Determining success factors for project with supervised machine learning

Grant Saxena
www.aetherarcade.com

Abstract

Every year, enormous project failure rates plague business companies across the globe, costing millions of dollars for each failed project. It is critical to understand the factors that influence project success. The goal of this study is to empirically discover the factors that lead to a successful project. This study used data from 469 projects to use three distinct supervised machine learning methods for classification: a) The Support Vector Machine (SVM), 2) The Probit regression, and 3) the Logistic regression. Five factors have been chosen: Support from the top management, Technical skills, Client’s acceptance, Communication, and Monitoring. The findings of the SVM study revealed that SVM could accurately predict successful and unsuccessful projects. The findings of Logistic and Probit regression revealed that project success is more likely if project managers get appropriate top-level support and if the project team has sufficient technical capabilities. This study also discovered that good communication and control improve the odds of a project’s success. However, according to the results of this research, client acceptability is a minor success factor for a project.

1. Introduction

Historically, project success has been described as a project that achieves its goals on time and on budget. In many sectors, this assessment criteria has remained the most frequent. However, in the case of a development project, success involves providing the benefits and satisfying the expectations of beneficiaries, stakeholders, donors, and funding agencies, in addition to achieving the schedule and budget objectives (Newton, 2013). However, identifying these
aspects of success is more challenging, and some may only be assessed years after the project is finished, which is problematic for many organizations owing to a lack of financing (Lester, 2007).

To assist companies in determining success, a difference between project accomplishment and project management accomplishment must be established. Project success is defined as a degree of effectiveness in which project deliverables are evaluated in terms of advantages and stakeholder satisfaction, or the amount to which the project's ultimate goals are met (Institute, 2019). The efficiency with which a project was able to meet its goals is a measure of project management success (Juli, 2011). Efficiency refers to how well a project uses its limited resources to achieve its objectives while maintaining positive connections with both internal and external players. On the other hand, a project may fail in a variety of ways. For example, a project might fail to achieve the budget, schedule, or scope goals but succeed in achieving the development objectives, or a project can reach the budget, schedule, and scope goals but fail to meet the eventual development goals (Wright & Lawlor-Wright, 2000.) (Padgett, 2009).

When a project fails to produce what was expected within the accepted budget and time frame, it is considered a failure. In most instances, however, the stakeholders determine if the project was successful or failure based on their assessment of the result and satisfaction with it. The same businesses agonize over the reasons for project failure, hire costly consultants to evaluate and recover failed projects, and often terminate what seemed to be well-planned, well-organized, and destined for success initiatives (Padgett, 2009). In every company and on any project, project failure may occur (Camilleri, 2016). Failure may have an endless number of causes. It is sometimes beyond the project manager's and/or team members' control. Failure can sometimes be managed (Camilleri, 2016).

2. Determinants

The following factors are critical to project’s success:

1) Support from the top management

This is a critical consideration. Any initiative that wants to thrive must have full support from management, since this support is critical to its success. Because project management provides the route for top management plans to be implemented, it also relies on upper executives for authority or direction, making top management support critical to project success (Gaer et al., 2010). Top management commitment to projects may go a long way toward reducing disputes in resource allocation and priority ranking. Support from top management shows itself in the deployment of time, people, and financial resources. It may go a long way toward boosting project managers' morale by assuring them of upper management's backing in contentious situations.
2) Technical skills

Personnel is required, as is technical knowledge, in order to guarantee that the project is adequately supervised. Technical skill refers to handling the situation in which the project implementation team has both the required people and the required expertise. Inefficiency and cost ineptitude have been linked to technical ineptitude (Project, 2018). To avoid hazards to people and equipment, users of any technology must be acquainted with the mode of operation.

3) Client’s acceptance

This has to do with the client's consultant. Regardless of how effectively a project is executed, the final acceptability of a project is determined by how well the customer is guided through the various phases of implementation. Acceptance is a critical step in project implementation that needs the project's final user to engage in its execution, especially in the early phases, but throughout the project's life cycle. If the user is involved in the early phases of the project, there is a good chance of acceptance.

The client is anybody who will eventually benefit from the project's result. In order to complete a project, it is essential to discuss with clients. The extent to which clients are involved in project execution influences their acceptance for the project. Client consultation is also the initial stage in change implementation. Client consultation, on the other hand, should be needed throughout the project, since there is a propensity to abandon this necessity after an early client counselling activity.

4) Communication

For a project to be effective, it must have proper and sufficient communication routes. Communication guarantees that the project team works effectively within itself while simultaneously keeping in touch and exchanging information with the rest of the business and the user. Though there are many forms of communication, it has been shown that verbal communication is more effective than written communication (Institute, 2019). A high level of administrative communication is one of the characteristics of high-performing teams. The sharing of essential information, not just feedback, is at the heart of communication. Policy updates and new processes, as well as objectives, updates, and other information, are communicated. Communication isn't always easy. It may be used on its own, but it's usually used in conjunction with other components.

