

Выпускная квалификационная работа по курсу «Data Science»

1. Общие сведения

В процессе выполнения выпускной квалификационной работы необходимо:

1. Подготовить проект, который необходимо будет представить на защите.

2. Подготовить пояснительную записку. Объем записки 30 страниц.

Структура пояснительной записки:

1. Аналитическая часть

1.1. Постановка задачи.

В данном разделе приводится в текстовом виде смысловое описание решаемой задачи анализа данных. Также дается характеристика датасета – число входных и выходных переменных (факторов), объем выборки, характеристика выборки с точки зрения ее особенностей (выбросы, пропуски и т.д.).

1.2. Описание используемых методов

В данном разделе приводится краткое описание методов (не более 1 страницы на каждый), которые предполагается использовать для решения поставленной задачи. Необходимо обратить особое внимание на достоинства, недостатки и области применения каждого из методов. Кроме того, желательно составление итоговой сравнительной таблицы с указанием среди прочего априорных предпосылок к работоспособности каждого метода.

1.3. Разведочный анализ данных

В данном разделе приводится краткое описание методов разведочного анализа данных, которые используются для первоначального анализа.

2. Практическая часть

2.1. Предобработка данных

В данном разделе приводится графики распределения для каждого признака до и после нормализации, указать максимальные и минимальные значения.

2.2. Разработка и обучение модели

В данной части приводится список моделей, которые будут использоваться для прогноза модуля упругости при растяжении и прочности при растяжении.

2.3. Тестирование модели

В данном разделе показывается ошибка каждой модели на тренировочной и тестирующей части выборки. Обосновывается выбор модели.

- 2.4. *Написать нейронную сеть, которая будет рекомендовать соотношение матрица.*

Описывается выбранная архитектура нейронной сети и ее результаты.

- 2.5. *Разработка приложения*

Описать функционал приложения и краткую инструкцию использования.

- 2.6. *Создание удаленного репозитория и загрузка результатов работы на него.*

Указывается страница слушателя, созданный репозиторий, коммиты в репозитории.

3. Подготовить презентацию до 15 слайдов с иллюстрациями к пояснительной записке и к самому проекту. Доклад на 7-10 минут.

2. Задание

Тема: Прогнозирование конечных свойств новых материалов (композиционных материалов).

Описание:

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т. е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом. Яркий пример композита - железобетон. Бетон прекрасно сопротивляется сжатию, но плохо растяжению. Стальная арматура внутри бетона компенсирует его неспособность сопротивляться сжатию, формируя тем самым новые, уникальные свойства. Современные композиты изготавливаются из других материалов: полимеры, керамика, стеклянные и углеродные волокна, но данный принцип сохраняется. У такого подхода есть и недостаток: даже если мы знаем характеристики исходных компонентов, определить характеристики композита, состоящего из этих компонентов, достаточно проблематично. Для решения этой проблемы есть два пути: физические испытания образцов материалов, или прогнозирование характеристик. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента).

На входе имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов. Кейс основан на реальных производственных задачах Центра НТИ «Цифровое материаловедение: новые материалы и вещества» (структурное подразделение МГТУ им. Н.Э. Баумана).

Актуальность: Созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

Датасет со свойствами композитов. Объединение делать по индексу тип объединения INNER

https://drive.google.com/file/d/1B1s5gBlvgU81H9GGolLQVw_SOivyNf2/view?usp=sharing

Требуется:

1) Изучить теоретические основы и методы решения поставленной задачи.

2) Провести разведочный анализ предложенных данных. Необходимо нарисовать гистограммы распределения каждой из переменной, диаграммы ящика с усами, попарные графики рассеяния точек. Необходимо также для каждой колонке получить среднее, медианное значение, провести анализ и исключение выбросов, проверить наличие пропусков.

3) Провести предобработку данных (удаление шумов, нормализация и т.д.).

4) Обучить нескольких моделей для прогноза модуля упругости при растяжении и прочности при растяжении. При построении модели необходимо 30% данных оставить на тестирование модели, на остальных происходит обучение моделей. При построении моделей провести поиск гиперпараметров модели с помощью поиска по сетке с перекрестной проверкой, количество блоков равно 10.

