Scorecard Model Acquisition Template

Indra Lukas Tjahaja

This template created for AWDA – Maucash Training

Contents

[1. Executive Summary 4](#_Toc85094349)

[2. Introduction 5](#_Toc85094350)

[3. Model Objectives 6](#_Toc85094351)

[3.1 Model History 6](#_Toc85094352)

[4. Data 7](#_Toc85094353)

[4.1 Data Source and Collection 7](#_Toc85094354)

[4.2 Exclusion 7](#_Toc85094355)

[4.3 Limitation of Data 7](#_Toc85094356)

[5. Model Development Methodology 8](#_Toc85094357)

[5.1 Development Overview 8](#_Toc85094358)

[5.2 Univariate Analysis 8](#_Toc85094359)

[5.2.1 Calculation Summary 8](#_Toc85094360)

[5.2.2 Analyzing Univariate Outputs 8](#_Toc85094361)

[5.3 Variable Selection 8](#_Toc85094362)

[5.4 Final Model Development 8](#_Toc85094363)

[5.4.1 Regression Analysis 8](#_Toc85094364)

[5.4.2 Final Model 8](#_Toc85094365)

[5.4.3 Ginia Analysis 8](#_Toc85094366)

[6. Validation Analysis 9](#_Toc85094367)

[6.1 Actual vs Predicted Reject Rate 9](#_Toc85094368)

[6.2 Stability Index 9](#_Toc85094369)

[6.3 Gini 9](#_Toc85094370)

[6.3.1 Gini Analysis 9](#_Toc85094371)

[7. Model Benefits Analysis 10](#_Toc85094372)

[8. Conclusion 11](#_Toc85094373)

[References 12](#_Toc85094374)

[Footnotes 13](#_Toc85094375)

[Tables 14](#_Toc85094376)

[Figures title: 15](#_Toc85094377)

# 1. Executive Summary

[Do executive summary after you complete all other parts]

[Talk about purpose of the scorecard]

[Put a simple table containing important information of model data and target, example as follow]

|  |  |
| --- | --- |
| MODEL SUMMARY | |
| Observation/Sample point | Point of sample is 3 months, from Jan-Mar’19 |
| Performance Window | April’19-April’20 |
| Observation Target |  |
| Data Sources | * Internal Application Data * SLIK/OJK data * Vendor Credit Score data |
| Data Exclusion | MOB <= 3 months |

[Talk about methodology being used]

[Talk about variables being used and why they are removed]

|  |  |  |
| --- | --- | --- |
| Reasons for Removal | Difference | Total Variables Remaining |
| Total |  | 200 |
| Low IV score | -50 | 150 |
| Multicollinearity | -75 | 75 |
| Logistic | -50 | 25 |
| Final |  | 25 |

[Talk about the result of development model]

# 2. Introduction

In the introduction, give explanation to the following questions:

* Why do you create this document?
* The business problem that require solution from model development

# 3. Model Objectives

The objective of developing the model must at least answer the following questions.

|  |  |
| --- | --- |
| Questions | Possible Answers |
| Type of Scorecard to build | Fraud, Credit (A-score, B-score, C-score), Reserve calculation, X-Sell/Up-Sell |
| Is the scorecard for specific product | Secured or Unsecured product (detail), new product targeting specific market, etc |
| What the company implement before to approve/reject? | Policy rules, expert rules, scorecard, combination, etc |
| Is the new scorecard will be used as sole decider? | Yes/No |
| What is the target variable that you are trying to predict? | Likelihood to default.  What is definition of default?  DPD30,60,90, ever DPD, late, ask discount, etc  How long the account on book to decide the final status?  To arrive to this decision, the document need to provide analysis and reasoning |
| If there is a scorecard before. Why do you need to build another scorecard? What is the new objective? | Reduce default rate for the same approval rate  Increase approval rate for the same default rate  Increase revenue (possibility of targeted pricing, etc)  Choose active customer (if revolving credit)  Or any other goals? |
| Is the development require support from external developer? | External vendor to create the model based on supplemented data |
| Is the model will require external data? | Yes/No, if yes explain. Example SLIK, vendor API, etc |
| Is fraud excluded from the scorecard? | Yes/No. How about suspected fraud cases? |

## 3.1 Model History

Answer the following questions if company previously implement scorecard

|  |  |
| --- | --- |
| Questions | Possible Answers |
| If there is scorecard before. Check for documentation | The documentation will allow scorecard developer to check all variables being used and performance period |
| Is the scorecard still being used? | Yes/no |
| What is the performance of previous scorecard | Answer the model performance currently, and previously |
|  |  |

If possible, for every changes in scorecard. It is better to score all applicants using all version of scorecard model. So we know the performance of all versions of model.

