

**HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**SCHOL OF INFORMATION TECHNOLOGY AND COMMUNICATION**

**-----🙞🙜🕮🙞🙜-----**



**DSS PROJECT REPORT**

**Project name**: Career counseling system for students

at Hust

**Supervisor Assoc. Prof. Phạm Văn Hải**

**Student names :** **Group 8**

|  |  |
| --- | --- |
| **Nguyễn Đức Minh** | **20176042** |
| **Nguyễn Trí Vinh** | **20176066** |
| **Hoàng Ngọc Nhất** | **20176048** |

***Hà Nôi, 1 / 2022***

***Hà Nội, tháng 10 năm 2012***

Table of Contents

[**I.** **Reality** 4](#_Toc94281887)

[**II.** **Design analysis** 4](#_Toc94281888)

[**III.** **Find out about the decision tree** 5](#_Toc94281889)

[1. Definitions 5](#_Toc94281890)

[2. Deciding tree types 5](#_Toc94281891)

[a. Regression tree 5](#_Toc94281892)

[b. Classification tree 5](#_Toc94281893)

[3. Pros/Cons of decision tree algorithm 5](#_Toc94281894)

[Advantage: 5](#_Toc94281895)

[The decision tree is a simple and popular algorithm. This algorithm is widely used because of its benefits: 5](#_Toc94281896)

[Disadvantage: 6](#_Toc94281897)

[Along with that, the decision tree also has specific disadvantages: 6](#_Toc94281898)

[         The tree model decides to depend greatly on your data. Even with a small change in the dataset, the decisive tree model structure can change completely. 6](#_Toc94281899)

[ The tree decides or has problems. overfitting 6](#_Toc94281900)

[**Tree algorithm that determines ID3 or C4.5** 6](#_Toc94281901)

[**1.** **Decision tree algorithm** 6](#_Toc94281902)

[**2.** **Construction of a decision tree consists of 2 steps:** 7](#_Toc94281903)

[**3.** **Evaluation of decision tree algorithms in the field of data mining** 7](#_Toc94281904)

[**4.** **General Problem** 8](#_Toc94281905)

[**5.** **Specific examples** 9](#_Toc94281906)

[**Data table** 9](#_Toc94281907)

[**Step 1: Consider the properties:** 12](#_Toc94281908)

[a. Considering industry attributes: 12](#_Toc94281909)

[c. Consider preference attributes 13](#_Toc94281910)

[d. Considering gifted attributes 13](#_Toc94281911)

[e. Considering soft skills attributes 13](#_Toc94281912)

[f. Consider language properties 14](#_Toc94281913)

[Step 2: Get the IG of the largest attribute 14](#_Toc94281914)

[a. Considering artificial intelligence 14](#_Toc94281915)

[b. Reviewing web programming 16](#_Toc94281916)

[c. Consider game programming 17](#_Toc94281917)

[e. Consider Cybersecurity Programmers 19](#_Toc94281918)

[f. Consider Phone App Programming 21](#_Toc94281919)

[g. Game Design Review 22](#_Toc94281920)

[h. Tester review 23](#_Toc94281921)

[i. Consider Bridge Engineer 24](#_Toc94281922)

[Step 3: Synthesize into a decision tree 25](#_Toc94281923)

[**IV.** **References** 25](#_Toc94281924)

1. **Reality**

Choosing a career is an important task, dominating the majority of the thinking and activities of university students, but in the process of learning as well as practical work experience is still limited. Most universities have the mindset that after finishing school, they will start looking for work, so the process of preparing information and knowledge for the selection of suitable professions is not good. On the other hand, the constantly changing annual job demands, the lack of active learning as well as balancing self-competence with the job you desire can lead to inappropriate industry thinking. When students have different concerns about the suitability of the profession, about the labor market, the ability of the chosen profession in the future ... There is almost no time to ask, or not to be properly explained. Therefore, there is a shortage of teachers and every year there are many graduates in many different professions but they are very difficult to participate in the labor market, partly because students do not meet the requirements of the profession, partly because their chosen profession is not in the needs of the labor market.

