PREDICTING FRAUDULENT FINANCIAL STATEMENTS USING SEMANTICS

PRESENTED BY

B.Archana(10z305)
R.K.Manisha(10z322)
S.Sasikala(10z340)
V.Subha Prabha(10z349)
A.Lalitha(11z462)

GUIDED BY

Mrs.C.Kavitha
Assistant Professor(Sr.Gr.)

MOTIVATION

- To improve foreign/corporate investments by identifying trust factor of companies from investor's perspective.

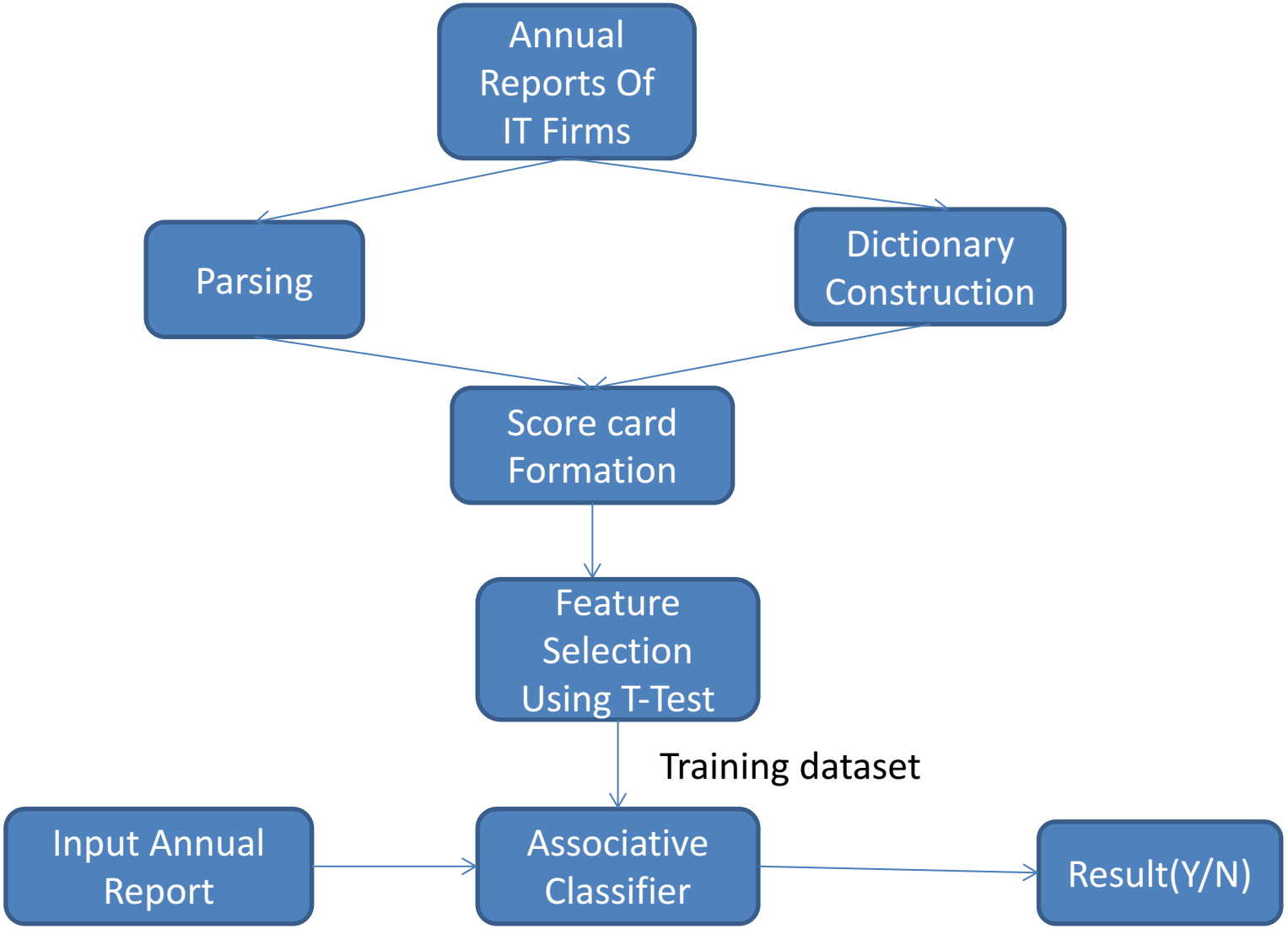
OBJECTIVE

- Finding fraudulent data provided by companies
- Helping customers in identifying the status of the company based on its annual reports

