

УДК 004.891.3

## КОНЦЕПЦИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ В ПРОГНОЗИРОВАНИИ ДЛИТЕЛЬНОСТИ ТРАНСПОРТНЫХ ОПЕРАЦИЙ

Саляев Илья Сергеевич<sup>1</sup>, аспирант

e-mail: [salyaev\\_i@mail.ru](mailto:salyaev_i@mail.ru)

<sup>1</sup> Волжский государственный университет водного транспорта, Нижний Новгород, Россия

**Аннотация.** В данной статье рассматривается возможность применения технологий машинного обучения для прогнозирования длительности транспортных операций, и, в частности, точного времени прибытия судна в порт.

**Ключевые слова:** транспортная логистика, проблемы логистики, цифровые технологии, машинное обучение, прогнозирование транспортных операций, управление флотом.

## THE CONCEPT OF USING MACHINE LEARNING TECHNOLOGIES IN PREDICTING THE DURATION OF TRANSPORT OPERATIONS

Salyaev Ilya Sergeevich<sup>1</sup>, Doctoral Student

e-mail: [salyaev\\_i@mail.ru](mailto:salyaev_i@mail.ru)

<sup>1</sup> Volga State University of Water Transport, Nizhny Novgorod, Russia

**Abstract.** This article discusses the possibility of using machine learning technologies to predict the duration of transport operations, and in particular the exact time of arrival of the vessel at the port.

**Keywords:** transport logistics, logistics problems, digital technologies, machine learning, forecasting of transport operations, fleet management.

### Введение

Прибытие судна в порт предваряет череду ответственных грузовых и коммерческих операций, и включает сложный процесс согласования очередности обработки груза и транспортного средства. Грузовые суда могут проводить в порту до 10-15% общего времени эксплуатации на якорной стоянке на рейде. Прогнозирование точного времени прибытия судов могло бы помочь портовым компаниям оптимально распределить свои ресурсы и сократить время простоя в интересах конкурентоспособности порта [2]. Сейчас использование информационных систем идентификации судов в режиме реального времени дало администрациям портов, судоходным линиям и другим субъектам возможность автоматизировать прогноз момента окончания текущей транспортной операции, а также отслеживать последовательные шаги, которые предпринимает судно при заходе в порт.

Расчётное время прибытия (англ., Estimated Time of Arrival –ETA) играет ключевую роль в оперативном планировании деятельности порта и операциями по обслуживанию судов [1]. Неопределённость или неточность продолжительности транспортных операций затрудняет возможность разрабатывать эффективные логистические схемы, подчёркивая решающую роль точных прогнозов времени прибытия.

Цифровые технологии и методы машинного обучения (МО) обладают большим потенциалом для поддержки процессов принятия решений, позволяя прогнозировать неизвестную информацию [10]. Способность извлекать и аппроксимировать системные взаимосвязи из обучающих данных без явных априорных знаний делает их очень подходящими для моделирования очень сложных и динамичных систем. Благодаря способности систем, основанных на МО, к обучению, задачи могут решаться более гибко, с меньшими усилиями и с более высокой точностью.

На этом фоне приложения на основе машинного обучения приобретают особое значение, особенно в области логистики [7]. Прогнозы ETA на основе машинного обучения могут внести важный вклад в совершенствование современных логистических сетей, на которые влияют растущие требования клиентов с точки зрения надёжности, прозрачности, устойчивости и экономической эффективности.

### **Требования логистики к прогнозированию точного времени прибытия судна**

Прибытие судна в вовремя в место назначения является серьёзной задачей для логистики. В настоящее время, прогнозирование времени прибытия речного судна в заданную точку основывается на сборе информации о прохождении этим судном некой контрольной точки (дислокации), например, шлюза или крупного населённого пункта. Далее, ответственно лицо - капитан или диспетчер, оценивая расстояние, называет предполагаемое время прибытия. Данный метод имеет огромную погрешность, измеряемую часами, время завершения рейса часто указывается календарными сутками.