5) Написать нейронную сеть, которая будет рекомендовать соотношение матрица-наполнитель.

6) Разработать приложение с графическим интерфейсом или интерфейсом командной строки, которое будет выдавать прогноз, полученный в задании 4 или 5 (один или два прогноза, на выбор учащегося).

7) Оценить точность модели на тренировочном и тестовом датасете.

8) Создать репозиторий в GitHub / GitLab и разместить там код исследования. Оформить файл README.

3. Подготовка к защите и защита работы

Защита выпускной квалификационной работы является завершающим этапом обучения. Защита проходит перед комиссией, имеющей в составе не менее 3 преподавателей и/или экспертов МГТУ им.Н.Э. Баумана.

При оформлении документации по выпускной работе необходимо подготовить для представления в комиссию следующие документы:

- 1) пояснительную записку в электронном виде (Word и PDF);
- 2) презентацию в электронном виде (Power Point и PDF);

За три дня до назначенной даты защиты выпускной квалификационной работы обучающийся должен представить всю документацию, в противном случае обучающийся может быть не допущен до защиты.

Весь материал, выносимый на слайды презентации, обязательно должен быть идентичен иллюстрациям, представленным в работе.

4. Оформление выпускной работы

4.1 Общие положения

Пояснительная записка к выпускной квалификационной работе (ВКР) состоит из следующих элементов:

- 1) титульный лист (Приложение);
- 2) содержание;
- 3) введение;
- 4) основная часть (разделы и подразделы);
- 5) заключение;
- 6) библиографический список;
- 7) приложения (по необходимости).

Выпускная работа должна быть оформлена на одной стороне листа белой бумаги формата А4. Допускается представлять таблицы и иллюстрации на листах бумаги формата не более А3.

Требования к форматированию текста:

- 1) поля: левое – 30 мм; правое – 10 мм; верхнее – 20 мм; нижнее 15 мм;
- 2) размер шрифта – 14;
- 3) стиль шрифта – Times New Roman;
- 4) цвет текста – черный;
- 5) междустрочный интервал – 1,5 строки;
- 6) абзацный отступ – 1,25 см;
- 7) расстановка переносов – автоматическая;
- 8) переносы в названиях разделов и подразделов не допускаются;
- 9) интервалы между абзацами не допускаются;
- 10) полужирное и курсивное начертание, а также подчеркивание текста не допускается.

В тексте пояснительной записки нельзя заменять слова и словосочетания их символьными обозначениями, а также применять математические знаки без числовых значений, например, $>$ (больше), $<$ (меньше), $=$ (равно), \geq (больше или равно), \leq (меньше или равно), № (номер), % (процент) и т.д.

Пример:

Правильно: *Наибольший процент соответствия имеет система...*

Неправильно: *Наибольший % соответствия имеет система...*

Если в тексте приводится диапазон числовых значений физической величины, выраженных в одной и той же единице физической величины, то обозначение единицы физической величины указывается после последнего числового значения диапазона.

Пример:

От 150 до 200 ч.

Основную часть пояснительной записки следует делить на разделы и подразделы, которые следует нумеровать арабскими цифрами и записывать с абзацного отступа.

Разделы должны иметь порядковую нумерацию в пределах всего текста, за исключением приложений.

Пример:

1, 2, 3 и т. д.

Номер подраздела включает номер раздела и порядковый номер подраздела, разделенные точкой.

Пример:

1.1, 1.2, 1.3 и т. д.

Разделы, подразделы должны иметь заголовки. Они должны четко и кратко отражать содержание разделов, подразделов.

Заголовки разделов в тексте следует печатать по центру страницы, симметрично тексту, с прописной буквы, с полужирным начертанием, размер шрифта – 16. Каждый раздел начинается с новой страницы.

Заголовки подразделов следует печатать с абзацного отступа с прописной буквы, с полужирным начертанием, размер шрифта – 14.

Между номером раздела и заголовком точка не ставится.

