|  |
| --- |
| МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное автономное образовательное учреждение  высшего образования |
| Национальный исследовательский ядерный университет «МИФИ»  IFES-logo |

ФАКУЛЬТЕТ

КИБЕРНЕТИКИ И ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Кафедра №75 «Финансовый мониторинг»

ОТЧЕТ

о домашнем задании

по курсу:

«Формы и методы подготовки аналитической информации»

Выполнил: Серебряков Александр,

студент группы С14-506

Проверила: Писарчик Е.Е.

Москва, 2017

## Постановка задачи

Требуется определить, какой язык программирования следует выбрать Python или R, чтобы проводить анализ данных. Выбор осуществляется с точки зрения студента-программиста, имеющем опыт работы с этими .

Принятие решения будет осуществляться с помощью двух методов:

1. Pros, Cons and Fixes
2. Weighted Ranking

## Pros, Cons and Fixes

Суть метода заключается в четком определении проблемы (сделано в постановке задачи), идентификации набора возможных альтернатив, позволяющих решить проблему, подборе «за» и «против» для каждой альтернативы. Эти данные оформляются в виде таблицы. Далее рассматриваются все «против» для каждой альтернативы и по мере возможности они объединяются либо к ним ищутся решения (fixes). После этого заполняется новая таблица с исправленными «против».

На основе новой таблицы делается решение в пользу той альтернативы, в которой «за» больше всего перевешивают «против».

Составим первую таблицу.

|  |  |  |  |
| --- | --- | --- | --- |
| язык программирования Python | | язык программирования R | |
| Pros | cons | Pros | cons |
| Простота в изучении | медленный | Результаты анализа более информативны | Сложен для изучения программисту с сложившимся представлением о том, как должен работать язык |
| Читаемость кода | Визуализация данных | Наличие встроенных мощных аналитических пакетов/библиотек | Без знания логики языка сложно разобраться в написанном коде |
| Множество библиотек для анализа данных(но в R больше) | Могут возникнуть проблемы с установкой библиотек для анализа данных | Открытый код | Для анализа необходимы заранее подготовленные данные |
| многоцелевой(возможность использования в других областях, отличных от анализа данных) | Отсутствие многих методов data mining | Огромное сообщество разработчиков | медленный |
| Огромное сообщество разработчиков | Результаты анализа менее информативны | Хорошая визуализация данных | Трудно внедрить в реальный программный продукт |

Таблица 1 – Первоначальные Pros, Cons

Далее пытаемся выявить возможные fixes для каких-либо «против».

|  |  |  |
| --- | --- | --- |
|  | Cons | Fixes |
| Язык программирования Python | Могут возникнуть проблемы с установкой библиотек для анализа данных | Нужно потратить побольше времени на установку |
| Язык программирования R | Сложен для изучения программисту с сложившимся представлением о том, как должен работать язык | можно решить эти две проблемы, потратив дополнительное время на более подробное изучение языка и его инструментов |
| Без знания логики языка сложно разобраться в написанном коде |

Таблица 2 – Cons и Fixes

Составляем итоговую таблицу Pros и Cons.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| язык программирования Python | | | язык программирования R | |
| Pros | cons | Pros | | cons |
| Простота в изучении | медленный | Результаты анализа более информативны | | Для анализа необходимы заранее подготовленные данные |
| Читаемость кода | Визуализация данных | Наличие встроенных мощных аналитических пакетов/библиотек | | медленный |
| Множество библиотек для анализа данных(но в R больше) | Результаты анализа менее информативны | Открытый код | | Трудно внедрить в реальный программный продукт |
| многоцелевой(возможность использования в других областях, отличных от анализа данных | Отсутствие многих методов data mining | Огромное сообщество разработчиков | |  |
| Огромное сообщество разработчиков |  | Хорошая визуализация данных | |  |

Таблица 3 – Итоговые Pros, Cons

На основании последней таблицы можно сделать вывод, что наилучшим выбором для анализа данных является язык программирования R.

## Weighted Ranking

Суть метода заключается в четком определении проблемы (сделано в постановке задачи), идентификации набора возможных альтернатив, позволяющих решить проблему, подборе значимых критериев выбора. Эти данные оформляются в виде таблицы.

На первом этапе мы попарно сравниваем критерии, определяя, насколько один приоритетнее другого в данной паре. При каждом сравнении победившему критерию записывается одно очко. В итоге критерии ранжируются в соответствии с набранными очками и средин их выбирается, например, три наиболее важных. Их приписываются определенные веса. Выбор количества важных критериев оценка их весов зависит от конкретной ситуации и точки зрения аналитика.

На втором этапе альтернативы сравниваются попарно между собой и также набирают «очки». После сравнения подсчитываются итоговые значения весомости каждой альтернативы, равные сумме произведений количеств очков за определенный критерий на вес этого критерия. Выбирается альтернатива, имеющая наибольшую весомость.

Составим таблицу критериев и попарно их сравним. Одному очку будет соответствовать символ «I».

|  |  |  |
| --- | --- | --- |
| Критерий | Количество очков | Ранг |
| Простота в изучении |  | 4 |
| Множество библиотек для анализа данных | I | 3 |
| Возможность использования в других областях, отличных от анализа данных | II | 2 |
| Качество анализа | III | 1 |

Таблица 4 – Ранжирование критериев

Самыми важными оказались критерии «Множество библиотек для анализа данных», «Возможность использования в других областях, отличных от анализа данных», «Качество анализа». Припишем им веса 0.2, 0.3, 0.5 соответственно.

Попарно сравним альтернативы по выбранным критериям и высчитаем итоговые значения весомостей.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Множество библиотек для анализа данных (0.2) | Возможность использования в других областях, отличных от анализа данных  (0.3) | Качество анализа (0.5) | Весомости | Ранг |
| Python |  | I |  | 1 \* 0.3 = 0.3 | 2 |
| R | I |  | I | 1\*0.2 + 1\*0.5 = 0.7 | 1 |

Таблица 5 – Ранжирование альтернатив

На основании последней таблицы можно сделать вывод, что наилучшим выбором для анализа данных является язык программирования R.