

# Probabilistic Model for Difficulty Assessment of Climbing Routes

**Mert Türedioglu**

Middle East Technical University  
mertturedioglu@gmail.com

## Abstract

The climbing route difficulty assessment is a challenging problem. In this research, the difficulty of the climbing route will be analyzed by probabilistic programming methods. Climbing routes will be chosen from the routes of the MoonBoard climbing board. MoonBoard is a popular climbing board with global prevalence. The use of boards in climbing training is becoming more common day by day. As climbing has become popular and has become an Olympic sport, new training boards have emerged. Hundreds of new routes are uploaded to the applications of these training boards by people from the climbing community every day. With the increase in the number of these boards and the number of routes, the need to automate the route difficulty assessment has arisen. In recent years, research in this area has begun to emerge. In this research, collection of MoonBoard routes and analysis of route difficulties will be implemented in Python. In this analysis, the Bayesian data analysis library PYMC will be used.

## Keywords

Climbing, MoonBoard, Training Board, Bayesian Data Analysis, Difficulty Grade.

Codebase can be found here: <https://github.com/mertturedioglu/moonboard-project-probabilistic>

## 1 Introduction

At the very beginning, indoor climbing emerged as a training for outdoor climbing. However, at the point reached today, indoor climbing has ceased to be a training for outdoor climbing and has become a sports discipline on its own. It recently gained Olympic sport status with the Tokyo Olympics in 2020.

The routes selected in the research will be chosen from the MoonBoard routes, which is a climbing training board. Board training is the most efficient type of training in climbing, with its simple structure and oriented towards the training needs of

a climber. The grid structure of the training board provides simplicity. The ledges of the training board held by the climbers are called the hold. Holds at specified distances and angles are placed on training boards around the world in the same way. This simple structure of the training board makes the difficult problem of routing difficulty assessment workable.

The fact that these training boards have a worldwide standard and have the same structure everywhere makes the output of this research generalizable. This research is intended to provide a formal aspect to the climbing route difficulty assessment. It is expected that the method in this research, which is made specifically for MoonBoard routes, can be applied to other climbing boards with some modifications.

### 1.1 Research Question

The following questions are planned to be answered in this research.

- What are the features that determine the difficulty of a climbing holds?
- What are the factors that determine the climbing route's difficulty level?
- What is the relationship between climbing holds and climbing route difficulty?

### 1.2 Methods and Results

A probabilistic approach was preferred in this study. A probabilistic approach to the complex relationship of factors determining the grade of the climb route was considered appropriate. Each of the factors that determine the difficulty level of the climb, when considered on their own, is not capable of determining the difficulty level of the route. The combination of these factors determines the route difficulty level. In this study, route and hold features that can represent these factors were determined. The effects of these factors on the route difficulty level were investigated.

The data of the routes to be used in the research were scraped from the MoonBoard website. A web bot is coded with the source code of Python's Selenium library. With the help of this bot, raw data of 451 climbing routes approved by MoonBoard's expert team were obtained.

In addition to the raw data of the routes scraped from the MoonBoard website, 140 climbing holds in the MoonBoard 2016 layout were examined and classified.

The raw data of these collected climbing routes and the raw data of the climbing holds were turned into features with the knowledge of an experienced climber.

A Directed Acyclic Graph (DAG) was designed to show the causal relationship between the generated features and the route difficulty level. Based on this graph, 2 statistical models were designed. With the help of these statistical models, the relationship between features and route difficulty levels was examined and the prediction capabilities of these two models were compared.

## 2 Related Work

The climbing route difficulty assessment is just beginning to appear in the literature. This task was studied with machine learning algorithms. Some of the research is only for the evaluation of climbing route difficulty, the other part is for creating new climbing routes. Due to its simple nature, the majority of research is directed towards training board routes. As in this research, it mainly works on MoonBoard routes.

