Individual Assignment Specification (2023/24)

In this individual assignment (Assessment Task A), you are given a dataset, and you are required to select, train, apply, evaluate, present and compare two different machine learning models using two of the contemporary machine learning algorithms the module covers, i.e., SVM and ensembles, on the dataset.

The learning outcomes: On successful completion of this assessment task students will achieve the following learning outcomes.

> **MO1** Compare and contrast the basic principles and characteristics of a range of contemporary machine Learning algorithms.

The individual assignment is about hands-on experience with classification tasks using a given real-world dataset and the appropriate Python libraries.

Weighting: 30%

The Individual Assignment Specification is **handed out in Week 4**.

Submission due (Week 9): 5pm Friday 29th March 2024

The Dataset: The dataset is given in a .csv file that is ready to use for the assignment. The dataset and its brief description are in the Assessment folder on the Blackboard:

`diabetes.csv`, `diabetes-description.txt`

In the .csv file, the first row contains the column names of the dataset.

The Requirements:

Select, train, apply, evaluate using appropriate performance metrics and present an SVM model using an appropriate kernel. Describe, explain and justify the selection of any appropriate hyper-parameters (i.e., $\boldsymbol{C}$, $\boldsymbol{d}$, $\boldsymbol{r}$, $\boldsymbol{\gamma}$) for the model.

Select, train, apply, evaluate using appropriate performance metrics and present an ensemble using an appropriate ensemble method. Describe, explain and justify the selection of any appropriate hyper-parameters (i.e., number of base classifiers, the size of the random subsets of data, the size of the random subsets of features, the level of the decision trees, learning rate, etc.) for the model.

Describe, explain and justify your approach to building each of the two models, including the key machine learning concepts involved, use of any library/function that is not covered in any practical notebook, the methods you have used and the design decisions you have made in each step.

Compare the performance of the two models using appropriate performance metrics, and contrast the advantages and disadvantages of the two models/machine learning algorithms.

Along with the code cells, you should use the appropriate markdown cells in the notebook to structure your notebook, and to describe, explain, justify and discuss your work.

Expected amount of time: 40 hours.

Feedback: Summative feedback will be provided by the marking grid and comments in the completed individual assignment assessment form as specified in the marking criteria. The assessment form is expected to be released within 2-3 weeks of the submission deadline (excluding 2 weeks Easter vocation).

Submission link: The submission link for the assignment will be available in the Assessment folder on the module site on the Blackboard.

Your submission consists of a Jupyter Notebook and a converted PDF file of the Jupyter Notebook. Please make sure that the submitted notebook runs successfully and ***all the cell outputs are cleared****. Please also make sure that* ***the submitted PDF file has the cell outputs visible****. Please name your files as follows: studentNumber.ipynb and studentNumber.pdf.*

Marking Criteria

	0-40%	**50-70%**	**50-70%**	**70-100%**
Analysis and treatment of dataset (10%)	Basic understanding, description and limited analysis and treatment of dataset	Good understanding, description and sufficient analysis and treatment of dataset	Clear understanding, description and good analysis and treatment of dataset	Excellent understanding and description, and excellent analysis and treatment of dataset
Model Selection and Training (40%)	Selection of models; Use of training, selection of some hyper-parameters	Sufficiently explained and justified selection of models; Sufficient use and description of training, satisfactory selection and description of hyper-parameters	Clearly explained and justified selection of models; Good use and description of training, good selection and description of hyper-parameters	Well explained and justified selection of models; Full use and excellent description of training, appropriate selection of hyper-parameters
Prediction and Evaluation (30%)	Generation and analysis of evaluation results	Satisfactory generation, analysis, explanation and description of evaluation results using explained and justified methods and performance metrics, appraisal of evaluation results	Good generation, analysis, explanation and description of evaluation results using well explained and justified methods and performance metrics, appraisal of evaluation results	Excellent generation and clear analysis and explanation of evaluation results using fully explained and justified methods and performance metrics, appropriate appraisal of evaluation results
Comparison of Two Models (10%)	Comparison of the models on performance; Discussion of advantages and disadvantages	Comparison of the models on performance using sufficiently explained and justified methods and performance metrics; Sufficient discussion of advantages and disadvantages	Good comparison of the models on performance using well explained and justified methods and performance metrics; Good discussion of advantages and disadvantages	Excellent comparison of the models on performance using fully explained and justified methods and performance metrics; Clear discussion of advantages and disadvantages
Presentation (10%)	Use of markdown, basic	Sufficient use of markdown and sections,	Good use of markdown and	Excellent use of markdown and sections, very

	descriptions, explanations and justifications	sufficient descriptions, explanations and justifications	sections, sufficient	clear descriptions, explanations and justifications with high level of clarity and completeness