УДК 311:631.1

ПРОГНОЗИРОВАНИЕ УРОВНЯ ДОХОДОВ КРЕСТЬЯНСКИХ (ФЕРМЕРСКИХ) ХОЗЯЙСТВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Дашиева Баярма Шагдаровна, *старший преподаватель кафедры статистики и кибернетики ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, dashieva.b.sh@rgau-msha.ru*

Аннотация: *в работе проводится прогнозирование уровня доходов крестьянских (фермерских) хозяйств с использованием методов машинного обучения, в результате проведенного исследования разработано веб-приложение, позволяющее получать прогнозные значения доходов КФХ на основе нейронной сети.*

Ключевые слова: *доходы, крестьянское (фермерское) хозяйство, машинное обучение, нейронная сеть, прогноз.*

С целью сохранения сельских территорий и целостности страны необходимо усилить меры по поддержке малого и среднего предпринимательства. В настоящее время требуется проведение подробного анализа больших массивов данных, собираемых Министерством сельского хозяйства России, Росстатом и другими ведомствами. Актуальность исследований в области анализа данных и искусственного интеллекта подтверждается национальной программой «Цифровая экономика Российской Федерации», ведомственным проектом «Цифровое сельское хозяйство». Ведомственный проект «Цифровое сельское хозяйство» предполагает выполнение задач по созданию и внедрению национальной платформы цифрового государственного управления сельским хозяйством «Цифровое сельское хозяйство» (ЦСХ), предполагающей разработку системы сбора, хранения и обработки данных о ресурсах и результатах сельскохозяйственного производства и разработку системы интеллектуального анализа данных и прогнозирования на основе технологий Advanced Analytics, Data Discovery, Data Mining, Machine Learning и искусственного интеллектах [1,2,3.4,5].

Для проведения исследования был выбран датасет, представленный обезличенными данными из формы № 1-КФХ ведомственной отчетности Минсельхоза России «Информация о производственной деятельности глав крестьянских (фермерских) хозяйств». В форме № 1-КФХ отражаются сведения о размере доходов КФХ, расходов КФХ, количестве членов КФХ, включая главу КФХ, о численности постоянных наемных работников КФХ, площади земельных участков и объектов природопользования, площади посевной площади, наличии сельскохозяйственной техники, сумме кредитов и займов, полученных хозяйством и др. Объем выборки составил 1202

крестьянских (фермерских) хозяйства Ставропольского края, специализирующиеся на производстве зерновых и зернобобовых культур, где удельный вес продукции зернопроизводства превышал 50% от общего размера выручки продукции сельского хозяйства. Численность всех работников КФХ по изначальному датасету определена как сумма членов КФХ и постоянных наемных работников. Среднегодовое число тракторов определено как среднеарифметическая из суммы тракторов на начало и на конец года. Таким же образом найдено среднегодовое число комбайнов, и среднегодовая общая площадь земли.

Статистическая обработка первичных данных позволила получить относительные показатели КФХ. В качестве входных переменных (факторов) выбраны признаки, характеризующие ресурсы производства: численность работников КФХ (чел.), наличие тракторов (шт.), комбайнов (шт.), общая площадь земли (га) в расчете на одно хозяйство и эффективность производства - урожайность зерновых (ц/га), так как данные представлены зерновыми хозяйствами, и с помощью данного показателя можно также прогнозировать и доходы КФХ. В качестве выходной переменной (результативной, целевой) – доходы КФХ в расчете на одно хозяйство. В качестве целевой переменной (результативной) выбран показатель «Доходы, тыс. руб.», так как данный показатель отражает результаты деятельности крестьянских (фермерских) хозяйств.

На первоначальном этапе определено количество пропусков в исходном датасете. В итоге выявлено, что нанимают работников 45,6% КФХ, не имеют тракторов 209 КФХ, или 17,4% всех хозяйств и не имеют в наличии комбайнов 500 КФХ, или 41,6% всех хозяйств. Все пропуски заменены на нулевые значения, так как это говорит об отсутствии наемных работников, или же об отсутствии сельскохозяйственной техники в КФХ.

По результатам вывода описательной статистики (рисунок 1) видно, что масштаб изучаемых признаков различен. Диапазон изменений значений признаков велик, так, например, по величине доходов КФХ видно, что минимальные значения доходов составили 50 тыс. руб., а максимальные 396570 тыс. руб., тогда как средняя величина доходов составила 9871,5 тыс. руб., а медианная средняя 4153,0 тыс. руб. Такие различия в величине среднеарифметической и медианной средней говорят об асимметричности ряда распределения КФХ по доходам. Таким же образом можно проанализировать каждый признак в наборе данных.

Таблица 1.

