How country-level indicators and project-level measures affect world bank project performance

Zhuo Leng*

June 2017

(version 17.04.a)

Abstract

This paper investigates country level indicators and project level measures correlated to world bank project, using data from 11000 World Bank projects evaluated by IEG group between 1983 and 2015. By running several machine learning predictive models, this research also use these features to analyzes World Bank project performance prediction at time of project approval. project-level measures, including project duration, lending cost and total commitment amount, etc. seems strongly correlated with project performances. However, country-level variables, including GDP and life span also matter significantly for project satisfactory.The result shows that overall, Random Forests performed the best according to precision at 5% (our preferred evaluation metric). Also the analysis highlight country level indicators, such as GDP and life span are more importance features than project level factors that could positively affect project performance.

keywords: World Bank project performance, Country-level indicators, project-level measures, performance prediction

JEL classification: D91, E21, H30

*University of Chicago, Social Science Division, zleng@uchicago.edu.

1 Introduction

The World Bank Group, the largest international development bank, awards about 20,000 - 30,000 contracts annually, which are worth over $60 billion. This funding supports development in areas such as education, health, and agriculture in order to improve the quality of life in impoverished and middle income countries. According to the Policy Reacher Working Paper from the Independent Evaluation Group, the monitoring and evaluation helps projects achieve their objectives. Usually there are a number of reasons that lead to unsatisfactory projects. Fraud, corruption, and collusion are a few select reasons that contribute to these underperforming projects. More than 5% of the world GDP ($2.6 trillion) is lost annually due to these factors. In addition to funding waste, underperforming projects also sap resources from the projects that most need development funding.

We should care about successful factors that could affect project performance for funding efficiency and project quality control. Commonly, the successful factors can be divided to two groups: Country level indicators and project level measures. If, with all the information of factors needed, we are able to predict the performance of the project in the decision making phase, the World Bank Group can veto the proposal and prevent the unsatisfactory project. Moreover, if a project is still approved with an unsatisfactory-project prediction, the World Bank Group can raise the monitoring level of the project to prevent it from failure. The project funding from the World Bank Group will be better used if we can understanding the correlated factors and predict the performance of projects in advance.

Giving the importance of project evaluation and prediction, I conduct a research on how country-level indicators such as GDP and life span and project level factors, such as project duration affect the potential performance of world bank projects. I use machine learning techniques to evaluate different models and eventually identify the best model to approach the research question. My findings offer important in-

sights to help world bank improve rate of substantiated investigation.

2 Literature Review

Before my research, I review works that related to the World Bank project performance study and try to gain insights in terms of what kind of features are more significant for predicting the project outcome and how those findings can generate practical implementation for the World Bank. Due to the nature of this research and data source structure, this literature review is divided into two parts: works that using project related factors to analysis world bank project performance and works that using country/government related factors to analysis the performance.

Related works on project level factors

Since the dataset available to me has more project related variables, I would like first to review academic works that focus on project related variables to analysis world bank performance outcomes. Many empirical studies have discussed the impacts that project level factors could have on the overall project satisfaction.

A series of researches have been conducted to examine the effect of World Banks data and performance on the project level. Some of them evaluate a broad range factors and tries to have an overall picture of the importance of each factor, while many other researches dig deep into one or two factors and provide really well understanding on the specific area. For example, in the article Does Participation Improve Performance? Establishing Causality with Subjective Data, Isham, Narayan and Pritchett explore the relationship between the participation of beneficiary and the project outcomes. When analyzing the data from world bank projects, Isham, Narayan and Pritchett believe that there is an upward bias towards the measure of participation, due to subject nature of this particular data. In order to estimate the true effect

of project participation, Isham, Narayan and Pritchett also add project level control variables, such as project timing into the model. This work reaches the conclusion that the participant of beneficiary does have a statistically significant positive effect on overall project outcomes. Even though this study focuses only on a narrow discipline, which is the rural water projects, this work offers a great example of how to study a specific project level factor within a particular context.

