ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

*Факультет бизнеса и менеджмента*

Юсифов Эмиль Вадимович

**Разработка итерационного алгоритма для оценки эффективности инвестиционного портфеля в процессе его формирования: предложения и рекомендации по IT сервису**

Выпускная квалификационная работа

по направлению подготовки *38.03.05 бизнес-информатика*

образовательная программа «Бизнес-информатика»

|  |  |
| --- | --- |
| Рецензент  к.т.н., профессор  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Е.А. Исаев | Научный руководитель  к.т.н., доцент  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Н.В. Сизых |

Москва 2020

**ОГЛАВЛЕНИЕ**

[ВВЕДЕНИЕ 3](#_Toc39429364)

[Глава 1. Анализ современных методов формирования портфеля и процесс их формирования 7](#_Toc39429365)

[*1.1 Теоретический обзор публикаций по теме финансового портфеля* 7](#_Toc39429366)

[*1.12 Теоретический обзор публикаций по теме методов формирования портфеля* 8](#_Toc39429367)

[Выводы к главе 1 18](#_Toc39429368)

[Глава 2 Реализация алгоритма, комбинирующего PCA, k-means и итеративный алгоритм 19](#_Toc39429369)

[*2.1 Отрасли финансовых рынков* 19](#_Toc39429370)

[*2.2 Концепция алгоритма* 20](#_Toc39429371)

[*2.3 Сбор данных и алгоритм* 21](#_Toc39429372)

[*2.4 Сравнение доходностей* 37](#_Toc39429373)

[Выводы к главе 2 37](#_Toc39429374)

[Глава 3 Бизнес-применение итеративного алгоритма 38](#_Toc39429375)

[3.1 Архитектура решения 38](#_Toc39429376)

[Выводы к главе 3 40](#_Toc39429377)

[Заключение 41](#_Toc39429378)

# ВВЕДЕНИЕ

В современном финансовом мире сбор наиболее эффективного портфеля в условиях сильной неопределенности событий является достаточно сложной и увлекательной сферой изучения. Есть много уважаемых работ на данную тему, но нельзя с полной уверенностью сказать, что данная тема изучена и по-прежнему много пространства для изучения остается свободным. Эта тема очень непростая, поскольку на рынке существует сильная неопределенность, на внешнем уровне – это макроэкономики государств, на внутреннем уровне – это микроэкономики различных институтов. Многие специалисты пытаются применить различные научные методы для укрепления инвестиционного портфеля и используют методы машинного обучения, статистические методы и глубокое обучение на основе данных. Всех специалистов в этой сфере объединяет одна общая цель – максимизировать доходность портфеля при наименьших возможных рисках.

Как и в прошлом столетии, так и сейчас, нужно своевременно реагировать на колебания рынка, хеджировать риски, не жертвуя доходностью портфеля. Чтобы успевать за рынком, специалисты прибегают к алгоритмам и автоматизации процессов.

В текущей работе передо мной, как исследователем стоит следующий вопрос: жизнеспособна ли комбинация машинного обучения и итерационного алгоритма для доходного и минимизированного по риску инвестиционного портфеля. Для усиления эффекта от алгоритма я также внедрил фундаментальный анализ финансовых метрик.

**Объектом** исследования является процесс формирования портфеля ценных бумаг.

**Предметом** исследования является разработка комбинации машинного обучения и итерационного алгоритма для оценки эффективности портфеля

**Целью работы** является алгоритм и рекомендации для создания IT-сервиса с применением разработанного алгоритма.

В данном исследовании выдвигаются следующие **гипотезы:**

1. Автоматизация процесса формирования портфеля ценных бумагс использованием итерационного алгоритма и машинного обучения способствует повышению эффективности (высокий доход, низкий риск) портфеля по сравнению с другими инвестиционными стратегиями.
2. Использование машинного обучения и итерационного алгоритма соответствует реалиям фондового рынка и способствует оперативному реагированию на колебания рынка.

**Задачи работы:**

В соответствии с поставленной целью в работе решаются следующие задачи:

* провести анализ методов формирования инвестиционного портфеля и методов оценки эффективности инвестиционного портфеля.
* Протестировать концепцию алгоритма на реальных данных, получить финальный портфель
* Сравнить результаты финального портфеля с некоторыми результатами инвестиционных стратегий

**Ограничения.** В контексте данного исследования были использованы акции. Также данный алгоритм не учитывает возможности коротких продаж.

**Допущение** исследовательской работы состоит в том, что данное исследование будет использоваться частным инвестором и мною лично в дальнейшем для укреплений его/моих решений об инвестировании.

**Методы исследования:**

В данной работе для достижения поставленных задач были использованы следующие методы:

* Сбор данных по финансовым мультипликаторам, заранее выбранных компаний из базы данных Fundamental Analysis ([FundamentalAnalysis](https://pypi.org/project/FundamentalAnalysis/), библиотека python) и расчет средней динамики изменения котировок, по данным yahoo.finance ([finance.yahoo](https://finance.yahoo.com/)), за выбранный промежуток времени;
* Построение блок-схемы алгоритма;
* Проведение PCA по финансовым мультипликаторам, метод главных компонент;
* Кластеризация K-means по финансовым мультипликаторам. Выделение наилучших кластеров;
* Итерационный алгоритм по доходности и риску всех возможных вариантов портфелей из оставшихся тикеров.
* Оценка эффективности комбинации ML и итерационного алгоритма для составления портфеля ценных бумаг с помощью показателей риска и доходности;
* Прогнозирование значения на несколько месяцев вперед.
* Использование блок-схемы, построенной в сервисе camunda modeler, для визуализации рекомендаций по IT-сервису.
* Использование языка программирования Python, библиотек Машинного Обучения, API-запросов.

**Исходные данные:**

Доходности были скачаны с yahoo.finance, фундаментальные метрики были скачаны с библиотеки python FundamentalAnalysis.

Случайным образом рассмотрено 200 тикеров из FundamentalAnalysis базы, из которых была сделана фильтрация (наличие последней актуальной информации по доходности и фундаментальным метрикам).

**Научная новизна и практическая значимость:**

Научная новизна настоящей работы заключается в использовании простой для понимания комбинации итерационного алгоритма с применением PCA и кластерного анализа для автоматизации процесса формирования портфеля ценных бумаг.

Кроме того, исследование будет полезно для дальнейшего развития теоретических работ, посвященных использованию итерационного алгоритма в комбинации с ML.

**Актуальность:**

В наши дни многие компании обращают внимание на пользу применения математических моделей и машинного обучения для повышения эффективности процессов (больше/столько же выгоды, меньше затрат), высвобождения человеческих ресурсов, ускорения, адаптации к современным реалиям. Особенно заметен интерес банков, страховых компаний и семейных офисов.

Структурно данная работа состоит из введения, глав, заключения и приложений.

В **первой главе** описываются предпосылки к исследованию, известные методы в сфере формирования портфеля.

Во **второй главе** описывается алгоритм, который был использован в работе. Приводятся выдержки из кода с пояснением.

В **третьей главе** рекомендации и бизнес-применение для алгоритма.

Основными источниками информации являются книга Marcos Lopez de Prado - Advances in Financial Machine Learning 2018 Wiley и сайт arxiv.org.

# Глава 1. Анализ современных методов формирования портфеля и процесс их формирования

## *1.1 Теоретический обзор публикаций по теме финансового портфеля*

Для начала стоит уточнить разницу между формированием и оптимизацией портфеля. Данное исследование из области формирования инвестиционных портфелей. После этого этапа как правило наступает следующий этап – это оптимизация инвестиционных портфелей, что не входит в данное исследование, т.к. для меня, как для исследователя, важно сперва сформировать эффективный портфель без дальнейшего учета распределения денежных средств по активам.

Мы живем в очень динамичном мире, и изобретение портфелей как раз было одним из решений проблемы динамичности и изменчивости финансового мира. Есть две сильные стороны портфеля:

* Полезные свойства (получение дивидендов и купонов, а также рост бумаг и пр.), которые появляются при вложении в набор бумаг, которые нельзя придать своим инвестициям при покупке бумаг только одного эмитента.
* Несмотря на изменчивость и динамичность мира можно рассчитывать на максимально возможный доход и наименьший риск несмотря на изменение состояния рынка.

В области формирования портфелей было много выдающихся исследований, и выработались основные подходы и понятия.

Термин "**портфель**" относится к любой комбинации финансовых активов, таких как акции, облигации и денежные средства. Портфели могут находиться у индивидуальных инвесторов и/или управляться профессионалами в области финансов, хедж-фондами, банками и другими финансовыми учреждениями. Основные принципы сбора портфеля, на основе которых собирается портфель – толерантность инвестора к риску, сроки инвестирования и инвестиционные цели. Главное понятие портфеля – соотношение риска и доходности портфеля.

Некоторые аналитики классифицируют по принципу источника доходов. Есть портфели роста, есть портфели дохода. Цель первого “вырастить” стоимость портфеля, цель второго получать дивиденды, купоны и пр.

Портфели, которые ориентированы на рост, также принято делить на портфели высокого/среднего/умеренного роста. Фокус высокого роста – быстрорастущие компании с большим риском. Фокус умеренного роста – хорошо известные компании, которые растут. Фокус среднего роста – комбинация портфелей первого и второго типа.

Портфель дохода собирается по принципу наибольших выплат и умеренному росту стоимости акций. Как правило, для такого портфеля выбираются высоконадежные активы. Тут также выделяют разные типы – постоянного и высокого дохода.

Одна из главных идей портфеля – эффект компенсации, потеря в одной доле актива может компенсироваться ростом в другой. Например, доля портфеля принесла доход за счет роста стоимости акций, а другая доля принесла доход с получением дивидендов и процентов.

Также это можно воспринимать как ставку на один из активов, где один показал умеренный рост, а другой проявил экспоненциальный рост, мат ожидание в таком случае выше, чем в случае инвестирования в один актив.

