
Моделируя рассуждения

— Байесовские сети: как и зачем —

0 рассуждениях

Использует ли он(а) Python
в своей профессиональной
деятельности?



0 рассуждениях

Использует ли он(а) Python
в своей профессиональной
деятельности?



— А если это студент
прикладной математики?

0 рассуждениях

Использует ли он(а) Python
в своей профессиональной
деятельности?



— А если это Data Scientist?

О рассуждениях

- у нас есть какое-то предварительное знание (prior)
- мы оцениваем событие с учетом предварительного знания (likelihood)
- мы обновляем наши оценки при поступлении новых данных (posterior)

Теорема Байеса

$$P(T|D) = \frac{P(D|T)P(T)}{P(D)}$$

Теорема Байеса

$$P(\text{Theory}|\text{Data}) = \frac{P(\text{Data}|\text{Theory})P(\text{Theory})}{P(\text{Data})}$$

Что такое байесовские сети

Байесовские сети

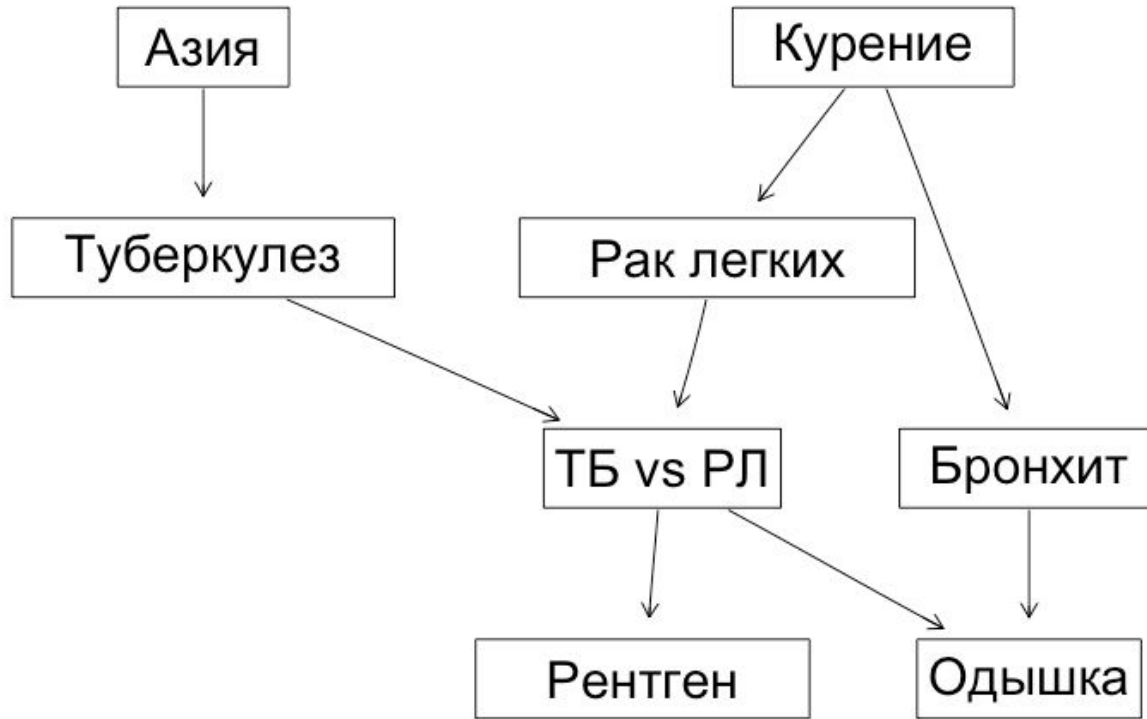


- Направленный ациклический граф
- Ребра показывают причинно-следственную связь
- Только непосредственные связи
- Заданы условные вероятности потомков при условии предков
- Ряд ограничений на связи

Байесовские сети



Рассуждения = распространение (пропагация) свидетельства -- появились новые данные, нужно оценить изменения в других вершинах



Вывод в двух направлениях (Lauritzen S, Spiegelhalter D)

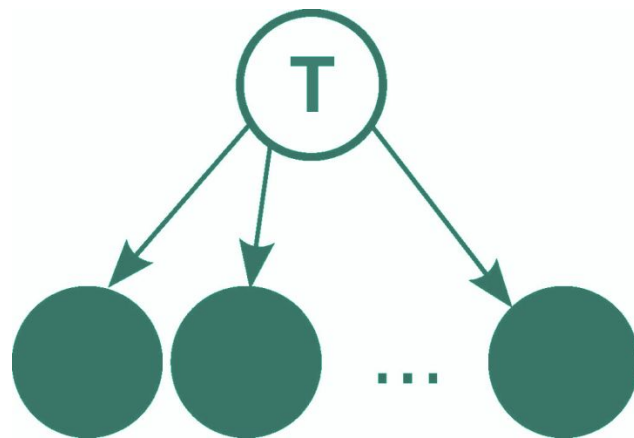
Пробуем!

<https://jiajin.shinyapps.io/shinyBN/>



Наивный байесовский классификатор

- нужно найти самую вероятную теорию T (или оценить параметр) при условии данных D
- в предположении, что отдельные элементы данных независимы при фиксированном значении T , сложная формула вероятности сводится к произведению отдельных элементов



Вывод в байесовских сетях

В общем случае сложнее, но идея та же -- разложить большое распределение на произведение маленьких

Как следствие -- ограничения на

- распределения
 - дискретные
 - многомерные нормальные
 - смешанные
- структуры

Вывод в байесовских сетях

Т.е. разбиваем сложные взаимосвязи между большим числом переменных на более простые части

Как следствие -- ограничения на

- распределения
 - дискретные
 - многомерные нормальные
 - смешанные
- структуры

Bayesian Networks (Байесовские сети)

- Вероятностные графические модели
- Основаны на теореме Байеса:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

> Когда мы получаем новые данные (свидетельства), мы пересчитываем исходные предположения

- Если мы знаем, что у пассажира высокий уровень в программе лояльности, мы пересчитываем его ожидаемую удовлетворенность

Что мы измеряем?

Пример: airlines

Какие факторы могут быть полезными при оценке авиалиний

- Какие вопросы вы бы включили в инструмент?
- Какие данные вы будете собирать?

Можно ли объединить факторы в какие-то группы?
(персональные / поведенческие/...)

Что мы измеряем

- Satisfaction dimensions
 - Seat comfort
 - Cabin service
 - Entertainment
- Segmentation variables:
 - Age
 - Loyalty program membership
 - Distance
 - Cabin class
 - Kind of travel: family / couple / solo / group, business / tourism
 - ...

Пример: [Skytrax](#)

Модели

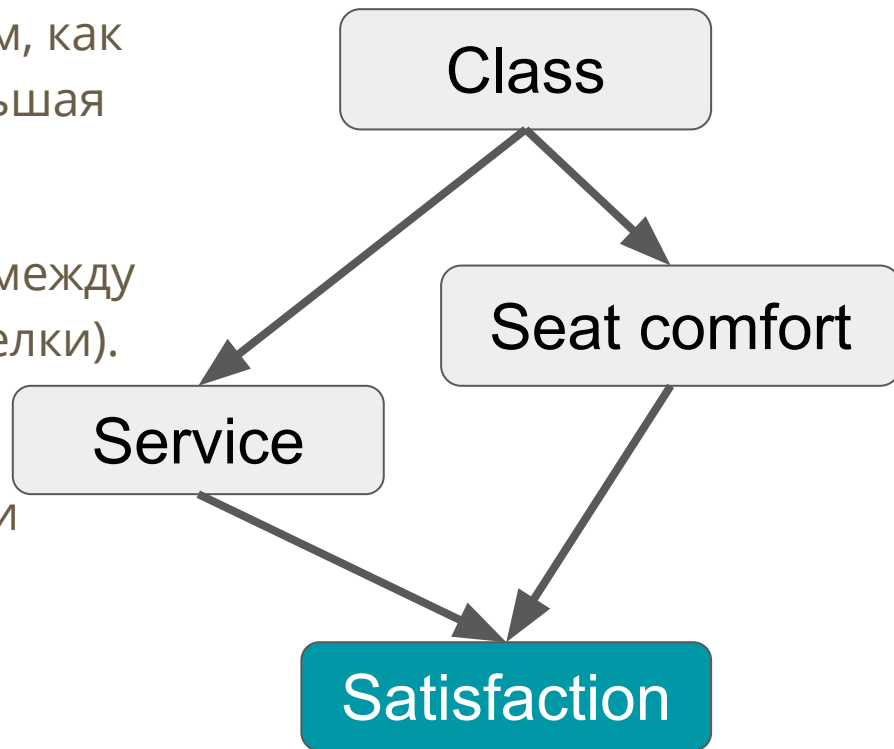
Вероятностные графические модели

“Модель”: наши предположения о том, как устроен мир (как минимум, его небольшая часть)

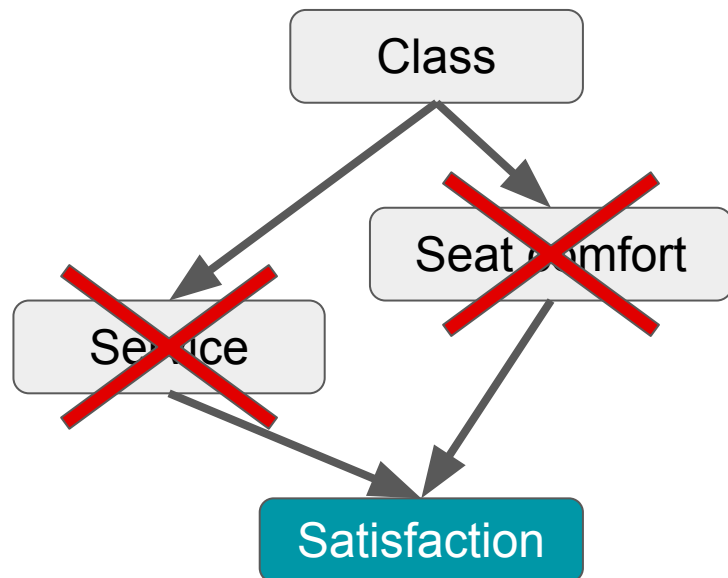
“Графические”: представляем связи между факторами как графы (вершины + стрелки).

- вершины = факторы
- стрелки = непосредственные связи

А как же косвенные связи?