Tai, Wu and Hinojosa(2020) approached the route difficulty assessment using neural networks in their research. They were inspired by the NLP (Natural Language Processing) domain in which Graph Convolutional Network architectures have proven success in text classification. They have preprocessed their problems into multi-hot encoded representations. To briefly explain the character of this representation, the route representations consist only of which of the 140 holds of the MoonBoard is on the route. They trained their models with these multi-hot encoded route representations and did not do any feature engineering. This representation does not contain any other information about the characteristics of the routes. They compared the accuracy scores of 17 different machine learning models. The most successful model in this comparison was the Graph Convolutional Model, inspired by the NLP domain. They compared the accuracy scores of 17 different machine learning models. In this comparison, the most successful model was the Graph Convolutional Model, inspired by the NLP domain, with an accuracy score of ~73%. In this study, unlike the research in this report, the number of routes used in the model is more than seventeen thousand. In addition to the routes whose difficulty levels and quality were approved by the MoonBoard team (Benchmark Routes), other MoonBoard routes added by the climbers were also used in their research. The 451 Benchmark routes are an insufficient sample size for developing machine learning models. To increase this number, they used the mentioned community routes while developing and testing their models.

Dobles, Sarmiento, and Satterwhite used machine learning techniques in their research for route difficulty estimation. Their preferred technique is the Convolutional Neural Network technique, which is widely used in pattern recognition.

This technique is generally used to detect patterns in images and to recognize the object in the image. The pattern recognition capability of this technique is promising in recognizing similar patterns in route difficulties. They represented the MoonBoard routes as multi-hot encoded as in the above research. As mentioned above, this representation only contains information about which holds are in the route and which are not. In their research, they experimented with three different classifiers: Naive Bayes classifier, Softmax Regression Classifier and Convolutional Neural Network classifier. They also used the Support Vector Machine but did not report the result due to its poor performance in the validation set. They compared the accuracy scores of 3 different models with people's predictions. They could not achieve an accuracy score above 36.5% from all 3 models. People's accuracy score was reported as 93.4%. They see one of the reasons behind this as people guessing routes not just from pictures but from their climbs.

In both studies, it can be thought that one of the most important factors reducing the accuracy score is that the difficulty levels of the route in the sample are not evenly distributed. There are more routes on easy difficulty levels, and fewer routes on hard difficulty levels. These imbalances have a detrimental effect on models' accuracy performance scores.

## 3 Dataset

The routes on which difficulty estimations will be made in the research were chosen among the MoonBoard routes. MoonBoard has multiple training boards. The routes used in this research were selected from the MoonBoard 2016 board. This MoonBoard is the first training board. Currently, there are more than fifty thousand routes made for this board. Climbing routes are set by climbers and can be shared with other climbers via the MoonBoard app. From the standardized structure of the training board, any climber can access this route information and climb these routes.

Access to MoonBoard routes is possible in two ways. Information on climbing routes can be accessed via MoonBoard's phone application or website. The routes to be used in this research were scrapped from the MoonBoard website. By means of a bot implemented in Python, the hold information and difficulty level of the climbing routes were obtained. The coding of this bot was done with Python's Selenium package.

The routes dataset has the 451 benchmark tagged climbing routes. Routes with benchmark tags are the ones approved by the MoonBoard team in terms of route difficulty and route quality. Due to the challenging and time-consuming aspect of route quality and difficulty assessment, only 451 of the fifty thousand routes mentioned have benchmark tags at the moment. This quality and difficulty assessment needs to be automated. It is expected that this research will support this possible automation. It is allowed to create routes in approximately 13 different difficulty levels on the MoonBoard training board. This range starts from 6B+ difficulty and goes up to 8B+. There are 13 difficulty levels, with 6B+ being the

easiest and 8B+ being the hardest. However, there are very few routes with benchmark tags at 8A+ and above difficulty levels. To be specific, there are currently 2 routes in 8A+ difficulty, 1 in 8B difficulty, and no routes in 8B+. The grade distribution of the routes is unbalanced. The distribution of routes with benchmark tags is as follows.