NEED FOR THE PROJECT

- Effective business analytical tools
- Extract insight of corporate data
- Deliver higher level of efficiency and profit

Proposed methodology



Implementation

Collection of annual reports of firms:

- Available in PDF format.

Identification of keywords:

- The keywords required for Dictionary are identified

Contd.,

GOVERNANCE PARAMETERS:

- Board matters
- Nomination matters
- Remuneration matters
- Audit matters
- Communication

Contd.,

Score card formation:

- Annual Report is parsed using PDF Parser.
- Score is assigned based on dictionary.

T-Test Evaluation:

- T-value is calculated as follows

$$t = (m1 - m2)/SD$$

where m1,m2=mean of sample1 and sample2 respectively

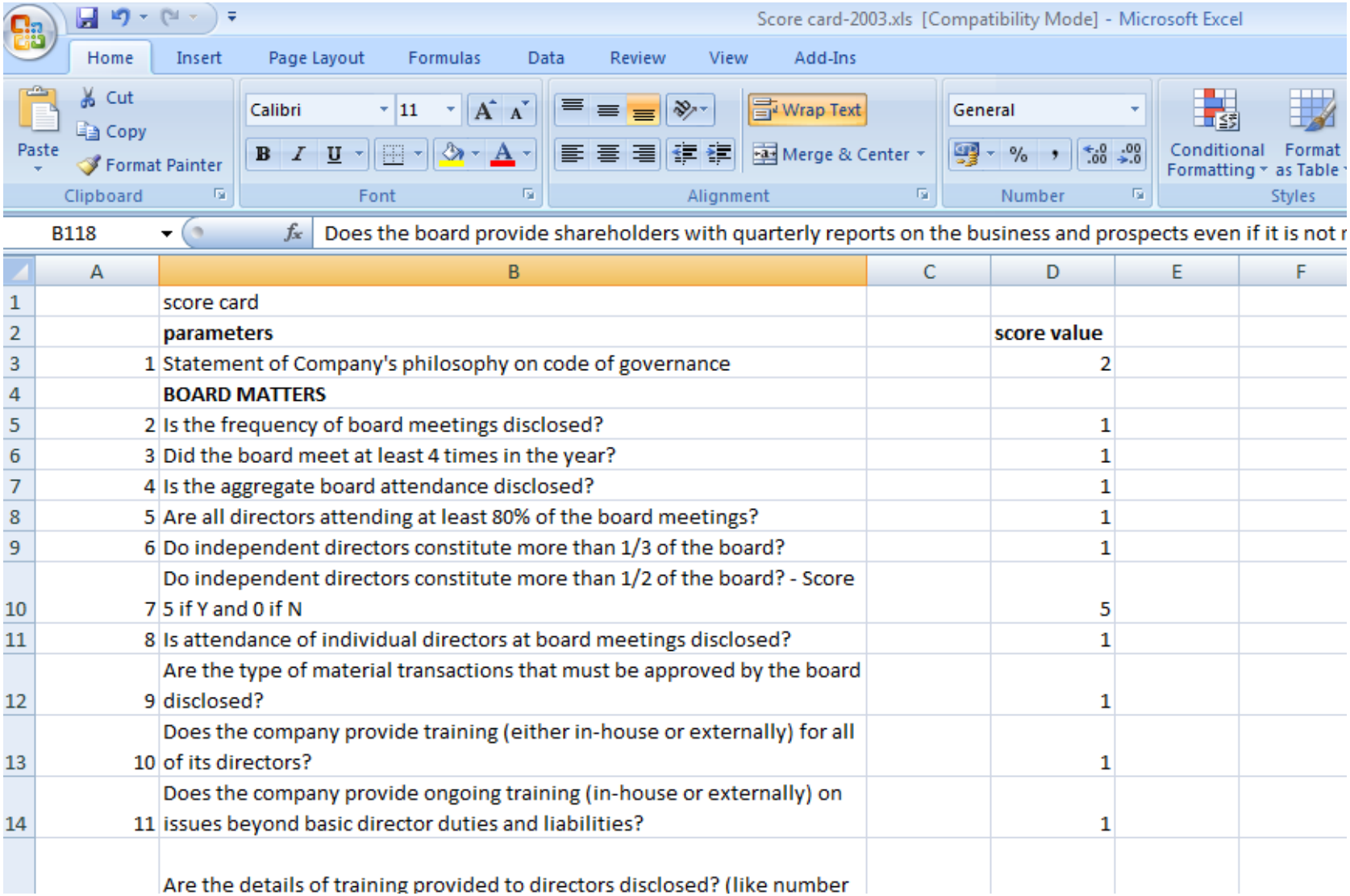
SD=Standard Deviation

- Feature selection is done.

Contd.,

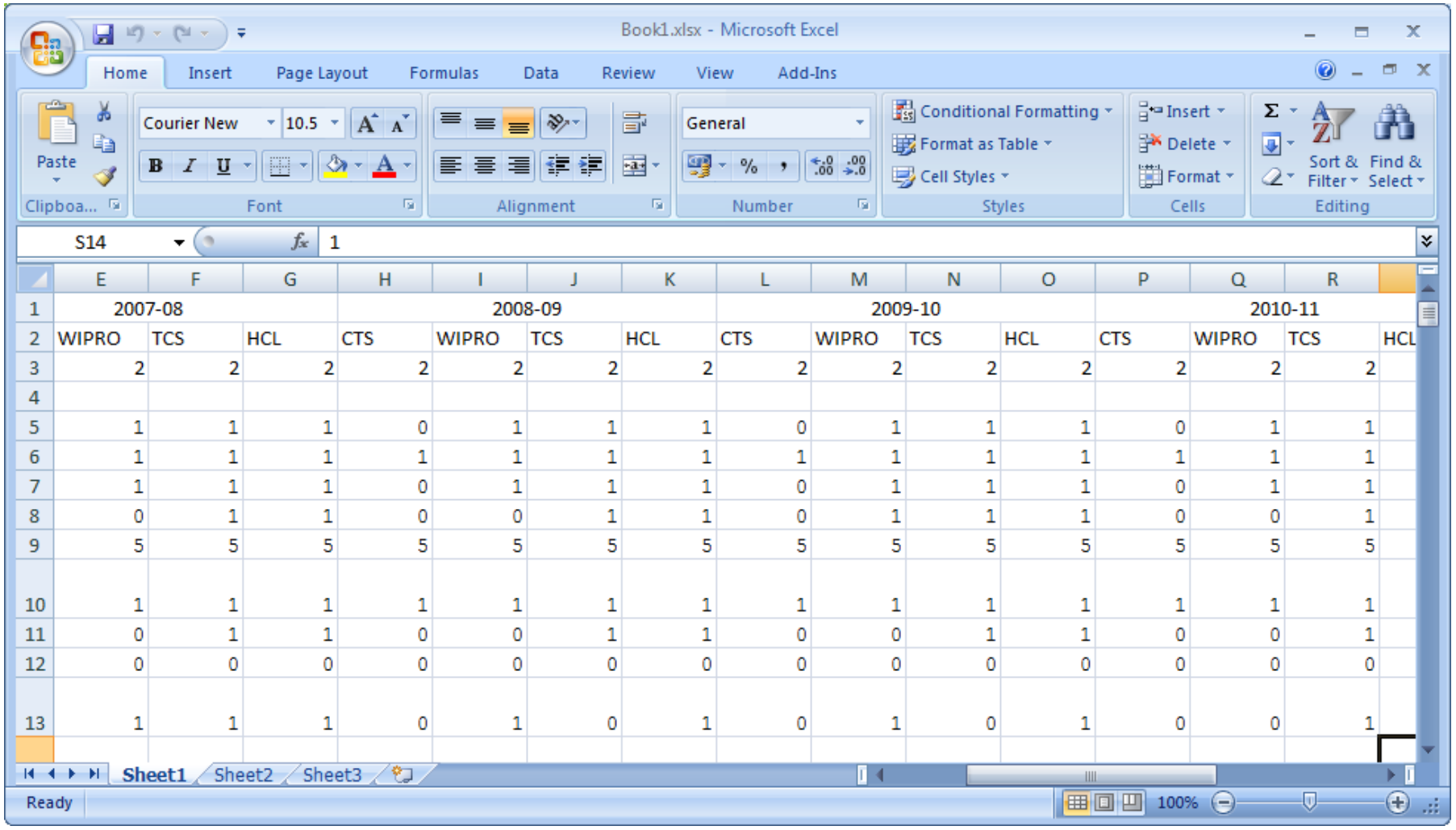
- Associative Classification :
  - Association Rule Mining + Classification.
- Associative classification involves 2 stages
  - Generation of Class Association Rules (CAR) from training dataset.
  - Classification of test dataset into predefined class labels.