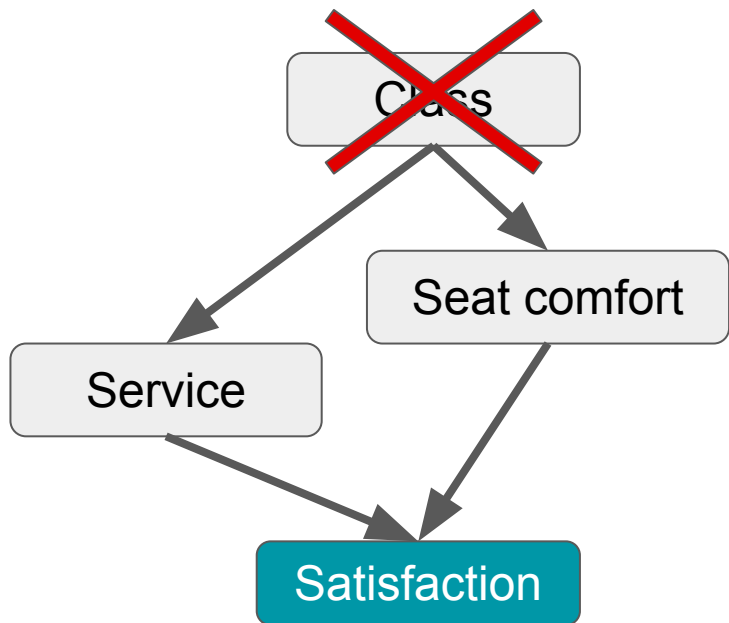
Непосредственные взаимосвязи (1)



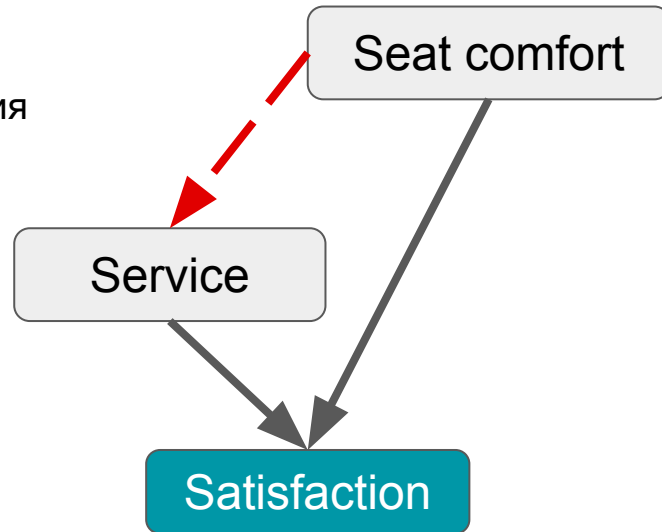
Попарные сравнения



Непосредственные взаимосвязи (2)



Попарные сравнения

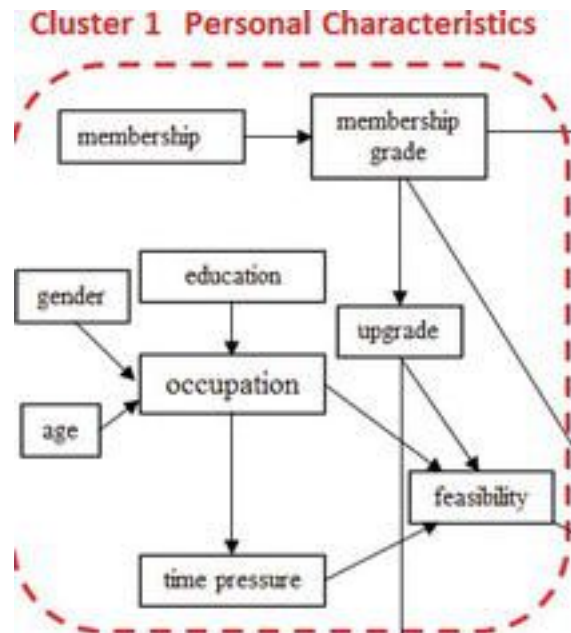


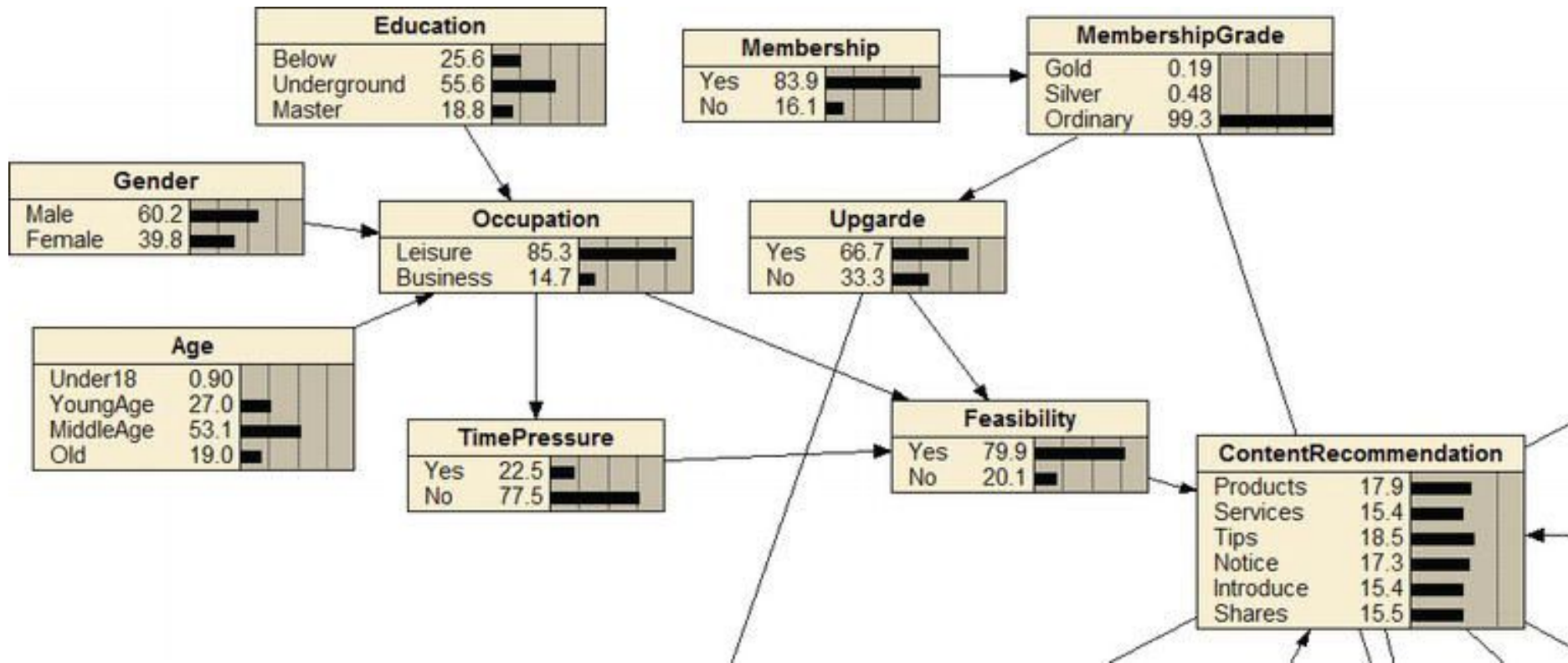
Вероятностные графические модели

“Модель” + “Графическая” (вершины + связи).

“Вероятностная”: наши предположения о связях, выраженные в цифрах (CPT - conditional probability table)

