

ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ ЗАБОЛЕВАНИЯ  
НА ОСНОВЕ МЕТОДОВ DATA MINING

*Горанина Сергей Игоревич, студент группы 345,  
Базилевич Ксения Алексеевна, доцент кафедры информатики  
Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»*

В современном мире, когда люди страдают различными болезнями, многие специалисты ищут способы и методы их лечения и диагностирования. Решение проблемы, которая заключается в ограниченных возможностях своевременного диагностирования, лежит в области методов Data Mining.

Первым из методов, рассматриваемых в исследовании, является метод оценивания параметров логистической регрессии на основе метода оценки шансов и вероятностей [1]. Метод выгоднее всего использовать для выборок с малым количеством параметров. На выборке с большим количеством параметров этот метод перестает быть актуальным и теряет свою точность. Вторым рассматриваемым методом является метод оценки вероятности заболевания с использованием байесовского классификатора. Этот метод выгоднее использовать на выборках с большим количеством параметров, т.к. метод не теряет своей точности при увеличении количества переменных, однако, несмотря на это количество признаков должно быть постоянным. В случае с переменным количеством признаков использование такого классификатора в явном виде приводит к потере его ковариантности.

Также в исследовании рассматривается метод оценивания параметров логистической регрессии на основе метода максимального правдоподобия. Этот метод уже долгое время является одним из лучших для решения задач подобного вида. Это обусловлено рядом причин: актуальность и возможность применения в различных областях, а также возможность реализации метода на современных производительных компьютерах. Недостатком метода является его трудоёмкость.

В результате исследования были проанализированы рассмотренные методы, а также построены соответствующие математические модели. Была проведена оценка точности полученных моделей на выборке данных. Методы были реализованы в программных модулях. При выполнении исследования также были рассмотрены модификации методов, их анализ и проверка требований к системе диагностики.

Список использованной литературы

1. Bailey Norman The mathematical approach to biology and medicine [Text] / Norman T. J. Bailey // John Wiley and Sons Ltd, 1967.