

## ПРИМЕНЕНИЕ DATA SCIENCE ИНСТРУМЕНТОВ В СЕЛЬСКОМ ХОЗЯЙСТВЕ: ПРОБИТ РЕГРЕССИЯ

*Кашиников Александр Владимирович, студент 2 курса института экономики и управления АПК, ФГБОУ ВО РГАУ–МСХА имени К. А. Тимирязева, e-mail: egkashnikov@gmail.com*

*Научные руководитель – Демичев Вадим Владимирович, к.э.н., доцент, доцент кафедры статистики и кибернетики ФГБОУ ВО РГАУ–МСХА имени К. А. Тимирязева, e-mail: demichev\_v@rgau-msha.ru.*

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

*Аннотация.* В статье показаны методы data science, используемые в сельском хозяйстве, также показан метод прогнозирования дефолта организации на основе финансовых показателей и модели пробит регрессии.

*Ключевые слова:* анализ данных, пробит-модель, логистическая регрессия, предприятие, дефолт, расчет, statsmodels, python.

В сельском хозяйстве существует большое количество сфер, релевантных для применения бинарных моделей прогнозирования. Формулировать гипотезу следует так, чтобы ответом на нее являлось «да» или «нет». Болеет ли растение или животное? Является ли особь половозрелой? Является ли организация банкротом?

Линейная вероятностная модель использует экономические и финансовые данные для оценки вероятности дефолта организации. В этой модели мы запускаем линейную регрессию, в которой объясняемая переменная  $Z$  может иметь значение 1 в случае дефолта или значение 0, когда фирма выплачивает свои долги.

Вид линейной вероятностной модели:

$$Y = \sum_{i=1}^n \beta_i \chi_i + \beta_0, \quad (1)$$

где  $\chi_i$  – объясняющие переменные (финансовые коэффициенты) организации  $i$ ,  $\beta_i$  – коэффициент, который измеряет важность переменной для объяснения дефолта.

После получения коэффициентов из регрессионного прогона можно умножить их на показатели организации, чтобы получить вероятность дефолта. Однако значение  $Y$ , которое измеряет вероятность дефолта организации, может отклоняться от диапазона  $[0, 1]$ , что является основным недостатком линейной модели.

Probit модель исправляет искажение, созданное в линейной вероятностной модели, и ограничивает вероятность дефолта между 0 и 1. Искомая переменная принимает только два значения: значение 1, которое представляет фирму, достигшую дефолта, и значение 0, которое представляет стабильную фирму.

Эта модель использует финансовые показатели для прогнозирования вероятности дефолта организации и предполагает, что эта вероятность имеет кумулятивное нормальное распределение, которое по определению ограничено диапазоном от 0 до 1.

Probit модель отличается от Logit модели в предположении, что вероятность дефолта организации имеет нормальное распределение, а не логистическое распределение [1]. Однако, умножив результаты логистического распределения на соответствующий коэффициент, можно получить распределение пробит-модели.

Перед тем как применять модель, стоит провести очистку данных от некорректных значений и привести данные к единому типу [2]. При этом, количество переменных не должно превышать одну третью часть от всего датасета [3]. Также стоит провести анализ корреляции переменных и убрать из анализа переменные, которые коррелируют между собой.

Для создания модели прогнозирования были получены необходимые показатели 62 организаций. В обучающей выборке половина организаций достигла дефолта  $Y = 1$ .

Список показателей, взятых для создания модели прогнозирования дефолта организации следующий (Таблица 1).

**Таблица 1 – Показатели организации**

Коэффициент рентабельности активов	Коэффициент рентабельности постоянного капитала	Коэффициент долгосрочного привлечения заемных средств	Коэффициент оборачиваемости активов	Коэффициент оборачиваемости денежных средств
Коэффициент рентабельности оборотных активов	Коэффициент рентабельности собственного капитала	Коэффициент реальной стоимости имущества	Коэффициент оборачиваемости собственного капитала	Доля чистого оборотного капитала в активах
Коэффициент рентабельности продаж	Коэффициент обеспеченности собственными средствами	Коэффициент автономии	Коэффициент оборачиваемости оборотных активов	Коэффициент текущей ликвидности
Коэффициент чистой рентабельности	Коэффициент обеспеченности запасов собственными средствами	Коэффициент финансовой напряженности	Коэффициент оборачиваемости запасов и затрат	Коэффициент быстрой ликвидности
Коэффициент валовой рентабельности	Коэффициент маневренности собственных оборотных средств	Коэффициент соотношения мобильных и иммобилизованных активов	Коэффициент оборачиваемости дебиторской задолженности	Коэффициент абсолютной ликвидности

Dep. Variable:	Y	No. Observations:	62
Model:	Probit	Df Residuals:	52
Method:	MLE	Df Model:	9
Date:	Чт, 17 ноя 2022	Pseudo R-squ.:	0.8369
Time:	04:14:28	Log-Likelihood:	-7.0110
converged:	True	LL-Null:	-42.975
Covariance Type:	nonrobust	LLR p-value:	6.368e-12

  

	coef	std err	z	P> z	[0.025	0.975]
const	2.5729	1.332	1.931	0.053	-0.038	5.184
X6	-0.0280	0.072	-0.389	0.697	-0.169	0.113
X7	-0.8711	0.638	-1.365	0.172	-2.122	0.380
X8	-0.5136	0.385	-1.335	0.182	-1.268	0.240
X11	0.0002	0.020	0.009	0.993	-0.039	0.039
X13	-0.0245	0.068	-0.362	0.717	-0.157	0.108
X19	0.8303	0.513	1.620	0.105	-0.174	1.835
X24	-0.7525	0.326	-2.306	0.021	-1.392	-0.113
X26	-3.1098	1.632	-1.906	0.057	-6.308	0.089
X27	-0.0787	0.414	-0.190	0.849	-0.891	0.734

**Рисунок 1 – Результат обучения пробит-модели**

В результате расчета (рисунок 1) получены необходимые коэффициенты, которые в дальнейшем можно использовать для прогнозирования дефолта сельскохозяйственных организаций. При работе использовался язык программирования python, библиотеки: pandas, statsmodels, seaborn, matplotlib.

### Библиографический список

1. **Ткачев, Р. А.** Логит и пробит-модели / Р. А. Ткачев // Молодежь – Барнаулу: Материалы XXII городской научно-практической конференции молодых ученых, Барнаул, 02–09 ноября 2020 года / Гл. редактор Ю.В. Анохин. – Барнаул : Алтайский государственный университет, 2021. – С. 168–169.

2. **Lisitsin, D. V.** Minimum gamma-divergence estimation for non-homogeneous data with application to ordered probit model / D. V. Lisitsin, A. G. Usol'tsev // Applied Methods of Statistical Analysis. Statistical Computation and Simulation – AMSA'2019 : Proceedings of the International Workshop, Novosibirsk, 18–20 сентября 2019 года / Editors: Boris Lemeshko, Mikhail Nikulin, Narayanaswamy Balakrishnan. – Novosibirsk: Новосибирский государственный технический университет, 2019. – pp. 227–234.

3. Simple and Multiple Linear Regression in Python. Available online: <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9> (дата обращения: 16.10. 2022).