5) Monitoring

For project success, monitoring each step of project execution is critical. This necessitates receiving input from key people on how the project is progressing in contrast to original estimates. Appropriate monitoring and evaluation channels between the project leader and the user, as well as between the project leader and senior management, must be established (Camilleri, 2016). It's critical to keep track of not just the project's timeline and budget, but also the actions of the project team.
3. Methodology

Support Vector Machine (SVM)

The basic aim of SVM classification is to effectively separate classes based on a set of characteristics. As shown in Figure 1, we can demonstrate this with a simple two-class issue. The simplest situation, in which the classes are linearly separable, is shown to the left. The SVM algorithm separates classes by maximizing the gap between the closest components in each class (Zhou, 2021). The nearest members of each class on each side of the line are called support vectors.

![Figure 1 Simple 2-class support vector machine](image)

Binary models

In this kind of model, the dependent variable can only have two potential values. Y may be a dummy factor that represents the probability of a scenario happening or a decision between two options. The results of each match in a league sample, for example, may be helpful in modeling (whether successful project or not). The teams vary in a number of measurable qualities that we call x. The goal is to establish a link between team qualities and the likelihood of project success. y is a binary variable having a 0-1 range. Because the observed model of the conditional average puts inadequate restrictions on the model's residuals, just virtually regressing y on x is insufficient. Furthermore, the value of y in a basic linear regression is not restricted to zero. Rather, to meet the basic needs of binary regressors, we utilize a specification (Long & Mustillo, 2021). Assume that the chance of seeing 1 is represented as follows:
The kind of binary model selected is specified by the function $F$, which is a pure increasing continuous function that takes a true value and produces an output between zero and one (Lei, 2021). As a consequence:

We may use the maximum likelihood methods to determine the values of this model given such a specification (Rass et al., 2021). The likelihoods function may be written as follows:

$$l(\beta) = \sum_{i=0}^{n} y_i \log(1 - F(-x_i^j \beta)) + (1 - y_i) \log(F(-x_i^j \beta)).$$

To produce parameter estimates, an iterative solution is needed due to the nonlinear character of first-order requirements for this probability. This standard contains two conflicting points of view that are worth considering. Binary models are often used to define latent variables. Assume there is a hidden latent factor $y^*$ that is linearly related to $x$.

$$y^*_i = x_i^j \beta + u_i$$

$y^*$ controls if the measured response variable reaches a predetermined value, where $u$ represents random fluctuations:

$$y_i = \begin{cases} 1 & \text{if } y^*_i > 0 \\ 0 & \text{if } y^*_i \leq 0. \end{cases}$$

The threshold is set to 0 in this instance, although the value of the threshold is unimportant if $x$ has a constant term (Musa 2013). The marginal effect of a characteristic $x$ factor on a single conditional probability linked with the goal $y$ is calculated in the following way:

$$\frac{\partial E(y_i|x_i^j, \beta)}{\partial x_{ij}} = f(-x_i^j \beta) \beta_j.$$
Other variables' data are PCA factors of various answers gathered on the basis of five characteristics. Statista was used to compile the data.

4. Results

The shapes of relationships.

Figure 2. Pairplot of features
Figure 3. correlation among features and target

The correlation coefficients between characteristics and project success are shown in Figure 3. The findings indicate that project success is positively linked to the variables selected. This indicates that these characteristics are beneficial to the project's success.
Figure 3. Features correlation
Figure 4 shows the outcome of the confusion matrix information. None of the initiatives expected to fail were labeled as having insufficient funding when they actually did (type one error). As a consequence, no type one errors were found. Furthermore, 29 of the projected project expenditures that did not materialize were deemed successful (type two error). Table 1 shows the categorization report. The f1-score, precision, support, and recall values are all shown in this graph. Furthermore, these results suggest that the model is quite accurate. In summary, the SVM Model proved pretty accurate in determining if a project was successful or not.

Table 1. Classification report

|       | precision | recall | f1-score | support |
|-------|-----------|--------|----------|---------|
| 0     | 0.97      | 0.53   | 0.68     | 55      |
| 1     | 0.77      | 0.99   | 0.87     | 88      |
| accuracy |          |        |          | 0.81    | 143     |
| macro avg | 0.87    | 0.76   | 0.77     | 143     |
| weighted avg | 0.85    | 0.81   | 0.80     | 143     |
Probit and Logit results.