	Работа	Отдых
“Давление” времени	0.8	0.1
Нет давления времени	0.2	0.9

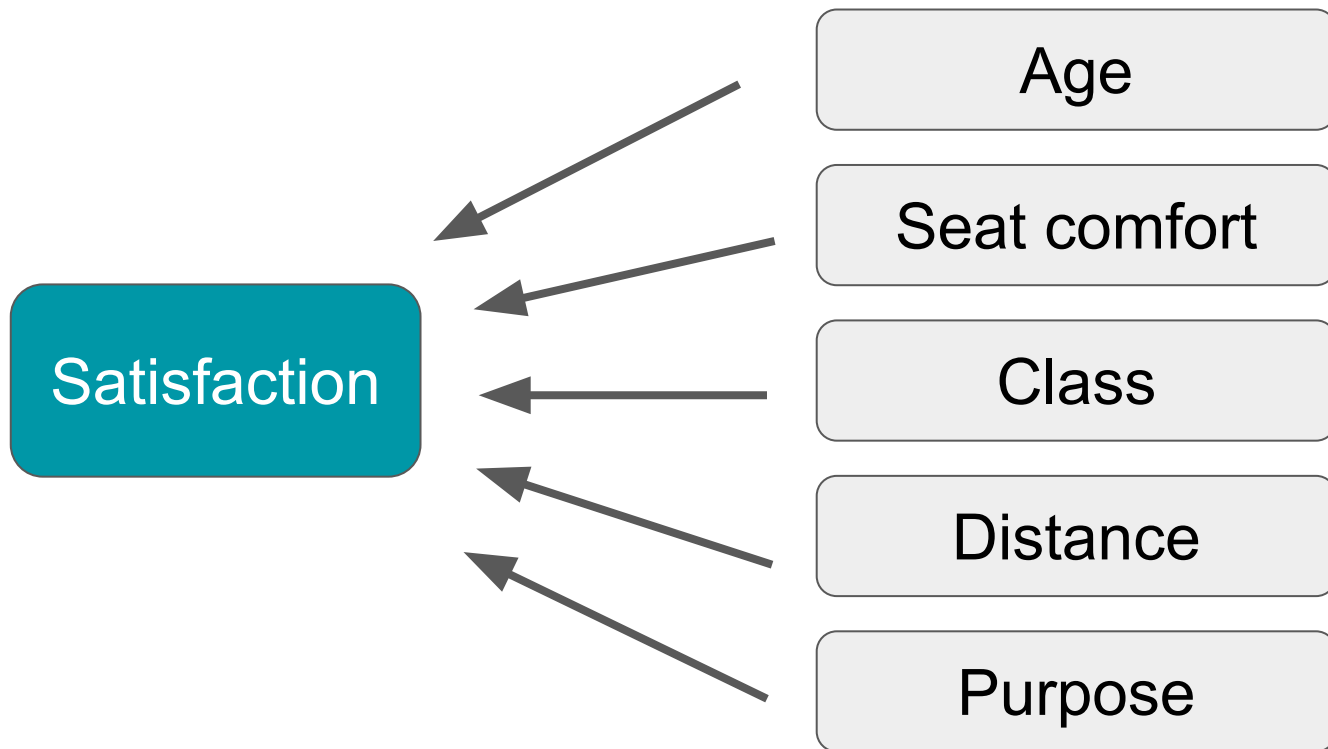




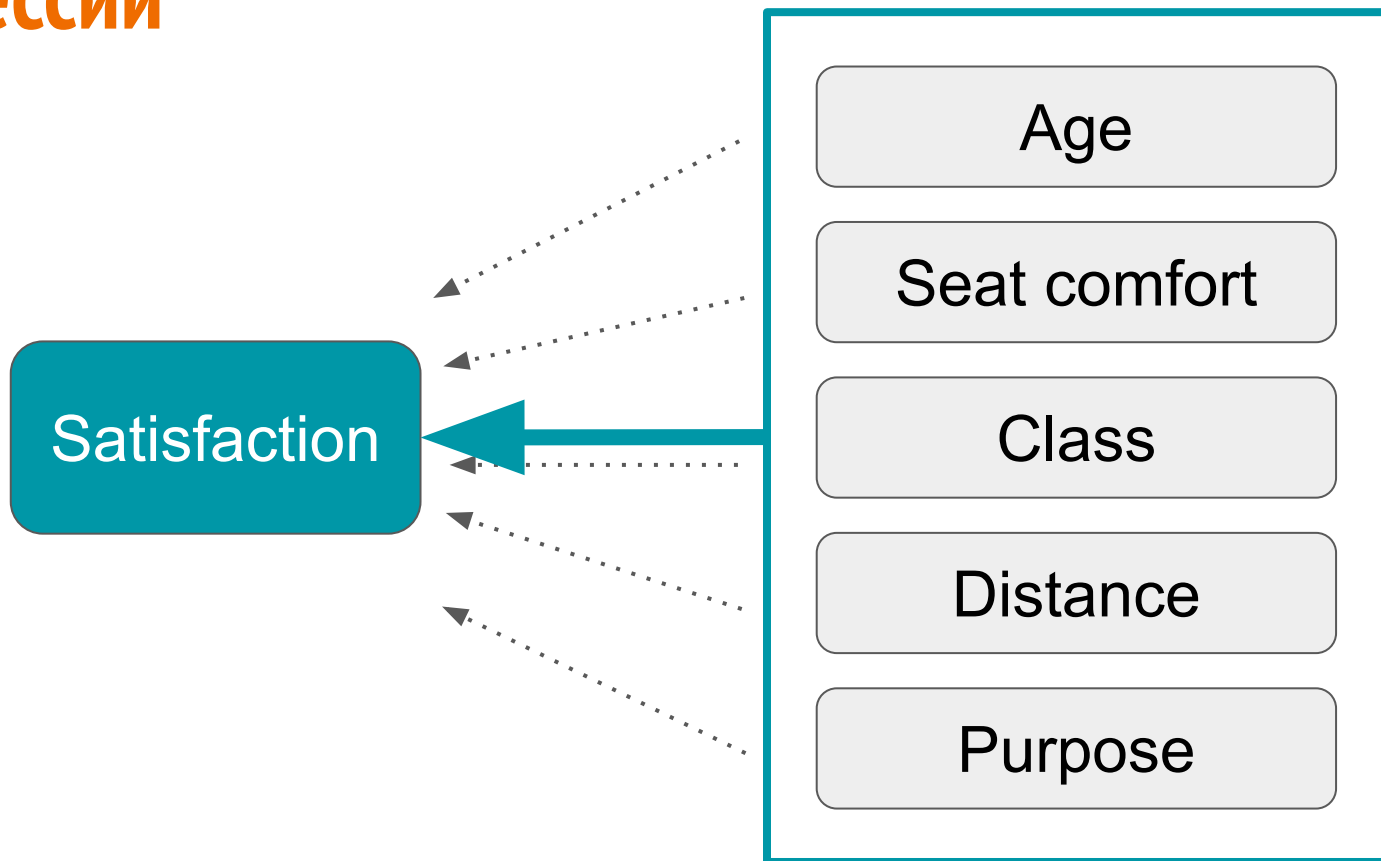
[Airlines Content Recommendations Based on Passengers' Choice Using Bayesian Belief Networks](#)

Отступление: моделирование удовлетворенности

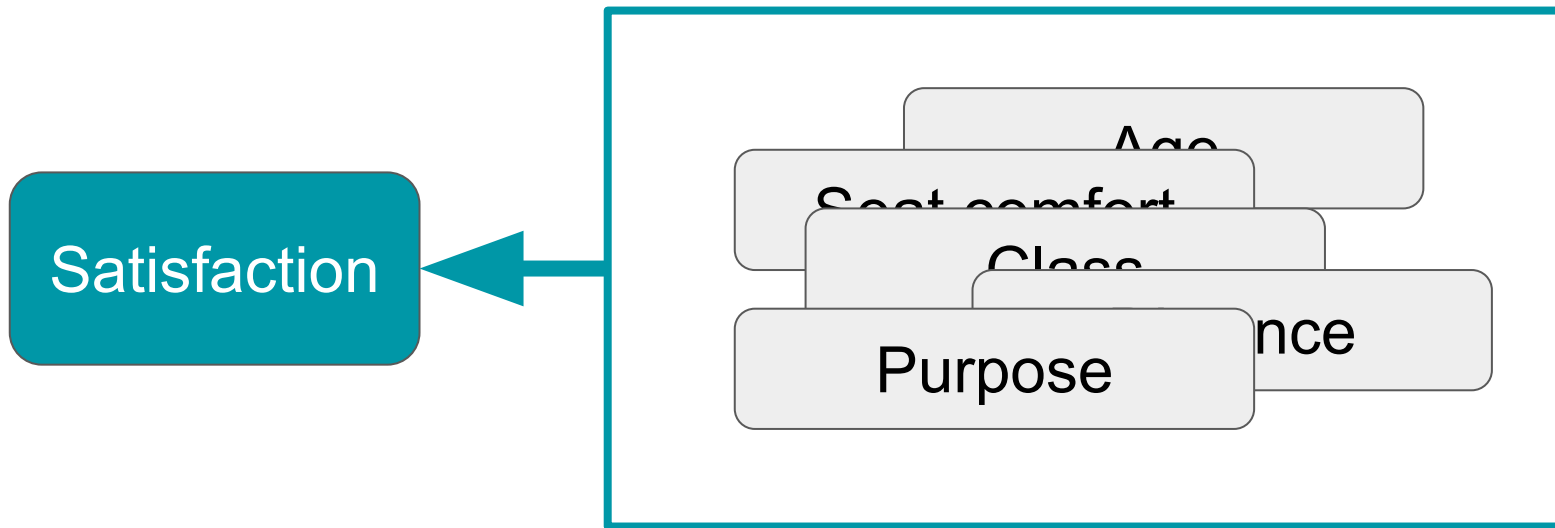
Попарные тесты



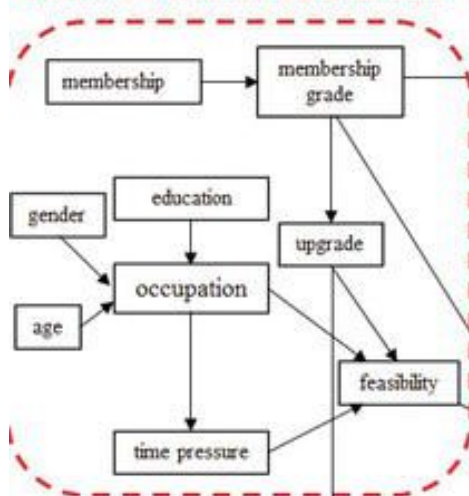
Регрессии



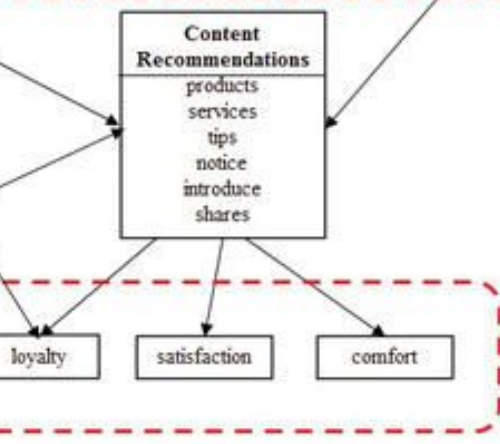
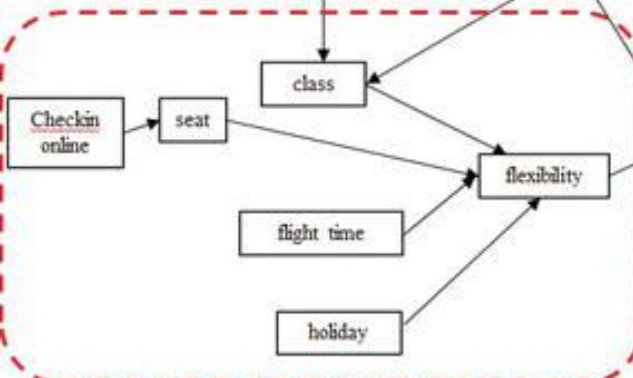
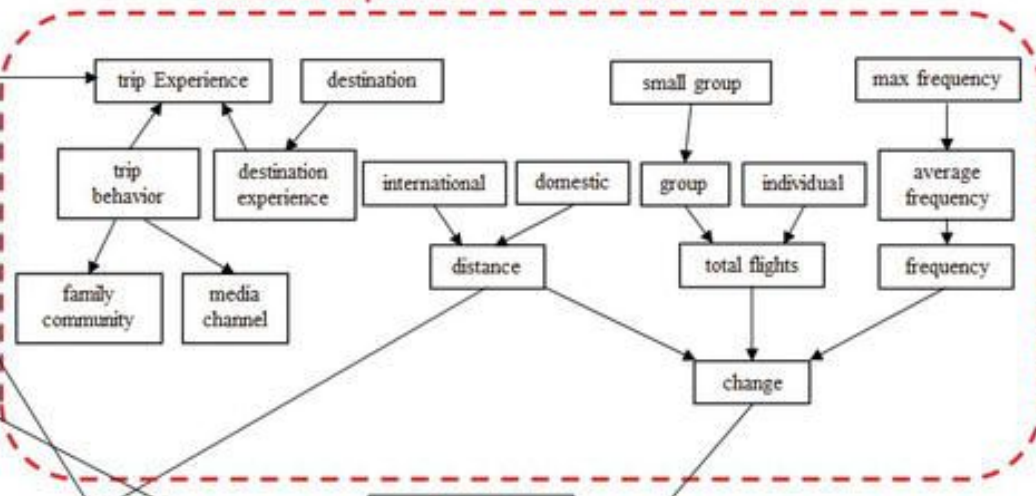
Машинное обучение (black-box models)



