

ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА ДАННЫХ И ПРОГНОЗИРОВАНИЯ

А.О. Карташов, О.В. Оксюта, А.Ю. Павлов

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. Работа по созданию эффективных моделей прогнозирования и аналитики связана с обработкой большого объема информации, называемого Big Data. Для эффективного взаимодействия с ними необходимо четко определить допустимые математические модели, корректно способные обработать такой объем данных и выбрать наиболее подходящую. Выбор правильной модели позволит в дальнейшем модифицировать программу без дополнительных затруднений.

Ключевые слова: математические модели, прогнозирование, большие данные.

CAUSES OF DATA LOSS, TREATMENT AND PREVENTION METHODS

A.O. Kartashov, O.V. Oksyuta, A.Y. Pavlov

Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The work on creating effective forecasting and analytics models involves processing a large amount of information, called Big Data. For effective interaction with them, it is necessary to clearly define acceptable mathematical models that are correctly able to process such a volume of data and select the most appropriate one. Choosing the right model will allow you to further modify the program without additional difficulties.

Keywords: mathematical models, forecasting, big data.

В условиях стремительно меняющегося рынка и цифровой экономики умение адаптироваться становится неотъемлемой частью деятельности

компаний во всех сферах жизни человека. Чем раньше изменения в экономической и социальной сферах будут замечены, тем больше времени на реагирование и принятие взвешенных решений будет у ответственных лиц. Соответственно, повышается вероятность успешной интеграции в бизнес-процессы новых систем и алгоритмов. Таким образом, использование методов прогнозирования является ключевым параметром для тех компаний, кто хочет успешно и наиболее безболезненно сталкиваться с различными трудностями и кризисами.

Использование больших данных для прогнозирования

Основой любой модели прогнозирования, будь то собственно созданные или наиболее популярные, является набор данных, которые эти модели будут анализировать, или на которых будут обучаться в дальнейшем. Если изначальный массив с информацией неполный, некорректный или подвержен внешнему вмешательству, получение рабочей и эффективной модели становится крайне сложной, а в некоторых случаях невозможной задачей. Объединяющим термином для таких массивов является Big Data или Большие данные. Главным критерием Big Data является использование в своих характеристиках всеобъемлющей информации об объекте исследования, это позволяет разрабатывать сложные и адаптивные функции для прогнозирования в информационных системах. К примеру, погодные условия могут стать ключевым фактором для прогнозирования спроса на услуги моек самообслуживания в крупных городах. А информациях о праздничных днях помогает определить всплески посещаемости в торговых центрах, соответственно, повышение продаж для магазинов, бутиков и лавок, расположенных в них.

Влияние прогнозов на бизнес деятельность

Влияние прогнозов в процессе ведения бизнеса сложно переоценить. Они позволяют формировать маркетинговые стратегии, исходя из потребностей компании и заинтересованности потребителей, заранее определять бюджеты на логистику, оплату труда и кредитные операции. Также понимание спроса может помочь при определении нагрузки на производства. Это особенно актуально для технически сложных товаров, выпуск которых может занимать значительное количество времени и ресурсов.

Основными методами для использования прогнозов по отраслям являются:

1. Для финансовой сферы – Регрессия и временные ряды. Они позволяют работать с данными о доходах и расходах. Исходя из этой информации, делаются выводы о прибыли или убытки компании за рассматриваемый промежуток времени.

2. Работа с клиентами – Машинное обучение, которое позволяет выработать алгоритм для системы вопрос-ответ в чат-ботах, позволяя экономить время и ресурсы работников.

3. Прогнозирование кризисов и рисков – Стохастические модели. Они помогают находить потенциально опасные тенденции на рынке, связанные с паданием курса валют, акций или спроса. Соответственно, предоставляя возможность для корректировки курса организации.

4. Анализ эффективности маркетинга – Логистическая регрессия. Позволяет бизнесу работать над оптимизацией маркетинга, поиска оптимальных методов для продвижения и рекламы своего товара или услуги.

5. Реализация запасов – ARIMA (Autoregressive Integrated Moving Average). Позволяет анализировать нестандартные временные ряды и приводить их к стационарному виду. Такой метод помогает предсказывать изменения в количестве товаров.

Математические модели

Для понимания работы методов прогнозирования необходимо рассмотреть наиболее популярные математические модели, которые используются в них.

1. Одной из самых простых, но при этом базовых моделей является линейная регрессия. Она имеет вид: $Y = a + bX + \varepsilon$,

где Y – это показатель, который необходимо спрогнозировать;

X – фактор;

a – свободный коэффициент;

b – коэффициент регрессии, демонстрирует взаимосвязь показателей Y и X ;

ε – случайная ошибка, которая показывает разницу между реальным значением и рассчитанным прогнозом.

Базово данная модель имеет вид $y = ax + b$. Однако использование случайной ошибки позволяет отойти от вида идеальной прямой и приблизить модель к реальным данным.

2. Модель ARIMA применяется в случаях, если в данных используется сезонность. Данная модель имеет вид: $Y_t = \varphi Y_{t-1} + \varepsilon_t$,

где φ – коэффициент авторегрессии;

ε_t – случайная ошибка, которая показывает разницу между реальным значением и рассчитанным прогнозом.

