**Выступление**

**Слайд 2**

Прогнозирование в экономике и финансах представляет собой систематический процесс оценки и предсказания будущих событий, тенденций и переменных, влияющих на экономическую деятельность.

Оно играет ключевую роль в принятии экономических решений, предоставляя необходимую информацию для определения оптимальных стратегий и тактик управления. На основе точных и надежных прогнозов управленцы и аналитики могут выстраивать эффективные бизнес-модели, минимизировать риски, адаптироваться к изменяющимся условиям рынка и принимать обоснованные решения, направленные на достижение поставленных целей компании.

Успешный прогноз в экономике и финансах характеризуется несколькими важными аспектами.

Во-первых, он должен быть точным, отражая реальные тенденции и явления в экономической сфере.

Во-вторых, успешный прогноз обладает высокой степенью достоверности, что позволяет минимизировать возможные ошибки и риски при принятии решений.

В-третьих, прогноз должен быть своевременным, предоставляя информацию заранее, чтобы предприятия и организации могли адаптироваться к предстоящим изменениям.

**Слайд 3-4**

Анализ временных рядов – это метод, который помогает понять закономерности изменения данных в разные моменты времени. Этот процесс включает в себя выделение трендов (долгосрочных изменений), сезонных колебаний (циклических изменений, повторяющихся в определенные периоды) и случайных флуктуаций. Графики временных рядов позволяют визуализировать эти компоненты, что полезно для предварительного анализа данных перед применением более сложных методов.

ARIMA (Autoregressive Integrated Moving Average):

Авторегрессия (AR): Этот компонент описывает зависимость текущего значения временного ряда от предыдущих значений. Например, если цены на акции росли в прошлом, есть вероятность, что они продолжат расти.

Интегрирование (I): Этот шаг включает в себя вычитание текущего значения временного ряда из предыдущего для стабилизации его колебаний.

Скользящее среднее (MA): Этот компонент учитывает зависимость текущего значения от предыдущих ошибок прогноза.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity):

Условная гетероскедастичность: Этот термин относится к изменчивости (волатильности), которая не является постоянной. Модель GARCH оценивает и учитывает эту изменчивость в условиях предыдущих ошибок прогноза.

SARIMA (Seasonal ARIMA): Эта модель добавляет компонент сезонности к ARIMA, что особенно полезно при работе с данными, имеющими циклические колебания в течение года.

Применение методов временных рядов в финансовых прогнозах позволяет анализировать и прогнозировать цены на активы, волатильность рынка и другие важные финансовые параметры. Например, модель GARCH может помочь в оценке риска инвестиций, учитывая переменную волатильность рынка.

**Слайд 5**

Эконометрика – это область экономической науки, которая использует математические и статистические методы для измерения и анализа экономических явлений. В прогнозировании эконометрические методы применяются для построения моделей, которые могут описывать взаимосвязи между различными переменными и использоваться для предсказания будущих значений.

Множественная регрессия: Этот метод позволяет учитывать влияние нескольких факторов на исследуемую переменную. Например, при прогнозировании продаж товара можно учесть не только цену, но и рекламные затраты, сезонные колебания и другие факторы. Модель множественной регрессии выражается уравнением, где зависимая переменная предсказывается на основе значений нескольких независимых переменных.

Панельные данные: Этот метод позволяет анализировать данные по нескольким объектам (например, компаниям) в разные периоды времени. В прогнозировании это позволяет учесть как временные, так и индивидуальные особенности объектов, улучшая качество предсказаний. Например, при изучении доходов различных компаний, можно учесть как изменения внутри компаний со временем, так и различия между компаниями.

**Слайд 6**

Машинное обучение – это подраздел искусственного интеллекта, который обучает компьютерные системы самостоятельно обрабатывать данные и выявлять закономерности.

Линейная регрессия — это метод, который моделирует линейную зависимость между зависимой переменной и одной или несколькими независимыми переменными. В задачах прогнозирования, где необходимо предсказать числовое значение, линейная регрессия может быть мощным инструментом.

Случайные леса — это метод, объединяющий несколько деревьев решений для более точного и устойчивого прогноза. Каждое дерево строится независимо, а затем их прогнозы усредняются. Этот метод подходит для прогнозирования в условиях сложных и нелинейных зависимостей.