## *1.12 Теоретический обзор публикаций по теме методов формирования портфеля*

При формировании портфеля ценных бумаг используется большое количество теорий и методов, можно выделить основные из них:

* 1) Метод Майкла О'Хиггинса и Гарднеров
* 2) Модель Г. Марковица
* 3) Модель Шарпа (модель САРМ)
* 4) Модель равновесной цены
* 5) Модель Блэка-Литермана

1. **Метод Майкла О'Хиггинса и Гарднеров**

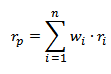
Основная идея данного метода – уменьшение инвестиционного риска при выборе определенных ценных бумаг, которые приносят доход в течение короткого времени. Также важно, что данный метод не требует анализа экономических и фундаментальных метрик, что скорее является слабостью данного метода, т.к. метод рассматривает лишь метрики временных рядов, не принимая во внимание воздействие других реальных факторов.

Упрощенная последовательность действий в данном методе.

* Основываясь на Dow Jones, отобрать 10 компаний с максимальными дивидендами.
* Из выбранных оставить топ-5 по наименьшей цене
* Убрать первую в списке, с самой минимальной ценой
* Разделить свои деньги на 5 частей. На 2/5 инвестиций купить компании с минимальными ценами, а на 3/5 остальные.

1. **Модель Г. Марковица**

Очень популярный метод в области формирования и оптимизации портфеля. Доходность в данной модели определяется как средневзвеш. суммарная доходность бумаг, которые входят в портфель. Формула (1) ниже, где wi – доля бумаг данного класса в портфеле, ri – доходность бумаги.



(1)

Риск в данной модели рассматривается как среднекв. откл. доходности бумаги. При расчете риска всего портфеля анализируется изменение отдельной бумаги, а также взаимное влияние между бумагами.

Марковиц разбил портфельный риск на 2 группы – группа систематического риска и специфический риск. Первая группа характеризуется экономикой, психологией и политикой, а именно как эти факторы влияют на составляющие портфеля. Ко второй группе относятся бумаги, которые содержат в себе свой уникальный риск.

Также важная часть модели — это критические линии, которые позволяют идентифицировать те портфели, которые дают нужный доход в соответствии со своим риском. Соответственно, изучая критические линии, можно выделить эффективно собранные портфели, которые минимальны по риску (или соответствуют приемлемому уровню риска инвестора) и максимальны по доходности.

Портфель доминирует над другим портфелем, если ожидаемая доходность \ мю  больше или равна доходности другого портфеля, а стандартное отклонение \ сигма  его стоимости меньше, чем у другого портфеля, или если ожидаемая доходность \ мю  больше и стандартное отклонение \ сигма  такое же. Это не означает, что портфель имеет одинаковый состав. Стандартное отклонение является результатом колебаний цены (спред) и, следовательно, является мерой риска портфеля.

Портфель называется **эффективным,** если в нем не доминирует какой-либо другой портфель, т.е. час если нет другого портфеля с более низким риском с такой же ожидаемой доходностью или более высокой доходностью с таким же риском.

Линия **эффективности** — это геометрическое расположение всех эффективных комбинаций риск-доходность.

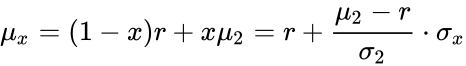
Оптимальное соотношение риск-доходность может быть проиллюстрировано с использованием двух ценных бумаг. В этой ситуации в зависимости от предпочтения инвестора к риску определяется оптимальная стратегия.

Мы считаем без риска р. Кроме того, мы хотим принять возможность коротких продаж (LV). В рассматриваемых случаях рассматривается рискованное обеспечение ценовым риском и риском дефолта (также валютным риском). Срок должен соответствовать плановому периоду. Таким образом, процентные риски и риски дефолта могут быть исключены для безрискового инструмента.

#### **Случай 1: *μ* 2 > *r,* без короткой продажи**

Рентабельность ценных бумаг выше, чем безрисковая ставка, и короткие продажи отсутствуют. Единственный выбор, который есть у инвесторов – это доля х \ в [0,1] их средств, которую они вкладывают в рискованные бумаги. Затем акция поступает в 1-х безрисковую облигацию.

*Эффективная граница* представляет собой прямую линию от комбинаций выход риска/



(2)

Где, \ sigma _ {x} = \ sigma _ {2} \ cdot xс х \ в [0,1].

Это означает:

* Отношение риск-доходность является линейным.
* Префактор риска портфеля соответствует стандартизированной премии за риск. Это избыточная доходность рискованной ценной бумаги, разделенная на ее риск.

(\ mu _ {x}, \ sigma _ {x}) соотношении компонентов типа (1-х, х)с 0 \ leq x \ leq 1.

\ sigma _ {x} = {\ sqrt {(1-x) ^ {2} \ sigma _ {1} ^ {2} + x ^ {2} \ sigma _ {2} ^ {2} +2 \ sigma _ {1} \ sigma _ {2} (1-x) x \ rho _ {{1,2}}}}\ mu _ {x} = r (1-x) + \ mu _ {2} x = r + (\ mu _ {2} -r) x(3)

Рассмотренный здесь особый случай вытекает из того факта, что первая бумага, облигация, является безрисковой, что математически \ sigma _ {1} = 0 и вытекает из нее \ sigma _ {x} = x \ sigma _ {2}.

#### **2-й случай: *μ* 2 > *r,* с короткими продажами**

Доходность по рискованным ценным бумагам больше, чем безрисковая ставка, и короткие продажи разрешены. Математически допустимость коротких продаж означает, что доля х средств, вложенных в рискованную бумагу, больше не ограничена интервалом [0,1].

Эффект левереджа заключается в том, что, если безрисковый инструмент продается с короткой ценой, ожидаемая стоимость портфеля увеличивается, но также и риск в форме большей диверсификации. Продажа безрисковой облигации означает х> 1что означает вкладывание больших средств в рискованную бумагу. Для полного возврата \ mu _ {x}результатов

\ mu _ {x} = r (1-x) + \ mu _ {2} x = \ mu _ {2} + (\ mu _ {2} -r) \ cdot (x-1)с х> 1. (4)

Продажа рискованной бумаги короткими продажами х <0. Таким образом, общий доход \ mu _ {x} = r + (\ mu _ {2} -r) \ cdot x меньше требуемого минимального дохода р.

Формальный процесс состоит из заимствования доли, продажи и вложения полученных средств в безрисковую бумагу. Ссуда ​​означает, что сторона, предоставляющая долю в кредит, возмещает все выплаты (дивиденды), возникшие в результате владения акцией, и что эта доля выкупается на рынке в конце срока и возвращается этой стороне.

Продавец коротких позиций несет тот же риск, что и акционер, и в этом случае приносит \ mu _ {2}> r более низкую доходность, чем можно было бы получить без риска. Таким образом, портфели, созданные путем короткой продажи рискованного актива, в этом случае неэффективны.

#### **3 случай: *μ* 2 < *r,* без короткой продажи**

Рискованная безопасность возвращает меньше, чем безрисковая ставка, и короткие продажи отсутствуют. То же самое относится и к общему доходу \ mu _ {x} = r- (r- \ mu _ {2}) \ cdot x, \, x \ in [0.1]. В этом случае эффективен портфель, который инвестирует только в безрисковый инструмент, поскольку, принимая повышенный риск, т. Е. х> 0Делая выбор, вы уменьшаете доход.

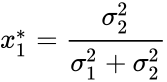
#### **4. Случай *μ* 2 < *r,* с короткими продажами**

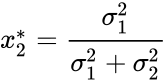
Допускаются короткие продажи: если рискованный инструмент продается в короткую, то есть х <0, доходность портфеля \ mu _ {x} = r + (r- \ mu _ {2}) \ cdot (-x)может быть увеличена по желанию, только если общий риск увеличивается из-за короткой продажи.

Снятие ограничения на короткие продажи не приводит к изменениям в портфеле минимальных отклонений, если корреляция Ро  принимает определенные значения, которые являются результатом отношения стандартных отклонений двух акций.

**Некоррелированные ценные бумаги**

Эффект \ rho = 0диверсификации всегда происходит с некоррелированными ценными бумагами. Оптимальное соотношение смешивания {\ displaystyle (x_ {1} ^ {*}, x_ {2} ^ {*} = 1-x_ {1} ^ {*})}:

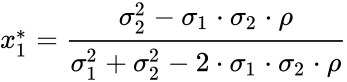
 (5)

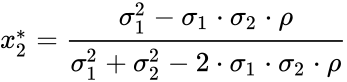


Коррелированные ценные бумаги

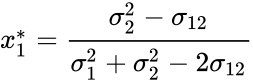
Диверсификация риска в зависимости от коэффициента корреляции Ро :

Выбор портфеля приводит к *минимальному отклонению портфеля* (сокращенно: *MVP*):

 (6)



Если ковариация известна, формула выглядит:

(7)

**Эффективный портфель, три рискованные ценные бумаги**

* Глобальное отклонение минимального портфеля с отрицательными пропорциями:

Ордината является местоположением всех комбинаций ценных бумаг 1 и 3, а абсцисса является комбинацией ценных бумаг 2 и 3.

* Глобальное отклонение минимального портфеля с положительными пропорциями:

Для x_ {1} + x_ {2} + x_ {3} = 1Есть две зависимые переменные.

