

Задания на олимпиаду «Миссия выполнима» (второй этап). Москва

Задача №1. Машинное обучение. (40 баллов)

Дан файл Excel (https://palchevsky.ru/uploads/Urovni2_1_1_new_1_2.xlsx) с данными о погодных условиях за последние 5 лет, включающий столбцы: «Уровень воды», «Температура», «Влажность», «Давление», «Скорость ветра».

1. Создайте локальную базу данных SQLite с таблицей `weather_data`, в которую загрузите данные из файла Excel.

2. Реализуйте программу, которая:

- выводит статистику пропущенных данных по каждому столбцу;
- восстанавливает пропущенные данные двумя методами: `IterativeImputer` и `SimpleImputer`;
- сохраняет восстановленные данные обратно в таблицу `weather_data`.

Результат: сохраните обновленные данные в базе данных SQLite и предоставьте исходный код программы. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Задача №2. Глубокое обучение. (30 баллов)

На основе данных из таблицы `weather_data` в базе данных SQLite постройте модель Recurrent Neural Networks (RNN) для прогнозирования температуры на следующий день:

1. Извлеките данные из базы.
2. Нормализуйте их перед обучением.
3. Разделите данные на тренировочный и тестовый наборы.
4. Представьте точность прогнозов не только выводом в консоль, но и графиком (`matplotlib`).

Результат: код модели, график точности прогнозов и обновленная база данных с добавленным прогнозируемым значением температуры в таблицу `Predictions`. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Задача №3. Постобработка данных. (30 баллов)

Выполните денормализацию прогнозируемых значений, добавленных в таблицу `predictions` (из задачи 2).

1. Денормализуйте значения в удобной для восприятия форме (например, температура в градусах Цельсия).
2. Обновите таблицу `predictions`, добавив колонку `denormalized_temp` с денормализованными значениями.

Результат: код программы и база данных с обновленной таблицей `Predictions`. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Задания на олимпиаду «Миссия выполнима» (второй этап). Регионы

Задача №1. Машинное обучение. (40 баллов)

Дан файл Excel (https://palchevsky.ru/uploads/Urovni2_1_1_new_1_2.xlsx) с данными о погодных условиях за последние 5 лет, включающий столбцы: «Уровень воды», «Температура», «Влажность», «Давление», «Скорость ветра».

3. Создайте локальную базу данных SQLite с таблицей `weather_data`, в которую загрузите данные из файла Excel.

4. Реализуйте программу, которая:

- выводит статистику пропущенных данных по каждому столбцу;
- восстанавливает пропущенные данные двумя методами: `IterativeImputer` и `SimpleImputer`;
- сохраняет восстановленные данные обратно в таблицу `weather_data`.

Результат: сохраните обновленные данные в базе данных SQLite и предоставьте исходный код программы. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Задача №2. Глубокое обучение. (30 баллов)

На основе данных из таблицы `weather_data` в базе данных SQLite постройте модель Recurrent Neural Networks (RNN) для прогнозирования температуры на следующий день:

5. Извлеките данные из базы.

6. Нормализуйте их перед обучением.

7. Разделите данные на тренировочный и тестовый наборы.

8. Представьте точность прогнозов не только выводом в консоль, но и графиком (`matplotlib`).

Результат: код модели, график точности прогнозов и обновленная база данных с добавленным прогнозируемым значением температуры в таблицу `Predictions`. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Задача №3. Постобработка данных. (30 баллов)

Выполните денормализацию прогнозируемых значений, добавленных в таблицу `predictions` (из задачи 2).

3. Денормализуйте значения в удобной для восприятия форме (например, температура в градусах Цельсия).

4. Обновите таблицу `predictions`, добавив колонку `denormalized_temp` с денормализованными значениями.

