Faktor Exacta

Vol. 19, No. 3, Sept 2026, pp. 130~140

eISSN: 2502-339X, pISSN: 1979-276X, https://doi.org/10.30998/faktorexacta.v19i3.5590

Optimization of Fiber Optic Network Feasibility Based on Link Budget and Total Loss Using the C4.5 Decision Tree Algorithm in a Web-Based System at PT. Qonita Teknologi Indonesia

Anang Dwi Prastyo[1], Anggreita Tiara Putri[2]
[1,2]Teknik Informatika, Fakultas Ilmu Komputer Universitas Pamulang

Article Info

Article history:

Received 10/09/2026
Revised 13/10/2026
Accepted 14/10/2026

Keywords:

Decision Tree C4.5
Fiber Optic
Link Budget
Total Loss
Transceiver

ABSTRACT

Optical fiber networks offer wide channel capacity with low attenuation, yet service quality remains frequently disrupted when Link Budget and Total Loss calculations are performed manually. This condition was observed at PT Qonita Teknologi Indonesia, where misdetermined network parameters often led to unsuitable transceiver selection. This study developed a web-based application using the Next.js framework and a MySQL database that combined Link Budget and Total Loss computations with the Decision Tree C4.5 algorithm to classify network status into "Network Feasible" or "Network Not Feasible." The C4.5 algorithm was selected to produce explicit, auditable if–then rules. The model was trained on thirty field records comprising structural and performance attributes. Verification involved comprehensive black-box, white-box (Basis Path method), and specification-based implementation testing, alongside leave-one-out and repeated stratified 5-fold cross-validation. All technical testing scenarios were validated, and the system classifications perfectly matched manual calculations in implementation tests, yielding 100% accuracy (95% confidence interval: 83.9–100%). Cross-validation estimated predictive accuracy at 90.0% for the discretized tree and 96.7% for continuous power margin splits, performing comparably to Naïve Bayes, K-Nearest Neighbor, and Random Forest benchmarks (86.7–93.3%). The results demonstrated that the system assesses optical fiber network feasibility quickly, transparently, and comprehensibly, thereby reducing manual calculation errors, although the learned rules remain specific to the operating conditions of a single company.

This work is licensed under a CC-BY-NC

Corresponding Author:

Anang Dwi Prastyo
Teknik Informatika, Fakultas Ilmu Komputer
Universitas Pamulang,
Jl. Surya Kencana No.1, Pamulang, Tanggerang Selatan.
Email: anangprastyo66@gmail.com

1. INTRODUCTION

The rapid growth of digital services has pushed transmission media toward levels of capacity, speed, and reliability that copper-based networks can no longer sustain. Optical fiber answers that demand through very wide bandwidth, low signal attenuation, and immunity to electromagnetic interference, which is why it

now forms the backbone of household internet access, industrial systems, and *Internet of Things* services [1]. Its performance in the field, however, depends on how carefully the optical power along the transmission path is planned before a link is put into service [2].

Two parameters govern that planning. The *Link Budget* expresses how much optical power is available between the transmitter and the receiver, while the *Total Loss* accumulates the attenuation contributed by fiber length, splices, and connectors along the route [3], [4]. When either value is estimated inaccurately, the received power may fall below the sensitivity threshold of the receiver, producing signal degradation, unstable links, and in the worst case a route that is not feasible at all [5], [6].

Internal documentation at PT. Qonita Teknologi Indonesia shows how such planning errors surface during operation. Measurements taken on one of the company's 100 Gbps backbone interfaces are summarized in Table 1.

Table 1. Technical losses recorded on a backbone interface

Parameter	Recorded value	Operational impact
Last link flapped	00:05:42, 10/11/2025	Link instability (flapping)
Interface resets	11	Repeated downtime
CRC error (RX)	148 errors	Corrupted incoming frames
Input errors	148	Small fault affecting the backbone
Interface speed	100 Gbps	High-capacity backbone route
Packet loss (ping test)	8% (3,303 tx / 3,019 rx)	Unreliable data delivery
RTT max	20 ms	Delay rises when loss occurs

Source: internal documentation of PT. Qonita Teknologi Indonesia

The 148 CRC errors recorded on the receiving side indicate that incoming frames were damaged by physical disturbance on the fiber medium, while eleven interface resets and a recorded link flap confirm that the link was not stable. A ping test on the same route returned 8% packet loss out of 3,303 transmitted packets, which shows that the fault was not confined to the physical layer but had already propagated to the network layer. On a high-capacity backbone, packet loss of that magnitude is a strong indication that the optical link is operating outside its feasible range.

Discussions with the engineering team traced these symptoms back to the way feasibility is assessed. *Link Budget* and *Total Loss* are still computed manually with spreadsheets, so a single mistyped parameter changes the conclusion; the decision to declare a route feasible rests on the individual judgement of each engineer rather than on a standardized rule, which makes the outcome inconsistent when several projects are evaluated at once; and the resulting technical figures are difficult to explain to customers who have no background in optical transmission.

Earlier studies on optical power planning have addressed the measurement side of this problem. Analyses of FTTH deployments [3] and of a campus optical network [5] confirmed that power link budget and rise time budget are reliable indicators of feasibility, and comparable results were reported for FTTH design [4] and for an FTTB GPON installation whose simulated and measured values both remained within standard tolerance [6]. Splice quality has likewise been shown to affect attenuation directly, with improved fusion splicing raising the received power measured at the customer terminal [2]. These studies stop at the computation itself, however, and leave the feasibility decision to manual interpretation.