Пример:

1.1 Постановка задачи

Расстояние между заголовками и текстом должно быть увеличено для выделения заголовка. До и после заголовка устанавливается интервал 15 пт. Заголовки не подчеркиваются, в конце их точки не ставятся. В заголовках следует избегать сокращений (за исключением общепризнанных аббревиатур), также не допускается перенос слова на следующую строку. Если заголовок состоит из двух предложений, их разделяют точкой.

Титульный лист, задание на ВКР, календарный план, последний лист ВКР оформляются самостоятельно по образцу.

В тексте пояснительной записки к ВКР рекомендуется чаще применять красную строку, выделяя законченную мысль в самостоятельный абзац. Излагать материал в работе следует четко, ясно, применяя принятую научную терминологию и избегая повторений.

Вне зависимости от способа выполнения пояснительной записки качество напечатанного текста, оформления иллюстраций и таблиц должно удовлетворять требованию их четкого воспроизведения.

4.2 Нумерация страниц

Все страницы пояснительной записки к ВКР (кроме бланков и содержания) обязательно должны быть пронумерованы. Нумерация страниц начинается с третьего листа (третий лист – введение) и заканчивается последним. На титульном листе, задании и содержании ВКР номера страниц не ставят.

Страницы следует нумеровать арабскими цифрами, соблюдая сквозную нумерацию по всему тексту. Номер страницы проставляют в центре нижней части листа без точки, используя шрифт Times New Roman размером 14 пт.

4.3 Оформление списков

В тексте ВКР могут быть приведены перечисления. Их выделяют абзацным отступом, который используют только в первой строке. После каждого перечисления ставится точка с запятой, после последнего – точка.

Перед каждым перечислением ставят арабскую цифру, строчную букву русского алфавита (за исключением букв ё, з, о, г, ь, й, ы, ъ) или дефис. При использовании цифр и букв после них ставят скобку. Текст перечисления пишут со строчной буквы. Между скобкой и текстом точку не ставят.

Пример:

- 1) концептуальная;**
- 2) логическая;**
- 3) физическая.**

а) концептуальная;

- б) логическая;*
- в) физическая.*

- концептуальная;*
- логическая;*
- физическая.*

В случае использования многоуровневых списков для пунктов перечисления применяют арабские цифры, а для подпунктов строчные буквы русского алфавита. Подпункты выделяют дополнительным абзацным отступом.

Пример:

- 1) _____;
 - а) _____;*
 - б) _____;*
- 2) _____;
 - а) _____;*
 - б) _____;*
 - в) _____;*
- 3) _____;
- 4) _____;
 - а) _____;*
 - б) _____.*

4.4 Оформление иллюстраций

Для наглядности в работу могут быть включены иллюстрации (графики, схемы, компьютерные распечатки, диаграммы). Все виды иллюстраций обозначаются словом «Рисунок». Их следует располагать непосредственно после текста, в котором они упоминаются впервые, или на следующей странице. На все рисунки должны быть даны ссылки в тексте.

Рисунки следует нумеровать арабскими цифрами, сквозной нумерацией.

Рисунок обозначается словом «Рисунок» с указанием номера (без знака №). Название рисунка следует помещать в одну строку с его номером через тире. Обозначение рисунка и его наименование выполняют шрифтом Times New Roman размером 14 пт. и располагают посередине строки, симметрично самому рисунку. Рисунок располагается по центру страницы.

При ссылках на иллюстрации следует писать «... в соответствии с рисунком 10». Иллюстрации, при необходимости, могут иметь пояснительные данные (подрисовочный текст), которые отделяются от

наименования рисунка двоеточием и располагаются на последующих строках. После пояснения рисунка точка не ставится.

Между рисунком и подрисуночной подписью (а также между текстом работы и рисунком) не должно быть пустых строк. После подрисуночного текста необходимо оставить одну пустую строку.

Графики, схемы, диаграммы выполняются четко, красиво, желательно в цвете, в соответствии с требованиями деловой документации.