А, между тем, точное и своевременное оперативное регулирование работы флота играет решающую роль в обеспечении успеха речных перевозок, что жизненно важно для поддержания эффективности транспортных операций и обеспечения стабильности всей цепочки поставок [5, 6].

В контексте терминальных операций время прибытия судов непосредственно влияет на распределение ресурсов и организацию транспортной деятельности, включая последовательность и продолжительность погрузочных операций, а также использование терминальных мощностей. Кроме того, пунктуальность прибытия судна имеет решающее значение для эффективного управления логистикой, поскольку это способствует своевременной доставке товаров и уменьшает перебои в цепочке поставок. Это особенно актуально для компаний, которые придерживаются системы доставки «точно в срок», где любая задержка с прибытием судна может вызвать цепную реакцию задержек и неэффективности во всей цепочке поставок.

Кроме того, своевременное прибытие судна может оказать положительное влияние на контроль затрат. Хотя внутренние водные перевозки часто рассматриваются как экономически эффективный вид транспорта по сравнению с другими альтернативами, такими как автомобильный или железнодорожный транспорт, задержка прибытия судна может привести к дополнительным расходам, таким как плата за простой, увеличение затрат на рабочую силу и плата за хранение. Эти расходы могут быстро накапливаться и оказывать негативное влияние на финансовые показатели компании, что ещё больше подчёркивает важность своевременного прибытия судов.

Погрешность между прогнозируемым и фактическим значениями времени прибытия судна в контрольную точку не должно превышать десятков минут.



## Система АИС (Автоматическая Идентификационная Система)

Автоматическая идентификационная система (англ., Automatic Identification System – AIS) используется для отслеживания судов и обеспечения безопасности судоходства [1]. Она предоставляет актуальные данные о судне, его местоположении, скорости, курсе, номере и т.д.

Частота передачи сообщений AIS с судов варьируется от 2 до 30 с в зависимости от их скорости. На дальность, на которой могут приниматься эти сообщения, влияют различные факторы, такие как условия распространения сигнала, состояние моря, а также высота и мощность передающей и приёмной антенн. Дальность приёма может варьироваться от 20 морских миль до 350 морских миль при оптимальных условиях. Как правило, ожидается, что сеть приёмников AIS достигнет среднего радиуса приёма около 40 морских миль.

AIS работает путём получения данных о местоположении и перемещении от судовой системы GPS или внутреннего датчика в блоке AIS. Эти данные, наряду с другой программируемой информацией из блока AIS (такой как идентификационный номер морской мобильной службы (MMSI), название судна, пункт назначения и тип груза), периодически передаются. Система не только отправляет информацию, но и получает данные из систем AIS других судов.

Каждое сообщение AIS содержит как статическую, так и динамическую информацию. Статическая информация включает атрибуты судна, в то время как динамическая информация охватывает пространственно-временные данные судна.

Существует множество служб, которые собирают и распространяют сообщения AIS с судов по всему миру с высокой частотой и относительно низкой задержкой [3]. Таким образом, данные AIS являются идеальным источником информации для прогнозирования ожидаемого времени прибытия судна (ETA).

### Ожидаемое время прибытия (ETA)

Ожидаемое время прибытия (ETA) — это прогнозируемое время прибытия судна в точку назначения [1]. Оно зависит от ожидаемой скорости судна на разных участках маршрута и ожидаемой продолжительности операций. Формула выглядит следующим образом:

$$ETA = T_{\text{тек}} + \sum \frac{L_{i\text{-го участка}}}{V_{i\text{ож}}} + \sum t_{i\text{ож}} \quad (1)$$

где  $T_{\text{тек}}$  – текущий момент времени (или время по дислокации судна).

$L_{i\text{-го участка}}$  – длина  $i$ -го участка, на котором будет соблюдаться ожидаемая скорость.

$V_{i\text{ож}}$  – ожидаемая скорость на  $i$ -м участке.

$t_{i\text{ож}}$  – ожидаемая продолжительность технологических операций. Например, шлюзование погрузочно-разгрузочные работы, бункеровка, комплексное обслуживание и т.д.