On the other hand, in the study of Good countries or good projects? Macro and micro correlates of World Bank project performance, Denizera, Kaufmannb and Kraayd investigate a wider range of micro level project related factors from over 6000 aid-financed development projects. This work uses the same dependent variable, which is the satisfaction of projects, and proof that country level macro measures such as policies and institutions do have a strong correlated with the performance outcome. However, after adding project level factors, Denizera, Kaufmannb and Kraayd believe those micro level factors such as project size and effort devoted to project preparation have an even strong impact power on the project performance. The result of study makes me more decisive to peruse a research that include and compare a difference set of country level and project level factors and testify if the finding can be applied.

Related works on country/government level factors

Beside project level factors, my research also evaluates the influence of country level factors on the performance outcome of World Bank Performance. In order to do this, I brought in country level factors, such as annual GDP, average life span and several indicators. Before analyzing the model, I conducted research on academic works that focusing on using country level factors to analyze project performance.

In the article Civil Liberties, Democracy, and the Performance of Government Projects,Isham, Kaufmann and, Pritchett focus on studying the relationship between World Bank project performance and country level political factors, such as civil strife and civil

liberties. Due to the topic of this study, Isham, Kaufmann and, Pritchett only include government investment projects from World Bank. After controlling other determinant of World Bank project performance such as democracy level, countries with better political and governance condition do perform better: countries with the strongest civil liberties have projects with an economic rate of return 822 percentage points higher than countries with the weakest civil liberties. This result demonstrates a strong link between governance performance and project performance, which inspire me to include and analyze more country level factors and compare them with project related factors. Other works have studied a broader range of country level factors. For example, in the study Project Evaluation and Uncertainty in Practice: A Statistical Analysis of Rate-of-Return Divergences of 1,015 World Bank Projects, Pohl and Mihaljek include more economical index on country level while controlling sectoral and geographical factors. The result of this study indicate that macro level economics factors have a strong influence on the project outcome, but the current project evaluators have not taken those factors into significant account. Besides, Pohl and Mihaljek also demonstrate that project performance significantly differs from each other depending on the geographic location of the project.

However, the dependent variable in this study is different from the study I conducted: this study analyzes rate of return of each project as the outcome variable, while I use the satisfaction rate of world bank project as a binary dependent variable. Furthermore, Pohl and Mihaljek focus on a relative smaller project sample size (1015 projects). This result in very different methodology in terms of statistic modelling techniques.

Contribution

Existing literature related to the performance of World Bank series can roughly devided into three categories: focused analysis on project level factors, focused analysis on country level factors and general analysis across both levels. Among these works,

most of them have uncovered the factors they choose to study have significant impact that the overall performance of World Bank projects. However, all of the studies listed above only analyze a specific category of project. The study I would like to begin would explore new factors that could potentially influence the project performance on both the country and project level, but have not been analyzed during the previous studies. Also by learning the methodology and techniques from literature, my study would compare difference factors on a macro level, using more generalized project data.

3 Data

I'm using three main data sources: (1) World Bank Projects summary, which is from WB projects API; (2) World Bank Projects performance, which is from WBs IEG dataset; (3) World economic development indicators. The last one is supplementary data relative to countries socio-economic context.

1) Projects Performance Dataset

It provides access to basic information on all of the World Banks lending projects from 1947 to the present. It includes information such as: project title, task manager, country, project id, sector, themes, commitment amount, product line, and financing.

2) World Bank Project Performance Ratings

This dataset contains all World Bank project assessment carried out by Independent Evaluation Group (IEG) in the 70s, spanning over 30 years. It includes more than 11,300 projects assessments, covering more than 9,600 complete projects.

3) World Economic Development Indicators.

This dataset is compiled from officially-recognized international sources. It presents the most current and accurate global development data available, and includes national, regional and global estimates. We extract countrys mean GDP per capital and

average life expectancy between the project period. Then I join these two features with our dataset by country and year.

