Coursework 1 (50% of module mark)

Data Classification + WEKA

Project Aim

The aim of the project is to develop a data classification model based on the selected classification algorithm for the selected issue and to compare the chosen method with other popular and available data classification models. The obtained results should be presented in the form of a synthetic, readable report. It should analyze selected methods, learning process and results, taking into account the specificity of algorithms and data.

Topics

The list of suggested project topics is presented below. The topics suggested below may be subject to any modifications. You can propose your own, ambitious project topic, however, you should bear in mind the timeframe for the completion of your work. By submitting your chosen topic, please contact me by email and provide your plan of work and possible modifications. When sending an e-mail, please enter the project title from the following list or suggest your own title (with description):

1. Letters recognition from black-and-white rectangular pixel displays (http://archive.ics.uci.edu/ml/datasets/Letter+Recognition)
2. Credit card application prediction for identifying customers that are likely to become insolvent (http://archive.ics.uci.edu/ml/datasets/Credit+Approval)
3. Detection of pathologic cardiac conditions from cardiotocograms measurements of fetal heart rate (FHR) and uterine contraction (UC) features (http://archive.ics.uci.edu/ml/datasets/Cardiotocography)
4. Hill-Valley detection on a two-dimensional graph (http://archive.ics.uci.edu/ml/datasets/Hill-Valley)
5. Chronic Kidney Disease detection (http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease)
6. Activity Recognition from Single Chest-Mounted Accelerometer (http://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer)
7. Activity Recognition Using Smartphones Data Set (http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones)
8. Image segmentation based on high-level numeric-valued attributes of the image (http://archive.ics.uci.edu/ml/datasets/Image+Segmentation)
9. Classification of wheat seeds based on their kernel's geometrical properties (http://archive.ics.uci.edu/ml/datasets/seeds)
10. Classification of urban land cover using high resolution aerial imagery intended to assist sustainable urban planning efforts (http://archive.ics.uci.edu/ml/datasets/Urban+Land+Cover)

Proposed Structure of the Report

1. First page with the project title and the name of the author.
2. **Table of Content.**
3. **Project abstract** (max 200 words).
4. **Introduction** – introduction and discussion on the research problem (subject) and justification for the choice of the topic.
5. **Methodology** – discussion on the methods used in the data classification experiment - What were the actions carried out?, What data/information/methods sources were used?, What data and from where it was collected? How the data was prepared/pre-processed?, etc.
6. **Discoveries and Results** - the most important part of the report, presenting the effects of work on the project. It should include information collected from various sources, both written sources and collected as a result of the conducted research. The information should be subject to selection and analysis and presented in such a way as to illustrate the overall work related to the implementation of the project. Any problems and their solutions should be discussed. It is advisable to use various types of drawings, screenshots, diagrams, summaries and tables in order to present the analysed problem in the most complete and transparent form. Please remember about the captions under the drawings and in front of the tables. It is recommended to divide this part of the project into sub-sections to improve the transparency of the report.
7. **Conclusions and recommendations** - contains a summary of the entire report and suggestions on what else could be done to further improve the model performance.
8. **References.**

(The whole report should have no more than 2,000-3,000 words)

Please Note:

Some of the topics are of research nature (for example, "Image recognition"). They may end in failure. If there is a failure of too low recognition rate, but attempts and efforts have been made, the project can be positively evaluated, and the reasons for the failure together with proposals for further attempts to improve effectiveness should be presented in the report. Please, do not be discouraged.

In case of any doubts, please ask the teacher.

If someone intends to write a diploma thesis under my supervision, then the subject of the project can be chosen so that it is similar to the subject of the diploma thesis (of course, if possible). And vice versa - if someone is interested in the project, it can become the basis for the diploma thesis.