In parallel, classification algorithms have been used to make technical decisions more measurable. The *Decision Tree C4.5* algorithm is frequently preferred because it handles numeric and categorical attributes at once and produces *if–then* rules that a human reader can inspect [7], [8]. It has been applied to the evaluation of security personnel [9], to network attack detection with 90% accuracy [11], to the prediction of fault resolution time in internet services [12], to internet package sales [13], to student study period prediction [14], and to nutritional status classification [15]. The work closest to the present study is the SCADA link feasibility recommendation system of Cahyono and Pratama [10], which reached 84% accuracy but operated on 20 kV electrical telemetry links rather than on optical power parameters.

The choice of *C4.5* over other widely used classifiers in this study rests on four requirements of the feasibility problem. First, the attributes combine continuous quantities such as length, loss, and margin with a nominal one, the transceiver type, and they are strongly dependent on each other, because *Total Loss* is computed from length, connectors, and splices, and the power margin is computed from *Total Loss*; *C4.5* handles mixed attribute types directly, whereas *Naïve Bayes* rests on an assumption of conditional independence that these attributes clearly violate [7], [8]. Second, the decision must be auditable, since an engineer has to see why a route was rejected before redesigning it; *C4.5* yields explicit *if–then* rules together with the *entropy* and *gain ratio* behind them, while *K-Nearest Neighbor* builds no explicit model and *Random Forest* aggregates many trees into a vote that cannot be read as a single rule [8], [10]. Third, the classifier runs

inside a web request, and a compact tree is evaluated in a handful of comparisons without keeping the training data, whereas *K-Nearest Neighbor* must compare every new route with every stored record. Fourth, the data available from a single company are limited, a setting in which *C4.5* has been reported to reach 84–100% accuracy in comparable classification tasks [9]–[11]. These arguments are not taken for granted: Section 3.6 compares *C4.5* empirically with *Naïve Bayes*, *K-Nearest Neighbor*, and *Random Forest* on the same data.

What remains largely unexplored is the combination of these two lines of work, namely optical power computation on one side and rule-based classification on the other, delivered as an operational tool rather than as an offline analysis. This research addresses that gap by integrating *Total Loss* and *Link Budget* computation with the *Decision Tree C4.5* algorithm inside a web application built on *Next.js* and *MySQL* [16], [17]. The contribution is threefold. The calculation is automated, so the risk of manual entry error is reduced. The feasibility decision is expressed as an explicit set of rules extracted from the company's own field data, so it becomes reproducible across engineers. Finally, the system exposes the *entropy* and *gain ratio* values behind each decision together with a transceiver recommendation, so that the reasoning is visible to engineers and customers alike.

2. METHOD

The research was carried out in five consecutive stages: collecting network data at PT. Qonita Teknologi Indonesia, preparing and discretizing that data, constructing the decision tree with *C4.5*, implementing the resulting rules in a web application, and testing the application functionally, structurally, and in terms of classification accuracy.

2.1. Data Collection and Data Source

Primary data were gathered through three channels. Field observation covered the planning and installation of optical routes, where route length was measured with an *Optical Time Domain Reflectometer*, connector and splice points were counted for each segment, and optical power was recorded with an *optical power meter*. Interviews with company engineers clarified the calculation procedure and the criteria applied when judging a route. Internal technical documentation supplied the transceiver datasheets, including transmit power, receiver sensitivity, operating wavelength, and supported reach, and also provided the feasibility label that engineers had assigned to each completed route.

Thirty records were obtained, covering routes from 1.0 to 22.0 km, with between two and thirteen connectors and between one and eighteen splices, served by four transceiver families (SFP, SFP+, QSFP+, and QSFP28), so that both small building installations and long backbone routes were represented. The class distribution was eighteen feasible records (60%) and twelve not feasible records (40%). The attenuation constants followed the cable specification used by the company: 0.35 dB/km for singlemode fiber, 0.5 dB per connector, and 0.1 dB per splice. The complete data set after cleaning is listed in Table 2 so that every computation reported in Section 3 can be reproduced.

Table 2. Field data set after cleaning (30 records)

No.	Length (km)	Conn.	Splices	Total Loss (dB)	Power Margin (dB)	Transceiver	Label	Subset
1	2.5	4	3	5.20	4.80	SFP 10 km	F	Train
2	8.0	6	7	12.40	−1.40	SFP 10 km	NF	Train
3	4.0	4	4	7.20	3.80	SFP+ 10 km	F	Train
4	15.0	8	12	21.80	−3.80	SFP+ 20 km	NF	Train
5	6.0	6	5	9.60	2.40	SFP 10 km	F	Train
6	20.0	10	15	28.50	−5.50	QSFP+ 20 km	NF	Train
7	3.0	4	3	5.80	6.20	SFP+ 10 km	F	Train
8	12.0	8	10	17.60	−0.60	QSFP+ 20 km	NF	Train
9	1.5	2	2	3.10	7.90	SFP 10 km	F	Train
10	10.0	6	8	14.80	1.20	SFP+ 20 km	F	Train
11	5.0	4	5	8.50	3.50	SFP 10 km	F	Train
12	18.0	10	14	25.60	−4.60	QSFP28 20 km	NF	Train
13	3.5	4	3	6.30	5.70	SFP 10 km	F	Train
14	11.0	8	9	16.20	−0.20	SFP+ 10 km	NF	Train
15	7.0	6	6	10.80	1.20	SFP+ 20 km	F	Train
16	1.0	2	1	2.20	8.80	SFP 10 km	F	Train
17	9.0	6	8	13.60	0.40	SFP+ 10 km	F	Train
18	22.0	10	18	31.40	−6.40	QSFP28 40 km	NF	Train
19	4.5	4	4	7.80	4.20	SFP 10 km	F	Train
20	13.0	11	11	19.20	−2.20	QSFP+ 20 km	NF	Train
21	2.0	2	2	3.80	7.20	SFP 10 km	F	Train
22	16.0	13	13	23.40	−4.40	QSFP+ 10 km	NF	Train
23	6.5	6	5	10.10	2.90	SFP+ 10 km	F	Train
24	8.5	6	7	13.00	0.00	SFP 10 km	NF	Train

