

RELIABILITY ASSESSMENT OF AN INTELLIGENT APPROACH TO CORPORATE SUSTAINABILITY REPORT ANALYSIS

Amir Mohammad Shahi

Dr Biju Issac

Dr Jashua Rajesh Modapothala

PROBLEM DEFINITION

- **Automatic scoring of Corporate Sustainability Reports**
- **Scoring framework by Global Reporting Initiative®**
- **CSR Report shall disclose certain information.**
- **More disclosures = Higher score**
- **Reports have no predefined format nor a fixed outline!
(We do not know where a disclosure is located, if ever mentioned!)**

GRI® FRAMEWORK

- **Company Profile Disclosures**
- **Disclosures of Management Approach to Sustainability**
- **Performance Indicators**
 - Economical (9 indicators)
 - Environmental (30 indicators)
 - Social (40 indicators)
 - Society (8 indicators)
 - Human Rights (9 indicators)
 - Labor (14 indicators)
 - Product Responsibility (9 indicators)

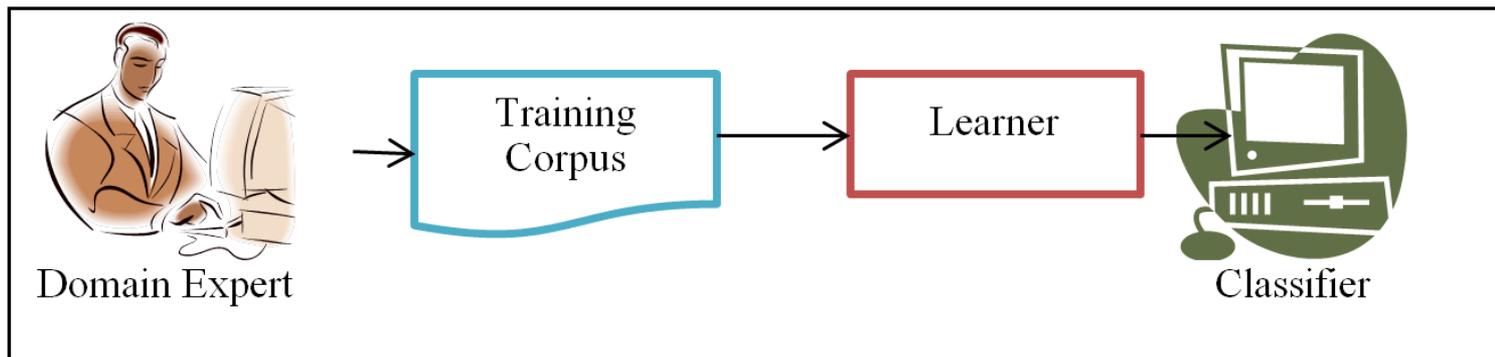
GRI G3.0

APPLICATION LEVELS

| Report Application Level | C | C+ | B | B+ | A | A+ |
|--------------------------|--|---------------------------|--|---------------------------|---|---------------------------|
| Standard Disclosures | Profile Disclosures OUTPUT Report on: 1.1 2.1 - 2.10 3.1 - 3.8, 3.10 - 3.12 4.1 - 4.4, 4.14 - 4.15 | Report Externally Assured | Report on all criteria listed for Level C plus: 1.2 3.9, 3.13 4.5 - 4.13, 4.16 - 4.17 | Report Externally Assured | Same as requirement for Level B | Report Externally Assured |
| | Disclosures on Management Approach OUTPUT Not Required | | Management Approach Disclosures for each Indicator Category | | Management Approach disclosed for each Indicator Category | |
| | Performance Indicators & Sector Supplement Performance Indicators OUTPUT Report fully on a minimum of any 10 Performance Indicators, including at least one from each of: social, economic, and environment.** | | Report fully on a minimum of any 20 Performance Indicators, at least one from each of: economic, environment, human rights, labor, society, product responsibility.*** | | Respond on each core and Sector Supplement* indicator with due regard to the materiality Principle by either: a) reporting on the indicator or b) explaining the reason for its omission. | |

OUR PROPOSED APPROACH

1. We treated each disclosure indicator as a mutually exclusive category.
 - Example: Indicator “EN1” discloses “Total Weight of Materials used”
2. Then we collected sample disclosures of “EN1” from hundreds of CSR reports
3. We trained our machine on how an “EN1” disclosure looks like 😊