Средняя ожидаемая скорость ( $V_{i\text{ож}}$ ) — это скорость, которая предполагается на  $i$ -м участке пути. Она зависит от нормативной скорости при номинальном режиме работы энергетической установки судна и от корректировочной скорости.

$$V_{\text{ож}} = V_{\text{норм.}} \pm V_{\text{корр.}} \quad (2)$$

где  $V_{\text{норм.}}$  – нормативная скорость. Средняя скорость конкретного судна по документации.

$V_{\text{корр.}}$  – корректировочная скорость. Она зависит от текущих судоходных условий и решения капитана или иного ответственного лица.

Нормативная скорость тоже может быть скорректирована, например, если судно по каким-то причинам движется с сильным отклонением от норматива. Соответственно алгоритм машинного обучения должен непрерывно анализировать нормативную и корректировочную скорости.



Теперь рассмотрим условия, которые влияют на корректировку ожидаемой скорости.

- Фактическая скорость судна. Судно часто может двигаться со скоростью, отличающейся от нормативной.
- Ветер и течение. Сила и направление ветра и течения могут существенно влиять на скорость судна.
- Глубина на участке. Глубина воды может ограничивать максимальную скорость судна, особенно в узких участках водных путей.
- Осадка (загрузка) судна. Чем больше загружено судно, тем выше сопротивление водной среды, тем меньше будет его скорость, особенно на мелководных участках.
- Близость к месту шлюзования. Чем ближе судно к месту шлюзования, тем меньше его скорость.
- Загруженность определённого участка. Количество судов, особенно встречных на конкретном участке маршрута может существенно уменьшить ожидаемую скорость.

### **Машинное обучение**

Машинное обучение (МО) является поддоменом искусственного интеллекта (ИИ) и включает в себя различные методы, которые позволяют компьютерным системам независимо извлекать закономерности из обширных данных [4, 10]. Таким образом, МО позволяет компьютерным системам обучаться индуктивно. Это означает, что логический вывод происходит на основе гипотетических корреляций, которые обучающий алгоритм приобрёл в ходе процесса обучения, адаптируясь к наблюдениям и обобщая содержащиеся в них закономерности. Это автоматическое извлечение закономерностей из данных позволяет распознавать сложные взаимосвязи, которые не распознаются человеком или требуют больших усилий, и использовать их в промышленности, например, для сегментации и прогнозирования информации, вывода правил и решения задач оптимизации. Подходы к МО можно условно разделить на три основных типа: контролируемое, неконтролируемое обучение и обучение с подкреплением. Для прогнозирования ETA, контролируемого и неконтролируемого обучения имеет особое значение, поскольку обучение в основном основано на исторических транспортных данных.

### **Схематичное ТЗ модели ETA на основе машинного обучения**

В настоящее время существует тысячи моделей грузовых судов. Каждая модель имеет индивидуальные эксплуатационные характеристики. Суда отличаются по грузоподъёмности, скорости габаритам, типу перевозимого груза и т.д. Следовательно, для каждой модели судна должна разрабатываться модель прогнозирования ETA индивидуально.

Разработка модели прогноза ETA, используя технологии машинного обучения осуществляется в несколько этапов. Ниже приведена схема технической реализации проекта [9, 10].



