ПРОГНОЗИРАНЕ НА ТЪРСЕНЕТО НА БЪРЗООБОРОТНИ СТОКИ: СРАВНИТЕЛЕН АНАЛИЗ НА КЛАСИЧЕСКИ И МОДЕРНИ МАШИННИ АЛГОРИТМИ

Иван Росенов Митков
Стопанска академия „Димитър А. Ценов“ – Свищов
Катедра „Статистика и приложна математика“
e-mail: ivan.rmitkov@gmail.com

Резюме: Тази публикация предоставя задълбочен анализ на различни методи за прогнозиране на търсенето в ритейл сектора, включващи както класически статистически модели, така и съвременни алгоритми за машинно обучение. Разгледани са класическите модели ARIMA и тройно експоненциално изглаждане, които са доказали своята ефективност при анализ на времеви редове, както и модерни алгоритми като Prophet, NeuralProphet, Random Forest, XGBoost и LSTM, които демонстрират висока гъвкавост и способност за обработка на сложни зависимости. Проведен е сравнителен анализ на предимствата и ограниченията на тези подходи в контекста на прогнозирането на търсенето на бързооборотни стоки (FMCG). На тази основа са избрани четири модела – ARIMA, тройно експоненциално изглаждане, Prophet и NeuralProphet – за последващ сравнителен анализ, с цел оценка на тяхната ефективност чрез метрики като MAE, MSE, RMSE и MAPE. Настоящото изследване предоставя насоки за оптимален избор на модел спрямо характеристиките на данните и специфичните бизнес нужди, както и посочва възможности за бъдещи изследвания, включително използването на хибридни модели и включването на допълнителни външни фактори за подобряване на прогнозната точност.

Ключови думи: прогнозиране на търсенето, ритейл, времеви редове, машинно обучение, бързооборотни стоки.
JEL: C22, C53.

DEMAND FORECASTING FOR FAST-MOVING CONSUMER GOODS: A COMPARATIVE ANALYSIS OF CLASSICAL AND MODERN MACHINE LEARNING APPROACHES

Ivan Rosenov Mitkov
„D. A. Tsenov“ Academy of Economics, Svishtov
Department of Statistics and Applied Mathematics
e-mail: ivan.rmitkov@gmail.com

Abstract: This paper provides an in-depth analysis of various demand forecasting methods in the retail sector, encompassing both classical statistical models and modern machine learning algorithms. The classical models ARIMA and triple exponential smoothing, known for their effectiveness in time series analysis, are reviewed alongside modern algorithms such as Prophet, NeuralProphet, Random Forest, XGBoost, and LSTM, which demonstrate high flexibility and capability to handle complex dependencies. A comparative analysis of the strengths and limitations of these approaches is conducted in the context of forecasting demand for fast-moving consumer goods (FMCG). Based on this analysis, four models - ARIMA, triple

exponential smoothing, Prophet, and NeuralProphet - are selected for subsequent comparative analysis to evaluate their performance using metrics such as MAE, MSE, RMSE, and MAPE. This study offers guidelines for selecting the optimal model based on data characteristics and specific business needs, while also identifying opportunities for future research, including the use of hybrid models and the inclusion of additional external factors to enhance forecasting accuracy.

Key words: demand forecasting, retailing, time series, machine learning, fast-moving consumer goods.

JEL: C22, C53.

Въведение

Прогнозирането на очакваното търсене е критичен процес за всеки търговец на дребно, особено в областта на бързооборотните стоки (FMCG), където се реализират продажби на продукти, характеризиращи се с кратък жизнен цикъл, висока честота на потребление и сравнително ниска продажна цена. Прецизните прогнози са от съществено значение за поддържането на конкурентоспособността на търговците, тъй като влияят директно върху ефективността на изградената верига на доставки, управлението на наличността, финансовото състояние и клиентската удовлетвореност. В последните години ритейл индустрията претърпява значителни промени поради растящото равнище на дигитализация и нарастващото потребление на данни, което прави процеса на прогнозиране все по-предизвикателен (Чобанов, 2020); (Марков, 2018). В условията на непрекъснати пазарни промени компаниите от всякакъв мащаб трябва бързо и точно да предвиждат потребителското търсене, за да могат ефективно да реагират на променящите се нужди и предпочитания на клиентите (Kotler, 2016).

Точността на прогнозите е не само предимство, но и необходимост за устойчиво присъствие на пазара. Във време, когато потребителите стават все по-взискателни към предложения продукт и очакват по-високо ниво на обслужване от страна на търговците, всяка неточност в прогнозите може да доведе до значителни финансови загуби. Липсата на прецизност в прогнозите за очаквано търсене могат да се окажат в основата на моменти с липси на наличност или свръхпредлагане на продукти, което намалява клиентското доверие и конкурентоспособността на компанията. Поддържането на оптимални нива на наличност е критично за минимализиране на разходите за съхранение и избягване на загуби поради излишни количества. Прогнозите за търсенето улесняват планирането на производството и дистрибуцията, позволяват на компаниите да използват ресурсите си по-ефективно и дават възможност за управление на маркетинговите кампании по прецизен начин (Колев, Н. Г., 2019).

Въпреки предимствата, които носят точните прогнози, ритейл бизнесът се сблъсква с редица предизвикателства, свързани с промените в

потребителските предпочитания, сезонните колебания, краткия жизнен цикъл на продуктите и икономическите условия. Едно от най-значимите предизвикателства е отчитането на сезонността в търсенето. Например празниците като Коледа или Великден се характеризират с увеличено търсене на специфични продукти, като декорации, храни и напитки, което изисква прогнозите да бъдат адаптирани към тези цикли (Fildes, 2008). Промоциите и ценовите намаления също водят до увеличено търсене за кратък период, което традиционните методи за прогнозиране често не могат адекватно да отразят (Ailawadi, 2009).

Променливият характер на търсенето в ритейл сектора често прави използването на класически статистически модели недостатъчно ефективно. Бързооборотните стоки са известни с това, че бързо губят популярност сред клиентите, което води и до по-ниска ефективност на алгоритмите (Montgomery, 2008). За разлика от традиционните методи новите алгоритми за машинно обучение могат да се адаптират по-бързо към промените и да предоставят по-точни прогнози в условията на динамични пазарни тенденции (Goodfellow, 2016).

Прогнозирането на търсенето в търговията на дребно е преминало през значителна еволюция през последните десетилетия. В миналото компаниите разчитаха основно на статистически модели като линейна регресия и експоненциално изглаждане, които предлагат стабилни и лесни за интерпретиране прогнози. Линейната регресия е предпочитана заради своята простота и способност да моделира зависимостта между фактори като цена, сезонност и обем на продажбите. Въпреки това този модел често се оказва недостатъчен при анализиране на сложни и нелинейни зависимости в данните (Hyndman, 2018). Експоненциалното изглаждане, от своя страна, предоставя по-голяма гъвкавост за краткосрочно прогнозиране и може да отчита скорошни промени в търсенето, като придава по-голяма тежест на последните стойности (Winters, 1960).

Класическите статистически методи на времевите редове, и по-специално моделите ARIMA, разработени от Box (1976) също играят важна роля в развитието на прогнозите. ARIMA моделите обединяват авторегресивни и компоненти на плъзгащи се средни, като биват широко използвани за прогнозиране на времеви редове с тренд и сезонни компоненти. Въпреки това ARIMA изисква стационарност на данните и е сравнително трудно приложим при големи и динамични редове с голяма волатилност.

