«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Реферат по теме: «Предиктивная модель оценки справедливой цены акций российских компаний на Московской бирже в разрезе методов машинного обучения»

Выполнил

Студент 1 курса

группы 09-115(3)

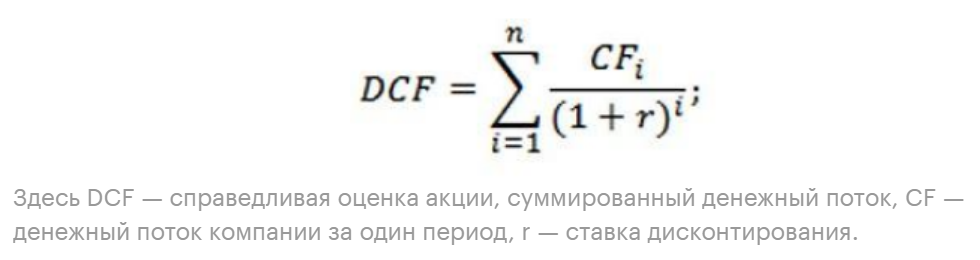
Зиновьев Е. А.

Казань 2021

**Введение**

В настоящее время существует большое количество методов для оценки справедливой цены акций компаний на фондовой бирже. Данное учение заложил Грэхем в 1920 – 1930-е года для нахождения недооцененных акций. Грэм рекомендовал выделять компании, акции которых котируются ниже их внутренней стоимости: торгуемые ниже балансовой стоимости, имеют высокую дивидендную доходность, низкий показатель цена/прибыль (P/E ratio) или коэффициент P/B и которые имеют относительно низкий уровень долга. Акции таких компаний Грэхем считал недооценёнными. То есть для анализа компании необходимо было из ее финансовой отчетности рассчитывать коэффициенты, описанные Грэхемом, и сравнивать с его рекомендациями и средними коэффициентами по рынку. [1]

Еще одним популярным методом было дисконтирование будущих денежных потоков, при котором размер будущих денежных потоков приводят к их справедливой величине в настоящем с использованием ставки дисконтирования, которая является ничем иным как требуемой доходностью или ценой капитала.  
То есть по модели DCF надо привести финансовые показатели к будущим значениям (продисконтировать) для вычисления денежных потоков, на основании которых уже высчитывается справедливая цена акции [2]:



При этих 2 методах учитывались только параметры, относящиеся только конкретно к этой компании (то есть ее фин.показатели), что может быть недостаточно для более широкой оценки рынка в целом, который может сильно повлиять на эту анализируемую компанию и исказить результаты этих 2 методов. Но для оценки такого большого кол-ва параметров нужны гигантские вычислительные мощности для применения классических статистических методов – регрессии, кластеризации и др. Современным методом решения такой проблемы является применение машинного обучения (нейронных сетей, деревьев решений и др.), которые способны сами обучаться на основании предоставленных данных и находить зависимости в них для обобщения на неизвестные случаи.

Оценка справедливой цены акций является актуальным вопросом в сфере инвестирования. Нерациональное поведение людей порождает неопределенность в оценке акций компании на фондовом рынке, но сам бизнес слабо реагирует на такие колебания цен. В связи с этим при положительном росте компании на фондовом рынке настроение у акционеров также, как правило, становится положительным к этой акции. Поэтому несмотря на дневные/недельные/месячные колебания цен акций, они обычно возвращаются к своим справедливым значениям, оцененным по показателям бизнеса.

Оглавление

**Исходные данные4**

**Постановка задачи6**

**Построение дерева7**

**Построение нейронной сети12**

**Пожелания и возможные усовершенствования модели17**

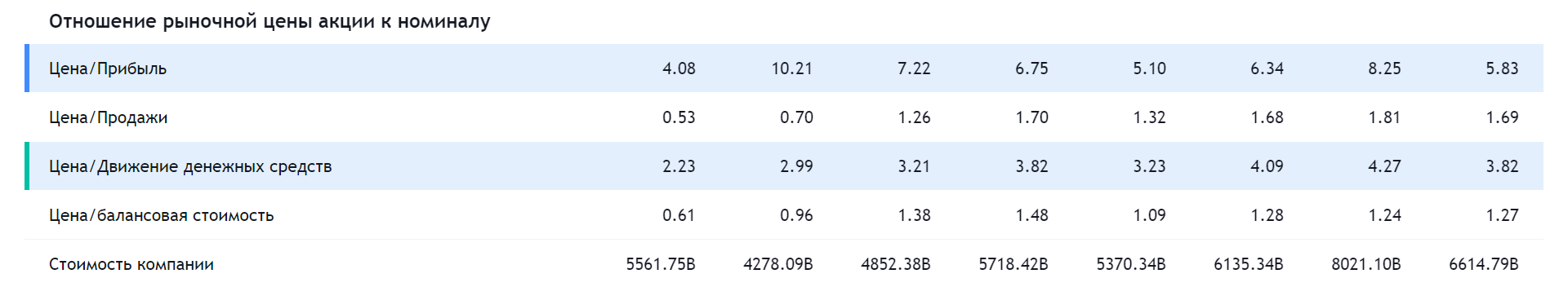
**Заключение19**

**Список использованной литературы20**

**Приложение21**

**Исходные данные**

Финансовые показатели компаний возьмем из консолидирующей инвестиционной площадки [3]. Там уже собраны и вычислены все необходимые коэффициенты. Осталось только создать бота, который бы прошелся по всем тикерам рос. компаний и собрал подобные данные воедино (рисунок 1):

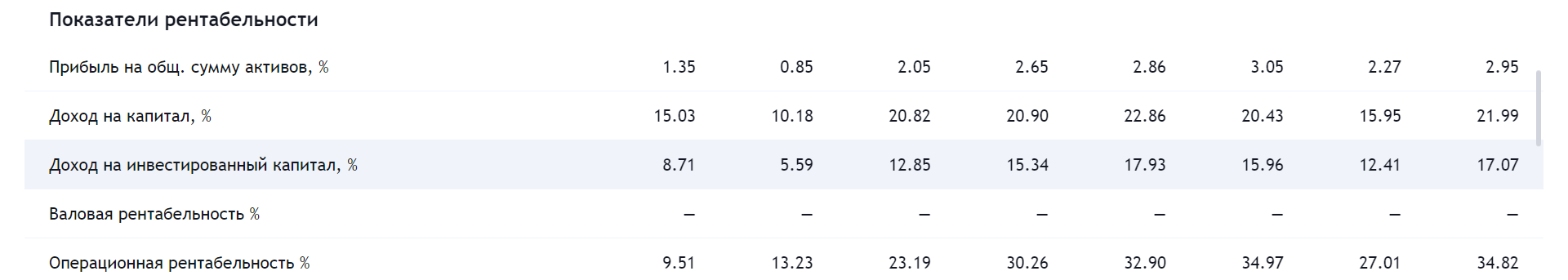






Рисунок 1 – Финансовые показатели компании Сбербанк

Скрипт по скачиванию html-кода страниц и извлечению из него таблиц будут в закреплении. Далее эти данные надо предобработать (привести к числовому формату все значения) и либо удалить, либо заменить пустые значения.

Далее необходимо добавить также актуальные цены самих акций, для этого подключим библиотеку investpy. Эти цены необходимо продисконтировать, чтобы учесть влияние инфляции (формула аналогична формуле выше).