Cluster 1 Personal Characteristics



Cluster 2 Experience or behavior Characteristics

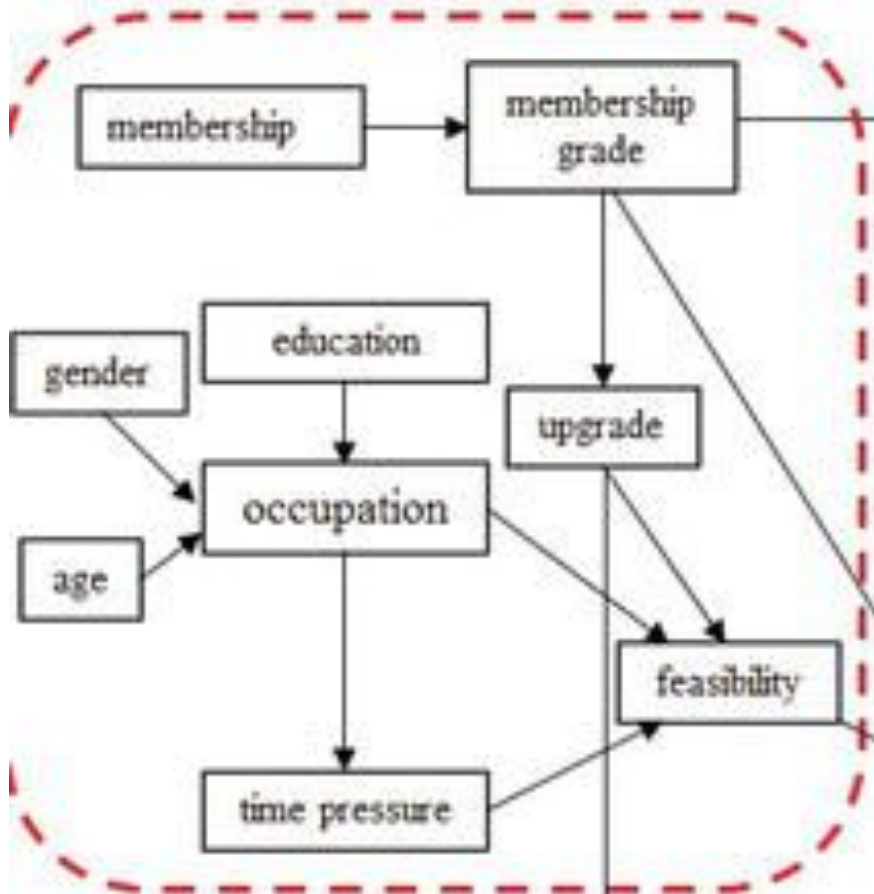


Cluster 3 Preference Characteristics

Cluster 4 Individual Perception

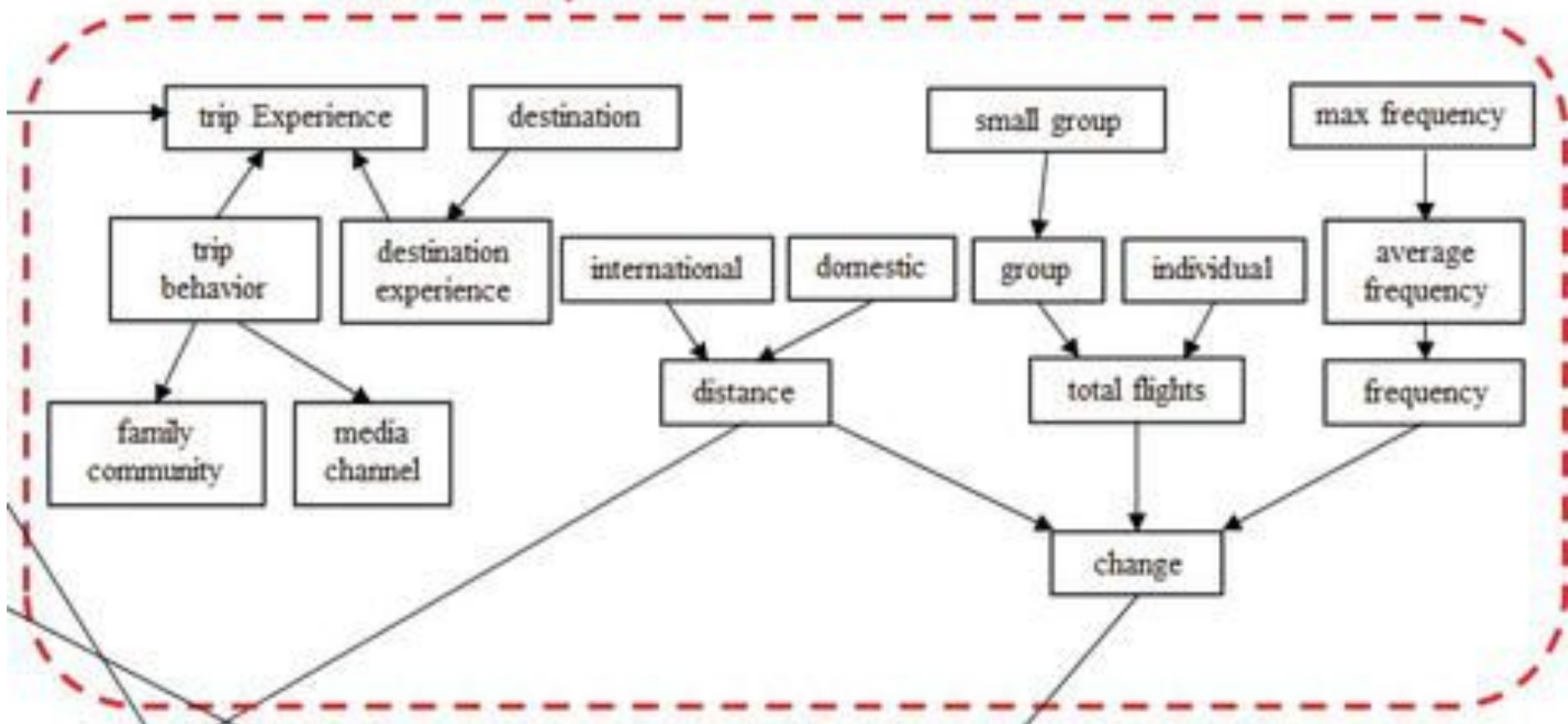
[Airlines Content Recommendations Based on Passengers' Choice Using Bayesian Belief Networks](#)

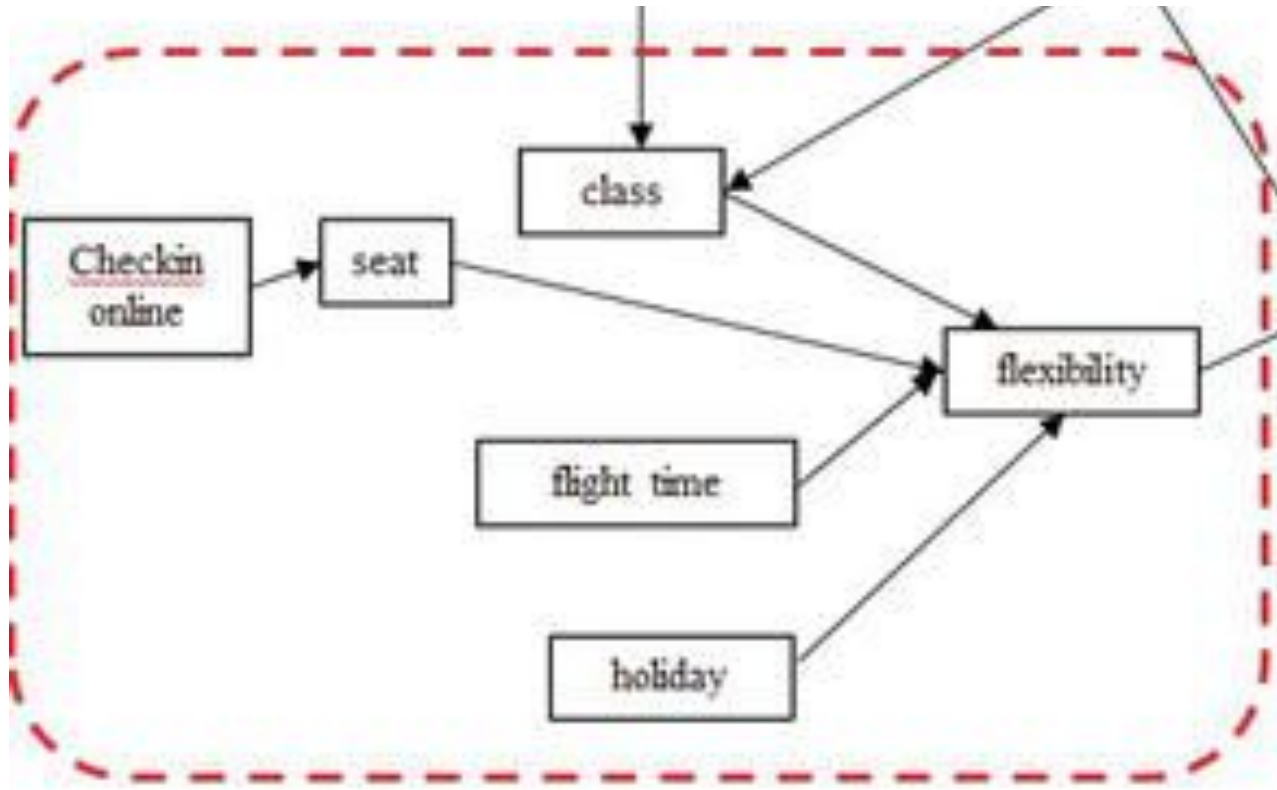
Cluster 1 Personal Characteristics



[Airlines Content Recommendations Based on Passengers' Choice Using Bayesian Belief Networks](#)

Cluster 2 Experience or behavior Characteristics

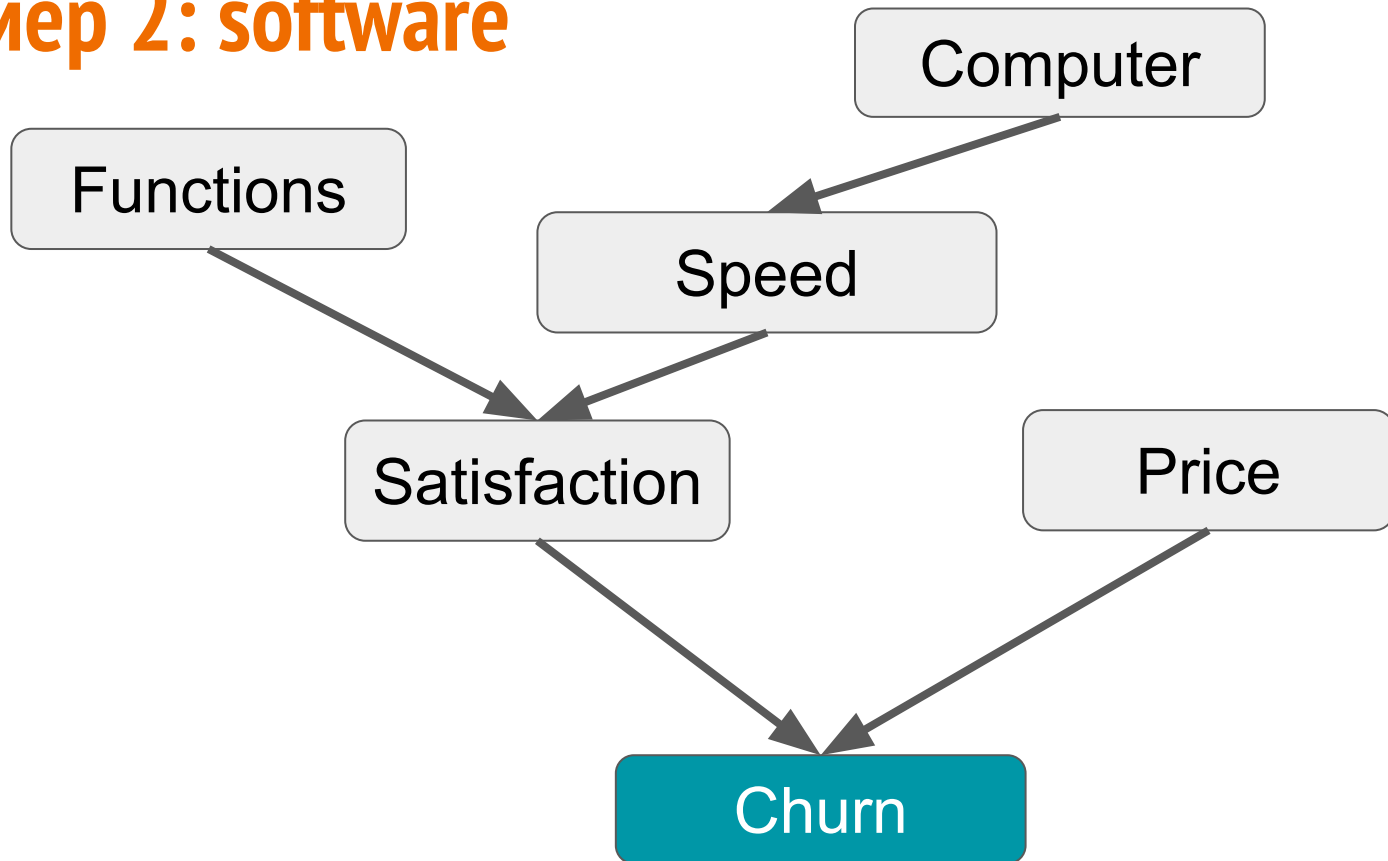




Cluster 3 Preference Characteristics

[Airlines Content Recommendations Based on Passengers' Choice Using Bayesian Belief Networks](#)

