

УДК 556.535.5:004.8

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В МОДЕЛИРОВАНИИ ДАТ ЛЕДОВЫХ ЯВЛЕНИЙ НА РЕКАХ БАССЕЙНА ОЗЕРА СЕВАН

Н. А. МАКАРОВ ^{1*}, Е. В. ГАЙДУКОВА ^{1**}, В. Г. МАРГАРЯН ^{2***}

¹ *Российский государственный гидрометеорологический университет (РГГМУ), Санкт-Петербург, Россия*

² *Кафедра общей географии ЕГУ, Ереван, Армения*

В исследовании предложено решение задачи модификации прогнозирования ледовых условий, выбраны два наиболее популярных направления: деревья решений и наивный байесовский классификатор. Объектом выступают реки бассейна оз. Севан, в частности Дзыкнагет, Драхтик, Памбак, Варденис, Бахтак, на которых отмечаются сложные процессы вскрытия и замерзания. Для схематизации и совмещения процесса формирования и разрушения льда в качестве прогнозной характеристики вместо дат начала и конца ледовых явлений используются dummy-переменные, а независимыми параметрами выступают среднедекадные значения температуры воздуха и уровни воды. Согласно удовлетворительным результатам исследования, предложенные модели хуже справляются с определением даты конца ледовых явлений, однако способны работать с малой выборкой и отсутствием функциональной связи.

<https://doi.org/10.46991/PYSUC.2026.60.2.290>

Keywords: ice conditions, classification decision trees, naive Bayes classifier, forecasts, machine learning.

Введение. В условиях современных изменений климата возрастает актуальность систематических исследований ледового режима рек [1, 2]. Обеспечение безопасного судоходства, защита гидротехнических сооружений, минимизация экономического ущерба от опасных явлений – все перечисленные задачи приоритетны в области изучения ледовой обстановки на водных объектах [3, 4]. К наиболее важным видам ледовых прогнозов относятся определение сроков начала и конца ледообразования, что является комплексным процессом, но существенно отличающимся по факторам образования (формирования). Главными параметрами начала активного ледостава принято считать уменьшение температуры воздуха и воды, скорость ветра, морфометрические особенности русла [5]. При рассмотрении процесса разрушения

* E-mail: nikitamakarov11@mail.ru

** E-mail: oderiut@mail.ru

*** E-mail: vmargaryan@ysu.am

покрова (начала весеннего снеготаяния) выделяют усиление солнечной радиации и атмосферного теплообмена, что чаще всего косвенно представлено температурой воздуха, и к этому добавляют еще резкий подъем уровня воды и увеличение скорости течения [6–8].

Целью исследования является внедрение методов машинного обучения – методы деревьев решений (ДР) и наивного байесовского классификатора (НБК) в практику моделирования ледовых процессов на реках бассейна оз. Севан. Предложенные алгоритмы основаны на вероятностном попадании отдельных признаков в определенный класс (группу).

Объект исследований. Для исследования были выбраны пять гидрологических пунктов в центральной части бассейна оз. Севан, где наблюдаются ледовые явления: р. Дзыгнагет (с. Цовагюх), р. Драхтик (с. Драхтик), р. Памбак (с. Памбак), р. Варденис (с. Варденис), р. Бахтак (п. Цаккар). На большинстве объектов в течение зимнего периода наблюдаются забереги и кратковременный ледостав. Преобладающая форма речной долины – V-образная и корытообразная. Берега крутые и обрывистые, где максимальная высота может достигать до 200 м. Площади водосборов, относящиеся к указанным гидрологическим постам, находятся в пределах от 22,5 до 144 км², средние высоты – от 2220 до 2680 м. Средний уклон реки от удаленной точки составляет от 34‰ до 104‰. Залесенность водосборов отсутствует, за исключением 15% от площади для р. Памбак. Распаханность почв составляет от 0 до 35%.

Методы исследования. В настоящий момент на смену классическим подходам определения сроков начала и конца ледообразования в гидрологическую практику моделирования все больше и больше внедряются алгоритмы на основе машинного обучения, а разнообразие представленных методов варьируется от традиционной регрессии до искусственного интеллекта [9].