Пример:

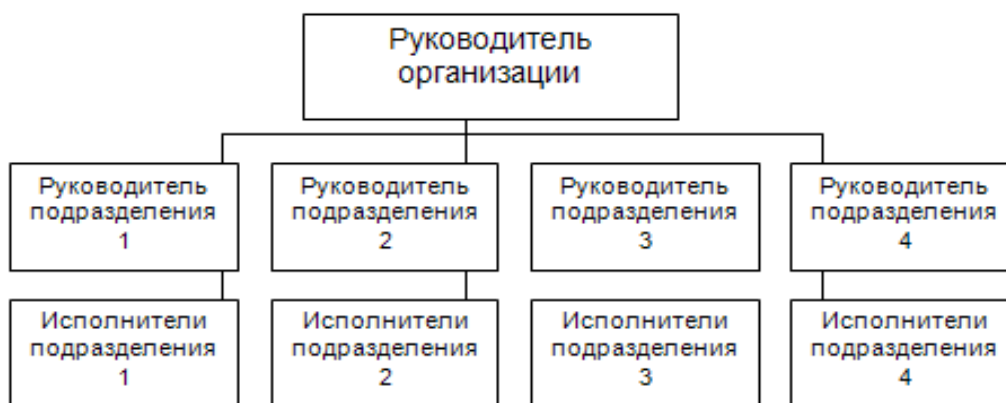


Рисунок 10 – Организационная структура предприятия

4.5 Оформление таблиц

Таблицы применяют для лучшей наглядности и удобства сравнения показателей. Название таблицы должно отражать ее содержание, быть точным и кратким.

Нумерация таблиц должна быть сквозной на протяжении всей работы. Название таблицы следует помещать над таблицей слева, без абзачного отступа в одну строку с ее номером (без знака №) через тире. Таблицу следует располагать непосредственно после текста, в котором она упоминается впервые, или на следующей странице. На все таблицы должны быть ссылки в тексте работы. При ссылке следует писать слово «Таблица» с указанием ее номера.

Пример:

Таблица 5 – Сроки реализации проекта

Период	Стадия
00.00.00 – 00.00.00	Создание технического задания
00.00.00 – 00.00.00	Изготовление
00.00.00 – 00.00.00	Тестирование/внедрение

00.00.00 – 00.00.00	Эксплуатация
---------------------	--------------

Разрывать таблицу и переносить часть ее на другую страницу можно только в том случае, если она целиком не уместается на одной странице. При переносе части таблицы название помещают только над первой частью таблицы, нижнюю горизонтальную черту, ограничивающую таблицу, не проводят.

После шапки таблицы следует поместить дополнительную строку с нумерацией столбцов. На следующей странице указывается заголовок «Продолжение таблицы» (с указанием порядкового номера), ориентированный по правому краю, а сама таблица начинается со строки с нумерацией столбцов.

Пример:

Таблица 5 – Сроки реализации проекта

Период	Стадия
1	2
00.00.00 – 00.00.00	Создание технического задания
00.00.00 – 00.00.00	_____

00.00.00 – 00.00.00	_____

-----Новая страница-----

Продолжение таблицы 5

1	2
00.00.00 – 00.00.00	Изготовление
00.00.00 – 00.00.00	_____
00.00.00 – 00.00.00	_____
00.00.00 – 00.00.00	Тестирование/внедрение
00.00.00 – 00.00.00	_____
00.00.00 – 00.00.00	_____
00.00.00 – 00.00.00	_____
00.00.00 – 00.00.00	Эксплуатация

00.00.00 – 00.00.00	
---------------------	--

Заголовки столбцов и строк таблицы следует писать с прописной буквы в единственном числе, а подзаголовки столбцов – со строчной буквы, если они составляют одно предложение с заголовком, или с прописной буквы, если они имеют самостоятельное значение. В конце заголовков и подзаголовков таблиц точки не ставят.

Таблицы слева, справа, сверху и снизу, как правило, ограничивают линиями. Допускается применять размер шрифта в таблице меньший на 2 пт., чем в тексте. Горизонтальные и вертикальные линии, разграничивающие строки таблицы, допускается не проводить, если их отсутствие не затрудняет пользование таблицей.