\ mu _ {x} = \ mu _ {1} x_ {1} + \ mu _ {2} x_ {2} + (1-x_ {1} -x_ {2}) \ mu _ {3}

\ sigma _ {x} ^ {2}= x_ {1} ^ {2} \ sigma _ {1} ^ {2} + x_ {2} ^ {2} \ sigma _ {2} ^ {2} + (1-x_ {1} -x_ {2} ) ^ {2} \ sigma _ {3} ^ {2}+2 \ sigma _ {1} \ sigma _ {2} x_ {1} x_ {2} \ rho _ {{1,2}} ~ (8)

+2 \ sigma _ {1} \ sigma _ {3} x_ {1} x_ {3} \ rho _ {{1,3}} ~ +2 \ sigma _ {2} \ sigma _ {3} x_ {2} x_ {3} \ rho _ {{2,3}} ~

**Эффективные портфели для бумаг n**

Это может быть определено только математически:

{\ displaystyle \ min _ {x_ {1}, \ ldots, x_ {n}} \ Sigma \ left [\ rho _ {ij} \ cdot \ sigma _ {i} \ cdot \ sigma _ {j} x_ {i } x_ {j} \ right]} (9)

Ограничения должны быть:

* Минимальная доходность
* Состояние бюджета
* Ограничения на короткие продажи также могут быть приняты во внимание.

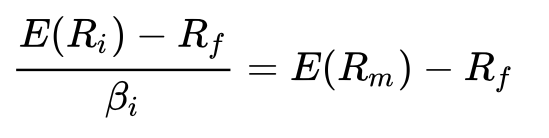
**Комбинация эффективных портфелей**

Например, в случае фондов средств возникает вопрос, приводит ли комбинация эффективных портфелей к эффективному портфелю. Это не должно быть правдой, потому что

* в случае, если короткие продажи не разрешены, линия эффективности ломается. Если, например, создать портфель из двух ценных бумаг в другой части линии, этот портфель больше не будет в строке эффективности.
* в случае если короткие продажи разрешены, короткие продажи эффективного портфеля могут создать неэффективные портфели.

**3) Модель CAPM**

Модель Марковица позволяет выбрать эффективные портфели, тем не менее модель не объясняет суть корреляции между риском и доходностью. Эту проблему решает модель CAPM, которая была разработана У. Шарпом и другими учеными.

 (10)

Где:

* E(Ri) — ожидаемая ставка доходности на долгосрочный актив;
* Rf — безрисковая ставка доходности;
* Bi — коэффициент чувствительности актива к изменениям рыночной доходности Rm, выраженный как ковариация доходности актива Ri с доходностью всего рынка Rm по отношению к дисперсии доходности всего рынка ст. откл
* E(Rm) — ожидаемая доходность рыночного портфеля

В данной модели упрощается модель Марковица ради упрощения вычислений и минимизации усилий за счет введения показателя β. Модель учитывает чувствительность актива к не диверсифицируемому риску (также известному как систематический риск или рыночный риск), часто представляемому в виде бета (β), а также ожидаемую доходность рынка и ожидаемую доходность теоретически безрискового актива. Риск измеряется дисперсией, доходность активов измеряется вероятностным распределением, которая в свою очередь полностью описывается первыми двумя моментами (например, нормальное распределение). Несмотря на неудачу многочисленных эмпирических тестов и наличие более современных подходов к ценообразованию активов и выбору портфеля (таких как арбитражная теория ценообразования и портфельная проблема Мертона), CAPM все еще остается популярным благодаря своей простоте и полезности в различных ситуациях.

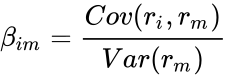
Для отдельных активов используется *линия рынка ценных бумаг (SML),* которая символизирует ожидаемую доходность всех активов рынка в зависимости от не диверсифицируемого риска и его связь с ожидаемой доходностью и систематическим риском (бета), показать, как рынок должен оценивать цену отдельного актива по отношению к классу, к которому он принадлежит.

Линия SML позволяет рассчитать отношение прибыли к риску для любого актива относительно общего рынка.

Соотношение равновесия, которое описывает CAPM:

{\ displaystyle E (r_ {i}) = r_ {f} + \ beta _ {im} (E (r_ {m}) - r_ {f}) \,} (11)

Где:

* {\ displaystyle E (r_ {i})}Ожидаемая норма доходности капитала по активу Я.
* {\ displaystyle \ beta _ {im}}это *бета* (сумма риска по отношению к рыночному портфелю), или 
* {\ displaystyle (\, E (r_ {m}) - r_ {f} \,)} Это премия за рыночный риск, то есть дополнительный доход, который инвесторы требуют инвестировать в рискованные активы, по сравнению с альтернативой инвестирования в безрисковые активы.
* {\ displaystyle (r_ {m})} Рыночная производительность.
* {\ displaystyle (r_ {f})} Доходность безрискового актива: обычно используется доходность облигации со сроком погашения, аналогичным сроку действия финансового актива, подлежащего оценке.

**Модель равновесной цены**

Здесь учитываются – инфляция, экономич. развитие, проц. ставки по кредитам банков и пр, что делает этот метод более обеспеченным, чем Метод Майкла О'Хиггинса и Гарднеров.

После добавления бумаг в портфель выполняется анализ бумаг и выявление набора бумаг, которые менее всего подвержены этим рискам и в совокупности дают приемлемый уровень доходности. Справедливости ради стоит заметить, что предпосылка о возможности учета всех типов риска является ложной, т.к. абсолютно все риски учесть практически невозможно.

**Модель Блэка-Литтермана**

Предполагается, что существует только три актива A, B и C. Превосходство каждого актива определяется рыночным равновесием и возможными неожиданностями, с одной стороны, фактором риска, общим для трех активов, а во-вторых, по трем отдельным факторам риска.

Математически мы могли бы описать модель следующим образом:

* {\ displaystyle R_ {A} \ = \ \ pi _ {A} + \ gamma _ {A} Z + \ nu _ {A}}
* {\ displaystyle R_ {B} \ = \ \ pi _ {B} + \ gamma _ {B} Z + \ nu _ {B}}
* {\ displaystyle R_ {C} \ = \ \ pi _ {C} + \ gamma _ {C} Z + \ nu _ {C}}

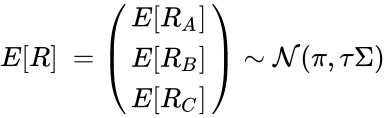
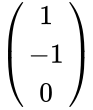
Где R _ {{i}} результат по активу i, {\ displaystyle \ pi _ {i}} это равновесный результат по активу i, {\ displaystyle \ gamma _ {i}}представляет влияние Z общего фактора риска и {\ displaystyle \ nu _ {i}} соответствует независимому случайному событию для актива i. В этой модели риск, связанный с избыточной доходностью активов A, B и C, измеряется ковариационной матрицей, \ Сигма  которая объединяет четыре фактора риска: общий фактор Z и три отдельных фактора, независимых и специфичных для каждого актива. Следует отметить, что невозможно изолировать различные факторы риска от простого наблюдения \ Сигма .

Вектор ожидаемой доходности активов является функцией равновесных премий за риск, математического ожидания общего фактора риска и математического ожидания вектора случайных событий, независимых от каждого актива. Таким образом, ожидание доходности актива А записывается следующим образом:

{\ displaystyle E [R_ {A}] \ = \ \ pi _ {A} + \ gamma _ {A} E [Z] + E [\ nu _ {A}]} (12)

Здесь {\ displaystyle E [R_ {A}]} сама по себе случайная переменная, которая обусловлена ​​ожиданиями по избыточным доходам. Ожидание {\ displaystyle E [R_ {A}]} соответствует премии за равновесный риск, тогда как неопределенность {\ displaystyle E [R_ {A}]} зависит от неопределенности общего фактора {\ displaystyle E [Z]} и отдельных факторов {\ displaystyle E [\ nu _ {i}]}. Black и Litterman сделали *предположение о том, что неопределенность случайных величин {\ displaystyle E [Z]} и {\ displaystyle E [\ nu _ {i}]} пропорциональны нестабильности Zи{\ displaystyle \ nu _ {i}}*. Это предположение подразумевает, что сформированный случайный вектор {\ displaystyle E [R_ {i}]} допускает ковариационную матрицу, пропорциональную \ Сигма , то есть {\ displaystyle \ tau \ Sigma}. Согласно Блэку и Литтерману, значение \ тау  близко к 0, так как согласно их терминам, неопределенность в ожидании случайной величины ниже, чем неопределенность в самой случайной переменной. Но эта ценность является предметом споров. Некоторые авторы иногда используют значение 0 для \ тау таких как Meucci, Krishnan и Mains, делая упрощающее предположение, что нет ошибки оценки в ожидании {\ displaystyle E [R_ {i}]}. Другие, такие как Satchell и Scowcroft, считают, что \ тау  часто равен 1. В качестве альтернативы Meucci оценивает ошибку оценки и, следовательно, \ тау  по количеству доступных данных. Согласно тому же автору, значение \ тау  обычно находитмя между 0 и 1. Тем не менее, Дж. Уолтерс указывает, что некоторые авторы рассматривают возможность значения {\ displaystyle \ tau> 1}8.

Если инвестор желает выразить *100% определенное* ожидание в рамках модели, он может сделать это с помощью линейного ограничения. Например, он предусматривает, что активы превосходят актив B на определенную величину Q, то это приведет к ограничению {\ displaystyle E [R_ {A}] - E [R_ {B}] \ = Q}.

Сделав более сильную гипотезу гауссовского распределения для случайного вектора: становится возможным определить новое распределение случайного вектора условно с учетом ограничения, в {\ displaystyle P ^ {T} \ cdot E [R] \ = Q} котором P находится вектор . Ожидание нового распределения, которое представляет избыточную доходность с учетом ожиданий:

{\ displaystyle {\ overline {E [R]}} \ = \ pi + \ tau \ Sigma \ cdot P \ cdot [P ^ {T} \ cdot \ tau \ Sigma \ cdot P] ^ {- 1} \ cdot [QP ^ {T} \ cdot \ pi]} (13)

Одним из нововведений, внесенных моделью Блэка-Литтермана, является возможность модулирования уровня достоверности ожидания и, следовательно, интеграции более или менее надежных ожиданий.