Деревья решений (ДР). Деревья решений – метод статистического анализа (регрессии или классификации) связи между зависимыми и независимыми переменными. В отличие от регрессионных модели, которые представлены в виде уравнения, подобный алгоритм представлен в виде древовидной структуры с иерархическим разделением обучающей выборки. Проще представить, что существует некоторое облако точек между двумя параметрами, производится отсечение при фиксированном значении предиктора для создания максимальных отличительных классов (узлов), а уже от получившихся групп продолжается разделение дальше до минимального количества членов в конечной выборке (листе). Таким образом, метод по своим свойствам является “жадным” и склонен к переобучению (полное повторение обучающей выборки без введения стохастической составляющей). Чтобы избежать подобного, вводятся статистические оценки качества разбиения – подсчет вероятности попадания в отдельный сегмент. К наиболее популярным критериям относится индекс *Gini*, который определяется по следующей формуле:

$$Gini(T) = 1 - \sum_{i=1}^n p_i^2,$$

где T – текущий узел; p_i – вероятность группы i в узле T ; n – количество групп [10, 11].

Наивный байесовский классификатор (НБК). Алгоритм основывается также на вероятностном подходе представления совместного пространства переменных, однако, с “наивным” предположением об их независимости, что может быть представлено формулой Байеса (переоценка вероятности события A с учетом факта, что наступило другое событие B):

$$P(A_k|B_1, B_2, \dots, B_n) = \frac{P(A_k) \prod_{i=1}^n P(B_i|A_k)}{P(B_1, B_2, \dots, B_n)},$$

где $P(A_k|B_1, B_2, \dots, B_n)$ – апостериорная вероятность события A при условии выполнения событий B ; $P(B_i|A_k)$ – условная вероятность событий B при условии выполнения события A , $P(A_k)$ и $P(B_1, B_2, \dots, B_n)$ – априорные вероятности событий A и B соответственно. В математическом смысле задача НБК сводится к установлению метки принадлежности объекта к определенной группе, описывающей его образ в виде многомерного вероятностного распределения с нахождением максимума произведения частных (одномерных) плотностей ($A_k \propto \arg \max P(A_k) \prod_{i=1}^n P(B_i|A_k)$) [12].

Результаты исследования и их обсуждение.

Первичный анализ. В работе были использованы данные гидрометеорологических наблюдений с 2015 по 2023 гг. Также для проверки качества методик исходные выборки были разделены на обучающую (2015–2022 гг.) и тестовую (2023 г.) части. Временной ряд 2015–2023 гг. сроков установления и разрушения ледового покрова на реках бассейна оз. Севан представлен на рис. 1. На графике видно, что для рек Варденис и Драхтик в большинстве лет дата окончания ледовых явлений наступает позже, чаще всего в марте. А для рек Бахтак и Дзыкнагет, наоборот, свойственно более раннее наступление ледообразования, в основном в третьей декаде декабря. Средняя продолжительность ледовых явлений составляет от 57 до 86 суток. Минимальная продолжительность наблюдается на реке Памбак, максимальная – Варденис. Отсутствие ледостава можно выделить на объектах Памбак, Бахтак (исключение 2017 г. – 87 суток) и Варденис, а средняя продолжительность для остальных пунктов составляет 50 суток.



Рис. 1. Сроки появления и окончания ледового покрова.

Результаты прогноза по ДР. На рис. 2 представлены результаты построения дерева решений для прогноза ледовых явлений на реке Варденис по обучающей выборке 2015–2022 гг. На первом (главном) ветвлении модель формирует две практически равные выборки ($N_1 = 62$, $N_2 = 57$). Температура воздуха, требуемая для разбиения, составляет больше (льда нет, возможность попадания в класс 55%) или меньше (лед есть, попадание в класс 45%) $-1,17^\circ\text{C}$. Это можно назвать точкой бифуркации в классификации для данного объекта, так левая часть полностью определена “наличием льда” при фиксированном значении температуры. Далее происходит уточнение по правой ветви, где также задается температура, но уже в двух случаях отмечается попадание в класс “отсутствие льда” больше (в 30% случаев из 20 значений) и меньше (в 70% случаев из 45 значений) значения $-2,28^\circ\text{C}$. Подобный вывод ветви говорит об разной значимости предиктора в ограниченном диапазоне исследуемых характеристик. Так для 33 значений обучающей выборки происходит за счет уровня – от 53,7 до 57,1 см прогнозируется “отсутствие льда” (18 значений), а больше или меньше заданного диапазона – “наличие льда” (15 значений), а оставшиеся 12 членов ряда – по температуре воздуха.