MACHINE LEARNING

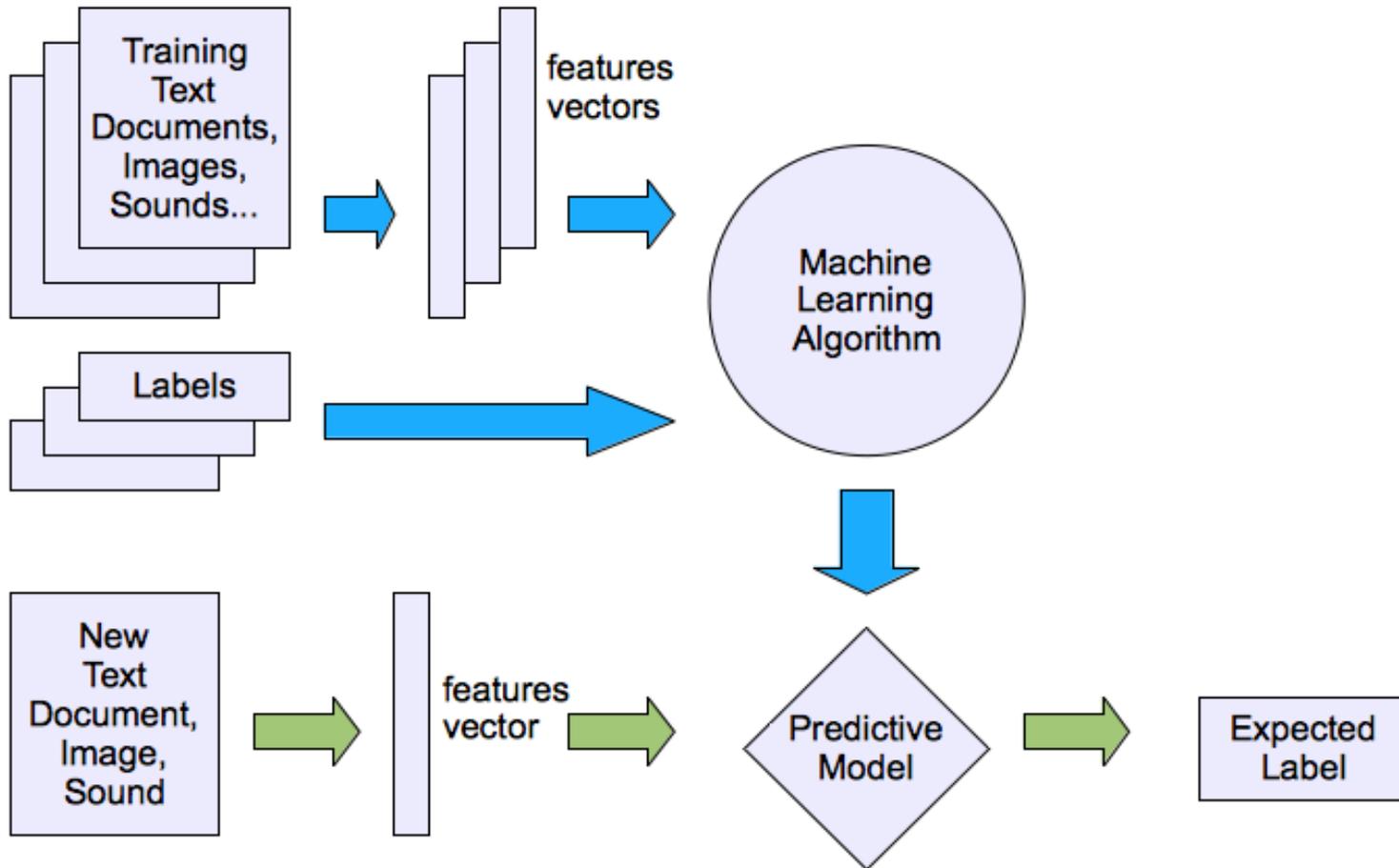
- **The dominant approach to text classification since the 90's**
- **Produces competitive results with Knowledge Based Approach**
- **Classification = Supervised Learning**
- **i.e. Learning inductively by studying the characteristics of pre-classified documents**