После этого данные необходимо нормализовать – привести к единому масштабу. Хорошим решением будет расчет относительного изменения коэффициентов. То есть теперь у нас будет относительное изменение, например, коэффициента оборачиваемости активов за год = (0,12 – 0,1)/0,1 \*100 % = 20 % (с 2014 по 2015 увеличился на 20 %). Итоговая консолидированная таблица будет выглядеть следующим образом:

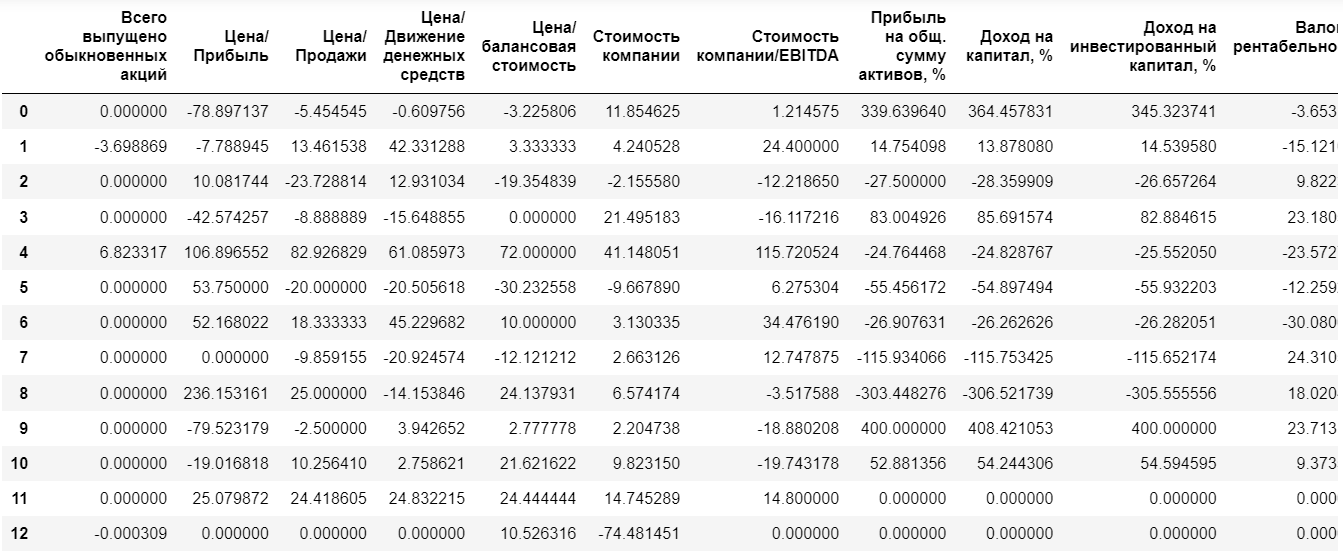




Рисунок 2 – Относительное годовое изменение части финансовых показателей компании Газпром в процентах

Так, например, с 2014 по 2015 оборачиваемость запасов уменьшилась на 5.19 %, а цена акции на 3.8.

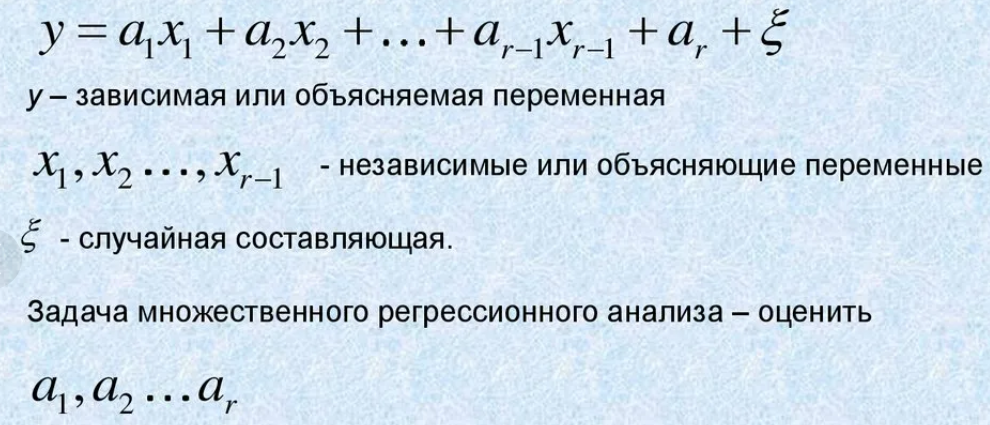
**Постановка задачи**

На основании представленных выше данных создать модель, которая на входе будет получать изменение фин. показателей компании и на выходе выдавать изменение цены акции относительно прошлой квартальной отчетности (то есть по ней можно вычислить и справедливую цену). Для этого рассмотреть следующие основные методы машинного обучения:

**Математическая постановка.**

**Общая для всех:** максимизироватьточность модели.

1. Постановка для множественной регрессии:

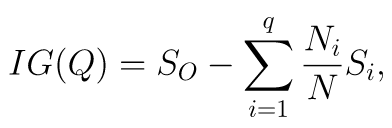


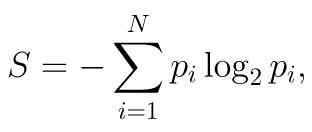
**Цель: Сумма квадратов отклонений - > 0.**

Рисунок 3 – Постановка задачи для множественной регрессии

Цель: Сумма квадратов отклонений - > 0.

1. Постановка задачи для дерева решений:



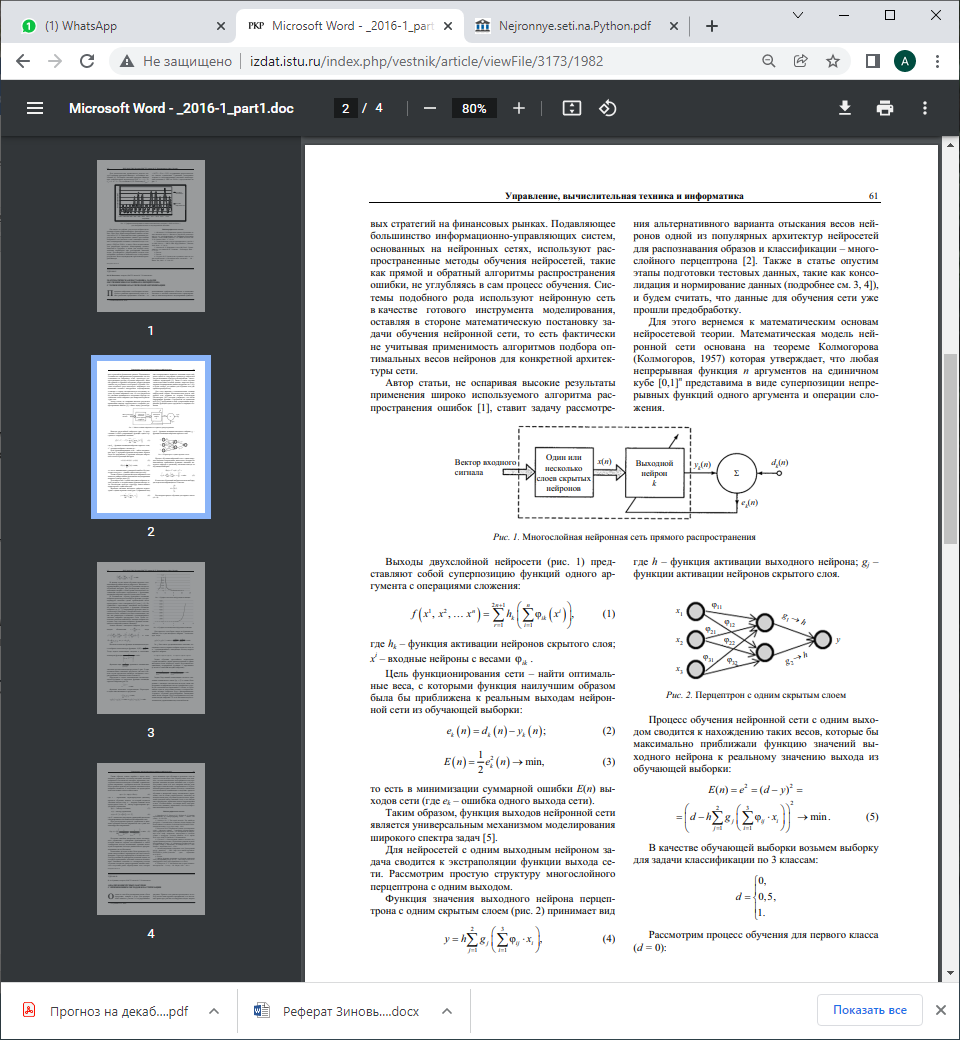


IG(Q) -> max: Q - ?