Table 2. Binary Probit model results

| Variable      | Coefficient | Std. Error | z-Statistic | Prob. |
|---------------|-------------|------------|-------------|-------|
| SUPPORT       | 0.824898    | 0.102930   | 8.014177    | 0.0000|
| TECHNICAL     | 0.218936    | 0.031456   | 6.960028    | 0.0000|
| MONITORING    | 0.008017    | 0.003591   | 2.232549    | 0.0256|
| COMMUNICATION | 0.783112    | 0.080208   | 9.763569    | 0.0000|
| ACCEPTANCE    | 8.71E-06    | 0.003744   | 0.002326    | 0.9981|
| C             | 24.18480    | 2.387374   | 10.13030    | 0.0000|

Table 2 shows the results of the Probit and Logit models. The findings indicate that, with the exception of “acceptance,” all of the characteristics in both models have a substantial effect on the likelihood of success. The coefficients for characteristics like top management support and communication are very high. This implies that these two variables are the most important elements in a project’s success. The McFadden R-square in both models is approximately 0.76, indicating that the models are statistically fit.
Table 3. Binary Logit model results

| Variable      | Coefficient | Std. Error | z-Statistic | Prob. |
|---------------|-------------|------------|-------------|-------|
| SUPPORT       | 1.491630    | 0.195312   | 7.637179    | 0.0000|
| TECHNICAL     | 0.398820    | 0.059381   | 6.716294    | 0.0000|
| MONITORING    | 0.013536    | 0.006415   | 2.109962    | 0.0349|
| COMMUNICATION | 1.417766    | 0.157218   | 9.017815    | 0.0000|
| ACCEPTANCE    | 0.001153    | 0.006790   | 0.169759    | 0.8652|
| C             | 43.74793    | 4.701918   | 9.304274    | 0.0000|

McFadden R-squared | 0.756928 | Mean dependent var | 0.627417
S.D. dependent var | 0.483918 | S.E. of regression | 0.223204
Akaikes info criterion | 0.342099 | Sum squared resid | 28.04873
Schwarz criterion | 0.387904 | Log likelihood | -91.32702
Hannan-Quinn criter. | 0.359972 | Deviance | 182.6540
Restr. deviance | 751.4400 | Restr. log likelihood | -375.7200
LR statistic | 568.7860 | Avg. log likelihood | -0.160504
Prob(LR statistic) | 0.000000 |                     |       

Obs with Dep=0 | 212 | Total obs | 569 |
Obs with Dep=1 | 357 |                     |     |

**Conclusion**

The variables that contribute to the success of a project were experimentally examined in this research. This research utilized 469 projects to test three different supervised machine learning methods for classification: Support Vector Machine (SVM), Logistic regression, and Probit regression.

When used to forecast project success, a technique based on SVM in kernel mode obtained 81 percent accuracy with acceptable true positive and false negative rates, suggesting that it is successful. There are many alternative methods, such as Naive Bayes, decision trees, and neural networks. When compared to the techniques mentioned above, the SVM has a better accuracy on the dataset utilized in this study.

The findings of the Probit and Logit models indicate that, with the exception of *Client acceptability*, all of the features/determinants in both models have a substantial effect on the likelihood of success. *Support from top-management level* and *communication* are two characteristics with high coefficients. This implies that these two factors are the most crucial elements in the success of a project.
References

Camilleri, E. (2016). Project success: critical factors and behaviours. 304.
Gaer, S., Lynn, S., Wong, B. L., Florez, M. C., Pomann, H., Pomann, J., Matsunaga, N. B., Macpherson, S. M., Bakin, B., Lockwood, R. B., Santamaria, J. C., Gwynne, S. (Steve), Greenberg, I. A., & Bixby, J. (n.d.). Project success.
Institute, P. M. (2019). A Guide to the Project Management Body of Knowledge (PMBOK(R) Guide-Sixth Edition. 1276.
Juli, T. (2011). Leadership principles for project success. 260.
Lei, C. (2021). Deep Learning Basics. In Deep Learning and Practice with MindSpore (pp. 17–28). Springer.
Lester, A. (2007). Project Management, Planning and Control. Project Management, Planning and Control.
Long, J. S., & Mustillo, S. A. (2021). Using predictions and marginal effects to compare groups in regression models for binary outcomes. Sociological Methods & Research, 50(3), 1284–1320.
Newton, R. (2013). The project management book.
Padgett, C. M. (2009). The project success method: a proven approach for achieving superior project performance in as little as 5 days. 244.
Project, C. Y. (2018). Chapter 15 Why Good Projects Fail Anyway. 15–16.
Rass, S., König, S., & Schauer, S. (2021). Semi-automated Parameterization of a Probabilistic Model Using Logistic Regression—A Tutorial. Game Theory and Machine Learning for Cyber Security, 438–484.
Wright, A. (Andrew J., & Lawlor-Wright, T. (n.d.). Project success and quality: balancing the iron triangle.
Zhou, Z.-H. (2021). Support vector machine. In Machine Learning (pp. 129–153). Springer.