

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

PHAN QUANG THÁI

**XÂY DỰNG HỆ TRỢ GIÚP RA QUYẾT ĐỊNH
TRONG CÔNG TÁC QUY HOẠCH CÁN BỘ
LÃNH ĐẠO, QUẢN LÝ TẠI TỈNH BÌNH ĐỊNH**

Chuyên ngành : Khoa học máy tính

Mã số : 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2012

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **PGS.TS. Phan Huy Khánh**

Phản biện 1 : **PGS.TSKH. TRẦN QUỐC CHIẾN**

Phản biện 2 : **TS. TRẦN THIÊN THÀNH**

Luận văn được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 20 tháng 01 năm 2013

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng;
- Trung tâm Học liệu, Đại học Đà Nẵng;

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Công nghệ thông tin là một trong các động lực quan trọng nhất của sự phát triển, cùng với một số ngành công nghệ cao khác đang làm biến đổi sâu sắc đời sống kinh tế, văn hoá, xã hội của thế giới hiện đại. Trước sự phát triển mạnh mẽ của công nghệ thông tin và những ứng dụng của nó, Chỉ thị số 58-CT/TW của Bộ Chính trị về đẩy mạnh ứng dụng và phát triển công nghệ thông tin phục vụ sự nghiệp công nghiệp hoá, hiện đại hoá, ngày 17 tháng 10 năm 2000 đã ra đời, tạo điều kiện cho sự phát triển công nghệ thông tin ở nước ta ngày càng mạnh mẽ.

Trong những năm gần đây , tại các cơ quan Đảng tỉnh Bình Định, việc ứng dụng công nghệ thông tin với mục tiêu “tin học hóa hoạt động các cơ quan Đảng” ngày càng được áp dụng rộng rãi. Hạ tầng kỹ thuật được đầu tư, nâng cấp hoàn thiện hơn. Các ứng dụng, cơ sở dữ liệu đã được xây dựng đưa vào khai thác và sử dụng, ví dụ như: Cơ sở dữ liệu đảng viên, cơ sở dữ liệu Văn kiện Đảng, Quy trình xử lý công văn trên mạng....

Tuy nhiên, so với kinh phí đã đầu tư, hiệu quả mang lại từ việc ứng dụng chưa cao, ứng dụng còn thiếu, và chưa đồng bộ, số cán bộ, chuyên viên lớn tuổi ngại sử dụng máy tính để giải quyết công việc chuyên môn. Trong công tác cán bộ nói chung và khâu quy hoạch đào tạo cán bộ chưa có sự trợ giúp đáng kể của công nghệ thông tin. Thông tin phục vụ cho công tác cán bộ còn thiếu , dẫn đến việc xử lý công việc chậm trễ, đôi lúc chưa đạt hiệu quả cao.

Công tác quy hoạch, đào tạo cán bộ trong thời gian vừa qua có lúc, có nơi tổ chức thực hiện chưa được tốt :

- Các cơ quan, đơn vị còn thiếu thông tin của cán bộ có thể đưa vào diện quy hoạch. Từ đó dẫn đến việc “quy hoạch động” theo yêu cầu là không đảm bảo. Tiến độ thực hiện quy hoạch một số cơ quan cấp tỉnh còn chậm, chưa bảo đảm thời gian theo quy định.

- Chưa đảm bảo tính công khai, chưa mở rộng thảo luận dân chủ. Có cơ quan, đơn vị đưa vào quy hoạch còn chưa đúng đối tượng theo quy định (không tuân thủ các tiêu chuẩn về quy hoạch cán bộ). Bỏ sót những người có trình độ chuyên môn tốt, có năng lực công tác không đưa vào diện quy hoạch.

- Độ tuổi dưới 40 đưa vào diện quy hoạch còn thấp.

Xuất phát từ những nhu cầu thực tế nêu trên tôi xin chọn đề tài ***“Xây dựng hệ trợ giúp ra quyết định trong công tác quy hoạch cán bộ lãnh đạo quản lý tại tỉnh Bình Định”*** với mong muốn đóng góp thêm một giải pháp về ứng dụng công nghệ thông tin , góp phần hạn chế những bất cập , tiêu cực, để công tác quy hoạch cán bộ trở nên minh bạch, khách quan, dân chủ, đội ngũ cán bộ được đưa vào diện quy hoạch là xứng đáng.

2. Mục đích nghiên cứu.

Trên cơ sở phân tích thực trạng về công tác quy hoạch, đào tạo các bộ và nhu cầu thực tiễn công việc. Đề tài sẽ tập trung nghiên cứu, vận dụng hệ trợ giúp ra quyết định để tạo ra sản phẩm hỗ trợ, giúp người dùng đưa ra quyết định phù hợp

Nghiên cứu các thuật toán liên quan đến cây quyết định. Phân tích, đánh giá và triển khai áp dụng thuật toán C4.5.

Phân tích thực trạng và nhu cầu ứng dụng công nghệ thông tin vào công tác quy hoạch cán bộ lãnh đạo quản lý. Đề ra giải pháp ứng dụng công nghệ thông tin nói chung và ứng dụng hệ trợ giúp ra quyết

định vào công tác quy hoạch cán bộ lãnh đạo, quản lý tại tỉnh Bình Định.

Áp dụng cơ sở lý thuyết nền tảng để xây dựng và triển khai ứng dụng.

Cài đặt và triển khai thử nghiệm tại cơ quan Văn phòng Tỉnh ủy và Ban tổ chức Tỉnh ủy Bình Định.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu

Nghiên cứu về các quy định, quy trình trong công tác cán bộ nói chung và công tác quy hoạch, điều động, bổ nhiệm cán bộ nói riêng.

Nghiên cứu về lý thuyết về hệ trợ giúp ra quyết định, thuật toán cây quyết định, trí tuệ nhân tạo.

Phạm vi nghiên cứu

Ứng dụng thuật toán C4.5 để xây dựng cây quyết định trong công tác quy hoạch cán bộ lãnh đạo, quản lý.

