

---

---

# Forever Gone

A classification exercise to identify books  
which are forever lost within NTHU Library.

## Group 5

James

Cynthia

Edward

Huynh Nhat To

Olive

---

---

# Business Problem

Someone comes to the library to borrow a book but its staff can't find it. Sometimes, this book is forever gone, most times, it gets found. How do you tell which is which?

# Opportunity

If the staff could instantaneously identify an item that is forever gone, then the library could initiate procurement for a replacement.

---

---

# Data Mining Problem

Supervised Task: We're trying to predict whether or not a book is forever lost, information we have.

Predictive: This is a classification exercise as items are either lost or found.

Deployment: Whenever library staff are unable to find an item, they can run this item through our model to get a likelihood of the item being missing.

---

# DATA DESCRIPTION

What we got from the library

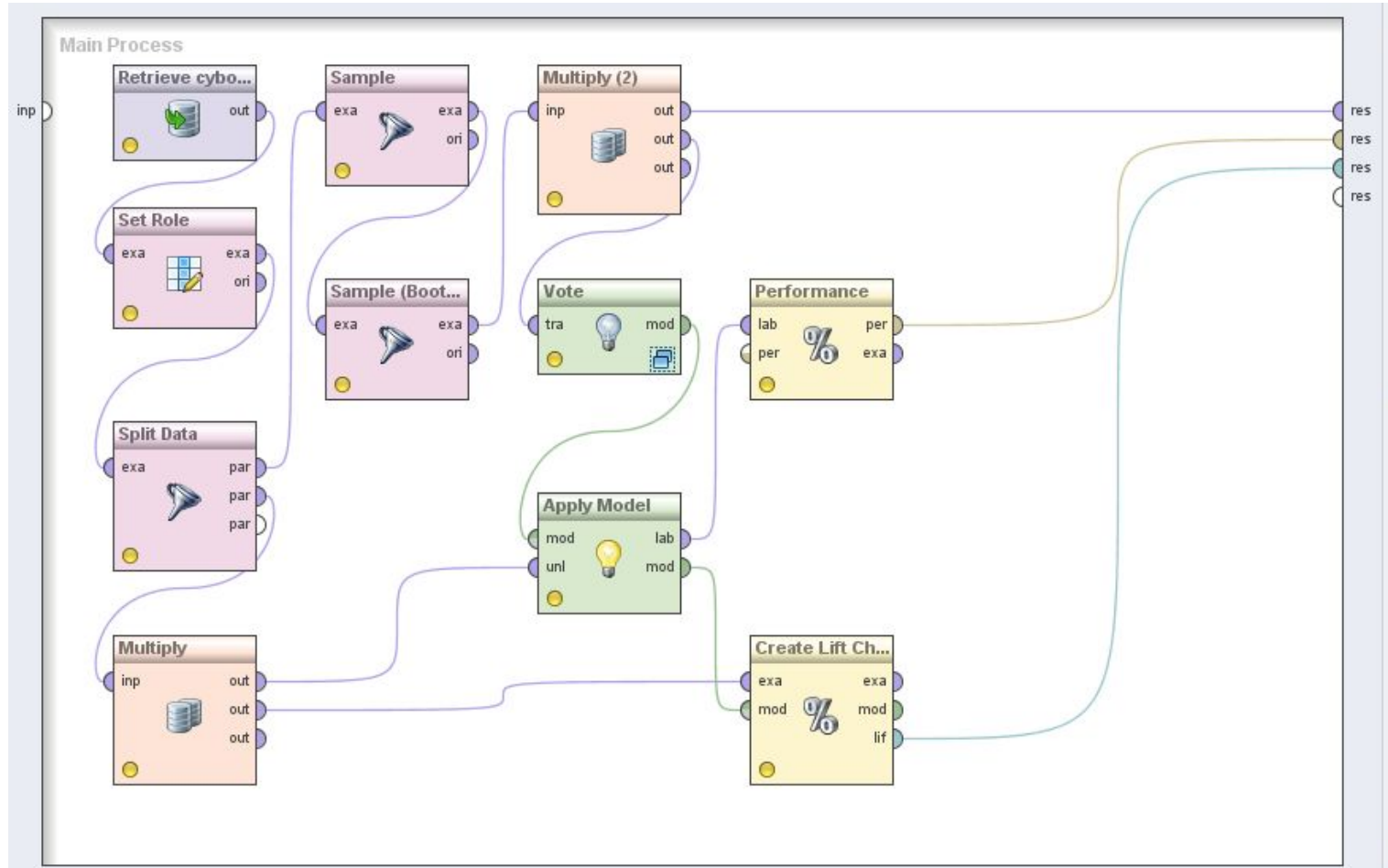
1. 33,070 records
2. Transactions (Borrow/Renew)