Table 1: variables table

Feature Type	Features Selected
Numerical	leading project cost, ibrd commitment amount, ida commitment amount, total commitment amount, life expectancy at birth, the GDP index
Categorical	region, country name, product line, leading instr type, agreement type
Aggregated	total number of projects done for the past years by the each country, total commitment amount for the past years, average GDP, average life span index, cost-commitment ratio, project duration time

Table 2: descriptive statistics

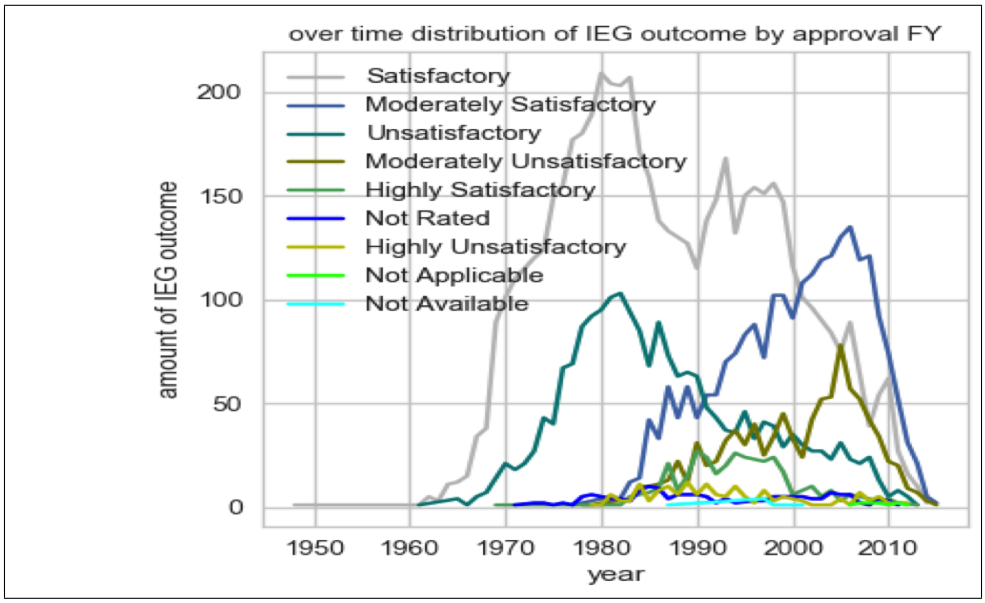
variables	count	mean	std	min	max
Approval FYs	11726.0	1991.0	11.45	1948	2015
Lending Project Cost	6.23e+03	1.66e+08	4.99e+08	6.00e+00	2.69e+10.0
Net Commitment	1.15e+04	6.43e+07	1.18e+08	-7.46e+08	3.00e+09
Exit FY	11726.00	1994.07	76.77	0.00	2016.00
IEG_{EvalFY}	111726	1999.19	10.55	1973.00	2017.00
ERR at Appraisal	4566.00	25.32	19.94	0.27	541.50
ERR at Completion	3979.00	22.23	28.85	-100.00	747.00
grantamt	1.17e+04	6.55e+05	4.69e+05	0.00e+00	1.19e+08
totalcommamt	1.17e+04	7.33e+07	1.25e+08	0.00e+00	3.00e+09
totalamt	1.17e+04	7.26e+07	1.25e+08	0.00e+00	3.00e+09
themecode	6.84e+03	2.07e+09	2.95e+09	2.10e+01	9.39e+09

The tables above present basic descriptive statistics for numerical features. We can already see some anomalies in the data, such as projects with a zero value for amount

features and Exit FY with value zero.I will do data visualization of some categorical features in next step.

From the project performance dataset made by the Independent Evaluation Group, we get the label variable IEG Outcome, which is the evaluation outcome of each project. There are three levels both for Satisfactory and Unsatisfactory, and we consider all three levels of "unsatisfactory" as bad performance projects. There are 180 projects that lack evaluation, and we chose to drop these rows since it is a fairly small number and we do not have a better way to fill them. From the graph below, we can see the changes of all levels of project performance over the years in **Figure 1** below.