С развитието на технологиите за събиране и анализ на данни, както и с нарастващото значение на големите данни, алгоритмите за машинно обучение като Random Forest, XGBoost и невронните мрежи придобиват популярност в прогнозирането. Те не само предлагат по-висока точност, но и възможността да обработват големи обеми данни и да откриват сложни нелинейни зависимости (Breiman, 2001); (Chen T. &., 2016); (Hochreiter, 1997). Проучванията показват, че методите за машинно обучение предоставят

значително по-точни прогнози в сравнение с класическите модели в случаите, когато пазарните условия се променят бързо. В хода на развитие и еволюция на методите от машинното обучение се появяват и Prophet (Taylor, 2018) и NeuralProphet (Triebe, 2021), като предлагат по-висока точност и лесна интерпретация. Разработен от Facebook, Prophet е особено подходящ за анализ на данни със сезонност и празнични ефекти, позволявайки бързо внедряване в бизнес контекста. NeuralProphet, който е базиран на Prophet, добавя елементи на дълбоко обучение (deep learning), като подобрява точността на прогнозите при сложни и нелинейни времеви зависимости.

С разширените възможности за събиране на данни от различни източници търговците на дребно разполагат и с повече инструменти за анализ на потребителското поведение. Големите данни промениха коренно прогнозите за търсенето, тъй като данните от социални медии, продажбени транзакции и дори информация за времето вече могат да бъдат включени в прогнозните модели (Chen, H. C., 2012). Тези нови възможности изискват аналитични методи, които са способни да обработват сложни и големи набори от данни и да извличат полезна информация за бизнес стратегиите.

Обект на настоящото изследване са различните методи за прогнозиране на търсенето, а предметът е свързан с извършването на сравнителен анализ на класическите статистически модели със съвременните алгоритми за машинно обучение. Основната цел е да се оценят предимствата и недостатъците на различните подходи и да се изясни приложението им в променливи търговски сценарии. В допълнение се разглеждат и хибридни модели, които съчетават предимствата на класическите статистически методи и машинното обучение и предлагат ново ниво на прецизност и адаптивност в прогнозите за търсенето (Khashei, 2011). Освен представянето на техническите аспекти на различните методи тук се дават и някои практически насоки за компаниите, търсещи по-ефективни начини за прогнозиране на търсенето в една конкурентна и динамична среда. В изследването се поддържа тезата, че, познавайки предимствата и ограниченията на традиционните (класическите) и съвременните алгоритми за прогнозиране, могат да се дадат насоки за оптимален избор на модел в зависимост от характеристиките на данните и специфичните бизнес нужди.

Изложението на разработката е следното. В следващата част се прави обзор на популярна и актуална литература, която застъпва поставения проблем. След това са представени теоретичните постановки при основните статистически алгоритми и тези от машинното обучение, като се изтъкват отличителните им характеристики, както и някои от силните и слабите им страни. Изследването продължава с четвъртата част, в която се прави задълбочена дискусия за приложимостта на представените модели в контекста на търговията с бързооборотни стоки. В последната част се прави заключение, базирайки се на представени факти, като се дават насоки за потенциални бъдещи изследвания.

1. Литературен преглед

Въведените в предходната част ключови предизвикателства и възможности пред прогнозиране на търсенето в търговията с бързооборотни стоки подчертават важността от детайлен анализ на съществуващите подходи. Използваните методи за прогнозиране са претърпели значително развитие през последните десетилетия, като всеки нов етап в технологиите за анализ на данни е допринасял за повишаване на точността и надеждността на прогнозите. Основната цел на тази секция е да представи еволюцията на прогнозните модели, както и тяхното значение за индустрията, като разгледани ще бъдат както класически статистически подходи, така и съвременни алгоритми за машинно обучение и хибридни модели. Чрез включването на изследвания от международни и български автори ще бъдат обсъдени предимствата и ограниченията на тези методи, като се постави акцент върху приложимостта им в контекста на бързооборотните стоки.

Литературният обзор започва с преглед на класическите модели за времеви редове, които продължават да играят важна роля в прогнозирането на търсенето. След това фокусът ще падне върху съвременните подходи, базирани на машинно обучение и големи данни, които предоставят по-голяма гъвкавост и точност в анализа на сложни и динамични пазарни условия. Във заключителната част на обзора биват обсъдени иновациите в хибридните модели и значението на човешкия фактор в прогнозите. Секцията приключва с обобщение на основните изводи от прегледа на литературата.

1.1. Класически методи и модели за времеви редове

Прогнозите за търсенето в ритейл сектора се базират на анализ на времеви редове, където класическите методи като линейната регресия и експоненциалното изглаждане продължават да играят основна роля, особено при краткосрочните прогнози. Линейната регресия представлява един от найранните модели за прогнозиране и намира приложение в условия на сравнително постоянна сезонност и тенденция. Според (Hyndman, 2018), линейната регресия е полезна при прогнозиране на данни с ясна сезонна структура, но не може ефективно да отрази сложни нелинейни зависимости. В българския контекст Марков и съавтори подчертават, че линейната регресия е приложима главно за анализ на структурирани и лесни за интерпретиране времеви редове, но не предлага достатъчна точност за прогнозиране на по-динамични данни, като тези на бързооборотните стоки (Марков, 2018).

Друг класически метод, използван в прогнозирането на времеви редове, е експоненциалното изглаждане. Winters (1960) въвежда този метод, който включва придаване на по-голяма тежест на по-близките стойности в данните и е подходящ за краткосрочни прогнози в условия на стабилни сезонни цикли (Winters, 1960). Според Winters експоненциалното изглаждане

позволява по-бързо реагиране на незначителни промени в тенденцията и сезонността, но при наличие на по-значителни промени методът не успява да осигури висока точност. В българската литература Колев и Иванов (2019) също така отбелязват, че експоненциалното изглаждане е често използван метод за прогнозиране на сезонни продукти в ритейл сектора, въпреки че липсата на адаптивност го прави неефективен за продукти с висока изменчивост в търсенето (Колев, Н. И., 2019).

Методите на времевите редове като ARIMA (Autoregressive Integrated Moving Average) предлагат по-структуриран подход към прогнозирането. Box и Jenkins (1976) разработват този модел, който съчетава авторегресивни компоненти и компоненти на плъзгащи се средни и позволява моделиране на по-сложни данни, включително сезонност и тренд. Въпреки ефективността си при данни с ясно изразена стационарност, ARIMA изисква значителна предварителна подготовка и настройка на параметрите, което ограничава приложимостта му за динамично променящи се данни (Hyndman, 2018). В България Марков и съавторите му (2018) посочват, че ARIMA моделите са подходящи за компании с достатъчно ресурси и експертиза за обработка на по-сложни исторически данни, но остава ограничен при прогнозиране на краткосрочни колебания в търсенето (Марков, 2018).

1.2. Съвременни подходи: машинно обучение и големи данни

Концепцията за големите данни променя начина, по който се събират, анализират и използват данните за прогнозиране на търсенето. Според Chen възможността за обработка на огромни обеми от данни в реално време осигурява на ритейл компаниите по-добро разбиране на клиентските предпочитания и динамиката на пазара (Chen, H. C., 2012). Чрез интегриране на данни от различни източници като социални медии, мобилни приложения и времеви редове, компаниите получават достъп до по-пълна информация за поведението на потребителите. В България анализът на големи данни се използва все по-често в ритейл сектора, за да подпомогне адаптирането на наличностите и планирането на стратегически кампании спрямо промените в потребителското търсене (Колев, Н. Г., 2019).

Съвременните технологии значително разширяват възможностите за прогнозиране на търсенето, като включват алгоритми за машинно обучение и инструменти за анализ на големи данни, които позволяват обработка на по-сложни и динамични данни. Една от основните цели на тези методи е да предложат по-гъвкаво и адаптивно прогнозиране, като вземат под внимание фактори като сезонност, променливи икономически условия и потребителски навици.