## Выводы к главе 1

Как можно заметить, методов и предложений по данной тематике очень много, некоторые играют с одной и той же концепцией риск/вознаграждение, другие фокусируются на различных вариациях риска, третьи на особенностях временных рядов и их метрик. Наиболее внушительным методом мне представляется модель CAPM за счет своего упрощения и усиления уважаемой модели Марковица за счет показателя бетта.

# **Глава 2 Реализация алгоритма, комбинирующего PCA, k-means и итеративный алгоритм**

## ***2.1 Отрасли финансовых рынков***

Я хотел бы поподробнее разобрать отрасли, которые использовал в этой работе. Финансовая отрасль состоит, в частности, из банков, инвестиционных фондов, страховых компаний и фирм по недвижимости. В целом, большая часть доходов, получаемых сектором, поступает от ипотеки и кредитов, которые приобретают стоимость по мере роста процентных ставок. Коммунальные предприятия состоят из электро-, газо- и водоснабжающих компаний, а также интегрированных провайдеров. В целом, сектор генерирует стабильный регулярный доход, взимая с потребителей и предприятий, которые обеспечивают более высокую, чем в среднем, дивидендную доходность. Потребительский сектор состоит из розничных продавцов, медиа-компаний, поставщиков потребительских услуг, швейных компаний и потребительских товаров длительного пользования. В целом, эти компании выигрывают от улучшения экономики, когда потребительские расходы ускоряются. Потребительский сектор состоит из компаний, занимающихся продуктами питания и напитками, а также компаний, создающих потребительские товары, которые не желают сокращать свои бюджеты. В целом, эти компании являются оборонительными предприятиями, способными противостоять экономическому спаду. Энергетика состоит из компаний, занимающихся разведкой и добычей нефти и газа, а также интегрированных энергетических компаний, нефтеперерабатывающих заводов и других предприятий. Как правило, эти компании получают доходы, привязанные к ценам на сырую нефть, природный газ и другие сырьевые товары. Здравоохранение состоит из биотехнологических компаний, компаний по управлению больницами, производителей медицинского оборудования и многих других. В целом, сектор считается одновременно и возможностью роста, и оборонительной игрой, поскольку люди всегда будут нуждаться в медицинской помощи. Промышленность состоит из аэрокосмических, оборонных, машиностроительных, строительных, производственных компаний. В целом, рост отрасли обусловлен спросом на строительные и промышленные товары, такие как сельскохозяйственное оборудование. Технологии состоят из производителей электроники, разработчиков программного обеспечения и фирм, занимающихся информационными технологиями. В целом, эти предприятия стимулируются циклами модернизации и общим состоянием экономики, хотя рост был устойчивым на протяжении многих лет. Телеком состоит из провайдеров беспроводной связи, кабельных компаний, провайдеров интернет-услуг и спутниковых компаний, в том числе. В целом, эти компании получают периодический доход от потребителей, но некоторые подгруппы отрасли сталкиваются с быстрыми изменениями. Материалы состоят из горнодобывающих, перерабатывающих, химических, лесных и смежных компаний, которые сосредоточены на обнаружении и разработке сырья. Поскольку эти компании находятся в начале цепочки поставок, они подвержены изменениям в бизнес-цикле. Недвижимость состоит из компаний, инвестирующих в жилую, промышленную и розничную недвижимость. Основным источником дохода для этих компаний является арендный доход и прирост капитала в сфере недвижимости. В результате, этот сектор чувствителен к изменениям процентных ставок.

## ***2.2 Концепция алгоритма***

Изображение выглядит как снимок экрана

Автоматически созданное описание

*Рисунок 1. Концепция алгоритма*

Алгоритм начинается со скачивания данных по фундаментальным метрикам и динамике тикеров. Далее фильтрация по наличию актуальной информации по тикерам. Далее преобразование большого количества нормализованных фундаментальных метрик по методу главных компонент. Далее, кластеризация по pca факторам. Затем, выбор наилучших (по доходности) кластеров. Генерация подмножеств, итеративный алгоритм. Получаем наилучший портфель и делаем прогноз значений по SARIMA.

## ***2.3 Сбор данных и алгоритм***

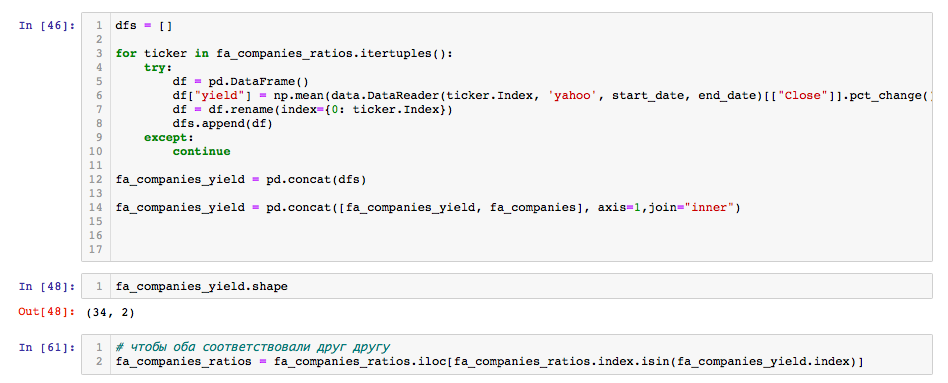
Вспомогательные библиотеки python, которые я использовал в работе можно найти в приложениях, приложение 2.

Доступные бумаги были скачаны с помощью python библиотеки

FundamentalAnalysis. Мною был взят sample из 200 тикеров. Затем данные тикеры были соотнесены с доступной информацией за последний год (период анализа). Также были скачаны данные с yahoo.finance по динамике котировок.



*Рисунок 2. Сбор данных*



*Рисунок 3. Сбор данных*

**PCA, principal component analysis (метод главных компонент).**

При наборе двух, трёх или более точек и высоких размерностях, прямая может быть определена как best-fit, которая минимизирует среднее квадратное расстояние от точки до линии. Следующая best-fit прямая может быть аналогично выбрана из направлений перпендикулярных первому. Повторение этого процесса дает ортогональную основу, в которой различные индивидуальные размеры данных не связаны между собой. Эти базисные векторы называются главными компонентами (PCA).

PCA в основном используется в качестве инструмента в исследовательском анализе данных и для создания прогностических моделей. Он часто используется для визуализации генетической дистанции и родства между популяциями.

Выполняется в два этапа:

* Вычисление ковариационной (или корреляционной) матрицы исходных данных.
* Выполнение стоимостного разложения на ковариационной матрице или сингулярным разложением матрицы проектирования.

Обычно исходные данные нормализуются перед выполнением PCA. Нормализация каждого атрибута состоит из вычитания среднего по центру – вычитания каждого значения данных из измеренного среднего значения его переменной таким образом, чтобы его эмпирическое среднее (среднее) было равно нулю. Некоторые поля, помимо нормализации среднего, делают это для дисперсии каждой переменной (чтобы она равнялась 1). Результаты PCA обычно обсуждаются в терминах оценки компонентов, иногда называемых оценками факторов (значения трансформированных переменных, соответствующие определенной точке данных), и нагрузок (вес, на который каждая стандартизированная исходная переменная должна быть умножена, чтобы получить оценку компонента). Если баллы за компоненты стандартизированы до единичной дисперсии, то нагрузки должны содержать дисперсию данных в них (величина собственных значений). Если баллы по компонентам не стандартизированы, то нагрузки должны быть единичными ("нормированными"), и эти веса называются собственными векторами; они являются косинусами ортогонального вращения переменных на основные компоненты или обратно.

Предположим, есть образец *п* экземпляров, каждый из которых измеряет *м* переменных (случайным образом) {\ displaystyle F_ {j}. \;}PCA может найти ряд базовых факторов *р* < *м* , объясняющий стоимости *м* переменных для каждого признака. Тот факт, что эти *p* лежат в основе факторов, может быть интерпретирован как уменьшение размерности данных: где раньше нам требовалось *m* значений для характеристики каждого отдельного человека, теперь *p* значений достаточно. Каждый из найденных *p* называется основным компонентом отсюда и название метода.

Существует два основных способа применения PCA:

1. Метод на основе матрицы корреляции, когда данные не являются однородными по размерам или порядок измеренных случайных величин не одинаков.
2. Метод на основе матрицы ковариации, используемый, когда данные однородны по размерам и имеют аналогичные средние значения.

**Корреляционный метод**

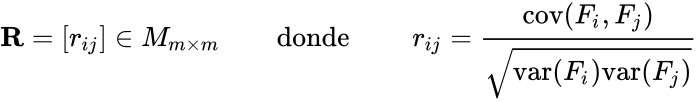
Метод начинается с корреляционной матрицы, давайте рассмотрим значение каждой из *m* случайных величин {\ displaystyle F_ {j} \,}. Для каждого из *n* наблюдений. Давайте возьмем значение этих переменных и запишем набор данных в матричной форме:

{\ displaystyle (F_ {j} ^ {\ beta}) _ {j = 1, ..., m} ^ {\ beta = 1, ..., n}} (14)

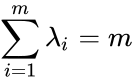
Обратите внимание, что каждый набор:

{\ displaystyle {\ mathcal {M}} _ {j} = \ {F_ {j} ^ {\ beta} | \ beta = 1, ..., n \}} (15)

можно считать случайной выборкой для переменной {\ displaystyle F_ {j} \,}. Из {\ displaystyle m \ times n} данных, соответствующих *m* случайным переменным, можно построить выборочную матрицу корреляции, которая определяется следующим образом:

 (16)

Поскольку матрица корреляции симметрична, то она диагонализируема и ее собственные значения \ лямбда _ {я} \, проверяют:



(17)

Основные факторы математически представлены в основании собственных векторов матрицы \ mathbf {R}. Понятно, что каждая из переменных может быть выражена как линейная комбинация собственных векторов или главных компонент.

