



Саратовский национальный исследовательский
государственный университет имени Н.Г. Чернышевского

Программа повышения квалификации «Интеллектуальная обработка данных с использованием технологий машинного обучения»

Кудрина Е.В.,

доцент кафедры информатики и программирования,
сотрудник центра непрерывной подготовки IT-специалистов,
заместитель декана факультета компьютерных наук и информационных технологий по
учебной работе.

2019 год

Требования к преподавателям вузов:

- руководствоваться в педагогической деятельности не только требованиями ФГОС ВО, но и требованиями профессиональных стандартов;
- быть ориентированными не только на методическое обеспечение учебного процесса и совершенствование методов обучения, но и на приобретение новых знаний в предметной области, и освоение инновационных технологий;
- не только сами активно участвовать в научных исследованиях, но и привлекать к ним студентов.

Квалификация преподавателя вуза определяется тем, насколько полно преподаватель соответствует вышеперечисленным требованиям.

Программы повышения квалификации для преподавателей на базе ИДПО СГУ

1. Деятельностная педагогика в профессиональном образовании.
2. Инновационное содержание гуманитарных дисциплин в условиях реализации ФГОС.
3. Инструменты маркетинга в системе управления.
4. Развитие профессиональных компетенций преподавателей в условиях системных изменений в высшем образовании.
5. Современные образовательные технологии в профессиональном образовании.
6. Современные подходы к управлению персоналом в вузе.
7. Создание инклюзивной образовательной среды в профессиональном образовании.
8. Социокультурные факторы и проблемы современного образования.
9. Технологии разработки, реализации и управления проектами.
10. Управление организационными изменениями.
11. Экономика и менеджмент в образовательной организации.
12. Формирование организационной культуры студентов в условиях современного университетского комплекса.
13. Правовые аспекты трудового законодательства.

Специализированные программы повышения квалификации для преподавателей факультета КНиИТ

- 1. Использование электронной информационно-образовательной среды вуза и информационно-коммуникационных технологий в образовательном процессе СГУ (108 часов, 2015-2017 г.г.)**
 - Обучено более 250 преподавателей факультета, а так же СГУ и Балашовского института СГУ.
 - Адаптирована под потребности других вузов Саратова: ФГБОУ ВО «Саратовский государственный медицинский университет имени В.И. Разумовского» - более 30 преподавателей; ФГКВОУ ВО «Саратовский военный Краснознаменный институт войск национальной гвардии Российской Федерации» - около 30 преподавателей.
- 2. Информационная безопасность. Научно-педагогическая деятельность по проектированию и реализации образовательного процесса (612 часов, 2017 г., совместно с ООО «Национальный центр по борьбе с преступлениями в сфере высоких технологий», ООО «Астрант»)**
 - Обучено 13 преподавателей факультета.
- 3. Язык программирования Python: базовый курс (108 часов, 2017 г.)**
 - Обучено 16 преподавателей факультета и 5 учителей информатики г. Саратова.
- 4. Промышленная разработка программного обеспечения (108 часов, 2017-2018 г.г., совместно с ООО «Эпам Системз»)**
 - Обучено 27 преподавателей факультета.

Новая программа повышения квалификации «Интеллектуальная обработка данных с использованием технологий машинного обучения»

(108 часов, 2019 год, совместно с Обособленным подразделением «Грид Динамикс 2» в г.Саратов)

Авторы программы:

- Лось Илья Викторович, ассистент кафедры математических основ информатики и олимпиадного программирования на базе МАОУ «ФТЛ №1» г. Саратова, старший инженер-программист Обособленного подразделения «Грид Динамикс 2» в г.Саратов.
- Семенов Михаил Сергеевич, доцент кафедры математической кибернетики и компьютерных наук, технический менеджер 2 категории Обособленного подразделения «Грид Динамикс 2» в г.Саратов.

