

*А. В. Галковский*

## МОДЕЛИРОВАНИЕ РЕЗУЛЬТАТА ФУТБОЛЬНОГО МАТЧА

*В работе рассматривается подход к моделированию результата футбольного матча на основе статистических данных. Проведен анализ факторов, влияющих на результат игры. Для разработки модели выигрыша команды используется множественная логистическая регрессия. Построенная модель может использоваться для определения вероятности выигрыша команды, что позволит делать более обоснованные футбольные прогнозы.*

Последнее десятилетие футбол как вид бизнеса вышел на новый уровень и то ведение дел, какое было уже не может гарантировать успех, как со спортивной стороны, так и финансовой точки зрения. С развитием вычислительной техники объём статистических данных в футболе начал стремительно расти и обрабатывать их по «старинке» ни тренерам, ни скаутам, ни владельцам клубов уже не удаётся. В бейсболе, баскетболе, американском футболе уже твёрдо укоренилась практика оценки статистических данных и прогнозирования результатов. В футболе «пионерам» в этой области стали лондонский «Арсенал», «Олимпик Лион», ФК «Милан», ФК «Динамо-Киев», их примеры показали, что расчёты способны принести не только «кубки», но и переломить аксиому «убыточности» футбольного бизнеса. Как показывает практика, чтобы постоянно участвовать в высшем футбольном дивизионе своей страны, команде необходимо минимизировать проигрыш. Из этого вытекает цель исследования: построить модель для чемпионата Республики Беларусь, которая могла бы по статистическим данным матчей исследуемой команды дать прогнозное значение на исход следующей игры – будет набор хоть одного очка или нет. Для исследования был выбран ФК «Шахтёр» из Солигорска по следующим причинам:

- клуб постоянно борется за призовые места;
- клуб не из крупного города;
- клуб не покидал Высшую Лигу.

Состав относительно стабилен в отличие от других клубов Высшей Лиги. Информационной базой исследований послужили статистические данные, публикуемые на сайте АБФФ. Было рассмотрено 30 туров Высшей лиги Чемпионата Республики Беларусь по футболу 2016 года.

Для прогнозирования результата игры была выбрана модель множественной логистической регрессии (бинарного отклика), в которой непосредственно оценивается вероятность события по формуле

$$p = \frac{e^z}{1 + e^z} \quad \text{или} \quad p = \frac{1}{1 + e^{-z}},$$

где  $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$ ;

$\beta_j$  – коэффициенты регрессии при объясняющей переменной  $x_j$ ;

$j = 1, \dots, m$ .

Отсюда получаем величину  $z = \ln \frac{p}{1-p}$ , которая называется логит. Величина  $\frac{p}{1-p}$

представляет собой шансы.

Вероятность успеха игры (получение очков) возрастает с усилением влияния совокупности факторов, характеризующих как саму игру, так и уровень взаимодействия

команд. Данная функция не линейная, а S-образная. В логистической регрессии для оценки параметров используется метод максимального правдоподобия и итерационный алгоритм (из-за нелинейности). Преимущество применения логистической регрессии для моделирования результата игры позволяет не только определить предполагаемый бинарный отклик игры, но и рассчитать вероятность получения очков.

При построении модели в качестве отклика будет выступать факт получения командой очка, хотя бы одного очка. Проигранная игра не дает команде ни одного очка. Логит-модель, используя логистическую функцию для аппроксимации, позволяет определить оценки результирующего признака, то есть в данном случае – вероятности получения хотя бы одного очка (благоприятного исхода для команды). В качестве инструмента для проведения анализ и построения модели использовался пакет Statistica 8.0.

Рассмотрим влияние на получение или не получение командой в игре очков следующие показатели – забитые голы ( $x_1$ ), пропущенные голы ( $x_2$ ), процент владения мячом ( $x_3$ ), количество ударов ( $x_4$ ), количество ударов в створ ( $x_5$ ), угловые ( $x_6$ ), фолы ( $x_7$ ), офсайды ( $x_8$ ), показанные жёлтые карточки ( $x_9$ ), показанные красные карточки ( $x_{10}$ ), процент точных передач ( $x_{11}$ ), процент выигранных единоборств ( $x_{12}$ ).

Оценка логистической регрессии имеет вид

$$\hat{p} = \frac{1}{1 + e^{-\hat{z}}},$$

где  $\hat{z} = -0,744 + 5,828 x_1 + 6,791x_2 - 0,134x_3 + 0,105x_4 + 0,914x_5 - 0,136x_6 - 0,164x_7 + 1,610x_8 + 0,278x_9 - 13,775x_{10} - 0,128x_{11} + 0,439x_{12}$ .

Статистика  $\chi^2(12) = 27,003$ ,  $p$ -уровень  $Pv = 0,007$  говорят о достаточной адекватности построенной модели.

Следовательно, при увеличении забитых и голов, количества ударов и количества ударов в створ, офсайдов, процента выигранных единоборств вероятность получения командой очка увеличивается, а чем больше угловых, фолов, показанных красных карточки, тем меньше вероятность получения очка.

Предсказанные результаты позволяют сделать следующие *выводы*.

В турах, где вероятность близка к 1, клуб сыграл в ничью, однако и в некоторых поражениях также не абсолютный ноль, что говорит о том, что в этих матчах ФК «Шахтёр» мог набрать очки, если бы показал лучшую игру с точки зрения статистики. Однако здесь есть одно исключение, оно касается 23 тура, тогда ФК «Шахтёр» встречался с ФК «Витебск» и хотя солигорчане набрали 3 очка, вероятность оказалась меньше 1. Данный факт обосновать тем, что с точки зрения статистических данных наблюдаемая команда сыграла хуже, чем в других матчах, где она стабильно набирала 3 очка.

Построенная модели позволяет предсказать результат матча со следующим соперником, но только с точки зрения наблюдаемой команды. Иначе говоря, не важно, как играет соперник – главное, как играет наша команда. Данная модель может использоваться только для прогнозирования результата в играх внутреннего первенства Республик Беларусь, поскольку она тестировалась только на играх Высшей лиги Республик Беларусь.

## Литература

1 Ассоциация Белорусская федерация футбола [Электронный ресурс]. – Режим доступа: [http://abff.by/index.php?option=com\\_content&view=article&id=58&Itemid=2934&lang=ru/](http://abff.by/index.php?option=com_content&view=article&id=58&Itemid=2934&lang=ru/) – Дата доступа: 18.04.2017.