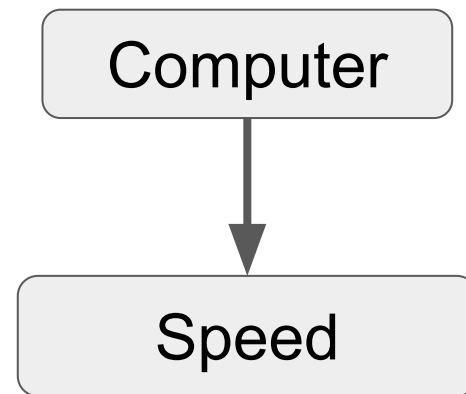
Пример 2: software



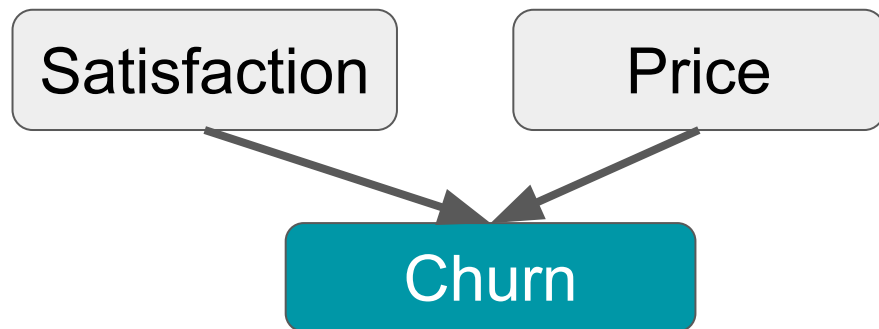
Еще раз про вероятности

Пример 2: software

	Computer		
Speed	Low	Medium	High
Low	0.8	0.5	0.1
High	0.2	0.5	0.9



Пример 2: software



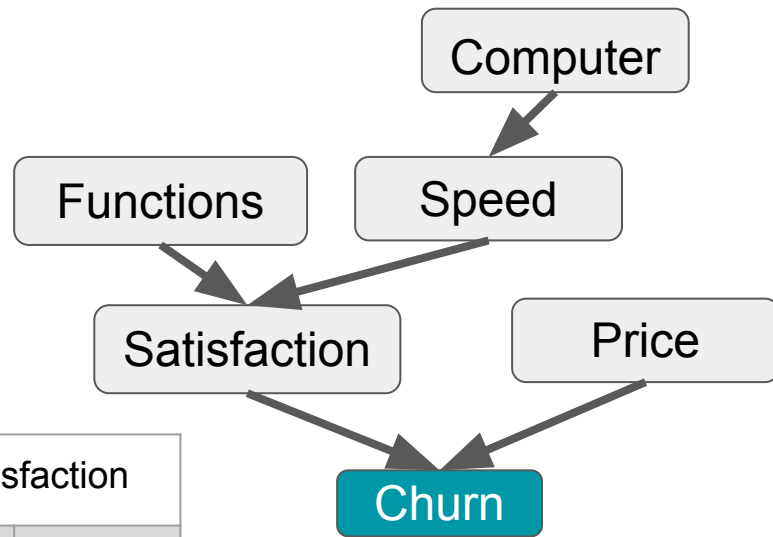
Price = Low	Satisfaction	
Churn	Low	High
Yes	0.4	0.2
No	0.6	0.8

Price = High	Satisfaction	
Churn	Low	High
Yes	0.8	0.5
No	0.2	0.5

Решения и предсказания

Предсказать отток (churn) пользователей

- Для упрощения вычислений:
 - 50:50 низких и высоких цен
 - 30:70 низкой и высокой удовлетворенности
- $P(\text{Churn} = \text{Yes}) = 0.4 \cdot 0.3 \cdot 0.5 + 0.2 \cdot 0.7 \cdot 0.5 + 0.8 \cdot 0.3 \cdot 0.5 + 0.5 \cdot 0.7 \cdot 0.5 = 0.425$

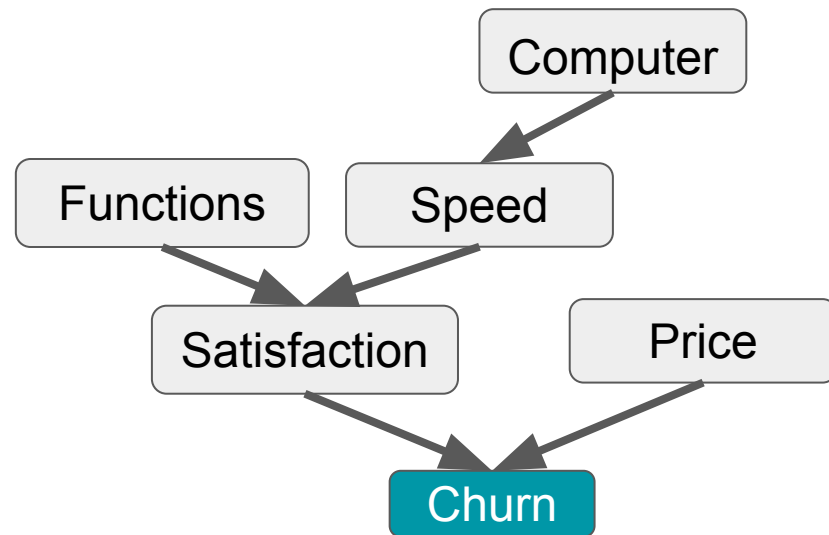


Price = Low	Satisfaction	
Churn	Low	High
Yes	0.4	0.2
No	0.6	0.8

Price = High	Satisfaction	
Churn	Low	High
Yes	0.8	0.5
No	0.2	0.5

Изменения, которые зависят от нас

- Изменить цену на низкую
 - Изменится ли структура?
 - Изменяются ли вероятности?
 - Изменится ли предсказание?



Изменения, которые зависят от нас

- Изменится ли структура? -- НЕТ
- Изменяются ли вероятности? -- ДА, одна из таблиц станет ненужной
- Изменится ли предсказание? -- ДА

Price = Low	Satisfaction	
Churn	Low	High
Yes	0.4	0.2
No	0.6	0.8

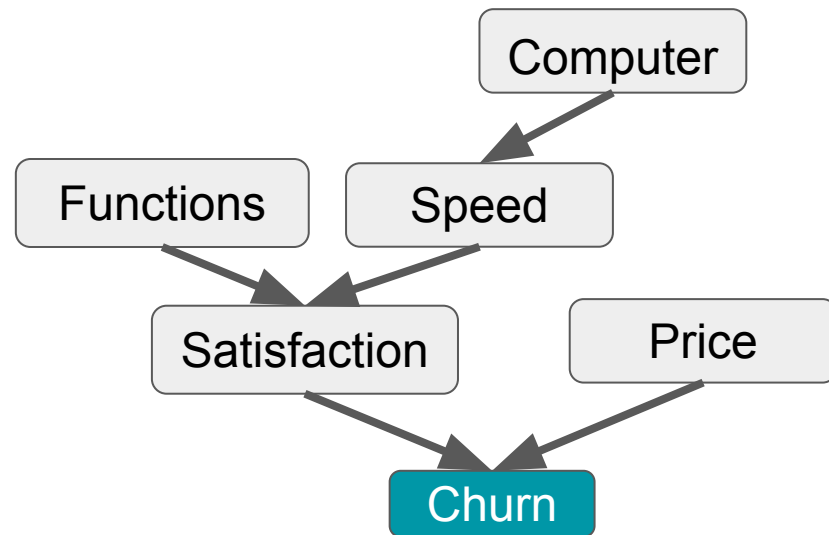
Price = High	Satisfaction	
Churn	Low	High
Yes	0.8	0.5
No	0.2	0.5



○ $P(\text{Churn} = \text{Yes}) = 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26$

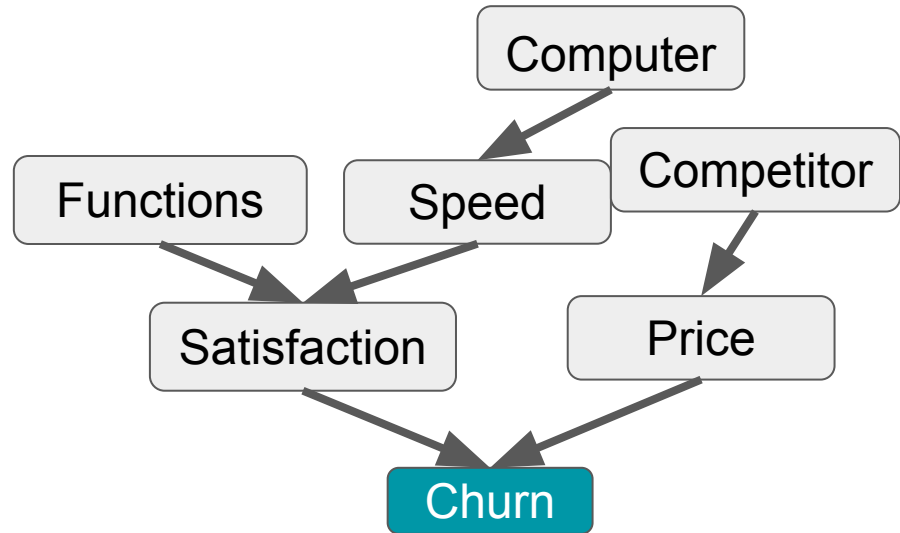
Изменения внешнего мира

- У конкурентов низкая цена
 - Изменится ли структура?
 - Изменяются ли вероятности?
 - Изменится ли предсказание?



