THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Strategies for pre-match trading on sports betting exchanges**
Author's name:	**Radek Bula**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Michal Sustr
Reviewer's department:	Department of Computer Science

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — challenging

How demanding was the assigned project?

The student's project was to explore a field of pre-match price development on sports exchanges and design betting strategies with a focus on providing a long-term profit. The assignment is quite challenging, as the study of horse race sport exchanges involves handling real-world data: The data had to be collected, cleaned and studied to find relevant subsets where there is a trading opportunity. Then the student had to apply machine learning algorithms and evaluate them properly.

Fulfilment of assignment — fulfilled

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The student met all of the requirements. He gave a good overview of the betting world and explained the basic terminology. He selected a specific sport for which he collected and cleaned a large amount of data (20GB) on which he could do significant analysis. He designed multiple betting strategies using machine learning algorithms, such as decision trees, gradient boosting, ridge regression or neural networks. He searched for model hyper-parameters to maximize the suitable metrics relevant for sports betting (return on investment or strategy efficiency, among other evaluations).

Methodology — partially applicab

Comment on the correctness of the approach and/or the solution methods.

The student used two targets for prediction of the ML models: price after time interval (as the ratios of open/close back/lay prices, see Eq. 5.3., 5.4.) or price inside the time interval (as comparing to the best back/lay price, see Eq. 5.5, 5.6). These targets were used for classification and regression tasks. He analyzed the data to select subsets that have the highest impact for the modeling and ran hyper-parameter tuning.

The methodology is mostly correct, however consider the following quote from the thesis: "For strategy evaluation, we used 10 000 **randomly** selected races from the whole dataset." (emphasis mine)

As the goal of the thesis is to design a betting strategy that would maximize profit, randomly selecting races from the whole dataset is a very subtle but grave error. Evaluation should be done **only** on future races (future from the training data), and great care must be taken to not leak any future information (i.e. test data) by mistake. All of the reported evaluations are therefore likely bogus, they should be treated only as an upper bound on the profits at best, and their true performance in the real world would likely be inferior.

A minor evaluation that I was missing was variance of profit over time: due to variance of the bet outcomes it is possible to run out of budget even though on average the strategy is profitable.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

Technical level	A - excellent.
Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?	
The thesis is technically sound and the student has applied expertise in his field of study, he clearly explained the motivations and analyzed data to support his decisions in making the models.	

Formal and language level, scope of thesis	B - very good
Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?	
The thesis is formally correct, the text is clear and understandable, the contents are well organized. There are minor English mistakes, such as incorrect use of (in)determinate articles, or capitalization of letters – Figure, Table, etc.	

Selection of sources, citation correctness	A - excellent.
Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?	
The thesis makes adequate research on earlier work in the topic, the original work is clearly distinguished and bibliographic citations meet the standards.	

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

The thesis is a practical work on designing betting strategies. While the thesis took non-trivial effort to accomplish, it does not provide a more general insights applicable outside of the chosen domain. It can be treated as a solid engineering / analysis work of the student.

A major drawback in the methodology is the *random* split of the whole dataset into training and evaluation subsets. While it is an understandable practice due to typical i.i.d. assumptions of the used models, it is nonetheless wrong for this domain. As it should be relatively easy to re-run the experiments with proper time-splits, I advise to reanalyze the experiments and present the updated results at the thesis defense. Betting strategies with/without time-splits can be compared, possibly over multiple time-window subsets to see how the evaluation also changes in time, with larger historical record of the training data.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The grade that I award for the thesis is B - very good

Date: 18/01/2021

Signature: Michal Sustr