Сред най-използваните алгоритми са Random Forest и XGBoost. Random Forest е ансамблов метод, който комбинира множество решаващи дървета за идентифициране на сложни зависимости между променливите.

Възможността за паралелна обработка на множество фактори го прави подходящ за анализ на търсенето в условия на динамични и често нелинейни зависимости. В своето изследване Чобанов подчертава, че Random Forest намира приложение в ритейл сектора като един от предпочитаните алгоритми за прогнозиране на търсенето (Чобанов, 2020). XGBoost допълнително усъвършенства техниките за градиентно усилване (gradient boosting) чрез оптимизиране на процеса и въвеждане на регуляризация, което намалява риска от свръхобучение (overfitting).

Дълбокото обучение, и по-специално моделите Long Short-Term Memory (LSTM), също се използват широко в ритейл прогнозите. LSTM мрежите предлагат възможности за улавяне на дългосрочни зависимости във времевите редове, като това ги прави подходящи за обработка на последователни и сезонни данни. В изследванията с български данни LSTM моделите се доказват като ефективни при прогнозиране на сезонни продукти, като плодове и зеленчуци, където точните прогнози могат да помогнат за оптимизация на наличностите (Тодорова, 2020).

Prophet е модел, който интегрира свойствата на класическите времеви редове с адаптивност към сезонни и седмични колебания, често срещани в ритейл данните. Ефективността на алгоритъма проличава особено при прогнозиране на времеви редове с пропуски и при данни с ясна сезонна или дневна структура. Авторите на модела отбелязват, че Prophet е лесен за настройка и позволява бързо обучение, което го прави подходящ за приложения, изискващи чести прогнози (Taylor, 2018). Възможността за автоматична обработка на сезонност и празнични ефекти осигурява гъвкавост, която го прави популярен в търговията на дребно.

Като надграждане на вече представените два подхода на невронни мрежи и Prophet, разработен бива NeuralProphet (Triebe, 2021). Моделът разширява възможностите на Prophet, като добавя адаптивност чрез използване на рекурентни невронни мрежи за обработка на дългосрочни зависимости и взаимодействие между променливите. NeuralProphet има потенциала да открива сложни зависимости в данните и да осигурява по-точни прогнози при динамични пазарни условия. Според (Triebe, 2021) моделът е ефективен при прогнозиране на времеви редове с много фактори като променливи в потребителското поведение и маркетингови кампании.

1.3. Хибридни модели за прогнозиране на търсенето

Хибридните модели за прогнозиране комбинират класическите статистически подходи с новите технологии за машинно обучение, съчетавайки предимствата на всеки от тях. Те са особено ефективни в контексти, при които данните имат както линейни, така и нелинейни зависимости. Klashei и съавтори (2011) предлагат хибриден модел, който интегрира ARIMA с изкуствени невронни мрежи, като постигат значително по-добра точност в

сравнение с отделните методи (Khashei, 2011). Те отбелязват, че комбинацията между авторегресивните компоненти на ARIMA и нелинейната обработка на данни чрез невронните мрежи позволява на модела да се справя с висока изменчивост на данните и сложни зависимости.

В българския контекст Марков и други автори (2018) изследват приложимостта на хибридните модели за прогнозиране на сезонните продукти. Те откриват, че комбинацията на ARIMA и Random Forest води до по-добри прогнози за продуктите с ясно изразени сезонни цикли и динамични промени в търсенето (Марков, 2018). Авторите посочват, че използването на хибридни модели може да подобри прецизността на прогнозите в ритейл сектора, тъй като тези модели съчетават адаптивността на машинното обучение и прецизността на класическите статистически подходи.

Други проучвания подчертават предимствата на хибридните модели, включващи LSTM и ARIMA. LSTM моделите могат да уловят дългосрочни зависимости и сезонни колебания, докато ARIMA позволява на модела да отчита краткосрочните колебания (Tashman, 2020). Тодорова и Иванов (2020) използват хибриден модел с LSTM и традиционни статистически методи за прогнозиране на търсенето на храни в различни сезони в България. Резултатите показват, че този подход е особено ефективен за продукти с изразени сезонни колебания, като например пресни плодове и зеленчуци, които изискват адекватно управление на наличностите.

1.4. Човешкият фактор и ролята на експертното мнение

Въпреки напредъка на технологиите за прогнозиране, човешкият фактор продължава да играе ключова роля в процеса на вземане на решения, свързани с управлението на търсенето, като дори най-добрите прогностични модели се нуждаят от човешка интерпретация и експертно мнение, за да бъдат адекватно адаптирани към извънредни ситуации и промени на пазара. Това е особено валидно за търговията с бързообортни стоки, където внезапни колебания в търсенето могат да бъдат предизвикани от фактори като промоции, промени в икономическите условия или социални тенденции.

В българския ритейл сектор, където автоматизацията на прогнозите е все още в процес на развитие, човешкият фактор играе важна роля за адаптирането на прогнозните модели към местните особености. В своето изследване Костова и съавтори (2020) отбелязват, че прогнозите в повечето български компании са базирани не само на моделите, но и на опита и познанията на мениджърите (Костова, 2020). Експертното мнение често се използва като коректив за прогнози, базирани на исторически данни, особено в случай на непредвидими събития. Това дава възможност на компаниите да се адаптират по-гъвкаво и ефективно към локалните нужди и предпочитания.

Една от значимите ползи от човешката намеса е способността за подобра реакция при извънредни събития и глобални кризи, които моделите

често не успяват да предвидят. По време на пандемията от COVID-19, например, ритейл компаниите по целия свят бяха изправени пред безпрецедентни промени в потребителското поведение, които изискваха значителна адаптация на прогнозните модели. Изследванията на Чобанов (2020) показват, че в България компаниите, които комбинират модели с експертно мнение, успяват да адаптират своите операции по-бързо към новите условия, като предвиждат нарастването на търсенето на определени категории стоки и същевременно ограничават наличностите на по-слабо търсените продукти.

Литературният преглед показва, че прогнозите за търсенето в ритейл сектора изискват комбинация от различни подходи и технологии, за да се постигне необходимата точност и адаптивност. Класическите методи като линейната регресия, експоненциалното изглаждане и ARIMA предоставят стабилна основа за анализ на времевите редове, но са ограничени в способността си да обработват динамични и комплексни данни. Същевременно алгоритмите за машинно обучение като Random Forest, XGBoost и LSTM позволяват обработка на големи и разнообразни данни и осигуряват по-добра прогностична точност в условия на променящо се потребителско поведение.

Хибридните модели, които съчетават статистически и машинно-обучителни подходи, осигуряват допълнителни предимства, тъй като интегрират линейни и нелинейни компоненти и предлагат по-голяма гъвкавост в прогнозите. Въпреки че тези методи изискват значителни изчислителни ресурси, те предоставят по-добра точност при сложни пазарни зависимости и сезонни колебания. Човешкият фактор също остава важен аспект, като осигурява критичен контрол при интерпретацията и адаптацията на прогнозите спрямо локалните и глобалните условия.

Българският ритейл сектор показва нарастваща осведоменост и прилагане на новите технологии за прогнозиране на търсенето, като се стреми към адаптация на международните практики към локалния пазар. Прегледът на литературните източници подчертава необходимостта от интеграция на различни подходи, като същевременно показва значимостта на локалния контекст и експертното мнение при вземането на прогностични решения. Успешната комбинация от традиционни и модерни модели, както и участието на експертите в процеса, се очаква да увеличи точността на прогнозите и да подпомогне ефективното управление на ритейл веригите в България и в глобален мащаб.