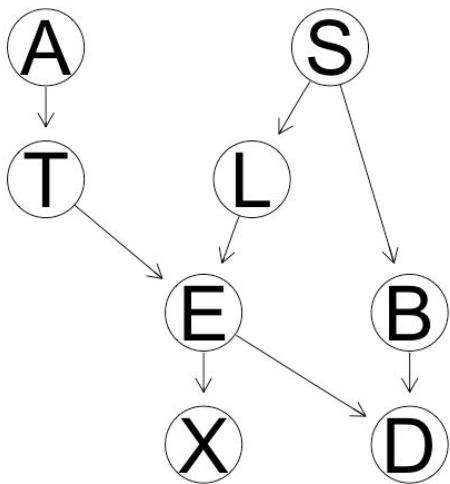
Изменения внешнего мира

- У конкурентов низкая цена
 - Изменится ли структура? --
Возможно
 - Изменяются ли вероятности? --
Возможно
 - Изменится ли предсказание?
-- ДА



Структура модели

Элементы байесовской сети

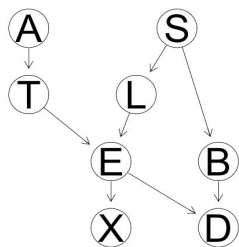


Структура



Условные вероятности

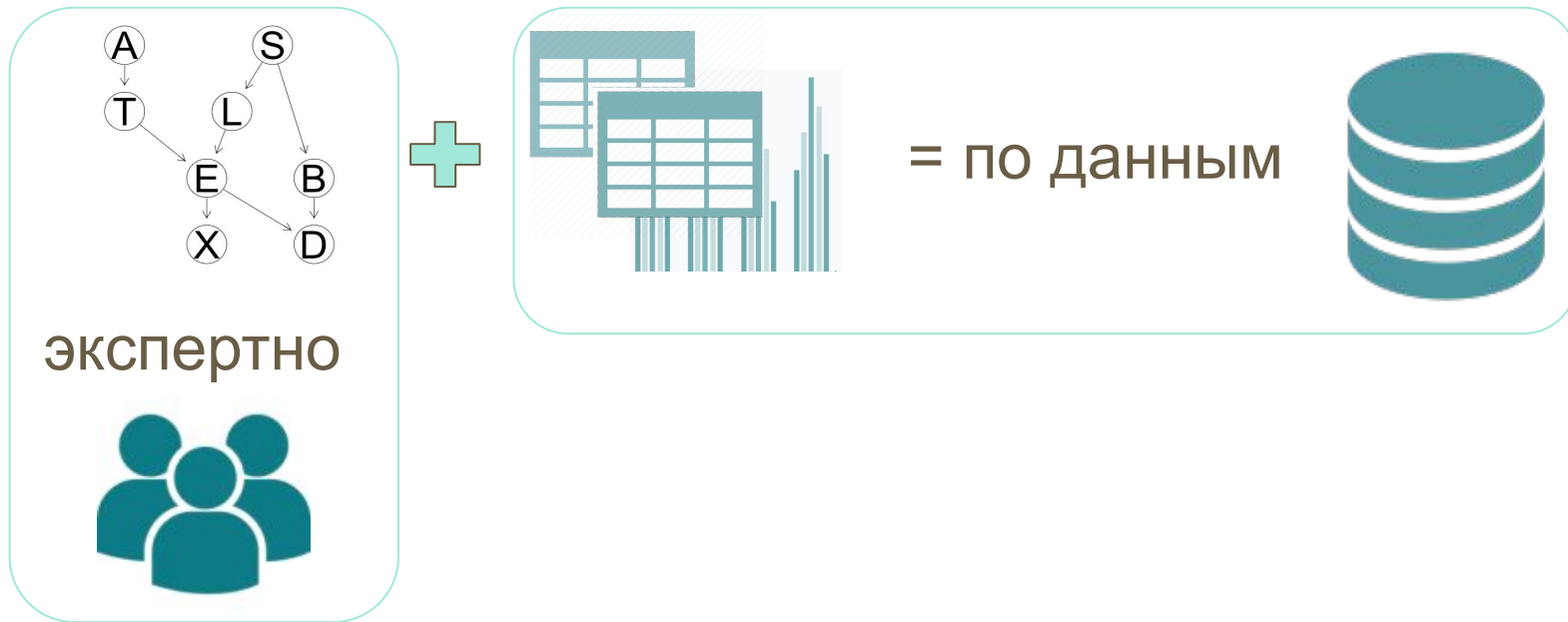
Варианты построения



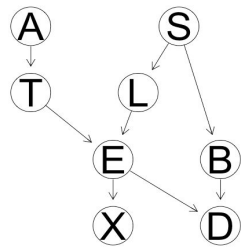
= экспертно



Варианты построения



Варианты построения



= по данным

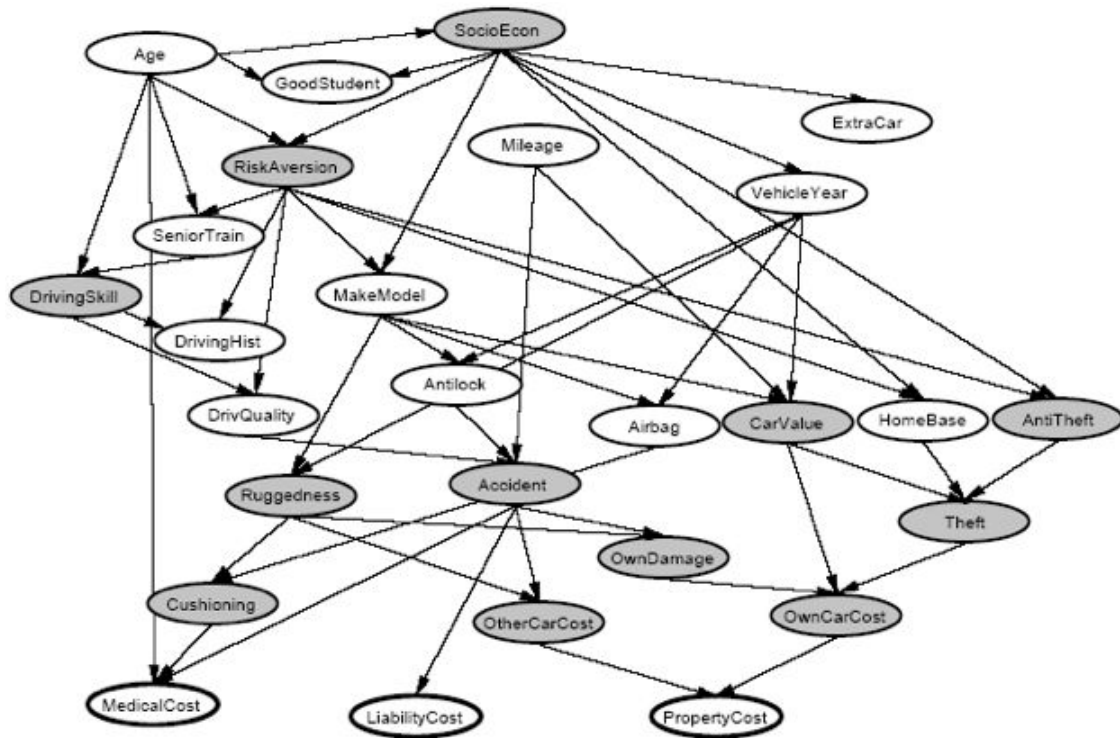


Как построить байесовскую сеть

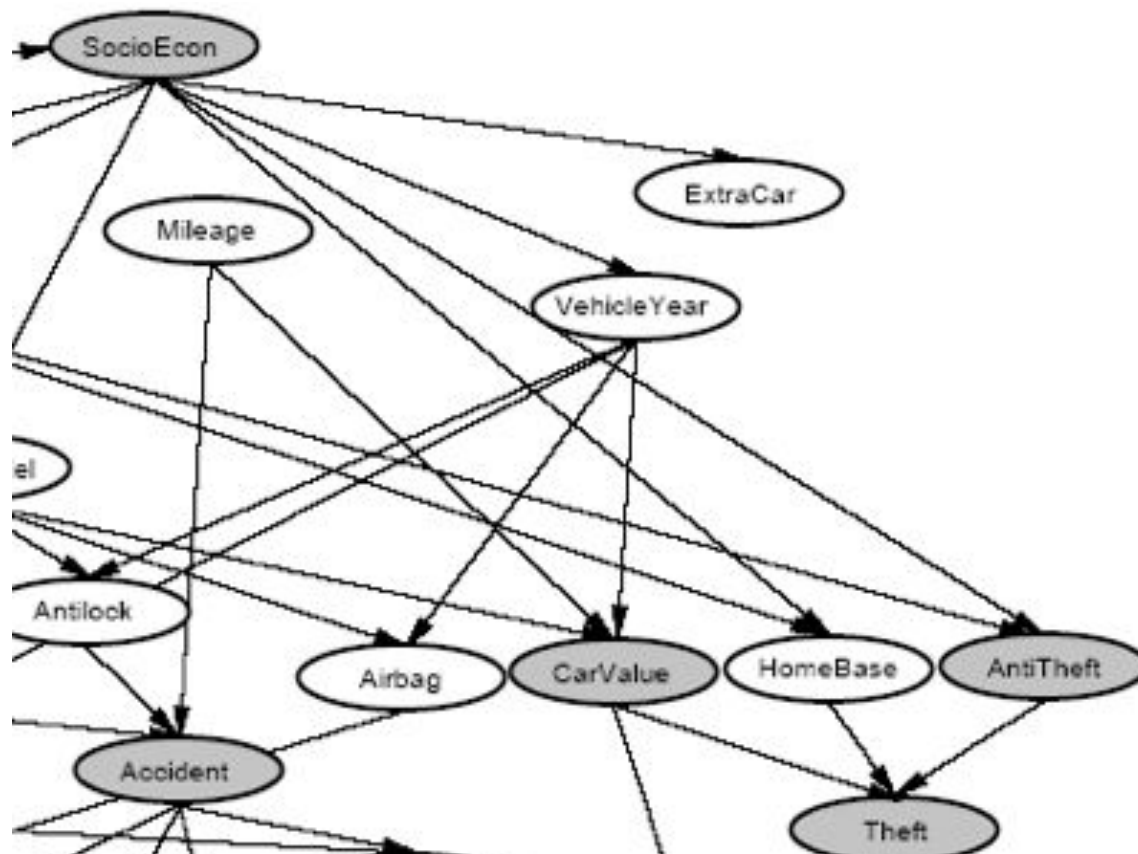
Bayesian Network: как построить модель

- Определите цель моделирования
- Разработайте структуру модели
 - Определите ключевые факторы
 - Определите остальные факторы
 - Определите, какие значения есть у факторов
 - Определите связи
- Определите параметры модели (вероятности)
- Исследуйте модель
 - анализ чувствительности
 - анализ влияния переменных
 - сценарный анализ или “what-if”
- Валидируйте модель

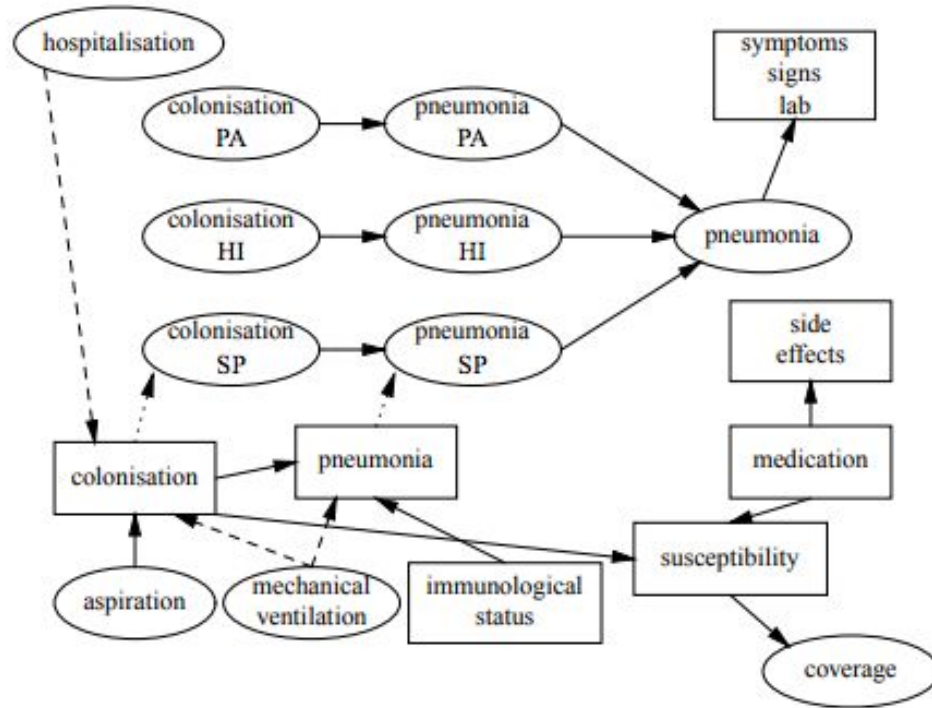
Примеры



Страхование Binder J, Koller D, Russell S, Kanazawa K (1997)