Адаптация программы к образовательным и научно-исследовательским задачам факультета:

- Кудрина Елена Вячеславовна, доцент кафедры «Информатика и программирование»

Учебный план программы

№ пп	Наименование модулей	Всего, час.	В том числе:			Форма промежуточной /итоговой аттестации
			Лекции	Практические занятия	Самостоя- тельная работа	
1.	Машинное обучение	36	6	6	24	практическое задание
2.	Интеллектуальная обработка изображений с использованием технологий машинного обучения	24	4	4	16	практическое задание
3.	Интеллектуальная обработка естественного языка с использованием технологий машинного обучения	24	4	4	16	практическое задание
4.	Подготовка к итоговой аттестации	24			24	зачет
Итого		108	14	14	80	

Структура и содержание программы

Неделя обучения	Модуль	Темы	Содержание
1-3 недели	Машинное обучение	Тема 1.1. Теоретические основы машинного обучения	Постановки основных классов задач в машинном обучении. Обучение с учителем: регрессия и классификация. Обучение без учителя: кластеризация. Решающие деревья и их применение для решения задач регрессии, классификации и кластеризации.
		Тема 1.2. Python-библиотеки для решения задач машинного обучения	Обзор Python-библиотек для решения задач машинного обучения: NumPy, SciPy, SciKit-Learn, Theano, TensorFlow, Keras, NLTK, OpenCV.
		Тема 1.3. Нейронные сети	Базовые определения и понятия теории нейронных сетей. Применение нейронных сетей в машинном обучении.

Тематика практических занятий:

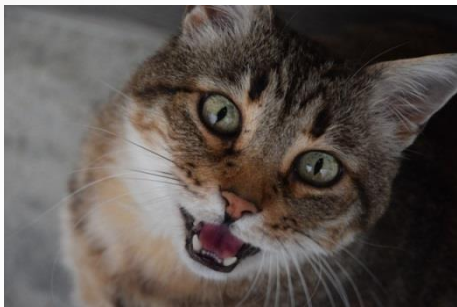
1. Конфигурирование пайплайна: Cats vs Dogs.
2. Логистическая регрессия и решающие деревья на фичах ResNet/VGG.
3. Обучение нейронной сети на MNIST

На лекционных занятиях особое внимание уделяется:

- математическому аппарату машинного обучения во взаимосвязи с математическим анализом, теорией вероятностей и математической статистикой, численными методами, методами оптимизации и т.д.;
- обзору функциональных возможностей Python-библиотек (NumPy, SciPy, SciKit-Learn, Theano, TensorFlow, Keras, NLTK, OpenCV) для решения задач машинного обучения;
- теоретическим и прикладным аспектам применения машинного обучения для проведения исследований в различных отраслях экономики;
- вопросам встраивания материала Программы в учебные курсы различных образовательных программ подготовки IT-специалистов в СГУ.

Практические задания:

➤ Что изображено на картинке – кот или собака?



Практические занятия:

направлены на формирование у слушателей
Программы навыков решения учебных и прикладных
задач машинного обучения с использованием Jupyter
NoteBook среды Anaconda для Python 3.6/3.7.

Скриншот 1: Чтение данных

Практика 1. Пайплайн для задачи cats-vs-dogs Last Checkpoint: 2 minutes ago (autosaved)

Обработка данных. Этот шаг я сделаю за вас в силу некоторой его трудоемкости. Используем библиотеки pandas, numpy и Pillow.

- Введение в pandas и numpy можно посмотреть в первом уроке курса mlcourse_open:
- https://github.com/Yorko/mlcourse.ai/blob/master/jupyter_russian/topic01_pandas_data_analysis/lesson1_part0_numpy.ipynb
- https://github.com/Yorko/mlcourse.ai/blob/master/jupyter_russian/topic01_pandas_data_analysis/lesson1_part1_pandas_intro.ipynb
- Pillow - одна из самых распространенных библиотек для работы с изображениями в python.

```
In [5]: import pandas as pd
import pathlib
import hashlib
import numpy as np
import random

train_directory = pathlib.Path("C:/Users/contest/Downloads/data")
sample_size = 5000

def initialize_random_seed():
    """Инициализирует ГПСЧ из STUDENT_ID"""
    sha256 = hashlib.sha256()
    sha256.update(STUDENT_ID.encode("utf-8"))

    fingerprint = int(sha256.hexdigest(), 16) % (2**32)

    random.seed(fingerprint)
    np.random.seed(fingerprint)

def read_target_variable():
    """Прочитаем разметку фотографий из названий файлов"""
    target_variable = {
        "filename": [],
        "is_cat": []
    }
    image_paths = list(train_directory.glob("*.jpg"))
    random.shuffle(image_paths)
    for image_path in image_paths[:sample_size]:
        filename = image_path.name
        class_name = filename.split(".")[0]
        target_variable["filename"].append(filename)
        target_variable["is_cat"].append(class_name == "cat")

    return pd.DataFrame(data=target_variable)

initialize_random_seed()

target_df = read_target_variable()
target_df
```