Các biểu mẫu, số liệu có liên quan đến công tác quy hoạch cán bộ. Mẫu thử nghiệm trong đề tài là các đối tượng có thể xem xét đưa và diện quy hoạch cán bộ lãnh đạo, quản lý cấp tỉnh (cấp trưởng, phó các sở, ban, ngành cấp tỉnh) tại Tỉnh ủy Bình Định.

4. Phương pháp nghiên cứu.

Luận văn sẽ kết hợp hai phương pháp nghiên cứu, đó là:

Phương pháp nghiên cứu lý thuyết

Nghiên cứu tài liệu, ngôn ngữ và các công nghệ có liên quan.

Tổng hợp, thu thập các tài liệu về công tác cán bộ nói chung và công tác quy hoạch cán bộ lãnh đạo quản lý nói riêng.

Phương pháp nghiên cứu thực nghiệm

Phân tích yêu cầu thực tế của của công việc , tìm ra giải pháp

và vận dụng lý thuyết, các thuật toán có liên quan để trợ giúp việc lập trình, xây dựng ứng dụng.

Thống kê, phân tích số liệu thực tế trong công tác quy hoạch trong giai đoạn 2005 -2010.

Xây dựng bộ dữ liệu mẫu dùng để kiểm tra, thử nghiệm chương trình và đưa ra nhận xét, đánh giá kết quả đạt được.

5. Ý nghĩa khoa học và thực tiễn

Về mặt khoa học

Đề tài sẽ đưa ra một phương thức ứng dụng cây quyết định trong công tác quy hoạch cán bộ, tạo tiền đề cho những nghiên cứu ứng dụng sau này.

Tìm hiểu và ứng dụng các công cụ, ngôn ngữ, công nghệ liên quan, đề ra giải pháp triển ứng dụng cây quyết định, có thể áp dụng lý thuyết này trong các nhu cầu công việc khác tại cơ quan.

Về mặt thực tiễn

Xây dựng hệ trợ giúp ra quyết định, triển khai hệ thống thử nghiệm tại Văn Phòng Tỉnh ủy và Ban tổ chức Tỉnh ủy nhằm đánh giá hiệu quả trước khi sử dụng rộng rãi.

Hệ thống thiết kế đơn giản, dễ sử dụng và có tính linh hoạt, có tính. Sản phẩm sẽ là hệ thống phục vụ đắc lực, kịp thời và có độ chính xác cao cho Ban Tổ chức Tỉnh ủy Bình Định, cho lãnh đạo các cơ quan đơn vị.

6. Bố cục luận văn

Sau phần mở đầu, giới thiệu..., nội dung chính của luận văn được chia thành 3 chương như sau:

Chương 1, trình bày cơ sở lý thuyết làm nền tảng để xây dựng ứng dụng, bao gồm: Hệ trợ giúp quyết định. Cây quyết định và giải thuật C4.5 xây dựng cây quyết định.

Chương 2, tìm hiểu, giới thiệu và phân tích thực trạng công tác quy hoạch cán bộ lãnh đạo quản lý tại tỉnh Bình Định, nêu những vấn đề hạn chế. Giải pháp ứng dụng cây quyết định phục vụ công tác quy hoạch cán bộ.

Chương 3, trình bày chi tiết về mô hình kiến trúc tổng thể của hệ thống và phương pháp xây dựng ứng dụng. Tiến hành kịch bản thử nghiệm trên số liệu thực tế, sau đó đánh giá kết quả đạt được và khả năng triển khai ứng dụng trên toàn hệ thống.

Cuối cùng là phần đánh giá, kết luận và hướng phát triển của đề tài.

CHƯƠNG 1

TỔNG QUAN VỀ HỆ HỖ TRỢ QUYẾT ĐỊNH

1.1. HỆ TRỢ GIÚP RA QUYẾT ĐỊNH

1.1.2. Khái niệm hệ trợ giúp ra quyết định

1.1.3. Quá trình ra quyết định

1.1.4. Các thành phần của hệ trợ giúp quyết định

1.2. RA QUYẾT ĐỊNH TRONG QUẢN LÝ

1.2.1. Mở đầu

1.2.2. Các phương pháp ra quyết định trong quản lý

1.3. HỆ TRỢ GIÚP RA QUYẾT ĐỊNH THÔNG MINH.

1.3.1. Tổng quan về trí tuệ nhân tạo

1.3.2. Một số vấn đề Trí tuệ Nhân tạo quan tâm.

1.4. CÂY QUYẾT ĐỊNH

1.4.1. Giới thiệu chung

1.4.2. Phân lớp dữ liệu dựa trên các kiểu cây quyết định

1.4.3. Giải thuật cơ bản xây dựng cây quyết định

1.4.4. Chọn thuật toán C4.5 xây dựng cây quyết định

1.5. THUẬT TOÁN C4.5

1.5.1. Giới thiệu

1.5.2. Giải thuật C4.5 xây dựng cây quyết định từ trên xuống.

a. Thuật toán C4.5

Function *xay_dung_cay*(*T*)

- ```
{
1. <Tính toán tần suất các giá trị trong các lớp của T>;
2. If <Kiểm tra các mẫu, nếu thuộc cùng một lớp hoặc có rất ít
 mẫu khác lớp> Then <Trả về 1 nút lá>
 Else <Tạo một nút quyết định N>;
3. For <Với mỗi thuộc tính A> Do <Tính giá trị Gain(A)>;
4. <Tại nút N, thực hiện việc kiểm tra để chọn ra thuộc tính có
 giá trị Gain tốt nhất (lớn nhất). Gọi N.test là thuộc tính
 có Gain lớn nhất>;
5. If <Nếu N.test là thuộc tính liên tục> Then <Tìm ngưỡng
 cho phép tách của N.test>;
6. For <Với mỗi tập con T' được tách ra từ tập T> Do
 (T' được tách ra theo quy tắc:
 - Nếu N.test là thuộc tính liên tục tách theo ngưỡng ở bước 5
 - Nếu N.test là thuộc tính phân loại rời rạc tách theo các giá
 trị của thuộc tính này.
7. { If <Kiểm tra, nếu T' rỗng>} Then
 <Gán nút con này của nút N là nút lá>;
 Else
8. <Gán nút con này là nút được trả về bằng cách gọi đệ qui
 lại đối với hàm xay_dung_cay(T'), với tập T'>;
 }
9. <Tính toán các lỗi của nút N>;
```



