1. Phân tích sự phụ thuộc của các thuộc tính bảng tính lương
2. Các thuộc tính không phụ thuộc:

+ Họ tên

+ Địa chỉ

+ Dân tộc

+ Sở thích không liên quan công việc

+ Tuổi

+ Giới tính

+ Ngoại ngữ

+ Ý thức, trách nhiệm với công việc công ty

+ Độ hót của công việc hiện tại

1. Các thuộc tính ràng buộc với các thuộc tính khác:

+ Bằng cấp: loại bằng, xếp loại, trường đào tạo

* Phụ thuộc vào tuổi: bằng đại học phải trên 22 tuổi, thạc sĩ phải trên 24 tuổi, tiến sĩ phải trên 27 tuổi
* Phụ thuộc ngoại ngữ: Tiến sĩ chắc chắn phải có chứng chỉ ngoại ngữ

+ Mảng công việc:

* Phụ thuộc bằng cấp: các công việc chuyên môn sâu cần bằng cấp cao, đúng chuyên ngành
* Phụ thuộc vào giới tính: tester sẽ chủ yếu là nữ, lập trình viên chủ yếu là nam giới
* Phụ thuộc vào tuổi

+ Năm kinh nghiệm:

* Phụ thuộc vào tuổi

+ Sở trường trong công việc:

* Phụ thuộc vào mảng công việc

+ Vị trí ứng tuyển:

* Phụ thuộc vào tuổi tác
* Phụ thuộc vào bằng cấp
* Phụ thuộc vào năm kinh nghiệm
* Phụ thuộc vào sở trường công việc
* Phụ thuộc vào ý thức, trách nhiệm với công việc công ty

+ Hệ số trách nhiệm:

* Phụ thuộc vào bằng cấp
* Phụ thuộc vào vị trí ứng tuyển

+ Giải thưởng, thành tích cá nhân:

* Phụ thuộc vào bằng cấp: thông thường bằng cấp cao sẽ có nhiều thành tích

+ Ngưỡng lương:

* Phụ thuộc vào bằng cấp
* Phụ thuộc vào mảng công việc
* Phụ thuộc vào vị trí ứng tuyển

+ Xếp loại trình độ (năng lực thực hiện công việc):

* Phụ thuộc bằng cấp: thường bằng cấp cao sẽ có năng lực hoàn thành tốt công việc
* Phụ thuộc vào vị trí: Vị trí ứng tuyển cao thường sẽ có năng lực thực hiện công việc tốt

+ Thời gian gắn bó với công ty:

* Phụ thuộc vào tuổi tác: VD: Thời gian gắn bó của người 23 tuổi với công ty tối đa chỉ được 5 năm, vì 18 tuổi mới đủ tuổi lao động
* Phụ thuộc vị trí ứng tuyển: Người có vị trí cao thường ở lại lâu dài

+ Mức lương công ty cũ trả:

* Phụ thuộc vào năm kinh nghiệm, nếu năm kinh nghiệm bằng 0 thì nhiều khả năng người đó không có lương cũ.
* Phụ thuộc vào ngưỡng lương: nếu mà mức lương cũ cao hơn mức ngưỡng lương của công ty hiện tại thì người đó không gia nhập công ty hiện tại.

+ Lương (nhãn đầu ra):

* Phụ thuộc vào 16 thuộc tính chính như đã phân tích, ngoài ra chú ý fake là mức lương không chênh lệch nhiều với mức lương cũ (nếu có)

1. Phân phối dữ liệu của các trường thuộc tính
2. Phân phối đều:

+ Họ tên

+ Sở thích không liên quan công việc (sở thích cá nhân)

+ Giới tính

+ Ý thức trách nhiệm với công việc công ty

1. Phân phối chuẩn và các loại phân phối khác: các thuộc tính còn lại
2. Fake dữ liệu: Random có ràng buộc các thuộc tính, phân phối dữ liệu phù hợp với từng trường dữ liệu

B1: Tạo các thuộc tính không phụ thuộc:

* Các thuộc tính phân phối đều tạo trước
* Các thuộc tính phân phối khác tạo sau

B2: Các thuộc tính phụ thuộc sau theo nguyên tắc:

* Thuộc tính B phụ thuộc thuộc tính A, thì thuộc tính A phải được tạo trước