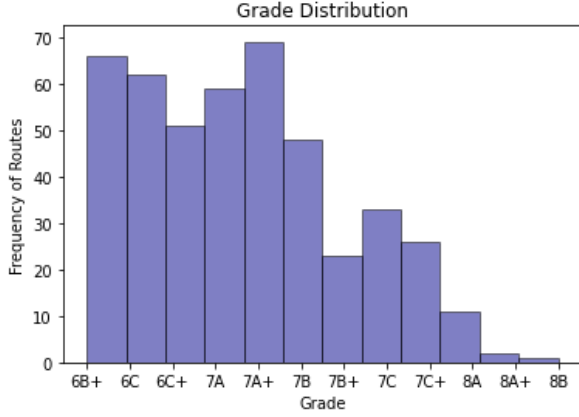


Figure 1: Grade Distribution

### 3.1 Feature Set

The raw data of these collected climbing routes and the raw data of the climbing holds were turned into features with the knowledge of an experienced climber. These feature sets are open to development.

### 3.2 Hold Features

Holds have been physically studied and classified. These classification headings are as follows: rotational angle set (direction of pull, in which direction one can grab the hold and apply force), z score (like incut size, the capacity to apply force outwards direction) and color. Multiple hold features are derived from this rotational angle information and transferred to the route features. The hold color may seem irrelevant, but the 3 different colors in the MoonBoard 2016 layout reflect some characteristics. Yellow holds are very small in size, black holds are relatively large and sloper, and whites are of various character. Hold's z score is used as a factor in generating route features to give an idea of possible climbing move difficulty. Multiple hold features are derived from the rotational angle feature obtained by physical examination. These features are categorical features of which direction the climber can apply force to the hold: only sideways, only upwards, downwards. On overhanging training boards such as the MoonBoard, the direction in which force can be applied to the hold provides information in terms of route difficulty estimation.

### 3.3 Route Features

Some of the route features are derived from the hold features, some of them are derived from the physical location of the

holds in the route, and some of them are derived from the combination of the two. MoonBoard has a grid structure. The distance between the holds is fixed. The route information scraped from the MoonBoard site contains information about which holds are included in the route. The 'number of holds' and 'mean distance of holds' features are derived from this information.

There are multiple route features related to the colors of the holds available in the route. 'Black count', 'white count', 'yellow count', 'black percent', 'white percent', 'yellow percent', 'b\_w percent'.

There are also multiple route features associated with the rotational angle of the holds in the route: 'downwards\_count', 'only\_upwards\_count', 'only\_sideways\_count', 'downwards\_percent', 'only\_upwards\_percent'.

Finally, two route features are derived from the difficulty of possible moves on the route: first\_move\_difficulty and last\_move\_difficulty. Some of the hold features were used to determine the difficulties of these moves. These hold features are the hold color, z\_score, and rotational angle. In determining these move difficulties, the information about the number of holds in the route was also used. The factors that determine these difficulties are associated with the average of the distances of the 3 holds included in these moves, how many of the 3 holds are yellow, their rotational angles and z-scores. The equation of this 'difficulty of first move' feature and the explanations of the variables are as follows:

Abbreviation	Meaning
FD	first move difficulty score
md	the mean distance between three holds
fy	is first hold yellow (0 or 1)
yc	yellow count of three holds (0,1 or 2)
tu	is third hold's rotational angle upwards (0 or 1)
zc	total z_score of three holds
ly	is the last hold yellow (0 or 1)
n	the number of holds in the route

Table 1: Abbreviations

$$FD = md^{(1+fy+yc+tu)/\ln(2.72+zc)}$$

The equation of this 'difficulty of last move' feature as follows:

$$LD = md^{(1+ly+yc)/\ln(n+zc)}$$

### 3.4 Data Overview

ADD DATA OVERVIEW FIGURES

## 4 Model

### 4.1 Causal Model - DAG

A Directed Acyclic Graph (DAG) was designed to show the causal relationship between the generated features and the route difficulty grade. Abbreviations and DAG as follows:

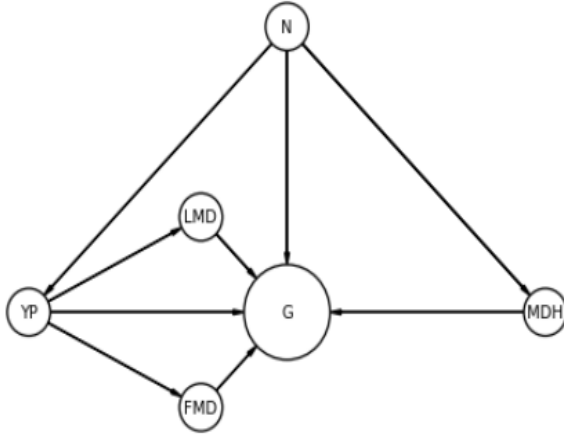


Figure 2: DAG

Abbreviations	Meaning
G	Difficulty Grade
N	Number of the holds
YP	Yellow percent
LMD	Last move difficulty
FMD	First Move difficulty
MD	Mean distance between holds

Table 2: DAG Abbreviations

### 4.2 Statistical Models

Based on DAG in Figure 2, two different statistical models were designed. The prior distributions of both models can be seen below. Refer to Table 2 for explanations of abbreviations.

$$\begin{aligned}
 G_i &\sim \text{OrderedLogit}(\phi_i, \alpha) \\
 \phi_i &= \beta_{fmd} FMD_i + \beta_{lmd} LMD_i + \beta_{YP} YP_i + \beta_N N_i \\
 \beta_{fmd}, \beta_{lmd}, \beta_{YP}, \beta_N &\sim \text{Normal}(0, 0.5) \\
 \alpha_j &\sim \text{Normal}(0, 1)
 \end{aligned}$$

Figure 3: Move Difficulty Model

$$\begin{aligned}
 G_i &\sim \text{OrderedLogit}(\phi_i, \alpha) \\
 \phi_i &= \beta_N N_i + \beta_{MDH} MDH_i + \beta_{YP} YP_i \\
 \beta_N, \beta_{MDH}, \beta_{YP} &\sim \text{Normal}(0, 1) \\
 \alpha_j &\sim \text{Normal}(0, 1)
 \end{aligned}$$

Figure 4: Mean Distance - Number of holds - Yellow Percent Model:

Ordered logit is preferred for estimated variable. Climbing difficulty levels are in order. There is no uniform difference between the climbing difficulty levels. For example, the increase in route difficulty levels between 6C+ and 7A is not the same as the increase in difficulty levels between 7C+ and 8A. For these two reasons, Ordered Logit was preferred for the difficulty level, G, variable for both statistical models.

As can be seen from the difficulty level distribution histogram in Figure 1 data is unbalanced in terms of difficulty grades. There are more easy routes than difficult routes. 2 different prior distributions for cut points (alphas) were tested. Assuming that it would be more appropriate with unbalanced data, tests were also performed with StudentT priors. However, StudentT's robust structure prevented the model from making grade predictions at 8B difficulty level. There is only 1 route in Dataset at 8B difficulty level. It was not possible to update the robust model effectively with a single instance.

## 5 Results and Discussion

### 5.1 Results

#### 5.1.1 Move Difficulty Model

The summary table for Move difficulty model can be seen below.

	mean	sd	hdi_3%	hdi_97%
alpha[0]	-2.219	0.139	-2.475	-1.957
alpha[1]	-1.246	0.108	-1.452	-1.048
alpha[2]	-0.634	0.098	-0.812	-0.443
alpha[3]	-0.007	0.096	-0.184	0.177
alpha[4]	0.742	0.101	0.545	0.929
alpha[5]	1.402	0.115	1.188	1.622
alpha[6]	1.859	0.128	1.612	2.092
alpha[7]	2.632	0.157	2.344	2.936
alpha[8]	3.706	0.218	3.297	4.111
alpha[9]	4.930	0.312	4.368	5.536
alpha[10]	6.494	0.476	5.665	7.451
bFMD	0.046	0.086	-0.113	0.212
bLMD	0.114	0.098	-0.073	0.294
bYP	0.720	0.103	0.529	0.914
bN	-0.366	0.085	-0.531	-0.209

