

МЕТОДИКА ОЦЕНКИ И ВЫБОРА ОПТИМАЛЬНОЙ МОДЕЛИ, ПРОГНОЗИРУЮЩЕЙ ОКОНЧАНИЕ ПЕРИОДА ЗАМОРОЗКОВ

В. А. Солозобов^а, С. А. Лысенкова^б

Сургутский государственный университет, г. Сургут, Российская Федерация

^а ✉ solo.val.al@yandex.ru

^б ORCID: <http://orcid.org/0000-0003-1007-7610>, lsa1108@mail.ru

Аннотация: одной из первых задач при разработке математической модели прогнозирования является определение критерия, по которому будет оцениваться и выбираться оптимальная модель из ряда возможных. При этом решаемые задачи моделирования могут требовать различных методов оценки в зависимости от поставленной цели. Именно от того, каким образом будет оцениваться модель прогнозирования, зависит конечный итог. В данном исследовании была разработана методика оценки и выбора модели, прогнозирующей окончание периода заморозков в весенне-летний период.

Ключевые слова: методы оценки, метрика качества, нейронная сеть, прогноз погоды, заморозок.

Для цитирования: Солозобов В. А., Лысенкова С. А. Методика оценки и выбора оптимальной модели, прогнозирующей окончание периода заморозков. *Успехи кибернетики*. 2025;6(2):100–107.

Поступила в редакцию: 19.05.2025.

В окончательном варианте: 04.06.2025.

ASSESSMENT AND SELECTION OF AN OPTIMAL MODEL FOR FORECASTING THE END OF THE FROST PERIOD

V. A. Solozobov^а, S. A. Lysenkova^б

Surgut State University, Surgut, Russian Federation

^а ✉ solo.val.al@yandex.ru

^б ORCID: <http://orcid.org/0000-0003-1007-7610>, lsa1108@mail.ru

Abstract: one of the first steps in developing a mathematical forecasting model is to define the criteria for evaluating and selecting the optimal model from among several candidates. Since modelling objectives can vary, the choice of evaluation method must align with the specific goal of the study. The accuracy and usefulness of the final result depend heavily on how the model is assessed. In this study, we developed a procedure for evaluating and selecting a model to forecast the end of the frost period in the spring–summer season.

Keywords: assessment methods, quality metric, neural network, weather forecast, frost.

Cite this article: Solozobov V. A., Lysenkova S. A. Assessment and Selection of an Optimal Model for Forecasting the End of the Frost Period. *Russian Journal of Cybernetics*. 2025;6(2):100–107.

Original article submitted: 19.05.2025.

Revision submitted: 04.06.2025.

Введение

Прогноз завершения периода отрицательных температур и наступления устойчивой положительной температуры в весенне-летний период является важным и определяющим для большого количества видов деятельности человека, таких как сельскохозяйственный, строительный, транспортный и другие. Для расчета такого прогноза требуется создание модели прогнозирования окончания периода заморозков (ОПЗ). Один из вариантов такой модели — строить прогноз на каждый день и выдавать вероятность завершения периода заморозков на этот день. При этом высокая вероятность завершения заморозков будет говорить о том, что с этого дня будет устойчивая положительная температура воздуха. Данная информация полезна для уменьшения рисков и увеличения эффективности деятельности человека.

Одними из первых задач, стоящих перед разработкой такой модели, являются выбор и определение критерия оценки качества, по которому будет определяться оптимальность этой модели [1]. Однако выбор оценочных критериев, или, по-другому, метрик качества, не является простой задачей. Нужно учитывать, к какому типу относится решаемая проблема, свойство входных данных, желаемое соотношение ложноположительной ошибки к ложноотрицательной ошибке, подходы к обучению и настройке модели, тип модели и другое [2, 3]. С помощью правильно разработанной методики оценки качества можно будет сравнивать различные модели.

