

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Vilém Krejcar
Advisor:	prof. PhDr. Ladislav Křišťoufek, Ph.D.
Title of the thesis:	Analysis of consumer behaviour among DotA 2 players and its effect on individual performance

OVERALL ASSESSMENT (provided in English, Czech, or Slovak):

Short summary

Author analyses data from a free online multiplayer game called Defence of the Ancients 2, commonly known as Dota 2. It is difficult to clearly summarize what it is that the author tries to achieve in this thesis since it is written in a very confusing way and there are different goals in introduction, conclusion and main body of text. Author writes that he tries to predict the rank of individual Dota 2 players with data regarding the players' consumer behaviour, however this does not match the variables included in the actual regression analysis. Most of the included independent variables do not represent the consumer behaviour of a player but rather some descriptive characteristics within the particular game. These variables are also clearly influenced by the actions of another team as well as teammates, clearly creating an endogeneity problem which is not discussed in the thesis at all. Author also does not actually do predictions, only simple in-sample fit, so the objective should perhaps rather be explanation than prediction.

Contribution

I fail to see a clear contribution of this thesis besides the improvement of the author's technical skills by collecting and cleaning the underlying dataset. Author himself also does not state his contributions clearly and mentions different objectives across the text. As far as I am concerned, this topic also does not fall under Economics in general and it is not clear why readers should care about the results. In spite of that, I certainly appreciate that the author decided to disclose his computational scripts as well as underlying datasets.

Methods

Author uses order logit and k-means clustering. Logistic regression is explained clearly. However, some claims about a distribution of the dependent variable *rank* are false. He claims that dependent variable *rank* has a distribution similar to the normal distribution, however does not perform any tests to verify this hypothesis. If he did, he would find out, as is clear from Figure 3.1, that the distribution is not normal as it is defined only on positive real numbers and is truncated at 35 (high rank).

K-means clustering is not properly explained, spanning only one paragraph on page 30. Half of the paragraph also consists of this sentence: "*After the initial segmentation, centroids are re-calculated to be situated in the actual centre of the cluster and re-calculates the points belonging to the new clusters.*", which makes little sense and is only a very vague description of the algorithm.

It is also not clear to me why the author chose unsupervised k-means clustering and not some supervised classification since he is assessing the algorithm performance by the accuracy with which this algorithm classifies the observations into the rank partitions (original dependent variable).

Literature

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Vilém Krejcar
Advisor:	prof. PhDr. Ladislav Křišťoufek, Ph.D.
Title of the thesis:	Analysis of consumer behaviour among DotA 2 players and its effect on individual performance

Author offers a detailed and extensive list of literature covering a wide area of topics related to DotA 2. These sources are properly referenced. However, there is no literature review covering the decision making in games or any other area more closely related to the field of economics.

On p. 11, author mentions use of machine learning in this area and points out their higher accuracy compared to classical linear or probabilistic models, however it is not mentioned what is actually being predicted and thus to what does mentioned accuracy refer to.

Even though a lot of space is dedicated to the description of the game, it is quite confusing for the non-initiated, e.g. what are "levels" on p.6 in sentence "Both HP and mana are being regenerated every second, increasingly with levels."

Manuscript form

There are grammatical errors littered all over the thesis. Sentences are also often poorly structured, making it difficult to understand the intended meaning. Writing style is clearly non-academic and often very vague. Organization of sections is also confusing. Dataset description is spread out over section 3 Data mining and collection as well as section 4 Dataset properties and description. In case of results, some of them are in subsection 5.2 Estimation and other results are in section 6 Results. Figures are also not self-explanatory as only short (one-sentence) descriptions are provided.

There are other problems with manuscript form as well such as mixing tenses, poor graphics quality or inconsistent usage of bold font to mark vectors (formula 5.7).

Overall evaluation and suggested questions for the discussion during the defense

This thesis looks more like exploratory data analysis of a non-traditional dataset rather than a complete research project with clear results. It is hard to read and presentation of results is confusing. There is almost no interpretation of main results in table 5.1. Table 5.2 contains results from the cluster analysis, however it is not clear to me how they were obtained or how to interpret them.

It is also not clear to me how results in tables 6.1 to 6.4 were calculated, how they are connected to the logistic regression model or other results in section 5. These tables also contain no standard errors or any indicator of confidence around these estimates.

Further, sentences like "*The estimation coefficients confirmed the hypotheses we set up in Section 3.2.*" on page 31 make me question the authors understanding of statistical hypothesis testing.

Questions for discussion during the defense:

- Explain the main goals and contributions of the thesis with the emphasis of relation to the field of economics.
- Discuss possible problem of endogeneity in your regression analysis.
- Interpret results in table 5.1. Why are all p-values in table 5.1 basically equal to zero? (Some explanation is provided in the text, however I have trouble understanding what the author means).
- Interpret results of tables 6.1 to 6.6 and describe how they were obtained.

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Vilém Krejcar
Advisor:	prof. PhDr. Ladislav Křišťoufek, Ph.D.
Title of the thesis:	Analysis of consumer behaviour among DotA 2 players and its effect on individual performance

- Explain clearly multicollinearity in general, how it affects your estimation of the logit model and how it was addressed.
- Explain the algorithm used in k-means clustering.
- Why is unsupervised approach (k-means clustering) used instead of supervised classification when it seems that the actual goal is the classification to low/mid/high rank groups?

The results of the Urkund analysis do / do not indicate significant text similarity with other available sources.

In my view, the thesis fulfills the requirements for a bachelor thesis at IES, Faculty of Social Sciences, Charles University. I recommend it for the defense, however only in case of successfully answering questions for discussion during the defense I suggest a grade E.

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	10
<i>Methods</i> (max. 30 points)	16
<i>Literature</i> (max. 20 points)	16
<i>Manuscript Form</i> (max. 20 points)	10
TOTAL POINTS (max. 100 points)	52
GRADE (A – B – C – D – E – F)	E

NAME OF THE REFEREE: Martin Hronec

DATE OF EVALUATION: 24.8.2021

Digitally signed, 24.8.2021, Martin Hronec

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F