

РЕГРЕССИОННЫЙ АНАЛИЗ ДАННЫХ

Грицацьева О.А.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Голда О.А. – ст. преп.

Данная работа посвящена построению регрессионной модели зависимости продуктивности (объясняемая переменная) от различных факторов (сон, шаги, погода и др.). Целью работы является прогнозирование продуктивности человека под влиянием различных факторов.

В последние годы происходят структурные изменения в рабочем процессе: переход на свободный график, удаленная работа и человек должен сам выбирать наиболее подходящее время для работы.

Компания RescueTime, занимающаяся тайм менеджментом, в 2019 провела исследование и получила следующие результаты:

- сотрудники в среднем работают всего 2 часа и 48 минут продуктивного времени в день;
- 26% работы выполняется вне рабочего времени;
- 40,1% рабочего дня тратится на многозадачность с помощью средств коммуникации [1].

Проанализировав эти данные, можно сделать вывод о том, что рабочий баланс далек от идеального и приложение, которое поможет пользователям отслеживать продуктивность работы и влияющие на нее факторы, будет особенно актуально не только сейчас, но и в будущем. Знание того, как каждый человек работает и какие факторы влияют на его эффективность, являются ключевыми для тех, кто подходит к жизни более осознанно, рационально и хочет поддерживать баланс между работой и личной жизнью.

Еще одной тенденцией нашего времени является развитие и внедрение в повседневную жизнь людей различных взаимодействующих между собой информационных систем, таких как фитнес браслеты, умные часы, трекеры привычек, отслеживание финансовых расходов и транзакций.

Подводя итог всему вышесказанному, было принято следующее решение о создании регрессионной модели зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с линейной функцией зависимости [2].

Совместно с испытуемым, были выбраны следующие входные параметры:

- продуктивность (зависимая переменная);
- количество пройденных шагов;
- сон (фиктивная переменная);
- глубокий сон (фиктивная переменная);
- такси на работу (фиктивная переменная);
- рабочий либо выходной день (фиктивная переменная);
- употребление алкоголя (фиктивная переменная).

Для анализа данных была выбран язык R, который обладает широкими возможностями для манипулирования и визуализации данных, а также статистического анализа и машинного обучения [3].

Этапы построения модели:

1. Проверка, насколько сильно атрибуты коррелируют друг и другом (насколько один атрибут зависит от другого).
2. Построение модели и анализ вычисленных коэффициентов.

3. Прогнозирование с помощью построенной модели.

Исходя из полученных данных можно сделать вывод о том, что линейная регрессия помогает найти зависимость между факторами и сделать прогноз.

Развитие проекта:

получение данных со сторонних сервисов, а не напрямую от испытуемого (Google api);

добавление в модель объективных факторов (например, облачность, атмосферное давление, политическая ситуация в стране, коэффициент благополучия в стране);

создание интегрируемого CRUD приложения, которое будет обмениваться данными с другими информационными системами.

Список использованных источников:

1. Blog.rescuetime.com [Электронный ресурс] – информационный портал. – Электронные данные. – Режим доступа: <https://blog.rescuetime.com/work-life-balance-study-2019/>

2. Wikipedia [Электронный ресурс] – электронная энциклопедия. – Электронные данные. – Режим доступа: <https://ru.wikipedia.org/wiki/>

3. R-statistics.co [Электронный ресурс] – информационный портал. – Электронные данные. – Режим доступа: <https://r-statistics.co/Linear-Regression.html>