Таким образом, цель данного исследования — определить методику оценки моделей, по которой будет происходить их сравнение и выбор оптимальной модели прогнозирования ОПЗ. Для достижения поставленной цели нужно решить следующие задачи:

- 1) рассмотреть метрики качества, используемые в подобного вида задачах;
- 2) определить основные свойства применяемого класса моделей и характеристики входных и выходных данных;
- 3) выбрать или разработать свою методику оценки и выбора оптимальной модели прогнозирования ОПЗ.

Методы оценки моделей, прогнозирующих явления

Разрабатываемая модель решает задачу классификации. Она должна определить на день прогноза, наступило ли ОПЗ в данный день, на основе предшествующей метеобстановки. Данные метеобстановки представляют собой временные ряды со значениями метеопараметров за последние несколько недель или месяцев. То есть модель классифицирует данные временные ряды на предмет того, относятся ли они к периоду заморозков или в них появились признаки наступления устойчивой положительной погоды.

Основные методы оценки качества при решении задачи классификации основаны на матрице ошибок (таблица 1). В нашем случае таких классов два: отрицательный (Negative) и положительный (Positive).

Таблица 1

Матрица ошибок

Классификация (прогнозная)	Классификация (действительная)	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

К позитивному классу относятся случаи продолжения заморозков, а случай окончания периода заморозков относится к негативному классу соответственно. Таким образом, FN — это случай, когда модель прогнозирует завершение периода заморозков, но на самом деле это не так. Данный случай несет наибольший ущерб, и его следует минимизировать в первую очередь.

Основные метрики, по которым делается оценка качества бинарной классификации, рассчитываются по значениям в матрице ошибок [1]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

Коэффициент корреляции Мэтьюса:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

$$F_\beta = \left(1 + \beta^2\right) \frac{Precision \cdot Recall}{(\beta^2 Precision) + Recall}, \quad (5)$$

где F -мера при $\beta = 1$ — среднегармоническая мера.

Также используют сравнение площадей под ROC-кривой (ROC-AUC) (англ. Receiver Operating Characteristics Curve — Area Under the Curve), оценку с помощью построения Precision-Recall (PR) кривой и расчет площади под PR-кривой.

Отдельно стоит упомянуть логистическую функцию ошибки, которая будет минимизироваться в процессе обучения модели. Она рассчитывается с помощью логарифмирования расхождений, рассчитанных вероятностей, принадлежности к классу и фактических значений. Казалось бы, минимальная логистическая ошибка должна указывать на оптимальную модель, но это не так, как выяснится дальше в данном исследовании.

Логистическая функция ошибки:

$$\text{LogLoss} = \frac{-1}{l} \sum_{i=1}^l (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)). \quad (6)$$

Одна из проблем, которая мешает выбрать стандартную метрику качества, заключается в том, что объем выборки будет различаться для моделей с разными входными параметрами. Например, это происходит при изменении скользящего окна — количества дней во временном ряде, которые подаются на вход прогнозирующей модели. Так, при изменении временного окна с 15 дней до 60 обучающая выборка уменьшится на 45 единиц. При этом уменьшение произойдет в положительном классе, который требуется прогнозировать в первую очередь. Изменение объема в выборке не позволит сравнивать различные модели по стандартным метрикам, так как объем класса будет изменяться, и при одинаковом количестве ошибок лучше себя покажет тот, у которого объем выборки больше.