Скриншот 2: Обработка изображений

Практика 1. Пайплайн для задачи cats-vs-dogs Last Checkpoint: 7 minutes ago (unsaved changes)

Используем изображения в качестве признакового описания. Признаками будут значения конкретных пикселей.

```
In [6]: from PIL import Image
from tqdm import tqdm_notebook

def read_data(target_df):
    """Читаем данные изображений и строим их признаковое описание"""
    image_size = (100, 100)
    features = []
    target = []
    for i, image_name, is_cat in tqdm_notebook(target_df.iterrows(), total=len(target_df)):
        image_path = str(train_directory / image_name)
        image = Image.open(image_path)
        image = image.resize(image_size) # уменьшаем изображения
        image = image.convert('LA') # преобразуем в ч/б
        pixels = np.asarray(image)[:, :, 0]
        pixels = pixels.flatten()
        features.append(pixels)
        target.append(is_cat)
    return np.array(features), np.array(target)

features, target = read_data(target_df)
features
```

Скриншот 3: Вывод данных

Практика 1. Пайплайн для задачи cats-vs-dogs Last Checkpoint: 5 minutes ago (autosaved)

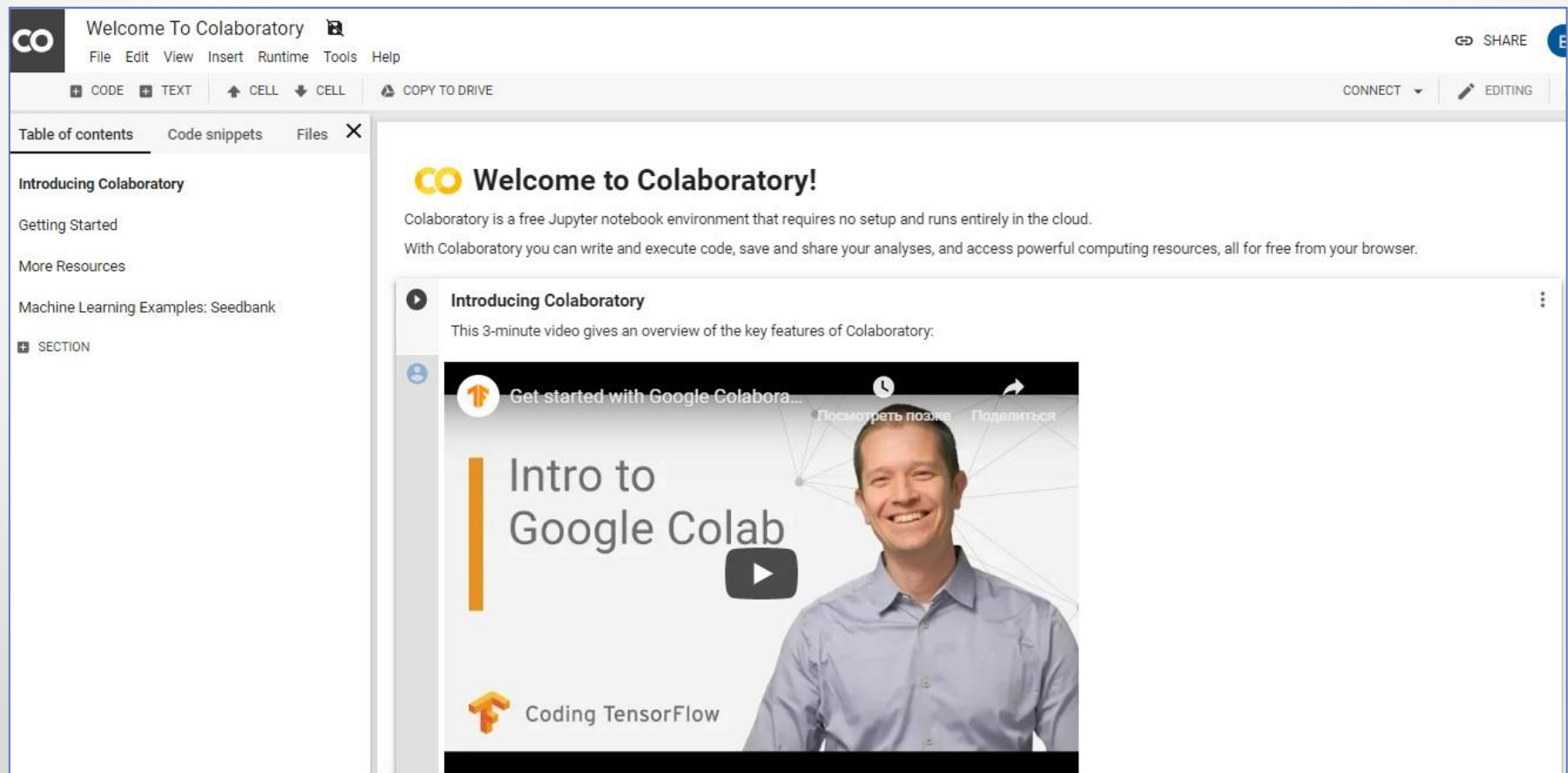
Out[5]:

	filename	is_cat
0	dog.12485.jpg	False
1	cat.11197.jpg	True
2	dog.1736.jpg	False
3	dog.8056.jpg	False
4	cat.3647.jpg	True
5	cat.7025.jpg	True
6	cat.6598.jpg	True
7	cat.3441.jpg	True
8	cat.11239.jpg	True
9	cat.7336.jpg	True

Out[6]: array([[101, 110, 119, ..., 38, 40, 42],
[48, 66, 24, ..., 127, 124, 131],
[0, 10, 18, ..., 51, 60, 8],
...,
[41, 41, 41, ..., 93, 93, 91],
[122, 79, 76, ..., 93, 92, 93],
[213, 210, 209, ..., 170, 173, 171]], dtype=uint8)

Практические занятия:

направлены на формирование навыков использования бесплатной среды Colaboratory (<https://colab.research.google.com/>) для Jupyter Notebook, которая не требует специальной настройки и полностью работает в облаке Google.



The screenshot displays the Google Colaboratory web interface. At the top, a navigation bar includes the 'co' logo, a 'Welcome To Colaboratory' message, and a menu with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. To the right of the menu are 'SHARE' and 'CONNECT' buttons. Below the navigation bar, a toolbar shows icons for 'CODE', 'TEXT', 'CELL' (with up and down arrows), and 'COPY TO DRIVE'. A sidebar on the left contains a 'Table of contents' section with links to 'Introducing Colaboratory', 'Getting Started', 'More Resources', and 'Machine Learning Examples: Seedbank', followed by a 'SECTION' header. The main content area features a 'Welcome to Colaboratory!' heading, a brief description of the service, and a video player titled 'Introducing Colaboratory'. The video thumbnail shows a man smiling and includes the text 'Intro to Google Colab' and 'Coding TensorFlow'.

co Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

CODE TEXT CELL CELL COPY TO DRIVE

CONNECT EDITING

Table of contents Code snippets Files

Introducing Colaboratory

Getting Started

More Resources

Machine Learning Examples: Seedbank

SECTION

co Welcome to Colaboratory!

Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.

With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.

Introducing Colaboratory

This 3-minute video gives an overview of the key features of Colaboratory:

Get started with Google Colabora...