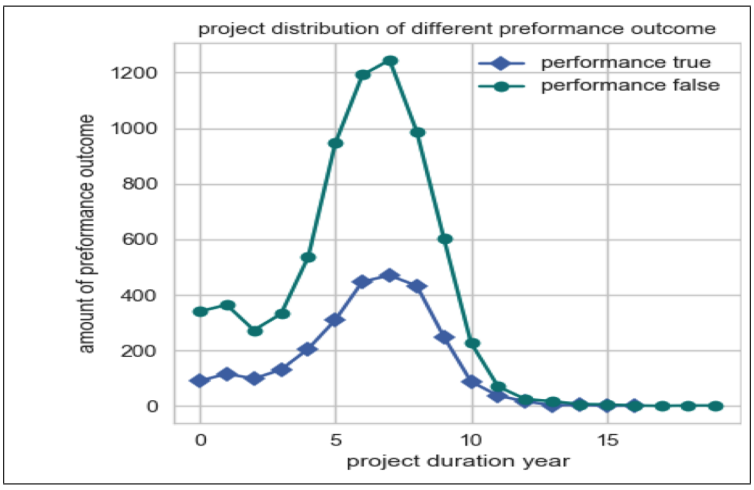
Figure 1: Distribution of IEG outcome by approval FY



There are 11726 projects with the approval year from 1948 to 2015. The average project number for each year is 280 with the standard deviation of 68. From the **Figure 2** below, we can see there is no statistical difference between the bad performance (true) and the good performance projects in terms of project duration time distribution. Projects generally have a long duration time, mostly more than five

years long.

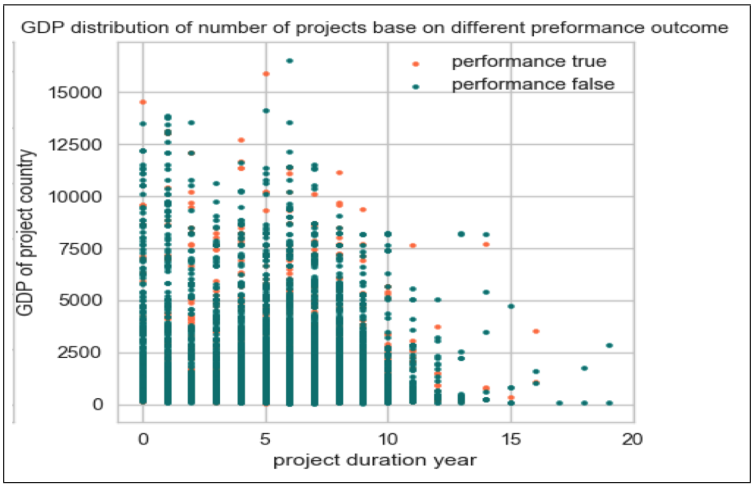
Figure 2:Distribution of project performance by project duration



From the project summary dataset, we get the theme or type related information, the geographical information and the cost-amount information of each project. Further aggregated features are developed mainly from this dataset.

And also, we join GDP and life span variables as country indicators. **Figure 3** below is the distribution of these two country level indicators.

Figure 3: GDP vs. life span



4 Method

Because world bank current project investigative process is complaint-driven and not accurate (2010), so I want to use machine learning system to detect world bank project bad performance by using classification system. Thus I could create automated mechanism to help them save cost and times as well as improve rate of substantiated investigation.

First, I need to perform features engineering to create my features. Beside 8 numerical features, there are 5 categorical features, I transfer them to dummy variables. Then I also aggregated new features as **Table 1** shows. For instance, I generate feature: project duration years by using project end years minus project Approval fiscal years.

The next step is to build a classifier to assign 'performance score' to World Bank Project in order to predict bank project performance. I will firstly try Random Forest, Adaboost, Decision Tree and random forest classifier. I will mainly do these two step:

1) Modeling:
   Adaboost ,
   Random Forest,
   Decision Tree,
   Logistic Regression
2) Rank model excellence based of these four metrics (AUC, Precision, Recall and f1).