<Trả về nút N>;

}

**b. Đánh giá độ phức tạp của thuật toán C4.5**

**c. Chọn thuộc tính phân loại tốt nhất**

**d. Entropy đo tính thuần nhất của tập ví dụ**

Khái niệm entropy của một tập S được định nghĩa trong lý thuyết thông tin là số lượng mong đợi các bit cần thiết để mã hóa thông tin về lớp của một thành viên rút ra một cách ngẫu nhiên từ tập S. Trong trường hợp tối ưu, mã có độ dài ngắn nhất. Theo lý thuyết thông tin, mã có độ dài tối ưu là mã gán  $-\log_2 p$  bits cho thông điệp có xác suất là  $p$  [9].

Trong trường hợp S là tập ví dụ, thì thành viên của S là một ví dụ, mỗi ví dụ thuộc một lớp hay có một giá trị phân loại.

Entropy có giá trị nằm trong khoảng  $[0..1]$ .

Entropy(S) = 0: tập S chỉ toàn ví dụ thuộc cùng một loại, hay S là thuần nhất.

Entropy(S) = 1: tập ví dụ S có các ví dụ thuộc các loại khác nhau với độ pha trộn là cao nhất.

$0 < \text{Entropy}(S) < 1$ : tập ví dụ S có số lượng ví dụ thuộc các loại khác nhau là không bằng nhau.

Để đơn giản ta xét trường hợp các ví dụ của S chỉ thuộc loại âm (-) hoặc dương (+).

Cho trước: Tập S là tập dữ liệu huấn luyện, trong đó thuộc tính phân loại có hai giá trị, giả sử là âm (-) và dương (+).

Trong đó:

$p_+$  là phần các ví dụ dương trong tập S.

$p_-$  là phần các ví dụ âm trong tập S.

Khi đó, entropy đo độ pha trộn của tập S theo công thức sau:

$$\text{Entropy}(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

Một cách tổng quát hơn, nếu các ví dụ của tập  $S$  thuộc nhiều hơn hai loại, giả sử là có  $c$  giá trị phân loại thì công thức entropy tổng quát là:

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

***e. Lượng thông tin thu được đo mức độ giảm Entropy mong đợi***

Entropy là một số đo đo độ pha trộn của một tập ví dụ, bây giờ chúng ta sẽ định nghĩa một phép đo hiệu suất phân loại các ví dụ của một thuộc tính. Phép đo này gọi là lượng thông tin thu được (hay độ lợi thông tin), nó đơn giản là lượng giảm entropy mong đợi gây ra bởi việc phân chia các ví dụ theo thuộc tính này.

Một cách chính xác hơn,  $\text{Gain}(S, A)$  của thuộc tính  $A$ , trên tập  $S$ , được định nghĩa như sau:

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Giá trị Value ( $A$ ) là tập các giá trị có thể cho thuộc tính  $A$ , và  $S_v$  là tập con của  $S$  mà  $A$  nhận giá trị  $v$ .

***f. Tỷ suất lợi ích Gain Ratio***

Khái niệm độ lợi thông tin  $\text{Gain}$  có xu hướng ưu tiên các thuộc tính có số lượng lớn các giá trị. Nếu thuộc tính  $D$  có giá trị riêng biệt cho mỗi bản ghi thì  $\text{Entropy}(S, D) = 0$ , như vậy  $\text{Gain}(S, D)$  sẽ đạt giá trị cực đại. Rõ ràng, một phân vùng như vậy thì việc phân loại là vô ích.

Thuật toán C4.5, một cải tiến của ID3, mở rộng cách tính Information Gain thành Gain Ratio để cố gắng khắc phục sự thiên lệch.

Gain Ratio được xác định bởi công thức sau:

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

Với SplitInformation(S, A) chính là thông tin do phân tách của A trên cơ sở giá trị của thuộc tính phân loại S. Công thức tính như sau:

Trong đó:

*Value(S)* là tập các giá trị của thuộc tính S

*A<sub>i</sub>* là tập con của tập A ứng với thuộc tính S = giá trị là

*v<sub>i</sub>*

### 1.5.3. Phương pháp đánh giá mức độ hiệu quả

## CHƯƠNG 2

### ỨNG DỤNG CÂY QUYẾT ĐỊNH TRONG CÔNG TÁC QUY HOẠCH CÁN BỘ LÃNH ĐẠO, QUẢN LÝ.

#### 2.1. CÔNG TÁC QUY HOẠCH CÁN BỘ LÃNH ĐẠO, QUẢN LÝ.

2.1.1. Quan niệm về công tác quy hoạch cán bộ lãnh đạo quản lý.

a. Quy hoạch cán bộ.

b. Cán bộ lãnh đạo, quản lý:

2.1.2. Nhận thức về công tác quy hoạch cán bộ.

2.1.3. Tiêu chuẩn cán bộ đưa vào quy hoạch.

2.1.4. Thực trạng về đội ngũ cán bộ tỉnh Bình Định.

2.1.5. Thực trạng công tác quy hoạch cán bộ lãnh đạo quản lý.

2.1.6. Phân tích thực trạng.

### 2.1.7. Giải pháp xây dựng hệ thống trợ giúp quyết định.

### 2.1.8. Kịch bản sử dụng hệ thống trợ giúp ra quyết định

## 2.2. ỨNG DỤNG CÂY QUYẾT ĐỊNH

### 2.2.1. Lượng hóa dữ liệu để đưa vào thuật toán C4.5

Để thuật toán C4.5 xây dựng được cây quyết định hiệu quả, đáng tin cậy, trước tiên ta phải lượng hóa dữ liệu nhằm có được tập dữ liệu huấn luyện tốt nhất.