TRAINING CORPUS

- 49 Categories
- 1610 sample documents
- ≈ 33 sample documents per class
- ≈ 17000 terms (attributes)
- Sum of all term weights = 10318 (using TFIDF)

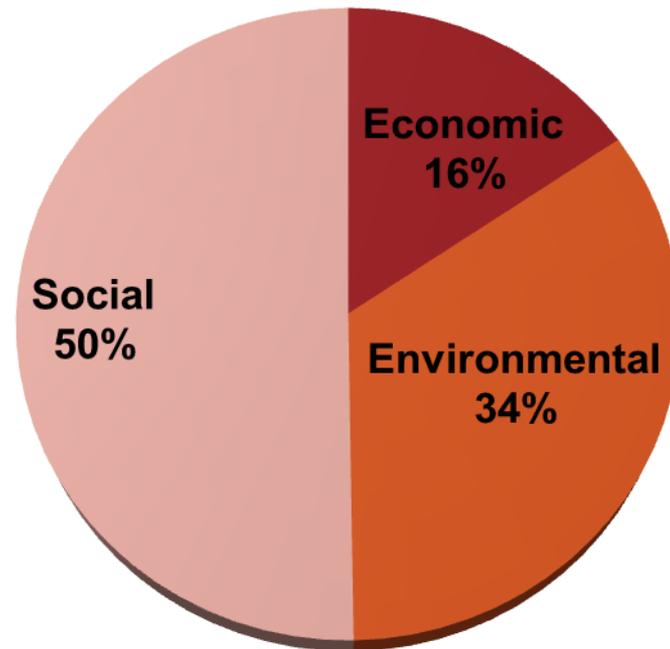
$$\rho(w_k|c_i) = \frac{Fre(w_k) + 1}{|V| + \sum_{s=1}^{|V|} Fre(w_s)}$$

