

А. В. Галковский

greatdru@gmail.com

Гомельский государственный университет им. Ф. Скорины, г. Гомель, Беларусь

ОЦЕНКА ФАКТОРОВ, ВЛИЯЮЩИХ НА РЕЗУЛЬТАТ ФУТБОЛЬНОГО МАТЧА

В работе дается оценка взаимосвязи данных послематчевой футбольной статистики. Проведена классификация значимых факторов, влияющих на результат матча. Показано, что все показатели разбиты на факторы атакующего и факторы оборонительного характера.

Сегодня любой спорт это бизнес, а футбол это бизнес планетарного масштаба. Футбол можно сравнить с IT-индустрией. Кто не пинал мяч хоть раз в жизни? А кто ни разу не сидел за компьютером и не пользовался благами мира высоких технологий? Таких людей мы сегодня вряд ли встретим. Футбол стал мощным социальным явлением, семейные выходные, отпуска и даже рабочие графики корректируются болельщиками в угоду футбольных матчей любой команды. Однако футбол до недавнего момента был крайне «консервативен» в отношении сторонних способов оценки как результата игры так и эффективности игроков. В NBA, NHL, MLB с 2002 года активно используется аппарат математической статистики для оценки различных игровых аспектов. Но футбол относительно недавно адаптировал и применил эти методы. Ярким примером этой работы могут являться английский Лестер Сити, и представители Бундеслиги – Хоффенхайм и РБ Лейпциг. Клубы с небольшими бюджетами резко вклиниваются в борьбу за чемпионство в своих чемпионатах.

Основной показатель финансового благополучия футбольного клуба – участие в высшем футбольном дивизионе своей страны. Это даёт возможность бороться за участие в матчах Лиги Чемпионов или Лиги Европы, получать более высокую прибыль от телевизионных трансляций и продажи билетов на стадион, приглашать лучших игроков, заключать лучшие спонсорские контракты. Как показывает практика, чтобы команда постоянно участвовала в Высшей футбольной лиги своей страны, достаточно не проигрывать матчи, даже игры в ничью, оставят команду в элитном дивизионе на следующий сезон. Поэтому, в этой работе, будет рассмотрена статистическая связь между послематчевой статистикой и фактом набора хотя бы одного очка командой за игру. Для анализа был выбран сезон 2016 года и первый круг сезона 2017 года Высшей лиги Чемпионата Республики Беларусь по футболу. Используемые данные находятся в открытом доступе на официальном сайте Белорусской федерации футбола. Изначально были исключены такие показатели как: владение мячом в минутах, процент точных передач, процент выигранных единоборств, количество красных карточек (из-за крайне малого количества), количество ударов (даёт мультиколлинеарность с показателем удары по воротам). Остальные показатели, такие как: забитые голы, пропущенные голы, процент владения мячом, удары в створ, угловые, фолы, офсайды, жёлтые карточки, точные передачи, выигранные единоборства, количество замен, игра дома или в гостях. Переменная «Набранные очки» принимает значение 0 – если команда проиграла и 1 – если была ничья или победа. Всего было рассмотрено 746 игр.

Для оценки влияния объясняющих переменных на результирующий фактор использовался тест ранговой корреляции Спирмена. Значимыми оказались следующие переменные с теснотой связи представленной в таблице 1.

Таблица 1 – Результаты теста ранговой корреляции Спирмена

Переменные	Набранные очки
Забитые голы	0,531
Пропущенные голы	- 0,681
Процент владения мячом	0,073
Удары в створ	0,351
Угловые	0,090
Фолы	0,107
Офсайды	0,102
Количество выигранных единоборств	0,151
Место проведения матча(дома или в гостях)	0,126

Из таблицы 1 видно, что забитые голы и пропущенные голы имеют умеренную тесноту связи. Однако пропущенные голы имеют значения больше чем забитые. Из чего можно предположить, что не пропустить гол важнее, чем забить противнику. Удары в створ имеет меньшую тесноту связи. Остальные переменные менее значимы для объясняющей переменной. Два наиболее слабо влияющие на исход матча показателя это угловые и владение мячом. Переменные: жёлтые карточки, точность передачи и замены, оказались статистически не значимы. Не последними показателями, влияющими на исход матча, оказались выигранные единоборства и фактор своего поля.

Для классификации показателей, влияющих на исход матча вне зависимости от места проведения, использовался кластерный анализ (метод «дальнего соседа», евклидово расстояние). Предварительно все данные были стандартизированы формуле:

$$z = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

В итоге факторы разбились на два кластера. В первый вошли выигранные единоборства, фолы, офсайды и пропущенные голы. Во второй – угловые, удары в створ, процент владения мячом и забитые мячи. Можно, грубо говоря, назвать первый кластер «кластером обороны», а второй «кластером атаки». Хотя вызывает сомнение попадание офсайдов и пропущенных голов в один кластер. Так как офсайд (в данном случае рассматривается полученные офсайды в атакующей фазе) происходит как минимум на половине поля противника. Здесь можно объяснить это тем, что после зафиксированного положения «вне игры», мяч переходит к команде противника. Тем самым противник начинает подготовленную атаку. В «атакующем» кластере всё предсказуемо.

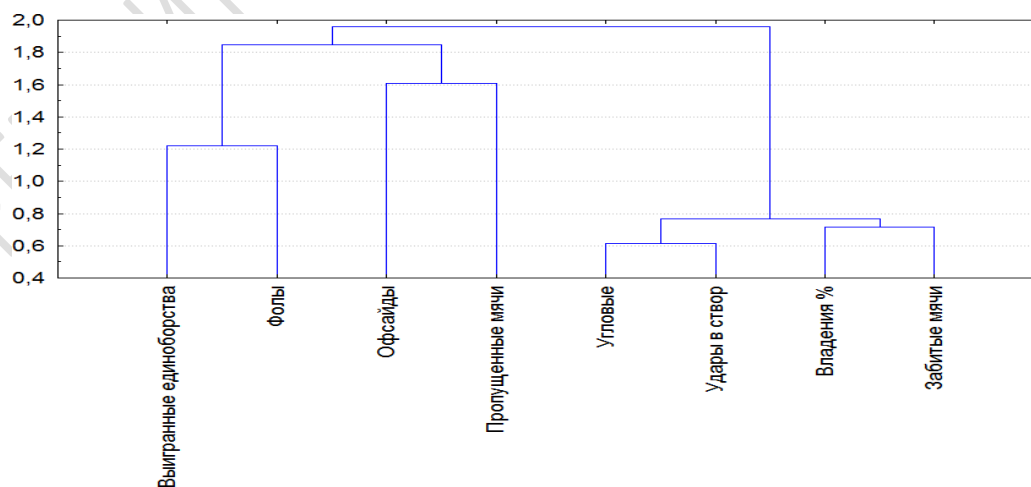


Рисунок 1 – Дерево кластеров

Результаты данной работы в дальнейшем будут использованы для моделирования набора командой очков путём построения модели множественной логистической регрессии. Данные исследования и модели актуальны исключительно для Высшей лиги Чемпионата Республики Беларусь по футболу. Так как различным странам присущи определённые стилистические особенности в футболе, тем более кубковым матчам.

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/ – Дата доступа: 18.04.2017.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