### 2.2.2. Phân tích dữ liệu.

Qua xem xét, phân tích bảng dữ liệu mẫu 2.3, bảng tóm tắt tiêu chuẩn cán bộ đưa vào quy hoạch 2.2, quy trình thực hiện quy hoạch hình 2.1 phần trên, ta thấy có nhiều yếu tố ảnh hưởng đến công tác quy hoạch cán bộ. Tuy nhiên, cần chú trọng phân tích những yếu tố chính sau:

*Độ tuổi (Dotuoi):* Có 2 độ tuổi được xem xét kỹ theo tiêu chuẩn quy hoạch cán bộ là : trên 40 tuổi và dưới 40 tuổi. Ta có thể chia khoảng như sau: dưới 40 tuổi (40-), từ 40 tuổi trở lên (40+).

*Trình độ chuyên môn (TDCM):* Là trình độ được tạo tạo của các nhân sự ta cần xem xét để đưa vào quy hoạch. Ta có thể chia theo 3 mức sau:

+ Dhcq : Những người có bằng đại học chính quy, bằng thạc sỹ, tiến sỹ.

+ Dhct: Những người có bằng đại học tại chức, từ xa, văn bằng 2 và các loại hình đào đại học khác.

+ Khong: là những người chưa có bằng đại học.

*Trình độ Lý luận chính trị (LLCT):* Ta chia làm làm 3 loại hình theo quy định hiện nay:

+ Cctt: Là những người có trình độ cao cấp lý luận chính trị hệ tập trung trở lên.(Cao cấp chính trị và cử nhân chính trị)

+ Cc0tt: Là những người có trình độ cao cấp lý luận chính trị không thuộc hệ tập trung.

+ Khong: những người chưa có trình độ lý chính trị cao cấp, trung cấp.

*Trình độ ngoại ngữ (NN): Chia làm 2 trường hợp*

+ Co: Là người có chứng chỉ ngoại ngữ trình độ B trở lên.

+ Khong: Chưa chứng chỉ ngoại ngữ trình độ B trở lên.

Các yếu tố trên chính là tập thuộc tính, dựa vào tập thuộc tính này để dự đoán giá trị cho thuộc tính đích Quyết định.

*Quyết định (QD):* là kết quả của việc áp dụng cây quyết định. Ta chia 2 tình huống: những người có thể đưa vào diện quy hoạch (Yes) và những người không nên đưa vào diện quy hoạch (No).

*Ta có các giá trị của các thuộc tính như sau:*

Dotuoi = {40-;40+}

TDCM = {Dhcq;Dhtc;Khong}

LLCT = {Cctt;Cc0tt;Khong}

NN = {Co;Khong}

QD = {Yes;No}

Từ phân tích trên ta xây dựng bảng dữ liệu huấn luyện mẫu như sau:

**Bảng 2.5. Bảng dữ liệu mẫu huấn luyện.**

| STT | Dotuoi | TDCM | LLCT  | NN    | QD  |
|-----|--------|------|-------|-------|-----|
| 1   | 40-    | Dhcq | Cc0tt | Co    | Yes |
| 2   | 40+    | Dhcq | Cc0tt | Co    | Yes |
| 3   | 40-    | Dhcq | Cc0tt | Khong | No  |
| 4   | 40+    | Dhcq | Cc0tt | Khong | No  |
| 5   | 40-    | Dhtc | Cc0tt | Co    | No  |
| 6   | 40+    | Dhtc | Cc0tt | Co    | Yes |

|    |     |       |       |       |     |
|----|-----|-------|-------|-------|-----|
| 7  | 40- | Dhtc  | Cc0tt | Khong | No  |
| 8  | 40+ | Dhtc  | Cc0tt | Khong | No  |
| 9  | 40+ | Khong | Cc0tt | Co    | No  |
| 10 | 40- | Dhcq  | Cctt  | Co    | Yes |
| 11 | 40+ | Dhcq  | Cctt  | Co    | Yes |
| 12 | 40- | Dhcq  | Cctt  | Khong | No  |
| 13 | 40+ | Dhcq  | Cctt  | Khong | No  |
| 14 | 40- | Dhtc  | Cctt  | Co    | Yes |
| 15 | 40+ | Dhtc  | Cctt  | Co    | Yes |
| 16 | 40- | Dhtc  | Cctt  | Khong | No  |
| 17 | 40+ | Dhtc  | Cctt  | Khong | No  |
| 18 | 40+ | Khong | Cctt  | Co    | No  |
| 19 | 40- | Dhcq  | Khong | Co    | No  |
| 20 | 40+ | Dhcq  | Khong | Co    | No  |
| 21 | 40- | Dhcq  | Khong | Khong | No  |
| 22 | 40+ | Dhcq  | Khong | Khong | No  |
| 23 | 40- | Dhtc  | Khong | Co    | No  |
| 24 | 40+ | Dhtc  | Khong | Co    | No  |
| 25 | 40- | Dhtc  | Khong | Khong | No  |
| 26 | 40+ | Dhtc  | Khong | Khong | No  |

Từ ví dụ trên, áp dụng thuật thuật toán C4.5, dùng tập dữ liệu huấn luyện để xây dựng cây quyết định cho bài toán trợ giúp quy hoạch cán bộ.

### 2.2.3. Triển khai giải thuật C4.5

Cây quyết định xây dựng bằng cách mở rộng cây xuất phát từ nút gốc. Nút được đặt tên là nút 1, và ứng với mức (level) 0 của cây. Các nút con của nút gốc sẽ ứng với nút 1, và tiếp tục như vậy. Để thuận tiện cho việc xây dựng cây quyết định, tại mỗi nút, chứa các thông tin cơ bản sau đây:

Dữ liệu huấn luyện ứng với nút hiện thời

Tên của nút cha của nút hiện thời

Thuộc tính dùng để phân tích nút hiện thời. Nếu nút hiện thời là nút lá, thuộc tính sẽ bằng rỗng. Nếu nút hiện thời không phải nút lá, nó sẽ chứa tên các nút con xuất phát từ nút hiện thời, cũng như các giá trị tương ứng của thuộc tính dùng để phân tích nút hiện thời. Nếu là nút lá thì có thêm quyết định đầu ra ở tại nút hiện thời. Ngược lại, nếu không phải là nút lá thì không có chứa thông tin này.