SUPERVISED LEARNING



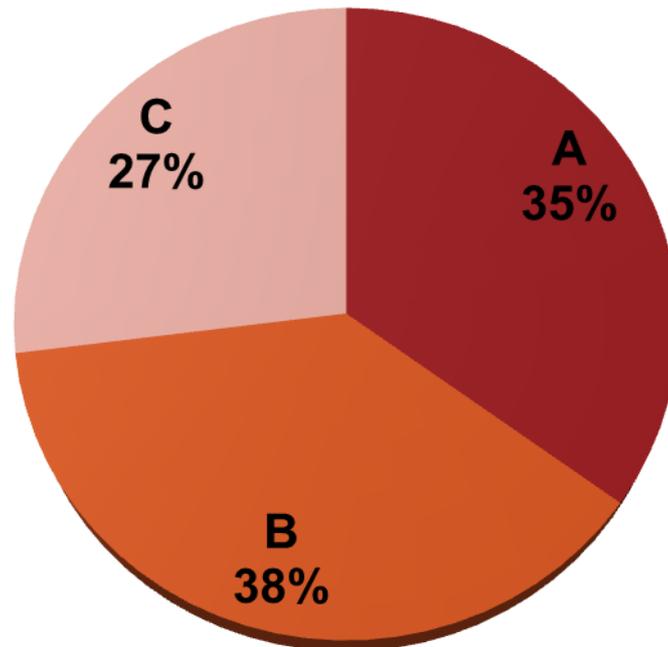
EXPERIMENT SETUP

Training Corpus

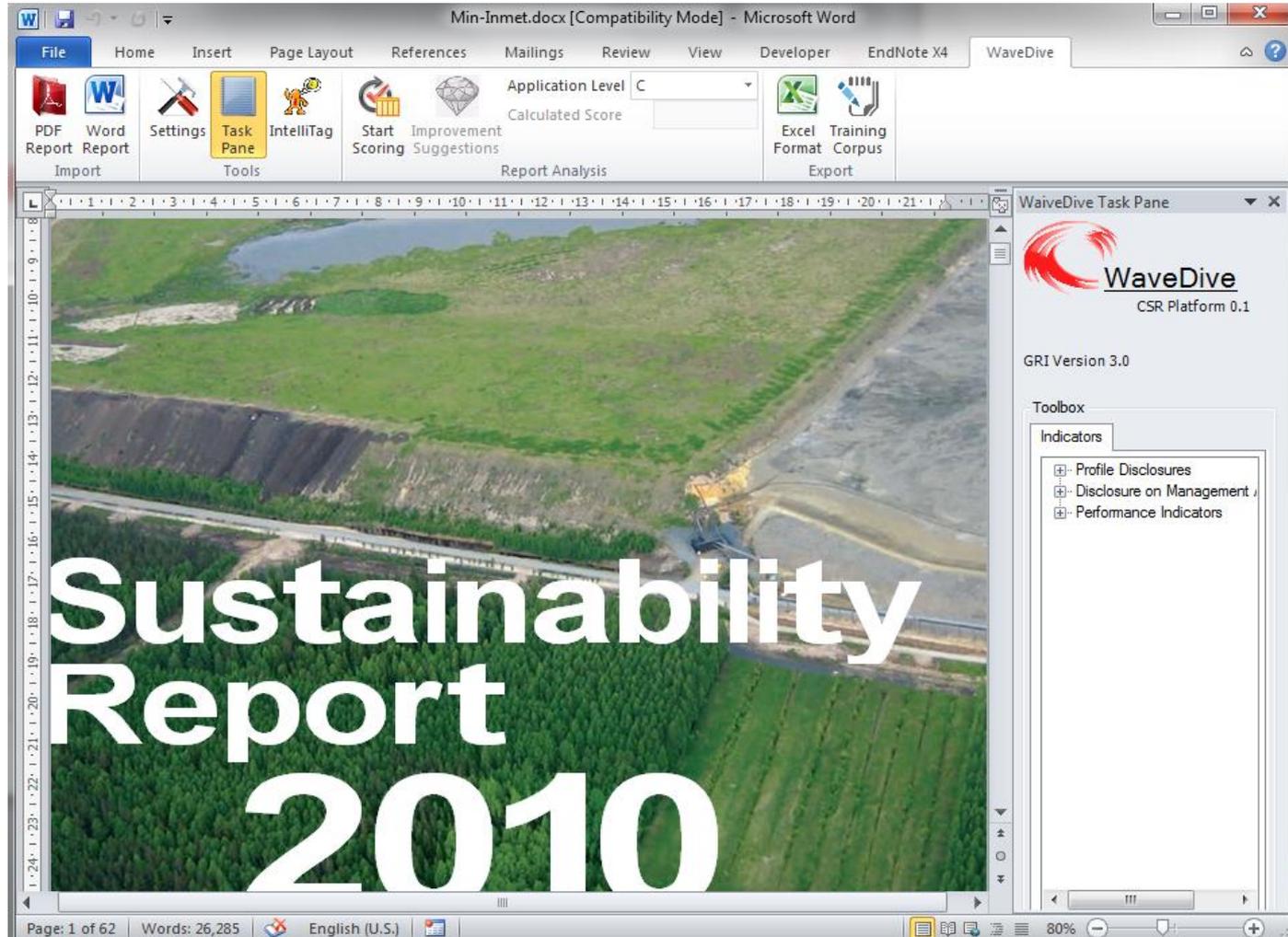


EXPERIMENT SETUP