Table 3: Summary Table - Move Difficulty Model

As it can be seen from the Table 3, ‘YP’, ‘FMD’ and ‘LMD’ have positive effects on ‘G’, whereas ‘N’ has negative effect on ‘G’. Yellow Percent of the holds (YP) has the largest effect on ‘G’. In addition, It can be seen from the table as the number of holds (N) increase in a route, the difficulty grade of the route decrease. Last move difficulty (LMD) has larger positive effect compared to first move difficulty (FMD).

### 5.1.2 Mean Distance – Number of Holds – Yellow Percent Model

The summary table for Mean Distance, Number of Holds, Yellow Percent model can be seen below.

	mean	sd	hdi_3%	hdi_97%
alpha[0]	-2.502	0.145	-2.765	-2.221
alpha[1]	-1.417	0.114	-1.636	-1.209
alpha[2]	-0.717	0.103	-0.919	-0.533
alpha[3]	0.002	0.101	-0.185	0.192
alpha[4]	0.855	0.108	0.655	1.060
alpha[5]	1.593	0.124	1.364	1.826
alpha[6]	2.102	0.137	1.843	2.356
alpha[7]	2.945	0.166	2.647	3.264
alpha[8]	4.040	0.220	3.642	4.464
alpha[9]	5.247	0.307	4.695	5.842
alpha[10]	6.779	0.461	5.962	7.712
bMDH	0.974	0.092	0.805	1.145
bYP	0.833	0.087	0.665	0.994
bN	-0.259	0.085	-0.420	-0.100

Table 4: Summary Table – Mean Distance, Yellow Percent, Number of holds Model

As it can be seen from the Table 4, ‘MDH’, ‘YP’ and ‘N’ have positive effects on ‘G’, whereas ‘N’ has negative effect on ‘G’. Mean distance between the holds (MDH) has the largest effect on ‘G’. In addition, It can be seen from the table as the number of holds (N) increase in a route, the difficulty grade of the route decrease.

### 5.1.3 Comparison of Predictive Performances of Models

The comparison table for PSIS scores of two models is below.

	rank	loo	p_loo	d_loo	weight	se	dse
Mean Distance-Number of holds Model-Yellow Percent	0	1853.441401	24.356650	0.000000	0.895985	52.310010	0.00000
Move Difficulty Model	1	1975.491508	26.118506	122.050107	0.104015	54.376863	25.76565

Table 5: Comparison table of Predictive Performances of Models

Mean Distance, Yellow Percent Number of Holds Model has better PSIS scores compared to the other model. From this table, it can be seen that MD, YP, N is better variable set than the set consists of FMD, LMD, YP, N variables.

## 5.2 Discussion

In two models, the results to test cases in posterior predictive simulations were examined. Although it cannot be included in this report due to the limited space here, the test cases and the graphics of the results of test cases can be accessed via the link given at the beginning of the report. Both models made promising predictions for the test cases.

To improve the performance of the models, new features can be produced, and existing features can be developed in order to improve the prediction scores of the models and to make more precise predictions.

These Causal DAG and related statistical models were developed for the MoonBoard 2016 board. The features used in these models were produced for MoonBoard 2016. However, similar models can be developed for other boards: MoonBoard 2017, MoonBoard 2019, Kilter Board, Tension Board etc.

It is thought that analyzes like the one made here can help systems that will automate route grade estimation. Detection of features that determine the route difficulty level can increase the accuracy of this automation.

## References

- Tai, C. H., Wu, A., & Hinojosa, R. (2020). Graph neural networks in classifying rock climbing difficulties. Student project report, CS, 230.
- Dobles, A., Sarmiento, J. C., & Satterthwaite, P. (2017). Machine learning methods for climbing route classification. Web link: <http://cs229.stanford.edu/proj2017/finalreports/5232206.pdf>.