1. **Design analysis**

* Software: use the J48 algorithm as the main algorithm for the data processing process.
* Build a sample dataset for training for software
* Users enter information to help the software decide and learn from user feedback

\* Input:

* Criteria:

+ Learning

+ Hobbies: gaming, watching movies, surfing the web, reading, science

+ Aptitude: Drawing, singing

+ Soft skills: communication, presentation, leadership, design and

administration

+ Programming language: Java, C, JavaScript, Python

\* Output:

Suitable job

1. **Find out about the decision tree**

## Definitions

In the field of machine learning,the decision tree is a predictive*model,*that is, a mapping from observations of an object/phenomenon to conclusions about the target value of things/phenomena. Each*internal node*corresponds to a variable; The seam between it and its sub-button represents a specific value for that variable. Each leaf node represents the predicted value of the target variable, given the values of the variables represented by the path from the root button to that leaf node. The machine learning technique used in the decision tree is called decision learning, or shortly called the decision tree.

Learning by the decision tree is also a common method of data mining. At that time, the tree decides to describe a tree structure, in which the leaves represent the classifications and the branches represent the combinations of properties that lead to that classification. A decision tree can be learned by dividing the source set into subsets based on a test of attribute values. This process is recursively for each derivative subset. Recursion is completed when the separation cannot be continued, or when a single classification can be applied to each element of the derivative subset. A*random forest*classification uses a number of deciding trees to improve the rate of classification.

The decision tree is also a descriptive means for calculating conditional probabilities.

The decision tree can be described as a combination of mathematical and computational techniques that support the description, classification, and generalization of a given dataset.

## Deciding tree types

### Regression tree

### Classification tree

# Pros/Cons of decision tree algorithm

# Advantage:

# The decision tree is a simple and popular algorithm. This algorithm is widely used because of its benefits:

-         The model produces easy-to-understand rules for the reader, creating a code with each branch of leaves being a law of the tree.

-         Input data can be missing data, without standardizing or creating fake variables

-         Can work with both digital and classification data

-         It is possible to authenticate the model using statistical tests

-         It's likely to be about big data.

# Disadvantage:

# Along with that, the decision tree also has specific disadvantages:

#          The tree model decides to depend greatly on your data. Even with a small change in the dataset, the decisive tree model structure can change completely.

#  The tree decides or has problems. [overfitting](https://trituenhantao.io/tu-dien-thuat-ngu/overfitting/" \t "_blank)

**Tree algorithm that determines ID3 or C4.5**

1. **Decision tree algorithm**

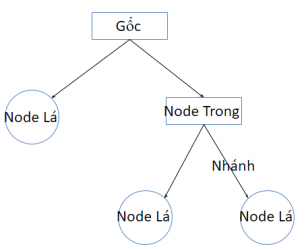
The tree algorithm that determines the result is a set of laws of training data with attributes. The decision tree is a common tool in data mining and layering. Characteristics of the deciding tree: is a structured tree in which:

Root is the top button of the tree.

Inner node: intermediate button on a single properties (Oval shape).

Branch: Show the results of the test on the button

Leaf node: Class or layer distribution (square or rectangular)



1. **Construction of a decision tree consists of 2 steps:**

**Decisive tree development:**go from the root, to the branches, grow inductive in the form of division to rule.

-         **Step 1.**Select the "best" attribute with a pre-specified measurement

-         **Step 2.**Grow trees by adding branches that correspond to each value of the selected properties

-         **Step 3.**Sort and divide training datasets to sub-nodes

-         **Step 4.**If the examples are clearly layered, stop.

**On the contrary:**repeat step 1 to step 4 for each sub-node

**Pruning trees:**to simplify, generalize trees, increase accuracy

1. **Evaluation of decision tree algorithms in the field of data mining**

**Advantageous:**

         The tree building process decides not to use knowledge of the field of data being studied or what input parameters.

         The results of the training process (learning) are represented in the form of trees so easy to understand and close to humans.

         In general, the deciding tree algorithms give fairly high accuracy.

**Difficult:**

         For data sets with many attributes, the deciding tree will be large (in depth both horizontally), thus reducing the ease of understanding.

         The ranking of properties for branching is based on the previous branching and ignores the interdependence between the properties.

         When information gain is used to determine forked properties, properties with multiple values are often preferred.

|  |  |
| --- | --- |
| Pseudocode:  · Check the basic case.  · For each A attribute, information is found by separating properties A  · Select a\_best is the attribute by which the measurement selects the "best" attribute  · Use a\_best as a attribute for the tree-splitting node.  · Recursion on sub-lists generated by division by a\_best, and adds these nodes as sub-nodes | (1)ComputerClassFrequency(T); (2) if OneClass or FewCases return a leaf; Create a decision node N; (3) ForEach Attribute A ComputeGain(A); (4)N.test=AttributeWithBestGain; (5) if (N.test is continuous) find Threshold; (6) ForEach T’ in the splitting of T (7) If ( T’ is Empty ) Child of N is a leaf else (8) Child of N=FormTree(T’); (9) ComputeErrors of N; return N |

1. **General Problem**

The two measurements used in C4.5 are information gain and gain ratio. RF(Cj,S) performs relative frequency of cases in S that belong to the Cj class.