I cross-validated my models using different train and test sets taking time into account. In particular, I tried a combination of train lengths of 5, 10, 20 and 30 years, and test lengths of 3, 5 and 8 years, to determine an appropriate training sample length.

Then in order to understand successful factors, including country level indicators and project level factors to affect world bank project performance, I use random forest to select very importance features to see which feature can strongly affect project performance.

I will use these features to build logistic regression model to better interpret the coefficient:

$$(\text{performance} = 1 \mid \text{S, P, I}) = \frac{\exp(\beta_0 + \beta_1 \text{feature1} + \beta_2 \text{feature2} + ...\beta_{12} \text{featuren})}{1 + \exp(\beta_0 + \beta_1 \text{feature1} + \beta_2 \text{feature2} + ...\beta_{12} \text{featuren})} \quad (1)$$

5 Result

My goal for the first step is to create a model which could use country-level indicators and project-level measures to predict and detect the potential bad performance projects of the World Bank at the time of the projects publication before I get important features of country. The dependent variable of the data set we care about is the performance outcome evaluated by the World Bank. It is a categorical variable with three levels of satisfactory and unsatisfactory. There are 180 projects out of 11,717 that lack of evaluation. I drop these rows since it is a fairly small number and we do not have better way to fill them. I divide the performance outcome into two groups, good performance and bad performance. And we convert the dependent variable into a Boolean variable with True for good performance and False for bad performance.

I evaluated the models according to different metrics. In my particular analysis, I gave a higher weight to models with higher precision at 5 percent. I think it is impossible for the World Bank to inspect all projects, so my proposal and assumption is that they care about the predicted 5% worse projects. I also used AUC as a tiebreaker metric. The following **Table 3** shows the results for above four kind of

classifiers according to precision at 5%:

Table 3: AUC-ROV evaluation

Model type	Parameter	AUC-ROC
RF	'max_depth': 10, 'min_samples_split': 5, 'n_estimators': 20, 'max_features': 'log2'	0.721645
DT	'criterion': 'entropy', 'max_features': 'sqrt', 'min_samples_split': 10, 'max_depth': 5	0.692314
AB	'algorithm': 'SAMME.R', 'n_estimators': 100	0.721576
LR	'C': 0.001, 'penalty': 'l1'	0.673455

Random Forests and Adaboost appear to be the best performing models according to precision at 5%.After analyzing my models, I decided to select a Random Forest model with the following parameters:

Number of estimators : 20

Min samples split : 5

Max features : log2

Max Depth : 20

The following figure shows specific evaluation measures for our model.From **Figure 4** we can see that precision is high for the projects with higher risk of being underperformed (before 20%). We also see that this model is better that random assignment (about 20% precision).

Next, we studied top 10 feature importance of our selected model, which is presented in the following **Figure 5** below. We see that countries indicators such as GDP and life expectancy are strong predictors for project performance. We also see that project related features such as amount involved and type of instruments are good predictors as well.