Screen shots



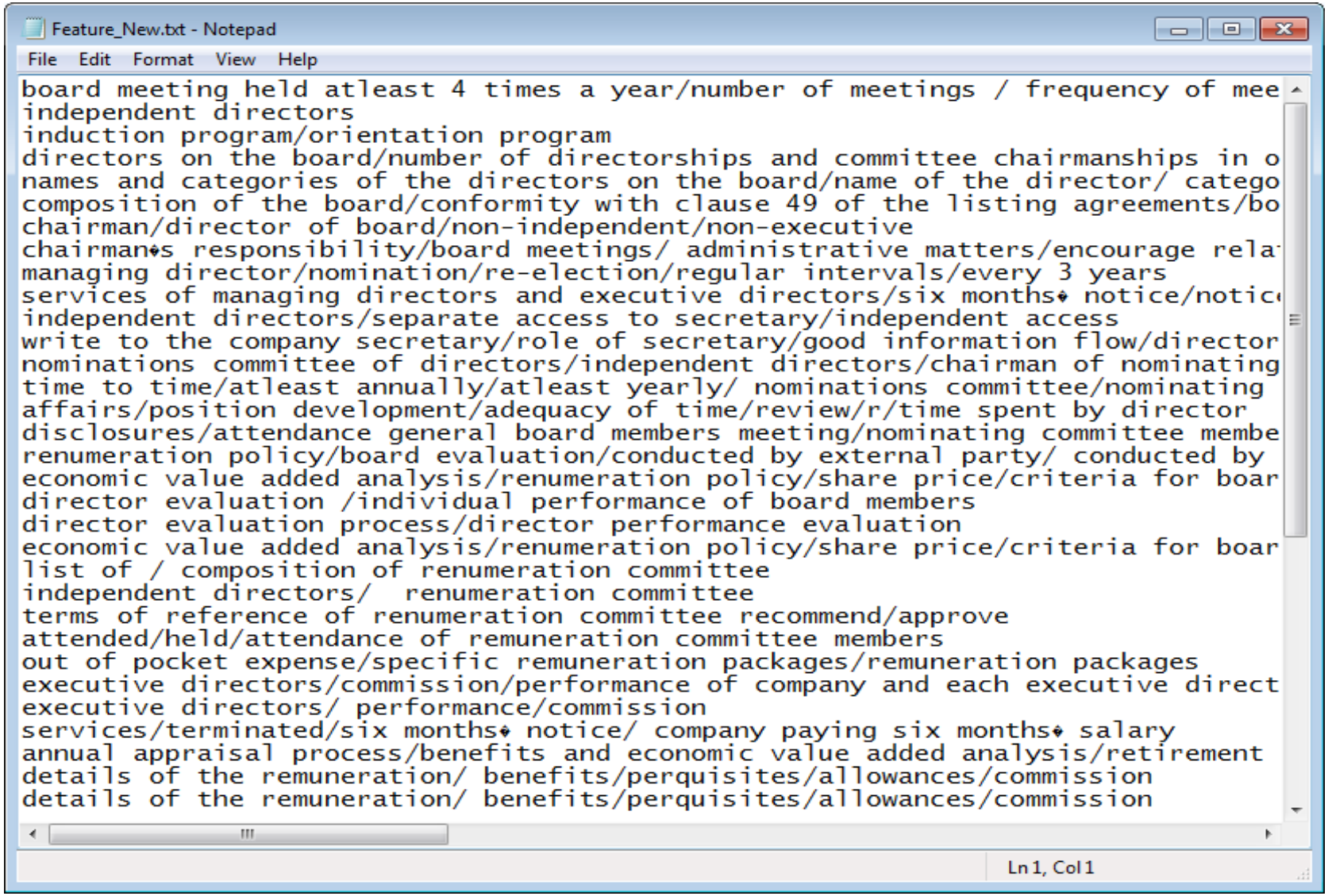
	A	B	C	D	E	F
1		score card				
2		**parameters**		**score value**		
3	1	Statement of Company's philosophy on code of governance		2		
4		**BOARD MATTERS**				
5	2	Is the frequency of board meetings disclosed?		1		
6	3	Did the board meet at least 4 times in the year?		1		
7	4	Is the aggregate board attendance disclosed?		1		
8	5	Are all directors attending at least 80% of the board meetings?		1		
9	6	Do independent directors constitute more than 1/3 of the board?		1		
10	7	Do independent directors constitute more than 1/2 of the board? - Score 5 if Y and 0 if N		5		
11	8	Is attendance of individual directors at board meetings disclosed?		1		
12	9	Are the type of material transactions that must be approved by the board disclosed?		1		
13	10	Does the company provide training (either in-house or externally) for all of its directors?		1		
14	11	Does the company provide ongoing training (in-house or externally) on issues beyond basic director duties and liabilities?		1		
		Are the details of training provided to directors disclosed? (like number				