Описательная статистика

	Доходы, тыс.руб.	Работники, чел.	Наличие тракторов, шт.	Наличие комбайнов, шт.	Общая площадь земли, га	Урожайность зерновых, ц/га
count	1202.0	1202.0	1202.0	1202.0	1202.0	1202.0
mean	9871.5	2.9	2.8	1.2	504.9	30.4
std	22013.3	4.7	3.0	1.8	1167.7	12.8
min	50.0	1.0	0.0	0.0	3.0	2.0
25%	1717.5	1.0	1.0	0.0	120.0	21.5
50%	4153.0	1.0	2.0	1.0	259.4	27.7
75%	9716.2	3.0	4.0	2.0	539.8	36.4
max	396570.0	76.0	31.0	22.0	23118.5	61.3

По диаграммам рассеяния выявлено, что направление связи между факторными признаками и результативным прямое, но имеются выбросы. Также графики boxplot показали наличие выбросов. Гистограммы и графики Q-Q по изучаемым признакам показали, что распределения КФХ по всем признакам отличаются от нормального. Нормирование данных было произведено по преобразованию Йео-Джонсона. Выбросы удалены с использованием межквартильного размаха. Масштабирование данных производилось с использованием Z-оценки, где среднее значение равно 0, а дисперсия 1.

Отбор факторов произведен на основе матрицы парных коэффициентов корреляции по данным после удаления выбросов. Так как связь между доходами КФХ и урожайностью зерновых близкая к слабой, то последний признак решено не добавлять в модель регрессии. Два факторных признака: число тракторов и число комбайнов – тесно связаны между собой (r=0,692). Решено удалить число комбайнов, так как сила связи между числом комбайнов и доходами слабее, и хозяйств, имеющих комбайны, встречается в совокупности гораздо реже, нежели хозяйств, не имеющих трактора.

Совокупность наблюдений разбита на две части: 30% наблюдений приходится на тестирование моделей, 70% – на обучение моделей. Поиск гиперпараметров моделей произведен с помощью поиска по сетке с перекрестной проверкой (GridSearchCV), количество блоков выбрано10.

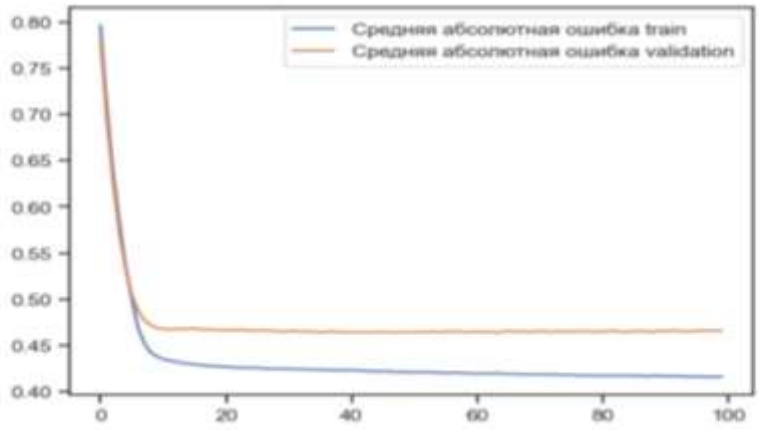
Для прогнозирования доходов КФХ было обучено несколько моделей машинного обучения. Вначале реализована множественная линейная регрессия, построенная с помощью библиотек statsmodels и sklearn. Скорректированный коэффициент детерминации R2=0,706 показал, что 70,6% вариации доходов КФХ объясняется изменением трех факторов, включенных в модель, а остальные 29,4% – это влияние других неучтенных факторов. Модель в целом статистически значима. Все параметры, кроме условного начала оказались статистически значимыми. Далее реализованы ридж и лассо-регрессии, которые показали одинаковый результат по качеству модели с множественной линейной регрессий по-обычному МНК. Оптимальное значение гиперпараметра для альфы по ридж-регрессии – 1, а по лассо-регрессии – 0,001. Далее применен метод K-ближайших соседей. Оптимальное значение 'n_neighbors' – 13. Качество модели получилось несколько хуже по сравнению с множественной линейной регрессией по

МНК. Затем построено дерево решений. Оптимальное значение гиперпараметров: 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 2. Модель Дерева решений по качеству получено самой низкой. Оптимальные гиперпараметры для случайного леса: 'criterion': 'squared_error', 'max_depth': 5, 'max_features': 'auto', 'n_estimators': 500. Качество модели по случайному лесу чуть лучше, чем по дереву решений и по множественной линейной регрессией по МНК. Градиентный бустинг показал хорошие результаты качества модели, как и случайный лес. По тестовой выборке самое лучшее качество показали модели, построенные по нейронным сетям, хотя различия с моделями случайного леса и градиентного бустинга незначительны. Наименьшие ошибки MAE и MSE у моделей нейронных сетей. Далее выбрано построение полносвязной нейронной сети. Построение нейронной сети было проведено с помощью библиотек sklearn (MLPRegressor) и tensorflow (keras. Sequential).

Поскольку исходные данные были ранее нормализованы в процессе предобработки, то на вход нейросеть поданы нормализованные значения от 0 до 1. Поэтому дополнительная нормализация данных не проводилась.