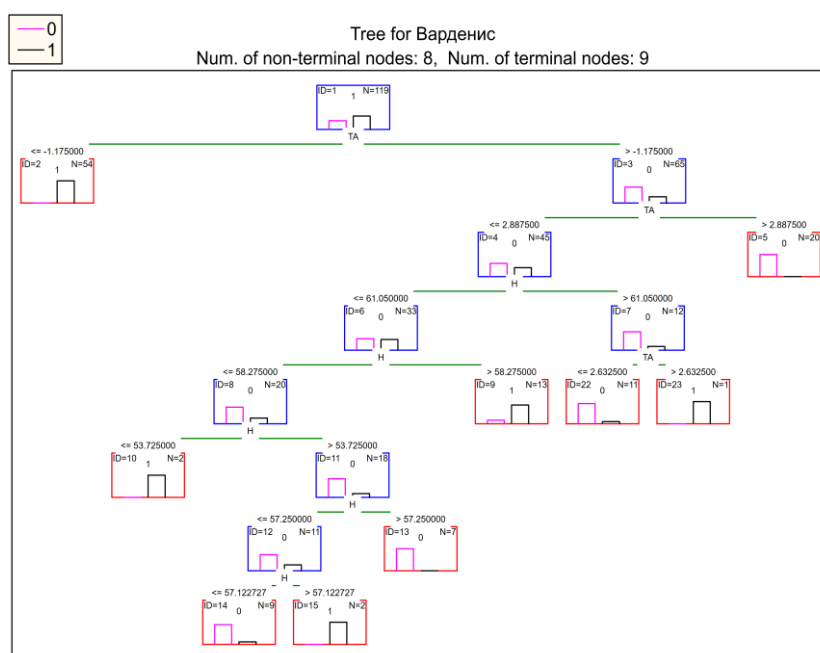


Рис. 2. Визуализация модели дерева решений классификации на примере р. Варденис.

Результаты прогноза по НБК. На рис. 3 представлены результаты байесовской классификации (апостериорная вероятность по уровням воды и температуре воздуха) также на примере р. Варденис по обучающей выборке 2015–2022 гг. Графики содержат кривые, где синяя линия отвечает за класс “отсутствие льда”, а красная – “наличие льда”. По оси абсцисс – значение предиктора (температура или уровень воды), ось ординат отвечает за отображение вероятности. По моделированию ледовых явлений методом НБК

наибольшая апостериорная вероятность попадания в класс “присутствие ледовых явлений” наблюдается при уменьшении температуры воздуха от -2°C до -12°C . При такой температуре воздуха вероятность попадания от 80 до 100%. Для класса “отсутствие ледовых явлений”, наоборот, кривая стремиться к своему максимуму при условии увеличения температуры от -3°C до 8°C (графики являются зеркальными относительно друг друга). Влияние уровня воды достаточно неоднозначно, так как можно отметить увеличение шансов попасть в класс “наличие льда” при росте уровня воды. Подобное явление происходит и на р. Драхтик, для других водных объектов – наоборот. Опорной точкой в классификации является уровень воды от 65 см.