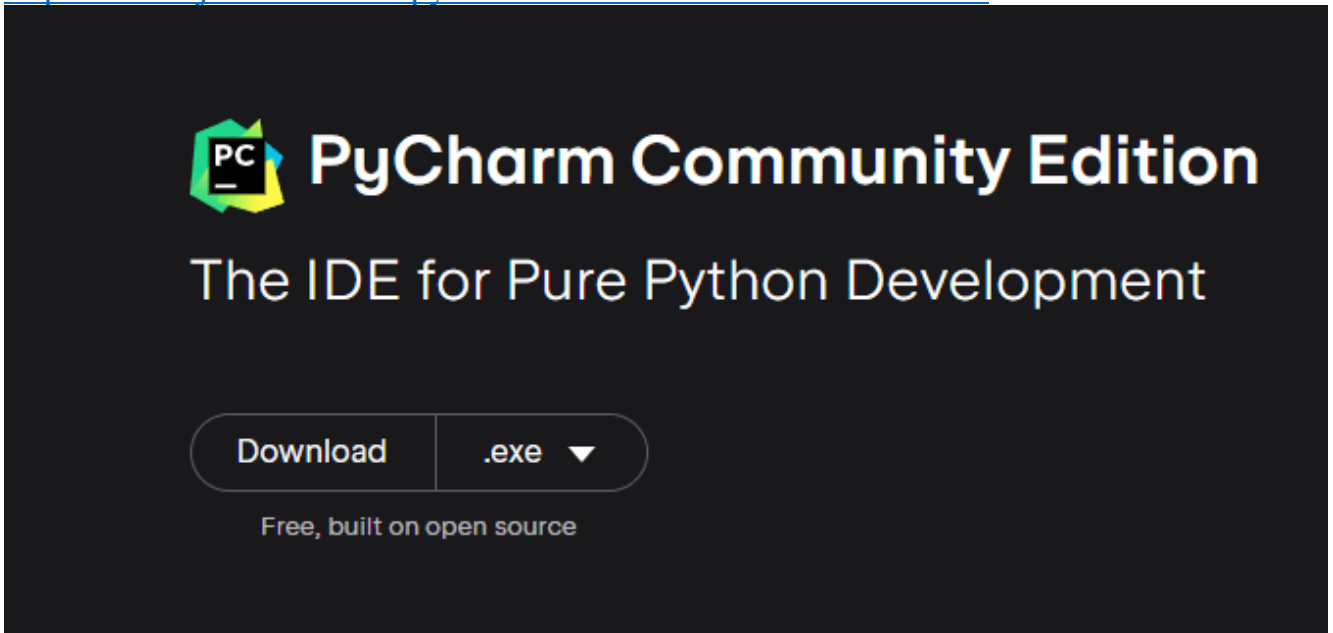
Результат: код программы и база данных с обновленной таблицей `Predictions`. В качестве ответа необходимо прислать zip-архив проекта, который содержит все