## Сравнение точности методов временных рядов, эконометрики и машинного обучения

**Методы временных рядов:**

Плюсы:

* Эффективны для моделирования временных зависимостей и сезонных колебаний.
* Хорошо интерпретируемы и обладают структурированным подходом.

Минусы:

* Ограничены в учете сложных нелинейных взаимосвязей.
* Требуют стационарности данных.

**Эконометрические методы:**

Плюсы:

* Обеспечивают структурный анализ влияния различных факторов.
* Интерпретируемы и применимы для анализа множественных переменных.

Минусы:

* Требуют предположения о структуре данных.
* Могут оказаться неэффективными при большом объеме данных и сложных нелинейных взаимосвязях.

**Методы машинного обучения:**

Плюсы:

* Способны моделировать сложные, нелинейные зависимости.
* Могут обрабатывать большие объемы данных и адаптироваться к изменениям.

Минусы:

* Могут быть сложными для интерпретации.
* Требуют больших объемов данных для обучения.

**Сравнительный анализ:**

Сравнение точности показывает, что методы машинного обучения, такие как нейронные сети и случайные леса, часто превосходят другие методы в сложных прогнозах, но при этом могут быть менее интерпретируемыми. Методы временных рядов и эконометрики подходят для задач с явными временными и структурными зависимостями, но могут проигрывать в сложных нелинейных сценариях.

**Слайд 8**

**Кейс магнита**

## Применение статистических моделей для прогнозирования продаж, кейс magnit tech

Под статистическими моделями подразумеваем набор моделей, использующих в основе скользящие средние с рядом преобразований разной степени сложности. От простых скользящих средних и экспоненциального сглаживания до ARIMA, SARIMA, ARIMAX и более сложных комбинированных вариантов.

Первые попытки использовать статистические модели в бизнесе уходят корнями в историю. Они предшествовали бурному развитию IT и росту вычислительных мощностей в силу относительной простоты и прозрачности применения, соразмерно уровню технической оснащённости тех времён.

Если внимательно посмотреть на процессы любого бизнес-подразделения, где возникает хотя бы минимальная потребность что-либо прогнозировать, высока вероятность встретить интуитивные попытки использования скользящих средних: «закажем, как в прошлые разы», «сделай по аналогии».

компания «Магнит»— не исключение. Более 20 лет назад был создан инструмент «Автозаказ» на базе скользящих средних.

Автозаказ – инструмент прогнозирования, при помощи которого осуществляется пополнение товаром торговых объектов компании.

В те годы это был настоящий прорыв: администрации магазинов больше не нужно было вести ручной учёт и размещать заказы товаров в конце рабочего дня. «Автозаказ» всё делал автоматически. Это сэкономило и высвободило тысячи человеко-часов за счёт эффекта масштаба, а также минимизировало человеческий фактор ручной ошибки прогноза.

Безусловно, требования к качеству метрик и сложность решаемых задач за прошедшие годы возросли на порядок. Однако статистические модели всё ещё в стеке и могут быть актуальны для решения отдельных задач: они хорошо себя зарекомендовали и прошли проверку временем.

Для прогнозирования промо использование статистических моделей в базовом виде не подходит из-за наличия набора сильно влияющих на целевую переменную факторов. Для решения этой проблемы компания использует мультипликативную модель прогнозирования (ММП) вида:

y = x \* k1 \* k2 \* … \* kn,

Где x – очищенная от влияния факторов базовая продажа, над которой строится статистическая модель.

k1…kn – учитываемые факторы (мультипликаторы).

Таким образом, декомпозиция ряда на составляющие позволяет приблизить ряд наблюдений X к стационарному виду, после чего над ними строится статистическая модель.

Один из ключевых моментов – правильная оценка факторов и учёт их влияния. Рассмотрим для примера самые важные из них:

* Коэффициент эластичности спроса – собственная разработка, аппроксимация эластичности спроса на товар в зависимости от изменения цены. Рассчитывается отдельным алгоритмом.
* Коэффициенты сезонности – сезонная составляющая изменчивости спроса. Для расчёта используем адаптированный под наши цели алгоритм на базе библиотеки Prophet.