R

With | Sj| is the size of the set of cases with a class value of CJ. | S| Is the size of the training data set.

The information required for classing: I(S) with S is the set that needs to consider class distribution as:

* Measurement of the selection of "best" properties:

After S is divided into subsets S1, S2,..., St by test B, information gain

is calculated as:

Information gain: Test B chia S={S1,S2,...,St)

* Test B will be selected if G(S, B) reaches the greatest value.

However, there is a problem when using G(S,B) priority tests with a large number of results, for example G(S,B) peaking with a test where each Si contains only one single case. The gain ratio standard solves this problem by including its own potential information.

*Your potential information is available for each plans:*

Gain ratio = G(S, B) / P(S, B) largest => chọn test B

In the C4.5 classification model, one of the two types of Information

Gain or Gain ratios can be used to determine the best attribute. Gain ratio is the default option.

1. **Specific examples**

**APPLY THE DECISION TREE ALGORITHM**

*In the problem of supporting career decision making for students*

**Data table**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 1 | Trí tuệ nhân tạo | G | Khoa học | Không | Phân tích | Python | Yes |
| 2 | Trí tuệ nhân tạo | G | Không | Không | Thiết kế | Python | Yes |
| 3 | Trí tuệ nhân tạo | K | Khoa học | Không | Thiết kế | Python | No |
| 4 | Trí tuệ nhân tạo | TB | Không | Không | Thuyết trình | Python | No |
| 5 | Lập trình web | G | Lướt Web | Có | Thiết kế | Javascript | Yes |
| 6 | Lập trình web | K | Lướt Web | Có | Thiết kế | Javascript | Yes |
| 7 | Lập trình web | TB | Không | Không | Thiết kế | Javascript | No |
| 8 | Lập trình game | G | Chơi Game | Không | Thiết kế | Java | Yes |
| 9 | Lập trình game | K | Chơi Game | Không | Thiết kế | Java | Yes |
| 10 | Lập trình game | TB | Lướt web | Không | Thiết kế | Java | No |
| 11 | Tư vấn viên | K | Lướt Web | Không | Thuyết trình | C | Yes |
| 12 | Tư vấn viên | TB | Lướt Web | Không | Giao tiếp | C | Yes |
| 13 | Lập trình viên an ninh mạng | G | Khoa học | Có | Thiết Kế | C | Yes |
| 14 | Lập trình viên an ninh mạng | K | Khoa học | Có | Thiết kê | C | No |
| 15 | Lập trình viên an ninh mạng | TB | Khoa học | Có | Không | C | No |
| 16 | Lập trình ứng dụng điện thoại | G | Không | Không | Thuyết trình | Javascript | No |
| 17 | Lập trình ứng dụng điện thoại | K | Khoa học | Có | Phân tích | Javascript | Yes |
| 18 | Lập trình ứng dụng điện thoại | TB | Khoa học | Không | Không | C | No |
| 19 | Thiết kế Game | G | Chơi game | Có | Không | Java | Yes |
| 20 | Thiết kế game | K | Chơi game | có | Thiết kế | C | Yes |
| 21 | Thiết kế Game | TB | Chơi game | Có | Phân tích | C | Yes |
| 22 | Tester | G | Không | Không | Phân tích | Python | No |
| 23 | Tester | K | Không | Không | Không | Python | No |
| 24 | Tester | TB | Chơi game | không | không | C | No |
| 25 | Kỹ sư cầu nối | G | Không | Không | Giao tiếp | Java | No |
| 26 | Kỹ sư cầu nối | K | Lướt web | Có | Giao tiếp | C | Yes |
| 27 | Kỹ sư cầu nối | TB | Lướt web | Có | Giao tiếp | Java | Yes |

**Algorithm:**

Set Yes: Y (with 15 results)

Set No: N (with 12 results)