Tested Reports' Application levels



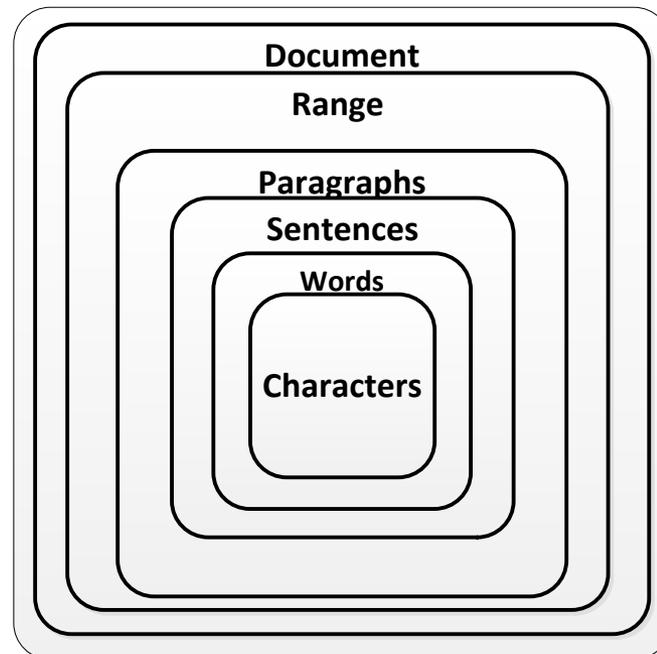
WORD 2010 ADD-IN



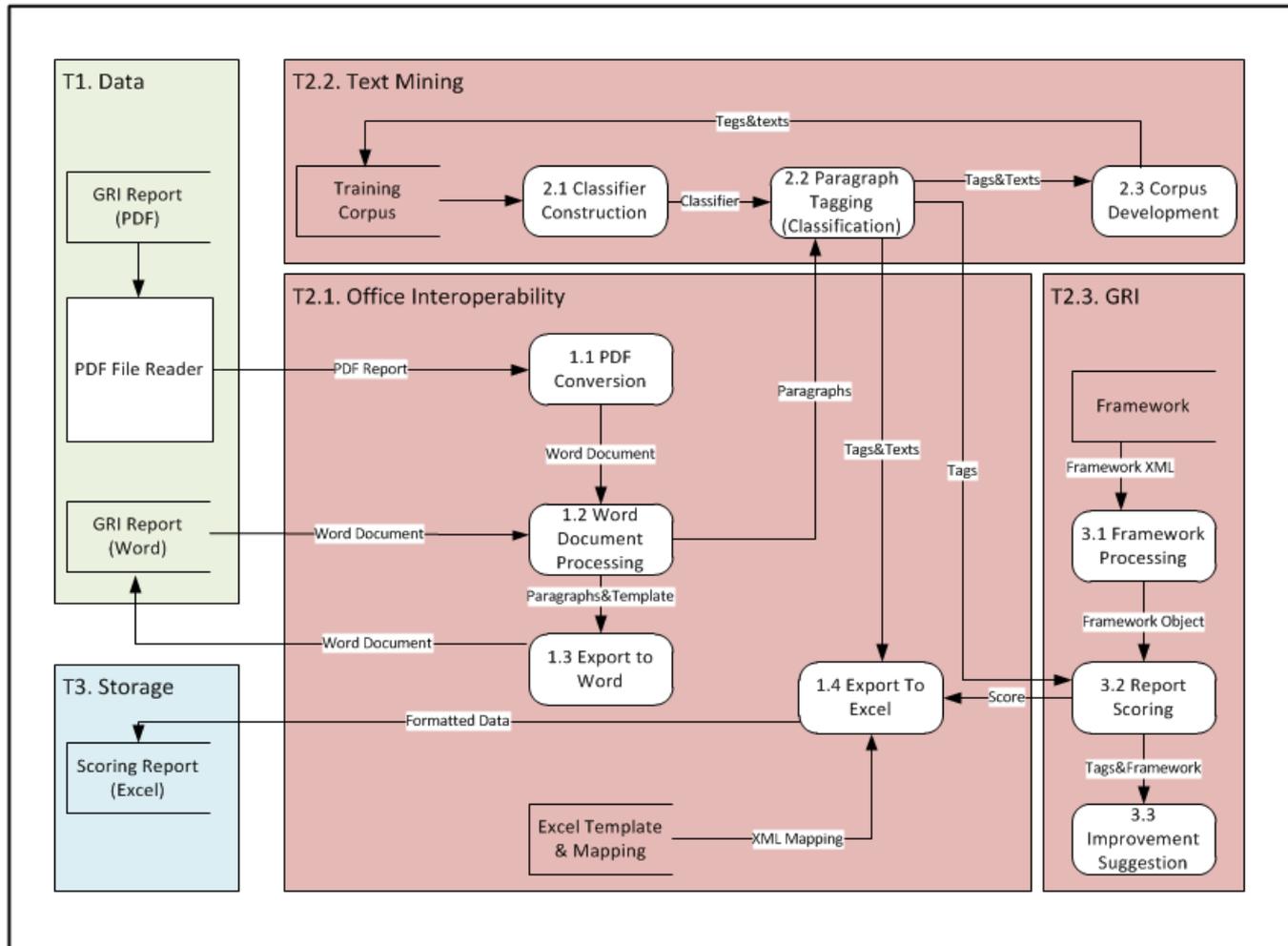
WHY WORD ADD-IN?