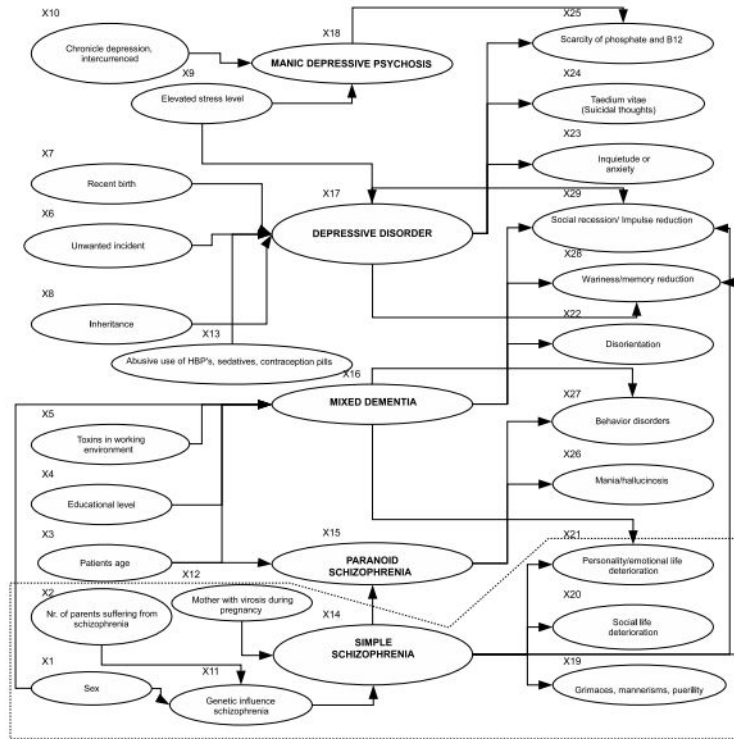


Страхование Binder J, Koller D, Russell S, Kanazawa K (1997)



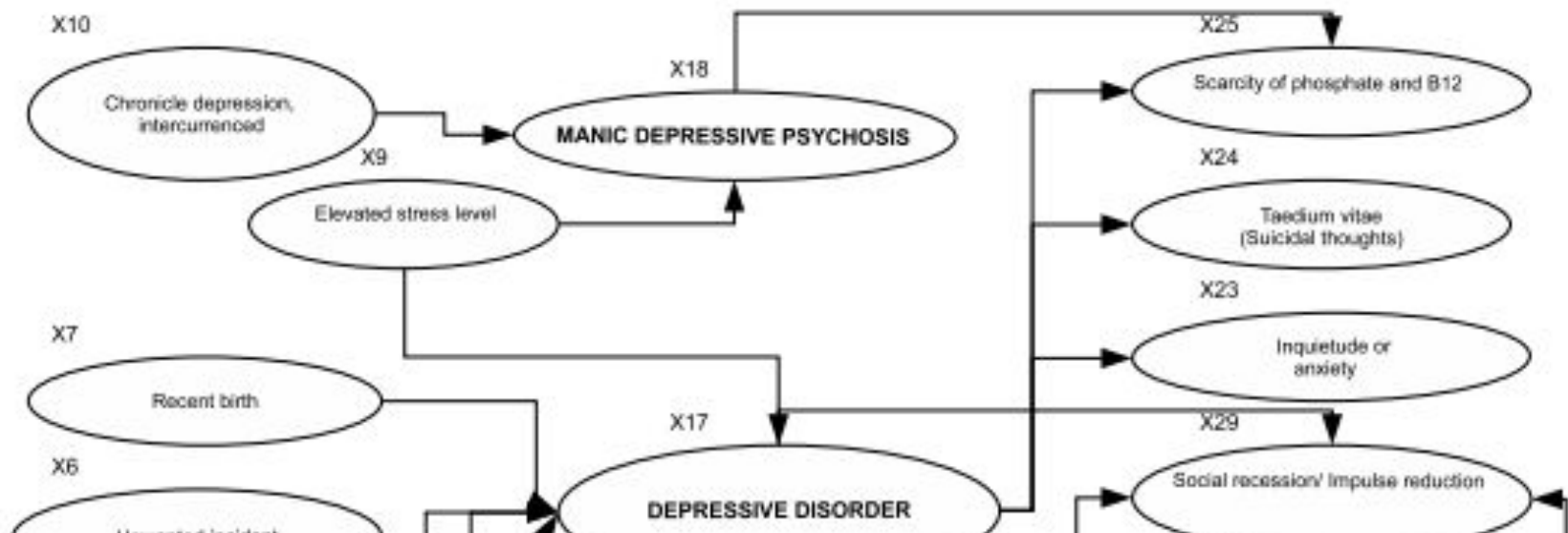
Медицина и диагностика

(Lucas P. Bayesian networks in medicine: a model-based approach to medical decision making)



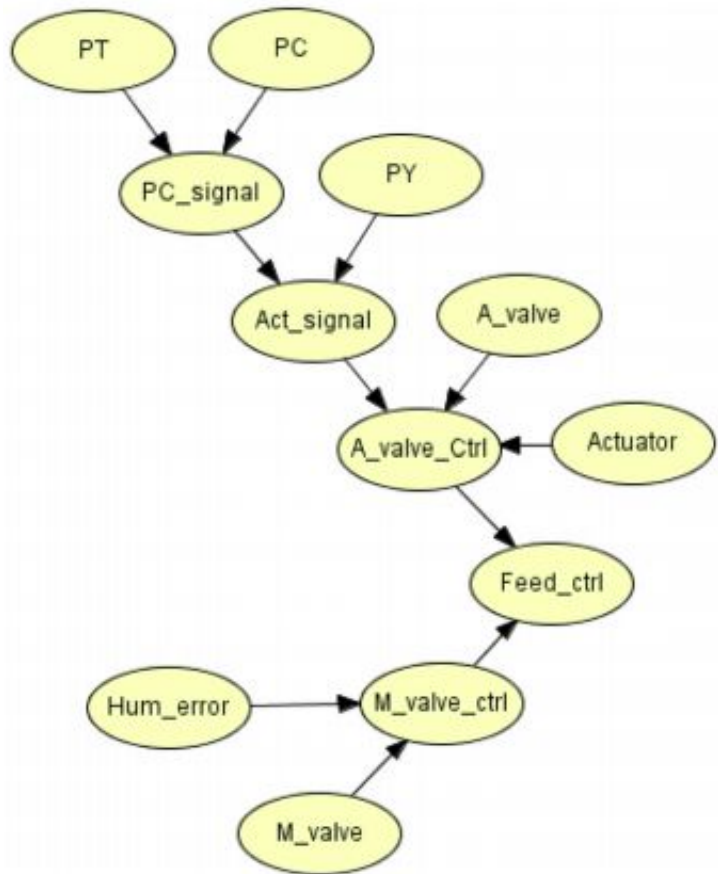
Медицина и диагностика - шизофрения

(Curiac D. I. et al. Bayesian network model for diagnosis of psychiatric diseases, 2009)



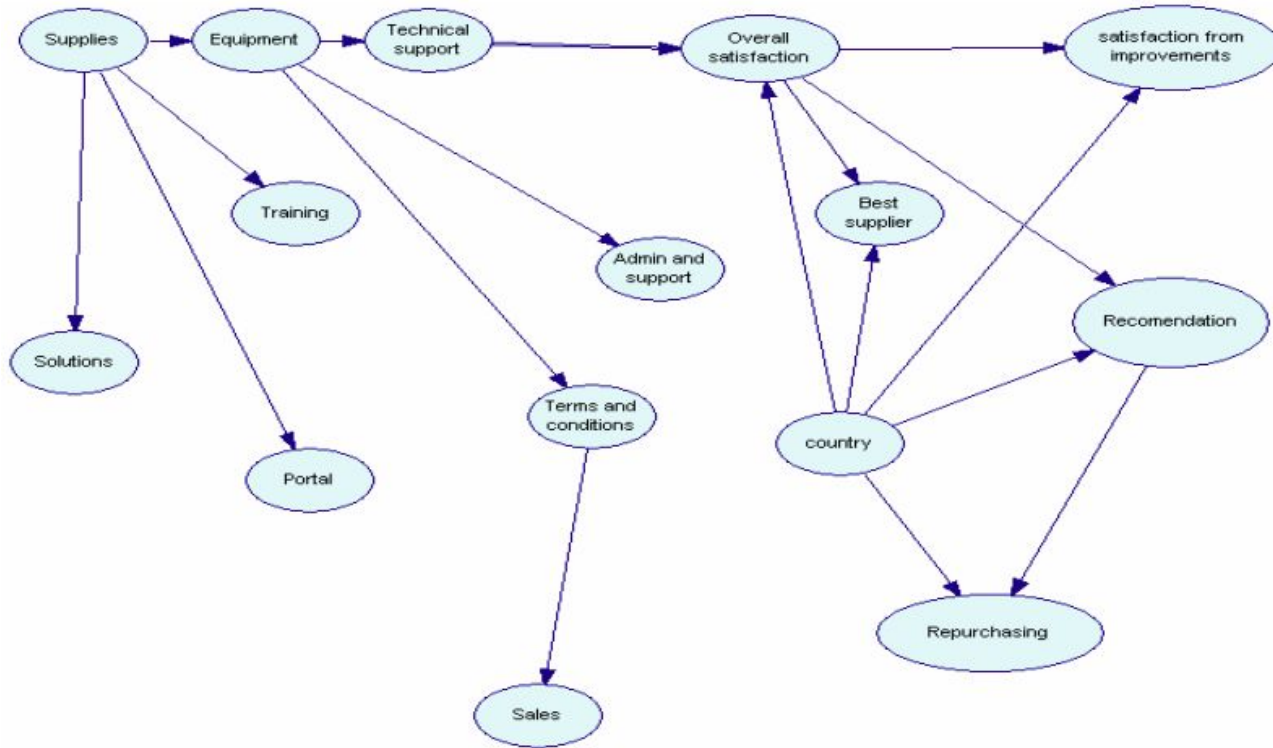
Медицина и диагностика - шизофрения

(Curiac D. I. et al. Bayesian network model for diagnosis of psychiatric diseases, 2009)