2. Основни характеристики на алгоритмите за прогнозиране

С напредъка на анализите на времеви редове и с нарастващия обем на данни прогнозите за търсенето в ритейл сектора преминаха през съществени трансформации. Литературният обзор представи историческото развитие и

текущите тенденции в областта, както и основните принципи на анализ на търсенето. Динамичният характер на потребителските предпочитания и разнообразието от външни влияния, включително сезонността, икономическите условия и промоционалните кампании, изискват точни и адаптивни подходи за прогнозиране. В този контекст се обособяват две основни категории модели: класическите статистически методи и съвременните алгоритми за машинно обучение. Тези модели се използват за откриване на закономерности в историческите данни и формиране на прогнози за бъдещото търсене.

Класическите модели като линейната регресия, експоненциалното изглаждане и ARIMA се отличават с простота и ефективност при по-малки набори от данни и стационарни времеви редове. Те позволяват на анализаторите да изграждат прогнози с базови математически зависимости, което ги прави особено полезни при по-малко комплексни случаи, където прогнозните времеви редове не включват значителни нелинейни зависимости или сложна сезонност. Тези модели са дълбоко изследвани и имат доказана ефективност при различни сценарии за прогнозиране на търсенето, но техните ограничения стават очевидни в условия на бързо променящи се пазарни условия или сложни взаимодействия между променливите.

От друга страна, с развитието на технологиите и нарастващата наличност на големи данни се появяват и модели, използващи машинно обучение и дълбоки невронни мрежи, като Random Forest, XGBoost; LSTM; Prophet и Neuralprophet. Тези алгоритми имат способността да откриват скрити зависимости и нелинейни взаимодействия в данните, което ги прави изключително подходящи за комплексни и динамични сценарии на прогнозиране. Алгоритмите за машинно обучение, и по-специално дълбоките невронни мрежи, предлагат голяма гъвкавост и могат да се адаптират към променящите се условия на пазара, като обединяват големи обеми от исторически и контекстуални данни. Въпреки че те изискват значителни изчислителни ресурси и внимателна настройка на параметрите, техните високи нива на точност ги правят предпочитан избор за сложни анализи на търсенето.

Тази теоретична част има за цел да разгледа по-подробно всеки от основните алгоритми, използвани за прогнозиране на търсенето, като се обърне внимание на техните теоретични основи, предимства и ограничения. Сравнението между класическите и съвременните методи ще предложи ясна представа за приложимостта на всеки модел и ще освети възможностите за комбиниране на различни техники, за да се постигнат по-точни прогнози. В края на теоретичната част ще бъдат разгледани и методите за оценка на точността на прогнозите, които са критични за избор на подходящ модел. Това ще създаде основа за дискусията, където ще бъдат обобщени основните предимства и ограничения на тези модели в контекста на прогнозирането в сектора за бързооборотни стоки.

2.1. Класически статистически модели за прогнозиране

2.1.1. Линейна регресия

Линейната регресия е един от основните статистически методи за прогнозиране, който изследва линейната зависимост между зависима променлива и една или повече независими променливи. По отношение на прогнозиране на търсенето линейната регресия е особено подходяща за анализ на тенденции и връзки между търсенето на даден продукт и фактори като цена, сезонност и обем на продажбите. Основното уравнение на линейната регресия е:

$$(2.1)\ y = \beta_0 + \beta_1 x + \epsilon,$$

където y е зависимата променлива, x е независимата променлива, β_0 е константата (или сечението), β_1 е наклонът и ϵ представлява случайния шум. Когато в модела се включат няколко независими променливи, моделът става многократна линейна регресия, описана със следната формула:

$$(2.2)\ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Линейната регресия се използва често заради своята простота и възможността за лесна интерпретация на параметрите. Например наклонът β_1 показва промяната в стойността на зависимата променлива y при промяна в x с една единица, което е полезно при прогнозиране на ефекта на определени фактори върху търсенето. Един от основните методи за оценка на параметрите β_0 и β_1 е методът на най-малките квадрати, който минимализира сумата на квадратите на отклоненията между наблюдаваните и прогнозните стойности на y.

Силните страни на линейната регресия са свързани с нейната лесна интерпретация и минимални изчислителни изисквания, което я прави подходяща за по-прости анализи и начални етапи на изследвания (Hyndman, 2018). Линейната регресия е особено полезна при малки набори от данни, които не съдържат сложни зависимости или нелинейни ефекти.

Въпреки това линейната регресия има няколко ограничения. Тя предполага, че зависимостта между променливите е линейна, което не винаги отразява реалността. Прогнозите на линейната регресия могат да бъдат неточни при наличие на нелинейни зависимости, които се срещат често в реалния свят, особено при прогнозиране на търсенето на бързооборотни стоки, където потребителските предпочитания често се променят нелинейно. Освен това линейната регресия е податлива на свръхобучение при включване на прекалено много независими променливи, което може да намали нейната способност за генерализация към нови данни.

След като разгледахме теоретичната основа и основните характеристики на линейната регресия, в следващата част ще продължим с експоненциалното изглаждане като още един класически модел за прогнозиране на търсенето.

2.1.2. Експоненциално изглаждане

Експоненциалното изглаждане е често използван метод за прогнозиране, който разчита на придаване на по-голяма тежест на по-новите данни. Това го прави подходящ за краткосрочни прогнози и за времеви редове с ясни тренд или сезонни зависимости. Основната идея на експоненциалното изглаждане е, че последните стойности от данните са по-информативни за бъдещото поведение на времевия ред и следва да се вземат под внимание с по-голяма тежест в сравнение с по-старите стойности.
Основното уравнение за експоненциално изглаждане е:

$$(2.3)\ S_t = \alpha Y_t + (1 - \alpha) S_{t-1},$$

където S_t е изгладената стойност за текущия период t, Y_t е реалната стойност за периода t, а α е изглаждащият коефициент, като $0 < \alpha < 1$. Стойността на α контролира тежестта, която се придава на по-новите наблюдения спрямо по-старите. По-високи стойности на α правят модела по-реактивен към последните изменения в данните, докато по-ниски стойности водят до по-гладко изглаждане, придавайки по-голяма тежест на историческите данни.

Освен основното експоненциално изглаждане, съществуват и вариации като двойното и тройното експоненциално изглаждане. Двойното експоненциално изглаждане се използва, когато времевият ред съдържа ясно изразен тренд, но няма сезонни вариации. Уравнението за двойно експоненциално изглаждане включва допълнителен компонент за тренда:

$$(2.4)\ S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$(2.5)\ b_t = \beta(S_t - S_{t-1}) + (1 - \beta) b_{t-1},$$

където b_t е компонентът на тренда, а β е параметър, който определя тежестта за трендовия компонент. Тази версия е подходяща за прогнозиране на данни, които показват тренд, но нямат сезонност.

Тройното експоненциално изглаждане, познато също като метода на Холд и Уинтърс (Winters, 1960), се използва, когато времевият ред включва както тренд, така и сезонни колебания. Уравнението му е по-сложно и включва компоненти за сезонност, които допълнително повишават точността на прогнозите при периодични данни.

Силните страни на експоненциалното изглаждане включват лесната му настройка и способността му бързо да адаптира прогнозите си към нови

изменения в данните. Това го прави много ефективен при краткосрочни прогнози, особено когато времевият ред не съдържа изразена сезонност. Основното предимство е способността му да реагира на промени в тренда чрез регулиране на параметъра α, като така предоставя адаптивност и точност в кратки периоди (Hyndman, 2018).

