

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОНИКИ И МАТЕМАТИКИ ИМ. А.Н.
ТИХОНОВА

ЗАДАНИЕ
для самостоятельной работы по индивидуальному плану
по курсу
«Проектный семинар “Python в науке о данных”»

Студента группы БИВ225
Пашенцева Павла Владимировича, rvpashentsev_1@edu.hse.ru, +79153309856

Тема работы:
Приложение для анализа роста компаний во время пандемии

Руководитель:
Полякова Марина Васильевна

Москва 2023

Участники договора

Настоящий документ является техническим заданием (далее ТЗ), которое формулируется преподавателем Поляковой Мариной Васильевной (далее Заказчик) для выполнения студентом Пашенцевым Павлом Владимировичем (далее Исполнитель).

Права исполнителя

Исполнитель полностью освобождается от посещения занятий и участия в любых контрольных мероприятиях. Оценка за курс исполнителя совпадает с оценкой за проект.

Полноценная работоспособная реализация данного ТЗ высокого качества гарантирует получение оценки девять или десять баллов (отлично).

Нарушение сроков и положений ТЗ, а также наличие ошибок, ведёт к адекватному снижению оценки.

Решаемая задача

Исполнитель создаёт приложение на платформе Windows 10 с графическим интерфейсом и обученной моделью искусственного интеллекта(модель обучения – контролируемое обучение) прогнозирует ориентировочную цену алмаза по входным характеристикам, а также производит анализ данных и построение графиков в исходном датасете.

Тема

Приложение для анализа роста компаний во время пандемии, основываясь на данных в рамках выбранного датасета.

(<https://www.kaggle.com/datasets/nguyenthicamlai/top-1000-fastest-growing-companies-in-europe>)

Вычислительная платформа и используемые библиотеки

Для реализации задачи используется язык программирования Python версии 3.9. Актуальный на начало курса дистрибутив Anaconda — Anaconda3- 2022.05-Windows-x86_64.exe будет использоваться в качестве основного.

Использованные библиотеки с указанием их версий:

- Pandas версия 1.5.3
- NumPy версия 1.23.0
- Matplotlib 3.6.3

Функционал приложения

Приложение обеспечивает:

- Возможность фильтрации данных по ключевым словам
- Прогноз роста компании по вводимым характеристикам
- Анализ обработанных данных, построение графиков на основе загруженного датасета

Архитектура приложения и требования к коду

Приложение должно запускаться из командной строки запуском на счет основного модуля или командного файла. Установка приложения, в частности создание при необходимости виртуальной среды, должна быть полностью автоматизирована, т.е. осуществляться запуском на счет модуля Python или командного файла. Установка приложения и его эксплуатация не должны требовать выполнения системных операций, связанных с

изменениями операционной системы, актуального дистрибутива Anaconda и прочих элементов вычислительной среды пользователя. В частности, развертывания ПО Docker и аналогичных инструментов.

Для работы приложения должна быть создана виртуальная среда исключительно средствами менеджера пакетов conda из актуального дистрибутива Anaconda. Таким образом, приложение должно работать в собственной виртуальной среде на любом компьютере, соответствующем требованиям данного документа, без необходимости добавления библиотек к актуальному дистрибутиву Anaconda. Приложение является локальным приложением с графическим интерфейсом, которое работает на изолированном компьютере. Это означает, что для работы приложения не нужны вычислительные ресурсы Интернета и не используется клиент-серверная архитектура (за исключением Web-интерфейса). В тоже время, приложение может анализировать данные из Интернета. Также приложение должно исполняться на компьютере средней мощности с 8Gb оперативной памяти и 2Gb видеопамея не очень мощной видеокарты. Приложение реализовано в виде основного модуля и библиотеки, оформленной в соответствии с требованиями Python и графического интерфейса, который использует функции данной библиотеки. Библиотека образована модулями, содержащими тематически близкие функции/классы. Каждый модуль содержит не более 300 строк. Код разработан в соответствии с требованиями «Python Enhancement Proposals (PEP) 8 -- Style Guide for Python Code» и «PEP 257 -- Docstring Conventions» (см. <https://www.python.org/dev/peps/>).

Описание интерфейса

Приложение имеет графический интерфейс, адаптированный для реализации указанного выше функционала. В частности:

- Графики, представляющиеся пользователю, для визуального понимания представленного анализа
- Формы для ввода в виде текста

Сопроводительная документация

Исполнитель создает два документа: «Руководство пользователя», содержащее подробное описание решаемой задачи, инсталляции, запуска и интерфейса приложения, и «Руководство разработчика», содержащее подробное описание архитектуры приложения — состав библиотек, функций и классов, связь между ними, структура каталогов и описание параметров, управляющих работой приложения. Документация формируется в виде двух файлов формата MS Word 2016 или Adobe Acrobat.

Сроки предоставления результатов

- Подбор алгоритмов
- Обучение модели
- Тестирование
- Анализ результатов