main question sequence, the following questions were asked on one final page:

```

gender      "What is your gender?"      1=Male 2=Female 3=Other
engnat      "Is English your native language?" 1=Yes 2=No
age         "What is your age in years?"
IE          "Do you identify as either an introvert or extravert?"
           1=Yes, introvert 2=Yes, extravert 3=No

```

On the final page, the users were also asked "Do you give accurate answers and can we store and use your data for research?". Only those who answered yes were recorded.

The following were determined from technical information:

```

country      user's network location
dateload    the time the user loaded the introduction page
introelapse the time spent in seconds on the introduction page
testelapse  the time spent in seconds on the test questions
surveyelapse the time spent in seconds on the final page

```

➤ **What is decision tree?**

- non-parametric supervised learning algorithm,
- hierarchical and tree structure
- it consists of a root node, branches, internal nodes, and leaf nodes.

➤ **Why is decision tree used here?**

- useful in situations where the data is complex and difficult to analyze using traditional statistical methods.
- model is easy to understand and can be visualized as a tree.
- easy to implement and can handle both categorical and continuous data.

➤ **How is decision tree applied?**

- **Input Used:**
 - Numerical Value: For scaling level of answer for each prompt 91 questions
 - 91 questions: Are features to decide personality type.
- **Output Obtained:**
 - Personality type is displayed stating introvert, extrovert, or ambivert.
- **Logic Used:**
 - **Gini index value:** Used for each feature question to determine whether user is of a particular personality type.

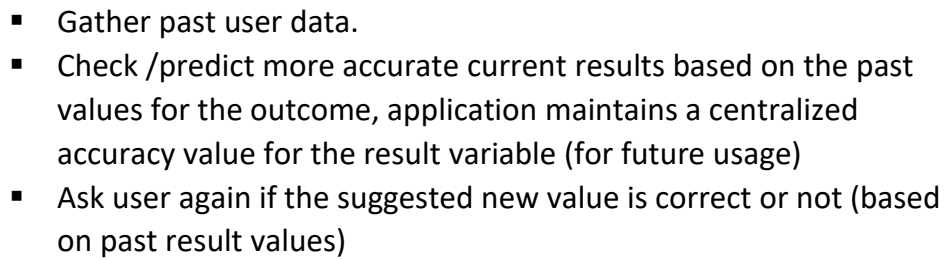
➤ **Application**

- Can be used in schools/universities while selecting subjects.
- Can be used by recruiting firms for taking personality tests.

- **Different parts of Future application:**

- **Data Source:** MongoDB (No SQL Database) Predefined values for feature questions for decision tree.
- **Python Django:** Framework to connect knowledgebase with user interface.
- **User Interface:** Web page which asks user the feature questions (page is designed in HTML and CSS)

- **Application Cycle:**



- **Results:** The findings of the analysis will be presented in the results section, including the decision tree model's accuracy in predicting personality traits as well as any insights gained from the feature importance rankings. To aid in the interpretation of the results.
- **Discussion:** The discussion section will place the findings in the context of the larger literature on personality prediction and decision tree modelling. The study's limitations will be discussed, as will potential future research directions. The findings' practical implications, particularly their relevance to real-world applications such as hiring and mental health treatment, will also be discussed.
- **Conclusion:** The report's conclusion will include a summary of the key findings and their implications, as well as any recommendations for future research or practical applications. Overall, the report seeks to provide a comprehensive overview of the use of decision trees in personality prediction, as well as its potential to contribute to our understanding of this complex and important construct.