Експоненциалното изглаждане обаче има някои недостатъци. Един от тях е, че методът не е подходящ за дългосрочни прогнози, тъй като се фокусира предимно върху последните стойности и не обхваща дългосрочните зависимости в данните. Освен това, за по-сложни данни с изразена сезонност или динамични трендове, експоненциалното изглаждане може да даде неточни прогнози, ако не се използват неговите двойни и тройни варианти (Winters, 1960).

След като разгледахме теоретичните основи и характеристиките на експоненциалното изглаждане, ще продължим с анализа на ARIMA моделите – един от най-широко използваните методи за прогнозиране на времеви редове.

2.1.3. ARIMA модели

ARIMA (Autoregressive Integrated Moving Average) моделите са един от най-популярните статистически подходи за прогнозиране на времеви редове, особено когато данните включват стационарни тенденции или сезонни компоненти. Разработени от Box и Jenkins (1976), ARIMA моделите комбинират три основни компонента: авторегресия (AR), диференциране (I) за постигане на стационарност и плъзгаща се средна (MA). Моделът ARIMA с ред (p, d, q) е описан с уравнението:

$$(2.6)\ y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

където y_t е стойността на времевия ред в момент t, ϕ_i са авторегресивните параметри, θ_j са параметрите на плъзгащата се средна, p е редът на авторегресията, d степента на диференциране за постигане на стационарност, а q е редът на плъзгащата се средна. Моделът ARIMA предоставя гъвкавост и точност при анализ на времеви редове с ясни зависимости, като често се използва за краткосрочни и средносрочни прогнози.

Силните страни на ARIMA включват способността му да улавя различни типове зависимости във времевите редове, както и доказаната му ефективност при стационарни данни (Box, 1976). Той е сравнително лесен за интерпретация и осигурява стабилни резултати при прогнози с ясно изразен тренд и сезонност.

Сред основните недостатъци на ARIMA са изискванията за предварителна обработка на данните и ограничената му способност да се справя с нестационарни и нелинейни зависимости. Настройката на параметрите p, d

и q също може да бъде предизвикателство и изисква специализирани знания за коректно приложение.

2.2.Съвременни подходи от машинното обучение

2.2.1. Random Forest

Random Forest е ансамблов метод за прогнозиране, който комбинира множество решаващи дървета, за да създаде стабилна и точна прогноза. Методът използва бутстрапинг и случайни подмножества от характеристики, за да създаде независими дървета и да намали свръхобучението. Основното уравнение за Random Forest е:

$$(2.7)\ \hat{y} = \frac{1}{N}\sum_{i=1}^{N} \hat{y_i},$$

където N е броят на дърветата в ансамбъла, а $\hat{y}$ е прогнозата на всяко дърво. Random Forest е подходящ за многомерни данни и има висока точност при прогнозиране на дълги времеви редове. Силните страни на Random Forest включват неговата устойчивост на шум, способността му да обработва сложни взаимодействия и многомерни характеристики, както и висока точност.

Основните ограничения на Random Forest са свързани с високите изисквания за изчислителни ресурси и необходимостта от настройка на параметрите при големи обеми от данни. Методът може да бъде чувствителен към прекомерна комплексност, което го прави подходящ за по-сложни, но не и за прекалено динамични данни (Breiman, 2001).

2.2.2. XGBoost

XGBoost (Extreme Gradient Boosting) е метод за градиентно усилване, който включва регуляризация за предотвратяване на свръхобучение и осигурява висока точност чрез добавяне на множество усилващи стъпки. Основната формула за XGBoost е:

$$(2.8)\ \hat{y} = \sum_{k=1}^{K} f_k(x),$$

където $f_k(x)$ представляват отделните дървета в ансамбъла, а K е броят на дърветата. XGBoost се отличава с висока точност и бързо обучение при големи набори от данни. Силните страни на XGBoost включват способността му да се справя с висока точност и ефективност при анализ на сложни времеви редове, както и гъвкавостта му при различни параметризации.

Недостатъците на XGBoost са свързани със сложността на настройка на параметрите и високите изисквания за изчислителни ресурси. Параметри като learning rate, depth и round count трябва да се настройват внимателно,

което може да забави процеса на обучение, но резултатите често оправдават усилията (Chen, T. &., 2016).

2.2.3. LSTM (Long Short-Term Memory) мрежи

LSTM (Long Short-Term Memory) мрежите са вид рекурентни невронни мрежи, които са създадени да преодолеят ограниченията на традиционните рекурентни мрежи при обработка на дълги последователности от данни. Тези модели включват специални клетки, които задържат информация за дълги времеви периоди чрез механизми за задържане и забравяне, което им позволява да улавят дългосрочни зависимости във времевите редове. Това ги прави изключително подходящи за данни със сложна сезонност и дългосрочни трендове, като например прогнозирането на търсенето в ритейл сектора, където могат да възникнат нелинейни и циклични модели на потребление.

LSTM клетките работят с няколко основни компонента: входен, забравящ и изходен порт, които контролират потока на информацията през клетката. Формулите за тези компоненти са следните:

(2.9) $f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big)$

(2.10) $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$

(2.11) $C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$

(2.12) $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$

LSTM моделите са особено полезни за прогнози с дългосрочни зависимости, като например при данни със силни сезонни и трендови компоненти. Техните основни предимства са възможността за улавяне на сложни времеви взаимодействия и способността да се адаптират към променящи се модели на потребление. Основните ограничения на LSTM са свързани с високите изисквания за изчислителна мощност, което може да затрудни тяхното обучение при много големи набори от данни. Също така настройването на хиперпараметрите на LSTM е трудоемък процес, който изисква задълбочени познания и често експериментиране (Hochreiter, 1997).

2.2.4. Prophet

Prophet е модел, характеризиращ се с лесна настройка за прогнозиране на времеви редове с ясно изразена сезонност и празнични цикли. Prophet използва адитивна структура, която включва компоненти за тренд, сезонност и празнични ефекти, за да предложи адаптивен подход към променящите се условия в данните. Моделът се отличава с гъвкавост и е подходящ за прогнози в търговията на дребно, където потребителското поведение често

следва сезонни модели, свързани с празници и промоционални периоди. Основното уравнение за Prophet е:

$$(2.13)\ y(t) = g(t) + s(t) + h(t) + \epsilon_t,$$

където $g(t)$ е функцията на тренда, $s(t)$ представлява сезонността, $h(t)$ описва празничните ефекти, а ϵ_t е случайният шум. Prophet е ефективен при прогнози с ясно изразени сезонни и празнични ефекти и не изисква сложни предварителни настройки.

Основното предимство на Prophet е способността му, автоматично да идентифицира и приспособява сезонните цикли и празничните ефекти, което го прави особено полезен за прогнози с дългосрочни и повтарящи се цикли. Недостатъкът му е, че е ограничен, когато данните не показват ясна сезонност или когато са налице нелинейни зависимости, които моделът трудно улавя.

2.2.5. NeuralProphet

NeuralProphet е хибриден модел, който съчетава елементите на Prophet с дълбоките невронни мрежи, като LSTM. Алгоритъмът е проектиран да обработва по-сложни и нелинейни времеви редове чрез добавяне на рекурентни слоеве, които помагат да се улавят дългосрочни зависимости. Това го прави подходящ за данни с комбинации от сезонност и други нелинейни ефекти, които са трудно предвидими чрез стандартни методи.

Основното уравнение за NeuralProphet включва добавка за нелинейния компонент:

$$(2.14)\ y(t) = g(t) + s(t) + h(t) + n(t) + \epsilon_t,$$

където $n(t)$ представлява нелинейните компоненти, обработени чрез невронна мрежа. Тази структура позволява на NeuralProphet да предоставя точни прогнози в ситуации, когато времевият ред показва сложни взаимодействия и многопластова динамика.

