Prediction of Success Construction Claims by Construction Provider Due to Delays in Completion Work

Novera Meylinda[1], Ayomi Dita Rarasati[2]
novera.meylinda@ui.ac.id[1], ayomi@ui.ac.id[2]

Civil Engineering Department, University of Indonesia, Depok, Indonesia[1,2]

Abstract. To compete in the construction business, it is also necessary to be careful in seeing the opportunity for claims due to delays in completion work, especially for construction providers who often fail in filing claims. Most of the previous research on claims was more about predicting the appearance of claims. In fact, the claim that arises does not stop until it is accepted or not. So this research is conducted so that construction providers know how to make claims successful. We conducted a literature review and Decision tree method C4.5 to determine the order of success factors for claims from the most influential, and the research showed administration of claims is the most influential factor in the success of a claim so claim can be accepted and with a model accuracy of 81,29.

Keywords: Claim, Prediction, Success factors, Decission tree, C4.5 algorithm.

1 Introduction

Delay in completion of work is one of the causes of claims and disputes. Delay in the completion of work is an almost common occurrence in a construction project, both private and government projects. For example, the delay that occurred in the Government's National Strategic Project (PSN). The Committee for the Acceleration of Priority Infrastructure (KPPIP) delivered 12 (twelve) PSN projects that were supposed to be completed by the end of 2019 but were postponed to 2020 [1]. The impact of delays is conflict and debate about what and who is the cause, also creates time demands, and added costs [2].

Claims for extention of time and additional cost or losses and expenses that submitted by construction providers often not optimal because of the position of construction users who are considered superior and construction providers are worried about getting bad reviews from clients and being seen as "claimants".

By knowing how claims can be accepted by construction users, judging from the success factors for claims, it will make the submission of claims made by construction providers more optimal. then a research is conducted on the prediction of the success of construction claims by construction providers due to delays in the completion of work.

The first research method used is a literature review to obtain the success factors for claims for work delays due to service users. To make a prediction model for claim success, the classification technique Decision Tree C4.5 is used, because it can convert data into decision rules. This decision tree can detail complex decision-making processes into simpler ones so that

ISTED 2021, July 17-18, Purwokerto, Indonesia

DOI 10.4108/eai.17-7-2021.2312156

problem solutions can be more easily obtained and interpreted. Decision trees are used to solve a problem where each node is a decision and leaves are the solution to the problem [3]. A tree is a data structure consisting of nodes and edges [4]. The algorithm to build the decision tree is C4.5 which used gain ratio as splitting criteria. It begins from the root node which contains the entire dataset and than split by an atrtribute the forming nodes and edges which connecting the nodes. This process maintains recursively with the rules of splitting (goodness of cut up criterion) till the criteria are met. A node that isn't splitting anymore is referred to as a leaf node and classified with the majority class.

2 Research Methods

2.1. Literature review

Literature review is used to obtain the success factors for claims for work delays due to service users. Furthermore, validation is carried out with a validation questionnaire. The questionnaire contains responses from experts in the form of yes or no that the success factors for existing claims are in accordance with the success factors for each claim submission made.

2.2. Decision tree C4.5

Methods Data analysis with the decision tree algorithm C4.5 in this study is depicted in **Fig. 1**.

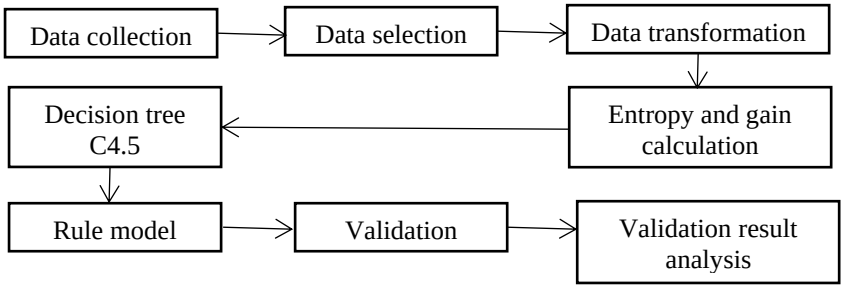

Fig 2. Research Procces Design

To find out the gain ratio, first calculate the entropy, entropy is used to determine how informative the attribute is. S is the Set of Cases, n is the number of partitions S. and pi is the number of cases on partition i. It is given by equation (1).

$$\text{Entopy (S)} = \sum_{i=0}^{n} - pi * \log_2 pi \quad (1)$$

After knowing the entropy, then calculate the information gain. S is the Set of Cases, n is the number of partitions atribute A, |Si| is the number of cases on partition i, and |S| is the number of cases in S. It is given by equation (2).

$$Gain\ (S,A) = Entropy\ (S) \ - \sum_{i-1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \quad (2)$$

To determine the root node with the C4.5 algorithm, the largest value of the gain ratio will be used, because the gain ratio can provide a more specific value than the gain or

information gain. However, it is necessary to first calculate the split information with the equation (3).

$$SplitInfo\ (S,A) = \sum_{i=1}^{c} -\frac{S_i}{S}\log_2\frac{S_i}{S} \quad (3)$$

Gain Ratio formula is found in equation (4), as follows.

$$Gain\ ratio = \frac{Gain(S,A)}{SplitInfomation(S,A)} \quad (4)$$

The gain in equation is the result of calculating the Information gain obtained after calculating the Entropy. Then select the gain ratio that produces the best practitioner or the greatest value to then be used as the root node in the C4.5 decision tree.