We have entropy(D) = - = 0.991

**Step 1: Consider the properties:**

1. Considering industry attributes:

|  |  |  |  |
| --- | --- | --- | --- |
| **Ngành nghề** | **Yi** | **Ni** | **I(Yi,Ni)** |
| Trí tuệ nhân tạo | 2 | 2 | 1 |
| Lập trình web | 2 | 1 | 0,918 |
| Lập trình game | 2 | 1 | 1 |
| Tư vấn viên | 2 | 0 | 0 |
| Lập trình viên an ninh mạng | 1 | 2 | 0,918 |
| Lập trình ứng dụng điện thoại | 1 | 2 | 0,918 |
| Thiết kế Game | 3 | 0 | 0 |
| Tester | 0 | 3 | 0 |
| Kỹ sư cầu nối | 2 | 1 | 0,918 |

Remainder (Ngành nghề) D = = 0,667

* IG(Ngành nghề) = Entropy(D) - Remainder (Ngành nghề) D = 0.991– 0.533 = **0.324**

1. Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
| **Học lực** | **Yi** | **Ni** | **I(Yi,Ni)** |
| G | 6 | 3 | 0,918 |
| K | 6 | 3 | 0,918 |
| TB | 3 | 6 | 0,918 |

Remainder (Học lực) D = = 0,918

* IG(Ngành nghề) = Entropy(D) - Remainder (Học lực) D = 0.991– 0.918= 0.**073**

1. Consider preference attributes

|  |  |  |  |
| --- | --- | --- | --- |
| **Sở thích** | **Yi** | **Ni** | **I(Yi,Ni)** |
| Khoa học | **3** | **4** | **0,985** |
| Không | **1** | **6** | **0,592** |
| Lướt web | 6 | 1 | 0,592 |
| Chơi game | 5 | 1 | 0,65 |

Remainder (Sở thích) D = = 0,707

* IG(Sở thích) = Entropy(D) - Remainder (Sở thích) D = 0.991– 0.707 = 0.**284**

1. Considering gifted attributes

|  |  |  |  |
| --- | --- | --- | --- |
| Năng khiếu | Yi | Ni | I(Yi,Ni) |
| Có | 9 | 2 | 0,684 |
| Không | 6 | 10 | 0,954 |

Remainder (Năng khiếu) D = = 0,844

* IG(Năng khiếu) = Entropy(D) - Remainder (Năng khiếu) D = 0.991– 0.844= **0.147**

1. Considering soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
| Kĩ năng | Yi | Ni | I(Yi,Ni) |
| Phân tích | 3 | 1 | 0,811 |
| Thiết kế | 7 | 4 | 0,946 |
| Thuyết trình | 1 | 2 | 0,918 |
| Không | 1 | 4 | 0,722 |
| Giao tiếp | 3 | 1 | 0,811 |

Remainder (Kĩ năng) D = = 0,861

* IG(Kĩ năng) = Entropy(D) - Remainder (Kĩ năng) D = 0.991– 0.861= **0.13**

1. Consider language properties

|  |  |  |  |
| --- | --- | --- | --- |
| Ngôn ngữ | Yi | Ni | I(Yi,Ni) |
| Python | 2 | 4 | 0,918 |
| Javascript | 3 | 2 | 0.971 |
| Java | 4 | 2 | 0,918 |
| C | 6 | 4 | 0,971 |

Remainder (Ngôn ngữ) D = = 0,948

* IG(ngôn ngữ) = Entropy(D) - Remainder (Kĩ năng) D = 0.991– 0.948= **0.043**

## Step 2: Get the IG of the largest attribute

We have: IG= 0.324 is the largest

There are 9 values in all professions, so we will consider each case.

1. Considering artificial intelligence

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 1 | Trí tuệ nhân tạo | G | Khoa học | Không | Phân tích | Python | Yes |
| 2 | Trí tuệ nhân tạo | G | Không | Không | Thiết kế | Python | Yes |
| 3 | Trí tuệ nhân tạo | K | Khoa học | Không | Thiết kế | Python | No |
| 4 | Trí tuệ nhân tạo | TB | Không | Không | Thuyết trình | Python | No |

We have Entropi(D1) = - =1

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 2 | 2 | 1 |
| K | 0 | 1 | 0 |
| TB | 0 | 1 | 0 |

Remainder (Học lực) D1 = = 0,148

* IG(Học lực) = Entropy(D1) - Remainder (Học lực) D1 = 1– 0.148= **0.852**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Khoa học | 1 | 1 | 1 |
| Không | 1 | 1 | 1 |