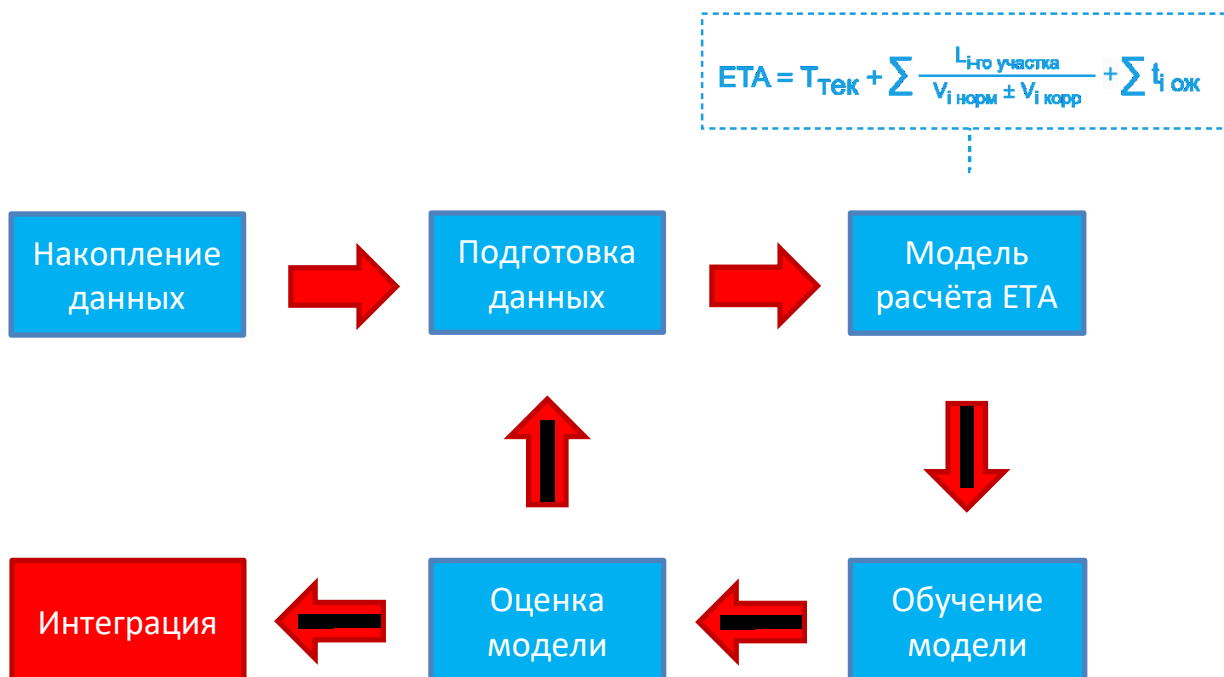


Рисунок 1 – Схема технической реализации модели расчёта ETA

#### 1) Накопление данных.

Для начала реализации модели нам необходимо иметь достаточное количество данных [10]. Они представляют собой детальную информацию о предыдущих рейсах. Эти данные мы будем собирать из разных источников, в том и описанной ранее системы AIS, и хранить в базе данных. Какие данные нам нужны:

- Исторические данные. Данные о предыдущих рейсах судна, включая время отправления, прибытия, и фактическое время в пути.
- Данные о судне. Тип судна, год выпуска, грузоподъёмность, скорость и техническое состояние.
- Данные о движении судна. Информация о фактической скорости судна на участках рейса, его загрузенности, задержках и т.д.
- Данные о погоде. Скорость и направление ветра, и течения, сезонность.
- Данные о маршруте. Данные о водном пути, пунктах отправления и назначения, остановки в портах, шлюзования, загрузенность участка.

#### 2) Подготовка и анализ данных.

Для выявления потенциальных проблем с данными, их необходимо проанализировать [8, 10]. Необходимо понимать структуру данных, распределений, корреляций и пропущенных значений.

Модель не может напрямую использовать необработанные данные. Следовательно, необходима предварительная обработка перед вводом данных в модель.

Первый этап в подготовке данных — это очистка мусорных данных. Удаляются данные с пропущенными значениями, выбросы и ошибки в данных.

Второй этап — это преобразование всех данных в числовой формат, с которым сможет работать модель. Поскольку машинное обучение не может работать с текстом. Это может включать в себя кодирование категориальных переменных.

Следующий этап включает в себя создание новых признаков на основе накопленных данных. Это может помочь извлечь из данных более информативные характеристики.

Заключительный этап подготовки — это разделение данных на обучающий, валидационный и тестовый наборы. Обучающий набор используется для оценки и проверки модели, валидационный для настройки гиперпараметров (например, скорость обучения) и тестовый для оценки производительности.

