



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)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ Информатика и системы управления _____

КАФЕДРА _____ Системы обработки информации и управления _____

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОМУ ПРОЕКТУ

НА ТЕМУ:

_____ решение комплексной задачи _____
_____ машинного обучения _____

Студент _____ ИУ5-65Б _____
(Группа)

_____ Камалов М. Р. _____
(Подпись, дата) (И.О.Фамилия)

Руководитель курсового проекта

_____ Гапанюк Ю.Е. _____
(Подпись, дата) (И.О.Фамилия)

Консультант

_____ _____
(Подпись, дата) (И.О.Фамилия)

2021 г.

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)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Заведующий кафедрой _____
(Индекс)

(И.О.Фамилия)
« ____ » _____ 20 ____ г.

З А Д А Н И Е
на выполнение курсового проекта

по дисциплине _____ Технологии машинного обучения _____

Студент группы ИУ5-65Б

_____ Камалов Марат Равилевич _____
(Фамилия, имя, отчество)

Тема курсового проекта _____
решение комплексной задачи машинного обучения _____

Направленность КП (учебный, исследовательский, практический, производственный, др.)
_____ учебный _____

Источник тематики (кафедра, предприятие, НИР) _____ кафедра _____

График выполнения проекта: 25% к ____ нед., 50% к ____ нед., 75% к ____ нед., 100% к ____ нед.

Задание _____ решение задачи машинного обучения на основе материалов дисциплины _____

Оформление курсового проекта:

Расчетно-пояснительная записка на 23 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

Дата выдачи задания « ____ » _____ 20 ____ г.

Руководитель курсового проекта

_____ Гапанюк Ю.Е. _____
(Подпись, дата) (И.О.Фамилия)

Студент

_____ Камалов М. Р. _____
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Содержание

1. Введение.....	4
2. Постановка задачи.....	4
3. Выполнение поставленной задачи	5
4. Заключение	Ошибка! Закладка не определена.
5. Список использованных источников информации	Ошибка! Закладка не определена.

Введение

Курсовой проект – самостоятельная часть учебной дисциплины «Технологии машинного обучения» – учебная и практическая исследовательская студенческая работа, направленная на решение комплексной задачи машинного обучения. Результатом курсового проекта является отчет, содержащий описания моделей, тексты программ и результаты экспериментов.

Курсовой проект опирается на знания, умения и владения, полученные студентом в рамках лекций и лабораторных работ по дисциплине.

В рамках курсового проекта возможно проведение типового или нетипового исследования.

- Типовое исследование - решение задачи машинного обучения на основе материалов дисциплины. Выполняется студентом единолично.
- Нетиповое исследование - решение нестандартной задачи. Тема должна быть согласована с преподавателем. Как правило, такая работа выполняется группой студентов.

Постановка задачи

Схема типового исследования, проводимого студентом в рамках курсовой работы, содержит выполнение следующих шагов:

1. Поиск и выбор набора данных для построения моделей машинного обучения. На основе выбранного набора данных студент должен построить модели машинного обучения для решения или задачи классификации, или задачи регрессии.
2. Проведение разведочного анализа данных. Построение графиков, необходимых для понимания структуры данных. Анализ и заполнение пропусков в данных.
3. Выбор признаков, подходящих для построения моделей. Кодирование категориальных признаков. Масштабирование данных. Формирование вспомогательных признаков, улучшающих качество моделей.
4. Проведение корреляционного анализа данных. Формирование промежуточных выводов о возможности построения моделей машинного обучения. В зависимости от набора данных, порядок выполнения пунктов 2, 3, 4 может быть изменен.
5. Выбор метрик для последующей оценки качества моделей. Необходимо выбрать не менее трех метрик и обосновать выбор.
6. Выбор наиболее подходящих моделей для решения задачи классификации или регрессии. Необходимо использовать не менее пяти моделей, две из которых должны быть ансамблевыми.

7. Формирование обучающей и тестовой выборок на основе исходного набора данных.
8. Построение базового решения (baseline) для выбранных моделей без подбора гиперпараметров. Производится обучение моделей на основе обучающей выборки и оценка качества моделей на основе тестовой выборки.
9. Подбор гиперпараметров для выбранных моделей. Рекомендуется использовать методы кросс-валидации. В зависимости от используемой библиотеки можно применять функцию GridSearchCV, использовать перебор параметров в цикле, или использовать другие методы.
10. Повторение пункта 8 для найденных оптимальных значений гиперпараметров. Сравнение качества полученных моделей с качеством baseline-моделей.
11. Формирование выводов о качестве построенных моделей на основе выбранных метрик. Результаты сравнения качества рекомендуется отобразить в виде графиков и сделать выводы в форме текстового описания. Рекомендуется построение графиков обучения и валидации, влияния значений гиперпараметров на качество моделей и т.д.