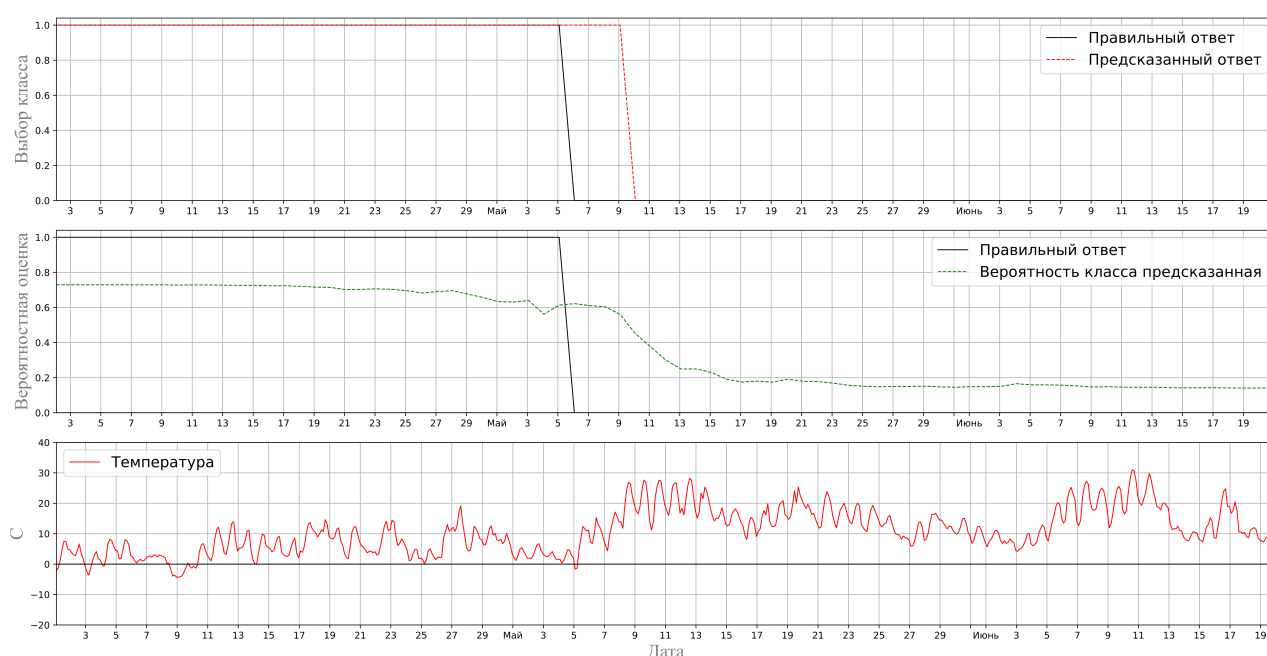


Рис. 1. Пример сравнения прогноза ОПЗ моделью и фактических данных

Еще одна причина того, почему стандартные метрики не подходят, — это неравномерное распределение ошибок по годовой выборке. Из рисунка 1 видно, что наиболее сложно для модели предсказать переходный период весна-лето, представляющий 20–40 дней, в котором она совершает ошибки. Начальная часть годовой выборки и конечная ее часть точно идентифицируются как зима и как лето, и от того, как много будет добавлено легких к прогнозированию дней (приходящихся на зиму или на лето), количество ошибок не изменится. Это еще одна причина, по которой из процесса вычисления метрики следует убрать общее количество экземпляров, по которым будет происходить обучение. Но и просто сравнивать количество ошибок нельзя, так как исчезнет возможность сравнивать модели, обученные на разных объемах выборок, ошибку следует привести к некоторому среднему показателю. Самый простой вариант — это рассчитывать отношение количества ошибок к количеству лет, участвующих в выборках. То есть мы получим, сколько в среднем за один прогнозируемый сезон модель совершит ошибок. Понятно, что некоторые года и их весенне-летний сезон прогнозируется хуже, чем другие, и на них будет приходиться больше ошибок, чем на другие. Но данная проблема будет решаться равномерным распределением сложных и простых случаев и использованием большего количества значений для обучения модели.

Еще одной важной проблемой является то, что ошибки первого и второго рода не равнозначны [4]. Следует в первую очередь снижать ошибку неверного прогнозирования ОПЗ. Но тут возможно возникновение проблемы: получение модели, у которой минимальная ошибка второго рода, но при этом ошибка первого рода настолько велика, что теряется смысл прогнозирования [3]. Требуется поиск сбалансированного решения. Поэтому при обучении модели будет производиться выбор по минимальной общей ошибке. А при сравнении моделей будет происходить отбор по минимальной ложнонегативной ошибке.

