PRACTICAL WORK
of
Artificial Intelligence

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Computer Engineering
January 30, 2021

Deadline of delivery: 13/01/2021

AUTOMATIC DETECTION OF THE OCCUPATION OF AN INTERIOR SPACE

Automatic building occupancy detection can be used successfully in the implementation of energy-saving strategies. As, for reasons of privacy, it is not feasible, in most of the case, the installation of surveillance cameras, it is necessary that the presence of occupants in an interior space is made through a more indirect monitoring. For this purpose, sensors were placed in space to measure humidity, temperature, light and carbon dioxide. For a period of time, the data recorded by those sensors were collected at various times of the day, together with an indication of the presence or not of occupants.
It is intended, with the present work, the development of a machine learning (ML) model, supported by the Random Forest algorithm, which will allow in the future to automatically detect the occupation of the space concerned based on the information given by these sensors.

Tasks to be done

- Implement in a Jupyter notebook the solution to the problem enunciated, developing, with the data made available in the `datasetEn.csv` file, a high-performance ML model, which fully meets the described specifications. [`generate solution.ipynb`]
- With the developed model, classify the observations of the dataset made available in the `dataset_unclassified.csv` file, for which the true classification is not made known. [`generate classification_estimated.csv`]
- Prepare a brief presentation of the work developed. [`generate presentation.pdf`]

Considerations to be taken into account in the implementation

- For the development of the requested ML model, Scikit-Learn and other Python support *packages* should be used.
- In addition to the construction of the model itself, the importance that each of the explanatory variables has in the ability of the model to detect the occupation of space should be evaluated.
- The datasetEn.csv and dataset_unclassified.csv files, containing, respectively, the dataset for induction of the model (with 12,000 observations) and the dataset that will put it to the test (with 5,000 observations), can be downloaded from the curricular unit area of the ipb.virtual platform, in Recursos/avaliacao/trabPratico/En.
- The presentation, after being prepared in an appropriate application, of choice, should be converted to pdf, contain between 3 and 7 slides (not counting the 1st), and the font size of the body of the text is between 16 and 20 pt. Do not forget to include in the 1st slide the name and number of each element that is part of the working group.
- The file to be submitted classification_estimated.csv must contain only the id_record and occupation columns and should therefore have the following:

```
id_record,occupation
17761,1
25264,1
30252,0
...
```

- In assessing the work, account will be taken, in particular, of the correctness, according to the F1 metric, of the forecasts contained in the file `classification_estimated.csv`, generated by the model for data with undisclosed classification.

General considerations

- This practical work should be carried out by groups of 3 students and is mandatory for approval to the curricular unit.
  (Suggestion for cooperation in the tasks to be performed: each member of the group can start by developing their own classification model; then jointly prepare a single model for submission, which takes advantage of the best ideas and options considered in the three proposals.)
- It is expressly forbidden to copy all or part of code from sources other than the documentation provided by the teachers of the curricular unit.
- The work should be delivered only by one of the members of the group, within the deadline established, mandatorily in the e-learning portal (in http://virtual.ipb.pt/, choose <Trabalho Pratico> from the <Atividades> tab, within the AI area), and in no situation can be sent by email.
- The 3 requested files (`solution`.ipynb, `classification_estimated`.csv and presentation.pdf) must be submitted, in separate and uncompressed attachments.
- The work can only be submitted with a maximum delay of 5 days, and a value to your note is subtracted for each day of delay.
- No resubmissions will be allowed (when submitting, make sure it is the final version).
- Students may have to defend their work, either in person or by videoconference.