2.3. Model validation with k-cross validation

Model validation with k-fold Cross Validation is a technique to ensure that the results found in the analysis can be generalized to an independent and invisible data set. This study uses 2 fold cross validation because of the small amount of the data. In two-fold cross-validation the data is partitioned into independent and similar subsets and using random assignment into the training data set and the test data set. The model is then constructed using data from the k - 1 subset, using the data subset as the test set. This is done iteratively until we have k different models, see **Figure 2**. The results from the k models are then combined using the average or selecting.

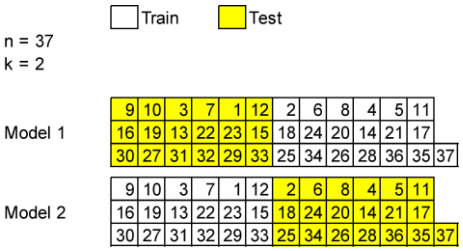

Fig 2. Splitting research data by 2 fold cross validation

In cross-validation k = 2, the data set is randomly generated and then divided into two sets (d0 and d1), so that both sets have the same size. Then training will be carried out on d0 and validation on d1, followed by training on d1 and validating on d0.

Performance measure is very important to indicate how well the model classifies information. In general, the confusion matrix table as in Table one is usedwhen verity values are notable to judge the model performance on a group of take a look at data [4]. Model performance is measured by conniving the worth of accuracy.

Table 1. Confusion matrix

		Predicted	
		Negative	**Positive**
Actual	**Positive**	False Negative (FN)	True Positive (TP)
	Negative	True Negative (TN)	False Positive (FP)

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

Accuracy is obtained by calculated verity foreseen observation divided by total observations. exactness additionally called positive predicted values is obtained by calculated the true predicted positive observations divided by total predicted positive observations.

3 Result and Discussion

Success factors for construction claims by construction providers due to delays in completing work

The questionnaire contains the success factors of claims based on previous research references which are then verified whether these claim success factors can be used as a reference to determine the prediction of the success of construction claims by service providers due to delays in completing work. The results of the recapitulation of phase 1 data collection can be seen in **Table 2**.

Table 2. Recapitulation of validation of construction claim success factors

Do you think the following factors affect the success of claims by service providers due to delays in completing work?					
No.	Claim success factor	P1	P2	P3	Status
1.	Does the submission of the claim state the claim requested (its rights and compensation)	Y	Y	Y	√
2.	What is cause of the claim	Y	Y	Y	√
3.	Complete or not the details of the claim	Y	Y	Y	√
4.	Complete or not the evidence for each claim	Y	Y	Y	√
5.	Is the claims analysis appropriate on a contractual/legal basis/based on evidence	Y	Y	Y	√
6.	Is there any prior notification of the submitted claim	Y	Y	Y	√
7.	Complete or not Project documentation	Y	Y	Y	√
8.	Complete or not Project administration	Y	Y	Y	√
9.	Persuasive presentation	Y	Y	Y	√
10	How does the submission document explain the cause and effect of the claim	Y	Y	Y	√
11.	Relationship between Construction Providers and Construction Users	Y	Y	Y	√
12.	Significance and criticality of claims	Y	Y	Y	√

Based on the results of data collection on the factors that affect the success of claims, all of these factors can be used to determine the prediction of the success of construction claims by construction providers due to delays in completing work. Of the 12 (twelve) factors that influence the success of the above claims, there are two factors that are difficult to measure, namely the relationship between construction providers and construction users and the significance and criticality of claims. Based on discussions with experts, the relationship between construction providers and construction users will be measured by the number of projects that have been worked on together, see **Tabel 3**.