При построении нейронной сети MLPRegressor был произведен поиск по сетке, где было предложено оптимальное число нейронов на каждом слое.

При построении нейронной сети с помощью tensorflow (keras. Sequential) было использовано два полносвязных скрытых слоя Dense с различным количеством нейронов, без дополнительного слоя дропаут и с добавлением дропаут, с функциями активации relu, tanh, sigmoid. В процессе обучения была предпринята попытка минимизировать потери функции, параметры обновлялись для повышения точности (рисунок 1).



Источник: разработано автором

Рисунок 1 – График функции потерь

Качество построенных моделей определялось с помощью следующих метрик: коэффициент детерминации, средняя квадратическая ошибка и средняя абсолютная ошибка. Ошибки каждой модели на тренировочной и тестирующей части выборки показаны в таблице 2.

Таблица 2.

Сравнение моделей машинного обучения

Таким образом, можно сделать вывод, что по ошибкам MAE, MSE и коэффициенту детерминации R^2 нейросеть показала лучший результат, чем любая из моделей регрессии.

С помощью фреймворка Flask было разработано одностраничное пользовательское веб-приложение, прогнозирующее доходы крестьянских (фермерских) хозяйств на основе нейронной сети. Для запуска приложения пользователь должен перейти по ссылке на сайт: http://127.0.0.1:5000/. Flask-приложение представляет собой форму, состоящую из трех входов, куда вводятся значения трех параметров: численность работников, чел., число тракторов, шт., площадь земли, га. Введенные значения должны быть больше или равны 0, в противном случае появится ошибка «ОШИБКА! Введенные значения должны быть больше или равны 0». После этого нужно нажать на кнопку «Submit», и модель выдаст прогнозное значение доходов КФХ при заданных параметрах [6,7,8].

Разработанное веб-приложение могут применять крестьянские (фермерские) хозяйства для прогнозирования их доходов в зависимости от таких существенных факторов, как численность работников, площадь земли и наличие тракторов. А также построенные модели регрессии могут быть использованы при разработке мер аграрной политики для развития малого предпринимательства.

Библиографический список

1. Билл Любанович. Простой Python. Современный стиль программирования. – СПб.: Питер, 2016. – 480 с.

2. Бринк, Х. Машинное обучение / Х. Бринк, Дж. Ричардс, М. Феверолф. – пер. с англ. Рузмайкина И. – Санкт-Петербург: Питер, 2017. – 336 с.

3. Брюс, П. Разведочный анализ данных / П. Брюс, Э. Брюс // Практическая статистика для специалистов Data Science. – СПб.: БХВ-Петербург, 2018. – 304 с.

4. Воронина, В. В. Теория и практика машинного обучения: учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святов. –

Ульяновск: УлГТУ, 2017. – 291 с.

5. Горбань, А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей / А.Н. Горбань // Сиб. журн. вычисл. математики. – 1998. – Т. 1, № 1. – 21 с.

6. Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. Пер. с англ. – СпБ.: ООО «Альфа-книга»: 2018. – 688 с.: ил

7. Плас, Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2023. – 576 с.

8. Khoruzhy, L.I., Katkov, Y.N., Romanova, A.A. Cloud Technologies in the Accounting Information System of Interorganizational Cooperation, Innovation, Technology and Knowledge Managementthis link is disabled, 2023, pp. 25–37 https://www.scopus.com/authid/detail.uri?authorId=57221331639 (Scopus)

9. Состояние социально-трудовой сферы села и предложения по ее регулированию : Ежегодный доклад по результатам мониторинга 2011 г. / Л. В. Бондаренко, А. В. Турьянский, Т. И. Наседкина [и др.] ; Ответственные за подготовку доклада: Д.И. Торопов. Том Выпуск 13. – Москва : Российский научно-исследовательский институт информации и технико-экономических исследований по инженерно-техническому обеспечению агропромышленного комплекса, 2012. – 220 с. – ISBN 978-5-7367-0903-8. – EDN UBCGAZ.

УДК 631.363

НАЛОГООБЛОЖЕНИЕ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ТОВАРОПРОИЗВОДИТЕЛЕЙ

***Огородникова Елена Петровна**, к.э.н., доцент кафедры экономической теории и управления ФГБОУ Оренбургский ГАУ, lena-dozent@mail.ru*
***Лысова Дарья Владимировна**, студент факультета экономики и права ФГБОУ Оренбургский ГАУ, knazevaaa.d@gmail.com*

***Аннотация:** В статье рассматриваются способы налогообложения в сельскохозяйственных предприятиях, произведен сравнительный анализ систем налогообложения, которые могут применять сельскохозяйственные товаропроизводители.*

***Ключевые слова:** Сельскохозяйственные организации, налоговый режим, система налогообложения, единый сельскохозяйственный налог.*

В современных условиях сельскохозяйственные организации имеют возможность планировать собственные налоговые расходы путем выбора системы налогообложения. Система налогообложения для