случаев обучение для системы «БЭСТ-5» не требуется, только лишь необходимы начальные знания о системе работы. Данный этап выполняется параллельно остальным этапам.

6. Промышленная эксплуатация и гарантийное сопровождение. После завершения работ по внедрению начинается этап, так называемой, промышленной эксплуатации. В рамках гарантийного сопровождения оказываются услуги по поддержке проекта. Это оперативные консультации, доработка неучтенных алгоритмов. Сопровождение программы в процессе эксплуатации, как правило, выполняется пользователями программы и не вызывает трудностей. Компания своевременно реагирует на изменения законодательства и реализует их в программе. Обновление системы осуществляется разработчиком централизовано (выпускаемые обновления подойдут к любой конфигурации системы, адаптированной пользователем под потребности каждого предприятия с сохранением всех его индивидуальных настроек), а можно получить обновление с корпоративного сайта компании «БЭСТ» и установить самостоятельно или воспользоваться услугами сопровождающей организации.

Таким образом, автоматизация бухгалтерского учета, безусловно, является необходимостью для предприятий, независимо от масштабов их деятельности. Внедрение автоматизированной системы бухгалтерского учета поможет обеспечить эффективность учета и дальнейшее развитие предприятия [2].

Источники литературы

1. Слободняк, И.А. Актуальные проблемы автоматизации бухгалтерского учета / И.А. Слободняк, И.В. Пискунов.– Иркутск: БГУЭП, 2014. – С. 29-34.

2. Филлипов, А.В. Автоматизация бухгалтерского учета / А. В. Филипов – М.: Экономика, 2006.– 127 с.

3. Хохлов, А.Е. Автоматизированные системы бухгалтерского учета: конспект лекций / А.Е. Хохлов. – Пенза: Пензенский государственный университет, 2002. – 108 с.

4. Официальный сайт «БЭСТ» [Электронный ресурс]. – Режим доступа: http://www.bestnet.ru. – Дата доступа: 20.11.2016.

5. Официальный сайт фирмы «1С» [Электронный ресурс]. – Режим доступа: https://1c.ru. – Дата доступа: 20.11.2016.

6. Официальный сайт «Инфо-Предприятие» [Электронный ресурс]. – Режим доступа: http://www.infop.ru. – Дата доступа: 20.11.2016.

Пузанова Ольга Юрьевна
Белорусский государственный экономический университет

Data Science и способы проверки корректности выводов

В связи с ежедневным накоплением огромного количества данных появляются проблемы, связанные со способами их хранения и обработкой. Когда для анализа и построения выводов нам просто перестает хватать доступных вычислительных мощностей, появляются новые алгоритмы работы с данными, которые реализуются в механизмах data mining и data science. Однако в тот момент, когда часть ручных расчетов мы передаем машине, важно понимать, насколько корректными будут полученные результаты.

Целью данной работы является разграничение понятий big data, data mining и data science и определение наиболее эффективных способов проверки исследований в ходе data science.

Big Data - совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети [1].

Буквально Big Data переводится как «большие данные». Хотя сейчас это скорее маркетинговый термин, но он означает, что имеется так много данных, которые невозможно анализировать все сразу, потому что объем оперативной памяти, необходимый для хранения и обработки данных, превышает объем доступной памяти.

Отличительные характеристики Big Data – физический объем, постоянный прирост и разнообразие информации, в результате чего можно говорить об имеющейся «неструктурированной» информации. Главная задача здесь – навести порядок в данных для дальнейшего анализа.

Data Mining – набор алгоритмов и методов, предназначенных для обнаружения ранее неизвестных свойств данных [2]. Включает в себя различные методы (регрессии, ассоциации, классификации, кластеризации и т.д.), которые имеет сильную статистическую и математическую базу, которая не принимает во внимание знания предметной области и предварительной обработки данных.

Тот механизм, который после получения всего массива данных превратит его в полезную форму, а затем с учетом предметной области определит механизм дальнейшей обработки – есть Data Science. Самым ярким направлением здесь является Machine Learning – машинное обучение – обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения.

Взаимосвязь понятий big data, data mining и data science показана на рисунке 1.

Рисунок 1 – Связь понятий big data, data mining и data science.
Примечание- Источник: собственная разработка автора

Применение алгоритмов data science на практике влечет за собой ряд проблем: мы можем подобрать алгоритм, нерепрезентативный для решения конкретной задачи, потерять данные при их сборе в связи различиями в хранении в разных базах данных или в связи с ограничениями на права доступа. В итоге мы получаем спорные результаты и неточные выводы.

Так происходит, потому что на практике у нас есть несколько гипотез, или нет точной гипотезы, или для начала нужны данные для формирования гипотезы. И зачастую для формирования и подтверждения гипотезы вы вынуждены будете использовать одни и те же данные.

Ваш доступ к данным может быть ограничен только статистическими данными - таким образом вы сможете предположить результаты вашего эксперимента и посмотреть, какие вариации он мог бы иметь в прошлом.

Следующая проблема – ваши данные на самом деле не имеют отношения к гипотезе, которую вы хотите проверить. Возможно, это просто единственные данные, которые у вас есть. Например, вы хотите проанализировать конкретный магазин своей сети, но данные есть только по магазину с подобным форматом, расположением и прочими признаками. Вам посчастливилось иметь хотя бы это, и с этим вам придется работать. Но другая более важная проблема, с которой вы столкнетесь гораздо чаще – вы не будете знать правильные и точные измерения, чтобы проверять ваши гипотезы.

Собрать данные тоже проблематично. В разных источниках одна и та же по смыслу информация будет храниться в разных форматах, при некорректном объединении строки могут пропадать или становиться неуникальными и т.д. Во избежание подобного требуется знания ETL-процессов, построения моделей данных и др., а также производительные вычислительные мощности.