Remainder (Sở thích) D1 = = 0,148

* IG(sở thích) = Entropy(D1) - Remainder (sở thích) D1 = 1– 0.148 = **0.852**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Không | 2 | 2 | 1 |

Remainder (Năng khiếu) D1 = = 0,148

* IG(Năng khiếu) = Entropy(D1) - Remainder (Năng khiếu) D1 = 1– 0.148 = **0.852**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Phân tích | 1 | 0 | 0 |
| Thiết kế | 1 | 1 | 1 |
| Không | 0 | 1 | 0 |

Remainder (Kĩ năng) D1 = = 0,074

* IG(Kĩ năng)D1 = Entropy(D1) - Remainder (Kỹ năng) D1 = 1-0,074 = **0.926**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Python | 2 | 2 | 1 |

Remainder (Ngôn ngữ) D1 = = 0,148

* IG(Ngôn ngữ)D1 = Entropy(D1) - Remainder (Ngôn ngữ) D1 = 1-0,148 = **0.852**

1. Reviewing web programming

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 5 | Lập trình web | G | Lướt Web | Có | Thiết kế | Javascript | Yes |
| 6 | Lập trình web | K | Lướt Web | Có | Thiết kế | Javascript | Yes |
| 7 | Lập trình web | TB | Không | Không | Thiết kế | Javascript | No |

We have Entropi(D2) = - =0,918

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 1 | 0 | 0 |
| K | 1 | 0 | 0 |
| TB | 0 | 1 | 0 |

Remainder (Học lực) D2= = 0

* IG(Học lực)D2 = Entropy(D2) - Remainder (Học lực) D2 = 0,918– 0= **0.918**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Lướt web | 2 | 0 | 0 |
| Không | 0 | 1 | 0 |

Remainder (Sở thích) D2 = = 0

* IG(sở thích)D2 = Entropy(D2) - Remainder (sở thích) D2 = 0,918 = **0.918**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Không | 0 | 1 | 0 |
| có | 2 | 0 | 0 |

Remainder (Năng khiếu) D2 = = 0

* IG(Năng khiếu)D2 = Entropy(D2) - Remainder (Năng khiếu) D2 = 0,918– 0 = **0.918**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Thiết kế | 2 | 1 | 0,918 |

Remainder (Kĩ năng) D2 = = 0,102

* IG(Kĩ năng)D2 = Entropy(D2) - Remainder (Kỹ năng) D2 = 0,918-0,102 = **0.816**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Javascript | 2 | 1 | 0,918 |

Remainder (Ngôn ngữ) D2 = = 0,102

* IG(Ngôn ngữ)D2 = Entropy(D2) - Remainder (Ngôn ngữ) D2 = 0,918-0,102 = **0.816**

1. Consider game programming

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 8 | Lập trình game | G | Chơi Game | Không | Thiết kế | Java | Yes |
| 9 | Lập trình game | K | Chơi Game | Không | Thiết kế | Java | Yes |
| 10 | Lập trình game | TB | Lướt web | Không | Thiết kế | Java | No |

We have Entropi(D3) = - =0,918

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 1 | 0 | 0 |
| K | 1 | 0 | 0 |
| TB | 0 | 1 | 0 |

Remainder (Học lực) D3= = 0

* IG(Học lực)D3 = Entropy(D3) - Remainder (Học lực) D3 = 0,918– 0= **0.918**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Lướt web | 0 | 1 | 0 |
| Chơi game | 2 | 0 | 0 |

Remainder (Sở thích) D3 = = 0

* IG(sở thích)D3 = Entropy(D3) - Remainder (sở thích) D3 = 0,918 = **0.918**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Không | 2 | 1 | 0,918 |

Remainder (Năng khiếu) D3 = = 0,102

* IG(Năng khiếu)D3 = Entropy(D3) - Remainder (Năng khiếu) D3 = 0,918– 0,102 = **0.816**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Thiết kế | 2 | 1 | 0,918 |

Remainder (Kĩ năng) D3 = = 0,102

* IG(Kĩ năng)D3 = Entropy(D3) - Remainder (Kỹ năng) D3 = 0,918-0,102 = **0.816**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Java | 2 | 1 | 0,918 |

Remainder (Ngôn ngữ) D3 = = 0,102

* IG(Ngôn ngữ)D3 = Entropy(D3) - Remainder (Ngôn ngữ) D3 = 0,918-0,102 = **0.816**