Для борьбы с переобучением, которое возможно, есть различные способы, но универсальный — это контроль по тестовой группе. Если по тестовой группе ошибка перестала падать и стала расти, начался процесс переобучения и запоминания выборки. Тут встает вопрос, какую все-таки выбрать модель: опираться на метрики тестовой модели или на средние метрики по обеим группам. В данном исследовании предлагается следующая система определения метрики качества по обучающей и тестовой выборке. Если тестовая ошибка выше обучающей, то она будет использоваться при оценивании модели. Если тестовая ошибка ниже обучающей, то будет находиться средняя ошибка по тестовой и обучающей выборкам, и она уже будет использоваться для оценки модели.

Также существует проблема наглядности и интерпретируемости метрик качества [5]. Хотя точность в 95% и полнота в 99% говорят о качественной модели, существует множество случаев, где таких цифр будет недостаточно в силу дисбаланса выборок как по классам, так и по количеству сложных случаев к общему объему выборки [6]. Также не всегда ясно, что выгоднее: снизить на полпроцента полноту, при этом увеличив на процент точность, или наоборот. Еще сложнее обстоят дела с другими метриками и сравнениями площадей под кривыми. С другой стороны, сравнивая именно количество ошибок, которые делает модель в среднем за сезон, ясно понимаешь возможности данной модели. И при регулировании порогового коэффициента сразу видно, как меняется баланс ложноположительной и ложноотрицательной ошибки, что дает четкое понимание результатов данной регуляции.

Методика оценки и выбора модели, прогнозирующей окончание периода заморозков

На основе вышеизложенного предлагается следующая методика сравнения и выбора оптимальной модели прогнозирования явления, в нашем случае ОПЗ:

1. Данные делятся на две выборки: обучающую и тестовую в пропорциях 2 к 1 соответственно.
2. В процессе обучения полученная матрица ошибок приводится к среднему количеству за один год-сезон.

$$FP_{\text{ср.тест}} = \frac{FP_{\text{тест}}}{Q_{\text{тест}}} \quad (7)$$

$$FP_{\text{ср.обуч}} = \frac{FP_{\text{обуч}}}{Q_{\text{обуч}}} \quad (8)$$

$$FN_{\text{ср.тест}} = \frac{FN_{\text{тест}}}{Q_{\text{тест}}} \quad (9)$$

$$FN_{\text{ср.обуч}} = \frac{FN_{\text{обуч}}}{Q_{\text{обуч}}}, \quad (10)$$

где $FP_{\text{тест}}$, $FP_{\text{обуч}}$ — общее количество ложноположительных результатов для тестовой и обучающей выборки соответственно,

$FN_{\text{тест}}$, $FN_{\text{обуч}}$ — общее количество ложнонегативных результатов для тестовой и обучающей выборки соответственно,

$Q_{\text{тест}}$ — количество лет в тесте,

$Q_{\text{обуч}}$ — количество лет в обучающей выборке.

3. Для расчета средней общей ошибки по обеим группам производится сравнение ошибок по обучающей и тестовой группе.

Общая ошибка модели считается равной ошибке по тестовой группе, если ошибка на тестовой группе больше или равна ошибке на обучающей группе. Считается, что произошло переобучение модели, поэтому оценка качества по обучающей группе не учитывается.

Общая ошибка модели равна средней ошибке по обеим группам (тестовой и общей), если ошибка тестовой группы меньше ошибки на обучающей группе. Это возможно в случае, если в обучающую

группу попали сложно классифицируемые случаи, а в тестовую группу — простые случаи. То есть требуется учет ошибок в обучающей группе.

$$FP_{\text{ср}} = \begin{cases} FP_{\text{тест}}, & FP_{\text{ср.тест}} \geq FP_{\text{ср.обуч}} \\ \frac{FP_{\text{тест}} + FP_{\text{обуч}}}{Q_{\text{тест}} + Q_{\text{обуч}}}, & FP_{\text{ср.тест}} < FP_{\text{ср.обуч}} \end{cases} \quad (11)$$

$$FN_{\text{ср}} = \begin{cases} FN_{\text{тест}}, & FN_{\text{ср.тест}} \geq FN_{\text{ср.обуч}} \\ \frac{FN_{\text{тест}} + FN_{\text{обуч}}}{Q_{\text{тест}} + Q_{\text{обуч}}}, & FN_{\text{ср.тест}} < FN_{\text{ср.обуч}} \end{cases} \quad (12)$$

4-1. В случае сравнения и выбора модели в процессе обучения будет выбираться модель с минимальной общей ошибкой: $FP_{\text{ср}} + FN_{\text{ср}}$. При равных величинах общей ошибки выбирается та, у которой меньше ошибка $FN_{\text{ср}}$. Если они тоже равны, выбирается та модель, у которой наименьшая логистическая функция ошибки.