Выполнение поставленной задачи

About this dataset

- Age : Age of the patient
- Sex : Sex of the patient
- exang: exercise induced angina (1 = yes; 0 = no)
- ca: number of major vessels (0-3)
- cp : Chest Pain type chest pain type
 - Value 1: typical angina
 - Value 2: atypical angina
 - Value 3: non-anginal pain
 - Value 4: asymptomatic
- trtbps : resting blood pressure (in mm Hg)
- chol : cholestoral in mg/dl fetched via BMI sensor
- fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- rest_ecg : resting electrocardiographic results
 - Value 0: normal
 - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
 - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalach : maximum heart rate achieved
- target : 0= less chance of heart attack 1= more chance of heart attack



ИУ5-65Б Камалов Марат КР

Данные загружены!

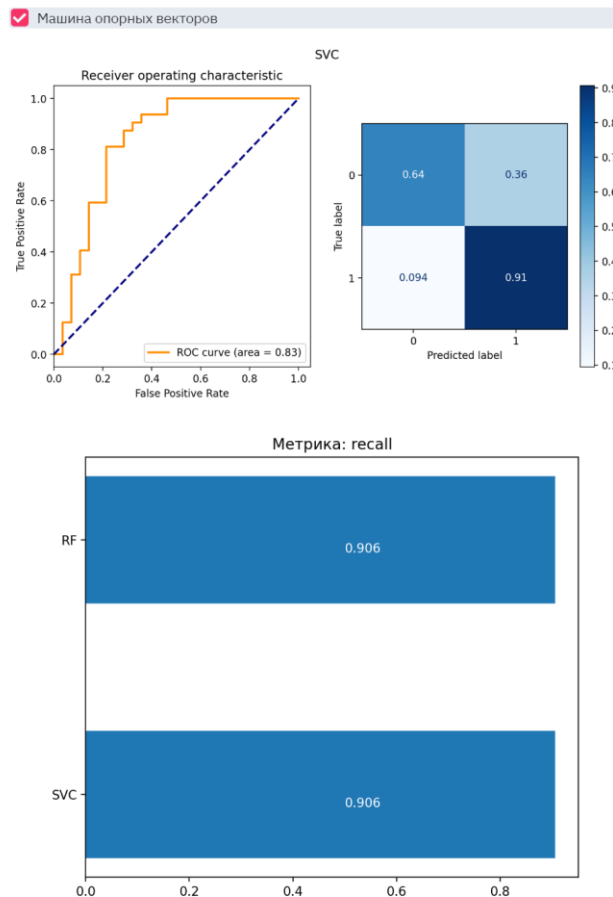
Первые 5 значений обучающей выборки

	cp	thalach	exang	oldpeak	slope	ca	thal	target
0	0	140	0	4.4000	0	3	1	0
1	2	162	0	0.5000	2	0	3	1
2	0	111	1	0	2	0	3	1
3	2	151	0	1	2	1	3	1
4	2	146	0	1.8000	1	3	3	1

Первые 5 значений тестовой выборки

	cp	thalach	exang	oldpeak	slope	ca	thal	target
0	0	132	1	1.2000	1	1	3	0
1	0	143	0	0.4000	1	0	2	1
2	1	175	0	0.6000	1	0	2	1
3	1	164	0	0	2	1	2	0
4	3	125	0	0	2	0	2	0

- ☐ Показать статистические характеристики
- ☐ Дисбаланс классов для Target
- ☐ Тепловая карта
- ☐ Логистическая регрессия
- ☐ Метод ближайших соседей
- ☐ Машина опорных векторов
- ☐ Решающее дерево
- ☐ Случайный лес
- ☐ Градиентный бустинг
- ☐ Оценка качества



Заключение

В ходе выполнения курсового проекта был сделан вывод о том, лучшей оказалась модель случайного леса

Список использованных источников информации

1. Видеотрансляции лекций курса «Технологии машинного обучения».
<https://www.youtube.com/playlist?list=PL9vFTJYocFHoRvWr-jiC9r3i9jayLzvPg>
2. Блог Александра Дьяконова. Стекинг (Stacking) и блендинг (Blending). <https://dyakonov.org/2017/03/10/стекинг-stacking-и-блендинг-blending>
3. Ансамблевые методы: бэггинг, бустинг и стекинг.
<https://neurohive.io/ru/osnovy-data-science/ansamblevye-metody-begging-busting-i-steking>
4. Автоматическое машинное обучение.
<https://github.com/mljar/mljar-supervised>

5. Блог Александра Дьяконова. Случайный лес (Random Forest).
<https://dyakonov.org/2016/11/14/случайный-лес-random-forest/comment-page-1>
6. Руководство scikit-learn. <https://scikit-learn.org/stable/index.html>