Ở đây, ta chú ý rằng, dữ liệu huấn luyện tại từng nút thay đổi tùy theo các nút khác nhau. Cây quyết định được xây dựng bằng cách xem xét tại mỗi nút liệu nó có thể được mở rộng tiếp hay không. Nếu nó không được mở rộng nữa thì được gọi là nút lá (leaf). Một nút sẽ được xem là nút lá nếu Entropy tại nút này bằng không (hoặc tất cả các thuộc tính đều đã được duyệt). Trên mỗi đường đi từ nút gốc đến nút lá, mỗi thuộc tính chỉ được duyệt tối đa một lần.

Dựa trên các thông tin chứa ở các nút và cách thức mở rộng cây quyết định như mô tả ở trên, ta thực hiện giải thuật C4.5 trong chương trình mô phỏng như sau. Ta sẽ xét 1 tập  $S_N$  chứa các nút cần được xử lý. Việc xử lý tại mỗi nút bao gồm việc xem xét nút đó có phải là nút lá hay không. Nếu là nút lá thì tìm quyết định đầu ra tương ứng. Ngược lại, nếu không phải là nút lá, ta sẽ tìm thuộc tính để phân tích nút này, và từ đó đưa ra thông tin của các nút con xuất phát từ đó. Ban đầu, tập  $S_N$  chứa một nút duy nhất là gốc, với các thông tin tương ứng được khởi động. Ta sẽ tiến hành quá trình xử lý các nút trên trong tập  $S_N$ . Với mỗi nút ta sẽ thực hiện việc xử lý như mô tả trên đây. Có 2 trường hợp sau: Nếu nó là nút lá, ta loại nó ra khỏi tập  $S_N$ . Nếu không phải là nút lá, ta loại nó ra khỏi tập  $S_N$ , đồng thời bổ sung các nút con của nó vào tập  $S_N$ . Như vậy, việc xử lý trên sẽ được thực hiện theo một vòng lặp cho đến khi tập  $S_N$  là rỗng.

#### **2.2.4. Xây dựng cây quyết định**

Để xây dựng cây quyết định ta phải xác định nút gốc để phân tách cây. Thuộc tính có độ lợi thông tin lớn nhất sẽ được chọn làm nút gốc.

Gọi S là tập thuộc tính đích. Có tất cả 26 ví dụ, trong đó:

- + Yes xuất hiện trong tập thuộc tính đích 7 lần,
- + No xuất hiện trong tập thuộc tính đích 19 lần.

Áp dụng công thức tính Entropy, ta có:

$$\text{Entropy}(S) = -(19/26)\log_2(19/26) - (7/26)\log_2(7/26) = 0,84$$

Đối với thuộc tính “Độ tuổi” (Dotuoi), ta tính Entropy của các tập con S được chia bởi các giá trị của thuộc tính “Dotuoi” như sau:

**Bảng 2.6. Bảng Entropy(S) phân theo độ tuổi**

| Dotuoi | Số lượng | QD  |    | Entropy |
|--------|----------|-----|----|---------|
|        |          | Yes | No |         |
| 40-    | 12       | 3   | 9  | 0,81    |
| 40+    | 14       | 4   | 10 | 0,86    |

**Entropy của S đối với thuộc tính “Độ tuổi” là:**

$$\text{Entropy}(S, \text{Dotuoi}) = (12/26) \times 0,81 + (14/26) \times 0,86 = 0,84$$

Độ lợi thông tin tương ứng là:

$$\begin{aligned} \text{Gain}(S, \text{Dotuoi}) &= \text{Entropy}(S) - \text{Entropy}(S, \text{Dotuoi}) = \\ &= 0,84 - 0,84 = 0,00 \end{aligned}$$

Tỷ suất lợi ích Gain Ratio:

$$\begin{aligned} \text{SplitInfor}(S, \text{Dotuoi}) &= - (12/26) \times \log_2(12/26) - (14/26) \times \\ &\quad \times \log_2(14/26) = 1 \end{aligned}$$

$$\text{GainRatio}(S, \text{Dotuoi}) =$$

$$= \text{Gain}(S, \text{Dotuoi}) / \text{SplitInfor}(S, \text{Dotuoi}) = 0,00/1 = 0$$

Một cách tương tự, ta tính độ lợi thông tin Gain và tỉ suất lợi ích GainRatio của các thuộc tính còn lại.



**Bảng 2.7. Độ lợi thông tin của thuộc tính “Trình độ chuyên môn”**

| TDCM  | Số lượng | QD  |    | Entropy |
|-------|----------|-----|----|---------|
|       |          | Yes | No |         |
| Dhcq  | 12       | 4   | 8  | 0,92    |
| Dhtc  | 12       | 3   | 9  | 0,81    |
| Khong | 2        | 0   | 2  | 0,00    |

$$\text{Gain}(S, \text{TDCM}) / \text{SplitInfor}(S, \text{TDCM}) = 0,03$$

**Bảng 2.8. Độ lợi thông tin của thuộc tính “Lý luận chính trị”**

| LLCT  | Số lượng | QD  |    | Entropy |
|-------|----------|-----|----|---------|
|       |          | YES | NO |         |
| Cctt  | 9        | 4   | 5  | 0,99    |
| Cc0tt | 9        | 3   | 6  | 0,92    |
| Khong | 8        | 0   | 6  | 0,00    |