файлы с исходным кодом + файл базы данных.

Инструкция

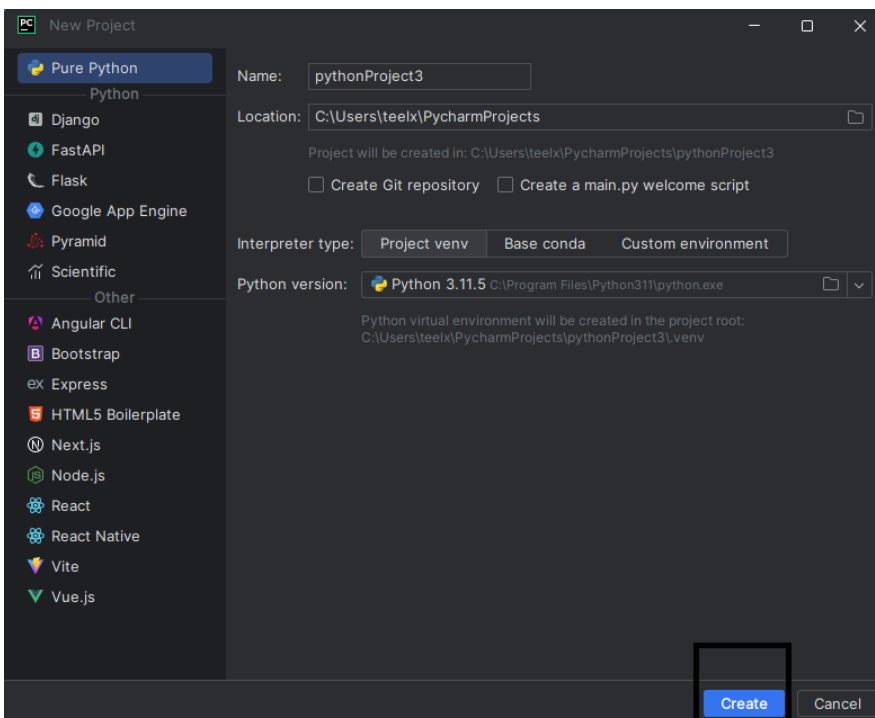
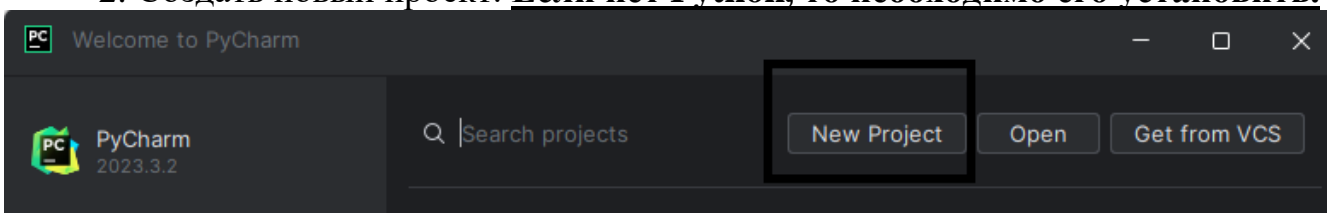
(настоятельно рекомендуется выполнить до начала олимпиады)