4-2. В случае сравнения и выбора среди уже обученных моделей выбираться будет с наименьшей $FN_{\text{ср}}$. При равных значениях выбирается с наименьшей $FP_{\text{ср}}$. Если они тоже равны, выбирается та модель, у которой наименьшая логистическая функция ошибки.

Стоит отметить, что подход, описанный в пункте 3, применим и для других оценочных метрик: точность, полнота, F_1 мера, логистическая функция ошибки, расчет площади под ROC и PR. Но только при условии постоянства объема выборок, участвующих в обучении модели.

Параметры модели прогнозирования окончания периода заморозков

Выходным параметром модели является случай ОПЗ. Обучающая выборка размечается «0» и «1», где «0» — это признак завершения периода заморозков с этого дня, а «1» — заморозки еще будут.

Входными данными для модели прогнозирования ОПЗ являются временные ряды с различными метеопараметрами. Для данного исследования были выбраны следующие метеопараметры: температура, давление, относительная влажность, скорость и направление ветра, точка росы, процент облачности. Эти метеопараметры были выбраны как наиболее коррелирующие с определением ОПЗ в ходе обзора исследований на данную тему [7, 8]. В качестве длины временных окон были взяты 30, 45, 60 дней. То есть будут сравниваться модели с разным количеством входных параметров, как по длине временного окна, так и по конкретным параметрам.

Имеющийся архив в 20 лет был поделен на три части: обучающая (12 лет), тестовая (6 лет), валидационная (2 года) [9]. Разделение осуществлялось строго по годам, так как при бесконтрольном перемешивании будет происходить утечка данных из одной группы в другую, что может изменить конечные результаты.

В качестве модели прогнозирования используется искусственная нейронная сеть (ИНС) с двумя скрытыми слоями. Это продолжение исследования [7], в котором решалась аналогичная задача, только с помощью методов линейной регрессии. Поиск нелинейных зависимостей с помощью ИНС должен привести к построению более точной модели прогнозирования. Два скрытых слоя в ИНС выбраны как стартовая точка, с которой будет проводиться исследование. Архитектура ИНС и параметры ее обучения здесь не являются определяющими, так как требуется, в первую очередь, разработать и оценить методику, по которой будут сравниваться получаемые модели. Уже после выбора оценочных критериев будет производиться поиск оптимальной архитектуры ИНС. Для оценки метрик качества вводится валидационная часть выборки.

Результаты

Процесс проверки методики оценки качества моделей был разделен на два этапа.

На первом этапе оценивалось применение методики в процессе обучения модели. Модель обучалась на входных данных: температура, давление, влажность воздуха, направление ветра и значение точки росы за последние 60 дней. На этом этапе вполне можно пользоваться стандартными метриками качества для выбора оптимальной модели в процессе обучения. Но, как показывают эксперименты, разработанная методика уверенно выбирает оптимальную модель из возможных (табл. 2). В таблице выделены оптимальные значения для конкретной метрики и можно оценить, какие значения при

