

Казнади Сергей Константинович

бакалавр пед. наук, магистрант

ФГБОУ ВО «Кубанский государственный университет»

г. Краснодар, Краснодарский край

PYTHON И R В АНАЛИЗЕ ДАННЫХ

Аннотация: в данной статье поднимается вопрос о выборе языка для изучения анализа данных, проведен сравнительный анализ между языками R и Python. Автором также рассмотрено место, которое они занимают в мире изучения данных и статистики.

Ключевые слова: языки программирования, Python, анализ данных.

Когда поднимается вопрос, какой из инструментов лучше использовать при статистической работе с данными, самым популярным ответом будет R или Python. Действительно, эти два языка программирования уже давно соперничают за звание лучшего инструмента для работы с данными. У каждого из них есть свои достоинства и недостатки, поэтому выбор того или иного языка зависит часто от конкретной ситуации, издержек на обучение, а также от того, какие еще программные средства требуются для решения задачи. В то время, как язык R разрабатывался с упором на потребности статистиков, Python популярен своим понятным и читаемым синтаксисом и универсальностью. И если в бизнесе есть возможность выбора нужной технологии исходя из решаемой задачи, то в образовании зачастую нет времени на изучение аналогов. В данной статье проведем сравнительный анализ между языками R и Python, а также рассмотрим место, которое они занимают в мире изучения данных и статистики.

Начнем с языка R. Разработчики языка Ross Ihaka и Robert Gentleman задумывали язык R как свободную реализацию языка программирования S. Целью авторов при разработке языка была обеспечение более качественного и понятного подхода к анализу данных, статистике и графическим моделям. С начала создания R использовался преимущественно в академической среде, но сравнительно недавно стал проникать и в мир компаний, использующие

информационные технологии. Поэтому R является наиболее бурно развивающимся статистическим языком, используемым в корпоративной практике. Язык R просто создан для визуализаций, а также обладает богатой экосистемой современных пакетов для анализа данных и располагает активным сообществом поддержки. В плане изучения является достаточно сложным языком ввиду своего компактного синтаксиса.

Обычно R применяется в тот момент, когда для анализа данных необходимы выделенные вычислительные мощности или отдельные сервера. Язык отлично подходит для исследовательской работы, удобен практически при любом варианте анализа данных, поскольку существует множество пакетов и готовых тестов, обеспечивающих нужный инструментарий для быстрого начала. Вплоть до того, что R может быть использован в качестве элемента решения в области больших данных [1].

Язык программирования Python – это высокоуровневый язык программирования общего назначения. Автором данного языка является нидерландский программист Гвидо ван Россум, приоритетной целью которого было разработать язык программирования, повышающий производительность разработчика и читаемость кода.

Python имеет минималистичный синтаксис. Основные архитектурные черты языка: динамическая типизация, автоматическое управление памятью, механизм обработки исключений, поддержка многопоточных вычислений и высокоуровневые структуры данных. Язык поддерживает несколько парадигм программирования, в том числе структурное, объектно-ориентированное, функциональное, императивное и аспектно-ориентированное.

Стандартная библиотека является одной из привлекательных сторон языка программирования Python. В ней представлены средства для разноплановой работы:

- со многими сетевыми протоколами и форматами интернета;
- с операционной системой;
- с регулярными выражениями;

- с текстовыми кодировками;
- с мультимедийными форматами;
- с криптографическими протоколами;
- с базами данных;
- с архивами;
- с сериализацией данных;
- с поддержкой юнит-тестирования.

Целесообразность применения Python обычно наступает в тех случаях, когда есть требование связать анализ данных с работой веб-приложений или внедрить статистический код в рабочую базу данных. Python, будучи полнофункциональным языком программирования, отлично подходит для реализации алгоритмов с их последующим практическим использованием. Совсем недавно пакеты для анализа данных на Python находились только в начальном состоянии, что представляло определенную проблему, но в последние годы ситуация значительно улучшилась [2].

С точки зрения анализа данных, R выглядит явным лидером. Если же рассматривать вопрос более комплексно, с образовательной точки зрения, то здесь больше всего подойдет именно Python, выигрывающий за счет своей простоты и универсальности. Поэтому, с нашей точки зрения, более целесообразно начинать изучение программирования и анализа данных именно с языка Python. Только так возможно обеспечить преемственность знаний на протяжении нескольких лет при изучении программирования, веб-дизайна и компьютерного моделирования.

Также в последнее время среди специалистов по анализу данных появилась тенденция к переходу с R на Python. Более того, растет доля тех программистов, которые владеют обоими языками и по мере необходимости используют то один, то другой.

Список литературы

1. Кабаков Р. R в действии. Анализ и визуализация данных на языке R. – М.: ДМК Пресс, 2014. – С. 198.
2. Маккинни У. Python и анализ данных. – М.: ДМК Пресс, 2015. – С. 276.