25	5.5	4	5	9.10	2.90	SFP+ 10 km	F	Test
26	14.0	8	12	20.60	−2.60	QSFP+ 20 km	NF	Test
27	3.2	4	3	6.00	5.00	SFP 10 km	F	Test
28	7.5	6	6	11.40	1.60	SFP+ 10 km	F	Test
29	17.0	10	14	24.80	−3.80	QSFP28 40 km	NF	Test
30	4.8	4	4	8.20	2.80	SFP 10 km	F	Test

F = Network Feasible, NF = Network Not Feasible; Conn. = number of connectors; records 1–24 form the training set and records 25–30 the holdout test set

Thirty records is a small sample by machine learning standards, so its adequacy was examined on three grounds. The first is coverage. Every category of every discretized attribute in Table 3 is populated, the routes span the installations the company actually handles, from 1.0 km in-building links to 22.0 km backbone routes, and the power margins range from −6.40 to 8.80 dB, so that both sides of the feasibility boundary are observed, together with nine records in the critical region between 0 and 3 dB. The second is the nature of the target. Feasibility is governed by a physical relationship, because a route can be operated only while its received power stays above the receiver sensitivity; the class boundary is therefore low-dimensional and can be located from comparatively few labelled examples, unlike targets that depend on many interacting behavioural factors. The third is the evaluation design. A single 24/6 holdout split leaves too few test records for a stable estimate, so it was complemented with leave-one-out and repeated stratified cross-validation, which use every record for testing and are the recommended practice when the sample is small [18], [19]. The records nevertheless come from one company, one cable specification, and a limited set of transceiver models, so they are representative of that operating environment rather than of optical networks in general; this limitation is revisited in Section 4.

2.2. Link Budget and Total Loss Computation

The optical power along a route was computed with (1) to (4):

$$Ltotal = (Lf \times d) + (Ns \times Ls) + (Nc \times Lc) \quad (1)$$

$$Link\ Budget = Ptx - Rx_sens \quad (2)$$

$$Prx = Ptx - Ltotal + M \quad (3)$$

$$Power\ Margin = Link\ Budget - Ltotal \quad (4)$$

where *Ltotal* is the total loss, *Lf* is the fiber attenuation per kilometre, *d* is the route length, *Ns* and *Ls* are the number of splices and the loss per splice, *Nc* and *Lc* are the number of connectors and the loss per connector, *Ptx* is the transmit power, *Rx_sens* is the minimum receiver sensitivity, and *M* is the safety margin applied to the received power. A positive *Power Margin* means the received power still exceeds the sensitivity threshold, so the route can be operated reliably.

2.3. Decision Tree C4.5

The classifier was built with the *C4.5* algorithm, which selects the splitting attribute at each node from the *entropy*, *information gain*, *split information*, and *gain ratio* values defined in (5) to (8):

$$Entropy(S) = -\sum pi \times log_2(pi) \quad (5)$$

$$Gain(S, A) = Entropy(S) - \sum (|Sv| / |S|) \times Entropy(Sv) \quad (6)$$

$$SplitInfo(A) = -\sum (|Sv| / |S|) \times log_2(|Sv| / |S|) \quad (7)$$

$$GainRatio(A) = Gain(S, A) / SplitInfo(A) \quad (8)$$

where *S* is the data set, *pi* is the proportion of records belonging to class *i*, *A* is the candidate attribute, and *Sv* is the subset of *S* whose value for *A* is *v*. The *gain ratio* rather than the raw *information gain* is used for the final selection, because dividing by the split information prevents the algorithm from favouring attributes that simply have many distinct values [8]. Continuous attributes can either be discretized beforehand, as done in this study, or split directly on the binary threshold that maximizes the criterion, which is the native treatment of numeric attributes in *C4.5* [20].

2.4. Data Preparation and Discretization

Four cleaning steps were applied to the thirty records. A missing value check found two records without a transmit power value, because the transceiver type had not been recorded completely during field work; both were completed from the project documentation and confirmed with the responsible engineer. A duplicate check found no identical entries. A consistency check compared each recorded *Total Loss* against a recomputation with (1) and revealed one record whose connector count had been mis-recorded, which was corrected against the project documentation. Finally, every feasibility label was verified against the sign of the corresponding *Power Margin*.

Because *C4.5* produces more compact and more interpretable trees when wide numeric ranges are grouped into technically meaningful intervals, the numeric attributes were discretized using the thresholds in Table 3. The thresholds were derived from the operating practice of the company rather than from equal-width binning, so that each category corresponds to a recognizable installation condition.

