

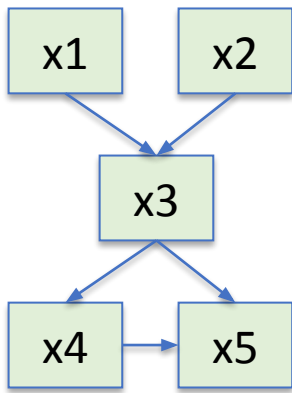


ІТМО

Обзор open source решения для прикладного использования байесовских сетей

Никита Ковалев
Инженер команды ProbAI в NSS Lab

Байесовская сеть – направленный ациклический граф, в узлах которого располагаются распределения признаков, а ребра обозначают условные зависимости между признаками.



$$x_1 \perp x_4 | x_3$$

$$p(\mathbf{X}) = \prod_i p(x_i | x_1, \dots, x_{i-1})$$

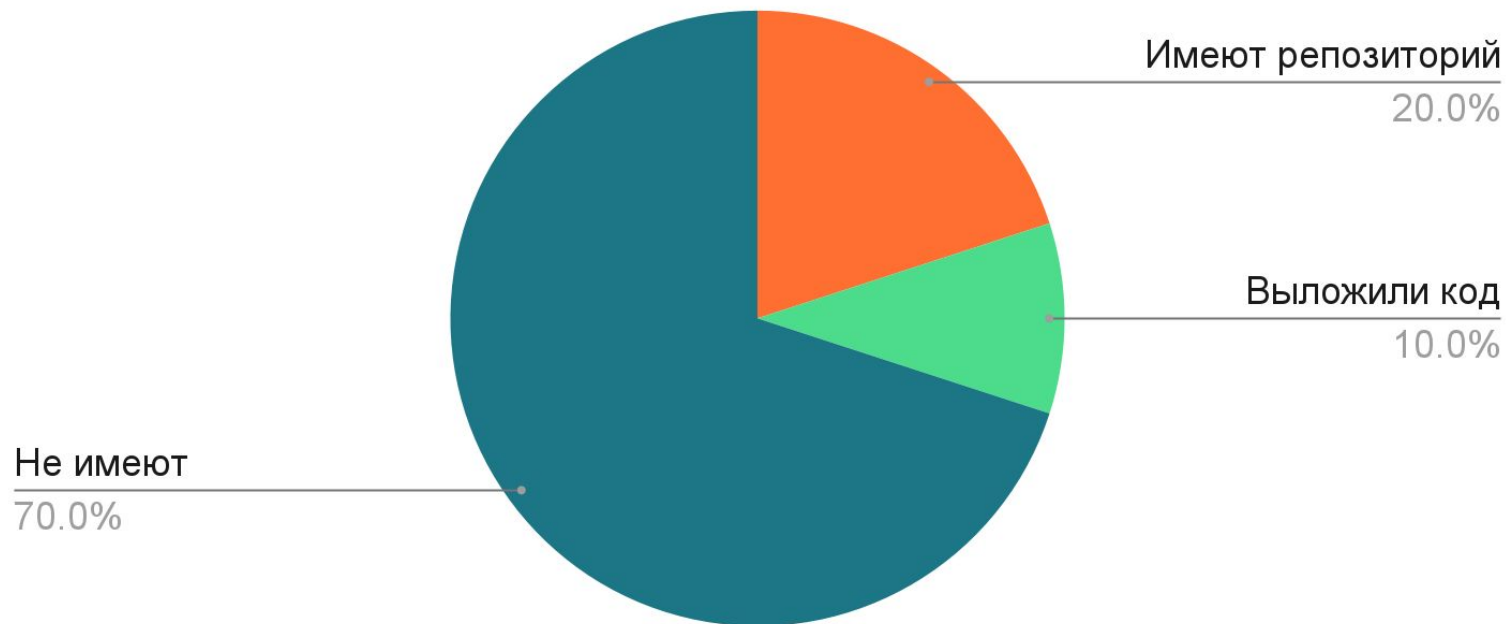
$$p(x_1, \dots, x_5) = p(x_1)p(x_2|x_1) \times \dots \\ \dots \times p(x_3|x_1, x_2)p(x_4|x_3) p(x_5|x_3, x_4)$$

Какого рода задачи мы решаем?