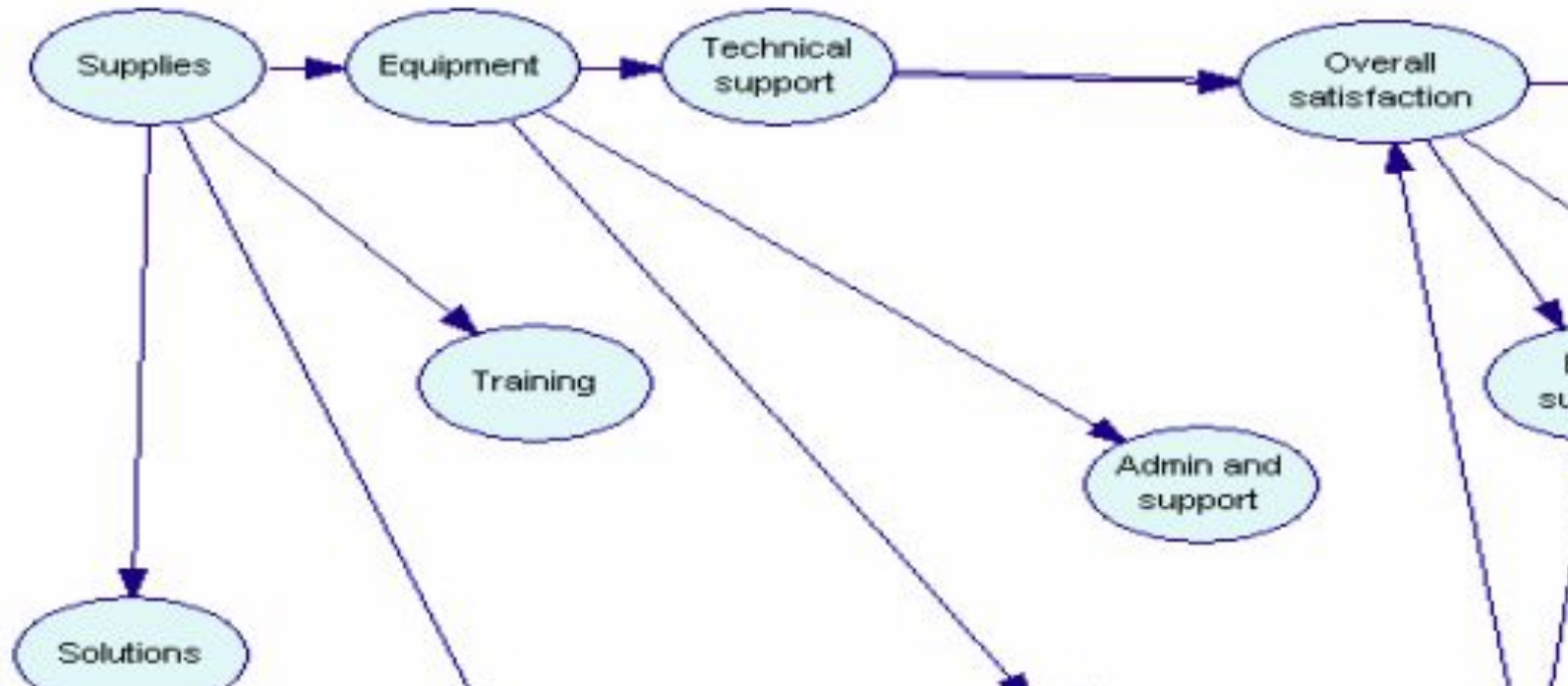
Надежность систем

(Khakzad N., Khan F., Amyotte P. Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches, 2012)



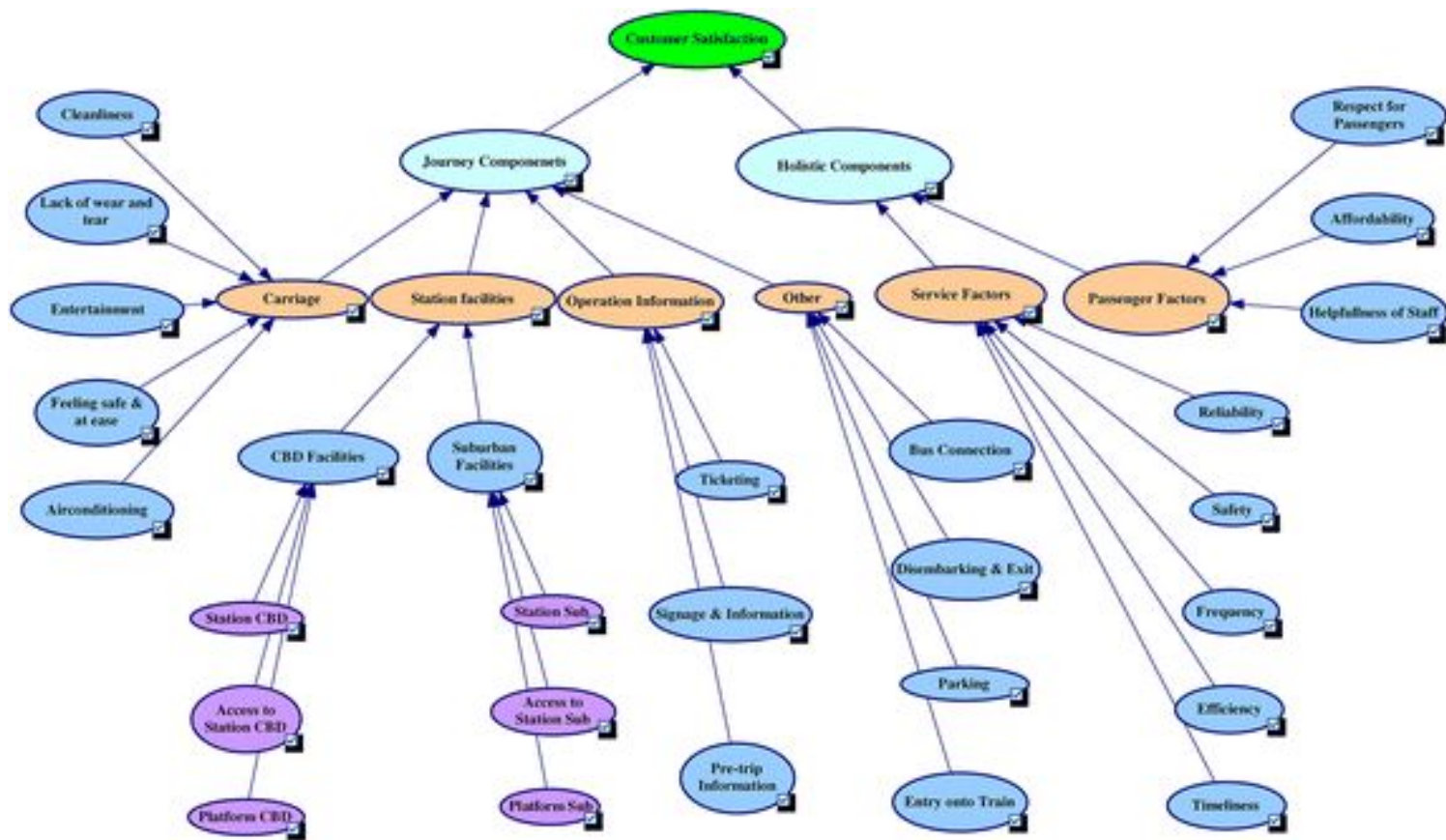
Customer Satisfaction Survey

(Salini S., Kenett R. S. Bayesian networks of customer satisfaction survey data, 2009)



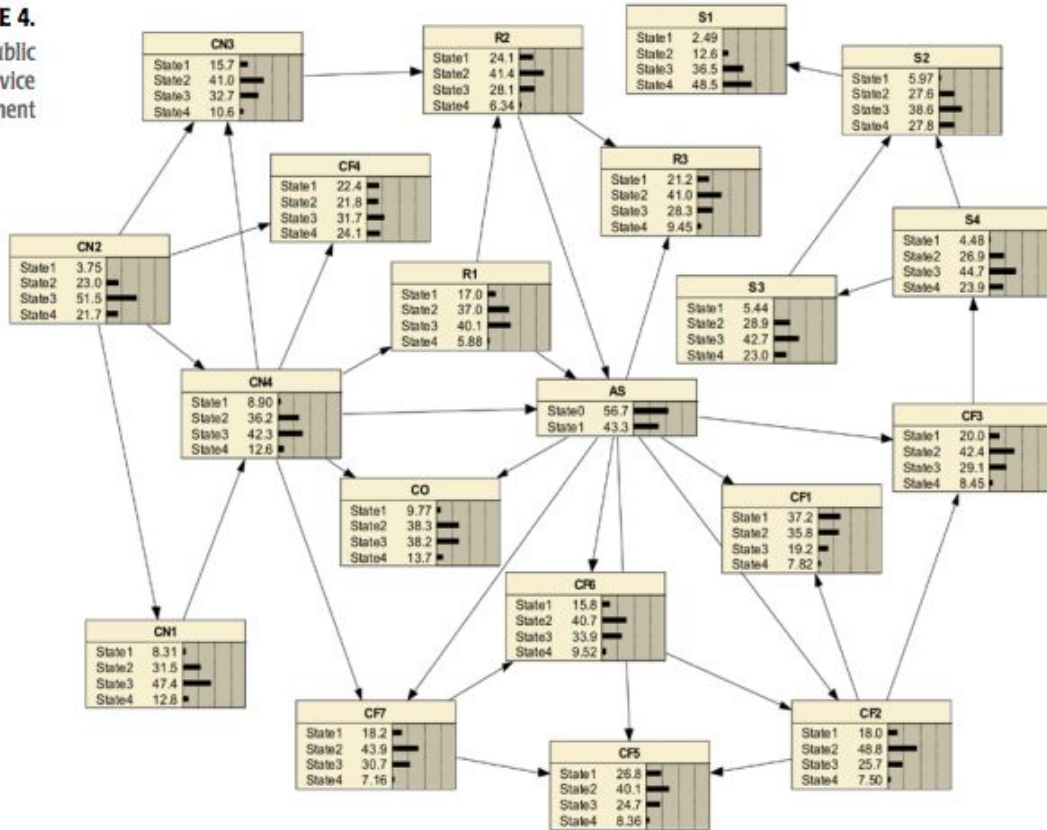
Customer Satisfaction Survey

(Salini S., Kenett R. S. Bayesian networks of customer satisfaction survey data, 2009)



A Bayesian Network-based customer satisfaction model: a tool for management decisions in railway transport

FIGURE 4.
Network of public
transit service
assessment



Пробуем!

<https://jiajin.shinyapps.io/shinyBN/>



Полезные ресурсы

- Nagarajan R., Scutari M., Lèbre S. Bayesian networks in R, 2013
- Højsgaard S., Edwards D., Lauritzen S. Graphical models with R, 2012
- Специализация Coursera [Probabilistic Graphical Models](#)
- Введение в DAG: Graphical Causal Models by Felix Elwert
<https://www.ssc.wisc.edu/soc/faculty/pages/docs/elwert/Elwert%202013.pdf> из [Handbook of Causal Analysis for Social Research](#), 2013
- Тьюториалы и интерактивные примеры <http://dagitty.net/learn/index.html>