

УДК 004:519.6:658.78

© *Т.В. Кожевникова, И.С. Манжула, 2021*

АПРОБАЦИЯ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО ДАННЫМ ДЗЗ ВЕГА-SCIENCE

Кожевникова Т.В. – н.с., «Лаборатория Информационных технологий» (ВЦ ДВО РАН), e-mail: ktvsl@mail.ru; *Манжула И.С.* – м.н.с., «Лаборатория Информационных технологий» (ВЦ ДВО РАН), e-mail: manzhula_94@mail.ru.

Предложен алгоритм разведочного анализа данных, реализованный в виде компьютерной программы, который выполняет прогноз урожайности на основе выборки из сервиса дистанционного зондирования Земли Вега-Science (для Белогорского района Амурской области). Полученные результаты позволяют оценить точность предложенной регрессионной модели с использованием метеорологических показателей и спутниковых данных более чем удовлетворительной для оценки степени влияния совокупности факторов на урожайность овса.

Ключевые слова: корреляционный и регрессионный анализ, информационная система, выборка, данные, дистанционное зондирование Земли.

Введение. Современные тенденции использования информационных технологий обеспечивают широкие возможности их применимости в естественных науках.

Научно-технический прогресс и развитие компьютерных сетей, создание высокоскоростных компьютеров и эффективного программного обеспечения привели к созданию глобального медиaprостранства. Внедрение данных и технологий глобального информационного пространства в жизнедеятельность человека действует как средство, уравновешивающее большой объем информации и ежедневно возрастающие объемы научных знаний.

VI Международная науч.-практ. конференция «Информационные технологии и высокопроизводительные вычисления»

Хабаровск, Россия, 14-16 сентября 2021 г. Сборник трудов. Отв. редактор Намм Р.В.

В данной работе представлен пример применения современных математических методов обработки данных для прогноза урожайности сельскохозяйственных культур, используя данные, хранящиеся в открытых источниках информации [1 - 3].

Исследование проводилось на основе рекомендаций специалистов Дальневосточного научно-исследовательского института сельского хозяйства (ДВ НИИСХ). В качестве социально-значимой культуры, влияющей на обеспечение населения продовольствием, была выбрана яровая культура - овёс. Построение прогнозируемой модели урожайности овса для Белогорского района Амурской области на основе данных, выгруженных из сервиса дистанционного зондирования Земли Vega-Science, является важным направлением исследования Дальнего Востока в рамках естественных и технических наук [4, 5]. На сегодняшний день Белогорский район является одним из крупнейших сельскохозяйственных районов Амурской области, по состоянию на 2020 год на его территорию приходится 8 сельскохозяйственных предприятий, 67 фермерских хозяйств и 5085 личных хозяйств.

Прогнозирование урожайности с/х культуры необходимо для решения практических задач, связанных с производством зерновых, а также в рамках планирования посевных площадей и экспортных операций субъектами ДФО.

Овёс является ключевой, экономически выгодной фуражной культурой Амурской области, основой для питания человека и животных. Белок овсяных круп богат незаменимыми аминокислотами, соединениями железа, кальция и фосфора, а также микроэлементами.

Целью исследования является разработать алгоритм разведочного анализа, который выполняет подбор и первичную обработку данных для прогноза урожайности сельскохозяйственных культур на основе статистических методов с использованием сервиса ВЕГА- Science.

Обработка данных для прогноза урожайности с/х культуры овёс на примере Амурской области. Для прогнозирования урожайности сельскохозяйственных культур использовались регрессионные модели, где в качестве независимых переменных выступают данные, полученные с помощью методов дистанционного зондирования Земли. Собранные в результате экологического мониторинга информация, перед проведением полноценного анализа, должна быть обработана, т.е. приведена в состояние, удобное для сравнений, обобщений и интерпретаций. Использование средств автоматизации не только упрощает процесс обработки данных, но и ускоряет его в несколько раз.

Первичная обработка исходной информации является фундаментом для проведения любого анализа или исследования. В первую очередь было произведено упорядочивание данных по дням, произведена сводка по каждому показателю для пахотных земель исследуемого муниципального образования Амурской области в период с 2010 по 2019 гг. По значениям еженедельных композитов NDVI были определены максимальные значения для каждого года, данный индекс считается наиболее эффективным для оценки состояния посевных площадей с воздуха, рассчитан ГТК Селянинова. В ходе обработки данных была обнаружена нехватка информации по объемам урожайности овса за 2010,2011 года. Данная проблема была решена путем линейной аппроксимации по методу наименьших квадратов.

В результате первичной обработки данных была составлена выборка с предикторами, непосредственно связанными с производимым исследованием: y - урожайность овса, x_1 - максимальный NDVI, x_2 - ГТК Селянинова, x_3 - количество дней активной вегетации, x_4 - атмосферное давление, x_5 - относительная влажность почвы, x_6 - нисходящая коротковолновая радиация и x_7 - влажность почвы (в слое 0-10 см). Данные показатели были выбраны исходя из рекомендаций специалистов предметной области – сотрудников ДВ НИИСХ. В качестве результирующего признака был выбран показатель объемов урожайности овса, в качестве факторных признаков были выбраны метеорологические характеристики, максимальный вегетационный индекс и гидротермический коэффициент увлажнения.