1. Consulting industry

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 11 | Tư vấn viên | K | Lướt Web | Không | Thuyết trình | C | Yes |
| 12 | Tư vấn viên | TB | Lướt Web | Không | Giao tiếp | C | Yes |

We have Entropi(D4) = =0

1. Consider Cybersecurity Programmers

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 13 | Lập trình viên an ninh mạng | G | Khoa học | Có | Thiết Kế | C | Yes |
| 14 | Lập trình viên an ninh mạng | K | Khoa học | Có | Thiết kê | C | No |
| 15 | Lập trình viên an ninh mạng | TB | Khoa học | Có | Không | C | No |

We have Entropi(D5) = - =0,918

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 1 | 0 | 0 |
| K | 0 | 1 | 0 |
| TB | 0 | 1 | 0 |

Remainder (Học lực) D5= = 0

* IG(Học lực)D5 = Entropy(D5) - Remainder (Học lực) D5 = 0,918– 0= **0.918**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Khoa Học | 1 | 2 | 0,918 |

Remainder (Sở thích) D5 = = 0,102

* IG(sở thích)D5 = Entropy(D5) - Remainder (sở thích) D5 = 0,918 – 0,102 = **0.816**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| có | 1 | 2 | 0,918 |

Remainder (Năng khiếu) D5 = = 0,102

* IG(Năng khiếu)D5 = Entropy(D5) - Remainder (Năng khiếu) D5= 0,918– 0,102 = **0.816**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Thiết kế | 1 | 1 | 1 |
| Không | 0 | 1 | 0 |

Remainder (Kĩ năng) D5 = = 0,074

* IG(Kĩ năng)D5 = Entropy(D5) - Remainder (Kỹ năng) D5 = 0,918-0,074 = **0.844**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| C | 1 | 2 | 0,918 |

Remainder (Ngôn ngữ) D5 = = 0,102

* IG(Ngôn ngữ)D5 = Entropy(D5) - Remainder (Ngôn ngữ) D5 = 0,918-0,102 = **0.816**

1. Consider Phone App Programming

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 16 | Lập trình ứng dụng điện thoại | G | Không | Không | Thuyết trình | Javascript | No |
| 17 | Lập trình ứng dụng điện thoại | K | Khoa học | Có | Phân tích | Javascript | Yes |
| 18 | Lập trình ứng dụng điện thoại | TB | Khoa học | Không | Không | C | No |

We have Entropi(D6) = - =0,918

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 0 | 1 | 0 |
| K | 1 | 0 | 0 |
| TB | 0 | 1 | 0 |

Remainder (Học lực) D6= = 0

* IG(Học lực)D6 = Entropy(D6) - Remainder (Học lực) D6 = 0,918– 0= **0.918**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Khoa Học | 1 | 1 | 1 |
| Không | 0 | 1 | 0 |

Remainder (Sở thích) D6 = = 0,074

* IG(sở thích)D6 = Entropy(D6) - Remainder (sở thích) D6 = 0,918 – 0,074 = **0.844**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| có | 1 | 0 | 0 |
| Không | 0 | 2 | 0 |

Remainder (Năng khiếu) D6 = = 0

* IG(Năng khiếu)D6 = Entropy(D6) - Remainder (Năng khiếu) D6= 0,918– 0 = **0.918**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Thuyết trình | 0 | 1 | 0 |
| Không | 0 | 1 | 0 |
| Phân tích | 1 | 0 | 0 |

Remainder (Kĩ năng) D6 = 0

* IG(Kĩ năng)D6 = Entropy(D6) - Remainder (Kỹ năng) D6 = 0,918-0 = **0.918**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| C | 0 | 1 | 0 |
| Javascript | 1 | 1 | 1 |

Remainder (Ngôn ngữ) D6= = 0,074

* IG(Ngôn ngữ)D6 = Entropy(D6) - Remainder (Ngôn ngữ) D6 = 0,918-0,074 = **0.844**

1. Game Design Review

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 19 | Thiết kế Game | G | Chơi game | Có | Không | Java | Yes |
| 20 | Thiết kế game | K | Chơi game | có | Thiết kế | C | Yes |
| 21 | Thiết kế Game | TB | Chơi game | Có | Phân tích | C | Yes |

We have Entropi(D7) = =0

1. Tester review

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 22 | Tester | G | Không | Không | Phân tích | Python | No |
| 23 | Tester | K | Không | Không | Không | Python | No |
| 24 | Tester | TB | Chơi game | không | không | C | No |