Предимствата на NeuralProphet включват неговата гъвкавост и способността му да обработва по-комплексни данни, като интегрира компоненти за сезонност и нелинейни зависимости. Основен недостатък е сложността на модела и необходимостта от повече изчислителни ресурси и време за обучение в сравнение със стандартния Prophet, което го прави по-малко подходящ за бързи и опростени анализи.

2.3. Измерители за грешката на прогнозата

За да се осигури максимална точност и стабилност на прогнозите, изборът на модел трябва да бъде съобразен с особеностите на данните и с

предварителната цел на анализа. При оценяването на ефективността на тези алгоритми ключова роля играят функциите на загуба, които предоставят измерители на отклонението между прогнозните и реалните стойности. Следващата част ще се фокусира върху различните видове подобни функции, тяхната роля в обучението на моделите и как изборът на подходяща функция може значително да повлияе на качеството на прогнозите.

2.3.1. Средна абсолютна грешка (MAE)

Средната абсолютна грешка (Mean Absolute Error, MAE) измерва средното отклонение между прогнозираните и реалните стойности, без да дава по-голяма тежест на по-големите грешки, което я прави полезна за сценарии, в които се изисква баланс между малки и големи отклонения. Формулата на MAE е:

$$(2.15)\ MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$$

Тъй като MAE е лесна за интерпретация и е приложима при различни типове данни, тя е подходяща за начален анализ на модела. Силата на MAE е в нейната устойчивост на големи грешки, което я прави подходяща за данни с много шум (Chai, 2014). Недостатъкът е, че не придава достатъчна тежест на значителните отклонения, които могат да са особено важни в бизнес приложенията, свързани с прогнозиране на търсенето.

2.3.2. Средна квадратична грешка на втора степен (MSE)

Средната квадратична грешка на втора степен (Mean Squared Error, MSE) измерва средното от квадратите на отклоненията между прогнозите и реалните стойности и е изключително чувствителна към големи отклонения. Тя е дефинирана чрез формулата:

$$(2.16)\ MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

MSE се използва широко, когато големите грешки имат критично значение, защото придава по-голяма тежест на отклоненията (Willmott, 2005). Това я прави подходяща за прогнози, при които е важно да се избегнат сериозни грешки. Недостатък на MSE е, че тя може да бъде силно повлияна от единични аномалии и по този начин да увеличи средната стойност, което понякога прави интерпретацията трудна.

2.3.3. Средна квадратична грешка (RMSE)

Средната квадратична грешка (Root Mean Squared Error, RMSE) се изчислява като квадратния корен от MSE, което улеснява интерпретацията,

тъй като стойността е в същите мерни единици като оригиналните данни. RMSE е дефинирана като:

$$(2.17)\ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

RMSE комбинира предимствата на MSE с лесната интерпретация на MAE, като е особено полезна за модели, които са чувствителни към големи грешки (Chai, 2014). Тази функция е ефективна при оценяване на прогнози, при които големите грешки са особено критични. Основният недостатък е, че тя също е податлива на аномалии и може да придаде прекомерна тежест на големите отклонения.

2.3.4. Средна абсолютна процентна грешка (MAPE)

Средната абсолютна процентна грешка (Mean Absolute Percentage Error, MAPE) измерва средния процент на грешката между прогнозните и реалните стойности, като така позволява да се сравняват грешки между различни мащаби на данните. Формулата на MAPE е:

$$(2.18)\ MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100$$

MAPE е полезна, когато е необходимо да се оценява грешката в процентно изражение, тъй като това прави сравнението между различни времеви редове по-лесно и интуитивно (Hyndman, 2018). Също така тя е лесна за интерпретация и позволява да се види какъв процент от реалните стойности се отклонява. Ограничение на MAPE е, че тя може да се изкриви при много малки стойности на y_i, което прави приложението ѝ по-подходящо за данни с по-големи стойности.

2.3.5. Huber Loss

Huber Loss е по-сложна функция, която съчетава предимствата на MAE и MSE, като е по-малко чувствителна към аномалии и запазва линейна реакция към малки грешки. Huber Loss е полезна в случаи, когато данните съдържат аномалии, тъй като тя не придава прекомерна тежест на грешките над определен праг. Тази функция е подходяща за ситуации, където трябва да се контролира влиянието на големите грешки, като същевременно се запазва добра точност за по-малките грешки (Huber, 1964).

2.3.6. Коефициент на определението (R-Squared)

Коефициентът на определението (R-Squared) измерва колко добре моделът обяснява вариацията в данните. Тя не е функция на загубата в класическия смисъл, но често се използва за оценка на качеството на моделите,

като представя процент от вариацията в данните, обяснена от модела. R^2 е дефинирана като:

$$(2.19)\ R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Основното предимство на R^2 е, че е лесен за интерпретация и показва колко от вариацията в данните е уловена от модела. Въпреки това показателят не предоставя информация за точността на конкретните прогнози, а по-скоро за общото качество на модела (Nagelkerke, 1991).

Изборът на подходяща функция на загубата зависи от специфичните изисквания на задачата за прогнозиране, характера на данните и допустимостта на аномалии. Тези функции служат като основни показатели за обучението и оптимизацията на моделите, което ги прави важен компонент в постигането на точни и надеждни прогнози. В следващата секция ще навлезем в задълбочен анализ при идентифицирането на подходящи класически алгоритми, както и на такива от машинното обучение. Въведен ще бъде и изборът на най-подходяща функция за сравнителен анализ.

3. Дискусия

Сравнителният анализ на методите за прогнозиране на търсенето в ритейл сектора изисква разглеждането както на класическите статистически подходи, така и на съвременните алгоритми за машинно обучение. Тази секция обосновава избора на конкретни алгоритми и функции на загуба, които ще бъдат използвани в бъдещ сравнителен анализ, като се опира на последната научна литература и конкретни публикации със сходни проблеми.

Класическите статистически модели като линейната регресия, ARIMA и експоненциалното изглаждане са доказали своята ефективност при анализ и прогнозиране на времеви редове. В изследването на (Hyndman, 2018) ARIMA е използван като един от основните подходи за прогнозиране на времеви редове, особено когато данните са стационарни или могат да бъдат трансформирани до стационарност. Методът на Холд и Уинтърс за тройно експоненциално изглаждане е изследван в контекста на краткосрочни прогнози, включително за търсенето на бързооборотни стоки, където сезонността играе важна роля (Winters, 1960).

ARIMA и методите за експоненциално изглаждане показват стабилна производителност при прогнозиране на времеви редове с линеарни зависимости (Makridakis, 2020). В резултат на това, тези модели остават основни кандидати за задачата по прогнозиране на търсенето, когато е налице сезонност или тренд. В този контекст, за нашия сравнителен анализ, ARIMA и тройното експоненциално изглаждане се открояват като основни пред-

ставители на класическите подходи. ARIMA може да улови сложни линейни зависимости и да отчита авторегресивни и сезонни компоненти, докато методът на Холд и Уинтърс се използва за бърза реакция при краткосрочни колебания в данните.

С развитието на машинното обучение и нарастващата достъпност до големи данни се появиха нови методи, които предоставят по-голяма гъвкавост и точност. Съвременните подходи като Random Forest и XGBoost са използвани успешно за прогнозиране в различни домейни, включително прогнозиране на търсенето в ритейл сектора. Random Forest е използван за прогнозиране на продажби и показва висока точност при сложни, многомерни данни, включващи влияние от промоции и сезонност (Yao, 2021), като разчита на ансамбли от решаващи дървета, които осигуряват устойчивост и помагат да се избегне свръхобучението.