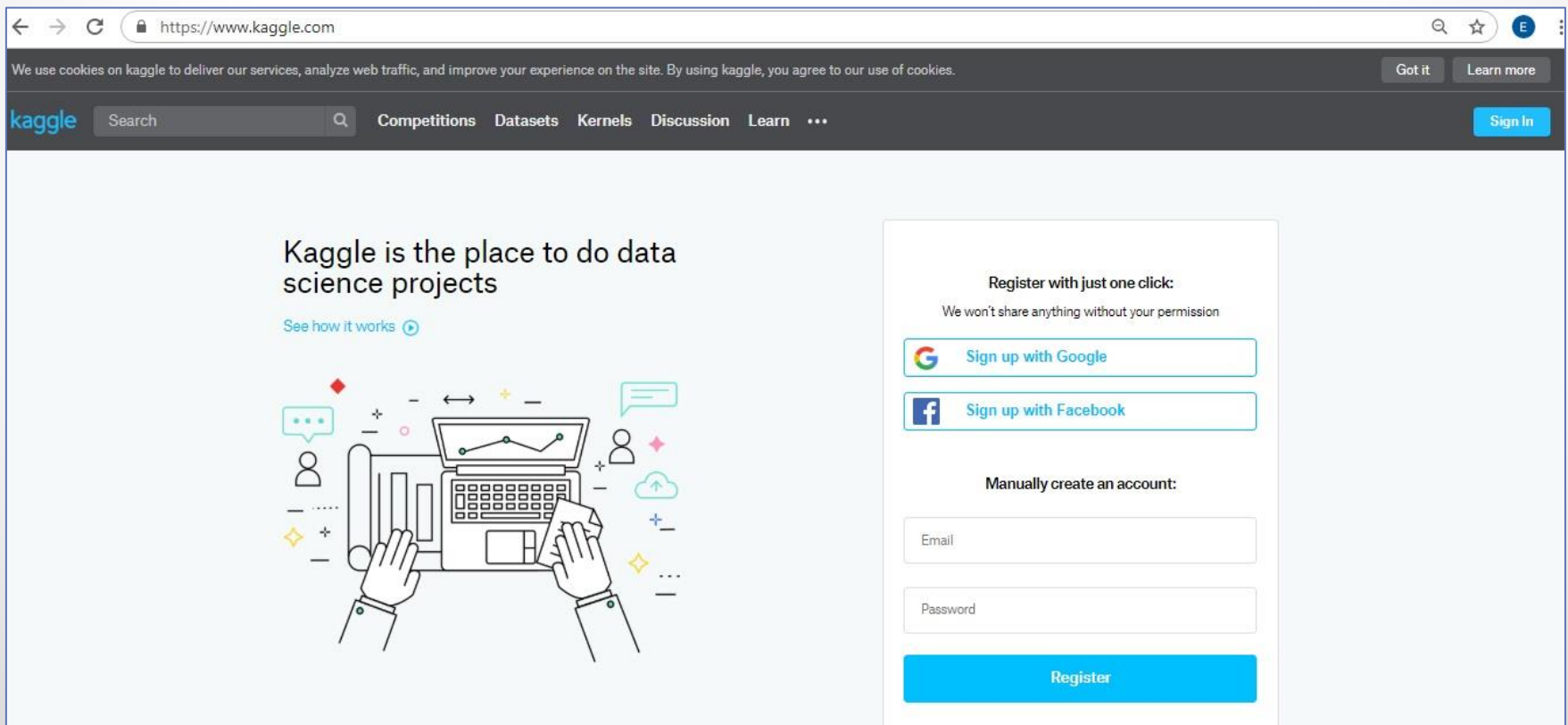
Intro to Google Colab

Coding TensorFlow

Практические занятия:

направлены на формирование навыков использования ресурсов платформы Kaggle (<https://www.kaggle.com/>):

- наборы размеченных данных;
- учебные курсы по машинному обучению
- конкурсы и исследовательские проекты по машинному обучению.