Table 3. Discretization of the numeric attributes

Attribute	Value range	Category	Technical basis
Cable_Length	≤ 5.0 km	Short	Building or campus installation
	5.01 – 10.0 km	Medium	Inter-building installation
	> 10.0 km	Long	Backbone or inter-site installation
Total_Loss	≤ 8.0 dB	Low	Below the typical SFP loss budget
	8.01 – 15.0 dB	Medium	Critical range, transceiver dependent
	> 15.0 dB	High	Above most transceiver loss budgets
Power_Margin	> 3.0 dB	Good	Sufficient reserve with tolerance
	0.0 – 3.0 dB	Fair	Minimal reserve, needs attention
	< 0.0 dB	Poor	Received power below sensitivity
Connector_Count	≤ 4	Few	Single-segment installation
	5 – 7	Medium	Multi-segment installation
	≥ 8	Many	Complex multi-point installation
Splice_Count	≤ 4	Few	Route shorter than two cable drums
	5 – 9	Medium	Route of two to four cable drums
	≥ 10	Many	Route longer than four cable drums

The discretized data set was then divided by the holdout method at a ratio of 80:20, producing a training set of twenty-four records (records 1–24; fourteen feasible, ten not feasible) and a testing set of six records (records 25–30; four feasible, two not feasible), as marked in Table 2. The split kept the class proportion comparable in both subsets, with 58.3% and 66.7% feasible records respectively.

2.5. System Development and Testing

The rules extracted from the tree were implemented in a multi-tier web application. The client layer is rendered by *Next.js* 16 with *TypeScript*, *Tailwind CSS*, and *Radix UI*; the server layer exposes REST endpoints protected by *JSON Web Token* middleware with two roles, *Administrator* and *Engineer*; and the data layer uses *MySQL* accessed through a connection pool [16], [17]. The computation logic resides in a single module containing *calculateLinkBudget()*, *calculateEntropy()*, *calculateGainRatio()*, *classifyNetwork()*, and *getOptimizationRecommendations()*.

Three testing methods were applied. *Black Box Testing* exercised every functional feature against its expected output, including negative scenarios such as invalid credentials and unauthorized access [21]. *White Box Testing* used *Basis Path Testing* on the three core functions: a flow graph was derived for each function, its *Cyclomatic Complexity* V(G) was computed, and one test case was designed for every independent path. Accuracy testing compared the output of *classifyNetwork()* with a manually computed ground truth on twenty scenarios built from real transceiver specifications across the 1G, 10G, 40G, and 100G categories, where the manual label applied a margin threshold of 3 dB; within each capacity category, routes inside and beyond the transceiver reach were included so that both sides of the decision boundary were exercised. The comparison was summarized in a confusion matrix from which *Accuracy*, *Precision*, *Recall*, and *F1-Score* were computed.

Because the twenty scenarios verify that the software reproduces the link budget rule, a separate evaluation was carried out to estimate how well the learned tree generalizes to records it has not seen. Three procedures were applied to the thirty field records of Table 2: the 24/6 holdout split described in Section 2.4; leave-one-out cross-validation (LOOCV), in which the tree is rebuilt thirty times and each record is tested once by a tree that never saw it; and stratified 5-fold cross-validation repeated ten times with different random partitions, giving 300 test predictions [18], [19]. In every fold the tree was rebuilt from the training portion only, using the gain ratio criterion of (8). Two variants of *C4.5* were evaluated: one on the discretized attributes of Table 3, as implemented in the system, and one in which the numeric attributes are split on thresholds learned from the data [20]. *Naïve Bayes* on the discretized attributes, *K-Nearest Neighbor* with k = 3 on standardized numeric attributes, and *Random Forest* with 100 trees were evaluated under the identical protocol for comparison. Because a proportion estimated from twenty or thirty cases carries considerable sampling uncertainty, 95% confidence intervals were computed with the Wilson score method [22].

3. RESULT AND DISCUSSION

This section presents the decision tree obtained from the training data, the rules extracted from it, the implementation of those rules in the web application, the results of the three testing methods, and a cross-validated comparison with other classifiers.

3.1. Attribute Evaluation and Root Node Selection

The training subset contains twenty-four records, fourteen feasible and ten not feasible, which gives an initial entropy of 0.9801 according to (5). A value that close to one confirms an almost balanced class distribution and therefore a high level of uncertainty before any split is made; it serves as the baseline against which every candidate attribute is measured. Each attribute was then evaluated with (6) to (8), and the results are recapitulated in Table 4.

Table 4. Information gain and gain ratio of all attributes

No.	Attribute	Information Gain	Split Info	Gain Ratio	Rank
1	Power_Margin	0.8176	1.5613	0.5237	1
2	Cable_Length	0.7284	1.5772	0.4619	2
3	Connector_Count	0.7284	1.5772	0.4619	2
4	Total_Loss	0.7097	1.5850	0.4478	4
5	Splice_Count	0.6357	1.5772	0.4030	5
6	Transceiver_Type	0.4089	1.7840	0.2293	6

Power_Margin obtained the highest gain ratio at 0.5237 and was therefore selected as the root node. The choice is also sound in engineering terms, because the power margin is the direct difference between the received power and the minimum sensitivity of the receiver, and thus determines whether the optical signal can still be decoded at all. *Cable_Length* and *Connector_Count* share an identical gain ratio of 0.4619, which reflects the strong correlation between them: a longer route generally requires more connection points, so the two attributes carry equivalent information about feasibility. *Transceiver_Type* ranks last at 0.2293 for two reasons. It has four distinct values, which raises its split information to 1.7840 and reduces the ratio, and its SFP and SFP+ subsets remain mixed, with entropies of 0.7219 and 0.8113 respectively, showing that the module type alone cannot determine feasibility without the other technical parameters.