**Ковариантный метод**

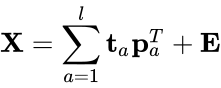
Цель состоит в том, чтобы преобразовать данный набор данных **X** измерения *nxm* в другой набор данных **Y** меньшего измерения *nxl* с наименьшей возможной потерей полезной информации с использованием ковариационной матрицы.

Он основан на наборе выборок *n,* каждая из которых имеет *m* переменных, которые их описывают, и цель состоит в том, чтобы каждая из этих выборок описывалась только с помощью переменных *I* , где *l* < *m*. Кроме того, число главных компонентов *l* должно быть меньше наименьшего из размеров X.

{\ displaystyle l \ leq \ min \ {n, m \}}

(18)

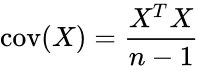
Данные для анализа должны быть центрированы по среднему значению 0 (вычитая среднее значение из каждого столбца) и / или автоматически масштабированы (центрированы по среднему значению 0 и, разделив каждый столбец на его стандартное отклонение).



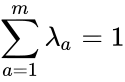
(19)

Векторы {\ displaystyle \ scriptstyle \ mathbf {t} _ {a}} известны как *оценки* и содержат информацию о том, как образцы связаны друг с другом, и они имеют свойство быть ортогональными. Векторы {\ displaystyle \ scriptstyle \ mathbf {p} _ {a}}называются *нагрузками* и сообщают о взаимосвязи между переменными и имеют качество ортонормированности. Сбор меньшего числа основных компонентов, чем переменных, и из-за ошибки подбора модели с данными приводит к ошибке, которая накапливается в матрице {\ displaystyle \ scriptstyle \ mathbf {E}}.

PCA основана на разложении собственных векторов ковариационной матрицы, который рассчитывается по следующему уравнению:



{\ displaystyle {\ mbox {cov}} (X) \ {\ mathbf {p} _ {a}} = {\ lambda_ {a}} \ {\ mathbf {p} _ {a}}} (20)



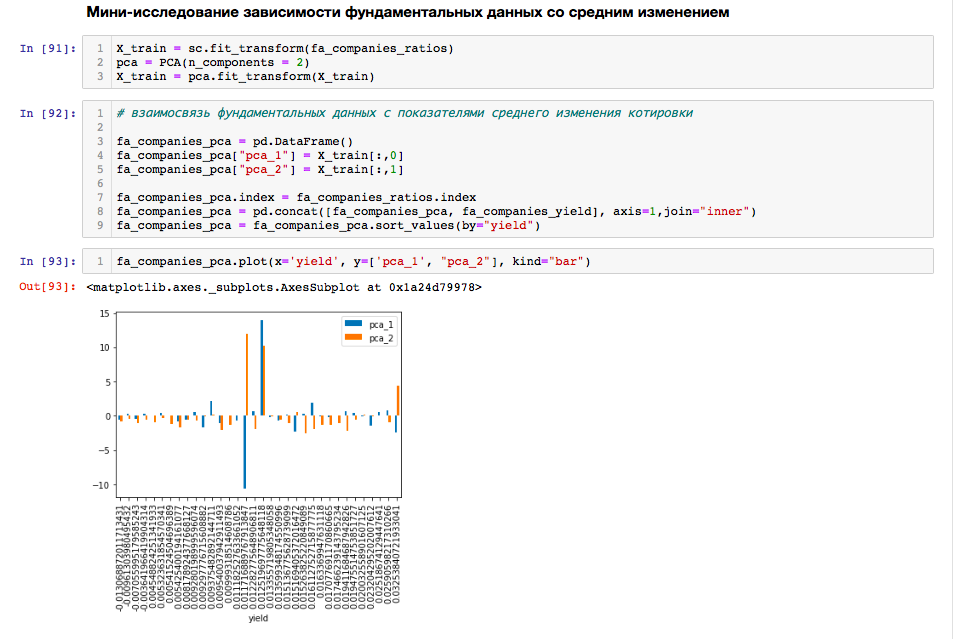
Где {\ displaystyle \ scriptstyle \ lambda _ {a}}это собственное значение, связанное с собственным вектором {\ displaystyle \ scriptstyle \ mathbf {p} _ {a}}. И, наконец,

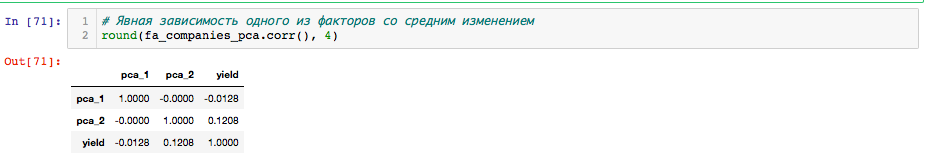
{\ displaystyle \ mathbf {t} _ {a} = X \ \ mathbf {p} _ {a}} (21)

Мы можем понимать это уравнение как {\ displaystyle \ scriptstyle \ mathbf {t} _ {a}}проекции X в {\ displaystyle \ scriptstyle \ mathbf {p} _ {a}}, где собственные значения {\ displaystyle \ scriptstyle \ lambda _ {a}} измеряют количество захваченной дисперсии, то есть информацию, которую представляет каждый из основных компонентов. Количество информации, которую захватывает каждый основной компонент, уменьшается в соответствии с их количеством, то есть основной компонент номер один представляет больше информации, чем два, и так далее.

Реализация PCA ограничена в ряде из предположений

* Предположение о линейности. Предполагается, что наблюдаемые данные являются линейной комбинацией определенной базы.





*Рисунок 4. PCA, метод главных компонент*

Небольшое исследование показало, что результаты PCA не коррелируют с доходностью. Возможно нужно выделить большее число факторов для выявления связи, т.к. из базового предположения должны быть факторы, которые положительно влияют на доходность, в то время как другие влияют отрицательно, но это не так важно.

**Кластеризация K-means**

Наиболее распространенный алгоритм использует метод итеративного уточнения. Из-за своей повсеместности его часто называют **алгоритмом *k-*средних**, он также известен как **алгоритм Ллойда**, особенно в компьютерном сообществе.

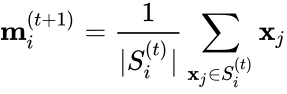
Учитывая начальный набор *K* центроиды **м** 1 (1) , ..., **т** *к* (1)

Шаг 1, **назначения**: назначьте каждому наблюдению группу с ближайшим средним значением (то есть разделением наблюдений согласно диаграмме Вороного, сгенерированной центроидами).

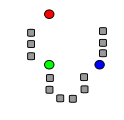
S_i ^ {(t)} = \ big \ {x_p: \ big \ |  x_p - m ^ {(t)} _ i \ big \ |  \ le \ big \ |  x_p - m ^ {(t)} _ j \ big \ |  \ \ forall \ 1 \ le j \ le k \ big \} (22)

Где каждый x_ {p}идет точно внутри одного S ^ {(t)} _ i, даже если он может войти в два из них.

Шаг **обновления**: Рассчитайте новые центроиды как центроиды наблюдений в группе.

 (23)

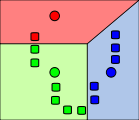
Считается, что алгоритм сходится, когда назначения больше не меняются. Обычно используются методы инициализации подделок и случайных разделов. Метод Forgy случайным образом выбирает *k* наблюдений из набора данных и использует их в качестве начальных центроидов. Метод случайного разделения сначала случайным образом назначает кластер для каждого наблюдения, а затем переходит к этапу обновления, таким образом, вычисляя начальный кластер как центр тяжести рандомизированных точек в кластере. Метод Форги имеет тенденцию рассеивать начальные центроиды, в то время как случайное разбиение определяет расположение центроидов вблизи центра набора данных. Для максимизации ожидания и стандартного алгоритма предпочтителен метод Форги.

****

*Рисунок 5. Кластеризация k-means*

1) *k* начальных центроидов (в данном случае *k* = 3) генерируются

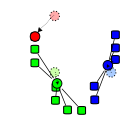
случайным образом в наборе данных (показано цветом).



*Рисунок 6. Кластеризация k-means*

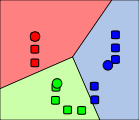
2) *k* групп генерируются путем связывания точки с ближайшим

средним. Разделение здесь представляет диаграмму Вороного, созданную центроидами.



*Рисунок 6. Кластеризация k-means*

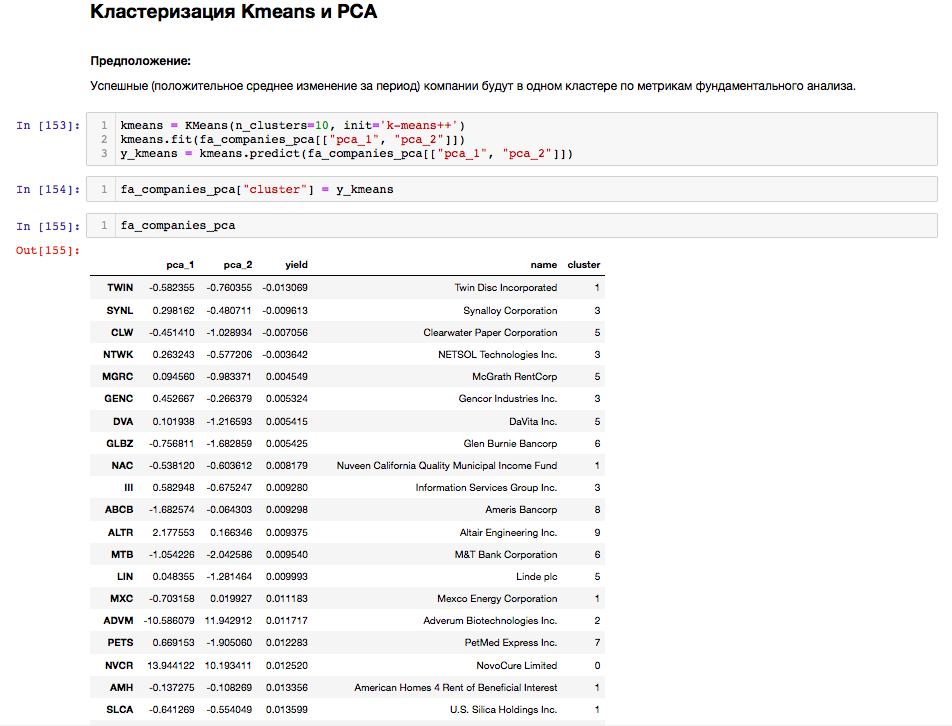
3) Центроид каждой из *k* групп пересчитывается.