### 3) Модель расчёта ETA.

Сама модель основывается на выведенной нами формуле расчёта ETA. У каждого компонента этой формулы, в контексте машинного обучения имеется свой вес. Вес — это показатель, который определяет насколько сильно тот или иной признак в конечном счёте влияет на ETA.

Для построения линейных связей между признаками (расстояние, скорость, погодные условия, загруженность и др.) и временем прибытия используется линейная регрессия [10].

В задачах регрессии для построения высокоточных моделей используется градиентный бустинг. Такая модель способна предсказывать непрерывные числовые значения на основе входных признаков.

Градиентный бустинг строит ансамбль (объединение) слабых (необученных) моделей последовательно, каждая новая модель исправляет ошибки предыдущей модели [4]. То есть, новая модель учится предсказывать остатки, оставшиеся после предыдущих моделей.

Существуют библиотеки, такие как XGBoost, LightGBM и CatBoost, которые предоставляют эффективные реализации градиентного бустинга.

### 4) Обучение модели.

Процесс обучения модели заключается в нахождении оптимальных параметров и весов этих параметров, которые минимизируют ошибку предсказания модели [1, 8, 10]. Обычно в качестве ошибки используется среднеквадратичная ошибка (MSE). Для её минимизации используется метод наименьших квадратов и градиентный спуск.

Метод наименьших квадратов (МНК) используется для нахождения оптимальных параметров модели путём минимизации суммы квадратов разностей между фактическими значениями и предсказанными моделью значениями.

Градиентный спуск — это итеративный метод обучения, который находит оптимальные веса для каждой модели таким образом, чтобы итоговое предсказание модели было наиболее близким к истинным значениям целевой переменной.

### 5) Оценка модели.

Чтобы определить качество прогнозируемого ETA используются разные способы оценки моделей [1, 2, 8]. Они показывают разницу между предсказанными и фактическим значениями. Для оценки модели линейной регрессии и градиентного бустинга используется среднеквадратичная ошибка (MSE) и коэффициент детерминации ( $R^2$ ).

Среднеквадратичная ошибка — это среднее значение квадратов разностей между фактическими и предсказанными значениями.

Коэффициент детерминации показывает, насколько хорошо модель соответствует данным. Если  $R^2 = 1$ , то модель идеально соответствует данным, если  $R^2 = 0$ , то модель не объясняет изменчивость данных вообще. Чем ближе  $R^2$  к 1, тем лучше модель объясняет данные, т.е. предсказывает значения, основываясь на входных данных.

В случае, если ошибка превышает заданные допустимые значения, то процесс обучения повторяется.

### 6) Интеграция.

При достижении достаточного качества прогнозирования ETA, модель необходимо преобразовать в программу, написанную на определённом языке программирования. И далее интегрировать с другими информационными системами на судне.



## Коммерциализация проекта

Рассмотрим целевую аудиторию, которой может быть выгодна и необходима информация о точном времени прибытия судна в речной порт.

- Транспортные компании. В первую очередь, данная идея будет интересна транспортным компаниям. Прогнозирование точного времени прихода судна в определённую точку позволит давать заказчику более точную информацию о доставке его груза.

- Администрация порта. Получение информации о точном прибытии судна позволяет порту планировать операционные вопросы и выстраивать очередность обработки судов, а также выполнять подготовительную работу на причале.

- Администрация бассейна. Как контролирующий орган Администрация бассейна должна отвечать за безопасность и понимать какое судно и в какое время будет следовать по участку в их зоне ответственности.

- Судоводители. Для понимания сколько времени будет длиться определённый маршрут.

- Грузоотправители. Потому что у них будет точная информация по времени доставки их груза.

- Сервисные компании. Компании, которые занимаются ремонтом, заправкой судов, различные плав магазины и т.д. будут владеть информацией о времени предложения своих услуг.

## Заключение

В этой статье рассмотрена актуальность и методология точного прогнозирования времени прибытия судна в речной порт технологиями машинного обучения.