Microsoft Word 2010 has a sophisticated Document Object Model (DOM), allowing random access to document elements such as pages, paragraphs, sentences and characters. .

This would be excellent for a text mining solution!

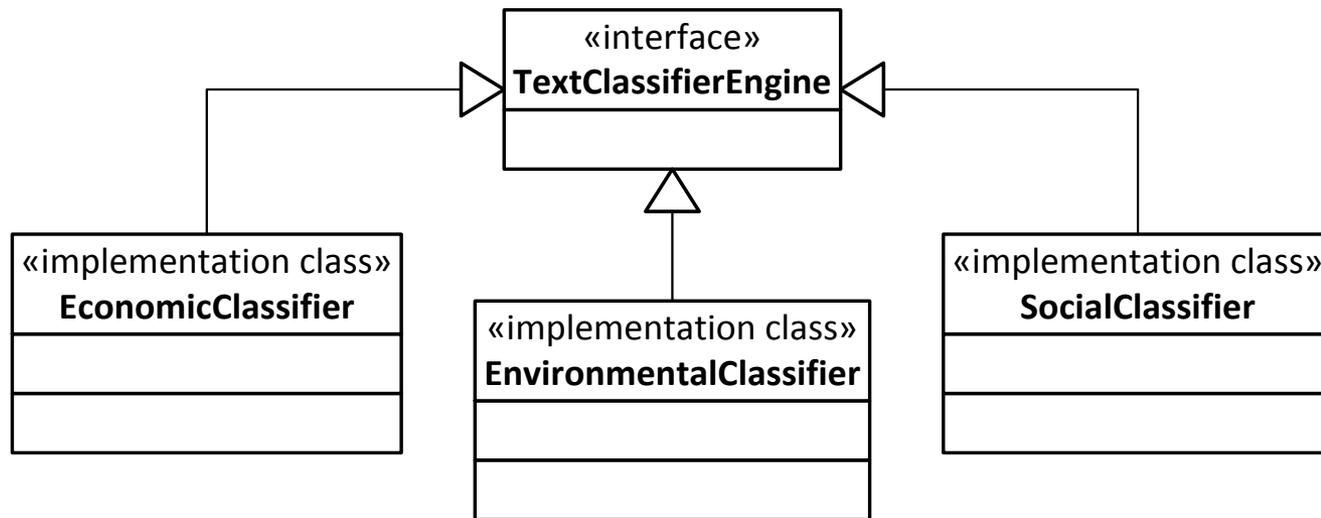


SOFTWARE ARCHITECTURE



INTELLIGENT TAGGING ENGINE

Consists of 3 classifiers (1 for each chapter)