Figure 4: Precision-Recall Curve

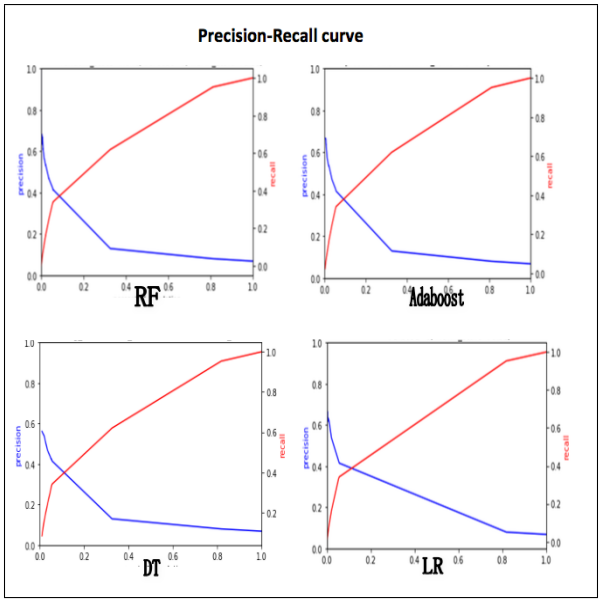
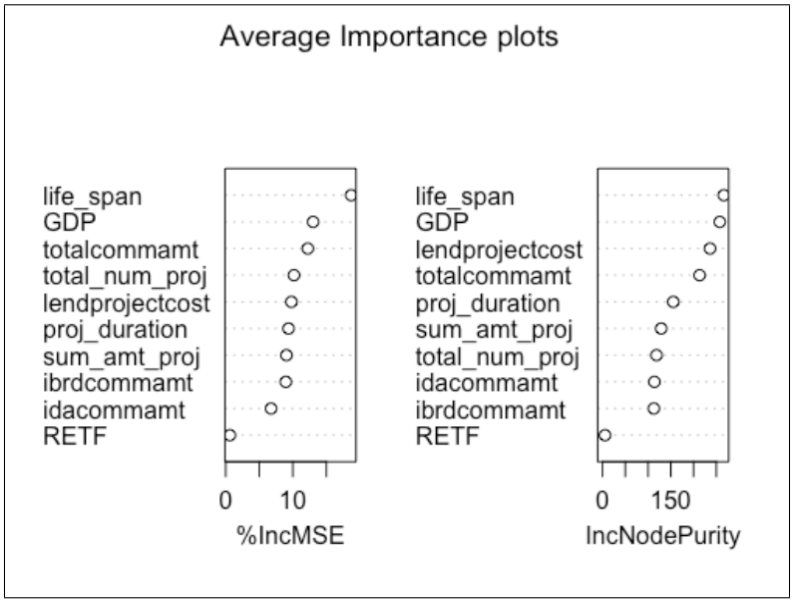
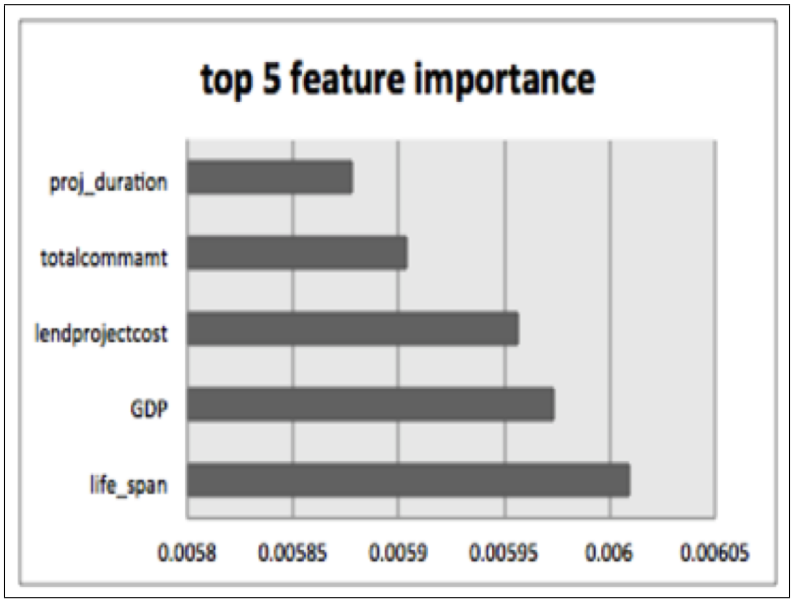


Figure 5: top 10 Feature importances for selected RF



I will select the top 5 as **Figure 6** shown to fit logistics regression model and interpret their coefficient to better understand how these features affect world bank project performance.

Figure 6: top 5 Feature importances for selected RF



In my preliminary logistic regression model, I use all five variables to predict performance satisfaction likelihood. From the summary table **Table 4**,we could know GDP is positive significant, life_span is negatively significant, and project_duration is positively significant. The model formula is as below:

glm(performance ~.,data=bank, family="binomial")

Then check skewness of variables, I find that GDP, lendprojectcost, totalcommamt are skewed (skewness¿5). Then I decide to take log on these three variables. From the summary table **Table 5** below, AIC decreases from 11465 to 11426 compared to model 1. Model 2 is preferred.