Score card



	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	2007-08			2008-09				2009-10				2010-11		
2	WIPRO	TCS	HCL	CTS	WIPRO	TCS	HCL	CTS	WIPRO	TCS	HCL	CTS	WIPRO	TCS
3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4														
5	1	1	1	0	1	1	1	0	1	1	1	0	1	1
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	1	1	0	1	1	1	0	1	1	1	0	1	1
8	0	1	1	0	0	1	1	0	1	1	1	0	0	1
9	5	5	5	5	5	5	5	5	5	5	5	5	5	5
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	0	1	1	0	0	1	1	0	0	1	1	0	0	1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	1	1	0	1	0	1	0	1	0	1	0	0	1

Selected Features



classification

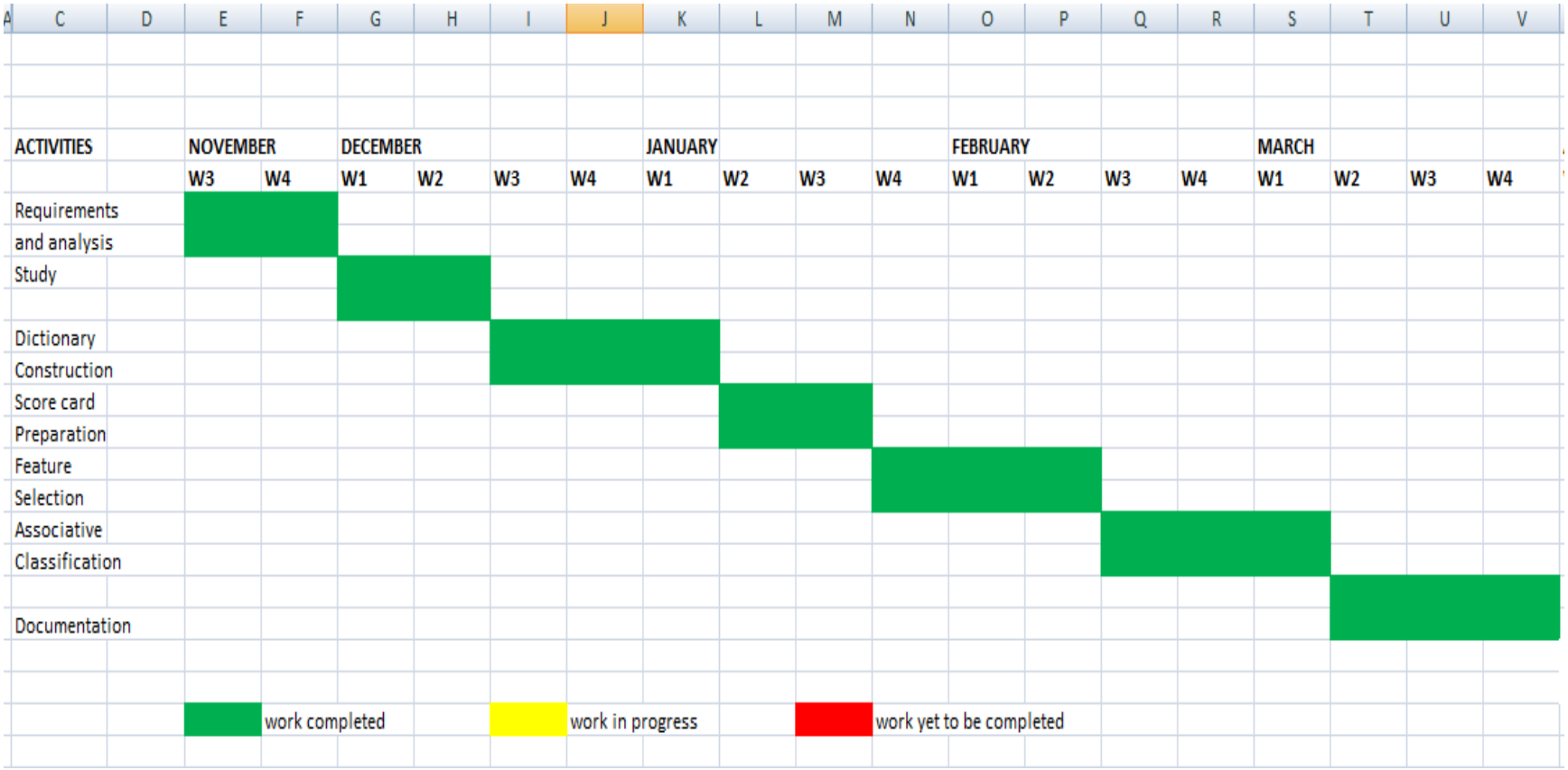
```
Output - MainProject1 (run)

run:
    *********Rules*******
[[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 5, 23]
, [1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1
, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 2, 0, 5, 23
], [1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 2, 0, 5, 2
4], [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 5,
23], [1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0
, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 2, 0, 5,
 24], [1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
0, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 2, 0, 5
, 24]]
