

ПРОВЕДЕНИЕ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА ПРОЕКТОВ КОМПАНИИ «ЯНДЕКС» НА ОСНОВЕ ДАННЫХ СЕРВИСА «ЯНДЕКС.РАДАР»

М.Д. Мацуганова

macuganovamaria5@gmail.com

О.Г Мандрик

mandrik_miit@rambler.ru

Т.П. Стасеня

tstasenia@mail.ru

УО Витебский Государственный Технологический Университет

Для решения различных экономических задач и проведения аналитических исследований выполняется экономико-математическое моделирование. Модель – это условный объект исследования, т.е. материальное или символично-информационное отображение реального объекта, процесса его функционирования в конкретной среде.

Целью данной работы является изучение методов расчета статистических параметров компанией «Яндекс». Исходные данные для моделирования компания получает со своих различных сервисов и на их основе разрабатывается многофакторная регрессионная модель.

Инструментарием исследования является пакет MS Excel и его приложения для моделирования.

Объектом исследования являются статистические данные компании «Яндекс».

Для научного исследования была выбрана крупнейшая технологическая компания – «Яндекс». Компания ведет бизнес как в России, так и еще в 21 стране. Бизнес-модель компании не подразумевает разработку программного обеспечения на заказ или прямую продажу его клиентам – компания делает упор именно на создание и развитие программного обеспечения различных пользовательских сервисов. «Яндекс» – один из ведущих игроков, организовавших свою экосистему. Также стоит упомянуть, что «Яндекс» не только самостоятельно разрабатывает новые сервисы с нуля, но и периодически покупает их для последующего встраивания в свою экосистему. Приобретаются как большие заметными сервисы, такие как «Кинопоиск» и «Авто-ру», так и более мелкие активы – например, «Партия еды» и часть активов «Везет».

Для разработки информационной модели и проведения анализа выбран сервис данных «Яндекс.Радар». «Яндекс.Радар» – инструмент для мониторинга популярности поисковых систем, браузеров и топ-проектов компании, который работает на обобщенных данных «Метрики». Для этого воспользуемся вкладкой «Проекты Яндекса», чтобы найти информацию о топ-проектах компании. Посещаемость интернет-проекта (и месячная, и дневная аудитория) определяется по кросс-девайсной склейке. Она позволяет «узнавать» пользователя во всех его браузерах и на всех устройствах. Например, если один и тот же пользователь открывал сайт в «Орега» и «Chrome» с десктопа, а потом в «Chrome» со смартфона, в статистике будет учтен один кросс-девайсный посетитель.

Склеюку анонимных идентификаторов браузеров и устройств одного и того же пользователя обеспечивает технология «Крипта». Система специально устроена так, что «Крипта» не получает личной информации о людях и тем более не передаёт её рекламодателям. Каждый пользователь для неё – это набор идентификаторов.

«Крипта» может с высокой вероятностью предположить, что пользователю с таким-то идентификатором может быть интересно такое-то предложение, – но кто этот человек, как его зовут и тому подобное, она не знает. «Крипта» умеет определять, что все эти идентификаторы принадлежат одному пользователю. В результате их «склейки» можно получить более полный портрет человека и распределять рекламные сообщения между его устройствами и приложениями.

Например, человеку, который со своего домашнего компьютера часто ищет что-то про средиземноморскую кухню, можно в «Навигаторе» показать рекламу греческого ресторана, когда он будет проезжать неподалёку.

Для «склейки» применяются два метода – вероятностный и факторный. Первый позволяет предположить, что два устройства принадлежат одному человеку, если поведение их пользователей очень похоже. Второй работает с более явными свидетельствами – например, фактором для «склейки» идентификаторов устройств может быть использование одного и того же логина на каком-нибудь сервисе. Правда, даже такой фактор нельзя трактовать однозначно, ведь одним логином часто пользуются несколько человек, члены семьи или коллеги. Поэтому на последнем этапе «склейки» производится обратный процесс, «расклейка» – устранение лишних и ненадёжных связей.