Таблица 2

Выборка моделей с оптимальными значениями метрик качества

Метрики качества	Варианты моделей прогнозирования ОПЗ					
	1	2	3	4	5	6
FN(L+T)	3.833	0.000	0.000	0.000	0.000	0.000
FP(L+T)	2.833	7.167	7.500	8.333	8.333	8.667
FP+FN(L+T)	6.667	7.167	7.500	8.333	8.333	8.667
FN(L+T+V)	3.222	0.333	0.444	0.000	0.000	0.000
FP(L+T+V)	2.556	8.111	8.333	9.333	9.333	9.556
FP+FN(L+T+V)	5.778	8.444	8.778	9.333	9.333	9.556
AUC_ROC(L+T)	0.992	0.991	0.991	0.997	0.997	0.997
AUC_ROC(L+T+V)	0.994	0.992	0.992	0.997	0.997	0.997
AUC_PR(L+T)	0.984	0.980	0.979	0.994	0.994	0.994
AUC_PR(L+T+V)	0.988	0.982	0.982	0.994	0.994	0.994
log_loss(L+T)	0.365	0.132	0.132	0.175	0.176	0.194
log_loss(L+T+V)	0.369	0.137	0.137	0.182	0.183	0.201
f_one(L+T)	0.932	0.932	0.929	0.922	0.922	0.919
f_one(L+T+V)	0.942	0.922	0.919	0.915	0.915	0.913
mcc(L+T)	0.900	0.901	0.897	0.887	0.887	0.883
mcc(L+T+V)	0.914	0.885	0.880	0.876	0.876	0.873

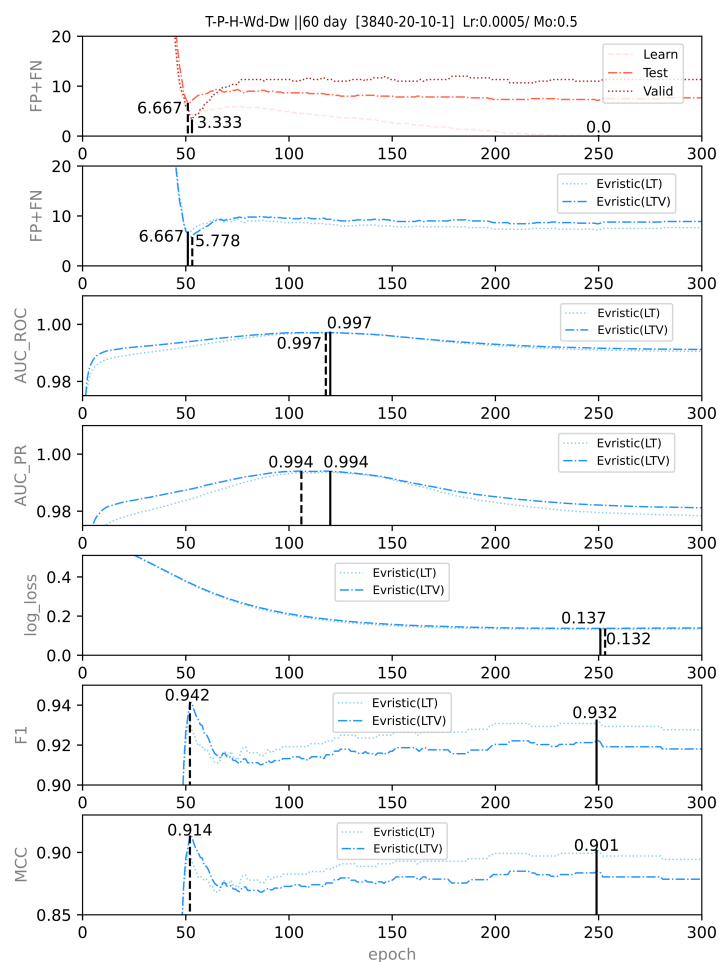


Рис. 2. Пример сравнения изменения значения метрик в процессе обучения модели

этом принимают другие метрики. На рисунке 2 представлены графики изменения значений различных метрик качества от эпох обучения. Важно отметить, что метрики MCC , F_1 , $FP_{cp} + FN_{cp}$ обладают схожим поведением и их оптимальные значения приходятся на модель со схожими коэффициентами и характеристиками. При этом оптимальные значения AUC_ROC и AUC_PR приходятся на модель с более высоким средним количеством ошибок, но нулевой ложноотрицательной ошибкой (табл. 2). То есть данные метрики также пригодны, если требуется снижение FN ошибки на стадии обучения. Однако причиной занижения FN является дисбаланс выборки в сторону негативного класса. Поэтому, если требуется снижение ложноотрицательной ошибки на стадии обучения, правильное будет использовать разработанную методику, минимизируя FN . Опять же это стоит делать внимательно, так как минимизация ошибки второго рода на этом этапе может сильно ухудшить общую ошибку.