Рисунок 4 – Постановка задачи для дерева решений

где S – энтропия Шеннона для системы с N возможными состояниями, S0  - начальная энтропия, pi – вероятности нахождения системы в i-ом состоянии. Поскольку энтропия – по сути степень хаоса (или неопределенности) в системе, уменьшение энтропии называют приростом информации. Формально прирост информации (information gain, IG) при разбиении выборки по признаку Q. То есть необходимо знать наиболее важные признаки Q, по которым больше всего наблюдается прирост информации. [4]

1. Постановка задачи для нейронной сети: найти оптимальные веса, с которыми функция dk наилучшим образом была бы приближена к реальным выходам нейронной сети yk из обучающей выборки:



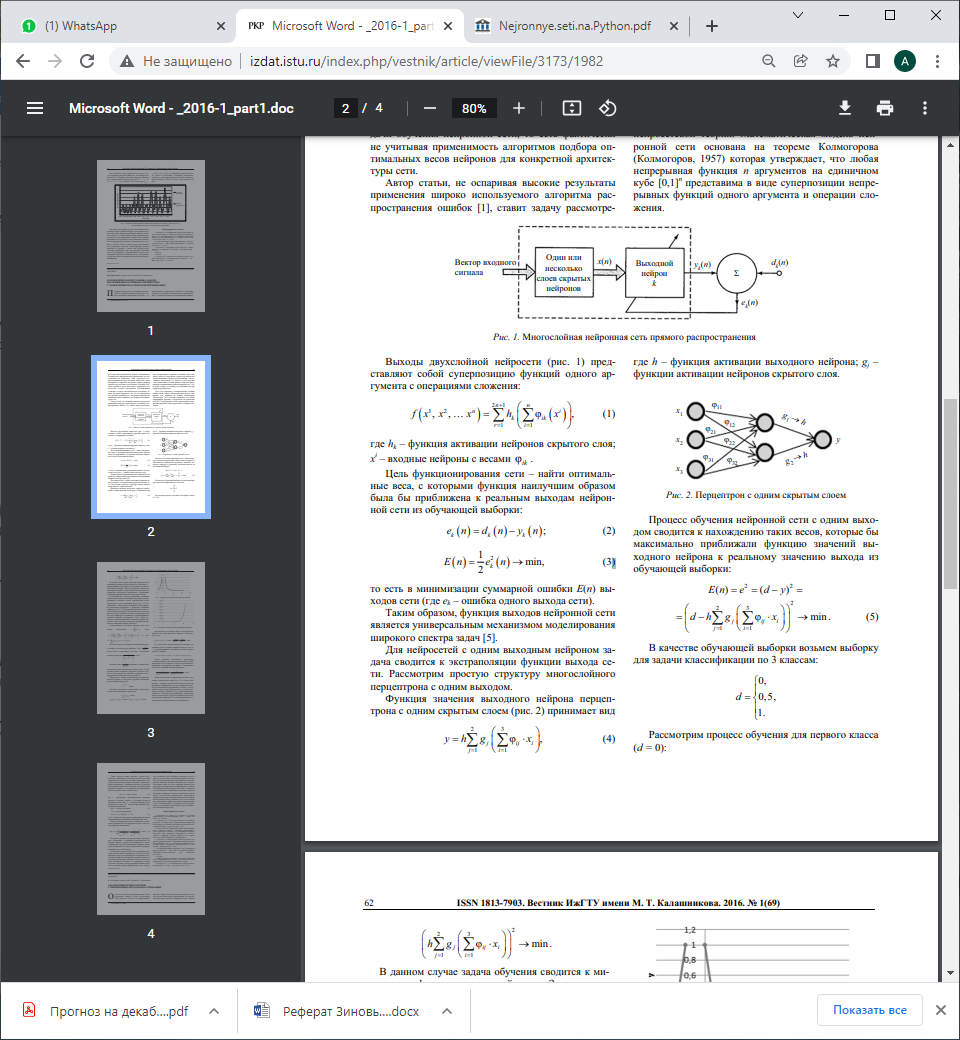


Рисунок 5 - Постановка задачи для нейронной сети [5]

Даны 3 постановки задачи, чтобы сравнить итоги простой линейной множественной регрессии, дерева решений и нейронной сети.

**Построение дерева**

Реализация дерева будет производиться на языке python с помощью библиотеки sklearn. Для более простого понимания разделим относительное изменение цены акций на категории, но сначала посмотрим частоту относительного изменения акций:

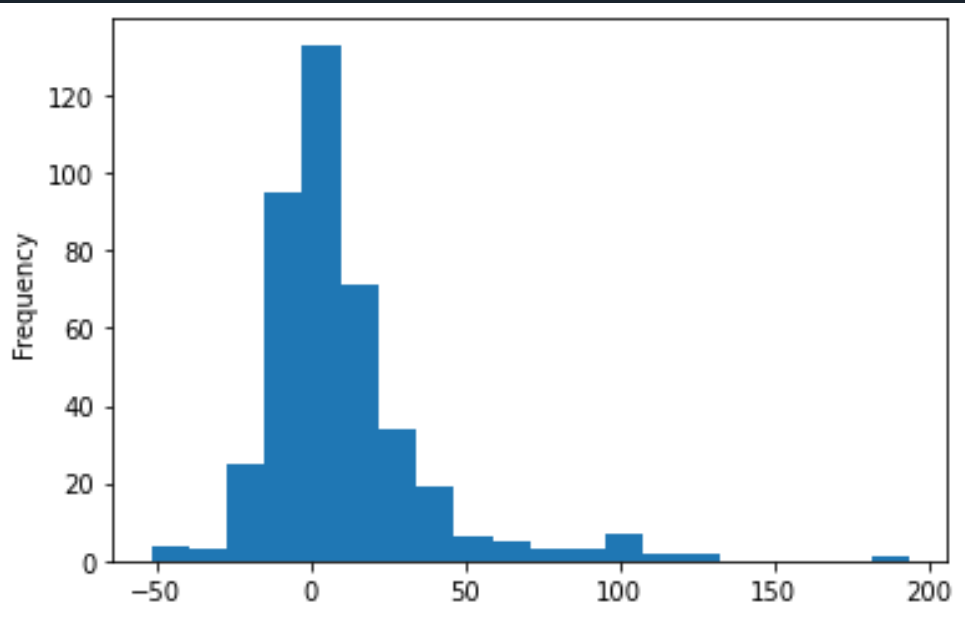


Рисунок 6 - Частота относительного изменения акций

Как видим, у большего кол-ва компаний цена акций меняется по абсолютной величине от 0 до 10 %.

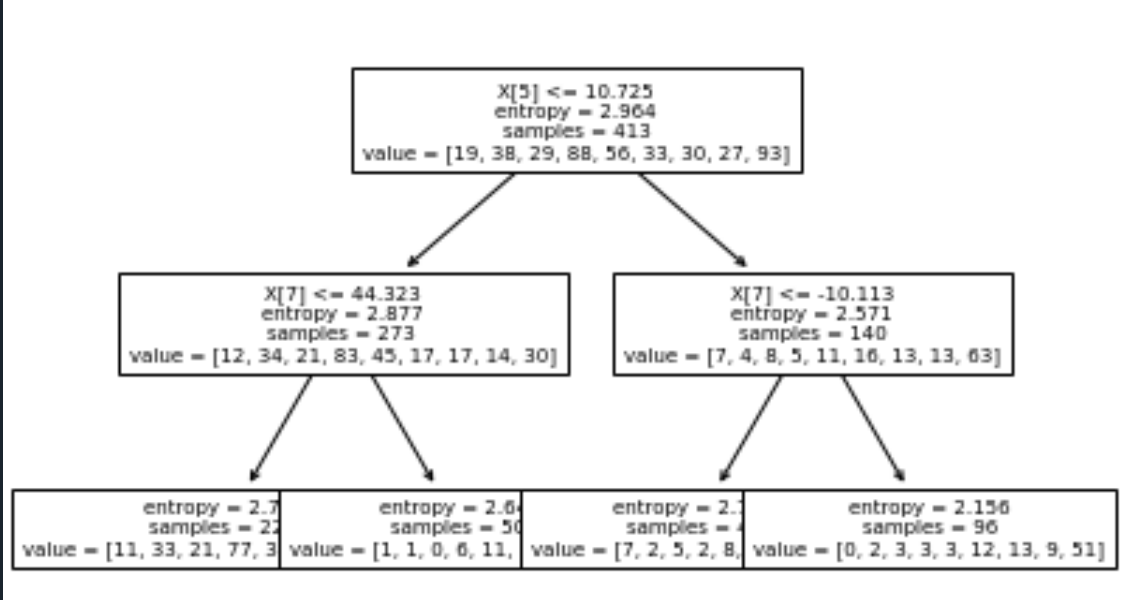
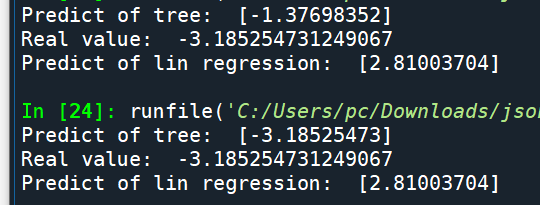


Рисунок 7 – Структура дерева решений с глубиной = 3

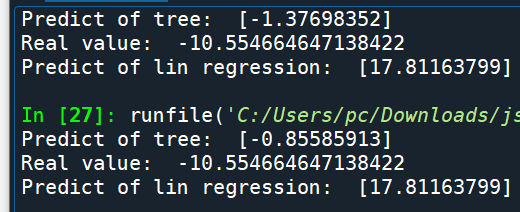
Выше показано обученное на ранее описанных данных дерево с глубиной, равной 3. Как видим, сверху вниз значение энтропии слабо уменьшается, поэтому нужно более глубокое дерево, но нельзя допустить переобучения, чтобы можно было корректно потом еще прогнозировать. Категории изменения цен акций: [-100, -20, -10, -5, 0, 5, 10, 15, 20, 500].