ІТМО

- Предсказание
- Генерация синтетики
- Заполнение пропусков
- Балансировка классов
- Извлечение признаков
- Генерация признаков
- Детектирование аномалий

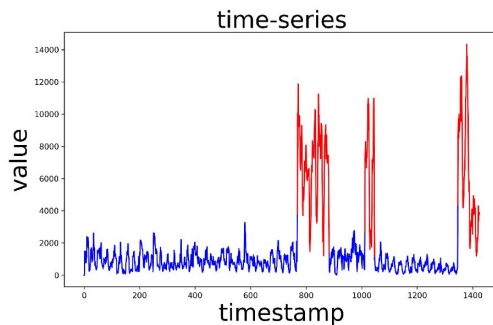
Количество проектов



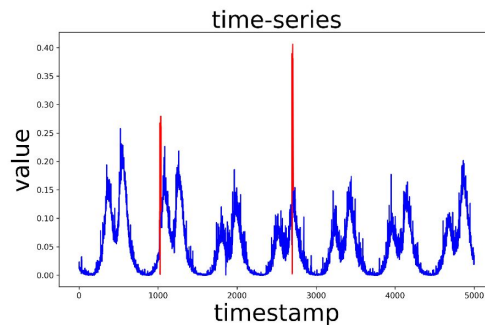
Опенсор проекты

TraceAnomaly¹
CPL1.0²

300 звезд, 62 форка
Язык: Python



(a)



(b)

Плюсы:

- Популярность
- Наличие докера
- Хороший код
- Инструкция по запуску

Минусы:

- Тесты
- Документация
- Сопровождение
- Формат – коробка

1 - <https://github.com/NetManAIOps/TraceAnomaly>

2 - <https://github.com/mihiranpathmika/CPL1.0>

Генерация синтетики

DataSynthesizer¹

PrivBayes²

Privbayes-nist-jpc³

BayNet⁴

Плюсы:

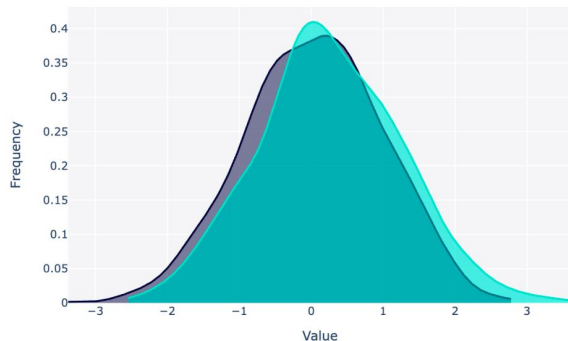
- Популярность
- CI
- PyPI
- WebUI

Минусы:

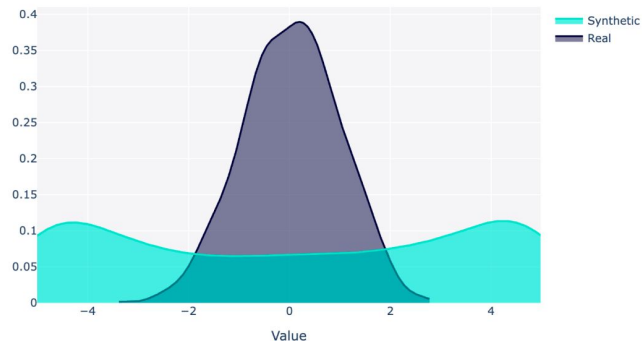
- Тесты
- Документация
- Формат – коробка

226 звезд, 78 форков, Python

Real vs. Synthetic Data (Score=0.9)



Real vs. Synthetic Data (Score=0.63)



1 – <https://github.com/DataResponsibly/DataSynthesizer>

2 – <https://sourceforge.net/projects/privbayes/>

3 – <https://github.com/journalprivacyconfidentiality/privbayes-nist-jpc/tree/jpc.776?tab=readme-ov-file>

4 – <https://github.com/Stoffle/BayNet>

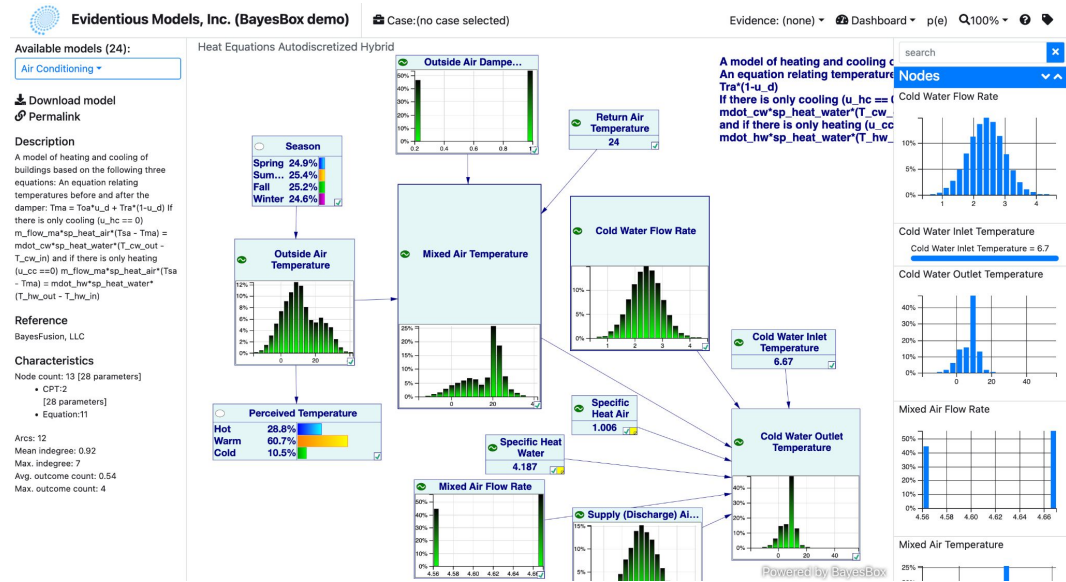
Чем пользуются прикладники?