Table 3. Construction provider and construction user relationship category table

The number of projects working together	Relationship between construction provider and construction user
> 2 projects	good
1-2 projects	enough

Meanwhile, the significance and criticality of claims are measured by the value claimed compared to the value of the construction contract, and also the amount of time claimed compared to the implementation period based on the contract. However, the significance and criticality of claims cannot be assessed only in one claim submission, because if there are further claims then the number can be large and significant, so that the construction provider always considers the claimed value to be significant and the time claimed is considered critical because it has exceeded the planned completion time

Decision tree model predicts the success of claims by construction providers due to delays in completing work

Each claim submission is identified the cause of the delay in completing its work to find out if the delay is caused by the construction user. Claims submitted include extention time and additional costs that follow which are submitted separately or submitted all at once in one submission. Each claim submission is carried out by collecting data according to the factors that affect the success of claims obtained in the previous data processing. The 37 data collected can be seen in **Table 4.**

Furthermore, calculations are carried out using the entropy, gain, split Information and gain ratio formulas for each attribute as follows :

Number of data = 37
Accepted = 20
Not Accepted = 17
Then the value ofi *entropy (S)* = ((-20/37)*Log2(20/37)+(-17/37)*Log2(17/37))
= (-0,541*-0,888)+(-0,459*-1,122)
= 0,47974339 + 0,51550916
= 0,995252549
and the calculation is continued with microsoft excel, see **Table 5.**

Table 4. Data on submission of claims by contractor providers along with factors that affect the success of claims

Code of claim	Submission Document Submission states the claim requested (its rights and compensation)	Claim Details	Evidence of each claim	Claim for each cause of claim	Element of claim cost	Analysis of claims based on contract/law/with evidence	The terms/conditions of the claim have been met
a	**b**	**c**	**d**	**e**	**f**	**g**	**h**
G1	Y	Y	complete	Time & cost	prelim	Y	Not listed
G2	Y	Y	complete	time	N	Y	Y
G3	Y	Y	complete	cost	N	N	Not listed
G4	Y	Y	complete	time	N	Y	Y
G5	Y	Y	complete	Cost	Over-head	N	N
...	...	...	...	...	...	...	...
G37							

Table 4. (cont.)

Code of claim	Claim notif	Project doc.	Project adm.	Persuasive presentatioa	Explaining cause and effect of claim	Contractor relationship with Owner	Significant claim	Claim accepted
a	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**
G1	Y	complete	complete	Y	Y	enough	Y	Y
G2	Y	complete	complete	N	Y	enough	Y	Y
G3	N	complete	complete	N	**N**	enough	Y	N
G4	Y	complete	complete	Y	Y	enough	Y	Y
G5	Y	complete	complete	Y	Y	enough	Y	N
...	...	...	...	...	...	...	...	...
G37								

Table 5. Calculation of the value of entropy, information gain, split information, and gain ratio for each attribute to determine the root node.

		Number of data	Accepted	Not Accepted	*Entropy*	*Gain*	*split information*	*gain ratio*
Total		37	20	17	0,99525			
Submission Document Submission states the claim requested (its rights and compensation						0,0000	0	0
	Ya	37	20	17	0,99525			
	Tidak	0	0	0	0,00000			
Claim Details						0,0057	0,75320	0,00767
	Ada	29	15	14	0,99914			
	Tidak ada	8	5	3	0,95443			
Evidence of each claim						0,0632	0,30337	0,20859
	lengkap	35	20	15	0,98523			
	tidak lengkap	2	0	2	0,00000			
Claim for each cause of claim						0,1868	1,24876	0,14964
	Biaya	18	6	12	0,91830			
	waktu	17	13	4	0,78713			
	waktu dan biaya	2	1	1	1,00000			
Element of claim cos						0,2405	1,10503	0,21770
	Fluktuasi dolar	2	1	1	1,00000			
	overhead	14	3	11	0,74960			
	preliminaries	4	2	2	1,00000			
	tidak ada	17	14	3	0,67229			
Analysis of claims based on contract/law /with evidence						0,0900	0,87796	0,10252
	Ya	26	17	9	0,93059			
	Tidak	11	3	8	0,84535			

Table 5. (cont.)

		data	Acc	Not Acc	Entropy	Gain	split infor-mation	gain ratio
The terms/conditions of the claim have been met						0,6605	1,50015	0,44033
	Ya	17	16	1	0,32276			
	Tidak	13	0	13	0,00000			
	Tidak tercantum	7	4	3	0,98523			
Claim notif						0,0642	0,99947	0,06424
	Ada	19	13	6	0,89974			
	Tidak ada	18	7	11	0,96408			
Project doc.						0,0970	0,40598	0,23914
	Lengkap	34	20	14	0,97742			
	tidak lengkap	3	0	3	0,00000			
Project adm						0,4476	0,87796	0,50982
	Lengkap	26	20	6	0,77935			
	tidak lengkap	11	0	11	0,00000			
Persuasive presentation						0,0009	0,63946	0,00144
	Ya	31	17	14	0,99323			
	tidak	6	3	3	1,00000			
Explaining cause and effect of claim						0,1300	0,84185	0,15446
	Ya	27	18	9	0,91830			
	tidak	10	2	8	0,72193			
Contractor relationship with Owner						0,0086	0,80039	0,01075
	Baik	9	4	5	0,99108			
	Cukup	28	16	12	0,98523			
Significant claim						0,0000	0	0
	Ya	37	20	17	0,99525			
	tidak	0	0	0	0,00000			

From the results of the calculation of the data table, it is known that the attribute with the highest gain ratio is "claims administration" which is 0.50982. Thus the attribute “Administration claims” can be a root node. There are 2 values for the “Claim administration” attribute, namely “Complete” and “incomplete”. The attribute value "incomplete" has classified the claim submission data as rejected so that no further calculations are needed, but for the attribute value "Claim administration" "yes" further calculations need to be carried out.