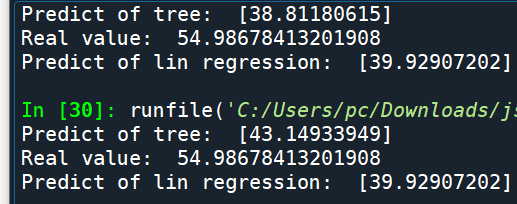
Также можно увидеть, например, что на первом шаге делили бинарно по 5 параметру (стоимость компании), то есть это один из самых важных параметров для предсказания цены акций.

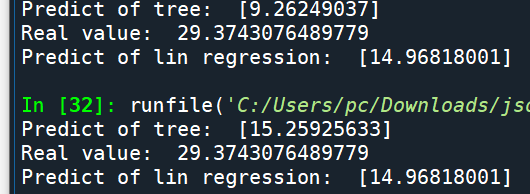
Для более наглядного представления отклонения прогноза результат дерева решений от реальных значений уйдем от категорий и возвратимся к относительным значениям цен акций. Сначала посмотрим на прогноз дерева на уже виденных им данных:



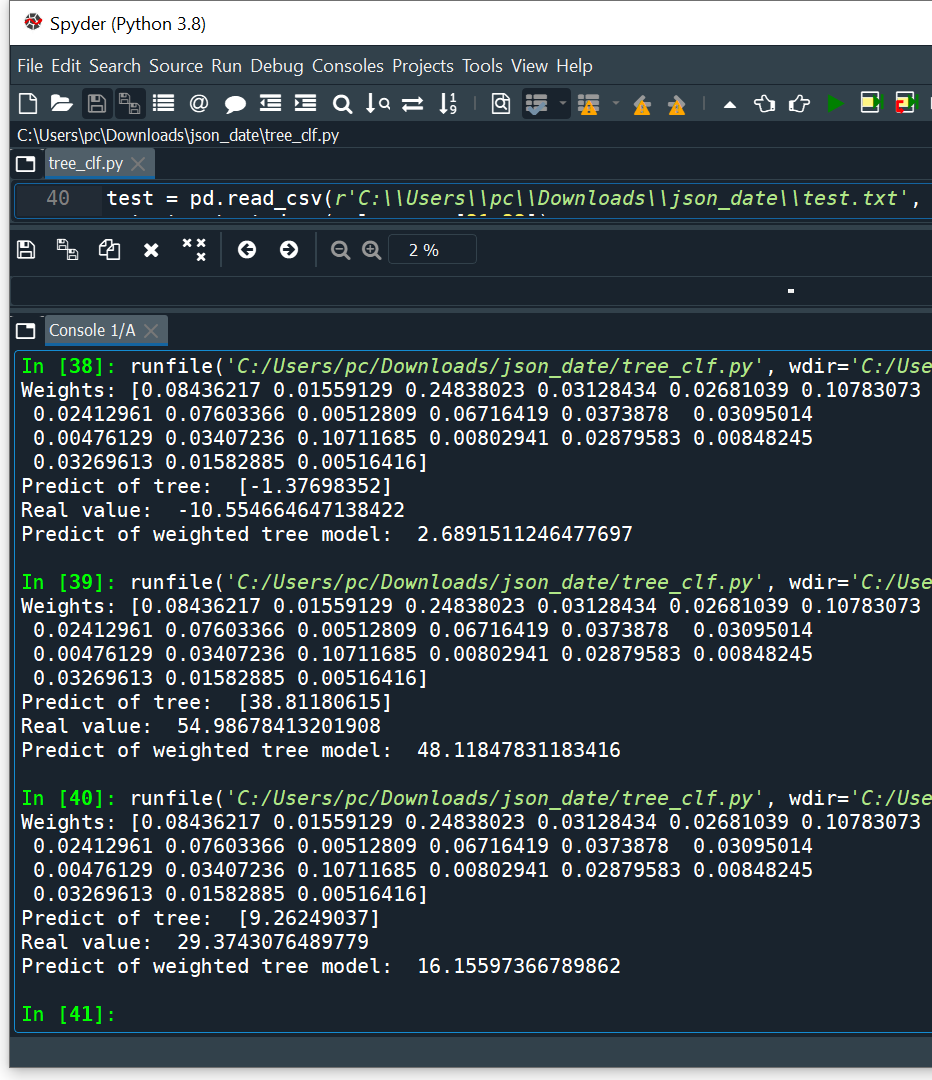
На верхнем случае, при глубине дерева = 10, а на нижнем – неограниченно, то есть на переобученном дереве. Прогноз дерева с небольшой глубиной точно спрогнозировал направление движения цены акций в отличие от линейной регрессии. Как видим, при переобучении дерево слишком зацикливается на этих данных, что может не позволить корректно обобщать данные. Посмотрим теперь на прогноз дерева на данных, которых оно не видело (3 различных компании, сверху – дерево с глубиной = 10, снизу – с неограниченной глубиной):







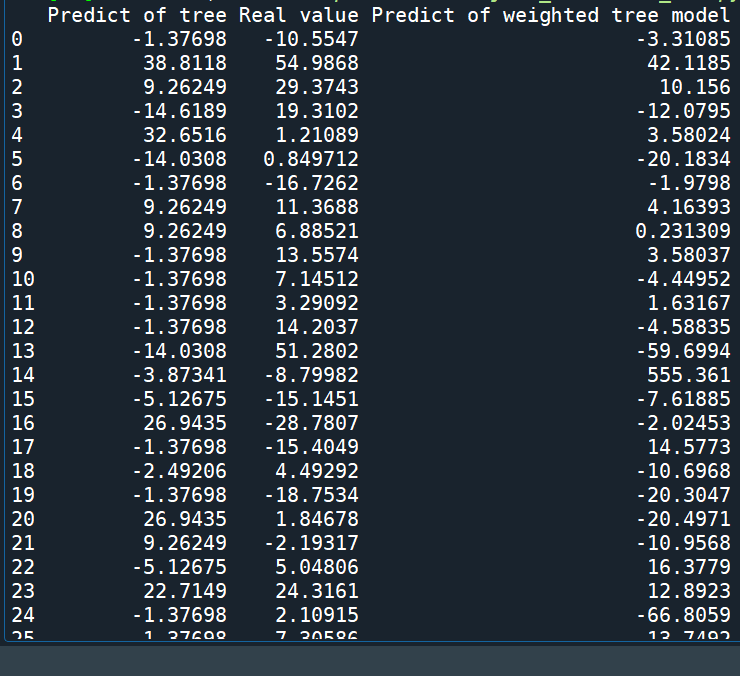
Как видим, дерево качественно во всех трех случаях при любой глубине угадало направление цен акций. Количественно же может довольно сильно ошибаться. Можно также перейти от бинарного деления к линейным коэффициентам, которые также рассчитаны в аттрибуте дерева, то есть скалярно перемножим вектора изменения фин показателей компании и вычисленные моделью веса weights:



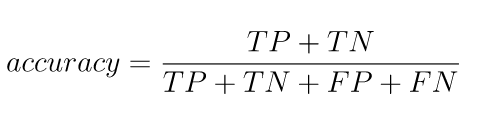
Как видим, 2 параметр – цена/продажи – самый важный (25 % веса), ранее упомянутый 5 - (стоимость компании) и 14 - быстрая ликвидность по 10 %.

Если взять ar взять в районе -6, например, то прогноз по взвешенным показателям будет даже еще точнее, чем прогноз самого дерева, во всех 3 случаях.

