

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования

ЯЦКОВ Владислав Николаевич

**РАЗРАБОТКА МОДЕЛЕЙ, МЕТОДОВ И ПРОГРАММНЫХ СРЕДСТВ
ДЛЯ ПРЕДСКАЗАНИЯ ВЫЖИВАЕМОСТИ ПАЦИЕНТОВ С
ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ**

Аннотация (реферат) магистерской диссертации

специальность 1-98 80 01 «Информационная безопасность»

Научный руководитель:
Скакун В. В.
кандидат физико-математических
наук, доцент

Допущен к защите

«__» _____ 2024 г.

Зав. кафедрой системного анализа и компьютерного моделирования
Яцков Н. Н.

Минск, 2024

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Диссертационная работа: 68 страницы, 18 иллюстраций, 11 таблиц, 55 источников, 3 приложения.

ПРЕДСКАЗАНИЕ ВЫЖИВАЕМОСТИ, БЛОЧНЫЙ ЛЕС, АЛГОРИТМ РАСЩЕПЛЕНИЯ УЗЛОВ EXPONENTIAL LOG-LIKELIHOOD LOSS, БУСТИНГ РЕГРЕССИИ КОКСА, МЕТОД ГЛАВНЫХ КОМПОНЕНТ, С-ИНДЕКС, ОЦЕНКА БРИЕРА, ЭКСПРЕССИЯ ГЕНОВ, ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ, КЛАСТЕРНЫЙ АНАЛИЗ.

Цель работы – разработка имитационных моделей, методов анализа и программных средств для предсказания выживаемости пациентов при онкологических заболеваниях с использованием алгоритмов машинного обучения.

Задачи: обзор литературы, разработка алгоритмов имитационного моделирования и предсказания выживаемости, сравнительный анализ алгоритмов предсказания выживаемости на смоделированных данных, разработка программного обеспечения и концепции системы защиты информации в реализованном программном обеспечении.

Объектом исследования являются генетические данные о выживаемости онкопациентов.

Предметом исследования являются алгоритмы имитационного моделирования и предсказания выживаемости онкопациентов.

Методы исследования: алгоритмы блочного леса с логарифмической вероятностной функцией потерь и регрессии Кокса для предсказания выживаемости; метод главных компонент; С-индекс и оценка Бриера для оценивания результатов предсказания выживаемости; алгоритм имитационного моделирования данных о выживаемости.

Разработан алгоритм имитационного моделирования данных о выживаемости пациентов с онкологическими заболеваниями, основанный на группировании данных. Алгоритм имитационного моделирования с учетом кластеров не увеличивает ошибку предсказания выживаемости онкопациентов. Выполнен сравнительный анализ алгоритмов предсказания выживаемости на смоделированных и экспериментальных данных.

АГУЛЬНАЯ ХАРАКТЕРЫСТЫКА ПРАЦЫ

Дысертацыйная праца: 68 старонак, 18 ілюстрацый, 11 табліц, 55 крыніц, 3 пракладання.

ПРАДКАЗАННЕ ВЫЖЫВАЛЬНАСЦІ, БЛОЧНЫ ЛЕС, АЛГАРЫТМ РАСШЧАПЛЕННЯ ВУЗЛОЎ EXPONENTIAL LOG-LIKELIHOOD LOSS, БУСТЫНГ РЭГРЭСІІ КОКСА, МЕТАД ГАЛОЎНЫХ КАМПАНЕНТ, С-ІНДЭКС, АЦЭНКА БРЫЕРА, ЭКСПРЭСІЯ ГЕНАЎ, ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ, КЛАСТЕРНЫЙ АНАЛИЗ.

Мэта працы – распрацоўка імітацыйных мадэляў, метадаў аналізу і праграмных сродкаў для прадказання выжывальнасці пацыентаў пры анкалагічных захворваннях з выкарыстаннем алгарытмаў машыннага навучання.

Задачы: агляд літаратуры, распрацоўка алгарытмаў імітацыйнага мадэлявання і прадказання выжывальнасці, параўнальны аналіз алгарытмаў прадказання выжывальнасці на змадэляваных дадзеных, распрацоўка праграмага забеспячэння і канцэпцыі сістэмы абароны інфармацыі ў рэалізаваным праграмным забеспячэнні.

Аб'ектам даследавання з'яўляюцца генетычныя дадзеныя аб выжывальнасці анкапацыентаў.

Прадметам даследавання з'яўляюцца алгарытмы імітацыйнага мадэліравання і прадказання выжывальнасці анкапацыентаў.

Метады даследавання: алгарытмы блочнага лесу з лагарыфмічнай імавернаснай функцыяй страт і рэгрэсіі Коксу для прадказання выжывальнасці; метады галоўных кампанент; С-індэкс і ацэнка Брыера для ацэньвання вынікаў прадказання выжывальнасці; алгарытм імітацыйнага мадэлявання дадзеных выжывальнасці.

Распрацаваны алгарытм імітацыйнага мадэлявання дадзеных аб выжывальнасці пацыентаў з анкалагічнымі захворваннямі, заснаваны на групуванні дадзеных. Алгарытм імітацыйнага мадэлявання з улікам кластараў не павялічвае памылку прадказання выжывальнасці анкапацыентаў. Выканан параўнальны аналіз алгарытмаў прадказання выжывальнасці на змадэляваных і эксперыментальных дадзеных.

GENERAL DESCRIPTION OF WORK

Thesis work: 68 pages, 18 illustrations, 11 tables, 55 sources, 3 applications.

SURVIVAL PREDICTION, BLOCK FOREST, EXPONENTIAL LOG-LIKELIHOOD LOSS NODE SPLITTING ALGORITHM, COX REGRESSION BOOSTING, PRINCIPAL COMPONENT ANALYSIS, C-INDEX, BRIER SCORE, GENE EXPRESSION, SIMULATION MODELLING, CLUSTER ANALYSIS.

The purpose – to develop of simulation models, analysis methods and software tools for predicting cancer patient survival using machine learning algorithms.

Objectives: to review the literature, to develop algorithms for simulation modeling and prediction of survival and, to perform a comparative analysis of algorithms for predicting survival on simulated data, to develop software and concepts for an information security system in the implemented software.

The object of the study is genetic data on the survival of cancer patients.

The subject of the study is algorithms for simulation modeling and prediction of survival of cancer patients.

Research methods: block forest algorithms with exponential log-likelihood loss node splitting method and Cox regression for predicting survival; principal component method; C-index and Brier score to evaluate the results of predicting survival; simulation algorithm for survival data..

An algorithm has been developed for simulation modeling of data on the survival of patients with cancer, based on data grouping. The simulation modeling algorithm taking into account clusters does not increase the error in predicting the survival of cancer patients. A comparative analysis of survival prediction algorithms using simulated and experimental data was performed.