And what we made of it

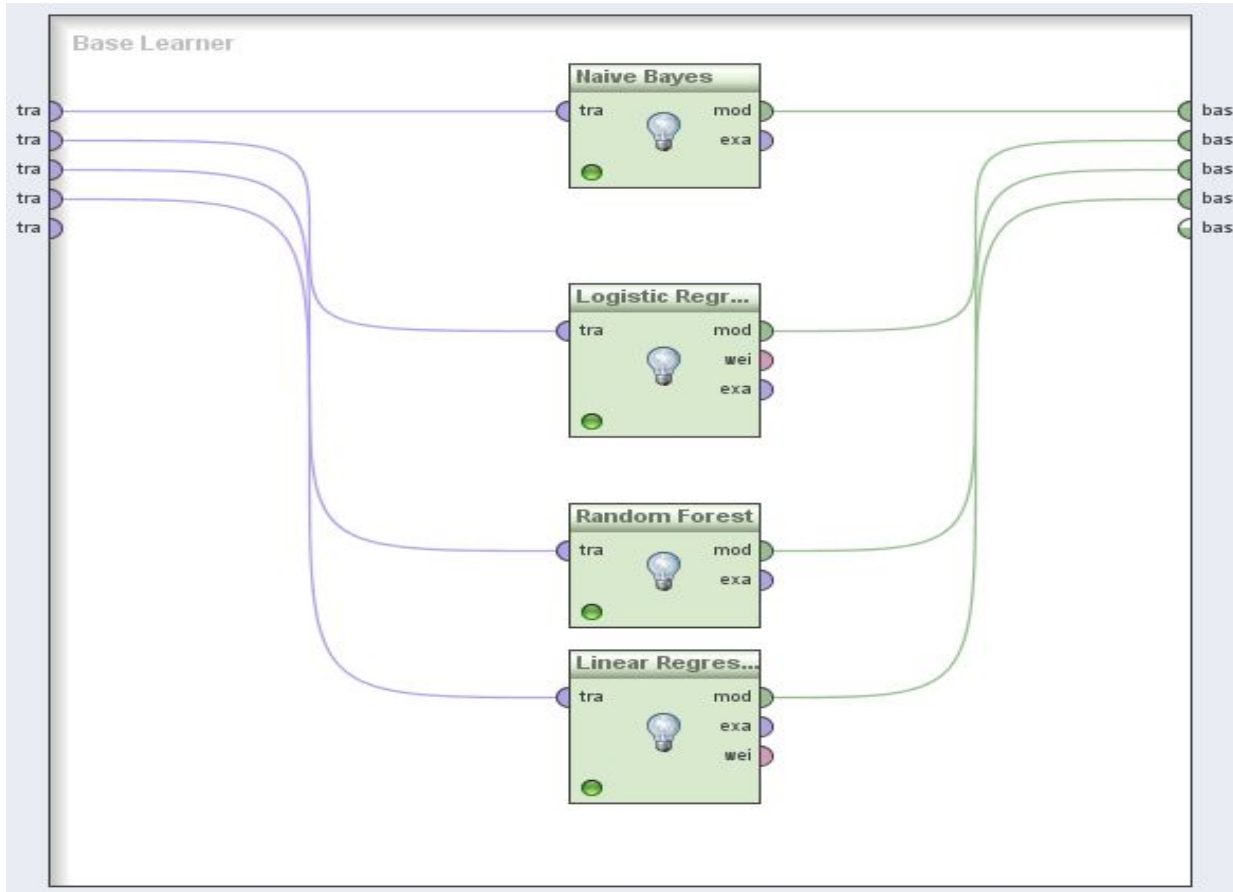
1. 787 records (with number of transactions recorded)
2. Items (Book, DVD, Library Card)
3. Transformed selected variables into dummy variables
  - a. Item Type (3); Location (7); Language (3); Merged Classification (9); **Lost/Found**.