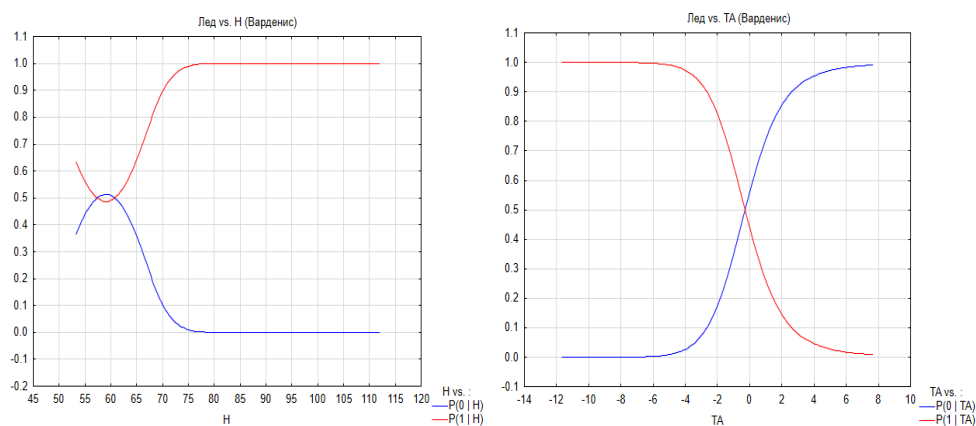


Рис. 3. Апостериорная вероятность по уровням воды и температуре воздуха на примере р. Варденис.

Результаты исследований и их обсуждение. Для всех исследуемых пунктов наблюдений также были построены модели ДР и НБК по обучающим выборкам 2015–2022 гг. и сделан прогноз ледовых явлений на 2023 г., который представлен в таблице.

Прогноз ледовых явлений по ДР классификации и НБК

Река	Начало/конец ледовых явлений в 2023 г.									
	начало				конец				весь период 2015– 2022 гг.	
	факт. знач.	прогноз ДР	прогноз НБК	Δ ДР/НБК	факт. знач.	прогноз ДР	прогноз НБК	Δ ДР/НБК	Р % ДР	Р% НБК
Дзыкнагет	январь 1 д	декабрь 3 д	декабрь 3 д	"10/10"	март 1 д	март 2 д	февраль 3 д	"10/10"	78	83
Драхтик	январь 1 д	декабрь 3 д	декабрь 3 д	"10/10"	март 1 д	март 2 д	февраль 3 д	"10/10"	89	89
Памбак	декабрь 3 д	декабрь 3 д	декабрь 3 д	"0/0"	февраль 2 д	февраль 3 д	февраль 3 д	"10/10"	83	89
Варденис	декабрь 3 д	декабрь 3 д	декабрь 3 д	"0/0"	март 1 д	февраль 3 д	февраль 3 д	"10/10"	94	94
Бахтак	декабрь 3 л	декабрь 3 л	декабрь 3 л	"0/0"	март 1 л	март 1 л	март 1 л	"0/0"	100	100

На двух гидропостах (Дзыкнагет и Драхтик) прогнозная дата начала ледообразования наступает раньше фактического значения, что в целом может подтверждаться наличием кратковременного ледостава. Кроме того, в четырех из пяти пунктов наблюдений алгоритмы хуже справляются с определением сроков конца ледовых явлений: ДР чаще всего прогнозирует позже, а НБК раньше факта. Если сравнивать относительное число оправдавшихся прогнозов за весь период тестовой выборки, то метод НБК обладает небольшим преимуществом в точности на таких водных объектах, как Дзыкнагет (78%/83%) и Бамбак (83%/89%), на остальных объектах результаты идентичны (от 89% до 100%).

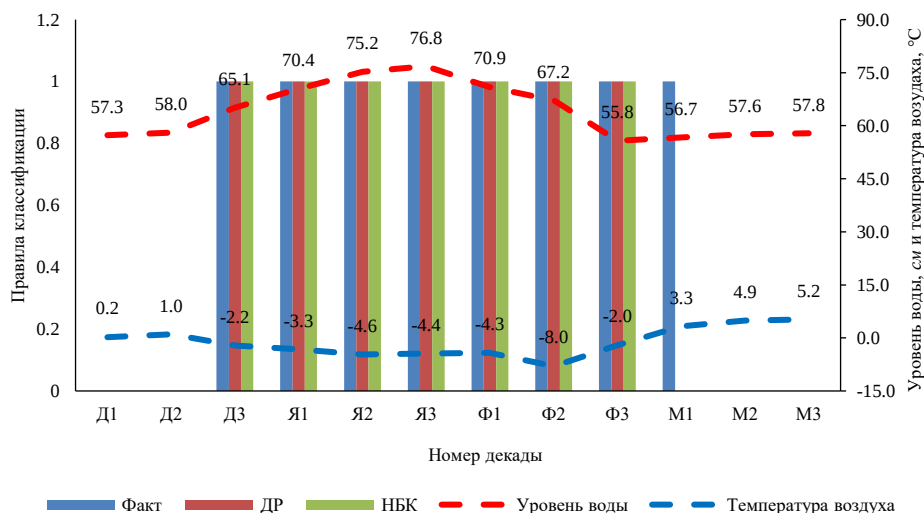


Рис. 4. Визуализация прогноза по моделям ДР и НБК для р. Варденис.