При использовании данной модели необходимо учитывать некоторые ограничения. Она работает, только если временной ряд стационарный, а также когда данные очищены и подобраны корректны. К тому же, модель не учитывает влияние внешних факторов и плохо адаптируется к резким изменениям в выборке.

3. Логистическая регрессия может применяться для предсказания поведения клиентов, сотрудников, заказчиков и так далее. Иными словами, логистическая регрессия используется для составления прогноза о наступлении события. Данная модель имеет вид:

$$P = \frac{1}{1 + e^{-(a+bX)}}$$

где P – вероятность того, что событие произойдет;

a – свободный коэффициент;

bX – коэффициенты при факторах и значение факторов;

e – экспонента.

В результате значение P варьируется от 0 до 1. Если $P > 0.5$, то событие может наступить, в обратном случае – нет.

4. Экспоненциальное сглаживание можно использовать в тех случаях, когда существует необходимость запоминать прошлые данные. То есть, в данном методе больший вес придается последним (новым) наблюдениям, относительно старых. Для того чтобы добавить или убавить значимость данных, используется коэффициент сглаживания. Например, в случае когда $\alpha = 1$ модель ориентируется на новые данные, а в обратном – на старые.

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$$

где F_{t+1} – результат прогноза для следующего периода;

Y_t – значение в текущем периоде;

F_t – прогноз текущего периода;

α – коэффициент сглаживания.

Исходя из всего этого, можно сделать вывод, что правильно подобранная математическая модель играет ключевую роль для составления прогнозов. Специфика той или иной сферы обязательно должна учитываться, так как изначальный массив данных должен претерпевать изменения в зависимости от того, какие задачи ставятся перед моделями. Например, линейная регрессия не подойдет в случаях, когда необходимо выяснить вероятность наступления события. А логистическая регрессия не может использоваться для учета сезонности.

В наши дни корректная работы с прогнозами становится ключом к высокой конкурентоспособности бизнеса, даже в тех случаях, когда компания не обладает большими финансовыми и человеческими ресурсами.

Список литературы

1. Поляков, Я. А. Использование различных подходов резервного копирования баз данных для предотвращения потери данных / Я. А. Поляков, Е. Д. Алехина // Научные технологии в приборостроении и развитии инновационной деятельности в вузе : Материалы Всероссийской научно-технической конференции. В 2-х томах, Калуга, 14–16 ноября 2023 года. – Москва: Издательство МГТУ им. Н.Э. Баумана, 2024. – С. 60-62.
2. Фридман, О. Д. Предотвращение риска разрушения и потери ценных данных в банковской сфере / О. Д. Фридман, Д. А. Трунина, Г. О. Чегуров // Научный электронный журнал Меридиан. – 2019. – № 16(34). – С. 42-44.
3. Танкаян, А. И. Информационная безопасность персонального компьютера и современные виды угроз потери данных / А. И. Танкаян, Д. Н. Савинская // Информационное общество: современное состояние и перспективы развития : Сборник материалов XI международного студенческого форума, Краснодар, 23–27 июля 2018 года. – Краснодар: Кубанский государственный аграрный университет имени И.Т. Трубилина, 2018. – С. 114-116.
4. Поляков, А. А. Скрытые риски массовых обновлений данных: как избежать потери и утечки информации / А. А. Поляков // Международный журнал информационных технологий и энергоэффективности. – 2024. – Т. 9, № 12(50). – С. 12-15.

5. Алешин, А. Ю. Эффективные способы борьбы с потерей корпоративных данных / А. Ю. Алешин // Защита информации: IT, правовые и экономические аспекты : Сборник научных трудов Межвузовской студенческой научно-практической конференции, Москва, 16–17 марта 2023 года. – Москва МИРЭА - Российский технологический университет, 2023. – С. 84-87.

References

1. Polyakov, Ya. A. The use of various database backup approaches to prevent data loss / Ya. A. Polyakov, E. D. Alyokhina // High-tech technologies in instrumentation and mechanical engineering and the development of innovation in higher education institutions : Proceedings of the All-Russian Scientific and Technical Conference. In 2 volumes, Kaluga, November 14-16, 2023. Moscow: Publishing House of Bauman Moscow State Technical University, 2024, pp. 60-62.

2. Friedman, O. D. Preventing the risk of destruction and loss of valuable data in the banking sector / O. D. Friedman, D. A. Trunina, G. O. Chegurov // Scientific electronic journal Meridian. – 2019. – № 16(34). – Pp. 42-44.

3. Tankayan, A. I. Information security of a personal computer and modern types of data loss threats / A. I. Tankayan, D. N. Savinskaya // Information Society: current state and development prospects : Collection of materials of the XI International Student Forum, Krasnodar, July 23-27, 2018. Krasnodar: I.T. Trublin Kuban State Agrarian University, 2018, pp. 114-116.

4. Polyakov, A. A. Hidden risks of massive data updates: how to avoid information loss and leakage / A. A. Polyakov // International Journal of Information Technology and Energy Efficiency. – 2024. – Vol. 9, No. 12(50). – pp. 12-15.

5. Alyoshin, A. Y. Effective ways to combat corporate data loss / A. Y. Alyoshin // Information protection: IT, legal and economic aspects : Proceedings of the Interuniversity Student Scientific and Practical Conference, Moscow, March 16-17 In 2023. – Moscow: MIREA