

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра теории вероятностей и математической статистики

Аннотация к магистерской диссертации

**РЕГРЕССИОННОЕ ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОДАЖ НА ОСНОВЕ
БОЛЬШИХ ДАННЫХ**

Кондратюк Никита Геннадьевич

Научный руководитель – профессор кафедры теории вероятностей и
математической статистики, доктор физико-математических наук,
Г.А.Медведев

Минск, 2017

РЕФЕРАТ

Магистерская диссертация, 115 страниц, 52 иллюстрации, 15 таблиц, 18 источников.

Ключевые слова: МАШИННОЕ ОБУЧЕНИЕ, СТАТИСТИЧЕСКОЕ ОБУЧЕНИЕ, КЛАССИФИКАЦИЯ, РЕГРЕССИЯ, МОДЕЛЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ, ЛИНЕЙНЫЙ ДИСКРИМИНАНТНЫЙ АНАЛИЗ, МОДЕЛЬ ЛИНЕЙНОЙ РЕГРЕССИИ, МОДЕЛЬ ДЕРЕВЬЕВ РЕГРЕССИИ, АНАЛИЗ ДАННЫХ, АНАЛИЗ УДОВЛЕТВОРЕННОСТИ КЛИЕНТОВ, КЛИЕНТООРИЕНТИРОВАННОСТЬ БАНКА, ПРОГНОЗИРОВАНИЕ ПРОДАЖ.

Объект исследований: модель взаимодействия банка Santander со своими клиентами, удовлетворенность клиентов услугами и сервисом банка, классификация клиентов, а также розничные продажи сети магазинов Rossmann.

Цель работы: анализ и прогнозирование удовлетворенности клиентов предприятия и прогнозирование суммарных продаж в торговой сети с использованием методов статистического обучения двух типов: классификации и регрессии.

Методы исследования: методы машинного обучения, статистические методы такие как логистическая регрессия, линейный дискриминантный анализ, линейная регрессия, деревья решений (включая ансамбли деревьев и случайные леса), основанные на использовании основ математической статистики и теории вероятностей; методы анализа исходных данных и оценки точности построенных моделей.

Результаты исследования: составлен вектор прогнозируемых результатов разбиения всего числа клиентов по заданным категориям при заданных начальных условиях на примере данных банка Santander; построен прогноз продаж на примере данных сети магазинов Rossmann; приведены алгоритмы построения моделей и способы их применения, оценка точности; проведен подробный анализ данных, их визуализация; разработана база данных для хранения и преобразования исходной информации.

Область применения: любое коммерческое предприятие занимающееся продажами и имеющее дело с клиентами.

РЭФЕРАТ

Магістарская дысертацыя, 115 старонак, 52 ілюстрацыі, 15 табліц, 18 крыніц.

Ключавыя словы: МАШЫННАЕ НАВУЧАННЕ, СТАТЫСТЫЧНАЕ НАВУЧАННЕ, КЛАСІФІКАЦЫЯ, РЭГРЭСІЯ, МАДЭЛЬ ЛАГІСТЫЧНАЙ РЭГРЭСІІ, ЛІНЕЙНЫ ДЫСКРЫМІНАНТНЫЙ АНАЛІЗ, МАДЭЛЬ ЛІНЕЙНАЙ РЭГРЭСІІ, МАДЭЛЬ ДРЭЎ РЭГРЭСІІ, АНАЛІЗ ДАДЗЕННЫХ, ПРАГНАЗАВАННЕ ПРОДАЖАЎ, АНАЛІЗ ЗАДАВОЛЕНАСЦІ КЛІЕНТАЎ.

Аб'ект даследаванняў: мадэль ўзаемадзеяння банка Santander са сваімі кліентамі, рознічныя продажу сеткі крам Rossmann.

Мэта працы: аналіз і прагназаванне задаволенасці кліентаў прадпрыемствы і прагназаванне сумарных продажаў у гандлёвай сеткі з выкарыстаннем метадаў статыстычнага навучання двух тыпаў: класіфікацыі і рэгрэсіі.

Метады даследавання: метады машыннага навучання, статыстычныя метады такія як лагістычная рэгрэсія, лінейны дыскрымінантны аналіз, лінейная рэгрэсія, дрэвы рашэнняў (уключаючы ансамблі дрэў і выпадковыя лесу), заснаваныя на выкарыстанні асноў матэматычнай статыстыкі і тэорыі верагоднасцяў; метады аналізу зыходных дадзеных і ацэнкі дакладнасці пабудаваных мадэляў.

Вынікі даследавання: складзены вектар прагназуемых вынікаў разбіцця усяго ліку кліентаў па зададзеных катэгорыях пры зададзеных пачатковых умовах на прыкладзе дадзеных банка Santander; пабудаваны прагноз продажаў на прыкладзе дадзеных сеткі крам Rossmann; прыведзены алгарытмы пабудовы мадэляў і спосабы іх прымянення, ацэнка дакладнасці; праведзены падрабязны аналіз дадзеных, іх візуалізацыя; распрацавана база дадзеных для захоўвання і пераўтварэнні зыходнай інфармацыі.

Вобласць ужывання: любое камерцыйнае прадпрыемства якое займаецца продажами і тое, якое мае справу з кліентамі.

ABSTRACT

Master's thesis, 115 pages, 52 illustrations, 15 tables, 18 sources.

Keywords: MACHINE LEARNING, STATISTICAL TRAINING, CLASSIFICATION, REGRESSION, LOGISTIC REGRESSION MODEL, LINEAR DISCRIMINANT ANALYSIS, LINEAR REGRESSION MODEL, REGRESSION TREE MODEL, DATA ANALYSIS, SALES FORECASTING, CUSTOMER SATISFACTION ANALYSIS.

The object of research: the model of Santander's interaction with its customers, retail sales of the Rossmann chain of stores.

Objective: analysis and prediction of customer satisfaction of the enterprise and forecasting of total sales in the trading network using the methods of statistical training of two types: classification and regression.

Methods: machine learning methods, statistical methods such as logistic regression, linear discriminant analysis, linear regression, decision trees (including tree ensembles and random forests), based on the basics of mathematical statistics and probability theory; methods for analyzing the initial data and estimating the accuracy of the models constructed.

Results: created a vector of predicted results of dividing the total number of clients into specified categories under given initial conditions (using the Santander bank data); sales forecast based on the example of the Rossmann network of stores; algorithms for constructing models and methods of their application, estimation of accuracy are given; detailed analysis of the data, their visualization; database for storing and transforming the initial information was developed.

Scope: any commercial enterprise engaged in sales and dealing with customers.