На втором этапе производилось сравнение различных моделей с помощью разработанной методики (рис. 3). Были обучены модели с различным диапазоном временного окна: 15, 30, 45, 60 дней и различным набором входных данных: Т — температура, Р — давление, Н — относительная влажность воздуха, Ws — скорость ветра, Wd — направление ветра, Dw — точка росы, Cp — процент облачности, DTU — коэффициент дня по году. Значение Evristic(LT) — это оценка модели по разработанной методике на основе тестовой и обучающей выборки. Значение Evristic(LTV) — это оценка с учетом валидационной выборки. Из графика видно (рис. 3), что если мы обучим модель ИНС на основе данных температуры, давления, влажности воздуха и скорости ветра за 15 дней, то получим модель со средней общей ошибкой 9 случаев за один год. Но если использовать данные за 30 дней и вместо скорости ветра использовать направление ветра, то модель получится со средней ошибкой за один год, равной 5,2.

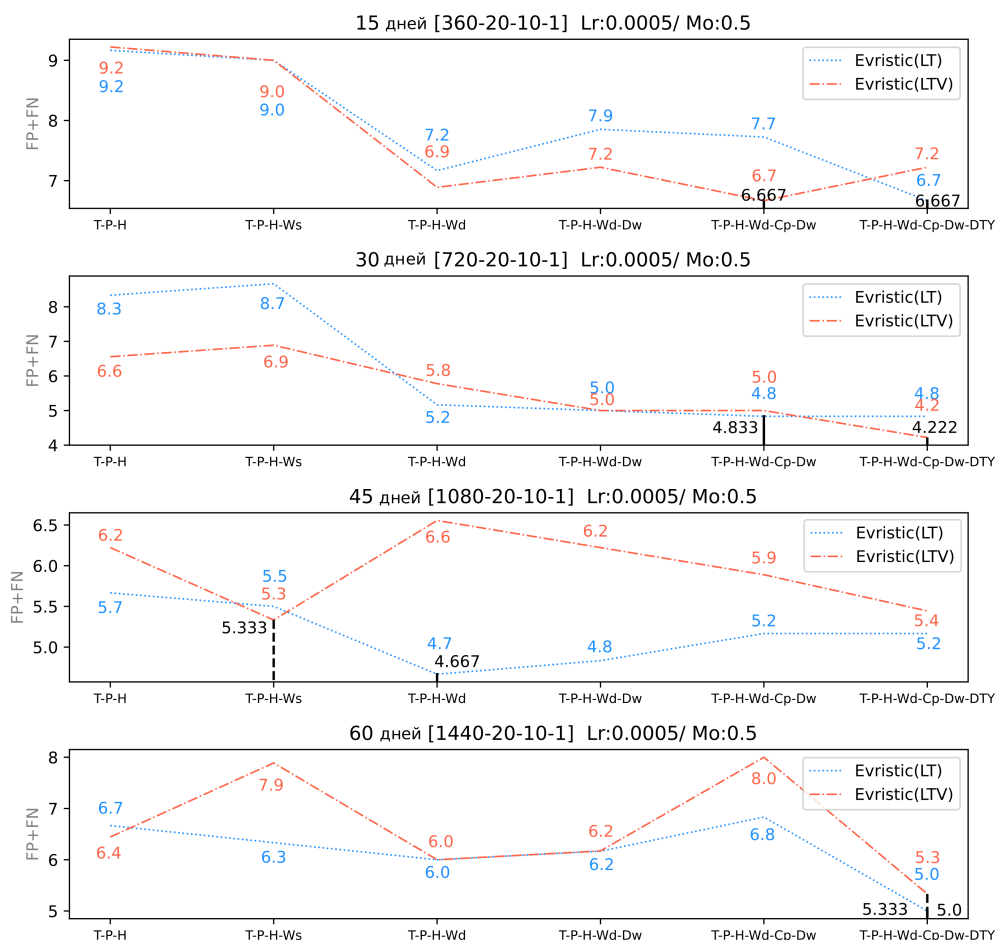


Рис. 3. Сравнение моделей, полученных с помощью разного набора входных данных

Отметим, что при 15- и 30-дневных временных окнах увеличение числа параметров приводит к снижению ошибки, что логично. Но при 45 и 60 днях это уже не наблюдается. Также можно