Для проведения исследования были приняты следующие обозначения: Y – это изменяющийся показатель под воздействием факторов X , а именно – месячная аудитория «Яндекса» в период с 01.01.2022г. по 01.06.2022г. В качестве независимых факторов X , под действием которых изменяется Y в период с 01.01.2022г. по 01.06.2022г., были определены: X_1 – «Портал», X_2 – «Дзен», X_3 – «Карты», X_4 – «Погода», X_5 – «Новости», X_6 – «КиноПоиск», X_7 – «Почта», X_8 – «Музыка», X_9 – «Авто.ру», X_{10} – «Расписание», X_{11} – «Путешествия», X_{12} – «Кью», X_{13} – «Услуги», X_{14} – «Спорт», X_{15} – «Услуги», X_{16} – «Телепрограмма», X_{17} – «Метро», X_{18} – «Недвижимость».

С помощью инструмента «Корреляция» пакета «Анализ Данных» была получена матрица коэффициентов парной корреляции. Анализ матрицы показал, что показатели X_5 – «Новости» и X_{13} – «Игры» в модели рассматриваться не будут, т.к. их коэффициенты корреляции меньше 0,4. У всех остальных показателей коэффициент корреляции больше 0,4, но учитывается мультиколлинеарность, т.е. наличие линейной зависимости между объясняющими переменными регрессионной модели. Следовательно, для дальнейшего анализа были выбраны только 6 факторов: X_1 , X_2 , X_3 , X_4 , X_{11} , X_{12} . Величины значимых факторов и значения показателя Y использовались в регрессионном расчете при помощи инструмента «Регрессия» пакета «Анализ Данных».

Анализ полученных данных показал, что модель является статистически значимой, а также значимыми являются все коэффициенты модели. В соответствии с полученным значением коэффициента множественной корреляции разработанной

модели (R) можно сделать вывод, что имеет место высокая степень тесноты связи между факторными и результативным признаками (100). Коэффициент детерминации (R^2) показывает, что вариация результативного признака на 100% обусловлена вариацией избранных факторных признаков.

В результате обработки и анализа многочисленных исходных данных получена следующая многофакторная регрессионная модель:

$$Y = 494673100 + 0,756965 \times X_1 - 0,44654 \times X_3 - 0,44654 \times X_4 - 0,14605 \times X_{11} + 1,103473 \times X_{12}.$$

Таким образом, можно сделать вывод о том, что исследование способов расчета статистики с сервисов, которые использует компания «Яндекс», позволяют разработать экономико-математическую модель многофакторной регрессии.

Моделирование помогает определить: тенденции развития результативного показателя; выделить основные влияющие факторы; оценить условия деятельности с учетом влияния определенных факторов; выполнить прогнозирование на основе параметров полученной модели.

"YANDEX" XİDMƏTİNİN MƏLUMATLARI ƏSASINDA "YANDEX" ŞİRKƏTİNİN LAYİHƏLƏRİNİN KORRELYASIYA VƏ REQRESSIYA TƏHLİLİNİN APARILMASI

M. D. Maçuganova

macuganovamaria5@gmail.com

O. G Mandrik

mandrik_miit@rambler.ru

T.P. Staseniya

tstasenia@mail.ru

Vitebsk Dövlət Texnologiya Universiteti

Bu işin məqsədi "Yandex" tərəfindən statistik parametrlərin hesablanması metodlarını öyrənməkdir. Şirkət modelləşdirmə üçün ilkin məlumatları müxtəlif xidmətlərindən alır və onların əsasında çoxfaktorlu reqressiya modeli hazırlanır. Tədqiqat alətləri MS Excel paketi və onun modelləşdirmə tətbiqləridir.

CONDUCTING CORRELATION AND REGRESSION ANALYSIS OF YANDEX PROJECTS BASED ON DATA FROM THE YANDEX SERVICE.RADAR

M.D. Matsuganova

macuganovamaria5@gmail.com

O.G. Mandrik

mandrik_miit@rambler.ru

T.P. Staseniya

tstasenia@mail.ru

Educational Institution Vitebsk State Technological University

The purpose of this work is to study the methods for calculating statistical parameters by "Yandex". The company receives the initial data for modeling from its various services, and a multifactor regression model is developed on their basis. Research tools are the MS Excel package and its modeling applications.

