```
[3]: # Global imports and settings
     %matplotlib inline
     from preamble import *
     plt.rcParams['savefig.dpi'] = 100 # Use 300 for PDF, 100 for slides
     HTML('''<style>html, body{overflow-y: visible !important} .CodeMirror{min-w
```

```
<IPython.core.display.HTML object>
```

Introduction to Machine Learning

2(+1) day intensive course

Resources

- All materials on GitHub: **https://git.io/vdTrB**
- Run through the prerequisites (Python package installations), you'll need them!
- We use Jupyter Notebooks, so you can interact (play around) with the materials
  - PDF exports for backup
- Also check the README for links to reference books

Assumptions

- This is not an introduction to Python. I assume you are broadly comfortable with writing Python code.
- This is not a statistics course. I assume you are familiar with basic statistical concepts.
- This is not a theory course. I am not going to derive anything. I will teach you machine learning tools, when to use them and how to use them correctly and most effectively.
- This is not a deep learning course. However, the machine learning concepts you'll learn here will help you study it much more effectively.

Brief Bio:

- Joaquin Vanschoren (j.vanschoren@tue.nl) MF 7.104a
- I lead the Open Machine Learning project (openml.org), an online platform to research machine learning algorithms
- Most of my work is on:
  - `automating` machine learning: automatically find good machine learning pipelines
  - learning to learn: how can machine learning algorithms learn from one problem to the next
- These topics won't be covered here explictly, though (no time)

Course Overview: Day 1

- 9:00 - 9:15. Welcome
- 9:15 - 10:45. Introduction: Basic concepts, k-Nearest Neighbors
- 10:45 - 11:00. Break
- 11:00 - 12:30. Linear models: Linear regression, Ridge, Lasso, Logistic regression
- 12:30 - 13:30. Lunch break

- 13:30 - 15:00. Evaluation and model selection: Avoiding overfitting. Cross-validation.
- 15:15 - 16:45. OpenML intro + Labs

Final/take-home assignment day 1

- Find a dataset that fascinates you
- Put it on OpenML, tag with 'TUe-ml'
- Build and share any machine learning model on it

This is to let you (and me) come in contact with issues not covered/foreseen in the course, and get feedback.

Course Overview: Day 2

- 9:00 - 9:15. Welcome, Q&A
- 9:15 - 10:45. Ensemble learning: Trees, Bagging, Boosting, Stacking
- 10:45 - 11:00. Break
- 11:00 - 12:30. Kernel methods and Bayesian Learning (intro)
- 12:30 - 13:30. Lunch break
- 13:30 - 15:00. Feature Engineering and Pipelines
- 15:15 - 16:45. Labs

Final/take-home assignment day 2

- Try to build the best possible model on your favourite dataset
- Also try to build good models on the datasets selected by other students
- Try to beat their best solution
- You can work in teams of ? students (or alone)
- Prepare to present your work on day 3