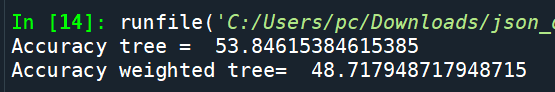
Запишем сравнения прогнозных значений цен акций и реальных значений в одну таблицу:



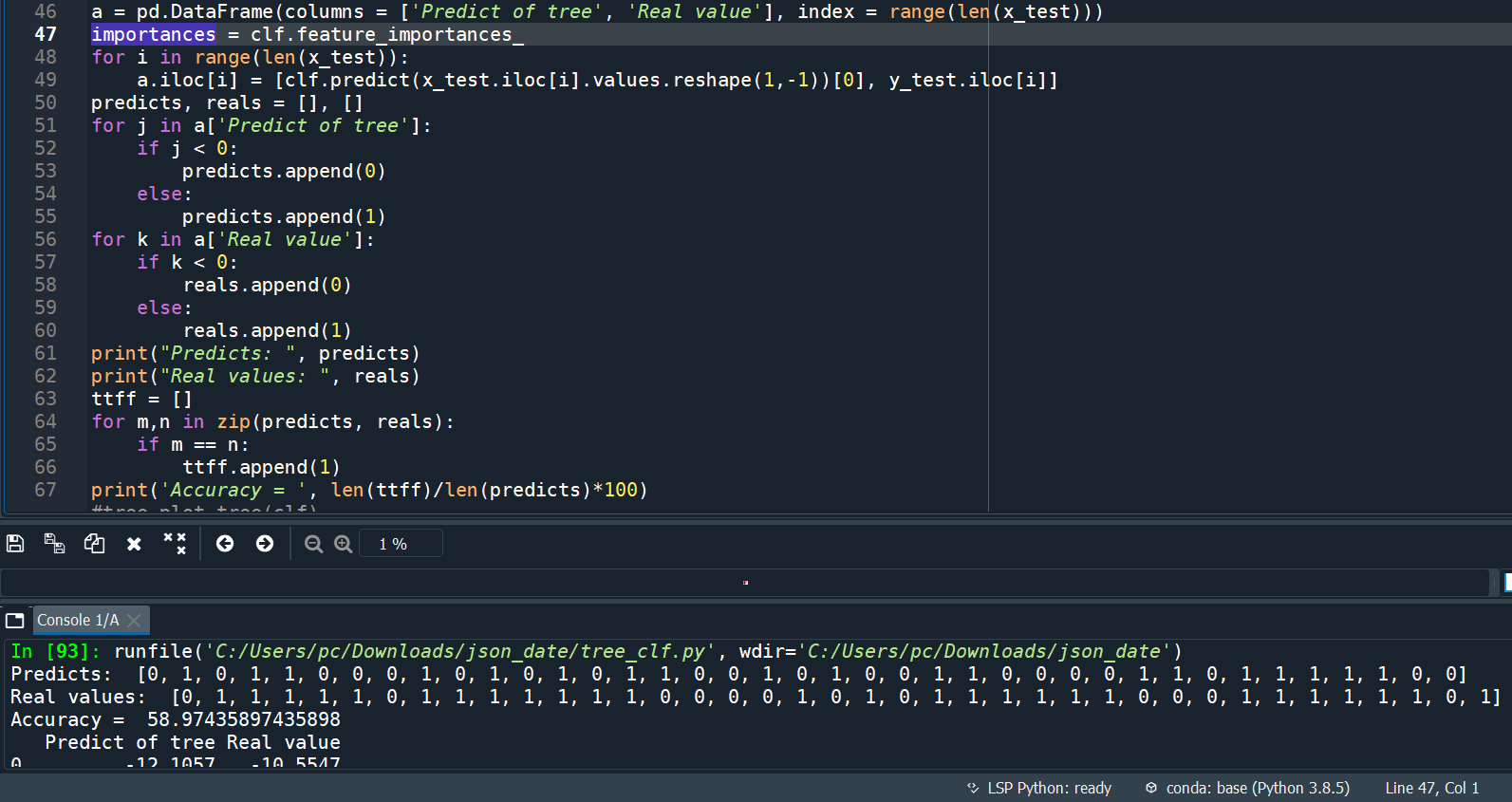
Для большей наглядности посчитаем точность модели (accuracy) – кол-во совпадений истинно-истинных и ложно-ложных значений ко всем выданным:



Для нашей модели будем считать 1, если у компании наблюдается за год рост, ибо инвестору в первую очередь важно понимать – потеряет ли деньги или нет, если вложит в этот актив, а уже потом какую прибыль с этого может иметь (для этого разделим на 2 категории):



То есть в среднем модель предсказывает так же, как и чистая случайность. Если попытаться улучшить модель, то всего немного достигнем прирост в точности:



Точность модели составило всего 59% (это лучшее значение, которое удалось получить на кросс-валидации). Это говорит о том, что модель, основанная на поочередном бинарном делении, не позволяет вскрыть более сложные зависимости представленных данных. Для установления таких связей необходимо применить нейронную сеть, которую с помощью регулирования ее размеров (кол-во нейронов в слое и кол-во слоев) и других парaметров можно сильно улучшить понимание данных моделью.

**Построение нейронной сети**

Простейшим примером нейронной сети является многослойный (в частном случае, однослойный) персептрон. Рассмотрим, что собой представляет такая модель.

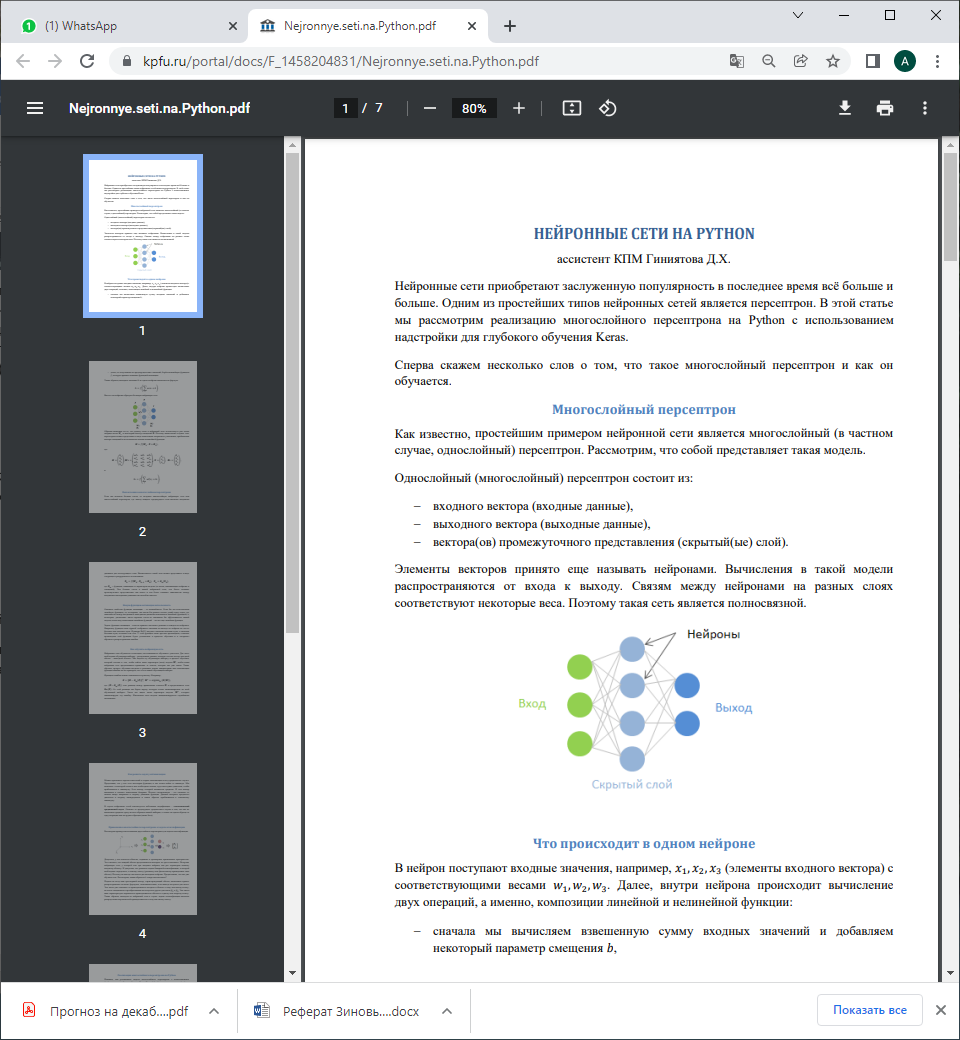
Однослойный (многослойный) персептрон состоит из:

− входного вектора (входные данные),

− выходного вектора (выходные данные),

− вектора(ов) промежуточного представления (скрытый(ые) слой).