FRAMEWORK XML

```
<?xml version="1.0" encoding="utf-8" ?>
<framework version="g3-0" fullName="GRI Version 3.0">
  <dimensions>
    <Dimension number="1" name="profileDisclosure" desc="Profile Disclosures">
      <sections>
        <Section number="1" name="strategyAnalysis" desc="Strategy and Analysis">
          <indicators>
            <Indicator code="1.1" desc="Statement from the most senior decision-maker of the organization" core="true" excelRowNumber="6"></Indicator>
            <Indicator code="1.2" desc="Description of key impacts, risks, and opportunities" core="false" excelRowNumber="7"></Indicator>
          </indicators>
        </Section>
        <Section number="2" name="organizationProfile" desc="Organizational Profile">
          <indicators>
            <Indicator code="2.1" desc="Name of the organization" core="true" excelRowNumber="10"></Indicator>
            <Indicator code="2.2" desc="Primary brands, products, and/or services" core="true" excelRowNumber="11"></Indicator>
            <Indicator code="2.3" desc="Operational structure of the organization, including main divisions, operating companies, subsidiaries, and joint ve" core="true" excelRowNumber="12"></Indicator>
            <Indicator code="2.4" desc="Location of organization's headquarters" core="true" excelRowNumber="13"></Indicator>
            <Indicator code="2.5" desc="Number of countries where the organization operates, and names of countries with either major operations or that are" core="true" excelRowNumber="14"></Indicator>
            <Indicator code="2.6" desc="Nature of ownership and legal form" core="true" excelRowNumber="15"></Indicator>
            <Indicator code="2.7" desc="Markets served (including geographic breakdown, sectors served, and types of customers/beneficiaries)" core="true" e" core="true" excelRowNumber="16"></Indicator>
            <Indicator code="2.8" desc="Scale of the reporting organization" core="true" excelRowNumber="17"></Indicator>
            <Indicator code="2.9" desc="Significant changes during the reporting period regarding size, structure, or ownership" core="true" excelRowNumber="18"></Indicator>
            <Indicator code="2.10" desc="Awards received in the reporting period" core="true" excelRowNumber="19"></Indicator>
          </indicators>
        </Section>
        <Section number="3" name="reportParameters" desc="Report Parameters">
          <indicators>
            <Indicator code="3.1" desc="Reporting period (e.g., fiscal/calendar year) for information provided" core="true" excelRowNumber="22"></Indicator>
            <Indicator code="3.2" desc="Date of most recent previous report (if any)" core="true" excelRowNumber="23"></Indicator>
            <Indicator code="3.3" desc="Reporting cycle (annual, biennial, etc.)" core="true" excelRowNumber="24"></Indicator>
            <Indicator code="3.4" desc="Contact point for questions regarding the content of the report" core="true" excelRowNumber="25"></Indicator>
          </indicators>
        </Section>
      </sections>
    </Dimension>
  </dimensions>
</framework>
```

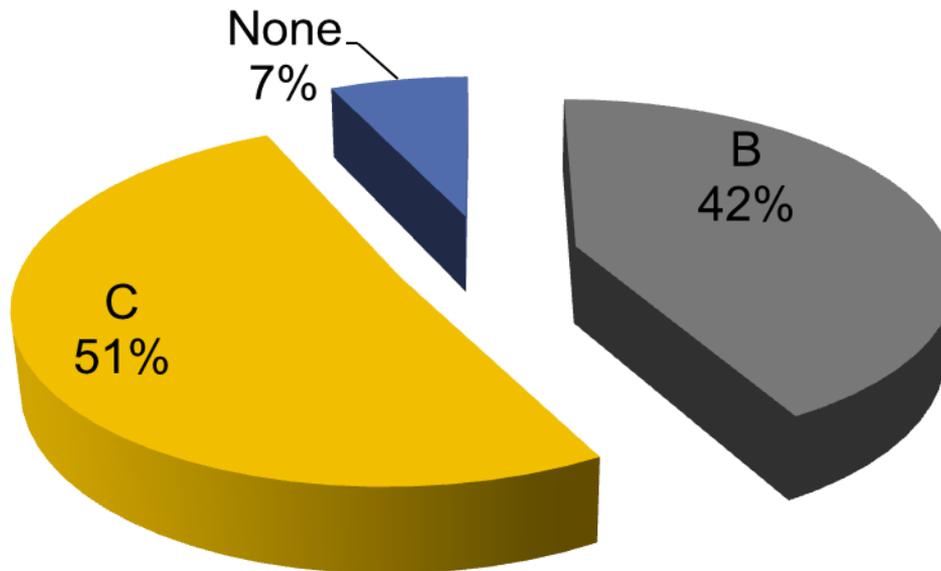
RESULTING EXCEL DOCUMENT

| C6 | | | | | |
|--------------------------|---|----------|-------------------------------|---|---------------------|
| A | B | C | D | E | F |
| 97 Economic | | | | | |
| 98 Performance Indicator | Description | Reported | Cross-reference/Direct answer | If applicable, indicate the part not reported | Reason for Omission |
| 99 Economic performance | | | | | |
| 00 EC1 | Direct economic value generated and distributed, including revenues, operating costs, employee compensation, donations and other community investments, retained earnings, and payments to capital providers and governments. | Yes | Page 14 | | |
| 01 EC2 | Financial implications and other risks and opportunities for the organization's activities due to climate change. | Yes | Page 14 | | |
| 02 EC3 | Coverage of the organization's defined benefit plan obligations. | | | | |
| 03 EC4 | Significant financial assistance received from government. | Yes | Page 14 | | |
| 04 Market presence | | | | | |