$$\text{GainRatio}(S, \text{LLCT}) = 0,18 / 1,58 = 0,11$$

**Bảng 2.9. Độ lợi thông tin của thuộc tính “Ngoại ngữ”**

| NN    | Số lượng | QD  |    | Entropy |
|-------|----------|-----|----|---------|
|       |          | Yes | No |         |
| Co    | 14       | 7   | 7  | 1,00    |
| Khong | 12       | 0   | 13 | 0,00    |

$$\text{GainRatio}(S, \text{Dotuoi}) = 0,30$$

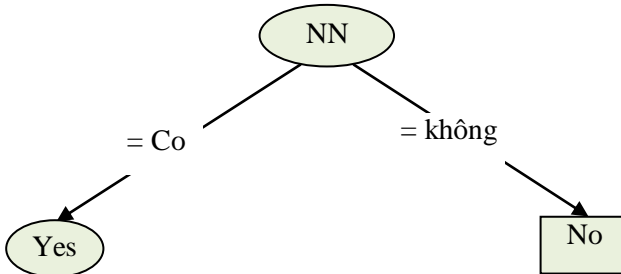
Từ những kết quả tính toán trên, ta đưa ra bảng so sánh GainRatio của các thuộc tính: Dotuoi, TDCM, LLCT, NN như sau.

**Bảng 2.10 Bảng so sánh kết quả tính GainRatio**

| Thuộc tính | Gain        | SplitInfor  | GainRatio   |
|------------|-------------|-------------|-------------|
| Dotuoi     | 0,00        | 1,00        | 0,00        |
| TDCM       | 0,04        | 1,31        | 0,03        |
| LLCC       | 0,18        | 1,58        | 0,11        |
| <b>NN</b>  | <b>0,30</b> | <b>1,00</b> | <b>0,30</b> |

Ta nhận thấy  $\text{GainRatio}(S, \text{NN}) = 0.30$ , đạt giá trị lớn nhất, do đó thuộc tính NN (Ngoại ngữ) có khả năng phân loại tốt nhất. Chính vì vậy ta sẽ chọn thuộc tính này làm nút gốc phân tách cây.

Ta có cây quyết định cấp 1 như hình vẽ sau:



**Hình 2.5. Cây quyết định cấp 1**

Ứng với giá trị NN = No; có Entropy = 0. Do đó, tại nhánh này sẽ là nút có phân lớp là No.

Đối với các nhánh còn lại NN = Co, ta tiếp tục vận dụng thuật toán để tìm thuộc tính làm nút của cây.

Xét nhánh NN = Co, ta có bảng dữ liệu như sau:

**Bảng 2.11. Bảng dữ liệu trường hợp Nn = Co**

| STT | Dotuoi | TDCM | LLCT  | NN | QD  |
|-----|--------|------|-------|----|-----|
| 1   | 40-    | Dhcq | Cc0tt | Co | Yes |
| 2   | 40+    | Dhcq | Cc0tt | Co | Yes |
| 3   | 40-    | Dhtc | Cc0tt | Co | No  |

|    |     |       |       |    |     |
|----|-----|-------|-------|----|-----|
| 4  | 40+ | Dhtc  | Cc0tt | Co | Yes |
| 5  | 40+ | Khong | Cc0tt | Co | No  |
| 6  | 40- | Dhcq  | Cett  | Co | Yes |
| 7  | 40+ | Dhcq  | Cett  | Co | Yes |
| 8  | 40- | Dhtc  | Cett  | Co | Yes |
| 9  | 40+ | Dhtc  | Cett  | Co | Yes |
| 10 | 40+ | Khong | Cett  | Co | No  |
| 11 | 40- | Dhcq  | Khong | Co | No  |
| 12 | 40+ | Dhcq  | Khong | Co | No  |
| 13 | 40- | Dhtc  | Khong | Co | No  |
| 14 | 40+ | Dhtc  | Khong | Co | No  |

Gọi  $S_1$  là tập thuộc tính đích. Có tất cả 14 ví dụ, trong đó:

+ Yes xuất hiện trong tập thuộc tính đích 7 lần,

+ No xuất hiện trong tập thuộc tính đích 7 lần.

Áp dụng công thức tính Entropy ta có:

$$\text{Entropy}(S_1) = -(7/14) \times \log_2(7/14) - (7/14) \times \log_2(7/14) = \mathbf{1.00}$$

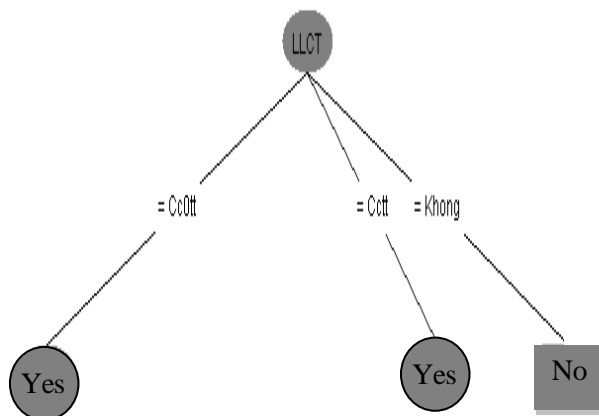
Tiếp theo ta lần lượt tính Gain của các thuộc tính như đã tính ở phần trên, kết quả như sau:

**Bảng 2.15. Kết quả tính SplitInfor và GainRatio**

| Thuộc tính  | Gain        | SplitInfor  | GainRatio   |
|-------------|-------------|-------------|-------------|
| Dotuoi      | 0.00        | 0.99        | 0.00        |
| TDCM        | 0.18        | 1.45        | 0.12        |
| <b>LLCT</b> | <b>0.40</b> | <b>1.58</b> | <b>0.25</b> |

Qua bảng so sánh kết quả trên, ta thấy  $\text{GainRatio}(S_1, \text{LLCT}) = 0.25$  đạt giá trị lớn nhất, do đó tại nhánh này ta chọn thuộc tính LLCT để làm nút phân tách cây.