Теперь детально разберем почему прогнозная модель плохо справилась с определением ледовых явлений на реке Варденис. На рис. 4 изображено моделирование ледовой обстановки для 2023 г., дополнительно на график нанесены значения температуры воздуха и уровня воды. Так как нас интересуют ошибки моделей, то сконцентрируемся только на определении срока разрушения льда. Прогноз по обоим алгоритмам вышел на декаду раньше фактического. Гидрометеорологические условия для первой декады марта составляют: температура воздуха – 3,3°C; уровень воды – 56,7 см. По ДР (рис. 2) результат классификации остановился на листе с температурой воздуха больше 2,88°C (3,3°C > 2,88°C, прогноз – “отсутствие льда”), другими словами, положительные температуры наступили раньше, чем разрушился лед. Обратимся к графикам НБК (рис. 3), сформированные метеоусловия также говорят об раннем разрушении ледовых условий; при данной температуре воздуха вероятность отсутствия льда составит более 90%. Таким образом, предложенные алгоритмы чаще всего ошибаются из-за временных сдвижек метеопараметров.

Заключение. Применение алгоритмов машинного обучения имеет ряд преимуществ относительно традиционных подходов, что выражается прежде всего в возможности работать с ограниченным набором гидрометеорологических данных, отсутствием выраженной линейной связи и применяется относительно простое построение. Представленные методы имеют свои плюсы и минусы, например, деревья решений классификации обеспечивают более наглядное представление процесса принятия решений, что может быть полезно для анализа влияния отдельных факторов на систему в целом, а наивный байесовский классификатор своей способностью лучше адаптироваться к изменчивости временных параметров. Из недостатков – обе модели имеют сильную чувствительность к входным параметрам. Например, температура воздуха может переходить в положительные значения раньше, чем лед начнет вскрываться. Однако точно можно сделать вывод, что данные методы являются достаточно перспективными для использования в прогнозах ледовых условий.

Исследование выполнено при финансовой поддержке Комитета по науке Республики Армения в рамках научного проекта 24WS-1E035.

Поступила 09.06.2026

Получена с рецензии 20.06.2026

Утверждена 26.08.2026

ЛИТЕРАТУРА

1. Маргарян В.Г., Гайдукова Е.В. и др. Климат и ледовый режим бассейна реки Драхтик (бассейн оз. Севан, Армения). *Тр. XIV Международ. научно-практ. Конф. “Морские исследования и образование (MARESEDU-2025)”*, I. Тверь, ООО “ПолиПРЕСС” (2025), 66–72.
2. Маргарян В.Г. Ледовый режим рек бассейна Дебед, Армения. *Лед и Снег* **61** (2021), 248–261. <https://doi.org/10.31857/S2076673421020086>
3. Margaryan V.G., Gaidukova E.V., Avetisyan G.D. Trends in the Minimum Winter River Flow on the Territory of the Basin of Lake Sevan. *Arid Ecosyst* **14** (2024), 269–279. <https://doi.org/10.1134/S2079096124700215>
4. Гайдукова Е.В., Маргарян В.Г., Седракян А.М. Долгосрочная оценка режима минимального зимнего стока рек бассейна оз. Севан при изменении климата. *Устойчивое развитие горных территорий* **16** (2024), 259–272. <https://doi.org/10.21177/1998-4502-2024-16-1-259-272>
5. Чижев А.Н. *Формирование ледяного покрова и пространственное распределение его толщины: Монография*. Ленинград, Гидрометеиздат (1990), 126.
6. Шуляковский Л.Г. *Появление льда и начало ледостава на реках, озерах и водохранилищах. Расчеты для целей прогнозов*. Ленинград, Гидрометеиздат (1960), 216.
7. Margaryan V.G., Frolova N.L., Gaidukova E.V. Minimum Winter Ten-Day Runoff in Rivers of the Lake Sevan Basin. *Water Resources* **52** (2025), 1108–1119. <https://doi.org/10.1134/S0097807825700228>
8. Маргарян В.Г., Гайдукова Е.В. К вопросу оценки зимнего минимального среднемесячного стока рек бассейна оз. Севан. *Изв. Тульского гос. ун-та. Науки о Земле* **2** (2023), 26–36.
9. Макаров Н.А., Гайдукова Е.В. *Метод долгосрочного прогнозирования среднемесячных расходов воды с использованием гибридной модели на основе копулы-функции и деревьев решений. Шестые Виноградовские чтения*. Гидрология нового поколения. Санкт-Петербург, ООО Издательство ВВМ (2026), 509–513.
10. Груздев А. *Прогнозное моделирование в IBM SPSS Statistics, R и Python. Метод деревьев решений и случайный лес*. ДМК Пресс (2017).
11. Гордеева С.М., Малинин В.Н. Использование *Data Mining* в задаче гидрометеорологического прогнозирования. *Ученые записки РГТУ* **44** (2016), 30–44.