От палитрата на алгоритми от машинното XGBoost е мощен градиентно усилващ метод, който е широко използван за прогнозиране в ритейл сектора, показвайки значително по-добри резултати в сравнение с класическите методи, особено в контекста на динамично променящи се условия (Wang, 2020). Този алгоритъм включва регуляризация, която намалява риска от свръхобучение и осигурява висока точност дори при големи набори от данни с множество взаимодействащи фактори.

Дълбокото обучение и вчастност LSTM мрежите са един от най-популярните подходи за прогнозиране на времеви редове, когато има нужда от моделиране на дългосрочни зависимости и сложни нелинейни взаимодействия. Този тип мрежи са доказано ефективни, когато данните включват множество фактори като сезонност, промоции и дългосрочни тенденции (Brownlee, 2021). Способността на LSTM да обработва дълги последователности от данни, го прави подходящ за прогнозиране на търсенето на продукти, които се влияят от дългосрочни цикли и неочаквани аномалии.

Разработеният от Facebook Prophet, от друга страна, е гъвкав модел за прогнозиране на времеви редове с ясно изразени сезонни и празнични компоненти. Основното предимство на Prophet е, че не изисква специални предварителни настройки на данните и позволява лесна интерпретация на прогнозите, което го прави подходящ за широк спектър от потребителски приложения (Saboo, 2022).

Като модел, хибрид между различните разгледани алгоритми, NeuralProphet се характеризира с предимствата на Prophet с рекурентни невронни мрежи за обработка на сложни зависимости, като показва значителна подобрена точност в сравнение със стандартния Prophet, особено при работа с данни, които включват както сезонни, така и нелинейни зависимости (Sun, 2022). Това прави NeuralProphet подходящ избор за анализ на търсенето в ситуации, когато е необходимо да се обработят множество зависимости и да се моделират различни видове взаимодействия.

След като разгледахме различните класически и модерни подходи, за бъдещ сравнителен анализ с български данни за търсенето на бързооборотни стоки, избрахме ARIMA и тройното експоненциално изглаждане от класическите методи и Prophet и NeuralProphet от модерните подходи за машинно обучение. Тези четири модела представляват баланс между традиционни и съвременни решения, като осигуряват висока точност и гъвкавост при прогнозирането на търсенето на бързооборотни стоки. ARIMA и тройното експоненциално изглаждане са избрани заради доказаната им ефективност при времеви редове със сезонност и трендови компоненти, докато Prophet и NeuralProphet предлагат съвременни решения за обработка на комплексни данни и нелинейни зависимости.

Изборът на метрики за оценка на тези модели също играе ключова роля в анализа. Средната абсолютна грешка (MAE) е избрана, защото предоставя лесна интерпретация и измерва средното отклонение между прогнозираните и реалните стойности. MAE е използвана като основна метрика за сравнение на прогностични модели и е показала, че е полезна при анализ на различни видове времеви редове.

За да бъде придадена голяма тежест на значителните отклонения, когато големите грешки трябва да бъдат минимализирани, средната квадратична грешка (MSE) ще бъде използвана също като метрика. MSE се прилага като основна метрика за оценка на точността на моделите при прогнози на времеви редове и е показала, че осигурява добра основа за оценка на моделите, които трябва да минимализират големи грешки. Коренната средна квадратична грешка (RMSE) също ще бъде използвана, тъй като предоставя резултати в същите мерни единици като оригиналните данни, което улеснява интерпретацията.

Като допълнителна метрика в анализа ще бъде включена и средната абсолютна процентна грешка (MAPE), която позволява измерването на грешките в процентно отношение и улеснява сравнението на модели при различни мащаби на данните. Hyndman и Koehler (2021) отбелязват, че MAPE е полезна при оценка на модели с различни мащаби на стойностите и осигурява разбиране на прогнозната грешка в контекста на общата стойност на времевия ред.

В заключение, избраните модели за сравнителен анализ – ARIMA, тройното експоненциално изглаждане, Prophet и NeuralProphet – предоставят комбинация от класически и модерни подходи, които осигуряват висока точност и гъвкавост при прогнозирането на търсенето в ритейл сектора. Метриките за оценка като MAE, MSE, RMSE и MAPE ще предложат пълна картина на ефективността на тези модели и ще помогнат да се оцени тяхната пригодност за различни типове прогнози. Този сравнителен анализ ще предостави задълбочено разбиране за приложимостта на различните подходи в ритейл прогнозите и ще даде насоки за избор на оптимален модел според спецификите на данните и бизнес целите.

Заключение

В обобщение, казаното по-горе потвърждава важността на прецизното прогнозиране на търсенето в ритейл сектора, особено когато става въпрос за бързооборотни стоки. С развитието на технологиите и нарастващото използване на големи данни, задачата за прогнозиране става все по-комплексна и изисква иновативни решения, които могат да отговорят на динамиката на съвременния пазар. В този труд разгледахме различни подходи за прогнозиране на търсенето – както класически, така и съвременни методи на машинното обучение – като обосновахме избора на конкретни модели и метрики за оценка, които ще бъдат използвани в бъдещ сравнителен анализ.

Класическите статистически модели като ARIMA и методът на Холд и Уинтърс са добре установени и доказани методи за прогнозиране на времеви редове. Тези подходи предлагат стабилност и лесна интерпретация, което ги прави особено подходящи за случаи, в които данните са линейни и не включват сложни зависимости или многомерни взаимодействия. ARIMA моделът е особено ефективен за прогнозиране на стационарни времеви редове, където сезонните и трендовите компоненти могат да бъдат ясно идентифицирани и моделирани. Методът на Холд и Уинтърс, от своя страна, осигурява добри резултати при краткосрочни прогнози на данни със сезонни колебания.

Модерните подходи от машинното обучение, като Prophet, NeuralProphet, Random Forest, XGBoost и LSTM, предлагат по-голяма гъвкавост и способност да се справят със сложни зависимости в данните. Prophet е предпочитан модел за прогнозиране на времеви редове с ясно изразени сезонни и празнични компоненти. Той е лесен за настройка и позволява моделиране на празнични ефекти и различни сезонни компоненти без сложна предварителна обработка. NeuralProphet, който е разширение на Prophet, добавя рекурентни слоеве за улавяне на сложни зависимости, като осигурява висока точност при динамични и нелинейни данни. Random Forest и XGBoost се открояват със своята способност да обработват големи обеми данни и да улавят нелинейни взаимодействия между множество фактори. LSTM мрежите, от своя страна, са подходящи за дългосрочно прогнозиране и улавяне на сложни последователни зависимости в данните.

Като конкурентни подходи бяха избрани четири модела, които представляват най-добрите практики както от класическите, така и от модерните подходи: ARIMA, тройното експоненциално изглаждане на Холд и Уинтърс, Prophet и NeuralProphet. Тези модели са избрани, за да осигурят баланс между стабилност, лесна интерпретация и гъвкавост при обработка на сложни данни. ARIMA и методът на Холд и Уинтърс представляват класически методи, които са добре установени и доказали своята ефективност при прогнозиране на времеви редове с линейни и сезонни компоненти. Prophet и

NeuralProphet, от своя страна, осигуряват гъвкавост и точност при динамични данни и са особено подходящи за обработка на сложни зависимости.

Метриките за оценка на тези модели включват Средната абсолютна грешка (MAE), Средната квадратична грешка (MSE), Коренната средна квадратична грешка (RMSE) и Средната абсолютна процентна грешка (MAPE). MAE предоставя информация за средното отклонение между прогнозите и реалните стойности, без да придава допълнителна тежест на големите грешки. MSE и RMSE придават по-голяма тежест на по-големите грешки, което е важно, когато е критично да се избегнат значителни отклонения. MAPE осигурява процентно изражение на грешките, което улеснява сравнението между различни времеви редове и различни мащаби на данните. Тези метрики ще осигурят задълбочена оценка на точността и ефективността на моделите и ще помогнат за избора на най-подходящия подход в зависимост от спецификите на данните и бизнес целите.