После формирования выборки, для автоматизации дальнейшей обработки данных, был разработан алгоритм для прогноза урожайности. Многофакторная регрессионная модель, с последующим удалением взаимно коррелирующих факторов, строилась в следующем виде:

$$y = b + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_7x_7$$

Следующим по порядку, но не по важности для проведения естественно-научного исследования является этап алгоритмизации задачи. Данный этап служит базой для написания кода будущей программы, на этом этапе устанавливается последовательность действий для решения поставленной задачи. Для наглядности алгоритм программы построения уравнения множественной регрессии для прогнозирования урожайности овса представлен в формате блок-схемы (рис. 1). Алгоритма реализован на языке программирования Python.



Рис. 1. Алгоритм разрабатываемого ПО.

С целью выявления тесноты связей между зависимой и независимыми переменными была построена корреляционная матрица, представленная на рис. 2.

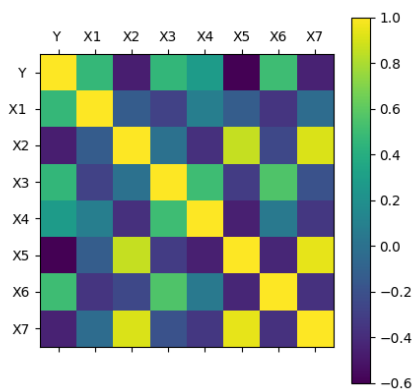


Рис. 2. Цветовое представление корреляционной матрицы.

Проанализируем полученную матрицу. Для построения регрессионной модели в качестве независимых предикторов целесообразно

оставить только один из двух показателей, характеризующих влажность почвы и в качестве интегральной климатической характеристики сохранить в модели ГТК.

Уравнение множественной регрессии, характеризующее зависимость урожайности овса от включенных в модель показателей, построенное по данным 2010-2019 гг., имеет следующий вид:

$$y = 2395.26 + 143.91x_1 - 12.38x_2 + 0.78x_3 - 2.68x_4 + 0.02x_6 + 1.83x_7$$

На рис. 3 показан график, на котором наглядно отображены фактические значения средней урожайности овса в 2010-2019 гг. и значения, полученные на основе регрессионной модели.

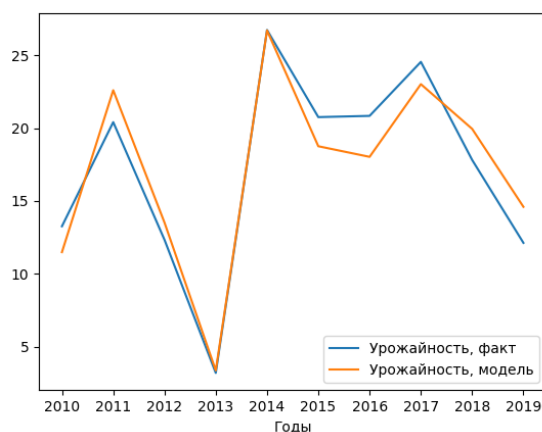


Рис. 3. Фактическая и смоделированная урожайность овса.

Заключение. Предложен алгоритм разведочного анализа данных, реализованный в виде компьютерной программы, который выполняет прогноз урожайности на основе выбранных данных из сервиса дистанционного зондирования Земли Vega-Science для Белогорского района. Результаты моделирования сопоставимы с фактическими данными. Точность результатов моделирования составляла от 76,7% до 92,2%.

Полученные результаты позволяют оценить точность предложенной регрессионной модели с использованием метеорологических показателей и спутниковых данных более чем удовлетворительной для оценки степени влияния совокупности факторов на урожайность овса в муниципальных образованиях.

Разработанное программное обеспечение рекомендовано для обработки экспериментальных данных экологического мониторинга специалистам, деятельность которых связана с проведением статистических исследований, а также напрямую связанных с изучением различных

сельскохозяйственных культур и оптимизацией использования почвенных ресурсов.

Библиографические ссылки

1. Кремер Н.Ш. Теория вероятностей и математическая статистика : учебник для вузов. М.: ЮНИТИ-ДАНА, 2007. 573 с.
2. Мандель И. Д. Кластерный анализ. М.: Финансы и статистика. 1988. 176 с.
3. Факторный, дискриминантный и кластерный анализ : сб. / под ред. И. С. Енюкова. М.: Финансы и статистика, 1989. 215 с.
4. Использование спутникового сервиса ВЕГА в региональных системах дистанционного мониторинга /Лупян Е.А., Барталев С.А., Толпин В.А. и др. // Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11, №. 3. С.215-232.
5. Спутниковый сервис мониторинга состояния растительности ("Вега") / Лупян Е.А., Савин И.Ю., Барталев С.А. и др. // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8, № 1. С. 190-198.