3.2. Decision Tree and Classification Rules

Splitting on *Power_Margin* produced two pure branches immediately. The nine records in the Good category are all feasible and the nine records in the Poor category are all not feasible, both with an entropy of zero, so each becomes a leaf node. Only the Fair branch, which contains six records with five feasible and one not feasible, retained an entropy of 0.6500 and required further splitting. Recomputing the gain ratio within that subset selected *Total_Loss* as the level-two node, and the recursion continued through *Cable_Length*, *Connector_Count*, and *Transceiver_Type* until every record was classified. The resulting tree, shown in Figure 1, has a maximum depth of five levels.

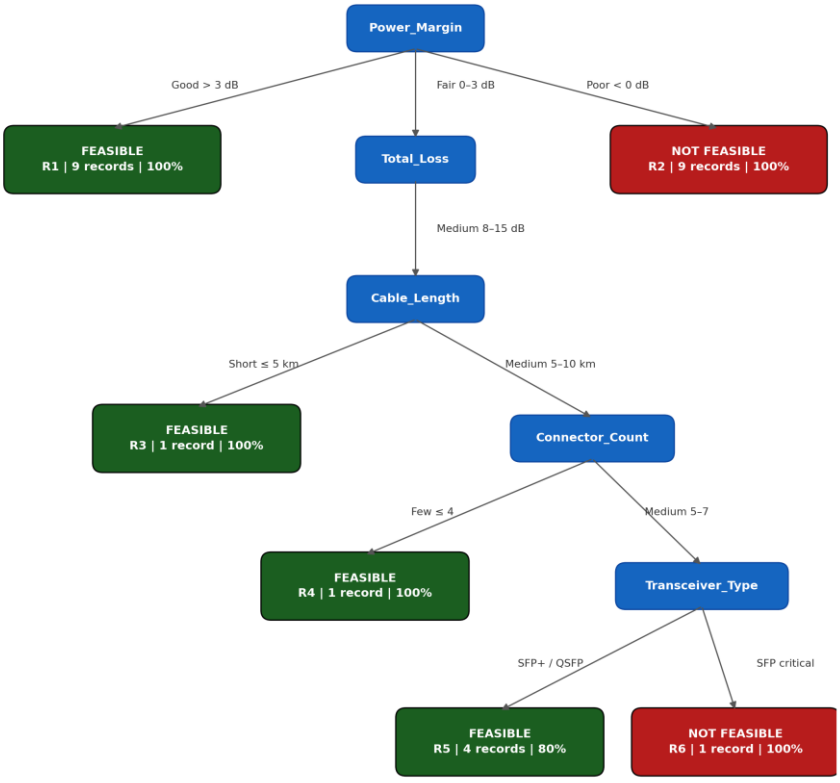


Figure 1. Decision tree produced by the C4.5 algorithm

Every path from the root to a leaf yields one *if–then* rule. Six rules were extracted in total, and they are summarized in Table 5.

Table 5. Classification rules extracted from the decision tree

Rule	Antecedent	Decision	Conf.	Support
R1	PM = Good	Feasible	100%	37.50%
R2	PM = Poor	Not feasible	100%	37.50%
R3	PM = Fair AND TL = Medium AND CL = Short	Feasible	100%	4.17%
R4	PM = Fair AND TL = Medium AND CL = Medium AND CC = Few	Feasible	100%	4.17%
R5	PM = Fair AND TL = Medium AND CL = Medium AND CC = Medium AND transceiver other than SFP at critical margin	Feasible	80%	16.67%
R6	PM = Fair AND TL = Medium AND CL = Medium AND CC = Medium AND SFP at critical margin	Not feasible	100%	4.17%

PM = Power_Margin, TL = Total_Loss, CL = Cable_Length, CC = Connector_Count

R1 and R2 together cover 75% of the training data, which means that in most cases the power margin alone settles the decision. The remaining four rules describe the marginal region in which the margin lies between 0 and 3 dB; there the outcome depends on how the loss was accumulated along the route rather than on its total value. R5 carries a confidence of 80% because of the single exception captured by R6, in which an SFP module with a lower transmit power cannot sustain transmission quality once every other factor is already at its limit.

3.3. System Implementation

The six rules were encoded in *classifyNetwork()*, which is invoked after *calculateLinkBudget()* has returned the loss breakdown. The analysis page collects eight parameters, namely fiber length, fiber attenuation per kilometre, connector count and loss per connector, splice count and loss per splice, transmit power, receiver sensitivity, safety margin, and transceiver type. After validation, the system returns the *Total Loss* breakdown, the *Link Budget*, the system margin, the feasibility status together with the reason behind it, and a transceiver

recommendation drawn from the company catalogue according to capacity and distance. Figure 2 shows the analysis page for a 25 km route.