| A1 Declared Application level | | |
|-------------------------------|---|-----------------------|
| | A | B |
| 1 | Declared Application level | A |
| 2 | Recieved Application level | C |
| 3 | Status | Not Checked |
| 4 | Company Size | SME |
| 5 | Listed Company | No |
| 6 | Organization Type | Non-Profit |
| 7 | Sector | Non-Profit / Services |
| 8 | Suppliments | Used |
| 9 | Region | Northern America |
| 10 | Number of Disclosed Items (NDI) | 10 |
| 11 | Number of Required Disclosures for Selected Application Level (NRDAL) | 49 |
| 12 | Total Number of Required Disclosure Items (TNRDI) | 49 |

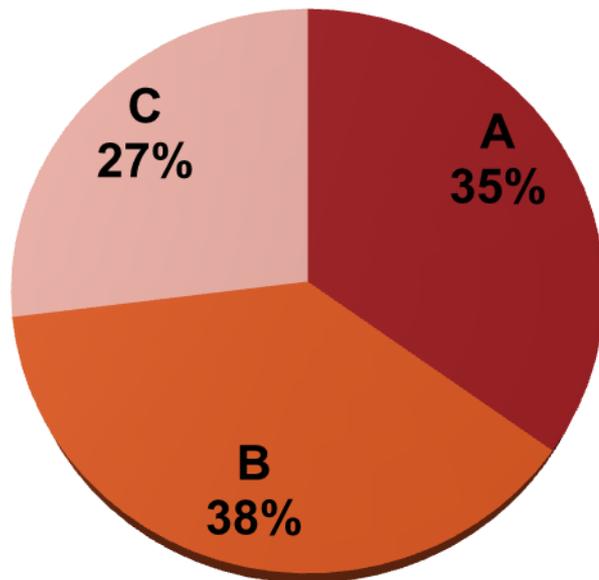
HOW DID WE DO?

After analyzing 100 reports, we had enough data to compare the produced scores with actual scores using Pearson Correlation theorem (r).

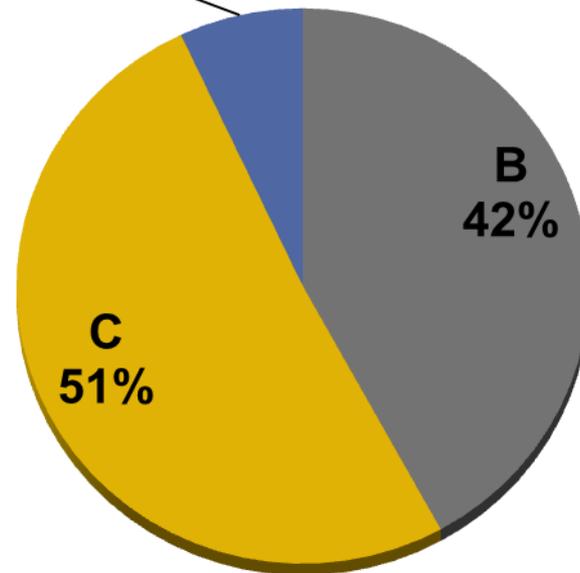


PUTTING THEM SIDE-BY-SIDE

Actual Scores



Calculated Scores



$(r) = 0.531$

Significant at 0.01

The correlation of 0.531 is considered 'moderate'.

DISCLOSURE DISCOVERY

Using Recall equation, the accuracy of discovering disclosures in reports was calculated on 25 externally assured reports.

$$R_i = \frac{TP_i}{TP_i + FN_i}$$

TP_i (or True Positives): Number of correctly identified disclosures

FN_i (or False Negatives): Number of not-identified information disclosures in report *i*.

| Dimension | No. of disclosure items | N | TP ¹ | FN ² | Recall (R) |
|---------------|-------------------------|----|-----------------|-----------------|------------|
| Economic | 7 | 25 | 129 | 46 | 73.71% |
| Environmental | 17 | 25 | 391 | 34 | 92.00% |
| Social | 25 | 25 | 485 | 140 | 77.60% |

LIMITATIONS

- Needs chapter start and end page numbers from user
- Unable to fetch external-to-report data for classification
- Does not analyze Profile Disclosure and DMA sections

THANK YOU!