Table 4: Summary table for model1

```
Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2316  -0.8210  -0.7325   1.4117   2.0341
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      7.469e-01  1.785e-01   4.185 2.85e-05 ***
## X                1.904e-05  7.680e-06   2.479 0.013175 *
## life_span       -3.441e-02  3.112e-03 -11.059  < 2e-16 ***
## GDP              7.174e-05  1.486e-05   4.828 1.38e-06 ***
## lendprojectcost -1.890e-10  1.221e-10  -1.548 0.121573
## totalcommamt    -1.057e-11  2.665e-10  -0.040 0.968361
## proj_duration    3.078e-02  9.151e-03   3.363 0.000771 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11601  on 9883  degrees of freedom
## Residual deviance: 11447  on 9877  degrees of freedom
## AIC: 11461
```

Table 5: Summary table for model2

```
# Call:
## glm(formula = performance ~ life_span + log(GDP + 0.1) +
log(lendprojectcost +
##     0.1) + log(totalcommamt + 0.1) + proj_duration, family =
"binomial",
##     data = bank)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3542  -0.8123  -0.7349   1.3810   1.9540
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                0.655110   0.330094   1.985  0.04719 *
## life_span                 -0.044861   0.003714 -12.080  < 2e-16
***
## log(GDP + 0.1)             0.257044   0.033019   7.785 6.99e-15
***
## log(lendprojectcost + 0.1)  0.037642   0.040570   0.928  0.35350
## log(totalcommamt + 0.1)    -0.082038   0.044736  -1.834  0.06668 .
## proj_duration               0.026915   0.008887   3.029  0.00246 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 11601  on 9883  degrees of freedom
## Residual deviance: 11414  on 9878  degrees of freedom
## AIC: 11426
##
## Number of Fisher Scoring iterations: 4
```

After I check correlation between variables, which is as showed in **Figure 7**, I find GDP and life_span are highly correlated. Also, lending project cost and total commitment amount are highly correlated. So I then add two interaction terms between life_span and GDP as well as lending project cost and total commitment amount. .From the significant level in **Table 6** below, all variables are more statistical signif-

icant than before. And R-square also increase a little bit. Then I decide to use this model to interpret the coefficient of features.

Figure 7: correlation plot

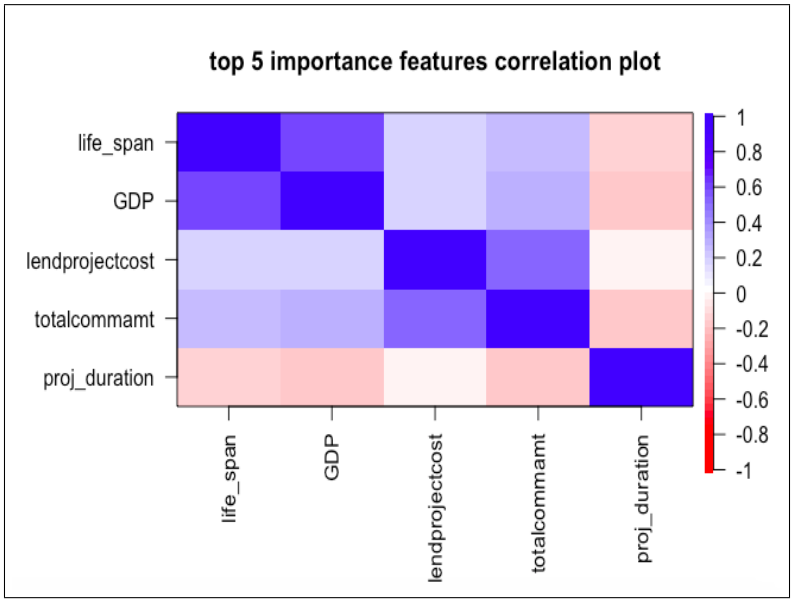


Table 6: model 3 significant level

	z value	Pr(>|z|)
(Intercept)	-1.040	0.29854
life_span	-0.082	0.09344
log(GDP)	4.299	1.72e-05
log(lendprojectcost)	0.590	0.55525
log(totalcommamt)	-0.132	0.00894
proj_duration	2.770	0.00561
log(lendprojectcost + 0.1):log(totalcommamt + 0.1)	-0.373	0.04089
life_span:log(GDP + 0.1)	-0.007162	0.002