Графа «Номер по порядку» в таблицу не включается. При необходимости нумерации показателей, их порядковые номера указывают в первой графе (боковике) таблицы, непосредственно перед их наименованием.

При отсутствии отдельных данных в таблице следует ставить прочерк (тире). Числовое значение показателя проставляют на уровне последней строки наименования этого показателя.

Содержание характеристики, приведенное в таблице виде текста, записывают на уровне первой строки наименования этой характеристики. После таблицы необходимо оставить пустую строку.

4.6 Оформление формул и уравнений

Уравнения и формулы следует выделять из текста в отдельную строку. Выше и ниже каждой формулы или уравнения должно быть оставлено по одной свободной строке. Расположение формулы – с абзачного отступа, выравнивание по левому краю. Если уравнение не умещается в одну строку, то оно должно быть перенесено после знака равенства (=) или после знаков плюс (+), минус (–), умножения (X), деления (:), или других математических знаков, причем знак в начале следующей строки повторяют. При переносе формулы на знаке, символизирующем операцию умножения, применяют знак «X».

Пояснение значений символов и числовых коэффициентов следует приводить непосредственно под формулой в той же последовательности, в которой они даны в формуле. Пояснение каждого символа приводят с новой строки, первую строку пояснения начинают со слова «где» без двоеточия после него и без абзачного отступа.

Формулы в работе следует нумеровать порядковой нумерацией в пределах всей выпускной работы арабскими цифрами в круглых скобках в крайнем правом положении на строке.

Пример:

$$NPV = \sum_{k=1}^n \frac{P_k}{(1+r)^k} - \sum_{k=1}^n \frac{C_k}{(1+r)^k},$$

где n – количество этапов;
 P_k – результаты на k -том этапе;
 r – коэффициент дисконтирования;
 C_k – затраты на k -том этапе.

Ссылки в тексте на порядковые номера формул дают в скобках.

Пример:

«... в формуле (3)»

4.7 Оформление приложений

Приложения помещают после списка использованных нормативно-правовых актов и научной литературы в порядке их упоминания в тексте.

В тексте работы на все приложения должны быть даны ссылки. Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова «Приложение», его обозначения и степени. Приложение должно иметь заголовок, который записывают симметрично относительно текста с прописной буквы отдельной строкой, размером шрифта – 16 с полужирным начертанием.

Приложения обозначают заглавными буквами русского алфавита, начиная с А, за исключением букв Ё, З, И, О, Ч, Ъ, Ы, Ъ. После слова «Приложение» следует буква, обозначающая его последовательность. Если в документе одно приложение, оно обозначается «Приложение А». Приложения должны иметь общую с остальной частью документа сквозную нумерацию страниц.

Иллюстрации каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Например, «рисунок А.3».

Формулы, помещаемые в приложениях, должны нумероваться отдельной нумерацией арабскими цифрами в пределах каждого приложения с добавлением перед каждой цифрой обозначения приложения, буква и цифра отделяются точкой, например «формула (В.5)».

4.8 Оформление списка использованной литературы

В начале списка приводятся законодательные и иные нормативные документы российского государства, которые располагаются в следующей последовательности:

- Конституция Российской Федерации;
- Кодексы
- Законы Российской Федерации;

- Указы Президента Российской Федерации;
- Постановления и др. акты Правительства Российской Федерации;
- Акты министерств и ведомств;
- Решения иных государственных органов и органов местного самоуправления;

Эти документы систематизируются по значимости, а внутри каждой выделенной группы документов по хронологии. В библиографии должно быть указано полное название акта, дата его принятия, номер, а также **официальный печатный источник**. Не допускается указывать в качестве источника справочные системы «Гарант», «Консультант Плюс» и др.

Научная, учебная, учебно-методическая литература, материалы периодической печати располагаются после нормативных документов единым списком. Сведения об источниках следует располагать в алфавитном порядке, нумеровать арабскими цифрами без точки и печатать с абзацного отступа.