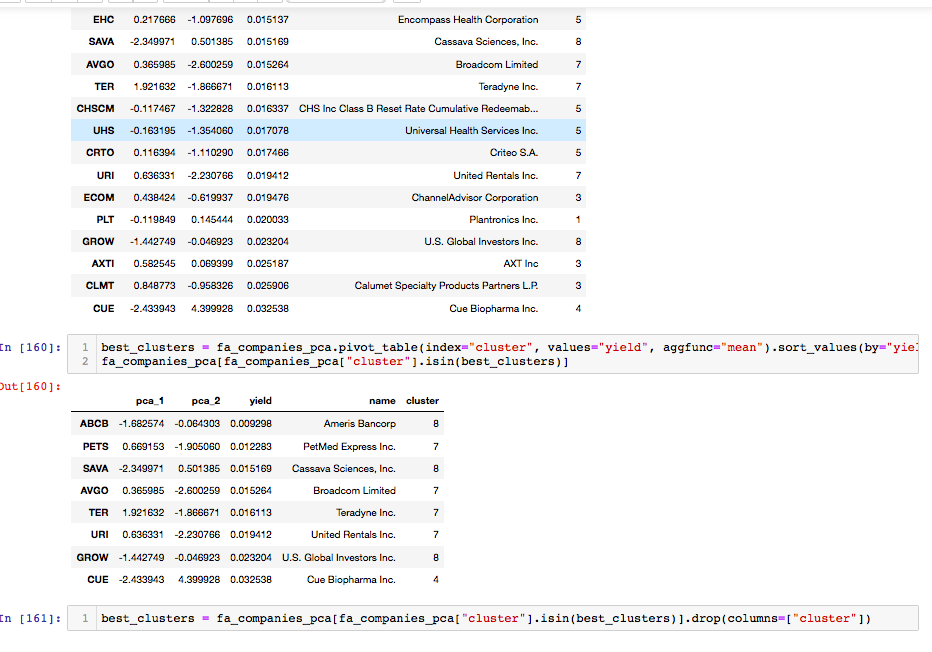


*Рисунок 7. Кластеризация k-means*

4) Шаги 2 и 3 повторяются до тех пор, пока не будет достигнута конвергенция.

Поскольку это эвристический алгоритм, нет гарантии, что они сходятся к глобальному оптимуму, и результат может зависеть от исходных групп. Поскольку алгоритм обычно очень быстрый, обычно запускается несколько раз с разными начальными условиями. Однако в наихудшем случае *k-средние значения* могут очень медленно сходиться: в частности, было показано, что существуют наборы определенных точек, даже в двух измерениях, в которых *k-средние значения* требуют экспоненциального времени, то есть 2 O ( n ) , чтобы сходиться. 9 Эти наборы точек, по-видимому, не возникают на практике: это подтверждается тем фактом, что в большинстве случаев время выполнения *k-означает* многочлен.

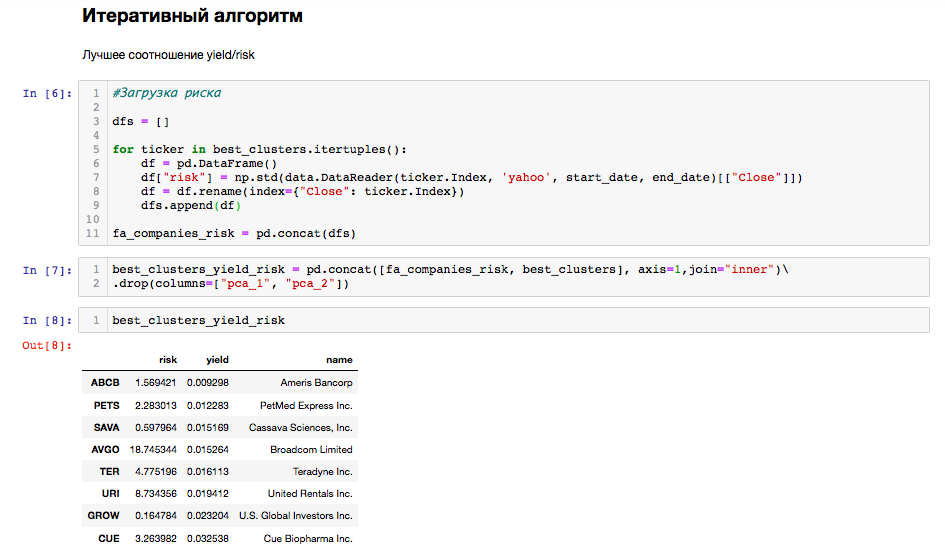




*Рисунок 8. Данные для k-means и наилучшие 3 кластера*

Далее тикеры были объединены в кластеры по похожей структуре фундаментальных метрик. Данная кластеризация проводится с целью нахождения наиболее “сильных” кластеров, которые будут обладать наибольшей доходностью.

**Итеративный алгоритм**





*Рисунок 9. Итеративный алгоритм.*



*Рисунок 10. Наилучший портфель.*

Распределение по тикерам поровну из наивного предположения и игнорирования вопроса внутренней оптимизации распределения активов. Стоит повториться, целью данной работы не является оптимизации портфеля, а сбор портфеля в условиях полной неизвестности. Доходность полученного портфеля URI, GROW, CUE тикеров в портфеле 2.5% в валюте, что неплохо учитывая современные реалии и гарантированный доход (по депозиту) 0.5%.

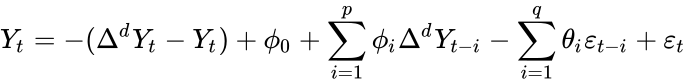
**Прогнозирование (S)ARIMA**

Модель ARIMA должна идентифицировать коэффициенты и количество регрессий, которые будут использоваться. Эта модель очень чувствительна к точности, с которой определяются ее коэффициенты.

Обычно это выражается как ARIMA (*p*, *d*, *q*), где параметры *p*, *d* и *q* являются неотрицательными целыми числами, которые указывают порядок различных компонентов модели – соответственно, авторегрессионных , интегрированных и скользящих средних . Когда любой из трех параметров равен нулю, обычно пропускаются соответствующие буквы аббревиатуры - AR для авторегрессионного компонента, I для интегрированного и MA для скользящего среднего. Например, ARIMA (0,1,0) можно выразить как I (1), а ARIMA (0,0,1) - как MA (1).

Модель ARIMA может быть дополнительно обобщена для учета влияния сезонности. В данном случае речь идет о модели SARIMA (*сезонная авторегрессия, интегрированная скользящая средняя*).

Модель ARIMA (p, d, q) может быть представлена ​​как:

 (24)

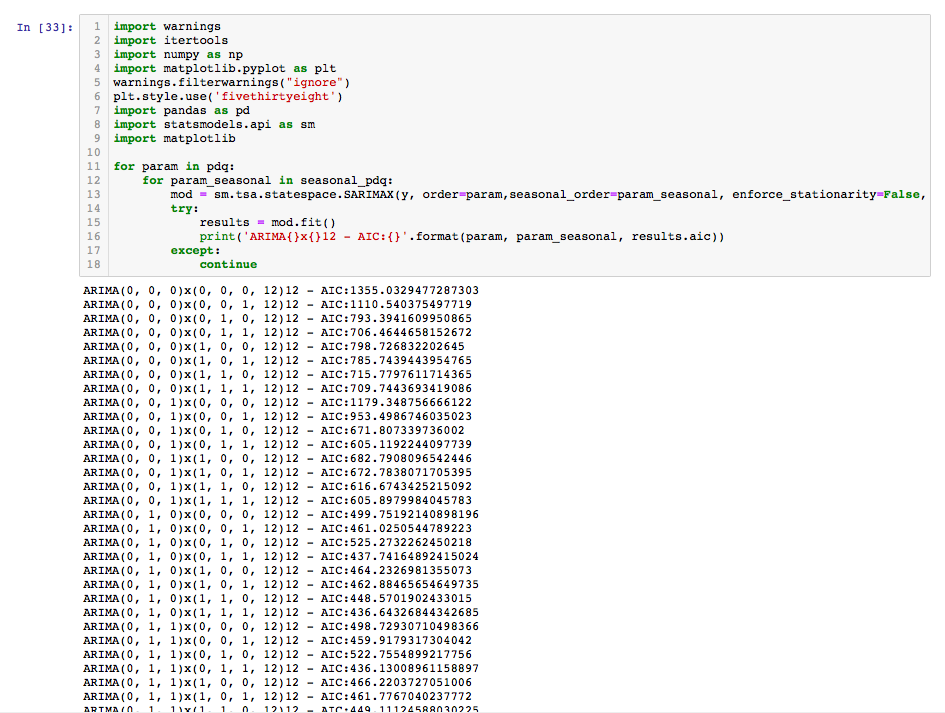
Где *d* соответствует разностям d, которые необходимы для преобразования исходного ряда в стационарный, {\ displaystyle \ phi _ {1}, \ ldots, \ phi _ {p}}- это параметры, относящиеся к « авторегрессионной » части модели, {\ displaystyle \ theta _ {1}, \ ldots, \ theta _ {q}} параметры, относящиеся к части « скользящих средних » модели, это {\ displaystyle \ phi _ {0}}константа, и {\ displaystyle \ varepsilon _ {t}}это термин ошибки (также называемый *новшеством* или *стохастическим возмущением,* последний больше ассоциируется для эконометрических моделей с одним или несколькими уравнениями).