The decision tree formation process stops because all data records in the last node have got the same class, there are no record attributes that can be partitioned again and no records in the branch are empty. So that the final decision tree based on manual calculations can be described as on **Figure 3.**

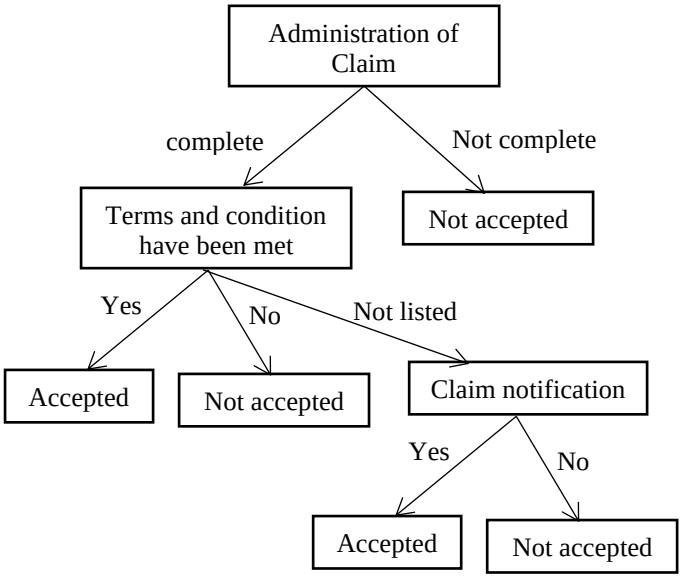

Fig 3. Decision tree model predicts the success of claims by construction providers due to delays in completing work

Model validation

Tabel 6. *Confusion Matrix* decision tree with *2-fold cross validation*

	true Diterima	true Ditolak
pred. Diterima	20	7
pred. Ditolak	0	10

From Table 6 can be explained that the number of data is 27 where 20 claim submissions are predicted correctly (true positive / TP) that the claim can be accepted, while 7 claims submissions are predicted incorrectly (false positive / FP) by the decision tree classifier C4.5, where in fact the submission of the claim was rejected. Tests on the submission of claims that are predicted to be rejected show that 10 claim submissions (true negative / TN) are correctly recognized that the claim submission is rejected, and no claim submissions are received according to the prediction (false negative / FN). So that the values of accuracy, precision, and recall are 81.29%, 100%, and 59.72%, respectively.

4 Conclution

Rules for accepted claims: 1) If the claim administration is complete and the claim terms/conditions have been met, the claim is accepted. 2) If the claim administration is complete,

the terms/conditions of the claim have been fulfilled not stated in the contract/law, there is a claim notification then the claim is accepted
Rules for rejected claims: 1) If the claim administration is incomplete then the claim is rejected. 2) If the claim administration is complete, the terms/conditions of the claim do not meet the requirements in the contract/law, then the claim is rejected. 3) If the claim administration is complete, the terms/conditions of the claim have been fulfilled not stated in the contract/law, there is no claim notification then the claim is rejected. Decision Tree C4.5 method can be implemented to predict the success of claim submission by service providers to service users with an accuracy rate of 81.29%.

References

[1] (n.d.). Retrieved from https://www.cnnindonesia.com/ekonomi/20191228165920-92-460687/penyelesaian-12-proyek-strategis-nasional-molor-ke-2020

[2] David, M. W. P. : Evaluation- From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation. Australia: School of Informatics and Engineering Flinders University of South Australia. 2007

[3] C. J. Mantas, J. Abellán, Francisco Javier García Castellano. Analysis of Credal-C4.5 for classification in noisy domains. Mathematics, Computer Science Expert Syst. Appl. 2016

[4] Proboyo, B. Keterlambatan Waktu Pelaksanaan Proyek : Klasifikasi Dan Peringkat Dari Penyebab-Penyebabnya, Dimensi Teknik Sipil. 1999. Vol. 1 No. 2

[5] Ghozali, M. Ferdi. Penggunaan Decision Tree Analysis dalam Perencanaan Bisnis. Institut Teknologi Bandung. 2017.

[6] Hermawati, F. A. Data Mining. Yogyakarta: Penerbit Andi. 2013.