We have Entropi(D8) = =0

1. Consider Bridge Engineer

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Ngành nghề | Học lực | Sở thích | Năng khiếu | Kỹ năng mềm | Ngôn ngữ lập trình | SPDecision |
| 25 | Kỹ sư cầu nối | G | Không | Không | Giao tiếp | Java | No |
| 26 | Kỹ sư cầu nối | K | Lướt web | Có | Giao tiếp | C | Yes |
| 27 | Kỹ sư cầu nối | TB | Lướt web | Có | Giao tiếp | Java | Yes |

We have Entropi(D9) = - =0,918

* Considering academic attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| G | 0 | 1 | 0 |
| K | 1 | 0 | 0 |
| TB | 1 | 0 | 0 |

Remainder (Học lực) D9= = 0

* IG(Học lực)D9 = Entropy(D9) - Remainder (Học lực) D9 = 0,918– 0= **0.918**
* Consider preference properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Lướt web | 2 | 0 | 0 |
| Không | 0 | 1 | 0 |

Remainder (Sở thích) D9 = 0

* IG(sở thích)D9 = Entropy(D9) - Remainder (sở thích) D9 = 0,918 – 0= **0.918**
* Consider gifted properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| có | 2 | 0 | 0 |
| Không | 0 | 1 | 0 |

Remainder (Năng khiếu) D9 = = 0

* IG(Năng khiếu)D9 = Entropy(D9) - Remainder (Năng khiếu) D9= 0,918– 0 = **0.918**
* Consider soft skills attributes

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| Phân tích | 2 | 1 | 0,918 |

Remainder (Kĩ năng) D9 = =0,102

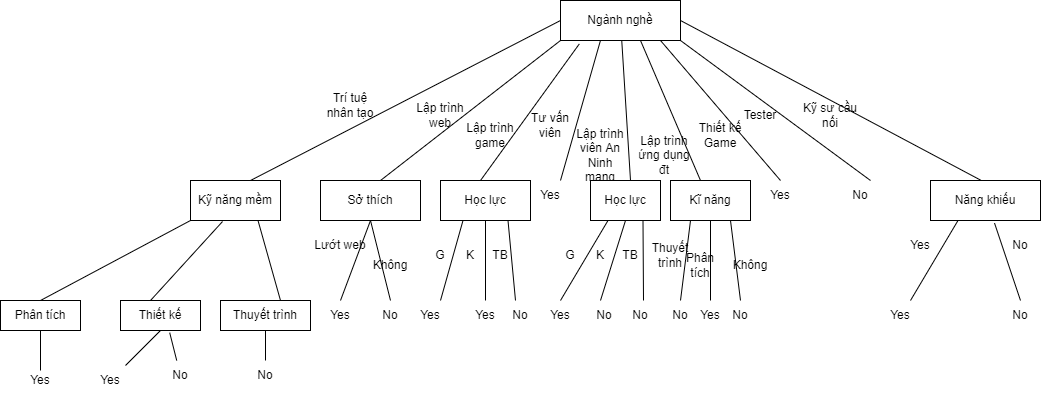
* IG(Kĩ năng)D9= Entropy(D9) - Remainder (Kỹ năng) D9 = 0,918-0,102 = **0.816**
* Consider programming language properties

|  |  |  |  |
| --- | --- | --- | --- |
|  | Yi | Ni | I(Yi,Ni) |
| C | 1 | 0 | 0 |
| Java | 1 | 1 | 1 |

Remainder (Ngôn ngữ) D9= = 0,074

* IG(Ngôn ngữ)D9 = Entropy(D9) - Remainder (Ngôn ngữ) D9 = 0,918-0,074 = **0.844**

## Step 3: Synthesize into a decision tree

After synthesizing and removing the most valuable properties, we synthesize the tree to decide: 

1. **References**

**1.**[**https://trituenhantao.io/kien-thuc/decision-tree/**](https://trituenhantao.io/kien-thuc/decision-tree/)

**2.** [**https://sinhvientot.net/thuat-toan-cay-quyet-dinh-c45/**](https://sinhvientot.net/thuat-toan-cay-quyet-dinh-c45/)

**3.** [**https://www.programmersought.com/article/55246223413/**](https://www.programmersought.com/article/55246223413/)

**4.** [**https://programmerall.com/article/3460838855/**](https://programmerall.com/article/3460838855/)