Прогнозирането на търсенето в ритейл сектора е от критично значение за ефективното управление на веригите на доставки, оптимизиране на наличностите и минимализиране на разходите за съхранение. В динамична пазарна среда, където потребителското поведение се променя бързо, точните прогнози могат да бъдат ключови за запазване на конкурентоспособността на бизнеса и задоволяване на клиентските нужди. Точността на прогнозите не само помага за оптимизиране на складовите наличности, но и подобрява планирането на производството, дистрибуцията и маркетинговите кампании.

Възможностите за следващи изследвания включват провеждането на сравнителен анализ на представените четири модела – ARIMA, тройно експоненциално изглаждане, Prophet и NeuralProphet – върху действителни пазарни данни с използване на метриките за оценка на точността. Резултатите от този анализ ще предоставят ясна представа за силните и слабите страни на всеки от моделите в различни сценарии на прогнозиране и ще помогнат за оптимизиране на процесите в ритейл сектора, свързани с управление на наличностите и планиране на продажбите. Също така може да бъде изследвано използването на хибридни модели, които комбинират класически и модерни подходи, с цел постигане на по-добра точност и гъвкавост при прогнозиране.

Друга възможност за бъдещи изследвания е включването на допълнителни променливи и външни фактори в моделите, като икономически индикатори, данни за времето, данни от социални медии и други източници, които могат да повлияят на търсенето. Включването на тези фактори може да подобри тяхната точност и да предостави по-задълбочена представа за факторите, които влияят на потребителското поведение. Освен това анализ на сезонните и празнични ефекти и тяхното влияние върху търсенето в различни региони и сегменти от ритейл сектора също може да бъде полезен за

по-добро разбиране на динамиката на търсенето и оптимизиране на бизнес процесите.

В заключение настоящата публикация подчертава важността на избора на подходящ модел за прогнозиране на търсенето в ритейл сектора в зависимост от характеристиките на данните и специфичните бизнес нужди. Класическите модели като ARIMA и метода на Холд и Уинтърс предлагат стабилност и лесна интерпретация, докато съвременните подходи като Prophet и NeuralProphet предоставят гъвкавост и способност да се справят с нелинейни зависимости и сложни взаимодействия. Изборът на подходящи метрики за оценка е от критично значение, за да се гарантира обективна и точна оценка на ефективността на моделите. Бъдещите изследвания ще бъдат насочени към сравнителен анализ на избраните модели и изследване на възможностите за включване на допълнителни фактори и хибридни подходи, които да подобрят точността и надеждността на прогнозите. Това ще помогне за оптимизиране на процесите в ритейл сектора и ще предостави практически насоки за използването на различни методи в зависимост от характеристиките на данните и бизнес целите.

Използвани източници

Колев, Н. Г. (2019). Прогнозиране на търсенето на бързооборотни стоки в ритейл сектора. *Управление и икономика*, 45-60.

Колев, Н. И. (2019). Моделиране на търсенето на бързооборотни стоки в ритейл сектора. *Икономически анализи*, 88-97.

Костова, В. И. (2020). Анализ на точността на прогнозиране на търсенето с различни статистически методи. *Икономически изследвания*, 201-215.

Марков, И. И. (2018). Приложение на регресионни модели в прогнозирането на търсенето. *Икономически изследвания, 27*(2), 102-115.

Тодорова, М. &. (2020). Приложение на машинното обучение в прогнозирането на продажби. *Търговски и икономически изследвания*, 145-158.

Чобанов, П. (2020). *Анализ на времеви редове в икономическите прогнози.* София, България: Издателство на БАН.

Ailawadi, K. L. (2009). The impact of promotions on consumer behavior. *Journal of Marketing Research*, 420-435.

Box, G. E. (1976). *Time series analysis: forecasting and control.* Holden-Day.

Breiman, L. (2001). Random forests. *Machine Learning*, 5-32.

Brownlee, J. (2021). Time series forecasting with LSTM neural networks. *Machine Learning Mastery*.

Chai, T. &. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247-1250.

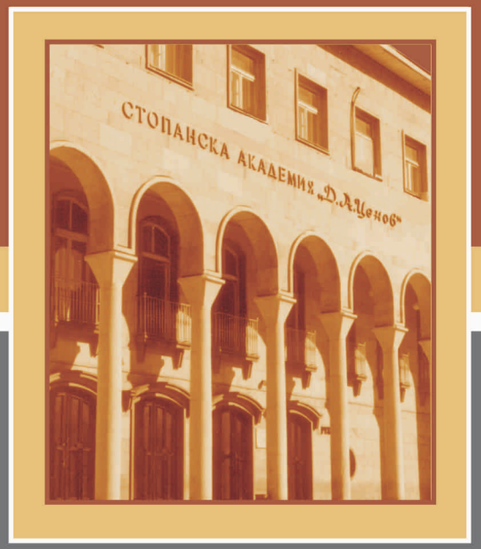
Chen, H. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 1165-1188.
Chen, T. &. (2016). Xgboost: A scalable tree boosting system. 785-794.
Fildes, R. &. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society, 59*(9), 1150-1172.
Goodfellow, I. B. (2016). *Deep Learning.* Cambridge, MA, USA: MIT Press.
Hochreiter, S. &. (1997). Long short-term memory. *Neural Computation*, 1735-1780.
Huber, P. J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, 73-101.
Hyndman, R. J. (2018). *Forecasting: principles and practice.* OTexts.
Khashei, M. &. (2011). A hybrid methodology for forecasting: integrating artificial neural networks and ARIMA models. *Applied Soft Computing, 11*(2), 2664-2675.
Kotler, P. &. (2016). *Marketing Management.* Pearson Education.
Makridakis, S. S. (2020). The M5 competition: Results, findings, and conclusions. *International Journal of Forecasting*, 54-74.
Montgomery, D. C. (2008). *Introduction to Time Series Analysis and Forecasting.* Wiley.
Nagelkerke, N. J. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, 691-692.
Saboo, A. R. (2022). The Impact of Machine Learning on Retail Demand Forecasting: A Comprehensive Review. *Journal of Retail and Consumer Services*, 102-115.
Sun, J. C. (2022). Enhanced retail demand forecasting using NeuralProphet. *Retail Data Science Journal, 7*(2), 99-110.
Tashman, L. J. (2020). Improving Forecast Accuracy in Retail with Machine Learning. *International Journal of Forecasting*, 150-165.
Taylor, S. J. (2018). Prophet: forecasting at scale.
Triebe, M. O. (2021). Advances in Demand Forecasting with Neural Networks. *Journal of Data Science and Analytics*, 230-245.
Wang, H. Z. (2020). Improving retail demand forecasting using XGBoost. *Journal of Business Forecasting, 38*(4), 45-53.
Willmott, C. J. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 79-82.
Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 324-342.
Yao, L. C. (2021). Application of Random Forest in retail sales forecasting.

СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ“ - СВИЩОВ

ГОДИШЕН

АЛМАНАХ

НАУЧНИ ИЗСЛЕДВАНИЯ НА ДОКТОРАНТИ



Том XVII, 2024
Книга 20

Академично издателство
„ЦЕНОВ“- Свищов

НАУЧНИ ИЗСЛЕДВАНИЯ НА ДОКТОРАНТИ

ГОДИШЕН АЛМАНАХ

Том XVII, 2024 г.
Книга 20

СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ“ • СВИЩОВ •

ISSN 1313-6542

Съдържание

Студии

Статии