Column	Description
id	Readers system number
reader identity	Type of reader
department	Department of reader
item type	Type of item
transaction date	Date of transaction
transaction type	Type of transaction
bar code	Barcode No
current item status	The current processing status of the item
location	Library Branch Code
updating date of current item status	The item processing status update
date of last returning	Last return date
call number	Call Number
rule of call number type	Language of Item
history status of item	Problem with item in the past
updating date of history item status	History of the item processing status update
item title	Title
item author	Author
item ISBN_ISSN	ISBN
call number type	item type based on Chinese System or Library of Congress

# Process Map



# Classification Techniques

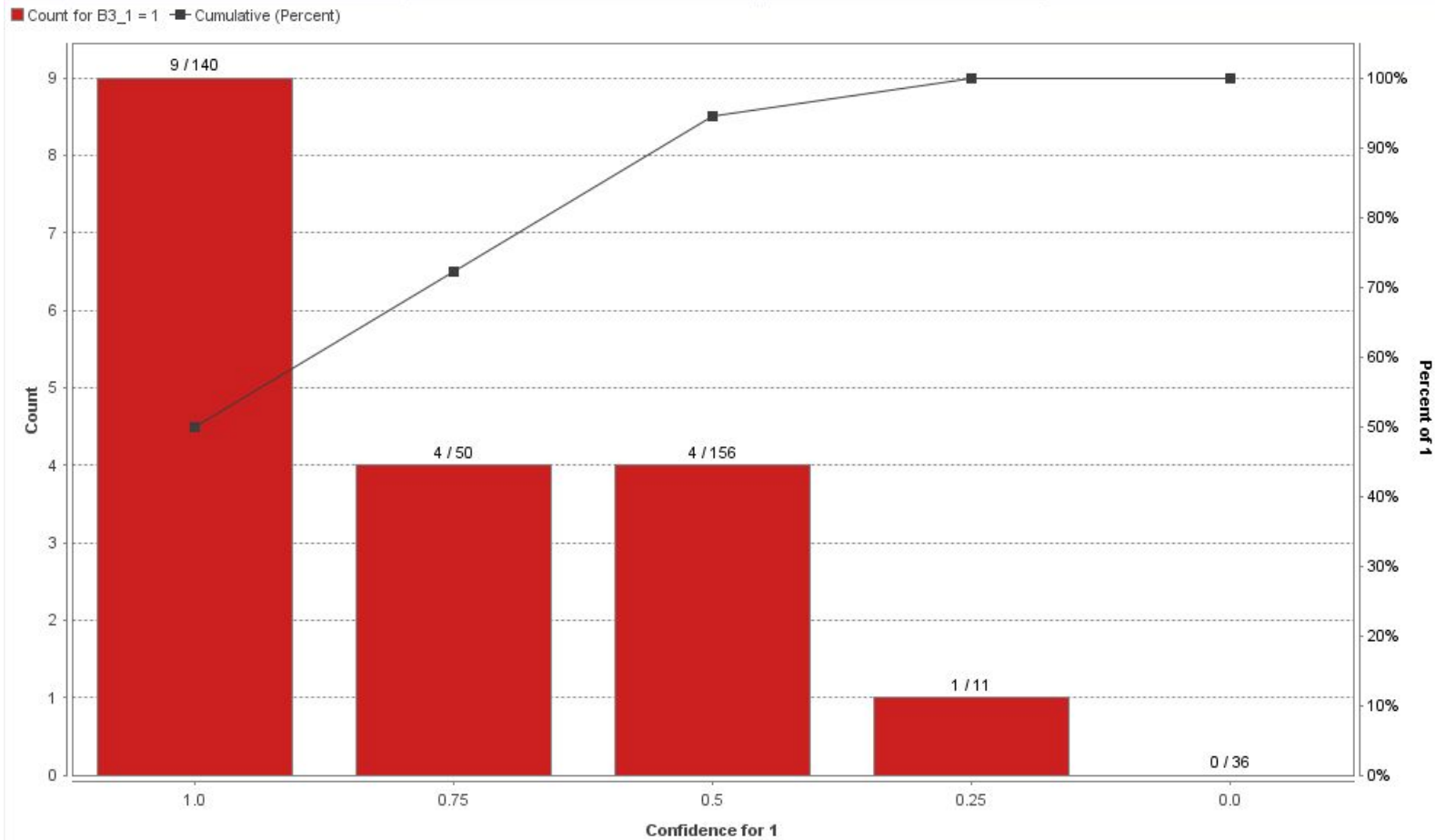


# Evaluation Methods I

**classification\_error: 46.31%**

	true 0	true 1	class precision
pred. 0	198	5	97.54%
pred. 1	177	13	6.84%
class recall	52.80%	72.22%	

# Evaluation Methods II





# Conclusion

1. Given the relatively few number of missing books, our model can not be guaranteed to reliably inform the library staff.
2. Compared to all books in library, the amount of missing books seems relatively very little. The information we want to have from missing books are not enough. Which means, they don't really either put effort on this problem or try solving it.