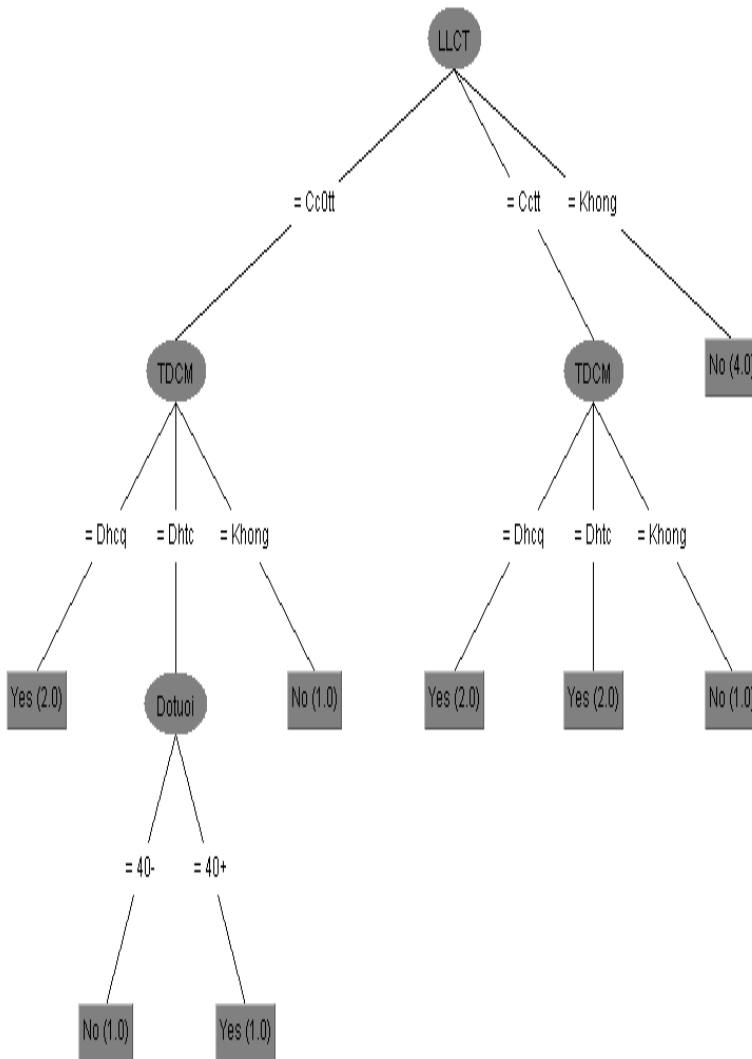
Ta có cây quyết định cấp ứng với nhánh NN =Co, thu được các nhánh con tương ứng như hình vẽ sau :



**Hình 2.6. Cây quyết định ứng với nhánh NN = Co**

Ta thấy, ứng với tất các giá trị LLCT = “Khong”, có entropy = 0, do đó, tại các nhánh này sẽ là nút lá với phân lớp là Yes (LLCT=Cctt và LLCT= Cc0tt).

Tương tự như cách tính ở trên ta xét cho các nhánh còn lại. Ta xây dựng được cây quyết định hoàn chỉnh như sau:



**Hình 2.10. Cây quyết định hoàn chỉnh**

#### 2.2.4. Tạo luật từ cây quyết định

Từ cây quyết định trên, ta có thể rút ra được một số luật sau:

|                                                                                                |
|------------------------------------------------------------------------------------------------|
| IF (NN=No) Then QD = No                                                                        |
| IF (NN=Co) and IF (LLCT= Khong) Then QD = No                                                   |
| IF (NN=Co) and IF ((LLCT= Cctt) or (LLCT= Cc0tt)) and<br>IF (TDCM = Khong) Then QD = No        |
| IF (NN=Co) and IF ((LLCT= Cctt) or (LLCT= Cc0tt)) and<br>IF (TDCM = Dhcq) Then QD = Yes        |
| IF (NN=Co) and IF (LLCT = Cctt) and (TDCM= Dhct)<br>Then QD = Yes                              |
| IF (NN=Co) and IF (LLCT = Cc0tt) and IF (TDCM=<br>Dhct) and and IF (Dotoi = 40+) Then QD = Yes |
| IF (NN=Co) and IF (TDCM= Dhct) and IF (LLCT =<br>Cc0tt) and IF (Dotoi = 40-) Then QD = No      |

Từ việc phân tích các mẫu dữ liệu ban đầu, luận văn đã xây dựng được một cây quyết định bằng cách ứng dụng thuật toán C4.5. Cây quyết định này có khả năng phân loại đúng đắn các ví dụ trong tập dữ liệu đã cho. Từ đó ta có thể thấy rằng đây là một công cụ biểu diễn rất trực quan, phù hợp trong việc phân tích các dữ liệu mà bài toán của luận văn đã đặt ra.

Tập luật được rút ra từ cây quyết định gồm 7 luật đơn giản, dễ hiểu và dễ cài đặt. Luận văn sẽ vận dụng tập luật này để xây dựng ứng dụng trợ giúp ra quyết định trong công tác quy hoạch cán bộ.

## **CHƯƠNG 3**

### **XÂY DỰNG ỨNG DỤNG THỬ NGHIỆM**

#### **3.1. XÂY DỰNG ỨNG DỤNG TRỢ GIÚP QUYẾT ĐỊNH**

**3.1.1. Chức năng của hệ thống**

**3.1.2. Kiến trúc tổng thể của hệ thống**

**3.1.3. Đặc tả chi tiết các thành phần**

**3.1.4. Cài đặt chương trình**

**3.1.5. Giao diện và các chức năng của chương trình.**

#### **3.2. THỬ NGHIỆM ỨNG DỤNG**

**3.2.1. Huấn luyện cây quyết định**

**3.2.2. Áp dụng cây quyết định**

#### **3.3. ĐÁNH GIÁ KẾT QUẢ THỰC HIỆN**

**3.3.1. Về ứng dụng hệ thống**

Sau khi đưa hệ thống trợ giúp quyết định trong công tác quy hoạch cán bộ tại tỉnh Bình Định với khoảng 200 mẫu, chúng tôi rút ra được những nhận xét, đánh giá một cách tương đối về hệ thống như sau:

Hệ thống được triển khai cài đặt, thử nghiệm dễ dàng, dễ sử dụng.