1. Скачать PyCharm Community по следующей ссылке:

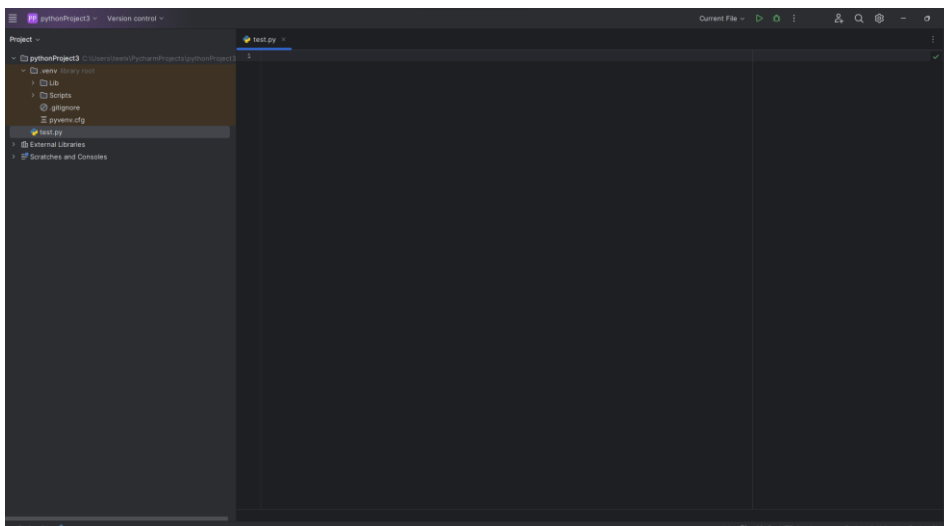
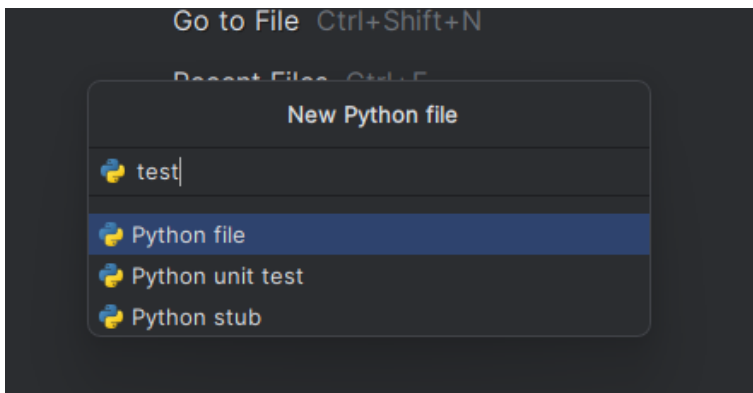
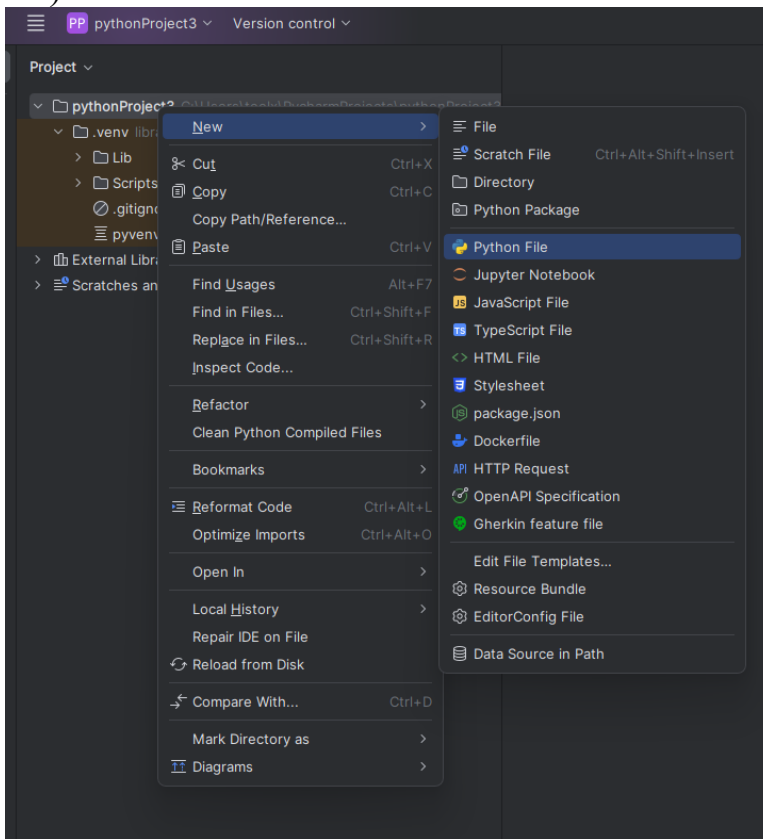
<https://www.jetbrains.com/pycharm/download/?section=windows>