Также важный пункт в реализации подобных алгоритмов – правильный выбор детализации расчёта, исходя из целей прогнозирования и горизонта сбора статистики для расчёта базовой компоненты. В их версии ММП они используют короткий тренд – около трёх месяцев: динамика продаж отдельных товаров в розничной торговле меняется достаточно быстро, а представленная модель не способна учитывать весь спектр факторов, влияющих на изменение целевой переменной.

Помимо прочего, ценная особенность алгоритма – независимая иерархическая структура расчёта каждой из компонент. Это позволяет использовать модель для прогнозирования товаров и магазинов, не имеющих собственных наблюдений.

Попробуем разобраться, почему это так важно. В таблице приведён пример наиболее часто встречающихся комбинаций наличия статистики по компонентам в упрощённом виде для одного условного магазина.

Если бы основная часть наблюдений соответствовала примеру 6–7, им не пришлось бы разделять модель на отдельные модули расчёта компонент. В реальности значительная доля случаев скорее «окрашена в красный цвет». Поэтому ММП в рамках расчёта последовательно проверяет, есть ли статистика каждой из компонент: начинает с нижнего уровня детализации, двигается вверх при отсутствии таковой и останавливается на том уровне иерархии расчёта, где выполняется критерий по достаточному количеству наблюдений.

Целью создания модели было решение двух задач:

* Быстро создать относительно простой и легко интерпретируемый для бизнеса алгоритм.
* Прогнозировать все поступившие на расчёт записи, включая те, у которых полностью отсутствует история продаж.

Помимо разработки и совершенствования основных модулей (мультипликаторов и компонент), мы реализовали ряд дополнительных веток расчёта для учёта нестандартных явлений разного рода, прогнозирование которых было критичным с точки зрения бизнеса. Это необходимо для случаев, где исходная версия модели не способна выдавать прогноз в силу архитектурных ограничений либо при появлении сильного внешнего фактора, вызывающего существенную просадку ключевых метрик.

Сильные стороны мультипликативной модели

* Простота применения, можно быстро адаптировать подход для разных задач.
* Проще интерпретировать результаты и искать причины ошибок прогноза благодаря декомпозиции ряда на отдельные, понятные бизнесу мультипликаторы.
* Применение модели на коротком тренде позволяет относительно быстро реагировать на внешние факторы, которые могут оказывать существенное влияние на спрос. Пример – ковид или резкое изменение структуры ассортимента из-за ухода старых и появления новых поставщиков.
* За счёт иерархической структуры расчёта компонент модель способна прогнозировать почти 100% от всех потенциальных входящих данных, вне зависимости от наличия статистики: например, для новых товаров и не встречавшихся ранее скидок и цен.

Слабые стороны мультипликативной модели

* Обратная сторона простоты – отсутствие «гибкости». В отличие от более сложных моделей нельзя просто разметить события в истории и пробросить признак в будущее.

Любое нестандартное явление в статистике продаж или реализация новой механики промо помимо явной разметки требует создания отдельной надстройки (ветки расчёта) и разработку логики для каждой уникальной ситуации (заданную последовательность действий: if…else…). Это приводит либо к бесконечному усложнению логики, утяжелению и разрастанию кода, либо к отказу от использования подхода в части случаев.

* Сложности при прогнозировании товарных категорий, присутствующих в ассортименте ограниченный период времени в течение года. Например, новогодних украшений, кремов для загара, сезонных овощей и фруктов. Подобные товары могут не иметь доверительной статистики продаж в короткой истории, а периоды их появления в ассортименте и выхода на «пиковые значения» меняются от года к году.
* Разработка новых признаков возможна только в виде мультипликаторов, что существенно ограничивает потенциал развития модели.

Статистические модели - безусловно являются фундаментом прогнозирования в силу высокой скорости расчёта, простоты реализации и интерпретации результатов, а также могут показывать неплохой результат для задач не высокой сложности.

Следуя принципу «Бритвы Оккама», при наличии ограниченного ресурса и времени на разработку, статистические модели могут оказаться отличным стартовым (а зачастую – достаточным) решением (вспоминаем про «самокаты» и «космические корабли»). Но не стоит забывать, что качество подобных моделей напрямую зависит от глубины понимания специфики бизнеса, качества и структуры имеющихся в наличии данных и способностей команды видеть связь между сухими цифрами в источниках и реальными событиями.