Машинное обучение обладает огромным потенциалом для точного прогнозирования продолжительности логистических процессов [8 – 10]. Однако на это качество в корне влияют доступность и объём обучающих данных. Особенно в сфере транспорта, где часто задействовано множество компаний, интеграция данных между участниками представляет собой серьёзную проблему для проектов машинного обучения. Оптимальный метод МО и подходящие функции заранее неизвестны и должны определяться в ходе многочисленных этапов разработки и оценки. Это приводит к высокой степени повторяемости процесса, в то время как привлечение экспертов предметной области было признано важным критерием успеха.

В целом, можно сделать вывод, что использование машинного обучения для логистики представляет собой высокий потенциал для улучшения технологических процессов с помощью интеллектуальных систем поддержки принятия решений. Это относится не только к представленному варианту использования прогноза ETA, который вносит важный вклад в повышение эффективности, надёжности и гибкости логистических сетей за счёт большей прозрачности сроков обработки транспортных заказов. Это также касается других областей применения с высокой неопределённостью и сложностью.

## Список литературы:

1) Nomayoon Arbabkhah, Atefe Sedaghat, Masod Jafari Kang, Maryam Hamidi. Automatic Identification System-Based Prediction of Tanker and Cargo Estimated Time of Arrival in Narrow Waterways // Chaos: JMSE Volume 12 - Issue 2. – URL: <https://doi.org/10.3390/jmse12020215> (дата обращения: 17.04.2024)

2) Albert Veenstra, Rogier Harmelink. On the quality of ship arrival predictions // Chaos: Marit Econ Logist 23, 655 – 673 (2021). – URL: <https://doi.org/10.1057/s41278-021-00187-6> (дата обращения: 16.04.2024)



- 3) Jussi Poikonen. AI for smart ports, part 1: Limitations of existing data sources for port call prediction. – URL: <https://www.awake.ai/post/ai-for-smart-ports-port-call-prediction> (дата обращения: 16.04.2024)
- 4) Jussi Poikonen. AI for smart ports, part 2: Optimizing vessel schedule predictions using machine learning. – URL: <https://www.awake.ai/post/ai-for-smart-ports-port-call-prediction-part2> (дата обращения: 16.04.2024)
- 5) Jinyu Lei, Zhong Chu, Yong Wu. Predicting vessel arrival times on inland waterways: A tree-based stacking approach. – URL: <https://doi.org/10.1016/j.oceaneng.2024.116838> (дата обращения: 16.04.2024)
- 6) Estimated Time of Arrival (ETA). – URL: <https://windward.ai/glossary/what-is-estimated-time-of-arrival-eta/> (дата обращения: 16.04.2024)
- 7) Peter Wenzel, Raka Jovanovic, and Frederik Schulte. A Neural Network Approach for ETA Prediction in Inland Waterway Transport // Chaos: ICCL 2023: [Computational Logistics](#) pp. 219 – 232. – URL: [https://doi.org/10.1007/978-3-031-43612-3\\_13](https://doi.org/10.1007/978-3-031-43612-3_13) (дата обращения: 16.04.2024)
- 8) Dejan Stepec, Domaz Marctincic, Fabrice Klein, Daniel Vladusic, Joao Pita Costa. Machine Learning based System for Vessel Turnaround Time Prediction // Chaos: 2020 21st IEEE International Conference on Mobile Data Management (MDM). – URL: <https://ieeexplore.ieee.org/document/9162235> (дата обращения: 16.04.2024)
- 9) Alfredo Alessandrini, Fabio Mazzarella, Michele Vespe. Estimated Time of Arrival Using Historical Vessel Tracking Data // Chaos: IEEE Transactions on Intelligent Transportation Systems (Volume: 20, Issue: 1, January 2019). – URL: <https://ieeexplore.ieee.org/document/8294051> (дата обращения: 16.04.2024)
- 10) Peter Poschmann, Manuel Weinke, Frank Straube. Predicting Estimated Arrival Times in Logistics Using Machine Learning // Chaos: Encyclopedia of Data Science and Machine Learning (pp. 2683 – 2701); October 2022. – URL: <https://www.igi-global.com/gateway/chapter/317703> (дата обращения: 16.04.2024)