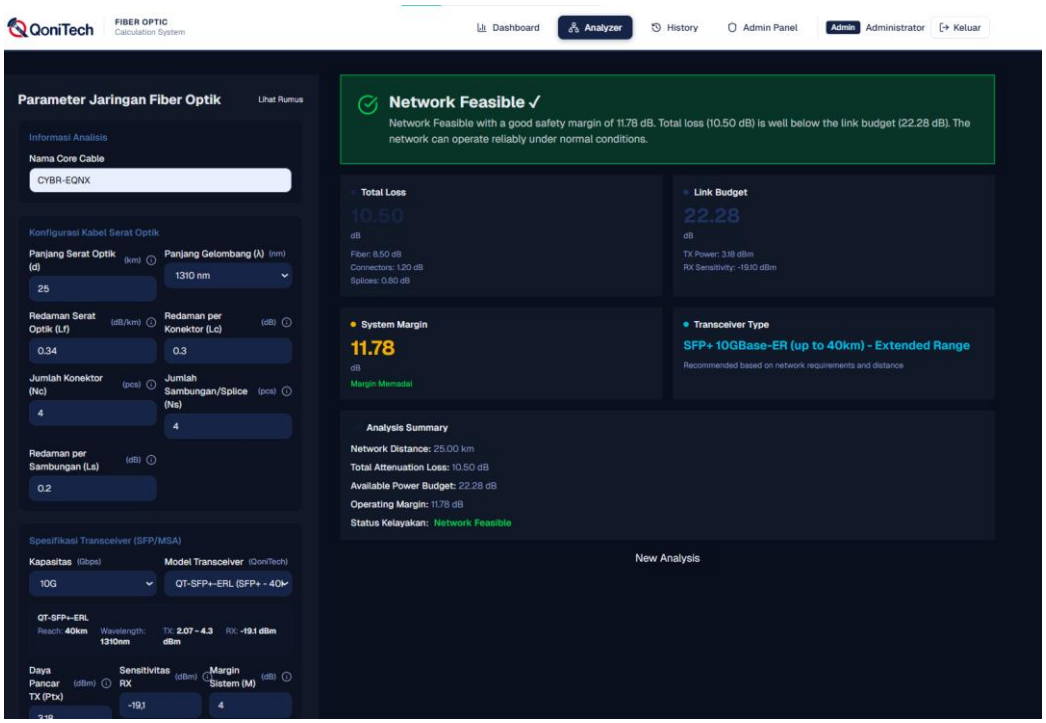


Figure 2. Network analyzer page of the developed system

In the example, a total loss of 10.50 dB against a link budget of 22.28 dB leaves a system margin of 11.78 dB, so the route is classified as *Network Feasible* and an extended-range SFP+ module is recommended. Every analysis is stored in the *analysis_history* table through the history endpoint, so that earlier evaluations remain traceable, while the *Administrator* role additionally manages user accounts and transceiver records.

3.4. Functional and Structural Testing

Black Box Testing covered authentication and role-based access, validation of the parameter form, the *Total Loss* and *Link Budget* computation, the feasibility classification, the history feature, user and transceiver management in the admin panel, and report download. Every scenario returned the expected output, including the negative cases in which invalid credentials were rejected and an *Engineer* account was denied access to the admin panel.

White Box Testing was applied to the three core functions of the calculation module. The results are recapitulated in Table 6.

Table 6. Recapitulation of white box testing

Function	Cyclomatic Complexity V(G)	Paths tested	Result
calculateLinkBudget()	2	2	100% valid
calculateEntropy()	3	3	100% valid
classifyNetwork()	10	10	100% valid

Fifteen independent paths were tested in total, covering simple branching, a loop, and the nested branching that implements the six classification rules. All of them produced the expected output, so the calculation and classification module can be considered free of logic errors at the level of the units tested.

3.5. Classification Accuracy

The twenty test scenarios yielded eleven cases labelled feasible and nine labelled not feasible by manual computation. The system label matched the manual label in every scenario, with no misclassification, as the confusion matrix in Table 7 shows.

Table 7. Confusion matrix of the C4.5 classification

	Predicted: Feasible	Predicted: Not Feasible	Total
Actual: Feasible	TP = 11	FN = 0	11
Actual: Not Feasible	FP = 0	TN = 9	9
Total	11	9	20

Table 8. Evaluation metrics

Metric	Formula	Computation	Result
Accuracy	(TP + TN) / (TP + TN + FP + FN)	(11 + 9) / 20	100%
Precision	TP / (TP + FP)	11 / (11 + 0)	100%
Recall	TP / (TP + FN)	11 / (11 + 0)	100%
F1-Score	2 × (Precision × Recall) / (Precision + Recall)	2 × (1 × 1) / (1 + 1)	100%

All four metrics reach 100%, as shown in Table 8, but the figure has to be read with care for two reasons. First, it measures implementation correctness, that is, the consistency between the rules encoded in *classifyNetwork()* and the link budget computation from which the manual labels were derived, rather than predictive accuracy on unseen field data. Second, twenty scenarios form a small test set: with no error observed in twenty cases, the 95% Wilson confidence interval for the true agreement rate is 83.9% to 100%, so the result rules out a gross implementation defect but cannot by itself establish error-free behaviour on every possible input. In this system *C4.5* is applied as a margin-threshold decision rule rather than as a model retrained on operational telemetry, which is precisely the distinction that explains the spread of accuracies reported elsewhere: 84% for SCADA link feasibility [10] and 90% for network attack detection [11] were obtained on noisy operational data, whereas 100% was also reported where the target variable followed an essentially deterministic rule [9]. The predictive performance of the learned tree is therefore assessed separately in Section 3.6.

3.6. Generalization on the Field Data and Comparison with Other Classifiers

On the 24/6 holdout split, the tree learned from records 1–24 classified all six test records correctly, namely the two not feasible routes with a Poor margin and the four feasible routes, three of which lie in the Fair margin region. Six records are too few for a reliable estimate, however, since the 95% Wilson interval for six correct out of six is 61.0% to 100%. The cross-validation results in Table 9 therefore give the more informative picture.