И после всего этого ваши выводы будут неопределенными. Иногда вы даже не сможете сказать, что было полезнее – проведение вашего теста или если бы вы ничего не делали.

Именно поэтому важно понимать, насколько полученная вами информация достоверна. Исходя из практики, наиболее подходящими способами проверки являются следующие [5]:

1. Построение сводных таблиц. Сводные таблицы очень полезны для обнаружения ошибок – правильность данных можно отследить уже на уровне единиц измерения, разрядности и самих значений в соответствии с показателями. При периодическом запросе одних и тех же таблиц и их сверке можно отследить динамику во времени. Аномальные отклонения будут очевидны.
2. Регрессионный анализ — метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной) [4]. При нахождении R^2 – меры определенности – мы видим качество полученной регрессии и можем оценить степень соответствия исходных данных и выбранной нами регрессионной модели.
3. Проблему со смешиванием данных можно решить на стадии проектирования эксперимента. Для проверки результатов используют A/B тестирование – способ рандомизации и деления данных на сегменты таким образом, чтобы их сравнение было сопоставимым. Это означает, что данные в сегментах не будут смешаны и будут уникальны для каждого сегмента.
4. Проблема выборки. В идеале для получения корректных результатов следует использовать в экспериментах случайные выборки. Однако вы не всегда можете вмешаться и как-то повлиять на этот процесс, поэтому ваша выборка может быть смещенной. Как следствие вам надо понимать, когда такая выборка будет репрезентативной, а когда нет, опираясь на погрешности.
5. Неоднозначные результаты. Надежность выводов определяет значимость всего эксперимента. Крайне сложно придумать универсальный алгоритм для определения размера выборки. Влияет ли больший объем на качество положительно или наоборот, меньшая выборка дала бы более точные результаты? Более того, на результат может повлиять неочевидная подмена измерений, поэтому важно понимать, в какой степени такие замены повлияют на конечный результат и можем ли мы этим пренебречь.

Перечисленные методики применялись компанией «А2 Консалтинг» для проведения анализа программы лояльности сети магазинов «Соседи».

Стартовал проект с сегментации клиентской базы по картам лояльности (более 300 тыс. карт). Для качественного сегментирования был использован RFM-анализ. Было формировано 125 групп по трем признакам (количество покупок в месяц, дата с последней покупки, накопленная сумма покупок за месяц), далее описаны 5 ведущих сегментов в базе.

Вторым этапом стал анализ акций и прямая работа по возвращению ушедших клиентов путем формирования целевых предложений. Были сформированы контрольные и

экспериментальные группы для различных предложений с помощью a/b тестирования. После оценки итогов анализа, были сформированы наиболее эффективные предложения.

Третий этап – это тонкий анализ поведения клиентов – анализировались смежные и похожие покупки с целью предложить клиенту попробовать новинки.

В ходе всего эксперимента все данные проверялись сверкой со сводными таблицами за предыдущие периоды, объективность выборки и полученных результатов оценивалась приглашенным экспертом.

Плановый результат подтвердился фактическим – после формирования точечных предложений в результате анализа групп, был получен возврат порядка 12% от числа клиентов, которые не покупали, но начали снова покупать после предложения. В результате проведенного анализа были получены новые знания о том, что рассылка клиентам со специальным предложением в их День Рождения увеличивает средний чек на 86 %, а купит по специальному предложению каждый второй участник программы лояльности. В результате рассылок с предложениями о покупках из зон смежных предпочтений прирост продаж составил 200% [6].

Таким образом, с распространением новых алгоритмов анализа данных и автоматизации этих процессов важно понимать, насколько корректные результаты мы получаем, тем самым определяя степень доверия к этой информации. Ошибки могут повлечь за собой потери для компании как в трудозатратах – на расчеты, так и в денежном выражении. Достоверная же информация позволяет принимать операционные и стратегические решения на качественно новом уровне.

Источники литературы*

1. Свободная энциклопедия Википедия [Электронный ресурс] . – Режим доступа: https://ru.wikipedia.org/wiki/Большие_данные. Дата доступа: 06.12.2016.
2. Свободная энциклопедия Википедия [Электронный ресурс] . – Режим доступа: https://ru.wikipedia.org/wiki/Data_mining. Дата доступа: 06.12.2016.
3. Профессиональный информационно-аналитический ресурс [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение. Дата доступа: 06.12.2016.
4. Профессиональный информационно-аналитический ресурс [Электронный ресурс] . – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Регрессия. Дата доступа: 17.11.2016.
5. Официальный сайт Coursera [Электронный ресурс]. – Режим доступа: https://www.coursera.org/learn/real-life-data-science. Дата доступа: 17.11.2016.
6. Официальный сайт А2 Консалтинг [Электронный ресурс]. – Режим доступа: http://a2c.by/press-tsentr/738-kupilka.html. Дата доступа: 10.12.2016.

**Статья опирается на свободную энциклопедию Википедия и Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных, т.к. по выбранной теме актуальные материалы можно найти только в источниках на оригинальных языках или в переводе энтузиастов на подобных ресурсах.*

Русакова Марина Михайловна
Белорусский государственный экономический университет

Аудит информационной безопасности предприятия

Данная тема актуальна в связи с тем, что сегодня информационные системы играют ключевую роль в обеспечении эффективности работы коммерческих и государственных предприятий. Повсеместное использование информационных систем для хранения, обработки и передачи информации делает актуальными проблемы их защиты, особенно учитывая глобальную тенденцию к росту числа информационных атак, приводящих к значительным финансовым и материальным потерям. Для эффективной защиты от атак