GeNIe
Netica
AgenaRisk
HuginExpert

- UI
- Nocode
- Инфографика
- Разные темы
- Поддержка

ПОРОГ ВХОЖДЕНИЯ



Part	Description	Price
120-7	Netica Application for Windows & Mac (single-user license)	\$685.00
310-7	Netica API for Windows, Linux, AIX, MacOS X, etc. (single-user license)	\$785.00

Что-то бесплатное? Ну, да

BNT¹
OpenMarkov²

Плюсы:

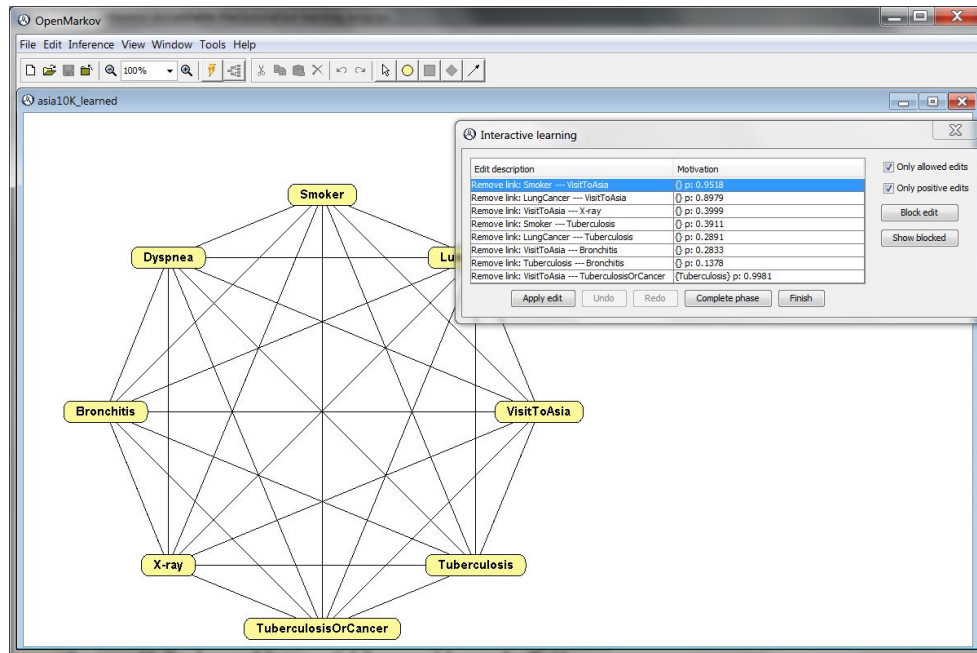
- Популярность
- Документация
- Инструкция по запуску
- Наличие GUI

Минусы:

- Тесты
- Документация
- Сопровождение

1 - <https://github.com/bayesnet/bnt>

2 - <http://www.openmarkov.org/>



452 звезды, 252 форка, MATLAB

- Построение структур сетей на основе данных различными алгоритмами, построение композитных БС с моделями МО;
- Обучение параметров распределений в узлах сети;
- Поддержка различных типов данных (дискретные, непрерывные);
- Генерация синтетических данных;
- Поиски нетривиальных зависимостей в данных (нелинейных);
- Модульность, расширяемость, интегрируемость с ML-инструментами;
- Сочетание легковесного API для конечного пользователя и расширенного конфигурирования для исследовательских задач.



Идея: Создать open-source проект с решением прикладных задач на основе БС.

- В основе – ВАМТ
- Многофункциональность – объединяем инструменты решений возможных задач
- Повышаем качество решений, в том числе и интерпретируемость
- Соблюдаем “стандарты” хорошего Open-Source продукта
- Снижаем порог вхождения

**Спасибо
за внимание!**

ITMO *re than a*
UNIVERSITY