

Наиболее объемными исследованиями являются изучение показателей экономической и социальной эффективности научной деятельности и ее влияния на развитие производства. Эти показатели рассчитываются как отношение полученного эффекта к сумме произведенных затрат.

Общим показателем эффективности научной деятельности является отношение фактически полученного годового эффекта от внедрения научных достижений в производство к фактическим затратам на научные изыскания и внедрение.

Предполагаемая цель для информации, собираемой в соответствии с этими инструкциями Руководства Фраскати, - помочь лицам, принимающим решения, особенно политикам. По мере того, как данные НИОКР становятся все более доступными и становятся все более заметными в публичных дискуссиях, эта информация также стала важным компонентом социального и политического диалога об использовании и воздействии этих ресурсов.

Широко известно, что данные НИОКР, отдельно или в сочетании с другими данными, могут обеспечить только часть основы для принятия решений. Однако до тех пор, пока правительства, лидеры бизнеса и общественность признают особенности НИОКР и наделяют их особым статусом, измерение человеческих и финансовых усилий, направленных на такие усилия, будет по-прежнему играть важную роль в необходимых статистических данных обществами по всему миру.

Библиографический список

1. Статистика научно-технического прогресса / Е. А. Кожевникова – Планово-экономический отдел. – №10. – 2011.

2. Guidelines for collecting and reporting data on research and experimental development URL: < <https://www.oecd-ilibrary.org/docserver/9789264239012-en.pdf?expires=1591380028&id=id&accname=guest&checksum=A450933D0F8FF0938180E37C01320671>> (дата доступа 25.05.2010)

3. Зинченко, А.П. Статистика / А.П. Зинченко. - М.: Изд-во РГАУ-МСХА, 2013. – 367 с.

УДК 519.688

РОЛЬ СТАТИСТИКИ В НАУКЕ О ДАННЫХ

Харитоновна Анна Евгеньевна, доцент кафедры статистики и эконометрики, ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, kharitonova.a.e@rgau-msha.ru

Аннотация: Определены основные понятия науки о данных. Рассмотрены наиболее популярные продукты, реализующие автоматизацию статистических расчетов и алгоритмов машинного обучения. Проведено сравнение методов классификации при обучении «с учителем» и «без учителя».

Ключевые слова: статистика, наука о данных, машинное обучение, data mining.

В настоящее время развитие любой науки происходит в условиях цифровизации. Из-за роста объемов информации многие программные продукты и методы анализа теряют свои преимущества. В результате развиваются новые направления. Так, в результате объединения информатики и статистики, появилось новое направление исследования - наука о данных.

Наука о данных (англ. data science; иногда даталогия — datalogy) — раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме.

В основе науки о данных лежит статистика, как основа методологии исследования. Однако эти методы реализуются с помощью языков программирования и все процессы автоматизированы. При этом методы статистики преобразовались в такие понятия как «машинное обучение», «глубокое обучение» и «интеллектуальный анализ данных». Однако многие ассоциируют эти методы со статистикой и не видят существенных различий, но эти понятия существенно отличаются друг о друга.

Статистика начала свое развитие еще с Древнего Рима. Сам термин был введен Готфридом Ахенваллем в 1746. Понятие «статистика» определяется как «самостоятельная общественная наука, изучающая массовые явления, происходящие в обществе, и закономерности их развития» [1]. Ежегодно статистика развивается, увеличивается комплекс методов обработки данных. Также в последние десятилетия стремительно развивается программирование, что существенно облегчает задачу обработки особенно больших массивов данных.

Первоначально компиляция статистики и программирования была выражена в развитии прикладных статистических программ. Сейчас широко используются при обработке данных такие пакеты прикладных статистических программ как Statistica, Eviews, STATA, SPSS и другие.

В дальнейшем развитие статистики и программирования привело к формированию такого понятия как искусственный интеллект и в его рамках машинного обучения. В 1959 году Артур Самуэль, ввел термин «машинное обучение» и определил его как процесс, в результате которого компьютеры способны показать поведение, которое в них не было явно запрограммировано. Сейчас машинное обучение считается ветвью искусственного интеллекта, основная идея которого заключается в том, чтобы компьютер не просто использовал заранее написанный алгоритм, а сам обучился решению поставленной задачи [1-3]. Цель машинного обучения – научить машину (точнее, программу) решать задачу, предъявив ей несколько примеров (с правильными и неправильными решениями).