Структура и содержание программы

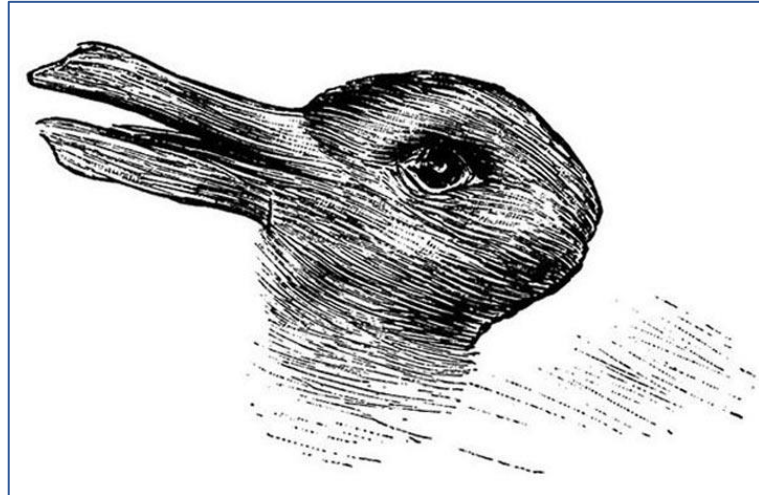
Неделя обучения	Модуль	Темы	Содержание
4-5 недели	Интеллектуальная обработка изображений с использованием технологий машинного обучения	Тема 2.1. Задачи распознавания и классификации объектов на изображениях	Теория и технологии интеллектуальной обработки изображений. Способы и методы решения задач распознавания и классификации объектов на изображения с использованием машинного обучения, получения данных с веб-страниц. Обучение нейронных сетей для интеллектуальной обработки изображений.
		Тема 2.2. Библиотека алгоритмов компьютерного зрения OpenCV	Обзор функциональных модулей библиотеки OpenCV (поиск особых точек, выделение границ, повышение резкости, фильтры Лапласа-Гаусс) для реализации компьютерного зрения. Применение библиотеки OpenCV для решения практических задач.

Тематика практических занятий:

1. Получение данных с веб-страниц с использованием технологий google scrapping
2. Поиск объектов на фотографиях с помощью OpenCV

Практические задание:

- К какому классу изображений относится данное изображение (Топ – N)?



Структура и содержание программы

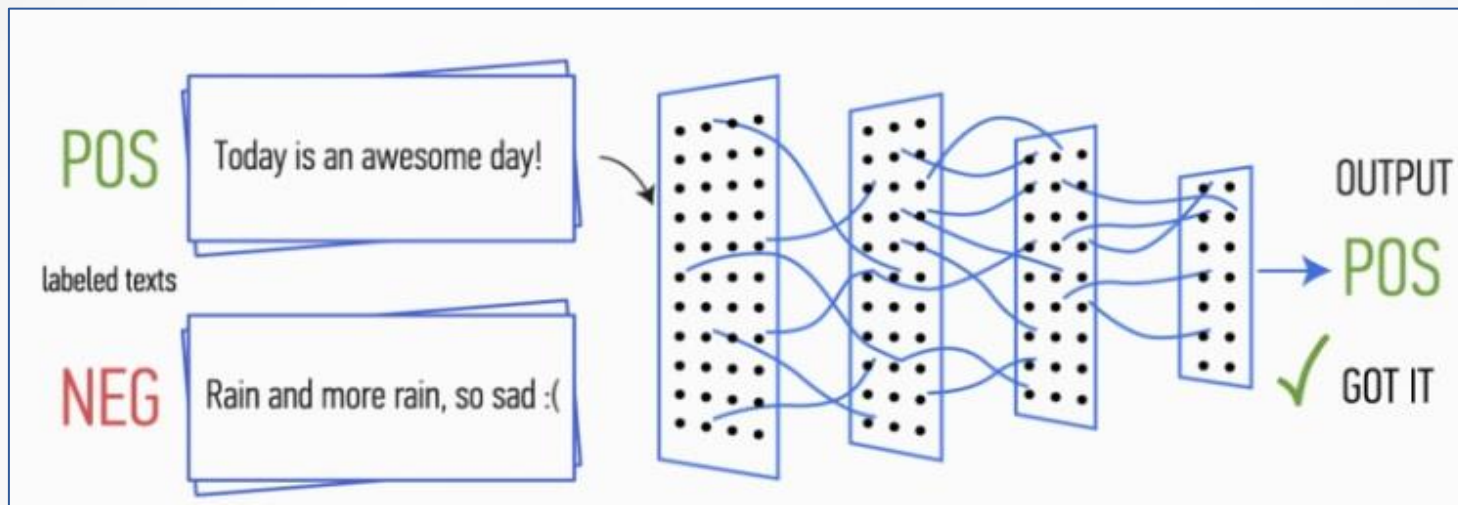
Неделя обучения	Модуль	Темы	Содержание
6-7 недели	Интеллектуальная обработка естественного языка с использованием технологий машинного	Тема 3.1. Задачи обработки естественного языка (NLP)	Задачи обработки естественного языка (NLP): распознавание речи, анализ речи (извлечение информации, информативный поиск, анализ высказываний, анализ тональности, вопрос-ответные системы), генерирование текста, синтез речи. Обучение нейронных сетей для интеллектуальной обработки текста.
		Тема 3.2. Технологии NLP	Препроцессинг и языковые модели. Технологии NLP: Word2Vec, GloVe, RNN, LSTM, GRU, Attention. Применение технологий NLP для решения практических задач.