Если при написании работы были использованы Ресурсы сети ИНТЕРНЕТ, их размещают в конце списка, в алфавитном порядке, указав авторов, адреса сайтов и дату обращения к документу.

В списке литературы используется сплошная сквозная нумерация для всех источников. При написании выпускной квалификационной работы необходимо использовать не менее 30 источников.

Пример при ссылке на законодательные материалы:

1 Российская Федерация. Конституция (1993). Конституция Российской Федерации [Текст] : офиц. текст. – М. : Маркетинг, 2001. – 39, [1] с. ; 10000 экз. – ISBN 5-94462-025-0.

2 Российская Федерация. Законы. Семейный кодекс Российской Федерации [Текст] : [федер. закон : принят Гос. Думой 8 дек. 1995 г. : по состоянию на 3 янв. 2001 г.]. – СПб. : Victory : Стаун-кантри, 2001. – 94, [1] с. ; 5000 экз. – ISBN 5-7931-0142-X.

Пример при ссылке на стандарты, правила:

1 ГОСТ Р 517721–2001. Аппаратура радиоэлектронная бытовая. Входные и выходные параметры и типы соединений. Технические требования [Текст]. – Введ. 2002–01–01. – М. : Изд-во стандартов, 2001. – 27 с.

2 ГОСТ 7. 53–2001. Издания. Международная стандартная нумерация книг [Текст]. – Взамен ГОСТ 7.53–86 ; введ. 2002–07–01. – Минск : Межгос. совет по стандартизации, метрологии и сертификации ; М. : Изд-во стандартов, 2002. – 3 с.

3 Правила безопасности при обслуживании гидротехнических сооружений и гидромеханического оборудования энергоснабжающих организаций [Текст] : РД 153-34.0-03.205–2001: утв. М-вом энергетики Рос. Федерации 13.04.01 : ввод. в действие с 01.11.01. – М. : ЭНАС, 2001. – 158 с. ; 5000 экз. – ISBN 5-93196-091-0.

Пример при ссылке на статью в журнале (газете):

1 Баутов, А. Общая оценка фондового рынка и некоторые тенденции [Текст] / А. Баутов // Страховое дело. – 2004. – № 4. – с. 2–6.

2 Гвызин, О. Л. Прогноз доходной составляющей федерального бюджета на территории субъекта РФ [Текст] : Финансы. – 2004. – № 2. – с. 18–19.

3 Голикова, Ю. С., Хохленкова, М.А. Независимость Банка России как гарантия стабильности финансово-кредитной системы [Текст] / Ю. С. Голикова, М. А. Хохленкова //: Банковское дело. – 2004. – № 4. – с. 12–15.

Пример при ссылке на книгу:

1 Бахвалов, Н. С. Численные методы [Текст] : учебное пособие для физ.-мат. специальностей вузов / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков ; под общ. ред. Н. И. Тихонова. – 2-е изд. – СПб. : Нев. диалект, 2002. – 630 с.

2 Белов, Е. Б. Основы информационной безопасности [Текст] : учебное пособие / Е. Б. Белов, В. П. Лось – М. : Горячая линия – Телеком, 2006. – 544 с.

3 Калверт, Ч. В. С / С++: Borland C++ Builder 6 [Текст] : учебник / Ч. В. Калверт, К. М. Рейсдорф – М.: Разработка программного обеспечения, 2005. – 275 с.

Пример при ссылке на WEB-ресурс:

1 Кручко, Л. А. Справочное руководство по MySQL [Электронный ресурс] : – Режим доступа: <http://www.mysql.ru/docs/man/Reference.html>. (дата обращения: 25.04.2011).

2 Мачин, Л. С. Помощь по С++: – Режим доступа: <http://www.cbuilder.ru/>. (дата обращения: 8.03.2011).

3 Рибаник, В. Л. Справочник для программиста: – Режим доступа: http://www.codenet.ru/webmast/php/php3/php3_45.php. (дата обращения: 29.02.2011).

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э.
Баумана
(национальный исследовательский университет)»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
по курсу
«Data Science»

Слушатель Грудинин Алексей Геннадьевич

Москва, 2022