Table 9. Cross-validated performance on the 30 field records

Classifier	Attributes	LOOCV Accuracy	LOOCV F1-Score	5-fold ×10 Accuracy	5-fold ×10 F1-Score
C4.5 (as implemented)	Discretized	90.0%	91.9%	90.3%	92.1%
C4.5 (continuous thresholds)	Numeric	96.7%	97.1%	94.3%	95.2%
Naïve Bayes	Discretized	93.3%	94.7%	93.3%	94.7%
K-Nearest Neighbor (k = 3)	Numeric, scaled	86.7%	90.0%	86.7%	90.0%
Random Forest (100 trees)	Numeric	93.3%	94.4%	93.0%	94.2%

Feasible is the positive class. LOOCV yields 30 test predictions; repeated stratified 5-fold cross-validation yields 10 × 30 = 300. 95% Wilson interval of LOOCV accuracy: 27/30 = 74.4–96.5%, 28/30 = 78.7–98.2%, 29/30 = 83.3–99.4%

The discretized tree reaches 90.0% under LOOCV, misclassifying three of the thirty records, and the repeated 5-fold estimate of 90.3% shows that this figure is not an artefact of one particular partition. The three errors are informative. Records 5 and 24 have identical values on every discretized attribute, both being SFP routes of medium length, medium loss, and a Fair margin, yet record 5, with a margin of 2.40 dB, is feasible while record 24, with a margin of 0.00 dB, is not. Once the margin is collapsed into the 0–3 dB Fair category, no tree built on the discretized attributes alone can separate the two, which is why the rules in Table 5 need a critical-margin condition in this region. Record 2 is the only not feasible route with a medium connector count; when it is withheld, *Connector_Count* overtakes *Power_Margin* at the root, which shows that with thirty records the choice of the root node is still sensitive to individual records. The continuous variant removes both problems by placing a single threshold on *Power_Margin*, between 0.0 and 0.4 dB in the full data set, and reaches 96.7%. Its only LOOCV error is record 17, whose margin of 0.40 dB is the smallest positive margin in the data and therefore falls below the threshold learned without it. That one margin split suffices is itself a meaningful finding: in the company data, feasibility is decided by whether the received power exceeds the receiver sensitivity, in agreement with link budget theory [3], [4], [6].

Compared with the alternative classifiers, the differences are small. *Naïve Bayes* and *Random Forest* reach 93.3% under LOOCV, one record more than the discretized tree, while *K-Nearest Neighbor* reaches 86.7%, all of its errors being not feasible routes predicted as feasible because they lie close to feasible routes in the scaled attribute space. With thirty records, a difference of one or two correct predictions falls well inside the confidence intervals in Table 9, so none of the alternatives can be said to be more accurate than *C4.5* on this evidence. What separates them is interpretability: the *C4.5* tree can be written out as the rules in Table 5 and displayed to the engineer together with its *entropy* and *gain ratio*, whereas a vote of one hundred trees or

of the nearest neighbours offers no comparable explanation. The cross-validated accuracies of 90.0% and 96.7% are in the same range as the 84% reported for SCADA link feasibility [10] and the 90% for network attack detection [11], and they are a more realistic indicator of predictive performance than the 100% of Section 3.5. The comparison thus supports the choice of *C4.5* on the grounds stated in Section 1 and points to a concrete improvement: splitting *Power_Margin* on its continuous value rather than on three fixed categories raises the estimated accuracy from 90.0% to 96.7% without any loss of readability.

4. CLOSING

A web-based system for computing the *Link Budget* and *Total Loss* of optical fiber routes was successfully designed and built with *Next.js* 16, *TypeScript*, and *MySQL*. The formulations in (1) to (4) were implemented in *calculateLinkBudget()* and proved valid under *Black Box Testing* across all calculation scenarios and under *White Box Testing* with a cyclomatic complexity of two over two independent paths, so the first research objective was met.

The *Decision Tree C4.5* algorithm was applied successfully in *classifyNetwork()* to classify route status as *Network Feasible* or *Network Not Feasible*. *Power_Margin* emerged as the root node with the highest gain ratio of 0.5237, six classification rules were extracted from the resulting tree, and the function passed white box testing with a cyclomatic complexity of ten over ten independent paths. Accuracy testing on twenty scenarios derived from real transceiver specifications produced *Accuracy*, *Precision*, *Recall*, and *F1-Score* values of 100% (95% confidence interval 83.9–100%), and cross-validation on the thirty field records estimated the predictive accuracy of the implemented tree at 90.0%, comparable to that of *Naïve Bayes*, *K-Nearest Neighbor*, and *Random Forest*, so the second objective was met.

The system also proved able to help engineers and non-expert customers reach a feasibility decision more quickly and with less ambiguity, since it presents the loss breakdown, the feasibility status, the transceiver recommendation, and the rule behind each decision in plain language, supported by a history feature and an admin panel. Together with the elimination of manual spreadsheet computation, this addresses the problems identified at the outset and fulfils the third objective.

The study has several limitations. The tree was learned from thirty records of a single company, with one cable specification and four transceiver families, so the rules describe that operating environment and should be re-validated before being applied elsewhere. The twenty implementation scenarios and the six-record holdout set are small, which is reflected in the width of their confidence intervals. The discretization of *Power_Margin* into three fixed categories merges routes on both sides of the 0 dB boundary and accounts for most of the errors observed in cross-validation. Finally, the feasibility labels were assigned by engineers and verified against the computed margin, so the model reproduces engineering judgement on link budget grounds rather than predicting observed link failures.