Элементы векторов принято еще называть нейронами. Вычисления в такой модели распространяются от входа к выходу. Связям между нейронами на разных слоях соответствуют некоторые веса. Поэтому такая сеть является полносвязной.

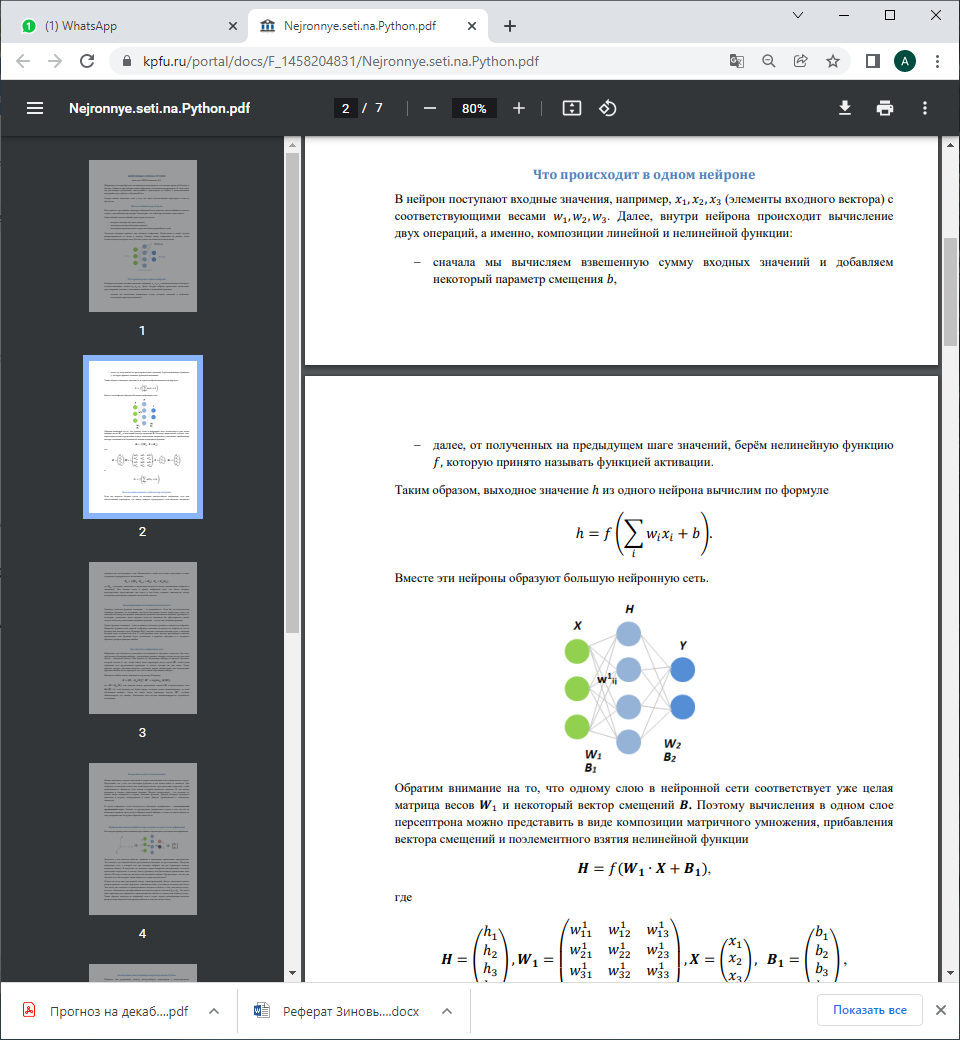


В нейрон поступают входные значения, например, 𝑥1, 𝑥2, 𝑥3 (элементы входного вектора) с соответствующими весами 𝑤1, 𝑤2, 𝑤3. Далее, внутри нейрона происходит вычисление двух операций, а именно, композиции линейной и нелинейной функции:

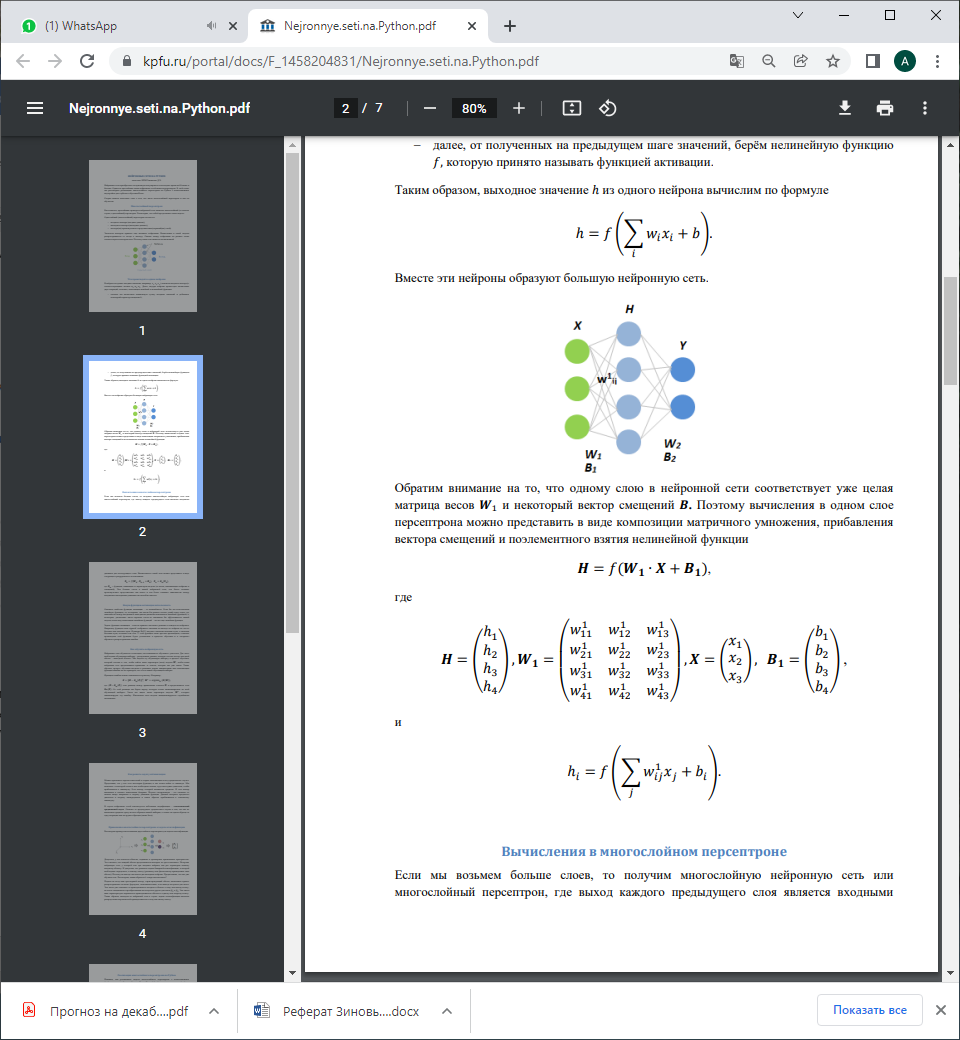
− сначала мы вычисляем взвешенную сумму входных значений и добавляем некоторый параметр смещения 𝑏,

− далее, от полученных на предыдущем шаге значений, берём нелинейную функцию 𝑓, которую принято называть функцией активации.