From the **Table 7** coefficient table, we could get result that GDP is the most significant feature that could affect project performance. 1% increase in GDP results in

0.73 increase in log-ration of performance satisfaction likelihood. 1% increase in GDP results in 1.08 times higher performance satisfaction likelihood.1 unit(year) increase in project duration results in 1.02 times higher performance satisfaction likelihood.

Table 7: coefficient

(Intercept)	0.03695109
life_span	0.99870102
log(GDP + 0.1)	1.07817265
log(lendprojectcost + 0.1)	1.11779723
log(totalcommamt + 0.1)	0.97615142
proj_duration	1.02524497
life_span:log(GDP + 0.1)	0.99286315
log(lendprojectcost + 0.1):log(totalcommamt + 0.1)	0.99613901

6 Conclusion

As we described above, Random Forests is the best model for my feature set, compared with Adaboost, Logistic Regression and Decision Tree.It helps us to predict world bank project performance and allows us to increase our confidence on evaluating which projects are likely to be bad performance projects.By looking at the feature importance, we can see that the main features appear to be countries characteristics such as GDP and life expectancy, and also the amount involved in the projects, total commitment amount and project duration, which are all project level indicators, are sub-importance.

GDP is the most significant feature. Usually, 1% increase in GDP results in 0.73 increase in log-ration of performance satisfaction likelihood. 1% increase in GDP

results in 1.08 times higher performance satisfaction likelihood. Other project level measures are all positively affect project performance.1 unit(year) increase in project duration results in 1.02 times higher performance satisfaction likelihood.

One of the limitation that makes our model performed not that good(AUC-ROC around 0.7), mostly because of missing information.After join country and government data, the number of projects data reduced about because of that. In future work, I expect to create new features from this dataset that allows us to count with more data (for example, using the changes of GDP and life span of countries during project instead of the mean value of the period).

Another limitation is that the country level features are not enough. In next stage, I will consider to join more country level indicator data, such as inflation rate, employment rate and crime rate etc..

7 Reference

World Bank what we do[Online]. Available: http://www:worldbank:org/en/about/what-we-do.

Transparency International.Corruption perceptions index[Online]. Avail- able: http://www:transparency:org/cpi2010/results,2010.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417441, and 498520.

Hotelling, H. (1936). *Relations between two sets of variates.* Biometrika, 28, 321377

Cevdet Denizera, Daniel Kaufmannb, c, Aart Kraayd, *Good countries or good projects? Macro and micro correlates of World Bank project performance*Journal of Development Economics Volume 105, November 2013, Pages 288302

Jonathan Isham, Deepa Narayan and Lant Pritchett, *Does Participation Improve Performance? Establishing Causality with Subjective Data*, The World Bank Economic Review Vol. 9, No. 2 (May, 1995), pp. 175-200

Jonathan Isham, Daniel Kaufmann, Lant H. Pritchett, *Civil Liberties, Democracy, and the Performance of Government Projects*, World Bank Econ Rev (1997) 11 (2): 219-242.

Walid Belassi, *A new framework for determining critical success/failure factors in projects*, International Journal of Project Management Volume 14, Issue 3, June 1996, Pages 141-151

E. Westerveld, *The Project Excellence Model: linking success criteria and critical*

success factors, International Journal of Project Management, Volume 21, Issue 6, August 2003, Pages 411418

Pinto, J. K. Slevin, D. P. (1988). *Critical success factors across the project life cycle.* Project Management Journal, 19(3), 6775.

APPENDIX