Further work should enlarge the data set with routes from further projects and from other operators; split *Power_Margin* on a continuous threshold learned from the data instead of fixed categories; train the *C4.5* model on historical link performance data such as CRC errors, packet loss, and link flapping, so that the system becomes predictive rather than deterministic; introduce a warning class for margins that sit close to the threshold; compare *C4.5* with *Support Vector Machine* and gradient boosting once a larger data set is available; extend the application to a native mobile client for field use; and integrate it with the network management system already running at the company, so that operational telemetry can feed the training data set automatically.

ACKNOWLEDGMENT

The authors thank PT. Qonita Teknologi Indonesia for granting access to its field data, transceiver documentation, and engineering staff during this research, and the Department of Informatics, Faculty of Computer Science, Universitas Pamulang, for the academic guidance provided throughout the work.

REFERENCES

[1] R. Perdana *et al.*, "Jenis Fiber Optik Berdasarkan Jumlah Mode dan Indeks Bias: Tinjauan dan Perbandingan," *Jurnal Pembelajaran IPA dan Aplikasinya (QUANTUM)*, vol. 2, pp. 61–68, 2022.

[2] P. Agustian, Z. Anthony, A. Anungrah, A. Premadi, and A. Y. Dewi, "Analisa Kualitas Penyambungan Kabel Fiber Optik Terhadap Kecepatan Jaringan Internet," *MOTEKAR: Jurnal Multidisiplin Teknologi dan Arsitektur*, vol. 2, pp. 852–857, 2024.

[3] S. H. Jeffri, "Analisis Power Link Budget pada Jaringan FTTH di Kelurahan Jatirasa Bekasi," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, 2024.

[4] M. Ulfah *et al.*, "Feasibility Analysis of FTTH Network Design with Link Budget and Rise Time Budget Methods," *Buletin Poltanesa*, vol. 26, pp. 283–291, 2025.
[5] F. Z. Algusti, "Analisa Performansi Jaringan Optik di UIN Suska Riau Menggunakan Metode Power Link Budget dan Rise Time Budget," Universitas Islam Negeri Sultan Syarif Kasim Riau, 2023.
[6] A. Mulia and F. Latifah, "Implementasi dan Optimasi Jaringan FTTB GPON pada Apartemen Sunter Park View," *Journal of Information System, Informatics and Computing*, vol. 9, pp. 246–256, 2025.
[7] A. Srirahayu *et al.*, "Review Paper Data Mining Klasifikasi Data Mining," *Jurnal Ilmiah Informatika Global*, vol. 14, pp. 7–12, 2023.
[8] Z. Gustiana, "Performance Evaluation Algoritma C4.5 pada Klasifikasi Data," *Djtechno: Jurnal Teknologi Informasi*, vol. 5, pp. 289–296, 2024.
[9] F. Pirmansyah and T. Wahyudi, "Implementasi Data Mining Menggunakan Algoritma C4.5 untuk Prediksi Evaluasi Anggota Satuan Pengamanan: Studi Kasus PT. YIMM Pulogadung," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, vol. 3, pp. 343–356, 2023.
[10] T. D. Cahyono and R. E. Pratama, "Sistem Rekomendasi Kelayakan Link SCADA Menggunakan Decision Tree C4.5," *Dinamika Informatika*, vol. 15, pp. 83–91, 2023.
[11] V. S. Esterlin, "Deteksi Serangan dalam Jaringan Komputer dengan Algoritma Pohon Keputusan C4.5," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, pp. 323–327, 2024.
[12] R. Fransdela, "Penggunaan Algoritma C4.5 untuk Prediksi Waktu Penyelesaian Gangguan pada Layanan Internet," *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 10, pp. 7083–7094, 2025.
[13] A. Simanjorang, A. M. H. Pardede, and S. Syahputra, "Implementasi Algoritma C4.5 untuk Memprediksi Penjualan Paket," *Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, vol. 6, pp. 106–115, 2024.
[14] K. A. Putri *et al.*, "Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4.5)," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, pp. 39–47, 2024.
[15] R. Ramdani *et al.*, "Implementasi Algoritma C4.5 untuk Klasifikasi Gizi Balita Berdasarkan Data Posyandu Flamboyan 1 Kampung di Pojok Hilir," *JRIIN: Jurnal Riset Informatika dan Inovasi*, vol. 3, pp. 2999–3006, 2026.
[16] R. Kurniah, "Penerapan Sistem Informasi Berbasis Web untuk Pengelolaan dan Pengarsipan Dokumen," *Infotek: Jurnal Informatika dan Teknologi*, vol. 6, pp. 258–267, 2023.
[17] R. D. Lestari *et al.*, "Sistem Basis Data Perpustakaan," *Jurnal Sistem Informasi dan Aplikasi*, vol. 1, pp. 71–79, 2023.
[18] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 2, pp. 1137–1143, 1995.
[19] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS ONE*, vol. 14, no. 11, e0224365, 2019.
[20] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[21] N. Mahrozi *et al.*, "Pengujian Aplikasi dengan Metode Blackbox Testing: Analisis Boundary Value dan Equivalence Partitioning pada Aplikasi Sistem Pakar Kucing," *Scientica: Jurnal Ilmiah Sain dan Teknologi*, vol. 2, pp. 257–265, 2024.
[22] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.