PRACTICAL WORK
of
Artificial Intelligence

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

Computer Engineering
January 10, 2022

Deadline of delivery: 17/01/2022

DETECTION OF OCCUPATION OF A SPACE USING DIFFERENT MACHINE LEARNING TECHNIQUES

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.

The present work intends to develop a classification model, which will allow, in the future, to automatically detect the occupation of the space in question based on the information given by the referred sensors. In the construction of this model, 3 Machine Learning (ML) techniques should be explored, namely: Logistic Regression, The K Nearest Neighbors and Decision Tree.

Tasks to be done

- Implement in a Jupyter Notebook document the solution to the problem enunciated, developing, with the data made available in the `dataset.csv` file, a high-performance ML model, which fully meets the described specifications. [`generate solucao.ipynb`]
- With the developed model, classify the observations of the dataset made available in the `dataset_naoclassificado.csv` file, for which the true classification is not made known. [`generate classificacao_estimada.csv`]
- Prepare a brief presentation of the work developed. [`generate apresentacao.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.
- The dataset.csv and dataset_naoclassificado.csv files, containing, respectively, the dataset for induction of the model (with 14,000 observations) and the dataset that will put it to the test (with 5,560 observations), can be downloaded from the curriculum unit area of the ipb.virtual platform, in Recursos/avaliacao/trabPratico.
- 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 classification_estimated.csv must contain only the id and `ocupado` columns and should therefore have the following:

```
id,ocupado
14001,1
14002,1
14003,0
...
```

- 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 `classificacao_estimada.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 of cooperation in the tasks to be carried out: each of the elements of the group can start by developing a classification model based on one of the 3 mentioned ML techniques; then, together, they integrate the three

types of models found in the same Jupyter Notebook document and propose the ideal model to solve the stated problem.)

- 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 < Assignments> tab, within the AI area), and in no situation can be sent by email.
- The 3 requested files (`solucao`.ipynb, `classificacao_estimada`.csv and `apresentacao`.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.