 ************Number Of Rules***********
6
  ********Test_data********
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 2, 0, 5]
*****RESULT**********
Given Annual Report is a GOOD One
BUILD SUCCESSFUL (total time: 0 seconds)
```

ACCURACY

$$\text{Accuracy} = \frac{\text{Number of correctly predicted test data}}{\text{Total number of test data}}$$

	Accuracy (in percentage)	Number of Rules
Without Feature Selection	75.0	16
With Feature Selection	87.5	16

Timeline


ACTIVITIES
NOVEMBER
DECEMBER
JANUARY
FEBRUARY
MARCH
W3
W4
W1
W2
W3
W4
W1
W2
W3
W4
W1
W2
W3
W4
W1
W2
W3
W4
Requirements and analysis
Study
Dictionary Construction
Score card Preparation
Feature Selection
Associative Classification
Documentation
work completed
work in progress
work yet to be completed

ADVANTAGES

- Increased profitability for customers
- Easier way of crediting a firm
- Identify fraudulent statements

REFERENCES

- “Predicting fraudulent financial statements with machine learning techniques” by Sotiris Kotsiantis , Euaggelos koumanakos, Dimitris Tzelepis and Vasilis Tampakas,2006.
- “Detecting financial statement fraud” by Johan L. Perols ,2008.
- “Prevention and Detection of Financial Statement fraud” by Rajan Gupta, Nasib Singh Gill, 2007.
- (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 8, 2012


Thank
You