отметить, что в среднем ошибка увеличивается по сравнению с 30-дневным временным окном, хотя должно происходить иначе. Это можно объяснить тем, что увеличение временного окна приводит к быстрому переобучению, поэтому требуется изменение архитектуры ИНС и параметров обучения. В дальнейших исследованиях можно провести поиск оптимальной архитектуры и гиперпараметров, при которых 60-дневное временное окно будет увеличивать точность, а не снижать ее. Также стоит отметить схожесть поведения значений *Evristic(LT)* и *Evristic(LTV)*, что говорит о сравнимом результате методики, основанной на тестовой и обучающей выборках и основанной на валидационной выборке.

Заключение

Полученная методика оценки и выбора оптимальной модели окончания периода заморозков позволяет сравнивать модели на основе ИНС с различной архитектурой и различным набором входных данных. Стоит отметить, что данный способ оценки может быть использован и для сравнения моделей, полученных различными методами машинного обучения.

Отметим следующие важные моменты разработанной методики:

- наглядность и интерпретируемость метрик;
- возможность сравнения различных видов моделей;
- возможность сравнения влияния входных параметров на прогнозируемое явление;
- возможность сравнения влияния параметров и методов оптимизации обучения на качество получаемых моделей;
- учитывается возможность переобучения модели.

Разработанная методика оценки качества в дальнейшем будет использоваться для поиска оптимального набора входных данных и получения модели с высокой прогнозирующей способностью.

ЛИТЕРАТУРА

1. Hossin M., Sulaiman M. N. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 2015;5(2):1–11. DOI: 10.5121/ijdkp.2015.5201.
2. Huang C. et al. Performance Metrics for the Comparative Analysis of Clinical Risk Prediction Models Employing Machine Learning. *Circulation: Cardiovascular Quality and Outcomes*. 2021;14(10):1076–1086. DOI: 10.1161/CIRCOUTCOMES.120.007526.
3. Vujović Ž. et al. Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*. 2021;12(6):599–606.
4. Архипов В. А. Сравнительный анализ метрик качества для моделей бинарной классификации на примере кредитного скоринга. *Вестник Алтайской академии экономики и права*. 2019;9(2):12–15. EDN: CFZQDC.
5. Hicks S. A. et al. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Scientific Reports*. 2022;12(1):5979. DOI: 10.1038/s41598-022-09954-8.
6. Owusu-Adjei M. et al. Imbalanced Class Distribution and Performance Evaluation Metrics: A Systematic Review of Prediction Accuracy for Determining Model Performance in Healthcare Systems. *PLOS Digital Health*. 2023;2(11):E0000290. DOI: 10.1371/journal.pdig.0000290.
7. Солозобов В. А. Поиск оптимальной регрессионной модели, предсказывающей завершение периода возвратных весенних заморозков. *Наука и инновации XXI века : материалы X Всероссийской конференции молодых ученых*. Сургут, 15 декабря 2023 года. В 3 ч. 2024;1:98–102. EDN: BHSRWN.
8. Солозобов В. А. Исследование возможности использования нейронных сетей для предсказания возвратных весенних заморозков, и поиск оптимального набора данных для этого вида прогнозирования. *Наука и инновации XXI века : сборник статей по материалам IX Всероссийской конференции молодых ученых*. В 4 томах. Сургут, 02 ноября 2022 года. 2023;1:150–155. EDN: ZXQFJS.
9. *Архив метеоданных*. Режим доступа: <https://tp5.ru/>.