

## Projekty FRMU řešené na Fakultě informatiky MU v roce 2020:

- MUNI/FR/1286/2019 - A2 - Podpora zavedení magisterského předmětu Graphic Design for Web  
MgA. Jana Malíková – 82 850 Kč
- MUNI/FR/1444/2019 - A3 - Inovace výukových materiálů předmětů programování hradlových polí (FPGA)  
RNDr. Zdeněk Matěj, Ph.D. – 83 050 Kč
- MUNI/FR/1268/2019 - B1 - Inovace seminární výuky předmětu IB031 Úvod do strojového učení  
RNDr. Jaroslav Čechák – 66 100 Kč

**Název projektu FR MU 2020:** Inovace seminární výuky předmětu IB031 Úvod do strojového učení

**Řešitel:** RNDr. Jaroslav Čechák

**Celková výše rozpočtu:** 66 100 Kč

**Dosažené cíle a výstupy:**

- 1) Aktualizované materiály pro praktickou seminární výuku.
- 2) Podklady pro studentské skupinové projekty (témata, zadání, popis, hodnocení).
- 3) Soupis aktivit pro oživení seminární výuky.

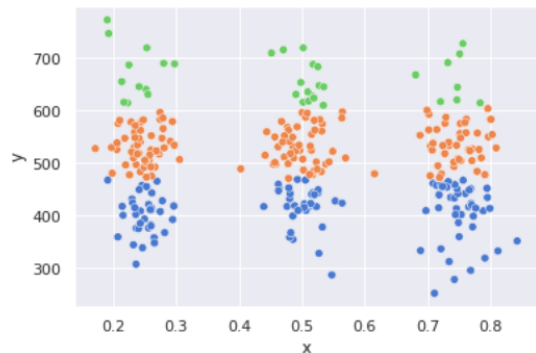
**Konkrétní dopad na výuku:** Studenti nově programují v jazyce Python, který je pro ně známý a mohou se tak soustředit více na strojové učení. V projektech si vyzkouší práci ve skupině a zároveň řešení problému bez předem daného postupu.

# Praktické ukázky výstupů

## Ukázka materiálů pro studenty

```
[3]: plot_clusters(data, clusters)
```

```
[3]: <AxesSubplot:xlabel='x', ylabel='y'>
```



That clustering is really bad. All three clusters are mixed together. Its because we forgot to normalize features. Distances between points in y-axis are much larger than in x-axis.

### Exercise 1

Make a pipeline that will scale both features into range [0, 1] and then run agglomerative clustering. If you did everything right, you should see the same clusters (except colors might be shuffled) as in the first figure.

```
[4]: # pipeline = TODO: your code goes here...  
# better_clusters = pipeline.fit_predict(X)  
# plot_clusters(X, better_clusters)
```

## Hodnotící rubrika k projektům

### 4 Hodnocení

Níže je rubrika shrnující co a jak budu na projektech hodnotit.

popis požadavku	body
projekt obsahuje explorační analýzu datasetu	2
vhodné předzpracování dat podle typu řešené úlohy a vybraných modelů	2
propracované předzpracování dat s využitím pokročilých technik	4
projekt obsahuje krátký popis pro každý vybraný model	2
všechny modely natrénované na datasetu	1
vhodná volba parametrů modelů a jejich ladění	4
vyhodnocení modelů pomocí několika vhodně zvolených mír	4
porovnání modelů s nějakým naivním „baseline“ modelem	2
projekt obsahuje krátké shrnutí výsledků	2
projekt obsahuje vysvětlující komentáře dokumentující jednotlivá rozhodnutí v projektu	2
správná metodologie učení a vyhodnocování modelů	5

# Praktické ukázky výstupů

Úvodní stránka pro seminární výuku

Fakultní Gitlab repozitář se všemi materiály

**IB031 Introduction to Machine Learning – Tutorials – spring 2020**

**Seminar Tutors**

- Jaroslav Čechák (cechakj@it.muni.cz)
- Filip Lusa (lusa@it.muni.cz)
- Amel Calibek (calibek@it.muni.cz)
- Adam Hora (hora@it.muni.cz)
- Mark Kadřák (kadrad@it.muni.cz)

Do not hesitate to email us should you encounter any difficulties. We will be also grateful for any feedback and suggestion on improving the course and tutorials.

**Software Requirements**

We will use Python as a programming language and Jupyter Lab as IDE. You may, however, choose your own IDE. To run Jupyter Lab on faculty computers just execute following command in the shell:

```
$ module add python3-3.7.6
$ jupyter lab
```

If you want to use Jupyter Lab on your personal computer you will need:

- Python of version  $\geq 3.7$
- install package `jupyterlab` (<https://pypi.org/project/jupyterlab>)

We will be using various Python packages throughout the tutorials. The module on faculty computers will have them already pre-installed. You will need to install them on your personal computer though. These packages are:

- `numpy` (<https://pypi.org/project/numpy/>)
- `gensim` (<https://pypi.org/project/gensim/>)
- `sklearn` (<https://pypi.org/project/scikit-learn/>)
- `scipy` (<https://pypi.org/project/scipy/>)
- `matplotlib` (<https://pypi.org/project/matplotlib/>)
- `seaborn` (<https://pypi.org/project/seaborn/>)
- `keras` (<https://pypi.org/project/keras/>)
- `tensorflow` (<https://pypi.org/project/tensorflow/>)

The list may change during the semester.

**Events**

Some of the events happening in Brno that we came across. We are not organizing them, it is not part of the course, and we cannot guarantee any quality.

- There are no upcoming events

**Projects**

The project assignment is available [here](#). It describes all required parts of the project, instructions on presentations and submission, and grading details. You also need to sign up for the [project topic in the IS](#).