Машинное обучение сочетает математическую статистику, методы оптимизации и классические математические дисциплины, но имеет и собственную специфику. Многие методы тесно связаны с извлечением информации и интеллектуальным анализом данных (Data Mining). Таким

образом, машинное обучение является развитием статистических методов с более широкими возможностями.

Основное отличие применения статистических методов от машинного обучения заключается в самом процессе анализа. Статистика работает с совокупностью, анализирует ее и на основании этого делает прогнозы. В машинном обучении исходная выборка делится на две части – обучающую и тестовую. Сам процесс построения моделей и поиска наилучших параметров проводится по обучающей выборке. Затем построенные закономерности и алгоритмы применяются к тестовой выборке и сравниваются результаты. На рис. 1 представлен процесс машинного обучения.



Рис. Процесс машинного обучения

В результате сразу проводится проверка качества построенной модели по данным, которые не участвовали в ее построении. В данном случае, на мой взгляд, качество модели проверяется надежнее, чем при использовании классических статистических методов.

Сочетание традиционных статистических методов, программирования, машинного обучения и предметной области и определяется как наука о данных (data science). Однако еще больше возможностей имеют языки программирования, которые позволяют прописать и автоматизировать не только методы, но и методику анализа.

В настоящее время разработаны готовые наборы методов машинного обучения. Они, в первую очередь, используются для общего понимания, какой метод лучше: Например, RapidMiner и Weka. Языки программирования, имеющие развитые библиотеки в области машинного обучения, такие как R, Python или Matlab, применяются в основном для экспериментов и выбора алгоритма. Также языки общего назначения могут реализовывать обработку данных, однако чаще всего уже когда алгоритм выбран и необходимо увеличить скорость обработки.

Широкое развитие получил язык статистической обработки данных R. Он появился в 1993 году и получил широкую популярность за счет свободного распространения и больших функциональных возможностей.

В целом следует отметить, что роль статистики в науке о данных остается существенной. Без фундаментальных основ и принципов статистики невозможно качественно и эффективно обработать данные. Однако развитие

статистики в науку о данных дает больше возможностей при работе с большими данными и делает эффективнее процесс обработки информации.

Библиографический список

1. Зинченко, А.П. Статистика / А.П. Зинченко. - М.: Изд-во РГАУ-МСХА, 2013. – 367 с.

2. Применение алгоритмов машинного обучения для решения задач количественного прогноза ФЕС по сейсмическим и скважинным данным / Егоров С.В., Приезжев И.И., Гладков Е.А. // Геофизика. – 2018. – № 3. – С. 33-38

3. Классификация, регрессия и другие алгоритмы Data Mining с использованием R / Шитиков В.К., Мастицкий С.Э. // Тольятти, 2017. – 351 с.

УДК 334.025

ДИСПЕРСИОННЫЙ АНАЛИЗ ВЛИЯНИЯ ВИДА ПОЛИЭТИЛЕНОВОГО ПАКЕТОВ НА ОБЪЕМ ПРОДАЖ

*Севастьянова Анастасия Сергеевна, студентка 2 курса магистратуры
seva.nastya@mail.ru*

Макунина Ирина Викторовна, к.э.н., доц. ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева

Аннотация: В данной статье приводится расчет влияния видов полиэтиленовых пакетов на объемы продаж путем дисперсионного анализа. Сделаны выводы о влиянии вида пакетов на объем продаж, высчитанный методикой дисперсионного анализа.

Ключевые слова: дисперсионный анализ, объем продаж, полиэтиленовый пакет, затраты.

Основная цель для проведения данного эксперимента – выявление зависимости влияния вида полиэтиленового пакета на объемы продаж. Чтобы подтвердить или отклонить нулевую гипотезу H_0 была произведена обработка данных, с использованием методики дисперсионного анализа

Дисперсионный анализ – статистический метод изучения взаимосвязи. Применяется для исследования влияния одной или нескольких качественных переменных на одну зависимую количественную переменную.[1.с33]

В основе дисперсионного анализа лежит предположение, что одни переменные могут рассматриваться как причины (независимые переменные), а другие как следствия (зависимые переменные или отклики).

Независимые переменные в дисперсионном анализе называют факторами, поскольку в ходе эксперимента исследователь может изменять их