2. Создать новый проект. Если нет Python, то необходимо его установить.

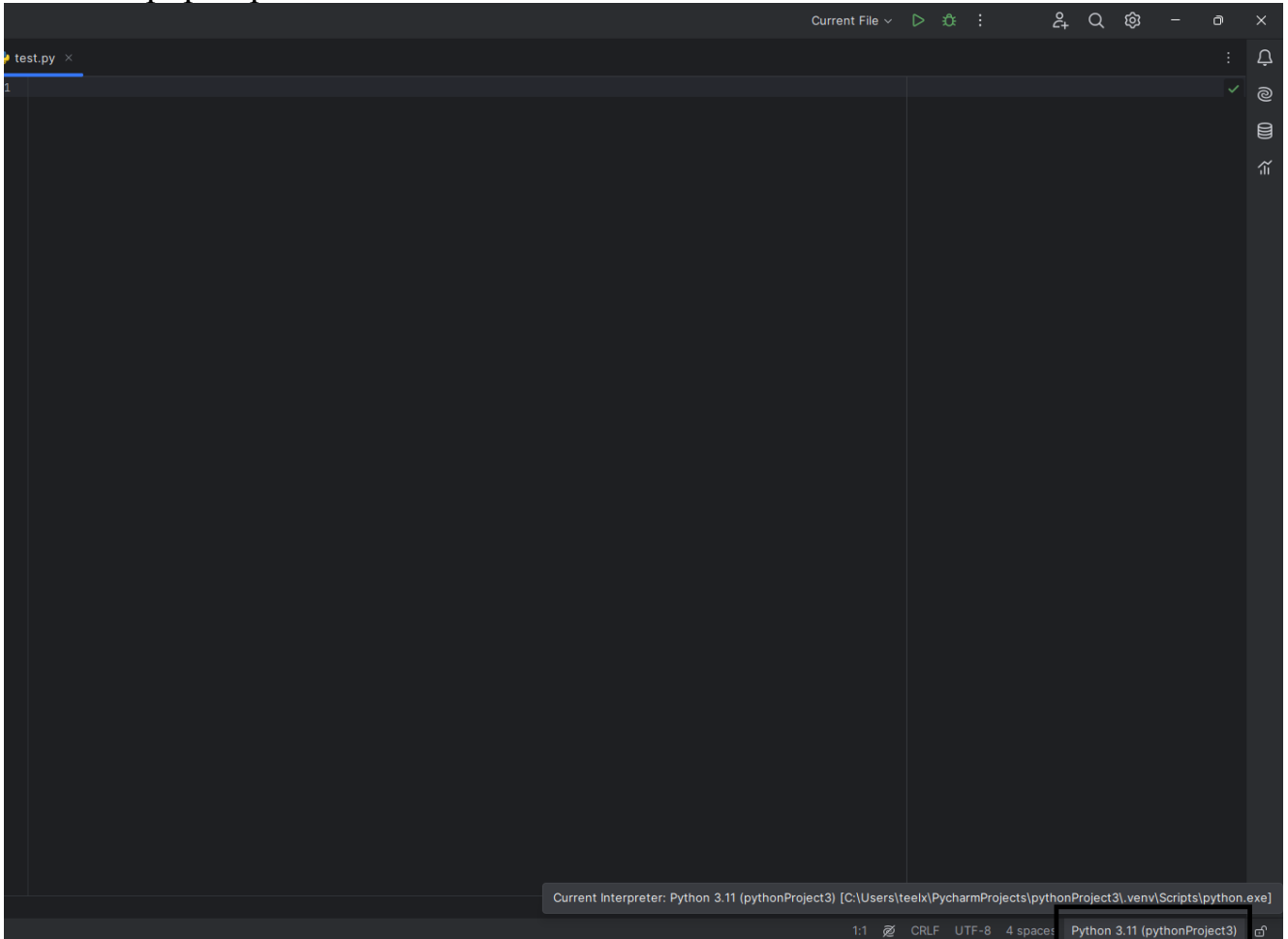


3. После создания проекта необходимо создать файл .py (если он не создан).

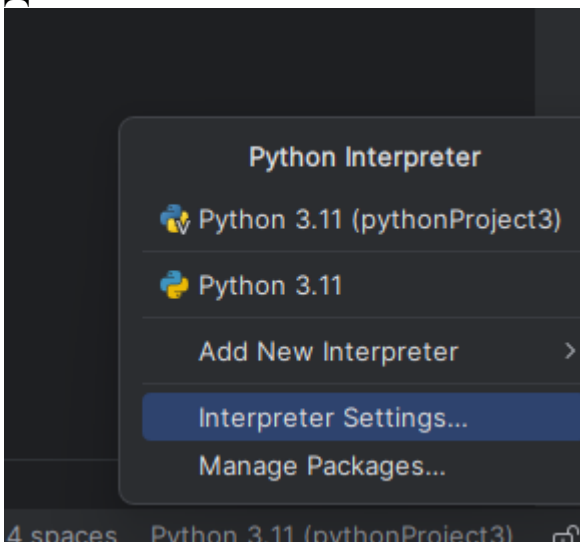


4. **Необходимо заранее установить следующие библиотеки:** scikit-learn, tensorflow, torch, pandas, opencv. Необходим выход в интернет. Перед установкой необходимо обновить pip до последней версии. Все библиотеки корректно работают с Python версии 3.11.5 (проверено).

