

УДК 004.9

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ АНАЛИЗА ПИЩЕВЫХ ПРИВЫЧЕК

Э. А. КАПЕЛЬКО

Научный руководитель Е. А. ЗАЙЧЕНКО

Белорусско-Российский университет

Могилев, Беларусь

Анализ влияния пищевых привычек людей на состояние здоровья выполнялся на наборе данных модуля «Питание и здоровье» (The Eating & Health Module Dataset), который содержит данные о потреблении различных категорий еды и общем состоянии здоровья. Данные находятся в свободном доступе, публикуются исследовательской организацией American Time Use Survey. Респондентам задавались вопросы о вторичном питании, т. е. приеме пищи во время выполнения другого вида деятельности; потреблении безалкогольных напитков; предпочтениях при покупке продуктов питания; методах приготовления пищи и безопасности пищевых продуктов; участии в продовольственной помощи; общем состоянии здоровья, росте и весе, физических упражнениях; доходе.

В качестве математической модели исследований был выбран множественный или многофакторный корреляционно-регрессионный анализ. Данный метод анализа решает три задачи: определяет форму связи результативного признака с факторными, выявляет тесноту этой связи и устанавливает влияние отдельных факторов. Зависимой переменной был показатель `eugenhth`, представляющий общее состояние здоровья, а все остальные показатели использовались в качестве независимых переменных.

В качестве средства разработки программного обеспечения для анализа пищевых привычек был использован язык программирования `python`, поскольку он имеет широкие возможности для проведения математического и статистического анализа данных. С помощью пакета `pandas` были выполнены очистка и первичная оценка данных по общим показателям. Пакеты `sklearn` и `scipy` применялись для реализации регрессионного анализа. С помощью перечисленных средств и графической библиотеки `matplotlib` выполнено построение матрицы корреляций между первыми по значимости индексами анализируемого набора данных.

Результаты анализа показали, что наиболее положительно коррелирующими коэффициентами с общим состоянием здоровья (показатель `eugenhth`) были вес респондента, количество мяса, птицы, рыбы и морепродуктов, съеденных респондентом за 24-часовой период, продолжительность физической активности, и, следовательно, они имеют положительную взаимосвязь с общим состоянием здоровья.

Основываясь на полученных результатах, можно говорить об успешной реализации данного метода, а также о его применимости в задачах анализа подобных задач в целом.