Тематика практических занятий:

1. Использование технологии Word2Vec/GloVe для определения тональности текста
2. Использование технологий RNN/LSTM/GRU/Attention для решения задачи нейросетевого машинного перевода текста.

Практическое задание:

➤ Позитивный или негативный текст?



Ололош,
Краснодар
22.06.2015

ТВ

Зашел в магазин за батарейками для пульта. Но мое внимание привлекла новая модель LG по такой доступной цене. Тут же оформили покупку и доставку телевизора. Картинка на высоте. Звук просто отличный. Только не было батареек в комплекте. (Но я как раз же за ними и пришел). Всем советую. Приятно удивило количество бонусов на карту. Спасибо, М.Видео!

POS

?

NEG

Структура и содержание программы

Неделя обучения	Модуль	Содержание
8-9 недели	Подготовка к итоговой аттестации	Разработка учебно-методических рекомендаций по организации студенческой научно-исследовательской работы, связанной с применением технологий машинного обучения для анализа данных. Разработка должна включать в себя: 1. тему НИР; 2. обоснование актуальности и практической значимости темы НИР; 3. цель и задачи НИР; 4. рекомендуемые источники информации по теме НИР.

Зачет проходит в форме семинара, на котором преподаватели должны представить свои рекомендации.

Регламент представления выступления на семинаре:

1. Представление слушателем учебно-методической разработки (4-5 минуты).
2. Ответы на вопросы членов комиссии и других слушателей (2-3 минуты).

Содержание курса базируется на:

- Исследованиях компании «Грид Динамикс» в области применения машинного обучения в электронной коммерции.
- Курсе «Специализация Яндекс по машинному обучению» (Яндекс, МФТИ) <https://www.coursera.org/learn/supervised-learning/>.
- Курс лекций К.В. Воронцова «Машинное обучение»
<http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5%D0%28%D0%BA%D1%83%D1%80%D1%81%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE%D0%B2%29>.
- Ресурсах порталов:
 - Colaboratory <https://colab.research.google.com/>
 - Kaggle <https://www.kaggle.com/>
 - GitHub (блог Yorko, посвященный использованию Jupyter Notebook)
https://github.com/Yorko/mlcourse.ai/tree/master/jupyter_russian

Результат:

- Обучение по программе «Интеллектуальная обработка данных с использованием машинного обучения» прошли 15 преподавателей факультета.

Перспективы:

- Встраивание новых курсов «Основы машинного обучения»/ «Машинное обучение и анализ данных»/ «Интеллектуальные системы и технологии»
- Расширение тематики научно-исследовательских работ, в том числе курсовых и выпускных квалификационных работ студентов факультета.
- Открытие на факультете центра машинного обучения совместно с компанией Грид Динамикс.

Спасибо за внимание!

- Сайт СГУ <http://www.sgu.ru/>
- Сайт факультета КНиИТ
<http://www.sgu.ru/structure/computersciences>
- E-mail факультета: kniit@info.sgu.ru
- Личный e-mail: kudrinaev@mail.ru