In this section, also explain on the implementation use of the previous scorecard

|  |  |
| --- | --- |
| Questions | Possible Answers |
| Explain in detail how the scorecard use in the overall approve/reject | Is scorecard used as sole decider, or combined with other methods? |
| What happened when we reject? | Do we ask for additional information? Or the reject given other products with higher pricing? |
| Are there any preferential treatment to those with high credit score? |  |

# 4. Data

Based on model objective mentioned in previous part. Explain in detail on the data being used to create the model:

## 4.1 Data Source

Source of data. Explain here on the source of the data. Explain in detail on how the data gathered. If the data gathered from internal DWH, then explain the source table

### 4.1.1 Sample and Performance Period

Discuss the time period when the data gathered

### 4.1.2 Internal data

Discuss about the use of internal data, and how/where to get the data

### 4.1.3 External data

Discuss about the use of external data (if any), and how/where to get the data

### 4.1.4 Data Quality

Discuss about the data and its related data quality. Preferably from 5 dimensions:

1. Completeness
2. Validity
3. Consistency
4. Uniqueness
5. Accuracy

### 4.1.5 Data Lineage and Feature Engineering

Discuss about the data and its related variables derived from the original variable.

It also need to clearly define on how to get the variable, the definition of the variable, if the data from source or derived. If it’s derived, what’s the logic to get the variable

### 4.1.6 Oversampling/Undersampling (if any)

If the data imbalance, describe the methodology used and reasoning

## 4.2 Exclusion

Are there any exclusion the data? Maybe inappropriate target customer? Exclude specific type of customers? Any exclusion need to be provided reasoning.

For example, customer indicated/proved to be fraudulent to be excluded from credit model.

If some customers type excluded because the business decided to target different type of customers in the future, then explain as well.

## 4.3 Limitation of Data

State the data limitation and the risk mitigation.

For example, the original data being used in development contains missing data. But for production, assume there is no missing data. Because we develop risk mitigation of any missing data will be handled manually (for example)

# 5. Model Development Methodology

Discuss parties involved in the discussion to create the scorecard. Please note that these departments do not necessarily know which variable finally used as part of modeling

|  |  |
| --- | --- |
| Departments | Involvement |
| Scorecard developer | The person creating the scorecard and writing down this document |
| Data Analyst / Data Scientist / Data Engineer | Collecting data, ETL, and preparing data for scorecard development |
| Product or Portfolio Risk manager / Credit Scoring Manager | Scorecard user from Risk Management side. These people also responsible for risk strategy and related policy.  Thus these people also contribute and decide what scorecard will be created  They also will be the one to decide how to use the scorecard as part of loan approval framework |
| Product Manager | Scorecard user from Business side. They responsible to create the target market of the product, and need to know the ideal customer for the product so they can prepare the marketing strategy as well |
| Operational Manager | Operational side. Need to ensure the process of application and scorecard implementation will be smooth |
| Validation team | The team/person to validate the result of the model. Need to be separate person than the developer to avoid bias / conflict of interest |
| IT Team (IT manager / project manager) | To ensure the system able to support the scorecard for production |
| Legal / Compliance / Enterprise Risk | To ensure compliance to regulation and avoid any other risk |

## 5.1 Development Overview

Provide explanation on the methodology being used to create the model

## 5.2 Univariate Analysis

Content -> written explanation from the Python result

### 5.2.1 Calculation Summary

Content -> written explanation from the Python result

### 5.2.2 Analyzing Univariate Outputs

Content -> written explanation from the Python result

## 5.3 Variable Selection

Content -> written explanation from the Python result

## 5.4 Segmentation (if any)

Describe here if data/team decide to create several segmentation with its own scorecard

## 5.5 Final Model Development

Content -> written explanation from the Python result

### 5.5.1 Regression Analysis

Content -> written explanation from the Python result

### 5.5.2 Final Model

Content -> written explanation from the Python result

### 5.5.3 Gini Analysis

Content -> written explanation from the Python result

# 6. Validation Analysis

Content -> written explanation from the Python result

## 6.1 Actual vs Predicted Reject Rate

Content -> written explanation from the Python result

## 6.2 Stability Index

Content -> written explanation from the Python result

## 6.3 Gini

Content -> written explanation from the Python result

## 6.4 Validation Review

Content -> written explanation from the Python result

### 6.4.1 Independent Validator

The validator is expected to be an independent person from the scorecard developer

### 6.4.2 Legal/Compliance/Risk

Get review from 2nd line of defense on the model. Note that only selected people (preferably higher ups) that know the content of the scorecard. If not possible, just ask them all the data used in the model, and ask them to exclude not appropriate data.

# 7. Model Benefits Analysis

Explain the result of model developed and give summary of benefit of using this new model. Example, the model able to give lower default rate for the same approval rate, etc.

# 8. Conclusion

This document does not specifically describe the implementation process. For a full detail of implementation process can refer to other document

References

Last Name, F. M. (Year). Article Title. *Journal Title*, Pages From - To.

Last Name, F. M. (Year). *Book Title.* City Name: Publisher Name.