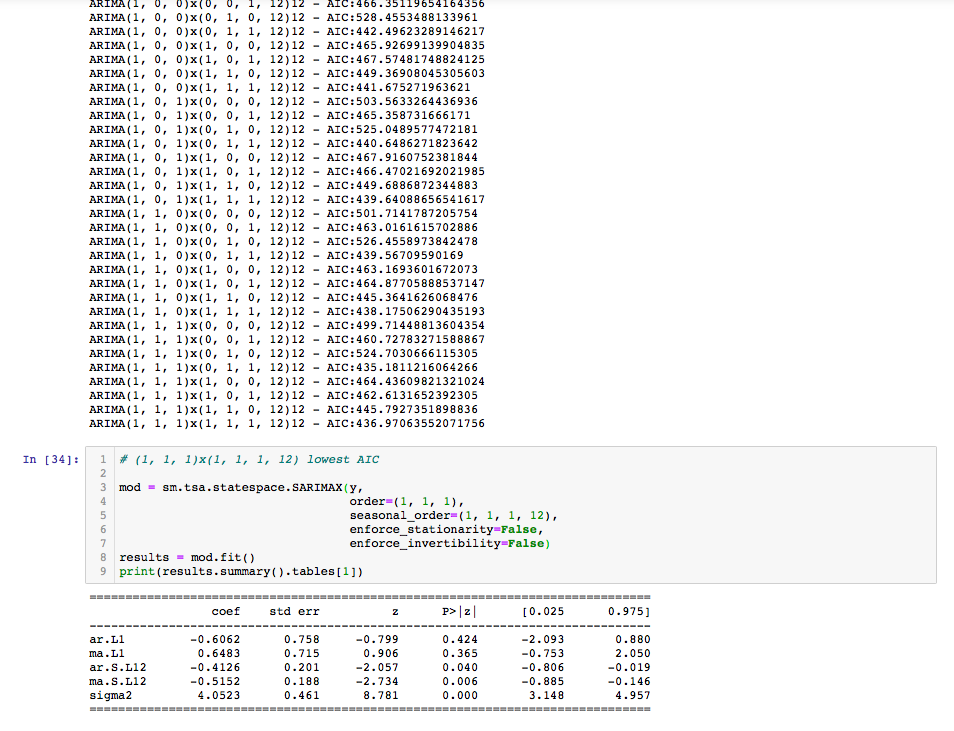
Следует учитывать, что:

{\ displaystyle \ Delta Y_ {t} = Y_ {t} -Y_ {t-1}}} (25)

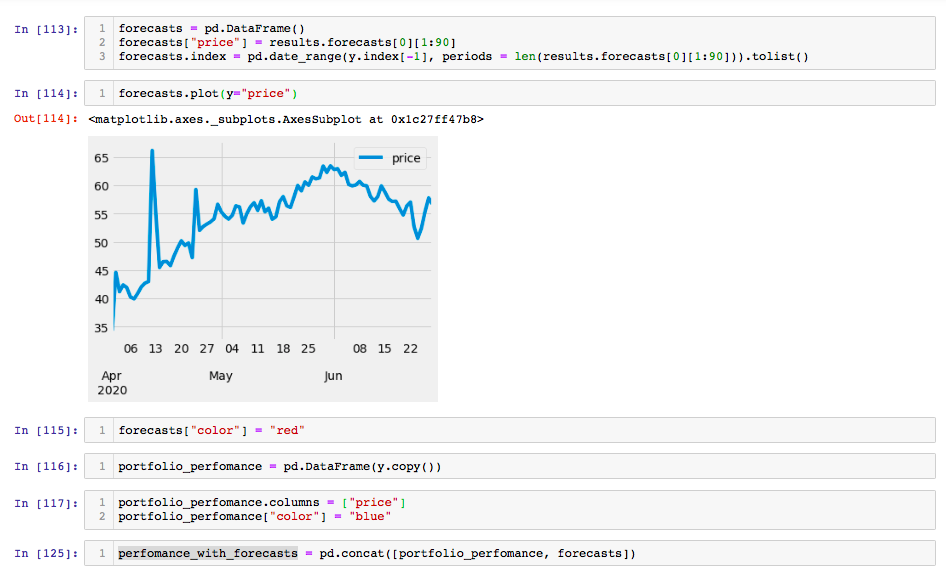


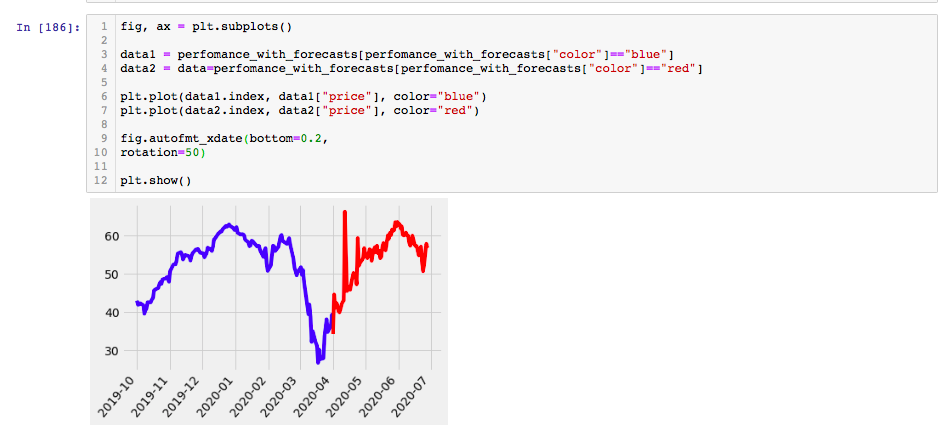
*Рисунок 11. Прогноз SARIMA*





*Рисунок 12. Прогноз SARIMA*





*Рисунок 13 Прогноз SARIMA*

Прогноз был сделан с помощью SARIMA, т.к. он учитывает сезонную составляющую (в отличие от ARIMA), а также совмещает сильные стороны AR и MA методов. Набор p,d,q исходя из наименьшего AIC. Был сделан прогноз на 3 месяца вперед. Как мы видим, в ближайшее время ожидается подъем и установка определенного ценового диапазона.

## ***2.4 Сравнение доходностей***

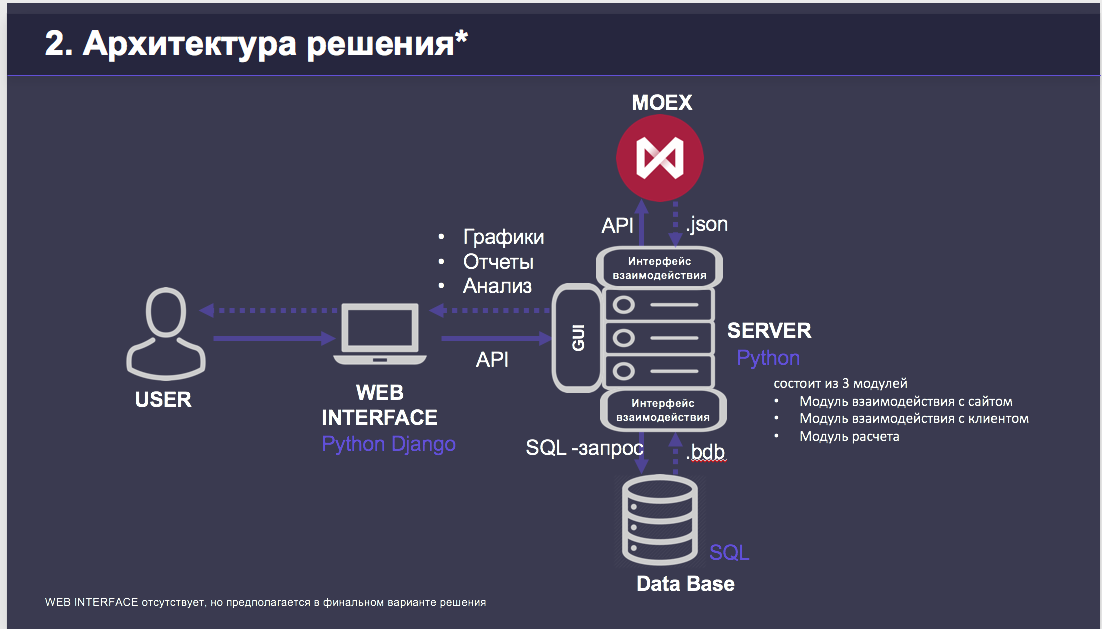
10-летние трежериз (03.05.2020, 19:33:00 москва) купон 1.5%, наш портфель 2.5%, есть смысл сформировать портфель. Доходность по вкладу в валюте составляет 1-1.5%, что опять-таки ниже. Помимо прочего, наш портфель также будет приносить дивиденды.

## **Выводы к главе 2**

В разработанном мною алгоритме есть 2 основных момента, где можно испробовать различные вариации. Первое – это метод уменьшения размерности, вместо PCA можно было бы попробовать Факторный анализ или LDA. На этапе кластерного анализа можно попробовать иерархический кластерный анализ (дендограмма) или DBSCAN. Основное преимущество данного алгоритма — это легкая интерпретируемость самой последовательности действий, совмещение наиболее сильных сторон машинного обучения и итеративного алгоритма.