Таким образом, выходное значение ℎ из одного нейрона вычислим по формуле:



Обратим внимание на то, что одному слою в нейронной сети соответствует уже целая матрица весов 𝑾1 и некоторый вектор смещений 𝑩. Поэтому вычисления в одном слое персептрона можно представить в виде композиции матричного умножения, прибавления вектора смещений и поэлементного взятия нелинейной функции:



Если мы возьмем больше слоев, то получим многослойную нейронную сеть или многослойный персептрон, где выход каждого предыдущего слоя является входными данными для последующего слоя. Вычисления в такой сети можно представить в виде следующего рекуррентного соотношения:

𝑿𝒌 = 𝑓(𝑾𝒌 ∙ 𝑿𝒌−𝟏 + 𝑩𝒌 ), 𝑿𝒏 = 𝑮𝒘(𝑿𝟎),

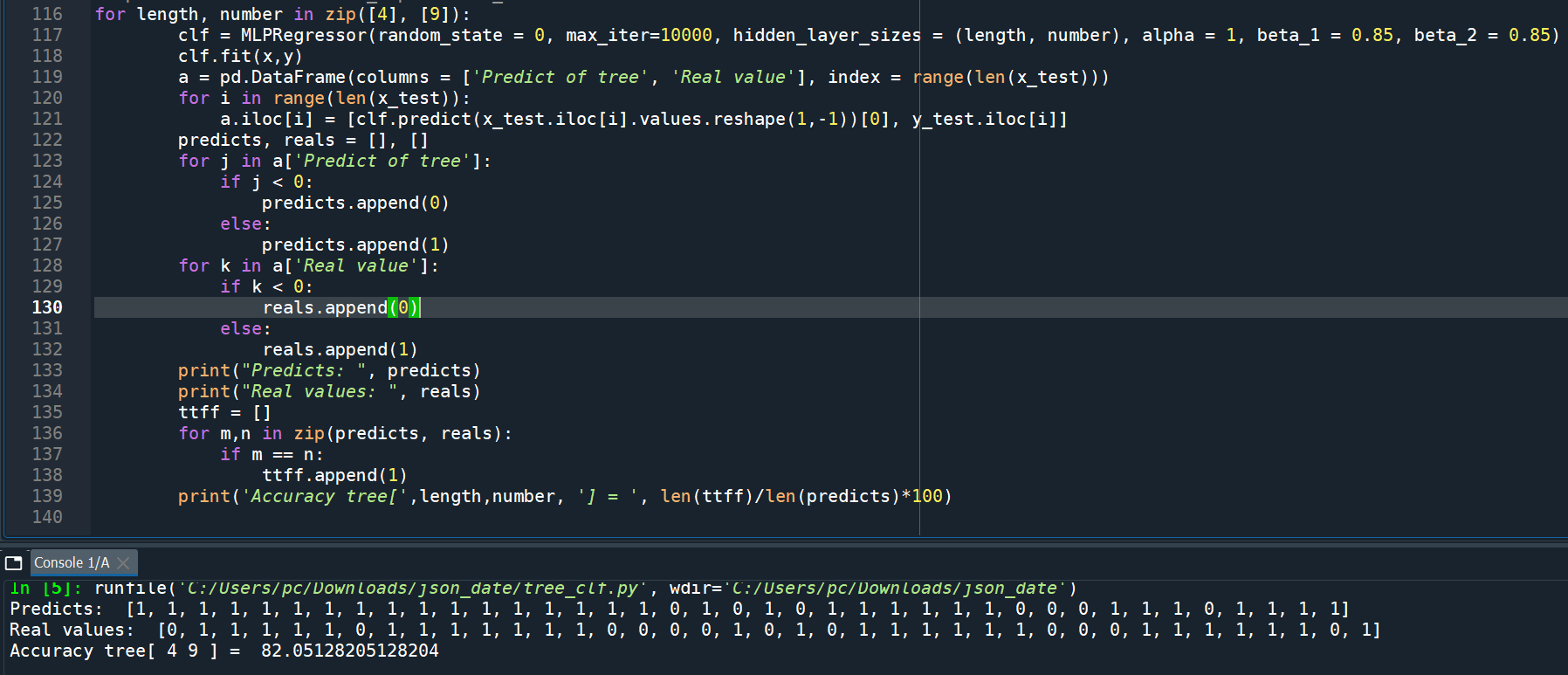
где 𝑮𝒘 – функция, зависящая от параметров модели (от весов, связывающих нейроны и смещений). Чем больше слоев в нашей нейронной сети, тем более сложное промежуточное представление она имеет и тем более сложные зависимости между входными и выходными данными она способна описать.

Основное свойство функции активации – ее нелинейность. Если бы мы использовали линейную функцию, то, во-первых, мы могли бы решать только узкий класс задач, где зависимость между входными и выходными данными описывается линейной функцией, а, во-вторых, увеличение числа скрытых слоев не повышало бы эффективность нашей модели, поскольку композиция линейных функций – это все еще линейная функция. Задача функции активации – помочь принять локальное решение в каждом из нейронов. Например, функция типа sigmoid отображает значения на выходе из нейрона во что-то большее или меньшее нуля. Функция ReLU зануляет значения меньшие нуля, а значения большие нуля, оставляет как есть. У этой функции очень простая производная, а именно производная этой функции будет участвовать в процессе обучения и в алгоритме обратного распространения ошибки. [6]

Перейдем к построению самой нейронной сети в python sklearn. Посмотрим сначала на результат сети из коробки (с настройками по умолчанию).

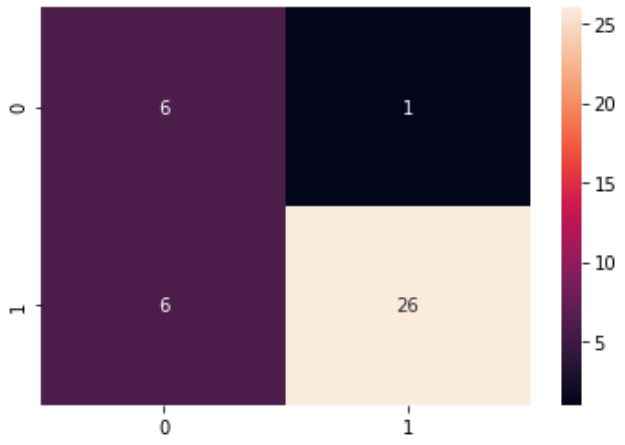


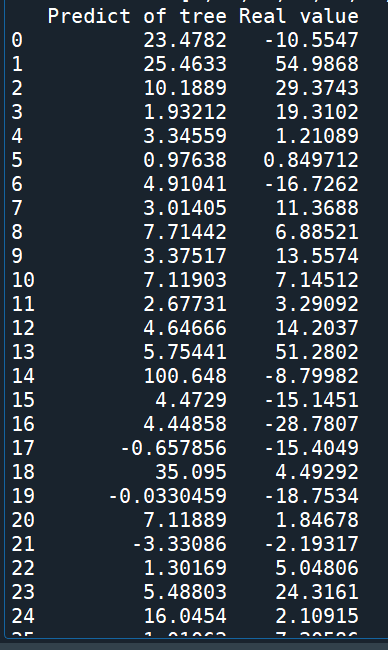
Как видим, даже без настройки параметров модель хорошо превосходит дерево решений в точности предсказания (66,6 %). Если же дополнительно использовать функцию GridSearch, которая будет проходить по сетке параметров с целью нахождения результатов работы нейронной сети и вывода параметров, при которых был достигнут наилучший результат при кросс-валидации, то получим следующее:



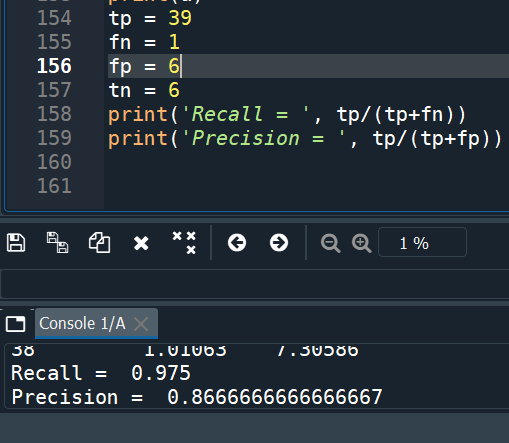
После итеративного прохода по всем параметрам была установлена следующая конфигурация нейронной сети:

activation = ‘relu’, solver = ‘adam’, max\_iter=10000, hidden\_layer\_sizes = (4,9), alpha = 1, beta\_1 = 0.85, beta\_2 = 0.85

Результат теперь значительно превзошел все итоги предыдущих моделей (82%). Такая модель уже вполне хороша для применения ее в практических целях. Но прежде надо еще посмотреть на ошибки, которые допускает модель. Построим для этого матрицу ошибок для выяснения количества ошибок 1 и 2 рода:



То есть модель допускает всего 1 ложноотрицательную ошибку (ошибку, при которой значение было спрогнозировано, как отрицательное, хотя на самом деле оно положительное) и 6 ложноположительных (ошибку, при которой значение было спрогнозировано, как положительное, хотя на самом деле оно отрицательное). Посчитаем precision и recall:



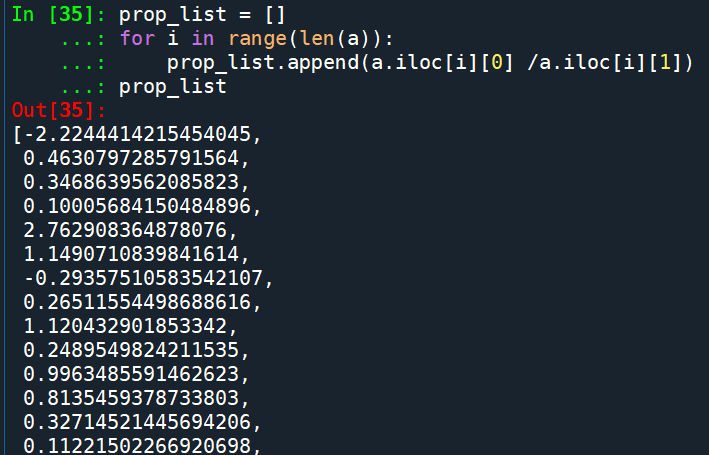
Recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм, а рrecision можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющимися положительными. Для нашей ситуации гораздо важнее recall, так как инвестору важнее не потерять деньги в принципе, чем пропустить возможность получить прибыль. Если захотим какой-то параметр увеличить, то придется пожертвовать другим, но значения обоих параметров и так очень хорошие, так что оставим эту модель [7].

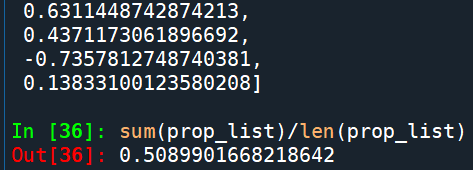
**Пожелания и возможные улучшения модели**

По найденной модели можно отобрать наиболее предрасположенные к росту акции. Для этого надо подать на вход обученной модели фин. показатели компании 2 последних рассчитываемых кварталов (точнее получить их абсолютное изменение за этот период). На выходе получим прогнозное значение, на сколько изменится цена акции относительно среднего значения за настоящий квартал, и сравним ее с настоящей ценой акции.

Например, мы хотим узнать, какая справедливая цена акции компании А. У нас есть ее отн.изменения фин.показателей 4-го квартала 21 года относительно 3-го, средняя цена акции за этот интервал (например, 100 руб) и цена акции на сегодня (например, 110 руб). Подавая на вход модели отн. изменения фин.показателей этой компании, получим относительное изменение цены акции (например, получили 20%). Это значит, что справедливая цена: 100 +100\*0.2 = 120 руб, а ее текущая цена 110 руб. То есть акции компании А недооценены и их можно рассмотреть к покупке. Сделав такие же вычисления для других компаний, можем найти наиболее перспективные компании.

Хоть модель качественна хороша, но количественно она сильно консервативна в своих прогнозах. Так прогнозируемые значения в среднем вдвое меньше соответствующих реальных:





Решением этой проблемы является большее обучение модели, но так как на [3] платформе представлено всего 45 российских компаний, то модель не смогла качественно натренировать себя. В дальнейшем будет использована гораздо большая база, выходящая за рамки МосБиржи и не ограничивающаяся только российскими компаниями.

Еще одной проблемой, связанной с недообучением модели, - это невозможность исследования отдельных кластеров из-за небольшого количества компаний. Так, например, можно было рассмотреть не 1 модель со всеми областями деятельности компаний, а разбить модели по секторам: создать модель по финансовому сектору, нефтегазовому, металлургическому и т.д, так как изменения финансовых показателей по-разному влияют на сектора. Так, например, изменение мультипликатора «капитализация к балансовой стоимости активов» значительно большее имеет влияние на компании нефтегазового сектора, чем на IT, так как в первой области просто необходимо иметь большое количество физического имущества в виде техники. Далее также при обновлении базы будут выделяться модели по отдельным сферам деятельности компании.

Другой проблемой, с которой может столкнуться модель, - это неучет изменчивости цены акции компании. Включение такого фактора покажет влияние риска на ценность данной ценной бумаги. Для примера: цена акции компании А (100 руб) по прогнозу должна увеличиться на 10 % (110 руб), а компании В (100 руб) на 5 %, но из-за того что коэффициент вариации у первой так же выше (например, также 10% и 5% сответственно), а страх потерь выше, чем желание выгоды (по многочисленным исследованиям, например, Митеша Патела [8, статья на английском], то статистически справедливые доходности этих 2 акций могут быть равнозначны. А если, например, есть акция, которая почти всегда стабильно растет, но очень медленно (Bank of America, например) и в противовес ей скачущая акция (Tesla, например), то как оценить количественно двух таких антиподов-акций? Возможные решения данной проблемы, я надеюсь, подскажите вы, так как значительно больше владеете информационной базой по поводу страховых расчетов и финансовой математики.

**Заключение**

Были рассмотрены построения и результаты одних из самых используемых методов в машинном обучении: линейная регрессия, деревья решений и нейронные сети, которые были применены для обучения моделей на данных изменения квартальных финансовых показателей компании для прогнозирования справедливой цены акции. Как показали исследования, результаты точности моделей составили 48, 58 и 82% соответственно, то есть нейронные сети сильно обошли в прогнозировании других методов.

Были высказаны недостатки модели такие, как малая база компаний, отсутствие кластеризации по их сферам деятельности и влияния риска колебаний цены на ее ценность, и возможные решения этих проблем.

В дальнейшем с вашей помощью в области вычисления рисков планируется создать модель, которая будет с приведением к настоящему моменту вычислять самые перспективные акции.

**Список использованной литературы**

1. Бенджамин Грэхем, Дэвид Додд. Анализ ценных бумаг = Security Analysis. — М.: Вильямс, 2016. — 880 с. — ISBN 978-5-8459-1945-8, 978-0-07-144820-8.
2. Интернет-источник <https://bcs-express.ru/novosti-i-analitika/stoimostnoe-investirovanie-chast-2-istoriia-sut-podkhoda-i-kriterii-otbora-aktsii>
3. Интернет-источник <https://tradingview.com/>.
4. Интернет-источник <https://habr.com/ru/company/ods/blog/322534/>
5. «Математическая постановка задачи обучения многослойного перцептрона с точки зрения классической оптимизации», Николаева Ю.В., вестник ИжГТУ, 2016 г., № 1(69). Интернет-источник <http://izdat.istu.ru/index.php/vestnik/article/viewFile/3173/1982>
6. Методическое руководство по дисциплине «Нейронные сети на python», Гиниятова Д.Х., КФУ. Интернет-источник <https://kpfu.ru/portal/docs/F_1458204831/Nejronnye.seti.na.Python.pdf>
7. Интернет-источник <https://habr.com/ru/company/ods/blog/328372/>
8. Интернет-источник <https://www.thedp.com/article/2016/03/med-school-fitness-study>

**Приложение**

Коротко о научных и профессиональных навыках (для справки):

Зиновьев Евгений Андреевич, 09 – 115 (3), Математические методы в экономике и финансах

2011 - 2014 г – проф. образование программиста-электроника (МИСиС)

2016 – 2020 – высшее образование разработчика нефтяных и газовых месторождений (СПГУ, бакалавриат)

Тема дипломной работы – анализ оптимизации контролируемых параметров гидроразрыва пласта Мамонтовского месторождения для увеличения его нефтеотдачи.

Основная тема научных публикаций в разных источниках – оптимизация численного решения при моделировании теплового процесса в нефтяном пласте.

2020 – 2021 – неоконченная магистратура экономики и менеджмента на предприятии (СПбПУ)

Тема дипломной работы – предиктивная модель направления движения цен акций компаний на фондовом рынке на основании применения методов машинного обучения к прогнозам инвестдомов.

Наличие инвест. портфеля, показывающего 7-15% годовых прибыли.

В проф.деятельности – практик , аналитик данных:

2019-2021 – аналитик в отделе планирования ООО «НТМ», Санкт-Петербург (осн. обязанность - прогнозирование потребления продукции сетью продуктовых точек и связанные с этим произв. задачи для оптимизации бизнеса, в том числе с применением методов машинного обучения)

2021 – н.в. – аналитик дирекции по аналитической работе Nefis Cosmetics, Казань (осн. обязанность – анализ динамики продаж по КЖК, прогноз продукции по торговым данным, оперативные данные по каналам сбыта)