**Outline**

All datasets used in the tutorials are either [here](#) or accessible through URLs specified in the notebooks.

**Tutorial 01 – Working with Data**  
(17. 2. and 20. 2.)

- Organizational information
- Introduction to Jupyter Lab and IPython notebooks
- Working with Data
- Working with Data (solutions)

**Documentation**

- Navbar
- Changelog

**Tutorial 02 – Exploration Data Analysis**  
(24. 2. and 27. 2.)

- Exploration Data Analysis
- Exploration Data Analysis (solutions)

**Documentation**

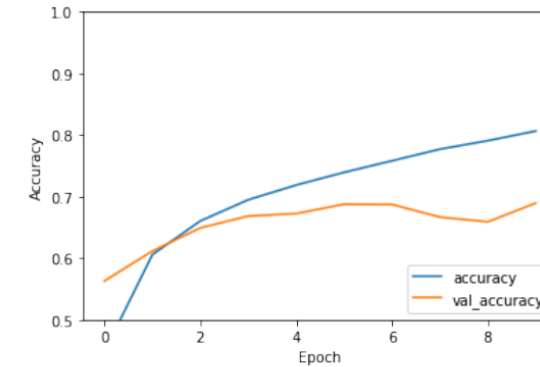
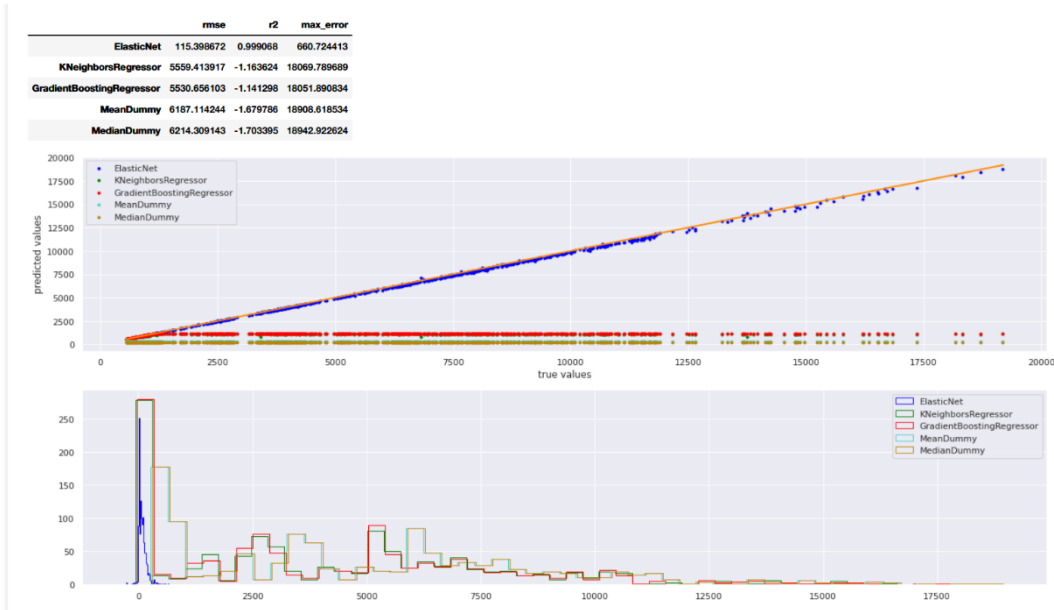
- MapReduce
- SparkCore

The screenshot shows the GitLab interface for the repository 'IB031\_tutorials\_2020'. The main content is a commit history table. The commit [tut04] change word2vec to word embeddings; types by Jaroslav Čechák is highlighted. The table lists 12 tutorial files and two workflow files.

Name	Last commit	Last update
datasets	[BACKUP] add forgotten materials and chna...	1 month ago
projects	[project] update project assignment for new ...	2 weeks ago
tutorial01	renumber tutorials for new semester	5 days ago
tutorial02	renumber tutorials for new semester	5 days ago
tutorial03	renumber tutorials for new semester	5 days ago
tutorial04	[tut04] change word2vec to word embedding...	2 days ago
tutorial05	renumber tutorials for new semester	5 days ago
tutorial06	renumber tutorials for new semester	5 days ago
tutorial07	renumber tutorials for new semester	5 days ago
tutorial08	renumber tutorials for new semester	5 days ago
tutorial09	renumber tutorials for new semester	5 days ago
tutorial10	renumber tutorials for new semester	5 days ago
tutorial11	renumber tutorials for new semester	5 days ago
tutorial12	renumber tutorials for new semester	5 days ago
envrc	[workflow] add envrc for direnv	6 months ago
gitignore	[workflow] add make file for converting mds ...	1 year ago
gitlab-ci.yml	Merge	1 year ago

# Praktické ukázky výstupů

## Ukázky studentských projektů



```
[52]: print(test_acc)
```

0.6895

### 5.6.6 Conclusion

The CNN that was trained on the compound data had the resulting accuracy of 80.65% on the training set and 68.95% on the testing set (after 10 epochs). Let's compare the results with the original images data, with which the CNN produced an accuracy of 90.02% on the training set and 70.96% on the testing set (after 10 epochs). The accuracy on the testing data got slightly worse, but it seems that overfitting on the training data got lowered to a rather large extent.

## 6 The conclusion of the project

In summary, the baseline model was weak as we expected, the reason could be the inability of Decision Tree in recognizing the general features of the objects in an image (e.g. the shape or color, which are related to a certain class, of the object). In determining the content of each image, pattern recognition plays an important role. However, instead of recognizing the general patterns, Decision Tree simply considers each element in the array representation of an image as an attribute and classifies the images in this way. On the contrary, neural network imitates human brain and is designed to recognize patterns.