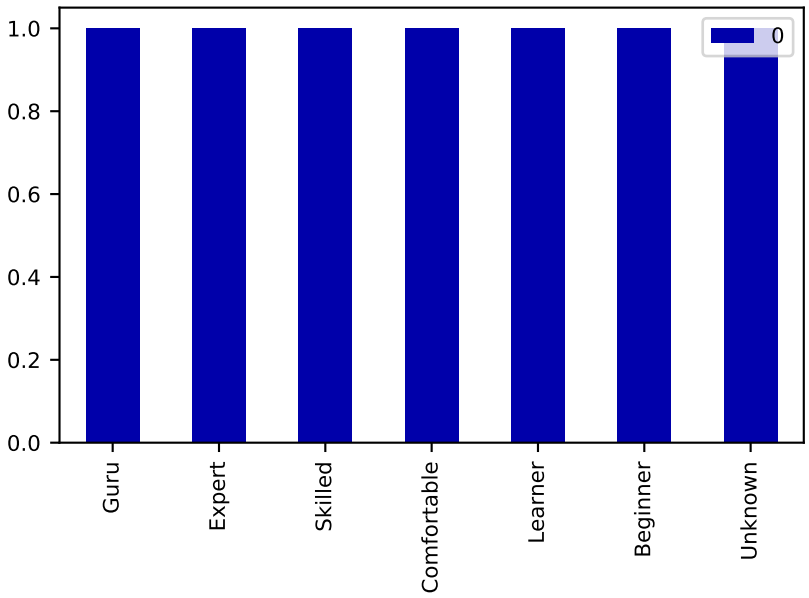
Course Overview: Day 3

- 9:00 - 9:15. Welcome, Q&A
- 9:15 - 10:45. Project presentations
- 10:45 - 11:00. Break
- 11:00 - 12:30. Project presentations

Machine Learning Proficiency Poll

1. Guru: I wrote the book, done it for decades, I present at all mayor conferences.
2. Expert: It has been my full time job for more than a decade.
3. Skilled: I have been doing it for years.
4. Comfortable: I understand it and have done it.
5. Learner: I am still learning.
6. Beginner: I have heard of it and would like to learn more.
7. Unknown: I'm just here to learn more about machines.

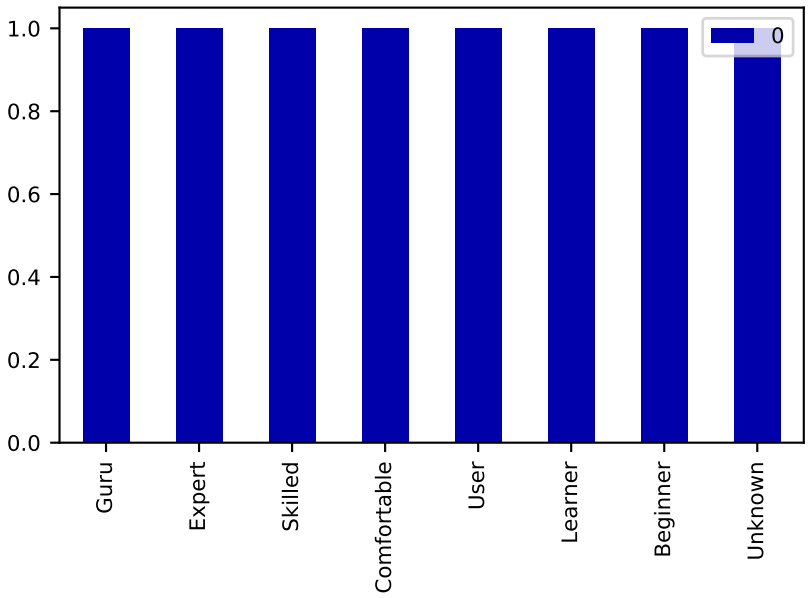
```
[12]: levels = ['Guru','Expert','Skilled','Comfortable','Learner','Beginner','Un
      data = np.array([1,1,1,1,1,1,1])
      pd.DataFrame(data, index=levels).plot(kind='bar');
```



Python Proficiency Poll

1. Guru: The Python core team comes to me for advice.
2. Expert: I have written several Python packages.
3. Skilled: I use it regularly and it is an important part of my job/study.
4. Comfortable: I use it often and am comfortable with the tool.
5. User: I use it sometimes, but I am often searching around for the right function.
6. Learner: I have used it a few times.
7. Beginner: I've managed to download and install it.
8. Unknown: I watched all Monty Python sketches. Is that sufficient?

```
[10]: levels = ['Guru','Expert','Skilled','Comfortable','User','Learner','Beginn
      data = np.array([1,1,1,1,1,1,1,1])
      pd.DataFrame(data, index=levels).plot(kind='bar');
```


1.0
0.8
0.6
0.4
0.2
0.0
0
Guru
Expert
Skilled
Comfortable
User
Learner
Beginner
Unknown