Hệ thống hoạt động ổn định, hiển thị kết quả nhanh chóng và rất thuận tiện.

Có thể quản lý, xử lý dữ liệu cán bộ trong diện quy hoạch nhanh chóng thuận tiện.

Với số lượng dữ liệu thử nghiệm, hệ thống đã xử lý và cho ra được những quyết định có độ chính xác cao, phù hợp với tính toán thực tế.

### **3.3.2. Về hiệu quả trong công tác quy hoạch cán bộ.**

Về thời gian: việc chuẩn bị nhân sự để đưa vào quy hoạch đối với các đơn vị (sở, ngành) và Ban tổ chức Tỉnh ủy theo cách làm hiện nay sẽ tốn nhiều thời gian. Thông thường đối với các đơn vị có số lượng quy hoạch ít, thời gian chuẩn bị khoảng 1-2 tuần; đơn vị có số lượng quy hoạch nhiều hơn, thời gian chuẩn bị khoảng 1 tháng. Công tác quy hoạch cán bộ vào Tỉnh ủy viên, Hội đồng nhân dân cần nhiều thời gian chuẩn bị hơn. Nếu sử dụng chương trình đã xây dựng sẽ rút ngắn thời gian chuẩn bị rất nhiều, tạo điều kiện cho người làm công tác tổ chức có thời gian nhiều hơn trong việc rà soát, cân nhắc, điều chỉnh quy hoạch.

Dựa vào thống kê của chương trình, ta có thể rà soát theo cơ cấu độ tuổi, tỷ lệ giới tính (đối chiếu so với quy định). Xem xét những người không đưa vào quy hoạch họ thiếu những tiêu chuẩn gì, những đơn vị nào có đội ngũ cán bộ kế cận còn mỏng. Từ đó, có kế hoạch đào tạo bồi dưỡng, tạo nguồn dự bị, tránh được tình trạng thiếu hụt cán bộ nguồn.

Dựa vào CSDL “quy hoạch cán bộ”, công tác quản lý cán bộ trong diện quy hoạch cũng thuận tiện hơn, hàng năm việc bổ sung, điều chỉnh quy hoạch được triển khai nhanh chóng, kịp thời loại ra khỏi quy hoạch những người không còn đủ điều kiện, bổ sung vào quy hoạch những người mới có năng lực, đáp ứng đầy đủ các tiêu chuẩn.



## KẾT LUẬN

### 1. Kết quả đạt được

#### *Về mặt khoa học*

Luận văn đã tiến hành nghiên cứu, phân tích, tìm hiểu được các bước quy trình của công tác quy hoạch cán bộ lãnh đạo, quản lý.

Nắm được các phương pháp và các mô hình toán học, áp dụng để giải quyết yêu cầu luận văn đã đặt ra.

Nghiên cứu và vận dụng giải thuật C4.5 để xây dựng mô hình bằng cây quyết định.

#### *Về mặt thực tiễn*

Luận văn đã nêu được giải pháp kỹ thuật để vận dụng và xây dựng hệ thống trợ giúp quyết định trong công tác quy hoạch cán bộ lãnh đạo quản lý tại tỉnh Bình Định.

Xây dựng được ứng dụng có khả năng phân tích các dữ liệu về công tác quy hoạch cán bộ lãnh đạo quản lý trong những năm qua.

Việc kết hợp lý thuyết về hệ trợ giúp quyết định và phân tích dữ liệu bằng cây quyết định là rất cần thiết, nó giúp giảm thiểu đáng kể về thời gian trong việc tìm kiếm, xác định thông tin, chuẩn bị nhân sự, tạo tính minh bạch trong công tác quy hoạch cán bộ. Công tác quản lý cán bộ trong diện quy hoạch trở nên thuận lợi, việc bổ sung, điều chỉnh quy hoạch sẽ tốt hơn.

Có thể nói, đây là một công cụ có tính thực tiễn, giúp ích cho những người làm công tác tổ chức cán bộ.

### 2. Hạn chế

Đặc thù của công tác tổ chức cán bộ nói chung và công tác quy hoạch cán bộ nói riêng có tính chất định tính, có nhiều yếu tố nhạy cảm, vì vậy hệ thống trợ giúp ra quyết định mang tính hỗ trợ là chính. Hệ thống cũng không thể giải quyết hết các trường hợp trong công tác quy

hoạch cán bộ (ví dụ đối với đối tượng là người dân tộc thiểu số, ưu tiên đối tượng nữ...).

Hiện tại hệ thống chỉ chạy thử nghiệm, chưa được phép kết nối vào CSDL đảng viên và CSDL cán bộ. Do đó cũng chưa đánh giá được hết những ưu điểm và nhược điểm của hệ thống cũng như tính tương thích với các chương trình “quản lý cán bộ và “quản lý đảng viên”.

### **3. Hướng phát triển**

Cần thử nghiệm hệ thống với khối lượng dữ liệu lớn để đánh giá lại độ tin cậy của cây quyết định trợ trong công tác quy hoạch cán bộ

Phối hợp với các đơn vị liên quan để tính toán, xây dựng hoàn thiện hơn tập dữ liệu huấn luyện nhằm đưa hệ thống vào sử dụng có hiệu quả, tận dụng được các CSDL có sẵn.

Nghiên cứu vận dụng thuật toán C5.0, là một cải tiến của C4.5, để giảm thiểu tỉ lệ lỗi, nâng cao hiệu suất và tăng cường độ tối ưu cho ứng dụng.

Tiếp tục phát triển, nâng cấp hệ thống có thể chạy trên hạ tầng mạng diện rộng (Wan) của Tỉnh ủy, thực hiện việc đồng bộ dữ liệu qua mạng, có thể tương thích và liên kết tốt một số các chương trình đang được sử dụng tại cơ quan đảng tỉnh Bình Định.