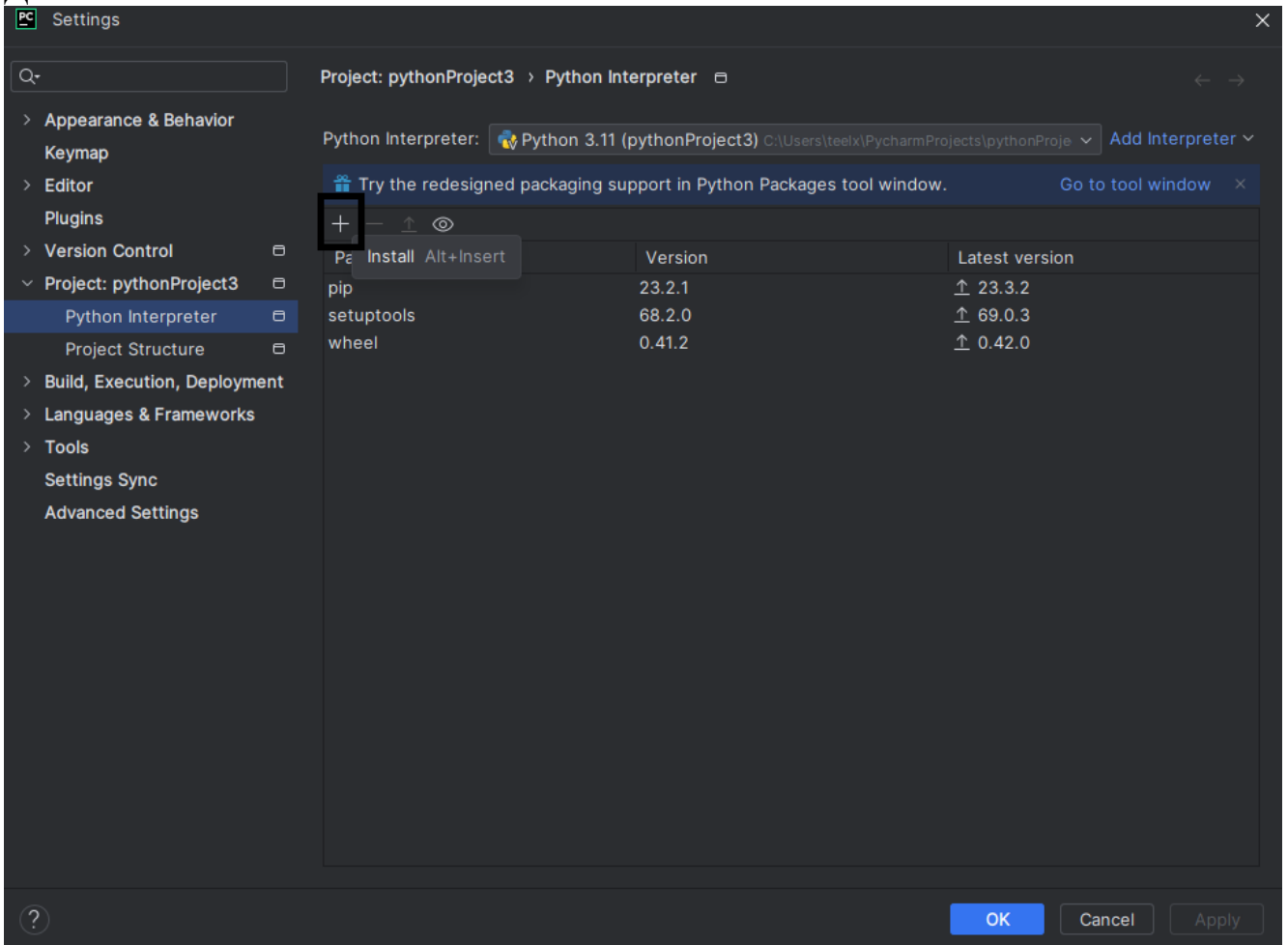
Для установки каждой библиотеки в готовом проекте необходимо выбрать наш интерпретер:



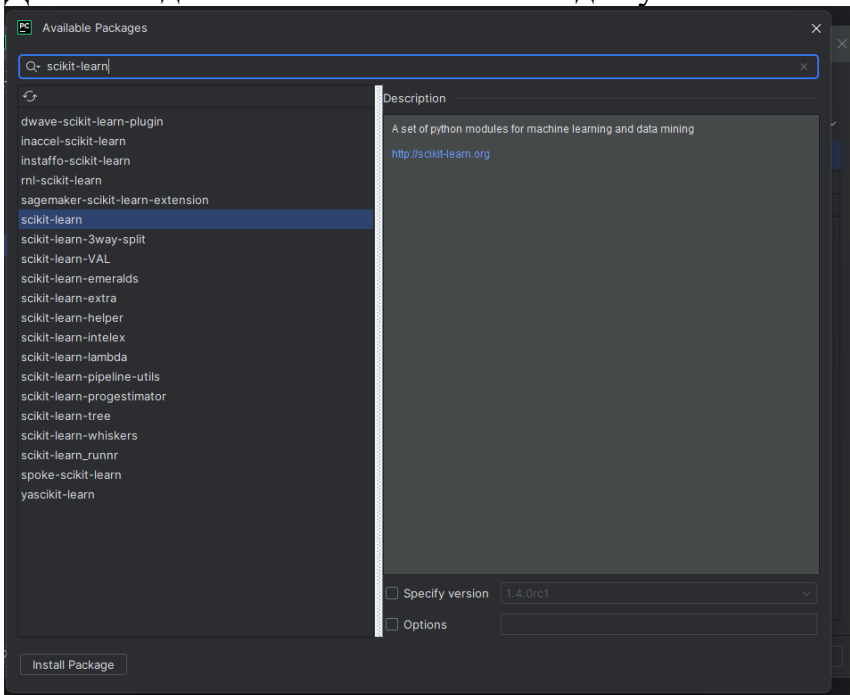
Далее:



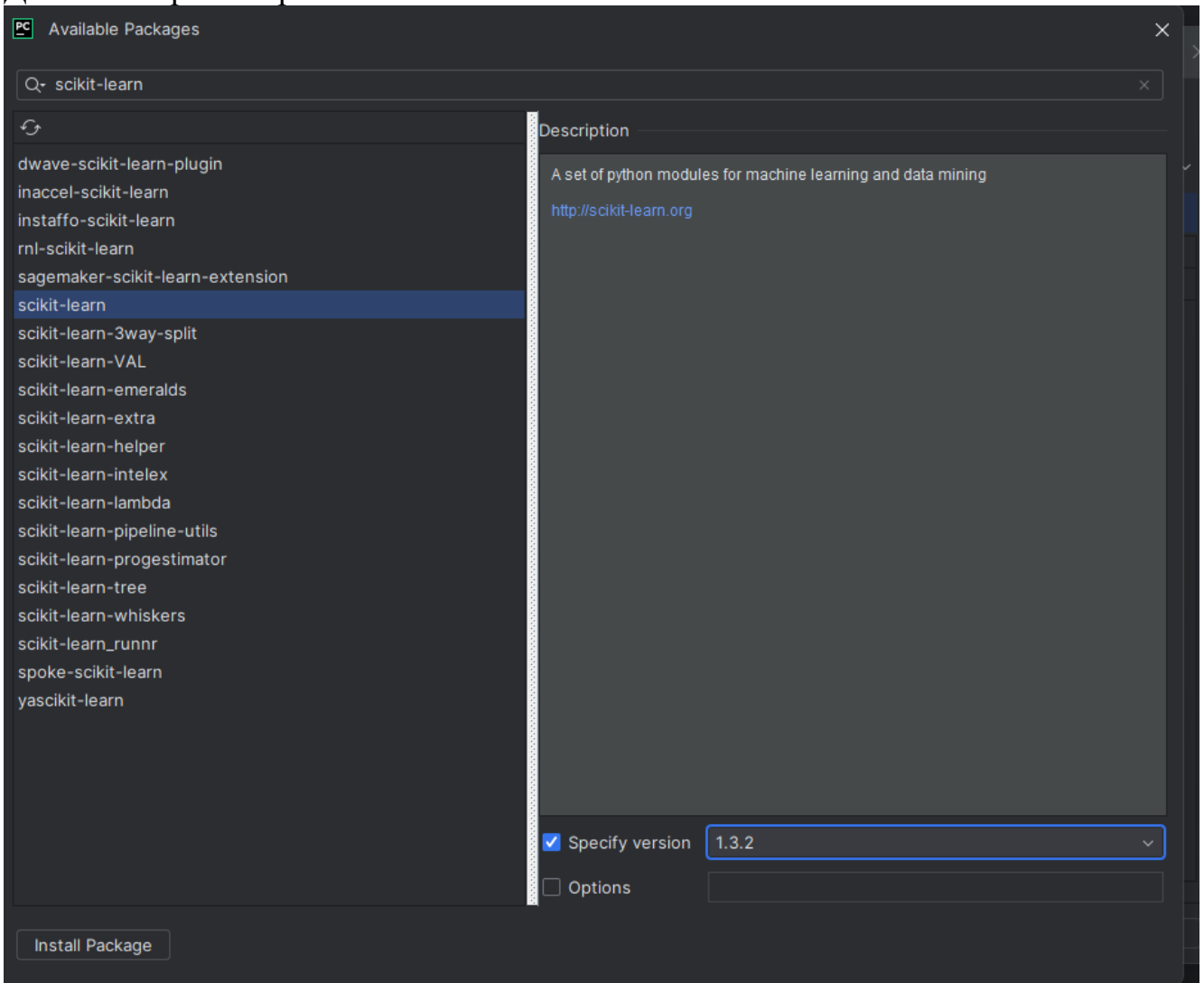
Далее:



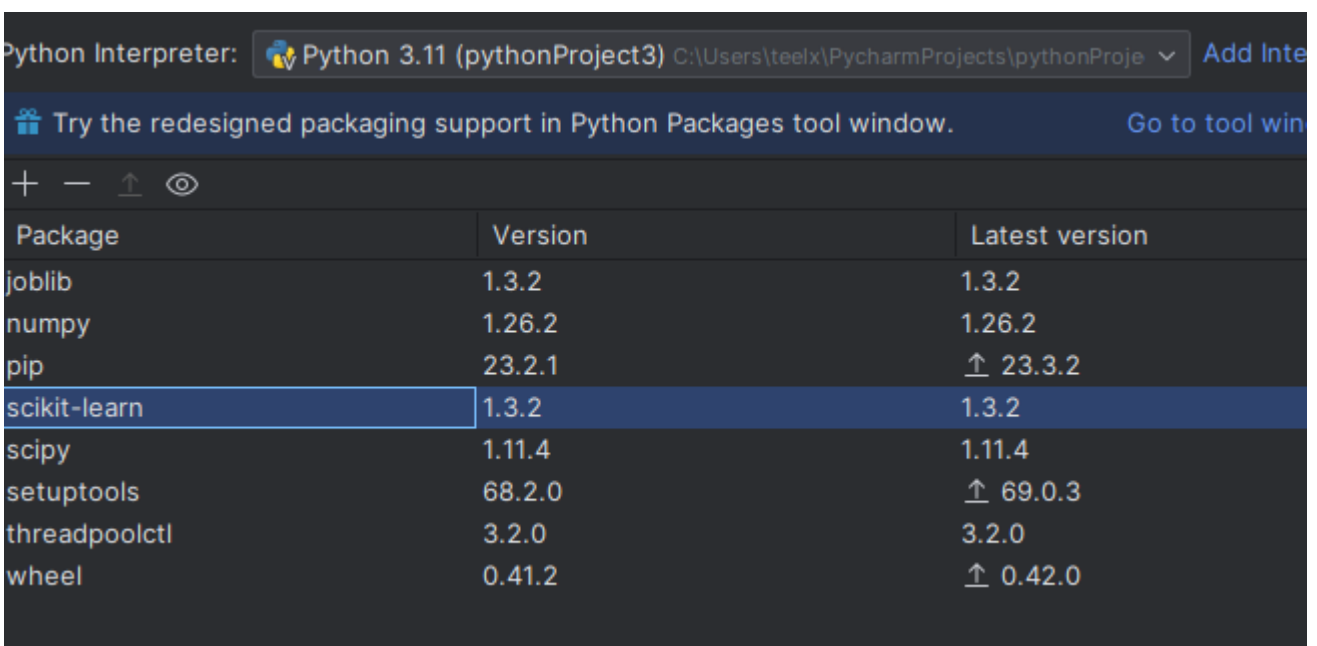
Далее вводим в окно поиска необходимую нам библиотеку:



Далее выбираем версию библиотеки:



И нажимаем «Install Package».



Готово. Библиотека установлена.