# Глава 3 Бизнес-применение итеративного алгоритма

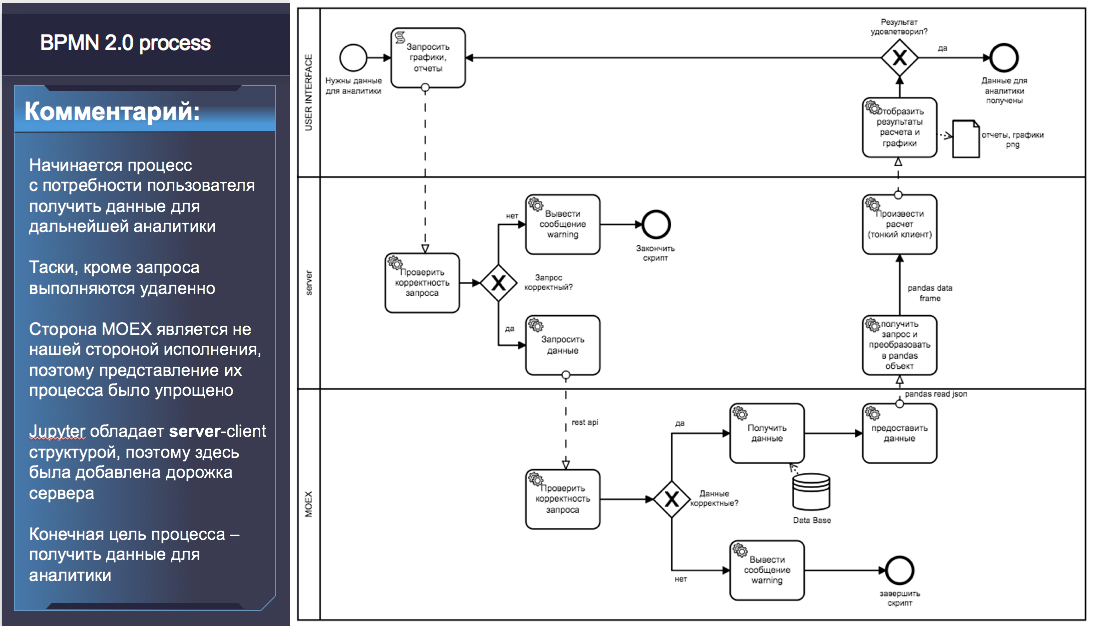
## **3.1 Архитектура решения**



*Рисунок 13. Функциональное представление решения*

В данной модели мы видим, как user делает запрос через web-interface, делая api-запрос к серверу, который в свою очередь делает свой запрос в базу данных, откуда берет данные о котировках. База данных реализована с помощью SQL Oracle. Сервер состоит из модулей – взаимодействия с сайтом, модуль взаимодействия с клиентом, модуля расчета. Общение с клиентом предполагается через web-interface. Исполнение сайта рекомендуется на python Django, из-за своей скорости создания, сильной комплектацией функционала, безопасности и масштабируемости по сравнению с другими языками.

## **3.2 BPMN 2.0 модель**



*Рисунок 14. BMPN 2.0 модель*

Начальное событие – это потребность клиента в аналитике для сбора портфеля. Финальная цель данного процесса является получение информации о портфеле, который стоит собрать клиенту.

Данный алгоритм рассчитан как на физических лиц, так и юридических. Идеальная ситуация, это средний бизнес, который будет инвестировать в рынок на часть заработанных средств.

Также вполне уместно в будущих итерациях подумать о внедрении данного алгоритма в чат-бот, добавить голосового помощника для дополнительных функций (узнать информацию о финансовых рынках, индексах, бизнес-новостях и т.д.).

Данный продукт, по моей задумке, рассчитан на малый и средний бизнес. Мне представляется, что у компаний малого и среднего масштаба должна быть выработана дисциплина откладывания небольшой части денежной средств для инвестирования в финансовые рынки для извлечения дополнительной прибыли, для минимизации затрат по времени и создан этот алгоритм. Достаточно быстрое, примерно за 15-20 минут можно создать достаточно прочный и хороший по доходности “в любую погоду” портфель для компании. Тот, кто предоставляет услугу может за дополнительную плату обеспечивать техническое сопровождение, дать доступ к дополнительным возможностям алгоритма, большей вариации моделей для достижения наилучшего эффекта. Можно работать по схеме управленца активами и брать маленький процент с прибыли (0.1%).

## **Выводы к главе 3**

Возможностей для улучшений в бизнес-среде достаточно много (чат-бот, голосовой помощник). Добавить различные вариации моделей. Тем не менее, я бы рекомендовал на скорости реализации и создания mvp модели для скорейшей отладки и тестирования.

# 

# Заключение

Автоматизация процесса формирования портфеля ценных бумагс использованием итерационного алгоритма и машинного обучения способствует повышению эффективности (высокий доход, низкий риск) портфеля по сравнению с другими инвестиционными стратегиями.

**Вывод по гипотезе 1:**

Действительно, применение комбинации ML и итеративного алгоритма позволило получить достаточно удачный по соотношению риска и доходности портфель по сравнению с наивным подходом инвестирования лишь в 1 актив, получая хорошую диверсификацию риска и лучшие показатели доходности и риска.

Использование машинного обучения и итерационного алгоритма соответствует реалиям фондового рынка и способствует оперативному реагированию на колебания рынка.

**Вывод по гипотезе 2:**

В самом деле, сам алгоритм занимает достаточно малое время и выполняется в среднем за 15-20 минут вместе с загрузкой данных, что говорит о возможности быстрого реагирования в условиях текущего рынка.

**СПИСОК ИСТОЧНИКОВ**

1. Котировки акций [Электронный ресурс] // Платформа Investing.com. URL: https://ru.investing.com/equities/
2. Bloomberg Terminal [Электронный ресурс] // Bloomberg Professional Services. URL: https://www.bloomberg.com/professional/solution/bloomberg-terminal/
3. Understanding the GitHub flow [Электронный ресурс] // GitHub Guides. URL https://guides.github.com/introduction/flow/
4. Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1), pp. 77-91. Retrieved from [https://onlinelibrary.wiley.com/doi/abs/10.1111](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x)
5. [/j.1540-6261.1952.tb01525.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x)
6. Ren, Z. (2005). *Portfolio construction using clustering methods* (Doctoral dissertation, Worcester Polytechnic Institute). Retrieved from https://web.wpi.edu/Pubs/ETD/Available/etd-042605-092010
7. /unrestricted/ZhiweiRen.pdf
8. Marvin, K. (2015). *Creating Diversified Portfolios Using Cluster Analysis*. USA: Princeton University. Retrieved from https://www.cs.princeton.edu/sites/default/files/uploads
9. Алгоритм k средних (k-means) [Электронный ресурс] // AlgoWiki. URL: http://algowiki-project.org/ru/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC\_k\_%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D0%B8%D1%85\_(k-means)
10. Дюран, Б. Кластерный анализ / Б. Дюран, П. Оделл ; пер. с англ. Е.З. Демиденко под ред. А. Я. Боярского. — М.: Изд-во «Статистика», 1977, 128 с.
11. Кластеризация: метод k-средних [Электронный ресурс] // глобальный интеллектуальный Портал statistica.ru. URL: http://statistica.ru/theory/klasterizatsiya-metod-k-srednikh/
12. Tutorial: How to determine the optimal number of clusters for k-means clustering [Электронный ресурс] // Cambridge Spark. URL: https://blog.cambridgespark.com/how-to-determine-the-optimal-number-of-clusters-for-k-means-clustering-14f27070048f
13. Метод главных компонент [Электронный ресурс] // Электронное пособие по прикладной статистике. URL: http://mystatbook.narod.ru/pc\_main.pdf (дата обращения: 15.03.2019)
14. An Approach to Choosing the Number of Components in a Principal Component Analysis [Электронный ресурс] // Towards Data Science. URL: https://towardsdatascience.com/an-approach-to-choosing-the-number-of-components-in-a-principal-component-analysis-pca-3b9f3d6e73fe
15. Диверсификация портфеля ценных бумаг [Электронный ресурс] // Инвесторов.Нет. URL: http://investorov.net/permanent/diversifikaciya-portfelya-cennyh-bumag
16. Поиск оптимального портфеля [Электронный ресурс] // Электронный учебник «ОЦЕНКА ФИНАНСОВЫХ ИНВЕСТИЦИЙ». URL: http://www.math.mrsu.ru/text/courses/invest2/2/2.3.4.html
17. Васнев, С.А. Статистика: учебное пособие / С.А. Васнев. — М.: МГУП, 2001. — 170 с.
18. Показатели вариации (колеблемости) признака [Электронный ресурс] // ВГУЭС: сайт цифровых учебно-методических материалов. URL: https://abc.vvsu.ru/books/pr\_stat1/page0010.asp
19. Что такое мультипликаторы и как ими пользоваться [Электронный ресурс] // Т—Ж: журнал про ваши деньги. URL: https://journal.tinkoff.ru/multilplicator/
20. Equity Screener [Электронный ресурс] // Yahoo Finance. URL: https://finance.yahoo.com/screener/new (дата обращения: 30.01.2019)
21. Оптимизация портфеля ценных бумаг средствами Python [Электронный ресурс] // Хабр. URL: https://habr.com/ru/post/341992/ (дата обращения: 1.05.2019)

**ПРИЛОЖЕНИЕ 1. ОТРАСЛИ ИНВЕСТИРОВАНИЯ ПО МОДЕЛИ YAHOO FINANCE EQUITY SCREENER**

|  |  |
| --- | --- |
| Отрасль (рус.) | Отрасль (англ.) |
| Basic Materials | Добывающая промышленность и металлургия |
| Utilities | Коммунальные услуги |
| Financial Services | Финансовые сервисы |
| Consumer Defensive | Потребительские товары краткосрочного пользования |
| Energy | Энергетика |
| Technology | Технологическая отрасль |
| Consumer Cyclical | Потребительские товары длительного пользования |
| Real Estate | Недвижимость |
| Communication Services | Услуги связи |
| Industrials | Промышленность |

**ПРИЛОЖЕНИЕ 2. БИБЛИОТЕКИ, ИСПОЛЬЗУЕМЫЕ В РАБОТЕ**

Изображение выглядит как снимок экрана

Автоматически созданное описание