12. Гефан Г.Д., Рудоминский В.А., Щелоков И.В. Байесовский метод и его использование в задачах классификации данных. *Информационные технологии и математическое моделирование в управлении сложными системами* 2 (2025), 26.

Ն. Ա. ՄԱԿԱՐՈՎ, Ե. Վ. ԳԱՅԴՈՒԿՈՎԱ, Վ. Գ. ՄԱՐԳԱՐՅԱՆ

ՄԵՎԱՆԱ ԼՃԻ ԱՎԱՉԱՆԻ ԳԵՏԵՐԻ ՍԱՌՑԱՅԻՆ ԵՐԵՎՈՒՅԹՆԵՐԻ ԱՄՍԱԹՎԵՐԻ ՄՈԴԵԼԱՎՈՐՄԱՆ ՀԱՄԱՐ ՄԵԶԵՆԱՅԱԿԱՆ ՈՒՍՈՒՅՄԱՆ ՄԵԹՈԴՆԵՐԻ ԿԻՐԱՌՈՒՄԸ

Ա մ փ ո փ ու մ

Ուսումնասիրությունը սառույցի կանխատեսման փոփոխման խնդրի համար առաջարկում է լուծում՝ ընտրելով երկու տարածված մոտեցում՝ որոշման ծառեր և միամիտ բայեսյան դասակարգիչ: Ուսումնասիրության համար ընտրվել են Սևանի ավազանի հինգ հիդրոլոգիական դիտակետեր, որտեղ դիտվում են սառցային երևույթներ՝ Ձկնագետ, Դրախտիկ, Փամբակ, Վարդենիս և Բախտակ գետերը: Սառույցի առաջացման և քայքայման գործընթացները սխեմատիկացնելու և համադրելու համար, որպես կանխատեսման բնութագրիչներ օգտագործվում են կեղծ փոփոխականներ՝ սառցե իրադարձությունների սկզբի և ավարտի ամսաթվերի փոխարեն, որտեղ օդի միջին տասնօրյա ջերմաստիճանը և ջրի մակարդակը ծառայում են որպես անկախ պարամետրեր: Բավարար արդյունքների համաձայն՝ առաջարկվող մոդելները պակաս լավ են աշխատում սառցե իրադարձությունների ավարտի ամսաթիվը որոշելու հարցում, բայց կարող են մշակել փոքր նմուշներ և ֆունկցիոնալ կապերի բացակայություն:

N. A. MAKAROV, E. V. GAIDUKOVA, V. G. MARGARYAN

USING MACHINE LEARNING METHODS TO MODEL ICE DATES ON RIVERS ON THE LAKE SEVAN BASIN

Summary

The study proposes a solution to the problem of modifying ice forecasting, choosing two popular approaches: decision trees and a Naive Bayes classifier. The study focuses on rivers in the Lake Sevan basin, specifically the Dzknaget, Drakhtik, Pambak, Vardenis, and Bakhtak, which experience complex ice breakup and freezing processes. To schematize and combine the ice formation and breakup processes, dummy variables are used as forecast characteristics instead of the onset and end dates of ice events, with average ten-day air temperatures and water levels serving as independent parameters. According to satisfactory results, the proposed models perform less well in determining the end date of ice events; however, they are capable of handling small samples and the absence of functional relationships.